

Open Edu Analytics Implementation Guide

Published: November, 2020

| | |
|--|---|
| Introduction | 1 |
| 1) Setup Data Share in School Data Sync..... | 1 |
| 2) Setup Resources in Azure Portal..... | 2 |
| 3) Setup Azure Data Share..... | 4 |
| 4) Setup Synapse | 6 |
| 4) Setup Power BI | 8 |

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>

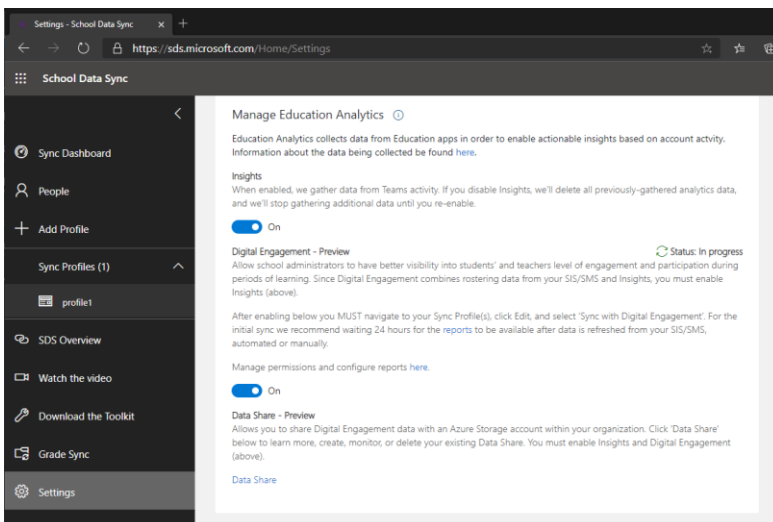
The following are required prior to implementing the steps in this guide:

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

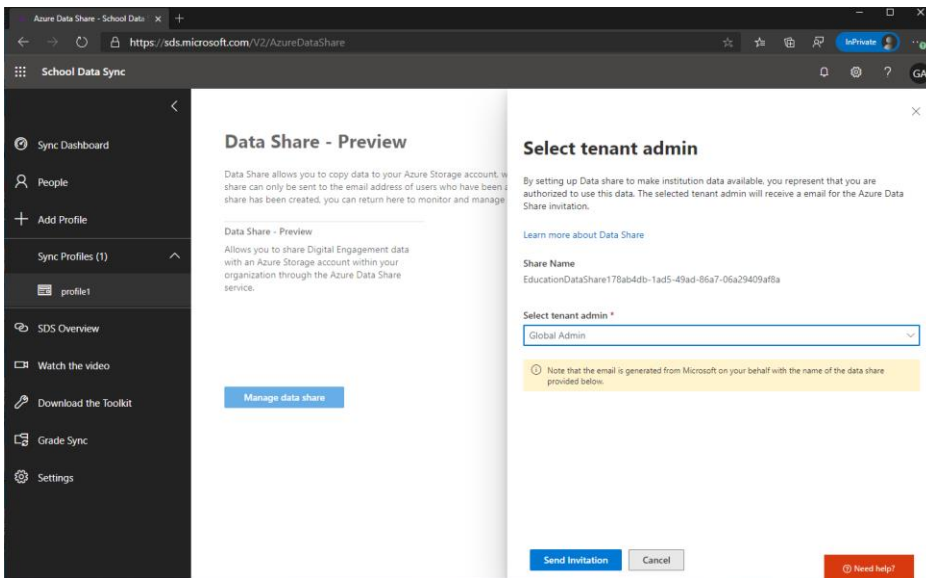
1) Setup Data Share in School Data Sync

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.



