Sri Lanka Institute of Information Technology

Data Warehousing & Business Intelligence

Assignment 01

Submitted by:

Rathnasooriya P.U (IT20156206)

Contents

1.Data Set Selection.

01.Background: -

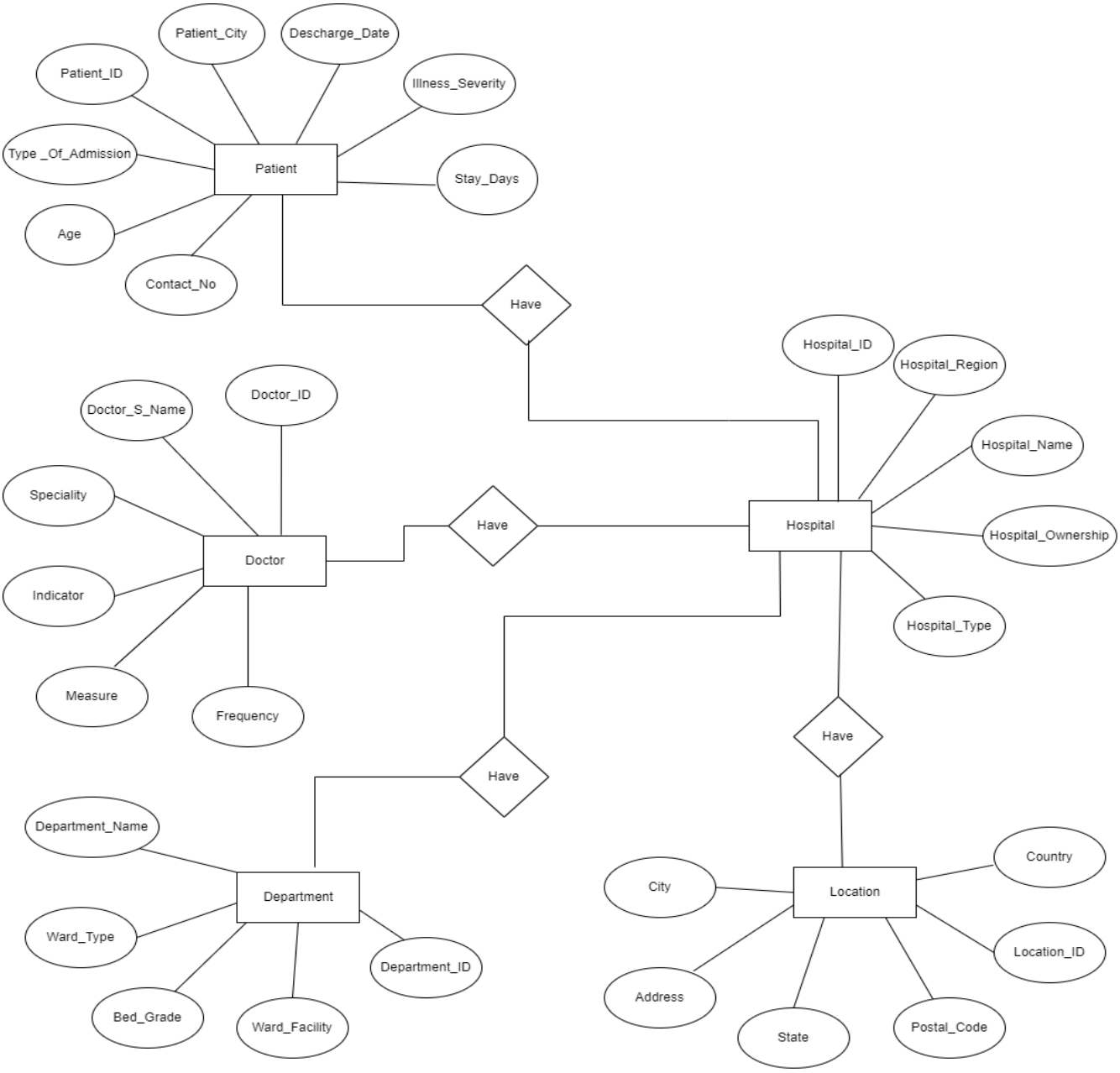
With rapidly growing demand for healthcare in hospitals and intensive care units (ICUs) throughout the world, the COVID-19 epidemic has put unprecedented strain on health systems. As the epidemic spreads, evaluating the related healthcare resource requirements (beds, personnel, and equipment) has become a top issue for many countries. Estimates of how long patients with COVID-19 require various levels of hospital care are needed to forecast future demand.

02.Content: -

The data set was downloaded from -:

COVID-19 Hospitals Treatment Plan | Kaggle

03. ER Diagram



2. Preparation of Data Sources.

In order to data extraction need to prepare the data sources. From my main data source, I have extracted to type of data sources.

1. Text file (.txt)
2. Excel files (.xlsx)

- Hospital Excel Data File.
- Department Excel Data File.
- Doctor Excel Data File.
- Customer Excel Data File.
- Patient Excel File.
- Accm_txn_complete_time Excel File.

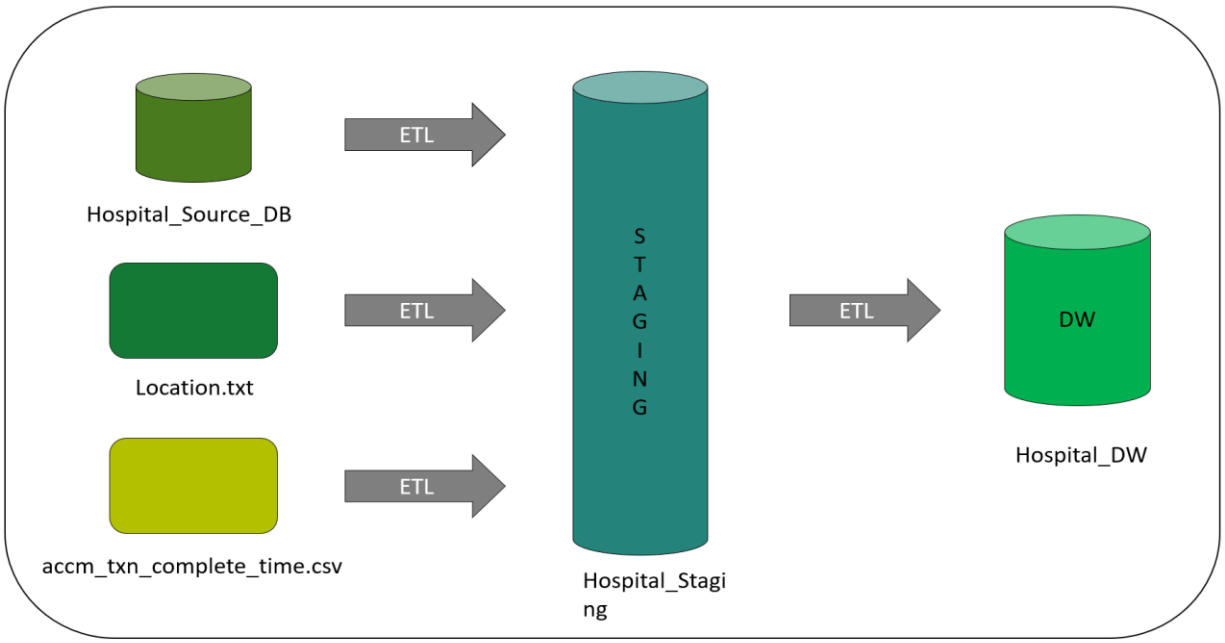
From these dataset, I import above excel files to SQL SERVER MANAGEMENT STUDIO and create the database called **Hospital_sourceDB** Database.

