SIEMENS

MindSphere

Integrated Data Lake

System Manual

05/2022

Legal information

Warning notice system

This manual contains notices you have to observe in order to ensure your personal safety, as well as to prevent damage to property. The notices referring to your personal safety are highlighted in the manual by a safety alert symbol, notices referring only to property damage have no safety alert symbol. These notices shown below are graded according to the degree of danger.

⚠ DANGER
indicates that death or severe personal injury **will** result if proper precautions are not taken.

⚠ WARNING
indicates that death or severe personal injury **may** result if proper precautions are not taken.

⚠ CAUTION
indicates that minor personal injury can result if proper precautions are not taken.

NOTICE
indicates that property damage can result if proper precautions are not taken.

If more than one degree of danger is present, the warning notice representing the highest degree of danger will be used. A notice warning of injury to persons with a safety alert symbol may also include a warning relating to property damage.

Qualified Personnel

The product/system described in this documentation may be operated only by **personnel qualified** for the specific task in accordance with the relevant documentation, in particular its warning notices and safety instructions. Qualified personnel are those who, based on their training and experience, are capable of identifying risks and avoiding potential hazards when working with these products/systems.

Proper use of Siemens products

Note the following:

⚠ WARNING
Siemens products may only be used for the applications described in the catalog and in the relevant technical documentation. If products and components from other manufacturers are used, these must be recommended or approved by Siemens. Proper transport, storage, installation, assembly, commissioning, operation and maintenance are required to ensure that the products operate safely and without any problems. The permissible ambient conditions must be complied with. The information in the relevant documentation must be observed.

Trademarks

All names identified by ® are registered trademarks of Siemens AG. The remaining trademarks in this publication may be trademarks whose use by third parties for their own purposes could violate the rights of the owner.

Disclaimer of Liability

We have reviewed the contents of this publication to ensure consistency with the hardware and software described. Since variance cannot be precluded entirely, we cannot guarantee full consistency. However, the information in this publication is reviewed regularly and any necessary corrections are included in subsequent editions.

Siemens AG
Digital Industries
Postfach 48 48
90026 NÜRNBERG
GERMANY

Ⓟ 05/2022 Subject to change

Table of contents

Integrated Data Lake is an application to import and store the historical IoT data, access cross accounts and perform analytics on the data.

Document history

1

Date	Changes	Link
2022-05-07	Updated topic on Schedule Import in Time Series Import chapter	Importing the scheduled Time Series data (Page 23)
2022-03-19	Added a topic on Schedule Import in Time Series Import chapter	Importing the scheduled Time Series data (Page 23)
	Updated topic "User interface Integrated Data Lake" with Schedule Import	User interface Integrated Data Lake (Page 11)
2021-12-09	Updated topic "Searching the files/objects in Data Explorer"	Searching the files/objects in Data Explorer (Page 55)
	Updated topic "Integrating Integrated Data Lake with Fleet Manager" with image format information	Integrating Integrated Data Lake with Fleet Manager (Page 54)
2021-10-30	Updated topic "Managing Cross Account accesses"	Managing Cross Account accesses (Page 30)
2021-09-04	Updated topic "Searching the files/objects in Data Explorer"	Searching the files/objects in Data Explorer (Page 55)
2021-08-07	Updated Data Explorer UI	Data Explorer (Page 47)
2021-07-10	Updated Event Subscription UI	Creating an Event Subscription (Page 59)
2021-05-15	Updated Time Series Import UI	Time Series Import (Page 21)
	Upfated Cross Account Access UI	Cross Account Access (Page 29)
2021-02-27	Updated Event Subscription UI	User interface Integrated Data Lake (Page 11)
	Updated Data Explorer UI	Exploring data in IDL (Page 47)
2021-01-30	Updated Event Subscription UI	Creating an Event Subscription (Page 59)
2020-11-28	Added a topic on Search view in Event Subscription	Creating an Event Subscription (Page 59)
2020-10-10	Added a topic on Search view in Data Explorer	Searching the files/objects in Data Explorer (Page 55)
	Added a chapter on Service Principal	Service Principal (Page 35)
	Added a topic on Integrated Data Lake plugin in Data Explorer	Integrating Integrated Data Lake with Fleet Manager (Page 54)
2020-09-05	Added a topic on "Adding metadata tags to the object"	Adding metadata tags to the object (Page 52)
2020-08-05	Added a topic on configuring AWS Athena in Integrated Data Lake	Configure AWS Athena in Integrated Data Lake (Page 32)
2020-07-11	UI changes	Data Explorer (Page 47)
2020-05-16	Updated a chapter on Data Explorer	Data Explorer (Page 47)
	Updated a chapter on Time Series Import	Time Series Import (Page 21)
2020-02-20	Updated a chapter on Data Explorer	Data Explorer (Page 47)

Date	Changes	Link
2020-01-08	Added a chapter on Data Explorer	Data Explorer (Page 47)
	Updated screenshots and Data Explorer UI	User interface Integrated Data Lake (Page 11)
2019-12-02	New document	-

Introduction to Integrated Data Lake 2

A Data Lake is a repository that allows you to store structured and unstructured data/ objects in its native format as needed. It handles large data pools for which the schema and data requirements are not defined until the data is queried. This offers more agility and flexibility than traditional data management systems. The stored data can be used for further analysis as per customer's requirement.

Integrated Data Lake provides following features

- Import time series data/ objects
- Generate signed URLs to upload, update or download data/ objects
- Delete data/ objects
- Add, update and delete tags for data/ objects
- Receive notifications
- Data access using cross account access
- Subtenancy support.

User interface Integrated Data Lake

3

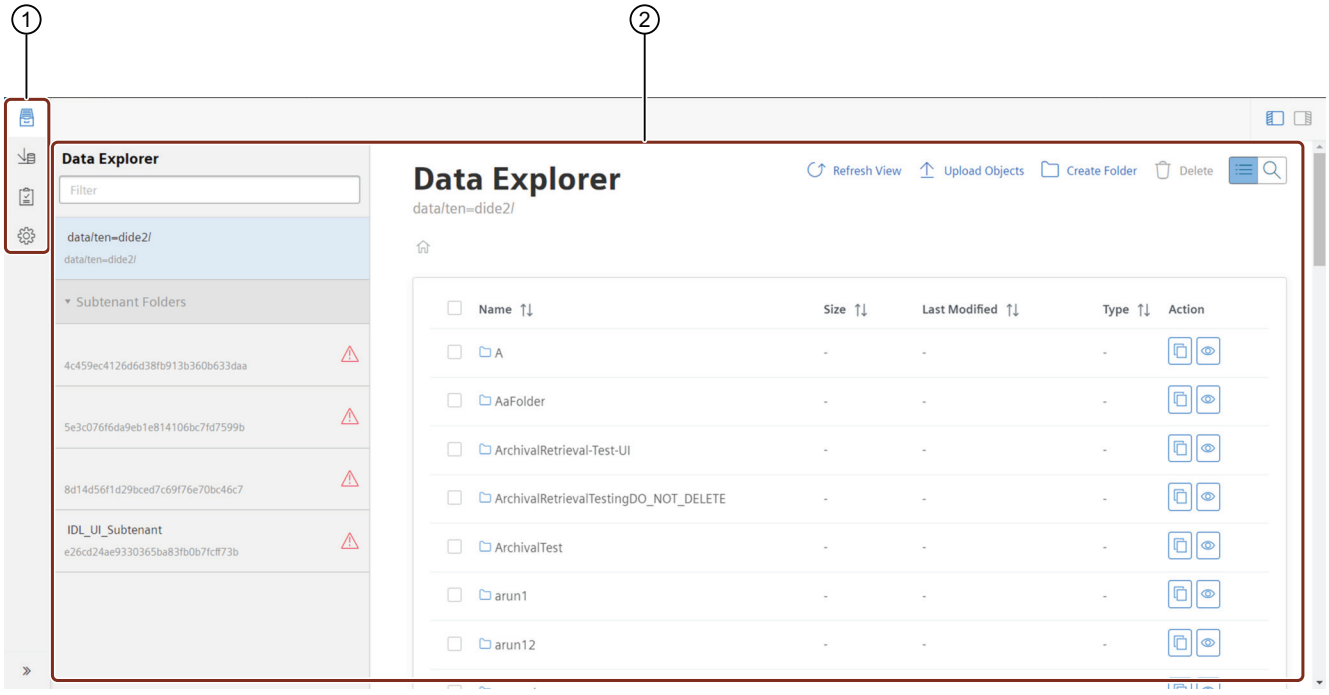
You can operate Integrated Data Lake using the "Navigation Tabs" and "Work area".

You can access Integrated Data Lake from MindSphere Launchpad. The following screenshot shows the Integrated Data Lake icon:

You can access the Integrated Data Lake icon to view the start screen.

Start screen

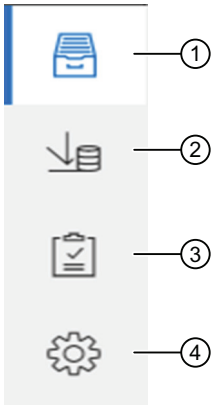
The Integrated Data Lake displays the "Data Explorer" navigation tab. The following screenshot shows the different elements of the Integrated Data Lake user interface:



① Navigation tabs
② Work area

Navigation tabs

You can access the navigation tabs which are available in the left side of the work area.



① Data Explorer
② Time Series Import
③ Event Subcription is to get notifications
④ Cross Account for Europe 1 and Service Prinicipal for Europe 2

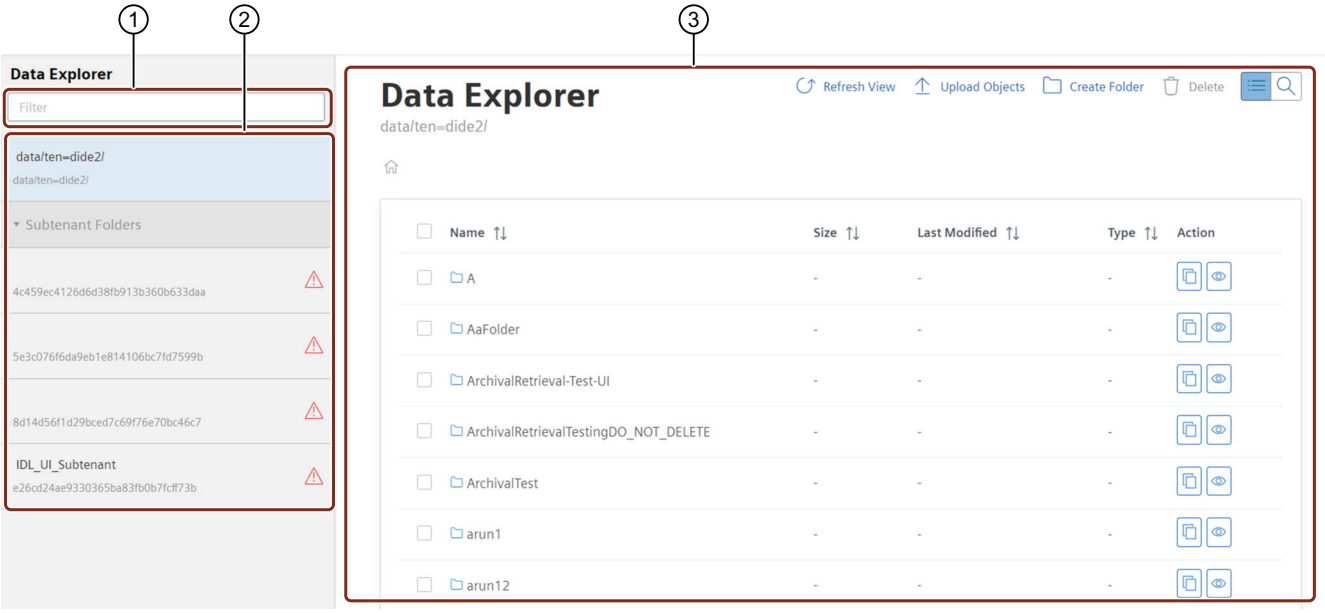
Note

Region deviation

- Feature "Cross account" is applicable for Region Europe 1.
- Feature "Service Principal" is applicable for Region Europe 2.

Data Explorer

The Data Explorer navigation tab is located on the left side of the Integrated Data Lake UI.



① Filter or search the folders

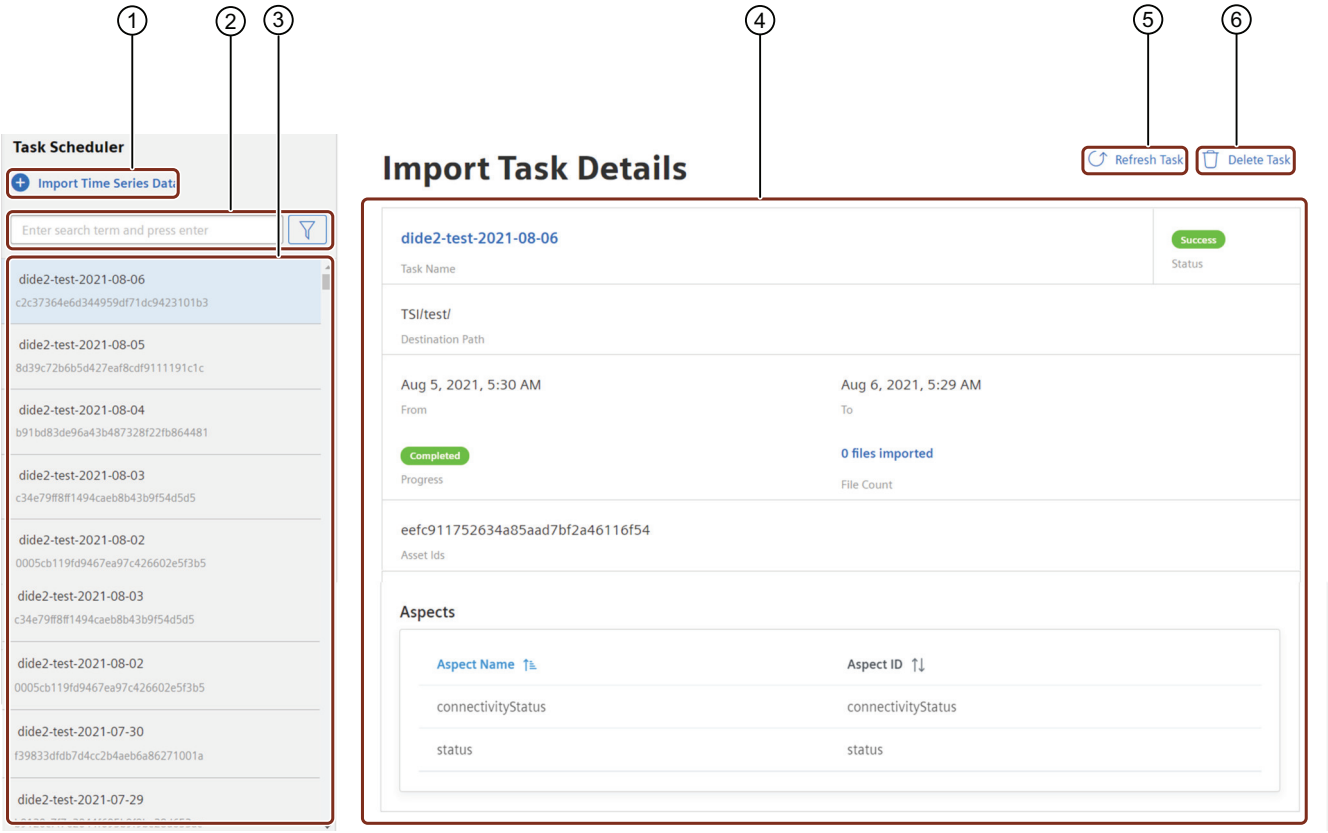
② List of the folders

③ Work area

Time Series Import

The Time Series Import navigation tab is located on the left side of the Integrated Data Lake UI. In Time Series Import, you can import the data by the following features:

