

Assignment 9 June 2022

NAME: ASHOK KUMAR

ROLL NUMBER: DXC-262-AB-1233

BATCH: DXC-262-ANALYTICS-B12-AZURE

COMPANY: DXC TECHNOLOGY

EMPLOYEE DOMAIN: AZURE ANALYTICS

TRAINING UNDER: MANIPAL PRO LEARN

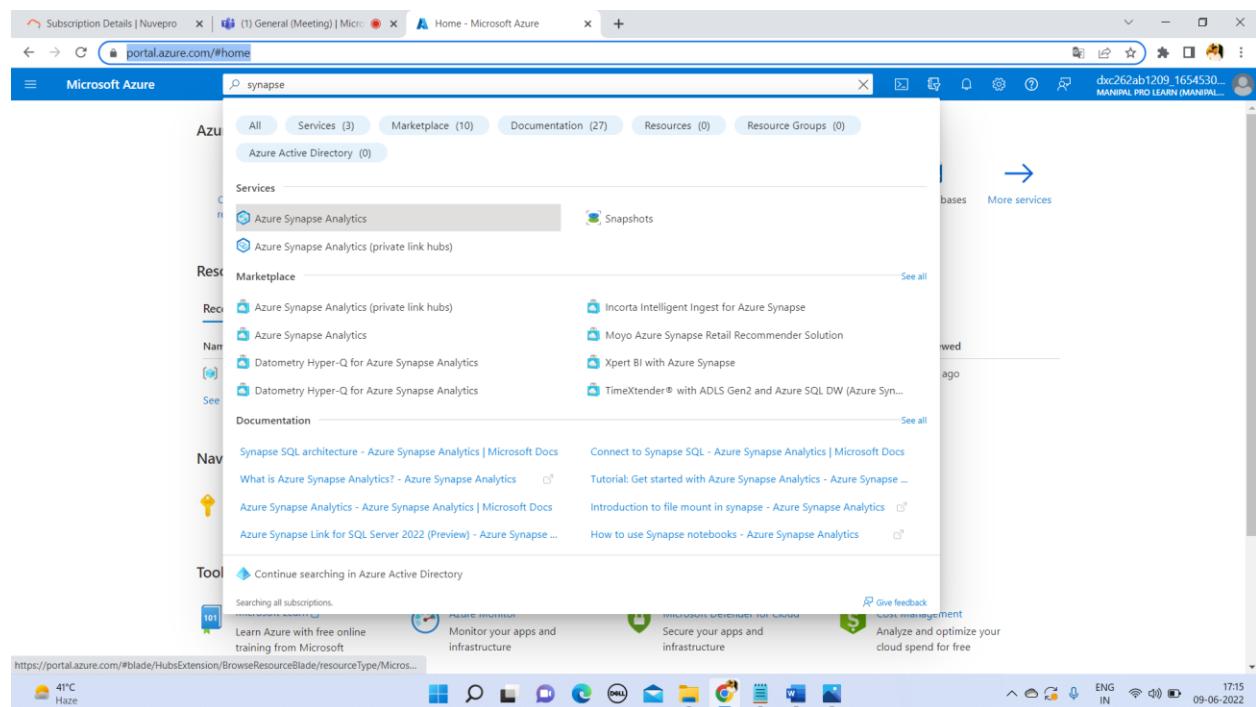
TRAINER NAME: MR. AJAY KUMAR

DATE OF SUBMISSION: 9 JUNE 2022

NO. OF QUESTIONS: 11

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

ANSWER: Go to <https://portal.azure.com/#home> and search azure synapse and click on it.



The screenshot shows the Microsoft Azure portal interface. The top navigation bar includes links for 'Subscription Details | Nuvepro', '(1) General (Meeting) | Micro...', and 'Azure Synapse Analytics - Micro...'. The main title is 'Azure Synapse Analytics' with a subtitle 'Manipal Pro Learn (manipalazure.onmicrosoft.com)'. Below the title are buttons for '+ Create', 'Manage view', 'Refresh', 'Export to CSV', 'Open query', and 'Assign tags'. A search bar says 'Search resources, services, and docs (G+ /)'. Filter options include 'Subscription == all', 'Resource group == all', 'Location == all', and 'Add filter'. The main content area has columns for 'Name', 'Type', 'Resource group', 'Location', and 'Subscription'. A large message 'No Azure Synapse Analytics to display' is centered, with a sub-message about Synapse Analytics being a fully-managed service for building modern data warehouses. It includes a 'Create Synapse workspace' button and a 'Service overview' link.

Click on create bottom to create synapse workspace.

And fill the details resource group, workspace name etc.

The screenshot shows the 'Create Synapse workspace' wizard in the Microsoft Azure portal. The 'Basics' step is selected. The 'Project details' section requires selecting a subscription (Azure-DXC262AB12Lab), a resource group (ashokrg99), and a managed resource group (Enter managed resource group name). The 'Workspace details' section requires entering a workspace name (ashoksynapse99), choosing a region (East US), selecting a Data Lake Storage Gen2 account (From subscription, account (New) ashokadlkstro99), and setting security (Review + create, < Previous, Next: Security >).

Go to aecurity tab and fill the login details of sql server admin.

The screenshot shows the 'Create Synapse workspace' page in the Azure portal. The 'Security' tab is selected. Under 'Authentication', 'Use both local and Azure Active Directory (Azure AD) authentication' is selected. SQL Server admin login is set to 'sqladminuser'. Under 'System assigned managed identity permission', there is a note about granting network access to Data Lake Storage Gen2. The 'Review + create' button is visible at the bottom.

Click on review+create

The screenshot shows the 'Create Synapse workspace' page in the Azure portal. The 'Review + create' tab is selected. A green bar at the top indicates 'Validation succeeded'. The 'Basics' section shows the subscription is 'Azure-DXC262A812Lab', resource group is 'ashokrg99', region is 'East US', and workspace name is '(new) ashoksynapse99'. The 'Create' button is highlighted at the bottom.

Click on create.

The screenshot shows the Microsoft Azure Synapse Analytics deployment progress page. At the top, there are three tabs: "Subscription Details | Nuvepro", "(1) General (Meeting) | Micro", and "Microsoft.Azure.SynapseAnalytics-20220609171637 | Overview". The current tab is "Microsoft.Azure.SynapseAnalytics-20220609171637 | Overview". The main content area displays the message "Deployment is in progress". Below this, deployment details are listed: Deployment name: Microsoft.Azure.SynapseAnalytics-20220609171..., Start time: 6/9/2022, 5:22:06 PM, Subscription: Azure-DXC262AB1Lab, Correlation ID: 350505bf-032d-4112-8c49-91601590b84a. A "Deployment details (Download)" section shows "No results.". On the right side, there are promotional banners for Microsoft Defender for Cloud, Free Microsoft tutorials, and Work with an expert.

The screenshot shows the Microsoft Azure Synapse Analytics deployment completion page. The layout is identical to the previous screenshot, with the same tabs at the top. The main content area now displays the message "Your deployment is complete". Below this, deployment details are listed: Deployment name: Microsoft.Azure.SynapseAnalytics-20220609171..., Start time: 6/9/2022, 5:22:06 PM, Subscription: Azure-DXC262AB1Lab, Correlation ID: 350505bf-032d-4112-8c49-91601590b84a. A "Next steps" section includes a "Go to resource group" button. On the right side, there are promotional banners for Cost Management, Microsoft Defender for Cloud, Free Microsoft tutorials, and Work with an expert.

Subscription Details | Nuvepro | (1) General (Meeting) | Microsoft | ashokrg99 - Microsoft Azure | +

portal.azure.com/#@manipalazure.onmicrosoft.com/resource/subscriptions/442b51c3-1fed-40a7-9d05-105684c1217c/resourceGroups/ashokrg99/overview

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

Home > Microsoft.Azure.SynapseAnalytics-20220609171637 > ashokrg99

ashokrg99 Resource group

Search (Ctrl+)

+ Create Manage view Delete resource group Refresh Export to CSV Open query Assign tags Move Delete Export template JSON View

Overview

Essentials

Subscription (move) : Azure-DXC262AB12Lab Deployments : Succeeded

Subscription ID : 442b51c3-1fed-40a7-9d05-105684c1217c Location : East US

Tags (edit) Click here to add tags

Resources Recommendations

Filter for any field... Type == all Location == all Add filter

Showing 1 to 2 of 2 records. Show hidden types

Name	Type	Location
ashokadikstro99	Storage account	East US
ashoksynapse99	Synapse workspace	East US

No grouping List view

< Previous Page 1 of 1 Next >

Give feedback

41°C Haze

17:32 09-06-2022 ENG IN

Subscription Details | Nuvepro | (1) General (Meeting) | Microsoft | ashoksynapse99 - Microsoft Azure | +

portal.azure.com/#@manipalazure.onmicrosoft.com/resource/subscriptions/442b51c3-1fed-40a7-9d05-105684c1217c/resourceGroups/ashokrg99/providers/Microsoft.Synapse/workspace/ashokrg99

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

Home > Microsoft.Azure.SynapseAnalytics-20220609171637 > ashokrg99 > ashoksynapse99

ashoksynapse99 Synapse workspace

Search (Ctrl+)

+ New dedicated SQL pool + New Apache Spark pool + New Data Explorer pool (preview) Refresh Reset SQL admin password Delete JSON View

Overview

Essentials

Resource group (move) : ashokrg99 Networking : Show firewall settings

Status : Succeeded Primary ADLS Gen2 account : https://ashokadikstro99.dfs.core.windows.net

Location : East US Primary ADLS Gen2 file s... : ashokfs99

Subscription (move) : Azure-DXC262AB12Lab SQL admin username : sqldadminuser

Subscription ID : 442b51c3-1fed-40a7-9d05-105684c1217c SQL Active Directory ad... : dxc262ab1209_165453008226@manipalazure.onmicrosoft.com

Managed virtual network : No Dedicated SQL endpoint : ashoksynapse99.sql.azuresynapse.net

Managed identity object id : f3f5a8fa-a9f8-444f-8a10-5891f6aac322 Serverless SQL endpoint : ashoksynapse99-ondemand.sql.azuresynapse.net

Workspace web URL : https://web.azuresynapse.net/workspace=%2bsubscriptions%2f44... Development endpoint : https://ashoksynapse99.dev.azuresynapse.net

Tags (edit) Click here to add tags

Getting started

Open Synapse Studio Start building your fully-integrated analytics solution and unlock new insights. Open

Read documentation Learn how to be productive quickly. Explore concepts, tutorials, and samples. Learn more

