PlagScan Results of plagiarism analysis from 2019-08-24 05:50 UTC

04_Jurnal-v01-up.pdf

5.1 %

Date: 2019-08-24 05:45 UTC

✱ All sources **13** | 🌐 Internet sources **7** | ≣ Plagiarism Prevention Pool **6**

☑ [0] ≣ from a PlagScan document dated 2019-03-22 03:36
`1.5%` 6 matches

☑ [1] ≣ from a PlagScan document dated 2018-12-06 15:31
`0.9%` 3 matches

☑ [2] ≣ from a PlagScan document dated 2019-01-29 17:50
`0.8%` 4 matches

☑ [3] 🌐 https://mariadb.com/kb/en/library/what-is-mariadb-galera-cluster/
`1.0%` 6 matches

☑ [4] ≣ from a PlagScan document dated 2019-03-13 03:14
`0.8%` 4 matches

☑ [5] 🌐 vtechie.com/2016/03/15/mariadb-galera-cluster-centos-7/
`0.7%` 5 matches

☑ [6] 🌐 https://www.fromdual.ch/taxonomy/term/63/feed
`0.7%` 4 matches

☑ [7] 🌐 https://www.fromdual.ch/de/taxonomy/term/116/feed
`0.4%` 3 matches

☑ [8] 🌐 https://www.academia.edu/36263456/THE_EFFECTIVENESS_OF_LEADERSHIP_MANAGEMENT_TRAINING
`0.6%` 1 matches

☑ [9] ≣ from a PlagScan document dated 2018-10-30 20:46
`0.3%` 2 matches

☑ [10] 🌐 https://www.govinfo.gov/bulkdata/CFR/2015/title-32/CFR-2015-title32-vol5.xml
`0.0%` 1 matches

☑ [11] ≣ from a PlagScan document dated 2018-07-31 10:47
`0.2%` 1 matches

☑ [12] 🌐 https://www.researchgate.net/publication...erformance_of_cement
`0.2%` 1 matches

7 pages, 3844 words

PlagLevel: 5.1% selected / 5.4% overall
   25 matches from 13 sources, of which 7 are online sources.

Settings
   Data policy: *Compare with web sources, Check against the Plagiarism Prevention Pool*
   Sensitivity: *Medium*
   Bibliography: *Consider text*
   Citation detection: *Reduce PlagLevel*
   Whitelist: --

EXPERIMENTS AND DESCRIPTIVE ANALYSIS IN THE MARIADB DATABASE CLUSTER SYSTEM TO PREPARE DATA AVAILABILITY

Suyud Widiono

Department of Computer System, Faculty of Information and Electrical Technology
University of Technology Yogyakarta
e-mail: suyud.w@uty.ac.id

ABSTRACT

A database server called the Database Management System (DBMS) that relates tables in a database is called the Relational Database Management System (RDBMS). DBMS/RDBMS is a computer program that provides data services for computers or other computer programs. One of the RDBMS type database server (hereinafter referred to as a database server) is MariaDB. The database server is in charge of managing and providing data, so data must always be ready, fast presented, accurate, and safe, it cannot be damaged or even lost. One way to provide this data is to install several database servers using the concept of replication in the Multiple Server Database system. [11] Replication in a cluster server database is a method of installing several database server nodes that allow between node servers to copy each other and distribute data from one node to another database server node, which then synchronizes data between server nodes to maintain data consistency. This study looks for the most optimal number of minimal database server nodes to provide accurate, fast and safe data on the MariaDB Cluster RDBMS. [0] From the results of the replication test from the cluster server database, it can be concluded that the number of 3 (three) node servers can be known to always synchronize and consistency of data between server nodes, so there are 3 (three) nodes of minimum database node with MariaDB RDBMS.

