**Name: Sesha Venkata Sriram Erramilli**

**ECode: FT805**

**Date Start: Oct 1, 2023**

**Date End: Oct 2, 2023**

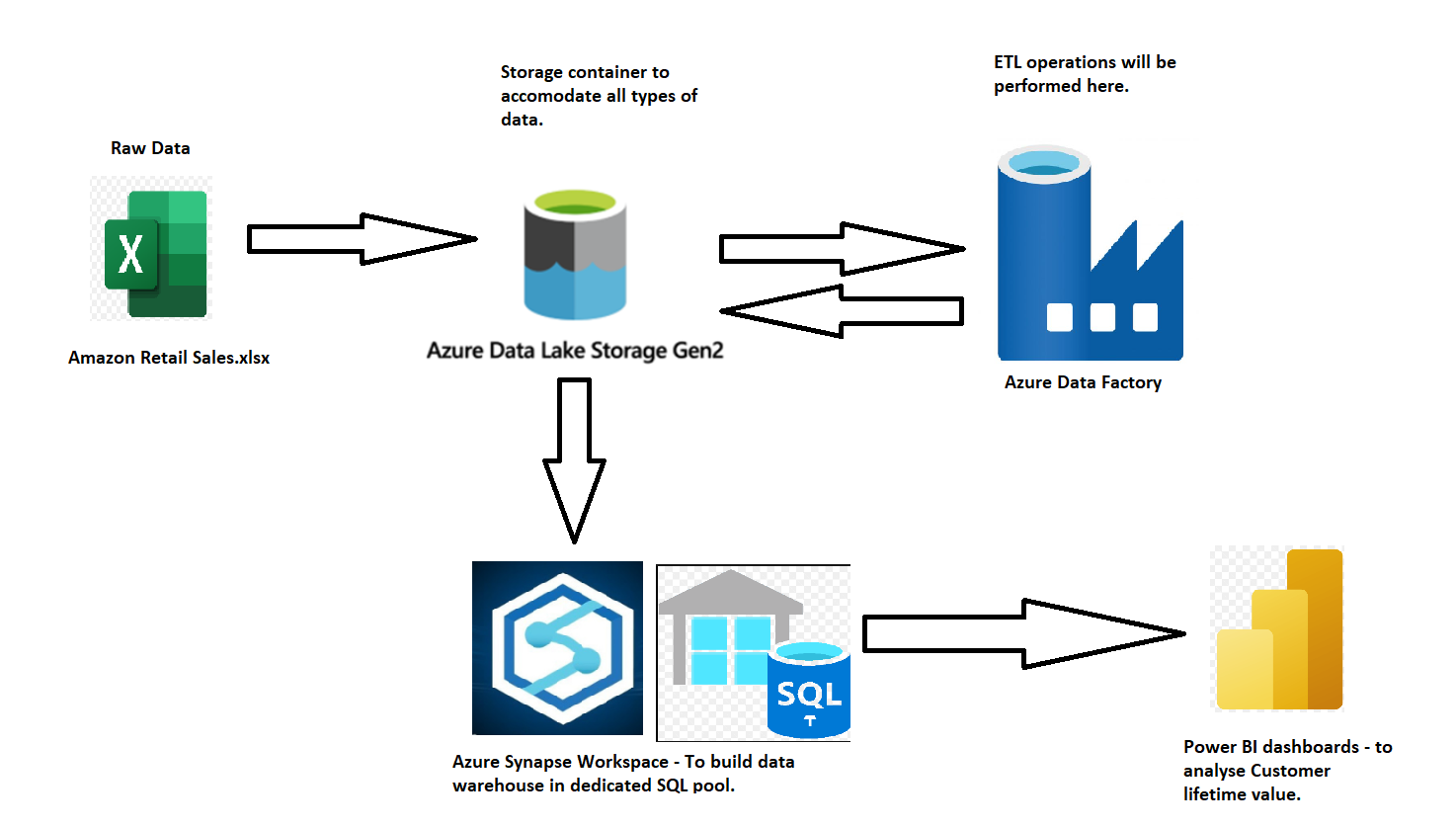
**Problem:**

Design an efficient Data engineering solution (Data Warehouse) for analyzing the retail sales data of Amazon.

