

ASSIGNMENT 9

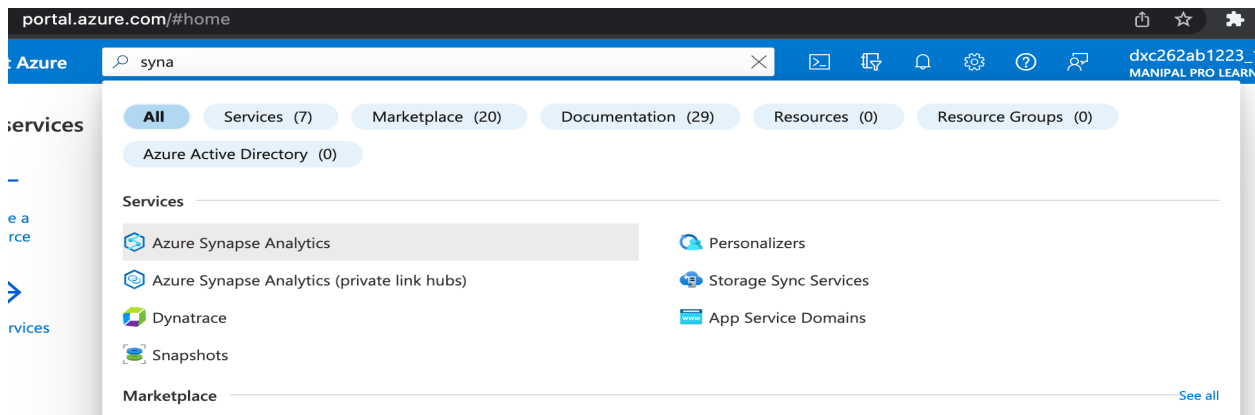
NAME – JYOTHI CHANDANA VOLETI
BATCH – DXC-262-ANALYTICS-B12-AZURE
EMPLOYEE DOMAIN –AZURE ANALYTICS
TRAINING UNDER – MANIPAL PRO LEARN
DATE OF SUBMISSION – 9TH JUNE 2022

ROLL NUMBER – DXC-262-AB-1218
COMPANY – DXC TECHNOLOGY
TRAINER NAME – MR. AJAY KUMAR
NO.OF QUESTIONS: 11

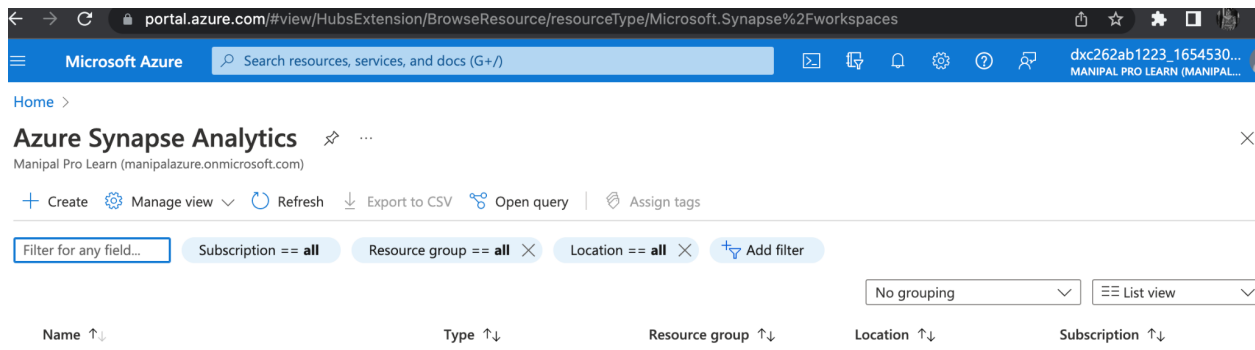
1. Explain the steps with screenshots how to create AzureSynapse analytics?

Ans:

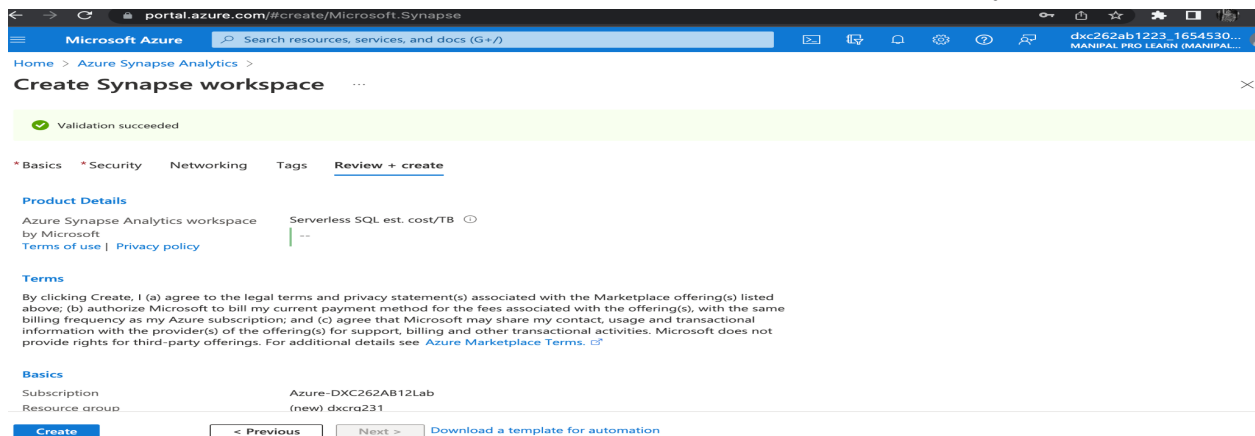
Step 1: Login to Azure portal and search synapse and click on “Azure Synapse Analytics”.



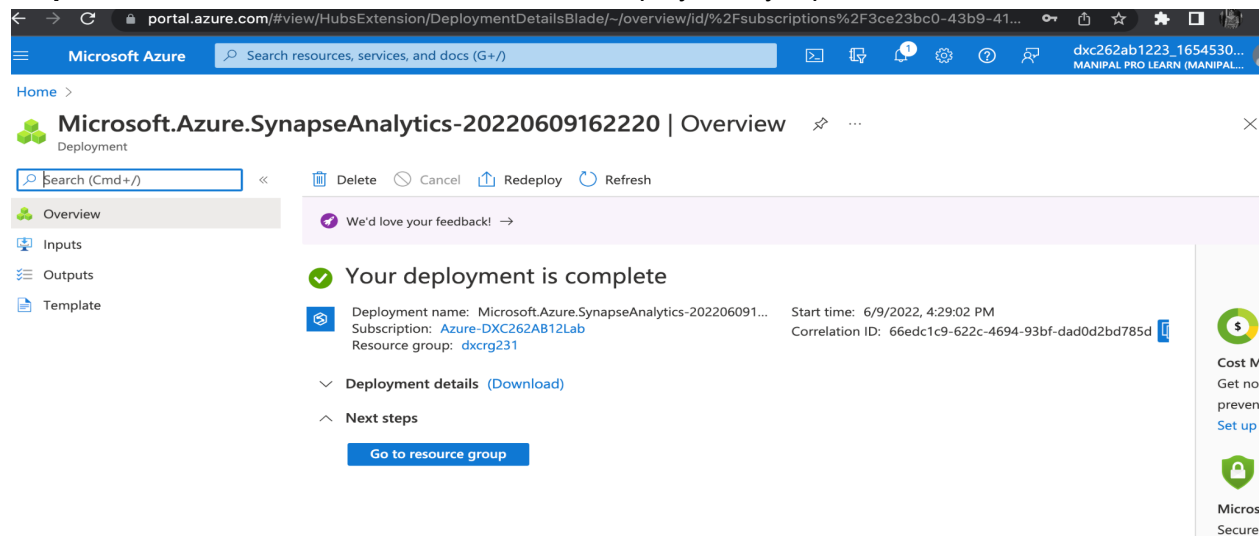
Step 2: Click on “create” to create azure synapse.



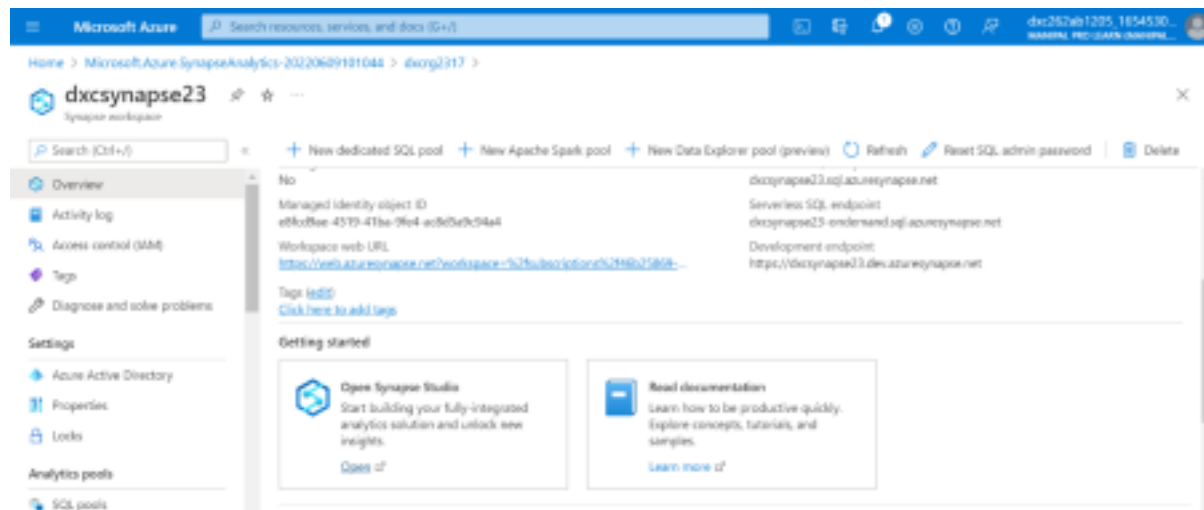
Step 3: Fill the required fields and click on “review + create” to validate the synapse.



