IBM Customer

IBM Content Analytics Install & Fast Start Service Description

January 1, 2011

Table of Contents

1.   Description of Services

Enterprise Content Management (ECM) Software Services is pleased to submit this Service Description to IBM Customer (<IBM Cust>). <IBM Cust> has requested the following services from IBM:

> ➢   IBM Content Analytics (Content Analytics) Install & Fast Start Package[1]

This service provides the expert resources to install the Content Analytics software at your site. It also provides hands-on assistance to address questions, recap key issues and to ensure that <IBM Cust>'s systems and resources are ready to begin applying the knowledge gained to <IBM Cust>'s business needs and operational requirements.

The Delivery Consultants will help guide <IBM Cust> on proven strategies and procedures that have been employed by other IBM Content Analytics installations. The completed installation also ensures that the software is supportable by IBM's Support organization.

The scope of this Install and Fast Start Package is limited to the installation and support of Content Analytics[2].

This Content Analytics Install & Fast Start Service will provide the following activities:

- Review <IBM Cust>'s environment

- Perform Document Analysis and Requirements Gathering
    - Create Data Source Configuration Document
    - Create Content Analytics System Design Document

- Install Content Analytics components:
    - Content Analytics Core Components
        - Parser
        - Crawler
        - Indexer
        - Document Processors
        - Annotators
    - Content Analytics Client Components
        - Text Miner Application
        - Search Application

- Install IBM Content Integrator (bundled with Content Analytics product)

- Install Content Analytics Agent Server

- Install Jetty Web Server

- Configure Crawler for a Simple Index

---

[1] This package may be used to implement either the IBM Content Analytics product or IBM Cognos Content Analytics product.
[2] This service includes the installation of Content Analytics in a "Master" (all on one server) architecture. Installation and configuration of any additional services other than those defined as part of this Install & Fast Start Package are considered additional options. Contact your Software Services Sales Leader for options.

- Configure for up to two Data Sources[3]

- Configure for up to two Collections, no more than 100,000 documents

- Index Content[4]

- Customize the content analytics system

  - Custom user dictionaries for one text analytics collection

  - Custom rules files for one text analytics collection

- Configure User Interface

  - Configure the facet tree for one text analytics collection

  - Configure Search Results list layout

  - Configure Document Ranking

- Tune Analytics Model

- Deliver System Administration Handoff

- Demonstrate the System to the Project Team

- Deliver Data Source Configuration Document

- Deliver Content Analytics System Design Document

- Project management and coordination to facilitate a professional project delivery

This Install and Fast Start Package includes a set of tasks that <IBM Cust> must complete to facilitate product installation and rollout. The actual duration of each task varies from customer to customer, based upon their knowledge of the products and the complexity of the anticipated Content Analytics environment.

The resources required for this project will include both the IBM and <IBM Cust> personnel to install and configure the Content Analytics software.

At the completion of the Content Analytics Install & Fast Start Package, <IBM Cust>'s system and resources will be ready for maintenance support and for initiating the process to configure the system for <IBM Cust>'s specific business requirements. This installation assumes individuals knowledgeable on IBM products will be performing the tasks to identify and expand the <IBM Cust> specific definition of the system after the completion of this project. Additional definition services beyond what is documented in this Service Description are outside the scope of deliverables of this Install & Fast Start Package[5].

Any change in the scope of work defined in this Services Description requested by <IBM Cust> may result in additional costs and must be documented in a Project Change Request (PCR).

2. Resources

The <IBM Cust> Project Team will consist of the resources from the ECM Account Team, the <IBM Cust> technical staff and, when applicable, an IBM Business Partner. Examples of Project Team members are listed below.

---

[3] The supported data sources for this implementation are listed in 7.5.1 Supported Data Sources.
[4] For demonstration purposes and to ensure the completion of indexing, the Delivery Consultant will select a subset of content which contains up to 100,000 documents of no more than 10 MG each.
[5] Services to configure additional site and user specific preferences can be accommodated through a Statement of Work. Contact your Software Services Sales Leader for options.

