Microsoft

Open Edu Analytics Implementation Guide

Published: November, 2020

Introduction

This document provides technical details on the step by step setup of the Open Edu Analytics solution.
Open Edu Analytics is an open source modern data warehouse solution for education, built on Synapse Analytics and supporting Azure platform services.
For a review of the solution itself, refer to the document Education Analytics Summary.

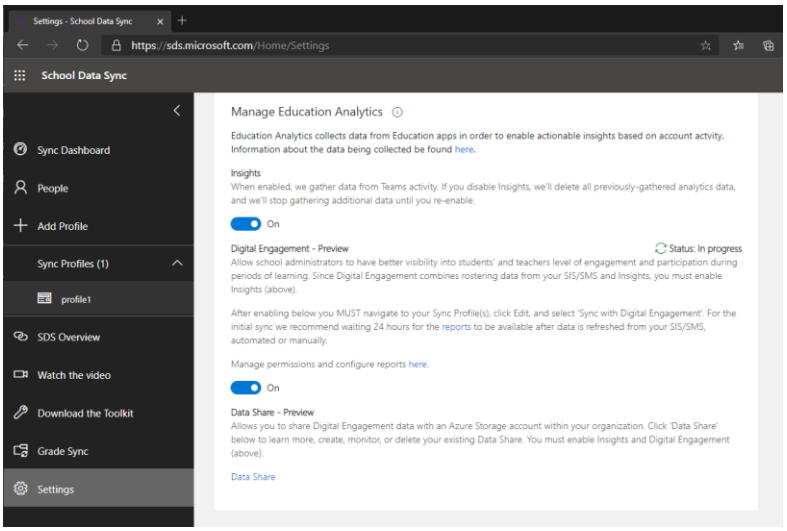
All scripts and documentation for the Open Edu Analytics solution can be found at: https://github.com/microsoft/OpenEduAnalytics

The following are required prior to implementing the steps in this guide:

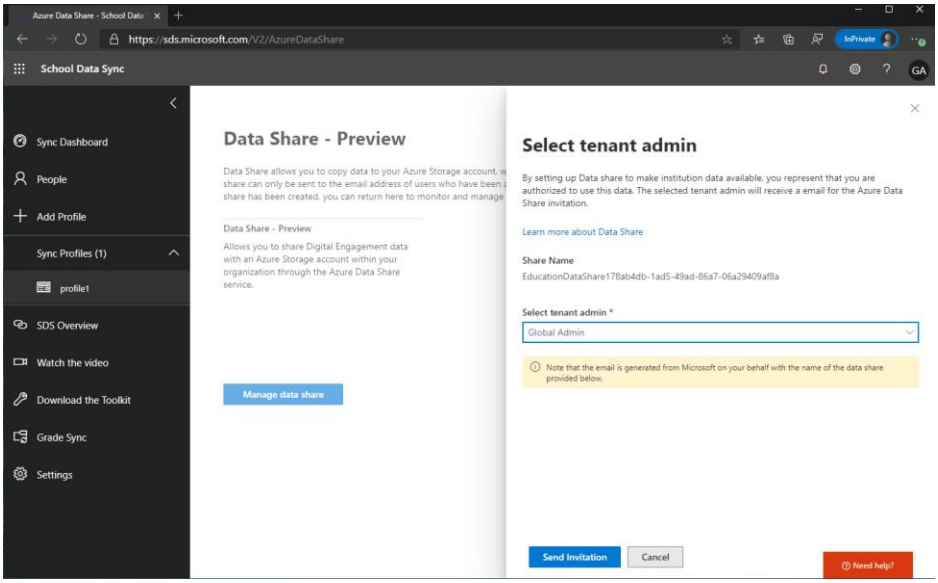
- Integration with School Data Sync
- Onboarding into the Education Analytics Azure Data Share TAP program
- An Azure subscription

1) Setup Data Share in School Data Sync

1.1) Click on Data Share in School Data Sync settings page
In order to begin receiving usage data from M365, the first step is to initiate the Data Share feature within School Data Sync. If the Data Share – Preview section is not visible in your account, check with your account manager to have the feature enabled for your tenant.



