

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

The screenshot shows the Microsoft Azure portal interface. The search bar at the top has 'synapse' typed into it. Below the search bar, there's a navigation menu with 'All' selected, followed by 'Services (3)', 'Resources (0)', 'Resource Groups (0)', 'Marketplace (10)', and 'Documentation (29)'. The main content area displays search results under 'Services'. It includes items like 'Azure Synapse Analytics' and 'Azure Synapse Analytics (private link hubs)'. Below this, there are sections for 'Resources', 'Marketplace', and 'Documentation'. A right-hand sidebar features a 'More services' button with an arrow icon.

The screenshot shows the 'Create Synapse workspace' wizard in the Microsoft Azure portal. The current step is 'Basics'. It requires selecting a subscription ('Subscription \*') and creating a new resource group ('Resource group \*'). The 'Managed resource group' field is optional ('Managed resource group'). Below these, 'Workspace details' are specified: 'Workspace name \*' and 'Region'. At the bottom, there are buttons for 'Review + create' and 'Next: Security >'. The status bar at the bottom shows the date and time as 09-06-2022 15:51.

General (Meeting) | Microsoft | [Subscription Details](#) | Nuvepro | Create Synapse workspace - Microsoft Azure | +

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

Microsoft Azure | Search resources, services, and docs (G+) | dxc262ab1216.1654530... | MANHAI PRO LEARN (MANHAI...)

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

\* Basics \* Security Networking Tags Review + create

Create a Synapse workspace to develop an enterprise analytics solution in just a few clicks.

**Project details**

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

Subscription \*  Azure-DXC262AB12Lab  dxorg1  Create new

Resource group \*  Managed resource group  Enter managed resource group name

**Workspace details**

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 \*  dxsy

Region \*  East US

**Review + create** < Previous Next: Security >

1000\_Companies.csv hour (1).csv hour.csv Show all

35°C Mostly sunny ENG INTL 15:52 09-06-2022

General (Meeting) | Microsoft | [Subscription Details](#) | Nuvepro | Create Synapse workspace - Microsoft Azure | +

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

Microsoft Azure | Search resources, services, and docs (G+) | dxc262ab1216.1654530... | MANHAI PRO LEARN (MANHAI...)

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 statements(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  dxorg1

Resource group  Create  < Previous Next > Download a template for automation

1000\_Companies.csv hour (1).csv hour.csv Show all

35°C Mostly sunny ENG INTL 15:52 09-06-2022

The screenshot shows the Microsoft Azure Synapse Analytics deployment status page. It displays a green checkmark indicating the deployment is complete. Deployment details include the name: Microsoft.Azure.SynapseAnalytics-20220609155121, subscription: Azure-DXC262AB12Lab, and resource group: dxcorg1. The start time was 6/9/2022, 3:52:54 PM, and the correlation ID is e7556e41-24b1-4f38-a0a3-cfba83e3ad7. A 'Deployment succeeded' message states that the deployment to resource group 'dxcorg1' was successful. 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', and 'Free Microsoft tutorials'. The bottom of the screen shows the Windows taskbar with various pinned icons and the date/time as 09-06-2022 15:56.

The screenshot shows the Microsoft Azure Resource Group overview for 'dxcorg1'. In the 'Essentials' section, it shows the subscription (move) to 'Azure-DXC262AB12Lab', subscription ID '406c09c5-382f-44be-b843-b50a69eadc92', and location 'East US'. There are 2 succeeded deployments. The 'Resources' section lists two resources: 'dxcdatabricks1' (Azure Databricks Service) and 'dxcsy' (Synapse workspace). Both resources are located in East US. The bottom of the screen shows the Windows taskbar with various pinned icons and the date/time as 09-06-2022 15:56.

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

### 3. Explain the steps with screenshots how to import COVID19 dataset in AzureSynapse analytics

**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.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.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.jsonl) [https://pandemicdatalake.blob.core.windows.net/public/curated/covid-19/bing\\_covid-19\\_data/latest/bing\\_covid-19\\_data.parquet](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](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/](https://pandemicdatalake.blob.core.windows.net/public/curated/covid-19/bing_covid-19_data/)

Add dataset Back Close

38°C Sunny 10:08 09-06-2022

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

Accept Reject More options

**Data** + <> Publish all

**Linked**

Filter resources by name

- Azure Blob Storage** 1
- Sample Datasets**
  - bing-covid-19-data
- Azure Data Lake Storage Gen2**
  - 2 New SQL script
  - New notebook
  - Edit
  - Delete
  - Properties

Select TOP 100 rows

38°C Sunny 10:11 09-06-2022

The screenshot shows the Microsoft Azure Synapse Analytics workspace interface. On the left, the 'Data' sidebar is open, showing 'Linked' datasets. Under 'Linked', there is a 'Sample Datasets' section containing 'bing-covid-19-data'. The main area displays a 'SQL script 1' editor with the following code:

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

The 'Properties' panel on the right shows the following details:

- Name: SQL script 1
- Type: sql script
- Size: 262 bytes
- Results settings per query:
  - First 5000 rows (default) (selected)
  - All rows

The status bar at the bottom indicates it's 10:12 on 09-06-2022.

The screenshot shows the Microsoft Azure Synapse Analytics workspace interface after executing the SQL script. The 'Results' tab is selected, displaying the output of the query. The table contains the following data:

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)

A message at the bottom states '00:02:12 Query executed successfully.' The status bar at the bottom indicates it's 10:14 on 09-06-2022.

Screenshot of Microsoft Azure Synapse Analytics workspace showing a query results chart.

