JURNAL SYNTAX ADMIRATION

Syntax Admiration: p-ISSN 2722-7782 | e-ISSN 2722-5356 Vol. 5, No. 10, October 2024

Comparison Performance Analysis of HDD, SSD, and NVME for OLTP Database Server

Ferdy Nirwansyah
Universitas Bina Nusantara, Indonesia
Email: ferdy.nirwansyah@binus.ac.id

Abstract

Databases are an essential component of life in modern society. The database has become an online Transaction Processing (OLTP) operational support company. This study discusses the storage configuration in a virtual server database for OLTP Database with a benchmark against three storage configurations for four databases: SQL Server, MySQL, PostgreSQL, and MariaDB for Windows Server virtualization. Benchmarks using TPC-C to get the best performance from all three configurations have been tested. The results of this study show that the storage configuration of HDD as a virtual disk drive OS and NVME as a virtual disk drive database get better performance as online transaction processing (OLTP) compared to HDD as a virtual disk drive OS and HDD or SSD as a virtual disk drive database. Using the TPC-C benchmark, the study tested the performance of four databases (SQL Server, MySQL, PostgreSQL, and MariaDB) in various virtual storage configurations. The results of 36 experiments involving different numbers of virtual users (10, 50, and 100) showed that a hybrid configuration that used HDDs for the operating system and NVME for database storage provided the best performance. On SQL Server, SSD configurations show better results for smaller user loads. The results suggest that the combination of HDD and NVME is the optimal configuration for most databases in OLTP tasks, and it provides useful guidance in efficient storage investment decision-making. Further research is expected to be able to conduct tests.

Keywords: storage configuration, database, OLTP, benchmark, TPC-c

Introduction

Databases and database systems are essential components of life in modern society: most of us encounter several activities every day that involve some interaction with a database (Kleppmann, 2017) (Elmasri, 2021). In the modern business world, the database has become an operational support company called online transaction processing (OLTP). The database is also used to assist companies in analyzing and making decisions, known as online analytical processing (OLAP). The database has transaction data and supporting data needed. The more the number of transactions on the application and the database, the more complex the application and database infrastructure that must be owned by the company (Laday et al., 2015). Infrastructure applications and databases are essential in order to improve high availability (Endo et al., 2016). To improve the performance done with performance and tuning, namely on: first, server environments such as mainboard, processor, RAM, LAN card, and others. Second, storage environment, third, database environment, fourth, network environment, fifth, desktop computer environment (Clark, 2014).

Along with the development of technology, high availability can be improved with virtualization. Management of server hardware and software has become more practical. Some of the features that can be done are the 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: Non-Volatile Memory Express (NVME), Solid State Drive (SSD), and magnetic drive (SAS and SATA), sometimes called Hard Disk Drive (HDD). NVME is a revolutionary new storage technology and has a positive impact on the performance of systems and databases. Investment in enterprise storage systems that use a full NVME storage system is still very expensive. On the other hand, there are many enterprise storage systems using systems as legacy systems.

Database servers store a database that is accessed either via the corporation network or via the Internet. Furthermore, the database must sometimes be connected to special application servers. Database servers have to simultaneously handle numerous parallel requests from various users, which makes low latency and high-reliability decisive factors. Multiple tiers of HDD and SSD can be effective for database servers to acquire better performance. In particular, PCIe SSD with NVMe Interface is the right choice to have high throughput rates and low latency. NVMe deploys flash storage on a PCIe bus and offers up to 64,000 parallel pathways from the CPU, thus overcoming serial limitations in storage I/O processing. The protocol capitalizes on the multiple parallel, low latency paths to flash storage to offer faster storage response times and higher throughput for speedier application performance. (Zhou et al., 2018)

The condition of existing storage 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 a difference between SSD and HDD configuration. This also applies to the configuration of storage on the database server. This study will do a hybrid database configuration on a virtualization server.

To get the best possible configuration, it should be performed benchmark. The criteria for a good benchmark for performance are as follows: first, representative, second, relevant, third, portable, fourth, scalable, fifth, and sixth verifiable, simple. There are some benchmarks that achieve active industry standards. The most commonly used are the TPC and SPEC. For this study, we used a TPC benchmark performance measurement. Because the focus of research is on the Database, OLTP will use TPC-C.

