# Register and Scan Google BigQuery source (preview)

This article outlines how to register a Google BigQuery project in Purview and set up a scan.

## Supported capabilities

The BigQuery source supports Full scan to extract metadata from a BigQuery project and fetches Lineage between data assets.

## Prerequisites

1. Set up the latest self-hosted integration runtime. For more information, see Create and configure a self-hosted integration runtime.
2. Make sure JDK 11 is installed on your virtual machine where self-hosted integration runtime is installed.
3. Make sure "Visual C++ Redistributable 2012 Update 4" is installed on the self-hosted integration runtime machine. If you don't yet have it installed, download it from here.
4. You will have to manually download BigQuery's JDBC driver on your virtual machine where self-hosted integration runtime is running from here
   **Note**: The driver should be accessible to all accounts in the VM. Do not install it in a user account.
5. Supported Google BigQuery version is 11.0.0

## Feature Flag

Registration and scanning of BigQuery source is available behind a feature flag. Append the following to your URL: &feature.ext.datasource={"bigQuery":"true"}

E.g., full URL https://web.purview.azure.com/?feature.ext.datasource={"bigQuery":"true"}

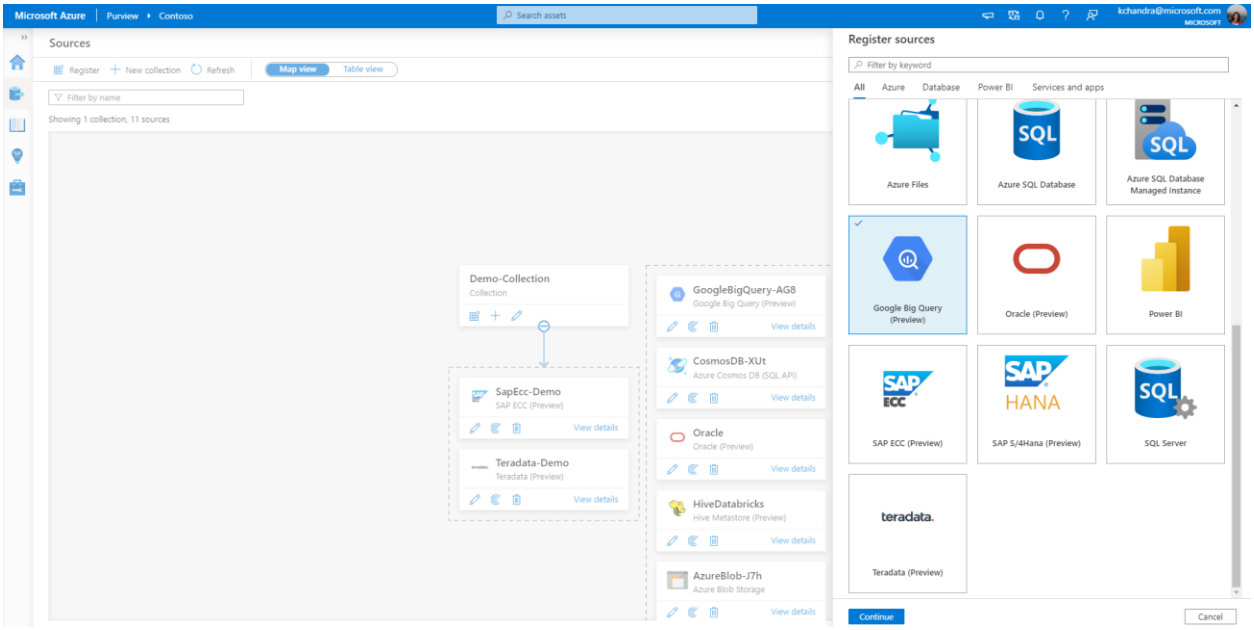
## Setting up authentication for a scan

The only supported authentication for a BigQuery source is **Service account key.**