Hospital_sourceDB have following tables -:

- Hospital table.
- Department Table.
- Doctor Table.
- Customer Table.
- Patient Table.

Text file –This text file include all the hospital address details including address, city, state, postal code and the country.

3. Solution Architecture.



Hospital_stagingDB

- Hospital_Staging
- Location_Staging
- Department_Staging
- Doctor_Staging
- Location_Staging
- Patient_Staging
- accm_txn_complete_time_Staging

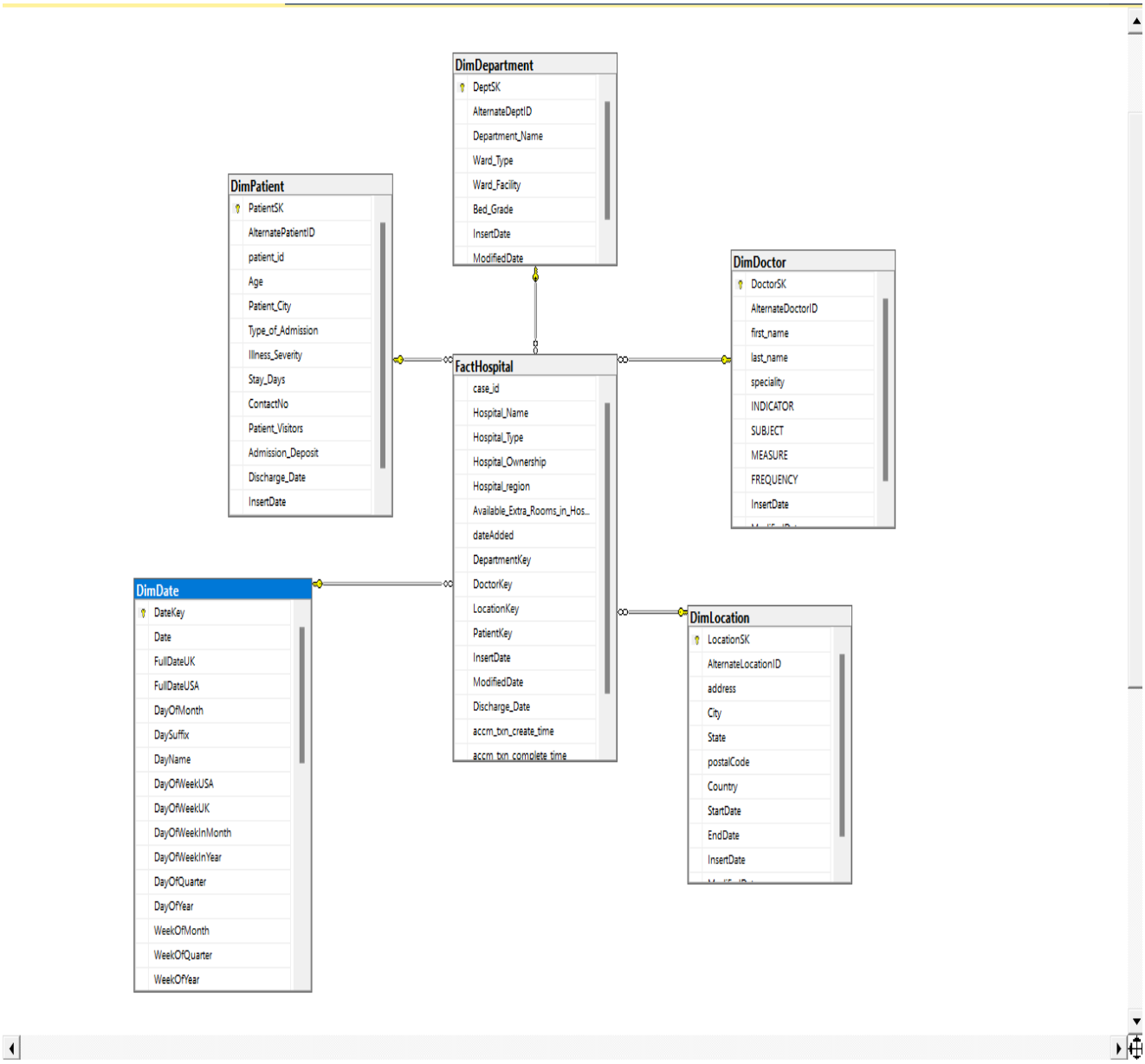
Hospital_DW

- DimLocation
- DimDepartmen

- DimDoctor
- DimLocation
- FactHospital

4. Data warehouse design & development.

Relational Diagram – Star Schema.

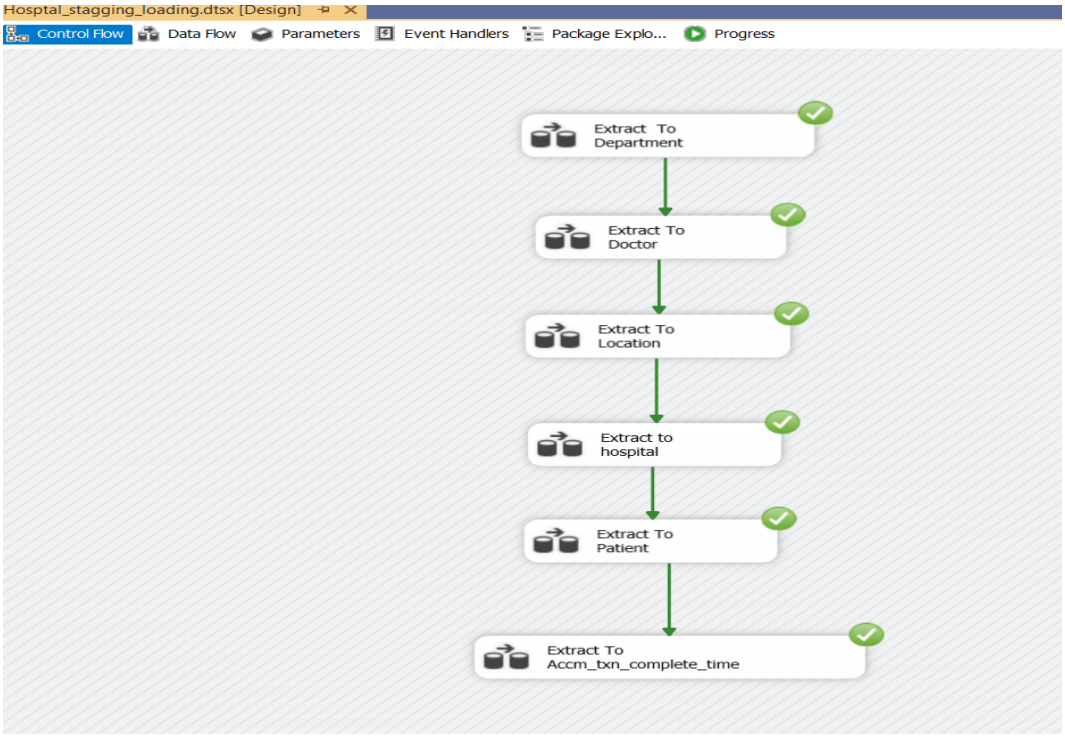


DimLocation is slowly changing dimension. Hospital Address may be changed in future. Therefore, I get it as slowly changing attribute.

