Lab 20 – Databricks Delta Lake

1. Provision Azure Databricks Workspace

A screenshot of a computer

AI-generated content may be incorrect.

1. Import Notebook into Delta Lake Notebook

A screenshot of a computer

AI-generated content may be incorrect.

1. Ingest csv file and view top 10 rows

A screenshot of a computer

AI-generated content may be incorrect.

1. Load the file data into a delta table

A screenshot of a computer

AI-generated content may be incorrect.

1. Update records in delta table

A screenshot of a computer

AI-generated content may be incorrect.

1. View log for updates

A screenshot of a computer

AI-generated content may be incorrect.

1. Create External Table

A screenshot of a computer

AI-generated content may be incorrect.

1. Create Managed Table

A screenshot of a computer

AI-generated content may be incorrect.

1. Compare External and Managed table post dropping them

A screenshot of a computer

AI-generated content may be incorrect.

1. Use Delta Lake for streaming data

A screenshot of a computer

AI-generated content may be incorrect.

1. Delete Azure Resources

A screenshot of a computer

AI-generated content may be incorrect.

Summary:

Delta Lake is an open-source project to build a transactional data storage layer for Spark on top of a data lake. Delta lake supports both batch and streaming data operations. We begin the lab by provisioning resources. We import the notebook and attach compute to run the notebook. We ingest products csv into delta table and update a few of the records and check the updates along with the logs. We also create external and managed tables based on the delta tables and look at the differences in where they are created and what happens if we delete these tables. Lastly we test delta tables along with streaming data and close the lab.