STM JOURNALS

Recent Trends in
Parallel Computing

ISSN: 2393-8749
Volume 12, Issue 3, 2025
September–December
DOI (Journal): 10.37591/RTPC

https://journals.stmjournals.com/rtpc

Enhancing Data Processing and Storage in Computing Environments: A Survey on the Use of NVMe SSDs

Yasoda Krishna Reddy Annapureddy[1],*, V. Krishna Reddy[2]

Abstract

In the current era of fast pace technological growth, the efficiency of data processing and storage systems has become a key factor of various computing environments. This survey explores the transformative role of Non-Volatile Memory Express (NVMe) Solid-State Drives (SSDs) across different domains, including Big Data processing, Cloud Computing, High Performance Computing (HPC), and containerized applications. The motivation behind this comprehensive review is to understand how NVMe SSDs, known for their superior speed and lower latency compared to traditional storage solutions, are redefining the landscape of data-intensive tasks. Through the analysis of five pivotal research papers, this survey encapsulates the evolving dynamics of NVMe SSDs in enhancing performance, addressing challenges and laying the foundation for next breakthroughs in storage technology. The key findings highlight NVMe SSDs' significant impact on improving throughput, reducing latency, and optimizing resource utilization in various settings. Additionally, the survey identifies existing bottlenecks and provides insights into potential areas of advancement. This survey underscores the essentialness of NVMe SSDs in contemporary and future computing infrastructures, indicating their importance for faster, more efficient data processing and storage solutions.

Keywords: NVMe SSDs, data-intensive computing, high performance computing (HPC), cloud storage optimization, big data processing

INTRODUCTION

Background

The NVMe SSDs is a breakthrough in the evolution of data storage technology. NVMe SSDs are designed to take advantage of the high-speed PCIe bus in modern computers, delivering a significant performance leap over traditional storage technologies like HDDs and SATA SSDs. This advanced storage solution has quickly become essential in modern computing for its ability to efficiently manage large volumes of data.

NVMe SSDs are specially recognized by their high throughput and low latency, making them ideally suited for environments where rapid data access and processing are very high. Unlike traditional HDDs, which use spinning disks to read and write data, and SATA SSDs, limited by older interface standards, NVMe SSDs utilize NAND flash memory without any moving parts. This fundamental difference in architecture and connectivity allows NVMe SSDs to deliver faster read/write speeds and reduced data access times.

The significance of NVMe SSDs extends across various domains of modern computing. In enterprise environments, they facilitate quick data retrieval and efficient handling of large-scale workloads, essential for big data analytics and

***Author for Correspondence**
Yasoda Krishna Reddy Annapureddy
E-mail: yasodaakrishna@gmail.com

[1]Student, Department of Computer Science engineering, Koneru Lakshmaiah education Foundation, Vaddeswaram, Andhra Pradesh, India
[2]Professor & Principal, Department of Computer Science Engineering, Koneru Lakshmaiah education Foundation, Vaddeswaram, Andhra Pradesh, India

Received Date: June 06, 2025
Accepted Date: June 23, 2025
Published Date: October 30, 2025

Citation: Yasoda Krishna Reddy Annapureddy, V. Krishna Reddy. Enhancing Data Processing and Storage in Computing Environments: A Survey on the Use of NVMe SSDs. Recent Trends in Parallel Computing. 2025; 12(3): 1–21p.

cloud-based services [1]. In the realm of high-performance computing (HPC), NVMe SSDs are crucial for handling compute-intensive tasks, such as scientific simulations and data-intensive research, where every millisecond of processing time counts [2].

In addition, the demand for NVMe SSDs is growing as container applications and microservices architectures become more common in software development [3]. In this context, NVMe SSDs provide the required IOPS performance in order to efficiently run multiple applications and services simultaneously without creating a bottleneck on system performance.

The integration of NVMe SSDs into these diverse computing environments is not just a trend but a response to the growing data-centric demands of the modern world. As we generate and process data at a rapid scale, the role of NVMe SSDs becomes increasingly central in driving efficiency, speed, and reliability in data storage and access, setting a new standard in the storage solutions.

MOTIVATION

The motivation for integrating NVMe SSDs into modern computing systems stems from the ever-increasing demand for performance optimization in data processing and storage. As we know the three V's (volume, velocity, variety) of data grow exponentially, traditional storage technologies are unable to catch the high-speed demands of the regular applications. NVMe SSDs are the best solution in this situation, primarily due to their ability to significantly enhance the system performance. The following factors have motivated me to research on this topic.

