Comparative Analysis of Database Query Storage Performance Between Stored Procedure and Function

Abednego Dwi Septiadi[1] , Lee jeong-bae[2]

[1] Amikom University Purwokerto, Indonesia
[2] Sun Moon University, South Korea
[1] abednego@amikompurwokerto.ac.id*; [2] bg.kim@mpcl.sunmoon.ac.kr;
* corresponding author

Abstract

This research was conducted to measure the data storage time carried out by the DBMS on data that has been prepared with an increasing number of data, the data provided is consistent student data which will be stored with Stored Procedure and Function. This study uses the action research method which has 4 stages, starting with planning, action, observation and reflection. From the results of the experiments that have been carried out, it appears that Stored Procedure is able to outperform Function in data storage time. The data provided are 2 different types of data, each of which consists of 500, 2500, 4500 and 6500 data stages. This study also compares data storage that is differentiated by the computer between the data provider computer and the data storage computer or server, the result of which is the Stored Procedure. able to outperform Function in data storage speed.

Keywords: DMBS; Stored Procedure; Function; Query.

1. Introduction

Reliable data sources are data sources that are able to present quickly, precisely and accurately, this era data sources are not only required to present data but also must have fast and precise performance. Data sources that evolved into databases in this technological era must have reliable performance so that users feel they are being served maximally, both in terms of quality and quantity, a database which is a collection of data stores stored in magnetic disks, optical disks, magnetic drums and storage media. Other secondary media that can accommodate large amounts of data [1], must be capable of serving data and information capable of addressing the needs of users.

To produce a good data server, it is necessary to have a good planning and also proper design so that the data server is able to work properly and optimally [10]. One thing that must be considered in the structure of service providers is the use of the right method so that a method that has been chosen is not wrong in its implementation. The method that will be chosen will affect the performance and the results that will be obtained so that the selection of this method or method is quite a crucial part.

In the programming world, there are several ways to store data into a database, namely with functions, triggers and stored procedures [9]. Each of these methods is capable of storing data, but there are advantages and disadvantages. SQL function is a module that returns data via return clause instead of using out or in out arguments [2]. Stored Procedure is a procedure like a subprogram (subroutine) in a regular programming language that is stored in a database catalog [3]. Several advantages offered by stored procedures include: increased performance, reduced network traffic, reusability, and increased security control. Behind these advantages, stored procedures also have drawbacks. Among them are the potential to increase server load and writing is not easy (requires specific knowledge) [3].

SQL functions can also be divided into several types, including single-row functions, which are commands that can be used as tasks such as changing uppercase or lowercase characters or vice versa or cutting text on a line. Aggregate Function is also a function that is used to calculate a number such as Max, Min, Average and

so on. The other types are Number and Data Function which are free input variables that can convert into date and numberic outputs. Next are the General and Conditional Functions which allow to expand the capabilities of a simple SQL command into a traditional programming language construct, while the conditional function helps to determine between several options [3].

Stored Procedures and Functions can be found in DBMS (Database Management System) which is a system capable of defining, creating, maintaining and controlling access to databases [4]. DBMS is a tool that can make it easier for users to access a database, so that it can accelerate the creation of a structured database. DBMS also provides access control such as security, system integrity, concurrency, recovery and user accessible catalog which are the advantages offered by DBMS.

2. Literature Review

This study will use the action research method which is an activity and or corrective action for something that is planned, implemented, and evaluated systematically so that its validity and reliability reach the research level [5]. Action research is good to use to analyze comparisons because in each cycle planning will be carried out followed by action, from the action in the form of a speed test, it will be observed why the output can occur and followed by reflection which will direct the next steps, whether this research is sufficient to conclusions are drawn or it is necessary to carry out the next stage of testing.

According to Grundy and Kemmis [6] in the book Educational Research In Australia, action research has two main objectives, namely to increase (improve) and involve (involve). Improve means, improving the field of practice, increasing understanding of the practice carried out by practitioners, and improving the situation in which the practice is carried out. Whereas Involving [8] means, it involves related parties, if the action research is carried out in the school, the parties involved are among others, the principal, teachers, students, employees, and parents of students or in this case are the stakeholders concerned.
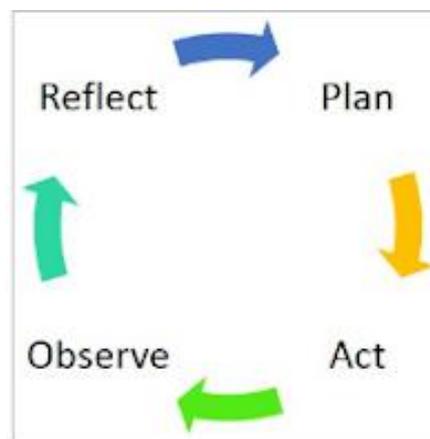

Fig. 1. Action Research

2.1. Plan

The first stage is planning, namely designing a database and compiling the data to be used. This stage requires a lot of time because the planning must be done in a detailed and structured manner so that the test design can run well.

