**Assessment-26**

**Name: Vinutha S**

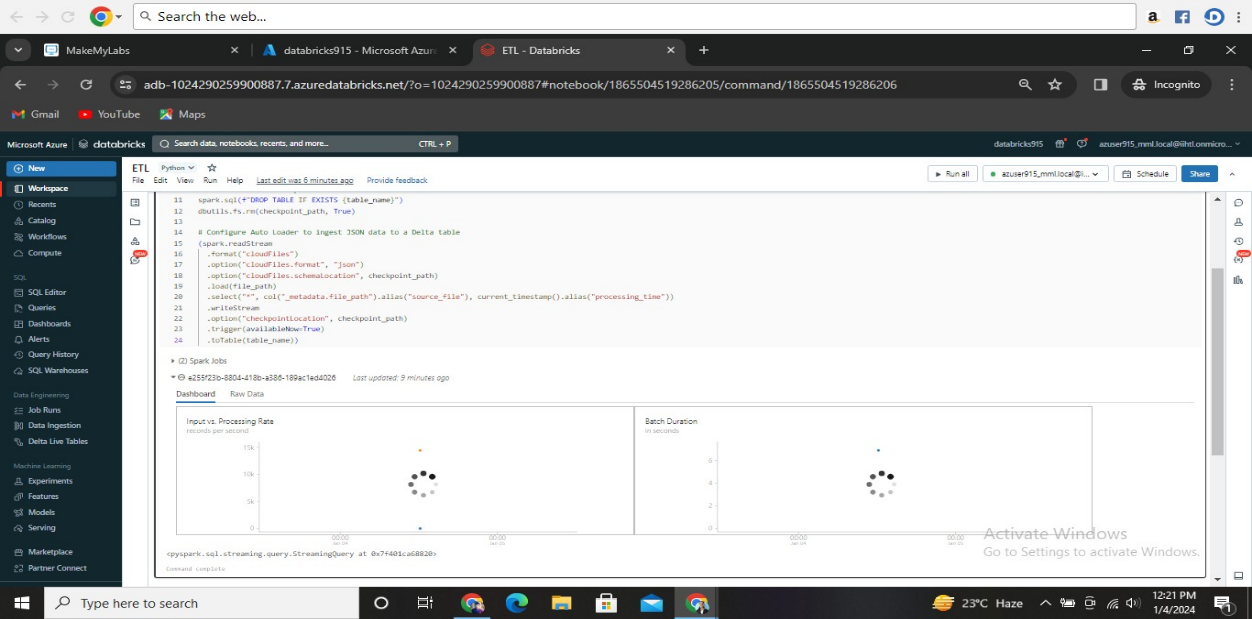
**Date: 04/01/2024 (Thursday)**

# ETL WORKLOAD ON AZURE DATABRICKS

1. Created a Azure Databricks Workspace. Followed by a cluster is created in my workspace.

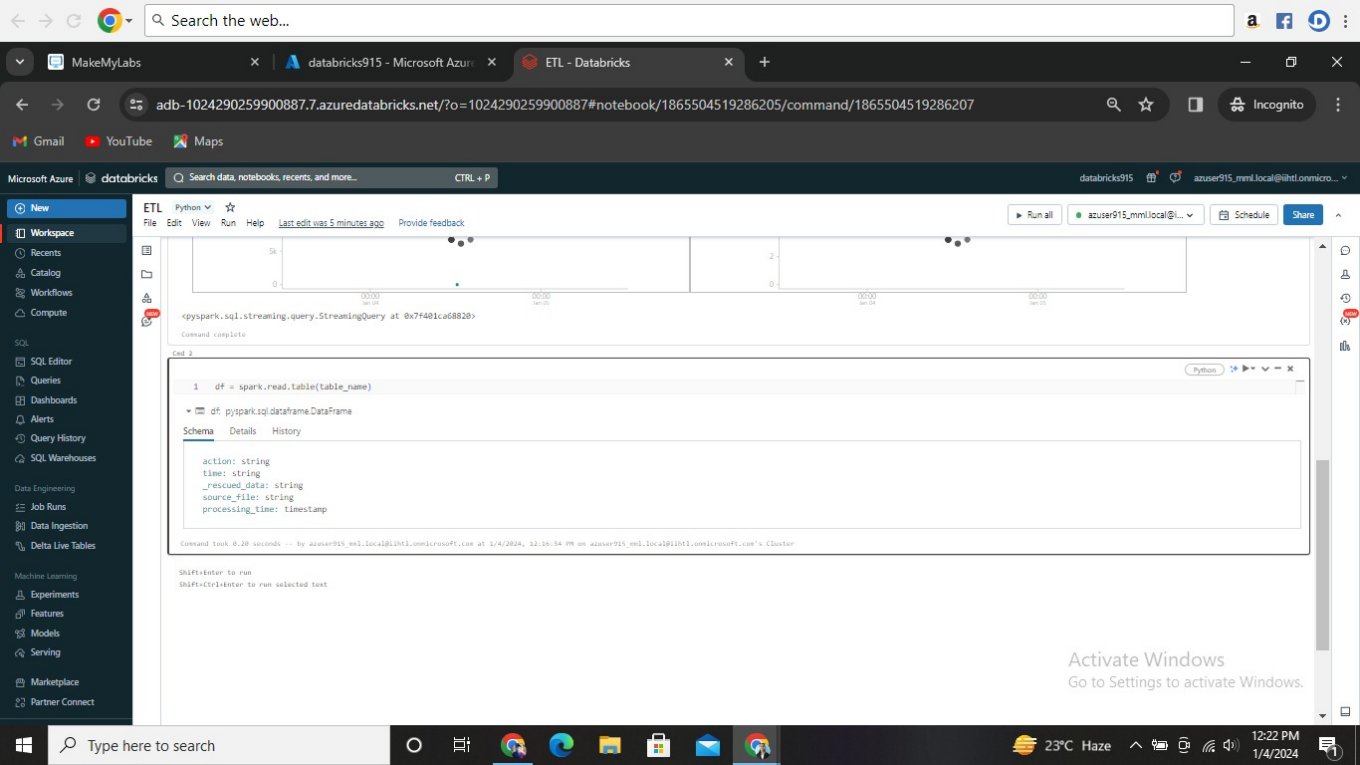
Later a new notebook is created with the name “ETL”.

1. Configured auto loader to ingest data into the delta lake. After entering the python code in the notebook the following images were created as shown below.

