

IBM StoredIQ

Integration with a governance catalog



Note

Before using this information and the product it supports, read the information in [Notices](#).

This edition applies to Version 7.6.0.22 of product number 5724M86 and to all subsequent releases and modifications until otherwise indicated in new editions.

© **Copyright International Business Machines Corporation 2017, 2020.**

US Government Users Restricted Rights – Use, duplication or disclosure restricted by GSA ADP Schedule Contract with IBM Corp.

Contents

- IBM StoredIQ product library..... iv**
- Contacting IBM StoredIQ customer support.....v**
 - Contacting IBM..... v
- Integrating with a governance catalog..... 1**
 - Discovery and classification of unstructured data..... 1
- Synchronizing data..... 3**
- Synchronized objects.....5**
 - Instance assets..... 6
 - Volume assets..... 7
 - InfoSet assets..... 8
 - Filter assets..... 11
 - The GovernanceDataClasses cartridge..... 13
 - The Governance Analytics action..... 14
- Synchronization API..... 15**
- Notices.....17**
 - Trademarks..... 18
 - Terms and conditions for product documentation..... 19
 - IBM Online Privacy Statement..... 19
- Index..... 21**

IBM StoredIQ product library

The following documents are available in the IBM® StoredIQ® product library.

- *IBM StoredIQ Overview Guide*
- *IBM StoredIQ Deployment and Configuration Guide*
- *IBM StoredIQ Data Server Administration Guide*
- *IBM StoredIQ Administrator Administration Guide*
- *IBM StoredIQ Data Workbench User Guide*
- *IBM StoredIQ Cognitive Data Assessment User Guide*
- *IBM StoredIQ Insights User Guide*
- *IBM StoredIQ Integration Guide*

The most current version of the product documentation can always be found online: https://www.ibm.com/support/knowledgecenter/en/SSSHEC_7.6.0/welcome/storediq.html

Contacting IBM StoredIQ customer support

For IBM StoredIQ technical support or to learn about available service options, contact IBM StoredIQ customer support at this phone number:

- 1-866-227-2068

Or, see the Contact IBM web site at <http://www.ibm.com/contact/us/>.

IBM Knowledge Center

The IBM StoredIQ documentation is available in [IBM Knowledge Center](#).

Contacting IBM

For general inquiries, call 800-IBM-4YOU (800-426-4968). To contact IBM customer service in the United States or Canada, call 1-800-IBM-SERV (1-800-426-7378).

For more information about how to contact IBM, including TTY service, see the Contact IBM website at <http://www.ibm.com/contact/us/>.

Integrating with a governance catalog

Data offers no competitive advantage until it can be put to use. Effective information governance strategies help enable the self-service information discovery that teams need to find, evaluate, understand, and use data. Get acquainted with the features that support a unified approach to enterprise information governance and to understand the significance of the integration of IBM StoredIQ with IBM Information Governance Catalog or IBM Cloud Private for Data for that strategy.

Overview of information governance

The significant growth in data volume, velocity, and variety requires enterprises to govern and manage the risk, quality, and cost of that data and provide higher confidence for its use. This is the domain of information governance. Information governance refers to the findability, usability, and integrity of the data in an enterprise.

Also, aligning the business and IT perspectives of the organization can be a particular challenge. Enforcing business guidelines on an ever amassing mountain of both structured and unstructured information can seem impossible.

Up to now, structured information, which means information that has a predefined data model or fits into relational tables, and unstructured information such as email messages, word processing documents, audio or video files, collaboration software, or instant messages had to be governed separately.

A unified governance architecture is designed to simplify and streamline information governance to support a connected and collaborative governance solution for the enterprise, enabling teams to find the data they need, evaluate its quality at a glance, and feel confident that it can be used to gain business value. Integrating IBM StoredIQ with IBM Information Governance Catalog or IBM Cloud Private for Data plays a significant role in achieving this goal by providing for a way to manage all enterprise data in one place.

Discovery and classification of unstructured data

Unify information governance of structured and unstructured information: make data that is governed by IBM StoredIQ available in a governance catalog for classification and have the classification reflected in IBM StoredIQ.

IBM StoredIQ artifacts such as infosets, volumes, and filters are represented as information assets within the governance catalog so that you can get a unified view of the content and can govern these assets through glossary terms, governance rules, business rules, and so on. In addition, you can trace the usage of those information assets through data lineage reports, for example, for compliance reporting.

Information classifications such as data class definitions are synchronized between IBM StoredIQ and the governance catalog so that those classifications are used uniformly across structured and unstructured content in the enterprise. Thus, you can find data of a certain data class in structured and unstructured content, for example, to detect policy violations for placement of sensitive data, so that you can apply the required actions such as preventing copying of sensitive information.