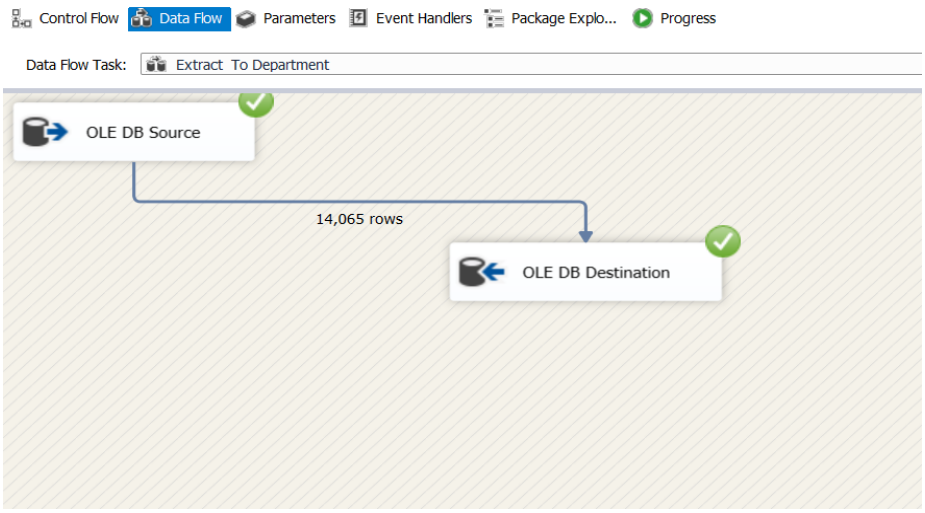
Country-> City-> Postal Code-> State-> Address = This is the Hierarchies.

5. ETL Development

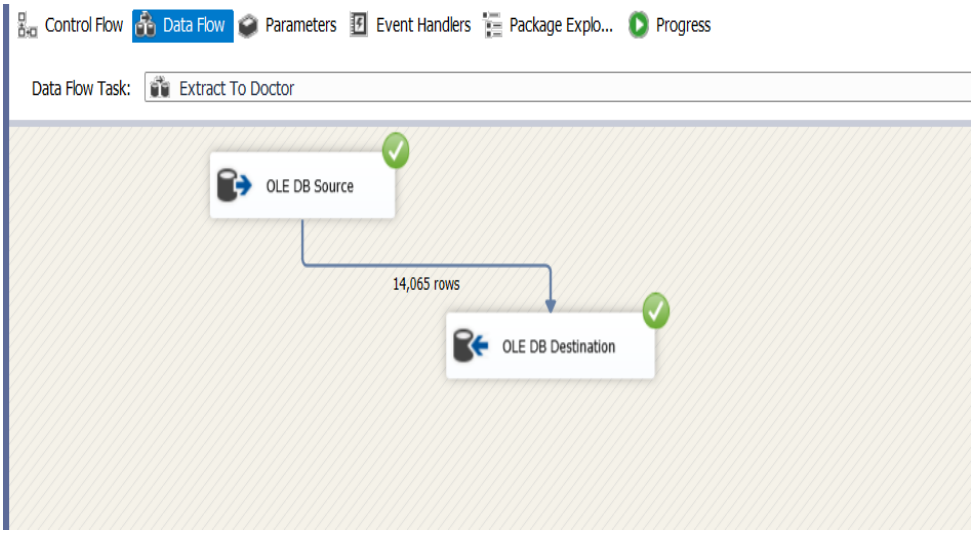
5.1 ETL –Source to Staging



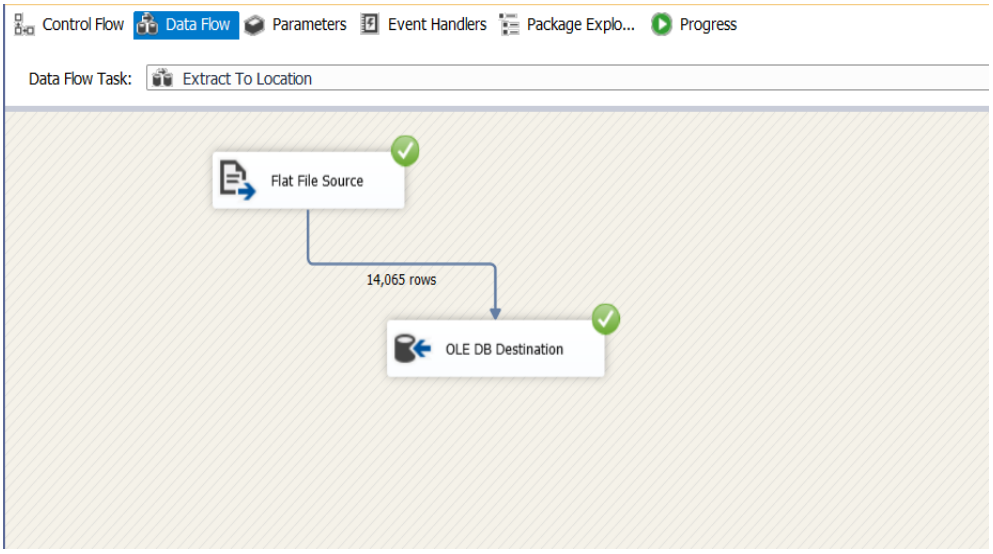
5.1.1 Load Data Department Staging Table