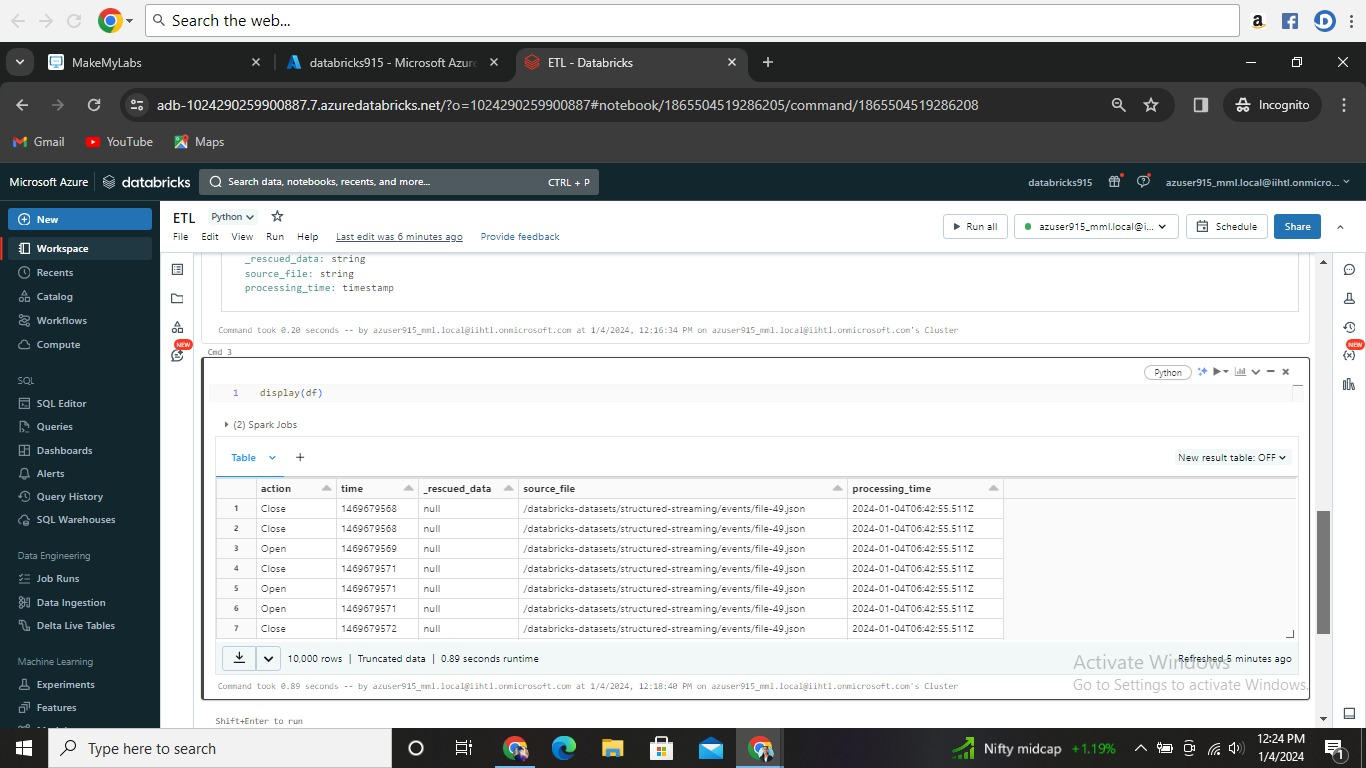


1. Process and interact with the data.

The checked whether the table is created or not by running a python code as shown below.