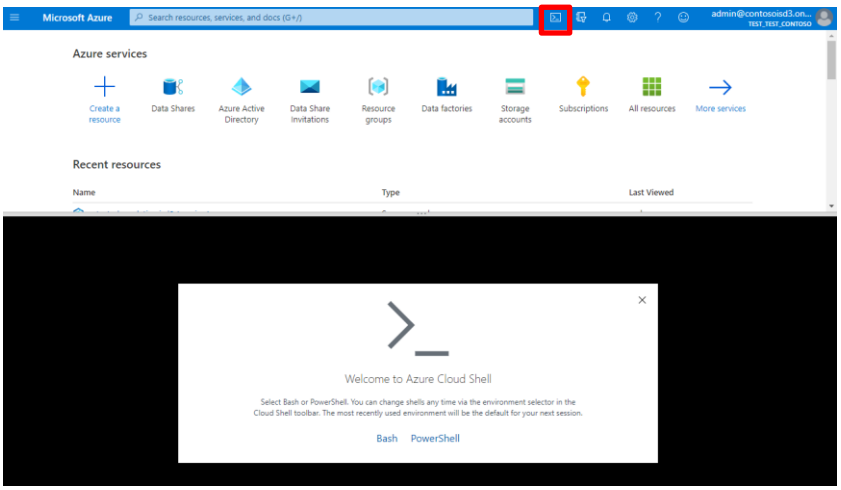
1.2) Click on “Manage data share”, then select the tenant admin who is doing the setup, and click “Send Invitation”



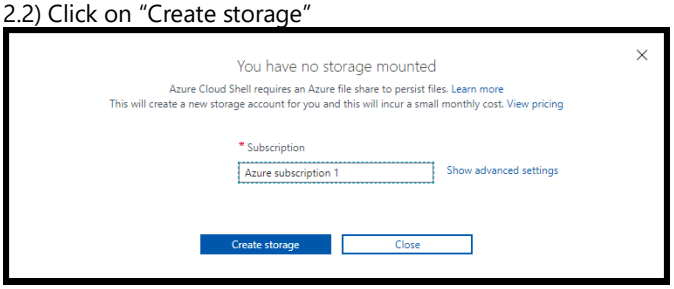
2) Setup Resources in Azure Portal

In this section you will use a script to provision the Azure resources that comprise the core of this solution.

2.1) In Azure portal, click on the Cloud Shell icon, then select "Bash".

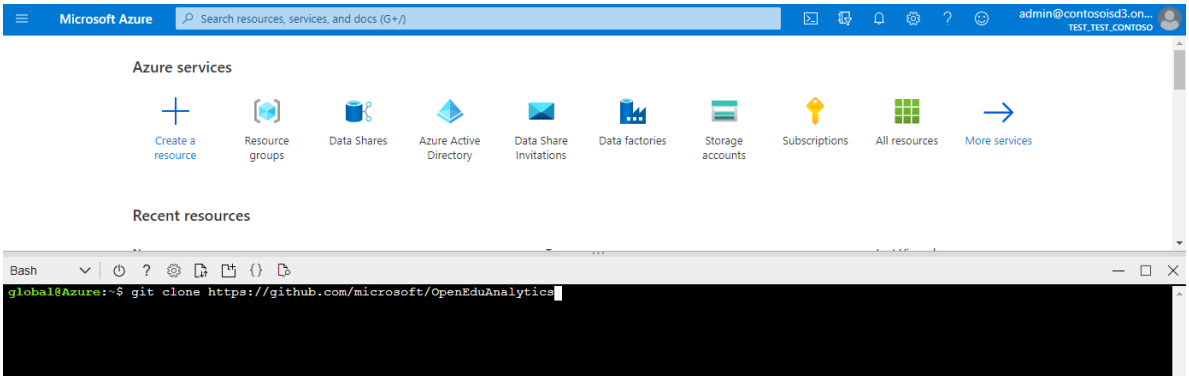


2.2) Click on "Create storage"