2.1   ECM Team Resources

The ECM Project Lead coordinates the appropriate ECM resources as required. The ECM Team typically includes the following members:

- Project Manager (PM) – Project Lead
- IT Consulting Specialist (Delivery Consultant)
- Software Sales Representative (SSR)
- Sales IT Specialist (ITS)

2.2   <IBM Cust> Team Resources

<IBM Cust> will need to identify internal personnel to fill the roles of the <IBM Cust> Project Team. <IBM Cust> is responsible for selecting a Project Lead to share lead responsibilities with the ECM Project Lead.

Depending upon, the project the <IBM Cust> Project Team may include any of the following roles:

- Project Lead
- IBM Content Analytics System Administrator
- Data Source Administrator
- IT Security Administrator
- User Interface Designer and/or Developer
- Infrastructure Specialist
- Business Lead Requirements Owner
- Text Analysis Administrator

3.   Role Definitions

<IBM Cust> will need to identify internal personnel to fill the roles as listed in Section 3, Resources. The following outlines the general roles for each of these individuals. Please note that additional resources may be required at various phases of the installation. Likewise, not all of the individuals may necessarily be involved in this install service.

Project Lead - This individual will represent <IBM Cust> resources and provide project coordination in conjunction with the ECM Project Lead.

Content Analytics System Administrator (1 to 3 people) - This individual will learn the administration functions necessary to oversee Content Analytics and will serve as the ongoing administrator for the system. This individual will work closely with the Delivery Consultant during installation and configuration of the Content Analytics software.

Text Analytics Administrator (1 person) – This individual owns the principal text analytics artifact – the Text Analytics Engine (TAE). This is maintained in a deployable archive (.pear file). The TAE comprises the text analytics model for the <IBM Cust>, at the time of this service. Should the model need changing, from the business perspective, the Text Analytics Administrator will be responsible for 1) making those changes in the TAE, using the LanguageWare Resource Workbench and 2) ensuring that the newly modified artifact 9.pear) is deployable and 3) produces the desired results.

Data Source Administrator (1 person) - This individual will assume responsibility for <IBM Cust>'s data sources. This individual will also provide proper passwords to allow the Delivery Consultant to perform his or her functions in the course of the installation when access to the data sources is required.

4. Responsibilities

4.1 <IBM Cust> Responsibilities

In order to facilitate delivery, <IBM Cust> has the responsibility to provide the following:

a.) Project Prerequisites

- All prerequisites must be met before IBM can begin the installation of the product.
- All hardware, software, and network configurations must meet Content Analytics requirements for system size and performance for the purchased software. This includes:
    - ➢ Installation of the operating system
        - Must comply with the Content Analytics operating system requirements for the appropriate Platform (Microsoft Windows, AIX, or Linux)
    - ➢ Installation and/or application of appropriate levels of patches to meet Content Analytics's minimum requirements
- All prerequisite hardware and software has been installed onto <IBM Cust>'s designated servers.
- Perform a network check (verify server connectivity to network, addressing, correct protocol).
- Download the IBM Content Analytics software to the appropriate servers prior to the resource arriving at <IBM Cust>'s facility.
- Complete the *Environment Prerequisite Checklist.*
- Attend the required training prior to the installation of any IBM software.

b.) <IBM Cust> acknowledges that:

i. Data Sources to be crawled must be installed, configured, and populated with data.

ii. This Content Analytics implementation is designated for single language use only.[6]

iii. Installation of Content Analytics software will not be in a highly-available manner (clustering and/or farming). A chargeable PCR may be required to perform the installation in a highly-available configuration.

iv. Additional resources and time for the development of "custom applications" are not included in this service.

c.) A Project Lead with sufficient authority to:

- Represent <IBM Cust>'s interests
- Serve as the focal point for all decisions and communications
- Coordinate activities assigned to <IBM Cust> resources
- Sign project-related documents

d.) Analysis and/or requirements gathering of business application rules, business application processes, application security access, text analysis scenarios, and sample documents as they pertain to the implementation of Content Analytics.

e.) Local system administrator(s) trained on the administration of the IBM products.