On the other hand, to complete this research, performance measurement testing of four databases is SQL Server, MySQL, PostgreSQL, and MariaDB. TPC-C benchmark, a widely used online transaction processing (OLTP) benchmark developed by the Transaction Processing Performance Council (TPC). Launched in 1992, TPC-C simulates order-entry business environments with various transactions, such as order entry, payment processing, and inventory management. The benchmark focuses on measuring performance in terms of transactions per minute (tpmC) and includes a comprehensive database schema to replicate real-world operations.

TPC-C is designed to enable fair performance comparisons across different hardware and software systems, adapting over time to incorporate technological advancements (Raasveldt et al., 2018) (Wang et al., 2022). Ouared, Amrani, Chadli, & Schobbens, (2024) Present a comparative study on improving the reproducibility of database performance tests, a crucial aspect for both developers and researchers. The authors introduce a deep variability modeling framework that leverages machine learning techniques to analyze complex interactions among various performance factors, including hardware configurations, workloads, and database settings.

By systematically varying parameters within a defined space, the proposed approach enhances the reliability of performance benchmarks. It allows for the identification of optimal configurations tailored to specific environments, thereby accounting for external factors that might influence performance, such as system load variations.

Experimental results demonstrate that this method produces consistent performance outcomes across different testing scenarios, significantly boosting confidence in the results. Ultimately, the paper contributes to establishing standardized performance testing methodologies, which are essential in the increasingly complex landscape of modern database systems. (Ouared et al., 2024).

He et al., (2023) Examines the performance issues that arise when modern databases interact with advanced storage technologies, such as SSDs and NVMe. The authors argue that traditional database configurations often fail to optimize the capabilities of these new storage devices, leading to significant performance bottlenecks. Through empirical studies, the paper identifies various performance mismatches between database workloads and the characteristics of contemporary storage solutions. The authors explore how factors like I/O patterns, data access methods, and storage latency contribute to inefficiencies. They propose a framework for understanding these interactions, which allows database administrators to analyze and adjust configurations to better align with the unique properties of modern storage devices.

The research includes practical recommendations for optimizing database performance, such as tuning caching mechanisms, adjusting query execution plans, and selecting appropriate storage tiers. By emphasizing the need for a holistic approach to database and storage management, the paper aims to help organizations improve performance and reduce latency in data-intensive applications. Ultimately, it calls for a more integrated perspective that recognizes the critical interplay between databases and storage technologies. (He et al., 2023).

George, Srikaanth, Sujatha, & Baskar, (2023) examine the transformative impact of the NVMe (Non-Volatile Memory Express) storage protocol on solid-state drives (SSDs). As flash-based SSDs have evolved, traditional interfaces like SATA and SAS have become performance bottlenecks, unable to leverage the full potential of modern hardware. NVMe addresses this by employing an efficient command interface that capitalizes on parallel processing capabilities.

The paper highlights NVMe's ability to handle numerous simultaneous I/O operations, significantly improving data access speed and efficiency. It showcases impressive performance metrics, such as achieving over 3,000 MB/s in sequential reads and up to 1,000,000 IOPS for random 4K operations, representing a 4-5x improvement over the fastest SATA SSDs. NVMe's design is particularly advantageous in multi-threaded environments and offers substantial benefits for applications requiring high performance, like real-time analytics and large database operations. The paper also discusses the accelerating adoption of NVMe in modern systems, while noting some compatibility challenges. (Zou et al., 2022).

Huang et al., (2023) Provides a comprehensive overview of various strategies to enhance the performance of database systems. It highlights key optimization techniques such as indexing, query optimization, and workload management. The survey discusses the trade-offs involved in these techniques, addressing how they can improve response times and throughput while managing system resources effectively. (Ramu, 2023).

Additionally, the paper reviews advancements in hardware technologies, including solid-state drives (SSDs) and memory systems, which play a crucial role in optimizing database performance. It emphasizes the importance of understanding the underlying hardware architecture to achieve better performance outcomes. The authors also explore emerging trends like machine learning and artificial intelligence, which offer new avenues for automated performance tuning.

Overall, the paper serves as a valuable resource for researchers and practitioners in the field, providing insights into both traditional and innovative approaches to database performance optimization. (Huang et al., 2023). Zhang, Li, Zhang, Zhang, & Feng, (2024) Which are designed to efficiently handle both transactional and analytical workloads. Traditionally, databases have been optimized for either Online Transaction Processing (OLTP) or Online Analytical Processing (OLAP), leading to the need for separate systems. HTAP databases seek to unify these functions, enabling real-time analytics on operational data without requiring complex ETL processes.

