Sri Lanka Institute of Information Technology

IT3021 – Data Warehousing and Business Intelligence

Assignment 01 Report

Student No: IT22060594

Name: Soyza W W K

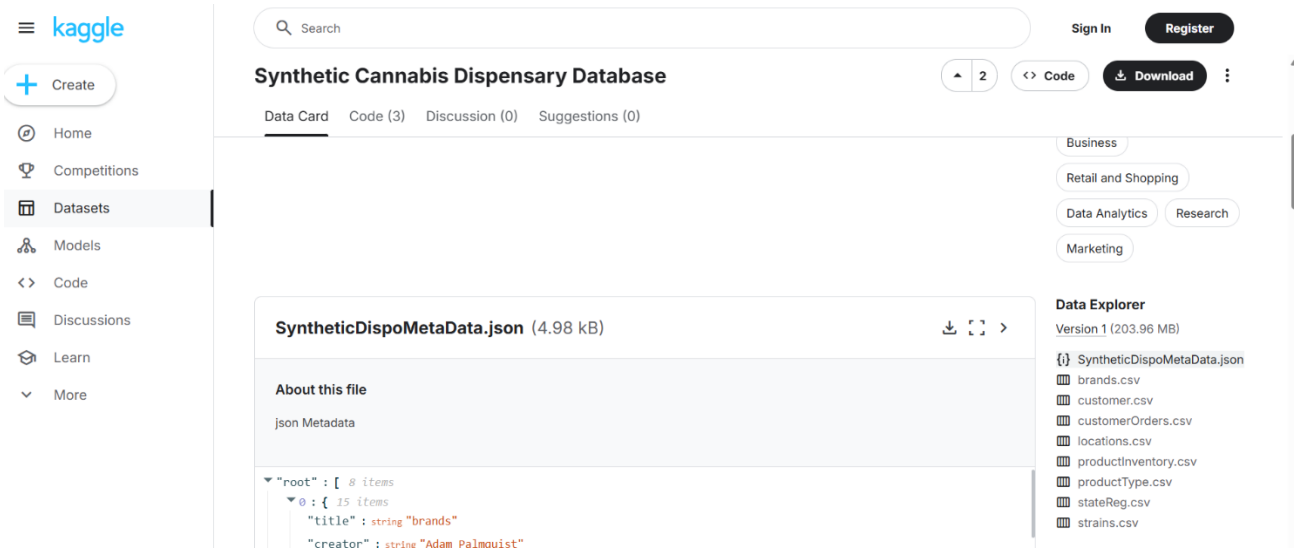
Batch: Y3.S1 DS WE 01

Introduction:

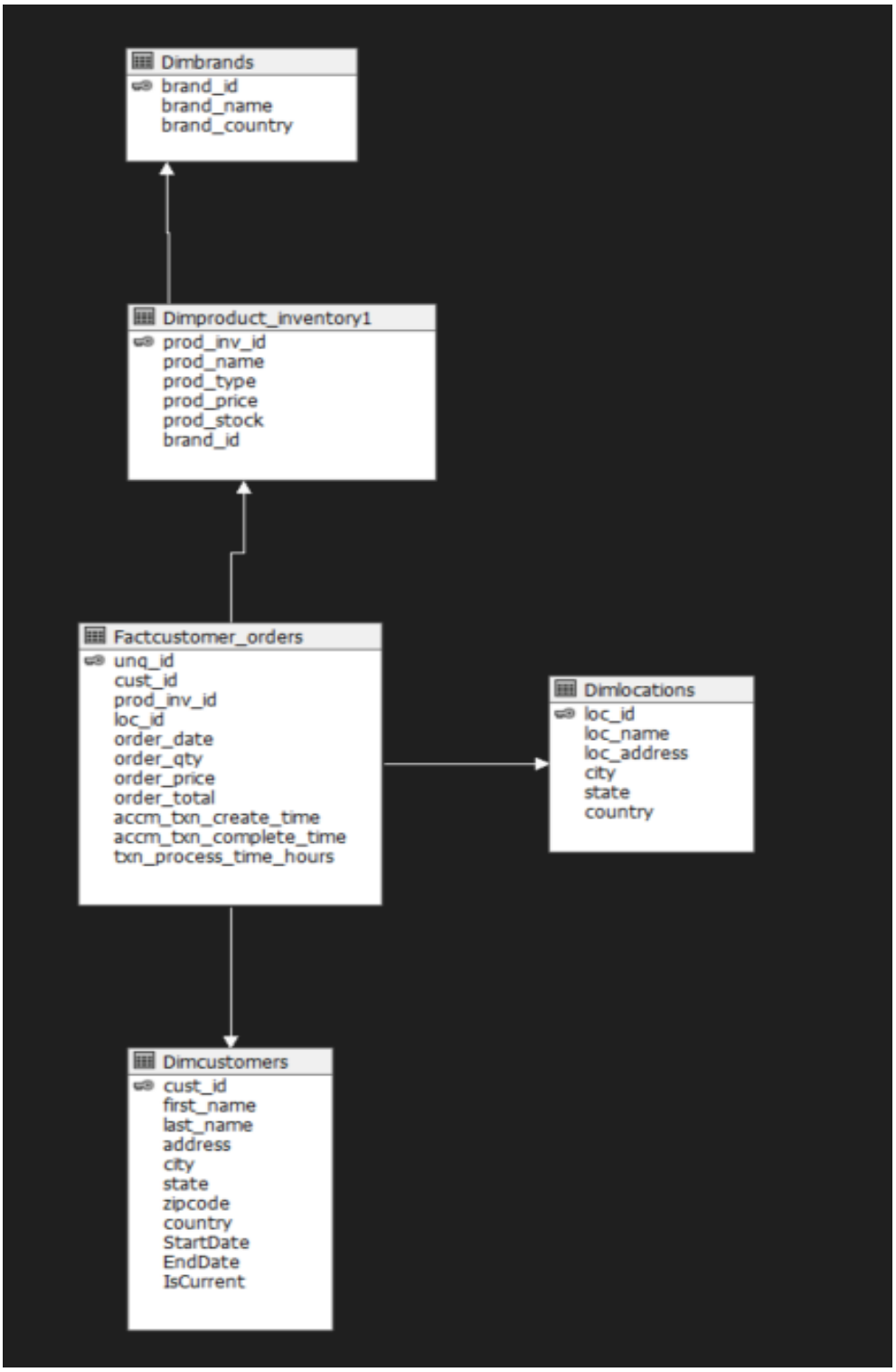
This dataset is a modified OLTP dataset originally sourced from Kaggle that simulates the operations of a **Cannabis Dispensary**. It has been enhanced using AI to better reflect a real-world business scenario and structured to support analytical processing. The dataset includes a **fact table containing transactional order data** (such as order totals) and multiple **dimension tables** that describe products, product types, geographical information (states and cities), and brands. Though initially designed for online transaction processing (OLTP), the dataset has been adapted to support **OLAP operations and Power BI reporting**, making it suitable for exploring business insights such as sales performance by product and region, customer purchasing trends, and more.

Dataset Selection

The selected dataset is the "Synthetic Cannabis Dispensary Dataset. It includes customer data, product inventory and type, customer orders, brand information. The dataset is well-suited for OLTP scenarios and suitable for DWBI use cases based on its relational schema and sufficient data variety.

