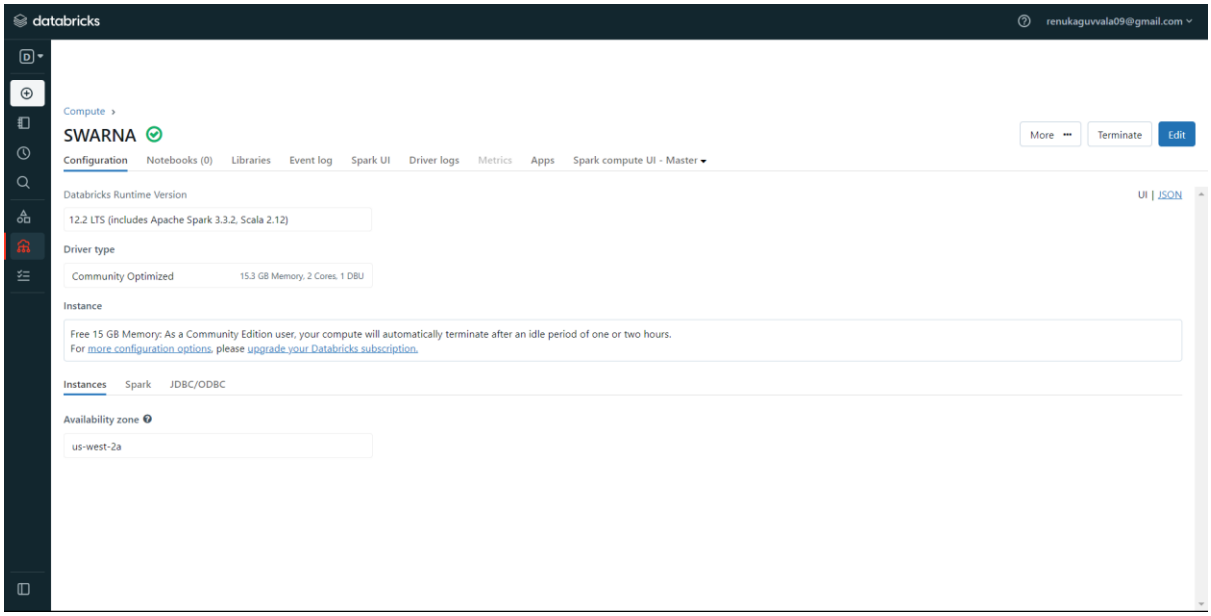
AZURE DATARICKS CODING ASSESSMENT

NAME-SWARNA LATHA GUVVALA

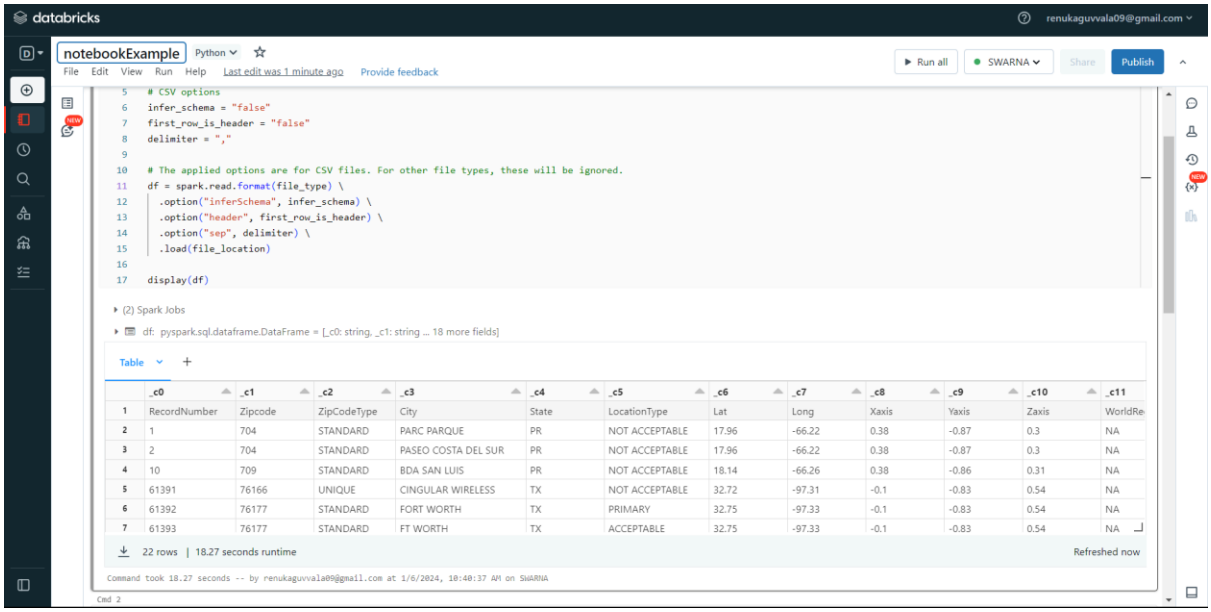
DATE-06/01/2024

1.create a cluster &attach the notebook to the cluster and run all commands