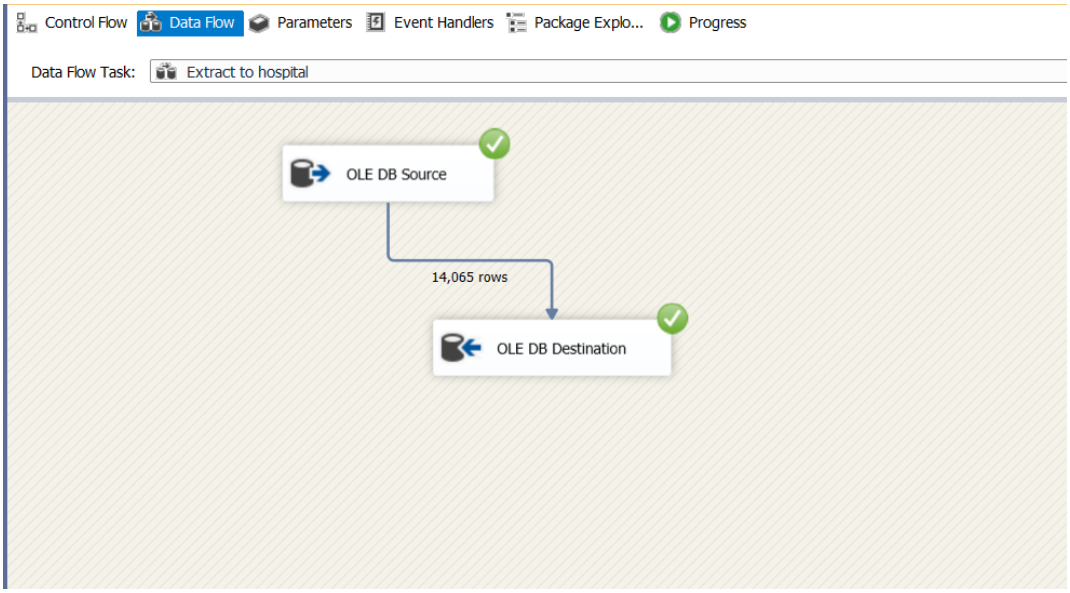
5.1.2 Load Data Doctor Staging Table



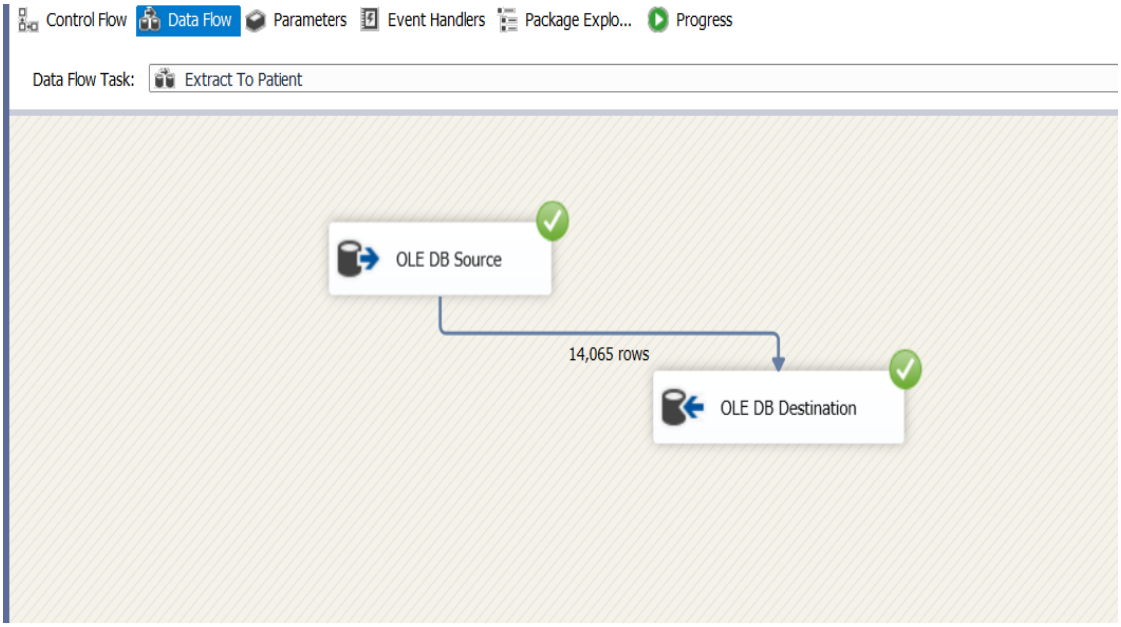
5.1.3 Load Data Location Staging Table



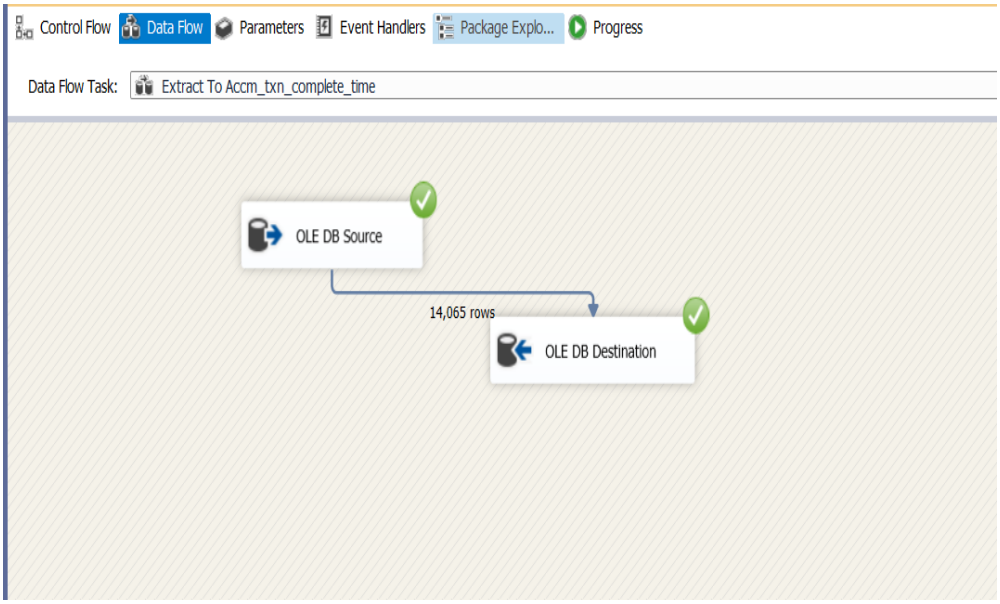
