COMPUTER SCIENCE AND DATA ANALYTICS

The comparative study of indexing techniques in different database systems

Report 3

Student: Sokrat Bashirov

GWID: G26315644.

Introduction

This report provides an update on the ongoing research project focused on the comparative study of indexing techniques in MySQL and PostgreSQL database systems. The project aims to evaluate the impact of indexing on query performance using a real-world dataset. This report highlights the measurement strategies, statistical analysis, and data visualization techniques that are going to be employed.

Measurement Strategies

Query Execution Time Measurement: I measure the time taken by each query to execute in both MySQL and PostgreSQL databases with and without any indexes. EXPLAIN command is going to be used to analyze query execution plans and capture the execution time.

Memory Usage Measurement: Memory consumption during query execution is another critical metric. I am going to use database-specific monitoring tools and system utilities to record memory usage for each query in both database systems.

Statistical Analysis

Comparing Query Execution Time: I apply statistical methods such as mean, median, and variation to compare the query execution times for each query on MySQL and PostgreSQL databases with and without indexes. I am going to assess if there are statistically significant differences in execution times.

Analyzing Memory Usage: For each query, I compute the average memory usage on both database systems. Statistical analysis is going to be used to determine if there are notable disparities in memory consumption between the two systems.

Data Visualization

Histograms: Histograms are presented to visualize the distribution of query execution times and memory usage for each query on both databases with and without indexes. The histograms provide an intuitive overview of performance differences.

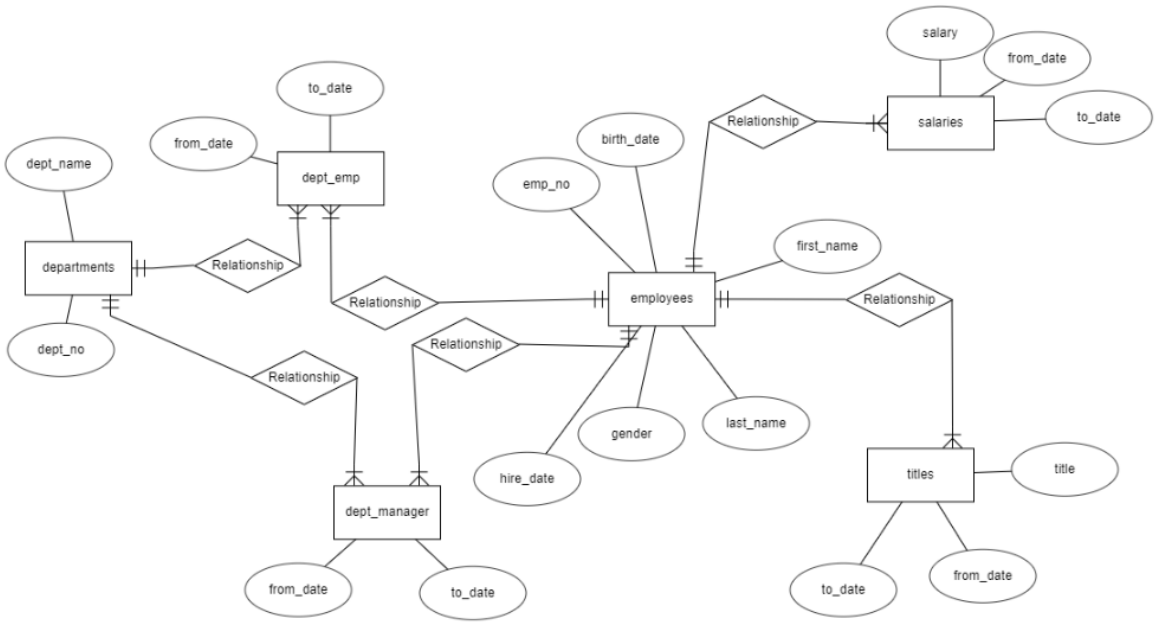

Figure1. Entity-Relationship Diagram (ERD) of the database used in the project.

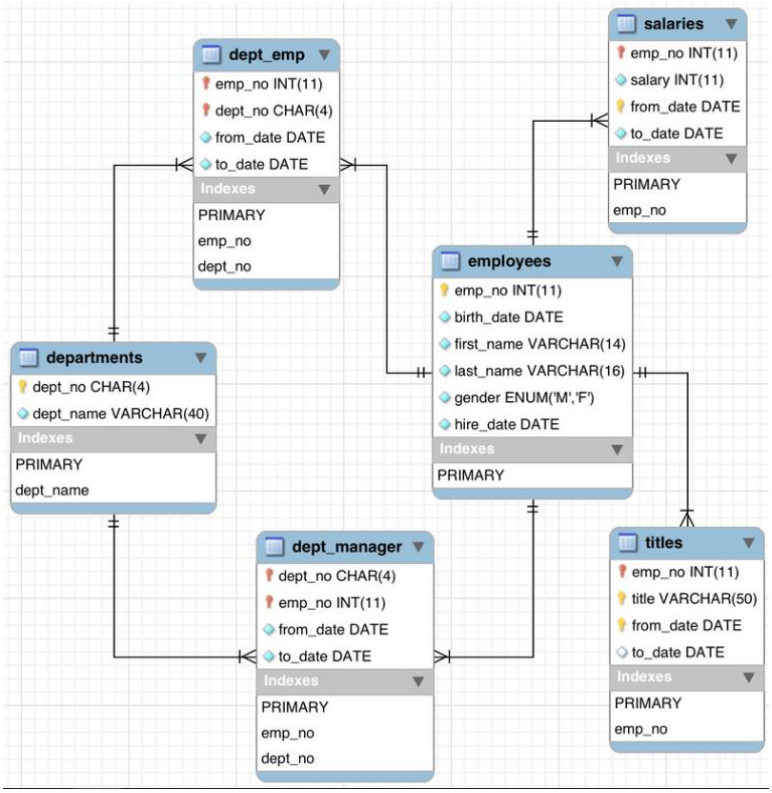

Figure2. Relational Schema of the database used in the project.