- *High Throughput and Low Latency:* NVMe SSDs provide high data transfer speeds and lower latency compared to regular HDDs and SATA SSDs. This performance boost is because of NVMe's direct connection to the system's PCIe bus, allowing for faster data transfer rates. In environments where time is critical, such as financial transactions or real-time data analytics, the speed of NVMe SSDs can dramatically enhance application performance and user experience.
- *Improved Data Access and Processing:* The advanced architecture of NVMe SSDs enables more efficient data access and processing. This is particularly beneficial for applications that require frequent read/write operations, such as database management and big data analytics. NVMe SSDs can handle large volumes of data more effectively, reducing bottlenecks in data access and processing.
- *Scalability and Flexibility:* The superiority of NVMe SSDs support a higher number of input/output operations per second, which is the most important factor required for scaling rapid applications based on user usage and the services. This is especially important in cloud data centers and virtualized environments where resources are dynamically allocated and shared across various applications. In such setups, the ability of NVMe SSDs is to deliver consistent performance even under heavy loads.
- *Energy Efficiency and Reliability:* NVMe SSDs specifically consume less power and have no moving parts, which leads to lower energy costs, reduced heat generation and less maintenance. This not only makes them more environmentally friendly but also enhances their reliability.
- *Support for Advanced Technologies:* The modern evolving technologies such as machine learning (ML), the Internet of Things (IoT), and artificial intelligence (AI) require rapid data processing and storing, so the need for high-speed, efficient storage becomes more critical. NVMe SSDs are the best solution to support these technologies, enabling faster data processing and real-time data analysis.

SCOPE

- *Big Data Processing:* Understanding how NVMe SSDs enhance the handling and analysis of large datasets, particularly in terms of speed and efficiency, crucial for data-intensive tasks in fields like business analytics, healthcare, and scientific research [4].

- *Cloud Computing:* Examining the role of NVMe SSDs in cloud environments, where they are used to improve data storage, retrieval, and processing capabilities, thus enhancing the performance of cloud services and applications [1].
- *High-Performance Computing (HPC):* Assessing the benefits of NVMe SSDs in HPC setups, where their highspeed data processing capabilities are essential for complex computations, such as simulations, modeling, and other resource-intensive tasks [2].
- *Containerized Applications:* Investigating how NVMe SSDs support the performance needs of containerized applications, particularly in environments using technologies like Docker and Kubernetes, where multiple applications and services are run concurrently [3].
- *Storage Systems:* Getting into the advancements of NVMe SSDs bring to storage systems, including improvements in data throughput, latency, and overall system efficiency, as well as their role in the evolving landscape of storage solutions [5].

OBJECTIVES

- *Comprehensive Analysis:* Our goal is to conduct an in-depth examination of selected research papers, delving into the methodologies employed, the design and execution of experiments, and the significant findings that emerge. This analysis focuses specifically on the application and performance of NVMe SSDs in diverse computing environments, ranging from cloud computing to high-performance and containerized environments.
- *Synthesis of Findings:* This objective involves integrating the insights obtained from the array of studies, comparing and contrasting their outcomes to construct a comprehensive view. By doing so, we aim to provide a nuanced understanding of how NVMe SSDs impact computing performance and efficiency, highlighting both their strengths and limitations [6].
- *Identifying Trends and Gaps:* We intend to pinpoint current trends and emerging patterns in the utilization of NVMe SSD technology, as well as uncover any significant research gaps or unexplored avenues. This exploration will help in anticipating future trajectories and potential innovations in the field of NVMe SSDs.
- *Guidelines and Recommendations:* Based on our comprehensive analysis, the aim is to formulate practical guidelines and strategic recommendations. These will serve as valuable resources for practitioners, researchers, and decision-makers in effectively integrating NVMe SSDs into various computing environments, ensuring optimized performance and future-proofing technology investments.

BACKGROUND: AN OVERVIEW

- *Technology Overview:* Non-Volatile Memory Express (NVMe) is a storage interface developed specifically for Solid State Drives (SSDs) to fully exploit the high-speed capabilities of NAND flash memory. Unlike its predecessors (such as SATA and SAS interfaces, which were originally optimized for mechanical hard disk drives), NVMe is built for the parallelism inherent in flash-based storage, allowing it to handle high levels of data throughput and reduce latency. The primary advantages of NVMe technology over traditional storage solutions are:
- *Higher Bandwidth:* NVMe operates over the PCI Express (PCIe) bus, which allows for a much wider and faster data transfer lane compared to SATA interfaces. This results in significantly higher bandwidth, enabling faster read/write speeds [7].
- *Reduced Latency:* NVMe reduces command overhead by streamlining the command sets, which allows for a significant reduction in data access times. This is particularly beneficial for applications requiring rapid data retrieval and processing.
- *Increased IOPS (Input/Output Operations Per Second):* NVMe can handle a greater number of I/O operations simultaneously, which is crucial for multi-threaded and high-demand applications.
- *Efficiency and Scalability:* NVMe's design minimizes CPU overhead for I/O operations, making it more efficient. It is also highly scalable, catering to the evolving needs of modern data centers and computing environments.

