IBM Spectrum Discover

Administration Guide

Note

Before using this information and the product it supports, read the information in “Notices” on page 85.

Edition notice

This edition applies to version 2 release 0 modification 0 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 xii. 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.

Contents

Figures

Tables

About this information

IBM Spectrum® Discover is 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 using built-in or custom agents.

Which IBM Spectrum Discover information unit provides the information you need?

The IBM Spectrum Discover library consists of the information units listed in Table 1 on page xi.

Table 1. IBM Spectrum Discover library 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: • 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: • 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.

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, which use policies that determine the metadata that is available to view.

User and group access can be authenticated by IBM Spectrum Discover, an LDAP server, or the IBM Cloud Object Storage System. The administrator role can manage the user access functions.

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, the user will have the privileges of the Data Admin role.

The following roles are available:

Admin
: This role can create users, groups, and collections and also manage LDAP connections. 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
: 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.

Data User
: This role can access metadata that is collected by IBM Spectrum Discover, but metadata access can be restricted by policies in 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

The default login for the IBM Spectrum Discover 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 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 -n authrbac 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 -n authrbac 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 or 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 using the following commands. If the username is not reset correctly, a "401 unauthorized error" is returned.

```
source keystone_sdadminrc
openstack user list
```

Managing user accounts

The administrator can create and manage local user accounts, which are authenticated by IBM Spectrum Discover, and 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 edit 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 Local User**. For more information, see “Creating user accounts” on page 3.

Editing a user account

You can edit account information for a local user. You cannot edit or delete either LDAP or IBM Cloud Object Storage user accounts, but you can assign these users to roles and collections.

To edit a local user account, click the icon in **Edit** column for a local user that is listed on the **Users** tab. Use the **Edit User** window to edit the local user account.

To edit an LDAP or IBM Cloud Object Storage user account, click the icon in **Edit** column for an LDAP or IBM Cloud Object Storage user that is listed on the **Users** tab. Use the **Edit LDAP User** window to assign these users to roles and collections.

Deleting a local user account

To delete a local user account, click the icon in **Delete** column for a local user that is listed on the **Users** tab.

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.

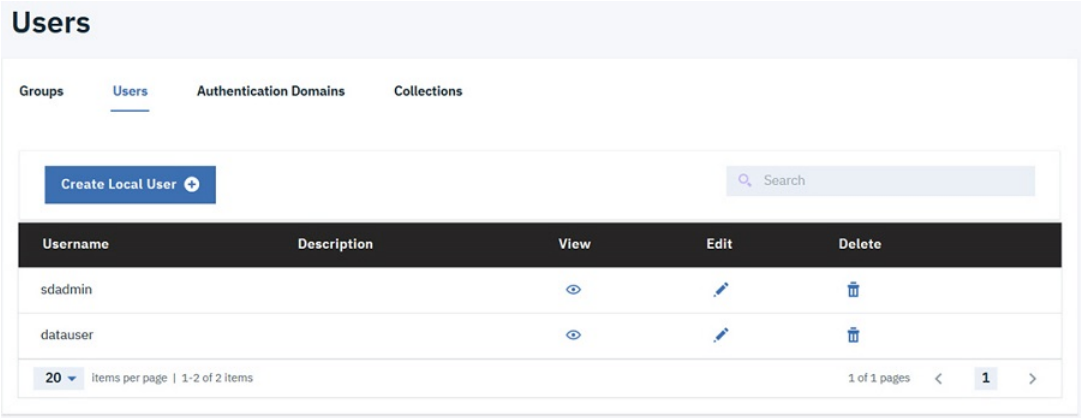
User Name
The user name for the account.

Role
The role that is assigned to the user.

Domain
The domain that provides authentication for the user. For authentication by IBM Spectrum Discover, the domain name is **Local**.

Collections
The collections that are assigned to the user.

Description
The description of the user.



Creating user accounts

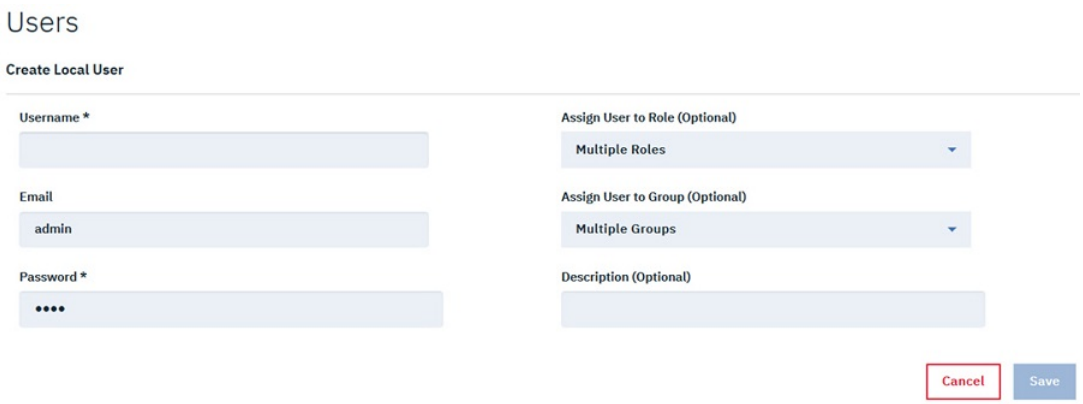
The administrator can create local user accounts, which are authenticated by IBM Spectrum Discover, and assign roles to users.

About this task

Use the **Users** tab on the **Management** page to create a local account. You can assign roles and passwords to users and also add a user to a group.

Procedure

1. From the **Management** page, click **Create Local User** to open the **Users** window.



2. Enter a **User Name** and **Email** address for the user.

3. Enter a **Password** for the user.
4. 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 role must also be associated with at least one collection.
5. 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. 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, and also assign local or 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 an LDAP server or the local domain. You can also use the tab to create, edit, or delete local groups. You cannot edit or delete LDAP or IBM Cloud Object Storage System groups, but you can assign these groups to roles and collections.

Creating a local group

To create a local group, click **Create Local Groups**. For more information, see “Creating groups” on page 5.

Editing a group

To edit a local group, click the icon in **Edit** column for a local group that is listed on the **Groups** tab. Use the **Edit Group** window to edit the local group.

To edit an LDAP group, click the icon in **Edit** column for an LDAP group that is listed on the **Groups** tab. Use the **Edit LDAP Group** window to assign an LDAP group to a role and collections.

Deleting a local group

To delete a local group, click the icon in **Delete** column for a local group that is listed on the **Groups** tab.

Group information

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

Group Name

The name for the group.

Role

The role that is assigned to the group.

Users

The number of users in the group.

Domain

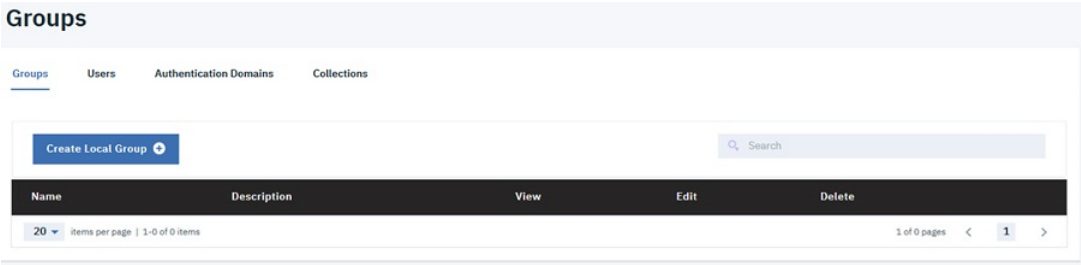
The domain that provides authentication for the group. For authentication by IBM Spectrum Discover, the domain name is **Local**.

Collections

The number of collections that are assigned to the group.

Description

The description of the group.



Creating groups

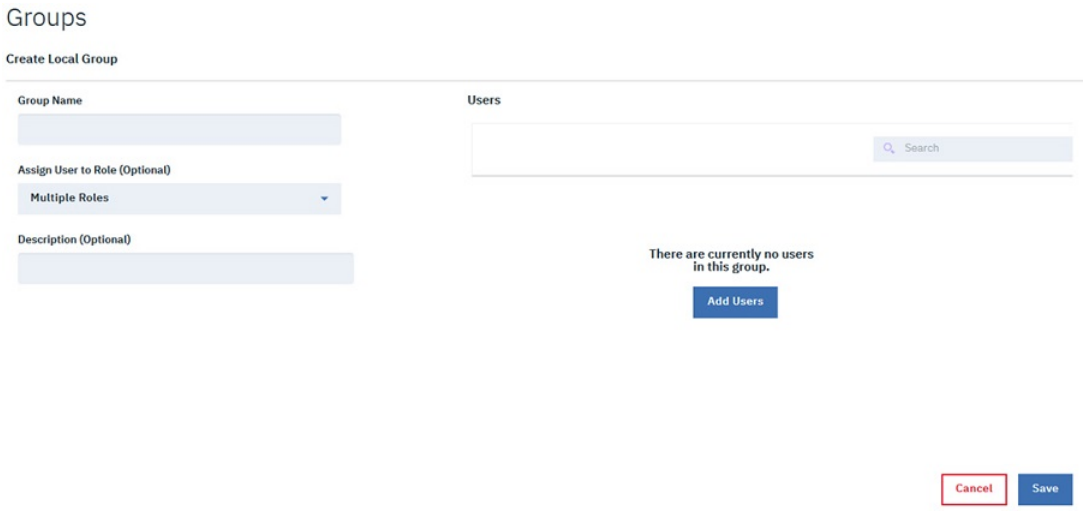
The administrator can create local groups, that are authenticated by IBM Spectrum Discover, and assign users to the groups.

About this task

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

Procedure

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



2. Enter a **Group Name**.
3. Optional. Use the **Assign User to Role** list to assign one or more roles to the group. For more information about roles, see “Roles” on page 1.
4. Click **Add Users** to open the **Add Users** window and add one or more local users to the collection.

   Enter a user name 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.

   User Name
   The user name or email address of the member.

   Domain
   The LDAP domain that provides authentication for the member.

   Role
   The role that is assigned to the member.

5. Optional. Use the **Assign Group to Collection** list to assign one or more collections to the group. You can also use the **Collections** tab to assign collections to groups. For more information about collections, see "Managing collections" on page 6.
6. Optional. Enter a **Description** for the group.
7. Click **Save**.

Managing collections

The administrator can create and manage collections, which 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 would not be able to see the metadata.

The administrator can create collections, assign users and groups to collections, and create a policy to associate specific metadata 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. 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.

[The CollectionAdmin can list any type of tag, create and modify Characteristic tags, but cannot create, modify, delete Open and Restricted tags. This is the same permissions as the DataUser role.]

[The CollectionAdmin can create, update and delete policies for the collections they administer.]

[The CollectionAdmin can view, update and delete policies of DataUsers for the collections they administer. They cannot delete a policy if it has a collection that they do not administer.]

[The CollectionAdmin can add Data User roles to users for collection(s) they administer. This gives DataUsers access to a particular collection, meaning access to the records 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 7.

Editing a collection
To edit a collection, click the icon in **Edit** column for a collection that is listed on the **Collections** page. Use the **Edit Collection** window to edit the collection.

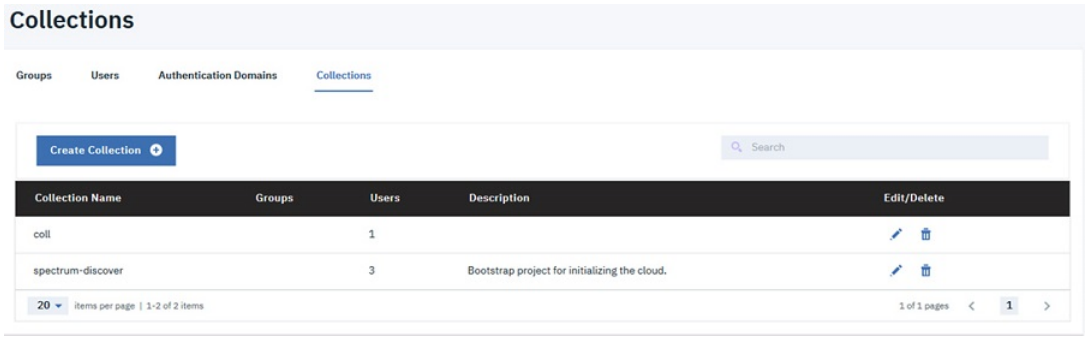
Deleting a collection
To delete a collection, click the icon in **Delete** column for a collection that is listed on the **Collections** page.

Collections information

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

Collection Name
The name of the collection.

Description
The description of the collection.



Creating collections

The administrator can create collections, assign users and groups to collections, and associate a collection with a policy. A CollectionAdmin cannot create a collection, but can assign users and groups to collections they administer. A user with the data admin role can associate metadata records with a collection by using an auto-tag policy.

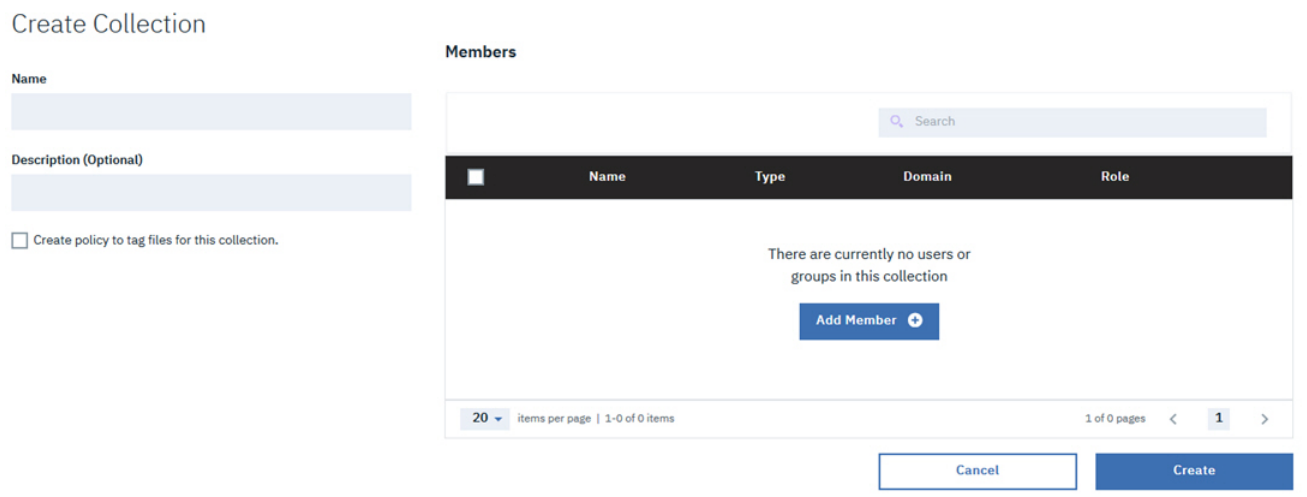
About this task

Collections are logical groups of records that can have access that 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 **Management** page to create collections.

Procedure

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



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

   Enter a user name, 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.

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

   Name
   The user name, group name, or email address of the member.

Type
The account type: user or group.

Domain
The LDAP domain that provides authentication for the member.

Role
The role 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 about defining policies, see Chapter 2, "Managing policies," on page 13.
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 an LDAP connection, click **Add Domain Connection**.

For steps to create a connection to an LDAP server, see "Creating an LDAP connection" on page 9.

For steps to create a connection to an IBM Cloud Object Storage System server, see "Creating an IBM Cloud Object Storage connection" on page 11.

Testing a connection
To test a connection, click **Test Connection**.