in the notebook &creates a dataframe from a databricks dataset&create a visualizations in databricks notebooks &rename, duplicate, or remove a visualization or data profile?

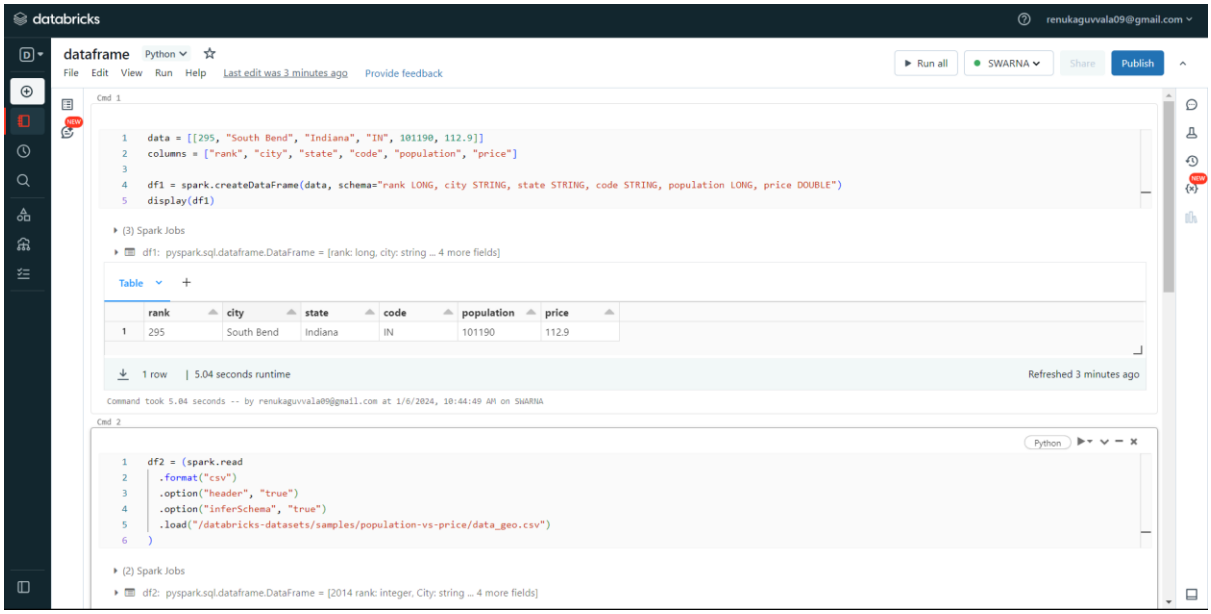
Create a Cluster:



