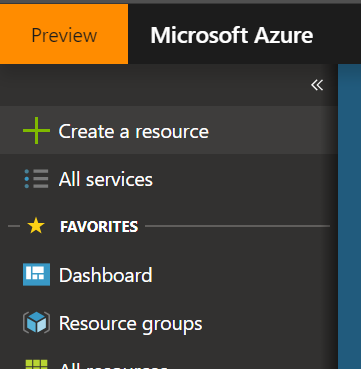
Loan Origination – Use-case Exploration

# Creating an Azure Databricks Workspace

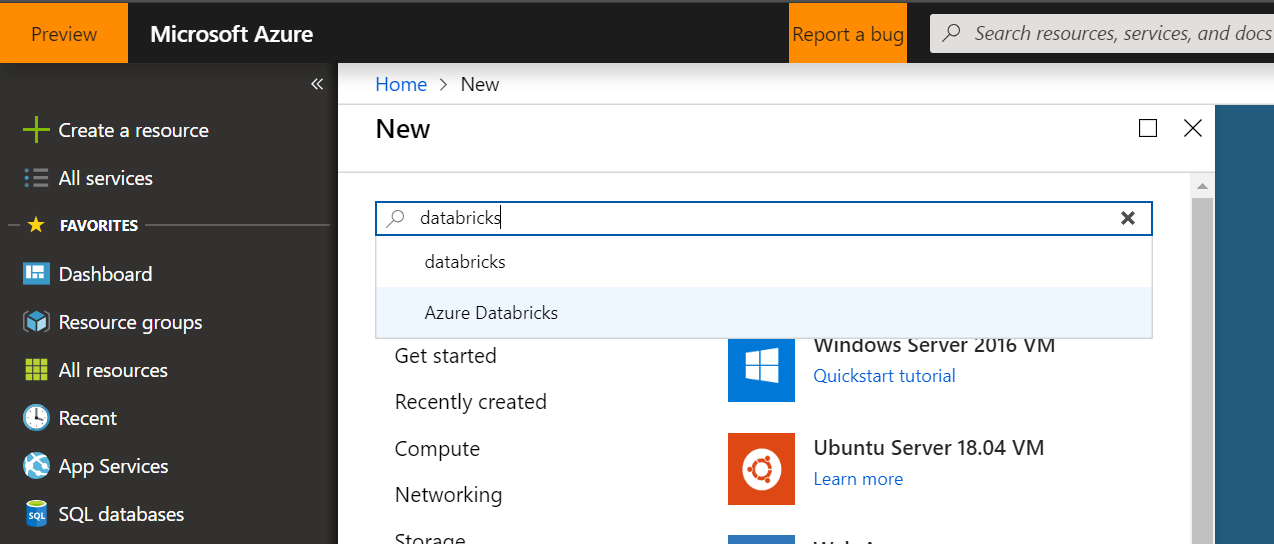
Log on to the Auzre Portal via: <https://ms.portal.azure.com>

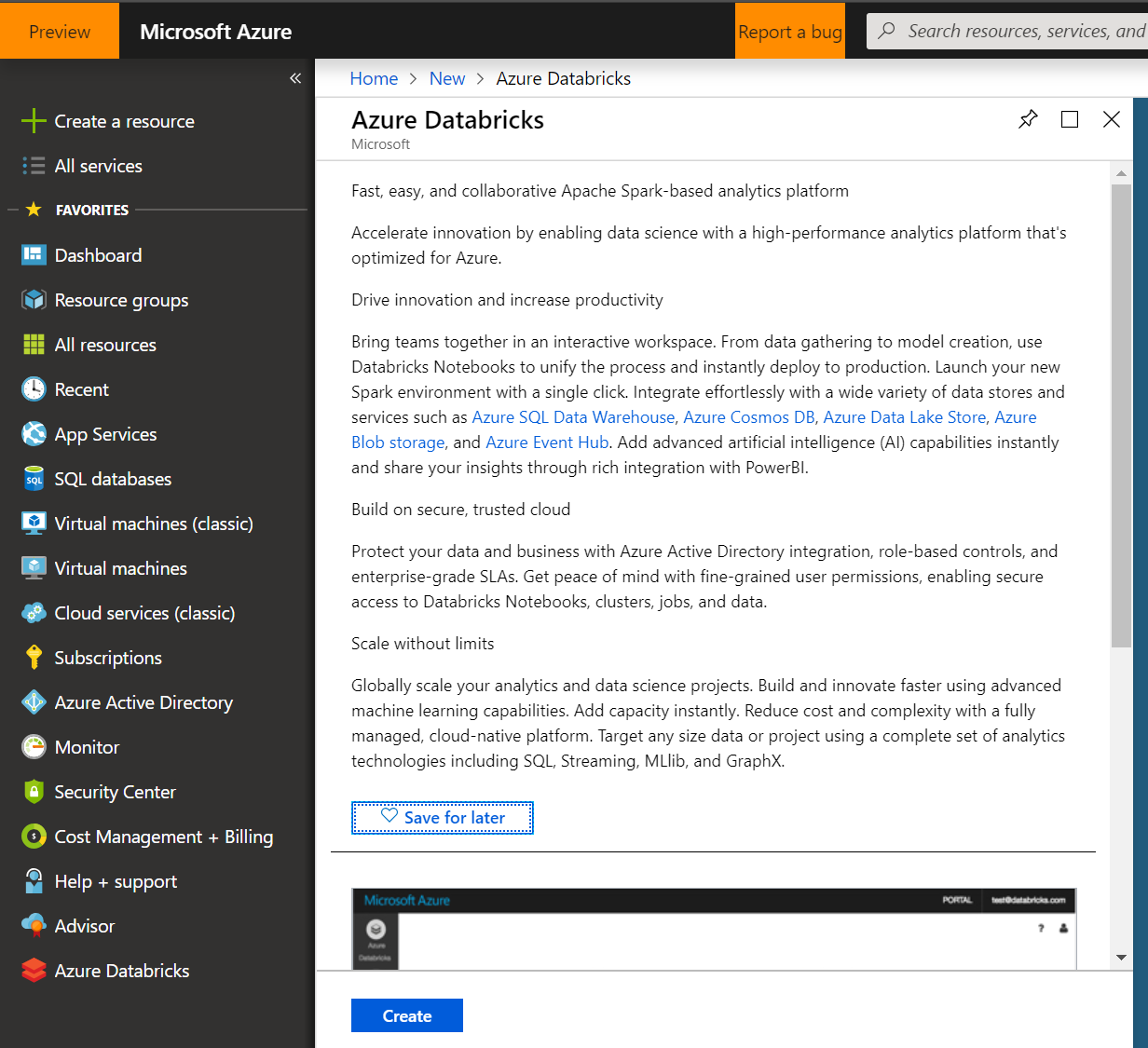
Use your corporate ID and password to log in.