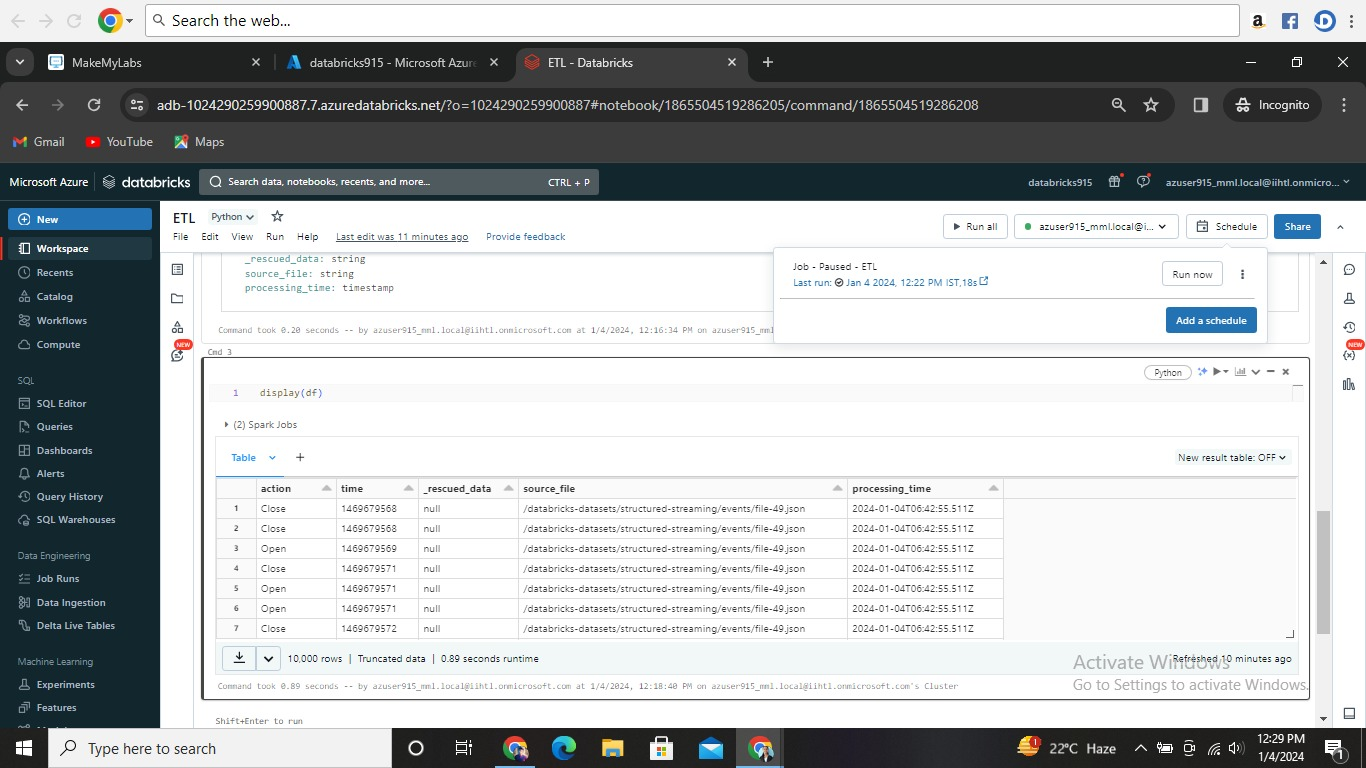


1. To display the data withing the table I ran the python code as shown below.



1. Schedule a job.

I scheduled a time to run the task by clicking schedule at right top and created as shown below.



