20th International Conference on Advances in ICT for Emerging Regions (ICTer 2020): 276 – 277

A comparative study between the capabilities of MySQL and ClickHouse in low-performance Linux environment

Akila Wickramasekara
University of Colombo School of Computing (UCSC)
Colombo 7, Sri Lanka
akilawickramasekara@gmai.com

M.P.P.Liyanage
Department Of ICT,
Faculty of Technology
University of Sri Jayewardenepura,
Nugegoda, Sri Lanka
prabani@sjp.ac.lk

Udayagee Kumarasinghe
Department of Biosystems Technology,
Faculty of Technology
University of Sri Jayewardenepura,
Nugegoda, Sri Lanka
udayagee@sjp.ac.lk

2020 20th International Conference on Advances in ICT for Emerging Regions (ICTer) | 978-1-7281-8655-9/20/$31.00 ©2020 IEEE | DOI: 10.1109/ICTer51097.2020.9325483

Abstract—Fast, reliable, and secure data processing is critical in the Information Technology (IT) sector as it is an indispensable requirement of all the growing technologies. Achieving this requires handling large amounts of data and massive computing power. Emerging technologies use devices under low-performance environments, such as raspberry-pi, which require highly efficient databases under low computing power. Inappropriate selection of databases leads to considerable resource wastage in numerous ways, such as waste of time, cost, and energy. The selection of a reliable, user-friendly database compatible with low-performance environments will downplay the energy requirement and cut down applications' carbon footprint. Thus, this study was conducted to compare the performance and query cost of using ClickHouse and MySQL databases as column-oriented and row-oriented databases in a low-performance environment. The effectiveness of using column-oriented or row-oriented databases with large amounts of data at different platforms was analyzed. The study was conducted by running different queries using an automated PHP script and checking the execution time, processing power, disk I/O, and memory usage for each query. The study was conducted with a virtual Linux environment of 1024MB RAM, 1.8Ghz single processor, and 25Gb storage, representing very low computing power, even lesser than the recommended minimum requirement of ClickHouse and MySQL databases. Both databases performed well in the above mentioned environment. Our experimental results show that ClickHouse execution time and disk write speed is tremendously low compared to MySQL. To obtain this performance, ClickHouse used more CPU processing power and memory utilization than MySQL.

Keywords— databases, MySQL, ClickHouse, column-oriented, row-oriented, Linux

I. INTRODUCTION

Currently available platforms or systems process huge amounts of data to give the required output for relevant events or end users. A system that processes a large bundle of data in a fast, reliable way requires a suitable environment and it should bind with a fitting database. However, poorly planned systems are commonly found in the sector which leads to system revamp or scaling with the growth of data. Choosing DBMS (Database Management System) is the key point of developing software systems and this needs to be decided by considering data growth rate and the data availability of the application. In the future, Isolated and intelligent devices will play a significant role in different sectors. At the same time, those devices will self-process the data without connecting to the internet—these require reliable databases that can perform under limited architecture and a low-performance environment. The objectives of this study are to compare the speed and efficiency of MySQL and ClickHouse with limited computing power and to find the most efficient DBMS which runs within limited resources. The study will also appraise behaviors of two databases and their capabilities in thin clients which act as the client and server by itself at the same time. MySQL and ClickHouse both Compatible with Reduced Instruction Set Computer (RISC) architecture, and we aimed to encounter the most efficient database, to run mini-computers with the RISC architecture.

II. RELATED WORKS

MySQL and ClickHouse are two open-source database management systems that are using most of the systems. MySQL is a row-oriented widely used open-source RDBMS in the world and became the most popular DBMS in 2019. [1],[2] ClickHouse is an open-source column-oriented DBMS mostly used for online analytical processing (OLAP).[3]

In the Row Oriented Database all the attributes of a specific record are added first and after it will go to the next record insertion. Therefore, Row-oriented systems are more productive when they need to select many columns of a single row at once. [4], [5]

A Column-Oriented Database will serialize all of the values of a column together, after the values of the next column, and so on. In a Column-Oriented Database, it will retrieve only the relevant columns in the query. Column-oriented databases are more productive when they need to process or compute a large amount of data from a large number of rows. [4], [5]

Both database systems consist of more similar features and functionalities which makes this comparison more reasonable with the two DBMS. Recommended minimum computing power to run ClickHouse is 2GB of RAM and 2 cores CPU with 1.3GB disk space and Recommended minimum computing power to run MySQL is 4GB RAM and 16 cores CPU with 20GB disk space.[6], [7]

The RISC architecture helps to produce more CPU power by simplifying the instruction set of the CPU. RISC is a type of microprocessor architecture that utilizes a small, highly optimized set of instructions, rather than a more specialized set of instructions often found in other types of architectures.[8]

III. METHODOLOGY

A virtual Linux environment with 1GB RAM and 1.8Ghz single-core processor and SSD (Solid State Drive) storage

978-1-7281-8655-9/20/$31.00 ©2020 IEEE

Authorized licensed use limited to: Univ of Calif Santa Barbara. Downloaded on May 18,2021 at 07:21:18 UTC from IEEE Xplore. Restrictions apply.

were used as the testing platform which represents low computing power. In order to get an idea about query behaviors we used query commands types such DDL(Data Definition Language) and DML(Data Modeling Language)[9], [10] We developed a PHP script to automate the data collection part and it will run 30 pre-define DDL and DML queries in both databases. Both databases consist of 1.5 million records with the same data and same tables. The PHP script will automatically record 17 parameters when the queries are executed and from that we mainly focus on average execution time, load average, CPU utilization, memory utilization, disk read I/O and disk write I/O.