Step 4: Now click on “create” at the bottom to deploy the synapse.



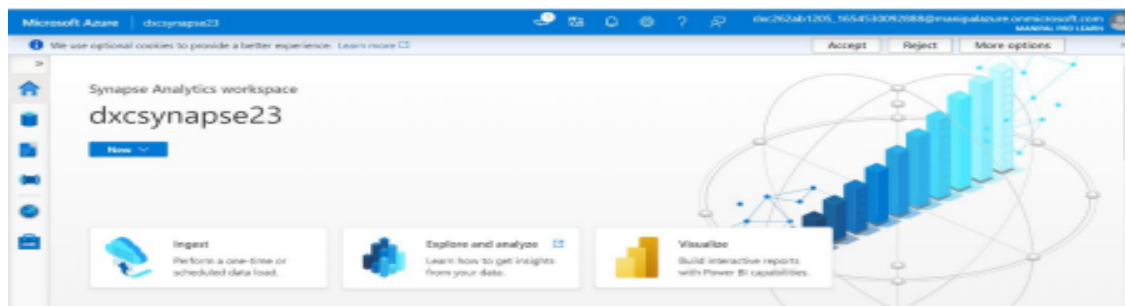
Step 5: Successfully created Azure Synapse Analytics.

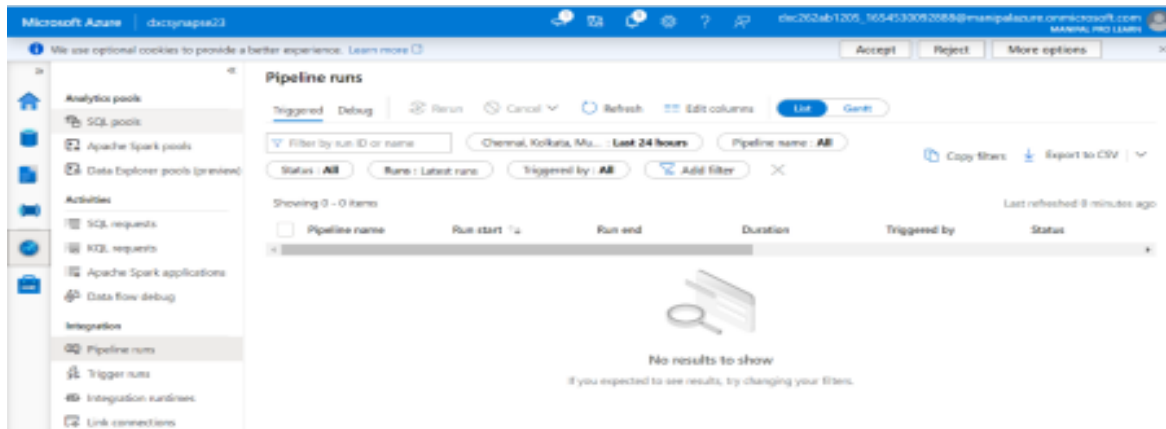


2. Explain the steps with screenshots how to SQL Pool in AzureSynapse analytics?

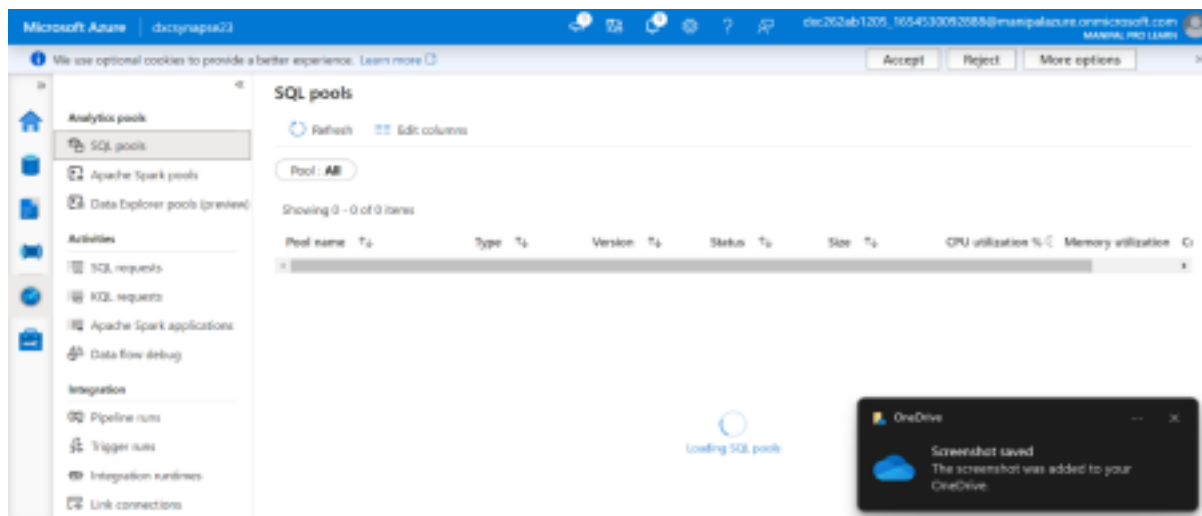
Ans:

Step 1: Open AzureSynapse and go to analytics.

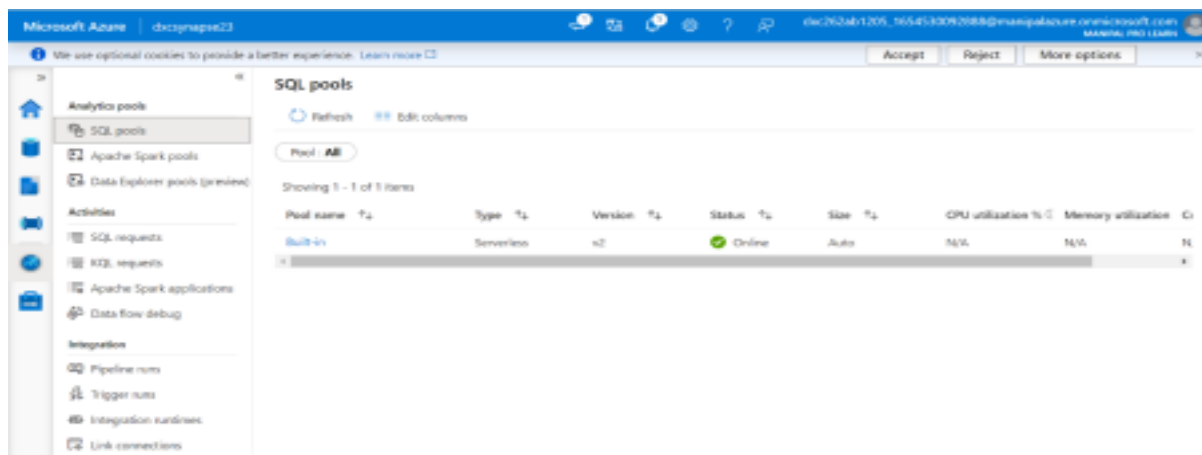




Step 2: In Analytics pools we can see SQL pools.



Step 3: Create a pool Built-in.



Step 4: after clicking on the new we need to finish the all details review and create.

New dedicated SQL pool

Validation succeeded.

Basics * Additional settings * Tags **Review + create**

Product details

Azure Synapse Analytics dedicated SQL pool by Microsoft
[Terms of use](#) | [Privacy policy](#)

Est. cost per hour
Failed to fetch billing info

Terms

By clicking "Create", I (a) agree to the legal terms and privacy statement(s) associated with the Marketplace offering(s) listed above; (b) authorize Microsoft to bill my current payment method for the fees associated with the offering(s), with the same billing frequency as my Azure subscription; and (c) agree that Microsoft may share my contact, usage and transactional information with the provider(s) of the offering(s) for support, billing and other transactional activities. Microsoft does not provide rights for third-party offerings. For additional details see [Azure Marketplace Terms](#)

Data source

Dedicated SQL pool name: dxcsql231
Performance level: DW100c

Additional settings

Use existing data: Blank
Collation: SQL_Latin1_General_CP1_CI_AS

Create < Previous Download template for automation

Step 5: the pool sql creation is done.

Synapse live Validate all Publish all

Home Data Develop Integrate Monitor Manage

Analytics pools
SQL pools
Apache Spark pools
Data Explorer pools (pre-)

External connections
Linked services
Microsoft Purview

Integration
Triggers
Integration runtimes
Security

SQL pools

The serverless SQL pool, Built-in, is immediately available for your workspace. Dedicated SQL pools can be configured to adapt to team or organizational requirements and constraints. [Learn more](#)

+ New Refresh

Filter by name

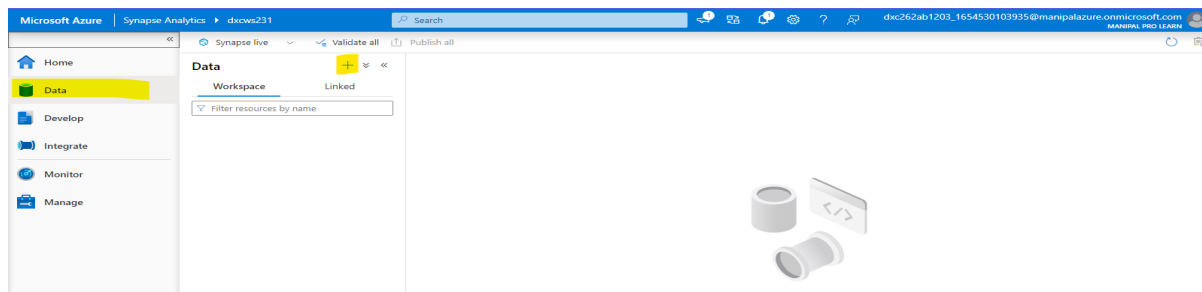
Showing 1-2 of 2 items (1 Serverless, 1 Dedicated)

