

PAPER • OPEN ACCESS

Inner Join Query Performance: MariaDB vs PostgreSQL

To cite this article: U Hairah and E Budiman 2021 *J. Phys.: Conf. Ser.* **1844** 012021

View the [article online](#) for updates and enhancements.

You may also like

- [The LSST DESC DC2 Simulated Sky Survey](#)
The LSST Dark Energy Science Collaboration (LSST DESC), Bela Abolfathi, David Alonso et al.
- [Column Store for GWAC: A High-cadence, High-density, Large-scale Astronomical Light Curve Pipeline and Distributed Shared-nothing Database](#)
Meng Wan, Chao Wu, Jing Wang et al.
- [Database on Demand: insight how to build your own DBaaS](#)
Ruben Gaspar Aparicio and Ignacio Coterillo Coz



The Electrochemical Society
Advancing solid state & electrochemical science & technology

241st ECS Meeting

May 29 – June 2, 2022 Vancouver • BC • Canada

Abstract submission deadline: Dec 3, 2021

Connect. Engage. Champion. Empower. Accelerate.
We move science forward



Submit your abstract



Inner Join Query Performance: MariaDB vs PostgreSQL

U Hairah¹ and E Budiman¹

¹Informatics Department, Universitas Mulawarman, Samarinda 75119, Indonesia

Email: ummihairah@gmail.com

Abstract. Research to test the performance of inner joint queries between MariaDB and the PostgreSQL database management system. Inner joint queries have the same output in MariaDB and PostgreSQL DBMS. This study aims to provide a comparative analysis of the response time performance of MariaDB and PostgreSQL databases in the use of inner join queries. The research test was conducted with 1,050,000 records using 21 response time data collection scenarios ranging from 50,000 data to 1,050,000 data collection with multiples of 50,000 data. The PostgreSQL DBMS was found to have better response times than MariaDB in all test scenarios. The results of the regression analysis to test the effect of using the F-test show that there is a significant effect on the amount of data and the relationship together on the response time in MariaDB and PostgreSQL. Then the T-test is carried out to strengthen the results of the study that PostgreSQL has a faster response time than MariaDB, testing 1 combined relationship shows a significant difference, testing 2 combined relationships shows a significant difference and testing 3 combined relationships shows a significant difference.

1. Introduction

PostgreSQL and MariaDB are database management systems that are free-opensource distributed under the BSD license agreement. This software is the most widely used database today, apart from MySQL and Oracle. January 2019 PostgreSQL is the database management system (DBMS) that gained more popularity in DB-Engine Ranking within the last year than any of the other 343 monitored systems and thus declare PostgreSQL as the DBMS of the Year 2018[1]. Meanwhile, the MariaDB has reached a new all-time high in the DB-Engine Ranking, rising to 14th place, up from 18th place last year[2]. The rank DBMS according to [3], is a list of DBMS sorted by how much they managed to increase their popularity in 2018, or in other words, how many additional people started communicating in one of the ways we measure in our methodology, e.g. job offers, entries professional profiles and quotes on the web[1]. In this regard, study motivations was directed to test the performance of PostgreSQL and MariaDB dbms based on response time to queries and data using a pharmaceutical database.

PostgreSQL was developed by the Berkeley Computer Science Department and is an Object-Relational Data Base Management System (ORDBMS). PostgreSQL also has a fairly complete set of features. These features support transactions, views, sub selects, foreign key support, checks, and others. Meanwhile, the MariaDB (MariaDB Corporation Ab and Foundation) initially had a focus on enhancing existing MySQL features, for example, its Aria storage engine is meant to be an improved MyISAM implementation. They quickly looked beyond MySQL for new feature development, offering functionality that was not available in MySQL at that time. More recently, MariaDB is very active in broadening its scope either by developing new components or by acquiring DBMS technologies and merging them into their system. Examples of that are the MariaDB ColumnStore, originally InfiniDB, the TokuDB high-performance storage engine, the Open Query GRAPH computation engine for supporting tree structures and graphs, the Sphinx storage engine to enable advanced text searching and



the Spider storage engine enhancing support for distributed databases[2]. From the various aspects of each of the advantages of the two DBMSs, therefore the authors are interested in analyzing the two DBMS into a research material based on the query performance of each DBMS.

