IBM Desktop Data Collector v7.5.1 documentation

Welcome to the IBM® Desktop Data Collector documentation, where you can find information about how to deploy, maintain, and use IBM Desktop Data Collector.

Getting started

Overview of the IBM StoredIQ components

Deployment of IBM StoredIQ Platform

IBM Desktop Data Collector hardware and software requirements

Common tasks

Configuring desktop settings

Utilizing desktop collection

IBM Desktop Data Collector installation

Configuring IBM Desktop Data Collector

Downloading the IBM Desktop Data Collector installer from the application

Troubleshooting and support

Delete policy: special notes

Support for IBM Desktop Data Collector

IBM Desktop Data Collector fixes

IBM Software Support home page

More information

Information lifecycle governance

Enterprise content management

Table of contents

IBM StoredIQ Platform overview

IBM® StoredIQ Platform provides scalable analysis and governance of unstructured data in-place across disparate and distributed email, file shares, desktops, and collaboration sites. Its products enable companies to discover, analyze, and act on data for eDiscovery; records retention and disposition; compliance; and storage optimization initiatives.

Powerful solutions for managing unstructured data in-place

IBM® Unstructured Data Identification and Management addresses the problems that challenge records management, electronic discovery, compliance, storage optimization, and data migration initiatives. By providing an in-depth assessment of unstructured data where it is, this software gives organizations visibility into data to make more informed business and legal decisions.

Unstructured Data Identification and Management delivers:

- In-place data management that allows an organization to discover, recognize, and act on unstructured data without moving it to a repository or specialty application.
- A powerful search function that accelerates the understanding of large amounts of unstructured content.
- Simplified analysis of large amounts of corporate data to provide detailed analysis faster and limit the impact on user productivity by analyzing and managing data in-place.
- Intelligence that supports many different policy actions such as copy, delete, move, copy to retention, or export.

An organized, systemic, and defensible approach to eDiscovery

IBM® eDiscovery Identification and Collection provides insight into enterprise data to help ease the costs and efforts that are involved in electronic discovery (eDiscovery) response. This software helps decrease the volume of unstructured data by targeting only the most relevant information to a particular case and providing forensically sound and defensible collections.

eDiscovery Identification and Collection delivers:

- Faster access to relevant information before collection, giving legal and IT teams the data that is needed to make more informed legal decisions.
- A powerful search function that accelerates the understanding of large amounts of unstructured content and encourages organizational alignment that can lead to reduced legal risks and costs.
- Simplified analysis of large amounts of corporate electronically stored information (ESI), providing faster detailed analysis, and limiting the impact on user productivity.
- Intelligence that allows companies to respond more quickly to litigation with the most relevant data.

Information governance to automate policy and compliance across unstructured data

IBM® Policy Assessment and Compliance helps organizations identify, classify, and manage enterprise information according to business value to reduce risk and cost. Corporations can gain a deeper and holistic understanding of their unstructured data to address business and regulatory requirements, compliance enforcement, data retention and respond to audit requests.

Policy Assessment and Compliance provides the following features and solutions:

- A powerful data assessment solution for discovering, recognizing, and acting on unstructured data without first moving it to a repository.
- Advanced search capabilities that are tailored to help legal, records, compliance, and IT staff discover data in accordance with corporate and regulatory policy.
- Detailed data analysis to simplify the analysis of large amounts of corporate data.
- In-place data management capabilities to remediate regulatory and corporate policy violations.

Flexible solution for identifying and collecting data from remote devices

IBM® Desktop Data Collector enables organizations to apply corporate governance policies to user desktops and notebooks. Users can identify corporate records or custodian data for legal matters and collect them to a central repository built for devices with intermittent connectivity.

Desktop Data Collector delivers:

- A powerful, flexible solution for identifying and collecting data for investigations, litigation matters, or records retention.
- A simplified collection of information on remote desktops and notebooks.
- Centralized management to minimize IT burden and improve efficiency when used with IBM eDiscovery Identification and Collection or IBM Policy Assessment and Compliance.
- Intelligent desktop data collection for identifying corporate records or custodian data and collecting them to a central repository.

- Overview of the IBM StoredIQ components
  The following sections describe three components of the IBM StoredIQ solution and the interfaces that IBM StoredIQ products provide to use the services of the IBM StoredIQ Platform.

Overview of the IBM StoredIQ components

The following sections describe three components of the IBM® StoredIQ solution and the interfaces that IBM StoredIQ products provide to use the services of the IBM StoredIQ Platform.

- [Components of the IBM solution](#)
  The three components of the IBM StoredIQ solution are the gateway, data servers, and application stack (AppStack).
- [IBM StoredIQ interface components](#)
  The IBM StoredIQ product components include IBM StoredIQ Administrator, IBM StoredIQ Data Workbench, IBM StoredIQ eDiscovery, IBM StoredIQ Data Script, IBM Stored IQ Policy Manager, and IBM Desktop Data Collector.
- [Notices](#)

Components of the IBM solution

The three components of the IBM® StoredIQ solution are the gateway, data servers, and application stack (AppStack).

The Gateway
> The gateway communicates between the data servers and the application stack. The application stack polls the gateway for information about the data on the data servers. The data servers push the information to the gateway.

Data Servers
> The IBM StoredIQ Data Server helps understand the data landscape of the enterprise. It obtains the data from supported data sources and indexes it. By indexing this data, you gain information about unstructured data such as file size, file data types, file owners. The data servers push the information about volumes and indexes to the gateway so it can be communicated to the application stack. Multiple data servers feed into a single gateway. In addition to an administrator user interface, administrators can deploy the IBM Desktop Data Collector and index desktops from the data server.

Application stack
> The application stack provides the user interface for the IBM StoredIQ Administrator, IBM StoredIQ Data Workbench, IBM StoredIQ eDiscovery, IBM StoredIQ Data Script, and the IBM Stored IQ Policy Manager products.

- **Parent topic:** [Overview of the IBM StoredIQ components](#)

IBM StoredIQ interface components

The IBM® StoredIQ product components include IBM StoredIQ Administrator, IBM StoredIQ Data Workbench, IBM StoredIQ eDiscovery, IBM StoredIQ Data Script, IBM Stored IQ Policy Manager, and IBM Desktop Data Collector.

IBM StoredIQ Administrator
> IBM StoredIQ Administrator monitors and manages the distributed infrastructure at a client site. IBM StoredIQ Administrator sits between the IBM StoredIQ Platform interface and the applications and facilitates the transfer and communication of information. IBM StoredIQ Administrator understands and manages IBM StoredIQ Platform concepts such as volumes, indexes, harvests, and configurations. At the same time, it manages the application concerns such as infoset lifecycle and creation, volume configuration, and action and target-set management. To this end, it is divided into two sections, that is, platform and application, so that the administrators know where to accomplish a task.

IBM StoredIQ Data Workbench
> With IBM StoredIQ Data Workbench, you can visualize the indexed data and identify potential red-flag issues to know how much and what types of data you have on different types of servers. It alerts people about potentially interesting or useful data. It helps ensure that the data of an enterprise is an asset, not a liability.

IBM StoredIQ eDiscovery
> IBM StoredIQ eDiscovery is integral to the eDiscovery process. It helps discover, identify, refine, preserve, and collect data that is relevant to legal matters. With IBM StoredIQ eDiscovery, you can conduct discovery work more efficiently and ensure that you capture the appropriate data at the same time.

IBM StoredIQ Data Script
> IBM StoredIQ Data Script automates execution within IBM StoredIQ Platform. Therefore, you can script, automate, and monitor processes that otherwise normally are manual processes that are run within IBM StoredIQ Data Workbench. IBM StoredIQ Data Script focuses on repeatable, understood, and approved processes for the purposes of culling and refining data in an approved manner.

IBM Stored IQ Policy Manager
> IBM Stored IQ Policy Manager acts on data in an automatic fashion at scale, running policies that affect data objects without requiring review.

IBM Desktop Data Collector
> IBM Desktop Data Collector deploys from the IBM StoredIQ Platform Data Server Administrator interface. It indexes desktops as volumes. The volumes appear in the Data Server Administrator interface and IBM StoredIQ Data Workbench, where the data can be analyzed and acted upon, and the IBM StoredIQ Administrator application.

IBM StoredIQ Platform uses a non-invasive deployment, specifically:

- Lightweight client deployment with no browser plug-ins and no client-side UI installations. All major browsers, such as Internet Explorer and Firefox, are supported natively.
- One (common) index that is shared across all use cases. This index supports a wide variety of data sources and is used by application dashboards for many use cases.
- No agents are placed on data sources. The native API/protocol is used whenever it is possible. Third-party or custom connectors are used only when necessary.
  Note: Agents are used for desktop collection.

- IBM StoredIQ Administrator
  IBM StoredIQ Administrator helps you to manage global assets common to the distributed infrastructure behind IBM StoredIQ Platform applications.
- IBM StoredIQ Data Workbench
  The following section provides a high-level explanation of what IBM StoredIQ Data Workbench is and its potential uses.
- IBM StoredIQ eDiscovery
  The following section provides a high-level explanation of what IBM StoredIQ eDiscovery

is, by whom it is used, prerequisites, and its potential uses.

- IBM StoredIQ Data Script
  Through the IBM StoredIQ Data Script interface, you can monitor each of the steps as they are run. Then, you view any defined outputs, such as reports, infosets, or exports that are generated as a result. By running IBM StoredIQ Data Workbench workflows, the user can reapply processes that were reviewed and approved to apply to a wide variety of different data problems.
- IBM Stored IQ Policy Manager
  IBM Stored IQ Policy Manager enables policy execution at scale.

- **Parent topic:** Overview of the IBM StoredIQ components

IBM StoredIQ Administrator

IBM® StoredIQ Administrator helps you to manage global assets common to the distributed infrastructure behind IBM StoredIQ Platform applications.

IBM StoredIQ Administrator provides at-a-glance understanding of the different issues that can crop up in the IBM StoredIQ Platform environment. These views are unique to the IBM StoredIQ Administrator application as they provide an overview of how the system is running. They allow access to various pieces of information that are being shared across applications or allow for the management of resources in a centralized manner.

The administrator is the person responsible for managing the IBM StoredIQ Platform. This individual has strong understanding of data sources, indexes, data servers, jobs, infosets, and actions. This list provides an overview as to how IBM StoredIQ Administrator works:

- Viewing data servers and volumes: Using IBM StoredIQ Administrator, the Administrator can identify what data servers are deployed, their location, what data is being managed, and the status of each data server in the system. Volume management is a central component of IBM StoredIQ Platform. IBM StoredIQ Administrator also allows the Administrator to see what volumes are currently under management, which data server is responsible for that volume, the state of the volume after indexing, and the amount and size of information that is contained by each volume. Administrators can also add volumes to data servers through this interface.
- Creating system infosets: System infosets that use only specific indexed volumes can be created and managed within IBM StoredIQ Administrator. Although infosets are a core component of IBM StoredIQ Data Workbench, system infosets are created as a shortcut for users in IBM StoredIQ Administrator.
- Managing users: The user management area allows administrators to create users and manage users' access to the various IBM StoredIQ applications.
- Configuring and managing actions: An action is any process that is taken upon the data that is represented by the indexes. Actions are run by data servers on indexed data objects. Any errors or warnings that are generated as a result of an action are recorded as exceptions in IBM StoredIQ Data Workbench.
  Note: Actions can be created within IBM StoredIQ Administrator and then made available to other IBM StoredIQ applications such as IBM StoredIQ Data Workbench.
- Managing target sets: Provides an interface that allows the user to set the wanted targets for specific actions that require a destination volume for their actions.
- Managing concepts: You can relate business concepts to indexed data.

- **Parent topic:** [IBM StoredIQ interface components](#)

IBM StoredIQ Data Workbench

The following section provides a high-level explanation of what IBM® StoredIQ Data Workbench is and its potential uses.

Big data is a pervasive problem, not a one-time occurrence. It is easy for most companies to realize that big data is problematic, but it is hard to identify what problems they have. Big data is all about the unknown, but the unknown cannot be off limits. IBM StoredIQ Data Workbench can help you learn about your data, make educated decisions with your most valuable asset, and turn your company's most dangerous risk into its most valuable asset.

IBM StoredIQ Data Workbench is a data visualization and management tool that helps you to actively manage your company's data. It helps you to determine how much data you have, where it is, who owns it, and when it was last used. When you have a clear understanding of your company's data landscape, IBM StoredIQ Data Workbench helps you take control of data. You can make informed decisions about your data and act on that knowledge by copying, copying to retention, or conducting a discovery export.

Here are just some of the examples that you can use IBM StoredIQ Data Workbench.

- You need to find all company email that is sent from or received by Eileen Sideways (esideways@thecompany.com). You can use IBM StoredIQ Data Workbench to find all email and then copy that data to a predefined repository. You can also use IBM StoredIQ Data Workbench to find all of the esideways@thecompany.com email that occurred between specific dates and then make that email available for review.
- As an administrator, you want to rid your networks and storage of unused data. You can use IBM StoredIQ Data Workbench to find all files that were not modified in more than five years.
- You want to find all image files that are created in 2007. Not only can IBM StoredIQ Data Workbench find all image files that were created in 2007. It also shows how much space they occupy on your network.
- A user needs to understand how data about Windows is being retained. Using IBM StoredIQ Data Workbench, you can provide that user with a visual overview of the number of objects that are retained and a breakdown of files per data source. Additionally, you can apply overlays to show the user if those files contain forbidden information such as credit-card numbers or Social Security numbers.

- IBM StoredIQ Data Workbench support

- **Parent topic:** IBM StoredIQ interface components

IBM StoredIQ Data Workbench support

IBM® StoredIQ Platform helps you to truly understand your company's data landscape. Using IBM StoredIQ Platform data server, all unstructured data in a company's enterprise network is indexed. By indexing this data, you gain information about unstructured data such as file size, file data types, file owners.

IBM StoredIQ Data Workbench is the tool that you use to visualize this indexed data, helping you to identify potential red-flag issues, to know how much data you have on different servers, and to alert people about potentially interesting or useful data.

- **Parent topic:** [IBM StoredIQ Data Workbench](#)

IBM StoredIQ eDiscovery

The following section provides a high-level explanation of what IBM® StoredIQ eDiscovery is, by whom it is used, prerequisites, and its potential uses.

IBM StoredIQ eDiscovery is a user application that helps legal users during the initial phases of the eDiscovery process. By helping you to capture needed electronic data, you can use IBM StoredIQ eDiscovery to communicate easily with other users about the status and review process of work that is being done. IBM StoredIQ eDiscovery does not drive the eDiscovery process, but instead helps legal users to control and communicate those processes more effectively.

Here are just some of the examples you can use IBM StoredIQ eDiscovery.

- You need to find all electronic information about an upcoming personal-injury lawsuit. You can use IBM StoredIQ eDiscovery to create a matter for the suit, and then create boxes to contain email and reports about the case.
- As a legal user, you want to see the status of all currently active matters. You can use the Matter Dashboard to visualize the different matters' states.
- As a data expert, you know that you need to respond to users in a timely fashion. Using IBM StoredIQ eDiscovery, people can email you directly from the application about a matter's questions.

Proactive eDiscovery consists of identifying and cataloging data or types of data that might eventually be responsive to legal matters. It collects and retains the data and then produces that data in a way that matches accepted or required legal practices.

Within IBM StoredIQ eDiscovery, the user is a non-technical user who needs to obtain certain pieces of data to complete a legal process. The IBM StoredIQ eDiscovery user knows the information or parameters for what kinds of data they want, but does not always know how to obtain that data.

Before you use IBM StoredIQ eDiscovery, ensure that prerequisites are met:

- The IBM StoredIQ Platform must be deployed, configured, and ready for use.
- IBM StoredIQ eDiscovery is dependent upon other IBM StoredIQ applications such as IBM StoredIQ Data Workbench and IBM StoredIQ Administrator.

- **Parent topic:** IBM StoredIQ interface components

IBM StoredIQ Data Script

Through the IBM® StoredIQ Data Script interface, you can monitor each of the steps as they are run. Then, you view any defined outputs, such as reports, infosets, or exports that are generated as a result. By running IBM StoredIQ Data Workbench workflows, the user can reapply processes that were reviewed and approved to apply to a wide variety of different data problems.

- **Parent topic:** [IBM StoredIQ interface components](#)

IBM Stored IQ Policy Manager

IBM® Stored IQ Policy Manager enables policy execution at scale.

IBM Stored IQ Policy Manager allows users to run mature policies and processes at scale across a wider range of data. The users can define and run systemwide policies, focusing on the execution of the process rather than understanding or reviewing affected data objects. Additionally, with IBM Stored IQ Policy Manager's reports, you can record what actions were conducted, when they were conducted, and what data was affected by the policy's execution.

- **Parent topic:** [IBM StoredIQ interface components](#)

Notices

This information was developed for products and services offered in the U.S.A. This material may be available from IBM in other languages. However, you may be required to own a copy of the product or product version in that language in order to access it.

IBM may not offer the products, services, or features discussed in this document in other countries. Consult your local IBM representative for information on the products and services currently available in your area. Any reference to an IBM product, program, or service is not intended to state or imply that only that IBM product, program, or service may be used. Any functionally equivalent product, program, or service that does not infringe any IBM intellectual property right may be used instead. However, it is the user's responsibility to evaluate and verify the operation of any non-IBM product, program, or service.

IBM may have patents or pending patent applications covering subject matter described in this document. The furnishing of this document does not grant you any license to these patents. You can send license inquiries, in writing, to:
IBM Director of Licensing
IBM Corporation
North Castle Drive
Armonk, NY 10504-1785
U.S.A.
For license inquiries regarding double-byte (DBCS) information, contact the IBM Intellectual Property Department in your country or send inquiries, in writing, to:
Intellectual Property Licensing
Legal and Intellectual Property Law
IBM Japan Ltd.
19-21, Nihonbashi-Hakozakicho, Chuo-ku
Tokyo 103-8510, Japan

The following paragraph does not apply to the United Kingdom or any other country where such provisions are inconsistent with local law: INTERNATIONAL BUSINESS MACHINES CORPORATION PROVIDES THIS PUBLICATION "AS IS" WITHOUT WARRANTY OF ANY KIND, EITHER EXPRESS OR IMPLIED, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF NON-INFRINGEMENT, MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE. Some states do not allow disclaimer of express or implied warranties in certain transactions, therefore, this statement may not apply to you.

This information could include technical inaccuracies or typographical errors. Changes are periodically made to the information herein; these changes will be incorporated in new editions of the publication. IBM may make improvements and/or changes in the product(s) and/or the program(s) described in this publication at any time without notice.

Any references in this information to non-IBM Web sites are provided for convenience only and do not in any manner serve as an endorsement of those Web sites. The materials at those Web sites are not part of the materials for this IBM product and use of those Web sites is at your own risk.

IBM may use or distribute any of the information you supply in any way it believes appropriate without incurring any obligation to you.

Licensees of this program who wish to have information about it for the purpose of enabling: (i) the exchange of information between independently created programs and other programs (including this one) and (ii) the mutual use of the information which has been exchanged, should contact:
IBM Corporation
J46A/G4
555 Bailey Avenue
San Jose, CA 95141-1003
U.S.A.

Such information may be available, subject to appropriate terms and conditions, including in some cases, payment of a fee.

The licensed program described in this document and all licensed material available for it are provided by IBM under terms of the IBM Customer Agreement, IBM International Program License Agreement or any equivalent agreement between us.

Any performance data contained herein was determined in a controlled environment. Therefore, the results obtained in other operating environments may vary significantly. Some measurements may have been made on development-level systems and there is no guarantee that these measurements will be the same on generally available systems. Furthermore, some measurements may have been estimated through extrapolation. Actual results may vary. Users of this document should verify the applicable data for their specific environment.

Information concerning non-IBM products was obtained from the suppliers of those products, their published announcements or other publicly available sources. IBM has not tested those products and cannot confirm the accuracy of performance, compatibility or any other claims related to non-IBM products. Questions on the capabilities of non-IBM products should be addressed to the suppliers of those products.

All statements regarding IBM's future direction or intent are subject to change or withdrawal without notice, and represent goals and objectives only.

This information contains examples of data and reports used in daily business operations. To illustrate them as completely as possible, the examples include the names of individuals, companies, brands, and products. All of these names are fictitious and any similarity to the names and addresses used by an actual business enterprise is entirely coincidental.

COPYRIGHT LICENSE:

This information contains sample application programs in source language, which illustrate programming techniques on various operating platforms. You may copy, modify, and distribute these sample programs in any form without payment to IBM, for the purposes of developing, using, marketing or distributing application programs conforming to the application programming interface for the operating platform for which the sample programs are written. These examples have not been thoroughly tested under all conditions. IBM, therefore, cannot guarantee or imply reliability, serviceability, or function of these programs. The sample programs are provided "AS IS", without warranty of any kind. IBM shall not be liable for any damages arising out of your use of the sample programs.

Each copy or any portion of these sample programs or any derivative work, must include a copyright notice as follows: © (your company name) (year). Portions of this code are derived from IBM Corp. Sample Programs. © Copyright IBM Corp. 2004, 2010. All rights reserved.

If you are viewing this information softcopy, the photographs and color illustrations may not appear.

- Privacy policy considerations
- Terms and conditions for product documentation
  Permissions for the use of these publications are granted subject to the following terms and conditions.
- Trademarks

- **Parent topic:** Overview of the IBM StoredIQ components

Privacy policy considerations

IBM Software products, including software as a service solutions, ("Software Offerings") may use cookies or other technologies to collect product usage information, to help improve the end user experience, to tailor interactions with the end user or for other purposes. In many cases no personally identifiable information is collected by the Software Offerings. Some of our Software Offerings can help enable you to collect personally identifiable information. If this Software Offering uses cookies to collect personally identifiable information, specific information about this offering's use of cookies is set forth below.

[Depending upon the configurations deployed, this Software Offering may use [session] [and] [persistent] cookies that collect each user's [name,] [user name,] [password,] [profile name,] [position,] or [other personally identifiable information] for purposes of [session management,] [authentication,] [enhanced user usability,] [single sign-on configuration] [or other usage tracking or functional purposes.] These cookies [cannot be disabled] [can be disabled, but disabling them will also [likely] eliminate the functionality they enable].

If the configurations deployed for this Software Offering provide you as customer the ability to collect personally identifiable information from end users via cookies and other technologies, you should seek your own legal advice about any laws applicable to such data collection, including any requirements for notice and consent.

For more information about the use of various technologies, including cookies, for these purposes, See IBM's Privacy Policy at http://www.ibm.com/privacy and IBM's Online Privacy Statement at http://www.ibm.com/privacy/details the section entitled "Cookies, Web Beacons and Other Technologies" and the "IBM Software Products and Software-as-a-Service Privacy Statement" at http://www.ibm.com/software/info/product-privacy.

- **Parent topic:** Notices

Terms and conditions for product documentation

Permissions for the use of these publications are granted subject to the following terms and conditions.

Applicability

These terms and conditions are in addition to any terms of use for the IBM website.

Personal use

You may reproduce these publications for your personal, noncommercial use provided that all proprietary notices are preserved. You may not distribute, display or make derivative work of these publications, or any portion thereof, without the express consent of IBM.

Commercial use

You may reproduce, distribute and display these publications solely within your enterprise provided that all proprietary notices are preserved. You may not make derivative works of these publications, or reproduce, distribute or display these publications or any portion thereof outside your enterprise, without the express consent of IBM.

Rights

Except as expressly granted in this permission, no other permissions, licenses or rights are granted, either express or implied, to the publications or any information, data, software or other intellectual property contained therein.

IBM reserves the right to withdraw the permissions granted herein whenever, in its discretion, the use of the publications is detrimental to its interest or, as determined by IBM, the above instructions are not being properly followed.

You may not download, export or re-export this information except in full compliance with all applicable laws and regulations, including all United States export laws and regulations.

IBM MAKES NO GUARANTEE ABOUT THE CONTENT OF THESE PUBLICATIONS. THE PUBLICATIONS ARE PROVIDED "AS-IS" AND WITHOUT WARRANTY OF ANY KIND, EITHER EXPRESSED OR IMPLIED, INCLUDING BUT NOT LIMITED TO IMPLIED WARRANTIES OF MERCHANTABILITY, NON-INFRINGEMENT, AND FITNESS FOR A PARTICULAR PURPOSE.

- **Parent topic:** Notices

Trademarks

IBM, the IBM logo, and ibm.com are trademarks or registered trademarks of International Business Machines Corp., registered in many jurisdictions worldwide. Other product and service names might be trademarks of IBM or other companies. A current list of IBM trademarks is available on the Web at "Copyright and trademark information" at http://www.ibm.com/legal/copytrade.shtml

Adobe, the Adobe logo, PostScript, and the PostScript logo are either registered trademarks or trademarks of Adobe Systems Incorporated in the United States, and/or other countries.

Linux is a registered trademark of Linus Torvalds in the United States, other countries, or both.

Microsoft, Windows, Windows NT, and the Windows logo are trademarks of Microsoft Corporation in the United States, other countries, or both.

Java and all Java-based trademarks and logos are trademarks or registered trademarks of Oracle and/or its affiliates.

UNIX is a registered trademark of The Open Group in the United States and other countries.

Other company, product, and service names may be trademarks or service marks of others.

- **Parent topic:** Notices

Deploying and configuring IBM StoredIQ Platform

Before you deploy and configure IBM® StoredIQ Platform, ensure that you meet all deployment prerequisites and requirements that are documented in the Deployment preparation chapter. You must also meet all hardware and software requirements in IBM StoredIQ Platform Overview Guide.

- IBM StoredIQ components
  The following sections describe three components of the IBM StoredIQ solution and the interfaces that IBM StoredIQ product provides to use the services of the IBM StoredIQ Platform.
- Deployment preparation
  To prevent any deployment problems, conduct the following planning activities that are appropriate for your environment.
- Deployment of IBM StoredIQ Platform
  The following sections describe how to deploy the virtual appliances.
- Installation of IBM StoredIQ Platform
  After the OVF deployment, proceed to install the IBM StoredIQ Platform.
- Configuration of IBM StoredIQ Platform
  After you deploy virtual appliances, you must configure IBM StoredIQ Platform before you can fulfill the product installation process.
- Creating a template from the IBM StoredIQ OVA
  OVAs can be modified to fit unique deployment scenarios.
- Editing the OVF file
  If you need to scale down the size of the index drive on the data servers, you can edit the OVF file.
- Upgrades
  You can obtain the upgrade ISO from the IBM Fix Central. Upgrades consist of upgrading the gateway and data server first and then the application stack.
- Customizable reports
  Reporting is a key step within the data-management process as it validates that processes were completed correctly within IBM StoredIQ.

IBM StoredIQ components

The following sections describe three components of the IBM® StoredIQ solution and the interfaces that IBM StoredIQ product provides to use the services of the IBM StoredIQ Platform.

- Components of the IBM StoredIQ solution
  The three components of the IBM StoredIQ solution are the Gateway, Data Servers, and Application Stack (AppStack).
- IBM StoredIQ products
  The IBM StoredIQ products include IBM StoredIQ Administrator, IBM StoredIQ Data Workbench, IBM StoredIQ eDiscovery, IBM StoredIQ Data Script, IBM Stored IQ Policy Manager, and IBM Desktop Data Collector.

Components of the IBM StoredIQ solution

The three components of the IBM® StoredIQ solution are the Gateway, Data Servers, and Application Stack (AppStack).

The Gateway
> The gateway communicates between the data servers and the application stack. The application stack polls the gateway for information about the data on the data servers. The data servers push the information to the gateway.

Data Servers
> IBM StoredIQ Platform Data Server helps understand the data landscape of the enterprise. It obtains the data from supported data sources and indexes it. By indexing this data, you gain information about unstructured data such as file size, file data types, and file owners.
>
> The data servers push the information about volumes and indexes to the gateway so it can be communicated to the application stack. Multiple data servers feed into a single gateway. In addition to an administrator user interface, administrators can deploy the IBM Desktop Data Collector and index desktops from the data server.

Application stack
> The application stack provides the user interface for the IBM StoredIQ Administrator, IBM StoredIQ Data Workbench, IBM StoredIQ eDiscovery, IBM StoredIQ Data Script, and the IBM Stored IQ Policy Manager products.

- **Parent topic:** IBM StoredIQ components

IBM StoredIQ products

The IBM® StoredIQ products include IBM StoredIQ Administrator, IBM StoredIQ Data Workbench, IBM StoredIQ eDiscovery, IBM StoredIQ Data Script, IBM Stored IQ Policy Manager, and IBM Desktop Data Collector.

IBM StoredIQ Administrator

IBM StoredIQ Administrator monitors and manages the distributed infrastructure at a client site. IBM StoredIQ Administrator sits between the IBM StoredIQ Platform interface and the applications and facilitates the transfer and communication of information.IBM StoredIQ Administrator understands and manages IBM StoredIQ Platform concepts such as volumes, indexes, harvests, and configurations. At the same time, it manages the application concerns such as infoset lifecycle and creation, data server and volume configuration, and action and target set management. To this end, it is divided into two sections, that is, platform and application so that the administrators know where to accomplish a task.

IBM StoredIQ Data Workbench

IBM StoredIQ Data Workbench enables visualizing the indexed data. You can identify potential red-flag issues to know how much and what types of data you have on different types of servers. It alerts people about potentially interesting or useful data.

IBM StoredIQ eDiscovery

IBM StoredIQ eDiscovery is integral to the eDiscovery process. It helps discover, identify, refine, preserve, and collect data that is relevant to legal matters. With IBM StoredIQ eDiscovery, you can conduct discovery work more efficiently and capture the appropriate data.

IBM StoredIQ Data Script

IBM StoredIQ Data Script enables automated execution within IBM StoredIQ Platform. It scripts, automates, and monitors processes that otherwise normally are a manual process that is run within IBM StoredIQ Data Workbench. IBM Data Script focuses on repeatable, understood, and approved processes for the purposes of culling and refining data in an approved manner.

IBM Stored IQ Policy Manager

IBM Stored IQ Policy Manager acts on data in an automatic fashion at scale, running policies that affect data objects without requiring review.

IBM Desktop Data Collector

The IBM Desktop Data Collector deploys from the IBM StoredIQ Data Server Administrator interface. It indexes desktops as volumes. The volumes appear in the Data Server Administrator interface, the IBM StoredIQ Data Workbench, where the data can be analyzed and acted upon, and theIBM StoredIQ Administrator application.

IBM StoredIQ Platform uses a non-invasive deployment, specifically:

- Lightweight client deployment with no browser plug-ins and no client-side UI installations. All major browsers, such as Internet Explorer and Firefox, are supported natively.
- One (common) index that is shared across all use cases. This index supports a wide variety of data sources and is used by application dashboards for many use cases.
- No agents are placed on data sources. The native API/protocol is used whenever it is possible. Third-party or custom connectors are used only when necessary.
  Note: Agents are used for desktop collection.

- **Parent topic:** [IBM StoredIQ components](#)

Deployment preparation

To prevent any deployment problems, conduct the following planning activities that are appropriate for your environment.

- [Deployment planning](#)
  When you plan a deployment of IBM® StoredIQ Platform, evaluate several infrastructure considerations.
- [Stack-provisioning prerequisites](#)
  Before a deployment, verify that you meet these prerequisites.

Deployment planning

When you plan a deployment of IBM® StoredIQ Platform, evaluate several infrastructure considerations.

- Open Virtual Appliance (OVA) configuration requirements
  IBM StoredIQ Platform is deployed as virtual appliances and currently only supported in VMware ESX(i) 5.x or greater environments. You must have a virtual infrastructure that meets theIBM StoredIQ Platform hardware requirements.
- Network and port requirements
  Data sources and types can drastically impact the scale and scope of what needs to be deployed. The complexity of the source directly affects the number of data servers to be deployed, for example, exchange versus simple text documents in a CIFS location.
- Environment sizing guidelines
  To size an environment precisely, you must understand the factors such as harvest frequency, complexity of the source, and use case scenarios that drive application use and action execution.

- **Parent topic:** Deployment preparation

Open Virtual Appliance (OVA) configuration requirements

IBM® StoredIQ Platform is deployed as virtual appliances and currently only supported in VMware ESX(i) 5.x or greater environments. You must have a virtual infrastructure that meets theIBM StoredIQ Platform hardware requirements.

Gateway server size

- vCPU: 1
- Memory: 8 GB
- Storage: 100 GB

Application stack size

- vCPU: 1
- Memory: 4 GB
- Storage: 20 GB

Data Server size

- vCPU: 4
  Note: Even though increasing the number of vCPUs increases performance, the actual benefits depend on whether the specific host is oversubscribed or not.
- Memory: 16 GB
  Note: While the minimum value works under light-load condition, as the load increases, the Data Server quickly starts consuming swap space. For high load situations, increasing RAM beyond 16 GB can benefit performance. Monitoring swap usage can provide insight.
- Storage:
  - Primary Disk (vmdisk1, SCSI 0:0): Default is 190 GB
    Note: This virtual disk has an associated VMDK that contains the IBM operating code. Do not change its size.
    Note: If you delete the primary disk, you delete the operating system, and the IBM software; the virtual machine might need to be scrapped.
  - Data Disk (vmdisk2, SCSI 0:1): Default is 1.9 TB
    Note: This virtual disk can be resized according to expectations on the amount of harvest data to be stored. For purposes of estimation, the index storage requirement for metadata is about 30 GB per TB of managed source data. Full-text indexing requires an extra 170 GB per TB. The default data disk size is therefore targeted for managing 10 TB of source information.
  - Swap Disk (vmdisk3, SCSI 0:2): Default is 40 GB
    Note: When under load, the Data Server can use many RAM; therefore, having ample swap space is prudent. The minimum swap size is equal to the amount of RAM configured for the virtual machine. For best performance under load, place this disk on the highest speed data store available to the host.

The general size limits for a data server are 150 million objects or 500 defined volumes, whichever limit is reached first. Assuming an average object size of 200 KB equals to about 30 TB of managed storage across 30 volumes of 5 million objects each, the index storage requirement for metadata on ~30 TB of storage that contains uncompressed general office documents is ~330 GB (11 GB per TB). Add 100 GB per TB of managed storage for full-text or snippet index. For example, to support 30 TB of storage that is indexed for metadata, you need 8 TB indexed for full-text search and extracted text (snippet cache) of 8 TB for auto-classification. A total of 1.9 TB of storage is required (metadata: 330 GB, full-text: 800 GB, snippet cache: 800 GB).

Data Server performance is impacted by the IOPS available from the storage subsystem. For each Data Server under maximum workload, at least 650 IOPS generally delivers acceptable performance. In the situations where there is a high load on the system, the

IOPS that is used can reach up to 7000 with main write operations.

- **Parent topic:** [Deployment planning](#)

Network and port requirements

Data sources and types can drastically impact the scale and scope of what needs to be deployed. The complexity of the source directly affects the number of data servers to be deployed, for example, exchange versus simple text documents in a CIFS location.

You must enable network connectivity from the following locations:

- The data server IP address to the gateway IP address on port 11103.
- The gateway IP address to and from the appstack IP address on ports 8765 and 5432.
- Ports 80, 443, and 22 from the administrative user's workstation (place from which the administrator is completing work with IBM® StoredIQ Administrator) to the application stack and data server IP addresses.
- Port 22 from the administrative workstation to the Gateway IP address.

- **Parent topic:** Deployment planning

Environment sizing guidelines

To size an environment precisely, you must understand the factors such as harvest frequency, complexity of the source, and use case scenarios that drive application use and action execution.

The general design guidelines for the IBM® StoredIQ release 7.x is as follows:

- One data server per 30 TBs of file shares. Varies depend on number of volumes, objects per volume, and object types
- One gateway per 50 data servers.
- One application server.
- NFS is slightly faster than CIFS for metadata only, but assume CIFS/NFS even for this exercise.
- Full-content processing of file (For example, .ZIP, .RAR, .GZ) and email archive (.PST, .NSF, .EMX) processing are slower as items must be extracted from the archives. If there is a significant number of these files in the file system and they are not excluded from content processing, the full-content processing rate can be too high. Until you have an initial index of the file system, you do not know how to weigh full-content processing of archives.
- An object/time metric is appropriate for metadata only NOT computing a hash, membership in the National Institute of Standards and Technology (NIST) or enumerating objects that are contained in archives. Converting it to a bytes/time metric is a function of the average object size and might vary tremendously. An average object size of 250 KB was used for the metric that is provided earlier.
- A bytes/time metric is appropriate for metadata-only computing a hash and full-content processing. The object per second rate can vary tremendously depending on the object type and sizes encountered. For example, processing an email or file archive is much more expensive than a PDF document.
- Metadata-only not computing a hash, membership in the NIST list, or enumerating objects that are contained in archives is requesting only the file-attribute information from the NAS. Individual files are not opened and read. The processing rate is high, but that does not translate into a large amount of data that traverses a network between the NAS and data server. The bytes/time rate does not translate into bytes served by the NAS and sent over the network.
- Metadata-only computing a hash, membership in the NIST list, or enumerating objects that are contained in archives opens and reads the contents of each file. The content of all requested files traverses the network between the NAS and data server. The maximum load that the data server can place on a NAS is metadata-only processing. It requires all file content to be read to compute a hash or enumerate objects that are contained in archives. The bytes/time rate translates into bytes served up by the NAS and network traffic that must be considered.
- Full-content processing opens and reads the contents of each file to extract all text. The content of all requested files traverses the network between the NAS and data server. The processing time to enumerate archives, extract text, index words, and extract entities on the data server reduces the rate that data is requested from a NAS compared to metadata-only with full hash. The bytes/time rate translates into bytes served up by the NAS and network traffic that must be considered.
- The interrogator process count on the data server for "metadata only not reading all content indexing" can be set to four for optimal performance.
- The interrogator process count for all other processing that involves reading all content is assumed to be two per data server.
- The interrogator count can be viewed as the number of client connections that are made to a data source actively requesting data. It is important for capacity planning for the data source.
- The data servers are assumed to be "network close" to the NAS data sources. Network latency under 10 ms with at least 1000 Mbps bandwidth is assumed (connected through local area network). The data servers need a low latency high-bandwidth connection to a NAS data source for acceptable indexing performance.
- The gateway and application stack can be located remotely from the data servers.

Network connections with latency greater than 10 ms and bandwidth of at least 2+ Mbps are acceptable.

VMware vSphere requirements

- VMware vSphere v5.x or greater.
- VMware virtual machine version 8.0 or later.
- VMware license to enable the required processor cores and memory for the virtual machine.

- **Parent topic:** [Deployment planning](#)

Stack-provisioning prerequisites

Before a deployment, verify that you meet these prerequisites.

- At least one physical server with sufficient processor, RAM, and hard disk configuration for the planned management project.
- VMware ESX or ESXi on CD/DVD or USB drive.
- IP addresses, cables, and physical switch ports for at least the ESXi/ESX interface, one data server, one gateway server, and one application stack.
- Network connectivity that is enabled from the following locations:
    - The data server IP address to the gateway IP address on port 11103
    - The gateway IP address to and from the application stack IP address on port 8765
    - Ports 80, 443, and 22 from the administrative workstation to the application stack and data server IP addresses
    - Port 22 from the administrative workstation to the gateway IP address.
- Network connectivity that is enabled from the data server IP address to any data sources to be harvested and managed.
- A management station computer or notebook from where the load-management work is done.

- **Parent topic:** [Deployment preparation](#)

Deployment of IBM® StoredIQ Platform

The following sections describe how to deploy the virtual appliances.

- Open Virtualization Format (OVF)
  An OVF is a file format that allows for the exchange of virtual appliances across products
  and platforms.
- Deploying the virtual appliances
  Use vSphere Client to deploy the virtual appliances to an ESX server. Three OVAs need to
  be deployed in this order: the gateway, data server, and application stack.

Open Virtualization Format (OVF)

An OVF is a file format that allows for the exchange of virtual appliances across products and platforms.

The OVF format offers these advantages:

- OVF files are compressed, allowing for faster downloads.
- The vSphere Client validates an OVF file before it imports it and ensures that it is compatible with the intended destination server. If the virtual appliance is incompatible with the selected host, it cannot be imported, and an error message appears.
- An OVF can encapsulate multitiered applications and more than one virtual machine.

The basic differences between OVFs and OVAs are as follows:

- An OVF is a collection of items in a single folder. Most commonly, it is a description file (.ovf), a manifest file (.mf), and virtual machine state files (*.vhd or *.vmdk).
- An OVA is a single file. The OVA is the OVF folder contents that are all compressed into a single file. The OVA is used when you want to take an OVF and share it, or give it as a download. The OVA needs to be opened into the OVF before it can be used, although it is not necessary when you use VMware. This step is automatically conducted during import. Both OVF and OVA files can be directly imported into vCenter and vSphere.

- **Parent topic:** [Deployment of IBM StoredIQ Platform](#)

Deploying the virtual appliances

Use vSphere Client to deploy the virtual appliances to an ESX server. Three OVAs need to be deployed in this order: the gateway, data server, and application stack.

Before you begin

Ensure that all prerequisites for the deployment are met before you start this task.

About this task

The number of data server OVAs deployed depends upon the number of data servers needed. Each data server requires a separate OVA deployment.

Procedure

1. Connect to ESX server or vCenter server.
2. From the File menu, select Deploy OVF Template.
3. Follow the Deploy OVF Template wizard.
4. On the Source page, complete these steps:
   a. Select File > Deploy.
   b. Browse and select the appropriate OVA, and then click Next.
5. Review the OVF Template Details.
   Note: The storage requirements are critical for later selecting a data store in the deployment.
6. Click Next.
7. On the Name and Location page, name the deployed template or use the default name.
8. On the Host/Cluster page, select the host and cluster where the deployed OVF template runs, and then click Next.
9. On the Resource Pool page, select the resource pool where the deployed OVF template runs, and then click Next.
10. On the Storage page, select a data store on which to store the deployed OVF template files, and then click Next.
11. On the Disk Format page, select the disk format to be deployed.
    Note: Though Thin Provision saves disk space, it can negatively affect performance. If possible, select Thick Provision Lazy Zeroed.
12. Click Next.
13. For Network Mapping, map the network to the appropriate network in the list, and then click Next.
14. On the Ready to Complete page, review the deployment settings and then click Finish.

- **Parent topic:** [Deployment of IBM StoredIQ Platform](#)

Installation of IBM StoredIQ Platform

After the OVF deployment, proceed to install the IBM® StoredIQ Platform.

Note: You can install IBM Desktop Data Collector only after you finish configuring the data servers. For information about installation of IBM Desktop Data Collector, see Installing IBM Desktop Data Collector

- Installing the gateway and data servers
  To install the gateway and data servers, you must turn on the virtual machine first.

Installing the gateway and data servers

To install the gateway and data servers, you must turn on the virtual machine first.

Procedure

1. Start the console for the virtual machine in vCenter.
2. Select I Agree to accept the CENTOS license agreement and then press Enter.
3. Go to I Agree to accept the IBM license agreement.
4. From the Corporate Network Configuration screen, select either Static or Obtain IP via DCHP.
   - To configure the static address,
     a. Gather this information to complete the static IP configuration for the system: Host name, Public Network IP, Netmask, Gateway, and Primary DNS Host.
     b. Go to the Hostname field and enter a fully qualified host name. Press Enter.
     c. Enter an IPv4 address in the Public IP box and press Enter.
     d. Enter the Netmask for the assigned IP address and press Enter.
     e. Enter the IP address of the default gateway for the IP subnet and press Enter.
     f. Enter the IP address for a domain name server and press Enter.
   - To obtain IP through DCHP,
     a. Gather the host name information to complete the DCHP IP configuration for the system.
     b. Select Obtain IP via DHCP.
     c. Select Hostname and enter a host name.
5. The change password prompt appears. Enter the new password twice. The installation of the IBM® StoredIQ operating system starts.
6. On the gateway only, a prompt appears asking for the appstack's IP address. Enter the IP address of the appstack. If you do not know or are uncertain regarding the appstack's IP address, enter `127.0.0.1`.

- **Parent topic:** [Installation of IBM StoredIQ Platform](#)

Configuration of IBM StoredIQ Platform

After you deploy virtual appliances, you must configure IBM® StoredIQ Platform before you can fulfill the product installation process.

- Configuring the gateway
  Use the SSH tool to configure the gateway.
- Configuring IBM StoredIQ Platform data server
  Use the SSH tool to log on to the data server as root.
- Configuring the application stack
  Follow these steps to configure the application stack.
- Installing IBM Desktop Data Collector
  IBM Desktop Data Collector deploys from the IBM Data Server Administrator interface. You cannot install it until you finish configuring the data server and bring it up running.

Configuring the gateway

Use the SSH tool to configure the gateway.

Procedure

1. Use the SSH tool to log on to the gateway server as root.
2. If you did not enter the correct appstack IP address when you deploy the OVA, run the /usr/bin/set-appstack-ip.sh script and then enter the appstack IP address when prompted. Note: This script adds an entry into the pg_hba.conf file, allowing trusted access to the gateway database from the AppStack IP. Without this access, reports cannot run from the appstack. This script must be run whenever the appstack IP address happens to change.

- **Parent topic:** [Configuration of IBM StoredIQ Platform](#)

Configuring IBM StoredIQ Platform data server

Use the SSH tool to log on to the data server as root.

About this task

You must complete this task for every data server.

Procedure

1. Run `service deepfiler restart`.
2. Enter `https://<IP of dataserver>` in the browser and log in to each data server with email `admin` and password `admin`.
3. Go to Administration > Configuration > DA Gateway settings.
4. Enter the IP of the gateway in the host field.
5. Enter the name of the data server in the node name field.
6. Restart services in either of the following ways:
     - Log in as root with PuTTY and then type `service deepfiler restart`.
     - Go to the data server dashboard, click the About Appliance link, and then click Restart Services.

- **Parent topic:** [Configuration of IBM StoredIQ Platform](#)

Configuring the application stack

Follow these steps to configure the application stack.

Before you begin

Ensure that the application stack is powered on. Start the vCenter console interface for the application stack.

About this task

There are two usage modes for the configuration tools. They are Bootstrap mode and Standalone mode.

In the Bootstrap mode, after the initial boot of an application stack, a series of screens appear.

Procedure

1. Accept the CENTOS and IBM license agreements. Press Enter. The change password prompt appears.
2. Enter the new password twice and press Enter. Installation of the application interface starts.
3. When the installation finishes, the user name and password prompt appears.
4. Enter the root user name and password that you previously changed. The following menu options are presented:
   - Network Configuration
   - Appstack Configuration
   - Certificate Configuration

   In the Standalone mode, the application stack tools can be used at any time to modify the application stack configuration.

   Table 1. Application stack tools in the Standalone mode

Tool	Binary
Network Configuration	/siq/bin/netcfg
Appstack Configuration	/siq/bin/appstackcfg
Certificate Configuration	/siq/bin/certcfg

- Network Configuration
  Use this tool to configure the application stack network.
- Appstack Configuration
  Use this tool to configure the general application stack settings, including the application stack's StoredIQ gateway, host name that is used by many applications, and the SMTP server.
- Certificate Configuration
  Use this tool to generate a certificate. You must complete all four following steps to ensure a successful configuration.

- **Parent topic:** Configuration of IBM StoredIQ Platform

Network Configuration

Use this tool to configure the application stack network.

Table 1. Static IP

Setting	Notes
Host name	
IP address	
Netmask	
Gateway	
Primary DNS host	
Restart network services	Restart network services settings are saved.

Table 2. Obtain IP through DHCP

Setting	Notes
Host name	
Primary DNS host	
Restart network services	Restarts network services to apply changes.

- **Parent topic:** Configuring the application stack

Appstack Configuration

Use this tool to configure the general application stack settings, including the application stack's StoredIQ gateway, host name that is used by many applications, and the SMTP server.

Table 1. Appstack Configuration

Setting	Notes
Host name	The host name for generated URLs
StoredIQ Gateway	The host or IP of the StoredIQ Gateway server
SMTP Server	Mail server host or IP address
SMTP Port	Defaults to 25
SMTP User	If supported by mail server, login username
SMTP Password	If supported by mail server, login password
Enable TLS	If supported by mail server, enables TLS encryption
Restart services	Restarts Appstack services to apply changes

- **Parent topic:** Configuring the application stack

Certificate Configuration

Use this tool to generate a certificate. You must complete all four following steps to ensure a successful configuration.

Note: Configuring a certificate is optional. To skip certificate configuration, select Finish.

Step 1: Creating self-signed root certificate
Generate a self-signed root certificate. The resulting certificate can be used as a certificate authority (CA).

Table 1. Creating self-signed root certificate

Setting	Notes
Common Name *	The name of the certificate (Required).
Email Address *	
Country (two-letter) *	An acceptable entry is an ISO-3166-1 alpha-2 code. A listing is available here.
State/Province *	
City *	
Organization *	
Department *	
Key length	The length of the key to be created. Default is 2048.
Days to expire *	The days before the certificate expiration. Default is 3650.
Root key location *	The location of the key. One is created if location does not exist.
Root certificate location	The output location of the root certificate.

Note: * means that the setting is required.

Step 2: Creating certificate signing request
Generate a certificate signing request (CSR) to be signed by a certificate authority. The process creates a key or uses a provided key and generate the CSR from it.
Note: Common name in Step 2 must be different from the common name of Step 1.

Table 2. Creating certificate signing request

Setting	Notes
Common Name *	The host name of the appstack. It must match the domain of the URL that you use.
Email Address *	
Country (two-letter) *	An acceptable entry is an ISO-3166-1 alpha-2 code. A listing is available here.
State/Province *	
City *	
Organization *	
Department *	
Key length	The length of the key to be created. Default is 2048.
Key location *	The location of the key. One is created if the location does not exist.
Certificate request location *	The output location of the CSR.

Note: * means that setting is required.

Step 3: Signing request with root certificate
Sign a provided certificate-signing request with a certificate authority eligible root certificate.

Table 3. Signing request with root certificate

Setting	Notes
Days to expire *	The days before certificate expiration. Default is 3650.

Certificate request location * The location of the CSR to sign.

Certificate location * The output location of the signed certificate.

Root key location * The location of the root key.

Root certificate location * The location of the root certificate.

Note: * means that setting is required.

Step 4: Updating Appstack HTTPS certificate

Update the application stack to use the provided certificate and key for HTTPS access.

Table 4. Updating Appstack HTTPS certificate

Setting	Notes
Key location *	The location of the key.
Certificate location *	The location of the certificate.

Note: * means that setting is required.

- **Parent topic:** [Configuring the application stack](#)

Installing IBM Desktop Data Collector

IBM® Desktop Data Collector deploys from the IBM Data Server Administrator interface. You cannot install it until you finish configuring the data server and bring it up running.

Procedure

1. Enter the https:// IP address of the data server in the browser.
2. Enter `admin` in the email field and password `admin` on the login page.
3. Under the Administration tab, click the Configuration tab.
4. Click the Desktop Settings link under the Application column on the System Configuration page.
5. Click the Download the desktop client installer link.
6. Click Run when the installer window prompts.
7. Complete the installation steps in the InstallShield Wizard.
8. After the installation is complete, IBM Desktop Data Collector appears as volumes on the primary data source page of the data server and in the IBM StoredIQ Administrator interface on the application stack.

- **Parent topic:** [Configuration of IBM StoredIQ Platform](#)

Creating a template from the IBM StoredIQ OVA

OVAs can be modified to fit unique deployment scenarios.

About this task

While building custom OVA configurations can be done, configuring the OVA can make the resulting virtual machine instance unstable or unusable. To ensure quality implementation of IBM® StoredIQ Platform, contact IBM support before you start the custom OVA configuration.

The following information provides a reference procedure for creating a new template from the IBM StoredIQ OVA before you deploy the data server. Use the vSphere Client to create the new template.

Procedure

1. Deploy OVA to create instance of the virtual machine.
   Note: Do not power on the virtual machine after successful deployment of the OVA.
2. Right-click on the virtual machine and select Edit Settings.
3. Change the processor and memory as needed.
4. To change the size of a disk, first remove its size and then add back the size you want.
   Note: Do not use this method to resize the primary disk.
5. Click File > Export > Export OVF Template.

Editing the OVF file

If you need to scale down the size of the index drive on the data servers, you can edit the OVF file.

Before you begin

To prepare your system for the change of the OVF file, you need:

- A released StoredIQ Data Server OVA
- A system with the `tar` and `openssl` utilities and a text editor
- Sufficient disk space to unpack the StoredIQ Data Server OVA (about 4 GB).

About this task

If you are comfortable with the OVF metadata standard and OVA packaging, reducing the size of the Data Server index drive can be accomplished by extracting the OVF file and manually editing the relevant parameters.

Procedure

1. Unpack the StoredIQ Data Server OVA archive with the tar utility in the current directory: tar xvf <filename>. You can see a total of three files, with extensions of ovf, mf, and vmdk.gz2.
2. Edit the .ovf file in one line with the text editor of your choice, specifying the size (ovf: capacity) in GB for disk (ovf: diskID) vmdisk.
   Note: The size for "Hard disk 2" for a production deployment must be 300 GB or greater. 300 GB is the minimum size for a metadata only index for 10 TB of managed storage.

   ```
   <Disk ovf:capacity="1984" ovf:capacityAllocationUnits="byte * 2^30" ovf:diskId=
   "vmdisk2" ovf:format="http://www.vmware.com/interfaces/specifications/
   vmdk.html#streamOptimized"/>
   ```

3. Update the mf file with the sha1 hash for the ovf and vmdk files with the openssl utility: openssl dgst -sha1 <filename>.ovf, <filename>.mf, <filename>-disk1.vmdk.gz .
4. Rebuild the OVA archive with the tar utility.
   Note: The order of the file list is important. The ovf file must be added first followed by the mf file and then the vmdk file: tar cvf <filename>.ova <filename>.ovf <filename>.mf <filename>-disk1.vmdk.gz
5. Clean up the files that are added to the ova archive. rm -f *.{ovf,mf,vmdk.gz}
6. Deploy the updated OVA. In vSphere, verify the size of "Hard disk 2" for the resulting virtual machine to ensure that it matches the size that is specified in the ovf file.
7. Power on the virtual machine and configure it from the console.
8. Confirm appropriate operation of virtual machine after initial configuration completes and the virtual machine restarts.

Application data migration

Application data can be moved from an old system to a new one in different operation systems.

You must have administrative privileges within your virtue machine environment to perform the following procedures.

- [Migrating data of the data server](#)
  You can move all data by swapping the virtual machine disk.
- [Migrating data of the gateway](#)
  Follow this procedure to migrate data in the gateway.
- [Migrating data of the application stack](#)
  Follow these steps to migrate data of the application stack.
- [Enabling public or private infosets](#)
  When you enable public or private infosets, you might want to restrict the All Data and All System infosets to only those users in the administrator role. It gives administrators more control over volume visibility. It does not make sense to have a private infoset that restricts access to certain volumes while users are still able to access those volumes through the All Data system infoset. To restrict access to these infosets, contact your technical support to run the appropriate procedure on the appstack.

Migrating data of the data server

You can move all data by swapping the virtual machine disk.

Procedure

1. Locate your old system and write down the name of the virtual machine of the system.
2. Create a clone of this system before you proceed to the next step.
3. Turn off the old system.
4. Deploy a new system with the latest OVA.
5. Turn on the new system.
6. Go through all the initial prompts to accept EULA. Wait for initialization script to run to completion.
   Note: It is important to go through this process before you proceed to the next step. Without this process, the initialization script can erase all the data. Ensure that your new system is installed and you can bring the user interface successfully.
7. Turn off the new system.
8. Right-click the virtual machine and select Edit Settings. Under the Hardware, you can see Hard Disk 1, Hard Disk 2, Hard Disk 3 installed on SCSI (0:0), SCSI (0:1), SCSI (0:2).
9. Select Hard Disk 2, which is 2 TB disk, and click Remove.
10. Under Removal options, select Remove from virtual machine and click OK. It takes a moment for the drive to be removed.
11. Right-click the new virtual machine and select Edit Settings again.
12. Under the Hardware tab, you can see Hard Disk 1, Hard Disk 2. Click Add....
13. Under Device Type, select Hard Disk and click Next.
14. Under Select a Disk, select Use an existing virtual disk or Reuse a previously configured virtual disk.
15. Under Select Existing Disk, browse to Disk File Path, where the old VM disk with the data to be migrated is. For example, [XX VM Datastore] centosX/centosX_vmdk
16. Under Advanced Options, select Virtual Device Node SCSI (0:1).
17. Under Ready to Complete, review the selection options carefully and click Finish to add the hardware. It takes a moment for the drive to be added.
18. Right-click the new system and select Edit Settings. You can see three hard disks.
19. Turn on your new system. Ensure that all the deepfiler services are up and running and start the user interface.
20. Verify that all the data was moved successfully. After verification, you can delete the old system. **Note:** The drive that was used for a swap is still located under old virtual machine folder in the VMware datastore.

- **Parent topic:** Application data migration

Migrating data of the gateway

Follow this procedure to migrate data in the gateway.

Procedure

1. Log on to your old system and run this command: `pg_dumpall -U dfuser -c -f db_file.out`. A file is created.
2. Bring the new virtual machine up and running.
3. Ssh into the new virtual machine. The file that is created in Step 1 can now be moved over to the new virtual machine and used to restore the data on the new system.
4. Follow these commands:
   a. Stop all services on the new virtual machine: `service siqinit stop`
   b. Verify that the services are running: `service deepfiler status`
   c. Start the restore process:`psql -U postgres -d postgres -f db_file.out`
   d. Restart the services: `service siqinit restart`
   e. Verify that the services are running: service deepfiler status
5. When the services are up and running on Redhat virtual machine, log in through the user interface and verify whether the data was restored successfully.

- **Parent topic:** [Application data migration](#)

Migrating data of the application stack

Follow these steps to migrate data of the application stack.

Procedure

1. Log on to the application stack system.
2. Run this command: /siq/env/postgresql/bin/pg_dumpall -U postgres -c -f db_file.out A file is created.
3. Bring the new virtual machine up and running.
4. Ssh into the new virtual machine. The file that was created in Step 2 can now be moved over to the new virtual machine and used to restore the data on the new system.
5. Follow these steps:
   a. Start the restore process: `/siq/env/postgresql/bin/psql -U postgres -d postgres -f db_file.out`
   b. Restart the services.
   c. When the services are up and running on Redhat virtual machine, log in through the user interface and verify whether the data was restored successfully.
   d. Copy /etc/siq/gateway/yaml and /etc/siq/local.yaml over to the new system.

- **Parent topic:** Application data migration

Enabling public or private infosets

When you enable public or private infosets, you might want to restrict the All Data and All System infosets to only those users in the administrator role. It gives administrators more control over volume visibility. It does not make sense to have a private infoset that restricts access to certain volumes while users are still able to access those volumes through the All Data system infoset. To restrict access to these infosets, contact your technical support to run the appropriate procedure on the appstack.

About this task

The procedure that is outlined here restricts access of the All Data and All System system infosets to only those users in the administrator role. The public or private feature does not apply to these two infosets. This script removes the convenience of these system infosets that are seen by all users. A good practice of enabling this feature is to create the system infoset, which contains all volumes that must be public.

Procedure

1. SSH onto the appstack.
2. Open Postgres and run this script:

```
#!/siq/bin/ipython
# This is a utility to generate a public mixed-level system infoset from all
currently active volumes
# make sure to chmod a+x this file to make it executable

from datetime import datetime
from mesh.standard import bind
from mesh.transport.http import HttpClient

class Bindings(object):
    def __init__(self):
        self.models = {}

    def bind(self, url, bundle, targets, mixins=None):
        client = HttpClient(url, bundle=bundle).register()
        for name, path in targets.iteritems():
            self.models[name] = bind(client.specification, path)

    def expose(self, instance_name='bindings'):
        ip = get_ipython()
        for name in self.models:
            ip.ex("%s = %s.models['%s']" % (name, instance_name, name))


bindings = Bindings()
bindings.bind('http://localhost:9998/', 'docket.entity', {
    'Infoset': 'docket.entity/1.0/enamel/1.0/infoset',
}, ['enamel.husk.infoset'])
bindings.bind('http://localhost:9999/api.enamel', 'enamel', {
    'DataSource': 'enamel/1.0/datasource',
})
bindings.expose()


formatted_time = datetime.now().strftime('%b %d, %Y %H:%M:%S')
public_all_data_objects_name = 'Public All Data Objects - {0}'.format
(formatted_time)
datasources = DataSource.query().filter(status='active').all()
composition = { 'type': 'datasource-list',
                'datasources' : [ {'id': ds.id} for ds in datasources ] }
```

```
public_all_data_objects = Infoset.create(
        name = public_all_data_objects_name,
        description = public_all_data_objects_name,
        type = 'mutable',
        composition = composition)

print 'Created infoset: ', public_all_data_objects_name
```

- **Parent topic:** [Application data migration](#)

Customizable reports

Reporting is a key step within the data-management process as it validates that processes were completed correctly within IBM® StoredIQ.

Note: You must enable specific configuration settings to utilize this functionality. Contact IBM Lab Services for further assistance.
You can customize reports in any of these scenarios:

- Modify reports to carry your organization's custom styles, logos, aligning them with other organization-based artifacts and documentation.
- Alter the format of the content reported in existing reports. For example, you can add more columns, switch axes in a graph, change the units for some values.
- Design reports to contain information that is not found in other, existing reports.

Using ODBC or any client that supports remote access to PostgreSQL, you can connect to either the IBM StoredIQ appserver or the gateway, locating a schema that is named `report_schema`. On both these server types, `report_schema` can be accessed remotely by logging in as a `reportuser`, which is an account that is created for the purposes of reporting on IBM StoredIQ data.

Note: Contact your administrator for `reportuser` credentials for the appstack and gateway.

Appstack
> These tables are available for reporting (with `SELECT` access only) from the `report_schema` on the `enamel` database on the appstack.
>
> - infoset
> - infoset_overlay
> - named_filter
> - query_filter
>
> These tables primarily carry information about infosets and filters that are created in IBM Data Workbench.

Gateway
> These tables are available for reporting (with `SELECT` access only) from the `report_schema` on the `dfdata` database on the gateway.
>
> - distribution_by_accessedtime
> - distribution_by_filetype
> - distribution_by_size
> - distribution_by_filesystem
> - distribution_by_modtime
> - distribution_by_objectclass
> - distribution_by_objectclass_by_owners_by_age
> - distribution_by_objectclass_by_owner_by_filetype
> - distribution_by_objectclass_by_owner_by_size
> - distribution_by_objectclass_by_owner
> - distribution_by_owner_by_createdtime
> - distribution_by_owner_by_modtime
> - distribution_by_owner_by_accessedtime
> - distribution_by_owner_by_filetype
> - distribution_by_owner_by_size
> - volumes
>
> The "volumes" table lists all the data sources that are defined on all data servers.
> The tables that have "distribution" in their prefixes carry summarized information for various metrics for each infoset. The infosets are referenced by their IDs here. The mapping to their names is available from the appstack tables. Each "distribution" table reports summaries for "node counts" (number of data objects in a grouping) and "total size" (total size in bytes of data objects in a grouping).

- [Installing the BIRT Report Designer](#)
  The BIRT Report Designer provides core reporting features such as report layout, data source configuration, and scripting features that render content and layout dynamically during report generation. Installing the BIRT Report Designer is a requirement for customizing reports.
- [Uploading the report design](#)
  IBM StoredIQ supports uploading a report design in a ZIP package. Report designs are expected to be in the rptdesign format within the compressed file. Currently, only one rptdesign is expected to be present within a compressed file. The design can use one or more .CSS style files and one or more image files, which can all be packaged within the compressed file.
- [Deploying fonts](#)
  By default, the IBM StoredIQ reporting engine supports only certain fonts that are provided with the product.
- [Changing the reportuser password](#)
  If needed, the password for the reportuser account on both the gateway and the appstack can be changed.

Installing the BIRT Report Designer

The BIRT Report Designer provides core reporting features such as report layout, data source configuration, and scripting features that render content and layout dynamically during report generation. Installing the BIRT Report Designer is a requirement for customizing reports.

Procedure

1. Install Eclipse, which can be downloaded from http://download.eclipse.org/.
2. Download BIRT from http://download.eclipse.org/birt/downloads/.
   Note: Several different packages contain BIRT. You need the all-in-one download, which must be installed on your workstation, specifically BIRT version 4.3.2.
   You can also place BIRT into your existing Eclipse environment with the Eclipse Update Manager - BIRT Update Site.

- Installing the JDBC driver
  A JDBC driver is needed to work with remote PostgreSQL databases with BIRT.
- Adding data sources
  The gateway and appstack databases act as the primary data sources for most reporting projects.
- Setting report parameters
  Some report parameters must be provided when you render the report.
- Adding data sets
  A report's charts and tables are built on data sets, which can either use data from the gateway or from the appstack. You can also define a joint data set that joins data sets from both the appstack and the gateway.
- Testing report designs
  Report designers can test their designs in the BIRT Report Designer itself using the preview feature, which is an HTML rendering of the report.

- **Parent topic:** Customizable reports

Installing the JDBC driver

A JDBC driver is needed to work with remote PostgreSQL databases with BIRT.

Procedure

1. Download the postgres_9.2 JAR file from the http://jdbc.postgresql.org/download.html.
2. Start Eclipse. Click File > New > Project, and in Business Intelligence and Reporting Tools, select Report Project.
3. Click New > Report, and in the Add a new Data Source dialog box, select JDBC Data Source.
4. Click Next, and then complete the text boxes within the Edit Data Source dialog box.
5. Click Manage Drivers.
6. Click Add to add the downloaded PostgreSQL JDBC .JAR file.

- **Parent topic:** Installing the BIRT Report Designer

Adding data sources

The gateway and appstack databases act as the primary data sources for most reporting projects.

About this task

When you create a data source for either the gateway or appstack database, be certain to select JDBC Data Source as the data source type.

Procedure

1. In the Driver Class text box, enter `org.postgresql.Driver (v9.2)`
2. In the Database URL text box, enter jdbc:postgresql://x.x.x.x/enamel, where x.x.x.x is the IP address or DNS resolvable host name of the appstack.
   Note: When the data source that is being created is for the gateway, this URL is used: jdbc:postgresql://x.x.x.x/dfdata, where x.x.x.x is the IP address or resolvable host name of the gateway.
3. In the User Name text box, enter `reportuser`.
4. In the Password text box, enter the password that is associated with this user name.
   Note: Your administrator might set different passwords for the appstack and the gateway. Contact your administrator to receive this password.
5. Click Test Connection to ensure that a connection can be established with the remote database.

- **Parent topic:** [Installing the BIRT Report Designer](#)

Creating a connection profile

When you create data sources repeatedly, you can create a connection profile as an alternative.

Procedure

1. When you create a new data source, click Create from a connection profile in the profile store.
2. From Select a Connection Profile, click New.
3. Click New and then select BIRT JDBC Data Source, providing the connection details specific to either the appstack or the gateway.
4. In the Specify file name text box, enter a name for your connection profile.

Results

After the connection profile is created, it can be reused when you create a new data source. Select Create from a connection profile in the profile store and then browse to select the profile from the available list.

Setting report parameters

Some report parameters must be provided when you render the report.

About this task

Currently, IBM® StoredIQ supports only one report parameter, which must map to InfoSet ID. It is the only parameter that can be supplied when you render the report.

Procedure

1. Create a report parameter.
2. Enter a useful name such as `infoset_id`.
3. For the message that must be displayed to the user that renders the report, enter the text in the Prompt text text box.
4. Select the appropriate Data Type, Display Type, and Is Required values. For an InfoSet ID parameter, these values must be set to String, Text Box, or True.

- **Parent topic:** [Installing the BIRT Report Designer](#)

Adding data sets

A report's charts and tables are built on data sets, which can either use data from the gateway or from the appstack. You can also define a joint data set that joins data sets from both the appstack and the gateway.

Procedure

1. To create an appstack data set:
    a. Create a data set.
    b. Select the previously named data source that references the appstack.
    c. Enter a valid name.
    d. Enter an SQL query that fetches some data back from the `report_schema`. If your data set targets infoset data, it carries a `where-clause` that looks like `where infoset_id=?`. The `?`denotes that a report parameter must be used here.
    e. If a report parameter is to be used in the SQL query, you must add a parameter to the data set.
        - Click Parameters > New.
        - Change the Linked to Report Parameter value from None to the name of your report parameter that is created during data set creation.
2. To create a gateway data set:
    a. Create a data set.
    b. Select the previously named data source that references the gateway.
    c. Enter a valid name.
    d. Enter an SQL query that fetches some data back from the `report_schema`. If your data set targets infoset data, it carries a `where-clause` that looks like `where infoset_id=?::uuid`. The `?` denotes that a report parameter must be used here. The `type-cast` to `uuid` is important on the gateway.
    e. If a report parameter is to be used in the SQL query, you must add a parameter to the data set.
        - Click Parameters > New.
        - Accept the default values in this dialog box; however, the Linked to Report Parameter value must be changed from None to the name of the report parameter that is created in the previous step.

- **Parent topic:** Installing the BIRT Report Designer

Testing report designs

Report designers can test their designs in the BIRT Report Designer itself using the preview feature, which is an HTML rendering of the report.

About this task

Before you upload the customized report, you must preview the report as both HTML and as a PDF. It helps to ensure that the report renders properly on the appstack.

Procedure

1. In Eclipse, open your custom design.
2. Click Preview to view the HTML preview of the custom report.
3. Click Run > Run as > PDF to view the PDF rendering of the report.

- **Parent topic:** [Installing the BIRT Report Designer](#)

Uploading the report design

IBM® StoredIQ supports uploading a report design in a ZIP package. Report designs are expected to be in the rptdesign format within the compressed file. Currently, only one rptdesign is expected to be present within a compressed file. The design can use one or more .CSS style files and one or more image files, which can all be packaged within the compressed file.

About this task

To better manage the artifacts such as the images and style sheets for a single report design package, you must do one of the following tasks. By selecting one of the report-design management options, you can manage all images and style sheets. They are used by the design in a single location, making the creation of the ZIP package easier as you compress a folder.

- Create a project for each report design. It means that each project contains only one rptdesign.
- Create a folder for each report design within the project. It means that each folder contains only one rptdesign.

Additionally, the uploaded compressed file must meet the following requirements:

- It contains only one rptdesign file and uses the file extension .rptdesign.
- All CSS and images that are referenced from the rptdesign are also present within the compressed file.
- Only one report parameter is defined in the rptdesign.
- Images use these extensions: .bmp, .jpg, .jpeg, .jpe, .jfif, .gif, .png, .tif, .tiff, .ico, or .svg.
- Style sheets must use the .css file extension.

- **Parent topic:** [Customizable reports](#)

Deploying fonts

By default, the IBM® StoredIQ reporting engine supports only certain fonts that are provided with the product.

About this task

By default, the IBM StoredIQ reporting engine supports the following fonts:

- icon-webfont.ttf
- ubuntu-bold-italic.ttf
- ubuntu-bold.ttf
- ubuntu-medium-italic.ttf
- ubuntu-medium.ttf
- ubuntu-regular-italic.ttf
- ubuntu-regular.ttf

Procedure

1. Deploy the fonts to the appserver. Copy the fonts to this folder on the appstack: /usr/share/fonts/default/TrueType.
   Note: Only True Type Fonts are supported. If a font is used in the report design and is not deployed correctly on the appstack, the default Times New Roman font is used when you render the report.
2. Restart IBM StoredIQ services. IBM StoredIQ services must be restarted on the appstack after new fonts are deployed.

- **Parent topic:** [Customizable reports](#)

Changing the reportuser password

If needed, the password for the reportuser account on both the gateway and the appstack can be changed.

Procedure

1. To change the password for the reportuser on the appstack:
    a. Log in to the appstack over ssh as `root`.
    b. Run the /siq/bin/change_reportuser_password script with the new password for the reportuser as a parameter.
2. To change the password for the reportuser on the gateway:
    a. Log in to the gateway over ssh as `root`.
    b. Copy the /siq/bin/change_reportuser_password from the appstack.
    c. Run this script with the new password for the reportuser as a parameter.

- **Parent topic:** Customizable reports

Installing the Livelink APIs

The Livelink API requires the installation of the lapi.jar, which is provided by its vendor, OpenText, under license to their customers. IBM cannot redistribute this file. At installation, you must copy this file to each data server in the /usr/local/IBM/ICI/vendor directory. It is likely to be found on the Livelink server in the C:\OPENTEXT\application\WEB-INF\lib directory. However, Livelink installations can vary; you must consult the customer's documentation to locate the lapi.jar file.

Procedure

1. Locate the lapi.jar file by going to C:\OPENTEXT\application\WEB-INF\lib.
    Note: Installations vary by customer. If the lapi.jar file is not in the C:\OPENTEXT\application\WEB-INF\lib directory, consult the customer's documentation.
2. Copy the lapi.jar file to each data server in the /usr/local/IBM/ICI/vendor directory.

Administering IBM StoredIQ

IBM® StoredIQ solutions provide services to manage the IBM StoredIQ Administrator application and data servers.

With the services for the IBM StoredIQ Administrator application, you can manage:

- Data servers and volumes
- System infosets
- Users
- Actions
- Target set
- Reports
- Auto-classification models
- Concepts
-

With the services for the data servers, you can manage:

- System administration
- IBM StoredIQ Platform configuration
- Volumes and data sources
- Data harvesting
- Job configuration
- Desktop collection
- Folders
- Audits and logs
- Deployment of customized web services

- [Administering IBM StoredIQ Administrator](#)
  Use IBM StoredIQ Administrator to configure shares, create system infosets, manage users, target sets, and actions, and create reports.
- [Administering IBM StoredIQ Platform Data Server](#)
  The following topics provide instructions on how to conduct administrative tasks such as administering appliance, configuring IBM StoredIQ Platform, and creating volumes and data source.

Administering IBM StoredIQ Administrator

Use IBM® StoredIQ Administrator to configure shares, create system infosets, manage users, target sets, and actions, and create reports.

- [Overview of IBM StoredIQ Administrator](#)
- [Data Servers and Volumes](#)
  The Data Servers and Volumes page of IBM StoredIQ Administrator contains the Data Dashboard, where you see all data currently under management, All Data Servers tab, All Volumes tab, and Performance tab.
- [System infosets](#)
  System infosets can be created in IBM StoredIQ Data Workbench to allow users to have a different starting point than the All Data Objects infoset.
- [User management](#)
  From the User Management page, users can be defined and provided with login information so that they can use the application stack.
- [Action management](#)
  Actions represent executable processes that act upon indexed data within data servers. Use the Action Management page to create and modify actions that are used by IBM StoredIQ Administrator.
- [Target sets](#)
  Some actions, namely copy actions, require a destination to complete, and a target set represents a mapping for each data server of the wanted target volume. To copy files, the data server must know where to copy the file to. Each data server must be configured with a list of target volumes, such as retention and discovery export volumes.
- [Reports](#)
  IBM StoredIQ contains a number of default reports and the ability to upload a report package.
- [Auto-classification models](#)
  Auto-classification models, also called automated document categorization, integrates the IBM® Content Classification's classification model into the IBM StoredIQ infoset-generation process.
- [Concept management](#)
  Within IBM StoredIQ Platform, you can use the concept-management feature to relate business ideas to indexed data.

- **Parent topic:** [Administering IBM StoredIQ](#)

Overview of IBM StoredIQ Administrator

IBM® StoredIQ Administrator provides at-a-glance understanding of the different issues that can crop up in the IBM StoredIQ Platform environment. These views are unique to the IBM StoredIQ Administrator application, providing an overview of how the system is running. These views also allow access to various pieces of information that are shared across applications or allow for the management of resources in a centralized manner.

The Administrator is the person responsible for managing the IBM StoredIQ Platform installation at a customer site. This individual has strong understanding of data sources, harvests, indexes, data servers, jobs, infosets, reports, concepts, and actions. This list provides an overview as to how IBM StoredIQ Administrator works.

Managing Data Servers
> Using IBM StoredIQ Administrator, the Administrator can identify what data servers are deployed, their location, what data is being managed, and the status of each data server in the system. Volume management is a central component of IBM StoredIQ Platform. With IBM StoredIQ Administrator, the Administrator can see:
>
> - What volumes are currently under management
> - Which data server is responsible for that volume
> - The state of the volume after indexing
> - The amount and size of information that is contained by each volume.
>
> Administrators can also add volumes to data servers through this interface.

Administering Harvests
> Administrators can initiate the incremental or full harvest of a single volume through the IBM StoredIQ Administrator user interface without having to log into IBM StoredIQ Platform Data Server.

Creating System Infosets
> System infosets that use only specific indexed volumes can be created and managed within IBM StoredIQ Administrator. Although infosets are a core component of IBM StoredIQ Data Workbench, system infosets are created as a shortcut for users in IBM StoredIQ Administrator.

Managing Users
> The user management area allows administrators to create users and manage users' access to the various IBM StoredIQ Platform applications.

Configuring and Managing Actions
> An action is any process upon the data that is represented by the indexes. Actions are run by data servers on indexed data objects. Any errors or warnings that are generated as a result of an action are recorded as exceptions in IBM StoredIQ Data Workbench. Note: Actions can be created within IBM StoredIQ Administrator and then made available to other IBM StoredIQ Platform applications such as IBM StoredIQ Data Workbench.

Managing target sets
> Provides an interface that allows the user to set the wanted targets for specific actions that require a destination volume for their actions.

Creating reports
> Helps you to create reports and upload report packages, helping you to produce reports about your environment.