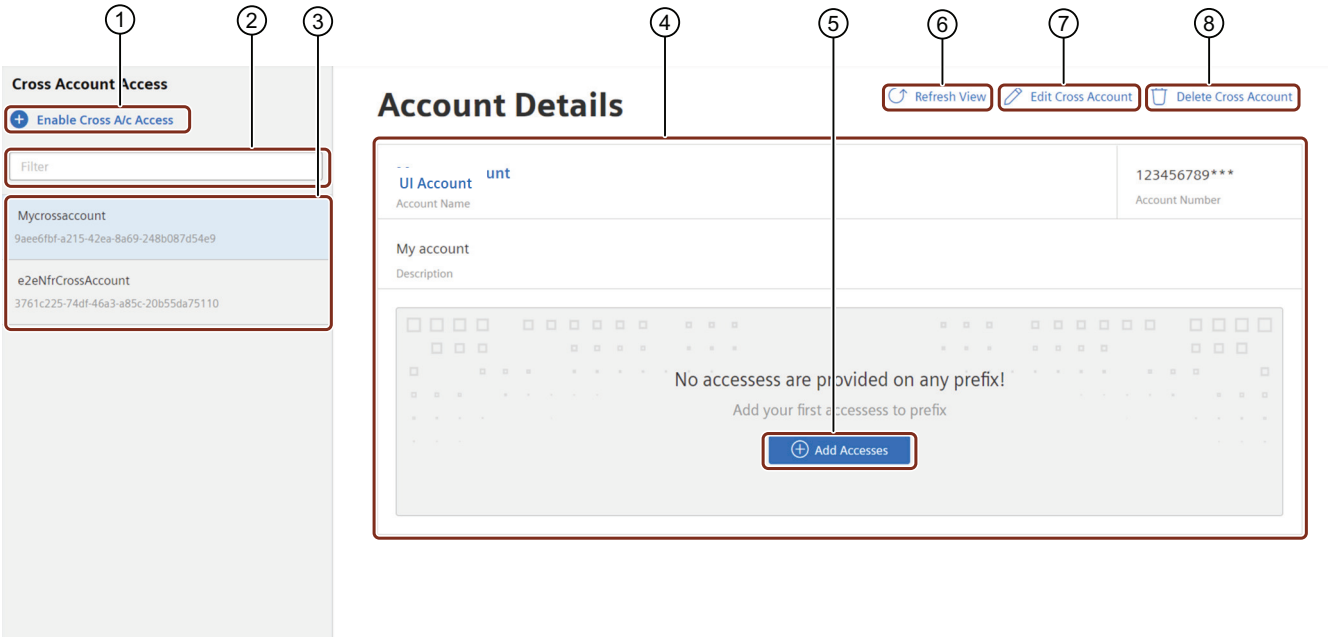
- One Time Import
- Schedule Import



① The Time Series Import creates a job to import time series data to Integrated Data Lake
② Filter or search of the task created for import data
③ Task which are created for imported data
④ Details of the selected task
⑤ Updates the task details
⑥ Deletes the selected task

Cross Account

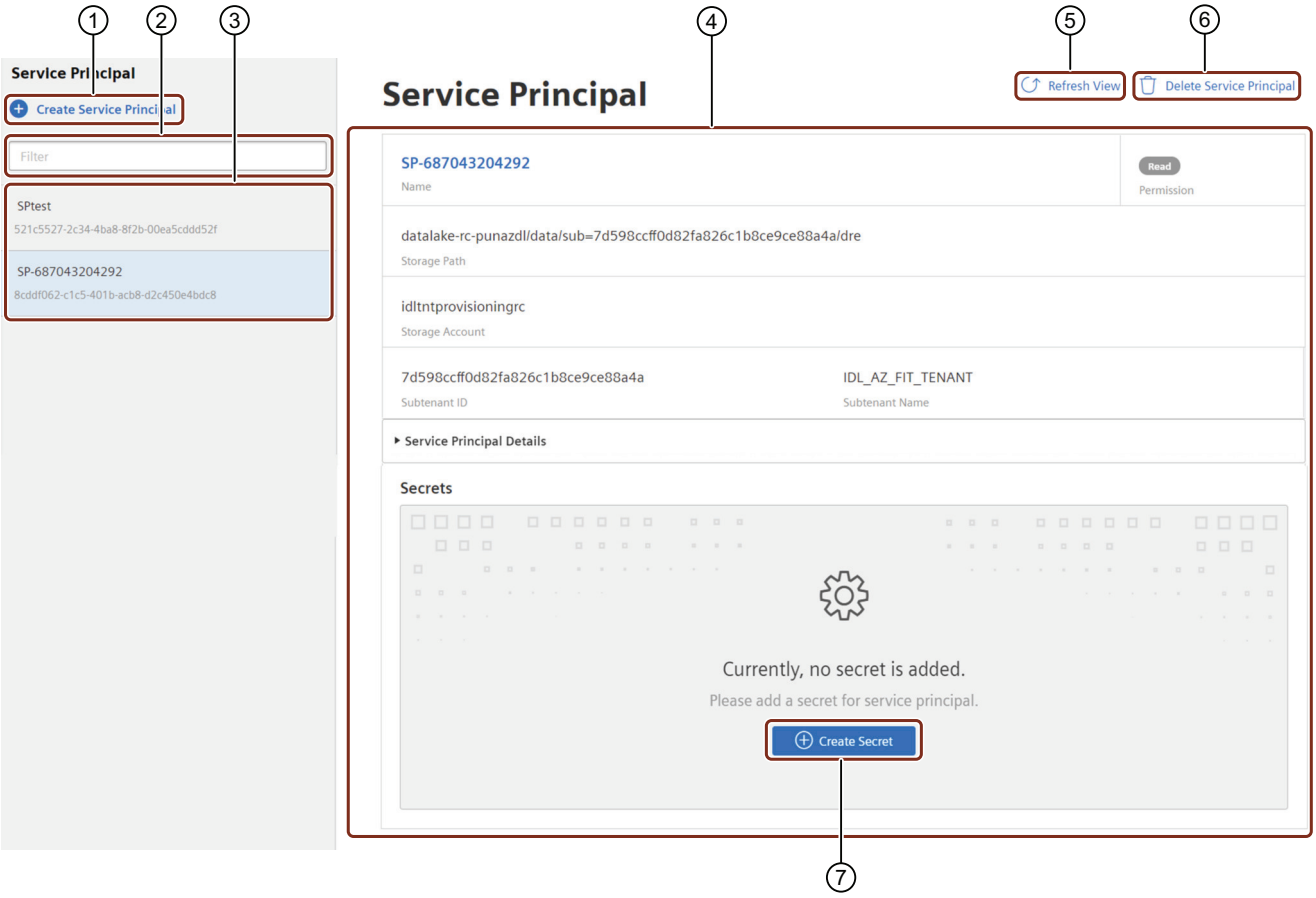
The Cross Account navigation tab is located on the left side of the Integrated Data Lake UI. You can enable the cross account access for the valid AWS accounts only.



1. Enable cross account access
2. Filter or search the created cross account
3. Cross account drop-down list
4. Cross account details
5. Add Accesses
6. Updates the latest changes
7. Edit cross account
8. Delete cross account

Service Principal

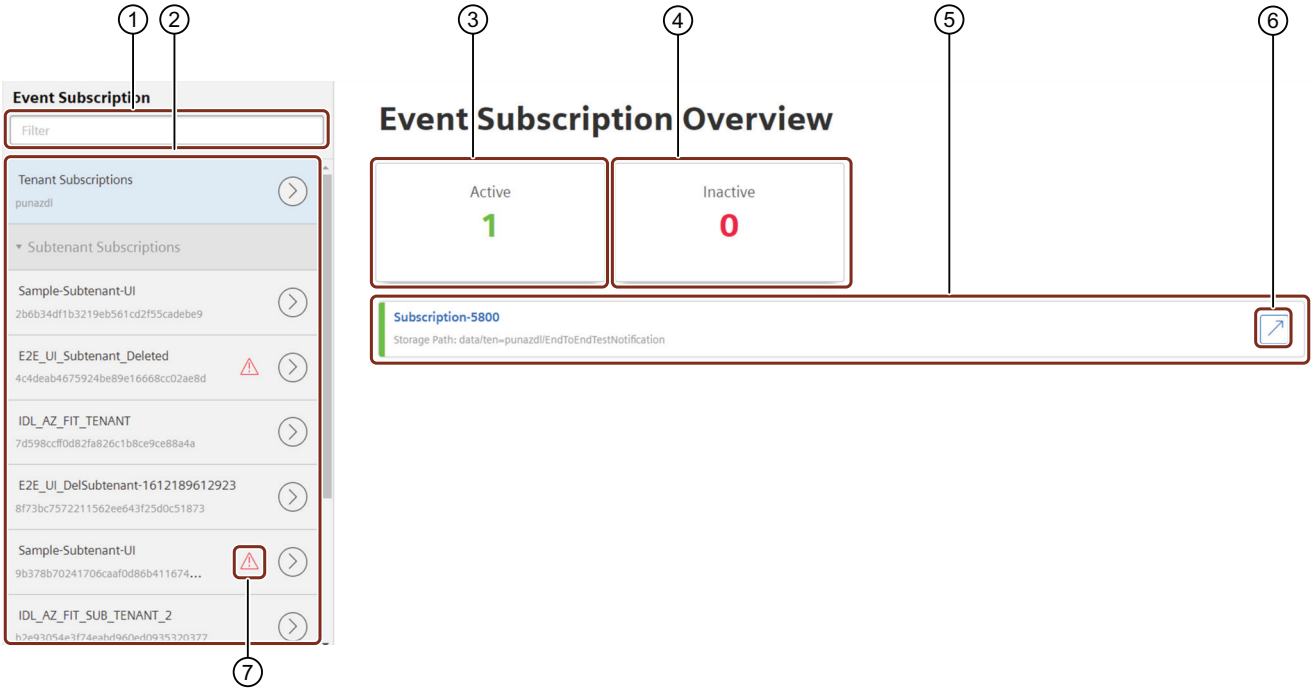
The Service Principal navigation tab is located on the left side of the Integrated Data Lake UI. You can enable the service principal for the valid Azure accounts only.



1. Creates new Service Principal
2. Filter or search the service principal
3. Service Principal drop-down list
4. Service Principal details
5. Updates service principal
6. Deletes service principal
7. Creates the secret for the service principal

Event Subscription

The Event Subscription navigation tab is located on the left side of the Integrated Data Lake UI.



1. Filter or search the event subscriptions
2. Creates new Event Subscription
3. Active Event Subscription list
4. Inactive Event Subscription list
5. List of event subscriptions
6. Redirects to Data Explorer stoarge path
7. Deleted Subtenant

User rights in Integrated Data Lake

4

Integrated Data Lake adopts the user rights from Settings (https://documentation.mindsphere.io/resources/html/settings/en-US/index.html). After activation of Integrated Data Lake, you receive all data lake related read and write permissions from the tenant administrator.

User rights

The user rights depend on the following user roles:

- Administrator
- User
- Viewer

The following table shows the permissions:

Permission	Role		
	Administrator	Standard user	Subtenant user
View all tenant's	✓	✓	✓
Time Series Import	✓	✓	✓
Cross Account	✓		
Event Subscription	✓		
Service Principal	✓		

Time Series Import

5

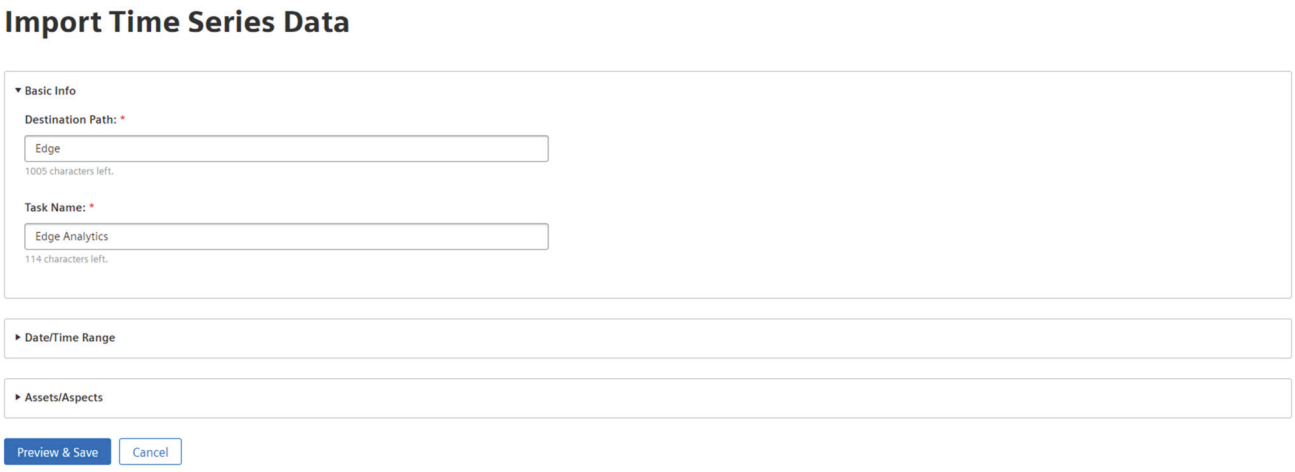
5.1 Importing Time Series data

The "Time Series Import" service allows the authorized users to import the historical IoT time series data to Integrated Data Lake. This enables an on-demand time series upload for analytics and machine learning tools.

Procedure

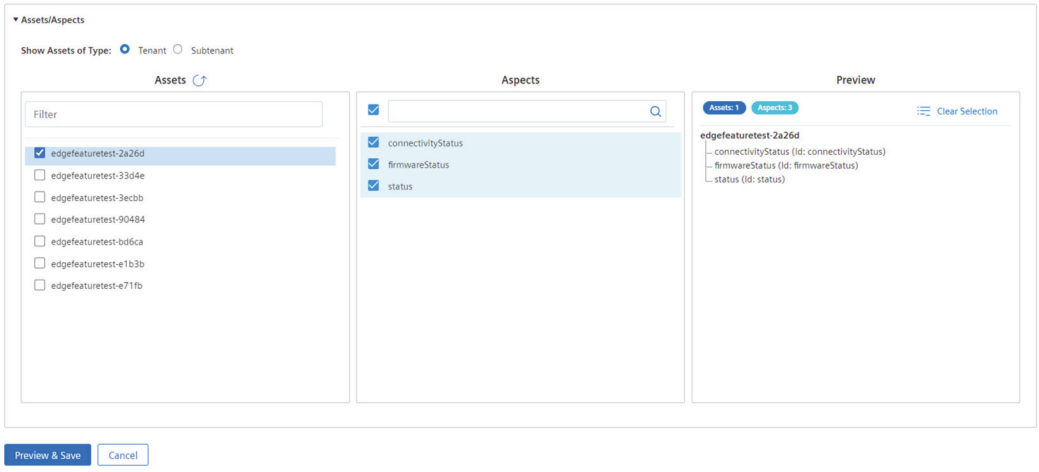
To import IoT time series data into Integrated Data Lake, follow these steps:

1. Click the "Time Series Import" tab.
2. Select the Import Time Series Data icon.
   The "Import Task Details" screen appears.