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

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

Connection details

The following connections details are available on the **Authentication Domains** tab:

LDAP Connection details

Directory name
The name of the directory that provides authentication.

Type
The directory type, which is LDAP.

Port
The LDAP server port that provides the connection.

Group Tree DN
The group tree distinguished name that uniquely identifies an entry in the directory.

User Object Class
The object class that is supported by the LDAP server.

LDAP Authentication details

Bind DN
The distinguished name that identifies entries in a directory.

Bind Password
The password for the **Bind DN**.

Base DN
The distinguished name that is the base of entry searches in the directory.

User Tree DN
The user tree name that uniquely identifies the users in the directory.

User Name Attribute
The attribute that is used to search for user names.

IBMCOS Connection details

Directory Name
The name of the directory that provides authentication.

Type
The directory type, which is IBMCOS.

IBMCOS Authentication details

User name
The IBM Cloud Object Storage System security administrator name.

Password
The IBM Cloud Object Storage System security administrator password.

Creating an LDAP connection

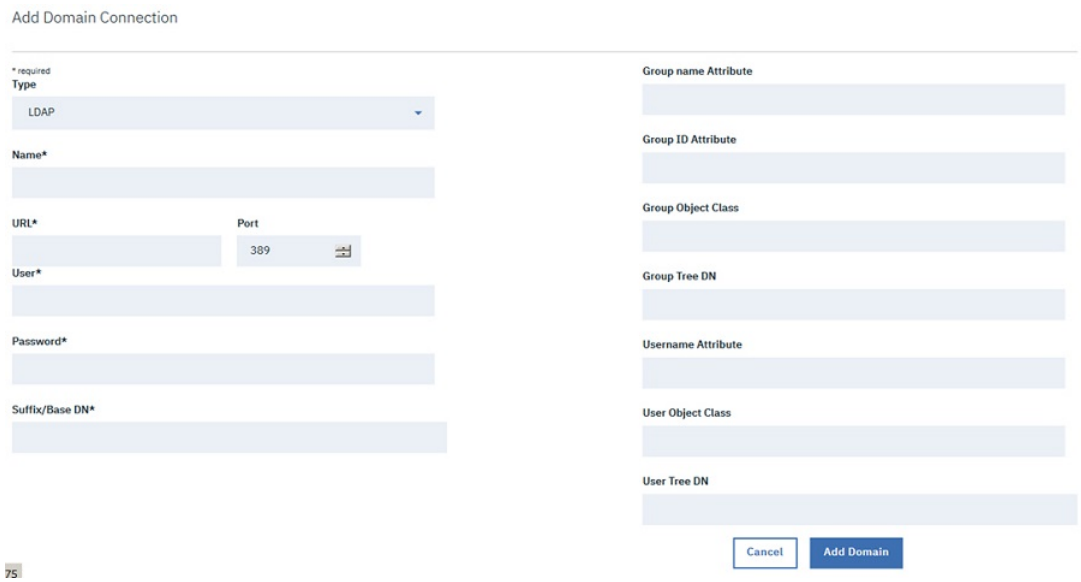
The administrator can create a connection to an LDAP server that provides authentication for IBM Spectrum Discover users.

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**.



3. Enter the following information for the LDAP directory:

   Directory Name
   The name of the LDAP directory that provides authentication.

   Host
   The host name of the LDAP server.

   Port
   The port number on the LDAP server.

   Bind DN
   Enter the bind distinguished name in the following format:

   ```
   cn=relative_distinguished_name dc=domaain_component
   ```

   For example,

   ```
   cn=John Doe,dc=example,dc=com
   ```

   Bind password
   Enter the bind password.

   Base DN
   Enter the base distinguished name.

4. Optional. Use the **LDAP Mapping** fields to allow access to users or groups within a specified range of LDAP members.

   LDAP server
   The LDAP server that manages the members in the range.

   Range
   Specify the range to include.

   User DN
   The distinguished name for users in the range.

   Group DN
   The distinguished name for groups in the range.

5. 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.

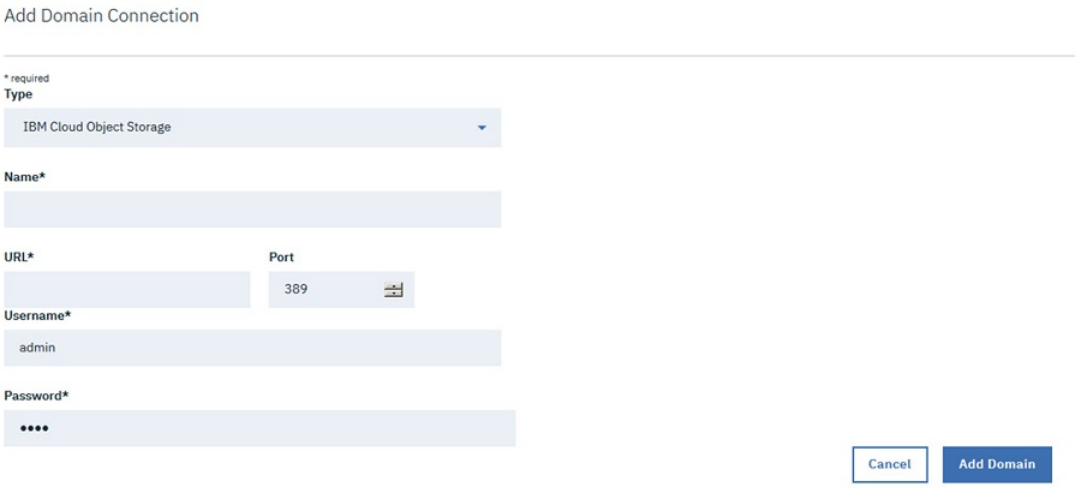
About this task

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.

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**.



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

   Directory Name
   The name of the directory that provides authentication.

   Domain Name or IP Address
   The domain name or IP address of the IBM Cloud Object Storage server.

   User name
   The IBM Cloud Object Storage security administrator name.

   Password
   The IBM Cloud Object Storage security administrator password.

4. Click **Connect**.

Chapter 2. Managing policies

Policies might be used to periodically automatically tag a set of documents. In addition, policies might be used to send sets of documents to be deep-inspected by a registered agent.

Roles and permissions

Data User
Create, modify, and view policies.

Note: A user with the role *Data User* can only create or modify a policy when **Schedule** is set to **Now**.

A user with the role Data User cannot use a COLLECTION tag when creating or modifying policies.

Data Administrator
Create, modify, view, and delete policies.

Security Administrators
Cannot create, modify, view, and delete policies.

Service User
Cannot create, modify, view, and delete policies.

Adding auto-tagging policies

About this task
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 filesystem or owner ID.

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

- 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 Rodriguez, the filter might be written as: **`owner='Rodriguez'`**
- To enact a policy on all PDF files in cluster cl1 and datasource fs1, the filter could be written as: **`cluster='cl1' and datasource='fs1' and filetype='pdf'`**

Procedure

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

   `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.
2. Enter a name for the policy in the **Name** box.
3. Select **AUTOTAG** from the **Policy Type** menu.
4. Enter a filter for the policy into the **Filter** box.
5. Associate one or more tags with this policy by using one of the following methods:
   a) Click the **+ Add Tag** control.
   b) Select a tag name from the **Field** menu.
   The Fields might be specified by going to **Metadata** > **Tags**.
   c) Add a value for the tag in the **Tag** box

Note: If you do not know the valid values for a tag, navigate 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.

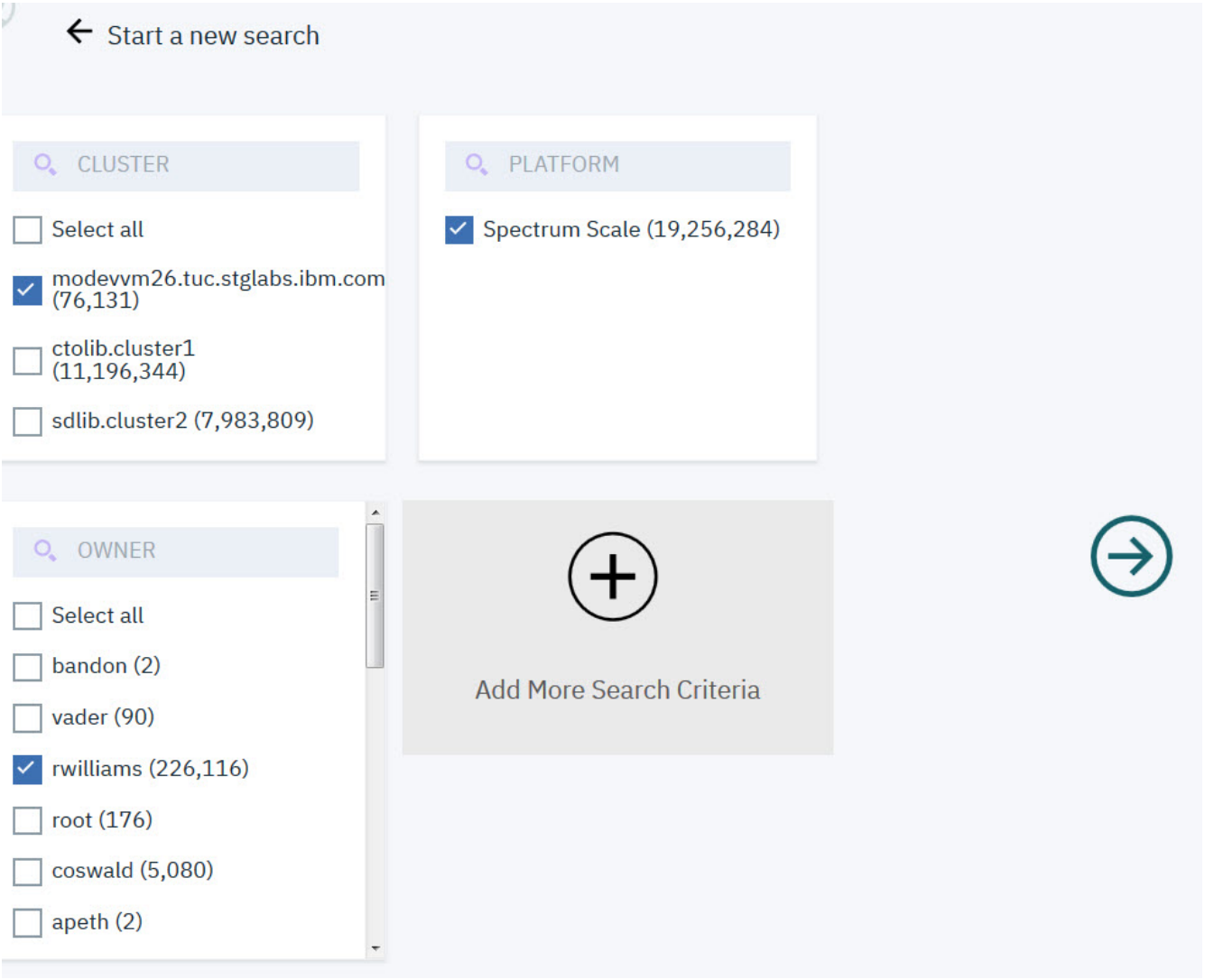


Figure 1. Tag values

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

a) Select the **Extract tag from path** checkbox.

b) Select a tag from the **Field** menu.

c) 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 17

6. Select a **Schedule** to apply the policy. Policy schedule times are entered in UTC.

Now
The policy is applied immediately, unless the policy's **Status** is **`Inactive`**

Daily

a. Enter the time of day to apply the policy by clicking the hour and minute from the widget that displays.

The policy is applied every day at the time specified.

Weekly

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 will be applied once a week on the day and time specified.

Monthly

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 once a month on the day and time specified.

7. Click **Submit Policy** to save the new policy.
   The new policy displays in the list of policies on the **Policies** tab.

Adding deep-inspection policies

About this task

You can enrich metadata through content inspection of source data. For example, you can extract patient names from medical records and index so you can search for files pertaining to patients by name. Deep-inspection policies can send lists of files to an action agent, which can examine the contents of files and return the values that it finds paired with defined tag keys.

A policy can contain a *filter* which is a search query that finds candidates to apply a policy. A filter uses the same syntax as the **where** clause in an SQL query. The filter must be constructed using standard SQL syntax. For example:

- To enact a policy on all files not accessed in one year, the filter could be written as: **`atime < (NOW() - 365 DAYS)`**
- To enact a policy on all files owned by Rodriguez, the filter could be written as: **`owner=Rodriguez`**
- To enact a policy on all PDF files in cluster cl1 and [datasource] fs1, the filter could be written as: **`[cluster='cl1' and datasource='fs1' and filetype='pdf']`**

Procedure

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

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

   `Inactive`
   An *Inactive* policy is will not run when its scheduling event is reached, including the **Now** event.
2. Enter a name for the policy in the **Name** box.
3. Select **`DEEP-INSPECT`** from the **Policy Type** menu.
4. Enter a filter for the policy into the **Filter** box.
5. Select an **Agent** from the menu.
6. Click **+ Add parameter** to assign a value to a parameter.
7. Select a **Parameter** from the menu.
8. Enter a **Value** for the parameter.
9. You can add more parameters by clicking the **+ Add parameter** control.
10. You can delete a parameter by clicking the **Delete** "minus" icon next to the parameter's **Value**.

11. Specify a **Schedule** to apply the policy. [Policy schedule times are entered in UTC.]

    Now

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

    Daily

    a. Enter the time of day to apply the policy by clicking the hour and minute from the widget that displays.

    The policy will be applied every day at the time specified.

    Weekly

    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 will be applied once a week on the day and time specified.

    Monthly

    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 will be applied once a month on the day and time specified.

12. Click **Submit Policy** to save the new policy.

    The new policy displays in the list of policies on the **Policies** tab.

Deleting policies

About this task

A policy can be deleted from the table on the Policies page. You cannot delete a policy that is currently running. A user with the role *data user* cannot delete a policy.

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 grayed-out, 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. 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 will not show up in the system metadata and the storage systems themselves do not provide management and reporting capabilities based on those organizational concepts. Custom tags allow you to store additional information and manage, report, and search for data 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
List any type of tag. 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 needed to describe or classify the object. Can contain very long descriptive values. Size limit is 4 KB

Creating 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.