Using Auto-classification Models
> Deploys natural language processing-based document classification to help you identify relevant documents in cases when relevance is difficult to characterize with traditional methods, such as queries and set operations.

Managing Concepts
> Helps you to relate business concepts to indexed data.

- [Switching between applications](#)
> With the IBM StoredIQ product, you can switch between different applications.

Switching between applications

With the IBM® StoredIQ product, you can switch between different applications.

About this task

These applications include:

- IBM StoredIQ Data Workbench
- IBM StoredIQ eDiscovery
- IBM StoredIQ Data Script
- IBM Stored IQ Policy Manager
- IBM StoredIQ Administrator

To switch to the application that you want to work on, follow these steps.

Procedure

1. Within the global navigation bar, click the white arrow next to the name of the application. A menu appears, listing available IBM StoredIQ Platform applications.
2. In the menu, click the name of the application to which you want to switch.
   Note: There are five roles for users: administrator, data user, discovery user, policy user, and SDK user. Different IBM StoredIQ Platform applications are available to the user, depending on what roles were selected. Administrators have access to IBM StoredIQ Administrator in addition to IBM StoredIQ Data Workbench, IBM StoredIQ eDiscovery, and IBM StoredIQ Data Script. Discovery users have access to IBM StoredIQ eDiscovery. Data users have access to IBM StoredIQ Data Workbench. Policy users have access to IBM Stored IQ Policy Manager. The SDK user role is enabled only for users of the IBM StoredIQ SDK. Individuals with the SDK user role must have both the administrator and SDK user roles in order to work with the IBM StoredIQ SDK. It should also be noted that users assigned only the SDK user role cannot log into any of the IBM StoredIQ applications.

- **Parent topic:** Overview of IBM StoredIQ eDiscovery

Data Servers and Volumes

The Data Servers and Volumes page of IBM® StoredIQ Administrator contains the Data Dashboard, where you see all data currently under management, All Data Servers tab, All Volumes tab, and Performance tab.

Data Dashboard

The description of the Data Dashboard fields is listed in the following table.

Table 1. Data Dashboard fields and descriptions

Data Dashboard	Description
Total Data Objects	This field lists the total number of data objects under management.
Total Data Size	This field lists the total size of data objects under management.
Number of Data Servers	This field lists the total number of data servers under management.
Number of Volumes	This field lists the total number of volumes under management.

All Data Servers

All Data Servers tab lists the detailed information about the data servers

Table 2. All Data Servers tab column names and descriptions

All Data Servers Column Name	Description
Data server name	Lists the name of the data server.
Status	Lists the data server's status, namely Healthy, Vulnerable, Critical, or Unavailable.
IP address	Lists the IP address of the data server.
Data objects	Lists the number of data objects that are found on the data server.
Total data object size	Lists the size of the data objects that are found on the data server.
Available space	Lists the amount of free space on the data server.
Load average	Lists the load average.

All Volumes

The tab provides the detailed information about the volumes.

Table 3. All Volumes tab column names and descriptions

Volumes Column Name	Description
Volume name	Lists the name of the volume.
Data server	Lists the name of the parent data server.
Type	Lists the volume's type, namely Primary, Retention, or Discovery Export. Primary indicates a primary volume. Retention indicates a retention volume. Discovery Export indicates a discovery export volume.
Source type	Lists the source or connection type for the volume.
Server name	Lists the name of the server where the volume is found.
Data objects	Lists the number of data objects that are found on the volume.
Total data object size	Lists the size of the data objects in the volume.
Last harvested	Denotes the last time that the volume was harvested.

Performance

This tab provides the performance information of the data servers.

Table 4. Performance tab column names and descriptions

Performance Column Name	Description
Data server name	Lists the name of the data server.
Status	Lists the data server's status, namely Healthy, Vulnerable, Critical, or Unavailable.
IP address	Lists the IP address of the data server.
Active db connection	Lists the number of database connections currently open for the selected data server.
Free memory	Denotes the amount of available memory (RAM) for the selected data server.
Free swap memory	Denotes the amount of available swap memory for the selected data server.
Load average	Lists the average process load for the selected data server.
Last communication	Lists the last time the data server that is connected to the gateway.
Uptime	Lists the total duration of time for which the data server ran.

- Searching for a specific data server
  When you have multiple data servers available from which to choose, you can search for a specific data server instead of using the slider to move through available options.
- Adding a primary volume
  A primary volume serves as a primary data source in IBM StoredIQ Platform. You must have at least one primary volume within your configuration.
- Adding a retention volume
  A retention volume stores data objects that are placed under retention, meaning that the object is retained for a specified period.
- Adding a discovery export volume
  A discovery export volume contains data that is produced from a discovery export action. That data is kept so that it can be exported as a load file and uploaded into a legal review tool.
- Viewing details of a data server
  Detailed information for any data server or volume can be viewed.
- Viewing performance of a data server
  The ability to check the status of all deployed data servers is critical to understanding the system's performance metrics.

Searching for a specific data server

When you have multiple data servers available from which to choose, you can search for a specific data server instead of using the slider to move through available options.

Procedure

1. In the Enter key term(s)... text box, enter the name of the data server for which you want to search, and then click Search. You can enter either the full data server name or a portion of it. To remove the search term, click the X to the left of the Enter key terms.... text box. Actions that match the entered search term are returned.
2. Select a data server from the returned list of servers.
3. View the data server's details by clicking View Details.

* **Parent topic:** [Data Servers and Volumes](#)

Related concepts

* [Data Servers and Volumes](#)

Adding a primary volume

A primary volume serves as a primary data source in IBM® StoredIQ Platform. You must have at least one primary volume within your configuration.

Procedure

1. Select Data Servers & Volumes > All Data Servers, select a data server, and then click Add Volume. The Add Volume dialog box appears.
2. In the Add Volume dialog box, complete these fields:

Table 1. Primary volume options

Add Volume dialog box option	Action	Notes	Applicable source types
Volume Type	In the Volume Type list, select Primary.		○ CIFS ○ Exchange ○ FileNet ○ Livelink ○ NewsGator ○ NFS ○ SharePoint
Source Type	In the Source Type list, select the server's connection or source type.		○ CIFS ○ Exchange ○ FileNet ○ Livelink ○ NewsGator ○ NFS ○ SharePoint
Server	In the Server text box, enter the fully qualified domain name of the server from which the volume is available for mounting.		○ CIFS ○ Exchange ○ FileNet ○ Livelink ○ NewsGator ○ NFS ○ SharePoint
Username	In the Username text box, enter the user name that is used to connect to and mount the volume.		○ CIFS ○ Exchange ○ FileNet ○ Livelink ○ NewsGator ○ SharePoint
Password	In the Password text box, enter the password that is used to connect to and mount the volume.		○ CIFS ○ Exchange ○ FileNet ○ Livelink ○ NewsGator ○ SharePoint
Assign to Data Server	In the Assign to Data Server list, select the data server.		○ CIFS ○ Exchange ○ FileNet ○ Livelink ○ NewsGator ○ NFS ○ SharePoint
Volume Name	In the Volume Name text box, enter the common name of this volume.		○ CIFS ○ Exchange ○ FileNet ○ Livelink

	volume.		○ NewsGator ○ NFS ○ SharePoint
FileNet Domain	In the FileNet Domain text box, enter the FileNet domain.		○ FileNet
Share	In the Share text box, enter the share name or this volume.		○ CIFS
Export	In the Export text box, enter the export name for this volume.		○ NFS
Site URL	In the Site URL text box, enter the site URL of the SharePoint server.		○ SharePoint
Object Store	In the Object Store text box, enter the object store.		○ FileNet
Connection Type	In the Connection Type list, select either HTTP or HTTPS.		○ FileNet
Port	In the Port text box, enter a port number.		○ FileNet ○ Livelink
Database	In the Database text box, enter the database name.		○ Livelink
Search slice	In the Search slice text box, enter the search slice.		○ Livelink
Server Version	In the Server Version list, select the version of Microsoft Exchange, choosing from 2000/2003, 2007, or 2010.		○ Exchange
Mailbox Server	In the Mailbox Server text box, enter the name of the mailbox server.		○ Exchange
Active Directory Server	In the Active Directory Server text box, enter the name of the Active Directory server. To use secure socket layer, select the Use SSL check box.		○ Exchange
Path	In the Path text box, enter the appropriate directory path.		○ FileNet
Stanza	In the Stanza text box, enter the appropriate stanza.		○ FileNet
Initial Directory	In the Initial Directory text box, enter the name of the initial directory from which the harvest must begin.		○ CIFS ○ Exchange ○ FileNet ○ Livelink ○ NewsGator ○ NFS ○ SharePoint
Virtual Root	In the Virtual Root text box, change the default name to match the Exchange server designation.	For Exchange, this option must be changed to match the server designation.	○ Exchange
Indexing Options	Select the check box for the indexing options that you want to include: ○ Include metadata for	These options are not selected by	○ CIFS ○ Exchange ○ FileNet ○ Livelink

| contained objects
○ Include content tagging
and full-text index | default. | ○ NewsGator
○ NFS
○ SharePoint |

3. Click Save to save your configurations and add the volume.
4. Click the All Volumes tab. Notice that the added volume appears therein, listed as a primary volume. To harvest this newly added volume, select that volume and then click Harvest.

- **Parent topic:** [Data Servers and Volumes](#)

Related concepts

- [Data Servers and Volumes](#)

Adding a retention volume

A retention volume stores data objects that are placed under retention, meaning that the object is retained for a specified period.

Procedure

1. Click Data Servers & Volumes > All Data Servers, and then click Add Volume. The Add Volume dialog box appears.
2. In the Add Volume dialog box, complete these fields:

Table 1. Retention volume options

Add Volume dialog box option	Action	Notes	Source type
Volume Type	In the Volume Type list, select Retention.		○ CIFS ○ NFS
Source Type	In the Source Type list, select the source or connection type.		○ CIFS ○ NFS
Server	In the Server text box, enter the fully qualified domain name of the server from which the volume is available for mounting.		○ CIFS ○ NFS
Username	In the Username text box, enter the user name that is used to connect to and mount the volume.		○ CIFS
Password	In the Password text box, enter the password that is used to connect to and mount the volume.		○ CIFS
Assign to Data Server	In the Assign to Data Server list, select a data server.		○ CIFS ○ NFS
Volume Name	In the Volume Name text box, enter the common name of this volume.		○ CIFS ○ NFS
Share	In the Share text box, enter the share name for this volume.		○ CIFS
Export	In the Export text box, enter the export name for this volume.		○ NFS
Indexing Options	Select the check box for the indexing options that you want to include: ○ Include metadata for contained objects ○ Include content tagging and full-text index	These options are not selected by default.	○ CIFS ○ NFS

3. Click Save to save your configurations and add the volume. Retention volumes do not have any data to harvest until a copy to retention runs.
4. Click the All Volumes tab. The added volume appears, listed as a retention volume.

- **Parent topic:** Data Servers and Volumes

Related concepts

- Data Servers and Volumes

Adding a discovery export volume

A discovery export volume contains data that is produced from a discovery export action. That data is kept so that it can be exported as a load file and uploaded into a legal review tool.

Procedure

1. Click Data Servers & Volumes > All Data Servers, and then click Add Volume. The Add Volume dialog box appears.
2. In the Add Volume dialog box, complete these fields:

Table 1. Discovery export volume options

Add Volume dialog box option	Action	Notes	Source type
Volume Type	In the Volume Type list, select Discovery Export.		○ CIFS ○ NFS
Source Type	In the Source Type list, select the source or connection type.		○ CIFS ○ NFS
Server	In the Server text box, enter the fully qualified domain name of the server from which the volume is available for mounting.		○ CIFS ○ NFS
Username	In the Username text box, enter the user name that is used to connect to and mount the volume.		○ CIFS
Password	In the Password text box, enter the password that is used to connect to and mount the volume.		○ CIFS
Assign to Data Server	In the Assign to Data Server list, select the data server.		○ CIFS ○ NFS
Volume Name	In the Volume Name text box, enter the name of this volume.		○ CIFS ○ NFS
Share	In the Share text box, enter the share name for this volume.		○ CIFS
Export	In the Export text box, enter the export name for this volume.		○ NFS

3. Click Save to save your configurations and add the volume.
4. Click the All Volumes tab. The added volume appears, listed as a discovery export volume.

- **Parent topic:** [Data Servers and Volumes](#)

Related concepts

- [Data Servers and Volumes](#)

Viewing details of a data server

Detailed information for any data server or volume can be viewed.

Procedure

1. Click Data Servers & Volumes > All Data Servers, and then click View Details. The <data server name> Details page appears, showing the All Data Servers and All Volumes tabs.
2. Select the All Data Servers tab, select the item for which you want to view its details, and then click View Details.
   Note: To view details for multiple items, press Ctrl and select those items, and then click View Details.
   A summary of the data server's details appears along with the Volumes and Performance tabs.

- **Parent topic:** [Data Servers and Volumes](#)

Related concepts

- [Data Servers and Volumes](#)

Viewing performance of a data server

The ability to check the status of all deployed data servers is critical to understanding the system's performance metrics.

Procedure

1. Click Data Servers & Volumes > All Data Servers, and then click View Details. The <data server name> Details page appears, showing the Volumes and Performance.
2. Select the Performance tab, and the data server's performance details appears.

- **Parent topic:** Data Servers and Volumes

Related concepts

- Data Servers and Volumes

System infosets

System infosets can be created in IBM® StoredIQ Data Workbench to allow users to have a different starting point than the All Data Objects infoset.

System infosets are the basis for user-created infosets. System infosets allow administrators to select which indexed volumes the infoset can draw upon, creating a shortcut for IBM StoredIQ Data Workbench users. Actions cannot be run on system infosets, but only on user-created infosets. Therefore, you must first create the system infoset here in IBM StoredIQ Administrator and then create a user infoset that can be acted upon in IBM StoredIQ Data Workbench.

Infosets can be public or private. The administrator determines users' access to system infosets at the time of creation. The All Data Objects and All System-Level Objects infosets, which are generated by the application, are set to admin by default and can be viewed only by an administrator.

Note: System infosets must not be used to manage volumes or indexes. Additionally, system infosets must be created judiciously. If there are several system infosets through which a user must cull, the shortcut aspect is removed.

Table 1. Infoset management column names and descriptions

Infoset Management Column Name	Description
Name	Lists the name of the system infoset. This name must be unique against all infosets.
Total objects	Lists the total number of data objects that are contained within the system infoset.
Infoset size	Lists size of the system infoset.
In use by	Indicates the infosets in which this system infoset is used. If the system infoset is in use, a link is shown, indicating the number of infosets by which the system infoset is being used. Click the link to see which infosets use this system infoset.
Access	Indicates whether the system infoset is public or private. If the system infoset is public, it can be viewed by any user. If the system infoset is private, it can be viewed by only those users granted permission by the administrator. If users are granted permission, the number of users is listed with a hyperlink to the Access Members modal, which lists the name and role of users who have access.
Description	This column lists the optional description of the system infoset.

- Searching for a specific system infoset
  When you have multiple system infosets available to choose, you can search for a specific system infoset instead of using the slider to move through available options.
- Creating a system infoset
  System infosets can be created with selected volumes, creating a shortcut for IBM StoredIQ Data Workbench users.
- Editing a system infoset
  Existing user-created system infosets can be modified as needed.
- Viewing infosets that use a system infoset
  Click the link to view which infosets use the selected system infoset.
- Viewing access members
- Deleting a system infoset
  System infosets can be deleted from the List of Infosets page if they are not in use by other infosets. When a system infoset is in use by a direct ancestor of other infosets, Delete is unavailable for use so that the system infoset cannot be deleted.

Searching for a specific system infoset

When you have multiple system infosets available to choose, you can search for a specific system infoset instead of using the slider to move through available options.

Procedure

1. In the Enter key term(s)... text box, enter the name of the system infoset for which you want to search, and then click Search. You can enter either the full system infoset name or a portion of it. To remove the search term, click the X to the left of the Enter key term(s).... text box. System infosets that match the entered search term are returned.
2. Select a system infoset from the returned list of system infosets.

- **Parent topic:** System infosets

Related concepts

- System infosets

Creating a system infoset

System infosets can be created with selected volumes, creating a shortcut for IBM® StoredIQ Data Workbench users.

Procedure

1. Click System Infosets, and then click Create Infoset. The Add System Infoset dialog box appears.
2. In the Add System Infoset dialog box, complete these fields.
   a. In the Infoset Name text box, enter the name of this system infoset.
   b. In the Description of Infoset text box, enter a brief description of this system infoset. This infoset might need to be used by several people, so an accurate description can be helpful to others.
   c. In the Access area, click either Public or Private. The All Data Objects and All System-Level Objects infosets, which are generated by the application, can be viewed only by an administrator. System infosets that are created by an administrator default to public. Public system infosets are visible and available to all users, including users that will be added at a future time. For public system infosets, the user list is disabled.
   d. For private system infosets, the list of available users becomes active. Select the users for whom this private infoset can be seen. If no users are selected for a private system infoset, then only administrators can see that system infoset because they have administrative rights to every infoset.
   Note: Access cannot be edited. It must be determined at the time of infoset creation.
   e. Add volumes to this infoset. In the Available Volumes area, select the volumes that you want to add to this system infoset, and then click Add. To remove volumes from the infoset, select the volume in Selected Volumes and then click Remove.
3. Click Save to save the system infoset. It now appears on the List of Infosets page.

- **Parent topic:** System infosets

Related concepts

- System infosets

Editing a system infoset

Existing user-created system infosets can be modified as needed.

Procedure

1. In the List of Infosets page, select the system infoset that you want to edit, and then click Edit Infoset. The Edit System Infoset dialog box appears.
   Note: Only created infosets can be edited.
2. In the Edit System Infoset dialog box, edit the infoset as needed.
   Note: Access cannot be edited. It must be determined at the time of infoset creation.
3. Click Save to save your changes.

- **Parent topic:** System infosets

Related concepts

- System infosets

Viewing infosets that use a system infoset

Click the link to view which infosets use the selected system infoset.

Procedure

1. Select the system infoset for which you would like what other infosets use the system infoset.
2. In the In use by column, click the link. The System infoset modal for the system infoset appears, listing the infosets that use it.
3. Click Close.

- **Parent topic:** [System infosets](#)

Related concepts

- [System infosets](#)

Viewing access members

Procedure

1. Select the system infoset for which you would like to access its members, and then click the link in the Access column. The Access Members window appears, listing both the user names and roles of all members of the system infoset.
   Note: The columns can be sorted to show members in ascending or descending order.
2. Click Close.

- **Parent topic:** [System infosets](#)

Related concepts

- [System infosets](#)

Deleting a system infoset

System infosets can be deleted from the List of Infosets page if they are not in use by other infosets. When a system infoset is in use by a direct ancestor of other infosets, Delete is unavailable for use so that the system infoset cannot be deleted.

Procedure

1. In the List of Infosets page, select the system infoset that you want to delete, and then click Delete Infoset.
   Note: The All Data Objects and All System-Level Objects system infosets cannot be deleted.
   The Delete Infoset window appears.
2. Click Delete to delete the selected system infoset. The system infoset is deleted from the list of available system infosets.

- **Parent topic:** [System infosets](#)

Related concepts

- [System infosets](#)

User management

From the User Management page, users can be defined and provided with login information so that they can use the application stack.

This table lists and describes all of the fields that are seen within the List of Users page. When a user is selected, that user can be viewed, edited, or deleted.

Table 1. List of Users column names and descriptions

List of Users column name	Description
User name	The user name that is assigned to the user
First name	The user's first or given name
Last name	The user's last or surname
Email address	The user's email address. System notifications are sent to this email address.
Role	The user's role. Options include Admin, Data User, Discovery User, Policy User, or SDK User.
Status	The user's status. Options include Active or Inactive. • Active users can log in and use the product. Active is the default status for a user. • Inactive users cannot log in to the product suite because their account is rendered inactive by the Administrator. When inactive users attempt to log in, they are notified that their login is invalid and that they must contact an Administrator.
Last login	The date and time of the most recent user login. If the user is not logged in to the system, N/A is shown.

- Viewing your user profile
  When you are viewing your user profile, you can see various aspects, including your assigned roles.
- Searching for a specific user
  When you have multiple users available to choose, you can search for a specific user instead of using the slider to move through available options.
- Creating a user
  Within IBM® StoredIQ Platform, a user authenticates to and interacts with the product suite. When a user is created, this user can log in and use the system.
- Editing a user
  You can modify existing users, including their status as needed. You can also activate the disabled users within IBM StoredIQ Administrator.
- Viewing details of a user
  Basic user information such as first and surname, email address, and role can be viewed by clicking View Details.
- Deleting a user
  You can delete existing users from the system as needed; however, the log entries of a deleted user are also deleted. In certain instances, it is a good practice that you change a user's status.
- Changing user passwords
  Either the user or the administrator can change passwords. Valid passwords are at least 8 characters in length and contain at least an uppercase letter, a lowercase letter, a number, and a special character.
- Resetting the user password
  User passwords can be set and reset as needed by users and administrators alike.
- Obtaining a new password
  Users occasionally need to obtain new passwords to be able to log in to the system.

Viewing your user profile

When you are viewing your user profile, you can see various aspects, including your assigned roles.

Procedure

1. Within the interface, click the down arrow to the right of your user name, and then select View Profile. The View Profile window appears, showing various pieces of data, including your name, role, and email address.
2. To change your password, click Change Password.
3. Click Close to close the View Profile window.

- **Parent topic:** [User management](#)

Related concepts

- [User management](#)

Searching for a specific user

When you have multiple users available to choose, you can search for a specific user instead of using the slider to move through available options.

Procedure

1. In the Enter key term(s)... text box, enter the user name of the user for which you want to search, and then click Search. To remove the search term, click the X to the right of the Enter key term(s)... text box. Users that match the entered search term are returned.
2. Select a user name from the returned list of user names.

- **Parent topic:** [User management](#)

Related concepts

- [User management](#)

Creating a user

Within IBM® StoredIQ Platform, a user authenticates to and interacts with the product suite. When a user is created, this user can log in and use the system.

Procedure

1. Click Users, and then click Create User. The Add User - Details dialog box appears.
2. In the Add User - Details dialog box, complete these fields.
    a. In the User name text box, enter the user name of the individual. It is generally a short, unique name, not the person's full name.
    b. In the First name text box, enter the first or common name of the user.
    c. In the Last name text box, enter the last or surname of the user.
    d. In the Email address text box, enter the user's email address.
       Note: Because User name and Email address are unique fields, users can log in to the system by entering either their email address or their user name.
    e. In the Password text box, enter the password for this user.
       Note: Passwords must be at least 8 characters in length and contain at least one of each of these characters: an uppercase letter, a lowercase letter, a number, and a special character.
    f. In the Re-enter password text box, enter the user's password again.
    g. In the Role list, select the user's role. Options include Admin, Data User, Discovery User, Policy User, or SDK User. A user can be assigned multiple roles by selecting the check boxes in the role list.
       Note: The default option during user creation is Data User.
    h. In the Status list, select the user's status. Options include Active or Inactive.
       Note: The default option during user creation is Active.
    i. Click Save, and the user is created.
   The newly created user appears within the User Management page.

- **Parent topic:** [User management](#)

Related concepts

- [User management](#)

Editing a user

You can modify existing users, including their status as needed. You can also activate the disabled users within IBM® StoredIQ Administrator.

About this task

Note: Users become disabled after three failed log-in attempts. It is different than a user rendered inactive, which can be done only by an administrator.

Procedure

1. Click Users, and then select the name of the user that you want to edit.
2. Click Edit User. The Edit User - Details dialog box appears.
3. In the Edit User - Detail dialog box, modify the fields as needed.
   Note: To modify a user's status, select the wanted, changed status in the Status list.
   Note: If you are an administrator and you are changing a user's password, click the Send reset password email link.
4. Click Save, and the edited user information is saved.

- **Parent topic:** [User management](#)

Related concepts

- [User management](#)

Viewing details of a user

Basic user information such as first and surname, email address, and role can be viewed by clicking View Details.

Procedure

1. Within the User Management page, select the name of the user for which you want to view details.
2. Click View Details, and the View User Details window appears.
3. Within the View User Details window, the following information can be seen:

Option	Description
User Name	The user name of the selected user.
First Name	The first or given name of the selected user.
Last Name	The last or surname of the selected user.
Email Address	The email address of the selected user.
Role	The role of the selected user. A role is a named collection of permissible activities within the system. Based on the assigned role of the selected user, that individual can or cannot do certain things within the system.
Status	The status of the selected user. Options include active or inactive.
Last Login	The last date and time at which the selected user logged in to the system.

4. To view the user's access log, click the View access log link. The View Access History Log window appears, showing the system entries for the selected user. Within the entries, you see system activities such as successful logins, unsuccessful logins, when a password was changed. To close the View Access History Log window, click Close.
Note: These columns can be sorted by clicking the up and down arrows to the right of the column name.
5. To close the View User Details window, click Close.

- **Parent topic:** [User management](#)

Related concepts

- [User management](#)

Deleting a user

You can delete existing users from the system as needed; however, the log entries of a deleted user are also deleted. In certain instances, it is a good practice that you change a user's status.

Procedure

1. Click Users, and then select the name of the user that you want to delete.
2. Click Delete User. The Delete User window appears.
   Note: If you opt to delete a user, then any log files that are associated with this user are also deleted.
3. To confirm that you want to delete the user, click Delete. The confirmation Delete User window appears, confirming that the user was deleted from the system.

- **Parent topic:** User management

Related concepts

- User management

Changing user passwords

Either the user or the administrator can change passwords. Valid passwords are at least 8 characters in length and contain at least an uppercase letter, a lowercase letter, a number, and a special character.

About this task

The password reset interaction can occur in four ways:

- A user clicks the Forgot password link on the login page, leading the user to change the password.
- A user clicks the Did not receive password details or Changed your email address link within the login screen from any of the applications.
- Users view their profile and opt to change the associated password.
- An administrator edits a user's profile and opts to reset the associated password. An email is then sent to the user, forcing individuals to change their password.

Procedure

1. Within the email your received regarding how to change your password, click the provided link. The Change Your Password dialog box appears.
2. In the Old password text box, enter the previously used password for this user.
   Note: If the password entered in this text box does not match the one listed in the system's database records, a red warning message appears. Reenter the old password.
3. In the New password text box, enter a new password for this user.
4. In the Re-enter new password text box, type the password again.
   Note: The passwords that are entered into the New password and Re-enter new password text boxes must be identical and meet the system's security requirements. If the entered data is not identical, a red warning appears.
5. Click Save. A congratulatory message appears, confirming that your password is changed. Click OK to close the window.

- **Parent topic:** [User management](#)

Related concepts

- [User management](#)

Resetting the user password

User passwords can be set and reset as needed by users and administrators alike.

About this task

There are two ways to reset a password:

- For administrators, click the Send reset password email link in the Edit User Details window.
- For users who reset their own passwords, use the following procedure.

Procedure

1. From the application login screen, click the Forgot your password link. The Reset Password dialog box appears.
2. In the Email Address text box of the Reset Password dialog box, enter the email address where a link to change the password must be sent. This email address must be linked with your account in IBM® StoredIQ Administrator.
3. Click Submit. A confirmation window appears, stating that an email with instructions for changing your password was sent.
4. Click OK to close the Reset Password window.

- **Parent topic:** [User management](#)

Related concepts

- [User management](#)

Obtaining a new password

Users occasionally need to obtain new passwords to be able to log in to the system.

Procedure

1. From the application's login screen, click the Did not receive a password link.
2. Contact your system administrator about obtaining a new password.

- **Parent topic:** [User management](#)

Related concepts

- [User management](#)

Action management

Actions represent executable processes that act upon indexed data within data servers. Use the Action Management page to create and modify actions that are used by IBM® StoredIQ Administrator.

With IBM StoredIQ Administrator, you can create these types of actions:

- Copy: The copy action copies infosets to a target set.
  - You can specify the directory structure and security settings for copied data objects.
  - If a data object exists with the same name, then the subsequent data objects that are encountered are renamed.
- Copy to Retention: The copy to retention action copies data objects from source volumes to a target set for a specified period.
- Delete: The delete action removes data objects from the source volume.
  - Only data objects, not directories, are deleted.
  - Use caution when you are deleting data objects.
- Discovery Export EDRM XML: The discovery export action copies data objects and generates an EDRM XML file for loading into third-party legal review tools.
- Discovery Export DAT: The discovery export action copies data objects and generates a Concordance DAT file for loading into third-party legal review tools.
- Modify Attribute: The modify attribute action simulates setting retention by manipulating specific attributes on objects.
- Move: The move action moves data objects in an infoset to another volume.
  - You can specify the directory structure and security settings for moved data objects.
  - If a data object exists with the same name, then the data object is renamed.
  - When the source equals the destination, the move is not allowed.
- Step-up Snippet: The step-up snippet action can be run on an infoset, extracting and storing text for any data object to which it is applied. These data snippets are then used by auto-classification. Infosets objects without snippets are not classified for auto-classification.
  Note: When scheduled harvests or policies are run against a data source, snippets are not updated or generated if they are missing.

Table 1. Actions column names and descriptions

Actions Column Name	Description
Action Name	Lists the name of the action.
Type	Lists the action type, namely Copy, Copy to Retention, Delete, Modify Attribute, Move, Discovery Export XML, Discovery Export DAT, and Step-up Snippet.
Last Modified	Lists the date and time at which the action was last modified.
Target Set Name	Lists the action's destined target set.
Description	Lists the action description.

- Searching for a specific action
  When you have multiple actions available to choose, you can search for a specific action instead of using the slider to move through available options.
- Creating an action
  Actions are processes that are enacted on data objects, which are defined by infosets. Actions must be defined here in IBM StoredIQ Administrator.
- Deleting an action
  Existing actions can be deleted from the list of available actions.
- Cloning an action
  Actions can be cloned or duplicated.
- Editing an action
  Existing actions can be modified.

Searching for a specific action

When you have multiple actions available to choose, you can search for a specific action instead of using the slider to move through available options.

Procedure

1. In the Enter key terms... text box, enter the name of the action for which you want to search, and then press Search. You can enter either the full action name or a portion of it. To remove the search term, click the X to the left of the Enter key terms.... text box. Actions that match the entered search term are returned.
2. Edit, clone, or delete the action as needed.

- **Parent topic:** Action management

Related concepts

- Action management

Creating an action

Actions are processes that are enacted on data objects, which are defined by infosets. Actions must be defined here in IBM® StoredIQ Administrator.

About this task

The step-up snippet action is a built-in action, meaning that no options are provided within the user interface. You cannot create or edit a step-up snippet action as it is provided for you as a prepopulated option on the Actions tab.

Procedure

1. Click Actions, and then click Create Action. The Add Action dialog box appears.
2. In the Add Action dialog box, complete these fields.
   a. In the Action Name text box, enter the name of this action.
   b. In the Description of Action text box, enter a brief description of this action. This action might need to be used by several people, so an accurate description can be helpful to others.
   c. In the Action Type list, select the action type.
   d. If you are creating a copy, copy to retention, discovery export, modify attribute, or move action, select a target set from the created actions or create a new target set. In Target Set, select a target set or click Create a new target set.
   e. Click Next to continue creating the action. Depending on the type of action that is selected, use the corresponding tables to better understand what options to select.

Copy action options	Action
Target Set	In the Target Set area, select a target set.
Destination Directory	In the Destination Directory text box, either enter a file path or use the list option to select a previously entered file path.
Directory Structure	In Directory Structure, select the check box if you want to Recreate directory structure on destination.
Hashes	In Hashes, select the check box if you want to Compute a hash value for each data object.
Harvesting	In Harvesting, select or clear the check box for Do not auto-harvest destination volume after copy.
Email Message Format	In the Email Message Format list, select the format of the email message.
SharePoint User Profile Items	In SharePoint User Profile Items, select the check boxes of application SharePoint items, including Notes, Libraries, Blog Posts, Wikis, and Misc.

Copy to Retention action options	Action
Target Set	In the Target Set area, select a target set.
Hashes	In Hashes, select the check box if you want to Compute a hash value for each data object.
Harvesting	In Harvesting, select the check box to Do not auto-harvest destination volume after copy if an automatic harvest of the retention volume is not wanted. If this option is selected, then the retention volume can be harvested with the discover retention volumes job on the data server.

Email Message Format	In the Email Message Format list, select the format of the email message.
Retention Tag	In Retention Tag, enter the key term for the tag.
Retention Period	In Retention Period, enter a number, and then select Days, Weeks, Months, or Years.
Delete action options	**Action**
Delete Parameters	In Delete Parameters, select the check box to Do not delete data objects that have been accessed since last harvest.
Discovery Export EDRM XML action options	**Action**
Destination Directory	In the Destination Directory text box, either enter a file path or use the list option to select a previously entered file path.
Unique ID Prefix and # of Digits	In Unique ID Prefix and # of Digits, enter a unique prefix and then the number of digits to follow that prefix. For example, if you enter a prefix of `feb2012` with 7 digits, this entry yields a Unique ID similar to `Feb2012-1234567`. Select the check box to Pad Zeros. Select the check box to Save a text copy.
Email Message Format	In the Email Message Format list, select the format of the email message.
Limit Export Process	Select the check box to Limit Export Process to and enter the number of data objects.
SharePoint User Profile Items	In SharePoint User Profile Items, select the check boxes of application SharePoint items, including Notes, Libraries, Blog Posts, Wikis, and Misc.
Discovery Export DAT action options	**Action**
Destination Directory	In the Destination Directory text box, either enter a file path or use the list option to select a previously entered file path.
Unique ID Prefix and # of Digits	In Unique ID Prefix and # of Digits, enter a unique prefix and then the number of digits to follow that prefix. For example, if you enter a prefix of `feb2012` with 7 digits, this entry yields a Unique ID similar to `Feb2012-1234567`. Select the check box to Pad Zeros.
Email Message Format	In the Email Message Format list, select the format of the email message.
Limit Export Process	Select the check box to Limit Export Process to and enter the number of data objects.
SharePoint User Profile Items	In SharePoint User Profile Items, select the check boxes of application SharePoint items, including Notes, Libraries, Blog Posts, Wikis, and Misc.

Items		
Modify Attribute action options	**Action**
Action Name	In the Action Name, enter the action's name.
Action Type	In the Action Type list, select the Modify Attribute option.
Description of Action	In the Description of Action text box, enter the action's description.
Attribute Type	Select from the following options: Read-only or Read-write.
Move action options	**Action**
Destination Directory	In the Destination Directory text box, either enter a file path or use the list option to select a previously entered file path.
Directory Structure	In Directory Structure, select the check box if you want to Recreate directory structure on destination.

3. Click Save to save your changes, click Back to return to the previous page, or click Cancel to cancel creating the action.

- **Parent topic:** [Action management](#)

Related concepts

- [Action management](#)

Deleting an action

Existing actions can be deleted from the list of available actions.

Procedure

1. In the List of Actions page, select the action that you want to delete, and then click Delete Action. The Delete Action window appears.
2. Click Delete to delete the selected action. The action is removed from the List of Actions page and it is no longer available for use.

- **Parent topic:** [Action management](#)

Related concepts

- [Action management](#)

Cloning an action

Actions can be cloned or duplicated.

Procedure

1. In the List of Actions page, select the action that you want to clone, and then click Clone Action. The Clone Action dialog box appears.
2. In the Clone Action dialog box, complete these fields.
   a. In the Action Name text box, enter the name of this action.
      Note: The name of the cloned action must be different from the name of the original action.
   b. In the Action Type list, select the action type, and then click Next to continue creating the action.
   c. In the Description of Action text box, enter a brief description of this action.
      Note: This action might be used by several people, so the description must be accurate.
   d. The selection that is made in the Action Type list determines what options are available after you click Next. If you are creating a copy, copy to retention, discovery export, or move action, select a target set from created target sets or create a new target set. In Target Set, select a target set or click Create a new target set.
   e. Click Next. Depending on the type of action that is being cloned, you can see various options.
3. Click Save to save your changes.

- **Parent topic:** [Action management](#)

Related concepts

- [Action management](#)

Editing an action

Existing actions can be modified.

Procedure

1. In the List of Actions page, select the action that you want to edit, and then click Edit Action. The Edit Actions dialog box appears.
2. In the Edit Actions dialog box, edit the action as needed.
3. Click Save to save your changes.

- **Parent topic:** [Action management](#)

Related concepts

- [Action management](#)

Target sets

Some actions, namely copy actions, require a destination to complete, and a target set represents a mapping for each data server of the wanted target volume. To copy files, the data server must know where to copy the file to. Each data server must be configured with a list of target volumes, such as retention and discovery export volumes.

Target Set Example
     Data Server 1 has three retention volumes: Volume A, Volume B, and Volume C.
     Data Server 2 has two retention volumes: Volume B, Volume X, and Volume Z.
     Data Server 3 has two retention volumes: Volume B and Volume Z.
Resulting Copy Action: You would like to create a target set for a copy action
     Data Server 1 copies to Volume A.
     Data Server 2 copies to Volume X.
     Data Server 3 copies to Volume B.

Another user might want to create a target set for a copy action where Data Server 1, Data Server 2, and Data Server 3 copies to Volume B.

Table 1. Target set management columns and descriptions

Target set column name	Description
Target Set Name	Lists the name of the target set.
Type	Lists the type of the target set, namely Primary, Retention, or Export.
Source Type	Lists the target set's source type, for example, CIFS, NFS, or SharePoint.
Volumes	Lists the number of volumes in that target set.
Last Modified	Lists the date and time that the target set was last modified.
Description	Lists the description of the target set.

- [Searching for a specific target set](#)
  When you have multiple target sets from which to choose, you can search for a specific target sets instead of using the slider to move through available options.
- [Creating a target set](#)
  Retention, copy, move, and discovery export actions require a named destination to which data objects can be copied, moved, or exported. Target sets associate data servers and appropriate volumes, providing your actions with a destination.
- [Editing a target set](#)
  Target sets can be edited, modifying the data servers that are associated with the target set.
- [Cloning a target set](#)
  Target sets can also be cloned or duplicated.
- [Deleting a target set](#)
  Existing target sets can be deleted from the list of available target sets.

Searching for a specific target set

When you have multiple target sets from which to choose, you can search for a specific target sets instead of using the slider to move through available options.

Procedure

1. In the Enter key terms... text box, enter the name of the target set for which you would like to search, and then press Search. You can enter either the full target set name or a portion of it. To remove the search term, click the X to the left of the Enter key terms.... text box. Target sets that match the entered search term are returned.
2. Edit, clone, or delete the target set as needed.

- **Parent topic:** Target sets

Related concepts

- Target sets

Creating a target set

Retention, copy, move, and discovery export actions require a named destination to which data objects can be copied, moved, or exported. Target sets associate data servers and appropriate volumes, providing your actions with a destination.

Procedure

1. Click Target Sets, and then click Create Target Set. The Create Target Set dialog box appears.
2. In the Create Target Set dialog box, complete these fields.
    a. In the Target Set Name text box, enter a name for this target set.
    b. In the Description of Target Set text box, enter a description.
    c. In the Type of Target Set list, select the type of this target set, namely Primary, Retention, or Export.
    d. In the Source Type list, select the source type of this target set, namely CIFS, NFS, FileNet, or SharePoint.
    e. In the Target Set List area, click Add Data Servers. Select a data server and then click Add to move it to the Target Set List box. Click Done when you finish adding data servers for the target set. Use the Ctrl key to add or remove several selections simultaneously. The added data servers appear within the Target Set List box. Click Add/Remove Data Servers to modify what data servers appear here.
    f. In the Create Target Set dialog box, click the added data server and then select a volume on that data server. Do this step for each added data server.
3. Click Save to save your configuration changes and create the target set.

- **Parent topic:** Target sets

Related concepts

- Target sets

Editing a target set

Target sets can be edited, modifying the data servers that are associated with the target set.

Procedure

1. On the List of Target Sets page, click the name of the target set that you would like to edit, and then click Edit Target Set. The Edit Target Set dialog box appears.
2. Edit the target set fields as needed.
3. Click Save to save your changes.

- **Parent topic:** Target sets

Related concepts

- Target sets

Cloning a target set

Target sets can also be cloned or duplicated.

Procedure

1. On the List of Target Sets page, click the name of the target set that you would like to clone or duplicate, and then click Clone Set. The Clone Target Set dialog box appears.
2. Edit the target set fields as needed.
   Note: The cloned target set's name must be unique.
3. Click Save to save your changes.

- **Parent topic:** Target sets

Related concepts

- Target sets

Deleting a target set

Existing target sets can be deleted from the list of available target sets.

Procedure

1. In the List of Target Sets page, select the target set you would like to delete, and then click Delete Set. The Delete Target Set window appears.
2. Click Delete to delete the selected target set. The target set is removed from the List of Target Sets page, meaning that it is no longer available for use.

- **Parent topic:** Target sets

Related concepts

- Target sets

Reports

IBM® StoredIQ contains a number of default reports and the ability to upload a report package.

IBM StoredIQ provides features to inventory unstructured data. It can then collect and evaluate unstructured date, compiling reports and jobs to act on the data according to an organization's needs. Reports can then be printed and disseminated, allowing teams to analyze data in an ad hoc manner. New reports can be added by uploading a report package.

IBM StoredIQ Administrator provides a number of built-in reports, such as summaries of data objects in the system, storage use, and the number of identical documents in the system. You can create custom reports, including Query Analysis Reports for e-discovery purposes, and automatically email report notifications to administrators and other interested parties. By default, these types of reports are available.

- Compliance Report
- CSV Exception List Export
- CSV Infoset Data Object Export
- Data Assessment Report
- Data Topology Report
- Duplication Summary Report
- Overlay Hit Report

Table 1. List of Reports options and descriptions

Report column option	Description
Name	Name of the report
Created	The date and time at which the report was created
Description	A description of the report

Note: For help in customizing IBM StoredIQ reports, contact your technical service representative.

- Searching for a specific report
  When you have multiple reports from which to choose, you can search for a specific report instead of using the slider to move through available options.
- Creating a report

Searching for a specific report

When you have multiple reports from which to choose, you can search for a specific report instead of using the slider to move through available options.

Procedure

1. In the Enter key terms... text box, enter the name of the report for which you would like to search, and then press Search. You can enter either the full report name or a portion of it. To remove the search term, click the X to the left of the Enter key terms.... text box. Reports that match the entered search term are returned.
2. Select the wanted report.

- **Parent topic:** [Reports](#)

Related concepts

- [Reports](#)

Creating a report

Procedure

1. Click Reports, and then click Create. The Create Report dialog box appears.
2. In the Create Report dialog box, complete these fields.
   a. In the Name text box, enter a name for this report.
   b. In the Description text box, enter a description of this report.
   c. Click Choose File, and then find and select the appropriate report-package file.
3. Click Create, and the newly created report is added to the List of Reports page.

- **Parent topic:** [Reports](#)

Related concepts

- [Reports](#)

Auto-classification models

Auto-classification models, also called automated document categorization, integrates the IBM® Content Classification's classification model into the IBM® StoredIQ infoset-generation process.

IBM Content Classification

IBM Content Classification helps organize unstructured content by analyzing the full text of documents and emails and applying rules that automate classification decisions. By managing documents and email, you organize and act on content every day. IBM Content Classification reduces the burden of manual decision making that is done by employees by accurately and automatically organizing information. Embedded with natural language processing and semantic analysis capabilities, IBM Content Classification determines the true intent of words and then uses that knowledge to automate decision making. Unlike other classification systems that are based on rules only, IBM Content Classification combines rules and contextual analysis to incorporate real-time learning that adapts to changing business needs. As a result, classification becomes accurate over time.

IBM Content Classification can organize information by policies or keywords, but it can also assign metadata that is based on the full context of the document. The classification process not only searches for a single word or phrase, but also analyzes the entire document. It then distills the main point of the text and assigns the text to a category. When it analyzes content, IBM Content Classification can recognize misspellings, abbreviations, jargon, and technical terms.

Accuracy improves over time because the system adapts to the unique nature of your business by identifying different categories from examples that you provide. When you provide feedback, the system adjusts in real time and immediately implements any corrections that you make. The accuracy of the classification results keeps pace with changes in your business.

IBM Content Classification combines this context-based approach with a rule-based, decision-making approach. The system can identify keywords, patterns, such as account numbers and phone numbers of case identifiers, and words within a certain proximity of each other. For example, the system can identify occurrences of the phrase "Attorney General" in the same sentence as the word "California". When content that matches a condition in a rule is detected, the action that is defined for the rule is applied, and the document or email is classified.

Examples of classification applications

You can use IBM Content Classification to achieve various business goals.

- Enterprise content standardization: To support document classification and taxonomy automation within your content management system, document properties or metadata can be automatically assigned when the content is classified. Documents can be automatically moved to the correct enterprise repository.
- Compliance and records management: Documents and email can be declared as records when they are classified and placed under the control of record retention policies and compliance standards.
- E-discovery readiness: Documents and email can be filtered to ensure that only items with business value are classified and archived. You can quickly and cost effectively prepare content for potential legal notices.
- Business process optimization: Automated decision making ensures more consistent outcomes and reduced costs. With content-based analysis, you accomplish these actions.
    - Insert documents into the workflow of a business plan.
    - Reroute email.
    - Suggest and apply agent responses within a customer-support center.

Using auto-classification, you can create and update classification models to help map data within infosets. The step-up snippet action extracts and stores text on data objects within an infoset. For auto-classification to work, you must use the step-up snippet action since infosets objects without extracted snippets do not have auto-classification scores.

The Data Expert can use IBM Content Classification to train a classification model. The classification model is then registered with the IBM StoredIQ Platform system with IBM StoredIQ Administrator. The registered classification model can be applied to an existing infoset in IBM StoredIQ Data Workbench to generate new metadata for the objects in the infoset. Finally, this metadata can be used in rule-based filters to create new infosets.

Note: To use the IBM StoredIQ Administrator auto-classification function, users must have a licensed installation of the IBM Content Classification on their desktop.

- [Applying the auto-classification feature](#)
  To use the auto-classification feature successfully, you must work in both IBM StoredIQ Administrator and IBM StoredIQ Data Workbench.
- [Data server sizing-configuration guidelines](#)
  The inclusion of the IBM Content Classification server on the data server changed the required memory and processors. Data server sizing-configuration guidelines are provided here.
- [Downloading the classification export utility](#)
  The classification export utility is a one-time-only download that must be installed before an importing a model.
- [Building an auto-classification model](#)
  Creating boolean-based rules to find documents that fit differing categories is time-consuming and requires constant, meticulous adjustments. However, importing a classification model with sets of training documents helps find other, similar documents.
- [Preparing an IBM Content Classification model for import](#)
  You must prepare the IBM Content Classification model for importing into IBM StoredIQ Platform with the IBM Content Classification tool BundleDPKB.exe.
- [Importing an auto-classification model](#)
  You must complete these tasks before you can import an auto-classification model.
- [Searching for a specific classification model](#)
  When you have multiple classification models to choose, you can search for a specific classification model instead of using the slider to move through available options.
- [Editing a model](#)
  The terms and formats that are associated with a particular classification model change over time, which means that the classification model must also change. Classification models can be edited or updated, modifying the model file that is associated with that particular classification model.
- [Deleting a model](#)
  Existing classification models can be deleted from the list of available models, if it was not used to enhance an infoset and that infoset still exists.

Applying the auto-classification feature

To use the auto-classification feature successfully, you must work in both IBM® StoredIQ Administrator and IBM StoredIQ Data Workbench.

Before you begin

Prior to preparing an IBM Content Classification model for import, you must finish downloading the classification export utility, which must be done only one time. The export utility is used for each IBM Content Classification model that you want to import.

About this task

IBM Content Classification works by taking a sample set of documents that you adjust, constructing a classification model, and then classifying a new document. The IBM StoredIQ Platform defines the classification model as a decision plan with at least one knowledge base that is referenced by the decision plan.

Procedure

1. In IBM StoredIQ Administrator, on the Action tab, the step-up snippet action is created automatically and appears in IBM StoredIQ Data Workbench.
2. In IBM StoredIQ Administrator, create an auto-classification model by importing a compressed file from IBM Content Classification.
3. In IBM StoredIQ Data Workbench, start the step-up snippet action, which is available on the Act tab, on an infoset.
   Note: The infoset must be a user infoset, not a system infoset.
4. While still in IBM StoredIQ Data Workbench, apply the wanted auto-classification enhancement, which is available on the Enhance tab, to the same infoset.
5. In IBM StoredIQ Data Workbench, apply the auto-class filter, which is available on the Create > Build Filter area to the infoset, refining and creating a new infoset.
6. Review the results of the newly created infoset.

- **Parent topic:** Auto-classification models

Related concepts

- Auto-classification models
- Data server sizing-configuration guidelines

Data server sizing-configuration guidelines

The inclusion of the IBM® Content Classification server on the data server changed the required memory and processors. Data server sizing-configuration guidelines are provided here.

To support the auto-classification function, some of the data server configuration parameters changed. You must increase the virtual machine settings for processors and memory on data servers to these new minimums as part of this update:

- Memory: 16 GB
- vCPUs: 4

Required storage on the data server for the extracted text (snippet cache) is roughly 13% of the size of the uncompressed content, or 130 GB per 1 TB. The size can vary depending on the object types. The maximum amount of extracted text per item is configured to not exceed 2 MB by default.

A data server with 2 TB of configured storage can support approximately 10 TB of managed storage, assuming 50 million data objects with an average size of 200 KB. Approximately 300 GB is metadata storage with the remaining 1,700 GB supporting a combination of full text and extracted text.

- **Parent topic:** Auto-classification models

Related tasks

- Applying the auto-classification feature
- Downloading the classification export utility
- Building an auto-classification model
- Preparing an IBM Content Classification model for import
- Importing an auto-classification model
- Searching for a specific classification model
- Editing a model
- Deleting a model

Downloading the classification export utility

The classification export utility is a one-time-only download that must be installed before an importing a model.

Procedure

1. In IBM® StoredIQ Administrator on the Auto-Classification tab, click Download Classification Export Utility. The siq-classification-export.zip file is downloaded.
2. Extract the export utility to the bin directory of your IBM Content Classification product installation. For Windows, it is typically in the C:\IBM\ContentClassification\Bin directory.

- **Parent topic:** Auto-classification models

Related concepts

- Auto-classification models
- Data server sizing-configuration guidelines

Building an auto-classification model

Creating boolean-based rules to find documents that fit differing categories is time-consuming and requires constant, meticulous adjustments. However, importing a classification model with sets of training documents helps find other, similar documents.

About this task

Using previously harvested customer data, you can create an auto-classification model.

Procedure

1. Determine the categories into which you want the auto-classification model to classify documents.
2. Using IBM® StoredIQ Data Workbench, create a filter for each category to capture documents that are representatives of the category.
3. For each filter, create an infoset. The members of the resulting infoset become the "training corpus" for the category.
4. For each infoset, run a copy action with IBM StoredIQ Data Workbenchonto a folder that is accessible by the IBM Content Classification application.
5. Use the IBM Content Classification application to create a decision plan and knowledge base by importing the training corps that you created.
   Note: A classification model consists of one decision plan and at least one knowledge base, which is a requirement of the IBM StoredIQ Platform auto-classification feature.

- **Parent topic:** [Auto-classification models](#)

Related concepts

- [Auto-classification models](#)
- [Data server sizing-configuration guidelines](#)

Preparing an IBM Content Classification model for import

You must prepare the IBM® Content Classification model for importing into IBM StoredIQ Platform with the IBM Content Classification tool BundleDPKB.exe.

Before you begin

IBM StoredIQ Platform defines the IBM Content Classification model as a set of a single decision plan and one or more knowledge bases that are referenced by the decision plan. You must prepare each classification model for import by bundling the relevant decision plan and knowledge base with the IBM Content Classification export utility BundleDPKB.exe.

Note: Before you prepare an IBM Content Classification model for import, you must finish downloading and installing the classification export utility.

Procedure

1. In a command interface, go to the BundleDPKB installation directory and then enter `cscript.exe ClassificationModelExporter.vbs <location of decision plan projects and knowledge bases> <decision plan directory> <optional empty directory> <optional zip file name>`. If you did not specify a directory to which the compressed file must be exported, an export directory is created in the <location of decision plan projects and knowledge bases> location.
2. Note the location and name of the generated compressed file.

- **Parent topic:** [Auto-classification models](#)

Related concepts

- [Auto-classification models](#)
- [Data server sizing-configuration guidelines](#)

Importing an auto-classification model

You must complete these tasks before you can import an auto-classification model.

Before you begin

- Download the classification export utility. The procedure must be done only one time.
- Prepare the IBM® Content Classification model for import. This procedure must be completed for every model you want to import.

About this task

To import a classification model, go to the Auto-classification tab of the IBM StoredIQ Administrator. The tab displays the already imported models as described in this table.

Table 1. Classification model column names and descriptions

Classification model column name	Description
Classification model name	Lists the name of the classification model.
Attribute name	Lists the name of the attribute that is used by the classification model.
Status	Lists the classification model's status, specifically Available, or Pending.
Creation date	Lists the date and time at which the classification model was created.
Description	Lists the description of the classification model.

Procedure

1. Within IBM StoredIQ Administrator, click Create Model. The Create Model dialog box appears.
2. In the Classification Model Name text box, enter a unique name for this auto-classification model.
3. In the Attribute Name text box, enter a unique name for this attribute. The Attribute Name must be unique within auto-classifications as it cannot be changed after the classification model is imported.
4. In the Description text box, enter a description of this model.
5. In the Auto-Classification Model File text box, click Browse and select the classification model file that you want to use when you are creating the model. Generally, this file is a compressed file that is created in the classification export utility. It consists of elements that are created in IBM Content Classification by the Data Expert.
6. Click Save, and the newly created model appears in the list of classification models.

- **Parent topic:** Auto-classification models

Related concepts

- Auto-classification models
- Data server sizing-configuration guidelines

Searching for a specific classification model

When you have multiple classification models to choose, you can search for a specific classification model instead of using the slider to move through available options.

Procedure

1. In the Enter key terms(s) text box, enter the name of the classification model for which you want to search, and then press Search. You can enter either the full classification model name or a portion of it. To remove the search term, click the X to the left of the Enter key term(s) text box. Classification models that match the entered search term are returned.
2. Edit, clone, or delete the classification model as needed.

- **Parent topic:** Auto-classification models

Related concepts

- Auto-classification models
- Data server sizing-configuration guidelines

Editing a model

The terms and formats that are associated with a particular classification model change over time, which means that the classification model must also change. Classification models can be edited or updated, modifying the model file that is associated with that particular classification model.

Procedure

1. On the List of Classification Models page of Auto-classification tab, click the name of the classification model that you want to edit, and then click Edit Model. The Edit Model window appears.
2. Edit the classification model fields as needed.
   Note: Attribute names cannot be changed, but you can edit the name of the model without specifying a new model (.ZIP) file. If you attempt to modify a classification model that is being used, a warning appears. Click OK to return to the List of Classification Models page.
3. Click Save to save your changes.

- **Parent topic:** Auto-classification models

Related concepts

- Auto-classification models
- Data server sizing-configuration guidelines

Deleting a model

Existing classification models can be deleted from the list of available models, if it was not used to enhance an infoset and that infoset still exists.

Procedure

1. In the List of Classification Models page, select the classification model that you want to delete, and then click Delete Model. The Delete Model window appears.
2. Click Delete to delete the selected model. The classification model is removed from the List of Classification Models page. It is no longer available for use.

- **Parent topic:** [Auto-classification models](#)

Related concepts

- [Auto-classification models](#)
- [Data server sizing-configuration guidelines](#)

Concept management

Within IBM® StoredIQ Platform, you can use the concept-management feature to relate business ideas to indexed data.

When you are using IBM StoredIQ Administrator, you use infosets, run actions, generate reports, and more generally, interact with IBM StoredIQ Platformas a whole. These technical aspects of using IBM StoredIQ Administratorare designed for more technical users. Concepts are designed for the business user and provide a user experience of creating a filter that finds documents that are owned by a particular user name, for example. This use of the product is oriented towards solving business problems. Using the product typically requires the user to translate business concepts like identity, department, vendor, and project into the various technical concepts of the product, such as filters and infosets.

A concept defines and represents an identity, custodian, vendor, and project that can be related to indexed data in a meaningful way. It is a set of attributes that describe and distinguish individual concept members. For example, an employee concept might define the attribute's user name, given name, surname, and email address.

Concept members are the granular building blocks of a concept. When concept members are defined, they can be gathered into groups for aggregate use.

The Identity concept is created as a predefined, preconfigured concept available within IBM StoredIQ eDiscovery.

The List of Concepts screen is simple, depicting concepts, identities, and aspects. From here, you can search for, delete, edit, and add identities.

Table 1. Concept columns and descriptions

Concept Column Name	Description
First name	Lists the given name.
Last name	Lists the surname.
Email address	Lists the email address.
Owner ID	Lists the owner identification number.
Company	Lists the member's company name.
Department	Lists the department of the member.

- [Adding a member to a concept](#)
  Members can be added to existing concepts.
- [Searching for a concept](#)
  When you have multiple concepts available to choose, you can search for a specific concept instead of using the slider to move through available options.
- [Editing a member of a concept](#)
  Existing concepts and their members can be modified.
- [Deleting a member of a concept](#)
  Existing concept members can be deleted from the list of available concepts.

Adding a member to a concept

Members can be added to existing concepts.

Procedure

1. Click Concepts.
2. From the list, select the type of concept you would like to add, and then click Create Member. The Identity concept was created as a predefined, preconfigured aspect. The Add Identity dialog box appears.
3. In the Add Identity dialog box, complete these fields.
   a. In the First Name text box, enter the identity's given name.
   b. In the Last Name text box, enter the identity's surname.
   c. In the E-mail Address text box, enter the identity's email address.
   d. In the Company text box, enter the name of the company that is associated with this identity.
   e. In the Department text box, enter the name of the identity's department.
   f. In the Owner ID text box, enter the identity's owner ID.
4. Click Cancel to cancel adding an identity and return to the List of Concepts screen, or click Save to save the added member and return to the List of Concepts page.

* **Parent topic:** [Concept management](#)

Related concepts

* [Concept management](#)

Searching for a concept

When you have multiple concepts available to choose, you can search for a specific concept instead of using the slider to move through available options.

Procedure

1. In the Enter key terms... text box, enter the name of the concept for which you want to search, and then press Search. You can enter either the full concept name or a portion of it. To remove the search term, click the X to the left of the Enter key terms.... text box. Concepts that match the entered search term are returned.
2. Edit or delete the concept as needed.

- **Parent topic:** Concept management

Related concepts

- Concept management

Editing a member of a concept

Existing concepts and their members can be modified.

Procedure

1. From the List of Concept page, select the concept that you want to edit, and then click Edit. The Identity concept was created as a predefined, preconfigured aspect. The Edit Identity dialog box appears.
2. Edit the concept member and click OK to save your changes.

- **Parent topic:** [Concept management](#)

Related concepts

- [Concept management](#)

Deleting a member of a concept

Existing concept members can be deleted from the list of available concepts.

About this task

Note: If you delete a member of a concept, existing filters that use that concept's members are affected.

Procedure

1. From the List Concepts page, select the concept member that you want to delete, and then click Delete. The Delete Confirmation window appears.
2. Click OK to delete the selected concept member. The member of the concept is removed from the table and is no longer available for use.

- **Parent topic:** [Concept management](#)

Related concepts

- [Concept management](#)

Administering IBM StoredIQ Platform Data Server

The following topics provide instructions on how to conduct administrative tasks such as administering appliance, configuring IBM® StoredIQ Platform, and creating volumes and data source.

- [IBM StoredIQ Platform user interface](#)
  The chapter provides an introduction to the IBM StoredIQ Platform web interface. It outlines the features within each tab and provides references to sections where you can find additional information on each topic.
- [System administration](#)
  System administration entails checking the system's status and restarting the appliance.
- [Configuration of IBM StoredIQ Platform](#)
  This section describes the Application and Network areas that an administrator can modify to configure IBM StoredIQ Platform.
- [Volumes and data sources](#)
  Volumes or data sources are integral to IBM StoredIQ Platform indexing your data.
- [Data harvesting](#)
  Harvesting (or indexing) is the process or task by which IBM StoredIQ Platform examines and classifies data in your network.
- [Job configuration](#)
  Jobs have different functions, and depending on the job type that is chosen, different system outcomes occur. Several jobs that are ready for use are also included.
- [Desktop collection](#)
  When you configure desktop settings, you are enabling or disabling encryption within IBM StoredIQ Platform. The IBM Desktop Data Collector (desktop client or client) enables desktops as a volume type or data source, allowing them to be used just as other types of added data sources. The IBM Desktop Data Collector is provided as a standard MSI file and is installed according to the typical method (such as Microsoft Systems Management Service (SMS)) used within your organization. The IBM Desktop Data Collector can collect PSTs, compressed files, and other data objects and can remove itself when its work is completed.
- [Folders](#)
  This section provides both conceptual and procedural information about folders and their usage.
- [Audits and logs](#)
  The following section describes the audit and log categories in the system, including descriptions of the various audit types and how to view and download details.
- [Deploying customized web services](#)
  This procedure highlights the basic steps that are required to deploy SharePoint custom web services.
- [Supported file types](#)
  The following section provides a comprehensive list of the file types that can be harvested and processed by IBM StoredIQ Platform, organized by name and by category. You can also view SharePoint attributes.
- [Supported server platforms and protocols](#)
  The following section lists the supported server platforms by volume type and the protocols for supported systems.
- [Event log messages](#)
  The following section contains a complete listing of all ERROR, INFO, and WARN event-log messages that appear in the Event Log of the IBM StoredIQ Platform console.

- **Parent topic:** [Administering IBM StoredIQ](#)

IBM StoredIQ Platform user interface

The chapter provides an introduction to the IBM® StoredIQ Platform web interface. It outlines the features within each tab and provides references to sections where you can find additional information on each topic.

- Navigation within IBM StoredIQ Platform
  The primary tabs and subtabs found within the user interface provide you with the access to data server functionality.
- Web interface icons
  The following tables describe the icons that are used throughout IBM StoredIQ Platform web interface.

Navigation within IBM StoredIQ Platform

The primary tabs and subtabs found within the user interface provide you with the access to data server functionality.

Primary IBM StoredIQ Platform tabs

IBM® StoredIQ Platform users do most tasks with the web interface. The menu bar at the top of the interface contains three primary tabs that are described in this table.

Table 1. IBM StoredIQ Platform primary tabs

Tab name	Description
Administration	Allows Administrators to do various configurations on these subtabs: Dashboard, Data Sources, and Configuration.
Folders	Create folders and jobs; run jobs.
Audit	Examines a comprehensive history of all harvests, run policies, imports, and event logs.

Administration tab

The Administration tab includes these subtabs: Dashboard, Data Sources, and Configuration.

- Dashboard: The Dashboard subtab provides an overview of the system's current, ongoing, and previous processes and its status. This table describes administrator-level features and descriptions.
- Data sources: The Data sources subtab is where administrators define servers and volumes. They can be places that are indexed or copied to. Various server types and volumes can be configured for use in managing data. Administrators can add Enterprise Vault sites, Centera pools, Dell DX Storage Clusters, Celerra, NetApp, and FileNet servers through the Specify servers area. Volumes are configured and imported in the Specify volumes section.
- Configuration: The administrator configures system and application settings for IBM StoredIQ Platform through the Configuration subtab.

Table 2. Dashboard subtab settings and descriptions

Dashboard setting	Description
Page refresh	Choose from 30-second, 60-second, or 90-second intervals to refresh the page.
Today's job schedule	View a list of jobs that are scheduled for that day with links to the job's summary.
System summary	View a summary of system details, including system data objects, contained data objects, volumes, and the dates of the last completed harvest.
Jobs in progress	View details of each job step as it is running, including estimated time to completion, average speed, total system and contained objects that are encountered, harvest exceptions, and binary processing information.
Harvest statistics	Review the performance over the last hour for all harvests.
Event log	Review the last 500 events or download the entire event log for the current date or previous dates.
Appliance status	Provides a status view of the appliance. Restart the appliance through the about appliance link. View cache details for volumes and discovery exports.

Table 3. Configuration settings and descriptions

Configuration setting	Description

- DA Gateway settings: Configure the DA Gateway host or IP address

System	• DA Gateway settings: Configure the DA Gateway host or IP address. • Network settings: Configure the private and public network interfaces. • Mail server settings: Configure what mail server to use and how often to send email. • SNMP settings: Configure Simple Network Management Protocol (SNMP) servers and communities. • System time and date: Set the system time and date on the appliance. • Manage users: Add, remove, and edit users. • Lotus Notes user administration: Add a Lotus Notes User.
Application	• Harvester settings: Set basic parameters and limits, data object extensions and directories to skip, and reasons to run binary processing. • Full-text settings: Set full-text search limits for length of word and numbers and edit stop words. • Data object types: Set the object types that appear in the disk use by data object type report. • Audit settings: Configure how long and how many audits are kept. • Hash settings: Configure whether to compute a hash when harvesting and which hash. • Desktop settings: Configure the desktop collection service.

Folders tab

Within the Folders tab, any type of user can create and manage application objects.

Audit tab

IBM StoredIQ Platform audit feature allows Administrators to review all actions that are taken with the data server, including reviewing harvests and examining the results of actions.

- **Parent topic:** [IBM StoredIQ Platform user interface](#)

Related concepts

- [IBM StoredIQ Platform user interface](#)

Web interface icons

The following tables describe the icons that are used throughout IBM® StoredIQ Platform web interface.

IBM StoredIQ Platform icons

Table 1. IBM StoredIQ Platform dashboard icons

Dashboard icon	Description
User account	The User account icon accesses your user account, provides information about version and system times, and logs you out of the system. For more information, see [Logging In and Out of the System](#).
Inbox	The inbox link provides you the access to the PDF audit reports.
Help	Clicking the Help icon loads IBM StoredIQ Platform technical documentation in a separate browser window. By default, the technical documentation is loaded as HTML help.

Folders icons

Table 2. IBM StoredIQ Platform Folders icons

Folder icon	Description
New	Use the New icon to add jobs and folders.
Action	Use the Action icon to act on workspace objects, including the ability to move, and delete jobs and folders.
Job	Jobs tasks such as harvesting are either a step or a series of steps. For more information, see [Job configuration](#).
Folder	Folders are a container object that can be accessed and used by administrators. For more information, see [Folders](#).
Folder Up	Folders are a container object that can be accessed and used by administrators. By default, you view the contents of the Workspace folder; however, by clicking this icon, you move to the parent folder in the structure. For more information, see [Using Folders](#).

Audit icons

No specialized icons are used on the Audit tab.

- **Parent topic:** [IBM StoredIQ Platform user interface](#)

Related concepts

- [IBM StoredIQ Platform user interface](#)

System administration

System administration entails checking the system's status and restarting the appliance.

- [Checking the system's status](#)
  You can check the system's status for information about various appliance details.
- [Restarting the system](#)
  Restarting the system must be done periodically.

Checking the system's status

You can check the system's status for information about various appliance details.

Procedure

1. Go to Administration > Dashboard > Appliance status.
2. Click About appliance to open the Appliance details page. The Appliance details page shows the following information:
   - Node
   - Harvester processes
   - Status
   - Software version
   - View details link
3. Click the View details link for the controller. The following table defines appliance details and describes the data that is provided for the node.

Option	Description
View appliance details	Shows software version and details of harvester processes running on the controller for the appliance.
Application services	Shows a list of all services and status, which includes this information: ○ Service: the name of each service on the appliance component ○ PID: the process ID associated with each service ○ Current memory (MB): the memory that is being used by each service ○ Total memory (MB): total memory that is being used by each service and all child services ○ Processor percentage: the percentage of processor usage for each service. This value is zero when a service is idle. ○ Status: the status of each service. Status messages include Running, Stopped, Error, Initializing, Unknown.
System services	Shows a list of basic system information details and memory usage statistics. System information includes this information: ○ System time: current time on the appliance component ○ GMT Offset: the amount of variance between the system time and GMT. ○ Time up: the time the appliance component was running since the last restart, in days, hours, minutes, and seconds ○ System processes: the total number of processes that are running on the node ○ Number of processors: the number of processors in use on the component ○ Load Average (1 Minute): the average load for system processes during a 1-minute interval ○ Load Average (5 Minutes): the average load for system processes during a 5-minute interval ○ Load Average (10 Minutes): the average load for system processes during a 10-minute interval Memory details include this information: ○ Total: total physical memory on the appliance component ○ In use: how much physical memory is in use ○ Free: how much physical memory is free ○ Cached: amount of memory that is allocated to disk cache ○ Buffered: the amount of physical memory that is used for file buffers ○ Swap total: the total amount of swap space available (in use plus free) ○ System services Swap in use: the total amount of swap space that is

	used
	○ Swap free: the total amount of swap space free
	○ Database Connections
	○ Configured
	○ Active
	○ Idle
	○ Network interfaces
	○ Up or down status for each interface
Storage	Storage information for a controller includes this information: ○ Volume ○ Total space ○ Used space ○ Percentage
Controller and compute node status	Indicator lights show component status ○ Green light: Running ○ Yellow light: The node is functional but is in the process of rebuilding; performance can be degraded during this time. Note: The rebuild progresses faster if the system is not being used. ○ Red light: not running Expand the node to obtain details of the appliance component by clicking the image.

- **Parent topic:** [System administration](#)

Related concepts

- [System administration](#)

Restarting the system

Restarting the system must be done periodically.

About this task

Note: The web application is temporarily unavailable if you restart it. Additionally, whenever volume definitions are edited or modified, you must restart the system.

Procedure

1. Select Administration > Dashboard > Appliance status. On the Appliance status page, you have two options:
   - Click the Controller link.
   - Click About Appliance.
2. The Restart services and Reboot icons appear at the bottom of the window. These icons are available on the View details page and on each of the tabs. Click either of these options:
   - Restart services: Restart all system services that are running on the node.
   - Reboot: to restart the components.

- **Parent topic:** System administration

Related concepts

- System administration

Configuration of IBM StoredIQ Platform

This section describes the Application and Network areas that an administrator can modify to configure IBM® StoredIQ Platform.

The Configuration subtab (Administration > Configuration) is divided into System and Application sections.

Table 1. System and Application configuration options

Section	Configuration options
System	- Configure the DA gateway. - View and modify network settings, including host name, IP address, NIS domain membership, and use. - View and modify settings to enable the generation of email notification messages. - Configure SNMP servers and communities. - Manage notifications for system and application events. - View and modify date and time settings for IBM StoredIQ Platform. - Set backup configurations. - Manage users. - Upload Lotus Notes user IDs so that encrypted NSF files can be imported into IBM StoredIQ Platform.
Application	- Specify directory patterns to exclude during harvests. - Specify options for full-text indexing. - View, add, and edit known data object types. - View and edit settings for policy audit expiration and removal. - Specify options for computing hash settings when harvesting. - Specify options to configure the desktop collection service.

- [Configuring DA Gateway settings](#)
  The DA Gateway settings are configured as part of general system-configuration options.
- [Configuring network settings](#)
- [Configuring mail settings](#)
  Mail settings can be configured as part of system configuration options.
- [Configuring SNMP settings](#)
  You can configure the system to make Object Identifier (OID) values available to Simple Network Management Protocol (SNMP) client applications. At the same time, you can receive status information or messages about system events in a designated trap. For information about environmental circumstances that are monitored by IBM StoredIQ Platform, see the following table.
- [Configuring notification from IBM StoredIQ Platform](#)
  You can configure the system to notify you using email or SNMP when certain events occur.
- [Setting the system time and date](#)
  The system's time and date can be modified as needed.
- [Setting system backup configurations](#)
  To prepare for disaster recovery, you can back up the configuration of a data server to a gateway. It backs up volume definitions, discovery export records, and data-server settings. It does not back up infosets, data maps, or indexes.
- [User management](#)
  User management is done by those users in the administrator role.
- [Importing encrypted NSF files from Lotus Notes](#)
  IBM StoredIQ Platform can decrypt, import, and process encrypted NSF files from IBM Lotus Domino v7. The feature works by comparing a list of user.id and key pairs that were imported into the system with the key values that lock each encrypted container or email. When the correct match is found, the file is unlocked with the key. After the emails

or containers are unlocked, IBM StoredIQ Platform analyzes and processes them in the usual fashion. This topic provides procedural information about how to import encrypted NSF files from IBM Lotus Domino.

- [Configuring harvester settings](#)
  You can use several different harvester settings to fine-tune your index process.
- [Configuring full-text index settings](#)
  Use the full-text index settings feature to customize your full-text index.
- [Specifying data object types](#)
  On the Data object types page, you can add new data object types and view and edit known data object types. These data objects appear in the Disk usage (by data object type) report. Currently, there are over 400 data object types available.
- [Configuring audit settings](#)
  Audit settings can be configured to determine the number of days and number of policy audits to be kept before they are deleted.
- [Configuring hash settings](#)
  Use the Hash Settings page to configure whether to compute a hash when harvesting. The page provides different types of hashes.
- [Configuring desktop settings](#)
  When you configure desktop settings, you are enabling or disabling encryption within IBM StoredIQ Platform. IBM Desktop Data Collector (desktop client or client) enables desktops as a volume type or data source, allowing them to be used just as other types of added data sources. The client is provided as a standard MSI file, and is installed according to the typical method (such as Microsoft Systems Management Service (SMS)) used within your organization. The client can collect PSTs and compressed files and other data objects and it is capable of removing itself when its work is completed.
- [Downloading the IBM Desktop Data Collector installer from the application](#)
  The IBM Desktop Data Collector installer can be downloaded from the application.
- [Upgrading the IBM Desktop Data Collector agent](#)
  The IBM Desktop Data Collector agent must be upgraded to open and harvest encrypted files.
- [Managing the Encrypted file system recovery agent](#)
  During IBM Desktop Data Collectorcollection, if IBM Desktop Data Collector finds an Encrypted File System-encrypted file, the IBM Desktop Data Collector installs a recovery agent certificate so that you can open the encrypted file.

Related tasks

- [Importing encrypted NSF files from Lotus Notes](#)

Configuring DA Gateway settings

The DA Gateway settings are configured as part of general system-configuration options.

Procedure

1. Go to Administration > Configuration > System > DA Gateway Settings.
2. In the Host text box, enter the gateway host. For example, enter `mgmt.example.com` or `192.168.10.10`.
3. Click OK. Services must be restarted.

- **Parent topic:** [Configuration of IBM StoredIQ Platform](#)

Related concepts

- [Configuration of IBM StoredIQ Platform](#)
- [User management](#)

Configuring network settings

Procedure

1. Go to Administration > Configuration > System > Network settings.
2. Click Controller Settings. Set or modify the following Primary Network Interface options.
   - IP type: Set to stic or dynamic. If it is set to dynamic, the IP address, Netmask, and Default Gateway fields are disabled.
   - IP address: Enter the IP address.
   - Netmask: Enter the network mask of the IP address.
   - Default gateway: Enter the IP address of the default gateway.
   - Hostname: Enter the fully qualified domain name that is assigned to the appliance.
   - Ethernet speed: Select the Ethernet speed.
   - Available ports: Indicate the available ports.
   - Separate network for file/email servers: Specify the additional subnet for accessing file/email servers. Select this check box if you are using the web application from one subnet and harvesting from another subnet.

   Restart the system for any primary network interface changes to take effect. See [Restarting and Rebooting the Appliance](#).
3. In Controller Settings, set or modify the following DNS Settings options.
   - DNS search order: Enter the DNS search order for multiple DNS servers.
   - Nameserver 1: Set the IP address of the primary DNS server for name resolution.
   - Nameserver 2: Set the IP address of the secondary DNS server for name resolution.
   - Nameserver 3: Set the IP address of the tertiary DNS server for name resolution.

   DNS settings take effect after they are saved. Changes to the server's IP address take effect immediately. Because the server has a new IP address, you must reflect this new address in the browser address line before next step.
4. Click OK.
5. Click Server name resolution. Set the following for the data server:
   a. Windows Share (CIFS) file server name resolution: These settings take effect upon saving.
      - LMHOSTS: Enter the IP host name format.
      - WINS Server: Enter the name of the WINS server.
   b. NIS (for NFS): These settings take effect upon saving.
      - Use NIS: Select this box to enable NIS to perform UID/GID to friendly name resolution in an NFS environment.
      - NIS Domain: Specify the NIS domain.
      - Broadcast for server on local network: Select this box if the NIS domain server is on the local network and can be discovered by broadcasting. This option does not work if the NIS domain server is on another subnet.
      - Specify NIS server: If not using broadcast, specify the IP address of the NIS domain server here.
   c. Active Directory: These settings take effect upon saving.
      - Active Directory server: Enter the name of the Active Directory server.
   d. Doc broker settings (for Documentum)
      - Enter Host name for doc broker.
      - Documentum global registry:
         - Registry name
         - User
         - Password
6. Click OK.

- **Parent topic:** [Configuration of IBM StoredIQ Platform](#)

Related concepts

- [Configuration of IBM StoredIQ Platform](#)
- [User management](#)

Configuring mail settings

Mail settings can be configured as part of system configuration options.

Procedure

1. Go to Administration > Configuration > System > Mail Server settings.
2. In Mail server, enter the name of the SMTP mail server.
3. In From address, enter a valid sender address. If the sender is invalid, some mail servers reject email. A sender address also simplifies the process of filtering email notifications that are based on the sender's email.
4. Click OK to save changes.

