Data Management & Business Intelligence
– Data Warehouse Project –

Academic Year 2025/2026

1 Objective

Imagine you are applying for a position as a data engineer or data analyst in the Business Intelligence department of a large corporation. As part of the hiring process, you are asked to demonstrate your technical and analytical skills by completing a practical task and presenting your results through a report and a live demo.

The main objective of this project is to design, implement, and query a data warehouse (DW). You may use real-world data from publicly available sources or generate synthetic data using online tools such as `https://generatedata.com`.

The project can be carried out individually or in groups of up to three members. Begin by creating an example database and proceed to design, implement, and populate a prototype data warehouse in PostgreSQL that can answer specific business questions. The DW should accurately represent the chosen application domain. Accordingly, you must carefully select relevant business processes and define meaningful analytical questions that reflect the needs of the company you are "applying" to.

2 Tasks

The project is structured into several tasks, described below.

2.1 Task 1: Domain Analysis and Description

- Select a company and describe its domain and data landscape. Explain the motivation for developing a DW in this context. Collect relevant background information and supporting documentation.
- Identify the business processes you intend to model, the business questions these processes should address, and the key performance indicators (KPIs) that the company would need to compute. Each group member should focus on at least one distinct business process. Discuss the level of granularity that is available and appropriate for the described processes.

2.2 Task 2: Conceptual Design

- Describe the chosen facts, their associated dimensions, and measures. Each fact should include at least four dimensions and one or more measures. Indicate the additivity properties of all measures.
- Create conceptual schemas using the Dimensional Fact Model (DFM). You may use the Indyco tool (students can apply for an educational license at `https://www.indyco.com/start-here/request-an-educational-license/`), any equivalent diagramming tool, or simply draw them by hand. Ensure that your schema clearly identifies the following: facts, dimensions with hierarchies, measures, descriptive attributes, convergence, shared hierarchies, multiple arcs, and optional arcs.

2.3 Task 3: Logical Design

- Develop logical models of your data warehouse using star schemas (or snowflake schemas if more suitable). Include all primary and foreign key relationships, attributes, and estimated cardinalities. Provide reasoning for any non-trivial modeling decisions (e.g., modeling of multiple arcs or recursive hierarchies).
- Choose two important business questions for each fact and write SQL queries that answer them. Provide example table instances (a few rows for each dimension and up to 15 rows for the fact table), execute the queries on these examples, and present the query results.

2.4 Task 4: Physical Design & ETL

- Write an SQL script to create your DW schema, including all fact and dimension tables.
- Populate your DW with data. Document and explain the steps taken for data cleaning, transformation, and loading (ETL). Include any assumptions or simplifications made during this process.

2.5 Task 5: Querying

- Write and execute SQL queries demonstrating the use of the following operators: `ROLLUP`, `CUBE`, and `GROUPING SETS`.
- Implement advanced analytical queries that include:
  - One ranking query using `NTILE`, `RANK`, or `DENSE_RANK`.
  - One windowing query using the `WINDOW` clause.
  - One period-to-period comparison query (e.g., comparing weekly or monthly sales for the current year with the same period of the previous year).

For each query, include the natural-language question, the SQL code, and the corresponding results.

2.6 Task 6: Data Analysis Tool

- Use a data visualization or business intelligence tool (e.g., Pentaho: `http://community.pentaho.com/`, or Grafana: `https://grafana.com`) to create an interactive dashboard that illustrates key results from your previous analyses.
- Alternatively, you may manually design a sample dashboard using tools such as Excel, gnuplot (`http://gnuplot.sourceforge.net`), or similar graphing software.

3 Deliverables

- A comprehensive project report of approximately 15 pages, suitable for both technical and non-technical audiences.
- All scripts required to create, populate, and query your data warehouse.
- A presentation (optional) and live demo aimed at both technical and non-technical stakeholders.

4 Project Evaluation

Project evaluation will be based on the correctness, complexity, completeness, and clarity of your implementation, report, and demo. Specific assessment criteria include:

- Complexity and soundness of your data warehouse design.
- Reasoning, expressiveness, and completeness of your dimensional modeling. (Note: not all attributes need to be implemented in the final system.)
- Quality and clarity of your justifications and arguments in both the report and presentation.
- Effectiveness of query design, including motivation, correctness, and visualization of results.

5 Learning Objectives

Upon successful completion of this project, students will be able to:

- Analyze business domains and identify relevant processes and KPIs for data warehousing.

- Design conceptual, logical, and physical data warehouse models.
- Design ETL processes for data integration and cleaning.
- Formulate analytical SQL queries using advanced operators and window functions.
- Develop dashboards to communicate business insights effectively.