International Journal of Internet, Broadcasting and Communication Vol.10 No.2 31-38 (2018)
http://dx.doi.org/10.7236/IJIBC.2018.10.2.6

IJIBC 18-2-6

Experiments of Search Query Performance for SQL-Based Open Source Databases

Meekyung Min*

Dept. of Computer Science, Seokyeong University, Seoul, Korea
mkmin@skuniv.ac.kr

Abstract

As the use of open source databases grows, so does need to evaluate, the performance of search queries for these databases. This paper compares the search query performance of SQL-based open source databases with commercial databases through experiments. The targets are MySql, MariaDB, and MS-SQL Server. In this study, the execution time of various types of search queries are measured. Also, search query performance was experimented according to change of index and number of tuples. Experimental results show that SQL-based open source databases have the potential to replace commercial databases when indexes are used and the number of tuples is not very large.

Keywords: *Search Query, SQL, Open Source Database, Query Performance*

1. Introduction

The Open Source Software is a software distributed with licenses that allow access to source code, free redistribution, creation of derivative works, and unrestricted use [1]. Open-source applications are increasingly used in almost every area of systems, such as operating systems and databases, software development tools, education, science, and engineering. They are increasingly being used as an alternative to expensive commercial software. There are many kinds of open source software. Database management system (DBMS) is one of them that can be classified as open source software for servers. The main purpose of database management system is to efficiently store and manage data for users and to efficiently search and use databases through search queries [2]. In recent years, not only conventional commercial databases but also open source databases have been widely used. Therefore, it is necessary to study the performance of search queries that search the database to find out whether an open source database is suitable as a substitute for existing commercial database.

In this paper, the performance of search queries of open source databases is demonstrated and compared with commercial database through experiments. Performance is defined as processing time for search queries. In the experiments, the most popular SQL-based open source databases, MySql and MariaDB are used. Also, the MS-SQL Server, which is the most popular database on Windows is used for the commercial database example. The types of search queries in the experiments are singleton query, multi-condition query, range

Manuscript Received: Mar. 2, 2018 / Revised: Mar. 11, 2018 / Accepted: Mar. 20, 2018
Corresponding Author: mkmin@skuniv.ac.kr
Tel:+82-2-940-7750, Fax: +82-2-919-0345
Dept. of Computer Science, Seokyeong University, Seoul, Korea

query, and join query. For each query, the experiments are done with index and also without index.

Section 2 reviews the current state of SQL-based open source databases. Section 3 shows the results of the query performance tests. Finally, section 4 concludes and discusses the results of this paper.

2. Related Works

Typical open source databases include MySql, PostgreSQL, MariaDB, SQLite, and MongoDB. The MySql is an SQL-based relational database management system. It is the #1 open source database for web-based applications, used by Facebook, Twitter, YouTube. MySql is now owned by Oracle Corporation, and is increasingly of interest to large enterprises as well as small and medium-sized businesses and personal sites.

PostgreSQL is an object-relational database management system (ORDBMS) that emphasizes extensibility and standards compliance. It can manage multiple loads, from small single machine applications to large internet applications with many concurrent users. Unlike relational database systems, PostgreSQL allows users to create arbitrary database objects to provide.

MariaDB is also an SQL-based open source relational DBMS based on the same source code as MySql. MariaDB has been created to counter the current uncertain MySql licensing status of Oracle. MariaDB is very popular as open source software because of its full compatibility with MySql.

SQLite is a relatively lightweight database used by applications rather than servers. It is also a database built in to the Google Android operating system. It is a representative mobile open source database.

MongoDB is a free and open-source cross-platform document-oriented database program. Classified as a NoSQL database program, MongoDB uses JSON-like documents with schemas.

The most widely used commercial databases include Oracle's Oracle and Microsoft's MS-SQL Server. Microsoft's MS-SQL Server is a relational database management system developed by Microsoft and represents a commercial database under the Windows operating system. With a market share of about 17% in Korea, it occupies the database market at the highest rate in the world.