The synchronization of data from IBM StoredIQ to the governance catalog happens on a predefined schedule.

Synchronization of data from the governance catalog to IBM StoredIQ is triggered by Kafka events.

Information discovery

Integrating IBM StoredIQ with a governance catalog lets governance users discover and work with unstructured data that is maintained in IBM StoredIQ. An IBM StoredIQ data expert can get insight into

content classifications, additional asset attributes, and term assignments to be able to apply the required actions.

Unstructured data sources in the asset catalog

The following IBM StoredIQ objects are made available as information assets in the governance catalog: volumes, infosets, and filters. These assets are grouped by originating IBM StoredIQ system, which is represented as the parent instance information asset. Thus, you can access unstructured data sources such as email servers, file shares, and so on through Information Governance Catalog or IBM Cloud Private for Data.

Content classifications, asset attribute information, and term assignments in IBM StoredIQ

Data class definitions that are applicable for unstructured content, any custom asset attributes and terms that are maintained and applied in the governance catalog are synchronized to IBM StoredIQ. Thus, the IBM StoredIQ data expert can identify and act on specific data.

Data classification

Integrating IBM StoredIQ with a governance catalog provides the means to classify enterprise content consistently.

You have a uniform view on data classifications in the enterprise and how they are defined. And you can apply data classes consistently to structured and unstructured content to identify and find key data, for example, to detect policy violations for placement of sensitive data.

In terms of information governance, a *data class* categorizes data according to type and usage. *Data classification* is the process of assigning a data class to an information asset. Applying the classification rules of a data class lets you detect data that contains elements of a given kind, for example, phone or credit card numbers.

Use of data class detection logic as cartridge actions in IBM StoredIQ

In the governance catalog, you create or update data classes of type Regex that can be applied to unstructured content. When synchronized, these data classes are reflected in IBM StoredIQ as filters.

By applying the regexes that are derived from the data classes, you analyze and classify data, and make the classified data searchable and detectable in IBM StoredIQ. Classification results for IBM StoredIQ infosets and volume contribution information are available in the governance catalog.

Classification of infoset assets

Data classifications of infoset assets can be derived from filter, union, expand, and collapse operations. Applying a filter is the basis for all classification. During synchronization, the number and size of objects in an infoset that match a certain data class is calculated (calculation of the so-called *data class contributions*). If you apply a filter that is linked to a data class, for example, all objects of the new infoset are classified by that data class.

The calculated classification is inherited upward and downward in the infoset's ancestry. The calculation results in either the match type Exact or the match type GreaterThan:

- Exact: The respective data class contributes exactly the given number and size of objects to the infoset.
- GreaterThan: The respective data class contributes at least the given number and size of objects to the infoset. In cases where the data class contributes to the infoset but the number and the size of objects cannot be determined (which might happen for infosets created through expand or collapse operations), both values are set to zero.

Synchronizing data

Enable the synchronization and determine which IBM StoredIQ objects are published to the governance catalog.

As a prerequisite, a working deployment of one of these products must be available:

- IBM Information Server. The minimum required version is Version 11.7.0. However, some synchronization features require Version 11.7.0 Fix Pack 1 to be installed.
- IBM Cloud Private for Data. The minimum required version for full support is Version 1.2.

When synchronizing with the governance catalog in an IBM Information Server deployment, these additional prerequisites apply:

- In Information Governance Catalog or IBM Cloud Private for Data, event notification must be enabled.
- When synchronizing with Information Governance Catalog, the deletion of import areas must be enabled in IBM Metadata Asset Manager and import of duplicate assets must be allowed. Previewing of express imports when the import contains duplicates or invalid identities must be disabled. For more information on IBM Metadata Asset Manager import settings, see [the topic about import settings](#) in the Information Server product documentation.

Some of the tasks must be completed as the system administrator, others as the data expert.

To have data synchronized between IBM StoredIQ and a governance catalog:

1. As the system administrator, enable the synchronization in IBM StoredIQ.

You can do it either during the initial configuration of the application stack or at any later time:

- [Installing and configuring the application stack](#)
- [Configuring the application stack to synchronize data with the governance catalog](#)

2. As system administrator, configure which volumes are not to be published to the governance catalog. Volumes are included in the synchronization by default. To exclude a volume from synchronization, edit the volume and clear the **Publish to catalog** checkbox.

If you want to exclude a volume of a type that cannot be edited from IBM StoredIQ Administrator, contact your IBM Services representative or IBM Support.

