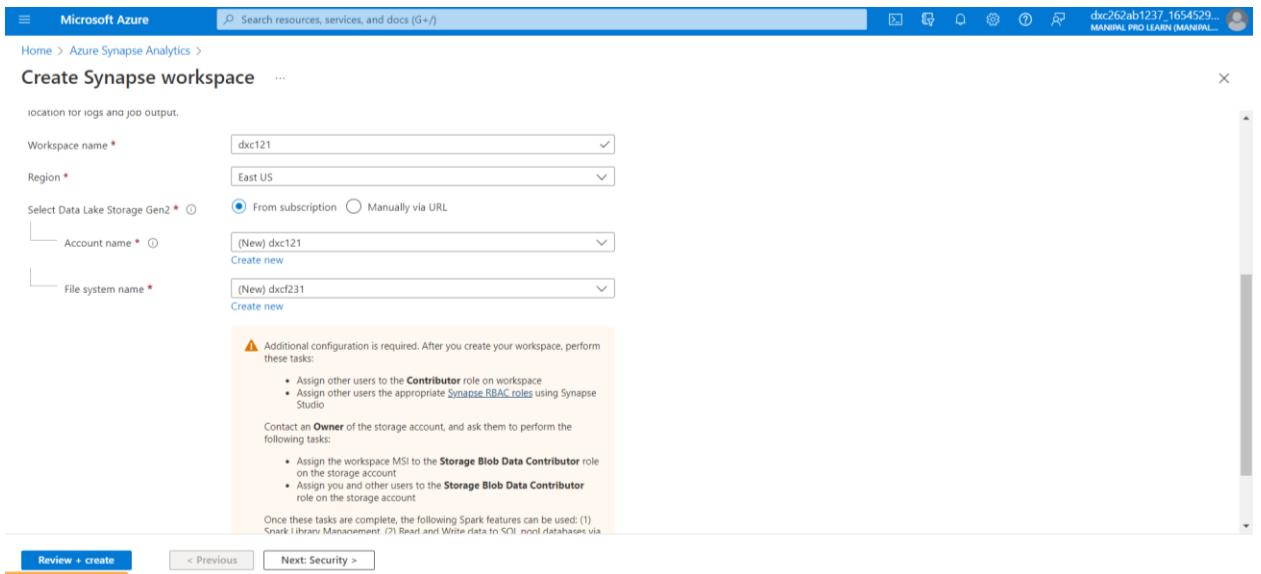


Name: MAHALAKSHMI S
Batch id:DXC-262-Analytics-B12-Azure
Domain: AZURE
Date:09.06.2022

ASSIGNMENT 9

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

Go to resources and search on Synapse ans then select AzureSynapse analytics



The screenshot shows the 'Create Synapse workspace' page in the Microsoft Azure portal. The workspace name is set to 'dxc121'. The region is 'East US'. Under 'Select Data Lake Storage Gen2', the 'From subscription' option is selected. A warning message box is displayed, stating: 'Additional configuration is required. After you create your workspace, perform these tasks: Assign other users to the Contributor role on workspace; Assign other users the appropriate Synapse RBAC roles using Synapse Studio.' Below this, it says: 'Contact an Owner of the storage account, and ask them to perform the following tasks: Assign the workspace MSI to the Storage Blob Data Contributor role on the storage account; Assign you and other users to the Storage Blob Data Contributor role on the storage account.' At the bottom, there are 'Review + create' and 'Next: Security >' buttons.

Add the name and select region and go to security

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

Home > Azure Synapse Analytics > Create Synapse workspace ...

* 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 • >Password and confirm password must match.

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. The selected Data Lake Storage Gen2 account does not restrict network access using any network access rules, or you selected a storage account manually via URL under Basics tab. [Learn more](#)

Review + create < Previous Next: Networking >

After Security create the workspace

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

Home > Azure Synapse Analytics > Create Synapse workspace ...

* Basics * Security Networking Tags Review + create

Tags are name/value pairs that enable you to categorize resources and view consolidated billing by applying the same tag to multiple resources and resource groups. [Learn more](#)

Note that if you create tags and then change resource settings on other tabs, your tags will be automatically updated.

Name	Value	Resource
<input type="text"/>	<input type="text"/> :	Synapse workspace

Save password?

Username: dxc121

Password: Passwords are saved in your Google Account so you can use them on any device

Review + create < Previous Next: Review + create >

Click review+create to create

The screenshot shows the Microsoft Azure 'Create Synapse workspace' wizard. At the top, there's a validation message: 'Validation succeeded'. Below it, the 'Review + create' tab is selected. The 'Product Details' section shows an Azure Synapse Analytics workspace by Microsoft, estimated cost per TB, and terms of use/privacy policy. The 'Terms' section contains legal text about agreeing to terms and privacy statements. The 'Basics' section shows subscription (Azure-DXC262AB12Lab), resource group (new_dxc213), and region (East US). At the bottom, there are 'Create', 'Previous', 'Next >', and 'Download a template for automation' buttons.

The deployment is completed and go to resource group.

The screenshot shows the Microsoft Azure 'Overview' page for a deployment named 'Microsoft.Azure.SynapseAnalytics-20220609100327'. It displays deployment details: name, subscription (Azure-DXC262AB12Lab), start time (6/9/2022, 10:07:38 AM), correlation ID, and resource group (dxc213). Below this, there are sections for 'Deployment details' (with a download link) and 'Next steps' (with a 'Go to resource group' button). On the right side, there are promotional cards for 'Cost Management', 'Microsoft Defender for Cloud', 'Free Microsoft tutorials', and 'Work with an expert'.

Microsoft Azure

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

Home > Microsoft.Azure.SynapseAnalytics-20220609100327 >

dxc213 Resource group

Overview

Essentials

Subscription (move) : Azure-DXC262AB12Lab
Subscription ID : 378cccd24-636d-4bed-8bde-fc66041e3b87
Tags (edit) : Click here to add tags

Deployments : 1Succeeded
Location : East US

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
dxc121	Storage account	East US
dxc121	Synapse workspace	East US

< Previous Page 1 of 1 Next >

No grouping List view

Give feedback

Finally it create the Azure synapse analytics.

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

Click on the AzureSynapse analytics and then open.

Microsoft Azure

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

Home > Microsoft.Azure.SynapseAnalytics-20220609100327 >

dxc213 Resource group

Overview

Essentials

Subscription (move) : Azure-DXC262AB12Lab
Subscription ID : 378cccd24-636d-4bed-8bde-fc66041e3b87
Tags (edit) : Click here to add tags

Deployments : 1Succeeded
Location : East US

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
dxc121	Storage account	East US
dxc121	Synapse workspace	East US

< Previous Page 1 of 1 Next >

No grouping List view

Give feedback

Open the SQL pool.

Microsoft Azure

Search resources, services, and docs (G+/-)

Home > Microsoft.Azure.SynapseAnalytics-20220609100327 > dxc213 >

dxc121 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

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

Essentials

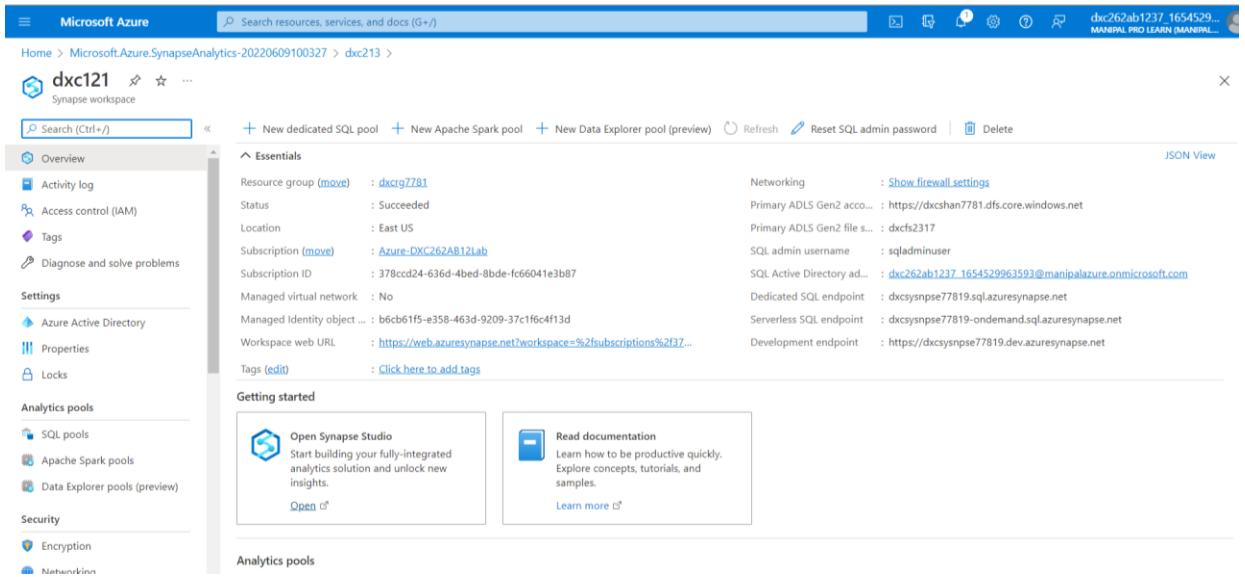
Resource group (move) : dxcrg7781 Status : Succeeded Location : East US Subscription (move) : Azure-DXC26AB12Lab Subscription ID : 378cd24-636d-4bed-8bde-fc66041e3b87 Managed virtual network : No Managed identity object id : b6cb61f5-e358-463d-9209-37c1f6c4f13d Workspace web URL : <https://web.azuresynapse.net/workspace=%2bsubscriptions%2f37...> Tags (edit) : Click here to add tags

