

# DXC AZURE ANALYTICS

## Assignment-9

Name: Mohana Likhitha Thotakura

ID: DXCAB1219

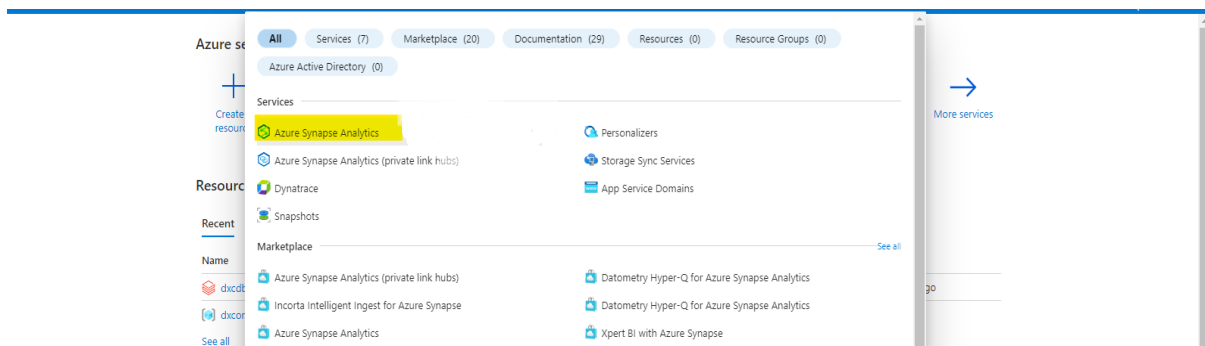
Date of submission: 09-06-2022

Batch: DXC-AZ-ANALYTICS

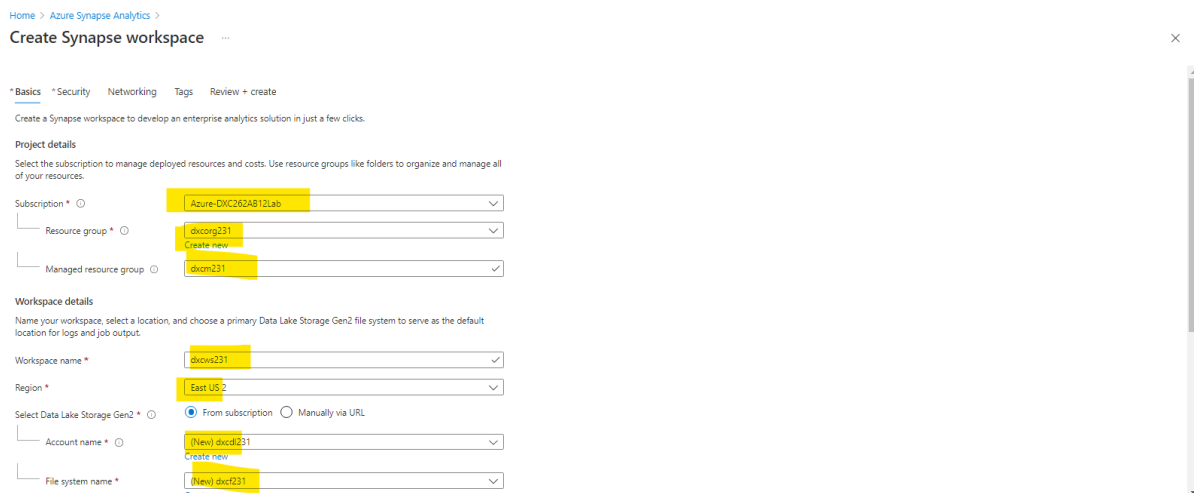
1. Explain the steps with screenshots how to create Azure Synapse analytics?

Ans) To create azure synapse analytics we have to follow the following steps

Step-1: login to the azure portal and search for Azure Synapse analytics



Step-2 : click on azure synapse analytics and click on create button



Step-3: next setup your sql administer and setup your passcode

\* Basics \* **Security** Networking Tags Review + create

Configure security options for your workspace.

**Authentication**

Choose the authentication method for access to workspace resources such as SQL pools. The authentication method can be changed later on. [Learn more](#)

Authentication method ☐ ☒ Use both local and Azure Active Directory (Azure AD) authentication ☐ Use only Azure Active Directory (Azure AD) authentication

SQL Server admin login \*

SQL Password

Confirm password

**System assigned managed identity permission**

Select to grant the workspace network access to the Data Lake Storage Gen2 account using the workspace system identity. [Learn more](#)

☐ Allow network access to Data Lake Storage Gen2 account. ☐

**Workspace encryption**

Step-4: after the validation is done click on create

Microsoft Azure Search resources, services, and docs (G+/)

Home > Azure Synapse Analytics >

## Create Synapse workspace ...

Validation succeeded

\* Basics \* Security Networking Tags **Review + create**

**Product Details**

Azure Synapse Analytics workspace by Microsoft Serverless SQL est. cost/TB --

[Terms of use](#) | [Privacy policy](#)

**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](#).

**Basics**

Subscription Azure-DXC262AB12Lab

Resource group dxcorg231

Region East US 2

(new) dxcws231

(new) <https://dxcrl231.dfs.core.windows.net>

**Create** < Previous Next > [Download a template for automation](#)

Step-5: The deployment is going on after deployment click on go to resource button

Home > Microsoft.Azure.SynapseAnalytics-20220609170109 | Overview

Deployment

Search (Ctrl+/) Delete Cancel Redeploy Refresh

Overview

We'd love your feedback! →

\*\*\* Deployment is in progress

Deployment name: Microsoft.Azure.SynapseAnalytics-20220609170... Start time: 6/9/2022, 5:07:21 PM  
Subscription: Azure-DXC262AB12Lab Correlation ID: cd6021f3-46a7-45bd-8cfd-4c3f5ad07e4  
Resource group: dxcorg231

Deployment details (Download)

Resource	Type	Status	Operation details
dxcws231	Microsoft.Synapse/workspaces	Created	<a href="#">Operation details</a>
dxcdl231/default/dxc231	Microsoft.Storage/storageAccounts/blobServices/...	Created	<a href="#">Operation details</a>
dxcdl231	Microsoft.Storage/storageAccounts	OK	<a href="#">Operation details</a>

Step-6: the deployment is completed and click on go to resources

Home > Microsoft.Azure.SynapseAnalytics-20220609170109 | Overview

Deployment

Search (Ctrl+/) Delete Cancel Redeploy Refresh

Overview

We'd love your feedback! →

✓ Your deployment is complete

Deployment name: Microsoft.Azure.SynapseAnalytics-20220609170... Start time: 6/9/2022, 5:07:21 PM  
Subscription: Azure-DXC262AB12Lab Correlation ID: cd6021f3-46a7-45bd-8cfd-4c3f5ad07e4  
Resource group: dxcorg231