**SQL Script:**

```

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

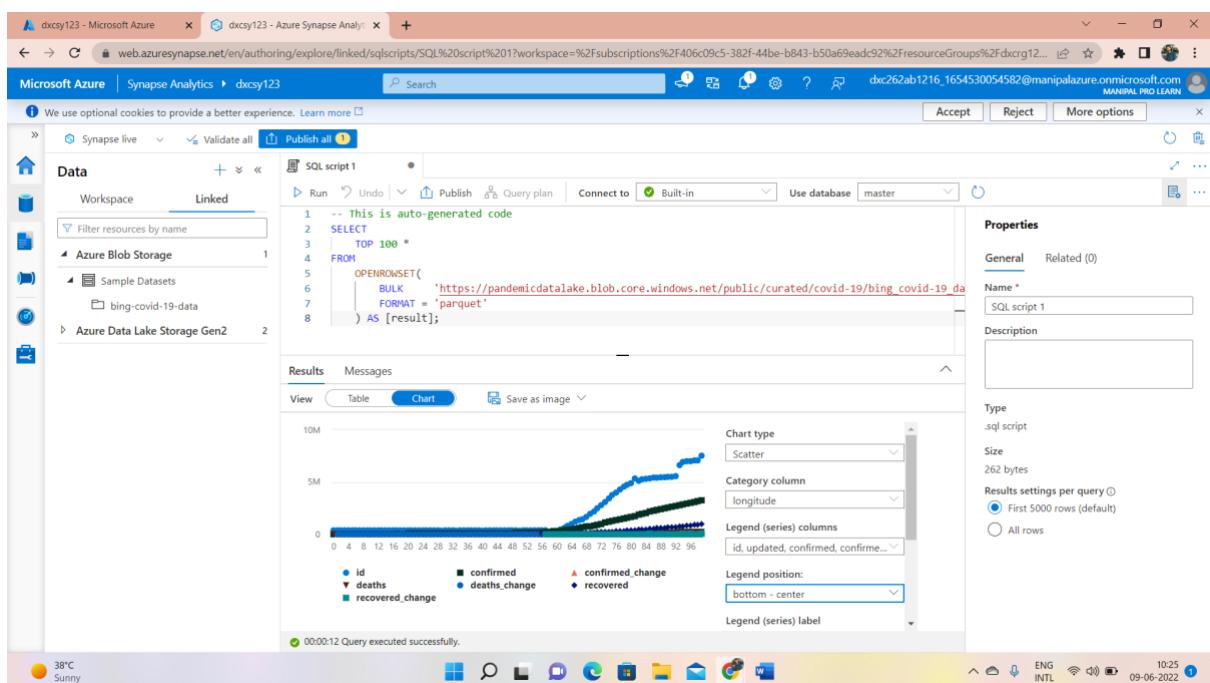
```

**Properties:**

- Name: SQL script 1
- Description:
- Type: sql script
- Size: 262 bytes
- Results settings per query:
  - First 5000 rows (default)
  - All rows

**Chart:**

Line chart showing data over time. The Y-axis ranges from 0M to 8M. The X-axis shows dates from 0 to 96. The chart shows a sharp increase starting around day 60, reaching approximately 7M by day 96.



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

[dxcsy123 - Microsoft Azure](#) [dxcsy123 - Azure Synapse Analytics](#)

Microsoft Azure | Synapse Analytics > dxcsy123

Boston Safety Data

Description

311 calls reported to the city of Boston.

Refer to this link to learn more about [BOS:311](#).

Volume and Retention

This dataset is stored in Parquet format. It is updated daily, and contains about 100K rows (10MB) in total as of 2019.

This dataset contains historical records accumulated from 2011 to the present. You can use parameter settings in our SDK to fetch data within a specific time range.

Storage Location

This dataset is stored in the East US Azure region. Allocating compute resources in East US is recommended for affinity.

Additional Information

This dataset is sourced from city of Boston government. More details can be found from here. Reference [Open Data Commons Public Domain Dedication and License \(ODC PDDL\)](#) for the license of using this dataset.

Notices

MICROSOFT PROVIDES AZURE OPEN DATASETS ON AN "AS IS" BASIS. MICROSOFT MAKES NO WARRANTIES, EXPRESS OR IMPLIED, GUARANTEES OR CONDITIONS WITH RESPECT TO YOUR USE OF THE DATASETS. TO THE EXTENT PERMITTED UNDER YOUR LOCAL LAW, MICROSOFT DISCLAIMS ALL LIABILITY FOR ANY DAMAGES OR LOSSES, INCLUDING DIRECT, CONSEQUENTIAL, SPECIAL, INDIRECT, INCIDENTAL

Preview

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	INTERSECT...	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

Add dataset Back Close

38°C Sunny 10:28 09-06-2022

[dxcsy123 - Microsoft Azure](#) [dxcsy123 - Azure Synapse Analytics](#)

Microsoft Azure | Synapse Analytics > dxcsy123

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

Synapse live Validate all Publish all

Data

Workspace Linked

Filter resources by name

Azure Blob Storage 1

- Sample Datasets
  - bing-covid-19-data
  - city\_safety\_boston

Azure Data Lake Storage Gen2 2

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

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)

00:00:06 Query executed successfully.

38°C Sunny 10:31 09-06-2022

The screenshot shows the Microsoft Azure Synapse Analytics workspace interface. On the left, the 'Data' sidebar is open, showing 'Linked' datasets under 'Azure Blob Storage'. One dataset, 'bing-covid-19-data', is selected. In the main area, a SQL script is displayed:

```

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

```

The results pane shows a chart titled 'COVID-19 Data' with an area chart representing cumulative confirmed cases over time. The legend includes 'confirmed', 'deaths', 'deaths\_change', 'id', 'recovered', and 'recovered\_change'. The chart has a Y-axis from 0 to 10M and an X-axis from 0 to 95. The data shows a sharp increase starting around week 50.

Properties panel on the right shows the following details:

- Name:** SQL script 1
- Type:** sql script
- Size:** 262 bytes
- Results settings per query:**
  - First 5000 rows (default)
  - All rows

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

The screenshot shows the Microsoft Azure Synapse Analytics workspace interface. On the left, the 'Data' sidebar is open, showing 'Linked' datasets under 'Azure Blob Storage'. One dataset, 'nyc\_tlc\_yellow', is selected. In the main area, a notebook cell is open with the following PySpark code:

```

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))

```

A warning message at the top of the cell says: 'Please select a Spark pool to attach before running cell!'. The cell status is 'Not started'.

**Microsoft Azure | Synapse Analytics > dxcsy123**

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

Validate all Publish all 1

**Apache Spark pool**

Apache Spark pools can be tuned to run different kinds of Apache Spark workloads using specific configuration libraries, permissions, etc. Learn more ▾

+ New Refresh

Filter by name

Showing 0-0 of 0 item

Name	Node size family	Size
No items to show		

Try changing your filter or create new Apache Spark pool

New Apache Spark pool

39°C Sunny ENG INTL 10:59 09-06-2022

**Microsoft Azure | Synapse Analytics > dxcsy123**

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

Validate all Publish all 1

**Apache Spark pool**

Apache Spark pools can be tuned to run different kinds of Apache Spark workloads using specific configuration libraries, permissions, etc. Learn more ▾

+ New Refresh

Filter by name

Showing 0-0 of 0 item

**Basics** Additional settings Tags Review + create

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

**Apache Spark pool details**

Name your Apache Spark pool and choose its initial settings.

