COMPUTER SCIENCE AND DATA ANALYTICS

The comparative study of indexing techniques in different database systems

Report 2

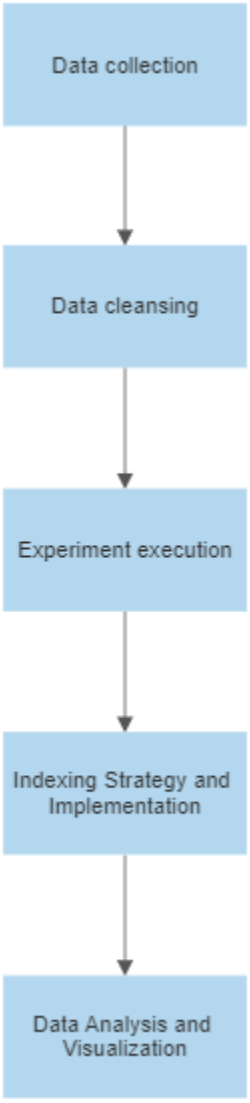
Student: Sokrat Bashirov

GWID: G26315644.

Research Strategy:

This research follows a quantitative approach, emphasizing empirical data collection and analysis. I will design controlled experiments by executing a carefully selected set of queries against both MySQL and PostgreSQL databases. The experiments will be performed multiple times to gather statistically significant metrics.

Research plan:



Data Collection: I will use a real-world dataset, the "employees" database, which contains multiple tables with millions of rows. The database has been imported into both MySQL and PostgreSQL, and all indexes, primary keys, and foreign keys have been removed to ensure a level playing field.

Data Cleansing Approaches: To ensure data consistency and quality, I will employ data cleansing approaches before conducting the experiments. Any necessary preprocessing steps will be documented.

Experiment Execution: I will design five representative queries that cover a range of typical use cases. Each query will be executed 1000 times against both MySQL and PostgreSQL databases without any indexes. Metrics, including query execution time and memory usage, will be collected for analysis.

Indexing Strategy and Implementation: After the initial experiments, I will add indexes one by one to the tables in both databases. Each time a new index is added, the queries will be re-executed, and performance metrics will be recorded.

Data Analysis and Visualization: Collected data will be analyzed, and performance metrics will be compared to assess the impact of indexing on query execution time and memory usage. Visualizations, such as histograms, will be used to present the results effectively.

Database Information

The "employees" database comprises six tables: salaries, titles, employees, dept_emp, departments, and dept_manager. The dataset contains millions of rows, providing a realistic environment to assess the indexing techniques' effectiveness.