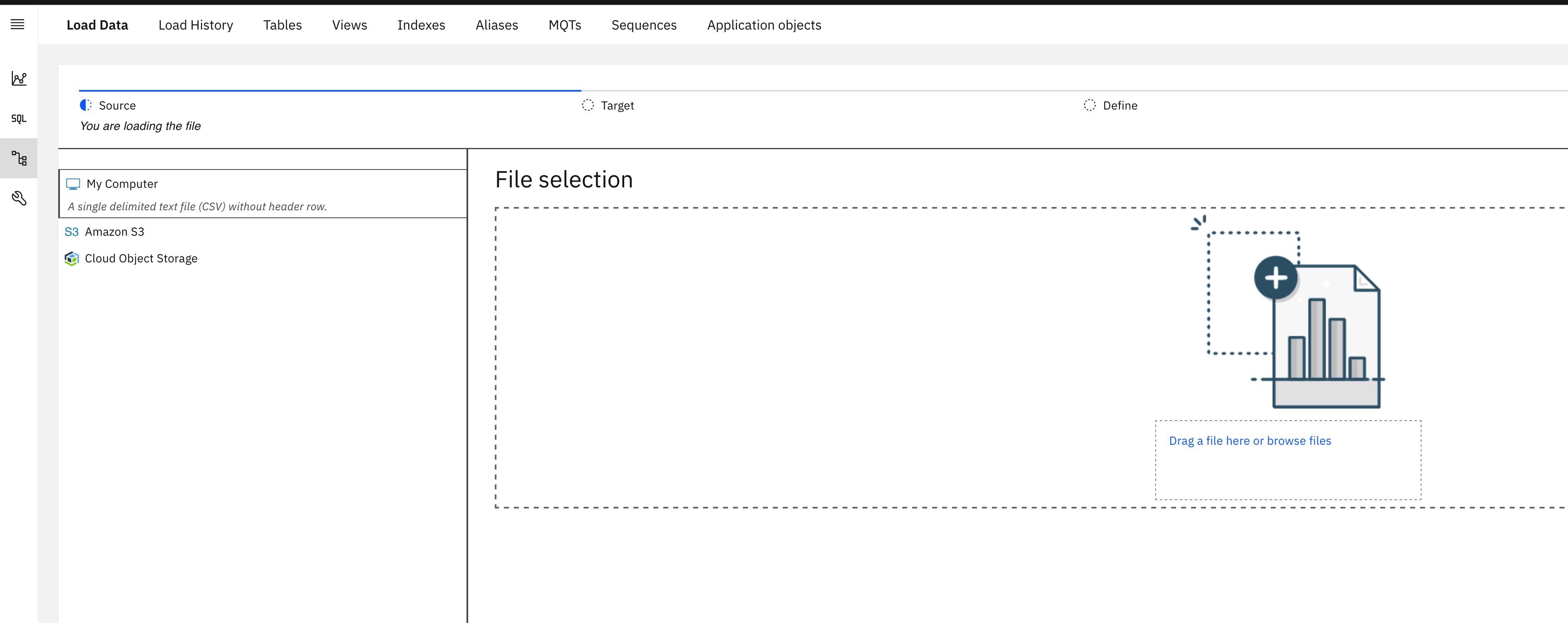
1. Upload csv files from Kaggle competition to some local folder

<https://www.kaggle.com/competitions/titanic/data>

1. Copy all of the files from wkc folder on GitHub
2. Db2 data upload

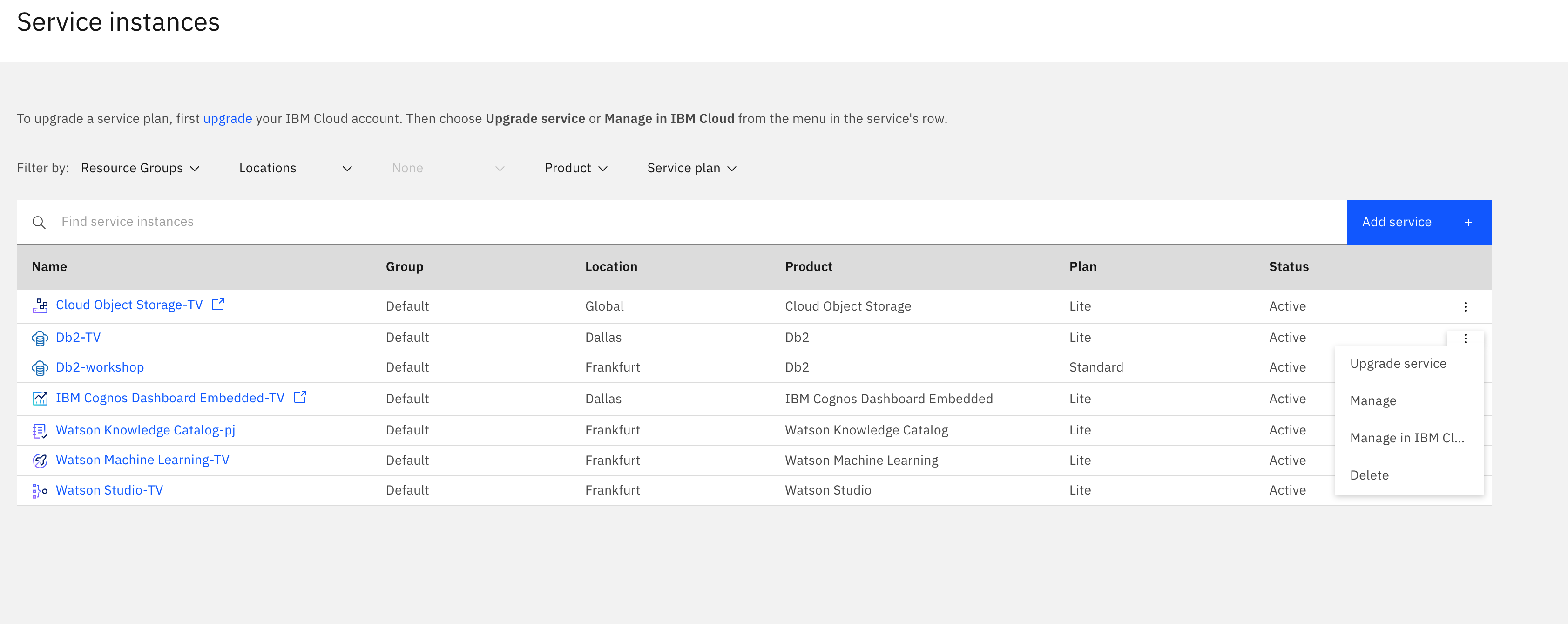
* From **menu/Services instances** check that you have **Db2 Lite** instance provisioned
* Start db2 by clicking on it
* Then **Go to IU**
* From **Data Part/Load Data** click on “***Drag a file here to browse files***” and upload train data set