In the past, research on the performance of open source software has been conducted [6,7,8]. The research [6] analyzed the cases of open source software with open source code at the code level. Among the performance evaluation features, reliability, security, efficiency, portability, were selected. The research [7] and [8] proposed a model that modified ISO/IEC quality model [3] to open source software and developed a program to automate quality evaluation based on this model. Research on search query performance has been going on for decades since the advent of databases[2,4]. This is also very important in open source databases. The research [5] analyzed the query performance based on a single query model in open source DBMSs.

This research measures the performance of SQL-based open source databases based on various search query models. MySql and MariaDB are used as open source databases, and MS-SQL Server is used as a commercial database.

3. Experiments of Search Query Performance

3.1 Experimental Environment

In this paper, the experiment measured the performance of search queries in three databases. MySql and MariaDB, which are the most widely used SQL-based open source databases are used in the experiments. Because this study measured performance in Windows environment, the MS-SQL Server, representing the Windows DBMS, was chosen as the commercial database. HeidiSQL was used as a front end to measure the performance of search queries in three databases under the same environment and same interface. HeidiSQL

is a free software and an open source client, MySql front-end product. It supports MySql, MariaDB and MS-SQL Server under Windows environment.

The experimental environment is as follows: CPU i7-4790 3.60GHz, 8GB RAM, NVIDIA GTX 750 Graphics, 64-bit Windows 10 home OS and HeidiSQL version 9.5.0.5196. The databases used in the experiments are MariaDB-10.2.14-winx64, MySQL-community-8.0.11.0 and Microsoft SQL Server 2017. Figure 1 is an instance of the performance experiment screen.

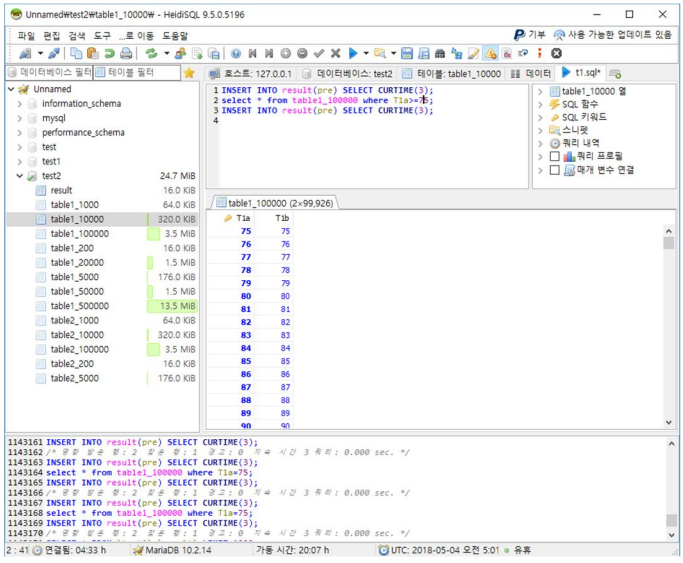

Figure 1. Experiment Screen

The data used in the experiment are arbitrarily generated tables. Tuples in the tables were randomly generated. The number of tuples in a table varied from 200 to half-million. The time to perform a search query was measured by varying the number of tuples in the table. Each result of every search query has a selectivity of 15%, except for the singleton query. For every query in each database, the processing time was measured 30 times, depending on the number of tuples. The average value of 30 searching times was used in the analysis.

3.2 Search Queries

The following queries are representative types of SQL-based search queries, and were used in experiments of this study.

- Singleton Query
- Multi-condition Query
- Range Query
- Join Query

The singleton query specifies one condition for the attribute, and the result of the query is a singular tuple. In this experiment, two cases have been tested: a query with an indexed primary key and a query with no indexed attribute. The multi-condition query is a query with more than one predicate in the search condition. The Range query is a search type that searches sequential data within a certain range. The Join query is a query that takes the longest processing time. This experiment measures two cases for every query, depending on the presence or absence of an index.

