STADVDB – Major Course Output 2: Data Warehousing

Using the same dataset from MCO1, execute the following to proceed with this project:

1. Dimensionality Modeling. Design a dimensional model containing at least one (1) fact table and three (3) dimension tables and following any of the multi-dimensional schema, namely *star*, *snowflake* or *constellation*. The schema can be implemented in MySQL or PostgreSQL.
   Criteria for grading the Schema:
   - Correctness and Completeness
   - Relevance.
   - Complexity.
2. ETL Process. Find an existing tool to extract, transform, and load the database content to the data warehouse (e.g. Rapid Miner, Weka, Python, etc.) Teams may also opt to write their own simple ETL tool to perform this task (your choice of programming language.)

   Write a simple application (using the tool or any programming language) that will allow the user to perform the following OLAP operations on the data warehouse: roll-up, drill-down, dice and slice. Appropriate design of a graphical user interface for the proper display of query results that will aid the user's analytical task is a must. You may reuse the application (probably with simple modifications) in MCO1.

3. Test your dimensional model, ETL processes and queries.

4. Technical Report. The outline is given below.

   I. Introduction. Give a brief description of the purpose of conducting this study. Describe the OLAP application that you built, its intended usage, and target users and beneficiaries.

   II. Dimensional Model. Present and discuss your model. (Use the following as a guide, not as sub-topics of this section.)
   - What are the contents of your fact table?
   - What are your dimensions? Justify your choice of dimensions and facts.
   - Which schema did you use – star, snowflake, or constellation? Why? What issues did you encounter in your model and schema design? How did you address these issues?

   III. ETL Process. Discuss the processes of extraction, transformation and loading. (Use the following as a guide, not as sub-topics of this section.)
   - If you used an existing tool, give a short overview of the tool. Why did you choose this tool?
   - If you developed your own ETL tool, describe your code.
   - For extraction, describe your data sources and their relevance to the model.
   - For transformation, present the rules or functions you applied on the extracted data. What made you decide to apply these rules and functions? What issues did you encounter during transformation? How did you address these?
   - For loading, discuss any additional constraints in the database schema. What issues did you encounter during loading? How did you address these?

   IV. OLAP Queries. Describe the main purpose of your application. What decision-making or analytical task(s) is the application intended for?
   - Give a sample query for each of the following operations - roll-up, drill down, dice and slice.
   - Describe each of the queries. Which dimensions are utilized (rolled up, drilled down, or constrained)?
   - Provide sample output as appropriate or relevant to the discussion.

V. Results and Analysis. Discuss your test process, test data and results.

VI. Conclusion. What are your learnings from this activity?

VII. References**.**

Criteria for grading the Technical Report:

- Correctness and Appropriateness of the Dimensional Model
- Correctness and Completeness of the ETL Process
- Quality of the OLAP queries
- Efficiency and Effectiveness of the methodology (Test Process)
- Clarity of the discussion of results and analysis
- Evidence of Critical Thinking presenting the learnings derived from the conduct of the project (Conclusion), such as understanding the impact of the dimensional model and the ETL process to the types of OLAP queries that can be sufficiently supported
- Overall document presentation, e.g., format (title page, page numbers, sections, tables and figures), references, and language (spelling, choice of words and grammar)

Final Deliverables

The following final deliverables are required to be uploaded to your Notion Workspace (identified in MCO1):

1. Database Schema for the Data Warehouse – to be uploaded in **Docs**
   ⇒ Sample filename: `STADVDB S17.8 MCO2 DB v1.pdf`
2. Source Code for the ETL process and the OLAP application – to be uploaded in **Git**
3. Project Presentation file
   ⇒ Sample filename: `STADVDB S17.8 MCO2 presentation`
4. Technical Report following ACM publication format
   ⇒ Sample filename: `STADVDB S17.8 MCO2 v1.pdf`

NOTE: Do not forget to update the To-Do and Meeting Notes, as in MCO1.