Important: You cannot exclude volumes contributing to infosets that are published to the governance catalog. If you deselect the **Publish to catalog** option for such a volume, this setting is automatically reverted in the next synchronization run.

For each volume, information is synchronized as described in [“Volume assets” on page 7](#).

3. As a data expert, configure which infosets are published to the governance catalog.

By default, infosets are not published to the governance catalog. To include a new infoset in the synchronization, select the **Publish to catalog** option when you create the infoset. At any time, you can change the setting for an existing infoset by editing the infoset. When you revert the setting so that the infoset is no longer published, the respective asset is removed from the governance catalog in the next synchronization run.

Some infosets might be published to the governance catalog as so-called *inferred* objects although the respective option is not set. This type of synchronization happens when the infoset serves as direct input to an object that is selected for publishing. The infoset is then required for proper object modeling in the governance catalog.

For each infoset, information is synchronized as described in [“Infoset assets” on page 8](#).

4. As a data expert, configure which filters are synchronized with the governance catalog.

Filters are automatically created for each new data class in the governance catalog that is reflected in the **GovernanceDataClasses** cartridge (filter names starting with **Contains governance data class**). These filters are synced back to the governance catalog by default.

Filters that you create manually in IBM StoredIQ are not automatically synchronized. To include such a filter in the synchronization, select the **Publish to catalog** option when you save the new filter. At any time, you can change the setting for an existing filter, even a system generated one, by editing the filter.

Filters might also be synchronized as inferred objects if they are used to create an info set that is published to the catalog. Here, too, the filter is required for proper object modeling in the governance catalog.

For each filter that is included in the synchronization, information is synchronized as described in [“Filter assets” on page 11](#).

The objects that are marked for publishing to the catalog (whether regularly or as an inferred object) are synchronized when the application stack services are started and afterward on the configured schedule.

However, when changing the set of objects to be synchronized, you might not want to wait for the next synchronization run but want to see the changes reflected instantly in the governance catalog. For this purpose, you can request immediate synchronization. For details, see [“Synchronization API” on page 15](#).

Remember that, after the synchronization, you might find assets in the governance catalog that were not marked for publishing but published as inferred objects.

If any other results in the governance catalog do not match your expectations, check your settings in IBM StoredIQ and check the `/var/siq/log/tomcat.log` file for any synchronization errors.

Synchronized objects

Find information about the objects that are synchronized between IBM StoredIQ and a governance catalog.

IBM StoredIQ objects in the governance catalog

These asset types are made available in the governance catalog to provide access to unstructured data sources such as email servers, file shares, and so on:

Instance

Each asset of this type represents an IBM StoredIQ server instance that governs unstructured data sources. For details about the information that is synchronized, see [“Instance assets” on page 6](#).

Volume

Each asset of this type represents an IBM StoredIQ volume. Volumes are published to the catalog by default, but you can exclude volumes from the synchronization. For details about the information that is synchronized, see [“Volume assets” on page 7](#).

Infoset

Each asset of this type represents an IBM StoredIQ infoset. For each infoset, you can decide whether it is to be published to the governance catalog. For details about the information that is synchronized, see [“Infoset assets” on page 8](#).

Filter

Each asset of this type represents an IBM StoredIQ filter, either a filter you created in IBM StoredIQ or a system generated filter that is associated with a data class that was synchronized from the governance catalog. For details about the information that is synchronized, see [“Filter assets” on page 11](#).

These asset types are populated when the synchronization is run for the first time. The new asset types and assets show up in the **Unstructured Data Sources** asset family in the governance catalog. They are specific to the instance of the catalog. They are also part of the metadata repository and can be accessed by other applications in Information Server or IBM Cloud Private for Data such as enterprise search.

After each synchronization, you might see so-called *inferred* assets in the governance catalog. Inferred assets are objects that were synchronized to the governance catalog although not selected for publishing because they serve as direct input to objects that are selected for publishing. Publishing such objects is required for proper object modeling in the governance catalog. In the IBM StoredIQ Data Workbench user interface, the value **Yes (Inferred)** for any **Publish to Catalog** entry indicates such objects.

Updates to published objects in IBM StoredIQ are propagated to the governance catalog on the configured schedule. Updates include adding objects, and deleting objects and all of their references from the governance catalog if, at any time, you decide to no longer publish specific objects to the catalog. Updates to IBM StoredIQ assets in the governance catalog are not necessarily propagated to IBM StoredIQ. For more information, see the respective asset descriptions.

