2017 Spring ITMD 526

Group Project

Overview

You have been provided with a fictitious OLTP dataset that consists of graduate studies applications and admissions information. Your group is to perform a dimensional modeling, ETL implementation, and reporting and analytics.

Dataset

Execute the file **oltp_admission_dump.sql** in your local MySQL DB server. This will create a database called **oltp_admission** with all the tables pre-populated. This will be the source dataset of your project. Create a new database connection called "**OLTP ADMISSION**" in your Kettle using the existing DB account information (ETL_USER_NAME/ETL_USER_PASS) to connect to the dataset. Use this connection for the extract job of your project.

Refer to the **PMD_oltp_admission.pdf** for the physical model diagram of the dataset.

Here are the brief descriptions of each table of the dataset:

Table Name	Description
adm_tag	Tags attached to applications by reviewers.
adm_decision	Admission decision and admited term for each application.
adm_review	Application review scores and status.
adm_term	Admission terms lookup.
adm_program	Programs lookup. Several program codes belong to a department.
adm_department	Departments lookup. Program and department are in hierarchy.
adm_messages	Email communications between applicants and school admissions offices. - **sender_type**: APP (Applicant), SCH (Admission Office) - **reply_to_id:** The **id** of the original request email
app_profile	Graduation school studies applications.
app_college	Previous education history of applicants.
app_recommender	Recommenders' list of applicants.

Deliverables and Points

1. Submission on time **10 points** (7 points deduction for missing the due. One point deduction per day afterward up to 3 points maximum).

2. **High-level technical architecture document** **5 points**
   (Refer to the Week 11 homework)

3. **Source-Target Mapping document** for each dimension **5 points**
   (Refer to the Week 11 homework)

4. **Dimensional modeling** **25 points**
   Provide a **physical model diagram** of your dimensional model (star schema) following the specifications below:
   - **Business Process**: Graduate School Admissions
   - **Grain of Fact**: One row per submitted (completed) application
   - **Dimensions**
     - Minimum 4 of dimensions for ETL implementation. Here are the suggested dimensions:
       - DIM_DATE (Role playing dimension)
       - DIM_APPLICANT
       - DIM_ETHNICITY (Junk dimension)
       - DIM_PROGRAM
     - Minimum 1 bridge table (**No implementation** is required)
   - **Measures**
     - Numeric flags called **IS_APPLIED, IS_ADMITTED**, and **IS_ACCEPTED.** The business logic for each flag is as below:

ADMISSION_DECISION	IS_APPLIED	IS_ADMITTED	IS_ACCEPTED
APPLIED	1	0	0
ADMITTED	1	1	0
DENIED	1	0	0
WITHRAW	1	1	0
MATRICULATED	1	1	1
NO DATA	1	0	0

     - Number of emails sent by each applicant
     - Number of emails replied (**adm_messages.reply_to_id** IS NOT NULL) by school admissions offices

5. **ETL implementation** **30 points**
   (Note: no implementation is required for the bridge table design)
   - Use the same folder structure, schemas, and environmental variables as those from the Individual Project (Week 11 homework version)
   - Compile all the tables created by your group into the **ddl-script.sql**. Make sure you include the schema names (source_db, stage_db, datamart_db) in the DROP TABLE and CREATE TABLE statement so that the scripts can run without having to manually switch schemas
   - Your code is expected to run successfully without an error by taking the steps below:

a. Set the **ETL_HOME_DIR** environmental variable
b. Run the DDL script (**ddl-scripts.sql**)
c. Run the main Kettle job

6. **Reporting and analytics 25 points**
   - Deploy your choice of a BI application on top of your star schema implementation.
     Here are the BI applications suggested (Any other BI reporting product you are familiar with is also welcomed):
     a. Tableau Desktop (Visualization Tool)
     - Download: http://www.tableau.com/academic/students
     - Free Training Videos: http://www.tableau.com/learn

     b. Saiku Analytics CE (OLAP)
     - Download: http://community.meteorite.bi/
     - Basic Demo: http://www.meteorite.bi/training/videos
     - Pentaho Schema Workbench (for meta data management)
       - Download: http://community.pentaho.com/
       - Usage Example Demo: https://www.youtube.com/watch?v=Tqw3oOk5jsM&list=PLIS-R80eiu1snl5wW893-BLiE0yDVhQAe
   - Add three calculated measures in the BI application layer as below:
     a. **Selectivity** = Number of Admitted / Number of Applied
     b. **Yield Ratio** = Number of Accepted / Number of Admitted
     c. **Email Response Rate** = Number of replies by the admissions office / Number of email inquiries
   - Explore the star schema via the BI application, and generate 5 reports

End.