- **Parent topic:** Configuration of IBM StoredIQ Platform

Related concepts

- Configuration of IBM StoredIQ Platform
- User management

Configuring SNMP settings

You can configure the system to make Object Identifier (OID) values available to Simple Network Management Protocol (SNMP) client applications. At the same time, you can receive status information or messages about system events in a designated trap. For information about environmental circumstances that are monitored by IBM® StoredIQ Platform, see the following table.

Procedure

1. Go to Administration > Configuration > System > SNMP settings.
2. To make OID values available to SNMP client applications, in the Appliance Public MIB area:
   a. Select the Enabled check box to make the MIB available, that is, to open port 161 on the controller.
   b. In the Community field, enter the community string that the SNMP clients use to connect to the SNMP server.
   c. To view the MIB, click Download Appliance MIB. This document provides the MIB definition, which can be provided to an SNMP client application.
3. To capture messages that contain status information in the Trap destination area:
   a. In the Host field, enter the common name or IP address for the host.
   b. In the Port field, enter the port number. Port number 162 is the default.
   c. In the Community field, enter the SNMP community name.
4. To modify the frequency of notifications, complete these fields in the Environmental trap delivery area:
   a. Send environmental traps only every __ minutes.
   b. Send environmental traps again after __ minutes.
5. Click OK. Environmental traps that are monitored by IBM StoredIQ Platform are described in this table.

Option	Description
siqConsoleLogLineTrap	A straight conversion of a console log line into a trap. It uses these parameters: messageSource, messageID, severity, messageText.
siqRaidControllerTrap	Sent when the RAID controller status is anything but normal. Refer to the MIB for status code information. It uses this parameter: nodeNum.
siqRaidDiskTrap	Sent when any attached raid disk's status is anything but OK. It uses this parameter: nodeNum.
siqBbuTrap	Battery Backup Unit (BBU) error on the RAID controller detected. It uses this parameter: nodeNum.
siqCacheBitTrap	Caching indicator for RAID array is off. It uses this parameter: nodeNum.
siqNetworkTrap	Network interface is not UP when it must be. It uses this parameter: nodeNum.
siqDbConnTrap	Delivered when the active Postgres connection percentage exceeds an acceptable threshold. It uses this parameter: nodeNum.
siqFreeMemTrap	Delivered when available memory falls too low. It uses this parameter: nodeNum.
siqSwapUseTrap	Sent when swap use exceeds an acceptable threshold. Often indicates memory leakage. It uses this parameter: nodeNum.
siqCpuTrap	Sent when processor load averages are too high. It uses this parameter: nodeNum.
siqTzMismatchTrap	Sent when the time zone offset of a node does not match the time zone offset of the controller. It uses this parameter: nodeNum.

- **Parent topic:** Configuration of IBM StoredIQ Platform

Related concepts

- [Configuration of IBM StoredIQ Platform](#)
- [User management](#)

Configuring notification from IBM StoredIQ Platform

You can configure the system to notify you using email or SNMP when certain events occur.

About this task

For a list of events that can be configured, see [Event log messages](#).

Procedure

1. Go to Administration > Configuration > System > Manage notifications.
2. Click Create a notification.
3. In the Event number: field, search for events by clicking Browse or by typing the event number or sample message into the field.
4. Select the event level by clicking the ERROR, WARN, or INFO link.
5. Scroll through the list, and select each event by clicking it. The selected events appear in the create notification window. To delete an event, click the delete icon to the right of the event.
6. In the Destination: field, select the method of notification: SNMP, or Email address, or both. If you choose email address, enter one or more addresses in the Email address field. If you choose SNMP, the messages are sent to the trap host identified in the SNMP settings window, with a trap type of siqConsoleLogLineTrap.
7. Click OK.
8. To delete an item from the Manage notifications window, select the check box next to the event, and then click Delete. You can also request a notification for a specific event from the dashboard's event log. Click the Subscribe link next to any error message and a prepopulated edit notification screen that contains the event is provided.

- **Parent topic:** [Configuration of IBM StoredIQ Platform](#)

Related concepts

- [Configuration of IBM StoredIQ Platform](#)
- [User management](#)

Setting the system time and date

The system's time and date can be modified as needed.

About this task

Note: A system restart is required for any changes that are made to the system time and date. See [Restarting and Rebooting the Appliance](#).

Procedure

1. Go to Administration > Configuration > System > System time and date.
2. Enter the current date and time.
3. Select the appropriate time zone for your location.
4. Enable Use NTP to set system time to use an NTP server to automatically set the system date and time for the data server. If NTP is used to set the system time, then the time and date fields set automatically. However, you must specify the time zone.
5. Enter the name or IP address of the NTP server.
6. Click OK to save changes.

- **Parent topic:** [Configuration of IBM StoredIQ Platform](#)

Related concepts

- [Configuration of IBM StoredIQ Platform](#)
- [User management](#)

Setting system backup configurations

To prepare for disaster recovery, you can back up the configuration of a data server to a gateway. It backs up volume definitions, discovery export records, and data-server settings. It does not back up infosets, data maps, or indexes.

Before you begin

The gateway must be manually configured to support this backup. For additional information, configure the gateway in the util menu. It can be done by contacting your service representative.

Procedure

1. Go to Administration > Configuration > System > Backup configuration.
2. Press Start backup.

- **Parent topic:** [Configuration of IBM StoredIQ Platform](#)

Related concepts

- [Configuration of IBM StoredIQ Platform](#)
- [User management](#)

User management

User management is done by those users in the administrator role.

User management entails logging in and out of the system and various account-administration tasks.

- [Logging in and out of the system](#)
  The system comes with a default administrative account named admin, the default password for which is admin. For security purposes, change the password as soon as possible.
- [Changing the administrative account](#)
  The administrator account and its password can be changed by an administrator.
- [Creating users](#)
  Users can be added only by administrators.
- [Editing users](#)
  User accounts can be edited by administrators.
- [Deleting a user's account](#)
  User accounts can be deleted by administrators.
- [Locking a user's account](#)
  User accounts can be locked by an administrator.
- [Managing a user's account](#)
  As part of basic user account administration, user accounts can be unlocked.

- **Parent topic:** [Configuration of IBM StoredIQ Platform](#)

Related concepts

- [Configuration of IBM StoredIQ Platform](#)

Related tasks

- [Configuring DA Gateway settings](#)
- [Configuring network settings](#)
- [Configuring mail settings](#)
- [Configuring SNMP settings](#)
- [Configuring notification from IBM StoredIQ Platform](#)
- [Setting the system time and date](#)
- [Setting system backup configurations](#)
- [Logging in and out of the system](#)
- [Changing the administrative account](#)
- [Creating users](#)
- [Editing users](#)
- [Deleting a user's account](#)
- [Locking a user's account](#)
- [Managing a user's account](#)

Logging in and out of the system

The system comes with a default administrative account named admin, the default password for which is admin. For security purposes, change the password as soon as possible.

Procedure

1. Open a browser and in the address line, enter the URL for your system.
2. On the Login page, enter `admin` for the email address and `admin` for the password the first time you log in.
3. Click Log In to enter the system.
   Note: Database Compactor: If someone tries to log in while the appliance is doing database maintenance, the administrator can override the maintenance procedure and use the system. For more information, see [Job configuration](#).
4. To log out of the application, click User Account in the upper right corner, and then click the Log out link.

- **Parent topic:** [User management](#)

Related concepts

- [User management](#)

Changing the administrative account

The administrator account and its password can be changed by an administrator.

Procedure

1. On the Administration > Configuration tab, click Manage users.
2. From the list, select The Administrator account, and then select Change the "admin" password.
3. Enter a new password, and then click OK to save the change.

- **Parent topic:** [User management](#)

Related concepts

- [User management](#)

Creating users

Users can be added only by administrators.

Procedure

1. Go to Administration > Configuration > Manage users, and then click Create new user.
2. In the First name field, enter the user's given name.
3. In the Last name field, enter the user's surname.
4. In the Email address field, enter the user's email address.
5. For Authentication, select Active Directory or Local.
6. In the Active Directory principal, enter the name of the Active Directory principal.
7. To receive Notification of reports by email, select either Yes or No.
8. Select Administrator and select the Role.
9. Select the View data objects in the viewer check box if you want to see data objects in the viewer.
10. Click OK to create the user.

- **Parent topic:** [User management](#)

Related concepts

- [User management](#)

Editing users

User accounts can be edited by administrators.

Procedure

1. Go to Administration > Configuration > Manage users, and then click the name of the user that you want to edit.
2. Click Edit User and edit the fields as needed.
3. Click OK to save your changes.

- **Parent topic:** [User management](#)

Related concepts

- [User management](#)

Deleting a user's account

User accounts can be deleted by administrators.

Procedure

1. Go to Administration > Configuration > Manage users, and then click the name of the user that you want to delete.
2. Click Delete, and then click OK.

- **Parent topic:** User management

Related concepts

- User management

Locking a user's account

User accounts can be locked by an administrator.

Procedure

1. On the Administration > Configuration page, click Manage users.
2. In the list, select the user name of the account you want to lock.
3. Click Lock user account to disable the account. A padlock icon appears next to the account name in the list, indicating that the account is now locked. The user cannot log in to the account while it is locked.

- **Parent topic:** [User management](#)

Related concepts

- [User management](#)

Managing a user's account

As part of basic user account administration, user accounts can be unlocked.

Procedure

1. On the Administration > Configuration page, click Manage users.
2. In the list, select the user name of the account you want to unlock.
3. Click Unlock account to re-enable the account.

- **Parent topic:** [User management](#)

Related concepts

- [User management](#)

Importing encrypted NSF files from Lotus Notes

IBM® StoredIQ Platform can decrypt, import, and process encrypted NSF files from IBM Lotus Domino v7. The feature works by comparing a list of user.id and key pairs that were imported into the system with the key values that lock each encrypted container or email. When the correct match is found, the file is unlocked with the key. After the emails or containers are unlocked, IBM StoredIQ Platform analyzes and processes them in the usual fashion. This topic provides procedural information about how to import encrypted NSF files from IBM Lotus Domino.

About this task

These use cases are supported:

- Multiple unencrypted emails within a journaling database that was encrypted with a single user.id key
- Multiple unencrypted emails in an encrypted NSF file
- Multiple encrypted emails within an unencrypted NSF file
- Multiple encrypted emails with the same or different user.id keys, contained in an encrypted NSF file
- Encrypted emails from within a journaling database

Procedure

1. On the primary data server, go to Administration > Configuration > Lotus Notes user administration.
2. Click Upload a Lotus user ID file.
   a. In the dialog that appears, click Browse, and go to a user file.
   b. In the Lotus Notes password: field, type the password that unlocks the selected file.
   c. Optional: In the Description: field, enter a description for the file.
   d. Click OK. Repeat until the keys for all encrypted items are uploaded. When the list is compiled, you can add new entries to it then.
   e. To delete an item from the list, from the Registered Lotus users screen, select the check box next to a user, and then click Delete. In the confirmation dialog that appears, click OK.
      Note: After you upload user IDs, restart services.

- **Parent topic:** [Configuration of IBM StoredIQ Platform](#)

Related concepts

- [Configuration of IBM StoredIQ Platform](#)

Configuring harvester settings

You can use several different harvester settings to fine-tune your index process.

Procedure

1. Go to Administration > Configuration > Application > Harvester settings.
2. To configure Basic settings:
    a. Harvester Processes: Select either Content processing or System metadata only.
    b. Harvest miscellaneous email items: Select to harvest contacts, calendar appointments, notes, and tasks from the Exchange server.
    c. Harvest non-standard Exchange message classes: Select to harvest message classes that do not represent standard Exchange email and miscellaneous items.
    d. Include extended characters in object names: Select to allow extended characters to be included in data object names during a harvest.
    e. Determine whether data objects have NSRL digital signature: Select to check data objects for NSRL digital signatures.
    f. Enable parallel grazing: Select to harvest volumes that were already harvested and are going to be reharvested. If the harvest completes normally, parallelized grazing enables harvests to begin where they left off when interrupted and to start at the beginning.
    g. Enable OCR image processing: Select this option to control at a global level whether Optical Character Recognition (OCR) processing is attempted on image files. If you select this option, you must restart services. See [Restarting and Rebooting the Appliance](#).
3. Specify Skip Content processing.
    ◦ Data object extensions to be skipped: Specify those file types that you want the harvest to ignore by adding data object extensions to be skipped.
4. To configure Locations to ignore, enter each directory that must be skipped. IBM® StoredIQ Platform accepts only one entry per line and that regular expressions can be used.
5. To configure Limits:
    a. Maximum data object size: Specify the maximum data object size to be processed during a harvest. During a harvest, iles exceeding the maximum data object size is not read. As a result, if full-text/content processing is enabled for the volume, they are audited as skipped: Configured max. object size. These objects still appear in the volume cluster along with all file system metadata. Since they were not read, the hash is a hash of the file-path and size of the object, regardless of what the hash settings are for the volume (full/partial/off).
    b. Max entity values per entity: For any entity type (date, city, address and the like), the system records, per data object, the number of values set in this field. The values do not need to be unique. For example, if the maximum value is 1,000, and the harvester collects 1,000 instances of the same date (8/15/2009) in a Word document, the system stops counting dates. This setting applies to all user-defined expressions (keyword, regular expression, scoped, and proximity) and all standard attributes.
    c. Max entity values per data object: Across all entity types, the total (cumulative) number of values that is collected from a data object during a harvest. A 0 in this field means "unlimited". This setting applies to all user-defined expressions (key-word, regular expression, scoped, and proximity) and all standard attributes.
6. Configure Binary Processing.
    a. Run binary processing when text processing fails: Select this option to run binary processing. The system runs further processes against content that failed in the harvesting. You can select options for when to start this extended processing and how to scan content. Binary processing does not search image file types such as .GIF and .JPG for text extraction.
    b. Failure reasons to begin binary processing: Select the check boxes of the options that define when to start extended processing. Binary processing can enact in extracting text from a file failure in these situations:

- When the format of the file is unknown to the system parameters
- When the data object type is not supported by the harvester scan
- When the data object format does not contain actual text

    c. Data object extensions: Set binary processing to process all data files or only files of entered extensions. To add extensions, enter one per line without a period.

    d. Text encoding: Set options for what data to scan and extract at the start of binary processing. This extended processing can accept extended characters and UTF-16 and UTF-32 encoded characters as text. The system searches UTF-16 and UTF-32 by default.

    e. Minimums: Set the minimum required number of located, consecutive characters to begin processing for text extraction. For example, if you enter 4, the system begins text processing when four consecutive characters of a particular select text encoding are found. This setting helps find and extract helpful data from the binary processing, reducing the number of false positives.

7. Click OK. Changes to harvester settings do not take effect until the appliance is rebooted or the application services are restarted.

- [Optical character recognition processing](#)
  Optical character recognition (OCR) processing enables text extraction from graphic image files that are stored inside archives where the Include content tagging and full-text index option is selected.

- **Parent topic:** [Configuration of IBM StoredIQ Platform](#)

Related concepts

- [Configuration of IBM StoredIQ Platform](#)
- [Optical character recognition processing](#)

Optical character recognition processing

Optical character recognition (OCR) processing enables text extraction from graphic image files that are stored inside archives where the Include content tagging and full-text index option is selected.

After content typing inside the IBM® StoredIQ Platform processing pipeline, enabling OCR processing routes the following file types through an optical character recognition engine OCR to extract recognizable text.

- Windows or OS/2 bitmap (BMP)
- Tag image bitmap file (TIFF) (TIFF)
- Bitmap (CompuServe) (GIF)
- Portable Network Graphics (PNG)

The text that is extracted from image files is processed through the IBM StoredIQ Platform pipeline in the same manner as text extracted from other supported file types. Policies with a specific feature to write out extracted text to a separate file for supported file types do so for image files while OCR processing is enabled.

The OCR processing rate of image files is approximately 7-10 KB/sec per IBM StoredIQ Platform harvester process.

- **Parent topic:** [Configuring harvester settings](#)

Related tasks

- [Configuring harvester settings](#)

Configuring full-text index settings

Use the full-text index settings feature to customize your full-text index.

About this task

Before you configure or search the full-text index, consider the following situations:

- Full-text filters that contain words might not return all instances of those words: You can limit full-text indexing for words that are based on their length. For example, if you choose to full-text index words limited to 50 characters, then no words greater than 50 characters are indexed.
- Full-text filters that contain numbers might not return all instances of those numbers: This situation can occur when number searches are configured as follows:
    - The length of numbers to full-text index was defined. If you configure the full-text filter to index numbers with 3 digits or more and try to index the numbers `9`, `99`, `999`, and the word `stock`, only the number `999` and the word `stock` are indexed. The numbers `9` and `99` are not indexed.
    - Number indexing in data objects that are limited by file extensions. For example, if you choose to full-text index the number `999` when it appears in data objects with the file extensions .XLS and .DOC, then a full-text filter returns only those instances of the number `999` that exist in data objects with the file extensions .XLS and .DOC. Although the number 999 can exist in other data objects that are harvested, these data objects do not have the file extensions .XLS or .DOC.

Procedure

1. Go to Administration > Configuration > Application > Full-text settings.
2. To configure Limits:
    a. Do not limit the length of the words that are indexed: Select this option to have no limits on the length of words that are indexed.
    b. Limit the length of words indexed to___characters: Select this option to limit the length of words that are indexed. Enter the maximum number of characters at which to index words. Words with more characters than the specified amount are not indexed.
3. To configure Numbers:
    - Do not include numbers in the full-text index: Select this option to have no indexed numbers. This option is selected by default.
    - Include numbers in the full-text index: Select this option to have numbers to be indexed.
    - Include numbers in full-text index but limit them by: Select this option to have only certain numbers indexed. Define these limits as follows:
        - Number length: Include only numbers that are longer than ____ characters. Enter the number of characters a number must contain to be indexed. The Number length feature indexes longer numbers and ignores shorter numbers. By not indexing shorter numbers, such as one- and two-character numbers, you can focus your filter on meaningful numbers. These numbers can be account numbers, Social Security numbers, credit card numbers, license plate numbers, or telephone numbers.
        - Extensions: Index numbers that are based on the file extensions of the data objects in which they appear. Select Limit numbers for all extensions to limit numbers in all file extensions to the character limits set in Number length. Alternatively, select Limit numbers for these extensions to limit the numbers that are selected in Numbers length only to data objects with certain file extensions. Enter the file extensions one per line that must have limited number indexing. Any data object with a file extension that is not listed has all indexed numbers.
4. Configure Stop words. Stop words are common words that are found in data objects.

These data objects are not indexed in the full-text index. To index, stop words like `the` or `but` would compromise full-text indexing speed and the relevancy of the results. By default, the following words are omitted from the full-text index: `a, an, and, are, as, at, be, but, by, for, if, in, into, is, it, no, not, of, on, or, such, that, the, their, then, there, these, they, this, to, was, will, with`.

- If you want the system to find a word that is listed, delete the word from the list to allow the full-text index to find it. You must reindex to find the word.
- To add a stop word, enter one word per line, without punctuation, which includes hyphens and apostrophes.

5. Configure Punctuation or special characters to index by entering punctuation characters in the index. Punctuation characters must be entered without spaces to separate them. By default, only some punctuation is indexed as a letter in a word by default. Most punctuation is turned into a space. To make certain that a specific punctuation character is indexed, add it to this list.

6. To configure Include word stems in index, select whether to stem words that were indexed. For example, `employ` is the stem word of words such as `employed`, `employment`, `employs`. If you use stemming and search for the term `employed`, IBM® StoredIQ Platform denotes any found instances of `employment`, `employ`, `employee`, and so on, when it views the data object.

- Do not stem words that are indexed (faster indexing): By not stemming indexed words, data sources are indexed faster.
- Stem words that are indexed (improved searching): By stemming indexed words, filters are more precise, although slower. Without stemming, a filter for `trade` would need to be written as `trade or trades or trading or traded` to get the same effect, and even then a user can miss an interesting variant.

Note: If stemming feature is enabled, the use of double quotation marks returns stemmed terms in results. To find exact words with no stemmed terms, use single quotation marks.

7. To Optimize wildcard suffix searches, select whether to optimize searches by suffix such as `*ology`, `*tion`, `*ious`, `*ive`, or `*less`.

- Do not optimize searches by suffix (faster indexing): By not optimizing searches by suffix, data sources are indexed faster.
- Optimize searches by suffix (faster searching): By optimizing searches by suffix, searches are conducted more quickly; however, non-suffix searches or standard wildcard searches (such as `bird*`) are not affected.

Note: Stemming is not conducted for search terms with wildcards (`"?"`, `"*"`) in them. It is true regardless of the term that is placed within single quotation marks.

8. Click OK.

- **Parent topic:** [Configuration of IBM StoredIQ Platform](#)

Related concepts

- [Configuration of IBM StoredIQ Platform](#)

Specifying data object types

On the Data object types page, you can add new data object types and view and edit known data object types. These data objects appear in the Disk usage (by data object type) report. Currently, there are over 400 data object types available.

Procedure

1. Go to Administration > Configuration > Application > Data object types.
2. In the add data object type section, enter one or more extensions to associate with the data object type. These entries must be separates by spaces. For example, enter `doc txt xls`.
3. Enter the name of the data object type to be used with the extension or extensions. For example, enter `Microsoft Word`.
4. Click Add to add the extension to the list.

- **Parent topic:** [Configuration of IBM StoredIQ Platform](#)

Related concepts

- [Configuration of IBM StoredIQ Platform](#)

Configuring audit settings

Audit settings can be configured to determine the number of days and number of policy audits to be kept before they are deleted.

Procedure

1. Go to Administration > Configuration > Application > Audit settings.
2. Specify the number of days to keep the policy audits before automatically deleting them.
3. Specify the maximum number of policy audits to keep before automatically deleting them.
4. Specify the file limit for drill-down in policy audits.
5. Click OK to save changes.

- **Parent topic:** [Configuration of IBM StoredIQ Platform](#)

Related concepts

- [Configuration of IBM StoredIQ Platform](#)

Configuring hash settings

Use the Hash Settings page to configure whether to compute a hash when harvesting. The page provides different types of hashes.

Procedure

1. Go to Administration > Configuration > Application > Hash settings.
2. Select Compute data object hash.
   Note: When you hash emails, click Choose email fields and select email attributes to be used. The email hash selections operate independently from the data object hash settings; that is, a data object can have a binary hash or an email hash, but not both.
3. When you hash data objects, except for emails, specify the following options:
   - Entire data object content (required for data object typing)
   - Partial data object content
4. Click OK.

- **Parent topic:** [Configuration of IBM StoredIQ Platform](#)

Related concepts

- [Configuration of IBM StoredIQ Platform](#)

Configuring desktop settings

When you configure desktop settings, you are enabling or disabling encryption within IBM® StoredIQ Platform. IBM Desktop Data Collector (desktop client or client) enables desktops as a volume type or data source, allowing them to be used just as other types of added data sources. The client is provided as a standard MSI file, and is installed according to the typical method (such as Microsoft Systems Management Service (SMS)) used within your organization. The client can collect PSTs and compressed files and other data objects and it is capable of removing itself when its work is completed.

Procedure

1. Go to Administration > Configuration > Application > Desktop settings.
2. In the Desktop Services area, select the Encrypt all traffic to/from desktops check box.
3. Select either Enabled or Disabled to enable or disable desktop services.
4. Click Apply.

- **Parent topic:** [Configuration of IBM StoredIQ Platform](#)

Related concepts

- [Configuration of IBM StoredIQ Platform](#)

Downloading the IBM Desktop Data Collector installer from the application

The IBM® Desktop Data Collector installer can be downloaded from the application.

Procedure

1. Go to Administration > Configuration > Application > Desktop settings.
2. In the Download the Desktop Agent installer area, click Download the desktop client installer.
3. When the download is complete, click Save File.

Results

After the desktop client is installed on a desktop and connected to and registered with the data server, that desktop is available as a data source within the list of primary volumes. Connectivity and the correct IP address are required.

- **Parent topic:** [Configuration of IBM StoredIQ Platform](#)

Related concepts

- [Configuration of IBM StoredIQ Platform](#)

Upgrading the IBM Desktop Data Collector agent

The IBM® Desktop Data Collector agent must be upgraded to open and harvest encrypted files.

About this task

During IBM Desktop Data Collector collection, if IBM Desktop Data Collector finds an Encrypted File System-encrypted file, then IBM Desktop Data Collector installs a recovery agent certificate, allowing the client to open the encrypted file and harvest from it.

Procedure

1. Go to Administration > Configuration > Application > Desktop settings.
2. In the Upgrades area, select either Automatic upgrade or Available versions.
3. For Automatic upgrade options:
    - Upgrades disabled: All upgrades are disabled, that is, none is applied.
    - Upgrade all workstations: All workstations are upgraded.
4. For Available versions options:
    - Select Manually publish new version, and then select that version.
    - Select Automatically publish the latest version.
5. Click Apply.

- **Parent topic:** [Configuration of IBM StoredIQ Platform](#)

Related concepts

- [Configuration of IBM StoredIQ Platform](#)

Managing the Encrypted file system recovery agent

During IBM® Desktop Data Collectorcollection, if IBM Desktop Data Collector finds an Encrypted File System-encrypted file, the IBM Desktop Data Collector installs a recovery agent certificate so that you can open the encrypted file.

Procedure

1. Go to Administration > Configuration > Application > Desktop settings.
2. In the Encrypted file system recovery agent users area, click Add encrypted file system user. The Upload Recovery Agent Certificate dialog box appears.
3. In the Select a .PFX file to upload: text box, click Browse to go to the wanted .PFX file. By default, the system takes a .PFX file.
4. Enter the .PFX password in the .PFX password: text box. This password protects the file itself.
5. In the Username: text box, enter the user name for the user, a SAM compatible/NT4 Domain name-style user name. For example, enter `MYCOMPANY\esideways`. This user name is the credential of the user to whom this recovery agent belongs.
6. In the Password: text box, enter the password for the user.
7. Optional: Enter a description in the Description: text box.
8. Click OK. The file is uploaded, and the added user is visible within the User name column.
   Note:
   After users are added, they can also be edited or deleted with the Edit or Delete options.

- **Parent topic:** [Configuration of IBM StoredIQ Platform](#)

Related concepts

- [Configuration of IBM StoredIQ Platform](#)

Volumes and data sources

Volumes or data sources are integral to IBM® StoredIQ Platform indexing your data.

A volume represents a data source or destination that is available on the network to the IBM StoredIQ Platform appliance. A volume can be a disk partition or group of partitions that is available to network users as a single designated drive or mount point. IBM StoredIQ Platform volumes have the same function as partitions on a hard disk drive. When you format the hard disk drive on your PC into drive partitions A, B, and C, you are creating three partitions that function like three separate physical drives. Volumes behave the same way that hard disk drive disk partitions behave. You can set up three separate volumes that originate from the same server or across many servers. Only administrators can define, configure, and add or remove volumes to IBM StoredIQ Platform.

- Volume indexing
  When you define volumes, you can determine the type and depth of index that is conducted.
- Server platform configuration
  Before you configure volumes on IBM StoredIQ Platform, you must configure the server platforms that you use for the different volume types. Each server type has requisite permissions and settings.
- Configuration of retention servers
  IBM StoredIQ Platform supports various types of retention servers, which must be configured before retention volumes are added due to unique requirements.
- Creating primary volumes
  Primary volumes can be created as data sources.
- Creating retention volumes
  Retention volumes store data objects that are placed under retention. It means that the object is retained.
- Creating discovery export volumes
  Discovery export volumes contain the data produced from a policy, which is kept so that it can be exported as a load file and uploaded into a legal review tool. Administrators can also configure discovery export volumes for managing harvest results from cycles of a discovery export policy.
- Creating system volumes
  System volumes support volume export and import. When you export a volume, data is stored on the system volume. When you import a volume, data is imported from the system volume.
- Export and import of volume data
  The system's export and import volume capabilities allow metadata and full-text indexed data to be collected or exported from separate locations, such as data servers located in various offices of the enterprise. Once the data is available, it can be imported to a single location, such as a headquarter's office data server, where selected files might be retained.
- Deleting volumes
  Administrators can delete volumes from the list of available data sources, provided that the data server is connected to the gateway.
- Policy limitations for volume types
  Volume types have different policy limitations.

Related tasks

- Creating retention volumes

Volume indexing

When you define volumes, you can determine the type and depth of index that is conducted.

There are three levels of analysis:

- System metadata index: This level of analysis runs with each data collection cycle and provides only system metadata for system data objects in its results. It is useful as a simple inventory of what data objects are present in the volumes you defined and for monitoring resource constraints (such as file size) or prohibited file types (such as .MP3s).
- System metadata plus containers: In a simple system metadata index, container data objects (compressed files, PSTs, emails with attachments, and the like) are not included. This level of analysis provides container-level metadata in addition to the system metadata for system data objects.
- Full-text and content tagging: This option provides the full local language analysis that yields the more sophisticated entity tags. Naturally, completing a full-text index requires more system resources than a metadata index. Users must carefully design their volume structure and harvests so that the maximum benefit of IBM® StoredIQ Platform's sophisticated analytics are used, but not on resources that do not require them. Parameters and limitations on "full-text" indexing are set when the system is configured.

- **Parent topic:** [Volumes and data sources](#)

Related concepts

- [Volumes and data sources](#)
- [Configuration of retention servers](#)
- [Server platform configuration](#)

Server platform configuration

Before you configure volumes on IBM® StoredIQ Platform, you must configure the server platforms that you use for the different volume types. Each server type has requisite permissions and settings.

Because IBM StoredIQ Platform supports several different types of volumes, server platforms must be configured to support those volume types.

- Configuring Windows Share (CIFS)
  Windows Share (CIFS) must be configured to harvest and run policies.
- Configuring NFS
  NFS must be configured to harvest and run policies.
- Configuration of Exchange servers
  When you configure Exchange servers, you must consider various connections and permissions.
- Configuration of SharePoint
  When you configure SharePoint, certain privileges are required by user account along with IBM StoredIQ Platform recommendations. Additionally, SharePoint 2007 and 2010 require the configuration of alternate-access mappings to map IBM StoredIQ Platform requests to the correct web sites.
- Configuring Documentum
  Documentum has configuration requirements when it is used as a server platform.
- Configuring Enterprise Vault sites
  Enterprise Vault sites must be configured before the addition of an Enterprise Vault volume.
- Configuration of Discovery Accelerator
  Before you configure Discovery Accelerator primary volumes, you must configure Discovery Accelerator customer information and Enterprise Vault sites sequentially so that certain configuration items can appear in the volume configuration lists.
- Configuring NewsGator
  When NewsGator is used as a server platform, several privileges must be configured.

- **Parent topic:** Volumes and data sources

Related concepts

- Volumes and data sources
- Configuration of retention servers
- Volume indexing
- Configuration of Exchange servers
- Configuration of SharePoint
- Configuration of Discovery Accelerator

Related tasks

- Configuring Windows Share (CIFS)
- Configuring NFS
- Configuring Documentum
- Configuring Enterprise Vault sites
- Configuring NewsGator

Configuring Windows Share (CIFS)

Windows Share (CIFS) must be configured to harvest and run policies.

Procedure

To harvest and run policies on volumes on Windows Share (CIFS) servers, the user must be in the backup operator group on the Windows Share server that shows the shares on IBM® StoredIQ Platform.

- **Parent topic:** Server platform configuration

Related concepts

- Server platform configuration
- Configuration of Exchange servers
- Configuration of SharePoint

Configuring NFS

NFS must be configured to harvest and run policies.

Procedure

To harvest and run policies on NFS servers, you must enable root access on the NFS server that is connected to IBM® StoredIQ Platform.

- **Parent topic:** Server platform configuration

Related concepts

- Server platform configuration
- Configuration of Exchange servers
- Configuration of SharePoint

Configuration of Exchange servers

When you configure Exchange servers, you must consider various connections and permissions.

- Secure connection: If you want to connect to Exchange volumes over HTTPS, you must add port number 443 after the server name, for example, `qa03exch2000.qaw2k.local:443`. In some cases, this secure connection can result in some performance degradation due to SSL running large. If you enter the volume information without the 443 suffix, the default connection will be over HTTP.
- Permissions for Exchange 2003: The following permissions must be set on the Exchange server to the mailbox store or the mailboxes from which you harvest.
  - Read
  - Execute
  - Read permissions
  - List contents
  - Read properties
  - List object
  - Receive as
- Permissions for Exchange 2007 and 2010: The Full Access permissions must be set on the Exchange server to the mailbox store or the mailboxes from which you will harvest.
- Recoverable items in Exchange 2010: To harvest the recoverable items folders in Exchange 2010, you must be logged in as an administrator role.
- Deleted items: To harvest items that were deleted from the Exchange server, enable Exchange's transport dumpster settings. For more information, see *Microsoft® Exchange Server 2010 Administrator's Pocket Consultant*. Configuration information is also available online at www.microsoft.com.
- Windows Authentication: For all supported versions, enable Integrated Windows Authentication on each Exchange server.

- [Enabling integrated Windows authentication on Exchange servers](#)
  Windows authentication can be integrated on Exchange servers.
- [Improving performance for IIS 6.0 and Exchange 2003](#)
  Within IBM StoredIQ Platform, performance can be improved for IIS 6.0 and Exchange 2003.

- **Parent topic:** [Server platform configuration](#)

Related concepts

- [Server platform configuration](#)

Related tasks

- [Configuring Windows Share (CIFS)](#)
- [Configuring NFS](#)
- [Enabling integrated Windows authentication on Exchange servers](#)
- [Improving performance for IIS 6.0 and Exchange 2003](#)

Enabling integrated Windows authentication on Exchange servers

Windows authentication can be integrated on Exchange servers.

Procedure

1. From Microsoft Windows, log in to the Exchange Server.
2. Go to Administrative Tools > Internet Information Services (IIS) Manager.
3. Go to Internet Information Services > Name of Exchange Server > Web Sites > Default Web Site.
4. Right-click Default Web Site, and then click the Directory Security tab.
5. In the Authentication and access control pane, click Edit.
6. Select Properties. The Authentication Methods window appears.
7. In the Authentication access pane, select the Integrated Windows authentication check box.
8. Click OK.
9. Restart IIS services.

- **Parent topic:** [Configuration of Exchange servers](#)

Related concepts

- [Configuration of Exchange servers](#)

Improving performance for IIS 6.0 and Exchange 2003

Within IBM® StoredIQ Platform, performance can be improved for IIS 6.0 and Exchange 2003.

Procedure

1. From Microsoft Windows, log on to the Exchange Server.
2. Go to Administrative Tools > Internet Information Services (IIS) Manager.
3. Select Internet Information Services > <Name of Exchange Server> > Web Sites > Application Pools.
4. Right-click Application Pools and select Properties.
5. On the Performance tab, locate the Web Garden section.
6. If the number of worker processes is different from the default value of 1, then change the number of worker processes to 1.
7. Click OK.
8. Restart IIS Services.

- **Parent topic:** [Configuration of Exchange servers](#)

Related concepts

- [Configuration of Exchange servers](#)

Configuration of SharePoint

When you configure SharePoint, certain privileges are required by user account along with IBM® StoredIQ Platform recommendations. Additionally, SharePoint 2007 and 2010 require the configuration of alternate-access mappings to map IBM StoredIQ Platform requests to the correct web sites.

To configure SharePoint, consider these connections and privileges:

Secure Connection
> If you want to connect to SharePoint volumes over HTTPS, you can use the Use SSL check box, or can add the port number 443 after the server name when you set up the volume on IBM StoredIQ Platform, for example, `qa01.company.com:443`. In some cases, this secure connection can result in some performance degradation due to Secure Socket Layer (SSL) running large. If you enter the volume information without the 443 suffix, the default connection is over HTTP.

Privileges
> To run policies on SharePoint servers, you must use credentials with Full Control privileges. Use a site collection administrator to harvest subsites of a site collection.

Privileges required by user account

IBM StoredIQ Platform is typically used with SharePoint for one of these instances: to harvest and treat SharePoint as a source for policy actions or to use as a destination for policy actions, which means that you can write content into SharePoint with . Consider these points:

- Attributes are not set or reset on a SharePoint harvest or if you copy from SharePoint.
- Attributes are set only if you copy to SharePoint.

You must denote the following situations:

- If you plan to read only from the SharePoint (harvest and source copies from), then you must use user credentials with Read privileges on the site and all of the lists and data objects that you expect to process.
- If you plan to use SharePoint as a destination for policies, you must use user credentials with Contribute privileges on the site.
- More Privileges for Social Data: If you want to index all the social data for a user profile in SharePoint 2010, then the user credentials must own privileges to Manage Social Data as well.
- Recommended Privileges: Use a site collection administrator to ensure that all data is harvested from a site or site collection.

Alternate-access mappings

Alternate-access mappings map URLs presented by IBM StoredIQ Platform to internal URLs received by Windows SharePoint Services. An alternate-access mapping is required between the server name and optional port that is defined in the SharePoint volume definition and the internal URL of the web application. If SSL is used to access the site, ensure that the alternate-access mapping URL uses `https://` as the protocol.

Refer to Microsoft SharePoint 2007 or 2010 documentation to configure alternate-access mappings. These mappings are based on the public URL that is configured by the local SharePoint administrator and used by the IBM StoredIQ Platform SharePoint volume definitions.

For example, you are accessing a SharePoint volume with the fully qualified domain name, `http://itweb.storediqexample.com`, from the intranet zone. An alternate-access mapping for the public URL `http://itweb.storediqexample.com` for the intranet zone must be configured for the SharePoint 2007 or 2010 web application that hosts the site to be accessed by the volume

definition. If you are accessing the same volume with SSL, the mapping added must be for the URL `https://itweb.storediqexample.com` instead.

Note: When you configure SharePoint volumes with non-qualified names, you are entering the URL for a SharePoint site collection or site that is used by IBM StoredIQ Platform in the volume definition. Consider the following conditions:

- The URL must be valid about the Alternate Access Mappings that are configured in SharePoint.
- If the host name in the URL does not convey the fully qualified domain to authenticate the configured user, an Active Directory server must be specified. The specified Active Directory must be a fully qualified domain name and is used for authentication.

- **Parent topic:** Server platform configuration

Related concepts

- Server platform configuration

Related tasks

- Configuring Windows Share (CIFS)
- Configuring NFS

Configuring Documentum

Documentum has configuration requirements when it is used as a server platform.

Procedure

To run harvests and copy from Documentum servers, you must use the Contributor role.

- **Parent topic:** [Server platform configuration](#)

Related concepts

- [Server platform configuration](#)

Configuring Enterprise Vault sites

Enterprise Vault sites must be configured before the addition of an Enterprise Vault volume.

Procedure

1. Go to Administration > Data sources > Specify servers.
2. Click Enterprise Vault sites, and then click Add new Enterprise Vault site.
3. In the Site name field, enter a unique logical site name. This name appears in the screens that are used to configure Enterprise Vault volumes.
4. In the Enterprise Vault site alias field, enter the FQDN of the Enterprise Vault Server. Each server can be added only one time.
5. In the User name field, enter the login name. If the user is a domain user, then enter the login name as domain\user. The Enterprise Vault Service Account or a user with equivalent privileges must be used.
6. In the Password field, enter the user's password to authenticate with Active Directory.
7. Click OK to save the site.

- **Parent topic:** [Server platform configuration](#)

Related concepts

- [Server platform configuration](#)

Configuration of Discovery Accelerator

Before you configure Discovery Accelerator primary volumes, you must configure Discovery Accelerator customer information and Enterprise Vault sites sequentially so that certain configuration items can appear in the volume configuration lists.

You must log in to the Discovery Accelerator server and run the `ImportExport.exe` tool in the installation folder to obtain the appropriate Customer IDs and customer database names.

- Discovery Accelerator Web-Services Interface: In order for IBM® StoredIQ Platform to interface with Enterprise Vault with the Discovery Accelerator web services, the following configurations must be made on the Discovery Accelerator service.
    1. Log in to the Discovery Accelerator client interface as a Vault User or Discovery Administrator.
    2. Click the Configuration tab and expand the API options on the Settings page.
    3. In the API settings group, ensure that the API Enabled setting is enabled.
    4. Configure a Temporary Storage Area as needed. Ensure that it has sufficient free space and that any authenticated users that define volumes against Discovery Accelerator have Full Control permissions on this storage area. Additionally, consider configuring the Temporary Storage Area Cleanup Interval. Depending on the size of the cluster that is deployed, typically, the default value of 30 minutes can be sufficient. If greater than four nodes in the cluster, this interval must be reduced for more frequent cleanups to free up storage space.
- Improve performance for IIS 6.0: If the Discovery Accelerator server runs over IIS 6.0, an existing bug in IIS causes severe performance degradation when used along with Kerberos authentication. The hotfix described in Microsoft Knowledge Base article 917557 (http://support.microsoft.com/kb/917557) can be applied to the server in this case.

Discovery Accelerator permissions

The credentials that are used for referencing the Enterprise Vault Site are credentials of the Vault User or any other administrator. IBM StoredIQ Platform validates that the credentials are strong enough for it to:

- Login remotely to the specified server
- Perform DCOM operations over RPC remotely

To harvest a Discovery Accelerator volume successfully, a user must have the following privileges:

- A role that is defined in the Discovery Accelerator Web Application
- Review messages permission for the case that is used in the volume definition
- Folder review permissions on a case, if a folder (sometimes also called a Research Folder) in the case is going to be harvested
- Permission to set all of the review marks that are selected for the volume definition

- [Configuring security settings for Enterprise Vault servers](#)
  The following procedure addresses only the standard security settings that must be configured on the Windows Servers hosting Enterprise Vault to allow it to interact with IBM StoredIQ Platform.

- **Parent topic:** [Server platform configuration](#)

Related concepts

- [Server platform configuration](#)

Related tasks

- [Configuring security settings for Enterprise Vault servers](#)

Configuring security settings for Enterprise Vault servers

The following procedure addresses only the standard security settings that must be configured on the Windows Servers hosting Enterprise Vault to allow it to interact with IBM® StoredIQ Platform.

Procedure

1. Go to Administration > Data sources > Specify servers.
2. Click Discovery Accelerator customers, and then click Add new Discovery Accelerator customer.
3. In the Customer name field, enter a unique display name of the DA customer. This name appears in the screens that are used to configure Enterprise Vault volumes.
4. In the Discovery Accelerator server field, enter the DNS name of the physical server that runs Discovery Accelerator.
5. In the Discovery Accelerator customer ID field, enter the customer ID value that is obtained from Discovery Accelerator. Each unique combination of Discovery Accelerator server name and customer ID can be used only one time.
6. In the Customer virtual directory field, enter the IIS Virtual Directory where the Discovery Accelerator web service is located.
7. In the Discovery Accelerator installation folder field, enter the path where Enterprise Vault Business Accelerator was installed on the Discovery Accelerator server. IBM StoredIQ Platform requires this path so that Discovery Accelerator can be accessed during configuration.
8. Click OK to save the site.

- **Parent topic:** Configuration of Discovery Accelerator

Related concepts

- Configuration of Discovery Accelerator

Configuring NewsGator

When NewsGator is used as a server platform, several privileges must be configured.

About this task

Privileges Required by User Account: The user account to harvest or copy from a NewsGator volume must have the Legal Audit permission on the NewsGator Social Platform Services running on the SharePoint farm.

Procedure

1. Log in as an administrator to your SharePoint Central Administration Site.
2. Under Application Management, select Manage Service Applications.
3. In the Manage Service Applications screen, select the NewsGator Social Platform Services row.
4. From the toolbar, select Administrators.
5. Add the user account that is used for the NewsGator harvest to the list of administrators. Ensure that the account has the Legal Audit permission.

- **Parent topic:** [Server platform configuration](#)

Related concepts

- [Server platform configuration](#)

Configuration of retention servers

IBM® StoredIQ Platform supports various types of retention servers, which must be configured before retention volumes are added due to unique requirements.

Retention servers must be configured before you add retention volumes to IBM StoredIQ Platform. As IBM StoredIQ Platform supports different types of retention servers, there are various configuration options.

- [Creating Centera pools](#)
  When a Centera pool is created, an empty volume set is automatically created and associated with the Centera pool. This feature enables the harvesting of unknown Centera volumes so business policies can be applied to data objects on already retained storage servers. The volume set cannot be edited or deleted from the manage volume sets page.
- [Configuring the Dell DX Object Storage Platform](#)
  A Dell DX Storage Cluster must be defined before a Dell DX Storage volume can be added. After you define the cluster, the cluster is available in the list of available options when a Dell DX Storage volume is added.
- [Configuring Enterprise Vault](#)
  When you configure Enterprise Vault as a retention volume, you must configure both security settings and DCOM.
- [Configuring IBM Information Archive Retention Servers](#)
  Before you create retention volumes, you need to configure settings on your IBM Information Archive retention server. Files that are sent to retention must have a corresponding management class to manage the retention period settings. These classes require a defined class, policy domain, policy set, and node on the IBM Information Archive retention server.
- [NetApp SnapLock](#)
  When you configure NetApp SnapLock as a retention server, certain configuration requirements must be met.
- [Hitachi HCAP configuration requirements](#)
  In order for IBM StoredIQ Platform to access the Hitachi HCAP server, certain configuration requirements must be met.

- **Parent topic:** [Volumes and data sources](#)

Related concepts

- [Server platform configuration](#)
- [Volume indexing](#)
- [Configuring Enterprise Vault](#)
- [NetApp SnapLock](#)
- [Hitachi HCAP configuration requirements](#)

Related tasks

- [Creating Centera pools](#)
- [Configuring the Dell DX Object Storage Platform](#)
- [Configuring IBM Information Archive Retention Servers](#)
- [Configuring FileNet](#)
- [Configuring Chatter messages](#)

Creating Centera pools

When a Centera pool is created, an empty volume set is automatically created and associated with the Centera pool. This feature enables the harvesting of unknown Centera volumes so business policies can be applied to data objects on already retained storage servers. The volume set cannot be edited or deleted from the manage volume sets page.

About this task

Advanced retention feature must be enabled.

- Permissions: To support all IBM® StoredIQ Platform features, the following effective access profile rights to a Centera pool must be enabled:
  - Read
  - (D)elete
  - (Q)uery
  - (E)xist
  - (W)rite
  - Retention
  - (H)old
- ***Centera Pools***: If you have an integrated Centera server, you need to create a Centera pool before you can add a Centera volume. Unlike other volumes, Centera servers are not placed into volume sets but into Centera Pools. After you create a pool, the Centera pool is in the list of available choices when a Centera volume is added.

Procedure

1. Go to Administration > Data sources > Specify volumes.
2. In Centera pools, click Create new pool to open the Centera pool editor.
3. Enter a unique name for the Centera pool in the IBM StoredIQ Platform pool profile name text box.
4. Choose how to define the pool. Select either Structured Output or Single text field and enter a connection string.
   - Specify Access: Enter a profile name and Secret for the pool.
   - Use .pea file: This file is a pool-entry authorization. For information about creating this type of file, see *Centera Monitor and Administration Guide*.
5. If you chose Structured input, click Add new connection in the Connections section.
6. Enter the IP address for an access point on the server, and then click Add. Multiple access points can be specified for failover in the event of a problematic access node.
7. Click OK to save the pool. The pool is now available in the Add volumes dialog when a Centera volume is added.

- **Parent topic:** [Configuration of retention servers](#)

Related concepts

- [Configuration of retention servers](#)
- [NetApp SnapLock](#)
- [Hitachi HCAP configuration requirements](#)

Configuring the Dell DX Object Storage Platform

A Dell DX Storage Cluster must be defined before a Dell DX Storage volume can be added. After you define the cluster, the cluster is available in the list of available options when a Dell DX Storage volume is added.

Procedure

1. Go to Administration > Data sources > Specify Servers > Dell DX Storage Clusters.
2. In the Dell DX Storage Cluster list, select Add new Dell DX Storage Cluster.
3. Enter a unique name for the Profile name.
4. Enter the Dell DX Storage Cluster name. This name must be the same when the DX CSN is configured.
5. Enter a port number that is used to communicate with the storage cluster. The default value is 80.
6. Optional: Specify alternative addresses to be used to communicate with the storage cluster. This address must be one or more IP addresses or DNS names that map to DX Storage Nodes or DX CSN in the cluster. Individual entries must be separated by a comma.
7. Optional: Specify the IP address or DNS name for a DR Site DX CSN that can be used for failover.
8. Enter a Starting IP address for cluster. Each node in the IBM® StoredIQ Platform cluster is assigned an IP address in the Dell DX Storage Cluster private network. The Starting IP address must be an unused, valid IP address within the Dell DX Storage Cluster private network. For example, if `172.17.17.100` is entered for a five-node IBM StoredIQ Platform cluster, `172.17.17.100`, `172.17.17.101`, `172.17.17.102`, `172.17.17.103`, and `172.17.17.104` are used by IBM StoredIQ Platform.
9. Enter the Netmask value for the Dell DX Storage Cluster private network. Typically, this value is 255.255.255.0.
10. The value of the Switch port, which must be used on the IBM StoredIQ Platform switches to link with the Dell DX Storage Cluster.

- **Parent topic:** [Configuration of retention servers](#)

Related concepts

- [Configuration of retention servers](#)
- [NetApp SnapLock](#)
- [Hitachi HCAP configuration requirements](#)

Configuring Enterprise Vault

When you configure Enterprise Vault as a retention volume, you must configure both security settings and DCOM.

Before you use Enterprise Vault to create retention volumes, you must configure Discovery Accelerator customers and Enterprise Vault sites sequentially so that certain configuration items can appear in the retention volume configuration lists. You must log in to Discovery Accelerator and run the `ImportExport.exe` tool to obtain the appropriate Customer IDs and customer database names.

Note: If you use a system other than IBM® StoredIQ Platform to import data into Enterprise Vault but still want to use IBM StoredIQ Platform for exporting out of Discovery Accelerator, you must define an Enterprise Vault site within IBM StoredIQ Platform and then use that site to define a Discovery Accelerator volume. When you define the Enterprise Vault site, configuring DCOM configuration is a prerequisite.

- **Parent topic:** [Configuration of retention servers](#)

Related concepts

- [Configuration of retention servers](#)
- [NetApp SnapLock](#)
- [Hitachi HCAP configuration requirements](#)

Enabling remote DCOM

Remote DCOM is required on all Enterprise Vault servers and Discovery Accelerator.

About this task

You can use the DCOM Config utility (`DCOMCnfg.exe`) found in Administrative Tools in Control Panel to configure DCOM settings as Component Services. This utility displays the settings that enable certain users to connect to the computer remotely through DCOM. Members of the Administrators group are allowed to connect remotely to the computer by default.

If the Enterprise Vault Service Account or the user whose credentials are used to define the Enterprise Vault Site IBM StoredIQ Platform does not have permissions to connect remotely for DCOM, then follow this procedure on the target server.

Procedure

1. Run dcomcnfg as a user with administrator privileges.
2. In the Component Services dialog box, expand Component Services, expand Computers, and then right-click My Computer > Default Properties. If not already enabled, select the Enable Distributed COM on this Computer check box.
3. Click the COM Security tab.
4. Under Launch and Activation Permissions, click Edit Limits.
5. In the Launch Permission dialog box, follow these steps if the user name does not appear in the Groups or user names list:
   a. In the Launch Permission dialog box, click Add.
   b. In the Select Users, Computers, or Groups dialog box, add the user name and then click OK.
   c. In the Launch Permission dialog box, select your user and in the Allow column under Permissions for User, select Remote Launch and select Remote Activation, and then click OK.

Allowing DCOM traffic through the Windows firewall

DCOM traffic must be allowed through the Windows firewall for Enterprise Vault servers and Discovery Accelerator.

About this task

To allow DCOM traffic over the network on the target server, the DCOM TCP port (135) must be open on the Firewall. This command opens this port if it is closed:

```
netsh firewall add portopening protocol=tcp port=135
name=DCOM_TCP135
```

You can also use the Firewall User Interface to open the port.

Procedure

1. In the Control Panel, double-click Windows Firewall.
2. Click Change Settings, and then click the Exceptions tab.
3. In the Exceptions window, select the check box for DCOM to enable DCOM traffic through the firewall. If there is no such check box, you can
   a. Click Add Port.
   b. In the dialog box, enter Name as `DCOM` and Port number as `135`. Ensure that TCP is selected.
   c. Click OK.

Configuring IBM Information Archive Retention Servers

Before you create retention volumes, you need to configure settings on your IBM Information Archive retention server. Files that are sent to retention must have a corresponding management class to manage the retention period settings. These classes require a defined class, policy domain, policy set, and node on the IBM Information Archive retention server.

About this task

Use the Tivoli Storage administrative client, dsmadmc, to enter commands. You can use the administrative client in either interactive or batch mode. See *Tivoli Storage Manager Administrator's Reference* for more details about using the administrative client.

Procedure

1. Create a policy domain. For example,

   `define domain Example_PD desc='Example Domain' archret=1`

2. Create a policy set for the policy domain.

   `define policyset Example_PD Example_PS`

3. Create one or more management classes. These classes are referred to as Retention Classes in the application's user interface. Each management class has only one copy group that is assigned to it.

   - `Define mgmtclass Example_PD Example_PS Example_MG_CR desc='Example Domain chronological retention management class'`

   - `Define mgmtclass Example_PD Example_PS Example_MG_0DAY desc='Example Domain zero day retention'`

   - `Define mgmtclass Example_PD Example_PS Example_MG_10DAY desc='Example Domain 10 day retention'`

4. Create an archive copy group for each management class that is defined before. The copy group must include `retinit=creation`. For example,

   - `Define copygroup Example_PD Example_PS Example_MG_CR STANDARD type=archive destination=archivepool retver=1 retinit=creation`

   - `Define copygroup Example_PD Example_PS Example_MG_0DAY STANDARD type=archive destination=archivepool retver=0 retinit=creation`

   - `Define copygroup Example_PD Example_PS Example_MG_10DAY STANDARD type=archive destination=archivepool retver=10 retinit=creation`

5. Assign one of the management classes that are created before as the default management class for the policy domain.
6. Assign `defmgmtclass Example_PD Example_PS Example_MG_CR`.
7. Complete the definition of the policy domain by validating and activating the policy set.
8. Validate `policyset Example_PD Example_PS`.
9. Activate `policyset Example_PD Example_PS`.
10. Create a node. It acts as a connection between IBM® StoredIQ Platform and the server and is used when you define IBM StoredIQ Platform volumes. For example,

    `register node Example Domain password domain=Example_PD`

    Note: You can also use the `query node` command to list nodes.

- **Parent topic:** [Configuration of retention servers](#)

Related concepts

- [Configuration of retention servers](#)
- [NetApp SnapLock](#)
- [Hitachi HCAP configuration requirements](#)

NetApp SnapLock

When you configure NetApp SnapLock as a retention server, certain configuration requirements must be met.

Note: Use IBM® StoredIQ Platform to manage files: NetApp SnapLock retention servers do not lock out user deletion and modification of files that do not have an active retention period. Always use IBM StoredIQ Platform to manage files on retention.

When you prepare to add SnapLock retention volumes, you must have licensed SnapLock. You must create at least one SnapLock volume and share it using either Windows Share or NFS (or both), and initialize the SnapLock compliance clock. See the NetApp administrator documentation for specific instructions.

When you configure the server, set the retention period settings. Set the minimum setting to zero while not entering a maximum setting. Access the SnapLock server and enter the following commands:

```
vol options <vol-name> SnapLock_minimum_period 0d
vol options <vol-name> SnapLock_default_period min
```

- **Parent topic:** [Configuration of retention servers](#)

Related concepts

- [Configuration of retention servers](#)
- [Configuring Enterprise Vault](#)

Related tasks

- [Creating Centera pools](#)
- [Configuring the Dell DX Object Storage Platform](#)
- [Configuring IBM Information Archive Retention Servers](#)
- [Configuring FileNet](#)
- [Configuring Chatter messages](#)

Hitachi HCAP configuration requirements

In order for IBM® StoredIQ Platform to access the Hitachi HCAP server, certain configuration requirements must be met.

The IBM StoredIQ Platform application accesses the Hitachi HCAP server with HTTP. The HTTP gateway must be enabled on the server. Depending on the current allow/deny lists for the HTTP gateway, you might need to add the IBM StoredIQ Platform data server's IP addresses to the Allow IP addresses list.

- **Parent topic:** [Configuration of retention servers](#)

Related concepts

- [Configuration of retention servers](#)
- [Configuring Enterprise Vault](#)

Related tasks

- [Creating Centera pools](#)
- [Configuring the Dell DX Object Storage Platform](#)
- [Configuring IBM Information Archive Retention Servers](#)
- [Configuring FileNet](#)
- [Configuring Chatter messages](#)

Creating primary volumes

Primary volumes can be created as data sources.

Procedure

1. Go to Administration > Data sources > Specify volumes > Volumes.
2. On the Primary volume list page, click Add primary volumes.
3. Enter the information that is described in the following table, which is based on your server type.
4. Click OK to save the volume.
5. Select one of the following options:
   - Add another volume on the same server.
   - Add another volume on a different server.
   - Finished adding volumes.

This table describes the fields that are available in the Add volume dialog box when you configure primary volumes.

Note: Case-sensitivity rules for each server type apply. Red asterisks within the user interface denote the fields.

Table 1. Primary volumes: Fields, required actions, and applicable volume types

Field name	Required action	Special notes	Applicable volume type
			- CIFS (Windows platforms) - NFS v2 and v3 - Excha

$-$ $-$

Server type	In the Server type list, select the server type.

- For Documentum, you must specify the doc broker. See [Configuring Documentum](#).
- For Discovery Accelerator, you must first create a Discovery Accelerator site and an Enterprise Vault site.
- For Domino, you must first upload at least one user.id. See [Adding Domino as a Primary Volume](#).
- For Desktop, the desktop agent must be installed on that desktop and then pointed to the data server. Desktop does not appear in the Server type list. Instead, it appears as an available server or volume.
- For FileNet, see [Configuring FileNet](#)
- For Chatter, see [Configuring Chatter messages](#).

- ...nge
- SharePoint
- Documentum
- Discovery Accelerator
- Domino
- FileNet
- NewsGator
- Livelink
- Jive

		○ Chatter
		○ Exchange
		○ CIFS (Windows platforms)
Version	In the Version list, select the appropriate version.	
		○ NFS v2 and v3
		○ NetApp
		○ Exchange
	○ For Exchange primary volumes, it is the fully qualified domain name where the OWA is. Multiple Client Access servers on Exchange 2007 are supported. The server must be load-balanced at the IP or DNS level.	
	○ For NetApp, select the predefined server from the list.	
Server	In the Server text box, enter the fully qualified name of the server where the volume is available for mounting.	
	○ For Domino, select the appropriate user name, which was entered with the Configuration subtab in the Lotus Notes user administration area.	
	○ When you add SharePoint volumes that contain spaces in the URL, see [Special Note: Adding SharePoint Volumes](#).	○ SharePoint
	○ For Desktop, the desktop agent must be installed on that	

	desktop and then pointed to the data server. Desktop does not appear in the Server type list. Instead, it appears as an available server or volume.	○ Domino ○ FileNet ○ NewsGator ○ Jive ○ Livelink ○ Discovery Accelerator
Customer	In the Customer list, select the customer for this volume.	
FileNet config	Use the FileNet config list to select the FileNet server you would like to use for this configuration.	For more information, see [Configuring FileNet](#). ○ FileNet
Mailbox server	When you configure multiple client access servers, enter the name of one or more mailbox servers, which are separated by a comma.	For Exchange primary volumes, it is the fully qualified domain name where the mailbox to be harvested is. ○ Exchange ○ Exchang

Active Directory server	In the Active Directory server text box, enter the name of the Active Directory server.	It must be a fully qualified Active Directory server.	e ○ Discovery Accelerator ○ SharePoint ○ Exchange ○ Discovery Accelerator
Protocol	To use SSL, select the Protocol check box.	API client uses HTTP over SSL to communicate with Discovery Accelerator Server configuration.	○ SharePoint ○ News s

Doc base	In the Doc base text box, enter the name of the Documentum repository.	A Documentum repository contains cabinets, and cabinets contain folders and documents.
Port	In the Port text box, enter the name of the port.	
Database	In the Database text box, enter the name of the database.	
Search slice	In the Search slice text box, enter the name of the search slice.	

- Gator
- Jive
- Documentum
- Livelink
- Livelink
- Livelink
- CIFS (Windows platforms)
- Exchange
- Shar

Conn ect as	In the Connect as text box, enter the logon ID that is used to connect and mount the defined volume.	For Domino, select the user name for the primary user ID. The user ID must be configured on the System Configuration screen under the Lotus Notes user administration link.	e P oi nt D o c u m e nt u m
○ D o m in o
○ D is c o v e ry A c c el e r at o r
○ Fi le N et
○ N e w s G at o r
○ Ji v e
○ C h at te r
○ Li v el in k
○ C I

			FS (Windows platforms)
			○ Exchange
			○ SharePoint
			○ Documentum
Password	In the Password text box, enter the password that is used to connect and mount the defined volume.	For Domino, enter the password for the primary user ID.	○ Domino
			○ Discovery Accelerator or

Field	Description	Notes	
			◦ FileNet
			◦ NewsGator
			◦ Jive
			◦ Chatter
			◦ Livelink
			◦ DiscoveryAccelerator
Site	In the Site list, select the site of the volume.		
Auth token	In the Auth token text box, enter the token that is used to authenticate the Chatter volume.	The auth token must match the user name that is used in the Connect as field. Auth tokens can be generated online on Salesforce. See [Configuring chatter messages](#).	◦ Chatter
Object store	In the Object store list, select the wanted object store.	The object store must exist before the creation of a FileNet primary volume.	◦ FileNet
			◦ CIFS

| Volume | In the Volume text box, enter the name or names of the volume to be mounted. |

- ○ (Windows platforms)
- ○ NFS v2 and v3
- ○ NetApp
- ○ Exchange
- ○ SharePoint
- ○ Documentum
- ○ Disc overy A

- ○ For Exchange, Domino, Documentum, Enterprise Vault (primary volume), and NewsGator, enter a friendly name for the volume.
- ○ When you add SharePoint volumes that contain spaces in the URL, see [Special Note: Adding SharePoint Volumes](#).

			c c el e r at o r
			○ D o m in o
			○ Fi le N et
			○ N e w s G at o r
			○ Ji v e
			○ C h at te r
			○ Li v el in k
			○ D is c o v e ry A c c el e r at o r
Disc over y Acce lerat or case	In the Discovery Accelerator case text box, enter the name of the Discovery Accelerator.	This text box is populated from Discovery Accelerator when connection is established.	○ D is c o v e ry
Scop	In the Scope of data retrieval product with will		

Scope of data	In the Scope of data options, select either All data in this case or Only selected review marks.		Accelerator ○ Documentum ○ Domino
Harvest	It is the Documentum harvest option: ○ To enable harvesting, select Harvest all document versions. ○ To harvest mailboxes, select the Harvest mailboxes option. ○ To harvest mail journals, select the Harvest mail journals option. ○ To harvest all applications, select the Harvest all applications option.	This option obtains the list of all known Domino users and their NSFs. It then harvests those mailboxes unless it was pointed to a single mailbox with the initial directory.	○ Exchange ○ CIFS (Windows platforms) ○ NFSv2 and v3
Folders	In Folder, select either of the Mailboxes or Public folders options.		
		○ With this feature, you can select a volume further down the directory	○ Net

Initial directory	In the Initial directory text box, enter the name of the initial directory from which the harvest must begin.	the directory tree rather than selecting an entire volume. ○ For Exchange, this field must be left blank if you are harvesting all mailboxes. If you are harvesting a single mailbox, enter the email address for that mailbox. ○ When you add SharePoint volumes that contain spaces in the URL, see [Special Note: Adding SharePoint Volumes](#).	AppExchange ○ ExchangeSharePoint ○ SharePoint ○ Document ○ FileNet ○ Jive ○ Chatter ○ Livelink ○ Exchange
Virtual root	In Virtual root, change the default name to match the Exchange server designation.	For Exchange, this option must be changed to match the server designation.	○ Exchange
Personal archives	To enable the collection of personal archives, select the Harvest personal archive check box.	This option pertains only to Exchange 2010 with SP1 applied.	○ Exchange

		2010
		○ CIFS (Windows platforms)
		○ NFS v2 and v3
		○ NetApp
		○ Exchange
		○ SharePoint
		○ Documentum
Index options	Select either or both of the Index options check boxes. ○ Include system metadata for data objects within containers. ○ Include content tagging and full-text index.	These options are selected by default.

- Discovery Accelerator
- Domino
- FileNet
- NewsGator
- Jive
- Chatter
- Desktop
- Livelink
- SharePoint

Subsites   Select Recurse into subsites.

Versions	Select Include all versions.	○ IBM® StoredIQ Platform supports indexing versions from SharePoint. For more information, see [Special Note: Adding SharePoint Volumes](#).	○ SharePoint
			○ Jive
Remove journal envelope	When selected, the journal envelope is removed.		○ Exchange
			○ CIFS (Windows platforms)
			○ NFS v2 and v3
			○ NetApp
			○ Exchange
			○ Sh

| Valid ation | To validate volume accessibility, select Validation. | When selected (the default state), IBM StoredIQ Platform tests to see whether the volume can be accessed. |

arePoint

- Documentum
- Discovery Accelerator
- Domino
- NewsGator
- Jive
- Chatter
- Livelink
- CIFS

| Include directories *or* Include research folders | In Include directories, specify a regular expression for included directories for each harvest (if it was specified). | ○ These directories are defined as sets of "first node" directories, relative to the specified (or implied) starting directory, that is considered part of the logical volume.
○ For Discovery Accelerator, this regular expression helps restrict the volume to one or more Research Folders in the case. | (Windows platforms)
○ NFS v2 and v3
○ NetApp
○ Exchange
○ SharePoint Document
○ Document
○ Discovery A |

ccelerator

- Domino
- FileNet
- Jive
- Chatter
- Desktop
- Livelink
- CIFS (Windows platforms)
- NFS v2 and

Start direc tory	In Start directory, designate a start directory for the harvest. The start directory involves volume partitioning to break up a large volume. If an initial directory is defined, the start directory must be underneath the initial directory. For example, in the case of directories E-H, E would be the start directory and H would be the end directory.

- v3
  - NetApp
  - Exchange
  - SharePoint
  - Domino
  - FileNet
  - Jive
  - Chatter
  - Livelink
  - CIFS (Windows platform

End direc tory	In End directory, determine the end directory for the harvest. The end directory is also part of volume partitioning and is the last directory harvested.

- s)
  - NFS v2 and v3
  - Exchange
  - NetApp
  - SharePoint
  - Domino
  - FileNet
  - Jive
  - Chatter
  - Livelink
  - CIFS (W

Access Times	In Access times, select one of these options:
	○ Reset access times but do not synchronize them. (This setting is the default setting.) ○ Do not reset or synchronize access times. ○ Reset and synchronize access times on incremental harvests.

indows platforms)

- NFS v2 and v3
- NetApp
- Domino
- CIFS (Windows platforms)
- NFS v2 and v3
- NetA

| Volume constraints or Constraints | In Constraints, select one of these options:
○ Only use __ connection process (es): Specify a limit for the number of harvest connections to this volume. If the server is also being accessed for attribute and full-text searches, you might want to regulate the load on the server by limiting the harvester processes. The maximum number of harvest processes is automatically shown. This maximum number is set on the system configuration tab.
○ Control the number of parallel data object reads: Designate the number of parallel data object reads.
○ Scope harvests on these volumes by extension: Include or exclude data objects that are based on extension. | The Scope harvests on these volumes by extension option pertains only to these volume types:
○ Windows Share
○ NFS v2 and v3
○ NetApp
○ Celerra
○ Discovery Accelerator | ○ pp Exchange
○ Domino
○ SharePoint
○ Document um
○ Discovery Accelerator
○ FileNet
○ Jive
○ Chatter
○ D |

- [Configuring Exchange 2007 Client Access Server support](#)
  The system supports the harvest of multiple Client Access Servers (CAS) when you configure Exchange 2007 primary volumes. This feature does not support redirection to other CAS/Exchange clusters or autodiscovery protocol.
- [Adding Domino as a primary volume](#)
  Domino volumes can be added as primary volumes.
- [Special note: SharePoint volumes](#)
  Certain fields must be configured when SharePoint volumes are added.
- [Configuring FileNet](#)
  By providing the configuration values for a FileNet domain, you are supplying the values that are needed to bootstrap into a domain.
- [Configuring Chatter messages](#)
  Within Chatter, the default administrator profile does not have the Manage Chatter Messages permission, but the appropriate permissions are required to harvest private messages.

- **Parent topic:** [Volumes and data sources](#)

Related concepts

- [Volumes and data sources](#)
- [Special note: SharePoint volumes](#)

Related tasks

- [Configuring Exchange 2007 Client Access Server support](#)
- [Adding Domino as a primary volume](#)
- [Configuring primary volumes by using Enterprise Vault](#)

Configuring Exchange 2007 Client Access Server support

The system supports the harvest of multiple Client Access Servers (CAS) when you configure Exchange 2007 primary volumes. This feature does not support redirection to other CAS/Exchange clusters or autodiscovery protocol.

Procedure

1. Go to Administration > Data sources > Volumes > Primary > Add primary volumes.
2. In the Server type list, select Exchange.
3. In the Version list, select 2007.
4. In the Server text box, type the name of the Exchange server. This server must be load-balanced at the IP or DNS level.
5. In the Mailbox server: text box, enter the name of one or more mailbox servers, which are separated by a comma and a space.
6. Complete the remaining fields for the primary volume, and then click OK.

- **Parent topic:** Creating primary volumes

Related concepts

- Special note: SharePoint volumes

Related tasks

- Creating primary volumes

Adding Domino as a primary volume

Domino volumes can be added as primary volumes.

Procedure

1. Add a Lotus Notes user by uploading its user ID file in Lotus Notes User Administration on the Administration > Configuration tab.
   - If you want to harvest a user's mailbox, add the user ID file for that user.
   - If you want to harvest multiple mailboxes within one volume definition, add the administrator's ID file.
   - If the mailboxes have encrypted emails or NSFs, then you need each user's user ID file to decrypt a user's data.
2. Point the volume to the Domino server. If a single mailbox must be harvested, set the initial directory to be the path to the mailbox on the Domino server, such as `mail\USERNAME`.
3. To harvest mailboxes, select the Harvest mailboxes option, which obtains the list of all known Domino users and their NSFs. It then harvests those mailboxes unless it was pointed to a single mailbox by using the initial directory.
4. To harvest all mail journals, select the Harvest mail journals option.
5. To harvest all mail applications, select the Harvest all applications option, which looks at all NSFs, including mail journals, on the Domino server.

- **Parent topic:** [Creating primary volumes](#)

Related concepts

- [Special note: SharePoint volumes](#)

Related tasks

- [Creating primary volumes](#)

Special note: SharePoint volumes

Certain fields must be configured when SharePoint volumes are added.

IBM® StoredIQ Platform supports the entire sites portion of a Sharepoint URL for the volume `/sites/main_site/sub_site` in the Volume field when you add a SharePoint volume. However, if the SharePoint volume URL contains spaces, then you must also use the Server, Volume, and Initial directory fields in the Add volume dialog box in addition to the required fields Server type, Server, Connect as, and Password. For example, the SharePoint volume with the URL `http://shpt2010.reglab5.local/sitestest/autoteamsite1/Attribute Harvest WikiPages Library/` would require the fields in the following table because of the spaces in the URL.

Table 1. SharePoint volumes as primary volumes: Fields and examples

Primary volume field	Example
Server	`shpt2010.reglab5.local`
Volume	`/sitestest/autoteamsite1`
Initial director	`Attribute Harvest Wiki Pages Library`

Performance conditions for using versions

When you add a primary volume, you define the volume by setting certain properties. If a SharePoint volume is added, you have the option of indexing different versions of data objects on that volume.

Since most versions of any object shares full-text content and attributes, The effort in processing them and maintaining an updated context for the version history of an object in the index is duplicated. Additionally, if you enable version feature on a SharePoint volume, the API itself causes extra overhead in fetching data and metadata for older versions.

- For each object, an extra API call must be made to get a list of all its versions.
- To fetch attributes for the older versions of an object, an API call must be made for each attribute that needs to be indexed.

- **Parent topic:** [Creating primary volumes](#)

Related tasks

- [Creating primary volumes](#)
- [Configuring Exchange 2007 Client Access Server support](#)
- [Adding Domino as a primary volume](#)
- [Configuring primary volumes by using Enterprise Vault](#)

Configuring FileNet

By providing the configuration values for a FileNet domain, you are supplying the values that are needed to bootstrap into a domain.

Procedure

1. Go to Administration > Data sources > Specify Servers > FileNet domain configurations.
2. Click Add new FileNet domain configuration The FileNet domain configuration editor page appears.
3. In the FileNet domain configuration editor page, configure these fields:
   a. In the Configuration name text box, enter the configuration name for this server.
   b. In the Server name text box, enter the server name.
   c. In the Connection list, select the connection type.
   d. In the Port text box, enter the port number.
   e. In the Path text box, enter the path for this server.
   f. In the Stanza text box, enter the stanza information for this server.
4. Click OK to save your changes.

- **Parent topic:** Creating primary volumes

Related concepts

- Configuration of retention servers
- NetApp SnapLock
- Hitachi HCAP configuration requirements

Configuring Chatter messages

Within Chatter, the default administrator profile does not have the Manage Chatter Messages permission, but the appropriate permissions are required to harvest private messages.

Before you begin

A user must have certain administrative permissions when that user account is used in the Connect as text box in Chatter. When you set up a Chatter user account to harvest and run actions against Chatter, you must use an account with the built-in system administrator profile. In general, however, these administrative permissions must be assigned to the account you use:

- API enabled
- Manager Chatter Messages (required if you want to harvest Chatter Private Messages)
- Manage Users
- Moderate Chatter
- View All Data

Procedure

For Chatter administrators who use the Auth token option, see how to set up a sandbox account.

- **Parent topic:** Creating primary volumes

Related concepts

- Configuration of retention servers
- NetApp SnapLock
- Hitachi HCAP configuration requirements

Creating retention volumes

Retention volumes store data objects that are placed under retention. It means that the object is retained.

Procedure

1. Configure your retention servers.
2. Create management or retention classes, or, if you are using Centera retention servers, create Centera pools.
3. If you are using Enterprise Vault, ensure that you defined Enterprise Vault Sites. You might also want to define Retention Categories on the Enterprise Vault server.
4. Create retention volumes.

- [Adding a retention volume](#)
  Retention volumes can be added and configured.
- [Configuring Enterprise Vault retention volumes](#)
  Enterprise Vault retention volumes can be configured on an as-needed basis.

- **Parent topic:** [Volumes and data sources](#)

Related concepts

- [Volumes and data sources](#)

Related tasks

- [Adding a retention volume](#)
- [Configuring Enterprise Vault retention volumes](#)

Adding a retention volume

Retention volumes can be added and configured.

Procedure

1. Go to Administration > Data sources > Volumes, and then click Retention.
2. Depending on the type of retention server you are adding, complete the fields as described in this table.
3. Click OK to save the volume.
   Note: Case-sensitivity rules apply. Red asterisks within the user interface denote required fields.

Table 1. Retention volume: Fields, required actions, and applicable volume types

Field name	Required action	Special notes	Applicable volume type
			○ Windows Share (CIFS) ○ NFS v3 ○ Centera ○ Hitachi ○ IBM Informat

Server type	In the Server type list, select the server type.	...ion Archive
		○ Del DX Storage
		○ Enterprise Vault
		○ NetApp SnapLock
		○ NetApp Standard
		○ Centera
Platform	In the Platform list, select the platform type.	
Pool	In the Pool list, select the StoredIQ pool profile name to provide access to a specific Centera pool.	

Site	In the Site list, select the appropriate site.
Vault store	In the Vault store list, select the wanted vault store for the volume.
Dell DX Storage Cluster	In the Dell DX Storage Cluster list, select the appropriate Dell DX Storage Cluster.
	○ Windows Share (CIFS) ○ NFS v3 ○ Hitachi ○ IBM Information Archive ○ N
Server	In the Server text box, assign the server a name.

Node name	In the Node nametext box, enter the name of the node.
Node port	In the Node port text box, enter the node's port number.
Node password	In the Node passwordtext box, enter the password for the node.

Con
nect In the Connect as text box, enter the login ID.
as

Pass
wor In the Password text box, enter the password for the login ID.
d

o
w
s
S
h
a
r
e
(
C
I
F
S
)
○ I
B
M
In
fo
r
m
at
io
n
A
rc
hi
v
e

○ W
in
d
o
w
s
S
h
a
r
e
(
C
I
F
S
)
○ I
B
M
In
fo
r
m
at
io
n
A
rc
hi
v
e
--

- NetAppSnapLock

- WindowsShare (CIFS)

- NFSv3

- Centera

- Hitachi

- IBM Information Archiv

Volume	In the Volume text box, enter the name or names of the volume to be mounted.

e

- ○ Del DX Storage

e

- ○ NetApp SnapLock
- ○ Windows Share (CIFS)
- ○ NFS v3
- ○ Centera
- ○ Hitachi
- ○ IBM-

These options are select

Index

Select either or both of the Index options check boxes.
- ○ Include system metadata for data objects within containers.

options	• Include content tagging and full-text index.	ed by default.	Information Archive ○ DellDX Storage ○ NetApp SnapLock ○ Enterprise Vault ○ Enterprise Vault
Matter archive name	In the Matter archive name text box, enter the name of the matter archive.		
Description	In the Description text box, enter a description.		