Name	Type	Status	Size
Built-in	Serverless	Online	Auto
dxcsql231	Dedicated	Deploying	DW100c

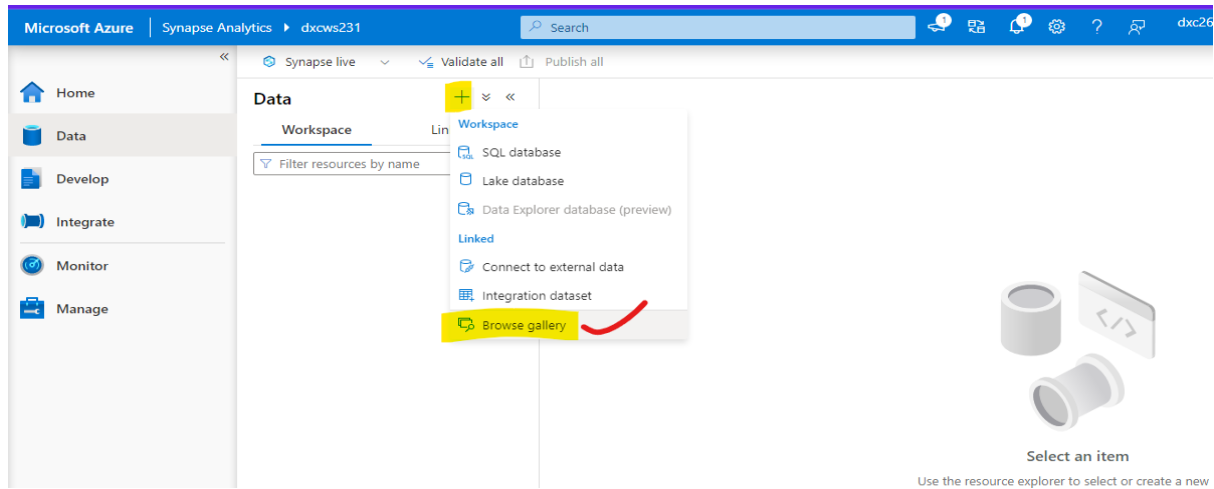
3. Explain the steps with screenshots how to import COVID19 dataset in AzureSynapse analytics and run sample 500 rows & display the output?

Ans: To insert the data in azure synapse analytics we have to follow the below mentioned steps

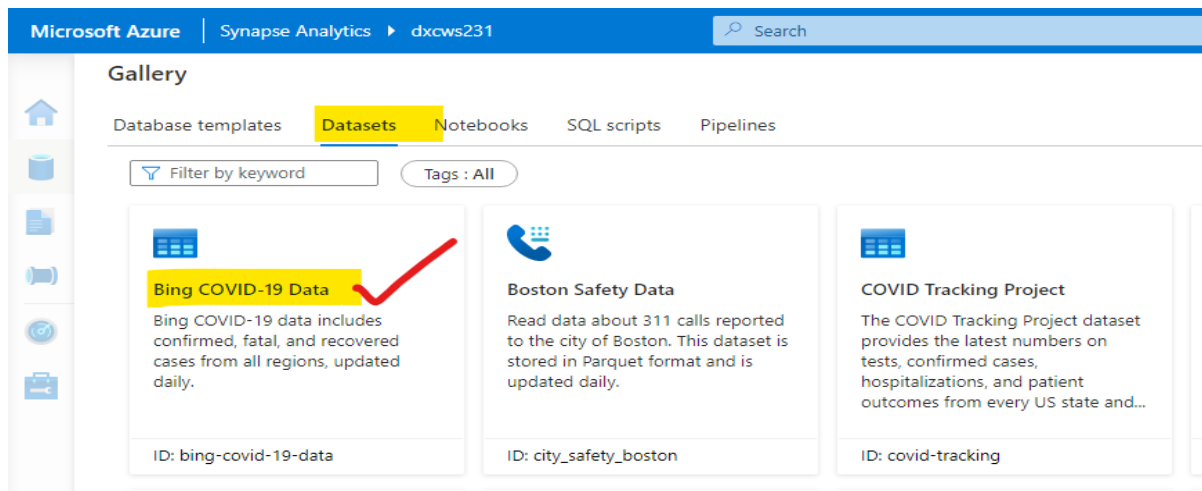
Step-1: navigate to the home page and click on data.



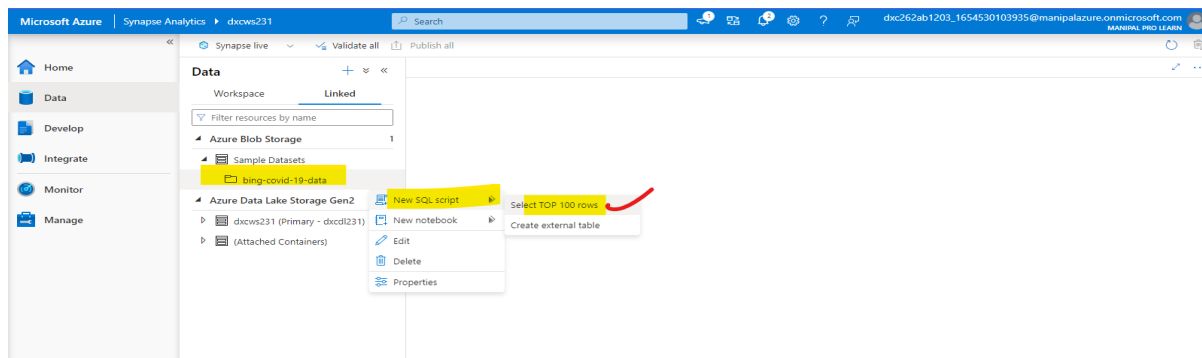
Step-2: click on the + button and choose browse gallery.



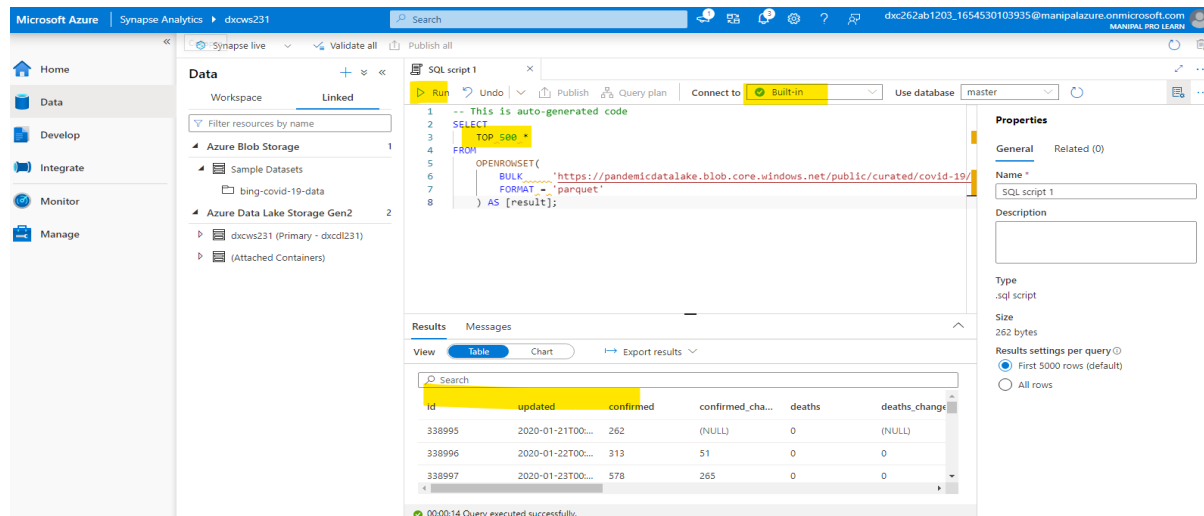
Step-3: select the covid 19 data from the menu from the data sets and add to the data sets after previewing the data.



Step-4: After that we will have to follow the following steps.



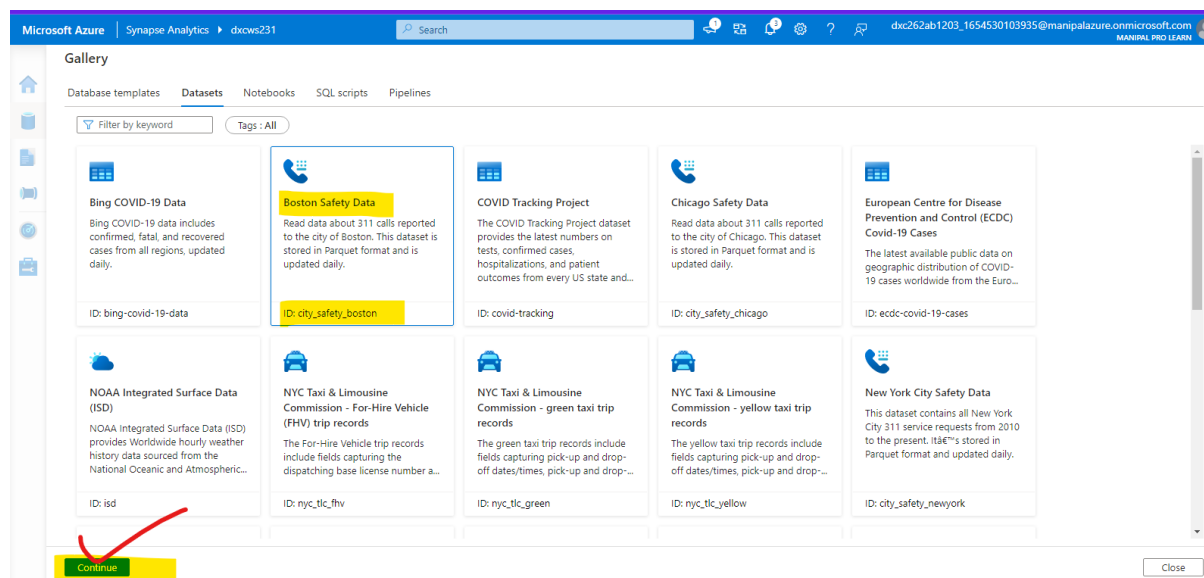
Step-5: After that run the command and the output data will be displayed.



