**Daily assessment** - 6

**Azure data bricks Assessment** 6

**Date:** 04/01/2024

**Name :** Abhiram Basa

**Topics**

* Run your first ETL workload on Azure Databricks.
* Scheduled jobs.
* Executing jobs.

**ETL :** ETL stands for extract, transform and load the data.

* These ETL can be done manually or by scheduling jobs.

**JOBS:** These are automatic mechanism to run code in Azure data bricks.

* These can be scheduled at particular time or can run at instant.

**Auto Loader:** It helps to load any format of data into delta table.

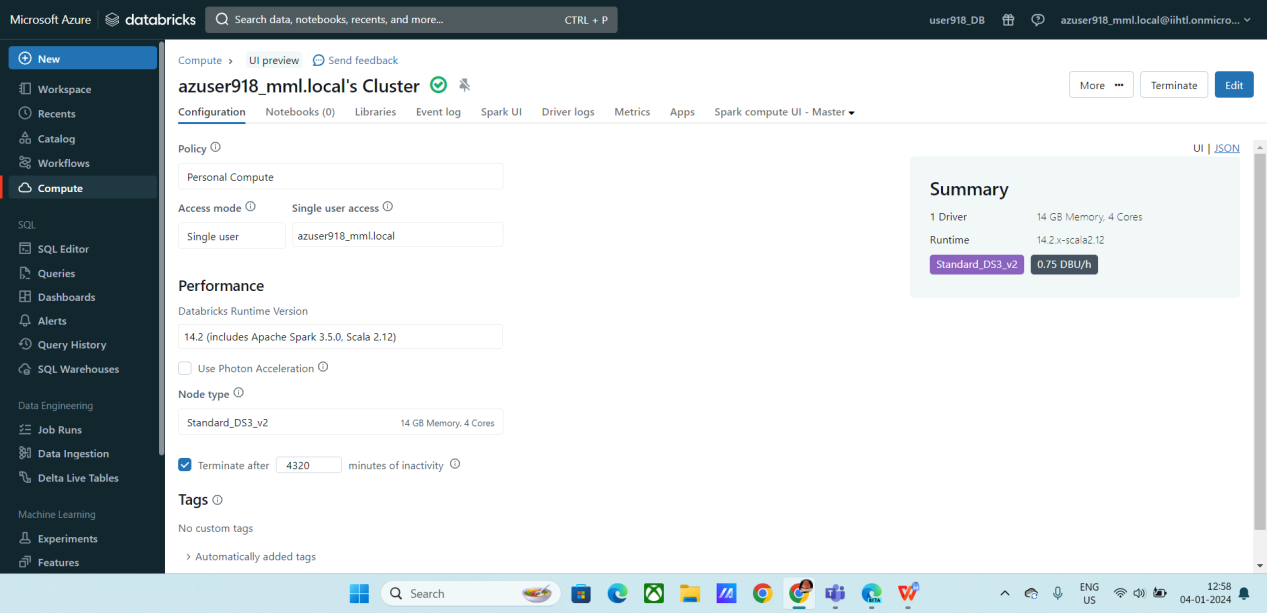
It used in data ingestion process.

In order to execute we need to create a cluster first.