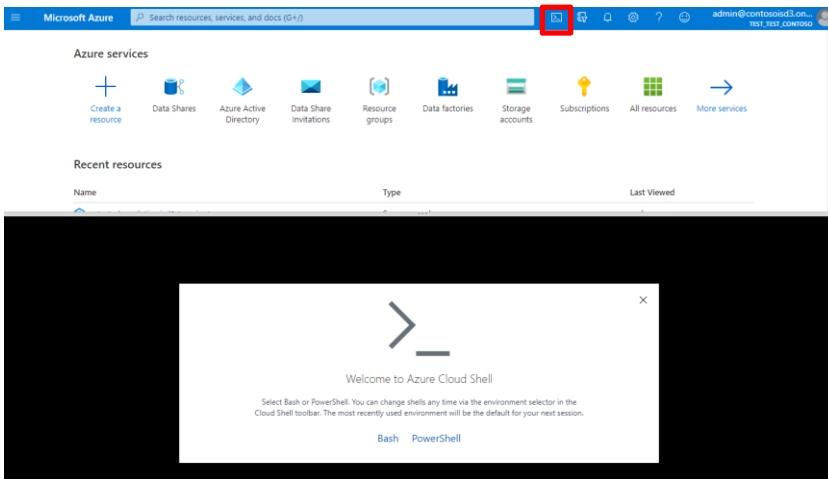
1.2) Click on “Manage data share”, then select the tenant admin who is doing the setup, and click “Send Invitation”



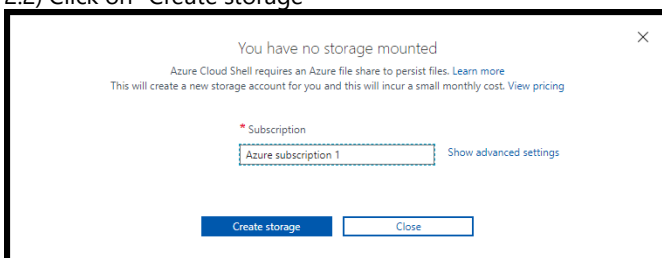
2) Setup Resources in Azure Portal

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".

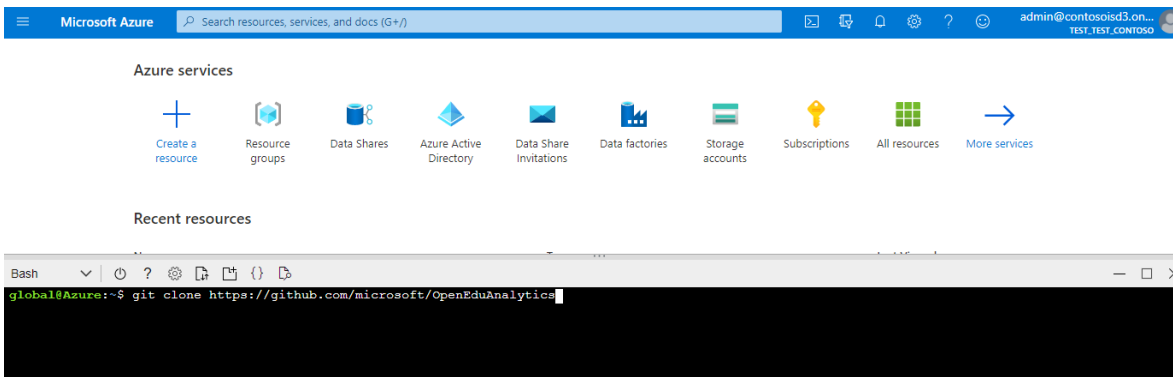


2.2) Click on "Create storage"



