

IBM Spectrum Discover
Version 2.0.3

Administration Guide



Note

Before using this information and the product it supports, read the information in [“Notices” on page 115.](#)

Edition notice

This edition applies to version 2 release 0 modification 3 of the following product, and to all subsequent releases and modifications until otherwise indicated in new editions:

- IBM Spectrum Discover ordered through Passport Advantage (product number 5737-I32)
- IBM Spectrum Discover ordered through AAS/eConfig (product number 5641-SG1)

IBM® welcomes your comments; see the topic [“How to send your comments” on page xi.](#) When you send information to IBM, you grant IBM a nonexclusive right to use or distribute the information in any way it believes appropriate without incurring any obligation to you.

© **Copyright International Business Machines Corporation 2018, 2020.**

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

Contents

Figures.....	vii
Tables.....	ix
About this information.....	xi
Prerequisite and related information.....	xi
How to send your comments.....	xi
Summary of changes.....	xiii
Chapter 1. Managing user access.....	1
Initial login.....	2
Resetting the sdadmin password.....	2
Password policies.....	2
Changing password.....	4
Managing user accounts.....	5
Creating user accounts.....	6
Managing groups.....	6
Creating groups.....	7
Managing collections.....	8
Creating collections.....	9
Managing LDAP and IBM Cloud Object Storage System connections.....	10
Creating an LDAP connection.....	10
Creating an IBM Cloud Object Storage connection.....	12
Chapter 2. Managing metadata policies.....	15
Adding policies.....	15
Adding auto-tagging policy parameters.....	17
Adding deep-inspection policy parameters.....	18
Adding content search policy parameters.....	18
Running policies.....	19
Viewing policies.....	19
Viewing policy log files.....	21
Modifying policies.....	22
Deleting policies.....	23
Chapter 3. Using content search policies.....	25
Identifying the required regex expressions.....	25
Creating a content search policy.....	26
Viewing content search application logs.....	27
Hints and tips for using content search.....	27
Chapter 4. Tiering data by using ScaleILM application.....	29
Viewing ScaleILM application logs.....	31
Chapter 5. Exporting Metadata to IBM Watson Knowledge Catalog.....	33
Exporting metadata to IBM Cloud Watson Knowledge Catalog.....	33
Exporting metadata to IBM on-premises Watson Knowledge Catalog.....	34

Exporting metadata from linked and non-linked data sources.....	36
Exporting metadata from linked IBM Spectrum Discover data source	36
Exporting metadata from non-linked IBM Spectrum Discover data source	36
Mapping similar source connections in Watson Knowledge Catalog	37
Troubleshooting export issues	38
Authentication failure with Watson Knowledge Catalog - both on-Premise and IBM Cloud.....	39
Incorrect WKC URL configuration.....	39
Invalid connection type in catalog.....	39
No linked connection type.....	39
WKC connector pod in CrashLoopBackoff state.....	40
S3 connection issues.....	40
Chapter 6. Managing tags.....	41
Creating tags.....	41
Viewing and searching tags.....	43
Editing tags.....	43
Deleting tags.....	43
Chapter 7. Discover data.....	45
Searching.....	45
Grouping data by file type.....	51
Searching system and custom metadata fields.....	52
System metadata fields to search on.....	52
Access control list metadata to search on.....	54
Search on custom metadata fields.....	55
Examples of search filters.....	56
Search results table.....	56
Refine search results.....	58
Sort search results.....	58
Tag search results manually.....	58
Chapter 8. Managing applications.....	61
Chapter 9. Using the IBM Spectrum Discover application catalog.....	63
Creating your own applications to use in the IBM Spectrum Discover application catalog.....	65
Chapter 10. Backup and restore.....	67
Initial setup configuration.....	67
Running a backup.....	68
Running an automated backup.....	69
Running a restore.....	69
Chapter 11. Reports.....	71
Chapter 12. High availability for an MPP deployment.....	73
Reintegrating a failed data node into an IBM Db2 Warehouse MPP cluster.....	74
Chapter 13. Monitoring data sources.....	75
Viewing data source status.....	75
Viewing data source connections.....	77
Recommended to move.....	78
Deleting or editing a connection.....	79
Chapter 14. Monitoring the IBM Spectrum Discover environment.....	83
Monitoring the status of the IBM Spectrum Discover environment.....	83
Monitoring the IBM Spectrum Discover virtual machine.....	84

Audit log.....	84
Using the FFDC script.....	85
Chapter 15. Updating the network configuration.....	87
Chapter 16. Using a third-party data movement application to move or copy data.....	89
Preserving tags during data movement.....	90
Chapter 17. Disaster recovery procedures.....	93
Preparations for disaster recovery.....	93
Running disaster recovery.....	93
Chapter 18. Troubleshooting.....	97
Best practices.....	97
Changing system time breaks jobs and pods.....	97
Recovering original SSH keys.....	97
How to recover a system after a yum update.....	98
Configuration issues.....	98
Healthy default pod list.....	98
Data issues.....	98
Delete markers from IBM Cloud Object Storage are ignored.....	98
Records are not ingested after reboot.....	99
Recovering from data ingestion consumer or producer issues.....	99
IBM Spectrum Discover scan data not landing in database.....	101
Viewing live reports might list incorrect results.....	101
Error diagnosis issues.....	102
Ansible® Warnings.....	102
ens160 activation errors in /var/log/messages.....	102
kubectl returns "error: You must be logged in to the server".....	102
Installation issues.....	103
Db2 Warehouse installation port conflict - Wait for Db2wh to initialize.....	103
IBM Cloud Private install logs are missing.....	103
Networking issues.....	103
CentOS reboots under load.....	103
IBM Cloud Object Store will not connect to the IBM Spectrum Discover kafka server by IP address.....	104
IBM Spectrum Scale can fail to load after an ESXi server is rebooted.....	104
Multi-node network settings get stuck while checking the Docker run status	104
Network configuration update: Failure recovery steps.....	104
Network configuration update: Error creating metaocean tables with Liquibase.....	105
Network settings change hangs while uninstalling IBM Cloud Private.....	106
Network settings change fails when you run mmconfigappliance.....	106
Pod stuck in CreateContainer error.....	107
Pod stuck in terminating state.....	107
Performance issues.....	107
Changed permissions for the current user are not effective until logout.....	108
Blank queries to the search API time out.....	108
Policy issues.....	108
A collection policy cannot be added to a collection or edited after the collection is created.....	108
Tagging policy failures under high load.....	108
Security issues.....	109
Upgrading issues.....	109
How to recover DB2 Warehouse from an unrecoverable state during an upgrade.....	109
Debugging a hung upgrade.....	110
Upgrader not logged in when you create metrics services.....	110

Accessibility features for IBM Spectrum Discover.....	113
Accessibility features.....	113
Keyboard navigation.....	113
IBM and accessibility.....	113
Notices.....	115
Trademarks.....	116
Terms and conditions for product documentation.....	116
IBM Online Privacy Statement.....	117
Index.....	119

Figures

- 1. Select the Account Settings option.....4
- 2. Change Password fields..... 4
- 3. The Users tab.....5
- 4. The Create Local User window.....6
- 5. The Groups Tab..... 7
- 6. The Create Local Group window..... 7
- 7. The Collections tab.....9
- 8. Create a Collection..... 9
- 9. Create an LDAP connection..... 11
- 10. Create an IBM Cloud Object Storage Connection..... 13
- 11. Policies table..... 19
- 12. Policies table..... 20
- 13. Policies table..... 21
- 14. Modify a policy..... 22
- 15. Preinstalled regular expressions..... 26
- 16. Tags table..... 42
- 17. New Organizational Tags..... 42
- 18. Start a visual exploration..... 45
- 19. Tag values.....46
- 20. Search - add groups..... 47
- 21. Search Results..... 47
- 22. Search Results Filters..... 48
- 23. Generate Report.....49

24. Add tags.....	50
25. Find settings icon.....	50
26. Click the settings icon.....	51
27. View used capacity.....	51
28. Example to generate a report sorted by file type and data source.....	57
29. Example of a search sorted by time since access and size range	58
30. Applications table.....	61
31. Reports table.....	71
32. View Data Report.....	71
33. Steady state for HA group.....	73
34. HA group after head node failover.....	74
35. Datasource capacity.....	75
36. Example of the capacity that is being used.....	76
37. Run table refresh button in the Discover database window.....	77
38. Example of a data source capacity widget.....	78
39. Example of a screen that shows the TEMPERATURE tag.....	79
40. Example of an autotag policy to identify files and objects that have not been accessed in more than one year.....	79
41. Example of a listing of existing connections.....	80
42. Starting the process to delete a data source connection.....	80
43. Example of a screen that shows how to delete a connection.....	81
44. Example of a screen that shows how to edit a connection.....	81

Tables

1. IBM Spectrum Discover library information units.....	xi
2. Data locations and File States.....	29
3. Report generated for moving files.....	90

About this information

IBM Spectrum® Discover is a metadata-driven management system for large-scale file and object environments. IBM Spectrum Discover maintains a real-time metadata repository for large-scale enterprise storage environments. Metadata can be searched, enhanced, discovered, and leveraged for data processing by using built-in or custom agents.

IBM Spectrum Discover - Information units

Information unit	Type of information	Intended users
IBM Spectrum Discover: Concepts, Planning, and Deployment Guide	This information unit provides information about the following topics: <ul style="list-style-type: none">• Product Overview• Planning• Deploying and configuring	Users, system administrators, analysts, installers, planners, and programmers of IBM Spectrum Discover.
IBM Spectrum Discover: Administration Guide	This information unit provides information about administration, monitoring, and troubleshooting tasks.	Users, system administrators, analysts, installers, planners, and programmers of IBM Spectrum Discover.
IBM Spectrum Discover: REST API Guide	This information unit provides information about the following topics: <ul style="list-style-type: none">• IBM Spectrum Discover REST APIs• Endpoints for working with a DB2 warehouse• Endpoints for working with policy management• Endpoints for working with connection management• Action agent management using APIs• RBAC management using APIs	Users, system administrators, analysts, installers, planners, and programmers of IBM Spectrum Discover.

Prerequisite and related information

For updates to this information, see IBM Spectrum Discover in IBM Knowledge Center (<https://www.ibm.com/support/knowledgecenter/SSY8AC>).

How to send your comments

You can add your comments in IBM Knowledge Center. To add comments directly in IBM Knowledge Center, you need to log in with your IBM ID.

You can also send your comments to ibmkc@us.ibm.com.

Summary of changes

The summary of changes compiles a list of changes that are implemented in the IBM Spectrum Discover licensed program and the IBM Spectrum Discover library. Within each topic, these markers () surrounding text or illustrations indicate technical changes or additions that are made to the previous edition of the information.

[

Summary of changes for IBM Spectrum Discover version 2.0.3.1 as updated, July 2020

This release of the IBM Spectrum Discover licensed program and the IBM Spectrum Discover library includes the following improvements. All improvements are available after an upgrade, unless otherwise specified.

Administering

Similar source connections can be mapped in Watson Knowledge Catalog for exporting metadata. For more information, see the topic *Mapping similar source connections in Watson Knowledge Catalog* in the *IBM Spectrum Discover: Administration Guide*.

Content Search support provided for SMB. For more information, see the topic *Hints and tips for using content search* in the *IBM Spectrum Discover: Administration Guide*.

You can preserve IBM Spectrum Discover tags while you are moving or copying data. For more information, see the topic *Preserving tags during data movement* in the *IBM Spectrum Discover: Administration Guide*.

A new troubleshooting article on creating metrics services during upgrades is now available. For more information, see the topic *Upgrader not logged in when you create metrics services* in the *IBM Spectrum Discover: Administration Guide*.

Deployment

Filegroup bucket information is added. For more information, see the topic *Editing and using the TimeSinceAccess and Size Range buckets* in the *IBM Spectrum Discover: Concepts, Planning, and Deployment Guide*.

A new command is included in Sudo access requirement for ScaleILM application. For more information, see the topic *Creating or identifying a user ID and password for scanning* in the *IBM Spectrum Discover: Concepts, Planning, and Deployment Guide*.

Information about deploying Discover on KVM is available. For more information, see the topic *Deploying the IBM Spectrum Discover open virtualization appliance on the Kernel-based Virtual Machine virtualization module* in the *IBM Spectrum Discover: Concepts, Planning, and Deployment Guide*.

Information about recovering DB2® Warehouse is available. For more information, see the topic *How to correct Db2 Warehouse if it is in an unrecoverable state when you upgrade* in the *IBM Spectrum Discover: Concepts, Planning, and Deployment Guide*.

Password-based authentication is introduced for Spectrum Scale cluster. For more information, see the topic *Creating an IBM Spectrum Scale data source connection* in the *IBM Spectrum Discover: Concepts, Planning, and Deployment Guide*.

REST API changes

The following REST API documentation is added:

How to get information on all buckets that are supported by using: /db2whrest/v1/bucket : GET

How to get information on a specific bucket with: /db2whrest/v1/buckets/<bucket> : GET

How to modify the bucket detail with: /db2whrest/v1/buckets/<bucket> : PUT

]

[

Summary of changes for IBM Spectrum Discover version 2.0.3 as updated, May 2020

This release of the IBM Spectrum Discover licensed program and the IBM Spectrum Discover library includes the following improvements. All improvements are available after an upgrade, unless otherwise specified.

Administering

Tier data by using ScaleILM application. For more information, see the topic *Tiering data by using ScaleILM application* in the *IBM Spectrum Discover: Administration Guide*.

Export metadata with WKC Connector application agent. For more information, see the topic *Exporting metadata* in the *IBM Spectrum Discover: Administration Guide*.

Create auto-tag policies to group data by file type. For more information, see the topic *Grouping Data by File Type* in the *IBM Spectrum Discover: Administration Guide*.

Collect Access Control List (ACL) metadata from SMB/CIFS data source search results. For more information, see the topic *Access Control list metadata to search on* in the *IBM Spectrum Discover: Administration Guide*.

Instructions are provided to run the **automatedBackup.py** script. For more information, see the topic *Running an automated backup* in the *IBM Spectrum Discover: Administration Guide*.

Instructions are provided for fixing network settings that fail. For more information, see the topic *Network settings change fails when you run mmconfigappliance* in the *IBM Spectrum Discover: Administration Guide*.

Instructions are provided to move or copy data by using a third-party data movement application. For more information, see the topic *Moving or copying data by using a third-party data movement application* in the *IBM Spectrum Discover: Administration Guide*.

Instructions are provided for troubleshooting to address live reports that might list incorrect data. For more information, see the topic *Viewing live reports might list incorrect results* in the *IBM Spectrum Discover: Administration Guide*.

Instructions are provided for troubleshooting when network settings changes cause the system to hang after you uninstall IBM Cloud Private. For more information, see the topic *Network settings change hangs while uninstalling IBM Cloud Private* in the *IBM Spectrum Discover: Administration Guide*.

Deployment

Enable bucket notifications for Ceph® Object Storage. For more information, see the topic *Enabling bucket notifications for Ceph Object Storage* in the *IBM Spectrum Discover: Concepts, Planning, and Deployment Guide*.

Scan IBM Spectrum Scale filesets. For more information, see the topic *Automated scanning of an IBM Spectrum Scale fileset* in the *IBM Spectrum Discover: Concepts, Planning, and Deployment Guide*.

Scan ESS filesets. For more information, see the topic *Scanning an Elastic Storage Server data source connection* in the *IBM Spectrum Discover: Concepts, Planning, and Deployment Guide*.

Create an SMB data source connection by using the IBM Spectrum Discover graphical user interface. For more information, see the topic *Creating an SMB datasource connection* in the *IBM Spectrum Discover: Concepts, Planning, and Deployment Guide*.

Validate code integrity. For more information, see the topic *Validating code integrity* in the *IBM Spectrum Discover: Concepts, Planning, and Deployment Guide*.

REST API changes

The following REST API documentation is added:

How to preview a policy with: `/policyengine/v1/policies/<policy_name>/preview`:
GET

How to get a policy status with: `/policyengine/v1/policies/<policy_name>/status`: GET

How to apply an action to a policy with: `/policyengine/v1/policies/<policy_name>/<action>`: POST

How to get a list of fileset or Shares with: `/connmgr/v1/scan/<connection>/partitions`:
GET

How to start a partial scan of the datasource connection with: `/connmanager/v1/scan/<connection>/partial`: POST

The following REST API documentation command requests and responses are updated:

`api/application/appcatalog/help`: DELETE

`api/application/appcatalog/help`: GET

`api/application/appcatalog/help`: PATCH

`api/application/appcatalog/help`: POST

]

Chapter 1. Managing user access

The IBM Spectrum Discover environment provides access to users and groups. The role that is assigned to a user or group determines the functions that are available. Users and groups can also be associated with collections that use policies that determine the metadata that is available to view.

User and group access can be authenticated by IBM Spectrum Discover, a Lightweight Directory Access Protocol (LDAP) server, or the IBM Cloud® Object Storage. The administrator role can manage the user access functions.

Note: [A local user that is created on the IBM Spectrum Discover system must use a user name and password to log in. Users from an external LDAP or IBM Cloud Object Storage domain must include the domain name as a prefix to the user name with a forward slash (/), such as "<domain>/<user>". The domain name is the name that is given to the external authentication domain in IBM Spectrum Discover.]

Roles

Roles determine how users and groups can access records or the IBM Spectrum Discover environment.

If a user or group is assigned to multiple roles, the least restrictive role is used. For example, if a user is assigned to the **Data User** role but is also included in a group that is assigned to the **Data Admin** role, that user has the privileges of the **Data Admin** role.

The following roles are available:

Admin

This role can create users, groups, and collections. This role can also manage connections to Lightweight Directory Access Protocol (LDAP) and IBM Cloud Object Storage domains. This role can use the Application Management APIs to install, upgrade, or delete IBM Spectrum Discover applications that use the IBM Spectrum Discover API service.

Data Admin

Users with this role can access all metadata that is collected by IBM Spectrum Discover and is not restricted by policies or collections. This role can also define tags and policies, including policies that assign a collection value to a set of records.

Note:

The built-in `Collection` tag is a special tag. This tag can be set only by users with the **Data Admin** role. All other tags can be set by any user with the **Data User** or **Data Admin** or **Collection Admin** role.

Users with this role can also edit local users and local groups and assign roles and collections to users and groups.

Collection Admin

The **Collection Admin** role is as a bridge between the **Data Admin** role and the **Data User** role. Users with the **Collection Admin** role can:

- Create, update, and delete the policies for the collections that they administer.
- View, update, and delete policies of data users for the collections they administer. They cannot delete a policy if it has a collection that they do not administer.
- Add users to collections that they administer. These data users can access to a particular collection, which means that they can access to the records marked with that collection value.
- List any type of tag and create or modify `Characteristic` tags. They cannot create, modify, or delete `Open` and `Restricted` tags. These permissions are the same as the ones associated with the **Data User** role.

Data User

Users with this role can access metadata that is collected by IBM Spectrum Discover, but metadata access can be restricted by the collections that are assigned to users in this role. This role can also define tags and policies, based on the collections to which the role is assigned.

Service User

This role is assigned to accounts for IBM service and support personnel.

Initial login

To log in with the default login information for the IBM Spectrum Discover graphical user interface, use the following information.

The default login for the IBM Spectrum Discover graphical user interface is:

Username

sdadmin

Password

Passw0rd

Note: It is strongly recommended that the administrator change the password during the initial login.

Resetting the sdadmin password

If the `sdadmin` password changes and you forget the password, you can access the keystone container and run the `reset_sdadmin_details.sh` script to reset the password to the original password.

Procedure

1. Get the keystone pod name.

```
kubectl get pods | grep keystone
```

2. Using the pod name, perform the following command to open a bash shell on the keystone container. Substitute `{pod name}` with the name returned from the previous command.

```
kubectl exec -it {pod name} bash
```

3. In the bash shell on the container, run the `reset_sdadmin_details.sh` script to reset the details back to the original password.

```
./reset_sdadmin_details.sh
```

4. Ensure that the password details are reset. When the password is reset, the list of users is displayed by using the following commands. If the username is not reset correctly, a "401 unauthorized error" is returned.

```
source keystone_sdadminrc  
openstack user list
```

Password policies

IBM Spectrum Discover, in the 2.0.2.1 release, introduces password policies for the local users who are configured in the default authentication domain.

The password policies that are introduced in IBM Spectrum Discover 2.0.2.1, for all local user accounts, enhance their security.

Note:

IBM Spectrum Discover does not enforce password policies for the user accounts that are imported to the IBM Spectrum Discover authentication scheme. These policies include all user accounts imported from

the external domains like LDAP or IBM Cloud Object Store domains that are configured with IBM Spectrum Discover. Any password policies that are configured for these external authentication providers (LDAP/IBM Cloud Object Store), would apply to the corresponding users from these authentication domains.

Password policies

IBM Spectrum Discover local users must follow the password policies that are defined in the 2.0.2.1 release.

Password strength requirements

- Passwords must have a minimum length of 7-characters.
- Passwords must contain at least one letter.
- Passwords must contain at least one digit.

Unique password history requirements

- Users must create a unique password each time the password is changed. The new password cannot be any of the last five passwords previously used.

Password expiration requirements

- The User password expires after 90 days from the time it is changed.

Password change requirements

IBM Spectrum Discover users with Admin roles (like the "sdadmin" user) can create a new user or reset the password of an existing user. However, this password expires when the user logs in for the first time and must be changed immediately.

Account lockout requirements

A user account is locked out for 1 hour after five successive failed login attempts.

Password upgrade for existing users

IBM Spectrum Discover deployments, upgraded from versions that precede the release 2.0.2.1, include the new password policies that are applied to local user accounts. Existing user accounts are also impacted in the following ways:

- Existing users can continue to use their current passwords to log in to the system.
- Passwords for existing user accounts expire only in the following situations:
 - Passwords expire when users change their password. In this scenario, the new password will expire after 90 days.
 - Passwords expire when the administrative user resets the user password. In this scenario, the updated password expires immediately after the first login and the user must create a new unique password.
- When the user password is changed, the following password policies are enforced:
 - **Password strength requirements** .
 - **Unique Password history requirements** - This policy restricts users from reusing any of the last five passwords.
- On completing the product upgrade, the **Account lockout requirements** policy is immediately enforced for all local users that includes all existing users.

Note: To apply all the password policies to the local user accounts after they upgrade to 2.0.2.1 release, follow the listed recommendations:

- The Admin user resets passwords for all the existing local user accounts and communicates the new password to the respective users.

- All the local users use the UI Password change REST API to change their passwords. For more information, see “[Changing password](#)” on page 4 and `/auth/vi/users/<user_ID>/password: Post` in *IBM Spectrum Discover: REST API Guide*.

Changing password

To change your password, follow the procedure that is mentioned.

About this task

To change your password, follow these steps.

Procedure

1. Click your user name on the upper-right corner of the screen and select **Account Settings**.

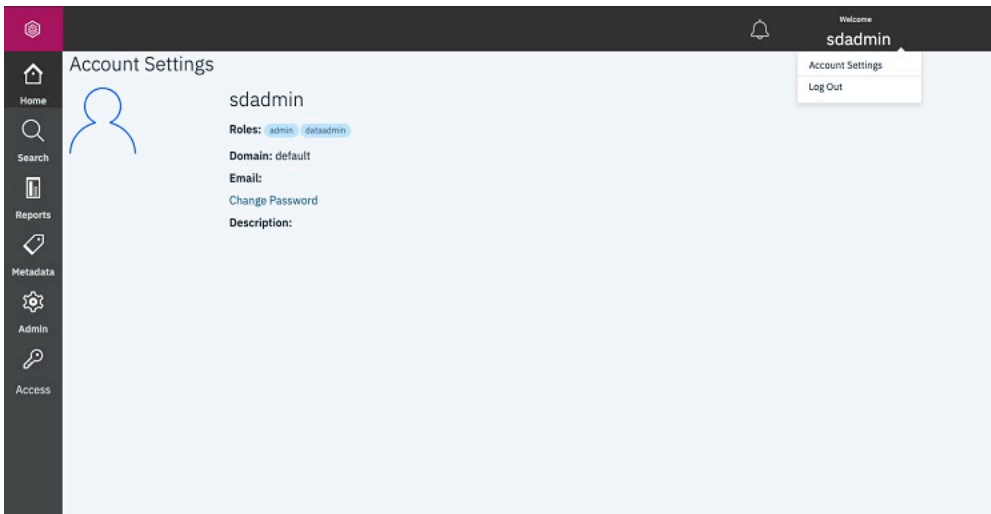


Figure 1. Select the Account Settings option

2. Select **Change Password**.

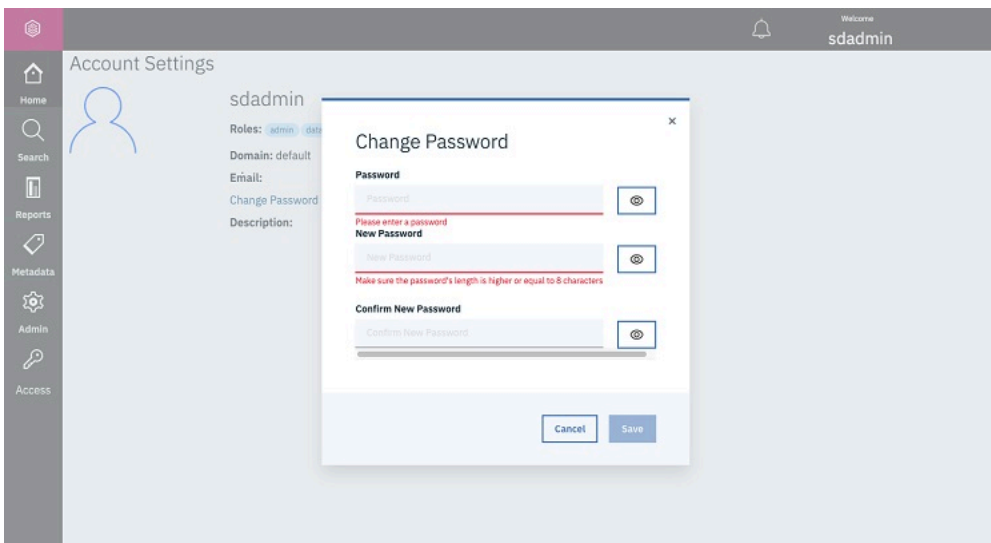


Figure 2. Change Password fields

3. Enter the existing password, new password, and new password confirmation.
4. Click **Save**.

Managing user accounts

The administrator can create and manage local user accounts, which are authenticated by IBM Spectrum Discover. The administrator can also assign local or LDAP and IBM Cloud Object Storage-managed users to roles and collections.

Use the **Users** tab on the **Access** page to view information about user accounts that are authenticated by the local domain or either an LDAP or IBM Cloud Object Storage server. You can also use the tab to create, edit, or delete local users. You cannot create or delete either LDAP or IBM Cloud Object Storage user accounts, but you can assign these users to roles and collections.

Creating a local user account

To create a local user account that is authenticated by IBM Spectrum Discover, click **Create new user**. To create a local user account that is authenticated by IBM Spectrum Discover, click **Create Local User**. For more information, see [“Creating user accounts” on page 6](#).

Editing a user account

You can edit account information for a local user. You cannot edit the details of either Lightweight Directory Access Protocol (LDAP) or IBM Cloud Object Storage user accounts, but you can assign these users to roles and collections.

To edit a local user account, select the user that you want to edit and click **Edit**. Use the **Edit User** window to edit the local user account.

To edit an LDAP or IBM Cloud Object Storage user account roles, select the user that you want to edit and click **Edit**. Use the **Edit User** window to assign these users to roles and collections.

Deleting a local user account

To delete a local user account, select the user that you want to delete and click **Delete**.

User information

The **Users** tab lists the users that are available from the local domain and from either LDAP or IBM Cloud Object Storage connections. The tab includes the following user account information.

User Name

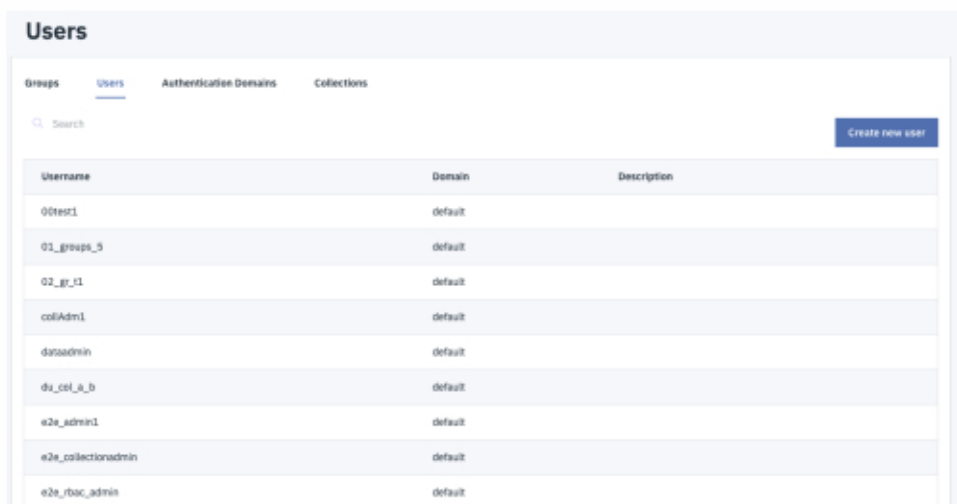
Indicates the username for the account.

Domain

Indicates the domain that provides authentication for the user. For authentication by IBM Spectrum Discover, the domain name is **Default**.

Description

Indicates the description of the user.



Username	Domain	Description
00test1	default	
01_group_5	default	
02_gr_11	default	
colAdmin	default	
dsadmin	default	
dl_col_a_b	default	
e2e_admin1	default	
e2e_collectionadmin	default	
e2e_tloc_admin	default	

Figure 3. The Users tab

Creating user accounts

The administrator can create local user accounts, which are authenticated by IBM Spectrum Discover, and assign roles to users.

About this task

Use the **Users** tab on the **Management** page to create a local account:

- You can also assign roles and passwords to users.
- You can also add a user to a group.

