Mutiara: Multidisciplinary Scientific Journal
Volume 2, Number 11 November 2024
p-ISSN 2988-7860 ; e-ISSN 2988-7992

Performance Comparison of Hybrid Storage on OLAP Database Server Virtualization

Ferdy Nirwansyah
Binus University, Indonesia
E-mail: nirwansyah@binus.ac.id

KEYWORD
database; OLAP; hybrid storage; TPC-H benchmark; virtualization

ABSTRACT
In today's digital era, efficient database management is important for companies to support analysis and decision-making processes. The use of On Line Analytical Processing (OLAP) in databases requires optimal storage configuration, especially in virtualization environments. This study aims to evaluate the performance of OLAP databases using hybrid storage configurations, combining Solid State Drives (SSD) and Hard Disk Drives (HDD), and identify the best configuration to improve database performance. The research method was carried out using the TPC-H benchmark on four database systems: Oracle, SQL Server, MySQL, and PostgreSQL. Four storage configurations were tested by allocating SSD for database storage and HDD for the operating system and temporary files, measured with 10 and 100 virtual users. The results showed that SQL Server provided the best performance in all configurations, with faster query execution times than other databases. The configuration with SSD for database storage and HDD for the operating system and temporary files proved to be the most efficient in reducing query latency. The conclusion of this study emphasizes the importance of using hybrid storage configurations in improving OLAP database performance in virtualization environments. Configuring with SSD as database storage and HDD for operating system and temporary files is the best choice to achieve high query performance.

INTRODUCTION

Database as an organized collection of data, structured to support efficient access, management, and retrieval. Databases enable users to store and interact with large amounts of information, ensuring data integrity, security, and consistency across various applications. Database becomes an integral part of our daily life. In the modern business world, the database became operational support company called On Line Transaction Processing (OLTP). The database is also used to assist companies in analyzing and making decisions, known as the On Line Analytical Processing (OLAP).

Database has transaction data and supporting data needed. More the number of transactions on the application and the database, more complex the application and database infrastructure that must be owned by the company. Performance tuning for database storage involves optimizing resource use and improving efficiency. Key methods include:

1. Indexing: Use indexes on frequently queried fields to speed up data retrieval.
2. Partitioning: Split large tables into smaller parts to enhance query performance.
3. Normalization: Organize data to reduce redundancy and improve consistency.
4. Caching: Store frequently accessed data in memory to minimize disk access.
5. Disk Optimization: Defragment disks and choose efficient file systems for faster data access.

These strategies ensure optimal storage and system performance, balancing resource usage. Along with the development of technology, high availability can be improved with virtualization. Management of server hardware and software becomes more practical. Some of the features that can be done are: management of hardware resources to each OS in a VM, for example, processor, RAM and storage, monitoring resource allocation to each OS and others.

At the storage level, we recognize several types of disk, namely: Solid State Drive (SSD) and magnetic drive (SAS and SATA), sometimes called the Hard Drive (HDD). SSD is a revolutionary new storage technologies and a positive impact on the performance of system and database. Investment of enterprise storage system that uses full SSD storage system is still very expensive, on the other hand there are many enterprise storage system using HDD as legacy system.

With the condition of existing storage, it influenced the design of the IT infrastructure in enterprise storage systems. In order for the utilization of all storage resources and the resulting optimal performance, there should be between SSD and HDD configuration. This also applies to the configuration of storage on the database server. This study will do a hybrid configuration database storage on the virtualization server.

To get the best possible configuration should be performed benchmark. The criteria for a good benchmark for the performance are as follows: first, representative, second, relevant, third, portable, fourth, scalable, fifth, and sixth verifiable, simple. There are some benchmarks that reach active industry standards. The most commonly used is the TPC and SPEC. For this study, using a TPC benchmark performance measurement. Because the focus of research in the Database OLAP will use the TPC-H. On the other hand, to complete this research, performance measurement testing of four databases are ORACLE, SQL Server, MySQL and PostgreSQL.

