IBM StoredIQ Administrator
Version 7.6.0.13

Administration Guide

SC27-5688-15

IBM StoredIQ Administrator
Version 7.6.0.13

Administration Guide

SC27-5688-15

Note
Before using this information and the product it supports, read the information in Notices.

This edition applies to Version 7.6.0.13 of product number 5724M86 and to all subsequent releases and modifications until otherwise indicated in new editions.

Table of contents

Tables

About this publication

IBM® StoredIQ Administrator Administration Guide describes how to manage the IBM StoredIQ Administrator application to configure shares, create system infosets, manage users, target sets, and actions, and create reports.

IBM StoredIQ Platform product library

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

- *IBM StoredIQ Platform Deployment and Configuration Guide*, SC27-6386
- *IBM StoredIQ Platform Overview Guide*, GC27-6398
- *IBM StoredIQ Platform Data Server Administration Guide*, SC27-5692
- *IBM StoredIQ Administrator Administration Guide*, SC27-5688
- *IBM StoredIQ Data Workbench User Guide*, SC27-5691
- *IBM StoredIQ eDiscovery User Guide*, SC27-5693
- *IBM StoredIQ Policy Manager User Guide*, SC27-5694

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:

Support and assistance

- 1-866-227-2068

To e-mail IBM StoredIQ customer support, use this email address:

- storediqsupport@us.ibm.com

For information about how to contact IBM, see the Contact IBM web site at http://www.ibm.com/contact/us/

IBM Knowledge Center

The IBM StoredIQ publications can be found from IBM Knowledge Center.

PDF publications

The IBM Publication Center site offers customized search functions to help you find all the IBM publications you need.

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/.

How to send your comments

Your feedback is important in helping to provide the most accurate and highest quality information.

Send your comments by using the online reader comment form at https://www14.software.ibm.com/webapp/iwm/web/signup.do?lang=en_US&source=swg-rcf.

Chapter 1. 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.

An 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, actions, and Mule scripts. This list provides an overview as to how IBM StoredIQ Administrator works.

Managing Data Servers

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 index.
- 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 in to IBM StoredIQ Platform Data Server.

Creating System Infosets

Administrators can create and manage System infosets that use only specific indexed volumes 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 create reports and upload report packages, helping you 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 relate business concepts to indexed data.

Managing Mule Scripts
Helps you to create Mule scripts and upload script packages. These Mule scripts are used by IBM StoredIQ Policy Manager to create policies by using the Automation Workflow.

Chapter 2. Logging in to IBM StoredIQ Administrator

Follow these steps to log in to IBM StoredIQ Administrator.

Before you begin

The administrator must ensure that these tasks are completed before you can use IBM StoredIQ Administrator:

- Install and configure IBM StoredIQ Platform.
- Ensure that the data servers are up and running.

Procedure

1. Open the IBM StoredIQ Platform user interface from a browser and enter `superadmin` in the email address text field and `admin` in the password text field.
2. Click **Log in** to open IBM StoredIQ Administrator.

Chapter 3. 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, **Data Servers** and **Volumes** sub-navigation. When a data server is selected, the System Status and System Activity panes provide detailed information about the selected data server.

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.

Data Servers

The **Data Servers** pane lists detailed information about a data server.

Table 2. The Data Servers pane 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, which are **Healthy**, **Vulnerable**, **Critical**, **Unavailable**, **Under maintenance**, **Rebooting**, or **Restarting**.
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 zize	Lists the size of the data objects that are found on the data server.

The **System Status** and **System Activity** panes provide performance information on a selected data server. These panels appear only once a data server is selected. The ability to check the status of all deployed data servers is critical to understanding the system's performance metrics.

Table 3. **System Status** *and* **System Activity** *panes field names and descriptions*

System Status Field Name	Description
Status	Lists the data server's status, namely **Healthy**, **Vulnerable**, **Critical**, **Unavailable**, **Under maintenance**, **Rebooting**, or **Restarting**.

Table 3. **System Status** *and* **System Activity** *panes field names and descriptions (continued)*

System Status Field Name	Description
Status message	Describes the data server's status. Status messages include these messages: • System and services running. • System is running but some services are still coming up. • Some query processes are not running. • There was a problem with the RAID controller. • Some services are in error. • Cannot connect to system and error.
IP address	Lists the IP address of the data server.
Software version	Lists the installed software version.
DataServer type	Lists the installed index type: Classic (DataServer - Classic) and Distributed (DataServer - Distributed).
DB version	Lists the installed database version.
System time	Lists the system's time, date, and utilized time zone.
System Activity Field Name	**Description**
Free RAM 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.
Available space	Lists the available space on the selected data server.
Active DB connections	Lists the number of database connections currently open for the selected data server.
System uptime	Lists the total duration of time for which the data server ran.

Volumes

The pane provides the detailed information about the volumes.

Table 4. The Volumes pane 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.

Table 4. The Volumes pane column names and descriptions (continued)

Volumes Column Name	Description
Last harvested	Denotes the last time that the volume was harvested.

Select the volume and click **View Details**, a bottom window opens with the number of data objects, total data object size, harvest status, last harvest date and time, and the harvest type.

DataServer - Distributed

Data servers can be categorized in two types: DataServer - Distributed and DataServer - Classic or regular data servers. DataServer - Distributed is different from the regular data servers in the following aspects:

- It provides better performance in search queries.
- The index is stored in a distributed Elasticsearch cluster rather than the embedded PostgreSQL database.
- Each data server can manage much larger amount of data, thus making the IBM StoredIQ deployments more scalable.
- Without adding more data servers, data that is managed by the IBM StoredIQ deployment can be increased by adding new nodes to the Elasticsearch cluster.

DataServer - Distributed supports the following features:

- Full and incremental harvest.
- Creating, deleting infosets and creating infosets from data map.
- Metadata and full text search.
- Supporting these actions: copy, move, delete, and export.
- Covering these datasources: NFS, CIFS, Box, Exchange, Sharepoint, Documentum.
- Generating all reports except Attribute Summary Report, Duplication Summary Report, Term Hit Report, and Custom CSV Export.
- Using the same StoredIQ gateway to manage both DataServer - Classic and DataServer - Distributed.

If an infoset contains volumes from both DataServer - Classic and DataServer - Distributed, then the infoset is a mixed infoset. For operations running on mixed infosets that are not supported on DataServer - Distributed, they only run on DataServer - Classic.

To deploy the virtual appliance that contains pre-installed Elasticsearch and populate the gateway and data server, see Deploying the virtual appliances in the deploying and configuring IBM StoredIQ documentation.

DataServer - Distributed vs. DataServer - Classic

As a different type of the data server, DataServer - Distributed works differently from DataServer - Classic.

A data server can either use the current PostgreSQL or Lucene index as an index or use Elasticsearch as the data index, but not both. However, both types of data servers can be added to the IBM StoredIQ gateway.

If an infoset contains volumes from both DataServer - Classic and DataServer - Distributed, then the infoset is a mixed infoset. For operations running on mixed infosets that are not supported on DataServer - Distributed, they only run on DataServer - Classic. See “DataServer - Distributed” on page 7 for supported operations on DataServer - Distributed.

Upon a reharvest, metadata in an infoset that was first created is not updated; it remains the same. Metadata after reharvest is different, therefore, the search result can be different accordingly.

With DataServer - Distributed, regular expression is altered. Some search regular expression syntax is not supported for DataServer - Distributed:

- POSIX classes such as `\p{ALPHA} \p{ALNUM}, \p{CNTRL}` and so on.
- Complex operators are not supported:
  - Back reference such as
    - (abc)\1 which is short for 'abcabc'
    - (\d)\1 which is short for same number repeating twice like 00, 11, 22, 33, 44, 55, 66, 77, 88 or 99
  - Look-ahead or Look-behind such as
    - [0-9]{3}(?<!000) for matching three digits number except 000

For more information about regular expression Elasticsearch or Lucene support, see Regular expression syntax.

Searching for a specific data server

When you have multiple data servers available 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. Data servers that match the entered search term are returned.
2. Select a data server from the returned list of servers. A window opens with the data server's details.

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. Click **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 in the following tables.

   For Box volume prerequisites and configuration information, see “Box volumes: configuration note” on page 20.

Table 5. Box volumes: Add Volume dialog box options for primary volumes

Box: Add Volume dialog box options	Action	Notes
Volume Type	In the **Volume type** list, select **Primary**.	
Source Type	Select **Box**.	
Server	Enter the fully qualified domain name of the server from which the volume is available for mounting. For Box volumes, the server name `api.box.com` is automatically entered.	
Authenticate with Box	Before a Box volume can be added, the user must be authenticated. Click the **Authenticate with Box** link, sign into the Box account, and select **Grant access**.	
Assign to Data Server	Select the data server.	
Volume Name	Enter the common, user-defined name of this volume.	
Include Users	Select this option to scope the volume. Regular expressions are supported.	
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.

Table 6. CIFS volumes: Add Volume dialog box options for primary volumes

CIFS: Add Volume dialog box options	Action	Notes
Volume Type	In the **Volume type** list, select **Primary**.	
Source Type	Select **CIFS**.	
Server	Enter the fully qualified domain name of the server from which the volume is available for mounting.	
Username	Enter the user name that is used to connect to and mount the volume.	
Password	Enter the password that is used to connect to and mount the volume.	
Assign to Data Server	Select the data server.	

Table 6. CIFS volumes: Add Volume dialog box options for primary volumes (continued)

CIFS: Add Volume dialog box options	Action	Notes
Volume Name	Enter the common name of this volume.	
Share	Enter the share name of this volume.	
Initial Directory	Optionally, enter the name of the initial directory from which the harvest must begin.	
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.

Table 7. Connections volumes: Add Volume dialog box options for primary volumes

Connections: Add Volume dialog box options	Action	Notes
Volume Type	Select **Primary** in the Volume type list.	
Source Type	Select **Connections** in the Source type list.	
Server	Enter the fully qualified domain name of the server from which the volume is available for mounting.	
User name	Enter the user name of the account that is set up with admin and search-admin privileges on the Connections server.	
Password	Enter the password of the account that is set up with admin and search-admin privileges on the Connections server.	
Assign to Data Server	Select the data server.	
Volume Name	Enter a name for the volume.	
Initial Directory	Optionally, enter the name of the initial directory from which the harvest must begin.	
Class name	Enter `deepfile.fs.template.impl.ibmconnections.ibmconnectionsconn.IBMConnections`	Required

Table 7. Connections volumes: Add Volume dialog box options for primary volumes (continued)

Connections: Add Volume dialog box options	Action	Notes
Repository	Enter `deepfile.fs.template.impl.ibmconnections.ibmconnectionsconn`	Required
Option string	Enter more option parameters.	
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.

Table 8. CMIS volumes: Add Volume dialog box options for primary volumes

CMIS: Add Volume dialog box options	Action	Notes
Volume Type	In the **Volume type** list, select **Primary**.	
Source Type	Select **CMIS**.	
Server	Enter the fully qualified domain name of the server from which the volume is available for mounting.	
Port	Enter the port number.	
Username	Enter the user name that is used to connect to and mount the volume.	
Password	Enter the password that is used to connect to and mount the volume.	
Assign to Data Server	Select the data server.	
Volume Name	Enter the common name of this volume.	
Use SSL	Select the **Use SSL** check box.	
Service	Enter the service name.	
Repository	Enter the name of the repository.	
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.

To add Documentum volumes, you must add the Documentum server first. For instructions about adding a Documentum server, see Adding a Documentum server as a data source.

Table 9. Documentum volumes: Add Volume dialog box options for primary volumes

Documentum: Add Volume dialog box options	**Action**	**Notes**
Volume Type	In the **Volume Type** list, select **Primary**.	
Source Type	Select **Documentum**.	
Doc base	Enter the name that was entered on the data server from the doc broker settings.	
Username	Enter the user name that is used to connect to and mount the volume.	
Password	Enter the password that is used to connect to and mount the volume.	
Assign To Data Server	Select the data server from the list.	
Volume Name	Enter the common name of this volume.	
Harvest all document versions	If you need to harvest all document versions, select the check box.	
Initial Directory	Optionally, enter the name of the initial directory from which the harvest must begin.	
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.

Table 10. Exchange volumes: Add Volume dialog box options for primary volumes

Exchange: Add Volume dialog box options	**Action**	**Notes**
Volume Type	In the **Volume type** list, select **Primary**.	
Source Type	Select **Exchange**.	
Server	Enter the fully qualified domain name of the server from which the volume is available for mounting.	If **Exchange Online** is selected as the **Source Type**, the server name is automatically entered.
Username	Enter the user name that is used to connect to and mount the volume.	

Table 10. Exchange volumes: Add Volume dialog box options for primary volumes (continued)

Exchange: Add Volume dialog box options	Action	Notes
Password	Enter the password that is used to connect to and mount the volume.	
Assign to Data Server	Select the data server.	
Volume Name	Enter the common name of this volume.	
Server Version	Select the version of Microsoft Exchange, choosing from **2000/2003**, **2007**, **2010/2013/2016**, and **Online**.	
Mailbox Server	Enter the names of the mailbox servers, which are separated by commas.	If **Exchange Online** is selected as the **Server Version**, this option is not available.
Active Directory Server	Enter the name of the Active Directory server.	If **Exchange Online** is selected as the **Server Version**, this option is not available.
Use SSL	To use secure socket layer, select the **Use SSL** check box.	If **Exchange Online** is selected as the **Server Version**, this option is grayed out.
Initial Directory	Optionally, enter the name of the initial directory from which the harvest must begin.	
Virtual Root	The name defaults to the correct endpoint for the selected Exchange version.	
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.

Table 11. FileNet volumes: Add Volume dialog box options for primary volumes

FileNet: Add Volume dialog box options	Action	Notes
Volume Type	In the **Volume type** list, select **Primary**.	
Source Type	Select **FileNet**.	
Server	Enter the fully qualified domain name of the server from which the volume is available for mounting.	

Table 11. FileNet volumes: Add Volume dialog box options for primary volumes (continued)

FileNet: Add Volume dialog box options	Action	Notes
Port	Enter the port number.	
Username	Enter the user name that is used to connect to and mount the volume.	
Password	Enter the password that is used to connect to and mount the volume.	
Assign to Data Server	Select the data server.	
Volume Name	Enter the common name of this volume.	
Object Store	Enter the object store.	
Connection Type	Select either **HTTP** or **HTTPS**.	
Path	Enter the appropriate directory path.	
Stanza	Enter the appropriate stanza.	
Scope	Optionally, enter the appropriate SQL where clause.	
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.

Table 12. HDFS volumes: Add Volume dialog box options for primary volumes

HDFS: Add Volume dialog box options	Action	Notes
Volume Type	In the **Volume type** list, select **Primary**.	
Source Type	Select **HDFS**.	
Server	Enter the fully qualified host name of the server or the IP address from which the volume is available for mounting.	
Port	Enter the port number.	
Username	Enter the user name that is used to connect to and mount the volume.	
Password	Enter the password that is used to connect to and mount the volume.	Authentication to HDFS is not supported. If your HDFS server requires a password, StoredIQ is not able to connect to it.

Table 12. HDFS volumes: Add Volume dialog box options for primary volumes (continued)

HDFS: Add Volume dialog box options	Action	Notes
Assign to Data Server	Select the data server.	
Volume Name	Enter the common name of this volume.	
Use SSL	To use SSL, select the check box.	See **Option String** for more certificate options.
Initial Directory	Optionally, enter the name of the initial directory from which the harvest must begin.	
Repository	Enter the name of the repository.	
Option String (optional)	This option is supported: VerifyCertificate=True.	This option is used to indicate that the validity of the HDFS server's SSL certificate is verified when SSL is used. Values are True, False, or default value. If no value is specified, value is False. To validate the certificate on the HDFS server, the user needs to specify this option and set the value to True.
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.

Table 13. IBM Content Manager volumes: Add Volume dialog box options for primary volumes

IBM Content Manager: Add Volume dialog box options	Action	Notes
Volume Type	In the **Volume type** list, select **Primary**.	
Source Type	Select **IBM Content Manager**.	
Server	Enter the fully qualified host name of the library server database.	
Port	Enter the port that is used to access the library server database.	
Username	Enter the user name that is used to connect to and mount the volume.	

Table 13. IBM Content Manager volumes: Add Volume dialog box options for primary volumes (continued)

IBM Content Manager: Add Volume dialog box options	Action	Notes
Password	Enter the password that is used to connect to and mount the volume.	
Connection String	Optional: Enter connection-string parameters.	
Assign to Data Server	Select the data server.	
Volume Name	Enter the common name of this volume.	
Repository	Enter the name of the library server database.	
Server Type	Select the type of server that is associated with the volume. Options include **DB2** and **Oracle**. By default, **DB2** is selected.	
Schema	Enter the schema for this library server database.	
Remote Database	Enter the name of the remote database.	
Harvest Itemtype	Enter the name of the item types to be harvested, separated by commas.	Harvest type is required to harvest the CM8 volume.
Copy to Itemtype	The **Copy to Itemtype** text box can be changed to either `SiqDocument` or to be an empty field. If this field is left empty, the volume cannot be used for copy-to actions.	In this release, the attribute lengths are increased for some. For more information, see "IBM Content Manager attributes" on page 27.
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.

Table 14. Livelink volumes: Add Volume dialog box options for primary volumes

Livelink: Add Volume dialog box options	Action	Notes
Volume Type	In the **Volume type** list, select **Primary**.	
Source Type	Select **Livelink**.	
Server	Enter the fully qualified domain name of the server from which the volume is available for mounting.	