Deployment details (Download)

Next steps

[Go to resource group](#)

**Cost Management**  
Get notified to stay within your budget and prevent unexpected charges on your bill.  
[Set up cost alerts >](#)

**Microsoft Defender for Cloud**  
Secure your apps and infrastructure  
[Go to Microsoft Defender for Cloud >](#)

**Free Microsoft tutorials**  
[Start learning today >](#)

Work with an expert

Step-7: after navigating to that click on synapse studio it will open in new tab

Home > dxcorg231 > dxcws231

Synapse workspace

Search (Ctrl+/) + New dedicated SQL pool + New Apache Spark pool + New Data Explorer pool (preview) Refresh Reset SQL admin password Delete

Overview

Activity log

Access control (IAM)

Tags

Diagnose and solve problems

Settings

Azure Active Directory

Properties

Locks

Analytics pools

SQL pools

Apache Spark pools

Data Explorer pools (preview)

Security

Encryption

Networking

Identity

Essentials

Resource group (move) : dxcorg231

Status : Succeeded

Location : East US 2

Subscription (move) : Azure-DXC262AB12Lab

Subscription ID : 48ef8049-33b2-4e37-bec3-5f8394096bd2

Managed virtual network : No

Managed identity object ... : 4c018d15-e517-4e7b-acc2-5a3893eb4e4a

Workspace web URL : <https://web.azure.synapse.net?workspace=%2fsubscriptions%2f48ef8049-33b2-4e37-bec3-5f8394096bd2>

Tags (edit) : [Click here to add tags](#)

Networking : [Show firewall settings](#)

Primary ADLS Gen2 account : <https://dxcdl231.dfs.core.windows.net>

Primary ADLS Gen2 file share : dxc231

SQL admin username : sqladminuser

SQL Active Directory admin : [dxc262ab1203\\_1654530103935@manjalazure.onmicrosoft.com](mailto:dxc262ab1203_1654530103935@manjalazure.onmicrosoft.com)

Dedicated SQL endpoint : dxcws231.sql.azuresynapse.net

Serverless SQL endpoint : dxcws231-ondemand.sql.azuresynapse.net

Development endpoint : <https://dxcws231.dev.azure.synapse.net>

Getting started

**Open Synapse Studio**  
Start building your analytics solutions.  
[Open](#)

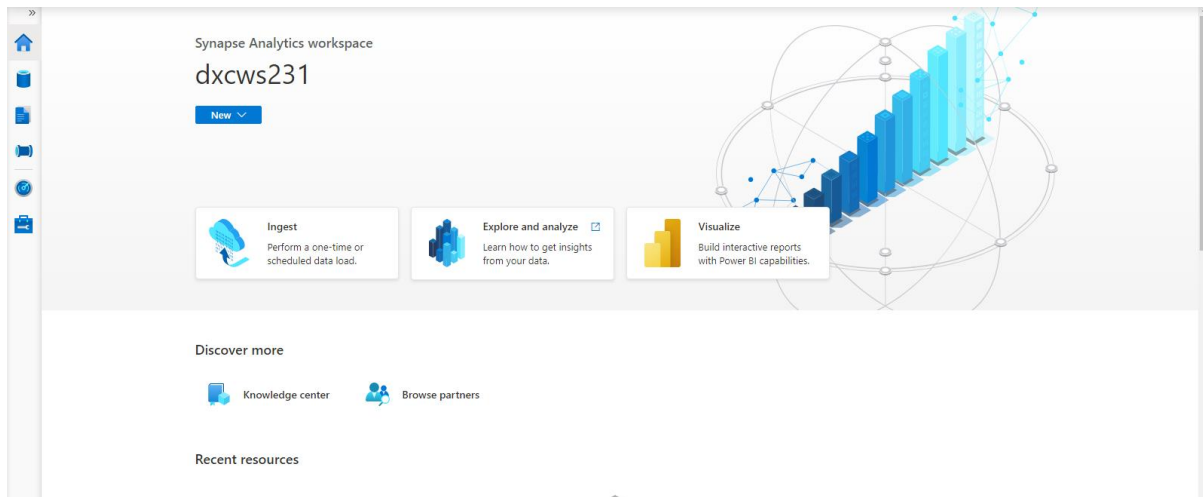
**Read documentation**  
Learn how to be productive quickly. Explore concepts, tutorials, and samples.  
[Learn more >](#)

Analytics pools

Search to filter items...

Name	Type	Size
------	------	------

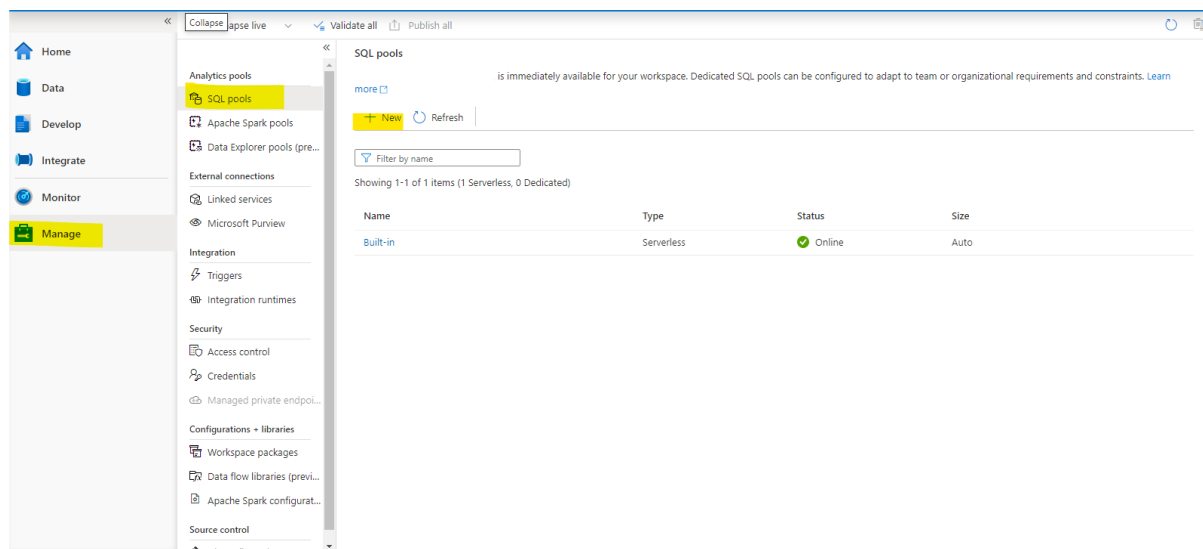
Step-8: you will navigate to this page, thus azure synapse analytics creation is done



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

Ans) in Azure synapse analytics the Sql pool can be created in the below mentioned way