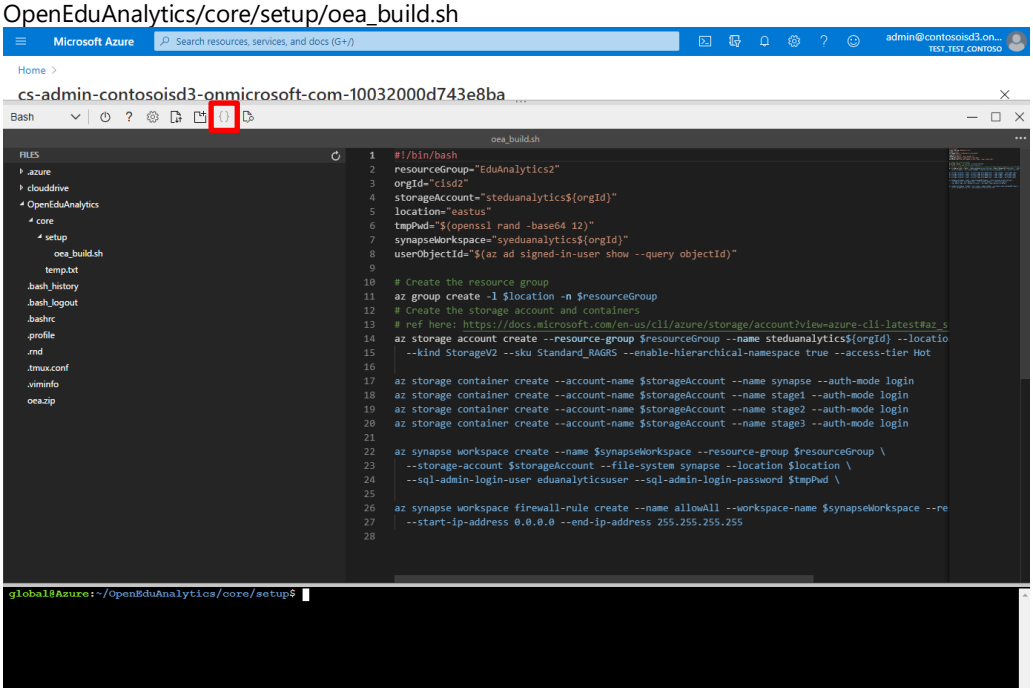
2.3) At the bash shell prompt, enter the following commands:

```
cd clouddrive
git clone https://github.com/microsoft/OpenEduAnalytics
```



2.4) Edit the build script in OpenEduAnalytics/core/setup/oea_build.sh by changing the settings for:
- "orgId" to be a value that represents your organization (this is a string that is used as a suffix in the naming of certain resources that require a globally unique id)
- "location" to be an appropriate Azure data center location

You can make this edit by using the built-in editor (click on the "Open Editor" icon), then navigate to OpenEduAnalytics/core/setup/oea_build.sh



2.5) Make the script executable, then execute it (use the following commands):

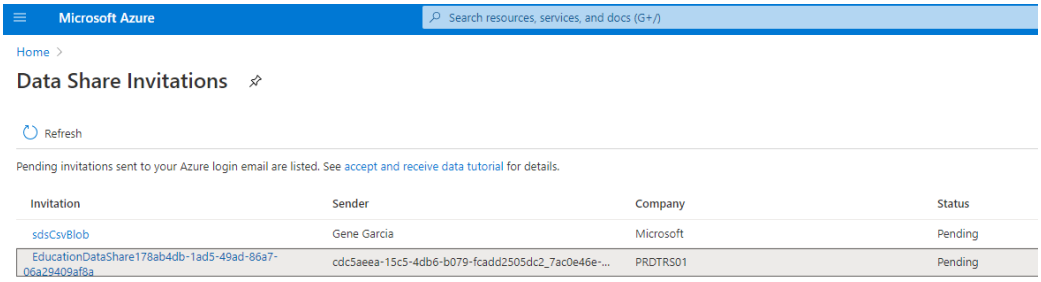
```
cd ./OpenEduAnalytics/core/setup
chmod +x oea_build.sh
./oea_build.sh
```

This will take a few minutes to complete, and will result in the provisioning of the following Azure resources:

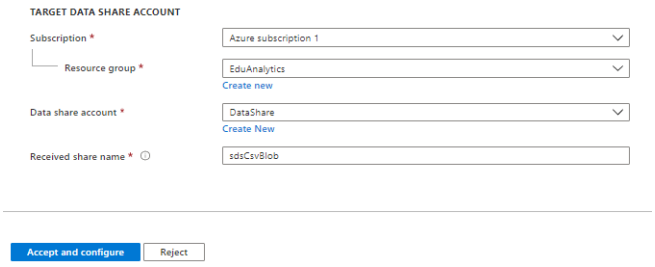
- EduAnalytics resource group
- storage account with 3 storage containers (named stage1, stage2, and stage3)
- Azure Synapse workspace