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

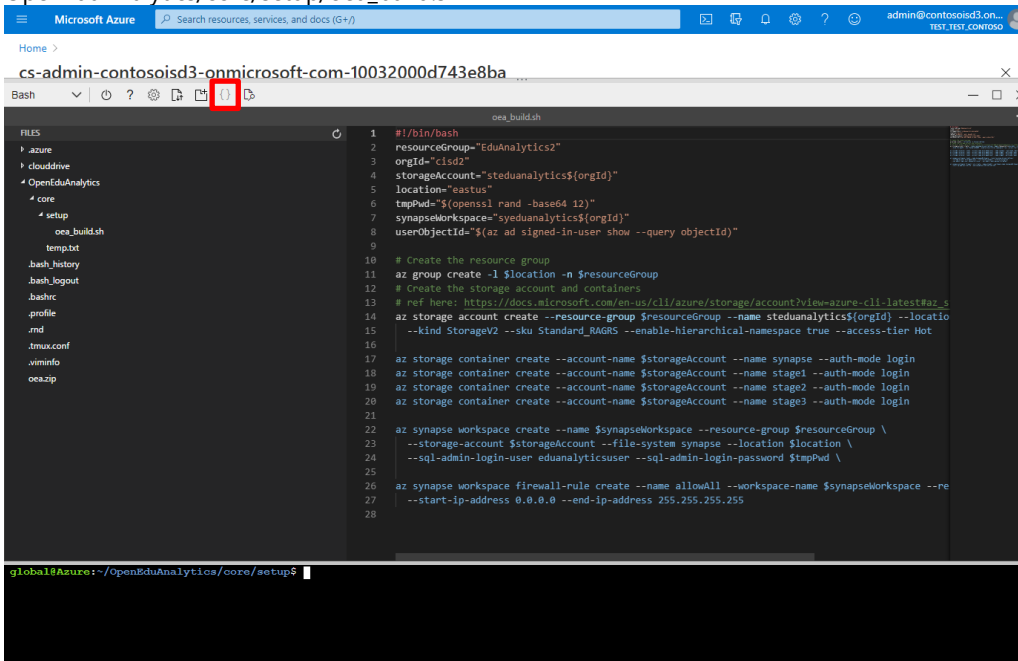
```
cd clouddrive
git clone https://github.com/microsoft/OpenEduAnalytics
```



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.

| Invitation | Sender | Company | Status |
|--|---|-----------|---------|
| sdsCsvBlob | Gene Garcia | Microsoft | Pending |
| EducationDataShare178ab4db-1ad5-49ad-86a7-06a29409af8a | cdc5aeea-15c5-4db6-b079-fcadd2505dc2_7ac0e46e-11fb-4010-8d84-286db5cf546b | PRDTRS01 | Pending |

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"

Subscription * Azure subscription 1

Resource group * EduAnalytics

Data share account * DataShare

Received share name * sdsCsvBlob

Create data share account

Data share account name * DataShare

Subscription Azure subscription 1

Resource group EduAnalytics

Location * East US

Create

Accept and configure Reject

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"

Overview Activity log Access control (IAM) Tags Diagnose and solve problems Settings Properties Locks Export template Data Share Sent Shares Received Shares Feedback Monitoring Alerts Metrics Diagnostic settings Logs Support + troubleshooting Resource health New support request

Received Shares > EducationDataShare178ab4db-1ad5-49ad-86a7-06a29409af8a

Configure Datasets to receive data into your target data store.

Configure Snapshot Schedule to receive automated updates.

Trigger snapshot Refresh

| Source share | Provider |
|--|---|
| EducationDataShare178ab4db-1ad5-49ad-86a7-06a29409af8a | cdc5aeea-15c5-4db6-b079-fcadd2505dc2_7ac0e46e-11fb-4010-8d84-286db5cf546b |

Provider company PRDTRS01

Shared on 3:46:15 PM, 8/13/2020

Number of source datasets 1

Accepted by Global Admin

Last run status Uninitiated

Received share status Active

Description Share for data for tenant 178ab4db-1ad5-49ad-86a7-06a29409af8a

Terms of use I confirm that I am a tenant administrator in the same institution that originally shared with me the data available using this invitation.

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

Home > DataShare > Received Shares

Search (Ctrl+J)

Received Shares > EducationDataShare173ab4d8-1a55-45ed-85a7-05a29405a6fa

Overview Activity log Access control (IAM) Tags Diagnose and solve problems

Settings Properties Locks Data Share Sent Shares Received Shares Feedback

Monitoring Alerts Metrics Diagnostic settings Logs Automation Tasks (preview) Export template Support + Troubleshooting Resource health New support request

Select datasets to map to target data stores. You can map datasets of the same type to the same target.