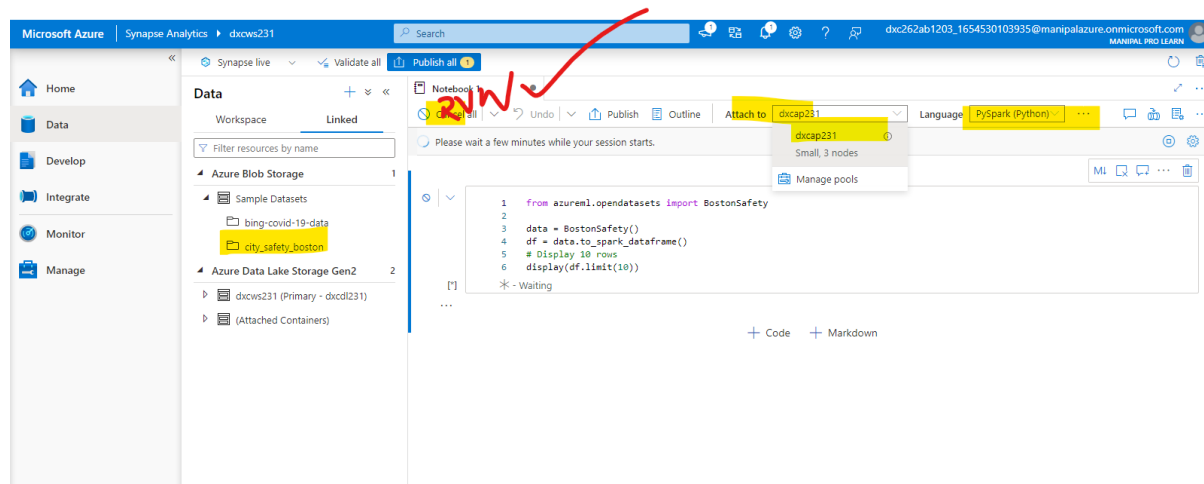
4. Explain the steps with screenshots how to input Boston Safety datasets into AzureSynapse analytics? Using Notebooks ?

Ans: In order to input Boston safety datasets into azure synapse analytics we have to follow the mentioned steps.

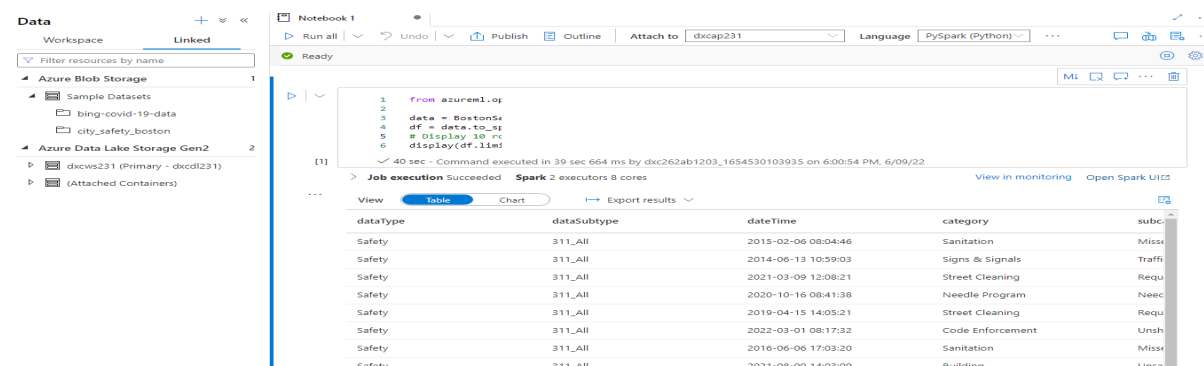
Step-1: select the Boston safety data sets from the data sets and click on continue.



Step-2: After getting data follow the steps refer attachment select the apache pool and Click on run it takes some time to initiate the first run.



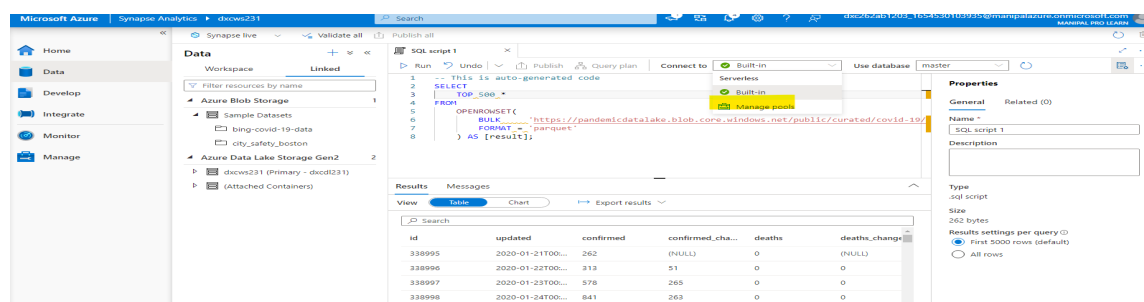
Step-3: after that the data will be displayed like this.



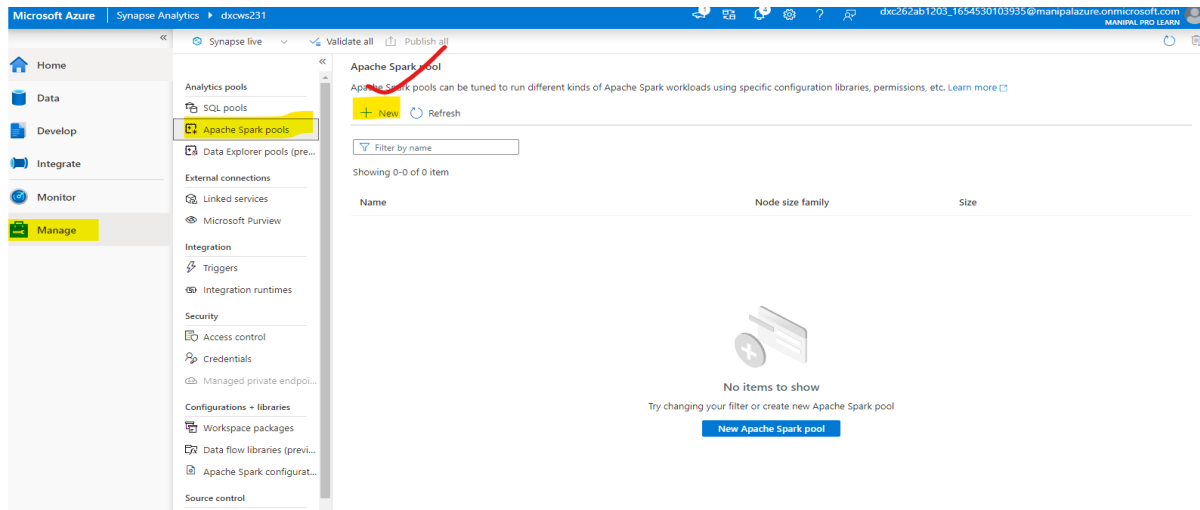
5. Explain the steps with screenshots how to create Spark pool in AzureSynapse analytics?

Ans: To create a spark pool we have to follow the steps mentioned below.

Step-1: click on manage pools.



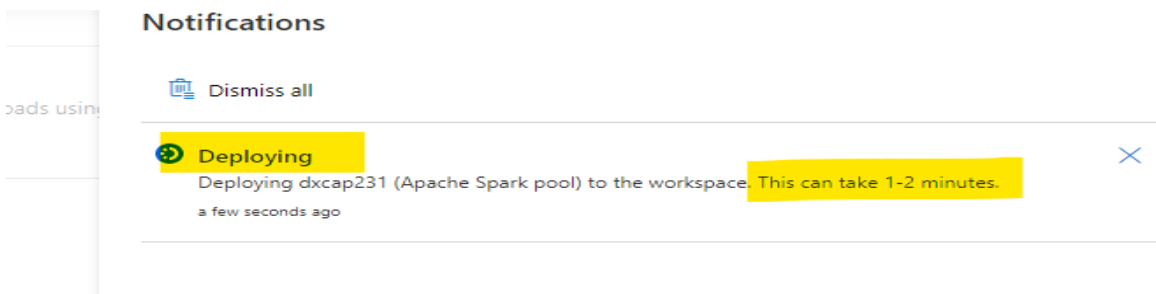
Step-2: After that it navigates to the manage page and select spark pool there refer screenshot.



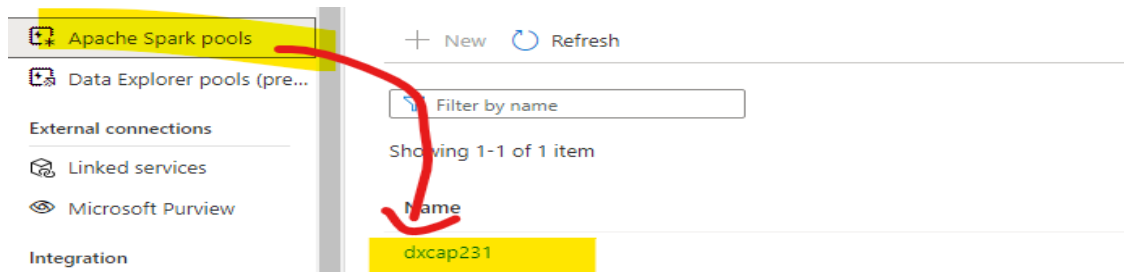
Step-3: fill all the requirements and click on review and create.

A screenshot of the 'New Apache Spark pool' form. The 'Basics' tab is active. The form contains fields for 'Apache Spark pool name' (dxcap231), 'Node size family' (Memory Optimized), 'Node size' (Small (4 vCores / 32 GB)), 'Autoscale' (Disabled), 'Number of nodes' (3), 'Estimated price' (Failed to fetch billing info), and 'Dynamically allocate executors' (Disabled). A red checkmark is placed over the 'Review + create' button.

