ASSIGNMENT------SOLUTION SUBMISSION
ON
AZURE ANALYTICS
BY

NAME : SAI KIRAN ANCHE
ROLL NO: DXC262AB12021
BATCH:DXC-262-ANALYTICS-B12-AZURE
COMPANY – DXC TECHNOLOGY
TRAINING UNDER : MANIPAL PRO LEARN
TRAINER NAME – MR. AJAY KUMAR
DATE OF SUBMISSION : 09-06-2022
NO OF QUESTIONS :11
EMPLOYEE DOMAIN - AZURE ANALYTICS

Assignement - 9th June 2022:
---

1. Explain the steps with screenshots how to AzureSynapse analytics?

2. Explain the steps with screenshots how to SQL Pool in AzureSynapse analytics?

3. Explain the steps with screenshots how to import COVID19 dataset in AzureSynapse analytics

and run sample 500 rows & dispaly the output?

4. Explain the steps with screenshots how to input Boston Safety datasets into AzureSynapse analytics?

using Notebooks ?

5. Explain the steps with screenshots how to create Spark pool in AzureSynapse analytics? ?

6. Explain the steps with screenshots how to create pipeline in AzureSynapse analytics? ?

7. Explain the steps with screenshots how to automate the pipelines in AzureSynapse analytics??

8. Explain the steps with screenshots how to Databricks ?

9. Explain the steps with screenshots how to create notebooks in Databricks ?

10. Explain the steps with screenshots how to insert data into databricks notebook & display the result?

11. Explain the steps with screenshots how to create cluster in databricks ?

Please create a word / pdf document, and send it to : avyuktitraining1@gmail.com

INTRODUCTION

This Assignment is given by manipal pro learn team on the basis of the training done in the forenoon session of this morning. The main objective behind this assignment is to master the theory and enhance knowledge over creating the azure synapse analytics, datafactory , databricks etc..

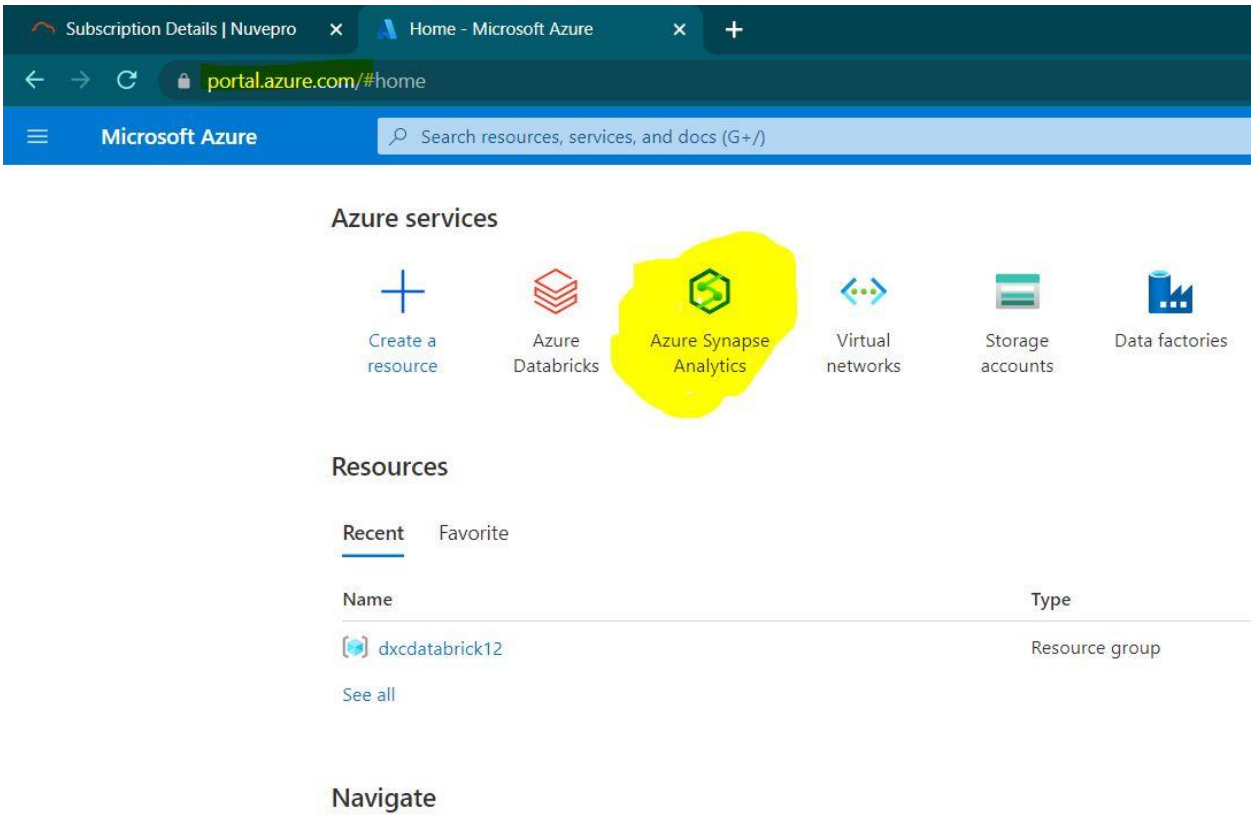
There are 11 questions and they are of easy to moderately difficult level. All the questions have been focused on what the trainer taught in the earlier sessions.All the demonstrations have been done successfully and documented except question 6 and 7 due to the unavailability of lab access after 6:00 pm I have tried a lot to do the task even till 6:33 PM but im unable to access the azure platform.

This assignment gave me immense confidence in mastering the domain that has been assigned to me. Special thanks to Unext team for providing the lab access.

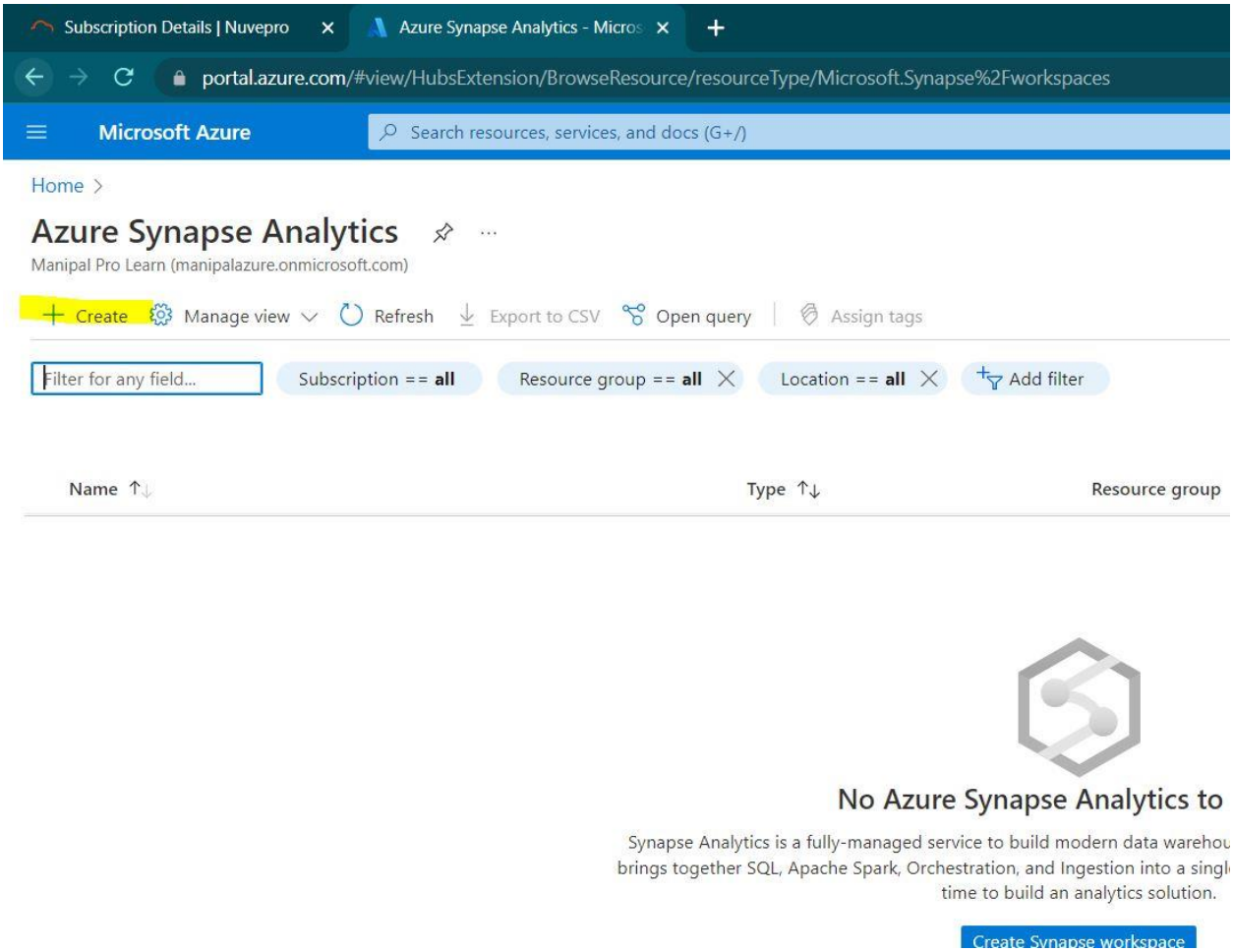
1. Explain the steps with screenshots how to AzureSynapse analytics?

A:

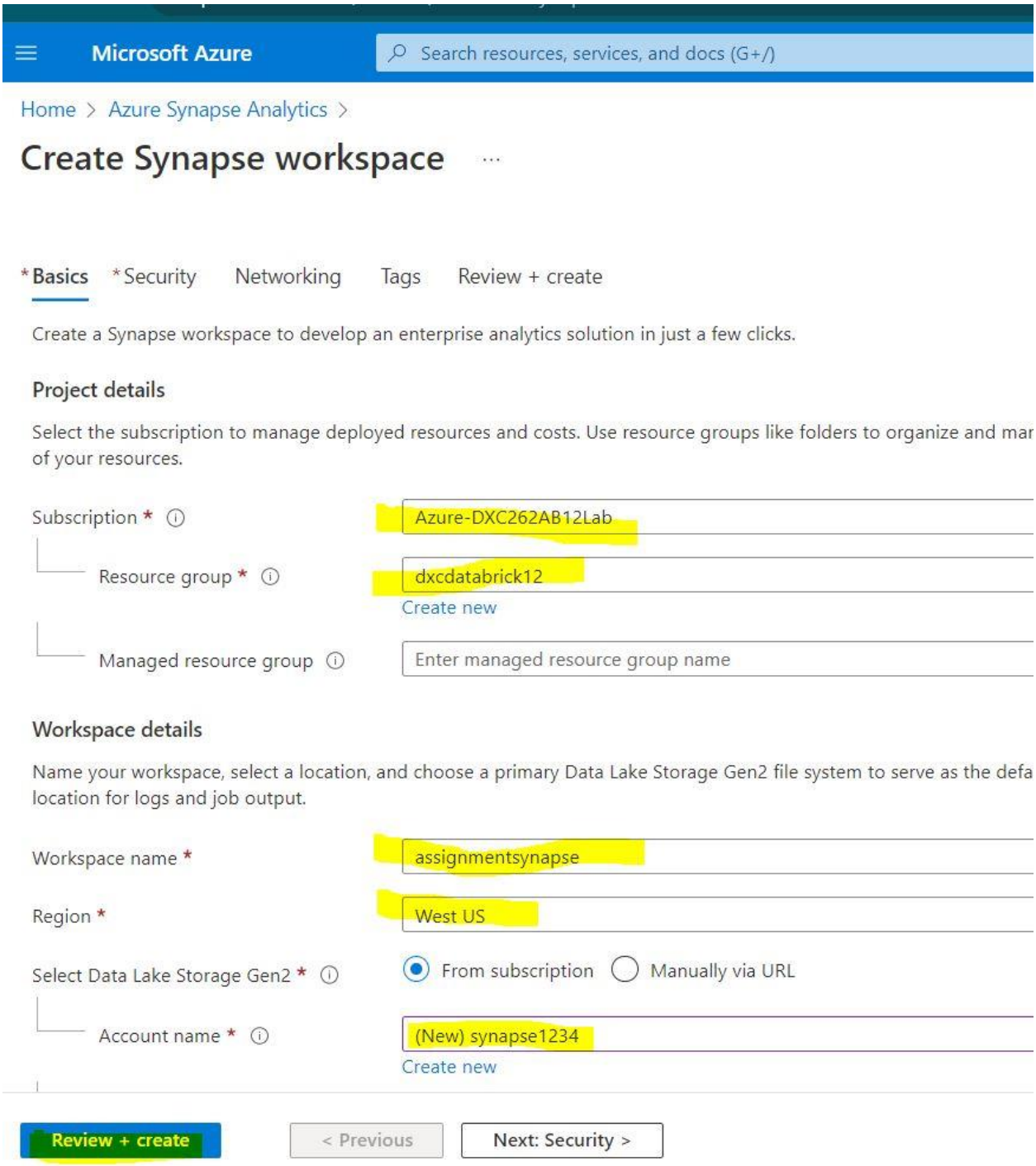
Step1: go to https://portal.azure.com/#home and select azure synapse analytics



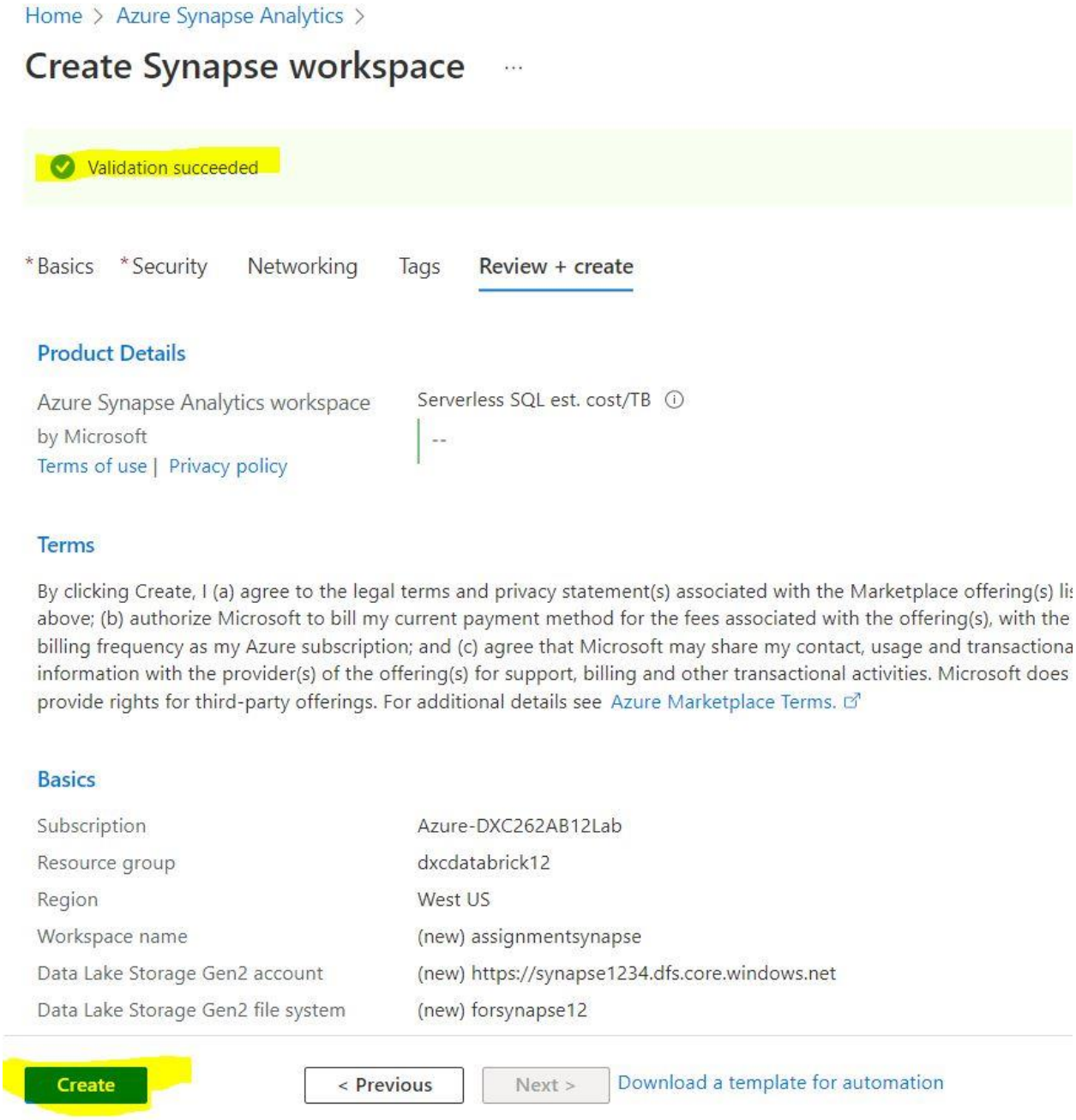
Step2: Click on create button



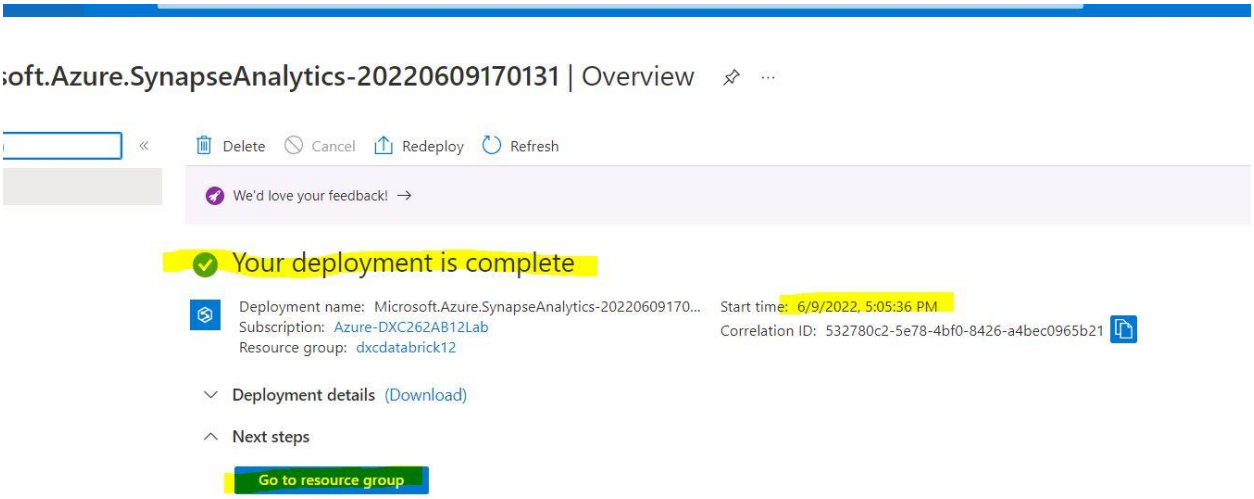
Step3: select the resource group , workspace, workspacename,account name etc and clickon review and create.



Step4: once the validation is completed click on create button.



