International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 07 Issue: 10 | Oct 2020 www.irjet.net p-ISSN: 2395-0072

SQL Inner Join: MySQL and PostgreSQL Performance

Ummul Hairah

Informatics Departments, Universitas Mulawarman
Samarinda 75119, East-Kalimantan, Indonesia

Abstract - *PostgreSQL and MySQL are two of the most popular open-source relational databases today. While they share many things in common, the differences between them are significant and can be a source of confusion both for newcomers to database management and for experienced DBAs. In this article, we offer a detailed comparison of these two database systems and review response time performance. It is intended for anyone interested in learning more about open-source databases but is invaluable for those looking to decide which database system is the right choice for their organization or application. We discussed response time, testing it with statistical analysis via the F-test and T-test. After this point-by-point analysis, we conclude that PostgreSQL is the superior choice based on a much faster PostgreSQL response time than MySQL, strengthened by the F-test which shows the effect of the amount of data and the number of relations on response time and comparative test testing. T-test which shows that there is a significant difference in the response time of the two DBMS in each relation.*

Key Words: mysql, postgresql, peroformance, inner-joint

1. INTRODUCTION

DB Engine Ranking[1] which was released in October 2020 recently again ranked MySQL in second place, followed by Microsoft SQL Server in third and PostgreSQL in fourth, while Oracle was in the first place. The DB engine score ranking is based on the current popularity ranking of the measurement system using the parameters of the number of results in search engine queries (Google and Bing), general interest in the system (search frequency on Google Trends), technical frequency of discussion about the system (number of questions interested in sites Stack Overflow and DBA Stack Exchange), number of job offers from job search engines (Indeed and Simply Hired), Number of profiles on professional networks (LinkedIn), and relevance on social networks (number of Twitter tweets) where the system is mentioned[1].

This study aims to measure the performance of MySQL and PostgreSQL against SQL relation parameters (inner join) and response time, these two DB engines are the study focus because they have similarities, MySQL and PostgreSQL are open-source licensed RDBMS, widely used and with the concept of ACID transactions. In addition, it's research also finds out how much the relationship between the two DB engines (MySQL vs PostgreSQL relationship) toward performance (response time) using Parametric statistical analysis: Homogeneity variance test Analysis with F-Test and T-Test for Comparative test analysis.

Study motivations: its intended for anyone interested in learning more about open-source databases but is invaluable for those looking to decide which database system is the right choice for their organization or application

PostgreSQL and MySQL are two of the most popular open-source relational databases today. While they share many things in common, the differences between them are significant and can be a source of confusion both for newcomers to database management and for experienced DBAs. MySql is an Oracle developer while PostgreSQL is developed by PostgreSQL Global Development Group.

Previous researchers have worked on comparisons of Mysql vs Postgresql performance, such as P Kathuria who tested the performance test with the Laravel API for Eloquent simple queries on 1 million items using Jmeter[2]. Andjelic et al the performed testing is the query (select, insert, delete and order by) execution times were measured for both DBMS[3], Database Query Performance and Optimization[4], Yang Xiaojie the research is empirically based on the analysis and comparison of statistical data collected from Internet authorities and reference manuals, the results of individual experiments conducted on different platforms in identifying the differences between MySQL and PostgreSQL[5], Database integrated[6], Benchmarking PostgreSQL vs. MySQL performance using Drupal 5.x[7], etc.

2. TESTING METHODOLOGIES AND ANALYSIS

An overview of the test results and analysis process design of the research performance DBMS is seen in Chart-1.