Procedure

1. From the **Management** page, click **Create new user** to open the **Users** window.

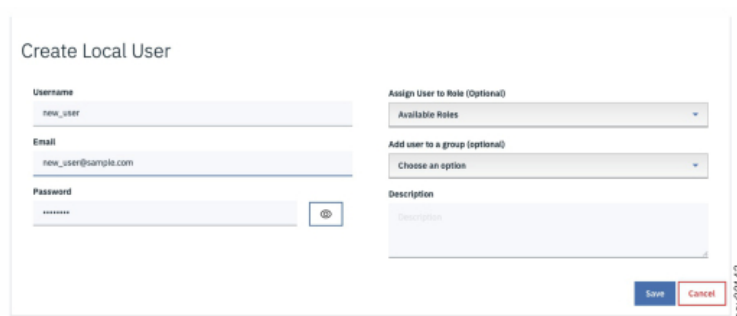


Figure 4. The Create Local User window

2. Enter a **User Name** and **Email** address for the user.
3. Enter a **Password** for the user.
4. This step is optional. Use the **Assign User to Role** list to assign one or more roles to the user. For more information about roles, see [“Roles”](#) on page 1.

Users that are assigned the **Data User** or **Collection Admin** role must also be associated with at least one collection.

5. This step is optional. Use the **Assign User to Group** list to assign the user to one or more user groups. You can also use the **Groups** tab to assign users to groups.
6. This step is optional. Enter a **Description** for the user.
7. Click **Save**.

Managing groups

The administrator can create and manage local groups that are authenticated by IBM Spectrum Discover. The administrator can also assign local or Lightweight Directory Access Protocol (LDAP) and IBM Cloud Object Storage system-managed groups to roles and collections.

Use the **Groups** tab on the **Access** page to view information about groups accounts that are authenticated by either a local domain, an LDAP server, or the IBM Cloud Object Storage server. You can also use the tab to create, edit, or delete local groups. You cannot edit or delete LDAP or IBM Cloud Object Storage groups, but you can assign these groups to roles and collections.

Creating a local group

To create a local group, click **Create new group**. For more information, see [“Creating groups”](#) on page 7.

Editing a group

To edit a group, select the group that you want to edit and click **Edit**. Use the **Edit Group** window to edit the local group.

Deleting a local group

To delete a local group, select the group that you want to delete and click **Delete**.

Group information

The **Groups** tab includes the following information.

Group Name

Indicates the name for the group.

Domain

Indicates the domain that provides authentication for the group. For authentication by IBM Spectrum Discover, the domain name is **Default**.

Description

Indicates the description of the group.

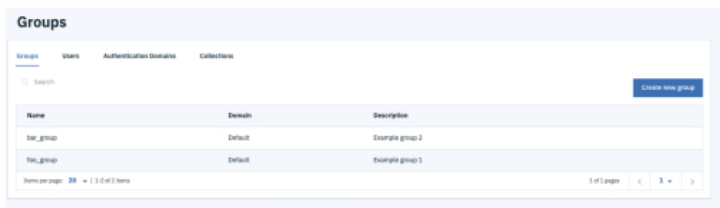


Figure 5. The Groups Tab

Creating groups

The administrator can create local groups that are authenticated by IBM Spectrum Discover, and assign users and roles to the groups.

About this task

Use the **Groups** tab on the **Access** page to create local groups. You can assign users and roles to the group and add the group to a collection.

Procedure

1. From the **Groups** tab of the **Access** page, click **Create new group** to open the **Create Local Group** window.

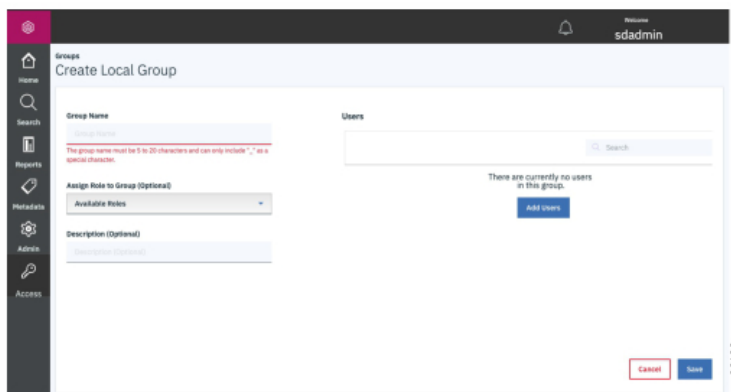


Figure 6. The Create Local Group window

2. Enter a **Group Name**.
3. Optional. Use the **Assign Role to Group** list to assign one or more roles to the group. Use the **Assign User to Role** list to assign one or more roles to the group. For more information, see [“Roles” on page 1](#).

Groups that are assigned the **Data User** role or **Collection Admin** role must be associated with at least one collection.

4. Click **Add Users** to open the **Add Users** window and add one or more local users to the collection.

Enter a username that you want to add to the group and press Enter. The window lists each name that you enter. Click a name to remove it from the list. Click **Add** to add the users to the group.

The **Users** list displays the following details for users that are added to the group.

Username

The username or email address of the member.

Domain

The domain that provides authentication for the member.

5. This step is optional. Enter a **Description** for the group.

6. Click **Save**.

Managing collections

Collections are logical groups of metadata that share a common member access list. For example, a collection can restrict metadata within a research project to the members of the project only. Members outside of the project cannot see the metadata.

The administrator can:

- Create collections.
- Assign users and groups to collections.
- Create a policy to associate specific metadata that is collected by IBM Spectrum Discover with the collection.
- Assign a collection to a connection to associate specific metadata from that connection data source that is collected by IBM Spectrum Discover with the collection.

Users with the **Data Admin** role can view all metadata that is collected by IBM Spectrum Discover and are not restricted by collections. Users with the **Data Admin** role can create policies that assign a collection value to a set of records, thus grouping a set of records under a collection.

Users with the **Collection Admin** role can add the **Data User** role to user for collections that they administer. Adding this role gives data users access to a particular collection, which allows the users to access the records that are marked with that collection value.

Use the **Collections** page to manage collections.

Creating a collection

To create a collection, click **Create Collection**. For more information, see [“Creating collections” on page 9](#).

Editing a collection

To edit a collection, select the collection that you want to edit and click **Edit Collection**. Use the **Edit Collection** window to edit the collection.

Deleting a collection

To delete a collection, select the collection that you want to edit and click **Delete Collection**.

Collections information

The **Collections** page includes the following information.

Collection Name

Indicates the name of the collection.

Description

Indicates the description of the collection.

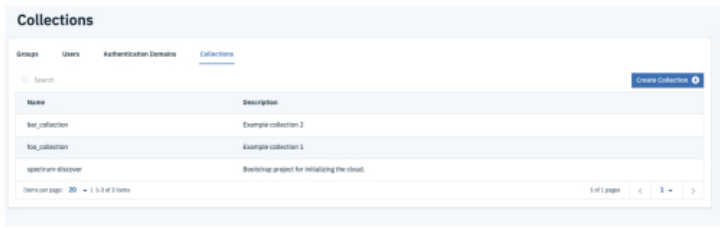


Figure 7. The Collections tab

Creating collections

The administrator can create collections or assign users and groups to collections. A **Collection Admin** administrator can assign users and groups only to collections that they administer. A **Data Admin** administrator can use the auto-tag policy to associate metadata records with a collection.

About this task

Collections are logical groups of records. Access to these record groups is restricted to specific users or groups. The administrator can associate policies with an appropriate collection value so that searches can be restricted to only the scope that a user or group has permissions to see.

Use the **Collections** tab on the **Access** page to create collections.

Procedure

1. From the **Collections** tab of the **Access** page, click **Create Collection** to open the **Create Collection** window.

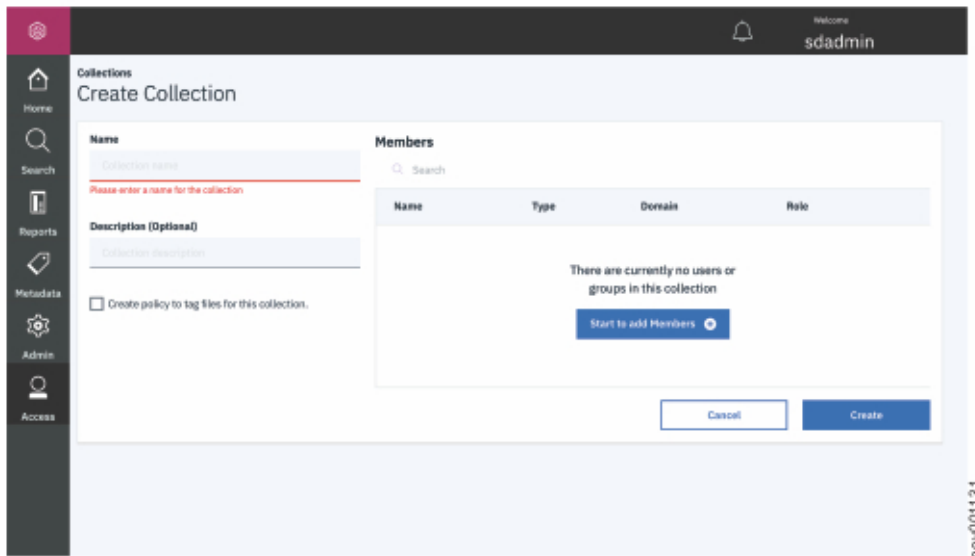


Figure 8. Create a Collection

2. Enter a collection **Name** and optional **Description**.
3. Click **Start to Add Members** to open the **Add Members** window and add one or more users or groups to the collection.

Enter a username, group name, or email address of a member to include in the collection and press **Enter**. The window lists each name or address that you enter. Click a name or address to remove it from the list. Click **Add** to add the members to the collection. Select the role for the member on the collection. The default role is **Data User**.

The **Members** area lists the following details for the members of the collection.

Name

The username, group name, or email address of the member.

Type

The account type: user or group.

Domain

The domain that provides authentication for the member.

Role

The role on the collection that is assigned to the member.

4. To create a policy for the collection, select **Create policy to tag files for this collection**. For more information, see [Chapter 2, “Managing metadata policies,” on page 15](#).
5. Click **Create**.

Managing LDAP and IBM Cloud Object Storage System connections

The administrator can create and manage connections to LDAP or IBM Cloud Object Storage System servers that provide authentication for IBM Spectrum Discover users.

Use the **Authentication Domains** tab on the **Access** page to create, test, manage, or delete LDAP connections.

You can create a connection that includes all users and groups that are authenticated by an LDAP server or only users or groups within a specified LDAP member range.

Note: You cannot specify a member range for users and groups that are managed by the IBM Cloud Object Storage System.

Creating a connection

To create a connection to an authentication domain, click **Add Domain Connection**.

For steps to create a connection to an LDAP server, see [“Creating an LDAP connection” on page 10](#).

For steps to create a connection to an IBM Cloud Object Storage system server, see [“Creating an IBM Cloud Object Storage connection” on page 12](#).

Editing a connection

To edit a connection, click **Edit**.

Deleting a connection

To delete a connection, click **Delete**.

Creating an LDAP connection

The administrator can create a connection to a Lightweight Directory Access Protocol (LDAP) server that provides authentication for IBM Spectrum Discover users.

About this task

Use the **Authentication Domains** tab on the **Access** page to create an LDAP connection. You can create a connection that includes all users and groups that are authenticated by an LDAP server or only users or groups within a specified LDAP member range.

Procedure

1. From the **Authentication Domains** tab of the **Access** page, click **Add Domain Connection** to open the **Add Domain Connection** window.
2. From the **Type** list, select **LDAP**.

Figure 9. Create an LDAP connection

3. Enter the following information for the LDAP directory:

Name

Indicates a name that IBM Spectrum Discover associates with the connection to the directory that provides authentication.

Type

Indicates the directory type, which is LDAP.

Port

Indicates the LDAP server port that provides the connection.

Username

Indicates the distinguished name (DN) for the user that is used to access directory name entries. Use the following format:

```
cn=relative_distinguished_name dc=domain_component
```

For example,

```
cn=Randy Marsh,dc=example,dc=com
```

Password

Indicates the password for the user name.

Suffix/Base DN

Indicates the DN that is the base of entry searches in the directory. For example:

- dc=test
- dc=org

Group Name Attribute

Indicates the LDAP attribute that is mapped to the group name.

Group ID Attribute

Indicates the LDAP attribute that is mapped to the group ID.

Group Member Attribute

Indicates the LDAP attribute that is mapped to show group membership.

Group Object Class

Indicates the LDAP object class for groups.

Group Tree DN

Indicates the DN that is the base for group searches.

Username Attribute

Indicates the LDAP attribute that is mapped to the user name.

User ID Attribute

Indicates the LDAP attribute that is mapped to the user ID.

User Object Class

Indicates the LDAP object class for users.

User Tree DN

Indicates the DN that is the base for user searches.

4. Click **Connect**.

Creating an IBM Cloud Object Storage connection

The administrator can create a connection to an IBM Cloud Object Storage server that provides authentication for IBM Spectrum Discover users and groups from the corresponding domain.

About this task

Use the **Authentication Domains** tab on the **Access** page to create a connection to an IBM Cloud Object Storage System server. You must provide credentials for the IBM Cloud Object Storage security administrator.

All users and groups that are managed by the IBM Cloud Object Storage are available for IBM Spectrum Discover. You cannot specify a member range for these connections.

Note: You cannot run scans unless you add override warnings in the configuration file.

Procedure

1. From the **Authentication Domains** tab of the **Access** page, click **Add Domain Connection** to open the **Add Domain Connection** window.
2. From the **Type** list, select **IBM Cloud Object Storage**.

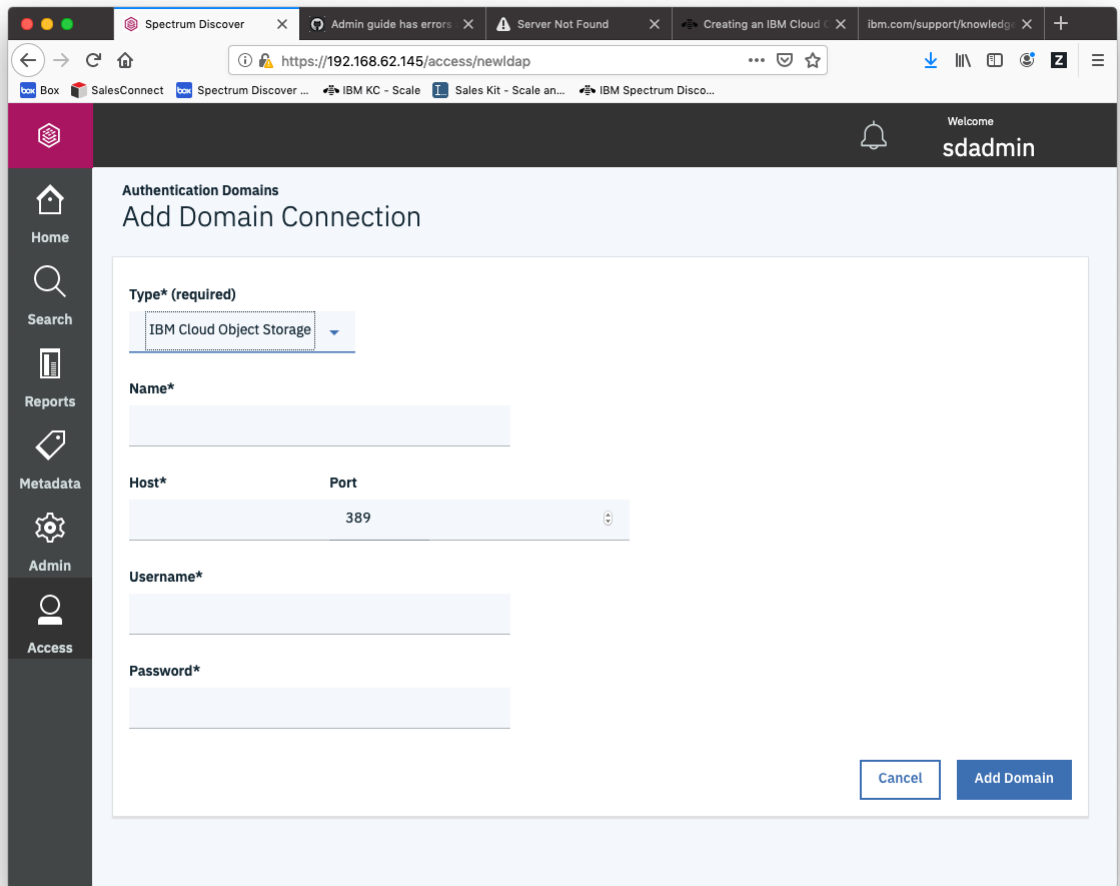


Figure 10. Create an IBM Cloud Object Storage Connection

3. Enter the following information for the IBM Cloud Object Storage connection:

Name

Indicates the IBM Cloud Object Storage domain name.

Host

Indicates the IBM Cloud Object Storage hostname.

Port

Indicates the IBM Cloud Object Storage port number.

User name

Indicates the IBM Cloud Object Storage security administrator name.

Password

Indicates the IBM Cloud Object Storage security administrator password.

4. Click **Connect**.

Chapter 2. Managing metadata policies

Policies might be used to automatically tag a set of documents on a periodic basis. In addition, policies might be used to send sets of documents to be deep-inspected by a registered application.

Roles and permissions

Data User

Users with this role can create, modify, and view their policies. Policies can be applied only to the collections the user has access to.

Users with the **Data User** role cannot use a `COLLECTION` tag when they create or modify policies.

Collection Admin

Users with this role can create, modify, and view their policies. Policies can be applied only to the collections that they administer. Users with the **Collection Admin** role cannot use a `COLLECTION` tag when they create or modify policies.

Data Administrator

Users with this role can create, modify, view, and delete policies.

Security Administrators

Users with this role cannot create, modify, view, or delete policies.

Service User

Users with this role cannot create, modify, view, and delete policies.

Adding policies

You can add policies to help with data management.

About this task

[To add a policy, you must define the policy and its parameters, establish a schedule to run the policy, and save the policy.]

You can add custom metadata values to all or a subset of the records based on filter criteria. For example, you can add a project name to records based on their location within the file system or owner ID.

You can add a policy filter, which is similar to the **where** clause in an SQL query. The filter must be constructed by using standard SQL syntax:

- To enact a policy on all files not accessed in one year, the filter might be written as:

```
atime < (NOW() - 365 DAYS)
```

- To enact a policy on all files owned by `Smithers`, the filter might be written as:

```
owner='Smithers'
```

- To enact a policy on all PDF files in cluster `c11` and data source `fs1`, the filter might be written as:

```
cluster='c11' and datasource='fs1' and filetype='pdf'
```

Procedure

1. [
Select the **Policies** page and click **Add Policy**.
]
2. [
Enter a name for the policy in the **Name** box (for example, `MyPolicy`).
]

-]
3. Select the required policy type from the **Policy Type** menu.
[The policy types are:
 - AUTOTAG
 - CONTENT SEARCH
 - DEEP-INSPECT]
 4. [Click **Next Step**.]
 5. [Complete the policy information:
 - a. Select the list of **Collections** the policy applies to. If no collections are selected, the policy applies to all collections available to the user when run.
 - b. You must use a filter for the policy. The filter defines which set of records the policy acts on. Enter your filter into the **Filter** box (for example, `size > 100`).
 - c. Complete the policy-specific parameters based on which policy type you select in option 3:
 - 1) You can set parameters for AUTOTAG policy types. For more information, see [“Adding auto-tagging policy parameters”](#) on page 17.
 - 2) You can set parameters for DEEPINSPECT policy types. For more information, see [“Adding deep-inspection policy parameters”](#) on page 18.
 - 3) You can set parameters for CONTENTSEARCH policy types. For more information, see [“Adding content search policy parameters”](#) on page 18.]
 6. [Now that your policy parameters are defined, you must schedule the frequency.]
 7. [Click the slider control to set the status to one of the following values:

Active
An *Active* policy is run whenever its scheduling event is reached.

Inactive
An *Inactive* policy is not run when its scheduling event is reached, including the **Now** event.

]
 8. Select a **Schedule** to apply the policy. [Indicate whether you want to schedule the policy Now, Daily, Weekly, or Monthly.]

Note: Policy schedule times are entered in Coordinated Universal Time (UTC).

Now
Indicates that the policy is applied immediately unless the policy's status is *Inactive*.

Daily
Indicates a specific time of day to apply the policy. Enter the time of day to apply the policy by clicking the hour and minute from the widget that is shown. The policy is applied daily at the specified time.

Weekly
Indicates a specific day and time in which to apply the weekly policy:

- a. Enter the time of day to apply the policy by clicking the hour and minute from the widget that is shown.
 - b. Select the day of the week from the list of days.
- The policy is applied once a week on the specified day and time.

Monthly

Indicates a specific day and time in which to apply the monthly policy:

- a. Enter the time of day to apply the policy by clicking the hour and minute from the widget that is shown.
- b. Select the date by clicking the month and day from the widget that is shown.

The policy is applied once a month on the specified day and time.

9. [Click **Next Step**.]

10. [Review the data and click **Save** to save the new policy.]
The new policy displays in the list of policies on the **Policies** tab.

Adding auto-tagging policy parameters

[You can add specific parameters for auto-tagging policies after you define the policy type and set up the mandatory filter or optional collection information.]

Procedure

1. Associate one or more tags with this policy by using one of the following methods:

- a) Click **+Add Tag**.
- b) Select a tag name from the **Field** menu.
The Fields can also be specified by going to **Metadata > Tags**.

- c) [Select which **Tag** to apply (for example, TEMPERATURE) and then enter the appropriate **Values**.

Note: If you do not know the valid values for a tag, go to **Search** in the main menu and select the tag from the **Start a visual exploration** list. Click the **Go** "circle arrow" icon. The valid values of the tag are displayed.

- d) To delete a Field, click the **Delete** "minus" icon next to a field.
- e) You can add more tag values by clicking the **+Add Tag** control.
Each new Field defaults to the next item in the **Field** menu.

[Or you can automatically extract the tag from the directory path by clicking the **Extract tag from box**. For example, you might have files in a directory structure by department and want to extract the department name into a tag. Automatically extract the tag:

- 1) Select the **Extract tag from path** checkbox.
- 2) Select a tag from the **Field** menu.
- 3) Specify the **Depth** in the path to be used as the value of the tag.

[To create a new tag, see [“Creating tags”](#) on page 41.

2. [Click **Next Step**.

]

Adding deep-inspection policy parameters

You can add deep-inspection policies.

About this task

You can enrich metadata through an external deep inspection application. For more information, see [Chapter 8, “Managing applications,” on page 61](#). For example, you can extract patient names from medical records and index the names. Indexing the names helps when you search for files that pertain to patients by name. Deep inspection policies can send lists of files to an application, which can examine the contents of files and return the values that it finds paired with defined tag keys.

Add specific parameters for deep-inspection policies by using the following information.

Procedure

1. [
Add the **Application** name (for example, example-application).
]
2. [
Click **+Add Tag**.
]
3. [
Select which **Parameter** to apply (for example, extract-tags), and then select the appropriate **Values** (for example, TEMPERATURE).
]
4. You can add more parameters by clicking the **+Add parameter** control. This is optional.
5. You can delete a parameter by clicking the **Delete** "minus" icon next to the parameter's **Value**. This is optional.
6. [
Click **Next Step**.
]

Adding content search policy parameters

You can add search content parameters for your policies.

About this task

You can enrich metadata through the built-in content search application. For more information, see [Chapter 3, “Using content search policies,” on page 25](#).

You can add specific parameters for content search policies by using the following steps:

Procedure

1. [
Select contentsearchagent for the **Application**.
]
2. Click **+Add Row** to open the **Parameter** dialog.
3. Enter a tag name in the **Parameter** box and select one or more **Search Expressions** from the dropdown list.
4. [For the **Value**, select either **True/False** or **Value matching expression**.]

5. Repeat steps 1 - 3 to set other tags that use the same policy.
6. [
 - Click **Next Step**.
]

Running policies

You can configure policies to run at specified or scheduled times, at policy creation time, or when the system is manually started, paused, restarted, or stopped.

About this task

To configure a policy to run:

Procedure

1. Go to **Metadata**
2. Click a policy to select it. The following screen displays:

Policy	Type	Schedule (UTC)	Status	Progress	Collections	Last Modified by	Last Modified
CS_ner_srm	DEEPIPECT	None	Active	380%	2 failures of 2000	admin	2019-10-28T17:57:28.000Z

Figure 11. Policies table

From the screen, you can perform the following actions on the selected policy:

Pause

Click the "vertical bars" icon to pause a policy that is running. When a policy is paused, it enters a Paused state. A policy cannot be paused until the current batch that is being processed finishes.

Start

Click the "right-arrow" icon to resume a paused policy or to start a stopped policy from its beginning. When a policy is started, it enters a Running state.

Stop

Click the "square" icon to stop a policy that is in progress. When a policy is stopped, it enters a Stopped state.

When a policy completes, it enters a Stopped state. The progress column indicates the success or failure status of the policy. If there are failures, you can examine the per policy execution log files to obtain more details of the failures.

For more information, see [“Viewing policies” on page 19](#).

Viewing policies

You can view your policy information.

About this task

You can see a list of the active and inactive policies and their status.

Policy	Type	Schedule (UTC)	Status	Progress	Collections	Last Modified by	Last Modified
archivepol	AUTOTAG	Done	active	100% 0 failed out of 73450		sdadmin	2020-01-21T12:13:31.000Z
archivepol2	AUTOTAG	Done	active	100% 0 failed out of 547		sdadmin	2020-01-21T12:38:12.000Z

Figure 12. Policies table

Procedure

1. Go to **Metadata > Policies**
2. View a table of the specified policies with the following aspects:

Policy

Displays the name of a policy

Type

Displays the policy type. There are three types:

- **AUTOTAG:** You can apply custom metadata values to some, or all of the records based on filter criteria.
- **CONTENTSEARCH:** You can enrich metadata by using the built-in CONTENTSEARCH application.
- **DEEPIINSPECT:** You can enrich metadata through content inspection of source data.

Schedule

Displays the frequency at which a policy is applied. Policy schedule times are displayed in Coordinated Universal Time (UTC).

Done

Indicates that the policy is applied.

Daily: 00:00

Indicates that the policy is applied one time a week on the displayed day and time.

Weekly:[day], 00:00

Indicates that the policy is applied one time a week on the displayed day and time.

Monthly:[date], 00:00

Indicates that the policy is applied one time a month on the displayed date and time.

Status

Displays the status and current state of the policy.

- A policy's **Inactive** status shows the **None** state.
- A policy's **Active** status can have a state of **Initialized**, **Running**, **Paused**, or **Stopped**.

Progress

Displays the percentage of completion of a policy.

If there is an **Error** "yellow triangle" icon, you can hover the cursor over it to see more information.

If there are some records that met the filter criteria but fail to be tagged, the number of failed records and the total number of records are displayed below the percentage of completion.

If there are failures, the per policy execution log files can be examined to obtain more details of the failures.

Collections

Displays the name of the collection that the policy applies to or the number of collections the policy applies to if there is more than one collection.

Last modified by

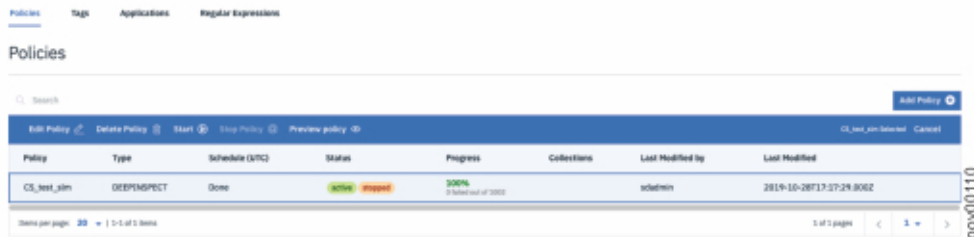
Displays the name of the user who last modified the collection.

Last modified

Displays the date and time that the collection was last modified.

You can add a policy by clicking **Add Policy +**. For more information, see [Adding auto-tagging policy parameters](#) or [Adding deep-inspection policy parameters](#).

3. Click a policy to select it. The following screen displays.



Policy	Type	Schedule (UTC)	Status	Progress	Collections	Last Modified by	Last Modified
CL_test_gim	DEEPSPECT	Done	Active	100%	1 total out of 2002	mlhadmin	2019-10-28T17:27:29.800Z

Figure 13. Policies table

[From the screen, you can perform the following actions:

Start, Pause, Stop, or delete

You can start, pause or stop a policy. For more information, see [“Running policies” on page 19](#).

Edit/Delete

Use the **Edit** "pencil" icon to modify a policy. Use the **Delete** "trashcan" icon to delete a policy. You cannot delete a policy that is running (The "trashcan" icon is made unavailable).

View

Preview the details of the selected policy.

Viewing policy log files

You can view policy execution log files to gain more information of the success or failure of the policy.

About this task

To view policy execution log files:

Procedure

1. Log in to the master node as the moadmin user.
2. Change directory to the location of the per policy execution logs:

```
cd /gpfs/gpfs0/policies
```

3. A subdirectory within this location is created for each policy execution.
The directory is named: <policy_name>_<policy_start_date>_<policy_start_time>
4. Change directory into the policy execution directory of interest. View the run.log file within this location. For example:

```
[2020-01-16 12:00:31] - Execution beginning for policy (testtag_policy)
[2020-01-16 12:00:31] - Policy stats update: {'pol_id': 'testtag_policy', 'execution_info':
{'start_time': "2020-01-16_12:00:31", "total_count": 0, "submitted_count": 0,
"failed_count": 0, "completed_count": 0, "skipped_count": 0, "autotag_count": 0,
"autotag_size": 0, "run_id": null}}
[2020-01-16 12:00:32] - Policy stats update: {'pol_id': 'testtag_policy', 'execution_info':
{'start_time': "2020-01-16_12:00:31", "total_count": 2002, "submitted_count": 0,
"failed_count": 0, "completed_count": 0, "skipped_count": 0, "autotag_count": 0,
"autotag_size": 0, "run_id": null}}
[2020-01-16 12:00:32] - Applying action 'AUTOTAG' to 2002 documents
[2020-01-16 12:00:32] - Policy stats update: {'pol_id': 'testtag_policy', 'execution_info':
```

```

{"start_time": "2020-01-16_12:00:31", "total_count": 2002, "submitted_count": 2002,
"failed_count": 0, "completed_count": 0, "skipped_count": 0, "autotag_count": 0,
"autotag_size": 0, "run_id": "a23e14c1ccaa49d9aefcde229c65ae0b"}
[2020-01-16 12:00:32] - Policy stats update: {'pol_id': 'testtag_policy', 'execution_info':
{"start_time": "2020-01-16_12:00:31", "total_count": 2002, "submitted_count": 2002,
"failed_count": 0, "completed_count": 2002, "skipped_count": 0, "autotag_count": 0,
"autotag_size": 0, "run_id": "a23e14c1ccaa49d9aefcde229c65ae0b"}}
[2020-01-16 12:00:33] - Policy stats update: {'pol_id': 'testtag_policy', 'execution_info':
{"start_time": "2020-01-16_12:00:31", "total_count": 2002, "submitted_count": 2002,
"failed_count": 0, "completed_count": 2002, "skipped_count": 0, "autotag_count": 0,
"autotag_size": 0, "run_id": "a23e14c1ccaa49d9aefcde229c65ae0b", "end_time":
"2020-01-16_12:00:33"}}
[2020-01-16 12:00:33] - Policy testtag_policy run ending
[2020-01-16 12:00:33] - Policy testtag_policy run completed

```

Modifying policies

You can modify your policies to more easily work with and use your data.