Create a new Azure DataBricks Resource



In the search bar, type *databricks*, then select **Azure Databricks**

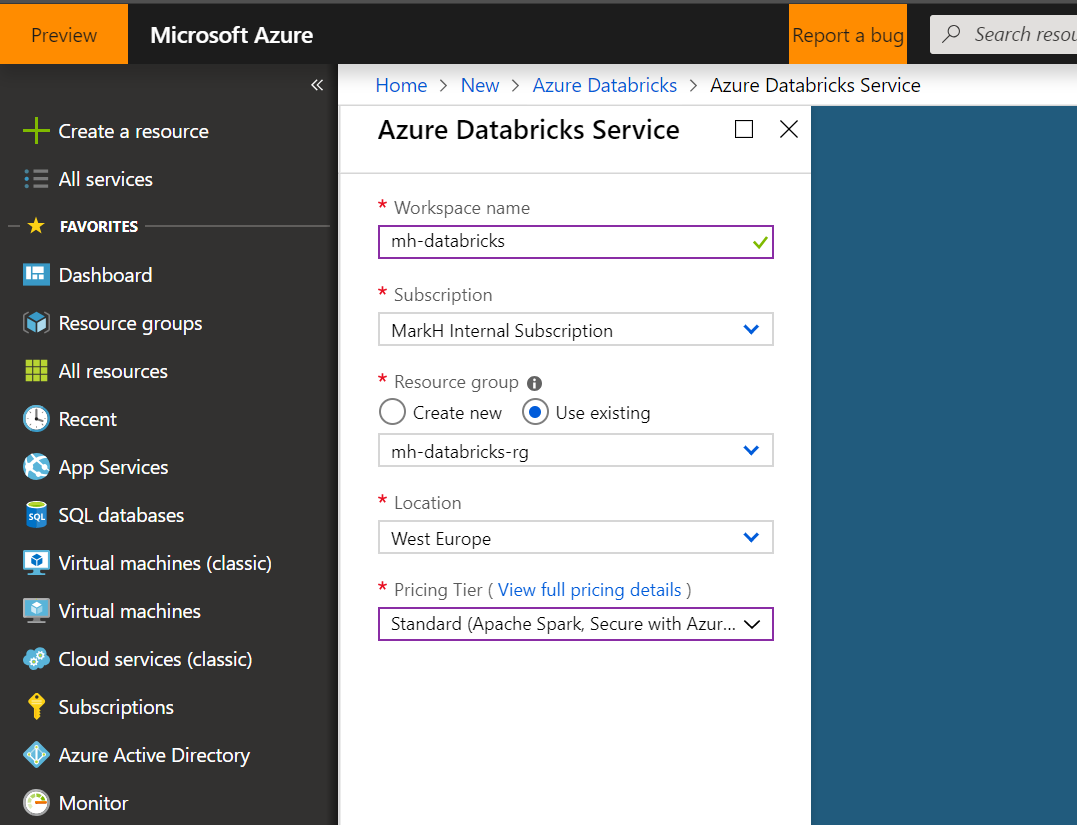




At the bottom, click on **Create**. A new blade opens with an empty form. Fill out the following fields:

