A Study on the Disk Performance Comparison

Hyun-Ju Song and Young-Hun Lee*

Dept. of Electronic Eng., Hannam University, Ojeong -dong, Daedeok-gu, Daejon 306-791, Korea
iisong86@nate.com, yhlee00@hnu.kr
**Corresponding Author: Phone: +82-42-629-7565, Email: yhlee00@hnu.kr*

Abstract

Nowadays, Attention about how to effectively manage data according to vast amount of data is increasing. Also, concern about power consumption and CO2 generation with regard to green IT business is trends which growing up. In this paper, performance of HDD and DRAM-based SSD was analyzed when used as data storage. Performance comparative-analyzed used Tool accordingly that make up a total of two kinds of environments and created three kinds of conditions, and comparative-analyzed power consumption amount and CO2 generated amount each of storage when experiment. Through this study result, when processing of low data I/O, use the product that using a HDD was judged beneficial. Used DRAM-SSD when stored and administered large amount data that is judged to be more effective which reduction in the time and power consumption reduction. Also, as a way which to handle of large amount data using a configured environment, SAN switch is considered more advantageous.

Keywords: *SSD, HDD, SAN, Mysql, TPC-H, postmark benchmark, CO2, power*

1. Introduction

Recently, as the amount of data is bulky, study is increasing about storage device and technology that stored and quickly process data of large amount. Difference has occurred in the weak development speed of the HDD and processing speed of CPU, therefore, serious data input/output bottleneck has occurred that was created to solve a problem is SSD. From the research of by applying HDD and SSD with storage, the difference of data I/O processing performance was progressedby comparing performance of storage device of each [1-3]. The purpose of this paper are investigated the analysis results that using a various tool to study about Beneficial to use storage device in any case. And As future research directions are presented.

The process of this paper is as follow. In Chapter 2 we will introduce used technology that evaluates performance of HDD and SSD, in Chapter 3 we will introduce analysis environment and condition by thinking each storage device and using tool. And in Chapter 4 analyzes test results will be shown with conditions described in Chapter 3, and Chapter 5 is concluded.

2. Related Technology

2.1. SSD

SSD is an Abbreviation for Solid State Driver.itis storage that is createdcombing nonvolatile NAND Flash Memory with Controller for control function. And it began to

replace the HDD that used as a storage device from pc because fast data reading-speed and shorten s/w starts up time of the pc.

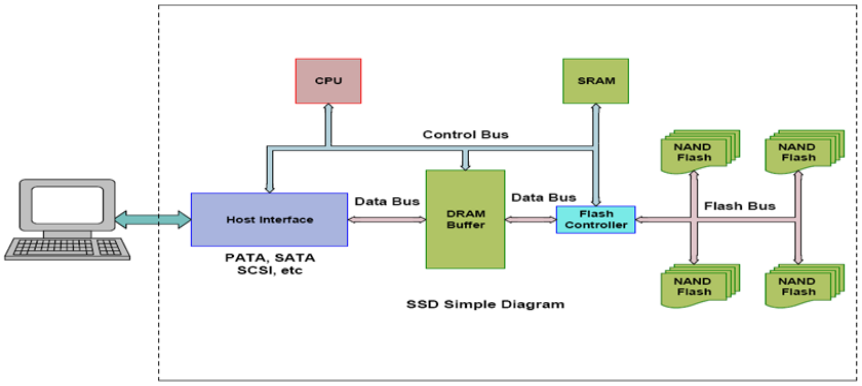

Figure 1. Block diagram of the SSD

2.1.1. Flash based on SSD: Flash based on SSD uses NAND Flash Memory as the primary storage device.It is similar memory stick at side using a Flash Memory,and it is technology that already forming market because information can freely input/output by electrical method, low power consumption, high-speed programmable.

2.1.2. DRAM based on SSD: DRAM-based SSD uses volatile DRAM as the primary storage device, and because it saves and access data directly from RAM-Chip, it has more thanfaster conventional magnetic devices.DRAM is basically volatile device but it is performed by integration of nonvolatile backup system as interior battery and HDD.