Default retention category	In the Default retention category list, select the default retention category.	
Retention category override	Select the Allow retention category to be overridden on policy.	
Constraints	In Constraints, select either or both of these options: - Only use __ connection process(es): Specify a limit for the number of harvest connections to this volume. If the server is also being accessed for attribute and full-text searches, you might want to regulate the load on the server by limiting the harvester processes. The maximum number of harvest	- Windows Share (CIFS) - NFS v3 - Center a - Hitachi - IBM Information

processes is automatically shown. This maximum number is set on the system configuration tab.
- Control the number of parallel data object reads: Designate the number of parallel data object reads.

- Archive Dell DX Storage
- NetApp SnapLock
- Enterprise Vault

- **Parent topic:** [Creating retention volumes](#)

Related tasks

- [Creating retention volumes](#)

Configuring Enterprise Vault retention volumes

Enterprise Vault retention volumes can be configured on an as-needed basis.

Procedure

1. Go to Administration > Data sources > Volumes > Retention.
2. Click Add retention volumes.
3. In the Server type list, select Enterprise Vault.
4. In thenSite list, select the Enterprise Vault site you created. See Enterprise Vault.
5. Enter the information described in Creating Retention Volumes, based on your server type.
6. Click OK to save the volume.
7. Select one of the following:
    - Add another volume on the same server
    - Add another volume on a different server
    - Finished adding volumes

- **Parent topic:** Creating retention volumes

Related tasks

- Creating retention volumes

Creating discovery export volumes

Discovery export volumes contain the data produced from a policy, which is kept so that it can be exported as a load file and uploaded into a legal review tool. Administrators can also configure discovery export volumes for managing harvest results from cycles of a discovery export policy.

Procedure

1. Go to Administration > Data sources > Specify volumes, and then click Volumes.
2. Click Discovery export, and then click Add discovery export volumes.
3. Enter the information described in the table below, and then click OK to save the volume. Note: Case-sensitivity rules for each server type apply. Red asterisks within the user interface denote required fields.

Table 1. Discovery export volumes: fields, required actions, and applicable volume types.

Field	Required action	Applicable volume type
Server type	Using the Server type list, select the type of server.	○ CIFS (Windows platforms) ○ NFS v2, v3 ○ CIFS (Windows platforms)
Server	In the Server text box, enter the name of the server where the volume is available for mounting.	pl at fo

r

rms)
- NFSv2, v3
- CIFS (Windows platforms)
- CIFS (Windows platforms)
- CIFS (Windows platforms)

Connect as — In the Connect as text box, enter the logon ID used to connect and mount the defined volume.

Password — In the Password text box, enter the password used to connect and mount the defined volume.

Volume — In the Volume text box, enter the name of the volume to be mounted.

rm

Constraints

To utilize Constraints, select this option:

- Only use __ connection process (es): Specify a limit for the number of harvest connections to this volume. If the server is also being accessed for attribute and full-text searches, you may want to regulate the load on the server by limiting the harvester processes. The maximum number of harvest processes is automatically shown. This maximum number is set on the system Configuration tab.

- **Parent topic:** [Volumes and data sources](#)

Related concepts

- [Volumes and data sources](#)

Creating system volumes

System volumes support volume export and import. When you export a volume, data is stored on the system volume. When you import a volume, data is imported from the system volume.

Procedure

1. Go to Administration > Data sources > Specify volumes > Volumes.
2. Select the System tab, and then click Add system volumes.
3. Enter the information described in the table below, and then click OK to save the volume. Note: Case-sensitivity rules apply. Red asterisks within the user interface denote required fields.

Table 1. System volume fields, descriptions, and applicable volume types

Field name	Required action	Special notes	Applicable volume type
Server type	Using the Server type list, select the type of server.		- CIFS (Windows platforms) - NFS v2, v3
Server	In the Server text box, enter the name of the server where the volume is available for mounting.		CIFS (Windows platforms)

r		forms)
		○ NFS v2, v3
		○ CIFS (Windows platforms)
Connect as	In the Connect as text box, enter the logon ID used to connect and mount the defined volume.	○ CIFS (Windows platforms)
Password	In the Password text box, enter the password used to connect and mount the defined volume.	○ CIFS (Windows platforms)
Volume	In the Volume text box, enter the name of the volume to be mounted.	

s)
- NFS v2, v3
- CIFS (Windows platforms)
- NFS v2, v3

Constraints — To utilize Constraints, select this option:
- Only use __ connection process (es): Specify a limit for the number of harvest connections to this volume. If the server is also being accessed for attribute and full-text searches, you may want to regulate the load on the server by limiting the harvester processes. The maximum number of harvest processes is automatically shown. This maximum number is set on the system Configuration tab.

- **Parent topic:** [Volumes and data sources](#)

Related concepts

- [Volumes and data sources](#)

Export and import of volume data

The system's export and import volume capabilities allow metadata and full-text indexed data to be collected or exported from separate locations, such as data servers located in various offices of the enterprise. Once the data is available, it can be imported to a single location, such as a headquarter's office data server, where selected files might be retained.

Only primary and retention volume data can be exported or imported using the export/import feature. Discovery export and system volumes cannot be imported or exported. The target location of an export or the source location of an import is always the IBM® StoredIQ Platform system volume.

Export and import volume processes run as jobs in the background. These jobs are placed into their prospective queues, and they are executed sequentially. When one job completes, the next one automatically starts. These jobs can be cancelled at any time while they are running. Cancelling one import or export job also cancels all the jobs that come after the one cancelled. Because the export jobs and import jobs reside in separate queues, cancelling one type of job does not cancel jobs in the other queue. The jobs cannot be restarted.

- **Parent topic:** [Volumes and data sources](#)

Related concepts

- [Volumes and data sources](#)

Volume data export to a system volume

When exporting volume data to a system volume, the export process creates two files: a binary file and a metadata file, which contains the exported data.

About this task

The export process creates two files: a binary file and a metadata file, which contains the exported data. These files' names contain the following information:

- Data server name and IP address
- Volume and server names
- Time stamp

Note the following:

- The exported data consists of data from the selected volume and any related information that describes that data with the exception of volume-specific audits.
- The exported data has to be made available to the import data server before it can be imported. This may require you to physically move the exported data to the system volume of the import data server.
- Licenses on the import appliances are enabled automatically if a feature of the imported volume requires it (such as Centera or Exchange licenses).

Procedure

1. Go to DSAdmin > Administration > Data sources > Volumes.
2. Select a volume of data to export from the list of volumes by clicking the export link in the far right-hand column.
3. Complete the Export volumes details dialog, described in this table.
4. Click OK. A dialog appears to let you know that the data is being exported.
5. To monitor the export progress, click the Dashboard link. To cancel the export process, under the Jobs in progress section of the dashboard, click Stop this job. Alternately, click OK to return to the Volumes page.

Note: The job cannot be restarted.

Option	Description
Server	The name of the server where the data resides.
Volume	The name of the volume where the data resides.
Export path (on system volume)	This path is where to save the data on the system volume. The default path is /exports. You can edit the export path. The specified location will be automatically created if necessary.
Description (optional)	Enter a description of the exported data.
Export full-text index	Select this option to export the volume's full-text index. (Available only if the volume has full-text index.)

Volume data import from a system volume

An imported volume looks, acts, and is, just like a volume originally defined and harvested on the data server with a few key differences.

About this task

For imported volumes, any action or relationship that is valid for a non-imported volume is valid for an imported volume, with a few exceptions:

- Logs and audit trails that capture the activity on the volume before the import are not available. However, the import itself is audited.
- The imported volume can be reharvested as long as the appliance has the proper network access and rights to the original source server and volume.
- The imported volume can be reharvested as long as the data server has the proper network access and rights to the original source server and volume.
- The data viewer works only if the appliance has the proper network access and rights to the source server and volume. You must have access and permission on export servers and volumes if the file you want to view has been migrated to a secondary server at the time of the export.

Note: When a volume with a licensed feature is imported into a data server that does not utilize licensing, the license is imported along with the volume. To see the licensed features, users will need to log out and then log back in to the data server.

Procedure

1. Make sure that exported data file is present in the system volume of the import appliance.
2. Go to Administration > Data sources > Volumes, and then click either the Primary or Retention tab.
3. Click the Import volume link at the top of either the primary or the retention volume lists. The Import volumes page appears, listing all of the volumes available for import. By default, the data server searches for available volumes to import in the `/imports` directory of the system volume. If you have placed the exported data to another path, click Change path and enter the appropriate path.
4. Click OK. The Import volumes page now displays the following information about the imported volumes. See the table below.
5. From the list of volumes, select a volume to import by clicking the Import link in the last column on the right.
6. Select the Import full-text index check box to import the selected volume's full-text index. (The check box is active only if full-text index is available.)
7. Select the Overwrite existing volume check box to replace the existing data of the volume with the imported data.
8. Click OK. A dialog appears to let you know that the volume is being imported. The OK button is not enabled if the volume already exists and unless the Overwrite existing

volume option is selected.

9. To view import progress, click the Dashboard link in the dialog. To cancel the import process, under the Jobs in progress section of the dashboard, click Stop this job, or click OK to return to the Manage volumes page.

Option	Description
Server and volume	The server name and volume name where the data physically resides.
Description	A description added when the volume was exported.
Volume type	The type of volume, namely Exchange, SharePoint, and so on.
Category	This is the category of volume, namely Primary or Retention.
Exported from	The server name and IP address of the server from which the data was exported.
Export date	The day and time the data was exported.
Total data objects	The total number of data objects exported for the exported volume.
Contains full-text index	Indicates whether or not the full-text index option was chosen when the data was exported.

Deleting volumes

Administrators can delete volumes from the list of available data sources, provided that the data server is connected to the gateway.

About this task

Note the following regarding deleted volumes:

- Deleted volumes are removed from target sets.
- Deleted volumes are removed from all volume lists, both from IBM® StoredIQ Platform and IBM StoredIQ Administrator.
- Within created jobs, steps that reference deleted volumes are implicitly removed, meaning that a job could contain no steps. The job itself is not deleted.
- Applicable object counts and sizes within IBM StoredIQ Administrator adjust automatically.
- Object counts and sizes within user infosets will remain the same. Remember, those user infosets were created at a specific point in time when this data source was still available.
- Users exploring a specific data source and any generated reports will no longer reference the deleted volume.
- No exceptions is raised on previously executed actions. Instead, the data is no longer available. For example, if an infoset is copied that contained data objects from a volume that has been deleted, no exception is raised.
- If you mark a desktop volume for deletion, it is automatically removed from the Primary volume list; however, the status of that workstation is set to uninstall in the background. When the desktop client next checks in, it will see that change in status and uninstall itself.

Note: If retention volumes such as Centera, Hitachi, and so on contain data, they cannot be deleted as IBM StoredIQ Platform is the source of record. Instead, you see the Under Management link.

To delete a volume:

Procedure

1. Go to Administration > Data sources > Specify volumes > Volumes.
2. Click the tab of the volume type you would like to delete: Primary, Retention, System, or Discovery export.
3. Click Delete, and in the confirmation dialog, click OK. The volume is deleted, removing it from the list of available volumes.

- **Parent topic:** <u>Volumes and data sources</u>

Related concepts

- <u>Volumes and data sources</u>

Policy limitations for volume types

Volume types have different policy limitations.

IBM® StoredIQ imposes some policy limitations on volume types, which are identified in this table.

Table 1. Policy limitations for volume types

Policy type	Source/Target Limitations	Other restrictions
Copy from	• Centera • Windows Share SnapLock • NFS SnapLock • Hitachi • IBM Information Archive • Windows Share • NFS • Documentum • SharePoint • Exchange • Enterprise Vault (Discovery Accelerator) • Dell DX Storage • NewsGator	
Copy to	Primary Volume • CIFS • NFS • SharePoint • Documentum • FileNet Retention Volume • Centera • NetApp SnapLock (CIFS/NFS) • CIFS Celerra FLR (CIFS/NFS) • Dell DX Object Store • CIFS • NFS • Documentum /w RPS • Hitachi HCAP • IBM Information Archive • Symantec Enterprise Vault	
Copy to (retention)	• Centera • Windows Share SnapLock • NFS SnapLock • Hitachi • IBM Information Archive • Windows Share • NFS • Documentum • Enterprise Vault • Dell DX Storage • Centera	

Discovery export from	- Windows Share SnapLock - NFS SnapLock - Hitachi - IBM Information Archive - Windows Share - NFS - Documentum - SharePoint - Exchange - Enterprise Vault (Discovery Accelerator - Dell DX Storage - NewsGator	
Discovery export to	- Windows Share - NFS	Considered discovery export volumes (category) not harvested
Move from	- Centera - Windows Share SnapLock - NFS SnapLock - Hitachi - IBM Information Archive - Windows Share - NFS - Documentum - Dell DX Storage	
Move to	- Centera - Windows Share SnapLock - NFS SnapLock - Hitachi - IBM Information Archive - Windows Share - NFS - Documentum - SharePoint - Dell DX Storage	
Delete	- Centera - Windows Share SnapLock - NFS SnapLock - Hitachi - IBM Information Archive - Windows Share - NFS - Documentum - Dell DX Storage	
Modify security	CIFS	Choose single volume (no volume sets)

- **Parent topic:** Volumes and data sources

Related concepts

- Volumes and data sources

Data harvesting

Harvesting (or indexing) is the process or task by which IBM® StoredIQ Platform examines and classifies data in your network.

Running a Harvest every volume job indexes all data objects on all volumes.

- A full harvest can be run on every volume or on individual volumes.
- An incremental harvest only harvests the changes on the requested volumes

These options are selected when you create a job for the harvest. A harvest must be run before you can start searching for data objects or textual content. An Administrator initiates a harvest by including a harvest step in a job.

Most harvesting parameters are selected from the Configuration subtab. You can specify the number of processes to use during a harvest, whether a harvest must continue where it left off if it was interrupted and many other parameters. Several standard harvesting-related jobs are provided in the system.

Harvesting with and without post-processing

You can separate harvesting activities into two steps: the initial harvest and harvest post-processing. The separation of tasks gives Administrators the flexibility to schedule the harvest or the post-process loading to run at times that do not affect system performance for system users. These users might, for example, be running queries. Here are examples of post-harvest activities:

- Loading all metadata for a volume.
- Computing all tags that are registered to a particular volume.
- Generating all reports for that volume.
- If configured, updating tags, and creating explorers in the harvest job.

Incremental harvests

Harvesting volumes takes time and taxes your organization's resources. You can maintain the accuracy of the metadata repository quickly and easily with incremental harvests. With both of these features, you can ensure that the vocabulary for all volumes is consistent and up to date. When you harvest a volume, you can speed up subsequent harvests by only harvesting for data objects that were changed or are new. An incremental harvest indexes new, modified, and removed data objects on your volumes or file servers. Because the harvests are incremental, it takes less time to update the metadata repository with the additional advantage of putting a lighter load on your systems than the original harvests.

Note: Harvesting NewsGator Volumes: Since NewsGator objects are just events in a stream, an incremental harvest of a NewsGator volume fetches only new events that were added since the last harvest. To cover gaps due to exceptions or to pick up deleted events, a full harvest might be required.

- [Harvesting properties and libraries](#)
  When harvesting private information, SharePoint volumes must use administrative roles for mounting permission.
- [Lightweight harvest parameter settings](#)
  To conduct a lightweight harvest, certain configuration changes can be made.

Harvesting properties and libraries

When harvesting private information, SharePoint volumes must use administrative roles for mounting permission.

Note: Administrative permissions are required to harvest personal information, libraries, and objects that are not designated as being visible to Everyone for user profiles.

The SharePoint volume needs to be mounted with administrative permissions. If the harvest is conducted without administrative permissions, then any of the user profile's properties that were marked as visible to the category other than Everyone is not visible in results. To harvest users' personal documents and information, volumes mounted without administrative permissions must use credentials that have full control on all SharePoint site collections. These collections are hosted by the user profile service application.

To override this restriction, see http://technet.microsoft.com/en-us/library/ee721057.aspx.

- **Parent topic:** Data harvesting

Related concepts

- Data harvesting
- Lightweight harvest parameter settings

Lightweight harvest parameter settings

To conduct a lightweight harvest, certain configuration changes can be made.

With IBM® StoredIQ Platform, you can conduct many types of harvests, depending on your data needs. While in-depth harvests are common, there are also instances where you need an overview of the data and a systemwide picture of files' types and sizes. For example, at the beginning of a deployment, you might want to obtain a high-level view of a substantial amount of data. It helps make better decisions about how you want to handle harvesting or other policies in the future. The following section provides possible system configurations for the system to process the volumes' data in the quickest manner possible.

- [Determining volume configuration settings](#)
  To conduct a lightweight harvest, you can make various configuration changes.
- [Determining harvester configuration settings](#)
  When you conduct a lightweight harvest, you can make certain harvester configuration changes.
- [Determining full-text settings](#)
  When you conduct a lightweight harvest, you can make full-text-index setting configuration changes.
- [Determining hash settings](#)
  When you conduct a lightweight harvest, you can make certain hash-setting configuration changes.

- **Parent topic:** [Data harvesting](#)

Related concepts

- [Data harvesting](#)
- [Harvesting properties and libraries](#)

Related tasks

- [Determining volume configuration settings](#)
- [Determining harvester configuration settings](#)
- [Determining full-text settings](#)
- [Determining hash settings](#)

Determining volume configuration settings

To conduct a lightweight harvest, you can make various configuration changes.

About this task

Volume Details: When you configure data sources for a lightweight harvest, you do not need to include content tagging and full-text indexes. By clearing this option, the system indexes the files' metadata, not the entire content of those files. The system can then run and complete harvests quickly. You can obtain much information about file types, the number of files, the age of the files, file ownership, and other information.

Procedure

1. Go to Administration > Data sources > Specify volumes > Volumes.
2. On the Primary volume list page, click Add primary volumes or Edit to edit an existing volume.
3. Verify that all of the Index options check boxes are cleared (some are selected by default).
4. Click OK and then restart services.
   - Show Advanced Details: In some cases, you might want to reduce the weight of a full-text harvest. In these instances, you can adjust the processing that is involved with the various harvest configuration controls.
     Within volume configuration, the advanced settings are used to control what is harvested within the volume. By harvesting only the directory structures that you are interested in, you can exercise some control over the harvest's weight.
     - Include Directory: If you want to harvest a subtree of the volume rather than the whole volume, then you can enter the directory here. It eliminates the harvest of objects that are not relevant to your project.
     - Start Directory/End Directory: You can select a beginning and end range of directories that are harvested. Enter the start and end directories.
     - Constraints/Scope harvest by extension: You can limit the files that are harvested through connection processes, parallel data objects, or scoping harvests by extension. For example, with the Scope harvest by extension setting, you can limit the files that you harvest by using a set of extensions. If you want to harvest only Microsoft Office files, you can constrain the harvest to .DOC, .XLS, and .PPT files.
     - It must be done from the Configuration tab.

- **Parent topic:** [Lightweight harvest parameter settings](#)

Related concepts

- [Lightweight harvest parameter settings](#)

Determining harvester configuration settings

When you conduct a lightweight harvest, you can make certain harvester configuration changes.

Procedure

1. Determine Skip Content Processing settings.
   Note: This setting is relevant only for full-text harvests.
   You might have many files that are types for which you do not need the contents such as .EXE files. In these instances, you can add these file types to the list of files for which the content is not processed. There are two points to consider when to skip content processing:
   - You do not spend time harvesting unnecessary objects, which can be beneficial from a time-saving perspective.
   - Later, you have the option of viewing the content of the skipped files. It creates more work, reharvesting these skipped files.
2. Determine which Locations to ignore.

   There might be instances where large quantities of data are contained in subdirectories, and that data is not relevant to your harvest strategy. For example, you might have a directory with a tree of source code or software archive that is not used as a companywide resource. In these cases, you can eliminate these directories from the harvests by adding the directory to the Locations to ignore. These locations are not specific to a volume, but can instead be used for common directories across volumes.

3. Determine Limits.
   - Maximum data object size: This setting is only relevant for full-text harvests. In cases many large files, you might want to eliminate processing those files by setting the Maximum data object size to a smaller number. The default value is 1,000,000,000. You can still collect the metadata on the large files, so you can search for them and determine which files were missed due to the setting of this parameter.
4. Determine Binary Processing.

   If the standard processing cannot index the contents of a file, binary processing is extra processing that can be conducted. For lightweight harvests, the Run binary processing when text processing fails check box must be cleared as this setting is only relevant for full-text harvests.

- **Parent topic:** [Lightweight harvest parameter settings](#)

Related concepts

- [Lightweight harvest parameter settings](#)

Determining full-text settings

When you conduct a lightweight harvest, you can make full-text-index setting configuration changes.

Procedure

1. Determine Limits. Limit the length of words to be harvested by selecting the Limit the length of words index to __ characters option. The default value is 50, but you can reduce this number to reduce the quantity of indexed words.
2. Determine Numbers. If there are large quantities of spreadsheet files, you can control what numbers are indexed by the system.

- **Parent topic:** [Lightweight harvest parameter settings](#)

Related concepts

- [Lightweight harvest parameter settings](#)

Determining hash settings

When you conduct a lightweight harvest, you can make certain hash-setting configuration changes.

Procedure

Determine the hash settings. A file hash is a unique, calculated number that is based on the content of the file. By selecting Partial data object content, you reduce the processing to create the hash. However, the two different data objects might create the same hash. It is a small but potential risk and is only relevant for full-text harvests.

- **Parent topic:** Lightweight harvest parameter settings

Related concepts

- Lightweight harvest parameter settings

Job configuration

Jobs have different functions, and depending on the job type that is chosen, different system outcomes occur. Several jobs that are ready for use are also included.

Jobs start tasks such as harvests. They can be run at the time of creation or scheduled to run at a designated future time and at regular intervals. Jobs consist of either a single step or a series of steps. The actions available at each step depend on the type of job that is being created. Several jobs are ready for use. They are described in this table.

Table 1. Out-of-the-box job types

Job	Description
Centera deleted files synchronizer	An unscheduled, one-step job that synchronizes the deleted Centera files. This job is in the Library/Jobs folder.
Windows Share (CIFS)/NFS retention volume deleted files synchronizer	An unscheduled, one-step job that harvests the Windows Share/NFS retention volumes, looking for files that require removal because the physical file was deleted from the retention file system. This job is in the Library/Jobs folder.
Database Compactor	A scheduled job that helps to limit "bloat" (unnecessary storage usage) in the database. While this job runs, it must have exclusive, uninterrupted access to the database. Administrators can override this job by logging in and then proceed to use the system. This job is in the Library/Jobs folder.
Enterprise Vault retention volume deleted files synchronizer	An unscheduled, one-step job that harvests the Enterprise Vault retention volumes, looking for files that require removal because the physical file was deleted from the retention file system. This job is in the Library/Jobs folder.
Harvest every volume	An unscheduled, one-step job that harvests all primary and retention volumes. This job is in the Workspace/Templates/Jobs folder.
Hitachi deleted files synchronizer	An unscheduled, one-step job that harvests the Hitachi volumes, looking for files that require removal because the physical file was deleted from the file system. This job is in the Library/Jobs folder.
IBM Information Archive deleted files synchronizer	An unscheduled, one-step job that harvests the IBM Information Archive volumes, looking for files that require removal because the physical file that was deleted from the file system. This job is in the Library/Jobs folder.
System maintenance and cleanup	A multistep job in the Library/Jobs folder. The system is configured to run a system maintenance and clean up job once a day and includes these items: • Email users about reports • Email Administrators about reports • Delete old reports • Delete old harvests • Load indexes • Optimize full-text indexes
Update Age Explorers	A one-step job in the Library/Jobs folder that recalculates these items: • Owner Explorer data for Access Date • Owner Explorer data for Modified Date • Created Date (API only) values

- [Creating a job](#)
  Creating a job is the first step to work with different job types.
- [Creating a job to discover retention volumes](#)
  Jobs can be created with various types of steps, including how to discover CIFS retention volumes.

- [Editing a job](#)
  Existing, created jobs can be edited.
- [Starting a job](#)
  Jobs can be started on an as-needed basis.
- [Saving a job](#)
  Jobs can be saved on an as-needed basis for future use.
- [Running a predefined job](#)
  Predefined jobs can be run on an as-needed basis.
- [Deleting a job](#)
  Jobs can be deleted when they are no longer needed.
- [Monitoring processing](#)
  You can track the system's processing on your harvest/policy and discovery export tasks with the View cache details feature. The appliance gathers data in increments and caches the data as it gathers it. If a collection is interrupted, the appliance can resume collection at the point that it was interrupted, instead of starting over from the beginning of the task.

Creating a job

Creating a job is the first step to work with different job types.

Procedure

1. From the Folders tab > Workspace folder, select New > Job.
2. Enter a unique job name.
3. In the Save in: list, select the appropriate folder, and the job is created.
4. Click OK. If you want to view the job and add steps, click Yes.
5. On the View job page, click Add step and select a step type from the list.
6. For Run harvest jobs, on the Specify harvest and load options page, configure the following options:
   - Harvest these volumes: Select a volume from the list.
   - Harvest type: Specify the type of harvest:
     - Run a full harvest: All data objects on this volume are indexed.
     - Run an incremental harvest (default): Only files or data objects that changed since the last harvest is indexed.
   - Harvest and load scheduling: You can separate harvest and load processes to limit resource use. Select one of the options:
     - Load indexes when harvest completes
     - Load indexes with next nightly system services job: This option delays the index loading to run with the next system-services job after the harvest is completed. The system-services job is scheduled to run at midnight by default.
   - Run harvest only: Select this option if you plan to load harvested data into indexes later.
   - Load indexes only: Select this option to load previously harvested data into indexes.
7. Harvest sampling: Select this option if you want to limit the harvest to a smaller sample. This option skips every second, third, 10th, or other number data object as entered in the text box.
8. Harvest limits: Limit the harvest by time or total number of data objects. Enter the number of minutes or number of data objects.
9. Click OK.

- **Parent topic:** [Job configuration](#)

Related concepts

- [Job configuration](#)

Creating a job to discover retention volumes

Jobs can be created with various types of steps, including how to discover CIFS retention volumes.

Procedure

1. From the Folders tab > Workspace folder, select New > Job.
2. Enter a unique job name.
3. In the Save in: list, select the appropriate folder, and the job is created.
4. Click OK. If you want to view the job and add steps, click Yes.
5. On the View job page, click Add step.
6. Select Discover Retention volumes.
7. In the Discover Retention volume list, select the retention volume to be used for this job.
8. Enter the duration that the harvest runs in Run harvest for __, and then select Minutes or Hours.
9. Enter the number of data objects to be harvested in Only harvest __data objects.
10. Click OK.

- **Parent topic:** [Job configuration](#)

Related concepts

- [Job configuration](#)

Editing a job

Existing, created jobs can be edited.

Procedure

1. From the Folders tab > Workspace folder, click the job that you want to edit. The Job details page opens.
2. Click Edit job details. The Edit job details dialog box appears. You to specify the time, date, and frequency for the job to run.
   - In the Time: field, enter the time that the job must start, or click Now to populate the time field with the current time. If you did not specify all of the job steps, you might want to add some time.
   - In the Date: field, enter the date on which to run the job, or click Today to populate the date field with the current date.
3. Using the options in the Frequency field, specify how often the job must run. If you select None, the job runs once, at the time and date provided.
4. Click OK.
5. To edit the job steps:
   a. Add a step to by job by clicking Add step.
   b. Edit an existing step by clicking Edit.
   c. Remove an existing step by clicking Remove.
   d. Change the order of existing steps by clicking Move up or Move down icons.
6. Click OK.

- **Parent topic:** [Job configuration](#)

Related concepts

- [Job configuration](#)

Starting a job

Jobs can be started on an as-needed basis.

Procedure

1. To start a job, do either of the following actions:
   - From the Folders tab > Workspace folder, click the name of the job you want to start, and in the Job details page, click Start job.
   - From the Folders tab > Workspace folder, right-click the job and select Start job.
2. In the Schedule area, a started job is displayed as This job is running now.
3. Click OK to return to the Folders tab. The started job displays Running in the Status column.

- **Parent topic:** [Job configuration](#)

Related concepts

- [Job configuration](#)

Saving a job

Jobs can be saved on an as-needed basis for future use.

Procedure

1. From the Folders tab > Workspace folder, click the name of the job you would like to save. The Job details page opens.
2. Click Save as, and the Save job as dialog box appears.
3. In the Job name text box, verify that the job's name is correct. If it is saved in the same folder as another job, each job must have a unique name.
4. In the Save in list, select the folder in which you would like to save the job.
5. Click OK to close the Save job as dialog box and to return to the Folders tab.

- **Parent topic:** Job configuration

Related concepts

- Job configuration

Running a predefined job

Predefined jobs can be run on an as-needed basis.

Procedure

1. Go to Folders > Library.
2. Click Jobs folder to open the job list.
3. Click the predefined job that you want to edit.
4. To set the schedule, click Edit job details, completed the changes, and then click OK.
5. Alternately, click Start job (selected jobs only) in the bottom right area of the pane to start the job immediately.

- **Parent topic:** Job configuration

Related concepts

- Job configuration

Deleting a job

Jobs can be deleted when they are no longer needed.

Procedure

1. From the Folders tab > Workspace folder, select Filter by... Jobs.
2. Click the job name to open the job details.
3. Click Delete in the lower-left hand corner of the screen. Click OK.

- **Parent topic:** [Job configuration](#)

Related concepts

- [Job configuration](#)

Monitoring processing

You can track the system's processing on your harvest/policy and discovery export tasks with the View cache details feature. The appliance gathers data in increments and caches the data as it gathers it. If a collection is interrupted, the appliance can resume collection at the point that it was interrupted, instead of starting over from the beginning of the task.

Procedure

From Administration > Dashboard, in the Appliance status pane, click View cache details. The View cache details page appears. To see harvest/policy progress, click the Volume cache tab. Or, to see discovery export job progress, click the Discovery export cache tab.
Note: Information for a job is only available while the job is running. After a task is completed, the job disappears from the list.

Table 1. Harvest/Volume cache details: Fields, descriptions, and values

Field	Description	Value
Name	The name of the volume that is being harvested.	
Start date	The time that the job started.	
Type	The type of job that is run.	• Copy • Harvest - full • Harvest - incremental
State	The status of the process.	• Caching: The volume cache is being created/updated by a harvest or policy • Cached: Creation or update of volume cache is complete (harvest only) • Loading: Volume cache contents are being successfully loaded into the volume cluster
Full-text	It indicates whether a full-text harvest is being conducted.	Yes or No
View audit link details	Link to the harvest/policy audit page.	

Table 2. Discovery export cache details: Fields, descriptions, and values

Field	Description	Value
Name	The name of the volume that is being processed.	
Start date	The starting date/time for the process.	
Type	Type of file that is being prepared for export.	Discovery export
State	The status of the discovery export job.	• Aborted: Discovery export policy was canceled or deleted by the user. • Caching: The volume cache is being created/updated by a harvest or policy. • Cached: Creation or update of volume cache is complete (harvest only). • Loading: Volume cache contents are being successfully loaded into the volume cluster.
Full-text	Whether a full-text harvest is being conducted.	Yes or No

- [Deleting a volume cache](#)
  Volume caches can be deleted as needed.
- [Determining whether a harvest is stuck](#)
  The speed of a harvest depends on volume size and processing speed; however, harvests do occasionally become stuck and are unable to complete successfully. Use the procedures that are outlined here to troubleshoot the harvest process.

- **Parent topic:** [Job configuration](#)

Related concepts

- [Job configuration](#)

Related tasks

- [Deleting a volume cache](#)
- [Determining whether a harvest is stuck](#)

Deleting a volume cache

Volume caches can be deleted as needed.

Procedure

1. From the volume cache or discovery export cache list, select the check box of the cache you want to delete.
2. Click Delete. A confirmation dialog appears. Click OK.

- **Parent topic:** [Monitoring processing](#)

Related tasks

- [Monitoring processing](#)

Determining whether a harvest is stuck

The speed of a harvest depends on volume size and processing speed; however, harvests do occasionally become stuck and are unable to complete successfully. Use the procedures that are outlined here to troubleshoot the harvest process.

Procedure

1. Click Administration > Dashboard > Jobs in Progress to verify that your job continues to run.
2. In Jobs in Progress, note the Total data objects encountered number.
3. Wait 15 minutes, leaving the harvest to continue to run.
4. Note the new value Total data objects encountered, and then compare it to that value denoted previously.
5. Go to Question 1: Question 1: Is the Total data object encountered counter increasing?
   - Yes: If the number of encountered data objects continues to increase, then the harvest is running correctly.
   - No: If the number of encountered objects remains the same, then go to Question 2: Question 2: Is the load average up?
   a. To view load averages, on Appliance status > About appliance > View details > System services, look at the load averages in the Basic system information area.
   - Yes: If the load averages number is up, the harvest might be stuck. Call technical support to report that the harvest is stuck on files.
   - No: The job is not really running. It means that the job must be restarted. Go to Question 3: Question 3: Did the job complete on the second pass?
   - Yes: If the job completed successfully after it was restarted, then the harvest is not stuck.
   - No: The job did not complete successfully. Call technical support to report a job that does not complete.

- **Parent topic:** [Monitoring processing](#)

Related tasks

- [Monitoring processing](#)

Desktop collection

When you configure desktop settings, you are enabling or disabling encryption within IBM® StoredIQ Platform. The IBM Desktop Data Collector (desktop client or client) enables desktops as a volume type or data source, allowing them to be used just as other types of added data sources. The IBM Desktop Data Collector is provided as a standard MSI file and is installed according to the typical method (such as Microsoft Systems Management Service (SMS)) used within your organization. The IBM Desktop Data Collector can collect PSTs, compressed files, and other data objects and can remove itself when its work is completed.

After the desktop client is installed on a desktop, you connect and register it with the data server. That desktop is available as a data source within the list of primary volumes. Additionally, while the snippet support and the snippet step-up action are supported by IBM Desktop Data Collector, a desktop cannot be the target or destination of an action.

- [IBM Desktop Data Collector client installation](#)
- [IBM Desktop Data Collector installation methods](#)
  During installation, the host name and IP address of the IBM StoredIQ Platform must be supplied. If the installation is conducted manually by users, you must provide this information to them using email, a text file, or another method.
- [Configuring the IBM Desktop Data Collector collection](#)
  Desktop collection configuration includes the IBM Desktop Data Collector installer and Encrypted File System.
- [Delete policy: special notes](#)
  When you use the delete policy, certain files cannot be deleted.

IBM Desktop Data Collector client installation

The IBM® Desktop Data Collector agent works with the following operating systems:

- Windows XP 32- and 64-bit, Service Pack 2 or later
- Windows 7 32- and 64-bit
- Windows Vista 32- and 64-bit
- Windows server 2003, 2008

Installation requires administrative privileges on the desktop. Before you use the IBM Desktop Data Collector, the you might want to notify users that desktop collection is going to be conducted and make them aware of the following items:

- The desktop must be connected over the network during data collection. If the connection is interrupted, IBM Desktop Data Collector resumes its work from the point at which it stopped.
- Users might notice a slight change in performance speed, but that they can continue working normally. Desktop collection does not interfere with work processes.
- That certain actions can be taken from the tray icon: Right-click for About, Restart, Status, and Email Logs (which packages logs in to single file and starts the email client so that the user can mail them to the IBM StoredIQ Platform administrator).

All communications are outbound from the client. The appliance never pushes data or requests to the desktop. The IBM Desktop Data Collector pings the appliance about every 60 seconds, and the Last Known Contact time statistic is updated approximately every 30 minutes. Additionally, the IBM Desktop Data Collector checks for task assignments every 5 minutes.

One can download the installer application from the application in the Configuration tab. Also, the Administrator can temporarily disable the client service on all desktops that are registered to the data server from the Configuration tab.

- **Parent topic:** [Desktop collection](#)

Related concepts

- [Desktop collection](#)

Related tasks

- [Configuring the IBM Desktop Data Collector collection](#)

IBM Desktop Data Collector installation methods

During installation, the host name and IP address of the IBM® StoredIQ Platform must be supplied. If the installation is conducted manually by users, you must provide this information to them using email, a text file, or another method.

The IBM Desktop Data Collector can be installed with the following methods.

Mass distribution method (SMS)

The appliance ID is part of the distribution configuration. This method supports passing installation arguments as MSI properties.

- Required
  - SERVERACTIONNODEADDRESS: IP address or host name for the Action node. When the installation is not silent, the user is prompted for IP address or host name. The default is the value of this argument. This field must be entered accurately or manual correction is required in the desktop configuration file.
- Optional
  - SERVERACTIONNODEPORT: Port number for the Agent on the Action node. Defaults to 21000, and can be only changed when the agent connects on a different port that is then mapped to 21000.
  - NOTRAYICON: Specifies whether the agent displays theIBM Desktop Data Collector tray icon while it is running. Changing the setting to 1 forces the agent to run silently and not display a tray icon.
  - SERVERACTIONNODEADDRESS: IP address or host name for the Action node. When the installation is not silent, the user is prompted for IP address or host name. The default is the value of this argument. This field must be entered accurately or manual correction is required in the desktop config file.
  - Emailing links: Send a link within an email such as `file:\\g:\group\install\Client-install.vbs`. The link can be to any executable file format such as .BAT, .VBS, or .MSI. The .BAT/.VBS formats can be used to pass client arguments to an .MSI. The user who clicks the link must have administrative privileges.
  - NT Logon Script, in which a .BAT file or .VBS script starts msiexec. Examples are given here:
    - `/i`: Install
    - `/x {7E9E08F1-571B-4888-AC08-CEA8A076F5F9}`: Uninstall the agent. The product code must be present.
    - `/quiet`: install/uninstall runs silently. When you specify this option, SERVERACTIONNODEADDRESS must be supplied as an argument.

      ```
      Set WshShell = CreateObject("WScript.Shell")

      WshShell.Run "%windir%\System32\msiexec.exe /i G:\group\install\
      desktopclient.msi
      NOTRAYICON=0 SERVERACTIONNODEADDRESS=clust017.test.local /q"Set WshShell
      = CreateObject("WScript.Shell")

      Set WshShell = Nothing
      ```
- MSI

```
Batch file
```

```
msiexec /i G:\group\install\desktopclient.msi NOTRAYICON=1
                    SERVERACTIONNODEADDRESS=clust017.test.local /q
```

- **Parent topic:** Desktop collection

Related concepts

- [Desktop collection](#)

Related tasks

- [Configuring the IBM Desktop Data Collector collection](#)

Configuring the IBM Desktop Data Collector collection

Desktop collection configuration includes the IBM® Desktop Data Collector installer and Encrypted File System.

Before you begin

For procedural information on downloading the IBM Desktop Data Collector installer, see [Downloading the IBM Desktop Data Collector installer from the application](#).

Procedure

1. Complete the procedure that is outlined in [Using the Encrypted File-System Recovery Agent](#).
2. Restart the service and begin collection, noting the following points:
   - If the computer is not part of a domain and it is running any version of Windows earlier than 7.0, then the user name must be the user name.
   - If the computer is not part of a domain and running Windows 7.0 or later, then the user name must be the name of the PC and the domain.

- **Parent topic:** [Desktop collection](#)

Related concepts

- [Desktop collection](#)
- [IBM Desktop Data Collector client installation](#)
- [IBM Desktop Data Collector installation methods](#)
- [Delete policy: special notes](#)

Delete policy: special notes

When you use the delete policy, certain files cannot be deleted.

When you use the IBM® Desktop Data Collector to delete IBM Desktop Data Collector files from a desktop, they are removed permanently. They are not transferred to the appliance or backed up to any other location. You must carefully review the infoset of affected data objects before you take a delete action. Your organization can use custom applications or other files that you might not want to delete. In reviewing the returned list, do not allow the following files to be deleted.

- Anything under this directory: `c:\Windows`
- Anything under this directory: `Documents and Settings`, with these extensions:

  `c:\Documents and Settings\<username>\UserData\` and extension *.xml

  `c:\Documents and Settings\<username>\Cookies\` and extension *.txt

  `c:\Documents and Settings\<username>\Start Menu\Programs\` and extension *.lnk

- Executable files
  - *.dll
  - *.exe
  - *.ocx
- Drivers
  - *.sys
  - *.inf
  - *.pnf
- Installers
  - *.msi
  - *.mst
- Important data files
  - *.dat
  - *.ini
  - *.old
  - *.cat
- These file names
  - desktop.ini
  - ntuser.dat
  - index.dat
  - ntuser.pol
  - ntuser.dat.log

- **Parent topic:** [Desktop collection](#)

Related concepts

- [Desktop collection](#)

Related tasks

- [Configuring the IBM Desktop Data Collector collection](#)

Folders

This section provides both conceptual and procedural information about folders and their usage.

The Folders tab displays two types of folders: Library and Workspace.

- Library: The Library folder contains the Jobs folder.
  Note: This folder cannot be renamed, moved, or deleted.
- Workspace: The Workspace folder is a custom folder that reflects your use of the system. By default, it contains a folder entitled Templates. If you are using the system for IT purposes, you might want to create folders for each locale or function.
  Note: These folders can be renamed, moved, or deleted, and you also have the options of setting folder security.

- [Creating a folder](#)
  Folders are created from the Folders tab.
- [Deleting a folder](#)
  When you delete folders, you can delete only empty folders.
- [Moving a folder](#)
  Folders can be moved from one location to another.
- [Renaming a folder](#)
  Folders can be renamed as needed.
- [Copying items to different folders](#)
  Items can be copied from one folder to another.
- [Saving items into different folders](#)
  You can rename jobs and save them into other folders.
- [Filtering items within the folder view](#)
  Items can be filtered within the Folder view.

Creating a folder

Folders are created from the Folders tab.

Procedure

1. From the Folders tab, select New > New Folder. The Create new folder dialog appears.
2. In the Name: field, give a name that represents the folder's purpose (legal matter, local, business unit, or the like).
3. In the Description: field, type a description for the folder.
4. In the Create in: field, use the list to select a place for the folder. All custom folders must be placed in the Workspace or a Workspace subdirectory; you cannot add folders to the Library folder.
5. Click OK. If you want to open the folder, click OK in the dialog that appears.

- **Parent topic:** [Folders](#)

Related concepts

- [Folders](#)

Deleting a folder

When you delete folders, you can delete only empty folders.

Procedure

1. From within the Folders tab, Workspace folder, do either of the following actions.
    - Select the check box next to the folder you want to delete, and in the Actions list, select Delete.
    - Right-click on a folder name and select Delete.
2. In the confirmation box that appears, click OK.

- **Parent topic:** [Folders](#)

Related concepts

- [Folders](#)

Moving a folder

Folders can be moved from one location to another.

Procedure

1. From within the Folders tab, Workspace folder, do either of the following actions.
    - Select the check box next to the folder you want to move, and in the Actions list, select Move.
    - Right-click on the folder name and select Move.
2. Click OK, and in the Move items dialog, select the new location from the list.
3. Click OK.

- **Parent topic:** [Folders](#)

Related concepts

- [Folders](#)

Renaming a folder

Folders can be renamed as needed.

Procedure

1. Right-click on the folder name and select Rename.
2. In the Rename folder dialog, change the Name and the Description.
3. Click OK.

- **Parent topic:** [Folders](#)

Related concepts

- [Folders](#)

Copying items to different folders

Items can be copied from one folder to another.

Procedure

1. From within Folders > Workspace, right-click the item that you want to copy, and then select Copy.
2. In the Copy dialog, assign a new name (if appropriate) in the Name field. You cannot reuse the same name for an item within a single folder.
3. In the Description field, type a description.
4. In the Save in list, choose the location for the copied item, and then click Save.

- **Parent topic:** Folders

Related concepts

- Folders

Saving items into different folders

You can rename jobs and save them into other folders.

Procedure

1. From the item's editor pane, click Save as.
2. In the Save [item] as... pane, type a name in the [Item] name: field.
3. In the Description: field, type a description.
4. In the Save in: field, use the list to select a location for the item.
5. Click Save.

- **Parent topic:** Folders

Related concepts

- Folders

Filtering items within the folder view

Items can be filtered within the Folder view.

Procedure

1. From within Folders, click Filter by...
2. Select the component that you want to display.

- **Parent topic:** [Folders](#)

Related concepts

- [Folders](#)

Audits and logs

The following section describes the audit and log categories in the system, including descriptions of the various audit types and how to view and download details.

- [Harvest audits](#)
  Harvest audits provide a summary of the harvest, including status, date, duration, average harvest speed, and average data object size. They can be viewed in two ways: by volume name or by the date and time of the last harvest.
- [Import audits](#)
  Volume-import audits provide information about the volume import. This information includes the number of data objects that are imported, the system that is exported from, the time and date of the volume import, whether the imported volume overwrote an existing volume, and status. The volume name links to the Import details page.
- [Event logs](#)
  Event logs captures every action that is taken by the system and its users. It documents actions that succeed and fail.
- [Policy audits](#)
  Policy audits provide a detailed history of the policy. It includes type of action, date last run, start and end dates with times, average speed, total data objects, and data object counts. They can be viewed by name, volume, time, and by discovery export.

Harvest audits

Harvest audits provide a summary of the harvest, including status, date, duration, average harvest speed, and average data object size. They can be viewed in two ways: by volume name or by the date and time of the last harvest.

Data objects can be skipped during a harvest for various reasons such as the unavailable object or a selected user option that excludes the data object from the harvest. The Harvest details page lists all skipped data objects that are based on file system metadata level and content level.

All skipped harvest-audit data and other files that are not processed can be downloaded for analysis.

Table 1. Harvest audit by volume: Fields and descriptions

Harvest audit by volume field	Description
Server	The server name.
Volume	The volume name.
Harvest type	The type of harvest that is conducted: Full Harvest, ACL only, or Incremental.
Last harvested	The date and time of the last harvest.
Total system data objects	The total number of system data objects encountered.
Data objects fully processed	The number of data objects that were fully processed.
Data objects previously processed	The number of data objects that were previously processed.
Processing exceptions	The number of exceptions that are produced during processing.
Binary processed	The number of processed binary files.
Harvest duration	The length of time of the harvest's duration.
Status	The harvest's status: Complete or Incomplete.
Average harvest speed	The average harvest speed, which is given in terms of data objects that are processed per second.
Average data object size	The average size of encountered data objects.

Table 2. Harvest audit by time: Fields and descriptions

Harvest audit by time field	Description
Harvest start	The time and date at which the harvest was started.
Harvest type	The type of harvest that is conducted: Full Harvest, ACL only, or Incremental.
Total system data objects	The total number of system data objects that were found.
Data objects fully processed	The total number of system data objects that were fully processed.
Data objects previously processed	The total number of system data objects that were previously processed.
Processing exceptions	The total number of encountered processing exceptions.
Binary processed	The total number of processed binary files.
Harvest duration	The length of time of the harvest's duration.
Status	The harvest's status: Complete or Incomplete.
Average harvest speed	The average harvest speed, which is given in terms of data objects that are processed per second.

Average data object size	The average size of encountered data objects.

Table 3. Harvest overview summary options: Fields and descriptions

Harvest overview summary options field	Description
Harvest type	The type of harvest: Full Harvest, ACL only, or Incremental.
Harvest status	The harvest's status. Options are Complete or Incomplete.
Harvest date	The date and time of the harvest.
Harvest duration	This duration is the length of time of the harvest's duration.
Average harvest speed	The average harvest speed, which is given in terms of data objects that are processed per second.
Average data object size	The average size of encountered data objects.

Table 4. Harvest overview results options: Fields and descriptions

Harvest overview results options field	Description
Total system data objects	The total number of system data objects that were found.
Total contained data objects	The total number of contained data objects.
Total data objects	The total number of encountered data objects.

Table 5. Harvest overview detailed results: Fields and descriptions

Harvest overview detailed results field	Description
Skipped - previously processed	The number of skipped objects that were previously processed.
Fully processed	The number of fully processed data objects.
Skipped - cannot access data object	The number of data objects that were skipped as they might not be accessed.
Skipped - user configuration	The number of data objects that were skipped because of their user configuration.
Skipped directories	The number of data objects in skipped directories.
Content skipped - user configuration	The number of data objects where the content was skipped due to user configuration.
Content type known, partial processing complete	The number of data objects for which the content type is known and partial processing is complete.
Content type known, but error processing content	The number of data objects for which the content type is known, but an error was produced while processing content.
Content type known, but cannot extract content	The number of data objects for which the content type is known, but the content might not be extracted.
Content type unknown, not processed	The number of data objects for which the content type is unknown and is not processed.
Binary text extracted, full processing complete	The number of data objects for which the binary text is extracted and full processing is completed.
Binary text extracted, partial processing complete	The number of data objects for which the binary text is extracted and partial processing is completed.
Error processing binary content	The number of data objects for which an error was produced while binary content is processed.

Total   The total number of data objects.

- [Viewing harvest audits](#)
  Harvest audits can be viewed from the Audit tab.
- [Downloading harvest list details](#)
  Harvest list details can be downloaded in a .CSV format.

- **Parent topic:** [Audits and logs](#)

Related concepts

- [Audits and logs](#)

Related tasks

- [Viewing harvest audits](#)
- [Downloading harvest list details](#)

Viewing harvest audits

Harvest audits can be viewed from the Audit tab.

Procedure

1. Go to Audit > Harvests > View all harvests. The Harvest audit by volume page opens, which lists recent harvests and includes details about them.
2. In the Volume column, click the volume name link to see the harvest audit by time page for that particular volume. The harvest audit by time page lists all recent harvests for the chosen volume and includes details about each harvest.
3. In the Harvest start column, click the harvest start time link to see the harvest overview page for the volume. You can also access the page by clicking the Last harvested time link on the Harvest audit by volume page. The Harvest overview page provides the following options.
   - Summary: Harvest type, status, date and time, duration, average harvest speed, and average data object size.
   - Results: Total system data objects, total contained data objects, total data objects.
   - Detailed results: Skipped - previously processed; fully processed; skipped - cannot access data object; skipped - user configuration; skipped directories; content skipped - user configuration; content type known, partial processing complete; content type known, but error processing content; content type known, but cannot extract content; content type unknown, not processed; binary text extracted, full processing complete; binary text extracted, partial processing complete; error processing binary content; error-gathering ACLs; and total.
4. To view details on data objects, click the link next to the data objects under Detailed results.
   - With the exceptions of skipped - previously processed, fully processed, and the total, all other results with more than zero results have links that you can view and download results.
   - The skipped data object list includes object name, path, and reason skipped. Data objects can be skipped at the file system metadata level or at the content level. Data objects skipped at the content level are based on attributes that are associated with the data object or its contents. Skipped Data Objects Results Details provides details about skipped data objects.

   If data objects were not harvested, you might want to download the data object's harvest audit list details for further analysis.

- **Parent topic:** Harvest audits

Related concepts

- Audits and logs

Related reference

- Harvest audits

Downloading harvest list details

Harvest list details can be downloaded in a .CSV format.

Procedure

1. From the Harvest details page, click the active link next to the data objects under Detailed results. A page named for the detailed result chosen (such as Skipped - user configuration or binary text extracted, full processing complete) appears.
2. Click the Download list in CSV format link on the upper left side of the page. A dialog informs you that the results are being prepared for download.
3. Click OK. A new dialog appears, prompting you to save the open or save the .CSV file. Information in the downloaded CSV file includes:
   - Object name
   - System path
   - Container path
   - Message explaining why data object was skipped
   - Server name
   - Volume name

- **Parent topic:** Harvest audits

Related concepts

- Audits and logs

Related reference

- Harvest audits

Import audits

Volume-import audits provide information about the volume import. This information includes the number of data objects that are imported, the system that is exported from, the time and date of the volume import, whether the imported volume overwrote an existing volume, and status. The volume name links to the Import details page.

Table 1. Imports by volumes details: Fields and descriptions

Imports by volumes details field	Description
Volume	The name of the imported volume.
Exported from	The source server of the imported volume.
Import date	The date and time on which the import occurred.
Total data objects imported	The total number of imported data objects.
Overwrite existing	If the import overwrote an existing volume, the status is Yes. If the import did not overwrite an existing volume, the status is No.
Status	The status of the import: Complete or Incomplete.

- [Viewing volume import audit details](#)
  Audit details of volume imports can be viewed.

- **Parent topic:** [Audits and logs](#)

Related concepts

- [Audits and logs](#)

Related tasks

- [Viewing volume import audit details](#)

Viewing volume import audit details

Audit details of volume imports can be viewed.

Procedure

1. Go to Audit > Imports, and click View all imports. The Imports by volume page opens, which lists volume imports and import information.
2. Click a volume name link in the Volume column to view the audit details for that particular import.

- **Parent topic:** Import audits

Related concepts

- Audits and logs

Related reference

- Import audits

Event logs

Event logs captures every action that is taken by the system and its users. It documents actions that succeed and fail.

These actions include creating draft and published queries and tags, running policies, publishing queries, deleting objects, configuring settings, and any other action that is taken through the interface. A detailed of list of log entries is provided in the event log messages.

You can view event logs for the current day or review saved logs from previous days, and up to 30 days worth of logs can be viewed through the interface. If you select and clear a day of logs, those logs are removed from the system.

- [Viewing event logs](#)
  Event logs can be viewed from the Dashboard or from the Audit tab.
- [Subscribing to an event](#)
  You can subscribe to and be notified of daily event logs.
- [Clearing the current event log](#)
  An event log can be cleared from the Dashboard.
- [Downloading an event log](#)
  Event logs can be downloaded and saved.

- **Parent topic:** [Audits and logs](#)

Related concepts

- [Audits and logs](#)

Related tasks

- [Viewing event logs](#)
- [Subscribing to an event](#)
- [Clearing the current event log](#)
- [Downloading an event log](#)

Viewing event logs

Event logs can be viewed from the Dashboard or from the Audit tab.

Procedure

1. Conduct either of the following actions:
    a. Click Administration > Dashboard, and then locate the Event log section on the dashboard. The current day's log displays there by default.
    b. Click the Audit tab and locate the Event logs section.
2. To view a previous day's log on the dashboard, use the View all event logs list to select the day for which you want to view an event log.
3. Select a different day from the view event log from the list. This menu displays the event log dates for the past 30 days. Each log is listed by date in YYYY-MM-DD format.

- **Parent topic:** Event logs

Related concepts

- Audits and logs
- Event logs

Subscribing to an event

You can subscribe to and be notified of daily event logs.

Procedure

1. Go to Audit > Event logs.
2. Click View all event logs, and the Event log for today page opens.
3. To the right of the event log to which you want to subscribe, click Subscribe. The Edit notification page appears.
4. In Destination, select the method by which you want to be notified of this event log. If you select Email address, be certain to use commas to separate multiple email addresses.
5. Click OK.
   Note: You can also subscribe to an event on the Dashboard. In the Event log area, click Subscribe to the right of the event.

- **Parent topic:** [Event logs](#)

Related concepts

- [Audits and logs](#)
- [Event logs](#)

Clearing the current event log

An event log can be cleared from the Dashboard.

Procedure

On the Administration > Dashboard, click Clear for the current view.

- **Parent topic:** Event logs

Related concepts

- Audits and logs
- Event logs

Downloading an event log

Event logs can be downloaded and saved.

Procedure

1. When you view an event log, click the Download link for saving the data to a text file.
2. Select to save the file from the prompt. Enter a name and select a location to save the file.

- **Parent topic:** [Event logs](#)

Related concepts

- [Audits and logs](#)
- [Event logs](#)

Policy audits

Policy audits provide a detailed history of the policy. It includes type of action, date last run, start and end dates with times, average speed, total data objects, and data object counts. They can be viewed by name, volume, time, and by discovery export.

Policy audit by name

Table 1. Policy audit by name: Fields and descriptions

Policy audit by name field	Description
Policy name	The policy name.
Policy status	The policy's status.
Number of times executed	The number of times that the policy was run.
Most recent date executed	The date on which the policy was last run.

Policy audit by volume

Table 2. Policy audit by volume: Fields and descriptions

Policy audit by volume field	Description
Volume	The name of the volume on which the policy was run.
Most recent date a policy was executed	The most recent date on which the policy was last run.
Number of policies executed	The number of policies that were run.

Policy audit by time

Table 3. Policy audit by time: Fields and descriptions

Policy audit by time field	Description
Policy name	The policy name.
Policy status	The status of the policy: Complete or Incomplete.
Start	The time at which the policy's execution was started.
End	The time at which the policy's execution was complete.
Success count	The number of processed messages that are classified as a success.
Failure count	The number of processed messages that are classified as a failure.
Warning count	The number of processed messages that are classified as a warning.
Other count	The number of processed messages that are classified as other.
Total data objects	The total number of data objects.
Action type	The type of policy that took place.
Ag. actions/second	The average number of actions per second.

Policy audit by discovery export

Table 4. Policy audit by discovery exports: Fields and descriptions

Policy audit by discovery export	Description

field	
Discovery export name	The name of the discovery export.
Number runs	The number of times the policy ran.
Most recent export status	The status of the most recent discovery export.
Most recent load file status	The status of the most recent load file.
Most recent date executed	The date of the most recent policy execution.

Discovery export runs by discovery export

Table 5. Discovery export runs by discovery export: Fields and descriptions

Discovery export runs by discovery export field	Description
Discovery export run	The name of the discovery export run.
Number of executions	The number of times the run was started.
Success count	The number of processed messages that are classified as a success.
Failure count	The number of processed messages that are classified as a failure.
Warning count	The number of processed messages that are classified as a warning.
Other count	The number of processed messages that are classified as other.
Total data objects	The total number of data objects.
Export status	The status of the export: Complete or Incomplete.
Load file status	The status of the load file: Complete or Incomplete.

Note: A warning in a policy audit trail is a success with the following conditions:

- If you copy an Exchange item such as `re:`, the `re` is copied, not `the:`. It generates a warning.
- The copied file is renamed.
- The file system to which you are copying does not accept characters in the file name.

- [Viewing policy audit details](#)
  Policy audits can be viewed by name, volume, time, or discovery export.
- [Search audit feature](#)
  With the search audit feature, you can search audit trails by entering either by Policy Details, Execution Details, or Data Object Details.
- [Saving results from an audit](#)
  You can save the results of policy executions into PDF and XML files. The information can be saved as PDF and XML files. The exporting of information appears as a running job on the dashboard until completed.
- [Policy audit messages](#)
  A policy audit shows the number of data objects that were processed during the policy execution. Processed data objects are divided into these categories: Success, Warnings, Failures, and Other (discovery export policies only).

- **Parent topic:** [Audits and logs](#)

Related concepts

- [Audits and logs](#)
- [Search audit feature](#)

Related tasks

Viewing policy audit details

Policy audits can be viewed by name, volume, time, or discovery export.

Procedure

1. Go to Audit > Policies, and then click Name. The Policy audit by name page provides policy name and status, the number of times it was run, and the time and date of the most recent execution.
2. Click a policy name to open the Policy executions by time page.
3. Click a policy name to open the Policy execution results page.
   Note: To view the list of data objects, click the [#] data objects link. To create a report, click Create XML or Create PDF.
   a. Click Volume to open the policy audit by volume page.
   b. Click a volume link to go to the Policy audit by time page.
   c. Click Time to see Audit by time page for the policy.
   d. On the Policy audit by time page, click the policy name to open the Policy execution results page. To view the list of data objects, click the [#] data objects link. To create a report, click Create XML or Create PDF.
   e. Click Discovery export.
   f. On the Policy audit by discovery export page, click the discovery export name to open the Discovery export runs by production page. The page details further information according to the incremental runs of the policy.
   g. Click a policy name to open the Policy executions by time page.
   h. Click a policy name to open the Policy execution results page. To view the list of data objects, click the [#] data objects link. To create a report, click Create XML or Create PDF.

As you review audit results through the pages, you can continue clicking through to review various levels of information, from the volume and policy execution level down to the data objects. To view more policy execution details, click the policy name in the execution summary page, which can be accessed by any of the policy views. As you continue browsing, IBM® StoredIQ Platform provides more detailed information such as:
   - Source and destination settings
   - Policy options: Details of the policy action. This section reflects the options that are selected when you create the policy. Most attributes that appear depend upon the type of policy run and the options available in the policy editor.
   - Query (either IBM StoredIQ Platform or user-defined)
   - View metadata link: The view metadata page describes security details for source and destination locations of the policy action.

- [Viewing a policy audit by name](#)
  Policy audits can be viewed by their name.
- [Viewing a policy audit by volume](#)
  Policy audits can be viewed by volume.
- [Viewing a policy audit by time](#)
  Policy audits can be viewed by time.
- [Viewing a policy audit by discovery export](#)
  Policy audits can be viewed by discovery export.

- **Parent topic:** [Policy audits](#)

Related concepts

- [Audits and logs](#)

Related reference

- [Policy audits](#)

Viewing a policy audit by name

Policy audits can be viewed by their name.

Procedure

1. Go to Audit > Policies.
2. Click Name. The Policy audit by name page provides policy name and status, the number of times it is run, and the time and date of the most recent execution.
3. Click a policy name to open the Policy execution by time page.

- **Parent topic:** Viewing policy audit details

Related concepts

- Audits and logs

Related reference

- Policy audits

Viewing a policy audit by volume

Policy audits can be viewed by volume.

Procedure

1. Go to Audit > Policies.
2. Click Volume. The Policy audit by page provides policy name, its most recent time, date of execution, and the number of policies that were run.
3. Click a policy name to open the Policy executions by time page.
4. Click the policy name to open the Policy executions by results page.

- **Parent topic:** [Viewing policy audit details](#)

Related concepts

- [Audits and logs](#)

Related reference

- [Policy audits](#)

Viewing a policy audit by time

Policy audits can be viewed by time.

Procedure

1. Go to Audit > Policies.
2. Click Time. The policy audit by time provides information about the policy.
3. On the Policy audit by time page, click the policy name to open the Policy execution results page.

- **Parent topic:** Viewing policy audit details

Related concepts

- Audits and logs

Related reference

- Policy audits

Viewing a policy audit by discovery export

Policy audits can be viewed by discovery export.

Procedure

1. Click Discovery export.
2. On the Policy audit by discovery export page, click the discovery export name to open the Discovery export runs by production page. The page details further information according to the incremental runs of the policy.
3. Click a policy name to open the Policy executions by time page.
4. Click a policy name to open the Policy execution results page.

- **Parent topic:** Viewing policy audit details

Related concepts

- Audits and logs

Related reference

- Policy audits

Search audit feature

With the search audit feature, you can search audit trails by entering either by Policy Details, Execution Details, or Data Object Details.

Policy details

Policy audits can be searched with any of these details.

Table 1. Policy audit details: Fields and descriptions

Policy detail field	Description
Audit search by policy details	In this area, select search criteria, define their values, and then add them to the list to search across all audits.
Specify search criteria	In this area, specify the Policy name, the Policy state, and the Action type.
Audit search criteria	In the Find audits that match list, select either Any of the following or All of the following.

Execution details

Policy audits can be searched with any of these execution details.

Table 2. Policy audit execution details: Fields and descriptions

Execution detail field	Description
Audit search by execution detail	In this area, select search criteria, define their values, and then add them to the list to search across all audits.
Specify search criteria	In this area, specify the Action type, Action status, Action start date, Action end date, Total count, Success count, Failure count, Warning count, Source volume, Destination volume, or Query name.
Audit search criteria	In the Find audits that match list, select either Any of the following or All of the following.

Data object details

Policy audits can be searched with any of these data-object details.

Table 3. Policy audit data object details: Fields and descriptions

Data object details field	Description
Audit search by data object details	In this area, select search criteria, define their values, and then add them to the list to search across all audits.
Specify search criteria	In this area, specify the Source volume, Destination volume, Source object name, Destination object name, Source system path, Destination system path, or Action result.
Audit search criteria	In the Find audits that match list, select either Any of the following or All of the following.

- **Parent topic:** Policy audits

Related reference

- Policy audits

Saving results from an audit

You can save the results of policy executions into PDF and XML files. The information can be saved as PDF and XML files. The exporting of information appears as a running job on the dashboard until completed.

Procedure

1. Go to Audit > Policies.
2. In the Browse by options, click Time.
3. Click the policy name.
4. In the Results pane, click Data objects to see items that were responsive to the policy. To download the material in .CSV, click CSV.
5. On the Policy execution results page, select Create PDF to generate a PDF or Create XML to generate an XML file of the results.
6. Access the report through the inbox on the navigation page.

- **Parent topic:** [Policy audits](#)

Related concepts

- [Audits and logs](#)

Related reference

- [Policy audits](#)
- [Policy audit messages](#)

Policy audit messages

A policy audit shows the number of data objects that were processed during the policy execution. Processed data objects are divided into these categories: Success, Warnings, Failures, and Other (discovery export policies only).

Table 1. Types of and reasons for policy audit messages

Audit message type	Reason
Success	- Data object is a duplicate of [object name] - Data object skipped but is loaded in load file. It applies to intermediate and files archives produced during a discovery export policy. - Data object is a duplicate produced in a previous run (discovery export only).
Warning	- Set directory attributes - Reset time stamps - Set attributes - Set time stamps - Set security descriptor (Windows Share) - Set access modes (Windows Share) - Set owner information - Set group information (NFS) - Set security permissions - Create a link after migration (Windows Share, NFS) - Find template to create a shortcut (Windows Share) - Extract text for the object (Discovery export policy)
Failure	- Failed to create target directory structure - Source does not exist - Failed to find a new name for the incoming object - Target is a directory - File copy failed - Cannot create target - Error copying data to target - Cannot copy due to network errors - Cannot delete source after move - Target disk is full - Source equals target on a copy or move - Insufficient permissions in general to conduct an action - All modify actions failed - File timed out waiting in the pipeline - File under retention; cannot be deleted (retention server) - Data object is a constituent of a container that already encountered failure (discovery export policy)
Other	Data objects are categorized in the other category during a discovery export policy when: - A data object is a member that makes its container responsive. - A data object is a non-responsive member of a container.

- **Parent topic:** Policy audits

Related concepts

- [Audits and logs](#)

Related tasks

- [Saving results from an audit](#)

Deploying customized web services

This procedure highlights the basic steps that are required to deploy SharePoint custom web services.

About this task

It applies only to SharePoint 2013 web services.

Procedure

1. Obtain the installation package. This package is created for you by IBM®.
2. Uninstall an existing instance of the web service.
3. Install the installation package.
4. Verify that the web service is hosted.
5. Configure admin knobs.

- [Uninstalling an existing instance of the web service](#)
  To install an upgrade to a web service, any previous, existing instance must first be uninstalled.
- [Verifying the location of the hosted web service](#)
  This procedure details how to verify the location of the hosted web service and access it in a browser.
- [Configuration of administration knobs](#)
  IBM StoredIQ has various administration knobs that can be customized to deploy web services. This section lists and describes the various knobs and provides usage notes and examples.

Uninstalling an existing instance of the web service

To install an upgrade to a web service, any previous, existing instance must first be uninstalled.

About this task

All steps within this procedure are completed on the SharePoint server.

Procedure

1. Within IIS, click Sites and find the website that was created by the previous installation. Right-click that website and remove it.
2. Within IIS, click Application Pools, find the web application that was created (it has the same name as the website). Right-click that web application and remove it.
3. In Windows Explorer, go to the folder where the web service was deployed and delete all content within this folder.
4. Reset IIS with the `iisreset` command.

- **Parent topic:** [Deploying customized web services](#)

Related tasks

- [Deploying customized web services](#)

Verifying the location of the hosted web service

This procedure details how to verify the location of the hosted web service and access it in a browser.

Procedure

1. Within IIS Manager, click Sites, and verify that you see the new site name that is listed along with the name that is entered into the installer.
2. Expand Sites and verify that you can see the new site name that is listed along with the name that is entered into the installer.
3. Select the newly created site and switch to the Content View, which is on the right pane.
4. An SVC file corresponds to the installed web service that is installed. Right-click the SVC file and click Browse. The web service URL is started in a browser, the address bar of which contains the HTTP location of the web service such as http://localhost:9000/UpdateItemService.svc.

- **Parent topic:** [Deploying customized web services](#)

Related tasks

- [Deploying customized web services](#)

Configuration of administration knobs

IBM® StoredIQ has various administration knobs that can be customized to deploy web services. This section lists and describes the various knobs and provides usage notes and examples.

Table 1. IBM StoredIQ administration knobs

Admin knob	Default value	Notes	Example
sharepoint_ custom_ webservice _location		Location of custom web-service used to facilitate migration of timestamps and owner information.	`INSERT INTO adminknobs (name, value, description, valuetype, use) SELECT 'sharepoint_custom_webservice_loca-tion', '', 'Location for custom web-service hosted on SharePoint, eg: 9100:/UpdateItemService.svc', 'str', 2 WHERE (SELECT name FROM adminknobs WHERE name = 'sharepoint_custom_webservice_loca-tion') IS NULL;`
sharepoint_ harvest_docs_ only	The default value is zero (0).	When set to 1, IBM StoredIQ will harvest only document libraries from SharePoint.	`INSERT INTO adminknobs (name, value, description, valuetype, use)SELECT 'sharepoint_harvest_docs_only', 0, 'Harvest only Sharepoint document type list objects', 'int', 1 WHERE (SELECT name FROM adminknobs WHERE name = 'sharepoint_harvest_docs_only') IS NULL;`
global_copy_ ignore_target _vc	The default value is zero (0).	When set to 1, IBM StoredIQ will not automatically harvest the destination volume when performing copies.	`INSERT INTO adminknobs (name, value, description, valuetype, use)SELECT 'global_copy_ignore_target_vc', 0, 'Global override to turn off auto-harvests on a copy', 'int', 1 WHERE (SELECT name FROM adminknobs WHERE name = 'global_copy_ignore_target_vc') IS NULL;`

- **Parent topic:** Deploying customized web services

Related tasks

- Deploying customized web services

Configuring administration knobs

Administration knobs must be configured to point at the installed web service.

About this task

Note: It is necessary to restart services after changes to admin knobs.
The value of the admin knob requires that you know the location of the web service. The format of the service location that is entered as the value for the admin knob is `port:servicelocation`.

Procedure

1. Verify the location of the web service.
2. Enter the location of the web service as part of the value for the admin knob in the format `port:servicelocation`. For example, if the web service URL is `http://localhost:9000/UpdateItemService.svc`, the admin knob value must be set to `9000:/UpdateItemService.svc`. It is done with the following SQL command: `update adminknobs`

```
set value='9000:/UpdateItemService.svc' where name='sharepoint_custom_webservice_location';
```

Supported file types

The following section provides a comprehensive list of the file types that can be harvested and processed by IBM® StoredIQ Platform, organized by name and by category. You can also view SharePoint attributes.

- [Supported file types by name](#)
  All file types by name that is supported by IBM StoredIQ are listed, including category, format, extension, category, and version.
- [Supported file types by category](#)
  All file types by category that is supported by IBM StoredIQ are listed, including category, format, extension, and version.
- [SharePoint attributes](#)
  The following section describes the various SharePoint data object types and their properties that are currently supported by IBM StoredIQ Platform.

Supported file types by name

All file types by name that is supported by IBM® StoredIQ are listed, including category, format, extension, category, and version.

Table 1. Supported file types by name

Format	Extension	Category	Version
Adobe Acrobat	PDF	graphic	2.1 3.0-7.0 - Japanese
Adobe FrameMaker Graphics	FMV	graphic	vector/raster through 5.0
Adobe FrameMaker Interchange Format	MIF	word processing	3.0-6.0
Adobe Illustrator		graphic	- Through 7.0 9.0
Adobe Photoshop	PSD	graphic	4.0
Ami Draw	SDW	graphic	all
ANSI	TXT	text and markup	7- and 8-bit
ASCII	TXT	text and markup	7- and 8-bit
AutoCAD	DWG	CAD	2.5-2.6 9.0-14.0 2002 2004 2005
AutoShade Rendering	RND	graphic	2.0
Binary Group 3 Fax		graphic	all
Bitmap	BMP, RLE, ICO, CUR, DIB, WARP	graphic	all
CALS Raster	GP4	graphic	Type I, II
Comma-Separated Values	CSV	spreadsheet	
Computer Graphics Metafile	CGM	graphic	- ANSI - CALS - NIST 3.0
Corel Clipart	CMX	graphic	5-6
Corel Draw	CDR	graphic	3.x-8.x
Corel Draw (CDR with Tiff header)		graphic	2.x-9.x
Corel Presentations	SHW	presentation	- Through 12.0 - X3
Corel WordPerfect		word	Through 12.0

Corel WordPerfect Windows	WPD	processing	• X3
DataEase		Database	4.X
dBase Database		Database	Through 5.0
dBXL		Database	1.3
DEC WPS PLUS	DX	word processing	Through 4.0
DEC WPS PLUS	WPL	word processing	Through 4.1
DisplayWrite (2 and 3)	IP	word processing	all
DisplayWrite (4 and 5)		word processing	Through 2.0
DOS command executable	COM	system	
Dynamic link library files	DLL	system	
EBCDIC		text and markup	all
ENABLE		word processing	• 3.0 • 4.0 • 4.5
ENABLE		Database	• 3.0 • 4.0 • 4.5
ENABLE Spreadsheet	SSF	spreadsheet	• 3.0 • 4.0 • 4.5
Encapsulated Post-Script (raster)	EPS	graphic	TIFF header
Executable files	EXE	system	
First Choice		Database	Through 3.0
First Choice		word processing	Through 3.0
First Choice		spreadsheet	Through 3.0
FoxBase		Database	2.1
Framework		Database	3.0
Framework		word	3.0

Name	Extension	Type	Version
Framework		word processing	3.0
Framework		spreadsheet	3.0
GEM Bit Image	IMG	graphic	all
Graphics Interchange Format	GIF	graphic	all
Graphics Environment Manager	GEM VDI	graphic	Bitmap and vector
Gzip	GZ	archive	all
Haansoft Hangul	HWP	word processing	• 1997 • 2002
Harvard Graphics (DOS)		graphic	• 2.x • 3.x
Harvard Graphics (Windows)		graphic	all
Hewlett-Packard Graphics Language	HPGL	graphic	2
HTML	HTM	text and markup	Through 3.0
IBM FFT		text and markup	all
IBM Graphics Data Format	GDF	graphic	1.0
IBM Picture Interchange Format	PIF	graphic	1.0
IBM Revisable Form Text		text and markup	all
IBM Writing Assistant		word processing	1.01
Initial Graphics Exchange Spec	IGES	graphic	5.1
Java class files	CLASS	system	
JPEG (not in TIFF format)	JFIF	graphic	all
JPEG	JPEG	graphic	all
JustSystems Ichitaro	JTD	word processing	• 5.0 • 6.0 • 8.0-13.0 • 2004
JustSystems Write		word processing	Through 3.0
Kodak Flash Pix	FPX	graphic	all
Kodak Photo CD	PCD	graphic	1.0
Legacy		word processing	Through 1.1

Legato Email Extender	EMX	Email	
Lotus 1-2-3	WK4	spreadsheet	Through 5.0
Lotus 1-2-3 (OS/2)		spreadsheet	Through 2.0
Lotus 1-2-3 Charts	123	spreadsheet	Through 5.0
Lotus 1-2-3 for SmartSuite		spreadsheet	1997-Millennium 9.6
Lotus AMI Pro	SAM	word processing	Through 3.1
Lotus Freelance Graphics	PRZ	presentation	Through Millennium
Lotus Freelance Graphics (OS/2)	PRE	presentation	Through 2.0
Lotus Manuscript		word processing	2.0
Lotus Notes	NSF	Email	
Lotus Pic	PIC	graphic	all
Lotus Snapshot		graphic	all
Lotus Symphony		spreadsheet	• 1.0 • 1.1 • 2.0
Lotus Word Pro	LWP	word processing	1996-9.6
LZA Self Extracting Compress		archive	all
LZH Compress		archive	all
Macintosh PICT1/2	PICT1/PICT1	graphic	Bitmap only
MacPaint	PNTG	graphic	NA
MacWrite II		word processing	1.1
Macromedia Flash	SWF	presentation	text only
MASS-11		word processing	Through 8.0
Micrografx Designer	DRW	graphic	Through 3.1
Micrografx Designer	DSF	graphic	Win95, 6.0
Micrografx Draw	DRW	graphic	Through 4.0
MPEG-1 Audio layer 3	MP3	multimedia	• ID3 metadata only • These files can be harvested, but there is no data in them that can be used in tags.
MS Access	MDB	Database	Through 2.0
MS Binder		archive	7.0-1997

MS Excel	XLS	spreadsheet	2.2-2007
MS Excel Charts		spreadsheet	2.x-7.0
MS Excel (Macintosh)	XLS	spreadsheet	3.0-4.0 1998 2001 2004
MS Excel XML	XLSX	spreadsheet	
MS MultiPlan		spreadsheet	4.0
MS Outlook Express	EML	Email	1997-2003
MS Outlook Form Template	OFT	Email	1997-2003
MS Outlook Message	MSG	Email	all
MS Outlook Offline Folder	OST	Email	1997-2003
MS Outlook Personal Folder	PST	Email	1997-2007
MS PowerPoint (Macintosh)	PPT	presentation	4.0-2004
MS PowerPoint (Windows)	PPT	presentation	3.0-2007
MS PowerPoint XML	PPTX	presentation	
MS Project	MPP	Database	1998-2003
MS Windows XML	DOCX	word processing	
MS Word (Macintosh)	DOC	word processing	3.0-4.0 1998 2001
MS Word (PC)	DOC	word processing	Through 6.0
MS Word (Windows)	DOC	word processing	Through 2007
MS WordPad		word processing	all
MS Works	S30/S40	spreadsheet	Through 2.0
MS Works	WPS	word processing	Through 4.0
MS Works (Macintosh)		word processing	Through 2.0

MS Works Database (Macintosh)		Database	Through 2.0
MS Works Database (PC)		Database	Through 2.0
MS Works Database (Windows)		Database	Through 4.0
MS Write		word processing	Through 3.0
Mosaic Twin		spreadsheet	2.5
MultiMate 4.0		word processing	Through 4.0
Navy DIF		word processing	all
Nota Bene		word processing	3.0
Novell Perfect Works		word processing	2.0
Novell Perfect Works		spreadsheet	2.0
Novell Perfect Works (Draw)		graphic	2.0
Novell WordPerfect		word processing	Through 6.1
Novell WordPerfect (Macintosh)		word processing	1.02-3.0
Office Writer		word processing	4.0-6.0
OpenOffice Calc	SXC/ODS	spreadsheet	• 1.1 • 2.0
OpenOffice Draw		graphic	• 1.1 • 2.0
OpenOffice Impress	SXI/SXP/ODP	presentation	• 1.1 • 2.0
OpenOffice Writer	SXW/ODT	word processing	• 1.1 • 2.0
OS/2 PMMetafile Graphics	MET	graphic	3.0
Paint Shop Pro 6	PSP	graphic	5.0-6.0
Paradox Database (PC)		Database	Through 4.0

Paradox (Windows)		Database	Through 1.0
PC-File Letter		word processing	Through 5.0
PC-File+Letter		word processing	Through 3.0
PC PaintBrush	PCX, DCX	graphic	all
PFS: Professional Plan		spreadsheet	1.0
PFS: Write		word processing	A, B, C
Portable Bitmap Utilities	PBM	graphic	all
Portable Greymap	PGM	graphic	NA
Portable Network Graphics	PNG	graphic	1.0
Portable Pixmap Utilities	PPM	graphic	NA
PostScript File	PS	graphic	level II
Professional Write		word processing	Through 2.1
Professional Write Plus		word processing	1.0
Progressive JPEG		graphic	NA
Q &A (database)		Database	Through 2.0
Q & A (DOS)		word processing	2.0
Q & A (Windows)		word processing	2.0
Q & A Write		word processing	3.0
Quattro Pro (DOS)		spreadsheet	Through 5.0
Quattro Pro (Windows)		spreadsheet	• Through 12.0 • X3
R:BASE 5000		Database	Through 3.1
R:BASE (Personal)		Database	1.0
R:BASE System V		Database	1.0
RAR	RAR	archive	
Reflex Database		Database	2.0
Rich Text Format	RTF	text and	all

SAMNA Word IV		markup word processing	
Smart Ware II		Database	1.02
Smart Ware II		word processing	1.02
Smart Ware II		spreadsheet	1.02
Sprint		word processing	1.0
StarOffice Calc	SXC/ODS	spreadsheet	• 5.2 • 6.x • 7.x • 8.0
StarOffice Draw		graphic	• 5.2 • 6.x • 7.x • 8.0
StarOffice Impress	SXI/SXP/ODP	presentation	• 5.2 • 6.x • 7.x • 8.0
StarOffice Writer	SXW/ODT	word processing	• 5.2 • 6.x • 7.x • 8.0
Sun Raster Image	RS	graphic	NA
Supercalc Spreadsheet		spreadsheet	4.0
Text Mail (MIME)	various	Email	
Total Word		word processing	1.2
Truevision Image	TIFF	graphic	Through 6
Truevision Targa	TGA	graphic	2
Unicode Text	TXT	text and markup	all
Unix TAR (tape archive)	TAR	archive	NA
Unix Compressed	Z	archive	NA
UUEncoding	UUE	archive	NA
vCard		word processing	2.1
Visio (preview)		graphic	4

Visio 2003		graphic	• 5 • 2000 • 2002
Volkswriter		word processing	Through 1.0 through 1.0
VP Planner 3D		spreadsheet	1.0
WANG PC		word processing	Through 2.6
WBMP		graphic	NA
Windows Enhanced Metafile	EMF	graphic	NA
Windows Metafile	WMF	graphic	NA
Winzip	ZIP	archive	
WML		text and markup	5.2
WordMARC word processor		word processing	Through Composer
WordPerfect Graphics	WPG, WPG2	graphic	Through 2.0, 7. and 10
WordStar		word processing	Through 7.0
WordStar 2000		word processing	Through 3.0
X Bitmap	XBM	graphic	x10
X Dump	XWD	graphic	x10
X Pixmap	XPM	graphic	x10
XML (generic)	XML	text and markup	
XyWrite	XY4	word processing	Through III Plus
Yahoo! IM Archive		archive	
ZIP	ZIP	archive	PKWARE-2.04g

- **Parent topic:** <u>Supported file types</u>

Related reference

- <u>Supported file types</u>

Supported file types by category

All file types by category that is supported by IBM® StoredIQ are listed, including category, format, extension, and version.