2.1 Equipment and Tools

Table -1: Equipment and Tools

Software	Harware
WAMP Server 3.0	Intel® Core™ i7-4712MQ 2.30 GHz
HeidiSQL 10.2.0.5599	NVIDIA® GeForce® 820M with 2 GB Dedicated VRAM
DBGenerator	RAM 8 GB DDR3 Memory
pgAdmin4	Harddisk 1 TB HDD

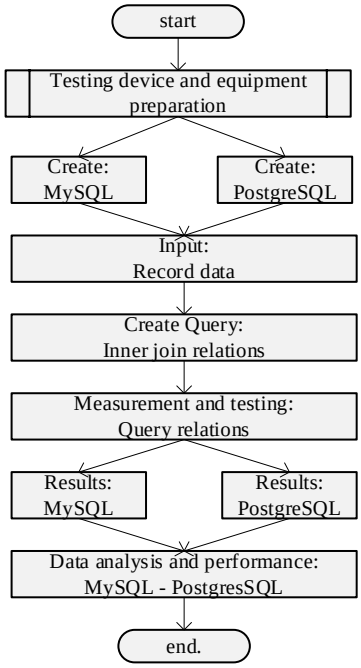

Chart -1: Flowchart analysis process design

2.1.1 Create: Input record and DBMS

Databases are created on both MySQL and PostgreSQL in accordance with the design design or Entity Relationship Diagram (ERD), entered (filled) 1,500,000 records.

2.2.3. 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 Fig.-1.

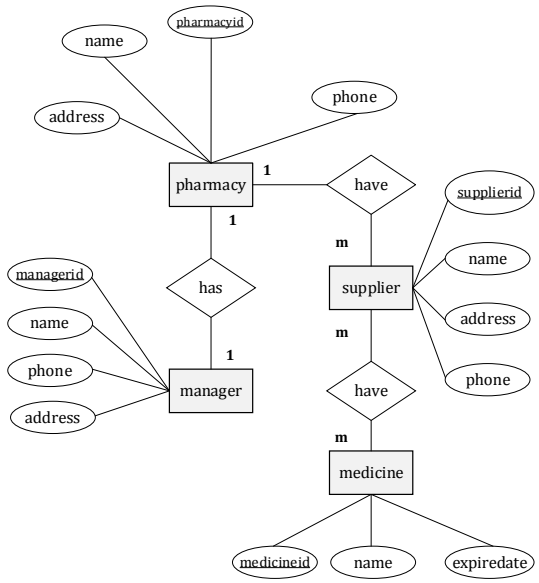

Fig -1: Entity Relation Diagram

The relationship diagram in Fig.-1 for measurements in MySQL and PostgreSQL has 4 tables, namely pharmacies, managers, suppliers, and medicine. The number of records in the pharmacy table is 1,050,000, manager table 1,050,000 records, supplier 1,050,000 records, and medicine table 1,050,000 records.

2.2 Measurement Variables and Parameters

Table -1: Variable and parameters

Variable	Parameters
Response time	Intel® Core™ i3-2348M 2.3GHz
Join relations	The large number of join queries used (Inner Join Query)
Record data	50.000 up to 1.050.000
DBMS	MySQL and PostgreSQL

2.3 Data Collection Methods

The data collection technique uses DBGenerator tools, with a pharmacy database from 50,000 records to 1,050,000 records that are generated automatically. Then we tested the response times of 3 inner union query relations on both Database Management Systems (MySQL and PostgreSQL).

2.4 Data Analysis Methods

The test results analysis using the F-test and T-test, F-test or the relationship model test in this study to see how all the independent variables influence the dependent variable together, or to test whether the regression model we make is good or significant, or non-significant[8]. If the model is significant, the model can be used for prediction[9] or forecasting, on the other hand, if it is non-significant, the regression model cannot be used for forecasting[10].

T-test or partial test, which is to test how the influence of each independent variable individually on the dependent variable.

2.5 Testing Method

This study was tested using multiple regression analysis using SPSS. In multiple regression analysis, several independent variables are used to predict the value of the dependent variable. The formulation for the multiple linear regression equation is as follows[11], [12]:

$$Y = \beta 0 + \beta 1X1 + \beta 2X2 + e \quad (1)$$

Where: Y = variable y, β0 = Constant β1, β2 = Multiple regression coefficient, x1 = Variable x1, x2 = Variable x2, e = Confounding variable (error).

