

Assigned on: October 25, 2022  
Due on: November 15, 2022

ISTM 637  
Fall 2022

Points: 75  
Joint Project

## Design and Implementation of a Data Warehouse for a Retail Store

### “Integrated” Report 3: Requirements Analysis, Data Design and ETL Design using SSIS

#### Overall Objective of this Project

The objective of this project is to design and develop a data warehouse for a retail store chain called Dominick’s Fine Food (DFF) that has many branches. We use their store level data to set up the data warehouse and help them analyze the data.

#### Project Scenario

The data for this project includes store-level data covering shelf management and pricing from all the branches of DFF. The data can be accessed from the following web site:

<http://gsbwww.uchicago.edu/kilts/research/db/dominicks/> or from Da Vinci.

#### Overall Objective of this Report

This report focuses on the design and implementation of the ETL operations and pumping the data into the warehouse tables. The tasks include:

1. All operations for this report must be done in SSIS environment. No use of Python, Excel, etc. for this report.
2. Data should be sourced from text files or CSV files as the current data sets are in CSV format.
3. Use only SSIS to store text/CSV data files first in a data base called “unique name-staging-area” in **SQL Server 2016**.
4. Create the independent data marts in a database called “unique name-dw-area” if you have not done so at the end of Report 2.
5. Write a report that shows
  - a. **Development of an ETL plan** (Include all these into the report.)
    - i. Write all the target data needed in the data warehouse.
    - ii. Write all the data sources (what is available from DFF).
    - iii. Prepare and write data mappings in Excel for data elements from sources to staging and then data mapping from staging to data warehouse (include all transformations).
    - iv. Write all comprehensive data extraction rules.
    - v. Write all data transformation and cleansing rules.
    - vi. Write plan for aggregate tables.
    - vii. Write out the organization of data staging area.
    - viii. Write procedures for all data extractions and loadings.
    - ix. Write the ETL for dimension table.
    - x. Write the ETL for fact tables.
  - b. **Implementation of the ETL**
    - Perform necessary cleansing, summarizations and other transformations as needed in the staging database. These operations will be different for each group.

- Look at the data granularity needed at the independent data mart level. Include these in the report.
- Start loading the data into the independent data marts.
- Once loading of each table is done, remove all temp tables from the staging area. Make sure to write a list of these temporary tables in the report.
- c. **Make sure to include ALL mapping table(s) in Excel in the report.**
- d. **Make sure to include ALL SQL statements that you have used for the ETL operations.**
- 6. **Submit** a report that shows all these tasks that you have done for ALL ETL operations and the snapshots of the before-after table contents and the ETL runs. This is important as it will show how effective your ETLs have been.
- 7. Report should also have text to tell the story and be self-explanatory.

### Submit

Your group needs to submit a report that shows all these tasks have been done. **Make sure to add the contents of reports 1 and 2 with this report. If you need to edit the previous reports, please do so now. The final report to be submitted should be an integration of all three reports.**

### Grading

- Presentation of the report 10%
- **Quality integration of three reports 10%**
  - Include summary of Reports 1 and 2. All three reports must feel like one integrated report
- Quality of the content 80%
  - **Provide the ETL plan as discussed in #5**
  - Make sure to show all works including the ETL **by capturing the screens.**
  - Detail out all your E, T and L processes also in text
  - Discuss and highlight any special things that you have done in your ETL tasks.