3. Enter the "Destination Path" and the "Task Name".
4. Expand the "Date/time Range" option to set the date and time.
5. Click "Date/Time Range" to select the date and time range and click "Set".

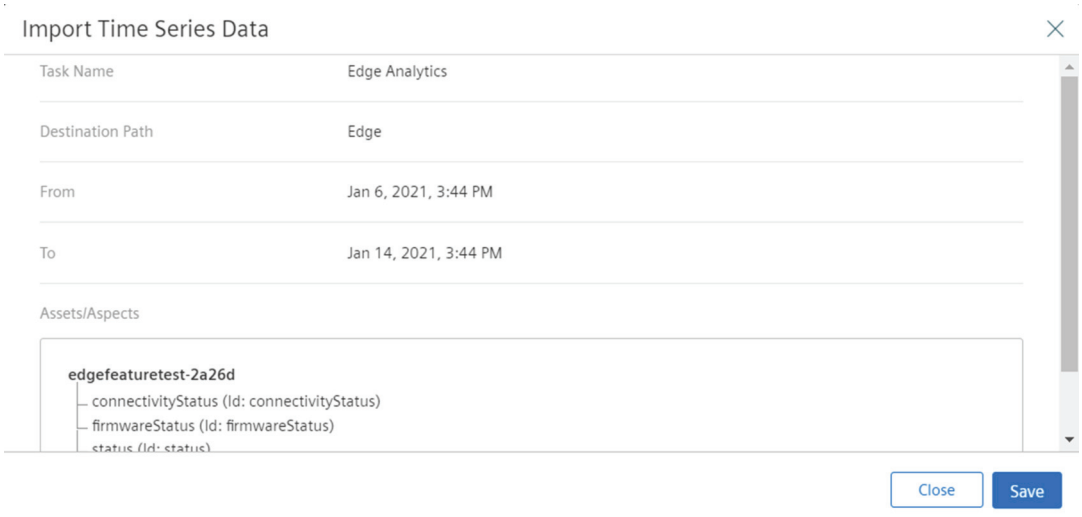
6. Expand the "Assets/Aspects" option to select the assets and corresponding aspects.



Note

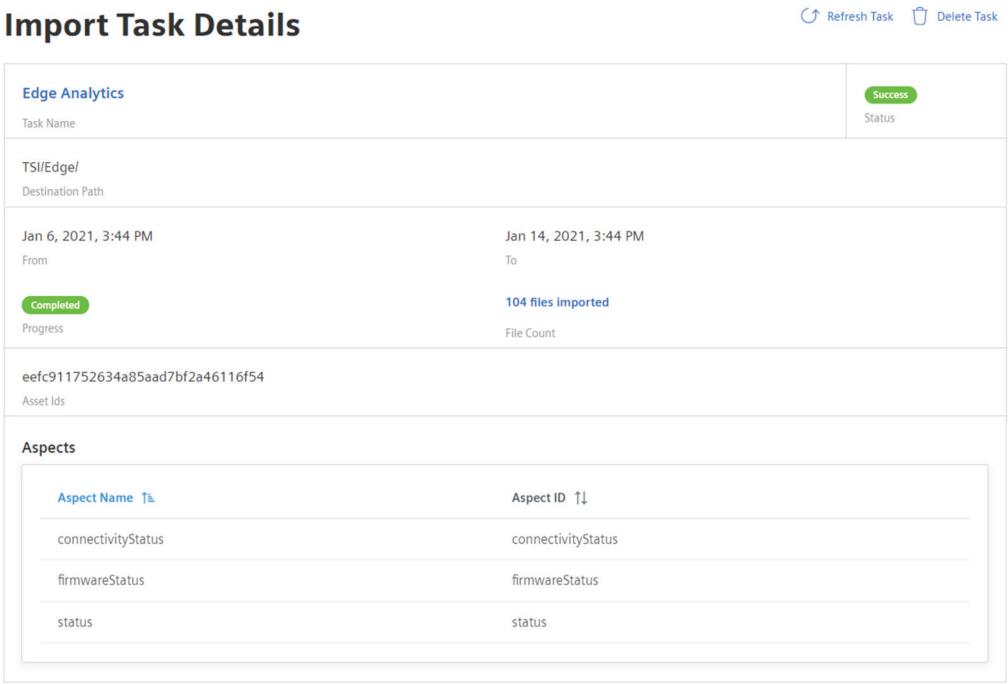
You can click ↻ to refresh the asset list.

7. Select tenant or subtenant to get the respective assets from the "Assets" list and corresponding aspects from the "Aspects" list.
8. Click "Preview & Save".



9. Click "Save".

10. The "Import Task Details" status "Success" screen appears.



11. Click "Refresh Task" button to get the latest status of the import task.
    Click "Delete Task" button to delete the task from the "Task Schedule" pane.

Note

- Integrated Data Lake application will not allow the user to import the duplicate data based on "Date/Time Range", "Assets" and "Aspects" information. The application will provide the information about the existing bulk imports with the "Asset ID".
- A user is not allowed to select more than 3 months data to import.
- The Time Series data (parquet files) transferred to cold storage (typically within 7 days) and will be available in the Data Lake.
- There is a change in the namespace of the Time Series data imported into Integrated Data Lake.

Result

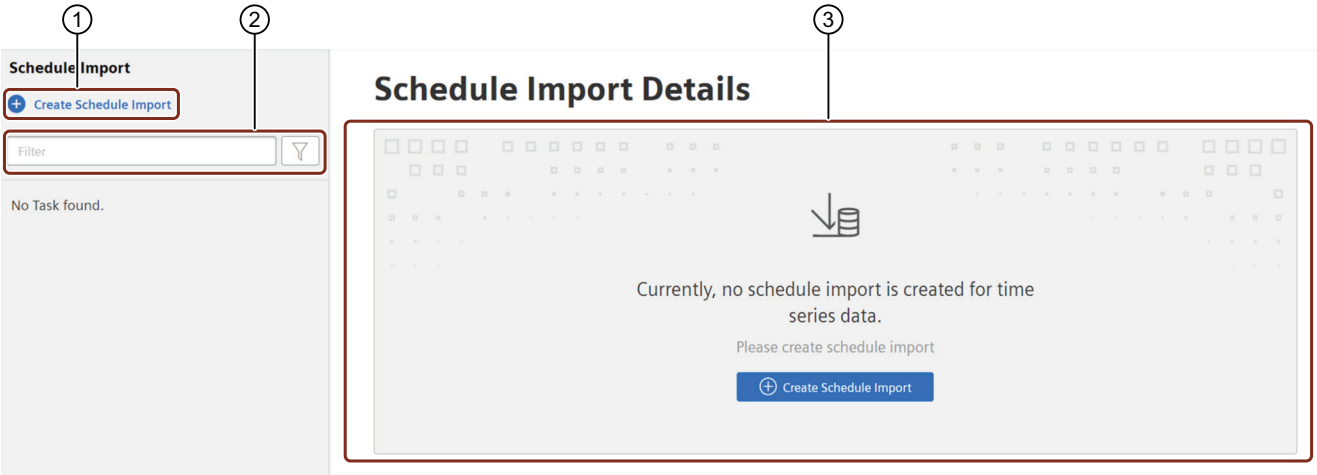
The historical IoT time series data will be imported to Integrated Data Lake and success status will be displayed with the number of file count.

5.2 Importing the scheduled Time Series data

In Time Series Import, you can schedule the import data. The import data can be scheduled with the frequency of "Daily", "Weekly" or "Monthly. Based on the frequency selection, your data will be imported as per the scheduled task.

User Interface

The "Schedule Import" is accessible from the Time Series Import tab.



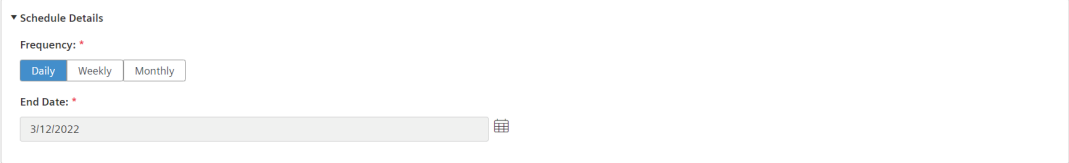
① Create Schedule Import jobs
② Filter or search the list of Schedule Import jobs
③ Displays the selected Schedule Import Details

Procedure

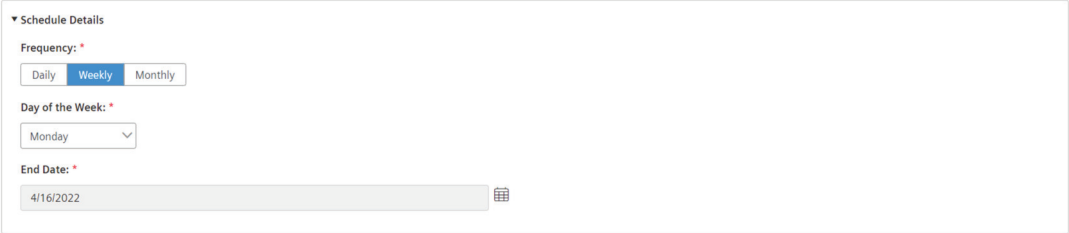
To create the scheduled IoT time series job, follow these steps:

1. In "Time Series Import" tab, select "Schedule Import".
2. Click "Create Schedule Import" to schedule the Time Series data import.
3. In "Basic Info"section, enter the details.

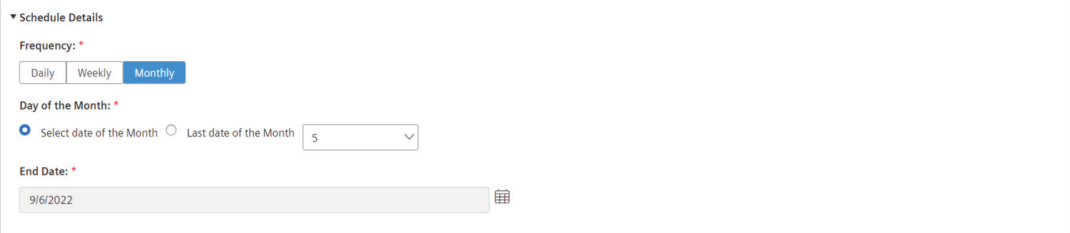
4. In "Schedule Details" section, choose any one "Frequency" to schedule the import:
   - **Daily**: This will schedule the job to run every day.



   - **Weekly**: This will schedule the job to run once in a week on selected day of the week. Ensure the end date of the job is far enough to execute the at least 1 execution of the job.



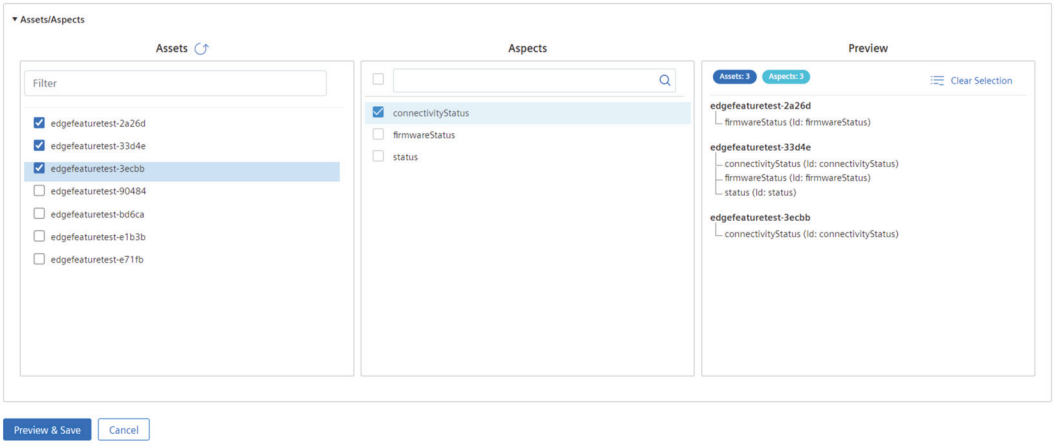
   - **Monthly**: This will schedule the job to run once every month. Ensure the end date of the job is far enough to execute the at least 1 execution on the job. There are two option available in monthly schedule:
     * On the selected day of the month
     * On the last day of every month.



Note

Job runs on the scheduled time and fetch the data one day prior to the execution date.

5. In "Assets/Aspects" section, select the Asset and Aspect.



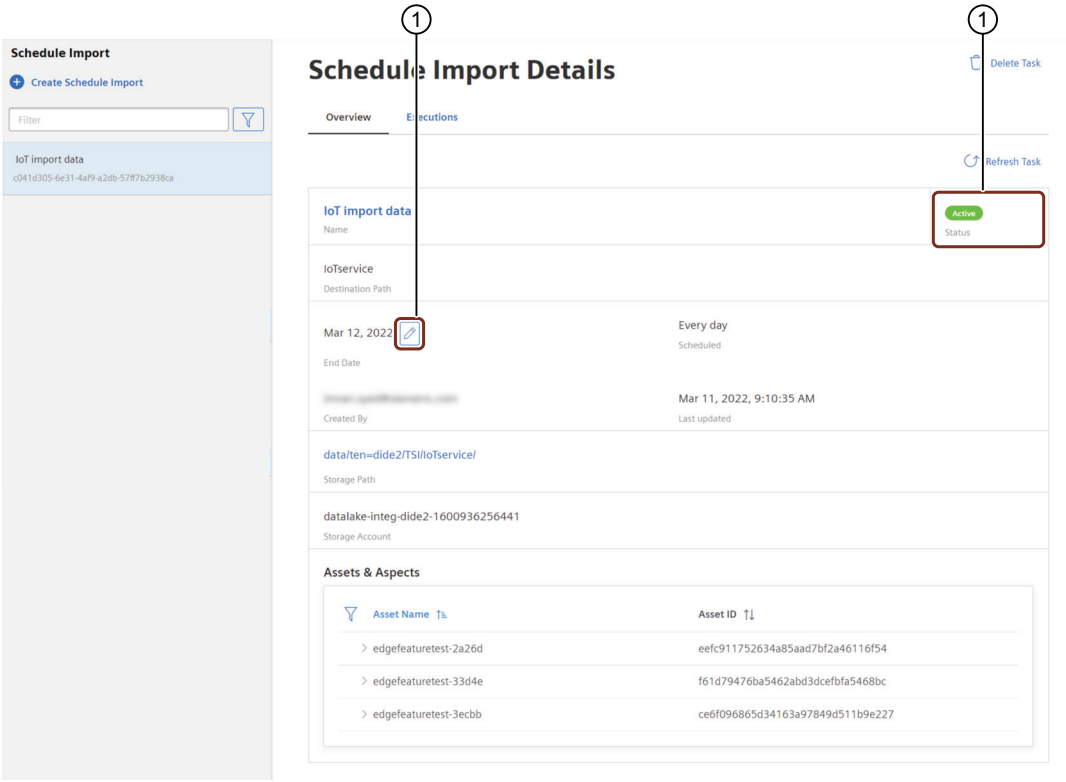
Note

Maximum 5 asset and aspect combination selection is allowed.

