Graduation Project Proposal

Project Title:

Design and Implementation of a Data Warehouse Using the Inmon Approach

Student Information:

Name: Melike Ay

Department: Software Development

Project Advisor: Ali Okatan

University: Istanbul Aydın University

1. Project Description

The goal of this project is to design and implement a data warehouse following the Inmon top-down approach. The warehouse will integrate data from a large-scale dataset (over one million rows) obtained from Kaggle in CSV format. The project aims to apply data warehousing principles, ETL (Extract, Transform, Load) processes, and business intelligence practices to create a structured and reliable data environment for analysis and reporting.

2. Project Scope and Methodology

Step 1: Data Collection and Profiling

- Obtain a large dataset from Kaggle with over a million rows.
- Perform data profiling to assess data quality, completeness, and structure.

Step 2: ETL Process (Extract, Transform, Load)

- Implement data cleansing and transformation logic using SQL and Python.
- Load processed data into the warehouse through multiple layers (staging, integration, and access layers) based on the Inmon architecture.

Step 3: Data Warehouse Design

- Design the enterprise data warehouse schema using normalization techniques.
- Optimize performance through indexing, partitioning, and data modeling best practices.

Step 4: Data Visualization and Reporting

- Connect the data warehouse to Power BI.
- Create interactive dashboards and business reports to visualize trends and key performance indicators (KPIs).

Step 5: Testing and Validation

- Conduct data integrity, accuracy, and performance tests.
- Validate that dashboards and reports reflect correct and up-to-date information.

3. Expected Outcomes

- A fully operational data warehouse built using the Inmon methodology.
- An automated ETL pipeline ensuring clean and consistent data flow.
- Power BI dashboards and reports providing actionable business insights.
- Documented test results and performance evaluations.

4. Technologies and Tools

- Database: PostgreSQL, DBeaver
- ETL: SQL, Python (Pandas, NumPy, Pytest)
- Visualization: Power BI
- Data Source: Kaggle (CSV dataset with >1M records)