2.2. Act

The next step after planning and designing, is followed by the testing process by entering a certain amount of data, starting with 500 data and continuing with multiples of 2000 data until enough output is produced to draw conclusions.

2.3. Observe

After the test produces results, an observation process is carried out on each test so that it can be concluded whether the test needs to be done again or enough [7]. when testing is needed again, the data will be added by 2000 so that the difference distance is more visible and so on.

2.4. Reflect

The next step is the reflection process, namely drawing a conclusion, whether this test needs to be done again or not, if necessary, planning will be done again, but if not then this research can draw conclusions from each testing process.

3. Method and Discussion

3.1. Planning

The initial stage of this research is to plan for repeated testing, to design the database as a place to accommodate it. The database design will be displayed in table 1 which is a student table consisting of 6 columns. This table will always be used to accommodate test data from 500 data to completion.

Table. 1. Student Table

Registration Number	Student Name	Gender	Birth Date	Birth Place	Religion
Holds Master Number data, which is numeric or numeric data type containing 8 characters and a primary key	Holds student name data with the same length to maintain consistency.	Contains one character	Data in the form of dates with DD-MM-YYYY format	Represents the same city data for consistency.	Contains the same religion so as to maintain consistency.

The next step is to create stored procedures and functions in the specified DBMS, which are made only to store data because the variables used are data storage. Figure 2 is the structure of the Stored Procedure used to store predetermined student data which is stored with parameters p1, p2, p3, p4, p5 and p6.

```
1    CREATE PROCEDURE Insert_Procs
2    AS
3        BEGIN
4            INSERT INTO student
5                            (.,
6                             ..,
7                             ...,
8                             ....,
9                             .....,
10                            ......)
11            VALUES       ( @p1,
12                           @p2,
13                           @p3,
14                           @p4,
15                           @p5,
16                           p6)
17        END
18   GO;
```

Fig. 2. Stored Procedure

Figure 3 is a function structure used to store data that calls the function for each execution of a sent data parameter.

```
1      INSERT INTO T1 (A1)
2         2  WITH FUNCTION Func_Insrt (.. VARCHAR2) RETURN NUMBER
3         3  IS
4         4  BEGIN
5         5  IF ... = '@p1' THEN RETURN a;
6         6  ELSE
7         7  RETURN b;
8         8  END IF;
9         9  END;
10       10  SELECT ....('...') FROM student;
11       11  /
12     WITH FUNCTION T11 (P_A1 VARCHAR2) RETURN NUMBER
13     ..
14     ..
15
```

Fig. 3. Insert Function

After finishing compiling tables for storing data and creating Stored Procedure and Function queries for storage, then preparing 500 data has been previously collected as initial data and can be added in multiples of 2000 data to be able to sharpen the comparison distance.

3.2. Action

a.   Stored Procedure Testing

The first stage in this test is to test the data speed for storing data in the amount of 500, 2500, 4500 and 6500 student data with Stored Procedure in a computer which is the same in other words data storage locally not with a network or without a data server. After testing, the results can be seen in Figure 4.

SP	Amount of data			
	500	2500	4500	6500
Data 1	344 ms	677 ms	872 ms	1092 ms
Data 2	306 ms	588 ms	878 ms	1028 ms
				*ms = millisecond

Fig. 4. Stored Procedure Testing Results

The test results above explain that when data storage 500 data with data 1 and 2 spent 344 milliseconds and 306 milliseconds, the next experiment with 2500 data resulted in 677 milliseconds for the first data and 588 milliseconds for the second data storage. In the next experiment, 2000 data were added to become 4500 and 6500, it can be seen that the first data took 872 milliseconds and 1092 milliseconds, while the second type of data took 878 milliseconds and 1028 milliseconds. It can be seen from the two data that the difference is not too far away.

b.   Function Testing

The next stage is to test the data speed for storing data in the amount of 500, 2500, 4500 and 6500 student data with SQL Function Insert in a computer which is the same in other words as data storage locally not with a network or without a data server. After testing, the results can be seen in Figure 5.