Table 14. Livelink volumes: Add Volume dialog box options for primary volumes (continued)

Livelink: Add Volume dialog box options	Action	Notes
Port	Enter the port number.	
Username	Enter the user name that is used to connect to and mount the volume.	
Password	Enter the password that is used to connect to and mount the volume.	
Assign to Data Server	Select the data server.	
Volume Name	Enter the common name of this volume.	
Database	Enter the database name.	
Search Slice	Enter the search slice.	
Initial Directory	Optionally, enter the search slice and name of the initial directory from which the harvest must begin.	
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.

Table 15. NFS volumes: Add Volume dialog box options for primary volumes

NFS: Add Volume dialog box options	Action	Notes
Volume Type	In the **Volume type** list, select **Primary**.	
Source Type	Select **NFS**.	
Server	Enter the fully qualified domain name of the server from which the volume is available for mounting.	
Assign to Data Server	Select the data server.	
Volume Name	Enter the common name of this volume.	
Export	Enter the export name for this volume.	
Initial Directory	Optionally, enter the name of the initial directory from which the harvest must begin.	

Table 15. NFS volumes: Add Volume dialog box options for primary volumes (continued)

NFS: Add Volume dialog box options	Action	Notes
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.

Table 16. NewsGator volumes: Add Volume dialog box options for primary volumes

NewsGator: Add Volume dialog box options	Action	Notes
Volume Type	In the **Volume type** list, select **Primary**.	
Source Type	Select **NewsGator**.	
Server	Enter the fully qualified domain name of the server from which the volume is available for mounting.	
Username	Enter the user name that is used to connect to and mount the volume.	
Password	Enter the password that is used to connect to and mount the volume.	
Assign to Data Server	Select the data server.	
Volume Name	Enter the common name of this volume.	
Initial Directory	Optionally, enter the name of the initial directory from which the harvest must begin.	
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.

Table 17. OneDrive volumes: Add Volume dialog box options for primary volumes

OneDrive: Add Volume dialog box options	Action	Notes
Volume Type	In the **Volume type** list, select **Primary**	
Source Type	Select **OneDrive**.	
Server	For OneDrive volumes, enter the server name.	

Table 17. OneDrive volumes: Add Volume dialog box options for primary volumes (continued)

OneDrive: Add Volume dialog box options	Action	Notes
Authenticate with OneDrive	Before a OneDrive volume can be added, the user must be authenticated. Click the **Authenticate with OneDrive** link, sign in as administrator.	
Assign to Data Server	Select the data server.	
Volume Name	Enter the common, user-defined name of this volume.	
Initial Directory	Select this option to specify the initial directory for OneDrive.	
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. Facets, personal drives, and notifications are not harvested.

Table 18. SharePoint volumes: Add Volume dialog box options for primary volumes

SharePoint: Add Volume dialog box options	Action	Notes
Volume Type	In the **Volume type** list, select **Primary**.	
Source Type	Select **SharePoint**.	
Server	Enter the fully qualified domain name of the server from which the volume is available for mounting.	
Username	Enter the user name that is used to connect to and mount the volume.	
Password	Enter the password that is used to connect to and mount the volume.	
Assign to Data Server	Select the data server.	
Volume Name	Enter the common name of this volume.	
Server Version	Select one of these servers: **2003**, **2007**, **2010**, **2013**, **2016**, and **Online**.	
Site URL	Enter the site URL of the SharePoint server.	
Recurse into subsites	Optional.	
Use SSL	Optional.	
Include all versions	Optional.	

Table 18. SharePoint volumes: Add Volume dialog box options for primary volumes (continued)

SharePoint: Add Volume dialog box options	Action	Notes
Initial Directory	Optionally, enter the name of the initial directory from which the harvest must begin.	
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. **Note:** For SharePoint Online, full-text indexing of OneNote notebook objects, that is, Notes, is not supported currently. FSMD-based searches for these files are supported.

3. Click **Save** to save your configurations and add the volume.
4. Click **View Volumes**. 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**.

Box volumes: configuration note

Box volumes can be added only from IBM StoredIQ Administrator and have several configuration prerequisites that must be met.

The prerequisites are as follows.

- Every application stack requires a dedicated Client ID. Each Client ID has its own Redirect URL that points to the IBM StoredIQ application stack. To create the application, log in to `developers.box.com`. Click **My Apps** > **Create Application** to get `client_id`, `client_secret`, and `redirect_uri`
- In the Box application, set `redirect_uri` to `https://x.x.x.x/proxy/enamel/1.0/oauthtokengenerator/box`, where `x.x.x.x` is the hostname/IP address of the application stack with which the user logs in.
- In the Box application, select the following **Scopes** options:
  - **Read and write all files and folders**
  - **Manage an enterprise**
  - **Manage an enterprise's managed users**
  - **Manage an enterprise's groups**
  - **Manage an enterprise's properties**
  - **Manage an enterprise's retention policies**
- Before Box volumes are added, the Client ID, Client Secret, and Redirect URI must be stored in an application stack `oauth.conf` file. This file can be found in the application stack directory in this location: `/etc/siq/oauth.conf`. After the `oauth.config` file is modified, restart the `uwsgi` service by logging in to the application stack with PuTTY and running this command: `/siq/bin/monit restart uwsgi`
- Each Client ID must have the **As-User** capability that is enabled to allow harvests and data discovery.

Note: Customers must contact their Box support representative to enable this permission. They also need their Client ID.

- The application stack must be configured for SSL access, and it must have access to the Box authentication endpoint on `api.box.com`. If the application stack was deployed without SSL enabled, run `/siq/bin/certcfg` and follow the instructions on the Certificate Configuration screen.
- A data server that manages a Box volume must have access to `api.box.com`.
- The **Application Settings** for the Box Enterprise Account must not disable unpublished applications from accessing the account. Clear the **Unpublished Applications** check box when you configure **Application Settings**.

When you copy to Box, by default the owner of a source content is mapped to the Box user only if the corresponding user is administered by the Box administration account. This way the contents can be copied to the mapped user account in Box. Additionally, the folder hierarchy of the source volume can be reproduced, but it is rooted in the home folder of the user.

When the user configures a copy action with Box as the target set from IBM StoredIQ Administrator, the user can optionally enter a destination directory. The user also can harvest the copied files by selecting or deselecting **Do not auto-harvest destination volume after copy**.

Two other options are available for the copy action, which has Box as target data source.

Mapping
: **Map permissions** from source to destination maps access permissions from the source to Box folder collaborators. Mapping preserves owners from source to destination volume.

Preserve version chains on destination
: Creates version chains in Box during the copy action.

Both options are not enabled by default and they can be modified based on your needs. If you clear **Preserve version chains on destination**, version chains are not created in Box, but all versions are copied as separate files.

If an infoset that is copied to Box contains an archive file along with members, then IBM StoredIQ copies only the archive file. The members of the archive file are skipped and audited in Policy Audits.

See the following retained Box metadata attributes to copy to Box volumes.

File name
: Same as source.

Description
: Same as source.

Owner
: Same as source based on mapping heuristics that apply only to CIFS and SharePoint. Box administrator's user name in all other cases.

  For CIFS, if single-sign-on (SSO) that is based on Active Directory is configured and the StoredIQ Data Server can locate the Active Directory Domain Controller by using DNS, StoredIQ uses the `E-mail` attribute from the Active Directory user profile of the owner of a source data object and tries to match it with the email address of a user in Box. If the owner of a

data object does not have an email address in Active Directory Server, or if owners of data objects that are not defined in Active Directory Server cannot be mapped, owner mapping fails.

If the StoredIQ Data Server cannot locate the Active Directory Domain Controller, StoredIQ uses the owner name itself and tries to map it to a Box User ID. Owner names in CIFS are typically in the form of `domain\username`, whereas email addresses in Box are in the form of `localpart@domainname`. In this particular case, StoredIQ maps only `username` to `localpart` of an email address in Box in a case insensitive manner.

For SharePoint, the mail address property of the source content owner must be matched with the Box User ID, where Box User ID must be managed by the Box administrator account and Box User ID is the mail address of the user. If the mail address is not available, the display name of the source content owner must be matched exactly with the display name of the Box user. If StoredIQ cannot find a display name for the SharePoint user, then the *username* part of a SharePoint login name is mapped to the *localpart* of an email address in Box. For example, *username* in `domain\`*`username`* is mapped to *localpart* in *`localpart`*`@domainname`.

No user mapping takes place if the conditions that are described are not met. Then, a content is copied to the administrator account of Box.

Size Same as source.

Created
Same as source.

Modified
Same as source.

Note: Box volumes can be added only from IBM StoredIQ Administrator, not from IBM StoredIQ Data Server.

Copy From Box is supported. IBM StoredIQ supports copying files from Box to CIFS, NFS, FileNet, and Box as target sets.

- Box Notes and Bookmarks are skipped during a Copy From Box. Skipped information is recorded in Policy Audits.
- If multiple versions exist, all versions are copied to the target. If the target data source does not support versioning, then the copied files have version numbers that are appended to the file names.
- Box metadata, except for **`Created At`** and **`Modified At`**, is not copied when files are copied from Box.
- Source file owner name is not mapped in the destination volume. Instead, all of the copied files have an `owner_name`: user name of the target volume.

IBM StoredIQ supports copying to Box from CIFS, NFS, and SharePoint source volumes. However, only copying documents and files to Box is supported. Copying social or collaboration content types such as Wiki pages, blog posts are not supported. IBM StoredIQ also supports Discovery Exports from Box. Box Notes and Bookmarks are exported as MHTML files. An export for a Box Note carries the textual content of the Box Note. An export of the Box Bookmark has no textual content but only metadata.

Each StoredIQ data server contains a CSV mapping file, which includes a set of CIFS and SharePoint standard properties that are mapped to the custom Box

properties. The mapping file is on the data server at `/deepfs/data/mapping_file/box_mapping.csv`. The mapping file can be edited as required but for the changes to take effect, services need to be restarted on the StoredIQ data server.

StoredIQ can index a list of collaborators who work with a document in Box. If an ancestor folder of a document is shared for collaboration, then each collaborator is indexed by StoredIQ with information about the User ID, email address, name, and role of the user. Incremental harvests in StoredIQ currently do not pick up changes to collaborators for a document. A full harvest might be required to get the index up-to-date concerning collaborators for documents.

Box notes cannot be viewed through the Data Object Viewer in IBM StoredIQ Data Workbench.

When an external user who is not native to the organization collaborates a folder with users who are managed by the administrator, the folder is not indexed if the Box volume was added with the credentials of the administrator. This kind of collaborated folder can be harvested only if the Box volume is added with the credentials of the managed users who were invited to collaborate on the folder by the external user.

When files or folders are deleted in Box, they are moved to the Trash folder in Box. Currently, contents in Trash are not indexed by StoredIQ.

Mapping of access permissions to Box folder collaborators

When you copy data into Box, besides the ownership of individual files, access permissions might also need to be retained based on the access control settings on the source volumes. The **Mapping: Map permissions from source to destination** option in the **Copy Action** is designed to support this use-case. IBM StoredIQ applies its proprietary mapping heuristic for mapping access control from CIFS and SharePoint volumes to Box folder collaborators.

Box supports only collaboration at the folder-level. Collaborators cannot be defined for individual files. The IBM StoredIQ proprietary mapping heuristic for mapping access control from CIFS and SharePoint volumes to Box folder collaborators works as follows.

Note: Currently, Box ACL is supported on the SharePoint 2013 server.

Individual file permissions within a source folder are mapped to Box collaborators that can be applied to the corresponding target folder. The folder collaborators on the target folder are an aggregate of the permissions for all files within the source folder. The permissions are aggregated in such a way that no user accidentally has permissions to a file on Box that they did not have access to on the source. However, this aggregation might, in certain cases, force some users to lose access to documents on Box that they were able to access on the source. For example, if User A had access to one file in the source folder but not to another file within the same folder, the aggregation forces User A to lose access to both files in Box. It happens because Box collaborators can be applied only at the folder-level.

Permissions are mapped from CIFS and SharePoint to Box Collaboration Roles by using the following mapping tables:

Table 19. Mapping permission from CIFS to Box Collaboration Roles

CIFS	BOX
Full Control	Co-owner, Owner
Modify	Editor
Read & Execute	Viewer
List Folder Contents	Viewer
Read	Viewer
Write	Uploader

Table 20. Mapping permission from SharePoint to Box Collaboration Roles

SharePoint	BOX
Full Control, Design	Co-owner, Owner
Edit, Contribute	Editor
Read	Viewer
View Only	Pre-Viewer

The **Mapping: Map permissions from source to destination** option works in tandem with the **Mapping: Preserve owners from source to destination** option:

- If only **Map permissions from source to destination** is selected, all files and folders are copied to the Box Administrator account and access control is mapped from the source volume to Box folder collaborators.
- If both **Map permissions from source to destination** and **Preserve owners from source to destination** are selected, then each file or folder is copied to Box user accounts that map to the source file owners and access control is mapped from the source volume to Box folder collaborators.
- If only **Preserve owners from source to destination** is selected, then each file or folder is copied to Box user accounts that map to the source file owners, but no folder collaborators are added on Box.
- If neither of these options are selected, the files and folders are all copied to the Box Administrator account and no folder collaborators are added on Box.

As an example, if a Public File Share is being copied to Box, **Map permissions from source to destination** can be selected while **Preserve owners from source to destination** can be deselected so that all the contents land in the Box Administrator account, while individual users still continue to have collaboration roles for this content. Similarly, if a particular User Share is being copied to Box, **Map permissions from source to destination** can be deselected while **Preserve owners from source to destination** can be selected so that all the contents land in the mapped Box user's account.

Before you run the actual copy action, it is a good practice to run a simulate action to preview the mappings and determine whether any access control changes are necessary before you run the copy action.

Note: Any files directly within a source file share (not contained within other folders), do not have collaboration enabled in Box since they do not have a parent folder to collaborate on. Setting a **Destination Directory** for the Copy Action ensures that these files have a parent folder and collaboration can be enabled in Box. Any files directly within a source file share (not contained within other folders), do not have collaboration enabled in Box since they do not have a parent

folder to collaborate on. Setting a **Destination Directory** for the Copy Action ensures that these files have a parent folder and collaboration can be enabled in Box.

Configuration of IBM Connections

IBM Connections can be harvested and the *Copy from* action to a CIFS target is supported. Discovery Exports are also supported.

Note: Not all Profile fields are harvested, such as mobile number, pager number, and fax number. Custom attributes are supported. Libraries in Connections are links to FileNet objects; these files can be harvested.

The *Copy from* action is supported only with a CIFS target. Any harvested Connections instance has the following directory structure. It is a logical structure of hierarchy, not the actual way that data is stored.

```
Home
  Communities
  Files
  Forums
  Wikis
  Activities
  Blogs
  Status
  Bookmarks
  Events
  Comments
  Profiles
```

Note: When you create a Connections volume, the use of an initial directory, Start directory or End directory beyond two levels of recursion, is not supported. For example, `Home/Files` is supported, but `Home/Files/User1` is not. Additionally, harvest scoping, which is the advanced option in IBM StoredIQ Platform Data Server, is not supported.

A Connections volume that is created in IBM StoredIQ version 7.6.0.10 must be fully reharvested after an upgrade for the objects to be viewed.

Each of the subdirectories has elements under the user name directory. So, if User A created a forum, the directory to find it is `home/forums/userA/<Forum Name>`. If a user created a forum inside a community that is owned by User B, the directory to find it is `home/communities/userB/<Community Name>/forums/<Forum Name>`.

For more information about Connections attributes and their use examples, see Connections attributes.

Setting up the administrator access on Connections

IBM Connections needs an actual user account, not `wasadmin`, to be set up with admin and search-admin privileges. The following procedure describes how to set up the administrator access on Connections.

About this task

This procedure needs to be done in the WebSphere® Application Server Administrative Console by the administrator.

Procedure

1. In the Administrative Console, follow these steps.
   a. Go to **Users and Groups** > **Administrative user roles**.
   b. Select **Add...** > **Administrator role**.
   c. Search for the Connections user account that is used to add the Connections volume in IBM StoredIQ and add it to the role.
   d. Click **OK** and select **Save directly to the master configuration**.
2. Follow these steps for each of these applications: Activities, Blogs, Communities, Dogear, Files, Forums, News, Profiles, RichTextEditors, Search, URLPreview, and Wikis.
   a. In the Administrative Console, go to **Applications** > **Application Types** > **WebSphere enterprise applications**.
   b. Select an application from the list.
   c. Select **Security role to user/group mapping** > **Search-admin** > **Map Users...**.
   d. Search for the Connections user account that is used to add the Connections volume in IBM StoredIQ and add it to the role.
   e. Click **OK** > **OK**.
   f. Select **Save directly to the master configuration**.
3. Follow these steps for each of these applications: Activities, Blogs, Common, Communities, Files, Forums, Homepage, Metrics, News, Profiles, PushNotification, RichTextEditors, Search, URLPreview, WidgetContainer, and Wikis.
   a. In the Administrative Console, go to **Applications** > **Application Types** > **WebSphere enterprise applications**.
   b. Select an application from the list.
   c. Select **Security role to user/group mapping** > **admin** > **Map Users...**.
   d. Search for the Connections user account that is used to add the Connections volume in IBM StoredIQ and add it to the role.
   e. Click **OK** > **OK**.
   f. Select **Save directly to the master configuration**.