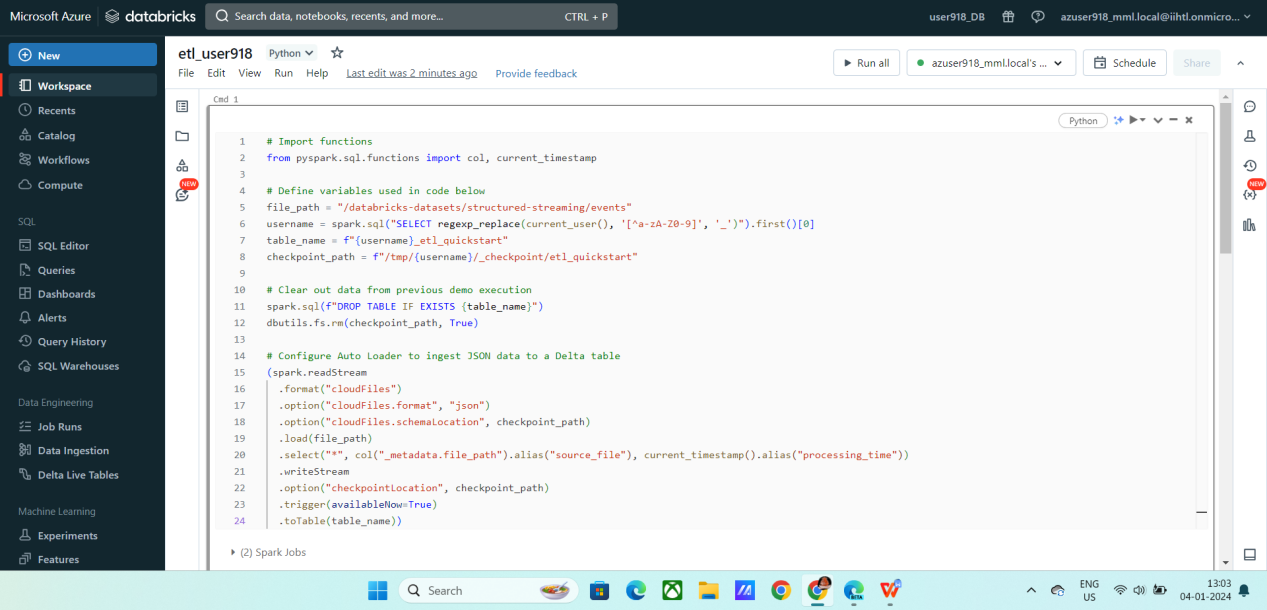
**Cluster:** It is a web based interface containing a series of commands to run on azure data bricks which operates on tables,files, etc.

* It is used to run jobs in Azure data bricks.

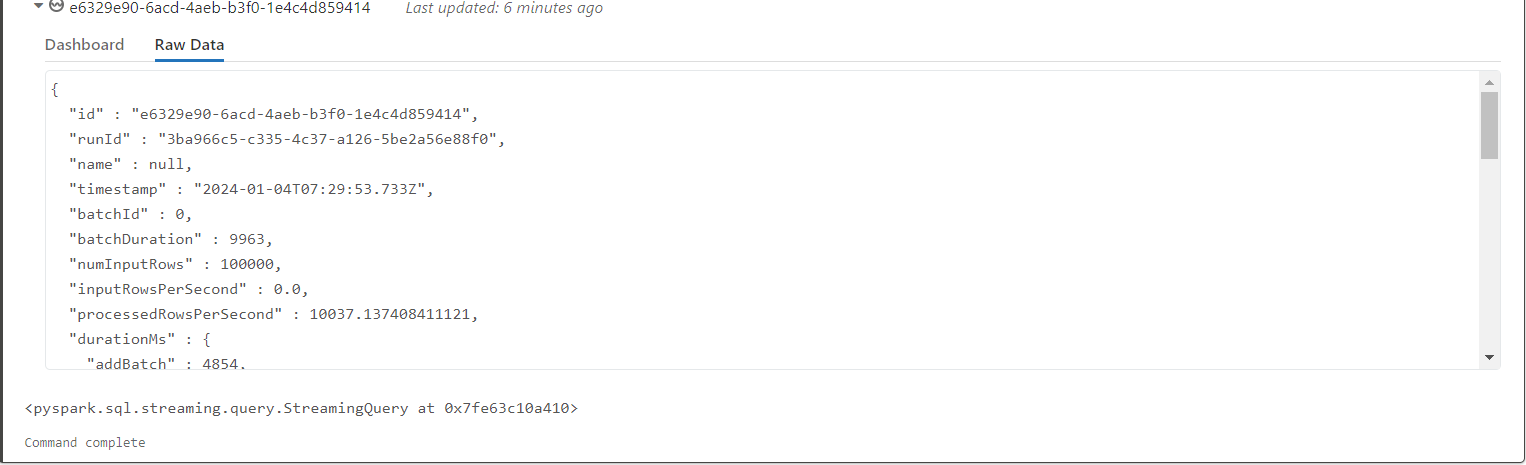
**Creating a cluster**:



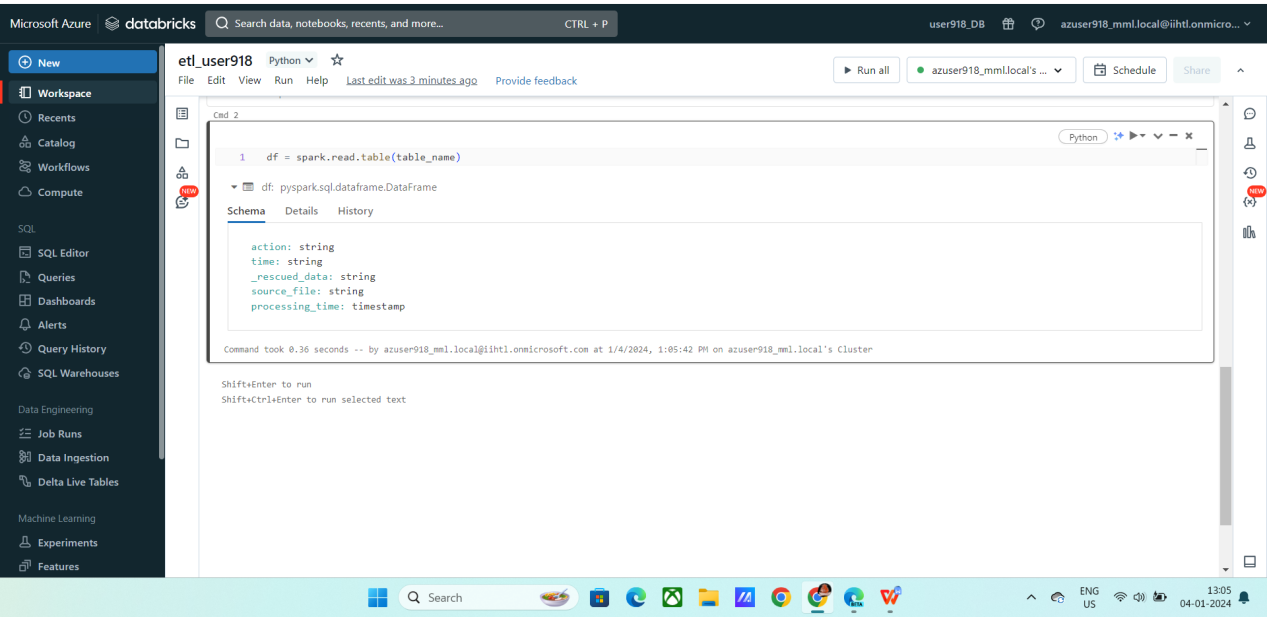
* Now Create a notebook
* Inside the note book paste the following code