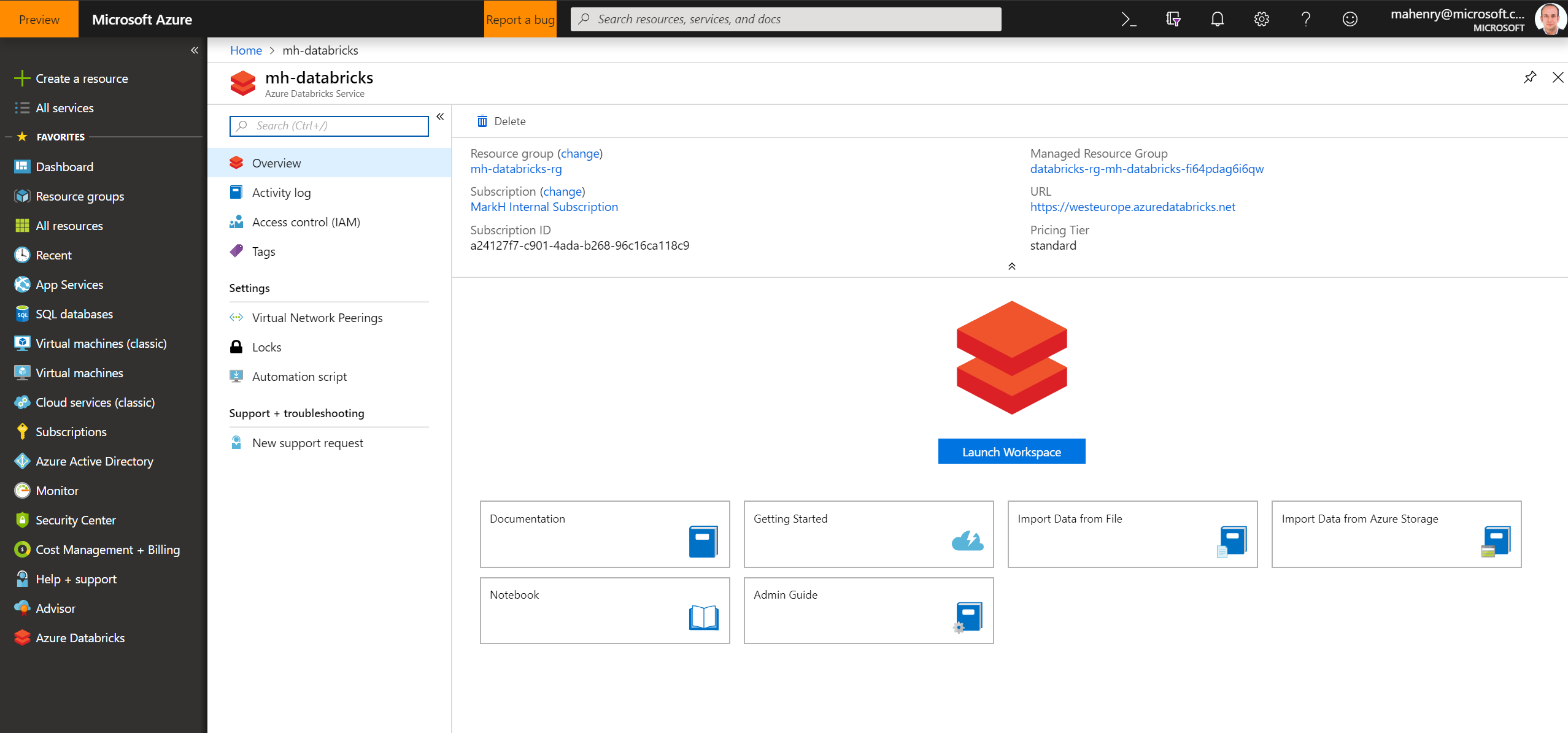
* Workspace name: provide a unique name for your cluster.
* Create a new Resource Group for todays work
* Leave the location as West Europe
* For the pricing tier, select: Premium (Apache Spark, Secure with Azure AD)

The details used in the following screenshot will be referenced when needed. For your work, switch out the values you have used with the ones in the screenshot when needed.

