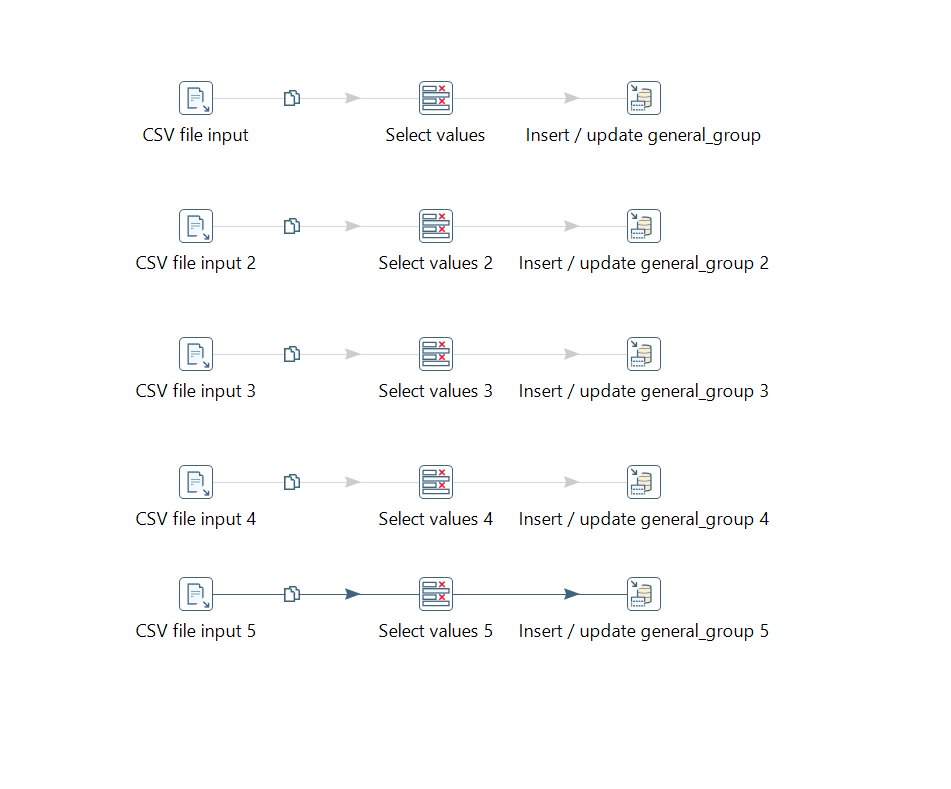
**Brief Process Documentation and Explanation**

**A. Tools I Used to include the following.**

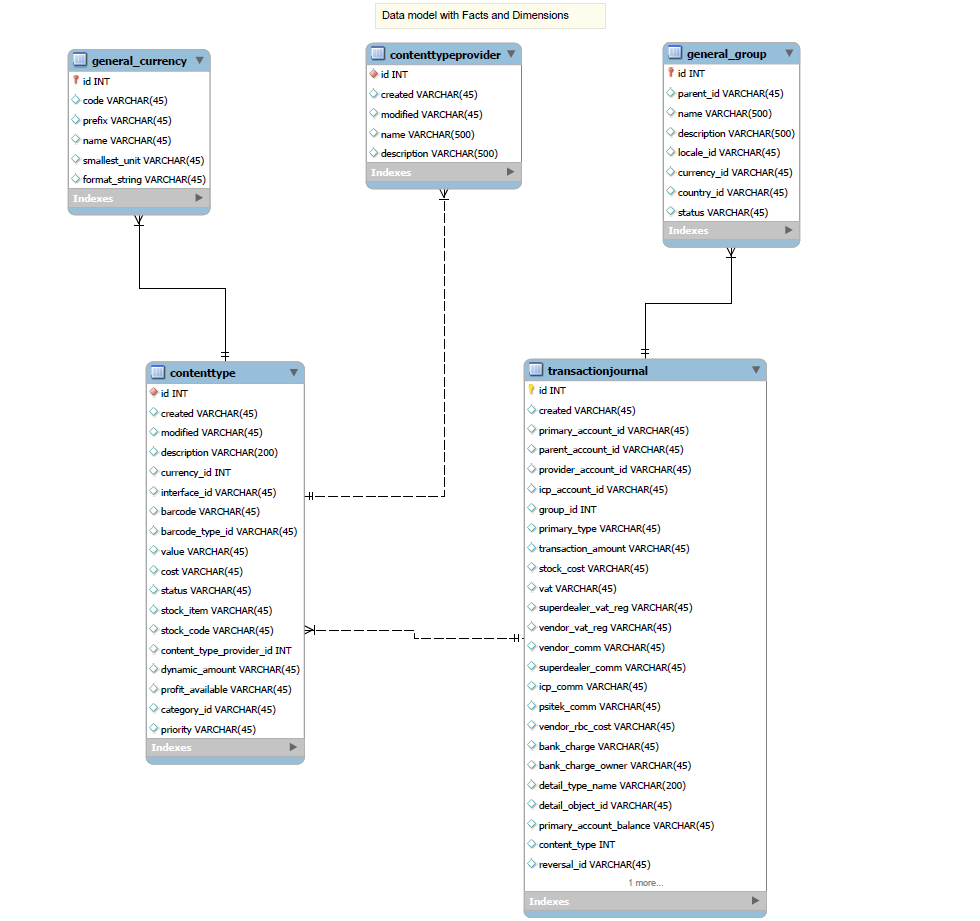
1. Pentahol Data Integration
2. Pyspark with Python
3. Pandas and matplotlib
4. Jupyter Lab (Anaconda)
5. Microsoft SQL Server Management Studio
6. Notepad++