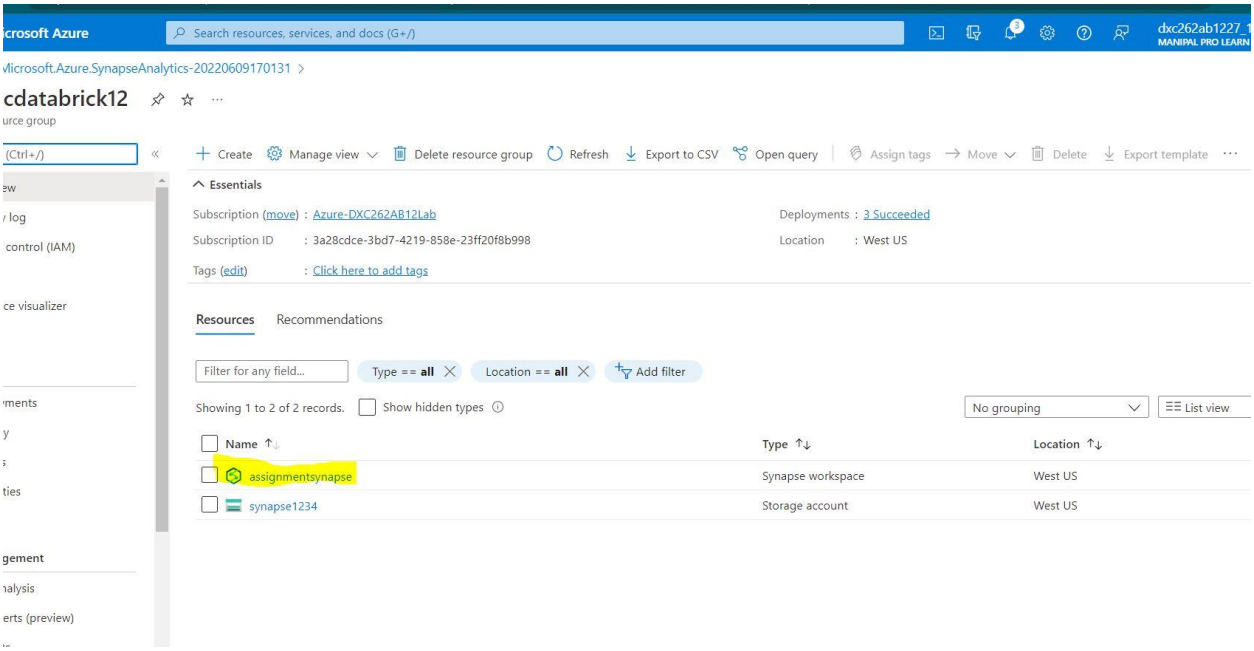
Step5: your deployment is completed and you can enjoy the services now.



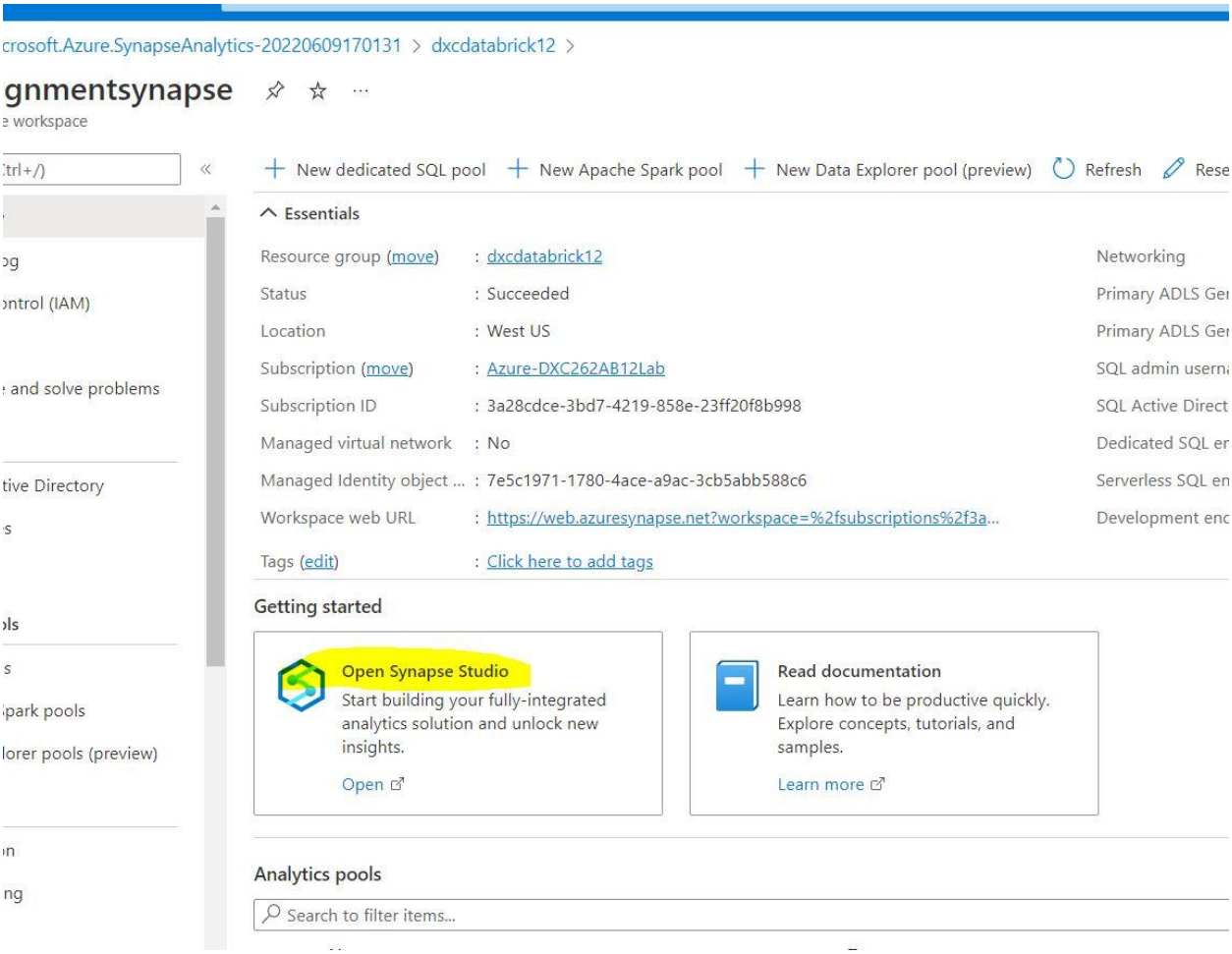
2. Explain the steps with screenshots how to SQL Pool in AzureSynapse analytics?

A:

Step1: open homepage of azure synapse and click on the synapse work space.

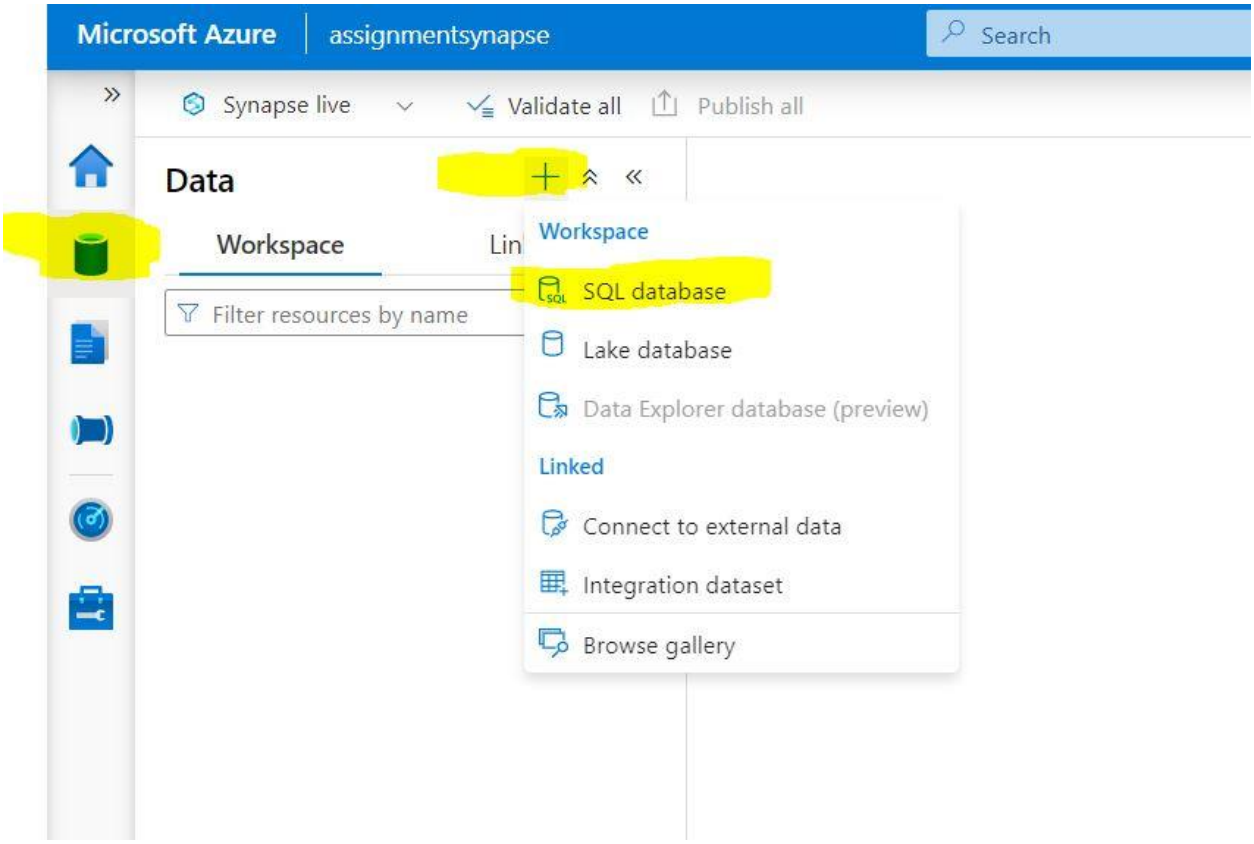


Step2: open the synapse studio.

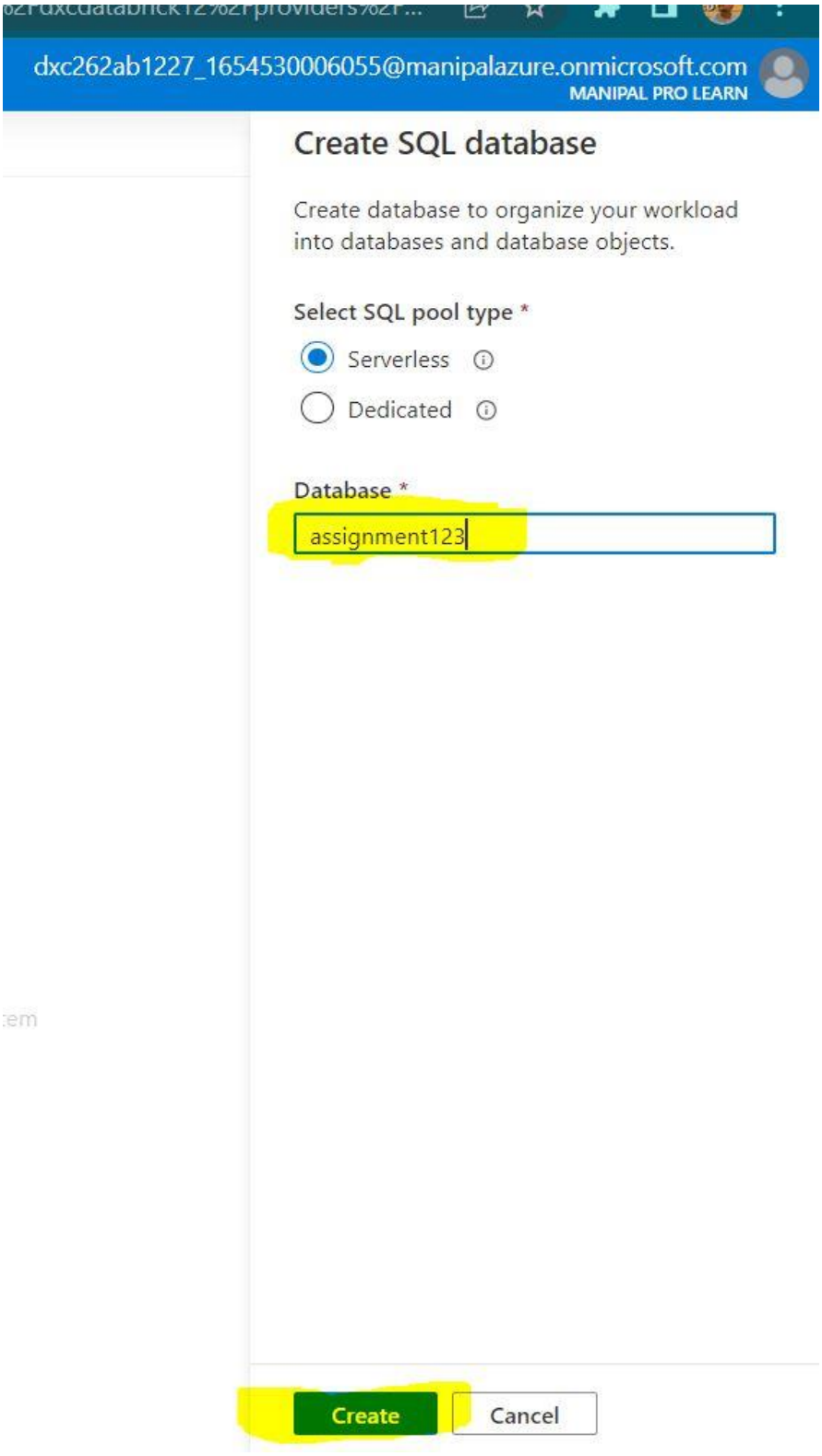


Step3:

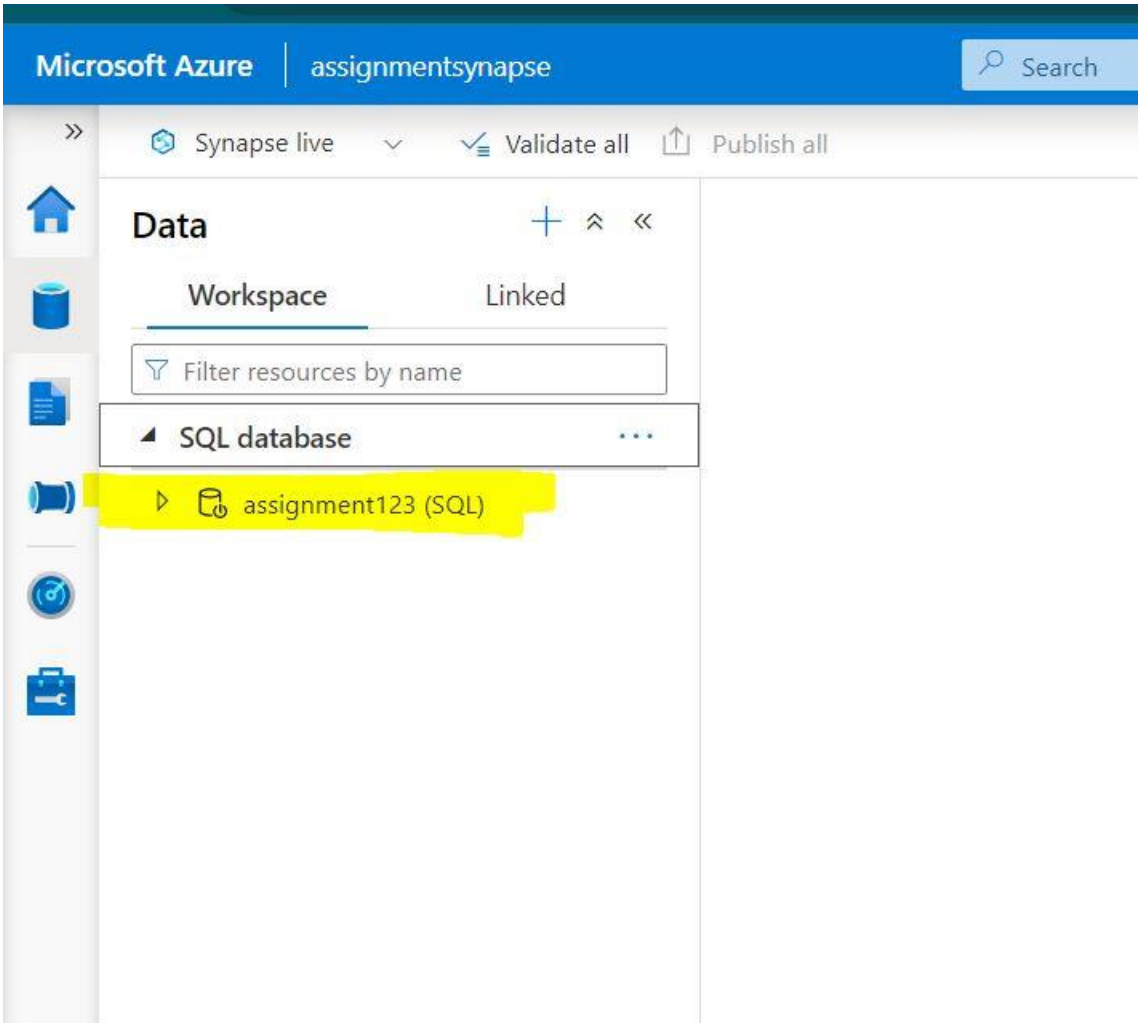
Click on data and then click on + and later sql data base.



Step4: name your database and click okay



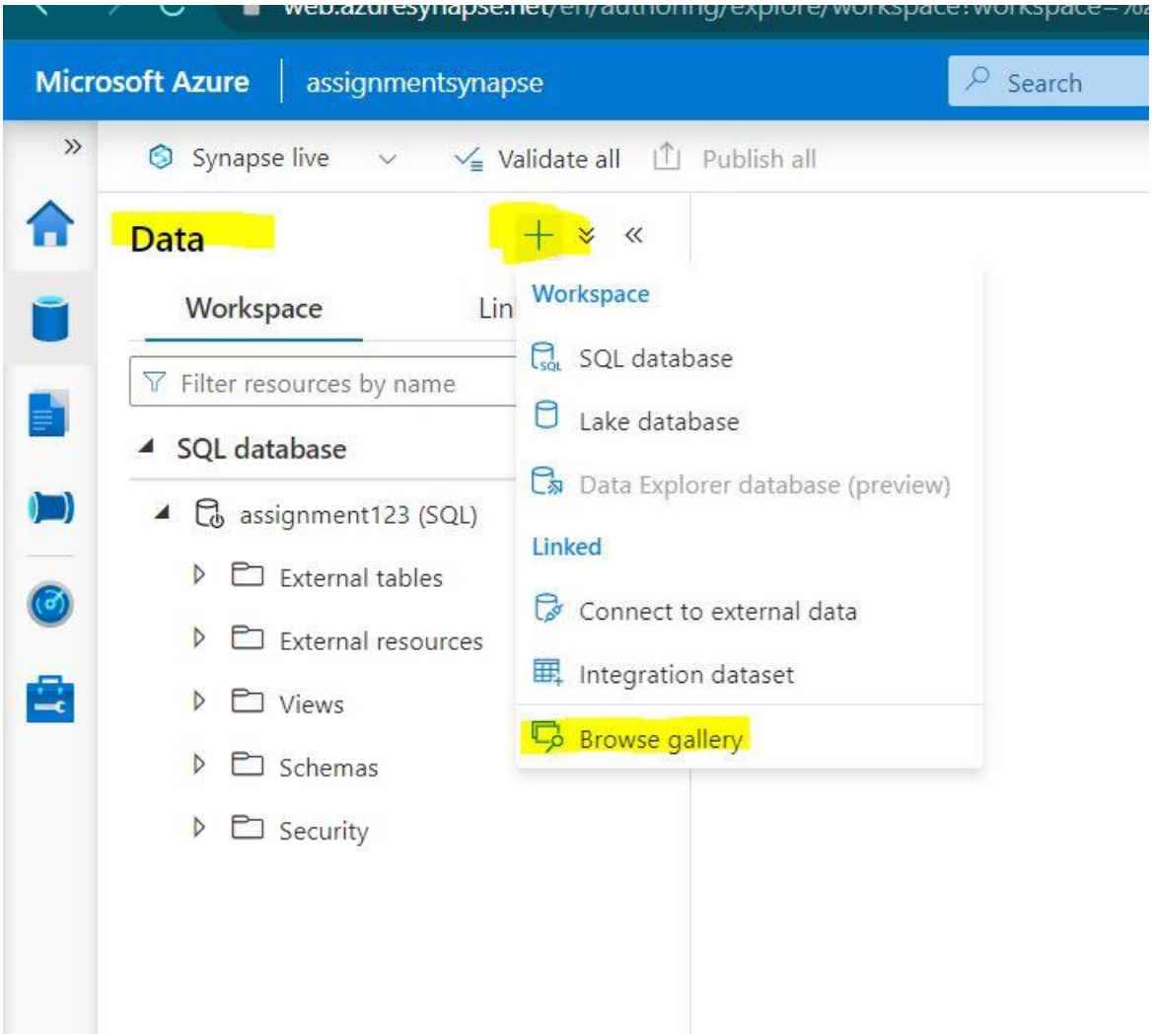
Step5: your sql database is ready


Microsoft Azure
assignmentsynapse
Search
Synapse live
Validate all
Publish all
Data
Workspace
Linked
Filter resources by name
SQL database
assignment123 (SQL)

