International Journal of Cyber and IT Service Management (IJCITSM) p-ISSN: 2797-1325
Vol. 2 No. 1 April 2022 e-ISSN: 2808-554X

Comparison Of Replication Strategies On Distributed Database Systems

Nur Azizah[1], Veranda Hartajaya[2], Sandi Riady[3]
University of Raharja, Indonesia[1,2,3]
e-mail: nur.azizah@raharja.info[1], veranda@raharja.info[2], sandi.riady@raharja.info[3]

To cite this document:
Azizah, N., Hartajaya, V., & Riady, S. (2022). Comparison Of Replication Strategies On Distributed Database Systems. *International Journal of Cyber and IT Service Management (IJCITSM)*, *2*(1), 20-29. Retrieved from https://iiast-journal.org/ijcitsm/index.php/IJCITSM/article/view/70

Author Notification
April 2022
Final Revised
April 2022
Published
April 2022

Abstract

Today's computer applications have ever-increasing database system capabilities and performance. The growing amount of data that has to be processed in a business company makes centralized data processing ineffective. This inefficiency shows itself as a long reaction time. This is in direct opposition to the purpose of utilizing databases in data processing, which is to reduce the amount of time it takes to process data. Another database design is required to tackle this problem. Distributed database technology refers to an architecture in which several servers are linked together, and each one may process and fulfill local queries. Each participating server is responsible for serving one or more requests. In a multi-master replication scenario, all sites are main sites, and all main sites communicate with one another. The distributed database system comprises numerous linked computers that work together as a single system.

Keywords *: Distributed Database, Multimaster Replication, Database.*

1. Introduction

1.1 Background

The usage of multi-process technology has resulted in advancements in Database Management System (DBMS) technology[1]. As a result, database management software that can handle multi-process and multi-user operations is required. This approach has become the de facto data management solution in huge data settings.

Integrating a distributed work environment results in a more efficient distribution function, with application programs running on specialized servers called database servers, while database operations are handled by dedicated machines called application servers. This system is made up of network cables connected via the network[2].

A parallel database system can make use of distributed database technologies. Parallel database systems use parallelism in data management to provide high-performance and high-availability database servers[3]. Data is spread between sites in a network of

computers or nodes of a multiprocessor system in distributed and parallel DBMS, which provide the same capabilities as centralized DBMS [4].

As a result, the system offers users a conceptually integrated view of physically scattered databases[5]. A distributed database is managed by a centralized database management system (DBMS) but has storage devices that are not connected. These storage locations can be physically close together (e.g., in the same building) or separated by a vast distance and connected via the internet network.

Distributed databases can be used on web servers, intranets, office extranets, and corporate networks[6]. There are two methods for keeping remote databases up to date: replication and duplication. Software is used in replication to identify or track changes that occur on a single base. Only after the modifications in one database have been recognized and known should updates be made to ensure that all databases are equivalent .

Because of its complexity, the replication process takes a long time and overburdens the computer. On the other hand, duplication is not the same as replication and is not nearly as complicated. One database is made the master in this procedure, after which it is spread to several copies. To avoid overwriting local data, changes should only be made to the master database during the duplication process[7].

A distributed database management system (DBMS) should be able to solve several information issues (islands of information). Databases are frequently regarded as restricted and inaccessible electronic collections, particularly in remote locations[8]. DDBMS is the solution to issues like geography, computer architecture, and communication protocols.

1.2 Problem Limitations

To stay on track with the subject, the author only compares two strategy replication approaches on distributed databases to prevent getting off track. With multi-master replication, you can have a single master.

1.3 Purpose and Benefits of Research

1.3.1 Research objectives

The goal of this study is to undertake unconventional database trials, in which we will construct and test a distributed prototype database.

1.3.2 Benefits of Research

The advantages of creating this journal include learning about the distributed database system and determining if it is worth it to be a future replacement for a centralized database system.

2. Research Methodology

This paper is an exploratory investigation into the fundamentals of distributed database technology. The objective of an exploratory study is for the author to experiment on how a distributed database system works to improve the author's understanding of the database system and serve as a reference for future research. "If the purpose is to obtain clarification on phenomena, get new understandings, or construct a more clear research challenge, then the research is called the exploratory study."