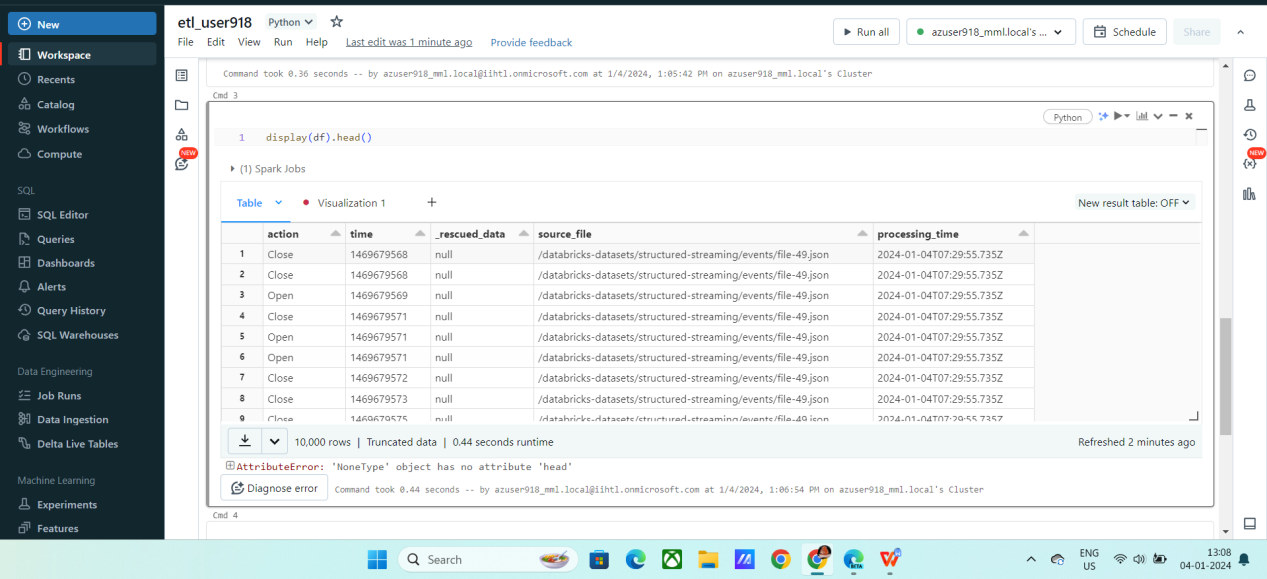
* It will ingest the data into delta lakes using Auto loader