Apache Spark pool name \* Enter Apache Spark pool name

Isolated compute \*  Enabled  Disabled

Node size family \* Memory Optimized

Node size \* Medium (8 vCores / 64 GB)

Autoscale \*  Enabled  Disabled

Number of nodes \* 3

Estimated price  Failed to fetch billing info

Dynamically allocate executors \*  Enabled  Disabled

Review + create Next: Additional settings > Cancel

39°C Sunny ENG INTL 10:59 09-06-2022

**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	dxcrg123
Apache Spark pool name	dxcspark1
Isolated compute	Disabled
Node size family	Memory Optimized

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

The screenshot shows the Microsoft Azure portal interface. The user is creating a new Apache Spark pool under the 'Synapse Analytics' section. The 'Basics' tab is selected. The 'Name' field is filled with 'dxcspark1'. Other configuration options like 'Resource group' (dxcrg123), 'Node size family' (Memory Optimized), and 'Isolated compute' (Disabled) are also visible. A validation message at the top says 'Validation succeeded.' Below the form, there's a 'Terms' section with legal disclaimers. At the bottom, there are 'Create', 'Cancel', and 'Download template for automation' buttons.

**Apache Spark pool**

Apache Spark pools can be tuned to run different kinds of Apache Spark workloads using specific configuration libraries, permissions, etc. [Learn more](#)

**Filter by name**

Showing 1-1 of 1 item

Name	Node size family	Size
dxcspark1	Memory Optimized	Small (4 vCores / 32 GB) - 3 to 3 nodes

The screenshot shows the Microsoft Azure portal interface. The user is viewing the list of existing Apache Spark pools under the 'Synapse Analytics' section. One pool, named 'dxcspark1', is listed. It is a 'Memory Optimized' pool with a size of 'Small (4 vCores / 32 GB) - 3 to 3 nodes'. The interface includes a sidebar with various service icons and a navigation bar at the top.

The screenshot shows the Microsoft Azure Synapse Analytics Notebook interface. On the left, the Data workspace sidebar lists 'Azure Blob Storage' with 'Sample Datasets' like 'bing-covid-19-data', 'city\_safety\_boston', 'covid-tracking', and 'nyc\_tlc\_yellow'. It also shows 'Azure Data Lake Storage Gen2' with 'dxcisy123 (Primary - dxcad12)' and 'dxcf12 (Primary)'. The main area displays a PySpark notebook titled 'Notebook 1' with the following code:

```

1 %%pyspark
2 blob_account_name = "pandemicdatalake"
3 blob_container_name = "public"
4 blob_relative_path = "curated/covid-19/bing_covid-19_data/latest/bing_covid-19_data.parquet"
5 blob_sas_token = r"\" 
6 # Allow SPARK to read from Blob remotely
7 wasbs_path = 'wasbs://{}{}.blob.core.windows.net/{}'.format(blob_container_name, blob_account_name, blob_relative_path)
8
9 spark.conf.set(
10     'fs.azure.sas.{}'.format(blob_container_name), blob_sas_token)
11 df = spark.read.parquet(wasbs_path)
12 display(df.limit(10))

```

The status bar indicates a successful job execution: 'Job execution Succeeded Spark 2 executors 8 cores'. Below the code, a table is displayed:

ID	updated	confirmed	confirmed_change	deaths
338995	2020-01-21	262	undefined	0
338996	2020-01-22	313	51	0
338997	2020-01-23	578	265	0
338998	2020-01-24	841	263	0

The screenshot shows the Microsoft Azure Synapse Analytics Notebook interface. The left sidebar is identical to the first screenshot. The main area displays a Python notebook titled 'Notebook 1' with the following code:

```

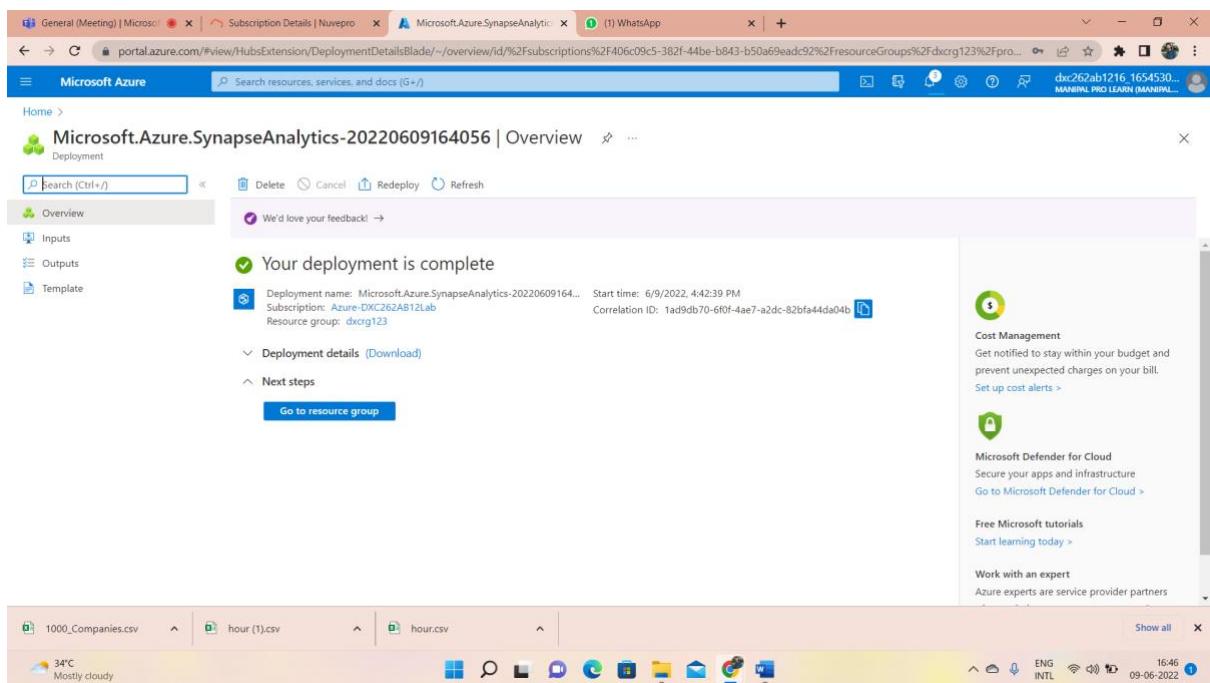
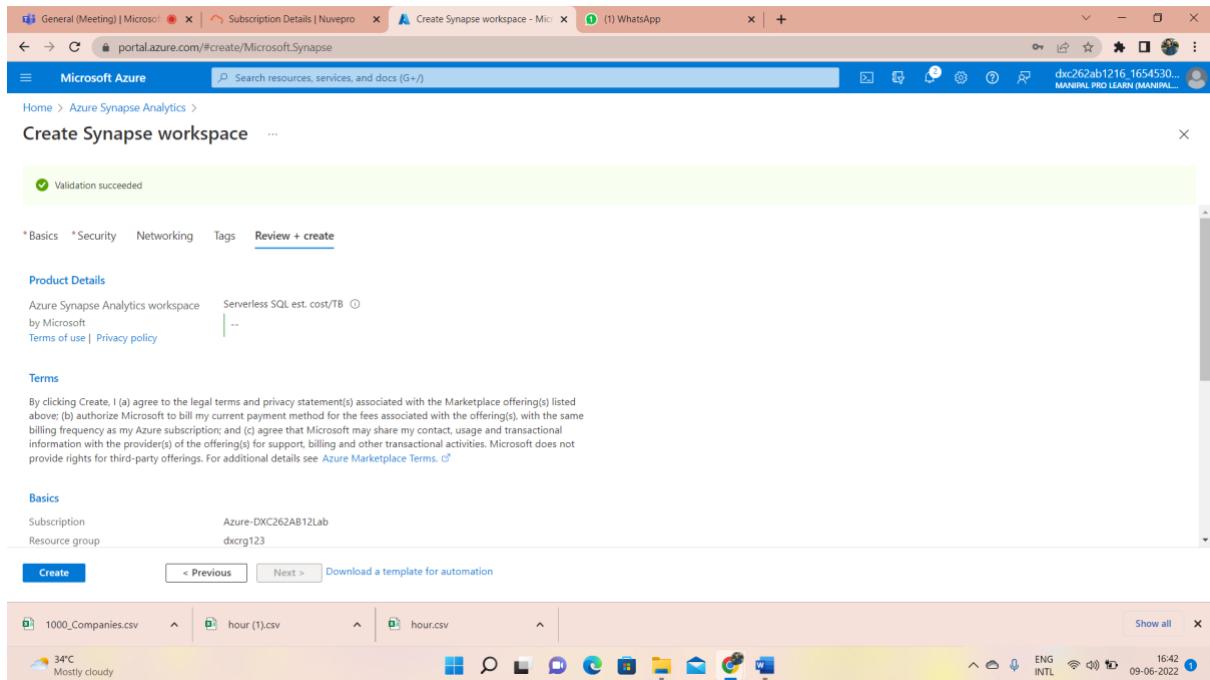
1 from azureml.opendatasets import BostonSafety
2
3 data = BostonSafety()
4 df = data.to_spark_dataframe()
5 # Display 10 rows
6 display(df.limit(10))

