Lab 10 – Synapse Pipeline

1. Provision an Azure Synapse Analytics workspace

A screenshot of a computer

AI-generated content may be incorrect.

1. View source and destination data stores

A screenshot of a computer

AI-generated content may be incorrect.

1. Configure data flow

A screenshot of a computer

AI-generated content may be incorrect.

1. Add Sources

A screenshot of a computer

AI-generated content may be incorrect.

1. Add lookup

A screenshot of a computer

AI-generated content may be incorrect.

1. Add an Alter row

A screenshot of a computer

AI-generated content may be incorrect.

1. Add a sink

A screenshot of a computer

AI-generated content may be incorrect.

1. Debug the data flow

A screenshot of a computer

AI-generated content may be incorrect.

1. Publish and run the pipeline

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:

Data flows in Synapse are visually designed data transformations that allow logic development without writing code, utilizing scaled-out Apache Spark clusters for data flow activities. To create a pipeline that pulls product data from a text file and loads it into a dedicated SQL pool, we configure two sources: the products text file and the Products Table. We set the data types of the columns and add a look-up transformation between these two sources to identify new and existing products. For rows matched based on the lookup, we update the existing rows, while for unmatched rows, we insert new ones. After publishing and running the pipeline, we can debug it to observe that some rows have been updated and others inserted. This process ensures efficient data management and integration within the SQL pool.