

Project 1: DBMS Performance Evaluation

DBMS can help us manage data conveniently and significantly improve the efficiency of data retrieval. PostgreSQL is a popular open-source RDBMS known for its robustness, advanced features, and strong compliance with SQL standards. openGauss is an enterprise-grade open-source RDBMS developed by Huawei, designed for high performance, security, and scalability in demanding business environments.

Design some experiments and try to answer the following questions:

1. What are the unique advantages of a DBMS compared with data operations in files?
2. Which DBMS is better? PostgreSQL or openGauss, and by which standard?

Hints

1. Find some data for experiments. You can use film data with titles, directors, actors, and other related information. Other data such as 'Stock data', 'Library data', etc. is also acceptable. However, the dataset size should be reasonably large. Store the data in both a PostgreSQL database table and a file.
2. Retrieval comparison: Use SELECT in SQL to find films with the word "XXX" in titles and record the execution time. You can obtain the execution time from a client after execution. Write a Java or C/C++ program to perform the same operation through the file. Compare the differences between the two methods.
3. Update comparison: Use UPDATE in SQL to change all "To" in person names to "TTOO". Write a Java or C/C++ program to perform the same operation in the file. Compare the differences between the two methods.
4. Conduct other comparisons and experiments as desired. For example, you can reorganize data into different formats for faster retrieval. You are recommended to study the mechanisms of DBMS for storing and retrieval.
5. The comparisons should be well-designed, comprehensive, and persuasive, backed by relevant experiments.
6. You can use Docker to install openGauss since openGauss is not well-supported by many operating systems.
7. Optional: You are encouraged to release the report and related source code online. You can use GitHub or any other preferred platform for hosting. Remember to include the URL in your submission and try to help others understand your findings.

The Report

1. This is a personal project. Each student should complete it separately and submit one report for the project.
2. The submitted report should contain the design, implementation, and evaluation of the functions.
3. Highlighting the key features and results.
4. You do not need to implement many complex features to improve efficiency since the motivation is not to implement a DBMS.

Rules:

1. The project report and the source code must be submitted before the deadline. Any submission after the deadline (even by 1 second) will result in **a score of 0**. The deadline is 23:59 on Oct. 19.
2. The files should be submitted as report.pdf, [filename].Java or [filename].cpp. The files should **NOT** be compressed into one.
3. You may use AI tools like DeepSeek or ChatGPT to help you understand concepts and configure software. However, do NOT use AI tools to generate low-quality or irrelevant text. Avoid producing verbose or meaningless content. When writing your report, always consider: "Will readers find the content interesting? Are the findings useful to readers?"
4. The score will depend on the quality of both the source code and the report. The report should be easy to understand and provide a clear description of the project, especially the highlights.