Networking Primary ADLS Gen2 account : https://dxcshan7781.dfs.core.windows.net Primary ADLS Gen2 file system : dxcfs2317 SQL admin username : sqadminuser SQL Active Directory administrator : dxc26ab1237_1654529963593@manipalazure.onmicrosoft.com Dedicated SQL endpoint : dxcsynapse77819.sql.azuresynapse.net Serverless SQL endpoint : dxcsynapse77819-ondemand.sql.azuresynapse.net Development endpoint : <https://dxcsynapse77819.dev.azuresynapse.net>

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



It will show workspace.

Microsoft Azure | dxcsynapse77819

We use optional cookies to provide a better experience. [Learn more](#)

Accept Reject More options

Synapse live Validate all Publish all

Data

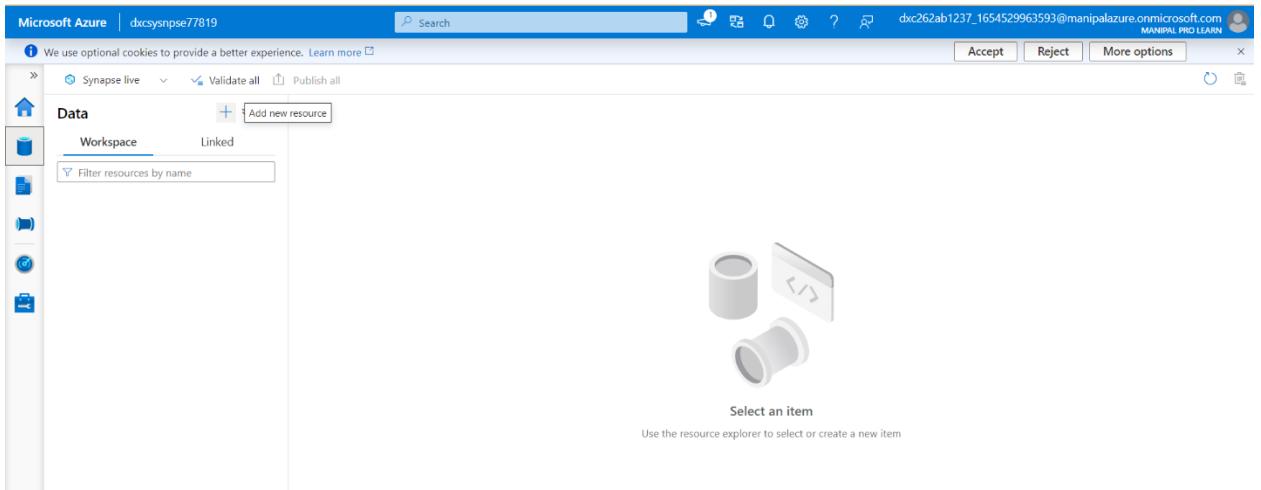
Workspace Linked

Add new resource

Filter resources by name

Select an item

Use the resource explorer to select or create a new item



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

After creating the AzureSynapse analytics then go to the datasets. Select the COVID19 dataset.

Click Add dataset and view the table.

Microsoft Azure | dxcysynpse77819

Bing COVID-19 Data

Description

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

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

For more information and original source data see this [link](#). For license terms see this [link](#).

Datasets:

Modified datasets are available in CSV, JSON, JSON-Lines, and Parquet.

- https://pandemicdatalake.blob.core.windows.net/public/curated/covid-19/bing_covid-19_data/latest/bing_covid-19_data.csv
- https://pandemicdatalake.blob.core.windows.net/public/curated/covid-19/bing_covid-19_data/latest/bing_covid-19_data.json
- https://pandemicdatalake.blob.core.windows.net/public/curated/covid-19/bing_covid-19_data/latest/bing_covid-19_data.jsonl
- https://pandemicdatalake.blob.core.windows.net/public/curated/covid-19/bing_covid-19_data/latest/bing_covid-19_data.parquet

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

Raw data:

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

Previous versions of modified and raw data:

https://pandemicdatalake.blob.core.windows.net/public/curated/covid-19/bing_covid-19_data/previous/

Add dataset Back Close

Click Azure Blob storage->New SQL script->select top 100

Microsoft Azure | dxcysynpse77819

We use optional cookies to provide a better experience. [Learn more](#) Accept Reject More options

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 Select TOP 100 rows

New notebook Create external table

Edit Delete Properties

It will shows the sql script. And click run.

Microsoft Azure | dxcysnps77819

We use optional cookies to provide a better experience. Learn more

Accept Reject More options

Data

Workspace Linked

Filter resources by name

Azure Blob Storage 1

Sample Datasets

bing-covid-19-data

Azure Data Lake Storage Gen2 Actions

SQL script 1

Run Undo Publish Query plan Connect to Built-in Use database master

```
1 -- This is auto-generated code
2 SELECT
3     TOP 100 *
4 FROM
5     OPENROWSET(
6         BULK 'https://pandemicdatalake.blob.core.windows.net/public/curated/covid-19/bing_covid-19_da...
7         FORMAT = 'parquet'
8     ) AS [result];
```

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

The output is displayed in table format.

Microsoft Azure | dxcysnps77819

We use optional cookies to provide a better experience. Learn more

Accept Reject More options

Data

Workspace Linked

Filter resources by name

Azure Blob Storage 1

Sample Datasets

bing-covid-19-data

Azure Data Lake Storage Gen2 2

dxcysnps77819 (Primary - dxcsha...)

(Attached Containers)

SQL script 1

Run Undo Publish Query plan Connect to Built-in Use database master

```
1 -- This is auto-generated code
2 SELECT
3     TOP 100 *
4 FROM
5     OPENROWSET(
6         BULK 'https://pandemicdatalake.blob.core.windows.net/public/curated/covid-19/bing_covid-19_da...
7         FORMAT = 'parquet'
8     ) AS [result];
```

Results Messages

View Table Chart Export results

Search

id	updated	confirmed	confirmed_change	deaths	deaths_change	recovered
338995	2020-01-21T00:00:00Z	262	(NULL)	0	(NULL)	(NULL)
338996	2020-01-22T00:00:00Z	313	51	0	0	(NULL)
338997	2020-01-23T00:00:00Z	578	265	0	0	(NULL)
338998	2020-01-24T00:00:00Z	841	263	0	0	(NULL)
338999	2020-01-25T00:00:00Z	1320	479	0	0	(NULL)

00:00:05 Query executed successfully.

