Data Warehouse
Second Semester 2024/2025
Homework Assignment # 2
Deadline: 17 May 2025

Question:
Part1:
Use the data in the following link as a starting point for your solution.
You first need to import it into any DBMS (like MySQL).

https://drive.google.com/file/d/1UxR65upm350BBJkEODJQ3h2LQKGrw5-L/view?usp=drive_link

The data represents transactional data for a rental film company.

Design the dimensional model:
1. Main Fact tables
   a. Monthly payment per staff per rent
      i. Dimension tables
         1. Date
         2. Staff
         3. rent
   b. Daily inventory per film per store
      i. Dimension tables
         1. Date
         2. Film
         3. Store
2. Which scheme fits the above dimension tables?
   a. Always consider the **performance** of the query.

Part2:
- Develop an ETL to fill the tables in the designed Dimensional model in part one.
    - The ETL contains the following tasks.
    - Extract the related data from the related tables to fill the first fact table
    - Make sure to handle (even if the data is clean)
        - Redundant data (if it exists)
        - missing values
    - Load the data into the (facts and dimensions tables)
- Make sure to run the ETL with current data, then try to add some data (not cleaned), to make sure the ETL is working well.
- Try to use automation tools like (Airflow) (**Bouns**)