Project Overview

1. Project Name: Relational Database System Implementation
2. Project Team: Nav Sanya Anand (Solo Project)
    a. Undergraduate Major: Computer Science
    b. Graduate Major: Computer Science - Data Science
    c. Skills: Databases: MySQL, Programming Languages: Java, Python, C++, Source and Version Control: Git, GitHub, Object-Oriented Programming (OOP), Data Analysis, Software Development

Project Description

The goal of this project is to design and implement a custom relational database system, complete with a unique query language and interactive command-line interface. The database system will enable users to define and manage structured data, perform various operations on the data, and support data insertion, deletion, and modification.

Project Objectives

1. Database System Design
    a. Status: *Completed*
    b. Description: The database system has been designed to meet the project requirements. The data storage and retrieval mechanisms have been planned.
2. Query Language
    a. Status: *In Progress*
    b. Description: The custom query language is under development, allowing users to perform common database operations like projection, filtering, joins, grouping, aggregation, and ordering.
3. Data Storage and Retrieval
    a. Status: *In Progress*
    b. Description: Data storage and retrieval mechanisms are being implemented, with a focus on data structures like B-trees for indexing.
4. Interactive CLI Interface
    a. Status: *In Progress*

    b. Description: An interactive command-line interface (CLI) is being developed to enable users to interact with the database system through commands and queries.
5. Data Modification Commands
    a. Status: *Not Started*
    b. Description: Commands for inserting, deleting, and updating data in the database will be implemented.
6. Real-World Dataset Integration
    a. Status: *Not Started*
    b. Description: The IMDB Top 250 Movies Dataset will be integrated into the database system to demonstrate its functionality.
7. Documentation
    a. Status: *Not Started*
    b. Description: Comprehensive documentation will be provided, explaining how to use the database system, including the query language, data modification commands, and real-world dataset integration.

Timeline

The project is currently on track, and the following timeline outlines the progress and planned activities:

Phase 1: Design and Planning [25 Sept - 29 Sept]
Define the data model (Completed).
Plan the database file storage mechanism (Completed).
Design the query language and CLI interface (In Progress).

Phase 2: Implementation [2 Oct - 18 Oct]
Develop the database system core (In Progress).
Implement the query language parser and executor (In Progress).
Create the interactive CLI (In Progress).

Phase 3: Testing and Debugging [8 Nov - 14 Nov]
Unit testing and debugging (Not Started).
Optimize data storage and retrieval mechanisms (Not Started).

Phase 4: Demonstration and Finalization [15 Nov - 20 Nov]

Showcase the db's functionality using a real-world dataset (Not Started).
Prepare for presentation (Not Started).

Phase 5: Documentation [29 Nov - 7 Dec]

Create user and technical documentation (Not Started).
Prepare a demonstration script (Not Started).

Issues Faced/Facing

The development of the custom query language is currently posing challenges in terms of parsing and execution. This is an ongoing issue that requires careful consideration and testing.
Ensuring efficient data storage and retrieval mechanisms is another challenge that will be addressed in the coming phases.
The integration of a real-world dataset, while exciting, may introduce complexities that need to be carefully managed.
Overall, the project is on track, and efforts are being made to address these challenges and deliver a functional relational database system.

Conclusion

The project is progressing as planned, with the design and initial development stages completed. The database system's core is being implemented, and work on the query language and interactive CLI is in progress. As the project advances, the remaining objectives, including data modification commands, dataset integration, and documentation, will be addressed.