

DBD Exam Project, Cphbusiness SOFT Spring 2025

Problem Statement

The modern enterprises and public IT systems rely on high-performance scalable applications processing large volumes of frequently collected data from variety of sources in support of their daily business operations and strategic decisions.

General-purpose data storage and single type databases are not sufficient for meeting the requirements and expectations of the business anymore.

The solutions come with [polyglot persistence](#) applications and [purpose-built](#) decoupled data stores, where different system components can be mapped to different data sources, rather than to a single database.

Operating multiple databases within the same application enables solving variety of problems related to performance optimization and usability of both the structured and unstructured data of a company.

Project Objective

The objective is providing database support to an application, where multiple diverse database models are built and implemented for different purposes.

Project Tasks and Requirements

- **Select a business or organization**, which operates with [large amounts of various data](#) and information extracted from it. Such examples are retail companies, food delivery, car rentals, books borrowing, IT training, travel related services, medical consultations, or any other business of your choice, where databases have essential role.
- **Define the application domain**, categories of users and use case, formulate functional and non-functional requirements to your project. The business and implementation scenarios must include various information retrieving [database operations and queries](#), not only database maintenance. Examples of such operations can be searching, filtering and aggregation, data quality control and session management, logging and log analysis, customer segmentation or product recommendation, pattern recognition and fraud detection, data transformation and information visualization, performance measurement and user experience optimization.
- **Select and motivate the selection** of appropriate database models that could support the requirements and the operability of the application. Consider using [minimum two different database types](#) from those introduced in class. As the databases operate in the same business domain, transformation and transition of data between the various database models is expected.
- **Design, create and populate the databases** with large amount of [real data collected from public Internet sources](#). The quantity of data should be sufficient to illustrate the database transactions management in a meaningful way. Divide the responsibilities of data operations between the databases and the application. Consider controlling the data quality, as well as validation and testing of your design and development process.

- **Deploy the databases** on `localhost` and enable accessing them from demo applications. Consider **deploying at least one of the databases in a cluster**, to ensure and demonstrate performance tuning and optimization, ACID or CAP qualities.
- **Create one or more simple client applications** and **implementation scenarios** to illustrate the communication of both end-users and database administrators with the specified database services and operations. No fancy graphical interfaces are expected. Developer-supporting tools, such as browsers, dashboards, testing applications like `Postman`, `curl`, or online platforms can be used as alternative. Keep in mind the data integration, though.

Apply the best of your analytical, programming, networking, and visualizing skills, but remember that the *database issues* are in focus at this project and exam.

Project Work and Delivery

It is a **group project**. The recommended size of a group is 3 students.

Hand-in a synopsis 3-5 pages plus the project solution artefacts, such as diagrams, scripts, the application code, and the data sources have to be collected in a **GitHub repository** and referred to in the Wiseflow submission according the exam schedule.

The application and the demonstration cases must be **reproducible**.

Exam

The **exam is individual** and censored, based on the project and the subject curriculum. The common exam rules apply.