Properties

General Related (0)

Name * SQL script 1

Description

Type .sql script

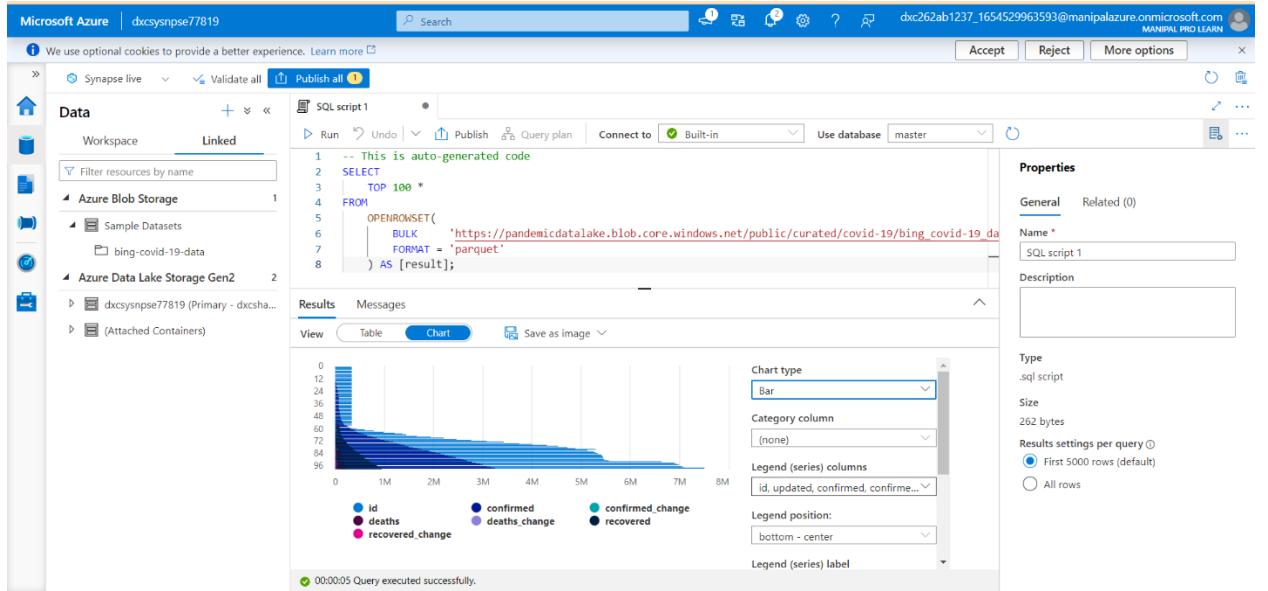
Size 262 bytes

Results settings per query

First 5000 rows (default)

All rows

The result is also in the chart format.



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

Go to dataset and then select Boston Safety datasets

The screenshot shows the Microsoft Azure dataset page for 'Boston Safety Data'. The left sidebar has 'Dataset' selected. The main area has 'Preview' selected. The preview table has the following columns:

dataType	dataSubtype	dateTime	category	subcategory	status	address	latitude	longitude	source
Safety	311_All	10/19/2021 ...	Enforcement...	Parking Enf...	Open	523 E Fourt...	42.3345	-71.0438	Citizens C
Safety	311_All	10/19/2021 ...	Enforcement...	Parking Enf...	Open	507 E Broad...	42.3353	-71.0444	Citizens C
Safety	311_All	10/19/2021 ...	Enforcement...	Parking Enf...	Open	45 Thomas ...	42.3323	-71.047	Citizens C
Safety	311_All	10/19/2021 ...	Enforcement...	Parking Enf...	Open	INTERSECTI...	42.3594	-71.0587	Citizens C
Safety	311_All	10/19/2021 ...	Enforcement...	Parking Enf...	Open	51-53 Wild...	42.2833	-71.0885	Citizens C
Safety	311_All	10/19/2021 ...	Enforcement...	Parking Enf...	Open	15 Wildwoo...	42.2839	-71.0872	Citizens C
Safety	311_All	10/19/2021 ...	Enforcement...	Parking Enf...	Open	44 Wildwoo...	42.2832	-71.0881	Citizens C
Safety	311_All	10/19/2021 ...	Enforcement...	Parking Enf...	Open	701 E Sixth ...	42.333	-71.0323	Citizens C
Safety	311_All	10/19/2021 ...	Enforcement...	Parking Enf...	Open	17 Buswell S...	42.3478	-71.1052	Citizens C
Safety	311_All	10/19/2021 ...	Enforcement...	Parking Enf...	Open	18 Tuckerm...	42.3315	-71.0546	Citizens C

The left sidebar also includes sections for 'Description', 'Volume and Retention', 'Storage Location', and 'Additional Information'. The 'Notices' section at the bottom contains a standard Microsoft disclaimer about dataset usage.

Click Azure Blob storage->New SQL script->select safety datasets.

The screenshot shows the Microsoft Azure Synapse Studio interface. On the left, the Data workspace navigation pane is open, showing 'Azure Blob Storage' and 'Sample Datasets' under 'Azure Blob Storage'. In the center, there are two tabs: 'SQL script 1' and 'SQL script 2'. The 'SQL script 1' tab contains a generated SQL script:`1 -- This is auto-generated code
2 SELECT
3 TOP 100 *
4 FROM
5 OPENROWSET(
6 BULK 'https://azureopendatastorage.blob.core.windows.net/citydatacontainer/Safety/Release/cit...
7 FORMAT = 'parquet'
8) AS [result];`The 'Results' tab is selected, displaying a table with the following data:| dataType | dataSubtype | dateTime | category | subcategory | status | address |
| --- | --- | --- | --- | --- | --- | --- |
| Safety | 311_All | 2016-10-11T14... | Environmental ... | Rodent Activity | Closed | 19 Charter St E |
| Safety | 311_All | 2017-02-14T15... | Code Enforcem... | Improper Stora... | Closed | 44 Fayette St E |

Click run and it shows the results.

The screenshot shows the Microsoft Azure Synapse Studio interface. The 'Publish all' dialog is open, prompting the user to publish pending changes to the live environment. The 'Pending changes' section lists two items:| NAME | CHANGE | EXISTING |
| --- | --- | --- |
| SQL script 1 | (New) | - |
| SQL script 2 | (New) | - |

The screenshot shows the Microsoft Azure Synapse Studio interface. In the top right corner, a message box displays "Publishing completed" and "Successfully published". The main workspace shows a query script titled "SQL script 1" containing the following T-SQL code:

```

1 -- This is auto-generated code
2 SELECT
3     TOP 100 *
4 FROM
5     OPENROWSET(
6         BULK 'https://azureopendatastorage.blob.core.windows.net/citydatacontainer/Safety/Release/cit...
7         FORMAT = 'parquet'
8     ) AS [result];

```

The results pane shows a table with two rows of data from the query. The properties pane on the right indicates the type is ".sql script" and the size is 244 bytes.

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

Click on Apache Spark pool

The screenshot shows the Microsoft Azure Synapse Studio interface with the left sidebar expanded to show "Analytics pools", "Apache Spark pools", and other options like "Data Explorer pools (pre...)" and "External connections". The main area is titled "Apache Spark pool" and contains the following text: "Apache Spark pools can be tuned to run different kinds of Apache Spark workloads using specific configuration libraries, permissions, etc. Learn more". Below this is a "New" button and a "Refresh" button. A search bar labeled "Filter by name" is present. The table below shows 0 items:

Name	Node size family	Size
No items to show Try changing your filter or create new Apache Spark pool		

A large blue "New Apache Spark pool" button is located at the bottom right of the table area.

Click NewApache Spark pool

The screenshot shows the Microsoft Azure portal interface for creating a new Apache Spark pool. The left sidebar navigation includes options like Synapse live, Validate all, Publish all, Analytics pools, SQL pools, Apache Spark pools (selected), Data Explorer pools, External connections, Linked services, Microsoft Purview, Integration (Triggers, Integration runtimes), Security (Access control, Credentials, Managed private endpoints), Configurations + libraries, Workspace packages, and Data flow libraries. The main content area is titled 'New Apache Spark pool' under the 'Basics' tab. It prompts the user to '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.' The 'Apache Spark pool details' section contains fields for Name (dxcspp1), Isolated compute (Disabled), Node size family (Memory Optimized), Node size (Small (4 vCores / 32 GB)), Autoscale (Enabled), Number of nodes (3 to 4), Estimated price (Est. cost per hour - Failed to fetch billing info), and Dynamically allocate executors (Enabled). Buttons at the bottom include 'Review + create' (highlighted in blue), 'Next: Additional settings >', and 'Cancel'.

Fill the details and then click create

The screenshot shows the Microsoft Azure portal interface for creating a new Apache Spark pool. The left sidebar navigation is identical to the previous screenshot. The main content area is titled 'New Apache Spark pool' under the 'Review + create' tab. A green success message states 'Validation succeeded.' Below it, the 'Product details' section shows 'Azure Synapse Analytics Apache Spark pool by Microsoft' and links for 'Terms of use' and 'Privacy policy'. The 'Terms' section contains legal text about agreeing to terms and conditions. The 'Basics' section displays the configuration: Subscription (Azure-DXC262AB12Lab), Resource group (dxcrg7781), Apache Spark pool name (dxcspp1), and Isolated compute (Disabled). At the bottom are 'Create' (highlighted in blue), '< Previous', 'Download template for automation', and 'Cancel' buttons.