2.2. SAN(Storage Area Network)

SAN stands for Storage Area network and say move available high-speed network for large-capacity between storage equipment which unrelated distributed sort of Host. Usually when we talk about the SANmeans transfer SCSI data over fibrechannel. But fibrechannel should support many number of protocol in addition to SCSI [4].

Figure 2. General storage server VS SAN structure

2.3. MYSQL

MYSQL is the relative database management system of open source that uses SQL which is standard database quality language, which is very fast, flexible, and easy to use. MYSQL

offers client/server environment, and a server installed MYSQL has MYSQL daemon called mysqld, so client program connects via network by this daemon so that it can control data.

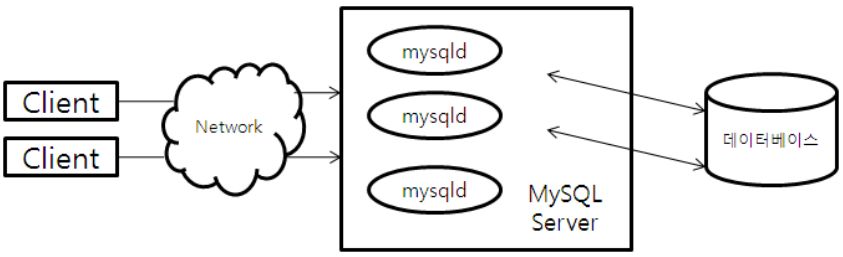

Figure 3. Performance Structure of MYSQL

2.4. TPC-H

TPC-H is benchmarking tool that to measure how quickly can handle complex SQL. It defines22 SQL statements and DB schema, and set of dataabout 1GB.TPC-H benchmark is public performance test that is used SQL that Business-oriented ad-hoc Query and concurrent data modifications made by the combination about large data.

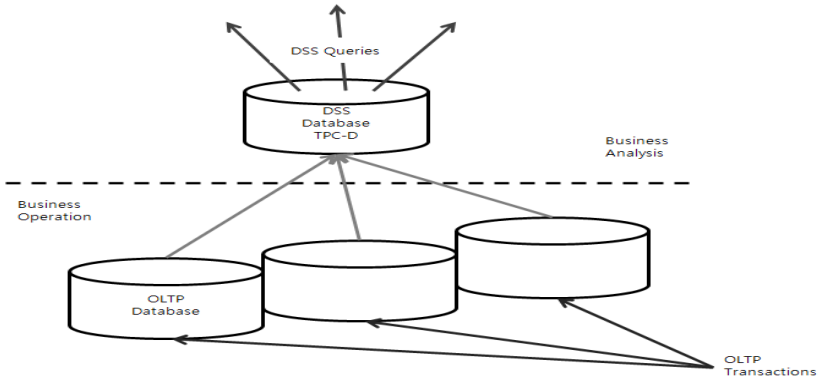

Figure 4. TPC-H's business environment

2.5. Postmark Benchmark

Postmark benchmark is benchmark program that used to measure the performance about files.It is creating having the size much of the text files. And then, it is not method that measure speed that reading and writing of traditional large amounts of files. However, it is benchmark to identify processing capacityabout small files performing Operation about deleting and reading the randomly selected files.

3. Experimental Environment and Conditions

3.1. Experimental environment

SAN environment has been implemented using postmark benchmark to analyze performance of each of storage, power consumption and CO2 emission of storage analyzed when raising using postmark benchmark data I/O. Data store of database (MYSQL) set in each of storage, performance analysis andpower consumption and CO2 emission of storage

using TPC-H Toolmeasurement-analyzed by constructing as using a SAN switch SAN environment and Local environment of general store concept.