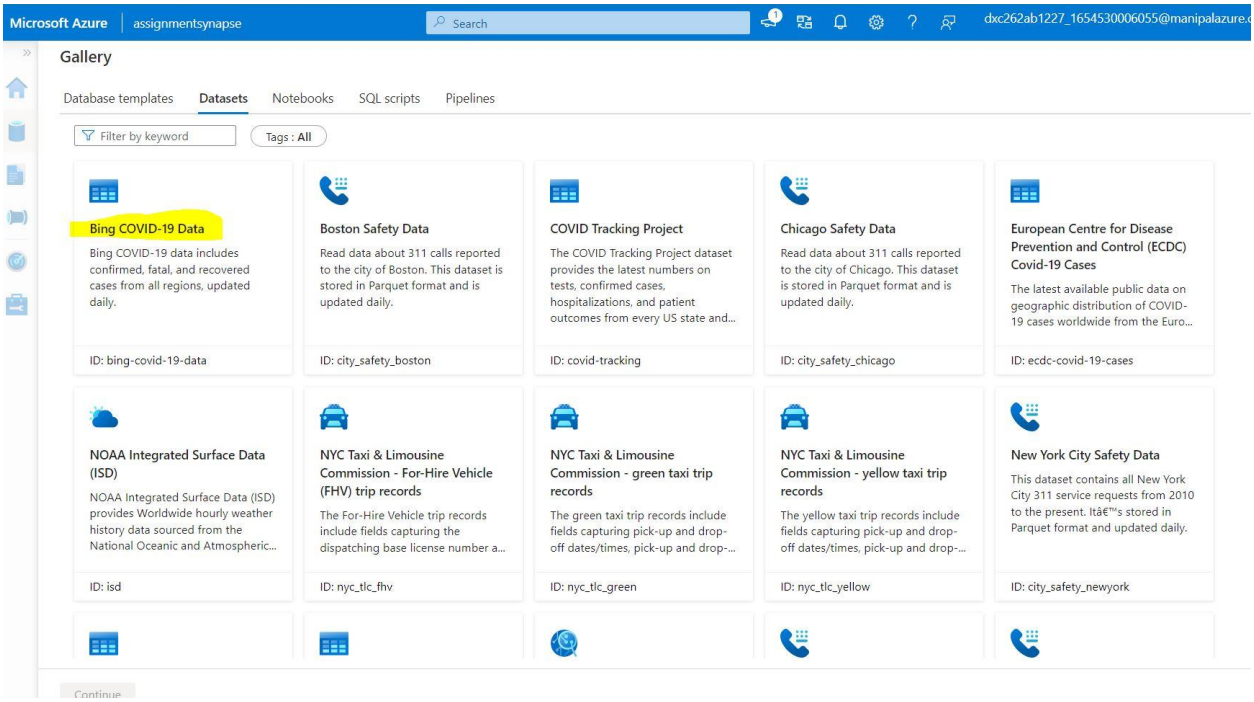
3. Explain the steps with screenshots how to import COVID19 dataset in AzureSynapse analytics and run sample 500 rows & dispaly the output?

A:

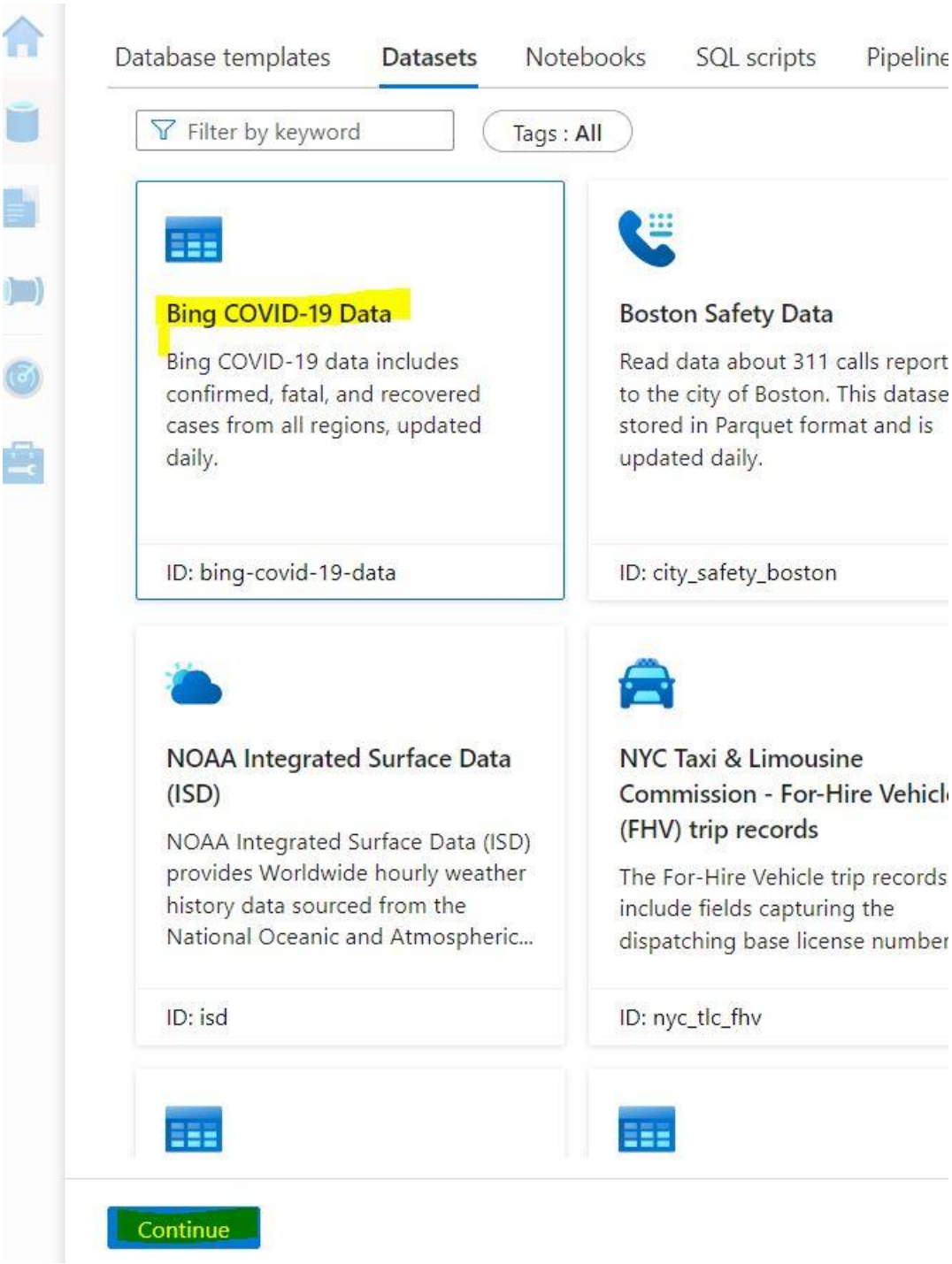
Step1: open azure synapse studio, click data and then on + later click on browse gallery.



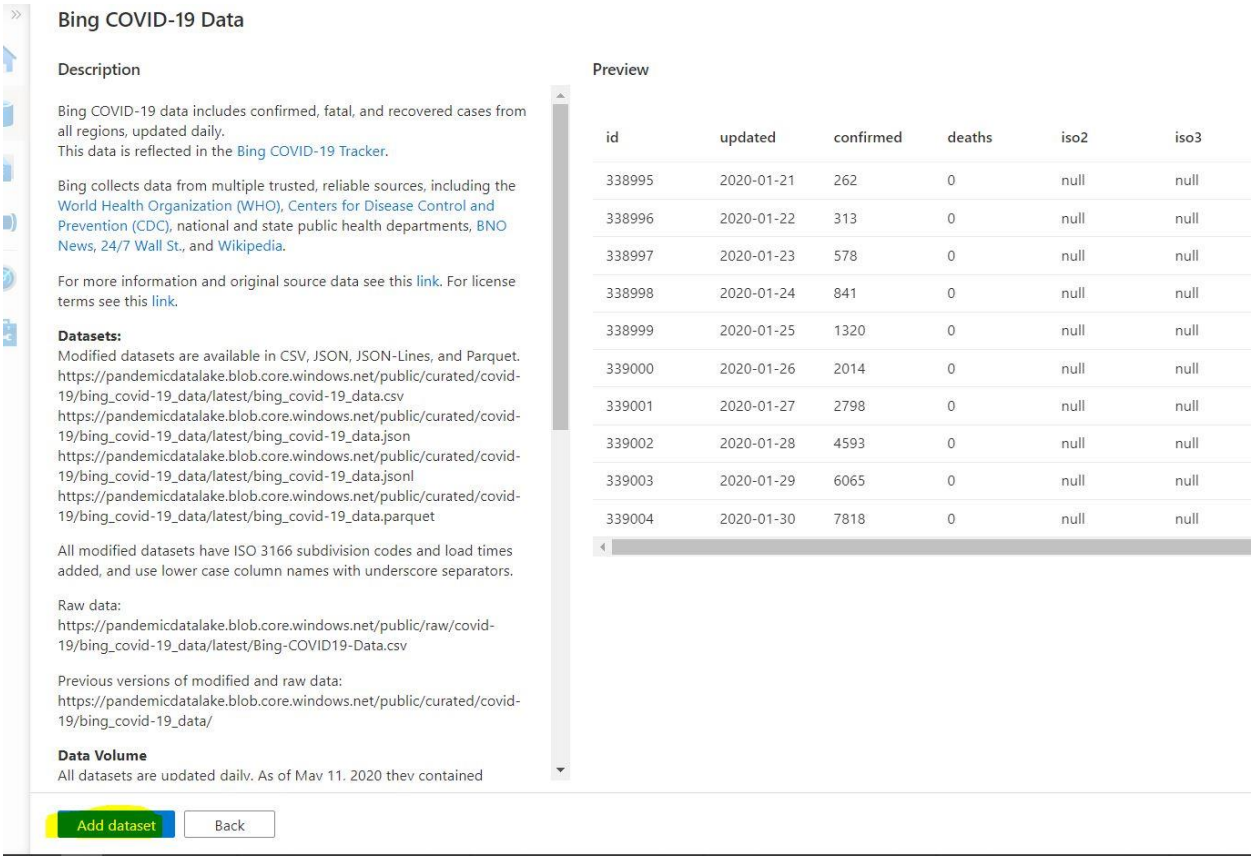
Step2: select bing covid data. And click continue



Step3: click continue



Step4: click on add dataset



Bing COVID-19 Data

Description

Bing COVID-19 data includes confirmed, fatal, and recovered cases from all regions, updated daily.
This data is reflected in the Bing COVID-19 Tracker.

Bing collects data from multiple trusted, reliable sources, including the World Health Organization (WHO), Centers for Disease Control and Prevention (CDC), national and state public health departments, BNO News, 24/7 Wall St., and Wikipedia.

For more information and original source data see this link. For license terms see this link.

Datasets:
Modified datasets are available in CSV, JSON, JSON-Lines, and Parquet.
https://pandemicdatalake.blob.core.windows.net/public/curated/covid-19/bing_covid-19_data/latest/bing_covid-19_data.csv
https://pandemicdatalake.blob.core.windows.net/public/curated/covid-19/bing_covid-19_data/latest/bing_covid-19_data.json
https://pandemicdatalake.blob.core.windows.net/public/curated/covid-19/bing_covid-19_data/latest/bing_covid-19_data.jsonl
https://pandemicdatalake.blob.core.windows.net/public/curated/covid-19/bing_covid-19_data/latest/bing_covid-19_data.parquet

All modified datasets have ISO 3166 subdivision codes and load times added, and use lower case column names with underscore separators.

Raw data:
https://pandemicdatalake.blob.core.windows.net/public/raw/covid-19/bing_covid-19_data/latest/Bing-COVID19-Data.csv

Previous versions of modified and raw data:
https://pandemicdatalake.blob.core.windows.net/public/curated/covid-19/bing_covid-19_data/

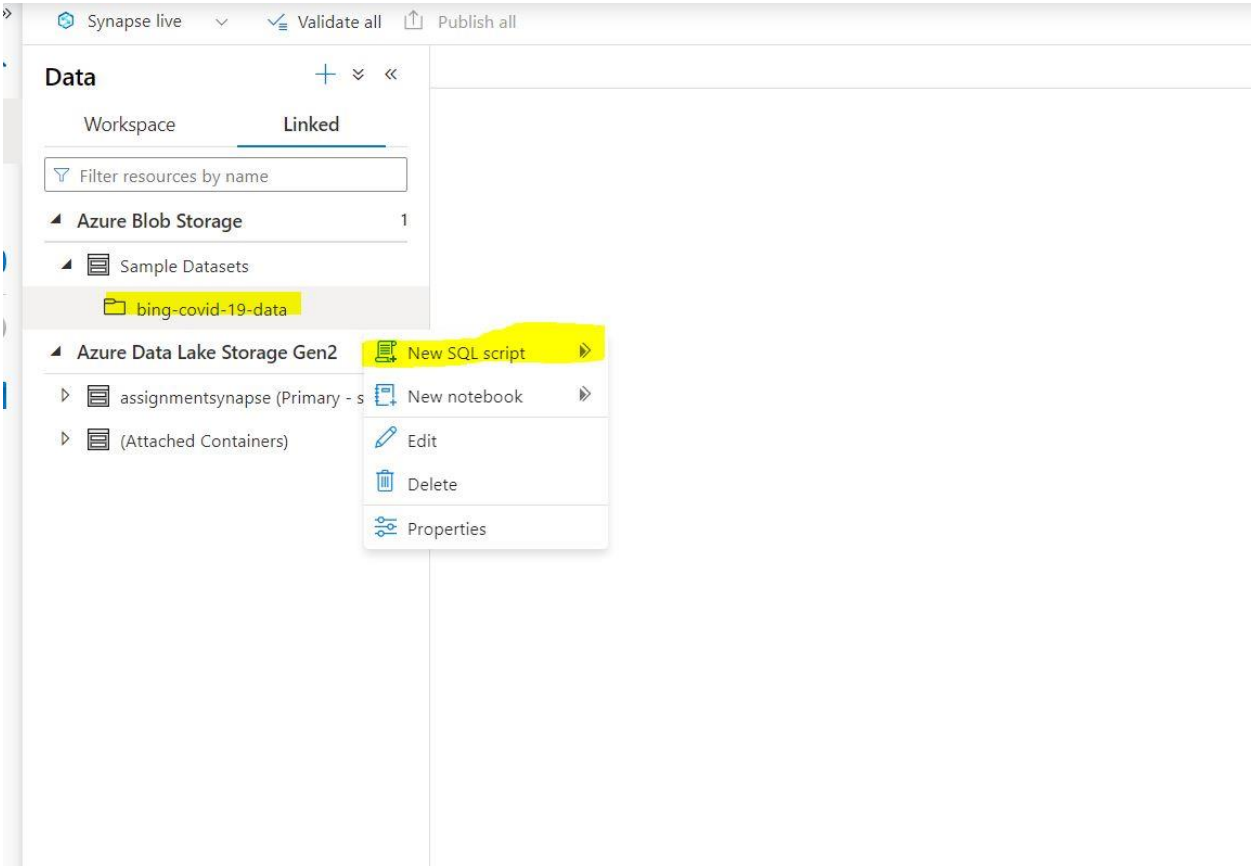
Data Volume
All datasets are updated daily. As of May 11, 2020 they contained

Preview

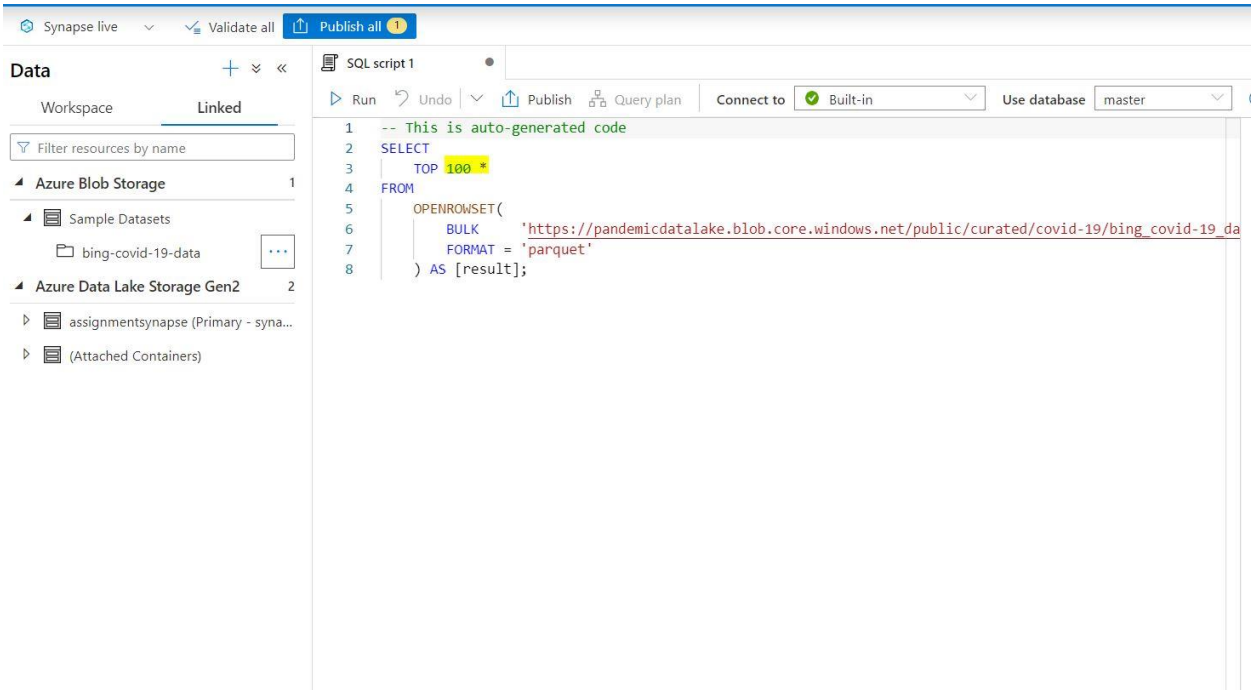
id	updated	confirmed	deaths	iso2	iso3
338995	2020-01-21	262	0	null	null
338996	2020-01-22	313	0	null	null
338997	2020-01-23	578	0	null	null
338998	2020-01-24	841	0	null	null
338999	2020-01-25	1320	0	null	null
339000	2020-01-26	2014	0	null	null
339001	2020-01-27	2798	0	null	null
339002	2020-01-28	4593	0	null	null
339003	2020-01-29	6065	0	null	null
339004	2020-01-30	7818	0	null	null

Add dataset | Back

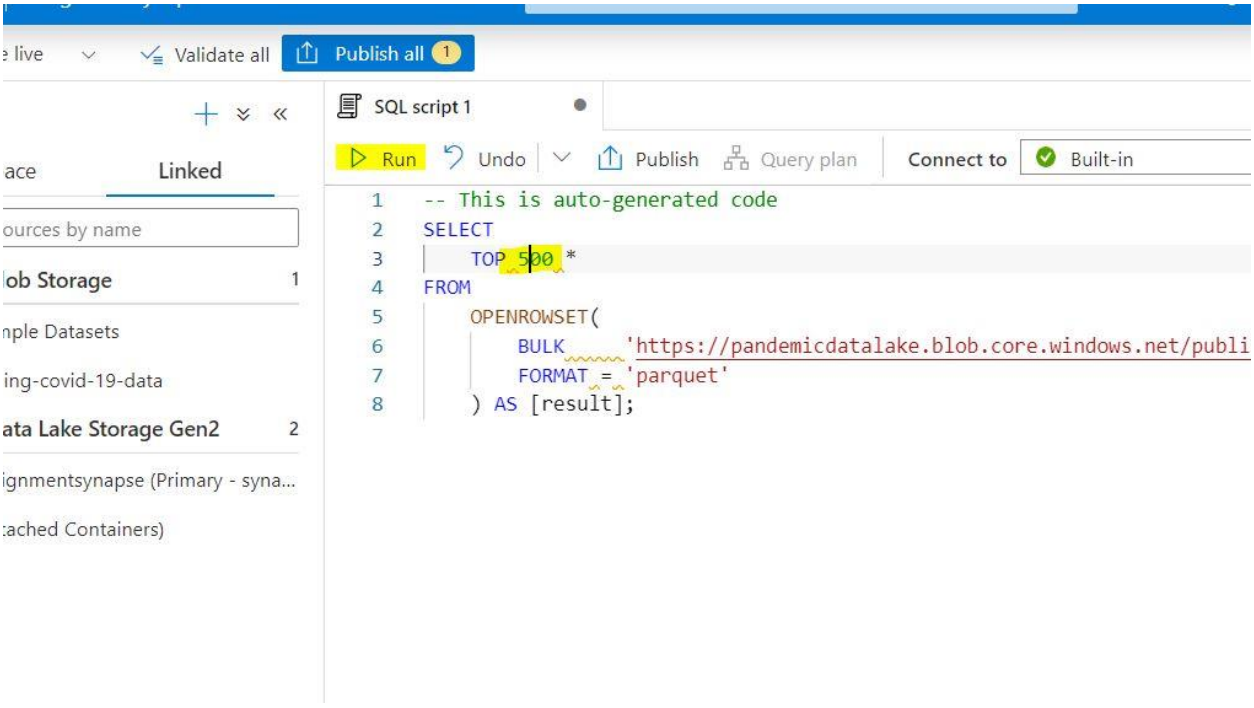
Step5: select data and then select our database and later click on new sql script. And elect top 100.



