

IBM StoredIQ Platform

Overview Guide



Note

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

This edition applies to Version 7.6.0.16 of product number 5724M86 and to all subsequent releases and modifications until otherwise indicated in new editions.

© **Copyright International Business Machines Corporation 2001, 2018.**

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

Contents

IBM StoredIQ Platform product library.....	iv
Contacting IBM StoredIQ customer support.....	v
Contacting IBM.....	v
Overview of IBM StoredIQ Platform.....	1
Component solutions.....	2
Applications of IBM StoredIQ Platform.....	3
IBM StoredIQ Platform Data Server.....	3
IBM StoredIQ Administrator.....	4
IBM StoredIQ Data Workbench.....	5
IBM StoredIQ eDiscovery.....	7
IBM StoredIQ Policy Manager.....	8
IBM StoredIQ Desktop Data Collector.....	8
User roles of IBM StoredIQ Platform.....	10
Key terms.....	11
Volumes.....	11
IBM StoredIQ Platform index.....	11
Information set.....	11
Filter.....	12
Overlay.....	12
Set Ops.....	12
Node ops.....	12
Duplicate operation.....	13
Enhancement.....	13
Action.....	13
Report.....	13
Exceptions.....	14
Getting started with the IBM StoredIQ Platform applications.....	15
Notices.....	16
Privacy policy considerations	17
Terms and conditions for product documentation.....	18
Trademarks.....	18
Index.....	20

IBM StoredIQ Platform product library

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

- *IBM StoredIQ Platform Deployment and Configuration Guide*
- *IBM StoredIQ Platform Overview Guide*
- *IBM StoredIQ Platform Data Server Administration Guide*
- *IBM StoredIQ Administrator Administration Guide*
- *IBM StoredIQ Data Workbench User Guide*
- *IBM StoredIQ eDiscovery User Guide*
- *IBM StoredIQ Policy Manager User Guide*

Contacting IBM StoredIQ customer support

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

- 1-866-227-2068

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

IBM Knowledge Center

The IBM StoredIQ publications can be found from [IBM Knowledge Center](#).

Contacting IBM

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

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

Overview of IBM StoredIQ Platform

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

Powerful solutions for managing unstructured data in-place

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

Unstructured Data Identification and Management delivers:

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

An organized, systemic, and defensible approach to eDiscovery

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

eDiscovery Identification and Collection delivers:

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

Information governance to automate policy and compliance across unstructured data

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

Policy Assessment and Compliance provides the following features and solutions:

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

Flexible solution for identifying and collecting data from remote devices

IBM® Desktop Data Collector enables organizations to apply corporate governance policies to user desktops and notebooks. Users can identify and collect corporate records or custodian data for legal matters.

Desktop Data Collector delivers:

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

Bridging structured and unstructured content for enterprise information governance

A unified governance architecture is the key to governing all enterprise content, be it structured or unstructured, on premises or in the cloud, thus helping organizations manage the findability, usability, and integrity of their data.

Integrating IBM StoredIQ with Information Governance Catalog bridges structured and unstructured content to enable enterprise information governance.

Component solutions

IBM StoredIQ provides three components of the solutions: the gateway, data servers, and application stack (AppStack).

The Gateway

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

Data Servers

The IBM StoredIQ Data Server helps understand the data landscape of the enterprise. It obtains the data from supported data sources and indexes it. By indexing this data, you gain information about unstructured data such as file size, file data types, file owners.

The data servers push the information about volumes and indexes to the gateway so it can be communicated to the application stack. Multiple data servers feed into a single gateway. In addition to an administrator user interface, administrators can deploy the IBM StoredIQ Desktop Data Collector and index desktops from the data server.

Application stack

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

The synchronization feature for integration with IBM Information Governance Catalog is also part of the application stack.

Applications of IBM StoredIQ Platform

IBM StoredIQ provides interface applications that help fulfill its solution goals.

IBM StoredIQ Platform Data Server

IBM StoredIQ Platform Data Server user interface provides access to data server functionality. It allows administrators to view the dashboard and see the status of the jobs and system details. Administrators can manage information about servers and conduct various configurations on the system and application settings.