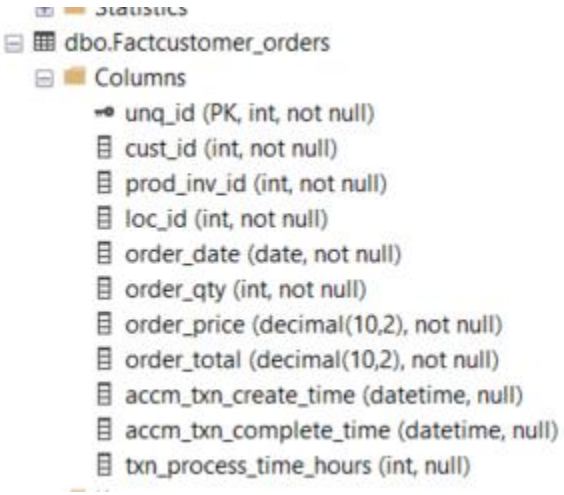


Relationships among Tables



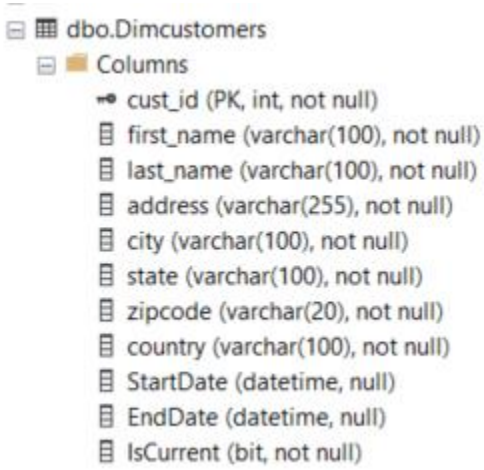
Preparation of data sources

Fact Tables- Factcustomer_orders Table

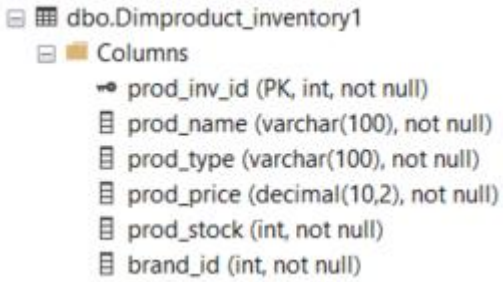


Dimension Tables-

- Customer Table - **Used in txt format**

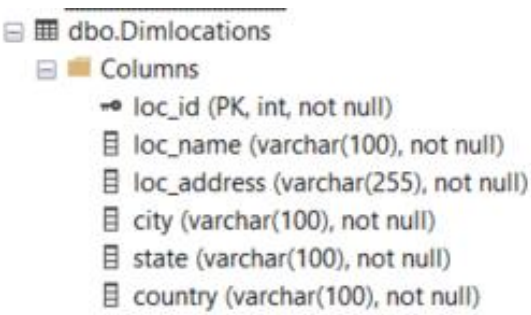


- Product Inventory Table- **Used in CSV format**

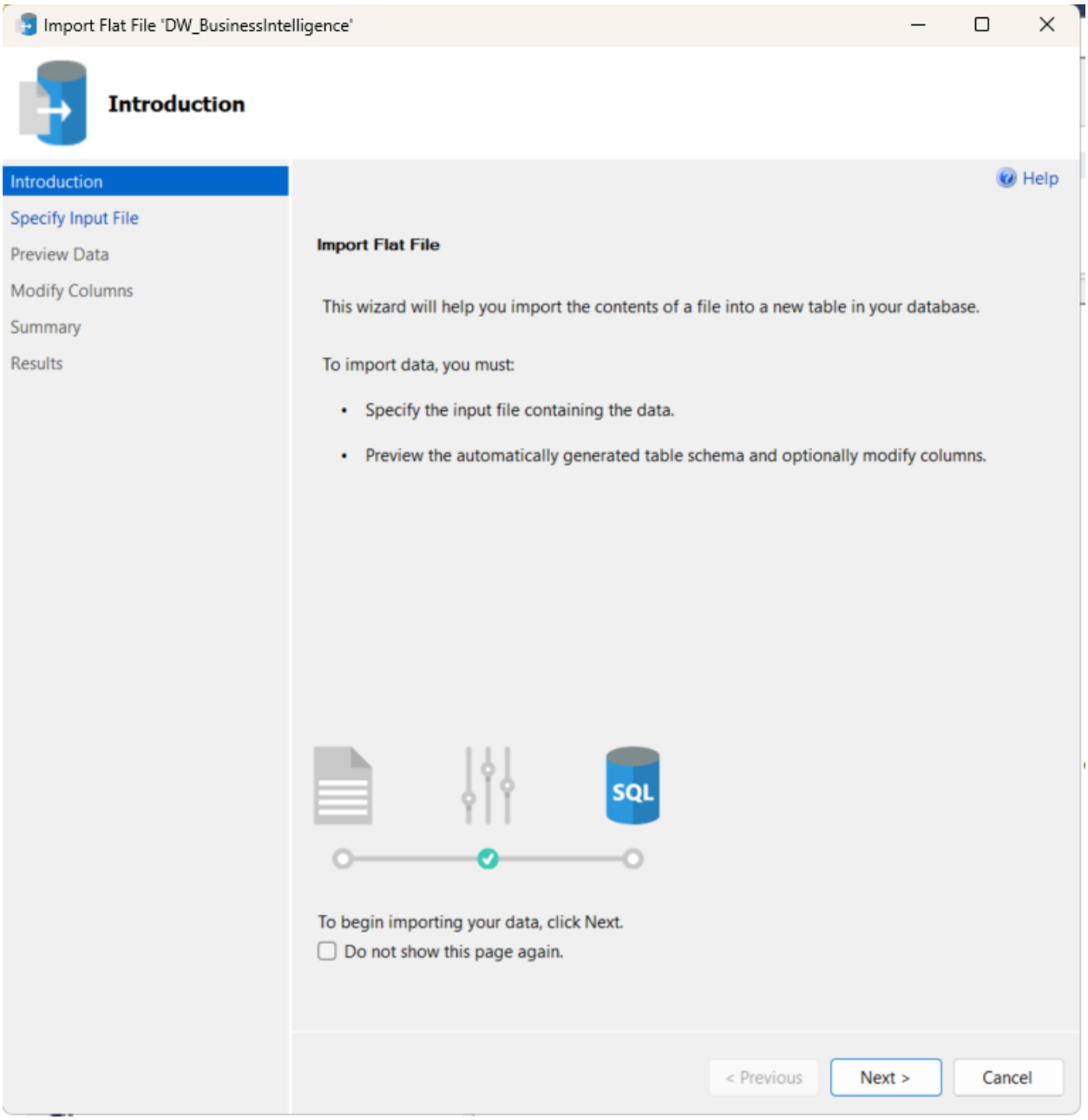


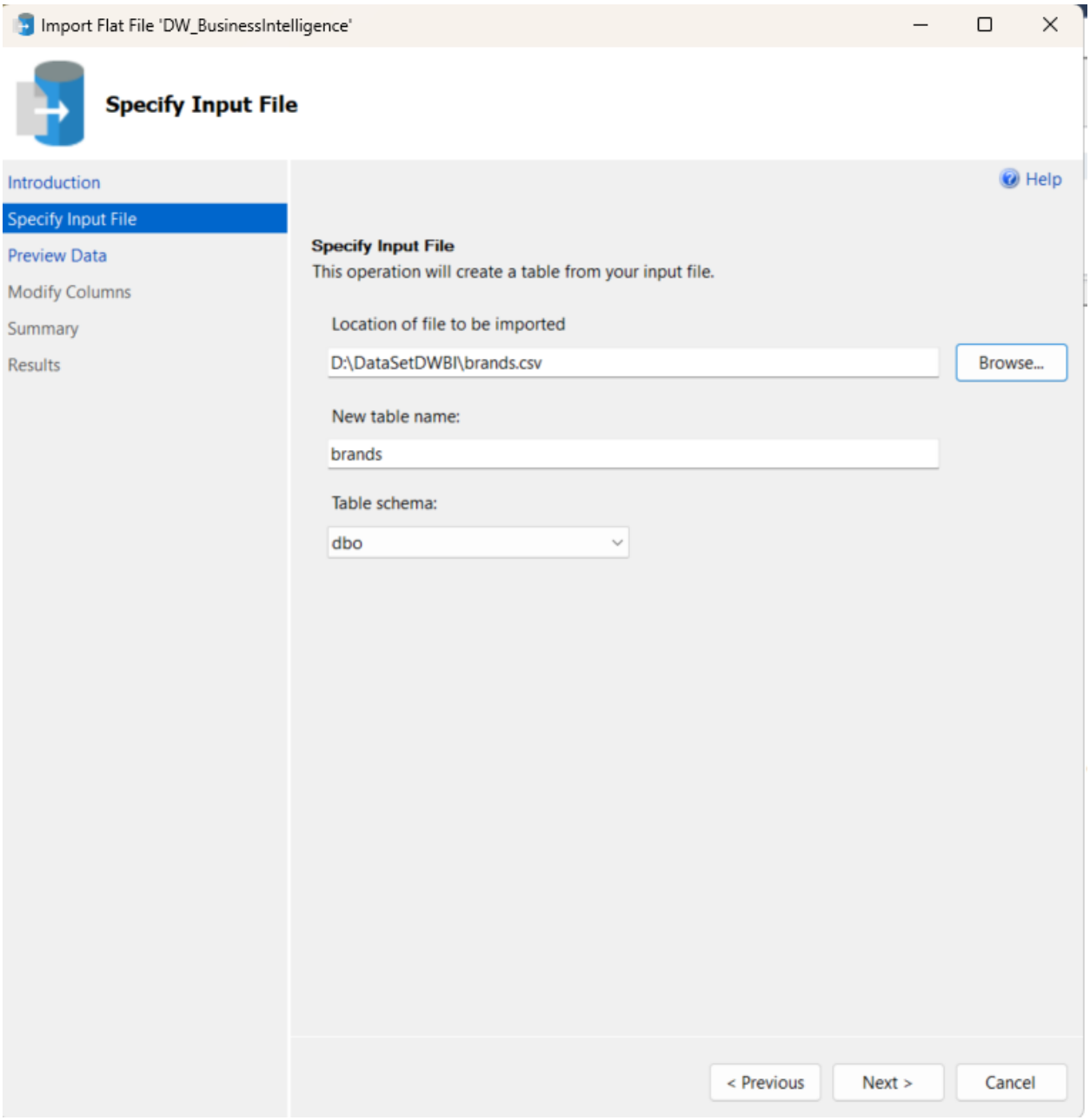
- Brands Table - **Used in CSV format**

- Locations Table - **Used in CSV format**



- CSV files and txt files were imported using Import Flat Files

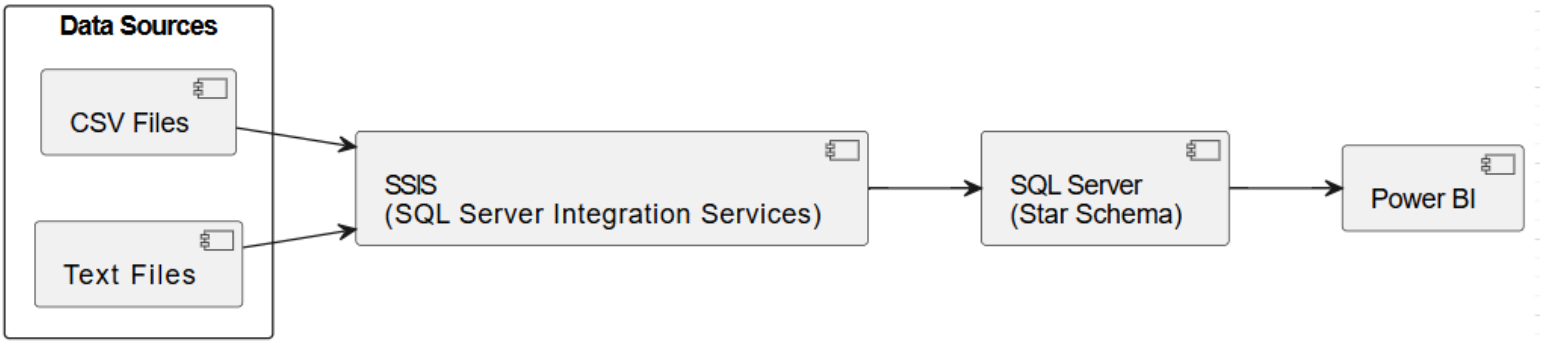



Import Flat File 'DW_BusinessIntelligence'
Specify Input File
Introduction
Specify Input File
Preview Data
Modify Columns
Summary
Results
Help
Specify Input File
This operation will create a table from your input file.
Location of file to be imported
D:\DataSetDWBI\brands.csv
Browse...
New table name:
brands
Table schema:
dbo
< Previous
Next >
Cancel

Solution architecture

A high-level solution architecture was designed to depict the flow of data from source to analytics:

- Data Sources: CSV and text files
- ETL Tool: SQL Server Integration Services (SSIS)
- Data Warehouse: SQL Server (Star schema)
- Reporting: Power BI



Data Sources (.csv/.txt)

↓

SSIS (ETL Tool)

