|  |  |
| --- | --- |
| **NAME** | **KAILASH S** |
| **DATE** | **04-01-24** |
| **ASS NO** | **6** |
| **TOPIC** | **ETL PIPELINE AND JOB SCHEDULING** |

**DATABRICKS ASSIGNMENT 6**

ETL PIPELINE AND WORLOAD SCHEDULING USING AUTOLOADER:

AUTOLOADER:

Auto Loader provides a Structured Streaming source called cloud Files. Given an input directory path on the cloud file storage, the cloud Files source automatically processes new files as they arrive, with the option of also processing existing files in that directory.

Azure Databricks Workflows orchestrates data processing, machine learning, and analytics pipelines on the Databricks Data Intelligence Platform.

STEP 1:

TO CREATE A CLUSTER:

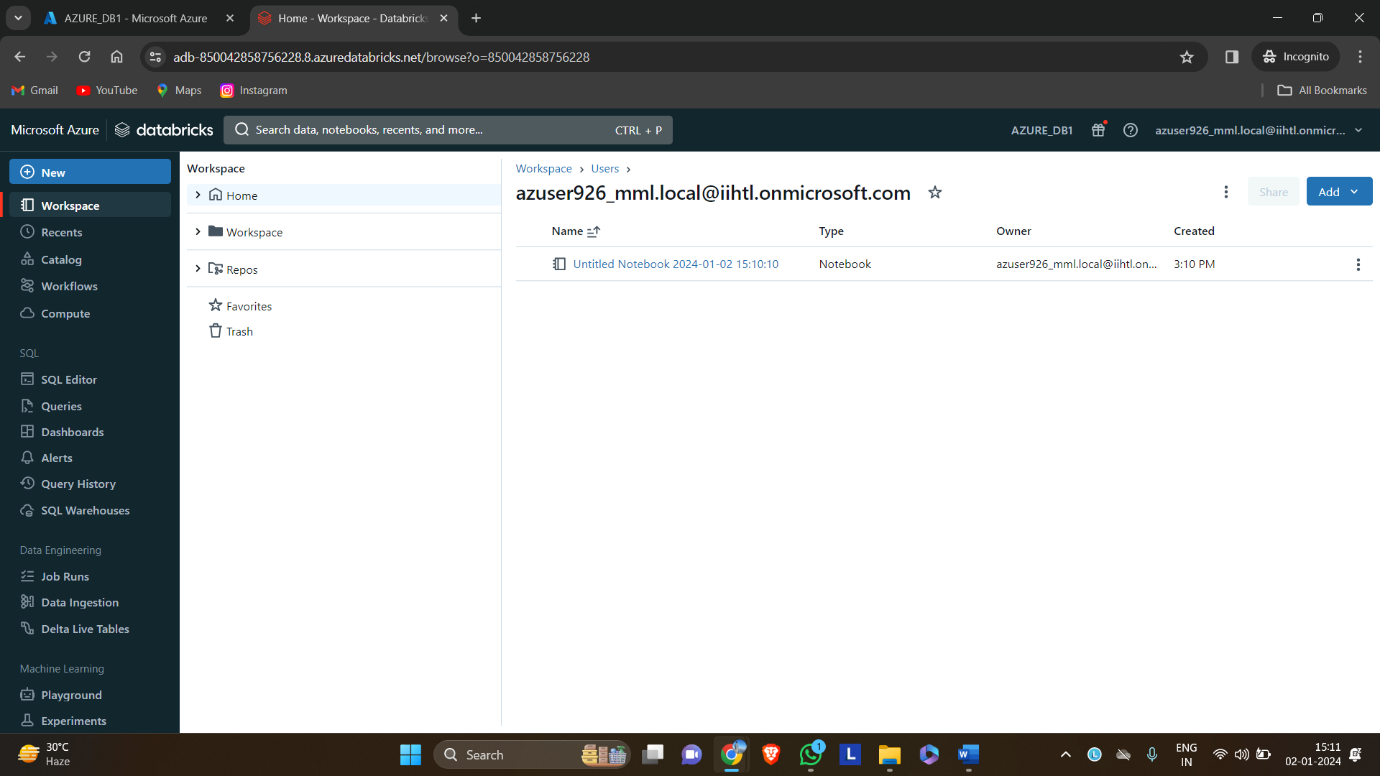
Create a cluster by clicking the compute tab in the menu and give a suitable name and initiate it.