6. Click "Preview and Save".
7. Check the job preview and click "Save".

Result

The scheduled Time Series job is created successfully.



① Edit end date in active state

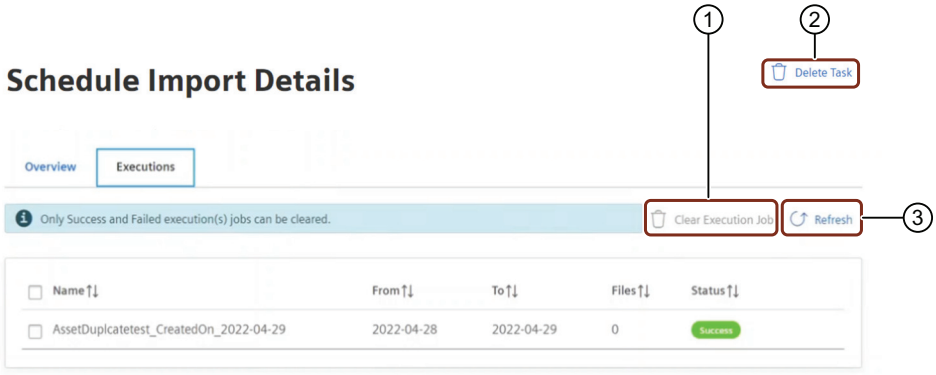
② Status of the related job:

- Active – It is the status of the job after creation. In "Active" status, you can extend "End Date" of the job.
- Inactive – It is the status of the job if any of asset/aspect from the job is deleted. You have to delete the job and create new one.
- Expired – It is the status of the job after 'End Date' is met. You have to delete the job and create new after it is expired.

Note

Maximum limit to schedule the jobs is 10 for all users (tenant + subtenant).

In "Schedule Import Details", click "Executions" tab for the outputs.



① Clears the selected execution jobs
② Deletes the selected tasks
③ Updates the execution tab

Note

Only "Success" and "Failed execution" jobs can be cleared.

Cross Account Access

6

6.1 Enabling Cross Account access

You (tenant administrator) can enable or delete the cross account access for the AWS account. You can also modify the relevant accesses to the specific prefixes for the folder of Integrated Data Lake. For more information refer User rights in Integrated Data Lake (Page 19).

Note

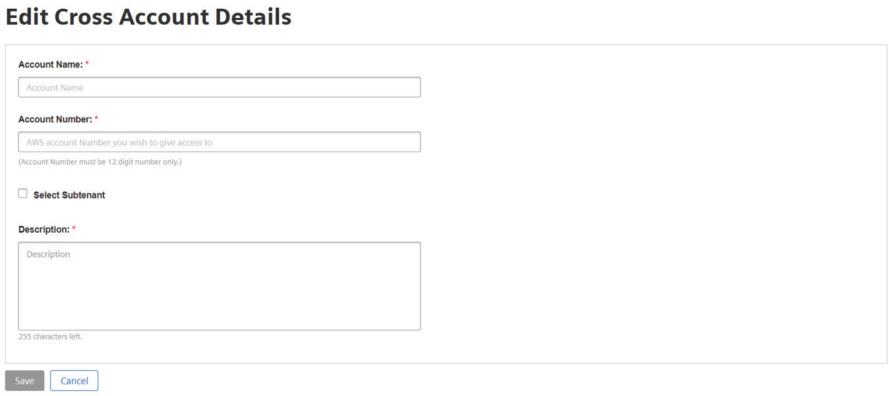
Region deviation

Cross account access is applicable for Region Europe 1.

Procedure

To provide cross account access for the AWS account, follow these steps:

1. Click on "Cross Account" tab.
2. In the "Cross Account Access" window, click on Enable Cross A/c Access icon.
   The "Cross Account Access" screen appears.

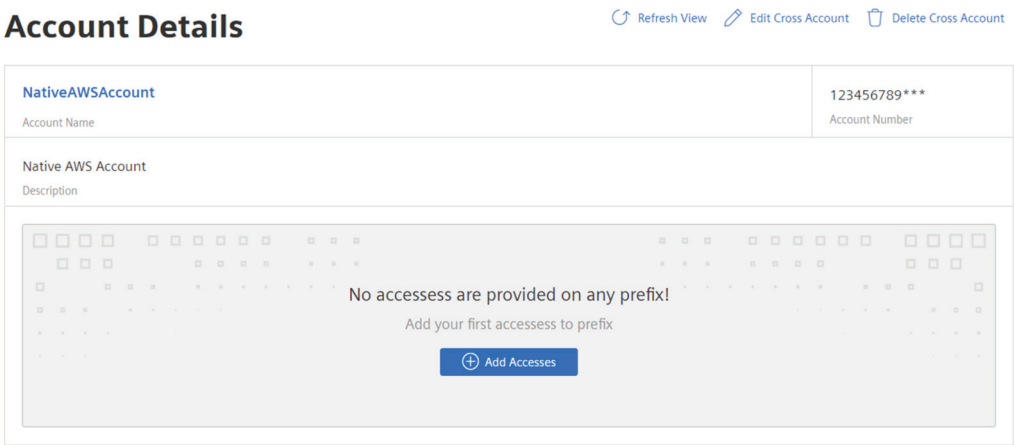


3. Enter the "Account Name".
4. Enter the valid AWS "Account Number" and "Description" .

   Note
   - Account name field is user defined.
   - A user should have a valid 12 digit AWS account to enable the "Cross Account".
   - Description field is mandatory to enable the cross account.

5. The subtenant option can be selected, if the AWS cross account access is for a subtenant.
6. From the "Select Subtenant" drop-down list, "Choose a subtenant" as per the requirements.

7. Click "Enable" button.
   The "Account Details" enabled screen appears.



8. Click "Edit Cross Account" button to edit the cross account.
9. Click "Delete Cross Account" button to delete the cross account.

Result

Cross Account access for a specific AWS account will be created. You can now create Cross Account Accesses to give access to specific folders.

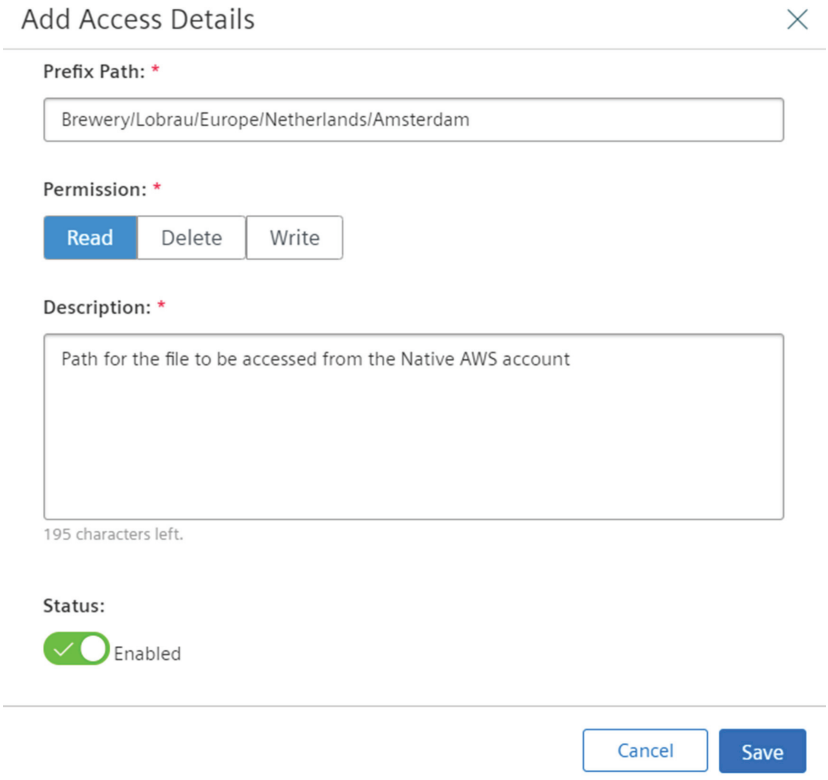
6.2 Managing Cross Account accesses

You can manage the "Cross Accounts accesses" by providing access to the specific folders based on the requirement. You can grant the permission to read, write or delete access as per the user requirements.

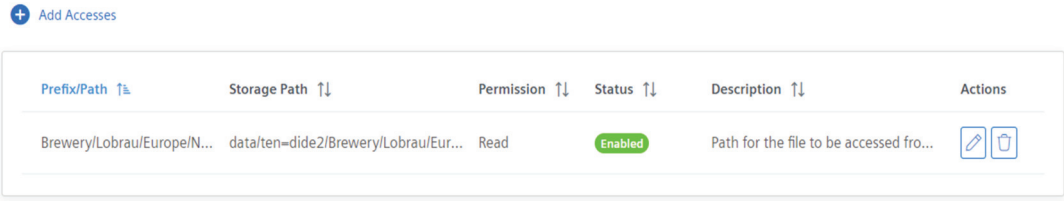
Procedure

To manage the cross accounts accesses, follow these steps:

1. In the "Account Details" screen, click the "Add Accesses" button.
   The "Add Access Details" screen appears.



2. Enter the "Prefix Path" to specific account access path on which the accesses to be provided.
3. Select the "Access Permission" (Read, Delete or Write).
4. Enter the "Description" regarding the access details.
5. Click "Status" button to enable or disable the "Access Permission".
6. Click "Save".
   The account access details will added to the list.



7. Click the icon to edit the account access permission.
8. Click the icon to delete the account access.

Note

- A maximum of 5 prefixes can be enabled across one or multiple cross accounts.
- A maximum 10 cross account acceses can be created in disabled state.
- Cross Account is set to be accessed from the native AWS tools.

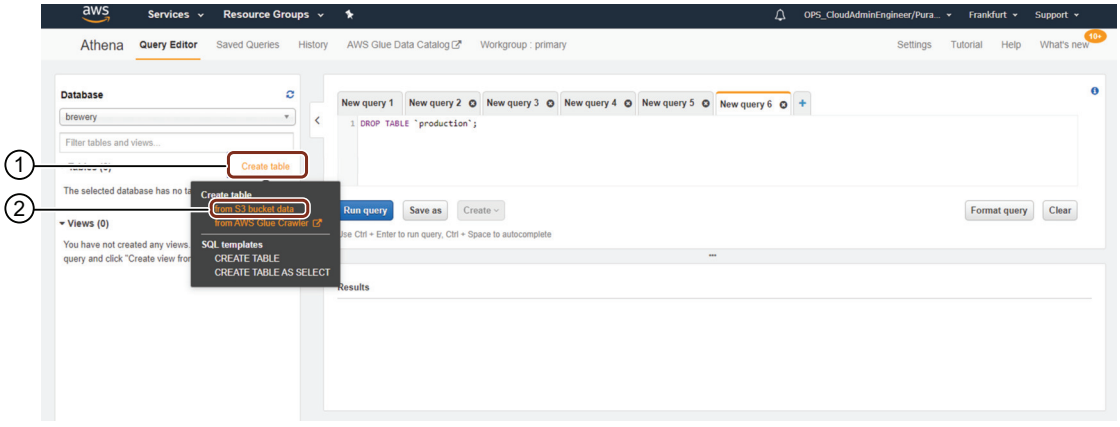
6.3 Example to integrate AWS Athena tool with Integrated Data Lake

This is an example to integrate AWS Athena tool with Integrated Data Lake after enabling cross account access.

Procedure to integrate AWS Athena tool

To integrate AWS Athena tool with Integrated Data Lake, follow these steps:

1. Open AWS Athena tool.
2. Create the table in Athena



① Click to create table

② Choose the option "from S3 bucket data"

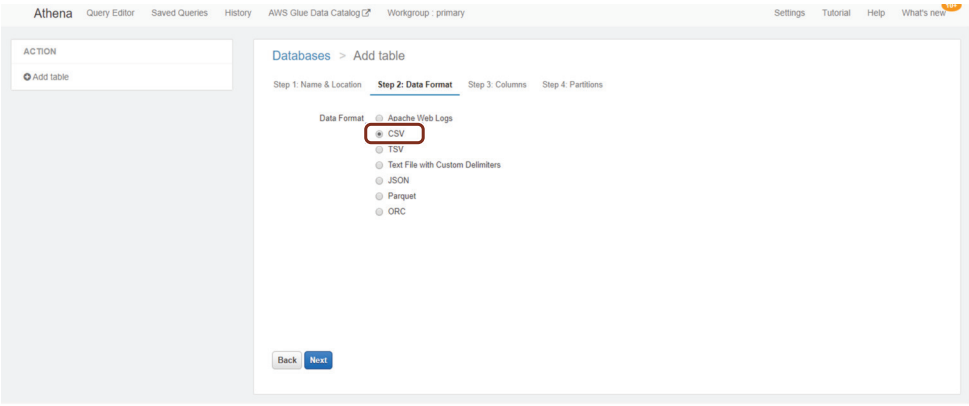
3. Add the details related to database, table and input data.
   - Database: Select the existing database or create a new by selecting "Create a new database".
   - Table Name: Enter the table name.
   - Location of Input Data Set: Path provided to access while configuring the cross account accesses.

   Note

   Location of input data set would look like "s3://+<<storage account>>+<<storage path>> and path should end with '/'.

   For example:
   - Storage Account = datalake-prod-a-starter-576071197214
   - Storage Path = data/ten=starter/Brewery/Lobrau/Europe/Netherlands/Amsterdam
   - Location of Input Data Set = s3://datalake-prod-a-starter-576071197214/data/ten=starter/Brewery/Lobrau/Europe/Netherlands/Amsterdam/

4. Click "Next".
5. Select "Data Format".



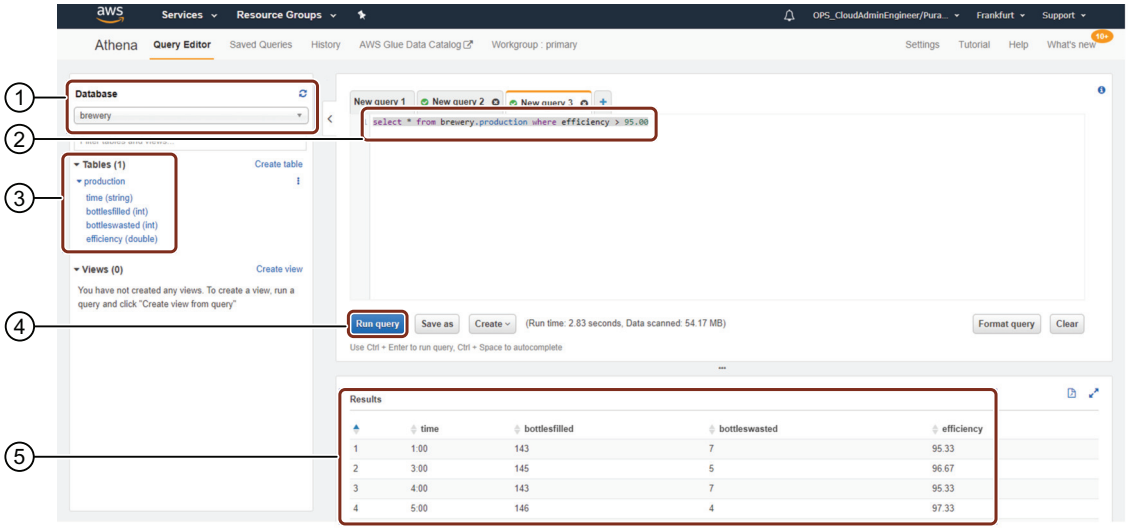
   Note

   File uploaded in the Data Explorer should be .csv file.