↓

SQL Server Data Warehouse

└── **Fact Table: FactCustomer_Order**

└── **Dimension Tables:**

- Customer

- Brands

- Product Inventory

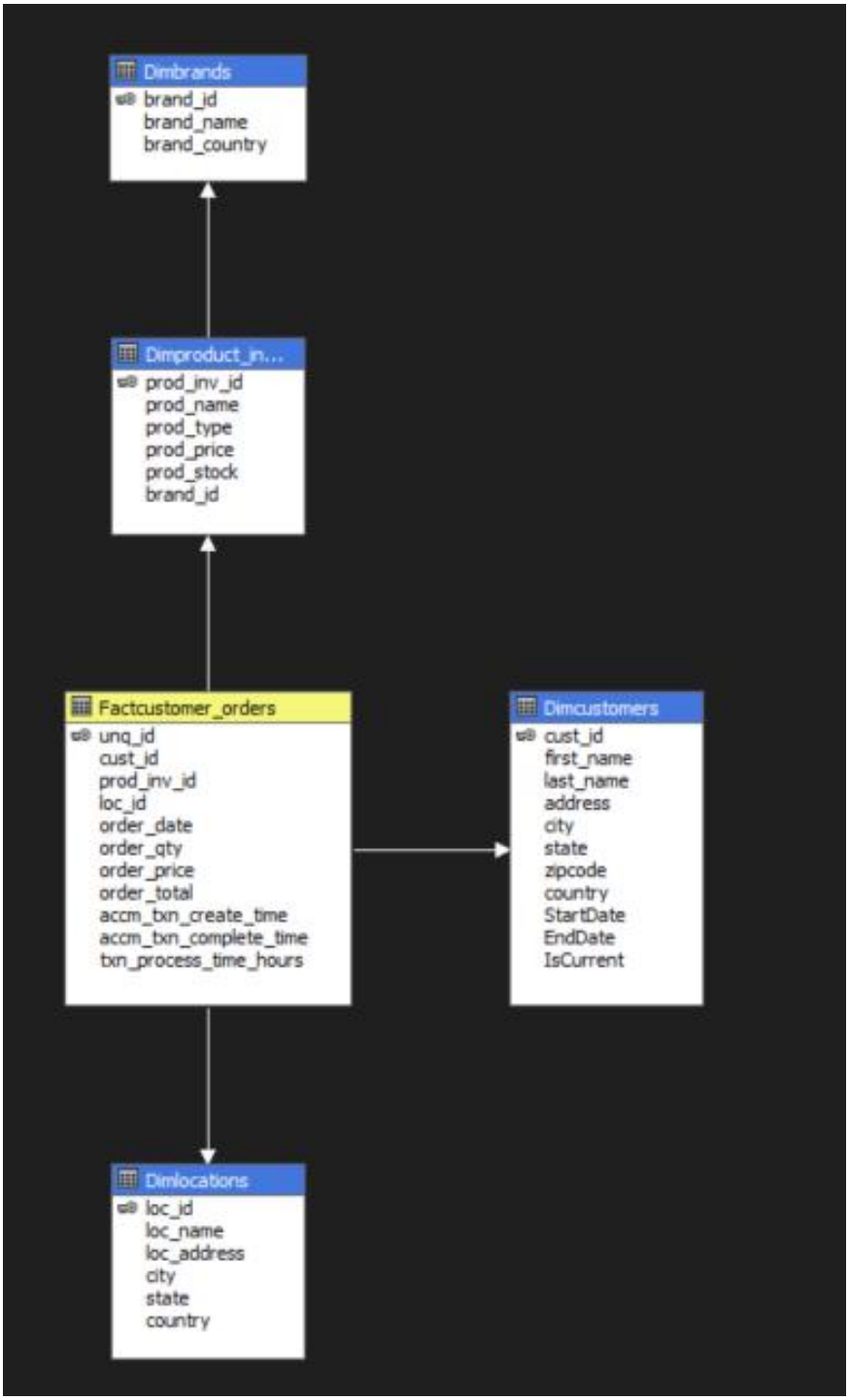
- Locations

↓

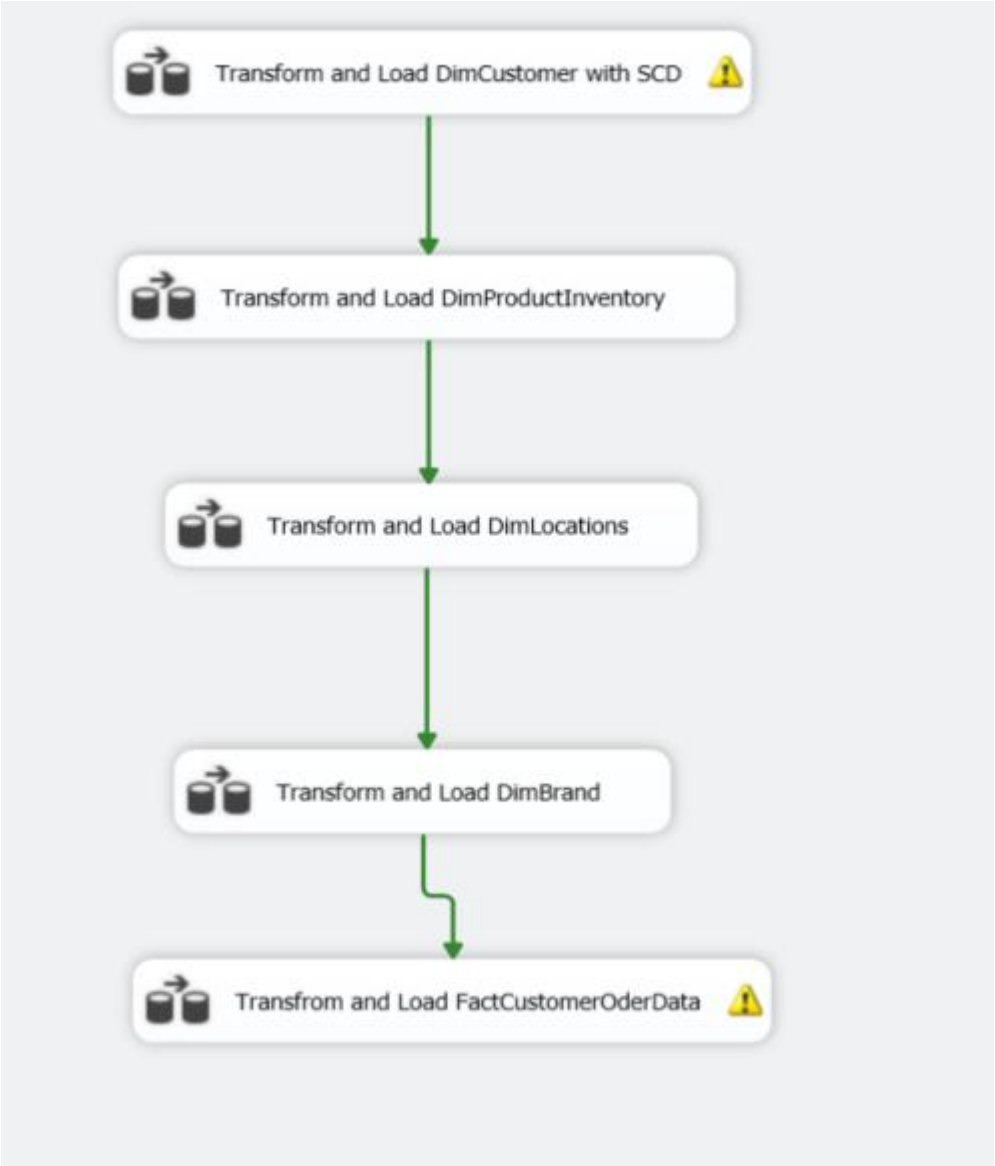
Power BI (Visualization & Analysis)

Data Warehouse Design & Development

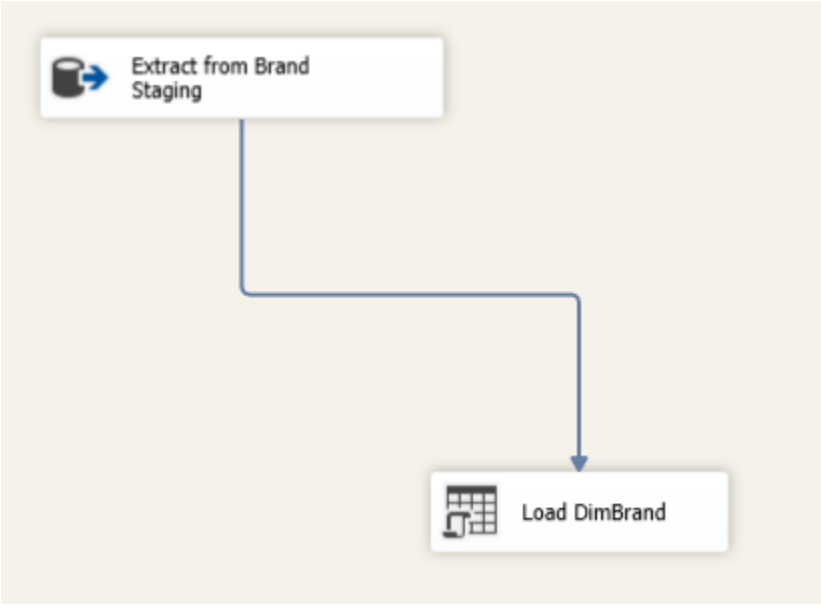
Star Schema Design:



ETL Process:

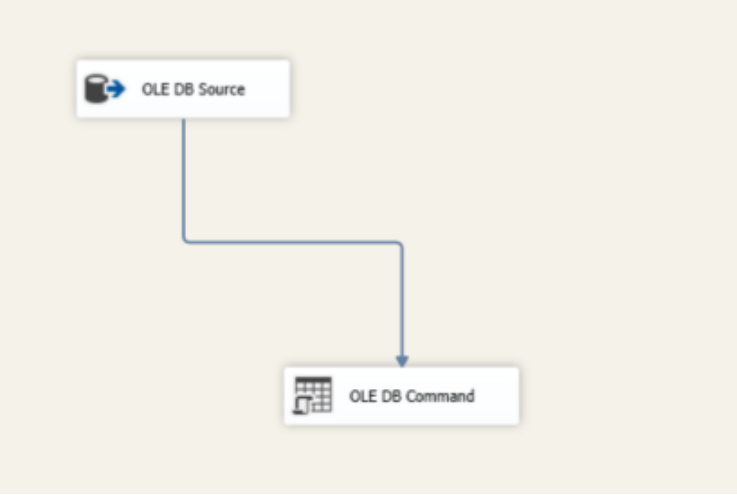

Transform and Load DimCustomer with SCD
Transform and Load DimProductInventory
Transform and Load DimLocations
Transform and Load DimBrand
Transfrom and Load FactCustomerOderData

