Lab 3 – Transform data with 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. Create an external data source and file format

A screenshot of a computer

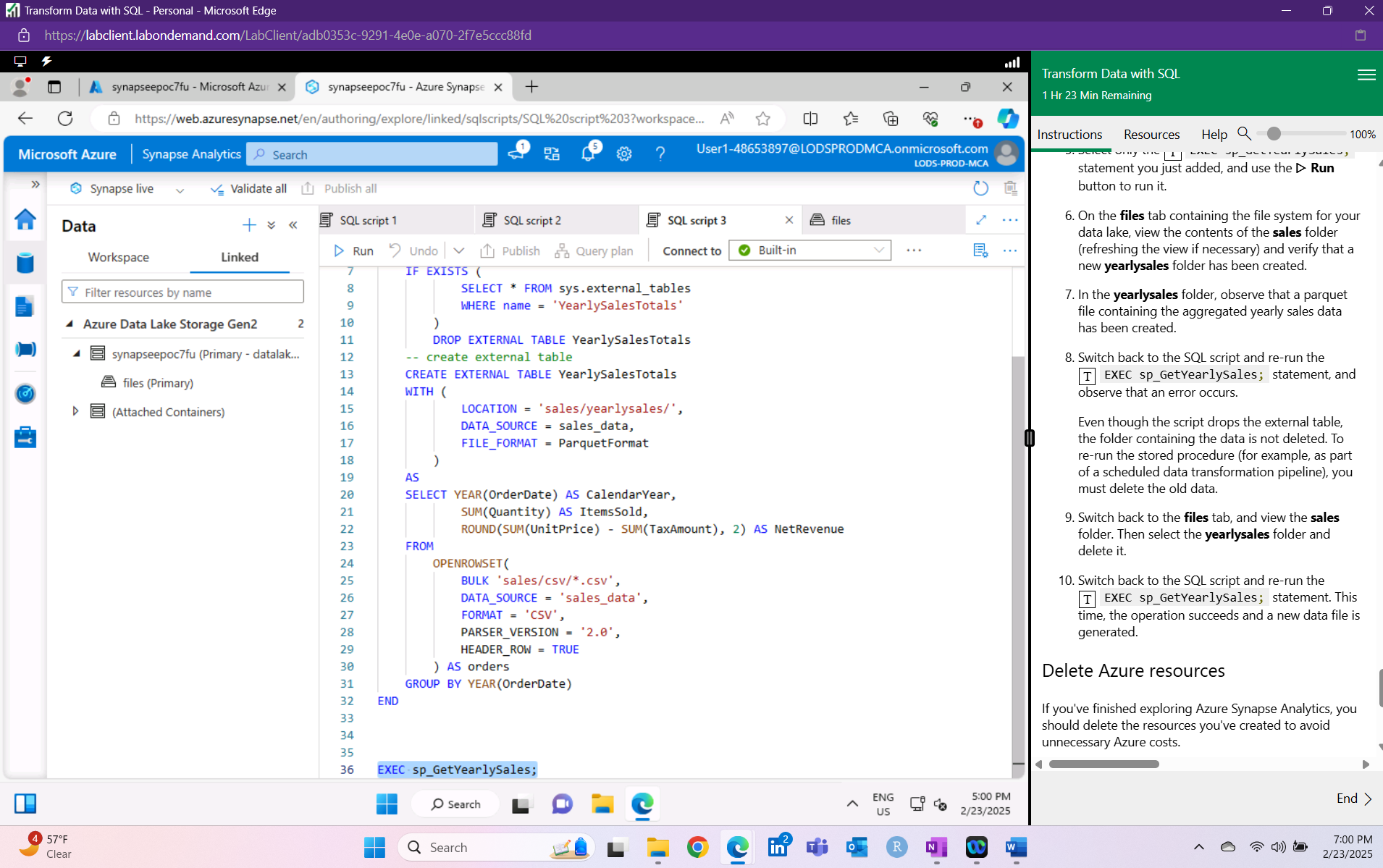
AI-generated content may be incorrect.

1. Create an external table

A screenshot of a computer

AI-generated content may be incorrect.

1. Encapsulate data transformation in a stored procedure



1. Delete Azure resources

A screenshot of a computer

AI-generated content may be incorrect.

Summary:

Data analysts and engineers use SQL for querying, transforming, and managing data. In this process, we use the serverless SQL pool in Azure Synapse to transform data. First, we provision an Azure Synapse Analytics workspace by cloning a Git repository and running a setup script. After exploring and querying CSV files in the sales folder of the data lake, we use CETAS (Create External Table as SELECT) to create external tables and export results. A stored procedure is defined to drop and recreate the external table while loading yearly sales data. Although the table is dropped, the data folder remains, so we deleted it manually before rerunning the procedure to generate updated data ready for querying. Once completed, we deleted the Azure resource group to stop the lab environment.