When you delete a published object in IBM StoredIQ, the respective asset is removed from the governance catalog as soon as possible. If other objects have dependencies on the asset, these objects must be updated. So, it might take some time until the asset is actually removed from the governance catalog. All information that you added to the asset in the governance catalog, such as custom attributes, is also deleted. Any IBM StoredIQ asset that you delete in the governance catalog is republished to the catalog in the next synchronization run.

Data classes synchronized to IBM StoredIQ

In the governance catalog, you create or update data classes of type Regex that can be applied to unstructured content. The regexes of all data classes that qualify for synchronization are bundled and synchronized to IBM StoredIQ in form of the **GovernanceDataClasses** cartridge.

Applying the logic of a cartridge requires defining and running a Step-up Analytics action. Therefore, the **Governance Analytics** action is created on the first synchronization run. Use this action to apply the logic that is contained in the **GovernanceDataClasses** cartridge to an infoaset.

For details, see [“The GovernanceDataClasses cartridge” on page 13](#) and [“The Governance Analytics action” on page 14](#).

In addition, a filter is created in IBM StoredIQ for each data class that is synchronized. See [“Filter assets” on page 11](#) for details. You can use such a filter to create infosets with data that matches the respective data class after you run the **Governance Analytics** action.

The initial propagation of data classes from the governance catalog to IBM StoredIQ also happens when the synchronization is run for the first time.

In subsequent synchronizations, only new or modified data classes are propagated. The **GovernanceDataClasses** cartridge is updated to contain the new detector logic. Depending on the changes, you might need to run the **Governance Analytics** action again to apply the new logic and to reclassify your data. Existing filters remain unchanged and new filters are added as appropriate. Filters that stem from data classes that no longer exist in the governance catalog or are no longer eligible for synchronization are still available but are marked as *deprecated* filters in the IBM StoredIQ Data Workbench user interface.

In case of conflicts during the synchronization, the information in the governance catalog takes precedence in resolving the conflict.

Links to objects in the originating application

Both applications provide links to synchronized objects in the originating application. When users follow such a link, they must log in to the application with the appropriate access rights:

- Login to Information Governance Catalog requires an account that has at least one Information Governance Catalog security role.
- Login to IBM Cloud Private for Data requires an account that has the permission to access the catalog.
- Login to IBM StoredIQ Administrator requires an account that has the Admin role.
- Login to IBM StoredIQ Data Workbench requires an account that has at least the Data User role.

If you are not logged in to IBM StoredIQ when you follow a direct link to an IBM StoredIQ object, you are redirected to the IBM StoredIQ login page and the original link information is lost. If you are already logged in, what you see depends on the IBM StoredIQ user role you have:

- Admin role: any object.
- Any Data User or Insights User role: the user infosets that were created by other data users and any filters. For links to non-public system infosets or user infosets that were created by an administrator, you are redirected to an infoaset you have access to.

Tip: The direct links to assets in Information Governance Catalog might not work in some web browsers. If you are not redirected to the asset after you log in to the application, install the fix that is mentioned at the end of the document about turning off the referer header checker in IBM Information Server and add the domain where the IBM StoredIQ user interface runs to the list of allowed domains as described.

Instance assets

Each asset of the type Instance represents an IBM StoredIQ server instance that governs unstructured data sources.

The following table lists the information available in the asset catalog after an IBM StoredIQ object is published to the governance catalog.

Property	Notes
Long description	System generated. It includes the URL to the application stack of this IBM StoredIQ instance.
Tool ID	The instance's internal ID in IBM StoredIQ.

The **Contains Assets** section lists all volumes, infosets, and filters that are synchronized for this instance.

In the governance catalog, you can add the following information:

- A short description of up to 255 characters
- Labels
- Stewards
- Information about term assignments
- Information about rules usage
- An alias (business name)
- A primary and a secondary contact, such as an administrator, the responsible manager, or a data expert supervisor
- Custom attributes

This information is not made available in IBM StoredIQ.

Volume assets

Each asset of the type Volume represents a IBM StoredIQ volume. Volumes are published to the catalog by default, but you can exclude a volume from the synchronization.

Some volume types cannot be edited from IBM StoredIQ Administrator. If you want to exclude such a volume, contact your IBM Services representative or IBM Support.

Volumes contributing to infosets that are published to the governance catalog cannot be excluded from the synchronization. Such volumes are marked as **Yes (Inferred)** in any **Publish to Catalog** entry in IBM StoredIQ.

The following table lists the information available in the asset catalog after an IBM StoredIQ object is published to the governance catalog.