```

The status bar indicates a successful job execution: 'Job execution Succeeded Spark 2 executors 8 cores'. Below the code, a table is displayed:

ID	Description	Status	Stages	Tasks	Submission Time	Duration
> Job 2	parquet at NativeMethodAccessorsImpl.java:0	<span style="color: green;">✓ Succeeded</span>	1/1	1/1 succeeded	11:59:49 AM, 6/09/22	< 1 ms
> Job 3	getRowsInJsonString at Display.scala:454	<span style="color: green;">✓ Succeeded</span>	2/2	9/9 succeeded	11:59:50 AM, 6/09/22	2 sec

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



8. Explain the steps with screenshots how to Databricks ?

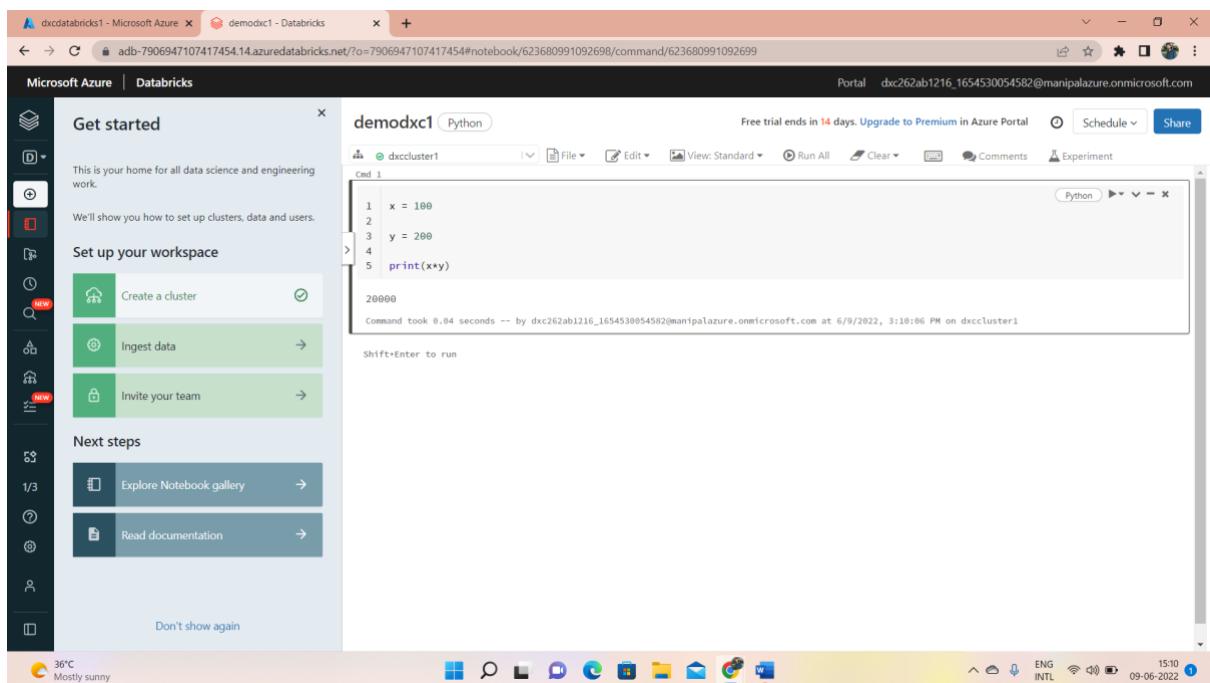
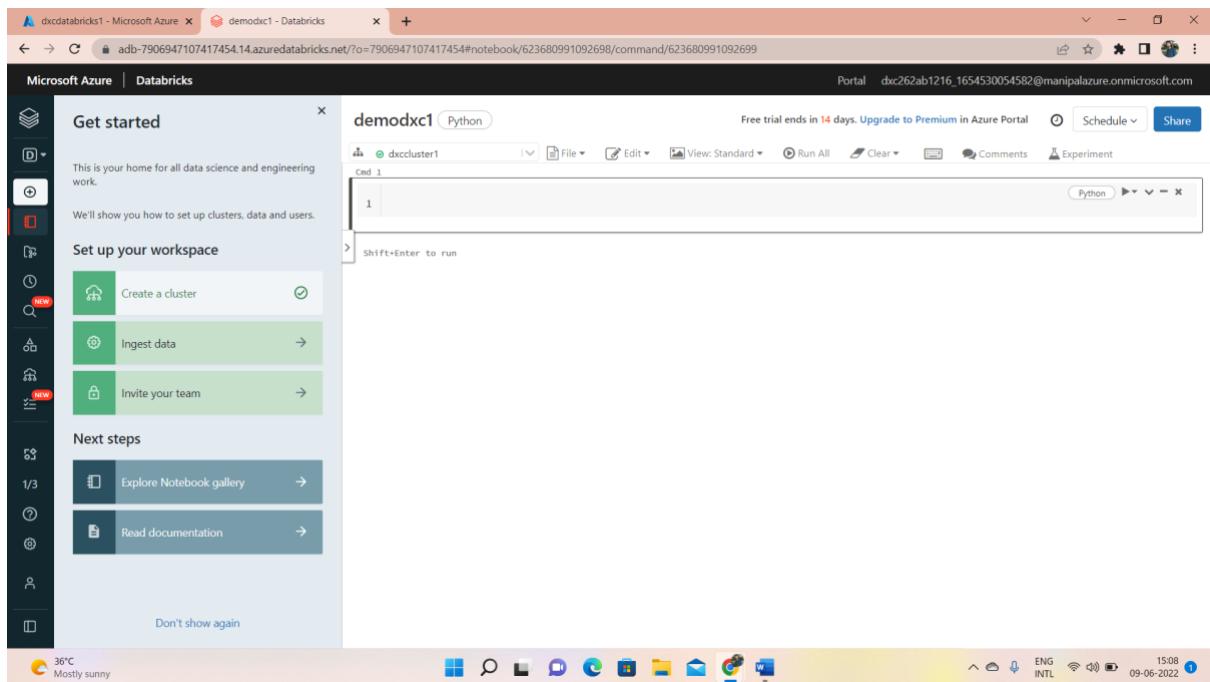
The screenshot shows the Microsoft Azure Deployment blade for a deployment named "dxcorg1\_dxcdatabricks1". The status is "Your deployment is complete". Deployment details include a start time of 6/9/2022, 2:48:51 PM, and a correlation ID of 9f334248-77fd-4af2-8bcc-1b90d6242eb1. A "Go to resource" button is present. On the right, there are promotional cards for Cost Management, Microsoft Defender for Cloud, and Free Microsoft tutorials.

The screenshot shows the Microsoft Azure Databricks Service blade for a service named "dxcdbricks1". The workspace URL is https://adb-7906947107417454.14.azuredatabricks.net. The workspace is active and located in the East US region, using the Azure-DXC262AB12Lab subscription. It is part of the Managed Resource Group databricks-rg-dxcdatabricks1-52thx4hlw3ju. The workspace is in Trial (Premium - 14-Days Free DBUs) tier. A "Launch Workspace" button is available, along with links for Documentation, Getting Started, Import Data from File, and Import Data from Azure Storage.

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

The screenshot shows the Microsoft Azure Databricks workspace interface. The left sidebar includes sections for Data Science & ML, Compute, and Workflows. The main area displays a cluster named "dxccluster1" with details like 10.4 Runtime, 14 GB Active memory, 4 cores Active cores, and 0.75 Active DBU / h. A context menu is open over this cluster, with the "Notebook" option being highlighted.

The screenshot shows the Microsoft Azure Databricks workspace interface. The left sidebar includes sections for Data Science & ML, Compute, and Workflows. A modal dialog box titled "Create Notebook" is open in the center. It contains fields for "Name" (set to "demodxc1"), "Default Language" (set to "Python"), and "Cluster" (set to "dxccluster1"). At the bottom of the dialog are "Cancel" and "Create" buttons.



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

**dxcdatabricks1 - Microsoft Azure** Create New Table - Databricks

adb-7906947107417454.14.azuredatabricks.net/?o=7906947107417454#tables/new

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

Create New Table

Data source ?

Upload File DBFS Other Data Sources

DBFS Target Directory ? /fileStore/tables/ (optional) Select

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

Files ?

Drop files to upload, or click to browse

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.

36°C Mostly sunny

ENG INTL 09-06-2022 15:12

The screenshot shows the 'Create New Table' wizard in the Microsoft Azure Databricks interface. On the left, there's a sidebar with 'Get started' and 'Set up your workspace' sections. The main area is titled 'Create New Table' with tabs for 'Data source', 'Upload File' (which is selected), 'DBFS', and 'Other Data Sources'. A 'DBFS Target Directory' input field contains '/fileStore/tables/'. Below it is a 'Files' section with a placeholder 'Drop files to upload, or click to browse'. A note about using Partner Connect for data ingestion is present. The bottom status bar shows the date and time as 09-06-2022 15:12.

**dxcdatabricks1 - Microsoft Azure** Create New Table - Databricks

adb-7906947107417454.14.azuredatabricks.net/?o=7906947107417454#tables/new

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

Create New Table

Data source ?

Upload File DBFS Other Data Sources

DBFS Target Directory ? /fileStore/tables/ (optional) Select

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

Files ?

1000\_Compa.  
48.1 KB Remove file

✓ File uploaded to /fileStore/tables/1000\_Companies.xlsx

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.