5.1.4 Load Data Hospital Staging Table



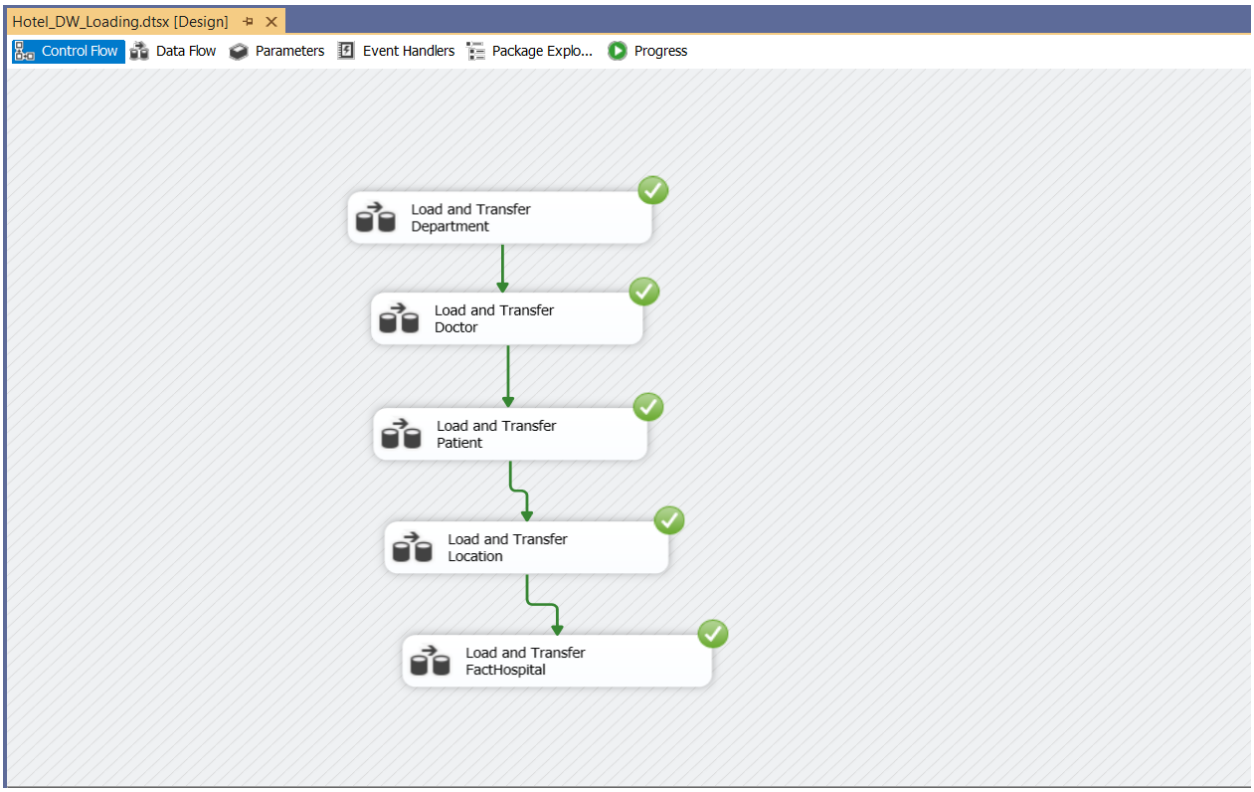
5.1.5 Load Data patient Staging Table



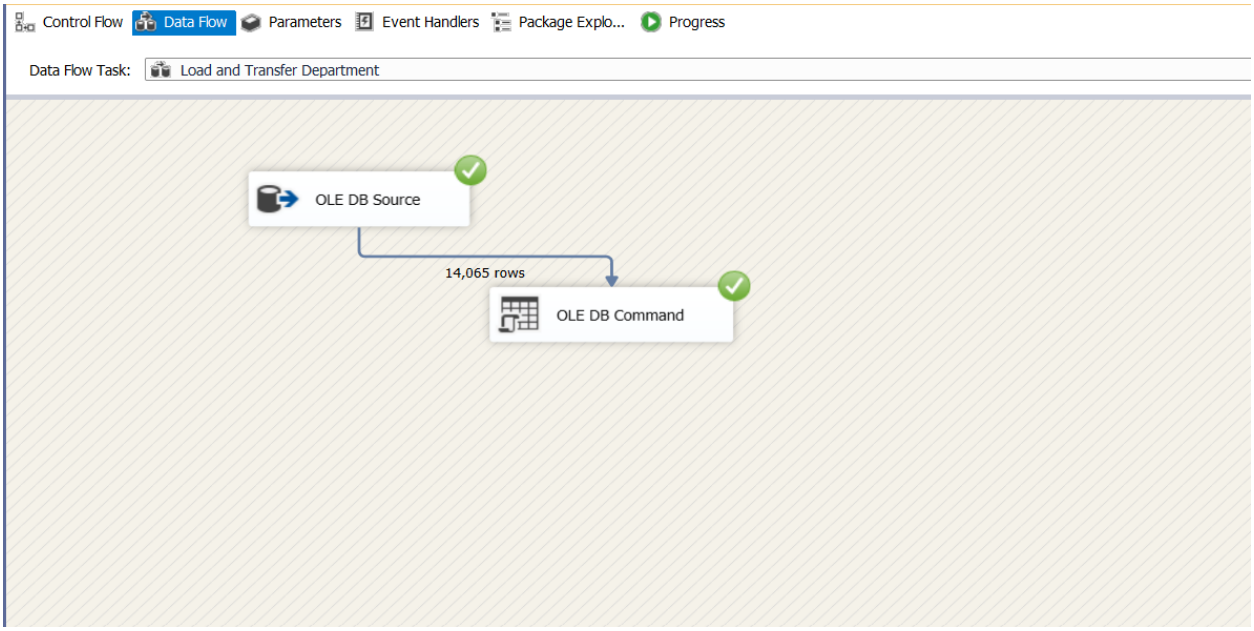
5.1.6 Load Data Branch Staging Table