The screenshot displays the IBM StoredIQ Platform Data Server user interface. The top navigation bar includes the IBM logo, a 'DS Admin' button, and tabs for 'Administration', 'Folders', and 'Audit'. Below this is a secondary bar with 'Dashboard', 'Data sources', and 'Configuration'. The main content area is divided into several sections:

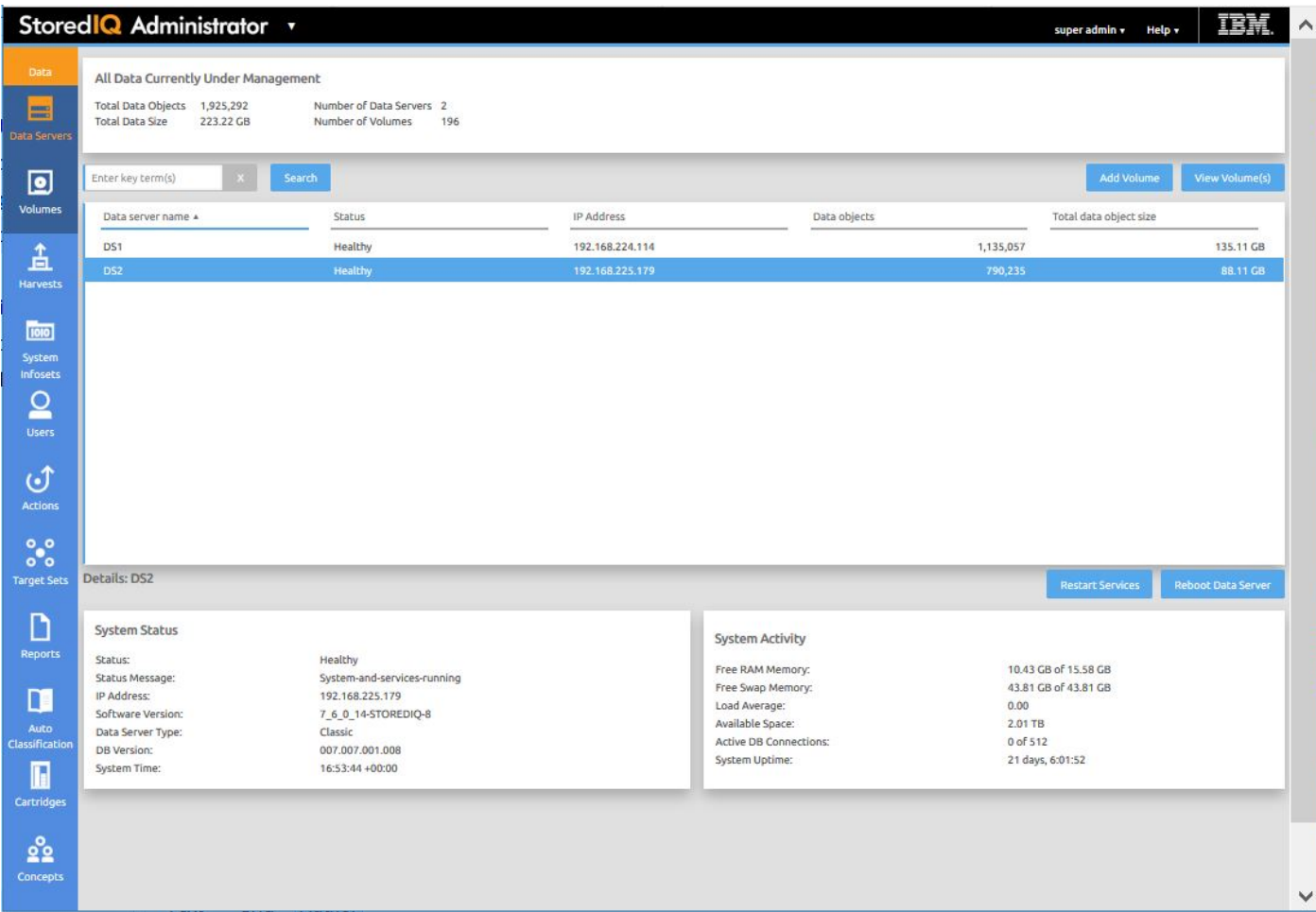
- Page refresh:** Off | 30 sec | 60 sec | 90 sec
- Today's job schedule:** No jobs scheduled for today.
- Jobs in progress:** No jobs are currently running.
- System summary:** View a summary of system details.

Total system data objects	10756
Total contained data objects	1081591
Total data objects	1092347
Number of volumes	13
Date of last completed harvest	No harvests run.
- Harvest statistics:** Review the performance over the last hour for all harvests.