3.3 Experimental Results

3.3.1 Singleton Query

Table 1 and Figure 2 show the experimental results for singleton query with no index and with primary key (PK) index. The horizontal axis of Figure 2 is the number of tuples in the relation, and the vertical axis is the average processing time of the query in seconds. In order to facilitate identification, only up to 100,000 tuples are displayed on the graph. The number of tuples in the tests were 200, 1000, 5,000, 10,000, 20,000, 50,000, 100,000, and 500,000.

Table 1. Singleton Query Experiments

	Average Searching Time (seconds) – NO INDEX							
	200 tuples	1,000	5,000	10,000	20,000	50,000	100,000	500,000
MariaDB	0.017367	0.020800	0.022167	0.022567	0.024700	0.032333	0.045067	0.141533
MS-SQL	0.021667	0.024367	0.022667	0.022600	0.024300	0.024933	0.026567	0.040467
mySql	0.029000	0.030800	0.035300	0.035900	0.036100	0.044400	0.056900	0.188500
	Average Searching Time (seconds) - PK INDEX							
	200 tuples	1,000	5,000	10,000	20,000	50,000	100,000	500,000
MariaDB	0.020467	0.020200	0.020500	0.020633	0.020500	0.020900	0.020933	0.026400
MS-SQL	0.024367	0.024500	0.024033	0.024933	0.025367	0.025900	0.025133	0.025067
mySql	0.028133	0.026400	0.028133	0.027667	0.030033	0.027800	0.027267	0.028833

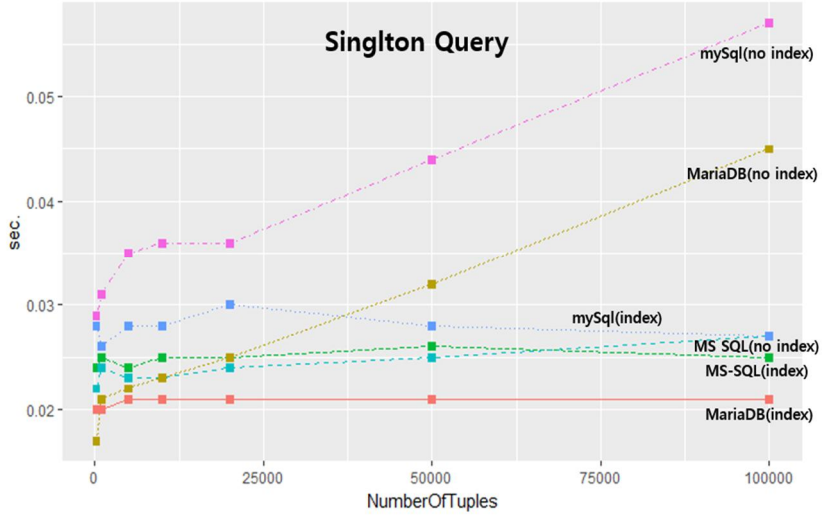

Figure 2. Singleton Query Performance

For singleton query without index, mySql has the worst performance regardless of the number of tuples, and the larger the number of tuples, the larger the difference to MS-SQL Server. Until 10,000 tuples, MariaDB is the fastest, then MS-SQL Server, and last is mySql. Starting with 20,000, MS-SQL is the fastest, the next one is MariaDB, and mySql is the slowest. The larger the tuple size, the bigger the performance difference between the MS-SQL Server and the open source database. MS-SQL performance is stable regardless of the number of tuples.

For singleton queries using PK indexes, there were few differences between the three databases. MariaDB almost always shows the best performance, followed by MS-SQL, and then mySql. MariaDB is

slower than MS-SQL only when the number of tuples is very large (500,000). All three databases were hardly affected in their performance depending on the number of tuples.

3.3.2 Multi-condition Query

Tables 2 and Figure 3 show the results of multi-condition query with no index and indexes. Figure 3 illustrates up to 100,000 tuples because the time value suddenly increases at 500,000.