This study aims to evaluate the performance of the OLAP (Online Analytical Processing) database system on various hybrid storage configurations that combine Solid State Drives (SSD) and Hard Disk Drives (HDD). In addition, this study also aims to identify the storage configuration that provides the best performance in a virtualization environment, with a focus on reducing query latency and optimizing overall database performance. This study compares the performance of four popular database systems, namely Oracle, SQL Server, MySQL, and PostgreSQL, in the context of hybrid storage.

The benefits of this study include improving database performance by providing insights into how proper storage configuration can improve database performance in OLAP applications, which is critical for business decision making. This study also provides guidance for IT practitioners in selecting and configuring efficient storage systems for virtualization environments. In addition, the results of this study can be a basis for further research on the effects of RAID configurations and other storage systems on database performance at a larger scale. Finally, this study contributes to the development of better hybrid storage technologies, considering cost and efficiency, which can be adopted by companies to improve their IT infrastructure.

METHODS

Step of Research

The steps of this study : study of literature, installation of the instrument, create a database and then loading data of fourth database for TPC-H testing as well as the replacement scheme configuration of the virtual drive, collecting the test data, the performance evaluation system configurations and then the conclusions and suggestions.

First step, the research begins by determining the background and purpose of the study as well as defining the scope. The literature study is done to deepen the understanding of the hybrid technique of virtual disks to virtual drives on Windows Server virtualization server. In addition, a literature study was also conducted to find out the results of hybrid storage technique has ever done.

Second step, the research software installation of

1. VMWare
2. Windows Server 2019
3. HammerDB
4. SQL Server
5. Personal ORACLE
6. MySQL
7. PostgreSQL

Hardware for the study were:

1. Intel Modular Server Chassis MFSYS25V2: 14 Drive carriers, 1 GbE switch, two power supplies, two power supply fan.
   - Node I: Intel MFS2600KI Compute Module, 2 x Intel (R) Xeon (R) E5-2660 CPU 0 @ 2.20GHz 8 Cores, 24 GB DDR3 RAM.
2. 2 Pieces hard drive Seagate Savvio HDD 300 GB 10K RPM 2.5 ".
3. 4 pieces harddisk Fource CORSAIR GS 240GB SSD 2.5 ".
4. 8 pieces SATA hard drives Seagate Momentus 500 GB 7.2K RPM 2.5 ".
5. Switch hub Cisco SG500-28 28 ports.

First, install VMWare on the server, then installation of 4 pieces disk SSD using RAID 10 for the virtual disk drive OS Windows Server in a VM and 8 pieces harddisk HDD using RAID 10 as a virtual disk drive for database ORACLE, SQL Server, MySQL and PostgreSQL. Then installation of Windows Server as a Virtual Machine (VM) in the VMWare virtual disk drive in settings using a configuration that has been provided. After that, do installation HammerDB on Windows Server. Installation of the four databases, namely: ORACLE, SQL Server, MySQL and PostgreSQL. Configure the database so that each database is directed to SSD on drive E. Temporary OS and database files are directed to SSD on drive F. This preparation is done for testing in the first research. First research infrastructure schemes can be seen in Figure 1.

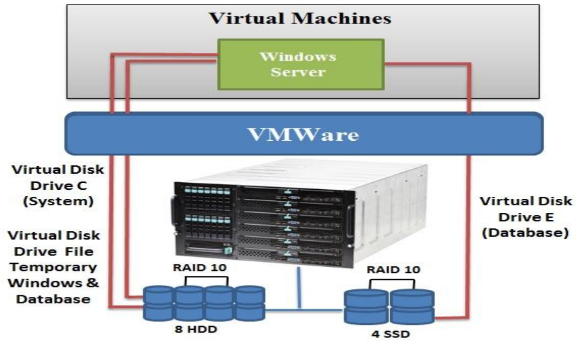


Fig 1. First Research Infrastructure Schemes

Third step, create databases and then loading data for testing. Backup image of Windows Server. The fourth step, testing the TPC-H. After completed, change virtual disk drive configuration by using a restore image of Windows Server. Make configuration of the hard disk with SSD for OS and the HDD for database . Temporary files OS and database are directed to HDD in drive F. Install a VM from a backup image before. After that do TPC-H test similar to previous research. The second research infrastructure schemes can be seen in Figure 2 .