The survey discusses the architectural designs of HTAP systems, emphasizing their ability to integrate transactional consistency with high-performance analytical queries. It reviews various existing HTAP solutions, focusing on their approaches to balancing dual workloads and addressing challenges related to data consistency and scalability.

Additionally, the paper highlights the growing importance of HTAP databases in modern data-driven applications, particularly in sectors like finance and IoT, where timely decision-making is crucial. Overall, the survey underscores the potential of HTAP systems to transform how organizations leverage data for strategic advantages. (Zhang et al., 2024).

Kandula, (2018) Investigates the performance differences between solid-state drives (SSDs) and traditional magnetic disks (HDDs) in the context of SQL Server. The study employs a series of benchmarks to evaluate how each storage type impacts various database operations, including read and write speeds, transaction processing, and overall query performance.

This study aims to compare the performance of three types of data storage configurations (HDD, SSD, and NVME) in supporting the performance of OLTP (Online Transaction Processing) database servers that use virtualization on Windows Server systems. By using hybrid techniques in storage configurations, this study seeks to determine the best configuration that can improve the performance of OLTP databases in handling parallel data requests efficiently and reducing latency. Another key objective is to find out which storage configuration is most efficient in terms of performance and cost, taking into account technical factors and budget constraints.

The benefits of this study provide database administrators with insight into the best storage configurations that can improve database server responsiveness in transaction-intensive environments. The results of this study can help companies consider more effective infrastructure investments, either by maintaining old storage devices or adopting new technologies that are more suitable for the company's specific needs and this research adds to the literature and knowledge about the performance and efficiency of various types of data storage (HDD, SSD, NVME), especially in the context of their use in virtualized database servers.

Guide to Database Infrastructure Management: With these results, companies can optimize database infrastructure management to achieve high performance and optimal reliability, especially for applications that require fast transaction processing.

The findings reveal that SSDs significantly outperform magnetic disks in most scenarios, particularly in random read and write operations, which are common in database environments. The authors highlight that SSDs reduce latency and increase throughput, leading to improved application responsiveness and higher transaction rates. Additionally, the study examines the cost-effectiveness of implementing SSDs versus HDDs, taking into account factors such as longevity and maintenance.

The research concludes that while SSDs provide superior performance, the decision to switch from magnetic disks should consider budget constraints and specific application requirements. Ultimately, it advocates a strategic approach to storage selection in SQL Server environments, emphasizing the need to balance performance benefits with cost considerations. (Panwar, 2024).

Method Research

In this research, we will adopt a hybrid storage approach to efficiently utilize the combination of HDD and SSD, with the aim of achieving optimal performance on database servers hosted in a virtualized environment. NVME and SSD technologies that are revolutionizing storage have the potential to change the principles of DBMS architecture (Bassil, 2012). However, NVME and SSD itself are still quite expensive. On the other hand, the company's storage system still uses the HDD like the old system. In connection with this, questions will arise such as: 1. How to improve database server performance with optimal storage configuration on a virtualization server? 2. How do I utilize all available storage resources without reducing database server performance?

To utilize all existing hard disk resources, we use hybrid techniques. Hybrid technique of setting up virtual drives in Windows Server as a database server and part of VMWare server virtualization. This research creates a hybrid technique by configuring virtual disks on the hard disk drive for the OS and database using different storage. The advantage of using hybrid techniques in virtualization is in terms of practicality and convenience compared to doing hybrid at the storage level.

The steps of this research: literature study, instrument installation, creating a database, and then loading the fourth database data for TPC-C testing as well as virtual drive placement scheme configuration, testing data collection, system configuration performance evaluation, and then conclusions and suggestions.

The first step, research begins with determining the background and objectives of the research and determining the scope. A literature study was carried out to deepen understanding of the hybrid virtual disk to virtual drive technique in Windows Server server virtualization. Apart from that, a literature study was also carried out to find out the results of hybrid storage techniques that have been carried out.

First, install VMWare on the server, then install the HDD for the Windows Server OS virtual disk drive in the VM and the HDD as the virtual disk drive for the SQL Server, MySQL, PostgreSQL, and MariaDB databases. Then install Windows Server as a Virtual Machine (VM) on the VMWare virtual disk drive and set it up using the configuration provided. After that, I install HammerDB on Windows Server. Installation of four databases, namely: SQL Server, MySQL, PostgreSQL, and MariaDB. Configure the databases so that each database is directed to the HDD on drive D

Results and Discussion

From the research carried out by doing 36 experiments, research results were obtained which described the data obtained from the three configurations with the four databases tested. Figure 5 shows the table of research on NoPM.

