

NAME: SHAIK ABDUL KHADAR JILANI

ROLLNO: DXC262AB12038

BATCH: DXC-262-Analytics-B12-Azure

SUBMISSION: 9-6-2022

COMPANY: DXC TECHNOLOGY

DAY-9

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

First we need to go to Azure and search for synapse analytics select it and + create

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

Home > Azure Synapse Analytics >

Create Synapse workspace

Name your workspace, select a location, and choose a primary Data Lake Storage Gen2 file system to serve as the default location for logs and job output.

Workspace name *

Region *

Select Data Lake Storage Gen2 * ☒ From subscription ☐ Manually via URL

Account name * [Create new](#)

File system name * [Create new](#)

Additional configuration is required. After you create your workspace, perform these tasks:

- Assign other users to the **Contributor** role on workspace.

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

<https://portal.azure.com/#>

Give the workspace name dxcsynapse23, storage account name-datastore235 and filesystem name as filesystem235

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](#)

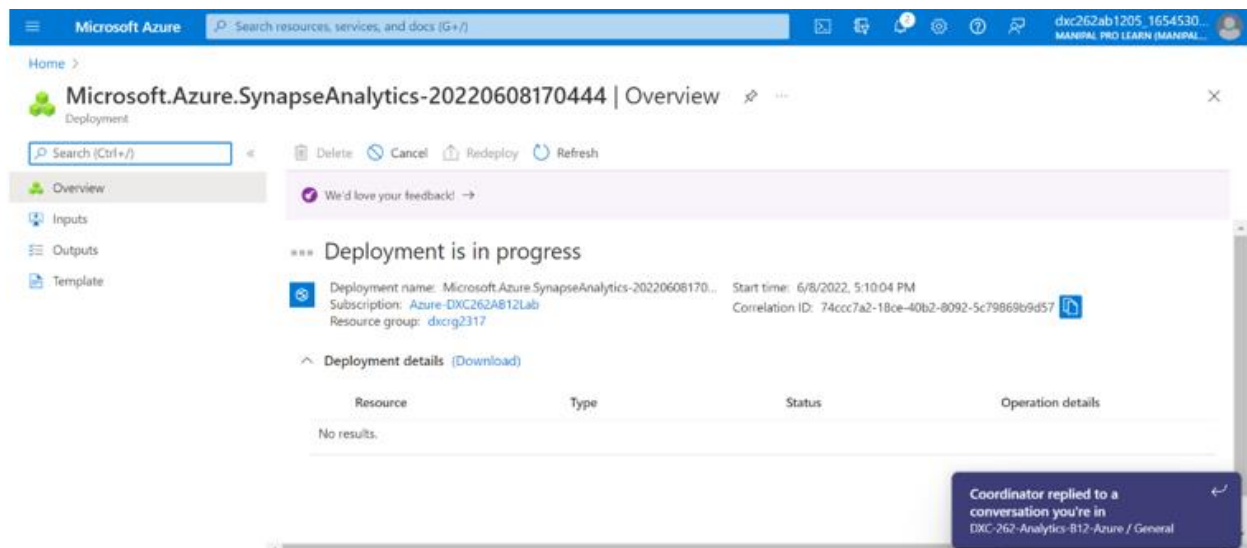
Terms

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

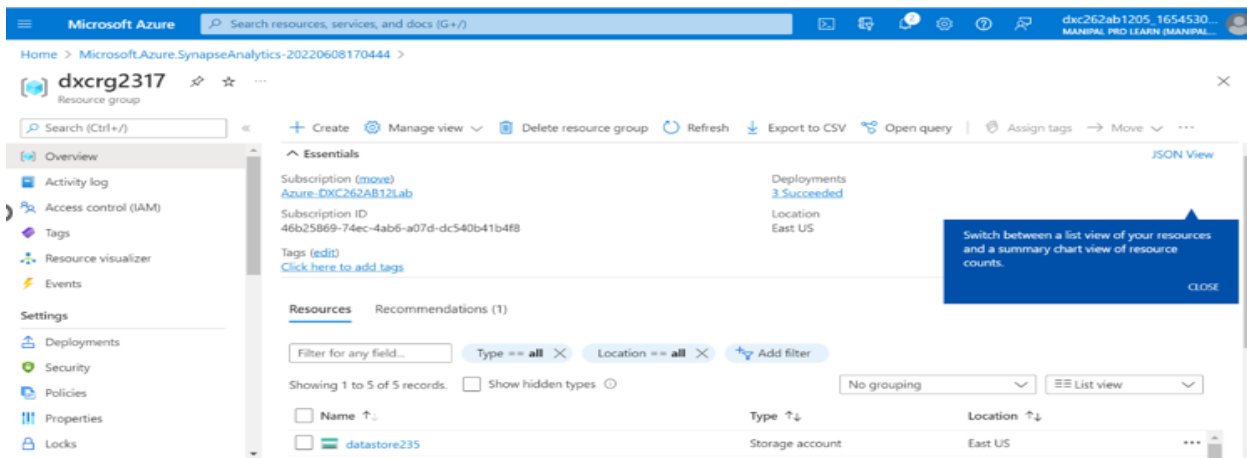
[Create](#) [< Previous](#) [Next >](#) [Download a template for automation](#)

<https://portal.azure.com/#>

All are default and create after validation



After that we get Deployment is complete and go to resource creating a synapse analytics is done



2) Explain the steps with screenshots how to SQL Pool in Azure Synapse analytics?

After creating the Synapse analytics go to resource page

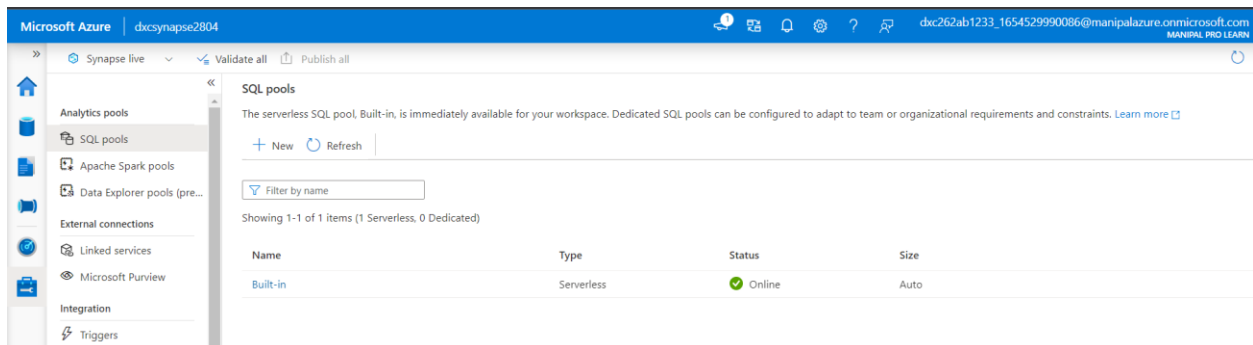
Open dxcsynapse2804 in that Getting started