* Viewing the created table

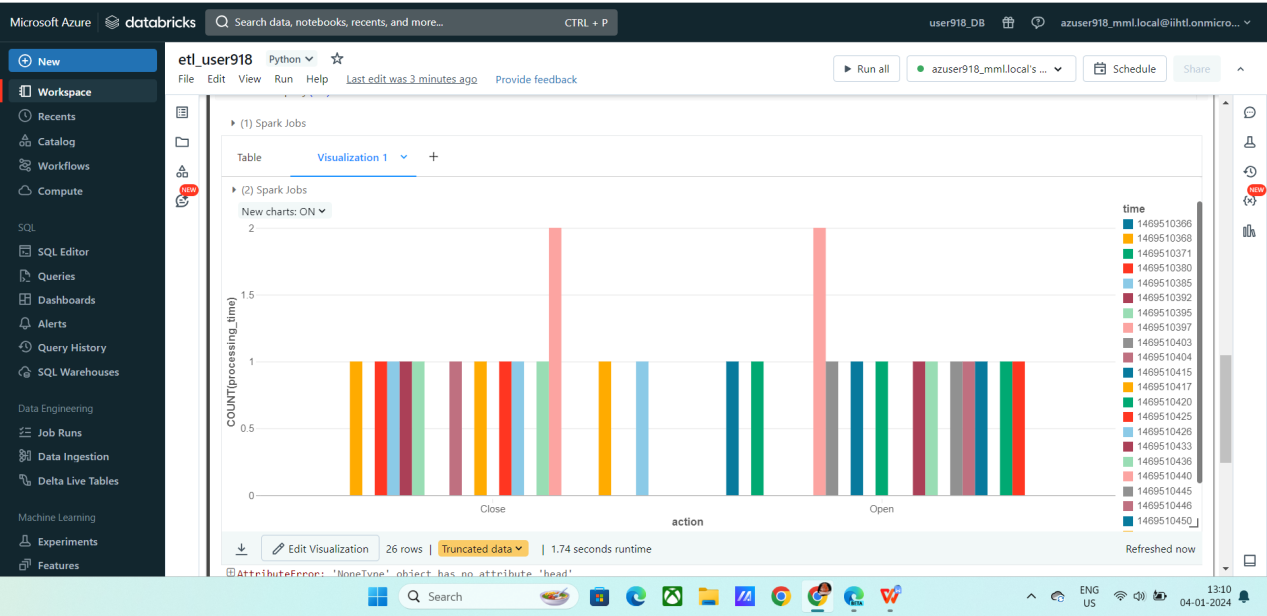


* Displaying the data



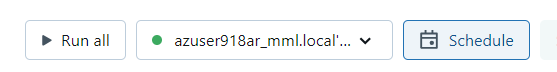
* Created a visualization

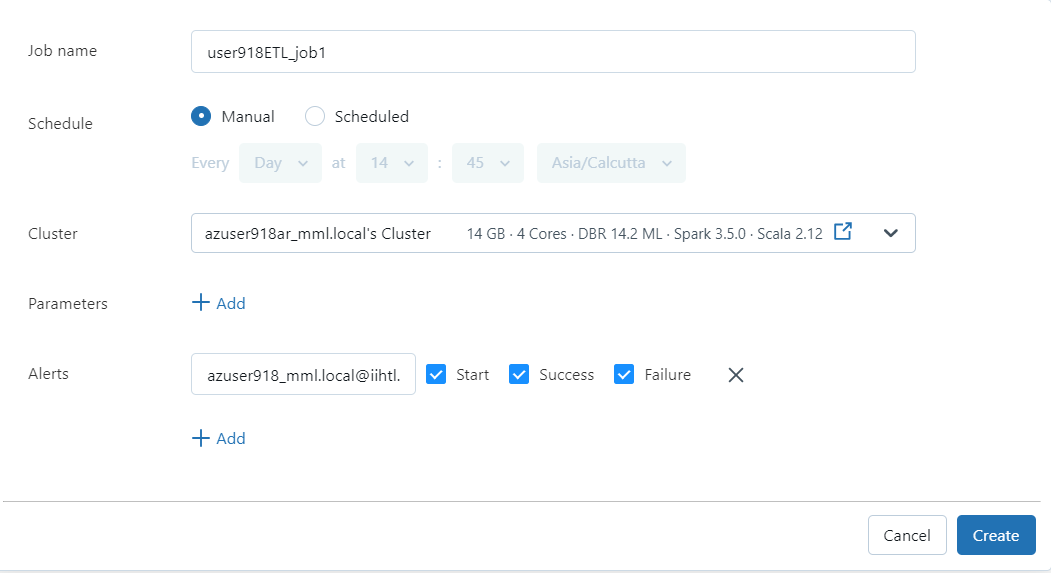
Used processing time on X-axis and action type on Y axis