Processes	4
Average data objects per second	0.0
Average data object size	0 bytes
Maximum data object size	0 bytes
Average data object processing time	0.0 sec
Maximum data object processing time	0.0 sec
- Event log:** The current event log as of 03/21/2018 05:03 PM. Includes links for 'Clear this view', 'Download today's event log', and 'View all event logs'. A list of 'Last 500 events' is shown with timestamps and messages, each with a 'Subscribe' link.
- Appliance status:** Shows a green status indicator and a 'Controller' link. Below it are links for 'About appliance' and 'View cache details'.

IBM StoredIQ Administrator

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



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

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

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

If IBM StoredIQ is configured for integration with Information Governance Catalog, the Administrator can also manage which volumes are published to the governance catalog.

- **Scheduling harvests:** Harvesting, which can also be referred to as indexing, is the process or task by which IBM StoredIQ Platform examines and classifies data in your network. Using IBM StoredIQ Administrator, harvests can be scheduled, edited, and deleted.

- **Creating system infosets:** System infosets that use only specific indexed volumes can be created and managed within IBM StoredIQ Administrator. Although infosets are a core component of IBM StoredIQ Data Workbench, system infosets are created as a shortcut for users in IBM StoredIQ Administrator.
- **Managing users:** The user management area allows administrators to create users and manage users' access to the various IBM StoredIQ applications.
- **Configuring and managing actions:** An action is any process that is taken upon the data that is represented by the indexes. Actions are run by data servers on indexed data objects. Any errors or warnings that are generated as a result of an action are recorded as exceptions in IBM StoredIQ Data Workbench.

Note: Actions can be created within IBM StoredIQ Administrator and then made available to other IBM StoredIQ applications such as IBM StoredIQ Data Workbench.

- **Managing target sets:** Provides an interface that allows the user to set the wanted targets for specific actions that require a destination volume for their actions.
- **Reports:** 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.
- **Auto-classification:** Automated document categorization, what IBM StoredIQ refers to as auto-classification models, integrates the IBM® Content Classification's classification model into the IBM StoredIQ Platform infoset-generation process. Data Experts can use IBM Content Classification to train a classification model, which 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. Metadata can be used in rule-based filters to create new infosets.
- **Cartridges:** Cartridges are compressed files that contain analysis logic. When you add 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.

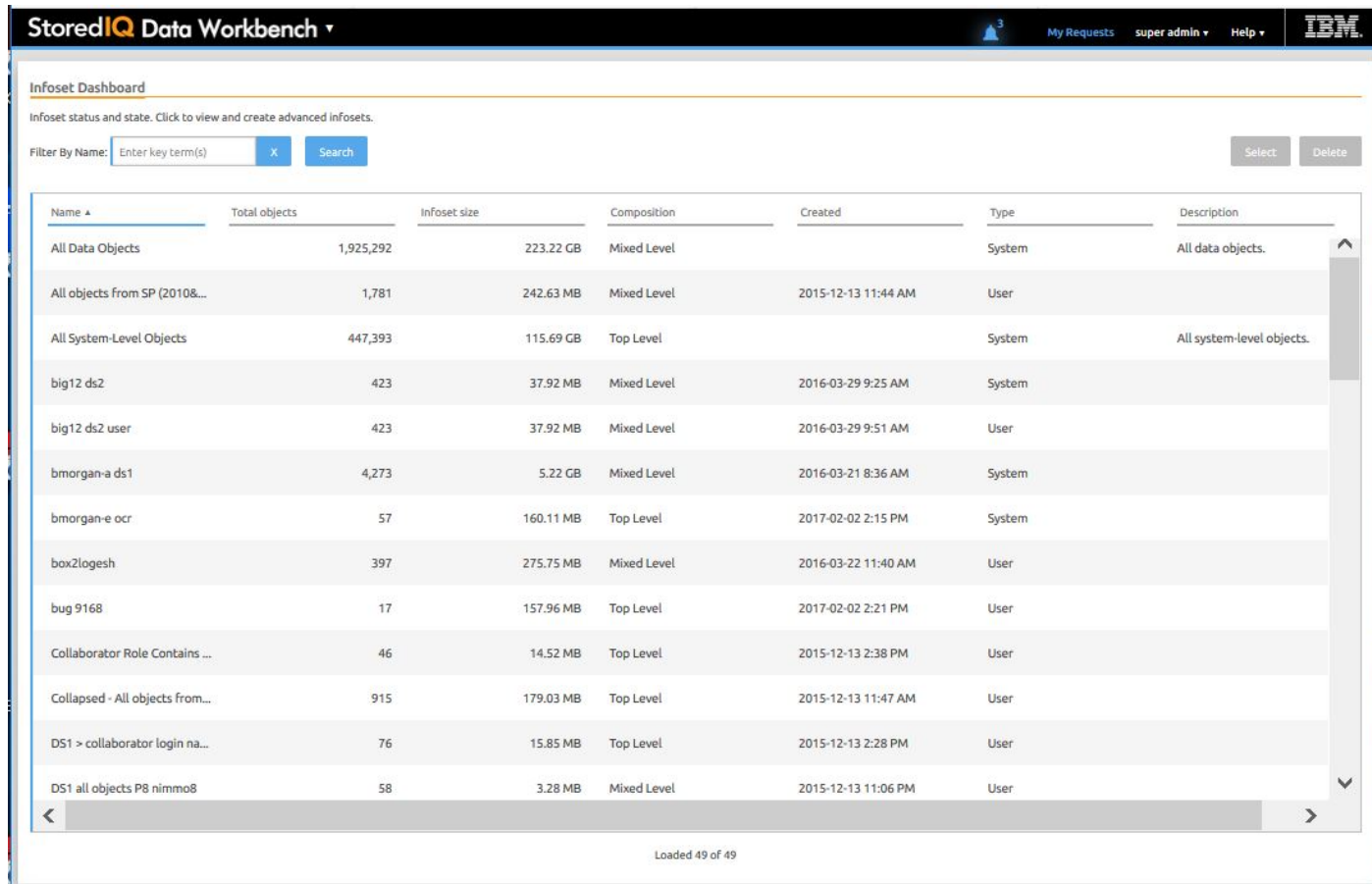
To apply the analysis logic contained in the cartridge, you must run a Step-up Analytics action that uses the cartridge on an infoset. IBM StoredIQ examines all documents in the infoset, applies the analytics, and then stores the analysis results in the IBM StoredIQ index.