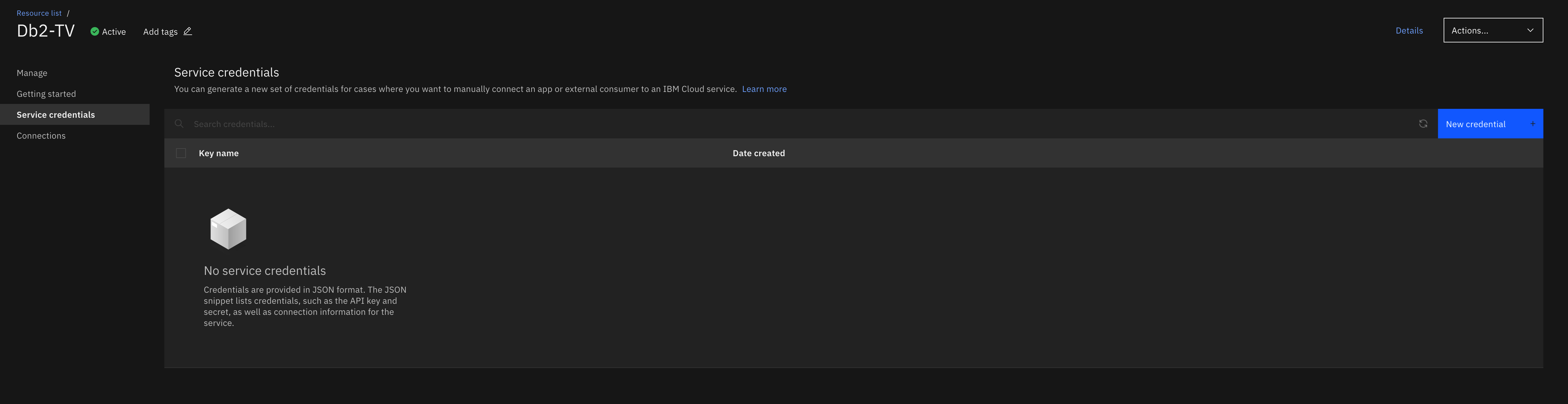
Within upload you will use the default Schema for data (within Lite version you are not able to create a new schema), but create a new table named **TITANIC\_TRAIN\_DATA**

1. Create credentials for db2 connection

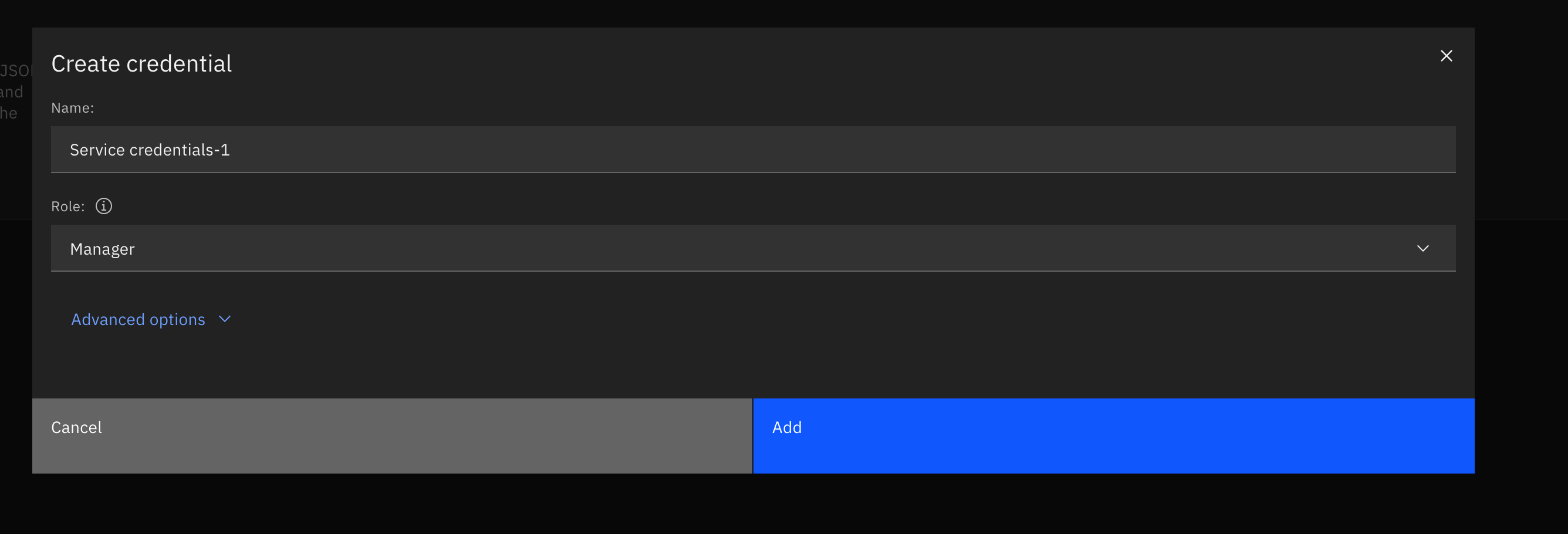
* Return to **Service instances**



* And by clicking on three dots on the right from DB2 instance select **Manage in IBM Cloud**
* From the left side menu select Service Credentials



* And on the right side click New Credential +
* Type in the Name of Credentials and select Manager role, click Add.

