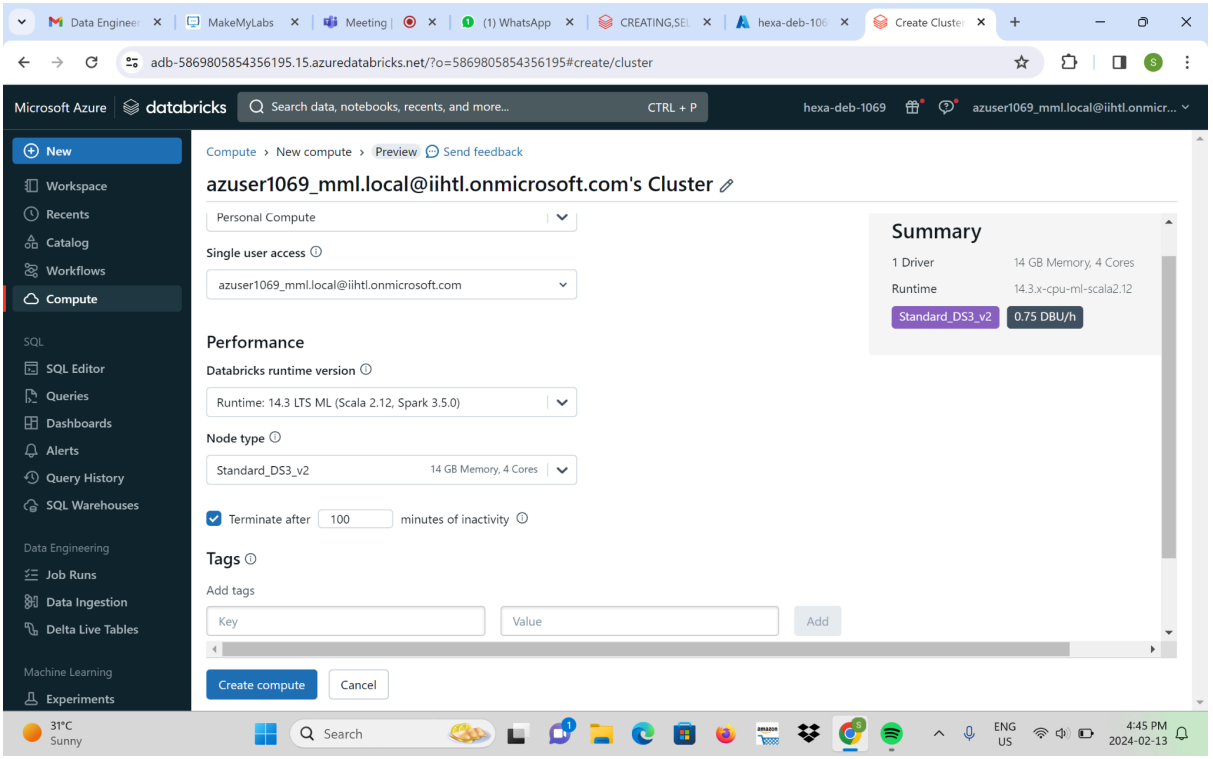
NAME : AKULA SHARATH CHANDRA
BATCH : DATA ENGINEERING
DATE : 13-02-2024
TOPIC:Cluster, Databricks Visualization