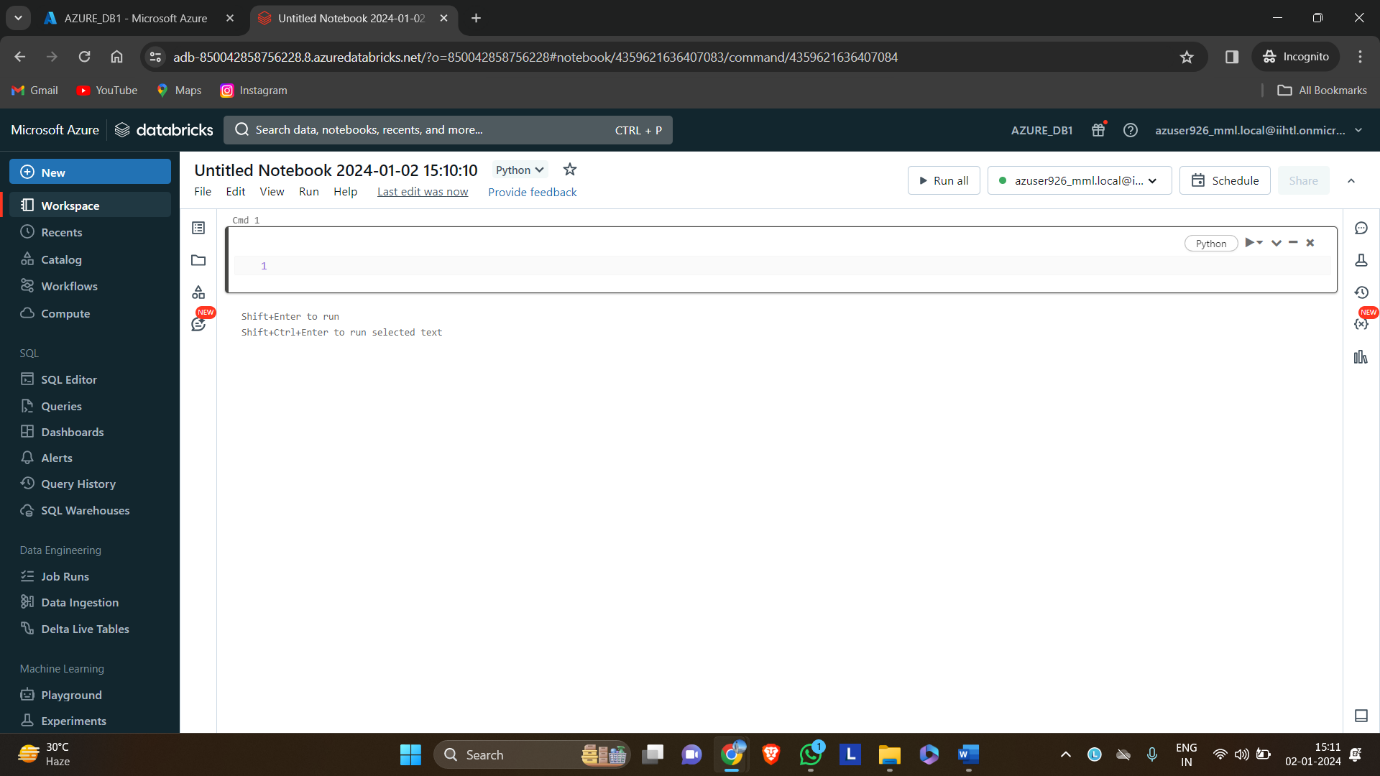
A screenshot of a computer

Description automatically generated

Step 2:

Create a notebook.