The study purpose is to provide a comparative analysis of the DBMS query performance PostgreSQL 13.0 and MariaDB 10.5.6. In particular the authors conducted a comparative analysis of inner join query optimization to determine the response time in the use of inner join queries by testing the effect of regression analysis using the F-test and comparison analysis using the T-test. The research test was conducted with 1,050,000 records using 21 response time data retrieval scenarios ranging from 50,000 data to 1,050,000 data retrieval with multiples of 50,000 data. The 1 join, 2 join and 3 join relations were tested using parametric statistical analysis (T-test), homogeneity testing of the two independent variables and the analysis method of the Linear Regression Effect Test with the F-Test.

Several studies related to the DBMS comparative analysis performance have been carried out, including Katarzyna Kroc et. al; Performance analysis of relational databases MySQL, PostgreSQL, MariaDB[4] it compares the performance of MySQL, PostgreSQL, MariaDB and H2 relational databases. Execution speed will be tested on simpler and more complex SQL queries. The test results will show which database is more efficient[4]. A database integrated system[5], a study on data input and output performance comparison of MongoDB and postgresql in the big data environment[6], database query performance and optimization[7], Comparison of performance between MariaDB and PostgreSQL in terms of CPU overhead[8], Web performance optimization[9], [10], A comparison of database performance of MariaDB and MySQL with OLTP workload[11]. A comparative study of relational database and key-value database for big data applications[12], Performance analysis of the resource loading time[13], etc.

2. Testing methodology and analysis

The research type used is quantitative, data is collected, recorded, compiled, and presented in the form of a frequency table where statistical values are measured to prove the correctness of the theory. The data used in the study were 1,050,000 records using software tools that were generated automatically with a predetermined number.

2.1. Metric Parameters

Performance measurement test parameters, ie.:

- Response time, MariaDB and PostgreSQL DBMS response time when executing a query in seconds
- Query relations, Inner join query which uses 1 relation, 2 relations, and 3 relations
- Data records, the data count has multiples of 50,000 starting with 50,000 records and ending up to 1,050,000 records

2.2. Analysis process and design

An overview of the testing phase and design of the DBMS research performance analysis process can be seen in Figure 1.

2.2.1. Preparations, this stage is the work of preparing the tools and equipment, the tools and equipment used are presented in Table 1.

Table 1. Equipment and Tools.

Software	Hardware
WAMP Server 3.2.0	Intel® Core™ i3-2348M 2.3GHz
HeidiSQL 10.2.0.5599	NVIDIA® GeForce® GT710M with 1 GB Dedicated VRAM
DBGenerator	RAM 8 GB DDR3 Memory
pgAdmin4	Harddisk 500 GB HDD

2.3. Create: Input record and DBMS. Databases are created on both MariaDB and PostgreSQL Database Management Systems (DBMS) in accordance with the design or Entity Relationship Diagram (ERD), entered (filled) 1,500,000 records.

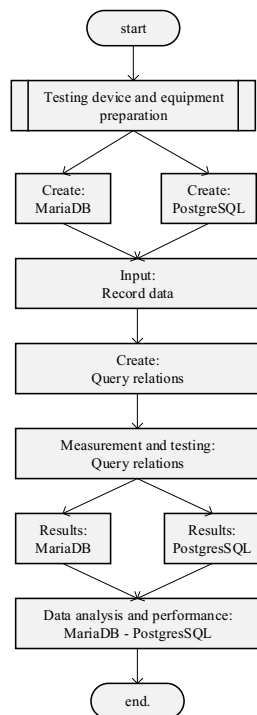


Figure 1. Analysis process design of testing.