About this task

You can modify a policy from the table on the **Policies** page. You cannot change the **Name** or **Type** of a policy.

Procedure

1. Go to **Metadata > Policies**
2. Click the **Edit** "pencil" icon in the **Edit/Delete** column of a policy that you want to modify. The **Modify a policy** window displays. The display of the policy parameters depends on the policy type. The policy configuration area changes depending on the **Extract tag from path** setting.

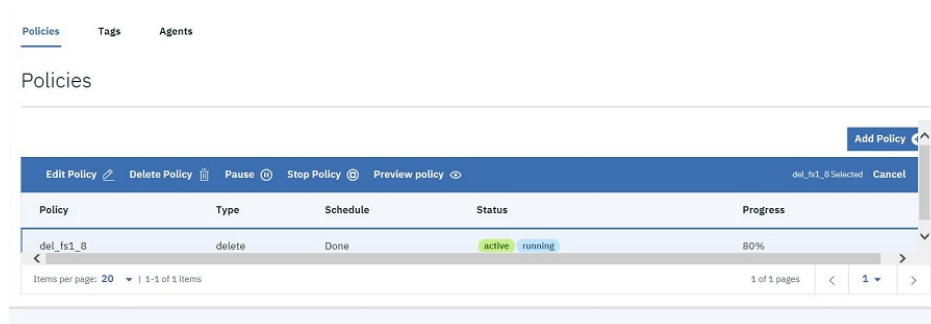


Figure 14. Modify a policy

3. Click the slider control to set the status to one of the following values:

Active

An *Active* policy is run whenever its scheduling event is reached.

Inactive

An *Inactive* policy is not run when its scheduling event is reached, including the **Now** event.

4. Check the **Name** field to verify that it is the policy you intend to modify. The name cannot be modified.
5. Modify the collections that the policy applies to (if required).
6. Modify or enter a filter in the **Filter** box. The filter must be constructed by using standard SQL syntax. For more information, see [example filters](#).
7. Modify the parameters specific to the policy type (if required).
 - For more information about AUTOTAG policy types, see [“Adding auto-tagging policy parameters” on page 17](#).
 - For more information about DEEPINSPECT policy types, see [“Adding deep-inspection policy parameters” on page 18](#).

- For more information about CONTENTSEARCH policy types, see [“Adding content search policy parameters” on page 18](#), [“Adding deep-inspection policy parameters” on page 18](#), or [“Managing user accounts” on page 5](#).
8. Specify a **Schedule** to apply the policy. Policy schedule times are entered in Coordinated Universal Time (UTC).

Now

The policy is applied immediately, unless the policy's status is **Inactive**

Daily

Indicates a specific time of day to apply the policy. Enter the time of day to apply the policy by clicking the hour and minute from the widget that displays. The policy is applied daily at the specified time.

Weekly

Indicates a specific day and time in which to apply the weekly policy:

- a. Enter the time of day to apply the policy by clicking the hour and minute from the widget that displays.
- b. Select the day of the week from the list of days.

The policy is applied one time a week on the specified day and time.

Monthly

Indicates a specific day and time in which to apply the monthly policy:

- a. Enter the time of day to apply the policy by clicking the hour and minute from the widget that displays.
- b. Select the date by clicking the month and day from the widget that displays.

The policy is applied one time a month on the specified day and time.

9. Click **Save** to save the policy.

The modified policy is displayed in the list of policies on the **Policies** tab.

Deleting policies

You can delete policy information.

About this task

A policy can be deleted from the table on the **Policies** page. You cannot delete a policy that is running. A user with the role **Data User** can delete only their own policies.

Procedure

1. Go to **Metadata > Policies**.
2. Click the **Delete** "trashcan" icon in the **Edit/Delete** column of the policy you want to delete.
If the "trashcan" icon is unavailable, then the policy is not available for deletion.
3. Click **Delete** in the confirmation window.
The policy is removed from the table in the **Policies** tab.

Chapter 3. Using content search policies

You can define regular expressions to search for and create policies that use these regular expressions.

About this task

You can enrich metadata through content inspection of source data by using the built-in CONTENTSEARCH application. To use this function, you can define regular expressions to search for and create policies that use these regular expressions.

When the policy runs, the documents are retrieved from the source system by the CONTENTSEARCH application, converted to text format if necessary, and searched by using the defined regular expressions. The results of the search are returned to IBM Spectrum Discover and the metadata of the files that are updated. To create a CONTENTSEARCH policy, see [“Adding policies” on page 15](#).

When you create a content search policy, you can select:

- Any tag type (including OPEN, RESTRICTED, and CHARACTERISTIC tags) for the CONTENTSEARCH application.
- Any regular expression.
- Either "True/False" or "Value matching expression". "True/False" sets the tag value to either True or False if a match is or is not found. "Value matching expression" sets the tag value to the extracted content match.

Remember:

- If you select a RESTRICTED tag with a defined set of values and choose to extract the value from a document, the value that is extracted must match one of the RESTRICTED values in the tag.
- If the content extracted exceeds the minimum tag value length, the extracted content is truncated.

Identifying the required regex expressions

The following information can help you identify required regex expressions.

About this task

Validate that the regular expressions that are required for the policy are present. You can create or modify them if necessary.

Procedure

1. On the metadata page, select the **Regular Expression** tab.
The list of available expressions displays.

Name	Description	Regular Expression
EmailID	Matching Email IDs like : John.Smith@example.com	<code>\b([w,.-]+@[w,.-]+\.[w]{2,3})b</code>
IPv4-Address	Matching IPv4 address like: 192.168.1.1	<code>\bd{1,3}\.d{1,3}\.d{1,3}\.d{1,3}b</code>
Dates-MM/DD/YYYY	Matching dates in MM/DD/YYYY format like: 05/21/2019	<code>\b((0 [0-9]) (1 0-2))\(\)\(0-2\)[0-9]((3 0-1))\(\)\d{4}b</code>
Dates-DD/MM/YYYY	Matching dates in DD/MM/YYYY format like: 15/10/2019	<code>\b(0-2)[0-9]((3 0-1))\(\)\(0-9\)[0-9]((1 0-2))\(\)\d{4}b</code>
MasterCard	Matching MasterCard number like: 5258704108753590	<code>\b(?:5[1-5][0-9]{2} 2221-9 22[3-9][0-9] 2[3-6][0-9]{2} 2701[0-9] 2720)[0-9]{12}b</code>
VisaCard	Matching Visa Card numbers like: 4563-7568-5698-4587	<code>\b([4]d{3}[s]d{4})\s\d{4}\s\d{4}\s\d{4}\s\d{4}\s\d{3}[-]\d{4}[-]\d{4}[-]\d{4}\s\d{3}\.d{4}\.d{4}\.d{4}\.d{4}\s\d{4}\s\d{4}\s\d{4}b</code>
AmexCard	Matching American Express Card numbers like: 340000000000009	<code>\b3[47][0-9]{13}b</code>
USZIPCode	Matching United States ZIP codes like: 97589	<code>\b((d{5}-d{4}) (d{5}))\((A-Z)\d[A-Z]\s\d[A-Z]\d)\b</code>
Geo-Coordinate	Matching Geo-Coordinates like: 51.498134, -0.201755	<code>\b([-+]?)\(\d{1,2}\(\(((d+\.))\(((1 ?)\(\d{1,3}\(((d+))?)\b</code>
CanadianSIN	Matching Canadian Social Insurance Number like: 123-456-789	<code>\b\d{3}\s\d{3}\s\d{3}\b \b\d{3}[-]\d{3}[-]\d{3}\b</code>
CreditCardExpirationDate	Matching Credit Card Expiration Date like 11/12	<code>\bd{2}\d{2}b</code>
CVV-Number	Matching Credit Card Verification Value number like: 670, 0927	<code>\b(0-9){3,4}b</code>
Currency	Matching currency like: 123, 25.50	<code>\b(d+(\.d{2})?)b</code>
US-SSN	Matching United States Social Security Numbers (SSN) like: 513-84-7329	<code>\bd{3}-d{2}-d{4}b</code>
URL	Matching URLs like: http://www.test.com/dir/filename.jpg?var1=foo#bar&var2=val2	<code>\b((http[s]? ftp):\/\/?((^\s +)(\w+)*\(\w - _ + ^#? s)+)(*)?#\(\w - _ +)?b</code>

Figure 15. Preinstalled regular expressions

- Search through the list of regular expressions to find any that match the content to be searched. As shown in Figure 15 on page 26, IBM Spectrum Discover includes a selection of regular expressions.

For example, an expression with an embedded value that might be extracted.

```
^([\w\.-]+@[[\w\.-]+\.[\w]{2,3})$
```

This regular expression matches an email address, and the value that is returned for the tag is the matched email address. This sort of regex is appropriate to use in a value type match.

Another example is an expression with no embedded value, for a straight match.

```
^Patient Name:.*$
```

This expression detects the presence of the literal string “Patient Name:” and subsequent string in a fixed format file, but it does not extract the value. This is appropriate for use in a Boolean find search.

- If there are no suitable regular expressions, select the **Create** icon.
- If a regular expression exists but requires modifications select the expression and click the “Modify” icon. If this regular expression is in use by any other policy, this modification affects those policies.
- Enter a suitable name, description, and the regular expression pattern.
- Select **Save**.

Creating a content search policy

For information about creating a content search policy, see Chapter 2, “Managing metadata policies,” on page 15 and “Adding policies” on page 15.

Viewing content search application logs

The following information describes how to view the content search application logs.

About this task

If you want to view the content search application logs, perform the following actions.

Procedure

1. Run the following command:

```
podlog spectrum-discover-contentsearchagent
```

2. Any failures of download or inspection are logged in the application log file.

In the following example, the file fails during the inspection stage. The message shows that the failure is because the application cannot contact the Tika server.

```
[2019-04-23,16:45:54.945] agent[22665][ERROR][INSPECT]: Inspection failure-origpath
metaocean1/alice.txt - error((HTTPConnectionPool(host='localhost',port=9998):Max retries
exceeded
with url:/tika(Caused by NewConnectionError('<urllib3.connection.HTTPConnection object at
0x7f4db813e0d0>:Failed to establish a new connection:[Errno 111] Connection refused',)))
```

```
kubectl logs $(kubectl get pods -n=spectrum-discover -l app=spectrum-discover-
contentsearchagent -o=jsonpath={.items[0].metadata.name}) -n spectrum-discover
```

Hints and tips for using content search

Following are some best practices for using the content search feature.

Testing on a subset of documents

Running a content search policy on a set of documents has several steps, including retrieving the document, formatting it as text, if necessary, and searching the document. Depending on the number, formatting, and size of the documents, searching the document can be a time-consuming process.

Therefore, it is best to test the policy and corresponding expressions on a small subset of documents to determine whether the policy and the regular expressions you select are correct. One way to run the test is to use a policy filter that selects only a small set of documents. After you confirm that the policy and search criteria is operating as expected, you can run it against the required set of documents.

The test on a subset of documents can also help you estimate how long the policy might take to run on the complete set of documents.

Avoiding retagging

When you rerun a policy against a set of documents that is previously tagged, the documents are retagged. If the values returned are different than the previous search, they are updated. This difference might occur if the policy or the set of expressions is modified, or if the set of documents is modified.

To avoid retagging the documents, add a criteria to the filter to not select documents that are already tagged.

Modifying regular expressions

If you modify a regular expression, it affects all policies that use that expression. Rerunning these policies might cause the documents to be tagged differently. To avoid changing the behavior of existing policies, create a new regular expression and use it in the specific policies where it is required.

Converting files with Apache Tika

IBM Spectrum Discover uses Apache Tika to convert files to text before it searches the content. This conversion has an impact on the overall content search performance.

Therefore, files that are text format must be configured in the `contentsearch` agent to prevent them from being processed unnecessarily by Apache Tika. The default configuration treats JavaScript Object Notation (JSON) and Variant Call Format (VCF) file types as text. To add more text file types to the configuration, edit the file:

```
/gpfs/gpfs0/agents/contentsearch/conf/contentsearch.conf
```

And add more types to the line:

```
text_filetypes=vcf,json
```

Apache Tika runs in a Kubernetes pod within IBM Spectrum Discover. You can increase Apache Tika pod instances to improve performance. For example, run this command to scale the number of Tika instances to three:

```
kubectl -n spectrum-discover scale --replicas=3  
deploy/spectrum-discover-tikaserver
```

Apache Tika is resource-intensive, so make sure that the number of Apache Tika instances does not exceed the host resources.

Supported connection types

Content search on IBM Spectrum Discover supports the following connection types.

- Spectrum Scale
- COS
- NFS
- S3
- [\[SMB\]](#)

For more information, see the topic *IBM Spectrum Scale data source connection* in *IBM Spectrum Discover: Concepts, Planning, and Deployment Guide*.

Chapter 4. Tiering data by using ScaleILM application

Use the IBM Spectrum Discover ScaleILM application to move data to different tiers (pools) that are configured on the IBM Spectrum Scale connection.

Before you begin

- Configure IBM Spectrum Scale with one or more internal pools. For more information, see *Internal Storage pools* in the *IBM Spectrum Scale: Administration Guide*.
- If you want to, you can configure IBM Spectrum Archive and its pools.
- Create the IBM Spectrum Scale connection in IBM Spectrum Discover. If the external pool that is managed by IBM Spectrum Archive is used as the destination tier, then the "host" setting of IBM Spectrum Scale connection must specify one of the IBM Spectrum Archive nodes.
- IBM Spectrum Discover ScaleILM application uses the same user, that is configured in IBM Spectrum Discover to scan the IBM Spectrum Scale and run the Data Movement policies. For this data movement, the IBM Spectrum Scale connection is used as the 'source_connection'. For more information, see the topic *Creating or identifying a user ID and password for scanning* in the *IBM Spectrum Discover: Concepts, Planning, and Deployment Guide*.

About this task

The IBM Spectrum Discover ScaleILM application provides the advanced data tiering function through the IBM Spectrum Scale connection. It uses the system metadata and custom metadata of documents that are collected by IBM Spectrum Discover for this advanced data tiering function. This capability eliminates the need to scan file systems during IBM License Manager (ILM) execution. It also provides the cognitive tiering capability by using the results of IBM Spectrum Discover's AUTOTAG, Content Search, and Deep Inspect policies.

IBM Spectrum Scale provides the built-in Information Lifecycle Management (ILM) capability that optimizes the cost-effectiveness of data by moving the physical location of data between the storage pools with different cost or performance characteristics. IBM Spectrum Scale's ILM supports the data movement between its internal storage pools and between both the internal pool and the external storage repository. The movement from external storage repository is managed by external applications, such as IBM Spectrum Archive Enterprise Edition (EE) and IBM Spectrum Protect for Space Management.

Make sure that you know the IBM Spectrum Archive Enterprise Edition (EE) software versions that are supported. For more information, see the topic *IBM Spectrum Storage software requirements* in the *IBM Spectrum Discover: Concepts, Planning, and Deployment Guide*.

While tiering data in external pools, the current location of the data is indicated in its file state. The data locations and the corresponding file states that indicate these locations, are listed in the following table:

Data locations	File states
The data location is only in the internal pool.	The file state is resident.
The data location is both internal and external pool.	The file state is premigrated.
The data location is only in the external pool.	The file state is migrated.

Note: The functions of the ScaleILM application depend on the capability of underlying storage hardware and its management software. For more information, see the following topics:

- *Information Lifecycle Management for IBM Spectrum Scale* in the *IBM Spectrum Scale: Administration Guide*

- *Introduction to IBM Spectrum Archive Enterprise Edition (EE) in the IBM Spectrum Archive Enterprise Edition (EE) IBM Knowledge Center*

Procedure

1. Log in to IBM Spectrum Discover GUI.
2. Go to **Admin > Management Policies** page.
3. Click **Add Policy** to create a policy.
4. Click the slider control and set the status to one of the following values:

Active

An active policy is run whenever its scheduling event is reached.

Inactive

An inactive policy is not run even when its scheduling event is reached, including the Now event.

5. Enter a policy name.
6. Enter a Policy filter. The policy filter includes the criteria for selecting the files for tiering, such as `filetype="pdf"`.

Note: If the destination tier is the external pool that is managed by IBM Spectrum Archive, define the policy filter criteria as "state in 'resdnt'". For example, `filetype = 'txt' and state in 'resdnt'`.

If the filter criteria does not include `state in 'resdnt'` while tiering data to an external pool, the ScaleILM application skips the files that are not in `resdnt` state.

7. Click **Next Step** to select the type of policy.
8. Select TIER as the **Policy type**.
9. Select ScaleILM as the **Agent**.
10. Select the `source_connection` from the drop-down list. The source connection is the source from where the data is being moved, such as the name of the defined IBM Spectrum Scale connection.
11. Enter the `destination_tier` where you want to move your data, such as:

- Internal Pools
 - Gold, silver, bronze, flash system
- External Pools
 - `archive:pool1@library1`
 - `archive:pool2@library2`
 - `archive:pool1@library1, pool2@library3`

Note: When you specify an IBM Spectrum Scale internal pool as the "destination_tier", you need to ensure that you specify a valid internal pool name. That internal pool name must be configured in the IBM Spectrum Scale source connection for the corresponding data source (file system). These internal pools can be listed by using the following IBM Spectrum Scale command

```
mmlspool <device> all
```

When you specify an external pool that is managed by IBM Spectrum Archive as the "destination_tier", you must specify a valid archive pool. This archive pool must be defined in the IBM Spectrum Archive that is configured on the IBM Spectrum Scale cluster node.

Additionally, you must specify the name of the pools, that are defined in the IBM Spectrum Archive configuration, with the prefix "archive:". The pool names must be specified in the same format as defined in the `-p` option of IBM Spectrum Archive Enterprise Edition (EE) CLI (`eadm`). The syntax must be as follows:

```
archive:<poolName>@<libraryName>
```

or

```
archive:<poolName1>@<libraryName1>,<poolName2>@<libraryName2>, ...
```

The policy execution fails when you do not follow the instructions.

12. Select **Next Step** to enter a schedule. The schedule indicates when you want to start the tiering.
13. Select **Next Step** to review the policy.
14. Select **Submit** to create the policy.
15. When the policy is created, view the files on the IBM Spectrum Discover Search catalog page to ensure that they are moved to the new tier. The following metadata are updated.

Tier

Displays the name of the internal pool in IBM Spectrum Scale where the data is stored.

Note: Even if the file is in the migrated state (migrted), the tier field shows the name of the original internal pool.

State

Displays the current state as one of the following values:

- resdnt
- premig
- migrted

Migloc

Displays the location information of the external pool when the file is in premigrated (premig) or migrated (migrted) state. If the file is in a resident (resdnt) state, this field shows NA.

SizeConsumed

Displays the actual size of the file in bytes, that is associated with the IBM Spectrum Scale file system field **SizeConsumed Bytes**.

Viewing ScaleILM application logs

The following section describes how to view the ScaleILM Data Mover application logs.

About this task

Follow the procedure to view the ScaleILM Data Mover application logs.

Procedure

1. Run the following command:

```
podlog spectrum-discover-scaleilmdatamover
```

2. Any failures that occur while the files are processed for tiering by ScaleILM Data Mover application, are logged in the application log file. A sample error log from the ScaleILM Data Mover application logs is shown in the following section. These error logs occur when the file is not on the specified 'source_connection' while the application is processing the Tiering policy.

```
2020-04-28 14:44:36,439 - __main__ - ERROR - File not found: /ibm/scale0/unicode_test/migrateDir.popFSDir.19824/unicodedir/test010.\n.test
2020-04-28 14:44:36,440 - __main__ - ERROR - File not found: /ibm/scale0/unicode_test/migrateDir.popFSDir.19824/unicodedir/test137..test
2020-04-28 14:44:36,440 - __main__ - ERROR - File not found: /ibm/scale0/unicode_test/migrateDir.popFSDir.19824/unicodedir/test147..test
2020-04-28 14:44:36,440 - __main__ - ERROR - File not found: /ibm/scale0/unicode_test/migrateDir.popFSDir.19824/unicodedir/test157..test
```

A sample error log from the ScaleILM Data Mover application logs for an invalid 'destination_tier', is shown in the following section:

```
2020-04-28 14:35:39,566 - __main__ - ERROR - Invalid destination_tier: 'gold:abcd'.
```

A sample error log from the ScaleILM Data Mover application logs when the 'destination_tier' does not exist on the IBM Spectrum Scale connection, is shown in the following section:

```
2020-05-12 14:32:38,436 - __main__ - ERROR - Provided internal pool goldflower does not exist on system: 9.11.212.29
```

3. Run the following command to view policy engine pod logs:

```
podlog spectrum-discover-policyengine
```

A sample policy engine pod log is shown in the following section. This log shows information that is received from the ScaleILM Data Mover application. It describes errors that are encountered while a Tiering policy is processed and when a file is not found on the specified 'source_connection':

```
2020-04-28 14:44:37,455 - policy.policyapiservice - ERROR - [policy_id: scale-tier-unicode-modevmm19-pol]: Agent reported 'status: failed' for fkey: 'modevmm19.tuc.stglabs.ibm.comscale0121547', path: '/ibm/scale0/unicode_test/migrateDir.popFSDir.19824/unicodedir/test237..test', reason: 'File not found: /ibm/scale0/unicode_test/migrateDir.popFSDir.19824/unicodedir/test237..test', errno: 'ENOENT'
2020-04-28 14:44:37,455 - policy.policyapiservice - ERROR - [policy_id: scale-tier-unicode-modevmm19-pol]: Agent reported 'status: failed' for fkey: 'modevmm19.tuc.stglabs.ibm.comscale0121557', path: '/ibm/scale0/unicode_test/migrateDir.popFSDir.19824/unicodedir/test247..test', reason: 'File not found: /ibm/scale0/unicode_test/migrateDir.popFSDir.19824/unicodedir/test247..test', errno: 'ENOENT'
```

A sample error log from the Policy engine pod logs for an invalid 'destination_tier', is shown in the following section:

```
2020-04-28 14:35:40,064 - policy.policyapiservice - ERROR - [policy_id: scale-tier-modevmm19-invalid-destn-pol]: Agent reported 'status: failed' for fkey: 'modevmm19.tuc.stglabs.ibm.comscale0300770', path: '/ibm/scale0/covid-19-dataset/inference/NWPU-Chest-X-Ray/2b16346f-8fe4-45ee-b3c7-aaa46fd26041.png', reason: 'Invalid destination_tier: 'gold:abcd'.', errno: 'EINVAL'
```

A sample error log from the Policy engine pod logs when the 'destination_tier' does not exist on the IBM Spectrum Scale connection, is shown in the following section:

```
2020-05-12 07:40:34,819 - policy.policyapiservice - ERROR - [policy_id: scale-tier-pol]: Agent reported 'status: failed' for fkey: 'nikkogpfs26728', path: '/ibm/gpfs2/scale0/customer_e2e/financial_report_file6.pdf', reason: 'Provided internal pool goldflower does not exist on system: 9.11.212.29', errno: 'EINVAL'
2020-05-12 07:40:34,819 - policy.policyapiservice - ERROR - [policy_id: scale-tier-pol]: Agent reported 'status: failed' for fkey: 'nikkogpfs26748', path: '/ibm/gpfs2/scale0/customer_e2e/HR_report_US_file6.pdf', reason: 'Provided internal pool goldflower does not exist on system: 9.11.212.29', errno: 'EINVAL'
```

Chapter 5. Exporting Metadata to IBM Watson Knowledge Catalog

IBM Spectrum Discover Watson Knowledge Catalog (WKC) connector supports the export of metadata to both the IBM Cloud instance of Watson Knowledge Catalog, and to an On-Premise Instance.

Note: IBM Spectrum Discover supports export to a single IBM Watson Knowledge Catalog instance and only one type of connection per instance.

The default deployment of the WKC connector contains the parameters that are associated with the IBM Cloud instance of Watson Knowledge Catalog. On startup, the IBM Spectrum Discover WKC connector checks for credential values in the following environment variables:

```
WKC_API_KEY
WKC_USER
WKC_PASSWORD
```

In the absence of appropriate credentials, IBM Spectrum Discover WKC connector goes into a sleep loop. In such a scenario, the **Export** button is not visible on the IBM Spectrum Discover user interface.

Exporting metadata to IBM Cloud Watson Knowledge Catalog

Export data along with relevant metadata tags from IBM Spectrum Discover to Watson™ Knowledge Catalog.

Before you begin

You must follow the procedure to add the IBM Cloud Watson catalog details to IBM Spectrum Discover.

1. Obtain the API key for accessing the Watson Knowledge Catalog (WKC). For WKC in IBM Cloud, go to <https://cloud.ibm.com/iam/overview> and select **Create an IBM Cloud API key**.
2. Export or copy the key.
3. Copy the base URI of your WKC instance. For IBM Cloud, the WKC instance URI is based on the geographical location. For example,

```
https://api.dataplatform.cloud.ibm.com/v2/
```

4. To configure the deployment parameters for the WKC connector app, log in to the IBM Spectrum Discover instance and run the following command:

```
kubectl -n spectrum-discover edit deploy/spectrum-discover-wkconnector
```

5. In the deployment editor, search and edit the environment variable *WKC_API_KEY*.
6. Add a value string with the API key that is acquired from IBM Cloud in step 2. For example,

```
name:WKC_API_KEY
```

```
value:<API-KEY-VALUE>
```

Note: You need to ensure that you maintain the correct indentation while you are editing in the deployment editor.

7. Set the base URI of the WKC instance to the one copied in step 3. For example:

```
name: WKC_BASE_URI
```

```
value:https://api.dataplatform.cloud.ibm.com/v2/
```

Note: If you switch to a different IBM Cloud account, you can edit the `WKC_API_KEY` environment variable in the deployment editor with the API key that is associated with the new account. The application automatically restarts and identifies the new account that is linked.

If you add new catalogs to your WKC instance, IBM Spectrum Discover retains the old registration information and continues to point to the old catalog IDs. To rectify this issue, follow the procedure:

- a. Go to **Metadata > Applications**.
- b. Restart the WKC application instance by using the following commands:

```
kubectl scale deployment --replicas=0 spectrum-discover-wkconnector
```

```
kubectl scale deployment --replicas=1 spectrum-discover-wkconnector
```

If you add a connection to IBM Spectrum Discover, then the WKC application has to be restarted to be able to use that connection. Use the following commands to restart the WKC:

```
kubectl scale deployment --replicas=0 spectrum-discover-wkconnector
```

```
kubectl scale deployment --replicas=1 spectrum-discover-wkconnector
```

About this task

The Watson Knowledge Catalog is a data cataloging system that is not always able to scan relevant data sources and capture the relevant metadata from those files. IBM Spectrum Discover helps to bridge this critical gap by helping to export data to Watson Knowledge Catalog with all relevant metadata tags.

Note: [If the source connection that IBM Spectrum Discover is accessing to export data is the one that WKC can also connect to, then you can configure the connection map within the WKC Connector App. For more information, see [“Mapping similar source connections in Watson Knowledge Catalog”](#) on page 37.]

Procedure

1. On the IBM Spectrum Discover Dashboard, search for the data to be exported by using a specific filter criteria.
2. Click **Export Data**. The **Export Data to Watson Knowledge Catalog** window appears.
3. Under **Destination Catalog**, select the catalog in Watson Knowledge Catalog where you want to export the data.
4. Select the tags that you want to export from the list in **Metadata Tags to Export**.
5. Click **Submit**.
6. After completion of the process, the exported data is displayed in the Watson Knowledge Catalog with the tags that are imported from IBM Spectrum Discover.

Note: Tags in IBM Spectrum Discover represent a name (for example, SizeRange) and a value (for example, small, large, or medium). In Watson Knowledge Catalog, the tags represent a value. The exported data, maps both of these attributes and it creates a single label. For example, SizeRange:Small.

Exporting metadata to IBM on-premises Watson Knowledge Catalog

Export data along with relevant metadata tags from IBM Spectrum Discover to IBM on-premises Watson Knowledge Catalog.

Before you begin

You must follow the procedure to add the IBM on premises Watson knowledge catalog details to IBM Spectrum Discover.

1. Obtain the values for the following Watson Knowledge Catalog (WKC) parameters from the systems administrator of the On-Premises instance.

```
WKC_USER
WKC_PASSWORD
WKC_BASE_URI
WKC_AUTH_URI
```

2. To configure the deployment parameters for the WKC connector, log in to the IBM Spectrum Discover instance and run the following command.

```
kubectl -n spectrum-discover edit deploy/spectrum-discover-wkconnector
```

3. In the deployment editor, add the parameter details in the sample format as shown

```
name: WKC_BASE_URI
  value: https://<wkc_hostname>/v2/
name: WKC_AUTH_URI
  value: https:// <wkc_hostname>/icp4d-api/v1/authorize
name: WKC_USER
  value: <wkc_admin_username>
name: WKC_PASSWORD
  value: <wkc_admin_password>
```

Note: If you switch to a different IBM on-premises account, you can edit the relevant WKC environment variables in the deployment editor to add the values that are associated with the new account. The application automatically restarts and identifies the new account that is linked.

If you add new catalogs to your WKC instance, IBM Spectrum Discover retains the old registration information and continues to point to the old catalog IDs. To rectify this issue, follow the procedure:

- a. Go to **Metadata > Applications**.
- b. Restart the WKC application instance by using the following commands:

```
kubectl scale deployment --replicas=0 spectrum-discover-wkconnector
```

```
kubectl scale deployment --replicas=1 spectrum-discover-wkconnector
```

If you add a connection to IBM Spectrum Discover, then the WKC application must be restarted to be able to use that connection. Use the following commands to restart the WKC:

```
kubectl scale deployment --replicas=0 spectrum-discover-wkconnector
```

```
kubectl scale deployment --replicas=1 spectrum-discover-wkconnector
```

About this task

The Watson Knowledge Catalog is a data cataloging system that is not always able to scan relevant data sources and capture the relevant metadata from those files. IBM Spectrum Discover helps to bridge this critical gap by helping to export data to Watson Knowledge Catalog with all relevant metadata tags.