---

[6] Services to configure other languages can be accommodated by creating an additional Statement of Work, contact your Software Services Sales Leader for options.

f.) After the <IBM Cust> System Administrator has observed the Delivery Consultant during the installation, he or she will be responsible for completing the following items:

- Defining and Indexing additional Collections & Data Sources

- Defining and maintaining crawler & indexing scheduling

- Defining and maintaining additional text analysis engines

- Defining and configuration additional facets

g.) Local help desk personnel trained on the IBM client products to be installed at <IBM Cust>.

h.) Adequate facilities for each Delivery Consultant including:

- cubicle, desk, telephone, and Internet access,

- remote access (e.g., VPN access),

- direct access to the keyboard.

i.) When performing remote work, <IBM Cust> will provide VPN access into <IBM Cust>'s network and remote access to a workstation in the network that:

- can connect to all required servers and

- is preloaded with all required connectivity tools appropriate to <IBM Cust>'s environment.

j.) Product Subscription and Support Agreement in place through IBM.

k.) <IBM Cust> will make available all <IBM Cust> personnel required for this project based on the agreed-upon project schedule. Any significant delays caused by <IBM Cust> personnel not being available or project prerequisites not being met may result in schedule delays, which can impact the project sufficiently to require a chargeable PCR addressing additional or extended IBM personnel requirements and additional costs.

4.2   IBM Responsibilities

IBM will provide the following activities:

a.) Project Management

b.) Review <IBM Cust>'s Environment

- Deliver & Review the Environment Prerequisite Checklist

- Review Content Analytics Architecture

c.) Gather Configuration Requirements

- Review of Content Analytics Functionality

- Create a Content Analytics Implementation:

  - Two collections, two Crawlers per Collection

  - Up to two Data Sources

  - One Text Analysis Engine

- Create Data Source Configuration Document

- Create Content Analytics System Design Document

d.) Install the following Content Analytics Components:

- One Parser Component per Collection

- Two Crawler Components

- Up to two Search Components
  - Search Application
  - Text Miner
- One Index Component per Collection
- One Text Analytics Engine
- IBM Content Integrator

e.) Configure Crawler(s)

- For up to two Data Sources
- Two Collections, two Crawlers, no more than 100,000 documents

f.) Index Content[7]

g.) Configure UI

- Configure the facet tree for one collection
- Search Results list layout
- Document Ranking

h.) Tune Analytics Model (16 hours time-boxed)

i.) Perform Content Analytics Administration Handoff

- Transfer of Information
- Routine Maintenance
- Backup Methodology
- Problem Reporting

j.) Demonstrate Product to the Project Team

k.) Deliver Final Data Source Configuration Document

l.) Deliver Final Content Analytics System Design Document

m.) Provide <IBM Cust> a completed Content Analytics Completion Checklist

5. Limitations

5.1 Service Limitations

The Content Analytics Install & Fast Start Package does not include repairs, corrections or troubleshooting of performance degradation, data corruption, errors and faults that are discovered during the process due to hardware failures, or non-qualified hardware or operating systems. Repairs, corrections or troubleshooting of such errors and faults will be <IBM Cust>'s responsibility; any IBM involvement will be charged on a Time-and-Materials (T&M) basis.

ECM resources will be provided Monday through Friday from 8:00 a.m. to 5:00 p.m., excluding holidays and vacation time. Specific details regarding schedules will be mutually agreed-upon prior to the project start-date.

---

[7] For demonstration purposes and to ensure the completion of indexing, the Delivery Consultant will select a subset of content which contains up to 100,000 documents of no more than 10 MG each.

5.1.1 Travel & Living Expenses

This package does **not** include travel and living expenses incurred by IBM resources traveling to <IBM Cust>'s site in order to fulfill IBM's tasks and responsibilities pertinent to delivering this service.

The Delivery Consultant may travel to <IBM Cust>'s site for four **(4)** visits for tasks associated in delivering this service.   <IBM Cust> will be billed actual travel and living costs.