Hypothesis testing is used to determine whether the independent variables simultaneously have a significant effect on the dependent variable. The degree of confidence

used is 5%, if the F-count > F-table, then the hypothesis states that all independent variables simultaneously have a significant effect on the dependent variable[13], [14].

$$F = \frac{R^2/(k-1)}{1-R^2/(n-k)} \quad (2)$$

Basic decision making:

If probability (significance)> 0.05 (α) or F count <F table means that the hypothesis is not proven, then H0 is accepted. Ha is rejected if done simultaneously[15].

If the probability (significance) <0.05 (α) or F count> F table means that the hypothesis is proven, then H0 is rejected and Ha is accepted if done simultaneously.

3. MEASUREMENT AND ANALYSIS RESULTS

The response time testing process uses HeidiSQL 10.2. software which connects to the two DBMS servers tested in this study.

3.1 The response time result for query 1 join relation

The results of testing query 1 join relationship between the manager table and the pharmacy table. The results of the query testing are presented in the **Table-2**.

Table -2: Response time result for query 1 join relation

test	Record	Response time (second)	
		MySQL	PostgreSQL
1	50000	0,6528	0,0626
2	100000	1,328	0,1092
4	200000	2,6402	0,2186
6	300000	3,9438	0,3188
7	350000	4,6064	0,3782
8	400000	5,2376	0,431
9	450000	5,9342	0,481
10	500000	6,5748	0,5374
11	550000	7,1748	0,5904
12	600000	7,8374	0,7816
13	650000	8,5592	0,847
14	700000	9,2218	0,9128
15	750000	9,875	0,978
16	800000	10,484	1,047
17	850000	11,1282	1,1064
18	900000	11,8564	1,181
19	950000	12,4252	2,1812
20	1000000	13,1248	2,2592
21	1050000	13,7442	2,3844
Average		7,218228571	0,820942857

Table-2 shows the average MySQL response time is 7.218228571 while the PostgreSQL average response time is 0.820942857. MySQL has linear increase response time as shown in Chart-2.

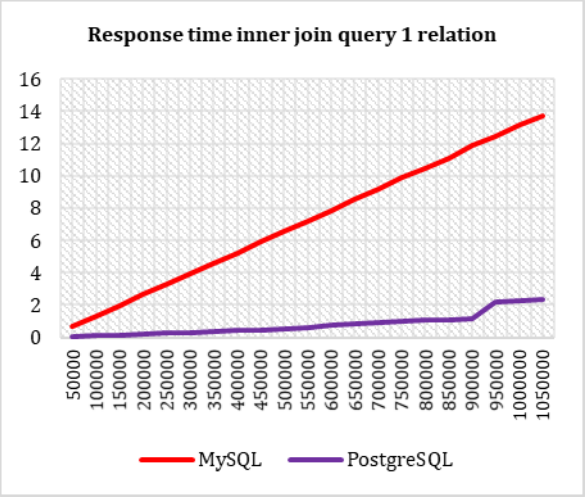

Chart -2: Response time inner join query 1

3.2 The response time result for query 2 join relation

The results of testing query 2 join relationship between manager table, pharmacy table, and supplier table. The results of the query testing are presented in the **Table-3.**

Table -3: Response time result for query 2 join relation

test	Record	Response time (second)	
		MySQL	PostgreSQL
1	50000	1,4874	0,2002
2	100000	3,0158	0,281
4	200000	4,4878	0,7064
6	300000	5,925	0,7376
7	350000	7,506	0,7782
8	400000	8,9624	1,3
9	450000	10,3408	1,3688
10	500000	11,8282	1,453
11	550000	13,4466	1,5282
12	600000	14,8692	1,6126
13	650000	16,5032	1,6778
14	700000	17,8718	1,7688
15	750000	19,4656	1,8374
16	800000	21,0906	1,9312
17	850000	22,5068	2,0126
18	900000	23,9658	2,0906
19	950000	25,328	2,1656
20	1000000	27,0966	2,2436
21	1050000	28,1532	2,3004
Average		7,218228571	29,7782