- **Managing concepts:** Provides the ability to 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 using the automation workflow.
- **DataSeter - Classic:** Data servers can be categorized in two types: DataSeter - Classic and DataSeter - Distributed. DataSeter - Classic refers to the regular data servers. It uses either the current PostgreSQL or Lucene index as an index.
- **DataSeter - Distributed:** The distributed data server uses an Elasticsearch cluster instead of an embedded Postgres database. It increases the scalability and flexibility of the IBM StoredIQ deployment in a way that it can manage much larger amounts of data. 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. Search queries perform better on DataSeter - Distributed.
- **Connector API SDK:** A connector is a software component of IBM StoredIQ that is used to connect to a data source such as a network file system and access its data. Using IBM StoredIQ Connector API SDK, developers of other companies can develop connectors to new data sources outside the IBM StoredIQ development environment. These connectors can be integrated with a live IBM StoredIQ application to index, search, manage, and analyze data on the data source.

IBM StoredIQ Data Workbench

Big data is a pervasive problem, not a one-time occurrence. It is easy for most companies to realize that big data is problematic, but it is hard to identify what problems they have. Big data is all about the

unknown, but the unknown cannot be off limits. IBM StoredIQ Data Workbench can help you learn about your data, make educated decisions with your most valuable asset, and turn your company's most dangerous risk into its most valuable asset.



StoredIQ Data Workbench

Infoset Dashboard

Infoset status and state. Click to view and create advanced infosets.

Filter By Name: X Search Select Delete

Name	Total objects	Infoset size	Composition	Created	Type	Description
All Data Objects	1,925,292	223.22 GB	Mixed Level		System	All data objects.
All objects from SP (20108...	1,781	242.63 MB	Mixed Level	2015-12-13 11:44 AM	User	
All System-Level Objects	447,393	115.69 GB	Top Level		System	All system-level objects.
big12 ds2	423	37.92 MB	Mixed Level	2016-03-29 9:25 AM	System	
big12 ds2 user	423	37.92 MB	Mixed Level	2016-03-29 9:51 AM	User	
bmorgan-a ds1	4,273	5.22 GB	Mixed Level	2016-03-21 8:36 AM	System	
bmorgan-e ocr	57	160.11 MB	Top Level	2017-02-02 2:15 PM	System	
box2logesh	397	275.75 MB	Mixed Level	2016-03-22 11:40 AM	User	
bug 9168	17	157.96 MB	Top Level	2017-02-02 2:21 PM	User	
Collaborator Role Contains ...	46	14.52 MB	Top Level	2015-12-13 2:38 PM	User	
Collapsed - All objects from...	915	179.03 MB	Top Level	2015-12-13 11:47 AM	User	
DS1 > collaborator login na...	76	15.85 MB	Top Level	2015-12-13 2:28 PM	User	
DS1 all objects P8 nimmo8	58	3.28 MB	Mixed Level	2015-12-13 11:06 PM	User	

Loaded 49 of 49

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

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

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

you can analyze and classify the data governed by IBM StoredIQ based on the data classes that are synchronized from the governance catalog.