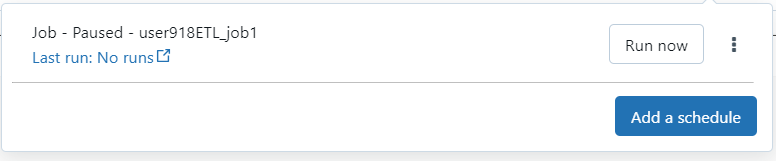


**Schedule a job :**

* Click on schedule on top right corner.
* Choose manual.
* Click create

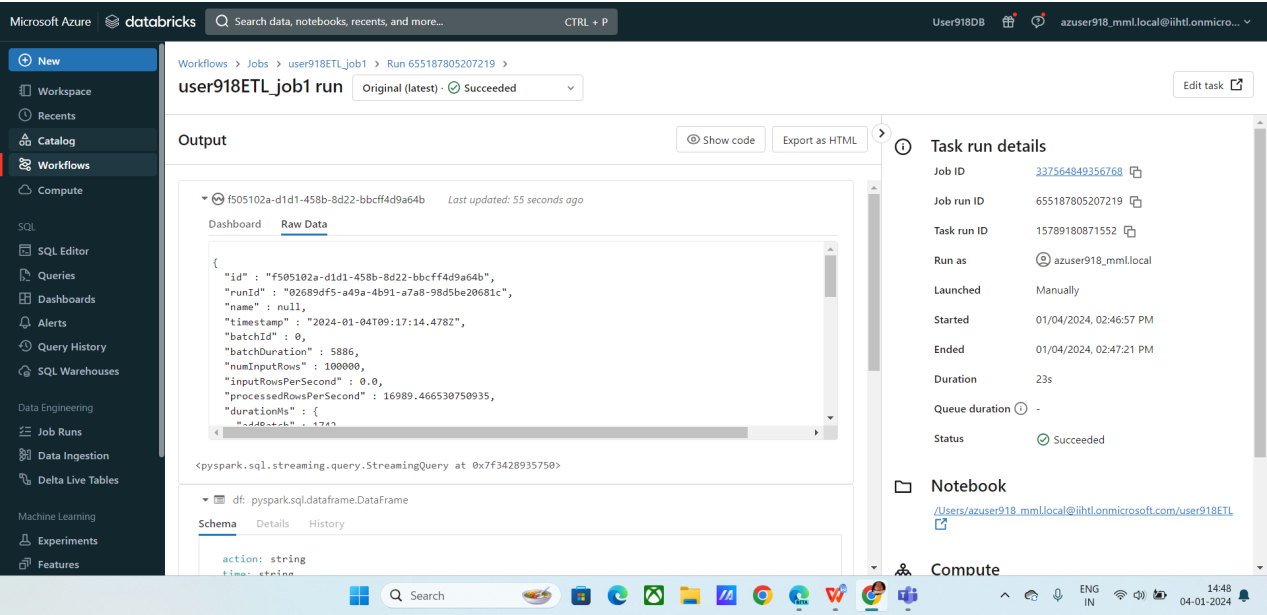


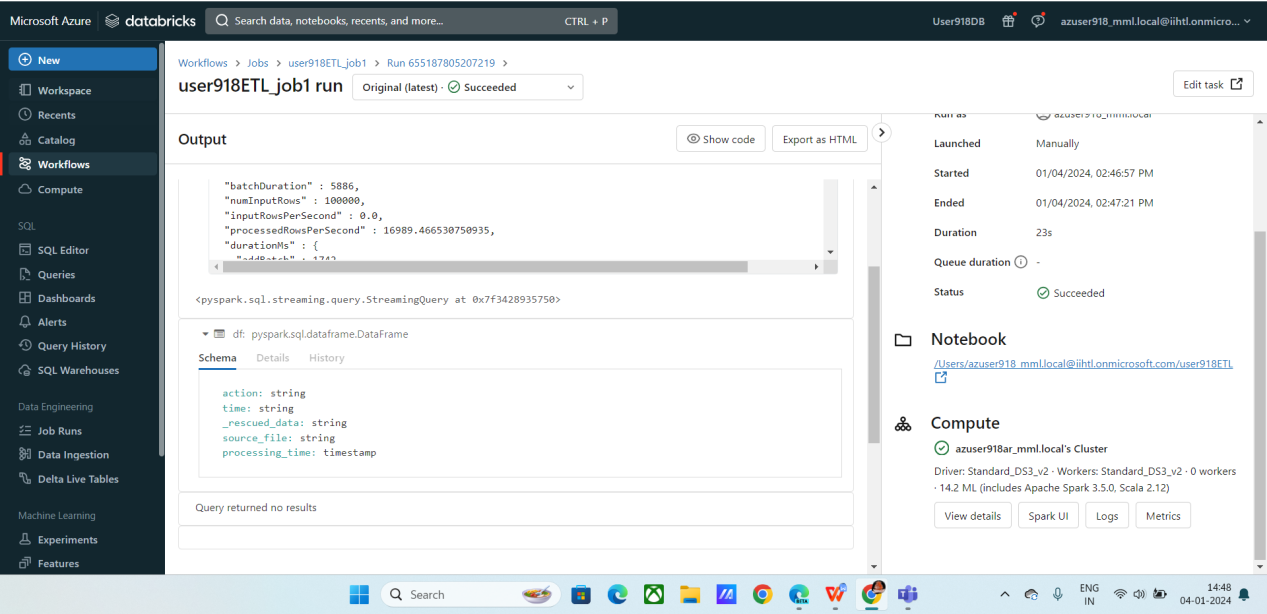


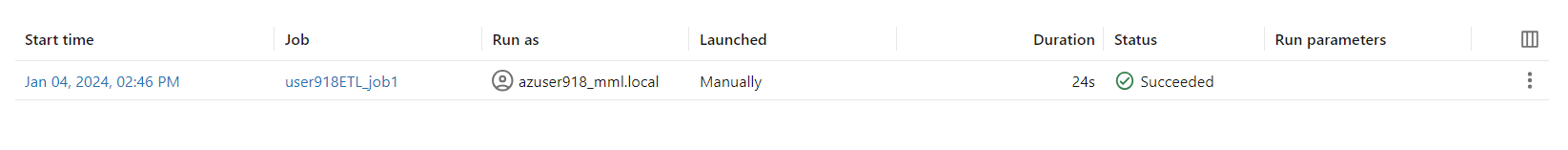


You can view the output, by going to job runs section.