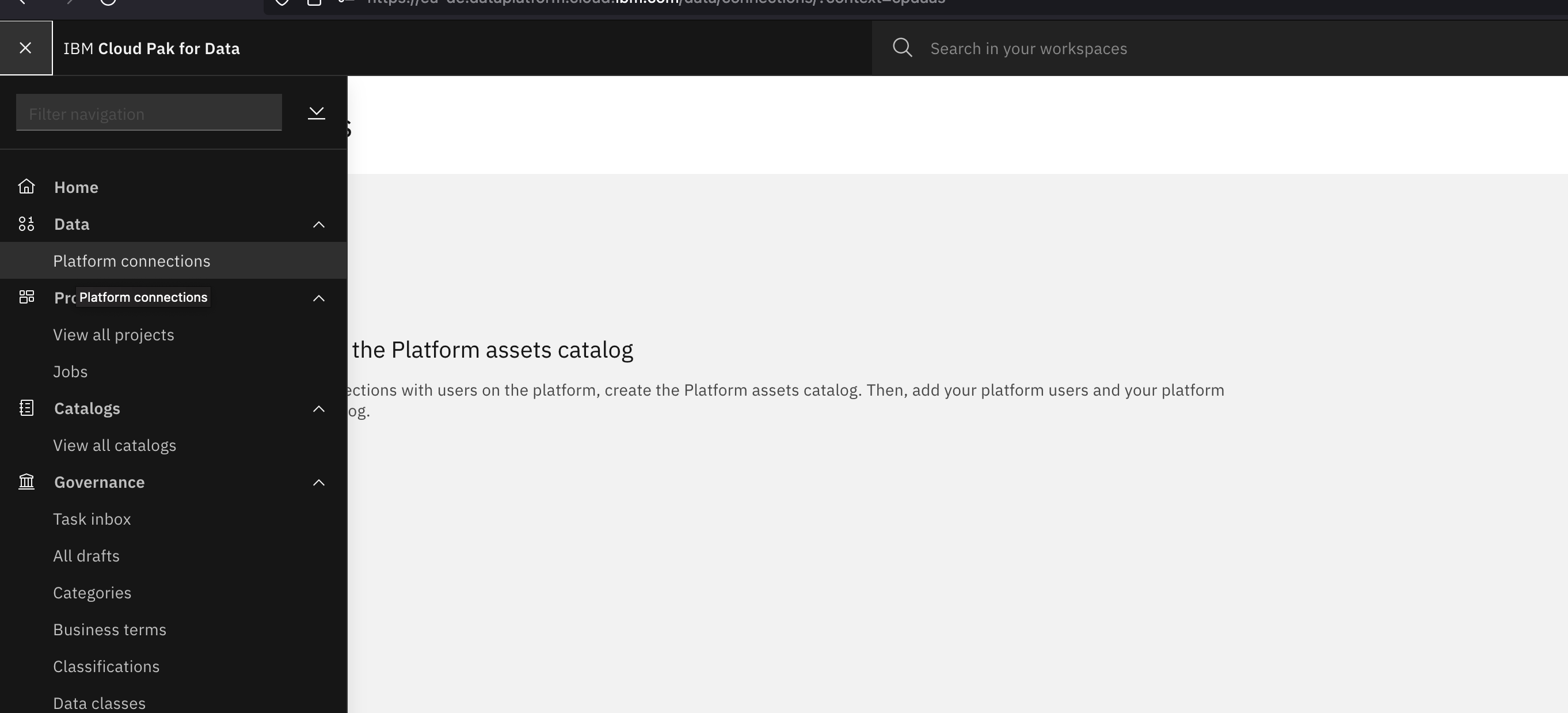


This will generate credentials for connection to db2, the most important are:

* Hostname
* Port
* Database (database name)
* Username
* Password

1. Connect to db2 via Platform Connections of Cloud Pak for Data

* Go to Platform connections from the main Cloud Pak for Data menu

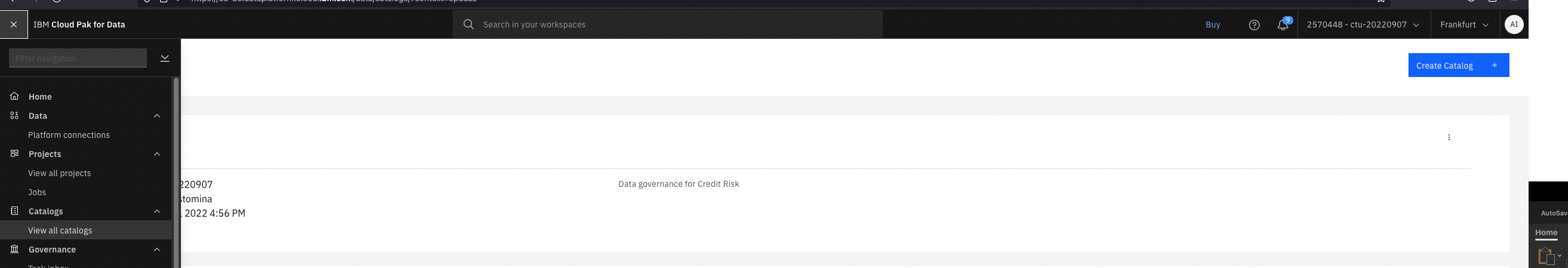


* It will ask you to create Platform assets catalog
* Click on Create catalog
* Select Overwrite original assets option and select Object Storage that you have initially created
* After that click in Create in the bottom right corner of the screen
* In the Access Control Tab that is opened by default after Platform assets catalog is created make sure that you are in the list of Collaborators
* Then from the Connections tab choose New connection +
* Select connection type, that is IBM Db2
* Fill in Text into Connection Name and Description
* Then add Connection details and Credentials from Service credentials created earlier, fill the box with Port is SSL-enabled
* After that test connection by choosing Test connection in the upper right corner of the screen
* After successful text Create by clicking on Create in the right down corner

1. Create a Catalog for Titanic data

We will create a catalog to share data across the whole account, with Lite version it is possible to create up to 2 Catalogs.

* From the Cloud Pak for Data navigation menu Navigation menu, choose Catalogs > View all catalogs.
* Click Create Catalog.
* For the Name, copy and paste the catalog name exactly as shown with no leading or trailing spaces: “Titanic Workshop Catalog”
* If prompted to associate the catalog to a Cloud Object Storage instance, select a Cloud Object Storage from the list.
* Select Enforce data policies, confirm the selection, and accept the defaults for the other fields.
* Click Create.



1. Add a category

* From the Cloud Pak for Data navigation menu Navigation menu, choose Governance > Categories.
* Click Add category > Import from file
* Add file
* Select glossary-organize-governance-titanic
* Select \* Replace all values and click Import
* And after successful import Close

It might take a few minutes for Categories to load

1. Add business terms

You will import business terms into a newly created Category **Titanic data**

* From the Cloud Pak for Data navigation menu, choose **Governance > Business terms.**
* Click Add business term > Import from file.
* Click Add file.
* Select the business-terms-titanic.csv file that you downloaded earlier.
* Click Open.
* Click Next.
* Select Replace all values, and click Import.
* Click Go to task to see the draft business terms. If you miss the notification, then from the Cloud Pak for Data as a Service navigation menu Navigation menu, choose Governance > Task inbox.
* Select the Publish business terms checkbox, and then click Publish. Click Publish to confirm.
* From the Cloud Pak for Data navigation menu Navigation menu, choose Governance > Business terms to the published business terms.

1. Create Titanic Project and add db2 connection to assets

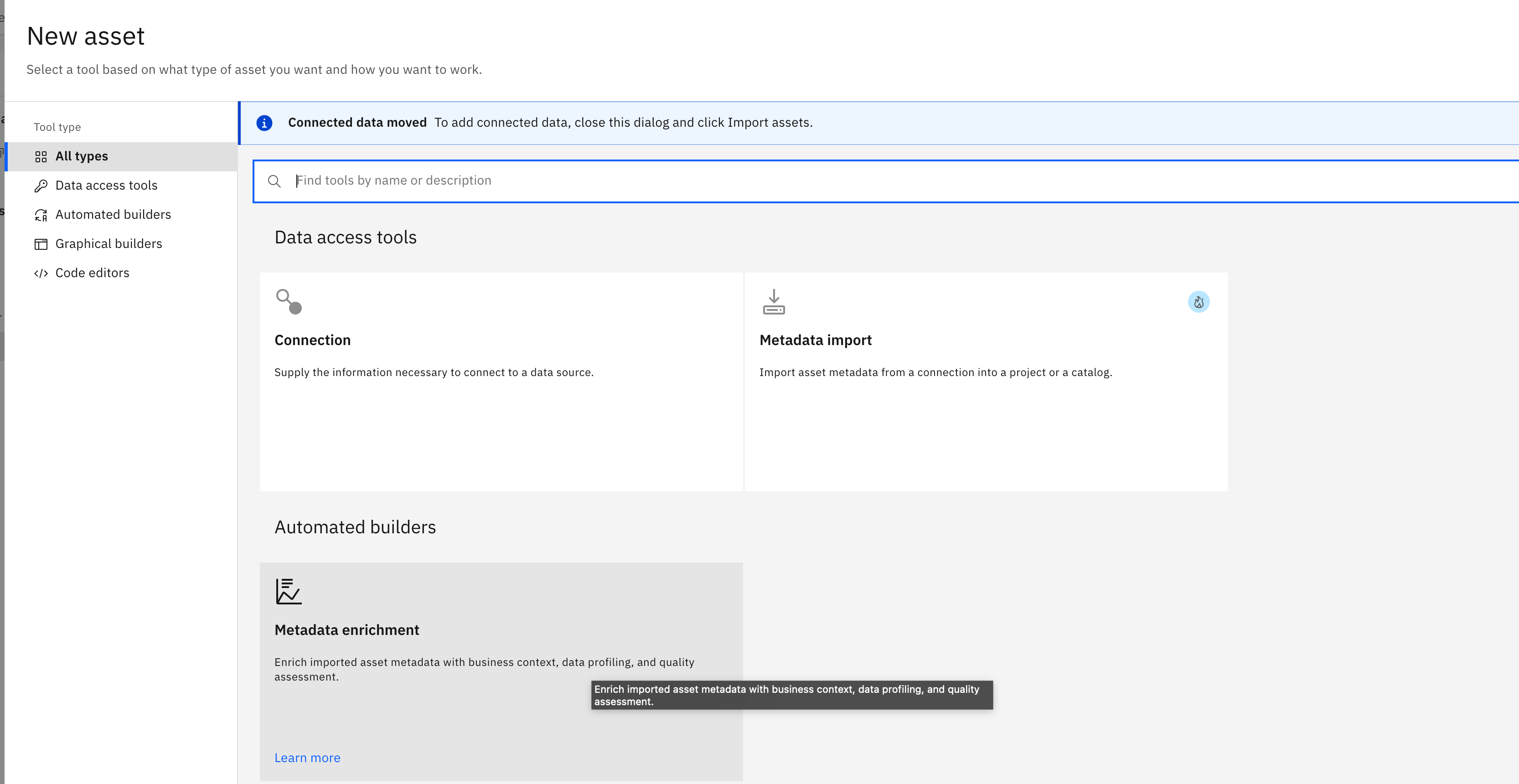
* From the navigation menu > Projects > View all projects
* Click New project+
* Select Create an empty project
* Fill in Project name, Project description (not obligatory) and select Object Storage then click Create.
* Click the Assets tab to view the project's assets.
* From Assets click New asset, select Connection and there from tab From Platform select DB2 connection that you have created earlier, click Select and then Create