2.1 Distributed Database Replication Method

The replication method is a method for replicating and distributing database items and objects from one database to another and synchronizing data between databases to ensure data consistency[9]. Replication is a coping database mechanism and the administration of object databases in a computer network that can automatically construct a distributed database system to preserve data consistency.

Data may be sent to other places utilizing this replication strategy via local and internal network connections[10]. Replication also enables application performance and the distribution of biological data according to its intended usages, such as online transaction processing and DSS (Decision Support System) or database processing across numerous servers [11]. Because data may be spread to different servers using this strategy, database dependability will be improved[12].

What will happen, for example, if the Bank only has a database server and the database server suffers from damage or a power outage? The banking transaction will almost certainly come to a halt[13]. If we utilize a distributed database, each transaction is replicated to another server; this is not the case[14].

2.1.1 Single Master Replicated Technique

One computer serves as the master and the other as the slave in this system. The server computer will access and write to the database during this operation. The slave computer, on the other hand, will only read from the database[15]. If we alter the data on the master, the data on the slave will also change automatically. However, if we alter data on the slave, the database on the master will remain unchanged.

2.1.2 Multi-Master Replication Techniques

In this technique, one computer serves as a master server, while the other also serves as a master server[16] . Every computer will write and read data in the database within the following year. If we alter the data on master server 1, the data on master server two will also change. Similarly, if we alter the master server data 2, the master server one database will change. This implies that each master one and master two will have the ability to modify and add data to the distributed database.

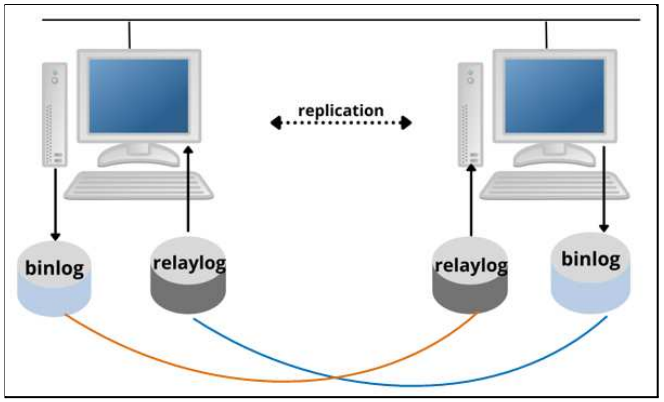

Figure 1. Multi Master Replication Techniques

2.2 PHP (Hypertext Preprocessor)

PHP (Hypertext preprocessor) is a server-side web programming language that is free source. PHP is server-side software that works in conjunction with HTML (server-side HTML embedded scripting). PHP is a script that allows you to generate dynamic web pages[17]. The term "dynamic" refers to the fact that the page to be shown is built as the client requests it. This approach ensures that a client's information is constantly fresh and current. All PHP scripts are run on the server where they are installed.

2.3 MySQL

MySQL (My Structured Query Language) is a database management system. It is a database management system (DBMS) used by various databases, including Oracle, MS SQL, PostgreSQL, and others. MySQL employs the SQL language to handle the database. We may use MySQL for free because it is open source. MySQL databases are also well-supported by PHP programming Anhar[18].

MySQL is the most popular database among web programmers because it is a robust database that is also stable enough to be utilized as a data storage medium. In comparison to other databases, MySQL is the most popular and extensively used database server capable of managing databases properly[19]. In addition to MySQL, several database servers have features that should not be overlooked, such as Oracle and PostgreSQL.

2.4 Apache

The Apache HTTP server, also known as Apache Web Server/WWW, is a web server that may serve and enable web pages on a variety of operating systems (Unix, BSD, Linux, Microsoft Windows, Novell Netware, and other platforms). HTTP is the protocol used to serve this web/www function. Apache's advanced features include customizable error messages, database-based authentication, and more. Several graphical user interfaces (GUIs) for Apache are available, making server management simple[20].

