

DATA 225

Database Systems for Analytics

Fall 2023
Section 21
Instructor: Ron Mak

Assignment #9

Assigned: Monday, November 6
Due: Monday, November 13 at 5:30 pm
Team assignment, 150 points max

Dimensional modeling and star schema

The purpose of this assignment is to give your team experience designing and creating a dimensional data model with a star schema.

- Create a dimensional model with a star schema based on the relational schema of your project's operational database.
- Your model should have at least four dimension tables and two fact tables.

Make at least two of your dimensions hierchical. Examples:

- Calendar: century, decade, quarter, year, month, day
- Location: system, university, school, department, class

Then you'll be ready to perform OLAP operations, next week's topic.

- Use ERDPlus to draw your star schema.
- Write SQL and Python code to perform an ETL to populate your dimension and fact tables from your operational tables.
 - For now, your dimensional model can contain data that don't come from your operational tables (i.e., you can generate some artificial data).
- Create at least two queries per fact table that join with the dimension tables.
 - Describe the query in English.
 - Write and execute the SQL.

NOTE: Use your team's new data warehouse whose name ends with _wh on the school server. You should be able to see and access it in MySQL Workbench and Python code.

What to submit

A zip file containing

- Relational schema of your operational database generated by ERD Plus.
- The star schema of your dimensional model generated by ERD Plus.
- A Python Jupyter notebook containing:
 - SQL and Python code that perform an ETL.
 - Descriptions of your fact table queries (as comments or markup).
 - The SQL queries as sent by Python code. Include their output.

Rubric

Criteria	Max points
Relational model	10
• ERD Plus: Relational schema of your operational database	• 10
Dimensional model	80
• Four dimension tables	• 40
• Two fact tables	• 20
• ERD Plus: Star schema of your analytical database	• 20
ETL	20
• Notebook: SQL and Python code	• 20
Queries	40
• Notebook: Two queries per fact table that join with the dimension tables	• 40