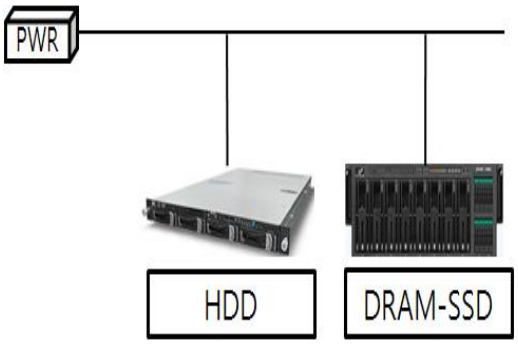

Figure 5.Local test environment configuration

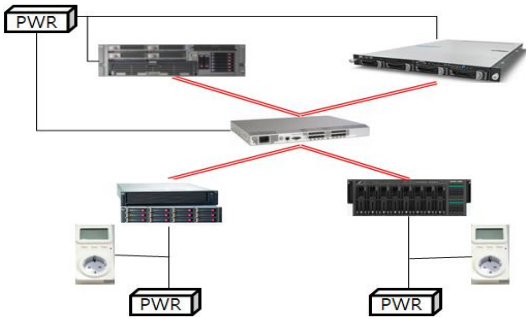

Figure 6. SAN test environment configuration

3.2. Experimental conditions

A total of three experiments were conducted having the two kinds of tool (postmark benchmark, TPC-H) at two kinds of environments of Section 3.1.

3.2.1. Postmark benchmark condition: As shown in Table 1, performance analysis of HDD storage and DRAM-SSD storage were measured simultaneously power consumption as the target three kinds of standard measurement criteria (Low, Medium, High) of postmark. Measuring the Block Size = 4KB, File Size = 327KB measured at a fixed total of 10 times, the average was calculated.

Table 1. Performance testing condition that using postmark Benchmark

Postmark Test Level	Test condition		
	Subdirectory number	File number	Transaction number
Low	10	10,000	50,000
Medium	10	30,000	50,000
High	100	70,000	100,000

3.2.2. TPC-H condition: Using the TPC-H tool to analyze the performance of three-step procedure has changed compared with values from the final analysis. Load Test is the first step, which make up database and step data store to generate. But, performance analysis did not include results. Next, when Single active userinthePower Test is run the Query that wasanalyzed bymeasuring the ability. Finally, when Multi active userintheThroughput Test is run Query at same time that was analyzed by measuring the ability. Power TEST and Throughput Test results have combined to analyze the performance of each storage device.

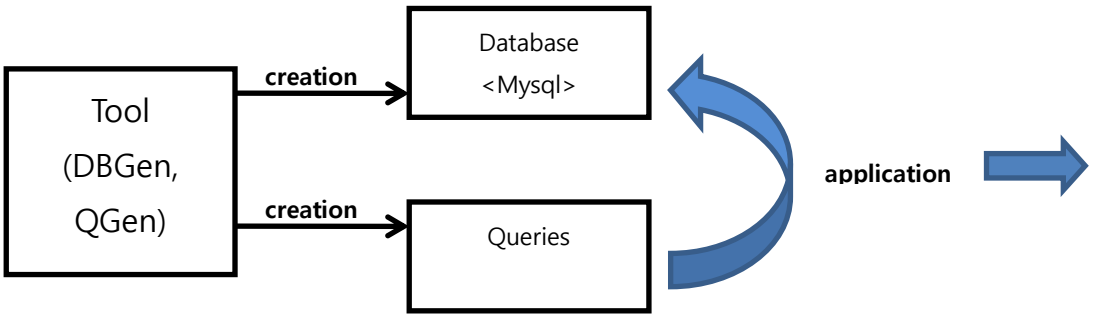

Figure 7. TPC-Happliesto get the results the Block Diagram

Table 2 is conditions for the through test, Figure 7 shows the results of the TPC-H applies will get the block diagram.

