**Final Project**

**Deployment and User Guide**

**Course: ITMD – 526 Data Warehousing**

**Project Developers**

1. **Cristina Quiroga Valcarcel**
2. **Joris Ekpangbo**
3. **Manikandan Ganesh**
4. **Saahil Sofat**
5. **Shushmitha Anthay Suthakakaran**

**Objective:**

To prepare a deployment document for carrying out the major steps involved in populating the Fact table (Fact\_application) from the given raw source table (admissions\_info).

**Pre-Requisite:**

* **SQLYog** is used for creating the tables, updating the table from the admissions\_info source table.
* **Pentaho Data Integration Platform** is installed for carrying out the ETL operations.
* **Tableau** is used for creating the reports.

**Environmental Setup:**

Connection Name: datamart\_application

Host Name: localhost

Port: 3306

User Name: ETL\_USER\_NAME in the Kettle Properties

Password: ETL\_USER\_PASS in the Kettle Properties

DIR\_SOURCE\_FILES: Place the local path of the source file location along with “\ETL packages” which contains our transformations.

**Process Flow:**

**1.Load Source File:**

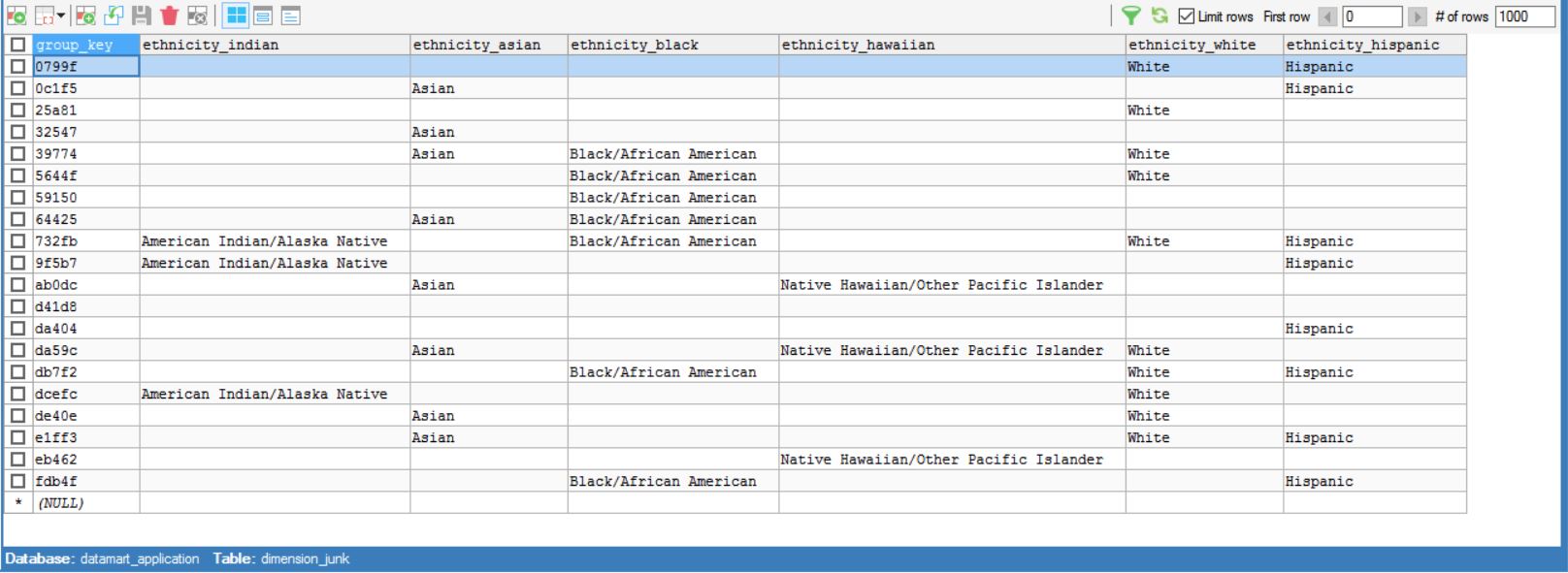
* Make sure that the admissions\_info\_dump has been executed and loaded to the database.
* We see a lot of NULL values for the regions in the data, please run the script for data cleaning. We have attached the cleaning script as well.



**2.Static Dimensions:**

The following dimension tables are made static:

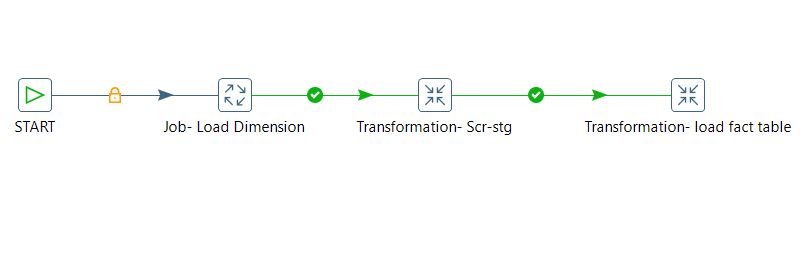
* **Junk Dimension:** It has group\_key (Primary Key) that defines all the distinct possible combinations of ethnicity for an applicant.