Note: [If the source connection that IBM Spectrum Discover is accessing to export data is the one that WKC can also connect to, then you can configure the connection map within the WKC Connector App. For more information, see [“Mapping similar source connections in Watson Knowledge Catalog” on page 37.](#)]

Procedure

1. On the IBM Spectrum Discover Dashboard, search for the data to be exported by using a specific filter criteria.
2. Click **Export Data**. The **Export Data to Watson Knowledge Catalog** window appears.
3. Under **Destination Catalog**, select the catalog in Watson Knowledge Catalog where you want to export the data.

4. Select the tags that you want to export from the list in **Metadata Tags to Export**.
5. Click **Submit**.
6. After completion of the process, the exported data is displayed in the Watson Knowledge Catalog with the tags that are imported from IBM Spectrum Discover.

Note: Tags in IBM Spectrum Discover represent a name (for example, SizeRange) and a value (for example, small, large, or medium). In Watson Knowledge Catalog, the tags represent a value. The exported data, maps both of these attributes and it creates a single label. For example, SizeRange:Small.

Exporting metadata from linked and non-linked data sources

Exporting metadata from IBM Spectrum Discover to the Watson Knowledge Catalog can be done by using either a linked or a non-linked data source.

To successfully export document-related metadata from IBM Spectrum Discover to Watson Knowledge Catalog (WKC), WKC must have access to the original documents.

Watson Knowledge Catalog can use the following options to access the target data source.

Exporting metadata from linked IBM Spectrum Discover data source

Exporting metadata from a linked data source indicates that both Watson Knowledge Catalog and IBM Spectrum Discover have access to the original data source location, eliminating the need to copy the original files.

Before you begin

- On Watson Knowledge Catalog, create a connection to the target data source.
- On the IBM Spectrum Discover, connect to the Watson Knowledge Catalog and collect the catalog's connection, including the catalog's connection details.
- On the IBM Spectrum Discover, scan the target data source and extract the metadata.

Procedure

1. Create an implicit policy in IBM Spectrum Discover.
2. Select the metadata to be exported to Watson Knowledge Catalog.
3. On the IBM Spectrum Discover user interface, click **Export**.

IBM Spectrum Discover exports the metadata and associates it with the documents that Watson Knowledge Catalog (WKC) can also access.

Exporting metadata from non-linked IBM Spectrum Discover data source

When Watson Knowledge Catalog does not have access to original IBM Spectrum Discover data source location, IBM Spectrum Discover copies the files that are associated with the export metadata to a location where the Watson Knowledge Catalog service can access these documents.

Before you begin

Note:

The process to export metadata from a non-linked data source must be used only when it is not possible for Watson Knowledge Catalog to connect to the same target data source that is used by IBM Spectrum Discover.

- On Watson Knowledge Catalog, create a connection to a IBM Cloud Object Storage(COS) account.
- On IBM Spectrum Discover, connect to the Watson Knowledge Catalog and collect the catalog's details that include the catalog's connection details.
- On IBM Spectrum Discover, scan the target data source and extract the metadata.

- Ensure that IBM Spectrum Discover can connect to the corresponding Watson Knowledge Catalog COS storage locations.

Procedure

1. Create an implicit policy in IBM Spectrum Discover.
2. Select the metadata to be exported to Watson Knowledge Catalog .
3. On the IBM Spectrum Discover user interface, click **Export**.

IBM Spectrum Discover exports the metadata and simultaneously copies the corresponding original data source files, to the IBM Cloud Object Storage location associated with the catalog.

[Mapping similar source connections in Watson Knowledge Catalog

IBM Spectrum Discover supports mapping source connections with Watson Knowledge Catalog connections (WKC) through WKC connector App.

When metadata is being exported to WKC, IBM Spectrum Discover might use a source connection that WKC can also connect to. In such a scenario, you can configure the connection mapping within the WKC Connector App.

The connections that can be linked to are:

- Amazon S3
- IBM Cloud Object Storage
- IBM Spectrum Scale

Note: Connections with IBM Spectrum Scale are established through an S3 connection as WKC does not support IBM Spectrum Scale directly.

The details for configuring the connection maps with each of these source connections are described.

S3

For an IBM Spectrum Discover S3 connection, the WKC connection must contain the following details:

Bucket

If the bucket name is configured, then you do not need to provide any further configuration details. The WKC Connector App can infer the details from the global namespace of Amazon S3 buckets.

If the bucket name is not provided, then configure the *WKC_Connection_Map* environment variable by using the following format: `<datasource>;<cluster>:<wkc connection name>`. A sample variable value is shown. All documents, that the WKC App receives in its work message corresponding to the data source and cluster pair that is defined in the variable, maps to the WKC connection of that name.

```
WKC_CONNECTION_MAP=testbucket1.sd.ibm.com;s3.eu-west-1.amazonaws.com:s3_con_no_bucket
```

Note: It is mandatory to provide the bucket name while you are configuring details in IBM Spectrum Discover but it is optional for WKC.

IBM Cloud Object Storage Infrastructure

For an IBM Spectrum Discover IBM Cloud Object Storage connection the WKC connection must contain the following details:

Login URL

The login URL must be that of the accessor defined in IBM Spectrum Discover. The data source for IBM Cloud Object Storage is the vault. However, since it is not possible to provide a vault here, you

must provide the mapping within the environment variable `WKC_CONNECTION_MAP` in the following format: `<datasource>;<cluster>:<wkc connection name>`. A sample variable value is shown.

```
WKC_CONNECTION_MAP=vault1;e09cdac0-80f8-73be-00ed-cb8edeede242:local_cos_con
```

Multiple IBM Cloud Object Storage connections to the same system can map to the same WKC connection (as it is at a higher level and can see all vaults).

IBM Spectrum Scale

Connection mapping with IBM Spectrum Scale must be done through an S3 connection as WKC cannot connect directly with it.

To establish an S3 connection, configure the following details in the WKC connector app:

Endpoint URL

The S3 Endpoint URL on the IBM Spectrum Scale mode. For example, `http://modevvm19.tuc.stglabs.ibm.com:9000`.

Access Key

Type the S3 access key.

Secret Key

Type the S3 secret key.

Bucket

Do not enter values in the bucket field.

Define the mapping within the environment variable `WKC_CONNECTION_MAP` in the following format: `<datasource>;<cluster>:<wkc connection name>`. A sample variable value is shown.

```
WKC_CONNECTION_MAP=scale0;modevvm19.tuc.stglabs.ibm.com:s3_scale
```

Mapping multiple connections

You can map multiple connections within the same environment variable by using commas to separate the values. A sample is shown:

```
WKC_CONNECTION_MAP=vault1;e09cdac0-80f8-73be-00ed-cb8edeede242:local_cos_con,scale0;modevvm19.tuc.stglabs.ibm.com:s3_scale,testbucket1.sd.ibm.com;s3.eu-west-1.amazonaws.com:s3_con_no_bucket
```

Run the following command to edit the WKC Connector app deployment and add the mapping:

```
kubect1 -n spectrum-discover edit deploy/spectrum-discover-wkconnector
```

The WKC connections are configured in the following format:

```
name: WKC_CONNECTION_MAP
value: vault1;e09cdac0-80f8-73be-00ed-cb8edeede242:local_cos_con,scale0;modevvm19.tuc.stglabs.ibm.com:s3_scale,testbucket1.sd.ibm.com;s3.eu-west-1.amazonaws.com:s3_con_no_bucket
```

Note: The **edit** command configures the WKC connection mapping and automatically restarts the WKC pod.

]

Troubleshooting export issues

This topic describes some issues faced while exporting data by using the Watson Knowledge Catalog.

Authentication failure with Watson Knowledge Catalog - both on-Premise and IBM Cloud

Resolve authentication failure with the Watson Knowledge Catalog (WKC).

The following error is displayed on the Watson Knowledge Catalog connector logs, when the connector fails to authenticate with the Watson Knowledge Catalog instance.

```
2020-07-16 12:46:04,870 - WKCCconnector - ERROR - User authentication failed with response code 401
2020-07-16 12:46:04,871 - WKCCconnector - INFO - No valid token available, cannot connect to WKC. Please set API Key or User Credentials
```

When the WKC authentication fails, check to ensure that the correct values are defined for the following environment variables for the on-premises connection:

```
WKC_USER, WKC_PASSWORD
```

If you are using an IBM Cloud connection type, you must check the value of the environment variable `WKC_API_KEY`.

Run the following command to check and confirm the configuration for the specified environment variables in the pod helm chart: **kubectl edit deployment spectrum-discover-wkccconnector**

Incorrect WKC URL configuration

Resolve the errors that occur due to an incorrect WKC URL configuration.

The following errors occur:

```
HTTPSConnectionPool(host='machine.ibm.com', port=443): Max retries exceeded with url: /v2/catalogs (Caused by NewConnectionError('<urllib3.connection.HTTPSConnection object at 0x7f63e7973c18>: Failed to establish a new connection: [Errno -2] Name or service not known',))
2020-07-09 22:59:29,412 - WKCCconnector - ERROR - Cannot communicate with WKC to retrieve catalog information. Check config URIs.
```

A possible reason for the error is that invalid values are defined for the environment variable `WKC_BASE_URI`. You can get similar errors if you define incorrect values for the `WKC_AUTH_URI` environment variable.

Run the following command to check and confirm the configuration of the said environment variables in the pod helm chart: **kubectl edit deployment spectrum-discover-wkccconnector**.

Invalid connection type in catalog

Resolve the error that occurs due to an invalid connection type in catalog.

When a valid connection type cannot be detected, the following error occurs.

```
2020-06-30 22:37:09,628 - WKCCconnector - ERROR - Could not find a supported connection in the supplied catalog
```

This error might occur in the scenario when the connection type within the Watson Knowledge catalog (WKC) does not match the connection type on Spectrum Discover. For example, if you try to export metadata from an IBM Spectrum Discover Cloud Object Storage (COS) source to a WKC catalog, which does not have an IBM Cloud Object Storage connection. The error can also occur if you have an IBM Cloud Object Storage bucket on WKC mapped to an on-Premises IBM Cloud Object Storage bucket that is configured on IBM Spectrum Discover.

No linked connection type

Resolve the errors that occur when links cannot be identified for the connection type.

The following errors can occur resulting in an export failure:

```
2020-07-01 13:58:34,129 - WKCCconnector - ERROR - Cannot export /Changing ACEs/Not for Luke and Amy/Full Gamora/Partial Luke/Partial Amy/No Clara/I mean not clara/TinyDoc.txt as document is not on a linked connection, and was unable to be downloaded from the SD connection for copy
```

```
2020-07-01 13:58:34,129 - WKConnector - ERROR - Failed to export /Changing ACEs/Not for Luke and Amy/Full Gamora/Partial Luke/Partial Amy/No Clara/I mean not clara/TinyDoc.txt
```

An exported document must either be on a linked connection, or must be available for copying to backup IBM Cloud Object Storage in the WKC catalog, if that exists.

WKC connector pod in CrashLoopBackoff state

Check the following settings if the Watson Connector pod is in the CrashLoopBackoff or Error state.

Check the settings for the following environment variables if you are using IBM Cloud Watson Knowledge Catalog.

```
WKC_API_KEY
```

Check the settings for the following environment variables if you are using an on-premises Watson Knowledge Catalog .

```
WKC_BASE_URI
WKC_AUTH_URI
WKC_USER
WKC_PASSWORD
WKC_CONNECTION_MAP
```

After successfully exporting records, if you change an environment variable to an invalid value and then export records again, you can end up with unprocessed messages in the Kafka buffer. You must clear these messages before you can change the invalid environment variable to the correct value. Run the following commands to clear the Kafka queue and then restart the Watson Knowledge Catalog connector application.

Clear Queue

```
/opt/kafka/bin/kafka-topics.sh --delete --zookeeper localhost:2181 --topic WKConnector_work
```

Recycle Pod

```
kubectl scale deployment --replicas=1 spectrum-discover-wkconnector
```

```
kubectl scale deployment --replicas=0 spectrum-discover-wkconnector
```

S3 connection issues

Resolve the issues faced owing to S3 connection.

To enable export of IBM Spectrum Scale file metadata without copying the actual file to Watson Knowledge Catalog, the latter must be connected to IBM Spectrum Scale through an S3 interface.

If you are experiencing issues with IBM Spectrum Scale file exports, check whether the connection is present in the connector connection map and that the details are made available in Watson Knowledge Catalog. You must check to see that correct values are defined for the IP, port, access, and secret key. To preview the documents in the catalog, you must check that the S3 interface is available on the IBM Spectrum Scale system.

Chapter 6. Managing tags

A tag is a custom metadata field that is used to supplement storage system metadata with organization-specific information. For instance, an organization might segment their storage by project or by chargeback department. Those facets do not show up in the system metadata. Additionally, the storage systems themselves do not provide management and reporting capabilities based on those organizational concepts. Use custom tags to store additional information and manage, report, or search for data by using that organizationally important information.

Permissions

Security Administrators

Cannot create, update, delete, or, list any type of tag.

Data Administrators

Create, modify, delete, and list **Open**, **Restricted**, and **Characteristic** types of tags.

Data Users

Can list any type of tag, and can create and modify **Characteristic** tags.

Cannot create, modify, delete **Open** and **Restricted** tags.

Types of tags

Categorization

Categorization tags contain values such as project, department, and security classification. **Open** and **Restricted** type of tags are **Categorization** tags. Size limit is 256 bytes.

Characteristic

Characteristic tags can contain any value that is needed to describe or classify the object. Can contain lengthy values. Size limit is 4 KB.

Important: You cannot use group values when you search for characteristics. Use this tag specifically, for values that are not grouped.

Creating tags

Use the following information when you create tags.

About this task

Use the **Tags** page to create new organizational tags. The table lists the tag name in the **Field Name** column, tag **Type**, and the tag values in the **Tags** column. Use the icons to **Edit** or **Delete** a tag.

Procedure

1. Go to **Metadata > Tags**
2. Click the **Add** button.

Field Name	Type	Tags	Edit/Delete
COLLECTION	Open		
TEMPERATURE	Open		

20 items per page | 1-2 of 2 items 1 of 1 pages 1

Figure 16. Tags table

- Enter the name of the tag in the **Name** field.

×

New Organizational Tags

COLLECTION

Type
Open

Values
Press "Enter" key to add the tag to the list

old

new ×

Cancel Submit

Figure 17. New Organizational Tags

- Select one of the following values from the **Type** menu:

Open

An **Open** tag can be anything that describes groups of records, but is non-restricted in value, such as project name, department, and sensor serial number.

Restricted

A **Restricted** tag can be anything that describes groups of records, but is restricted to a set of pre-defined values, such as data classification or billing department number.

Characteristics

A **Characteristics** tag is something that is specific in value for each record. They are typically used for content extraction, such as patient name, VIN, or GPS location.

- Enter one or more values for the tag into the **Values** box.
Press the **Enter** key to save each value. Each saved tag is displayed below the **Values** box.
- Click the **Submit** button.
The tags, types, and values are displayed in the table in the **Tags** tab.

Viewing and searching tags

About this task

You can see a list of all tags or search for a subset of them on the **Tags** tab of the **Metadata** page.

Procedure

1. Go to **Metadata > Tags**.
2. A listing of tag **Names**, tag **Types**, and tag **Values** displays.
3. Click the headings of each column to sort in ascending or descending alphabetical order.
4. Enter text into the **Search** box to find tags that begin with the text.
As you enter text, a subset of the tags that contain the text string is automatically displayed.
5. **Edit** or **Delete** a tag by clicking the appropriate icon at the end of the row.

Editing tags

About this task

You can edit tags on the **Tags** tab of the **Metadata** page.

Procedure

1. Go to **Metadata > Tags**
2. Click the **Edit** "Pencil" icon at the end of the row that contains the tag to be edited.
3. The **Modify Organizational Tags** box displays with the **Name** and **Type** unavailable. You cannot change these fields.
4. Remove a tag value by clicking the value displayed in the blue bubbles.
5. Enter one or more values for the tag into the **Values** box.
Press the **Enter** key to save each value.
6. Click the **Submit** button.
The tags, types, and modified values are listed in the table in the **Tags** tab.

Deleting tags

About this task

You can delete tags on the **Tags** tab of the **Metadata** page.

Procedure

1. Go to **Metadata > Tags**.
2. Find a tag by using the **Search** box, by sorting a column, or by navigating by using the page arrows at the bottom of the table.
3. Click the **Delete** "trashcan" icon next to the tag that you want to delete.
4. Click **Delete** in the confirmation box.
The tag is removed from the table in the **Tags** tab.

Chapter 7. Discover data

By discovering your data, you can apply policies that assign tags to your data. You can apply tags to the results of a single search, or you can use policies to automatically apply tags on a periodic basis.

There are three ways to discover data:

- Content-based keyword and tagging. The search is based on regular expression patterns that are defined within IBM Spectrum Discover. For more information, see [“Creating a content search policy” on page 26](#).
- Create a policy by using tags with known values. A policy is automatically run against all data that meets criteria that are specified in a filter. For more information about creating and using policies, see [Searching](#).
- Search your data by using a query in standard SQL grammar or do a visual exploration of tags by point-and-click. For more information, see [“Searching system and custom metadata fields” on page 52](#).

Searching

Procedure

1. Perform the following steps:

- Navigate to **Start a visual exploration** to build your query.
 - a. Check one or more categories to search. Click the **Go** "circle-arrow" icon on the right side of the window to expand the categories.

Discover what's in your Data Search

or start a visual exploration

<input checked="" type="checkbox"/> Cluster	<input type="checkbox"/> Datasource	<input checked="" type="checkbox"/> Owner
<input checked="" type="checkbox"/> Platform	<input type="checkbox"/> Site	<input type="checkbox"/> Tier
<input type="checkbox"/> SizeRange	<input type="checkbox"/> TimeSinceAccess	<input type="checkbox"/> COLLECTION
<input type="checkbox"/> TEMPERATURE	<input type="checkbox"/> UnusedTag	<input type="checkbox"/> Project
<input type="checkbox"/> Department	<input type="checkbox"/> Classification	

Figure 18. Start a visual exploration

- b. Check one or more boxes in the list of groups, policies, and tags. The following figure shows examples. Your data might be different. Then, click the **Go** "circle-arrow" icon on the right side of the window. The valid values for the groups, tags, and policies you selected are displayed.

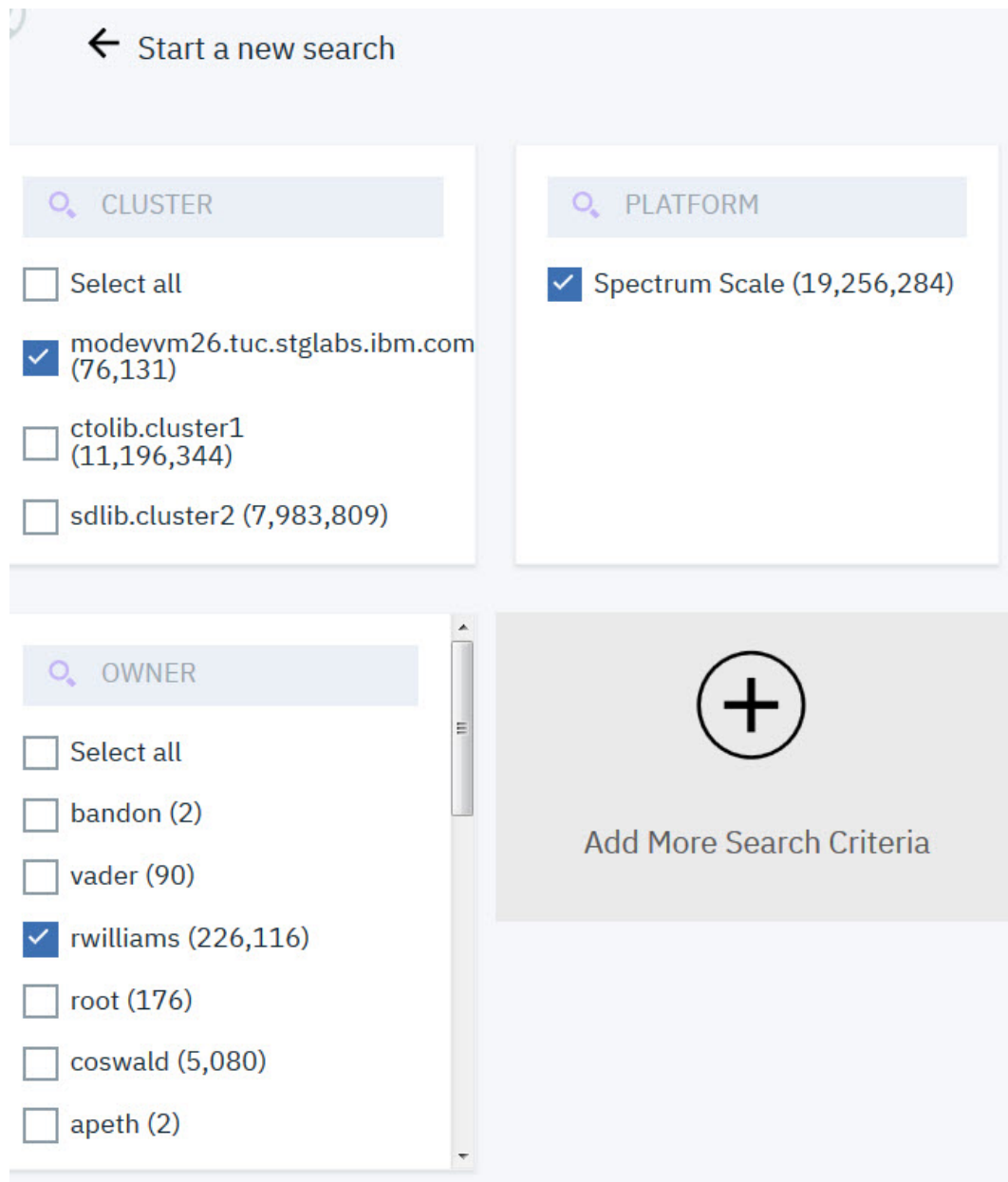


Figure 19. Tag values

- c. Select one or more values for each of the groups, policies, and tags that are displayed.
- d. If needed, click **Add More Search Criteria**. Select one or more items from the **Add groups to your visual search** dialog box. The groups that are displayed in the dialog box are implementation-dependent.

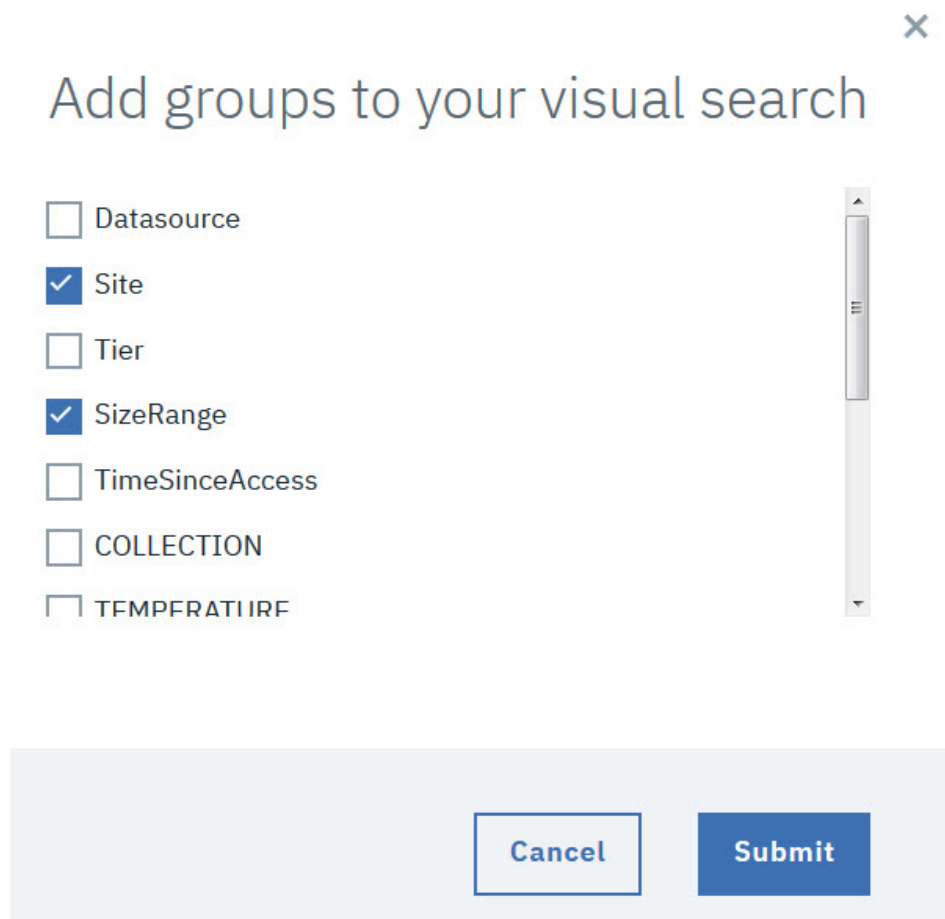


Figure 20. Search - add groups

- e. Click **Submit**.
 - f. Click the **Go** "circle-arrow" icon on the right side of the window. Your query is built and displayed in the **Discover what's in your Data** box.
 - g. You can modify the query, if necessary. Click **Search**.
2. The **Results** of the search are displayed.

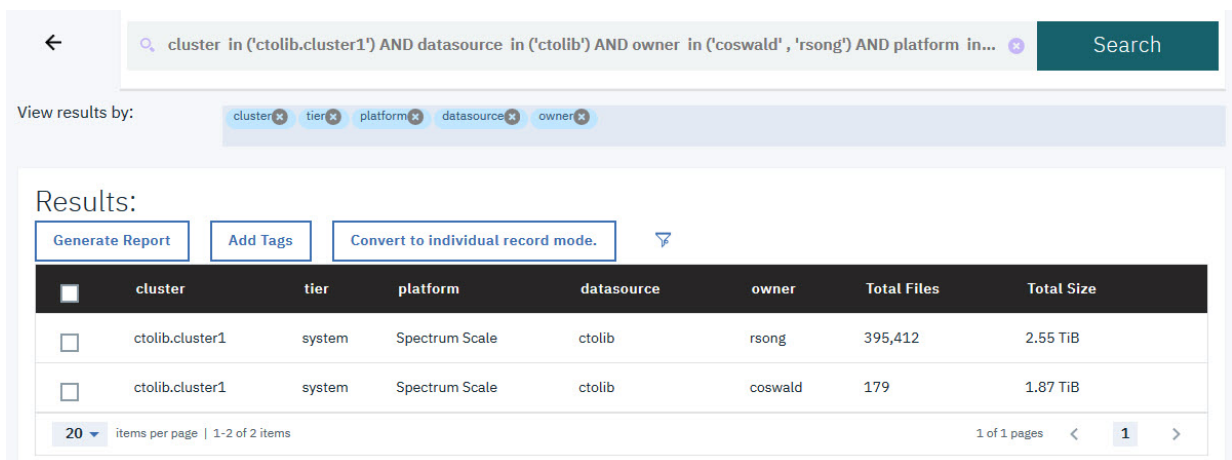


Figure 21. Search Results

- You can change the sort order of a column in the **Results** table by clicking a column's header. Currently, the sort is limited to local data supported by your web browser, with up to a maximum of 10,000 records per query.

- You can change the columns of the **Results** table by clicking columns to remove in the **View results by** list. This list groups the results by applying a new search criteria to the original results.

Note: There might be mismatched results when you are doing an initial search, followed by grouping (**View results by**) and ungrouping (**Convert to individual record mode**). For instance, if the initial search was **size>90000** but the results are grouped, by example, or by **datasource**, you might see a different number of records. If you click **Convert to individual record mode**, the initial search is replaced by the **datasource** grouping, and the results reflect the entire contents of your vault, instead of only the initial results.

If different number of results are returned when you do the search, use the **Search** box to reenter the original query from scratch (you might have copied it to your clipboard in step 1) and the filters to the original query.

- You can add columns to the **Results** table by clicking columns in the **Suggested options** list.

The **Suggested options** menu is available after a column is removed.

3. Filter the search results, if required:

- Click the **Filters** icon. The filters display in the panel to the right of the **Results**.
- Click one or more filters to expand it and select or input values.
- Click **Apply** and the filtered results display in the table.

The screenshot shows a search interface with a search bar containing the query: `cluster in ('ctolib,cluster1') AND datasource in ('ctolib') AND owner in ('coswald', 'rsong') AND platform in (' ...`. Below the search bar, the results are grouped by 'View results by: cluster, tier, platform, datasource, owner'. The main results table has columns: cluster, tier, platform, datasource, owner, Total Files, and Total Size. Two rows are visible, both for 'ctolib.cluster1' with 'system' tier and 'Spectrum Scale' platform. The first row has 'rsong' as owner, 395,412 files, and 2.55 TiB size. The second row has 'coswald' as owner, 179 files, and 1.87 TiB size. On the right, a filter panel is open for 'DATASOURCE', showing a checkbox for 'ctolib (395,591)'. Other filter categories like SIZERANGE, PLATFORM, and TIER are collapsed. An 'Apply' button is at the bottom of the filter panel.

Figure 22. Search Results Filters

- To generate a report, check the box on the left of each row of data that is required. Then, click **Generate Report**.

Generate Report

Name

ctolib.cluster files

Current selected: 5
Current report query: cluster IN ('ctolib.cluster1') AND platform IN ('Spectrum Scale', 'undefined') AND owner IN ('dnoble', 'apeth', 'coswald')

Group By: Cluster Owner

View Individual Records

Cancel Submit

Figure 23. Generate Report

- a. Enter a name for the report in the **Name** box.
 - b. Click the **View Individual Records** box to display the individual files that meet the search criteria in the report.
 - c. Click **Submit** to generate the report. Reports might be viewed by navigating to **Reports** on the main menu.
5. To apply tags to the search results, complete the following tasks:
- a. Select the checkbox of the records you want to add tags to.
 - b. Click **Add Tags**.