36°C Mostly sunny

ENG INTL 09-06-2022 15:13

The screenshot shows the 'Create New Table' wizard after a file has been uploaded. The 'Files' section now displays '1000\_Compa.' with a size of '48.1 KB' and a 'Remove file' link. A green checkmark indicates the file was successfully uploaded to '/fileStore/tables/'. Below this, there are two buttons: 'Create Table with UI' and 'Create Table in Notebook'. The rest of the interface and status bar are identical to the first screenshot.

dxcdatabricks1 - Microsoft Azure | 2022-06-09 - DBFS Example - Databricks

adb-7906947107417454.14.azuredatabricks.net/?o=7906947107417454#notebook/623680991092700/command/623680991092701

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

Detached Cell 1

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.xlsx"
3 file_type = "xlsx"
4
5 # CSV options
6 infer_schema = "false"
7 first_row_is_header = "false"
8 delimiter = ","
9
10 # The applied options are for CSV files. For other file types, these will be ignored.
11 df = spark.read.format(file_type) \
12 .option("inferSchema", infer_schema) \
13 .option("header", first_row_is_header) \
14 .option("sep", delimiter) \
15 .load(file_location)
16
17 display(df)
```

Cmd 2

Python

ENG INTL 09-06-2022 15:15

dxcdatabricks1 - Microsoft Azure | 2022-06-09 - DBFS Example - Databricks

adb-7906947107417454.14.azuredatabricks.net/?o=7906947107417454#notebook/623680991092700/command/623680991092701

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

Detached Cell 1

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.xlsx"
3 file_type = "xlsx"
4
5 # CSV options
6 infer_schema = "false"
7 first_row_is_header = "false"
8 delimiter = ","
9
10 # The applied options are for CSV files. For other file types, these will be ignored.
11 df = spark.read.format(file_type) \
12 .option("inferSchema", infer_schema) \
13 .option("header", first_row_is_header) \
14 .option("sep", delimiter) \
15 .load(file_location)
16
17 display(df)
```

Cmd 2

Python

ENG INTL 09-06-2022 15:16

**2022-06-09 - DBFS Example** Python

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

Portal dxc26ab1216\_1654530054582@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

Don't show again

Cmd 3

```
# Create a view or table
temp_table_name = "Companies_xlsx"
df.createOrReplaceTempView(temp_table_name)
```

Cmd 4

```
%sql
/* Query the created temp table in a SQL cell */
select * from `1000_Companies_xlsx`
```