METHODOLOGY

1. *Leveraging NVMe SSDs for Big Data Processing in Enterprise Data Centers:* This study explains the effectiveness of NVMe SSDs in big data processing within Apache Spark environments, comparing their performance against traditional Hard Disk Drives (HDDs) [4]. The central question is whether the high I/O capabilities and cost of NVMe SSDs justify their use in all large-scale workloads.

   The authors structured their research around a detailed experimental setup, utilizing Apache Spark on a virtual machine hypervisor with Yarn for resource management and End-to-End Dataflow in a Spark Environment with YARN and HDFS Integration (Figure 1). They conducted a series of tests comparing NVMe SSDs and HDDs across various applications, focusing on I/O throughput and shuffling impact [4].

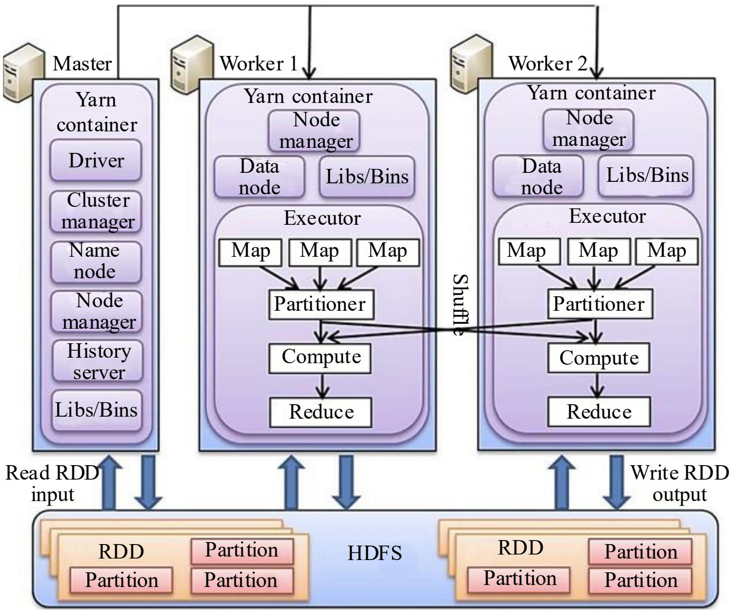


Figure 1. End-to-End dataflow in a spark environment with YARN and HDFS Integration [2].

Key Observations:

1. *Performance Variation Across Applications:* The performance enhancement with NVMe SSDs varied depending on the application. For example, K-means clustering showed a modest improvement (Figure 2), while database applications like TPC-H-Spark exhibited more significant benefits (Figure 3) [4].