Open Synapse Studio

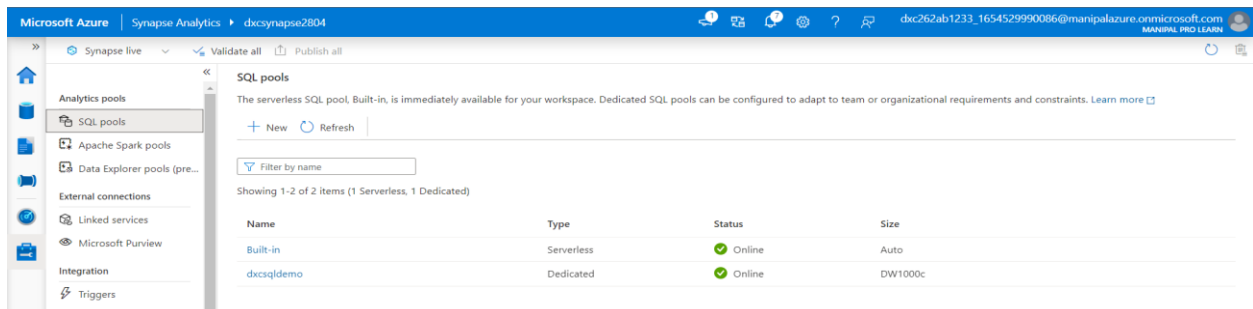
In this open the Data below home symbol



Go to manage tab and go to SQL Pool and + New



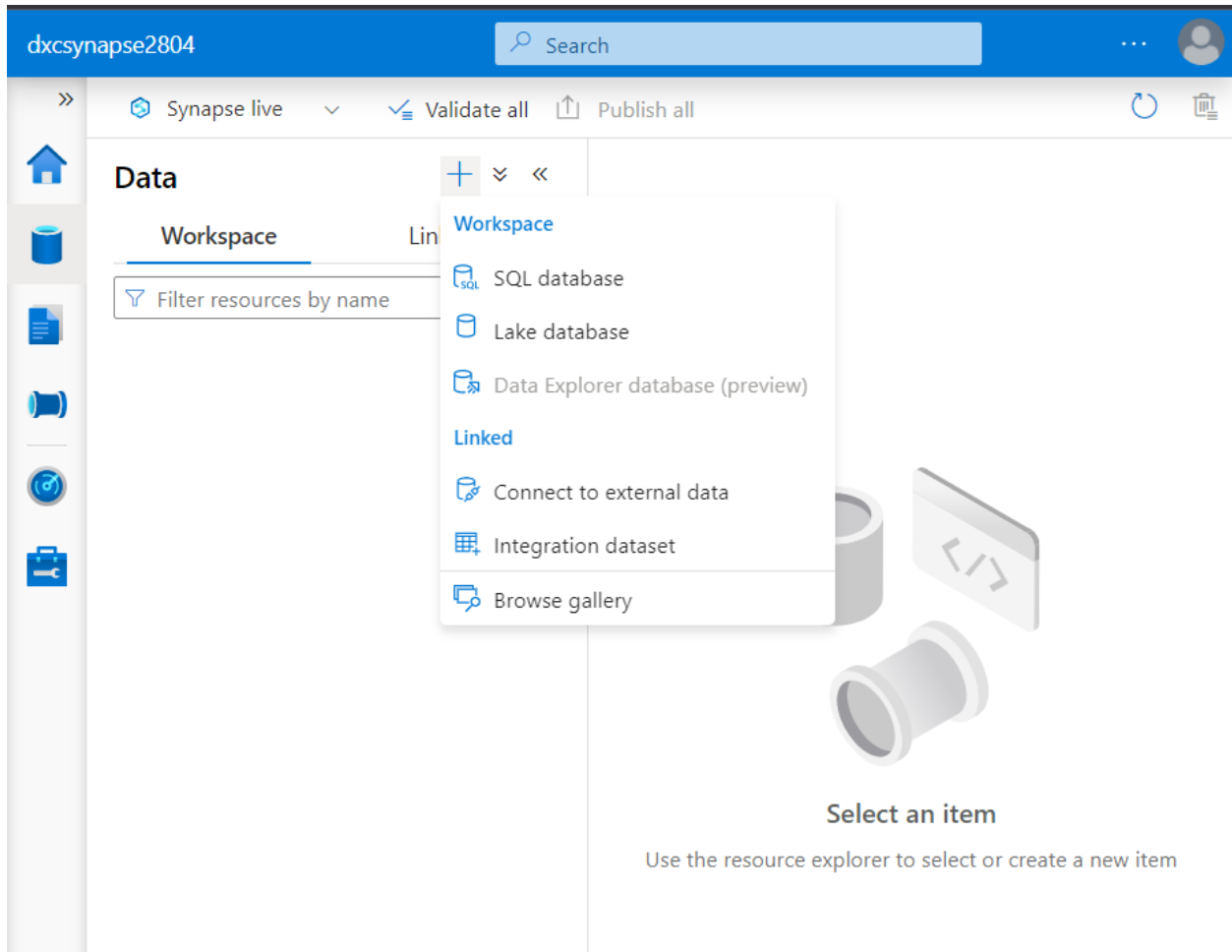
Give the name of the sql pool and create



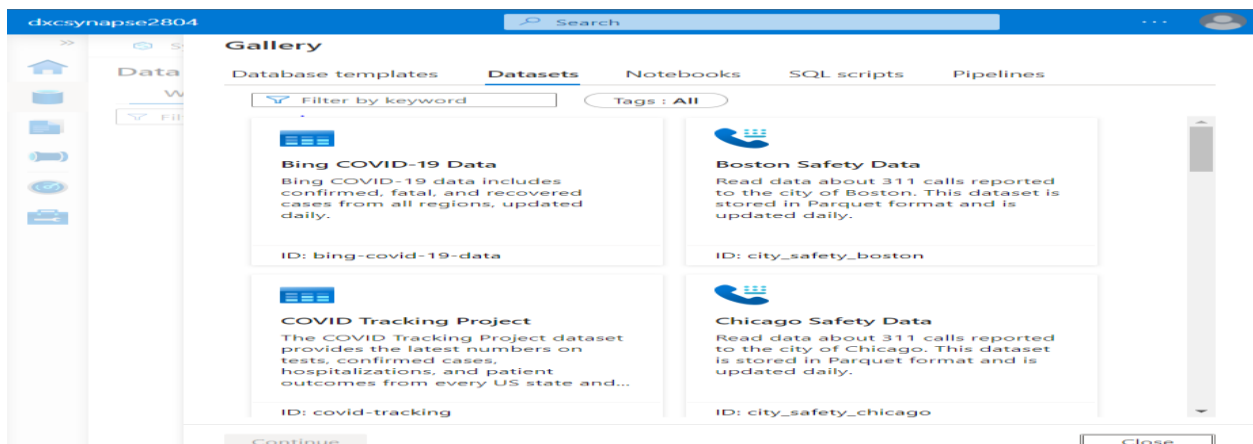
3) Explain the steps with screenshots how to import COVID19 dataset in AzureSynapse analytics and run sample 500 rows & display the output?

In synapse studio

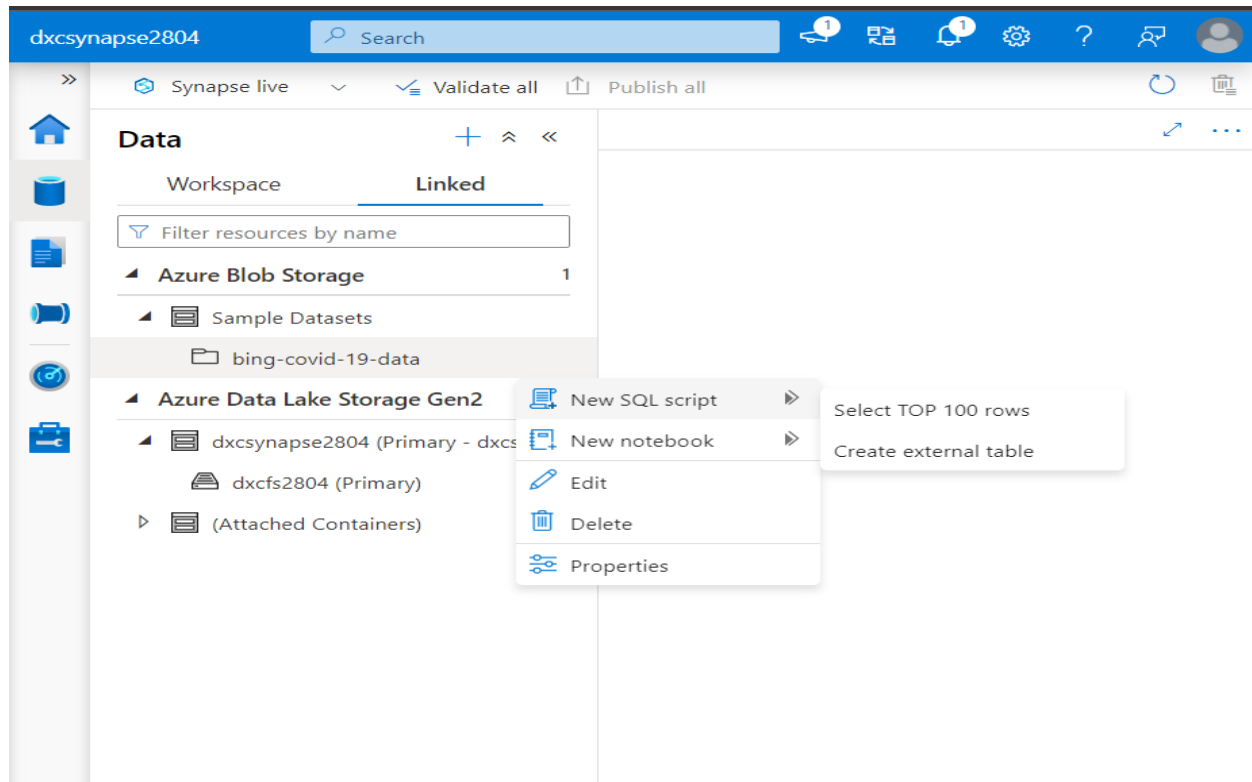
In this Data Tab click on + sign and select Browse gallery