6. Add column name and column type corresponding to the data in the .csv file.
7. Click "Next".
8. Click "Create Table".

Result

The table is created successfully and now you can query the data in AWS Athena. To query the data, follow the below steps:



① Select the table from the database
② Specify the table name in the query as `database.tablename`
③ Columns of the table
④ Click to run the query
⑤ Query result

Service Principal

7

7.1 Creating Service Principal

You can create Service Principal to configure the native tools of Microsoft Azure to access the files/objects in Integrated Data Lake. Using Service Principal, you can call Azure Data Lake Storage (ADLS) APIs directly. You can generate Service Principal for each tenant with a specific access to your own data only.

The Service Principal's ClientId/Secret can be managed as per MindSphere guidelines.

Note

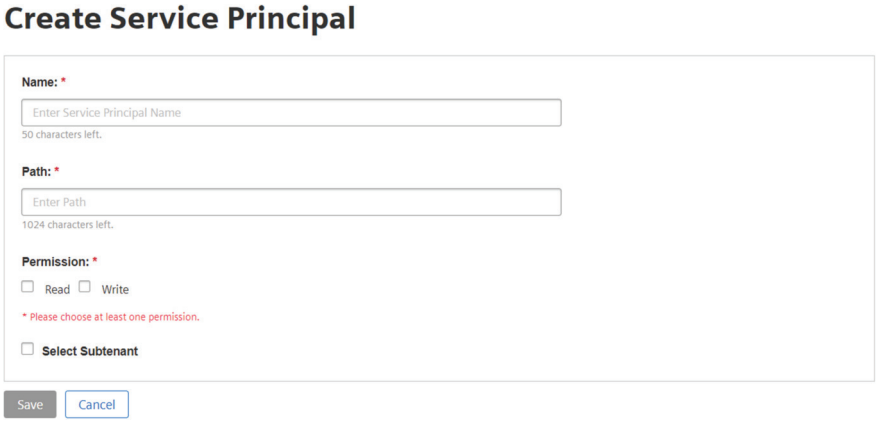
Region deviation

Service Principal is applicable for Region Europe 2.

Procedure

To configure the native tools of Microsoft Azure with Service Principal, follow these steps:

1. Click on "Service Principal" tab.
2. In the "Service Principal" window, click on Create Service Principal icon.

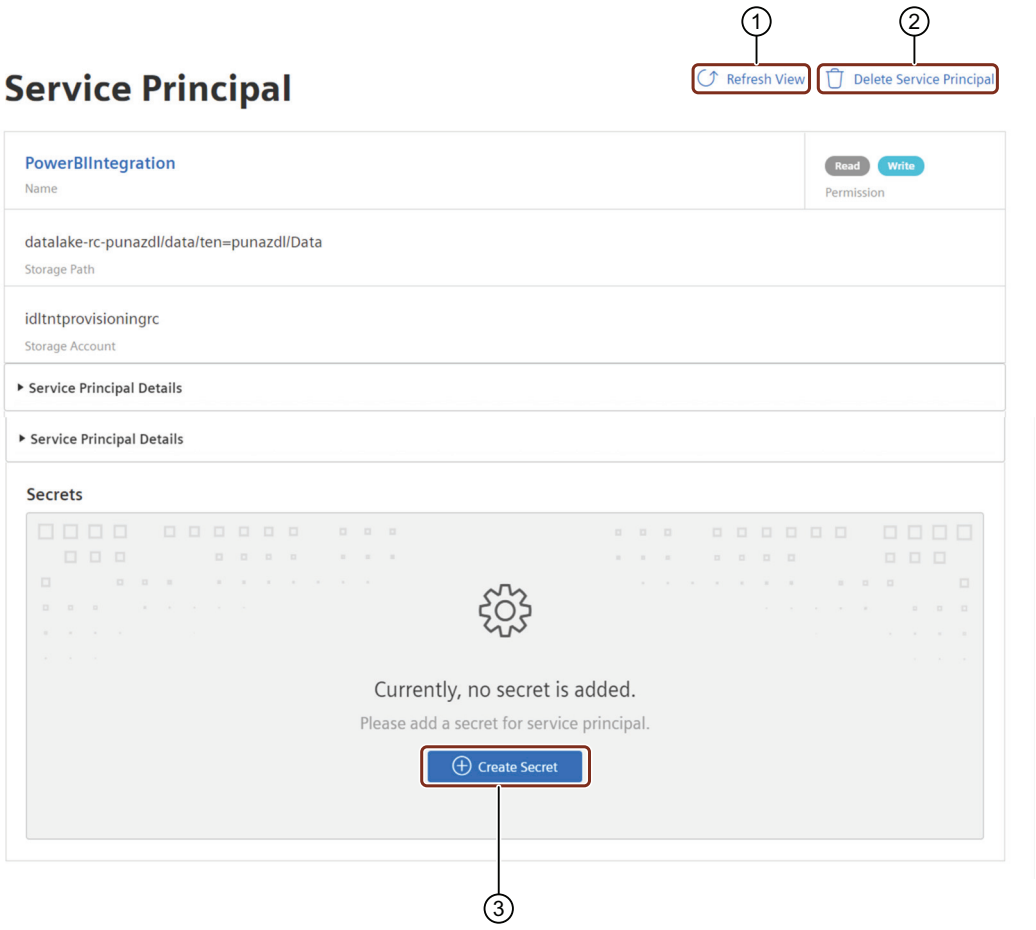


   – Enter the "Name".
   – Enter the path for "Service Principal".
   – Specify the access permission (Read, Write or Read and Write)

   Note

   The subtenant option can be selected, if the Service Principal is for a subtenant.

3. Click Create.



① Refresh to update Service Principal
② Deletes the Service Principal
③ Create the seceret to the Service Principal

Result

You can now create a secret for Service Principal to access the files/objects from the specified path.

Note

Maximum 5 Service Principals can be created in Integrated Data Lake.

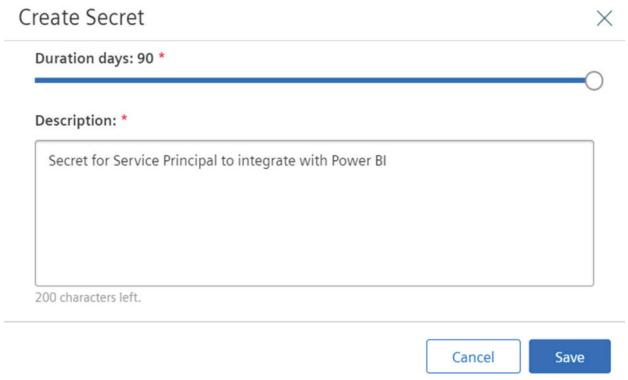
7.2 Managing Service Principal

You can manage the "Service Principal" by adding a secret to configure with Microsoft Azure Native tools.

Procedure

To manage the service principal, follow these steps:

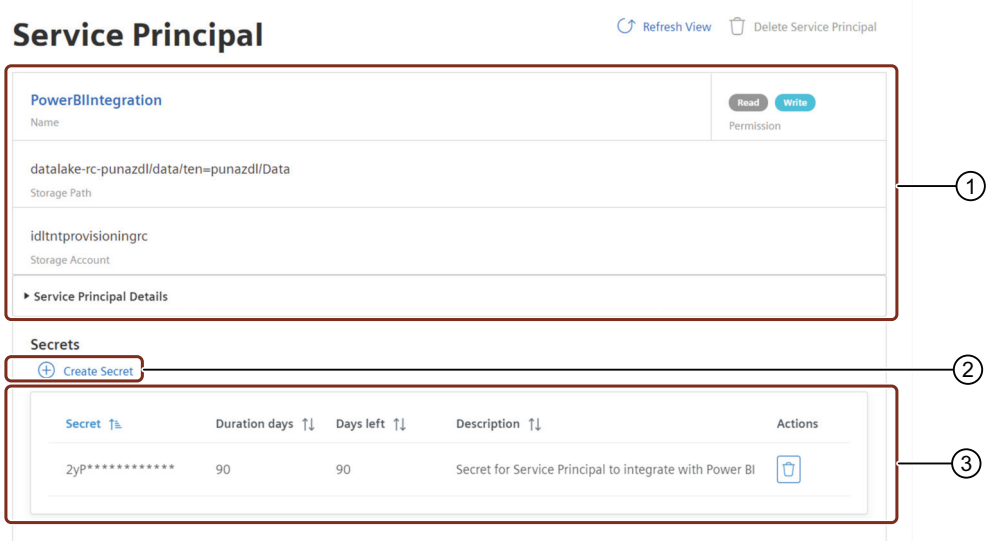
1. In the "Service Principal" screen, click the "Add Secret".



   – Select the duration from minimum 1 to maximum 90 day.
   – Enter the description.
2. Click Save.
3. Click to copy the generated secret and confirm the secret is copied.
4. Click Close.

Result

The secret is generated for Service Principal successfully.



① Additional details of Service Principal
② Adds a secret to the Service Principal
③ Service Principal details

You can now configure the Service Principal with Microsoft Azure Native tools.

Note

- Maximum 2 Secrets can be added for each Service Principal in Integrated Data Lake.
- If the Secret is added, the Service Principal cannot be deleted.
- The duration of the secret is left with 30 to 10 days, then symbol will appear.
- The duration of the secret is left with 10 to 1 days, then symbol will appear.

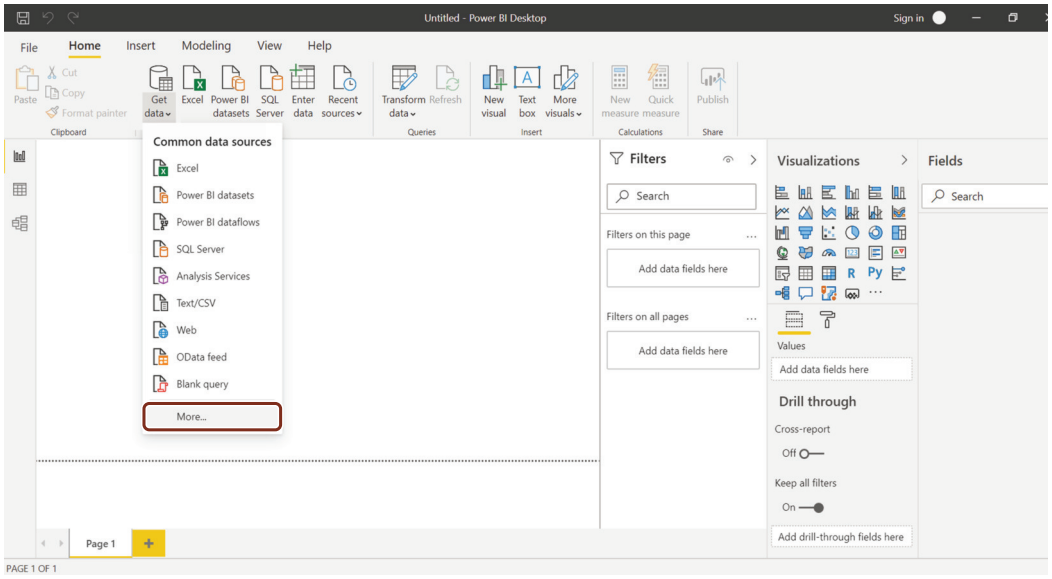
7.3 Example to configure Azure Power BI tool with Integrated Data Lake

This is an example to configure Azure Power BI tool with Integrated Data Lake after creating service principal.

Procedure to configure Azure Power BI tool

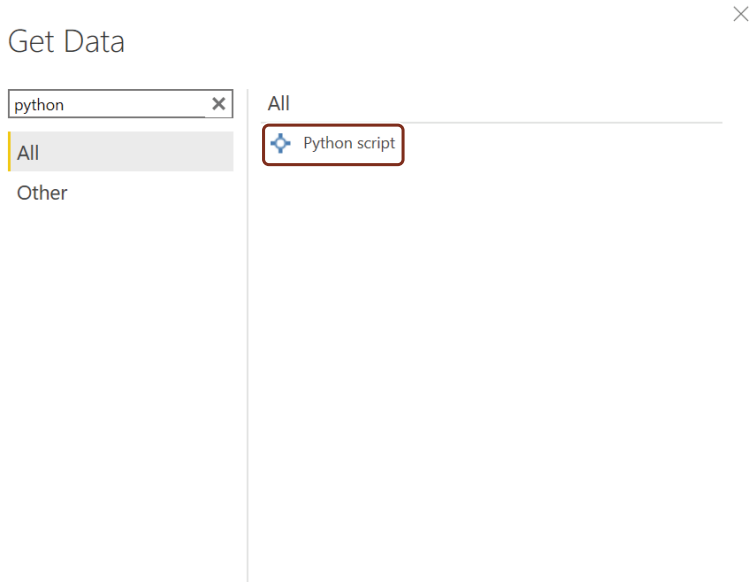
To configure Azure Power BI tool with Integrated Data Lake, follow these steps:

1. Open Power BI tool.



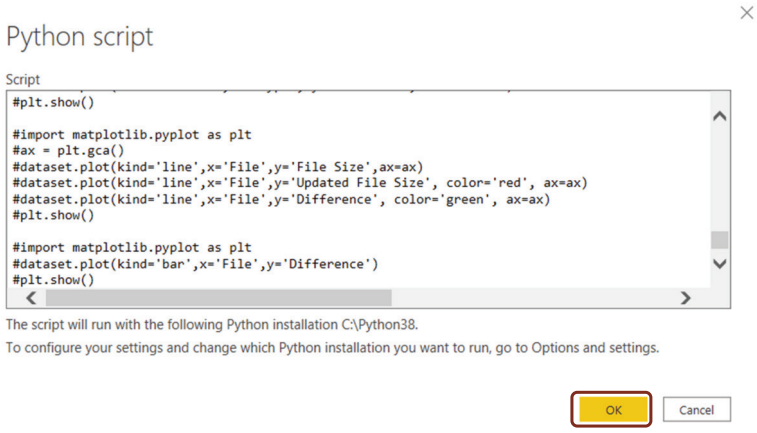
2. In Power BI tool, open "Get data" and select "More".