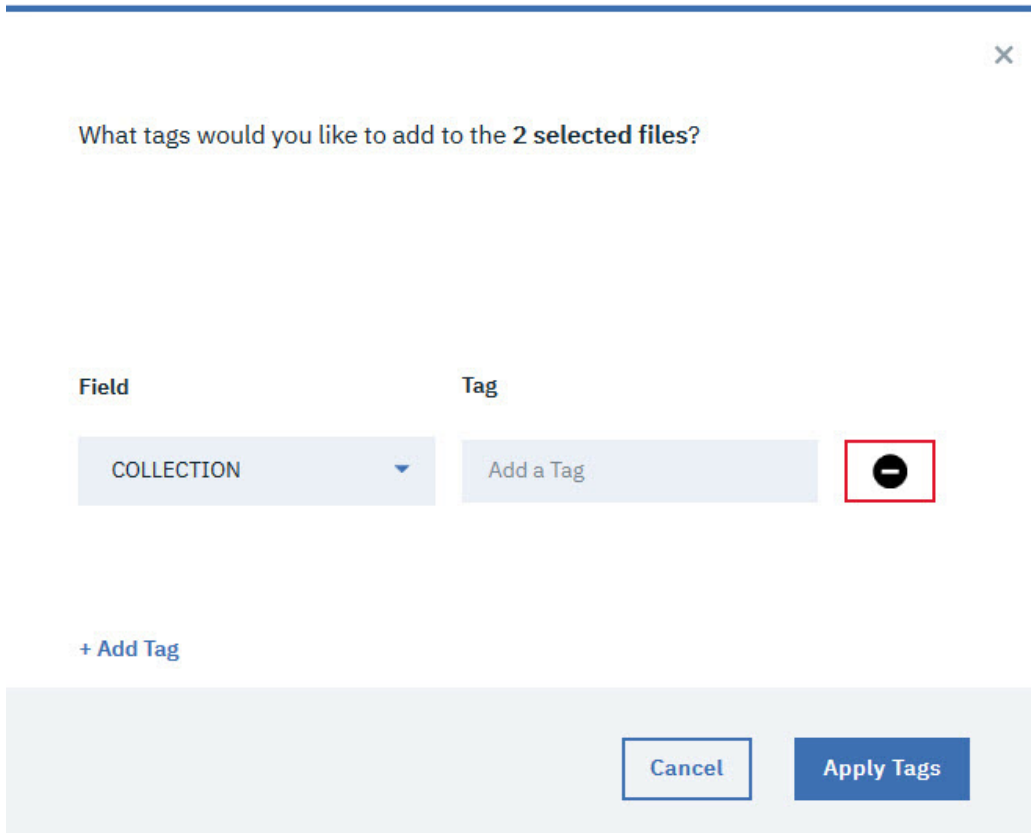


Figure 24. Add tags

- c. If there are no tags that are listed, click **Add tag**.
- d. If you want to delete a tag, click the **Delete** "minus" icon to the right of the tag.
- e. Select the tag to add from the **Field** dropdown menu.
- f. Enter the tag value into the **Tag** box, and press **Enter** on your keyboard.
- g. Continue adding tags as needed.
- h. Click **Apply Tags** when you finish entering all the tags that you need.

When you add tags without a policy, an *Implicit policy* is created. You might view Implicit policies by clicking the [bell] icon in the window's title bar.

6. To view individual records that meet the search criteria, click **Convert to individual record mode**.
7. To view used capacity on the visual search results table, use the settings icon that is located next to **Convert to individual record mode**.
 - a. Find the settings icon:

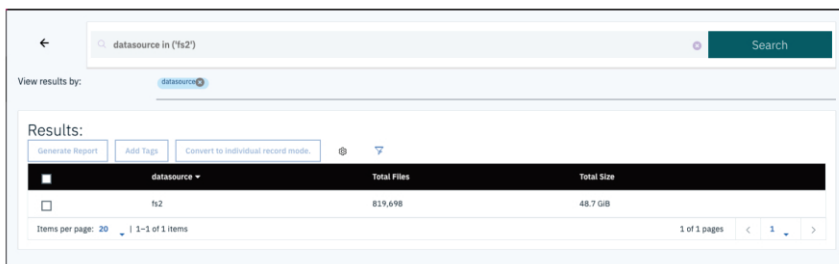


Figure 25. Find settings icon.

- b. Click the settings icon:

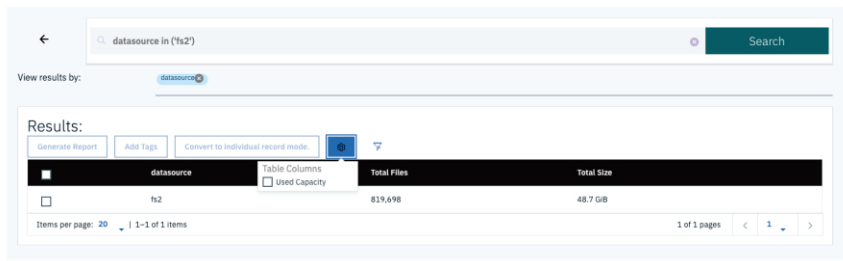


Figure 26. Click the settings icon

c. View the used capacity:

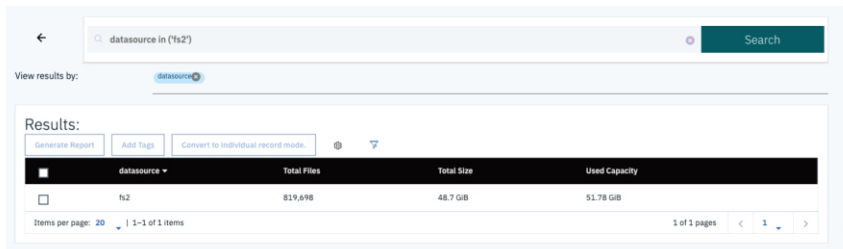


Figure 27. View used capacity

Grouping data by file type

Create policies to group data by file type.

About this task

You can create auto-tag policies and run them on a periodic basis to group data by specific file types.

Procedure

1. Create an **Open** tag to group data. For example, **FileTypeGroup**
 - a. Go to **Metadata > Tags**
 - b. Click **Add**.
 - c. Select **Open** in the **Type** menu.
 - d. Click **Submit**.
2. Create an auto-tag policy with a relevant filter criteria. For example, `Filetype = 'pdf'`.

- a. Go to **Metadata > Policies**.
- b. Click **Add Policy**.
- c. Enter a name for the policy in the **Name** box. For example, `FileTypeGroupTagging`.
- d. Click the slider control to set the policy status to active.
- e. Select **Autotag** from the **Policy Type** menu.
- f. Type `filetype = 'pdf'` as the filter criteria.

Note: You can modify the filter criteria based on the file type you want to group. For example, if you want to group all JPG images, type `filetype = 'jpg'`.

- g. Click **Add Tag**.
- h. Select the file type group tag that you have previously defined. In this example, the tag name is `FileTypeGroup`.
- i. Match the tag value with the filetype in the filter criteria. For example, if the filter criteria is `filetype = 'pdf'` use `pdf` as the tag value.
- j. Click **Save** to run the auto-tag policy.

Note: You can manually refresh the IBM Spectrum Discover database to confirm whether the policy is run properly. Refreshing the database is optional and not an essential prerequisite for the process. To manually refresh the database, follow the procedure shown:

- a. Go to **Admin > Discover Database**.
- b. Click **Run table refresh** on the **Metadata Summarization table**.
3. Create a policy to group data with a different file type tag value or modify the existing policy.
4. Define a schedule to rerun policies on a periodic basis.

Searching system and custom metadata fields

1. Click the **Discover what is in your Data box**.
 - a. Enter your query directly and click **Search**. Use the standard grammar that is used in a SQL query.
 - b. Select a **Suggested Field** from the dropdown menu, complete your query, and click **Search**.
 - c. Select a **Recent Searches** from the dropdown menu, modify the query if necessary, and click **Search**. Click **Show all history** to reveal more in the list of Recent Searches.

The query language is SQL. The underlying code takes care of certain semantics, for example,

- Keyword
- Columns to select
- Name of the databases
- Where clause
- Limits
- Offsets
- Order by clauses

The search clause that is input by the user is only the body of the query that would appear after the where clause and before the limit/offset/order qualifiers.

System metadata fields to search on

The list in this section provides definitions of items on which you can search system metadata fields.

The list below shows search filters that you can use for a search.

Datasource

The name of the datasource where the record originated. The datasource refers to the label of the source storage system that was defined in the IBM Spectrum Discover connection management panel.

Platform

The type of storage system from which this record originated.

Site

The physical site for the data as input by the user at scan time.

Cluster

The name of the IBM Spectrum Scale cluster to which the record belongs. This term applies only to IBM Spectrum Scale.

NodeName

For IBM Spectrum Protect, this indicates the node or client system to which the backup or archive record belongs.

Fileset

The file set to which the record belongs for IBM Spectrum Scale. This term applies only to IBM Spectrum Scale.

MgmtClass

For IBM Spectrum Protect, this indicates the management class to which the backup or archive record belongs.

Owner

The system metadata owner of the record (file only).

Group

The system metadata group owner of the record (file only).

UID

The numeric ID of file owner (file only).

GID

The numeric ID of file group (file only).

Path

The file path or object storage bucket of the file that is represented by this record.

Filename

The name of the file or object represented by the record.

Filespace

For IBM Spectrum Protect, this indicates the file space to which the backup or archive record belongs.

Filetype

The type of the file or object. MtimeLast modified time for the file (file only).

State

The state of the file or object. Possible values for IBM Spectrum Scale are:

premig

Premigrated

migrtd

Migrated

resdnt

Resident

Possible values for IBM Spectrum Protect are:

ACTIVE

Active backup copy

INACTIVE

Inactive backup copy

ARCHIVE

Archive copy

Mtime

Last modified time for the file (file only).

Atime

Last accessed time for the file (file only).

Time

Creation time of the file (file only).

Size

Size of the file or object.

Inode

The inode of the file (file only).

Permissions

The permissions of the file (file only).

sizeConsumed

The size of the consumed capacity (file only).

Access control list metadata to search on

Access Control List (ACL) metadata is collected from SMB/CIFS data source search results.

For SMB/CIFS data sources, IBM Spectrum Discover collects Access Control List (ACL) metadata in addition to the standard system metadata. This information is stored in two tables, which are the Access Control Owner and Group (ACOG) and the Access Control Entries (ACES).

Type the criteria in the search bar to search the ACL metadata. Possible fields to search on are:

acog.ownername

Indicates the owner of the file

acog.ownerid

Indicates the security identifier for the owner of the file.

Note: To search for files based on owner name or security identifier, see the following example:

```
acog.ownername='DOMAIN\user'
```

```
acog.ownerid='S-1-1-1-1000'
```

acog.groupname

Indicates the group of owner of the file.

acog.groupid

Indicates the security identifier for the group of the owner of the file.

aces.username

Indicates the user or group name for which this ACE applies.

aces.userid

Indicates the security identifier for the user or group name for which this ACE applies.

aces.entrytype

Indicates the entry type can either be DACL or SACL.

aces.accesstype

Indicates the access type can be either one of the following options:

- ALLOWED
- DENIED
- AUDIT

Note: To search files with an access control entry that allows everyone to access, see the following example:

```
aces.username='\Everyone' and aces.entrytype='DACL' and aces.access_type='ALLOWED'
```

To search for files that are on a particular data source and display a Deny Access Control Entry, see the following example:

```
datasource in ('smb1') and aces.entrytype='DACL' and aces.accesstype='DENIED'
```

aces.permissions

Indicates the possible permission levels that include:

- R- Read
- W-Write
- X - Execute
- D - Delete
- P - Write access controls
- O - Owner

The valid permission combinations are:

- READ - R + X
- CHANGE - R + W + X + D
- FULL - R + W + X + D + P + O

aces.flags

Indicates the flags displaying the type of access control entry (ACE).

Note: A file can have more than one access control entry (ACE) associated with it. Search results that contain ACL metadata, repeat the file metadata for each ACE. Therefore, reports are the preferred method for using IBM Spectrum Discover search results with ACL metadata.

Search on custom metadata fields

You can do a search on custom metadata fields.

Comparators

To do a search, you can also use the following comparators.

=

Is equal to.

<>

Is not equal to.

<

Is less than.

>

Is greater than.

<=

Is less than or equal to.

>=

Is greater than or equal to.

is

When you search for null values:

is null

Indicates a null (or no) value.

is not null

Indicates a valid value that is not null.

Conjunctions

You can also use conjunctions.

AND

Tie together multiple filter criteria.

OR

Meet at least one of multiple filter criteria.

Helpers

You can also use helpers.

NOW()

Get the current TIMESTAMP.

DAYS/MONTHS/YEARS

Compares TIMESTAMP/DATE values.

Wildcards

You can also use a wildcard.

%

You can use a wildcard like % with the keyword LIKE to form a wildcard search.

Note: Both single quotation marks (') and backslashes (\) need to be escaped with a leading backslash. For example, if the search statement is: `mytag in ('first', 'can't', 'with\slash')`, the following convention needs to be followed to escape it:

```
mytag in ('first', 'can\t', 'with\\slash').
```

]

Examples of search filters

This section provides a list of examples for search filters.

Note: You must wrap string values in single quotation marks but you cannot wrap numeric values in single quotes.

Owner= 'bob'

All files owned by 'bob'.

Fileset= 'bobs project'

All files in the file set bobs_project.

Filetype = 'pdf' AND size > 500000

All PDF files that are larger than 500000 bytes.

Filetype is ('txt', 'pdf', 'doc')

All files of type TXT, PDF, or DOC.

Atime < (NOW() - 180 DAYS)

All files not accessed in the last 180 days.

Filesystem = 'big_fs ' AND owner <> 'root'

All files in the big_fs filesystem that are not owned by root.

collection = 'proj_xylem'

Search for all records that are tagged with the user-defined tag 'Project' set to 'proj_xylem'.

collection <> ''

Search for all records that have a collection that is assigned.

filename LIKE 'the_quick_brown_%'

Returns all records for which the file name begins with "the_quick_brown_".

department= 'department_xylem'

Search for all records that are tagged with the user-defined tag 'Department' set to 'proj_xylem'.

custom_tag is null

All files for which custom_tag is not set to any value.

Search results table

The search results table displays information about the records that met the search criteria.

By default, certain columns are shown, and others are hidden. You can customize the fields in the view in the **Headers** column of the **Advanced Search Options**.

Figure 28 on page 57 shows an example of a search by file type and data source.

sdadmin

cluster IN ('9.11.201.5')

View results by:

Results:

Generate Report

Add Tags

checkbox	filename	filetype	datasource	size
<input type="checkbox"/>	.htaccess	htaccess	datadump	6527.000
<input type="checkbox"/>	._collections.pyc	pyc	datadump	13663.000
<input type="checkbox"/>	._collections.py	py	datadump	10553.000
<input type="checkbox"/>	tarfile.pyc	pyc	datadump	78796.000
<input type="checkbox"/>	__init__.py	py	datadump	274.000
<input type="checkbox"/>	__init__.py	py	datadump	0.000
<input type="checkbox"/>	__init__.pyc	pyc	datadump	533.000
<input type="checkbox"/>	url.py	py	datadump	5879.000
<input type="checkbox"/>	langthaimodel.py	py	datadump	11275.000

- > Size
- > Creation Time
- > Last Accessed Time
- > Last Updated Time
- Columns
 - path
 - filetype
 - datasource
 - owner
 - group
 - revision
 - site
 - platform
 - cluster
 - inode
 - permissions
 - fileset
 - uid

pov00079

Figure 28. Example to generate a report sorted by file type and data source

Figure 29 on page 58 shows an example of the search results for the time since access and size range.

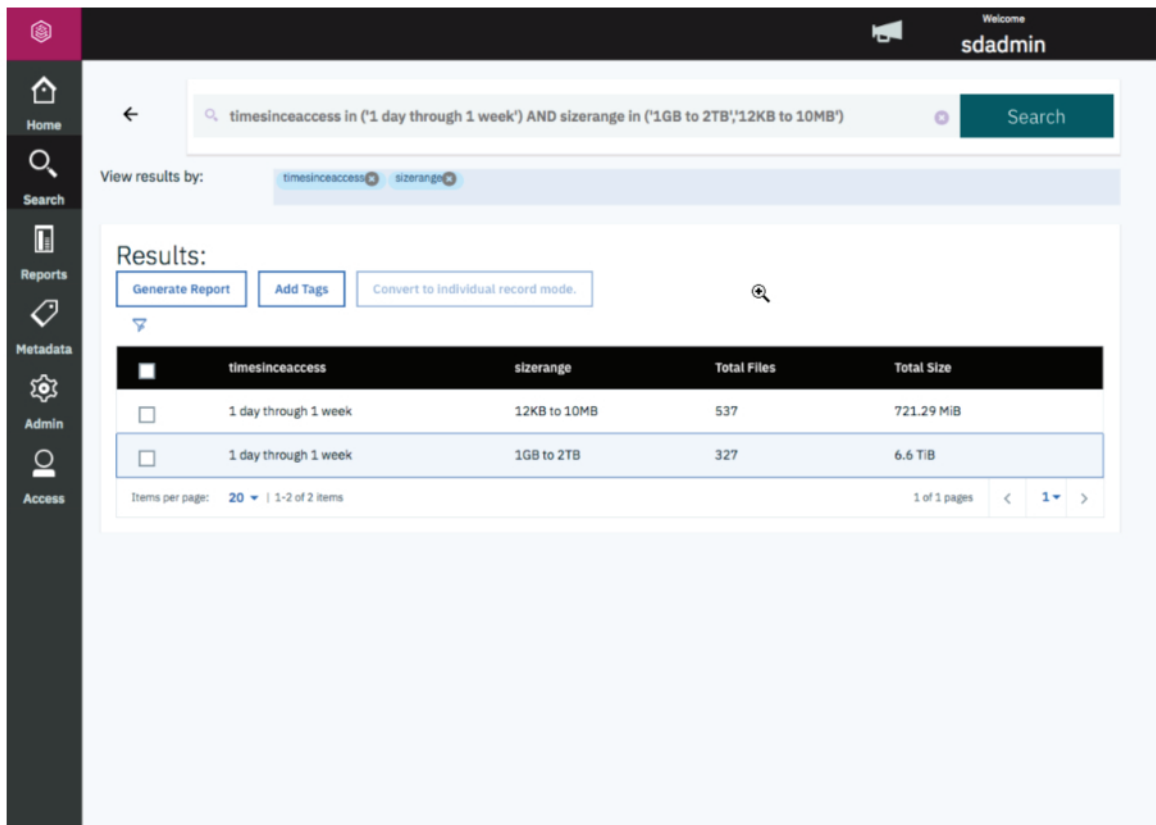


Figure 29. Example of a search sorted by time since access and size range

Refine search results

After a set of search results is returned, you can use the **Advanced Search Options** refine the data.

The selection box for **File System** provides you with a way to select one or more source storage systems to restrict the search. The selection box for **Time** provides you with a way to select a range of time since access for the records. The selection box for **Size** provides you with a way to select a minimum or maximum file size footprint for the records to meet the criteria.

You can use a combination of any or all of the filtering criteria. To apply the filters to the current search results, click **Apply**. To reset the filtering criteria and return to the base search click **Reset**.

Sort search results

You can sort search results by column.

When you click the column header, you can sort the results in ascending order. When you click the column a second time, you can sort the column in descending order. The time it takes to sort depends on the size of the result set.

Note: Sorting by a second column loses the order of the data in the first column. A combination sort view is not supported.

Tag search results manually

After a filtered set of records has been identified through the search pane, the user has the ability to select all or some of those documents and apply organizational tags to them.

For example, if a drill-down search results in identifying all of the records for a particular project, you can click **Add Tags** and specify that an organizational tag called '**Project**' be set to the name of the project represented by the filtered set. The tag application runs as a background task and you get a notification when processing has completed.

You can apply more than one tag at same time.

Chapter 8. Managing applications

An application is a program that interfaces with IBM Spectrum Discover and can access the source storage. There are many use cases for application, including data content inspection for enriching metadata, data movement or migration, data scrubbing or sanitation, and more. Data is identified by IBM Spectrum Discover by policy filter and passed to the application as pointers through a messaging queue. Then, the application performs whatever work is appropriate on the source data and returns a completion status back to IBM Spectrum Discover, which might or might not include enriched metadata for the records. If it does include enriched metadata, IBM Spectrum Discover catalogs that metadata and makes it immediately searchable.

Permissions

Data Administrator

Create (register), update, delete (unregister), and view the applications.

Data User


View the applications created by a Data Administrator.

Security Administrator

Cannot create, modify, view, or delete any application.

Management

Applications might be viewed and deleted by navigating to **Metadata > Applications**. You can define an application when you are creating a new **DEEP-INSPECT** policy. In addition, you can add **Parameters** for an application during the process of creating a **DEEP-INSPECT** policy. For more information, see [“Adding deep-inspection policy parameters” on page 18](#).



Application	Parameters	Action ID	View/Delete
contentsearchagent		"CONTENTSEARCH"	 

Figure 30. Applications table

The **Applications** table displays the following information:

Application

The name of the application.

Parameters

The parameters that were assigned to the application when the policy was created.

Action ID

AUTOTAG or **deepinspect** - the policy type that the application is assigned to.

View/Delete

Use the **View** "eye" icon to view the contents of the application:

- Application
- Action ID
- Params

Use the **Delete** "trashcan" icon to remove the application from the system.

For more information, see the *Application Registration REST API Guide*.

Chapter 9. Using the IBM Spectrum Discover application catalog

Use the IBM Spectrum Discover application catalog to search, download, or deploy applications (which are provided by IBM, customers, or third parties) to use in IBM Spectrum Discover.

To use the commands in the examples throughout this document, you must use Secure Shell (SSH) to log in to the IBM Spectrum Discover. You also must have an authentication token that is generated from the command-line interface (CLI). (The token expires after 1 hour.) Run the following command to generate a token:

```
ssh moadmin@<your IP entered during mmconfigappliance>
# Enter in the password you set during the mmconfigappliance
export SD_USER=<sdadmin or another user with dataadmin privileges>
export SD_PASSWORD=<password for SD_USER above>
export OVA=images
gettoken
```

Note: In this example, `gettoken` is an alias under the `moadmin` user. Using an alias saves the token in an environment variable that is called `TOKEN`.

Note: The examples in the sections throughout this document use the aliases `tcurl` and `tcurl_json` under the `moadmin` user, which also uses the `TOKEN` environment variable.

Information about the endpoints

Follow the procedure to access information on endpoints:

1. Go to [IBM Spectrum Discover Knowledge Center](#).
2. Choose the version of IBM Spectrum Discover that you are running.
3. Go to **Table of Contents > REST API > Application management using APIs**.

Querying the available applications

Run this command to query the applications that are available on `dockerhub`:

```
tcurl https://${OVA}/api/application/appcatalog/publicregistry | jq
```

The output that is generated contains information that is gathered from the image itself (and from `dockerhub`).

Downloading an application image

Run the following command after you identify an application to download:

```
tcurl_json https://localhost/api/application/appcatalog/image/ibmcom/spectrum-discover-example-application -X POST | jq
```

Note: In this example, `ibmcom/spectrum-discover-example-application` is the `repo_name` used in the `publicregistry` command.

Running an application

After you download an application to your local `docker` cache, you can use it as a Kubernetes pod within IBM Spectrum Discover. Create a JSON-formatted file with the following information (the file that is created is named `example.json`):

```
[{
  "repo_name": "ibmcom/spectrum-discover-example-application",
  "version": "1.2.3",
```

```

"description": "Unique description about your use of this application",
"application_name": "example",
"my_env_var": "my_value",
"LOG_LEVEL": "DEBUG"
}

```

Note: In this example:

- The `repo_name` is the same `repo_name` that you used to download the application image.
- The `version` is the same as the version from the output of the `publicregistry` command.
- The `description` is a unique description that is based on your application use.
- The `application_name` is the name that gets registered within the policyengine. The system automatically appends a `-application` to the end of the file name for identification.

Run the following command to start the application as a Kubernetes pod:

```

tcurl_json https://localhost/api/application/appcatalog/helm -d@example.json -X POST | jq

```

You can add environment variables to the JSON example. These environment variables can be ones that your application needs or they can be ones that can override some software development kit (SDK) values. The application SDK supports the following environment variables that can override default settings:

LOG_LEVEL - INFO (default), DEBUG

Specifies the log level for the application to run with.

MAX_POLL_INTERVAL - 86400000 (in milliseconds)(default - 1 day)

Specifies when the Kafka consumer becomes unresponsive. Set this value higher than the time it takes for the application to process up to 100 records before it sends the reply to IBM Spectrum Discover. The default allows approximately 15 minutes for each record.

PRESERVE_STAT_TIME - False (default), True

Specifies whether to preserve atime or mtime when you run the deep-inspection application. If the application processes records from Network File System (NFS), Server Message Block (SMB), or local IBM Spectrum Scale connections, the system preserves the exact atime or mtime (in nanoseconds).

If the application processes records from a remote IBM Spectrum Scale connection, the system preserves atime or mtime up to and including seconds (with no subsecond preservation). The connection user must also have write access to the files. If the connection user does not have write access to the files, the system skips restoration of the atime or mtime because of permission errors. If DEBUG is on, you can see the original atime or mtime in the logs, so you can potentially manually restore any that fail.

Scaling an application

An application by design processes each of the records one at a time. You can scale the number of replicas the pod is running to process records in parallel. You can scale up to 10 replicas based on the number of partitions available for the Kafka topics. Create a JSON-formatted file with the following information (the file that is created is named `replicas.json`):

```

{
  "replicas": 10
}

```

Then, run the following command to scale the replicas:

```

tcurl_json https://localhost/api/application/appcatalog/helm/interesting-anaconda-example-application -d@replicas.json -X PATCH

```

Note: In this example, `interesting-anaconda-example-application` is the combination of `deployment_name` and `chart_name` from the `Running an application` section.

Stopping an application

Run the following command to stop an application (no matter how many replicas you scale):

```
tcurl_json https://localhost/api/application/appcatalog/helm/interesting-anaconda -X DELETE | jq
```

Note: In this example, `interesting-anaconda` is the `chart_name` when the application was started.

Deleting an application image

Run the following command (after you stop the application) to delete the application from your local docker cache:

```
tcurl https://localhost/api/application/appcatalog/image/ibmcom/spectrum-discover-example-application -X DELETE | jq
```

Creating your own applications to use in the IBM Spectrum Discover application catalog

Use this information to create applications for the IBM Spectrum Discover application catalog.

The following reference locations provide the source materials to help you start creating applications for IBM Spectrum Discover application catalog:

https://github.com/IBM/Spectrum_Discover_App_Catalog

This link contains the source code of the IBM-provided applications.

https://github.com/IBM/Spectrum_Discover_Application_SDK

This link contains the source code for the IBM Spectrum Discover Application Software Development Kit (IBM Spectrum Discover Application SDK). The link also describes how to build a test image for use in creating your own applications.

https://github.com/IBM/Spectrum_Discover_Example_Application

This link contains the source code for the template application. Start here when you create your own applications.

Chapter 10. Backup and restore

IBM Spectrum Discover includes a set of scripts for safely backing up and restoring your database and file system.

The scripts that are used to back up and restore databases and file systems are located in the `opt/ibm/metaocean/backup-restore` directory, and must be run as root user:

```
sudo python3 /opt/ibm/metaocean/backup-restore/backup.py
```

It is a good practice to back up your system at least one time a week.

Note: [The backup that you use to restore an IBM Spectrum Discover system must be at the same code level as the IBM Spectrum Discover system that is being restored. For example, you must be restoring a 2.0.2.1 system if you want to use a 2.0.2.1 backup.]

IBM Spectrum Discover provides the `automatedBackup.py` script that can be used to configure a `cron` job that backs up your system and offloads a `tar` file to your selected storage server. The default configuration is daily at 12 midnight; however, you can configure the backup frequency by running the **automatedBackup.py** script after the initial setup.

Remember:

- If any files or a database becomes corrupted, run the `restore.py` script to recover your file system and database back to the date of your last successful backup.
- When you start a backup or restore operation, remember it can take up to 1 hour or more time to complete. Make sure that you plan for this possibility.

Initial setup configuration

Follow these steps to set up initial configuration.

Procedure

1. Run the `initialSetup.py` script as root.
2. Enter the type of storage you're using:
 - IBM Cloud Object Storage ("cos")
 - a) Enter the Accesser Device (or Load Balancer) IP address.
 - b) Enter the Manager Device IP address.
 - c) Enter the name of IBM Cloud Object Storage vault to store backup **tar** files.
 - d) Enter the user name for IBM Cloud Object Storage account configured with read/write access to storage vault.
 - IBM Spectrum Protect ("spectrum")
 - a) You must have a IBM Spectrum Protect server and a backup-archive client that is installed and properly configured. For more information, see [IBM Spectrum Protect Knowledge Center](#).
 - External FTP server ("ftp")
 - a) Enter the SFTP server IP or host name,
 - b) Enter the user name for read/write authorized SFTP user.
 - c) Enter the password for read/write authorized SFTP user.
 - d) Enter the path to the directory for storing and retrieving backup **tar** files (Example: `/var/backups/daily/`).
3. Enter a maximum number of backup `tar` files to be retained in storage.

The default number of backups is 30, but you can enter any number in the range 1 - 999. After the selected number of backups is exceeded, the oldest backup tar file is deleted.

Example

Log or console output from **initialSetup.py**:

```
Tue, 28 Aug 2018 14:26:02 INFO Setup is validating user inputs, this might take a while....
Tue, 28 Aug 2018 14:26:02 INFO Checking manager credentials are valid: successful
Tue, 28 Aug 2018 14:26:02 INFO Checking whether the specified account exists: successful
Tue, 28 Aug 2018 14:26:02 INFO Checking whether the specified vault exists: successful
Tue, 28 Aug 2018 14:26:02 INFO Checking whether the specified user has read-write access to the specified vault: successful
Tue, 28 Aug 2018 14:26:02 INFO Generating access key and secret for the username provided: successful
Tue, 28 Aug 2018 14:26:02 INFO Configuration file is created successfully
Tue, 28 Aug 2018 14:26:02 INFO Setup is successful, please continue running backup or restore scripts as a root.
```

Note: If a backup or restore procedure is interrupted or unexpectedly stops, a checkpoint is logged that you can use to rerun the script and pick up from where the process was halted. To override these functions and force a fresh restart of the backup or restore procedure, run the **backup.py** or **restore.py** script with an extra **--override** parameter. For example:

```
sudo python3 restore.py -r "2018-08-28" --override
```

Running a backup

Follow these steps to run a backup.