2.5 phpMyAdmin

phpMyAdmin is a free piece of software created in the PHP programming language that allows you to manage MySQL databases through the Internet[21]. phpMyAdmin can manage primary data, tables, fields, relationships, indexes, users, and permissions, among other MySQL activities.

2.6 Xampp

Free software is a collection of applications that runs on various operating systems. It is a stand-alone server (localhost) with Apache HTTP Server programs, MySQL databases, and PHP and Perl language translators[22]. XAMPP stands for X (four different operating systems), Apache, MySQL, PHP, and Perl. The program is free and distributed under the GNU General Public License.

2.7 Database Plan

Table 1. Phone List

No	Field Name	Type	Width	Information
1	Name	Varchar	30	User Name
2	Phone	Varchar	50	Phone Number
3	Address	Varchar	40	Address

2.8 Application Plan

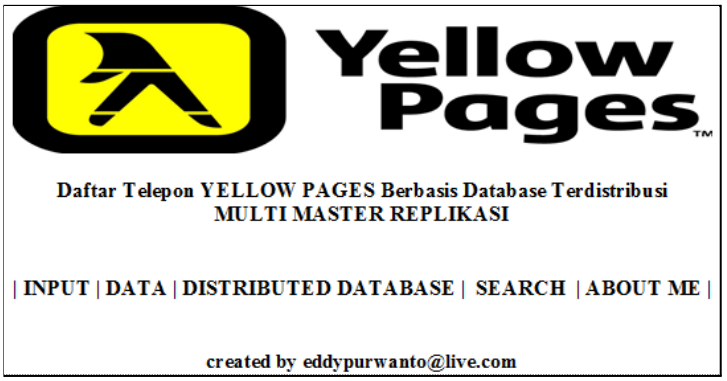

Figure 2. Application Design

3. Results

"Comparison of Replication Strategies on Distributed Database Systems," according to the study. We will compare and contrast two replication methods: single-master and multi-master replication. The procedure is carried out from network configuration through database configuration to testing. We will go through each one individually below:

3.1 Single Master Replication Prototype Testing

Following the completion of the database system prototype, the next step is to evaluate the CPU, Memory, and Time Response utilized by a server computer and slave to replicate the contents of the database once we input the data into the server database[23]. Here, we test a client with 5, 10, and 15 PCs, all accessing apps to change the contents of the server's database simultaneously. Later, the user client and the database server collaborate to alter data on the database server via applications that have been developed.

PC 1 (Master Server) 10.237.7.200

Table 2. Test Master Server Single Master Replication

User Client (PC/Laptop)	**Single Master Replication**		
	CPU (%)	**Memory (%)**	**Replication Time (Seconds)**

5	5	13	1.5
10	14	18	2
15	21	23	3
Average	**13.33**	**18**	**2.16**

Table 2 shows the test results on the master computer for the single master prototype. When accessed 5.10 and 15 users, the average CPU usage, memory, and replication time results are CPU response 13.33 %, memory usage 18 %, and replication time 2.16 seconds. The specs of the machine used as a server influence CPU and memory response, whereas replication time is governed by network access speed.

PC 2 (Master Slave) 10.237.7.210

Table 3. Test Master Slave Single Master Replication

Single Master Replication		
CPU (%)	**Memory (%)**	**Replication Time (Seconds)**
3	10	1
6	12	1.5
8	16	2
Average = 5.6	**Average = 12.66**	**Average = 1.5**

Table 3 shows the test results on the slave computer for the single master prototype. If accessed 5.10 and 15 users, the average CPU use, memory, and replication time results are CPU response 5.6 %, memory usage 12.66 %, and replication time 1.5 seconds. The specs of the machine used as a server impact CPU and memory response, while network access speed determines replication time.

The results in master and slave are pretty significant because in computer slave if there is a change in data on the master, the procedure is conducted and copied in the slave.

3.2 Multi-Master Replication Prototype Testing

After completing the database system prototype, we evaluate the CPU response, memory response, and time response utilized by computer servers 1 and 2 to replicate the contents of the database after we enter data into each server.

Three examples were tested, with a varying number of customers accessing data simultaneously through the apps constructed in each case. Each case has a total of five to ten clients[24]. Table 4 shows the measurements against memory utilization, CPU, and replication speed in master 1 utilizing multi-master approaches.

