Assignment-18

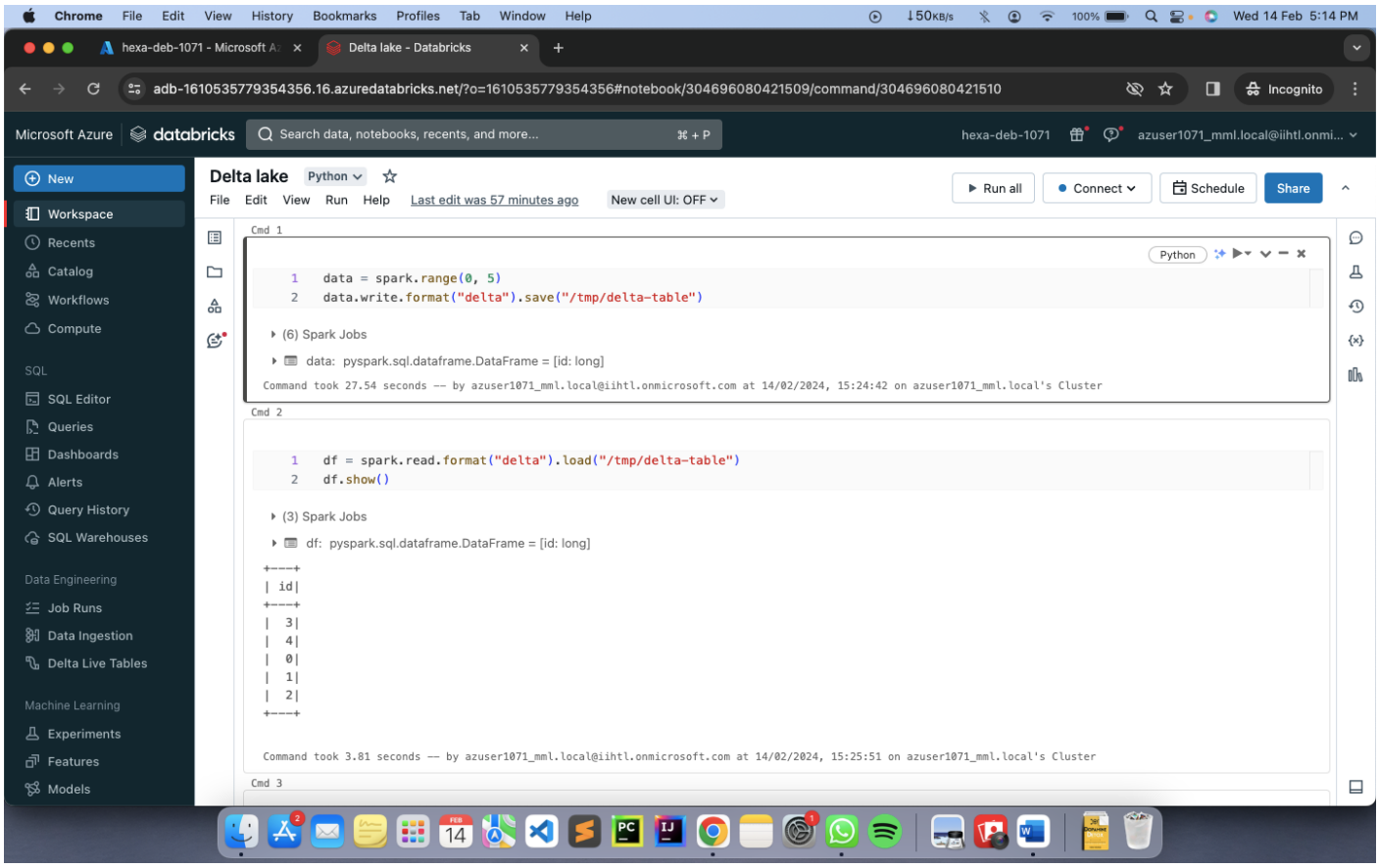
Name-Ajay Chaudhary
Batch-Data Engineering (Batch-1)

Create a table

To create a Delta table, write a DataFrame out in the `delta` format. You can use existing Spark SQL code and change the format from `parquet`, `csv`, `json`, and so on, to `delta`.