IV. RESULTS AND DISCUSSION

Fig. 1 depicts an overview of the comparison obtained from data related to query execution time, system load CPU, Memory and Disk. According to the results obtained, a significant difference was observed between the execution time of MySQL and ClickHouse except for bulk export. The execution time of ClickHouse is relatively lower than MySQL. Comparison of Load averages of two databases clearly shows that the ClickHouse produces a higher load to the server than MySQL. With the information derived from CPU utilization and Memory allocation, a noticeable difference was observed; ClickHouse consumes more processing power than MySQL. MySQL CPU utilization is very low, and it is an ignorable amount compared to ClickHouse. Further, it uses a considerable memory than MySQL. Disk specific data shows that ClickHouse disk I/O write time is meager compared to MySQL disk I/O write time, and the disk read I/O time of ClickHouse is higher than MySQL.

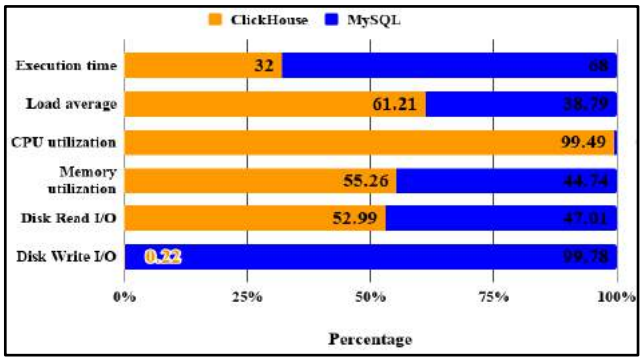

Fig. 1. Overall Comparison

According to Fig. 1, the query execution time for the ClickHouse is only about 32%, out of the total time consumption for both DBMSs. Further, it is only half of the time from the time it has taken for MySQL. This reveals that ClickHouse is faster than MySQL in a limited resource low-performance environment. With this, ClickHouse can be recommended for the low resource environments which handle a large set of data, while acting as isolated devices. As ClickHouse supports RISC architecture with limited resources this study emphasizes that ClickHouse is more suitable for applications which run within mini computers and portable computers such as Raspberry Pi, Banana Pi, etc. will be an advantage for critical data manipulation applications such as applications in different industries. This outcome emphasized that ClickHouse is more potent for fast data manipulation in a low-performance environment.

As per the result, the efficiency of ClickHouse is much higher than MySQL in a limited resource low-performance environment and this will lead to maintaining low cost and portable database servers. At the same time, it expresses that ClickHouse is more environmentally friendly than MySQL as it has a high response time and gives high-speed results by using limited power consumption. This finding will lead to a high impact on energy saving as there are millions of databases using MySQL in large database servers.

V. CONCLUSIONS

This study was conducted to compare and observe the most efficient DBMS that runs in low performance and low power computer systems that can be used for further research areas. The results showed the capability to use these databases in the small microprocessor, leading to running high-speed stand-alone AI and ML applications in miniature environments. Both Databases performed well under limited computing power, even lesser than the recommended minimum requirement of ClickHouse and MySQL databases. Our experimental results show that ClickHouse execution time and disk write speed is tremendously low compared to MySQL still, to obtain this performance, ClickHouse used more CPU processing power and memory utilization than MySQL.

VI. REFERENCES

[1] "Ranking of the most popular database management systems worldwide, as of June 2020." statista.com. https://www.statista.com/statistics/809750/worldwide-popularity-ranking-database-management-systems (accessed May 31, 2020).

[2] "2019 Database Trends – SQL vs. NoSQL, Top Databases, Single vs. Multiple Database Use." scalegrid.io. https://scalegrid.io/blog/2019-database-trends-sql-vs-nosql-top-databases-single-vs-multiple-database-use (accessed May 31, 2020).

[3] "ClickHouse Overview." clickhouse.tech. https://ClickHouse.tech/docs/en (accessed May 31, 2020).

[4] V. Bhagat and A. Gopal, "Comparative study of row and column oriented database," *Int. Conf. Emerg. Trends Eng. Technol. ICETET*, pp. 196–201, 2012, doi: 10.1109/ICETET.2012.56.

[5] A. KumarDwivedi, C. S. Lamba, and S. Shukla, "Performance Analysis of Column Oriented Database Vs Row Oriented Database," *Int. J. Comput. Appl.*, vol. 50, no. 14, pp. 31–34, 2012, doi: 10.5120/7841-1050

[6] "ClickHouse Requirements." clickhouse.tech. https://clickhouse.tech/docs/en/operations/requirements (accessed May 31, 2020).

[7] "System Requirements." mysql.com. https://dev.mysql.com/doc/mysql-monitor/4.0/en/system-prereqs-reference.html (accessed May 31, 2020).

[8] I. B. M. Risc and P. Chow, "RISC-(reduced instruction set computers)," IEEE Potentials, vol. 10, no. 3, pp. 28–31, Oct. 1991, doi: 10.1109/45.127642.

[9] J. G. Ramakrishnan Raghu, R. Ramakrishnan, J. Gehrke, and J. Gehrke, *Database management systems*, vol. 3. McGraw-Hill New York, 2003.

[10] A. Silberschatz, H. F. Korth, and S. Sudarshan, "System Concepts," 2008.

Authorized licensed use limited to: Univ of Calif Santa Barbara. Downloaded on May 18,2021 at 07:21:18 UTC from IEEE Xplore. Restrictions apply.