PC 1 (Master Server 1) 10.237.7.200

Table 4. Test Master Server 1 Multi Master Replication

User Client (PC/Laptop)	Multi Master Replication		
	CPU (%)	Memory (%)	Replication Time (Seconds)
5	6	14	1.5
10	13	19	2.5
15	23	25	3
Average	**14**	**19.33**	**2.33**

Table 4 shows the testing results on the master in the case of the multi-master prototype on master computer 1. CPU response 14 %, memory use 19.33 %, and replication time 2.33 seconds are the average CPU usage, memory, and replication time. The particular type of computer used as a server affects CPU and memory responsiveness, whereas replication time is influenced by network access speed.

PC 2 (Master Server 2) 10.237.7.210

Table 5. Test Master Server 2 Multi Master Replication

User Client (PC/Laptop)	Multi Master Replication		
	CPU (%)	Memory (%)	Replication Time (Seconds)
5	5	13	1
10	14	17	2
15	22	22	2.5
Average	**13.67**	**17.33**	**1.83**

Table 5 shows the master two measurement results for memory use, CPU, and replication speed. When accessed 5.10 and 15 users, the average CPU usage, memory, and replication time results are CPU response 13.67 %, memory usage 17.33 %, and replication time 1.83 seconds. The machine's specs are used as a server adjust for CPU and memory response, while replication time is governed by network access speed.

When comparing single master and multi-master replication approaches, it can be stated that single master replication techniques outperform multi-master replication techniques in terms of memory use, CPU usage, and replication speed[25]. This is because single master replication approaches have a more straightforward procedure than multi-master replication strategies. A single master replication process works in just one direction (one way), whereas a multi-master replication method works in two directions.

4. Conclusions

Following our step-by-step testing of a prototype distributed database using both single and multi-master replication approaches, we can conclude that :

1. Building a distributed database system is tough since we will be dealing with highly complicated challenges. Starting with database configuration, configuring database network access, and other technical and non-technical issues.
2. The single master and multi-master replication systems are essentially the same in that they both replicate all of the database's contents to another computer database in a dispersed way. The single master approach differs in that one of the masters serves as a server (read and write) and the other as a slave (read). All computers are master servers in multi-master approaches (read and write).
3. Compared to multi-master replication systems, single master replication uses less memory, uses less CPU, and replicates faster. This is because single master replication approaches have a more straightforward procedure than multi-master replication strategies. When utilizing a single master, replication occurs in just one direction; however, when using multiple masters, replication occurs in two directions.
4. The pace of replication on single master and multi-master systems is mainly determined by the LAN or internet network's access speed. The server computer's characteristics greatly influence time response and CPU consumption.

References

[1] B. Min, Y. O.-I. J. of Contents, and undefined 2019, “Design and Implementation of Incremental Learning Technology for Big Data Mining,” *koreascience.or.kr*, vol. 15, no. 3, 2019, doi: 10.5392/IJoC.2019.15.3.032.

[2] F. Alfiah, S. Warohmah, M. F. Z. Z. Wahin, M. N. Annisa, and A. Kurniawan, “Social Impact Of Social Media Use Intensity On Mental Health In The Form Of Apatic Attitude,” *Int. J. Cyber IT Serv. Manag.*, vol. 1, no. 2, pp. 226–232, Oct. 2021, Accessed: Jan. 25, 2022. [Online]. Available: https://iiast-journal.org/ijcitsm/index.php/IJCITSM/article/view/62.

[3] S. Z. Friandi, I. Handayani, and F. K. Ratih, “Instagram Social Media As A Means Of Supporting Information And Promotion At PT. AIK Cheong Indonesia,” *Int. J. Cyber IT Serv. Manag.*, vol. 1, no. 2, pp. 151–158, Oct. 2021, Accessed: Jan. 25, 2022. [Online]. Available: https://iiast-journal.org/ijcitsm/index.php/IJCITSM/article/view/41.

[4] D. Fernandes, J. B.- Data, and undefined 2018, “Graph Databases Comparison: AllegroGraph, ArangoDB, InfiniteGraph, Neo4J, and OrientDB.,” *scitepress.org*, 2018, doi: 10.5220/0006910203730380.