Procedure

1. Place the system in maintenance mode:

```
[sudo /opt/ibm/metaocean/helpers/maintenance.sh on]
```

2. From the backup-restore directory, run the **backup.py** script as root:

```
sudo python3 backup.py
```

The following example log or console output is from **backup.py**:

```
Tue, 28 Aug 2018 14:26:57 INFO The COS Endpoint is 172.19.17.34, writing to the vault: mo_backups
Tue, 28 Aug 2018 14:26:57 INFO Suspending writes on container (1a8420e6dd85)...
Tue, 28 Aug 2018 14:27:11 INFO Creating snapshot 2018-08-28T14.26.57_snapshot...
Tue, 28 Aug 2018 14:27:11 INFO Snapshot 2018-08-28T14.26.57_snapshot created.
Tue, 28 Aug 2018 14:27:11 INFO Resuming writes on container (1a8420e6dd85)...
Tue, 28 Aug 2018 14:27:15 INFO Converting snapshot to tar
Tue, 28 Aug 2018 14:28:18 INFO Snapshot tar /gpfs/gpfs0/2018-08-28T14.26.57_snapshot.tar created
Tue, 28 Aug 2018 14:28:18 INFO Beginning upload of /gpfs/gpfs0/2018-08-28T14.26.57_snapshot.tar to
storage...
Tue, 28 Aug 2018 14:33:21 INFO Upload of file /gpfs/gpfs0/2018-08-28T14.26.57_snapshot.tar complete.
Tue, 28 Aug 2018 14:33:21 INFO Beginning cleanup...
Tue, 28 Aug 2018 14:33:21 INFO Deleted snapshot 2018-08-28T14.26.57_snapshot
Tue, 28 Aug 2018 14:33:21 INFO Deleted tar /gpfs/gpfs0/2018-08-28T14.26.57_snapshot.tar
Tue, 28 Aug 2018 14:33:21 INFO Backup procedure is complete.
```

Note: If a backup or restore procedure is interrupted or unexpectedly stops, a checkpoint is logged so you can rerun the script and pick up where the process was halted. To override these functions and force a fresh restart the backup or restore procedure, run the **backup.py** or **restore.py** script with an extra **--override** parameter:

```
sudo python3 restore.py -r "2018-08-28" --override
```

3. Remove the system from maintenance mode:

```
[sudo /opt/ibm/metaocean/helpers/maintenance.sh off]
```

Running an automated backup

Follow these steps to run an automated backup.

Procedure

1. Run the initial setup.
2. Run the **automatedBackup.py** script as root from the backup-restore directory:

```
sudo python3 automatedBackup.py
```

3. Type yes or y to confirm that you want to schedule an automated backup next to the following message:

```
Please enter (yes) or (y) if you want to schedule an automated cron job for backup?
```

4. Type yes or y to run a non-default schedule next to the following message:

```
Please enter (yes) or (y) if you want to setup an non-default automated cron job for backup.
Choosing (n) would result to default setup (default is a daily at 12 AM)
```

Note: Type n if you want to configure the automated backup to the default schedule. All default schedules run daily backups at 12 midnight.

5. Type the frequency at which you want to run the backup. The frequency options that are provided include:

Daily

For all daily backups, type hours of the day (0-23).

Weekly

For all weekly backups, type the day of week (MON-SUN) and the hour (0-23). The values that you enter must be coma-separated and without any space. For example: MON,15.

Monthly

For all monthly backups, type the day of the month (0-31) and the hour (0-23). The values that you enter must be coma-separated and without any space. For example: 27,12.

6. When the automated backup script is successfully run, the completion message appears as shown:

```
A cron job setup is successful.
```

Running a restore

Follow these steps to run a restore.

Procedure

1. Make sure that all backup processes or backup scripts stop, and then place the system in maintenance mode:

```
[sudo /opt/ibm/metaocean/helpers/maintenance.sh on]
```

2. Run a restore. From the backup-restore directory, run the `restore.py` script as root, with a parameter for date to restore back to (`--restore-date` or `-r`) in YYYY-MM-DD format:

```
sudo python3 restore.py -r "2018-08-28"
```

3. Remove the system from maintenance mode:

```
[sudo /opt/ibm/metaocean/helpers/maintenance.sh off]
```

4. If IBM Cloud Object Storage notifications are lost during the period when the system was disabled, the notifications can be recovered by using the IBM Cloud Object Storage Replay.

Chapter 11. Reports

Reports can be generated upon applying tags to a set of data.

Procedure

1. Reports can be generated by using the following methods:
 - **Discover data** by performing a **Search** in IBM Spectrum Discover. The search results provide an option to **Generate Reports**. For more information, see [“Searching” on page 45](#) for details.
 - Use the Graphical User Interface (GUI) to automatically run the reports during deployment.
2. Go to **Reports** in the IBM Spectrum Discover main menu.

Report	Last Run	Duration (seconds)	Status	Output Size	Actions
Cluster Report - Individual Records	2018-10-18T22:12:05.000Z	0	failed		
Cluster Report	2018-10-18T22:11:42.000Z	0	complete	199 Bytes	
apond	2018-10-11T14:14:07.000Z	0	complete	41 Bytes	

20 items per page | 1-3 of 3 items 1 of 1 pages < 1 >

Figure 31. Reports table

3. The following actions can be completed in a table:

View

- a. Click the "eye" icon to view a report. The report's statistics are displayed in a box.

×

View Data Report

Report: Cluster Report
Last Run: 2018-10-18T22:11:42.000Z
Duration: 0
Status: complete
Output Size: 199
Query: { "group_by": ["cluster", "sizerange", "site", "timesinceaccess", "Platform"], "name": "Cluster Report", "sort_by": "", "filters": [], "query": "cluster IN ('ctolib.cluster1') AND sizerange IN ('extra large') AND site IN ('') AND timesinceaccess IN ('very short', 'very long') AND platform IN ('Spectrum Scale') " }

[See on table.](#)

Cancel

Figure 32. View Data Report

- b. Click **See on table** to view all the records of a report. The **Search** window displays the results of the search.

Download

Click the **Download** button to open a report with a text editor, or to save the report to local storage.

Rerun report

Click the **Go** "right arrow" icon to rerun the report.

Delete

Click the **Delete** "trashcan" icon to remove the report.

Chapter 12. High availability for a Db2 Warehouse MPP deployment

For an MPP deployment, Db2® Warehouse provides high availability, offering you the ability to have your data warehouse carry on with its activities if failures occur.

The HA solution is based on a heartbeat mechanism, automatic restart of services, and node failover. The heartbeat detects when a node, a database partition, or the web console is down, and the cluster manager takes the appropriate action. For instance, the cluster manager attempts to restart any failed data partitions or the web console. [Figure 33 on page 73](#) shows a Db2 Warehouse HA group in a healthy state. The file system is not a part of the HA group, so use whatever HA solution that is appropriate for the technology you are using. Similarly, you can use a method such as a load balancer to make head node failures not apparent to connected applications.

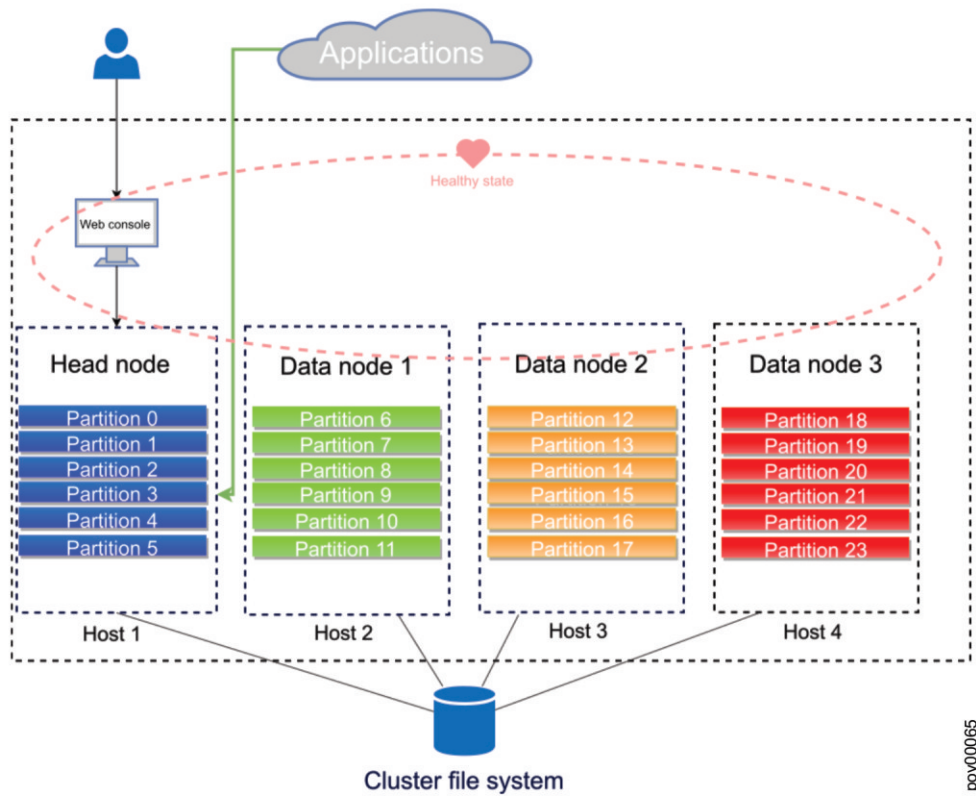


Figure 33. Steady state for HA group

If a data node fails and does not restart within the heartbeat interval, all services are stopped on that node. The data partitions (and their workload) that are assigned to that node are automatically redistributed across the surviving nodes in the cluster. There is no way to automatically reintegrate failed nodes; you must perform some manual steps to have a failed node rejoin the cluster.

If the head node fails and does not restart within the heartbeat interval, its data partitions are redistributed, and an election occurs. In the election, a new head node is selected from the first seven active data nodes in the cluster. As you can see in [Figure 34 on page 74](#), the web console is restarted on the new head node.

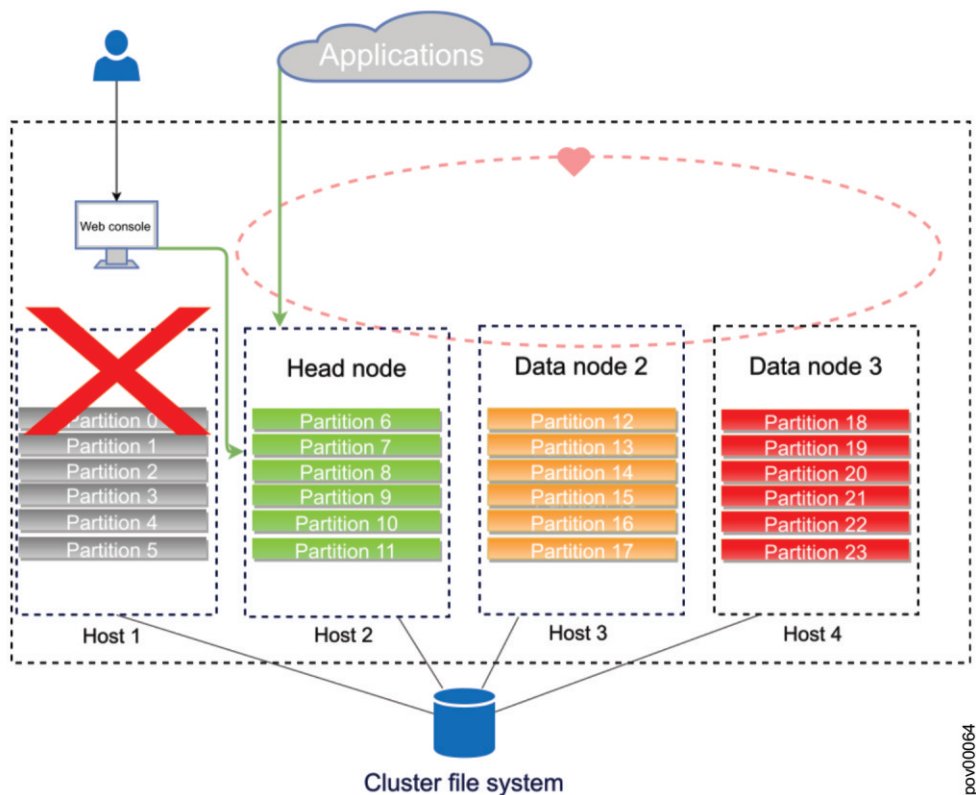


Figure 34. HA group after head node failover

After a head node failover, if the original head node becomes reachable again, restart the system for the original head node to become the current head node again.

Reintegrating a failed data node into an IBM Db2 Warehouse MPP cluster

You must perform some manual steps to have a failed data node rejoin its cluster.

About this task

To perform this task, you need to have root authority.

Procedure

1. Address whatever issue caused the node host failure.
2. Start the Db2 Warehouse container on the node you want to rejoin to the cluster.

```
docker start Db2wh
```

3. On the head node, stop the Db2 Warehouse services for the cluster.

```
docker exec -it Db2wh stop
```

4. On the head node, start the Db2 Warehouse services.

```
docker exec -it Db2wh start
```

The cluster should come up with the same topology as before the data node failure, with the data partitions distributed across all nodes.

Chapter 13. Monitoring data sources

You can use the **Home** page to monitor the data sources that are connected to your IBM Spectrum Discover environment. Use the **Data Source Connections** page view details about data source connections.

Viewing data source status

Use the **Home page** to monitor your environment for storage system capacity, used capacity, records indexed, and duplicate files. You can also view data usage for a specific area of your organization.

The data in the **home page** is updated periodically. The last update is indicated by a time stamp.

Viewing storage system capacity

Use the **Data source Capacity** area to view capacity usage compared to the allocated capacity for all data sources that are registered with IBM Spectrum Discover. The data sources can be a mixture of file systems and object vaults. A graph provides a convenient view of the current capacity of data sources and whether any are close to running out of space. This view also indicates the number of files to move or archive, based on user-defined policies.

Hover over a data source in the graph to view details about the data source. Click a data source to open the **Search** page and perform a search of the selected data source.

Note: [Data sources that do not have data residing in them are not displayed in the graph.]

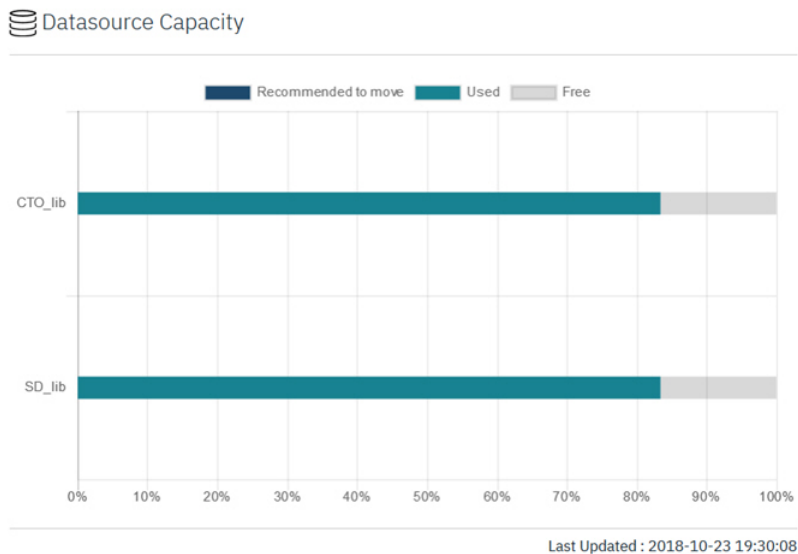


Figure 35. Datasource capacity

Understanding size and capacity differences

IBM Spectrum Discover collects size and capacity information. Generally:

- Size refers to the size of a file or object in bytes.
- Capacity refers to the amount of space the file or object consumes on the source storage in bytes.

For objects, size and capacity values always match. For files, size and capacity values can be different because of file system block overhead or sparsely populated files.

Note: Storage protection overhead (such as RAID values or erasure coding) and replication overhead are not captured in the capacity values.

Viewing used capacity

Use the **Capacity Used by** area to view graphs with an aggregated display of capacity usage for selected metadata attributes. You can view capacity for both primary and backup sources. The graphs provide details about capacity usage by aggregating across different attributes that are available from standard system metadata.

Use the **Capacity Used by** list to select an attribute and display the capacity consumers of that attribute in the graphs.

The **Used** graph displays the highest consumers of capacity for the selected attribute, in order of consumption.

The data source graph displays the percentage of overall usage per data source for the selected attribute. You can select a specific capacity consumer to display in the graph.

Hover over a value in a graph to view details. Click a value in a graph to open the **Search** page and perform a search of the selected item.



Figure 36. Example of the capacity that is being used

Viewing records indexed

Use the **Records Indexed** area to view both the total number of records and the capacity of the records that are indexed by IBM Spectrum Discover. This view provides a summary view of total storage usage.

Records Indexed

19,180,153

Total Records Indexed

322.41 TiB

Total Capacity Indexed

Last Updated : 2018-10-23 19:30:08

Click the **Total Records Indexed** value to open the **Search** page and perform a search of the indexed records.

Viewing duplicate file information

Use the **Duplicate File Information** area to view information about possible duplicate files within the storage environment. Possible duplicate files are files with the same name and size but different paths or object names. The number of duplicates and the capacity that is consumed by these files is displayed. You can also use a report that provides detailed and sorted information for the potential duplicates.

Click the **Duplicate Records** value to open the **Search** page and perform a search of duplicate records.

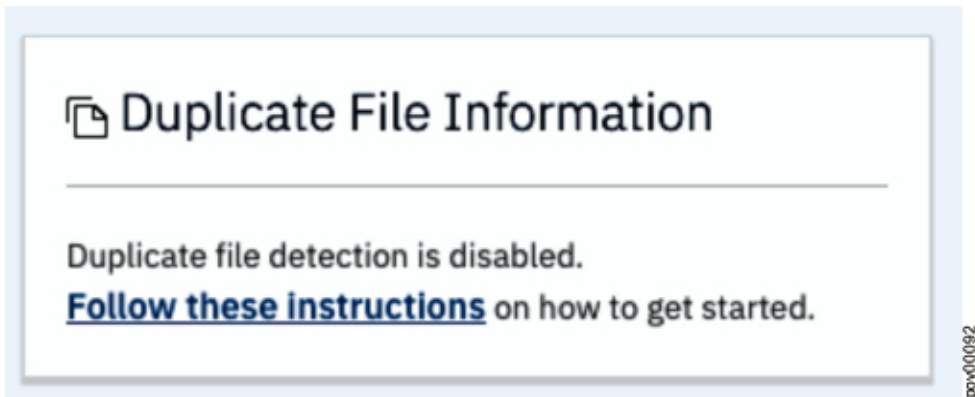
Duplicate File Information

10,913,954
Duplicate Records

1.11 TiB
Total Capacity Consumed

Last Updated : 2018-10-23 00:15:39

Identifying potential duplicates can be resource-intensive on IBM Spectrum Discover. By default, the background task that refreshes potential duplicate information is disabled. However, you can update potential duplicate information either on demand or on a specific schedule. If you disable duplicate background task, the dashboard shows the following message:



To view and manage how often data in the home page is updated, navigate to **Admin > Discover database**.

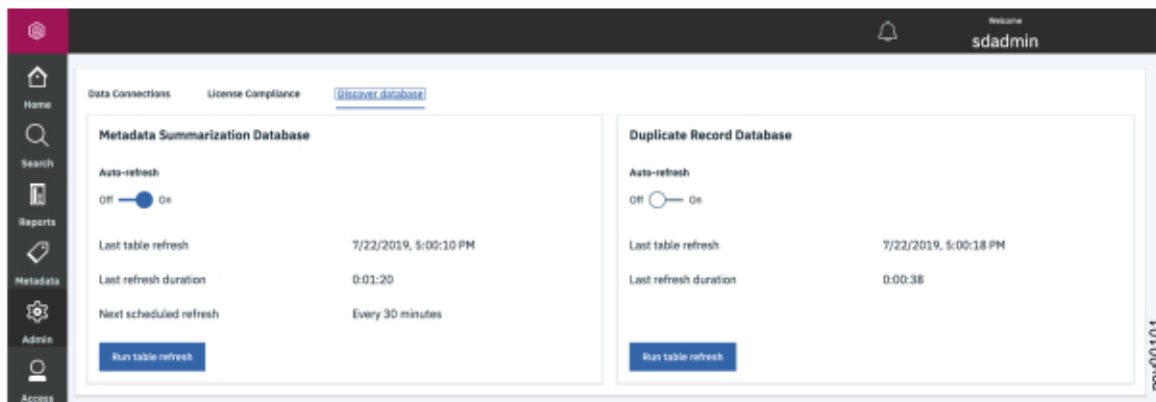


Figure 37. Run table refresh button in the Discover database window

From here, you can enable or disable the automatic updating of summary information. You can update information on the home page on demand by clicking the **Run table refresh** button.

Viewing data source connections

Use the **Data Source Connections** page to view connection information for the data sources that are connected to your IBM Spectrum Discover environment.

The following connections details are available:

Source Name

A name that uniquely identifies the connection to the data source. A data source can have multiple connections.

Platform

The platform of the data source - IBM Spectrum Scale system or IBM Cloud Object Storage system.

Cluster

The cluster address of the data source.

Data source name

The full name of the data source.

Site

The physical location of the data source.

Recommended to move

In the IBM Spectrum Discover dashboard, you can categorize data as **Recommended to move**.

Figure 38 on page 78 shows an example of a data source capacity widget.

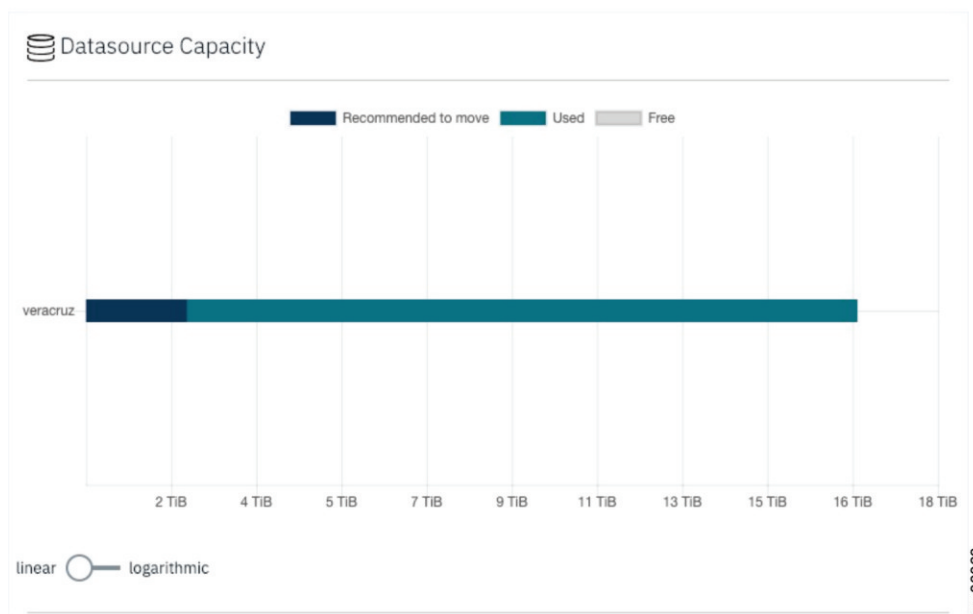


Figure 38. Example of a data source capacity widget

Use the **data source capacity** area to view capacity usage compared to the allocated capacity for all data sources that are registered with IBM Spectrum Discover. The data sources can be a mixture of file systems and object vaults. A graph provides a convenient view of the current capacity of data sources and whether any are close to running out of space. This view also indicates the number of files to move or archive, based on user-defined policies.

Hover over a data source in the graph to view details about the data source. Click a data source to open the Search page and perform a search of the selected data source.

The data source capacity widget displays any files or objects that have the **TEMPERATURE** tag set to a value of **ARCHIVE** as **Recommended to move**. You can create an autotag policy to look for files and objects, which meet your archive criteria and set the **TEMPERATURE** tag to a value of **ARCHIVE**.

Any files that match the criteria of the autotag policy filter are tagged as **ARCHIVE**. The filter might be age-based or more complex. For example, the filter might match only certain file types, or files over some size threshold.

Figure 39 on page 79 shows an example of a screen that shows the **TEMPERATURE** tag.

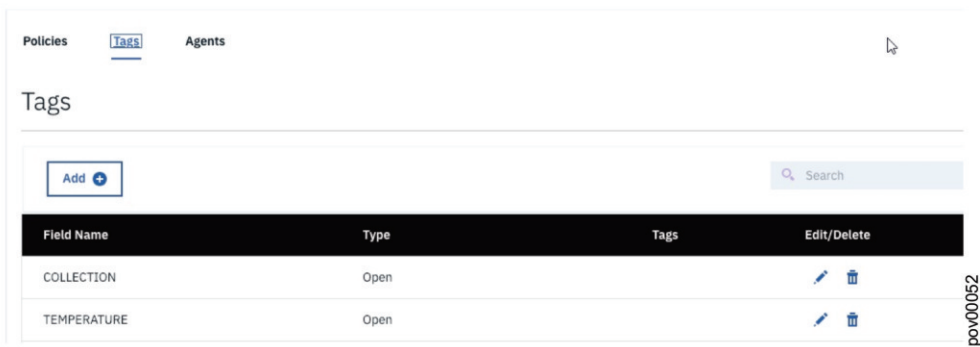


Figure 39. Example of a screen that shows the **TEMPERATURE** tag

Figure 40 on page 79 shows an example of an autotag policy to identify files and objects that have not been accessed for more than a year.

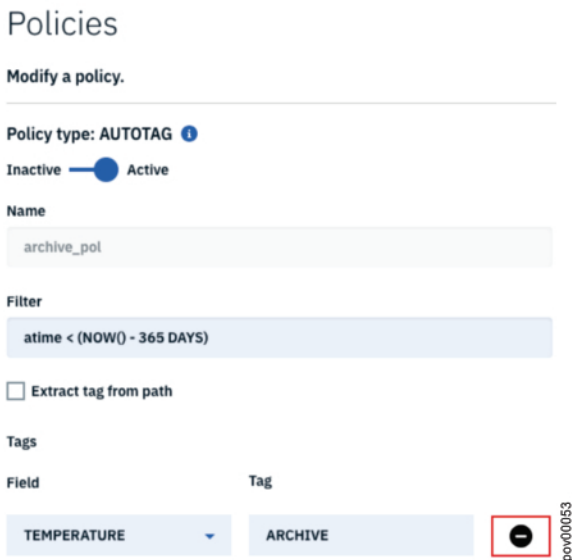


Figure 40. Example of an autotag policy to identify files and objects that have not been accessed in more than one year

Deleting or editing a connection

Use the following information to delete or edit a connection.

About this task

You can delete or edit a connection by using the graphical user interface.

Procedure

1. Click **Admin** to display a listing of existing connections as shown in [Example of a listing of existing connections](#)

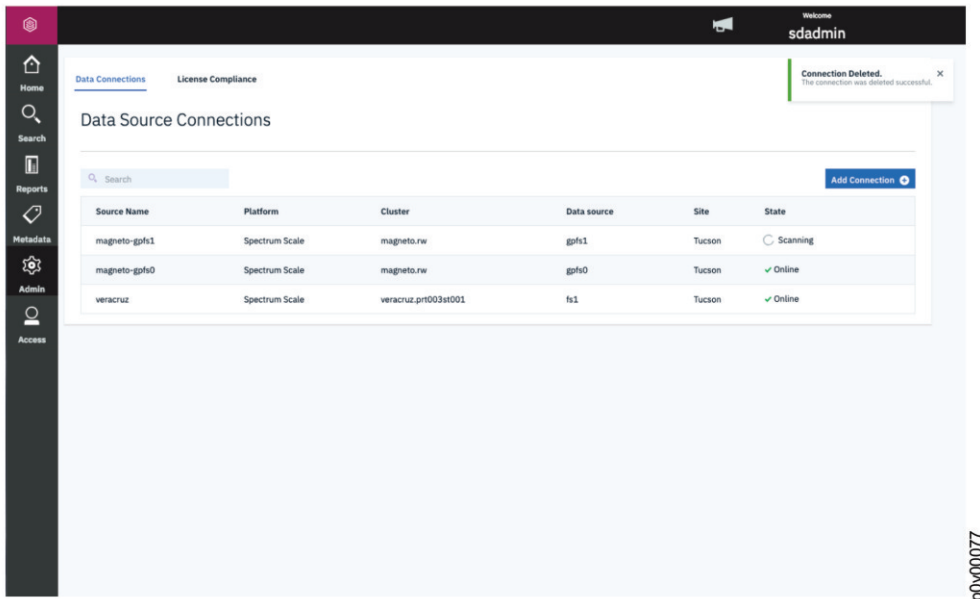


Figure 41. Example of a listing of existing connections

2. Click **Remove** to start the process to remove the data source connection as shown in Figure 42 on page 80.

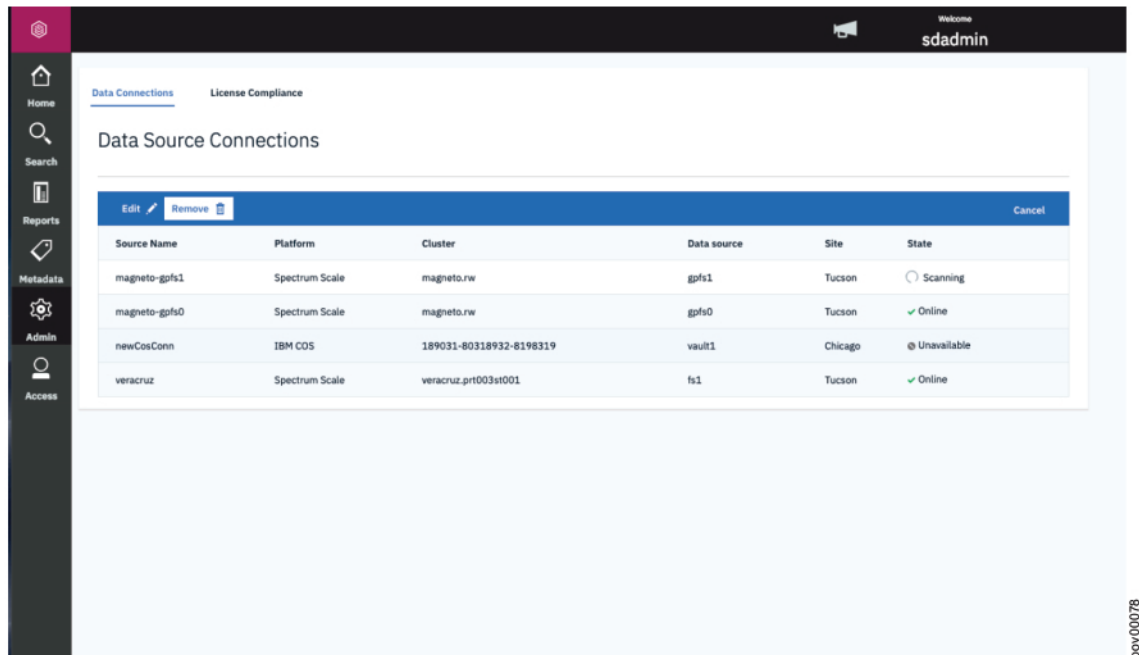


Figure 42. Starting the process to delete a data source connection

3. Clicking **Remove** displays a screen as shown in Figure 43 on page 81. If you are sure you want to delete the connection, click **Delete**.