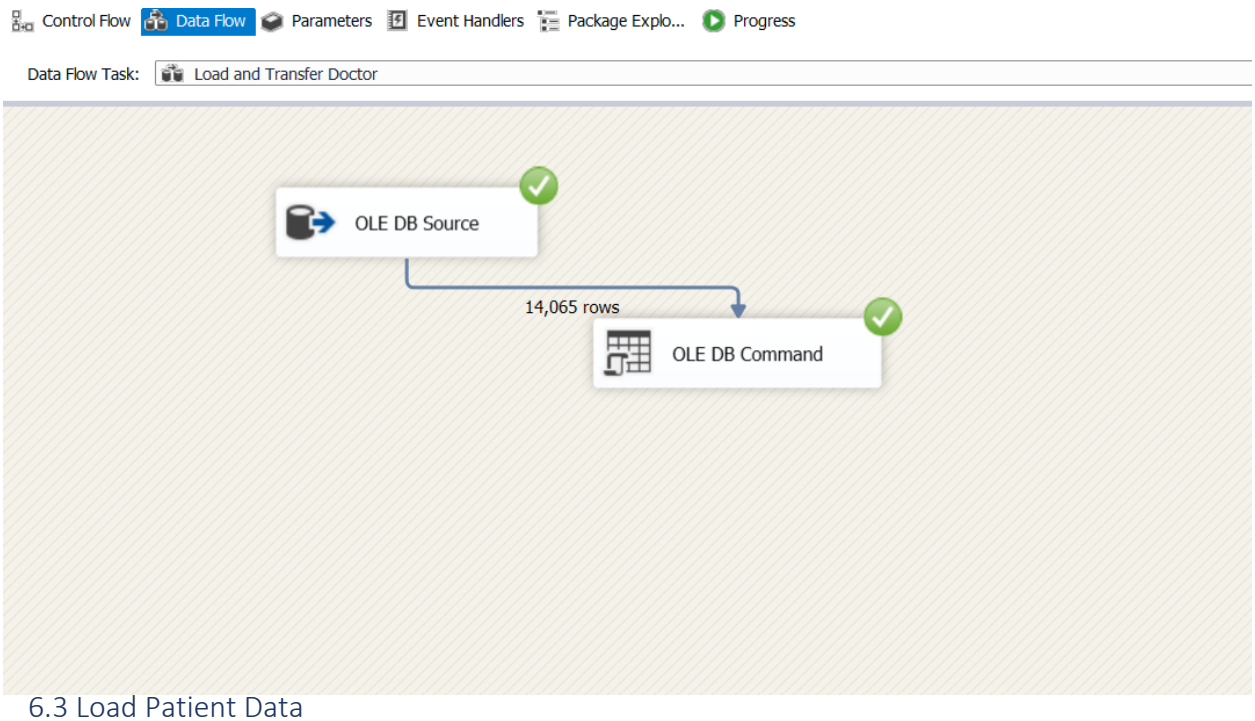
6. Staging To Data Warehouse



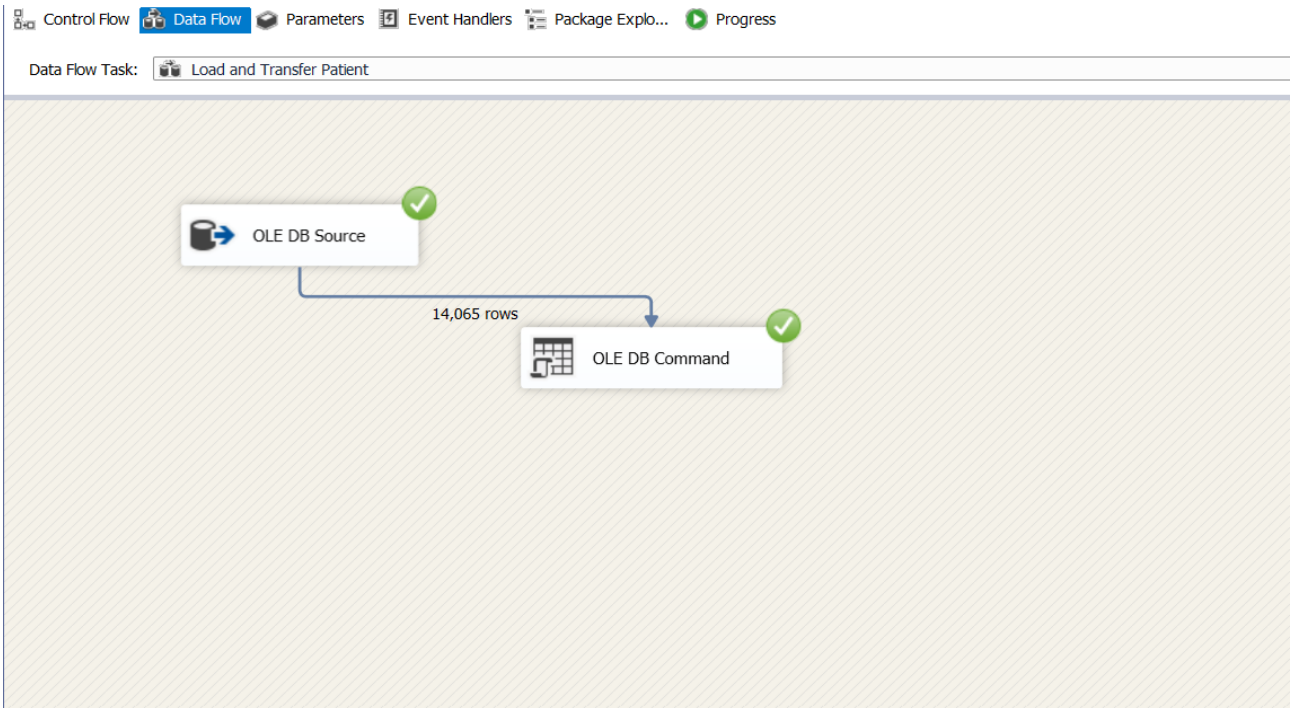
6.1 Load Department Data



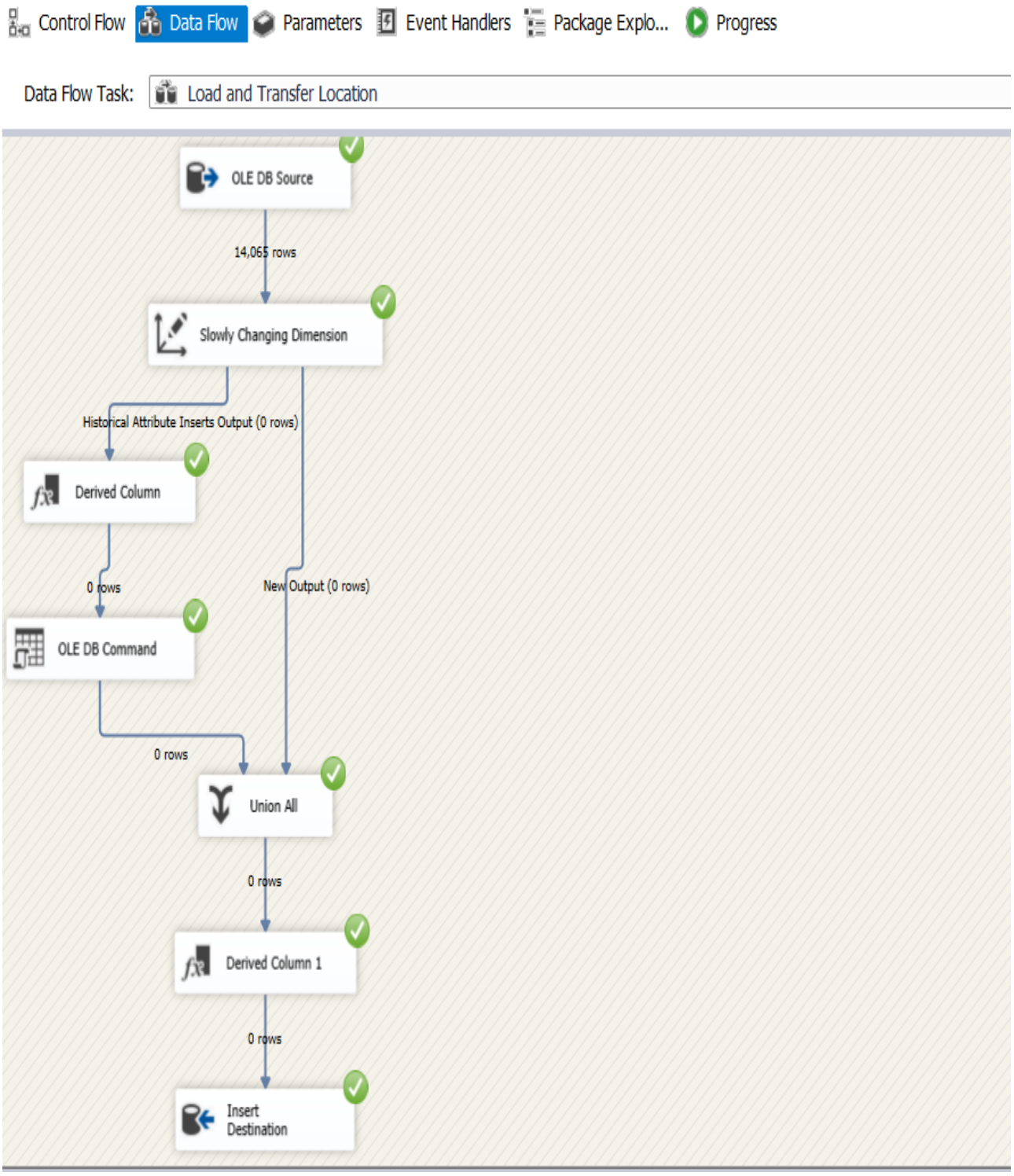
6.2 Load Doctor Data



