Feiyang Yu

INFO 498 C

2019/08/26

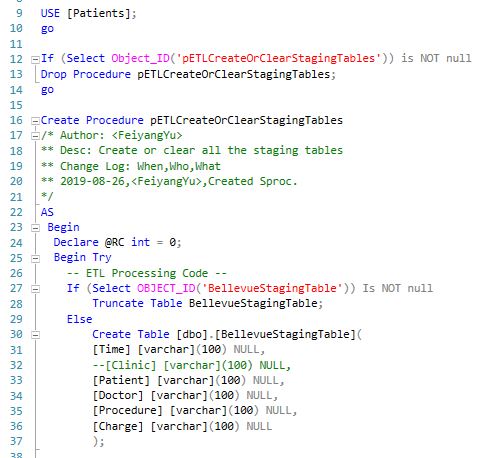
INFO 498 Final Project Technical Overview

Introduction:

For this final project, the task is to create a database for a patient-doctor appointment system and load all the data into it. The new patient visit data are created and needed to be imported to the database consistently from the csv files. This ETL process would be divided into two parts: from csv files to OLTP database, from OLTP database to database warehouse.

Topic:

First, staging tables should be created in Patients database(Figure 1) for storing newly inserted data. All data would be varchar 100 type when they are imported into the database.



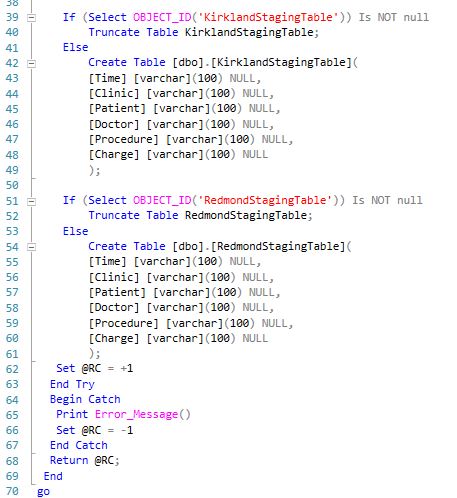


Figure 1

Next, create a procedure to import new data to staging tables(Figure 2).



Figure 2

Create a view for these data(Figure 3).

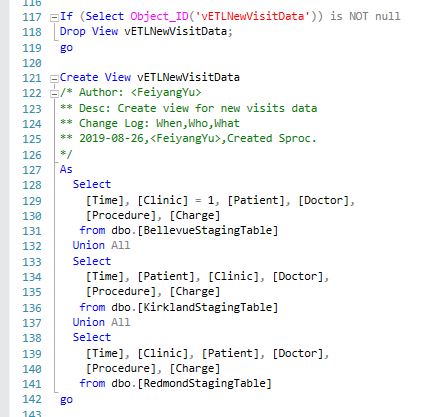
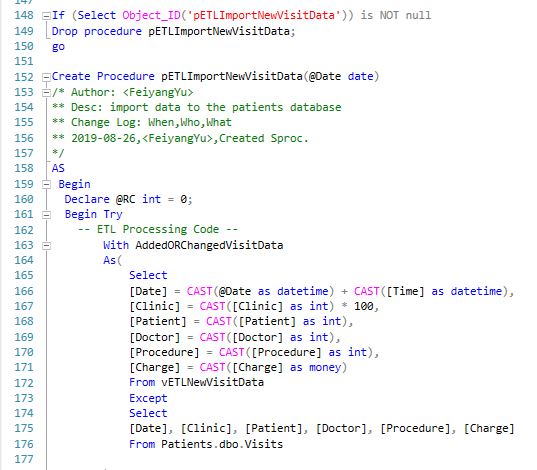


Figure 3

After that, create a procedure to import the new data to the Visits table in Patients database by using the view we just created(Figure 4).