**B. Steps in Extracting and Transforming the data**

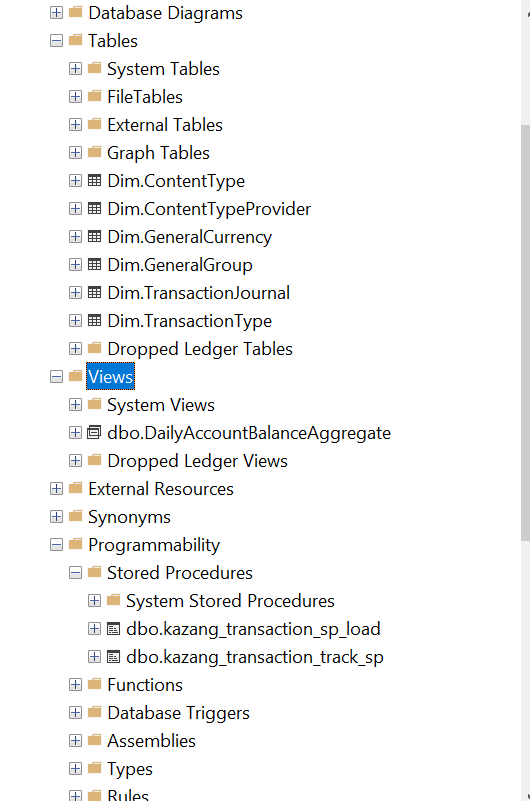
1. I created the various tables for the CSV’s. See create and script in the folder
2. Using Pentaho to extract, transformation and loading the CSV into the MSSQL database



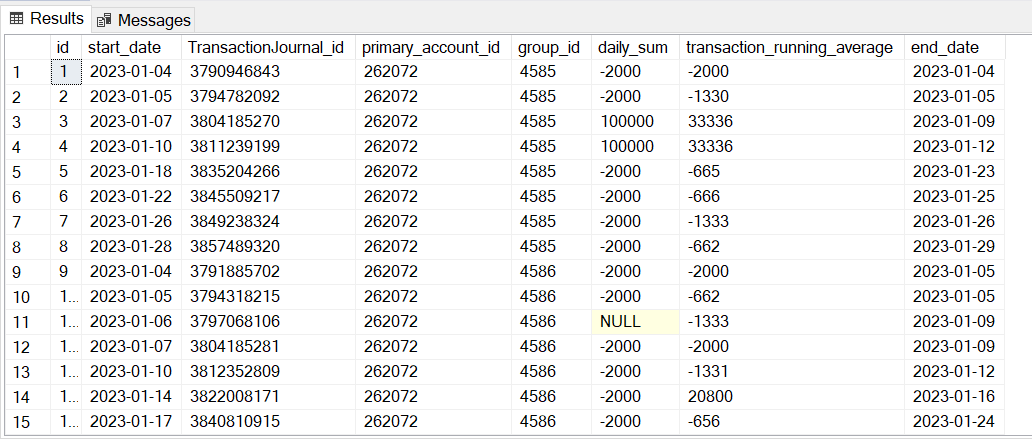
1. Using Pyspark and Pandas as well as Jupyter lab (IDE) to process by extracting, transforming, and loading the kazang\_transaction\_types.py data into the MSSQL ArchiveRawData database server. Script used in the assessment folder namely Loading\_kazang\_transaction\_data.ipynb
2. Creation of Dim and Fact Table in the KazangTransaction MSSQL database. Scripts for DIM and Fact in the assessment folder. Also, script for creating the entire database in the folder as well.
3. Data Model diagram to indicate the relationship as well as every other attributes. See diagram below.