Analytics pools

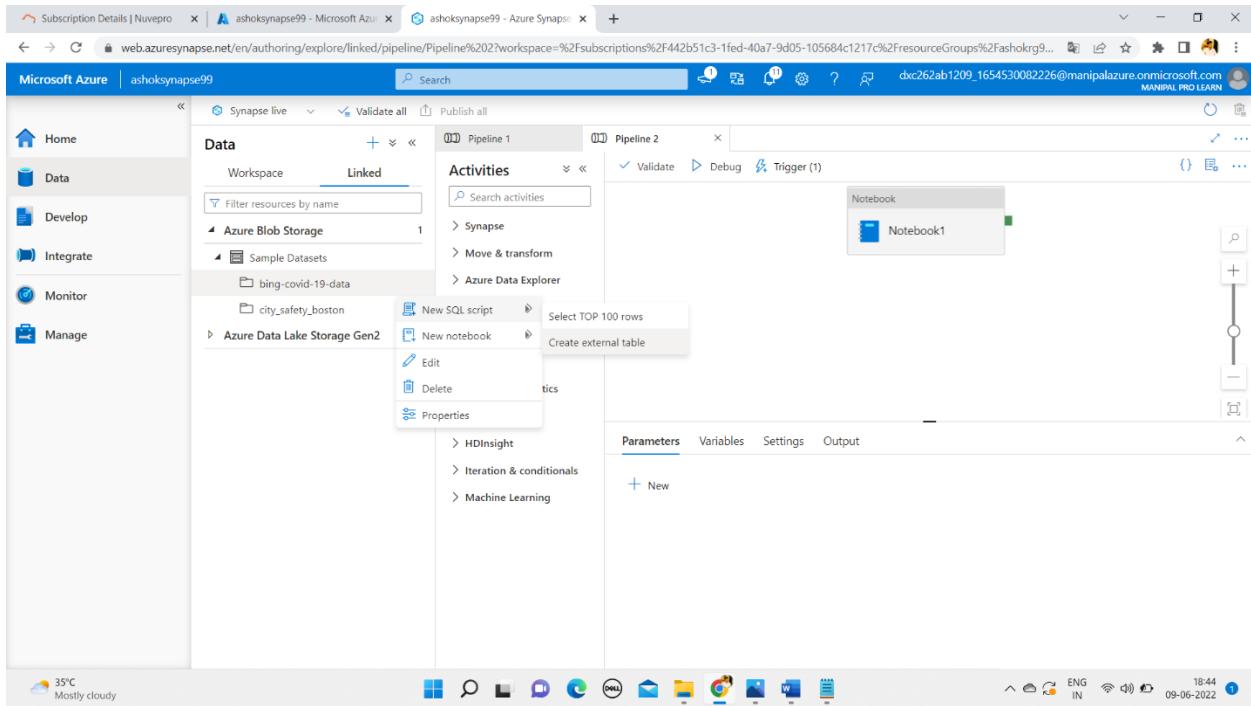
Name Type Size

41°C Haze

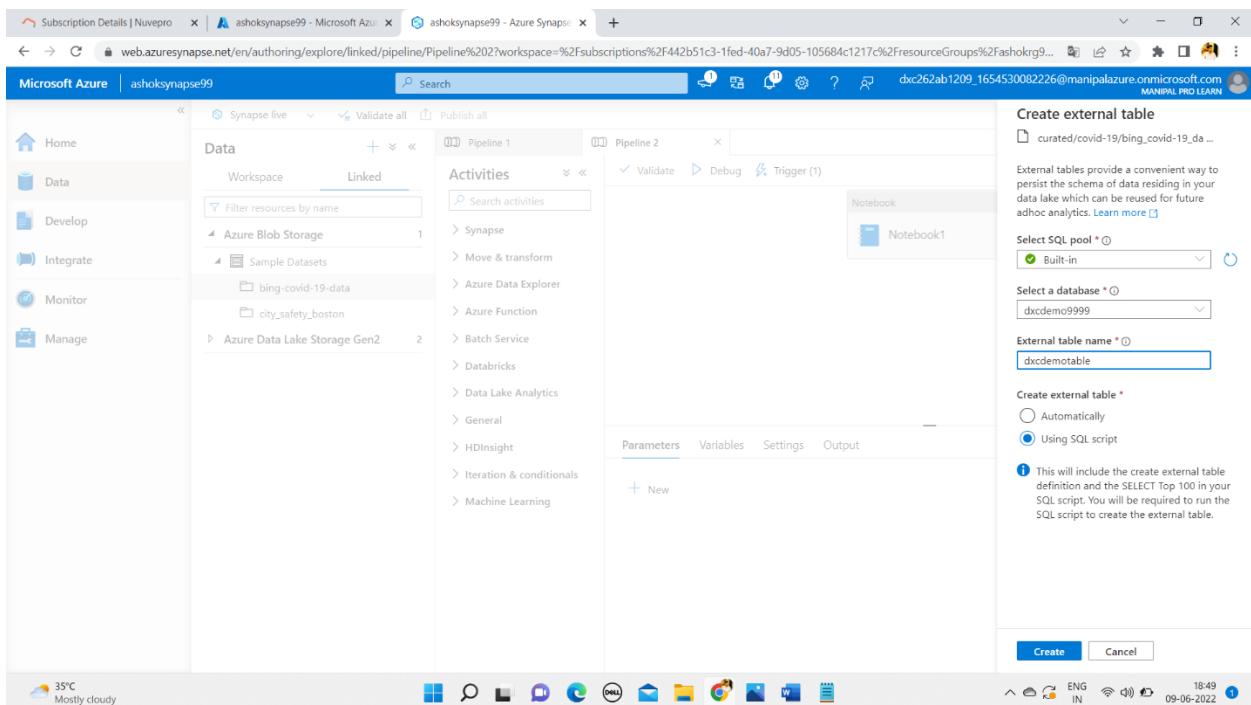
17:32 09-06-2022 ENG IN

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

ANSWER:



Click on create external table.



The screenshot shows the Microsoft Azure Synapse Studio interface. On the left, the navigation bar includes Home, Data, Develop, Integrate, Monitor, and Manage. Under the Data section, 'Linked' is selected, showing resources from Azure Blob Storage (Sample Datasets: bing-covid-19-data, city_safety_boston) and Azure Data Lake Storage Gen2. The main workspace displays a SQL script titled 'SQL script 2' under 'Publishing'. The script creates an external file format for Parquet files and an external table named 'dxcdemotable' with columns for id, updated, confirmed, confirmed_change, deaths, deaths_change, recovered, recovered_change, latitude, longitude, iso2, iso3, country_region, admin_region_1, iso_subdivision, admin_region_2, and load_time.

```
1 IF NOT EXISTS (SELECT * FROM sys.external_file_formats WHERE name = 'SynapseParquetFormat')
2 CREATE EXTERNAL FILE FORMAT [SynapseParquetFormat]
3 WITH (FORMAT_TYPE = PARQUET)
4 GO
5
6 IF NOT EXISTS (SELECT * FROM sys.external_data_sources WHERE name = 'public_pandemic')
7 CREATE EXTERNAL DATA SOURCE [public_pandemic]
8 WITH (
9     LOCATION = 'https://pandemicdatalake.blob.core.windows.net/public',
10 )
11 Go
12
13 CREATE EXTERNAL TABLE dxcdemotable (
14     [id] int,
15     [updated] date,
16     [confirmed] int,
17     [confirmed_change] int,
18     [deaths] int,
19     [deaths_change] smallint,
20     [recovered] int,
21     [recovered_change] int,
22     [latitude] float,
23     [longitude] float,
24     [iso2] varchar(8000),
25     [iso3] varchar(8000),
26     [country_region] varchar(8000),
27     [admin_region_1] varchar(8000),
28     [iso_subdivision] varchar(8000),
29     [admin_region_2] varchar(8000),
30     [load_time] datetime2(7)
31 )
32
```

The screenshot shows the Microsoft Azure Synapse Studio interface. The left navigation bar includes Home, Data, Develop, Integrate, Monitor, and Manage. Under the Data section, 'Analytics pools' is selected, showing 'SQL pools' (with one item: 'Built-in'), 'Apache Spark pools', and 'Data Explorer pools (preview)'. The main workspace displays the 'SQL pools' blade, which lists a single pool named 'Built-in' with details: Type: Serverless, Version: v2, Status: Online, Size: Auto, CPU utilization: N/A, Memory utilization: N/A, and Created on: N/A. The blade also includes sections for 'Activities' (SQL requests, KQL requests, Apache Spark applications, Data flow debug) and 'Integration' (Pipeline runs, Trigger runs, Integration runtimes, Link connections).

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

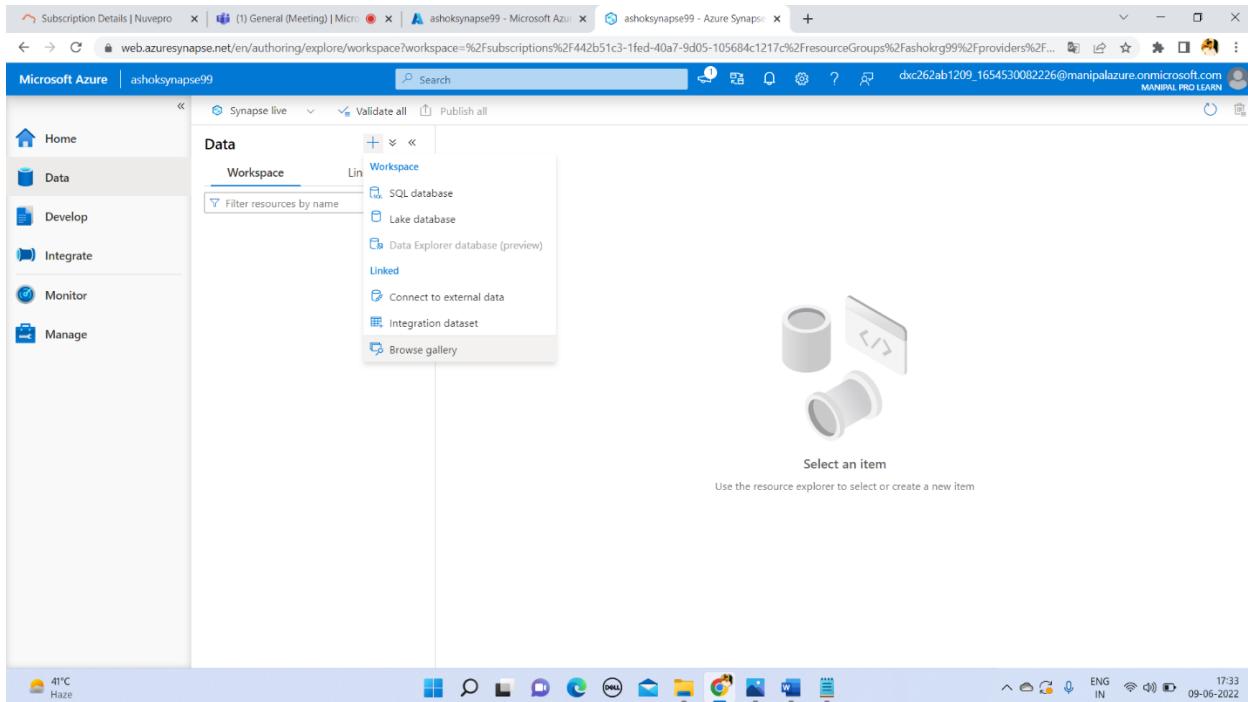
ANSWER:

The screenshot shows the Microsoft Azure portal interface. The top navigation bar includes tabs for 'Subscription Details | Nuvepro', '(1) General (Meeting) | Micro...', 'ashoksynapse99 - Microsoft Azure', and 'portal.azure.com'. The main content area is titled 'ashoksynapse99' and 'Synapse workspace'. On the left, there's a sidebar with sections like 'Overview', 'Essentials', 'Getting started', 'Analytics pools', 'Security', and 'Azure Active Directory'. The 'Essentials' section displays various workspace details such as Resource group, Status, Location, Subscription, Managed virtual network, Managed identity object, Workspace web URL, and Tags. Below this, there are two cards: 'Open Synapse Studio' and 'Read documentation'. The 'Analytics pools' section shows a table with columns 'Name', 'Type', and 'Size'. The bottom of the screen shows the Windows taskbar with icons for File Explorer, Task View, Start, Edge, Mail, Photos, OneDrive, Word, Excel, and Power BI.

click on open synapse studio and login.

The screenshot shows the 'ashoksynapse99' Synapse Analytics workspace home page. The top navigation bar has tabs for 'Subscription Details | Nuvepro', '(1) General (Meeting) | Micro...', 'ashoksynapse99 - Microsoft Azure', and 'ashoksynapse99 - Azure Synapse'. The main content area is titled 'Synapse Analytics workspace' and 'ashoksynapse99'. On the left, there's a sidebar with 'Home', 'Data', 'Develop', 'Integrate', 'Monitor', and 'Manage' options. The central area features three main buttons: 'Ingest' (Perform a one-time or scheduled data load), 'Explore and analyze' (Learn how to get insights from your data), and 'Visualize' (Build interactive reports with Power BI capabilities). To the right, there's a large, stylized 3D bar chart visualization. Below these buttons, there are sections for 'Discover more' (Knowledge center, Browse partners) and 'Recent resources'. The bottom of the screen shows the Windows taskbar with icons for File Explorer, Task View, Start, Edge, Mail, Photos, OneDrive, Word, Excel, and Power BI.

Go to DATA tab on the left side



Now click on + bottom then Browse gallery to access datasets.

A screenshot of the Microsoft Azure portal showing the 'Datasets' tab in the 'Gallery' section. The page lists various datasets such as 'Bing COVID-19 Data', 'Boston Safety Data', 'COVID Tracking Project', 'Chicago Safety Data', 'European Centre for Disease Prevention and Control (ECDC) Covid-19 Cases', 'NOAA Integrated Surface Data (ISD)', 'NYC Taxi & Limousine Commission - For-Hire Vehicle (FHV) trip records', 'NYC Taxi & Limousine Commission - green taxi trip records', and 'New York City Safety Data'. Each dataset card includes a preview icon, name, description, and ID. At the bottom, there are 'Continue' and 'Close' buttons.

Now click on bing covid-19 data then continue to import it.

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](#).

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/

Add dataset Back Close

Click on add dataset.

Synapse live Validate all Publish all

Data

Workspace Linked

Filter resources by name

- Azure Blob Storage 1
- Sample Datasets
 - bing-covid-19-data

Azure Data Lake Storage Gen2

- New SQL script
- New notebook
- Edit
- Delete
- Properties

Select TOP 100 rows
Create external table

Then Sample datasets → bing-covid-19-data → new sql script → select top 100 rows

```
-- This is auto-generated code
SELECT
    TOP 100 *
FROM
    OPENROWSET(
        BULK 'https://pandemicdatalake.blob.core.windows.net/public/curated/covid-19/bing-covid-19-data',
        FORMAT = 'parquet'
    ) AS [result];
```

Properties

Name: SQL script 1

Type: .sql script

Size: 262 bytes

Results settings per query:

- First 5000 rows (default)
- All rows

Write **TOP 500 *** at the place of **TOP 100 *** to run sample 500 rows.

Then publish all and run to display the output.