3. In Get Data screen, search for "python" and choose "Python script".



4. Click Connect.

5. Configure the python script connector



– Sample Python Script to be used as giveb below:

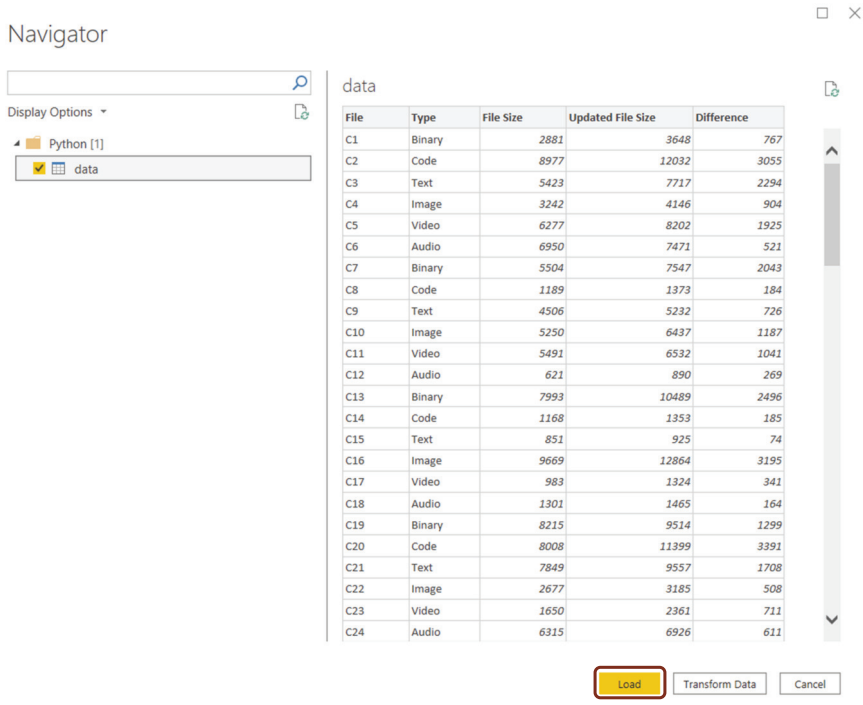
```
from pprint import pprint
import os, uuid, sys
from azure.identity import ClientSecretCredential
from azure.storage.filedatalake import DataLakeServiceClient
import pandas as pd
import matplotlib.pyplot as plt
from io import StringIO
def load_file_in_powerbi(file_system_client):
    try:
        directory_client =
file_system_client.get_directory_client(sampledata_directory)
        file_client =
directory_client.get_file_client(sampledata_filename)
        download = file_client.download_file()
        downloaded_bytes = download.readall()
        bytes_string = str(downloaded_bytes, 'utf-8')
        data_stringio = StringIO(bytes_string)
        global data
        data = pd.read_csv(data_stringio)
        print(data)
    except Exception as e:
        print(e)
def initialize_storage_account_ad(storage_account_name,
client_id, client_secret, tenant_id):
    try:
        global datalake_service_client
        credential = ClientSecretCredential(tenant_id,
client_id, client_secret)
        datalake_service_client = DataLakeServiceClient(
            account_url="{}://
{}.dfs.core.windows.net".format("https", storage_account_name),
credential=credential)
    except Exception as e:
        print(e)
def get_file_system_client(file_system_name):
    try:
        file_system_client =
datalake_service_client.get_file_system_client(file_system_name
)
        return file_system_client;
    except Exception as e:
        print(e)
def get_directory_client(file_system_client, directory_path):
    try:
        directory_client =
file_system_client.get_directory_client(directory_path)
        return directory_client;
    except Exception as e:
        print(e)
def upload_file_to_directory(directory_client, src_file_path,
file_name):
```

```
    try:
        file_client = directory_client.create_file(file_name)
        local_file = open(src_file_path, 'r')
        file_contents = local_file.read()
        file_client.append_data(data=file_contents, offset=0,
length=len(file_contents))
        file_client.flush_data(len(file_contents))
    except Exception as e:
        print(e)
def list_directory_contents(file_system_client, directory_path):
    try:
        paths = file_system_client.get_paths(directory_path)
        return paths
    except Exception as e:
        print(e)
def download_file_from_directory(directory_client, file_name):
    try:
        local_file = open(file_name, 'wb')
       file_client = directory_client.get_file_client(file_name)
        download = file_client.download_file()
        downloaded_bytes = download.readall()
        # bytes_string = str(downloaded_bytes,'utf-8')
        # data_stringio = StringIO(bytes_string)
        # data = pd.read_csv(data_stringio)
        # completedData = data.fillna(method='backfill',
inplace=False)
        # data["completedValues"] =  completedData["SMI missing
values"]
        # pprint(data)
        # pprint(vars(data))
        # pprint(data.head())
        # data.plot(kind='bar', x='Day', y='completedValues',
color='red')
        # plt.show()
        local_file.write(downloaded_bytes)
        local_file.close()
    except Exception as e:
        print(e)
def download_hierarchy_directory(file_system_client,
directory_path):
    try:
        directory_client =
get_directory_client(file_system_client, directory_path)
       paths_list = list_directory_contents(file_system_client,
directory_path)
        for path in paths_list:
            if path.is_directory == True:
               download_hierarchy_directory(file_system_client,
path.name)
            else:
                # print(path.name + '\n')
                # pprint(vars(path))
```

```
                path_split = path.name.rsplit("/", 2)
               if path_split[1] == directory_path.rsplit("/", 1)
[1]:
                    file_name = path_split[2]
                    print(file_name + '\n')
                download_file_from_directory(directory_client,
file_name);
    except Exception as e:
        print(e)
if __name__ == "__main__":
    tenant_id = "xxxxxxxx-xxxx-xxxx-xxxx-xxxxxxxxxxxx"
    client_id = "xxxxxxxx-xxxx-xxxx-xxxx-xxxxxxxxxxxx"
    client_secret = "xxxxxxxx-xxxx-xxxx-xxxx-xxxxxxxxxxxx"
    storage_account_name = "idltntprovisioningrc"
    file_system_name = "datalake-rc-punazdl"
    directory_path = "data/ten=punazdl/powerbi"
    sampledata_directory = "data/ten=punazdl/powerbi/sample-
data"
    sampledata_filename = "file-size-upload.csv"
    initialize_storage_account_ad(storage_account_name,
client_id, client_secret, tenant_id)
    file_system_client =
get_file_system_client(file_system_name)
    paths = list_directory_contents(file_system_client,
directory_path)
    directory_client = get_directory_client(file_system_client,
directory_path)
    load_file_in_powerbi(file_system_client)
```

Application ID	Location
tenant_id	Tenant ID will be available on the Service Principle page in Data Lake.
client_id	This is the application ID that can be copied from the Service Principle page in Data Lake.
client_secret	This is the Service Principle secret that was generated and copied.
sampledata_directory	Path on which Service Principle is generated.
sampledata_filename	File to be pulled into Azure Power BI.

6. Click OK to load data into Azure Power BI.

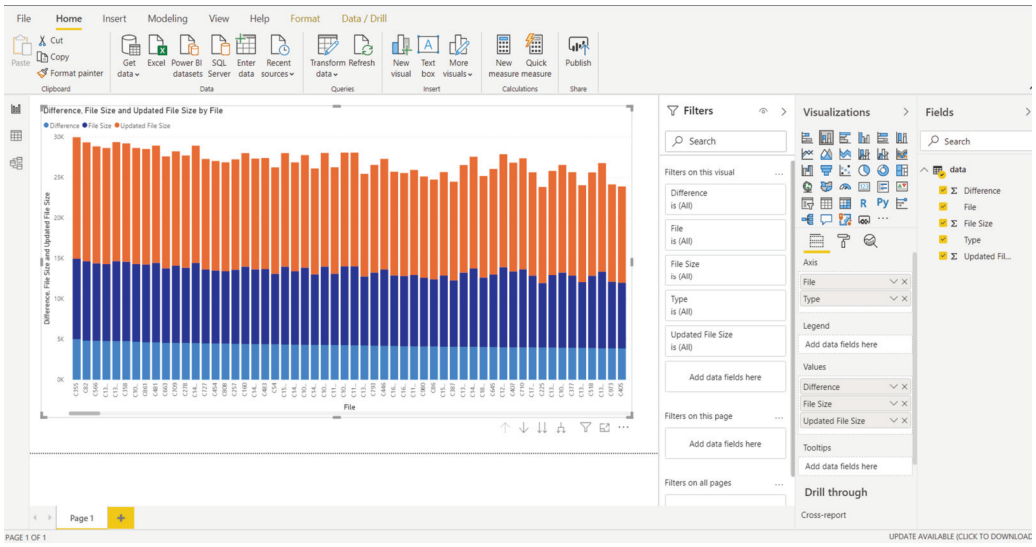


File	Type	File Size	Updated File Size	Difference
C1	Binary	2881	3648	767
C2	Code	8977	12032	3055
C3	Text	5423	7717	2294
C4	Image	3242	4146	904
C5	Video	6277	8202	1925
C6	Audio	6950	7471	521
C7	Binary	5504	7547	2043
C8	Code	1189	1373	184
C9	Text	4506	5232	726
C10	Image	5250	6437	1187
C11	Video	5491	6532	1041
C12	Audio	621	890	269
C13	Binary	7993	10489	2496
C14	Code	1168	1353	185
C15	Text	851	925	74
C16	Image	9669	12864	3195
C17	Video	983	1324	341
C18	Audio	1301	1465	164
C19	Binary	8215	9514	1299
C20	Code	8008	11399	3391
C21	Text	7849	9557	1708
C22	Image	2677	3185	508
C23	Video	1650	2361	711
C24	Audio	6315	6926	611

7. Click Load.

Result

The data is successfully loaded in Azure Power BI. You can now visualize your data in Azure Power BI.



Data Explorer

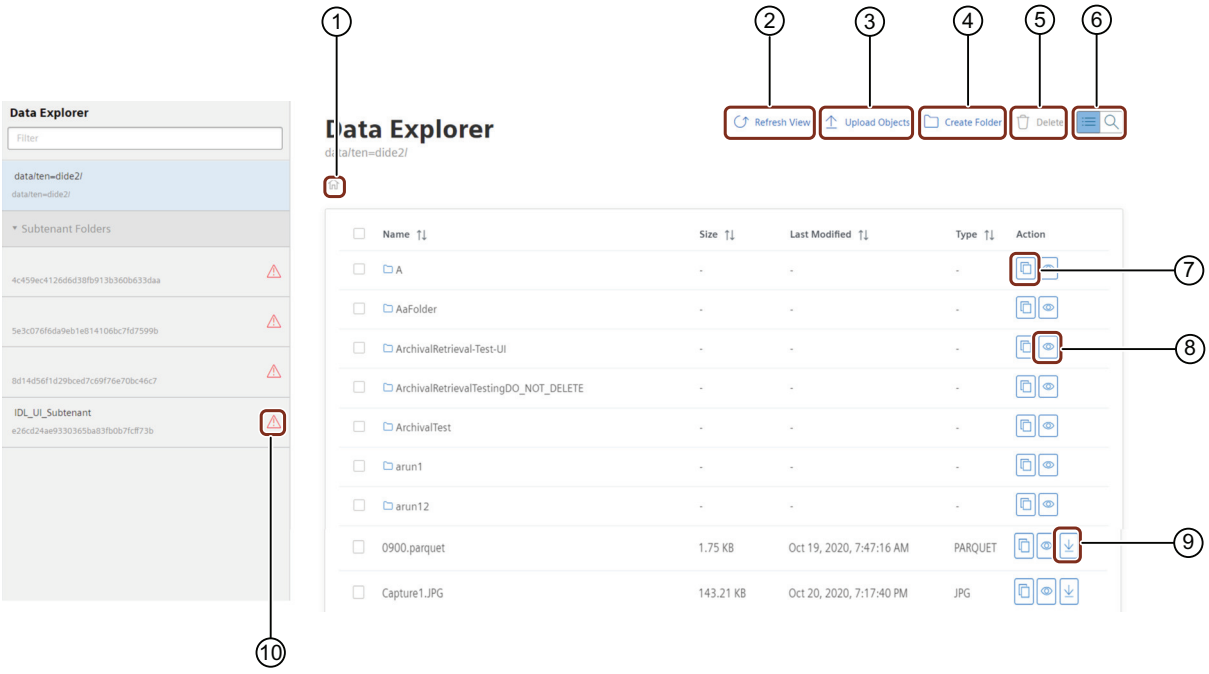
8

8.1 Exploring Data Explorer

Data Explorer feature enables you to explore the folders created by you and other subtenants. From the application, you can use the functionality of breadcrumb, upload, download, create the folder, sort and pagination.

You can view the data in the folders and objects available in Integrated Data Lake. The following screenshot displays the folders and objects in Data Explorer screen:

Tenant admin view

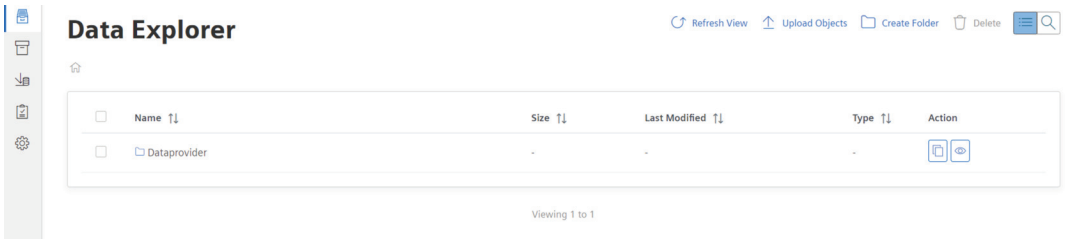


① Breadcrumb
② Refresh the page to get the latest data
③ Upload the objects
④ Creates the new folder
⑤ Delete the objects
⑥ Search objects/files with object name, metadata or both
⑦ Copy the object path
⑧ Displays Object View details and adds the metadata tags to the folder and object
⑨ Download the object
⑩ Deleted subtenant symbol

Note

- The subtenant users are permitted to access only those folders created by themselves, but not the folders created by other users.
- Tenant admin can see the folder/files created by subtenant user and access them from the Data Explorer folder pane.
- If the subtenant is deleted, uploading objects and adding metadata tags are not allowed.

Subtenant user view



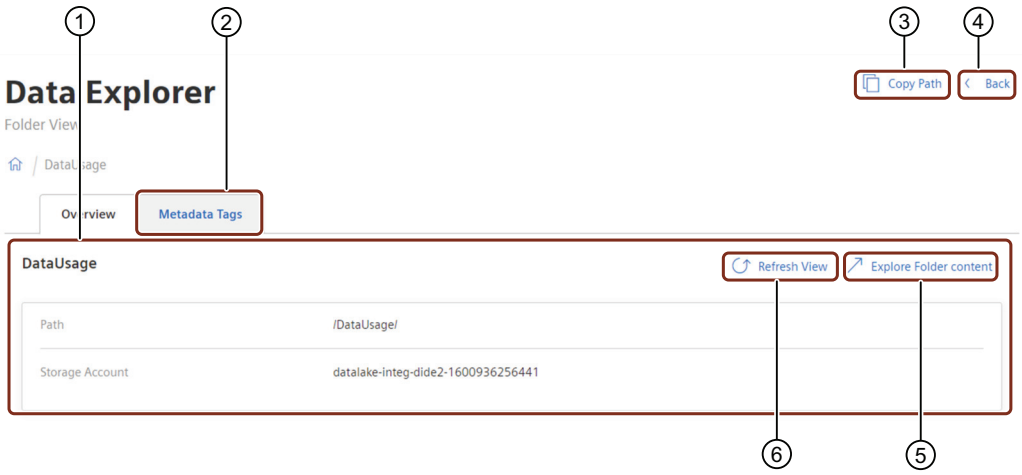
Note

Data Explorer folder pane is not accessible for the subtenant user.