Step-1: after opening the home page navigate to manage and choose sql pool



Step-2: after clicking on the new we need to finish the all details review and create

## New dedicated SQL pool

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



< Previous

[Download template for automation](#)

Step 5: the pool sql creation is done

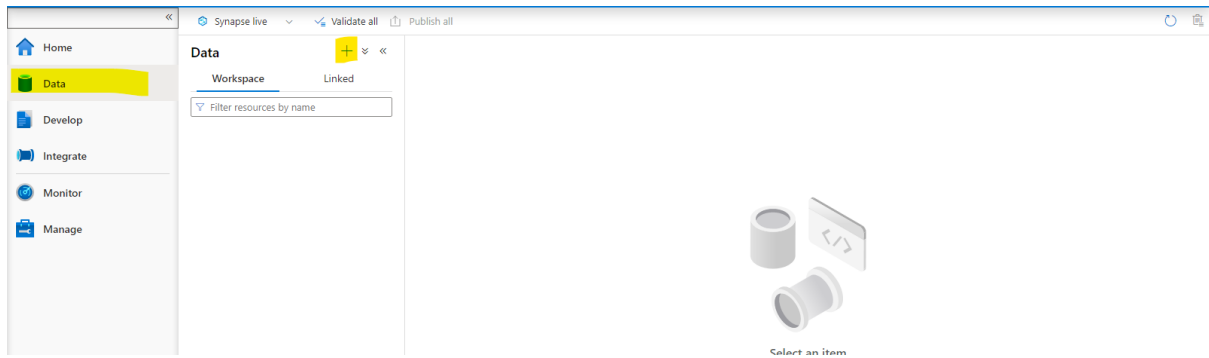
The screenshot shows the Azure Synapse Analytics interface. On the left is a navigation pane with options like Home, Data, Develop, Integrate, Monitor, and Manage. The 'SQL pools' section is highlighted. The main area displays a table of SQL pools. The table has columns for Name, Type, Status, and Size. Two pools are listed: 'Built-in' (Serverless, Online, Auto) and 'dxcsql231' (Dedicated, Deploying, DW100c). The 'dxcsql231' pool is highlighted with a yellow background.

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

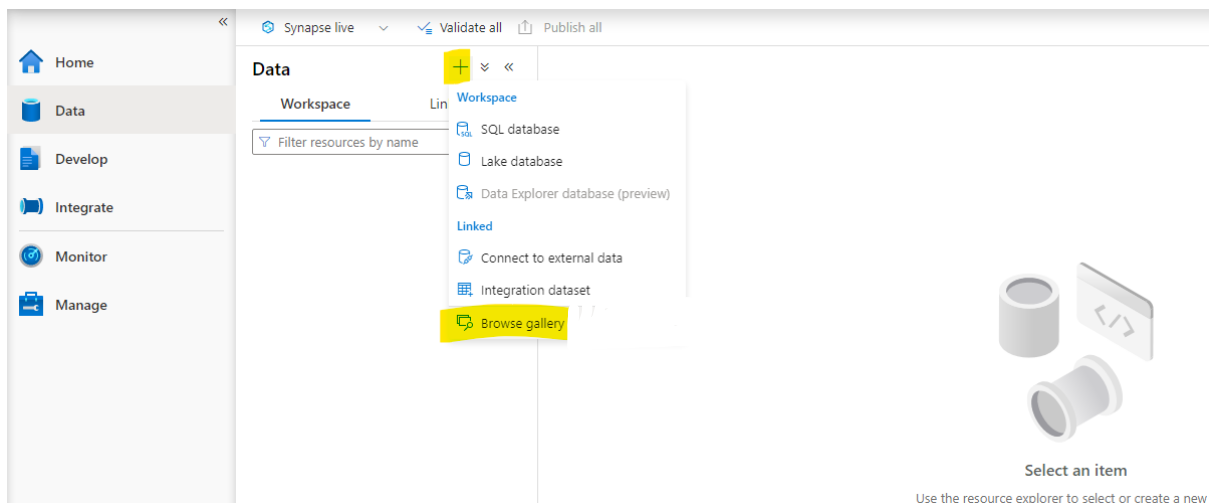
3. Explain the steps with screenshots how to import COVID19 dataset in AzureSynapse analytics and run sample 500 rows & display the output?

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

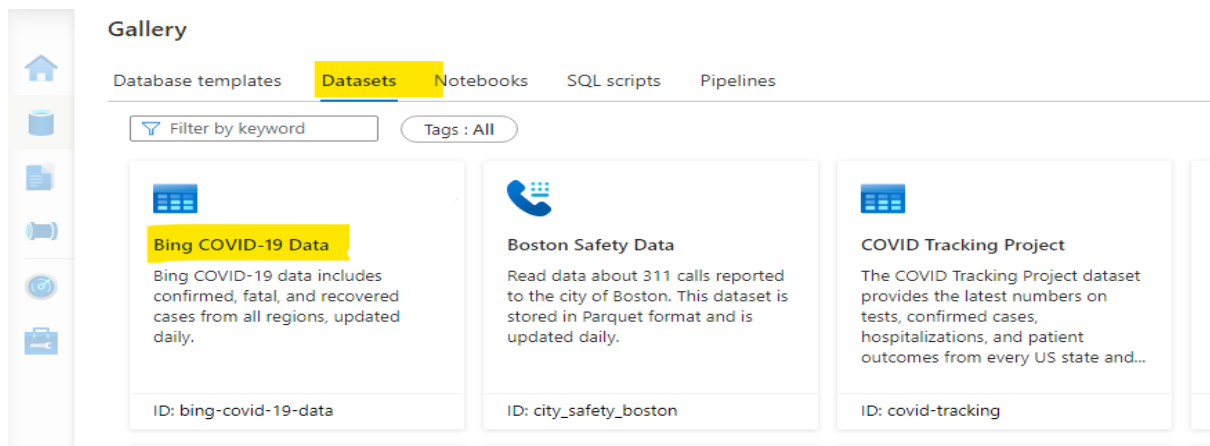
Step-1: navigate to the home page and click on data