Key words: server, database, replication, cluster, node, MariaDB

1. INTRODUCTION

1.1 Background

A database is a collection of interrelated data that is stored together with controlled redundancy to serve one or more applications optimally (Elmasri R. & Navathe B.S., 2011). Data is stored in such a way that it does not depend on the program used by people to access data. The approach used to add new data, modify and retrieve data in the database is general and controlled to ensure the security of data from people who access data. Systems that manage databases are often referred to as database servers, one type of database server is relational database, therefore it is often called the Relational Database Management System (RDBMS).

RDBMS is a computer program designed to manage the database as a collection of data stored in a structured manner and carry out operations on data at the request of its users. The use of RDBMS (database server) for an institution / organization both small, medium and large scale already uses applications in the form of management information systems to assist its operational activities. The entire information system requires a database that can perform optimally. [12] Various techniques have been developed to improve the performance of a database system, one of which is the replication technique.

When a database server fails in processing data, all services will experience interference and all system business processes will not work properly. The importance of this database server role that makes a database server must have good performance. Data availability services by database servers must not be interrupted and must always provide valid and consistent data, therefore it is important to conduct research on data readiness and availability in the MariaDB database server cluster replication system.

1.2 Formulation of the problem

Based on the explanation in the Introduction and Background section above, formulation of the problem can be taken as follows:

a. How to design and implement computer networks for MariaDB database cluster system?

b. How are the results of system testing based on the interruption activities of several database servers connected in a cluster?

c. What is the recommendation of a minimum number of database servers that must be connected in a cluster, so that they are always ready to provide data?

1.3 Research purposes

The main objectives to be achieved in this research opportunity are:

a. Designing and implementing MariaDB database cluster system.

b. Test and analyze replication data synchronization on all computer database servers that are clustered.

c. Provide solutions and recommendations for the minimum number of database server computers that must be connected in a cluster, to remain consistent in providing data.

2. BASIC THEORY

2.1 Experiments and Descriptive Analysis [1]

According to Sugiyono (2014), descriptive analysis method is a statistic used to analyze data by describing or describing data that has been collected as it is without intending to make conclusions that apply to general or generalizations. Whereas according to Nazir.M (2003), the descriptive method is a method in examining the status of a human group, an object, a condition, a thought, or a class of events in the present.

Experiments according to Sugiyono (2014) explain that "research methods are used to find the effect of certain treatments on others in controlled conditions.

Experimental research aims to investigate the possibility of mutual causal relations, it is explained by Arikunto (2006) that: [10] Experiments are studies that are intended to determine whether there is a result of "something" that is imposed on the subject in question.

2.2 Database

Database is a collection of integrated data storage groups that are centrally regulated and controlled. Databases also store descriptive information such as attribute names, giving limits to values, and access control for sensitive data (Satzinger, Robert, & Stephen, 2005). The database is also interpreted as a collection of files connected to each other. This file is set according to the similarity of the elements, so that the desired data can be easily searched (Williams & Sawyer, 2007). [4] Based on the definitions described by the experts above, it can be concluded the database (English: [4] database), is a collection of information stored on a computer systematically so that it can be checked using a computer program to obtain information from the database. The software used to manage and call database queries is called the Database Management System (DBMS).

2.3 Database Replication

Database replication is a method of multiplying data at several different physical locations for the same logical data. The goal is for each server to be a backup, so that when there is damage or failure in one physical location it does not cause a failure of the entire system. Replication is also to increase the readiness of data availability in database systems by dividing the workload of each database server (Aditya & Juhana, 2015). The implementation of replication is technically different, depending on the Database Management System (DBMS) used. Replication technology has the function of copying data from one location to another location remotely, after the data has been replicated to the remote target, then only the changed data is replicated, thus saving bandwidth.