[5] B. Rawat and S. Purnama, “MySQL Database Management System (DBMS) On FTP Site LAPAN Bandung,” *Int. J. Cyber IT Serv. Manag.*, vol. 1, no. 2, pp. 173–179, 2021.

[6] T. Nurhaeni, N. Lutfiani, A. Singh, W. Febriani, and M. Hardini, “The Value of Technological Developments Based on An Islamic Perspective,” *Int. J. Cyber IT Serv. Manag.*, vol. 1, no. 1 SE-Articles, pp. 1–13, Apr. 2021, [Online]. Available: https://iiast-journal.org/ijcitsm/index.php/IJCITSM/article/view/4.

[7] D. W. Archer *et al.*, “From keys to databases—real-world applications of secure multi-party computation,” *academic.oup.com*, vol. 12, 2018, doi: 10.1093/comjnl/bxy090.

[8] D. Apriani, A. Williams, U. Rahardja, A. Khoirunisa, and S. Avionita, “The Use of Science Technology In Islamic Practices and Rules In The Past Now and The Future,” *Int. J. Cyber IT Serv. Manag.*, vol. 1, no. 1 SE-Articles, pp. 48–64, Apr. 2021, [Online]. Available: https://iiast-journal.org/ijcitsm/index.php/IJCITSM/article/view/16.

[9] M. R. Anwar, R. Panjaitan, and R. Supriati, “Implementation Of Database Auditing By

Synchronization DBMS," *Int. J. Cyber IT Serv. Manag.*, vol. 1, no. 2 SE-Articles, pp. 197–205, Oct. 2021, [Online]. Available: https://iiast-journal.org/ijcitsm/index.php/IJCITSM/article/view/53.

[10] T. Ayuninggati, N. Lutfiani, and S. Millah, "CRM-Based E-Business Design (Customer Relationship Management) Case Study : Shoe Washing Service Company S-Neat-Kers," *Int. J. Cyber IT Serv. Manag.*, vol. 1, no. 2 SE-Articles, pp. 216–225, Oct. 2021, [Online]. Available: https://iiast-journal.org/ijcitsm/index.php/IJCITSM/article/view/58.

[11] A. Alwiyah, S. Husin, P. Padeli, M. Anggaraeni, and S. Sulistiawati, "Alignment of Science and Technology With Islamic Principles Using Quantum Theory," *Int. J. Cyber IT Serv. Manag.*, vol. 1, no. 1 SE-Articles, pp. 115–120, May 2021, [Online]. Available: https://iiast-journal.org/ijcitsm/index.php/IJCITSM/article/view/32.

[12] P. Edastama, A. Dudhat, and G. Maulani, "Use of Data Warehouse and Data Mining for Academic Data : A Case Study at a National University," *Int. J. Cyber IT Serv. Manag.*, vol. 1, no. 2 SE-Articles, pp. 206–215, Oct. 2021, [Online]. Available: https://iiast-journal.org/ijcitsm/index.php/IJCITSM/article/view/55.

[13] W. Setyowati, R. Widayanti, and D. Supriyanti, "Implementation Of E-Business Information System In Indonesia : Prospects And Challenges," *Int. J. Cyber IT Serv. Manag.*, vol. 1, no. 2 SE-Articles, pp. 180–188, Oct. 2021, [Online]. Available: https://iiast-journal.org/ijcitsm/index.php/IJCITSM/article/view/49.

[14] T. Ramadhan, Q. Aini, S. Santoso, A. Badrianto, and R. Supriati, "Analysis of the potential context of Blockchain on the usability of Gamification with Game-Based Learning," *Int. J. Cyber IT Serv. Manag.*, vol. 1, no. 1 SE-Articles, pp. 84–100, Apr. 2021, [Online]. Available: https://iiast-journal.org/ijcitsm/index.php/IJCITSM/article/view/24.