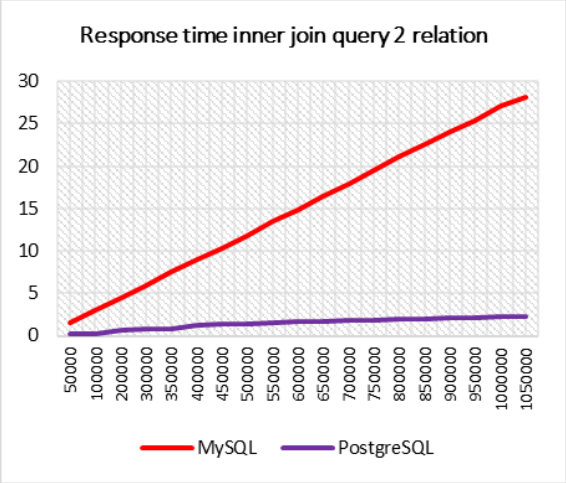

Chart -3: Response time inner join query 2

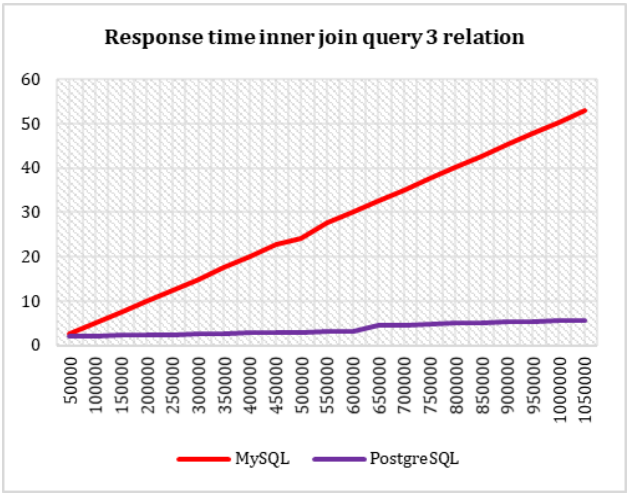

Chart-4: Response time inner join query 3

Table 3 and Chart-3 shows the average MySQL response time is 16.43040952 seconds while the PostgreSQL average response time is 1.569514286 seconds.

3.3 The response time result for query 3 join relation

The results of testing query 3 join relationship between the manager table, pharmacy table, supplier table, and medicine table. The query test results are presented in Table -4.

Table -4: Response time result for query 2 join relation

test	Record	Response time (second)	
		MySQL	PostgreSQL
1	50000	1,4874	0,2002
2	100000	3,0158	0,281
4	200000	4,4878	0,7064
6	300000	5,925	0,7376
7	350000	7,506	0,7782
8	400000	8,9624	1,3
9	450000	10,3408	1,3688
10	500000	11,8282	1,453
11	550000	13,4466	1,5282
12	600000	14,8692	1,6126
13	650000	16,5032	1,6778
14	700000	17,8718	1,7688
15	750000	19,4656	1,8374
16	800000	21,0906	1,9312
17	850000	22,5068	2,0126
18	900000	23,9658	2,0906
19	950000	25,328	2,1656
20	1000000	27,0966	2,2436
21	1050000	28,1532	2,3004
Average		7,218228571	29,7782

Table 4 and Chart-4 shows the average MySQL response time is 7,218228571 seconds while the PostgreSQL average response time is 29,7782 seconds

The test results for query 1, 2 and 3 join relationship show that the difference between MySQL and PostgreSQL response times is getting bigger. The greater the amount of data being tested, the greater the resulting response time.