Step-4 : It takes a few minutes to deploy.



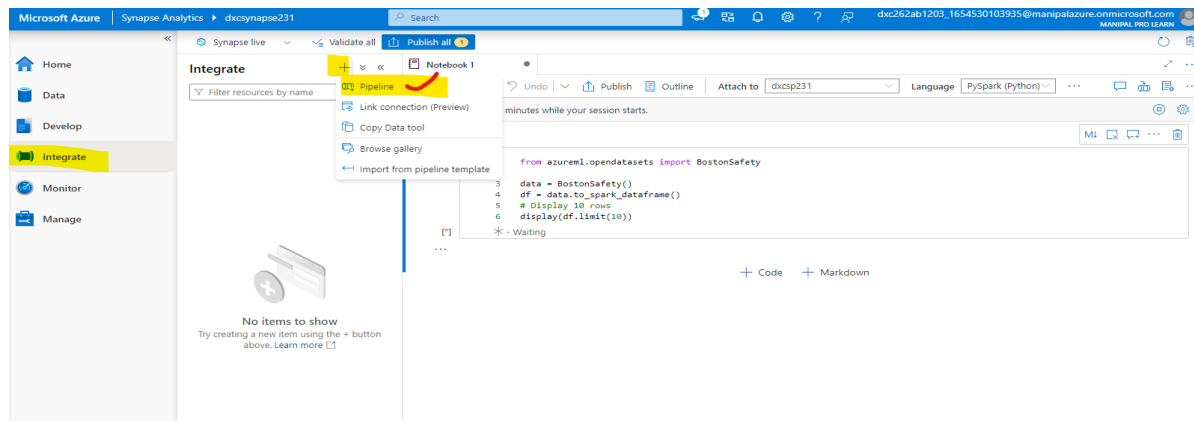
Step-5: Done.



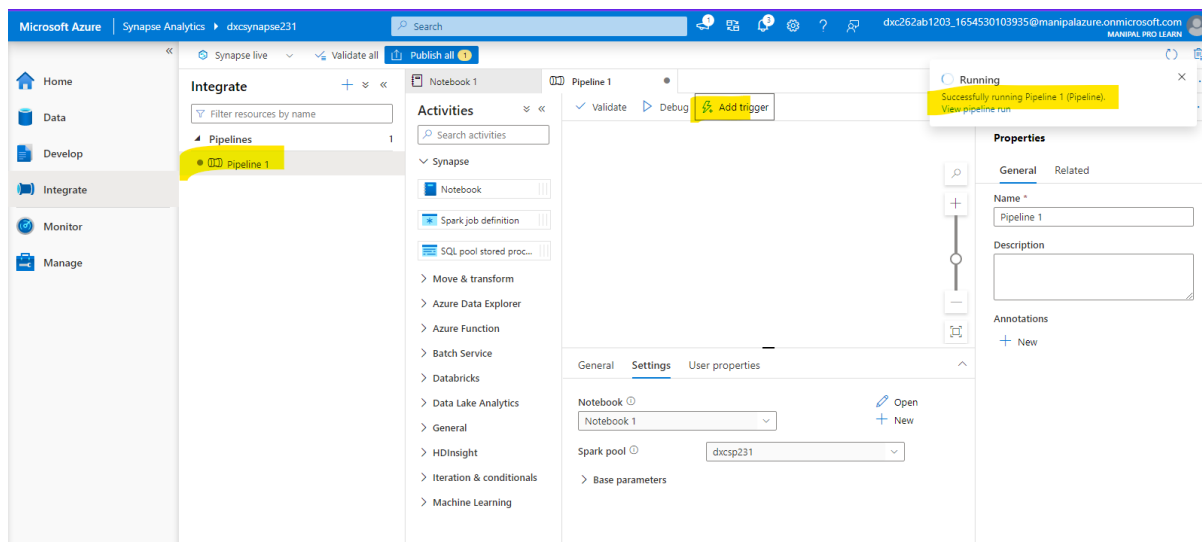
6. Explain the steps with screenshots how to create a pipeline in AzureSynapse analytics?

Ans: In azure synapse analytics we have to follow these steps to create a pipeline.

Step-1: click on integrate,”+” icon and after that click on pipeline.



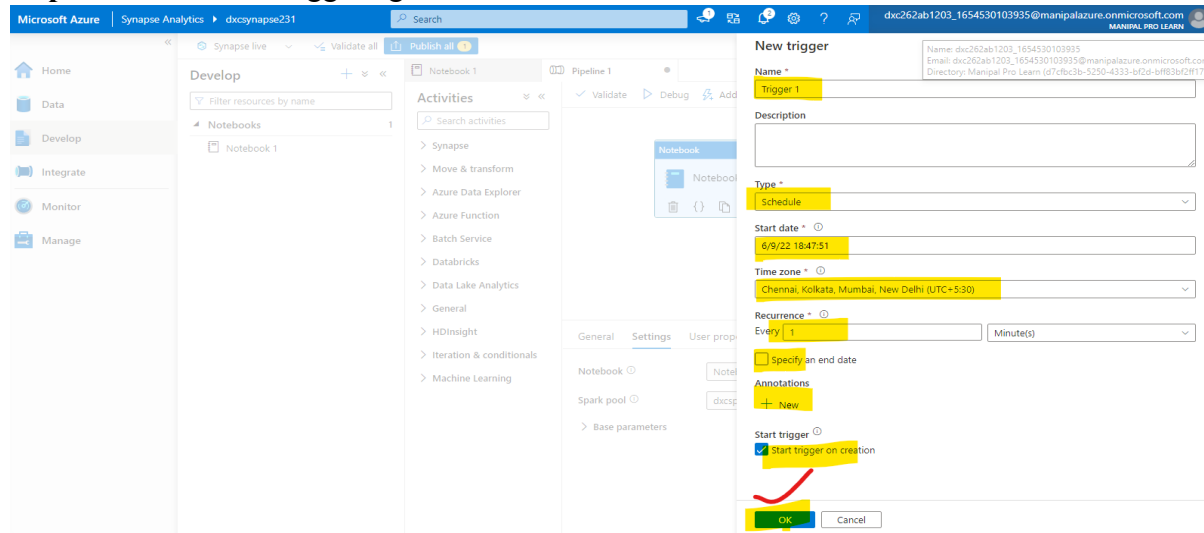
Step-2: The pipeline has been created successfully and triggered.



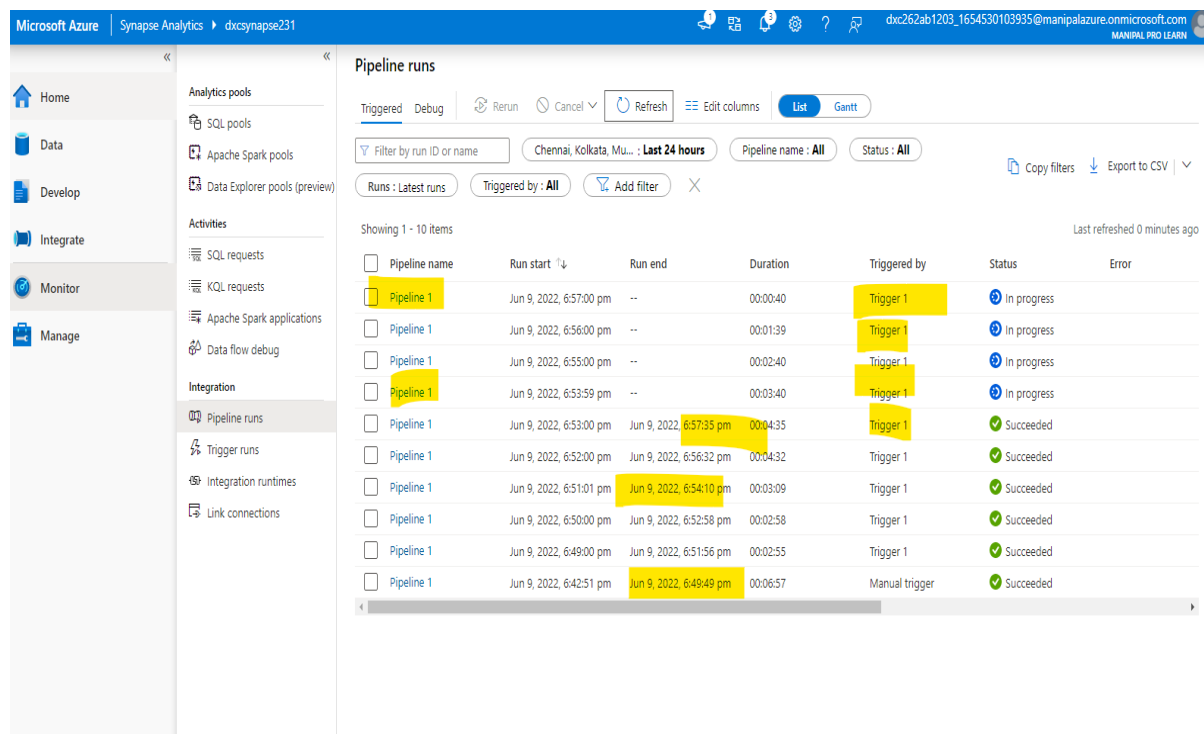
7. Explain the steps with screenshots how to automate the pipelines in Azure Synapse analytics?

Ans: Automation of pipelines is nothing but scheduling the pipelines to trigger at the time interval given.

Step-1: schedule the triggering time and click on ok.



Step-2 : The pipeline will Automatically trigger as per the limit that you have set.