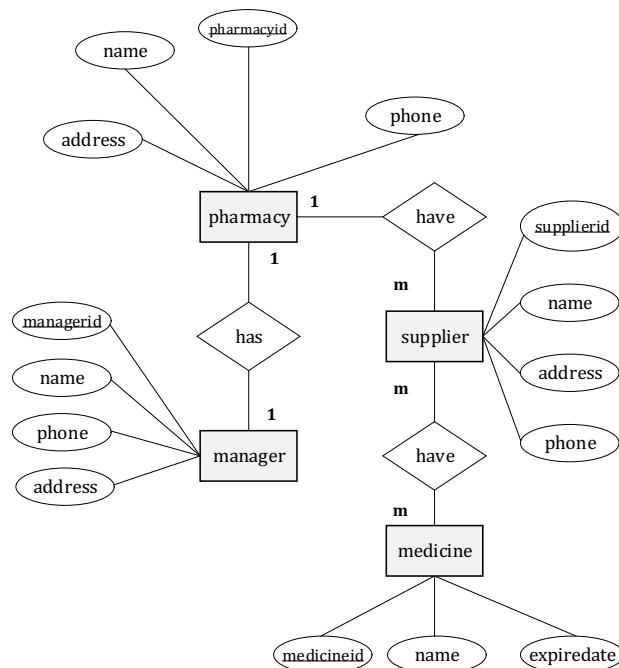


Figure 2. Entity relational diagrams.

2.3.1. Create: Query relations

The testing data used a pharmacy database which consists of 3 tables along with their attributes such as id, name, address, telephone, and others. As shown in Figure 2.

- Inner Join Query 1 Relation

Measurement of the join query 1 relation is applied to the manager table and pharmacy table with the query script.

```

SELECT manager.manager_name AS manager_name, pharmacy.name AS
pharmacy_name, pharmacy.address AS pharmacy_addres FROM manager
INNER JOIN pharmacy ON manager.managerid = pharmacy.managerid;
  
```

- Inner Join Query 2 Relation

Measurement of the join query 2 relation is applied to the manager table, pharmacy table and supplier table with the query code:

```

SELECT manager.name AS manager_name, pharmacy.name AS
pharmacy_name, supplier.name AS supplier_name FROM manager INNER
JOIN pharmacy ON manager.managerid = pharmacy.managerid INNER JOIN
supplier ON pharmacy.pharmacyid = supplier.pharmacyid;
  
```

- Inner Join Query 3 Relation

Measurement of the join query 3 relation is applied to the pharmacy table and supplier table with the query code:

```
SELECT manager.name AS manager_name, pharmacy.name AS
pharmacy_name, supplier.name AS supplier_name, medicine.name AS
medicine_name FROM manager INNER JOIN pharmacy ON manager.managerID
= pharmacy.pharmacyid INNER JOIN Supplier ON pharmacy.pharmacyid =
supplier.supplier_id INNER JOIN medicine_supplier ON
supplier.supplierid = medicine_supplier.supplierid INNER JOIN
medicine ON medicine_supplier.medicineid = medicine.medicineid;
```

2.4. Data analysis methods

Homogeneity testing technique for two independent variables with F-test to determine the homogeneity of the two groups of variances. The T-test used to determine the effect between MariaDB and PostgreSQL response times. Two-tailed T test (two tail) at a significant level of alpha 0.05, with hypothesis results H0 for indicate "there is no significant difference, and H1 there is significant difference. The way to prove it is if the value of t count > T table then H0 is rejected and H1 is accepted, whereas if T table > than T count then H0 is accepted H1 is rejected. The equation used from [14],[15] ie:

$$F\text{-Test}[14] = \frac{R^2/(n-1)}{(1-R^2)/(n-k)} \quad (1)$$

$$T\text{-Test}[15] = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{n_1+n_2-2} \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}} \quad (2)$$

3. Results

3.1. Response time inner join query 1 relations

The PostgreSQL and MariaDB query relations has 4 tables, ie pharmacies, managers, suppliers, and medicine. The total number of records in the pharmacy table is 1,050,000, the manager's table 1,050,000 records, 1,050,000 supplier records, and the medicine table 1,050,000 records. the measurement results can be presented in Table 2.

Table 2. Inner join query 1 relations.