**Output:**







Ex: 2 :

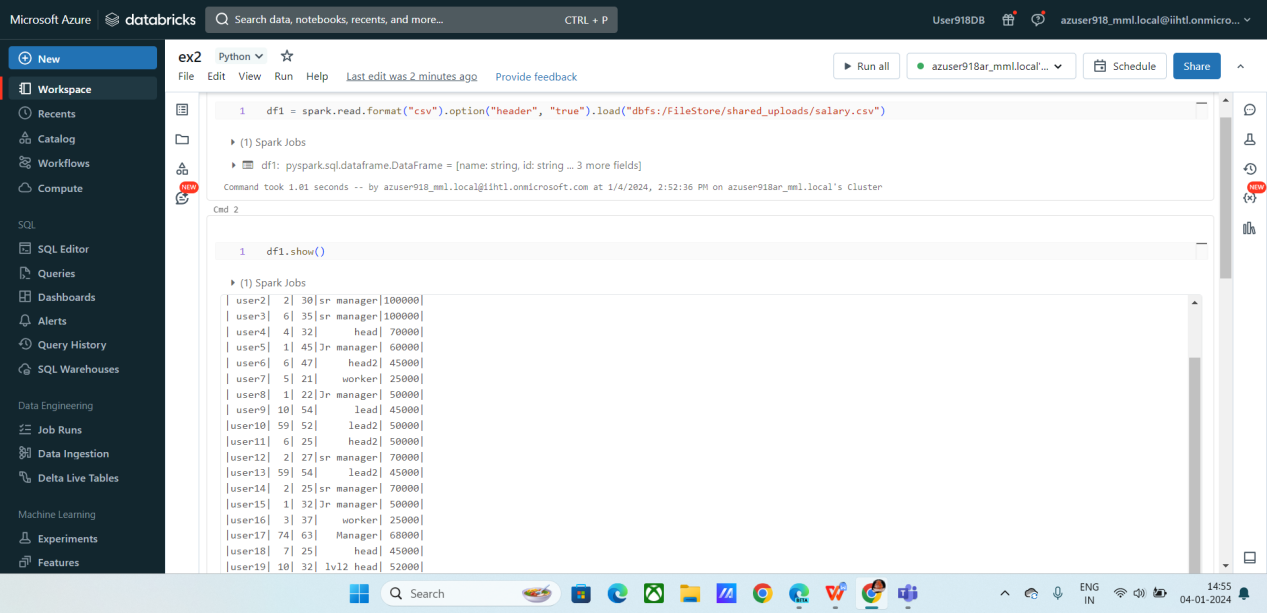
* In order work with the uploaded data
* Create a note book
* Click on upload
* Click on upload to dbfs
* Select the location.
* Upload the files