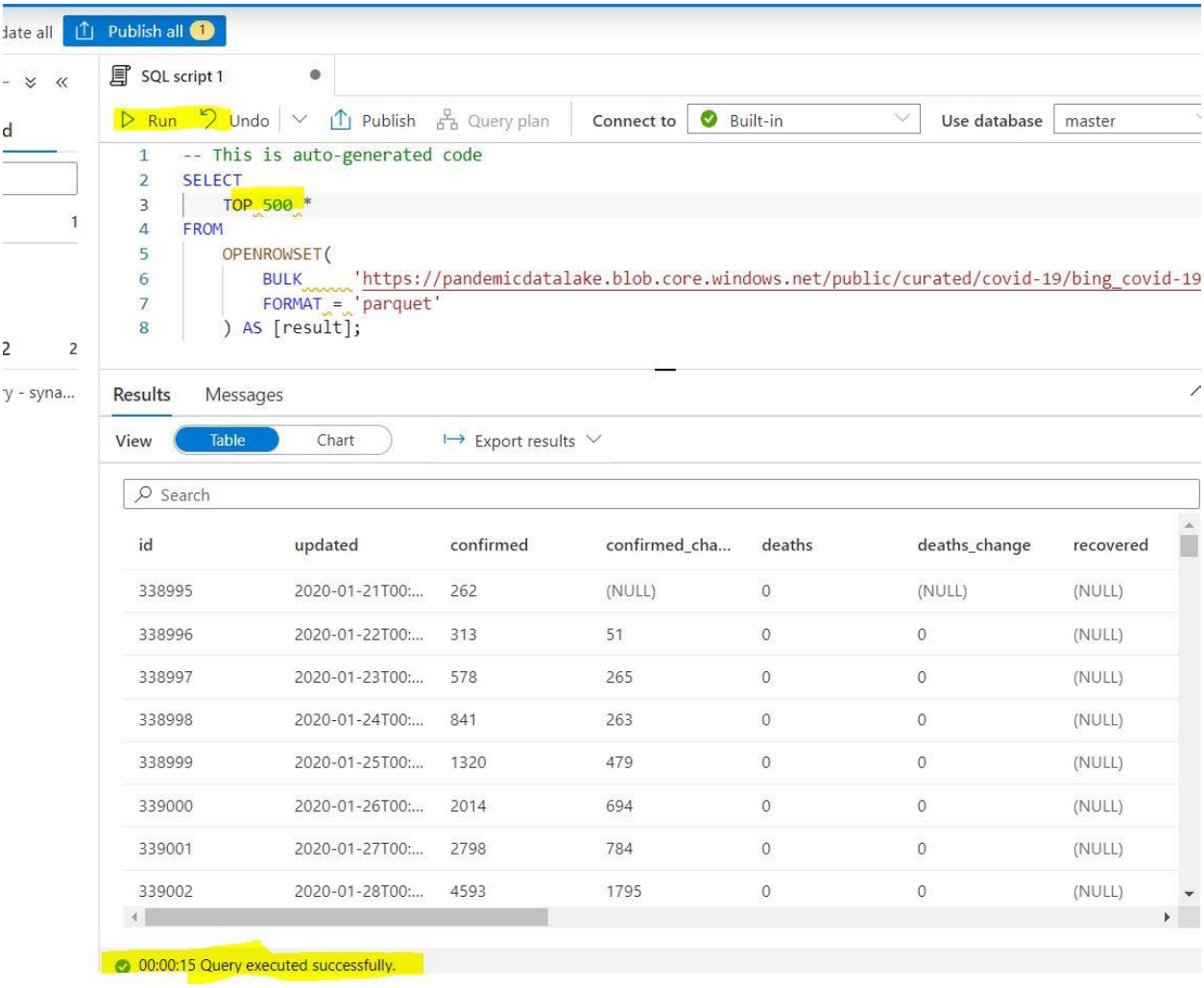
Step6: change the 100 in the code to 500



Step7: click on run button



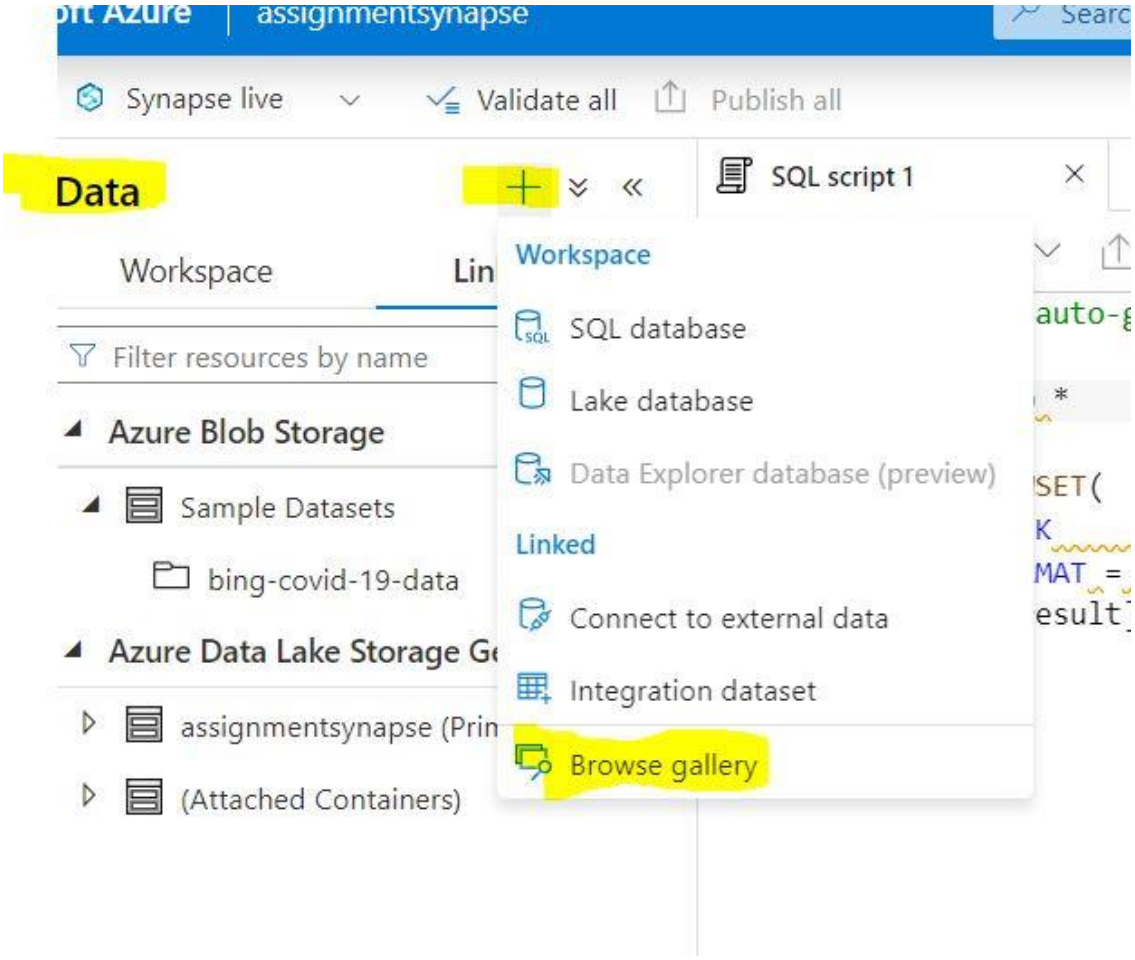
Step8: here are the results.



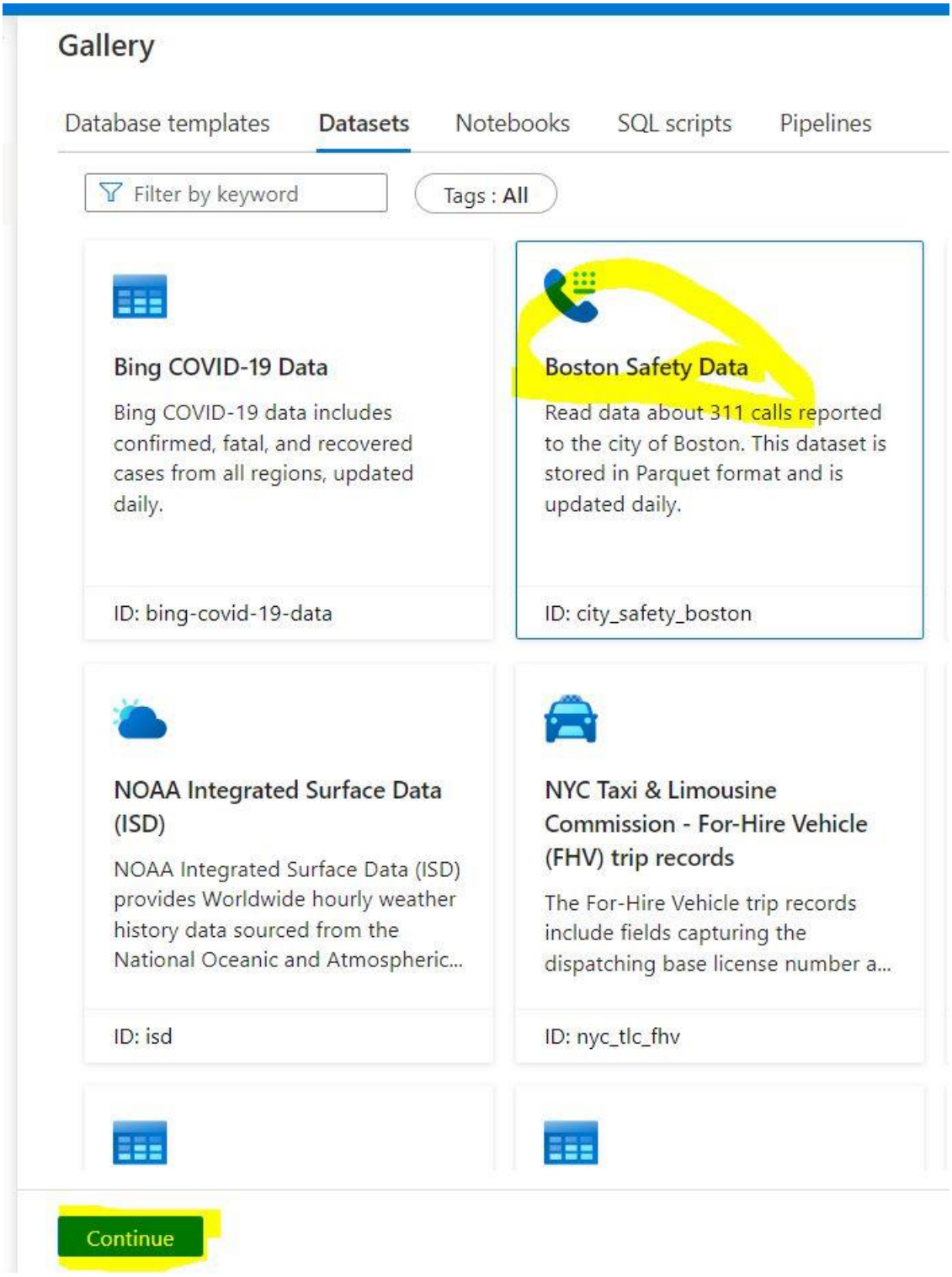
4. Explain the steps with screenshots how to input Boston Safety datasets into AzureSynapse analytics? using Notebooks ?

A:

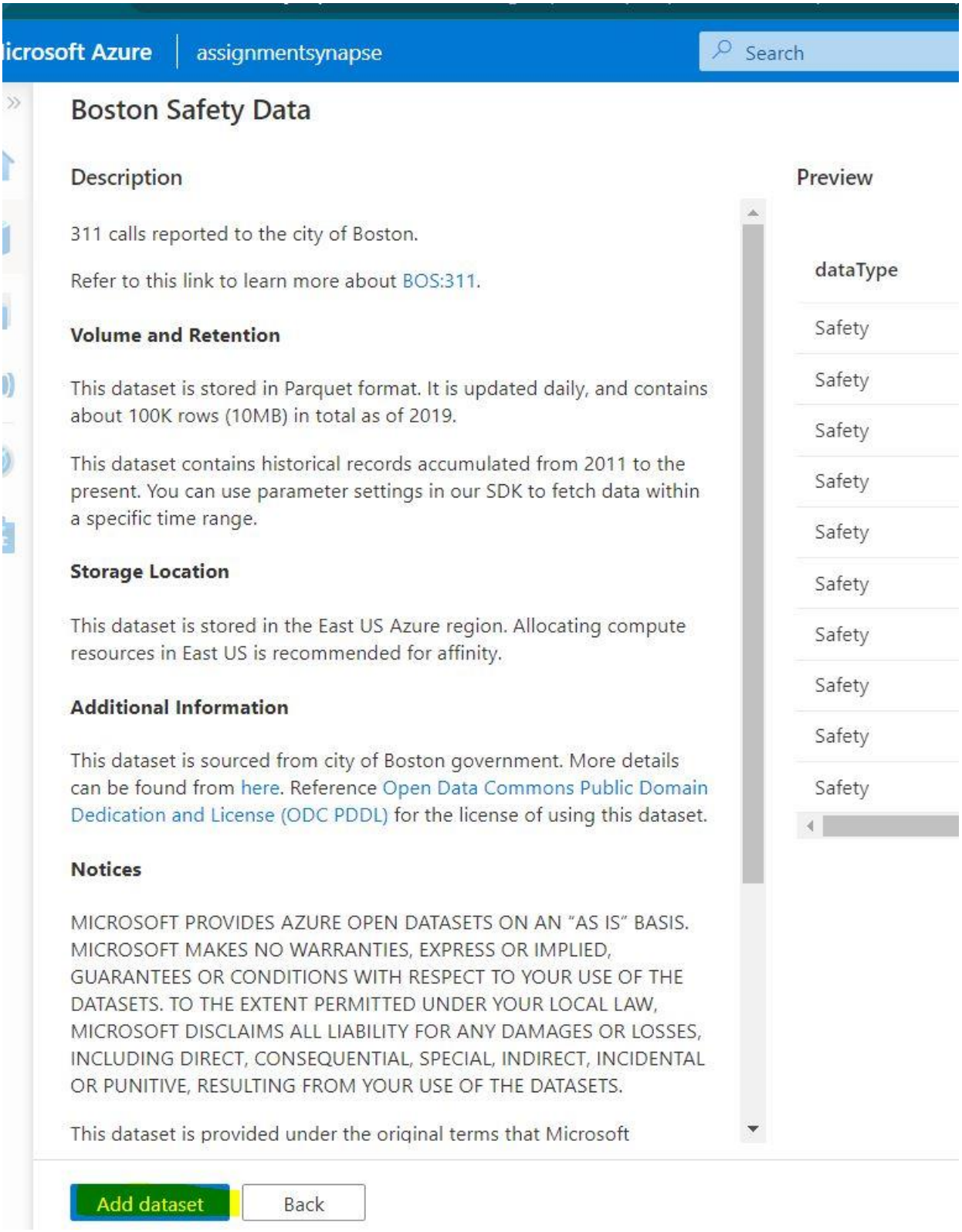
Step1: click on data , +and browse gallery



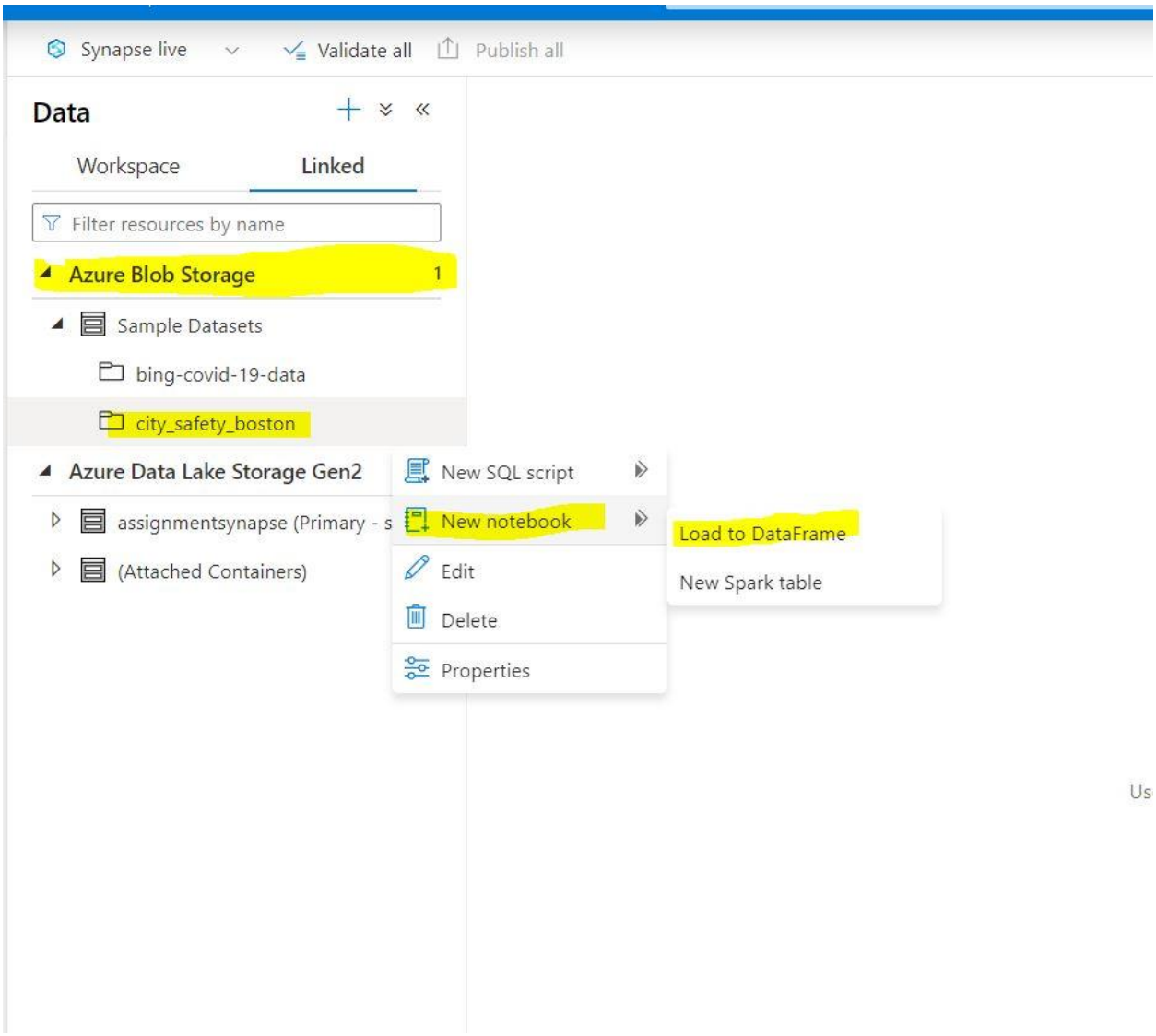
Step2: select boston secyrity data and click continue.