1. Transform and load Brand Data

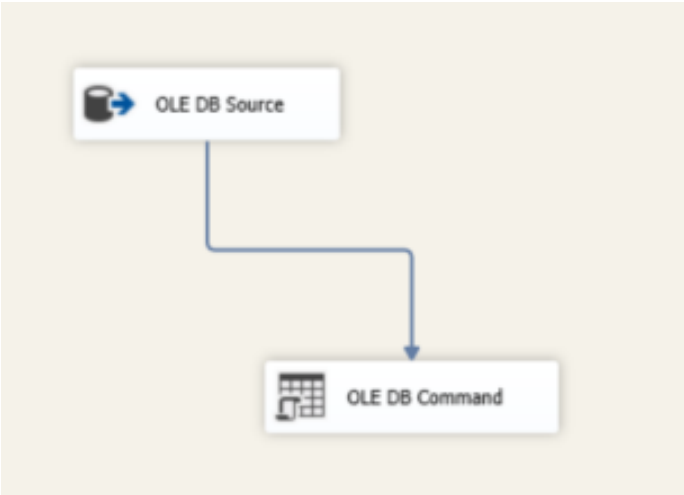


- Extracting brand-related data from data source
- Transforming the data (cleaning)
- Loading it into the target data warehouse

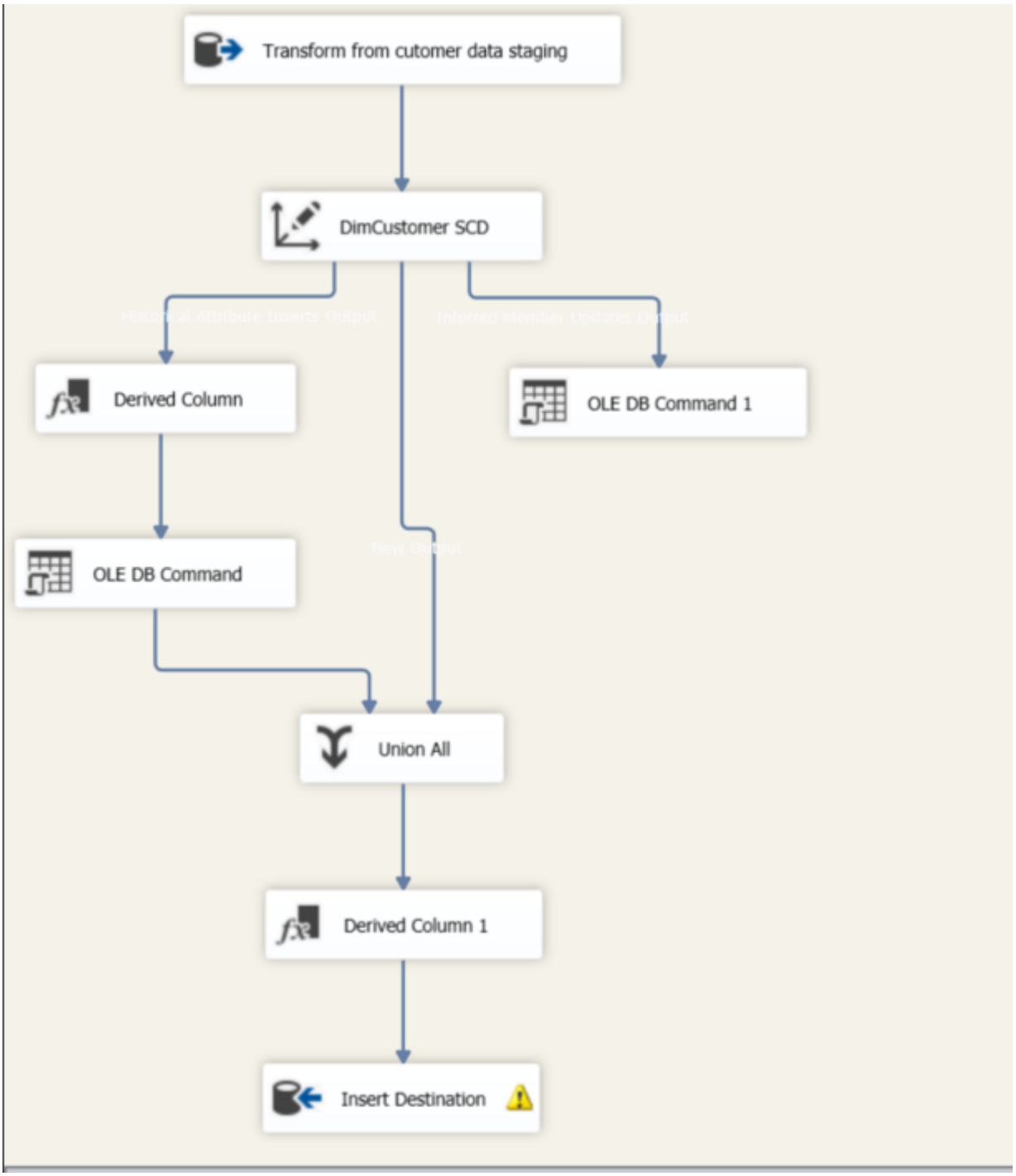
2. Transform and load location Data



3. Transform and load Product Inventory Data

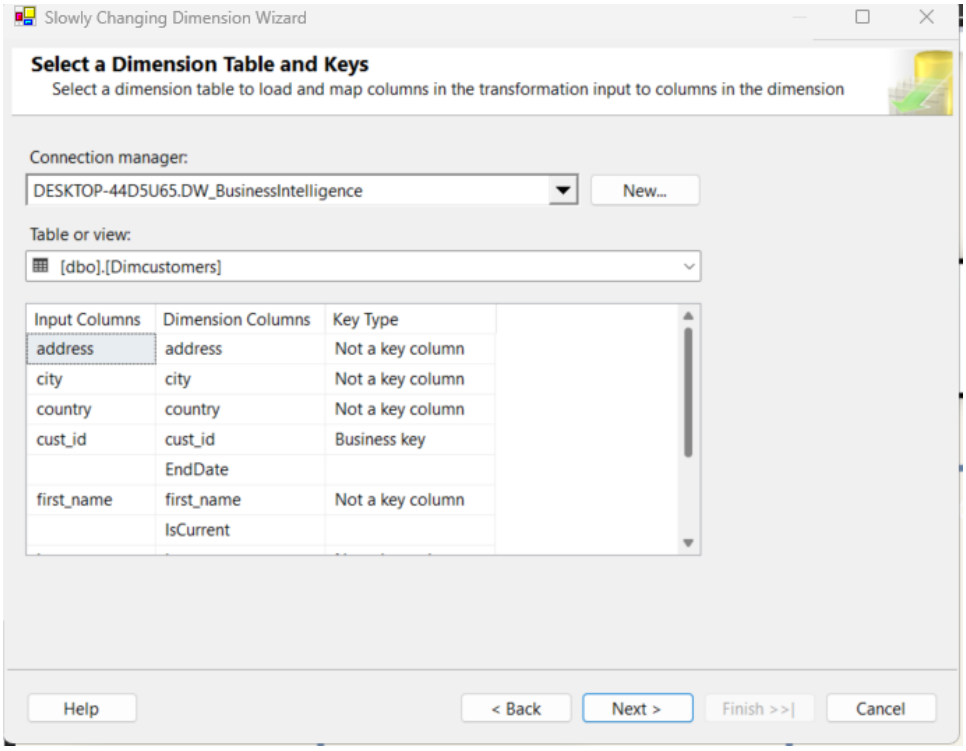


4. Transform and load Customer Data

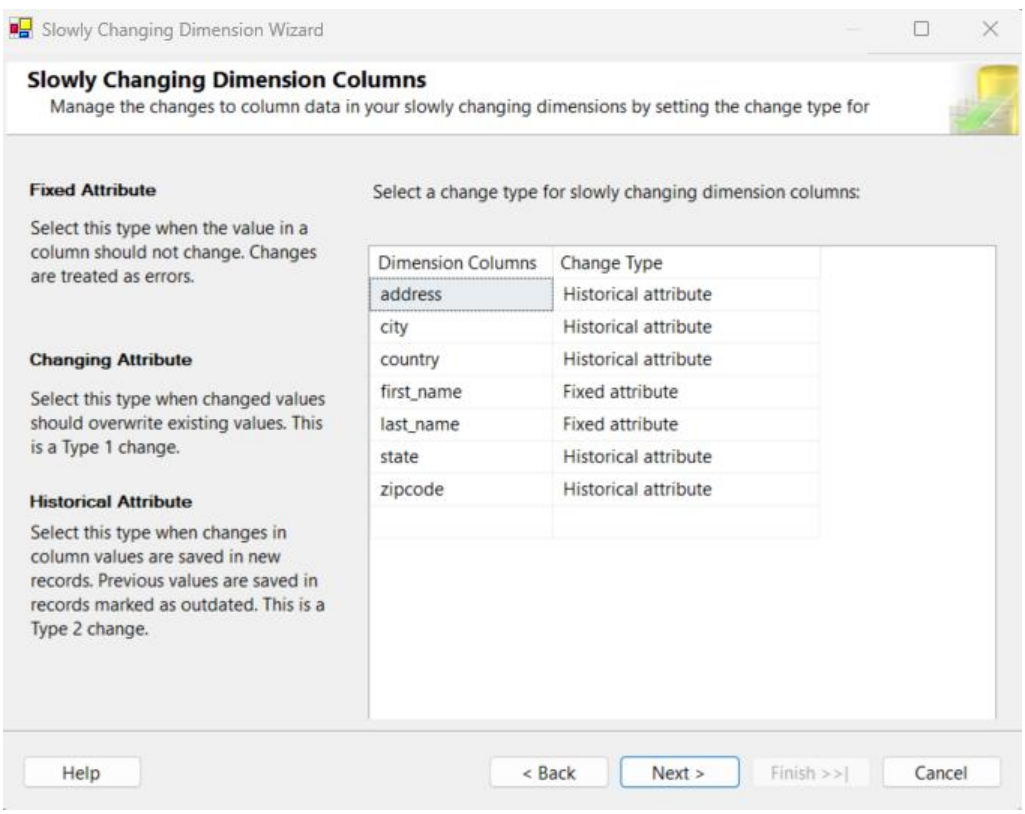


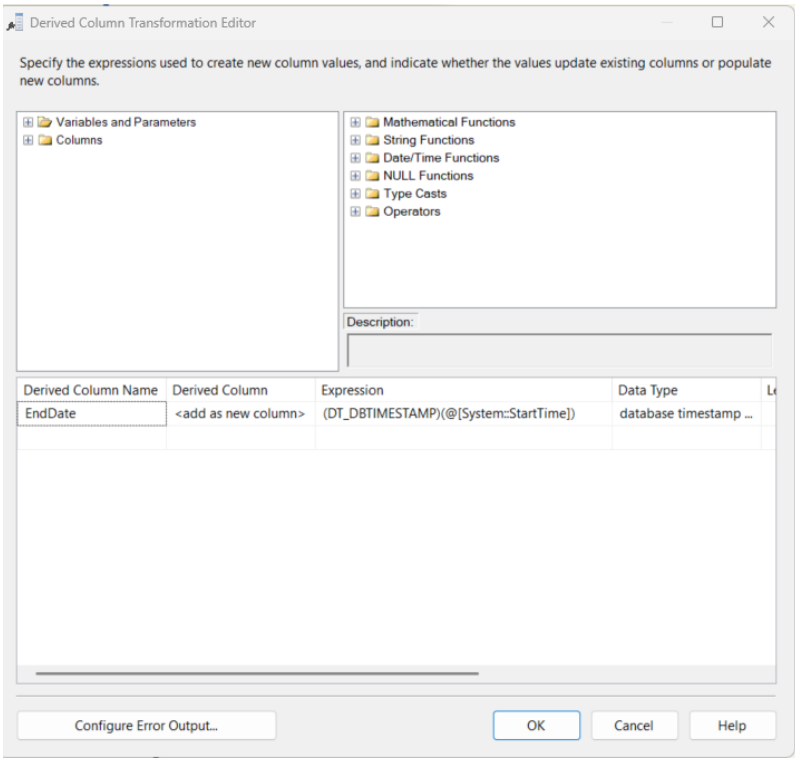
Implementation of Slowly Changing Dimension (SCD) for Customer Data

- In the data warehouse, the DimCustomer table was configured as a Slowly Changing Dimension (SCD) to track historical changes in customer address details while preserving fixed attributes such as first and last names. This approach ensures that whenever a customer's location (address, city, state, country, or zip code) changes, a new record is created with updated details, while the previous record is retained with an expiration timestamp. This design allows for accurate historical reporting, ensuring that past transactions are linked to the correct customer location at the time of purchase. By marking address-related fields as historical, we maintain a complete audit trail of customer movements, which is crucial for regional sales analysis, compliance, and customer behavior tracking. Meanwhile, name fields (first_name, last_name) were set as fixed attributes as they rarely change and overwriting them ensures consistency in customer identification across reports