Records	Response time(s)	
	MariaDB	PostgreSQL
50.000	0,569	0,078
100.000	1,134	0,153
200.000	2,269	0,297
300.000	3,381	0,447
400.000	4,494	0,581
500.000	5,660	1,140
600.000	6,772	1,275
650.000	7,375	1,319
700.000	7,912	1,385
750.000	8,488	1,429
800.000	9,053	1,506
850.000	9,625	1,578
900.000	10,172	1,647
950.000	10,687	1,709
1.000.000	11,275	1,772
1.050.000	11,806	1,844
Average	6,211	1,006

Table 2 shows the average MariaDB response time is 6,211 seconds, while the PostgreSQL average response time is 1,006 seconds. And it can be concluded that the response time of the MariaDB and PostgreSQL DBMS using the inner join query on 1 join relation in the manager - pharmacy table shows a difference. The initial test of MariaDB and PostgreSQL response times did not have a large enough difference, namely a difference of about 0.5 seconds, but after entering the number of 100,000 records, the difference between the two began to look bigger, where at the response time from MariaDB the greater the number of data tested, the greater response time generated. different from PostgreSQL is the increase in response time is not large.

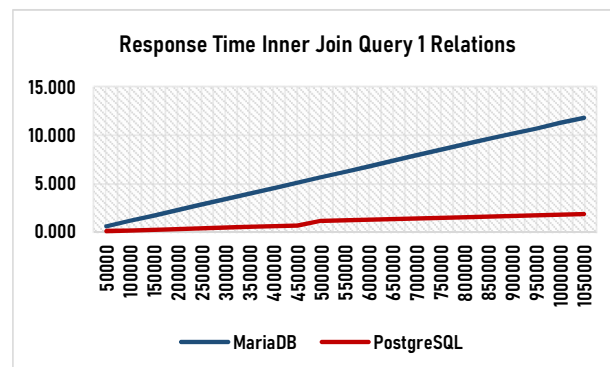


Figure 3. Response time inner join query 1 relations

Figure 3, it is found that MariaDB's response time is much greater than that of PostgreSQL, the time increase in PostgreSQL testing is quite normal, which is around 0.1 second. At PostgreSQL, when testing the number of data from 400,000 to 500,000, there was an increase in response time of about 0.5 seconds but after that the increase in response time was stable at 0.1 second. Meanwhile, the response time increase in MariaDB is very stable, which is around 0.6 seconds. The graph above is based on time analysis in the test table query 1 join relation in the manager and pharmacy table.

3.2. Response time inner Join query 2 relations

The testing the 2 join relation query results of the manager table, pharmacy table, and supplier table. The results of testing the inner join query seeing in Table 3.

Table 3. Inner join query 2 relations.

Records	Response time(s)	
	MariaDB	PostgreSQL
50.000	1,809	0,431
100.000	3,584	0,678
200.000	7,194	1,441
300.000	10,765	1,869
400.000	14,344	2,041
500.000	16,122	2,135
600.000	19,750	2,338
700.000	23,369	2,528
800.000	26,947	2,741
900.000	30,531	2,913
1.000.000	34,153	3,119
1.050.000	35,915	3,219
Average	37,569	3,325

Table 4 shows the average MariaDB response time is 19.747 seconds, while the PostgreSQL average response time is 2.226 seconds. The results of the query inner join 2 join relations in the table manager and pharmacy-supplier show a difference. Initial testing of MariaDB and PostgreSQL response time data showed a difference, where in the MariaDB response time, the greater the amount of data tested, the greater the response time generated with an increase of about 1.7 seconds in each test. different from PostgreSQL which has a stable response time increase of about 0.1 second.

Figure 3 shows that MariaDB's response time is much larger than that of PostgreSQL, proving that there is a significant difference. The response time increase in MariaDB is very stable, which is around 1.7 seconds, so it takes 37.569 seconds to display 1,050,000 data. In PostgreSQL, there was an increase in response time of about 0.1 seconds for each test, except for testing the number of data from 250,000 to 300,000 an increase of 0.3 seconds so that displaying 1,050,000 data took 3,325 seconds. Figure 4 is based on time analysis on query test table 2, combining relationships in the manager, pharmacy and supplier table.

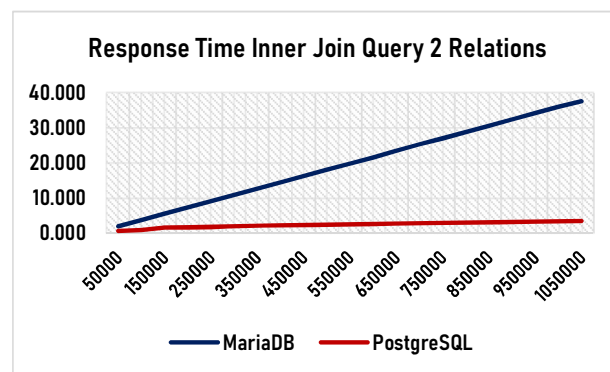


Figure 4. Response time inner join query 2 relations

3.3. Response time inner Join query 3 relations

The testing the 3 join relation query results of the manager table, pharmacy table, supplier table, and medicine table. The results of testing the inner join query seeing in Table 4.