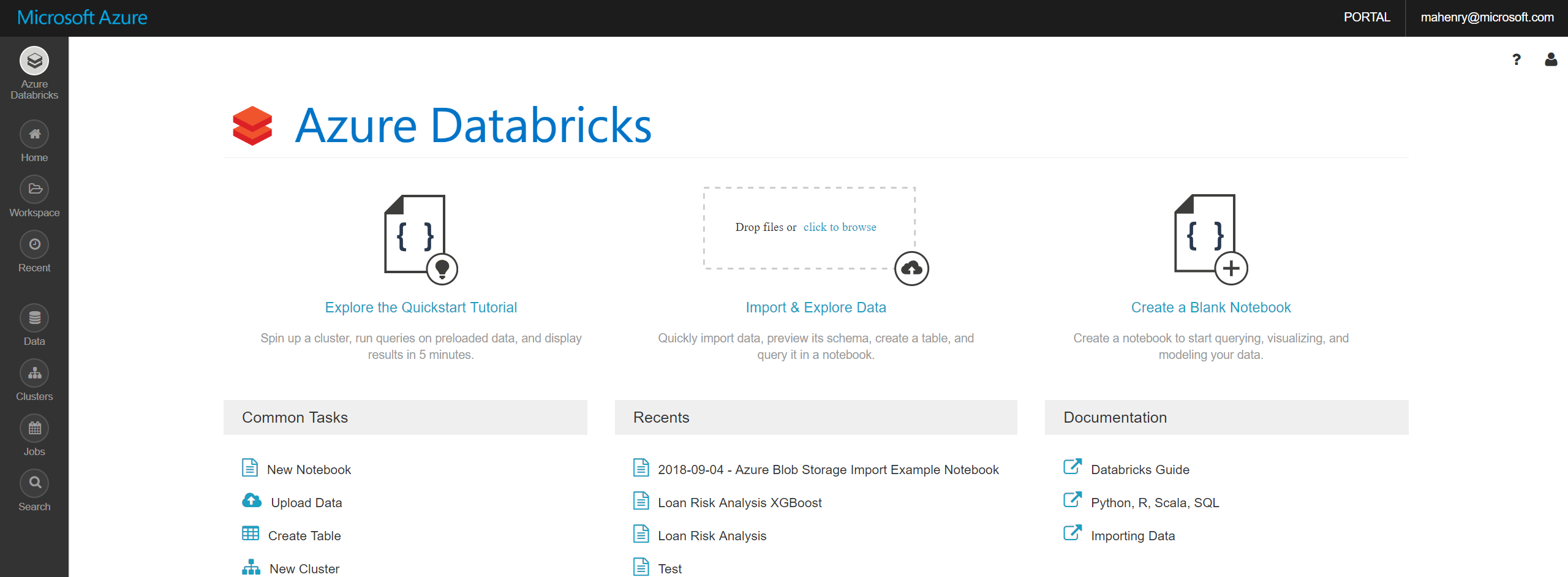


Click on create at the bottom.

It would be easier to add it as a shortcut to your dashboard. When deployment is complete, open the resource, you should get a screen like the below. Click on the **Launch Workspace** button in the middle.

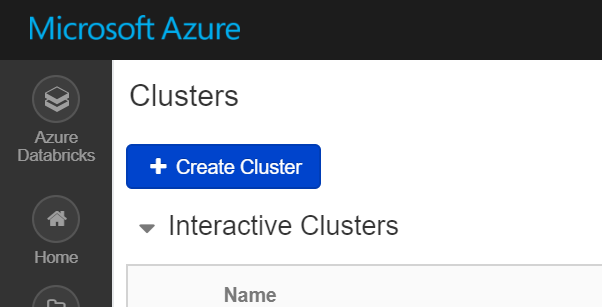
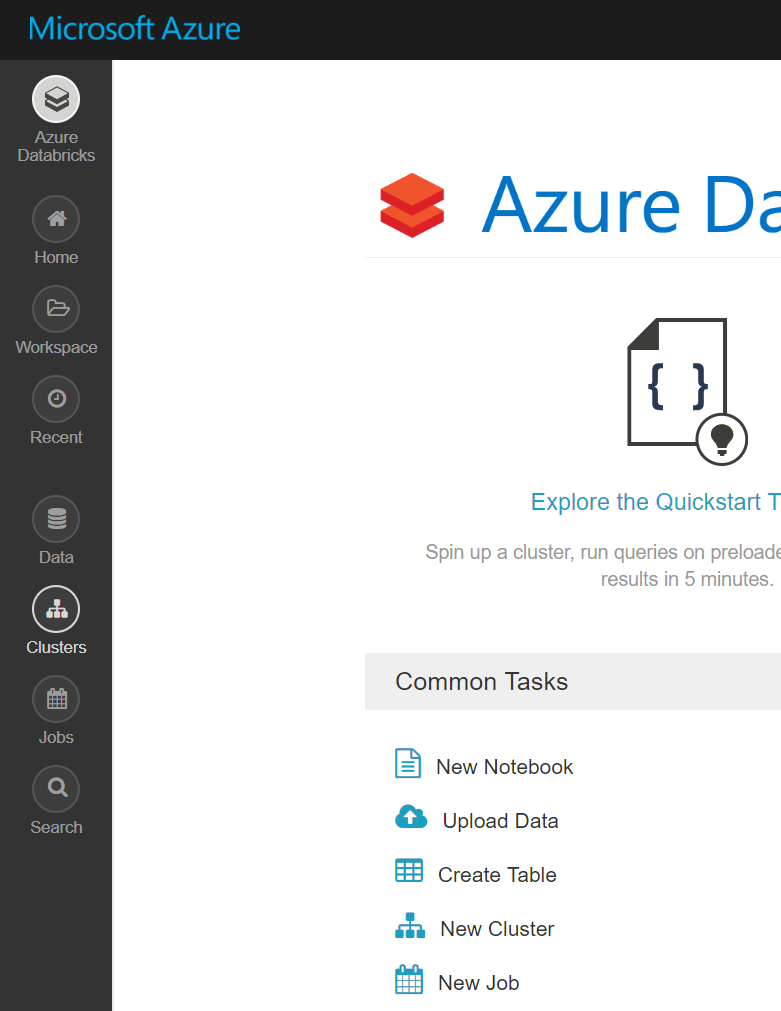


The service should automatically log you on and then show you the home page below.