SQL Server	10 User	50 User	100 User
1st Config	3.112	3.096	3.025
2nd Config	118.525	102.793	37.430
3rd Config	111.960	84.099	63.197

MySQL	10 User	50 User	100 User
1st Config	123	125	128
2nd Config	3.012	2.618	3.253
3rd Config	8.593	11.960	11.846

PostgreSQL	10 User	50 User	100 User
1st Config	8.318	6.145	5.423
2nd Config	66.613	37.839	6.606
3rd Config	69.136	39.069	12.359

MariaDB	10 User	50 User	100 User
1st Config	11.193	15.000	22.763
2nd Config	30.640	34.655	35.534
3rd Config	60.753	64.196	66.528

Figure 1. Table of Research

Three hard disk configurations for OLTP databases using the TPC-C scheme obtained a comparison of database performance for each configuration for virtual users. For 10 virtual users, configuration II was ranked first for SQL Server, then the second configuration III was also for SQL Server. Ranked third and fourth for PostgreSQL in Configurations II and III. The best performance is configuration III for all databases except SQL Server, where configuration II is much better by 6% compared to configuration III.

With 50 virtual users, it has the same results as 10 virtual users for first and second place. The best performance is also from SQL Server in configuration II, second from SQL server in configuration III. The third performance of MariaDB in configuration III. Ranked fourth of PostgreSQL on configuration. The best performance is configuration III for all databases except SQL Server, where configuration II is much better by 22% compared to configuration III.

In 100 virtual users, the best performance surprisingly came from MariaDB in configuration III, second to SQL Server in configuration III. Third, the performance of SQL Server in configuration II. MariaDB ranked fourth in configuration. The best performance is configuration III for all databases. Figure 6 shows the OLTP database performance on each disk configuration for all virtual users.

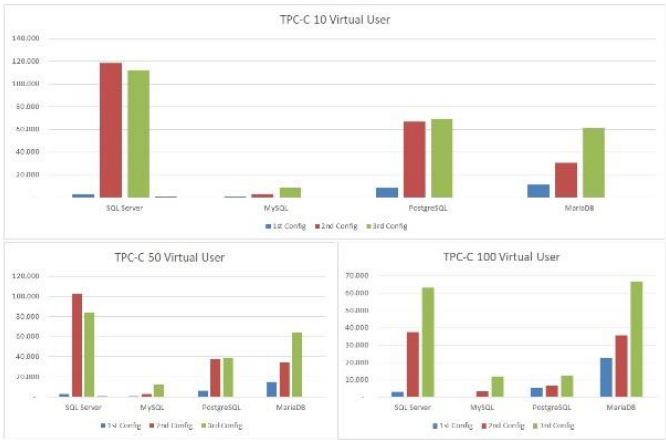

Figure 2. Graph of TPC-C per Virtual User

Three hard disk configurations for OLTP databases using the TPC-C scheme. A comparison of the database performance of each configuration was obtained. The performance of the SQL Server II configuration on 10 virtual users slightly outperforms configuration III. For 50 virtual users, the conditions are the same as for 10 virtual users. Configuration II slightly outperforms Configuration III. For 100 virtual users, Configuration III is much better at 68.84% than Configuration II.

MySQL occurs in different conditions, for 10 virtual users in configuration III, followed by configuration II. For 50 virtual users, configuration III is 356.84% superior to configuration II. At 100 virtual users, the winner is also configuration III. PostgreSQL on 10 virtual users, configuration III ranks first with a performance increase of 3.79% compared to Configuration II. At 50 virtual users, configuration III is 3.25% superior to configuration II. Of 100 virtual users, the winner was also Configuration III and the second for Configuration II. Overall, configuration III always has the best performance.

MariaDB on 10 virtual users, configuration III is ranked first, and the second performance is configuration II. With 50 virtual users, configuration III is 85% superior to Configuration II. Of 100 virtual users, the winner was also Configuration III and second for Configuration II. Overall, configuration III always has the best performance. Figure 7 shows the performance of the OLTP database on each disk configuration.

Figure 3. Graph of TPC-C per Database

Conclusion

Based on TPC-C data generated from this research, OLTP gets the best performance in the third configuration where the OS uses HDD, the database uses NVME. Except for SQL Server for virtual 10 and 50 users of the second configuration, which uses SSD, have better performance than the third configuration by 5.86% and 22.23%. The first configuration is not recommended for use as an OLTP database server. To get the best performance on the OLTP database server, you can use a hybrid configuration with configuration III. Further testing needs to be done on more and more virtual users. However, this test will require higher hardware than in this study.