Property or section	Notes
Context	A link to the details of the instance with which the asset is associated.
Long description	System generated. It lists the type of volume and the name of the parent data server.
Data server	The name of the parent data server.
Type	The volume's type: DiscoveryExport , Primary , Retention , or System
Source type	The source or connection type for the volume. For a list of possible source or connection types for primary volumes, see Adding a primary volume . Possible source types for retention or discovery export volumes are CIFS and NFS.
Server	The name of the server where the volume is found.
Total objects	The number of data objects that are found in the volume.

Property or section	Notes
Size	The total size of the data objects in the volume.
Last harvest date	The date and time that the volume was last harvested.
Tool ID	The volume's internal ID in IBM StoredIQ.
URL	The URL to the list of volumes in IBM StoredIQ Administrator.
Infosets	A list of the synchronized infosets to which this volume contributes.

In the governance catalog, you can add the following information:

- A short description of up to 255 characters
- Labels
- Stewards
- Information about term assignments
- Information about rules usage

This information is not made available in IBM StoredIQ.

In addition, you can apply custom attributes to the asset. These custom attributes are made available in IBM StoredIQ.

You can access this catalog information from the volume details in the IBM StoredIQ Administrator user interface:

- The **Volume in Catalog** field provides a link for accessing the volume asset in the governance catalog. When you follow this link, you must sign in to Information Governance Catalog or IBM Cloud Private for Data, which requires an account with appropriate security roles.
- The **Asset Attributes** field provides a link for viewing synchronized custom attributes in IBM StoredIQ if any were assigned to the asset in the governance catalog.

Infoset assets

Each asset of the type Infoset represents an IBM StoredIQ user infoset. For each user infoset, you can decide whether it is to be published to the governance catalog.

System infosets cannot be selected for publishing. However, they might be added to the governance catalog as inferred objects for reflecting the ancestry of an infoset. The same applies to user infosets that are not marked for publishing but serve as input for infosets that are selected for synchronization. You can edit an inferred asset's details like any other's.

In the IBM StoredIQ Data Workbench user interface, the value **Yes (Inferred)** for any **Publish to Catalog** entry indicates such objects.

If you require system infosets to be synchronized as regular objects, contact your IBM representative.

The following table lists the information available in the asset catalog after an IBM StoredIQ object is published to the governance catalog.

Property or section	Notes
Context	A link to the details of the instance with which the asset is associated.

Property or section	Notes
Long description	This information originates in IBM StoredIQ. Therefore, ensure that the object's description there is as detailed as possible to maximize the value for a governance user.
Assigned to terms	<p>Any terms that might be assigned automatically during the synchronization from IBM StoredIQ and their respective category hierarchies. The term assignment is based on infoSet or classification contribution. In addition, you can directly assign terms to the asset, which are then made available in IBM StoredIQ as asset attributes.</p> <p>If a term that is inherited from a data class is removed from an asset in the governance catalog, this term is not automatically reassigned. If such a term is removed from the related data class, it is also removed from the asset.</p>
Include for governance	<p>One of the following values:</p> <p>Regular All infoSets and their associated objects in this infoSet's ancestry are synchronized to the governance catalog per configuration.</p> <p>Inferred Partial To be referenced in the information about the operations that were used to create the infoSet, the direct ancestor infoSets must be present in the governance catalog. For this purpose, these infoSets are synchronized to the governance catalog even if they are not configured for being published. These infoSets have the type Inferred Partial. For infoSets of this type, no references to the originating operations are created.</p> <p>If the infoSet is included in the synchronization later, any missing information is added to the asset in the catalog and its type is changed to Regular.</p>
Type	The infoSet type as created in IBM StoredIQ: System or User
Composition	The composition of the data that is contained within an infoSet: MixedLevel , if the infoSet contains top-level and child-level objects, or TopLevel .
Total objects	The number of data objects in the infoSet.
Size	The size of the infoSet.
Creator	The person who originally created the infoSet.
Created	The date and time the infoSet was created.

Property or section	Notes
Data classification	Classification contribution. A list of the data classes that contribute to the infoaset. The information includes the number and the overall size of the objects that matched each data class and whether all or part of the objects in the infoaset matched. For details, see “Classification of infoaset assets” on page 2.
Tool ID	The infoaset's internal ID in IBM StoredIQ.
URL	The URL to the infoaset's ancestry in the IBM StoredIQ Data Workbench.
Volumes	Volume contribution. A list of the volumes on which the data in the infoaset can be found. The information includes the number of objects and the size of the data that each volume adds to the infoaset. This information is available only for volumes that are synchronized.
Originating operation	A link to the details of the operation that was used to create the infoaset, such as the filter that was used or the infoasets to which the operation was applied.
Applied operations	A list of all operations that were applied to the infoaset to create other infoasets, with the respective links to the operation details.