The screenshot shows the Microsoft Azure Synapse Analytics workspace interface. On the left, there is a navigation sidebar with various options like Analytics pools, SQL pools, Apache Spark pools, Data Explorer pools, External connections, Linked services, Microsoft Purview, Integration, Triggers, Integration runtimes, Security, Access control, Credentials, Managed private endpoints, Configurations + libraries, Workspace packages, and Data flow libraries. The main area is titled "Apache Spark pool" and contains a table with one item:

Name	Node size family	Size
dxcsp1	Memory Optimized	Small (4 vCores / 32 GB) - 3 to 4 nodes

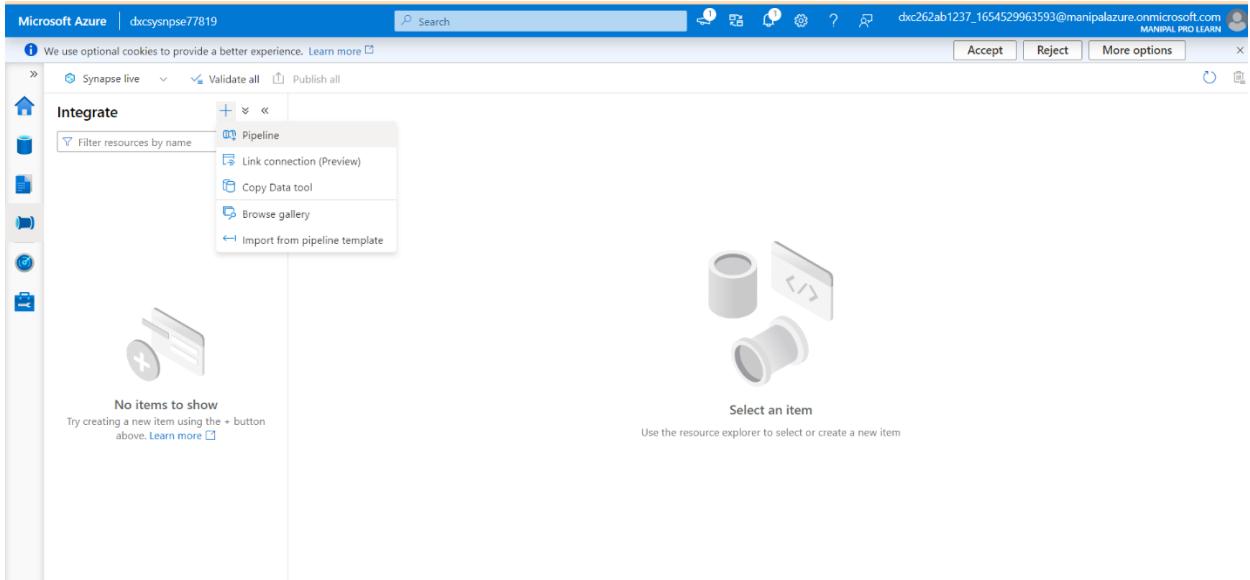
The spark pool is created.

The screenshot shows the Microsoft Azure Synapse Analytics workspace interface. The left sidebar is visible with the "Data" section selected. In the main area, there is a "Notebook 1" tab open, showing a PySpark (Python) code cell. The code is as follows:

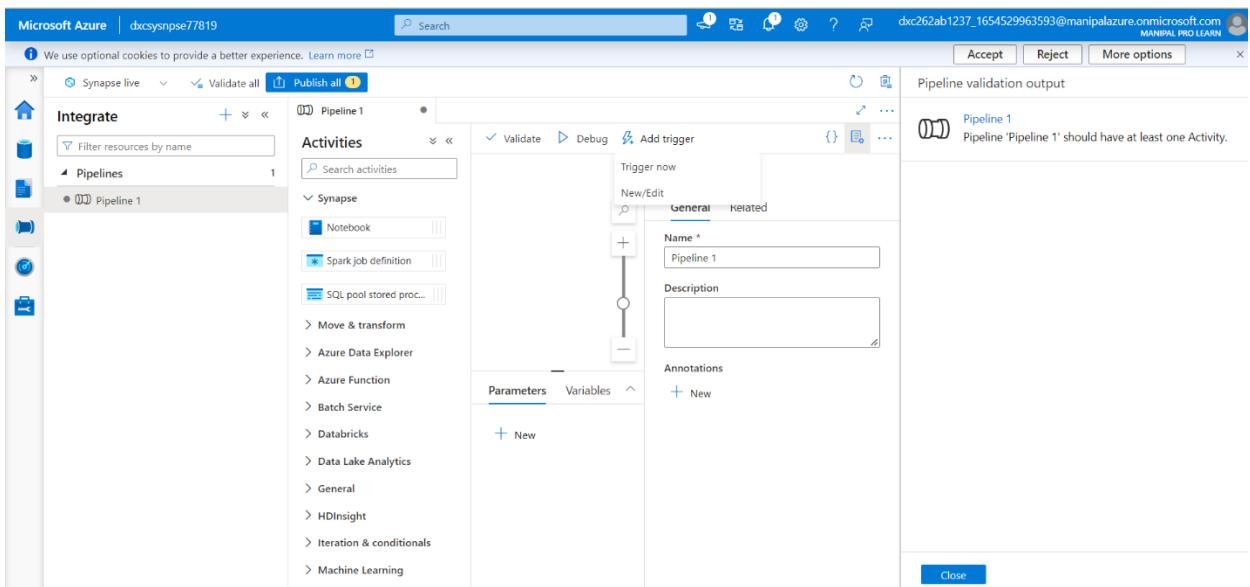
```
1 from azureml.opendatasets import NyctlcYellow
2
3 data = NyctlcYellow()
4 df = data.to_spark_dataframe()
5 # Display 10 rows
6 display(df.limit(10))
```

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

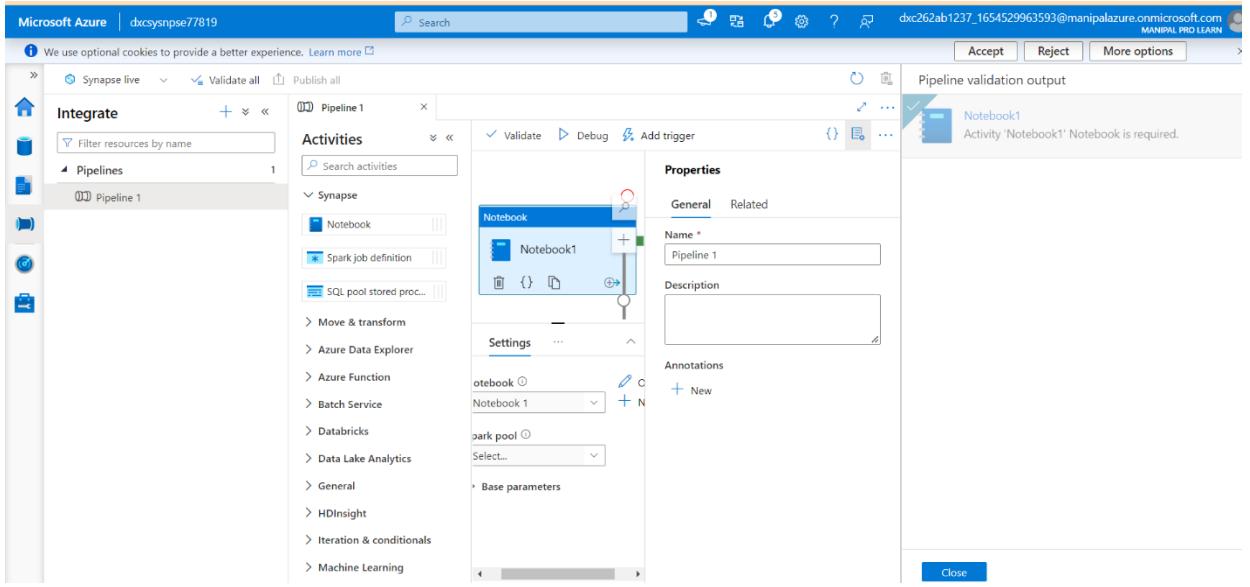
Go to Integrate and then click + button to create pipeline.



The pipeline will be created.