OneDrive for Business volumes: configuration note

OneDrive for Business volumes can be added only from IBM StoredIQ Administrator and have several configuration prerequisites that must be met.

- To create the application, log in to `apps.dev.microsoft.com`. Click **Add an App** to get `client_id`, `client_secret`, and `redirect_uri`.
- In the OneDrive application, click **Add Platform** and make sure **Allow Implicit Flow** is set and set **Redirect URLs** to `https://x.x.x.x/proxy/enamel/1.0/oauthtokengenerator/onedrive`, where x.x.x.x is the IP address of the application stack. Save the changes in the IBM StoredIQ application.
- Before OneDrive volumes are added, the Client ID, Client Secret, and Redirect URI must be stored in an application stack `oauth.conf` file. Its scope is set as `scope=offline_access Files.ReadWrite.All Group.ReadWrite.All Notes.ReadWrite.All`. The token and authentication URL are specified as `auth_url=https://login.microsoftonline.com/common/oauth2/v2.0/authorize token_url=https://login.microsoftonline.com/common/oauth2/v2.0/token`. The key in the `oauth` file for OneDrive configuration is `OAUTHLIB_RELAX_TOKEN_SCOPE=1`. This file can be found in the application stack directory in `/etc/siq/oauth.conf`.

The OneDrive configuration must be a separate configuration under the box configuration. A sample oauth.conf file is as follows.

```
[onedrive]
client_id=d1106b42-ec8e-40b9-9832-d5c970aaa7bc
client_secret=bmSuRLefzaDKeMmrDNmdFS5
redirect_uri=https://9.30.52.69/proxy/enamel/1.0/oauthtokengenerator/onedrive
auth_url=https://login.microsoftonline.com/common/oauth2/v2.0/authorize
token_url=https://login.microsoftonline.com/common/oauth2/v2.0/token
scope=offline_access Files.ReadWrite.All Group.ReadWrite.All Notes.
ReadWrite.All Sites.ReadWrite.All User.ReadWrite.All
OAUTHLIB_RELAX_TOKEN_SCOPE=1
```

After the `oauth.config` file is modified, restart the `uwsgi` service by logging in to the application stack with PuTTY and run service appstack stop and then service appstack start.

- The application stack must be configured for the SSL access; it must have access to the OneDrive authentication endpoint on `Microsoft.graph.com`. If the application stack was deployed without SSL enabled, run `/siq/bin/certcfg` and follow the instructions on the Certificate Configuration screen.
- A data server that manages a OneDrive volume must have access to `Microsoft.graph.com`.

IBM StoredIQ supports index, copying from, and exporting from OneDrive for Business to CIFS and NFS shares. However, OneNote is not supported in the OneDrive harvest.

IBM Content Manager attributes

In the `SiqDocument` item type, various attributes are increased when you run copy to IBM Content Manager.

In the `SiqDocument` item type, the length of the following attributes is increased 128 - 256 bytes when you run copy to IBM Content Manager:

- `SiqServer`
- `SiqShare`
- `SiqInitialDirectory`
- `SiqFileName`
- `SiqContainerPath`
- `SiqOwner`

This change is handled automatically if you do not already have an `SiqDocument` item type in your IBM Content Manager server. However, if this item type exists, it must be recreated with the new attribute lengths for this change to take effect.

Note: If you run a working CopyTo IBM Content Manager without issues or if you know that your attribute lengths are not greater than 128 in length, then you can defer this action as you did not encounter the attribute length issue.

Note: The attribute length is in bytes. The number of actual characters this length holds varies based on the database code page that is used. For example, if ASCII is used, then the number of characters is equal to the number of bytes. If UTF-8 is used, the number of bytes per character varies depending on the characters. Without this change, you see errors if the source attributes for a copy to IBM Content Manager exceed 128 bytes. If you see these errors, you need to take the following actions.

To extend the length of these attributes, take the following actions:

- If the SiqDocument item type does not exist in the IBM Content Manager server, create a new IBM Content Manager volume with a CopyTo option to select SiqDocument. It creates the item type and its attributes with the correct lengths.
- If SiqDocument item type exists in the IBM Content Manager server and you need to fix the attribute length problem, then the administrator must delete or drop the SiqDocument item type and recreate or update the IBM Content Manager volume that is used for copy. It automatically creates the item types and attributes desired.

  Note: Take a backup of the database before you drop and recreate SiqDocument item type. When you drop the item type, you permanently lose all the items (documents) stored in it. If the source documents are still available, you can run copy again to copy the data back into this item type. If no items exist in the item type, then it is not an issue.
  To drop the SiqDocument item type,

  1. Delete all the items in the SiqDocument. This delete is permanent and you lose all of the data.
  2. Delete the SiqDocument item type.
  3. Delete all the attributes that belong to the SiqDocument

If the item type exists and contains data that you need to keep, and you need to extend these attributes, this process is possible through direct database manipulation. However, this process is not supported and issues that derive from it cannot be covered by IBM support. If you want this process, services must be employed to make these database changes.

Adding a retention volume

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

Procedure

1. Click **Data Servers** and select a data server, then click **Add Volume**. The Add Volume dialog box appears.
2. In the Add Volume dialog box, complete these fields:

Table 21. 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

Table 21. Retention volume options (continued)

Add Volume dialog box option	Action	Notes	Source type
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 **View Volumes**. The added volume appears, listed as a retention volume.

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 22. 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

Table 22. Discovery export volume options (continued)

Add Volume dialog box option	Action	Notes	Source type
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.

Editing a volume

You can edit the existing volumes from the **Volumes** pane.

About this task

To edit a volume, follow these steps.

Procedure

1. Select a volume from the existing volume list and click **Edit Volume**. The Edit Volume window appears.
2. Make changes as needed.

   Note: These volumes cannot be edited from IBM StoredIQ Administrator: Chatter, Discovery Accelerator, Domino, Enterprise Vault, and Jive. They can be edited from the IBM StoredIQ data server.
3. Click **Save**.

Viewing details of multiple volumes

Detailed, combined information for multiple volumes can be viewed.

Procedure

1. Click **View Volumes**, and then press and hold the **CTRL** key or **SHIFT** key to select the volumes for which you want to view combined detailed information. The Combined Volume Details page appears, showing the combined details of the selected volumes, including the number of data objects, the size of the total number of data objects, the number of data servers, and the number of volumes.

2. Add or remove volumes from the Combined Volume Details view using the **CTRL** key. The volume is added or removed, and the Combined Volume Details view is updated automatically.

Restarting services

Services on a data server can be restarted when viewing that data server's details.

Procedure

1. Click **Data Servers**, then select a data server by clicking on the data server row. The **Details: <data server name>** page appears and the **Restart Services** and **Reboot Data Server** buttons become active.
2. Click **Restart Services**. The Confirm Restart Services window appears, verifying that you want to restart services on the selected data server.
3. Click **Restart Services**, and services restart. When the services are restarting, its status is changed to **Restarting**, and the status message is one of the following:
   - System and services running.
   - System is running but some services are still coming up.
   - Some query processes are not running.
   - There was a problem with the RAID controller.
   - Some services are in error.
   - Cannot connect to system and error.

   All other details appear as **N/A** until the services restart, at which point the data server's details are once again depicted.

Rebooting a data server

Data servers can be rebooted when viewing that data server's details.

Procedure

1. Click **Data** and select a data server by clicking the data server row.
2. The **Restart Services** and **Reboot Data Server** buttons become active.
3. Click **Reboot Data Server**. The Confirm Data Server Reboot window appears, verifying that you want to reboot the selected data server.
4. Click **Reboot Data Server**, and the data server reboots. When the data server is rebooting, its status is changed to **Maintenance Mode**, and the status message is **Rebooting-system**. All other details appear as **N/A** until the reboot completes, at which point the data server's details are once again depicted.

Removing a volume

Volumes can be deleted from the list of available data sources. However, volume deletion can cause data loss in existing infosets or indices that reference the deleted volumes.

Procedure

1. Select **Volumes** on the side navigation bar or **View Volumes** on the Data Servers pane, and then select the volume you want to delete.
2. Click **Remove Volume**. The volume is deleted, removing it from the list of available volumes.

Chapter 4. Harvests

The Harvests page of IBM StoredIQ Administrator displays the complete list of harvests, where you see all harvests currently under management.

When you click **Harvests** from the side navigation bar, the menu bar expands to show the **Scheduled**, **Current**, and **Completed** options. Select either **Scheduled**, **Current**, or **Completed**. Depending on the option that is selected, only those harvests in that category are shown Scheduled in the List of Harvests table.

The description of the different harvest fields is listed in the following table.

Table 23. List of Harvests fields and descriptions

List of Harvests	Description
Name	Lists the name of the harvest instance.
Type	Lists the type of harvest, either **Full** or **Incremental**.
Start time	Lists the date and time at which the harvest was initiated.
End time	Lists the date and time at which the harvest was ended.
Total time	Lists the time that the harvest takes.
Owner	Lists the owner of the harvest.

Harvesting a volume

By clicking **Harvest**, an administrator can initiate a full or incremental harvest of a primary volume without having to utilize IBM StoredIQ Platform Data Server.

Procedure

1. Click **Data** on the side navigation bar. The menu expands.
2. Click **Volumes**. Or, from the Data Servers pane, click **View Volumes**.
3. Select a volume and click **Harvest**. The Harvest Volume dialog box appears.
4. In the **Harvest Name** text box, enter a unique name for this harvest.
5. In the **Schedule Harvest** area, select either of the following options.
   - Select **Immediate** to harvest the volume immediately.
   - Click **Schedule** to schedule a date and time for the harvest to occur. Enter the date (YYYY-MM-DD) and time in the text boxes.
6. In the area of **Harvest Options**, select **Incremental** or **Full**.
   - An **Incremental** harvest only harvests any changes on the selected volume since its last harvest.
   - A **Full** harvest performs a full harvest of the selected volume, regardless of when it was last harvested.
7. Click **Save**. The volume is harvested.

Searching for a specific harvest

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

Procedure

1. Click **Harvests** on the side navigation bar, select either **Scheduled**, **Current**, or **Completed**. Depending on the option that is selected, only harvests in the selected category are shown in the table.
2. In the **Enter key term(s)** text box, enter the name of the harvest for which you want to search, and then click **Search**. You can enter either the full harvest name or a portion of it. To remove the search term, click the **X** to the right of the **Enter key term(s)....** text box. Harvests that match the entered search term are returned.
3. Select a harvest from the returned list of harvests.

Stopping a harvest

Harvests that are not yet completed can be stopped.

Procedure

1. In the **Scheduled Harvests** list, select **Current Harvests**. The list of harvests that are currently running opens.
2. Select the harvest that you want to stop and click **Stop Harvest**. The **Confirm Stopping Harvest** window appears.
3. Click **Stop Harvest** to confirm that the harvest needs to be stopped and not completed.

Editing a harvest

Existing harvests can be modified as needed.

Procedure

1. In the Scheduled harvests pane, select the harvest that you want to edit, and then click **Edit Harvest**. The Edit Harvest dialog box appears.

   Note: Only scheduled harvests can be edited.
2. In the Edit Harvest dialog box, edit the harvest as needed.
3. Click **Save** to save your changes.

Deleting a harvest

Scheduled harvests can be deleted from the List of Harvests page.

Procedure

1. In the Scheduled harvests pane, select the scheduled harvest that you want to delete, and then click **Delete Harvest**. The Delete Harvest window appears.
2. Click **Delete**. The harvest is deleted from the list of scheduled harvests.

Auditing a harvest

Harvest audit information is available in the **Completed** harvests page. It lists the number of the exceptions that the selected harvest has. This exception list serves as an information trace of what errors occurred during the harvest process.

Before you begin

You need to know this information before you audit a harvest:

- Harvests must be completed and ready for view from the **Completed** harvest pane.
- Only harvests that are initiated on the IBM StoredIQ Administrator server reflect audits on IBM StoredIQ Administrator. Harvests that are done from the Data Server do not reflect audits on IBM StoredIQ Administrator.
- Step-up Full-Text and Step-up Snippet are not reflected on the IBM StoredIQ Administrator server.
- For harvests that were completed before IBM StoredIQ release 7.6.0.8, the **Skipped directories** and **Skipped - user configuration** categories display zero.
- **Skipped directories** are counted in the **Data Objects Not Harvested** field in the **Harvest Details** pane. They are not counted on the Data Server as processing exceptions.

About this task

The Harvest Details view presents summary information about a harvest. A number greater than zero for **Data Objects Not Harvested** is a link to show the specific files that have exceptions during the harvest process. The following list provides the categories of exceptions that you can audit.

- **Binary text extracted, full processing complete**
- **Binary text extracted, partial processing complete**
- **Content Skipped - user configuration**
- **Content type known, but cannot extract content**
- **Content type known, but error processing content**
- **Content type known, partial processing complete**
- **Content type unknown, not processed**
- **Error gathering ACLs**
- **Error processing binary content**
- **Skipped - cannot access data object**
- **Skipped directories**
- **Skipped - user configuration**

Procedure

1. Select **Harvest** > **Completed** on the side navigation bar. A list of completed harvests displays.
2. Select a harvest that you want to audit. A **Harvest Details** pane displays at the bottom of the page.
3. Look for the number next to **Data Objects Not Harvested** . If the number is zero, it means that there are no exception for this harvest. If you see a number greater than zero, click the number to proceed to the harvest exception page.

   Note: If a harvest from the link expired or a volume was deleted, the exception page shows the categories and their respective counts, but no exceptions are listed. An appropriate message also appears.
4. Select the categories from the list to audit the harvest. Selecting or clearing the categories updates the list of the exceptions.
5. Click the breadcrumb to go back to the **Completed** harvest page.

Chapter 5. System infosets

System infosets can be created in IBM StoredIQ Administrator 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 several system infosets through which a user must cull exists, the shortcut aspect is removed.

Table 24. 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.

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 right 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.

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 are 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**. It now appears in the System Infosets pane.

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.
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.

Viewing infosets that use a system infoset

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

Procedure

1. Select a system infoset.
2. In the **In use by** column, click the link. The Infoset List window appears, listing the infosets that use this system infoset.
3. Click **Close**.

Viewing access members

Procedure

1. Select a system infoset for which you want to view its access 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**.

Deleting a system infoset

System infosets can be deleted from the System Infosets pane 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 System Infosets pane, 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**. The system infoset is deleted from the list of available system infosets.

Chapter 6. User management

From the Users pane, 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 in the Users pane. When a user is selected, that user can be viewed, edited, or deleted.

Table 25. User names and descriptions

User 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.

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.

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.

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 dialog box appears.
2. In the Add User dialog box, complete these fields.
   a. In the **User name** text box, enter 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 **Active** or **Inactive**.

      Note: The default option during user creation is **Active**.
   i. Click **Save**, and a user is created.

   The newly created user appears within the Users pane.

Connecting to the LDAP server

Before users can be imported and authenticated from an LDAP or LDAP-supported authentication system, you must connect to the LDAP server.

Procedure

1. Exit IBM StoredIQ Administrator.
2. Verify that your IBM StoredIQ application stack is upgraded to version 7.6.0.6 or higher.

3. In an SSH tool, enter `/siq/bin/ldapcfg`. The LDAP Configuration window appears.
4. In the LDAP Configuration window, enter these configuration details.
   a. Press the space bar to select **Allow External LDAP User**. It enables LDAP.
   b. In the **LDAP URL** text box, enter the IP address of your server.
      - For the LDAP URL, enter

        ```
        ldap://ldap-server-hostname
        ldap://<ip-address>
        ```

      - For SSL mode, enter

        ```
        ldaps://ldap-server-hostname
        ldaps://<ip-address>
        ```

   c. In the **LDAP User** text box, enter the LDAP user. For example, enter `cn=user,dc=example,dc=com` .
   d. In the **LDAP Password** text box, enter the LDAP user's password.
   e. In the **Base DN** text box, enter the base DN. For example, enter `dc=siqdomain,dc=com`
   f. Select **Test connection** at the bottom of the window to test whether you are connected to the LDAP server. The **LDAP Status** text box shows whether the connection test passed or failed.
5. Required: The LDAP Configuration window also contains attribute-mapping details.

   Note: Do not modify these attribute-mapping changes unless your schema is different. If you have questions about the schema or these changes, contact your company's LDAP administrator.
   a. The **First Name** text box is automatically filled with **givenName**.
   b. The **Last Name** text box is automatically filled with **sn**.
   c. The **Email** text box is automatically filled with **mail**.
   d. The **Username** text box is automatically filled with **cn**. This field cannot contain spaces or special characters. Additionally, it must be a part of your DN for users to import successfully.
6. Restart the `uwsgi` by using this command: `/siq/bin/monit restart uwsgi`
7. Check the status by using this command: `/siq/bin/monit summary`
8. Log in to IBM StoredIQ Administrator, click **Users** > **Import User**. The list of users that were imported from the LDAP server is visible within the Import User dialog box.

Synchronization of the LDAP server and application stack database

To keep users' details properly synchronized between the LDAP server and application stack database, a daily synchronization process occurs.

Onboarded users' details in the application stack database can become stale, which means that they are not synchronized properly with the LDAP server. To keep users' details properly synchronized between the LDAP server and application stack database, a recurring background process runs once per day. This daily task avoids overloading a synchronization request on the LDAP server. However, the synchronization can also be done manually. For example, if the administrators want to update users' details to be in application stack database, they can run the `sync-ldap` script manually. The users' updates are synchronized immediately from the LDAP server.