Table 2. Conditions for Throughput Test

Database capacity	User number
1GB	2
5GB	2
10GB	3

4. Experimental Results

Performance of each of storage comparative analyzed in terms of total three. In addition, analyzed measuring along with power consumption of each storage device and analyzed emission of CO2 simultaneously at experimental environment connected SAN switch.

4.1. Results and analysis using postmark

Table 3. Postmark performance measurement results

Postmark Test Level	Test results(performancetime : Sec)		
	HDD	DRAM-SSD	Rate
Low	15	15	0
Medium	153	34	4.5
High	388	74.33	5.2

Table 3 is result measuring total performance time during load occurs from each of level. At the Low load conditions (Low Level) there was no difference in the performance.But as the data load is increasing, known increased that performance difference according to run time of DRAM-SSD storage and HDD storage.

4.2. Results and analysis using TPC-H

Experiment results using TPC-H tool, comparative-analyzed performance of each storage device when same conditions at Local environment and SAN environment, and analyzed difference of performance according to environment.

4.2.1. Local environment: As can be seen as a result of figure 8, QphH@size value that can handlead-hocqueries, find out the difference between less to handlead-hocquery capabilities of HDD Storage and DRAM-SSD Storage in a small database capacity, but With increasing amount of data to the DRAM-SSD Storage HDD Storage for more than an hour to handlead-hocquery capabilities that can be seen that much higher.

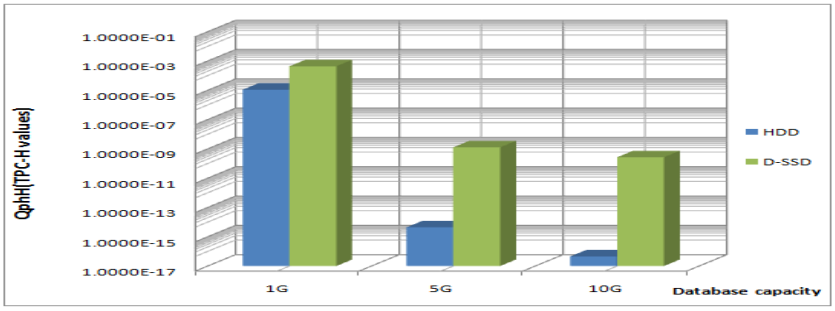

Figure 8. QphH@size value in the Local Environment

4.2.2. SAN environment: Performance test results also at SAN environment, performance of storage were few difference in low load database capacity, but DRAM-SSD storage can be seen that much higher which per hour to handlead-hocquery capabilities more than HDD storage when amounts of data is increasing.

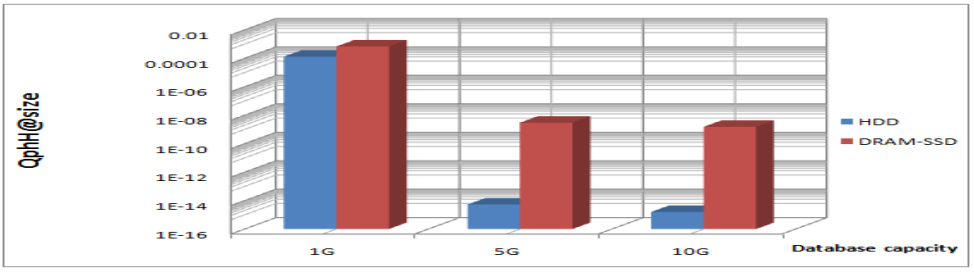

Figure 9.QphH@size value in the SAN environment

4.2.3. Results and analysis according to environment: Per hour ad-hoc query capabilities known good HDD storage and DRAM-SSD storage regardless of the database capacity in a SAN environment than the Local environment. In addition, DRAM-SSD storage can be known far better performance by increasing database capacity in SAN environment.