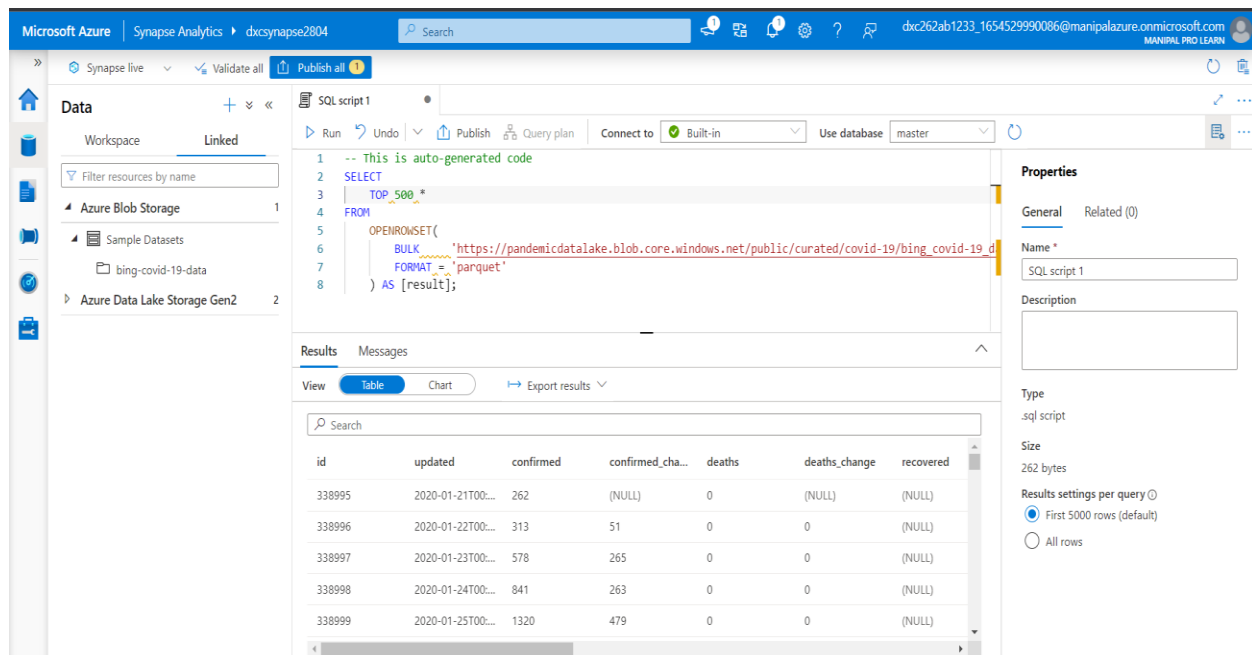
In that Datasets-> Select Bing COVID-19 Data and Continue



After that it will ask to add dataset we click add dataset in that dataset go to bing-covid-19-data
New SQL script-> Select Top 100 rows



Here we change the Run the SQL Script which shows the table of top 500 rows in the table



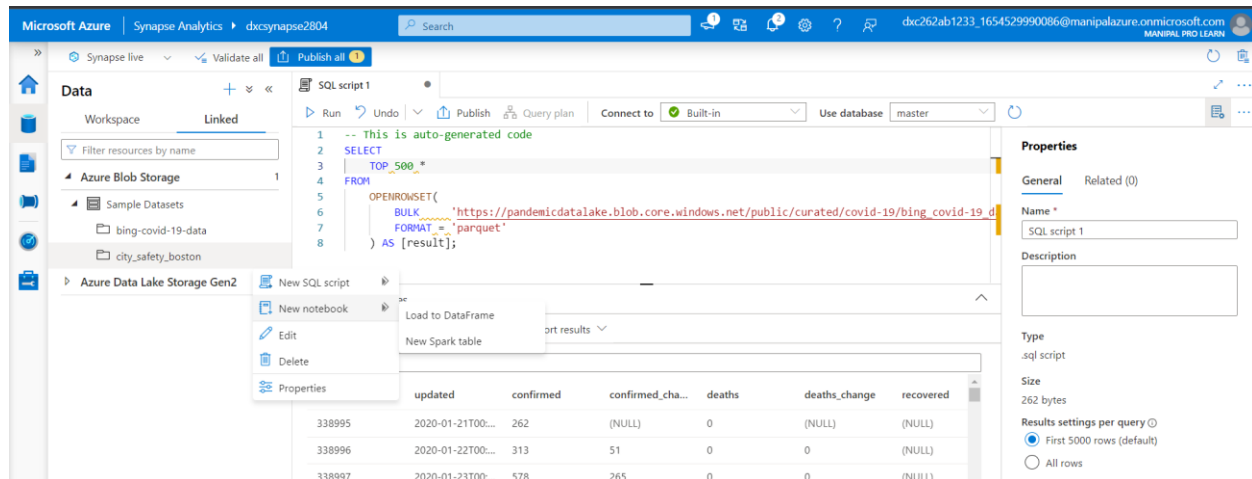
4) Explain the steps with screenshots how to input Boston Safety datasets into Azure Synapse analytics using Notebooks ?

In synapse studio

In this Data Tab click on + sign and select Browse gallery

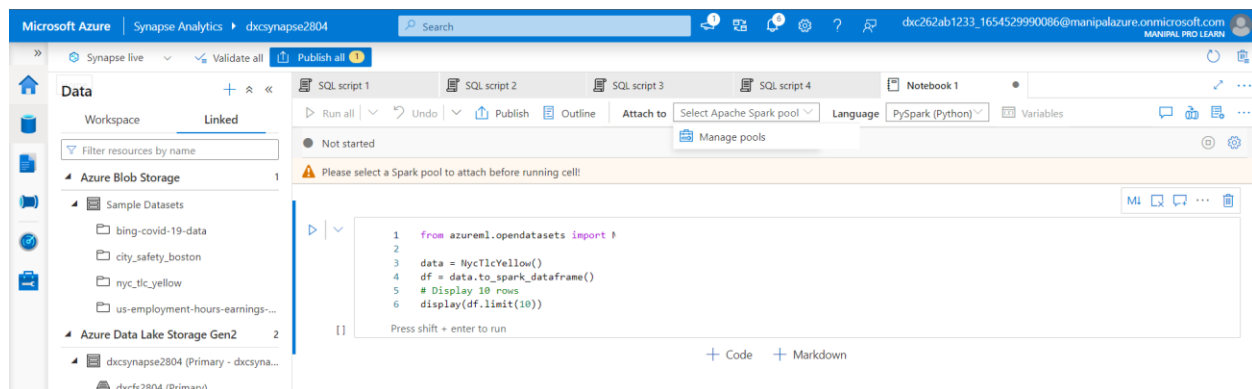
In that Datasets-> Select Boston Safety dataset Data and Continue

