

# Transferring Attributes of E-Commerce using Replication and Load Balancing Algorithm

Bain Khusnul Khotimah  
Departement of Informatic Engineering,  
Faculty of Engineering  
University of Trunojoyo Madura,  
Bangkalan, East Java, Indonesia

**Abstract**—The Usage of information and communication technology for SMEs is especially access and mobilitas very urgent. Information technology can encourage SMEs to obtain export opportunities and other business opportunities. The distribution of data is now happening between customers the shopping center requires a long transfer access. The process of data distribution will be done by a distributed database by dividing data base into two: a master database and database slave using asynchronous replication methods. Where the database is located in the master database will be replicated to the slave database to cope if a problem occurs in the master database. So that the process of distribution of data between branches and head office can continue. Analysis of the current testing distribution of data to multiple servers was using the throughput, response time, request and reply to each scenario using 4 algorithm than that Round Robin (RR), Least Connection (LC), Weighted Round Robin, and Weighted Least Connection. The results of this study indicate that the output produced by the load balancing algorithm to produce output that varies. The highest average throughput generated by the algorithm SED, WRR, and the WLC and the smallest throughput generated by RR algorithm. Response time is generated by an algorithm which is the smallest and the largest NQ produced by RR algorithm. Request and reply highs generated by the algorithm WLC and WRR.

**Keywords:** Database Replication; Load Balance; Data Center; Information Technology; Database

## I. INTRODUCTION

Data center is the central repository of data on storage media and virtual data management used in certain businesses to usage of community center that cares for the largest archive of information [1]. Data center on information systems is a combination of working procedures, information, people, and information technology organized to achieve goals within an organization. Data center need to service distribute networks with lowers latency between communication user other [8]. Process data exchange approach Office Automation (OA) to support the work firmly spacious, usually used to improve the flow of communication between co-workers, no matter whether the worker must be in the same location or outside [2]. Data centers used for cloud services represents a significant investment in the corporate business world. These problems typically call for extensive communication between servers, so the speed of the computation would drop as the propagation delay between servers increases. Further, the dollar cost of communication would go up if the servers

were spread out across multiple data centers separated by long distance links, as the market price for these far exceeds the cost of intra-building links. Cloud service applications often build on one another. The cloud service providers are built for data center distribution network and improve the reliability of taking the entire site. Design and management of data center networks that right can improve the services designed to benefit from it. To restore this problem, the study of joint optimization of network and data center resources [3]. Having large numbers of servers in the same location eases systems design and lowers the cost of efficiently supporting applications with multiple dependencies and associated communication needs [8].

Each user can access the internet anywhere without difficulty find data and backing up data manually. The data center is the newest form of technology development in the field of cloud computing networks which are designed can present data online to facilitate the planning, implementation and monitoring of data of SMEs. Data Center is a technology to overcome the problems of a rapidly growing industry and work automatically, using the Distributed Database Replication. Replication technique is a technique for duplicating and distribution of data from the Data Base Data Base to another [4]. This system also as a storage medium from one place to another storage medium by performing synchronization so that more consistency of the Data Base. This research resulted in data center applications by creating a web server side and client side web SMEs which will make the process of data distribution and replication between databases. Improvement data center is expected to help the local government for the unification of data and data access among SMEs [5].

Replication is also possible to support the deployment application performance with physical data in accordance with its use, such as for online transaction processing, decision support system (DSS) processing requires database distributed through the relationship between server and client. The data center is needed where the presence of this system is expected to help the process of centralization of data and maintain the availability of data continuous and up to date. This research is to build data center applications that can bring together data from multiple clients by organizing into a management server with diverse database automatically. Process data record SMEs do with data center applications, the planning system in the data center will be connected on-line of some SMEs that acts as a data recorder

to the center of the Department. According Setiadi, et al (2007), the process of data access can be done on-line and off-line configuration is done locally with multiple network devices. If there is a connection between the Data Center and SMEs, then the connection is done in a virtual private network. As for the areas of SMEs have to bear a huge work load, although the data record is done on-line. While in some SMEs record process data off-line, where data is temporarily stored in the external media (flash/CD/HD external). Then admin will be placed at the district/city for the duration terterntu update to the Data Center via a VPN. Configuration data will be described as parsialsebagai data systems in a single PC application with a server. The system would perform extrat and process data interoperability, either from the data server and the data server in the Office of SMEs. While, the user is necessary overview of the network system that can integrate service access information based on the relation of heterogeneous data into complete information.