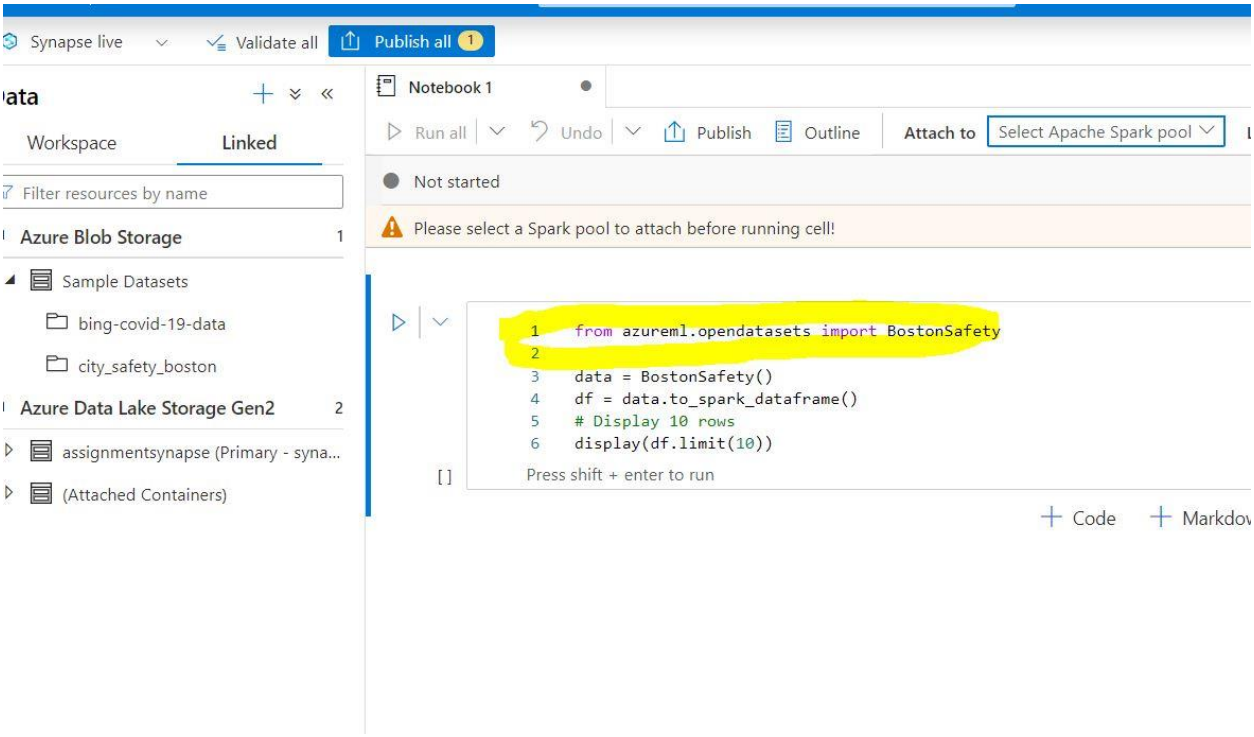
Step3: click on add dataset



Step4: click on azure blobstorage and then city_safety_boston and click on new note book and later load to Dataframe.



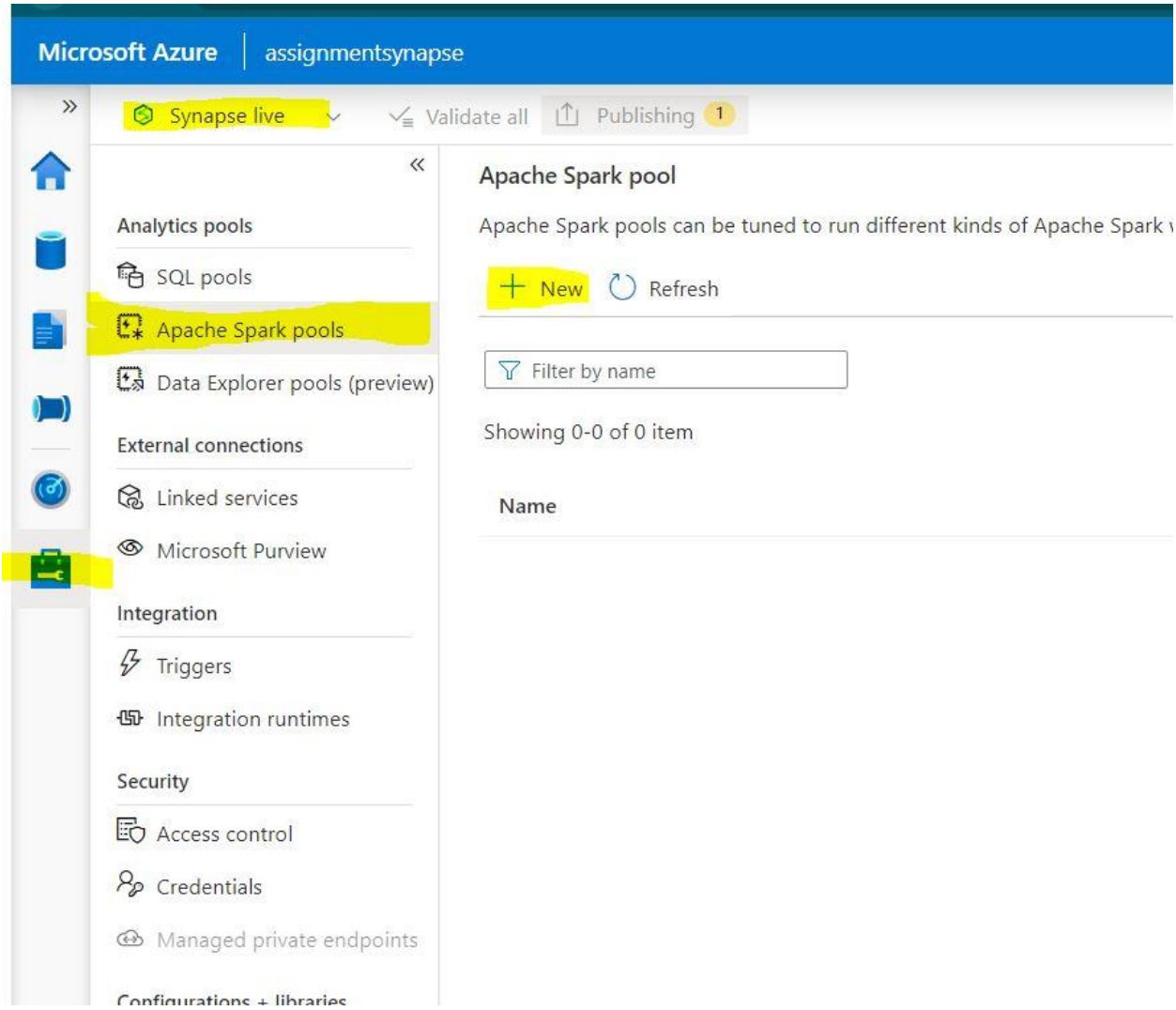
Step5: we can see the input dataframe in the notebook



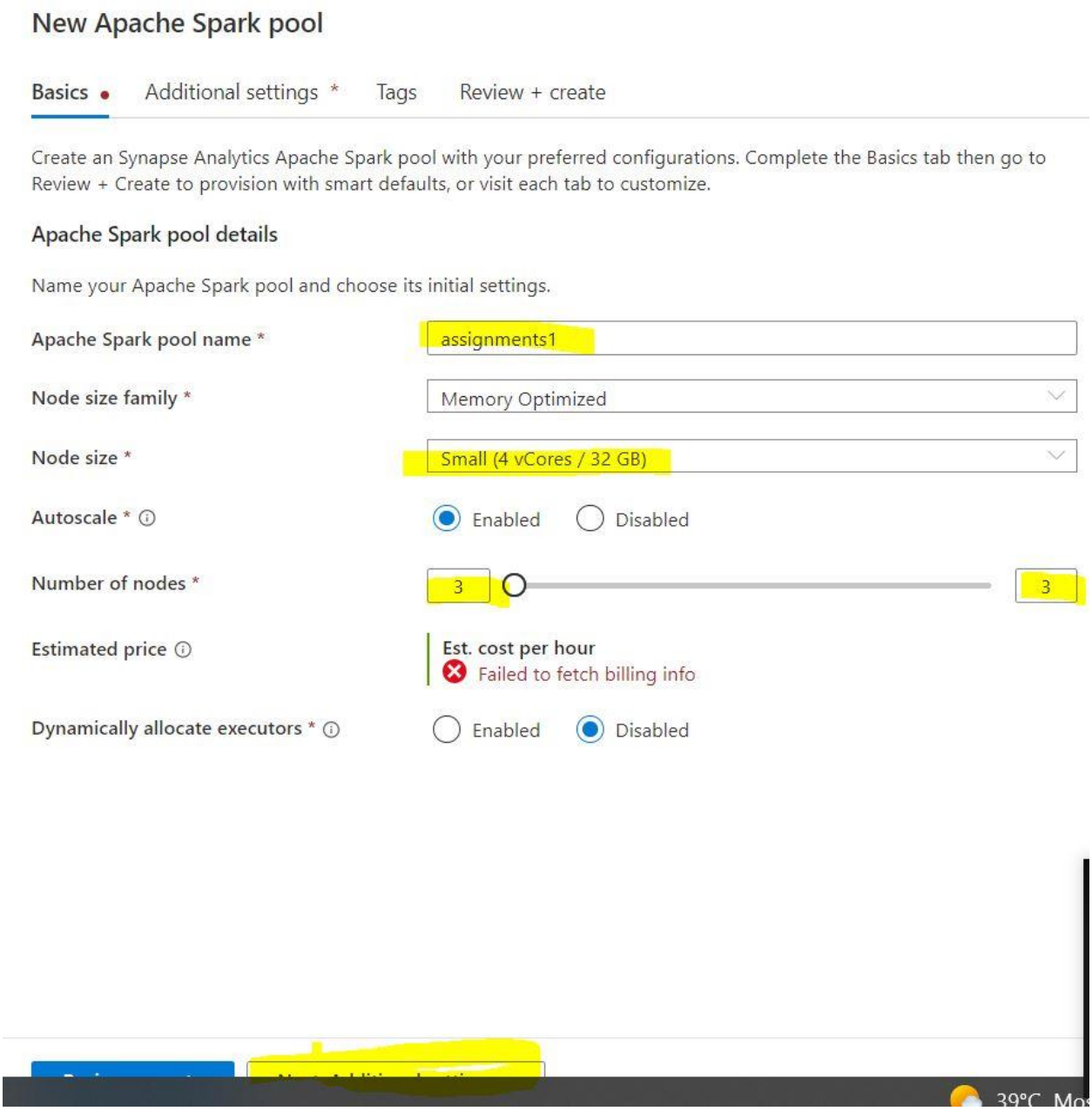
5. Explain the steps with screenshots how to create Spark pool in AzureSynapse analytics? ?

A:

Step1: open synapse studio and click on synapse live and click on apache spark pools and on new button.

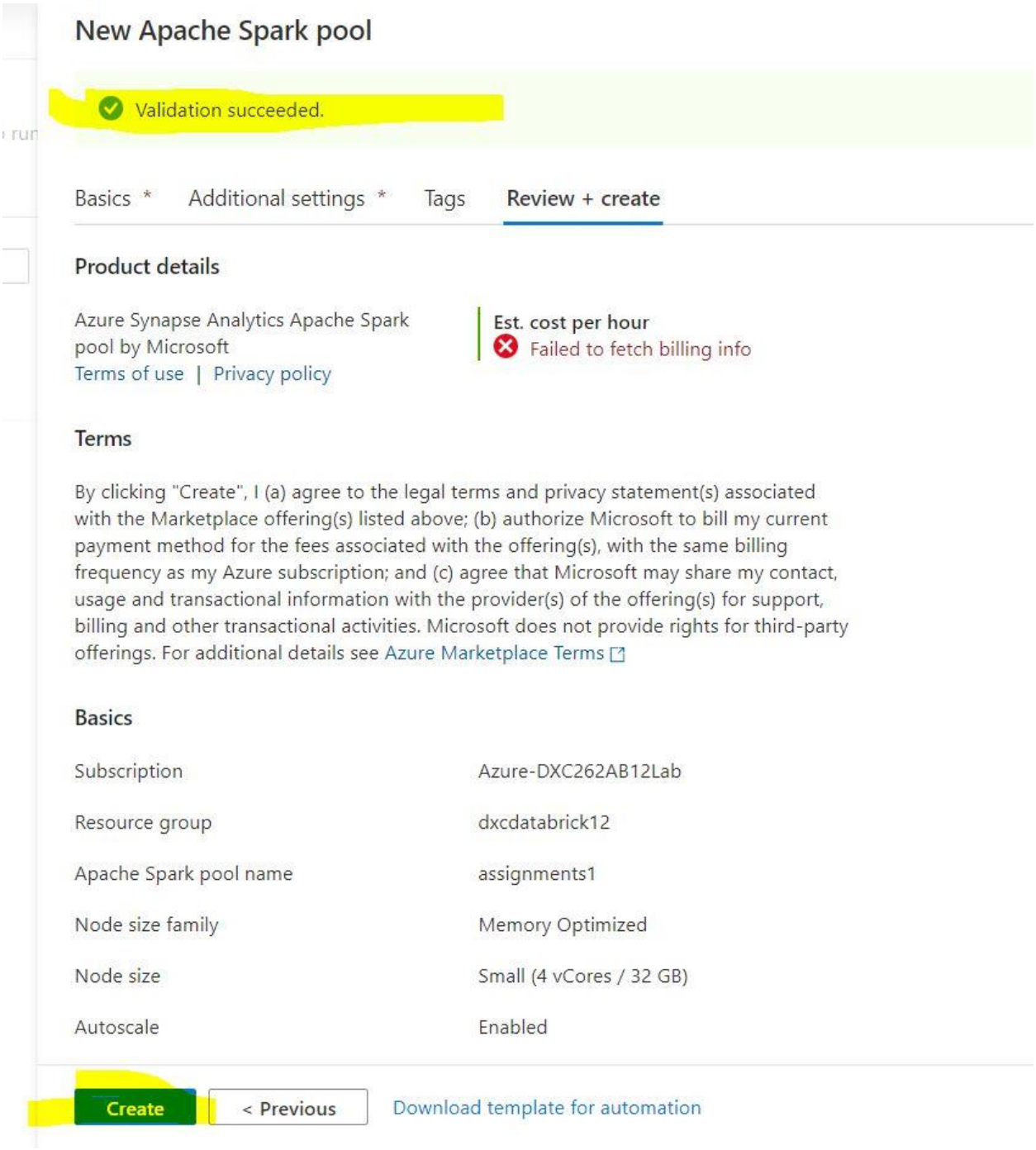


Step2: name the spark pool and select node size , pool name and number of nodes.

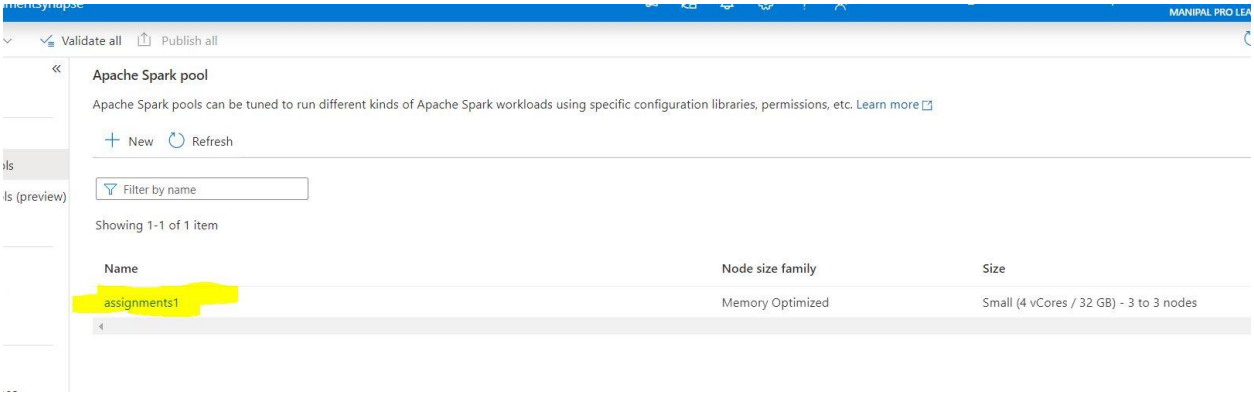


Step3:

After successful validation click on create .



Step4: your spark pool is created.



6. Explain the steps with screenshots how to create pipeline in AzureSynapse analytics? ?

A:

Step1:

Step1:

Step1:

Step1:

Step1:

Step1:

Step1:

Step1:

Step1:

Step1:

Step1:

Step1:

Step1:

Step1:

Step1:

Step1:

Step1:

Step1:

Step1:

Step1:

Step1:

Step1:

Step1:

7. Explain the steps with screenshots how to automate the pipelines in AzureSynapse analytics??

A:

Step1:

Step1:

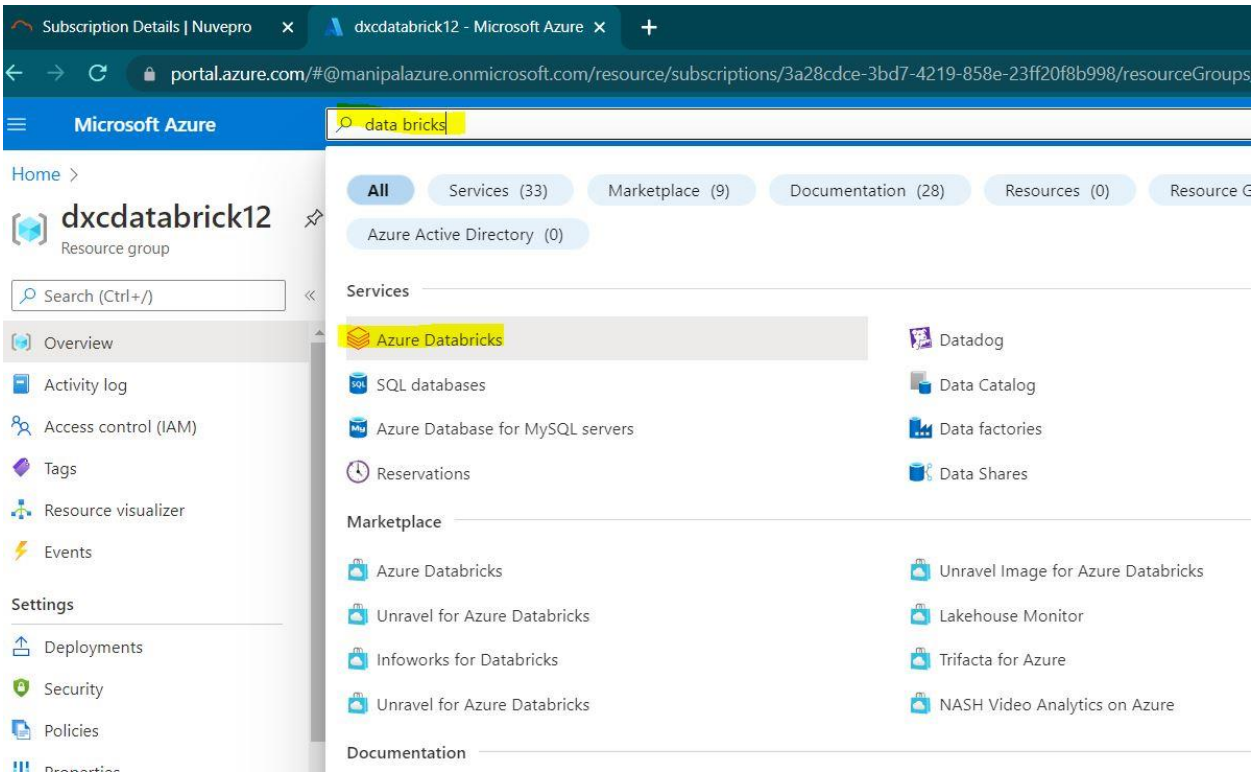
Step1:

Step1:

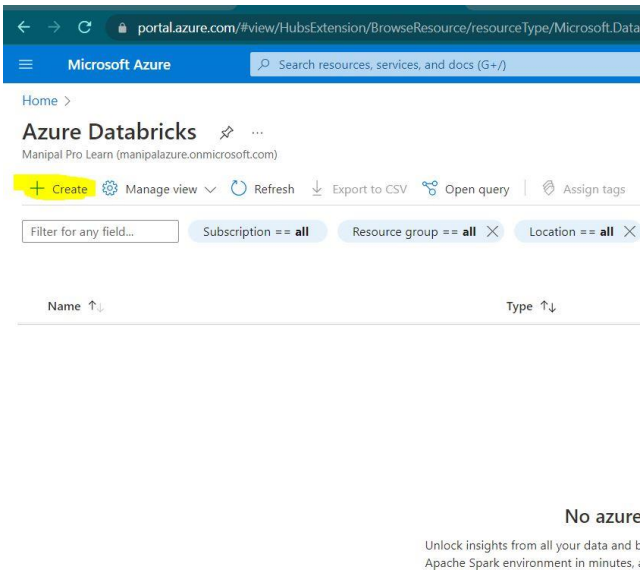
8. Explain the steps with screenshots how to Create Databricks ?