Go to city_safety_boston-> New NoteBook-> Load to Data Frame

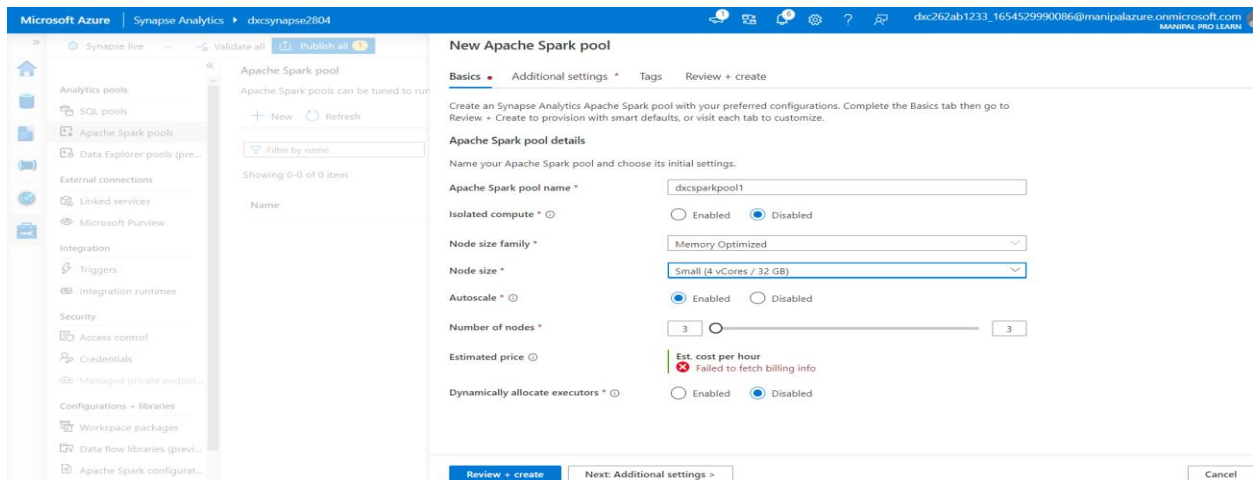


To run the data frame 1st we need to create Apache spark

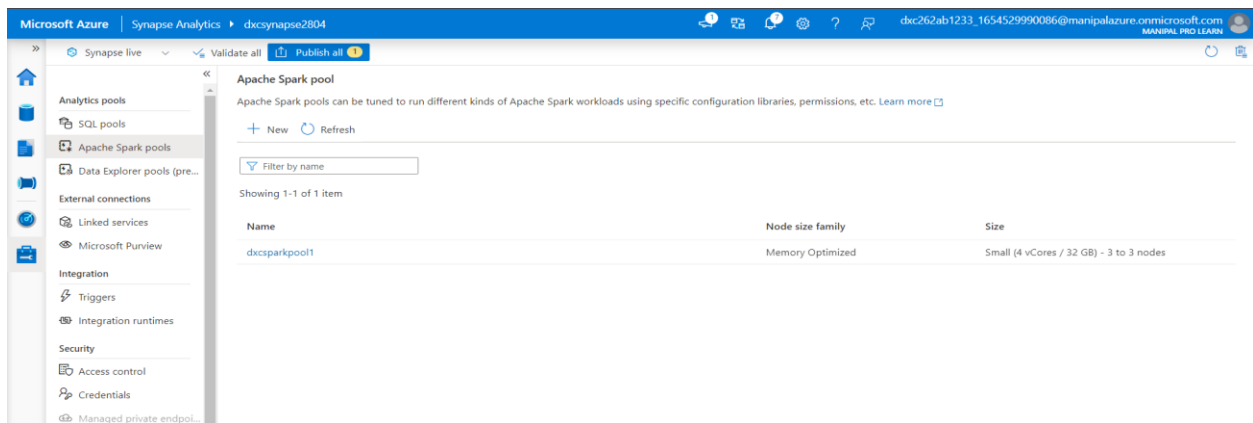
Go to Manage pools and go to sparkpool



Add new sparkpool name and node size small and no od nodes max 3 min 3 and review and create

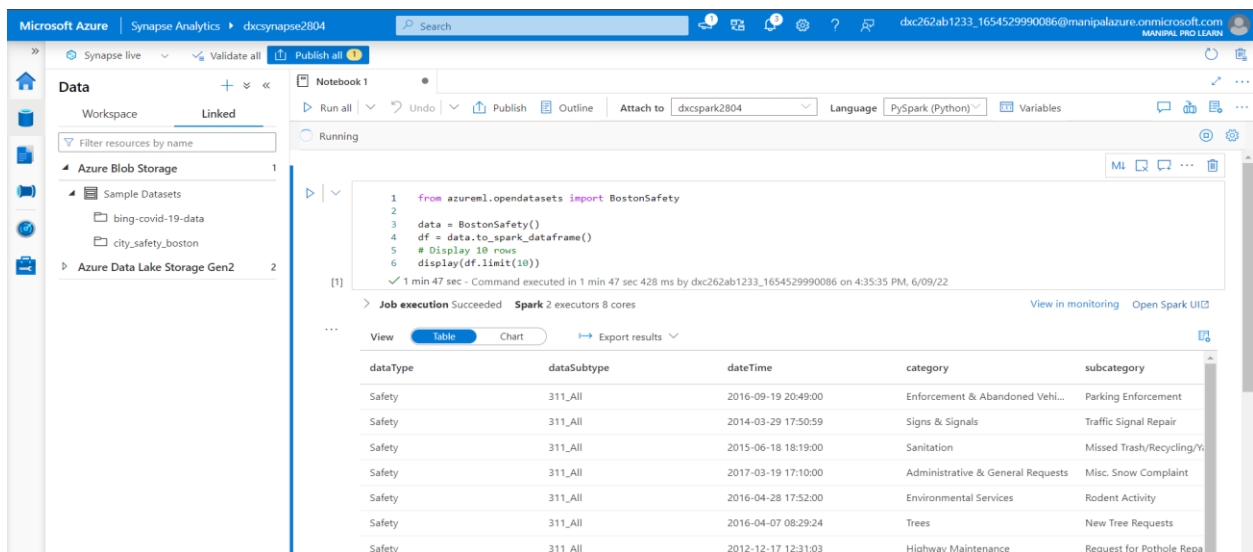


Apache spark pool is created



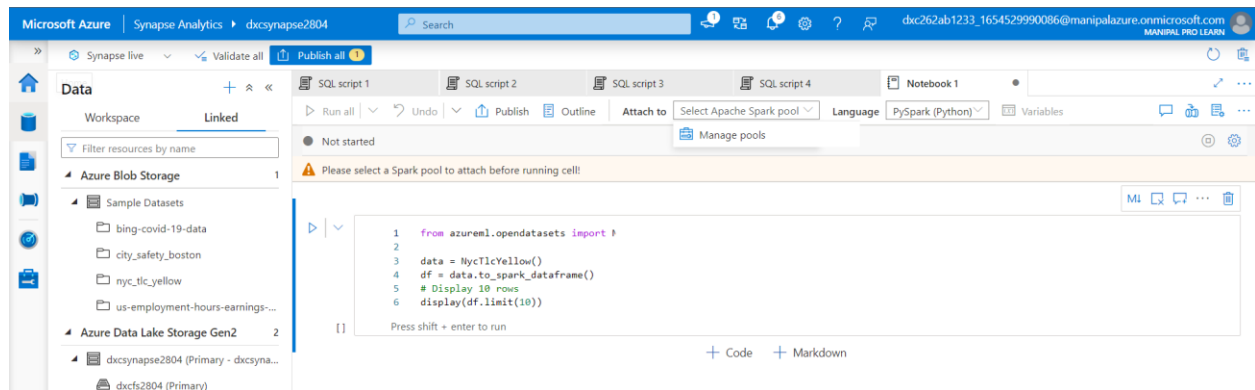
Now go to the data frame page where we can now run attach dxcspark2804 to the data frame

And click on run all language Pyspark while running the notebook don't keep more than 1 notebook

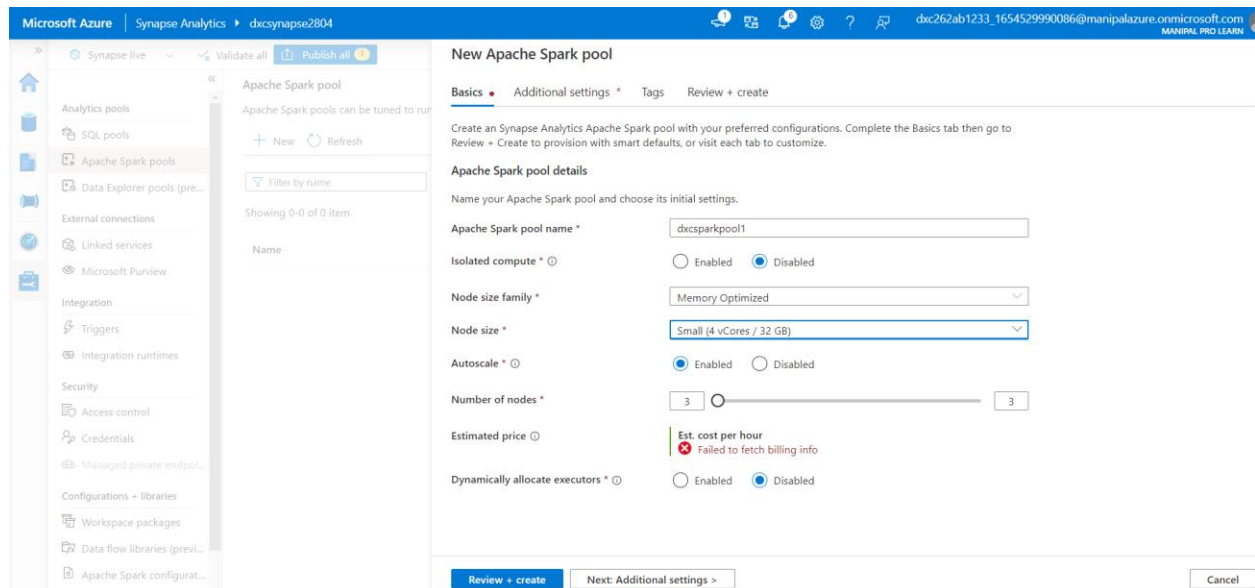


5) Explain the steps with screenshots how to create Spark pool in Azure Synapse analytics?