Table 1. Supported archive file types by category

Format	Extension	Version
Gzip	GZ	all
LZA Self-Extracting Comparess		all
LZH Compress		all
MS Binder		7.0-1997
RAR	RAR	
Unix TAR (tape archive)	TAR	NA
Unix Compressed	Z	NA
UUEncoding	UUE	NA
Winzip	Zip	
Yahoo! IM Archive		NA
ZIP	ZIP	PKWARE-2.04g

Table 2. Supported CAD file types by category

Format	Extension	Version
AutoCAD	DWG	2.5-2.6 9.0-14.0 2002 2004 2005

Table 3. Supported database file types by category

Format	Extension	Version
DataEase		4.x
dBase DataBase		Through 5.0
dBXL		1.3

- ENABLE
  - 3.0
  - 4.0
  - 4.5

- First Choice
  - Through 3.0

- FoxBase
  - 2.1

- Framework
  - 3.0

- MS Access
  - MDB
  - Through 2.0

- MS Project
  - MPP
  - Through 2.0

- MS Works Database (Macintosh)
  - 2.0

- MS Works Database (PC)
  - Through 2.0

- MS Works Database (Windows)
  - Through 4.0

- Paradox Database (PC)
  - Through 4.0

- Paradox Database (Windows)
  - Through 1.0

- Q&A (database)
  - Through 2.0

- R:BASE 5000
  - Through 3.1

- R:BASE (personal)
  - 1.0

- R:BASE System V
  - 1.0

- Reflex Database
  - 2.0

- Smart Ware II
  - 1.02

Table 4. Supported email file types by category

Format	Extension	Version
Legato Email Extender	EMX	
Lotus Notes	NSF	
MS Outlook Express	EML	1997-2003

- MS Outlook Form Template
- OFT
- 1997-2003

- MS Outlook Message
- MSG
- all

- MS Outlook Offline Folder
- OST
- 1997-2003

- MS Outlook Personal Folder
- PST
- 1997-2007

- Text Mail (MIME)
- various

Table 5. Supported graphic file types by category

Format	Extension	Version
• Adobe Acrobat	• PDF	• 2.1 • 3.0-7.0 • Japanese
• Adobe Framemaker Graphics	• FMV	• vector/raster-5.0
• Adobe Illustrator		• Through 7.0 • 9.0
• Adobe Photoshop	• PSD	• 4.0
• Ami Draw	• SDW	• all
• AutoShade Rendering	• RND	• 2.0
• Binary Group 3 Fax		• all
• Bitmap	• BMP, RLE, ICO, CUR, DIB, WARP	• all
• CALS Raster	• GP4	• Type I, II
• Computer Graphics Metafile	• CGM	• ANSI • CALS • NIST 3.0
• Corel Cliart	• CMX	• 5-6
• Corel Draw	• CDR	• 3.x-8.x
• Corel Draw (CDR with TIFF		• 2.x-9.x

header

• Encapsulated Post Script (raster)	• EPS	• TIFF header
• GEM Bit Image	• IMG	• all
• Graphics Interchange Format	• GIF	• all
• Graphics Environment Manager	• GEM VDI	• Bitmap • vector
• Harvard Graphics (DOS)		• 2.x • 3.x
• Harvard Graphics (Windows)		• all
• Hewlett-Packard Graphics Language	• HPGL	• 2
• IBM Graphics Data Format	• GDF	• 1.0
• IBM Picture Interchange Format	• PIF	• 1.0
• JPEG (not in TIFF format)	• JFIF	• all
• JPEG	• JPEG	• all
• Kodak Flash PIX	• FPX	• all
• Kodak Photo CD	• PCD	• 1.0
• Lotus Pic	• PIC	• all
• Lotus Snapshot		• all
• macintosh PICT1/2	• PICT1/PICT2	• Bitmap only
• MacPaint	• PNTG	• NA
• Micrografx Designer	• DRW	• Through 3.1
• Micrografx Draw	• DRW	• Through 4.0

• Novell Perfect Works (Draw)		• 2.0
• OpenOffice Draw		• 1.1 • 2.0
• OZ/2 PM Metafile Graphics	• MET	• 3.0
• Paint Shop Pro 6	• PSP	• 5.0-6.0
• PC Paintbrush	• PCX, DCX	• all
• Portable Bitmap Utilities	• PBM	• all
• Portable Network Graphics	• PNG	• 1.0
• Portable Pixmap Utilities	• PPM	• NA
• Postscript	• PS	• Level II
• Progressive JPEG		• NA
• StarOffice Draw		• 5.2 • 6.x • 7.x • 8.0
• Sun Raster Image	• RS	• NA
• Truevision Image	• TIFF	• Through 6
• Truevision Targa	• TGA	• 2
• Visio		• 4
• Visio 2003		• 5 • 2000 • 2002
• WBMP		• NA
• Windows Enhanced Metafile	• EMF	• NA
• Windows Metafile	• WMF	• NA

Format	Extension	Version
• WordPerfect Graphics	• WPG, WPG2	• Through 2.0 • 7 • 10
• X Bitmap	• XBM	• x10
• XDump	• XWD	• x10
• X Pixmap	• XPM	• x10

Table 6. Supported multimedia file types by category

Format	Extension	Version
• MPEG-1 Audio Layer 3	• MP3	• ID3 metadata only Note: These files can be harvested, but there is no data in them that can be used in tags.

Table 7. Supported presentation file types by category

Format	Extension	Version
• Corel Presentations	• SHW	• Through 12.0 • X3
• Lotus Freelance Graphics	• PRZ	• Through Millennium
• Lotus Freelance Graphics (OS/2)	• PRE	• Through 2.0
• Macromedia Flash	• SWF	• text only
• MS PowerPoint (Macintosh)	• PPT	• 4.0-2004
• MS PowerPoint (Windows)	• PPT	• 3.0-2007
• MS PowerPoint XML	• PPTX	
• OpenOffice Impress	• SXI/SXP/ODP	• 1.1 • 2.0
• StarOffice Impress	• SXI/SXP/ODP	• 5.2 • 6.x • 7.x • 8.0

Table 8. Supported spreadsheet file types by category

Format	Extension	Version
• Comma-Separated Values	• CSV	

• ENABLE Spreadsheet	• SSF	• 3.0 • 4.0 • 4.5
• First Choice		• Through 3.0
• Framework		• 3.0
• Lotus 1-2-3	• WK4	• Through 5.0
• Lotus 1-2-3 (OS/2)	•	• Through 2.0
• Lotus 1-2-3 Charts	• 123	• Through 5.0
• Lotus 1-2-3 for SmartSuite		• 197-9.6
• Lotus Symphony		• 1.0 • 1.1 • 2.0
• MS Excel	• XLS	• 2.2-2007
• MS Excel Charts		• 2.x-7.0
• MS Excel (Macintosh)	• XLS	• 3.0-4.0 • 1998 • 2001 • 2004
• MS Excel XML	• XLSX	
• MS MultiPlan		• 4.0
• MS Works	• S30/S40	• Through 2.0
• Mosaic Twin		• 2.5
• Novell Perfect Works		• 2.0
• OpenOffice Calc	• SXC/ODS	• 1.1 • 2.0
• PFS: Professional Plan		• 1.0

• Quattro Pro (DOS)	• Through 5.0
• Quattro Pro (Windows)	• Through 12.0 • X3
• Smart Ware II	• 1.02
• StarOffice Calc • SXC/ODS	• 5.2 • 6.x • 7.x • 8.0
• Supercalc Spreadsheet	• 4.0
• VP Planner 3D	• 1.0

Table 9. Supported system file types by category

Format	Extension
• Executable files	• .EXE
• Dynamic link library files	• .DLL
• Java class files	• .class
• DOS command executables	• .COM

Table 10. Supported text and markup file types by category

Format	Extension	Version
• ANSI	• .TXT	• 7- and 8-bit
• ASCII	• .TXT	• 7- and 8-bit
• EBCDIC		• all
• HTML	• .HTM	• Through 3.0
• IBM FFT		• all
• IBM Revisable Form Text		• all
• Rich Text Format	• RTF	• all
• Unicode Text	• .TXT	• all

- WML

- XML

- 5.2

- .XML

Table 11. Supported word-processing file types by category

Format	Extension	Version
Adobe FrameMaker Interchange Format	MIF	3.0-6.0
Corel WordPerfect Windows	WPD	Through 12.0 X3
DEC WPS PLUS	DX	Through 4.0
DEC WPS PLUS	WPL	Through 4.1
Display Write (2 and 3)	IP	all
Display Write (4 and 5)		Through 2.0
ENABLE		3.0 4.0 4.5
First Choice		Through 3.0
Framework		3.0
Haansoft Hangul	HWP	1997 2002
IBM Writing Assistant		1.01
JustSystems Ichitaro	JTD	5.0 6.0 8.0-13.0 2004
JustSystems Write		Through 3.0
Legacy		Through 1.1
Lotus AMI Pro	SAM	Through 3.1
Lotus Manuscript		2.0

• Lotus Manuscript		• 2.0
• Lotus Word Pro	• LWP	• 1996-9.6
• MacWrite II		• 1.1
• MASS-11		• Through 8.0
• MS Windows XML	• DOCX	
• MS Word (Macintosh)	• DOC	• 3.0-4.0 • 1998 • 2001
• MS Word (PC)	• DOC	• Through 6.0
• MS Word (Windows)	• DOC	• Through 2007
• MS WordPad		• all version
• MS Works	• WPS	• Through 4.0
• MS Works (Macintosh)		• Through 2.0
• MS Write		• Through 3.0
• MultiMate 4.0		• Through 4.0
• Navy DIF		• all versions
• Nota Bene		• 3.0
• Novell Perfect Works		• 2.0
• Novell WordPerfect		• Through 6.1
• Novell WordPerfect (Macintosh)		• 1.02-3.0
• Office Writer		• 4.0-6.0
• OpenOffice Writer	• SXW/ODT	• 1.1 • 2.0
• PC File Letter		• Through 5.0

- PC-File Letter
  - Through 5.0

- PC-File + Letter
  - Through 3.0

- PFS Write
  - A
  - B
  - C

- Professional Write Plus
  - Through 2.1

- Q&A (DOS)
  - 2.0

- Q&A (Windows)
  - 2.0

- Q&A Write
  - 3.0

- SAMNA Word IV

- Smart Ware II
  - 1.02

- Sprint
  - 1.0

- StarOffice Writer
  - SXW/ODT
  - 5.2
  - 6.x
  - 7.x
  - 8.0

- Total Word
  - 1.2

- **Parent topic:** [Supported file types](#)

Related reference

- [Supported file types](#)

SharePoint attributes

The following section describes the various SharePoint data object types and their properties that are currently supported by IBM® StoredIQ Platform.

Supported SharePoint object types

These types of SharePoint objects are supported:

Table 1. Supported SharePoint object types

Supported SharePoint object types	Supported SharePoint object types	Supported SharePoint object types
• Blog posts and comments	• Discussion board	• Calendar
• Tasks	• Project tasks	• Contacts
• Wiki pages	• Issue tracker	• Announcements
• Survey	• Links	• Document libraries
• Picture libraries	• Records center	

Notes regarding SharePoint object types

- Calendar: Recurring calendar events are indexed as a single object in IBM StoredIQ Platform. Each recurring calendar event has multiple `Event Date` and `End Date` attribute values, one pair per recurrence. For instance, if there is an event defined for `American Independence Day` and is set to recur yearly, it is indexed with Event Dates `2010-07-04`, `2011-07-04`, `2012-07-04`, and so on.
- Survey: Only individual responses to a survey are indexed as system-level objects. Each response is a user's feedback to all questions in the survey. Each question in the survey that was answered for a response is indexed as an attribute of the response in the IBM StoredIQ Platform index. The name of the attribute is the string that forms the question while the value is the reply entered. Surveys have no full-text indexable body, and they are always indexed with `size=0`.

Hash computation

The hash of a full-text indexed object is computed with the full-text indexable body of the object. However, in the case of SharePoint list item objects (excluding documents and pictures), the full-text indexable body might be empty or too simplistic. It means that you can easily obtain duplicate items across otherwise two different objects. For this reason, other attributes are included in the hash computation algorithm.

These attributes are included while the hash is computed for the SharePoint data objects, excluding documents and pictures.

Table 2. Attribute summary

Attribute	Types
Generic attributes	• Title (SharePoint) • Content type (SharePoint) • Description (SharePoint)

Blog post attributes	• Post category (SharePoint)
Wiki page attributes	• Wiki page comment
Calendar event attributes	• Event category (SharePoint) • Event date (SharePoint) • Event end date (SharePoint) • Event location (SharePoint)
Task or project task attributes	• Task start date (SharePoint) • Task due date (SharePoint) • Task that is assigned to (SharePoint)
Contact attributes	• Contact full name(SharePoint) • Contact email (SharePoint) • Contact job title (SharePoint) • Contact work address (SharePoint) • Contact work phone (SharePoint) • Contact home phone (SharePoint) • Contact mobile phone (SharePoint)
Link attributes	• Link URL (SharePoint)
Survey attributes	• All survey questions and answers in the response are included in the hash.

- **Parent topic:** [Supported file types](#)

Related reference

- [Supported file types](#)

Supported server platforms and protocols

The following section lists the supported server platforms by volume type and the protocols for supported systems.

Primary volume
> A primary volume is storage knowledge workers access to create, read, update, and delete unstructured content. Unstructured content is stored in standard formats such as office documents, text files, system logs, application logs, email, compressed archives that contain email, documents, or enterprise social media content.

Retention volume
> Generally, a retention volume is immutable storage that enforces retention and hold policies. While data is under management, it cannot be modified or deleted, and knowledge workers typically do not access retention storage directly. Specific applications such as IBM® StoredIQ eDiscovery move data to this storage to manage it. This storage typically has its own special API/protocol, although NAS vendors implemented retention/hold features that use standard CIFS and NFS protocols (NetApp SnapLock or EMC Celerra FLR are specific examples). The storage platform typically does not implement a hierarchical namespace to store content, but instead relies on a globally unique identifier as a handle to metadata and content. Applications are free to write metadata and binary content in any internal format to satisfy their requirements. Typically, IBM StoredIQ does not attempt to discover and manage data that is written by other applications to retention volumes. Application-specific knowledge is often required to interpret metadata and content. Retention storage is used by IBM StoredIQ to manage data on compliant immutable storage for retention and holds. It does so in a way that does not interfere with knowledge workers that create, update, and access content on primary volumes.
> IBM StoredIQ preserves source metadata when data is written to a retention volume. The original source metadata is important for governance and legal discovery (custodian, time stamps, and so on) to replicate the content and metadata from the retention volume when needed.

Export volume
> Export volumes are unmanaged (not indexed) storage location where content is copied along with metadata and audit detail in a format that can be imported by other applications. A common usage of export volumes is to stage native documents to be imported into a legal review tool in a format such as standard EDRM or a Concordance-compatible format.

System volume
> System volumes are a storage location where files can be written to and read by IBM StoredIQ. It can be used to export volume metadata that is contained in the index on a Data Server. Exported volume data can be imported from a system volume to populate a volume index.

Table 1. Supported platforms and protocols by IBM StoredIQ volume type

Platform/protocol	IBM StoredIQ primary volume	IBM StoredIQ retention volume	IBM StoredIQ export volume	IBM StoredIQ system volume	Notes
CIFS	x	x	x	x	
EMC Celerra FLR		x			
EMC Centera		x			
EMC Documentum	x	x			Customer must supply DFC files to enable the connector.
Hitachi HCAP		x			
IBM Information					

Platform/protocol					Notes
Information Archive (DR550)		x			
IBM Domino	x				Email only.
IBM FileNet	x				
Jive	x				
Microsoft Exchange	x				
Microsoft SharePoint	x				
NetApp SnapLock		x			
NewsGator	x				
NFS	x	x	x	x	
OpenText Livelink/Content Server	x				
Salesforce Chatter	x				
Symantec Discovery Accelerator	x				
Symantec Enterprise Vault	x	x			

Table 2. Supported operations and limitations by platform and protocol

Platform/protocol	Operation type: read	Operation type: write	Operation type: delete	Notes
CIFS	R	W	D	
EMC Celerra FLR	R	W	D	Treated as retention volume; only content written by IBM StoredIQ is supported.
EMC Centera	R	W	D	Metadata and content written by IBM StoredIQ or EMC EmailXtender is supported. EmailXtender support is limited to email archived from Exchange.
EMC Documentum	R	W	D	Only Current documents in standard cabinets are indexed.
Hitachi HCAP	R	W	D	
IBM Information Archive (DR550)	R	W	D	
IBM Domino	R			Email only. Email is converted to the .MSG format for processing.
IBM FileNet		W		
Jive	R			
Microsoft Exchange	R			Messages (email), Contacts, Calendar items, Notes, Tasks, and Documents

All document versions are optional.

	R	W	D	Notes
Excluding e				Documents
Microsoft SharePoint	R	W		SharePoint 2010/2013 supported list types: User profiles, User notes, Blog Post, Blog Comment, Discussion Post, Discussion Reply, Wiki Page, Calendar, Task/Project Task, Contact, Issue Tracker, Survey, Link, and Announcements. Content of custom lists is indexed generically as text. It is not modeled specifically, like standard list types.
NetApp SnapLock	R	W	D	Treated as retention volume; only content written by IBM StoredIQ is supported.
NewsGator	R			No API support for Poll responses.
NFS	R	W	D	
OpenText Livelink/Content Server	R			
Salesforce Chatter	R			
Symantec Discovery Accelerator	R			
Symantec Enterprise Vault	R	W	D	The delete action is only supported for retention volumes and only metadata and content written by IBM StoredIQ.

Table 3. Supported CIFS server platform and protocol versions (read/write/delete)

CIFS server	Notes
Windows XP, Vista, 7	
Windows Server 2003, 2008, 2008 R2	
Samba	
NetApp ONTAP 7.0.0 or later	
EMC Celerra DART 5.5 or later	
Mac OS X, 10.7, 10.8, 10.9	

Table 4. Supported EMC Centera server platform and protocol versions (read/write)

EMC Centera server	Notes
Centera API v3.1	

Table 5. Supported EMC Documentum server platform and protocol versions (read/write)

EMC Documentum server	Notes
Documentum 6.0, 6.5, 6.7	Customer must supply DFC files to enable the connector.
Documentum with Retention Policy Services (RPS) 6.0, 6.5	Customer must supply DFC files to enable the connector.

Table 6. Supported IBM FileNet server platform and protocol versions (write)

FileNet content services	Notes
FileNet 5.3	WebDAV protocol

Table 7. Supported Hitachi HCAP server platform and protocol versions (read/write/delete)

Hitachi HCAP service	Notes
Hitachi Content Platform v1.8	

Table 8. Supported IBM Domino/Notes server platform and protocol versions (read)

IBM Domino/Notes server	Notes
Domino/Notes 6.x, 7.x, 8.x	

Table 9. Supported IBM Information Archive server platform and protocol versions (read/write/delete)

IBM Information Archive	Notes
TSM Client 5.3.4.0	

Table 10. Supported Jive server platform and protocol versions (read)

Jive service	Notes
Jive 5.0.2	

Table 11. Supported Microsoft Exchange server platform and protocol versions (read)

Microsoft Exchange server	Notes
Exchange 2003	WebDAV protocol
Exchange 2007, 2010	Exchange web service interface

Table 12. Supported Microsoft SharePoint server platform and protocol versions (read/write)

Microsoft SharePoint server	Notes
SharePoint 2003	WebDAV protocol
SharePoint 2007, 2010, 2013	SharePoint web service interface

Table 13. Supported NewsGator server platform and protocol versions (read)

NewsGator API	Notes
NewsGator Social 2.1.1229	Installed on SharePoint 2010 or later.

Table 14. Supported NFS server platform and protocol versions (read/write/delete)

NFSv3	Notes
Red Hat Enterprise Linux 5.x, 6.x	
CentOS 5.x, 6.x	
NetApp ONTAP 7.0.0 or later	
EMC Celerra DART 5.5 or later	
Mac OS X, 10.7, 10.8, 10.9	

Table 15. Supported OpenText Livelink Enterprise Server platform and protocol versions (read)

Livelink Enterprise Server	Notes

| OpenText Livelink Enterprise Server 9.7, 9.7.1 | Connector is based on IBM Content Integrator (ICI) version 8.6. |

Table 16. Supported OpenText Content Server platform and protocol versions (read)

Content Server	**Notes**
Content Server 10.0.0	Connector is based on IBM Content Integrator (ICI) version 8.6.

Table 17. Supported Salesforce Chatter server platform and protocol versions (read)

Salesforce Chatter service	**Notes**
Salesforce Partner API v26.0	

Table 18. Supported Symantec Discovery Accelerator server platform and protocol versions (read)

Symantec Discovery Accelerator server	**Notes**
Enterprise Vault DCOM API 8.0.3, 9.0, 10.0	Query content and metadata from Discovery Accelerator cases.

Table 19. Supported Symantec Enterprise Vault server platform and protocol versions (read/write)

Symantec Enterprise Vault server	**Notes**
Enterprise Vault DCOM API 8.0.3, 9.0, 10.0	

Event log messages

The following section contains a complete listing of all ERROR, INFO, and WARN event-log messages that appear in the Event Log of the IBM® StoredIQ Platform console.

- [ERROR event log messages](#)
  The following table contains a complete listing of all ERROR event-log messages, reasons for occurrence, sample messages, and any required customer action.
- [INFO event log messages](#)
  The following table contains a complete listing of all INFO event-log messages.
- [WARN event log messages](#)
  The following table contains a complete listing of all WARN event-log messages, reasons for occurrence, sample messages, and any required customer action.

Related reference

- [ERROR event log messages](#)
- [INFO event log messages](#)
- [WARN event log messages](#)

ERROR event log messages

The following table contains a complete listing of all ERROR event-log messages, reasons for occurrence, sample messages, and any required customer action.

Table 1. ERROR event log messages. This table lists all ERROR event-log messages..

Type	Event number	Reason	Sample message	Required customer action
ERROR	1001	Harvester was unable to open a socket for listening to child processes.	Harvester could not allocate listen port after <number> attempts. Cannot kickstart interrogators. (1001)	Log in to UTIL and restart the application server. Restart the data server. Contact Customer Support.
ERROR	9083	Unexpected error while it is exporting a volume.	Exporting volume 'dataserver:/mnt/demo-A' (1357) has failed (9083)	Contact Customer Support.
ERROR	9086	Unexpected error while it is importing a volume	Importing volume 'dataserver:/mnt/demo-A' (1357) failed (9086)	Contact Customer Support.
ERROR	15001	No volumes are able to be harvested in a job. For instance, all of the mounts fail due to a network issue.	No volumes harvested. (15001)	Make sure IBM® StoredIQ Platform still has appropriate permissions to a volume. Verify that there is network connectivity between the data server and your volume. Contact Customer Support.
ERROR	15002	Could not mount the volume. Check permissions and network settings.	Error mounting volume <share> <start-dir> on server <server-name>. Reported <reason>. (15002)	Make sure that the data server still has appropriate permissions to a volume. Verify that there is network connectivity between the data server and your volume. Contact Customer Support.
ER 1			Failed to	

R 5 R 0 O 2 R 1	Error saving harvest record.	save HarvestRecord for qa1:auto-A (15021)	Contact Customer Support. This message occurs due to a database error.
E 1 R 7 R 0 O 0 R 1	Unhandled unrecoverable exception in Centera Discovery.	Centera Harvester fatal failure: <exception description> (17001)	Contact Customer Support.
E 1 R 7 R 0 O 1 R 2	An error occurred when a volume was created during Centera Discovery.	Unable to create Centera Volume Company_jpool_ 2009_ FEB_1 in pool jpool. Error: <database error description> (17012)	Contact Customer Support. This message occurs due to a database error.
E 1 R 7 R 5 O 0 R 1	Generic retention discovery failed in a catastrophic manner.	Generic retention discovery fatal failure: <17501>	Contact Customer Support.
E 1 R 7 R 5 O 0 R 3	Generic retention discovery creates volume sets associated with primary volumes. When that fails, IBM StoredIQ Platform sends this message. This failure likely occurred due to database errors.	Error creating/loading volumeset for <server>: <share>	Contact Customer Support.
E 1 R 7 R 5 O 0 R 5	Unable to query object count for a discovered volume due to a database error.	Unable to determine object count for <server>: <share>	Contact Customer Support.
E 1 R 7 R 5 O 0 R 6	Generic retention discovery could not create discovered volume.	Error creating volume <server>: <share:>	Contact Customer Support.
E 1 R 8 R 0 O 0 R 1	SMB connection fails.	Windows Share Protocol Exception when connecting to the server <server-name> : <reason>. (18001)	Make sure IBM StoredIQ Platform still has appropriate permissions to a volume. Verify that there is network connectivity between the data server and your volume. Contact Customer Support.

E 1 R 8 R 0 O 0 R 2	The SMB volume mount failed. Check the share name.	Windows Share Protocol Exception when connecting to the share <share-name> on <server-name> : <reason>. (18002)	Verify the name of the server and volume to make sure that they are correct. If this message persists, then contact Customer Support.
E 1 R 8 R 0 O 0 R 3	There is no volume manager.	Windows Share Protocol Exception while initializing the data object manager: <reason>. (18003)	Contact Customer Support.
E 1 R 8 R 0 O 0 R 6	Grazer volume crawl threw an exception.	Grazer._run : Unknown error during walk. (18006)	Verify the user that mounted the specified volume has permissions equivalent to your current backup solution. If this message continues, contact Customer Support.
E 1 R 8 R 0 O 2 R 1	An unexpected error from the server prevented the harvest to reach the end of the activity stream on the NewsGator data source that is harvested. The next incremental harvest attempts to pick up from where the current harvest was interrupted.	Unable to fetch trailing activity stream from NewsGator volume. Will retry in next harvest. (18021)	Check to ensure the NewsGator server has sufficient resources (disk space, memory). It is likely that this error is transient. If the error persists across multiple harvests, contact Customer Support.
E 1 R 8 R 0 O 1 R 8	Start directory has escape characters, and the data server is configured to skip them.	Cannot graze the volume, root directory Nunez has escape characters (18018)	Consider turning off escape character checking.
E 1 R 9 R 0 O 0 R 1	An exception occurred during interrogator initialization.	Interrogator._init__exception: <reason>. (19001)	Contact Customer Support.
E 1 R 9 R 0 O 0 R 2	An unknown exception occurred during interrogator initialization.	Interrogator._init__exception: unknown. (19002)	Contact Customer Support.
E 1 R 9	An exception occurred	Interrogator.pro-cess exception	

R 0 O 0 R 3	during interrogator processing.	(<volumeid>, <epoch>): <reason>. (19003)	Contact Customer Support.	
E 1 R 9 R 0 O 0 R 4	An unknown exception occurred during interrogator processing.	Interrogator. pro-cess exception (<volumeid>, <epoch>). (19004)	Contact Customer Support.	
E 1 R 9 R 0 O 0 R 5	An exception occurred during viewer initialization.	Viewer.__init__: Exception - <reason>. (19005)	Contact Customer Support.	
E 1 R 9 R 0 O 0 R 6	An unknown exception occurred during viewer initialization.	Viewer.__init__: Unknown exception. (19006)	Contact Customer Support.	
E 3 R 3 R 0 O 0 R 4	Volume could not be unmounted.	Unmounting volume failed from mount point : <mount point>. (33004)	Restart the data server. If the problem persists, then contact Customer Support.	
E 3 R 3 R 0 O 0 R 5	Data server was unable to create a local mounting point for the volume.	Unable to create mount_point using primitive.thread-Safe-Makedirs(). (33005)	Restart the data server. If the problem persists, then contact Customer Support.	
E 3 R 3 R 0 O 1 R 0	Failed to make SMB connection to Windows Share server.	Mounting Windows Share volume failed with the error : <system error message>. (33010)	Verify user name and password that is used for mounting the volume are accurate. Check the user data object for appropriate permissions to the volume. Make sure that the volume is accessible from one of the built-in protocols (Windows Share). Verify that the network is properly configured for the data server to reach the volume. Verify that the data server has appropriate DNS settings to resolve the server name.	
E 3 R 3 R 0 O 1 R 1	Internal error. Problem accessing local /proc/mounts	Unable to open /proc/mounts. Cannot test if volume was already mounted. (33011)	Restart the data server. If the problem persists, then contact Customer Support.	
E 3 R 3 R 0 O 1	Database problems when a volume was deleted.	An exception occurred while working with HARVESTS_ TABLE in	Contact Customer Support.	

R 2		Volume._delete(). (33012)	
E 3 R 3 R 0 O 1 R 3	No volume set was found for the volumes set name.	Unable to load volume set by its name. (33013)	Contact Customer Support.
E 3 R 3 R 0 O 1 R 4	System could not determine when this volume was last harvested.	An error occurred while performing the last_harvest operation. (33014)	Contact Customer Support.
E 3 R 3 R 0 O 1 R 8	An error occurred mounting the Exchange share.	Mounting Exchange Server failed : <reason>. (33018)	Verify user name and password that is used for mounting the share are accurate. Check for appropriate permissions to the share. Make sure that the share is accessible. Verify that the network is properly configured for the data server to reach the share. Verify that the data server has appropriate DNS settings to resolve the server name.
E 3 R 3 R 0 O 1 R 9	Failed to connect and authenticate to the Hitachi Archivas Content Archive server.	Mounting HCAP volume failed : Cannot connect to HCAP share. (33019)	Ensure the connectivity, credentials, and permissions to the Hitachi volume and try again.
E 3 R 3 R 0 O 2 R 0	Failed to connect and authenticate to the IBM Information Archive retention volume.	Mounting IBM Information Archive volume failed with the error: Server unreachable. (33020)	Ensure the connectivity, credentials, and permissions to the IBM Information Archive volume and try again.
E 3 R 3 R 0 O 2 R 2	Failed to connect to the Discovery Accelerator with the information for the Volume.	Mounting Discovery Accelerator volume failed with the error: insufficient permissions to review CaseOne (33022)	Verify the information that is used to add the volume and ensure that all details are entered correctly before you try it again. If the error points to network issues with connectivity, address them and try again.
E 3 R 3 R 0 O 2 R 7	The attempt to connect and authenticate to the IBM FileNet server failed.	Mounting IBM FileNet volume failed : <reason>. (33027)	Ensure the connectivity, credentials, and permissions to the FileNet volume and try again.
E 3 R 3	Failed to create the volume	Exceeded maximum	Contact Customer Support.

R 0 O 2 R 9	as the number of active volume partitions exceeds the limit of 500.	number of volume partitions (33029).	~~Contact Customer Support.~~
E 3 R 4 R 0 O 0 R 2	Could not complete the copy action because the target disk was full.	Copy Action aborted as the target disk has run out of space (34002)	Verify that there is space available on your policy destination and try again.
E 3 R 4 R 0 O 0 R 9	Could not complete the move action due to full target disk.	Move Action aborted as the target disk has run out of space. (34009)	Verify that there is space available on your policy destination, then run another harvest before you run the policy. When the harvest completes, try running the policy again.
E 3 R 4 R 0 O 1 R 5	The policy audit could not be deleted for some reason.	Error Deleting Policy Audit: <error message> (34016)	Contact Customer Support.
E 3 R 4 R 0 O 2 R 0	The copy to Centera action could not be run because of insufficient permissions.	Copy to Centera failed as we do not have read/write permissions on the access profile used. (34020)	Check permissions on the access profile that is provided for the Centera pool on which the volume is defined, and check whether the appropriate permissions are provided.
E 3 R 4 R 0 O 2 R 1	The move to Centera action could not be run because of insufficient permissions.	Move to Centera failed as we do not have read/write permissions on the access profile used. (34020)	Check permissions on the access profile that is provided for the Centera pool on which the volume is defined, and check whether the appropriate permissions are provided.
E 3 R 4 R 0 O 3 R 0	Discovery export policy is started since it detected the target disk is full.	Production Run action aborted because the target disk has run out of space. (34030)	Create sufficient space on target disk and run discovery export policy again.
E 3 R 4 R 0 O 3 R 4	The target volume for the policy could not be mounted. The policy is started.	Copy objects failed, unable to mount volume: QA1.COMPANY. COM:SHARE . (34034)	Ensure the connectivity, login credentials, and permissions to the target volume for the policy and try again.
E 4 R 1 R 0	The job is ended	<job-name> ended	Try to run the job again. If it fails again,

Code	Description	Message	Action
O 0 R 4	abnormally.	unexpectedly. (41004)	contact Customer Support.
E 4 R 1 R 0 O 0 R 7	Job failed.	[Job name] has failed (41007).	Look at previous messages to see why it failed and refer to that message ID to pinpoint the error. Contact Customer Support.
E 4 R 2 R 0 O 0 R 1	The copy action could not run because of parameter errors.	Copy data objects did not run. Errors occurred: <error-description>. (42001)	Contact Customer Support.
E 4 R 2 R 0 O 0 R 2	The copy action was unable to create a target directory.	Copy data objects failed, unable to create target dir:<target-directory-name>. (42002)	Check permissions on the target. Make sure the permissions that are configured to mount the target volume have write access to the volume.
E 4 R 2 R 0 O 0 R 4	An unexpected error occurred.	Copy data objects terminated abnormally. (42004)	Contact Customer Support.
E 4 R 2 R 0 O 0 R 6	The move action could not run because of parameter errors.	Move data objects did not run. Errors occurred: <error-description>. (42006)	Contact Customer Support.
E 4 R 2 R 0 O 0 R 7	The move action was unable to create a target directory.	Move data objects failed, unable to create target dir:<target-directory-name>. (42007)	Check permissions on the target. Make sure the permissions that are configured to mount the target volume have write access to the volume.
E 4 R 2 R 0 O 0 R 9	An unexpected error occurred.	Move data objects terminated abnormally. (42009)	Contact Customer Support.
E 4 R 2 R 0 O 1 R 7	An unexpected error occurred.	Delete data objects terminated abnormally. (42017)	Contact Customer Support.
E 4 R 2	The policy action could not	Policy cannot execute.	

R 0 O 2 R 5	run because of parameter errors.	Attribute verification failed. (42025)	Contact Customer Support.
E 4 R 2 R 0 O 2 R 7	An unexpected error occurred.	Policy terminated abnormally. (42027)	Contact Customer Support.
E 4 R 2 R 0 O 5 R 0	The data synchronizer could not run because of an unexpected error.	Content Data Synchronizer synchronization of <server-name>: <volume-name> failed fatally.	Contact Customer Support.
E 4 R 2 R 0 O 5 R 9	Invalid set of parameters that are passed to discovery export policy.	Production Run on objects did not run. Errors occurred: The following parameters are missing: action_limit. (42059)	Contact Customer Support.
E 4 R 2 R 0 O 6 R 0	Discovery export policy failed to create target directory for the export.	Production Run on objects (Copying native objects) failed, unable to create target dir: production/10. (42060)	Verify that the discovery export volume has write permission and re-execute policy.
E 4 R 2 R 0 O 6 R 2	Discovery export policy was ended abnormally.	Production Run on objects (Copying native objects) terminated abnormally. (42062)	Contact Customer Support.
E 4 R 2 R 0 O 8 R 8	The full-text optimization process failed; however, the index is most likely still usable for queries.	Full-text optimization failed on volume <volume-name> (42088)	Contact Customer Support.
		Time allocated to	

E 4 R 5 R 8 O 0 R 2	A full-text index is already being modified.	gain exclusive access to in-memory index for volume= 1357 has expired (45802)	Contact Customer Support.
E 4 R 5 R 8 O 0 R 3	The index for the specified volume does not exist. This message can occur under normal conditions.	Index '/deepfs/full-text/volume_index/volume_1357' not found. (45803)	No user intervention is required.
E 4 R 5 R 8 O 0 R 4	Programming error. A transaction was never initiated or was closed early.	Transaction of client: node.client.com_ FINDEX_ QUEUE_ 1357_117251 522 _3_ 2 is not the writer (45804)	Contact Customer Support.
E 4 R 5 R 8 O 0 R 5	The query is not started or expired. The former is a programming error. The latter is normal.	Query ID: 123 does not exist (45805)	No user intervention is required.
E 4 R 5 R 8 O 0 R 6	The query expression is invalid or not supported.	Failed to parse 'dog pre\3 bar' (45806)	Revise your full-text query.
E 4 R 5 R 8 O 0 R 7	Programming error. A transaction was already started for the client.	Client: node.client.com_ FINDEX_QUEUE _1357_11725 15 222_3_2 is already active (45807)	Contact Customer Support.
E 4 R 5 R 8 O 0 R 8	A transaction was never started or expired.	No transaction for client: node.client.com_ FINDEX_ QUEUE_1357_ 1172515222_ 3_2 (45808)	No user intervention is required. The system handles this condition internally.
E 4		Invalid	

Code	Description	Message	Action
R 5 R 8 O 1 R 0	Programming error.	volumeId. Expected: 1357 Received:2468 (45810)	Contact Customer Support.
E 4 R 5 R 8 O 1 R 2	A File I/O error occurred while the system was accessing index data.	Failed to write disk (45812).	Try your query again. Contact Customer Support for more assistance if necessary.
E 4 R 5 R 8 O 1 R 4	The query expression is too long.	Query: 'a* b* c* d* e*' is too complex (45814)	Refine your full-text query.
E 4 R 5 R 8 O 1 R 5	The file that is being indexed is too large or the query expression is too complex. The engine temporarily ran out of memory.	Java heap exhausted while indexing node with ID: '10f4179cd5f f22f 2a6b 79a1bc3aef2 47 fd94ccff' (45815)	Check the skipped file list in the audit log for files that failed to load due to their sizes. Revise your query expression and try again.
E 4 R 6 R 0 O 2 R 3	Tar command failed while it persists full-text data to Windows Share or NFS share.	Failed to back up full-text data for server:share. Reason: <reason>. (46023)	Check disk space and permissions.
E 4 R 6 R 0 O 2 R 4	Unhandled unrecoverable exception while persisting full-text data into a .tgz file.	Exception <exception> while backing up fulltext data for server:share (46024)	Contact Customer Support.
E 4 R 6 R 0 O 2 R 5	Was not able to delete partial .tgz file after a failed full-text backup.	Failed to unlink incomplete backup image. Reason: <reason>. (46025)	Check permissions.
E 4 R 7 R 0 O 0 R 2	Synchronization failed on a query.	Synchronization failed for query '<query-name>' on volume '<server-and-volume> (47002)	Contact Customer Support.
E 4		Cannot process full-	

R 7 R 1 O 0 R 1	An error occurred during the query of a full-text expression.	text expression (Failed to read from disk (45812) (47101)	Restart services and contact Customer Support.
E 4 R 7 R 2 O 0 R 3	No more database connections are available.	Database connections exhausted (512/511) (47203)	Contact Customer Support.
E 4 R 7 R 2 O 0 R 7	User is running out of disk space.	Disk usage exceeds threshold. (%d)	Contact Customer Support. In rare cases, this message can indicate a program error leaking disk space. In most cases, however, disk space is almost full, and more storage is required.
E 4 R 7 R 2 O 1 R 2	Interrogator failed while the system processed a file. The current file is missing from the volume cluster.	Harvester 1 Does not exist. Action taken : restart. (47212)	If the problem persists (that is, the system fails on the same file or type of files), contact Customer Support.
E 4 R 7 R 2 O 1 R 4	SNMP notification sender is unable to resolve the trap host name.	Unable to resolve host name nomachine.n o where.com (47214)	Check spelling and DNS setup.
E 5 R 0 R 0 O 1 R 1	The DDL/DML files that are required for the database versioning were not found in the expected location on the data server.	Database version control SQL file not found. (50011)	Contact Customer Support.
E 5 R 0 R 0 O 1 R 8	Indicates that the pre-upgrade database restoration failed, which was attempted as a result of a database upgrade failure.	Database restore is unsuccessful. Contact Customer Support. (50018)	Contact Customer Support.
E 5 R 0 R 0 O 2 R 0	Indicates that the current database requirements do not meet those requirements that are specified for the upgrade and cannot proceed with the upgrade.	Versions do not match! Expected current database version: <dbversion> . (50020)	Contact Customer Support.
E 5 R 0 R 0 O 2 R 1	Indicates that the full database backup failed when the system attempts a data-object level database backup.	Database backup failed. (50021)	Contact Customer Support.
E 6 R 1 R 0 O 0 R 3	Discovery export policy failed to mount volume.	Production policy failed to mount volume. Aborting.	

Code	Explanation	Message	Action
ERROR 61005	The discovery export load file generation fails unexpectedly. The load files can be produced correctly, but post-processing actions like updating audit trails and generating report files might not complete.	(61003) Production load file generation failed. Load files may be produced, but post-processing may be incomplete. (61005)	Contact Customer Support.
ERROR 61006	The discovery export load file generation was interrupted because the target disk is full.	Production load file generation interrupted. Target disk full. (61006)	Free up space on the target disk, void the discovery export run and run the policy again.
ERROR 68001	The gateway and data server must be on the same version to connect.	Gateway connection failed due to unsupported data server version.	Update your data server to the same build number as the gateway and restart services. If your encounter issues, contact Customer Support.
ERROR 68003	The data server failed to connect to the gateway over an extended period.	The data-server connection to the gateway cannot be established.	Contact Customer Support.
ERROR 80002	The system failed to open a connection to the database.	Failed to connect to the database (80002)	The "maximum database connections" configuration parameter of the database engine might need to be increased. Contact Customer Support.

- **Parent topic:** [Event log messages](#)

Related reference

- [Event log messages](#)

INFO event log messages

The following table contains a complete listing of all INFO event-log messages.

Table 1. INFO event log messages

Type	Event number	Reason	Sample message	Required customer action
INFO	9001	No conditions were added to a query.	Harvester: Query <query name> cannot be inferred because no condition for it has been defined (9001).	Add conditions to the specified query.
INFO	9002	One or more conditions in a query were incorrect.	Harvester: Query < query name> cannot be inferred because of regular expression or other condition error (9002).	Verify that regular expressions are properly formed.
INFO	9003	Volume Harvest is complete and explorers are being calculated.	Volume statistics computation started (9003).	No user intervention is required.
INFO	9004	Explorer calculations are complete.	Volume statistics computation completed (9004).	No user intervention is required.
INFO	9005	Query membership calculations started.	Query inference will be done in <number> steps (9005).	No user intervention is required.
INFO	9006	Query membership calculations progress information.	Query inference step <number> done (9006).	No user intervention is required.
INFO	9007	Query membership calculations completed.	Query inference completed (9007).	No user intervention is required.
INFO	9012	Indicates the end of dumping the content of the volume cache.	Dump of Volume cache(s) completed (9012).	No user intervention is required.

INFO	9013	Indicates the beginning of the load process.	Postprocessing for volume 'Company Data Server:/mnt/demo-A' started (9013).	No user intervention is required.
INFO	9067	Indicates load progress.	System metadata and tagged values were successfully loaded for volume 'server:volume' (9067).	No user intervention is required.
INFO	9069	Indicates load progress.	Volume 'data server: /mnt/demo-A': System metadata, tagged values and full-text index were successfully loaded (9069).	No user intervention is required.
INFO	9084	The volume export finished.	Exporting volume 'data server:/mnt/demo-A' (1357) completed (9084)	No user intervention is required.
INFO	9087	The volume import finished.	Importing volume 'dataserver:/mnt/demo-A' (1357) completed (9087)	No user intervention is required.
INFO	9091	The load process was ended by the user.	Load aborted due to user request (9091).	No user intervention is required.
INFO	15008	The volume load step was skipped, per user request.	Post processing skipped for volume <server>:<vo-lume>. (15008)	No user intervention is required.
INFO	15009	The volume load step was run but the harvest step was skipped, per user request.	Harvest skipped for volume <server>: <vol-ume>. (15009)	No user intervention is required.
INFO	15012	The policy that ran on the volume is complete and the volume load can	Volume <volume> on server <server> is free now. Proceeding with load. (15012)	No user intervention is required.

			now proceed.	
INFO	15013	The configured time limit on a harvest was reached.	Harvest time limit reached for server:share. Ending harvest now. (15013)	No user intervention is required.
INFO	15014	The configured object count limit on a harvest was reached.	Object count limit reached for server:share. Ending harvest now. (15014)	No user intervention is required.
INFO	15017	Check box is selected for nightly load job.	Deferring post processing for volume server:vol (15017)	No user intervention is required.
INFO	15018	Harvest size or time limit is reached.	Harvest limit reached on server:volume. Synthetic deletes will not be computed. (15018)	No user intervention is required.
INFO	15019	User stops harvest process.	Harvest stopped by user while processing volume dpfsvr:vol1. Rest of volumes will be skipped. (15019)	No user intervention is required.
INFO	15020	The harvest vocabulary changed. Full harvest must run instead of incremental.	Vocabulary for dpfsvr:jhaide-A has changed. A full harvest is recommended (15020).	Full harvest must be run instead of an incremental harvest.
INFO	15022	The user is trying to run an ACL-only harvest on a volume that is not a Windows Share or SharePoint volume.	Permission-only harvest: permission checks not supported for <server>: <share>	No action is needed as the volume is skipped.
INFO	15023	The user is trying to run an ACL-only harvest on a volume that was assigned a user list.	Permission-only harvest: volume <server>:<share> has no associated user list.	No action is needed as the volume is skipped.
		Sent when Centera	Centera External Iterator : Starting to	No user

INFO	17002	Discovery sends the query to the Centera server	populate using pool QAPOOL. (17002)	intervention is required.
INFO	17003	Sent when Centera Discovery autocreates a new volume	<servername>: Created new volume QAPOOL:_QA-POOL_2009_JAN_1 (17003)	No user intervention is required.
INFO	17004	A Centera Discovery auto-created volume reached the preconfigured limit, starting a new one.	Object limit reached for _QAPOOL_2009 _JAN_1, starting new volume (17004)	No user intervention is required.
INFO	17007	Pending data return from Centera.	Centera Harvester: No items returned from Centera for over 5 minutes. Still waiting. (17007)	Check whether a Centera node is down. A Centera cluster might be overloaded.
INFO	17009	Configured time limit is reached for Centera discovery step.	Centera Discovery : time limit for discovery reached. Ending this run. (17009)	No user intervention is required.
INFO	17010	Configured object count limit that is reached for Centera discovery step.	Centera Discovery: configured item count limit for discovery reached. Ending this run. (17010)	No user intervention is required.
INFO	17507	Limit (time or object count) reached for generic retention discovery.	Retention discovery limit reached for <server>:<share>	Contact Customer Support.
INFO	17508	Generic retention discovery found no new items for this master volume.	No new items discovered. Post-processing skipped for volume <server>: <share>	No user intervention is required unless the user is certain that new items must be discovered.
INFO	17509	Generic retention discovery created	Created new discovered volume <server>:<share> in	No user intervention is

INFO	17909	a new volume.	volume set <autodis-covered volume set name>.	required.
INFO	18004	Job was stopped.	Walker._process File: Grazer Stopped. (18004)	No user intervention is required.
INFO	18005	Grazer queue was closed.	Walker._processFile: Grazer Closed. (18005)	No user intervention is required.
INFO	18016	Displays the list of top-level directories that are selected by matching the start directory regular expression. Displays at the beginning of a harvest.	Choosing top-level directories: <directories> (18016)	No user intervention is required.
INFO	34001	Marks current progress of a copy action.	<volume>: <count> data objects processed by copy action. (34001)	No user intervention is required.
INFO	34004	Marks current progress of a delete action.	<volume>: <count> data objects processed by delete action. (34004)	No user intervention is required.
INFO	34008	Marks current progress of a move action.	<volume>: <count> data objects processed by move action. (34008)	No user intervention is required.
INFO	34014	A policy audit was deleted.	Deleting Policy Audit # <audit id> <policy name> <start time> (34014)	No user intervention is required.
INFO	34015	A policy audit was deleted.	Deleted Policy Audit # <audit id> <policy name> <start time> (34015)	No user intervention is required.
INFO	34031	Progress update of the discovery export policy, every 10000 objects	Winserver:top share : 30000 data objects processed by production action.	No user intervention is required.

		processed.	(34031)	
INFO	41001	A job was started either manually or was scheduled.	\<jobname> started. (41001)	No user intervention is required.
INFO	41002	The user stopped a job that was running.	\<jobname> stopped at user request (41002)	No user intervention is required.
INFO	41003	A job is completed normally with or without success.	\<jobname> completed. (41003)	No user intervention is required.
INFO	41006	Rebooting or restarting services on the controller or compute node causes all jobs to stop.	Service shutdown. Stopping outstanding jobs. (41006)	Rerun jobs after restart if you want the jobs to complete.
INFO	41008	Database compactor (vacuum) job cannot run while there is database activity.	Database compactor was not run because other jobs are active (41008).	Set the database compactor's job schedule so that it does not conflict with long-running jobs.
INFO	42005	The action completed or was ended. Shows results of copy action.	Copy complete: \<number> data objects copied, \<number> collisions found. (42005)	No user intervention is required.
INFO	42010	The action completed or was ended. Shows results of move action.	Move complete: \<number> data objects moved, \<number > collisions found. (42010)	No user intervention is required.
INFO	42018	The action completed or was ended. Shows results of deleted action.	Copy data objects complete: \<number> data objects copied, \<number> collisions found. (42018)	No user intervention is required.
		Third-party	Centera Deleted Files Synchronizer	

INFO	42020	application deleted clips on discovered Centera volumes.	complete: 10000 objects passed. 2 objects were missing from the Centera cluster. (42020)	No user intervention is required.
INFO	42024	The synchronizer was completed normally.	Content Data Synchronizer complete. (42024)	No user intervention is required.
INFO	42028	The action completed or was ended. Shows results of policy action.	Policy completed (42028).	No user intervention is required.
INFO	42032	The action completed or was ended. Shows results of report action.	<report name> completed (42032).	No user intervention is required.
INFO	42033	The synchronizer started automatically or manually with the GUI button.	Content Data Synchronizer started. (42033)	No user intervention is required.
INFO	42048	Reports that the synchronizer is skipping a volume if synchronization is determined not to be required.	Content Data Synchronizer skipping <server-name>:<volume-name> as it does not need synchronization. 42048)	No user intervention is required.
INFO	42049	Reports that the synchronizer started synchronization of a volume.	Content Data Synchronizer starting synchronization for volume <server-name>:<volume-name>	No user intervention is required.
INFO	42053	The policy that was waiting for participant volumes to be loaded before it continues, is now starting.	Proceeding with execution of <policy-name>.	No user intervention is required.
		Report on completion of	Production Run on objects (Copying	No user

INFO	42063	discovery export policy execution phase.	native objects) completed: 2003 data objects copied, 25 duplicates found. (42063)	intervention is required.
INFO	42065	A discovery export policy that was held up for want of resources, is now done waiting, and begins execution.	Proceeding with execution of 'Production case One'. (42065)	No user intervention is required.
INFO	42066	A new discovery export run started.	New run number 10 started for production Production Case 23221. (42066)	Note the new run number to tie the current run with the corresponding audit trail.
INFO	42067	Discovery export policy is preparing the audit trail in XML format. It might take a few minutes.	Production Run producing Audit Trail XML. (42067)	No user intervention is required.
INFO	42074	A query or tag was replicated to a member data server successfully.	Successfully sent query 'Custodian: Joe' to member data server San Jose Office (42074)	No user intervention is required.
INFO	46001	The backup process began. Any selected backups in the system configuration screen are run if necessary.	Backup Process Started. (46001)	No user intervention is required.
INFO	46002	The backup process did not complete all its tasks successfully. One or more backup types did not occur.	Backup Process Failed: <error-description>. (46002)	Check your backup volume.
		The backup process		

INFO	46003	completed attempting all the necessary tasks successfully. Any parts of the overall process add their own log entries.	Backup Process Finished. (46003)	No user intervention is required.
INFO	46004	The Application Data backup, as part of the overall backup process, needed to run but did not succeed.	Application Data backup failed. (46004)	Check your backup volume. Look at the setup for the Application Data backup. If backups continue to fail, contact Customer Support.
INFO	46005	The Application Data backup, as part of the overall backup process, needed to run and succeeded.	Application Data backup finished. (46005)	No user intervention is required.
INFO	46006	The Application Data backup, as part of the overall backup process, was not configured.	Application Data backup not configured, skipped. (46006)	No user intervention is required.
INFO	46007	The Harvested Volume Data backup, as part of the overall backup process, needed to run but did not succeed.	Harvested Volume Data backup failed. (46007)	Check your backup volume. Look at the setup for the Harvested Volume Data backup. If backups continue to fail, contact Customer Support.
INFO	46008	The Harvested Volume Data backup, as part of the overall backup process, needed to run	Harvested Volume Data backup finished. (46008)	No user intervention is required.

		and succeeded.		
INFO	46009	The Harvested Volume Data backup, as part of the overall backup process, was not configured.	Harvested Volume Data backup not configured, skipped. (46009)	No user intervention is required.
INFO	46010	The System Configuration backup, as part of the overall backup process, needed to run but did not succeed.	System Configuration backup failed. (46010)	Check your backup volume. Look at the setup for the System Configuration backup. If backups continue to fail, contact Customer Support.
INFO	46011	The System Configuration backup, as part of the overall backup process, needed to run and succeeded.	System Configuration backup finished. (46011)	No user intervention is required.
INFO	46012	The System Configuration backup, as part of the overall backup process, was not configured.	System Configuration backup not configured, skipped. (46012)	No user intervention is required.
INFO	46013	The Audit Trail backup, as part of the overall backup process, needed to run but did not succeed.	Policy Audit Trail backup failed. (46013)	Check your backup volume. Look at the setup for the Audit Trail backup. If back-ups continue to fail, contact IBM support.
INFO	46014	The Audit Trail backup, as part of the overall backup process, needed to run and succeeded.	Policy Audit Trail backup finished. (46014)	No user intervention is required.

INFO	46015	The Audit Trail backup, as part of the overall backup process was not configured.	Policy Audit Trail backup not configured, skipped. (46015)	No user intervention is required.
INFO	46019	Volume cluster backup failed.	Indexed Data backup failed: <specific error> (46019)	Contact Customer Support.
INFO	46020	Volume cluster backup finished.	Indexed Data backup finished. (46020)	No user intervention is required.
INFO	46021	Volume is not configured for indexed data backups.	Indexed Data backup not configured, skipped. (46021)	No user intervention is required.
INFO	46022	Full-text data was successfully backed up.	Successfully backed up full-text data for server:share (46022)	No user intervention is required.
INFO	47213	Interrogator was successfully restarted.	Harvester 1 is now running. (47213)	No user intervention is required.
INFO	60001	The user updates an object on the system. It includes any object type on the data server, including the updating of volumes.	Query cities was updated by the administrator account (60001).	No user intervention is required.
INFO	60002	The user creates an object. It includes any object type on the data server, including the creation of volumes.	Query cities was created by the administrator account (60002).	No user intervention is required.
INFO	60003	The user deletes an object. It includes any object type on the data server, including the deletion of	Query cities was deleted by the administrator account (60003).	No user intervention is required.

		volumes.		
INFO	60004	The user publishes a full-text query set or a query.	Query cities draft was published by the administrator account (60004).	No user intervention is required.
INFO	60005	The user tags an object. It includes a published query, a draft query, or tag.	Query tagging for cities class was started by the administrator account (60005).	No user intervention is required.
INFO	60006	A user restarted services on the data server.	Application services restart for all data servers was requested by the administrator account (60006).	No user intervention is required.
INFO	61001	Concordance discovery export is now preparing the load files.	Preparing for upload of load file(s). (61001)	No user intervention is required.
INFO	61002	Concordance discovery export is ready to upload the load files.	Load file(s) ready for upload. (61002)	No user intervention is required.
INFO	65000	The log file finished downloading.	Log file download complete (65000)	No user intervention is required.

- **Parent topic:** [Event log messages](#)

Related reference

- [Event log messages](#)

WARN event log messages

The following table contains a complete listing of all WARN event-log messages, reasons for occurrence, sample messages, and any required customer action.

Table 1. WARN event log messages

Type	Event number	Reason	Sample message	Customer action
WARN	1002	An Interrogator process failed because of an unknown error. The data object that was processing is skipped. A new process is created to replace it.	Processing could not be completed on object, interrogator died : <data object name>. (1002)	Classify the document manually and Contact Customer Support.
WARN	1003	Interrogator child process did not properly get started. There might be problems to access the volume to be harvested.	Interrogator terminated before accessing data objects. (1003)	Try to readd the volume that is harvested. If that fails, contact Customer Support.
WARN	1004	Interrogator child process was ended because it was no longer responding. The data object that was processing is skipped. A new process is created to replace it.	Processing was not completed on object, interrogator killed : <data object name>. (1004)	Contact Customer Support.
WARN	6001	A user email might not be sent. The mail server settings are incorrect.	Failed to send an email to user <email address>; check mail server configuration settings (6001).	Verify that your SMTP server is configured correctly. Make sure that the IP address that is configured for the data server can relay on the configured SMTP server.
WARN	8001	The database needs to be vacuumed.	The Database is approaching an operational limit. Please run the Database maintenance task using the Console interface (8001)	Run the database maintenance task to vacuum the database.
WARN	9	Tagged values were loaded, but full-	System metadata and tagged values were loaded successfully for	Contact Customer

RN 68	text index loading failed.	volume 'server:volume', but loading the full-text index failed (9068)	Support.
WARN 9070	Tagged values and full-text index loading failed.	Loading system metadata, tagged values and the full-text index failed for volume 'server:volume' (9070)	Contact Customer Support.
WARN 150003	The volume mount appeared to succeed, but the test for mount failed.	Volume <volume name> on server <server name> is not mounted. Skipping. (15003)	Contact Customer Support.
WARN 150004	A component cleanup failure on stop or completion.	[<component>] Cleanup failure on stop. (15004)	Contact Customer Support.
WARN 150005	There was a component run failure.	[<component>] Run failure. (15005)	Contact Customer Support.
WARN 150006	Cleanup failed for component after a run failure.	[<component>] Cleanup failure on abort. (15006)	Contact Customer Support.
WARN 150007	A component that is timed out needs to be stopped.	Component [<component>] unresponsive; autostopping triggered. (15007)	Try your action again. If this error continues, contact Customer Support.
WARN 150010	The same volume cannot be harvested in parallel. The harvest is skipped and the next one, if any are in queue, started.	Volume <volume-name> on server <server-name> is already being harvested. Skipping. (15010)	No user intervention is required. You might want to verify that the volume harvest is complete.
WARN 150011	A volume cannot be harvested if it is being used by another job. The harvest continues when the job is complete.	Volume <volume-name> on server <server-name> is being used by another job. Waiting before proceeding with load. (15011)	No user intervention is required.

WARN	15015	Configured harvest time limit is reached.	Time limit for harvest reached. Skipping Volume v1 on server s1.1 (15015)	Reconfigure harvest time limit.
WARN	15016	Configured harvest object count limit is reached.	Object count limit for harvest reached. Skipping Volume v1 on server s1 (15016)	Reconfigure harvest data object limit.
WARN	17008	Query that ran to discover Centera items ended unexpectedly.	Centera External Iterator : Centera Query terminated unexpectedly (<error description>). (17008)	Contact Customer Support.
WARN	17011	Running discovery on the same pool in parallel is not allowed.	Pool Jpool appears to have another discovery running. Skipping. (17011).	Make sure that two jobs are not running at the same time that discovers the same pool.
WARN	17502	Generic retention discovery is already running for this master volume.	Volume <server>: <share> appears to have another discovery running. Skipping.	No user intervention is required as the next step, if any, within the job is run.
WARN	17504	Sent when a retention discovery is run on any volume other than a Windows Share retention volume.	Volume <server>: <share> is not supported for discovery. Skipping.	Contact Customer Support.
WARN	18007	Directory listing or processing of data object failed in Grazer.	Walker._walktree: OSError - <path><reason> (18007)	Make sure that the appliance still has appropriate permissions to a volume. Verify that there is network connectivity between the appliance and your volume. Contact Customer Support.
WARN	18008	Unknown error occurred while processing data object or listing directory.	Walker._walktree: Unknown exception - <path>. (18008)	Contact Customer Support.
	1			

WARN	18009	Grazer timed out processing an object.	Walker._process File: Grazer Timed Out. (18009)	Contact Customer Support.
WARN	18010	The skipdirs file is either not present or not readable by root.	Unable to open skipdirs file: <filename>. Cannot skip directories as configured. (18010)	Contact Customer Support.
WARN	18011	An error occurred reading the known extensions list from the database.	Grazer._run: couldn't read extensions - <reason>. (18011)	Contact Customer Support.
WARN	18012	An unknown error occurred reading the known extensions list from the database.	Grazer._run: couldn't read extensions. (18012)	Contact Customer Support.
WARN	18015	NFS initialization warning that NIS is not available.	NIS Mapping not available. (18015)	User name and group names might be inaccurate. Check that your NIS server is available and properly configured in the data server.
WARN	18017	A folder in an Enterprise Vault case that is harvested was skipped because of insufficient permissions.	Skipping Folder (ID=3) in volume evdiscaccel.company.com: CaseOne due to insufficient permissions. Check Review permissions on the folder. (18017)	If all folders were expected to be harvested in the volume, verify that the user name used to add the Volume has Folder Review permission on all folders.
WARN	18019	The checkpoint that is saved from the last harvest of the NewsGator data source failed to load. Instead of conducting an incremental harvest, a full harvest is run.	Unable to load checkpoint for NewsGator volume. A full harvest will be performed instead. (18019)	If the message repeats in subsequent harvests, contact Customer Support.
WARN	18020	The checkpoint noted for the current harvest of the NewsGator data source might not be saved. The next incremental harvest of the data source is not able to pick up from this checkpoint.	Unable to save checkpoint for NewsGator harvest of volume. (18020)	If the message repeats in subsequent harvests, contact Customer Support.
				Server administrators

WARN 33016	System might not unmount this volume.	Windows Share Protocol Session teardown failed. (33016)	can see that connections are left hanging for a predefined time. These connections will drop off after they time out. No user intervention required.
WARN 33017	System encountered an error while it tries to figure out what query uses this volume.	An error occurred while retrieving the query instances pointing to a volume. (33017)	Contact Customer Support.
WARN 33021	The teardown operation of the connection to a IBM Information Archive volume failed. Some connections can be left open on the IBM Information Archive server until they are timed out.	IBM Information Archive tear-down failed. (33021)	None
WARN 33023	A connection to the Discovery Accelerator might not be torn down completely, and a few resources on the server might be tied up for a short while.	HTTP Connection tear-down to Discovery Accelerator failed. (33023)	None
WARN 33028	The tear-down operation of the connection to a FileNet volume failed. Some connections can be left open on the FileNet server until they time out.	IBM FileNet tear-down operation failed. (33028)	None
WARN 34003	Skipped a copy data object because disk full error.	Copy action error :- Target disk full, skipping copy : <source volume> to <target volume>. (34003)	Verify that there is space available on your policy destination and try again.
WARN 34010	Skipped a move data object because disk full error.	Move action error :- Target disk full, skipping copy : <source volume> to <target volume>. (34010)	Verify that there is space available on your policy destination. After verifying that space is available, run another harvest before you run your policy. Upon harvest completion, try running the policy again.
WAR 34002	Discovery export policy detects the target disk is full and skips production of an object.	Discovery export Run action error: Target disk full, skipping discovery export: `share-1/saved/years.pdf to production/10/documents/1`	Create sufficient space on target disk and run discovery export policy

N	9		/0x0866e 5d6c898d9ffdbea 720b0 90a6f46d3058605 .txt. (34029)	again.
W A R N	3 4 0 3 2	The policy that was run has no volumes in scope that is based on the configured query and scoping. The policy cannot be run.	No volumes in scope for policy. Skipping policy execution. (34032)	Check policy query and scoping configuration, and re-execute policy.
W A R N	3 4 0 3 3	Celerra data mover error. There are many possible causes; the actual one is listed in the error message. The sample provides the most common one.	Data mover returned NO_MATCHING_ CONNECTION : The specified OFFLINE_PATH was not able to map to an existing connection to a secondary server. (34033)	Consult the Celerra Administrator Manual.
W A R N	3 4 0 3 5	If the global hash setting for the system is set to not compute data object hash, no hash can be computed for the target objects during a policy action.	Copy objects : Target hash will not be computed because Hashing is disabled for system. (34035)	If target hashes need to be computed for the policy audit trail, turn on the global hash setting before you run the policy.
W A R N	3 4 0 3 6	The policy has no source volumes in scope, which means that the policy cannot be run.	The policy has no source volume(s) in scope. Wait for the query to update before executing the policy. (34036)	Confirm that the query used by the policy has one or more volumes in scope.
W A R N	4 2 0 0 3	The job in this action is stopped by the user.	Copy data objects stopped at user request. (42003)	No user intervention is required.
W A R N	4 2 0 0 8	The job in this action is stopped by the user.	Move data objects stopped at user request. (42008)	No user intervention is required.
W A R N	4 2 0 1 6	The job in this action is stopped by the user.	Delete data objects stopped at user request. (42016)	No user intervention is required.
W A	4 2 0	The job in this action is stopped by	Policy stopped at user	No user intervention is

R 2 N 6	the user.	request. (42026)	required.
W A R N 4 2 0 3 5	When the job in this action is stopped by the user.	Set security for data objects stopped at user request. (42035)	No user intervention is required.
W A R N 4 2 0 5 1	Two instances of the same policy cannot run at the same time.	Policy <policy-name> is already running. Skipping. (42051)	No user intervention is required.
W A R N 4 2 0 5 2	Policies cannot be run until after volumes are loaded, if those volumes are participants to the policy by virtue of being in the query.	One or more volume(s) needed by policy <policy-name> are being loaded. Waiting for the bulk load(s) to finish. (42052)	No user intervention is required.
W A R N 4 2 0 6 1	Discovery export policy was stopped by user.	Discovery export run on objects (Copying native objects) stopped at user request. (42061)	No user intervention is required.
W A R N 4 2 0 6 4	Discovery export policy execution is delayed because a conflicting discovery export run is in progress.	A Discovery export run related to policy 'Discovery export case One' is in progress. Waiting for it to finish. (42064)	The discovery export policy execution is held up for required resources. Execution must begin as soon as resource becomes available.
W A R N 4 2 0 6 8	Policy failed to set appropriate permissions on the target directory. Objects that are created from the policy might not have appropriate permissions set.	Copy objects warning, unable to set permissions on target directory: share-1/saved. (42068)	Policy might not be able to set appropriate permissions on the objects it creates. If it is not acceptable, verify that target volume has appropriate write permissions and re-execute.
W A R N 4 2 0 6 9	If the "Copy data objects modified since last harvest" option is selected for a discovery export policy, it is valid only if the discovery export itself is defined to act on the original file/email archive, as opposed to their members. If it is not true, the warning tells the user that modified	Discovery export DAT_Export is configured to act on members of containers, and cannot act on objects modified after the last harvest. Discovery export run X	If the modified objects need to be acted upon, either use a discovery export action only on the original file/email archive, or conduct an incremental harvest on

		objects are skipped.	will skip modified objects. (42069)	the source volumes.
WARN	46026	Volume is being harvested or policies are running against it. If there are other full-text indexes to be backed up, the system works on those actions. Try this volume again.	Volume volume:share is in use. Unable to back up full-text index. Will retry later. (46026)	Rerun backup when volume is not in use.
WARN	47201	Database connections are down to a normal level.	Database connections at normal level again (512/100) (47201)	No user intervention is required.
WARN	47202	The system is starting to run low on database connections. This situation is abnormal. An indication of process restarts and connections are not being cleared.	Database connections usage seems excessive (512/415) (47202)	Contact Customer Support.
WARN	47215	Someone internally or externally is trying (and failing) to SSH into the data server.	SSHD: Failed password for root from 172.17.18.185 port 57982. (47125)	Contact your local IT manager. It might be either a mistyped password by a legitimate user or in the worst case scenario, a genuine break-in attempt.
WARN	61003	One of the load files cannot be uploaded because the compute node might not be accessed to obtain.	Failed to mount transaction cache dump '/deepfs/postgres/pro-duction_cache'. (61003)	Some of the load files will be missing after the discovery export completes. These load files are reproduced on a new run. If problem persists across runs, Contact Customer Support.
WARN	61004	Warns the user that one of the transaction cache memory dump processes encountered an error. If a discovery export runs, it means that the discovery export fails to produce one of the load files.

Note: If multiple memory dumps fail, there is one warning per failed memory dump. | Transaction Cache Dump failed with error - Validation failed during creation of load file. (61004) | Run the discovery export policy that saw the error again. If the error persists, and you cannot find any cluster/data server configuration issues, contact Customer Support. |

- **Parent topic:**

Related reference

- [Event log messages](#)

Managing the IBM StoredIQ products

IBM® StoredIQ provide interfaces and tools to manage data in your system. The following topics describe in detail how to use these interfaces and tools to achieve your system goals.

- [Managing IBM StoredIQ Data Script](#)
  IBM StoredIQ Data Script provides an interface that you can manage data from the Document Dashboard. The following topics describe in detail how to work with the documents in IBM StoredIQ Data Script.
- [Managing IBM StoredIQ Data Workbench](#)
  IBM StoredIQ Data Workbench is a data management tool that helps you learn about data in your system and manage it for your business purposes.
- [Managing IBM StoredIQ eDiscovery](#)
  IBM StoredIQ eDiscovery captures and manages the electronic data in your system. IBM StoredIQ eDiscovery uses "matter" to create and organize information so that you can always easily access the data that you need and review the status or process of your work.
- [Managing IBM Stored IQ Policy Manager](#)
  IBM Stored IQ Policy Manager provides an interface for you to view available policies, create, edit, delete, or run a policy or a report.

Managing IBM StoredIQ Data Script

IBM® StoredIQ Data Script provides an interface that you can manage data from the Document Dashboard. The following topics describe in detail how to work with the documents in IBM StoredIQ Data Script.

- [Overview of IBM StoredIQ Data Script](#)
  IBM StoredIQ Data Script enables automated execution within IBM StoredIQ Platform. You can script, automate, and monitor processes that are normally run manually within IBM StoredIQ Data Workbench. IBM StoredIQ Data Script focuses on repeatable, understood, and approved processes for the purposes of culling and refining data in an approved manner.

- **Parent topic:** [Managing the IBM StoredIQ products](#)

Overview of IBM StoredIQ Data Script

IBM® StoredIQ Data Script enables automated execution within IBM StoredIQ Platform. You can script, automate, and monitor processes that are normally run manually within IBM StoredIQ Data Workbench. IBM StoredIQ Data Script focuses on repeatable, understood, and approved processes for the purposes of culling and refining data in an approved manner.

Through the IBM StoredIQ Data Script interface, you can monitor each of the data management steps as they are run. You can then view any defined outputs, such as reports, infosets, or exports that are generated as a result.

You can choose predefined workflows that run a well-defined set of steps. These steps mimic a process that a user usually runs manually within IBM StoredIQ Data Workbench. These defined steps, or workflows in IBM StoredIQ Data Script, are meant to address repeatable processes for data management. By running workflows, the user can reapply processes that have been reviewed and approved to apply to a wide variety of different data problems.

Workflows originate in IBM StoredIQ Data Workbench, by selecting Details > Infoset Ancestry, or from custom scripts. Infosets can be saved as workflows in the documents of IBM StoredIQ Data Script, the starting and end points for which are the first and last operations in the infoset ancestry.

- Document dashboard view
  The Document Dashboard provides high-level overviews of existing documents and their status within IBM StoredIQ Data Script.
- Workflow origination
  Workflows originate in IBM StoredIQ Data Workbench by selecting Details > Infoset Ancestry, or from custom scripts.
- Adding a document
  To manage data with IBM StoredIQ Data Script, you must first add a document to make it available within IBM StoredIQ Data Script.
- Searching for a document
  You can search existing documents from the Document Dashboard of .
- Viewing a document
  After you added documents in IBM Data Script, you can view them from the Document Dashboard.
- Document details
  The IBM Data Script interface provides the Document Detail page and the Workflow Status and Outputs page, which presents detailed information about documents and workflow status.
- Running a workflow
  You can run workflows on existing documents to modify your documents as needed.

- **Parent topic:** Managing IBM StoredIQ Data Script

Document dashboard view

The Document Dashboard provides high-level overviews of existing documents and their status within IBM® StoredIQ Data Script.

The Document Dashboard provides three critical pieces of information for each added document: the document's name, a brief description of that document, and the date and time when the document was created. From the Document Dashboard, you can do the following tasks:

- Add a document by clicking the Add Document button.
- View a document by double-clicking the document that you want to see.
- Search for a document by filtering the existing documents.

Related concepts

- [Document details](#)

Related tasks

- [Adding a document](#)
- [Searching for a document](#)
- [Viewing a document](#)

Workflow origination

Workflows originate in IBM® StoredIQ Data Workbench by selecting Details > Infoset Ancestry, or from custom scripts.