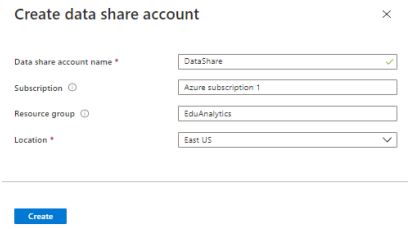
3) Setup Azure Data Share

3.1) In Azure portal, search for “Data Share Invitations” then click on the link to arrive at the Data Share Invitations page, and click on the invitation for EducationDataShare.

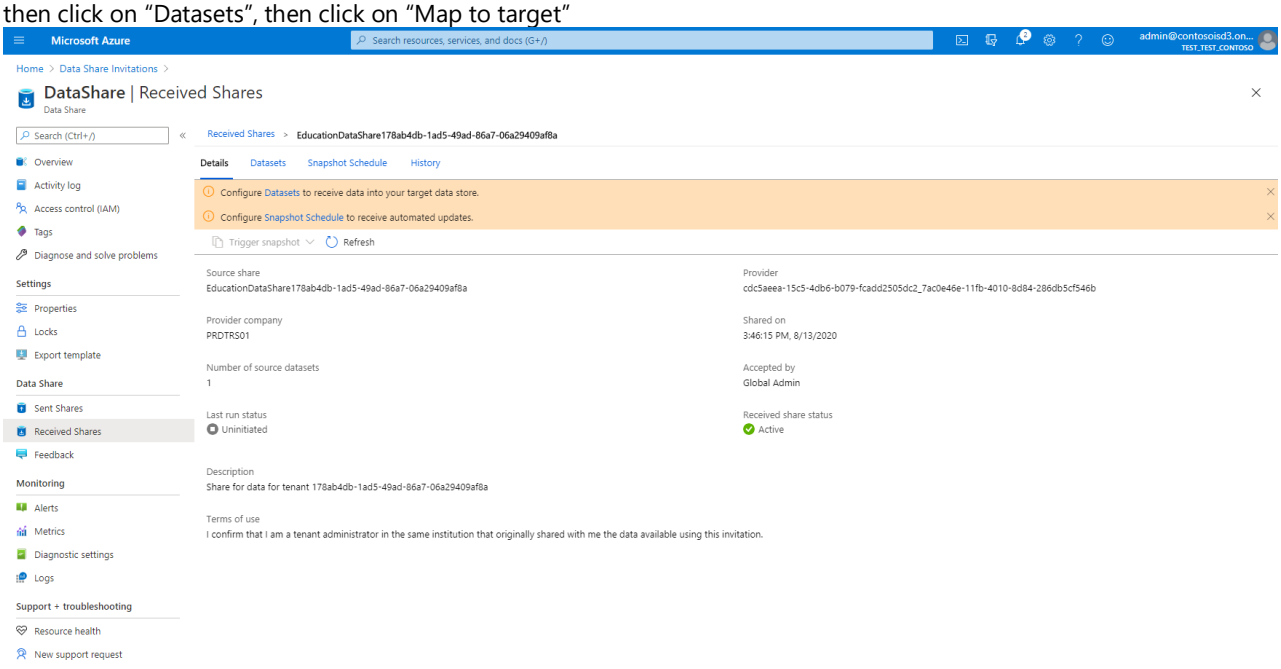


2.5) For “Resource group”, select “EduAnalytics” for the resource group.
For “Data share account”, click “Create new”, type “DataShare” for the name of the data share account and click “Create”.
Then click on “Accept and Configure”