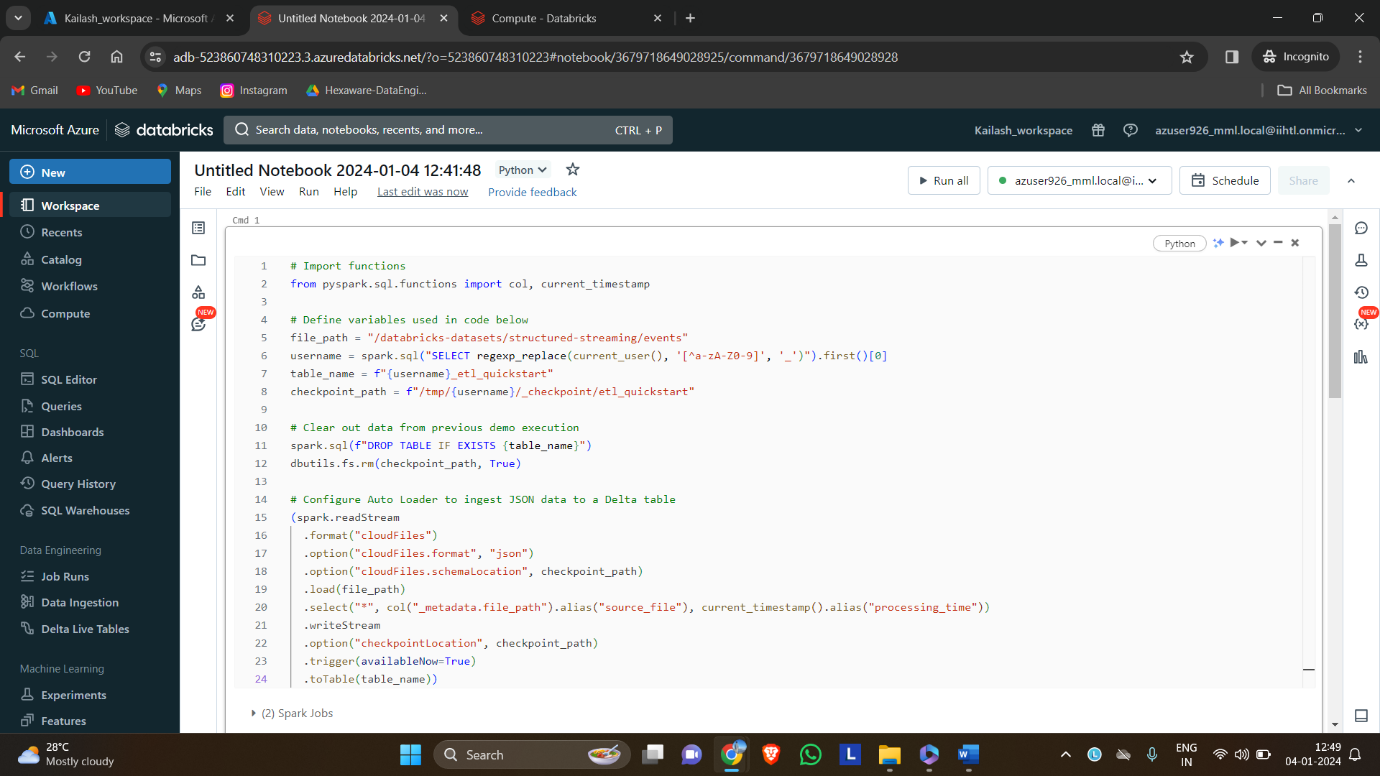




Step 3:

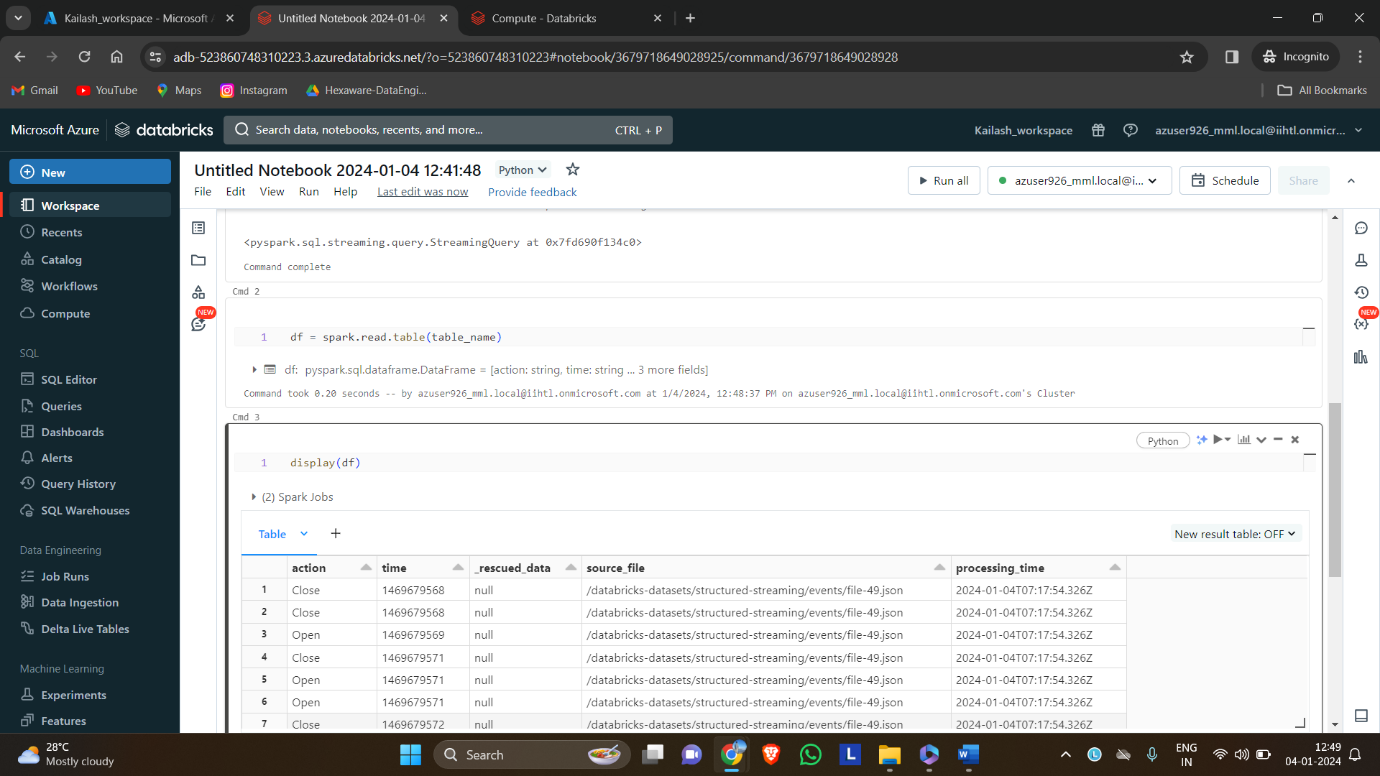
**Configure Auto Loader to ingest data to Delta Lake.**