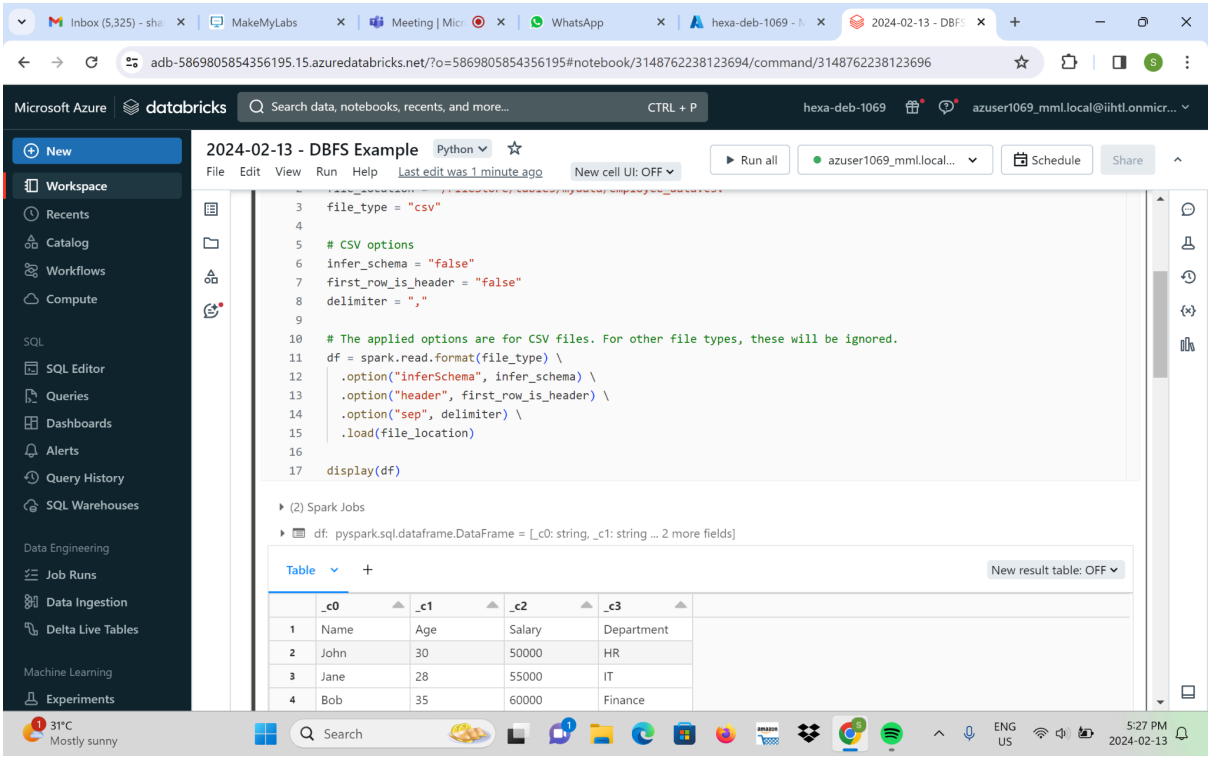
1)CLUSTER: In Azure Databricks, cluster is a series of Azure VMs that are configured with Spark, and are used together to unlock the parallel processing capabilities of Spark. In short, it is the compute that will execute all of our Databricks code.
There are two main types of clusters in Databricks :
1)**Interactive**: An interactive cluster is a cluster which we manually create through the cluster UI, and is typically shared by multiple users across multiple notebooks.
2)**Job**: A job cluster is a cluster that is tied to a Databricks Job. It spins up and then back down automatically when the job is being run.