Table 2. Multi-condition Query Experiments

	Average Searching Time (seconds) for Range Query with NO INDEX							
	200 tuples	1,000	5,000	10,000	20,000	50,000	100,000	500,000
MariaDB	0.023933	0.024300	0.025933	0.025333	0.027167	0.033500	0.041967	0.126533
MS-SQL	0.027233	0.025900	0.026033	0.024833	0.027133	0.026300	0.037533	0.042667
mySql	0.032500	0.033100	0.036600	0.036900	0.031600	0.028933	0.028133	0.149867
	Average Searching Time (seconds) for Range Query with INDEX							
	200 tuples	1,000	5,000	10,000	20,000	50,000	100,000	500,000
MariaDB	0.031367	0.030103	0.029233	0.029167	0.022900	0.024533	0.024767	0.025033
MS-SQL	0.026167	0.027400	0.028833	0.027300	0.025800	0.029367	0.024433	0.027200
mySql	0.033933	0.034567	0.031433	0.033840	0.033933	0.030767	0.032667	0.033267

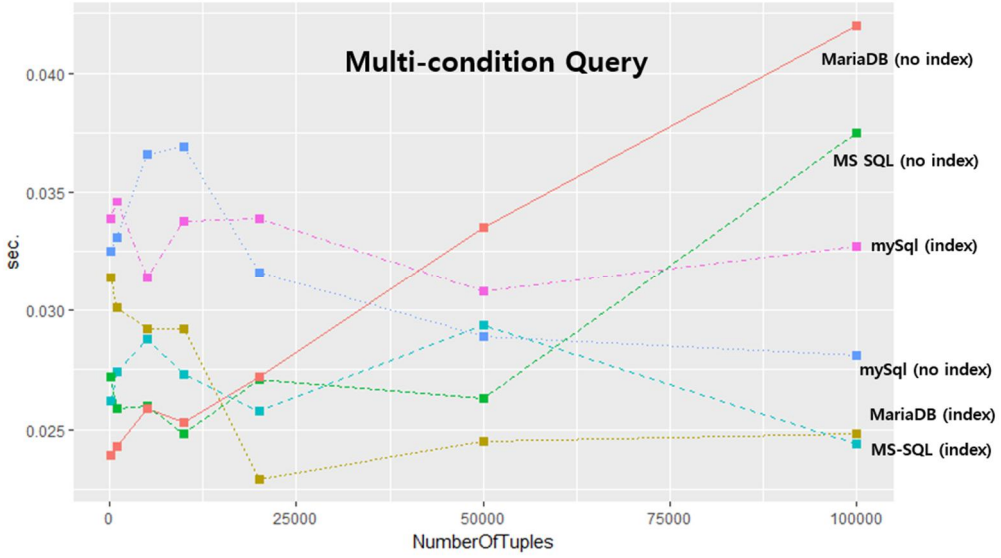

Figure 3. Multi-condition Query Performance

For multi-conditional query without index, the performance order of the three databases continues to change until the number of tuples reaches 100,000. After then, at 500,000, mySql and MariaDB become very slow compared to MS-SQL, which experiences little degradation.

For multi-conditional query with indexes, MS-SQL is the fastest until the number of tuples reaches 10,000, followed by MariaDB, and mySql. When the tuple is 20,000, the order is reversed and MariaDB is the fastest, MS-SQL is second and mySql is third. The three databases showed stable performance with indexes.

3.3.3 Range Query

Table 3 and Figure 4 show the experimental results of range query with no index and with index. In the case of range query with no index, MS-SQL shows the best performance regardless of the number of tuples.

MariaDB is next, and then mySql. MySql is faster than MariaDB only if the tuple size is very small (200 tuples). In MS-SQL, the performance degradation rate from the increase of tuples is relatively moderate.