Auto Loader automatically detects and processes new files as they arrive in cloud object storage.



**Step 4:**

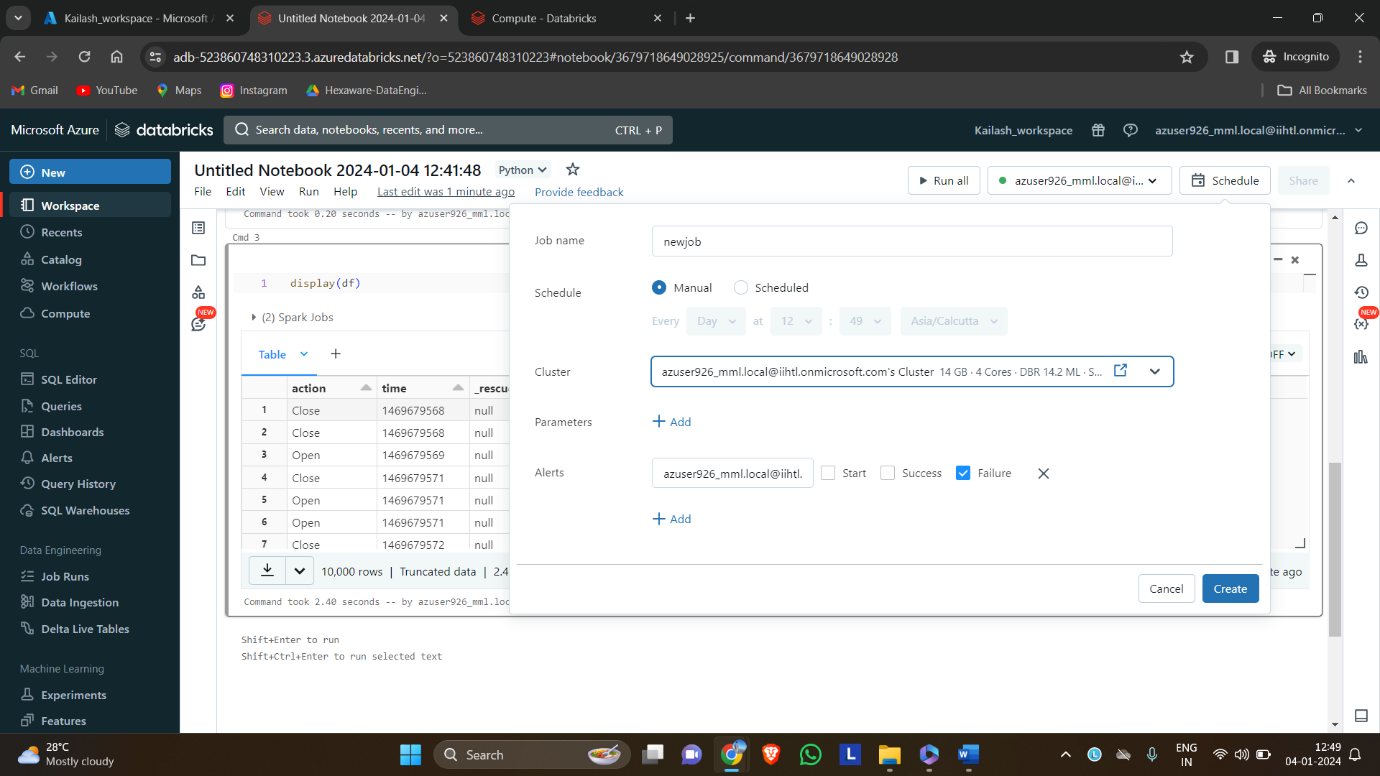
Process and interact with data



Step 6:

To create a job, select schedule in the right corner.

Give a suitable name, select manual in the schedule and select the created cluster. Then the job can be created.



After scheduling the job we run that code in any time we wanted.