Note: During synchronization, only active LDAP users' details are updated. If LDAP users are deleted from the LDAP server, then those users' details are marked as inactive in the application stack database.

This command starts a manual synchronization between the LDAP server and the application stack database: `/siq/bin/sync-ldap`

Importing a user

Users can be imported and authenticated from an LDAP or LDAP-supported authentication systems, enabling enterprise security policy enforcement.

Procedure

1. Click **Users**, and then click **Import User**. The Import User dialog box appears.
2. In the Import User dialog box, the **User name**, **First name**, **Last name**, and **Email address** text boxes are read-only fields. By default, these fields are filled automatically by using content from the **LDAP Users**.
3. To load a user into the form, select a user from the **LDAP Users** list. When a user is selected for import, the **Save** button is enabled. If a user was already imported, a message appears, informing you to select another user.
4. In the **Role** list, select **Admin**, **Data User**, **Legal Ops User**, **Policy User**, or **SDK User**.
5. In the **Status** list, select **Active** or **Inactive**.
6. Click **Save**, and then click **Close**. The **Save** button resets the Import User dialog box. To import another user, select that user within the **LDAP Users** table.
7. Click **Save**, and then click **Close**.

Editing an imported user

When editing LDAP imported users, only the user's role and status can be modified.

Procedure

1. Click **Users**. In the list of users, select a user, and then click **Edit User**. The Edit Imported User dialog box appears. The **User name**, **First name**, **Last name**, and **Email address** text boxes are read-only fields.
2. In the **Role** list, select **Admin**, **Data User**, **Discovery User**, **Policy User**, or **SDK User**.
3. In the **Status** list, select **Active** or **Inactive**.
4. Click **Save**.

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 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**.

Viewing details of a user

Basic user information such as first and surname, email address, and role can be viewed from the **View User Details** pane.

Procedure

1. Click **Users** on the side navigation bar and select a user. The View User Details pane appears to display the detailed information about the user.
2. Within the View User Details pane, 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.

3. 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.

   Note: These columns can be sorted by clicking the up and down arrows to the right of the column name.

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 a user.
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.

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. Re-enter 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.

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.

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.

Chapter 7. 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 in IBM StoredIQ Data Workbench.

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 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.
- **Discovery Export DAT Light**. The discovery export action copies data objects and generates a Concordance DAT (Light) 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 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.
- **Watson Curation**. The Watson Curation action copies data objects to the Watson Curator repository. The Watson Curation action is available in IBM StoredIQ Policy Manager.
- **Step-up Full-Text**. The step-up full-text action can be run on an infoset, providing content to the full-text index for objects that are infoset members.
- **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 26. Actions column names and descriptions

Actions Column Name	Description
Action Name	Lists the name of the action.

Table 26. Actions column names and descriptions (continued)

Actions Column Name	Description
Type	Lists the action type. They are **Copy**, **Copy to Retention, Delete**, **Modify Attribute**, **Move**, **Discovery Export XML**, **Discovery Export DAT**, **Discovery Export DAT Light**, **Step-up Full-Text**, **Step-up Snippet**, and **Watson Curation**.
Last Modified	Lists the date and time at which the action was last modified.
Target Set Name	Lists the destined target set of the action.
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.

Procedure

1. In the **Enter key term(s)** 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 right of the text box. Actions that match the entered search term are returned.
2. Edit, clone, or delete the action as needed.

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 and step-up full-text actions are built-in actions, which means that no options are provided within the user interface. You cannot create or edit the step-up snippet or step-up full-text action as they are provided as prepopulated options 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. Enter the name of this action.
   b. Enter a brief description of this action. This action might be used by other people, so an accurate description can be helpful to others.
   c. 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	Select a target set. If the action is copy to Box, you must select or create a target set with **Box** as its source type.
Destination Directory	Either enter a file path or use the list option to select a previously entered file path. **Note:** For the Watson Curation action, the destination directory is not supported.
Directory Structure	If you want to **Recreate directory structure on destination**, select the check box . **Note: Recreate directory structure on destination** is enabled by default when copy to Box target set is configured.
Hashes	If you want to **Compute a hash value for each data object**, select the check box.
Harvesting	Select or clear the check box for **Do not auto-harvest destination volume after copy**.
Mapping	If you select Box as the target data source, you get the following mapping options: • **Map permissions from source to destination** **Note:** It is supported only for CIFS source volumes. • **Preserve owners from source to destination** **Note:** It is supported only for SharePoint and CIFS source volumes.
Others	**Preserve version chains on destination**. It is available for copy to Box target sets only.
Email Message Format	Select the format of the email message.
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	Select a target set.
Hashes	If you want to **Compute a hash value for each data object**, select the check box.
Harvesting	If an automatic harvest of the retention volume is not wanted, select the check box to **Do not auto-harvest destination volume after copy**. 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	Select the format of the email message.
Retention Tag	Enter the key term for the tag.
Retention Period	Enter a number, and then select **Days**, **Weeks**, **Months**, or **Years**.

Delete action options	Action
Delete Parameters	Select the check box **Do not delete data objects that have been accessed since last harvest**.

Discovery Export EDRM XML action options	Action
Destination Directory	Either enter a file path or use the list option to select a previously entered file path. **Note:** For the Watson Curation action, the destination directory is not supported.
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	Select the format of the email message.
Limit Export Process	Select the check box to **Limit Export Process** and enter the number of data objects.
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	Either enter a file path or use the list option to select a previously entered file path. **Note:** For the Watson Curation action, the destination directory is not supported.
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 **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 **Limit Export Process** and enter the number of data objects.
SharePoint User Profile Items	Select the check boxes of application SharePoint items, including **Notes**, **Libraries**, **Blog Posts**, **Wikis**, and **Misc**.

Discovery Export DAT-Light action options	Action
Destination Directory	Either enter a file path or use the list option to select a previously entered file path. **Note:** For the Watson Curation action, the destination directory is not supported.

Discovery Export DAT-Light action options	Action
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`. **Note:** This field is a required field. Select the check box to **Pad Zeros**.
Email Message Format	Select the format of the email message.
Limit Export Process	Select the check box **Limit Export Process** and enter the number of data objects.
SharePoint User Profile Items	Select the check boxes of application SharePoint items, including **Notes**, **Libraries**, **Blog Posts**, **Wikis**, and **Misc**.

Modify Attribute action options	Action
Action Name	Enter the action's name.
Action Type	Select the **Modify Attribute** option.
Description of Action	Enter the action's description.
Attribute Type	Select from the following options: **Read-only** or **Read-write**.

Move action options	Action
Destination Directory	Either enter a file path or use the list option to select a previously entered file path. **Note:** For the Watson Curation action, the destination directory is not supported.
Directory Structure	If you want to **Recreate directory structure on destination**, select the check box.

Watson Curation action options	Action
Destination Directory	Either enter a file path or use the list option to select a previously entered file path. **Note:** For the Watson Curation action, the destination directory is not supported.
Directory Structure	If you want to **Recreate directory structure on destination**, select the check box.
Hashes	If you want to **Compute a hash value for each data object**, select the check box.
Harvesting	If an automatic harvest of the volume is not wanted, select the check box to **Do not auto-harvest destination volume after copy**.
Collection ID	Enter the name of the `collection` in Watson Curator.
SharePoint User Profile Items	Select the check boxes of application SharePoint items, including **Notes**, **Libraries**, **Blog Posts**, **Wikis**, and **Misc**.

Step-up Analytics options	Action
Cartridge list which contains the following fields:	List of cartridges that are uploaded in the upload cartridge user interface. Select single or multiple cartridges.

Cartridge fields	Description
Name	Name that is entered by an administrator when a cartridge is uploaded.
Supported results	Lists the supported results for the cartridge.
Languages supported	Lists the languages of the documents that are annotated during the analytic step-up. If no language is listed, all documents that are responsive are annotated. If specific languages are listed, then only documents that are responsive in those languages are annotated.

Note: A cartridge must be uploaded in the cartridge user interface by the administrator before a Step-up Analytics is configured. For instructions for uploading cartridges, see "Uploading cartridges" on page 82.

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

Creating a Step-up Analytics action

Follow these steps to create a Step-up Analytics action.

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. Enter a name for the Step-up Analytics in the **Action Name** field.
   b. Select Step-up Analytics from the **Action Type** list.
   c. Type a description for the action in the **Description of Action** field.
   d. Click **Next**. A list of available cartridges displays for you to choose. One or more cartridges must be selected. Otherwise, you cannot save the action.
   e. Click **Save**.

   Note: You cannot change the cartridges that are selected for a Step-up Analytics action after you create the Step-up Analytics action. If you do, an error message displays and the action cannot be saved.

Deleting an action

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

Procedure

1. In the action list pane, select the action that you want to delete, and then click **Delete Action**. The Delete Action window appears.
2. Click **Delete**. The action is removed from the action list and it is no longer available for use.

Cloning an action

Actions can be cloned or duplicated.

Procedure

1. In the action list pane, 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. Enter the name of this action.

      Note: The name of the cloned action must be different from the name of the original action.
   b. Select the action type, and then click **Next** to continue creating the action.
   c. Enter a brief description of this action.

      Note: This action might be used by other 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**.

Editing an action

Existing actions can be modified.

Procedure

1. In the action list pane, select the action that you want to edit, and then click **Edit Action**. The Edit Actions dialog box appears.
2. Edit the action as needed.
3. Click **Save**.

Chapter 8. Target sets

Some 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 27. 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: **Box**, **CIFS**, **CMIS**, **Documentum**, **FileNet**, **HDFS**, **IBM Content Manager**, **NFS**, and **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.

Procedure

1. In the **Enter key terms** text box, enter the name of the target set for which you want 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 right of the text box. Target sets that match the entered search term are returned.
2. Edit, clone, or delete the target set as needed.

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** from the side navigation bar, and then click **Create Set**. The Add Target Set dialog box appears.
2. In the Add Target Set dialog box, complete these fields.
   a. Enter a name for this target set.
   b. Enter a description.
   c. Select the type of this target set from **Primary**, **Retention**, or **Export**.
   d. Select the source type of this target set from **Box**, **CIFS**, **CMIS**, **Documentum**, **FileNet**, **HDFS**, **IBM Content Manager**, **NFS**, and **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. Repeat this step for each added data server.
3. Click **Save**.

Editing a target set

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

Procedure

1. In the target set list pane, click the name of the target set that you want 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**.

Cloning a target set

Target sets can also be cloned or duplicated.

Procedure

1. In the target set list pane, click the name of the target set that you want 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**.

Deleting a target set

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

Procedure

1. In the target set list pane, select the target set you want to delete, and then click **Delete Set**. The Delete Target Set window appears.
2. Click **Delete**. The target set is removed from the target list pane. It is no longer available for use.

Chapter 9. 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
- Content Collector Manifest File CSV Export
- CSV All Audited Object Exports
- CSV Attribute Summary
- CSV Exception List Export
- CSV Infoset Data Object Export
- Data Assessment Report
- Data Topology Report
- Duplication Summary Report
- Overlay Hit Report
- Term Hit Report

Table 28. 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 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 that you want 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 right of the text box. Reports that match the entered search term are returned.

2. Select the wanted report.

Creating a report

Besides the built-in reports, you can create custom reports of your own.

About this task

For information about configuring custom reports, see "Customizable reports."

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. Enter a name for this report.
   b. Enter a description of this report.
   c. Click **Choose File**, and then find and select the appropriate report-package file.
3. Click **Create**. The newly created report is added to the report list pane.

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 function. 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, and 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 StoredIQ 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.

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.

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.

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.

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**.

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.

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.

Uploading the report design

IBM StoredIQ supports uploading a report design in a compressed file. 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` 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 stylesheets for a single report design package, you must do one of the following tasks:

- Create a project for each report design. Each project contains only one `rptdesign` file.
- Create a folder for each report design within the project. Each folder contains only one `rptdesign` file.

By selecting one of the report-design management options, you can manage all images and stylesheets. They are used by the design in a single location, making the creation of the compressed file easier as you compress a folder. 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` file are also present within the compressed file.
- Only one report parameter is defined in the `rptdesign` file.
- Images use these extensions: `.bmp`, `.jpg`, `.jpeg`, `.jpe`, `.jfif`, `.gif`, `.png`, `.tif`, `.tiff`, `.ico`, or `.svg`.
- Stylesheets must use the `.css` file extension.

Procedure

1. Log in to IBM StoredIQ Administrator.
2. Select **Reports** from the side navigation bar.
3. Click **Create**.

4. Type the name of the report.
5. Optional: Enter the description of the report.
6. Select a report compressed file from your local drive and click **Create** to upload a report design to IBM StoredIQ Administrator. The report design is then also shown in IBM StoredIQ Data Workbench.

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 application stack: `/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 application stack, 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 application stack after new fonts are deployed.
   a. To restart services, issue the command `bootstrap server.mode idle` and wait until `bootstrap server.mode` shows idle.
   b. Run the command `bootstrap server.mode production`
   c. To check the status, run the command `bootstrap server.mode`

Changing the reportuser password

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

Procedure

1. To change the password for the reportuser on the application stack:
   a. Using an SSH tool, log into the application stack 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. Using an SSH tool, log into the gateway as `root`.
   b. Enter `cd /usr/local/storediq/bin/util/`
   c. Run `change_reportuser_password.sh` with the new password.

Chapter 10. Auto-classification models

Auto-classification models, also called automated document categorization, integrates the IBM® Content Classification's classification model into the IBM StoredIQ Platform 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 synchronous 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. 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. You provide feedback, and a deferred retraining operation adjusts the model and implements corrections based on that feedback. 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 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 infoset 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 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.

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, in the **Action** pane, 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 as **Action** on the side navigation bar.

   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. After the enhancement completes in IBM StoredIQ Data Workbench, go to **Create > Build** and apply the auto-class filter to create a new infoset.
6. Review the results of the newly created infoset.

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, you must increase the virtual machine settings for processors and memory on data servers to these new minimums:

- 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 type. 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.

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.

Building an auto-classification model

Creating 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 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 Workbench onto 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.
   For best practices to create an Auto-classification model, see "Best practices for creating an Auto-classification model" on page 76.

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.

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, select **Auto Classification** from the side navigation bar. The right pane displays the imported models as described in this table.

Table 29. 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. In the Auto Classification pane, 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.

Selecting a classification model

By selecting a classification model, you can view the details of that model, including the status of the last retrain, how much feedback it has received, and how many infosets have been enhanced using this model.

Procedure

1. Within IBM StoredIQ Administrator, click **Auto-Classification**. The List of Classification Models page appears.
2. Select the model for which you want to view details. The <model name> Details panel appears.

Table 30. Row names and descriptions of the <model name> Details panel

Row name	Description
Status	Lists the classification model's status, specifically **Available**, **Defunct**, **Deleting**, **Failed**, **Invalid**, **Pending**, **Processing**, or **Retraining**. **Note:** In the **Processing** or **Retraining** states, models cannot be edited.
Last retrained	Lists the date and time this classification model was last retrained.
Last retrain status	Lists the classification model's retrain status. Status messages are **Success**, **Failed**, and **N/A**.
Total feedbacks	Lists the number of times the model has received feedback.
Total feedbacks after last retrain	Lists the number of times this model received feedback after its last retrain date.
Total infosets created using this model	Lists the number of infosets that use this classification model.
Total infosets enhanced by this model	Lists the number of infosets that have been enhanced by this 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.

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 right of the text box. Classification models that match the entered search term are returned.
2. Edit, delete, or retrain the classification model as needed.

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. In the Auto Classification pane, click the name of the classification model that you want to edit, and then click **Edit**. 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 Edit Model window.
3. Click **Save**.

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 Auto Classification pane, select the classification model that you want to delete, and then click **Delete Model**. The Delete Model window appears.
2. Click **Delete**. The classification model is removed from the Auto Classification pane. It is no longer available for use.

Retraining a classification model

Data models are deployed regularly against data sets. When a classification model is deployed, it can be applied against millions of data objects as opposed to the few hundred that were used for the initial training of the model. By applying the model to so many data objects, it gives you much more granular data points. Because it takes time and effort to "train" or prepare a classification model against a set, data experts' positive or negative feedback is useful in retraining a classification model. Does the model work satisfactorily? Where does the model not seem to function well? Are there inconsistencies? Data experts have intimate knowledge of particular data sets and classification models, and they can provide positive and negative feedback about a model's accuracy. This feedback can be used to reinforce or improve a model's behavior; you can retrain the model to improve its accuracy.

Before you begin

With the assistance of IBM software support, modify the data learning settings.

1. Click **Project > Project Options**. The Project Options dialog box appears.
2. In the Project Options window, click **Advanced**. The Advanced Project Options window appears.
3. Select the **Maintain learning data set with knowledge base** option. Accept all other default values.
4. Click **OK**.

The data-learning settings can also be modified in the Create, Analyze, and Learn dialog box by clicking **Settings**. For more information, see http://www-01.ibm.com/support/knowledgecenter/SSBRAM_8.8.0/com.ibm.classify.workbench.doc/c_WBG_Saving_Learning_Data.htm.