Publish all

You are about to publish all pending changes to the live environment. [Learn more](#)

Pending changes (1)

NAME	CHANGE	EXISTING
SQL script	SQL script 1 (New)	-

Publish Cancel

Subscription Details | Nuvepro | (1) General (Meeting) | Microsoft Edge | ashoksynapse99 - Microsoft Azure | ashoksynapse99 - Azure Synapse | +

Microsoft Azure | ashoksynapse99

Synapse live | Validate all | Publish all

Data | Workspace | Linked | Filter resources by name

Azure Blob Storage | Sample Datasets | bing-covid-19-data | Azure Data Lake Storage Gen2

SQL script 1

```

1 -- This is auto-generated code
2 SELECT
3     TOP_500 *
4 FROM
5     OPENROWSET(
6         BULK 'https://pandemicdatalake.blob.core.windows.net/public/curated/covid-19/bing_covid-19_data.parquet'
7     ) AS [result];

```

Results | Messages | View: Table | Chart | Export results

id	updated	confirmed	confirmed_change	deaths	deaths_change	recovered
338995	2020-01-21T00...	262	(NULL)	0	(NULL)	(NULL)
338996	2020-01-22T00...	313	51	0	0	(NULL)
338997	2020-01-23T00...	578	265	0	0	(NULL)
338998	2020-01-24T00...	841	263	0	0	(NULL)
338999	2020-01-25T00...	1320	479	0	0	(NULL)
339000	2020-01-26T00...	2014	694	0	0	(NULL)
339001	2020-01-27T00...	2798	784	0	0	(NULL)

Properties

General | Related (0)

Name: SQL script 1

Description:

Type: .sql script

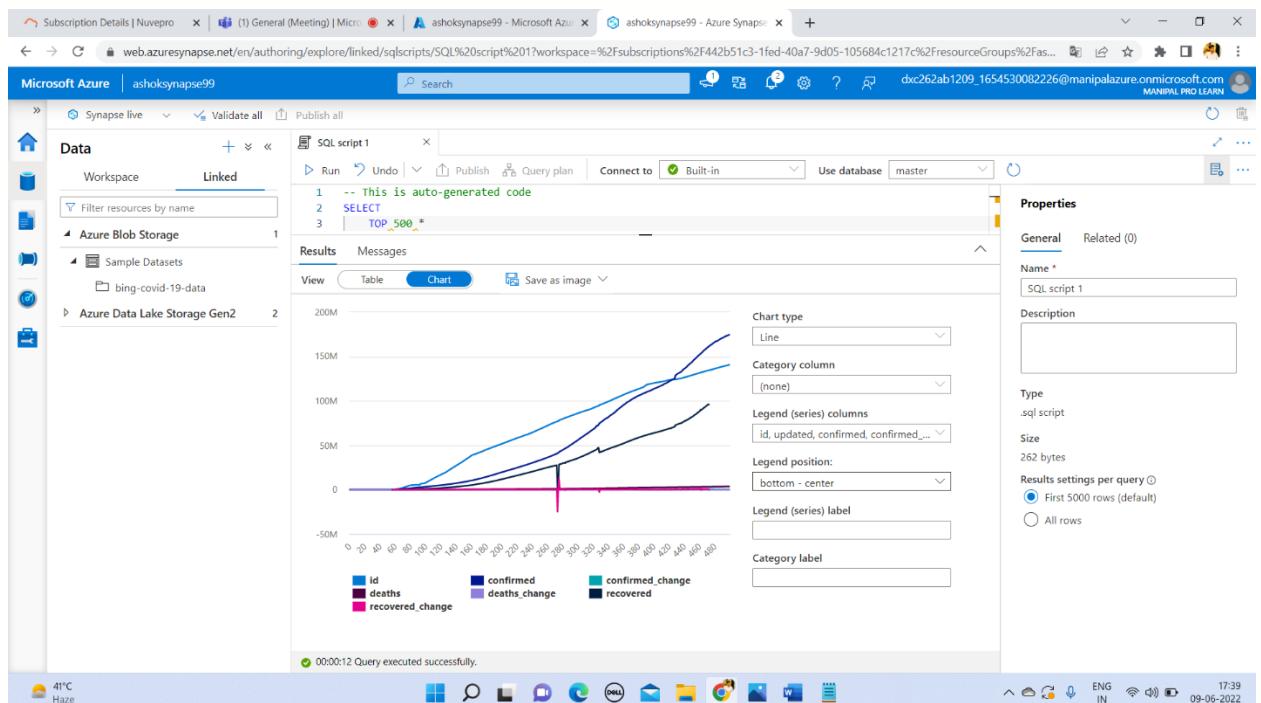
Size: 262 bytes

Results settings per query:

- First 5000 rows (default)
- All rows

00:00:12 Query executed successfully.

41°C Haze



Screenshot of Microsoft Azure Synapse Studio showing a SQL script named "SQL script 1" running against the master database. The script contains the following code:

```

1 -- This is auto-generated code
2 SELECT
3 | TOP_500 *

```

The results are displayed in a chart view, showing a pie chart with categories "deaths" and "0". Below the chart is a scatter plot of "deaths" values (ranging from 0 to 4012) against an unnamed column (ranging from 0 to 3380). The status bar indicates "00:00:12 Query executed successfully." and the system tray shows the date as 09-06-2022.

Screenshot of Microsoft Azure Synapse Studio showing the same SQL script "SQL script 1" running against the master database. The results are displayed in a chart view, showing an area chart of "recovered" values over time. The chart has a Y-axis ranging from -50M to 200M. The legend includes "id", "confirmed", "deaths_change", "confirmed_change", and "recovered_change". The status bar indicates "00:00:12 Query executed successfully." and the system tray shows the date as 09-06-2022.

Like this we can see output in all formats.

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

ANSWER:

Screenshot 1: Microsoft Azure Datasets page showing various datasets. The 'Datasets' tab is selected. The 'Boston Safety Data' dataset is highlighted.

ID	Name	Description
ID: bing-covid-19-data	Bing COVID-19 Data	Bing COVID-19 data includes confirmed, fatal, and recovered cases from all regions, updated daily.
ID: city_safety_boston	Boston Safety Data	Read data about 311 calls reported to the city of Boston. This dataset is stored in Parquet format and is updated daily.
ID: covid-tracking	COVID Tracking Project	The COVID Tracking Project dataset provides the latest numbers on tests, confirmed cases, hospitalizations, and patient outcomes from every US state and...
ID: city_safety_chicago	Chicago Safety Data	Read data about 311 calls reported to the city of Chicago. This dataset is stored in Parquet format and is updated daily.
ID: ecdc-covid-19-cases	European Centre for Disease Prevention and Control (ECDC) Covid-19 Cases	The latest available public data on geographic distribution of COVID-19 cases worldwide from the Euro...
ID: isd	NOAA Integrated Surface Data (ISD)	NOAA Integrated Surface Data (ISD) provides Worldwide hourly weather history data sourced from the National Oceanic and Atmospheric...
ID: nyc_tlc_fhv	NYC Taxi & Limousine Commission - For-Hire Vehicle (FHV) trip records	The For-Hire Vehicle trip records include fields capturing the dispatching base license number a...
ID: nyc_tlc_green	NYC Taxi & Limousine Commission - green taxi trip records	The green taxi trip records include fields capturing pick-up and drop-off dates/times, pick-up and drop...
ID: nyc_tlc_yellow	NYC Taxi & Limousine Commission - yellow taxi trip records	The yellow taxi trip records include fields capturing pick-up and drop-off dates/times, pick-up and drop...
ID: city_safety_newyork	New York City Safety Data	This dataset contains all New York City 311 service requests from 2010 to the present. It's stored in Parquet format and updated daily.

Screenshot 2: Detailed view of the 'Boston Safety Data' dataset. The 'Preview' section shows a table of data.

dataType	dataSubtype	dateTime	category	subcategory	status	address	latitude	longitude	source
Safety	311_All	10/19/2021 ...	Enforcement	Parking Enforcement	Open	523 E Fourth St	42.3345	-71.0438	Citizens C
Safety	311_All	10/19/2021 ...	Enforcement	Parking Enforcement	Open	507 E Broadway	42.3353	-71.0444	Citizens C
Safety	311_All	10/19/2021 ...	Enforcement	Parking Enforcement	Open	45 Thomas Street	42.3323	-71.047	Citizens C
Safety	311_All	10/19/2021 ...	Enforcement	Parking Enforcement	Open	INTERSECTION	42.3594	-71.0587	Citizens C
Safety	311_All	10/19/2021 ...	Enforcement	Parking Enforcement	Open	51-53 Wildwood Street	42.2833	-71.0885	Citizens C
Safety	311_All	10/19/2021 ...	Enforcement	Parking Enforcement	Open	15 Wildwood Street	42.2839	-71.0872	Citizens C
Safety	311_All	10/19/2021 ...	Enforcement	Parking Enforcement	Open	44 Wildwood Street	42.2832	-71.0881	Citizens C
Safety	311_All	10/19/2021 ...	Enforcement	Parking Enforcement	Open	701 E Sixth Street	42.333	-71.0323	Citizens C
Safety	311_All	10/19/2021 ...	Enforcement	Parking Enforcement	Open	17 Buswell Street	42.3478	-71.1052	Citizens C
Safety	311_All	10/19/2021 ...	Enforcement	Parking Enforcement	Open	18 Tuckerman Street	42.3315	-71.0546	Citizens C

The screenshot shows the Microsoft Azure Synapse Studio interface. On the left, the navigation bar includes Home, Data, Develop, Integrate, Monitor, and Manage. The main area is titled 'Data' and shows a 'Linked' workspace. A 'SQL script 1' tab is open, containing the following code:

```
-- This is auto-generated code
SELECT
    TOP_500*
FROM
    OPENROWSET(
        BULK 'https://pandemicdatalake.blob.core.windows.net/public/curated/city_safety_boston.parquet',
        FORMAT = 'parquet'
    ) AS [result];
```

The results pane shows a DataFrame with columns: id, updated, confirmed, confirmed_change, deaths, and death_change. The data is as follows:

id	updated	confirmed	confirmed_change	deaths	death_change
338995	2020-01-21T00:00:00Z	262	(NULL)	0	(NULL)
338996	2020-01-22T00:00:00Z	313	51	0	0
338997	2020-01-23T00:00:00Z	578	265	0	0

The status bar at the bottom indicates "00:00:12 Query executed successfully." The system tray shows the date as 09-06-2022.

Do right click on city_safty_boston → new notebook → load to DataFrame

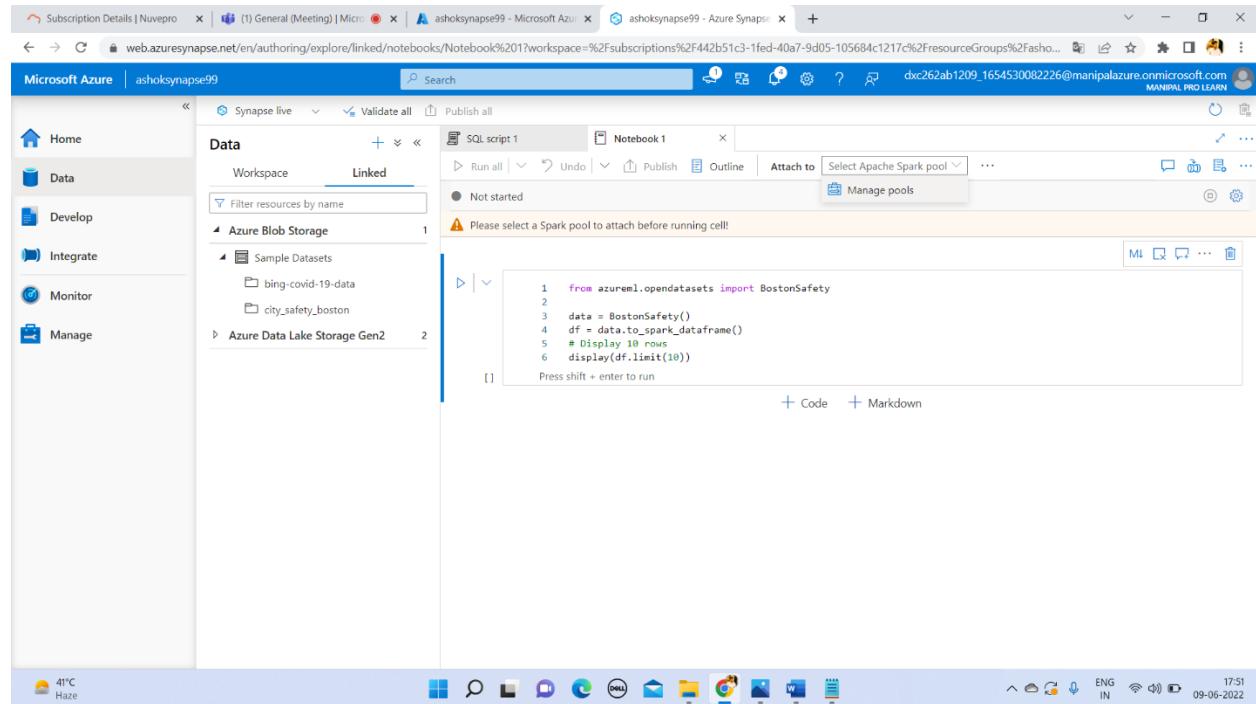
The screenshot shows the Microsoft Azure Synapse Studio interface with a new notebook titled 'Notebook 1'. The notebook contains the following Python code:

```
from azureml.opendatasets import BostonSafety
data = BostonSafety()
df = data.to_spark_dataframe()
# Display 10 rows
display(df.limit(10))
```

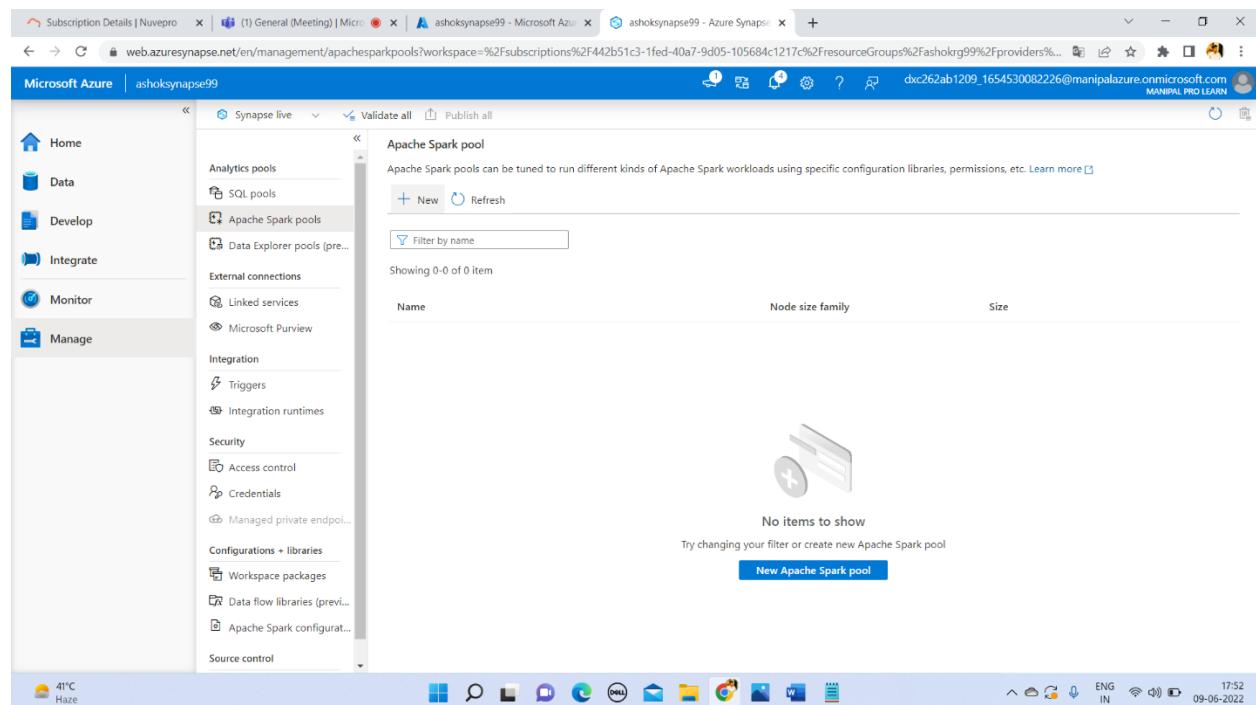
A warning message at the top of the notebook says: "Please select a Spark pool to attach before running cell!". The system tray at the bottom shows the date as 09-06-2022.

