**SBD Project 2023/2024**
**Delivery date: June 9th, 2024**

The objective of the Project is to provide you with practical experience regarding benchmarking database systems using standard and open tools. For this purpose, we will use HammerDB (https://www.hammerdb.com/) the leading open benchmarking tool for several commercial and open source database management systems.

The team must have 3 elements, and ideally should have different types of machine configurations so that more diversity can be explored.

The team may have several approaches, and all are welcome:

- Configure a DBMS to achieve better performance in the benchmark(s)
- Tune the database to achieve better performance in the benchmark(s)
- Analyze the performance when the number of users/size of the database increases
- Compare the different characteristics of the benchmarks and add extra characteristics to the benchmark (e.g. new transactions, new queries)
- Compare two or more database systems for a particular setting
- …

You must deliver with at most 30 pages, comprising:

- Introduction
- Overview of HammerDB
- Problem to be addressed and brief summary of DBMS that will be used, including a description of the features that you will use in the benchmarking
- Description of the benchmark that will be used
- Methodology to perform the benchmarking, including machine configuration, DBMS configurations, experimental setup, and metrics used
- Results of tests, including used metrics and figures showing the behaviour of the system under test
- Discussion of the results
- Conclusions

You also should deliver all the scripts that you have produced (NO DATA PLEASE!), and results gathered in the form of CSV files and figures.

It would be preferable that in your team there is at least one Windows or LINUX laptop since MacOS systems do not run easily the GUI interface. The GUI client is nice to start, but afterwards you might not use it anymore.

The following plan is a good starting point for your project:

1) Read the documentation of HammerDB and install the benchmarking tools for your system (regarding MacOS systems, you need to use the Docker images available. It might be difficult to put the client GUI working, but you have a command line client available…)
2) Run a simple test following the steps in the Quick Start chapter of the HammerDB documentation. If it is taking a lot of time, reduce the number of virtual users, warehouses and transactions to test (start slowly!)
3) You have two benchmarks available: TPROC-C and TPROC-H, one for OLTP and OLAP workloads. Try to understand them and select the one(s) to use. TPROC-C is presented in Chapters 3 and 4, with the schema of the database in section 3.5 and the kind of workload used to test the system. TPROC-H is described in Chapter 11, and is more appropriate to the students with background in Datawarehouses (eg. Took the Data Modelling course in the 1st semester).
4) Define the problem/approach you want to follow and select the DBMS(s) to benchmark
5) Read chapters 7, 9 and 13 from the HammerDB documentation. These are very helpful if you wish to tailor your benchmarks and database configuration or schema. Try to do a test manually so that you understand all the details involved.
6) Perform the testing of your project (you will need 2 to 3 weeks to do this). Gather the results and do not limit yourselves to a single run (try at least 5 times each, if possible), and keep all the results. When presenting any metrics do not forget to mention the number of executions, the average times, the standard deviation, and minimum and maximum times of execution. The HammerDB has other metrics that make possible to compare different systems (see Documentation!)
7) Write the report (do not leave this to the end!). While you collect the data try to analyze it since this might give you clues on how to improve the system

**GOOD LUCK!**