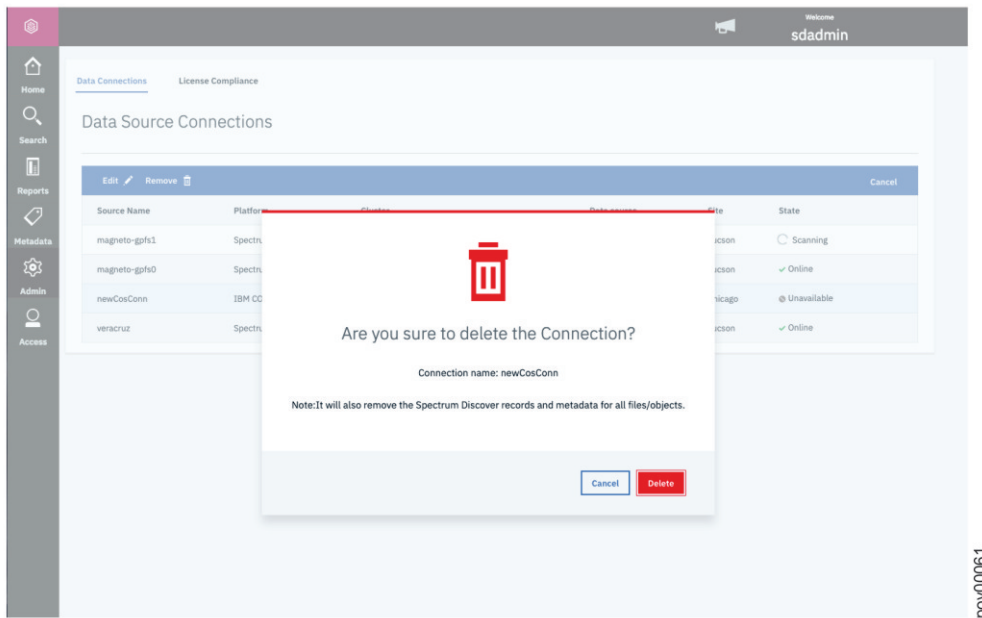


Figure 43. Example of a screen that shows how to delete a connection

4. To edit a connection, click **Edit**.

- a) Edit the appropriate fields in the window for **Edit Data Source Connection**.
- b) Click **Update Connection**.

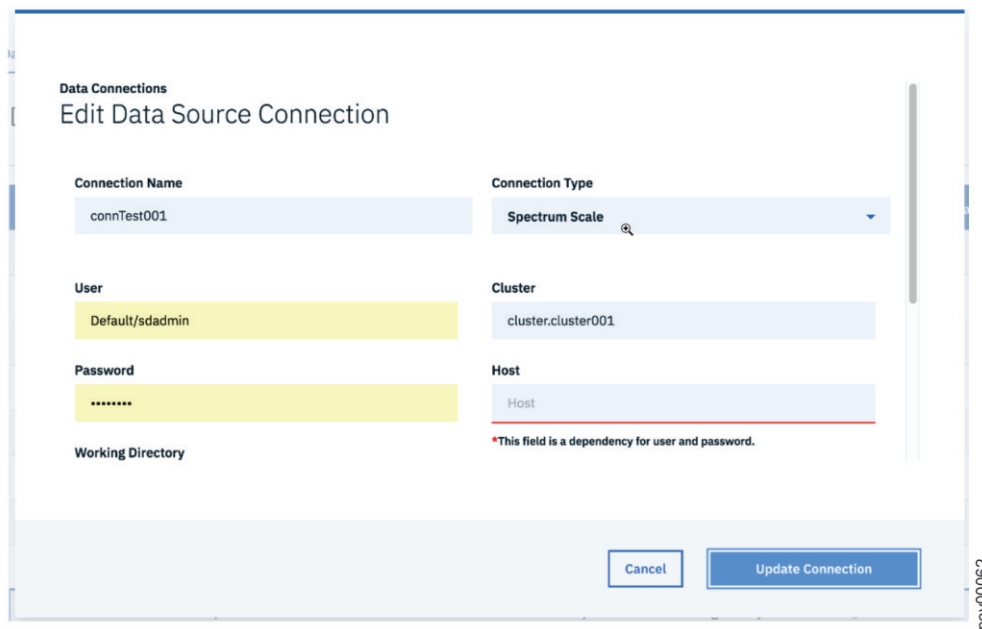


Figure 44. Example of a screen that shows how to edit a connection

Chapter 14. Monitoring the IBM Spectrum Discover environment

You can monitor the health and status of the IBM Spectrum Discover environment and obtain audit log information.

Monitoring the status of the IBM Spectrum Discover environment

You can monitor the health status of the IBM Spectrum Discover environment by using the monitoring dashboards in IBM Cloud Private. IBM Cloud Private is installed on the computer that is running IBM Spectrum Discover.

Opening IBM Cloud Private

Open IBM Cloud Private in a web browser by entering the following URL: `https://sd_computer_address:8443`, where `sd_computer` is the address of the computer that is running IBM Spectrum Discover.

Log on to IBM Cloud Private with the user name `admin` and a password that is stored on the computer that is running IBM Spectrum Discover.

Accessing the IBM Cloud Private password

To access the password for IBM Cloud Private, open a command line on the computer that is running IBM Spectrum Discover and enter the following command:

```
awk '/^default_admin_password/ {print $2}' /opt/ibm/ibm-cloud-private/3.1.2/cluster/config.yaml
```

The command displays the IBM Cloud Private password. Copy the password to access IBM Cloud Private.

Viewing the Dashboards

You can use the IBM Cloud Private dashboard and the Grafana cluster monitoring dashboards to monitor your IBM Spectrum Discover environment.

IBM Cloud Private dashboard

Use the IBM Cloud Private **Dashboard** page to review current system and resource metrics. To open the **Dashboard** page, select **Dashboard** from the IBM Cloud Private menu.

For more information, see [System and resource monitoring](#) in the IBM Cloud Private online documentation.

Grafana cluster monitoring dashboards

Use the Grafana cluster monitoring dashboards to monitor the status of your cluster and applications. To open the monitoring dashboards, select **Platform > Monitoring** from the IBM Cloud Private menu.

For more information, see [IBM Cloud Private cluster monitoring](#) in the IBM Cloud Private online documentation.

Importing the Container and pod status dashboard

The Container and pod status dashboard provide status information for containers and pods in your IBM Spectrum Discover environment. Within the dashboard, green rows indicate pods and containers that are running, and orange rows indicate pods and containers that are not running.

You can import the Container and pod status dashboard into IBM Cloud Private and view it with the other cluster monitoring dashboards.

The Container and pod status dashboard are stored in a JSON file, `pod_container_status_tables.json`, that is located in the following path on the IBM Spectrum Discover computer: `/opt/ibm/metaocean/grafana`. Use a tool such as SmartCloud Provisioning or FTP to copy `pod_container_status_tables.json` to your local computer before you import the file into IBM Cloud Private.

To import the Container and pod status dashboard, open the Grafana cluster monitoring dashboards page and select **Dashboard > Import** to open the **Import dashboard** window. Use the window to select and import `pod_container_status_tables.json` from your local computer.

Monitoring the IBM Spectrum Discover virtual machine

Use the **Monitoring** tab in the VMware vSphere Client to monitor the performance of the IBM Spectrum Discover virtual machine.

To open the **Monitoring** tab, complete the following steps:

1. In the **Navigator** list, click the IBM Spectrum Discover virtual machine to display the details for the machine.
2. From the details view, select the **Monitor** tab.
3. Click **Performance** to view details, including CPU and memory usage.

Audit log

Use the audit log entries to monitor activity of REST API calls within the IBM Spectrum Discover environment, including the API endpoint that was used.

You can obtain the audit log entries by using the FFDC script. For more information, see [“Using the FFDC script”](#) on page 85.

Note: The FFDC script redacts user account and IP address information in the audit log entries.

To view audit log entries, extract the output from the compressed file that is generated by the FFDC script. You can use a text editor to read the FFDC output. Audit log entries are in JSON format and are identified in the FFDC output by the string **AUDIT** in the **type** field.

For more information about API endpoints in the IBM Spectrum Discover environment, see *REST API* in *IBM Spectrum Discover: REST API Guide*.

The audit log includes the following fields:

service

The service that processed the request. The service and node name are included. The following details are optional: namespace, serviceInstance, and containerId.

requestId

The request ID that is returned back to the client, or a correlation tag that is used for internal tracking.

timestampStart

The time that the request was received.

request

The API endpoint that made the request.

serverAddress

The IP address of the server or node that processed the request.

userAgent

The identification string of the user agent that made the request.

type

The log entry type: AUDIT

responseSize

Size of the response, in bytes, sent back to the client.

hostname

The IP address from which the request originates.

protocol

The protocol of the request.

requestLatency

The latency of the request in milliseconds.

responseStatus

The return code that is provided to the client.

auth

The user name and the authentication scheme, bearer (for LDAP) or basic (for local authentication).

Using the FFDC script

The first failure data capture (FFDC) script collects diagnostic and log information about events and conditions in your IBM Spectrum Discover environment. Use the FFDC script to obtain diagnostic and log information or to collect data that can be used by IBM service personnel to analyze problems in your environment.

The FFDC script must be run as the root user on the IBM Spectrum Discover master node.

The FFDC script creates an archived output file within the current working directory. The output file uses the following format: `mo-ffdc-datestamp.tar.xz`. For example, the output might look like this: `mo-ffdc-20180430074006548.tar.xz`.

Note: The FFDC script redacts user account and IP address information.

FFDC script syntax

Log in to the IBM Spectrum Discover node via ssh and enter your password when prompted:

```
[ssh moadmin@spectrum-discover-hostname]
```

The script is located in the directory `/opt/ibm/metaocean/helpers`.

Syntax for use with IBM support

Use the **all** option to collect diagnostic information for use with IBM service personnel.

```
# cd /opt/ibm/metaocean/helpers
# sudo./ffdc all
```

Syntax for collecting audit log entries

Use the **namespaces** option to collect audit log entries.

```
# cd /opt/ibm/metaocean/helpers
# sudo./ffdc namespaces
```

FFDC script options

The FFDC script includes the following options. You can use only one option with the script.

To display a list of options for the script, use the **ffdc** command without an option: `# ./ffdc`

all

The "all" is the standard option that must be used when you are reporting a failure situation to IBM Service. Use this option, unless asked to do otherwise by IBM service personnel.

helm

The "helm" option collects a list of deployed helm charts together with the deployment histories for each of these deployments.

logs

The "logs" option archives some of the log directories, which are located under `/var/log`, from all nodes in the IBM Spectrum Discover cluster, including the DB2 Warehouse logs.

namespaces

The "namespaces" option collects information about all namespaces in Kubernetes. The information that is collected includes description of all the pods within the namespace, logs for all containers within the pods, and a log of events within the namespace. Use this option to collect audit log entries.

services

The "services" option collects service information for a number of services, including Docker, Kafka, and NFS, from all nodes in the IBM Spectrum Discover cluster.

system

The "system" option captures operating system statistics about details such as free disk space, the time since last restart, memory usage, and network ports.

versions

The "versions" option captures the version information for the operating system,

- Docker
- Cloudant®
- Kafka
- Kubernetes

Chapter 15. Updating the network configuration

This topic describes how to update the network configuration.

Follow this procedure to update the network configuration of either master node or any of the worker nodes for IBM Spectrum Discover. The process might take several hours because only one node can be completed at a time. For example:

1. Log into the master node as the moadmin user.
2. Run the following command to change to the configuration directory:

```
cd /opt/ibm/metaocean/configuration
```

3. Run the following command to update your old_fully_qualified_hostname:

```
sudo ./mmconfigappliance -a <old_fully_qualified_hostname>
```

4. Update the network configuration of either the master or worker node to the new network configuration and make sure that the VM starts.

You can use the `sudo ./mmconfigappliance` command to update the network configuration. For example:

- a. Log into node that is acquiring a new network configuration.
- b. Run the following command:

```
cd /opt/ibm/metaocean/configuration
```

- c. Run the following command to change your configuration data:

```
sudo ./mmconfigappliance -n <new FQDN hostname>:<interface>:<new IP>:<netmask>:<gateway>:<dns>
```

5. Run the following command to update your new_fully_qualified_hostname:

```
sudo ./mmconfigappliance -b <new_fully_qualified_hostname>
```

You are prompted for the moadmin password.

The `old_fully_qualified_hostname` must be the old FQDN of either master or worker node that is to be updated. The `new_fully_qualified_hostname` must be the new FQDN of either master or worker node that is to be updated. Additionally, you must run both the `sudo ./mmconfigappliance -a <old_fully_qualified_hostname>` and `sudo ./mmconfigappliance -b <new_fully_qualified_hostname>` commands on the master node.

[Chapter 16. Using a third-party data movement application to move or copy data

Use IBM Spectrum Discover with third-party applications to move or copy data between data sources.

Before you begin

- The third-party application must be registered with IBM Spectrum Discover.
- IBM Spectrum Discover must scan both the source and destination data sources before the data movement.
- The application must be configured to access the same source and destination data sources.

You can see the third-party application documentation for details.

About this task

IBM Spectrum Discover can interact with third-party data movement applications to move or copy data between data sources. Create data management policies on IBM Spectrum Discover, and specify the set of documents to process by using a policy filter. The policy filter can be based on the system metadata or the custom metadata of documents that are collected by IBM Spectrum Discover.

The third-party application registers with IBM Spectrum Discover providing the operations that it supports (move, copy, or both). For each operation, it gives the list of parameter values that it needs to perform the operation.

When you create the data management policy in IBM Spectrum Discover the user defines the filter, the parameter values, and when the policy must run. When the policy runs, IBM Spectrum Discover sends the list of files to the data movement application and any additional parameters. The application processes the files and returns a status summary to IBM Spectrum Discover. The summary is displayed to the user.

Note: [During data migration, the migrated files need to preserve the IBM Spectrum Discover tags. You can follow a manual procedure to preserve the tags. For more information, see [“Preserving tags during data movement”](#) on page 90.]

Procedure

1. Log in to the IBM Spectrum Discover GUI.
2. Go to **Admin > Management Policies**.
3. To create a policy, click **Add Policy**.
4. Click the slider control and set the status to one of the following values:

Active

An active policy runs whenever its scheduling event is reached.

Inactive

An inactive policy does not run even when its scheduling event is reached (including a NOW event).

5. Enter a policy name.
6. Enter a policy filter.
The policy filter includes the criteria for selecting the files for moving or copying. For example, `filetype="pdf"` selects all files of type PDF.
7. To select the policy type, click **Next Step**.
8. Select **MOVE** or **COPY** as the policy type.
9. Select the agent name as the **Agent**.

- Enter the remaining parameters. The parameters that are displayed depend on the application, and these parameters might include:

Source connection type

Indicates the type of connection that the files currently reside on.

Source connection

Indicates the name of the connection that the files currently reside on.

Destination connection type

Indicates the type of connection that the files are being moved or copied to.

Destination connection

Indicates the name of the connection name that the files are being moved or copied.

Force migrate

Indicates whether to force demigration or recall of the file at source location when it is migrated to other location before you perform the operation.

Overwrite

Indicates what value to give when a file exists at the destination.

Preserve attributes, timestamp, or permissions

Indicates the parameters to control whether the files metadata is preserved.

- To enter a schedule, select **Next Step**.

The schedule indicates when you want to start the move of the copy.

- To review the policy, select **Next Step**.

- To create the policy, select **Submit**. The policy runs at the scheduled time.

- When the policy runs or completes an execution status summary, view it by clicking **Policy Preview**.

]

Preserving tags during data movement

Before you begin

Generate a report of the files to be moved or copied before you start the data movement process. This report includes the relevant tags for each source file. A sample report is shown in the following table:

Table 3. Report generated for moving files

Path	Filename	Filetype	Datasource	tagA	tagB
/	Chrysanthemum.jpg	jpg	dir1	image	value1
/	Lighthouse.jpg	jpg	dir1	image	value1
/	newtextfile.txt	txt	dir1	text	value2
/	Hydrangeas.jpg	jpg	dir1	jpg	value1
/	Thumbs.db	db	dir1	unknown	value1
/	Koala.jpg	jpg	dir1	jpg	value1
/	Desert.jpg	jpg	dir1	jpg	value1

About this task

If IBM Spectrum Discover tags must be preserved during data migration, then you can follow a manual procedure to ensure that the tags are preserved in the moved or the copied files.

Procedure

- Run the data movement process.

2. Rescan the destination connection after the data movement completes successfully.
3. Refer to the information provided in the report, shown in [Table 1](#) that was generated before you started the data movement process, to identify the files for which you need to preserve the tags.

Note: You can define a filter to identify and select the correct set of files that must be tagged. For example, if we want to move files shown in [Table 1](#) to a new datasource (ddsource), you can define filters that include the actions to be taken for the tag values. For a filter defined as shown:

```
datasource in ('ddsource') and filetype in ('jpg')
```

You can create an AUTOTAG policy that sets the value of tagA to image.

This process assumes either of the two things.

- No other files belonging to these filetypes exist in the destination datasource.
 - Those files can also be tagged with the tagA values.
4. Reapply the tags (for example, tagA) on the moved or copied files by using the auto tagging capability of IBM Spectrum Discover.

Note:

You must perform a search using the above filters to check that it retrieves the correct documents that are being selected. You can also sort, based on the report headings in MS Excel, to identify a suitable filter.

It is not easy to identify a suitable filter to reapply the tagB values shown in [Table 1](#) which indicates the difficulty in manually reapplying the tags.

This solution is feasible only where you can define a small number of filters that identify the groups of files to which you need to apply the tags. If filters cannot be defined then you must run several autotag policies, which might be an impractical and tedious process.

Chapter 17. Disaster recovery procedures

Use this process to recover from a disaster that involves an IBM Spectrum Discover system discusses the following scenarios:

- Recovery from the entire loss of a single node IBM Spectrum Discover deployment.
- Recovery from the loss of a single node in a multi-node IBM Spectrum Discover deployment or the entire multi-node system.

Preparations for disaster recovery

Before you need to perform disaster recovery, there are several tasks that must be accomplished to ensure the ability to recover.

About this task

Procedure

1. Take a backup of the IBM Spectrum Discover system as described in [Chapter 10, “Backup and restore,”](#) on page 67.
2. Record the installation configuration, including:
 - Network settings
 - Storage settings
 - CPU and memory
 - IBM Spectrum Discover version
3. Ensure that the physical and virtual infrastructure is available to replace the system that might fail.
4. You cannot recover to a different version of IBM Spectrum Discover. If a change of version is required, you need to recover to the current version before you install the upgrade.

If you perform a recovery on a system that is upgraded, it is recovered directly to the upgraded IBM Spectrum Discover version. For example, if IBM Spectrum Discover version 2.0.0.3 is deployed and then upgraded to IBM Spectrum Discover version 2.0.1 and then recovered from a failure, the recovery goes directly to the 2.0.1 version.

Running disaster recovery

You can recover single-node and multi-node systems (disaster recovery).

About this task

Follow these steps to recover single-node and multi-node systems.

Procedure

1. Record the time of the failure.
2. If the virtual machine that hosts IBM Spectrum Discover is running, shut it down.
3. Redeploy the system as described in these sub steps.
 - a) For a single node system, redeploy with the same parameters as the failed deployment. For more information, see *.For more information, see the topic [Configure networking and performing provisioning of a single node trial or single node production IBM Spectrum Discover virtual appliance in IBM Spectrum Discover: Concepts, Planning, and Deployment Guide.](#)*
 - b) For a multi-node system, redeploy with the same parameters as the failed deployment. For more information, see the topic *[Configure networking and perform provisioning for the IBM Spectrum](#)*

Discover multi-node virtual appliance cluster in *IBM Spectrum Discover: Concepts, Planning, and Deployment Guide*.

4. For more information, see the topic *Initial setup configuration* in the *IBM Spectrum Discover: Administration Guide*.
5. For more information, see the topic *Running a restore* in the *IBM Spectrum Discover: Administration Guide*.
6. Do not remove the system from maintenance mode until you complete the following sub steps.
 - a) For a multi-node system, confirm that DB2 Warehouse is running, and use this command to determine the HEAD node. This command provides output for a multi-node system similar to the following table.

```
docker exec -it Db2wh status
```

NodeName	IP	Type	Role	State
ch3-gc1000-11535	172.26.7.223	DATA	ACTIVE	UP
ch3-gc1000-11536	172.26.7.221	DATA	ACTIVE	UP
ch3-gc1000-11537	172.26.7.222	HEAD	ACTIVE	UP

- b) For multi-node, log in to the HEAD node IP address.
- c) Change the database password to the new value of the deployment.
 - 1) Record the value of the database password. The encrypted value is stored in the `/opt/ibm/db2wh/password` directory. Run the following command to decrypt the value:

```
PYTHONPATH=/opt/ibm/metaocean/provisioning/filter_plugins python -c  
"from metaocean import password_decode; print password_decode('$(sudo cat  
/opt/ibm/db2wh/password)')"
```

- 2) Run the following command to record the value of the database password that is stored in `/opt/ibm/db2wh/password`:
 - 3) Run the following command to update the DB password with this value:

```
docker exec -it Db2wh setpass <DB password>
```

- 4) Run the following command to log in to the DB docker container and change to the DB user:

```
docker exec -it Db2wh bash  
su - db2inst1
```

- 5) [Run the following commands to restart the DB:

```
/opt/ibm/dsserver/bin/stop.sh  
/opt/ibm/dsserver/bin/start.sh
```

]

- d) If you use Cloud Object Storage events, update the IBM Cloud Object Storage notification certificate:
 - 1) Take a copy of the Kafka SASL password. The password is stored in `/etc/kafka/sasl_password`
 - 2) Take a copy of the CA PEM certificate. The certificate is stored in `/etc/kafka/ca.crt`
 - 3) Apply these details to the IBM Cloud Object Storage notifications. For more information, see the topic *Configure IBM Cloud Object Storage notifications for IBM Spectrum Discover* in *IBM Spectrum Discover: Concepts, Planning, and Deployment Guide*.
7. Remove the system from maintenance mode and record the recovery time.

8. If IBM Cloud Object Storage notifications are being used to keep the metadata up to date, run the Replay procedure for the time in which the system is unavailable. For more information, see For more information, see the topic *Replay* in *IBM Spectrum Discover: Concepts, Planning, and Deployment Guide*.
9. For any other data source, repeat the scan procedure if it was due to be run during the time IBM Spectrum Discover was unavailable.

Chapter 18. Troubleshooting

[Best practices

This topic describes certain best practices that make the IBM Spectrum Discover troubleshooting process easier.

]

Changing system time breaks jobs and pods

Changing the system time of IBM Spectrum Discover nodes causes problems with jobs and pods within IBM Cloud Private. Ensure that the system time is correctly set and ntp is configured before installing the IBM Spectrum Discover cluster.

Configure the IBM Spectrum Discover virtual appliance network time protocol (NTP) settings by using the following command:

```
sudo /opt/ibm/metaocean/configuration/mmconfigappliance -t <NTPServer>
```

To test that the time has been correctly configured, use the following command:

```
date
```

If the system is not correctly configured before deployment and needs to be corrected after the system is installed, following configuration test the system with the following command:

```
kubectl get pods --all-namespaces
```

If the command hangs, reboot the server and allow up to 30 minutes for the system to come back online fully.

Recovering original SSH keys

You might have to recover your original Secure Shell (SSH) key pair.

If the SSH authentication keys for IBM Cloud Private are compromised in any way, it might result in an unrecoverable error that uninstalls IBM Cloud Private during the upgrader process:

```
UNREACHABLE! => {"changed": false, "msg": "Failed to connect to the host via ssh: Permission denied (publickey,gssapi-keyex,gssapi-with-mic,password).\r\n", "unreachable": true}
```

You must recover the original SSH keys before you run the upgrader again:

1. Log in to the master node as the moadmin user.
2. Run this command: `cd /opt/ibm/metaocean/provisioning`
3. Run this command: `sudo cp ./ssh_key /opt/ibm/ibm-cloud-private/<version>/cluster/`
4. Run this command: `sudo chmod 700 /opt/ibm/ibm-cloud-private/<version>/cluster/ssh_key`
5. Run this command: `sudo ssh-copy-id -i ssh_key -o StrictHostKeyChecking=no moadmin@<hostname of master node>`
6. Enter the password when you are prompted.

How to recover a system after a YUM update

You might have to recover your system after you run a Yellowdog Updater, Modified (YUM) update.

For example:

1. If you update the kernel module for IBM Spectrum Scale after you run a YUM update, you must restart the system to make sure that the new kernel is functioning correctly.
2. Run this command to rebuild the kernel module:

```
sudo /usr/lpp/mmfs/bin/mmbuildgpl
```

(The **mmbuildgpl** command manages or verifies prerequisite packages for Linux® and also builds the GPFS portability layer.)

3. Restart the system to make sure IBM Spectrum Scale starts up correctly.

Important: For more information about kernels that are supported by IBM Spectrum Scale, see: <https://www.ibm.com/support/knowledgecenter/STXKQY/gpfsclustersfaq.pdf?view=kc>

[Configuration issues

This topic describes configuration issues that you might encounter when you use IBM Spectrum Discover.

Healthy default pod list

To list all pods, execute this command:

```
kubectl get pods --all-namespaces
```

If any of the following pods are not running on your system, contact IBM Support.

- *-auth-ibac-auth-*
- *-auth-ibac-keystone-*
- *-consumer-cos-consumer-* (x10)
- *-consumer-scale-le-consumer-* (x10)
- *-consumer-scale-scan-consumer-* (x10)
- *-db2wh-rest-*
- *-db2warehouse-mpp-prod-* (at least one for each node)
- *-metaocean-api-*
- *-producer-cos-producer-*
- *-producer-scale-le-producer-*
- *-producer-scale-scan-producer-*
- *-ui-backend-*
- *-ui-frontend-*

[Data issues

This topic describes data issues that you might encounter when you use IBM Spectrum Discover.

Delete markers from IBM Cloud Object Storage are ignored

When a delete marker is created within IBM Cloud Object Storage, a `CreateDeleteMarker` or `CreateDeleteMarker:NullVersionDeleted` notification is emitted. These notifications are currently not processed by IBM Spectrum Discover.

Records are not ingested after reboot

After rebooting a IBM Spectrum Discover server, the producer pods might need to be restarted using `kubectl` in order for them to ingest data. The problem is caused by a race condition with the connection management service.

Perform the following steps at the command line.

1. List the producer pods:

```
kubectl get pods --all-namespaces | grep producer
```

The namespace is in the first column and the pod name is in the second column of the `kubectl get pods` output.

2. For each producer pod, use `kubectl` to delete the pod.

```
kubectl delete pod -n=<namespace> <pod name>
```

Failure to do this might result in valid data being discarded at ingest with no obvious notification to the user.

Recovering from data ingestion consumer or producer issues

When a producer or consumer application running in a pod encounters an error that causes the application to halt, the pod restarts. When a recovery action is carried out, that means you must restart the pods. The following actions might be taken by the IBM Spectrum Discover administrator on the IBM Spectrum Discover cluster master node.