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.
# To do so, choose your table name and uncomment the bottom line.
permanent_table_name = "1000_Companies_xlsx"
# df.write.format("parquet").saveAsTable(permanent_table_name)
```

36°C Mostly sunny 15:17 ENG INTL 09-06-2022

**Create New Table**

Preview Table

**Specify Table Attributes**

Specify the Table Name, Database and Schema to add this to the data UI for other users to access

**Invalid characters (.,{} =& found in the columns) have been replaced with underscores**

Table Name companies\_csv

Create in Database default

File Type CSV

Column Delimiter ,

First row is header

Infer schema

Multi-line

**Create Table**

**Create Table in Notebook**

Table Preview

R_D_Spend	Administration	Marketing_Spend	State	Profit
FLOAT	STRING	STRING	Column Name	STRING
165349.2	136897.8	471784.1	New York	192261.83
162597.7	151377.59	443898.53	California	191792.06
153441.51	101145.55	407934.54	Florida	191050.39
144372.41	118671.85	383199.62	New York	182901.99
142107.34	91391.77	366168.42	Florida	166187.94
131876.9	99814.71	362861.36	New York	156991.12
134615.46	147198.87	127716.82	California	156122.51

36°C Mostly sunny 15:31 ENG INTL 09-06-2022

The screenshot shows the Microsoft Azure Databricks interface. At the top, there are two tabs: "dxcdatabricks1 - Microsoft Azure" and "Create New Table - Databricks". The URL in the address bar is "adb-7906947107417454.14.azuredatabricks.net/?o=7906947107417454#table/hive\_metastore/default/companies\_csv". The title bar says "Microsoft Azure | Databricks". On the right, it says "Portal dxc262ab1216\_1654530054582@manipalazure.onmicrosoft.com" and "Free trial ends in 14 days. Upgrade to Premium in Azure Portal".  
**Description:**  
Created at: 2022-06-09 10:01:29  
Last modified: 2022-06-09 10:01:35  
Partition columns:  
Number of files: 1  
Size: 37.7 kB  
**Schema:**

	col_name	data_type	comment
1	R_D_Spend	float	
2	Administration	string	
3	Marketing_Spend	string	
4	State	string	
5	Profit	string	
6			
7	# Partitioning		

Showing all 8 rows.

**Sample Data:**

	R_D_Spend	Administration	Marketing_Spend	State	Profit
1	165349.2	136897.8	471784.1	New York	192261.83