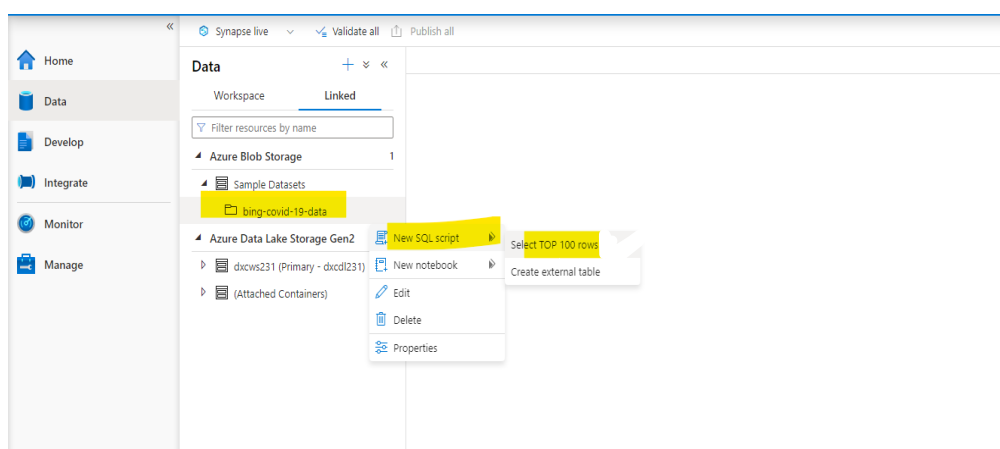
Step-2: click on the + button and choose browse gallery



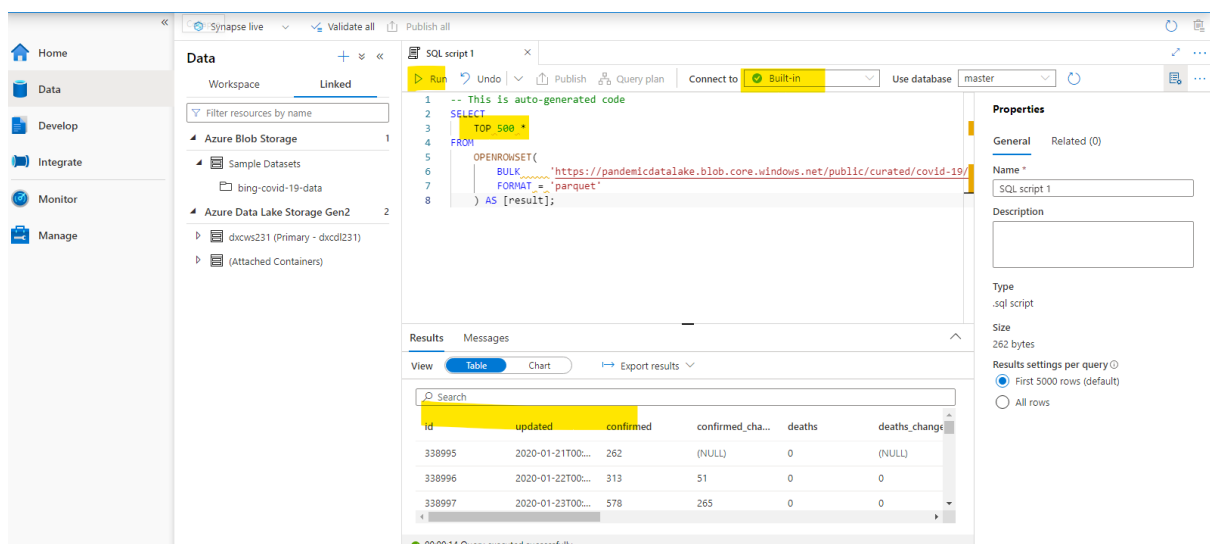
Step-3: select the covid 19 data from the menu from the data sets and add to the data sets after previewing the data



Step-4: after that we will have to following steps



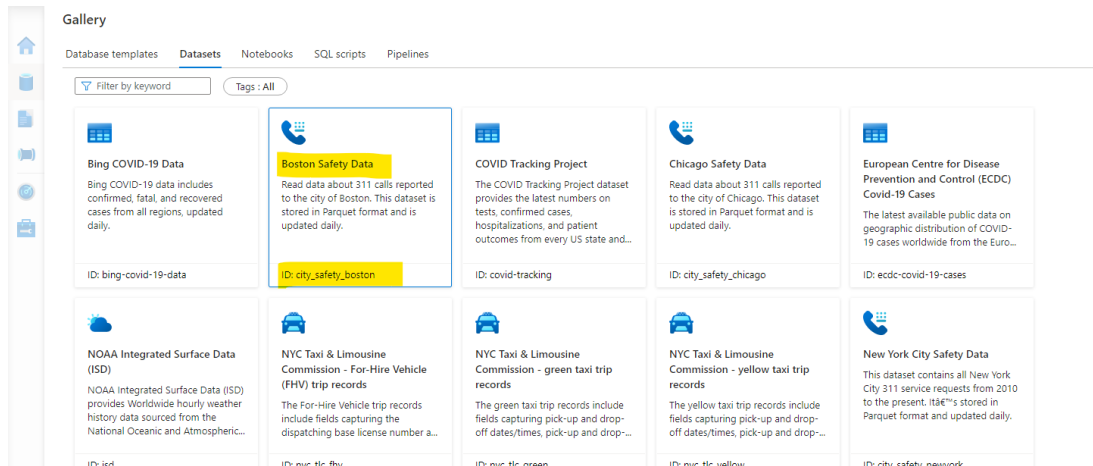
Step-5: after that run the command and the output data will be displayed.



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 first run.

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

```
1 from azureml.op
2
3 data = BostonSe
4 df = data.to_sp
5 # Display 10 r
6 display(df.limi
```

[1] ✓ 40 sec - Command executed in 39 sec 664 ms by dxc262ab1203\_1654530103935 on 6:00:54 PM, 6/09/22

Job execution Succeeded Spark 2 executors 8 cores

View Table Chart Export results

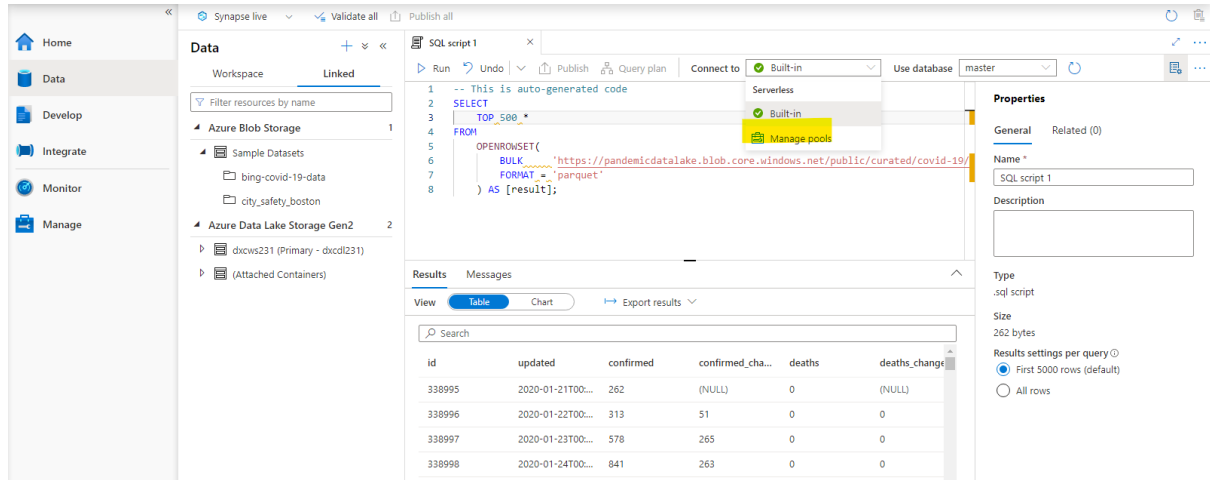
dataType	dataSubtype	dateTime	category	subc
Safety	311_All	2015-02-06 08:04:46	Sanitation	Miss
Safety	311_All	2014-06-13 10:59:03	Signs & Signals	Traffi
Safety	311_All	2021-03-09 12:08:21	Street Cleaning	Requ
Safety	311_All	2020-10-16 08:41:38	Needle Program	Neec
Safety	311_All	2019-04-15 14:05:21	Street Cleaning	Requ
Safety	311_All	2022-03-01 08:17:32	Code Enforcement	Unsh
Safety	311_All	2016-06-06 17:03:20	Sanitation	Miss
Safety	311_All	2021-08-09 14:03:00	Building	Unsa