1. View the status of running pods for consumer and producers as follows:

```
$ kubectl get pods -n Namespace
```

Where Namespace might be one of the following:

namespace	: Description
producercos	: IBM Spectrum Discover COS Producer
producerscalescan	: IBM Spectrum Discover Scale Scan Producer
producerscalele	: IBM Spectrum Discover Scale Live Event Producer
consumercos	: IBM Spectrum Discover COS Consumers
consumerscalescan	: IBM Spectrum Discover Scale Scan Consumers
consumerscalele	: IBM Spectrum Discover Scale Live Event Consumers

You can expect to see 10 running pods per consumer deployment, and 1 running pod per producer deployment. For example:

```
$ kubectl get pods -n consumerscalescan
```

NAME	READY	STATUS
anxious-fly-consumer-scale-scan-consumer-56b6c4ff9c-d4k8v	1/1	Running
anxious-fly-consumer-scale-scan-consumer-56b6c4ff9c-h862t	1/1	Running
anxious-fly-consumer-scale-scan-consumer-56b6c4ff9c-j4649	1/1	Running
anxious-fly-consumer-scale-scan-consumer-56b6c4ff9c-ksbh4	1/1	Running
anxious-fly-consumer-scale-scan-consumer-56b6c4ff9c-kt9sc	1/1	Running
anxious-fly-consumer-scale-scan-consumer-56b6c4ff9c-lk8jz	1/1	Running
anxious-fly-consumer-scale-scan-consumer-56b6c4ff9c-p2lr6	1/1	Running
anxious-fly-consumer-scale-scan-consumer-56b6c4ff9c-qqhfd	1/1	Running
anxious-fly-consumer-scale-scan-consumer-56b6c4ff9c-wknbc	1/1	Running

```

0    anxious-fly-consumer-scale-scan-consumer-56b6c4ff9c-zrp6k    1/1    Running
    46m

$ kubectl get pods -n producercos

NAME                                READY   STATUS
RESTARTS   AGE
0    exacerbated-tarsier-producer-cos-producer-64748764cf-pctwz    1/1    Running
    3h

```

You can view the logs for a pod as follows:

```
$ kubectl -n Namespace logs Name
```

Where Namespace is one of the items from this listing and Name is the name of a specific pod from the get pods output. For example:

```

$ kubectl -n consumerscalescan logs anxious-fly-consumer-scale-scan-consumer-56b6c4ff9c-zrp6k

Options provided:
-----
Application = scale
DB Protocol = http
DB IP = db2whrest.db2whrest:80
Broker IP = 203.0.113.15:9093
DB name = metaocean
Topic = scale-scan-topic
Group ID = mo1
DB User = bluadmin
Batch size = 50000
Log directory = none
initial scan = true
mode = update mode
-----
                          Starting MetaOcean Consumer
-----
PID: 7
Construct InFromKafka object
broker=203.0.113.15:9093 topic=scale-scan-topic group=mo1
Create DatabasePayload object
Construct Db2whOutputStream object
created kafka consumer rdkafka#consumer-1
librdkafka version is 0.11.0(721151)
Successfully opened connection to Kafka
Create DatabasePayload object
Construct OutToKafka object
Found Kafka SSL Client Certificate
Found Kafka SSL Client Key
Created producer : rdkafka#producer-2
librdkafka version is 0.11.0(721151)
created topic_handle
Created topic handle: 0x20f87c8 with name consumer-debug-topic
Construct Logger object, log directory not specified, direct output to STDOUT
Create ConsumerLogger object
Construct MessageConsumer object
Construct ScaleConsumer object
Create DatabasePayload object
No throttle control file
2018-10-18 23:22:29.153 > rebalance_cb: partitions_assigned:[{topic: scale-scan-topic,
part: 9, offset: -1001}]

```

2. Delete and reinstall consumers and producers as follows:

- a. Obtain a list of active application deployments using the following curl commands to communicate with the API server.
 - a1. Obtain the bearer token used to authenticate the REST calls to the API server endpoints. For more information, see Authentication process in IBM Spectrum Discover: REST API Guide
 - a2. Obtain a list of application charts and their deployments in the IBM Spectrum Discover cluster. The first column of the output is the chart name and the second column is the deployment name.

```
$ curl -s -k -H "Authorization: Bearer $TOKEN" -H "Accept: application/json" -X GET https://localhost/api/application/ |jq '.[] | "\(.chart) : \(.deployments[.deployment)">'
```

%	Total %	Received %	Xferd	Average Dload	Speed Upload	Time Total	Time Spent	Time Left	Current Speed		
100	3194	100	3194	0	0	298	0	0:00:10	0:00:10	--:--:--	911
"auth-ibac : invited-boxer"											
"connmgr : washing-mole"											
"consumer-cos : lazy-vulture"											
"consumer-scale-le : plinking-quoll"											
"consumer-scale-scan : anxious-fly"											
"db2wh-rest : quarrelsome-hamster"											
"metaocean-api : mean-bronco"											
"policyengine : hopping-blackbird"											
"producer-cos : intended-dingo"											
"producer-scale-le : foppish-donkey"											
"producer-scale-scan : quelling-dachshund"											
"ui : worn-hummingbird"											

b. Delete a deployment as follows:

```
$ curl -k -H "Authorization: Bearer $TOKEN" -H "Accept:application/json" -X DELETE
https://localhost/api/application/Deployment_Name
```

c. Restart a deployment by reinstalling the associated chart.

```
$ curl -k -H "Authorization: Bearer $TOKEN" -H "Content-Type:application/json" -H
"Accept: application/json" -X OST -d "{\"chart\":\"Chart_Name\", \"repository\":\"metaocean
\", \"version\":\"\"}" https://localhost/api/application/
```

Where Deployment_Name is the name of the deployment associated with the application. For example, in "producer-cos : intended-dingo", the deployment name for the COS producer is intended-dingo.

ATTENTION: Do not install more than one instance of a specific chart at one time.

IBM Spectrum Discover scan data not landing in database

An error can occur if the clocks on the IBM Spectrum Discover node are ahead (in time) of the clocks on the IBM Spectrum Scale system that is being scanned.

This error occurs because the security certificates that are used for in-flight data encryption have a time stamp that must not contain a future value.

If this situation occurs, you get a message like the following message in the scan log, which is kept in the IBM Spectrum Discover node under the /gpfs/gpfs0/connections/scale directory:

```
[1558076225.325]FAIL|rdkafka#producer-1| [thrd:ssl://9.11.201.53:9093/bootstrap]: ssl://9.11.201.53:9093/
bootstrap: Failed to verify broker certificate: certificate is not yet valid (after 21ms in state CONNECT)
```

Note: You must use the Network Time Protocol (NTP) on both systems to avoid this situation. This means that you must make sure that the clocks of the host IBM Spectrum Scale node are not significantly different from the clocks on IBM Spectrum Discover system.

[Viewing live reports might list incorrect results

When you view live reports, the reports might list incorrect results if original report criteria is lost.

Contents from any saved reports can be generated live by using the See on table hyperlink within the report summary. Summary (or grouped) reports also provide a table of results that you can further refine.

During the process of refining a summary report, the original report criteria might become lost or unavailable. If the criteria is lost or unavailable, incorrect results display in the search table. To prevent this situation, start the search from the search panel rather than a report.

]

[Error diagnosis issues

This topic describes error diagnosis issues that you might encounter when you use IBM Spectrum Discover.

Ansible® Warnings

This information describes how to handle ansible warnings.

You might encounter either of the following messages:

- [WARNING]: Could not match supplied host pattern, ignoring: non-controller
- [WARNING]: Could not match supplied host pattern, ignoring: controller

For example, you see the following warning in the upgrader logs - but the upgrader seems to be working:

```
TASK [Create kafka non-controller group] *****
Monday 06 January 2020 11:50:00 +0000 (0:00:00.027) 0:15:38.292 *****
skipping: [9.37.136.50]
[WARNING]: Could not match supplied host pattern, ignoring: non-controller
```

If Kafka does not require an update during the upgrader process, you see this warning while the upgrader code is looking for either a non-controller group or controller group. The upgrader does not run tasks that target either of these groups if the groups do not exist.

Important: You can safely ignore these warnings.

ens160 activation errors in /var/log/messages

ens160 activation errors appearing in /var/log/messages can be safely ignored.

As an example:

```
NetworkManager[1039]: [1540162458.1216] device (ens160): activation-stage: schedule
activate_stage5_ip6_config_commit,10 which replaces activate_stage5_ip6_config_commit,10 (id
171022 -> 171024).
```

kubectl returns "error: You must be logged in to the server"

There is a bug in ICP version 2.1.0.3 that can cause authentication to stop working when the authorization service starts before the mongodb service that it depends on. This can also cause the helm list command to fail.

You can confirm this error at the command line by running the following command:

```
sudo /etc/cron.hourly/icp_login.sh
```

and checking for output similar to:

```
Logging into ICP spectrumdiscover Cluster
API endpoint: https://10.3.23.168:8443
Authenticating...

OK

FAILED
Error response from server. Status code: 500; message: {"error":
{"statusCode":500,"message":"Internal Server Error"}}

Configuring Cluster spectrumdiscover
FAILED

Cannot connect to a back-end service. Try again later. (E0004)
Incident ID: 90cb3e84-935a-4a8e-9687-c8ab641c11dd
```

To fix the issue, first use an alternative method to enable kubectl:

```
mkdir ~/.kube
sudo cp /var/lib/kubelet/kubelet-config ~/.kube/config
sed -i -e 's/kubelet.crt/kubecfg.crt/' -e 's/kubelet.key/kubecfg.key/g' ~/.kube/config
```

Next, restart the auth-idp pod:

1. `kubectl get pods -n kube-system | grep auth-idp`
2. `kubectl delete pod -n kube-system <pod name from previous command>`

In some cases, this still leaves the `helm list` command failing with the error `Error: the server could not find the requested resource (get configmaps)`. To fix this error, restart the tiller-deploy pod:

1. `kubectl get pods -n kube-system | grep tiller-deploy`
2. `kubectl delete pod -n kube-system <pod name from previous command>`

[Installation issues

This topic describes installation issues that you might encounter when you use IBM Spectrum Discover.

Db2 Warehouse installation port conflict - Wait for Db2wh to initialize

Occasionally, ports required by Db2 Warehouse will be used by ICP services that select a random port in a high range. When this happens, IBM Spectrum Discover installation fails at the step "Wait for DB2WH to initialize", and the DB2 Warehouse logs contains the error "FATAL RUNTIME ERROR DETECTED".

To recover the installation:

1. Reboot the node.
2. Delete previous Db2wh container:

```
sudo docker rm -f Db2wh
```

3. Re-run ansible:

```
cd /opt/ibm/metaocean/configuration
sudo ./launch_ansible
```

IBM Cloud Private install logs are missing

The IBM Cloud Private installation logs are not included in `/opt/ibm/metaocean/provisioning/ansible.log`. If you have trouble with the `launch_ansible.sh` script installing IBM Cloud Private, use the IBM Cloud Private install logs in `/opt/ibm/ibm-cloud-private/<version>/cluster/`.

[Networking issues

This topic describes networking issues that you might encounter when you use IBM Spectrum Discover.

CentOS reboots under load

CentOS might reboot under load due to a kernel bug.

For more information, see <https://access.redhat.com/solutions/3492911>.

IBM Cloud Object Store will not connect to the IBM Spectrum Discover kafka server by IP address

When connecting to the kafka server, IBM Cloud Object Store uses TLS to validate the certificate presented by the server. IBM Spectrum Discover includes the hostname in the certificate but not the IP address.

To fix the problem, use the IBM Spectrum Discover hostname within the IBM Cloud Object Store configuration instead of the IP address.

IBM Spectrum Scale can fail to load after an ESXi server is rebooted

When an ESXi server is rebooted, it is possible that the MAC address associated with the virtual machine can change. This stops IBM Spectrum Scale from starting within the IBM Spectrum Discover cluster. It can be corrected by updating the MAC address.

Check for the following error in the IBM Spectrum Scale logs found in `/var/adm/ras/mmfs.log.latest`:

```
mmautoload: Unable to determine the local node identity.  
Mon Jun 25 22:32:45 UTC 2018 mmautoload: GPFS is waiting for daemon network
```

To address the issue:

1. Get the network configuration file MAC address from the file `/etc/sysconfig/network-scripts/ifcfg-ens<n>`, in the `HWADDR` property.
2. Get the MAC address for the network interface using the `ip a` command, in the `link/ether` property.
3. Update the network configuration file with the new MAC address.
4. Reload the connection: `nmcli con reload /etc/sysconfig/network-scripts/ifcfg-ens<n>`
5. Bring up the connection: `nmcli con up ens<n>`

Multi-node network settings get stuck while checking the Docker run status

Your multi-node network settings can get stuck while you check whether the Docker is running when the `mmconfigappliance` command is run. You must run the script if it gets stuck during the Docker service restart.

Follow this procedure:

1. Run this command to end all Docker containers:

```
sudo docker kill $(docker ps -q)
```

2. Run this command to stop the Docker:

```
sudo systemctl stop docker
```

3. Remove the Docker lock files:

```
sudo rm -f /var/run/docker /var/run/docker.*
```

4. Restart the Docker:

```
sudo systemctl start docker
```

Network configuration update: Failure recovery steps

Occasionally, the network configuration can become unstable when performing a network configuration update. For example, if incorrect options have been used, network connection to the master virtual machine is broken. This information helps recover the system in these cases.

The network configuration update is a two-step process.

- pre: `sudo ./update_network -a <old_FQDN>`
- post: `sudo ./update_network -b <new_FQDN>`

Recover from failure during pre

Before re-running the pre-steps, check the following:

hosts files

1. `/etc/hosts`
Confirm that aliases still point to the old configuration
2. `/opt/ibm/metaocean/provisioning/hosts`
Confirm that appropriate IPs point to the old configuration

Recover from failure during post

Before re-running the post steps, check the following:

hosts files

During the pre-steps, a backup is made of the hosts file. Ensure the backups are correct and point to the old IPs and hostnames where appropriate. The backups can be used to reset the hosts files using the commands below.

1. `sudo cat /etc/hosts.orig > /etc/hosts`
Confirm that aliases point to the new configuration
2. `sudo /opt/ibm/metaocean/provisioning/hosts.orig > /opt/ibm/metaocean/provisioning/hosts`
Confirm that appropriate IPs point to the new configuration

The post steps will update the hosts files again with the new configuration.

ICP

The post steps might have attempted an install of ICP. Before re-running post steps, uninstall ICP first. If ICP is already uninstalled, these steps will produce a message to that effect, which can be ignored.

To uninstall ICP manually run these steps:

1. Log into master node using the moadmin user
2. `cd /opt/ibm/ibm-cloud-private/3.1.2/cluster/`
3. `sudo docker run -e LICENSE=accept --net=host -t -v "/opt/ibm/ibm-cloud-private/3.1.2/cluster":/installer/cluster ibmcom/icp-inception:3.1.2-ee-sd uninstall`

Network configuration update: Error creating metaocean tables with Liquibase

During a network configuration update, Liquibase can fail to create metaocean tables.

The error presented looks like this:

```
Unexpected error running Liquibase:  
com.ibm.db2.jcc.am.DisconnectNonTransientConnectionException: [jcc][t4][2043]  
[11550][3.72.30] Exception java.net.ConnectException: Error opening socket to
```

To address this issue, run the following commands:

1. `cd /opt/ibm/metaocean/provisioning`
2. `ansible-playbook -s mo_config_post_icp.yml`

```
3. ansible-playbook -s network_config_master_cleanup.yml --extra-vars
"old_ip=<old_ip>"
```

[Network settings change hangs while uninstalling IBM Cloud Private

Uninstalling IBM Cloud Private from IBM Spectrum Discover can take several minutes.

Uninstalling IBM Cloud Private can cause the system to hang.

If the docker process is blocked for more than 120 seconds, warnings are reported in the system logs.

If the system hangs, restart it and rerun the network change command.

]

[Network settings change fails when you run mmconfigappliance

The following information describes what to do if your network settings change fails when you run the IBM Spectrum Discover **mmconfigappliance** command.

Important: An unrecoverable error ("Unexpected Client Error") can occur during a database restore when you run this command to administer network changes:

```
"TASK [network_update_i_restoredb : Restore db2 backup]"
```

Complete the following steps so your network setting changes do not fail when you run the **mmconfigappliance -b** command:

1. Run this command to create an empty database:

```
docker exec -i Db2wh bash -c "su db2inst1 -c '. /mnt/blumeta0/home/db2inst1/sqllib/
db2profile && db2 create db BLUDB;'"
```

2. Run this command to stop the **log_chopper** service:

```
sudo systemctl stop log_chopper
```

3. Run this command to stop the Db2 Warehouse container:

```
docker stop Db2wh
```

4. Run this command to stop GPFS:

```
sudo /usr/lpp/mmfs/bin/mmshutdown
```

5. Run this command to stop GPFS (gpfs0) from auto-mounting:

```
sudo /usr/lpp/mmfs/bin/mmchfs gpfs0 -A no
```

6. Run the **mmconfigappliance -b** command:

```
sudo ./mmconfigappliance -b <new_hostname>
```

7. Run this command to enable GPFS (gpfs0) to automount:

```
sudo /usr/lpp/mmfs/bin/mmchfs gpfs0 -A yes
```

8. Run this command to start the **log_chopper** service:

```
sudo systemctl start log_chopper
```

]

Pod stuck in CreateContainer error

When you change the system time, you might encounter an issue where the start time is earlier than the finish time for the pod.

You might also see a message similar to this:

```
The container name "XXX" is already in use by container "YYY". You have to remove (or rename) that container to be able to reuse that name.
```

Try to delete whichever pod is stuck in the terminating state to see whether that fixes it. For example, run this command:

```
kubectl delete pod -n <namespace> <pod name>
```

If the pod returns to the same state, run this command:

```
`kubectl describe pod -n | grep "Container ID"
```

From this listing of container IDs, delete the IDs one by one. For example, run this command:

```
"docker container rm <container id>"
```

Note: You might see a message that states that the container does not exist. If you do, continue deleting the rest of the containers.

After completion, verify that the stuck pod is now removed. For example, run this command:

```
kubectl get pods --all-namespaces
```

Pod stuck in terminating state

You might have to delete a pod that is stuck in a terminating state.

When you change the system time, you might encounter an issue where the start time is earlier than the finish time for the pod. Try to delete whichever pod is stuck in the terminating state to see whether that fixes it. For example, run this command:

```
kubectl delete pod -n <namespace> <pod name>
```

If the pod returns to the same state, run this command:

```
`kubectl describe pod -n | grep "Container ID"
```

From this listing of container IDs, delete the IDs one by one. For example, run this command:

```
"docker container rm <container id>"
```

Note: You might see a message that states that the container does not exist. If you do, continue deleting the rest of the containers.

After you are done, verify that the stuck pod is now removed. For example, run this command:

```
kubectl get pods --all-namespaces
```

[Performance issues

This topic describes performance issues that you might encounter when you use IBM Spectrum Discover.

]

Changed permissions for the current user are not effective until logout

When adding permissions to a user using a group (for example, adding the data admin role to the sdadmin user), the new permissions will not be effective in the user interface until the user logs out of Spectrum Discover and logs back in again.

Blank queries to the search API time out

An unqualified query to the IBM Spectrum Discover REST API /search endpoint might result in a timeout. To workaround this problem, include a query.

Example unqualified request:

```
{"query": "", "filters": [], "group_by": [], "sort_by": [], "limit": 100000}
```

Example query to search for all files:

```
{"query": "filename like '%'", "filters": [], "group_by": [], "sort_by": [], "limit": 100000}
```

[Policy issues

This topic describes policy issues that you might encounter when you use IBM Spectrum Discover.

A collection policy cannot be added to a collection or edited after the collection is created

You must create a collection policy at the same time that you create a collection. You cannot add a collection policy after you create the collection. You can enter the policy details and save the collection with the UI, but the policy is never created.

The workaround is to always create a policy when you create a collection. If the details of the policy are unknown at the time, you can create an inactive policy and update it when the details are known.

Note: You cannot edit a collection policy through the edit collection screen. The collection policy must be edited in the metadata screens.

Tagging policy failures under high load

When running multiple tagging policies in parallel or in periods of high stress load on the database, tagging batches can fail due to transaction timeout at the database.

In the event of such a failure, the user receives a notification on the GUI policy status table in the Progress column, below the percent completed.

When this happens, an administrator can run the policy or policies again to clean up the missed records, however, it can be desirable to reduce the working set to just the records that were missed. This can be accomplished by modifying the filter to exclude records that have already been tagged.

As an example, take a tagging policy to set the **TEMPERATURE** tag to *ARCHIVE* for the filter:

```
project = 'my_proj' and atime < (NOW() - 365 DAYS)
```

Modify the filter to add a new condition:

```
project = 'my_proj' and atime < (NOW() - 365 DAYS) and TEMPERATURE <> 'ARCHIVE'
```

This will only apply the tag value to records that have not already been tagged with TEMPERATURE set to ARCHIVE, which will be a much smaller set assuming a low percentage of failed records.

If the tagging policy was used to set a different value for each record, you can check for the value being empty instead by adding <tag field> = '' to the original filter. Use this method for an extract from path policy where you cannot check for a specific tag value being set.

[Security issues

This topic describes security issues that you might encounter when you use IBM Spectrum Discover.

[Upgrading issues

This topic describes upgrading issues that you might encounter when you use IBM Spectrum Discover.

[How to recover Db2 Warehouse from an unrecoverable state during an upgrade

An upgrade might loop (while you wait for the UI backend) because DB2 Warehouse is in an unrecoverable state.

As the last step of an upgrade, a task waits for the UI backend to be ready by looping it up to 180 times:

```
TASK [common : Wait for pod app=spectrum-discover-ui-backend in namespace spectrum-discover] ***
Wednesday 01 July 2020 12:55:17 +0000 (0:00:00.246) 0:59:36.325 *****
FAILED - RETRYING: Wait for pod app=spectrum-discover-ui-backend in namespace spectrum-discover
(180 retries left).
FAILED - RETRYING: Wait for pod app=spectrum-discover-ui-backend in namespace spectrum-discover
(179 retries left)
...
FAILED - RETRYING: Wait for pod app=spectrum-discover-ui-backend in namespace spectrum-discover
(2 retries left).
FAILED - RETRYING: Wait for pod app=spectrum-discover-ui-backend in namespace spectrum-discover
(1 retries left).
FATAL: [x.x.x.x]

PLAY RECAP *****
x.x.x.x : ok=172 changed=59 unreachable=0 failed=1 skipped=74
rescued=0 ignored=0
```

After 180 retry attempts, if the upgrade fails you can check the DB2 Warehouse logs. The failed upgrade might indicate that DB2 Warehouse is in an unrecoverable state:

```
docker logs Db2wh
```

The log output might contain information similar to the following statements:

```
[
21.659412] start_dashDB_local.sh[70]: IBM Db2 Warehouse stack is NOT initialized yet.
23.840237] start_dashDB_local.sh[70]: cp: cannot stat '/etc/openldap/certs/': No such file
or directory
[
24.210802] start_dashDB_local.sh[70]: certutil: function failed: SEC_ERROR_LEGACY_DATABASE:
The certificate/key database is in an old, unsupported format.
[
24.285892] start_dashDB_local.sh[70]: cp: cannot stat '/etc/openldap/certs/rootCA.pem': No
such file or directory
[
27.321018] start_dashDB_local.sh[70]: configure_user_management completed successfully.
[
30.325604] start_dashDB_local.sh[70]: Creating IBM Db2 Warehouse instance and directories
[
38.339772] start_dashDB_local.sh[70]: Initialize IBM Db2 Warehouse instance to use up to 80%
of the memory assigned to container namespace
[
43.435741] start_dashDB_local.sh[70]: cp: cannot stat '/tmp/bluhelix.properties': No such
file or directory
[
43.477866] start_dashDB_local.sh[70]: mv: cannot stat '/tmp/dswebserver.properties': No such
file or directory
[
43.479807] start_dashDB_local.sh[70]: chown: cannot access '/opt/ibm/dsserver/Config/
dswebserver.properties': No such file or directory
[ 104.024301]
start_dashDB_local.sh[70]: !!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!
!!!!!!!!!!
[ 104.027795] start_dashDB_local.sh[70]: FATAL RUNTIME ERROR DETECTED
[ 104.031407] start_dashDB_local.sh[70]: REASON: SLAPD failed to start in 60 seconds
[ 104.034966]
```

```
start_dashDB_local.sh[70]: !!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!  
!!!!!!!!
```

Run the following commands in these steps to correct this problem:

1. Run the following command to restart DB2 Warehouse to recover:

```
docker stop Db2wh  
docker start Db2wh
```

2. Run the following command to monitor the DB2 Warehouse logs and confirm a successful start:

```
docker logs Db2wh --follow
```

3. After DB2 Warehouse starts successfully, run the following command and the IBM Spectrum Discover pods recover and enter a running state:

```
kubectl get pods -n spectrum-discover
```

]

Debugging a hung upgrade

You must debug an upgrade if a failed task causes that upgrade to hang or stop.

If a new Secure Shell (SSH) session displays the UPGRADE IN PROGRESS message but does not display any additional output, a failed task causes the upgrade to stop.

To identify the cause for the upgrade to stop, issue this command:

```
tail -n 200 /opt/ibm/metaocean/logs/upgrade.log
```

After you fix any issues identified in the upgrade log, reboot and then restart the upgrader. Use the **systemctl** command to restart the upgrader:

```
[root@ ~]# systemctl start upgrader
```

The upgrader output is not displayed in the console after you restart the upgrader. Either use the tail of the `/opt/ibm/metaocean/logs/upgrade.log` file or reconnect your SSH connection to automatically see the progress.

If the upgrader does not restart when you upgrade from a same-level version, run the upgrade command again:

```
[root@ metaocean]# ./upgrade
```

[Upgrader not logged in when you create metrics services

If the upgrader is not logged in when you create metrics services, perform the following actions.

The following error displays if the upgrader fails when you create metrics services during the "TASK [icp_post : Create metrics services]" step:

```
error: You must be logged in to the server (the server has asked for the client to provide  
credentials)
```

This error occurs because IBM Cloud Private is not running fully.

Note: This issue generally occurs only in low spec machines.

Use the following steps to recover:

1. Run the following command to ensure that the IBM Cloud Private is running fully:

```
sudo /usr/local/bin/kubectl --namespace=kube-system get pods --no-headers 2>&1 | egrep -v  
'Completed|Running|Succeeded'
```

Note: The recovery process is ready to start only when no output is displayed after running the command. Wait for IBM Cloud Private to be running fully to recover.

2. Run the following command to validate that your IBM Cloud Private login is working. Validate that your IBM Cloud Private login works: `sudo /etc/cron.hourly/icp_login.sh`
3. Start the upgrade by using the most recent restart: `sudo systemctl start upgrader.service`
4. Reconnect the Secure Shell (SSH) connection to automatically monitor the upgrade progress.

]

Accessibility features for IBM Spectrum Discover

Accessibility features help users who have a disability, such as restricted mobility or limited vision, to use information technology products successfully.

Accessibility features

The following list includes the major accessibility features in IBM Spectrum Discover:

- Keyboard-only operation
- Interfaces that are commonly used by screen readers
- Keys that are discernible by touch but do not activate just by touching them
- Industry-standard devices for ports and connectors
- The attachment of alternative input and output devices

IBM Knowledge Center, and its related publications, are accessibility-enabled. The accessibility features are described in [IBM Knowledge Center \(www.ibm.com/support/knowledgecenter\)](http://www.ibm.com/support/knowledgecenter).

Keyboard navigation

This product uses standard Microsoft Windows navigation keys.

IBM and accessibility

See the [IBM Human Ability and Accessibility Center \(www.ibm.com/able\)](http://www.ibm.com/able) for more information about the commitment that IBM has to accessibility.

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.

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:

© (your company name) (year).

Portions of this code are derived from IBM Corp.

Sample Programs. © Copyright IBM Corp. _enter the year or years_.

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

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 www.ibm.com/legal/copytrade.shtml](http://www.ibm.com/legal/copytrade.shtml).

Intel is a trademark of Intel Corporation or its subsidiaries in the United States and other countries.

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

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

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

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

Terms and conditions for product documentation

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

Applicability

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

Personal use

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

Commercial use

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

Rights

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

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

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

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

IBM Online Privacy Statement

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

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

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

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

Index

A

- accessibility features for IBM Spectrum Discover [113](#)
- Adding auto-tagging policies
 - Managing policies [17](#)
- Adding content search policy parameters
 - Using content search policies
 - Administering [18](#)
- Adding deep-inspection policies
 - Managing policies [18](#)
- Adding policies
 - Managing policies [15](#)
- Administering [89](#)
- Ansible Warnings
 - Troubleshooting [102](#)

B

- Backup and restore
 - Planning [67](#)
- bash shell
 - on container [2](#)

C

- container
 - bash shell [2](#)
- Creating a content search policy
 - Using content search policies [26](#)
- Creating an IBM Cloud Object Storage System connection
 - Managing LDAP and IBM Cloud Object Storage System connections [12](#)
- Creating applications for the IBM Spectrum Discover Application Catalog [65](#)
- Creating collections
 - Managing collections
 - Managing user accounts [9](#)
- Creating groups
 - Managing groups
 - Managing user accounts [7](#)
- Creating user accounts
 - Managing user accounts [6](#)

D

- Deleting policies
 - Managing policies [23](#)

I

- IBM Spectrum Discover (Backup and restore) [67](#)
- IBM Spectrum Discover information units [xi](#)
- Identifying the required regex expressions
 - Using content search policies [25](#)
- Initial setup configuration
 - Backup and restore [67](#)

- Installation issues
 - Troubleshooting [103](#)

K

- keystone
 - pod name [2](#)

M

- Managing applications
 - Administering [61](#)
 - Creating applications for the IBM Spectrum Discover Application Catalog [65](#)
 - Using the IBM Spectrum Discover Application Catalog [63](#)
- Managing collections
 - Managing user accounts [8](#)
- Managing groups
 - Managing user access [6](#)
- managing LDAP and IBM Cloud Object Storage connections
 - Managing user accounts [10](#)
- Managing [metadata] policies
 - Administration [15](#)
- Managing user access
 - Administering [1](#)
- Managing user accounts
 - Administering [5](#)
- Modifying policies
 - Managing policies [22](#)
- Monitoring data sources
 - Viewing data source status [75](#)
- Multinode network settings get stuck while checking whether the Docker is running [104](#)

N

- network configuration
 - updating [87](#), [97](#)
- Network settings change fails when you run mmconfigappliance
 - Troubleshooting [106](#)
- Network settings change hangs while uninstalling IBM Cloud Private
 - Troubleshooting [106](#)

P

- pod name
 - keystone [2](#)
- policy
 - autotag [78](#)
- Preparations for disaster recovery
 - Administering [93](#)

R

- recommended to move
 - autotag policy
 - creating [78](#)
 - definition [78](#)
- Reintegrating a failed data node into an IBM Db2 Warehouse MPP cluster
 - Highly available for MPP deployment [74](#)
- resetting
 - sdadmin password [2](#)
- Running a backup
 - Backup and restore [68](#)
- Running a restore
 - Backup and restore [69](#)
- Running an automated backup
 - Backup and restore [69](#)
- Running an restore [69](#)
- Running disaster recovery
 - Administering [93](#)
- Running policies
 - Managing policies [19](#)

S

- sdadmin password
 - resetting [2](#)

T

- temperature tag
 - example [78](#)
- Troubleshooting
 - Best practices [97](#)
 - Configuration issues [98](#)
 - Data issues [98](#)
 - Debugging a hung upgrade [110](#)
 - Error diagnosis issues [102](#)
 - How to recover a system after a yum update [98](#)
 - How to recover DB2 Warehouse from an unrecoverable state during an upgrade [109](#)
 - Networking issues [103](#)
 - Performance issues [107](#)
 - Pod stuck in terminating state [107](#)
 - Policy issues [108](#)
 - Scanning [101](#)
 - Security issues [109](#)
 - Upgrader not logged in when creating metrics services [110](#)

U

- updating
 - network configuration [87](#), [97](#)
- Upgrading
 - Troubleshooting [109](#)
- Using a third-party data movement application to move or copy data [89](#)
- Using content search policies
 - Managing policies [25](#)
- Using the FFDC script
 - Administering [85](#)
- Using the IBM Spectrum Discover Application Catalog [63](#)

V

- Viewing content search application logs
 - Using content search policies [27](#)
- Viewing data source status [75](#)
- Viewing live reports might list incorrect results
 - Troubleshooting [101](#)
- Viewing policies
 - Managing policies [19](#)
- Viewing policy log files
 - Viewing policies [21](#)
 - Viewing policiesManaging policies [21](#)
- Viewing scaleless data mover application logs [31](#)
- viewing scaleless data mover error logs [31](#)



Product Number: 5737-I32
5737-SG1

SC27-9602-08

