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: My SQL, Programming Languages: Java, Python, C++, Source and Vision Control: Git, GitHub, Object-Oriented Programming (OOP), Data Analysis, Software Development

Project Description

This project aims to design and implement a custom relational database system, as outlined in the provided requirements. The database system will provide users with the ability to define and manage structured data using a custom query language, while also supporting data insertion, deletion, and modification commands.

Project Objectives

1. *Database System Design:* Design a relational database system that adheres to the project requirements and provides efficient data storage and retrieval.
2. *Query Language:* Develop a custom query language that allows users to perform common database operations like projection, filtering, joins, grouping, aggregation, and ordering.
3. *Data Storage and Retrieval:* Implement mechanisms for storing and retrieving data efficiently, considering data structures like B-trees for indexing.
4. *Interactive CLI Interface:* Create an interactive command-line interface (CLI) that enables users to interact with the database system through commands and queries.
5. *Data Modification Commands:* Implement commands for inserting, deleting, and updating data in the database.
6. *Real-World Dataset Integration:* Integrate a real-world dataset, called IMDB Top 250 Movies Dataset*, into the database system to demonstrate its functionality.
7. *Documentation:* Provide comprehensive documentation explaining how to use the database system, including the query language, data modification commands, and real-world dataset integration.

* G, C. R. (n.d.). *IMDB Top 250 Movies Dataset*. Kaggle. Retrieved September 18, 2023, from https://www.kaggle.com/datasets/rajugc/imdb-top-250-movies-dataset

Proposed Timeline

The project will be divided into the following phases:

1. *Design and Planning [25 Sept -29 Sept]:*
    a. Define the data model.
    b. Plan the database file storage mechanism.
    c. Design the query language and CLI interface.
2. *Implementation:*
    a. Develop the database system core. *[2 Oct - 18 Oct]*
    b. Implement the query language parser and executor. *[19 Oct - 27 Oct]*
    c. Create the interactive CLI. *[30 Oct - 7 Nov]*
3. *Testing and Debugging: [8 Nov - 14 Nov]*
    a. Perform unit testing and debugging to ensure system functionality.
    b. Optimize data storage and retrieval mechanisms.
4. *Demonstration and Finalization: [15 Nov - 20 Nov]*
    a. Showcase the database system's functionality using a real-world dataset.
    b. Prepare for presentation.
5. *Documentation: [29 Nov - 7 Dec]*
    a. Create user and technical documentation.
    b. Prepare a demonstration script.

Conclusion

This project aims to create a custom relational database system that aligns with the provided requirements, offering a unique query language and an interactive CLI interface. The successful completion of this project will demonstrate the ability to design and implement a functional database system as a solo developer.

* G, C. R. (n.d.). *IMDB Top 250 Movies Dataset*. Kaggle. Retrieved September 18, 2023, from https://www.kaggle.com/datasets/rajugc/imdb-top-250-movies-dataset