Table 4. Inner join query 3 relations.

Records	Response time(s)	
	MariaDB	PostgreSQL
50.000	1,809	0,431
100.000	3,584	0,678
200.000	7,194	1,441
300.000	10,765	1,869
400.000	14,344	2,041
500.000	16,122	2,135
600.000	19,750	2,338
700.000	23,369	2,528
800.000	26,947	2,741
900.000	30,531	2,913
1.000.000	34,153	3,119
1.050.000	35,915	3,219
Average	37,569	3,325

Table 4 shows the average MariaDB response time is 38.176 seconds, while the average PostgreSQL response time is 3.646 seconds, that the MariaDB and PostgreSQL DBMS response times use the 3 inner join query join relation in the manager - pharmacy - supplier - drug table. show that difference. From the beginning of testing the MariaDB and PostgreSQL response time data, there is a difference,

where at the MariaDB response time the greater the amount of data tested, the greater the resulting response time with an increase of about 3.5 seconds in each test. different from PostgreSQL which has a stable response time increase of about 0.1 seconds

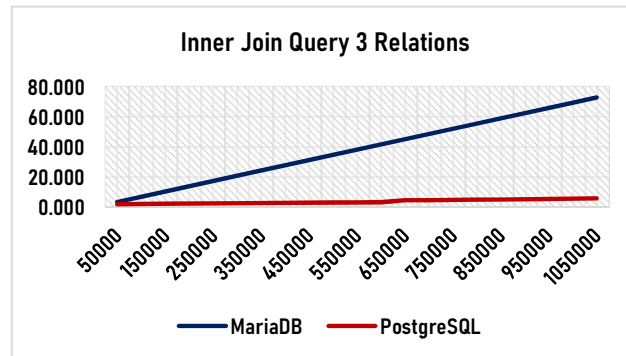


Figure 5. Response time inner join query 3 relations

Figure 5 obtained MariaDB response time is much greater than PostgreSQL, proving that there is a significant difference. The increase in response time on MariaDB is very stable, which is about 3.5s, so it takes 73.053s to display 1,050,000 records. In PostgreSQL, the increase in response time was around 0.1 second in each test, except for testing the amount of data from 600,000 records to 650,000 records, there was an increase of 1.4s so that to display 1,050,000 record of 5.722s.

4. Test Results Analysis

Based on the response time obtained from all query results, a comparative test is needed to determine whether the response time between MariaDB and PostgreSQL has a significant difference or there is no significant difference. This analysis was performed by calculating the F-Test and T-Test using SPSS statistical analysis.

- The testing results of the effect of Linear Regression using the F-Test

Table 5. F-Test MariaDB linear regression results

ANOVA	MariaDB linear regression results				
	df	SS	ms	F	Significance F
Regression	2	27820,9607	13910,48	232,74375	1,549972E
Residual	123	7351,3856	59,76736		
Total	125	35172,3463			

Table-5, shows that the value of Fcount = 232.7437541. Compare with the F-count value with the Ftable value, where Fcount > Ftable (232.7437541 > 3.07) then H0 is rejected, meaning that there is a significant effect of the amount of data and the join relation on MariaDB response time.