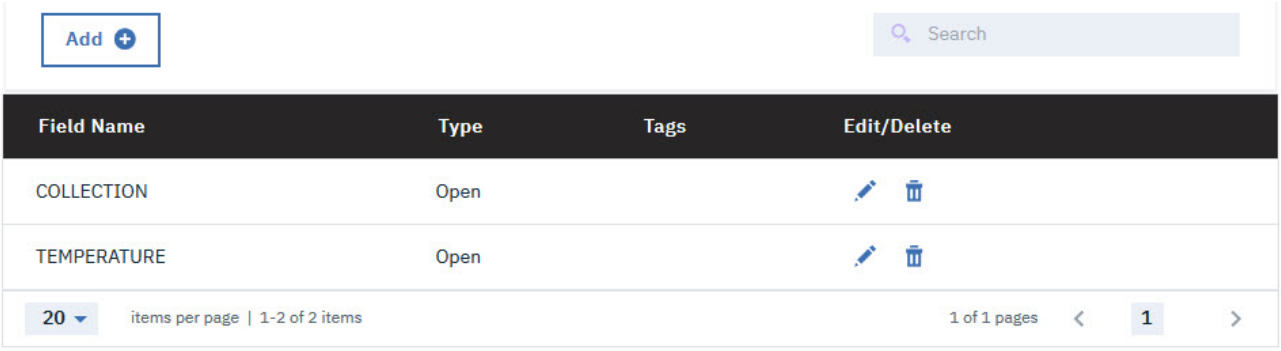


Figure 2. Tags table

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

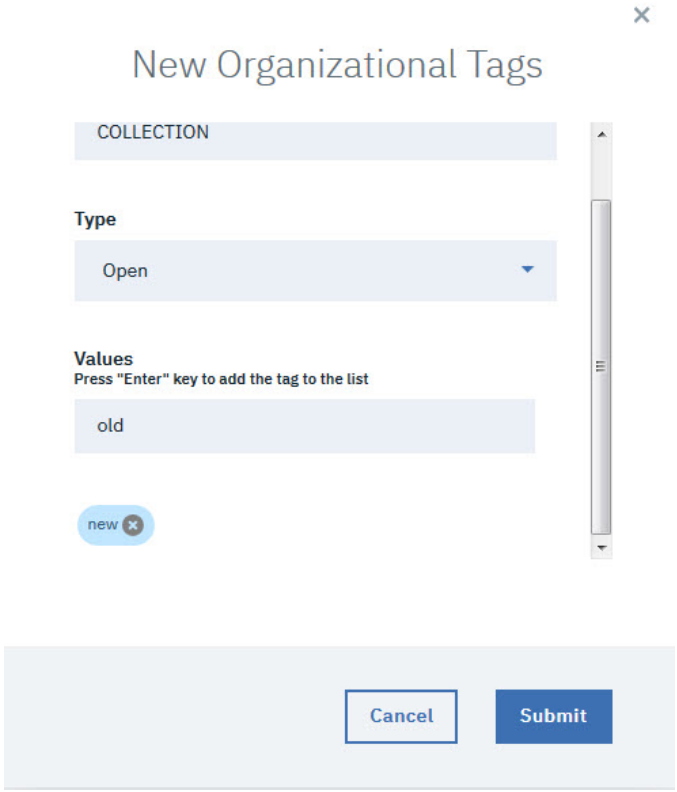


Figure 3. New Organizational Tags

4. Select one of the following 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.

5. 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.

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

Chapter 4. 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 defined within Spectrum discover. For more information, see Creating a CONTENTSEARCH policy.
- Create a policy using tags with known values. A policy is automatically run against all data that meet criteria specified in a filter. For more information about creating and using policies, see Searching
- Search your data 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.

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.

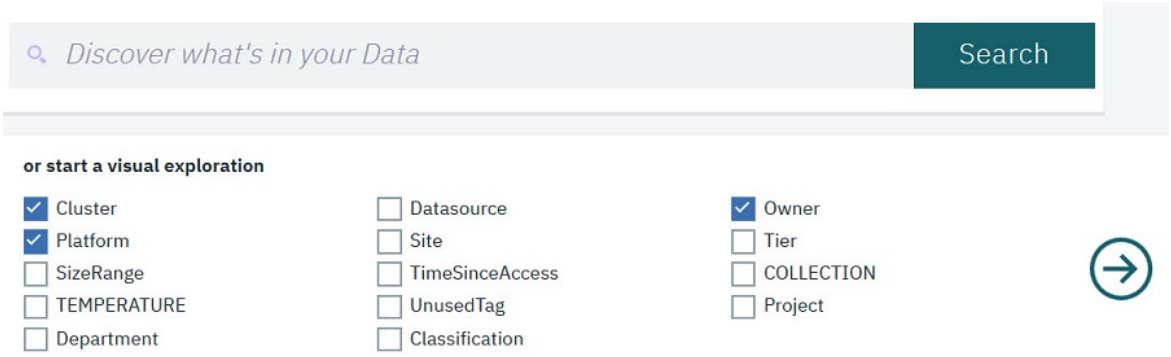


Figure 4. Start a visual exploration

     b. Check one or more boxes in the list of groups, policies, and tags. The figure below 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.

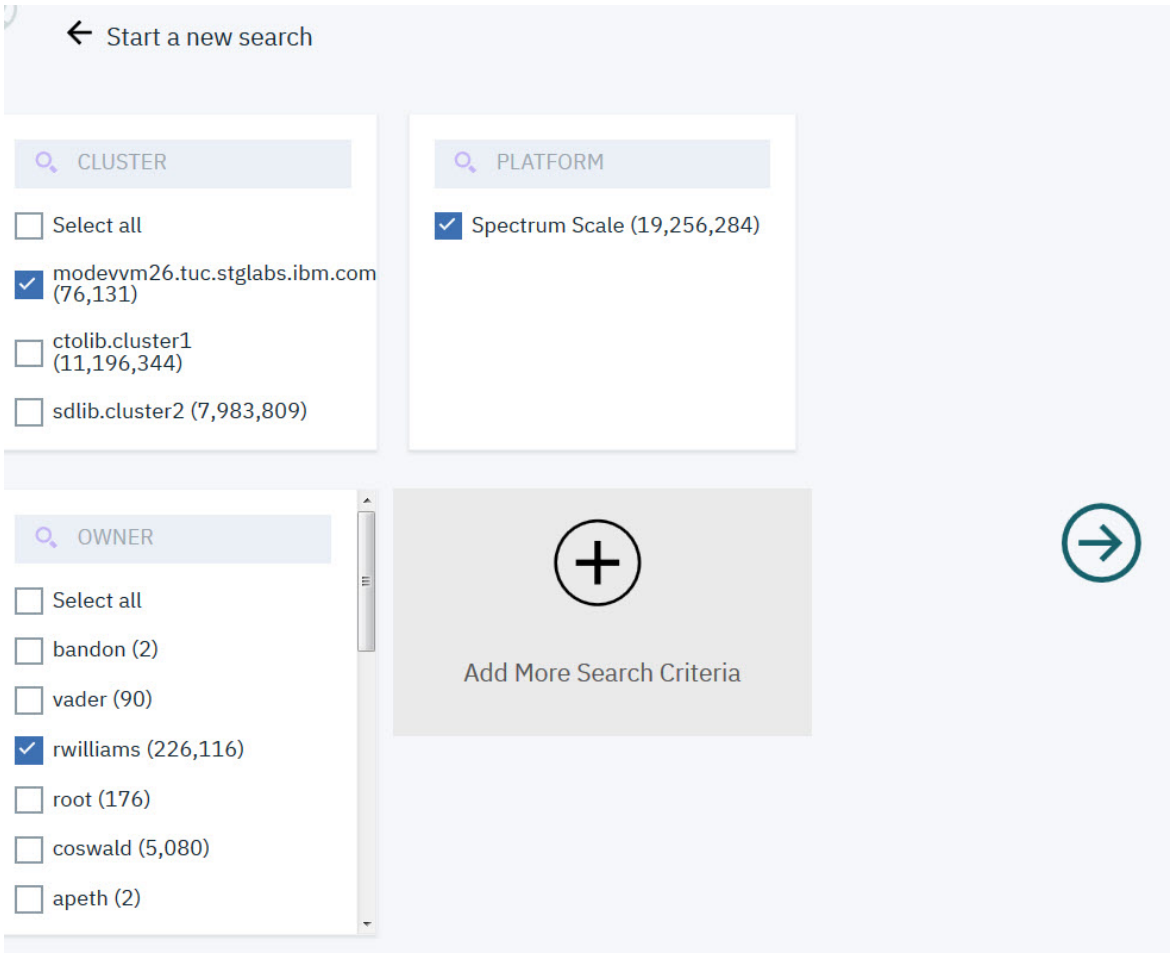


Figure 5. Tag values

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.

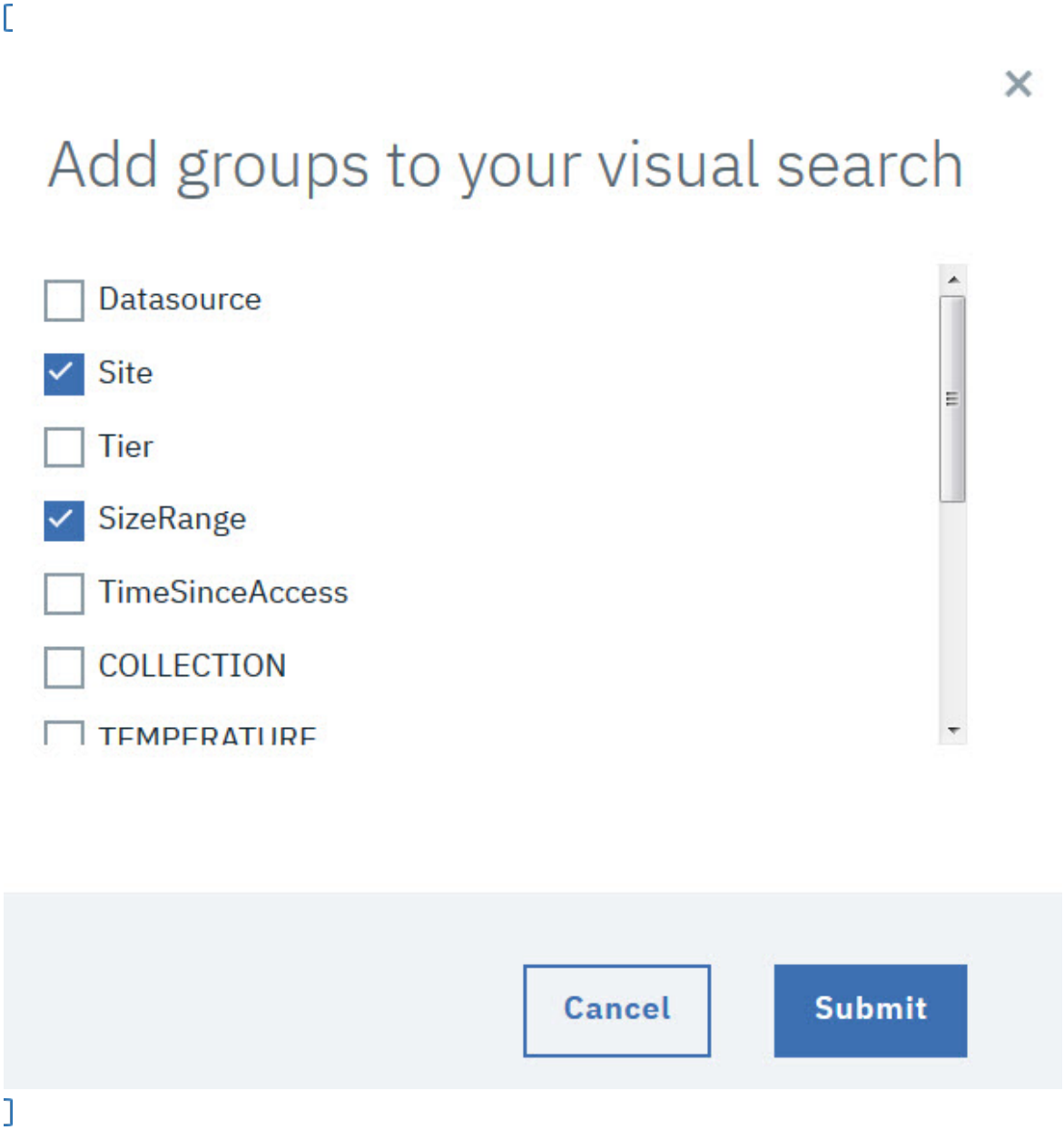


Figure 6. Search - add groups

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.

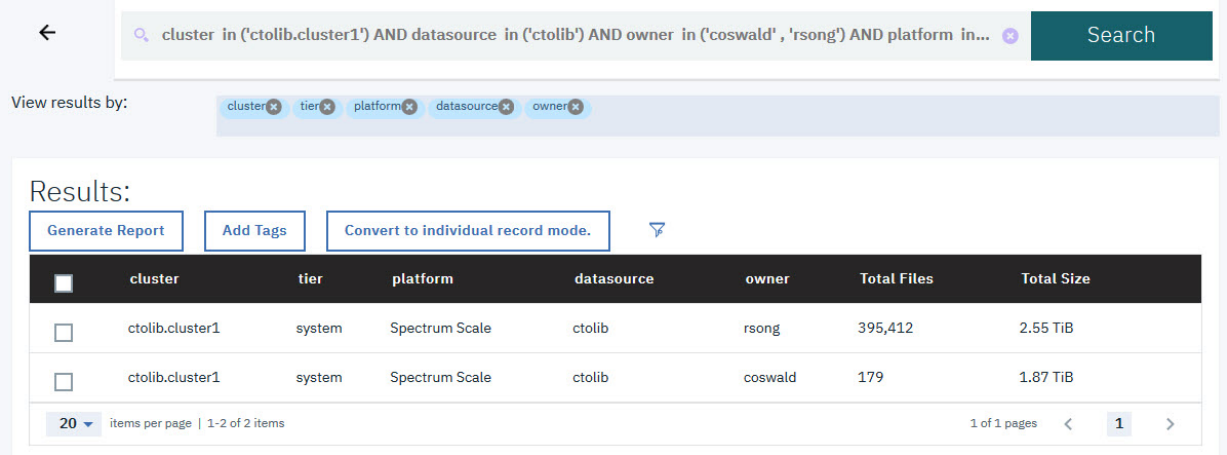


Figure 7. Search Results

- 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 essentially groups the results by applying a new search criteria to the original results.

  Note: There might be mismatched results when doing an initial search, then 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 a different number of results is returned when you do the search, use the **Search** box to re-enter 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 has been removed.

3. Filter the search results, if required:

   a. Click the **Filters** icon. The filters display in the panel to the right of the **Results**.

   b. Click one or more filters to expand it and select or input values.

   c. Click **Apply** and the filtered results display in the table.