8.2 Creating folder with metadata tags in Data Explorer

In Data Explorer, you can create multiple folders. After creating the folder, you can view the folder details and add the metadata tags. The folders are quickly accessible by using metadata tags. You can view the folder details by clicking and add the metadata tags to the folder.

View folder details

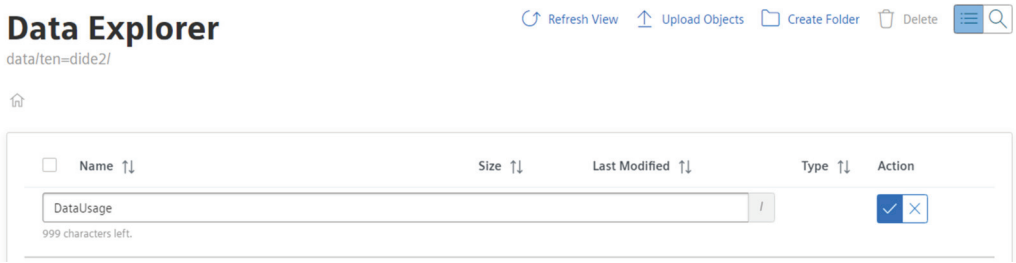


① Folder overview details
② Add the metadata tags to the folder
③ Copy the folder path
④ Navigates to the parent folder or Data Explorer
⑤ Navigates to the folder
⑥ Refresh to update the folder details

Procedure

To create the folder with metadata tag, follow these steps:

1. In Data Explorer, click "Create Folder".



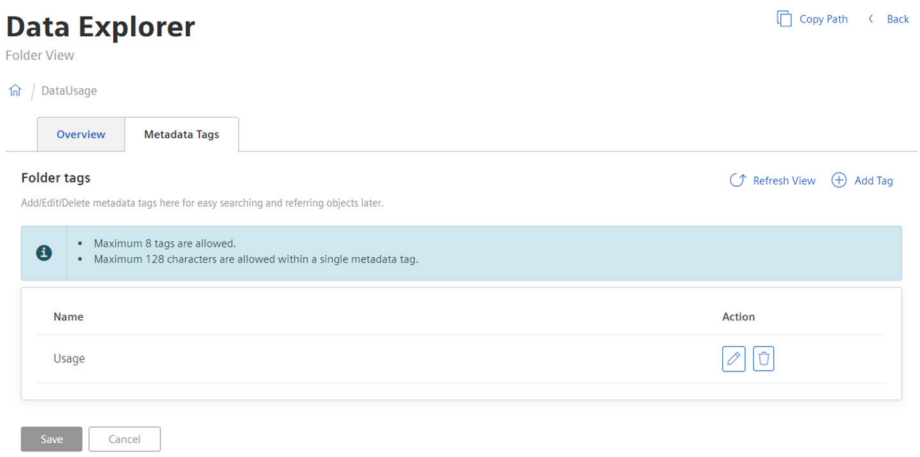
2. Enter a folder name and click ✓ to save the folder.
3. Click 👁 to add the metadata tags for the created folder.
4. Click "Add Tag".
5. Enter the tag name.
6. Click "Save".

Note

- If the folder is added with metadata tag, then deleting the folder is not possible.
- Folder name cannot contain '/'.
- Folder name can contain only alphanumeric with * ' & $ @ : ; + , < > ~ [ ] " # | ? { } ^ % ` ( ) _ = - characters.

Result

The folder is created successfully with metadata tag in Data Explorer.



Note

Maximum 8 multiple tags can be added to the folder.

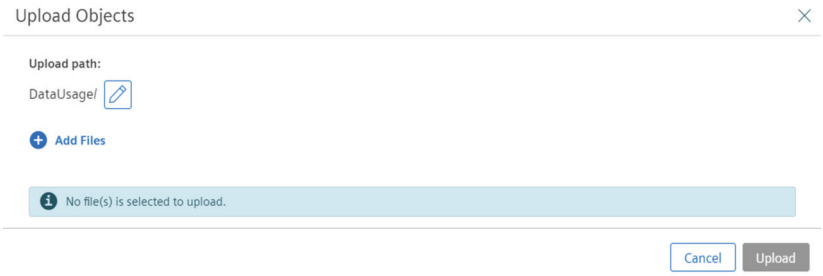
8.3 Uploading the files/objects to the folder or Data Explorer

You can upload the files/objects to the folder or "Data Explorer". The files/objects are uploaded with different formats.

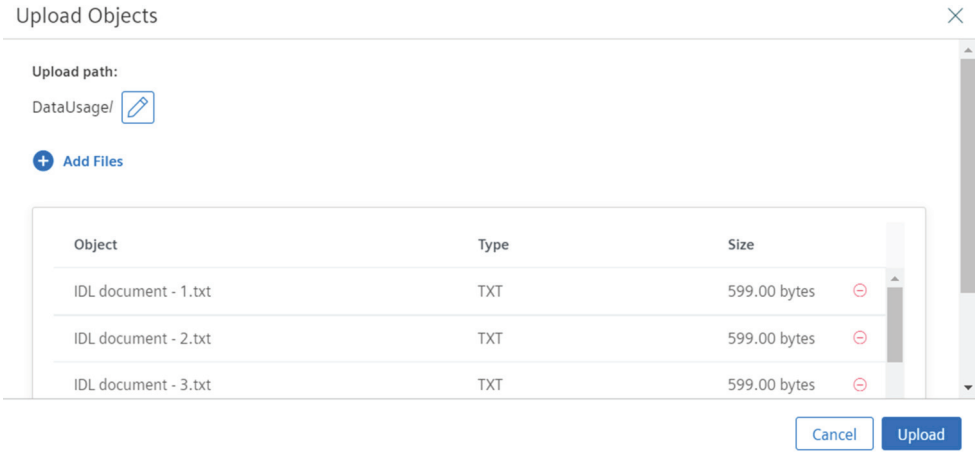
Procedure

To upload the files/objects into Integrated Data Lake, follow the steps below:

1. In the "Data Explorer" window, click "Upload Objects".



2. In the Upload Objects, click "Add Files" and select the files/objects to be uploaded.



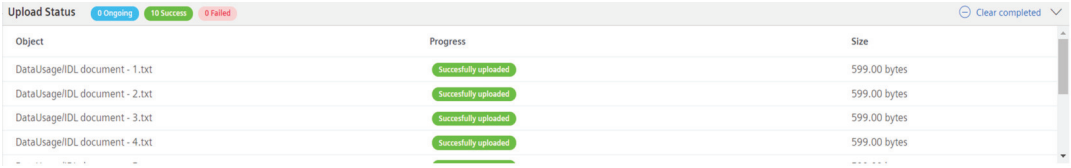
3. Click "Upload".

Note

- A maximum of 30 files/objects with 5 GB per file can be uploaded.
- If duplicate files/objects are uploaded (with same name or different size with same name), the recently uploaded file/object will overwrite the existing file/object.
- If you refresh the page while ongoing upload operation, the browser will prompt the confirmation message.
- You will be notified with a warning message, when you try to access other pages/module during upload.
- The file size greater than 400mb is recommended to download through API.
- Click button to change the path.
- Click button to remove the file from the list.
- Upload into "Time Series Import" folder will not be allowed in Integrated Data Lake.

Result

Uploading of files/objects is successful and the upload status can be viewed.

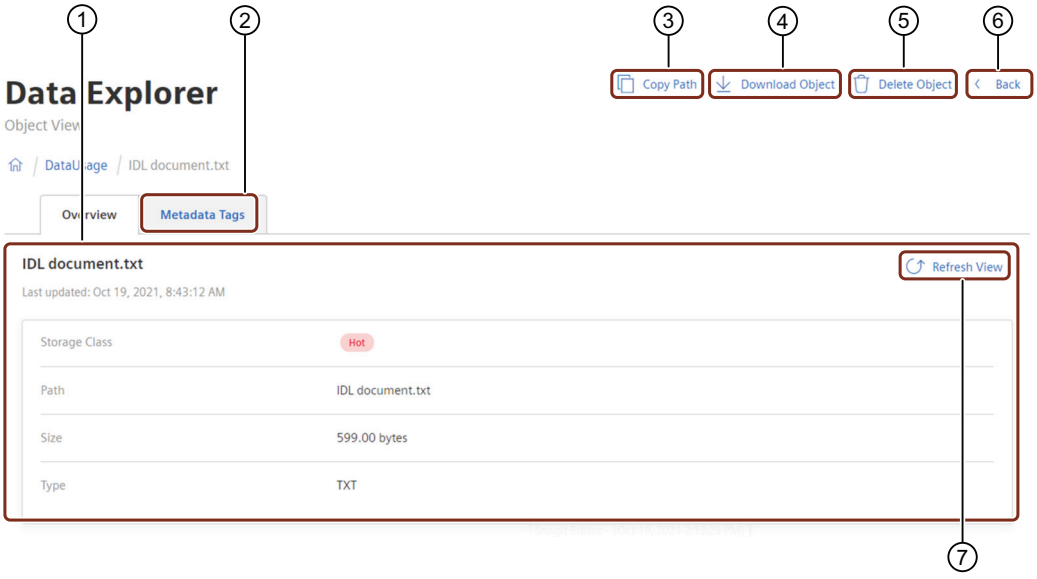


In Data Explorer, click "Refresh View" button to view the uploaded files/objects in the table.

8.4 Adding metadata tags to the object

After creating a folder, you can upload the objects to the folder. After uploading the object, you can view the object details and add the metadata tags. You can view the object details by clicking and add the metadata tags to the object.

View object details

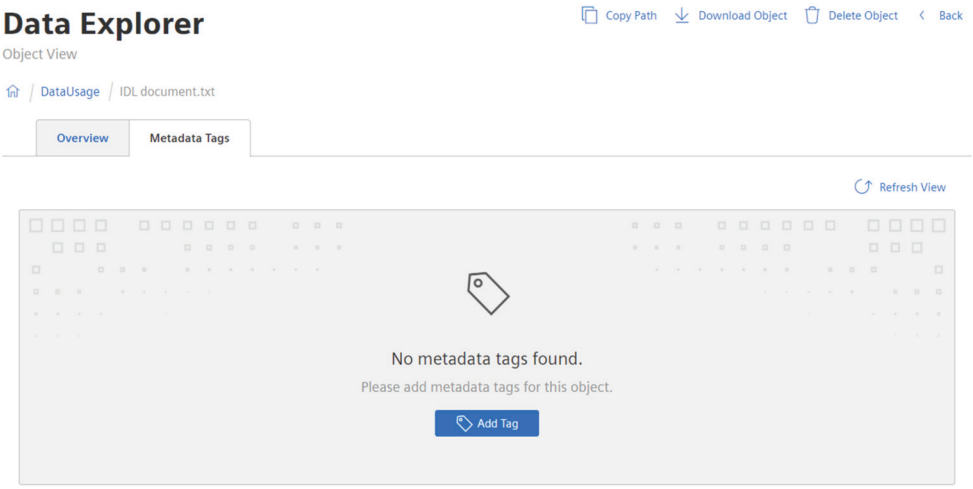


1. Object overview details
2. Add the metadata tags to the object
3. Copy the object path
4. Download the object
5. Delete the object
6. Navigates to the parent folder or Data Explorer
7. Refresh to update the object details

Procedure

To add metadata tag to the object, follow these steps:

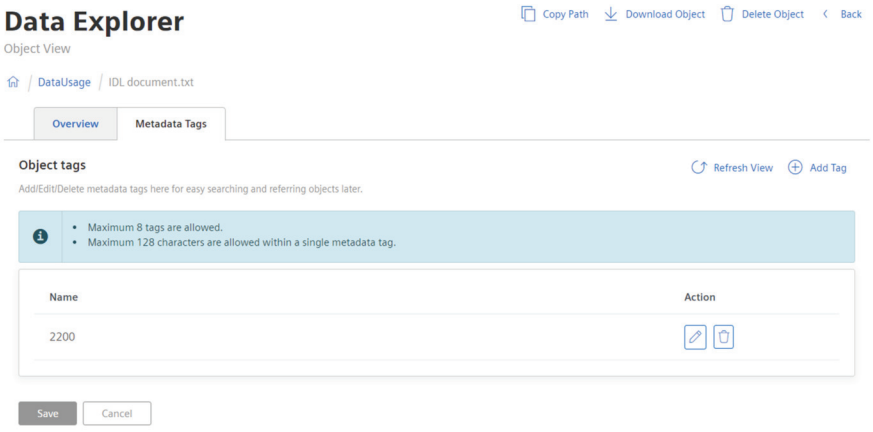
1. In Data Explorer or in folder, click to view the object details.
2. Click "Metadata Tags" tab.



3. Click "Add Tag".
4. Enter the tag name.
5. Click "Save"

Result

The "Metadata Tags" is added successfully to an object:



Note

- Maximum 8 multiple tags can be added to the object.
- If metadata added to the file which contains special characters (including space), it will be considered in metadata search.
- If metadata tags are added to folder then only first level files are tagged, second level or sub-folder files will not be tagged with the metadata.

8.5 Integrating Integrated Data Lake with Fleet Manager

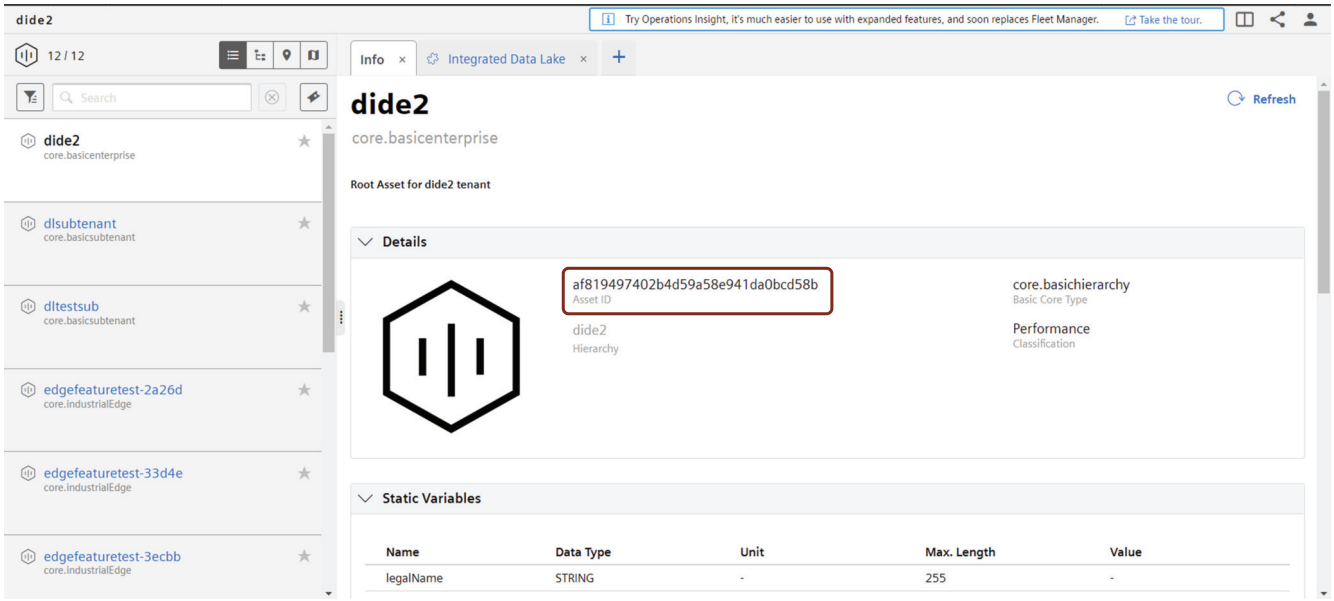
You can add Integrated Data Lake pluign to Fleet Manager and tag the asset related files/objects using Data Explorer. The assetID of an asset in Fleet Manager is used for tagging the files/objects from Integrated Data Lake.