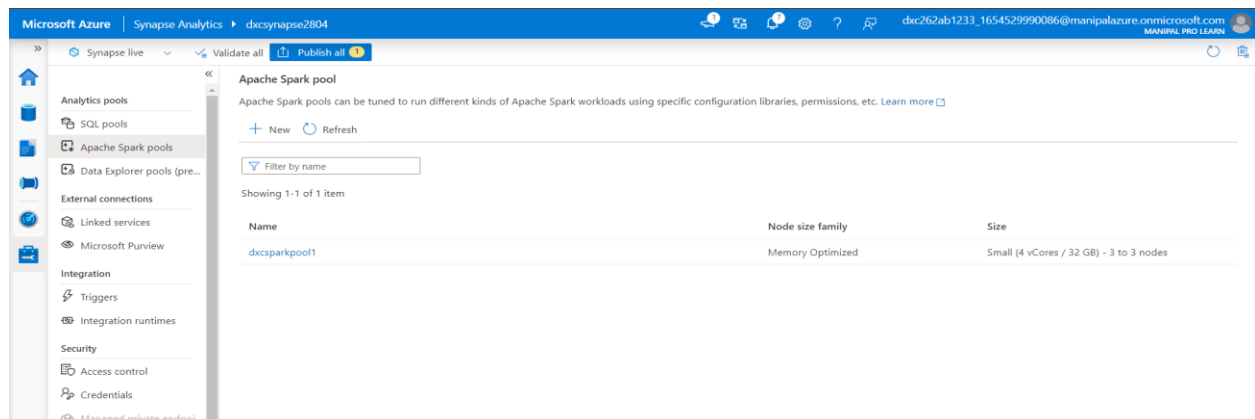
Go to Manage pools and go to sparkpool



Add new sparkpool name and node size small and no od nodes max 3 min 3 and review and create

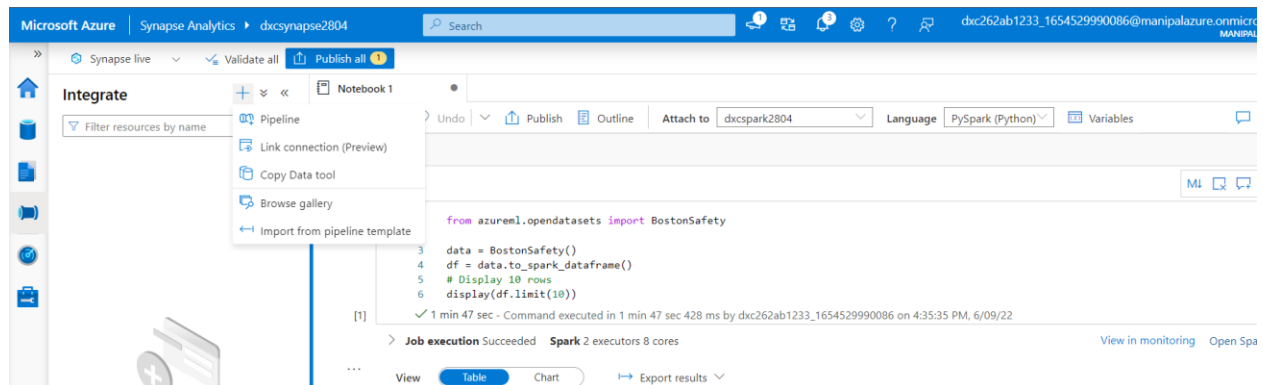


Apache spark pool is created

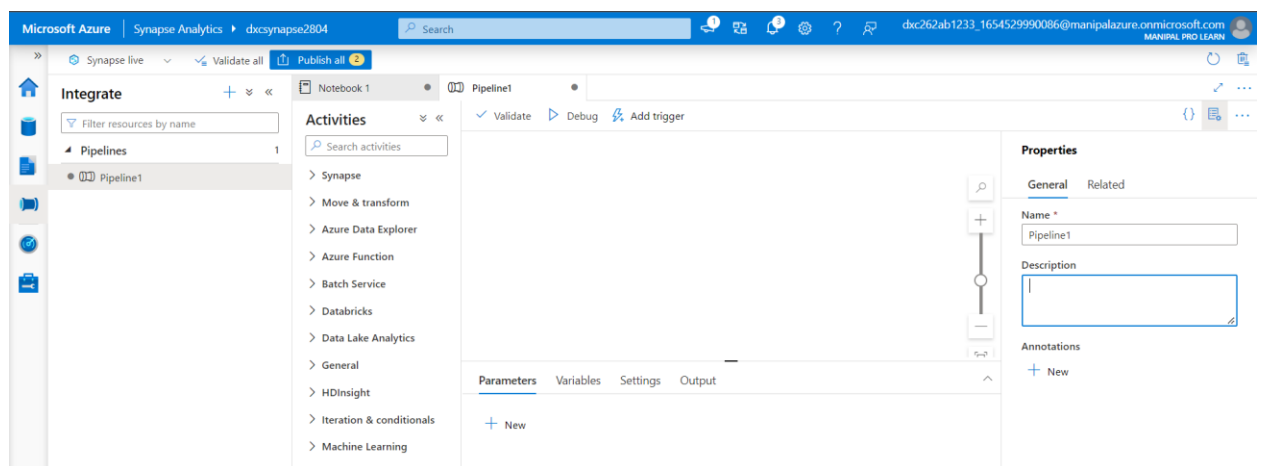


6) Explain the steps with screenshots how to create pipeline in Azure Synapse analytics?

Go to Integrate tab-> + click on Pipeline

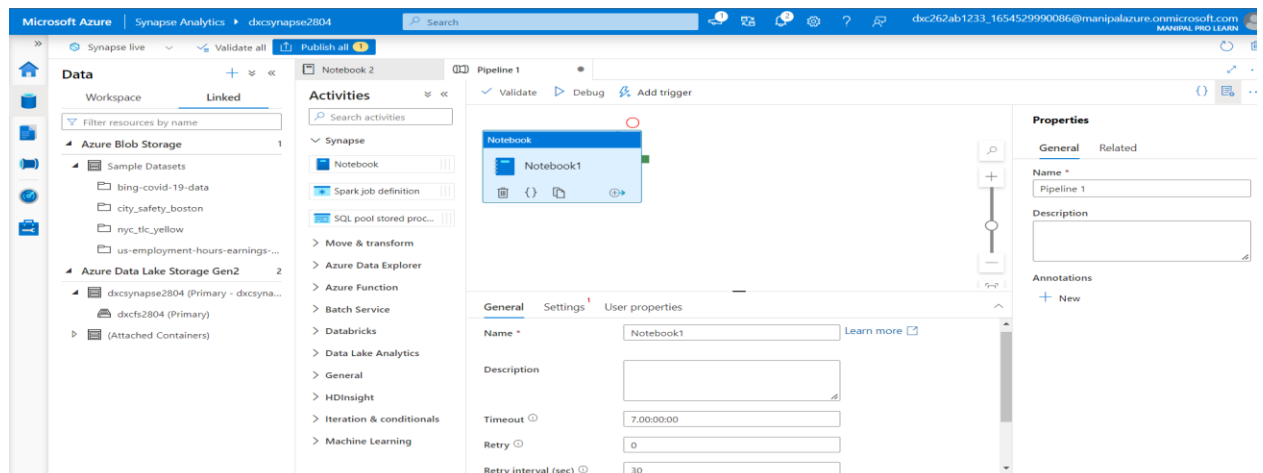


Give the name of the pipeline

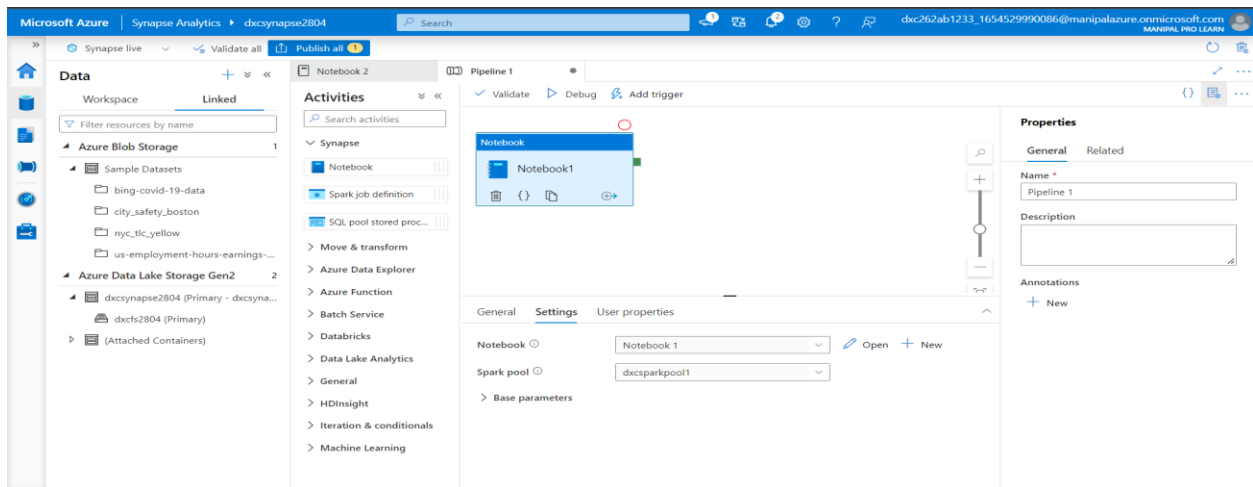


7) Explain the steps with screenshots how to automate the pipelines in Azure Synapse analytics?

