Lab 2 – Serverless SQL

1. Provision an Azure Synapse Analytics workspace

A screenshot of a computer

AI-generated content may be incorrect.

1. View files in the data lake

A screenshot of a computer

AI-generated content may be incorrect.

1. Use SQL to query CSV files

A screenshot of a computer

AI-generated content may be incorrect.

1. Use SQL to query parquet files

A screenshot of a computer

AI-generated content may be incorrect.

1. Use SQL to query JSON files

A screenshot of a computer

AI-generated content may be incorrect.

1. Create an External data source

A screenshot of a computer

AI-generated content may be incorrect.

1. Create an External table

A screenshot of a computer

AI-generated content may be incorrect.

1. Visualize query results

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:

Azure Synapse Analytics serverless SQL pools enable querying data in formats like CSV, Parquet, and JSON, making them ideal for data lakes without requiring infrastructure setup or cluster maintenance; charges are based on data processed. To begin, a Synapse workspace and data lake storage account are created using a PowerShell script and ARM template, with source files cloned from GitHub. Using Synapse Studio, files can be explored (though Parquet files cannot be previewed), and queries on CSV files are performed with OPENROWSET, adding column headers as needed. Parquet files were queried using wildcard matching and filtering, while JSON queries require additional parameters like fieldterminator and rowterminator. For advanced use cases, PolyBase capabilities allow creating external databases, defining external data formats, and referencing external tables for CSV or Parquet files. Depending on the scenario, users can choose between OPENROWSET or external storage access with serverless SQL pools.