Infosets can be saved as workflows and used in IBM StoredIQ Data Script documents. They serve as the first and last operations in the infoset ancestry.

Related tasks

- [Running a workflow](#)

Adding a document

To manage data with IBM® StoredIQ Data Script, you must first add a document to make it available within IBM StoredIQ Data Script.

Procedure

1. From the Document Dashboard, click Add Document. The Create Document dialog box appears. The Create Document dialog box has three steps: Name and Description, Workflows, and Infosets.
2. Within the Name and Description subtab, complete these fields.
    a. In the Name text box, enter a unique name for this document.
    b. Optional: In the Description text box, enter a brief description of this document.
    c. Click Next to access the Workflows subtab.
3. Within the Workflows subtab, complete these fields.
    a. Use Add and Remove to apply the workflows to be available for this document.
    b. If needed, search for an available workflow in the Enter terms... text box and Search.
    c. Click Next to access the Infosets subtab.
4. Within the Infosets subtab, complete these fields.
    a. Select the infoset name to choose the starting infoset for this document.
    b. If needed, search for an available infoset in the Enter terms... text box and Search.
    c. Click Save to complete the process.
    The added document appears in the Document Dashboard.

Related concepts

- [Document dashboard view](#)

Searching for a document

You can search existing documents from the Document Dashboard of .

Procedure

1. View the Document Dashboard from IBM® Data Script.
2. In the Enter terms here.... text box, enter the name of the document for which you want to search, and then press Search. You can enter either the full document name or a portion of its name. Documents that match the entered search term are returned.
3. To remove the search term, click the X to the left of the Enter terms here.... text box.

Related concepts

- [Document dashboard view](#)

Viewing a document

After you added documents in IBM® Data Script, you can view them from the Document Dashboard.

Procedure

From the Document Dashboard, double-click the document that you want to view. The document opens within the Document Details page.

Related concepts

- [Document dashboard view](#)

Document details

The IBM® Data Script interface provides the Document Detail page and the Workflow Status and Outputs page, which presents detailed information about documents and workflow status.

Document details

The Document Detail page contains five distinct areas:

- Document Dashboard
- Workflows
- Infosets
- Actions
- Reports

Four icons are used to indicate status within the Document Detail page: Success, Fail, Processing, and Processing Not Yet Started.

- Success: This icon indicates that processing is finished and that the workflows are complete.
- Fail: This icon indicates that the workflow's processing failed.
- Processing: This icon indicates that the workflow is still processing.
- Processing Not Yet Started: This icon indicates that the workflow's processing is not yet started.

The Document Dashboard provides an abbreviated view of the selected document, including the document's name, its creation date and time, and a brief description. It provides the option to edit the document.

Workflow Status and Outputs

Workflows demonstrate the individual business processes of a larger procedure: collect data, collect custodians, run actions, generate a report, move to preservation, and export. The sequence of workflows can be reordered as needed to create different workflows, which then result in different outputs.

User infosets that are based on their ancestry can be created as workflows within IBM Data Script.

If no workflows appear in the Workflow Details area, you can add a workflow to a document by clicking Add Workflow.

The Workflow Details includes these fields.

- Name: Name of the workflow that was provided when the workflow was added.
- Description: Description of the workflow that was provided when the workflow was added.
- Started: Date and time when the workflow began processing the document.
- Overall Status: Information about the workflow's status, including which steps are still in progress and which steps are completed.

The status icons are used to provide at-a-glance views of the workflow's status. Infosets, actions, and reports are all outputs of a workflow; the artifacts are created as the result of the workflow's execution.

- Infosets: Provides high-level information about infosets and their status.
- Actions: Provides high-level information about actions and their completion status.
- Reports: provides high-level information about reports and the status of report generation.

Related concepts

- [Document dashboard view](#)

Running a workflow

You can run workflows on existing documents to modify your documents as needed.

Procedure

1. From the Workflow Status and Outputs page, click Run Workflow. The Run Workflow dialog box appears.
2. Select a workflow and click Next.
3. In the Run Workflow dialog box, complete these fields.
    - Enter the Base Name of this workflow in the Base Name field.
    - Enter a brief description of this workflow in the Base Description field.
    - Enter the email address to which notification of the workflow result is sent.
4. Click Next.
5. In the Starting Infoset window, select the button of the starting infoset for this workflow and then click Run. When the workflow runs successfully, a green check marks the process complete. Otherwise, a red x marks the process as failed.

Related concepts

- [Workflow origination](#)

Managing IBM StoredIQ Data Workbench

IBM® StoredIQ Data Workbench is a data management tool that helps you learn about data in your system and manage it for your business purposes.

- [Overview of IBM StoredIQ Data Workbench](#)
  IBM StoredIQ Data Workbench is a data visualization and management tool that helps you actively manage your company's data.
- [Infoset management](#)
  The infoset window provides the following high-level information for a selected infoset.
- [Filter administration](#)
  Filters are objects that can be applied to infosets. You can apply one filter with multiple filter attributes to an infoset, continually refining that infoset until it represents responsive data that you need.
- [Enhancements to infoset](#)
  The Enhance tab provides you with a visual layout of available enhancements. Enhancements are created as models within IBM StoredIQ Administrator. When an enhancement is applied to an infoset, it updates that infoset's index.
- [Actions on data](#)
  Actions are preconfigured options that help you to manage data objects that are contained within infosets.
- [Exploration of infosets and data](#)
  The Explore tab provides you with a visual layout of an infoset. At a glance, you can see how much data you have, its type, its overlays, and detailed information about specific data. By clicking within the data map itself, you can explore the data and analyze various details.
- [Exceptions](#)
  Exception lists contain various data objects such as files, emails, which encountered an error when an action was conducted on the infoset. This list helps you to know exactly what data needs attention due to some error or exception. It also serves as an informational trace of what went wrong during the process.
- [Filter condition reference and expression language](#)
  This appendix provides in-depth information about full-text and attribute conditions and how to use the expression language.
- [IBM Discovery Export EDRM XML](#)
  IBM EDRM XML is based on the specification that is described in the EDRM XML Interchange Format Schema Documentation, January 23, 2008. The appendix describes the discovery export feature for EDRM XML.
- [Metadata by attribute type](#)
  This reference guide describes metadata that is organized by attribute type.

- **Parent topic:** [Managing the IBM StoredIQ products](#)

Overview of IBM StoredIQ Data Workbench

IBM® StoredIQ Data Workbench is a data visualization and management tool that helps you actively manage your company's data.

IBM StoredIQ Data Workbench helps you find out how much data you have, where it is, who owns it, and when it was last used. When you have a clear understanding of your company's data landscape, IBM StoredIQ Data Workbench helps you take control of it and make informed decisions about your data and act on that knowledge.

Big data is a pervasive problem for the business. IBM StoredIQ Data Workbench is developed to help you learn about your data, make educated decisions about that data, and turn a dangerous risk into a valuable asset.

IBM StoredIQ Data Workbench can be used in the following ways:

- Find all emails from a specific sender in any time frame that you specify.
- Identify unused data that stays in the system for a long time.
- Search for specific format of files on the network for a defined time period. For example, search image files that were created in 2007, and see how much space these files take on the network.
- Provide a visual overview of how data is retained. Break down the data source so that forbidden information such as credit-card numbers or Social Security numbers can be identified.

- [Key concepts of IBM StoredIQ Data Workbench](#)
  Key concepts of IBM StoredIQ Data Workbench, such as infosets, filters, refinements, and indexes are explained. These concepts are key to understanding IBM StoredIQ Data Workbench as a whole.

Key concepts of IBM StoredIQ Data Workbench

Key concepts of IBM® StoredIQ Data Workbench, such as infosets, filters, refinements, and indexes are explained. These concepts are key to understanding IBM StoredIQ Data Workbench as a whole.

- Information set
  An information set, commonly referred to as an infoset, is a snapshot of data at a certain point in time. An infoset is consisted of one or many indexes. The data that is contained within each index is presented within a single user interface on which a user can act or conduct extra refinements. Infosets allow users to avoid the details of where the data is found and focus on what matters: the data itself.
- IBM StoredIQ Platform index
  Infosets are derived from theIBM StoredIQ Platform index of your data sources. Therefore, to take advantage of infosets, the administrator must first use IBM StoredIQ Platform to index the data sources. Contact your administrator for more information.
- Filter
  Filters are a mix of user-created and system-provided objects that can be applied to infosets. You can apply several filter attributes to an infoset, continually refining it until the infoset represents the exact amount of data that you need.
- Overlay
  Overlays are configurable filters that display hits or matches in a selected infoset.
- Scope operation
  Contained data, such as data within .ZIP, .TAR, or .PST files, is hierarchical and can have many different relationship and connections. Depending on how that data is viewed, that data can give a different perception than what is represented. The Scope Operations tab helps users to understand more fully what data is represented by data sets.
- Action
  An action is an activity that is created by an administrator and conducted on an infoset. Available IBM StoredIQ Data Workbench actions include copy, copy to retention, delete, discovery export, modify attribute, move, and snippet step-up.
- Report
  IBM StoredIQ Data Workbench reporting function provides external views of infosets and validates IBM StoredIQ processes. Reports can also be customized with the BIRT Report Designer.

- **Parent topic:** Overview of IBM StoredIQ Data Workbench

Related concepts

- Overview of IBM StoredIQ Data Workbench

Information set

An information set, commonly referred to as an infoset, is a snapshot of data at a certain point in time. An infoset is consisted of one or many indexes. The data that is contained within each index is presented within a single user interface on which a user can act or conduct extra refinements. Infosets allow users to avoid the details of where the data is found and focus on what matters: the data itself.

Infoset Basics

An infoset is a named, persistent object that any user can refer to by referencing its name. Consider these basic infoset concepts:

- Infosets are derived from data sources that are indexed by IBM® StoredIQ Platform.
- Infosets are a reference to the indexed data at the time that the infoset was created. Because an infoset is composed of that indexed data, it is more rightly viewed as a snapshot in time.
- After an infoset is selected, a list of data objects can then be viewed.
- Infosets can be referred to when you are conducting actions such as copying data.

Table 1. Frequently asked questions about infosets

Question	Answer
Can infosets be deleted?	Yes. User-created infosets and administrator-created system infosets can be deleted; however, the All Data Objects and All System-Level Objects system infosets cannot be deleted.
Are IBM indexes real-time indexes?	No. Indexes are not real-time indexes. They are representative of data only at the time it was harvested.
Can an infoset have more than one filter applied to it?	Filters can be made up of multiple attributes; however, only one filter can be applied to an infoset at a time. Keep in mind that after every application of a filter, a new infoset is created.

Types of infosets

There are two types of infosets within IBM StoredIQ Data Workbench: system and user.

Table 2. IBM StoredIQ Data Workbench Infoset Types

Information Set Type	Description
System	A system information set is an infoset that is generated and kept current by IBM StoredIQ Platform. All objects within that system are kept up to date with their current index. System infosets are the starting point for user information sets. - The All Data Objects and All System-Level Objects infosets are automatically generated by IBM StoredIQ Platform. Each time IBM StoredIQ Platform indexes a data source, system infosets are regenerated. - Cannot be edited by users and can be viewed by the administrator in the IBM StoredIQ Data Workbench user interface - The All Data Objects and All System-Level Objects system infosets provide users with a starting point upon which to base all other refinements. - Administrators can create system infosets that target certain volumes in the IBM StoredIQ Administrator application. These system infosets can be edited and deleted.

User | This infoset is generated by a user, which is called a user infoset, and is a snapshot in time of the underlying objects. The objects that are represented by the user infoset might change, but the infoset itself does not change.

Infosets are derived from any number of indexed data sources. They are also derived through the application of filters, which results in a refined infoset.

- **Parent topic:** [Key concepts of IBM StoredIQ Data Workbench](#)

Related concepts

- [Key concepts of IBM StoredIQ Data Workbench](#)

IBM StoredIQ Platform index

Infosets are derived from theIBM® StoredIQ Platform index of your data sources. Therefore, to take advantage of infosets, the administrator must first use IBM StoredIQ Platform to index the data sources. Contact your administrator for more information.

IBM StoredIQ Platform index types

Indexing or harvesting is the process or task by which IBM StoredIQ Platform examines and classifies data in your network. The result of this indexing process is your IBM StoredIQ Platform index, which is the basis for your work in IBM StoredIQ Data Workbench. These indexes are generated by the IBM StoredIQ data server, and each data server is configured to index data at specific network locations. Individual data servers are responsible for different sections of data that might be contained on different data sources such as Windows, Exchange, Sharepoint, Documentum, Domino, or Centera. This data can be at different network addresses or even different physical locations.

Note: The administrator configures and defines the index type. If you have questions about volume types, data sources, or harvests, contact your administrator.

Table 1. IBM StoredIQ Platform index types

IBM StoredIQ Platform Index Type	Description
Metadata Index	IBM StoredIQ Platform indexes contain all the information about the data at a specific location on a network, including descriptive information or attributes about the data such as file name, file size, created date, and owner.
Full-Text Index	IBM StoredIQ Platform can also generate a more detailed index on the contents of the data itself. By reading the contents of the data, the words or characters that are contained within the data can be referred to and searched against.

- **Parent topic:** [Key concepts of IBM StoredIQ Data Workbench](#)

Related concepts

- [Key concepts of IBM StoredIQ Data Workbench](#)

Filter

Filters are a mix of user-created and system-provided objects that can be applied to infosets. You can apply several filter attributes to an infoset, continually refining it until the infoset represents the exact amount of data that you need.

Filters are persistent objects that can be named and referenced. When a filter is applied to an infoset, you can refine the infoset into a smaller infoset. Like infosets, details about how the filters work are abstracted away from the user. The filter can create a raw expression. The underlying API then creates the appropriate queries to break down the infoset into its appropriate subset.

Filters can be reused across multiple infosets. The ability to replicate the process of filtering allows the users to replay workflow processes. In this fashion, business users or knowledge workers can capture refinement techniques that they might want to apply to various data.

Infoset example

You have a system infoset that contains all files, emails, and documents that pertain to all company employees, including a user that is named Josh Smith. Since this system infoset is large and unwieldy, you need to refine it so that you have two infosets. One infoset represents all of Josh Smith's email that contains an email subject of `stock options` and another that represents Josh's files that are larger than 1 GB.

Using this information, you can refine the system infoset by applying a filter with these attributes:

- Anything that Josh Smith owns. To find this information, apply a name filter where the owner is Josh Smith.
- Only Josh Smith's email. To find those emails, apply a file filter that looks for the .MSG extension.
- All of Josh Smith's emails that had an email subject of `stock options`. To find this kind of emails, apply an email filter that finds emails with a subject of `stock options`.

These filters and attributes give you the result that you want: All of Josh Smith's emails with an email Subject of Stock Options.

Additionally, you also need to create an infoset that represents Josh's files. These files are larger than 1 GB. Using the previous information, the same infoset represents all of the files and emails that are associated with Josh Smith. Therefore, you can again refine the infoset by applying a filter with these attributes:

- Anything that Josh Smith owns. To find this information, apply a name filter where the owner is Josh Smith.
- All of Josh Smith's files that are larger than 1 GB. To find these files, apply filter with an attribute of size larger than 1 GB.

The application of that last filter attributes is what gives you the result that you want: All of Josh Smith's Files Larger than 1 GB.

- **Parent topic:** [Key concepts of IBM StoredIQ Data Workbench](#)

Related concepts

- [Key concepts of IBM StoredIQ Data Workbench](#)

Overlay

Overlays are configurable filters that display hits or matches in a selected infoset.

Within the data map, color intensities for data objects that match the overlay change. The greater the overlay matches, the more red that tile appears within the data map.

- **Parent topic:** [Key concepts of IBM StoredIQ Data Workbench](#)

Related concepts

- [Key concepts of IBM StoredIQ Data Workbench](#)

Scope operation

Contained data, such as data within .ZIP, .TAR, or .PST files, is hierarchical and can have many different relationship and connections. Depending on how that data is viewed, that data can give a different perception than what is represented. The Scope Operations tab helps users to understand more fully what data is represented by data sets.

Within scope operations, you can conduct expansion or collapse operations.

- In an expand operation, all files within an infoset are expanded, causing the creation of a new, potentially more accurate infoset.
  Note: If the files within the infoset are not container files, then the Expand operator has no effect.
- In a collapse operation, all opened or expanded files within an infoset are collapsed, causing the creation of a new, smaller infoset.
  Note: If the files within the infoset are not container files, then the Collapse operator has no effect.

- **Parent topic:** [Key concepts of IBM StoredIQ Data Workbench](#)

Related concepts

- [Key concepts of IBM StoredIQ Data Workbench](#)

Action

An action is an activity that is created by an administrator and conducted on an infoset. Available IBM® StoredIQ Data Workbench actions include copy, copy to retention, delete, discovery export, modify attribute, move, and snippet step-up.

Although actions are the only object to touch live data, they do not alter infosets. An infoset is a grouping of data, not the data itself, and actions are applied to the actual objects, not the infoset. When you copy, you copy the actual file, not the infoset. The same is true for copying to retention or discovery export. Additionally, actions can be scheduled to run immediately or at a predetermined time and date.

- **Parent topic:** Key concepts of IBM StoredIQ Data Workbench

Related concepts

- Key concepts of IBM StoredIQ Data Workbench

Report

IBM® StoredIQ Data Workbench reporting function provides external views of infosets and validates IBM StoredIQ processes. Reports can also be customized with the BIRT Report Designer.

You can share the information that is contained within infosets with the reporting component, which allows infosets to be transferred to other media types for review and analysis. These reports do not affect existing infosets, but provide you with more usable formats in which to understand the files and data that is captured by an infoset.

Reporting is a key step within the data-management process as it validates that processes were completed correctly within IBM StoredIQ. You can customize reports in any of these scenarios:

- Modify reports to carry your organization's custom styles and logos, aligning them with other organization-based artifacts and documentation.
- Alter the format of the content reported in existing reports. For example, you can add more columns, switch axes in a graph, or change the units for some values.
- Design reports to contain information that is not found in other, existing reports

IBM StoredIQ Data Workbench contains these types of reports:

- Compliance Report
- Data Assessment Report
- Data Topology Report
- Overlay Hit Report
- CSV Exception List Export Report
- CSV Data Object Export Report
- Duplication Summary Report

- **Parent topic:** [Key concepts of IBM StoredIQ Data Workbench](#)

Related concepts

- [Key concepts of IBM StoredIQ Data Workbench](#)

Infoset management

The infoset window provides the following high-level information for a selected infoset.

Table 1. Infoset heading names and descriptions

Infoset Column Name	Description
Name	This list reflects the name of the selected information set. An infoset must be selected before you can view detailed information, conduct analyses, or act on it.
Total objects	The number of data objects that are contained within the infoset. If you place the mouse over the number of total objects, a window appears. It shows the number of top-level and child-level objects that are contained within the total number of objects. Note: If the infoset is still being created, only the Details window is active. Infosets that are still being created list 0 in this column.
Infoset size	The size of the infoset.
Composition	The composition of the data that is contained within an infoset, either Mixed or Top Level.
Created	The date and time at which the infoset was created. Note: System infosets are listed as N/A in this field, and user infosets list the creation date, even if their creation status shows pending.
Type	The type of infoset, either User or System.
Access	This option lists whether the infoset is Public or Private or the name of person or number of people who have access to the infoset. If you place your mouse over the number of people, a tooltip appears with the names of those users who have access. If you click the link, a window with the users' names and roles appears.
Creator	The person that originally created the infoset. For older infosets, the creator is listed as N/A.
Description	A brief description of the infoset, which was provided during infoset creation.

- [Creating an infoset](#)
  Infosets can be created to include the data map and its function, or they can exclude data maps, resulting in a faster creation time. Additionally, you can interrupt infosets during their creation.
- [Interrupting infoset creation](#)
  Infoset creation can sometimes exceed the time that you have available. Infoset interruption makes it possible for you to opt out of the creation process.
- [Searching for a specific infoset](#)
  When you have multiple infosets available to choose, you can search for a specific infoset instead of using the slider to move through available options.
- [Creating an overlay](#)
  Created filters appear as overlay options during infoset creation.
- [Editing an infoset](#)
  When you are editing an existing infoset, you can modify only the infoset's name and description.
- [Deleting an infoset](#)
  Infosets can be deleted if they are interrupted during their creation.
- [View of the infoset details](#)
  You can use the Details tab to view detailed information about a selected infoset and any updates that are made to the data sources uses by the infoset.

Creating an infoset

Infosets can be created to include the data map and its function, or they can exclude data maps, resulting in a faster creation time. Additionally, you can interrupt infosets during their creation.

About this task

Infosets are created from other existing system infosets.

To create an infoset from an existing infoset:

Procedure

1. Select the infoset from which you want to create a new infoset. It can be either a system or user infoset. The name of the selected infoset appears in the infoset window.
2. Click Create, and then select a filter in the Name column in the Library tab.
   Note: If this creation is your first time to create an infoset, you must create a custom filter. See Creating a Custom Filter.
3. Click Create New Infoset. The Create New Infoset dialog box appears.
4. In the New Infoset Name text box, enter the name of this infoset.
5. In the Description text box, enter the description of this infoset.
6. Select or clear the Include Data Map check box.
   - If you select this check box, the data map is generated for this infoset. The infoset takes longer to generate, but you are able to use the data map and its function for this infoset.
   - If you clear this check box, the infoset generates more quickly, but it does not include the data map.
     Note: If you opt to create an infoset without a data map, some of the reports that can be run against that infoset are affected. If the Compliance, Data Assessment, Data Topology, Overlay, and Customized Reports are run against an infoset without a data map, they contain no data.
7. Click Select overlays for this infoset to determine which overlays must be applied to this infoset. When you are exploring this infoset, these selected overlays are available to be applied within the data map.
   Note: If the data map is not generated, then overlays cannot be selected.
8. To send a notification email that the infoset was created, select Send email when the process is complete check box and enter the email address at which the individual must be notified.
9. Click Save and View to go to the Details/Infoset Ancestry tab, or click Save to begin infoset creation and close the dialog box.

- **Parent topic:** Infoset management

Related concepts

- Infoset management

Interrupting infoset creation

Infoset creation can sometimes exceed the time that you have available. Infoset interruption makes it possible for you to opt out of the creation process.

About this task

Infoset creation contains different forms. In some instances, you might want to interrupt the infoset's creation, preserve what was found, and then log the unfinished work as an exception. In other instances, you might want to halt infoset creation and then delete that infoset altogether and deleting any artifacts from the creation process.

- After an interrupted and partially created infoset was moved from Created > Interrupted > Completed, you can use that infoset. However, you must keep in mind that the completion state depends on when the infoset is stopped.
- When you are deleting an infoset partway through its creation, click Delete. When an infoset is deleted, it is no longer available within the user interface.
- If an infoset is processing, a Stop Processing link is available within the dashboard. During this processing state, this link can be used to interrupt its creation and access the Stop Processing confirmation window.

To stop infoset creation:

Procedure

1. During infoset creation, in the Create New Infoset dialog box, click Save and View or Save.
   Note: In the Total objects and Infoset size columns, the infoset is listed as Pending.
2. To stop or interrupt infoset creation, you can do either of the following steps:
   - Select the pending infoset, and then click Stop.
   - If the pending infoset is active, click Stop Processing.
3. Click OK to verify your action.
   Note: The infoset's creation is now changed to Interrupting. It then moves to a Completed state.
   Note: If you opt to use this interrupted infoset, it is only a partial infoset as its creation was interrupted.
4. Explore or delete your infoset.
   - Click Delete, deleting the infoset from the user interface and the infoset library as a whole.
   - Select the infoset and explore its content just as you do with any other infoset. You can refine its content, apply actions, and view exceptions on interrupted infosets.

- **Parent topic:** [Infoset management](#)

Related concepts

- [Infoset management](#)

Searching for a specific infoset

When you have multiple infosets available to choose, you can search for a specific infoset instead of using the slider to move through available options.

Procedure

1. In the infoset window, click the down arrow. A complete listing of all available infosets appears.
2. In the Enter key terms to search.... text box, enter the name of the infoset for which you want to search, and then press Search. You can enter either the full infoset name or a portion of it. To remove the search term, click the X to the left of the Enter key terms to search.... text box. Infosets that match the entered search term are returned.
3. Click the infoset that you want to select, and then click Select Infoset. The selected infoset appears at the top of the infoset window.

- **Parent topic:** [Infoset management](#)

Related concepts

- [Infoset management](#)

Creating an overlay

Created filters appear as overlay options during infoset creation.

Procedure

1. Select the infoset for which you want to create an overlay. The name of the selected infoset appears within the infoset window.
2. Click Create, and then click Build Filter.
3. Create and save a filter.
4. Click Create New Infoset and name the newly created infoset.
   Note: Saved filters can be used as overlays for multiple infosets.
   Note: The data map must be enabled for overlays to be added.
5. Select an overlay or several overlays from the overlay list, and then click either Save and View or Save.

- **Parent topic:** [Infoset management](#)

Related concepts

- [Infoset management](#)

Editing an infoset

When you are editing an existing infoset, you can modify only the infoset's name and description.

Procedure

1. In the Infoset window, use the down arrow to select the infoset that you want to edit.
2. Click Edit, which is in the upper right corner of the infoset window.
   Note: Because system infosets cannot be edited, this Edit link is disabled if a system infoset is selected within the infoset window.
   The Edit Infoset dialog box appears.
3. In the Edit Infoset dialog box, you can edit only the infoset's name and description. Click within either the Name or Description text boxes to edit the infoset.
4. Click Save to save your modifications.

- **Parent topic:** [Infoset management](#)

Related concepts

- [Infoset management](#)

Deleting an infoset

Infosets can be deleted if they are interrupted during their creation.

About this task

Infosets can be deleted in both the Completed and Pending states. An infoset cannot be deleted under these circumstances:

- If an infoset was created with another infoset as its source, then the child infoset must be deleted before you delete the parent infoset.
- If an action was run or is running against the infoset, then the infoset cannot be deleted.

To delete an infoset:

Procedure

1. In the Infoset window, use the down arrow to select the infoset you want to delete.
2. Click Delete, and then click OK.
   Note: Deleting an infoset permanently removes its history.

- **Parent topic:** Infoset management

Related concepts

- Infoset management

View of the infoset details

You can use the Details tab to view detailed information about a selected infoset and any updates that are made to the data sources uses by the infoset.

There are three tabs within the Details area: Infoset Ancestry, Execution Log, and Data Objects.

The Infoset Ancestry tab shows the history and genealogy of the selected infoset. For example, it displays the system infoset from which the user infoset was created and which filter was used.

Additionally, the Infoset Details section provides the following information about a selected infoset.

Table 1. Infoset Ancestry Column Definitions

Infoset Ancestry Column Heading	Definition
Name	The name of the item selected.
Created	The date and time the selected item was created.
Type	The item's type, either system or user infoset, filter.
Description	The description of the selected item.

- Viewing infoset ancestry
  View infoset ancestry from the Details window.
- Saving infoset ancestry as a workflow
  Infosets can be saved as workflows in IBM® StoredIQ Data Script documents, the starting and end points for which are the first and last operations in the infoset's ancestry.
- Execution log
  When the Execution Log tab is selected, you can see the name, status, the run start time, and the length of time of an action ran.
- Viewing action log of infosets
  By selecting an action in the Name column, the Action Details section provides information about that action, including name, description, status, and which infosets it was last used on.
- Generating an audit report
  Audit reports can be generated for actions selected within the Execution Log.
- Viewing data objects
  When the Data Objects tab is selected, the data object list returns objects in the infoset. You can also spot-check data objects within an infoset with the data object viewer.

- **Parent topic:** Infoset management

Related concepts

- Infoset management

Viewing infoset ancestry

View infoset ancestry from the Details window.

Procedure

1. Verify that you selected the infoset for which you want to view its ancestry.
2. In the Details window, click the Infoset Ancestry tab. This tab is selected by default.
3. In the Name column, click the name of the item for which you want to view ancestry information. The Ancestor Infoset Details area populates with the name, description, total objects, top-level objects, child objects, infoset size, composition, created date, type, and creator. The Ancestor Filter Details area displays name, description, a link to view the filter code, created date, and type.

- **Parent topic:** View of the infoset details

Related concepts

- View of the infoset details

Saving infoset ancestry as a workflow

Infosets can be saved as workflows in IBM® StoredIQ Data Script documents, the starting and end points for which are the first and last operations in the infoset's ancestry.

Procedure

1. In the Details window, click the Infoset Ancestry tab. This tab is selected by default.
2. Click Save Ancestry as Workflow. The Create Workflow dialog box appears.
3. In the Create Workflow dialog box, complete these fields.
    - In the Name text box, enter a unique name for this workflow.
    - In the Description text box, enter a description of this workflow.
4. Click Save. After this ancestry is saved as a workflow, it is available in other related applications such as IBM StoredIQ Data Script.

- **Parent topic:** View of the infoset details

Related concepts

- View of the infoset details

Execution log

When the Execution Log tab is selected, you can see the name, status, the run start time, and the length of time of an action ran.

The Execution Log column is defined as follows.

Table 1. Execution Log Column Definitions

Execution Log Column Heading	Definition
Name	The name of the action on the selected infoset.
Status	The status of the action on the selected infoset.
Type	The type of action being run on the selected infoset.
Run started	The time at which the action was started. By clicking the arrow to the right of the column name, you can list the column in ascending or descending order.
Run duration	The length of run time of the selected action.

- **Parent topic:** View of the infoset details

Related concepts

- View of the infoset details

Viewing action log of infosets

By selecting an action in the Name column, the Action Details section provides information about that action, including name, description, status, and which infosets it was last used on.

Procedure

1. Verify that you selected the infoset for which you want to view its actions log.
2. In the Details window, click the Execution Log tab.
3. In the Name column, click the name of the action for which you want to view log information. The Action Details area populates, displaying name, description, status, the infoset on which the action was last used, any exceptions, and the Generate Report link. If you click the exceptions percentage, the Exceptions tab appears, providing you with greater context and detail.
   - When the Generate Report link is clicked and an audit report is generated, a Source Action detail becomes available, showing the report's name, the date and time it was created, and the length of time necessary to generate the report. If the report is being generated, its status is shown as Running.
   - When the report has finished generating, it can be viewed by clicking the Download Report link within Status.

- **Parent topic:** [View of the infoset details](#)

Related concepts

- [View of the infoset details](#)

Generating an audit report

Audit reports can be generated for actions selected within the Execution Log.

About this task

Action Audit Report
    The Action Audit Report provides detailed information in these areas:

- Execution Summary
- Data Object Result Count Summary
- Infoset Summary
- Participating Data Server and Volume Summary
- Action Parameters

Procedure

1. Verify that you selected the infoset for which you want to generate an audit report, and then click Details > Execution Log.
2. In the Action Details area, click Generate Report. The Generate Audit Report dialog box appears.
3. In the Report Name text box, the report's name is prefilled, based on the selected action in the Execution Log.
4. In the Send To: text box, enter the email addresses to whom this report is sent. Separate multiple email addresses with a comma.
5. Click Generate, and the audit report generates. When the audit report is generating, its status changes to Running in both the Action Details area and the Execution Log. When the report is generated, it is sent to the email addresses entered in the Send To: text box, and it can be opened from that email. The report can also be downloaded within the Execution Log.

- **Parent topic:** [View of the infoset details](#)

Viewing data objects

When the Data Objects tab is selected, the data object list returns objects in the infoset. You can also spot-check data objects within an infoset with the data object viewer.

About this task

Note: By clicking a data object, you can view the object itself using the data object viewer.

Table 1. Data objects default column definitions

Data Objects Column Heading	Definition
File name	The name of the file within the elected infoset.
File size	The size of the file.
File path	The directory path of the file.
Created	The creation date and time of the file.
Last modified	The date and time at which the file was last modified.

Procedure

1. In the Details window, click theData Objects tab. Data objects within the infoset are returned; you can see the data objects that are contained within the infoset.
2. Click an object's file name to view and verify that object in the data object viewer.
3. Click Close to close the data object viewer.
4. Configure columns when you are reviewing data objects:
    a. In the Details window, click the Data Objects tab.
    b. Click the blue gear icon, and the appears.
    c. In the Configure Column - Data Objects dialog box, click the down arrow to the right of options 2-6, selecting the name of the column. Choose from these options: None, File size, File path, File type, Container path, Last accessed, Created, Last modified, Owner name, or Group name. The File name option cannot be changed. By default, it is the name of the first column when you are viewing data objects. Additionally, as you rename columns, more columns are automatically added to the Configure Column - Data Objects dialog box.
    d. To reset the columns to the default selections, click Reset to default column selections. The column configuration options return to their default settings.
    e. Click Save. The layout of the Data Objects area reflects the chosen options.

- **Parent topic:** [View of the infoset details](#)

Related concepts

- [View of the infoset details](#)

Filter administration

Filters are objects that can be applied to infosets. You can apply one filter with multiple filter attributes to an infoset, continually refining that infoset until it represents responsive data that you need.

Note: Attributes within a filter category are ORed. For example, when you are creating a file filter that is defined by file name and extension contains .DOC, the filter syntax is `"file name OR file extension contains doc"`. Adding another file attribute category, the categories are ANDed. Therefore, the file category filter attribute `"extension OR file name contains doc"` and adding the name category with `"owner name John Smith,"` the syntax is `"file name"` OR `"extension contains doc"` AND `"owner names contains John Smith."`

There are four tabs on the Create page: Library, Build Filter, Set Operations, and Scope Operations.

- [Library of available filters](#)
  The Library tab lists all filters that are created and available for use with infosets.

Library of available filters

The Library tab lists all filters that are created and available for use with infosets.

Within the Library tab, you can view the names and descriptions of all saved filters. Additionally, you can edit and delete those filters.

- [Searching for an existing filter](#)
  Search for an existing filter within the Library tab.
- [Editing a filter](#)
  After filters are created and saved, they can be edited.
- [Viewing filter code](#)
- [Deleting a filter](#)
  Filters can be deleted when they are no longer used by an infoset.
- [Building a filter](#)
  Create a filter to customize your search for specific data.
- [Previewing filter results](#)
  You can preview sample, returned data before you save that infoset.
- [Configuring columns when previewing filter results](#)
  When you preview data objects that are contained within an infoset, you can select what type of information is displayed, ensuring that the returned data is meaningful to you.
- [Set operations](#)
  A set operation allows infosets to be combined in different ways to produce another infoset.
- [Scope operations](#)
  Contained data, such as data within .ZIP, .TAR, or .PST files, is hierarchical and can have many different relationship and connections. Depending on how that data is viewed, that data can give a different perception than what is represented. The Scope Operations tab helps users to understand more fully what data is represented by data sets.

- **Parent topic:** [Filter administration](#)

Related concepts

- [Filter administration](#)
- [Set operations](#)
- [Scope operations](#)

Related tasks

- [Searching for an existing filter](#)
- [Editing a filter](#)
- [Viewing filter code](#)
- [Deleting a filter](#)
- [Building a filter](#)
- [Previewing filter results](#)
- [Configuring columns when previewing filter results](#)

Searching for an existing filter

Search for an existing filter within the Library tab.

Procedure

1. Click Create, and then click the Library tab.
2. In the Enter key terms here.... text box, enter the name of the filter for which you want to search, and then press Search. You can enter either the full filter name or a portion of it.
   - To remove the search term, click the X to the left of the Enter key terms here.... text box.

   Filters that match the entered search term are returned. The Filter Details area displays the filter's name, description, a link to view the filter's code, and the name of the infoset in which the filter was last used.

- **Parent topic:** [Library of available filters](#)

- **Parent topic:** [Building a filter](#)

Related concepts

- [Library of available filters](#)

Related tasks

- [Building a filter](#)

Editing a filter

After filters are created and saved, they can be edited.

Procedure

1. Click Create > Library, and then select the name of the filter that you want to edit.
2. Click Edit. The Edit Filter dialog box appears.
3. Edit the filter as needed, determining the filter's scope and validating the code.
4. Click Save to save your changes.

- **Parent topic:** [Library of available filters](#)

Related concepts

- [Library of available filters](#)

Viewing filter code

Procedure

1. Within the Filter Details area, click View Filter Code. The Filter Code window appears.
2. Click Close to close the Filter Code window.

- **Parent topic:** [Library of available filters](#)

Related concepts

- [Library of available filters](#)

Deleting a filter

Filters can be deleted when they are no longer used by an infoset.

Procedure

1. Click Create > Library, and then select the name of the filter that you want to delete.
2. Click Delete. The Filter Delete Confirmation dialog appears.
3. Click OK to delete the filter, or click Cancel to cancel the deletion. If the filter is deleted, it is removed from the list of available filters within the Library tab.

- **Parent topic:** [Library of available filters](#)

Related concepts

- [Library of available filters](#)

Building a filter

Create a filter to customize your search for specific data.

Procedure

1. Click Create, and then click the Build Filter tab.
2. Select the check box or boxes of the type of filter attribute that you want to create. Depending on the type of filter you are creating, select from these options: File, Date,Size, Name, Email, Full Text, and Auto-Classification.

Option	Description
File	File filters are based on common file fields such as file name, file path, container path, and extension.
Date	Date filters are based on common date information such accessed, modified, or created dates.
Size	Size filters are based on common data-object file size field such as bytes, KB, MB, and GB.
Name	Name filters are based on name parameters such as the owner name or SharePoint owner.
Email	Email filters are based on common email fields such as To, From, CC.
Full Text	Full-text filters are based on a simple search string against all data objects. This option returns results only for full-text harvests.
Auto-classify	Auto-classify filters are based on the step-up snippet action. They can be run on an infoset, extracting and storing text for any data object to which it is applied, and the auto-classification enhancement.

3. In the Filter Details area, enter a name and description for the newly created filter, and then click Save Filter. The newly created filter can be seen within the Library tab.

- [Creating a file filter](#)
  File filters are a type of customized filters. File filters are based on common file fields such as name, extension, and path.
- [Creating a date filter](#)
  Date filters are based on common date information such as accessed, modified, or created dates.
- [Creating a size filter](#)
  Size filters are based on common data-object file size field such as bytes, KB, MB, and GB.
- [Creating a name filter](#)
  Name filters are based on name parameters such as the owner name or SharePoint owner.
- [Creating an email filter](#)
  Email filters are based on common email fields such as To, From, CC, BCC, and email subject.
- [Creating a full-text filter](#)
  Full-text filters are based on a simple search string for full-text indexed objects. Only objects that were full-text indexed return results.
- [Creating an auto-classification filter](#)
  Auto-classification filters are based on the auto-classification model that is created in IBM® StoredIQ Administrator.
- [Searching for an existing filter](#)
  Search for an existing filter within the Library tab.
- [Creating customized searches with the code view](#)
  Use the code view to create customized searches with search syntax.

- **Parent topic:** [Library of available filters](#)

Related concepts

- [Library of available filters](#)

Creating a file filter

File filters are a type of customized filters. File filters are based on common file fields such as name, extension, and path.

Procedure

1. Select the infoset and then click Select Infoset. The selected infoset appears in blue.
2. Click Create, and then click the Build Filter tab.
3. Select the File check box, and then complete following steps.
   a. In the Contains list, select either Contains, Is, or Begins with.
   b. In the text box, enter the text.
   c. Select the check box for what the particular filter uses, choosing from File Name, Extension, File Path, or Container Path. You can select any or all of the file type filter options.
4. In the Filter Details area, complete the following steps.
   a. In the Filter Name text box, enter the name of this new filter.
   b. In the Description text box, enter a brief description of this filter.
   c. Click Save Filter.

Results

The custom filter is created and saved, and it can now be accessed from the Library tab.

- **Parent topic:** Building a filter

Related tasks

- Building a filter

Creating a date filter

Date filters are based on common date information such as accessed, modified, or created dates.

Procedure

1. Click Create, and then click the Build Filter tab.
2. Select the Date check box, and then complete following steps:
   a. In the list of operators, select After, Before, On, or Between.
   b. In the Date text box, enter the date for this custom filter.
   c. Select the check box as to what fields to which the date filter applies, choosing from Accessed, Modified, or Created. You can select any or all of the date-type filter options.
   d. In the Filter Details window, enter the name of this new filter.
   e. In the Description text box, enter a brief description of this filter.
   f. Click Save Filter.

Results

The custom filter is created and saved, and it can now be accessed from the Library tab.

- **Parent topic:** [Building a filter](#)

Related tasks

- [Building a filter](#)

Creating a size filter

Size filters are based on common data-object file size field such as bytes, KB, MB, and GB.

Procedure

1. Click Create, and then click the Build Filter tab.
2. Select the Size check box, and then complete following steps:
   a. In the list of operators, select More Than, Less Than, Is, or Between.
   b. In the empty text box between the lists, enter the number that applies to this custom size filter.
   c. In the size value list, select bytes, KB, MB, or GB.
   d. In the Filter Details window, enter the name of this new filter.
   e. In the Description text box, enter a brief description of this filter.
   f. Click Save Filter.

Results

The custom filter is created and saved, and it can now be accessed from the Library tab.

- **Parent topic:** [Building a filter](#)

Related tasks

- [Building a filter](#)

Creating a name filter

Name filters are based on name parameters such as the owner name or SharePoint owner.

Procedure

1. Click Create, and then click the Build Filter tab.
2. Select the Name check box, and then complete following steps:
   a. In the blank text box, enter the owner or SharePoint owner name.
      Note: If you are entering more than one name, separate the names with commas. For example, you would enter multiple names as `John Smith, Mary Jones, Ellen Sideways`.
   b. Select the check box as to what fields to which this name filter applies, choosing from Owner or SharePoint Owner. You can select any or both of the name-type filter options.
   c. In the Filter Details window, enter the name of this new filter.
   d. In the Description text box, enter a brief description of this filter.
   e. Click Save Filter.

Results

The custom filter is created and saved, and it can now be accessed from the Library tab.

- **Parent topic:** Building a filter

Related tasks

- Building a filter

Creating an email filter

Email filters are based on common email fields such as To, From, CC, BCC, and email subject.

Procedure

1. Click Create, and then click the Build Filter tab.
2. Select the Email check box, and then complete these options.
   a. In the blank text box, enter the names or email addresses that is used by this filter. Separate multiple names or email address with commas. You might enter `John Smith`, `Mary Jones` or `jsmith@company.com`, `mjones@company.com`.
   b. Select the check box or boxes to which this email filter applies, choosing from To, From, CC, BCC, or Email Subject.
   Note: If you select the Email Subject check box, you must also select an operator: Contains, Is, or Begins With, and then enter some text about the email's subject into the empty text box.
3. In the Filter Details area, complete the following steps.
   a. In the Name text box, enter the name of this new filter.
   b. In the Description text box, enter a brief description of this filter.
   c. Click Save Filter.

Results

The custom filter is created and saved, and it can now be accessed from the Library tab.

- **Parent topic:** [Building a filter](#)

Related tasks

- [Building a filter](#)

Creating a full-text filter

Full-text filters are based on a simple search string for full-text indexed objects. Only objects that were full-text indexed return results.

Procedure

1. Click Create, and then click the Build Filter tab.
2. Select the Full Text check box.
3. In the blank text box, enter the search string that is run against data objects in a full-text index. You might enter `stock options AND CEO` or `Apple NOT IBM`.
4. In the Filter Details area, complete the following steps.
   a. In the Name text box, enter the name of this new filter.
   b. In the Description text box, enter a brief description of this filter.
   c. Click Save Filter.

Results

The custom filter is created and saved, and it can now be accessed from the Library tab.

- **Parent topic:** [Building a filter](#)

Related tasks

- [Building a filter](#)

Creating an auto-classification filter

Auto-classification filters are based on the auto-classification model that is created in IBM®
StoredIQ Administrator.

Procedure

1. Click Create, and then click the Build Filter tab.
2. Select the Auto-classification check box.
3. What appears in this area is dependent upon the model that is created in IBM StoredIQ
   Administrator.
    a. Select the check box that you want: And the highest score in the selected category
       or And where score is. Then, use the lists to select Greater than, Less than, or Equal
       to and the appropriate decimal amount.
4. In the Filter Details area, complete the following steps:
    a. In the Name text box, enter the name of this new filter.
    b. In the Description text box, enter a brief description of this filter.
    c. Click Save Filter. The Save Created Filter confirmation appears, verifying that you
       want to save this filter. Click OK.

- **Parent topic:** [Building a filter](#)

Related tasks

- [Building a filter](#)

Creating customized searches with the code view

Use the code view to create customized searches with search syntax.

Procedure

1. Click Create, and then click the Build Filter tab.
2. Click Code View.
   Note: To return to the filter's form view on the Build Filter tab, click Filter Form View. Then, you can create and save filters with the prebuilt forms.
3. In the large text box, create and enter a customized search with correct filter syntax. For more information about appropriate filter syntax, including tips and example, see [Syntax Primer](#).
4. Click Validate Code. If the entered expression is valid, a message that statesThe filter expression passed validation is returned. If it is invalid, a message that statesThe filter expression is not valid is returned.
5. In the Filter Details area, complete the following steps.
   a. In the Name text box, enter the name of this new filter.
   b. In the Description text box, enter a brief description of this filter.
   c. Click Save Filter.

- [Syntax primer](#)
  This primer offers syntax examples and tips for using Boolean operators, multiple terms or term phrases, wildcard, and proximity filters for the full-text filter or full-text source code filter.

- **Parent topic:** [Building a filter](#)

Related tasks

- [Building a filter](#)

Syntax primer

This primer offers syntax examples and tips for using Boolean operators, multiple terms or term phrases, wildcard, and proximity filters for the full-text filter or full-text source code filter.

Note: Saved filters are available as overlays when you are creating new infosets. For more information, see [Creating Infosets](#).

Boolean operator syntax

AND	The `AND` operator ensures that both terms appear in the data object. For example, the query `chairman AND stock` returns data objects where `chairman` and `stock` appear. `&&` functions the same as `AND`.
OR	The `OR` operator returns data objects where either term appears. The query `chairman OR stock` returns data objects that have `chairman`, `stock`, or both in the data object. `||` functions the same as `OR`.
NOT	The `NOT` operator ensures that a term does not appear in any data objects returned. The query `Apple NOT IBM` returns data objects where word `Apple` appears, but not `IBM`. `!` functions the same as `NOT`.

Multiple terms or term phrase syntax

`" "` and `' '`	Quotation marks around a term phrase return data objects that contain that phrase in that order. For example, "`September 20 1998`" returns only data objects with the exact phrase of `September 20 1998`. Double quotation marks allow for wildcards in terms whereas single quotation marks do not.
Multi-item, no Boolean operators	A search for multiple words without Boolean operators or quotation marks is an implied `OR` search. For example, `September 20 1998` returns data objects that have `September` OR `20` OR `1998`. Stop words are removed.
Parentheses	- Grouping with Boolean operators and parentheses controls how operators are evaluated. For example, `($10 OR billion) AND options` returns data objects where `$10` or `billion` might be in the data object, but `options` must also exist. - Boolean operators (`AND, OR, NOT`) work in a grouping or in a proximity group. So `("Sahara Desert" AND "Southeast Asia") OR ("Pacific Ocean" AND "Bavaria")` returns data objects that have `Sahara Desert` and `Southeast Asia` or `Pacific Ocean` and `Bavaria`.

Wildcard syntax

- A single asterisk finds multiple characters or letters at beginning of a term. A filter for `*olol` returns results for `betaxolol`, `carteolol`, `levobetaxolol`, `levobunolol`, `metipranolol`, and `timolol`.
- An asterisk matches any number of characters in the middle of a term.

*(Asterisk)	`te*t` returns results for `text`, `test`, `tent`, and `tempest`. • An asterisk that is placed after a partial word finds terms that begin with those letters at the end of a term. A filter for `ap*` returns results for `Apple`, `April`, `applied`, and `application`. • An asterisk alone finds all objects.
?	The question mark matches a single character. For example, `compl?` returns results for `comply` but not `complies`. To filter for more than one character, use a question mark for each letter, for example, `ap??e`.

Proximity syntax

w/x	• Filtering uses the `w/x` syntax (`x` represents the number of terms between the two terms) to use the relationship between search terms. `w/0` matches terms that are next to each other in any order, and `w/1` matches terms with one word between them. The terms can appear in any order. • Grouping with proximity allows for even more targeted results. For example, `("stock" OR "option") w/3 (2007)` finds either `stock` or `option` within three terms of `2007`. • Terms can have wildcards or phrases.
pre/x	• The `pre/x` syntax (`x` represents the number of words between the two terms) enables users to target proximity filters by dictating which term comes before another term. For example, `chairman pre/15 options` returns data objects where `chairman` comes up to 15 terms before `options`. • pre/0 matches terms that are next to each other and in the same order; pre/1 matches terms with at most one word between them and in the same order. • Terms can have wildcards or phrases.

Notes regarding syntax

When you are examining text, IBM® StoredIQ evaluates the requested operators and gives precedence to some over others.

• When parentheses are used to group operators, the expression in parentheses is given the highest precedence.
• NOT: `"OctoberNOT November OR December"` means `"(October AND (NOT November)) OR December"`.
• Proximity operators (`"w/"` and `"pre/"`): `"October w/3 November pre/5 December"` means `"(October w/3 November) pre/5 December"`. (`"October"` and `"November"` must be within three words of each other, but both must be before `"December"` by no more than five words.)`"October w/3 November and December"` means `"(October w/3 November) and December"`.
• The `AND` operator: `"October AND November OR December"` means `"(October AND November) OR December"`.
• The `OR` operator or no operator: `"October November OR December"` means the same as `"(October OR November)OR December"`.

• **Parent topic:** [Creating customized searches with the code view](#)

Related tasks

• [Creating customized searches with the code view](#)

Previewing filter results

You can preview sample, returned data before you save that infoset.

Procedure

1. Select the infoset from which you would like to create a new infoset.
2. Apply a filter to this infoset by either using an existing filter or creating a new filter.
3. Click Preview Filter Results, and data that is contained within that infoset appears within the Filter Results. It includes the name, file size, file path, its creation date and time, and the date and time it was last modified.

- **Parent topic:** [Library of available filters](#)

Related concepts

- [Library of available filters](#)

Configuring columns when previewing filter results

When you preview data objects that are contained within an infoset, you can select what type of information is displayed, ensuring that the returned data is meaningful to you.

Procedure

1. Select the infoset from which you want to create a new infoset.
2. Apply a filter to this infoset by either using an existing filter or creating a new filter.
3. Click Preview Filter Results, and a list of data objects that are contained within that infoset appears.
4. Click the blue gear icon. The Configure Column - Preview Filter Results dialog box appears.
5. In the Configure Column - Preview Filter Results dialog box, click the down arrow to the right of options 2-6, selecting the name of the column. Choose from these options: None, File size, File path, File type, Container path, Last accessed, Created, Last modified, Owner name, or Group name.
   Note: The File name option cannot be changed. By default, it is the name of the first column when you are previewing data. Additionally, as you rename columns, more columns are automatically added to the Configure Column - Preview Filter Results dialog box.
6. To reset the columns to the default selections, click Reset to default column selections. The column configuration options return to their default settings.
7. Click Save. The layout of the Filter Results area reflects the chosen options.

- **Parent topic:** [Library of available filters](#)

Related concepts

- [Library of available filters](#)

Set operations

A set operation allows infosets to be combined in different ways to produce another infoset.

Set operations allow for the combination of infosets in new and unique ways that you can highlight specific information that is contained within the infoset. You can apply these types of set operations to infosets: union, intersection, symmetric difference, and subtraction.

Table 1. Set operations

Operation	Description	Example
Union	A union infoset is the combination of two or more infosets: the primary infoset and the check boxes of whatever available infoset you select.	For example, you have three infosets, two of which contain PSTs by owner and a third that contains all Exchange email. You want to combine all three infosets to make it easier to run actions on all of the data objects therein. Notice that all three infosets combine in their entirety, resulting in a new, larger infoset. The combination of the affected areas is turned gray within the interface.
Intersection	An intersected infoset is the result of two or more infosets that contain some of the same objects. This overlap, which is where identical objects are in all infosets, is called the intersection, and it is this intersection that results in the new infoset.	For example, you have two custodians, each with a significant amount of data, some of which is shared. You would like to ascertain which of the shared data objects are identical. In this case, you would create an intersection infoset as you want to find identical, common data objects between the two. Notice that only the area of similarity, the gray intersection between the two white circles, is found in the new infoset.
Symmetric Difference	A symmetric difference infoset combines a primary and selected infoset and removes identical objects from them. The remaining data objects from both filtered infosets then combine to create a new infoset.	For example, you again have two custodians, each with a significant amount of shared data. You would like to ascertain which of the shared data objects are not identical. In this case, you would create a symmetric difference infoset as you want to find data objects that are different between the two. Notice that only the area of difference between the two infosets is the white intersection. This area of difference results in a new infoset of everything but the data that is shared between the two circles.
Subtraction	A subtraction infoset combines a primary and selected infoset, removes identical objects, and then subtracts the remaining objects from the primary infoset. The remainder data objects result in the subtraction infoset.	For example, you have two infosets, both of which contain similar PSTs by owner. You would like to combine both infosets and retain only the unique data objects to make a new infoset. Notice that the two infosets combine, identical data objects are removed, and only the remaining objects, that area in gray, create

the new infoset.

This table describes the options available for set operations for each infoset.

Table 2. Available operation infoset options and descriptions

Available Operation Infoset Option	Description
Name	The name of the infoset against which a set operation can be conducted.
Total Objects	The total number of data objects that are contained within the infoset.
Infoset Size	The size of the infoset.
Composition	The composition of the infoset, top level, or mixed level.
Created	The date and time at which the infoset was created.
Type	The type of infoset. Remember, set operations can be conducted only against user infosets. System infosets are not available in the Available Operation Infosets pane.
Search	Search for a particular infoset with the Enter terms... text box and Search.
Details	Click Details to view the Infoset Details window. Information regarding the infoset's name, description, total number of objects, number of parent and child objects, size, composition, created date and time, type, and percentage of exceptions appears.

Note: For union and intersection set operations, notice that check boxes appear to the left of the infoset names but that for symmetric difference and subtraction operations, radio buttons appear.

- [Creating a union infoset](#)
  Union infosets are the combination or joining in their entirety of two or more infosets.
- [Creating an intersection infoset](#)
  Intersected infosets are the result of the combination two or more infosets where identical data objects are the new infoset.
- [Creating an infoset with a symmetric difference](#)
  Symmetric difference infosets remove identical objects from two filtered infosets. The remaining data objects from both filtered infosets then combine to create a new infoset.
- [Creating a subtraction infoset](#)
  Subtraction infosets remove objects that are contained within the secondary infoset from the primary infoset. The remaining objects then create a subtraction infoset.

- **Parent topic:** [Library of available filters](#)

Related concepts

- [Library of available filters](#)

Creating a union infoset

Union infosets are the combination or joining in their entirety of two or more infosets.

Procedure

1. Select the infoset to which you would like to combine one or more infosets. This infoset is the primary infoset.
   Note: Set operations can only be conducted on user infosets, not system infosets.
   The name of the primary infoset appears within the infoset window.
2. Click Create, and then click Set Operations.
3. Click Union, and then select the check boxes of the available infosets to which you would like to join to the primary infoset. To select all of the available operation infosets, select the check box to the left of Name.
4. Click Create New Infoset. The Create New Infoset dialog box appears.
5. In the New Infoset Name text box, enter the wanted name of this infoset.
6. In the Description text box, enter the wanted name of this infoset.
7. Click Select overlays for this infoset to determine which overlays are to be applied to this infoset. (When you are exploring this infoset, these selected overlays are available to be applied within the Data Map.)
8. To receive an email after the infoset is created, select Email me when the process is complete check box and enter the email address at which you would like to be notified.
9. Click Save and View to watch the infoset create, or click Save.

- **Parent topic:** Set operations

Related concepts

- Set operations

Creating an intersection infoset

Intersected infosets are the result of the combination two or more infosets where identical data objects are the new infoset.

Procedure

1. Select the infoset to which you want to combine one or more infosets. This infoset is the primary infoset.
   Note: Set operations can only be conducted on user infosets, not system infosets.
   The name of the primary infoset appears within the infoset window.
2. Click Create, and then click Set Operations.
3. Click Intersection, and then select the check boxes of the available infosets in which you would like to find identical data objects. Data objects that are the same in each of the selected infosets are then used to form a new, intersection infoset. To select all of the available operation infosets, select the check box to the left of Name.
4. Click Create New Infoset. The Create New Infoset dialog box appears.
5. In the New Infoset Name text box, enter the wanted name of this infoset.
6. In the Description text box, enter the wanted name of this infoset.
7. Click Select overlays for this infoset to determine which overlays are to be applied to this infoset. (When you are exploring this infoset, these selected overlays are available to be applied within the Data Map.)
8. To receive an email after the infoset is created, select Email me when the process is complete check box and enter the email address at which you would like to be notified.
9. Click Save and View to watch the infoset create, or click Save.

- **Parent topic:** [Set operations](#)

Related concepts

- [Set operations](#)

Creating an infoset with a symmetric difference

Symmetric difference infosets remove identical objects from two filtered infosets. The remaining data objects from both filtered infosets then combine to create a new infoset.

Procedure

1. Select the infoset from which you want to remove the symmetric difference of another infoset. This infoset is the primary infoset.
   Note: Set operations can only be conducted on user infosets, not system infosets.
   The name of the primary infoset appears within the infoset window.
2. Click Create, and then click Set Operations.
3. Click Symmetric Difference, and then select the infoset in which you want to find identical data objects. Data objects that are the same in the primary and selected infoset are removed, and the remaining data objects are then combined to form the new infoset.
4. Click Create New Infoset. The Create New Infoset dialog box appears.
5. In the New Infoset Name text box, enter the name of this infoset.
6. In the Description text box, enter the name of this infoset.
7. Click Select overlays for this infoset to determine which overlays must be applied to this infoset. When you are exploring this infoset, these selected overlays are available to be applied within the Data Map.
8. To receive an email after the infoset is created, select Email me when the process is complete check box and enter the email address at which you want to be notified.
9. Click Save and View to watch the infoset create, or click Save.

- **Parent topic:** Set operations

Related concepts

- Set operations

Creating a subtraction infoset

Subtraction infosets remove objects that are contained within the secondary infoset from the primary infoset. The remaining objects then create a subtraction infoset.

Procedure

1. Select an infoset. This infoset is the primary infoset.
   Note: Set operations can only be conducted on user infosets, not system infosets.
   The name of the primary infoset appears within the infoset window.
2. Click Create, and then click Set Operations.
3. Click Subtraction, and then select the infoset that you want to subtract from the primary infoset. Data objects that are the same in the primary and selected infoset are removed, and the remaining data objects form the new infoset.
4. Click Create New Infoset. The Create New Infoset dialog box appears.
5. In the New Infoset Name text box, enter the name of this infoset.
6. In the Description text box, enter the name of this infoset.
7. Click Select overlays for this infoset to determine which overlays must be applied to this infoset. When you are exploring this infoset, these selected overlays are available to be applied within the Data Map.
8. To receive an email once the infoset is created, select Email me when the process is complete check box and enter the email address at which you want to be notified.
9. Click Save and View to watch the infoset create, or click Save.

- **Parent topic:** [Set operations](#)

Related concepts

- [Set operations](#)

Scope operations

Contained data, such as data within .ZIP, .TAR, or .PST files, is hierarchical and can have many different relationship and connections. Depending on how that data is viewed, that data can give a different perception than what is represented. The Scope Operations tab helps users to understand more fully what data is represented by data sets.

There are two actions available within scope operations:

Expand
> This operator expands all of the files within an infoset. Any nested hierarchies or contained files are opened, expanding into a new infoset that is a potentially more accurate representation of the contained data.
> Note: If the files within the infoset are not container files, then the Expand operator has no effect.

Collapse
> This operator collapses or retracts all of the opened or expanded files within an infoset. It results in a new, smaller infoset that is a potentially more accurate representation of the data.
> Note: If the files within the infoset are not container files, then the Collapse operator has no effect.

- [Expanding a hierarchy](#)
  When you are using scope operations to expand a hierarchy, nested hierarchies and associations such as message threads, email attachments, and compressed archive contents are no longer compressed. Data object counts likely either remain the same or increase.
- [Collapsing a hierarchy](#)
  When you are using scope operations to collapse a hierarchy, nested hierarchies and associations such as message threads, email attachments, and compressed archive contents are no longer expanded. Data object counts can possibly either remain the same or decrease.

- **Parent topic:** [Library of available filters](#)

Related concepts

- [Library of available filters](#)

Expanding a hierarchy

When you are using scope operations to expand a hierarchy, nested hierarchies and associations such as message threads, email attachments, and compressed archive contents are no longer compressed. Data object counts likely either remain the same or increase.

Procedure

1. Select the infoset in which you want to expand hierarchies. This infoset is the primary infoset. The name of the primary infoset appears within the infoset window.
2. Click Create.
3. Click Scope Operations.
4. Click Expand. Nested hierarchies and associations are expanded, and data object counts likely either remain the same or increase.
5. Click Create New Infoset. The Create New Infoset dialog box appears.
6. In the New Infoset Name text box, enter the name of this infoset.
7. In the Description text box, enter the name of this infoset.
8. Click Select overlays for this infoset to determine which overlays must be applied to this infoset. When you are exploring this infoset, these selected overlays are available to be applied within the Data Map.
9. To receive an email after the infoset is created, select Email me when the process is complete check box and enter the email address at which you want to be notified.
10. Click Save and View to watch the infoset create, or click Save.
11. Review and validate the objects within the new infoset as needed.

- **Parent topic:** [Scope operations](#)

Related concepts

- [Scope operations](#)

Collapsing a hierarchy

When you are using scope operations to collapse a hierarchy, nested hierarchies and associations such as message threads, email attachments, and compressed archive contents are no longer expanded. Data object counts can possibly either remain the same or decrease.

Procedure

1. Select the infoset in which you want to collapse hierarchies. This infoset is the primary infoset. The name of the primary infoset appears within the infoset window.
2. Click Create.
3. Click Scope Operations.
4. Click Collapse. Nested hierarchies and associations are collapsed, and data object counts likely either remain the same or decrease.
5. Click Create New Infoset. The Create New Infoset dialog box appears.
6. In the New Infoset Name text box, enter the name of this infoset.
7. In the Description text box, enter the name of this infoset.
8. Click Select overlays for this infoset to determine which overlays must be applied to this infoset. When you are exploring this infoset, these selected overlays are available to be applied within the Data Map.
9. To receive an email after the infoset is created, select Email me when the process is complete check box and enter the email address at which you want to be notified.
10. Click Save and View to watch the infoset create, or click Save.
11. Review and validate the objects within the new infoset as needed.

- **Parent topic:** [Scope operations](#)

Related concepts

- [Scope operations](#)

Enhancements to infoset

The Enhance tab provides you with a visual layout of available enhancements. Enhancements are created as models within IBM® StoredIQ Administrator. When an enhancement is applied to an infoset, it updates that infoset's index.

Table 1. Enhance options and descriptions

Enhance Option	Description
Name	The name of the enhancement that can be applied to an infoset.
Type	The enhancement's type. Currently, the only option is Auto-class.
Description	An optional description of the enhancement.

In the Enhancements Details area, a summary of the selected enhancement is displayed, including the enhancement's name, description, and its last usage date.

- Running an enhancement
  Enhancements are another way to refine or distill an existing infoset. Within an enhancement, existing patterns and rules are used to define the criteria for which the enhancement is comprised. By applying an enhancement to an infoset, the same criteria that make up the enhancement is applied to the selected infoset, creating a new, refined infoset.
- Searching for a specific enhancement
  When you have multiple enhancements from which to choose, you can search for a specific enhancement instead of using the slider to move through available options.

Running an enhancement

Enhancements are another way to refine or distill an existing infoset. Within an enhancement, existing patterns and rules are used to define the criteria for which the enhancement is comprised. By applying an enhancement to an infoset, the same criteria that make up the enhancement is applied to the selected infoset, creating a new, refined infoset.

About this task

Note: Enhancements such as an auto-classification model must be created and defined within IBM® StoredIQ Administrator by an administrator. If an enhancement is not available within IBM StoredIQ Administrator, contact the administrator.

To run an enhancement:

Procedure

1. Select the infoset that you want to enhance.
2. Click Enhance.
3. Select the enhancement that you want to apply to the selected infoset.
4. Click Run Enhancement. The Run Enhancement Confirmation window appears. Click Run and View or Run to run the enhancement.

- **Parent topic:** [Enhancements to infoset](#)

Related concepts

- [Enhancements to infoset](#)

Searching for a specific enhancement

When you have multiple enhancements from which to choose, you can search for a specific enhancement instead of using the slider to move through available options.

Procedure

1. Select the infoset that you want to enhance. The selected infoset appears at the top of the infoset window.
2. Click Enhance.
3. In the Enter key terms... text box, enter the name of the enhancements for which you want to search, and then press Search. You can enter either the full enhancement name or a portion of it. To remove the search term, click the X to the left of the Enter key terms.... text box. Enhancements that match the entered search term are returned.
4. Select the enhancement, and then click Run Enhancement.

- **Parent topic:** [Enhancements to infoset](#)

Related concepts

- [Enhancements to infoset](#)

Actions on data

Actions are preconfigured options that help you to manage data objects that are contained within infosets.

Note: Actions must be created and defined within IBM® StoredIQ Administrator by the administrator. They can be scheduled to run immediately or at a later time and date. If an action is not available within IBM StoredIQ Data Workbench, contact your administrator. Within IBM StoredIQ Data Workbench, you can perform these actions on data objects within infosets:

Basic Move
> A move action moves data objects from one location to another. When you are moving an object, the designated location must be specified.

Basic Copy
> A copy action copies data objects from a source volume to a destination volume.

Retention Copy
> This action copies data objects from a source to a retention location for a specified retention period.

Discovery Export
> This action copies data objects and generates files for loading into third-party legal review tools.

Delete
> This action removes data objects from the source volumes.

Within the Modify Attribute policy, these options are available within the policy:

Set read-only
> Sets the read-only bit on files in an infoset for CIFS and NFS data sources.

Set read/write
> Sets the read/write bit on files in an infoset for CIFS and NFS data sources.

Note: For more information and assistance in using actions, contact your administrator.

- [Searching for a specific action](#)
  When you have multiple actions available to choose, you can search for a specific action instead of using the slider to move through available options.
- [Scheduling and running an action](#)
  From the Actions tab, actions can be scheduled to run immediately or at a predetermined schedule, providing you with greater flexibility and control over your data.
- [Canceling a scheduled action](#)
  Scheduled or queued actions' executions can be canceled before they run.

Searching for a specific action

When you have multiple actions available to choose, you can search for a specific action instead of using the slider to move through available options.

Procedure

1. Select the infoset on which you want to act. The selected infoset appears at the top of the infoset window.
2. Click Act.
3. In the Enter key terms... text box, enter the name of the action for which you want to search, and then press Search. You can enter either the full action name or a portion of it. To remove the search term, click the X to the left of the Enter key terms.... text box. Actions that match the entered search term are returned.
4. Click the action that you want to select, and then click Run.

- **Parent topic:** [Actions on data](#)

Related concepts

- [Actions on data](#)

Scheduling and running an action

From the Actions tab, actions can be scheduled to run immediately or at a predetermined schedule, providing you with greater flexibility and control over your data.

Procedure

1. Within the Available Actions area, select the action that you want to schedule, and then click Run Action. The Action Scheduling dialog box appears.
2. Within the Action Scheduling dialog box, you must determine an action schedule. Select from these options:
   - Immediate: If you select this option, the action is queued to run immediately.
   - Scheduled: If you select this option, the action is queued to run according to the schedule you set. In the On text box, enter the date, and in the at text box, enter the time that this action is scheduled to run.
3. After the action schedule is set, click Run and View or Run. The action can be seen in the Details > Execution Log tab. Actions that were scheduled to run immediately show as Running in the Type column. Actions that are scheduled to run later are listed as Queued.

- **Parent topic:** Actions on data

Related concepts

- Actions on data

Canceling a scheduled action

Scheduled or queued actions' executions can be canceled before they run.

Procedure

1. Within the Details > Execution Log tab, select the action that you want to cancel.
   Note: In order for an action to be canceled, it must be scheduled or queued to run. You cannot cancel an action that is running.
2. In the Action Details area, click Cancel Queued Action. The Cancel confirmation window appears.
3. Click OK.

- **Parent topic:** Actions on data

Related concepts

- Actions on data

Exploration of infosets and data

The Explore tab provides you with a visual layout of an infoset. At a glance, you can see how much data you have, its type, its overlays, and detailed information about specific data. By clicking within the data map itself, you can explore the data and analyze various details.

- [Data maps](#)
  The IBM® StoredIQ Data Workbench data map provides a visual layout of your data while giving you in-depth information regarding data source types, data categories, size or amounts, the number of data objects, and more details.
- [Grouping an infoset by data source type](#)
  When you are analyzing what data source types are used to comprise infosets, you can select data source types to display in the data viewer.
- [Viewing data with different views](#)
  Objects within infosets can be viewed by different data views, ranging from category or size to its last accessed or modified date.
- [Selecting a data overlay](#)
  Data overlays, which are configured during infoset creation, can be applied to an infoset.
- [Downloading CSV data](#)
  Data for a selected tile of an infoset can be downloaded as comma-separate-value files.

Data maps

The IBM® StoredIQ Data Workbench data map provides a visual layout of your data while giving you in-depth information regarding data source types, data categories, size or amounts, the number of data objects, and more details.

The data map contains these areas.

Table 1. Data maps

Data Map Option	Description	See Also
Group By	The Group By area contains these options: • Data Source Type: This view displays the different data source types that make up the selected infoset. By selecting or clearing the data source types, the data map automatically reconfigures, providing you with a visual overview of your data. • None: This option removes data source types from the data map.	For more information, see [Grouping an Infoset by Data Source Type](#).
View By	The View By area contains these options: • Category View: You can view data by category, including container files and archives, images, spreadsheets, email, and so on. • Size View: You can view data by size, from less than 1 MB to over 100 GB. The Size View is only active for infosets that have a top-level composition. • Created View: You can view data by created date from within the past 30 days to more than five years ago. • Last Modified View: You can view data by the last modified date, from within the past 30 days to more than five years ago. • Last Accessed View: You can view data by the last accessed date, from within the past 30 days to more than five years ago. By selecting or clearing the viewing options, the data map automatically reconfigures, providing you with a visual overview of your data.	For more information, see [Viewing Data with Different Views](#).
Select Data Overlay	Data overlays are a visual representation of a saved filter. For example, you might create filters that can show you where objects that contain credit-card numbers, Social Security numbers, or prohibited words can be found within your system. Note: To apply a data overlay, you must first create and save a filter. For more information, see [Building a filter](#). Within the data map, you have these viewing	For more information, see [Selecting a Data Overlay](#).

Data Map	options. • Object Count: If this option is selected, the data map displays the number of objects and their percentage of that data source when you place the pointer over a category. • Object Size: If this option is selected, the data map displays the size and percentage of the objects for that data source when you place the pointer over a category. Object Size is only active for infosets that have a top-level composition. You also have these viewing options: • Squarify: This view depicts data as unordered squares rather than rectangles. • Slice & Dice: This view depicts data as ordered rectangles. • Strip: This view strips data, simply depicting it in the easiest viewing aspect possible.	For more information, see [Grouping an Infoset by Data Source Type](#).
Data Map Details	This area displays details about a tile that is selected in the data map. By clicking an individual map tile, details about that specific tile are displayed. You can also see the total number of data objects referred to by the map tile. Within the Data Map Details, you have these viewing options: • Volumes: If the Volumes option is selected, the volume's name and view selected displays in the Data Map Details for the selected tile. The bar graph populates automatically based on the count or size of the selected tile. • Owners: If the Owners option is selected, the owners of content displays within that tile are displayed in the Data Map Details. The bar graph populates automatically based on the count or size of the selected tile. Note: Selected tiles have a green border. Tiles over which the pointer was placed have a black border. The Data Map Details also shows you the Total Size of the selected tile. The bar graph automatically populates, providing a visual representation of how the size of the selected tile relates to that of the size of the infoset as a whole By clicking View Full List, the Data Map Details dialog appears. From here, you can download information about the selected tile	

to a CSV file.

Under the data map, this field displays the following options.

Viewing x of x total object	Under the data map, this field displays the following options. • If Object Count is selected, this figure indicates the number of data objects that are being viewed compared to the total number of data objects. • If Object Size is selected, this figure indicates the data objects' size in bytes compared to the total byte size of the data objects.

Note: Selected options such as Data Source Type,Object Count, and Slice & Dice within the user interface are blue.

• **Parent topic:** [Exploration of infosets and data](#)

Related concepts

• [Exploration of infosets and data](#)

Grouping an infoset by data source type

When you are analyzing what data source types are used to comprise infosets, you can select data source types to display in the data viewer.

Procedure

1. Select the infoset for which you want to explore its data. The name of the selected infoset appears within the infoset window.
2. Click Explore. The data map appears, providing a visual overview of the data within the selected infoset.
3. To view the selected infoset and the different data sources it uses, click Data Source Type in the Group By area. To view only the infoset without its types of data sources, click None. A complete listing of the different data source types that make up the selected infoset appears within the Group By area.
4. Select and clear the check boxes to the left of the data source types. Data sources that are cleared are removed from the data map, and the view is automatically reconfigured to reflect the data sources that are selected or cleared within the Group By area.
5. Use the data map to better understand your infoset and the data sources that are used to make up that infoset.

- **Parent topic:** [Exploration of infosets and data](#)

Related concepts

- [Exploration of infosets and data](#)

Viewing data with different views

Objects within infosets can be viewed by different data views, ranging from category or size to its last accessed or modified date.

Procedure

1. Select the infoset for which you want to explore its data. The name of the selected infoset appears within the infoset window.
2. Click Explore. The data map appears, providing a visual overview of the data within the selected infoset.
3. To view the selected infoset by View By categories, select one of these categories in the list: Category View, Size View, Created View, Last Modified View, or Last Accessed View. The infoset is displayed within the data map, and the map tiles correspond to the selected view type.

Table 1. View By Selection Options

View By Option	Available Category Options
Category View	These options are available when Category View is selected: ○ Container files and archives ○ Images ○ Spreadsheets ○ Miscellaneous ○ System files ○ Multimedia ○ Computer programming ○ Word processing ○ Presentation ○ Email ○ Collaboration
Size View	These options are available when Size View is selected: ○ Less than 1 MB ○ 1 MB to 10 MB ○ 10 MB to 100 MB ○ 100 MB to 1 GB ○ 1 GB to 10 GB ○ 10 GB to 100 GB ○ Over 100 GB Note: This option is only active when you are viewing a top-level infoset.
Created View	These options are available when Created View is selected: ○ Last 30 days ○ 1 - 6 months ago ○ 6 - 12 months ago ○ 1 - 3 years ago ○ 3 - 5 years ago ○ Over 5 years ago
Last Modified View	These options are available when Last Modified View is selected: ○ Last 30 days ○ 1 - 6 months ago ○ 6 - 12 months ago ○ 1 - 3 years ago ○ 3 - 5 years ago ○ Over 5 years ago
Last Accessed View	These options are available when Last Accessed View is selected: ○ Last 30 days ○ 1 - 6 months ago ○ 6 - 12 months ago

- 1 - 3 years ago
- 3 - 5 years ago
- Over 5 years ago

4. Select and clear the check boxes to the left of the viewing options. The viewing options that are cleared are removed from the data map, and the view is automatically reconfigured to reflect the options that were selected or cleared within the View By area.
5. Use the data map to better understand your infoset and the data sources that are used to make up that infoset.

- **Parent topic:** [Exploration of infosets and data](#)

Related concepts

- [Exploration of infosets and data](#)

Selecting a data overlay

Data overlays, which are configured during infoset creation, can be applied to an infoset.

Procedure

1. Select the infoset for which you want to explore its data. The name of the selected infoset appears within the infoset window.
2. Click Explore. The data map appears, providing a visual overview of the data within the selected infoset.
3. Select the data overlay that you want to apply to the infoset. For more information, see [Creating an overlay](#).

- **Parent topic:** [Exploration of infosets and data](#)

Related concepts

- [Exploration of infosets and data](#)

Downloading CSV data

Data for a selected tile of an infoset can be downloaded as comma-separate-value files.

Procedure

1. Select the infoset for which you want to download CSV data.
2. Click Explore, and then select a tile in the Data Map.
3. In Data Map Details, click View Full List. The Data Map Details dialog box appears.
4. Click Download to CSV, and then open or save the CSV file as needed.

- **Parent topic:** [Exploration of infosets and data](#)

Related concepts

- [Exploration of infosets and data](#)

Exceptions

Exception lists contain various data objects such as files, emails, which encountered an error when an action was conducted on the infoset. This list helps you to know exactly what data needs attention due to some error or exception. It also serves as an informational trace of what went wrong during the process.

The Exceptions tab is divided into three areas: Event, Types, and Exception Objects. When you select an exception within either the Event or Types areas, contextual details for the selected exception appear within the Exception Objects area.

Events

Table 1. Events column definitions

Events Column Heading	Definition
Name	The name of the exception.
Started	The date and time at which the exception started.
Type	The action type.

Types

Table 2. Types column definitions

Types Column Heading	Definition
Name	The name of the exception that was applied to the data objects. Options include Export Error, File Access Error, File Copy Error, Network Error, Other, Permissions Error, Server Error, System Error, or Unsupported Action.
Severity	The severity level of the exception.
% of Infoset	The percent of the infoset with errors.
Count	The number of data objects with an error.
Size	The size of data objects with an error.