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

ANSWER:



Click on manage pools



Click on + New to create new apache spark pool

Subscription Details | Nuvapro | (1) General (Meeting) | Microsoft Azure | ashoksynapse99 - Microsoft Azure | ashoksynapse99 - Azure Synapse | +

Microsoft Azure | ashoksynapse99

New Apache Spark pool

Basics • Additional settings Tags Review + create

Create a 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 * ashoksparkpool9

Isolated compute * Enabled Disabled

Node size family * Memory Optimized

Node size * Small (4 vCores / 32 GB)

Autoscale * Enabled Disabled

Number of nodes * 3 10

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

Dynamically allocate executors * Enabled Disabled

Review + create **Next: Additional settings >** **Cancel**

41°C Haze

Subscription Details | Nuvapro | (1) General (Meeting) | Microsoft Azure | ashoksynapse99 - Microsoft Azure | ashoksynapse99 - Azure Synapse | +

Microsoft Azure | ashoksynapse99

New Apache Spark pool

Validation succeeded.

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

Product details

Azure Synapse Analytics Apache Spark pool by Microsoft **Est. cost per hour** Failed to fetch billing info
[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	ashokrg99
Apache Spark pool name	ashoksparkpool9
Isolated compute	Disabled
Node size family	Memory Optimized

Create **< Previous** **Download template for automation** **Cancel**

41°C Haze

The screenshot shows the Microsoft Azure portal interface. The left sidebar has 'ashoksynapse99' selected. The main area is titled 'Apache Spark pool' under 'Analytics pools'. It displays a single item: 'ashoksparkpool9', which is 'Memory Optimized' and 'Small (4 vCores / 32 GB) - 3 to 10 nodes'. A success message box is open, stating 'Successfully deployed ashoksparkpool9 (Apache Spark pool) to the workspace.' The bottom right corner shows system status: ENG IN, 17:55, 09-06-2022.

Spark pool `ashoksparkpool9` is created!

Now we can attach our notebook1 to this sparkpool and run them and can see the outputs.

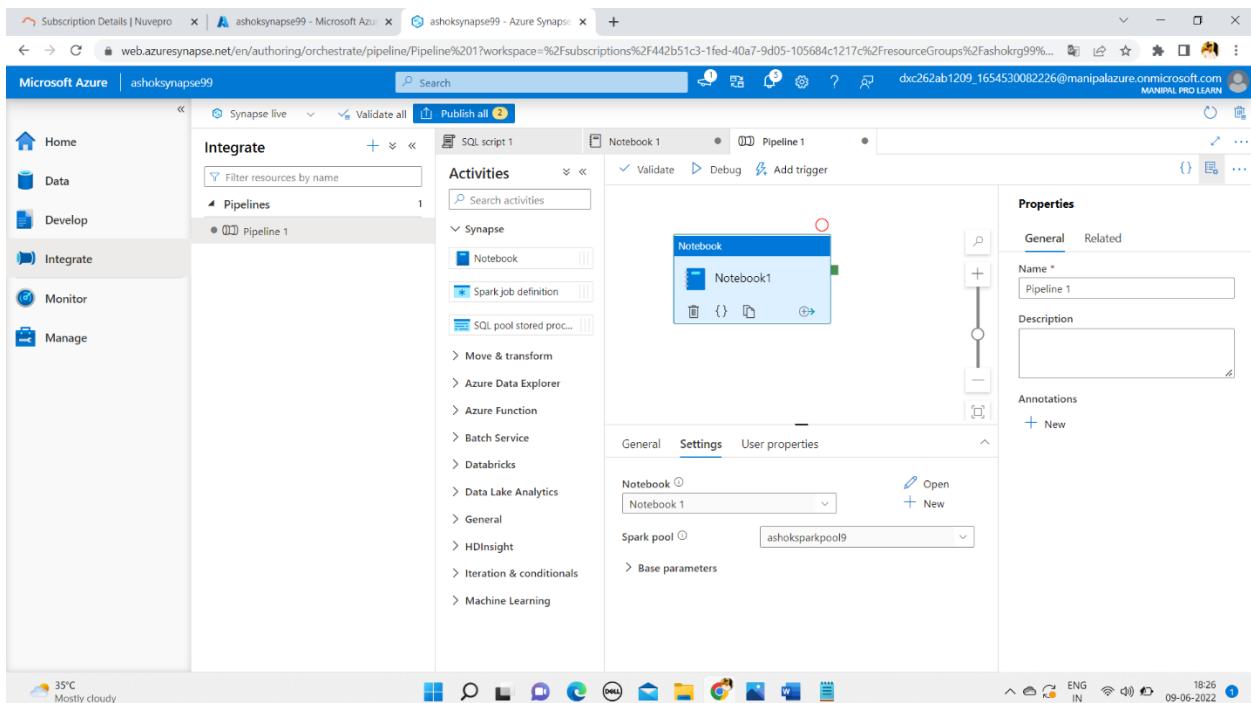
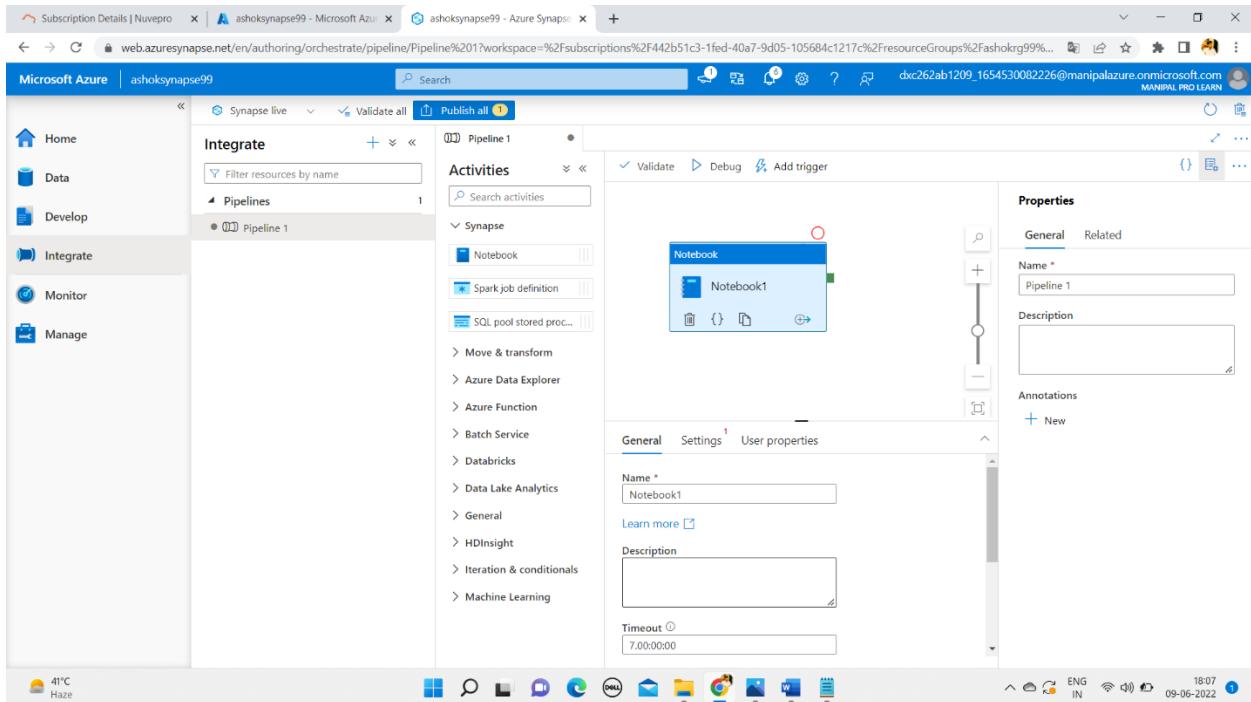
QUESTIONS 6. Explain the steps with screenshots how to create pipeline in AzureSynapse analytics?

ANSWER: click on Integrate → click on + → click on pipeline

The screenshot shows the Microsoft Azure portal interface for creating a pipeline in Azure Synapse Analytics. The left sidebar has 'Integrate' selected. In the main area, under 'Integrate', there is a 'Pipeline' section with a '+' button. A dropdown menu is open, showing options like 'Pipeline', 'Link connection (Preview)', 'Copy Data tool', 'Browse gallery', and 'Import from pipeline template'. Below this, a list of activities is shown: 'Synapse', 'Move & transform', 'Azure Data Explorer', 'Azure Function', 'Batch Service', 'Databricks', 'Data Lake Analytics', 'General', 'HDInsight', 'Iteration & conditionals', and 'Machine Learning'. On the right, the 'Properties' panel is open for 'Pipeline 1', showing fields for 'Name' (Pipeline 1), 'Description', and 'Annotations'. The bottom status bar shows the date and time as 09-06-2022 18:04.

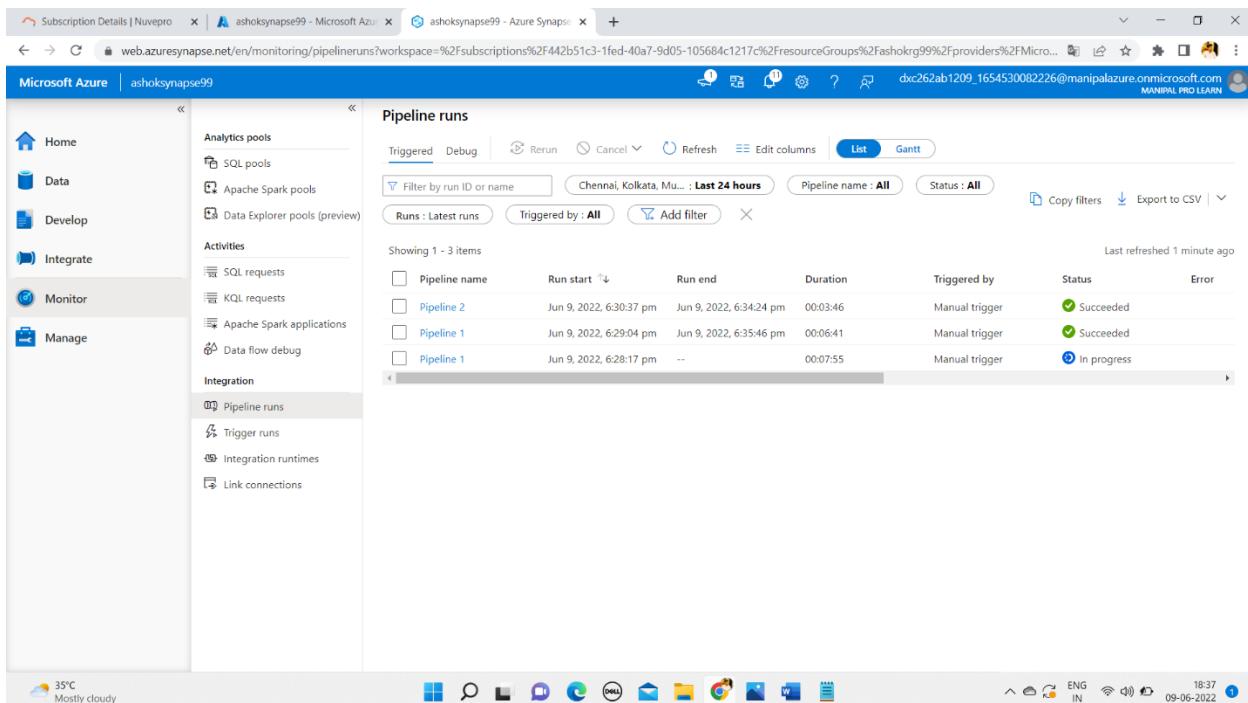
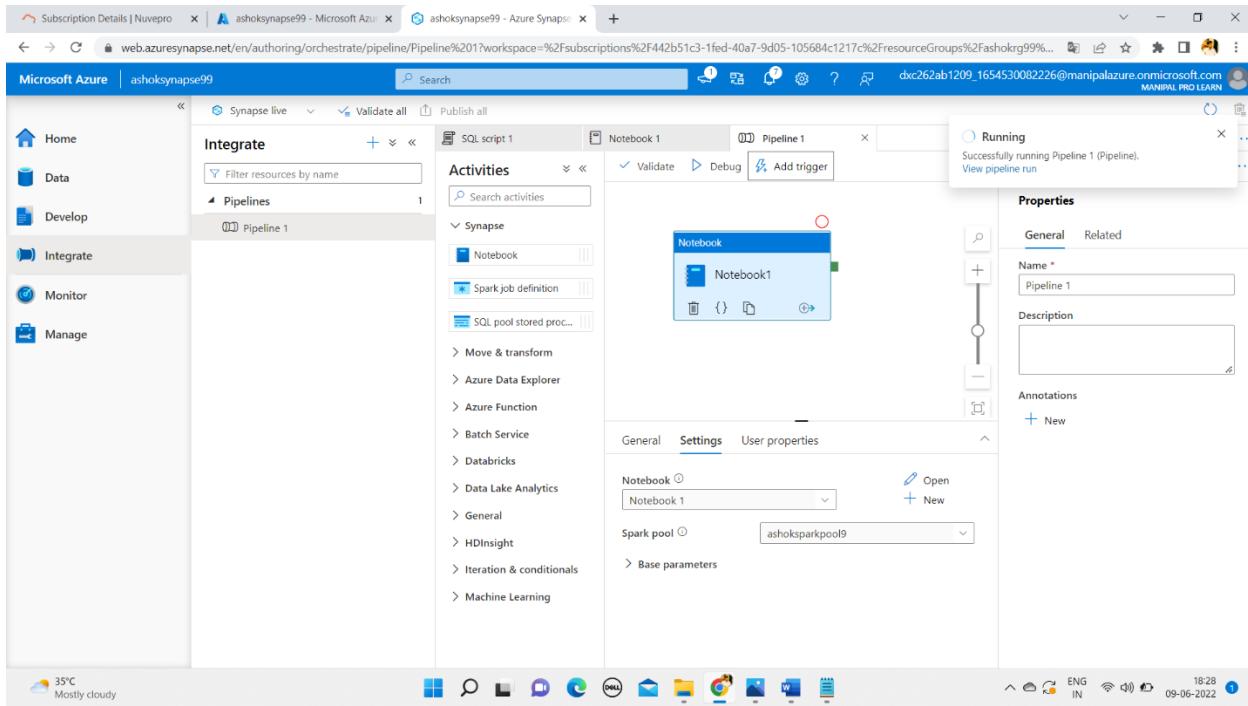
This screenshot shows the continuation of the pipeline creation process. The 'Activities' section is now active, and the '+' button is highlighted with the text 'Add new resource'. The rest of the interface is identical to the previous screenshot, including the list of activities and the properties panel for 'Pipeline 1'.