[15] P. Edastama, A. S. Bist, and A. Prambudi, "Implementation Of Data Mining On Glasses Sales Using The Apriori Algorithm," *Int. J. Cyber IT Serv. Manag.*, vol. 1, no. 2 SE-Articles, pp. 159–172, Oct. 2021, [Online]. Available: https://iiast-journal.org/ijcitsm/index.php/IJCITSM/article/view/46.

[16] I. Y. Ruhiawati, A. P. Candra, and S. N. Sari, "Design and Build a Multimedia System for Indonesian Religious Activities Based on Android," *Int. J. Cyber IT Serv. Manag.*, vol. 1, no. 2 SE-Articles, pp. 233–239, Oct. 2021, [Online]. Available: https://iiast-journal.org/ijcitsm/index.php/IJCITSM/article/view/64.

[17] D. Immaniar, N. Azizah, D. Supriyanti, N. Septiani, and M. Hardini, "PoTS: Proof of Tunnel Signature for Certificate Based on Blockchain Technology," *Int. J. Cyber IT Serv. Manag.*, vol. 1, no. 1 SE-Articles, pp. 101–114, May 2021, [Online]. Available: https://iiast-journal.org/ijcitsm/index.php/IJCITSM/article/view/28.

[18] J. Fjällid, K. Skolan, F. Elektroteknik, and O. Datavetenskap, "A comparative study of databases for storing sensor data," 2019, Accessed: Jan. 25, 2022. [Online]. Available: https://www.diva-portal.org/smash/record.jsf?pid=diva2:1325707.

[19] Y. Durachman, A. Sean Bein, E. Purnama Harahap, T. Ramadhan, and F. Putri Oganda, "Technological and Islamic environments: Selection from Literature Review Resources," *Int. J. Cyber IT Serv. Manag.*, vol. 1, no. 1 SE-Articles, pp. 37–47, Apr. 2021, [Online]. Available: https://iiast-journal.org/ijcitsm/index.php/IJCITSM/article/view/12.

[20] N. Neves, "Database Replication applied to Network Management," 2018, Accessed: Jan. 25, 2022. [Online]. Available: https://run.unl.pt/handle/10362/61550.

[21] D. Cahyadi, A. Faturahman, H. Haryani, E. Dolan, and S. millah, "BCS : Blockchain Smart Curriculum System for Verification Student Accreditation," *Int. J. Cyber IT Serv. Manag.*, vol. 1, no. 1 SE-Articles, pp. 65–83, Apr. 2021, [Online]. Available:

https://iiast-journal.org/ijcitsm/index.php/IJCITSM/article/view/20.

[22] G. Maulani, G. Gunawan, L. Leli, E. Ayu Nabila, and W. Yestina Sari, "Digital Certificate Authority with Blockchain Cybersecurity in Education," *Int. J. Cyber IT Serv. Manag.*, vol. 1, no. 1 SE-Articles, pp. 136–150, May 2021, [Online]. Available: https://iiast-journal.org/ijcitsm/index.php/IJCITSM/article/view/40.

[23] N. P. Puji, A. Dudhat, H. Teja Sukmana, A. Mardiansyah, and M. Hardini, "Modern Technology Advances with Benefits for Humanity to Demonstrate Design with Conventional Sources Islamic," *Int. J. Cyber IT Serv. Manag.*, vol. 1, no. 1 SE-Articles, pp. 14–36, Apr. 2021, [Online]. Available: https://iiast-journal.org/ijcitsm/index.php/IJCITSM/article/view/8.

[24] M. Mulyati, I. Ilamsyah, A. Aris, iketut gunawan, and M. Suzaki Zahran, "Blockchain Technology: Can Data Security Change Higher Education Much Better?," *Int. J. Cyber IT Serv. Manag.*, vol. 1, no. 1 SE-Articles, pp. 121–135, May 2021, [Online]. Available: https://iiast-journal.org/ijcitsm/index.php/IJCITSM/article/view/36.

[25] S. Watini, T. Nurhaeni, and L. Meria, "Development Of Village Office Service Models To Community Based On Mobile Computing," *Int. J. Cyber IT Serv. Manag.*, vol. 1, no. 2 SE-Articles, pp. 189–196, Oct. 2021, [Online]. Available: https://iiast-journal.org/ijcitsm/index.php/IJCITSM/article/view/51.