Func	Amount of data			
	500	**2500**	**4500**	**6500**
Data 1	458 ms	761 ms	899 ms	1112 ms
Data 2	362 ms	599 ms	989 ms	1131 ms

*ms = millisecond

Fig. 5. Function Testing Results

From the data storage experiment with Function, it can be seen that the time taken to store data of 500, 2500, 4500 and 6500 is 458 milliseconds, 761 milliseconds, 899 milliseconds and 1112 milliseconds for the first data and the second data is 362 milliseconds, 599 milliseconds, 989 milliseconds, 1131 millisecond.

c. Testing Stored Procedures on the Network

To be more certain about the tests that have previously been carried out, it is necessary to test them with storage with service providers or data servers. So that it will be seen the influence between the presence of service providers or not, these results will be more convincing with the time interval between the performance of Stored Procedure and Function. Figure 6 describes data storage for 500, 2500, 4500 and 6500 using a computer server.

Serv.SP	Amount of data			
	500	**2500**	**4500**	**6500**
Data 1	562 ms	892 ms	949 ms	1377 ms
Data 2	533 ms	783 ms	1090 ms	1583 ms

*ms = millisecond

Fig. 6. Results of Stored Procedure Testing with Server

Testing is done with two computers, user data is on the first computer and will be stored into the second computer, namely the server computer, from the results in Figure 6 it seems that it takes a longer time to be able to store the same amount of data. But there are inconsistent intervals affected by sending data over the network.

d. Testing Function in Network

The same test is carried out on functions, namely storing 500, 2500, 4500 and 6500 data stored on different server computers, the test results will be shown in Figure 7.

Serv.Func	Amount of data			
	500	**2500**	**4500**	**6500**
Data 1	829 ms	1233 ms	1789 ms	1825 ms
Data 2	898 ms	1341 ms	1590 ms	1783 ms

*ms = millisecond

Fig. 7. Function Testing Results with Server

Figure 7 shows that the function takes longer than the previous test, this can be seen from the test results that have been shown in Figure 7.

3.3. Observe

After conducting experiments and testing, the next process is to observe or observe the results that have been tested, from these results can be illustrated in a graph like Figure 8.
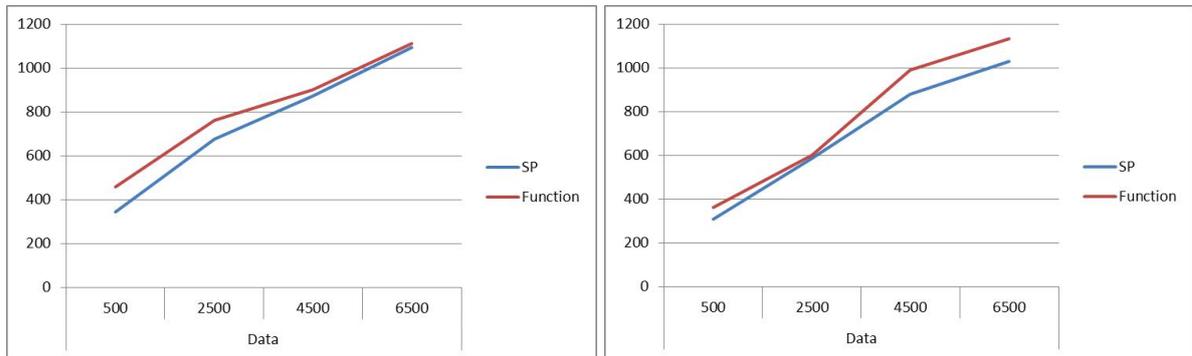
Fig. 8. Comparison of Stored Procedures with Functions
(Left-side : Data 1, Right-side : Data 2)

From Figure 8 is a comparison graph between data storage speed between Stored Procedure and Function, it can be seen that the red line representing the Function takes a long time to store data in each range of the amount of data. These results can be seen in the experiment for the second data, where the function has a long time to store data that has been determined. From these observations it can be concluded that Stored Procedures are faster in storing data, both the first data and the second data.