Expand synapse and drag notebook



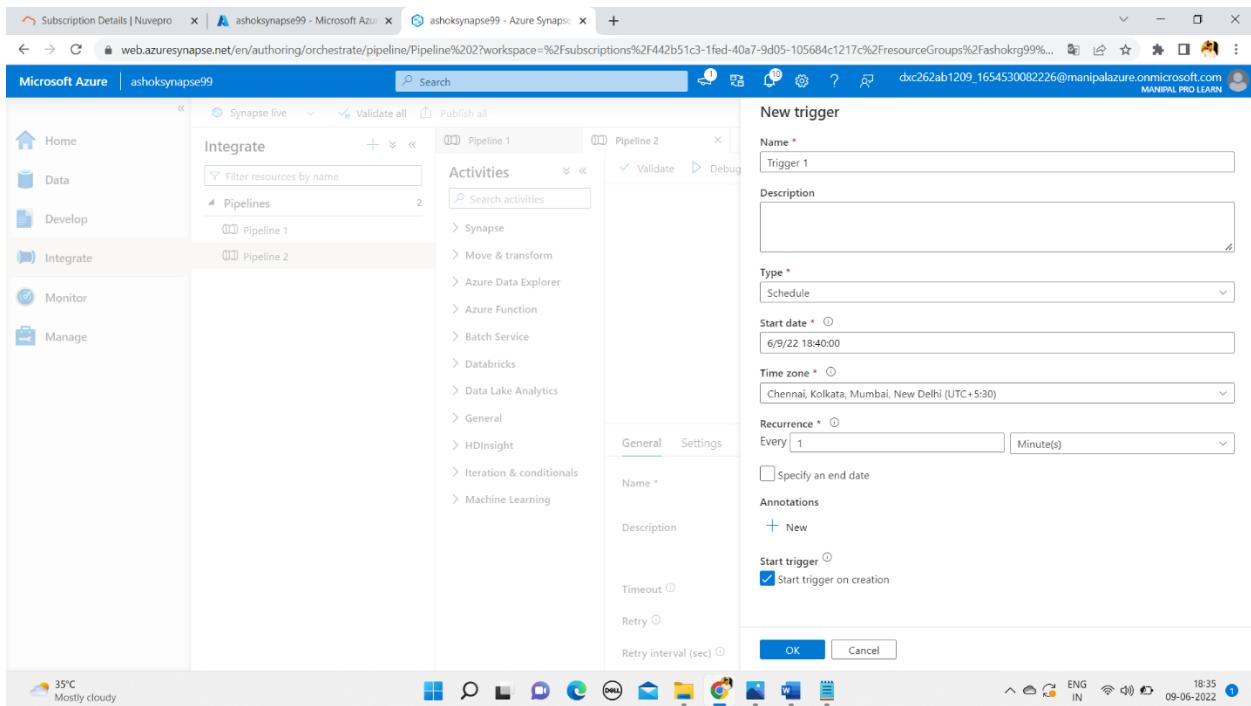
Click on publish all to publish the pipeline

Now click on add trigger to run pipeline.



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

ANSWER: Now click on new/edit trigger



Now at 18:40:00 pipeline will be automatically triggered in every minute.

Subscription Details | Nuvepro | ashoksynapse99 - Microsoft Azure | ashoksynapse99 - Azure Synapse | +

Microsoft Azure | ashoksynapse99

Pipeline runs

Triggered Debug Rerun Cancel Refresh Edit columns List Gantt

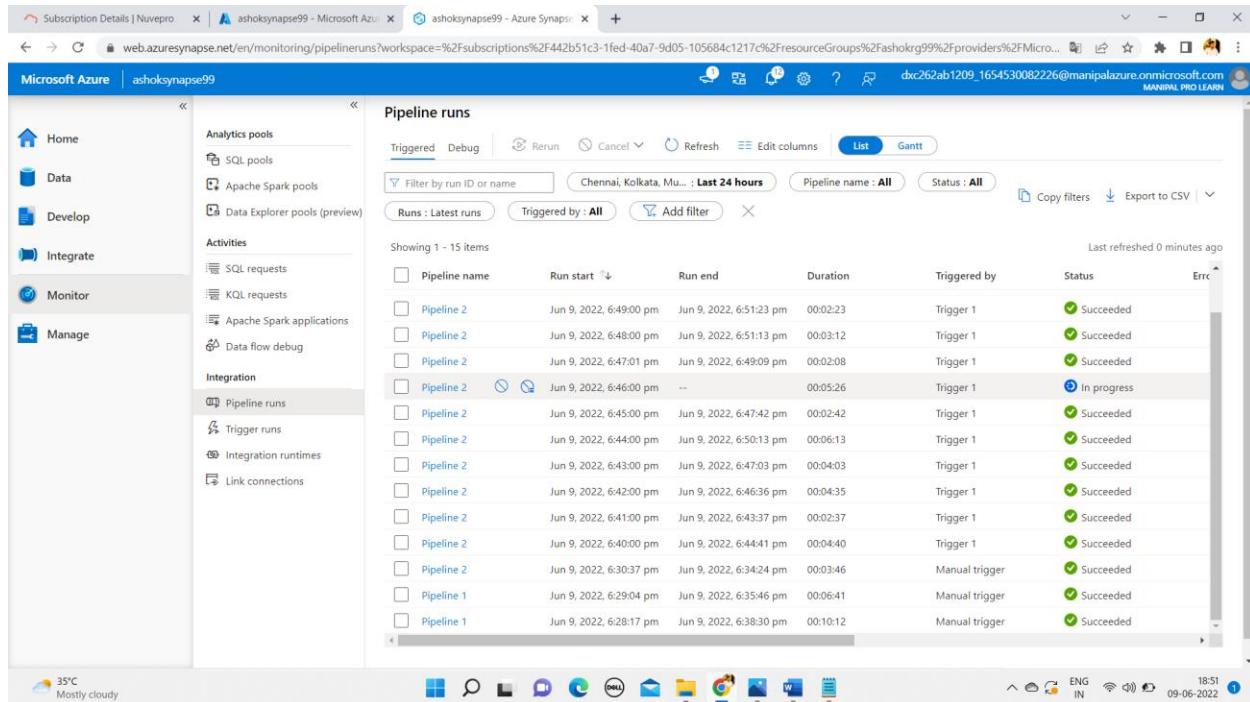
Filter by run ID or name Chennai, Kolkata, Mu... : Last 24 hours Pipeline name : All Status : All

Runs : Latest runs Triggered by : All Add filter

Last refreshed 0 minutes ago

Pipeline name	Run start	Run end	Duration	Triggered by	Status
Pipeline 2	Jun 9, 2022, 6:49:00 pm	Jun 9, 2022, 6:51:23 pm	00:02:23	Trigger 1	Succeeded
Pipeline 2	Jun 9, 2022, 6:48:00 pm	Jun 9, 2022, 6:51:13 pm	00:03:12	Trigger 1	Succeeded
Pipeline 2	Jun 9, 2022, 6:47:01 pm	Jun 9, 2022, 6:49:09 pm	00:02:08	Trigger 1	Succeeded
Pipeline 2	Jun 9, 2022, 6:46:00 pm	--	00:05:26	Trigger 1	In progress
Pipeline 2	Jun 9, 2022, 6:45:00 pm	Jun 9, 2022, 6:47:42 pm	00:02:42	Trigger 1	Succeeded
Pipeline 2	Jun 9, 2022, 6:44:00 pm	Jun 9, 2022, 6:50:13 pm	00:06:13	Trigger 1	Succeeded
Pipeline 2	Jun 9, 2022, 6:43:00 pm	Jun 9, 2022, 6:47:03 pm	00:04:03	Trigger 1	Succeeded
Pipeline 2	Jun 9, 2022, 6:42:00 pm	Jun 9, 2022, 6:46:36 pm	00:04:35	Trigger 1	Succeeded
Pipeline 2	Jun 9, 2022, 6:41:00 pm	Jun 9, 2022, 6:43:37 pm	00:02:37	Trigger 1	Succeeded
Pipeline 2	Jun 9, 2022, 6:40:00 pm	Jun 9, 2022, 6:44:41 pm	00:04:40	Trigger 1	Succeeded
Pipeline 2	Jun 9, 2022, 6:30:37 pm	Jun 9, 2022, 6:34:24 pm	00:03:46	Manual trigger	Succeeded
Pipeline 1	Jun 9, 2022, 6:29:04 pm	Jun 9, 2022, 6:35:46 pm	00:06:41	Manual trigger	Succeeded
Pipeline 1	Jun 9, 2022, 6:28:17 pm	Jun 9, 2022, 6:38:30 pm	00:10:12	Manual trigger	Succeeded

35°C Mostly cloudy 18:51 ENG IN 09-06-2022



Subscription Details | Nuvepro | ashoksynapse99 - Microsoft Azure | ashoksynapse99 - Azure Synapse | +

Microsoft Azure | ashoksynapse99

All pipeline runs > Pipeline 2 - Activity runs

Pipeline 2

List Gantt

Rerun Rerun from activity Rerun from failed activity Refresh Update pipeline

Notebook Notebook1

Activity runs

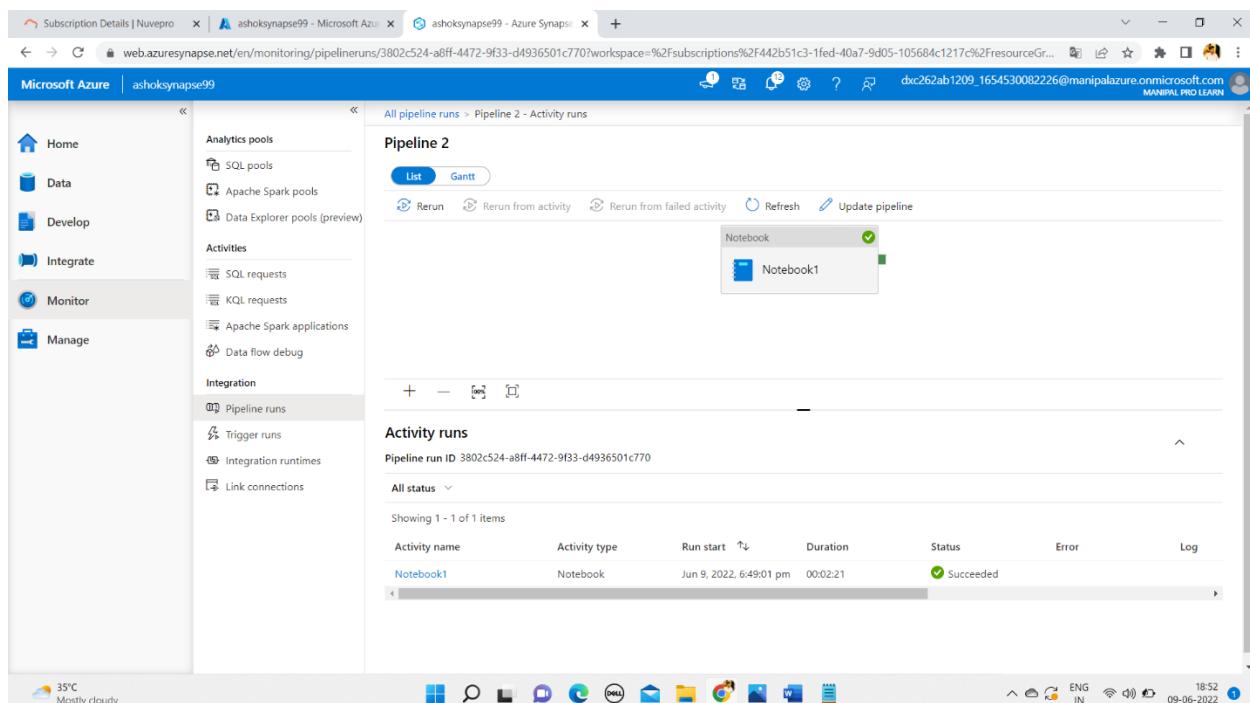
Pipeline run ID: 3802c524-a8ff-4472-9f33-d4936501c770

All status

Showing 1 - 1 of 1 items

Activity name	Activity type	Run start	Duration	Status	Error	Log
Notebook1	Notebook	Jun 9, 2022, 6:49:01 pm	00:02:21	Succeeded		

35°C Mostly cloudy 18:52 ENG IN 09-06-2022



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

ANSWER: go to databricks section on azure portal

The screenshot shows the Microsoft Azure portal search results for 'databricks'. The search bar at the top contains the query 'databricks'. Below the search bar, there are several sections:

- Services**: Shows 'Azure Databricks' as the top result, followed by 'Azure Database Migration Projects', 'Azure Database Migration Services', 'SQL databases', 'Azure Database for MySQL servers', 'Managed databases', 'Azure Database for MariaDB servers', and 'See all'.
- Marketplace**: Shows 'Azure Databricks', 'Unravel Image for Azure Databricks', 'Infoworks for Databricks', 'Lakehouse Monitor', 'Unravel for Azure Databricks', 'Trifactor for Azure', 'NASH Video Analytics on Azure', and 'See all'.
- Documentation**: Shows links to 'What is Azure Databricks? | Microsoft Docs', 'Azure Databricks documentation | Microsoft Docs', 'What is Databricks Machine Learning? | Microsoft Docs', 'Databricks Runtime 10.4 LTS for Machine Learning - Azure Databricks...', 'Databricks Utilities - Azure Databricks | Microsoft Docs', and 'Databases and tables - Azure Databricks | Microsoft Docs'.