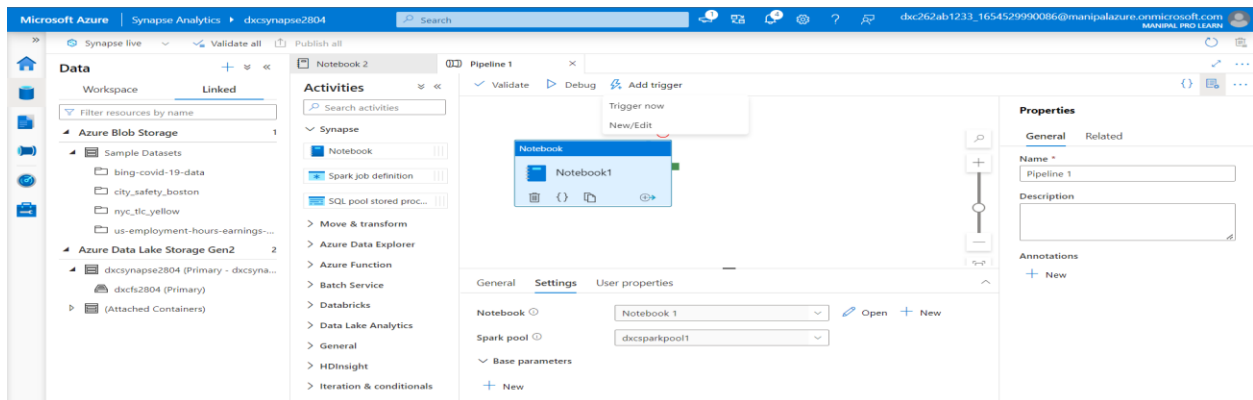
Pipeline->In Activities ->Synapse->Notebook select notebook1 and drag and drop



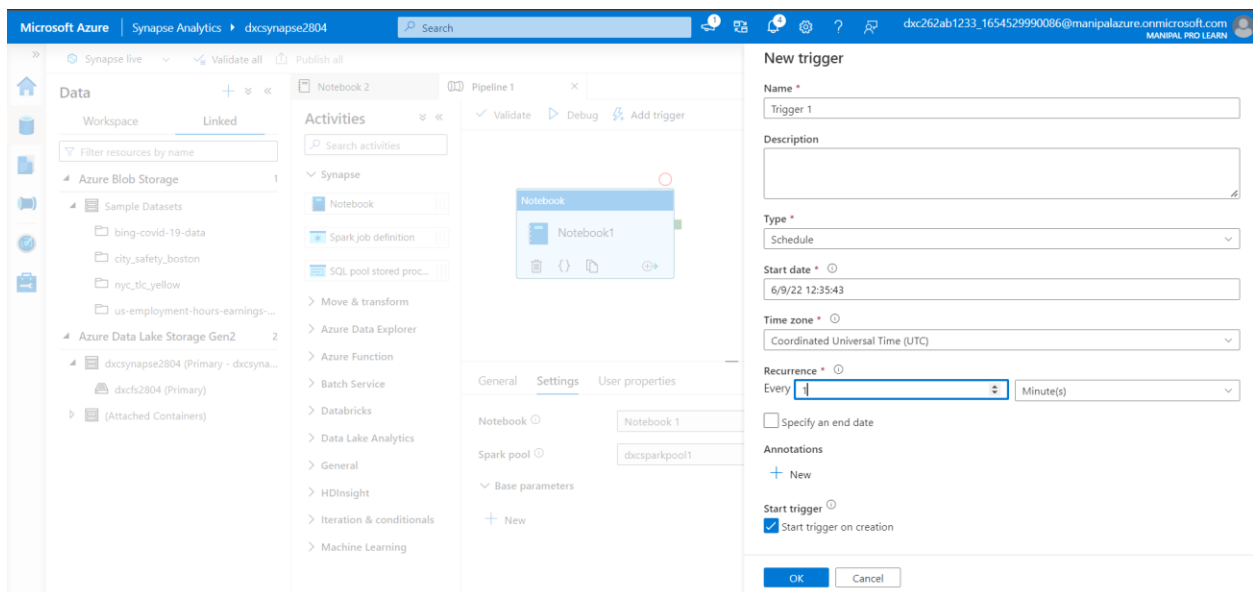
In setting give notebook1 and dxcspark2804 and publish the pipeline and validate



Add trigger->trigger now



If we want to trigger new/edit in Add trigger we can give the timing it will automatically trigger



First we need to publish the trigger To check go to monitor tab Trigger run refresh every one minute keep the time zone Chennai it will trigger every time

Microsoft Azure | Synapse Analytics | dxcynapse2804

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 Copy filters Export to CSV

Showing 1 - 3 items Last refreshed 0 minutes ago

Pipeline name	Run start	Run end	Duration	Triggered by	Status	Error	Run
Pipeline1	Jun 9, 2022, 4:45:31 pm	--	00:02:59	Manual trigger	In progress		Original
Pipeline1	Jun 9, 2022, 4:47:01 pm	--	00:01:29	Trigger 1	In progress		Original
Pipeline1	Jun 9, 2022, 4:48:00 pm	--	00:00:30	Trigger 1	In progress		Original

Microsoft Azure | Synapse Analytics | dxcynapse2804

Trigger runs

All Schedule Tumbling window Storage events Custom events Refresh Edit columns

Chennai, Kolkata, Mu... Last 24 hours Trigger name: All Status: All Runs: Latest runs Export to CSV

Showing 1 - 3 items

Trigger name	Trigger type	Trigger time	Status	Pipelines	Run	Message	Properties	Run ID
Trigger 1	Schedule trigger	Jun 9, 2022, 4:48:59	Succeeded	1	Original			08585468333458041
Trigger 1	Schedule trigger	Jun 9, 2022, 4:48:00	Succeeded	1	Original			08585468334049776
Trigger 1	Schedule trigger	Jun 9, 2022, 4:47:00	Succeeded	1	Original			08585468334646105

Like this we go the automation

8) Explain the steps with screenshots how to Databricks ?

Go to Azure and search for Data bricks give the resource group and workspace name and pricing tier trial

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

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 * dxcrg2317 [Create new](#)

Instance Details

Workspace name * dxcd2804

Region * East US

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

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

We get deployment is complete and go to resource

The screenshot shows the Microsoft Azure portal interface. At the top, there's a search bar and navigation icons. The main header indicates the user is logged in as 'dxc262ab1233_1654529...'. The page title is 'dxcrg2317_dxcdb2804 | Overview'. On the left, there's a sidebar with navigation options: Overview, Inputs, Outputs, and Template. The main content area shows a green checkmark and the text 'Your deployment is complete'. Below this, it lists deployment details: Deployment name: dxcrg2317_dxcdb2804, Subscription: Azure-DXC262AB12Lab, Resource group: dxcrg2317, Start time: 6/9/2022, 4:57:31 PM, and Correlation ID: 5b019318-1f6d-489f-a1dc-e2ae77c637f4. There are buttons for 'Delete', 'Cancel', 'Redeploy', and 'Refresh'. A 'Go to resource' button is also present. On the right, there's a 'Cost Management' section with a green shield icon and text: 'Get notified to stay within your budget and prevent unexpected charges on your bill. Set up cost alerts >'. There's also a 'We'd love your feedback!' message with a link to provide feedback.

Here Launch WorkSpace

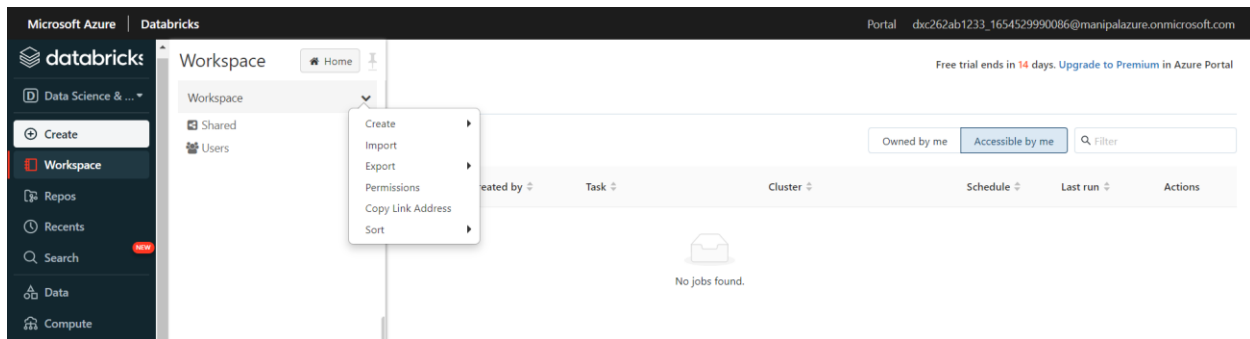
The screenshot shows the Microsoft Azure portal interface for the resource 'dxcdb2804' (Azure Databricks Service). The left sidebar shows navigation options: Overview, Activity log, Access control (IAM), Tags, Settings, Virtual Network Peerings, Encryption, Properties, Locks, Automation, Tasks (preview), Export template, Support + troubleshooting, and New Support Request. The main content area shows the 'Essentials' section with details: Status: Active, Resource group: dxcrg2317, Location: East US, Subscription: Azure-DXC262AB12Lab, Subscription ID: 393eef23-258e-4780-b646-3ff567d08e5e, and Tags (edit) with a link to 'Click here to add tags'. It also shows the Managed Resource Group: databricks-rg-dxcdb2804-42j3gy6kh6k3s, URL: https://adb-575702103626145.5.azure.databricks.net, and Pricing Tier: Trial (Premium - 14-Days Free DBUs). There are buttons for 'Launch Workspace' and 'Upgrade to Premium'. The right side shows a 'JSON View' button.