Procedure

1. Create a classification model. In IBM StoredIQ Administrator, click **Create Model** to create a classification model. You must set an attribute when you create a classification model.
2. Create a user infoset. In IBM StoredIQ Data Workbench, create a user infoset, and then select that user infoset so that it is active.
3. Run the step-up snippet action. In IBM StoredIQ Data Workbench, run the step-up snippet action on the newly created user infoset.
4. Apply an enhancement. In IBM StoredIQ Data Workbench, select an enhancement to apply to the newly created user infoset, and then click **Run Enhancement**.
5. Apply a filter. In IBM StoredIQ Data Workbench, create a filter.

a. Select the **Auto-classify** option.
b. In the list under the **Auto-classify** option, select the classification model for which results must be filtered.
c. Select the categories. By default, the **Category** option is selected, as are all of the available categories.
d. Determine how the results are displayed, selecting either **And the highest score in the selected category** or **And where the score <is/Greater than/Less than/Not equal/Greater than or equal/Less than or equal> <1>**.

6. View the objects in the Data Object Viewer. Click **Preview Filter Results**, and the objects are displayed in the **Filter Results** area. Select one of the returned objects to view it in the Data Object Viewer.
7. Provide feedback for this data object. Click the button to the right of **Auto-class Scores**. It displays the list of categories and the object's score for each category. The category with the highest score is listed at the top. As the Data Expert, select the categories to which the selected data object belongs. Data objects with a value closer to 1 are more closely associated with that category. Data objects with a value closer to 0 are not closely associated with that category. Click **Submit**. This feedback helps to improve the accuracy and validity of the classification model. Repeat this step for as many different data objects as you like. The number of feedback submissions can be seen within the <model name> Details window in IBM StoredIQ Administrator.
8. In IBM StoredIQ Administrator, click **Auto-Classification** and select the classification model. The <model name> Details window shows that it received feedback. Click **Retrain** to start the retrain process for this model. If the model was uploaded without a learning archive (SARC file) or if it received no feedback, then the **Retrain** button is disabled.
9. Run the enhancement against the infoset again, noting the new scores within the Data Object Viewer. The improved scores indicate greater validity and accuracy.

Best practices for creating an Auto-classification model

The following topics address some best practices specific for creating an Auto-classification model.

IBM Content Classifier Workbench

IBM Content Classifier Workbench is an application of the IBM Content Classification product. When you create an Auto-classification model with this application, you must ensure that you follow these practices.

Preparing the sample content sets for training

- The more number of samples per category in the content set you provide, the better it is for training. However, the documents in the content set for a single category in the sample must not be self-contradictory as it can confuse the classifier.
- The best number is 40 - 50 items per category.
- The sample content set must be prepared or reviewed by some subject matter experts.

Creating a Knowledge Base

You must specify at least one field from the content set to be analyzed by the Classification Module. This field must contain meaningful text.

To tell the Classification Module that this field is to be analyzed, you assign a content type value to the field. Do not assign a content type value for the fields that contain non-textual values, such as account numbers or telephone numbers, or non-meaningful text, such as a content field that contains arbitrary administrative comments about each content item.

The number of documents per category

The number of documents per category in the Knowledge Base must be inversely proportional to the number of categories. The fewer categories that you have, the more documents you want to maintain within the Knowledge Base. The default value is 80. To change this value, follow these steps:

- Open an existing Knowledge Base or create a new one. Click **Project Options**.
- Click **Advanced** and edit **Optimally maintain X documents per category**.
- Increase the number as necessary. If you have five categories, for example, this number can be set to a range of 200 to 300. If you have many categories, it is best to keep this number around 100 rather than 200 or 300.

Training your Knowledge Base

Follow these steps to train your Knowledge Base:

- Click **Create, analyze and learn**.
- Use any other option besides **Create using all, analyze using all**.
- Use **Create using all, analyze using all** to assess only the data consistency of the document set contents and to check whether the information in the content can statistically predict the categories. A good document set can yield in this test more than 95% for the top ranking category.
- Use one of the other methods on the list, such as **Create using even, analyze using odd** to have a more realistic prediction estimation after the assessment is done.
- The right option depends on how the content set was prepared and shared by the Subject Matter Experts. Some degree of randomness must be introduced to show how the content set is partitioned for training and analysis. In general, the best option is **Create Using Even, Analyze using Odd**.
- In the next page, enable **Save learning data (SARC File) with Knowledge base** for the model to accept feedback.

Analyzing and reviewing your Knowledge Base by using the built-in reports

- Click **View Reports** on the toolbar to open the View Reports window.
- To understand the overall accuracy of your Knowledge Base, view **Knowledge Base Data Sheet**, **Cumulative Success** summary reports, and **Total Precision vs. Recall graph**.
- In **Knowledge Base Data Sheet**, set the first column to report a number greater than 95. It indicates that performance of the Knowledge Base is good during analysis. The less this number is than 95, the more self-contradictory the Knowledge Base becomes. This expectation is related to the data consistency verification when you use **Create using all, analysis using all**, while for the regular creating and analysis case, the number can be lower.

- In the **Total Precision vs. Recall graph**, the curve must be in the upper-right portion. It indicates that performance of the Knowledge Base is good.
- If the curve is on the lower-left portion, it means that performance of the Knowledge Base is poor.

Choosing to reserve items from the training set in the Knowledge Base

- Open an existing Knowledge Base or create a new one.
- Click **Knowledge Base**.
- Right click **Reserve Items**. It ensures that no matter how much feedback you give to this Knowledge Base in the future, the reliable sample data set that is used for training is always retained within the Knowledge Base.

Another way to reserve items is to use the **Freeze** option when you use the BundleDPKB utility to prepare a model and upload it into StoredIQ. An example of adding the **Freeze** option to the BundleDPKB command is:

```
BundleDPKB "C:\IBM\ContentClassification\Classification
Workbench\Projects_Unicode" project_DP Output freeze
```

Working with Learning Data (SARC)

In the Knowledge Base Workbench projects, you can set the Knowledge Base to work with the Learning Data method. During the training process, a file that retains the important training content information (SARC) is generated along with the Knowledge Base.

Assessing the data quality in IBM Content Classifier Workbench

Run **Create using all, analysis using all** on the data to verify that it is statistically consistent. Results are expected to be unrealistically high, which is about 95% correct for the top category. If you get less than that it means the data sets of the categories are contradictory.

When you pass that initial test, you must run a real test by training a part of the data and matching the other part. It gives you an idea of the results that you get in deployed systems. You must use either even or odd sets or random cuts.

For more information about how to work with IBM Content Classification, see this Redbook at http://www.redbooks.ibm.com/redbooks/pdfs/sg247707.pdf

IBM StoredIQ

When you create an Auto-classification model with IBM StoredIQ, you must ensure that you follow these practices.

- Run a Step-up Snippet enhancement on the infoset before you run an auto-classification enhancement with your model.
- Every time a model is retrained, another auto-classification enhancement must be run on the infoset to compute the classification results of the newly trained model.
- Step-up Snippet needs to be run only once for an infoset.
- After you run a Step-up Snippet, exceptions can be generated from the data servers that participate in the Step-up Snippet operation. Exceptions affect not only the classification, but also any feedback on the documents.
- It is assumed that auto-classification training and feedback work without human error. You must not provide feedback with any doubt about which category the document belongs to.

- Single feedback does not have a significant impact on a model. For example, giving feedback for the `foo.doc` file as Category A does not classify the `foo.doc` file as A after a retrain and subsequent enhancement.
- Repetitive feedback for the same document does not increase impact on the Knowledge Base. If you give feedback for the `foo.doc` file as Category A, after every retrain you keep giving the same feedback, the subsequent feedback has no additional effect on the Knowledge Base.

Chapter 11. Cartridges

Cartridges are units of deployment and management for analytic plugin extensions in the IBM StoredIQ Platform. Cartridges enable IBM StoredIQ to detect additional information in documents.

Cartridges are basically compressed files that contain analysis logic. By adding a cartridge to IBM StoredIQ AppStack, it can detect new data in documents during indexing and make these new insights searchable. For example, a sensitive pattern cartridge can enable IBM StoredIQ to detect passport numbers, phone numbers, and other IDs. Step-up Analytics runs a selected cartridge on an infoset. IBM StoredIQ examines all documents in the infoset, applies the analytics that is contained in the cartridge to the document, and then stores the analysis results in the IBM StoredIQ index.

Cartridge types

Cartridges can contain analysis logic based on different technologies that range from simple regular expressions to full blown cognitive approaches like natural language processing (NLP).

Note: Cartridges that are provided by IBM for use with the IBM StoredIQ Platform can be downloaded from IBM Fix Central. IBM Business Partners and IBM Services can also create cartridges for specific purposes and provide them through their channels.

Deploying cartridges

Running cartridges in the documents go through the following process.

Before you begin

You must deploy a cartridge into the IBM StoredIQ AppStack and make it available as a new Step-up Analytics action before you use it to detect new data in the documents.

Procedure

1. Upload a cartridge in the cartridge deployment interface.

   See “Uploading cartridges” on page 82 for instructions to upload a cartridge. After the cartridge is uploaded, the cartridge is validated and information that is pulled from the cartridge is populated in the cartridge page grid.
2. Create a Step-up Analytics action by using one or more uploaded cartridges.

   Create a new Step-up Analytics action in the Action section. Select one or more cartridges as part of the Step-up Analytics action creation. After the Step-up Analytics is created, the cartridges that are specified for the Step-up Analytics action cannot be changed.
3. Run the Step-up Analytics on an infoset.

   During this process, the analysis logic that is encapsulated in the cartridge is applied to the documents in the infoset that is selected for the Step-up Analytics. For example, the regex in the cartridge is run against the document content. The results of the analysis are then indexed to be available for quick

searches through filters. To search for analysis results from a cartridge, the data expert can create a filter by using the cartridge syntax.

Uploading cartridges

Upload a cartridge before you can create a Step-up Analytics action that uses a cartridge.

Before you begin

Download IBM StoredIQ cartridges from IBM Fix Central to your local drive. Cartridges can also be acquired from IBM Services or IBM Business Partners.

About this task

To upload a cartridge, select **Cartridges** from the side navigation bar in the IBM StoredIQ Administrator user interface. The grid displays the cartridges as described in the following table.

Table 31. Cartridges

Cartridge column name	Description
Name	Name that is entered by the user when a cartridge is created.
Cartridge name	Lists the cartridge name. It is populated in the grid by cartridge information after the cartridge is uploaded.
Supported results	Lists the supported results for the cartridge to be used in search and filter creation. It is populated in the grid by cartridge information after the cartridge is validated.
Status	Lists the cartridge status, such as Pending, Processing, Failed, and Available.
Creation date	Lists the date that a cartridge is uploaded.

Procedure

1. Click **Upload Cartridge** in the Cartridge pane. The Upload Cartridge window appears.
2. Enter a display name for the cartridge in the **Name** field.
3. Select the check box to accept the terms of the license that is contained in the cartridge. The **Choose File** button is not active until the license is accepted.
4. Click **Save**.

   The cartridge validates automatically. After validation is complete, the grid populates with data from the cartridge.

   Note: Every uploaded cartridge must have a unique cartridge name. The cartridge name is given in the `cartridge.properties` file in the cartridge. Specifying a cartridge that has the same name as a previously uploaded cartridge causes failure of the upload operation.
5. Select a cartridge after validation is complete.

   The **Details** section is populated with this information:

Details field	Description
Status	Lists the cartridge status, such as Pending, Processing, Failed, and Available.
Last run	Date that the cartridge was last run by using a Step-up Analytics action.
Last error	Error message that was generated if upload or update failed to validate the cartridge.
Version	Lists the version of the cartridge that is populated by information within the cartridge.
Description	Lists the description of the cartridge that is populated by information within the cartridge.
Language supported	Lists languages of the documents that the cartridge annotates and is populated by information within the cartridge.
Total infosets using this cartridge	Lists the total number of infosets that are used by this cartridge.
Total active Step-up Analytics executions using this cartridge	Lists the number of active Step-ups Analytics executions using this cartridge.
Total Step-up Analytics actions using this cartridge	The number of Step-up Analytics actions that use this cartridge.

Updating cartridges

Follow these steps to update a cartridge from an older version.

About this task

Procedure

1. Select a cartridge from the list.
2. Click **Update** from the top of the window. The Update Cartridge window opens.
3. Enter a name for the new cartridge and check the agreement box.
4. Click **Choose File** to select a new version of the previously uploaded cartridge from the directory of your desktop.
5. Click **Save**.

   If the cartridge that is provided in the update is invalid or if an unexpected error occurs during the update process, an error message is returned and displayed in the **Last Error** field of the Details pane. If an error occurs, the previously uploaded version of the cartridge remains available.

   If Step-up Analytics executions are active, the **Update** button is not active. The number of active Step-up Analytics executions is given in the **Total active Step-up Analytics executions using this cartridge** field in the Details pane.

   The new cartridge must have the same cartridge name as the cartridge that it is being updated. The cartridge name is given in the `cartridge.properties` file that is contained in the cartridge.

Deleting cartridges

Follow these steps to delete a cartridge.

Procedure

1. Select a cartridge from the list.
2. Click **Delete**.

   Note: The delete function is active only for failed cartridges or cartridges that are not in use by a Step-up Analytics action and infoset.
3. Click **Delete** to confirm your action on the Delete confirmation dialog.

Chapter 12. 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 Platform as a whole. These technical aspects of using IBM StoredIQ Administrator are 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 32. 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.

Procedure

1. Click **Concepts** from the side navigation bar.
2. Click **Create Member**. The Add Identity dialog box appears. The **Identity** concept was created as a predefined, preconfigured aspect.
3. In the Add Identity dialog box, complete these fields.

a. Enter the identity's given name.
b. Enter the identity's surname.
c. Enter the identity's email address.
d. Enter the name of the company that is associated with this identity.
e. Enter the name of the identity's department.
f. Enter the identity's owner ID.

4. Click **Save** or **Cancel** to return to the concept list pane.

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 that 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 right of the text box. Concepts that match the entered search term are returned.
2. Edit or delete the concept as needed.

Editing a member of a concept

Existing concepts and their members can be modified.

Procedure

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

Deleting a member of a concept

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

About this task

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

Procedure

1. From the concept list pane, 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.

Chapter 13. Mule scripts

Using the **Mule Script** tab, administrators can upload the Mule archive from their local directories. Mule archives are externally authored scripts, which are used by the IBM StoredIQ system to automate workflows.

This table lists and describes all of the fields that are seen within the Available Scripts page. When a Mule script is selected, it can be viewed in greater detail, edited, or deleted.

Table 33. Mule script columns and descriptions

Mule script column name	Description
Name	Lists the name of the Mule script.
Created	Lists the creation date and time of the Mule script.
Script package	Lists the name of the Mule script package.
Description	Lists the description of the Mule script.

Note: For information regarding how to set up the IBM StoredIQ environment to work with Mule Anypoint Studio and Mule scripts, see Appendix A, "IBM StoredIQ Mule script," on page 91.

Creating a Mule script

Created Mule scripts are available for use from the **Mule Scripts** page.

Procedure

1. Click **Mule scripts**, and then click **Create**. The Create Script dialog box appears.
2. In the Create Script dialog box, complete these fields. Newly created Mule scripts cannot be saved until these fields are complete.
   a. In the **Name** text box, enter the name of the Mule script.
   b. In the **Description** text box, enter a description of this newly created Mule script.
   c. Click **Choose File**, and then navigate to and select the script package file. The name of the selected script package file appears to the right of the **Choose File** button.
   d. Click **Create**, and Mule script appears within the list of available Mule scripts. Once Mule scripts are available within the Available Script list, the **Delete** and **Edit** buttons are active.

Searching for a specific Mule script

When you have multiple Mule scripts available to choose, you can search for a specific Mule script 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 Mule script for which you want to search, and then click **Search**. You can enter either the full Mule script 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. Mule scripts that match the entered search term are returned.
2. Select a Mule script from the returned list of scripts.

Viewing details of Mule scripts

Detailed information for any Mule script can be viewed.

Procedure

1. In Available Scripts, select the Mule script for which you need more information.
2. The Script Details pane appears at the bottom of the page, providing detailed information about that Mule script.

   Name Lists the name of the Mule script.

   Created
   Lists the creation date and time of the Mule script.

   Script Package
   Lists the name of the script package. By clicking the link, you can open a browser and download the file.

   Endpoint URL
   Lists the endpoint URL.

   Description
   Lists a description of the Mule script.

   Used by
   Lists the policies that use the Mule script.

Editing a Mule script

Existing Mule scripts can be modified as needed.

Procedure

1. In the Available Scripts page, select the Mule script that you want to edit, and then click **Edit**. The Edit Script dialog box appears.
2. In the Edit Script dialog box, edit the Mule script as needed.
3. Click **Save** to save your changes.

Deleting a Mule script

Existing Mule scripts can be deleted from the Available Scripts page.

About this task

Mule scripts that are being used by policies cannot be deleted, and deleted Mule scripts are not available for newly created policies. To see what policies use a Mule script, review the **Used By** section within the **Script Details**.

Procedure

1. In the Mule Script page, select the Mule script that you want to delete, and then click **Delete**. The Confirm Deletion window appears.
2. Click **Delete Script** to delete the selected Mule script. The Mule script is deleted from the list of available scripts.

Importing the IBM StoredIQ application stack SSL certificate

For the Mule Standalone engine to work with the SSL-enabled IBM StoredIQ application stack, you must first import your IBM StoredIQ application stack SSL root certificate to the JRE used by the Mule Standalone engine.

Before you begin

Verify that your application stack is updated to the most recent version.