A:

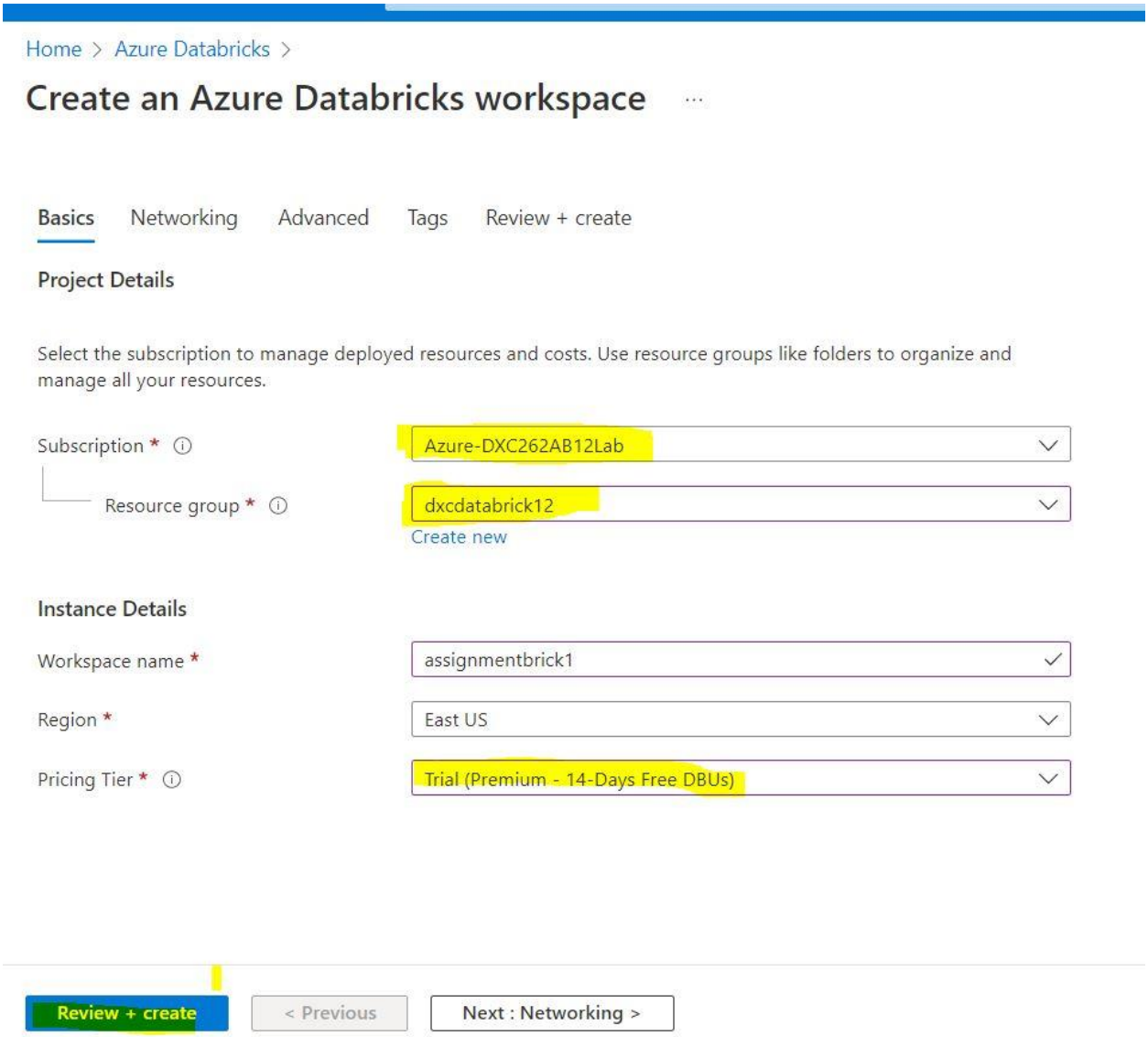
Step1: Go to https://portal.azure.com/ and search for Databricks.



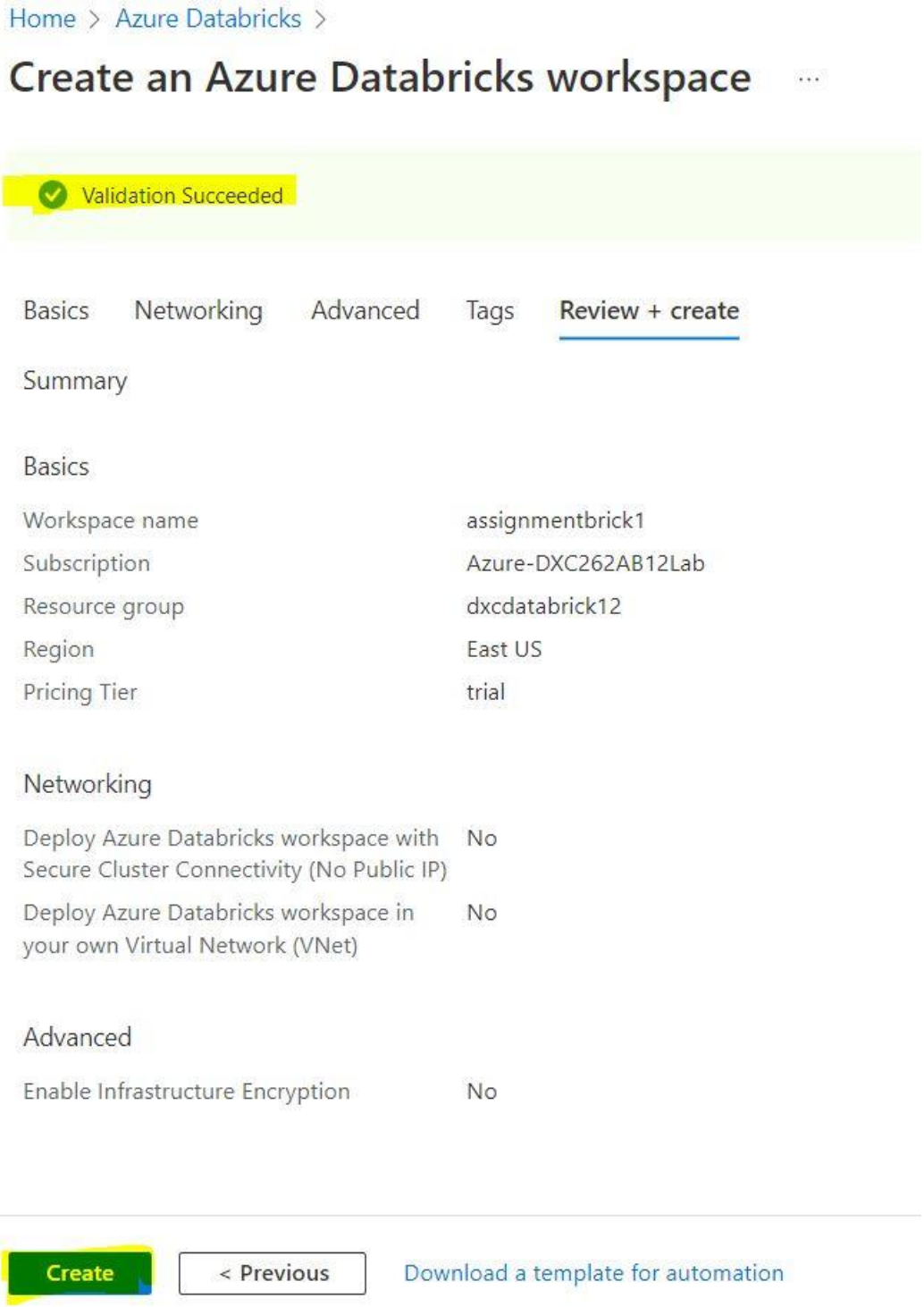
Step2: click on create button



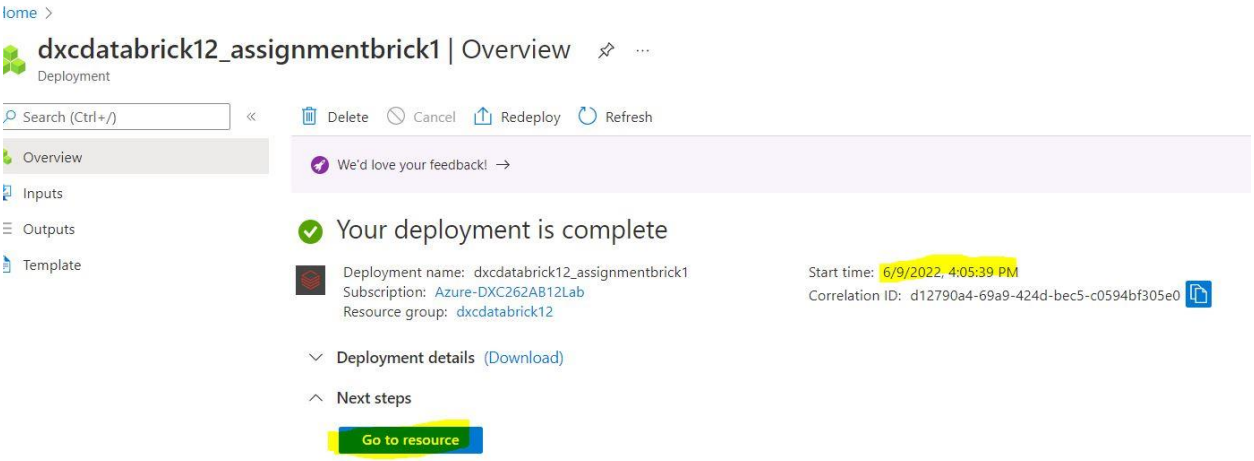
Step3: select resource group, name the workspace , select the pricing tier and click on review and create



Step4: after validation succeeded click on create button.



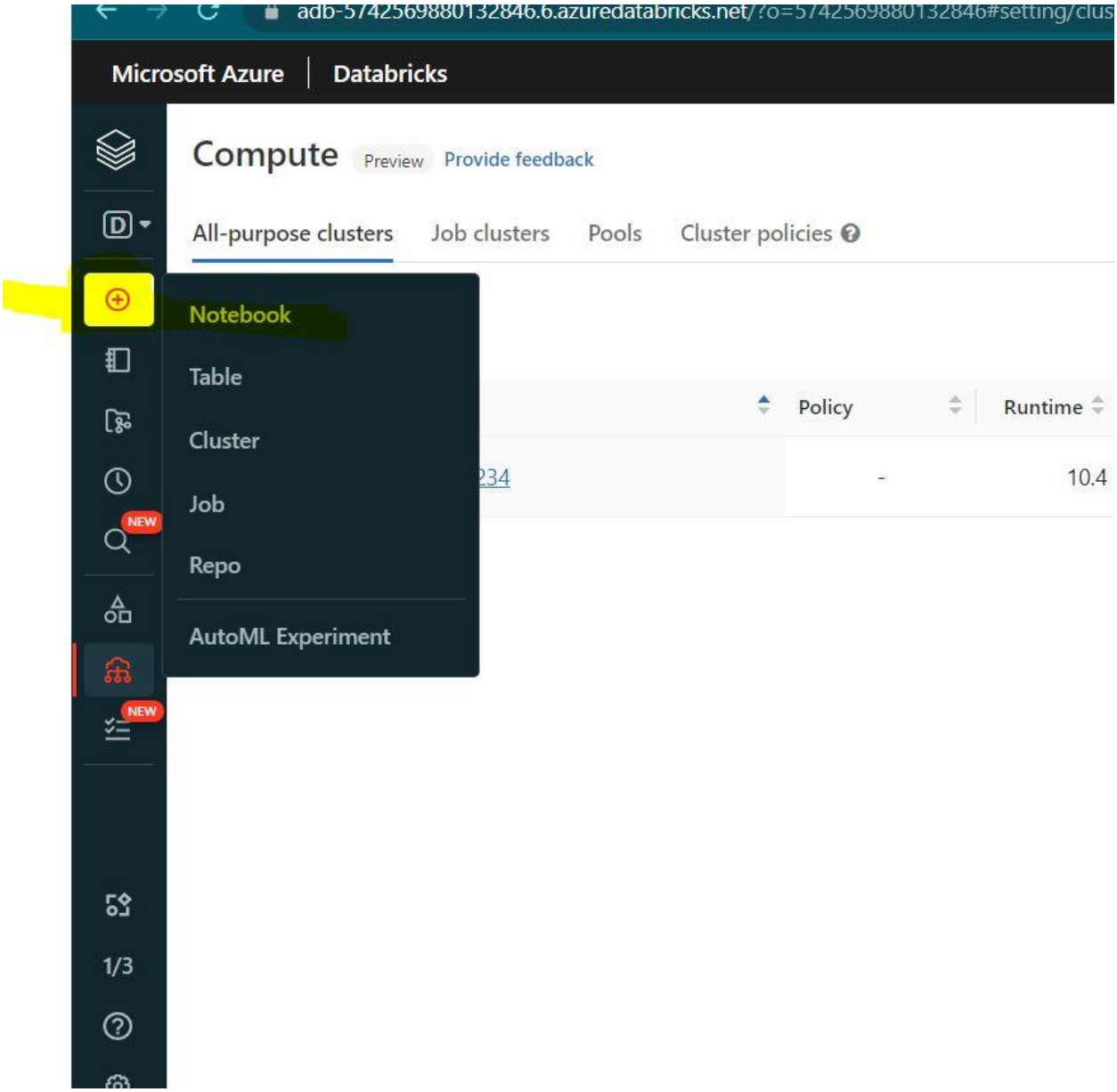
Step5: your deployment is completed.click goto resource.



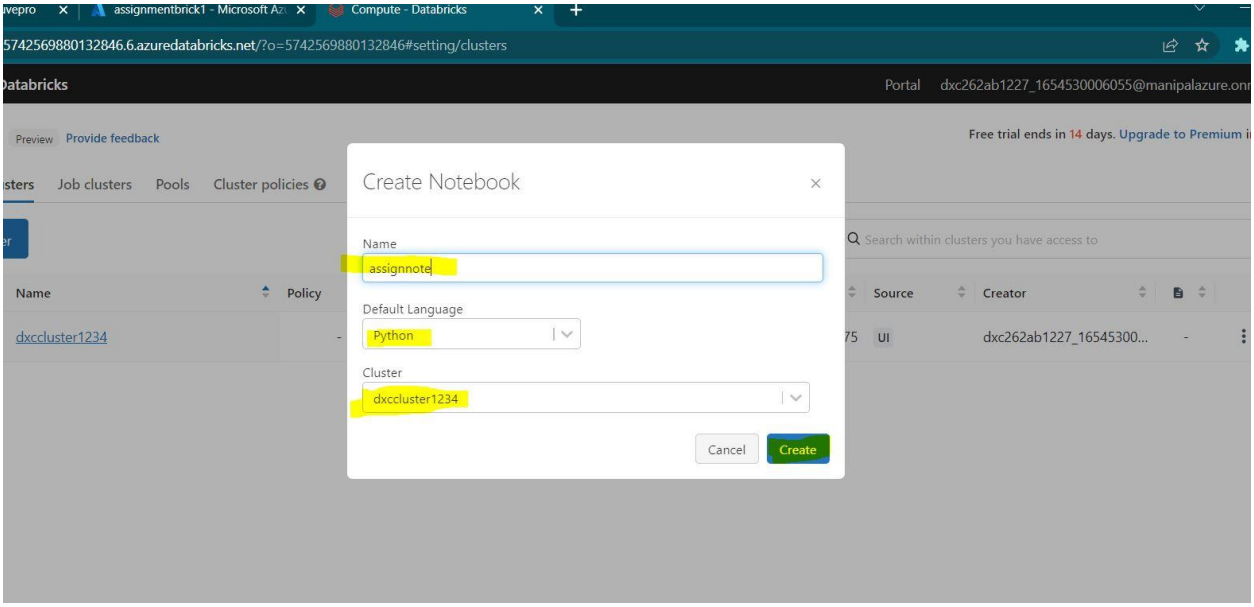
9. Explain the steps with screenshots how to create notebooks in Databricks ?

A:

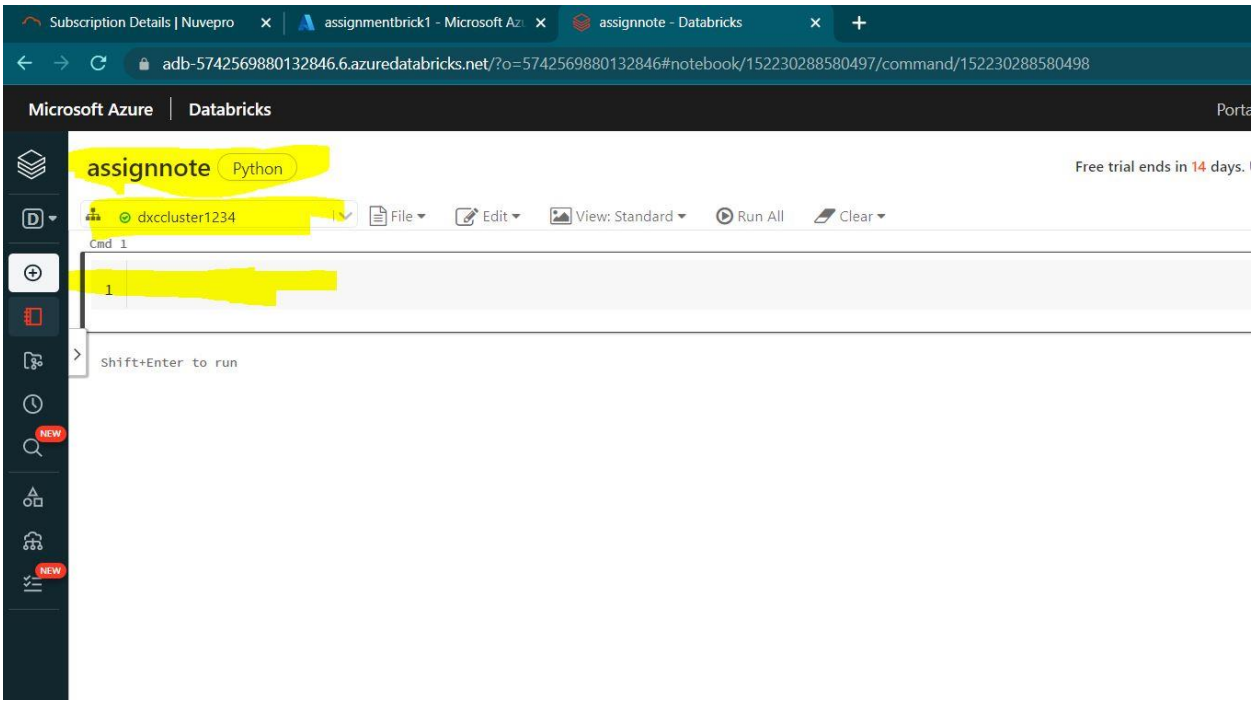
Step1: Click on create button in the workspace and then on notebook



Step2:give the name , slect the cluster and also language and finally click on create button.