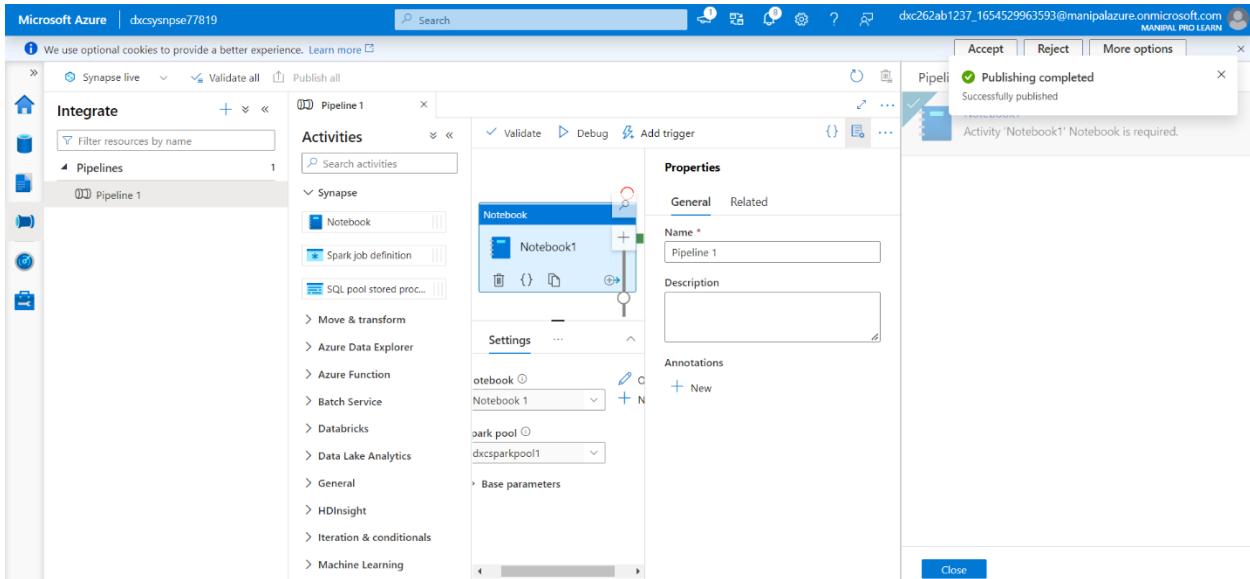


Click synapse to drag the notebook and run the pipeline. The attachment is necessary to run the pipeline.

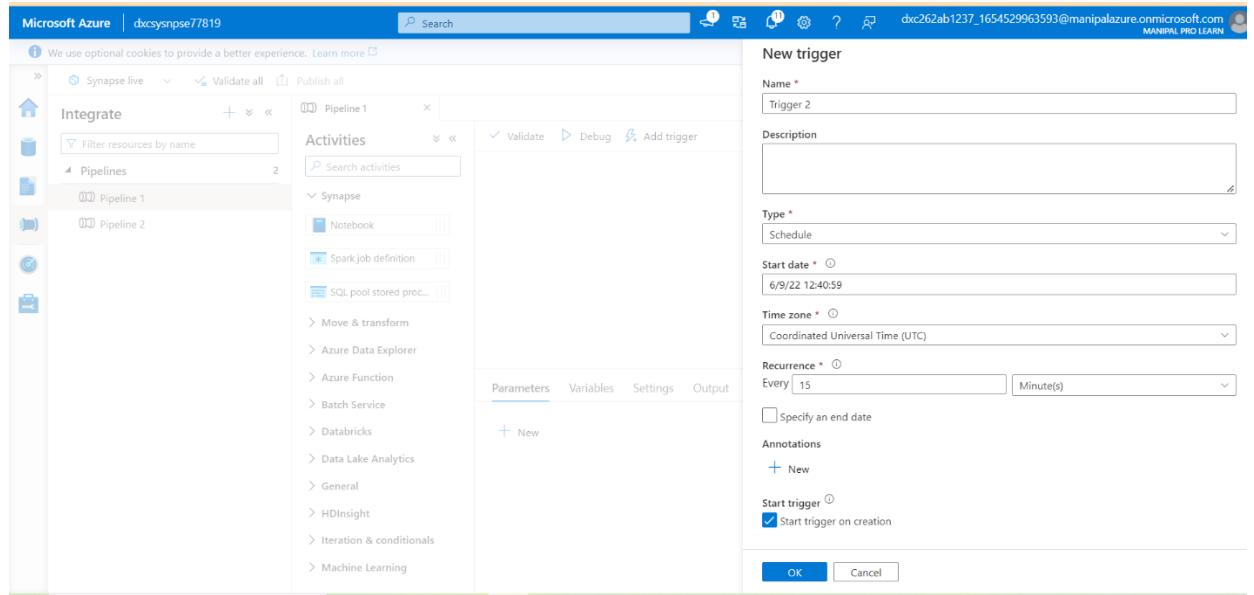


7. Explain the steps with screenshots how to automate the pipelines in Azure Synapse analytics??

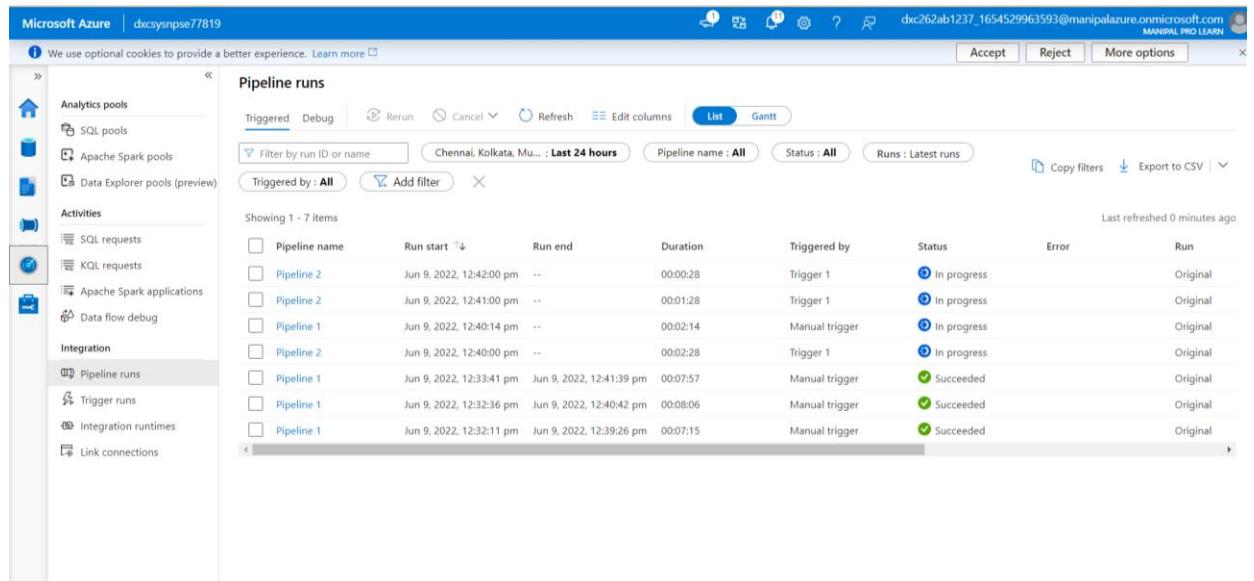
Once the pipeline is created it is need to publish it.



After publishing click the trigger option.



The pipeline runs successfully.



The screenshot shows the Microsoft Azure Databricks interface. On the left, there's a sidebar with categories like Analytics pools, Activities, Integration, and Pipeline runs. The 'Trigger runs' section is selected. The main area displays a table titled 'Trigger runs' with the following columns: Trigger name, Trigger type, Trigger time, Status, Pipelines, Run, Message, Properties, and Run ID. The table lists 11 entries, all of which have a status of 'Succeeded'. The 'Run' column indicates they are all 'Original' runs. The 'Run ID' column shows a sequence of IDs starting from 0858546847685648.

8. Explain the steps with screenshots how to Databricks?

Search on Azure Databricks and then click create Azure Databricks service

The screenshot shows the Azure portal search results for 'Azure Databricks'. At the top, there are buttons for 'Create', 'Manage view', 'Refresh', 'Export to CSV', 'Open query', and 'Assign tags'. Below this are filter options for 'Subscription', 'Resource group', 'Location', and 'Add filter'. The main area displays a message: 'No azure databricks services to display'. It includes a callout for 'Create azure databricks service' and a link to 'Learn more'. At the bottom right, there's a 'Give feedback' button.

