Data Warehouse

Project

The goal of this project is to create a Data Warehouse, to analyse and to process the database using the OLAP.

The project should contains the following parts:

1. **Choose** and **analyse** the dataset and the domain application. Point out the context and the difficulties.
Use the following link: https://lipn.univ-paris13.fr/~grozavu/DW/dataProject/
You are free to choose another database from https://archive.ics.uci.edu/ml/datasets.html

2. **Logical design** of the database (star schema or snowflake schema), you are free to add hierarchies and more aggregates;

3. **Physical creation** of the Data Warehouse;

4. **Insert data** into the DW, you can randomly generate data if needed;

5. **Query** the Data Warehouse using OLAP (CUBE,…) queries in order to obtain in output a rectangular matrix **X** with **n** lines and **p** columns containing in lines the objects and in columns theirs characteristics. These queries depends on the data problem and application field. You can obtain several matrices. Explain the queries and the results.

6. Use **Phyton** in order to load the obtained matrix (matrices) and **visualize** the data using scatter plot. Analyse the results. Apply a **clustering** model using k-means method and visualize the results. Explain the results.

Prepare a report and upload it on the moodle.

The report should contain all these parts, and you should explain the goal of this DW, the motivation of the used queries, and discuss the obtained results.