IBM will invoice <IBM Cust> for travel and living expenses and other reasonable expenses incurred in connection with the package delivery. Amounts are due upon receipt of invoice and payable within 30 days. <IBM Cust> agrees to pay accordingly, including any late payment fee. <IBM Cust> does not have to pay any amount disputed in good faith while it is being investigated, but <IBM Cust> remains obligated to pay any amount not in dispute.

Additional on-site assistance requested by <IBM Cust> beyond the number of visits indicated above will be billable subject to the execution of a PCR. A visit is four or fewer days at your site by an IBM or IBM subcontract employee. IBM also reserves the right to bill <IBM Cust> for any additional trips that are the direct result of <IBM Cust>'s failure to meet or perform its obligations outlined herein.

6.   Project Tasks & Deliverables

The Project Team will perform the following tasks. Deliverables for each task are detailed within the task descriptions.

6.1   Task 1: Kickoff Meeting

The purpose of this meeting is to prepare both the <IBM Cust> team and the ECM team for the Content Analytics Install & Fast Start Package. Topics of discussion will include <IBM Cust>'s requirements, expectations, prerequisites, and schedules.

6.1.1 Review Project Services Description

The ECM Project Lead will review the package Service Description, the installation tasks, and roles and responsibilities with <IBM Cust>'s Project Team.

6.1.2 Define and Review Project Plan

Once selected, the joint <IBM Cust> and IBM Project Team customizes the sample project plan defining mutually agreed-upon dates for delivery (see Sample Project Plan). Each task is assigned a resource and a time window based on the effort required to complete the task. IBM strongly encourages the practice of joint task assignments as an excellent means of training and knowledge transfer for the team members.

6.1.3 Project Status

The PM will track all issues and document their resolution in an Issues Log. Any issue whose resolution requires a change will be documented in a PCR.

6.2   Task 2: Environment Review

The PM and the Delivery Consultant will review the IBM software purchased, <IBM Cust>'s server configuration (hardware, operating system and site specific software), IBM specific network configuration, and <IBM Cust>'s

System Diagram. This review will encompass the items indicated in the Environment Prerequisite Checklist that is delivered to <IBM Cust>. Upon completion of <IBM Cust> portion of this document, the PM will then schedule the software installation.

ECM recommends reviewing any questions arising from the completion of the *Environment Perquisite Checklist* during the weekly status calls.

6.3   Task 3: Document Analysis & Requirements Gathering

The Install & Fast Start Package is specially designed to ask the necessary questions and guide <IBM Cust> through the design and implementation process of a Content Analytics system. The Delivery Consultant will perform a data analysis in regards to the configuration of the Content Analytics system to match the defined needs of the business utilization. Tasks include defining Content Analytics collections data sources, indexing identified data, and performing text analysis on unstructured text using text analysis engine. The Delivery Consultant and <IBM Cust>'s Content Analytics System Administrator will initiate the Data Source Configuration Document (documents the connection details for up to two data sources to connect to Content Analytics) and the Content Analytics System Design Document (to outline how the Content Analytics system has been designed and configured) with the information gathered during this analysis[8]. This document will be used to communicate with <IBM Cust> personnel for additional information and clarification.

6.4   Task 4: Software Installation

6.4.1 Install Content Analytics Core Components

The Delivery Consultant, with the assistance of <IBM Cust>'s Content Analytics System Administrator, schedules the server software installation. <IBM Cust>'s Information Technology staff must review the release notes included on the software media, which provides the latest information regarding operating system software release levels and patches required for software installation.

Together the team will install the Parser, Crawler, Indexer runtime, and text analysis engine components on the designated servers.

6.4.2 Search Client Components

The Search Application and Text Miner Application will be installed on the Jetty Web Server.

6.4.3 Install IBM Content Integrator

Content Integrator provides bidirectional, real-time federated access to disparate content repositories. Content Analytics uses Content Integrator to enable the crawling of many of its supported content sources, such as Microsoft Exchange Server and FileNet Content Manager, for example.