Procedure

1. Locate the JRE used by the Mule Standalone engine. It is located at `/siq/env/java/jre`.
2. Import the SSL root certificate `/etc/siq/ssl/root.crt` into `/siq/env/java/jre/lib/security/cacerts`.
3. Save the cacerts file.
4. Restart Mule using this command: `/siq/bin/monit restart mule`
5. Check status using this command: `/siq/bin/monit summary`

Appendix A. IBM StoredIQ Mule script

Mule scripts are used to create automation workflow policies from the dashboard of IBM StoredIQ Policy Manager. Before you create a Mule script policy from the dashboard, a user who understands the customer's needs and requirements must help create the automation flow through Mule Anypoint Studio.

From the user interface of Anypoint Studio, the Mule scripts can be edited, tested, imported, and exported.

After Mule scripts are created, the administrator uses the IBM StoredIQ Administrator user interface to manage the Mule scripts to get them ready for the policy user to create policies.

For information about how to use the Mule products, see documentation on the official website of MuleSoft Inc.

Setting up Mule Anypoint Studio with IBM StoredIQ

Mule Anypoint Studio is a graphic design environment that provides functions of creating, editing, testing, importing, and exporting Mule scripts. You must set up the Mule Anypoint Studio environment before you create Mule scripts.

About this task

Follow these steps to set up the Mule Anypoint Studio environment.

Procedure

1. Select the correct operating system and then download Mule Anypoint Studio from https://www.mulesoft.com/lp/dl/studio
2. Install and start Mule Anypoint Studio.
   a. Extract the installation files into the local directory.
   b. Double-click `AnypointStudio.exe` to start Mule Anypoint Studio.
   c. Choose the workspace location to work with this tool.
3. Set up the Java environment from **Windows** > **Preferences** > **Java** > **Installed JREs**. Ensure that it points to JRE7 or JDK7.
4. Install Mule Runtime Community Edition (CE). By default, Mule Anypoint Studio is packaged with Enterprise Engine (EE), which has a 30-day license. You need to install Community Engine with the free license to create the flow with Mule Anypoint Studio.
   a. From **Help** > **Install New Software**, enter http://studio.mulesoft.org/r4/studio-runtimes.
   b. Select and install Mule Server 3.6.1 CE.
5. Install the Mule StoredIQ Connector.
   a. From **Help** > **Install New Software**, enter the URL of the connector update, `http://{StoredIQ_appStack host}/mule-connector-update`, to install `Community/StoredIQ Connector (Mule 3.5.0+)`.
   b. Verify that the StoredIQ Connector is installed, create a Mule project, and type `storediq` in the search box of the right window. A new icon **StoredIQ** displays.

c. Drag this connector to construct StoredIQ scripts. You can see a list of operations with the corresponding parameters in the Basic Settings window.

6. Install the `storediq_domain` project. For more information, see "Importing a domain project in Anypoint Studio" on page 109.

   Note: The `storediq_domain` is used to share the HTTP inbound endpoint across different StoredIQ Mule scripts. The shared listening port is 8081. All IBM StoredIQ Mule Scripts must be associated with `storediq_domain` to use this shared `HTTP` port. Download `storediq_domain.zip` from `http://{StoredIQ_appstack host}/download/mule-flows/storediq_domain.zip`. Import the domain project into Mule Anypoint Studio.

7. Create a Mule project.
   a. Select **New** > **Mule Project**.
   b. Enter the name of the Mule project and select **Mule Server 3.6.1 CE** as the Runtime engine.
   c. Associate Mule Projects with `storediq_domain`. See "Associating a Mule project with a domain project" on page 110.
   d. Drag the icons from the search box to create a Mule script with the HTTP endpoint.

      Note: You can test the Mule project within Mule Anypoint Studio.

8. Export the Mule projects to the Mule archive, that is, script package. This Mule archive can be passed to IBM StoredIQ administrator, who checks in Mule scripts to IBM StoredIQ Administrator.
   a. Click **Export** to export the existing Mule project into archive.
   b. Clear **Export parent domain project**.
   c. Save the archive as a `.zip` file in the local directory.

9. Import the Mule projects from the Mule archive. If you want to modify an existing Mule script, download the Mule archive from `http://{StoredIQ_appstack host}/download/mule-flows/<mule script>.zip`, and then import the Mule archive into Mule Anypoint Studio.

Importing the IBM StoredIQ application stack SSL certificate

For Mule Anypoint Studio to work with the SSL enabled IBM StoredIQ application stack, you must first import your IBM StoredIQ application stack SSL root certificate to your local JDK.

Before you begin

Make sure that your application stack is updated.

Procedure

1. Download the SSL root certificate file from IBM StoredIQ application stack, for example, `/etc/siq/ssl/root.crt`, to your local directory.
2. Check the version of JRE that is used by Mule Anypoint Studio.
3. Use keytool to open the `cacerts` file under `<JRE>/lib/security` folder.
4. Import the SSL root certificate into the cacerts file and save the `cacerts` file.
5. Restart Mule Anypoint Studio.

Mule script management

For information about use of Mule platform, see Mule documentation from its official website.

Mule script restrictions

You must follow these rules when you work with the Mule scripts:

- Mule scripts must have an inbound end-point `http` with the `storediq_domain` shared port 8081. For example, `http://localhost:8081/mule_policy1`.
- Mule scripts have a readme file that provides details about the script. This readme file is written by the user and meets these criteria:
  - Must be in the `.pdf` format and its file name can be accessible by URL.
  - Shows the diagram, description, and input data, if needed
  - Is put under the top level of the Mule archive
- Mule scripts might require input data. If so, that input data must be in the `.JSON` format with key/value pairs. For example,

```
{
  "sysinfoset_id": "d8242a7e-2658-4ca9-bf27-d54f30dae085",
  "filter_id": "48e225fb-b9d0-4ce4-9706-636c0832b1f1",
  "report_id": "00000000-0000-0000-0000-000000001000"
}
```

IBM StoredIQ Connector operations

You can construct a Mule script with the IBM StoredIQ operation blocks.

Table 34. IBM StoredIQ Connector operations

Connector operation	Parameter	Description
createFilter	**filter_name**	Name of the filter
	description	Description of the filter
	expression	Expression of the filter
createSystemInfoset	**infoset_name**	Name of the infoset
	description	Description of the infoset
	datasource_ids	List of data source IDs separated by commas
	access_type	Access type of infoset, either public or private
	user_list	List of user IDs that can be accessed to the infoset
	waitforCompletion	Flag to indicate sync and async

Table 34. IBM StoredIQ Connector operations (continued)

Connector operation	Parameter	Description
createUserInfosetFromFilter	**infoset_name**	Name of the infoset
	description	Description of the infoset
	infoset_id	ID of the infoset source
	filter_id	ID of the filter
	overlays	Overlay filter IDs separated by comma
	notify	Notification mail
	generate_datamap	Flag to indicate whether a datamap is generated
	waitforCompletion	Flag to indicate sync or async
createUserInfosetByOperations	**operations**	Infoset operations
	infoset_name	Name of the infoset
	description	Description of the infoset
	sources	List of source infoset IDs separated by commas
	overlays	Overlay filter IDs separated by commas
	notify	Notification email
	generate_datamap	Flag to indicate whether a datamap is generated
	waitforCompletion	Flag to indicate sync or async
createUserInfosetFromExceptions	**infoset_name**	Name of the infoset
	description	Description of the infoset
	infoset_id	ID of the infoset source
	event_ids	Event IDs separated by commas
	categories	Category names that are separated by commas
	overlays	Overlay filter IDs separated by commas
	notify	Notification email
	generate_datamap	Flag to indicate whether a datamap is generated
	waitforCompletion	Flag to indicate sync or async
createTargetset	**targetset_name**	Name of the target set
	description	Description of the target set
	type	Type of the target set
	datasource_ids	Data source IDs separated by comma

Table 34. IBM StoredIQ Connector operations (continued)

Connector operation	Parameter	Description
createDatasourceBox	**datasource_name**	Name of the data source
	dataserver_id	ID of the data server
	box_server	Box server
	box_include_users	Box included users for scoping the volume
	box_username	Box user name
	box_password	Box password
	index_options_container_metadata	Include metadata for contained objects
	index_options_full_text_content	Include content tagging and full-text index
createDatasourceNfs	**datasource_name**	Name of the data source
	dataserver_id	ID of the data server
	datasource_type	Type of the data source
	nfs_server	NFS server IP or host name
	nfs_export	NFS server export directory
	scoping_information_initial_directory	NFS server initial directory
	scoping_information_include_directories	NFS server include directories
	index_options_container_metadata	Include metadata for contained objects
	index_options_full_text_content	Include content tagging and full-text index
createDatasourceCifs	**datasource_name**	Name of the data source
	dataserver_id	ID of data server
	datasource_type	Type of the data source
	cifs_server	CIFS server IP or host name
	cifs_username	CIFS server user name
	cifs_password	CIFS server password
	cifs_shared	CIFS server share directory
	scoping_information_initial_directory	CIFS server initial directory
	scoping_information_include_directories	CIFS server include directories
	index_options_container_metadata	Include metadata for contained objects
	index_options_full_text_content	Include content tagging and full-text index

Table 34. IBM StoredIQ Connector operations (continued)

Connector operation	Parameter	Description
createDatasourceCm8	**datasource_name**	Name of the data source
	dataserver_id	ID of data server
	cm8_server	CM8 server IP or host name
	cm8_port	CM8 server port
	cm8_username	CM8 server user name
	cm8_password	CM8 server password
	cm8_repository	CM8 server repository
	cm8_dbtype	CM8 server remote DB type
	cm8_remotedb	CM8 server remote DB
	cm8_schema_name	CM8 server schema name
	cm8_connection_string	CM8 server connection string
	scoping_information_initial_directory	CM8 server initial directory
	scoping_information_include_directories	CM8 server include directories
	index_options_container_metadata	Include metadata for contained objects
	index_options_full_text_content	Include content tagging and full-text index
createDatasourceCmis	**datasource_name**	Name of the data source
	dataserver_id	ID of the data server
	cmis_server	CMIS server IP or host name
	cmis_port	CMIS server port
	cmis_username	CMIS server user name
	cmis_password	CMIS server password
	cmis_use_ssl	Boolean value to indicate ssl or not
	cmis_service	CMIS server service
	cmis_repository	CMIS server repository
	index_options_container_metadata	Include metadata for contained objects
	index_options_full_text_content	Include content tagging and full-text index

Table 34. IBM StoredIQ Connector operations (continued)

Connector operation	Parameter	Description
createDatasourceConnections	**datasource_name**	Name of the data source
	dataserver_id	ID of the data server
	connections_server	Connections server host name
	connections_username	Connections server user name
	connections_password	Connections server password
	connections_classname	Connections class name
	connections_repository	Connections repository
	connections_optionstring	Connections option string
	scoping_information_initial_directory	Connections server initial directory
	index_options_container_metadata	Include metadata for contained objects
	index_options_full_text_content	Include content tagging and full-text index
createDatasourceExchange	**datasource_name**	Name of the data source
	dataserver_id	ID of the data server
	exchange_server	Exchange server host name
	exchange_username	Exchange server name
	exchange_password	Exchange server password
	exchange_active_directory_server	Exchange server active directory server
	exchange_use_ssl	Exchange server use SSL flag
	exchange_version	Exchange server version
	exchange_folder	Exchange server folder
	exchange_mailbox_server	Exchange server mailbox server
	exchange_personal_archive	Exchange server personal archive
	scoping_information_initial_directory	Exchange server initial directory
	scoping_information_include_directories	Exchange server include directories
	index_options_container_metadata	Include metadata for contained objects
	index_options_full_text_content	Include content tagging and full-text index

Table 34. IBM StoredIQ Connector operations (continued)

Connector operation	Parameter	Description
createDatasourceSharepoint	**datasource_name**	Name of the data source
	dataserver_id	ID of the data server
	sharepoint_server	SharePoint server host name
	sharepoint_port	SharePoint server port
	sharepoint_username	SharePoint server user name
	sharepoint_password	SharePoint server password
	sharepoint_use_ssl	SharePoint server use SSL flag
	sharepoint_version	Available supported versions are 2003, 2007, 2010, 2013, 2016, online
	Subsites	Recurse into subsites
	Versions	Include all versions
	sharepoint_active_directory_server	SharePoint server active directory server
	sharepoint_site_url	SharePoint server site URL
	scoping_information_initial_directory	SharePoint server initial directory
	scoping_information_include_directories	SharePoint server include directories
	index_options_container_metadata	Include metadata for contained objects
	index_options_full_text_content	Include content tagging and full-text index
createDatasourceLivelink	**datasource_name**	Name of the data source
	dataserver_id	ID of the data server
	livelink_server	Livelink server IP or host name
	livelink_port	Livelink server port
	livelink_username	Livelink server user name
	livelink_password	Livelink server password
	livelink_database	Livelink server database
	livelink_search_slice	Livelink server search slice
	scoping_information_initial_directory	Livelink server initial directory
	scoping_information_include_directories	Livelink server include directories
	index_options_container_metadata	Include metadata for contained objects
	index_options_full_text_content	Include content tagging and full-text index

Table 34. IBM StoredIQ Connector operations (continued)

Connector operation	Parameter	Description
createDatasourceFilenet	**datasource_name**	Name of the data source
	dataserver_id	ID of the data server
	filenet_server	FileNet server IP or host name
	filenet_port	FileNet server port
	filenet_username	FileNet server user name
	filenet_password	FileNet server password
	filenet_connectionType	FileNet connection type
	filenet_path	FileNet server path
	filenet_stanza	FileNet server stanza
	filenet_object_store	FileNet server object store
	filenet_wheresql	FileNet sql where clause
	filenet_domain	FileNet server domain
	filenet_create_document	FileNet server create document flag
	scoping_information_initial_directory	FileNet server initial directory
	scoping_information_include_directories	FileNet server include directories
	index_options_container_metadata	Include metadata for contained objects
	index_options_full_text_content	Include content tagging and full-text index
createDatasourceNewsgator	**datasource_name**	Name of the data source
	dataserver_id	ID of the data server
	newsgator_server	NewsGator server host name
	newsgator_username	NewsGator server user name
	newsgator_password	NewsGator server password
	newsgator_use_ssl	Boolean value to indicate ssl or not
	index_options_container_metadata	Include metadata for contained objects
	index_options_full_text_content	Include content tagging and full-text index
harvestDatasource	**harvest_name**	Name of harvest
	datasource_id	ID of the data source
	harvest_type	Type of harvest
	load_type	Type of load
	schedule	Harvest schedule
	waitforCompletion	Boolean value to indicate sync or async

Table 34. IBM StoredIQ Connector operations (continued)

Connector operation	Parameter	Description
generateReport	**report_name**	Name of the report
	user_infoset_id	ID of the user infoset
	action_id	ID of the report
	terms	List of terms that are separated by commas for Term Hit Report only
	filter_ids	List of filter IDs that are separated by commas for Term Hit Report only
	execution_id	ID of execution for Audit Summary Report only
	notify	Notification email
	waitforCompletion	Boolean value to indicate sync or async
executeAction	**execution_name**	Name of the action
	user_infoset_id	ID of the user infoset
	action_id	ID of the action
	waitforCompletion	Boolean value to indicate sync or async
cancelExecution	**id**	ID of execution
createActionCopyPlain	**action_name**	Name of the action
	description	Description of the action
	targetset_id	Target set ID
	destination_path	Destination path
	copy_sharepoint_userprofile_notes	Boolean value to copy SharePoint user profile notes
	copy_sharepoint_userprofile_libraries	Boolean value to copy SharePoint user profile libraries
	copy_sharepoint_userprofile_posts	Boolean value to copy SharePoint user profile posts
	copy_sharepoint_userprofile_wikis	Boolean value to copy SharePoint user profile wikis
	copy_sharepoint_userprofile_misc	Boolean value to copy SharePoint user profile misc
	compute_hash_value	Boolean value to compute the harsh value
	recreate_directory_structure	Boolean value to re-create directory structure
	no_auto_harvest	Boolean value to indicate that there is no auto harvest

Table 34. IBM StoredIQ Connector operations (continued)

Connector operation	Parameter	Description
createActionCopyRetention	**action_name**	Name of the action
	description	Description of the action
	targetset_id	Target set ID
	retention_type	Type of retention
	retention_value	Retention value
	retention_tag	Tag of retention
	compute_hash_value	Boolean value to compute hash value
	no_auto_harvest	Boolean value to indicate that there is no auto harvest
createActionDiscoveryExportEdrmxml	**action_name**	Name of the action
	description	Description of the action
	targetset_id	Target set ID
	destination_path	Destination path
	data_object_limit	Data object limit
	copy_sharepoint_userprofile_notes	Boolean value to copy SharePoint user profile notes
	copy_sharepoint_userprofile_libraries	Boolean value to copy SharePoint user profile libraries
	copy_sharepoint_userprofile_posts	Boolean value to copy SharePoint user profile posts
	copy_sharepoint_userprofile_wikis	Boolean value to copy SharePoint user profile wikis
	copy_sharepoint_userprofile_misc	Boolean value to copy SharePoint user profile misc
	unique_document_id_prefix	Unique document ID prefix
	unique_document_id_digits	Unique document ID digits
	unique_document_id_pad_digits	Boolean value to unique document ID pad digits
	save_text_copy	Boolean value to save the text copy
	email_item_disposition	Email item disposition

Table 34. IBM StoredIQ Connector operations (continued)