Pipeline runs

Triggered Debug Rerun Cancel Refresh Edit columns List Gantt

Filter by run ID or name Chennai, Kolkata, Mu... : Last 24 hours Pipeline name : All Status : All

Runs : Latest runs Triggered by : All Add filter

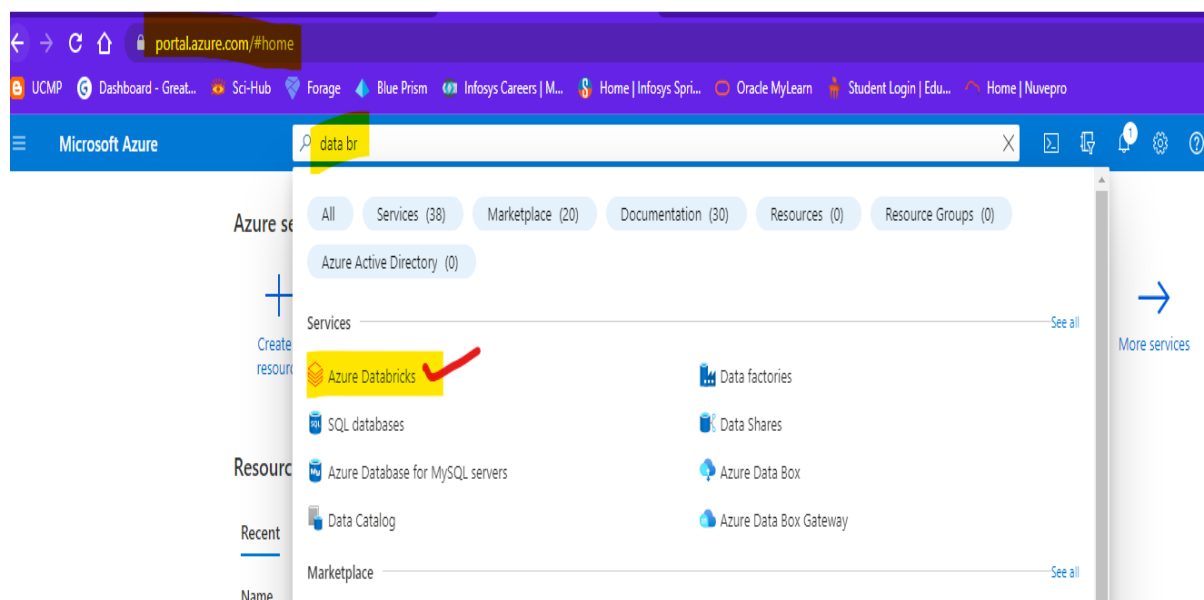
Showing 1 - 10 items Last refreshed 0 minutes ago

<input type="checkbox"/>	Pipeline name	Run start	Run end	Duration	Triggered by	Status	Error
<input type="checkbox"/>	Pipeline 1	Jun 9, 2022, 6:57:00 pm	--	00:00:40	Trigger 1	In progress	
<input type="checkbox"/>	Pipeline 1	Jun 9, 2022, 6:56:00 pm	--	00:01:39	Trigger 1	In progress	
<input type="checkbox"/>	Pipeline 1	Jun 9, 2022, 6:55:00 pm	--	00:02:40	Trigger 1	In progress	
<input type="checkbox"/>	Pipeline 1	Jun 9, 2022, 6:53:59 pm	--	00:03:40	Trigger 1	In progress	
<input type="checkbox"/>	Pipeline 1	Jun 9, 2022, 6:53:00 pm	Jun 9, 2022, 6:57:35 pm	00:04:35	Trigger 1	Succeeded	
<input type="checkbox"/>	Pipeline 1	Jun 9, 2022, 6:52:00 pm	Jun 9, 2022, 6:56:32 pm	00:04:32	Trigger 1	Succeeded	
<input type="checkbox"/>	Pipeline 1	Jun 9, 2022, 6:51:01 pm	Jun 9, 2022, 6:54:10 pm	00:03:09	Trigger 1	Succeeded	
<input type="checkbox"/>	Pipeline 1	Jun 9, 2022, 6:50:00 pm	Jun 9, 2022, 6:52:58 pm	00:02:58	Trigger 1	Succeeded	
<input type="checkbox"/>	Pipeline 1	Jun 9, 2022, 6:49:00 pm	Jun 9, 2022, 6:51:56 pm	00:02:55	Trigger 1	Succeeded	
<input type="checkbox"/>	Pipeline 1	Jun 9, 2022, 6:42:51 pm	Jun 9, 2022, 6:49:49 pm	00:06:57	Manual trigger	Succeeded	

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

Ans: To create data bricks we have to follow the following steps.

Step-1: login into the azure portal and search for the Data Bricks.



Step-2: click on azure data bricks and it will navigate you to the page. And click on create-to-create data bricks.

Home > Azure Databricks >

Create an Azure Databricks workspace

Basics Networking Advanced Tags Review + create

Project Details

Select the subscription to manage deployed resources and costs. Use resource groups like folders to organize and manage all your resources.

Subscription * ⓘ Azure-DXC262AB12Lab

Resource group * ⓘ dxcorg231
[Create new](#)

Instance Details

Workspace name * dxcdb231

Region * East US

Pricing Tier * ⓘ Standard (Apache Spark, Secure with Azure AD)

[Review + create](#) < Previous Next : Networking >

Step-3: After checking all these reviews and creating and waiting for deployment after deployment we will get like this.

Notifications

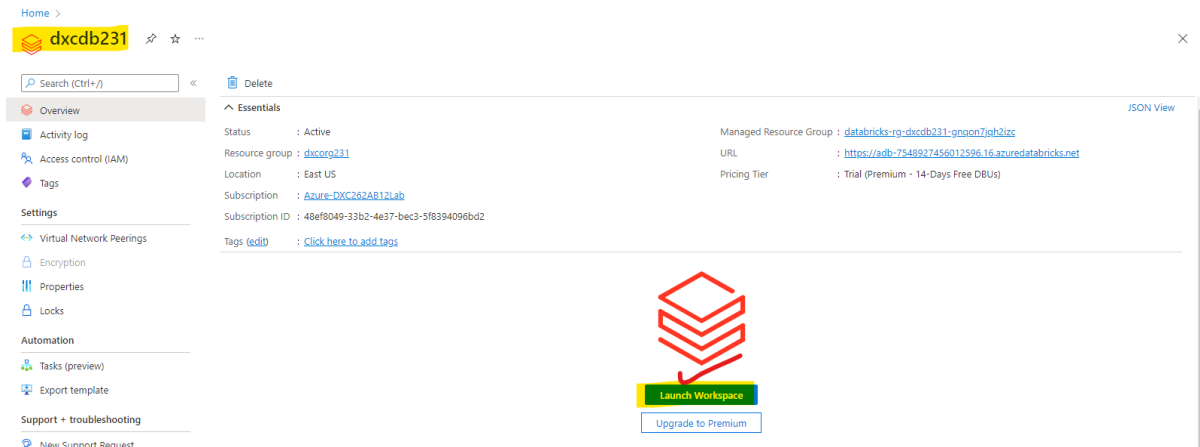
[More events in the activity log →](#) Dis

✓ **Deployment succeeded**

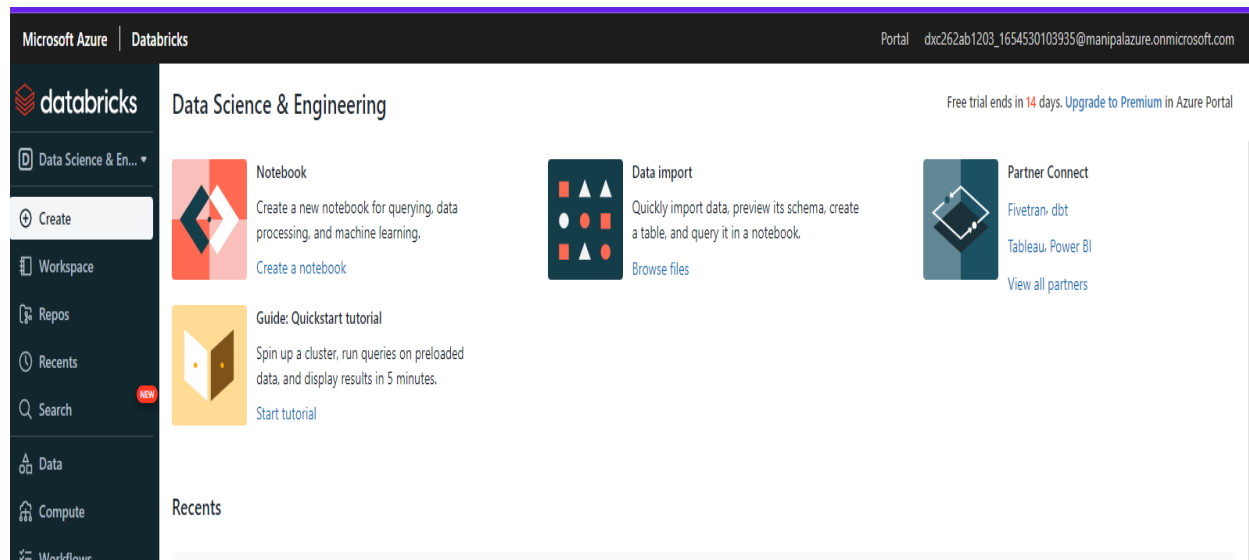
Deployment 'dxcorg231_dxcdb231' to resource group 'dxcorg231' wa

[Go to resource](#) [Pin to dashboard](#)

Step-4: after clicking the go to resource button you are navigated to the data bricks.