6.5   Task 5: Configure Crawlers

With the software installed, the Delivery Consultant will configure the product for up to two selected data source(s) for a simple index using the Collections and Crawlers defined in the Data Source Configuration Document.   In order to test the configuration, the Delivery Consultant and <IBM Cust>'s Content Analytics System Administrator will index up to 100,000 documents of no more than 20MB each.

---

[8] The Delivery Consultant time required for performing the analysis and creating the Data Source Configuration Document should not exceed one day.

6.5.1 Data Sources

Select up to 2 Data Sources from the options[9] at the following web site (Content Analytics 2.1 only):

http://www-01.ibm.com/support/docview.wss?rs=4173&uid=swg27015094

6.6   Task 6: Create and Install Text Analysis Engine

6.6.1 Create Text Analysis Engine

The Delivery Consultant in conjunction with the Text Analysis Administrator will create a Text Analysis Engine (TAE), using IBM Text Analytics tooling. The TAE will model the business needs scenario outlined in the information gathering steps.

The output of the TAE creation will be a Processing Engine ARchive (PEAR). The Content Analytics Systems Administrator will be guided in the installation of this PEAR into the Content Analytics System, and its subsequent configuration as a TAE in the Parser of the Text analytics Collection.[10]

6.6.2 Install Text Analysis Engine

The Delivery Consultant, in conjunction with the Content Analytics System Administrator, will install the Text Analysis Engine (TAE) into the Content Analytics System, then configure the TAE to run in the Parser component of one Collection.

6.7   Task 7: Configure Search

6.7.1 Search Results Layout

There are numerous options as to the content which is displayed in the search results:

- any field in the structured document can be either displayed or suppressed;
- the content of the search item summary can also be controlled, as well as the summary length;
- the clickable link can be set to any text contained in the result.

Working with the <IBM Cust>, the Delivery Consultant will configure these options to the <IBM Cust>'s choices.

6.7.2 Document Ranking

Document ranking, i.e. "relevancy", in Content Analytics is configurable. For this package, static ranking will be applied, with the basic choices of ranking by number of links to the document, or ranking by date. The ranking can be extended, if <IBM Cust> chooses, with the following options:

1. Custom boost word dictionary: the Delivery Consultant will create, with feedback from the <IBM Cust>, a custom boost word dictionary, and add it to the Content Analytics configuration. This dictionary will be considered a sample only, with the intention of instructing the <IBM Cust>'s Content Analytics Administrator in the process of creating and maintaining boost word dictionaries.

---

[9] Implementing Data Source options other than those indicated in this Service are considered out of scope. Contact your Software Services Sales Leader for options.
[10] Effort to create Text Analysis Engine not to exceed 5 business days (8 hours per day).

2. Assign boost factors to URI patterns: the Delivery Consultant, with feed back from the <IBM Cust> will demonstrate how to assign boost factors to URI patterns of the <IBM Cust>'s choosing (up to five patterns will be configured with the assistance of the Delivery Consultant). This configuration will be considered a sample only, with the intention of instructing the <IBM Cust>'s Content Analytics Administrator in the process of assigning boost factors to URI patterns.

6.7.3 Facets

Faceted search will be configured in the text analytics collection.

In a text analytics collection, facets will be derived from documents in the following ways:

1. Fields and metadata in the source content will be associated with facets by annotators that add facet values to content as metadata or explicitly configure facets for the collection. Facets will be created from search fields that you assign to the facet or you can specify which fields are to be used as facets when you configure facets or crawler options.

2. Custom text analysis rules and dictionaries will be associated with keywords or facets. Dictionaries will also associate keywords with facets and synonyms will be defined for keywords. The rules and dictionaries are applied when unstructured content is added to a collection.

6.7.4 Test

Upon completion of the configuration, the Delivery Consultant will perform initial testing to validate that the system is operational. The Project Team members can use this phase as a learning process to ensure they understand the complete installation and testing process of the Content Analytics system.