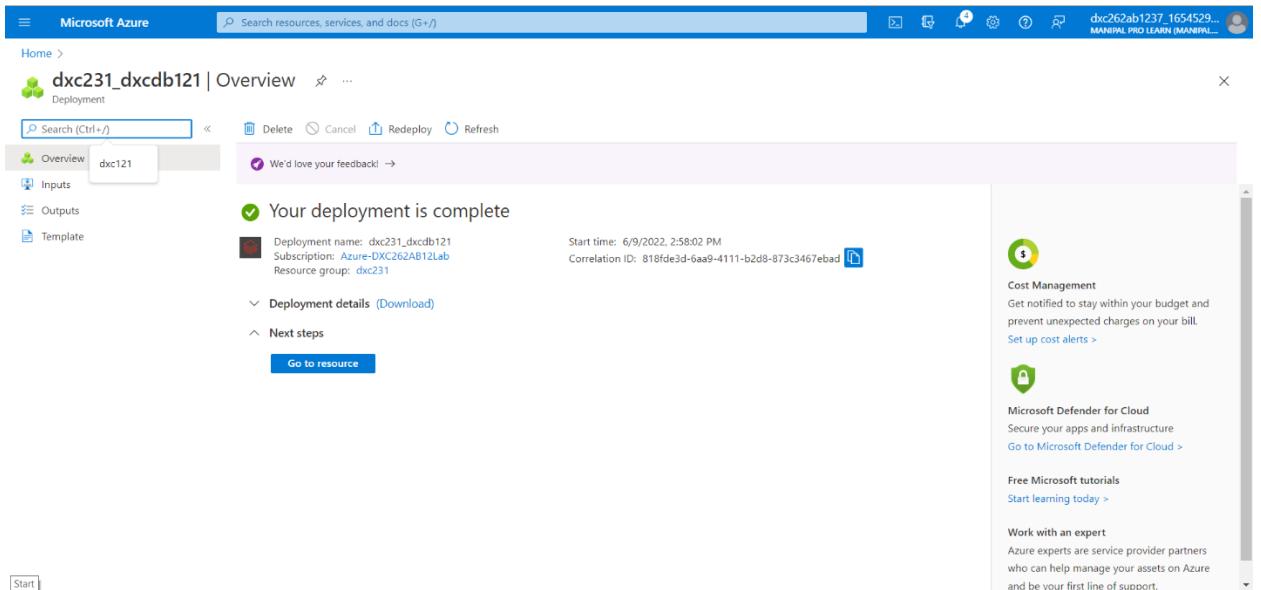
Give the necessary detail to create Azure Databricks.

The screenshot shows the 'Create an Azure Databricks workspace' wizard on the 'Basics' step. The 'Project Details' section is active, showing the selected subscription ('Azure-DXC262A812Lab') and resource group ('(New) dxc231'). The 'Instance Details' section shows the workspace name ('dxcdb121'), region ('East US'), and pricing tier ('Standard (Apache Spark, Secure with Azure AD)').

After validation is success click on create button.

The screenshot shows the 'Create an Azure Databricks workspace' wizard on the 'Review + create' step. A green success message at the top states 'Validation Succeeded'. The 'Summary' section displays the configuration details: Workspace name (dxcdb121), Subscription (Azure-DXC262A812Lab), Resource group (dxc231), Region (East US), and Pricing Tier (trial). The 'Networking' section shows that 'Deploy Azure Databricks workspace with Secure Cluster Connectivity (No Public IP)' is set to 'No'. The 'Advanced' section is partially visible below.

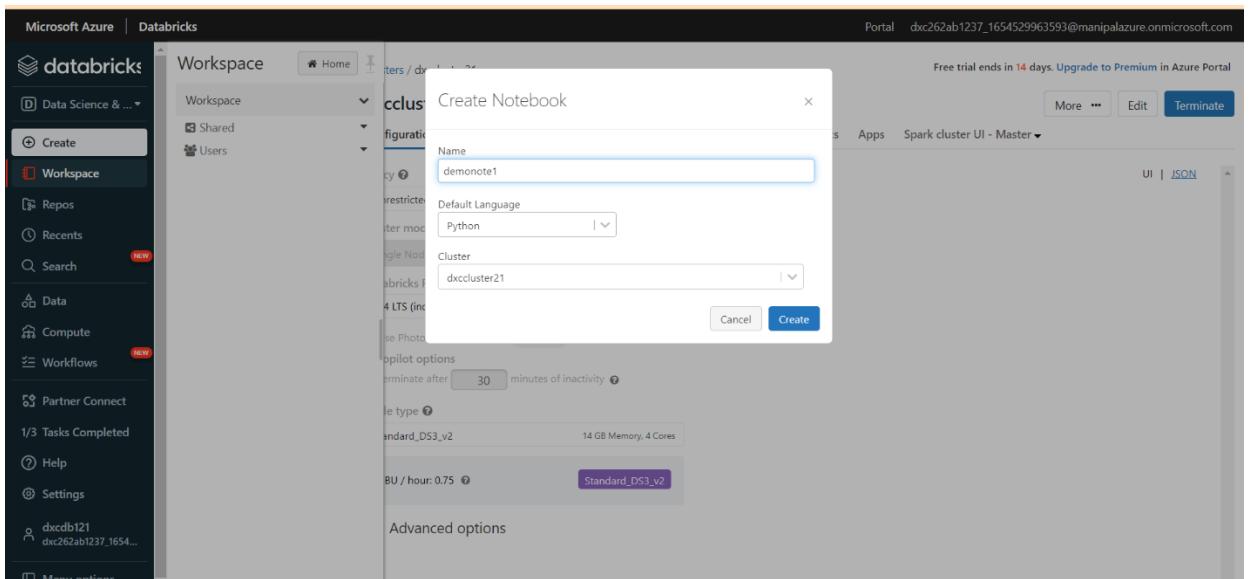
The deployment is completed.



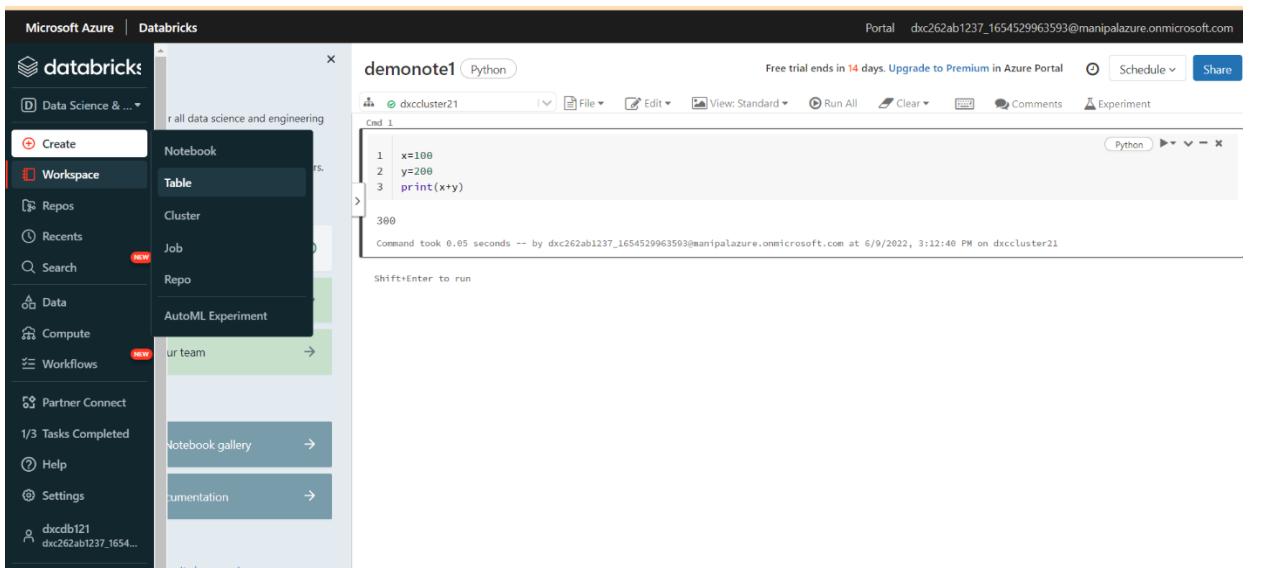
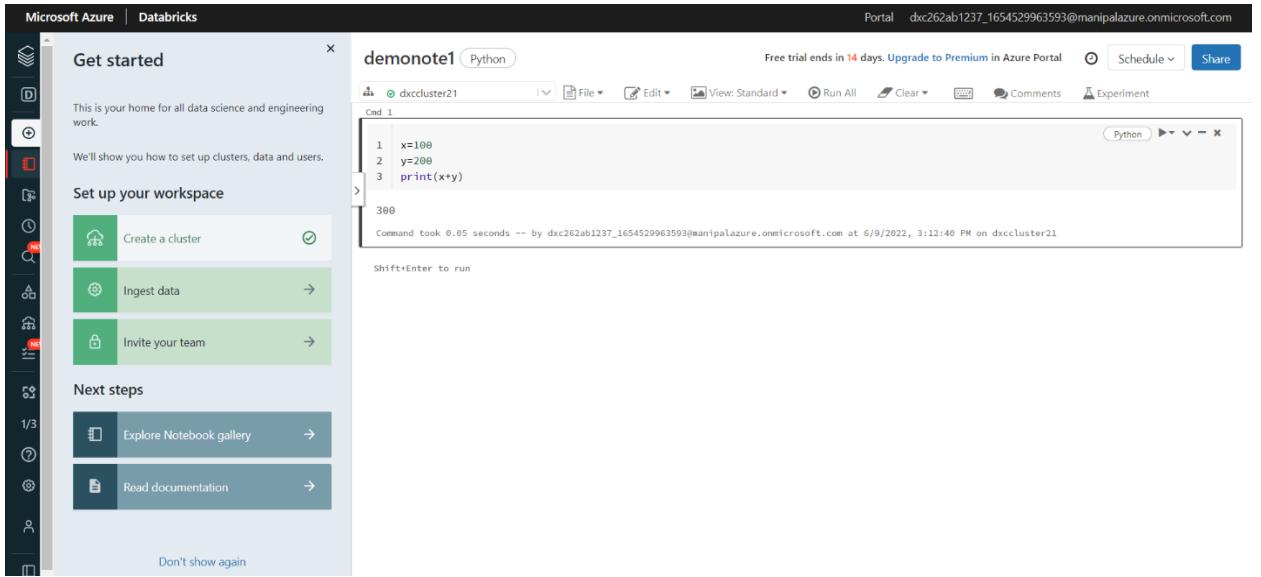
The Azure Databricks is created.