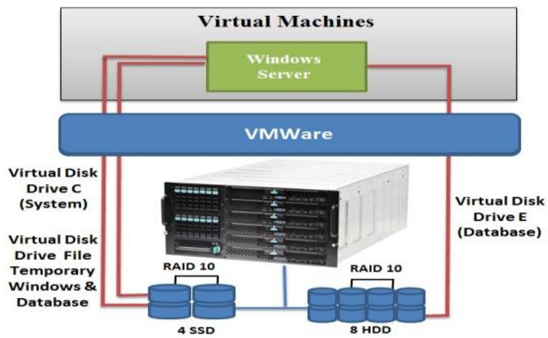


Fig 2. Second Research Infrastructure Schemes

Next third research. Make configuration of HDD for OS and SSD for database. Temporary files OS and database are directed to the SSD on drive F. After that do TPC-H test similar to previous research. Third research infrastructure schemes can be seen in Figure 3.

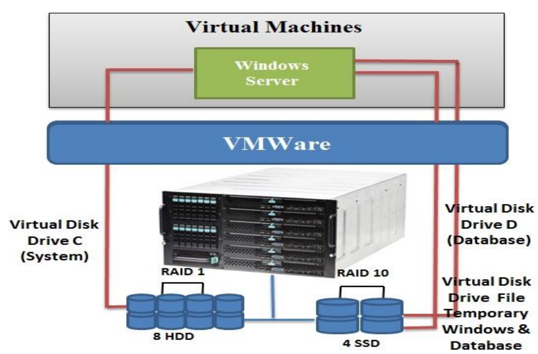


Fig 3. Third Research Infrastructure Schemes

After that go to fourth research. Make configuration of SSD for OS and HDD for database. OS and database temporary files are directed to the HDD in drive F. Do the testing of TPC-H similar to previous research. fourth research infrastructure schemes can be seen in Figure 4.

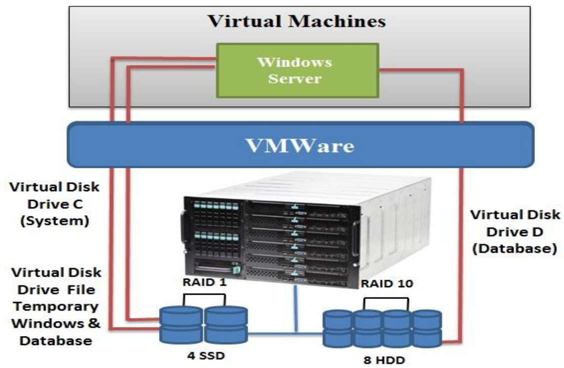


Fig 4. Fourth Research Infrastructure Schemes

Used Method

Used method for this study from create databases to output generated is using HammerDB . There are four databases that will be examined, ie ORACLE XE, SQLServer , MySQL and PostgreSQL. There are four configurations of virtual disk drives to be tested, first, a virtual disk drive for the OS using SSDs and virtual disk drive database using the HDD but the temporary files of Windows and database using the HDD, second, virtual disk drive for the OS using the HDD and virtual disk drive database using SSD but the temporary files of Windows and database using the HDD, third, virtual disk drive for the OS using the HDD and virtual disk drive database using the SSD but the temporary files of Windows and database using SSD, fourth, virtual disk drive for the OS using SSDs and virtual disk drive database using HDD but Windows temporary files and databases using SSD. There are two schemes used by HammerDB, TPC-H for OLAP.