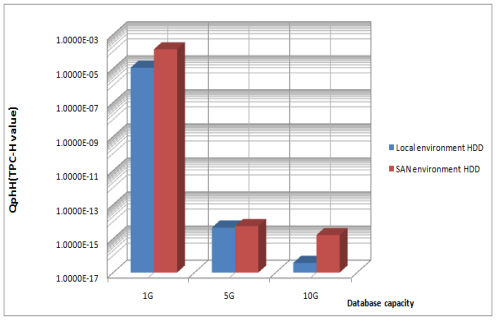

Figure 10.The QphH @ size value comparison of HDD storage in each of environment

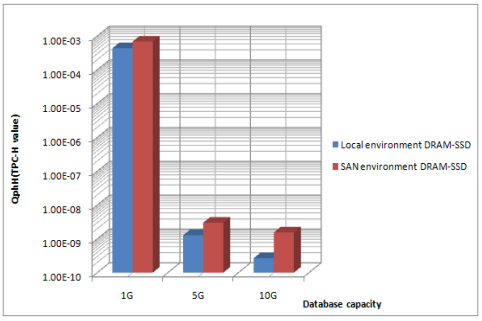

Figure 11. The QphH @ size value comparison of DRAM-SSD storage in each of environment

4.3. CO2 emission comparison

When power spend of each storage device comparative analyzed emission of CO2.

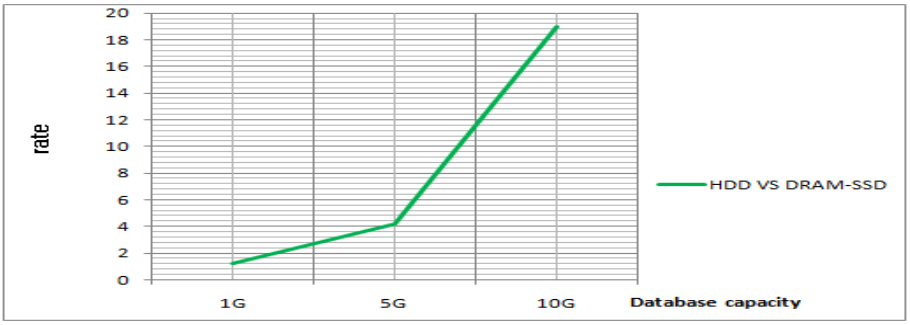

Figure12. CO2 occur ratio of HDD and DRAM-SSD

CO2 emissions analysis result during analysis of performance of each storage using Postmark and TPC-H, as shown in Figure 12 HDD storage can be considered less CO2 emissions due to the power consumption when non-load. But, increasing the load to DRAM-SSD storage generated less CO2 emission.

4.4. Analysis and comparison of power consumption

At this section analyzed power consumption having a test results by attaching the measuring instrument.

4.4.1. Using a Postmark benchmark: Based on the results of Table 4 and Table 5, case of low data I/O occurs, power consumption of HDD storage evaluate less occurs by difference of non-load power consumption, but, power consumption of DRAM-SSD is less than HDD storage to difference of run time by increasing database capacity.

Table 4. Power consumption measurement results

Postmark Test Level	Test result(W)		
	HDD	DRAM-SSD	Rate
Non-load	263.8	475.7	0.55
Low	264.18	476.55	0.55
Medium	264.63	484.74	0.55
High	281.26	508.01	0.55

Table 5. Total power consumption comparison

Postmark Test Level	Test result(run time X power W)		
	HDD	DRAM-SSD	Rate
Low	3,962.7	7,148.25	0.55
Medium	40,488.39	16,481.16	2.46
High	109,128.88	37,760.38	2.89

4.4.2. Using a TPC-H: As can be seen from the result of Table 6 and Table 7, like the preceding the difference of power consumption of HDD and DRAM-SSD is can be seen increased by increasing database capacity.