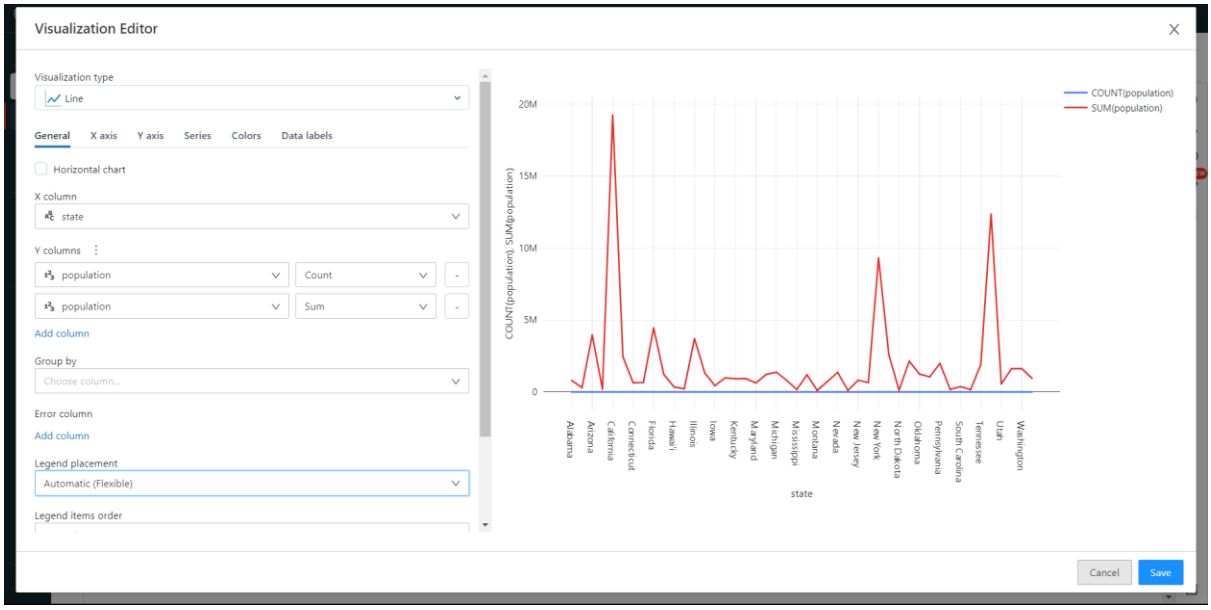
Attach the notebook to the cluster and run all commands in the notebook:

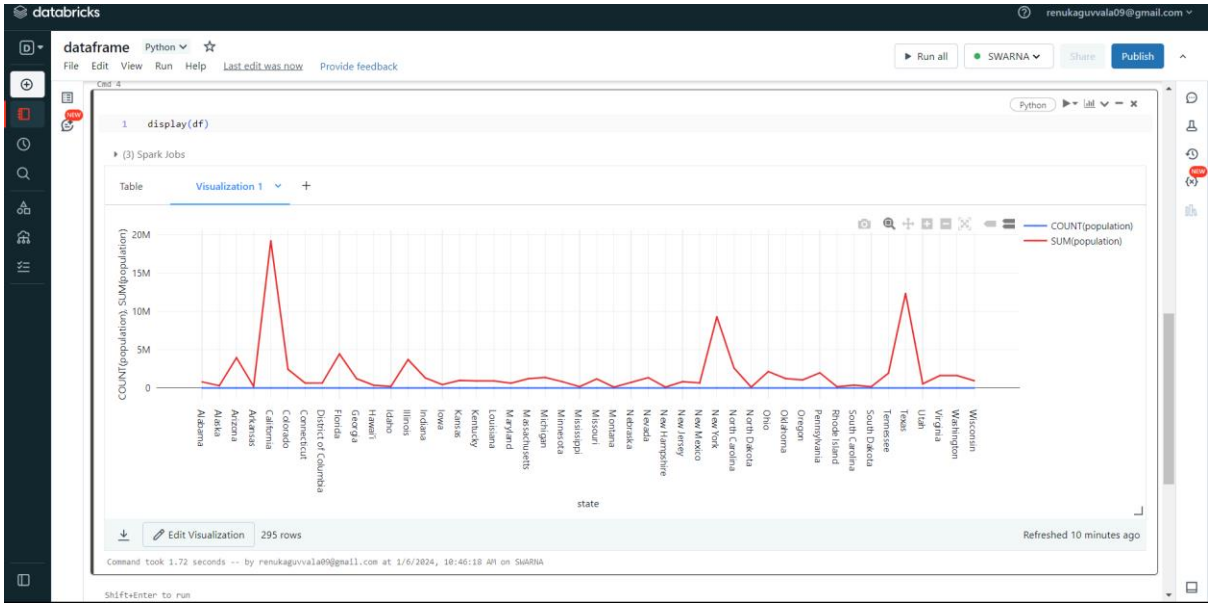


Creates a dataframe from a databricks dataset:

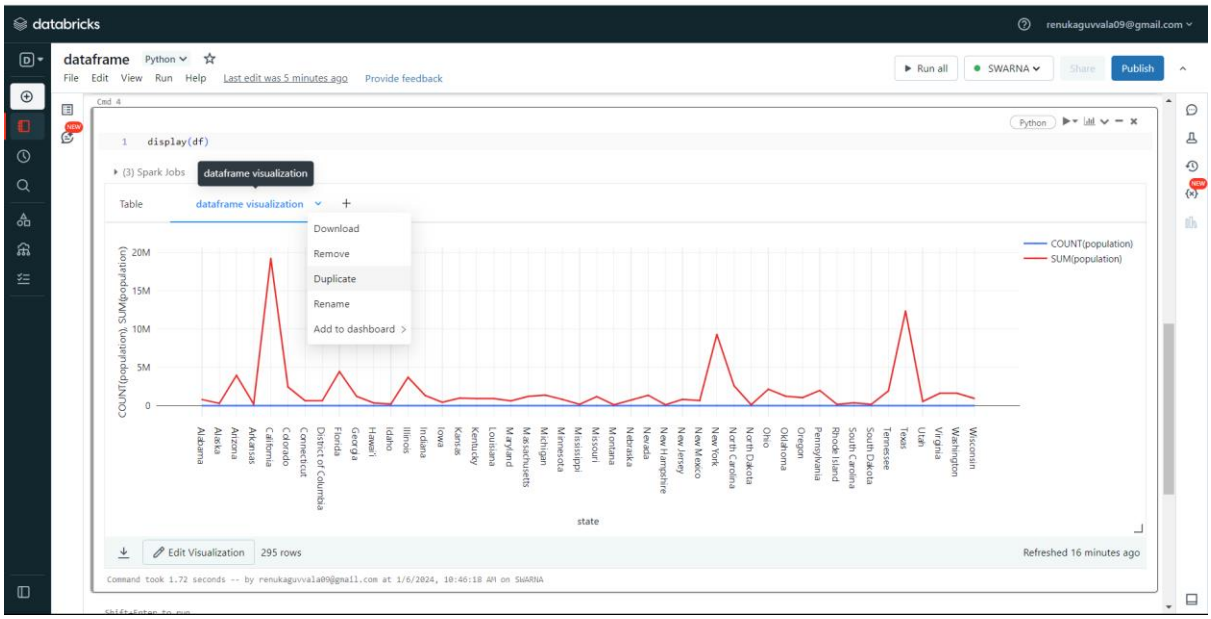


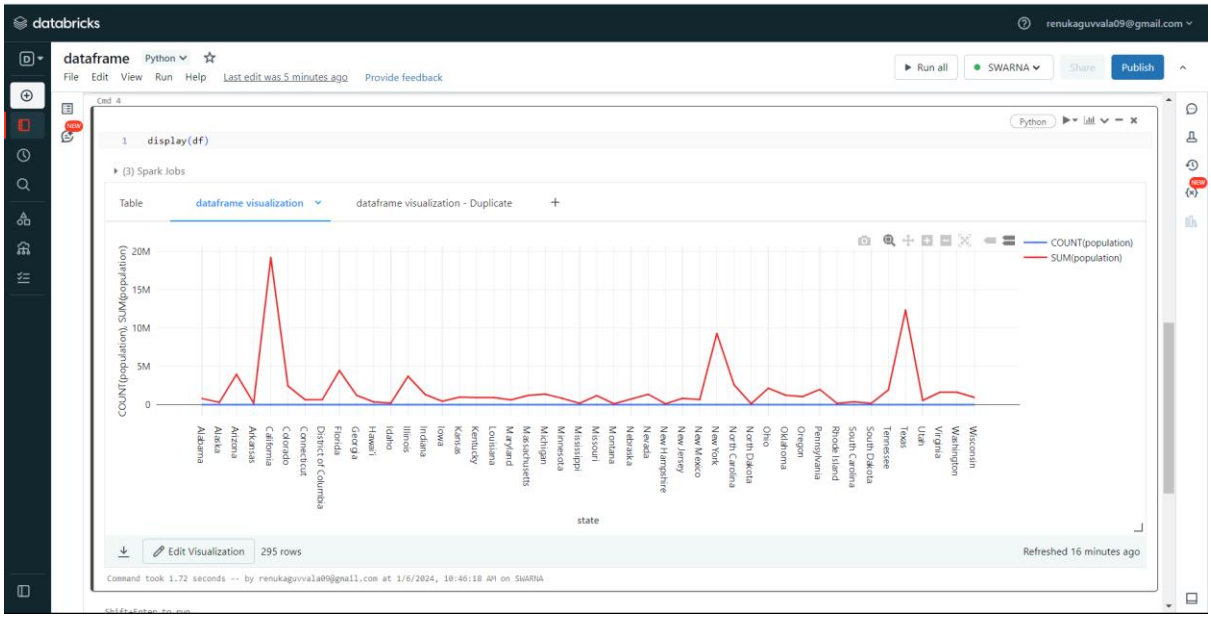
Create a visualizations in databricks notebooks:

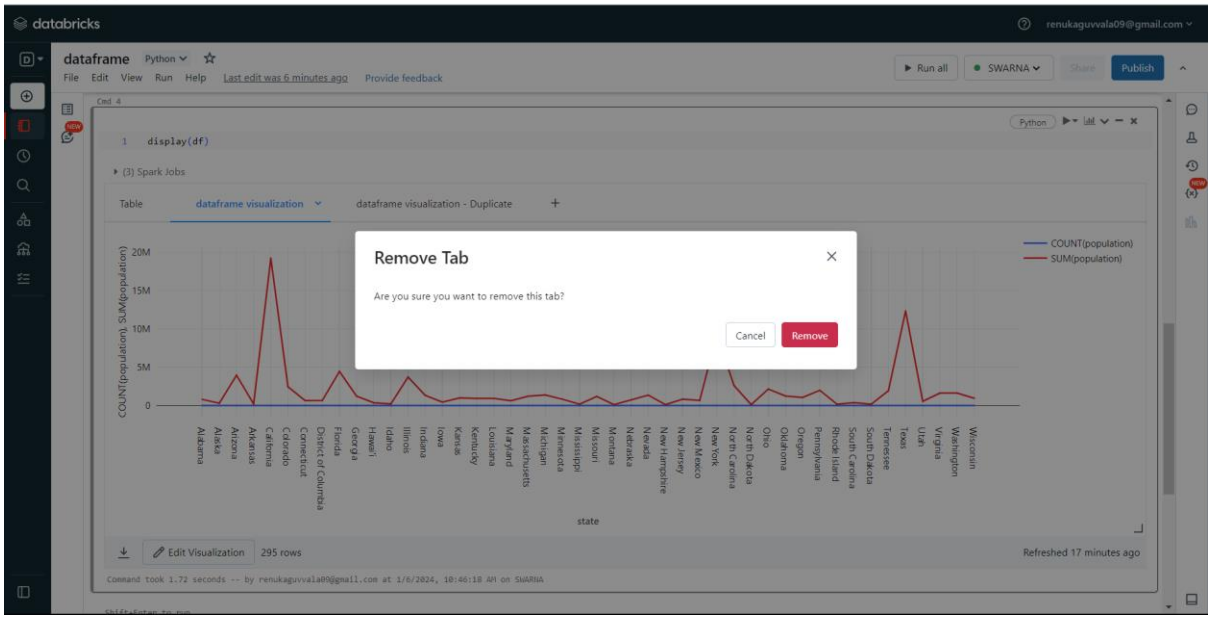


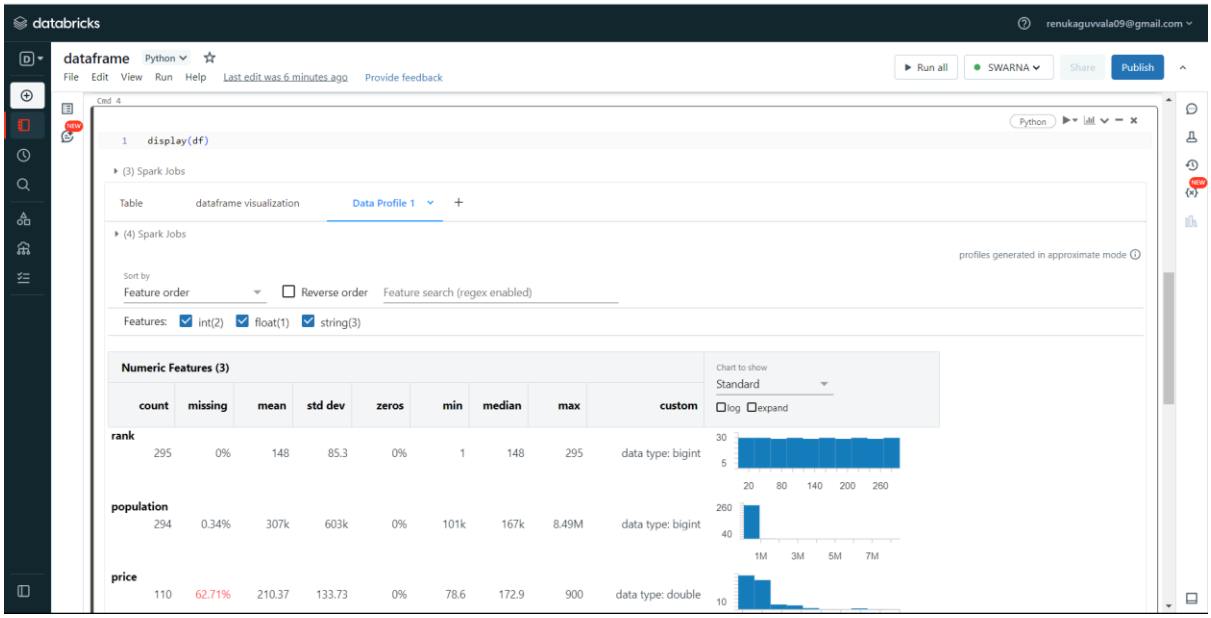


Rename, duplicate, or remove a visualization or data profile:




databricks
renukaguvvala09@gmail.com
dataframe
Python
File Edit View Run Help
Last edit was 5 minutes ago
Provide feedback
Run all
SWARNA
Share
Publish
Cmd 4
1 display(df)
(3) Spark Jobs
Table
dataframe visualization
dataframe visualization - Duplicate
COUNT(population)
SUM(population)
COUNT(population), SUM(population)
20M
15M
10M
5M
0
Alabama Alaska Arizona Arkansas California Colorado Connecticut District of Columbia Florida Georgia Hawaii Idaho Illinois Indiana Iowa Kansas Kentucky Louisiana Maryland Massachusetts Michigan Minnesota Mississippi Missouri Montana Nebraska Nevada New Hampshire New Jersey New Mexico New York North Carolina North Dakota Ohio Oklahoma Oregon Pennsylvania Rhode Island South Carolina South Dakota Tennessee Texas Utah Virginia Washington Wisconsin
state
Edit Visualization
295 rows
Refreshed 16 minutes ago
Command took 1.72 seconds -- by renukaguvvala09@gmail.com at 1/6/2024, 10:46:18 AM on SWARNA


databricks
renukaguvvala09@gmail.com
dataframe
Python
File Edit View Run Help
Last edit was 6 minutes ago
Provide feedback
Run all
SWARNA
Share
Publish
Cmd 4
1 display(df)
(3) Spark Jobs
Table
dataframe visualization
dataframe visualization - Duplicate
Remove Tab
Are you sure you want to remove this tab?
Cancel
Remove
COUNT(population)
SUM(population)
COUNT(population), SUM(population)
20M
15M
10M
5M
0
Alabama Alaska Arizona Arkansas California Colorado Connecticut District of Columbia Florida Georgia Hawaii Idaho Illinois Indiana Iowa Kansas Kentucky Louisiana Maryland Massachusetts Michigan Minnesota Mississippi Missouri Montana Nebraska Nevada New Hampshire New Jersey New Mexico New York North Carolina North Dakota Ohio Oklahoma Oregon Pennsylvania Rhode Island South Carolina South Dakota Tennessee Texas Utah Virginia Washington Wisconsin
state
Edit Visualization
295 rows
Refreshed 17 minutes ago
Command took 1.72 seconds -- by renukaguvvala09@gmail.com at 1/6/2024, 10:46:18 AM on SWARNA



2.Explain the copy activity in Azure data factory.

In Azure Data Factory(ADF),the copy activity is a crucial component that enables the movement of data from source to destination. The primary purpose of the copy activity is to facilitate the extraction, transformation, and loading (ETL) of data.

Here is a breakdown of the Key aspects of the copy activity in Azure Data Factory:

1.Source and Destination:

- Source Dataset: This is the data store from which data is to be copied .It could be an on-premises database, a cloud-based storage solution (like Azure Blob Storage or Azure SQL Database)
- Destination Dataset: This is the target data store where the data will be copied .Similar to the source dataset, it can be on-premises or cloud-based data store.

2.Data Movement:

- The copy activity efficiently moves data from source to destination. It handles data transfer in a secure and performant manner, optimizing the process for large-scale data movement.

3.Supported Formats:

- The copy activity supports various date formats ,including CSV,JSON...etc. It can handle structured, semi-structured, and unstructured data.

4.Mapping and Transformation:

- The copy activity allows for simple data mappings between source and destination columns. It also supports limited transformations during the copying process.

5.Integration Runtimes:

- Integration Runtimes are used to provide the necessary compute infrastructure to execute the copy activity .Azure data factory supports various types of integration runtimes, including Azure .These runtimes ensure that data movement occurs securely and efficiently.

6.Monitoring and Logging:

- Azure Data Factory provides monitoring capabilities for the copy activity. You can monitor the process, view logs and gain insights into the success or failure of the data movement operation.

7.Error Handling:

- The copy activity includes features for handling errors during the data movement process. This includes retry policies and configurable settings to manage how the system responds to failures.

8.Scheduling and Orchestration:

- The copy activity is often used within pipelines ,where it can be scheduled to run at specific intervals or triggered by external events. This enables you to orchestrate complex ETL workflows within Azure Data Factory.

In summary, the copy activity in Azure Data Factory is a versatile tool for efficiently moving and transforming data between various sources and destinations ,making it a key component for building data integration solutions in the Azure cloud.