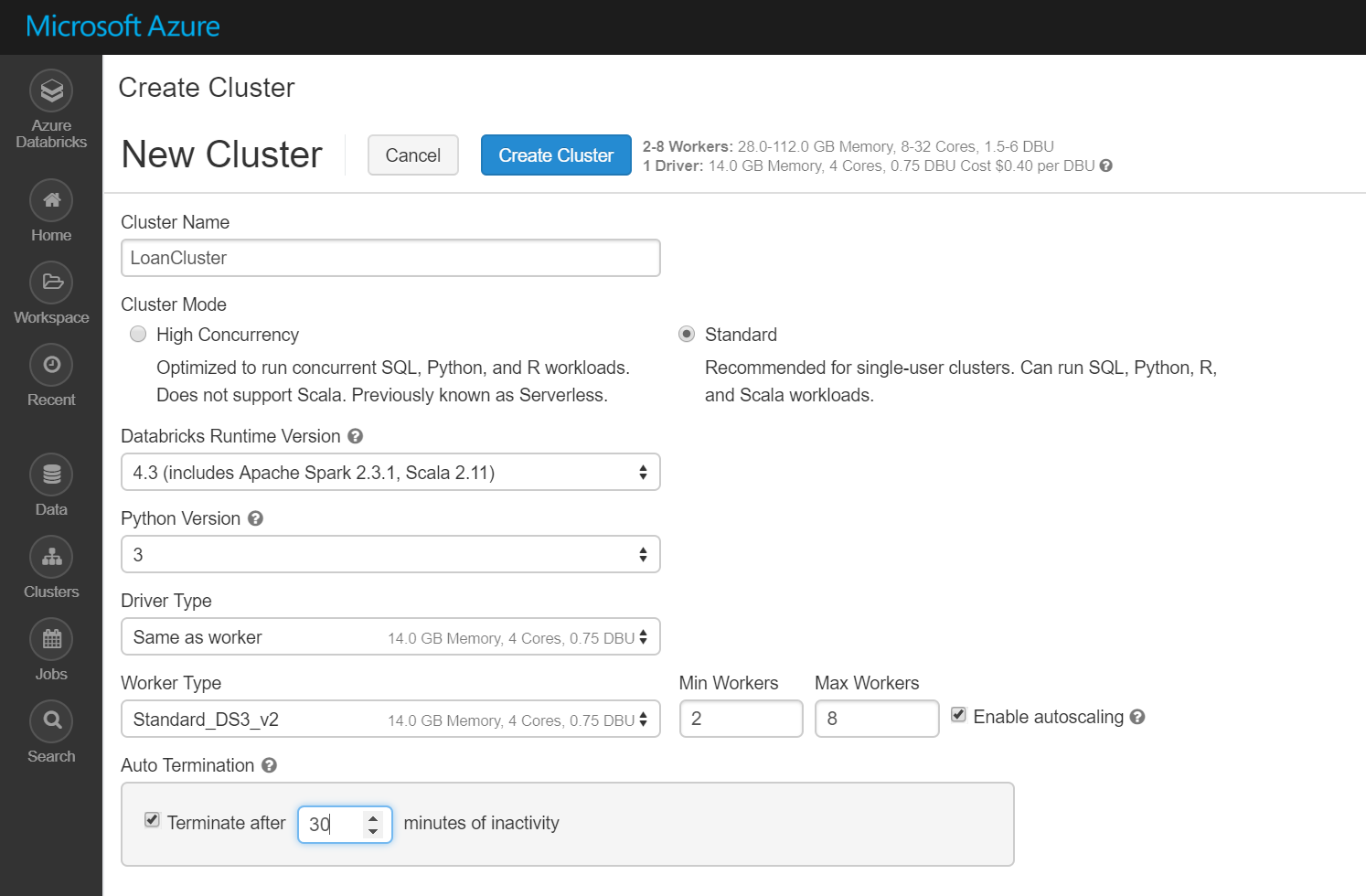


# Spinning up a Cluster

In the menu on the left, click on Clusters



Then click the Create Cluster button. When filling up the details in the new form, make sure to use the same details as in the screenshot below, they are highlighted in red boxes for your convenience. This cluster is called LoanCluster.