1. Once the scheduled job is executed , the page will be redirected to a result page which contains Task run details, outputs as shown below.

# 

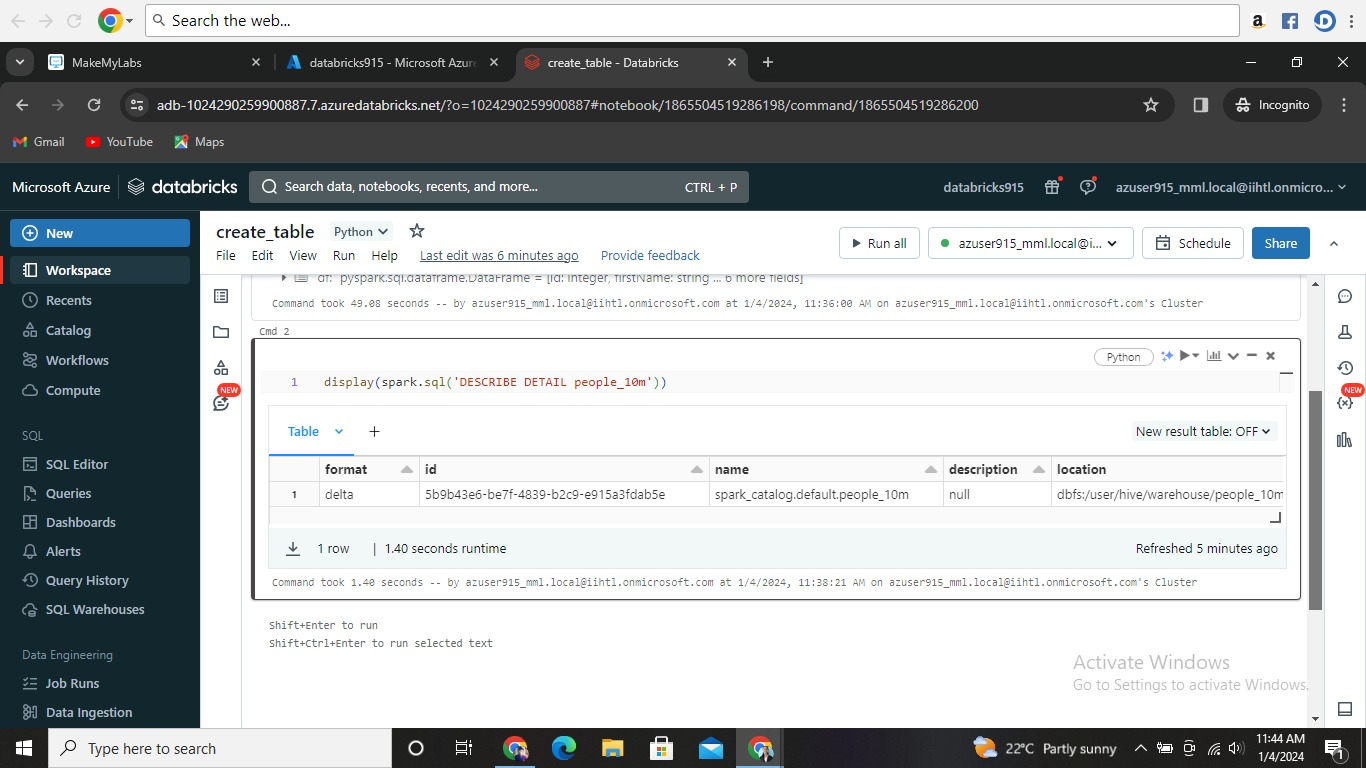
# **Data pipelines using Delta Lake and Delta Live Tables**