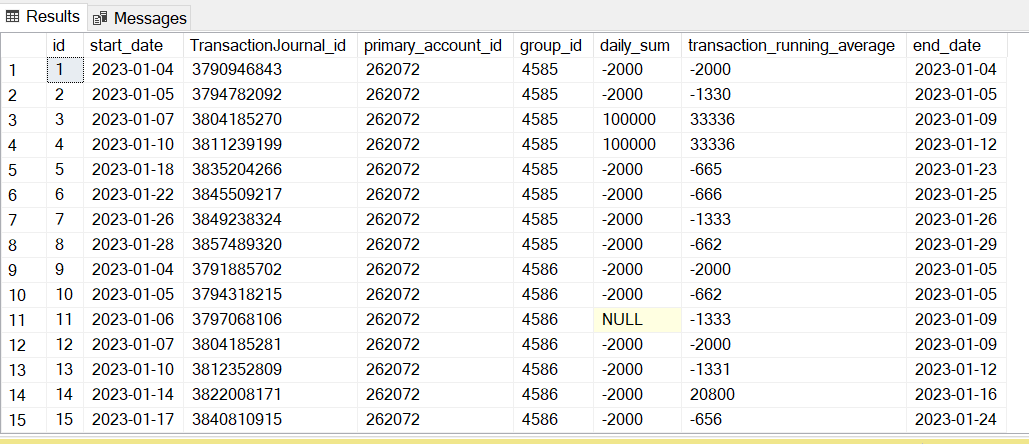
1. Both store proc and Data view were created. Scripts also can be found in the same folder.



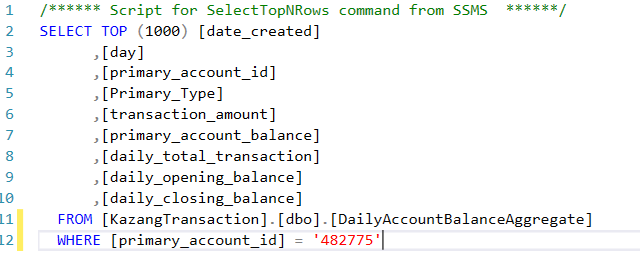
1. Dimensional table and procedure that tracks Primary Accounts’ group movements. Script saved in the assessment folder

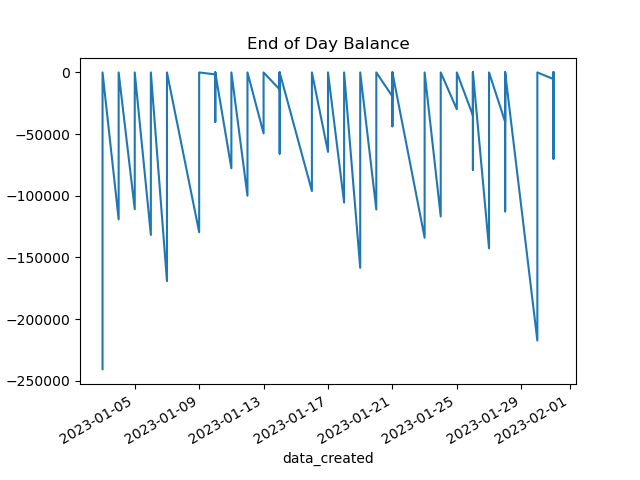


For results of Group dimension table for primary\_account\_id = 262072

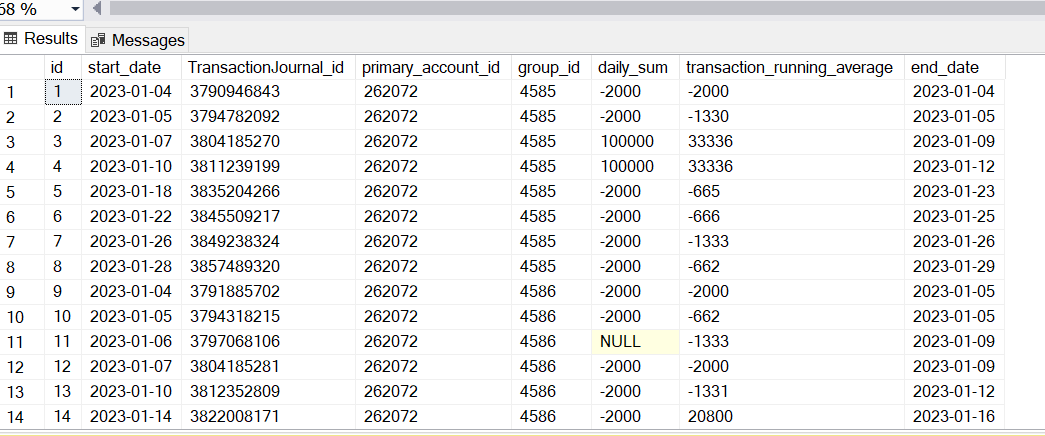


1. Daily aggregated table or data view along with procedure that tracts the opening and closing balances of accounts (primary\_account\_balance) along with debited and credited amounts. Daily Balance table (file) and script are in the assessment folder





1. results of Group dimension table for primary\_account\_id = 262072



1. full results transaction type dimension table.