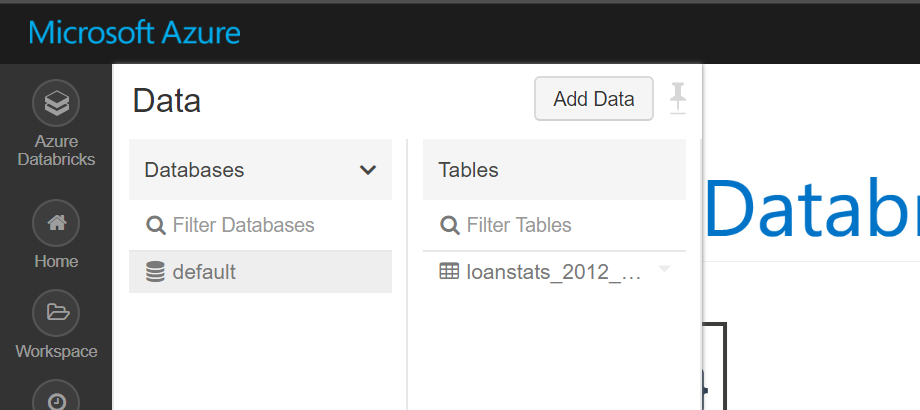
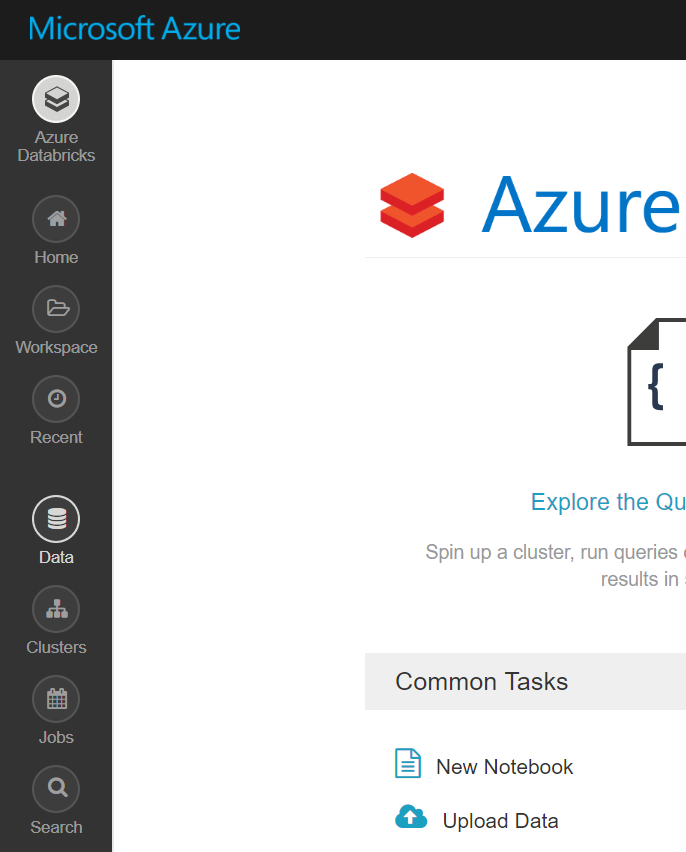


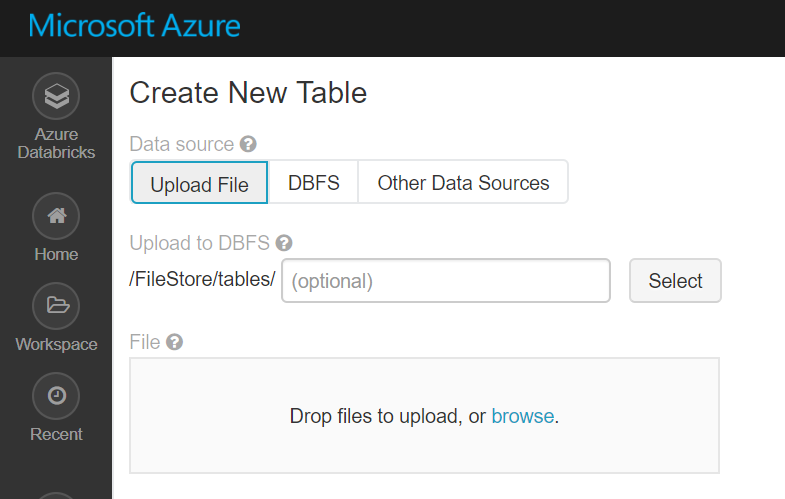
Make sure you select the 4.2 ML (Beta) version, without GPU. This is the only one that comes with XGBoost preinstalled.

Click on the blue Create Cluster button at the top.

# Uploading the data

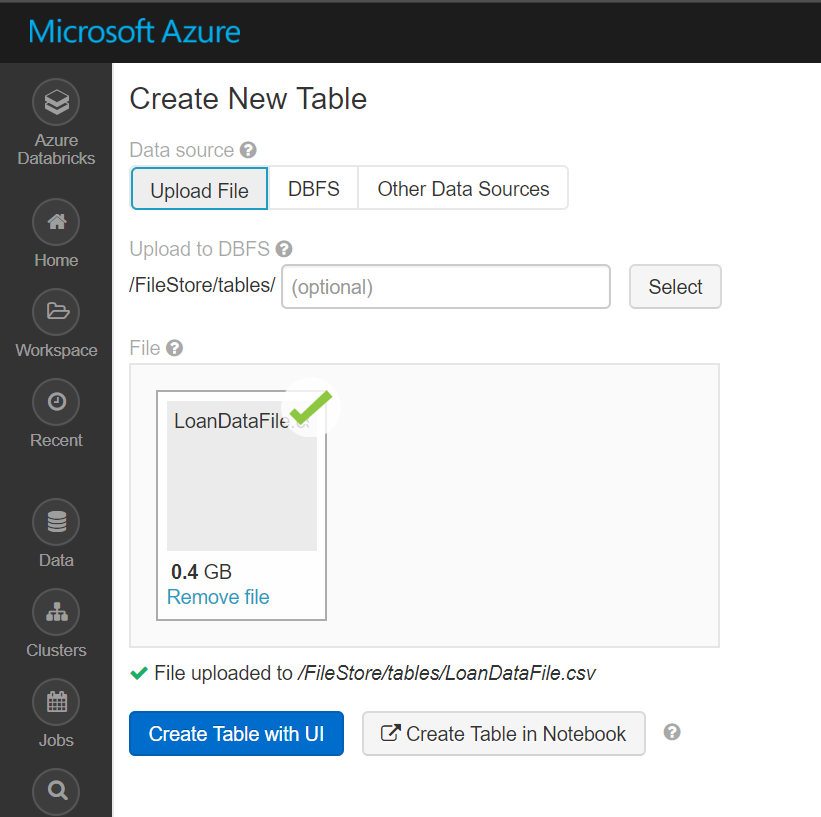
Next we will upload the data and create a Spark Table that will allow us to use the data natively in the code, without worrying about connections.