2	162597.7	151377.59	443898.53	California	191792.06
3	153441.52	101145.55	407934.54	Florida	191050.39
4	144372.4	118671.85	383199.62	New York	182901.99
5	142107.34	91391.77	366168.42	Florida	166187.94

At the bottom, there is a taskbar with icons for Start, Search, Task View, File Explorer, Edge, Task Manager, Mail, Photos, and Google Sheets. The system tray shows the date (09-06-2022), time (15:31), battery level (ENG INTL), and weather (36°C, Mostly sunny).

The screenshot shows the Microsoft Azure Databricks interface. At the top, there are two tabs: "dxcdatabricks1 - Microsoft Azure" and "Workflows - Databricks". The URL in the address bar is "adb-7906947107417454.14.azuredatabricks.net/?o=7906947107417454#job/runs". The title bar says "Microsoft Azure | Databricks". On the right, it says "Portal dxc262ab1216\_1654530054582@manipalazure.onmicrosoft.com" and "Free trial ends in 14 days. Upgrade to Premium in Azure Portal".  
**Workflows**  
**Jobs**   **Job runs**   **Delta Live Tables**  

Start time	Job	Run as	Launched	Duration	Status	Actions
Jun 9 2022, 15:34 PM IST	dxcnote1	dxc262ab1216_1654530054582@...	Manually	3m 9s	Running	[More]

At the bottom, there is a taskbar with icons for Start, Search, Task View, File Explorer, Edge, Task Manager, Mail, Photos, and Google Sheets. The system tray shows the date (09-06-2022), time (15:38), battery level (ENG INTL), and weather (35°C, Mostly sunny).

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

**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

36°C Mostly sunny

Data Science & Engineering

Notebook

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

Create a notebook

Guide: Quickstart tutorial

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

Start tutorial

Data import

Quickly import data, preview its schema, create a table, and query i...

Browse files

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

Partner Connect

Fivetran.dlt  
Tableau - Power BI  
View all partners

Recent

Name Last viewed

There are no recent yet

Documentation

Get started guide  
This tutorial gets you going with Azure Databricks  
Data Science & Engineering

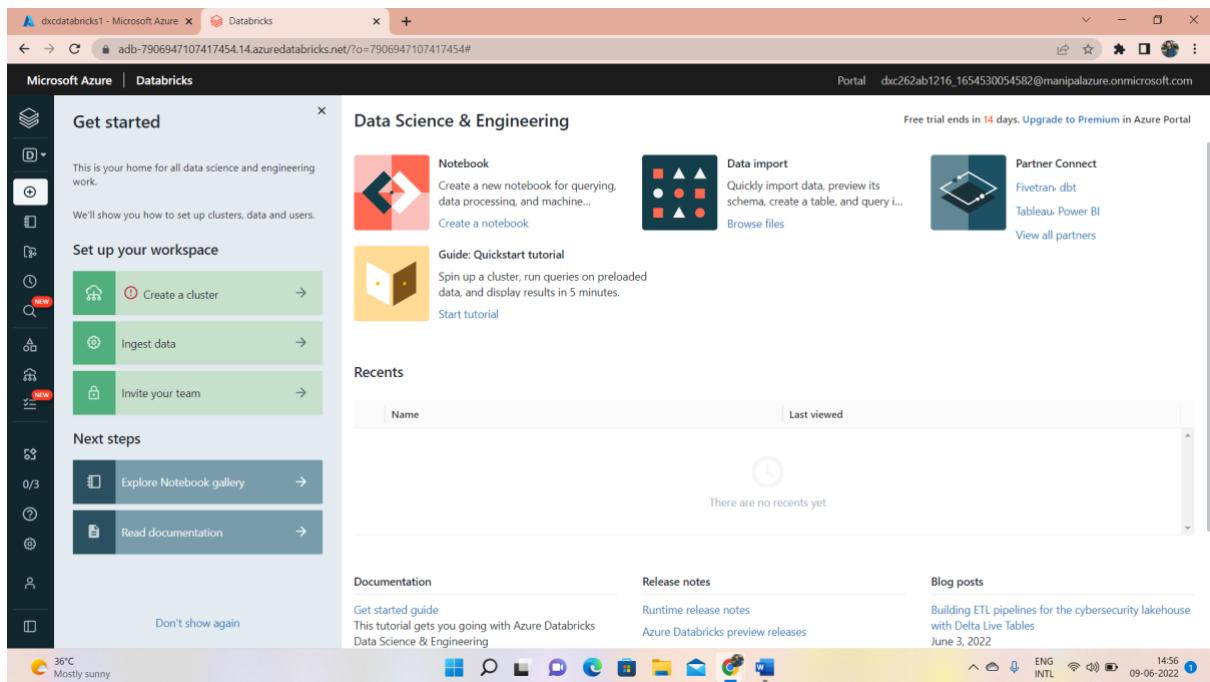
Release notes