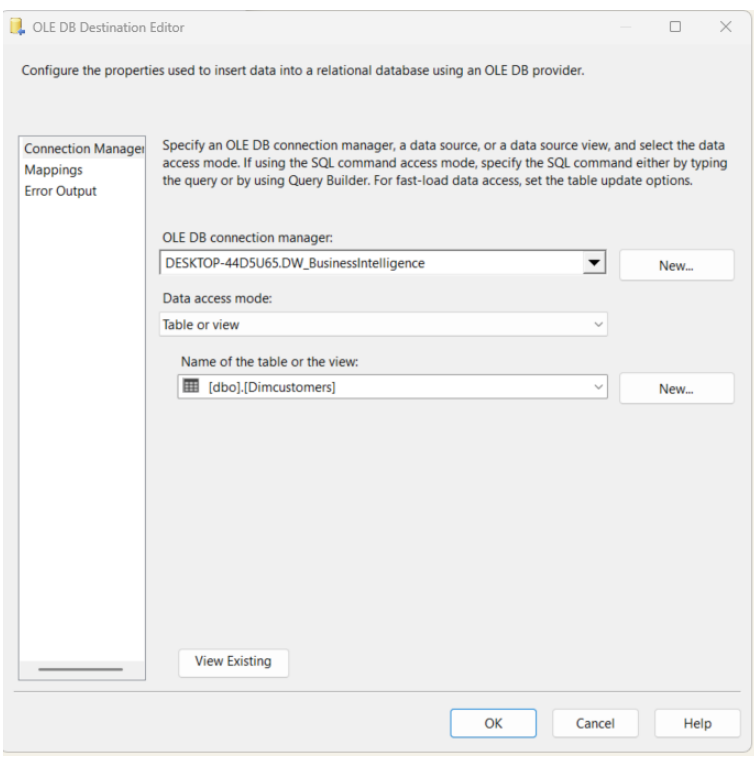


- Cust_id is placed as business key

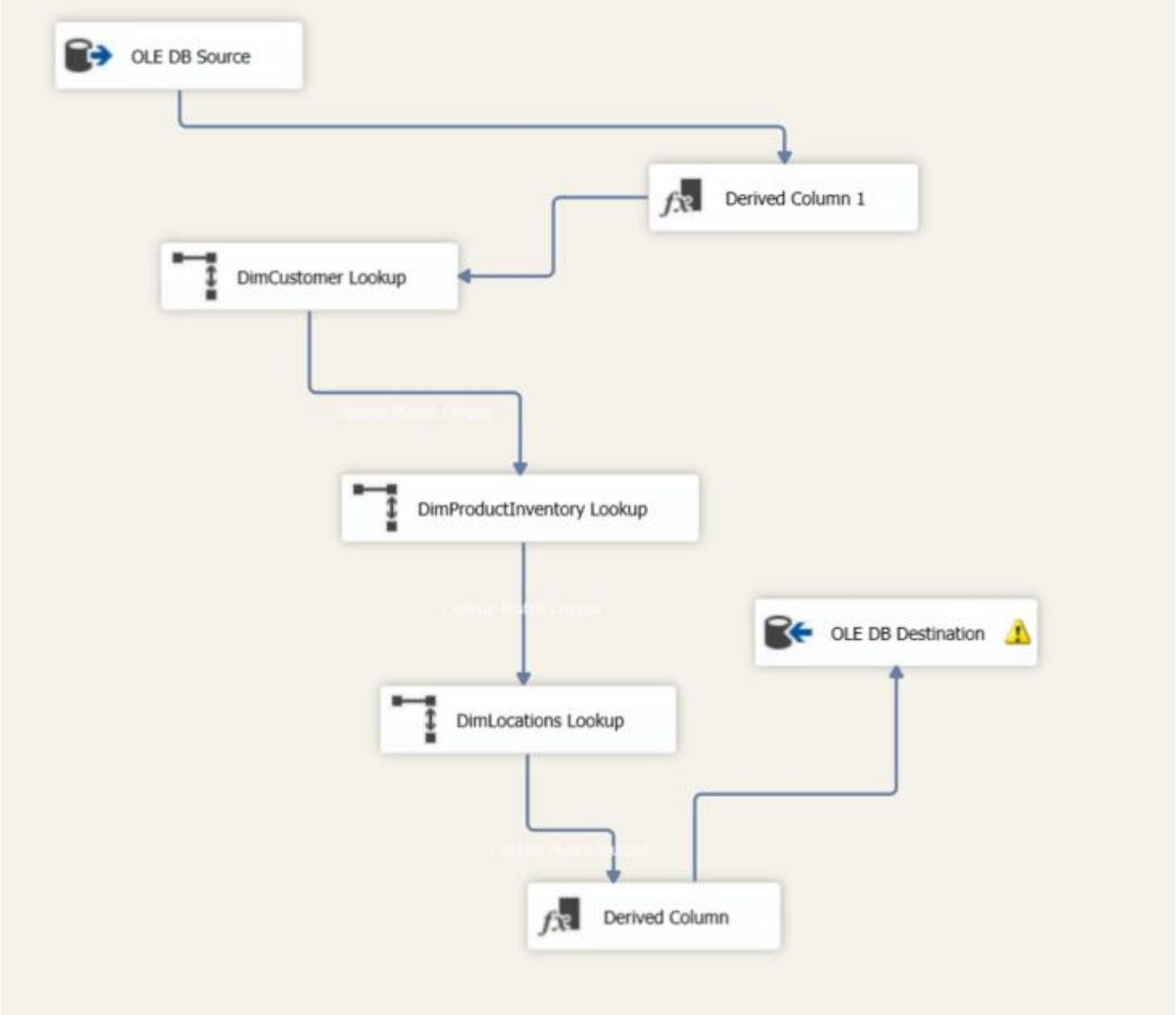




Finally loaded to the data warehouse

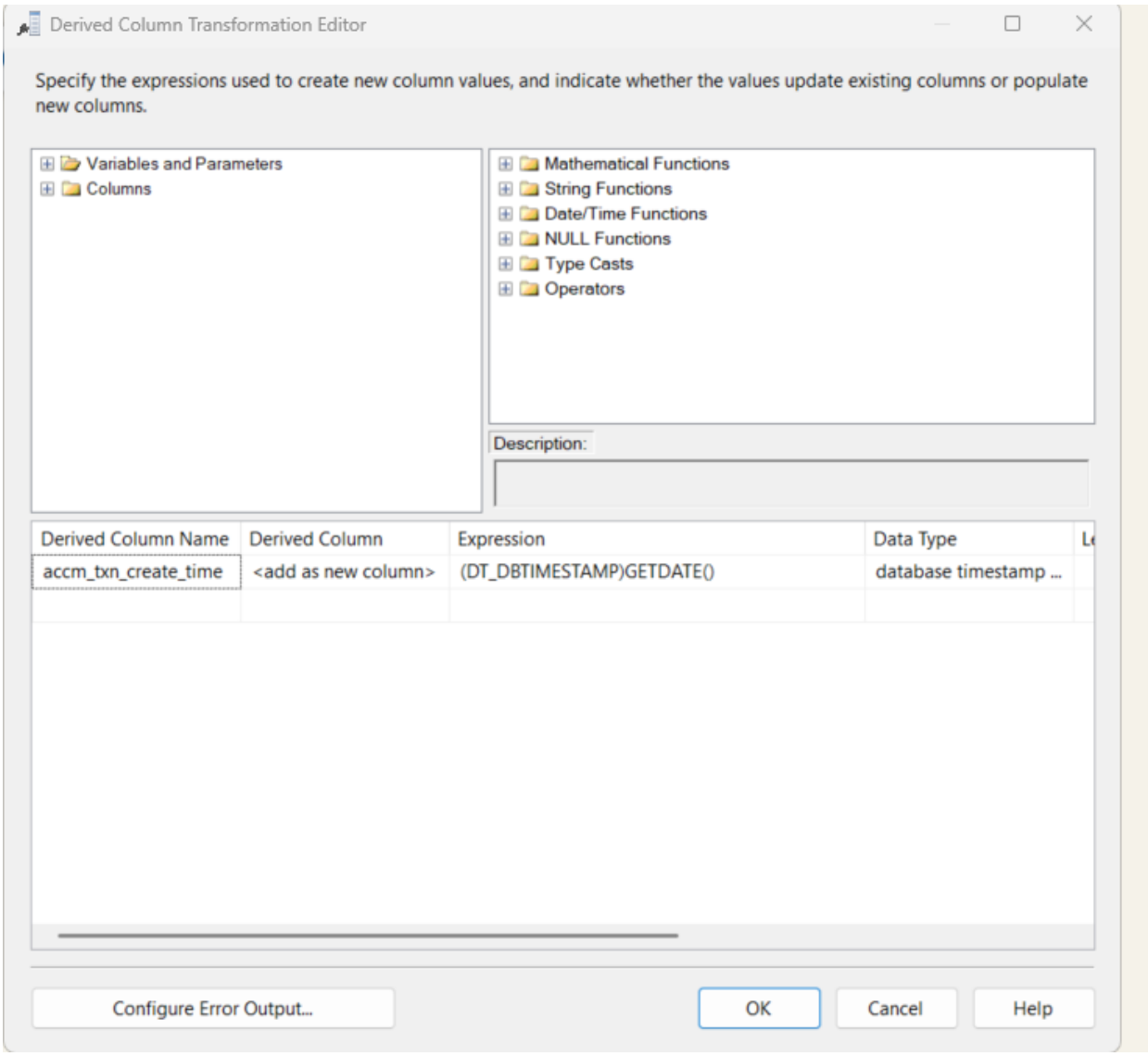


5. Transform and load FactCustomerOrder table

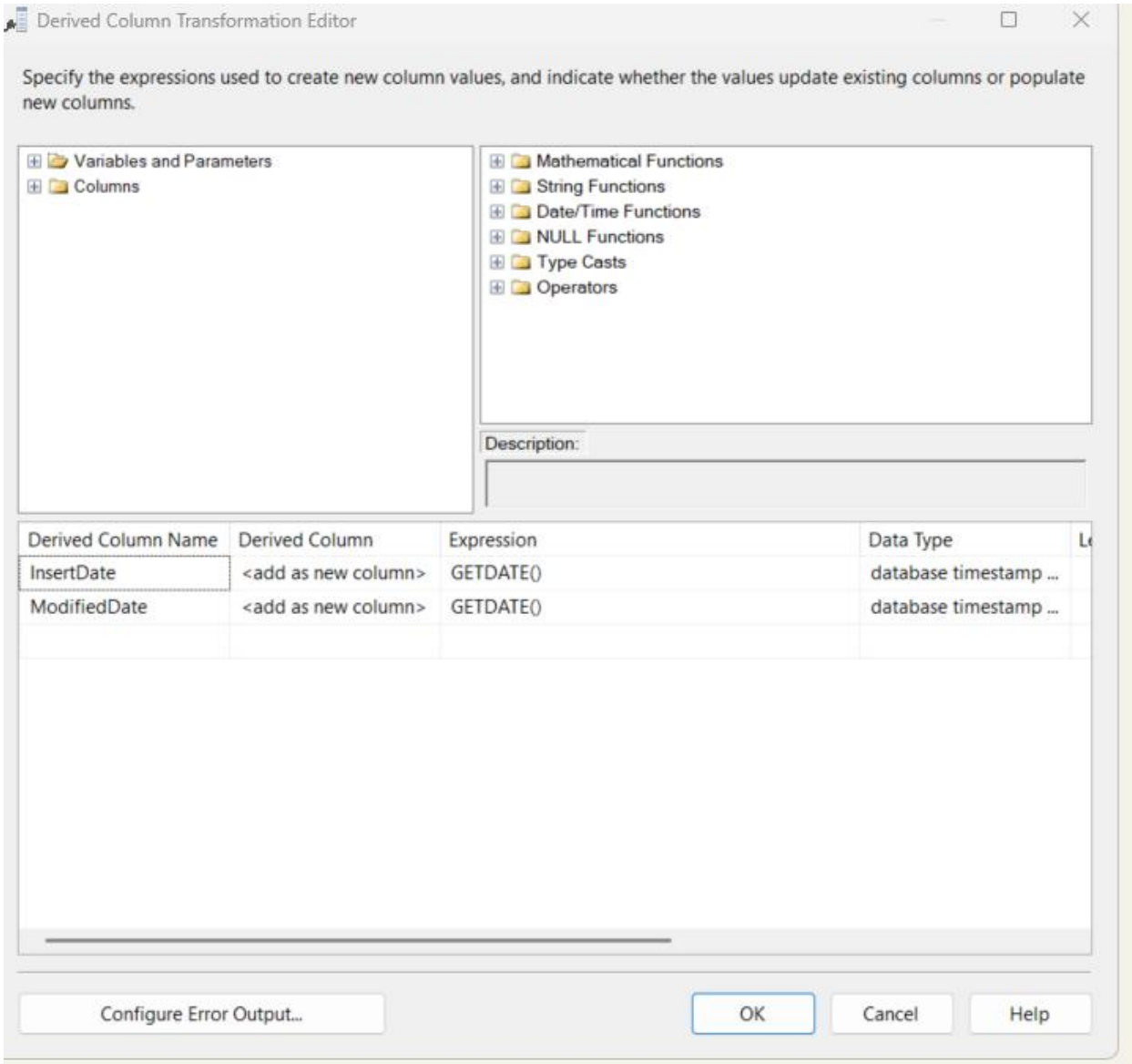


- The fact table is linked to the corresponding dimension tables using lookup transformations.
- Specifically, the CustomerOrder fact table is joined with the Customer, Location, and Product Inventory dimension tables to establish referential integrity and enable comprehensive analytical queries.

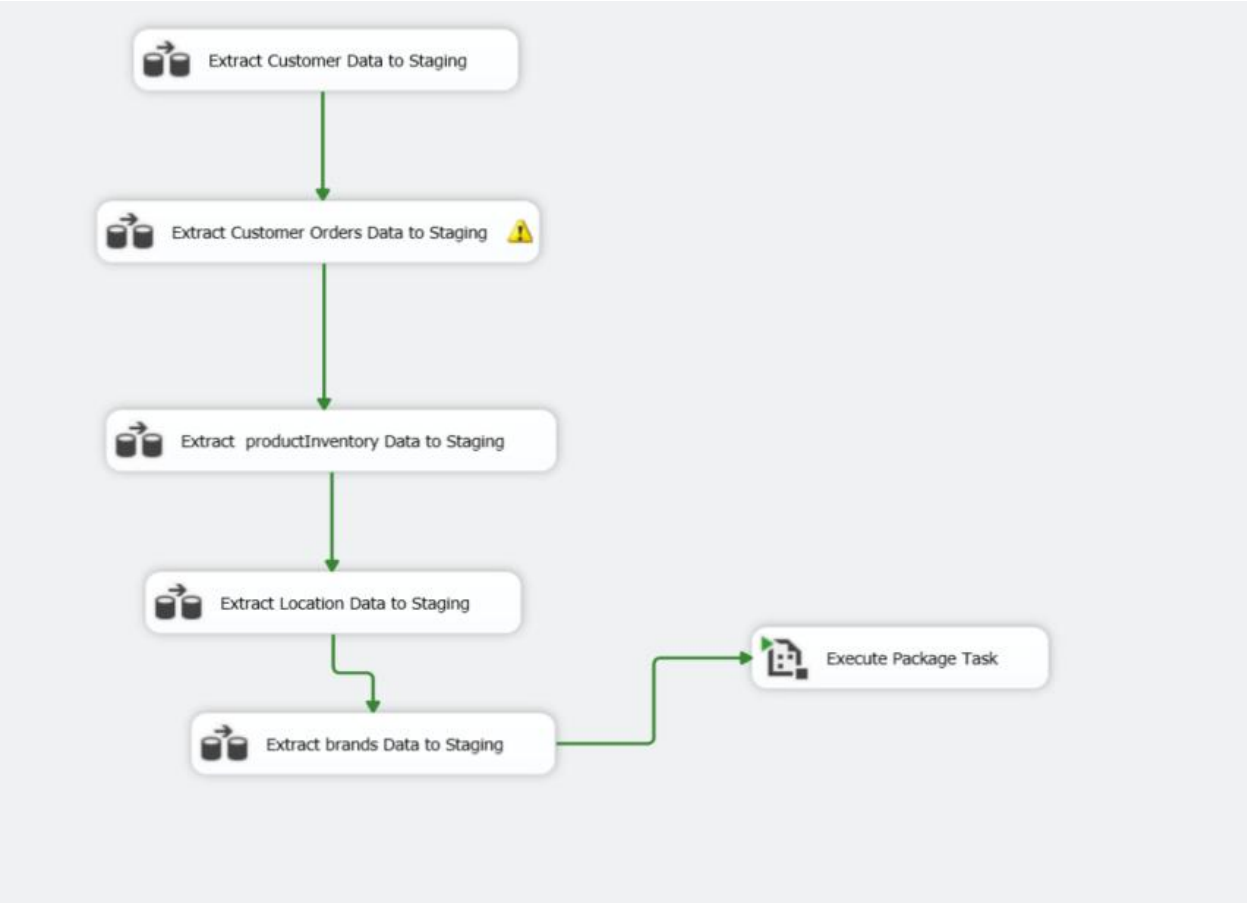
- Additionally, the created_date field is derived using the GETDATE() function, based on the value of accm_txn_create_time.



Additionally, the InsertDate and ModifiedDate field is derived using the GETDATE() function, based on the value of



Load to staging



- We have implemented a structured data pipeline where the data is first loaded into the staging area before being transferred to the data warehouse.

- The Execute package task option demonstrates the integration between the staging and data warehouse packages, ensuring that once the staging process is completed, the data is automatically loaded into the data warehouse.