The initial data replication in remote storage is called seeding. After the data has been successfully seeded, the replication function can then run on the following two models:

a. Synchronous Replication Model

This model allows the exchange of data in real-time so that the data will always be maintained, because when a transaction is writing something to the source storage, then at the same time writing is also done on the target store in the remote location. The entire writing process on the source disk and target disk must be completed first before moving on to the next operational transaction and be acknowledged for both if it is finished. In this replication mode, the need for high system performance must be considered. In addition, the distance between the source disk and the target disk is also the main prerequisite, that the parties involved in this replication mode must be 100km between the two. The advantage of this replication mode is to provide consistent and complete recovery for all time periods.

b. Asynchronous Replication Model

This replication model is done by exchanging data using a buffer, that is, the data transaction is placed on the container, and at a certain time period is replicated to the target. So that this replication model does not guarantee the synchronization of data on the two parties involved. If this happens then heavy recovery is carried out. The advantage of this replication model is cost effectiveness.

2.4 DBMS MariaDB

MariaDB was developed by Michael "Monty" Widenius, (one of the founders of MySQL AB and Monty Program AB) and a community of developers who previously contributed to the MySQL database.

(Source: www.codepolitan.com)
Figure 1. MariaDB RDBMS logo

MariaDB is an implementation of a relational database management system (RDBMS) that is distributed under the GPL (General Public License) license and can be distributed as open source software under the GPL license, can also be used by multiple users at the same time, and can process more SQL per unit time.

The main features of MariaDB are written on the page: https://mariadb.com/kb/en/library/what-is-mariadb-galera-cluster/, are as follows:

a. Synchronous replication
b. Active-active multi-master topology
c. Read and write to any cluster node
d. Automatic membership control, failed nodes drop from the cluster
e. Automatic node joining
f. True parallel replication, on row level
g. Direct client connections, native MariaDB look & feel.
h. No slave lags
i. No lost transactions
j. Both read and write scalability
k. Smaller client latencies

MariaDB has replaced InnoDB with XtraDB which has received improvements from Google and Percona. XtraDB is claimed to be significantly able to run faster than InnoDB, and without the need to install plugins as well as MySQL.

2.5 MariaDB Cluster

Along with the launch of MariaDB Galera Clusters 5.5.33a, the MariaDB Foundation enhances its portfolio with the "Synchronous Multi Master Replication" feature. For this reason, the developer MariaDB has implemented the "Galera Cluster" technology developed by Coder ship and embedded directly into the MariaDB database management system. This solution allows synchronous replication with Multi Master topology. The process of reading and writing in the database can be done in all nodes, where the system can also automatically detect nodes that are not affordable and ignore it from the Cluster. Previously in the MariaDB system only the "semi-synchronous Replication" feature was available.
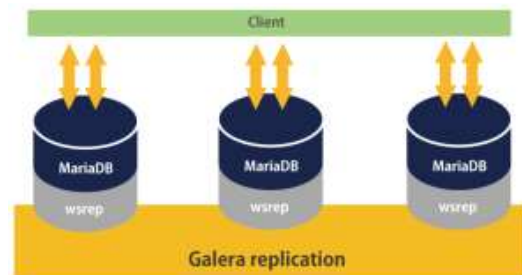
(Source: http://vtechie.com)
Figure 2. MariaDB Cluster Architecture

3. RESEARCH METHODS

The research method used in this study is library study, design of server network architecture, experiments and observations accompanied by descriptive analysis.

3.1 Literature Study

Read and study some literature about the basic theories that support this research by sourcing books on computer networks and database servers, replication, database clusters, journals and internet articles, tutorial and video notes that have to do with the mechanism of clusters and MariaDB database server.