We will get the home page of data bricks

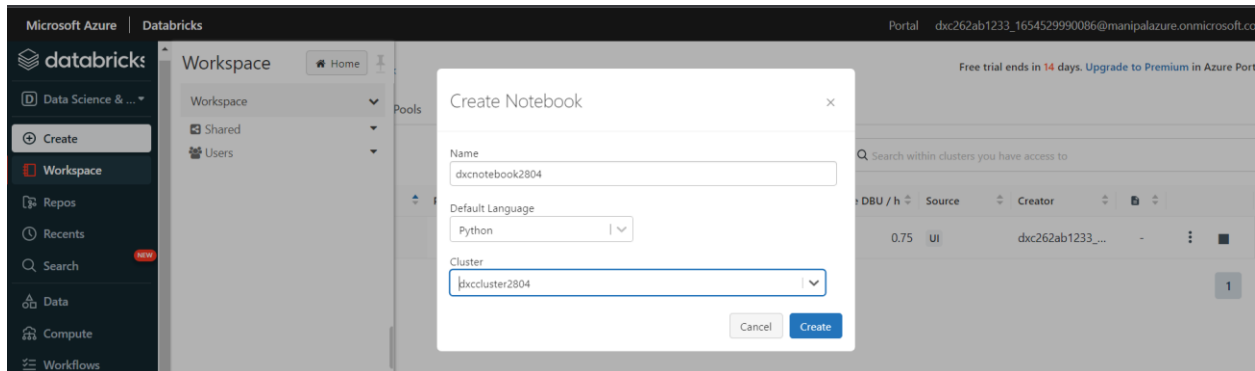
The screenshot shows the Databricks home page. The top navigation bar includes 'Microsoft Azure | Databricks' and a 'Portal' link. The main content area is divided into several sections. On the left, there's a 'Get started' section with a 'Set up your workspace' area containing buttons for 'Create a cluster', 'Ingest data', and 'Invite your team'. Below this is a 'Next steps' area with buttons for 'Explore Notebook gallery' and 'Read documentation'. The main section is 'Data Science & Engineering', which includes a 'Notebook' section with a 'Create a notebook' button, a 'Guide: Quickstart tutorial' section with a 'Start tutorial' button, a 'Data import' section with a 'Browse files' button, and a 'Partner Connect' section with a 'View all partners' button. There's also a 'Recents' section with a table showing 'Name' and 'Last viewed' columns, and a message 'There are no recents yet'. At the bottom, there are links for 'Documentation', 'Release notes', and 'Blog posts'.

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

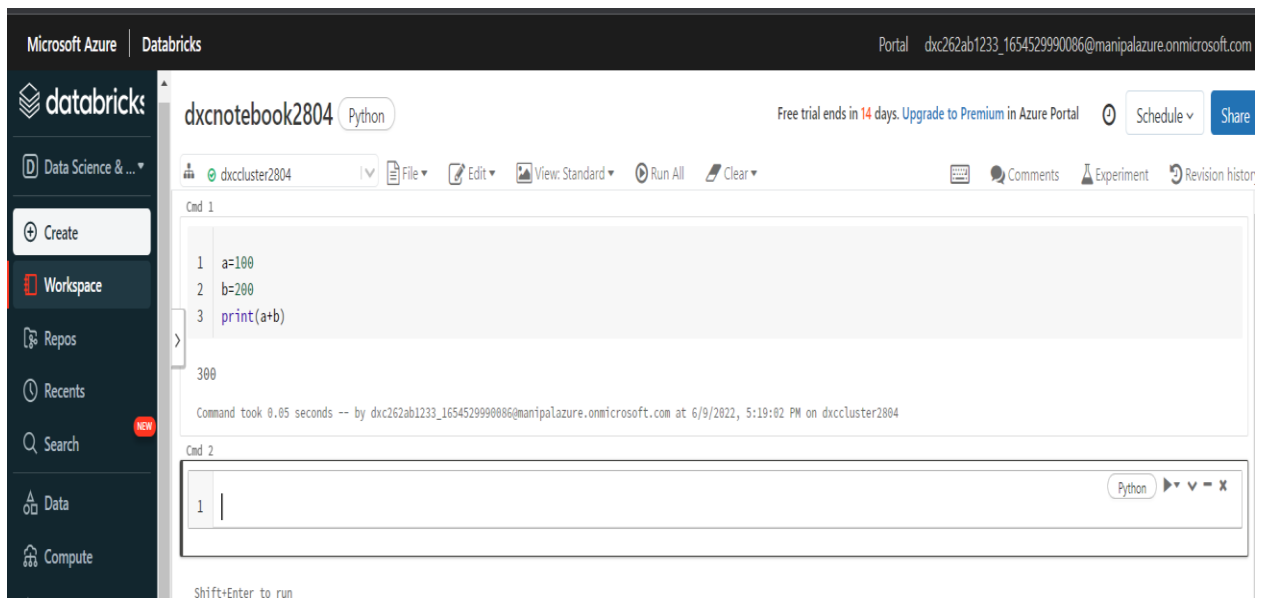
Go to workspace-> Create->notebook



Give the name of the notebook and give the cluster before you created and create the notebook