Table 6: F-Test PostgreSQL linear regression results

ANOVA	PostgreSQL linear regression results				
	df	SS	ms	F	Significance F
Regression	2	164,00213	82,00107	287,406393	4,35733E
Residual	123	35,093621	0,285314		
Total	125	199,09575			

Table-6 shows that the value of Fcount = 287.4063929. Compare it with the F-count value with the F-Table value, where Fcount > Ftable (287.4063929 > 3.07) then H0 is rejected, meaning that there is a significant effect of the amount of data and the join relation on the response time of PostgreSQL

- The T-test results of the inner join query 1 relations

Table 7. T-test inner join query 1 relations results

Metrics	Inner Join MariaDB	Inner Join PostgreSQL
Mean	6,210714286	1,006104762
Variance	12,23503558	0,364887426
Observations	21	21
df	21	
t Stat	6,719142533	
P(T<=t) two-tail	1,19714E-06	
t Critical two-tail	2,079613845	

Table 7 that the t-Stat value (tcount) is greater than the two sides t-Critical (T-table) and the p-value is more (P (T <= t) two sides) is smaller than alpha which there is a significant difference in the comparison of the results of 1 inner joint relationship in the manager table - pharmacies using MariaDB and PostgreSQL are accepted. This means that there is a significant difference between the results of the comparison of 1 joint relation on the manager - pharmacy table using MariaDB and PostgreSQL.

- The F-test results of the inner join query 2 relations

Table 8. T-test Inner join query 2 relations results

Metrics	Inner Join MariaDB	Inner Join PostgreSQL
Mean	19,74699048	2,226047619
Variance	124,0674324	0,637683736
Observations	21	21
df	20	
t Stat	7,189935345	
P(T<=t) two-tail	5,8219E-07	
t Critical two-tail	2,085963447	

Table 8 shows that the t Stat value (tcount) is greater than the two-tail tCritical (ttable) and the p-value is more (P (T <= t) two-tail) is less than the alpha value that H0 is rejected and H1 is accepted. This means that there is a significant difference between the results of the comparison of 2 joint relations in the manager-pharmacy-supplier table using MariaDB and PostgreSQL DBMS.

- The F-test results of the inner join query 3 relations

Table 9: T-test Inner join query 3 relations results

Metrics	Inner Join MariaDB	Inner Join PostgreSQL
Mean	38,24646667	3,646285714
Variance	465,1474032	1,777683438
Observations	21	21
df	20	
t Stat	7,337775383	
P(T<=t) two-tail	4,31248E-07	

t Critical two-tail

2,085963447

Table 9 shows that the t Stat value (tcount) is greater than the two-tail tCritical (Ttable) and the p-value is more (P (T ≤ t) two-tail) is less than the alpha value that H0 is rejected and H1 is accepted. This means that there is a significant difference between the results of the comparison of 3 joint relations in the manager-pharmacy-supplier-medicine table using MariaDB and PostgreSQL DBMS.

5. Discussion

Response time series measurement processes between MariaDB and PostgreSQL with the same table structure and number of records, namely the number of records of 1,050,000. Testing of MariaDB and PostgreSQL with testing techniques 1 join, 2 join, and 3 join relations, which is done with 21 experiments using the amount of data with multiples of 50,000 starting with 50,000 up to 1,050,000 records.

The test results show that the calculation of the average response time at 1 join relation for MariaDB is 6.211s and PostgreSQL is 1.006s. The 2 join relations for MariaDB are 19,747s and PostgreSQL are 2,226s. and 3 join relations for MariaDB is 38,246s and PostgreSQL is 3,646s, and it can be concluded that in testing 1 join relation, 2 join relations, and 3 join relations, it shows that PostgreSQL response time is less than MariaDB, it proves that PostgreSQL response time is faster than MariaDB. From the average range between MariaDB response time and in PostgreSQL the join relation 2 shows the least range when compared to join relation 1 and join relation 3.

The effect value between the amount of data and the joint relation on the response time of regression analysis testing using the f-test, which is based on the test results on the MariaDB DBMS because Fcount > Ftable or 232.7437541 > 3.07 then H0 is rejected. That is, there is a significant effect of the amount of data and join relations on the response time on the MariaDB DBMS. Because Fcount > Ftable or 287.4063929 > 3.07, H0 is rejected. That is, there is a significant effect of the amount of data and join relations on the response time in the PostgreSQL DBMS. The purpose of the test results above is that the variable number of data and join relations has an influence on the response time results from MariaDB and PostgreSQL.

6. Conclusions

PostgreSQL vs MySQL when choosing between these two database management systems, sometimes we have to determine which one is better first. Some developers often fall into the trap of familiarity and comfort in determining the database management system. A good developer should always make an informed decision among the options, benefits and drawbacks.