3.2 Experiments and Observations

This experimental study will try to implement the design of a computer network server with a cluster system, trying to disrupt the server so that the server does not work, noting the condition of the availability of output data generated by the system, analysing the presence or absence of causal relationships arising from the treatment.
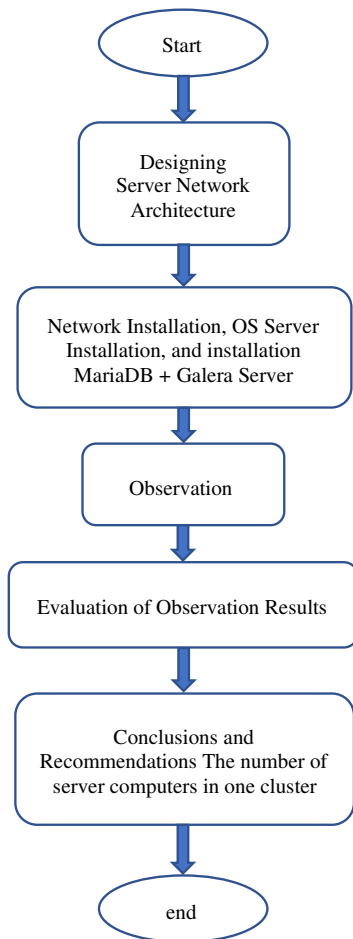
Figure 3. Series of Research Stages

To support the smoothness of the research carried out, the first thing to do is to find and collect various literatures that lead to various methods of cluster database servers and other supporting aspects. The design of the server computer network scheme which will further interfere with the running of the server and observed.[2] Data evaluation and data analysis in accordance with the scenario that will be given to the server cluster in preparing data.[2] In conducting this research in more detail, the method to be carried out is shown in Figure 3. Series of Research Stages.

4. SYSTEM DESIGN

System design in this study consists of two parts, namely the first part is the process of designing a server computer network, and in the second part is the design of a system testing scenario.

4.1 Designing a Server Computer Network

Computer networks in this study were built using 5 (five) physical computers, consisting of:

- 1 (one) Laptop
- 1 (one) Web Server
- 3 (three) Database Servers

Laptops are used for remote servers and operate system testing stages.

Web Server is used to install and run applications that are used to manipulate data in the database server.

Three database servers each use Ubuntu 16.04.6-LTS Linux, including already installed MariaDB Server and Client version 10.6.4 and automatically have Galera cluster that is used to synchronize data in each database server.
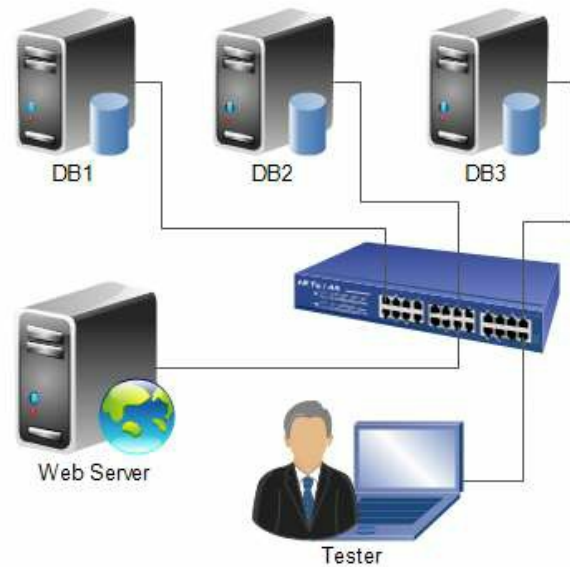
Figure 4. Computer Network Design

4.2 Designing Experimental Tables and Observation Replication

This section is a table design for experimentation and replication observation. From this design it is expected to produce test values that are able to provide recommendations on the number of MariaDB database servers that apply the Cluster concept.

Table 1. Experimental Tables

No. Test	Srv DB1	Srv DB2	Srv DB3
1	On	On	On
2	On	On	Normal Down
3	On	Normal Down	On
4	Normal Down	On	On
5	On	Normal Down	Normal Down
6	Normal Down	Normal Down	On
7	Normal Down	On	Normal Down
8	On	On	disconnect
9	On	disconnect	On
10	disconnect	On	On
11	On	disconnect	Disconnect
12	disconnect	Disconnect	On
13	disconnect	On	disconnect
14	On	On	not normal down
15	On	not normal down	On
16	not normal down	On	On
17	On	not normal down	not normal down
18	not normal down	not normal down	On
19	not normal down	On	not normal down