1. Import metadata to the project

* Click the Assets tab.
* Click New asset.
* Select Metadata Import.
* Type in Name and Click Next to continue.
* On the Select target page, select This project, and click Next to continue.
* On the Select scope page, click Select connection.
* Select the Data Fabric Trial - Db2 Warehouse connection.
* Click the default SCHEMA and select table name
* Click Next to continue to the schedule.
* Click Next to continue to the review.
* Review the summary of the import, and click Create. The metadata import job starts.
* Refresh to see the status

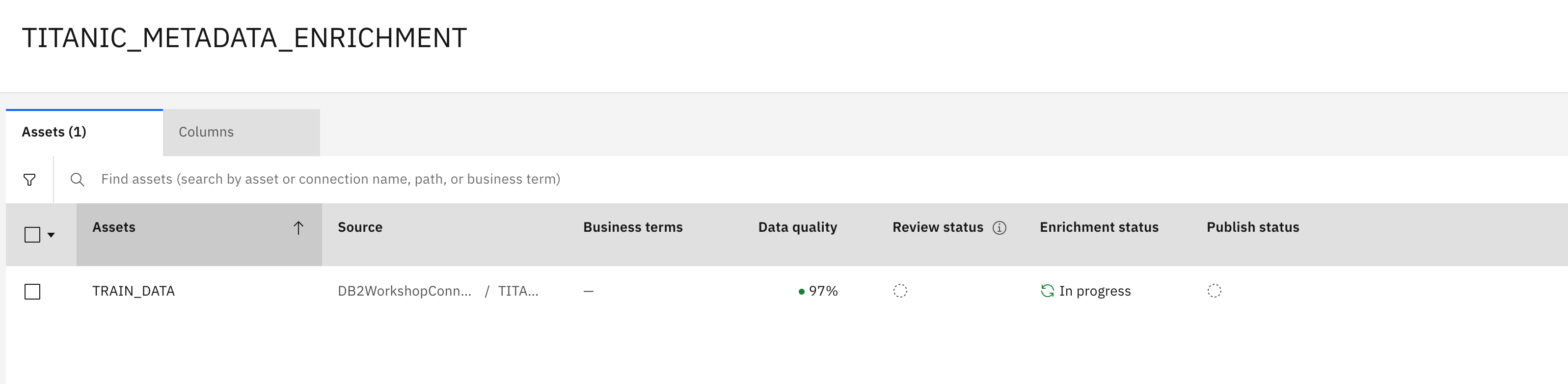
1. Create and run the job of metadata enrichment

* On the Assets tab, click New asset.
* Select Metadata Enrichment.