This Research develops a model of analysis to predict the throughput and response time of the replicated database using a measurement workload on a standalone database. This study will develop analytical models that predict large scale application workloads on a database replication system. Performance data center system that has been replicated depends on the parameters of the workload. Model data center designed for system-based replication middleware in a LAN environment requires estimates predict throughput and response time. We validate the models by comparing their predictions to the measured performance of the prototype for the second multi-master and the master system. We are aware of the many complexities and availability trade off between single and multi-master replication. Data Center is a facility which used for the placement of a collection of servers and components, such as telecommunications systems, and data storage. The data center is managed by an administrator to support the entire performance of the network, application usage, and measurement standards.

## II. DISTRIBUTED DATABASE LITERATURE

Distributed database is a collection of data that is shared logical connection with each other but physically scattered on a computer network. Database is shown table and divided into several fragments, where fragment is stored on computers under the control of a separate Distubed Database Management System (DDBMS), to connect computers using access networks". The site will have the ability to access a user requests on local data and are able to process data stored on other computers connected to the network. DDBMS is collection of data that is used with the are connected logically but physically scattered on a network computer. All the problems that exist in the system Distributed must be solved internally, not externally problems or user levels in the operation of the database using data manipulation language (DML) such as SELECT, INSERT, UPDATE, and DELETE logically should not be changed. Perhaps the definition language (DDL) experienced some developments, such as when a table is created in Torino, then the manufacturer can determine for

example the data stored in Site B . We call these principles as the main Rule of distributed systems [9].

### A. Load Balancing Server

load balancing as a result of the evolution of clustering that there is strong integration between the set of servers with different OS. Advantages of server load balancing to balance the number of requests among the servers. load balancing server can also support a variety of network protocols that are more flexible and have a more simple design of technology, there is no interaction between the nodes and clear delineation function. The design of server load balancing in accordance with its functional shown in Fig.1. According Fachrurrozi that the Load Balance is one of the features developed from the cluster architecture. The system aims to balance the load balancing system work in between nodes in a cluster. According to Tony Collins (2001), load balancing is a technology that transmits data communications (traffic) in several network servers. The server load balancing device will send data traffic to an address directly to many servers [10].

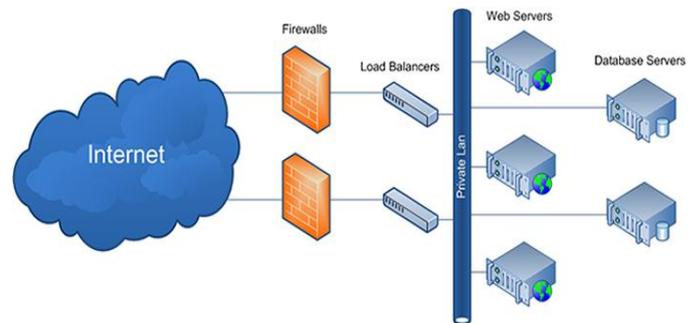


Figure 1. Load balancing architecture

Load balancing will share the load equally in computer internetworking in Fig. the distribution of the burden on a service that is on the set of servers or network devices when there is a request from the user. the concept of load balancing techniques used are subnetting is dividing the two lanes internet connection to many computer in balancing the load of connection by IP host.

### B. Round Robin (RR)

Round Robin is one of the simplest scheduling algorithms with certain scheduling, where every incoming request to the server will be entered into the list without priorities (cyclic executive). For example three cluster server (server A, B and C) with the first request will be given to the server A, 2nd request will be granted to server B, the request 3 will be given to the server C, and 4 requests would be given to the server A again, and so on. Round Robin algorithm treats all equally server does not depend on the number of incoming connections or response time of each server. In the software algorithm is an algorithm default. RR easy to understand and work effectively. The drawback does not consider server load, quite possibly one of the servers will be overloaded because of the high number of requests and the low processing capacity, while the server which has a larger capacity does not do anything [12].

### C. Weighted Round Robin (WRR)

Weighted Round Robin algorithm is designed for better handling on different servers processing capacity. Each server is weighted integer that indicates the capacity of processing. Servers that have a higher weight will receive a first connection of the server that has a lower weight. Servers that have more weight will get a lot more connection requests than servers have a smaller weight. For example, there are three servers, A, B, and C which each have a weight of 4, 3 and 2. So, the good scheduling order for this case is AAB ABC ABC. In the implementation of weighted round robin scheduling, order scheduling will be generated based on the weight of the server after the virtual server configuration is created. The network connection will be forwarded to different servers based on the scheduling order used Round Robin. In short, round robin scheduling is Weighted Round Robin scheduling that considers all the same server weights. Scheduling of Weighted Round Robin is better than round robin scheduling when the processing capacity of each server node is different. This will lead to dynamic load imbalance among servers if the request is dating a very varied. request requires a response that likely will be directed to the same server node (which made the greatest weight). The administrator will correctly determine weight for balance division performance. The better the administrator determine the weight of each server for resulting the better the performance generated by this algorithm [11].