IBM StoredIQ eDiscovery

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

StoredIQ eDiscovery super admin Help IBM

Matter Dashboard

View Matter status. Click to view details.

Enter key term(s) X Search Create Matter

Matter Name	Contents	Contacts
Company B's tax case Created: 2018-03-21 10:44 AM	Boxes: 0 Reports: 0	Data Expert: super admin Legal Ops: super admin
John's inheritance case Created: 2018-03-21 10:39 AM	Boxes: 0 Reports: 0	Data Expert: super admin Legal Ops: super admin
Company A Created: 2018-03-21 10:37 AM	Boxes: 0 Reports: 0	Data Expert: super admin Legal Ops: super admin
测试 Created: 2015-09-26 11:20 AM	Boxes: 1 Reports: 0	Data Expert: super admin Legal Ops: super admin

Loaded 4 of 4

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

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

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

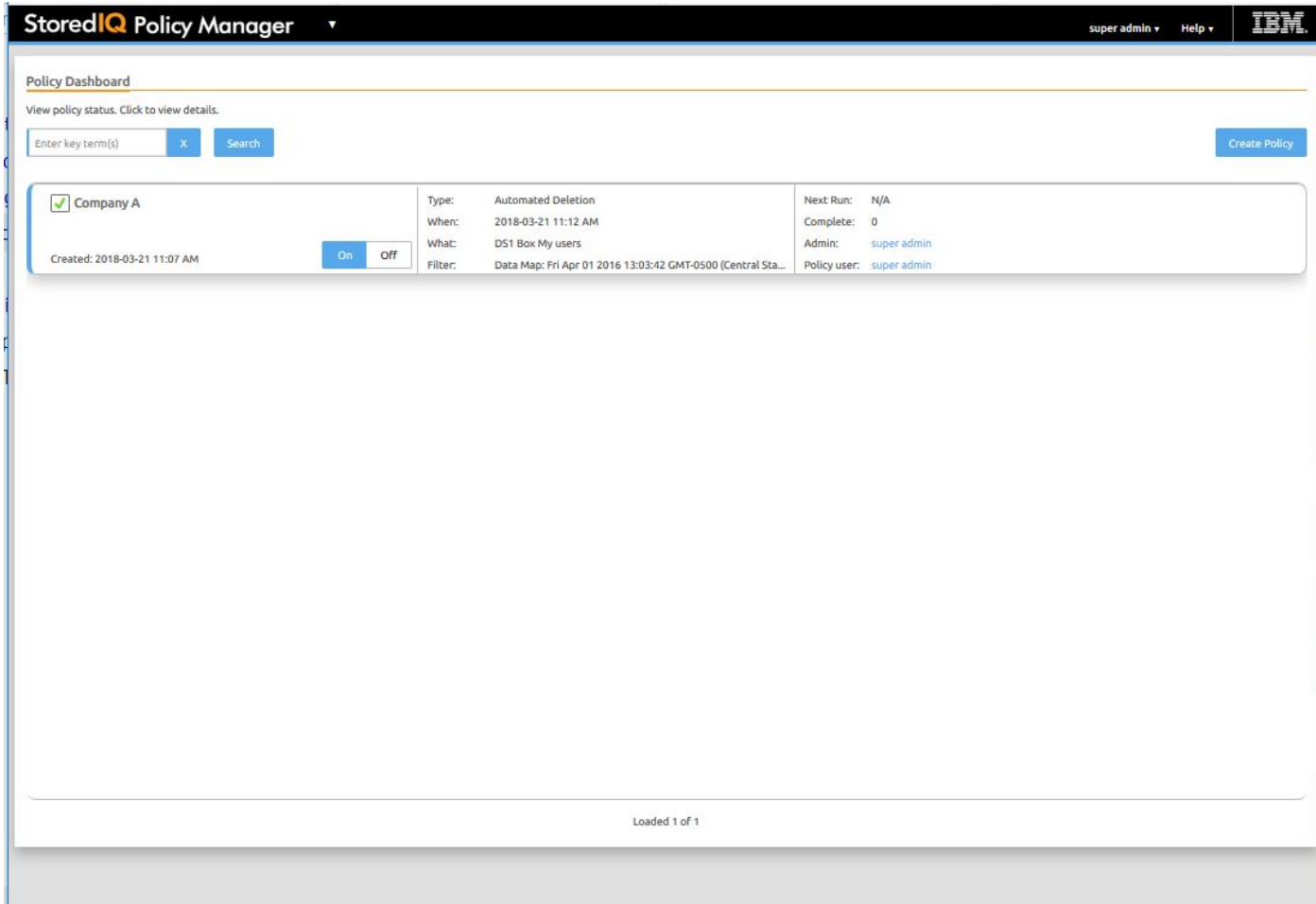
Within IBM StoredIQ eDiscovery, the user is a non-technical user who needs to obtain certain pieces of data to complete a legal process. The IBM StoredIQ eDiscovery user knows the information or parameters for what kinds of data they want, but does not always know how to obtain that data.