3.4 Results test analysis

Based on the response time test results, then a regression model test is carried out to determine whether the response time has a significant difference or there is no difference. This analysis is performed using equation (2) to calculate the F-test and the T-test.

Table -5: T-test results of 1, 2, and 3 Inner Join Relations

	1 Join		2 Join		3 Join	
	MySQL	Postgre-SQL	MySQL	Postgre-SQL	MySQL	Postgre SQL
Mean	14.76	1.483	21.6	2.161	27.555	3.646
Variance	69.45	0.331	148.6	0.667	243.77	1.778
Observations	21	21	21	21	21	21
Df	20		20		20	
t Stat	7.285		7.30		6.992	
t Critical two-tail	2.086		2.08		2.086	

- F-test to determine the effect of linear regression on MySQL where the value of Ftable = F (0.05; 2; 3.07). Because Fcount > Ftable or 342.572 > 3.07, Ho is rejected, meaning that there is a significant effect on the amount of data and relations on the response time in MySQL DBMS.

- F-test to determine the effect of linear regression on PostgreSQL where the value of F table = F (0.05; 2; 3.07). Because Fcount > Ftable or 266.0716 > 3.07, Ho is rejected, meaning that there is a significant effect on the amount of data and relations on the response time in PostgreSQL DBMS.
- T-test value (T-count) for query 1 inner join of the relation obtained a value of 6.90, these results indicate that there is a significant difference in the results of the comparison of 1 joint relations in the pharmacy-supplier table using MySQL and PostgreSQL DBMS.
- T-test value (T-count) for query 2 inner join of the relation obtained a value of 7.32, these results indicate that there is a significant difference in the results of the comparison of 1 joint relations in the pharmacy-supplier table using MySQL and PostgreSQL DBMS.
- T-test value (T-count) for query 3 inner join of the relation obtained a value of 6.99, these results indicate that there is a significant difference in the results of the comparison of 3 joint relations in the manager, pharmacy, supplier and medicine table using MySQL and PostgreSQL DBMS.

4. DISCUSSION

Response time series measurement processes between MySQL and PostgreSQL with the same table structure and number of records, ie the number of records of 1,050,000, because if a test is carried out with a small number of records, the value cannot be clearly measured and the differences in each. Testing of two DBMS using inner join queries, MySQL and PostgreSQL with testing techniques 1 join relationship, 2 join relations, and 3 join relations, which is done with 21 experiments using the amount of data with multiples of 50,000 starting with 50,000 and ending. up to 1,050,000 records.

The results of the inner join query relationship test show that the calculation results of the average response time for 1 join relation in the pharmacy-manager table for MySQL is 7,218 second and PostgreSQL is 0.820. Second inner join of 2 relations in the pharmacy-supplier-drug table for MySQL is 21,624 second and PostgreSQL is 1,569 second, and 3 join relations in the manager-pharmacy-supplier-drug table for MySQL is 27.554 second and PostgreSQL is 3.030 second. These results show that testing 1 join relation, 2 join relations and 3 join relations for PostgreSQL response time than MySQL results obtained PostgreSQL response time results are faster than MySQL.

In the analysis of the linear regression analysis for the MySQL f-test, the value of Fcount = 342.572 and significant F = 5.231 so that if it is significant <α with α = 0.05, then Ho is rejected. So that in MySQL there is a significant effect of the amount of data and relations on the response time value. Meanwhile, at PostgreSQL. It is known that the value of Fcount = 266.071 and significant F = 2.11 so that if it is significant <α with α = 0.05 then Ho is rejected. So that PostgreSQL there is a significant effect of the amount of data and relations on the response time value.

5. CONCLUSIONS

MySQL vs PostgreSQL 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 MySQL and PostgreSQL. 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 MySQL and PostgreSQL. The results of the t-test comparative test analysis compared the response time between the MySQL and PostgreSQL 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 MySQL and PostgreSQL response times. Based on the results of the t-test above, there is a significant difference between the results of MySQL and PostgreSQL response time, where the PostgreSQL response time results are faster than MySQL.