### D. Least Connection (LC)

Least Connection scheduling algorithm has direct connection from the network to the server that has the fewest number of active connections. This scheduling algorithm is dynamic because it must calculate the number of active connections on each server dynamically. For the virtual server scheduling algorithm Least Connection good enough to smooth distribution when the connection request is very diverse [12]. Virtual server will direct the request to a server that has an active connection at least. If the RR algorithm calculates the number of connections when you share a connection, then the LC algorithm takes into account the number of active connections which means it is related to the processing time. Each unit of time LC algorithm has to check the connections is active on each server and then create a priority order division of labor with a server that has the fewest active connections to become the first. If there are two or more servers that have the same number of active connections, the LC will share the work as well as ways in which the RR algorithm. First glance it does seem the least connection scheduling will also run well although there are differences between the processing capacity of servers in the cluster, because the server has greater processing capacity will get more network connections. In fact, it did not work well because the TIME\_WAIT state-owned TCP. The mean of TCP TIME\_WAIT is typically access about 2 minutes. During this time a busy site often receives thousands of connections. For example, server A is twice as powerful than server B. A server processing thousands of requests and keep it in its TCP TIME\_WAIT state, but server B must

crawl to complete thousands of such connections. So, scheduling Least Connection may not run well on several different servers processing capacity. Advantages: the balance of the division of labor based on the number of active connections on each server. Therefore, the connections of users are taken into account. Disadvantages: do not consider the processing capability of each server. Additionally, these algorithms are technically blocked by one of the procedures that the TCP TIME\_WAIT. This algorithm is suitable for use if the server handles the connection varied [13].

### E. Network and Server Performance

Throughput is the number of bits per second received successfully through a system or a communication medium in a particular observation time interval. Generally throughput represented in units of bits per second (bps). The main aspect throughput ranged on the availability of sufficient bandwidth to run the application. It determines the amount of traffic that can be obtained by an application as it passes through the network. Another important aspect is the error rate and the losses associated with a buffer capacity are stated in bandwidth [14].

Response time can be broken down into several factors, namely latency, processing time, and think time. Think time is the time it takes the server to read a request and enter the order into the queue. Processing time is the time required by the server to process a request. In the measurement of web servers, latency is the time required by the request to be sent from the client to the server to process the request plus the time to send back from the server to the client.

$$\text{Response time} = \text{latency} + \text{processing time} + \text{think time} \quad (1)$$

Latency can then be elaborated back into the transfer time plus queue time. Transfer time is the time required to send and receive packets. Queue time is the time spent by a packet to wait in the queue before being processed.

$$\text{Latency} = \text{transfer time} + \text{queue time} \quad (2)$$

Processing time is of course influenced by the processor, memory and cache capability of server. The speed of the processor and RAM capacity as the same as server process that will impact on the lower processing time. Cache capability is essential when the server should provide the same service in large numbers. Time to queue (queue time) is influenced by a large number of requests from the client, the number of threads that process the requests on the web server engine, and the maximum amount of memory for each thread. The number of requests is received by the server by sending a notice to the client that the request has been accepted. Besides effect on latency, the connection is affected the high throughput. It is the average amount of content (number of bits) received multiplied by the number of content requests Replied connection and divided by the measurement time.

### III. RESULT AND DISCUSSION

The design of the network used for the experiment using a local network by building a database system at each site can communicate with each other, and two distribution models as a means of communication and a replica process. As for the process needs to be made a replica database link on each site, as seen in the figure below:

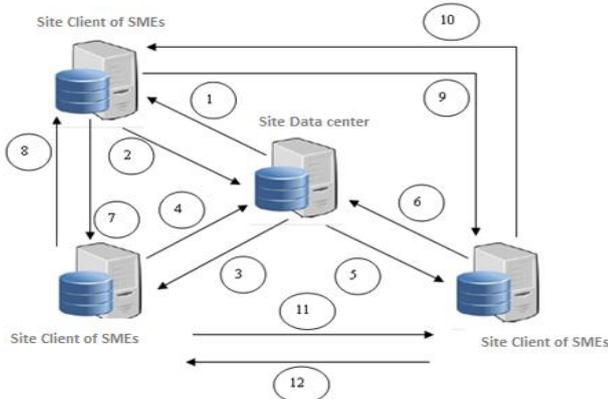


Figure 2. Transfer data between SMEs

TABLE 1. Details on the client and the IP address of each site database

No	Site Database SMEs	Client	IP Address	explanation
1	datacenter.com		Eth0:10.5.2.241	Server database
2	nusaindahmadura.com	Client 1	Eth0:10.5.2.242	Server web
3	hmf-hjunyil.com	Client 2	Eth0:10.5.2.244	Server web
4	alfatih-kkg.com	Client 3	Eth0:10.5.2.245	Server web
5.	Pgpool Server		Eth0:10.5.2.230	Server load balance and replikasi Postgresql
6.	Server load balance web		Eth0:10.5.2.232	Server load balance sekaligus Router

#### A. Setting Computer Networking

This scenario will conduct experiments on three web server running Apache2 and use load balancing using a dedicated server. These servers run specialized software Ipvadm. Here do load balancing algorithm change attempts at Ipvadm. PostgreSQL database server used was 3 servers that are already in the load balancing and replication using Pgpool and PgpoolAdmin.

The algorithm load balancing for Web servers do several attempts using different algorithms for each scenario and then retrieved data from the experiments. To add to the server load balancing algorithm Ipvadm use the following command:

```
#ipvsadm -A -t 11.12.13.1:80 -s wrr
```

The above command means that the server will run its service on the interface using the IP Address 10.11.12.1 of port 80. The algorithm used is WRR (Weighted Round Robin). So, we need to define the IP address or server which will receive a request from a user who forwarded by Ipvadm server.

Datacenter.com web applications in the main menu there are four of them, namely the menu profiles, news, messages and SMEs. Where once the main menu has a submenu, as in Figure 3.9: In datacenter.com web application, there are four main menus among which menu profiles, news, messages and SMEs which will be distributed from the data center to the Web site each SME using dblink\_dc\_SMEs. Then the table of product categories, products, orders, receipts and customers distributed from site to site datacenter.com SMEs using dblink\_SME\_dc. In figure 3 for Data Center database associated with the site, while the on site SMEs associated with the product category, product, city, orders, and customers' receipts. Server on SMEs at each site after the database is installed, then the next step to configure the IP address on each computer site server koputer department of cooperatives and SMEs, as shown in table 1.

The next server will divert incoming request to the IP address 11.12.13.1 on port 80 to the IP address 11.12.13.10 port 80 as well. These transfers using MASQUERADE option (-m). -w. Parameter of the server will be given a value weight of 100. Improvement this value using decimal numbers (integers). Giving this parameter done because if the algorithm used is WRR and wlc.

#### B. Response Time Trial Replication

The database system at each site can communicate with each other, so that the database link needs to be made at each site. The results of measurements carried out in accordance with a scenario that has been designed. The measurement results are presented in two tables, namely, initial trials and Measurement result table. Label the initial trial includes the results of experiments in order to measure the number of connections that can be accommodated by the scenario. The number of connections is then used in the actual measurement results are presented in the table of measurement results.

The Round Robin and Least Connection algorithm are began of the experiment that one of the web server shortage of resources. This is a result of the incompatibility division of labor performed by Ipvadm to servers. The specifications, when using a weighted algorithm Least Connection, needs to be done several time resetting Award.

TABLE 2. Test Results of Response time

Client	Table test	Data	Time access (second)	Client	Data	Time access (second)
Client of SME 1	TBINFORMATION (site datacenter)	4	0,14	Client of SME 2	4	0,17
	TBPRICE (site datacenter)	6	0,18		6	0,16
	TBUSER (site datacenter)	3	0,21		3	0,20
	TBAGENDA (site datacenter)	4	0,25		4	0,31
	TBSME (site datacenter)	1	0,19		1	0,17
	TBFORUM (site datacenter)	10	0,25		10	0,21
	TBKCATEGORYPRODUCT (2 site UKM)	6	0,27		9	0,21
	TBPRODUCT (2 site UKM)	24	0,25 detik		24	0,17

TABLE 3. Time analysis comparison of several methods

Algorithm	Total Connections	Throughput (KB/s)	Response Time (s)	Request	Reply
Round Robin	77	129.9	6717.5	626	609
Least Connection	90	174.04	5747.6	751	740
Weighted Round Robin	97	181.5	6196.01	816	801.1
Weighted Least Connection	97	181.02	6373.52	831.9	818.7

Pgpool began to run out of RAM when tested using the above connections 95 connections. Thus, the number of connections used is 95 connections. The enhanced connection is of up to 90 connections, one experiencing overload web servers. The number of connections used is 85 connections.