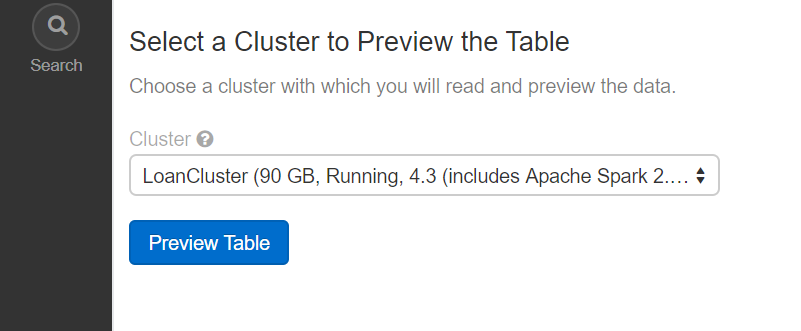


When the first tab is selected, click on Browse and select the **LoanDataFile.csv** file accompanying this file. Leave it to upload.

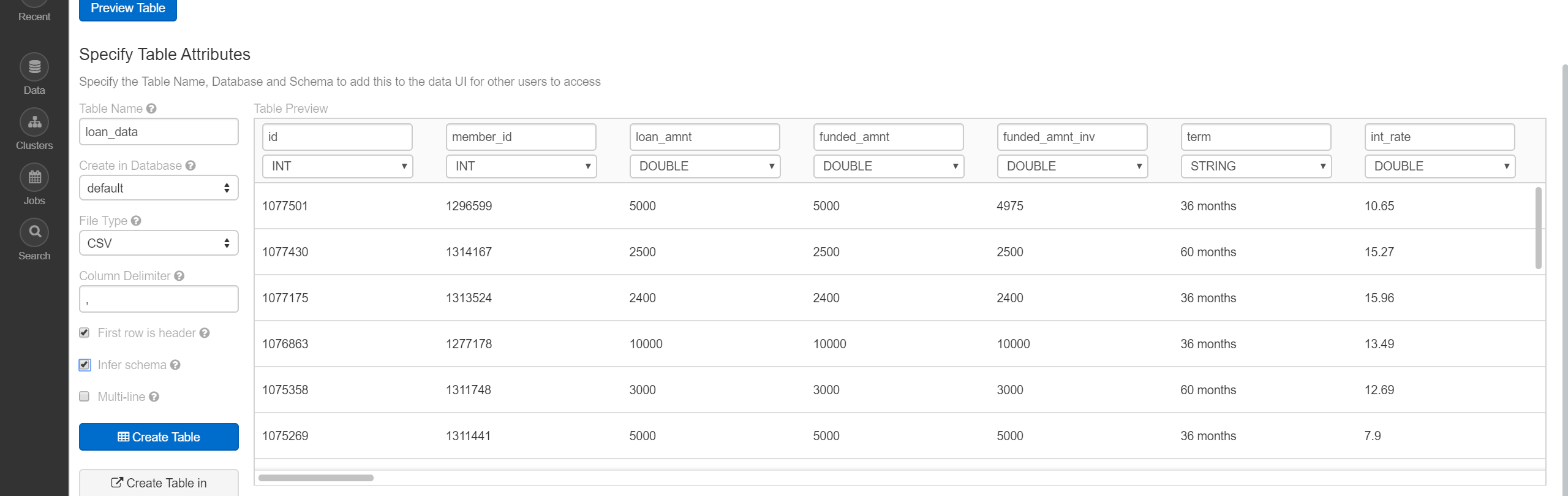
When the upload is complete click on the Create Table with UI button.



You will get a new section, to select a cluster. Select the cluster created in the previous section:



Click on Preview Table. Wait for it to load then scroll down:



Make sure to name the table **loan\_data**. If you choose another name, then please keep it so that you could refer to it in the following sections of the code. Highlighted on the screenshot above are the settings required.

Before moving forward, ensure that the column names and types are similar to the ones above. Check the settings in the red boxes if they aren’t.