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

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

IBM StoredIQ Policy Manager

IBM StoredIQ Policy Manager allows users to run mature policies and processes at scale across a wider range of data.



The users can define and run systemwide policies, focusing on the execution of the process rather than understanding or reviewing affected data objects. Additionally, with reports of IBM StoredIQ Policy Manager, you can record what actions were conducted, when they were conducted, and what data was affected by the policy's execution.

IBM StoredIQ Desktop Data Collector

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

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

- Lightweight client deployment with no browser plug-ins and no client-side UI installations. All major browsers, such as Internet Explorer and Firefox, are supported natively.

- One (common) index that is shared across all use cases. This index supports a wide variety of data sources and is used by application dashboards for many use cases.
- No agents are placed on data sources. The native API/protocol is used whenever it is possible. Third-party or custom connectors are used only when necessary.

Note: Agents are used for desktop collection.

User roles of IBM StoredIQ Platform

The IBM StoredIQ Platform provides applications and interfaces for four user types.

System administrator

This person is responsible for IBM StoredIQ Platform installation setup, system and network configuration and maintenance, and administration activities. These activities are required to be done before other users of the IBM StoredIQ applications can start their work. The administrative activities mainly include:

- Getting the data servers and data centers ready for use.
- Adding volumes and ensuring security of data and data source.
- Harvesting volumes and generating system info sets.
- Managing users, actions, target sets and concepts.
- Managing cartridges.
- Creating reports and using auto-classification models.
- Creating Mule scripts.
- Setting up the integration with Information Governance Catalog.

For more information about what system administrators do, see [Administering IBM StoredIQ Platform Data Server](#) and [Administering IBM StoredIQ Administrator](#).

Data expert

A data expert understands both business processes and technical implementation. This person is responsible for responding to requests from the business user in a timely fashion, managing data in the IBM StoredIQ applications, such as IBM StoredIQ Data Workbench. This person also decides which data is published to the governance catalog.

For more information about what a data expert does, see [Managing IBM StoredIQ Data Workbench](#).

Legal user

The legal user is an end user who is looking to complete some business processes. This person is responsible for managing matters, such as filtering, adding, deleting matters, configuring matters, or creating and managing boxes. The legal user's work is dependent upon the system administrator and data expert. The system administrator sets up the IBM StoredIQ Platform environment and authorizes the right and scale of work that the legal user can do with the IBM StoredIQ eDiscovery application. The data expert controls and manages the data in the legal system.

For more information about what a legal user does, see [Managing IBM StoredIQ eDiscovery](#).

Policy user

The policy user is responsible for creating and running a set of pre-determined steps to automatically execute against data that is being managed by the IBM StoredIQ solution. The policy user uses the IBM StoredIQ Policy Manager application to create new policy runs. Each policy run is based off pre-configured scripts that were deployed to IBM StoredIQ Policy Manager.

For more information about what a policy user does, see [Managing IBM StoredIQ Policy Manager](#).

Key terms

The following terms are key to understanding IBM StoredIQ as a whole.

Volumes

A volume represents a data source or destination that is available on the network to IBM StoredIQ.

Within IBM StoredIQ, these volume types exist:

Primary

The storage where the unstructured content resides (*data source*) and is harvested from.

Retention

Storage for information to be retained for a set amount of time or for litigation hold. Typically, a retention volume is immutable.

Export

Storage to keep the data produced from a policy so that it can be exported as a load file and uploaded into a legal review tool. Administrators can also configure export volumes for managing harvest results from cycles of a discovery export policy.