Test the initial configuration validating:

- User Interface
- Search Results List Layout
- Response Retrieval
- Response Presentation
- Accuracy of Text Analytics

6.8 Task 8: System Administration Handoff

6.8.1 Transfer of Information

The Delivery Consultant will review the product with the <IBM Cust> Content Analytics System Administrator to help prepare <IBM Cust> for further expansion of the system. The Delivery Consultant will reinforce topics covered in the Learning Services courses required for this Install and Fast Start Package.

6.8.2 Routine Maintenance

Routine maintenance involves reviewing the daily duties of <IBM Cust>'s Content Analytics System Administrator with respect to maintenance of the system. These are tasks that should be performed daily and will be specific to <IBM Cust>'s usage of the system and its configuration. IBM recommends that the System Administrator develop a schedule of "housekeeping" procedures and tasks to be performed.

6.8.2.1 Text Analytics Modeling Maintenance

As the results of the initial modeling become apparent, adjustments may be needed in the model itself. For example, the model may need additional entities, dictionaries, or rules, in order to capture additional insights. In so far as modeling is an iterative process, <IBM Cust> will need to be capable of making modifications to the modeling code artifacts and re-deploying them.

6.8.3 Review Backup Methodologies

The Delivery Consultant will review possible backup strategies with the <IBM Cust> Content Analytics System Administrator. This Install and Fast Start Package does not provide for the Delivery Consultant to analyze <IBM Cust>'s backup objectives nor does it develop the plan to meet those objectives.

6.8.4 Problem Reporting

Before calling IBM Customer Support, full documentation of the problem at hand always expedites resolution and helps maintain internal system up-time records. The Delivery Consultant will review troubleshooting and problem reporting procedures and provide a trouble report template to <IBM Cust>. Over time the <IBM Cust> Content Analytics System Administrator will learn to determine the nature of problems that arise will be able to resolve or report them as necessary.

For integrity issues, IBM requests that <IBM Cust> contact IBM Customer Support by using the IBM toll free Customer Service telephone number 1-800-426-7378 or 1-800-IBM-SERV.

6.9 Task 9: Demonstrate System to <IBM Cust> Project Team

After the initial samples have been tested, the system is demonstrated to the <IBM Cust> Project Team. The demonstration is an opportunity to share ideas on how to best configure the system, gather immediate feedback on issues that may have been overlooked, and solicit suggestions for system improvements. Other uses for the system are often identified in such meetings, giving impetus for follow-on projects.

6.10 Task 10: Data Source System Configuration Document

As <IBM Cust> tests the system, the Delivery Consultant will provide <IBM Cust> the completed Data Source System Configuration Document. This document represents a starting point for ongoing maintenance and modification of the system.

6.11 Task 11: Content Analytics System Design Document

The purpose of the Content Analytics System Design Document is to provide a reference-able "topography" of the final, installed Content Analytics system. This document is to be used by the <IBM Cust> Content Analytics System Administrator 1) for management of the system and 2) to plan for future expansion.

The Content Analytics System Design Document will include:

- A visual diagram of the Master Crawler Server and the connected data sources including machine hostnames, IP addresses, and firewalls.

- A list of user ID's and passwords required for the system

- A list of the directories used in 1) installation 2) Content Analytics "runtime" and 3) data directory. Including

    o Complete paths and machine names, if necessary.

        o   Directory sizes and intended data corpus sizes.

- A list of the data sources – machine names, IP addresses, and user ID's and passwords. [11]

- A description of the Text Analysis components and their locations. This description should list the Annotations used, in a pseudo-code style, e.g. Named Entity: Person.

6.12 Task 12: Project Completion Checklist

Upon demonstration of the product the Delivery Consultant will deliver to <IBM Cust> an Install Completion Checklist which constitutes completion of this project. Please refer to the *IBM Content Analytics Install & Fast Start Completion Checklist*.

If issues arise after delivery of the Project Completion Checklist, <IBM Cust> should contact IBM Customer Support by using the IBM toll free Customer Service telephone number 1-800-426-7378 or 1-800-IBM-SERV.

---

[11] Note: in the interest of security, it is only necessary to indicate the user ID AND the name of a system administrator who can provide the password.