Lab 7 –Spark Delta Lake

1. Provision an Azure Synapse Analytics workspace

A screenshot of a computer

AI-generated content may be incorrect.

1. Explore data in the data lake

A screenshot of a computer

AI-generated content may be incorrect.

1. Load the file data into a delta table

A computer screen with a message box

AI-generated content may be incorrect.

1. Create external catalog tables

A screenshot of a computer

AI-generated content may be incorrect.

1. Create managed catalog tables

A screenshot of a computer

AI-generated content may be incorrect.

1. Create a table using SQL

A screenshot of a computer

AI-generated content may be incorrect.

1. Use delta tables for streaming data

A screenshot of a computer

AI-generated content may be incorrect.

1. Query a delta table from the serverless SQL pool

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 designed to create a transactional data storage layer on top of a data lake, supporting both batch and streaming data to enable the Lakehouse architecture. After setting up the workspace, delta tables are created, which offer the advantage of version tracking for updates. Additionally, external tables stored in file storage and managed tables defined in the Hive metastore are explored. Managed tables are removed from the metastore when dropped. Delta tables are particularly useful for streaming data, as they allow for constant tracking of changes. The lab concludes with querying delta tables using a Serverless SQL pool.