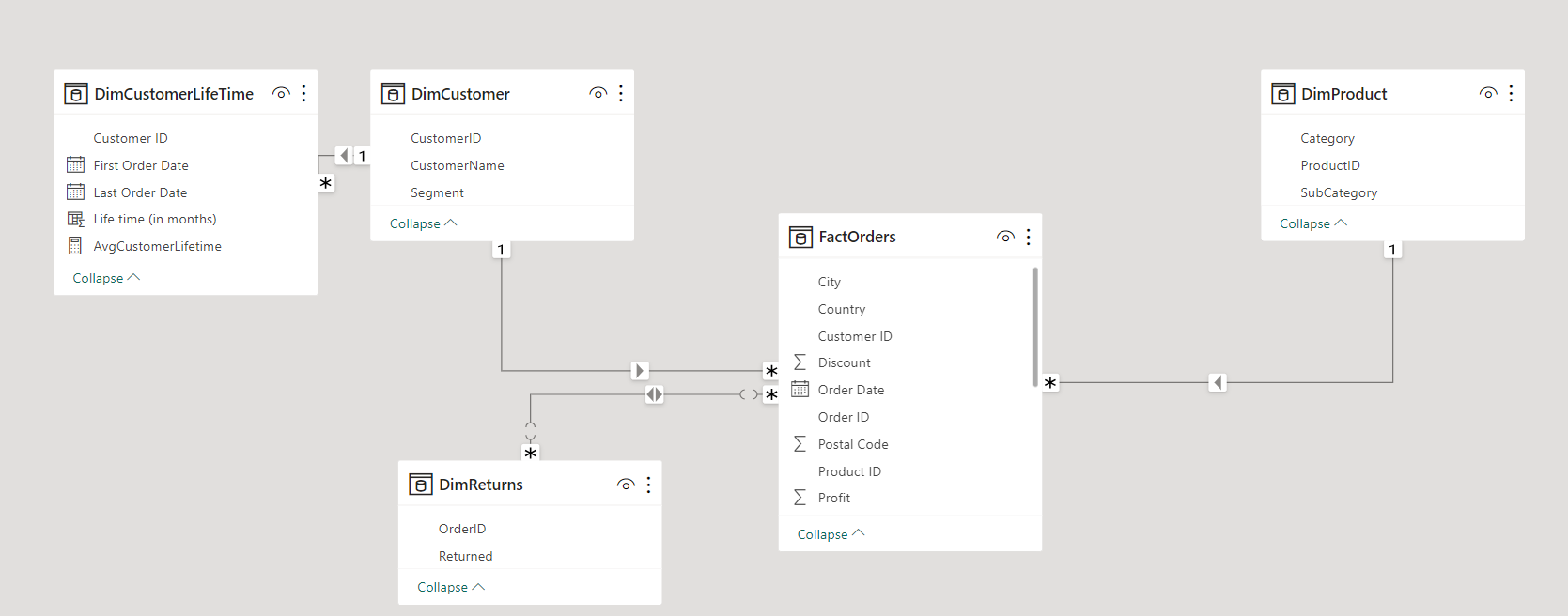
**Dataset link:**

<https://docs.google.com/spreadsheets/d/1Dg2jK3mrA8JYw417DS_CZ6jkeWMGfBlx/edit#gid=1579268976>

**Architecture:**

****

**Final Data model:**

****

**Solution Steps:**

1. Azure resource group: esv\_retail

A screenshot of a computer

Description automatically generated

Synapse authorization: SQL Authentication

1. Upload the raw “Amazon Retail Sales” excel file into the “raw” directory inside the Datalake storage.

A screenshot of a computer

Description automatically generated

1. Now inside Data factory, design data flows that creates staging files from the given excel file. I.e.,
   1. orders.csv
   2. returns.csv
   3. customer.csv
   4. product.csv

All these files have to be stored in a different container “transformed” inside the same Data lake storage.

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

1. Provisioned Dedicated SQL pool inside the Synapse workspace. This can be used to create a data warehouse.

A screenshot of a computer

Description automatically generated

1. Launch Synapse workspace. Inside the “Data” tab, we find the transformed container inside the Datalake storage which contains all the stage data.