In the governance catalog, you can add the following information:

- A short description of up to 255 characters
- Labels
- Stewards
- Information about rules usage
- An alias (business name)

This information is not made available in IBM StoredIQ.

In addition, you can apply custom attributes to the asset. This information and any assigned terms are made available as asset attributes in IBM StoredIQ.

You can access the catalog information from the infoaset details in the IBM StoredIQ Data Workbench user interface:

- The **Infoaset in Catalog** field provides a link for accessing the volume asset in the governance catalog. When you follow this link, you must sign in to Information Governance Catalog or IBM Cloud Private for Data, which requires an account with appropriate security roles.
- The **Asset Attributes** field provides a link for viewing synchronized terms and custom attributes in IBM StoredIQ if any were assigned to the asset in the governance catalog.

Filter assets

Each asset of the type Filter represents a IBM StoredIQ filter, either a filter you created in IBM StoredIQ or a system generated filter that is associated with a data class that was synchronized from the governance catalog.

Synchronization with the governance catalog

When data classes that qualify for synchronization are propagated to IBM StoredIQ, a filter for each data class is created automatically in IBM StoredIQ. No user action is required to create such filters. These system generated filters are by default also published to the governance catalog. The names of such filters follow the pattern **Contains governance data class *igc-data-class***. In case of naming conflicts when these system generated filters are added to IBM StoredIQ, existing filters are renamed by appending (user defined).

If you delete such a system generated filter from IBM StoredIQ, the filter is not re-created in later synchronization runs although the underlying data class might still be included and updated in the **GovernanceDataClasses** cartridge. The matching catalog asset is also removed.

If the data class for which a filter was generated is no longer synchronized, that filter is marked in IBM StoredIQ as being deprecated. It is not deleted so that you can still use it on previously generated analysis results in the index. The filter might also be used in active infosets.

Synchronization in general

Filters that are not marked for publishing might still be synchronized to the governance catalog. This is the case for filters that are in use by one or more infosets that are published to the governance catalog and is required to reflect the contributing operations properly. You can edit such an asset's details like any other's.

In the IBM StoredIQ Data Workbench user interface, the value **Yes (Inferred)** for any **Publish to Catalog** entry indicates such objects.

The following table lists the information available in the asset catalog after an IBM StoredIQ object is published to the governance catalog.

Property	Notes
Context	A link to the details of the instance with which the asset is associated.
Long description	This information originates in IBM StoredIQ. Therefore, ensure that the object's description there is as detailed as possible to maximize the value for a governance user. Tip: For a system generated filter, the description in IBM StoredIQ initially contains just the filter expression. You might want to add more details that are then included in the long description in the governance catalog on synchronization.

Property	Notes
Assigned to terms	<p>Any terms that might be assigned automatically during the synchronization from IBM StoredIQ and their respective category hierarchies. This term assignment is class based. In addition, you can directly assign terms to the asset, which are then made available in IBM StoredIQ as asset attributes.</p> <p>If a term that is inherited from a data class is removed from an asset in the governance catalog, this term is not automatically reassigned. If such a term is removed from the related data class, it is also removed from the asset.</p>
Filter expression	<p>For system generated filters, the indexed annotation with which the data is tagged. The annotation usually matches the data class name where spaces or special characters are replaced with an underscore (_).</p> <p>For any other filter, the filter code as defined in IBM StoredIQ.</p>
Data classification	The assigned data class.
Creator	The name and the email address of the user who originally created the filter. For system generated filters, the value is System .
Tool ID	The filter's internal ID in IBM StoredIQ.
Infosets	A list of infosets that were created by using this filter.

In the governance catalog, you can add the following information:

- A short description of up to 255 characters
- Labels
- Stewards
- Information about rules usage

This information is not made available in IBM StoredIQ.

In addition, you can apply custom attributes to the asset and you can add a data classification. The custom attributes, the classification information, and any assigned terms are made available in IBM StoredIQ.

You can access this catalog information from the filter details in the IBM StoredIQ Data Workbench user interface:

- The **Filter in Catalog** field provides a link for accessing the filter asset in the governance catalog. When you follow this link, you must sign in to Information Governance Catalog or IBM Cloud Private for Data, which requires an account with appropriate security roles.
- The **Data Class in Catalog** field provides a link for accessing the respective data class in the governance catalog. When you follow this link, you must sign in to Information Governance Catalog or IBM Cloud Private for Data, which requires an account with appropriate security roles.
- The **Asset Attributes** field provides a link for viewing synchronized terms and custom attributes in IBM StoredIQ if any were assigned to the asset.