After tagging, these files/objects of an asset can be viewed in Fleet Manager using Integrated Data Lake plugin extension.

Procedure

To tag the files/objects from Integrated Data Lake to an asset in Fleet Manager:

1. Open Fleet Manager from MindSphere Launchpad.
2. Select an asset.
3. Click "Info" extension.

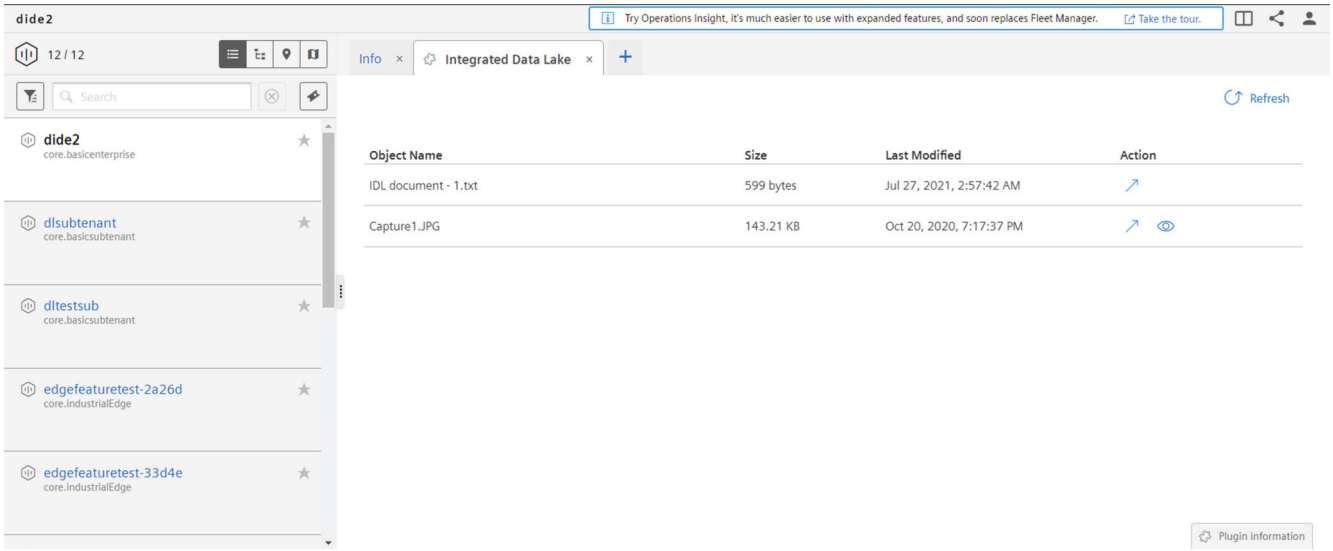


4. Copy Asset ID.
5. In Integrated Data Lake, open Data Explorer.
6. Click of the file/object and click "Metadata Tags" tab.

7. Click "Add Tag", paste the Asset ID.
8. Click "Save".

Result

To view the results, click "Integrated Data Lake" plugin extension in Fleet Manager to view the file/object tagged to an asset.



For more information about how to add a plugin to Fleet Manager, see Add plugins to Fleet Manager. (https://documentation.mindsphere.io/resources/html/fleet-manager/en-US/119809047947.html)

You can view the image file/object by clicking 👁 in Fleet Manager.

Note

Currently, "jpg", "jpeg", "png", "gif" and "bmp" image formats are supported.

8.6 Searching the files/objects in Data Explorer

The files/objects can be searched in Data Explorer by providing object name, metadata tag or both parameters with condition and operator. You can perform actions like copy path to clipboard, view object details and download the object from the "Search view".

Note

Metadata search is not possible, if metadata added to the file contains special characters (including space).

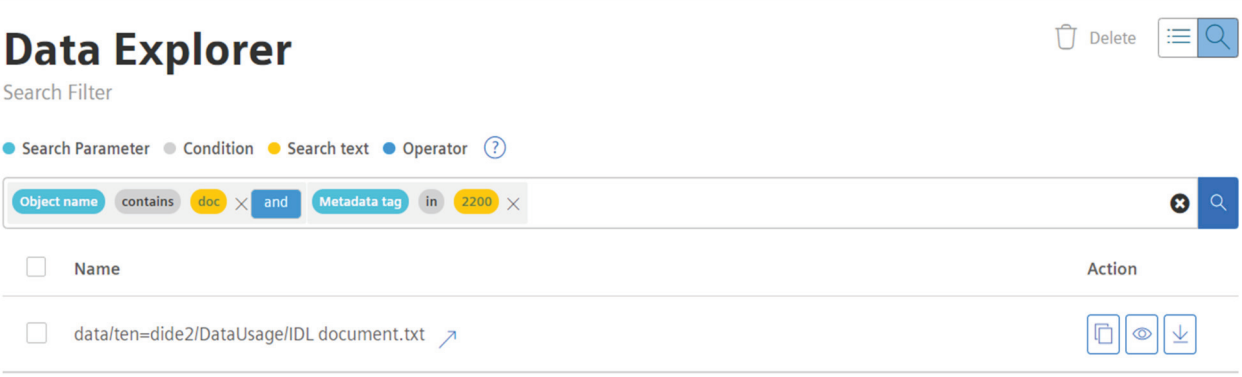
Procedure to search

To search the files/objects with Object name in Data Explorer, follow these steps:

1. Click to open "Search view".
2. Select object name or metadata tag with a condition or both parameters using an operator.
3. Enter file/object name or search text.
4. Click "Search".

Result

The file/object details will be displayed and you can perform the actions from the Search view.



Note

- If the file/object with the appropriate option is not recognized, then search by selecting both Object name and Metadata tag parameters with condition and operator.
- Search in Integrated Data Lake is case sensitive. For example, the search query will treat the string “Datalake” differently to the string “datalake”.

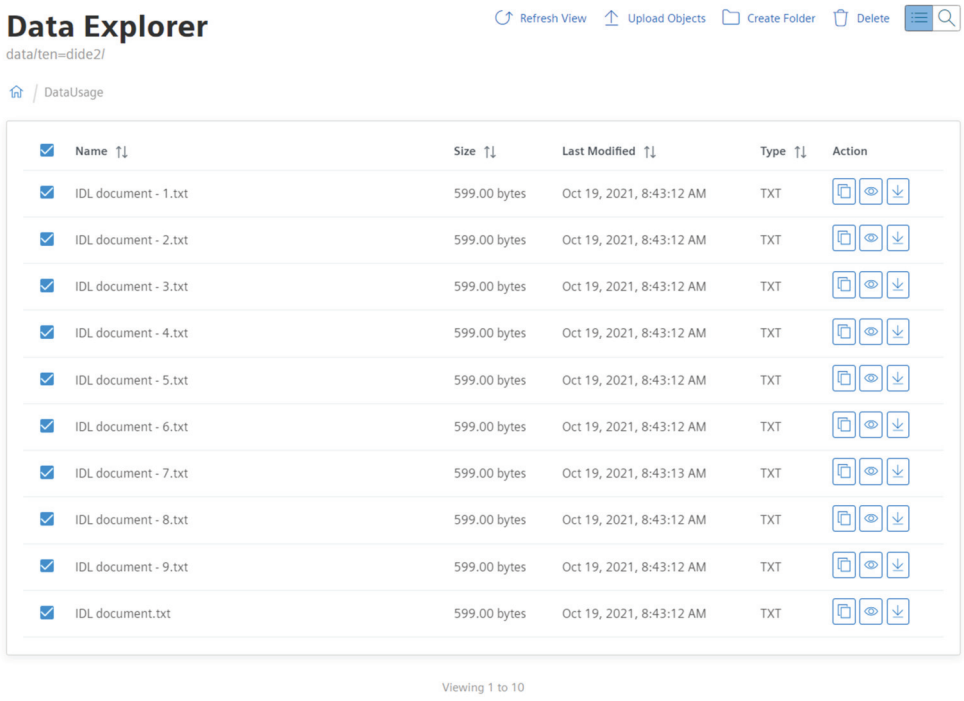
8.7 Deleting the folder or object in Data Explorer

You can delete the folder or objects in Data Explorer.

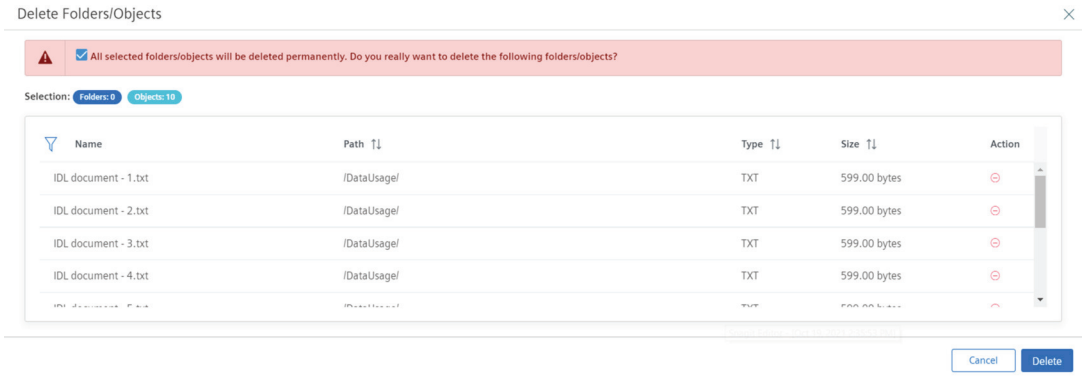
Procedure

To delete the folder or object, follow these steps:

1. Select the folder or object.



2. Click Delete.
3. Check to confirm.



4. Click "Delete".

Result

The folder or object is successfully deleted from Data Explorer.

Note

- The folder will be deleted automatically, if you delete all files/objects from Data Explorer.
- Deleting a folder is allowed only if the folder is empty and no metadata tag is added on the folder level.
- You can select single/multiple files/objects.
- Click ⊖ button to remove the file from the list.

Event Subscription

9

9.1 Creating an Event Subscription

Event Subscription allows you to subscribe the events to get notifications. You can register the destination Service Bus for Azure and Simple Notification Service for AWS to get notifications, which will be published by Integrated Data Lake service. These notifications include object events like create, update or delete in tenant prefix. You can add, view, edit and delete event subscriptions.

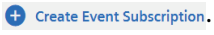
If the permission to send notification to Service Bus/ Simple Notification Service topic is removed or there is any misconfiguration, tenantAdmin wil be notified through an email to check and respond. Even the status of the subscription will be changed to "Inactive" on the user interface. You can change the status to "Active" by resolving the configuration in the user interface.

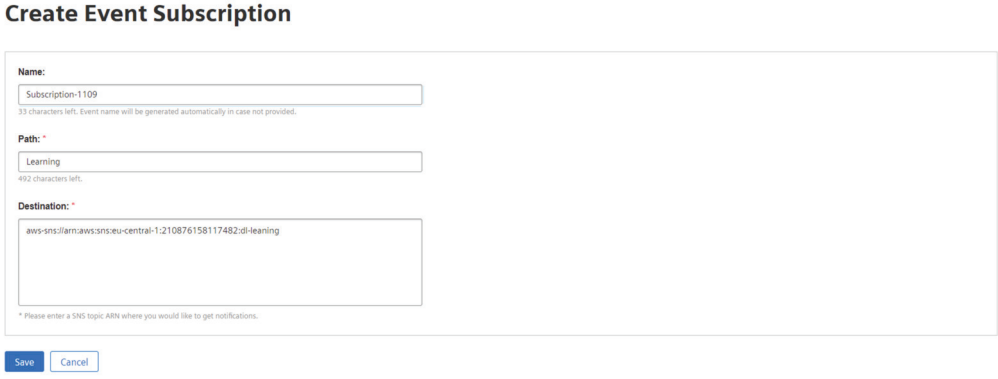
Note

Subscription name is not mandatory. Incase the name is not provided, default name for the subscription will be chosen as (subscription_<<unique_identifier>>).

Procedure

To subscribe the events with the notifications, follow these steps:

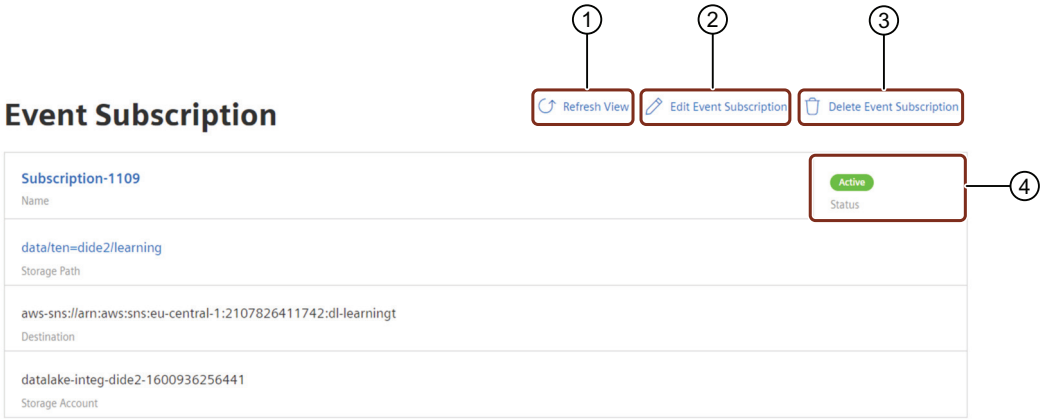
1. Click on "Event Subscription" tab.
2. In "Event Subscription", click  .
3. Click  .



   – Enter the "Name".
   – Enter the path for storage.
   – Enter the path for destination.
4. Click Save.

Result

You are now successfully subscribed to get the notifications for Integrated Data Lake services.



1. Edits the Event Subscription
2. Deletes the Event Subscription
3. Status of th Event Subscription

Note

- You can only change the status from 'Inactive' to 'Active'.
- Maximum 15 Event Subscriptions can be created for each tenant.
- If the subtenant is deleted, you can only view its existing event subscriptions and delete them if not required.

Glossary

Destination path

The desired path where the user wants to store the historical data in Integrated Data Lake.

Task name

The user reference for the created import task.