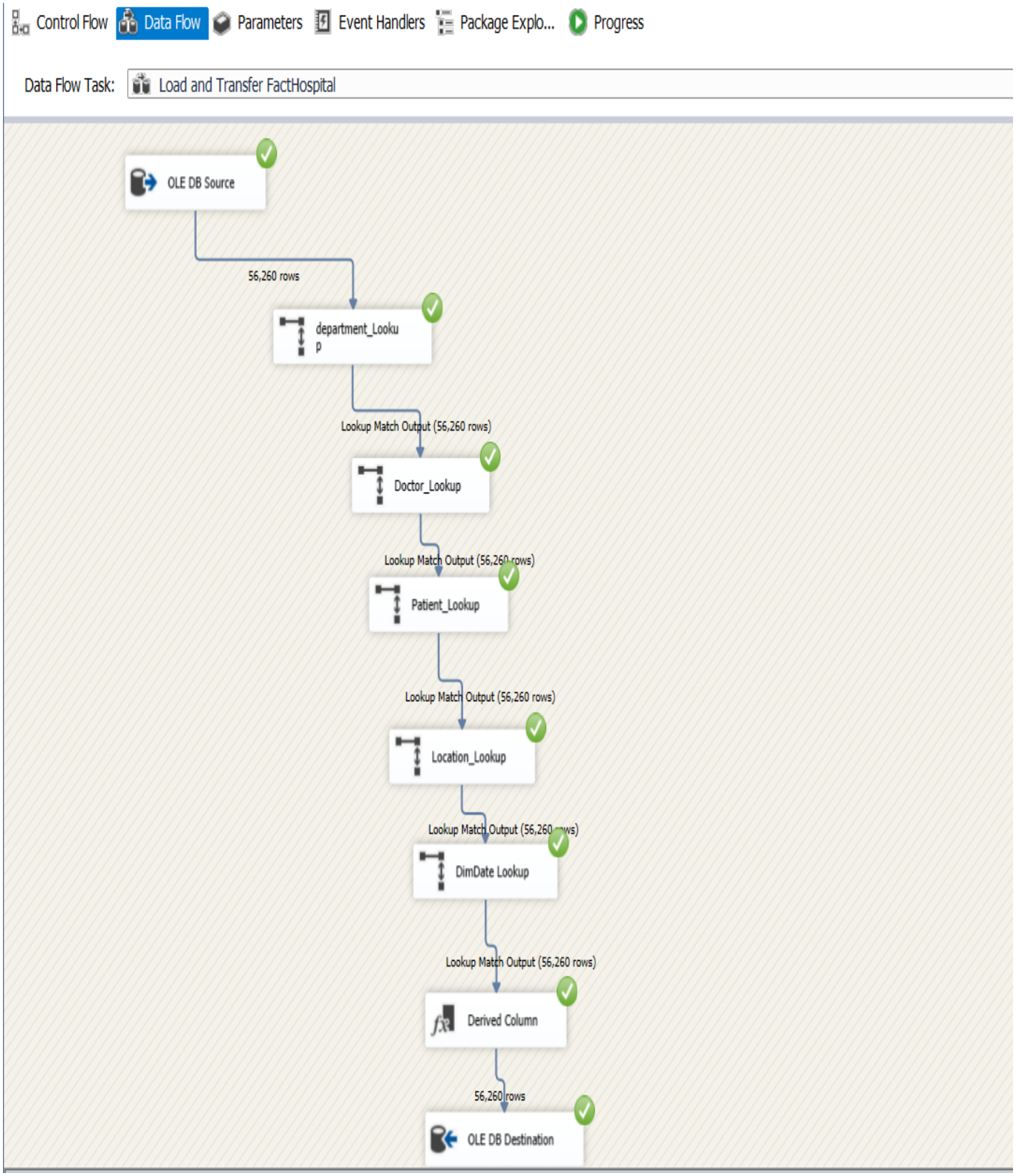
6.3 Load Patient Data



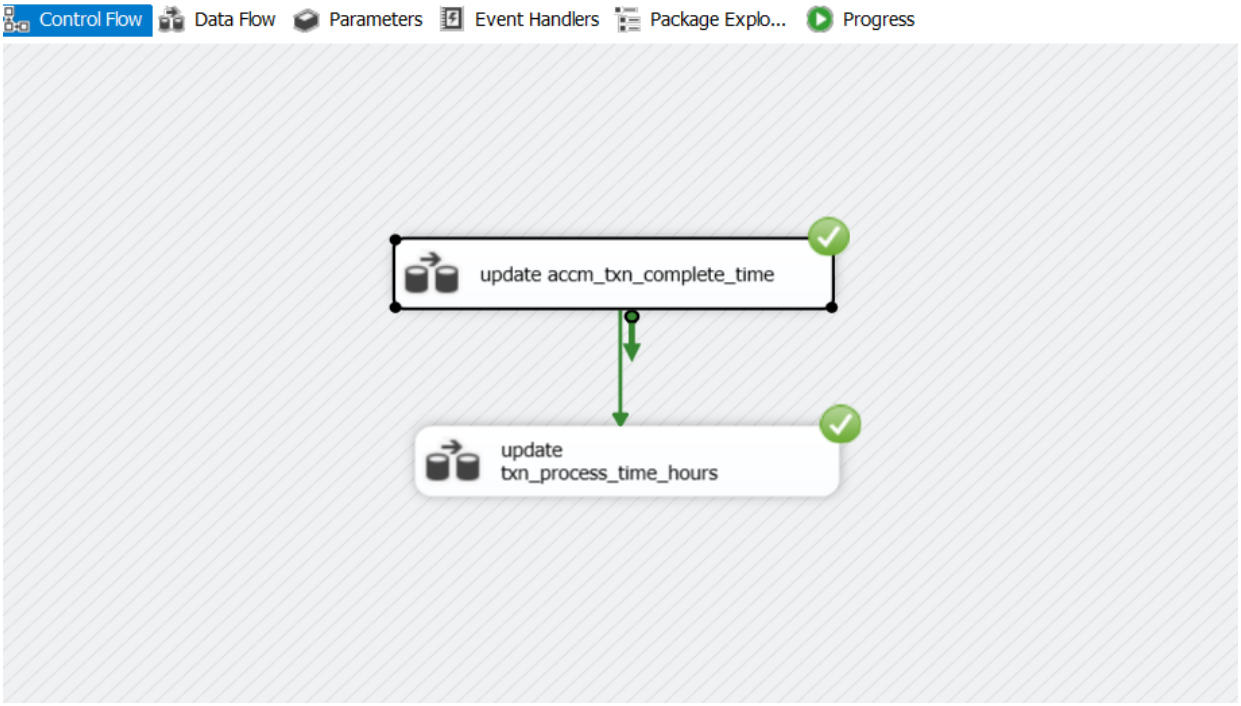
6.4 Load Location Data



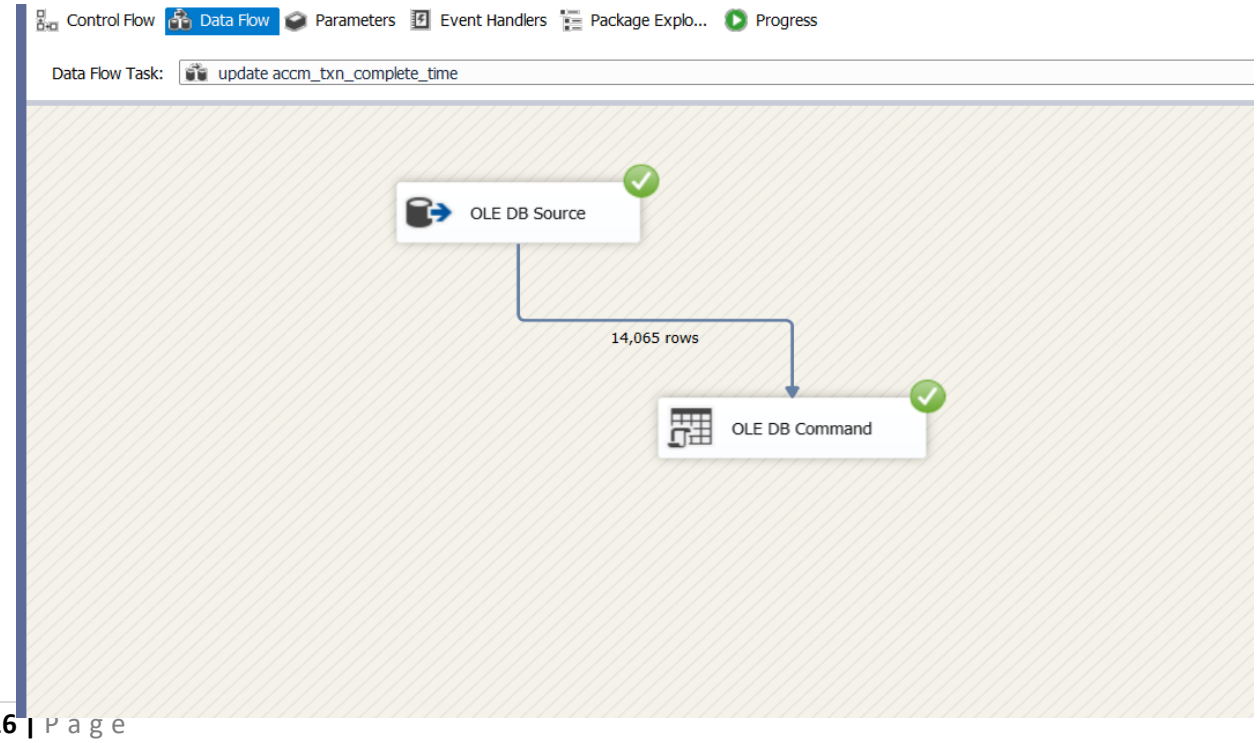
6.5 Load FactHospital Data