Refresh Unmap Map to target Datasets Trigger snapshot

Filter by name

| DATASETS | SOURCE TYPE | SOURCE PATH | STATUS | TARGET TYPE | PATH |
|-------------------------------------|-------------|-------------------------------------|------------|-------------|------|
| <input checked="" type="checkbox"/> | DPData | Azure Data Lake Storage Gen2 Folder | Not Mapped | - | - |

Map datasets to target

Target data type: Azure Data Lake Storage Gen2

Subscriptions: Azure subscription 1

Resource groups: Subscription1RG1

Storage accounts: Subscription1Storage002

Create new: none

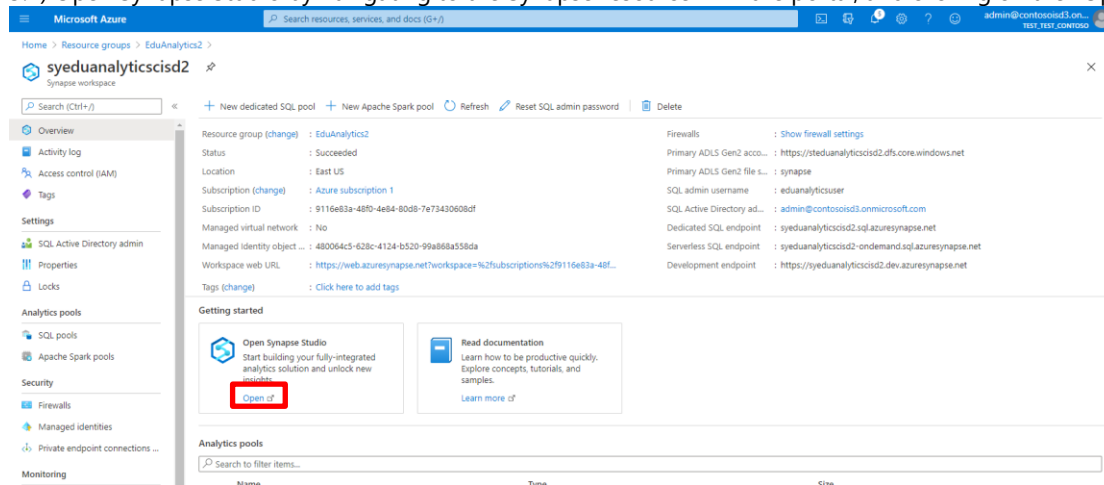
Path: none

| DATASETS | SOURCE TYPE | SOURCE PATH |
|-------------------------------------|-------------|-------------------------------------|
| <input checked="" type="checkbox"/> | DPData | Azure Data Lake Storage Gen2 Folder |

Cancel Map to target

4) Setup Synapse

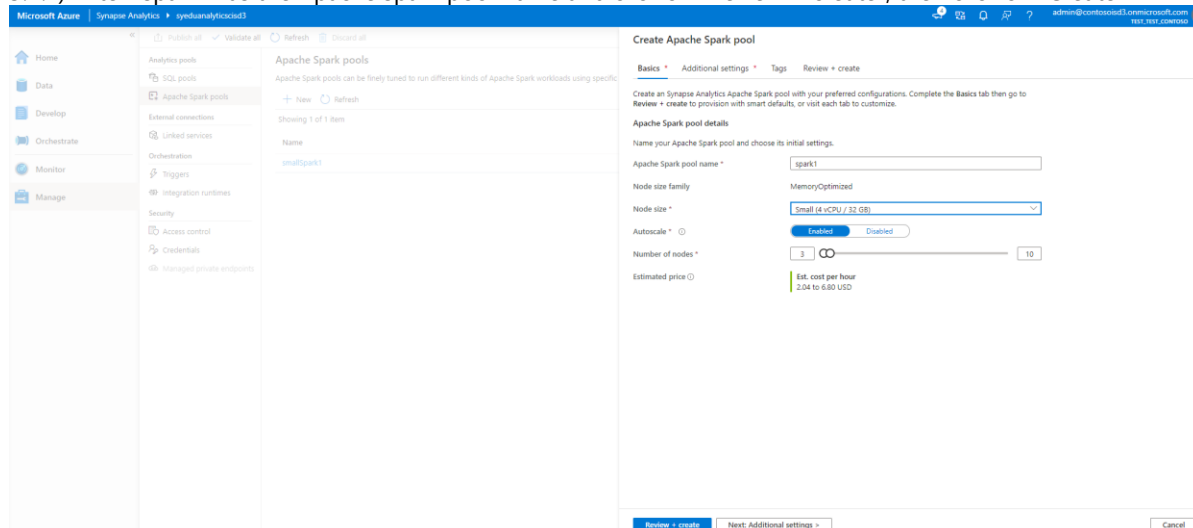
3.1) Open Synapse Studio by navigating to the Synapse resource in Azure portal, and clicking on the “Open” link:



3.2) Create an Apache Spark pool

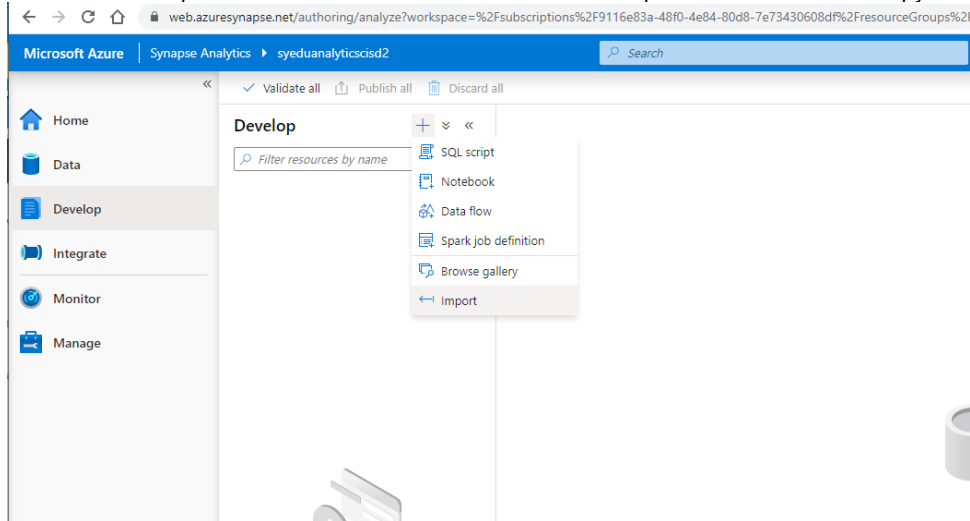
3.2.1) Click on “Manage”, then click on “Apache Spark pools”, then on “New”

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.



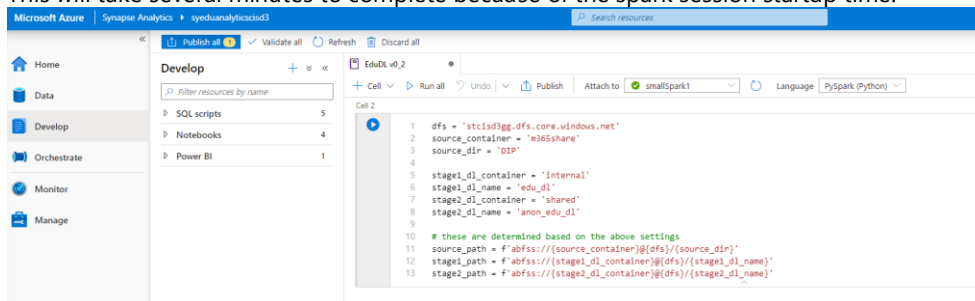
3.4) Ingest m365 data into stage 2 and stage 3 data lakes

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.

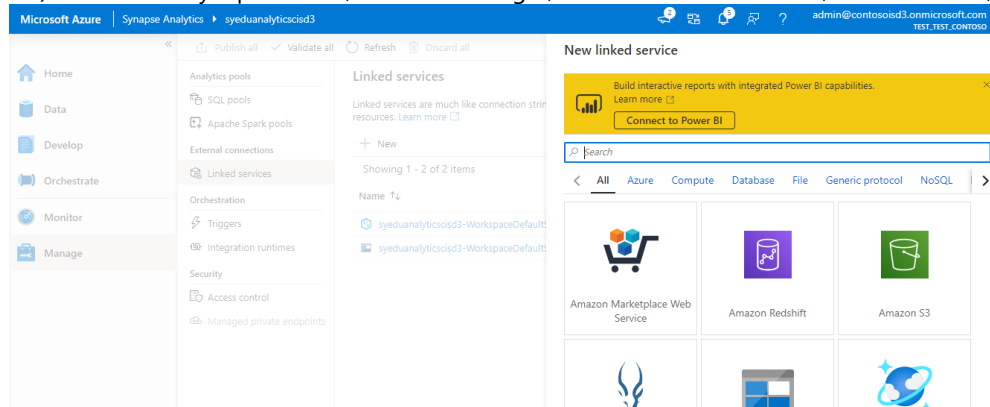
3.4.6) Click on the play icon for Cell 5 and wait for it to complete.