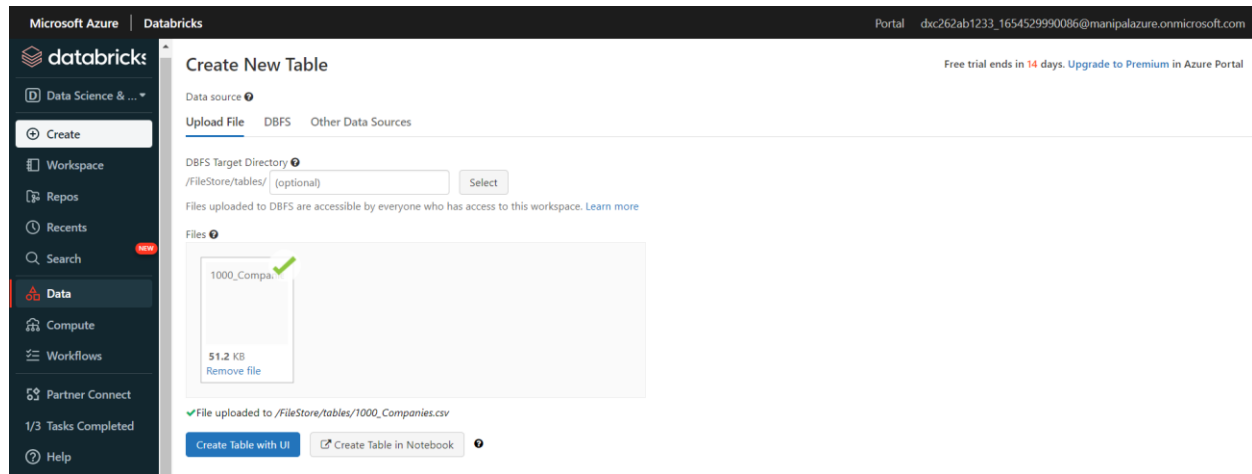


Notebook will be created and we should attach the cluster with out forgetting

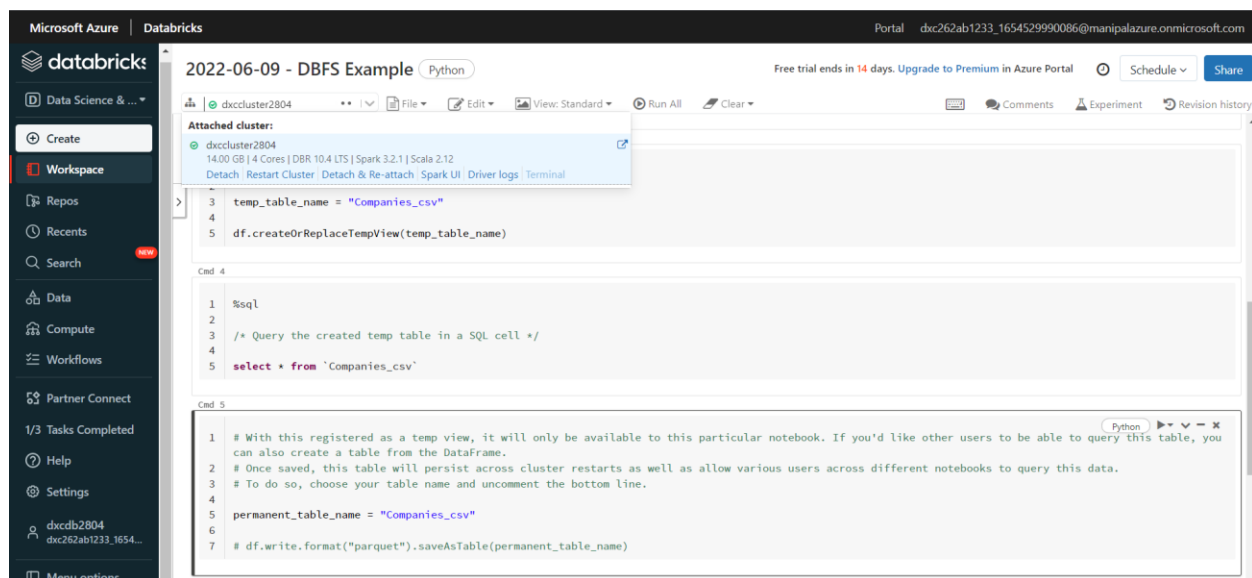


10) Explain the steps with screenshots how to insert data into data bricks notebook & display the result?

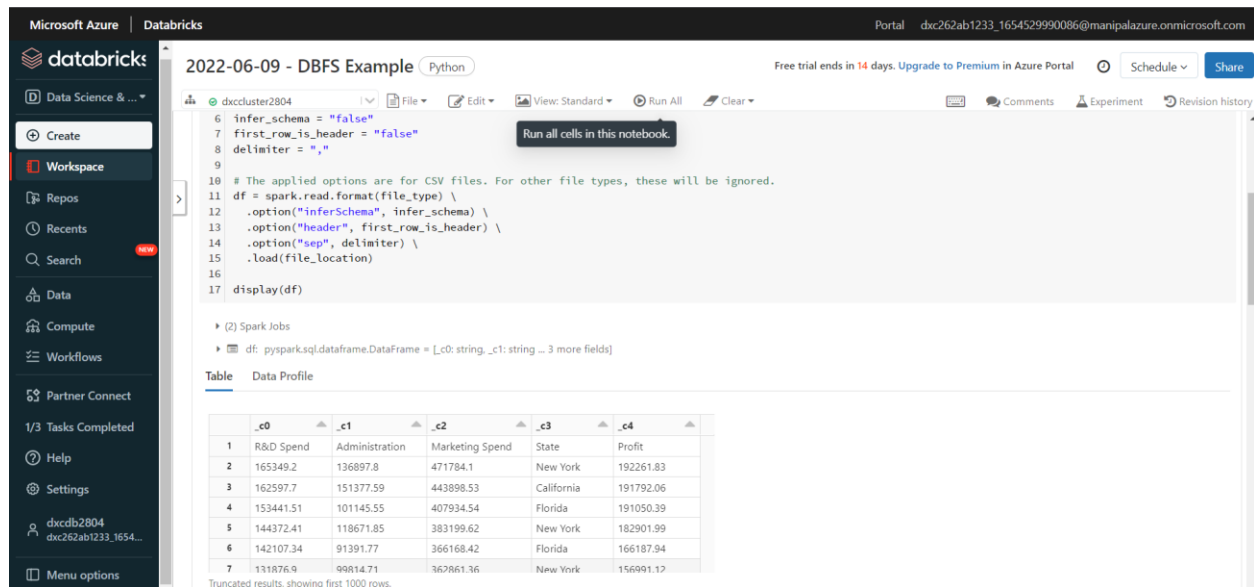
Go to data tab and click create table and drag and drop the data you want to insert



Click on the create Table in Notebook



Attach the cluster at the top that you create and remove the 1000_ in the table name and 3 places
And click on run all



The screenshot shows a Databricks notebook titled "2022-06-09 - DBFS Example" in Python. The code reads a CSV file from DBFS and displays the first 1000 rows. Below the code, a table view shows the 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	90814.71	362861.36	New York	156991.12

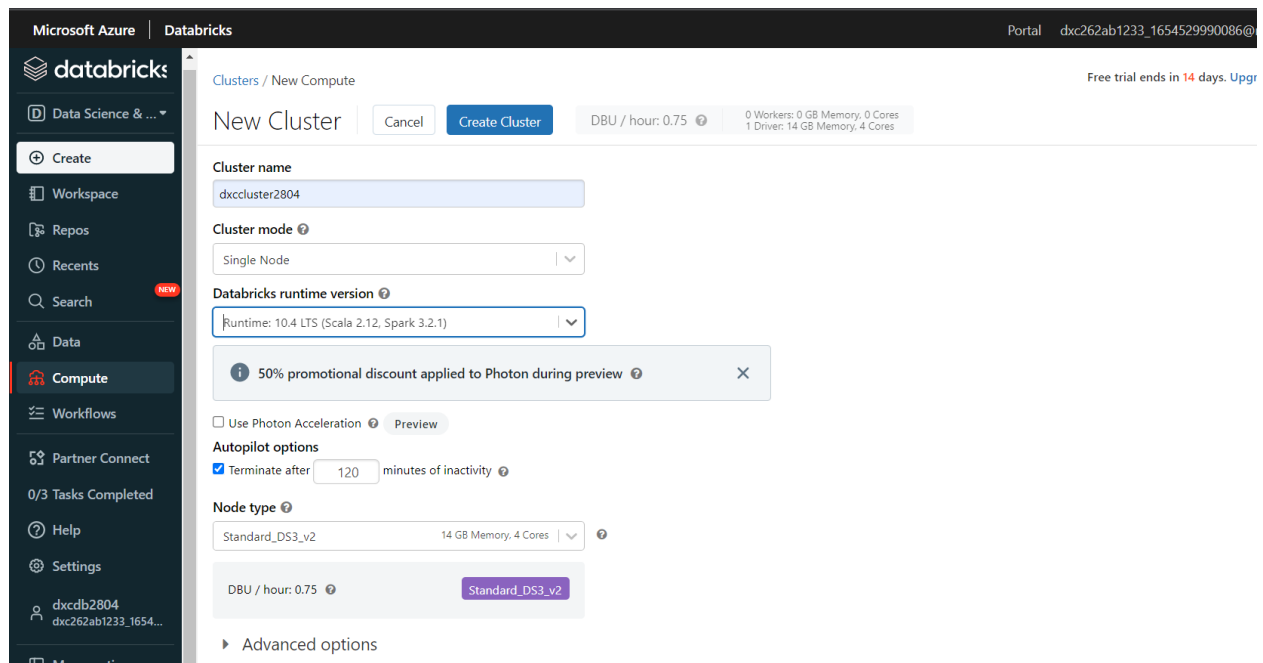
Truncated results, showing first 1000 rows.

In this we can see the data in tables which we have inserted

11) Explain the steps with screenshots how to create cluster in data bricks ?

Go to Compute -> Click Create Cluster

Give the name and cluster mode as single node and click Create Cluster



The screenshot shows the "New Cluster" configuration page in Databricks. The cluster name is "dxcluster2804", the cluster mode is "Single Node", and the Databricks runtime version is "runtime: 10.4 LTS (Scala 2.12, Spark 3.2.1)". The node type is "Standard_DS3_v2" with 14 GB Memory and 4 Cores. The DBU / hour is 0.75. A promotional discount of 50% is applied to Photon during preview. The Autopilot options are set to "Terminate after 120 minutes of inactivity".

We get a tick mark that we have created a cluster

The screenshot shows the Databricks Clusters page for a cluster named **dxcccluster2804**. The cluster is in a 'Running' state, indicated by a green checkmark. The configuration details are as follows:

- Policy:** Unrestricted
- Cluster mode:** Single Node
- Databricks Runtime Version:** 10.4 LTS (includes Apache Spark 3.2.1, Scala 2.12)
- Autopilot options:** ☒ Terminate after 120 minutes of inactivity
- Node type:** Standard_DS3_v2 (14 GB Memory, 4 Cores)
- DBU / hour:** 0.75

The left sidebar shows the navigation menu with 'Compute' selected. The top right corner indicates a free trial ending in 14 days.

To check the cluster go to compute here it will show the complete details of the cluster

The screenshot shows the Databricks Compute page, which displays a list of clusters. The cluster **dxcccluster2804** is listed with the following details:

Name	Policy	Runtime	Active memory	Active cores	Active DBU / h	Source	Creator
dxcccluster2804	-	10.4	14 GB	4 cores	0.75	UI	dxcc262ab1233_...

The left sidebar shows the navigation menu with 'Compute' selected. The top right corner indicates a free trial ending in 14 days.