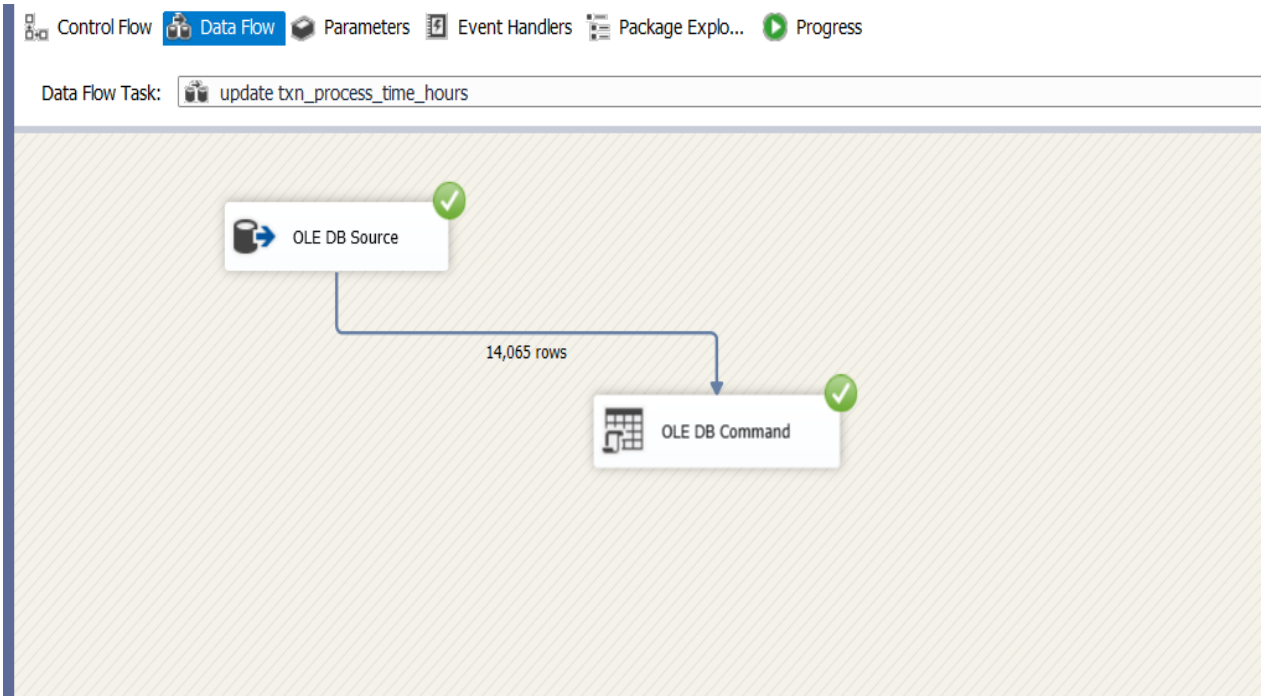
7. Data Warehouse Update SISS Package



7.1 Load accm_txn_complete_time Data



7.2 Load_txn_process_time_houres Data



8. Accumulated Fact Table