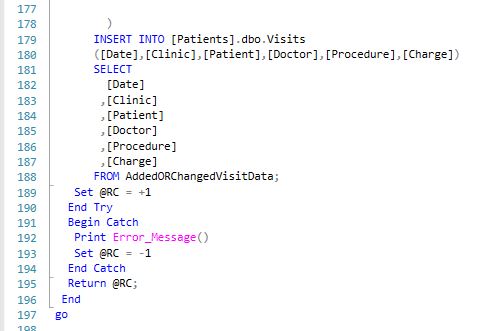
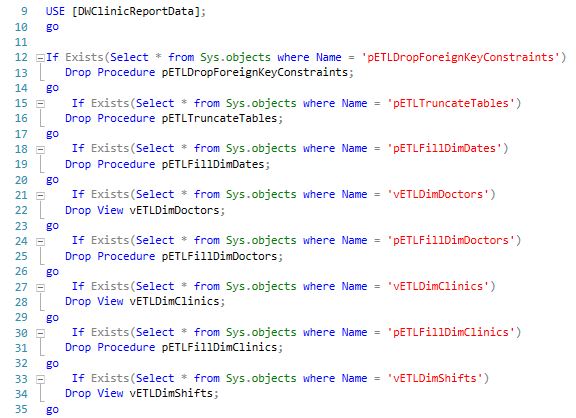


Figure 4

By executing these codes, the new data would be successfully loaded to the Patients database. The first part of ETL process is finished.

First, in DWClinicReportData database, we need to check and clear any existing views and stored procedures(Figure 5).



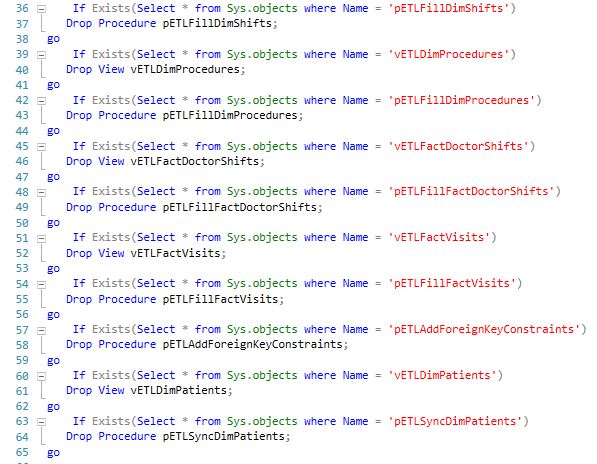


Figure 5

And then we need to drop all the foreign keys(Figure 6) and truncate all the tables(Figure 7) for the flush and fill process.

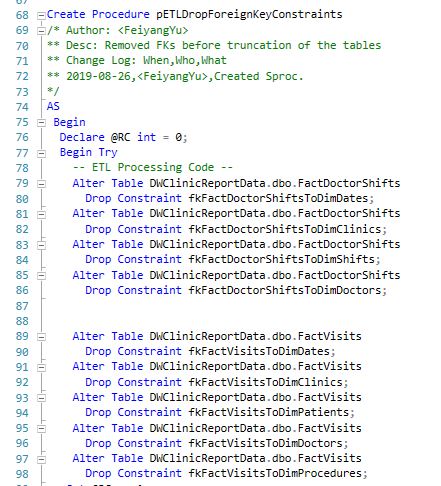


Figure 6

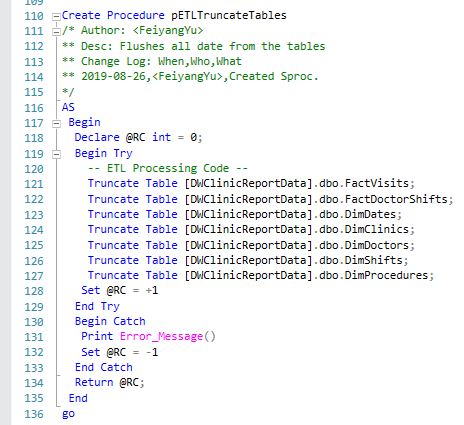


Figure 7

After that, create a view for each data warehouse table and a stored procedure to import the data to the data warehouse table(Figure 8-12).

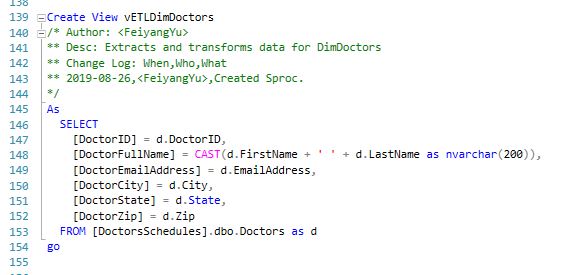




Figure 8

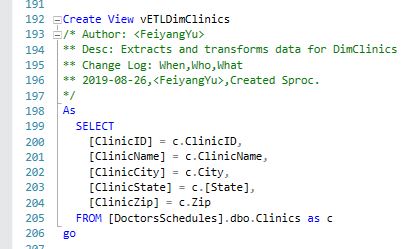
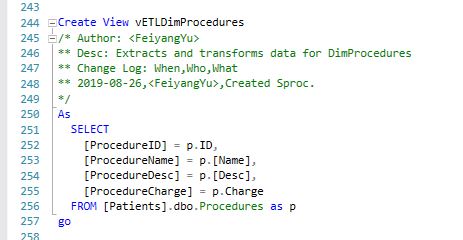