## Register a Google BigQuery project

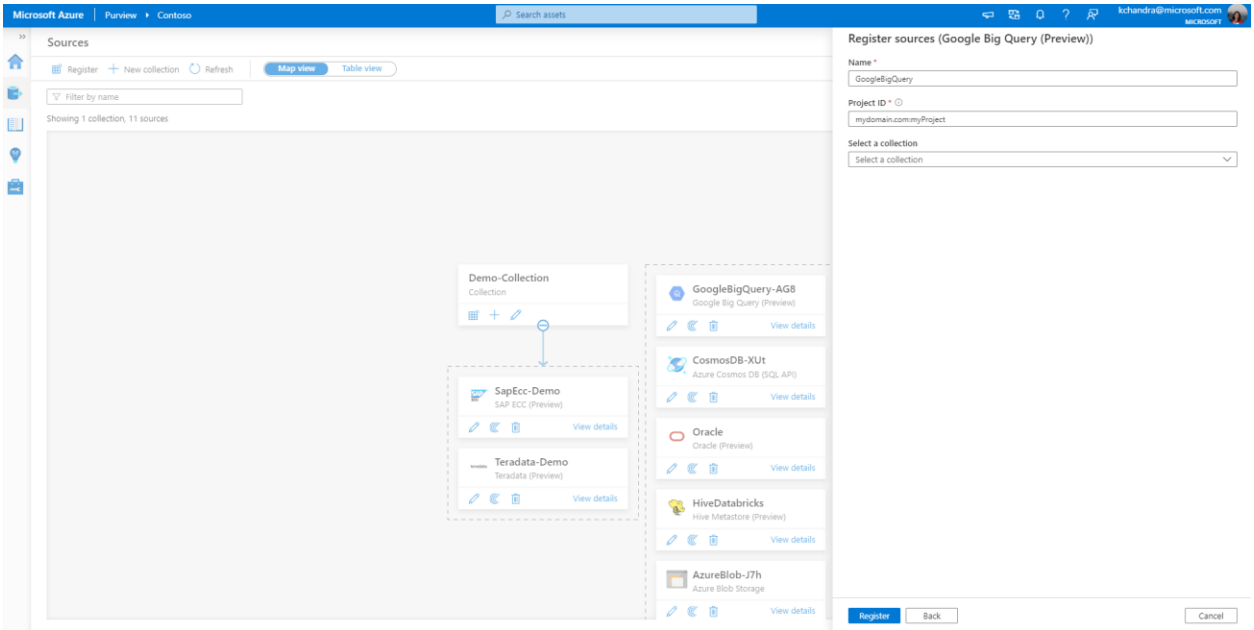
To register a new Google BigQuery project in your data catalog, do the following:

1. Navigate to your Purview account.
2. Select **Sources** on the left navigation.
3. Select **Register.**
4. On Register sources, select **Google BigQuery** . Select **Continue.**




On the Register sources (Google BigQuery) screen, do the following:

1. Enter a **Name** that the data source will be listed within the Catalog.
2. Enter the **ProjectID.** This should be a fully qualified project Id. For example, mydomain.com:myProject
3. Select a collection or create a new one (Optional)
4. Finish to register the data source.




## Creating and running a scan

To create and run a new scan, do the following:

1. In the Management Center, click on Integration runtimes. Make sure a self-hosted integration runtime is set up. If it is not set up, use the steps mentioned here to setup a self-hosted integration runtime
2. Navigate to **Sources**.
3. Select the registered **BigQuery** project.
4. Select **+ New scan**.
5. Provide the below details:
    a. **Name**: The name of the scan
    b. **Connect via integration runtime**: Select the configured self-hosted integration runtime
    c. **Authentication method**: Only "Service account key file" authentication is supported currently. This option will be selected by default.
    d. **Private key file path**: Specify the absolute path to the private key file of the service account. The private key file should either be a JSON or P12 format (with either JSON or P12 file extension).
    To create a new private key from Google's cloud platform, in the navigation menu, click on IAM & Admin -> Service Accounts -> Select a project -> Click the email address of the service account that you want to create a key for -> Click the **Keys** tab -> Click the **Add key** drop-down menu, then select Create new key. Now choose either JSON or P12 format.
    Place this file in your virtual machine where self-hosted integration runtime is running and provide the absolute path to this file.
    e. **Client email ID**: Provide the email ID of the service account. For example, xyz@developer.gserviceaccount.com
    f. **Driver location**: Specify the path to the JDBC driver location in your VM where self-host integration runtime is running. This should be the path to valid JAR folder location
    g. **Dataset**: Specify a list of BigQuery datasets to import. For example, dataset1; dataset2. When the list is empty, all available datasets are imported.
    Acceptable dataset name patterns using SQL LIKE expressions syntax include using %, e.g. A%; %B; %C%; D
        - start with A or
        - end with B or
        - contain C or
        - equal D

    Usage of NOT and special characters are not acceptable.
    h. **Maximum memory available**: Maximum memory (in GB) available on customer's VM to be used by scanning processes. This is dependent on the size of Google BigQuery project to be scanned.

## Scan "GoogleBigQuery"

Name *

Scan-BigQuery

Connect via integration runtime * ⓘ

IntegrationRuntime-BZu (Running)

Authentication method

Service account key file

Private key file path * ⓘ

D:\Drivers\BigQuery\testproject.json

Client email ID * ⓘ

serviceaccount@testproject-30.iam.gserviceaccount.com

Driver location * ⓘ

D:\Drivers\BigQuery\SimbaJDBCDriverforGoogleBigQuery42_1.2.14.1017

Dataset ⓘ

Enter dataset

Maximum memory available ⓘ

8

☑ Use default cache location ⓘ

Continue | Test connection | Cancel

6. Click on **Test connection.**
7. Click on **Continue**.
8. Choose your **scan trigger**. You can set up a schedule or ran the scan once.
9. Review your scan and click on **Save and Run**.