Table 3. Range Query Experiments

	Average Searching Time (seconds) for Range Query with NO INDEX							
	200 tuples	1,000	5,000	10,000	20,000	50,000	100,000	500,000
MariaDB	0.032733	0.034133	0.035900	0.035533	0.036500	0.043867	0.055367	0.139533
MS-SQL	0.029233	0.033467	0.033800	0.034033	0.036200	0.040700	0.048467	0.114633
mySql	0.032567	0.035133	0.036533	0.038533	0.041367	0.047100	0.057323	0.171000
	Average Searching Time (seconds) for Range Query with INDEX							
	200 tuples	1,000	5,000	10,000	20,000	50,000	100,000	500,000
MariaDB	0.033167	0.033933	0.035600	0.036500	0.036033	0.038367	0.041567	0.075167
MS-SQL	0.025133	0.033633	0.034367	0.034833	0.035900	0.038900	0.046733	0.096723
mySql	0.033233	0.034833	0.034900	0.035233	0.039133	0.039033	0.045033	0.075027

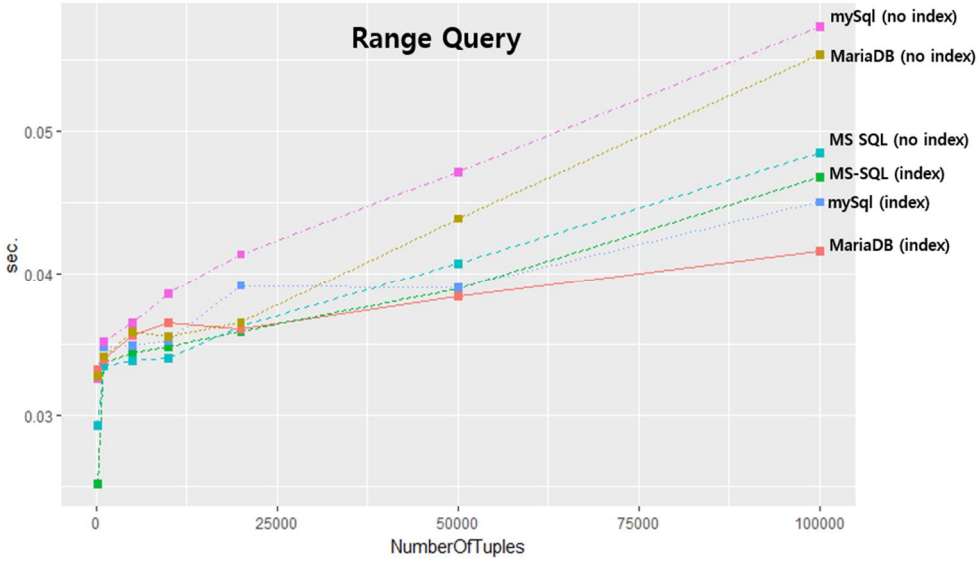

Figure 4. Range Query Performance

In the case of range query with index, when the number of tuples is from 200 to 20,000, MS-SQL performs best. If the number of tuples is 50,000 or 100,000, MariaDB has the best performance, and when it is 500,000, mySql has the best performance. That is, when the number of tuples is more than 50,000, the performance of the open source database is better.

3.3.4 Join Query

Table 4 and Figure 5 shows the result of join query performance with index and no index. With an index, two tables are joined on foreign key. As in table 4, the test was not performed at 500,000 because the searching time was too long. As the range of values is too wide, the vertical axis of the graph is represented by logarithm.

In the case of join query with index, until the tuple size reaches 1,000, MariaDB is the fastest. Starting from 5,000, the performance of MS-SQL is the best. And the more the number of tuples thereafter, the greater the difference between MS-SQL and open source databases. The gap between MariaDB and mySql was relatively small.

In the case of join query with index, MS-SQL Server is the fastest only except at 200 tuples. In this case, the gap between the open source databases and MS-SQL is very large. Especially if the number of tuples is large, it causes a fatal slowdown in the open source database. On the other hand, the performance of MS-SQL is stable against increasing tuples.