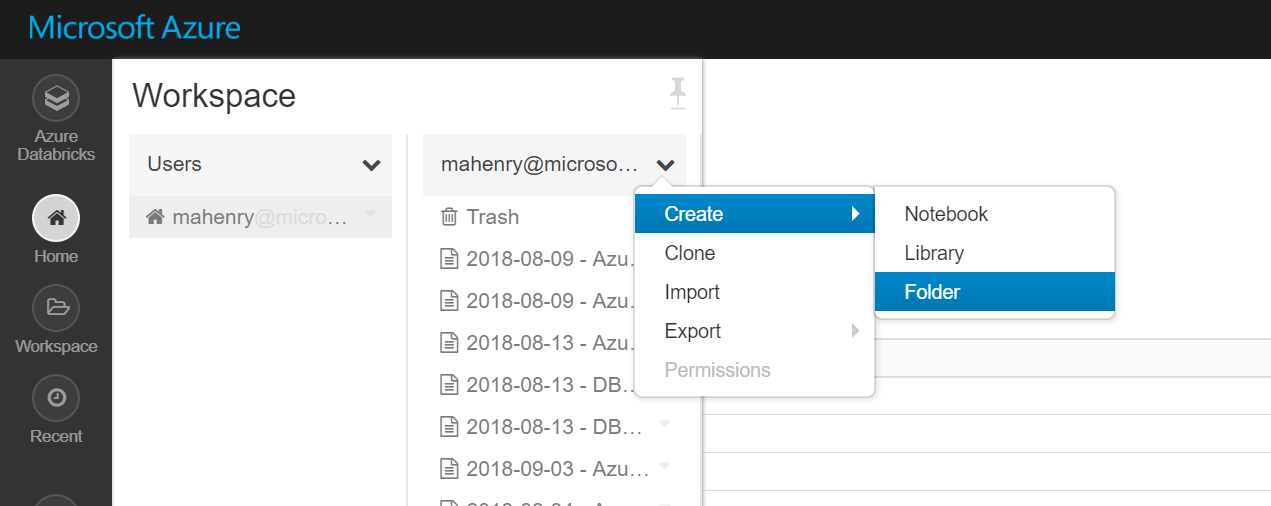
Click on the Create Table button.

A new screen shows up with the tables metadata and a sample of its content.

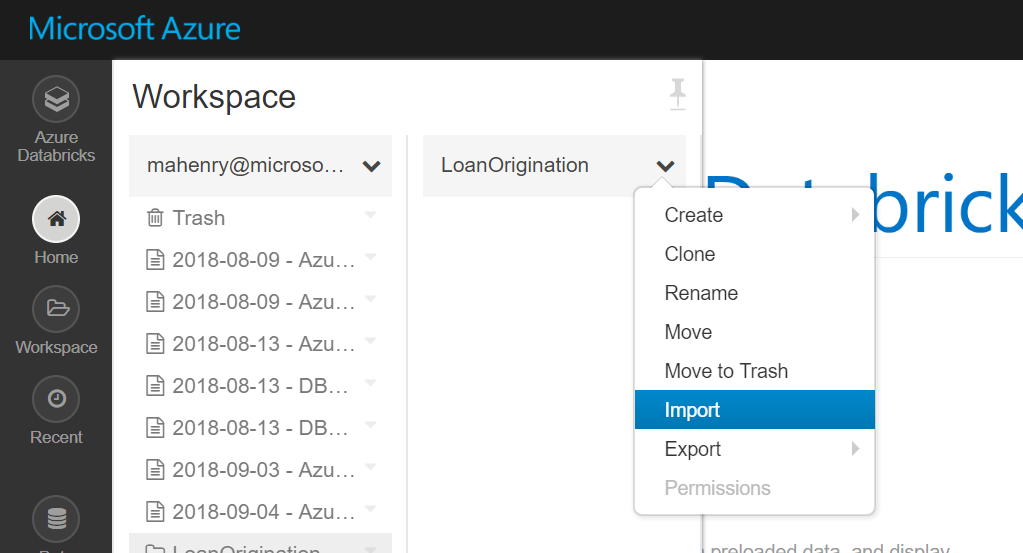
# Upload the code

On the menu on the left, click on Home. You should see a view similar to the one below but with an empty list of files. In the second pane, click the arrow at the top next to your account email. Select Create then Folder.

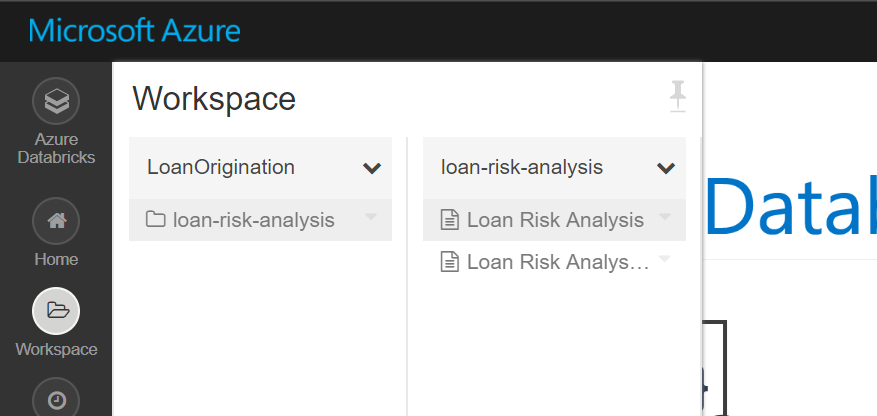
Call this folder Loan Origination.



Click on the newly created folder, it will open a new pane. In the new pane, click on the down arrow and click on import. Now select the dbc file accompanying this file.



You should get the following view on your screen:



Work through the 2 files of code. The first is an exploration for the data. The second uses XGBoost to solve the problem.