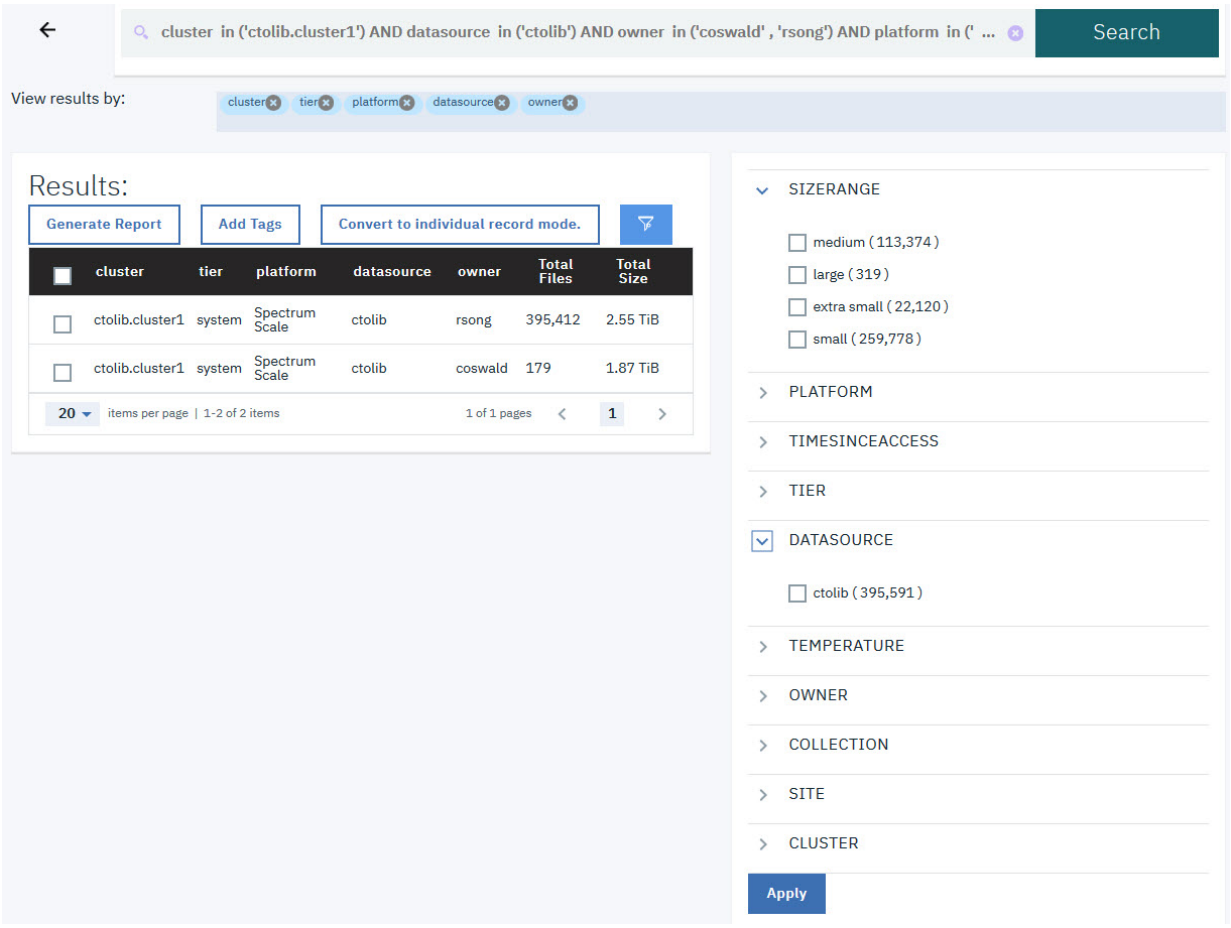


Figure 8. Search Results Filters

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

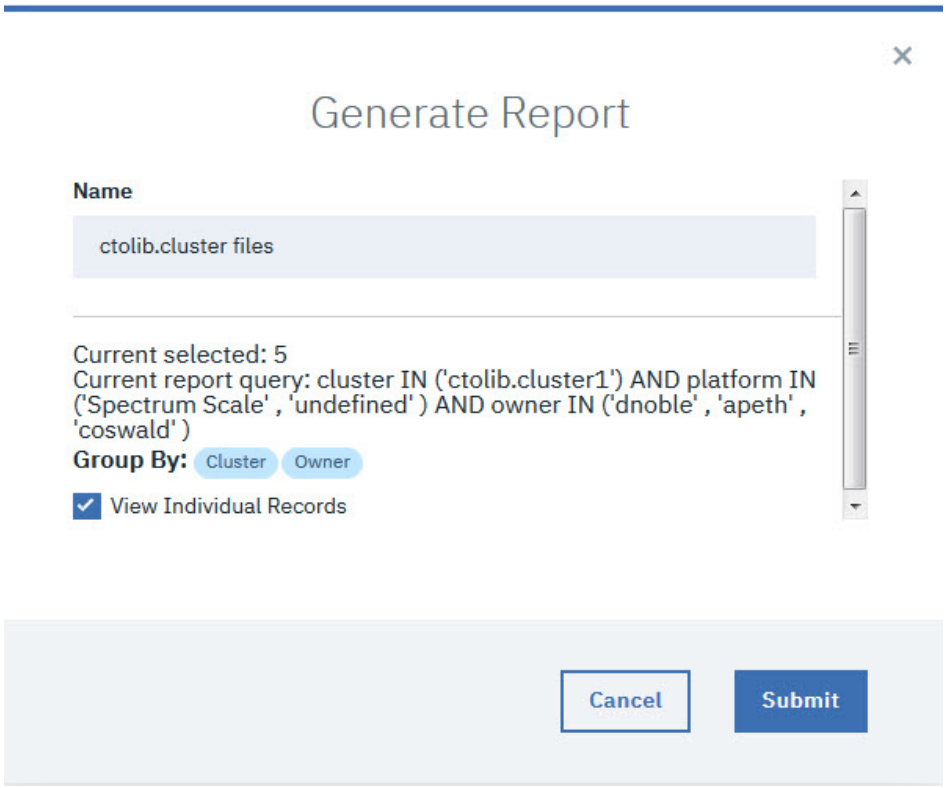


Figure 9. Generate Report

a. Enter a name for the report in the **Name** box.

b. Click the **View Individual Records** box if you want 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 do the following:

a. Select the checkbox of the records you want to add tags to.

b. Click **Add Tags**.

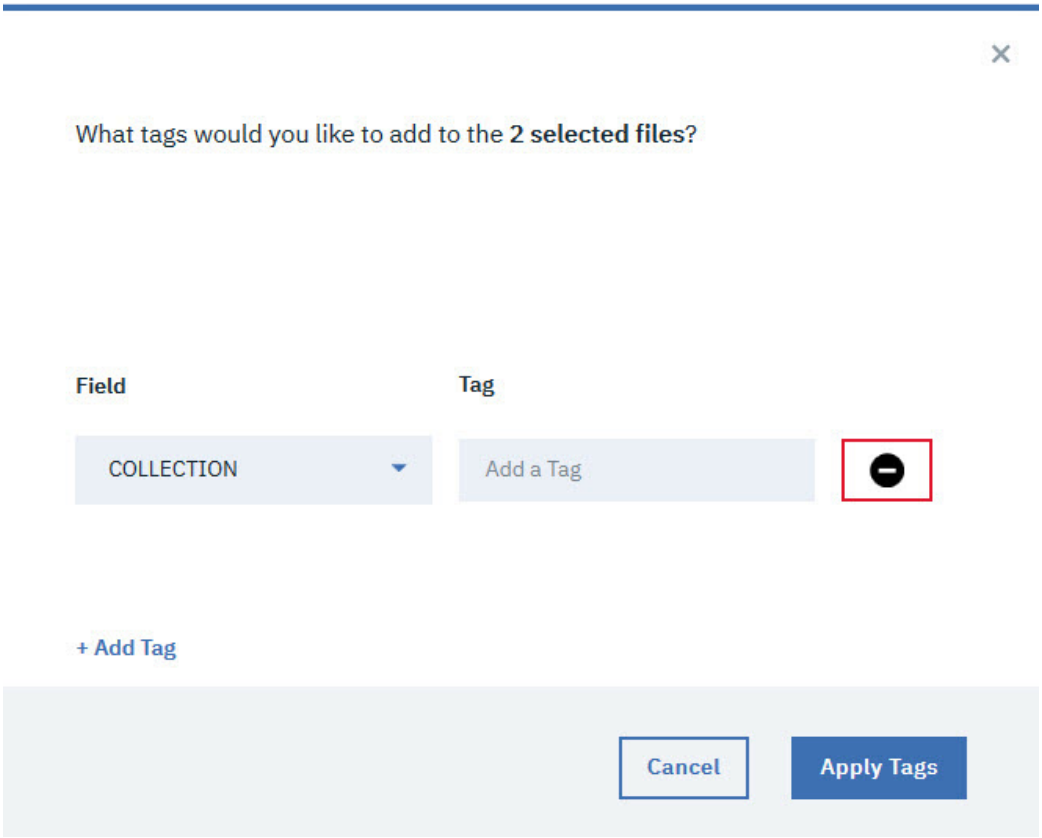


Figure 10. Add tags

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 have finished 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**.

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.

Fileset
The fileset to which the record belongs for IBM Spectrum Scale. This term applies only to IBM Spectrum Scale.

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.

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

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).

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.

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
For use when you do TIMESTAMP/DATE comparisons.

Wildcards

You can also use a wildcard.

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

Examples of search filters

This section provides a list of examples for search filters.

Note: You must wrap string values in single quotes 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.

Atime < (NOW() - 180 DAYS)
All files that have not been 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’.

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 can be customized in the **Headers** column of the **Advanced Search Options**.

Figure 11 on page 30 shows an example of a search by file type and data source.

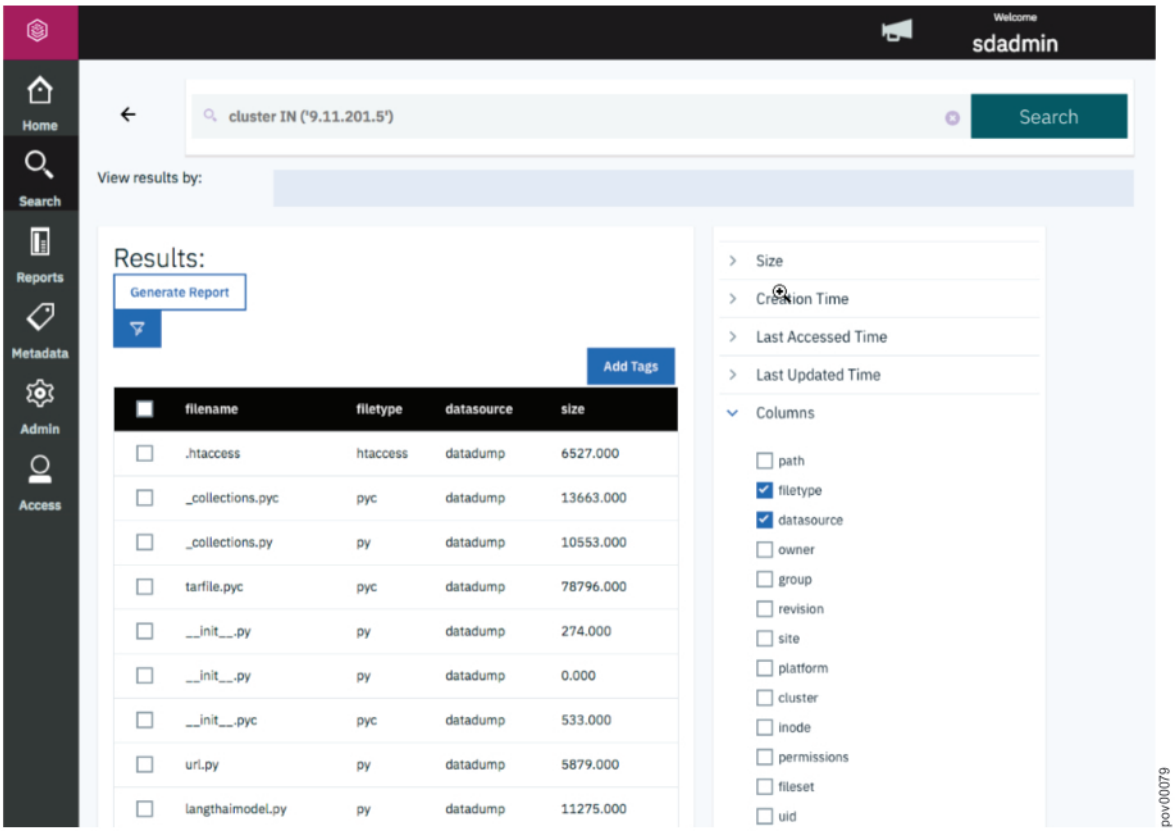


Figure 11. Example to generate a report sorted by filetype and datasource

Figure 12 on page 31 shows an example of the search results for timesinceaccess and sizerange.

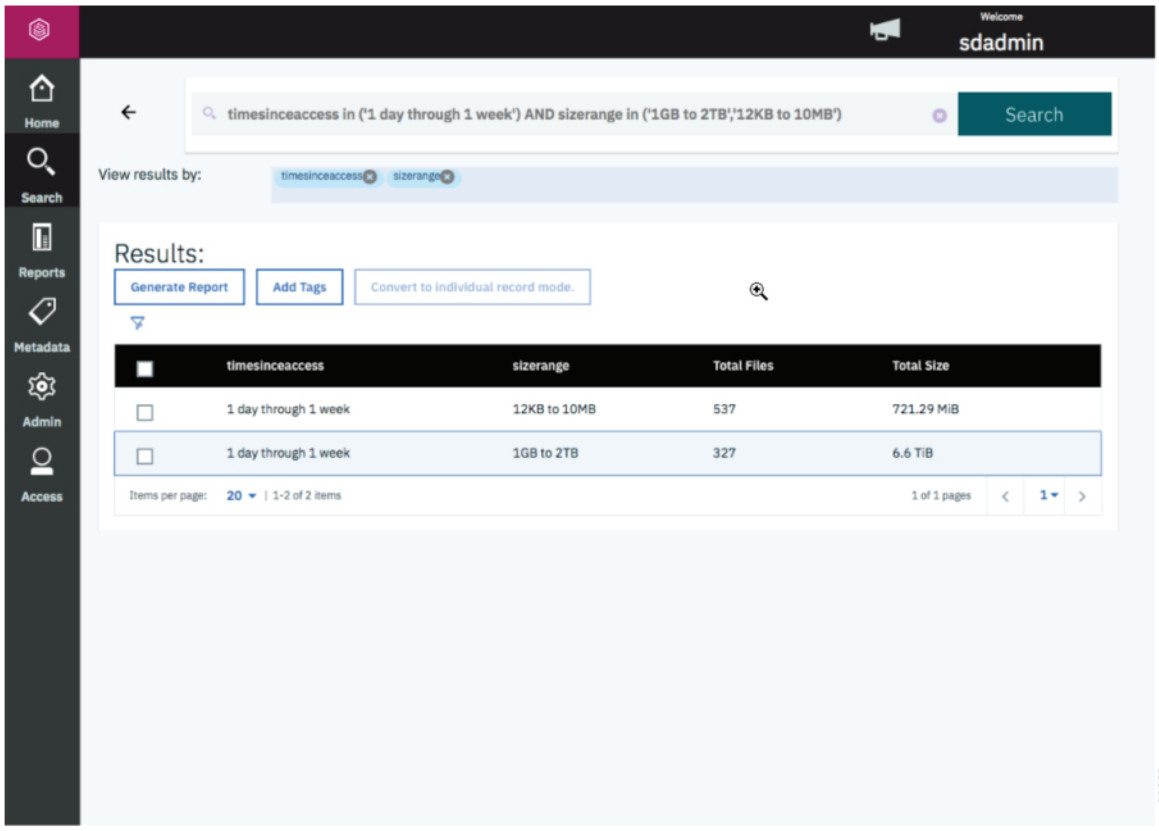


Figure 12. Example of a search sorted by timesinceaccess and sizerange

]

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 5. Managing action agents

An action agent is a program that interfaces with IBM Spectrum Discover and has access to the source storage. There are many use cases for action agents, including data content inspection for enriching metadata, data movement/migration, data scrubbing/sanitation, and more. Data is identified by IBM Spectrum Discover by policy filter and passed to the action agent as pointers through a messaging queue. Then, the action agent 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 (de-register), and view the action agents.

Data User
View the action agents created by a Data Administrator.

Security Administrator
Cannot create, modify, view, or delete any action agent.

Management

Agents might be viewed and deleted by navigating to **Metadata** > **Agents**. You can define an agent when you are creating a new **DEEP-INSPECT** policy. In addition, you can add **Parameters** for an agent during the process of creating a **DEEP-INSPECT** policy. See “Adding deep-inspection policies” on page 15 for more information.

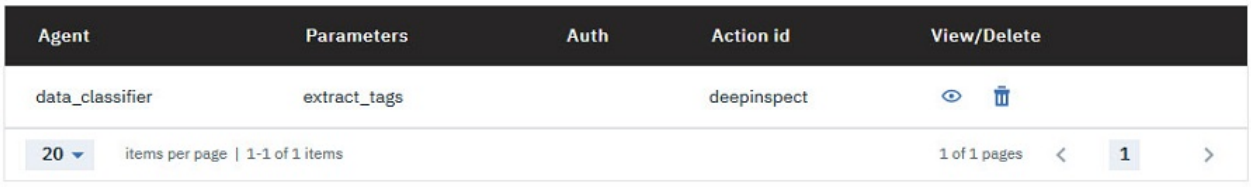


Figure 13. Agents table

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

Agent
The name of the agent.