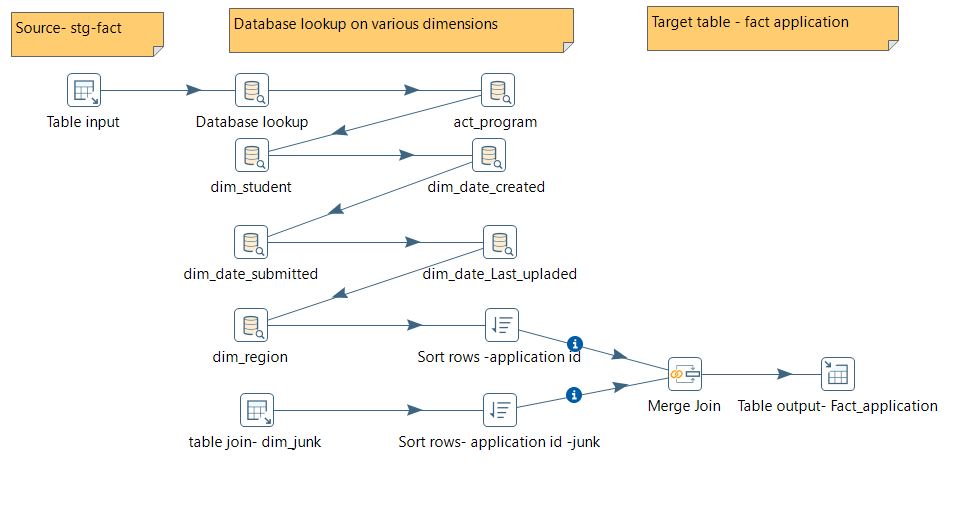


* **Program Action:** It has a program action id that defines the applicant’s application status. It has numeric flags on admission status.



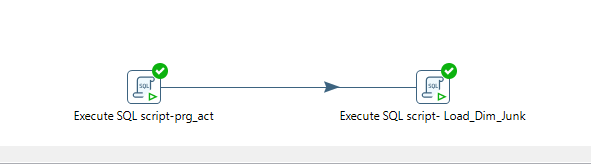
**2.Steps involved in deploying the Jobs:**

* The first step in the deployment process is to run the static transformation which creates the Program Action dimension and Junk Dimension. These dimensions are static as they don’t have any impact in their data with the change in source tables.
* Next run the “Jb\_load\_datamart” job.
* The first job “Jb\_load\_dim\_data” carries the following transformation stages to populate all the dimensions:
  + The first stage is to populate the department dimension.
  + The second stage is to populate the student application dimension.
  + The third stage is to populate the date dimension
  + The fourth stage is to populate the views for created, updated and submitted dates in the application info.
  + Lastly, the transformation is mapped to a job (jb\_dim\_region) to update permanent address and the current address of the applicant into the dim\_region dimension.
* The job is then mapped on to the staging transformation “scr-stg” that prepares the staging table stg\_admissions\_info with the required fields for the final fact table.
  + Take input from the application\_info dump.
  + Select the values we want in our staging area. This can be helpful in case we want to alter or not select some values for the fact table.
  + Lastly, we are just updating all the selected values in the staging table.
* The staging transformation is then mapped on to the load fact table transformation.
* It takes the input from the staging table and undergoes various database lookup to feed the fact table “fact\_application”.
  + We are basically looking up for the various Primary key’s in the dimension tables after mapping them with the “application\_id” from the “admissions\_info” table.

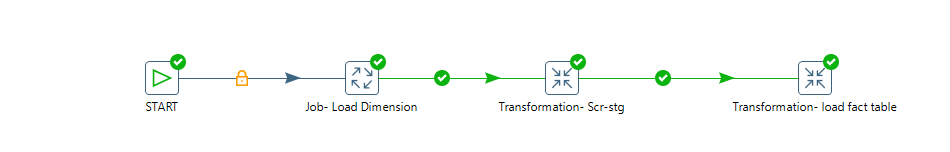


**User Guide to run the transformations and Jobs:**

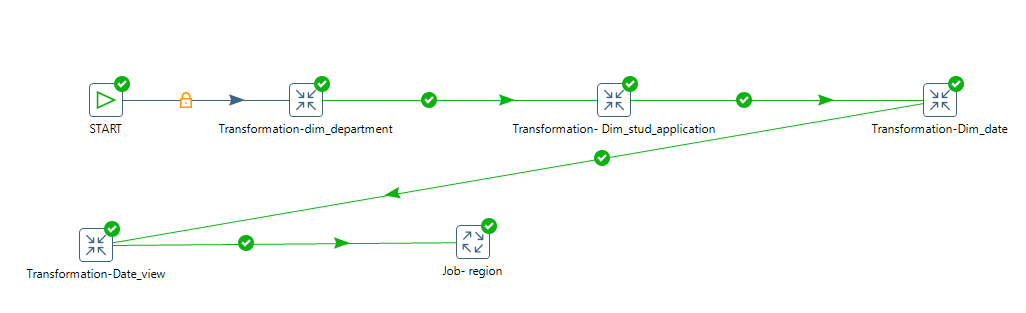
Static Transformation



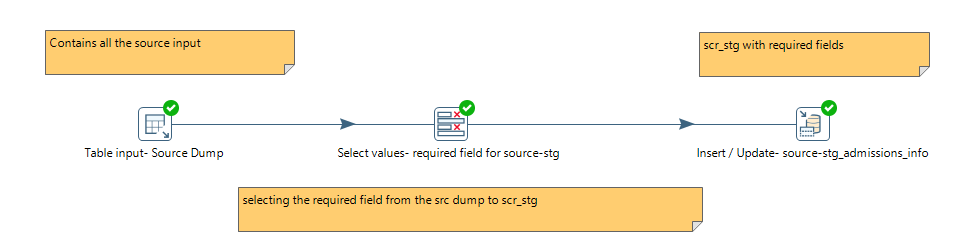
Jb\_load\_datamart



Jb\_load\_dim\_data



Scr\_stg



Fact\_appl