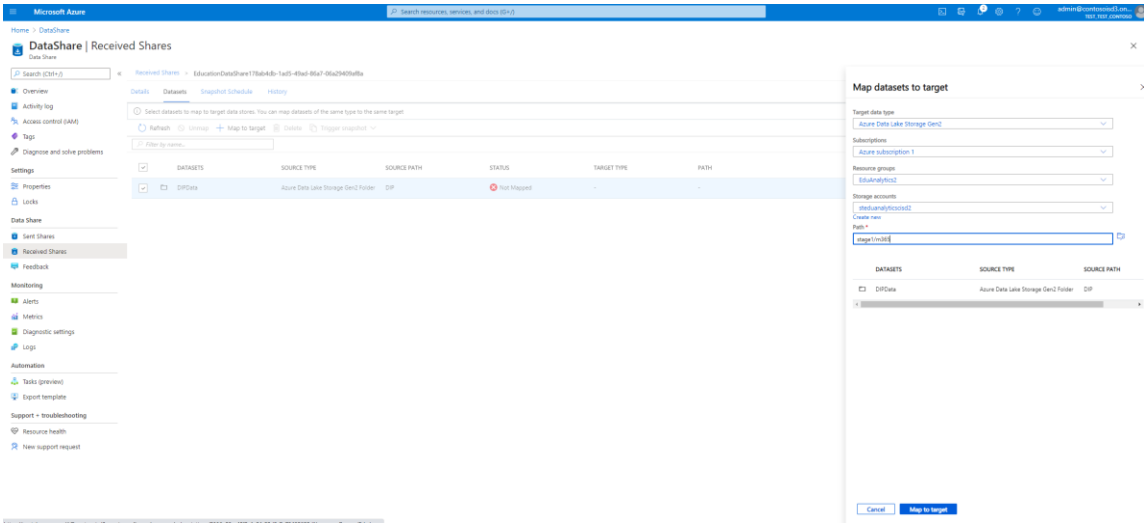




2.6) Go to the DataShare resource created in the last step, click on “View received shares”, then click on the EducationDataShare link, then click on “Datasets”, then click on “Map to target”

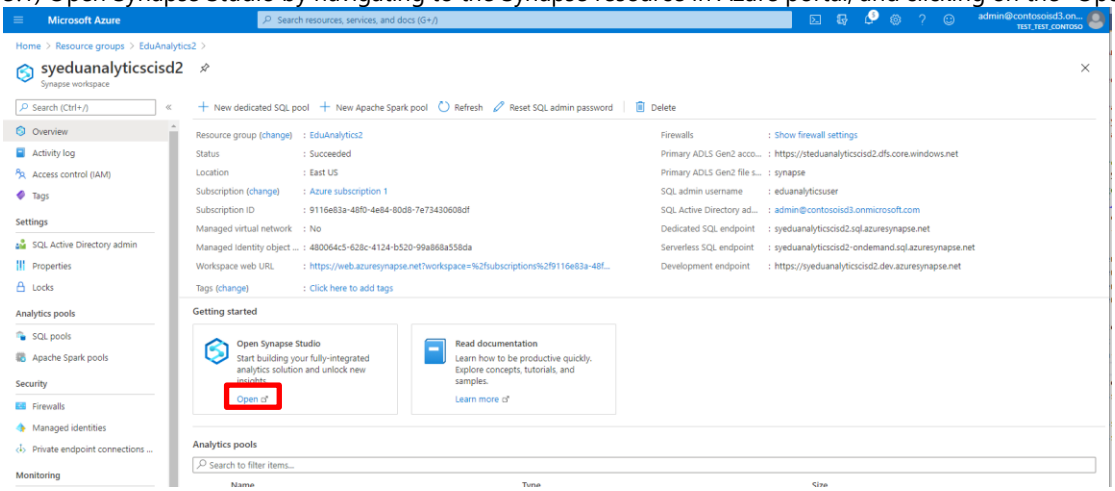


2.7) Select the resource group and storage account created earlier and enter “stage1/m365” for Path (this container is for the purpose of receiving usage data from M365), then click “Map to target” button.


Microsoft Azure
DataShare | Received Shares
Map datasets to target
Target data type
Azure Data Lake Storage Gen2
Subscriptions
Azure subscription 1
Resource groups
EduAnalytics2
Storage accounts
Path *
DATASETS
SOURCE TYPE
SOURCE PATH
DIPData
Azure Data Lake Storage Gen2 Folder
DIP
Cancel
Map to target