Methods of data collection is to record the results of tests on four databases with four virtual drive configuration of two different schemes. TPC-H is calculated based QPhH of 22 queries are executed. At each TPC scheme conducted two experiments on each disk configuration and database. TPC-H using SF 1 for 10 and 100 virtual user and 1 query set. In each test produce results in log files. On PostgreSQL, TPC-H for query 17, 20, 21 did not include in test because it requires quite a long time for execute the query.

RESULT AND DISCUSSION

Four studies conducted for OLAP database schema using the TPC-H test results, obtained are illustrated by the graph on SF 1 with 10 virtual users and SF 10 with 100 virtual user. Table 1 shows the test results of TPC-H SF 1 with 10 and 100 virtual users.

Table 1. Table of TPC-H SF 1 for 10 and 100 Virtual Users

Result TPC-H in QppH

10 Users				
Database	**1st Config**	**2nd Config**	**3rd Config**	**4th Config**
ORACLE	629	228	664	328
SQL Server	7.733	3.334	7.539	2.849
MySQL	503	203	512	208
PostgreSQL	2.456	1.158	2.595	1.615

100 Users				
Database	**1st Config**	**2nd Config**	**3rd Config**	**4th Config**
ORACLE	22.22	41.55	25.34	46.32
SQL Server	1006.31	985.68	959.60	874.73
MySQL	56.32	37.81	54.58	47.23
PostgreSQL	287.28	21.29	183.61	24.29

Discussion

10 Users

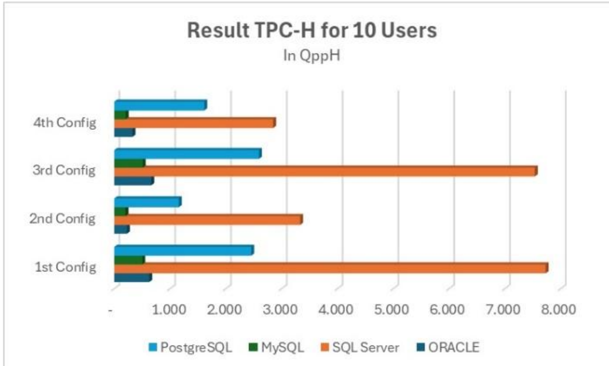
The data provided shows the performance of four different database systems—Oracle, SQL Server, MySQL, and PostgreSQL—across four distinct configurations. SQL Server consistently demonstrates the highest performance across all configurations. The performance of SQL Server peaks at 7.733 in the 1st configuration and declines to 2.849 in the 4th configuration. Despite this decrease, it remains the top performer among the databases tested.

Oracle's performance varies significantly. It reaches its peak in the 3rd configuration with 664 and drops to 228 in the 2nd configuration, indicating fluctuations in performance depending on the system configuration.

MySQL also shows moderate variability in its performance. The 1st configuration records 503, while the 3rd configuration is slightly better at 512. However, the 2nd and 4th configurations show lower results, 203 and 208, respectively.

PostgreSQL follows a similar pattern, with the highest performance in the 3rd configuration at 2.595 and the lowest in the 2nd configuration at 1.158. The other configurations (1st and 4th) show intermediate results. Based on the provided data, here's the ranking of the configurations from highest to lowest performance:

1. 1st Configuration performs the best overall, with SQL Server leading by a large margin. PostgreSQL, Oracle, and MySQL follow with solid performance, but SQL Server stands out significantly.
2. 2nd Configuration ranks the lowest overall, with SQL Server still leading but with the biggest drop in performance. PostgreSQL, Oracle, and MySQL perform poorly, with MySQL showing the weakest performance here.
3. 3rd Configuration is also high-performing, with SQL Server again in the lead. PostgreSQL shows the highest performance for this configuration, followed by Oracle and MySQL, which still perform well, but not as much as in the 1st configuration.
4. 4th Configuration sees a noticeable decline in performance across all databases. SQL Server is still the top performer, but the gap between it and the other databases is much narrower compared to the previous configurations. Graph 1. shows Result TPC-H For 10 Users on each disk configuration.