Connector operation	Parameter	Description
createActionDiscoveryExportDat	**action_name**	Name of the action
	description	Description of the action
	targetset_id	Target set ID
	destination_path	Destination path
	data_object_limit	Data object limit
	copy_sharepoint_userprofile_notes	Boolean value to copy SharePoint user profile notes
	copy_sharepoint_userprofile_libraries	Boolean value to copy SharePoint user profile libraries
	copy_sharepoint_userprofile_posts	Boolean value to copy SharePoint user profile posts
	copy_sharepoint_userprofile_wikis	Boolean value to copy SharePoint user profile wikis
	copy_sharepoint_userprofile_misc	Boolean value to copy SharePoint user profile misc
	unique_document_id_prefix	Unique document ID prefix
	unique_document_id_digits	Unique document ID digits
	unique_document_id_pad_digits	Boolean value to unique document ID pad digits
	email_item_disposition	Email item disposition
createActionDiscoveryExportDatLight	**action_name**	Name of the action
	description	Description of the action
	targetset_id	Target set ID
	destination_path	Destination path
	data_object_limit	Data object limit
	copy_sharepoint_userprofile_notes	Boolean value to copy SharePoint user profile notes
	copy_sharepoint_userprofile_libraries	Boolean value to copy SharePoint user profile libraries
	copy_sharepoint_userprofile_posts	Boolean value to copy SharePoint user profile posts
	copy_sharepoint_userprofile_wikis	Boolean value to copy SharePoint user profile wikis
	copy_sharepoint_userprofile_misc	Boolean value to copy SharePoint user profile misc
	unique_document_id_prefix	Unique document ID prefix
	unique_document_id_digits	Unique document ID digits
	email_item_disposition	Email item disposition

Table 34. IBM StoredIQ Connector operations (continued)

Connector operation	Parameter	Description
createActionWatsonCuration	**action_name**	Name of the action
	description	Description of the action
	targetset_id	Target set ID
	destination_path	Destination path
	collection_id	Collection ID
	copy_sharepoint_userprofile_notes	Boolean value to copy SharePoint user profile notes
	copy_sharepoint_userprofile_libraries	Boolean value to copy SharePoint user profile libraries
	copy_sharepoint_userprofile_posts	Boolean value to copy SharePoint user profile posts
	copy_sharepoint_userprofile_wikis	Boolean value to copy SharePoint user profile wikis
	copy_sharepoint_userprofile_misc	Boolean value to copy SharePoint user profile misc
	compute_hash_value	Boolean value to compute hash value
	recreate_directory_structure	Boolean value to recreate directory structure
	no_auto_harvest	Boolean value to indicate that there is no auto harvest
createActionExportExceptions	**action_name**	Name of the action
	description	Description of the action
	export_format	Export format
createActionModifyAttribute	**action_name**	Name of the action
	description	Description of the action
	attribute	Modification type
createActionMovePlain	**action_name**	Name of the action
	description	Description of the action
	targetset_id	ID of the target set
	destination_path	Destination path
	recreate_directory_structure	Boolean value to re-create a directory structure
createActionMoveRetention	**action_name**	Name of the action
	description	Description of the action
	targetset_id	ID of the target set
	retention_type	Type of the retention
	retention_value	Retention value
createActionDeleteNodes	**action_name**	Name of the action
	description	Description of the action
	ignore_accessed_objects	Boolean to ignore accessed objects

Table 34. IBM StoredIQ Connector operations (continued)

Connector operation	Parameter	Description
search	**resourceType**	Type of the resource
	id	ID of the resource
	nameOptions	Name search options
	resource_name	Name of the resource
	status	Status of the resource
	max	Maximum number of the result to be returned
	delete	Flag to indicate whether the resource is to be deleted

Creating the IBM StoredIQ Mule script

Use Mule Anypoint Studio to create the IBM StoredIQ Mule script.

Before you begin

Before you use a Mule script to access the IBM StoredIQ resources, you must authenticate to the IBM StoredIQ application stack server. You must configure three parameters for authentication:

- StoredIQ application stack hostname
- StoredIQ application stack user name
- StoredIQ application stack password

You can use either of the following ways to configure connections to the IBM StoredIQ application stack server:

- Select the **Global Mule Configuration Elements** tab from Mule script. Click **Create** and search for `storediq`.
- Drag an IBM StoredIQ connector in the Mule script. Click the **+** sign to the left of **Connector Configuration**. Type the value of the IBM StoredIQ hostname, user name, and password. Click **Test Connection** to ensure that the authentication is successful.

Note: The authentication is valid for all IBM StoredIQ connectors within the same flow.

Procedure

1. Search and delete IBM StoredIQ resources.
   a. Enter a name in the **Display Name** field. For example, `list dataservers`.
   b. Select **Search** from the **Operation** list.
   c. Select **dataserver** from the **Resource Type** list.
   d. Enter information in the **Search** option for specific resources.
   e. To delete the selected resources, select the **Delete the resource** checkbox.
2. Create and harvest the data source.
   - To create a data source:
     a. Enter a name in the **Display Name** field, for example, `create nfs datasource`.
     b. Select **StoredIQ** from **Connector Configuration**.

c. Enter the parameters for **Name**, for example, `mule_datasource`, **DataServer Id**, for example, UUID of the data server, and select the type.

- To harvest a data source:
  a. Enter a name in the **Display Name** field, for example, `harvest datasource`.
  b. Select **StoredIQ** from the **Connector Configuration** list.
  c. Select **Harvest datasource** from the **Operation** list.
  d. Enter the parameters for **Name**, for example, `mule_harvest`, **DataSource Ids**, for example, UUID of the data source, and select **Harvest Type** and **Load Type** from the lists.
  e. Select **Wait for Completion** under **Flag**.

3. Create a system infoset.
   a. Enter a name in the **Display Name** field. For example, create system infoset.
   b. Select **StoredIQ** from **Connector Configuration**.
   c. Select **Create system infoset** from **Operation**.
   d. Enter the parameters for **Infoset Name**. For example, `mule_sys_infoset`.
   e. Enter the parameter for **DataSource Ids**. For example, UUID of the data source.
   f. Select **Wait for Completion** under **Flag**.
4. Create a filter.
   a. Enter a name in the **Display Name** field. For example, `create user filter`.
   b. Select **StoredIQ** from **Connector Configuration**.
   c. Select **Create filter** from **Operation**.
   d. Enter the parameters for the filter name. For example, `mule_filter`.
   e. Enter the expression: `att:"/Library/Attributes/System metadata/Extension" IS any ("txt") IN all`.
5. Create a user infoset.
   a. Enter a name in the **Display Name** field. For example, `create user infoset`.
   b. Select **StoredIQ** from **Connector Configuration**.
   c. Select **Create user infoset from filter** from **Operation**.
   d. Enter the parameters for **Infoset Name**. For example, `mule_user_infoset`.
   e. Enter **Filter Id**. For example, UUID of the filter.
   f. Enter the **Infoset Source Ids**. For example, if the operation is to create the infoset `usrifsC` from the operations between the infosets `usrifsA` and `usrifsB` and the `sessionVars.usrifsA` and `sessionVars.usrifsB` hold the IDs respectively, the **Infoset Source Ids (separated by comma):** field is to be updated as `#[sessionVars.usrifsA + ',' + sessionVars.usrifsB]`
   g. Select **Wait for Completion** under **Flag**.
6. Create a target set.
   a. Enter a name in the **Display Name** field. For example, `create targetset`
   b. Select **StoredIQ** from **Connector Configuration**.
   c. Select **Create targetset** from **Operation**.
   d. Enter the parameters for **Name**. For example, `mule_targetset`.
   e. Enter the parameter for **DataSource Ids**. For example, UUID of the data source.
   f. Select **primary** as its **Type**.

7. Create an action.
   a. Enter a name in the **Display Name** field. For example, `create action`.
   b. Select **StoredIQ** from **Connector Configuration**.
   c. Select **Create action copy plain** from **Operation**.
   d. Enter the parameters for **Name**. For example, `mule_action`.
   e. Enter the parameter for **TargetSet Id**. For example, UUID of the target set.
8. Run an action.
   a. Enter a name in the **Display Name** field. For example, `run action`.
   b. Select **StoredIQ** from **Connector Configuration**.
   c. Select **Execute action** from **Operation**.
   d. Enter the parameters for **Name**. For example, `mule_execution_action`.
   e. Enter the parameter for **User Infoset Id**. For example, UUID of the user infoset.
   f. Enter the parameter for **Action Id**. For example, UUID of the action.
9. Generate a report.
   a. Enter a name in the **Display Name** field. For example, `generate report`.
   b. Select **StoredIQ** from **Connector Configuration**.
   c. Select **Generate report** from **Operation**.
   d. Enter the parameters for **Name**. For example, `mule_report`.
   e. Enter the parameter for **User Infoset Id**. For example, UUID of the user infoset.
   f. Enter the parameter for **Report Id**. For example, UUID of the report action.

   Note: If you select **Wait for Completion**, the StoredIQ flow waits for the report to be successfully generated and then returns the report URL in the result string. If you do not select **Wait for Completion**, the StoredIQ flow does not wait for completion of the report generation and the result string does not contain the report URL.

Deploying the Mule script

Mule Standalone Engine is preinstalled in IBM StoredIQ application stack. It hosts the Mule scripts without Mule Anypoint Studio.

About this task

Use these commands to start or stop the engine:

```
/siq/bin/monit start mule
/siq/bin/monit stop mule
```

When the engine is up and running, use the following steps to deploy or undeploy the Mule script.

Procedure

1. Export the Mule script from Anypoint Studio into an archive file in the `.zip` format. For example, `mule_helloworld.zip`
2. Write a readme file to explain the Mule script diagram. Provide a description and example input data in the `.pdf` format and with the same name as the Mule script. For example, `mule_helloworld.pdf`.

   A readme file can look like the following example.

Description:

This Mule script creates a user infoset from the system infoset and filter and then it generates a report. The unique ID of the system infoset, filter, and report can be passed in from the input file in the `.json` format.

Example of input data:

```
 {
"sysinfoset_id": "d8242a7e-2658-4ca9-bf27-d54f30dae085",
"filter_id": "48e225fb-b9d0-4ce4-9706-636c0832b1f1",
"report_id": "00000000-0000-0000-0000-000000001000"
}
```

3. Save this readme file into the Mule archive file that was created in Step 1.
4. Pass the Mule script archive file to the administrator of IBM StoredIQ Administrator. For example, `mule_helloworld.zip`.
5. The administrator can create a Mule script with the provided Mule archive from IBM StoredIQ Administrator. See *Creating a Mule script*. After the Mule script is created successfully, it is automatically deployed into Mule Standalone Engine.
6. When the administrator deletes an existing Mule script from IBM StoredIQ Administrator user interface, the Mule script is automatically undeployed from Mule Standalone Engine.

Results

The Mule script log traces the Mule script process. It is in the `/var/siq/log` directory.

Input files and confidential data encryption

Using IBM StoredIQ Mule script with the JSON input file provides flexibility to input data and secure confidential information.

Using the JSON input file

You can design and deploy Mule scripts to allow users to input the data later in IBM StoredIQ during the process of creating a IBM StoredIQ policy. The input file must be in the JSON format.

See examples of the JSON input file and how to use it to retrieve the data from Mule Anypoint Studio as follows.

`Example1.json`

JSON input file in key or value pair as follows.

```
{
          "SIQ_DataserverID": "c7d5b4af-ee1b-41fe-9f27-2a52f382eacc",
          "SIQ_DataserverIP": "192.168.0.20",
          "CIFS_Prefix": "sample1",
          "CIFS_Volume_Name": "CIFS",
          "CIFS_Description": "Sample to create volume",
          "CIFS_DatasourceProtocol": "cifs",
          "CIFS_DatasourceServer": "192.168.225.2",
          "CIFS_DatasourceUsername": "domain\\user",
          "CIFS_DatasourcePassword": "userpassword",
          "CIFS_DatasourceShare": "share1",
          "CIFS_DatasourceADServer": "",
          "CIFS_DatasourceInitialDir": "testfiles",
          "CIFS_DatasourceIncludeDirs": ""
      }
```

To retrieve the value in the Mule Anypoint Studio flow, use this JSON path expression:

```
#[json:CIFS_Volume_Name]
```

```
Example2.json
```

JSON input file in key or value pair as follows.

```
{
      "inputdata":
      [
          {
              "SIQ_DataserverID": "c7d5b4af-ee1b-41fe-9f27-2a52f382eacc",
              "SIQ_DataserverIP": "192.168.0.20",
              "CIFS_Prefix": "sample2",
              "CIFS_Name": "CIFS",
              "CIFS_Description": "Volume CIFS using mule script",
              "CIFS_DatasourceProtocol": "cifs",
              "CIFS_DatasourceServer": "192.168.0.3",
              "CIFS_DatasourceUsername": "domain\\user",
              "CIFS_DatasourcePassword": "userpassword",
              "CIFS_DatasourceShare": "share2",
              "CIFS_DatasourceADServer": "",
              "CIFS_ExchangeVersion": "",
              "CIFS_ExchangeVirtualRoot": "",
              "CIFS_ExchangeUseSSL": "",
              "CIFS_DatasourceInitialDir": "testfiles",
              "CIFS_DatasourceIncludeDirs": ""
          },
          {
              "SIQ_DataserverID": "c7d5b4af-ee1b-41fe-9f27-2a52f382eacc",
              "SIQ_DataserverIP": "192.168.0.20",
              "EXCH_Prefix": "sample2",
              "EXCH_Name": "EXCHANGE",
              "EXCH_Description": "Volume EXCHANGE using mule script",
              "EXCH_DatasourceProtocol": "exchange",
              "EXCH_DatasourceServer": "192.168.0.19",
              "EXCH_DatasourceUsername": "domain\\user",
              "EXCH_DatasourcePassword": "userpassword",
              "EXCH_DatasourceField": "",
              "EXCH_DatasourceADServer": "",
              "EXCH_ExchangeVersion": "",
              "EXCH_ExchangeVirtualRoot": "",
              "EXCH_ExchangeUseSSL": "true",
              "EXCH_DatasourceInitialDir": "",
              "EXCH_DatasourceIncludeDirs": ""
          }
        ]
     }
```

To retrieve the value in the Mule Anypoint Studio flow, use this JSON path expression:

```
#[json:inputdata[0]/CIFS_Description]
```

Encrypting confidential data

For confidential input data, IBM StoredIQ encrypts the data before it stores the input data into the database. At design time, your confidential data in the JSON input file can be in text format. After you upload it to IBM StoredIQ during the process of creating a policy, the confidential data is encrypted. Users see only the encrypted confidential data if they download the input file from the Policy Details page. To use StoredIQ to encrypt confidential data from input file, the input key must contain a `password` or `Password` keyword; it is case-insensitive.

For example, the plain JSON input data is:

```
{"username":"user1", "password":"p@ssw0rd"}
```

When you create or edit a policy or create a new run, you upload the input file. The password fields are encrypted to be, for example, `{"username":"user1", "password":"001U2FsdGVkXwpyTH1hM+i0QP3bpqLE05EvtbhuBLnhfw=="}`.

Configuring Mule script timeout lengths

By default, the Mule script timeout value is 432000, or five days. This value can be modified by administrators.

About this task

By default, the Mule script timeout value is 432000, or five days. This value can be modified by administrators.

Procedure

1. Verify that your IBM StoredIQ application stack is upgraded to the most recent, supported version.
2. In an SSH tool, enter `/siq/bin/mulecfg`. The Console window appears.
3. In the Console window, the **Mule Script timeout (seconds)** parameter is shown. Modify the default value of 432000.
4. Press **Tab** to select **Save** or **Save and exit**, and then press **Enter**.
5. Restart the uwsgi using this command: `/siq/bin/monit restart uwsgi`
6. Check the status using this command: `/siq/bin/monit summary`

Mule domain project

Mule defines selected connectors as common resources and they are accessible to all applications that are deployed under the same domain. These resources are known as shared resources.

To use these shared resources, you must associate the Mule project with the Mule domain project, then reference the shared resource in the Mule project.

The following sections use HTTP Listener Connector as an example to demonstrate how to share HTTP host and listener port across different Mule projects.

Importing a domain project in Anypoint Studio

The following procedure uses `storediq_domain.zip` as an example to describe how to import a Mule domain project.

Procedure

1. From Anypoint Studio, right-click **Package Explorer**, and from the menu select **Import > Anypoint Studio > Anypoint Studio-generated Deployable archive (.ZIP)**.
2. Specify the location of **storediq_domain.zip**, and then click **Finish**. The `mule-domain-config.xml` file displays automatically. The shared resource `http:listner-config` is defined inside.

Associating a Mule project with a domain project

Follow these steps to associate any new or existing Mule project with the domain project.

Procedure

1. Select an existing Mule project that has an HTTP listener connector as the inbound endpoint. Open the project's Mule configuration xml file and select the proper domain under **Runtime > Domain**. For example, in Mule project, open `mule-project.xml` and then select `storediq_domain` as the domain.

   Note: You can add multiple Mule projects under the same domain, but you can associate only one project with one domain at a time.
2. In the Mule project, directly refer to a connector configuration in the HTTP listener connector. For example, in Mule project, use `HTTP_Listener_storediq` from the domain project `storediq_domain` without creating a new HTTP listener configuration.
3. Use the path value in the HTTP listener connector to differentiate Mule projects that share the host and port within the same domain.

Appendix B. IBM StoredIQ Connector operations

The following table provides information about IBM StoredIQ Connector operations, their parameters, and descriptions.