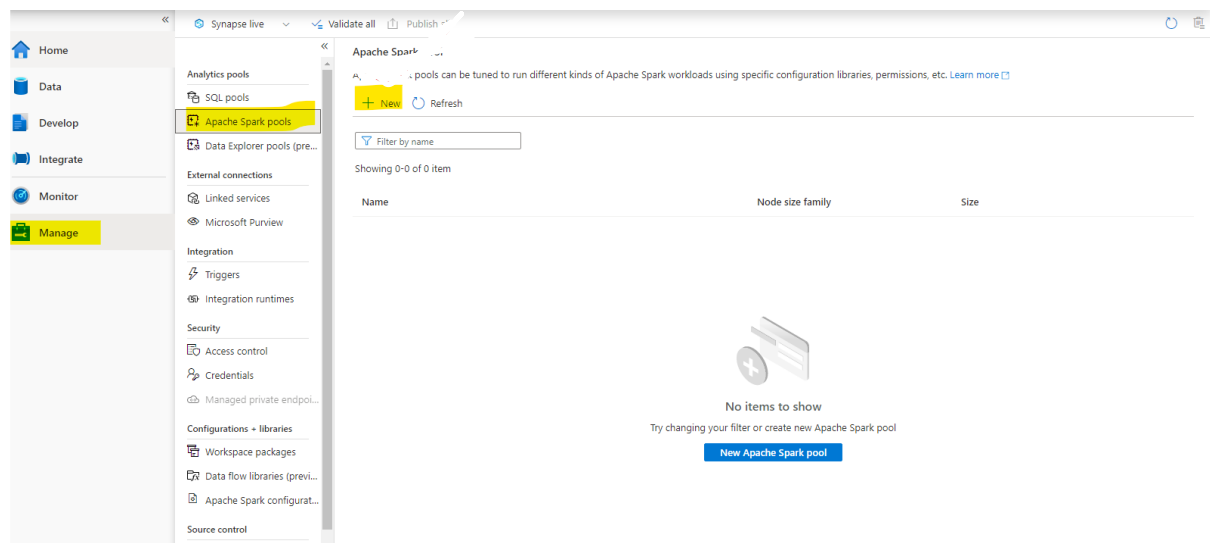
5. Explain the steps with screenshots how to create Spark pool in AzureSynapse analytics?

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

## New Apache Spark pool

Basics \*   Additional settings \*   Tags   Review + create

Create an Synapse Analytics Apache Spark pool with your preferred configurations. Complete the Basics tab then go to Review + Create to provision with smart defaults, or visit each tab to customize.

### Apache Spark pool details

Name your Apache Spark pool and choose its initial settings.

Apache Spark pool name \*

Node size family \*

Node size \*

Autoscale \* ☐ Enabled ☒ Disabled

Number of nodes \*  3


Estimated price ⓘ   
 Est. cost per hour   
 ❌ Failed to fetch billing info


Dynamically allocate executors \* ⓘ ☐ Enabled ☒ Disabled

[Review + create](#) [Next: Additional settings >](#) [Cancel](#)


Step-4 : it takes few minutes to deploy

## Notifications

 Dismiss all

 Deploying

Deploying dxcap231 (Apache Spark pool) to the workspace. This can take 1-2 minutes.