Graph 1. Of TPC-H Per Database For 10 Users

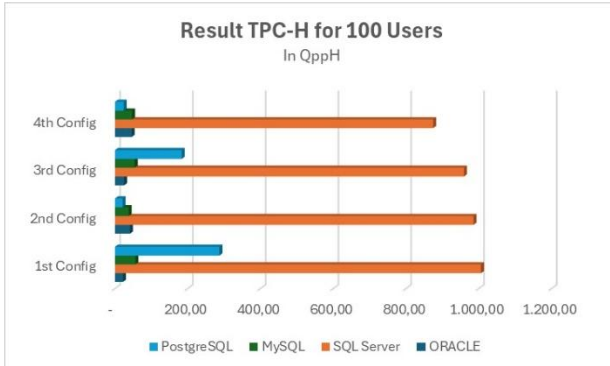
IV.2.1 100 Users

From the data, we can observe that SQL Server consistently shows the highest values across all configurations, with figures ranging from 874.73 to 1,006.31. In comparison, PostgreSQL performs the weakest, especially in the 2nd and 4th configurations, with values as low as 21.29 and 24.29, respectively. MySQL's performance is more consistent, with values ranging from 37.81 to 56.32, showing mid-range performance compared to the other databases. Oracle's performance fluctuates, showing a slight decrease in performance from the 1st configuration (22.22) to the 4th configuration (46.32), but generally stays within a moderate range.

Overall, SQL Server outperforms all other databases across all configurations, followed by Oracle, MySQL, and PostgreSQL in terms of performance. These results suggest that SQL Server might be more efficient or suited for the given configurations, while PostgreSQL appears to underperform. The choice of database may depend on specific use case requirements, with SQL Server being a more robust option in this instance.

Based on the provided data, here is the ranking of configurations from best to worst, considering the highest performance values across all four database systems:

1. 1st Configuration. This configuration performs the best, particularly due to SQL Server's high value of 1,006.31. It also shows moderate results from Oracle (22.22), MySQL (56.32), and PostgreSQL (287.28).
2. 2nd Configuration. This configuration performs the lowest, with SQL Server still leading at 985.68, but the very low results from PostgreSQL (21.29) and MySQL (37.81) drag it down in the overall ranking. Figure IV.3 shows Result TPC-H For 10 Users on each disk configuration
3. 3rd Configuration. This configuration ranks second, with SQL Server showing a strong value of 959.60. MySQL and Oracle also perform reasonably well, but PostgreSQL again shows a lower value (183.61), which affects its overall standing.
4. 4th Configuration. This configuration sees a drop in SQL Server performance (874.73), but it still ranks higher than the 2nd configuration. MySQL (47.23), Oracle (46.32), and PostgreSQL (24.29) all show reduced performance compared to the 1st and 3rd configurations.



Graph 2. Of TPC-H Per Database For 100 Users

CONCLUSION

In 10 Users, SQL Server outperforms the other databases across all configurations. Oracle, MySQL, and PostgreSQL show more noticeable fluctuations in performance depending on the configuration, with PostgreSQL performing relatively well overall but with significant variability. Base on configuration, The 1st Configuration performs the best, followed by the 3rd Configuration. The 4th Configuration ranks third, and the 2nd Configuration has the lowest performance. In 100 Users, SQL Server outperforms all other databases across all configurations, followed by Oracle, MySQL, and PostgreSQL in terms of performance. These results suggest that SQL Server might be more efficient or suited for the given configurations, while PostgreSQL appears to underperform. The choice of database may depend on specific use case requirements, with SQL Server being a more robust option in this instance. Base on configuration, The 1st Configuration performs the best across all systems, followed by 3rd Configuration, 4th Configuration, and finally the 2nd Configuration. Overall for 10 and 100 users, best configuration on 1st Configuration which is used HDD with RAID 10 for OS and Temporary Windows File and SSD with RAID 10 For Database. However, this experiment will different when we use another RAID configuration an also kind of storage. The ratio of the hard drive is used in this study is 1:2. But these results will not be relevant when used for the ratio was raised to 1 : 3 or 1 : 4 and more. Due to the increase in the ratio of use of hard disk, the speed of the HDD storage will be increasingly offset by the performance of SSD storage.