A computer screen shot of a computer

Description automatically generated

1. And under the Workspace tab, we find the SQL database inside Dedicated SQL pool.

A screenshot of a computer

Description automatically generated

1. Created Staging tables as external tables for all the csv files inside “esv\_sql” database.

A screenshot of a computer

Description automatically generated

1. Created Fact and Dimension tables inside the “esv\_sql” database.

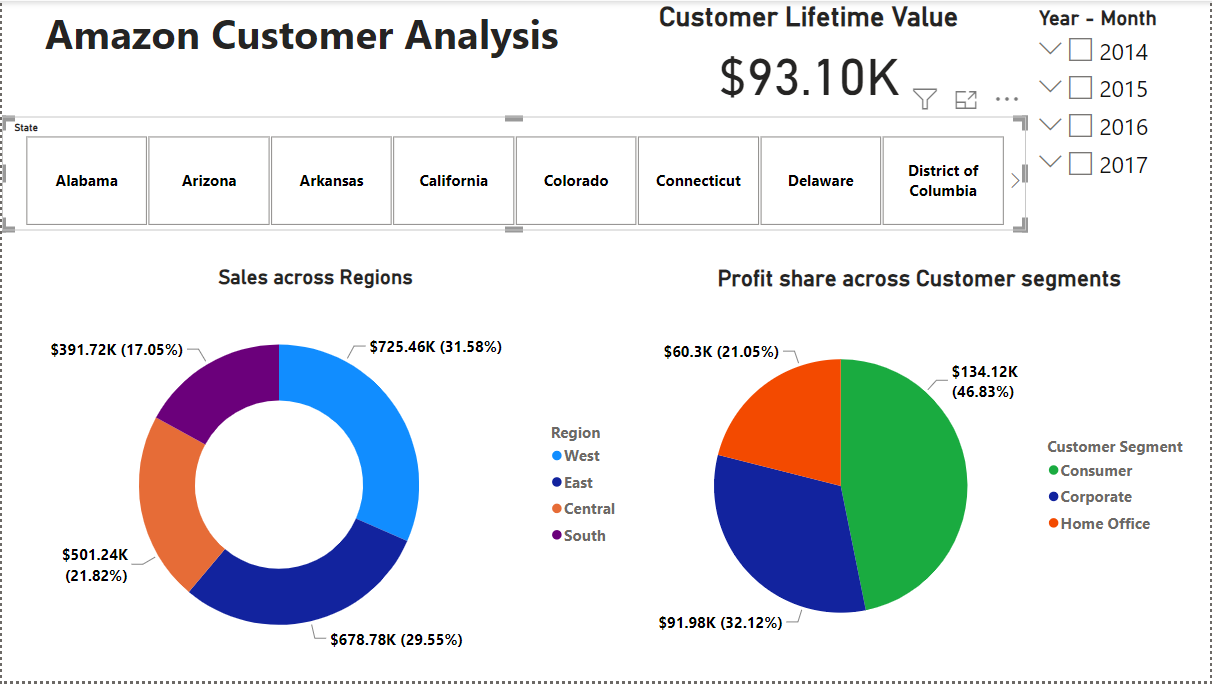
A screenshot of a computer

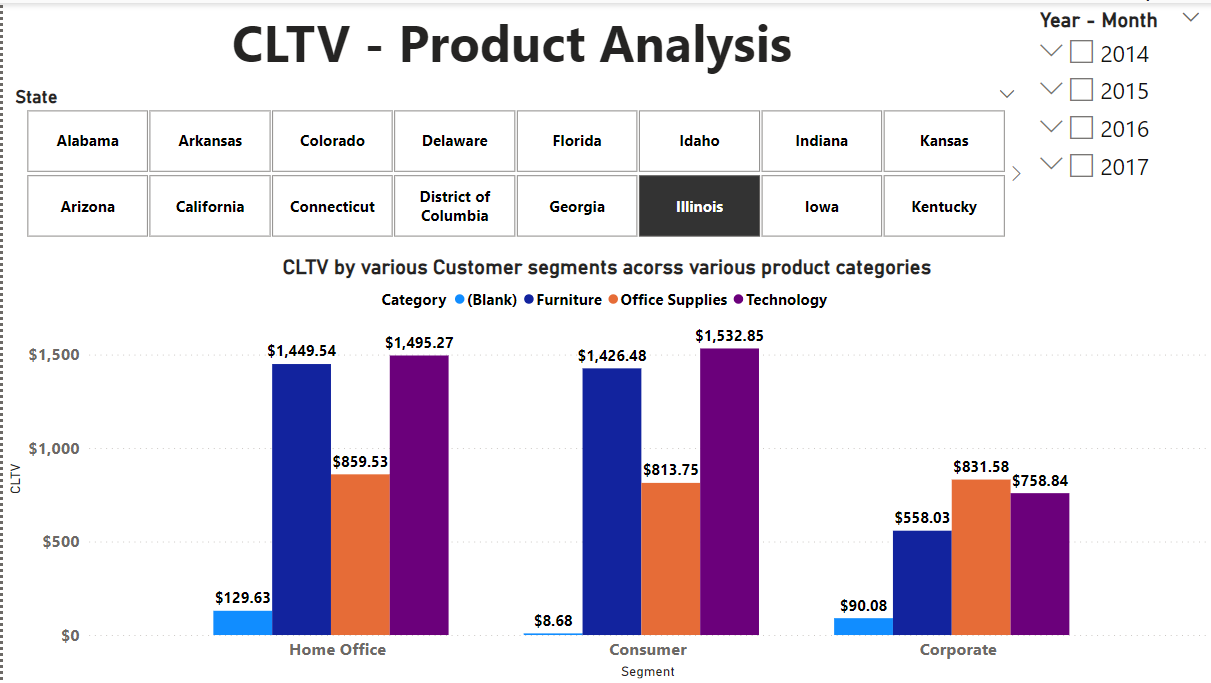
Description automatically generated

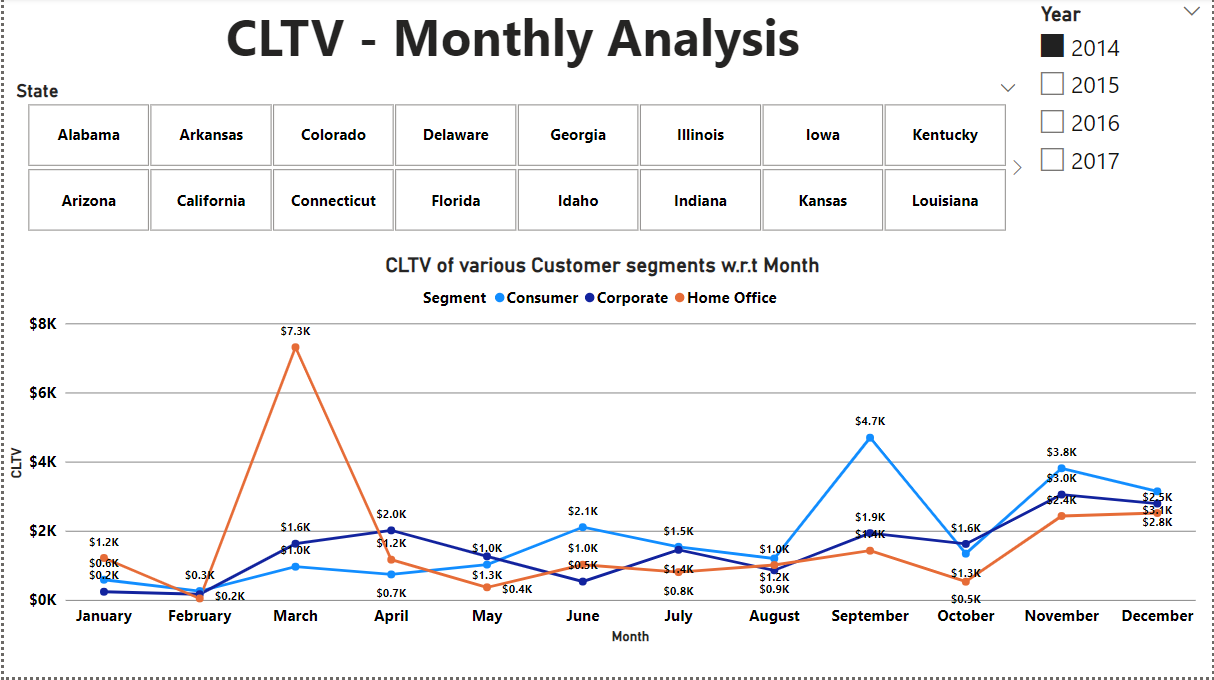
This is our final Data warehouse. Now connect this Data warehouse to PowerBI desktop.

**Data Analysis using Power BI:**

**Performed Customer Lifetime value analysis:**

****

****

****