Before discussing the results, it is better to first find out the cause of the things that impede service server to a client request. The things that become an obstacle server service on this study, namely the capacity of memory, processor speed, and In scenario 1 when using algorithms Round Robin and Least Connection at the beginning of the experiment that one of the web server resource shortage. This is a result of the incompatibility of the division of work done by Ipvssadm on different servers this specification. Similarly, when using WLC algorithm, need to be done several times a rearrangement weighting each server to obtain the truly appropriate weighting. The throughput is generated by a web server Apache2. So, it is generated by Nginx does. The average throughput generated by the highest Apache2 is using SED algorithm, WRR, and WLC. The average throughput of the three algorithms is almost the same. So, it is generated by Nginx no visible difference in average throughput significantly. The smallest throughput generated by RR algorithm when using a web server Apache2.

### CONCLUSION

1. NQ algorithm produces an average response time low when implemented using Apache2 web server.
2. Algorithms WRR and WLC generates an average throughput, number request and reply highest number when implemented using Apache2 web server.
3. The algorithm produces output RR bottom of every parameter when implemented using Apache2 web server.
4. Sixth load balancing algorithm produces output that is relatively the same on every measurement parameter when implemented using Nginx web server.

### REFERENCES

- [1] Molla, A. & Licker, P.S. E-commerce Systems Success: An Attempt to Extend and ReSpecify the Delone and McLean Model of IS Success, "Journal of Electronic Commerce Research," vol 2, No. 4, 2001, pp.48-58.
- [2] Sanayeie, A. "Electronic commerce in the Third Millennium," Jihad Daneshgahi Publication, Isfahan, 2002.
- [3] Seddon, P.A. "Re-specification and extension of the Delone and McLean model of IS success," Information Systems Research, vol 8, no. 3, 2010, pp.24-53.
- [4] Senn, J.A.. "Business-To-Business E-Commerce," Information Systems Management, 2000.
- [5] M. Bari, R. Boutaba, R. Esteves, L. Granville, M. Podlesny, M. Rabbani, Q. Zhang, and M. Zhani, "Data center network virtualization: A survey," Communications Surveys Tutorials, IEEE, vol. 15, no. 2, pp. 909-928, 2013.

- [6] M. Chetty, R. Buyya, "Weaving computational grids: How analogous are they with electrical grids?," *Computing in Science and Engineering*, vol. 4, no. 4, 2002, pp.61–71.
- [7] D.S. Milojicic, V. Kalogeraki, R. Lukose, K. Nagaraja, J. Pruyne, B. Richard, S. Rollins, Z. Xu, , "Peer-to-peer computing," Technical Report HPL-2002-57R1, HP Laboratories, Palo Alto, USA, 3 July 2003.
- [8] Claudio Fiandrino, Dzmityr Kliazovich, Pascal Bouvry, Albert Y. Zomaya, C, "Performance Metrics for Data Center Communication Systems," IEEE 8th International Conference on Cloud Computing, European Union, 2015
- [9] Kaiji Chen, Yongluan Zhou, Yu Cao, "Online Data Partitioning in Distributed Database Systems," 18th International Conference on Extending Database Technology (EDBT), March, 23-27, 2015, Brussels, Belgium
- [10] Mohammadreza Mesbahi, Amir Masoud Rahmani, "Load Balancing in Cloud Computing: A State of the Art Survey," *I.J. Modern Education and Computer Science*, 2016, 3, 64-78, Published Online March 201 6 in MECS
- [11] Zuber Patel and Upena Dalal, "Design And Implementation Of Low Latency Weighted Round Robin (Lwrr) Scheduling For High Speed Networks," *International Journal of Wireless & Mobile Networks (IJWMN)* Vol. 6, No. 4, August 2014
- [12] Mustafa ElGili Mustafa, Amin Mubark Alamin Ibrahim, "Load Balancing Algorithms Round-Robin (RR), Least-Connection and Least Loaded Efficiency," *International Journal of Computer and Information Technology*, Vol. 4 no. 2, March 2015, pp.255-257
- [13] Brajendra Kumar, Vineet Richhariya, "Load Balancing of Web Server System Using Service Queue Length, *International Journal of Emerging Technology and Advanced Engineering*," *International Journal of Emerging Technology and Advanced Engineering*, Vol 4, No 5, May 2014, pp. 73-81.
- [14] *Gottfried Schimunek, Erich Noriega, Greg Paswindar, George Weaver*, "AS/400 HTTP Server Performance and Capacity Planning," IBM International Technical Support Organization, January 2000.