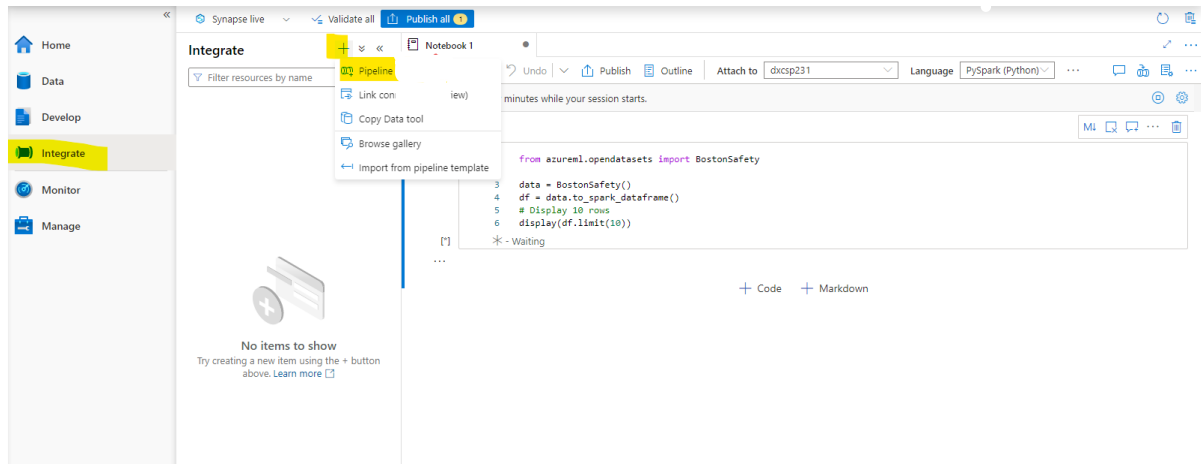
a few seconds ago

Step-5: and done

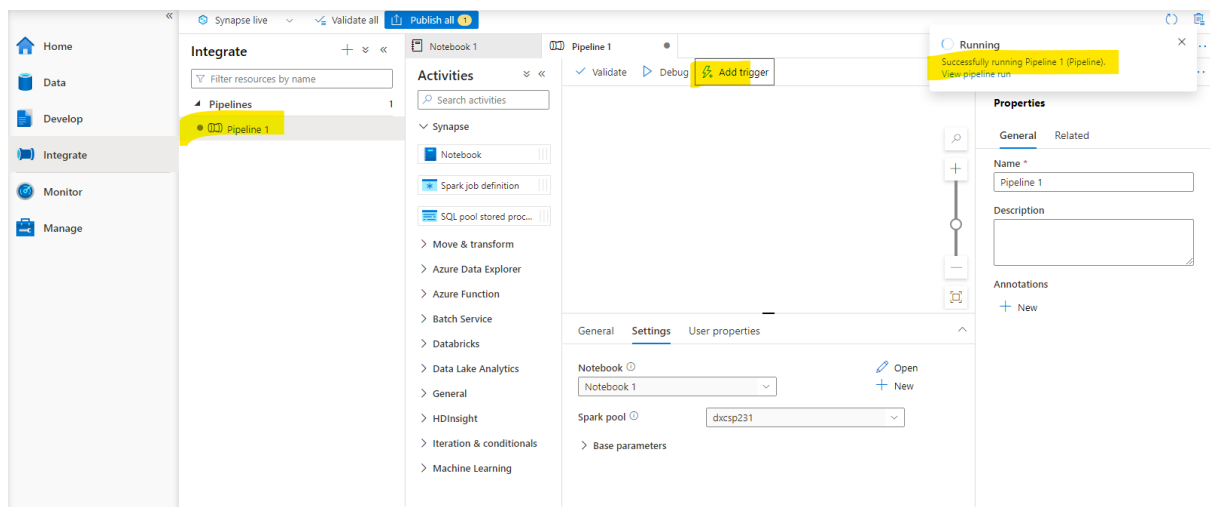
6. Explain the steps with screenshots how to create pipeline in AzureSynapse analytics?

Ans) In azure synapse analytics we have to follow to these steps to create a pipe line

Step-1: click on integrate,”+” icon and after that click on pipe line



Step-2: The pipe line has created successfully and triggered



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

Ans) Automation of pipelines is noting but scheduling the pipelines to trigger at time interval given

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

**New trigger**

Name \*  
Trigger 1

Description

Type \*  
Schedule

Start date \*  
6/9/22 18:47:51

Time zone \*  
Chennai, Kolkata, Mumbai, New Delhi (UTC+5:30)

Recurrence \*  
Every 1 Minute(s)

☐ Specify an end date

Annotations  
+ New

Start trigger  
☒ Start trigger on creation

OK Cancel

Step-2 : The pipeline will Automatically triggers 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

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

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

Copy filters Export to CSV

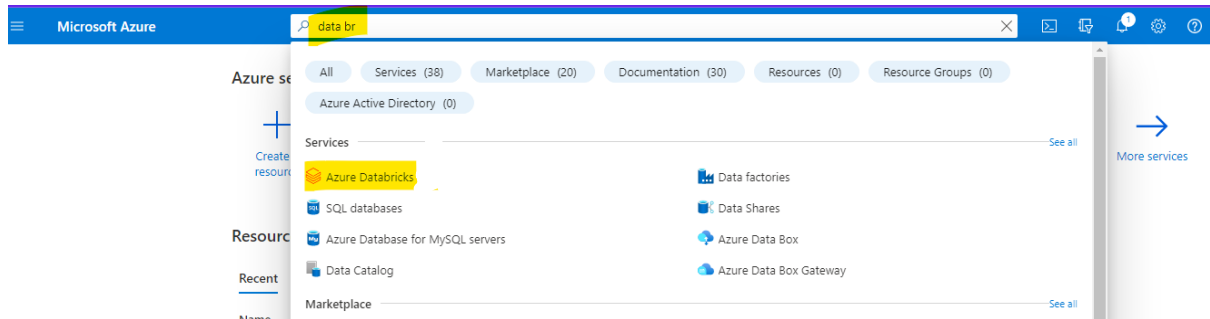
Showing 1 - 10 items Last refreshed 0 minutes ago

<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 create to 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)

Step-3: after checking all these reviews and create and wait for deployment after deployment we will get like this

# Notifications

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

[Dismiss](#)

✓ **Deployment succeeded**

Deployment [\[redacted\]](#) to resource group 'dxcorg231' was successful.

[Go to resource](#)

[Pin to dashboard](#)

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

Search (Ctrl+/) Delete

Overview

Activity log

Access control (IAM)

Tags

Settings

Virtual Network Peerings

Encryption

Properties

Locks

Automation

Tasks (preview)

Export template

Support + troubleshooting

New Support Request

Essentials

Status : Active

Resource group : [dxcorg231](#)

Location : East US

Subscription : [Azure-DXC262A8121a0](#)

Subscription ID : 48ef8049-33b2-4e37-bec3-5f8394096bd2

Tags ([edit](#)) : [Click here to add tags](#)

Managed Resource Group : [databricks-rg-dxcdb231-gnqon7ipb2zic](#)

URL : [https://adb-7548927456012596.16.azure.databricks.net](#)

Pricing Tier : Trial (Premium - 14-Days Free DBUs)

[Launch Workspace](#)

[Upgrade to Premium](#)

Step-5: after click on launch workspace, you are in data bricks

**databricks**

Data Science & Engineering

Free trial ends in 14 days. [Upgrade to Premium in Azure Portal](#)

[Create](#)

[Workspace](#)

[Repos](#)

[Recents](#)

[Search](#)

[Data](#)

[Compute](#)

[Workflows](#)

**Notebook**

Create a new notebook for querying, data processing, and machine learning.

[Create a notebook](#)

**Guide: Quickstart tutorial**

Spin up a cluster, run queries on preloaded data, and display results in 5 minutes.

[Start tutorial](#)

**Data import**

Quickly import data, preview its schema, create a table, and query it in a notebook.