Table 35. IBM StoredIQ Connector operations

Connector operation	Parameter	Description
cancelExecution	id	ID of execution
createActionCopyPlain	action_name	Name of the action
	description	Description of the action
	targetset_id	Target set ID
	destination_path	Destination path
	copy_sharepoint_userprofile_notes	Boolean value to copy SharePoint user profile notes
	copy_sharepoint_userprofile_libraries	Boolean value to copy SharePoint user profile libraries
	copy_sharepoint_userprofile_posts	Boolean value to copy SharePoint user profile posts
	copy_sharepoint_userprofile_wikis	Boolean value to copy SharePoint user profile wikis
	copy_sharepoint_userprofile_misc	Boolean value to copy SharePoint user profile misc
	compute_hash_value	Boolean value to compute the harsh value
	recreate_directory_structure	Boolean value to re-create directory structure
	no_auto_harvest	Boolean value to indicate that there is no auto harvest
createActionCopyRetention	action_name	Name of the action
	description	Description of the action
	targetset_id	Target set ID
	retention_type	Type of retention
	retention_value	Retention value
	retention_tag	Tag of retention
	compute_hash_value	Boolean value to compute hash value
	no_auto_harvest	Boolean value to indicate that there is no auto harvest
createActionDeleteNodes	action_name	Name of the action
	description	Description of the action
	ignore_accessed_objects	Boolean to ignore accessed objects

Table 35. IBM StoredIQ Connector operations (continued)

Connector operation	Parameter	Description
createActionDiscoveryExportDat	action_name	Name of the action
	description	Description of the action
	targetset_id	Target set ID
	destination_path	Destination path
	data_object_limit	Data object limit
	copy_sharepoint_userprofile_notes	Boolean value to copy SharePoint user profile notes
	copy_sharepoint_userprofile_libraries	Boolean value to copy SharePoint user profile libraries
	copy_sharepoint_userprofile_posts	Boolean value to copy SharePoint user profile posts
	copy_sharepoint_userprofile_wikis	Boolean value to copy SharePoint user profile wikis
	copy_sharepoint_userprofile_misc	Boolean value to copy SharePoint user profile misc
	unique_document_id_prefix	Unique document ID prefix
	unique_document_id_digits	Unique document ID digits
	unique_document_id_pad_digits	Boolean value to unique document ID pad digits
	email_item_disposition	Email item disposition
createActionDiscoveryExportDatLight	action_name	Name of the action
	description	Description of the action
	targetset_id	Target set ID
	destination_path	Destination path
	data_object_limit	Data object limit
	copy_sharepoint_userprofile_notes	Boolean value to copy SharePoint user profile notes
	copy_sharepoint_userprofile_libraries	Boolean value to copy SharePoint user profile libraries
	copy_sharepoint_userprofile_posts	Boolean value to copy SharePoint user profile posts
	copy_sharepoint_userprofile_wikis	Boolean value to copy SharePoint user profile wikis
	copy_sharepoint_userprofile_misc	Boolean value to copy SharePoint user profile misc
	unique_document_id_prefix	Unique document ID prefix
	unique_document_id_digits	Unique document ID digits
	email_item_disposition	Email item disposition

Table 35. IBM StoredIQ Connector operations (continued)

Connector operation	Parameter	Description
createActionDiscoveryExportEdrmxml	action_name	Name of the action
	description	Description of the action
	targetset_id	Target set ID
	destination_path	Destination path
	data_object_limit	Data object limit
	copy_sharepoint_userprofile_notes	Boolean value to copy SharePoint user profile notes
	copy_sharepoint_userprofile_libraries	Boolean value to copy SharePoint user profile libraries
	copy_sharepoint_userprofile_posts	Boolean value to copy SharePoint user profile posts
	copy_sharepoint_userprofile_wikis	Boolean value to copy SharePoint user profile wikis
	copy_sharepoint_userprofile_misc	Boolean value to copy SharePoint user profile misc
	unique_document_id_prefix	Unique document ID prefix
	unique_document_id_digits	Unique document ID digits
	unique_document_id_pad_digits	Boolean value to unique document ID pad digits
	save_text_copy	Boolean value to save the text copy
	email_item_disposition	Email item disposition
createActionExportExceptions	action_name	Name of the action
	description	Description of the action
	export_format	Export format
createActionModifyAttribute	action_name	Name of the action
	description	Description of the action
	attribute	Modification type
createActionMovePlain	action_name	Name of the action
	description	Description of the action
	targetset_id	ID of the target set
	destination_path	Destination path
	recreate_directory_structure	Boolean value to re-create a directory structure
createActionMoveRetention	action_name	Name of the action
	description	Description of the action
	targetset_id	ID of the target set
	retention_type	Type of the retention
	retention_value	Retention value

Table 35. IBM StoredIQ Connector operations (continued)

Connector operation	Parameter	Description
createActionWatsonCuration	action_name	Name of the action
	description	Description of the action
	targetset_id	Target set ID
	destination_path	Destination path
	collection_id	Collection ID
	copy_sharepoint_userprofile_notes	Boolean value to copy SharePoint user profile notes
	copy_sharepoint_userprofile_libraries	Boolean value to copy SharePoint user profile libraries
	copy_sharepoint_userprofile_posts	Boolean value to copy SharePoint user profile posts
	copy_sharepoint_userprofile_wikis	Boolean value to copy SharePoint user profile wikis
	copy_sharepoint_userprofile_misc	Boolean value to copy SharePoint user profile misc
	compute_hash_value	Boolean value to compute hash value
	recreate_directory_structure	Boolean value to recreate directory structure
	no_auto_harvest	Boolean value to indicate that there is no auto harvest
createDatasourceBox **Note:** The Box data source creation that uses a Mule script works only for Box OAuth without redirection and two-factor authentication. Before you add a Box data source, see "Box volumes: configuration note" on page 20.	datasource_name	Name of the data source
	dataserver_id	ID of the data server
	box_server	Box server IP/hostname (default=`api.box.com`)
	box_include_users	Box server option to include users in a regular expression
	box_username	Box OAuth user name
	box_password	Box OAuth password
	index_options_container_metadata	Include metadata for contained objects
	index_options_full_text_content	Include content tagging and full-text index
createDatasourceCifs	datasource_name	Name of the data source
	dataserver_id	ID of data server
	datasource_type	Type of the data source
	cifs_server	CIFS server IP or host name
	cifs_username	CIFS server user name
	cifs_password	CIFS server password
	cifs_shared	CIFS server share directory
	index_options_container_metadata	Include metadata for contained objects
	index_options_full_text_content	Include content tagging and full-text index

Table 35. IBM StoredIQ Connector operations (continued)

Connector operation	Parameter	Description
createDatasourceCm8	datasource_name	Name of the data source
	dataserver_id	ID of data server
	cm8_server	CM8 server IP or host name
	cm8_port	CM8 server port
	cm8_username	CM8 server user name
	cm8_password	CM8 server password
	cm8_repository	CM8 server repository
	cm8_dbtype	CM8 server remote DB type
	cm8_remotedb	CM8 server remote DB
	cm8_schema_name	CM8 server schema name
	cm8_connection_string	CM8 server connection string
	cm8_harvest_itemtypes	Item types to be harvested, separated by comma
	cm8_copy_to_itemtype	Copy to the item type: **`SiqDocument`** or leave empty
	index_options_container_metadata	Include metadata for contained objects
	index_options_full_text_content	Include content tagging and full-text index
createDatasourceCmis	datasource_name	Name of the data source
	dataserver_id	ID of the data server
	cmis_server	CMIS server IP or host name
	cmis_port	CMIS server port
	cmis_username	CMIS server user name
	cmis_password	CMIS server password
	cmis_use_ssl	Boolean value to indicate ssl or not
	cmis_service	CMIS server service
	cmis_repository	CMIS server repository
	index_options_container_metadata	Include metadata for contained objects
	index_options_full_text_content	Include content tagging and full-text index

Table 35. IBM StoredIQ Connector operations (continued)

Connector operation	Parameter	Description
createDatasourceDocumentum **Note:** Before you add a Documentum data source, see the Documentum data source setup procedures in the Installing Documentum client jars to the data server and Adding a Documentum server as a data source topics.	datasource_name	Name of the data source
	dataserver_id	ID of the data server
	documentum_docbase	Documentum server document base
	documentum_username	Documentum server user name
	documentum_password	Documentum server password
	documentum_include_versions	Documentum server flag to indicate whether to harvest all document versions
	index_options_container_metadata	Include metadata for contained objects
	index_options_full_text_content	Include content tagging and full-text index
createDatasourceExchange20002003	datasource_name	Name of the data source
	dataserver_id	ID of the data server
	exchange_server	Exchange server host name
	exchange_username	Exchange server user name
	exchange_password	Exchange server password
	exchange_mailbox_server	Exchange server mailbox server
	exchange_active_directory	Exchange server active directory server
	exchange_use_ssl	Exchange server use SSL flag
	exchange_folder	Exchange server folder
	exchange_virtual_root	Exchange server virtual root
	index_options_container_metadata	Include metadata for contained objects
	index_options_full_text_content	Include content tagging and full-text index
createDatasourceExchange2007	datasource_name	Name of the data source
	dataserver_id	ID of the data server
	exchange_server	Exchange server host name
	exchange_username	Exchange server user name
	exchange_password	Exchange server password
	exchange_mailbox_server	Exchange server mailbox server
	exchange_active_directory	Exchange server active directory server
	exchange_use_ssl	Exchange server use SSL flag
	exchange_folder	Exchange server folder
	exchange_virtual_root	Exchange server virtual root
	index_options_container_metadata	Include metadata for contained objects
	index_options_full_text_content	Include content tagging and full-text index

Table 35. IBM StoredIQ Connector operations (continued)

Connector operation	Parameter	Description
createDatasourceExchange20102013	datasource_name	Name of the data source
	dataserver_id	ID of the data server
	exchange_server	Exchange server host name
	exchange_username	Exchange server user name
	exchange_password	Exchange server password
	exchange_mailbox_server	Exchange server mailbox server
	exchange_active_directory	Exchange server active directory server
	exchange_use_ssl	Exchange server use SSL flag
	exchange_folder	Exchange server folder
	exchange_virtual_root	Exchange server virtual root
	exchange_personal_archive	Exchange server personal archive
	index_options_container_metadata	Include metadata for contained objects
	index_options_full_text_content	Include content tagging and full-text index
createDatasourceExchangeonline	datasource_name	Name of the data source
	dataserver_id	ID of the data server
	exchange_server	Exchange server host name
	exchange_username	Exchange server user name
	exchange_password	Exchange server password
	exchange_folder	Exchange server folder
	exchange_virtual_root	Exchange server virtual root
	exchange_personal_archive	Exchange server personal archive
	index_options_container_metadata	Include metadata for contained objects
	index_options_full_text_content	Include content tagging and full-text index

Table 35. IBM StoredIQ Connector operations (continued)

Connector operation	Parameter	Description
createDatasourceFilenet	datasource_name	Name of the data source
	dataserver_id	ID of the data server
	filenet_server	FileNet server IP or host name
	filenet_port	FileNet server port
	filenet_username	FileNet server user name
	filenet_password	FileNet server password
	filenet_connectionType	FileNet connection type
	filenet_path	FileNet server path
	filenet_stanza	FileNet server stanza
	filenet_object_store	FileNet server object store
	filenet_wheresql	FileNet sql where clause
	filenet_domain	FileNet server domain
	filenet_create_document	FileNet server create document flag
	index_options_container_metadata	Include metadata for contained objects
	index_options_full_text_content	Include content tagging and full-text index
createDatasourceIBMConnections	connections_server	IBM Connections server IP or host name
	connections_username	IBM Connections server user name
	connections_password	IBM Connections server password
	connections_classname	IBM Connections class name
	connections_repository	IBM Connections repository
	connections_optionstring	IBM Connections additional options
createDatasourceLivelink	datasource_name	Name of the data source
	dataserver_id	ID of the data server
	livelink_server	Livelink server IP or host name
	livelink_port	Livelink server port
	livelink_username	Livelink server user name
	livelink_password	Livelink server password
	livelink_database	Livelink server database
	livelink_search_slice	Livelink server search slice
	index_options_container_metadata	Include metadata for contained objects
	index_options_full_text_content	Include content tagging and full-text index

Table 35. IBM StoredIQ Connector operations (continued)

Connector operation	Parameter	Description
createDatasourceNewsgator	datasource_name	Name of the data source
	dataserver_id	ID of the data server
	newsgator_server	NewsGator server host name
	newsgator_username	NewsGator server user name
	newsgator_password	NewsGator server password
	newsgator_use_ssl	Boolean value to indicate ssl or not
	index_options_container_metadata	Include metadata for contained objects
	index_options_full_text_content	Include content tagging and full-text index
createDatasourceNfs	datasource_name	Name of the data source
	dataserver_id	ID of the data server
	datasource_type	Type of the data source
	nfs_server	NFS server IP or host name
	nfs_export	NFS server export directory
	index_options_container_metadata	Include metadata for contained objects
	index_options_full_text_content	Include content tagging and full-text index
createDatasourceSharepoint	datasource_name	Name of the data source
	dataserver_id	ID of the data server
	sharepoint_server	SharePoint server host name
	sharepoint_port	SharePoint server port
	sharepoint_username	SharePoint server user name
	sharepoint_password	SharePoint server password
	sharepoint_use_ssl	SharePoint server use SSL flag
	sharepoint_active_directory_server	SharePoint server active directory server
	sharepoint_site_url	SharePoint server site URL
	sharepoint_version_type	SharePoint version
	index_options_container_metadata	Include metadata for contained objects
	index_options_full_text_content	Include content tagging and full-text index
createFilter	filter_name	Name of the filter
	description	Description of the filter
	expression	Expression of the filter

Table 35. IBM StoredIQ Connector operations (continued)

Connector operation	Parameter	Description
createSystemInfoset	infoset_name	Name of the infoset
	description	Description of the infoset
	datasource_ids	List of data source IDs separated by commas
	access_type	Access type of infoset, either public or private
	user_list	List of user IDs that can be accessed to the infoset
	waitforCompletion	Flag to indicate sync and async
createTargetset	targetset_name	Name of the target set
	description	Description of the target set
	type	Type of the target set
	datasource_ids	Data source IDs separated by comma
createUserInfosetByOperations	operations	Infoset operations
	infoset_name	Name of the infoset
	description	Description of the infoset
	sources	List of source infoset IDs separated by commas
	overlays	Overlay filter IDs separated by commas
	notify	Notification email
	generate_datamap	Flag to indicate whether a datamap is generated
	waitforCompletion	Flag to indicate sync or async
createUserInfosetFromExceptions	infoset_name	Name of the infoset
	description	Description of the infoset
	infoset_id	ID of the infoset source
	event_ids	Event IDs separated by commas
	categories	Category names that are separated by commas
	overlays	Overlay filter IDs separated by commas
	notify	Notification email
	generate_datamap	Flag to indicate whether a datamap is generated
	waitforCompletion	Flag to indicate sync or async

Table 35. IBM StoredIQ Connector operations (continued)

Connector operation	Parameter	Description
createUserInfosetFromFilter	infoset_name	Name of the infoset
	description	Description of the infoset
	infoset_id	ID of the infoset source
	filter_id	ID of the filter
	overlays	Overlay filter IDs separated by comma
	notify	Notification mail
	generate_datamap	Flag to indicate whether a datamap is generated
	waitforCompletion	Flag to indicate sync or async
executeAction	execution_name	Name of the action
	user_infoset_id	ID of the user infoset
	action_id	ID of the action
	waitforCompletion	Boolean value to indicate sync or async
generateReport	report_name	Name of the report
	user_infoset_id	ID of the user infoset
	action_id	ID of the report
	terms	List of terms that are separated by commas for Term Hit Report only
	filter_ids	List of filter IDs that are separated by commas for Term Hit Report only
	execution_id	ID of execution for Audit Summary Report only
	notify	Notification email
	waitforCompletion	Boolean value to indicate sync or async
harvestDatasource	harvest_name	Name of harvest
	datasource_id	ID of the data source
	harvest_type	Type of harvest
	load_type	Type of load
	schedule	Harvest schedule
	waitforCompletion	Boolean value to indicate sync or async

Table 35. IBM StoredIQ Connector operations (continued)

Connector operation	Parameter	Description
search	resourceType	Type of the resource
	id	ID of the resource
	nameOptions	Name search options
	resource_name	Name of the resource
	status	Status of the resource
	max	Maximum number of the result to be returned
	delete	Flag to indicate whether the resource is to be deleted

Notices

This information was developed for products and services offered in the US. This material might 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, MD-NC119
Armonk, NY 10504-1785
US

For license inquiries regarding double-byte character set (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 websites are provided for convenience only and do not in any manner serve as an endorsement of those

websites. The materials at those websites are not part of the materials for this IBM product and use of those websites is at your own risk.

IBM may use or distribute any of the information you provide 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.

The client examples cited are presented for illustrative purposes only. Actual performance results may vary depending on specific configurations and operating conditions.

The performance data and client examples cited are presented for illustrative purposes only. Actual performance results may vary depending on specific configurations and operating conditions.

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-IBMproducts. 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.

All IBM prices shown are IBM's suggested retail prices, are current and are subject to change without notice. Dealer prices may vary.

This information is for planning purposes only. The information herein is subject to change before the products described become available.

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 actual people or business enterprises 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.

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.

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.

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.

Index

Printed in USA

SC27-5688-15