Figure 9



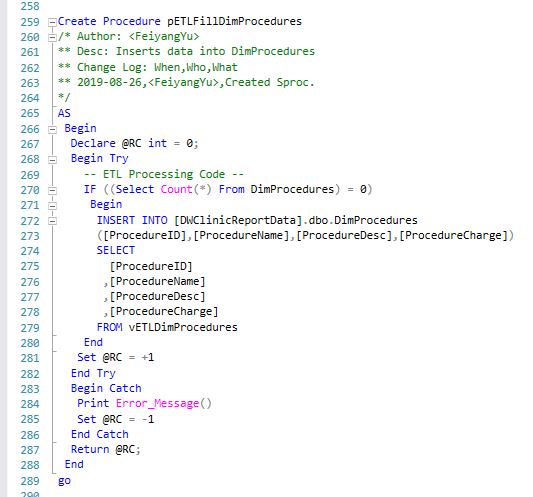
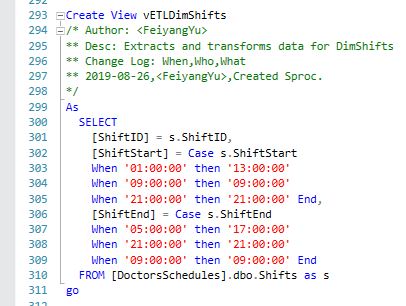


Figure 10



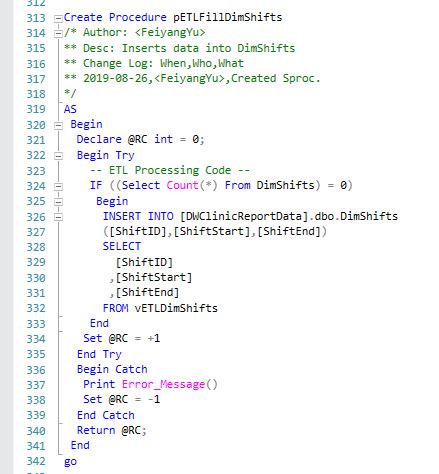


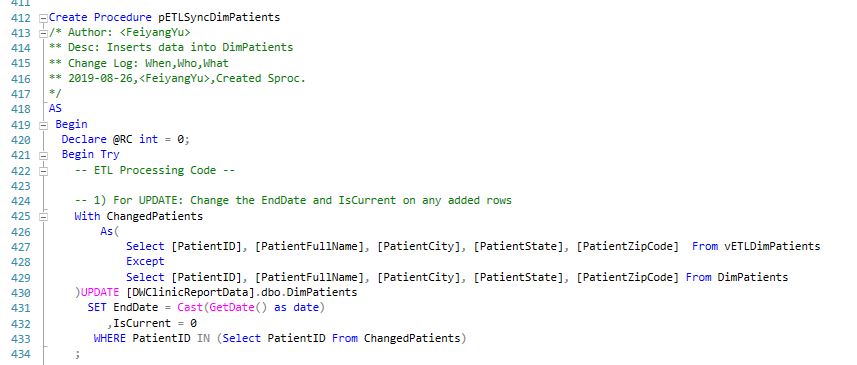
Figure 11

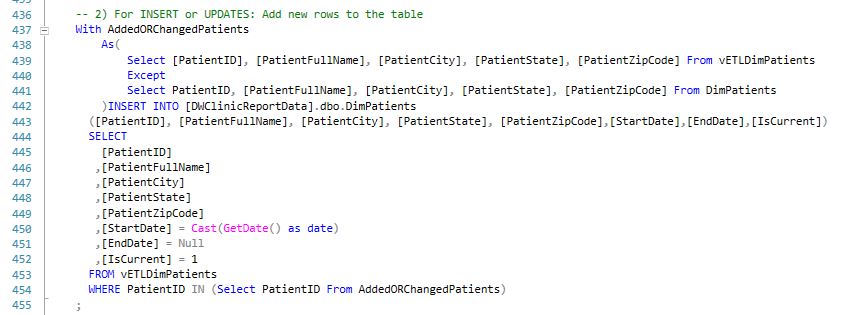
For DimDates table, only a stored procedure is needed since there is no data would be loaded into it. It’s more like a calendar reference functional table.