Exception objects

Table 3. Exception objects column definitions

Exception Objects Column Heading	Definition
Name	The name of the exception object.
Timestamp	The date and time of the exception.
Event	The run action that resulted in the exception.
Exception	The type of exception encountered.
Message	Based on the exception type, this message provides additional information about the exception itself using information that is pulled from the audit log.
File size	The exception object's file size.
File path	The file path to the exception object's location.

Regarding Step-Up Snippets: The Exceptions tab does not report exceptions for a step-up snippet. To see those exceptions, check the harvest audits on individual data servers.

- [Searching for a specific exception object](#)

When you have multiple exception objects to choose, you can search for a specific exception object instead of using the slider to move through available options.

- [Viewing exception object details](#)
  The Exception Object Details dialog box displays exception details that result from different data objects.
- [Exporting the exception objects list](#)
  Use this procedure to export a CSV list of the exception objects that are listed in the Exception Objects area.

Searching for a specific exception object

When you have multiple exception objects to choose, you can search for a specific exception object instead of using the slider to move through available options.

Procedure

1. Click Exceptions.
2. In the Enter key terms here... text box of the Exceptions Objects, enter the name of the exception object and then press Search. You can enter either the full exception object name or a portion of it. To remove the search term, click the X to the left of the Enter key terms.... text box. Exception objects that match the entered search term are returned.
3. Click the exception object that you want to select.

- **Parent topic:** [Exceptions](#)

Related concepts

- [Exceptions](#)

Viewing exception object details

The Exception Object Details dialog box displays exception details that result from different data objects.

Procedure

1. Select the exception objects for which you want to view details.
2. Click View Details, and the Exception Object Details dialog box appears. This dialog box displays details about the data object that caused the exception.
3. Click Next or Previous to view different data objects that caused the exceptions.
4. After you finish viewing exception object details, click Close.

- **Parent topic:** [Exceptions](#)

Related concepts

- [Exceptions](#)

Exporting the exception objects list

Use this procedure to export a CSV list of the exception objects that are listed in the Exception Objects area.

Procedure

1. Select the event, type, and exception objects that you want to export.
2. Click Export Exception Objects List. The Export Exceptions Object List dialog box appears.
3. To send an extra copy of the CSV file to another person, select the Also send a copy to the following address check box and enter the email address in the text box. The exceptions object list is exported as a CSV file, and a link to the location is emailed.
4. Click OK.

- **Parent topic:** Exceptions

Related concepts

- Exceptions

Filter condition reference and expression language

This appendix provides in-depth information about full-text and attribute conditions and how to use the expression language.

- [Full-text syntax](#)
  The full-text editor supports full-text syntax.
- [Full-text macros](#)
  The IBM® StoredIQ Platform full-text filter supports a set of predefined macros.
- [Extended ASCII characters](#)
  Extended ASCII characters are handled several different ways in the full-text filter. Some are special characters while others are treated as spaces. Some are treated as letters that depend on the spaces around the characters and the configured full-text settings.
- [Attribute condition reference](#)
  The following section explains how to type in a reference to an attribute condition.
- [Specification of attribute conditions](#)
  Attributes, entity, tag, or query can be divided up in several ways.
- [Type of object to match](#)
  For any attribute condition, you can specify what object that condition must match.
- [Filter condition reference and expression language](#)
  This appendix provides in-depth information about full-text and attribute conditions and how to use the expression language.
- [Multiple matching values for operators](#)
  You can give multiple matching values that use the following operators.
- [At-least expressions](#)
  On occasion, it is helpful to specify that some number of conditions must be true before an object is considered to match. These expressions are called `at-least` expressions because at least `X` subexpressions must match an object before the object matches the overall expression.
- [Parent and child conditions](#)
  This parent and child conditions represent advanced function that must be undertaken only by experienced administrators.
- [No-inheritance conditions](#)
  The no-inheritance conditions represent advanced function that must be undertaken only by experienced administrators.
- [Analytics filter expressions](#)
  The analytics filter expressions represent advanced function that must be undertaken only by experienced administrators.

- **Parent topic:** [Filter condition reference and expression language](#)

Full-text syntax

The full-text editor supports full-text syntax.

The full-text feature supports a wide range of full-text syntax to broaden or narrow filters. A one-word term like `September` returns results from all harvested data objects that include the term in their content. Word matching is case-insensitive. A list of words match any of those words. For example, `September OR October` match all objects that contain either `September` or `October`.

Grouping

When you are examining text, IBM® evaluates the requested operators and gives precedence to some over others. These rules are described in the following list.

Note: Right-associative and left-associative means that when parentheses are used to group operators, the expression in parentheses is given the highest precedence. If no parentheses are used in the expression, associativity rules are used instead. Right-associative means that the operator is grouped initially with the operand to its right. Left-associative means that the operators/operands to the left are evaluated before the operators/operands to the right.

1. `NOT`, right-associative. `"October NOT November OR December"` means `"(October AND (NOT November)) OR December"`.
2. Proximity operators (`"w/"` and `"pre/"`), left-associative. `"October w/3 November pre/5 December"` means `"(October w/3 November) pre/5 December"`. (`"October"` and `"November"` must be within three words of each other, but both must be before `"December"` by no more than five words.) `"October w/3 November and December"` means `"(October w/3 November) and December"`.
3. The `AND` operator, left-associative. `"October AND November OR December"` means `"(October AND November) OR December"`.
4. The `OR` operator or no operator, left-associative. `"October November OR December"` means the same as `"(October OR November) OR December"`.

Boolean Operators

This table lists Boolean operator syntax and descriptions.

Table 1. Boolean operator syntax

Item	Description
AND &&	- When used between two terms, the AND operator ensures that both terms appear in the data object. For example, the filter `chairman AND stock` returns only data objects where `chairman` and `stock` both appear. - The double ampersand (&&) can be used instead of the AND operator. AND can be uppercase or lowercase.
OR ||	- When used between two terms, the OR operator returns data objects where either term appears. For example, the filter `chairman OR stock` returns data objects that have `chairman`, `stock`, or both in the data object. - The double pipe (||) can be used instead of the OR operator. OR can be uppercase or lowercase.
NOT !	- When used before a term, the NOT operator ensures that the following term does not appear in any data objects returned. For example, `Apple NOT IBM` returns data objects where word `Apple` appears, but not `IBM`. You can think of the NOT item as `"and not"` or `"but not"`. - The exclamation point (!) can be used instead of the NOT operator. NOT can be uppercase or lowercase.

Multiple Terms or Term Phrase

This table lists multiple terms or term phrase syntax items and descriptions.

Table 2. Terms

Item	Description
" " and *	- Quotation marks around a term phrase return data objects that have that phrase in order in a data object. For example, `"September 20 1998"` returns only data objects with the exact phrase of `September 20 1998`. Double quotation marks allow for wildcards in terms whereas single quotation marks do not. - A term phrase has the stop words removed. For phrases with double quotation marks, all wildcard characters in the phrase are matched as wildcards. For phrases with single quotation marks, the words are divided on white space and filtered for exactly as they are entered. For example, `"stock AND option*"` returns results with both `stock option` and `stock options`. The word `AND` is ignored as a stop word.
Multi-term, no Boolean operators	A filter for a multiple word without Boolean operators or quotation marks is an implied OR filter. For example, `September 20 1998` returns data objects that have `September OR 20 OR 1998`. Stop words are removed.
Grouping	- Grouping with Boolean operators and parentheses allows users to narrow results by using attributes. For example, `($10 OR billion) AND options` returns data objects where `$10` or `billion` might be in the data object, but `options` must exist. - Boolean operators (AND, OR, NOT) work in a grouping or in a proximity group. So `("Sahara Desert" AND "Southeast Asia") OR ("Pacific Ocean" AND "Bavaria")` returns data objects that have `Sahara Desert` and `Southeast Asia` or `Pacific Ocean` and `Bavaria`.

Wildcard

This table lists wildcard syntax items and descriptions.

Table 3. Wildcard

Item	Description
* (Asterisk)	- A single asterisk finds multiple characters or letters at beginning of a term. An asterisk that is placed before a partial word finds terms that end with those letters. For example, a filter for `*olol` returns results for `betaxolol`, `carteolol`, `levobetaxolol`, `levobunolol`, `metipranolol`, and `timolol`. - An asterisk matches any number of characters in the middle of a term. An asterisk that is placed in the middle of a term finds terms that have those characters on either side of the asterisk. For example, `te*t` returns results for `text`, `test`, `tent`, and `tempest`. - An asterisk that is placed after a partial word finds terms that begin with those letters at the end of a term. For example, a filter for `ap*` returns results for `Apple`, `April`, `applied`, and `application`. - An asterisk alone finds all objects.
?	The question mark matches a single character. For example, `compl?` returns results for `comply` but not `complies` or `compliance`. A filter for `ap?` finds `ape`, but not `apple`. To filter for more than one character, use a question mark for each letter, for

example, `ap??e`.

Proximity

This table lists proximity syntax items and descriptions.

Table 4. Proximity syntax

Item	Description
w/x	- Filtering uses the `w/x` syntax (x represents the number of terms between the two terms) to use the relationship between terms to get more targeted results. A two-term filter with one term that appears at the beginning of a data object and the other at the end might not be as relevant as when the terms appear in the same paragraph or sentence. For example, in an SEC data object, `chairman w/15 options` is a more relevant filter than `chairman AND options` appearing in the same data object that might be several pages long. `w/0` matches terms that are next to each other in any order. `w/1` matches for terms with one word between them. The terms can appear in any order in the document. - Grouping with proximity allows for even more targeted results. For example, `("Sahara Desert" "Southeast Asia") w/5 ("Pacific Ocean" Bavaria)` finds either `Sahara Desert` or `Southeast Asia` within five terms of `Pacific Ocean` or `Bavaria`. - Terms can have wildcards or phrases.
pre/x	- The `pre/x` syntax (x represents the number of terms between the two terms) enables users to further target proximity filters by dictating which term comes before another term in the data object and how far before it is. For example, `chairman pre/15 options` returns data objects where `chairman` comes up to 15 terms before `options`. `pre/0` matches terms that are next to each other and in the same order. `pre/1` matches terms with at most one word between them and in the same order. - Terms can have wildcards or phrases.

- **Parent topic:** Filter condition reference and expression language

Related concepts

- Filter condition reference and expression language

Full-text macros

The IBM® StoredIQ Platform full-text filter supports a set of predefined macros.

A macro is an abbreviated set of instructions that are used in common filter patterns. For example, the filter `{SSN}` finds any document that contains Social Security numbers.

Full-text macros have the following characteristics:

- A macro is an alphanumeric name that is enclosed in curly braces, for example, `{SSN}`.
- A macro's name is case insensitive. For example, the macro `{ssn}` is the same as the macro `{SSN}`.
- The { (left curly brace) can have a space after it. The }, right curly brace, can have a space before it. For example, `{ SSN }` is the same as `{SSN}`.
- The macro's value is replaced in the macro's full-text query. Then, the filter is parsed and run.
- Any unrecognized macro results in a syntax error.
- Regular expressions, that is named character classes, do not match macros. That is, a regular expression like `\p{Lower}` (matches any lowercase ASCII character) is not recognized as a macro.

The following macros are defined within IBM StoredIQ Platform's full-text feature. Each of these macros is a high-confidence macro. It means that the filter `{SSN}` finds any document that contains Social Security numbers with high confidence that the matches really are Social Security numbers. To find anything that might be a Social Security number, use the low-confidence macro `{SSN-lc}`.

Macros can be combined with any other full-text expression, such as `{SSN} w/5 "John Doe"`.

- {SSN}: Social Security numbers, which can include dashes and numbers
- {USPhone}: US phone numbers, which can include dashes, dots, parentheses, numbers, and letters. This macro can filter for mnemonics like 1-800-flowers.
- {InternetAddress}: Legal IP addresses, either as numbers or as names, which can include dots, numbers, and letters.
- {EmailAddress}: Legal email addresses, which can include @, dots, numbers, and letters.
- {CCAmex}: American Express credit-card numbers, which can include dashes and numbers.
- {CCDiners}: Diners Club credit-card numbers (including enRoute credit-card numbers), which can include dashes and numbers.
- {CCDiscover}: Discover credit-card numbers, which can include dashes and numbers.
- {CCJCB}: JCB credit-card numbers, which can include dashes and numbers.
- {CCMC}: MasterCard credit-card numbers, which can include dashes and numbers.
- {CCVisa}: Visa credit-card numbers, which can include dashes and numbers.
- {CreditCard}: Any credit-card number from the listed providers.

- **Parent topic:** Filter condition reference and expression language

Related concepts

- Filter condition reference and expression language

Extended ASCII characters

Extended ASCII characters are handled several different ways in the full-text filter. Some are special characters while others are treated as spaces. Some are treated as letters that depend on the spaces around the characters and the configured full-text settings.

- If the full-text setting is configured to index with special characters or punctuation, those special characters or punctuation marks are indexed and can be searched for in filters.
- If that character is not configured under the setting that is mentioned in the previous bullet point, refer to the following table to see what behavior applies to the character.

Table 1. Extended ASCII Character syntax

Extended ASCII Character Syntax	Description
Special characters	Some special characters generate syntax errors because they cannot be processed. Either remove the character from the filter or surround the filter with single quotation marks. For example, a filter for `hello!` without quotations generates an error. However, if the filter is `'hello?'` (with single quotation marks), the query can be processed. Whether you get results that match the term depends on the full-text indexing settings with which the volume was harvested.
	If `!` is on the list of special characters that are indexed, then the filter treats the `!` as a letter, and the filter finds results for the documents that contain `!`. If `!` is not on the list of special characters, the filter treats `!` as a space and documents that contain `!` are not found.
	Most special characters can also be searched for with quotation marks. For example, `"hello!"` finds `hello!`. The difference between quotation marks and single quotation marks is that quotation marks also have another function: to find words that are next to on another. Because of this, special characters that are used elsewhere in the syntax, such as `*` and `?`, are treated as the full-text operators for wildcards instead of as the characters `*` and `?`.
	When you are filtering for single quotation marks, you can surround the expression that you are filtering for with single quotation marks to find it. For example, if you configured the appliance to treat quotation marks as letters, you can filter for `"Joe's"` to find the word `Joe's`.
	These special characters generate errors unless surrounded by quotation marks: `@ - / & : ! ( ) { } [ ] ^ ~ : , # $ % = _`
Period (.)	A period with no space on either side is treated like a letter. For example, a filter for `172.16.0.46` finds the whole number of `172.16.0.46`, not partial numbers of `172`, `16`, `046`.
	A period with a space on either side is treated as a space and ignored. For example, a filter for `"1234 Mahana St."` ignores the period and just finds `1234 Mahana St` only without a period.
At (@)	An `@` symbol is indexed like a letter when there are no spaces on either side. For example, a filter for `jdoe@yourcompany.com` finds the entire email address. An `@` symbol with a space on either side or one side is treated as a space and ignored. For example, `@home` finds home in the data object.

Hyphen (-)	A hyphen is indexed like a letter when there are no spaces on either side. For example, a filter for `012-44-5668` finds the entire number, not `012`, `44`, `5668` as separate numbers. A hyphen with a space on either side is treated as a space and ignored, except when followed by numerals, such as `"-1"`.
Slash (/)	A slash is indexed like a letter when there are no spaces on either side. For example, a filter for `10/30/1940` finds the entire expression `10/30/1940` (and not `10 30 1940`) as separate numbers.
Ampersand (&)	An ampersand is indexed like a letter when there are no spaces on either side. For example, a filter for `S&P` finds `S&P` in the data objects. However, a filter for `Proctor & Gamble` without quotation marks finds either `Proctor` or `Gamble`. A filter with quotation marks `"Proctor & Gamble"` finds the phrase `Proctor Gamble` only.
Colon (:)	A colon is considered a special character. Terms with colons can be filtered either by surrounding the term with single quotation marks, `'21:35:15'`. Specifying terms with colons without the single quotation marks results in an invalid syntax error.
Apostrophe (')	The apostrophe is treated as a space when it is a possessive. For example, a filter for patient's returns results for patient. However, embedded apostrophes, like in the name `O'Hara`, are not considered a space and are indexed as a whole word. An apostrophe is treated as a single quotation mark, so it must always be quoted with double quotation marks.
Comma (,)	Commas are left in when surrounded by numerals, and those numerals make a legitimate number. For example, `"1,000"` is indexed as `1,000`, but `"10,00"` is indexed as `10` and `00`. In all other cases, a comma is treated as a space.
Number sign (#), Dollar sign ($), Percent (%), Equal sign (=), Underscore (_)	These characters are treated as white space in the filter. For example, a search for `$100` returns results for the number `100` whether `100` refers to dollars, a percentage, or a whole number.

- **Parent topic:** Filter condition reference and expression language

Related concepts

- Filter condition reference and expression language

Attribute condition reference

The following section explains how to type in a reference to an attribute condition.

If you always select an attribute from the custom filters or one of the attribute groups, xxx takes care of the attribute reference for you. If you do not select an attribute from the custom filters or one of the attribute groups, then you need to review and use this section.

A typed attribute reference is composed of several parts: `attribute-type path/attribute-name`. The `attribute-type` gives the specific type of attribute to which the `path` and `attribute-name` refer. This prefix is a short prefix of the attribute type. Attributes can have duplicate names, even within one folder.

Table 1. Attribute type prefix

Attribute-type Prefix	Description
"att:" or "@"	This prefix is a system-defined attribute or user-defined entity.

The `"path"` describes the folder path to the attribute. Generally, it is optional, but it is required if the attribute name alone is not unique. You do not need to type the full folder path to an attribute.

If an attribute name or folder path name has spaces or punctuation in it, then you must enclose the name in double or single quotation marks. However, you do not need to quote the forward slash ('/') between folders or before the attribute name. It is not legal to have part of an attribute name quoted; all of it must be quoted. You must never quote the attribute type prefix.

- **Parent topic:** [Filter condition reference and expression language](#)

Related concepts

- [Filter condition reference and expression language](#)

Specification of attribute conditions

Attributes, entity, tag, or query can be divided up in several ways.

- What type of values does the attribute have? Attributes can have values of `number`, `string`, `enumeration`, `boolean`, or `date`. An enumeration attribute is an attribute that has a limited set of possible values, which are shown in a list.

  When an attribute is a compound attribute, each of its subattributes also has a value type: `string`, `enumeration`, `boolean`, or `date`. The subattributes can also be either `sparse` or `dense`. Currently, all subattributes are single-valued. Compound attributes can also be `sparse` or `dense`. (Currently all compound attributes are sparse.)

- Do all objects have a value for this attribute? If the objects all have a value, it is known as a dense attribute. If the objects do not have a value, it is known as a sparse attribute. All objects have an `Object name` value, so `Object name` is a dense attribute. Only some objects have a `CC` attribute value, so `CC` is a sparse attribute.
- Can an attribute have multiple values within one object? Each object has one name, so `Object name` is a single-valued attribute. There might be multiple `To` values on an email, so `To` is a multivalued attribute. Compound attributes can be single-valued or multivalued. A compound attribute never exists where all of its subattribute values are missing.

These condition operators are applicable to each attribute category.

Note: For illustration purposes, some operators are shown in uppercase, and some are shown in lowercase. The parser recognizes operators in either uppercase or lowercase.

- [Conditions for string attributes](#)
  Tags are considered single-valued, sparse string attributes. They are not considered enumeration attributes, even though they have a limited set of possible values.
- [Conditions for enumeration attributes](#)
  Some attributes have a limited set of predefined values that they can match, which are called enumeration attributes.
- [Conditions for date attributes](#)
  The conditions for date attributes are listed as follows.
- [Conditions for number attributes](#)
  Conditions for number attributes are listed as follows.
- [Conditions for sparse attributes](#)
  The conditions for multivalued and single valued sparse attributes are listed as follows.
- [Conditions for compound attributes](#)
  Conditions for compound attributes are listed as follows.
- [Conditions for multivalued attributes](#)
  Conditions for multivalued sparse or dense attributes are listed as follows.
- [Access mode attribute](#)
  Specifies an attribute that can be used in a search expression for creating infosets that are based on certain file attributes that are collected during harvests.

- **Parent topic:** [Filter condition reference and expression language](#)

Related concepts

- [Filter condition reference and expression language](#)

Conditions for string attributes

Tags are considered single-valued, sparse string attributes. They are not considered enumeration attributes, even though they have a limited set of possible values.

```
attribute = value or attribute is value
```

```
attribute != value or attribute is not value
```

```
attribute contains value
```

```
attribute does not contain value
```

```
attribute starts with value
```

```
attribute does not start with value
```

```
attribute ends with value
```

```
attribute does not end with value
```

```
attribute matches value
```

```
attribute does not match value
```

The `value` can be a single word, a double-quoted string, or a single-quoted string. For matches, the value is a regular expression within quotation marks.

- **Parent topic:** [Specification of attribute conditions](#)

Related concepts

- [Specification of attribute conditions](#)

Conditions for enumeration attributes

Some attributes have a limited set of predefined values that they can match, which are called enumeration attributes.

In this case, only '=' and '!=' operators are supported.

`attribute = value or attribute is value`

`attribute != value or attribute is not value`

A list of enumeration attributes is as follows:

- Container type
- Is-part-of
- Box reason
- Encryption type
- System processing state
- Content processing state
- Binary processing state
- System error code
- Content error code
- Binary error code
- Message type
- Latest review status
- Object category

- **Parent topic:** [Specification of attribute conditions](#)

Related concepts

- [Specification of attribute conditions](#)

Conditions for date attributes

The conditions for date attributes are listed as follows.

`attribute = value` or `attribute is value`

`attribute !+ value` or `attribute is not value`

`attribute < value`

`attribute <= value`

`attribute > value`

`attribute >= value`

`attribute between value1 and value2`

`attribute not between value1 and value2`

The `value` can have one of two forms: an `explicit date` or a `relative date that is offset from today`.

An explicit date is in the form of YYYY-MM-DD, for example, 2009-03-08.

- Other date formats are supported.
- Time is not supported.
- The dates are all interpreted as midnight of the listed date.

A relative date offset is specified as a positive number, followed by a required date units, and the `ago` or `ahead` keyword. The `ago` keyword is the default and means a date in the past. The `ahead` keyword means a date in the future. For example, `3 months ago`, `3 days ahead`, or `3 weeks ahead` are legal relative dates. The allowed date units are as follows:

- days
- weeks
- months
- years

This function is also supported:

- -3 days = 3 days ago
- +3 weeks = 3 weeks ahead
  Note: Months are considered as 30 days; years are considered as 365 days.
- Normally, relative date offsets are in the past, but some dates can be in the future, such as a `Retention expiration date`. For these references, use the `ahead` keyword instead of `ago`.
- Relative date offsets can also be given as a positive or negative number, such as `-3 days` or `+3 months` without the `ago` or `ahead` keyword. In this case, negative numbers are considered in the past and positive numbers are considered in the future.

- **Parent topic:** [Specification of attribute conditions](#)

Related concepts

- [Specification of attribute conditions](#)

Conditions for number attributes

Conditions for number attributes are listed as follows.

`attribute = value` or `attribute is value`

`attribute != value` or `attribute is not value`

`attribute < value`

`attribute <= value`

`attribute > value`

`attribute >= value`

`attribute between value1 and value2`

`attribute not between value1 and value2`

The `value` is a positive or negative integer. Size attributes (`Size` and `Size (on disk)`) can take an optional size unit after the number. These amounts are the legal size units.

- `b` or `bytes` for single bytes
- `k` or `kb` for kilobytes
- `m` or `mb` for megabytes
- `g` or `gb` for gigabytes
- `k` or `kb` for kilobytes

Note: The default size unit is bytes. Size units are based on multiples of 1024, not 1000. Event duration can have time units. These time amounts are the legal size units:

- `m` or `minutes` for minutes
- `h` or `hours` for hours
- `d` or `days` for days

- **Parent topic:** [Specification of attribute conditions](#)

Related concepts

- [Specification of attribute conditions](#)

Conditions for sparse attributes

The conditions for multivalued and single valued sparse attributes are listed as follows.

```
attribute APPEARS
attribute NOT APPEARS or attribute DOES NOT APPEAR
```

These conditions specify whether the attribute has any values for matching objects. In this case, it means that one of the subattributes has a value on this object. Tags are considered single-valued, sparse string attributes.

It is not an error to specify these conditions on dense attributes. However, if you do, the parser warns you that the condition is always `true` or `false`. These operators can also apply to compound attributes. In this case, it means that one of the subattributes has a value on this object.
Note: As a backward compatible measure, if `att: attribute appears` is given on a dense, boolean attribute, it is converted into `att: attribute = true`. If `not appears` is given, it is converted into `att: attribute = false`.

- **Parent topic:** [Specification of attribute conditions](#)

Related concepts

- [Specification of attribute conditions](#)

Conditions for compound attributes

Conditions for compound attributes are listed as follows.

```
attribute WHERE <sub-attribute-condition>
attribute WHERE (<sub-attribute expression>)
```

A `<sub-attribute-condition>` is just like any of the attribute conditions that are described except it is on a subattribute of the compound attribute. Additionally, a `<sub-attribute-condition>` can be just the subattribute name. It is equivalent to the subattribute `APPEARS`.

A `<sub-attribute-expression>` is any combination of `<sub-attribute-condition>`'s combined with AND, OR, or NOT. This expression must be enclosed in parentheses. The `<sub-attribute-expression>` can also be empty. In this case, it is the same as attribute APPEARS.

This example uses WHERE expressions:

```
att: "Site owner (SharePoint)" WHERE name = "John Smith"
att: "Site owner (SharePoint)" WHERE (name = "John Smith" AND url APPEARS )
att: "Site owner (SharePoint)" WHERE (name = "John Smith" OR url NOT APPEARS)
att: "Site owner (SharePoint)" WHERE ()
```

Note: This expression is the same as `att: "Site owner (SharePoint)" APPEARS`.

```
att: "Site owner (SharePoint)" WHERE (name)
```

Note: This expression is the same as `att: "Site owner (SharePoint)" WHERE name APPEARS`.

- **Parent topic:** [Specification of attribute conditions](#)

Related concepts

- [Specification of attribute conditions](#)

Conditions for multivalued attributes

Conditions for multivalued sparse or dense attributes are listed as follows.

`attribute APPEARS operator number`

`attribute APPEARS UNIQUE operator number`

These operators can also apply to compound attributes, and the `operator` can be one of the following operators:

- `= >`
- `>=`
- `<`
- `<=`

The `operator number` specifies a condition on the number of values present for the attribute on any matching object.

- If `operator number` is not given, then `> 0` is assumed.
- If `UNIQUE` is given, it specifies the number of unique values present for the attribute on any matching object.

It is not an error to specify these conditions on single-valued attributes. However, if you do, the parser warns you that the condition is always `true` or `false`.

- **Parent topic:** [Specification of attribute conditions](#)

Related concepts

- [Specification of attribute conditions](#)

Access mode attribute

Specifies an attribute that can be used in a search expression for creating infosets that are based on certain file attributes that are collected during harvests.

Access mode attribute supports these operators: `APPEARS`, `NOT APPEARS`, and `WHERE`.

The `Subject` subattribute supports these operators: `IS`, `IS NOT`, `=`, `!=`, `any()`, and `all()`.

The `Readable` and `Writable` subattributes both support these operators: `IS`, `IS NOT`, `=`, and `!=`.

For example,

- To find data objects for which you collect or maintain information about the read or write file system flags, enter `att:"Access Mode" appears`. This search returns every system-level object on CIFS or NFS volumes.
- To find data objects that are writable by either of the three types of subjects, enter `att:"Access Mode" where (Subject is any (Owner, Group,Others) and Writable is True)` or `att:"Access Mode" where (Writable is False)`. These two examples are semantically identical. If you think of the r/w/x flags as separate groups of flags (that is, rwx | rwx | rwx), where one is for the `Owner`, one is for the `Group` that a user might be associated with, and one is for `Others` in that order, then the `Readable` or `Writable` subattributes are evaluated independently for each three group of flags.

  The `Subject` subattribute (if present) then just defines which group out of the three must be considered or ignored. The` lack of the `Subject` subattribute indicates that the objects must be considered responsive to the expression if the `Readable` or `Writable` conditions (or any combination of them) evaluate to `True` in ANY of the three groups.

- To find data objects that are read-only by everyone, enter `att:"Access Mode" where (Subject is all (Owner, Group, Others) and Readable is True and Writable is False)`. The subattributes are all `ANDed` together and the `Subject` subattribute lists every possible option with the `all` qualifier.
- To find data objects that are read-only by at least one type of subject, enter `att:"Access Mode" where (Subject is any (Owner, Group, Others) and Readable is True and Writable is False)` or `att:"Access Mode" where (Readable is True and Writable is False)`. Again, the two examples here produce identical results for the same reason as explained in the previous bullet point.
- To find data objects that are read-only by the owner, enter `att:"Access Mode" where (Subject = Owner and Readable is True and Writable is False)`. You care only about the owner in this case, so you explicitly call `Owner` and only `Owner` out.
- To find data objects that owner can either read or write (not necessarily both), enter `att:"Access Mode" where (Subject = Owner and (Readable is True or Writable is True))`.
- To find data objects that can be written by the owner but is read-only by the group, enter `att:"Access Mode" where ((Subject = Owner and Writable = True) and (Subject = Group and Writable is False))`.

- **Parent topic:** [Specification of attribute conditions](#)

Related concepts

- [Specification of attribute conditions](#)

Type of object to match

For any attribute condition, you can specify what object that condition must match.

You can specify `IN object-type` after the expression. For example, in `att: Extension = doc=in system`, the object must be a doc file at the system level. The `condition-scope` can be any of these object types here.

Table 1. Available object types

Available Object Types	Description
all	The matching object can be any type of object. This object is the default.
container	The matching object must be a container object that can have other objects nested inside it. File-archive, email-archive, email, calendar, contact, and task objects are included.
simple	The matching object must not be a container object.
system	The matching object must be a system-level object.
contained	The matching object must be nested within some other container object.

Only one condition-scope can apply to any condition. If you give multiple condition-scopes, the parser issues a warning.

- **Parent topic:** Filter condition reference and expression language

Related concepts

- Filter condition reference and expression language

Multiple matching values for operators

You can give multiple matching values that use the following operators.

- `=`
- `!=`
- `contains`
- `not contains`
- `matches`
- `not matches`
- `starts with`
- `not starts with`
- `ends with`
- `not ends with`

To specify multiple matching values for these operators, give the keyword `any` or `all` after the operator, then a list of values in parentheses. The `any` keyword means that any of the values are a match. The `all` keyword means that all the values must be present. The `all` keyword is only meaningful for multivalued attributes. For example:

Table 1. Keyword examples

Example	Result
`att:"System metadata/Owner" = any (joe, bob, sally)`	Owner can be `joe`, `bob`, or `sally`.
`att:Size = any (1mb, 2mb, 1gb)`	Size can be `1mb`, `2mb`, or `1gb`.
`att: "Modified date" = any (1 week ago, 2 months ago)`	
`att: "Accessed date" = any (2009-09-19, 2009-09-20, 2009-09-21)`	
`att:To = all (joe, bob)`	"To" must contain both `joe` and `bob`

For `any`, the operator must be true for at least one of the values. It is as if the user spelled out `attribute oper val1 OR attribute oper val2 OR ...`

- For `all`, the operator must be true for all the values. It is as if the user spelled out `attribute oper val1 AND attribute oper val2 AND....`
- For the `NOT` various operators, this syntax is the same as using `NOT` around the entire expression and the corresponding positive operator. For example, `att:Size != any( 1mb, 2mb, 2gb )` is the same as `NOT (att:Size = any( 1mb, 2mb, 1gb ))`. The `NOT` matches all documents that do not match the positive operator.
- It is not an error to specify `all` on single-valued attributes. However, if you do, the parser warns you that the condition is always false.

- **Parent topic:** [Filter condition reference and expression language](#)

Related concepts

- [Query expression language](#)

At-least expressions

On occasion, it is helpful to specify that some number of conditions must be true before an object is considered to match. These expressions are called `at-least` expressions because at least `x` subexpressions must match an object before the object matches the overall expression.

The syntax for an at-least expression is to give the keywords `"at least"` followed by a number, and then by a list of expressions in parentheses. For example:

```
at least 2 (att: size=1mb, att: "system metadata/owner"=joe, att:
"object name"=foo.doc, att: "modified date">2009-03-28 )
```

This example specifies that at least two of the four conditions must be true before any object is considered to match the overall expression.

Each subexpression in the list can be arbitrarily complex, but normally they are one condition. Each subexpression in the list must be separated from the others with a comma.

- **Parent topic:** [Filter condition reference and expression language](#)

Related concepts

- [Query expression language](#)

Parent and child conditions

This parent and child conditions represent advanced function that must be undertaken only by experienced administrators.

You might want to filter for objects whose parents or children match certain conditions. For example, find all emails to `Joe` in a `saved-emails.pst` file or find emails to `Joe` that have an attachment that is named `big-contract.pdf`. To specify these kinds of queries, give the conditions for the matching objects as normal and add a prefix `parent` or `child` to the conditions that apply to the parent or child objects. Here are the expressions for the two previous examples:

```
att:"any address"=Joe and parent att: "object name"="saved-
emails.pst"
```

```
att:To=Joe and child att:"object name"="big-contract.pdf"
```

If there is more than one expression that must match the parent or child objects, you must put all the conditions inside parentheses, for example:

```
att:"any address"=Joe and parent(att: "object name"="saved-
emails1.pst" or att:"object name"="saved-emails2.pst")
```

```
att:To=Joe and child (att:"object name"="big-contract.pdf" and
att:"modified date">2009-09-07)
```

It is also possible to specify conditions on the object level of parent or child objects. An object level is the nesting level of an object. System-level objects are at level one , immediately nested objects are at level two, and so on. For example, if you want to find the non-nested emails to `Joe` in `saved-emails.pst`, you can do:

```
att:"any address"=Joe and att:"object level" = 2
and parent
att:"object name"="saved-emails.pst"
```

It specifies that the emails to Joe must be at object level 2 and be inside `saved-email.pst`.

It is also possible to specify the object level of the parent or child object relative to the other. For example, to find the parent object of any file that is named `nested-file.doc`, you can specify it this way:

```
child att: "object name" = "nested-file.doc" parent-level = child-level - 1
```

It means that `nested-file.doc` can itself be at any level, and the query results are the direct parent of those files.

The general form of relative parent-child level conditions is listed here.

```
parent-level operator child-level +|- number
child-level operator parent-level +|- number

parent-level operator child-level +|- number
child-level operator parent-level +|- number
```

- Plus (+) in the +|- indicates that the parent-child must be more deeply nested than the other objects.
- Minus (-) means the parent-child must be less deeply nested than the other objects.

The operator can be any of these comparisons:

Numeric Comparisons for Operators

```
=       !=       <       <=       >       >=
```

- **Parent topic:** [Filter condition reference and expression language](#)

Related concepts

- [Query expression language](#)

No-inheritance conditions

The no-inheritance conditions represent advanced function that must be undertaken only by experienced administrators.

Normally, the search engine allows some objects to inherit the values of their parent objects. Most of the time this behavior is intuitive, but there are times when this behavior must be turned off. When attribute inheritance is not wanted, prefix the expression with `noinheritance`, and attribute inheritance is turned off for that expression. For example, this expression finds emails to `Joe` containing `stock option` in the body or any attachment:

```
att: To=joe and "stock option"
```

However, if you want to find emails to `Joe` with `stock option` in the email body itself, then you must turn off inheritance for that part of the expression:

```
noinheritance ( att: To=joe and "stock option" )
```

- **Parent topic:** [Filter condition reference and expression language](#)

Related concepts

- [Query expression language](#)

Analytics filter expressions

The analytics filter expressions represent advanced function that must be undertaken only by experienced administrators.

Query analytics provides a way to filter objects with certain values out of the query results. These filter expressions are specified as part of the query expression. To specify filtering, you specify the `"filtered"` keyword after the query expression. For example, query: `Stock-OptionMatterBoxJoe filtered`. It uses whatever filters are saved for the `StockOptionMatterBoxJoe` query. The `StockOptionMatterBoxJoe` query must have analytics that are associated with it.

It is also possible to exactly enumerate what filters must be applied to an analytics query. You specify this query by specifying what dimension values must be included or excluded. Here is the general form:

```
query: queryName filtered on include dimension=value, exclude all
dimension=value, exclude unique dimension=value...
```

Filters come in three types: included (`include`), excluded (`exclude all`), or excluded if the values are unique (`exclude unique`). If no dimension value is given for `include`, `exclude all`, or `exclude unique`, then the saved filters of that type is used. For example, if the query expression were `query: queryName filtered` on include `att:extension=doc`, then only the objects with `extension=doc` is in the results. Additionally, any objects that were explicitly excluded with `exclude-all` or `exclude-unique` are filtered out.

The `value` for a dimension-value can also be `none` or `other`.

- The `none` keyword means those objects that do not have that dimension-value. For example, `query: queryName filtered` on include `tag: department=none`. This query would show only those objects that do not have some tag-value from `department` on them. (This case is the same as `tag: department` not appears.)
- The `other` keyword means those objects that have some other value than what is shown in the analytics for the dimension. Analytics typically show the top 10-25 values for any dimension. If you want to see those objects that have a value other than what is shown in the analytics, you can use an expression like `query: queryName filtered on include att: extension=other`. It shows those objects with an extension that is not shown in the analytics.

- **Parent topic:** [Filter condition reference and expression language](#)

Related concepts

- [Query expression language](#)

IBM Discovery Export EDRM XML

IBM® EDRM XML is based on the specification that is described in the EDRM XML Interchange Format Schema Documentation, January 23, 2008. The appendix describes the discovery export feature for EDRM XML.

- [File system directory layout](#)
  The following Discovery Export Parameters have a significant impact on the content of the EDRM XML run export.
- [EDRM XML considerations](#)
  To load data from an IBM EDRM XML run, begin by using the `siq_loadfile_report.csv` to locate all the load files successfully produced in the current run.
- [EDRM XML metadata tags](#)
  Describes valid data types and EDRM XML v1.0 metadata tags.

File system directory layout

The following Discovery Export Parameters have a significant impact on the content of the EDRM XML run export.

- Destination Directory
- Limit Export process to number of Data Objects
- Save a text copy (that is extracted from data files, otherwise known as plaintext)
- Unique ID prefix, Number of Digits, and Pad Zeros

- [Source volume files](#)
  The following examples are based on harvesting the source volume \\qa01\grogers, starting with the directory home for one of two users: `flexus`.

- **Parent topic:** [IBM Discovery Export EDRM XML](#)

Related concepts

- [IBM Discovery Export EDRM XML](#)

Source volume files

The following examples are based on harvesting the source volume \\qa01\grogers, starting with the directory home for one of two users: `flexus`.

Although all directories cannot be seen, this list is the complete list of source volume files and directories:

- `home`
- `home/flexus`
- `home/flexus/backup`
- `home/flexus/backup/carlota.zip`
- `home/flexus/backup/Spaulding.doc`
- `home/flexus/carlota.zip`
- `home/flexus/jbowers_identity.doc`
- `home/flexus/mybcc`
- `home/flexus/mybcc/to_jbowers.msg`
- `home/flexus/Spaulding.doc`
- `home/jbowers`
- `home/jbowers/Copy of identity.doc`
- `home/jbowers/from_flexus.msg`
- `home/jbowers/identity.doc`
- `home/jbowers/philosophy.txt`
- `home/flexus.pst`

Interesting features of the source documents include the following files and directories:

- The file `home/flexus/carlota.zip` contains two text files, `friends.txt` and `enemies.txt`.
- The directory `home/flexus/backup` contains copies of two files that are found in `home/flexus`.
- `home/flexus/mybcc/to_jbowers.msg` contains a saved email message that `flexus` sent to `jbowers`. That message has a copy of `Spaulding.doc` as an attachment.
- `home/flexus.pst` contains two emails, one with subject `Letter from Spaulding` to lawyers and the other with the subject `Philosophy of Life`. Each email further has an attachment; the former has the attachment `Spaulding.doc`, and the latter has the attachment `identity.doc`.
- `home/jbowers/Copy of identity.doc` is a copy of `home/jbowers/identity.doc`. The contents of both are the same as `home/flexus/jbowers_identity.doc`.
- `home/jbowers/from_flexus.msg` contains the same email message as `home/flexus/mybcc/to_jbowers.msg`.

EDRM XML Production provides examples that illustrate how the discovery export feature exports files and metadata.

- EDRM XML production
  The following section describes the details about the files that are generated for an EDRM XML discovery export.

- **Parent topic:** File system directory layout

Related concepts

- File system directory layout

EDRM XML production

The following section describes the details about the files that are generated for an EDRM XML discovery export.

Note: The discovery export action identifies duplicate files that are based on content-based digest and only exports one copy of a file per data server. The duplicate files are identified in the audit trail.

A run creates directories and files on the discovery export volume. All files for a run are located under a discovery export- and run-specific directory on the destination volume under the destination directory.

For each subsequent run, a new subdirectory is created. The directories are named by the run numbers 2, 3, 4, and so on.

Any number of discovery exports can use the same export directory. The output is kept separate by the use of the discovery export name, such as case_123, as the parent directory to the run directories.

This example configuration uses the following information:

Target Volume	auto-P
Destination Directory	exports
Document ID Prefix	NS, number of digits 6, with 0 padding

When a discovery export policy is configured to export EDRM XML, it exports two types of output files:

- Native Files: Copies of original source files such as doc, ppt, txt, and, optionally, plain-text versions of the original source files. The native files are found in the documents subdirectory.
- Run Files: Files that cover the entire run. These files include the EDRM XML run siq load file, the audit trail xml, the EDRM tag definition list, a csv load file report, and a user-ID-to-name csv mapping file. These files are found at the top of the `discovery export run` subdirectory.

The `run` directory contains these files:

- `siq_audit_trail.xml`: The audit trail for the discovery export policy execution. It shows the details of each source and destination object. Essentially, it is an XML encoding of the information available using the IBM® StoredIQ interface to the audit trail.
- `siq_edrm_tag_definitions.csv`: This file is a csv file that lists all the tags in the EDRM XML load file that is produced by IBM StoredIQ and related information. The file has the same contents as the table of metadata tags in [EDRM XML metadata tags](#).
- `siq_load_file-0-0.xml`: This file is an instance of an EDRM XML load file that is produced in the run.

Note: For performance reasons, more than one instance of an EDRM XML load file can be produced during a single run.

- `siq_loadfile_report.csv`: This csv file lists all load files that are produced by the run and status information that states whether they are valid for a load or not. The file also reports the range of document IDs produced in each of the load files and other metadata information such as size and date created.
- `siq_userid_map`: This csv file contains a list of user IDs called uids in the NFS domain and SID in the Windows Share domain with corresponding user names that map to the user

IDs. The user IDs and the user names come from the produced native files for the run.

Under the `documents` subdirectory, an intermediate directory is named `1` that contains the files. To avoid putting too many files and directories in a single subdirectory, the files are placed into numbered subdirectories that contain no more than 10,000 files and directories.

In an EDRM XML discovery export, containers are always expanded to clearly articulate their internal structure. A directory structure is built with each level named after the container and any intermediate containers therein. The path to a data object within the container is achieved by flattening the structure on the file system. You can see this path by looking at the directory structure.

For instance, in this image, the data objects `enemies-0.txt` and `friends-1.txt`, which are contained within the `carlota.zip` container, are flattened out under a directory named `carlota.zip.members`. Their corresponding plain-text files are next to them in the same directory.

Plaintext file is optional, but when enabled, all documents have a corresponding plain-text file. The plain-text files are named according to their SHA-1 hash value and suffix, which is .txt.

The complete list of files and directories for the EDRM XML export for the example is as follows:

```
//qa1/auto-P/exports/edrm global/
//qa1/auto-P/exports/edrm global/1/
//qa1/auto-P/exports/edrm global/1/documents
//qa1/auto-P/exports/edrm global/1/documents/1
//qa1/auto-P/exports/edrm global/1/documents/1/
0x724b937d5a2750038325c8f4b48d1b5dcba77853.txt
//qa1/auto-P/exports/edrm global/1/documents/1/
0xf4dd8fe9b09c834a2033849ff8a5d63f67823787.txt
//qa1/auto-P/exports/edrm global/1/documents/1/jbowers_identity.doc
//qa1/auto-P/exports/edrm global/1/documents/1/Spaulding.doc
//qa1/auto-P/exports/edrm global/1/documents/1/
0x8d1e7cef561d7d81585af82d9f79f8a703bd96ee.txt
//qa1/auto-P/exports/edrm global/1/documents/1/
0xd7eb7675f2d201c79cf821debaaeba5a208cba29.txt
//qa1/auto-P/exports/edrm global/1/documents/1/from_flexus.msg
//qa1/auto-P/exports/edrm global/1/documents/1/philosophy.txt
//qa1/auto-P/exports/edrm global/1/documents/1/carlota.zip.members
//qa1/auto-P/exports/edrm global/1/documents/1/carlota.zip.members/
0x22e2105930a1ddc75b9bc9dc3d5408c3b0bac8b1.txt
//qa1/auto-P/exports/edrm global/1/documents/1/carlota.zip.members/
0xfce44897add0b77c25fea85118fe47164dc828d4.txt
//qa1/auto-P/exports/edrm global/1/documents/1/carlota.zip.members/enemies-
0.txt
//qa1/auto-P/exports/edrm global/1/documents/1/carlota.zip.members/
friends-1.txt
//qa1/auto-P/exports/edrm global/1/documents/1/from_flexus.msg.members
//qa1/auto-P/exports/edrm global/1/documents/1/from_flexus.msg.members/
0x1564f56aecf315cdecde563605fb151356f8dcc2.txt
//qa1/auto-P/exports/edrm global/1/documents/1/from_flexus.msg.members/
Spaulding-1.doc
//qa1/auto-P/exports/edrm global/1/documents/1/flexus.pst.members
//qa1/auto-P/exports/edrm global/1/documents/1/flexus.pst.members/Top
of Personal Folders.members/
//qa1/auto-P/exports/edrm global/1/documents/1/flexus.pst.members/Top
of Personal Folders.members/Story.members/
//qa1/auto-P/exports/edrm global/1/documents/1/flexus.pst.members/Top
of Personal Folders.members/Story.members/Production.members
//qa1/auto-P/exports/edrm global/1/documents/1/flexus.pst.members/Top
of Personal Folders.members/Story.members/Production.members/
flexus.members
//qa1/auto-P/exports/edrm global/1/documents/1/flexus.pst.members/Top
of Personal Folders.members/Story.members/Production.members/
flexus.members/0x277bab09a965f3840cd8a14300e0fccf59f6f5a9.txt
//qa1/auto-P/exports/edrm global/1/documents/1/flexus.pst.members/Top
of Personal Folders.members/Story.members/Production.members/
```

```
flexus.members/Philosophy of Life-7.msg
//qa1/auto-P/exports/edrm global/1/documents/1/flexus.pst.members/Top
of Personal Folders.members/Story.members/Production.members/
flexus.members/bcc.members/
//qa1/auto-P/exports/edrm global/1/documents/1/flexus.pst.members/Top
of Personal Folders.members/Story.members/Production.members/
flexus.members/bcc.members/Letter from Spaulding to his lawyers-9.msg
//qa1/auto-P/exports/edrm global/1/documents/1/flexus.pst.members/Top
of Personal Folders.members/Story.members/Production.members/
flexus.members/bcc.members/
0x2a5bccfda550233ce9bac5d1f25e52590702a539.txt
//qa1/auto-P/exports/edrm global/1/documents/1/flexus.pst.members/Top
of Personal Folders.members/Story.members/Production.members/
flexus.members/bcc.members/.siq_overflow
//qa1/auto-P/exports/edrm global/1/documents/1/flexus.pst.members/Top
of Personal Folders.members/Story.members/Production.members/
flexus.members/bcc.members/.siq_overflow/
0xd7eb7675f2d201c79cf821debaaeba5a208cba29.txt
//qa1/auto-P/exports/edrm global/1/documents/1/flexus.pst.members/Top
of Personal Folders.members/Story.members/Production.members/
flexus.members/bcc.members/.siq_overflow/Spaulding-8.doc
//qa1/auto-P/exports/edrm global/1/documents/1/flexus.pst.members
//qa1/auto-P/exports/edrm global/1/documents/1/flexus.pst.members/Top
of Personal Folders.members/Story.members/Production.members/
flexus.members/Philosophy of Life-7.mail.members
//qa1/auto-P/exports/edrm global/1/documents/1/flexus.pst.members/Top
of Personal Folders.members/Story.members/Production.members/
flexus.members/Philosophy of Life-7.mail.members/
0xf4dd8fe9b09c834a2033849ff8a5d63f67823787.txt
//qa1/auto-P/exports/edrm global/1/documents/1/flexus.pst.members/Top
of Personal Folders.members/Story.members/Production.members/
flexus.members/Philosophy of Life-7.mail.members/identity-6.doc
//qa1/auto-P/exports/edrm global/1/siq_audit_trail.xml
//qa1/auto-P/exports/edrm global/1/siq_edrm_tag_definitions.csv
//qa1/auto-P/exports/edrm global/1/siq_load_file-0-0.xml
//qa1/auto-P/exports/edrm global/1/siq_loadfile_report.csv
//qa1/auto-P/exports/edrm global/1/siq_userid_map.csv
```

Note: Duplicate Files: Although 20 data objects were originally identified for export, only 11 native files were eventually produced, along with corresponding plaintext files. Duplicates are not copied and are marked in the policy audit trail. The EDRM XML load file lists all the exported documents and related metadata according to standard EDRM specifications. Any exceptions are noted in the IBM StoredIQ EDRM metadata tags list.

- **Parent topic:** [Source volume files](#)

Related concepts

- [Source volume files](#)

EDRM XML considerations

To load data from an IBM® EDRM XML run, begin by using the `siq_loadfile_report.csv` to locate all the load files successfully produced in the current run.

Load these files one by one, preferably in the order of increasing document IDs mentioned in the report, into the target application. All file references in each of the load files are relative to the location of the load files themselves. Therefore, it is necessary to ensure that this relative location is maintained for a successful load operation.

The EDRM XML produced in each load file is compliant with the EDRM XML Version 1.0 XML specification and can successfully validate against the EDRM XML Validation Tool version 1.0.

- [XML root information](#)
  The name of the discovery export is used as the CaseID that describes the load file itself. The discovery export description, which is entered from IBM StoredIQ application interface, is also published in the Description attribute.
- [Batch information](#)
  The batch element is associated with a single run in the application. By this definition, there is only one batch element in every load file produced. The name of the batch is the run number.
- [Document entries in the load file](#)
  Intermediate archives inside file archives are only present in the `export all members case` or if these intermediate archives are responsive to the query used to drive the run.
- [Tag information](#)
  In addition to the EDRM XML standard metadata tags, the following information is also present in the load files in the form of IBM StoredIQ EDRM XML tags.
- [Relationship between documents](#)
  The relationships that are described between documents are as follows.

- **Parent topic:** [IBM Discovery Export EDRM XML](#)

Related concepts

- [IBM Discovery Export EDRM XML](#)
- [XML root information](#)
- [Batch information](#)
- [Document entries in the load file](#)
- [Tag information](#)
- [Relationship between documents](#)

XML root information

The name of the discovery export is used as the CaseID that describes the load file itself. The discovery export description, which is entered from IBM® StoredIQ application interface, is also published in the Description attribute.

- **Parent topic:** [EDRM XML considerations](#)

Related concepts

- [EDRM XML considerations](#)

Batch information

The batch element is associated with a single run in the application. By this definition, there is only one batch element in every load file produced. The name of the batch is the run number.

- **Parent topic:** EDRM XML considerations

Related concepts

- EDRM XML considerations

Document entries in the load file

Intermediate archives inside file archives are only present in the `export all members` case or if these intermediate archives are responsive to the query used to drive the run.

Email archives are never expressed in the XML as documents. Only the emails that are produced from these email archives are described in the XML.

- **Parent topic:** [EDRM XML considerations](#)

Related concepts

- [EDRM XML considerations](#)

Tag information

In addition to the EDRM XML standard metadata tags, the following information is also present in the load files in the form of IBM® StoredIQ EDRM XML tags.

Tag information	Description
Responsive	Whether IBM StoredIQ deemed the document to be responsive or not, based on the query that drives the policy.
Internal path	Path leading to the object that starts from below the file system object, if it was contained within a container.
Internet Message ID	Unique ID identifying every email.

- **Parent topic:** EDRM XML considerations

Related concepts

- EDRM XML considerations

Relationship between documents

The relationships that are described between documents are as follows.

- Attachment: Direct attachments to an email are related to the email.
- Container: A member of a file archive is related to its immediate parent only if the immediate parent is exported in the load file and has a DocID of its own. In the absence of the immediate parent, the member is related to the file system-level file archive that is always present in the load file.

If an email has archives or other emails that are attached to it, the members of the archives or emails are not related to the top email. Instead, they are related to the archives or emails to which they directly belong.

There is no container relationship between an email in an email archive and the email archive itself because the email archive is not produced as a document in the load file.

- **Parent topic:** [EDRM XML considerations](#)

Related concepts

- [EDRM XML considerations](#)

EDRM XML metadata tags

Describes valid data types and EDRM XML v1.0 metadata tags.

The following data types are valid for IBM® StoredIQ EDRM XML tags.

Table 1. Valid data types for EDRM XML Tags

Type	Comment
Text	Text (255 characters or fewer)
Integer	32-bit integer
DateTime	YYYY-MM-DDThh:mm:ss.sTZD
Decimal	Standard float
Boolean	{True/False}
LongInteger	64-bit integer
LongText	Text (32,767 characters or fewer)

EDRM-standard tag names are prefaced with a #. IBM StoredIQ proprietary names are prefaced with an SIQ-. In certain cases, IBM StoredIQ altered the data types for some tags from what is prescribed in the original EDRM XML v1.0 metadata tags list to support a larger range of values. It does not violate EDRM compatibility.

Table 2. EDRM data tags

Document Type	Metadata Property	Tag Name	EDRM Data Type	Data Type	Description
All	Language	#Language	Text	Text	This document is the ISO 639-1 (two-character) primary language of the document. If omitted, then English (EN) can be assumed.
	Responsive-ness	SIQ-Responsive-ness	N/A	Boolean	It indicates whether IBM StoredIQ deemed the object to be responsive for the discovery export run.
	Internal Path	SIQ- Internal Path	N/A	LongText	This path is the path that leads to an object contained inside a container, relative to the file system-level container.
	URL (SharePoint)	SIQ-SP-URL	N/A	LongText	It represents information about the server-relative URL for the specified Windows SharePoint Services object.
	Version (SharePoint)	SIQ-SP-Version	N/A	Decimal	It represents the version string that is associated with the user interface of the specified Windows SharePoint Services object.
	Author (SharePoint)	#Author	N/A	Text	It represents the specified author of the Windows SharePoint Services object.
	Title (SharePoint)	#Title	N/A	Text	It represents information about the occupational title of a specified

				Windows SharePoint Services contact object.
Checked out to (SharePoint)	SIQ-SP-Checked-OutTo	N/A	Text	It represents the designated user who checked out the Windows SharePoint Services object by using version control.
Editor (SharePoint)	SIQ-SP-Editor	N/A	Text	This information is the editor name or information that is associated with a person who is referenced by a Windows SharePoint Services contact object.
Approval status (SharePoint)	SIQ-SP-Moderation Status	N/A	Text	It provides information about the moderation status of the specified Windows SharePoint Services object, one of Approved, Denied, Draft, Pending, or Scheduled.
Content type (SharePoint)	SIQ-SP-ContentType	N/A	Text	It is the content type of the specified Windows SharePoint Services object.
Check in comment (SharePoint)	SIQ-SP-Checkin-Comment	N/A	LongText	It is the check in comments of the specified Windows SharePoint Services object.
Publishing level (SharePoint)	SIQ-SP-Publishing-Level	N/A	Text	It is the version-control alert level information of the specified Windows SharePoint Services object, one of Published, Draft, or Checkout.
Managed keywords (SharePoint)	#Keywords	N/A	Text	It describes all managed keywords that are applied to specified Windows SharePoint Services object. Multiple keywords are separated into their individual tags.
Description	SIQ-SP-Description	N/A	LongText	It provides alternative text for specified Windows SharePoint Services object.
Keywords (SharePoint)	#Keywords	N/A	Text	It is the keyword summary of the specified Windows SharePoint Services object. Multiple keywords are separated into their individual tags.
Comments (SharePoint)	#Comments	N/A	Text	It provides comments that are associated with the specified Windows SharePoint Services object. Multiple

Virus status (SharePoint)	SIQ-SP-VirusStatus	N/A	Text	comments are separated into their individual tags. It is the virus-scan status of a specified Windows SharePoint Services object: • Clean • Virus reported • Virus reported; can be removed • Virus previously reported; removed since • Virus reported; clean failed • Virus reported; delete recommended • Timeout last processing for Virus
Declared record time (SharePoint)	SIQ-SP-Declared RecordDate	N/A	DateTime	It is the date at which the item was declared a record.
Original expiration date (SharePoint)	SIQ-SP-Record Expiration-Date	N/A	DateTime	It is the date at which the retention on the object is set to expire.
Routing rule description (SharePoint)	SIQ-SP-Routing-Rule-Description	N/A	LongText	This description is associated with the object that matched a routing rule.
Litigation hold (SharePoint)	SIQ-SP-Litigation-Holds	N/A	Text	It is the list of all legal holds placed on an object. Multiple litigation holds are separated into their individual tags.
Record restrictions (SharePoint)	SIQ-SP-Record-Restrictions	N/A	Text	Describes the restrictions that are applied to the record by the policies in the Records Center: • None • BlockDelete • BlockEdit
Blog or list name (SharePoint)	SIQ-SP-BlogOr ListName	N/A	Text	It is the name of the blog or list to which the object belongs.
Approver comments (SharePoint)	SIQ-SP-Approver Comments	N/A	LongText	These comments are the moderation comments of the specified Windows SharePoint Services weblog object.
	SIQ-SP- CUSTOM-Attribute			It provides all custom

Custom properties (SharePoint)	Name, where Attribute-Name is the name of the Custom Attribute with white spaces replaced by '_'	N/A	LongText	to provide all custom properties that are applied to specified Windows SharePoint Services object.
Post Category (SharePoint)	#Category	N/A	Text	This category is associated with the post. Categories are defined as managed objects on the blog site for use with posts.
Published Date (SharePoint)	SIQ-SP-Published-Date	N/A	DateTime	The date on which the post was published.
Publishing Contact (SharePoint)	SIQ-SP-Publishing Contact	N/A	Text	The user name of the contact.
Wiki Page Categories (SharePoint)	#Category	N/A	Text	They are the managed metadata values that describe categories for the wiki page. Multiple categories are separated into their individual tags.
Publishing Contact Email (SharePoint)	SIQ-SP-Publishing Contact- Email	N/A	Text	The email address of the contact.
Wiki Page Comment (SharePoint)	#Comments	N/A	LongText	This comment describes the Wiki page. There can be only one.
Publishing Contact Name (SharePoint)	SIQ-SP-Publishing-Contact-Name	N/A	LongText	The friendly name of the contact.
Average Rating (SharePoint)	SIQ-SP-Average-Rating	N/A	Decimal	The average of all ratings that are applied to the object.
Rating Count (SharePoint)	SIQ-SP-Rating-Count	N/A	Integer	The total number of ratings that are applied to this object.
Message ID (SharePoint)	SIQ-SPMessage-ID	N/A	Text	The unique message ID (similar to those on an email) associated with a post on a discussion board.
Discussion Last Updated (SharePoint)	SIQ-SP-Discussion-Last-Updated	N/A	DateTime	The time stamp for when the discussion was last updated.
Thread Index (SharePoint)	SIQ-SP-Thread-Index	N/A	LongText	The unique index into the discussion to help determine the position of the post in the threading hierarchy of the discussion. It identifies a field that

Status Bar (SharePoint)	SIQ-SP-StatusBar	N/A	DateTime	contains information about the status bar for a specified SharePoint Foundation object.
Discussion Title (SharePoint)	SIQ-SP-Discussion-Title	N/A	Text	The title of the post for which the reply was written.
Recurrence Event (SharePoint)	SIQ-SP-Recurring-Event	N/A	Boolean	It indicates whether an event is a recurring event.
All Day Event Flag (SharePoint)	SIQ-SP-AllDay-Event	N/A	Boolean	It indicates whether the Event is an all-day Event or not.
Event Date (SharePoint)	SIQ-SP-EventDate	N/A	DateTime	The start date for the event.
Event Location (SharePoint)	SIQ-SP-Event-Location	N/A	Text	The location for the event.
Event Category (SharePoint)	#Category	N/A	Text	The category that is associated with the event.
Event Type (SharePoint)	SIQ-SP-EventType	N/A	Text	The event type, which has one of the following values: • Single Event • Recurring Event • Deleted Recurring Event • Edited Recurring Event
Event End Date (SharePoint)	SIQ-SP-EventEnd-Date	N/A	DateTime	The end date for the event.
Event Duration (SharePoint)	SIQ-SP-Duration	N/A	Integer	The duration in seconds for the event.
Task Due Date (SharePoint)	SIQ-SP-TaskDue-Date	N/A	DueTime	The due date for a task.
Assigned To (SharePoint)	SIQ-SP- Task AssignedTo	N/A	Text	The user name to which the task was assigned.
Task Priority (SharePoint)	SIQ-SP-TaskPriority	N/A	Text	The priority of the task, which is usually one of " (1) High," "(2) Normal," or "(3) Low." The Administrator has the option of editing the choices.
Task Start Date (SharePoint)	SIQ-SP-TaskStart-Date	N/A	DateTime	The start date for a task.
				It is either the task status or the issue status. These options for task status: • "Not Started" • "In Progress"

Status (SharePoint)	SIQ-SP-TaskOr-Issue-Status	N/A	Text	"Completed" "Deferred" "Waiting on someone else" These options are for issue status: "Active" "Resolved" "Closed"
Predecessors (SharePoint)	SIQ-SP- Task Predecessors	N/A	Text	The names of tasks that have a defined precedence to the task. Multiple predecessors are separated into their individual tags.
Percent Complete (SharePoint)	SIQ-SP-TaskPercent Complete	N/A	Integer	The completion percentage of the task as an integer.
Related Issues (SharePoint)	SIQ-SP-Related-Issues	N/A	Text	Issues that are marked as related to an issue. Multiple issues are separated into their individual tags.
Issue ID (SharePoint)	SIQ-SP-IssueID	N/A	LongInteger	The unique ID for the issue in the list.
Contact Full Name (SharePoint)	SIQ-SP-Contact FullName	N/A	Text	The full name of the contact.
Work Country (SharePoint)	SIQ-SP-Contact-Work-Country	N/A	Text	The country of the contact.
Contact Web Page (SharePoint)	SIQ-SP-Contact WebPage	N/A	Text	The web page URL of the contact with its associated description.
Contact Email (SharePoint)	SIQ-SP-Contact-Email	N/A	Text	The email address of the contact.
Contact Job Title (SharePoint)	SIQ-SP-Contact-JobTitle	N/A	Text	The job title of the contact.
Work City (SharePoint)	SIQ-SP-Contact WorkCity	N/A	Text	The city of the contact.
Contact Company (SharePoint)	#Company	N/A	Text	The company of the contact.
Work ZIP (SharePoint)	SIQ-SP-Contact-WorkZip	N/A	Text	The postal code of the contact.
Contact First Name (SharePoint)	SIQ-SP-Contact-FirstName	N/A	Text	The given name of the contact.
Contact Business Address	SIQ-SP-Contact-Work-Address	N/A	LongText	The work address of the contact.

(SharePoint)				
Work State (SharePoint)	SIQ-SP-Contact-WorkState	N/A	Text	The work state of the contact.
Contact Business Phone (SharePoint)	SIQ-SP-Contact-WorkPhone	N/A	Text	The work phone of the contact.
Contact Home Phone (SharePoint)	SIQ-SP-Contact HomePhone	N/A	Text	The home phone of the contact.
Work Fax (SharePoint)	SIQ-SP-Contact-WorkFax	N/A	Text	The work fax number of the contact
Contact Mobile Phone (SharePoint)	SIQ-SP-Contact-Mobile- Phone	N/A	Text	The mobile number of the contact.
Announce-ment Expiry Date (SharePoint)	SIQ-SP-Announce-ment-Expiry	N/A	DateTime	The expiry date set for the announcement, if any.
Survey Completed (SharePoint)	SIQ-SP-Survey Completed	N/A	Boolean	It indicates whether the survey response was completed.
Link URL (SharePoint)	SIQ-SP-LinkURL	N/A	LongText	The URL that forms the basis of the link, plus an optional description. In the absence of the description, the URL is repeated.
ProfileId (SharePoint)	SIQ-SP-User-ProfileId	N/A	Text	The unique identifier for the user profile.
Contact last name (SharePoint)	SIQ-SP-UserLast-Name	N/A	Text	The surname of the user.
Preferred Name (SharePoint)	SIQ-SP- User Preferred Name	N/A	Text	The preferred name of the user.
Department (SharePoint)	SIQ-SP-User Department	N/A	Text	The department of the user.
Manager (SharePoint)	SIQ-SP-User-Manager	N/A	Text	The manager of the user.
User name (SharePoint)	SIQ-SP-UserName	N/A	Text	The user name that is associated with a SharePoint user profile.
Quicklinks (SharePoint)	SIQ-SP-UserQuick links	N/A	Text	The Quick Links entry that is configured on a user profile. Multiple links are separated into their individual tags.
Public URL (SharePoint)	SIQ-SP-UserProfile-PublicUrl	N/A	LongText	The public URL to access a SharePoint user profile.
Member Of (SharePoint)	SIQ-SP-UserMem-berOf	N/A	Text	It indicates the sites for which a SharePoint User is a member. Multiple sites are separated into

				their individual tags.
Dotted-line Manager (SharePoint)	SIQ-SP-UserDotted LineManager	N/A	Text	The dotted-line manager for the user.
Ask me about (SharePoint)	SIQ-SP-UserAsk MeAbout	N/A	Text	The query topics for which a SharePoint User is open or available. Multiple topics are separated into their individual tags.
Office Communica-torAddress (SharePoint)	SIQ-SP-UserOffice-Communi-catorAddress	N/A	Text	The office communicator address for user.
Hire Date (SharePoint)	SIQ-SP-UserHire-Date	N/A	DateTime	The hire date for SharePoint user.
Office (SharePoint)	SIQ-SP-UserOffice	N/A	Text	The office information for SharePoint user.
Office Location (SharePoint)	SIQ-SP-UserOffice Location	N/A	Text	The office-location information for SharePoint user.
Assistant (SharePoint)	SIQ-SP-User Assistant	N/A	Text	The assistant for the user.
Past Projects (SharePoint)	SIQ-SP-User PastProjects	N/A	Text	The list of past projects that are conducted by SharePoint user. Multiple projects are separated into their individual tags.
Skills (SharePoint)	SIQ-SP-UserSkills	N/A	Text	The skills that are listed by SharePoint user. Multiple skills are separated into their individual tags.
Schools (SharePoint)	SIQ-SP-User-Schools	N/A	Text	The schools that are listed by SharePoint user. Multiple schools are separated into their individual tags.
Birthday (SharePoint)	SIQ-SP-User Birthday	N/A	Text	The birth date of SharePoint user (MM-DD).
Status Message (SharePoint)	SIQ-SP-UserStatus Message	N/A	Text	The most recent status message on SharePoint user profile.
Interests (SharePoint)	SIQ-SP-User Interests	N/A	Text	The interests that are listed by SharePoint user. Multiple interests are separated into their individual tags.
Organization (SharePoint)	SIQ-SP-User-Organization	N/A	Text	The organization to which the SharePoint user belongs.
Colleague (SharePoint)	SIQ-SP-User Colleague	N/A	Text	The colleagues for a user. Multiple colleagues are separated into their individual tags.
				The sender of the

Message	From	#From	Text	LongText	message in the EDRM-recommended RFC2821 format (either "your-name@domain.com" or "Your Name <your-name@domain.xyz>") Note: Email addresses can have 64-character account names and 255-character domain names.
	To	#To	Text	LongText	The recipient of the message in the same format as "From." Multiple recipients are separated into their individual tags.
	CC	#CC	Text	LongText	The carbon-copied recipient of the message in the same format as "From." Multiple recipients are separated into their individual tags.
	BCC	#BCC	Text	LongText	The blind carbon-copied recipients of the message in the same format as "From." Multiple recipients are separated into their individual tags.
	Subject	#Subject	Text	Text	The subject of the message.
	Full Email Header	#Header	Text	LongText	The message header of the message. Note: The header can be longer than 64,000 characters; however, it is truncated in that case.
	Sent Date	#DateSent	DateTime	DateTime	The date the message was sent.
	Received Date	#DateReceived	DateTime	DateTime	The date the message was received.
	Has Attachments	#HasAttachments	Boolean	Boolean	It indicates whether the email has attachments.
	Attachment Count	#AttachmentCount	Integer	Integer	The number of attachments the email has.
	Attachment Names	#AttachmentNames	Text	LongText	The concatenated list of attachment names that are separated by semicolons. Note: The list of attachment names can be longer than 255 characters.
	Read Flag	#ReadFlag	Boolean	Boolean	It indicates if the email was read.
	Importance	#ImportanceFlag	Boolean	Boolean	Indicates whether the email was sent with high importance.

	Message Class	#Message-Class	Text	Text	The (Outlook) message class.
	Flag Status	#FlagStatus	Text	Text	The (Outlook) flag status.
	Internet MessageID	SIQ-Internet MessageID	N/A	Text	The universal Internet message ID assigned to the email. Note: The message ID is usually in the form unique_id@domain.
File	FileName	#FileName	Text	LongText	The name of the original file. Note: File names can be longer than 255 characters.
	File Extension	#File-Extension	Text	Text	The extension of the original file.
	File Size	#FileSize	LongInteger	LongInteger	The size of the original file in bytes.
	Creation Time	#Date-Created	DateTime	DateTime	The date the file was created.
	Last Accessed Time	#Date-Accessed	DateTime	DateTime	The date the file was last accessed.
	Last Modified Time	#Date-Modified	DateTime	DateTime	The date the file was last modified.
	Date Last Printed	#Date-Printed	DateTime	DateTime	The date the file was last printed.
	Title	#Title	Text	Text	The (Office) document title.
	Subject	#Subject	Text	Text	The (Office) document subject.
	Author	#Author	Text	Text	The (Office) document author.
	Company	#Company	Text	Text	The (Office) document company.
	Category	#Category	Text	Text	The (Office) document category.
	Keywords	#Keywords	Text	LongText	The (Office) document keywords.
	Comments	#Comments	Text	LongText	The (Office) document comments. Note: All Office document metadata can be longer than 255 characters.

- **Parent topic:** [IBM Discovery Export EDRM XML](#)

Related concepts

- [IBM Discovery Export EDRM XML](#)

Metadata by attribute type

This reference guide describes metadata that is organized by attribute type.

The column headings in the attribute tables are listed as follows:

Table 1. Column headings in attribute table

Name	Name of the attribute.
Dense	Dense or Sparse, signifying whether an attribute occurs on an object.
Multiple	Single or Multiple Instance, signifying whether the attribute can occur multiple times (Y) or one time (N).
Type	Indicates whether the attribute is a string, integer, Boolean, or date.
Normalized	Indicates whether the pipeline has a normalization liability when you are extracting the attribute. In general, normalizing the format of an attribute produces richer results.
Description	Provides a description of the attribute (optional).

- System metadata attributes
  This table describes system metadata for Windows Share and NFS attributes.
- Documentum server metadata attributes
  Documentum attributes are listed in the following table.
- Centera server metadata attributes
  Metadata for Centera server attributes is listed in the following table.
- Dell DX Object Storage Platform metadata attributes
  Metadata for Dell DX Object Storage Platform attributes is listed in the following table.
- IBM Information Archive Server metadata attributes
  This table describes metadata for IBM Information Archive Server attributes.
- Hitachi HCAP Server metadata attributes
  This table describes metadata for Hitachi HCAP Servers attributes.
- NetApp SnapLock Server metadata attributes
  This table describes metadata for NetApp SnapLock Server attributes.
- Exchange Server metadata attributes
  This table describes metadata for Exchange Server attributes.
- Discovery Accelerator metadata attributes
  This table describes metadata for Discovery Accelerator attributes.
- FileNet metadata attributes
  This table describes metadata for FileNet attributes.
- Email metadata attributes
  This table describes metadata for email attributes.
- Data object property attributes
  This table describes Data Object Property Attributes. Data object property attributes are stored inside the data object.
- SharePoint attributes
  This table describes SharePoint metadata attributes.
- Calendar metadata attributes
  This table describes metadata attributes for calendar items.
- Contact metadata attributes
  This table describes metadata attributes for contacts.
- Digital signatures metadata attributes
  This table describes attributes that are used in IBM® StoredIQ data analysis.
- Processing metadata attributes
  This table describes attributes that are used in data processing.
- Task metadata attributes
- Attribute value types
  This table describes the value and instance-based expressions that can be used with a query condition, depending on the type of attribute.
- IBM Lotus Notes metadata attributes
  This table describes metadata for IBM Lotus Notes attributes.
- Enterprise vault metadata attributes
  This table describes metadata for enterprise vault attributes.

- [Jive metadata attributes](#)
  This table describes attributes that are used in Jive items.
- [Chatter metadata attributes](#)
  This table describes attributes that are used in Chatter items

System metadata attributes

This table describes system metadata for Windows Share and NFS attributes.

Table 1. System metadata attributes

Name	Dense	Multiple	Type	Normalized	Description
Accessed Date	Y	N	Date	Y	A time stamp that indicates the last accessed time of the file (`YYYY-MM-DD hh:mm:ss`)
Container path	Y	N	String	No	Path of an object relative to the container it lives in
Content type description	Y	N	String	No	Derived from the content of an object; a description of the application with which the object was created
Content type mismatch	Y	N	Boolean	No	True if the object's extension does not match the content type that is derived from the object
Created Date	Y	N	Date	Y	A time stamp that indicates the creation time of the file (`YYYY-MM-DD hh:mm:ss`)
Data object type (is known)	Y	N	Boolean	N	True if the data type of the object can be mapped to a known file extension in the system
Extension	Y	N	String	N	The extension of the data object
					The name of the group owner of the file as resolved by a user directory service. If the group

					identifier might not be resolved, perhaps
Group	Y	N	String	N	because no directory service was configured or an error occurred, the field contains the string representation of the group ID.
Modified Date	Y	N	Date	Y	A time stamp that indicates the last modified time of the file (`YYYY-MM-DD hh:mm:ss`)
Object category	Y	N	String	N	Object general file type
Object ID	Y	N	String	N	Name of the data object (file) located within the context of the specified volume
Object name	Y	N	String	N	Unique data object identifier that provides the data object's "handle" to the system where the data object resides.
Owner	Y	N	String	N	The user name of the file owner as resolved by a user directory service. If the user identifier might not be resolved, perhaps because no directory service was configured or an error occurred, the field contains the string representation

Owner (is known)	Y	N	Boolean	N	True if the object's owner attribute can be mapped to a known user name in the system
Size	Y	N	Int	N	The size of the file in bytes
Size (on disk)	Y	N	Int	N	The amount of physical cluster space a data object occupies
Source container path	Y	N	String	N	The container path of the object on the source volume
Source Volume	Y	N	String	N	The original primary source location
System Path	Y	N	String	N	All attributes that reference the data object's location

- **Parent topic:** [Metadata by attribute type](#)

Related concepts

- [Metadata by attribute type](#)

Documentum server metadata attributes

Documentum attributes are listed in the following table.

Name	Dense	Multiple	Type	Normalized	Description
ACL Domain					The repository owner name or user name that created the Access Control List
ACL Name					Name uniquely identifying the Access Control List on the Documentum server
Authors	N	N	String	N	The creator of the data object, as shown in Document Properties.
Checked out to	N	N	String	N	Name of person the data object is checked out to for editing.
Creation Date	Y	N	Date	Y	A time stamp that indicates the creation time of the file.
File Size	Y	N	Int	N	The size of the file in bytes.
Format					Application type of the data object's content.
Group	Y	N	String	N	The group name of the file owner as resolved by a Documentum server.
Immutable	N	N	String	N	Denotes whether a data object can be change or deleted.
Keywords	N	N	String	N	Keywords that are associated with the data object.
Lock Date	Y	N	Date	N	Date that machine was locked (YYYY-

					MM-DD)
Lock Machine	N	N	String	N	Name of the client machine on which users are working when they lock an object (by a check out or branch operation). If the server is unable to resolve the name, r_lock_attribut e is set to the client machine's IP address instead.
Modified Date	Y	N	Date	Y	A time stamp that indicates the modification time of the file.
ObjectID	Y	N	String	N	Unique data object identifier provides the data object's "handle" to the system where the data object is found.
Object Name	Y	N	String	N	Name of the data object (file) located within the context of the specified volume.
Owner	Y	N	String	N	The user name of the file owner as resolved by a Documentum server.
Parent document ID	Y	N	String	N	Data object parent ID
Parent document name	Y	N	String	N	Data object parent name
Retain until	Y	N	Date	N	Date data object is released from retention. Set on the Documentum server not through the

					application. Date data object was placed on hold.
Retention date	Y	N	Date	N	Set on the Documentum server not through the application.
RPS markup name					Retention Policy Services retention policy markup name
RPS policy name					Retention Policy Services retention policy name
Subject	N	N	String	N	The subject of the data object. By default, it is the subject of the Office document as shown in Document Properties. Configurable by the user when it is checked in.
Title	N	N	String	N	The title of the data object. By default, it is the title of the Office document as shown in Document Properties. Configurable by the user when it is checked in.
Version	Y	N	String	N	The version of the data object from a Documentum server perspective.