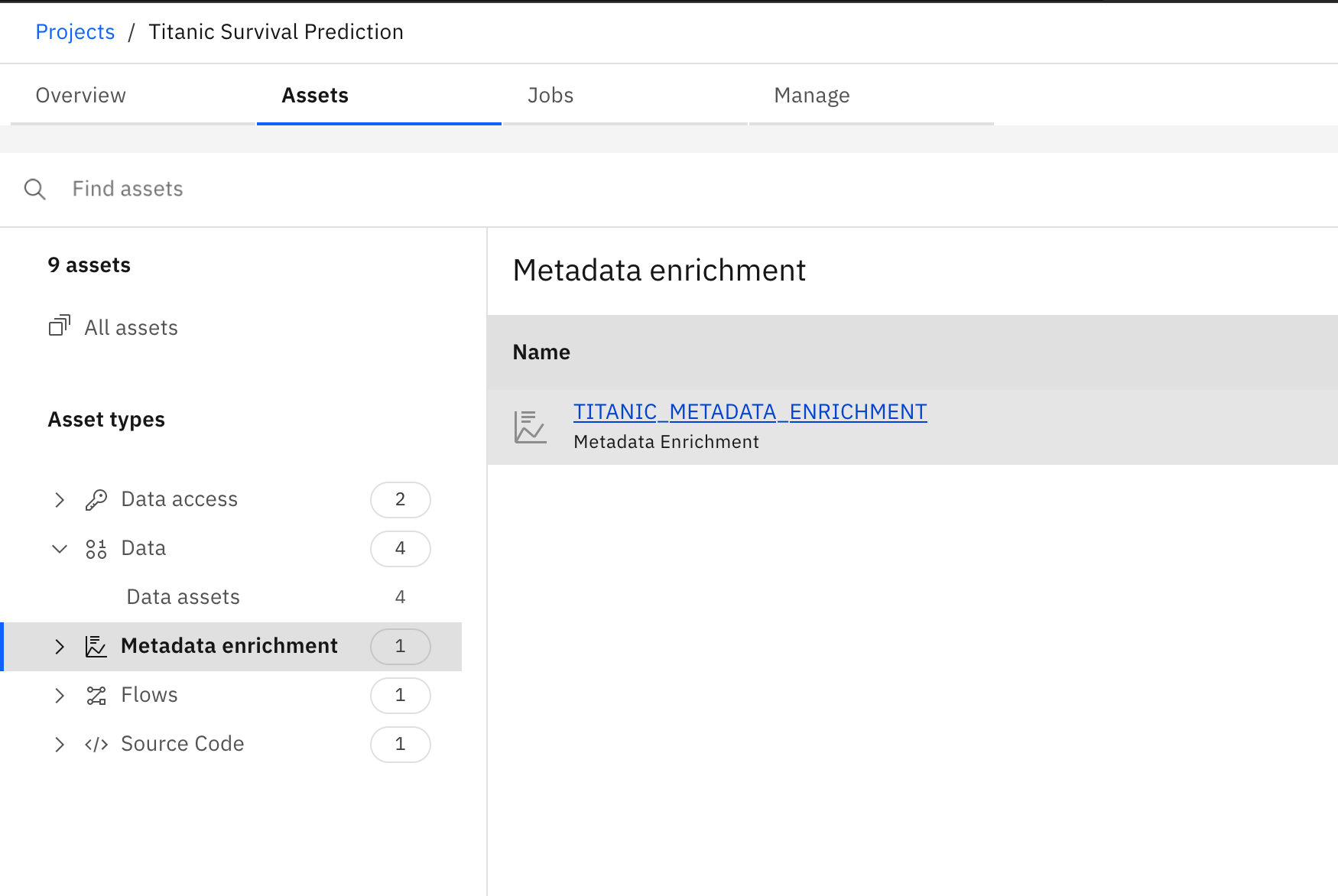
* Type in some name and Next to continue
* Click Select data from project.
* Select Metadata import.
* Click the checkbox next to the imported metada
* Click Next to continue to the enrichment objective.
* Select all enrichment objectives:
  + Profile data
  + Analyze quality
  + Assign terms
* For Categories, click Select categories.
* Select only [uncategorized] and Titanic data.
* For the Sampling, select Basic.
* Click Next to continue to the schedule.
* Click Next to continue to the review.
* Click Create.

The job might take several minutes to complete. Refresh to update the progress of the job.