Table 6.Power consumption of each of storage in loads

Database capacity	Test result(W)		
	HDD	DRAM-SSD	Rate
Non-load	262.81	423.85	1.61
1GB	263.76	424.03	1.61
5GB	264.66	424.79	1.61
10GB	265.37	425.89	1.60

Table 7. Total power consumption

Database capacity	Test result:Kw		
	HDD	DRAM-SSD	Rate
1GB	5513.638	7009.741	0.79
5GB	59334.309	22801.186	2.60
10GB	161237.434	13614.083	11.84

5. Conclusion

In this paper, we show the performance analysis results by looking for experimentin many ways DRAM-SSD and HDD.As a resultsof a postmark benchmark and TPC-H analyzed about the HDD storage and DRAM-SSD storage for performance analysis, the performance difference of HDD and DRAM-SSD was little in low data I/O. But, the DRAM-SSD had a better performance than the HDD in large amounts of data I/O.

In addition, the database become larger than10G, DRAM-SSD had less power consumption and emissions about 92% than the HDD in analysis of power consumption and CO2 emissions. Also, performance of HDD and DRAM-SSD showedimprove in using SAN environment than the general environment.

In the future, the new direction is needed through the more reliable comparison of performance with SSD storage system for to implement the testing. Thus, such tests respect to actually use mail Service, VOD Service. So, besides SAN, we should confirm the advantages of the SSD through new analysis by DAS or NAS configuration of an environment.

Acknowledgements

This paper has been supported by 2013Hannam University Research Fund.

References

[1] K. -H. Park, J. -k. Choe, Y. -H. Lee, S. -K. Cheong and Y. -S. Kim, “Performance Analysis for DRAM-based SSD system”, Korean information technical academic society dissertation, vol. 7, no. 4, **(2009)**, pp. 41–47.
[2] Y. -H. Kang, J. -H. Yoo and S. -K. Cheong, “Performance Evaluation of the SSD based on DRAM Storage System IOPS”, Korean information technical academic society dissertation, vol. 7, no. 1, **(2009)**, pp. 265–272.

[3] S. -K. Cheong, Y. -W. Jeong, Y. -J. Jeong and J. -j. Jeong, "Input-Output Performance Analysis of HDD and DDR-SSD Storage under the Streaming Workload", Korean information technical academic society dissertation, **(2010)**, pp. 322-325.
[4] J. Judd and C. Beaucbamp, "Building SANs Witch Brocade", syngess, **(1996)** November, pp. 9-16.
[5] H. Kim, LearntoMYSQLDatabaseProgramming, Young-jin.com.
[6] http://www.tpc.org/tpch/spec/tpch2.13.0.pdf.
[7] PostMark: A New File System Benchmark <Jeffrey Katcher>.
[8] http://epoch11.blog.me/140107554392.
[9] H.-j. Song, Y. -h. Lee, "CO2 & power consumption comperision Analysis of HDD and DRAM-SSD", UCMA 2012 proc., **(2012)**, pp. 147-151.
[10] A. Oudah, T. Abd.Rahman and N. Seman, "On The Impact of MIMO Antennas on Collocation and Coexistence Requirements of LTE Networks in 2.6 GHz Frequency Band", IJMUE, vol. 8, no. 1, **(2013)** January, pp. 1-10.
[11] D. -M. Seo, N. -h. Kim, D. -H. Lee, J. -D. Kim and J. S. In, "Social Network Service-based Impact Analysis of CustomerRequirements", IJMUE, vol. 7, no. 4, **(2012)** October, pp. 245–252.
[12] H. I. Abdalla, "A New Data Re-Allocation Model for Distributed DatabasSystems", IJDTA, vol. 5, no. 2, **(2012)** June, pp. 45–60.

Authors

Hyun-Ju Song
2010. Electronics Engineering, Hannam University
2012. M.S. Electronics Engineering, Hannam University
Present. Clerk of Strategic Business Development in DB Communication & Systems
*major Field: SSD, Server

Young-Hun Lee
1973. Electronics Engineering, Donga University
1980. M.S. Electronics Engineering, Soongsil University
1984. Ph. D. Electronics Engineering, KyoungHee University
Present. Professor in Electronics Engineering, Hannam University
*major Field: Hybrid Circuit Design