Data Warehouses – Project

The process of creating a data warehouse should be preceded by understanding the "business needs" and the reality (problem domain) represented by available data resources. The implementation of the task below is to make the problems occurring in a specific (selected) fragment of reality visible, and then enable the identification (definition) of the needs, purpose and possibilities of business analyses to support decision-making processes (making the right business decisions).

The final project should contain at least one Analysis Services cube related to **the data from at least two data sources** (one data set is to be selected from those presented during laboratory classes, and the other – freely chosen by the student).

The created cube should:

- contain **at least 5 dimensions,** including at least two with a hierarchical structure (e.g. time, place, etc.)
- have **at least 3 measures,** including at least one non-additive one
- the corresponding fact table should have **at least 10,000 records.**

Stage I – 19/05/2025

Topic suggestion

Please prepare the scope of the project according to the specification below and discuss the project proposal with the person conducting the classes. Register the agreements made in the form of applications.

Scope of HD project development

1. Project Title
2. Characteristics of the problem domain, short description of the analysis area, problems and needs
3. The purpose of the project (expectations) and the scope of the analysis – aspects examined
4. Data sources (location, format, availability), preliminary analysis of data sources

No.	File Type	Number of records	Size [MB]	1.	Description
2.					
…					

5. Data profiling (analysis of data quality and its usefulness in the project)

No.	Attribute	Data type	Value range	1.	Notes - Data Quality Assessment
2.					

6. Definition of entity/class types (including properties) and relationships between them, class diagram (proposal of dimensions, hierarchies, additive and non-additive measures)
7. **At least 10 multidimensional reports that will be created after the cube is implemented**
8. Implementation of the database according to the proposed conceptual data model

Conclusions:

Stage II – 26/05/2025

ETL process

1. Populate the tables created in the previous step with data in accordance with the established design assumptions using SQL queries, available tools in Integration Services or other available tools.

The following elements of the package(s) will be taken into account in the assessment:

- proper structure of the ETL process – proper division of the ETL process into tasks/packages, well-chosen names of individual tasks
- **complexity of the operations performed** – for example, if the source data is already fully denormalized, please do not expect the maximum number of points for this element
- **introduced automation, e.g. incremental feeding of the warehouse, automation of data cleaning (fuzzy methods), error handling,** etc.
- stability and correct, flawless execution
- the documentation should include a short description of each task, which will allow you to understand what its purpose is (e.g. task Z copies data from tables X and Y to table T, performing denormalization) and a logical map of the ETL process.

Conclusions:

Stage III – 02.06.2025

Ankle:

1. Prepare a cube project, edit dimensions, add calculated measures. Prepare reports from stage I (p. 7) and show other interesting relationships in the analyzed data (depth analysis, not just pivot tables).

The following cube elements will be taken into account in the evaluation:

- correct cube structure – the cube model should allow the analyst to use the data intuitively and easily
- calculated measures, KPI
- documentation which should include a brief description of all dimensions, all their attributes and all measures
- conclusions from data analysis

Conclusions:

Stage IV – 16/06/2025

Presentation

The presentation should contain 4-8 slides (last about 8 minutes) and explain what data is being analyzed. The presentation should end with a short demonstration that shows the most interesting relationships between the data in the cube.

Note. The project will be finally passed after submitting a written report containing descriptions of the individual stages of work.