At the bottom of the page, there are toolbars for Microsoft Learn, Azure Monitor, Microsoft Defender for Cloud, and Cost Management, along with system status indicators like weather (35°C, Mostly cloudy) and date/time (09-06-2022, 18:59).

The screenshot shows the Microsoft Azure portal resource list for 'Azure Databricks'. The title bar says 'Subscription Details | Nuvepro' and 'Azure Databricks - Microsoft Azure'. The search bar at the top contains 'Search resources, services, and docs (G+)'. The main area displays the following information:

- Home > Azure Databricks
- Actions: + Create, Manage view, Refresh, Export to CSV, Open query, Assign tags.
- Filters: Filter for any field..., Subscription == all, Resource group == all, Location == all, Add filter.
- Grouping: No grouping, List view.
- Content: A large 'No azure databricks services to display' message with a 'Create azure databricks service' button.

At the bottom of the page, there are toolbars for Microsoft Learn, Azure Monitor, Microsoft Defender for Cloud, and Cost Management, along with system status indicators like weather (35°C, Mostly cloudy) and date/time (09-06-2022, 19:00).

Subscription Details | Nuvepro x Create an Azure Databricks workspace x +

portal.azure.com/#create/Microsoft.Databricks

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 * ashokrg99 Create new

Instance Details

Workspace name * dxcdatabricks99 Region * East US Pricing Tier * Trial (Premium - 14-Days Free DBUs)



Subscription Details | Nuvepro x Create an Azure Databricks workspace x +

portal.azure.com/#create/Microsoft.Databricks

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

Home > Azure Databricks > Create an Azure Databricks workspace ...

Validation Succeeded

Basics Networking Advanced Tags Review + create

Summary

Workspace name: dxcdatabricks99
Subscription: Azure-DXC262AB12Lab
Resource group: ashokrg99
Region: East US
Pricing Tier: trial

Networking

Deploy Azure Databricks workspace with Secure Cluster Connectivity (No Public IP): No
Deploy Azure Databricks workspace in your own Virtual Network (VNet): No

Advanced

Enable Infrastructure Encryption: No

Create < Previous Download a template for automation

35°C Mostly cloudy ENG IN 19:01 09-06-2022

Subscription Details | Nuvepro x ashokrg99_dxdatabricks99 - Microsoft Azure

portal.azure.com/#view/HubExtension/DeploymentDetailsBlade/~/overview/id%2Fsubscriptions%2F442b51c3-1fed-40a7-9d05-105684c1217c%2FresourceGroups%2Fashokrg99... MANAGER PRO LEARN (MANAGER)

Home > ashokrg99_dxdatabricks99 | Overview

Deployment

Search (Ctrl+ /) Delete Cancel Redeploy Refresh

We'd love your feedback!

Your deployment is complete

Deployment name: ashokrg99_dxdatabricks99
Subscription: Azure-DXC262AB12Lab
Resource group: ashokrg99

Start time: 6/9/2022, 7:01:40 PM
Correlation ID: 1bb51a90-928e-4630-86be-aa506aae3647

Deployment details (Download)
Next steps
Go to resource

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
Azure experts are service provider partners who can help manage your assets on Azure and be your first line of support.
Find an Azure expert >

35°C Mostly cloudy

ENG IN 19:04 09-06-2022

Subscription Details | Nuvepro x dx databricks99 - Microsoft Azure

portal.azure.com/#@manipalazure.onmicrosoft.com/resource/subscriptions/442b51c3-1fed-40a7-9d05-105684c1217c/resourceGroups/ashokrg99/providers/Microsoft.Databricks/... MANAGER PRO LEARN (MANAGER)

Home > ashokrg99_dxdatabricks99 >

dx databricks99 - Azure Databricks Service

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
Managed Resource Group : databricks-rg-dxdatabricks99-7usunjadrcnos
Resource group : ashokrg99 URL : https://adb-7504800665053194.14.azuredatabricks.net
Location : East US Pricing Tier : Trial (Premium - 14-Days Free DBUs)
Subscription : Azure-DXC262AB12Lab
Subscription ID : 442b51c3-1fed-40a7-9d05-105684c1217c
Tags (edit) : Click here to add tags

JSON View

Launch Workspace

Upgrade to Premium

Documentation Getting Started Import Data from File Import Data from Azure Storage

35°C Mostly cloudy

ENG IN 19:05 09-06-2022

Subscription Details | Nuvepro x | adb-7504800665053194 - Microsoft Azure x | Databricks x +

Microsoft Azure | Databricks Portal dxc262ab1209_1654530082226@manipalazure.onmicrosoft.com

Subscriptions                                      

Data Science & Engineering

Notebook  Create a new notebook for querying, data processing, and machine learning...  Notebook  Data import  Quickly import data, preview its schema, create a table, and query it...  Notebook  Browse files 

Guide: Quickstart tutorial  Spin up a cluster, run queries on preloaded data, and display results in 5 minutes.  Guide: Quickstart tutorial  Partner Connect  FiveTran+ dbt  Tableau, Power BI  View all partners 

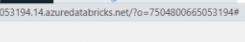
Recent items

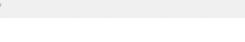
Name	Last viewed
There are no recent items yet.	

Documentation  Get started guide  This tutorial gets you going with Azure Databricks Data Science & Engineering Documentation  Release notes  Runtime release notes  Azure Databricks preview releases  Release notes  Blog posts  Building ETL pipelines for the cybersecurity lakehouse with Delta Live Tables  June 3, 2022 

Cloud services  

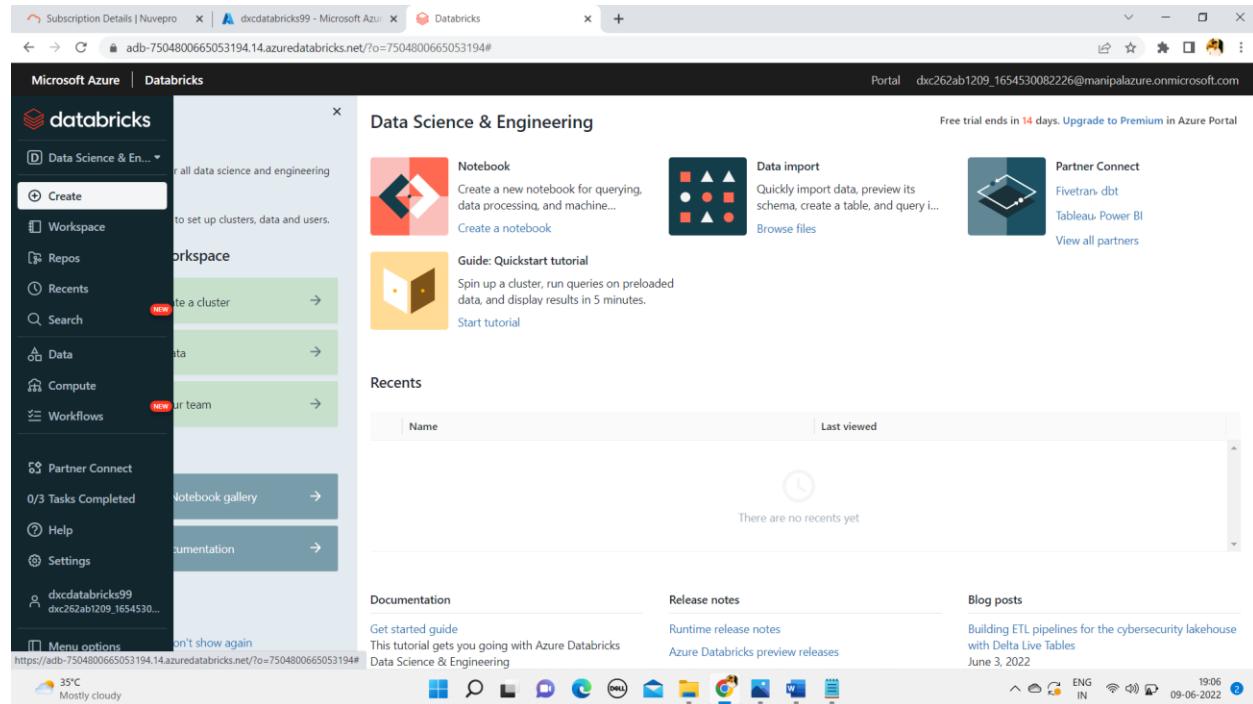
System status  ENG IN  19:06 09-06-2022 

Menu options             

35°C Mostly cloudy             

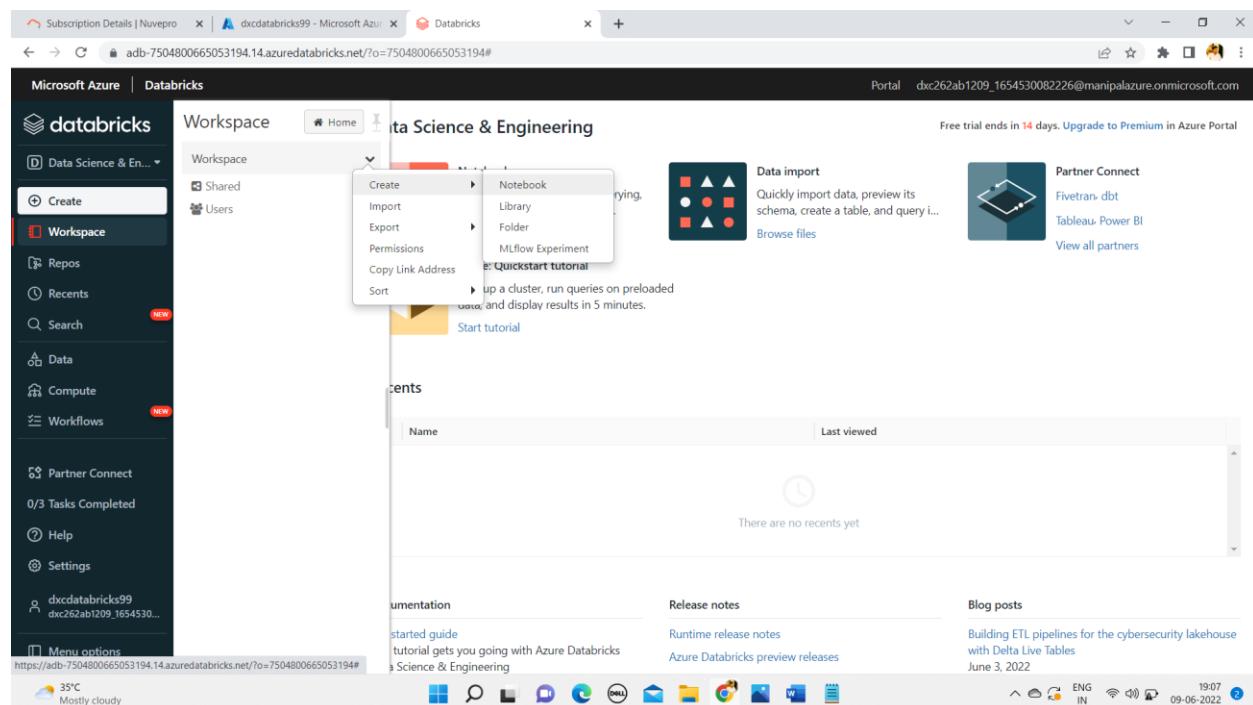
QUESTIONS 9. Explain the steps with screenshots how to create notebooks in Databricks ?