System

Storage for data that is used for application specific purposes.

IBM StoredIQ Platform index

Based on the volumes that are added in the data servers, indexes are generated through IBM StoredIQ Administrator or IBM StoredIQ Data Server to examine, classify, and map data in the network. They are used to search for data and find out what and how much data you have in your system.

There are two types of indexes: metadata index and full-text index.

Metadata index contains all the information about the data at a specific location on a network. It includes descriptive information or attributes about the data such as a file name, file size, created date, and owner.

Full-text index is a more detailed index on the contents of the data itself. By reading the contents of the data, the words or characters that are contained within the data can be referred to and searched against.

Information set

An information set, abbreviated as an info set, is the core concept in using the IBM StoredIQ applications. It is created and used to collect specific data to manage the business system.

There are two types of info sets: system info sets and user info sets.

System info set

These info sets are generated after volumes are harvested from IBM StoredIQ Administrator or IBM StoredIQ Data Server. They can be viewed from the IBM StoredIQ Data Workbench user interface but users cannot edit or delete them. They can also be manually created by the administrator to target certain volumes in the IBM StoredIQ Administrator application. These system info sets can be edited or deleted by the system administrator. System info sets must be generated or created before any user info sets can be created.

User info set

A user info set is created out of an existing system info set by a user. It is defined to contain specific data that a user needs to operate upon the system. For example, you can create a user info set that contains a person's emails in Year 2000. Then, you can act on the data within this info set: to move, copy, or delete it from your system.

Data Map

Data Map is the visualization of an infoset. It provides a visual layout of the data and in-depth information about data source types, data categories, size or amounts, the number of data objects and details. For more information, see [Data maps](#).

Filter

A filter is created upon the available information that was populated by the index. It is used to classify or refine the existing infoset to create a new infoset.

A filter can have multiple attributes. You can apply several attributes of a filter to one infoset to create a new infoset that consists of the data that you need.

Example of filtering an infoset

A system infoset contains all files, emails, and documents of all company employees. You need to retrieve some specific data about Josh Smith: Josh's emails with a subject of stock option and Josh's files that are larger than 1 GB. To get these two sets of data, you can use the filter to refine the system infoset into two user infosets.

To create the first infoset about Josh's emails with the subject of stock options, you need to take the following steps:

- Apply a name filter attribute to find Josh Smith.
- Apply a file filter attribute to find Josh's emails with the .MSG extension.
- Apply an email filter attribute to find Josh's emails with a subject of stock options.

To create the second infoset to get Josh's files of larger than 1 GB, you need to take the following steps:

- Apply a name filter attribute to find Josh Smith.
- Apply a filter attribute of size larger than 1 GB to all of Josh Smith's files that are larger than 1 GB.

Overlay

Overlays are configurable filters that display hits or matches in a selected infoset.

Within the data map, color intensifies for data objects that match the overlay change. The greater the overlay matches, the more red that tile appears within the data map.

Set Ops

Set Ops allows infosets to be combined in different ways to produce another infoset.

You can select a primary infoset and use Set Ops to combine one or more infosets to create a union, intersection, symmetric difference, or subtraction infoset.

Node ops

Contained data, such as data within .ZIP, .TAR, or .PST files, is hierarchical and can have different relationship and connections with other data. Depending on how that data is viewed, that data can give a different perception than what is represented.

The **Node ops** pane helps you understand more about what data is represented by data sets. Within Node ops, you can conduct expansion or collapse operations.

- In an expand operation, all files within an infoset are expanded, so creation of an infoset can be more accurate.

- In a collapse operation, all opened or expanded files within an info set are collapsed, so an info set that is created can be small.

Note: If the files within the info set are not container files, then the **Expand** operator or **Collapse** operator has no effect.

Duplicate operation

With Duplicate operations, you can identify varieties of duplicate data in your system. Apply filters, operations, reports, or actions to a new duplicate identification info set to start the data deduplication process.

Duplicate operation compares objects of two info sets that are based on each other's hash value. If an object's hash value matches, the system can flag that object as a duplicate object.

Enhancement

An enhancement is a way of refining or distilling an info set. Enhancements are created as models within IBM StoredIQ Administrator. When you apply an enhancement to an info set, it updates that info set's index.

Action

An action is an activity that is created by an administrator and conducted on an info set. Available IBM StoredIQ Data Workbench actions include copy, copy to retention, delete, discovery export, modify attribute, move, Watson Curation, Step-up Snippet, Step-up Full-text, and Step-up Analytics.