There are 3 (three) types of disorders tested:

- Normal Down database server, which is a condition in which the database service on a server computer is turned off and is deliberately shut down normally. This condition simulates when the server is under maintenance so that the database service needs to be stopped normally;

- Disconnection, which is a condition where the network connection on the server is accidentally disconnected. This simulates the condition when the network connection on the database server is lost;

- Server Down not Normal, which is a condition where the server is accidentally turned off, causing the database service to stop abnormally. This simulates the condition when the server computer is experiencing an error which causes the server computer to stop working without being shut down;

The three types of interference are tested on clusters with a number of different servers, namely a combination of three existing database servers. The purpose of this condition is to find out the impact of increasing the number of server computers on the readiness and availability of data on MariaDB's database server cluster system. To conduct a trial by giving a disruption to the database cluster system, the Web Server is used to manipulate data that is connected to one of the servers in the database cluster. Then the impact of each sequence number is monitored.

The next step is to return the disturbance to normal conditions (no more interference) and see the condition of the cluster database after the disturbance condition is normalized. This step simulates the condition when the interference is successfully repaired. With experiments like this it is expected to be used to find out the number of database server computers that implement cluster systems.

[0] ▸

5. EXEPRIMENT AND OBSERVATION

Experiments and observations are adjusted to the stages of conditions in Table 1, the results of experiments and observations can be seen in Table 2. The data is obtained from several trials at each stage of the disturbance. The results obtained each time the

experiment did not change, so it can be concluded that the impact of each failure scenario is consistent.

The conditions of test 1, 2, 3, 4, 8, 9, and 10 are failure scenarios in the form of normal database service down and network database disconnect, with the number of nodes 2 and 3 still active, causing no disruption to the database cluster system. Active nodes can still serve access from users.

In scenario number 5, 6, 7, 11, 12, and 13, that is, when a failure occurs in 2 (two) database servers in the cluster system, other active nodes cannot operate. When testing the manipulation of data using the script in the Web Server application into the node an error message appears 'Deadlock found when trying to get lock '. This is because the network connection on the system cluster is disconnected when the State Snapshot Transfer (SST) is taking place on both nodes, so MariaDB prevents the addition of new data to avoid data inconsistencies.

Table 2. Results of Experiments and Observations

No. Test	condition there is a disturbance	Disruption is normalized
1, 2, 3, 4, 8, 9, 10 2 (two) servers ON	A server that has no interference can still provide data	Back to normal. Interrupted server can do auto recovery
5, 6, 7, 11, 12, 13 2 (two) servers OFF	A server that has no interference cannot operate and an error message appears: Deadlock found when trying to get lock	Back to normal. With 'mysqld --wsrep-recover' command to recover data.
14, 15, 16 2 (two) servers ON	A server that has no interference can still replicate.	1407 error message appears: WSREP has not yet prepared node for application use. Some are normal again and an error does not appear.
17, 18, 19 2 (two) servers OFF ABNORMAL	A server that has no interference cannot operate and an error message appears: Deadlock found when trying to get lock	1407 error message appears: WSREP has not yet prepared node for application use. Continuously. Cluster System Failed.

However, this condition does not occur in the cluster database system with the previous conditions, namely the condition of the test numbers 1, 2, 3, 4, 8, 9, and 10. In a database cluster system with three nodes, the SST process can still take place on two nodes that are still active, so that interference on one node does not cause data synchronization to the other nodes in the cluster.

In the test conditions test numbers 17, 18, and 19, namely the failure scenario in the form of a normal server down-up on 2 (two) database servers, the other nodes are not interrupted and can operate normally. However, when the normal down-up node is turned on again an error message 1407 appears that says "WSREP has not yet prepared the node for application use". This error continues so that all nodes in the cluster stop working and cannot synchronize data. This is because there is corrupt data when the database server is down-normal.