4) Setup Synapse

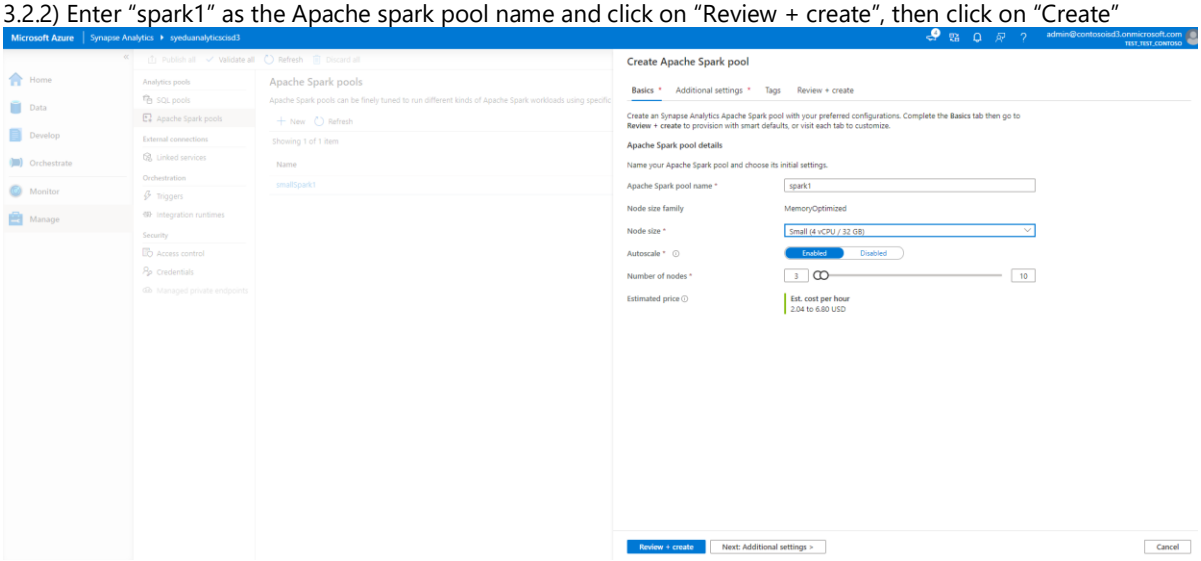
3.1) Open Synapse Studio by navigating to the Synapse resource in Azure portal, and clicking on the “Open” link:



3.2) Create an Apache Spark pool

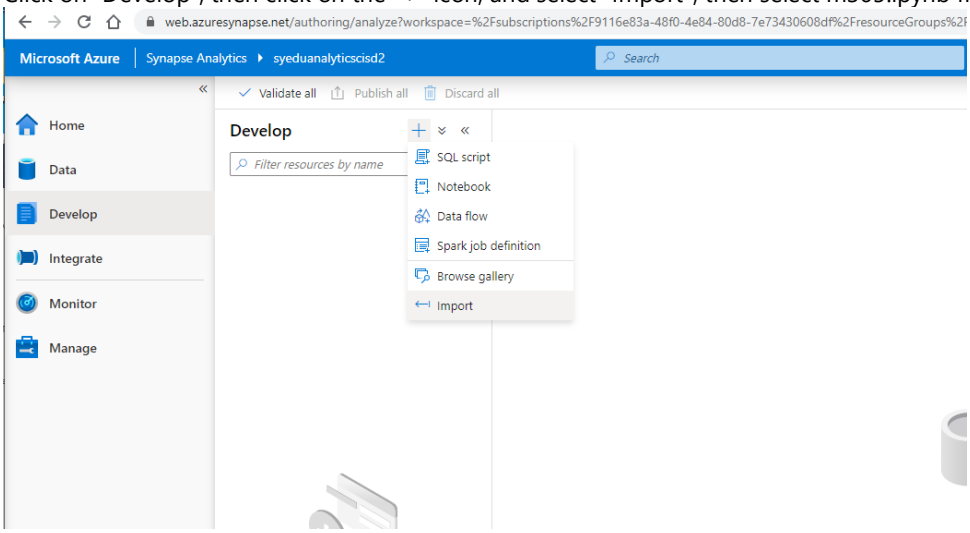
3.2.1) Click on “Manage”, then click on “Apache Spark pools”, then on “New”

3.2.2) Enter “spark1” as the Apache spark pool name and click on “Review + create”, then click on “Create”