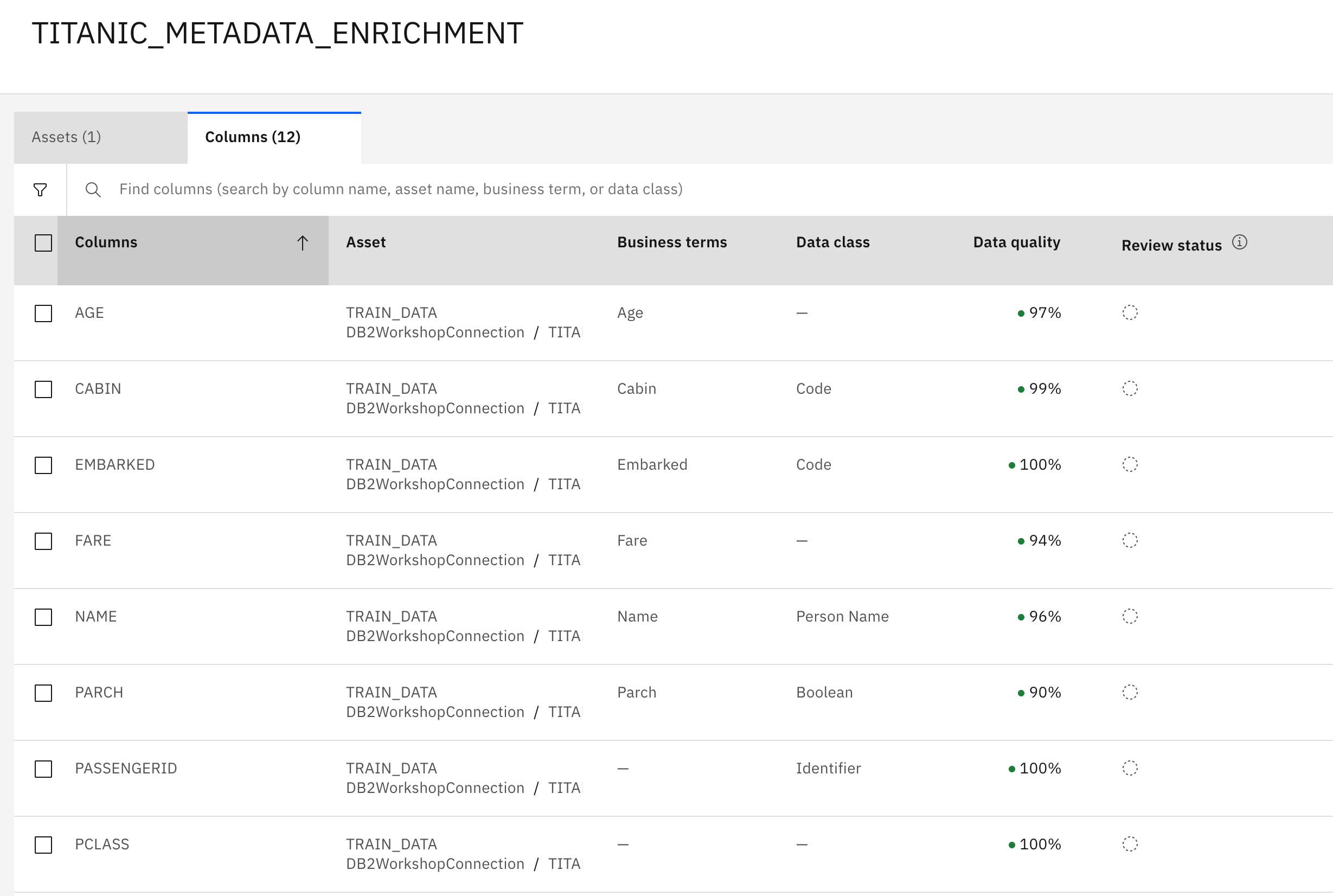


1. View the enriched data

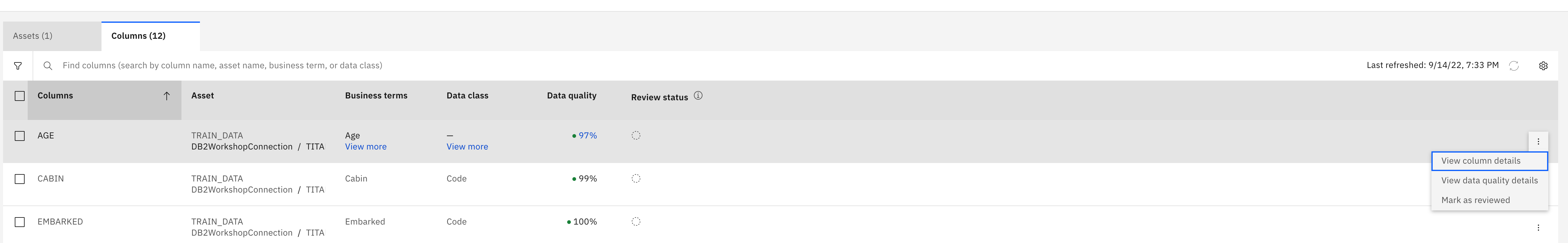
* From the project select metadata enrichment that you have created before.



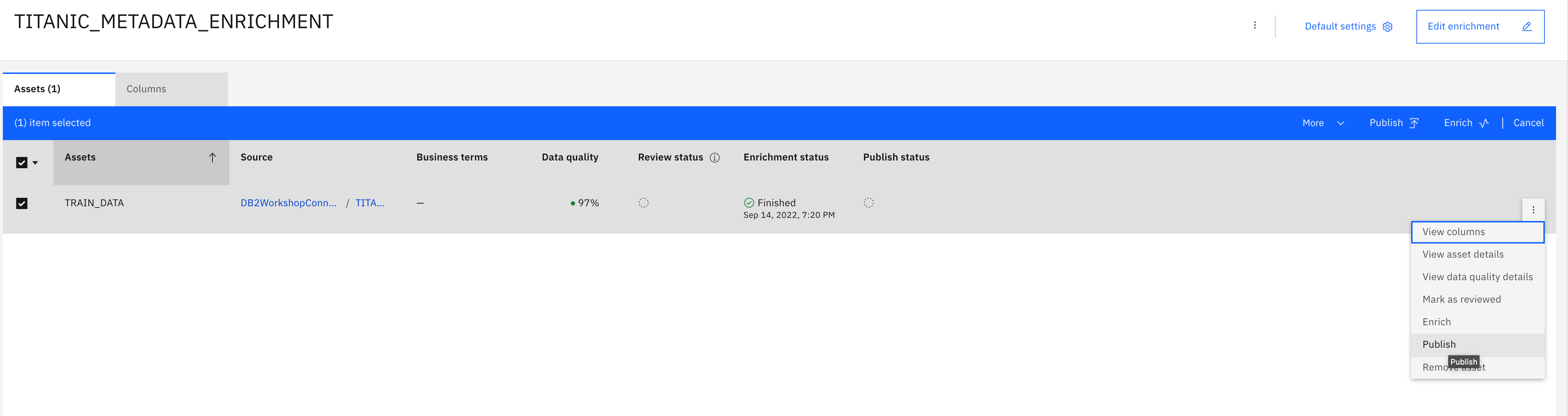
* From the Columns tab review Business terms and Data classes assigned as well as Data Quality