ANSWER:

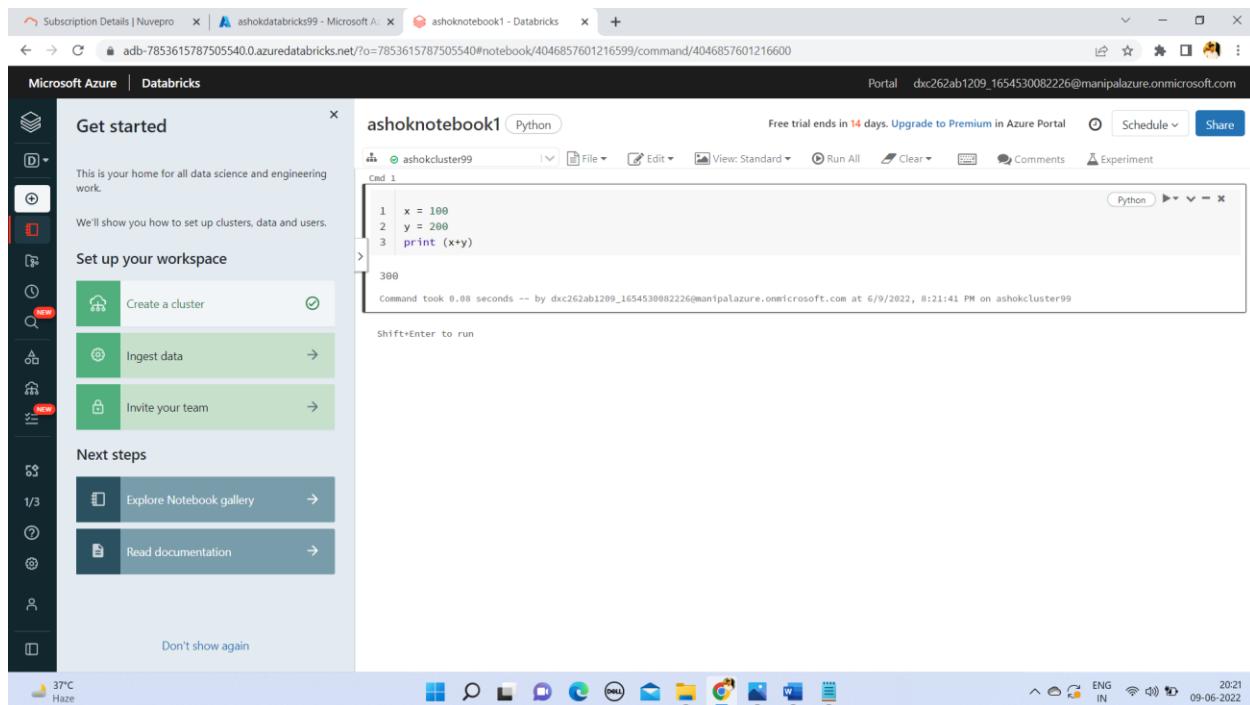
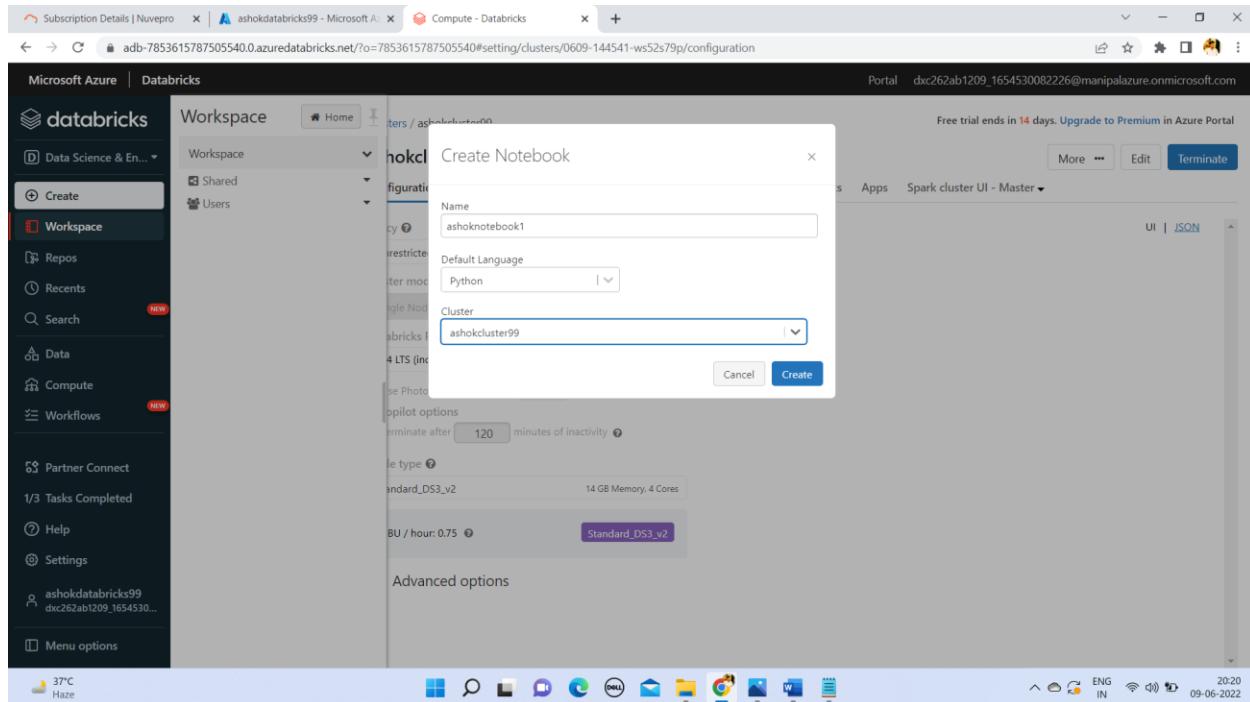


The screenshot shows the Microsoft Azure Databricks workspace. On the left, there's a sidebar with navigation links such as Data Science & Engineering, Create, Workspace, Repos, Recents, Search, Data, Compute, Workflows, Partner Connect, Help, Settings, and a user profile. The main area is titled "Data Science & Engineering". It features several cards: "Notebook" (Create a new notebook for querying, data processing, and machine learning), "Data import" (Quickly import data, preview its schema, create a table, and query it), "Guide: Quickstart tutorial" (Spin up a cluster, run queries on preloaded data, and display results in 5 minutes), and "Recents" (a list that currently says "There are no recents yet"). Below these are sections for Documentation, Release notes, and Blog posts.

Go to workspace → click on create → click on notebook

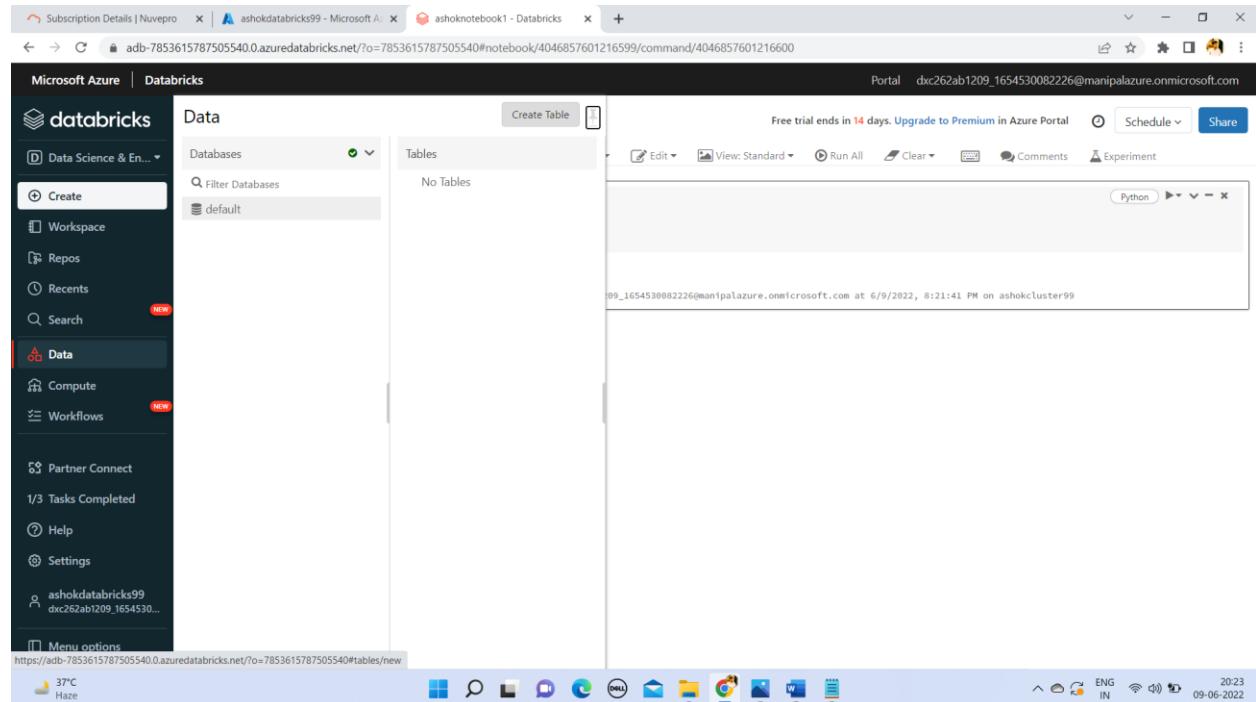


This screenshot is similar to the previous one but focuses on the "Create" menu. The "Create" option in the sidebar is highlighted, and a dropdown menu is open, showing "Notebook" as the selected item. The rest of the interface is identical to the first screenshot, displaying the Data Science & Engineering dashboard with its various components and links.



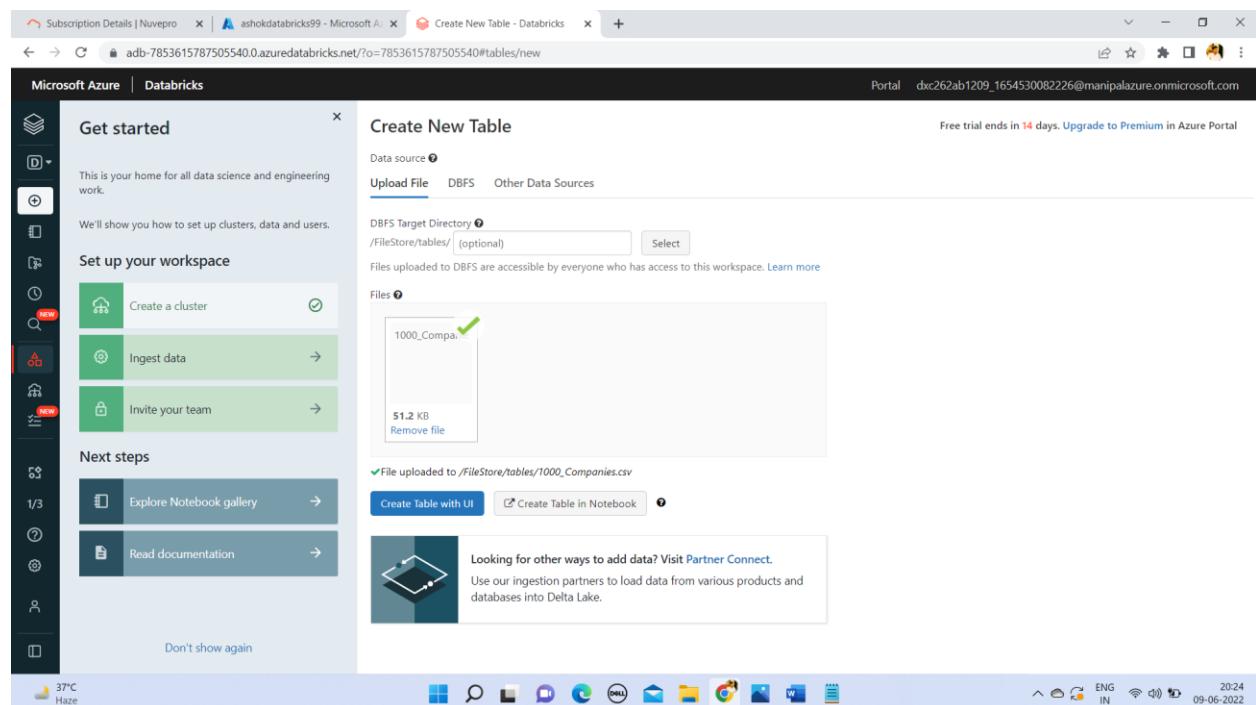
QUESTIONS 10. Explain the steps with screenshots how to insert data into databricks notebook & display the result?

ANSWER:



The screenshot shows the Microsoft Azure Databricks interface. On the left, the sidebar has sections like 'Data Science & Engineering', 'Workspace', 'Repos', 'Recents', 'Search', 'Data' (which is selected), 'Compute', 'Workflows', 'Partner Connect', 'Tasks Completed', 'Help', 'Settings', and a user profile. The main area is titled 'Data' and shows 'Databases' and 'Tables'. Under 'Tables', it says 'No Tables'. There is a 'Create Table' button at the top right of the table list. The status bar at the bottom indicates 'Free trial ends in 14 days. Upgrade to Premium in Azure Portal'.

Upload file from your local storage and click on create table in notebook



The screenshot shows the 'Create New Table' dialog box. It has tabs for 'Data source' (DBFS is selected), 'Upload File', and 'Other Data Sources'. Under 'Upload File', there is a 'DBFS Target Directory' field containing '/FileStore/tables/' and a 'Select' button. Below it, a file named '1000_Compan...' is listed with a size of '51.2 KB' and a 'Remove file' link. A message at the bottom says 'File uploaded to /FileStore/tables/1000_Companies.csv'. At the bottom of the dialog, there are buttons for 'Create Table with UI' and 'Create Table in Notebook'. A note on the right says 'Looking for other ways to add data? Visit Partner Connect.' and 'Use our ingestion partners to load data from various products and databases into Delta Lake.' The status bar at the bottom indicates 'Free trial ends in 14 days. Upgrade to Premium in Azure Portal'.

Subscription Details | Nuvepro | ashokdatabricks99 - Microsoft A | 2022-06-09 - DBFS Example - D | +

adb-7853615787505540.azuredatabricks.net/?o=7853615787505540#notebook/4046857601216601/command/4046857601216606

Microsoft Azure | Databricks Portal dxc26ab1209_1654530082226@manipalazure.onmicrosoft.com

Get started

This is your home for all data science and engineering work.