3.4.7) Click on the play icon for Cell 6 and wait for it to complete.

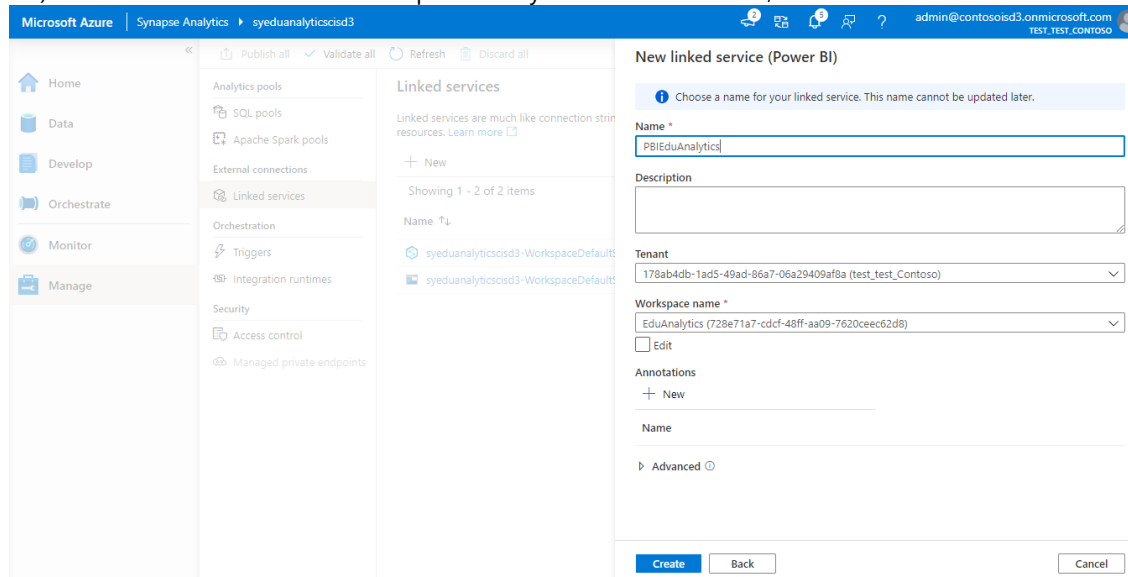
4) Setup Power BI

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"