Parameters
The parameters that were assigned to the agent when the policy as created.

Action ID
AUTOTAG or **deepinspect** - the policy type that the agent is assigned to.

View/Delete
Use the **View** "eye" icon to view the contents of the agent:

- Agent
- Action ID
- Params

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

For more information see the *Agent Registration REST API Guide*

Chapter 6. Backup and restore

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

The scripts used to backup and restore databases and file systems are located in the **/opt/ibm/metaocean/backup-restore** directory, and must be run as root user (Example: **sudo python /opt/ibm/metaocean/backup-restore/backup.py**).

It is a good practice to back up your system at least once a week. 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 12:00AM, however you can configure the backup frequency by running the **automatedBackup.py** script following the initial setup.

Remember: If any files or a database become corrupted, run the **restore.py** script to recover your file system and database back to the date of your last successful backup.

Initial setup 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 COS vault to store backup **tar** files.
   d) Enter the username for COS account configured with read-write access to storage vault.
   e) Enter the username for COS 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 installed and properly configured. For more information on IBM Spectrum Protect, see the IBM Knowledge Center.
   - External FTP server ("ftp")
   a) Enter the SFTP server IP or hostname
   b) Enter the username 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 between 1 and 999. Once the selected number of backups is exceeded, the oldest backup **tar** file will be deleted.

Example

Log/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-rite 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 which allows you to re-run the script and pick up where the process was halted. To override this functionality and force a fresh restart the backup or restore procedure, run the **backup.py** or **restore.py** script with an additional **--override** parameter (Example: **sudo python restore.py -r "2018-08-28" --override**)

Running a backup

Procedure

From the backup-restore directory, run the **backup.py** script as root (Example: **sudo python backup.py**).

Example log/console output 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 which allows you to re-run the script and pick up where the process was halted. To override this functionality and force a fresh restart the backup or restore procedure, run the **backup.py** or **restore.py** script with an additional **--override** parameter (Example: **sudo python restore.py -r "2018-08-28" --override**)

Running a restore

Procedure

1. Place the system in maintenance mode:

   `/opt/ibm/metaocean/helpers/maintenance.sh on`
2. Execute 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. For example: `sudo python restore.py -r "2018-08-28"`).
3. Remove the system from maintenance mode:

   `/opt/ibm/metaocean/helpers/maintenance.sh off`
4. If COS notifications have been lost during the period when the system was unavailable, these can be recovered using the COS Scanner replay function.

Chapter 7. 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**. See “Searching” on page 21 for details.
   - Use the Command Line Interface to execute REST API commands. In the **IBM Knowledge Center**, navigate to **IBM Spectrum Discover 2.0.0** > **REST API** > **Endpoints for working with a DB2 warehouse** for some JSON examples.
2. Go to **Reports** in the IBM Spectrum Discover main menu.

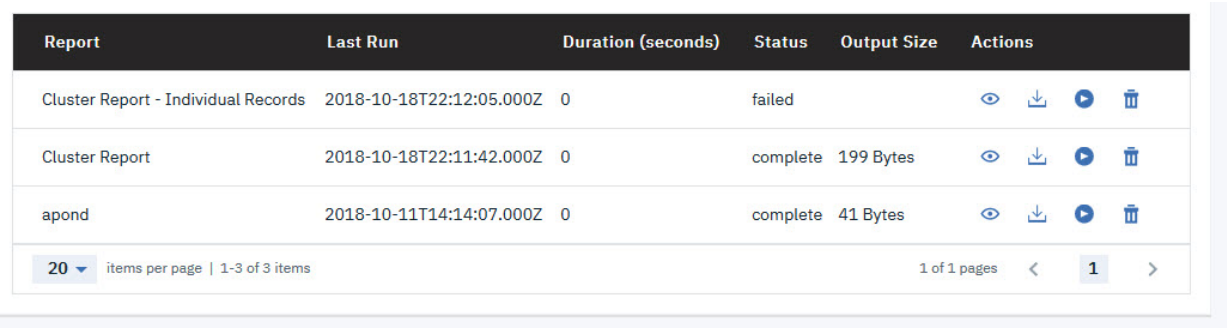


Figure 14. Reports table

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.

Figure 15. 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.

Re-run report
Click the **Go** "right arrow" icon to re-run the report.

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

Chapter 8. High availability for an Db2 Warehouse MPP deployment

Navigation title: High availability for an MPP deployment

For an MPP deployment, Db2® Warehouse provides out-of-the-box 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 16 on page 39 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 should use a method such as a load balancer to make head node failures transparent to connected applications.

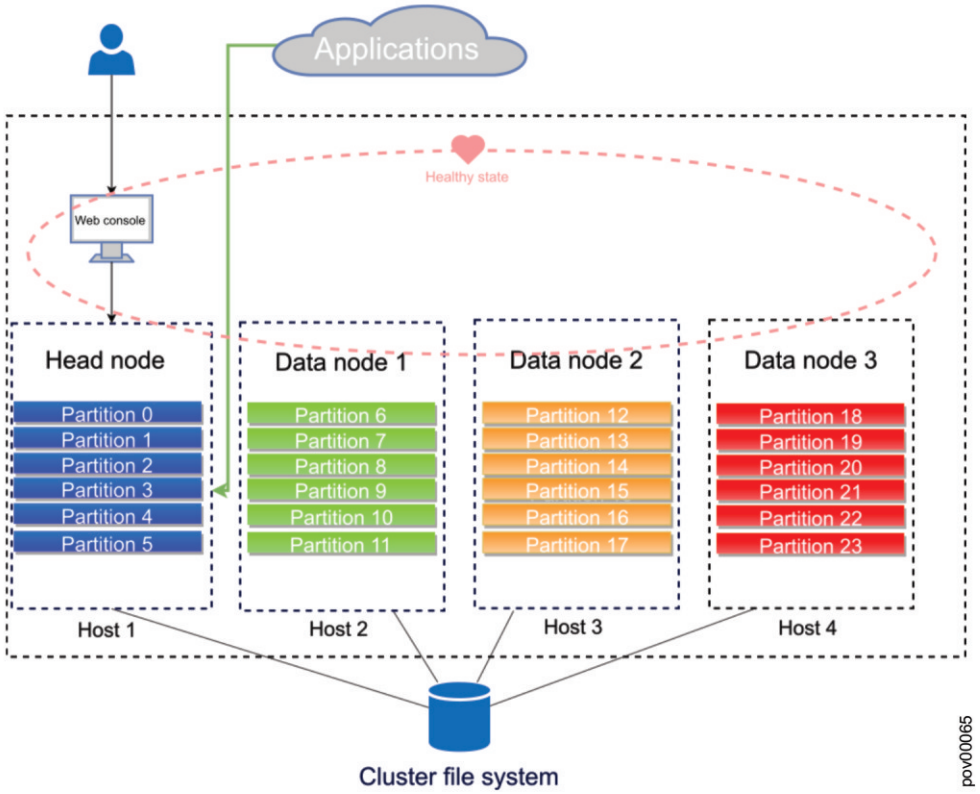


Figure 16. Steady state for HA group

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 17 on page 40, the web console is restarted on the new head node.

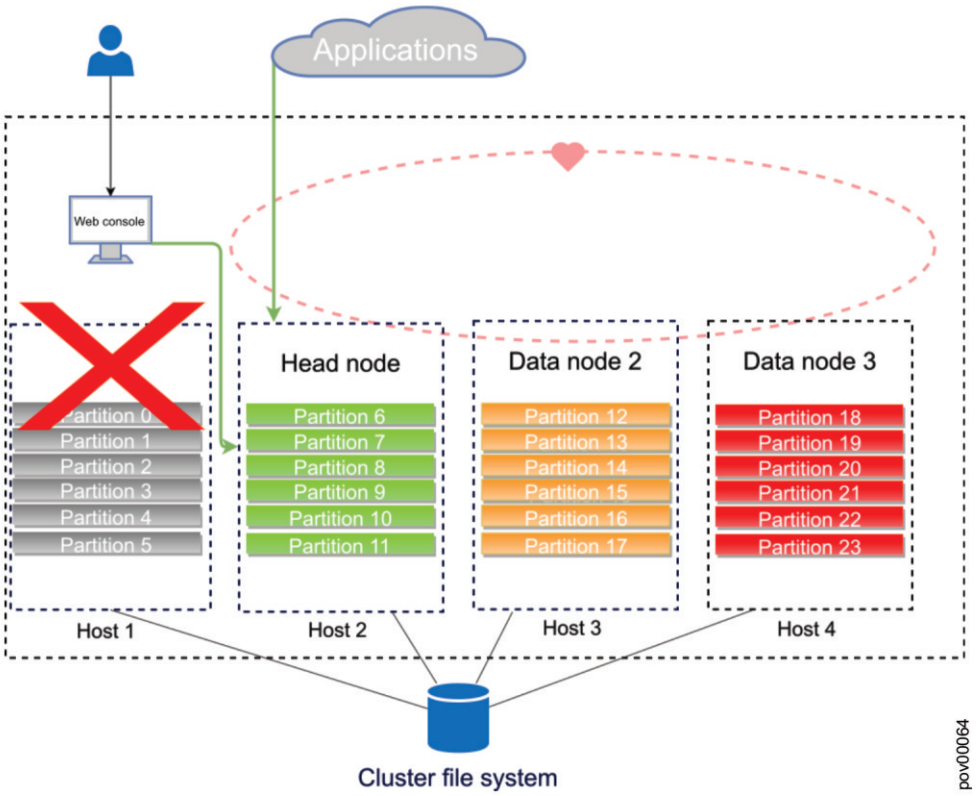


Figure 17. HA group after head node failover

After a head node failover, if the original head node becomes reachable again, restarting the system causes 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 9. 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

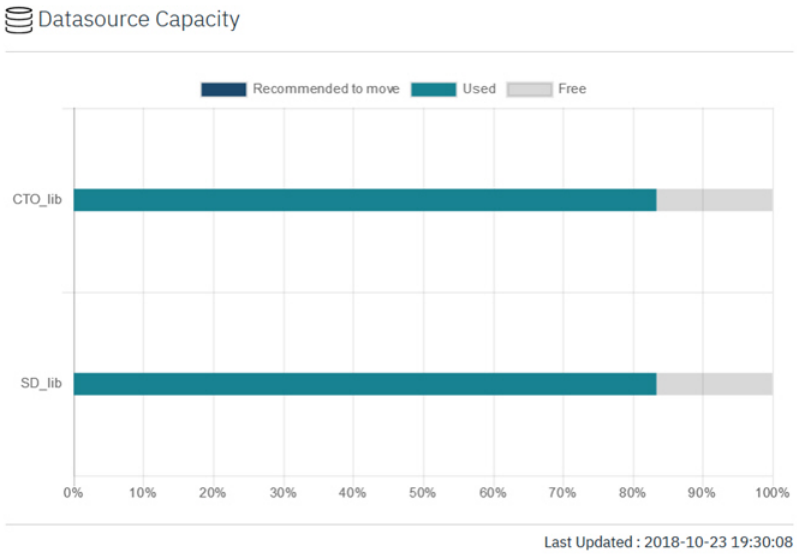
Use the **Home page** to monitor your environment for storage system capacity, used capacity, records indexed, and duplicate files. You can also view data usage for a specific area of your organization.

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.



Viewing used capacity

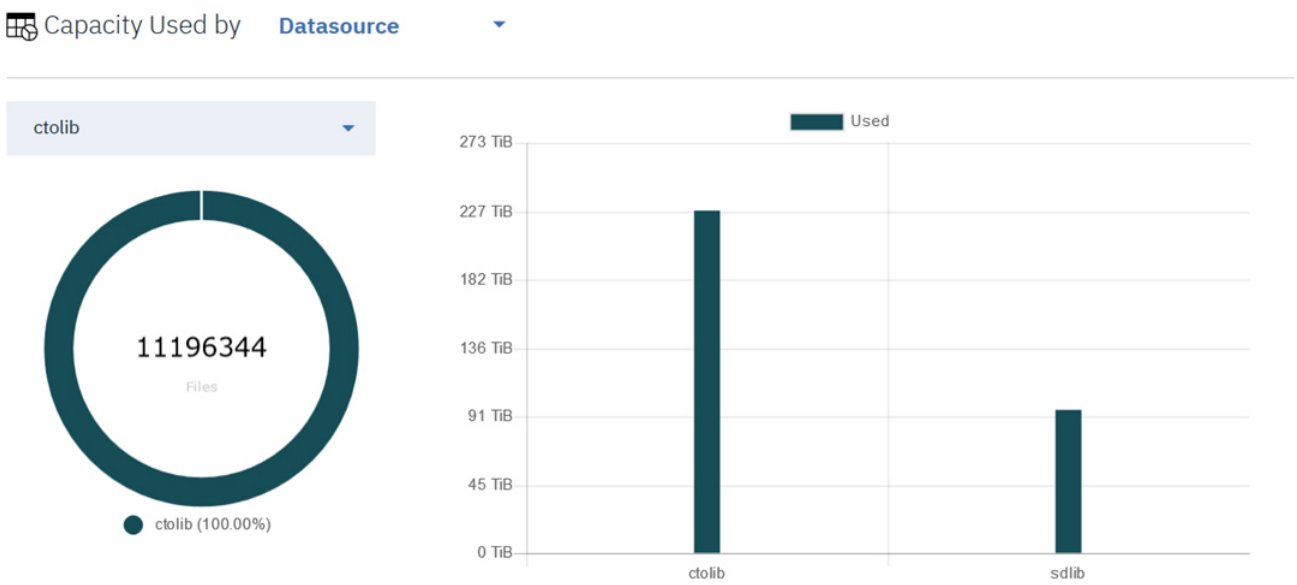
Use the **Capacity Used by** area to view graphs with an aggregated display of capacity usage for selected metadata attributes. The graphs provide details about capacity usage by aggregating across different attributes that are available from standard system metadata.

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.

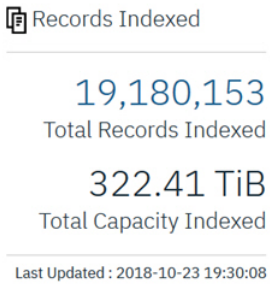
The data source graph displays the percentage of overall usage per data source for the selected attribute. You can select a specific capacity consumer to display in the graph.

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.



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 utilization.

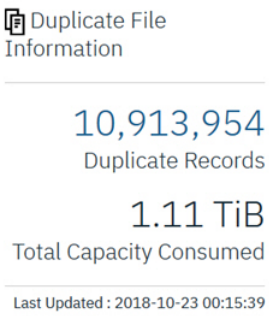


Click the **Total Records Indexed** value to open the **Search** page and perform a search of the indexed records.

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 these files consume 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.



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 18 on page 46 shows an example of a data source capacity widget.