Note: Access Time not Reset: When harvesting Documentum volumes, IBM® StoredIQ does not reset the access time back to the original value before the harvest. The reason is due to a known issue. Resetting the volume access time on IBM StoredIQ causes Documentum to reset its files' modification time and date to the time and date of the harvest.

- **Parent topic:** [Metadata by attribute type](#)

Related concepts

- [Metadata by attribute type](#)

Centera server metadata attributes

Metadata for Centera server attributes is listed in the following table.

Table 1. Metadata for Centera server attributes

Name	Dense	Multiple	Type	Normalized	Description
Access Date	Y	N	Date	Y	A time stamp that indicates the last accessed time of the data object.
Creation Date	Y	N	Date	Y	A time stamp that indicates the creation time of the data object.
File Extension	Y	N	String	N	The data object's extension.
File Name	Y	N	String	N	Name of the data object located within the context of the specified volume.
File Path	Y	N	String	N	All attributes that reference the data object's location.
File Size	Y	N	Int	N	The size of the data object in bytes.
Litigation Hold	N	Y	String	N	Name of the litigation hold that is applied to the data object. There can be multiple litigation holds applied to a data object.
Modified Date	Y	N	Date	Y	A time stamp that indicates the modification time of the data object.
Object ID	Y	N	String	N	Unique data object identifier provides the data object's "handle" to the system where

						the data object is found.
Retention Class	N		N	String	N	Name of the retention class that is applied to the data object.
Retention Period	N		N	String	N	The length of time a data object is set to be retained.
Retention Policy Name	N		N	String	N	The name of the policy that last set retention for a data object.
Retention Tag	N		N	String	N	Tag defined in copy policy that is retained in the copied and harvested data object.
Source Volume	Y		N	String	N	The originating server and share name from which an object was copied to Centera.

- **Parent topic:** [Metadata by attribute type](#)

Related concepts

- [Metadata by attribute type](#)

Dell DX Object Storage Platform metadata attributes

Metadata for Dell DX Object Storage Platform attributes is listed in the following table.

Table 1. Metadata fro Dell DX Object Storage Platform attributes

Name	Dense	Multiple	Type	Normalized	Description
Access Date	Y	N	Date	Y	A time stamp that indicates the last accessed time of the data object.
Creation Date	Y	N	Date	Y	A time stamp that indicates the creation time of the data object.
File Extension	Y	N	String	N	The data object's extension.
File Name	Y	N	String	N	Name of the data object located within the context of the specified volume.
File Path	Y	N	String	N	All attributes that reference the data object's location.
File Size	Y	N	Int	N	The size of the data object in bytes.
Litigation Hold	N	Y	String	N	Name of the litigation hold that is applied to the data object. There can be multiple litigation holds applied to a data object.
Modified Date	Y	N	Date	Y	A time stamp that indicates the modification time of the data object.
Object ID	Y	N	String	N	Unique data object identifier provides the data object's

"handle" to the system where |

					the data object is found.
Retention Period	N	N	String	N	The length of time a data object is set to be retained.
Retention Policy Name	N	N	String	N	The name of the policy that last set retention for a data object.
Retention Tag	N	N	String	N	Tag defined in copy policy that is retained in the copied and harvested data object.
Source Volume	Y	N	String	N	The originating server and share name from which an object was copied to the Dell server.

- **Parent topic:** [Metadata by attribute type](#)

Related concepts

- [Metadata by attribute type](#)

IBM Information Archive Server metadata attributes

This table describes metadata for IBM Information Archive Server attributes.

Table 1. Metadata for IBM Information Archive Server

Name	Dense	Multiple	Type	Normalized	Description
Access Date	Y	N	Date	Y	A time stamp that indicates the last accessed time of the data object.
Creation Date	Y	N	Date	Y	A time stamp that indicates the creation time of the data object.
File Extension	Y	N	String	N	The data object's extension.
File Name	Y	N	String	N	Name of the data object located within the context of the specified volume.
File Path	Y	N	String	N	All attributes that reference the data object's location.
File Size	Y	N	Int	N	The size of the data object in bytes.
Litigation Hold	N	Y	String	N	Name of the litigation hold that is applied to the data object. There can be multiple litigation holds applied to a data object.
Modified Date	Y	N	Date	Y	A time stamp that indicates the modification time of the data object.
Object ID	Y	N	String	N	Unique data object identifier provides the data object's "handle" to the system where

					the data object resides.
Retention Class	N	N	String	N	The name of the IBM Information Archive management class that is bound to the data object.
Retention Period	N	N	String	N	The length of time a data object is set to be retained.
Retention Policy Name	N	N	String	N	The name of the policy that last set retention for a data object.
Source Volume	Y	N	String	N	The originating server and share name from which an object was copied to the IBM Information Archive.

- **Parent topic:** Metadata by attribute type

Related concepts

- Metadata by attribute type

Hitachi HCAP Server metadata attributes

This table describes metadata for Hitachi HCAP Servers attributes.

Table 1. Metadata for Hitachi HCAP

Name	Dense	Multiple	Type	Normalized	Description
Access Date	Y	N	Date	Y	A time stamp that indicates the last accessed time of the data object.
Creation Date	Y	N	Date	Y	A time stamp that indicates the creation time of the data object.
File Extension	Y	N	String	N	The data object's extension.
File Name	Y	N	String	N	Name of the data object located within the context of the specified volume.
File Path	Y	N	String	N	All attributes that reference the data object's location.
File Size	Y	N	Int	N	The size of the data object in bytes.
Litigation Hold	N	Y	String	N	Name of the litigation hold that is applied to the data object. There can be multiple litigation holds applied to a data object.
Modified Date	Y	N	Date	Y	A time stamp that indicates the modification time of the data object.
Object ID	Y	N	String	N	Unique data object identifier provides the data object's handle to the system where

Retention Period	N	N	String	N	The length of time a data object is set to be retained.
Retention Policy Name	N	N	String	N	The name of the policy that last set retention for a data object.
Source Volume	Y	N	String	N	The originating server and share name from which an object was copied to the Hitachi HCAP.
Access Date	Y	N	Date	Y	A time stamp that indicates the last accessed time of the data object.

- **Parent topic:** <u>Metadata by attribute type</u>

Related concepts

- <u>Metadata by attribute type</u>

NetApp SnapLock Server metadata attributes

This table describes metadata for NetApp SnapLock Server attributes.

Table 1. Metadata for NetApp SnapLock Server

Name	Dense	Multiple	Type	Normalized	Description
Access Date	Y	N	Date	Y	A time stamp that indicates the last accessed time of the data object.
Creation Date	Y	N	Date	Y	A time stamp that indicates the creation time of the data object.
File Extension	Y	N	String	N	The data object's extension.
File Name	Y	N	String	N	Name of the data object located within the context of the specified volume.
File Path	Y	N	String	N	All attributes that reference the data object's location.
File Size	Y	N	Int	N	The size of the data object in bytes.
Litigation Hold	N	Y	String	N	Name of the litigation hold that is applied to the data object. There can be multiple litigation holds applied to a data object.
Modified Date	Y	N	Date	Y	A time stamp that indicates the modification time of the data object.
Object ID	Y	N	String	N	Unique data object identifier provides the data object's

"handle" to the system where |

					the data object is found.
Retention Period	N	N	String	N	The length of time a data object is set to be retained.
Retention Policy Name	N	N	String	N	The name of the policy that last set retention for a data object.
Source Volume	Y	N	String	N	The originating server and share name from which an object was copied to the NetApp SnapLock Server.
Access Date	Y	N	Date	Y	A time stamp that indicates the last accessed time of the data object.

- **Parent topic:** Metadata by attribute type

Related concepts

- Metadata by attribute type

Exchange Server metadata attributes

This table describes metadata for Exchange Server attributes.

Table 1. Metadata for Exchange Server

Name	Dense	Multiple	Type	Normalized	Description
Access Date	Y	N	Date	Y	A time stamp that indicates the last accessed time of the file.
Attachments	Y	N	String	N	Whether the data object has attachments.
BCC	Y	N	String	N	Recipients blind-copied on the email data object.
CC	Y	N	String	N	Recipients that are copied on the email data object.
Creation Date	Y	N	Date	Y	A time stamp that indicates the creation time of the file.
File Extension	Y	N	String	N	The extension of the data object. For Exchange, this extension is always .EML.
File Name	Y	N	String	N	Name of the data object (file) located within the context of the specified volume.
File Path	Y	N	String	N	All attributes that reference the data object's location.
File Size	Y	N	Int	N	The size of the file in bytes.
From	Y	N	String	N	The sender of the email data object.
Importance	Y	N	String	N	The importance level that is assigned by the sender of the email data object: High, Normal, Low.

Modified Date	Y	N	Date	Y	A time stamp that indicates the modification time of the file.
Sensitivity	Y	N	String	N	The sensitivity level that is assigned by the sender of the email data object: Normal, Personal, Private, Confidential.

- **Parent topic:** [Metadata by attribute type](#)

Related concepts

- [Metadata by attribute type](#)

Discovery Accelerator metadata attributes

This table describes metadata for Discovery Accelerator attributes.

Table 1. Metadata for Discovery Accelerator

Name	Dense	Multiple	Type	Description
Latest Review Status	N	N	String	All system-level Accelerator objects have one of the following statuses: Produced, Reviewed, Unreviewed, Pending, Questioned.
Review Mark Names	N	Y	String	Mark names set for each Review Status that the object was put into. Review marks are user-defined, and an object might have multiple marks.
Review Comments	N	Y	String	Comment (text) added by a Reviewer in Discovery Accelerator for this object. An object might have multiple review comments.
Reviewers	N	Y	String	Log in ID of a Reviewer who either added a Comment or a Status/Mark on the object. An object might have multiple Reviewers.
Message Type	N	N	String	All system-level Accelerator objects have one of the following messages: Unknown, Exchange, Instant Messaging, Bloomberg, Fax, Lotus, FSA, SMTP

- **Parent topic:** [Metadata by attribute type](#)

Related concepts

- [Metadata by attribute type](#)

FileNet metadata attributes

This table describes metadata for FileNet attributes.

Table 1. Metadata for FileNet

Name	Dense	Multiple	Type	Description
FoldersFiledIn	Y	Y	Object	The list of folders where this document is filed.
Id	N	Y	GUID	The unique object ID. References FileNet `Id`.
Name	N	N	String	This attribute returns the value of the designated name property for the object or its ID if there is no name property. References FileNet `Name`.
DateContentLastAccessed	N	N	Date/time	The date and time when this content of a document was last accessed. References FileNet `DateContentLastAccessed`.
DateCreated	N	N	Date/time	The date and time this object was created. References FileNet `DateCreated`.
DateLastModified	N	N	Date/time	The date and time when this object was last modified. References FileNet `DateLastModified`.
ContentSize	N	N	Float	The size (in bytes) of the captured content that is associated with this object. References FileNet `ContentSize`.
Owner	N	N	String	The security owner of the object. References FileNet `Owner`.
ContentRetentionDate	N	N	Date/time	The date until which the object must be retained, as determined by the content storage subsystem. References FileNet `ContentRetentionDate`.
Creator	N	N	String	The name of the user who created this object. References FileNet `Creator`.
DateCheckedIn	N	N	Date/time	The date and time this object was checked in. References FileNet `DateCheckedIn`. Different from `Document.name`.
DocumentTitle	N	N	String	References FileNet

				DocumentTitle.
IsCurrentVersion	N	N	Boolean	Indicates whether it is the current version in the version series. It is not a separate attribute, but rather used to set the `Current` string in attribute `VersionStatus` (`FileNet`). References FileNet `IsCurrentVersion`.
IsFrozenVersion	N	N	Boolean	Indicates whether user properties or content can be modified. It is not a separate attribute, but rather used to set the `Frozen` string in attribute `VersionStatus` (`FileNet`). References FileNet `IsFrozenVersion`.
IsInExceptionState	N	N	Boolean	Indicates whether the lifecycle of a document is in exception state. References FileNet `IsInExceptionState`.
IsReserved	N	N	Boolean	Indicates whether a user reserved the right to check in a version after this version. References FileNet `IsReserved`.
LastModifier	N	N	String	The name of the user who last modified this object. References FileNet `LastModifier`.
MajorVersionNumber	N	N	Integer	The major version number for this document version. References FileNet `MajorVersionNumber`.
MimeType	N	N	String	The MIME type of the document. References FileNet `MimeType`.
MinorVersionNumber	N	N	Integer	The minor version number for this document version. References FileNet `MinorVersionNumber`.
ReleasedVersion	N	N	Boolean	Indicates whether user properties or content can be modified.
VersionStatus	N	N	Integer	The version status of this document version. These status options are listed here. • In Process • Current version • Released

- Superseded
- Frozen

- **Parent topic:** [Metadata by attribute type](#)

Related concepts

- [Metadata by attribute type](#)

Email metadata attributes

This table describes metadata for email attributes.

Table 1. Metadata for email

Name	Dense	Multiple	Type	Available for Non Microsoft Email	Description
Any address	N	Y	String	Y	Any recipient or the sender of an email object
Attachments					Whether the data object has attachments. True if message has attachments
BCC					Recipients that are blind-copied on the email data object
CC					Recipients that are copied on the email data object
Client submit time	Y	N	Date	Y	
Conversation index	N	N	String	N	A value that indicates the relative position of this message within a conversation thread
Conversation topic	N	N	String	N	The topic of the first message in a conversation thread
Expire time	N	N	Date	N	The date and time at which the messaging system can invalidate the content of a message (YYYY-MM-DD)
From					The sender of the email data object

Importance					The importance level that is assigned by the sender of the email data object: High, Normal, Low
Internet message ID	N	N	String	Y	Corresponds to the message ID field as specified in [RFC2822]
Is-part-of					TRUE if object is part of email body, email attachment, miscellaneous email item's body or miscellaneous email item's attachment
Message create time	N	N	Date	Y	Date the message was created (YYYY-MM-DD)
Message delivery time	N	N	Date	Y	
Message modify time	N	N	String	Y	Date the message was last modified (YYYY-MM-DD)
Priority	N	N	String	Y	The order or speed at which the message must be sent by the messaging system software: Urgent, Normal, Low
Proof of Delivery Requested	N	N	Boolean	N	TRUE if a message sender requests proof of delivery for a particular recipient
Proof of					TRUE if a message sender requests proof that the message

Submission Requested	N	N	Boolean	N	transfer system was submitted a message for delivery to the originally intended recipient
Provider Submit Time	N	N	Date	Y	The date and time a transport provider passed the message to its underlying messaging system (YYYY-MM-DD)
Received by	N	N	String	N	Message recipient
Received date					Date the message was received (YYYY-MM-DD)
Received Representing	N	N	String	N	The display name for the messaging user that is represented by the recipient
Reply Message ID	N	N	String	Y	The original message's PR_INTERNET_MESSAGE_ID property value
Reply Message time	N	N	Date	N	Time when the reply was sent (YYYY-MM-DD)
Reply Requested	N	N	Boolean	Y	TRUE if a message sender requests a reply from a recipient
Sensitivity					The sensitivity level that is assigned by the sender of the email data object: Normal, Personal, Private, Confidential
Sent date					Date the message was

Sent date					sent (YYYY-MM-DD)
Sent Representing	N	N	String	N	Contains the display name for the messaging user that is represented by the sender
Subject					The subject of the email data object
To					The recipient/s of the email data object
Transport Message Headers	N	N	Date	Y	Contains transport-specific message envelope information

- **Parent topic:** [Metadata by attribute type](#)

Related concepts

- [Metadata by attribute type](#)

Data object property attributes

This table describes Data Object Property Attributes. Data object property attributes are stored inside the data object.

Table 1. Data object property attributes

Name	Dense	Multiple	Type
Comment	N	Y	String
Keyword	N	Y	String
Last saved by	N	Y	String
Primary author	N	Y	String
Subject	N	Y	String
Title	N	Y	String
Abstract	N	Y	String
Account	N	Y	String
Address	N	Y	String
Attachments	N	Y	String
Authorization	N	Y	String
Backup date	N	Y	String
Bill to	N	Y	String
Blind copy	N	Y	String
Carbon copy	N	Y	String
Category	N	Y	String
Checked by	N	Y	String
Client	N	Y	String
Completed date	N	Y	String
Characters	N	Y	String
Pages	N	Y	String
Words	N	Y	String
Created	N	Y	String
Department	N	Y	String
Destination	N	Y	String
Disposition	N	Y	String
Division	N	Y	String
Total editing time	N	Y	String
Editor	N	Y	String
Forward to	N	Y	String
Group	N	Y	String
Language	N	Y	String
Printed	N	Y	String
Mailstop	N	Y	String
Matter	N	Y	String
Office	N	Y	String
Operator	N	Y	String
Owner	N	Y	String
Project	N	Y	String
Publisher	N	Y	String
Purpose	N	Y	String
Received from	N	Y	String
Recorded by	N	Y	String

Recorded date	N	Y	String
Reference	N	Y	String
Revision date	N	Y	String
Revision notes	N	Y	String
Revision number	N	Y	String
Secondary author	N	Y	String
Section	N	Y	String
Security	N	Y	String
Source	N	Y	String
Status	N	Y	String
Type	N	Y	String
Typist	N	Y	String
Version date	N	Y	String
Version notes	N	Y	String
Version number	N	Y	String
Base file location	N	Y	String
Manager	N	Y	String
Company	N	Y	String
Last saved	N	Y	String
Presentation format	N	Y	String
Bytes	N	Y	String
Lines	N	Y	String
Paragraphs	N	Y	String
Slides	N	Y	String
Hidden slides	N	Y	String
Notes	N	Y	String
Multimedia clips	N	Y	String
Characters (with spaces)	N	Y	String
Scale crop	N	Y	String
Links dirty	N	Y	String
Title of parts	N	Y	String
Heading pairs	N	Y	String
Document number	N	Y	String
Application version	N	Y	String
User defined	N	Y	String

- **Parent topic:** [Metadata by attribute type](#)

Related concepts

- [Metadata by attribute type](#)

SharePoint attributes

This table describes SharePoint metadata attributes.

Table 1. Metadata for SharePoint

Name	Dense	Multiple	Type	Normalized	Description
Checked Out To	Y	N	String	N	Name of the user who currently has the data object checked out. This attribute is empty if the file is not checked out.
Created Date	Y	N	Date	Y	A time stamp that indicates the creation time of the file.
Extension	Y	N	String	N	The extension of the data object on the SharePoint server.
File Name	Y	N	String	N	Name of the data object located within the context of the specified volume.
System Path	Y	N	String	N	All attributes that reference the data object's location.
File Size	Y	N	Int	N	The size of the data object in bytes.
Modified Date	Y	N	Date	Y	A time stamp that indicates the modification time of the data object.
Status	Y	N	String	N	The state of the data object: published, checked out, or checked in.
					The title of the data object. By default, this attribute is the title of

Title	Y	N	String	N		the Office document as shown in Document Properties. Configurable by the user during check-in.
URL	Y	N	String	N		The URL that points to the data object on SharePoint.
Version	Y	N	String	N		The version of the data object, from a SharePoint server perspective

- **Parent topic:** [Metadata by attribute type](#)

Related concepts

- [Metadata by attribute type](#)

Calendar metadata attributes

This table describes metadata attributes for calendar items.

Table 1. Metadata for calendar

Attribute Name	Dense	Multiple	Type	Normalized	Description
Event duration	No	No	Integer	No	Time in minutes
Event end date	No	No	String	Yes (ISO 8601)	Date a calendar entry end (YYYY-MM-DD)
Event location	No	No	String	Yes (lowercase)	Location field in a calendar entry
Event start date	No	No	String	Yes (ISO 8601)	Date that a calendar entry begins (YYYY-MM-DD)
is Meeting	No	No	Boolean	No	True if the calendar entry is a meeting (Any entry that has attendees)

- **Parent topic:** <u>Metadata by attribute type</u>

Related concepts

- <u>Metadata by attribute type</u>

Contact metadata attributes

This table describes metadata attributes for contacts.

Table 1. Metadata for contact

Attribute Name	Dense	Multiple	Type	Normalized	Description
Business address	No	No	String	Yes (lowercase)	Business address that is listed for a contact entry
Business phone	No	No	String	Yes (lowercase)	Business phone that is listed for a contact entry
Company	No	No	String	Yes (lowercase)	Company that is listed for a contact entry
Contact job title	No	No	String	Yes (lowercase)	Job title that is listed for a contact entry
Contact keywords	No	No	String	Yes (lowercase)	Keywords that are listed for a contact entry
Department	No	No	String	Yes (lowercase)	Department that is listed for a contact entry
Email display name	No	No	String	Yes (lowercase)	Email display name that is listed for a contact entry
Full name	No	No	String	Yes (lowercase)	Full name that is listed for a contact entry
Home address	No	No	String	Yes (lowercase)	Home address that is listed for a contact entry
Home phone	No	No	String	Yes (lowercase)	Home phone that is listed for a contact entry
IM address	No	No	String	Yes (lowercase)	Instant Messenger address that is listed for a contact entry
Mobile phone	No	No	String	Yes (lowercase)	Mobile phone number that is listed for a contact entry
Office location	No	No	String	Yes (lowercase)	Office location that is listed for a contact entry
					Pager number

Pager	No	No	String	Yes (lowercase)	that is listed for a contact entry
Web page	No	No		Yes (lowercase)	Web page that is listed for a contact entry

- **Parent topic:** [Metadata by attribute type](#)

Related concepts

- [Metadata by attribute type](#)

Digital signatures metadata attributes

This table describes attributes that are used in IBM® StoredIQ data analysis.

Table 1. Metadata for digital signatures

Attribute Name	Dense	Multiple	Type	Normalized	Description
Encryption type	No	No	Enum	Yes (lowercase)	Lotus Notes' or 'General'
Hash	Yes	No	String	Yes (lowercase)	Hash value for the object
NSRL digital signature	Yes	No	Boolean	No	True if the object's hash value matches an item on the NSRL hash list

- **Parent topic:** [Metadata by attribute type](#)

Related concepts

- [Metadata by attribute type](#)

Processing metadata attributes

This table describes attributes that are used in data processing.

Table 1. Metadata for processing data

Attribute Name	Dense	Multiple	Type	Normalized	Description
Binary processing state	Yes	No	Integer	Yes	State of binary processing: Not Performed, Successful, Partial, Failed
Binary error code	No	Yes	String	Yes (lowercase)	When binary processing is performed, the following codes apply: Indexing of binary text failed, Binary text partially indexed, Binary content tagging failed, Max entities reached in binary text, Binary text extraction failed, Binary full processing completed
Content processing state	Yes	No	Integer	Yes	State of content processing: Not Performed, Successful, Partial, Failed
Content error	No	Yes	String	Yes	When the content processing state is "Partial" or "Failed," the following codes apply: Configured skipped data object, Unsupported file format, Configured max. object size, Indexing failed, Partially indexed,

code				(lowercase)	Unknown file format, Content tagging failed, Max entities reached, Encrypted or password protected, Text extraction failed, No text in this file format, Could not read source object, Max entity value length reached
Object	Yes	No	Integer	Yes	
System processing state	Yes	No	Integer	Yes	State of system processing: Successful, Partial
System error code	No	Yes		Yes (lowercase)	When system processing state is "Partial," the following code applies: User/Group lookup failed

- **Parent topic:** [Metadata by attribute type](#)

Related concepts

- [Metadata by attribute type](#)

Task metadata attributes

Table 1. Metadata for task items

Attribute Name	Dense	Multiple	Type	Normalized	Description
Task complete	No	No	Boolean	No	True if the task is marked as complete
Task due date	No	No	String	Yes (ISO 8601)	Date the task is marked as being due (YYYY-MM-DD)
Task owner	No	No	String	Yes	Creator of the task
Task percent complete	No	No	Integer	Yes	Percent complete of a task
Task start date	No	No	String	Yes (ISO 8601)	Date the task is marked as being started (YYYY-MM-DD)
Task status	No	No	String	Yes	Status of a task

- **Parent topic:** Metadata by attribute type

Related concepts

- Metadata by attribute type

Attribute value types

This table describes the value and instance-based expressions that can be used with a query condition, depending on the type of attribute.

Table 1. Attribute value types

Value Type	Dense		Sparse	
	Single Instance	**Multi-instance**	**Single Instance**	**Multi-instance**
			Appears	Appears
			Appears uniquely	Appears uniquely
			Does not appear	Does not appear
String	Is	Is	Is	Is
	Is not	Is not	Is not	Is not
	Contains	Contains	Contains	Contains
	Does not contain	Does not contain	Does not contain	Does not contain
	Begins with	Begins with	Begins with	Begins with
	Ends with	Ends with	Ends with	Ends with
	Matches regular expression	Matches regular expression	Matches regular expression	Matches regular expression
		at least n times (n>0)		at least n times
		at most n times (n>0)		at most n times
		exactly n times (n=0)		exactly n times
			Appears	Appears
			Appears uniquely	Appears uniquely
			Does not appear	Does not appear
Date	Is	Is	Is	Is
	Between x and y days ago	Between x and y days ago	Between x and y days ago	Between x and y days ago
	At least x days ago	At least x days ago	At least x days ago	At least x days ago
	At most x days ago	At most x days ago	At most x days ago	At most x days ago
		at least n times (n>0)		at least n times
		at most n times (n>0)		at most n times
		exactly n times (n=0)		exactly n times
Int			Appears	Appears
			Appears	Appears
			Appears uniquely	Appears uniquely
			Does not appear	Does not appear
	Equals	Equals	Equals	Equals
	At least	At least	At least	At least
	At most	At most	At most	At most
	Between	Between	Between	Between
		at least n times (n>0)		at least n times

	at most n times (n>0)		at most n times
	exactly n times (n=0)		exactly n times
Not performed		Not performed	
Successful		Successful	
Partial		Partial	
Failed		Failed	

- **Parent topic:** [Metadata by attribute type](#)

Related concepts

- [Metadata by attribute type](#)

IBM Lotus Notes metadata attributes

This table describes metadata for IBM Lotus Notes attributes.

Table 1. Metadata for IBM Lotus Notes

Attribute Name	Dense	Multiple	Type	Normalized	Description
Attachments					Whether the data object has attachments. True if message has attachments
Bcc, BlindCopyTo, EnterBlindCopyTo					Recipients that are blindcopied on the email data object
CC, CoptyTo, EnterCopyTo					Recipients that are copied on the email data object
Thread_Topic	N	N	String	N	The topic of the first message in a conversation thread
From					The sender of the email data object
Importance					The importance level that is assigned by the sender of the email data object: High, Normal, Low
Message-ID	N	N	String	Y	Corresponds to the message ID field as specified in [RFC2822]
Is-part-of					TRUE if object is part of email body, email attachment, miscellaneous email item's body or miscellaneous email item's attachment
					Date the

Name			Type		Description
Create, ComposedDate	N	N	Date	Y	message was created (YYYY-MM-DD)
DeliveredDate	N	N	Date	Y	
Modify	N	N	String	Y	Date the message was last modified (YYYY-MM-DD)
DeliveryPriority	N	N	String	Y	The order or speed at which the message must be sent by the messaging system software: Urgent, Normal, Low
ReturnReceipt	N	N	Boolean	N	TRUE if a message sender requests proof of delivery for a particular recipient
Sensitivity					The sensitivity level that is assigned by the sender of the email data object: Normal, Personal, Private, Confidential
Sent, Posted					Date the message was sent (YYYY-MM-DD)
SentRepresenting-Name	N	N	String	N	Contains the display name for the messaging user that is represented by the sender
Subject					The subject of the email data object
To, SendTo					The recipient/s of the email data object
Attachments					Whether the data object has attachments. True if

- **Parent topic:** [Metadata by attribute type](#)

Related concepts

- [Metadata by attribute type](#)

Enterprise vault metadata attributes

This table describes metadata for enterprise vault attributes.

Table 1. Metadata for enterprise vault

Name	Dense	Multiple	Type	Description
Archive date (Enterprise Vault)	N	N	Date	Date that the object was archived (YYYY-MM-DD).
Author (Enterprise Vault)	N	N	String	The creator of the object, as shown in the archive properties
Content fingerprint (Enterprise Vault)	N	N	String	Content fingerprint of the object
Custom ID (Enterprise Vault)	N	N	String	Application-specific ID for the object
Custom properties (Enterprise Vault)	N	Y	String	Application-specific information for the object
Dedupe fingerprint (Enterprise Vault)	N	N	String	Deduplication fingerprint of the object
Domino UNID (Enterprise Vault)	N	N	String	Domino Universal NoteID
FSA original name (Enterprise Vault)	N	N	String	File system archive original name
Keywords (Enterprise Vault)	N	Y	String	Keywords that are associated with the data object
Message type (Enterprise Vault)	N	N	String	Data type of the object (DOC, XLS, MSG)
Missing content reason (Enterprise Vault)	N	N	Numeric String	Reason that object content is not available in the archive
Original location (Enterprise Vault)	N	N	String	Original location of the archived object
Retention category (Enterprise Vault)	N	N	String	Name of retention category that is applied to the object
Retention expires	N	N	Date String	Date that the retention on the

| (Enterprise Vault) | object expires (YYYY-MM-DD) |

- **Parent topic:** [Metadata by attribute type](#)

Related concepts

- [Metadata by attribute type](#)

Jive metadata attributes

This table describes attributes that are used in Jive items.

Table 1. Metadata for Jive

Attribute Name	Dense	Multiple	Type	Normalized	Description
Author ID (Jive)	No	No	String	No	The unique ID of the author within Jive.
Author Name (Jive)	No	No	String	No	The author's full name within Jive.
Author Username (Jive)	No	No	String	No	The author's user name within Jive.
Content ID (Jive)	No	No	String	No	The unique ID of content within Jive.
Content URL (Jive)	No	No	String	No	The content's URL within Jive.
Discussion Question (Jive)	No	No	Boolean	No	Indicates whether the Jive discussion was marked as a question.
Like Count (Jive)	No	No	Date	No	Indicates that the number of times that content was marked as "liked" on Jive.
Place ID (Jive)	No	No	String	No	The unique ID of the parent place on Jive.
Place Name (Jive)	No	No	String	No	The name of the parent place on Jive.
Place URL (Jive)	No	No	String	No	The URL of the parent place on Jive.
Published Date (Jive)	No	No	Date	No	The post's published date on Jive.
Reply Count (Jive)	No	No	Date	No	Indicates that the number of replies posted to content on Jive.
Status (Jive)	No	No	String	No	Indicates a discussion's status on Jive.
Subject (Jive)	No	No	String	No	Indicates content's subject on Jive.

Version (Jive)	No	No	String	No	Indicates a document's version on Jive.
View Count (Jive)	No	No	Date	No	Indicates that the number of times content was viewed on Jive.

- **Parent topic:** [Metadata by attribute type](#)

Related concepts

- [Metadata by attribute type](#)

Chatter metadata attributes

This table describes attributes that are used in Chatter items

Table 1. Metadata for Chatter

Attribute Name	Dense	Multiple	Type	Normalized	Description
Object ID (Chatter)	No	No	String	No	The unique ID of the object within Chatter.
Object type (Chatter)	No	No	String	No	The type of object within Chatter.
Parent group ID (Chatter)	No	No	String	No	The unique ID of the parent group or profile within Chatter.
Parent group name (Chatter)	No	No	String	No	The name of the parent group or profile within Chatter.
Parent post ID (Chatter)	No	No	String	No	The unique ID of the parent post within Chatter.
Parent post type (Chatter)	No	No	Boolean	No	The object type of the parent post within Chatter.
Created by user ID (Chatter)	No	No	Date	No	The unique ID of the creator within Chatter.
Created by username (Chatter)	No	No	String	No	The user name of the creator within Chatter.
Created date (Chatter)	No	No	Datetime	No	The posted date within Chatter.
Comment count (Chatter)	No	No	Integer	No	The number of comments for a post within Chatter.
Like count (Chatter)	No	No	Integer	No	The number of "likes" for a post or comment within Chatter.
Title (Chatter)	No	No	String	No	The title for links and attachments within Chatter.
Link URL (Chatter)	No	No	String	No	The content of links posted within Chatter.

- **Parent topic:** <u>Metadata by attribute type</u>

Related concepts

- <u>Metadata by attribute type</u>

Managing IBM StoredIQ eDiscovery

IBM® StoredIQ eDiscovery captures and manages the electronic data in your system. IBM StoredIQ eDiscovery uses "matter" to create and organize information so that you can always easily access the data that you need and review the status or process of your work.

The following topics describe in detail how to use IBM StoredIQ eDiscovery to achieve your business goals.

- [Overview of IBM StoredIQ eDiscovery](#)
  IBM StoredIQ eDiscovery is designed for the legal users to manage their electronic data. With this application, you can capture and control the data to communicate with other users about the work status and case review process.
- [Actions on a matter](#)
  Provides instructions about how to manage matters on the Matter Dashboard.
- [Matter configuration](#)
  Matters are the basis of any e-discovery work. Through matter configuration, you can tune them to meet your legal needs.
- [Actions on a box](#)
  When an Identification Box is created, multiple actions are available to help you categorize and move data objects as part of the e-discovery process.

- **Parent topic:** [Managing the IBM StoredIQ products](#)

Overview of IBM StoredIQ eDiscovery

IBM® StoredIQ eDiscovery is designed for the legal users to manage their electronic data. With this application, you can capture and control the data to communicate with other users about the work status and case review process.

IBM StoredIQ eDiscovery identifies and catalogs data or types of data that can eventually be responsive to legal matters. The data that is collected and retained by IBM StoredIQ eDiscovery is then produced in a way that matches accepted or required legal practices.

IBM StoredIQ eDiscovery can be used to accomplish any of these tasks:

- Create a matter for a law suit.
- Create boxes to contain emails and reports about cases.
- Visualize the status of the different active matters.
- Communicate with emails directly from the application to discuss any questions about a matter.

Before you use IBM StoredIQ eDiscovery, ensure that you meet these prerequisites:

- The IBM StoredIQ Platform must be deployed, configured, and ready for use.
- IBM StoredIQ eDiscovery is dependent upon other IBM StoredIQ applications such as IBM StoredIQ Administrator and IBM StoredIQ Data Workbench.

- Switching between applications
  With the IBM StoredIQ product, you can switch between different applications.

Switching between applications

With the IBM StoredIQ product, you can switch between different applications.

About this task

These applications include:

- IBM StoredIQ Data Workbench
- IBM StoredIQ eDiscovery
- IBM StoredIQ Data Script
- IBM StoredIQ Policy Manager
- IBM StoredIQ Administrator

To switch to the application that you want to work on, follow these steps.

Procedure

1. Within the global navigation bar, click the white arrow next to the name of the application. A menu appears, listing available IBM StoredIQ Platform applications.
2. In the menu, click the name of the application to which you want to switch. Note: There are four roles for users: administrator, data user, discovery user, and policy user. Different IBM StoredIQ Platform applications are available to the user, depending on what roles were selected. Administrators have access to IBM StoredIQ Administrator in addition to IBM StoredIQ Data Workbench, IBM StoredIQ eDiscovery, and IBM StoredIQ Data Script. Discovery users have access to IBM StoredIQ eDiscovery. Data users have access to IBM StoredIQ Data Workbench. Policy users have access to IBM StoredIQ Policy Manager.

- **Parent topic:** Overview of IBM StoredIQ eDiscovery

Actions on a matter

Provides instructions about how to manage matters on the Matter Dashboard.

- [Matter Dashboard view](#)
  The Matter Dashboard provides a high-level overview of matters within IBM® StoredIQ
  eDiscovery. It helps you to find and identify matters of interest.

Matter Dashboard view

The Matter Dashboard provides a high-level overview of matters within IBM® StoredIQ eDiscovery. It helps you to find and identify matters of interest.

A matter is a central container object within IBM StoredIQ eDiscovery. You can organize all information that pertains to a potential or active legal matter. Matters can be seen on the matter dashboard. It shows a matter's status while it provides critical pieces of information for each matter: its matter and box navigation trail, its name, a brief description, its creation date and time, and a visual cue about its status. It also provides insight into the contents of the matter, including what boxes and reports it contains and relevant contacts for the matter. Matters are listed according to the time sequence of their creation. From the Matter Dashboard, you can do the following tasks:

- View a matter by clicking the matter that you want to see.
- Search for a matter by filtering the existing matters.
- Add a matter by clicking Create Matter.

- [View of a matter](#)
  Through accessing the matter, you can view the contents of an existing matter such as its boxes and reports.
- [Emailing the contact person of a matter](#)
  You can email the contact person of a matter by clicking the name of that person on the Matter Dashboard.
- [Filtering for a matter](#)
  You can filter for a specific matter within existing matters from the Matter Dashboard.
- [Adding a matter](#)
  A matter is a container that encapsulates all data that you are gathering for a specific legal case. It can also contain some constraining parameters such as matter dates and identifying information like the matter's name and description. Matters are an essential component of using IBM StoredIQ eDiscovery and can be added from the Matter Dashboard.
- [Deleting a matter](#)
  Matters can be deleted from the Matter Dashboard.

- **Parent topic:** [Actions on a matter](#)

Related concepts

- [Actions on a matter](#)

View of a matter

Through accessing the matter, you can view the contents of an existing matter such as its boxes and reports.

On the Matter Dashboard, double-click the matter that you want to view. The matter opens on the Matter Details page.

- **Parent topic:** Matter Dashboard view

Related concepts

- Matter Dashboard view

Emailing the contact person of a matter

You can email the contact person of a matter by clicking the name of that person on the Matter Dashboard.

Procedure

1. From the Matter Dashboard, find the matter that contains the contact person you want to email.
2. In the Contacts area of that matter, click the name of the individual you want to email. An email to that individual automatically opens in your email client with the matter's name as the subject of the email.
   Note: The Data Expert for a box can also be emailed by clicking Contact Data Expert on the matter's Boxes and Reports page.

- **Parent topic:** [Matter Dashboard view](#)

Related concepts

- [Matter Dashboard view](#)

Filtering for a matter

You can filter for a specific matter within existing matters from the Matter Dashboard.

Procedure

1. From the Matter Dashboard, in the Enter terms here.... text box, enter the name of the matter for which you want to filter, and then press Search. You can enter either the full matter name or a portion of it. Only matter names that match the filter entry are displayed in the list. Matters that match the entered search term are returned.
2. View the matter by double-clicking it within the Matter Dashboard.
3. To remove the search term, click the X to the left of the Enter terms here.... text box.

- **Parent topic:** Matter Dashboard view

Related concepts

- Matter Dashboard view

Adding a matter

A matter is a container that encapsulates all data that you are gathering for a specific legal case. It can also contain some constraining parameters such as matter dates and identifying information like the matter's name and description. Matters are an essential component of using IBM® StoredIQ eDiscovery and can be added from the Matter Dashboard.

Procedure

1. Within the Matter Dashboard, click Create Matter. The Create Matter dialog box appears. The Create Matter dialog box has four fields:Name, Description, Data Expert Contact, and Legal Ops Contact.
2. In the Name text box, enter a unique name for this matter. This name appears on the Matter Dashboard page, providing you an easy way to identify the matter.
3. In the Description text box, enter a unique description of this matter.
4. In the Data Expert Contact text box, select the user that serves as the Data Expert for this matter. The users in this list are populated with the IBM StoredIQ Administrator user management feature.
5. In the Legal Ops Contact text box, select the user that serves as the Legal Operations contact for this matter. The users in this list are populated with the IBM StoredIQ Administrator user management feature.
6. Click Save. The newly created matter appears within the Matter Dashboard. When you have a matter, then you can create boxes and begin to manage your e-discovery process.

- **Parent topic:** Matter Dashboard view

Related concepts

- Matter Dashboard view

Deleting a matter

Matters can be deleted from the Matter Dashboard.

Procedure

1. Within the Matter Dashboard, place your mouse over the matter you want to delete.
2. Click Delete, which appears within the Contact tile. The Delete Matter confirmation box appears.
   Note: You cannot delete a matter that has a box in a pending state.
3. Within the Delete Matter confirmation box, click Delete, and the matter is removed from the Matter Dashboard

- **Parent topic:** Matter Dashboard view

- **Parent topic:** Matter configuration

Related concepts

- Matter Dashboard view
- Matter configuration

Matter configuration

Matters are the basis of any e-discovery work. Through matter configuration, you can tune them to meet your legal needs.

- Editing a matter
  Use the following procedure to edit an existing matter.
- Deleting a matter
  Matters can be deleted from the Matter Dashboard.
- Creating a box
  A box is a container for requested data within a matter. It serves as the object of reference as you move from managing information through to data identification, collection, and export.
- Editing the summary of a box
  When you are editing a box, you can modify the base information or metadata that is entered during box creation.

Editing a matter

Use the following procedure to edit an existing matter.

Procedure

1. Within the Matter Dashboard, double-click the matter that you want to edit. The matter opens, showing both the Matter Dashboard for the selected matter and the matter's boxes and reports.
2. In the upper right-hand corner, click Edit Matter. The Edit Matter dialog box appears.
3. Edit the matter as needed, changing the name or description of the matter.
4. Click Save.

- **Parent topic:** Matter configuration

Related concepts

- Matter configuration

Creating a box

A box is a container for requested data within a matter. It serves as the object of reference as you move from managing information through to data identification, collection, and export.

Boxes are created within matter from the Boxes and Reports page, which has four areas: Boxes, Summary, Reports, and Available Actions.

Box

This table describes the fields that appear on the left side of the Matter Overview screen, labeled Boxes.

Table 1. Box table fields and descriptions

Field Name	Description
Name	The name of the box within the matter.
Type	A list of the type of the box. Options include Identification, Export, and Collection.
Created	A list of the date and time at which the box was created.
Status	The system status of the box. Options include Active or Pending. If the box is in a Pending state, certain links such as File List and Data Map can be disabled.
Total Box Objects	This field lists the total number of data objects in the selected box. If a box is selected in the table, then the counts in this field update to reflect the selected box.
Estimated Total Box Size	This field lists the estimated total size of data objects in the selected box. If a box is selected in the table, then the counts in this field update to reflect the selected box.

Selecting a box

1. In the All Boxes list, select the type of box you want to view. Options include All Boxes, Identification Boxes, Collection Boxes, and Export Boxes.
   - Identification Boxes: Identifying data is a key step in the e-discovery process that is based on several methods, including custodians and matter terms. An Identification Box is the initial type of box that is created when you are configuring a matter, and it is key to defining the data within a matter.
   - Collection Boxes: A Collection Box collects data for preservation when it is approved.
   - Export Boxes: An Export Box retains specific data when it is approved for export.
2. Depending on the selection that you made, boxes of that type appear in the Boxes table.
   Note: By default, the only type of box you can create is an Identification Box. When you create an Identification Box, you can create a Collection Box or Export Box that is based on the original Identification Box.

Summary

This table describes the Summary area on the right side of the Matter Overview screen.

Table 2. Summary options and descriptions

Field Name	Description
Name	The name of the selected box.
Description	The description of the selected box.
Details	Provides links to detailed data for the box, including File List and Data Map. When a box is in a pending state, these links are disabled.
Started	The created date and time of the box.

Status   The system status of the selected box: Active or Pending.

Reports

This table describes the Reports area on the right side of the Matter Overview screen.

Table 3. Reports options and descriptions

Field Name	Description
Name	The name of the report.
Created	The creation date and time of the report.
Overall Status	The creation status of the report. Options include Complete or Pending.

Available Actions

This area provides links to what actions can be taken on a selected box. From here, you can click a link to act on a box. These link options are available:

- Create Identification Box
- Refine this box
- Restart this box
- Create a Collection Box that is based on this box
- Create an Export Box that is based on this box
- Clone this box
- Run a report on this box
- Deleting this box

Creating an identification box

1. Select the matter in which you want to create a box.
2. Click Create Identification Box. The Create Box dialog box appears. There are five subtabs in the Create Identification Box dialog box: Base Information, Where and When, Who, Terms, and Notes.
3. Within the Base Information subtab, complete these fields.
    a. In the Name text box, enter the name for the box and its outputs.
    b. Optionally, in the Description text box, enter a brief description of the box.
    c. Click Next to access the Where and When subtab.
4. Within the Where and When subtab, complete these fields.
    a. In the File Locale check boxes, select at least one locale for this box. Options include All User Data, File Shares, Emails, Desktops, and Collaboration Files. Note: If you select the All User Data check box, all file-locale options are automatically selected.
    b. In the Time Interval area, enter the start and end dates for this box.
    c. Click Next to access the Who subtab.
5. Within the Who subtab, complete these fields.
    a. Select identities to act as custodians for the box. Use Add and Remove to move custodians from the All Identities list to the Box Custodians list. Note: Identities are created by an administrator in the concept portion of IBM® StoredIQ Administrator. Note: Search for a specific identity by entering all or a portion of the user's name in the Enter terms... text box and clicking Search.
    b. Select Any Identities or All Identities. This field is a Boolean AND/OR value. Note: By default, Any Identities is selected. By changing the Any Identities/All Identities setting, you might obtain unexpected results.
    c. Click Next to access the Terms subtab.
6. Within the Terms subtab, complete these fields. Note: Entered terms are used as part of a full-text search within an infoset. The data objects within the infoset must be part of the full-text index to return results.
    a. Optionally, enter litigation terms for this box. In the Enter term and click Add

text box, enter the wanted litigation term, and then click Add.
Note: Use Delete and Delete All to remove added litigation terms.

    b. Select Any Term or All Terms. This value is a Boolean `AND/OR` value.
Note: By default, Any Term is selected. By changing the Any Term/All Terms setting, you might obtain unexpected results.

    c. Click Next to access the Notes subtab.

7. Within the Notes subtab, optionally enter notes or extra instructions for the box's data expert. In the Notes to Data Expert text box, enter the information.
Note: This text box has a limit of 2,000 characters.

8. Click Finish. The Boxes and Reports page of the matter populates with the parameters set during box creation with the box in a Pending state.

9. When an Identification Box is created, it is placed in a Pending state. An email is sent to the Data Expert, indicating that the Data Expert must generate a response that matches the parameters that are indicated in the box request. To complete the infoset request, the Data Expert must click the URL in the system-generated email.

10. When the infoset is addressed, the box moves from a Pending to Available state, at which point more actions become available to that box. The actions available in the Available Actions area that is changed from a single option, Create Identification Box, to multiple options.

- **Parent topic:** [Matter configuration](#)

Related concepts

- [Matter configuration](#)

Editing the summary of a box

When you are editing a box, you can modify the base information or metadata that is entered during box creation.

Procedure

1. Within the matter, select the box that you want to edit in the Boxes table.
2. In the Summary area, click Edit.
3. Modify the summary options of the box as needed.

- Data maps
  The data map provides a visual layout of your data and in-depth information about data source types, data categories, size or amounts, the number of data objects, and details.
- Viewing the file list or data map
  The file list and data map helps you to explore your data, its details, and relationships.

- **Parent topic:** Matter configuration

Related concepts

- Matter configuration

Data maps

The data map provides a visual layout of your data and in-depth information about data source types, data categories, size or amounts, the number of data objects, and details.

The data map contains these areas:

Data Map Option	Description
Group By	The Group By area contains these options: • Data Source Type: This view displays the different data source types that make up the selected infoset. By selecting or clearing the data source types, the data map automatically reconfigures, providing you with a visual overview of your data. • None: This option removes data source types from the data map.
View By	The View By area contains these options: • Category View: View data by category, including container files and archives, images, spreadsheets, and email. • Size View: View data by size, from less than one MB to over 100 GB. The Size View is only active for infosets that have a top-level composition. • Created View: View data by created date from within the past 30 days to more than five years ago. • Last Modified View: View data by its last modified date, from within the past 30 days to more than five years ago. • Last Accessed View: View data by its last accessed date, from within the past 30 days to more than five years ago. By selecting or clearing the viewing options, the data map automatically reconfigures, providing you with a visual overview of your data.
Data Map	Within the data map, you have these viewing options. • Object Count: The data map displays the number of objects and their percentage of that data source when you place the pointer over a category. • Object Size: The data map displays the size and percentage of the objects for that data source when you place the pointer over a category. Object Size is only active for infosets that have a top-level composition. You also have these viewing options: • Squarify: This view depicts data as unordered squares rather than rectangles. • Slice & Dice: This view depicts data as ordered rectangles. • Strip: This view strips data, depicting it in the easiest viewing aspect possible. This area displays details about a tile that is selected in the data map. By clicking an individual map tile, details about that specific tile are displayed. You can also see the total number of data objects referred to by the map tile. Within the Data Map Details, you have these viewing options. • Volumes: If the Volumes option is selected, the volume's name and view selected displays in the Data Map Details for the selected tile. The bar graph populates automatically based on the count or size of the selected tile.

Data Map Details	• Owners: If the Owners option is selected, the owners of content displays within that tile are displayed in the Data Map Details. The bar graph populates automatically based on the count or size of the selected tile.
	Note: Selected tiles have a green border. Tiles over which the pointer was placed have a black border.
	The Data Map Details also shows you the Total Size of the selected tile. The bar graph automatically populates, providing a visual representation of how the size of the selected tile relates to that of the size of the infoset as a whole.
	By clicking View Full List, the Data Map Details dialog appears. From here, you can download information about the selected tile to a CSV file.
	Under the data map, this field displays the following options.
Viewing x of x total object	• If Object Count is selected, this figure indicates the number of data objects that are being viewed compared to the total number of data objects. • If Object Size is selected, this figure indicates the data objects' size in bytes compared to the total byte size of the data objects.

Note: When you select one of the Data Source Type, Object Count, or Slice & Dice options, that option is colored blue within the user interface.

- **Parent topic:** [Editing the summary of a box](#)

Related tasks

- [Editing the summary of a box](#)

Viewing the file list or data map

The file list and data map helps you to explore your data, its details, and relationships.

About this task

Note: Switch between the file list and the data map by clicking the List or Map tabs. A selected, active tab is blue.

To view a box's file list or data map:

Procedure

1. Within the matter's box, click File List in the Summary area.
   Note: The File List link is disabled when the box is in a pending state.

   The Settings area depicts selections that are entered during box creation. The four subtabs, that is, Where and When, Who, Terms, and Notes, all appear as headings within the Settings area.

   The File List area shows information about the box's data objects, including the Name, File Size, File Path, Created date and time, and the Last Modified date and time. To view a particular object, click the object's name and it opens in the Data Object Viewer page.

   a. In a box's File List, click the name of the data object you would like to view. The data object appears in the Data Object Viewer.
   b. Move from one object to the next or within the object itself using the navigation options at the top of the viewer.
   c. Close the Data Object Viewer by clicking Close.
2. Alternately, within the matter's box, click Data Map in the Summary area.
   Note: The Data Map link is disabled when the box is in a pending state.
   The data map is identical to the data map in IBM® StoredIQ Data Workbench. Within the data map, you can explore your data to better understand it and its details.

- **Parent topic:** [Editing the summary of a box](#)

Related tasks

- [Editing the summary of a box](#)

Actions on a box

When an Identification Box is created, multiple actions are available to help you categorize and move data objects as part of the e-discovery process.

These box actions are available in the Available Actions area:

- Refine this box
- Restart this box
- Create a new box based on this box
- Create a Collection Box based on this box
- Create an Export Box based on this box
- Clone this box
- Run a report on this box
- Delete this box

- [Refining a box](#)
  After a box is created, it can be refined by changing the scope of the box. This revision modifies the box's parameters.
- [Restarting a box](#)
  If a box undergoes several revisions, it is necessary to restart the box.
- [Cloning a box](#)
  Created boxes can be cloned to create other, identical boxes.
- [Creating a collection box](#)
  Collection boxes can be created with another box's original settings.
- [Creating an export box](#)
  Export boxes can be created with another box's original settings.
- [Reports](#)
  IBM® StoredIQ eDiscovery provides various types of reports.
- [Running a report on a box](#)
  Follow this procedure to run various types of reports.
- [Deleting a box](#)
  Boxes that are no longer needed within a matter can be deleted.

Refining a box

After a box is created, it can be refined by changing the scope of the box. This revision modifies the box's parameters.

Procedure

1. Select the box that you want to refine, and then click the Refine this box link in the Available Actions area. The Refine Box dialog box appears. There are four subtabs in the Refine Box dialog box: Where and When, Who, Terms, and Notes.
2. Within the Where and When subtab, complete these fields.
   a. In the File locale check boxes, select at least one locale for this box. Options include File Shares, Emails, Desktops, and Collaboration.
   b. In the Time interval area, use the list and enter the start and end dates for which this box must be revised.
   c. Click Next to access the Who subtab.
3. Within the Who subtab, complete the following fields. By default, the fields are populated with current custodian selections.
   a. Select identities to act as custodians for the box. Use Add and Remove to move custodians from the All Identities list to the Current Box Custodians list.
      Note: Search for a specific identity by entering all or a portion of the user's name in the Enter terms... text box and clicking Search.
   b. Select All Identities or All Identities. The value is a Boolean AND/OR value.
      Note: If you change the All Identities/All Identities setting, you might obtain unexpected results.
   c. Click Next to access the Terms subtab.
4. Within the Terms subtab, complete the following fields. By default, the fields are populated with current matter-term selections.
   a. In the Enter term and click Add text box, enter the wanted litigation term, and then click Add.
   b. Select either Any Term or All Term. This value is a Boolean AND/OR value.
      Note: If you change the Any Term/All Terms setting, you might obtain unexpected results.
   c. Use Delete and Delete All to remove added litigation terms.
   d. Click Next to access the Notes subtab.
5. Within the Notes subtab, complete these fields.
   a. In the Notes to Data Expert text box, enter instructional information for the box's data expert.
      Note: This text box has a limit of 2,000 characters.
6. Click Finish. The box is now in a Pending state on the Matter Overview page.

- **Parent topic:** [Actions on a box](#)

Related concepts

- [Actions on a box](#)

Restarting a box

If a box undergoes several revisions, it is necessary to restart the box.

Procedure

1. Select the box that you want to restart, and then click the Restart this box link in the Available Actions area. The Restart Box dialog box appears. There are four subtabs in the Restart Box dialog box: Where and When, Who, Terms, and Notes.
2. Within the Where and When subtab, complete the following fields. By default, the dates that are entered during box creation populate this box.
   a. In the File locale check boxes, select at least one locale for this box. Options include All User Data, File Shares, Emails, Desktops, and Collaboration.
   b. In the Time interval area, use the list and start and end dates for when this box must be restarted.
   c. Click Next to access the Who subtab.
3. Within the Who subtab, complete the following fields. By default, the fields are populated with current custodian selections.
   a. Select identities to act as custodians for the box. Use Add and Remove to move custodians from the All Identities list to the Box Custodians list.
      Note: Search for a specific identity by entering all or a portion of the user's name in the Enter terms... text box and clicking Search.
   b. Select All Identities or All Identities. This value is a Boolean AND/OR value.
      Note: If you change the All Identities/All Identities setting, you might obtain unexpected results.
   c. Click Next to access the Terms subtab.
4. Within the Terms subtab, complete the following fields. By default, the fields are populated with current matter-term selections.
   a. In the Enter term and click Add text box, enter the wanted litigation term, and then click Add.
   b. Select either Any Term or All Term. This value is a Boolean AND/OR value.
      Note: If you change the Any Term/All Terms setting, you might obtain unexpected results.
   c. Use Delete and Delete All to remove added litigation terms.
   d. Click Next to access the Notes subtab.
5. Within the Notes subtab in the Notes to Data Expert text box, enter information for the box's data expert.
   Note: This text box has a limit of 2,000 characters.
6. Click Finish. The box is now in a Pending state on the Matter Overview page.

- **Parent topic:** Actions on a box

Related concepts

- Actions on a box

Cloning a box

Created boxes can be cloned to create other, identical boxes.

About this task

Note: When you are cloning a box, the settings and values of the original's box are used to prepopulate the values of the cloned box.

Procedure

1. Select the box that you want to clone, and then click the Clone this box link in the Available Actions area. The Clone Identification Box dialog box appears. There are five subtabs in the Clone Identification Box dialog box: Base Information, Where and When, Who, Terms, and Notes.
2. Within the Base Information subtab, complete these fields.
   a. In the Base Name text box, enter the name to be used for the box and its outputs.
   b. In the Description text box, enter a brief description of the box.
   c. Click Next to access the Where and When subtab.
3. Within the Where and When subtab, complete these fields:
   a. In the File locale check boxes, select at least one locale for this box. Options include All User Data, File Shares, Emails, Desktops, and Collaboration.
   b. In the Time interval area, use the list and enter the start and end dates for which this box must be revised.
   c. Click Next to access the Who subtab.
4. Within the Who subtab, complete these fields. By default, the fields are populated with current custodian selections.
   a. Select identities to act as custodians for the box. Use Add and Remove to move custodians from the All Identities list to the Current Box Custodians list.
      Note: Search for a specific identity by entering all or a portion of the user's name in the Enter terms... text box and clicking Search.
   b. Select All Identities or All Identities. This value is a Boolean AND/OR value.
      Note: If you change the All Identities/All Identities setting, you might obtain unexpected results.
   c. Click Next to access the Terms subtab.
5. Within the Terms subtab, complete the following fields. By default, the fields are populated with current matter-term selections.
   a. In the Enter term and click Add text box, enter the wanted litigation term, and then click Add.
   b. Select either Any Term or All Term. This value is a Boolean AND/OR value.
      Note: If you change the Any Term/All Terms setting, you might obtain unexpected results.
   c. Use Delete and Delete All to remove added litigation terms.
   d. Click Next to access the Notes subtab.
6. Within the Notes subtab, complete these fields.
   a. In the Notes to Data Expert text box, enter information for the box's data expert.
      Note: This text box has a limit of 2,000 characters.
7. Click Finish. The box is now in a Pending state on the Boxes and Reports page.

- **Parent topic:** [Actions on a box](#)

Related concepts

- [Actions on a box](#)

Creating a collection box

Collection boxes can be created with another box's original settings.

About this task

Note: When you are creating a box that is based on another box, the settings and values of the original's box's are used to prepopulate the collection box's values.

Procedure

1. Select the box from which you want to create a collection box, and then click the Create a Collection Box based on this box link in the Available Actions area. The Create Collection Box dialog box appears. There are two subtabs in the Create Collection Box dialog box, Base Information and Notes.
2. Within the Base Information subtab, complete these fields.
   a. In the Base Name text box, enter the name to be used for the box and its outputs.
   b. In the Description text box, enter a brief description of the box.
   c. Click Next to access the Notes subtab.
3. Within the Notes subtab, complete these fields.
   a. In the Notes to Data Expert text box, enter information for the box's data expert. Note: This text box has a limit of 2,000 characters.
4. Click Finish. The box is now in a Pending state on the Matter Overview page.

- **Parent topic:** [Actions on a box](#)

Related concepts

- [Actions on a box](#)

Creating an export box

Export boxes can be created with another box's original settings.

Procedure

1. Select the box from which you want to create an export box, and then click the Create an Export Box based on this box link in the Available Actions area. The Create Export Box dialog box appears. There are two subtabs in the Create Export Box dialog box, Base Information and Notes.
2. Within the Base Information subtab, complete these fields.
    a. In the Base Name text box, enter the name to be used for the box and its outputs.
    b. In the Description text box, enter a brief description of the box.
    c. Click Next to access the Notes subtab.
3. Within the Notes subtab in the Notes to Data Expert text box, enter information for the box's data expert.
   Note: This text box has a limit of 2,000 characters.
4. Click Finish. The box is now in a Pending state on the Matter Overview page.

- **Parent topic:** [Actions on a box](#)

Related concepts

- [Actions on a box](#)

Reports

IBM® StoredIQ eDiscovery provides various types of reports.

The following table describes all of the report types that you can create within IBM StoredIQ eDiscovery.

Table 1. IBM StoredIQ eDiscovery report types

Report Type	Description
Compliance Report	The Compliance Report indicates whether data is in compliance with established guidelines, specifications, or governance policies. The generated report provides detailed information about data-object compliance in these areas: • Compliance Overview • Distribution by Category • Distribution by Data Source • Distribution by Size of Data Object • Distribution by Data Owner • Distribution by Last Modified Date
Data Assessment Report	The Data Assessment Report assesses data in terms how it interacts with IT policies. The generated report provides detailed information in these areas: • Distribution by Data Source • Distribution by Size of Data Object • Distribution by Data Owner • Distribution by Last Accessed Date • Distribution by Last Modified Date • Distribution by Category
Data Topology Report	The Data Topology Report describes how data is distributed across your organization. The generated report provides detailed information in these areas: • Overall Summary • Distribution by Data Source • Distribution by Category • Distribution by Data Owner • Distribution by Last Accessed Date • Distribution by Last Modified Date
Overlay Hit Report	The Overlay Hit Report compares data's responsiveness for various overlays. The generated report provides a detailed analysis about overlay information in these areas: • Overlay Summary • Overlay Results
CSV Exceptions	The CSV Exceptions Export Report lists the encountered exceptions. The generated report provides a detailed analysis about exceptions that are generated during an action: • ID • Source Object Path • Source Object Container Path • Source Object Name • Source Object Version • Source Object Created Date

Export Report	• Source Object Last Modified Date • Destination Object Path • Destination Object Container Path • Destination Object Name • Destination Object Version • Destination Object Created Date • Destination Object Last Modified Date • Category • Token
CSV Infoset Data Object Export Report	The CSV Infoset Data Object Export Report lists the data objects within a box. The generated report provides the following columns: • Name • Path • Container Path • Volume Name • Object Type • Owner • Size • Ctime • Mtime • Atime • Hash

- **Parent topic:** [Actions on a box](#)

Related concepts

- [Actions on a box](#)

Related tasks

- [Deleting a box](#)

Running a report on a box

Follow this procedure to run various types of reports.

Procedure

1. Select the box on which you want to run a report, and then click the Run a report on this box link in the Available Actions area. The Run a Report Box dialog box appears. There are two subtabs in the Run a Report Box dialog box, Select Report and Report Options.
2. Within the Select Report subtab, select the report that you want to run. When a report is selected in the Select Report subtab, details for that report appear within the dialog box, including the report's name and description.
   a. Search for a specific report by entering all or a portion of the report's name in the Enter terms... text box and clicking Search.
   b. Click Next to access the Report Options subtab.
3. Within the Report Options subtab, complete these fields.
   a. In the Report Name text box, verify that you are running the wanted report. If not, this field can be edited, so you can enter the name of the report you want to run.
   b. To have an email that is sent after the report is complete, select the Send an email when the process is complete check box. Then, enter the email addresses of the individuals to whom this report is sent in the Enter notification emails text box. Note: If you enter multiple email addresses, they must be separate with a comma.
4. Click Run to run the report. The box reverts to the Pending state when the report is running. When the report is complete, click the report's name under Reports, the report can be viewed from the user interface.

- **Parent topic:** [Actions on a box](#)

Related concepts

- [Actions on a box](#)

Deleting a box

Boxes that are no longer needed within a matter can be deleted.

Procedure

1. Select the box that you want to delete, and then click the Delete this box link in the Available Actions area. The Delete Box Confirmation window appears.
2. Click OK to delete the box. The box is deleted and removed from the matter. It is no longer seen within the Matter Overview page.
   Note: Deleting a box does not affect any boxes that are created with it or cloned from it. However, its deletion means that it is no longer available to create new boxes or reports.

- **Parent topic:** Actions on a box

Related concepts

- Actions on a box

Managing IBM Stored IQ Policy Manager

IBM® Stored IQ Policy Manager provides an interface for you to view available policies, create, edit, delete, or run a policy or a report.

The following topics describe in detail how to use IBM Stored IQ Policy Manager to act on policies or reports.

- [Overview of IBM Stored IQ Policy Manager](#)
  IBM Stored IQ Policy Manager enables policy execution at scale within IBM StoredIQ Platform.
- [IBM Stored IQ Policy Manager administration](#)
  IBM Stored IQ Policy Manager administration describes how to use IBM Stored IQ Policy Manager to manage policies.

- **Parent topic:** [Managing the IBM StoredIQ products](#)

Overview of IBM Stored IQ Policy Manager

IBM® Stored IQ Policy Manager enables policy execution at scale within IBM StoredIQ Platform.

The IBM StoredIQ policies set rules and criteria for the system to act on data. The policies are defined by these factors:

- Type: The type of action that the user wants the system to act upon the data.
- Criteria: Well-defined expressions that outline exactly what data is eligible for the policy to act upon.
- Scope: A mechanism to define what set of data is to be acted upon.
- Schedule: A schedule to define when the system must run the policy.

IBM Stored IQ Policy Manager's reports record what and when certain actions were completed and what data was affected by the policy's execution.
IBM Stored IQ Policy Manager provides the Policy Dashboard, which provides a visual means for the user to quickly find and identify the policy to work on. The Policy Dashboard has these functions:

- Lists all policies available. This list can be sorted by the Created Date of the Policies and can be filtered by Name.
- Shows high-level policy details. Details include policy name, description, created date, number of runs associated with the policy, the assigned personas to the box, and details about the policy's type, criteria, scope, and schedule.
- Allows for policies to be selected for view of details of the specific policies.
- Provides a way to create a new policy.
- Provides a way to enable or disable the policy from running on its defined schedule.
- Allows for policies to be deleted from the Policy Dashboard.

Related concepts

- [IBM Stored IQ Policy Manager administration](#)

IBM Stored IQ Policy Manager administration

IBM® Stored IQ Policy Manager administration describes how to use IBM Stored IQ Policy Manager to manage policies.

- [Policy Dashboard view](#)
  Policy Dashboard provides high-level information about a policy, its name, description, creation time, creation date, and whether the policy is turned on or off. It helps you to find and identify policies of your interest.
- [Actions on a policy](#)
  When a policy is created, multiple actions are available to help you act on data objects as part of the policy-management process.

Related concepts

- [Overview of IBM Stored IQ Policy Manager](#)

Policy Dashboard view

Policy Dashboard provides high-level information about a policy, its name, description, creation time, creation date, and whether the policy is turned on or off. It helps you to find and identify policies of your interest.

You can view a policy's status on the Policy Dashboard, which also provides critical pieces of information for each policy.

- A policy's state, its name, description, and the date and time at which it was created
- Whether the policy is On or Off
- Its assigned action
- When the policy is scheduled to run, and what it runs on next
- What filters are used
- Time and date of the policy's next run
- How many times the policy is successfully completed
- The Admin and Policy User for that policy.

From the Policy Dashboard, you can accomplish these tasks.

- View a policy in detail by clicking the policy that you want to see.
- Enable or disable a policy by turning it On or Off.
- Contact either the Administrator or Policy User about a specific policy by clicking that person's name.
- Delete a policy by clicking X within the policy.
- Search for a policy by filtering the existing policies.
- Create a policy by clicking Create Policy.

Table 1. Policy Manager: Policy Dashboard Options and Descriptions

Option	Description
Icon.	An icon is displayed to the left of the policy name, indicating its status: Active 　An active policy is correctly configured and ready to start any defined scheduled and recurring runs. You can also use the Run Now option. Pending 　A pending policy is incomplete and not yet available for execution. Failed 　A failed policy is one that failed creation and is not available for use.
Name	Name of the policy.
Created	Time and date at which the policy was created.
On/Off	Indicates whether a policy is enabled or disabled. - If On is blue, the policy is enabled and available for use. - If Off is blue, the policy is disabled and unavailable for use. Click On or Off to turn the policy On or Off.
Action	The action that the policy takes. Automated Deletion 　Deletes files on an automated schedule. This policy also has a manual option.
When	The policy's execution schedule. Specifically, it lists whether the policy runs daily, weekly, or monthly, on what day, and at what time.
What	What infoset is used as part of this policy's execution.
Filter	What filter is used as part of this policy's execution.
Next Run	The time and date at which the policy runs next time.

Complete	The number of times that the policy completed its execution.
Admin	The policy's administrator is listed here. To email this individual, click the name.
Policy user	The policy user is listed here. To email this individual, click the name.

- **Creating a policy**
  You can create policies from the Policy Dashboard.
- **Viewing a policy**
  You can view the contents of an existing policy, including an abbreviated dashboard, its runs, reports, and available actions by accessing the policy itself.
- **Enabling or disabling a policy**
  Within IBM Stored IQ Policy Manager, enabled policies mean that they are available for use, while disabled policies are unavailable for use.
- **Emailing a contact person**
  You can email a contact person of a policy by clicking the name of that person on the Policy Dashboard.
- **Deleting a policy**
  Deleting a policy deletes all pending and completed runs of that policy and its reports.

- **Parent topic:** IBM Stored IQ Policy Manager administration

Related concepts

- IBM Stored IQ Policy Manager administration

Creating a policy

You can create policies from the Policy Dashboard.

Procedure

1. From the Policy Dashboard screen, click Create Policy. The Create Policy dialog box appears.
2. Select the Automated Deletion action (it is the only available action) for the policy, and then click Next. When an action is selected for the policy, the Create Policy dialog box shows four fields:
   - Base Information
   - What
   - Filter
   - When
3. In the Base Information tab, complete these options.
   a. In the Name text box, enter a unique name for this policy.
   b. Optional: In the Description text box, enter a description of this policy.
   c. In the Admin Contact list, select the individual that acts as the administrative contact for this policy. The users in this list are populated from the IBM® StoredIQ Administrator user management feature.
   d. In the Policy Contact list, select the individual that acts as the policy user contact for this policy. The users in this list are populated from the IBM StoredIQ Administrator user management feature.
   e. Click Next.
4. In the What tab, select the infoset for the policy run. If needed, enter the name of the infoset in the Enter terms... text box and then click Search to filter the infosets. Administrators can build more infosets with IBM StoredIQ Administrator.
5. Click Next.
6. In the Filter tab, select a filter for the policy. If needed, enter the name of the filter in the Enter terms... text box and then click Search to search the infosets. Administrators can build more filters with IBM StoredIQ Data Workbench.
7. Click Next.
   Note: View the filters source code by clicking View Source Code. A separate dialog window opens, showing the attributes that are used to build the selected filter.
8. In the When tab, set the time and date when the policy runs. Select either One time or Repeating.
   a. If you select the One time option, complete the On and At text boxes. Select the date and time on which the policy runs.
   b. If you select the Recurring option, select either the Weekly or Monthly list option, its frequency, and the day, time, and start dates the policy runs.
9. Click Finish. The newly created policy appears in the Runs and Reports area of the Policy Dashboard and reflects the values set during the policy creation.

- **Parent topic:** [Policy Dashboard view](#)

Related concepts

- [Policy Dashboard view](#)

Viewing a policy

You can view the contents of an existing policy, including an abbreviated dashboard, its runs, reports, and available actions by accessing the policy itself.

Procedure

1. Within the Policy Dashboard, click the policy that you want to view. The policy opens within the Policy View page.
2. Within the Policy View page, these areas can be seen.

Table 1. Policy View page runs options and descriptions

Run Option	Description
Name	Name of the policy, which was configured during the policy creation.
Started	Start time and date at which the policy was last run.

Type	Type of the policy's run:
	
	Run Now
	    The Run Now icon indicates that a policy runs immediately when Run Now is clicked.
	Recurring
	    The Recurring icon indicates that a policy is scheduled to run many times.
	Scheduled
	    The Scheduled icon indicates that a policy is schedule to run one time.

Status	Policy's status about its run.
	
	Completed
	    The policy completed its run successfully.
	Queued
	    The policy is scheduled to run and is queued to do so.
	Failed
	    The policy failed its run.

Table 2. Policy View Summary Options and Descriptions

Summary Option	Description
Name	Name of the policy, which was configured during the policy creation.
Type	Type of the policy's run.
Description	Description of the policy.
Started	Start time and date at which the policy was last run.
Duration	Duration of the policy's most recent execution.

Table 3. Policy View Report Options and Descriptions

Report Options	Descriptions
Name	Name of the report.
Created	Time and date that the report was created.
Status	Report status. Options include Report processing, Success, and Failed.

Table 4. Policy View Available Actions and Descriptions

Available Action Options	Descriptions
	Available actions: Run now

Run now
> By clicking this option, the selected policy runs when it is available in the queue.

Action Execute CSV exception report on run
> By clicking this option, the CSV Exception Report can be created.

Delete the run
> By clicking this option, the selected run along with its associated pending and completed reports is deleted.

3. Click the Policy Dashboard link within the breadcrumb trail to return to the Policy Dashboard.

- **Parent topic:** [Policy Dashboard view](#)

Related concepts

- [Policy Dashboard view](#)

Enabling or disabling a policy

Within IBM® Stored IQ Policy Manager, enabled policies mean that they are available for use, while disabled policies are unavailable for use.

About this task

To enable or disable a policy:

Procedure

In the Policy Dashboard, locate the policy that you want to enable or disable.

a. To enable a policy, click On. Verify that it is blue. The policy is now available for use.
b. To disable a policy, click Off. Verify that it is blue. The policy is now unavailable for use.

- **Parent topic:** Policy Dashboard view

Related concepts

- Policy Dashboard view

Emailing a contact person

You can email a contact person of a policy by clicking the name of that person on the Policy Dashboard.

About this task

To email a contact person of a policy from the Policy Dashboard:

Procedure

1. Find the policy that contains the contact person you want to email.
2. Click the name of the Admin or Policy User that you want to email. An email to that individual automatically opens in your email client with the name of the matter as the subject of the email.

- **Parent topic:** [Policy Dashboard view](#)

Related concepts

- [Policy Dashboard view](#)

Deleting a policy

Deleting a policy deletes all pending and completed runs of that policy and its reports.

Procedure

1. On the Policy Dashboard, select the policy that you want to delete.
2. Click X in the right corner of that selected policy. The Delete Policy confirmation dialog appears.
   Note: If you delete the policy, all of the pending and completed policy runs and its reports are deleted.
3. In the Delete Policy confirmation dialog, click Delete.

- **Parent topic:** Policy Dashboard view

Related concepts

- Policy Dashboard view

Actions on a policy

When a policy is created, multiple actions are available to help you act on data objects as part of the policy-management process.

These actions are available in the Available Actions area:

Run now
  By clicking this option, the selected policy can be run after it is available in the queue.
Execute CSV exception report on run
  By clicking this option, the CSV Exception List Report can be created.
Delete the run
  By clicking this option, the selected run along with its associated pending and completed reports is deleted.

Additionally, you can also edit a policy by clicking Edit Policy.

- Editing a policy
  Editing a policy is similar to creating policy in the way that you can modify any of the fields during editing.
- Running a policy
  Follow these steps to run a policy.
- Running a report
  To work with the IBM StoredIQ policies, you can generate reports.
- Viewing a policy report
  You can view Execute CSV exception report on run as a .CSV file.
- Deleting a run
  Policy executions and their associated reports can be deleted.
- Stopping a run
  You can stop policies that are currently being run. However, actions that are already started by the run persist and the run cannot be restarted.

- **Parent topic:** IBM Stored IQ Policy Manager administration

Related concepts

- IBM Stored IQ Policy Manager administration

Editing a policy

Editing a policy is similar to creating policy in the way that you can modify any of the fields during editing.

Procedure

1. In the Policy View screen, click Edit Policy within the Policy Dashboard. The Edit Policy dialog box appears.
2. Edit the fields of the policy during the policy creation.

- **Parent topic:** [Actions on a policy](#)

Related concepts

- [Actions on a policy](#)

Running a policy

Follow these steps to run a policy.

Procedure

1. In the Policy View screen, click Run Now. The Run Policy dialog box appears.
2. In the Run Policy dialog box, enter a base name and a description for the policy run. This base name is used to name the box and its reports.
3. Click Run.

- **Parent topic:** [Actions on a policy](#)

Related concepts

- [Actions on a policy](#)

Running a report

To work with the IBM® StoredIQ policies, you can generate reports.

Procedure

1. In the Available Actions area, select Execute CSV exception report on run. The Run Exception Report dialog box appears.
2. In the dialog box, complete these fields.
   a. The Name text box is auto-filled, but the report name can be edited or changed by entering your text into the Name text box.
   b. To send the generated report to different email addresses, select the Send an email when the process is complete check box.
   c. In the Enter notification emails text box, enter the email addresses to which the generated report is sent.
3. Click Run.

- **Parent topic:** Actions on a policy

Related concepts

- Actions on a policy

Viewing a policy report

You can view Execute CSV exception report on run as a .CSV file.

Procedure

1. In the Policy View window, select the run for which you want to view the report.
2. In the Reports area, click the report that you want to view. A folder downloads.
   Note: Only reports in the Completed state can be viewed.
3. Click the folder and then open the .CSV file to view the report.

- **Parent topic:** [Actions on a policy](#)

Related concepts

- [Actions on a policy](#)

Deleting a run

Policy executions and their associated reports can be deleted.

Procedure

1. In the Available Actions area, click Delete run. The Delete Run confirmation appears, verifying that you want to delete the selected run along with its associated pending and completed reports.
2. Click Delete.

- **Parent topic:** [Actions on a policy](#)

Related concepts

- [Actions on a policy](#)

Stopping a run

You can stop policies that are currently being run. However, actions that are already started by the run persist and the run cannot be restarted.

Procedure

1. Click Stop. The Stop Run confirmation box appears.
2. Click Stop.

- **Parent topic:** [Actions on a policy](#)

Related concepts

- [Actions on a policy](#)