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

Click create->Create Notebook



After give the name and cluster it will create the notebook.



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

Microsoft Azure | Databricks

Portal dxc262ab1237_1654529963593@manipalazure.onmicrosoft.com

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

Workflows

Jobs Job runs Delta Live Tables

Create Job

No jobs found.

Select notebook name and confirm it.

Microsoft Azure | Databricks

Portals dxc262ab1237_1654529963593@manipalazure.onmicrosoft.com

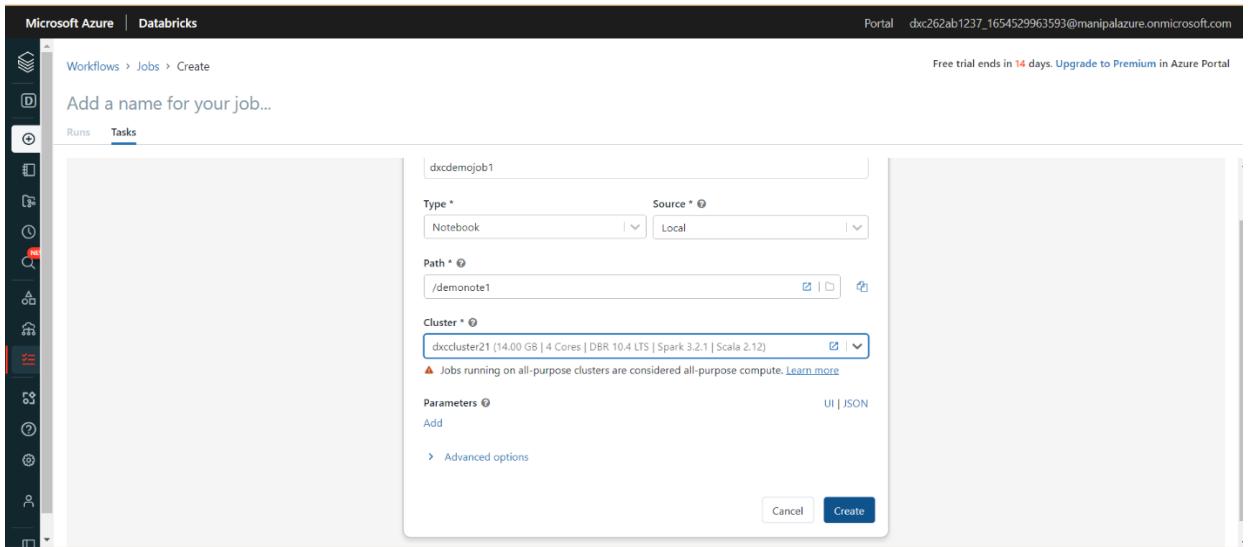
Add a name for your job...

Workflow > Jobs > Create

Select Notebook

Runs Tasks

Cancel Confirm



Click the run to run the notebook.

The screenshot shows the 'Runs' page for the job 'dxcdemojob1'. It lists 'Active runs' and 'Completed runs (past 60 days)'. On the right, there is a sidebar with sections for 'Job details', 'Git', and 'Schedule'. The 'Job details' section shows the job ID, creator, and run-as information. The 'Git' section has a 'Add Git settings' button. The 'Schedule' section indicates 'None' and has an 'Edit schedule' button.

Microsoft Azure | Databricks

Workflows > Jobs > dxcdemojob1

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

Run now

dxcdemojob1

Runs

Active runs

Start time	Run ID	Launched	Duration	Spark	Status	Actions
Run now / Run now with different parameters						

Completed runs (past 60 days)

Latest successful run (refreshes automatically)

Start time	Run ID	Launched	Duration	Spark	Status	Actions
Jun 9 2022, 15:36 PM IST	840	Manually	19s	Spark UI / Logs / Metrics	Succeeded	

Job details

Job ID: 120994987967673

Creator: dxc262ab1237_1654529963593@manipalazure.onmicrosoft.com

Run as: dxc262ab1237_1654529963593@manipalazure.onmicrosoft.com

Tags: + Tag

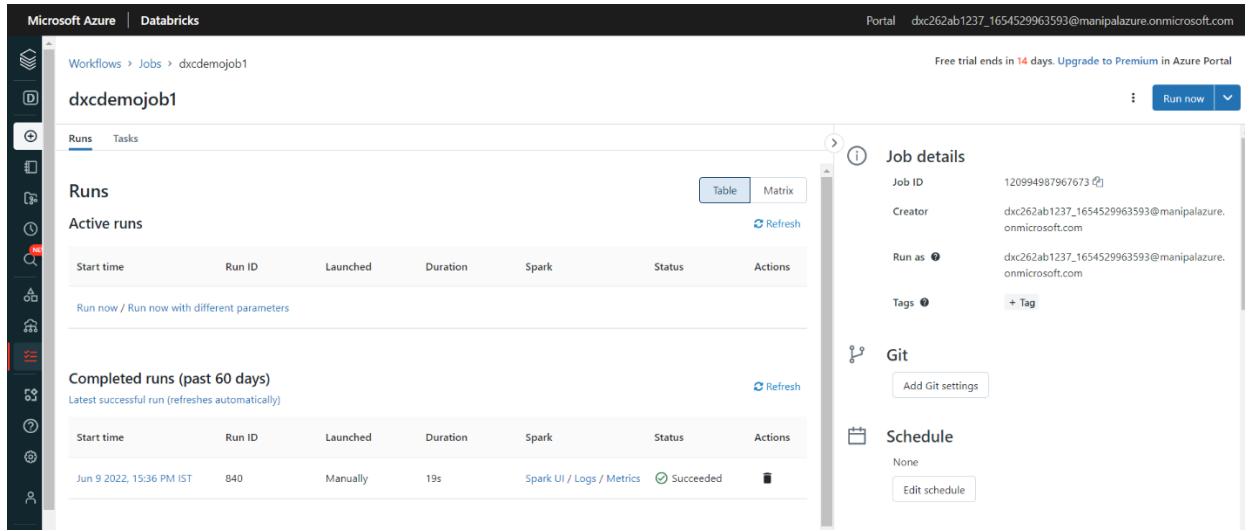
Git

Add Git settings

Schedule

None

Edit schedule



Under create table is to insert the file.

Microsoft Azure | Databricks

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

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

Create New Table

Data source: DBFS

Upload File

DBFS Other Data Sources

DBFS Target Directory: /FileStore/tables/ (optional)

File uploaded to DBFS are accessible by everyone who has access to this workspace. Learn more

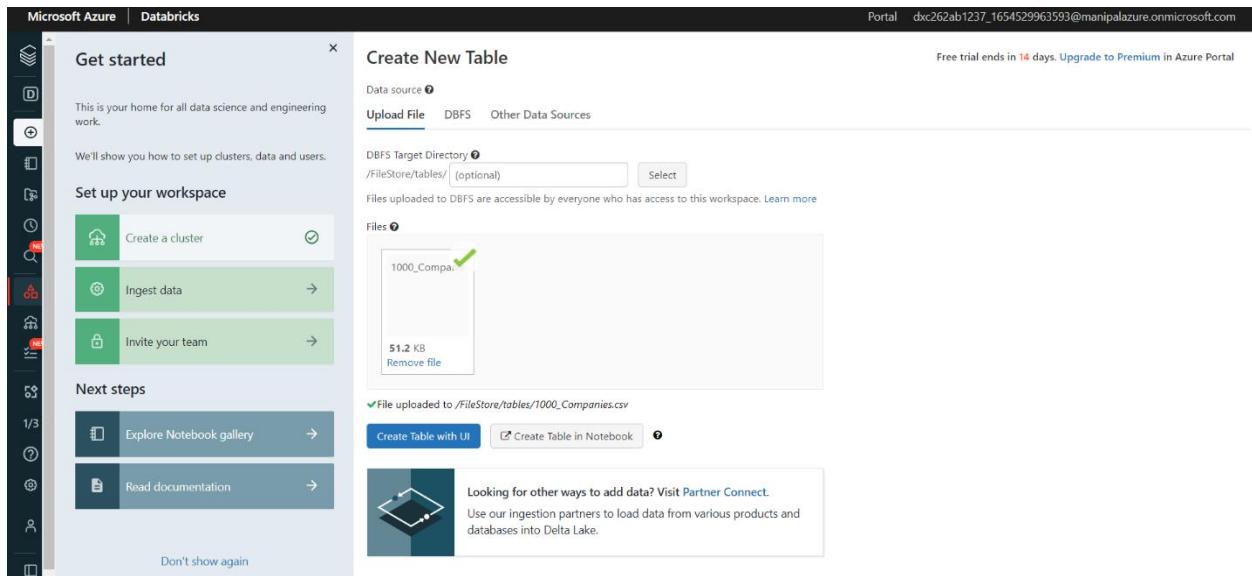
Files:

- 1000_Compa.csv (51.2 KB) Remove file

✓ File uploaded to /FileStore/tables/1000_Companies.csv

Create Table with UI Create Table in Notebook

Looking for other ways to add data? Visit Partner Connect. Use our ingestion partners to load data from various products and databases into Delta Lake.



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

Overview

This notebook will show you how to create and query a table or DataFrame that you uploaded to DBFS. DBFS is a Databricks File System that allows you to store data for querying inside of Databricks. This notebook assumes that you have a file already inside of DBFS that you would like to read from.

This notebook is written in **Python** so the default cell type is Python. However, you can use different languages by using the `%LANGUAGE` syntax. Python, Scala, SQL, and R are all supported.

```

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

```

It will be run and create the result in table.

2022-06-09 - DBFS Example Python

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

(2) Spark Jobs

df: pyspark.sql.dataframe.DataFrame = [c0: string, c1: string ... 3 more fields]

Table Data Profile

	_c0	_c1	_c2	_c3	_c4
1	R&D Spend	Administration	Marketing Spend	State	Profit
2	165349.2	136897.8	471784.1	New York	192261.83
3	162597.7	151377.59	443898.53	California	191792.06
4	153441.51	101145.55	407934.54	Florida	191050.39
5	144372.41	118671.85	383199.62	New York	182901.99
6	142107.34	91391.77	366168.42	Florida	166187.94
7	131876.9	99814.71	362861.36	New York	156991.17

Truncated results, showing first 1000 rows.
Click to re-execute with maximum result limits.

Command took 4.22 seconds -- by dxc262ab1237_1654529963593@manipalazure.onmicrosoft.com at 6/9/2022, 3:19:39 PM on dxcluster21

Cmd 3

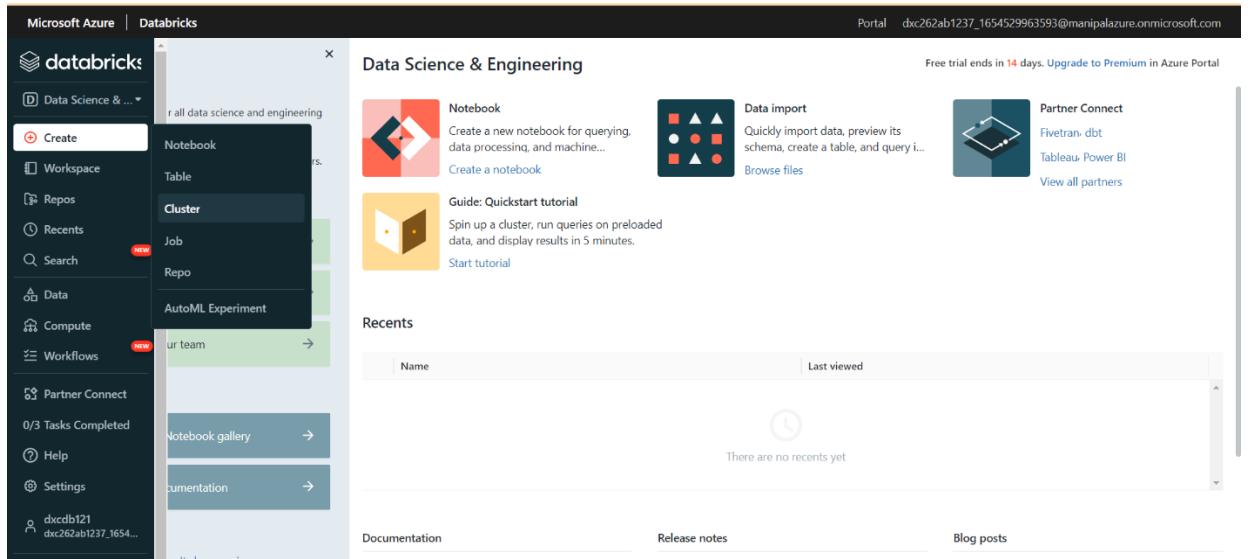
```

1 # Create a view or table
2
3 temp_table_name = "1000_Companies_csv"
4

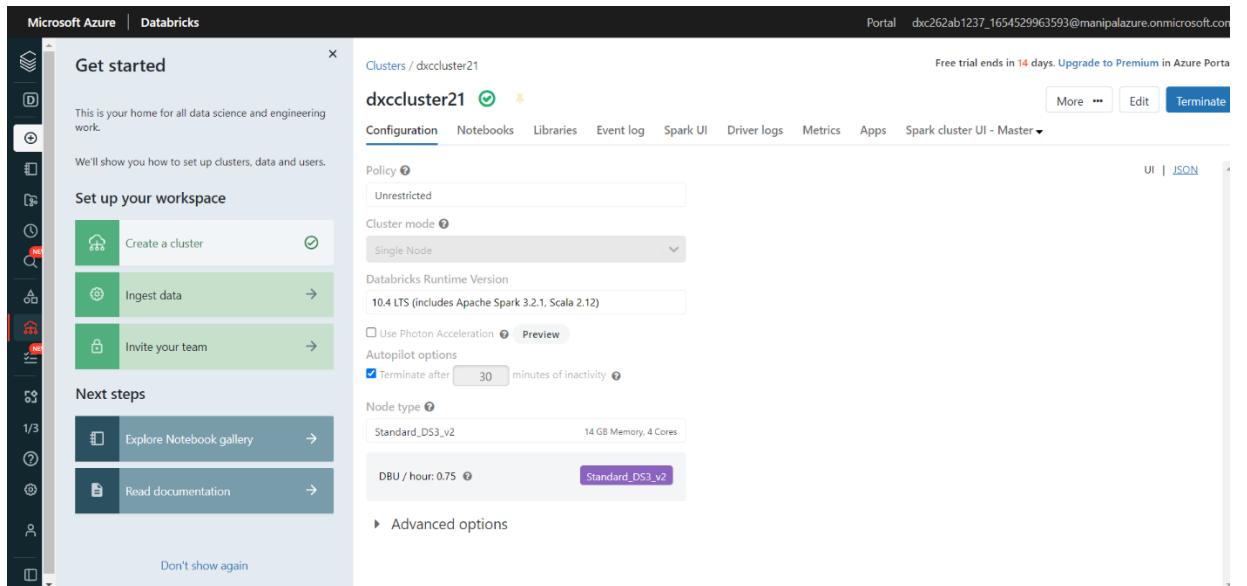
```

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

Click Create->cluster.



Give the cluster name.



Select on the respected cluster name.

Microsoft Azure | Databricks

Portal dxc262ab1237_1654529963593@manipalazure.onmicrosoft.com

The screenshot shows the Microsoft Azure Databricks Compute interface. On the left, there's a sidebar with options like Create, Workspace, Data, Compute, Workflows, Partner Connect, Help, Settings, and a user profile. The main area is titled 'Compute' with tabs for All-purpose clusters, Job clusters, Pools, and Cluster policies. A table lists existing clusters, including 'dxccluster21' which is highlighted. The cluster details show it was created by 'dxc262ab1237_1654529963593@manipalazure.onmicrosoft.com' and is accessible by 'me'. It has a runtime of 10.4, 14 GB memory, 4 cores, and a cost of 0.75 BU/h.

The cluster is created successfully.

Microsoft Azure | Databricks

Portal dxc262ab1237_1654529963593@manipalazure.onmicrosoft.com

This screenshot shows the Microsoft Azure Databricks workspace for the cluster 'dxccluster21'. The cluster card is visible with its configuration details. A context menu is open over the cluster name, showing options like Create, Notebook, Library, Folder, Permissions, Copy Link Address, Sort, and Advanced options. The 'Create' option is highlighted.