The GovernanceDataClasses cartridge

The **GovernanceDataClasses** cartridge is a system generated cartridge.

It is automatically created, when data is synchronized for the first time between IBM StoredIQ and the governance catalog if IBM StoredIQ is configured accordingly.

The **GovernanceDataClasses** cartridge bundles regexes that are derived from data classes in the governance catalog. A data class is synchronized to IBM StoredIQ if it is applicable for unstructured data, is enabled, and its regex passes validation. The following regex definitions, although valid, are not accepted because their results would not be very useful:

```

.
*" .*"
"^ .*"
".*$"
"^ .*$"
"\\b.*\\b"
"\\s.*\\s"
"\\w*"
"\\D*"

```

Empty regexes are also not accepted. For each otherwise qualifying data class that contains any of these regex definitions, an error message is logged.

In addition to the cartridge, a corresponding Step-up Analytics action (**Governance Analytics**) to apply those regexes and filters for analysis results from the **GovernanceDataClasses** cartridge are also automatically created.

The **GovernanceDataClasses** cartridge is updated whenever relevant information in the governance catalog changes:

- Data classes are added.

This usually means new supported results. Rerun the **Governance Analytics** action on any info set of interest to ensure that the content is annotated with the new terms. You can then use the new filters that were created for the additional data classes to create new info sets.

- Data classes are updated.

Such updates include regex or class code changes.

To reclassify your content and to ensure that analysis results are up-to-date, you can rerun the **Governance Analytics** action on all info sets to which you applied those data classes so far. Due to the data class changes, analysis results might change. Therefore, create new info sets that represent the latest analysis results by reapplying the filters that correspond to the modified data classes.

The following changes to a data class are treated the same way as the creation or deletion of a data class:

Data class property	From	To	Handled as creation or deletion
Type	Any type but Regex	Regex	Creation
	Regex	Any other type	Deletion
Applicable for	Structured Data Only	Unstructured Data Only All Data	Creation
	Unstructured Data Only All Data	Structured Data Only	Deletion
Enabled	False	True	Creation
	True	False	Deletion

- Data classes are deleted.

To reflect those changes, you can run the **Governance Analytics** action on all infosets to which you applied those data classes so far to ensure consistent results. However, as harvests always use the newest cartridges, infosets will automatically be cleaned up over time even if you do not rerun the **Governance Analytics** action.

As a data expert, you receive email notifications for these changes. Use your best judgment to decide whether you must rerun the **Governance Analytics** action.

The Governance Analytics action

The **Governance Analytics** action is a system generated action.

It is automatically created, together with the **GovernanceDataClasses** cartridge, when data is synchronized for the first time between IBM StoredIQ and the governance catalog.

Run the **Governance Analytics** action to apply the analytics in the **GovernanceDataClasses** cartridge to an infoset. The action identifies and tags all files in the infoset that contain data that matches the regexes in the cartridge with the appropriate `ia:term` indexed annotation term. Thus, you classify data, and make the classified data searchable and detectable in IBM StoredIQ.

After you run the action, you can search on the indexed annotation terms to find data of a specific data class, either by building and applying your own full-text filter or by using the filters that are generated to reflect the synchronized data classes.

If the **GovernanceDataClasses** cartridge is updated in subsequent synchronization runs, you might need to run the **Governance Analytics** action again to apply the new logic and to reclassify your data.

Synchronization API

Use this API to request immediate synchronization of objects to the governance catalog.

Before you can invoke the synchronization API, you must enable the APIs and then authenticate to IBM StoredIQ by invoking the logon API. On successful logon, this API returns a session identifier cookie that is used for authentication in each subsequent API call.

Note: The commands in the following sections might be wrapped for readability.

Enable the APIs

Log in to the AppStack as `siqadmin` user and enable the APIs with this command:

```
storediqapi -enable
```

This command must be run on the AppStack while the API calls can be issued on any client that can reach the AppStack via network.

CURL command syntax to invoke the logon API

```
curl -k -I -L https://appstack-host/storediq/1.0/logon
-H "X-SPIRE-CREDENTIAL-TENANT-ID: appstack"
-H "X-SPIRE-CREDENTIAL-TYPE: password"
-H "X-SPIRE-CREDENTIAL-USERNAME: username"
-H "X-SPIRE-CREDENTIAL-PASSWORD: userpassword"
-H "Referer: https://appstack-host/storediq/1.0/logon"
-H "X-Requested-With: XMLHttpRequest"
-c cookies-file
```

Replace the variables as follows:

appstack-host

The host name or IP address of the AppStack server

username

A valid IBM StoredIQ user name

userpassword

The password of the user

cookies-file

The name of the file where the session ID cookie of the logon request is to be stored. This file is created automatically. This file name needs to be provided in all subsequent API calls.