Runtime release notes  
Azure Databricks preview releases

Blog posts

Building ETL pipelines for the cybersecurity lakehouse with Delta Live Tables  
June 3, 2022

ENG INTL 09-06-2022 14:56



**Microsoft Azure | Databricks**

Compute - Databricks

All-purpose clusters Job clusters Pools Cluster policies

Create Cluster

Created by me Accessible by me Q Search within clusters you have access to

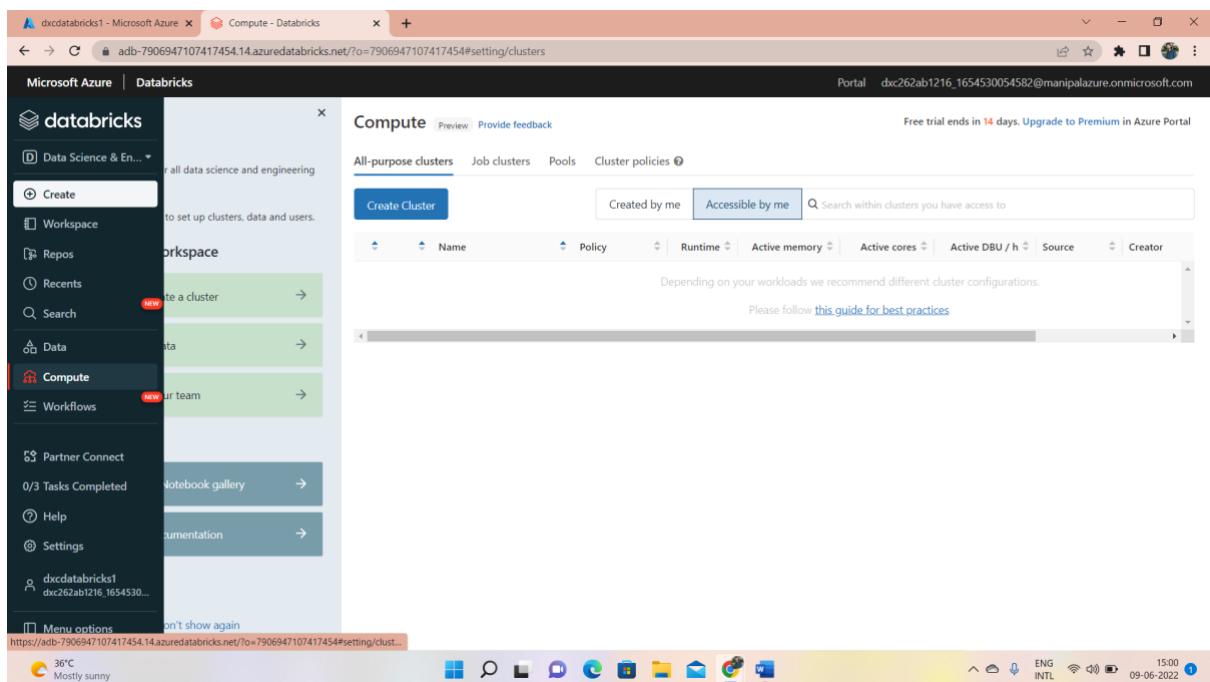
Name Policy Runtime Active memory Active cores Active DBU / h Source Creator

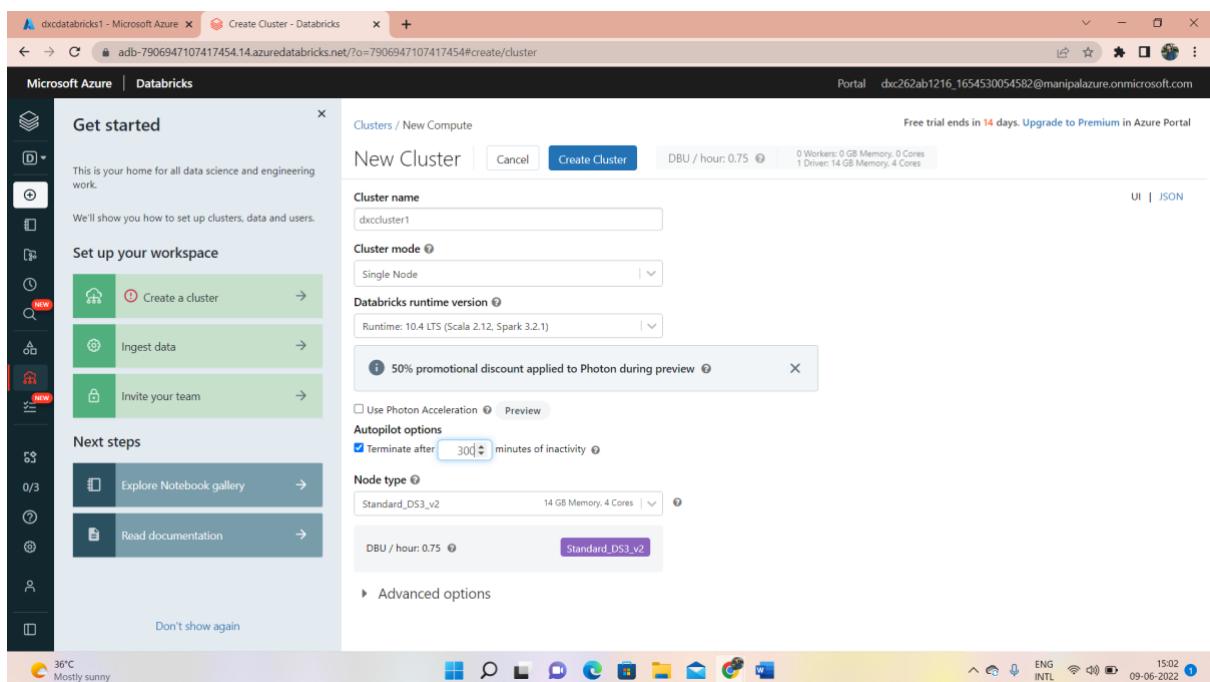
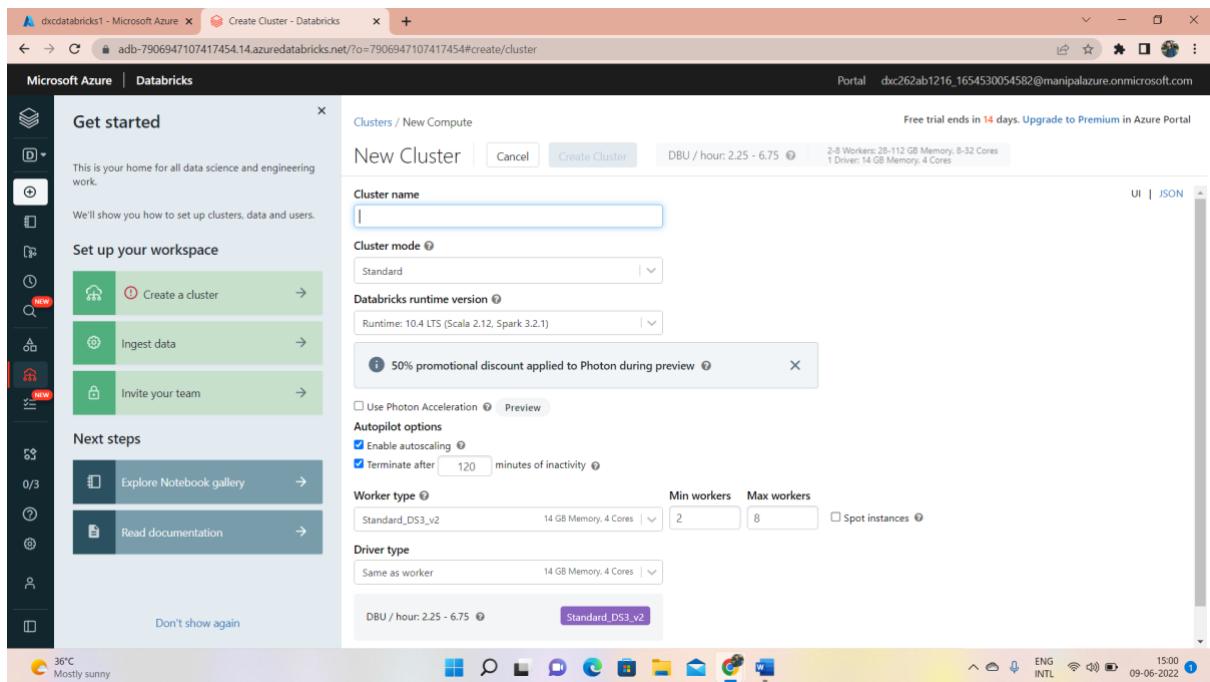
Depending on your workloads we recommend different cluster configurations.

Please follow: [this guide for best practices](#)

Menu options

36°C Mostly sunny





**dxddatabricks1 - Microsoft Azure** Create Cluster - Databricks

adx-7906947107417454.14.azuredatabricks.net/?o=7906947107417454#setting/clusters/0609-093207-5j4eh7q/configuration

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

Clusters / dxcluster1

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

dxcluster1

Configuration Notebooks Libraries Event log Spark UI Driver logs Metrics Apps Spark cluster UI - Master

More Edit Terminate

Policy: Unrestricted

Cluster mode: Single Node

Databricks Runtime Version: 10.4 LTS (includes Apache Spark 3.2.1, Scala 2.12)

Use Photon Acceleration Preview

Autopilot options:  Terminate after 300 minutes of inactivity

Node type: Standard\_DS3\_v2 (14 GB Memory, 4 Cores)

DBU / hour: 0.75 Standard\_DS3\_v2

Advanced options

UI | JSON

36°C Mostly sunny

ENG INTL 09-06-2022 15:04

**dxddatabricks1 - Microsoft Azure** Compute - Databricks

adx-7906947107417454.14.azuredatabricks.net/?o=7906947107417454#setting/clusters

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

Compute

Preview Provide feedback

All-purpose clusters Job clusters Pools Cluster policies

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

Created by me Accessible by me Search within clusters you have access to

Name	Policy	Runtime	Active memory	Active cores	Active DBU / h	Source	Creator
dxcluster1	-	10.4	14 GB	4 cores	0.75	UI	dxc262ab1

36°C Mostly sunny

ENG INTL 09-06-2022 15:05