BIBLIOGRAPHY

T. Connolly and C. Begg, “Database Systems.”

R. Schiesser, “IT systems management,” pp. 110–126, 2010.

Ouared, A., Amrani, M., Chadli, A., & Schobbens, P. Y. (2024). Deep variability modeling to enhance reproducibility of database performance testing. *Cluster Computing*, *27*(8), 11683-11708. https://doi.org/10.1007/s10586-024-04533-0

George, A. S., Srikaanth, P. B., Sujatha, V., & Baskar, T. (2023). *Flash Fast : Unleashing Performance with NVMe Technology Partners Universal International Research Journal (PUIRJ)*. *September*, 71–81. https://doi.org/10.5281/zenodo.8350245

Huang, S., Qin, Y., Zhang, X., & Tu, Y. (2023). *Survey on performance optimization for database systems Survey on performance optimization for database systems*. *January*. https://doi.org/10.1007/s11432-021-3578-6

Zhang, C., Li, G., Zhang, J., Zhang, X., & Feng, J. (n.d.). *HTAP Databases : A Survey*. *ii*.

B. Mao, H. Jiang, S. Wu, L. Tian, D. Feng, J. Chen, and L. Zeng, "Hpda," *ACM Trans. Storage*, vol. 8, no. 1, pp. 1–20, 2012.

Y. Bassil, "A Comparative Study on the Performance of the Top DBMS Systems," *arXiv Prepr. arXiv1205.2889*, pp. 20–31, 2012.

Y. Kim, A. Gupta, B. Urgaonkar, P. Berman, and A. Sivasubramaniam, "HybridStore: A cost-efficient, high-performance storage system combining SSDs and HDDs," *IEEE Int. Work. Model. Anal. Simul. Comput. Telecommun. Syst. - Proc.*, pp. 227–236, 2011.

D. Bausch, I. Petrov, and A. Buchmann, "On the performance of database query processing algorithms on flash solid state disks," *Proc. - Int. Work. Database Expert Syst. Appl. DEXA*, pp. 139–144, 2011.

J. Do, D. Zhang, J. M. Patel, D. J. DeWitt, J. F. Naughton, and A. Halverson, "Turbocharging DBMS buffer pool using SSDs," *Proc. 2011 Int. Conf. Manag. data - SIGMOD '11*, p. 1113, 2011.

R. Shaull, T. Ron, and A. Littman, "Enterprise Storage Provisioning with Flash Drives," 2010.

H. Jo, Y. Kwon, H. Kim, E. Seo, J. Lee, and S. Maeng, "SSD-HDD-hybrid virtual disk in consolidated environments," *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 6043 LNCS, no. 2009, pp. 375–384, 2010.

X. Wu and A. L. N. Reddy, "Managing storage space in a flash and disk hybrid storage system," *Proc. - IEEE Comput. Soc. Annu. Int. Symp. Model. Anal. Simul. Comput. Telecommun. Syst. MASCOTS*, pp. 610–613, 2009.

S.-W. Lee, B. Moon, and C. Park, "Advances in flash memory SSD technology for enterprise database applications," *Proc. 35th SIGMOD Int. Conf. Manag. data SIGMOD 09*, vol. 14, no. 3, pp. 863–870, 2009.

K. Park, D. H. Lee, Y. Woo, G. Lee, J. H. Lee, and D. H. Kim, "Reliability and performance enhancement technique for SSD array storage system using RAID mechanism," *2009 9th Int. Symp. Commun. Inf. Technol. Isc. 2009*, pp. 140–145, 2009.