1. Data from 3 different sources imported in 3 different tables

This is achieved through Snowflake, Using Internal Stage all 3 files are imported.

Then using COPY INTO command, data from 3 files is imported to 3 different raw tables.

Database created - data\_analyse\_pei

Schema Created – Raw\_Data

Tables Created – Customer\_Raw, Order\_Raw, Shipping\_Raw

1. Accuracy, Completeness and Reliability checks done on Raw data from Tables

To check no negative order amounts

To check special characters in names

Count of duplicate customers in shipping table

To check No missing customer\_id in Order and Shipping Raw table

Count of NULLs per table

To check orders with non-existent customers

To check shipping with non-existent customers

1. Transform Data and Create Dimension and Fact Tables

Here for analytics Purpose, we have created a start schema where

Order table is fact table

Customers and Shipping is dimension table

Following Transformations are performed –

Replacing special characters in first and last name columns of Customer table

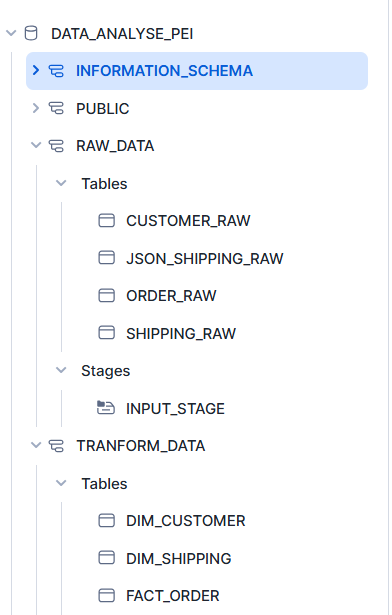
In shipping dimension table, Customer IDs records are created along with Latest Shipping Id and Shipping Status removing duplicate records

In Fact Order Table, for Latest Shipping Status column, value is fetched from shipping dimension table and if record not found column is filled with “Yet to Ship”

1. Business Reporting Requirements

Added SQL queries for business report requirements mentioned in document by using created dimension and fact tables.

Attaching screenshot of Snowflake’s created database for reference –



SQL script written in Snowflake environment is loaded in Github link