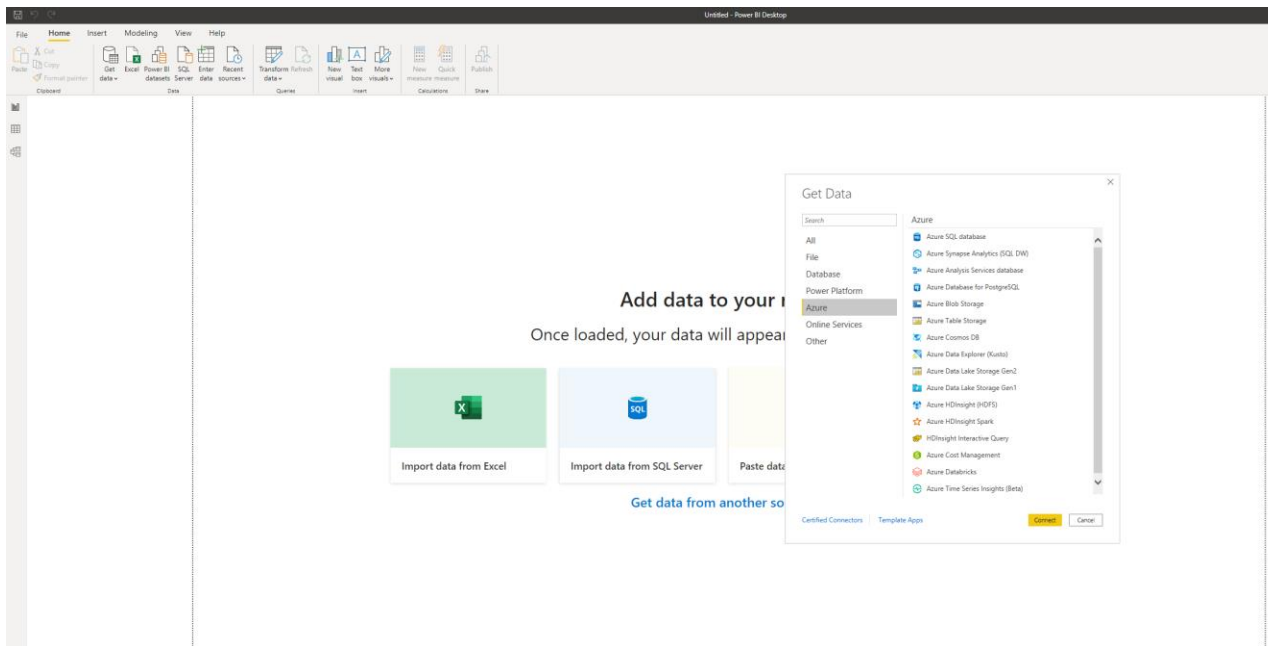


4.2) Enter a name and select the workspace from your Power BI account, then click "Create"

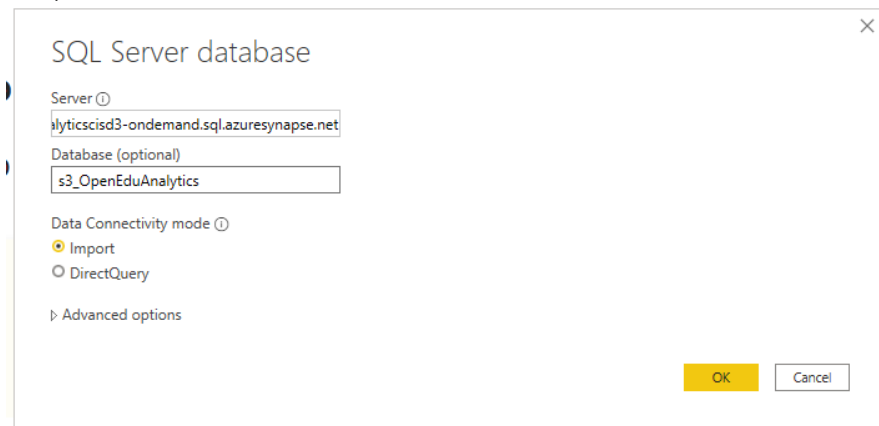


4.3) Create a PowerBI Report

4.3.1) Open Power BI Desktop, select "Get data"



4.3.2) Select "Azure SQL database".



This is a preliminary document and may be changed substantially prior to final commercial release of the software described herein. The information contained in this document represents the current view of Microsoft Corporation on the issues discussed as of the date of publication. Because Microsoft must respond to changing market conditions, it should not be interpreted to be a commitment on the part of Microsoft, and Microsoft cannot guarantee the accuracy of any information presented after the date of publication. This white paper is for informational purposes only. Microsoft makes no warranties, express or implied, in this document. Complying with all applicable copyright laws is the responsibility of the user. Without limiting the rights under copyright, no part of this document may be reproduced, stored in, or introduced into a retrieval system, or transmitted in any form or by any means (electronic, mechanical, photocopying, recording, or otherwise), or for any purpose, without the express written permission of Microsoft Corporation. Microsoft may have patents, patent applications, trademarks, copyrights, or other intellectual property rights covering subject matter in this document. Except as expressly provided in any written license agreement from Microsoft, the furnishing of this document does not give you any license to these patents, trademarks, copyrights, or other intellectual property.

© 2020 Microsoft Corporation. All rights reserved