Figure 12

Specifically, for DimPatients table, we need use incrementally loading approach. So only a stored procedure is needed(Figure 13).





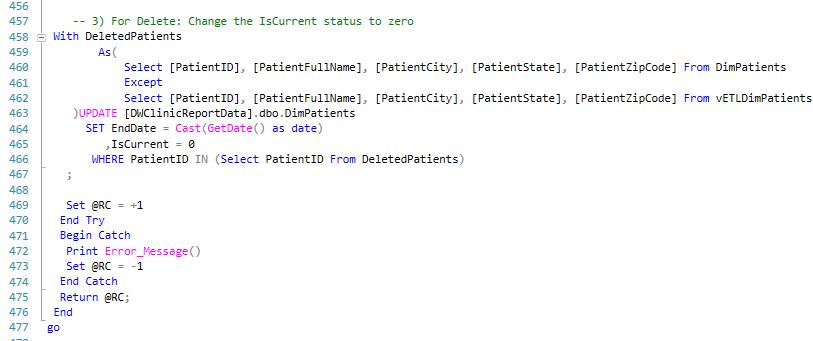
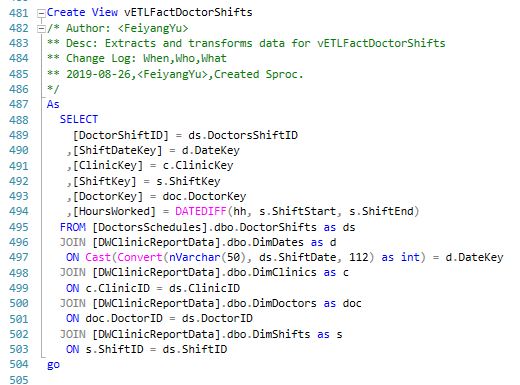


Figure 13

In the end, create views and stored procedures for two fact tables(Figure 14-15).



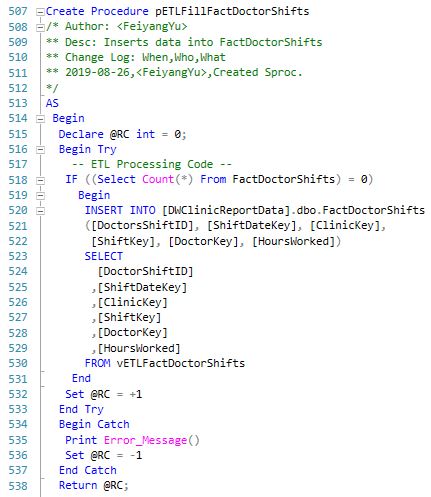
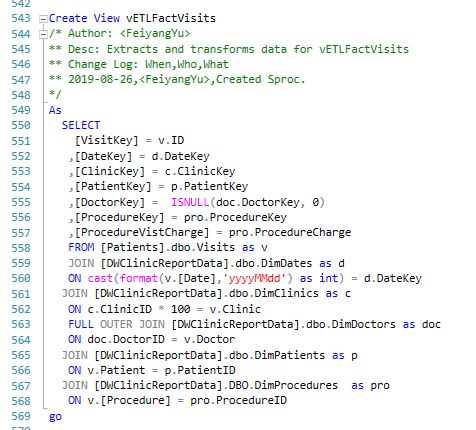


Figure 14



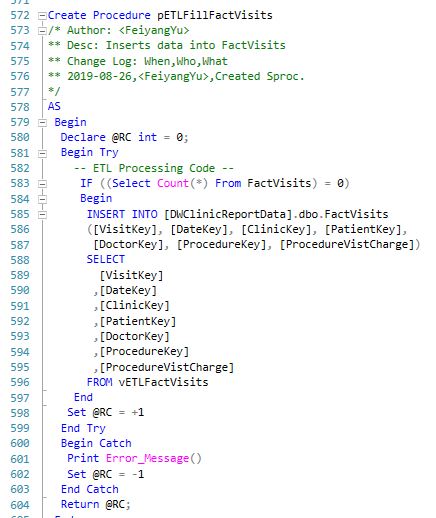


Figure 15

After we finish all the flush and fill processes, all foreign keys are added back(Figure 16).

