Project Title: Data Persistence and Performance Benchmarking Suite

Project Description:

This project aims to comprehensively evaluate and compare the performance characteristics of four major database systems: SQL, MongoDB, Neo4J, and Hadoop. The central focus is on three key aspects: time efficiency, memory usage, and CPU utilization during data operations. To ensure a uniform approach, Python scripts are employed to automate the loading of data into each database and to execute a variety of queries that simulate real-world data manipulation and retrieval tasks.

The benchmarking process involves measuring the time taken to run each query, alongside monitoring memory and CPU usage, to provide a multi-dimensional view of each database's performance under different conditions. The queries are carefully designed to test a wide range of database functionalities, including but not limited to data insertion, retrieval, updates, and complex joins or relationships in the case of Neo4J.

This project not only serves as a performance assessment tool but also guides in understanding how each database system responds to different data loads and query complexities. The outcomes of this project are anticipated to be a valuable resource for database administrators, system architects, and developers, providing them with critical insights to make informed decisions about the right database technology for their specific needs and scenarios.

By the end of this project, a comprehensive benchmarking suite will be created, showcasing a detailed comparative analysis of SQL, MongoDB, Neo4J, and Hadoop in terms of data persistence and performance efficiency. This suite aims to be a pivotal resource in the database technology selection process, aiding in the optimization of data management strategies across various industries and applications.