REFERENCES

[1] DB-Engines. DB-Engines Ranking. 2020.

[2] Kathuria P. Mysql vs Postgresql performance test with Laravel API for simple Eloquent queries on 1 million items. May 3, 2019 2019.

[3] Andjelic S, Obradovic S, Gacesa B. A performance analysis of the DBMS - MySQL vs PostgreSQL. Komunikacie 2008.

[4] Budiman E, Tejawati A, Hairah U. Bioinformatics Database Query Performance and Optimization. International Journal of Recent Technology and Engineering 2020; **9**(3): 581–588. DOI: 10.35940/ijrte.c4666.099320.

[5] Xiaojie Y. Analysis of DBMS: MySQL Vs PostgreSQL. 2011.

[6] Haviluddin H, Budiman E, Hidayat NF. A database integrated system based on SOAP web service. TEM Journal 2019; **8**(3). DOI: 10.18421/TEM83-12.

[7] 2bits.com. Benchmarking PostgreSQL vs. MySQL performance using Drupal 5.x. 2bits.com.

[8] Akbar MA, Yendra, Nursyamsi I, Budiman E. Role Analysis: Trust in Mediating Informal Learning towards Customer Behavior to use Electronic Banking. Journal of Physics: Conference Series 2019; **1230**: 012057. DOI: 10.1088/1742-6596/1230/1/012057.

[9] Budiman E, Haviluddin, Dengan N, Kridalaksana AH, Wati M, Purnawansyah. Performance of Decision Tree C4.5 Algorithm in Student Academic Evaluation., 2018. DOI: 10.1007/978-981-10-8276-4_36.

[10] Dengen N, Haviluddin, Andriyani L, Wati M, Budiman E, Alameka F. Medicine Stock Forecasting Using Least Square Method. In Proceedings - 2nd East Indonesia Conference on Computer and Information Technology: Internet of Things for Industry, EIConCIT 2018, 2018. DOI: 10.1109/EIConCIT.2018.8878563.

[11] Brown S, Melamed L. T Test. In Experimental Design and Analysis, 2012. DOI: 10.4135/9781412984218.n3.

[12] Budiman E, Parassa Y, Haerullah H, Moeis D, Soekarta R, Jamil M. The effect of mobile learning media on student learning motivation in data structure courses. In Proceedings - 2018 3rd International Conference on Information Technology, Information Systems and Electrical Engineering, ICITISEE 2018, 2018. DOI: 10.1109/ICITISEE.2018.8720963.

[13] Ibrahim MBH, Jufri MT, Alam SN, Zakaria, Akbar MA, Budiman E. Statistical Analysis of Performance Goals Effect to Lecturer Work Achievement in Higher Education. In Proceedings - 2nd East Indonesia Conference on Computer and Information Technology: Internet of Things for Industry, EIConCIT 2018, 2018. DOI: 10.1109/EIConCIT.2018.8878571.

[14] Schumacker R, Tomek S, Schumacker R, Tomek S. F-Test. In Understanding Statistics Using R, 2013. DOI: 10.1007/978-1-4614-6227-9_11.

[15] Budiman E, Parassa Y, Haerullah H, Moeis D, Soekarta R, Jamil M. The effect of mobile learning media on student learning motivation in data structure courses. In Proceedings - 2018 3rd International Conference on Information Technology, Information Systems and Electrical Engineering, ICITISEE 2018, IEEE, 2018; 60–64. DOI: 10.1109/ICITISEE.2018.8720963.

BIOGRAPHIES

Ummul Hairah, Department of Informatics, Mulawarman University, Samarinda, Indonesia. She is a member of IEEE and APTIKOM (Association Higher Institution Informatics and Computer Indonesia).