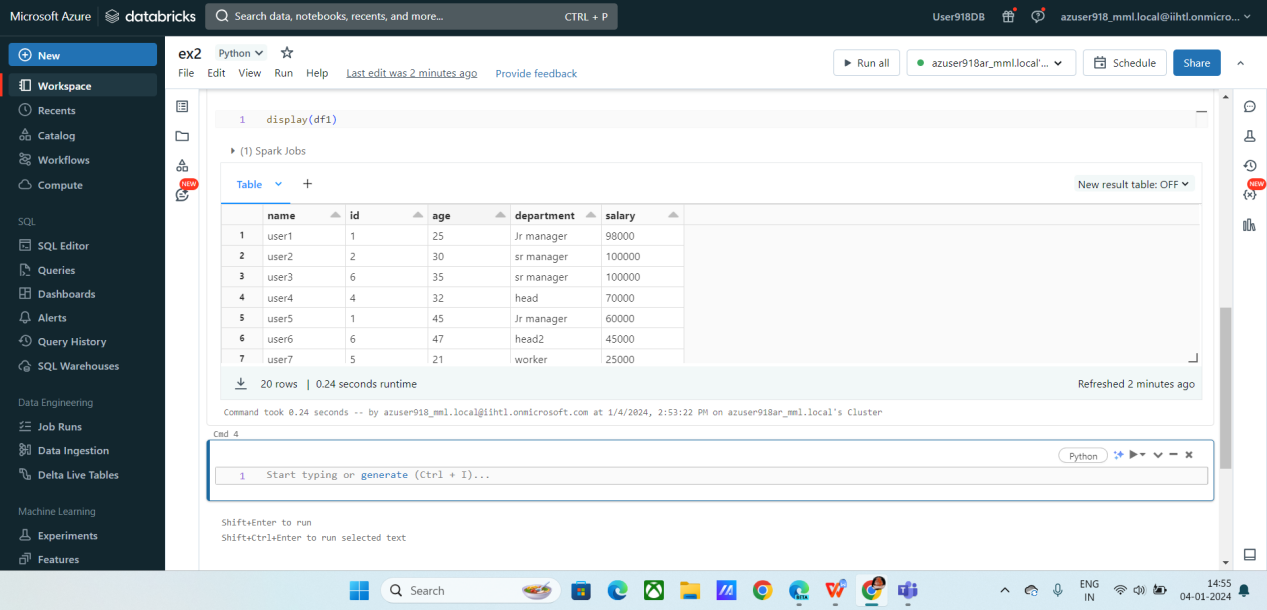
Access it in your notebook using below syntax

**Syntax:**

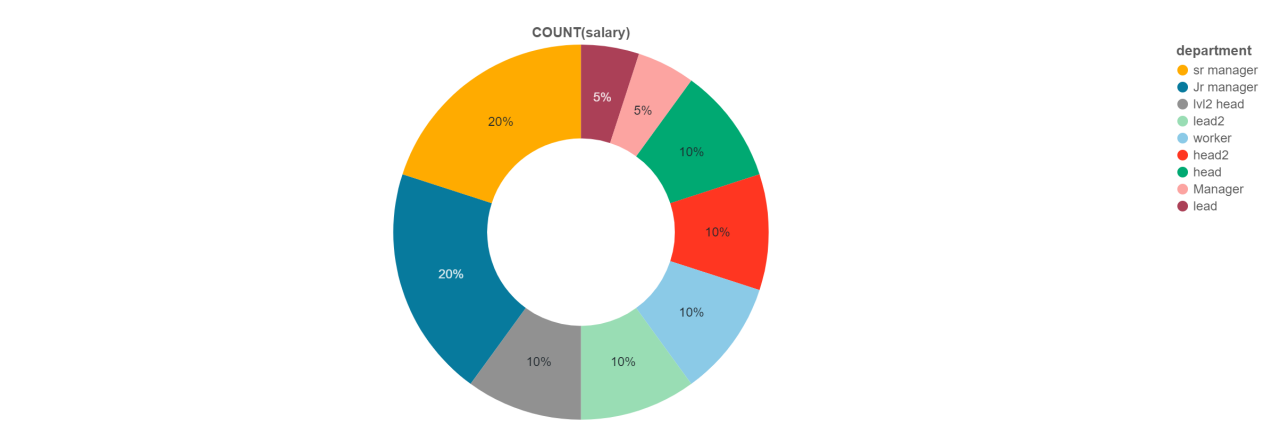
df1 = spark.read.format("csv").option("header", "true").load("dbfs:/FileStore/shared\_uploads/salary.csv")

Path: dbfs:/FileStore/shared\_uploads/salary.csv





**Created a visualization based on department and its salaries.**



**Analysis:**

As you can as per the data 20% are sr manager with high rate of salaries.