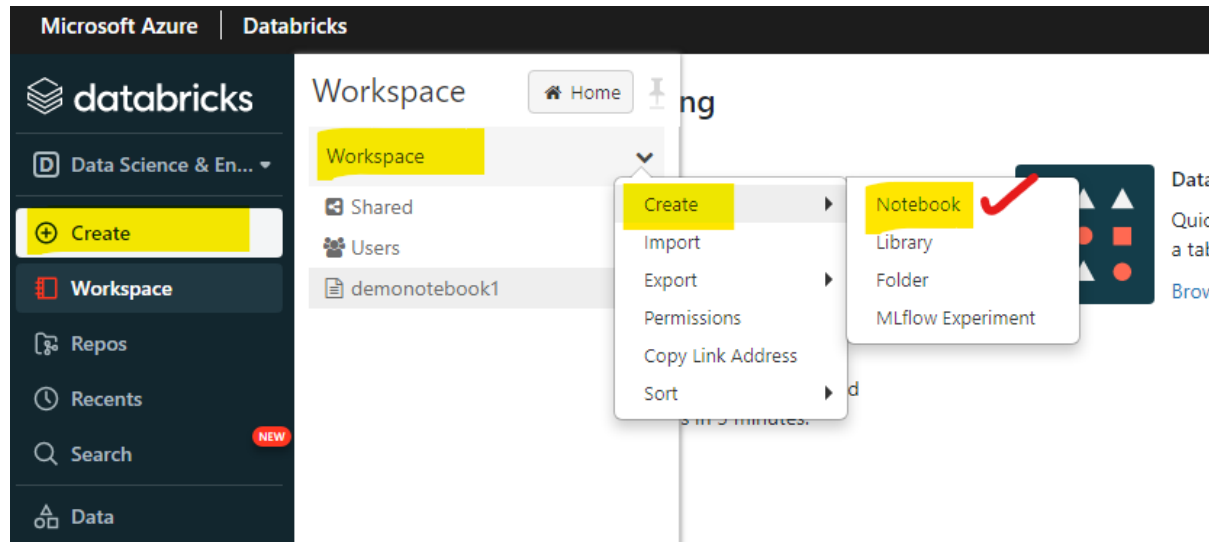
Step-5: After clicking on launch workspace, you are in data bricks.



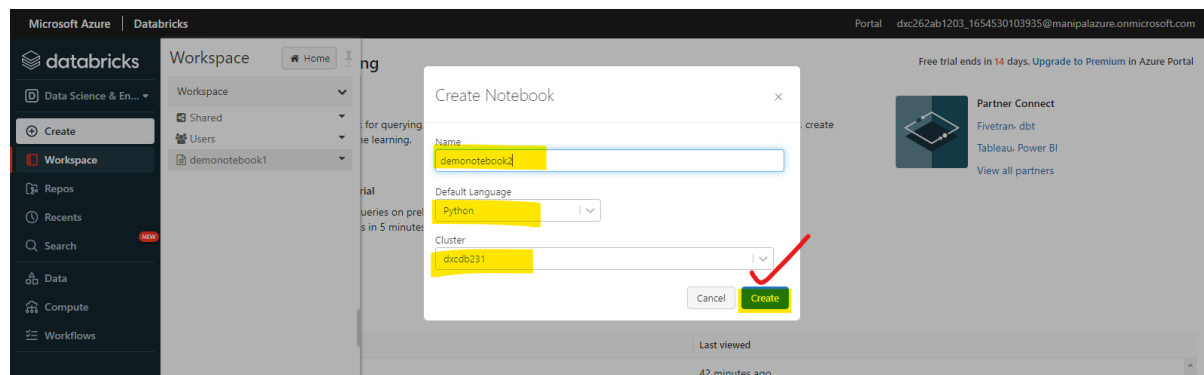
9. Explain the steps with screenshots how to create notebooks in Databricks?

Ans: To create Notebooks in Data bricks we have to follow the below mentioned steps.

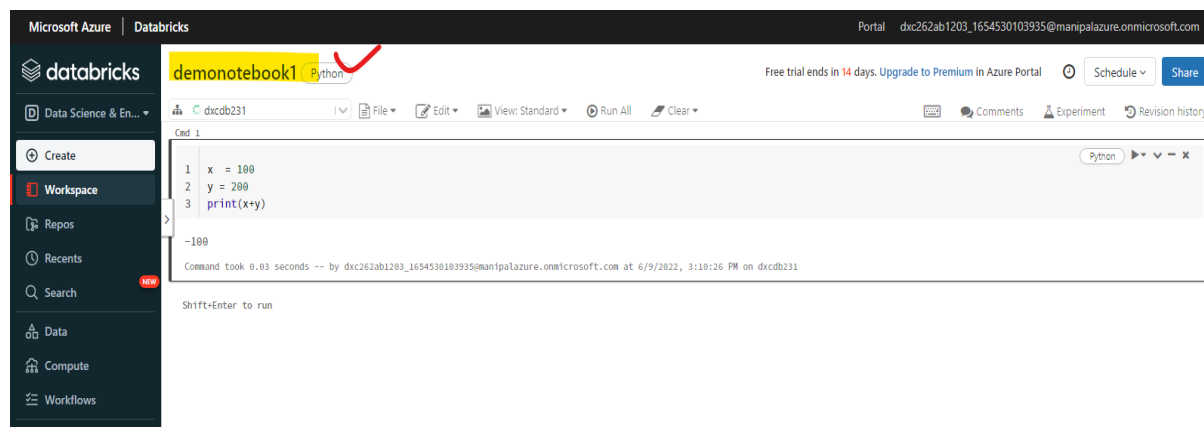
Step-1 : open azure portal and navigate to azure data bricks after that follow the attached screenshots.



Step-2: click on the note book and name it and choose the language cluster.



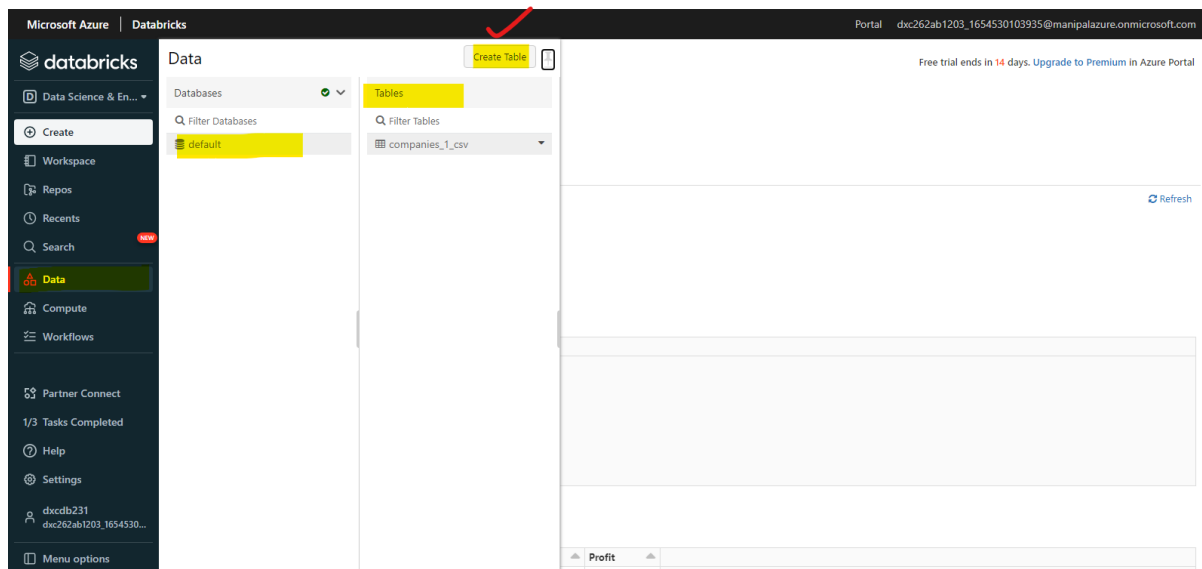
Step-3: The notebook is created in data bricks.



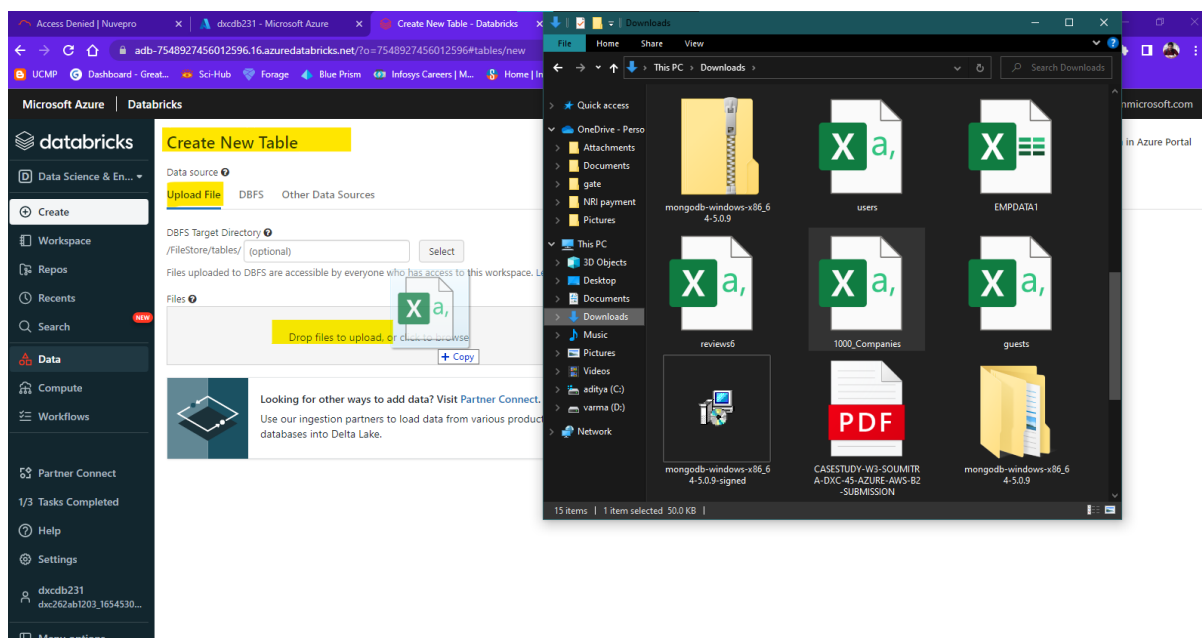
10. Explain the steps with screenshots how to insert data into databricks notebook & display the result?