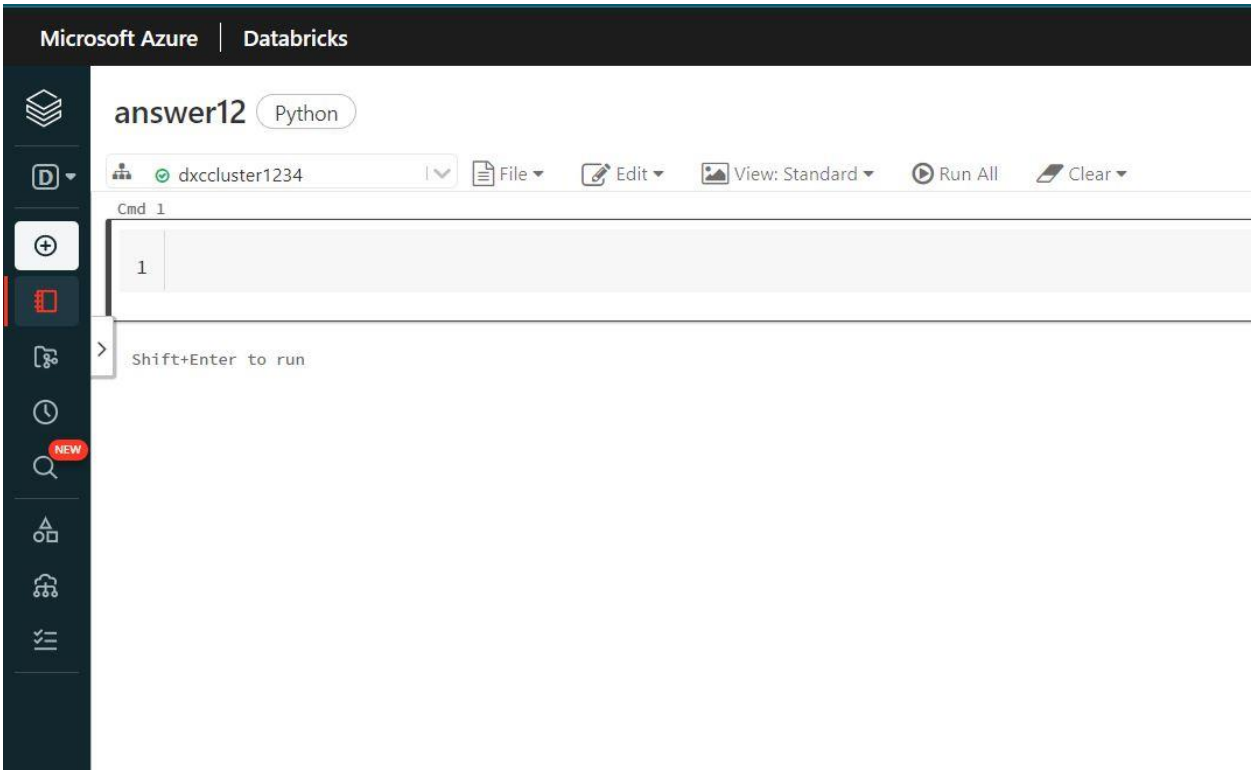
Step3: your notebook is created.



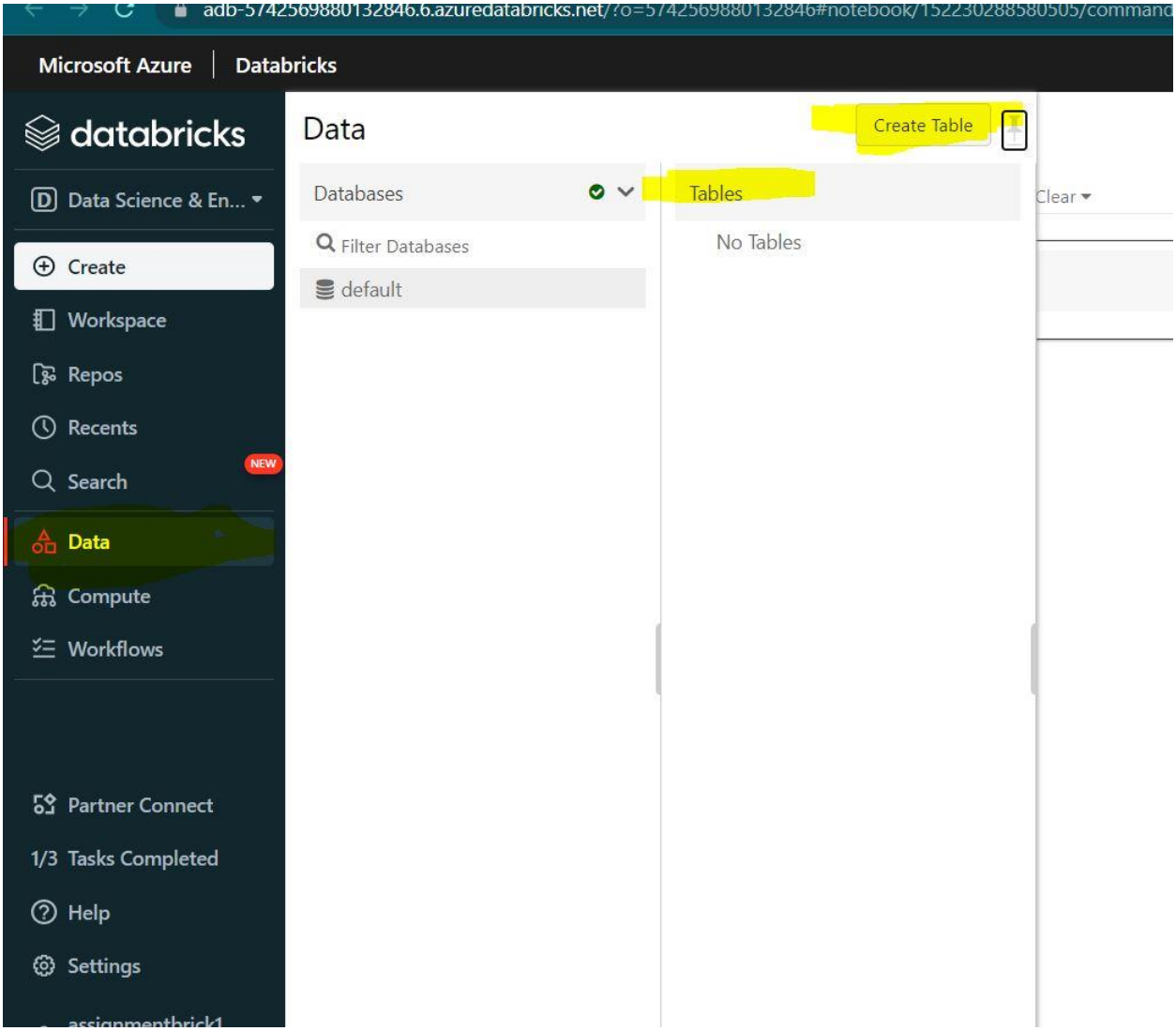
10. Explain the steps with screenshots how to insert data into databricks notebook & display the result?

A:

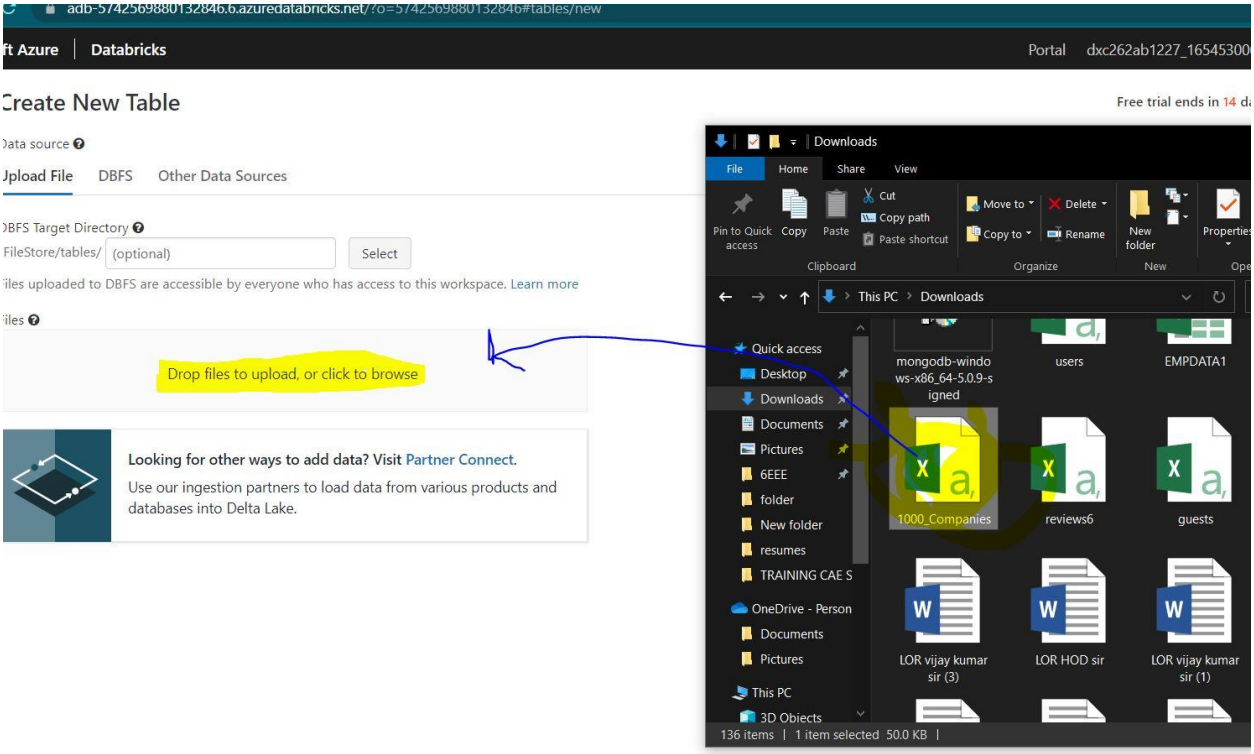
Step1: open the note book



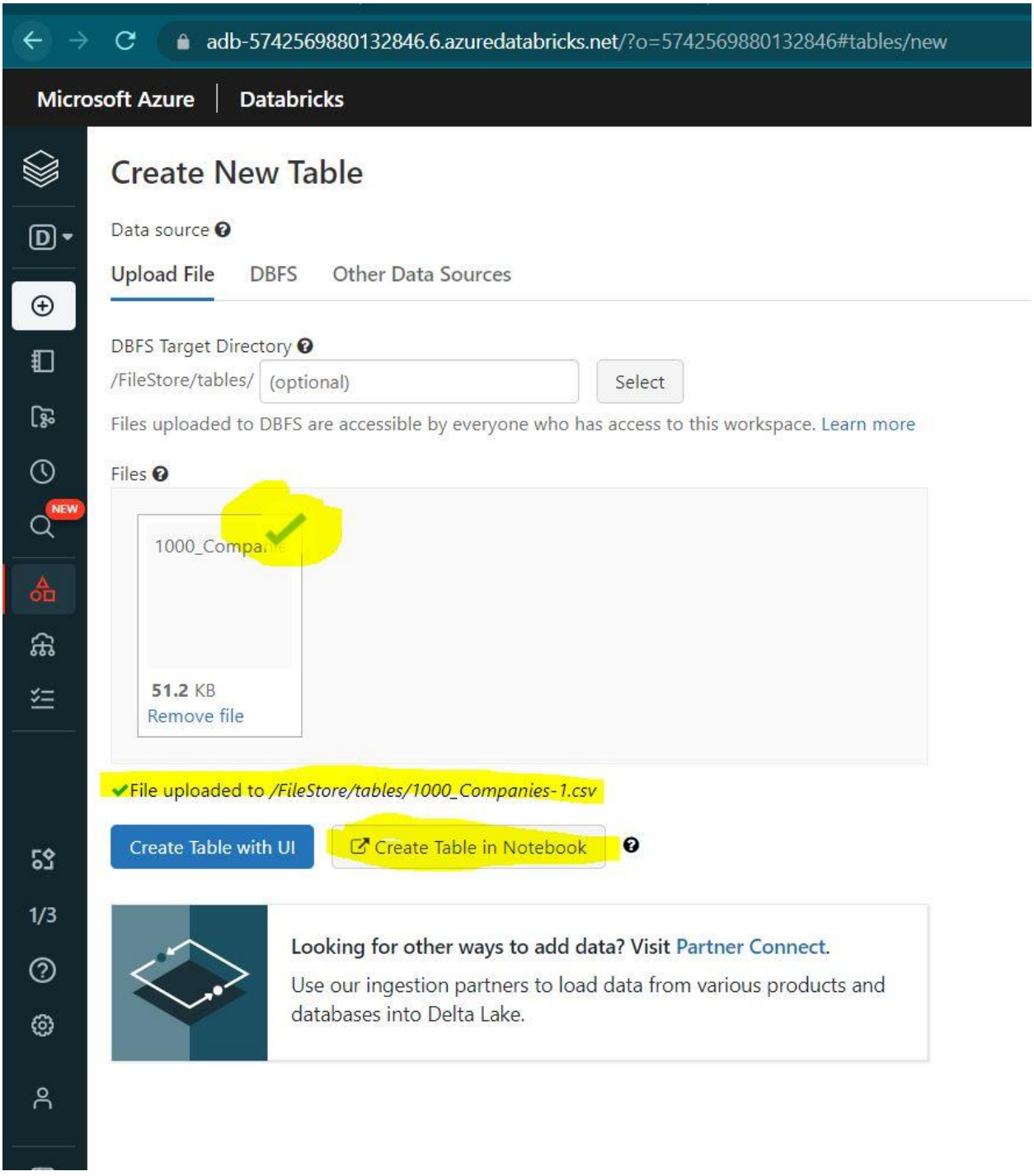
Step2: click on create and then table , create table



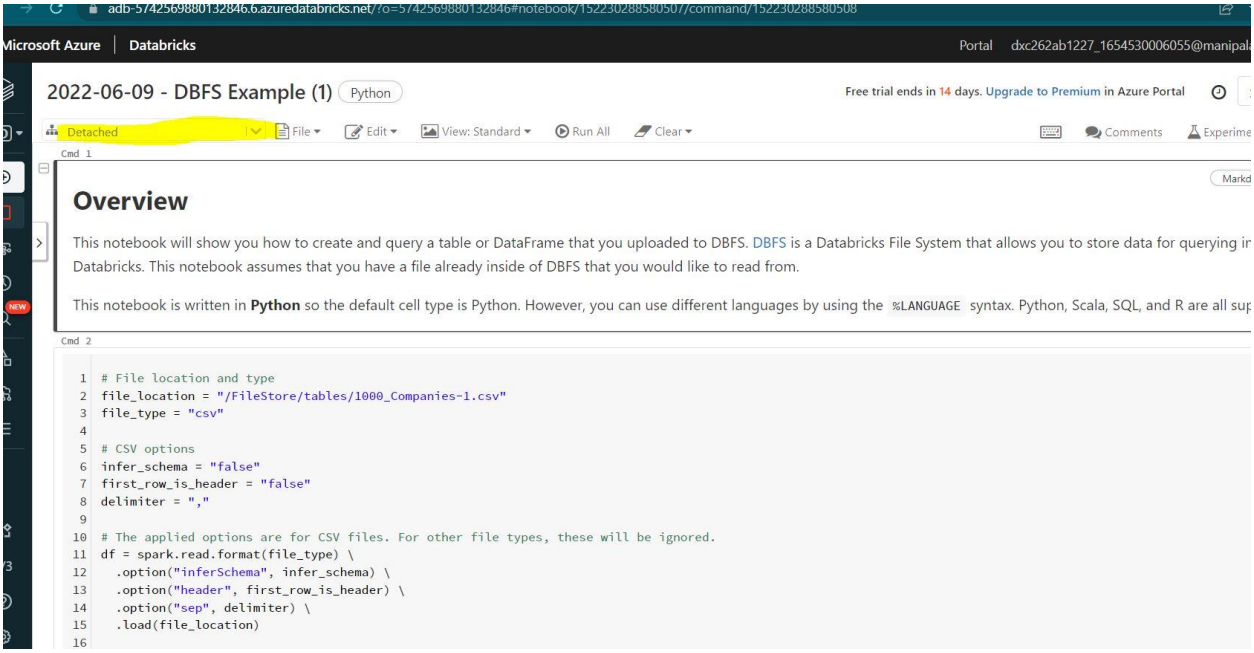
Step3: upload the data which you want.



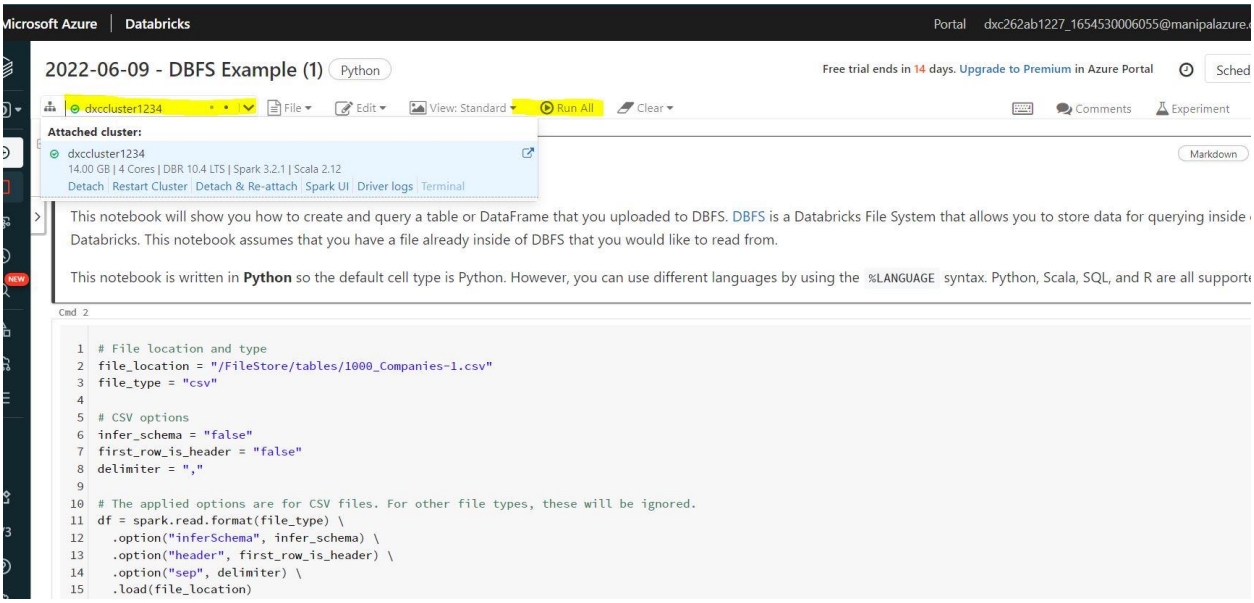
Step4: after uploading successfully click on create table in notebook



Step5: attach the cluster in the menu bar.



Step6: after attaching the cluster click on run all



Step7: hence the results are executed.

```
5    .load(file_location)
6
7  display(df)
```

▸ (2) Spark Jobs

▸ df: pyspark.sql.dataframe.DataFrame = [_c0: string, _c1: string ... 3 more fields]

able | Data Profile

	_c0	_c1	_c2	_c3	_c4
1	R&D Spend	Administration	Marketing Spend	State	Profit
2	165349.2	136897.8	471784.1	New York	192261.83
3	162597.7	151377.59	443898.53	California	191792.06
4	153441.51	101145.55	407934.54	Florida	191050.39
5	144372.41	118671.85	383199.62	New York	182901.99
6	142107.34	91391.77	366168.42	Florida	166187.94
7	131876.9	99814.71	362861.36	New York	156991.12

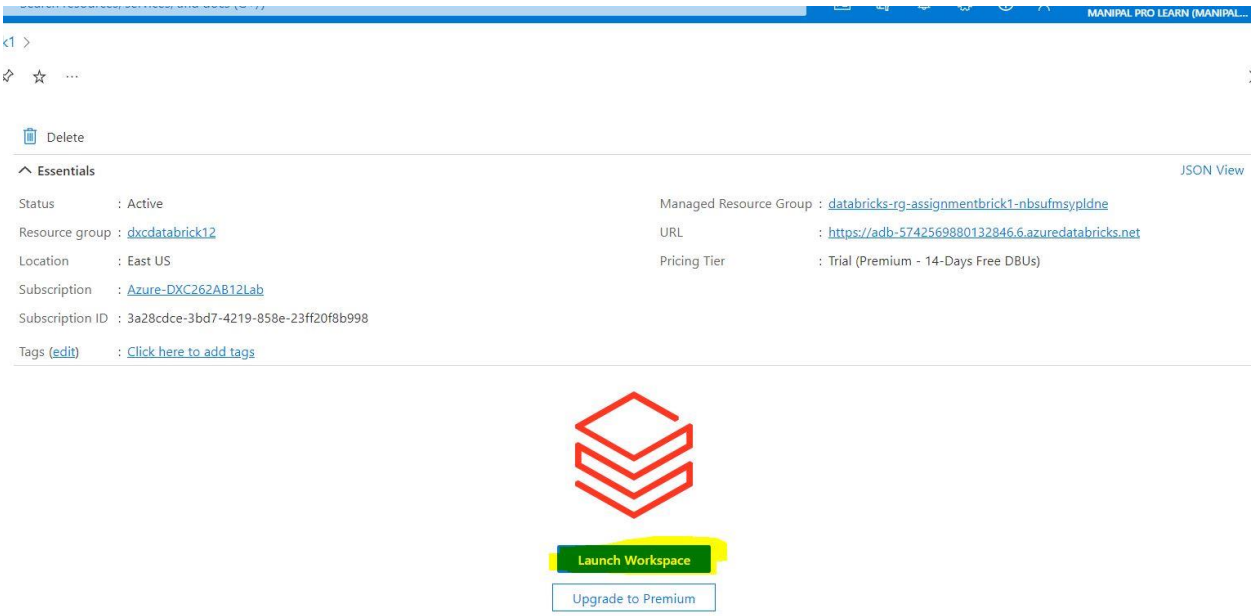
Truncated results, showing first 1000 rows.