[Browse files](#)

**Partner Connect**

[Fivetran® dbt](#)

[Tableau. Power BI](#)

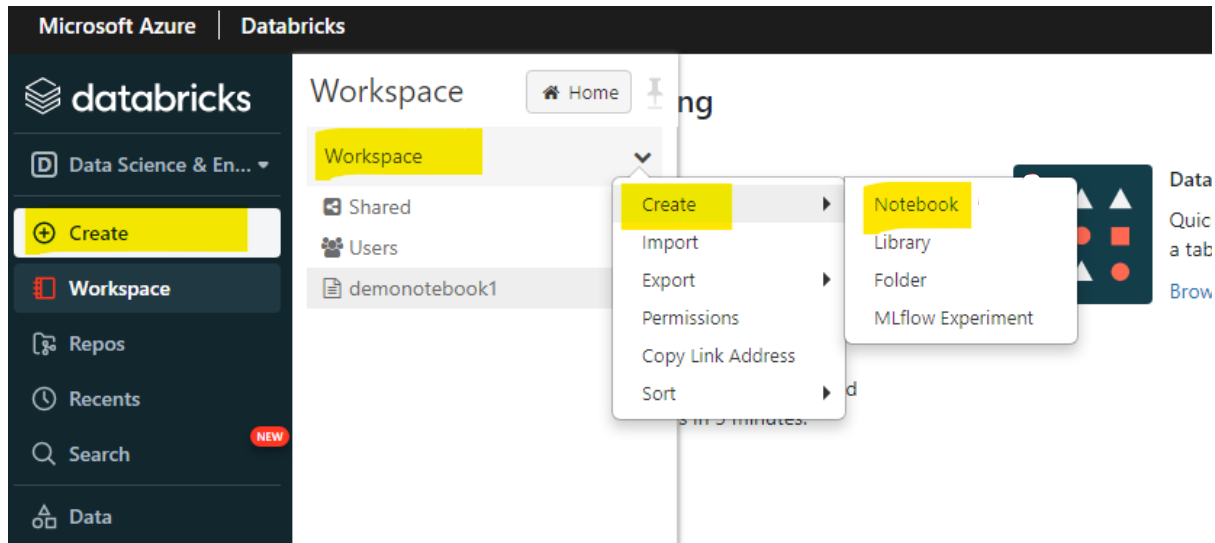
[View all partners](#)

**Recents**

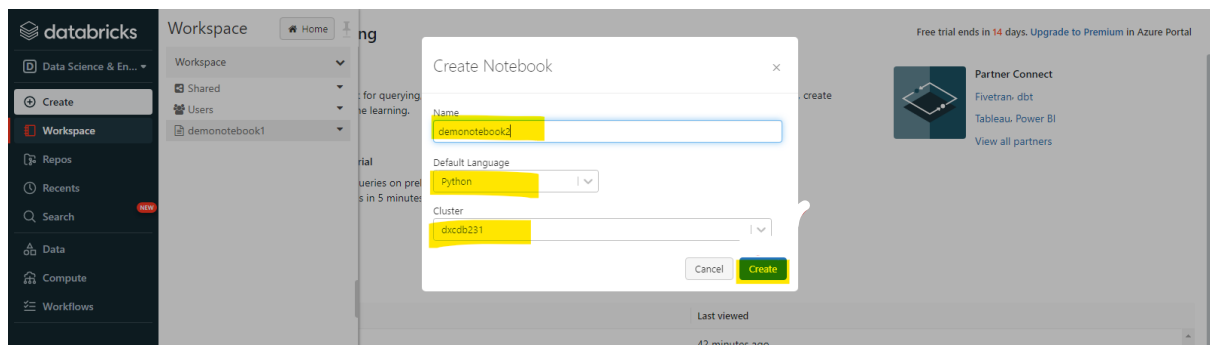
9. Explain the steps with screenshots how to create notebooks in Databricks?

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

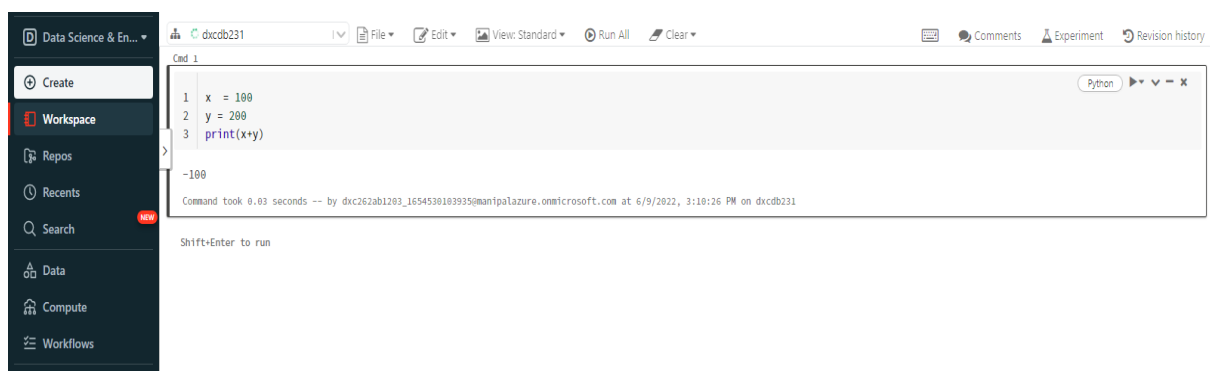
Step-1 : open azure portal and navigate to azure data bricks after that follow the attached screen shots