This research reveals that SSD and NVME significantly outperform HDD in all scenarios, especially in write operations that occur in database environments. Additionally, the study did not examine the cost-effectiveness of implementing SSDs versus HDDs, considering factors such as longevity and maintenance. The research concludes that although SSD and NVME provide superior performance, the decision to switch from HDD must consider budget constraints and specific application requirements. Ultimately, we advocate a strategic approach to storage selection, emphasizing the need to balance performance benefits with cost considerations.

BIBLIOGRAPHY

Bassil, Y. (2012). A Comparative Study on The Performance of The Top DBMS Systems. *Arxiv Preprint Arxiv:1205.2889*.

Clark, R. L. (2014). *Leadership Trust in Virtual Teams Using Communication Tools: A Quantitative Correlational Study*. University Of Phoenix.

Elmasri, R. (2021). *Fundamentals of Database Systems Seventh Edition*.

Endo, P. T., Rodrigues, M., Gonçalves, G. E., Kelner, J., Sadok, D. H., & Curescu, C. (2016). High Availability In Clouds: Systematic Review and Research Challenges. *Journal of Cloud Computing*, *5*, 1–15.

George, A. S., Srikaanth, P. B., Sujatha, V., & Baskar, T. (2023). Flash Fast: Unleashing Performance with NVME Technology. *Partners Universal International Research Journal*, *2*(3), 71–81.

He, H., Xu, E., Li, S., Jia, Z., Zheng, S., Yu, Y., Ma, J., & Liao, X. (2023). When Database Meets New Storage Devices: Understanding and Exposing Performance Mismatches Via Configurations. *Proceedings of The VLDB Endowment*, *16*(7), 1712–1725.

Huang, S., Qin, Y., Zhang, X., Tu, Y., Li, Z., & Cui, B. (2023). Survey on Performance Optimization for Database Systems. *Science China Information Sciences*, *66*(2), 121102.

Kandula, V. (2018). *Analysis Of SSD's Performance in Database Servers*.

Kleppmann, M. (2017). *Designing Data-Intensive Applications: The Big Ideas Behind Reliable, Scalable, and Maintainable Systems*. " O'Reilly Media, Inc."

Laday, R. K., Sukoco, H., & Nurhadryani, Y. (2015). Distributed System and Multimaster

Replication Model on Reliability Optimation Database. *TELKOMNIKA Indonesian Journal of Electrical Engineering*, *13*(3), 529–536.

Ouared, A., Amrani, M., Chadli, A., & Schobbens, P.-Y. (2024). Deep Variability Modeling to Enhance Reproducibility of Database Performance Testing. *Cluster Computing*, 1–26.

Panwar, V. (2024). Optimizing Big Data Processing In SQL Server Through Advanced Utilization of Stored Procedures. *Journal Homepage: Http://Www. Ijmra. Us*, *14*(02).

Raasveldt, M., Holanda, P., Gubner, T., & Mühleisen, H. (2018). Fair Benchmarking Considered Difficult: Common Pitfalls in Database Performance Testing. *Proceedings of The Workshop on Testing Database Systems*, 1–6.

Ramu, V. B. (2023). Optimizing Database Performance: Strategies For Efficient Query Execution and Resource Utilization. *International Journal of Computer Trends And Technology*, *71*(7), 15–21.

Wang, Y., Yu, M., Hui, Y., Zhou, F., Huang, Y., Zhu, R., Ren, X., Li, T., & Lu, X. (2022). A Study Of Database Performance Sensitivity To Experiment Settings. *Proceedings Of The VLDB Endowment*, *15*(7).

Zhang, C., Li, G., Zhang, J., Zhang, X., & Feng, J. (2024). HTAP Databases: A Survey. *IEEE Transactions On Knowledge And Data Engineering*.

Zhou, D., Fang, V., Xie, T., Pan, W., Kesavan, R., Lin, T., & Patel, N. (2018). Empirical Evaluation and Enhancement of Enterprise Storage System Request Scheduling. *ACM Transactions On Storage (TOS)*, *14*(2), 1–27.

Zou, Y., Awad, A., & Lin, M. (2022). Directnvm: Hardware-Accelerated NVME SSDS for High-Performance Embedded Computing. *ACM Transactions on Embedded Computing Systems (TECS)*, *21*(1), 1–24.