Lab 9 – Load Data Warehouse

1. Provision an Azure Synapse Analytics workspace

A screenshot of a computer

AI-generated content may be incorrect.

1. Start dedicated SQL Pool

A screenshot of a computer

AI-generated content may be incorrect.

1. Preview files in the data lake

A screenshot of a computer

AI-generated content may be incorrect.

1. Load data from data lake using the COPY statement

A screenshot of a computer

AI-generated content may be incorrect.

1. Load data using the COPY statement and log errors

A screenshot of a computer

AI-generated content may be incorrect.

1. View the errors logs

A computer screen with a black pen

AI-generated content may be incorrect.

1. Use a CREATE TABLE AS (CTAS) statement

A screenshot of a computer

AI-generated content may be incorrect.

1. Combine INSERT and UPDATE statements to load a slowly changing dimension table

A screenshot of a computer

AI-generated content may be incorrect.

1. Post load optimization

A computer screen with a computer screen

AI-generated content may be incorrect.

1. Delete Azure Resources

A screenshot of a computer

AI-generated content may be incorrect.

Summary:

ELT (Extract, Load, Transform) is a process where data is extracted from a source system, loaded into a dedicated SQL pool, and then transformed. To demonstrate data loading into a data warehouse, we start by provisioning a Synapse Analytics workspace. Staging tables are used to validate or transform data before moving it to existing dimension tables, either by appending or upserting. The Copy statement facilitates easy data loading and redirects invalid rows. It is used to load data from a data lake into staging tables. Subsequently, data is loaded into the DimProduct table from the StageProduct table. When loading data into dimension tables, it is essential to consider slowly changing dimensions. Type 1 dimensions involve updating the existing row, while Type 2 dimensions involve inserting a new row. This process is demonstrated using INSERT and UPDATE statements with stage and dimension tables. Post-load optimization can be performed by creating indexes and statistics on key columns to enhance performance.