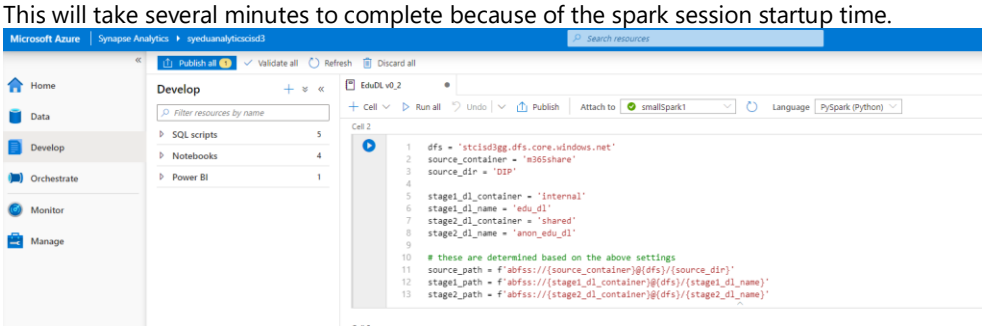


3.3) Import m365 Notebook

Click on “Develop”, then click on the “+” icon, and select “Import”, then select m365.ipynb from your local file system.



3.4) Ingest m365 data into stage 2 and stage 3 data lakes
3.4.1) Click on “Develop“, then click on m365 to open the notebook
3.4.2) Select “spark1” in the “Attach to” dropdown and select “PySpark (Python)” in the Language dropdown.
3.4.3) Click on the play icon under the heading “Cell 2” to execute the contents of the script in Cell 2.
This will take several minutes to complete because of the spark session startup time.

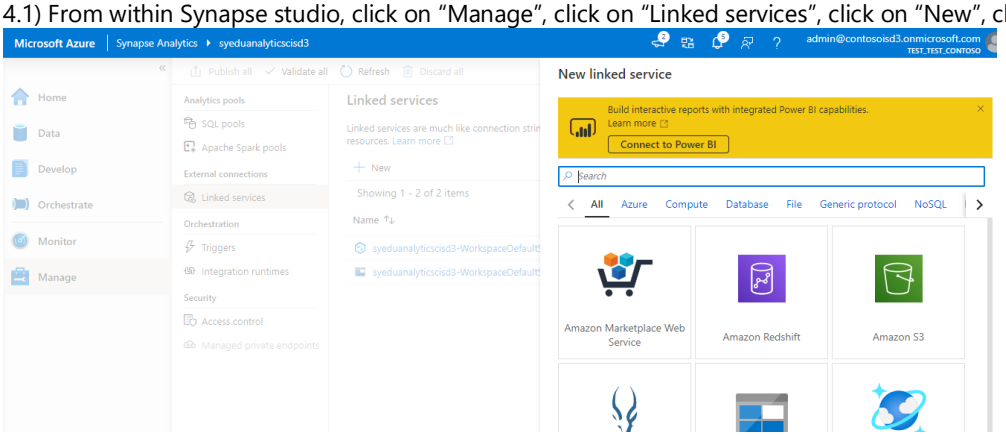


3.4.4) Click on the play icon for Cell 3 and wait for it to complete.
3.4.5) Click on the play icon for Cell 4 and wait for it to complete.
3.4.6) Click on the play icon for Cell 5 and wait for it to complete.
3.4.7) Click on the play icon for Cell 6 and wait for it to complete.