Read data

You read data in your Delta table by specifying the path to the files: `"/tmp/delta-table"`:



Update table data

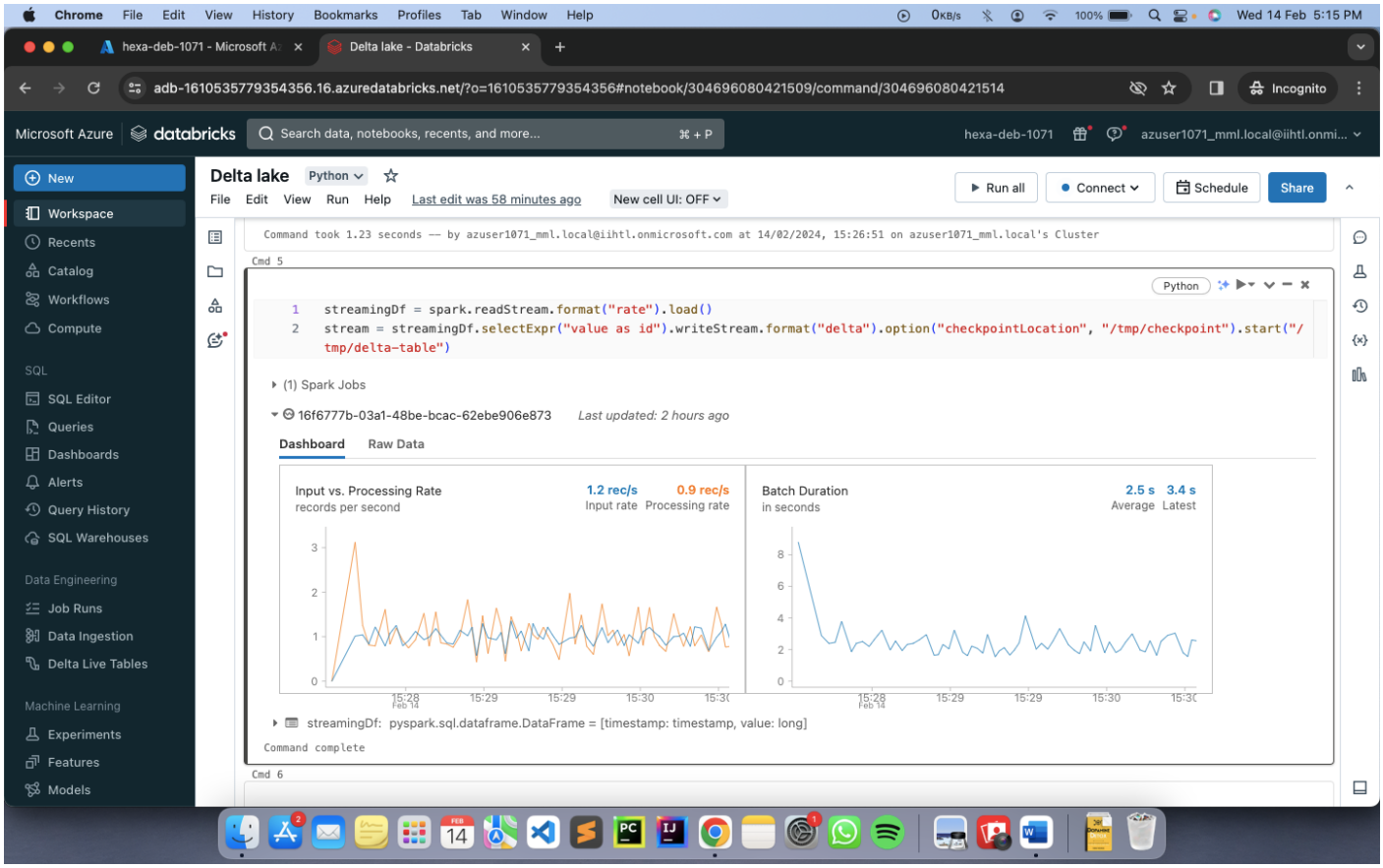
Delta Lake supports several operations to modify tables using standard DataFrame APIs. This example runs a batch job to overwrite the data in the table:

Read older versions of data using time travel

You can query previous snapshots of your Delta table by using time travel. If you want to access the data that you overwrote, you can query a snapshot of the table before you overwrote the first set of data using the `versionAsOf` option.

Write a stream of data to a table

You can also write to a Delta table using Structured Streaming. The Delta Lake transaction log guarantees exactly-once processing, even when there are other streams or batch queries running concurrently against the table. By default, streams run in append mode, which adds new records to the table:



Read a stream of changes from a table

While the stream is writing to the Delta table, you can also read from that table as streaming source. For example, you can start another streaming query that prints all the changes made to the Delta table.