* After review mark those columns as reviewed

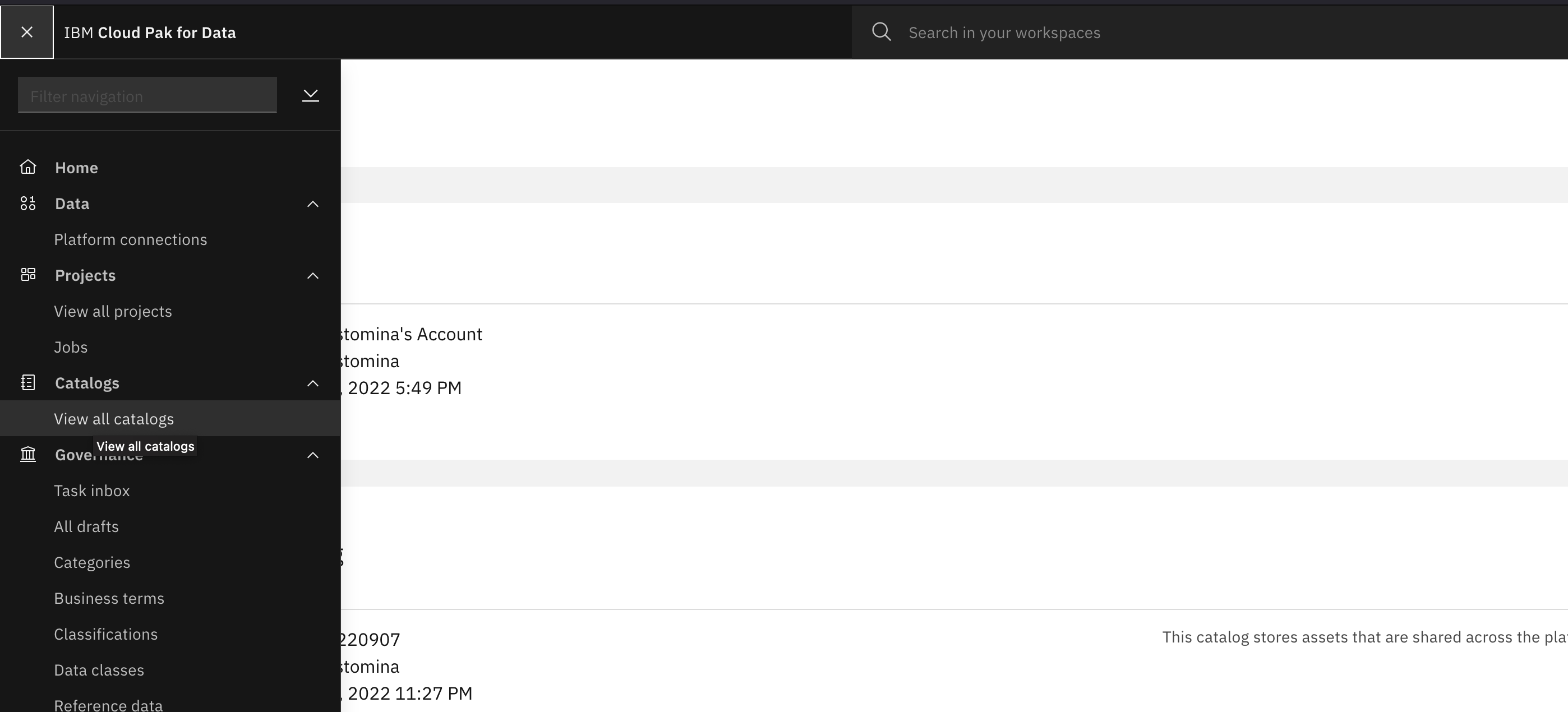


* From Assets Tab publish reviewed assets into the Catalog

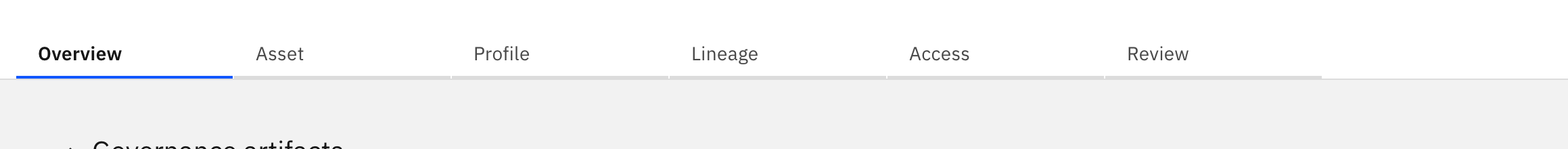


* Access published data in the Catalog

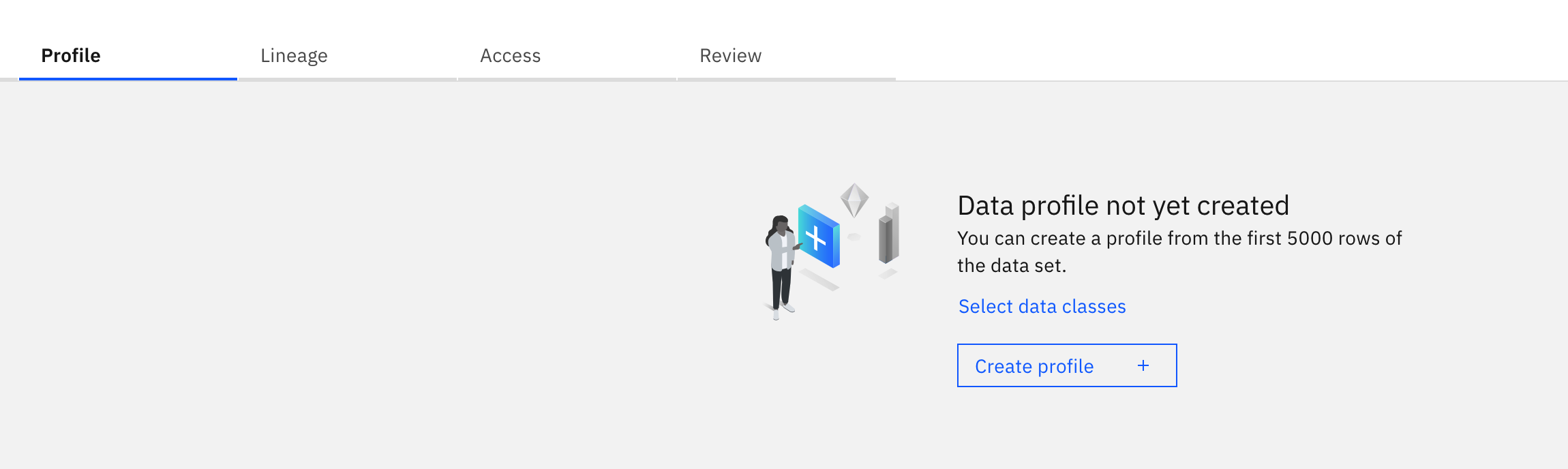
Catalog > View all catalogs > Your Catalog Name



* From there select the published asset and go over the different tabs



* Create profile from profile tab



* Make a review of the data