* Delta lake is a storage layer in Azure Databricks which is used to store data and the tables.
* Delta Lake is a default storage layer. Until unless we give a particular path or folder the data and tables will be stored in Delta Lake default.
* Delta Lake is open-source software.
* Few operations of Delta Lake are:

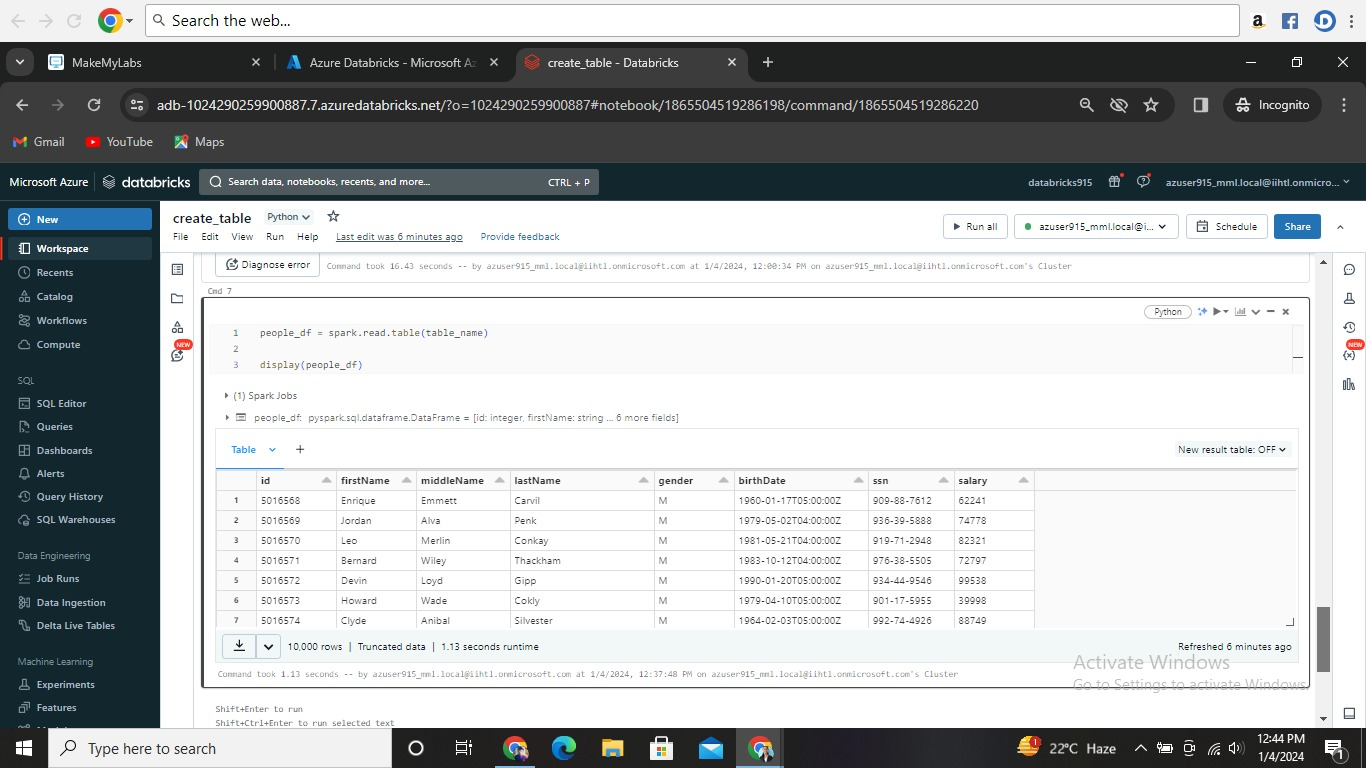
1. **Create a table.**
2. **Upsert to a table.**
3. **Read from a table.**
4. **Display table history.**
5. **Query an earlier version of a table.**
6. **Optimize a table.**
7. **Add a Z-order index.**
8. **Vacuum unreferenced files.**
9. **Create a table.**