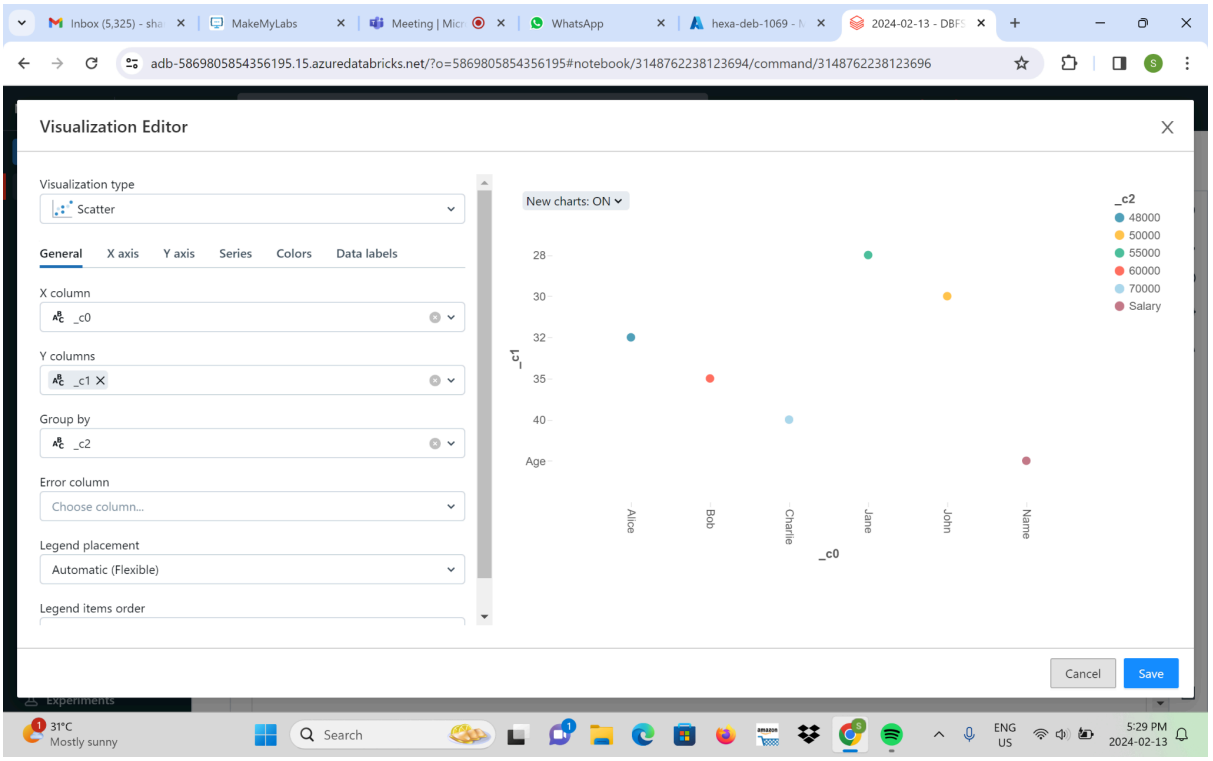


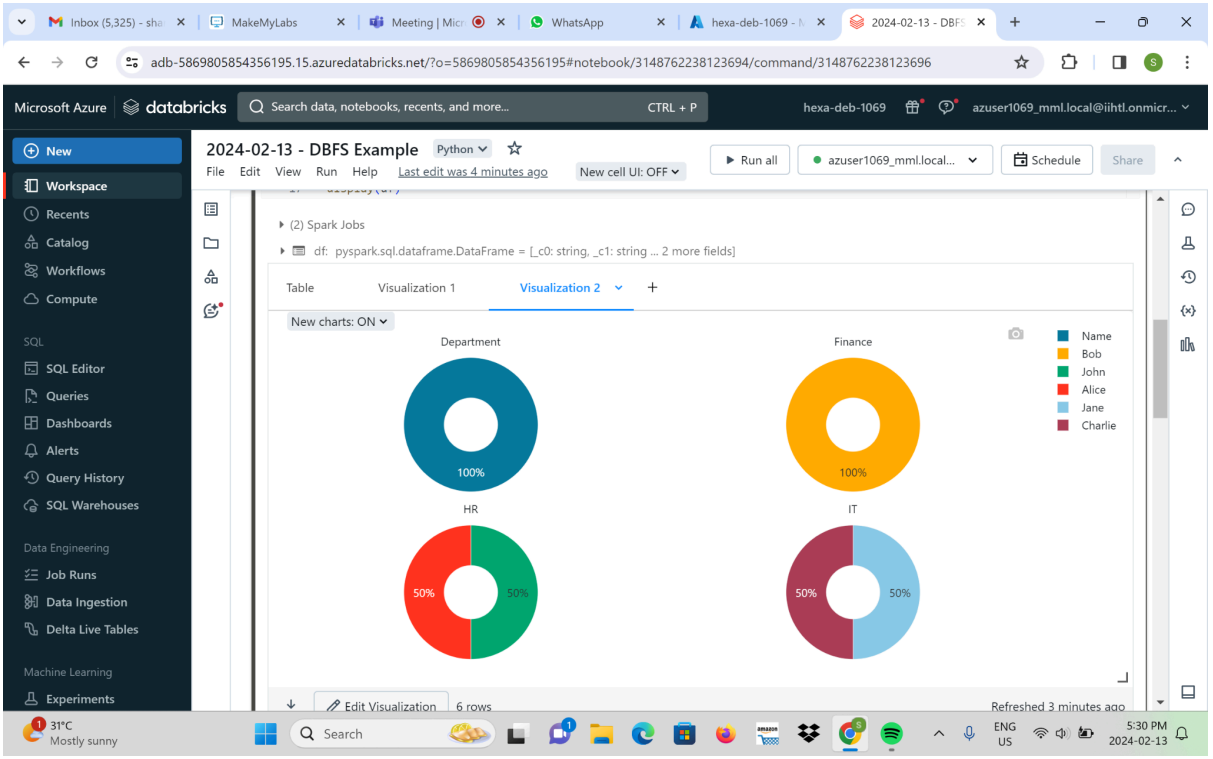
2)Create a new Visualization:

In order to create visualizations, we need to have data. Here I provided data and used the display() method.

→After creating a table
→Click on + symbol
→Click on visualization.
→Select the type of visualization, here I selected “Scatter” and “pie chart”.


Microsoft Azure
databricks
Search data, notebooks, recents, and more...
CTRL + P
hexa-deb-1069
2024-02-13 - DBFS Example
Python
File Edit View Run Help
Last edit was 1 minute ago
New cell UI: OFF
Run all
Schedule
Share
New
Workspace
Recents
Catalog
Workflows
Compute
SQL
SQL Editor
Queries
Dashboards
Alerts
Query History
SQL Warehouses
Data Engineering
Job Runs
Data Ingestion
Delta Live Tables
Machine Learning
Experiments
file_type = "csv"
# CSV options
infer_schema = "false"
first_row_is_header = "false"
delimiter = ","
# The applied options are for CSV files. For other file types, these will be ignored.
df = spark.read.format(file_type) \
.option("inferSchema", infer_schema) \
.option("header", first_row_is_header) \
.option("sep", delimiter) \
.load(file_location)
display(df)
(2) Spark Jobs
df: pyspark.sql.dataframe.DataFrame = [_c0: string, _c1: string ... 2 more fields]
Table
New result table: OFF
_c0 _c1 _c2 _c3
1 Name Age Salary Department
2 John 30 50000 HR
3 Jane 28 55000 IT
4 Bob 35 60000 Finance
31°C Mostly sunny
Search
ENG US
5:27 PM 2024-02-13


Visualization Editor
Visualization type
Scatter
General X axis Y axis Series Colors Data labels
X column
_c0
Y columns
_c1
Group by
_c2
Error column
Choose column...
Legend placement
Automatic (Flexible)
Legend items order
New charts: ON
_c2
48000
50000
55000
60000
70000
Salary
28
30
32
35
40
Age
_c1
Alice
Bob
Charlie
Jane
John
Name
_c0
Cancel
Save
31°C Mostly sunny
Search
ENG US
5:29 PM 2024-02-13


Microsoft Azure
databricks
Search data, notebooks, recents, and more...
CTRL + P
hexa-deb-1069
2024-02-13 - DBFS Example
Python
File Edit View Run Help
Last edit was 4 minutes ago
New cell UI: OFF
Run all
Schedule
Share
New
Workspace
Recents
Catalog
Workflows
Compute
SQL
SQL Editor
Queries
Dashboards
Alerts
Query History
SQL Warehouses
Data Engineering
Job Runs
Data Ingestion
Delta Live Tables
Machine Learning
Experiments
(2) Spark Jobs
df: pyspark.sql.dataframe.DataFrame = [_c0: string, _c1: string ... 2 more fields]
Table
Visualization 1
Visualization 2
New charts: ON
Department
Finance
HR
IT
100%
50%
Name
Bob
John
Alice
Jane
Charlie
Edit Visualization
6 rows
Refreshed 3 minutes ago
31°C
Mostly sunny
Search
ENG
US
5:30 PM
2024-02-13