We'll show you how to set up clusters, data and users.

Set up your workspace

- Create a cluster
- Ingest data
- Invite your team

Next steps

- Explore Notebook gallery
- Read documentation

Attached cluster: ashokcluster99
14.00 GB | 4 Cores | DBR 10.4 LTS | Spark 3.2.1 | Scala 2.12
Detach Restart Cluster Detach & Re-attach Spark UI Driver logs Terminal

```

16
17 display(df)

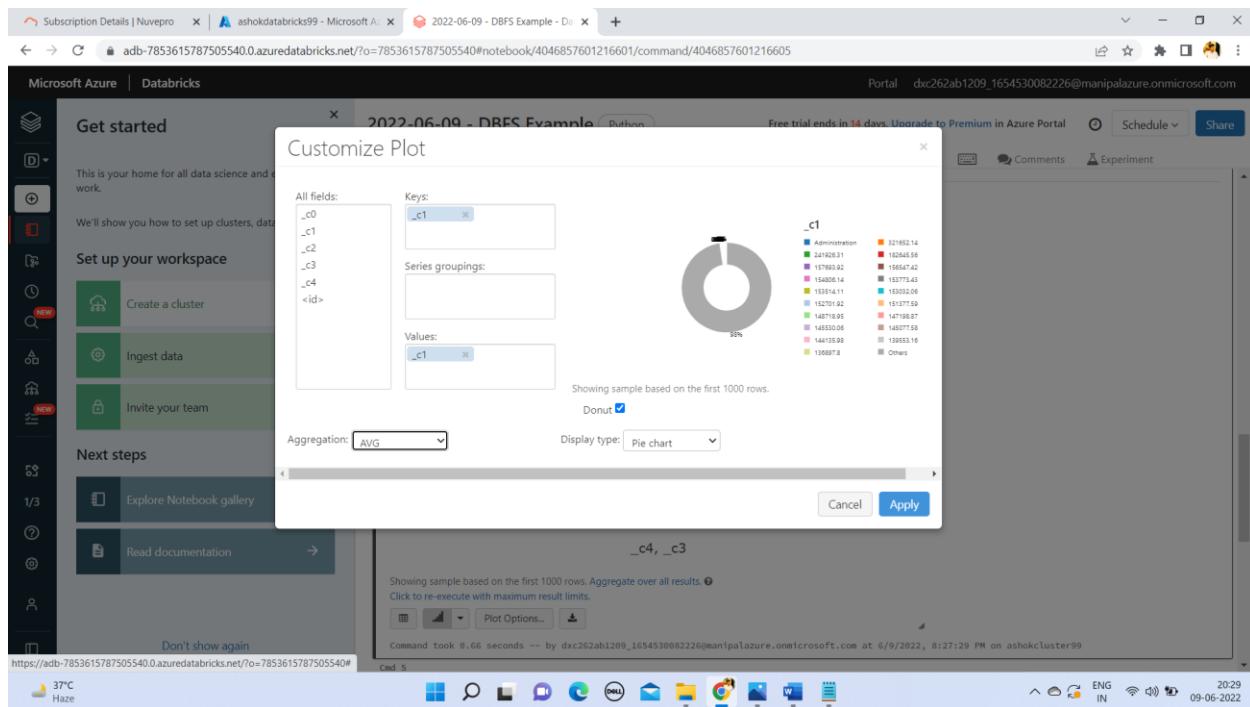
Cmd 3
1 # Create a view or table
2
3 temp_table_name = "Companies_csv"
4
5 df.createOrReplaceTempView(temp_table_name)

Cmd 4
1 %sql
2
3 /* Query the created temp table in a SQL cell */
4
5 select * from 'Companies_csv'

Cmd 5
1 # With this registered as a temp view, it will only be available to this particular notebook. If you'd like other users to be able to query this table, you can also create a table from the DataFrame.
2 # Once saved, this table will persist across cluster restarts as well as allow various users across different notebooks to query this data.
3 # To do so, change your table name and uncomment the bottom line.

```

Python 37°C Haze 2027 09-06-2022



Subscription Details | Nuvepro | ashokdatabricks99 - Microsoft A... | 2022-06-09 - DBFS Example - D... | +

adb-7853615787505540.azuredatabricks.net/?o=7853615787505540#notebook/4046857601216601/command/4046857601216605

Microsoft Azure | Databricks

Get started

This is your home for all data science and engineering work.

We'll show you how to set up clusters, data and users.

Set up your workspace

- Create a cluster
- Ingest data
- Invite your team

Next steps

- Explore Notebook gallery
- Read documentation

Don't show again

2022-06-09 - DBFS Example (Python)

Free trial ends in 14 days. Upgrade to Premium in Azure Portal

ashokcluster99

File View Standard Run All Clear Comments Experiment

SQL

```
%sql
/* Query the created temp table in a SQL cell */
select * from `Companies_csv`
```

(1) Spark Jobs

Chart Data Profile

Showing sample based on the first 1000 rows. Aggregate over all results. Click to re-execute with maximum result limits.

Plot Options

Command took 0.66 seconds -- by dxc262ab1209_1654530082226@manipalazure.onmicrosoft.com at 6/9/2022, 8:27:29 PM on ashokcluster99

Cmd 5

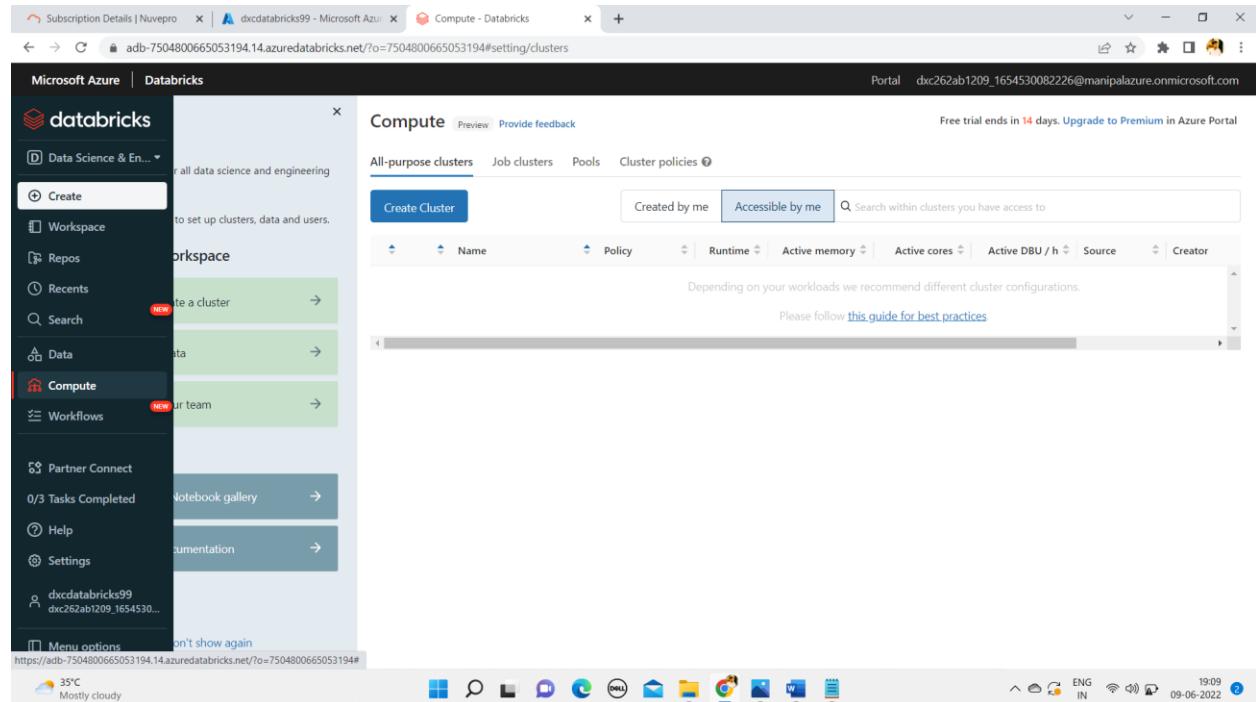
```
# With this registered as a temp view, it will only be available to this particular notebook. If you'd like other users to be able to query this table, you can also create a table from the DataFrame.
# Once saved, this table will persist across cluster restarts as well as allow various users across different notebooks to query this data.
```

37°C Haze

ENG IN 20:29 09-06-2022

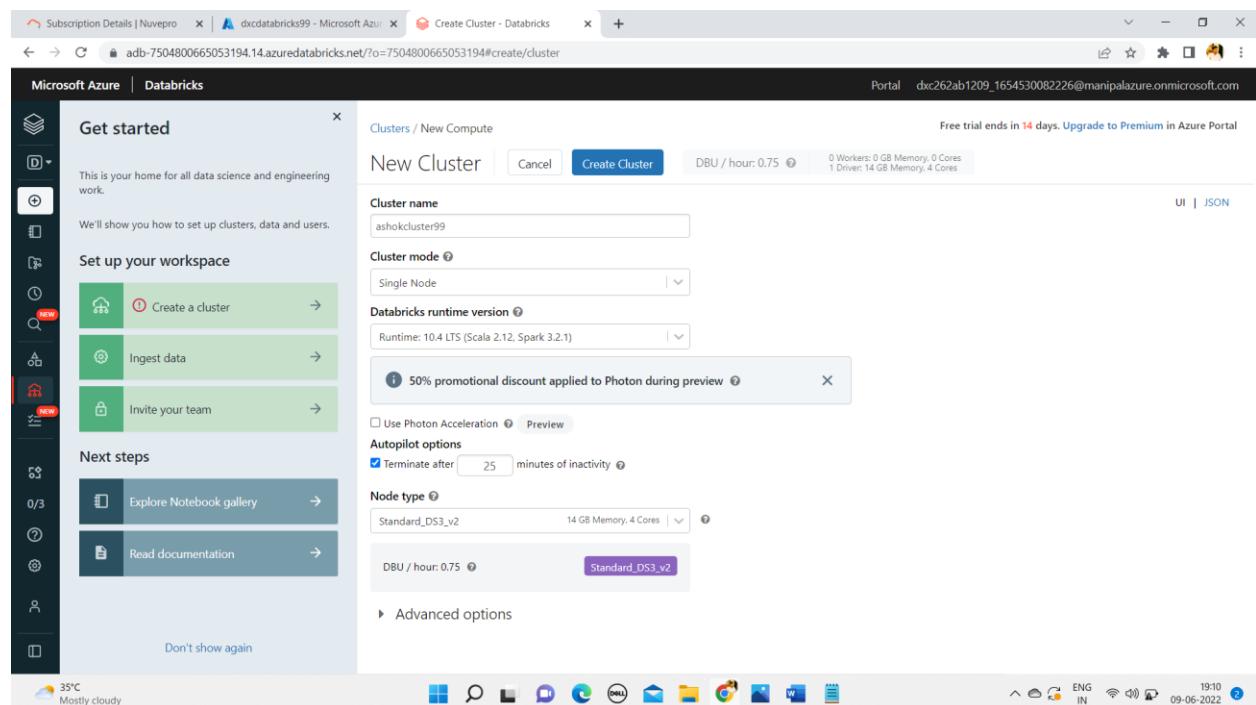
QUESTIONS 11. Explain the steps with screenshots how to create cluster in databricks ?

ANSWER:



The screenshot shows the Microsoft Azure Databricks Compute interface. On the left, there's a sidebar with various options like Data Science & Engineering, Workspace, Repos, Recents, Search, Data, Compute (which is selected), Workflows, Partner Connect, Help, Settings, and a user profile. The main area is titled 'Compute' and has tabs for 'All-purpose clusters', 'Job clusters', 'Pools', and 'Cluster policies'. A prominent blue button labeled 'Create Cluster' is visible. Below it is a table with columns for Name, Policy, Runtime, Active memory, Active cores, Active DBU / h, Source, and Creator. A note at the bottom says 'Depending on your workloads we recommend different cluster configurations.' and links to 'this guide for best practices'.

Click on compute then create cluster



The screenshot shows the 'Create Cluster' page. The left sidebar is identical to the previous one. The main area is titled 'Clusters / New Compute' and has a sub-section 'New Cluster'. It includes fields for 'Cluster name' (set to 'ashokcluster99'), 'Cluster mode' (set to 'Single Node'), 'Databricks runtime version' (set to 'Runtime: 10.4 LTS (Scala 2.12, Spark 3.2.1)'), and a note about a '50% promotional discount applied to Photon during preview'. There are also sections for 'Autopilot options' (with a checked checkbox for 'Terminate after 25 minutes of inactivity') and 'Node type' (set to 'Standard_DS3_v2'). At the bottom, there's a link to 'Advanced options' and a note about 'Photon Acceleration'.

This screenshot shows the 'Get started' page for a Databricks cluster named 'ashokcluster99'. The 'Configuration' tab is active. Key settings visible include:

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

The page also includes sections for 'Set up your workspace' (Create a cluster, Ingest data, Invite your team) and 'Next steps' (Explore Notebook gallery, Read documentation). A 'Don't show again' button is at the bottom.

This screenshot shows the 'Compute' page in the Microsoft Azure Databricks interface. The 'All-purpose clusters' section is displayed, listing the single cluster 'ashokcluster99' created earlier. The cluster details are as follows:

Name	Policy	Runtime	Active memory	Active cores	Active DBU / h	Source	Creator
ashokcluster99	-	10.4	14 GB	4 cores	0.75	UI	dxc262ab1

The page also includes sections for 'Get started' (Create a cluster, Ingest data, Invite your team) and 'Next steps' (Explore Notebook gallery, Read documentation). A 'Don't show again' button is at the bottom.

Cluster is created!

Thank you!

Ashok kumar