Based on the results of the study, it is concluded that the comparison of the response time tested using the inner join query with the testing technique of 1 join relation, 2 join relations, and 3 join relations, in which each test technique is tested 21 times the amount of data which has a multiple of 50,000 and starts with 50,000 up to 1,050,000 records on all tables in MariaDB and PostgreSQL DBMS. The calculation time or query speed tends to change at each test session, but does not change the position of the fastest query and the longest access time.

From the results of the regression analysis for testing the effect using the F-test, it shows that there is a significant effect of the amount of data and the joint relation on the response time in the MariaDB and PostgreSQL DBMS. The results of the t-test comparative test analysis compared the response time between the MariaDB and PostgreSQL DBMS using data from the test 1 join relation, 2 join relations, and 3 join relations. From the results of the analysis it is concluded that in testing 1 join relation, 2 join relations, and 3 join relations show a significant difference between MariaDB and PostgreSQL response times. Based on the results of the t-test above, there is a significant difference between the results of MariaDB and PostgreSQL response time, where the PostgreSQL response time results are faster than MariaDB, in contrast to MariaDB which has slower response time results than PostgreSQL. The research findings refer to the www.enterprisedb.com report[16] showing that PostgreSQL database performance is superior for enterprise applications, both new applications are developing. Compared to MariaDB, EDB Postgres offers a richer physical design option that meets the performance needs of a wider range of enterprise applications[16].

Future studies will focus on performance optimization with a variety of DBMS based on an object-oriented framework, the development and competition of free-open source services is a motivation to continue creating and developing.

7. References

- [1] Andlinger P and Matthias G 2019 PostgreSQL is the DBMS of the Year 2018 *db-engines.com*
- [2] Gelbmann M 2020 MariaDB strengthens its position in the open source RDBMS 5 2018
- [3] DB-Engines 2020 DB-Engines Ranking
- [4] Kroc K, Kizun O and Skublewska-Paszowska M 2020 Performance analysis of relational databases MySQL, PostgreSQL, MariaDB and H2 *Journal of Computer Sciences Institute*
- [5] Haviluddin H, Budiman E and Hidayat N F 2019 A database integrated system based on SOAP web service *TEM Journal*
- [6] Jung M G, Youn S A, Bae J and Choi Y L 2016 A study on data input and output performance comparison of MongoDB and PostgreSQL in the big data environment *Proceedings - 8th International Conference on Database Theory and Application, DTA 2015*
- [7] Budiman E, Tejawati A and Hairah U 2020 Bioinformatics Database Query Performance and Optimization *International Journal of Recent Technology and Engineering* **9** 581–8
- [8] Lee D-H, Song M-C, Cho Y-T and Kim S-W 2018 Comparison of performance between MariaDB and PostgreSQL in terms of CPU overhead *Proceedings of the Korea Information Processing Society Conference* 297–9
- [9] Budiman E, Puspitasari N, Wati M, Widians J A and Haviluddin 2019 Web Performance Optimization Techniques for Biodiversity Resource Portal *Journal of Physics: Conference Series* 1230
- [10] Budiman E, Puspitasari N, Haerullah, Jamil M, Wati M and Saudek A 2018 Evaluation of the bioinformatics resource portal *Proceedings - 2018 3rd International Conference on Information Technology, Information Systems and Electrical Engineering, ICITISEE 2018* (IEEE) pp 54–9
- [11] Tongkaw S and Tongkaw A 2017 A comparison of database performance of MariaDB and MySQL with OLTP workload *ICOS 2016 - 2016 IEEE Conference on Open Systems*
- [12] Puangsaijai W and Puntheeranurak S 2017 A comparative study of relational database and key-value database for big data applications *2017 International Electrical Engineering Congress, iEECON 2017*
- [13] Budiman E, Puspitasari N, Alam S N, Akbar T M A, Haeruddin and Indra D 2018 Performance analysis of the resource loading time for borneo biodiversity information system *Proceedings of the 3rd International Conference on Informatics and Computing, ICIC 2018*
- [14] Schumacker R, Tomek S, Schumacker R and Tomek S 2013 F-Test *Understanding Statistics Using R*
- [15] Brown S and Melamed L 2012 T Test *Experimental Design and Analysis*
- [16] Peng H 2020 Comparative analysis *Springer Geography* 135–59