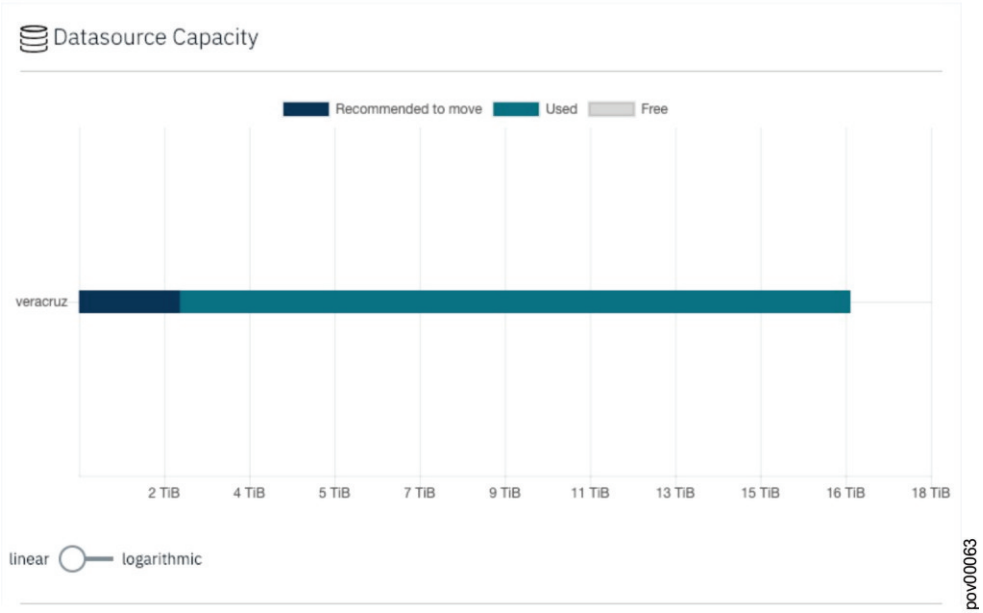


Figure 18. Example of a data source capacity widget

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 19 on page 46 shows an example of a screen that shows the **TEMPERATURE** tag.

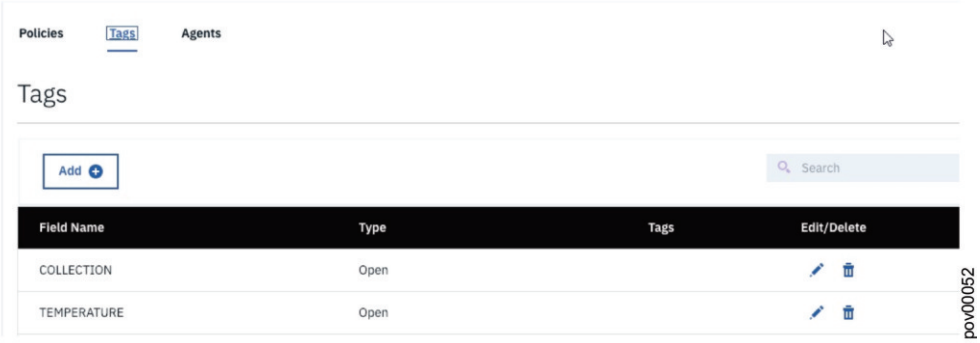


*Figure 19. Example of a screen that shows the **TEMPERATURE** tag*

Figure 20 on page 47 shows an example of an autotag policy to identify files and objects that have not been accessed for more than a year.

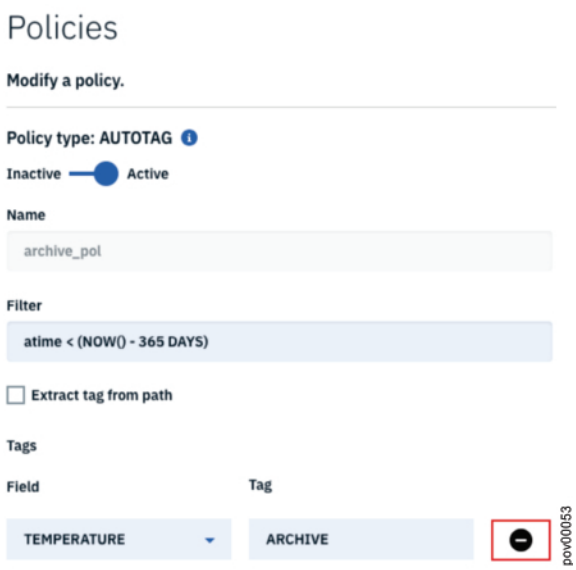


Figure 20. Example of an autotag policy to identify files and objects that have not been accessed in more than one year

]

[Deleting or editing a connection

About this task

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

Procedure

1. Click **Admin** to display a listing of existing connections as shown in Example of listing of existing connections

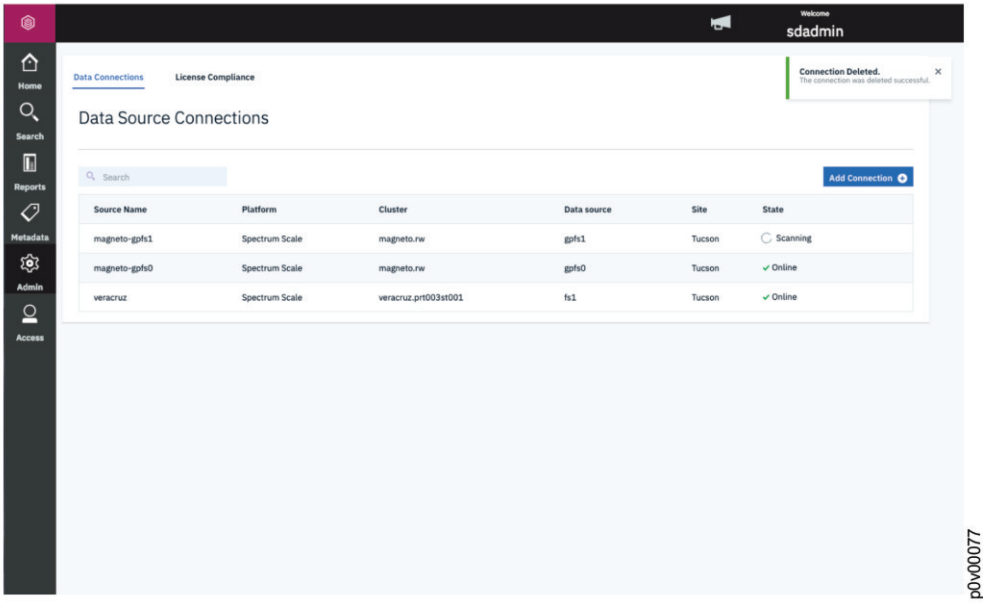


Figure 21. Example of a listing of existing connections

2. Click **Remove** to start the process to remove the data source connection as shown in Figure 22 on page 48.

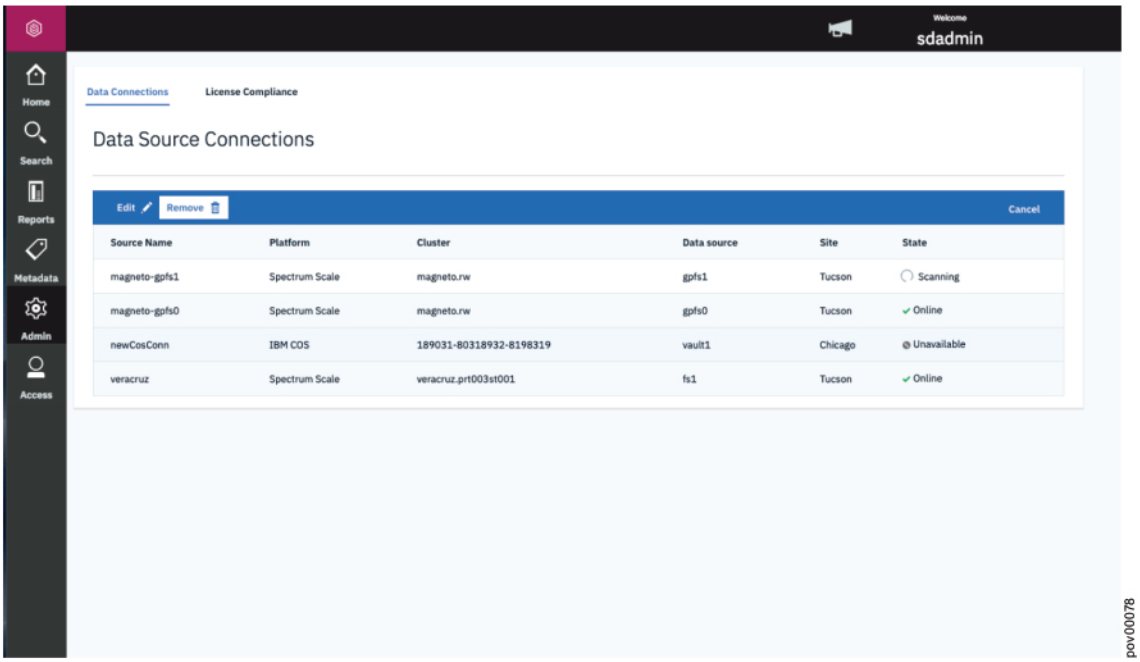


Figure 22. Starting the process to delete a data source connection

3. Clicking **Remove** displays a screen as shown in Figure 23 on page 48. If you are sure you want to delete the connection, click **Delete**.

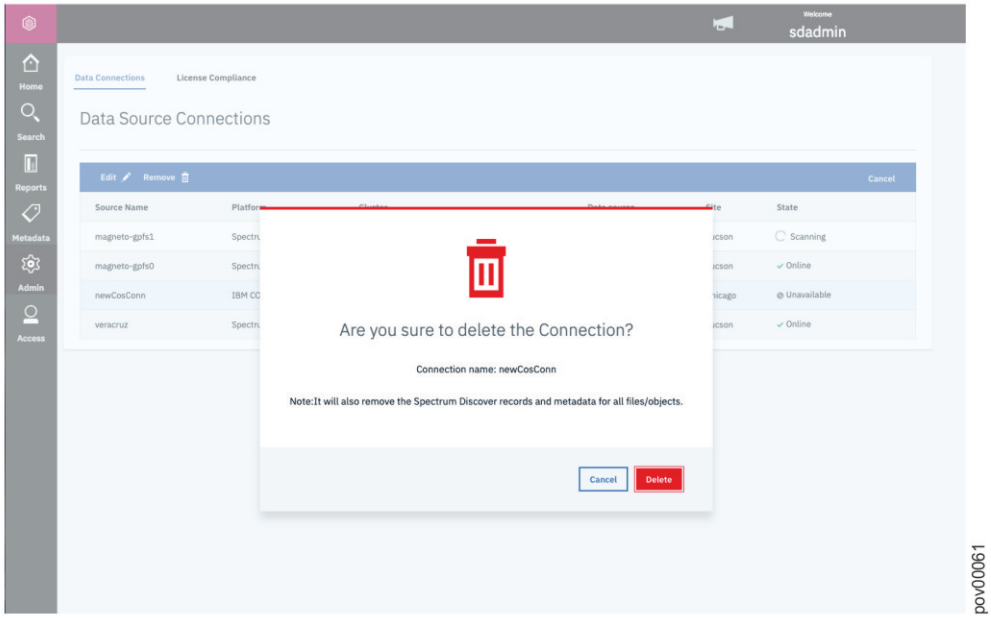


Figure 23. Example of a screen that shows how to delete a connection

4. To edit a connection, click **Edit**.
   a) Edit the appropriate fields in the window for **Edit Data Source Connection**.
   b) Click **Update Connection**.

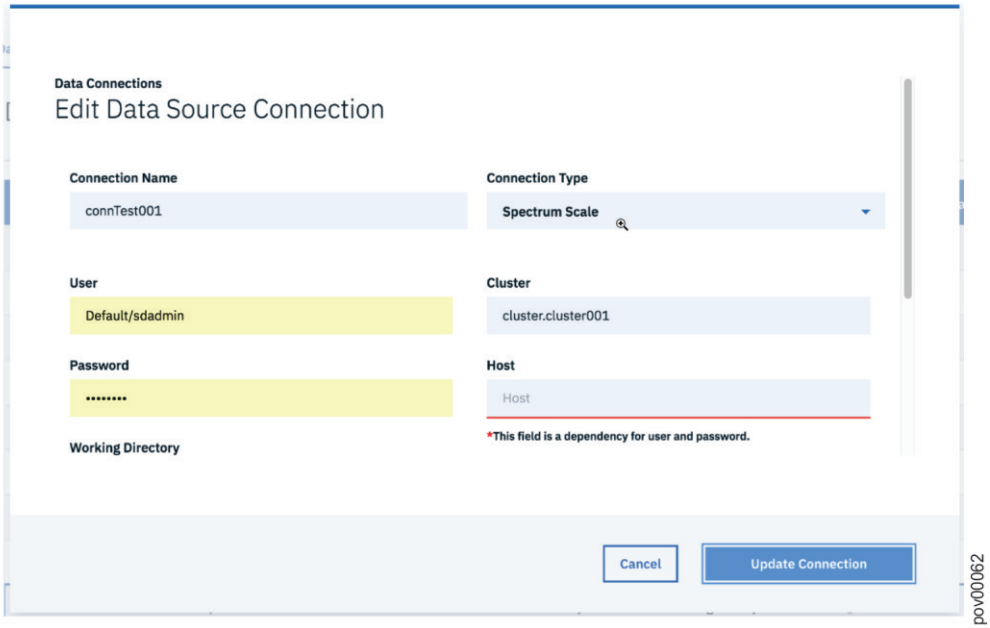


Figure 24. Example of a screen that shows how to edit a connection

Chapter 10. 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:

```
cat /opt/ibm/ibm-cloud-private/password
```

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 about using the **Dashboard** page, 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 about using the Grafana cluster monitoring dashboards, 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 provides 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 is 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 53.

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 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: `mo-ffdc-20180430074006548.tar.xz`.

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

FFDC script syntax

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
# ./ffdc all
```

Syntax for collecting audit log entries
Use the **namespaces** option to collect audit log entries.

```
# cd /opt/ibm/metaocean/helpers
# ./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
This is the standard option that should be used when you are reporting a failure situation to IBM Service. Use this option, unless asked to do otherwise by IBM service personnel.

cloudant
This option runs the Cloudant® "must gather" script.

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

logs
This 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
This option collects information about all namespaces in Kubernetes, including 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
This option collects service information for a number of services, including Docker, Kafka, and NFS, from all nodes in the IBM Spectrum Discover cluster.

system
This option captures operating system statistics about details such as free disk space, the time since last restart, memory usage, and network ports.

versions
This option captures the version information for the operating system,

- Docker
- Cloudant
- Kafka
- Kubernetes

Chapter 11. Updating the network configuration

These following topics describe how to update the network configuration of the master node and worker node for IBM Spectrum Discover. This section also describes how to back up and restore the database.

Before you update the network the configuration, go to "Backup database - run first" on page 55 to back up the database. After you back up the database, go to either "Master node" on page 55 or "Worker node" on page 56 depending upon which of the nodes require a network update. Lastly, go to "Restore database" on page 56.

Backup database - run first

Follow this procedure to back up the database before running any network configuration for the IBM Spectrum Discover.

Procedure

1. Log in to the master node using the *moadmin* user.
2. Run the following command: **cat /gpfs/gpfs0/db2wh/nodes**.
3. Take note of the *head* node.
4. Log in to the head node using *moadmin* user.
5. Run the following command: **cd /opt/ibm/metaocean/configuration**.
6. Run the following command: **./db2_offline_backup.sh**.
7. After you back up the database, go to either "Master node" on page 55 or "Worker node" on page 56 depending upon which of the nodes require a network update.

Master node

Follow this procedure to update the network configuration of the master node for IBM Spectrum Discover.

The following steps describe how to run the commands to update the master node. The process to update the master node takes several hours.

1. Log in to the master node as *moadmin* user.
2. Run the following command: **cd /opt/ibm/metaocean/configuration**.
3. Run the following command: **sudo ./update_network -p master:<old_FQDN>:<new_FQDN>**.
4. Follow the on-screen instructions.
5. After the VM restarts, and the new network configuration is applied to the master node, run **sudo ./update_network -t master:<old_FQDN>:<new_FQDN>**.