Table 4. Join Query Experiments

	Average Searching Time (seconds) for Join Query with INDEX							
	200 tuples	1,000	5,000	10,000	20,000	50,000	100,000	500,000
MariaDB	0.035633	0.043833	0.058600	0.073800	0.090967	0.178700	0.293600	1.385867
MS-SQL	0.037067	0.046800	0.048933	0.053067	0.054800	0.072800	0.088667	0.198733
mySql	0.038033	0.047733	0.058267	0.072467	0.095700	0.202667	0.333067	1.679600
	Average Searching Time (seconds) for Join Query with NO INDEX							
	200 tuples	1,000	5,000	10,000	20,000	50,000	100,000	500,000
MariaDB	0.0274	0.0680	0.9504	3.6772	14.5854	85.3372	363.9608	-
MS-SQL	0.0336	0.0376	0.0376	0.0430	0.0540	0.0702	0.1412	-
mySql	0.0316	0.0704	0.9602	3.7684	14.7672	92.9534	371.6894	-

Figure 5. Join Query Performance

4. Conclusion

In this paper, search query performance was measured for MariaDB and mySql, which are SQL-based open source databases, and for MS-SQL Server, which is a commercial database, using various types of queries. The most representative types of queries, singleton query, multi-condition query, range query, and join query were used in the experiments. In each case, the experiment was conducted with and without an index. The number of tuples was experimented with 200, 1,000, 5,000, 10,000, 20,000, 50,000, 100,000, and

500,000 cases. In the experiment, the processing time of the search query was measured. The performance was analyzed using the average value of the 30 experiments.

For singleton query without index, it is better to use MariaDB when the tuple size is smaller than 10,000, and to use MS-SQL Server when the tuple is larger than 10,000. MariaDB showed the best performance for singleton query with index. In the case of a multi-condition query, the performance rankings between the three databases were constantly changing, so it was hard to see any significant difference in performance. In addition, there was little change in performance depending on the number of tuples. However, if the number of tuples is 500,000 and there is no index, the open source database was not appropriate. In the case of range query, the performance of open source databases was low without index. However, the performance of open source databases was better when there were indexes and more than 50,000 tuples for range query. In the case of join query, MS-SQL Server showed performance stability with or without indexes. An indexless join is a weakness of the open source database. Especially when the number of tuples is large and there is no index, it is almost impossible to use an open source database.

The following conclusions can be made from this experiments. The performance of open source databases was most affected by the presence or absence of indexes. When indexes are used, the performance of open source databases often surpasses that of commercial databases depending on the type of search query. Next, performance was affected by the number of tuples. The larger the number of tuples, the more often the performance of the open source database is degraded. Finally, in terms of query type, in join query, the performance of open source databases did not reach the performance of the commercial database. The results of this study can be used as a basis for selecting a database.

Acknowledgement

This research was supported by Seokyeong University in 2018.

References

[1] S. Androutsellis-Theotokis, D. Spinellis, M. Kechagia, and G. Gousios, *Open Source Software: A Survey from 10,000 Feet*, Foundations and Trends in Technology Information and Operations Management and Publish, 2011.

[2] R. Elmasri, and S. B. Navathe, *Fundamentals of Database Systems, 7th Edition*, Pearson, 2016.

[3] ISO/IEC, *System and software engineering – Systems and software Quality Requirements and Evaluation*, ISO/IEC TR-25010, 2011.

[4] N. Kim, "Experimental Evaluation Method for Performance Analysis in Web Application Services", *The Journal of the Institute of Internet, Broadcasting and Communication*, Vol. 16, No.1, pp. 69-75, 2016. DOI: http://dx.doi.org/10.7236/JIIBC.2016.16.1.69

[5] M. Min, "Evaluating the Performance Quality of Open Source Database Management Systems", *Journal of the Korean Society for Quality Management*, Vol. 45, No. 4, pp. 933-942, 2017. DOI: https://doi.org/10.7469/JKSQM.2017.45.4.933

[6] I. Samoladas, G. Gousios, D. Spinellis, and I. Stamelos, "The SQO-OSS quality model: measurement-based open source software evaluation", *Open source development, communities and quality*, pp. 237-248, 2008.

[7] D. Spinellis, *Code Quality: The Open Source Perspective*, Addison Wesley Professional, 2006.

[8] D. Spinellis, G. Gousios, V. Karakoidas, P. Louridas, and P. Adams, "Evaluation the Quality of Open Source Software", *Electronic Notes in Theoretical Computer Science*, Vol. 233, pp. 5-28, 2009.