Actions do not alter info sets. An info set is a grouping of data, not the data itself, and actions are applied to the actual objects, not the info set. When you copy, you copy the actual file, not the info set. The same is true for copying to retention or discovery export. Actions can be scheduled to run immediately or at a predetermined time and date.

Report

The reporting function provides external views of info sets and validates IBM StoredIQ processes. Reports can also be customized with the BIRT Report Designer.

You can share the information that is contained within info sets with the reporting component, which allows info sets to be transferred to other media types for review and analysis. These reports do not affect existing info sets, but provide you with more usable formats in which to understand the files and data that is captured by an info set.

Reporting is a key step within the data-management process as it validates that processes were completed correctly within IBM StoredIQ. You can customize reports in any of these scenarios:

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

IBM StoredIQ provides a number of preconfigured system reports, such as summaries of data objects in the system, storage use, and the number of identical documents in the system.

Exceptions

When you conduct an action and encounter errors, an exception list occurs. The list helps you trace and understand what errors are so that you can correct them.

Exceptions are presented with contextual details in three areas: Events, Types, and Exception objects.

Getting started with the IBM StoredIQ Platform applications

Follow these steps to log in to the applications.

The administrator must ensure that these tasks are completed before you can use any applications:

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

To use IBM StoredIQ Data Workbench, IBM StoredIQ eDiscovery, and IBM StoredIQ Policy Manager, more tasks must be completed by the administrator:

- Add volumes and create indexes through IBM StoredIQ Administrator or IBM StoredIQ Data Server.
- Harvest volumes and generate system info sets.
- Create users so to interact with the IBM StoredIQ products.
- Create and define actions for the user to manage data objects that are contained within info sets.

Note: IBM StoredIQ eDiscovery and IBM StoredIQ Policy Manager are dependent upon other IBM StoredIQ products such as IBM StoredIQ Administrator and IBM StoredIQ Data Workbench.

1. Open the IBM StoredIQ Platform user interface from a browser and enter your user name or email address and password. To log in for the first time after the deployment, use the default administrative account.
2. Click **Log in**.

If your user account is associated with more than one user role, you have access to several applications. To switch between the applications:

- a. Within the global navigation bar, click the white arrow next to the name of the application. A menu appears, listing the available applications.
- b. Select the application that you need to work with, for example, IBM StoredIQ Administrator.

Product legal 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-IBM products. Questions on the capabilities of non-IBM products should be addressed to the suppliers of those products.

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

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.

Each copy or any portion of these sample programs or any derivative work, must include a copyright notice as follows: © Copyright IBM Corp. 2004, 2015. All rights reserved.

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

A

action [5](#), [13](#)
appstack [2](#)

C

contained data [12](#)
customized reports [13](#)

D

data server [2](#)
Data Server dashboard [3](#)
Data Workbench
 about [6](#)
 potential uses of [6](#)
Desktop Agent [2](#)

E

eDiscovery
 potential uses of [7](#)
 prerequisites for using [8](#)
enhancement [13](#)
exception objects [14](#)
exceptions [5](#), [14](#)

F

filter [12](#)

G

gateway [2](#)
getting started with the IBM StoredIQ applications [15](#)

H

harvest [11](#)

I

IBM StoredIQ Administrator [4](#)
IBM StoredIQ applications [15](#)
IBM StoredIQ Data Workbench [5](#)
IBM StoredIQ Desktop Data Collector [8](#)
IBM StoredIQ eDiscovery [7](#)
IBM StoredIQ Platform Data Server [3](#)
IBM StoredIQ Platform index [11](#)
IBM StoredIQ Policy Manager [8](#)
index types
 full-text index [11](#)
 metadata index [11](#)
information set [11](#)

infoaset
 system infoaset [11](#)
 user infoaset [11](#)
intersection [12](#)

K

key terms
 action [11](#)
 filter [11](#)
 index [11](#)
 information set [11](#)
 infoaset [11](#)
 overlay [11](#)
 report [11](#)
 scope operation [11](#)
 set operation [11](#)

L

legal
 notices [16](#)
 trademarks [18](#)

N

Node ops
 Collapse [12](#)
 Expand [12](#)
notices
 legal [16](#)

O

overlay [12](#)

R

report
 report types [13](#)
reports [13](#)

S

Set Ops [12](#)
subtraction [12](#)
symmetric difference [12](#)

T

trademarks [18](#)

U

union [12](#)