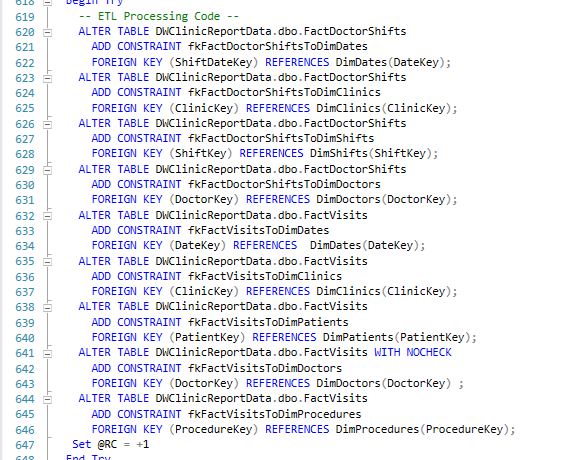
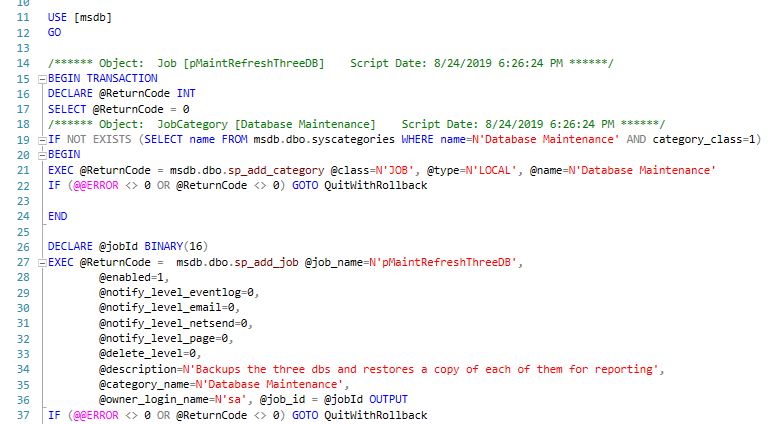
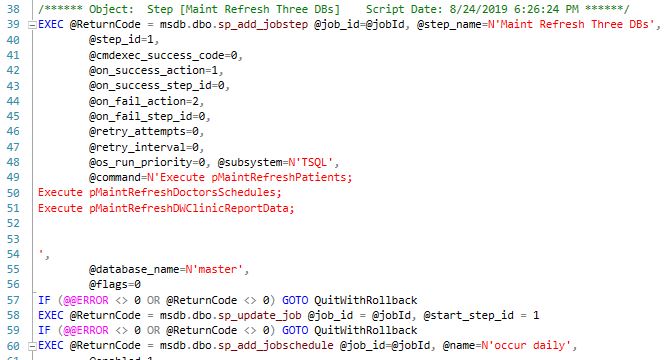


Figure 16

By executing these codes, the data from source databases should be imported to the data warehouse. New Visits data would be added to the FactVisits table.

And then, the whole ETL process is finished. In addition, SSIS package and backup job(Figure 17) is added to this project to make it more applicable.





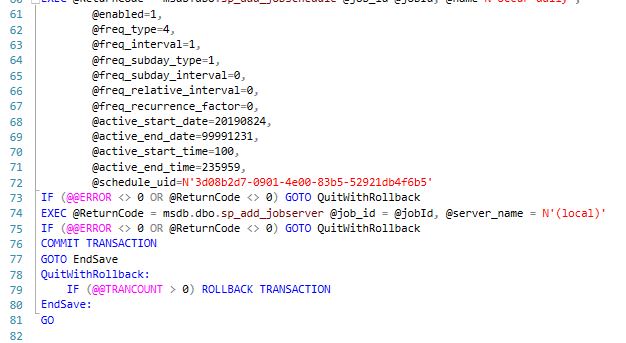


Figure 17

Summary:

In this assignment, we are focusing on data movement and database architecture. Dealing with massive real world data is a new experience and we have to solve related problems which we never faced before. Both flush and fill and incrementally loading methods are used in this project. SSIS package and backup job is added after ETL process.