Next is to observe and compare the results of experiments involving two computers, namely a data provider computer and a computer that functions to store data which is assumed to be a computer serve which is both connected by a network cable, here are the results shown in Figure 9.
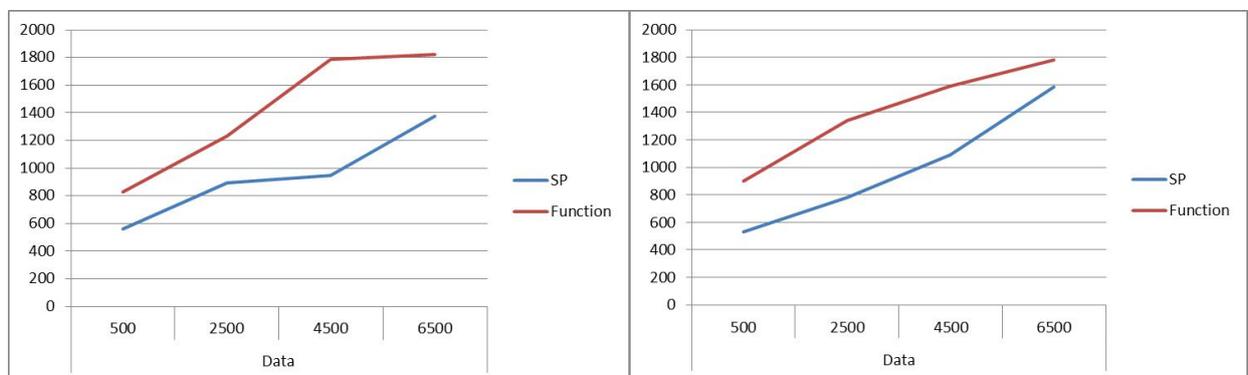
Fig. 9. Comparison of Stored Procedure and Function with a computer server with a network
(Left-side : Data 1, Right-side: Data 2)

From Figure 9, it can be seen that data storage is 500, 2500. 4500 and 6500 in each data and each method has the same relationship, namely Stored Procedure is able to be superior to function in data storage with a server computer connected to the network. It seems that the time interval is wide enough, Stored Procedure is able to complete this data storage time better than Function.

4. Result

From the results obtained from each experiment and pairing data to compare data storage speeds, the next step is to make conclusions on the basis of these experimental results. It can be concluded that Stored Procedures can be faster in data storage, which probably occurs because Stored Procedures do not carry a return value, in

contrast to Functions which always carry a return value after sending a query or carrying parameters that need to be sent.

This distance is increasingly felt when the data provider computer with the storage computer is separated or differentiated, the storage speed distance is increasingly felt when using the function in data storage on the server computer, because the function also carries the return value that always has to be carried over the network when completing the task of sending data to server.

5. Conclusion

From the results of the above research, it can be concluded that Stored Procedures are able to store data faster than functions, this is evidenced by the experiments that have been presented previously. Stored Procedures which also have a return value do not carry the return value one by one but carry the return value simultaneously, in contrast to the function which always carries the return value in each process of sending data via parameters.

References

[1] A. Rahadi, "ANALISIS DAN DESAIN SISTEM INFORMASI PERSEDIAAN BARANG BERBASIS KOMPUTER (Studi Kasus pada Toko Arta Boga)," J. Adm. Bisnis S1 Univ. Brawijaya, vol. 8, no. 2, p. 79908, 2014.

[2] D. Frishman, K. Heumann, A. Lesk, and H. Mewes, "Intelligent Databases : Current Status," vol. 14, no. 7, pp. 551–561, 1998.

[3] S. Peck Lee and D. Zildzic, "Oracle Database Workload Performance Measurement and Tuning Toolkit," Issues Informing Sci. Inf. Technol., vol. 3, pp. 371–381, 2006, doi: 10.28945/898.

[4] D. C. P. Smith and N. Goodman, Data-Base Systems, vol. 11, no. 9. 1978.

[5] A. Yusuf, "Strategi Genius Learning Dalam Pembelajaran Maharatul Kitabah," Pendidik. Bhs. Arab, vol. 9, no. 20, p. 162, 2018.

[6] B. W. M. Boog, "The Emancipatory Character of Action Research, its History and the Present State of the Art," J. Community Appl. Soc. Psychol., vol. 13, no. 6, pp. 426–438, 2003, doi: 10.1002/casp.748.

[7] Y. Li and S. Manoharan, "A performance comparison of SQL and NoSQL databases," IEEE Pacific RIM Conf. Commun. Comput. Signal Process. - Proc., no. November, pp. 15–19, 2013, doi: 10.1109/PACRIM.2013.6625441.

[8] S. Agrawal, V. Narasayya, and B. Yang, "Integrating vertical and horizontal partitioning into automated physical database design," Proc. ACM SIGMOD Int. Conf. Manag. Data, pp. 359–370, 2004, doi: 10.1145/1007568.1007609.

[9] T. Wahyuningsih, "Problems , Challenges and Opportunities Visualization on Big Data," J. Appl. Data Sci., vol. 1, no. 1, pp. 20–28, 2020.

[10] N. H. Trang, "Limitations of Big Data Partitions Technology," J. Appl. Data Sci., vol. 1, no. 1, pp. 11–19, 2020.