Ans: In order to insert data into databricks notebook we can insert data as mentioned below.

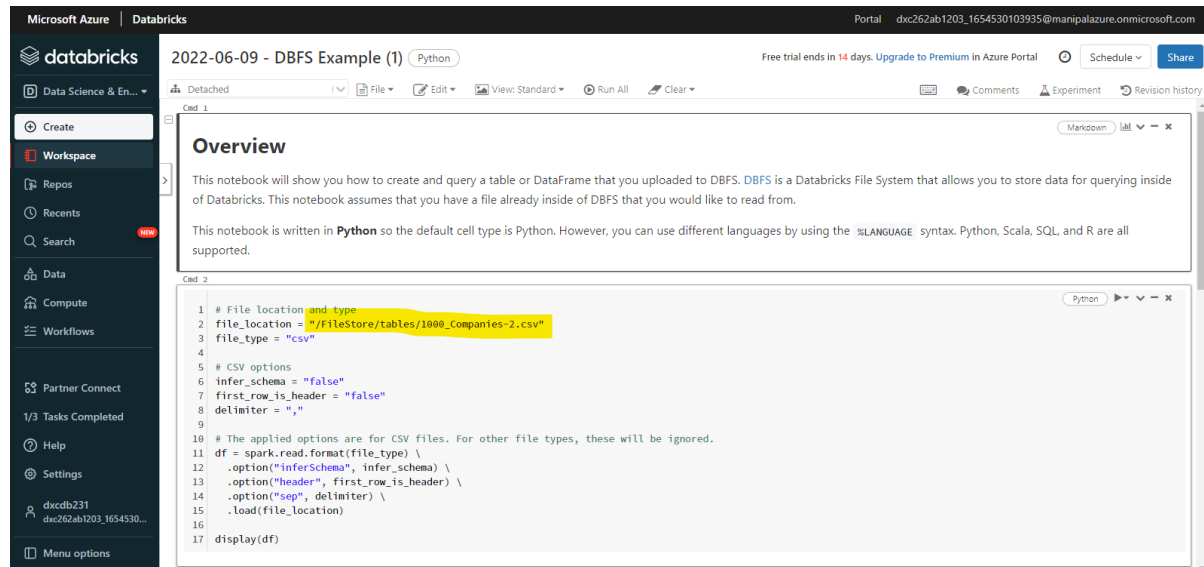
Step-1: open the data bricks workspace and click on data and it will navigate to a tab and click on create table.



Step-2 : after clicking on create table. We can drag and drop the files which we need to insert into the notebook refer attachment for better understanding.



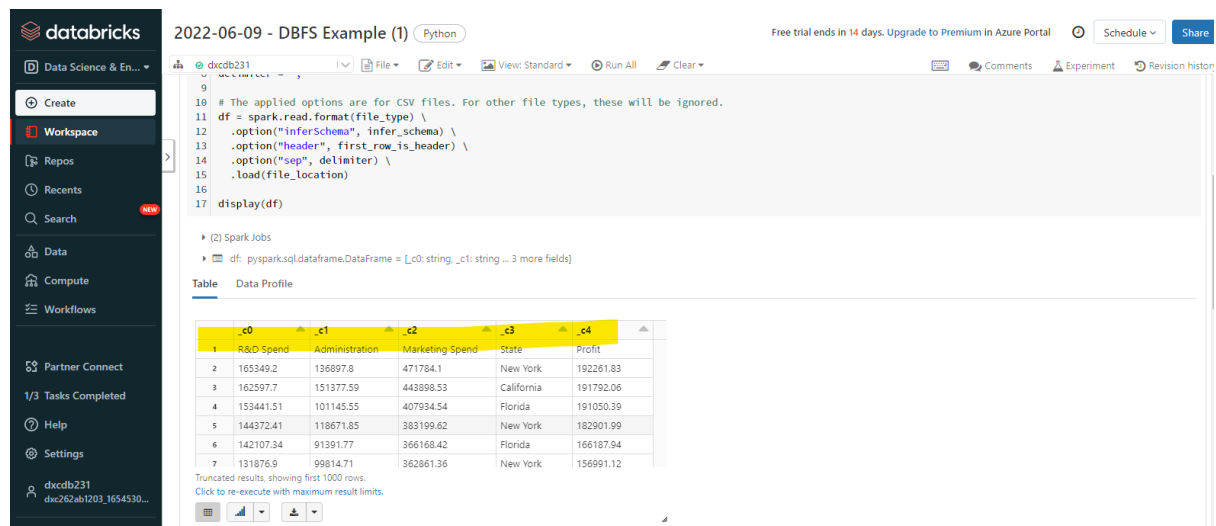
Step-3: After uploading the table click on create table in notebook.



The screenshot shows the Databricks notebook interface. The left sidebar contains navigation options like 'Workspace', 'Repos', 'Recents', 'Search', 'Data', 'Compute', 'Workflows', 'Partner Connect', 'Tasks Completed', 'Help', 'Settings', and a user profile. The main area is titled '2022-06-09 - DBFS Example (1)' and is in 'Python' mode. It shows an 'Overview' section with text about DBFS and a code cell with the following Python code:

```
1 # File location and type
2 file_location = "/FileStore/tables/1000_Companies-2.csv"
3 file_type = "csv"
4
5 # CSV options
6 infer_schema = "false"
7 first_row_is_header = "false"
8 delimiter = ","
9
10 # The applied options are for CSV files. For other file types, these will be ignored.
11 df = spark.read.format(file_type) \
12     .option("inferSchema", infer_schema) \
13     .option("header", first_row_is_header) \
14     .option("sep", delimiter) \
15     .load(file_location)
16
17 display(df)
```

Step-4: After running, run all commands, note that cluster is attached.



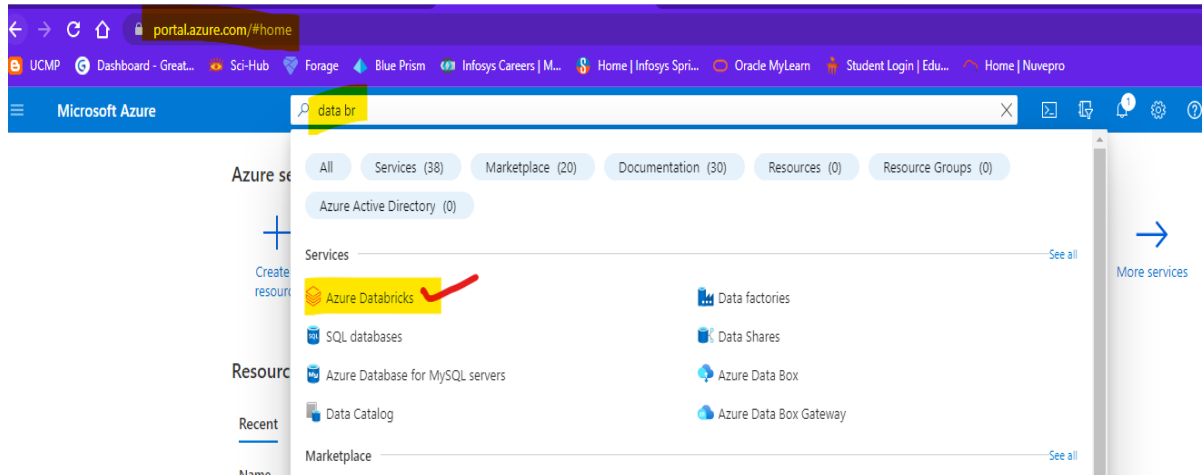
The screenshot shows the Databricks notebook interface after running the code. The code cell is now empty, and the output section shows the execution results. The output includes a message indicating that the cluster is attached and a data profile of the DataFrame. The data profile shows the following columns: '_c0', '_c1', '_c2', '_c3', and '_c4'. The data is truncated to the first 1000 rows.

	_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

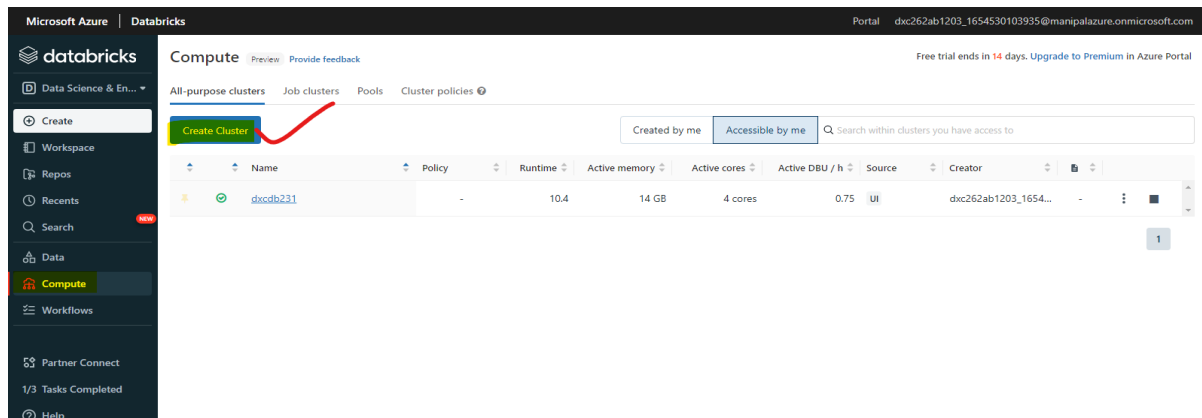
11. Explain the steps with screenshots how to create clusters in databricks?

Ans: To create a cluster in data bricks we have to follow the steps mentioned below.

Step-1: login into the azure portal and navigate to the data bricks which we already created before.



Step-2: After navigating to the data bricks portal and click on compute and click on create cluster.



Step-3: after that name the cluster and select requirements and click on create-to-create cluster.