****

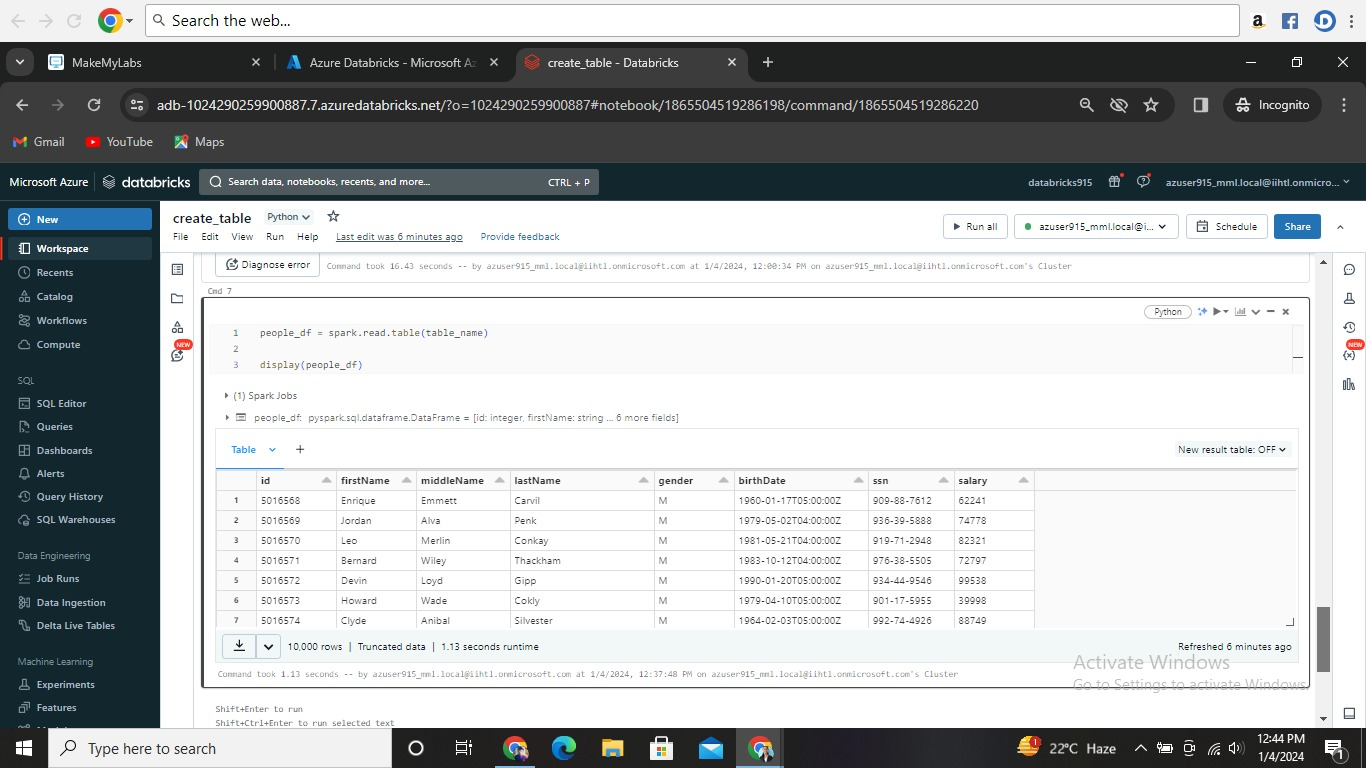
1. **Display the table. (Location of the table)**

****

1. **Display the data of the table.**

****

1. **Update the table.**

****