Note: You will be prompted to enter *moadmin* password shortly after you run the last command.

Configure a new network address

To get a list of interfaces, perform the following steps.

1. Run the following command: **ls /sys/class/net/**
2. Determine whether you are logging in to the master node or the worker node, then log in to the master node or worker node with the *moadmin* user.
3. Run the following command: **cd /opt/ibm/metaocean/configuration**.

4. Run the following command: **sudo ./mmconfigappliance -n <FQDN>:<interface>:<ip>:<netmask>:<gateway>:<dns>**.

Worker node

Follow this procedure to update the network configuration of the worker node for IBM Spectrum Discover.

The following steps describe how to run the commands to update the worker. The process to update the master node takes several hours..

1. Log in to the worker node as the *mo admin* user.
2. Run the following command to navigate to the configuration folder: **cd /opt/ibm/metaocean/configuration**.
3. Run the following command: **sudo ./update_network -p worker:<old_FQDN>:<new_FQDN>**.
4. Follow the on-screen instructions.
5. After the VM has restarted, and the new network configuration has been applied to the worker node run the following command: **sudo ./update_network -t worker:<old_FQDN>:<new_FQDN>**.

Note: You will be prompted to enter *moadmin* password shortly after running the last command.

Configure a new network address

To get a list of interfaces, perform the following steps.

1. Run the following command: **ls /sys/class/net/**
2. Determine whether you are logging in to the master node or the worker node, then log in to the master node or worker node with the moadmin user.
3. Run the following command: **cd /opt/ibm/metaocean/configuration**.
4. Run the following command: **sudo ./mmconfigappliance -n <fqdn>:<interface>:<ip>:<netmask>:<gateway>:<dns>**.

Restore database

After you run the database backup and do either the steps for the master node or worker node, you must restore the database.

Procedure

1. Log in to the master node using *moadmin* user
2. Run the following command: **cat /gpfs/gpfs0/db2wh/nodes**.
3. Take note of the head node.
4. Log in to the head node using the *moadmin* user.
5. Run the following command: **cd /opt/ibm/metaocean/configuration**.
6. Run the following command: **./db2_offline_restore.sh**

Chapter 12. IBM Spectrum Discover Content Inspection with Apache Tika

Introduction

This section describes how to use an IBM Spectrum Discover action agent to extract content from unstructured documents in data sources catalogued by IBM Spectrum Discover and automatically add the extracted content as custom tags in the IBM Spectrum Discover catalog.

The action agent leverages a customer provided Apache Tika server. Apache Tika is a popular open source tool for converting unstructured documents of various formats into a raw text stream.

Content is extracted from the unstructured documents according to the following workflow. The numbers in the list correspond to the numbers in Figure 25 on page 58.

1. The IBM Spectrum Discover policy engine generates a message containing a list of documents to inspect based on the user specified filtering criteria and places the message on a work queue.
2. The action agent obtains a list of documents to inspect by reading the message from the IBM Spectrum Discover policy engine work queue.
3. The action agent reads the documents from the source storage systems.
4. The action agent passes documents to a customer provided Apache Tika server.
5. The Apache Tika server converts the document to a raw text stream.
6. The raw text stream is passed to example agent parsers that search for the specified keywords in the documents. The following example parsers are provided as part of the action agent.

   Personally identifiable information (PII) detector
   Searches for user-defined terms in the text stream. If a term is found, a classification custom tag is set to PII. If no terms are found, the classification custom tag is set to non-PII.

   Word count parser
   Searches for the terms that are specified by the user and counts the number of occurrences of the term. The example action agent parsers may be customized by the user and users may also create new parsers and plug them into the action agent.

   The raw text stream is intercepted by example action agent parsers that search for the specified keywords in the documents.

   The example action agent parsers may be customized by the user and users may also create new parsers and plug them into the action agent.
7. Provides a response message based on the results of the document parsing where the response message contains the tag or tags that were found.
8. The IBM Spectrum Discover policy engine processes the response message and adds the tags as enrichments to the catalog for IBM Spectrum Discover.

Figure 25 on page 58 shows an example of the architecture for the action agent.

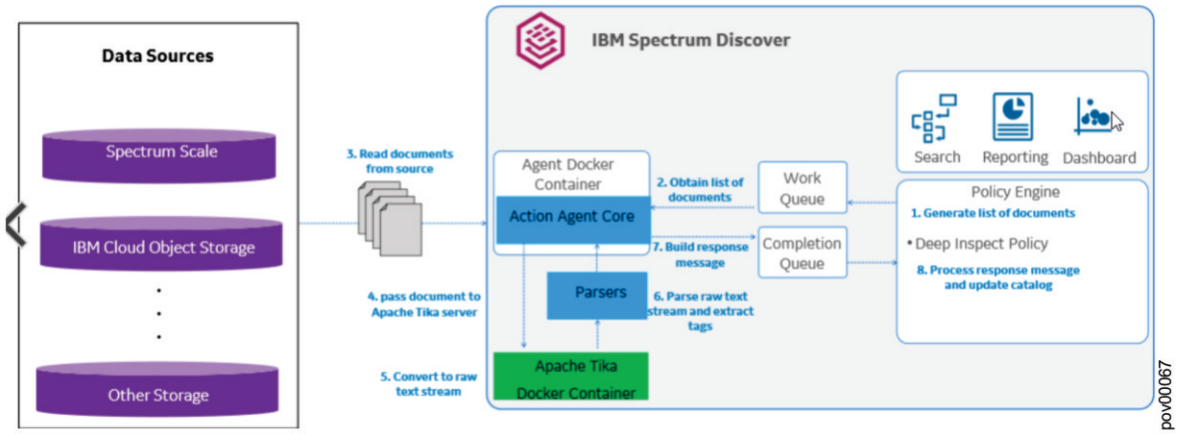


Figure 25. Example of the architecture for the action agent

]

[Apache Tika

Apache Tika is an open source tool that extracts metadata and text from over a thousand different file types, for example, PPT, XLS, and PDF.

You can parse the file types through a single interface, which makes Tika useful for search engine indexing, content analysis, conversion, and more. Apache Tika provides methods to extract metadata from well-defined document headers, for example, headers found in JPG and PDF files. Apache Tika also provides a method to convert the body of documents to a raw text string.

Go to https://tika.apache.org for an overview of Apache Tika. Go to https://tika.apache.org/1.20/formats.html for a description of the document formats supported by Apache Tika.

]

[Deployment considerations

The IBM Spectrum Discover content inspection action agent ships with the IBM Spectrum Discover virtual appliance. The action agent integrates with a customer supplied Apache Tika server.

The content inspection action agent and Apache Tika server can be configured to run inside a IBM Spectrum Discover single node or multi-node cluster. Both the action agent and the Apache Tika server are multi-threaded and a single instance of the action agent and Apache Tika server are deployed.

]

[Installation

This section describes the installation of the content inspection action agent and the Apache Tika server.

]

Location of action agent inside IBM Spectrum Discover virtual appliance

The action agent is pre-installed as a docker container in the IBM Spectrum Discover virtual appliance.

About this task

The action agent configuration file and other persistent data is located in the `/gpfs/gpfs0/agents/deepinspect` directory on the host machine.

The action agent parsers are located in `/gpfs/gpfs0/agents/deepinspect/parsers`. The action agent will look within this directory and attempt to load any python module within that complies with the correct interface.

Installation of the Apache Tika server on an IBM Spectrum Discover node

This section describes how to install the Apache Tia server on a IBM Spectrum Discover node.

Before you begin

The IBM Spectrum Discover node must have access to the Docker Hub Container Image repository at https://www.docker.com/products/docker-hub.

Procedure

Install the Apache Tika server as a docker container.

```
docker pull logicalspark/docker-tikaserver
```

Configuration and runtime of the IBM Spectrum Discover action agent and Apache Tika server

This section describes how to configure and run the IBM Spectrum Discover content inspection action agent and Apache Tika server.

Action agent registration

This section describes how to configure and run the IBM Spectrum Discover content inspection action agent and Apache Tika server

About this task

Perform the following steps to register an action agent.

Procedure

1. Obtain an auth token by using credentials of the data admin user as shown in the following example:

   ```
   curl-k https://<spectrum_discover_host>:443/auth/v1/token-u“<user_name>:<password>
   ```

   For a valid user, the auth token is returned in the “X-Auth-Token” response header.
2. Create a .json file with action agent details:

   ```
   {
   "action_agent": "extractagent",
   "action_id": "deepinspect",
   "action_params":
   ["extract_tags"]
   }
   ```

Note: Only ‘deepinspect’ is supported as the valid ‘action_id’ ‘action_agent’ parameter needs to be maximum 64 characters long and can contain alpha-numeric characters and ‘.’, ‘_’ or ‘-’ characters. Action parameters, when converted to string, can be maximum 256 characters long.

3. Submit the following request:

```
curl-k https://<spectrum_discover_host>/policyengine/v1/agents-H "Content-Type:
application/json"-X POST-d @agentregmsg.json-H “Authorization: Bearer <token>"
```

Response:

```
curl -k https://<spectrum_discover_host>policyengine/v1/agents H "Content-type:
application/json" -X POST -d @agent.json H "Authorization bearer <token>"
```

```
Content-Type: application/json
{
"broker_ip":
10.10.10.11,
"broker_port": 9093,
"work_q": "extractagent_work",
"completion_q": "extractagent_compl"
}
```

The following response codes are displayed based on the result of operation.

200
OK - For success.

401
Unauthorized - For invalid user name / password while registering the agents.

403
Forbidden - For an unauthorized user. That is, a user who does not have the ‘dataadmin’ role.

409
Conflict - If an agent already exists with the same name.

Configuration of the action agent

This section describes how to configure the action agent parameters by modifying the configuration file called deep inspect.conf located in the `/gpfs/gpfs0/agents/deepinspect/conf/` directory.

Procedure

1. Make a backup copy of the configuration file.

```
sudo cp /gpfs/gpfs0/agents/deepinspect/conf/deepinspect.conf /gpfs/gpfs0/agents/deepinspect/
conf/deepinspect.conf.bak
```

2. Edit the configuration file.

```
sudo vi /gpfs/gpfs0/agents/deepinspect/conf/deepinspect.conf
```

3. Configure the action agent specific parameters in the agent section of the file.
   a) Set the value of work_topic to match the output from the action agent registration response message. For example, if the work_q response from the action agent registration message was extractagent_work, set work_topic to this value
   b) Set the value of reply_topic to match the output from the action agent registration response message. For example, if the completion_q response from the action agent registration message was extractagent_compl, set reply_topic to this value
   c) Set the actions parameter to extract_tags

d) Set the Kafka hosts to be the fully qualified host name or IP address of the IBM Spectrum Discover node along with the :9093 port designation

e) Provide the name of the consumer group to be used for the action agent. The consumer group name is used when checking the Kafka topic lag

```
[agent]
work_topic=agent-testing_work
reply_topic=agent testing_compl
actions=extract_tags
kafka_hosts=luke.tuc.stglabs.ibm.com:9093
consumer_group=jobreader
```

f) Configure the Cloud Storage Object stanza if the action agent will be inspecting objects from the Cloud Storage Object storage.

1) Specify the endpoint URL to the value of the Cloud Storage Object accessor node

```
[cos]
endpoint_url=$$ACCESSOR_URL
```

g) Configure the agent_sdk parameters for the action agent

1) Specify the IBM Spectrum Discover hostname for the spectrum_discover_host

2) Specify the name of the action agent that was specified during the action agent registration process. For example, if extractagent was specified for the action_agent in the action agent registration JSON, set agent_name to extractagent.

```
[agent_sdk]
spectrum_discover_host=https://<spectrum discover host>:443/
agent_name=extractagent
```

Start the Apache Tika server

This section describes how to start the Apache Tika server.

Procedure

Start the Apache Tika server.

```
docker run --rm -p 9998:9998 logicalspark/docker-tikaserver
```

Note: With the --rm option, the Tika server will be deleted as soon as container stopped.

Customizing the PII detector parser

Users may customize the terms that are searched for along with how there are mapped to classification tag values by modifying the `/gpfs/gpfs0/agents/deepinspect/parsers/pii_parser.py` file.

About this task

As an example, the send_data function looks for the term ‘name’ in the raw text stream provided by the Apache Tika server and sets the classification tag value to pii.

```
def
send_data (self,
file_content):
```

```
"""
Receive file content data from action agent
"""
print "Parser: received TIKA file data (%d) bytes "%len(file_content)
# pipe this to grep, and if we find PII ('name'), mark result as pii

p=Popen(['grep','-i','-c','name'],stdout=PIPE,stdin=PIPE,stderr=STDOUT)
p_stdout=p.communicate(input=file_content)[0]
count=int(p_stdout.decode())
print"Found%dexamples of PII"%count
self.result.update({'classification':'pii'ifcount>0else'non-pii'})
if count > 0:
    # if address is also present, mark as confidential
    p = Popen(['grep','-i','-c','address'],stdout=PIPE,
stdin = PIPE,stderr=STDOUT)
             count=int(p_stdout.decode())
        self .result.update({'confidential':'true'if count>0 else 'false'})
        print "Result so far is %s" % self.result
```

This code can be customized to search for different terms and map the values to the appropriate tags based on customer needs.

]

[Writing a customer parsing plugin

Users can write a custom parser in the python language and put it in the following directory: `/gpfs/gpfs0/agents/deepinspect/parsers`

About this task

The agent loads its parsers dynamically upon startup. It will look within this user defined directory and attempt to load any python module that complies with the correct interface, which is specified in the `/gpfs/gpfs0/deepinspect/parsers/parser_interface.py` directory.

It must extend the classes AgentParserClientInterface and AgentDocumentParserInterface. When loaded it must be ready to receive data from the agent via the send_metadata and send_data methods, and to return results for that data as a dictionary of tag names and values via the get_result methods. Aside from those constraints, the parser is free to process the data as it requires. This includes passing to a different process entirely in order to build a custom text processing pipeline.

Note: Only one parser should reside in the `/gpfs/gpfs0/agents/deepinspect/parsers` directory at a time.

]

[Viewing content search agent logs

This section describes how to view the content search agent logs using the **deepinspect-agent-logs** alias.

About this task

Procedure

1. View the content search agent logs by invoking the **deepinspect-agent-logs** alias.

   ```
   deepinspect-agent-logs
   ```

2. Enter valid credentials for an IBM Spectrum Discover user with dataadmin role.

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

   Search the log for the string ERROR to find information about any failures. For example:

   ```
   [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',)))
```

In this example, the file failed in the inspection stage. The message shows that it is due to the agent not being able to contact the Tika server.

Tag and policy management for the IBM Spectrum discover action agent

This section describes how to specify the keywords that will be automatically searched for and in the source documents and added as custom tags to the IBM Spectrum Discover catalog by the IBM Spectrum Discover policy engine.

Tag management

This section describes how to create the tags to be used for the count words action agent.

About this task

When using the count_words parser you must create the characteristics tags that you want the parser to search for. In this example, the parser is configured to search for the number of occurrences of the term **thorax**.

Procedure

1. From the IBM Spectrum Discover graphical user interface, go to **Metadata** > **Tags** > **Add**.

   The count_words parser will update the tag value to indicate the number of occurrences of the tag key.
2. Type the name of the tag and select the **Characteristics** type.

   Figure 26 on page 63 shows an example of a tag named thorax.