Click to re-execute with maximum result limits.

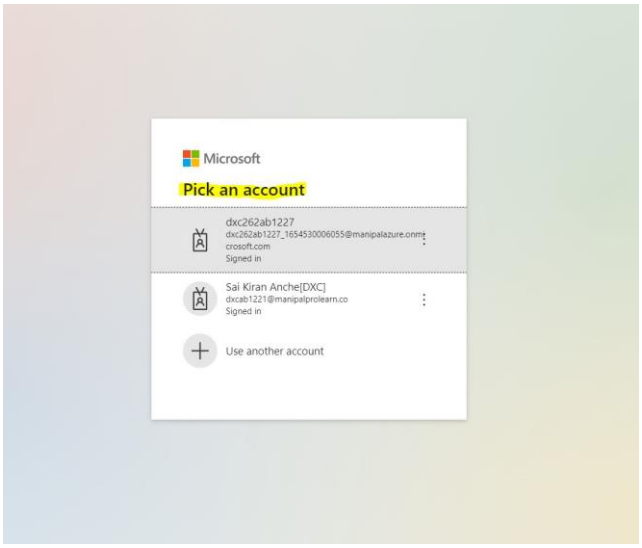
11. Explain the steps with screenshots how to create cluster in databricks ?

A:

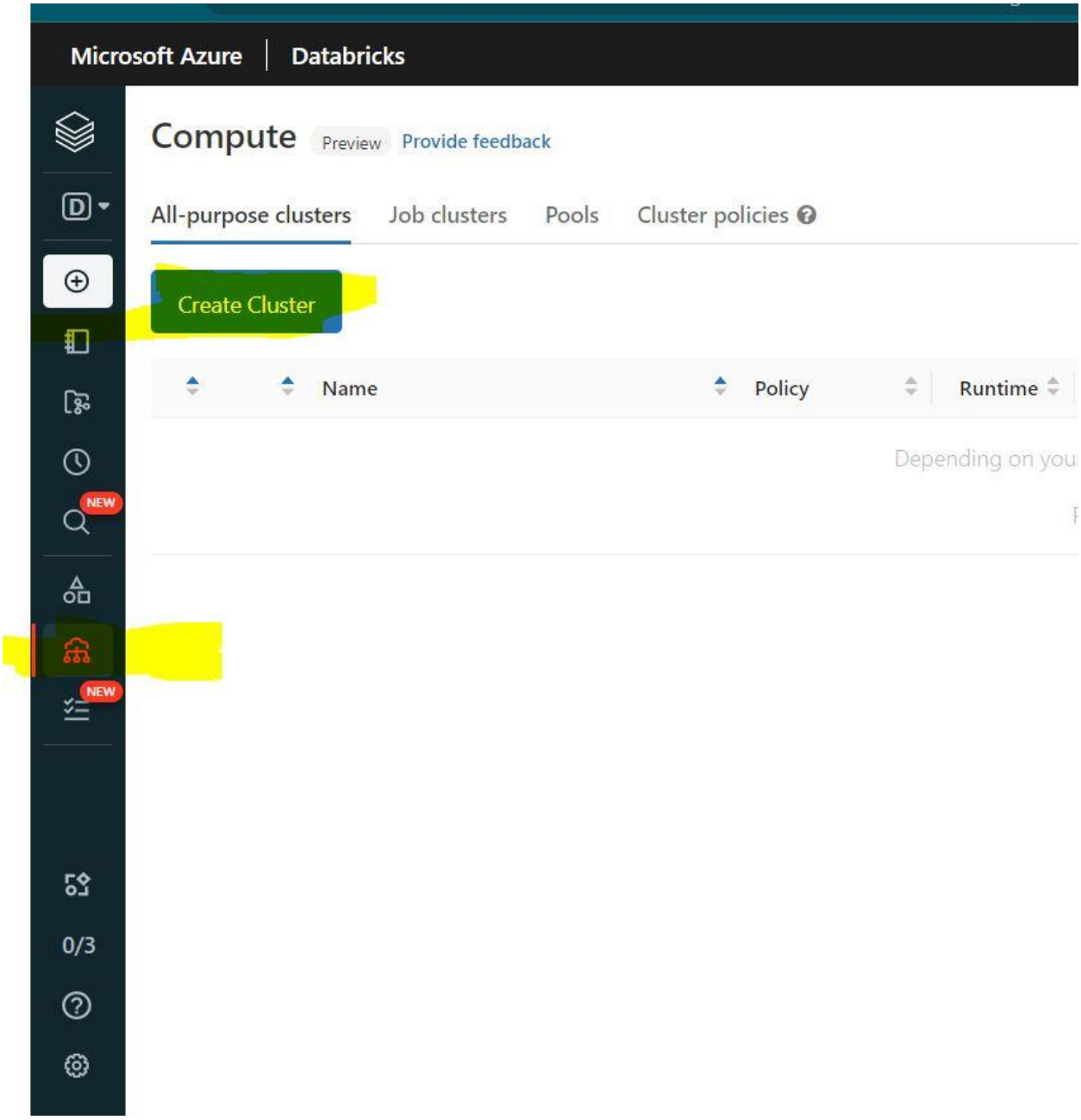
Step1: open overview page of data bricks and click on launch work space



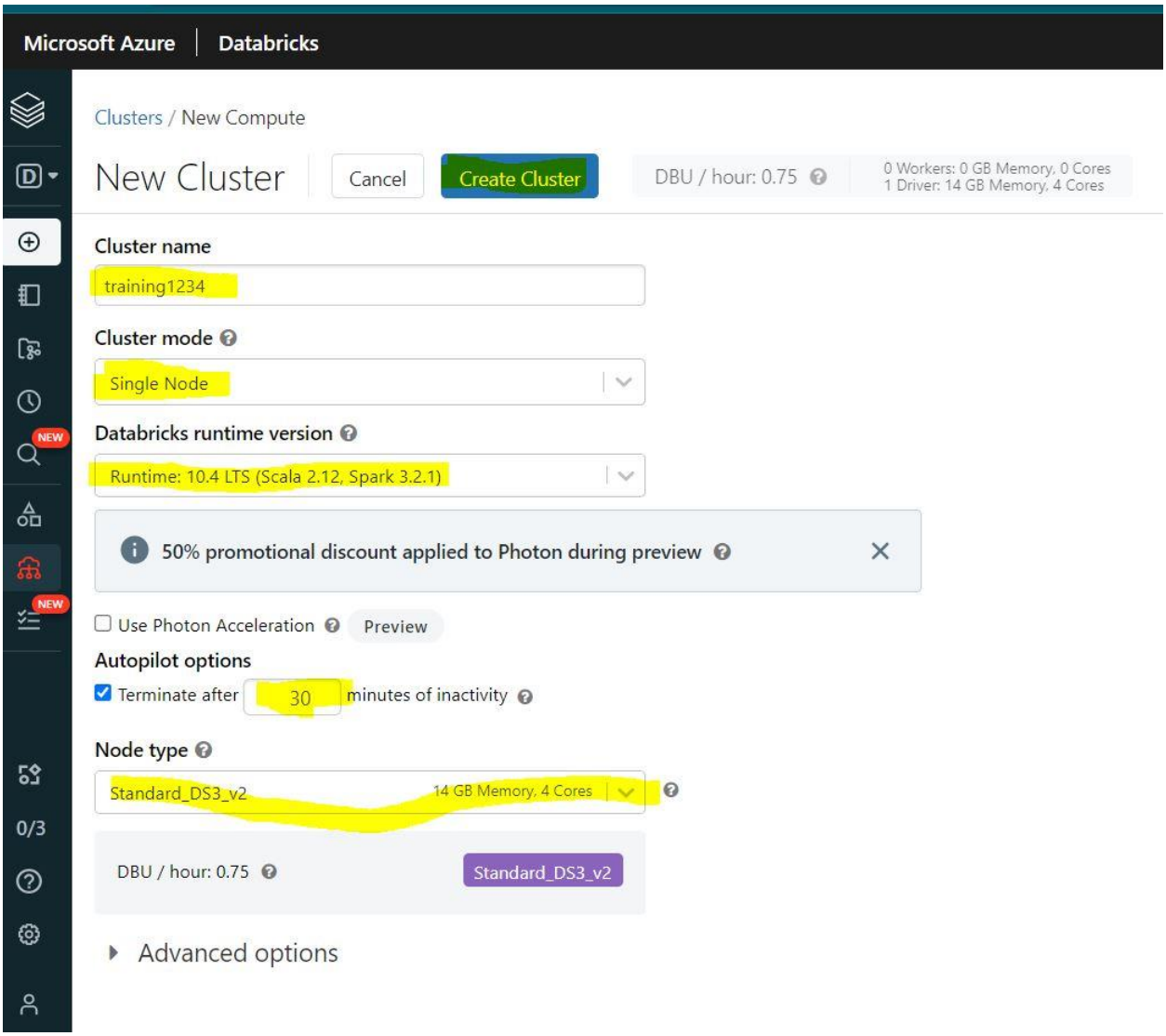
Step2: pick an account for the work space



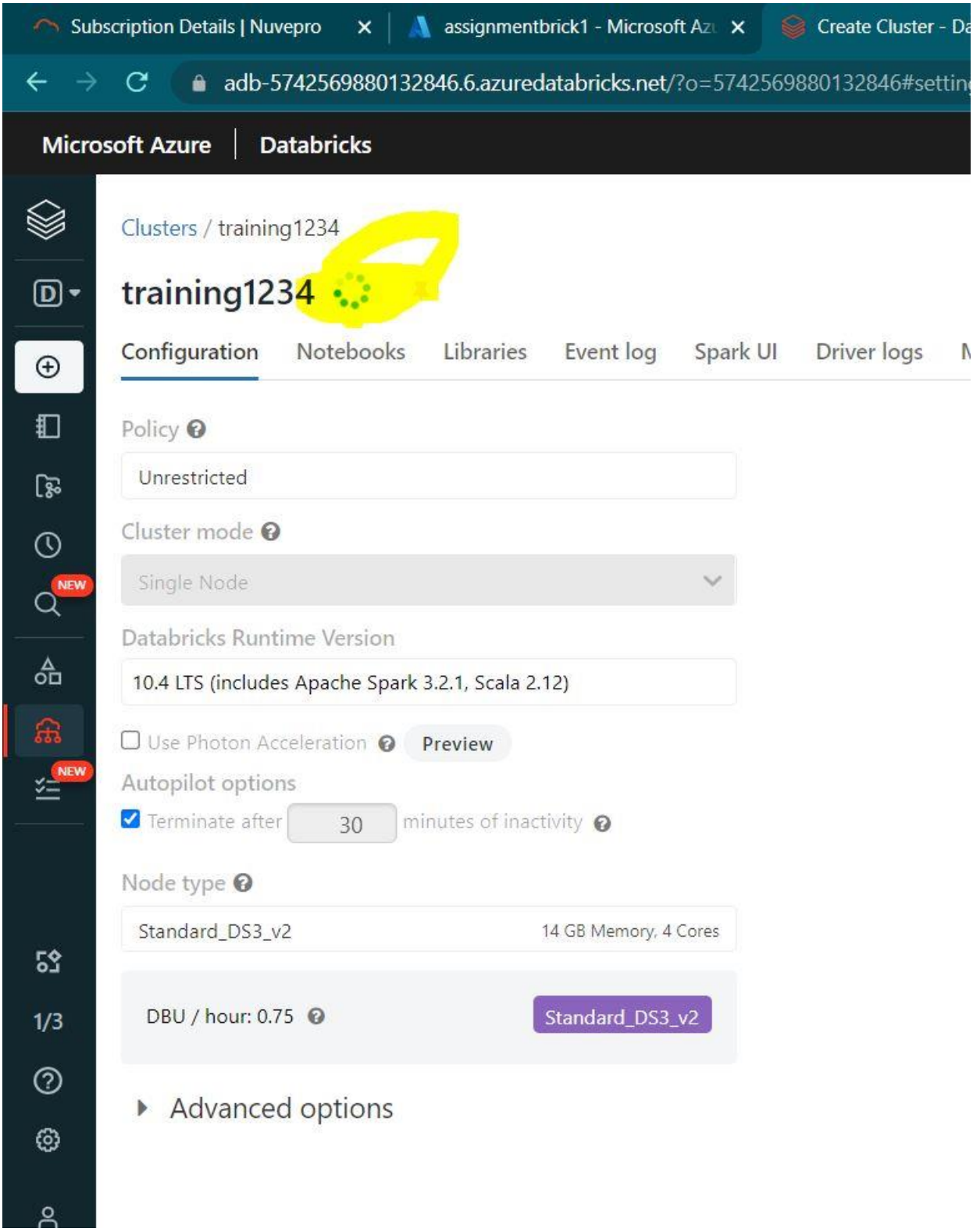
Step3: click on compute and then click on create cluster.



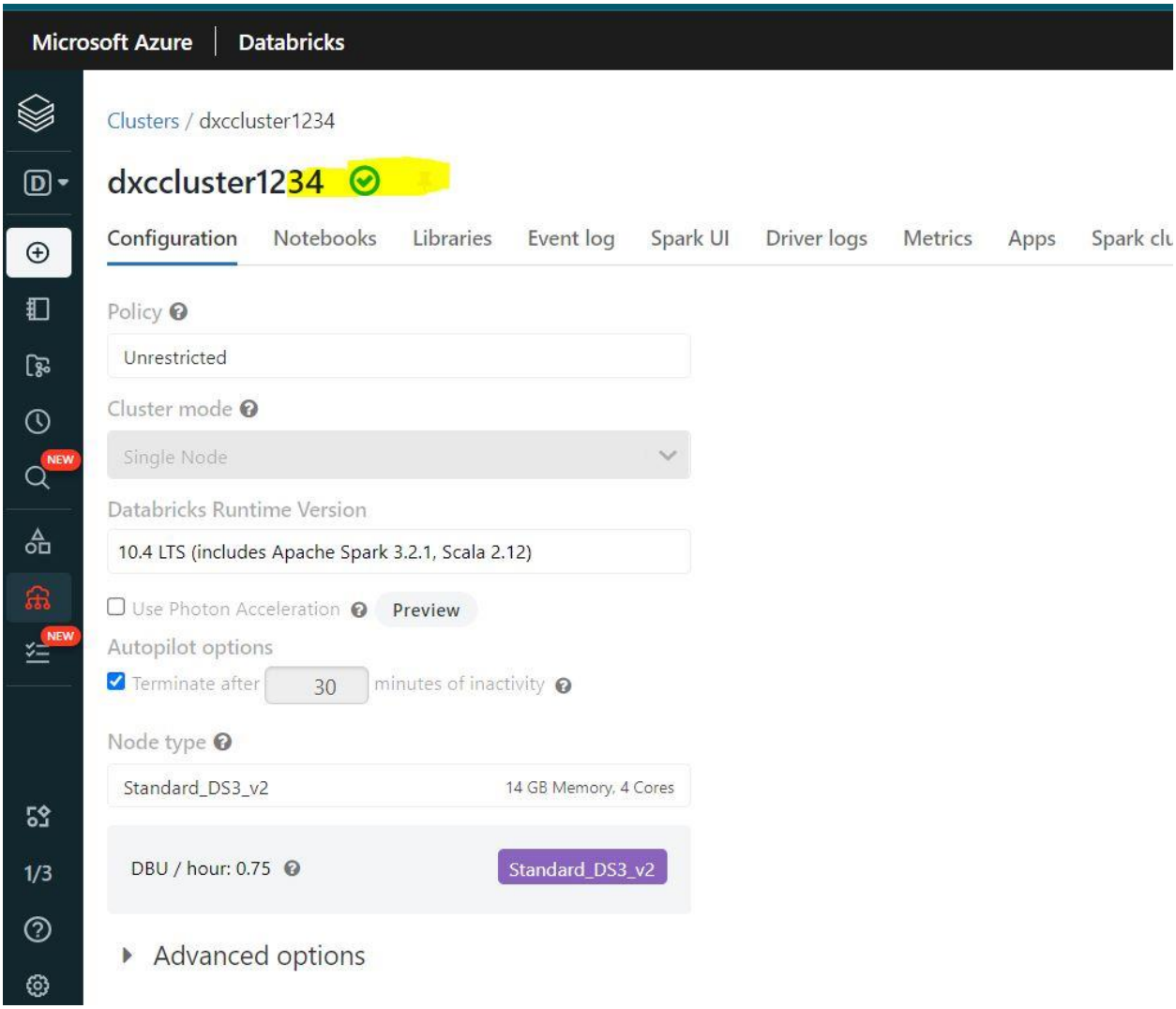
Step4: give the name and select the cluster mode , termination time etc and click on create cluster .



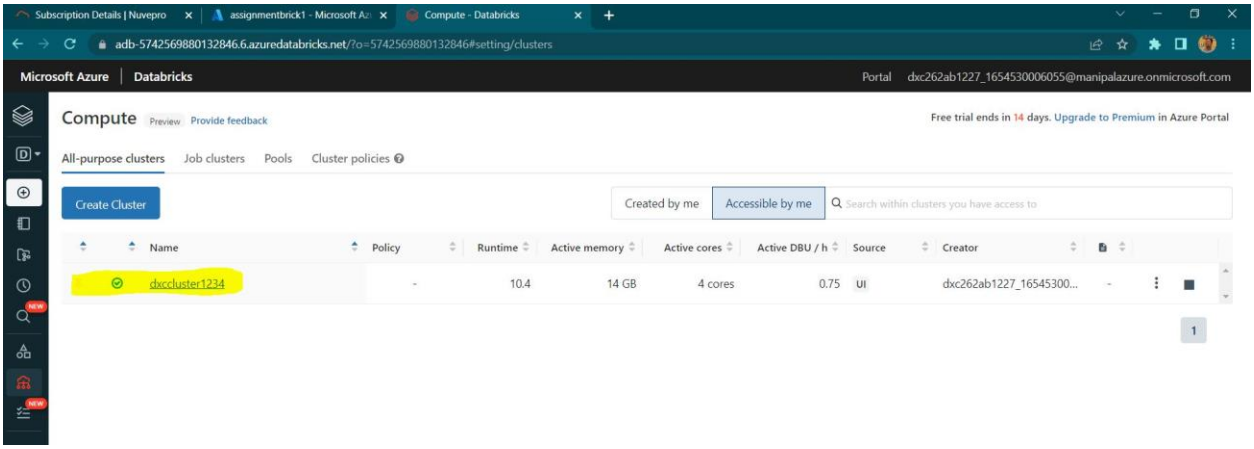
Step5: your cluster will be validating.



Step6: after successful creation of a cluster a green tick will appear



Step7: your cluster is successfully created.



RESULT

Almost all the test questions have been solved and presented successfully in the present document.

CONCLUSIONS

All the questions have been solved successfully with all the concepts that have been covered in the training session. It's really a great experience of learning while solving the cases. This assignment gave me immense confidence regarding my ability to upskill in new technologies.