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



Step-3: The notebook is created in data bricks

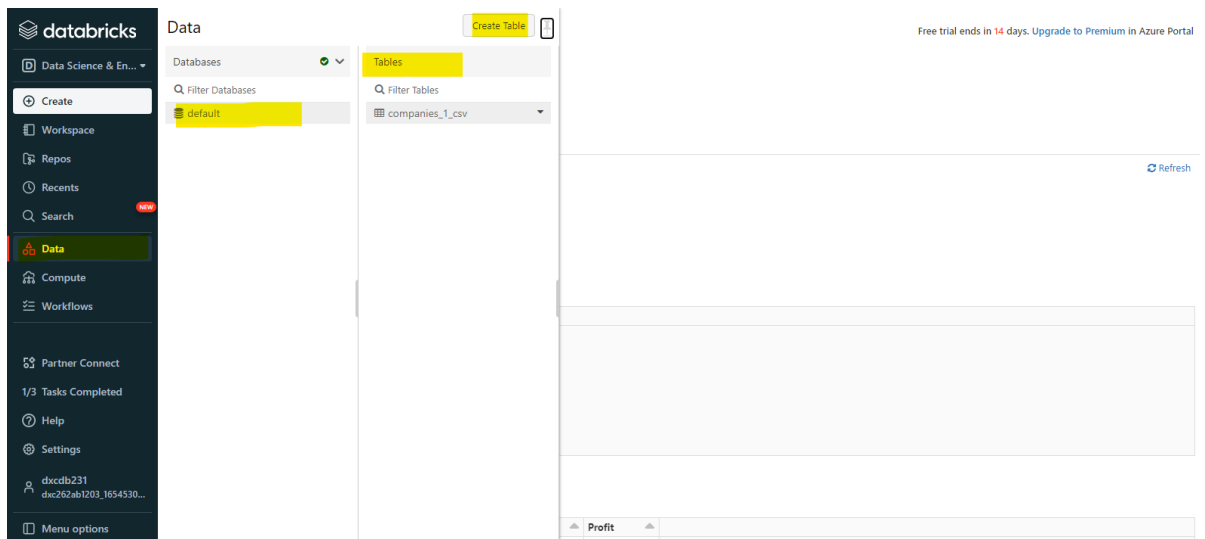




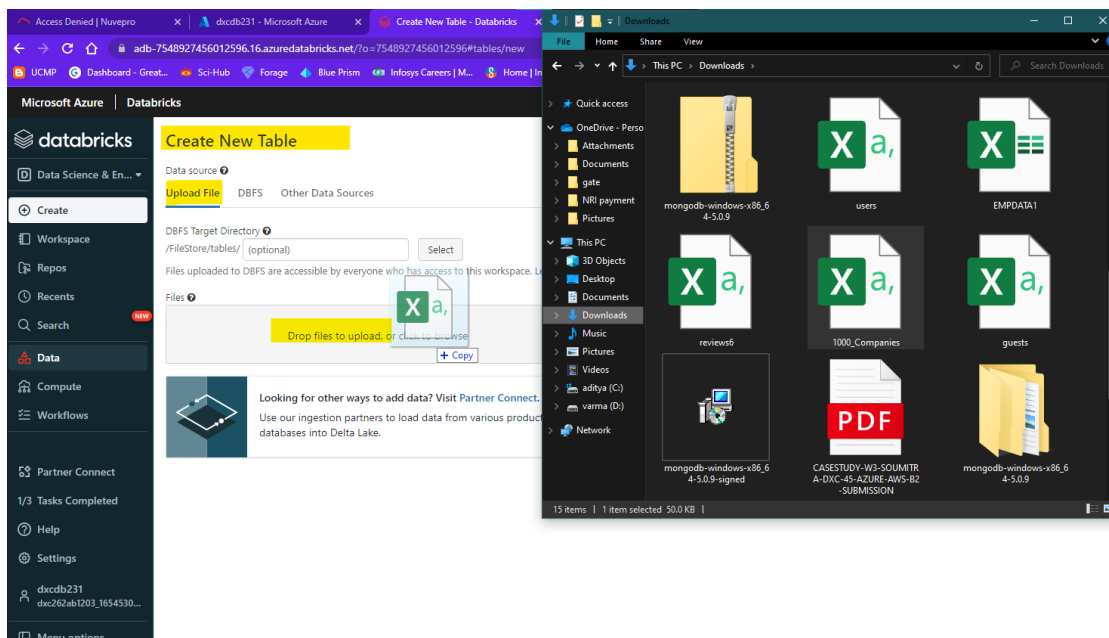
10. Explain the steps with screenshots how to insert data into databricks notebook & display the result?

Ans) in order to insert data into databricks note book we can insert data as mentioned below

Step-1: open the data bricks work space and click on data and it will navigates to a tab and click on create table



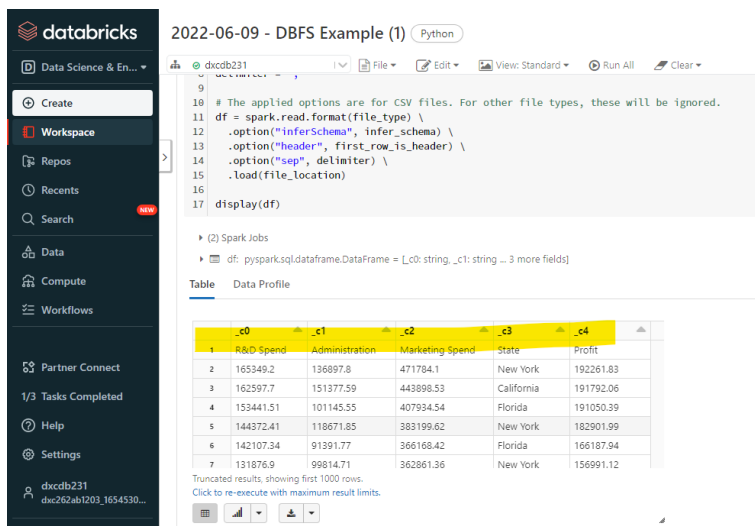
Step-2 : after clicking on create table. We can drag and drop the files which we need to insert int the note book refer attachment for better understanding.



Step-3: after uploading the table click on create table in note book



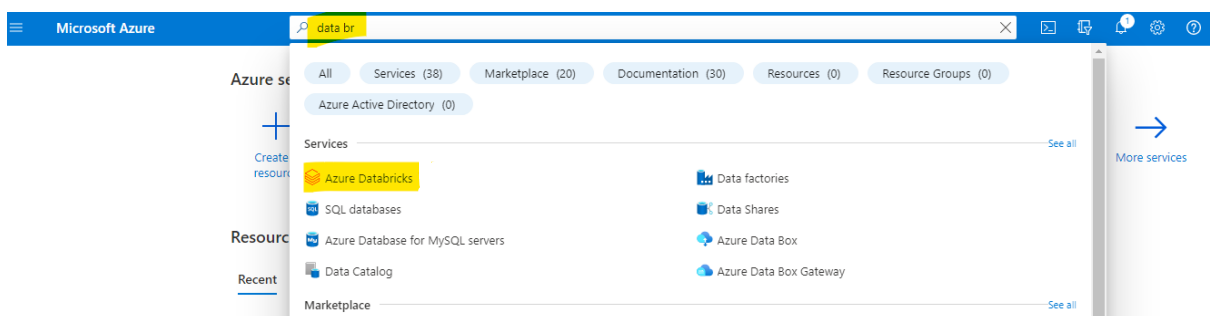
Step-4: after running the run all command. note that cluster is attached



11. Explain the steps with screenshots how to create cluster 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

**databricks** Compute [Preview](#) [Provide feedback](#) Free trial ends in 14 days. [Upgrade to Premium in Azure Portal](#)

All-purpose clusters Job clusters Pools Cluster policies ⓘ

[Create Cluster](#) Created by me Accessible by me 🔍 Search within clusters you have access to

Name	Policy	Runtime	Active memory	Active cores	Active DBU / h	Source	Creator	
dxcdb231	-	10.4	14 GB	4 cores	0.75	UI	dxc262ab1203_1654...	-

1

Step-4: after that name the cluster and select requirements and click on create-to-create cluster.

**databricks** Clusters / New Compute Free trial ends in 14 days. [Upgrade to Premium in Azure Portal](#)

New Cluster [Cancel](#) [Create Cluster](#) DBU / hour: 0.75 ⓘ 3 Workers: 0 GB Memory, 3 Cores 1 Driver: 14 GB Memory, 4 Cores

Cluster name  UI | JSON

Cluster mode ⓘ

Databricks runtime version ⓘ

50% promotional discount applied to Photon during preview ⓘ

☐ Use Photon Acceleration ⓘ [Preview](#)

Autopilot options ☒ Terminate after  minutes of inactivity ⓘ

Node type ⓘ  14 GB Memory, 4 Cores ⓘ

DBU / hour: 0.75 ⓘ [Standard\\_DS3\\_v2](#)

Advanced options