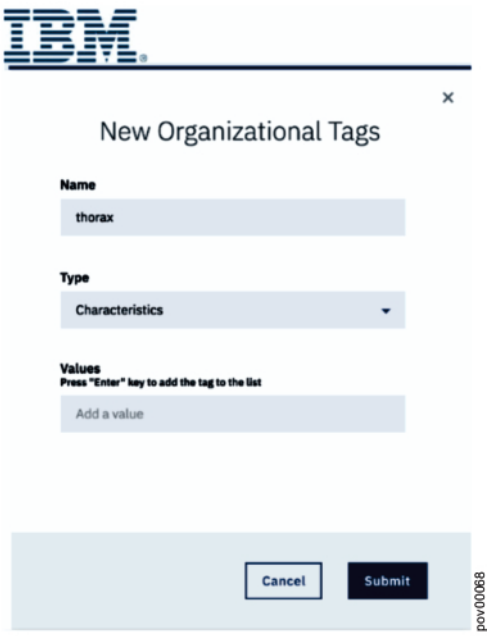


Figure 26. Example of a tag named thorax

PII detector tag management

This section describes how to create tags to be used for the PII detector action agent.

About this task

When using the PII detector parser you must create a restricted tag and the associated classification values that you want the parser to set. For example, create a restricted tag named classification and set the values to pii and confidential.

The tag name classification and the values must match the values set in the pii detector parser program file called pii_parser.py. You may customize the tag values based on your scenario.

Figure 27 on page 64 shows an example of

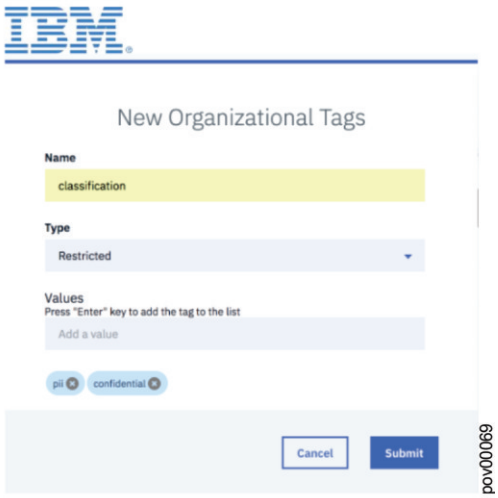


Figure 27. Example of new organizational tags

Policy management

This section describes how to create an IBM Spectrum Discover policy.

Procedure

1. Select the DEEPINSPECT policy type.
2. Select the name of the action agent that was registered in “Configuration of the action agent” on page 60.
3. Specify the filtering criteria that controls which documents will be passed to the DEEPINSPECT agent.
4. Select the tag or tags to search for and their associated values.
5. Select the schedule when you want the policy to run.

   Figure 28 on page 65 shows an example of how to create an IBM Spectrum Discover policy.

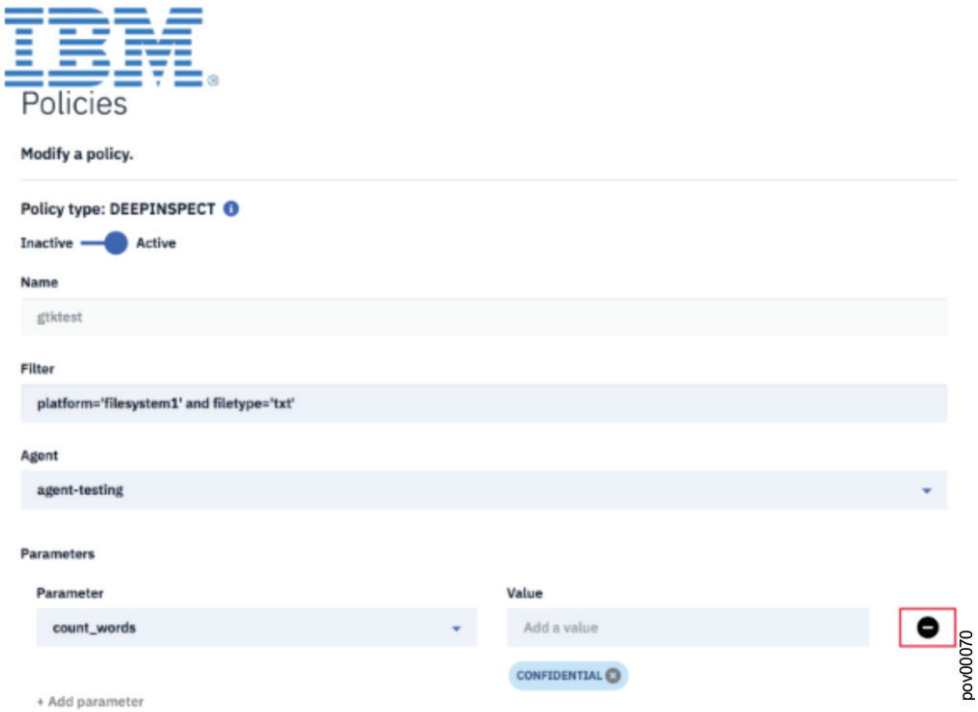


Figure 28. Example of how to create an IBM Spectrum Discover policy.

Viewing enrichments

The capacity consumed by the extracted tags can be viewed by the Capacity Used dashboard widget by selecting the desired tag.

Procedure

Search for specific extracted tags by going to Search and typing the name of the tag and the value to search for or by selecting the appropriate tags from the ‘start a visual exploration’ section followed by clicking the arrow button.

Chapter 13. Disaster recovery procedures

This process to recover from a disaster involving a 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 multinode system.

Running disaster recovery

About this task

Follow these steps to recover single-node and multinode systems.

Procedure

1. Record the time of the failure.
2. If the virtual machine hosting IBM Spectrum Discover is still running, shut it down.
3. Redeploy the system as described in these substeps.
   a) For a single node system, redeploy with the same parameters as the failed deployment as described in *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 as described in *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. Run the initial setup of the backup/restore procedure as described in"Initial setup configuration" on page 35
5. Restore the system using a previous backup as discussed in "Running a restore" on page 36
6. Do not remove the system from maintenance mode until you complete the following substeps.
   a) For a multi-node system confirm that DB2 is running, and use this command to determine the HEAD node.

   ```
   docker exec -it Db2wh status
   ```

   This provides output for a multi-node system similar to the following table.

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 into 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. Its encrypted value is stored in `/opt/ibm/db2wh/password` and is decrypted with the following command

```
PYTHONPATH=/opt/ibm/metaocean/provisioning/filter_plugins python -c
          "from metaocean import password_decode; print password_decode('$(sudo cat
          /opt/ibm/db2wh/password)')"
```

2) Record the value of the database password. Its value is stored in `/opt/ibm/db2wh/password`

3) Update the DB password with this value.

```
docker exec -it Db2wh setpass  <DB password>
```

4) Log into the DB docker container, and change to the DB user.

```
docker exec -it Db2wh bash
su - db2inst1
```

5) 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 as described in *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 Cloud Object Store notifications were being used to keep the SD metadata up to date, run the Replay procedure for the period of time for which the system was unavailable. For more information refer to *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 SD was unavailable.

Chapter 14. Troubleshooting

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.

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.

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.

Sorting search results does not sort using all results

You can change the sort order of a column in the Results table, by click a column's header. Currently, the sort is limited to local data supported by your web browser, up to a maximum of 10,000 records per query. If there are more than 10,000 records, the sorted data will be incomplete.

Cannot filter issues after search

After performing a search, using the filter button to narrow down the results does not change the results in the table. To add additional filters to a search, change the search query to include the new filter. For example, add `mtime` to the query to filter based on modification time.

Converting a grouped search to individual record mode doesn't work for null values

After performing a grouped search including at least one "Empty Value" option and selecting **Convert to individual record mode** on that group, the search returns no results.

This is caused by the query having single quotes around the null value. To fix the issue, remove the single quotes from around the null value. For example, change from:

```
temperature in ('null') AND sizerange in ('extra large')
```

to:

```
temperature in (null) AND sizerange in ('extra large')
```

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.

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}
```

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.

DB2 Warehouse installation port conflict - Wait for DB2WH to initialise

Occasionally, ports required by DB2 Warehouse will be used by ICP services that select a random port in a high range. When this happens, Spectrum Discover installation will fail at the step "Wait for DB2WH to initialise", and the DB2 Warehouse logs will contain 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
```

Network configuration update: Please read before attempting

The network configuration update for worker node consists of the following commands:

1. `sudo ./update_network -p worker:<old_FQDN>:<new_FQDN>`
2. `sudo ./update_network -t worker:<old_FQDN>:<new_FQDN>`

Before running the second command please complete the following steps:

1. `rm -f /opt/ibm/ibm-cloud-private/password`
2. Replace the Scale section in /opt/ibm/metaocean/provisioning/network_config_worker_post.yml with the following:

```
# Scale
- hosts: master
  tasks:
    - name: SCALE - Stop GPFS on all nodes
      command:  /usr/lpp/mmfs/bin/mmshutdown -a
      changed_when: True
      become: yes
      register: result
      failed_when: "result.rc != 0 and 'Unable to reach any quorum node' not in result.stderr"

    - name: Remove the cached key for old node
      shell: "ssh-keygen -R {{ old_ip }}"
      changed_when: True

    - name: SCALE - Update interfaces for new node
      command: /usr/lpp/mmfs/bin/mmchnode --admin-interface {{ new_ip }} --daemon-interface
{{ new_ip }} -N {{ old_ip }}
      changed_when: True
      become: yes
      failed_when: "result.rc !=0 and 'No nodes were found that matched the input
specification' not in result.stderr"

    - name: SCALE - Start GPFS on all nodes
      command:  /usr/lpp/mmfs/bin/mmstartup -a
      changed_when: True
      become: yes

    - name: Check status
      shell: /usr/lpp/mmfs/bin/mmgetstate -a | grep active
      register: result
      changed_when: False
      until: result.stdout_lines|length == groups['all']|length
      retries: 30
      delay: 10
```

```
  - name: SCALE - Add quorum
    command: /usr/lpp/mmfs/bin/mmchnode --quorum -N {{ new_ip }}
    changed_when: True
    become: yes
    when: is_single_node|bool == false

  - name: SCALE - Stop GPFS on all nodes
    command:  /usr/lpp/mmfs/bin/mmshutdown -a
    changed_when: True
    become: yes
```

NOTE: The file is owned by root, you will need to use sudo when modifying this file.

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 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`

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-rbac-auth-\*`
- `\*-auth-rbac-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-\*`

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
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>`

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/`.

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.

Exception in DB2WH-REST if authorization token has expired

When using an expired authorization token, requests to the IBM Spectrum Discover REST API will fail with an Internal Server Error (status code 500). The request should fail with an Unauthorized Error (status code 401). To workaround the issue, get a new token and resubmit the REST API request.

CentOS reboots under load

CentOS might reboot under load due to a kernel bug.

For more information, see https://access.redhat.com/solutions/3492911.

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).
```

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 will stop 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>`

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
RESTARTS   AGE
    anxious-fly-consumer-scale-scan-consumer-56b6c4ff9c-d4k8v   1/1       Running
0          46m
    anxious-fly-consumer-scale-scan-consumer-56b6c4ff9c-h862t   1/1       Running
0          46m
    anxious-fly-consumer-scale-scan-consumer-56b6c4ff9c-j4649   1/1       Running
0          46m
    anxious-fly-consumer-scale-scan-consumer-56b6c4ff9c-ksbh4   1/1       Running
0          46m
    anxious-fly-consumer-scale-scan-consumer-56b6c4ff9c-kt9sc   1/1       Running
0          46m
    anxious-fly-consumer-scale-scan-consumer-56b6c4ff9c-lk8jz   1/1       Running
0          46m
    anxious-fly-consumer-scale-scan-consumer-56b6c4ff9c-p2lr6   1/1       Running
0          46m
    anxious-fly-consumer-scale-scan-consumer-56b6c4ff9c-qqhfd   1/1       Running
0          46m
    anxious-fly-consumer-scale-scan-consumer-56b6c4ff9c-wknbc   1/1       Running
0          46m
    anxious-fly-consumer-scale-scan-consumer-56b6c4ff9c-zrp6k   1/1       Running
0          46m
```

```
    $ kubectl get pods -n producercos

    NAME                                                          READY     STATUS
RESTARTS   AGE
    exacerbated-tarsier-producer-cos-producer-64748764cf-pctwz   1/1       Running
0          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 Db2whOutStream 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 Speed   Time     Time     Time      Current
                                          Dload   Upload  Total    Spent    Left      Speed
     100  3194  100  3194     0    0      298     0       0:00:10  0:00:10  --:--:--  911
     "auth-rbac : 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.

Chapter 15. mmconfigappliance command

Manages configuration and network settings of IBM Spectrum Discover virtual appliance nodes.

Synopsis

```
mmconfigappliance -n {hostname}:{adapter}:{ip}:{netmask}:[gateway]:[dns]
```

or

```
mmconfigappliance -m ip
```

or

```
mmconfigappliance -s ip
```

or

```
mmconfigappliance -t ntp
```

or

```
mmconfigappliance -d {device}:{partition}:{vg}:{lv}
```

or

```
mmconfigappliance -p password
```

Availability

Available on all IBM Spectrum Discover editions

Description

Use the **mmconfigappliance** command to configure IBM Spectrum Discover virtual appliance nodes by adding the IP address of the master and slave nodes in the hosts file. You can configure network settings, such as IP address, netmask, and gateway, of an virtual appliance node for a given adapter.

Additionally, you can use this command to configure Network Time Protocol (NTP), resize the root disk, and also update the administrator password.

Parameters

-n

Configures the network settings and the IBM Spectrum Discover master. The variable values need to be separated by a '**:**' without any space.

hostname
The name to be used for the host virtual appliance.

adapter
The name to be used for the Ethernet adapter device. For example, `ens192`.

ip
The system IP address of the Virtual Appliance Node. For example, `192.168.56.101`.

netmask
The system netmask of the Virtual Appliance Node. For example, `255.255.255.0`.

gateway
The system gateway of the Virtual Appliance Node. For example, `192.168.56.1`.

dns
The DNS server. For example, `192.168.56.1`.

-m
Specifies the IBM Spectrum Discover master. The -m option adds or updates the IP address of the master system in the hosts file.

ip
The IP address to bind the services. For example, `192.168.56.101`.

-s
Specifies the IBM Spectrum Discover slave to be added and configured. The slave must be accessible on the root account with passwordless SSH.

ip
The IP address of the slave.

-t
Configures the NTP settings of the system.

ntp
The NTP server. For example, `192.168.56.1`.

-d
Resizes the root disk. The variable values need to be separated by a '**:**' without any space.

device
Specifies the device on which to extend the partition.

partition
Specifies the partition number to expand the root disk.

vg
Specifies the volume group containing the logical volume to expand.

lv
Specifies the logical volume to expand.

-p
Updates the administrator password.

password
Specifies the password.

Exit status

0
Successful completion.

nonzero
A failure has occurred.

Security

You must have root authority to run the `mmconfigappliance` command.

Examples

1. To configure the network settings and specify the IBM Spectrum Discover master, issue this command:

```
mmconfigappliance -n momaster:ens192:192.168.56.101:255.255.254.0:192.168.56.1:192.168.56.1
```

2. To specify the IBM Spectrum Discover master, issue this command:

```
mmconfigappliance -m 192.168.56.101
```

3. To specify the IBM Spectrum Discover slave, issue this command:

```
mmconfigappliance -s 192.168.56.102
```

4. To configure the NTP settings, issue this command:

```
mmconfigappliance -t 192.168.56.1
```

5. To resize the root disk, issue this command:

```
mmconfigappliance -d sdb:1:c1:root
```

Location
/opt/ibm/metaocean/configuration/

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).

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) 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.

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

T

U

V

W