CURL command syntax to invoke the syncnow API

```
curl -i -k -L -X POST https://appstack-host/storediq/1.0/govsyncnow
-H "Referer: https://appstack-host/storediq/1.0/govsyncnow"
-H "X-Requested-With: XMLHttpRequest"
-b cookies-file
```

Replace the variables as follows:

appstack-host

The host name or IP address of the AppStack server

cookies-file

The name of the file where the session ID cookie is stored.

Examples

The following examples show how to invoke the APIs and the responses.

Logon request

```
curl -k -I -L https://appstack/storediq/1.0/logon
-H "X-SPIRE-CREDENTIAL-TENANT-ID: appstack"
-H "X-SPIRE-CREDENTIAL-TYPE: password"
-H "X-SPIRE-CREDENTIAL-USERNAME: user1"
-H "X-SPIRE-CREDENTIAL-PASSWORD: Passw0rd!"
-H "Referer: https://appstack/storediq/1.0/logon"
-H "X-Requested-With: XMLHttpRequest"
-c cookies
```

Logon response

```
HTTP/1.1 200 OK
Server: nginx
Date: Wed, 16 May 2018 20:04:47 GMT
Content-Length: 0
Connection: close
Set-Cookie: sessionid=bf3243e2b77f4d8be424691574fccc082432e530; Path=/; secure; HttpOnly
X-Bastion-Subject-Id: 79dabf39-f76b-4d5d-b6cb-5362c7c81081
Cache-Control: must-revalidate, no-cache
X-Frame-Options: SAMEORIGIN
Content-Security-Policy: default-src 'self'
```

Synchronization request

```
curl -i -k -L -X POST https://appstack/storediq/1.0/govsyncnow
-H "Referer: https://appstack/storediq/1.0/govsyncnow"
-H "X-Requested-With: XMLHttpRequest"
-b cookies
```

Synchronization response

```
HTTP/1.1 200 OK
Server: nginx
Date: Wed, 16 May 2018 20:08:44 GMT
Content-Type: application/json
Content-Length: 43
Connection: close
Cache-Control: must-revalidate, no-cache
X-Frame-Options: SAMEORIGIN
Content-Security-Policy: default-src 'self'

{"status": "started: 2018-05-16T20:08:44Z"}
```

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

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 jurisdictions 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 Director of Licensing
IBM Corporation
North Castle Drive, MD-NC119
Armonk, NY 10504-1785
US

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.

The performance data discussed herein is presented as derived under specific operating conditions. Actual results may vary.

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.

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. _enter the year or years_.

Trademarks

IBM, the IBM logo, and [ibm.com](http://www.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" <http://www.ibm.com/legal/copytrade.shtml>.

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

Microsoft and Windows 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.

VMware, VMware vCenter Server, and VMware vSphere are registered trademarks or trademarks of VMware, Inc. or its subsidiaries in the United States and/or other jurisdictions.

The registered trademark Linux® is used pursuant to a sublicense from the Linux Foundation, the exclusive licensee of Linus Torvalds, owner of the mark on a worldwide basis.

Red Hat and OpenShift are trademarks or registered trademarks of Red Hat, Inc. or its subsidiaries in the United States and other countries.

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.

IBM Online Privacy Statement

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.

This Software Offering does not use cookies or other technologies to collect personally identifiable information.

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>.

Index

A

action
 Governance Analytics [13](#), [14](#)
 system generated [13](#), [14](#)
API
 log in [15](#)
 synchronization [15](#)

synchronization (*continued*)

 immediate [15](#)
 infoset [8](#)
 instance [6](#)
 monitoring [2](#)
 StoredIQ objects [5](#)
 volume [7](#)

C

cartridge
 GovernanceDataClasses [13](#), [14](#)
 system generated [13](#), [14](#)
classification [2](#)

D

detection logic
 regex [2](#)

L

legal
 notices [17](#)

M

monitoring
 API [2](#)
 component status [2](#)
 health [2](#)
 synchronization [2](#)

N

notices
 legal [17](#)

R

regex [2](#)
regular expressions [2](#)

S

Step-up Analytics action
 Governance Analytics [14](#)
 system generated [14](#)
synchronization
 configuring objects [3](#)
 enabling [3](#)
 filter [11](#)
 force [15](#)
 governance catalog objects [5](#)