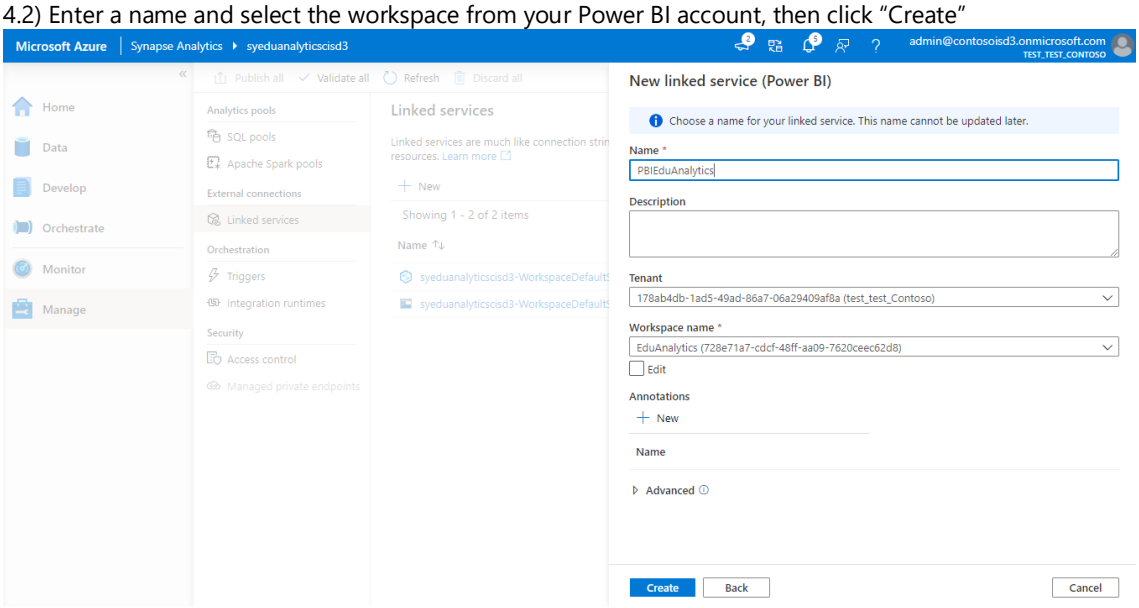
4) Setup Power BI

In order to setup Power BI for use within Azure Synapse Analytics, you must already have a Power BI account, and a Power BI workspace created within that account.

4.1) From within Synapse studio, click on “Manage”, click on “Linked services”, click on “New”, click on “Connect to Power BI”

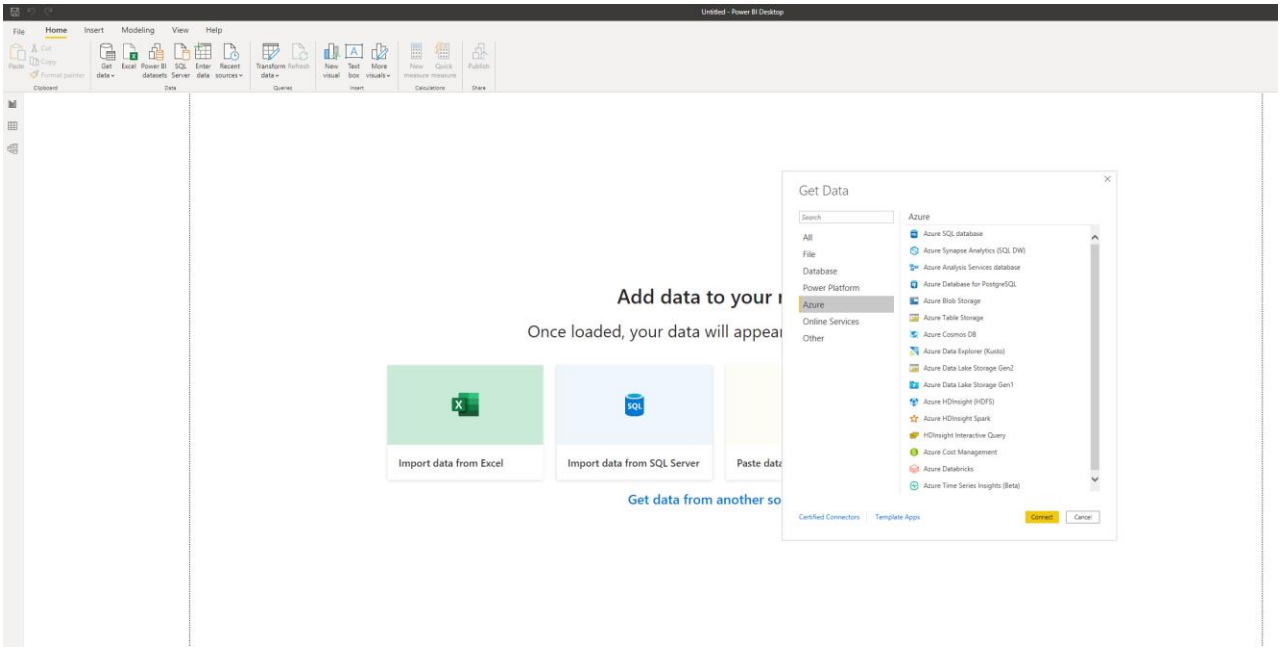


4.2) Enter a name and select the workspace from your Power BI account, then click “Create”



4.3) Create a PowerBI Report

4.3.1) Open Power BI Desktop, select “Get data”



4.3.2) Select “Azure SQL database”.