The above conditions do not occur in the test conditions number 14, 15, and 16. In the system cluster with a number of nodes 3, the 1407 error message only appears momentarily when the forced stop node is restarted, but after a while the error message disappears and synchronizes data in the cluster can run normally again.

Note to do data recovery during the MariaDB service condition after the server experiences a normal

down-up error. When the server is restarted, the MariaDB service cannot be turned on. So, the step that must be done is to do data recovery with the command 'mysqld - wsrep-recover'. This command will repair data that failed to synchronize.

6. CONCLUSION

From the results of the experiments and observations above, a table can be made as follows:

Table 3. Conclusion of Experimental Results

Disorders	Number of Servers is ON	
	2 DB Servers	3 DB Servers
Server Normal Down	Available Data	Available Data
Network disconnect	System error	Available Data
Server not Normal down	System error	Available Data

Experiments and observations made on the cluster system on the MariaDB Galera Cluster server in this study there are several things that need to be considered by the database server administrator when implementing MariaDB Galera Cluster. As written in Table 3, the first thing that should be considered is the number of database server computers in the cluster system that was built. The experimental results and observations prove that clusters that only have two database servers will be very susceptible to interference if one server has a problem. Even the interference can cause the cluster system to fail. This causes a very fatal interference because the database is a crucial part of a system, where database failures will cause the entire system to be disrupted. To prevent this, the recommended minimum number of computers must be available when building a cluster database system are three database server computers. From the experimental results it is proven that the cluster system built with 3 computers can always provide data from all the failure tests tested, and the database cluster system can still work well. The second thing is when the server crashes, the MariaDB service cannot run when the server returns. You need to do recovery manually manually with the 'mysqld - wsreprecover' command..

In general, this study can provide recommendations that by using three computer servers in a cluster database system, the availability of data can be maintained properly, noting that there are no more than two computers experiencing interference at the same time.

7. DAFTAR PUSTAKA

[1] ADITYA, B. & JUHANA, T., 2015. A high availability (HA) MariaDB Galera Cluster across data center with optimized WRR scheduling algorithm of LVS - TUN, in 2015 9th International Conference on Telecommunication Systems Services and Applications (TSSA).

[2] Arikunto, S., 2006. Prosedur Penelitian Suatu Pendekatan Praktik Edisi Revisi VI. Jakarta : Rineka Cipta.

[3] Elmasri R. & Navathe B.S., 2011. undamental of Database Systems (6th ed.)., Pearson Education, Inc., publishing as Addison-Wesley.

[4] Moh. Nazir, 2003. Metode Penelitian. Cetakan Kelima, Jakarta, Ghalia Indonesia.

[5] Satzinger, John W, Jackson, Robert B, Burd, Stephen D., 2005, Object-Oriented Analysis and Design With The Unified Process. Thomson Course Technology.

[6] Sugiyono, 2014. Metodologi Penelitian Bisnis. Bandung: CV Alfabeta.

[7] Williams, B. K., & Sawyer, S. C., 2007. Using Information Technology: Practical Introduction to the World of Computers and Communication (7th ed.) (Translator: Nur Wijayaning Rahayu & Th. Arie Prabawati). Yogyakarta: ANDI.

[8] CODEPOLITAN, 2019. Mengenal MariaDB. Available in: https://www.codepolitan.com/mengenal-mariadb/, [Accessed July 01, 2019].

[9] MARIADB, 2019. What is MariaDB Galera Cluster? Available in: https://mariadb.com/kb/en/library/what-is-mariadb-galera-cluster/, [Accessed June 04, 2019].

[10] MARIADB, 2019. About MariaDB. Available in: https://mariadb.org/about/, [Accessed June 04, 2019].

[11] VTECHIE, 2019. Install MariaDB Galera Cluster on Centos 7. Available in: http://vtechie.com/2016/03/15/mariadb-galera-cluster-centos-7/, [Accessed July 01, 2019].