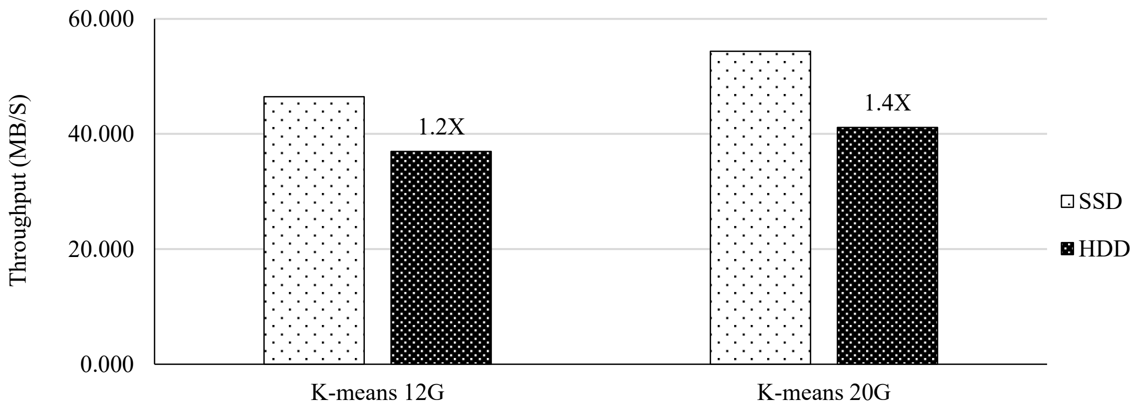
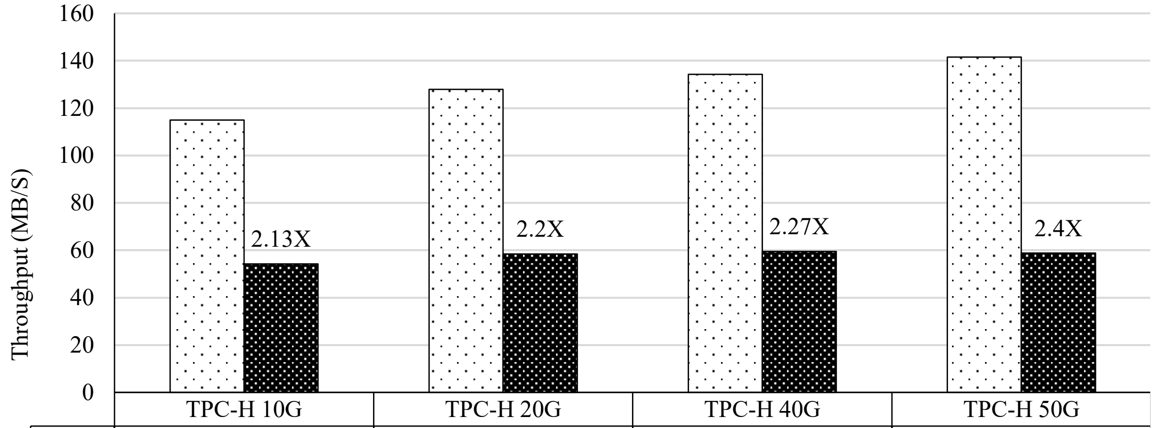


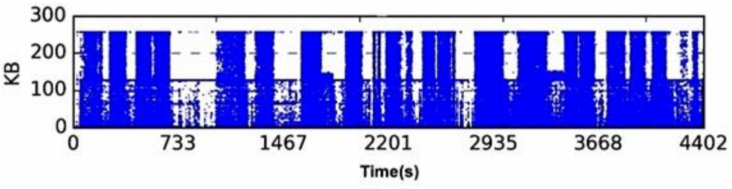
Figure 2. K-Means Clustering Performance with Variable Workloads on High-Speed and Traditional Storage.



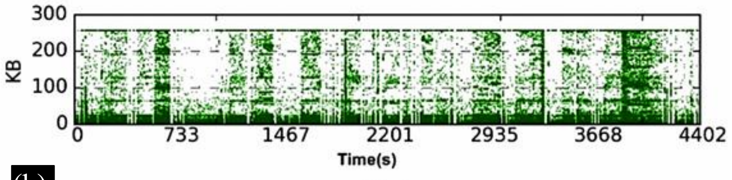
	TPC-H 10G	TPC-H 20G	TPC-H 40G	TPC-H 50G
☐ SSD	114.99	127.85	134.3	141.59
▩ HDD	54.2	58.4	59.5	58.805

Figure 3. TPC-H Spark Throughput Comparison Across Varying Workload Sizes on NVMe SSDs and HDDs.

2. *I/O Access Behavior:* The study also examined I/O characteristics for database applications [4], revealing different I/O size patterns between NVMe SSDs (Figure 4) and HDDs (Figure 5).

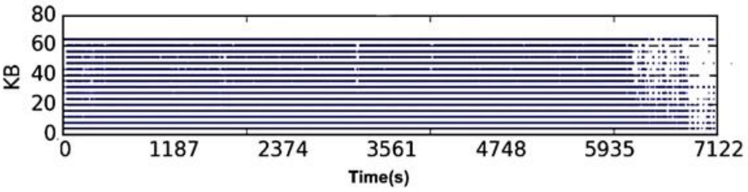


(a) Read

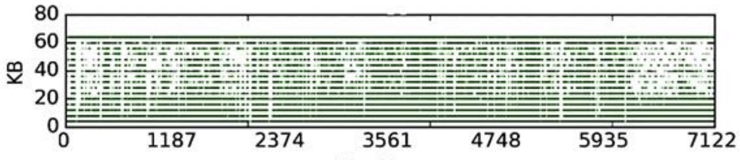


(b) Write

Figure 4. (a, b) Read and Write I/O Characteristics of TPC-H Spark with 50 GB workload on NVMe SSD.



(a) Read



(b) Write

Figure 5. (a, b) Read and Write I/O Characteristics of TPC-H Spark with 50 GB workload on HDD.

3. *Optimal Parallelism:* Another key aspect of their analysis was the impact of parallelism in Spark on SSD and HDD performance [4]. They identified an optimum level of parallelism for both storage types (Figure 6).

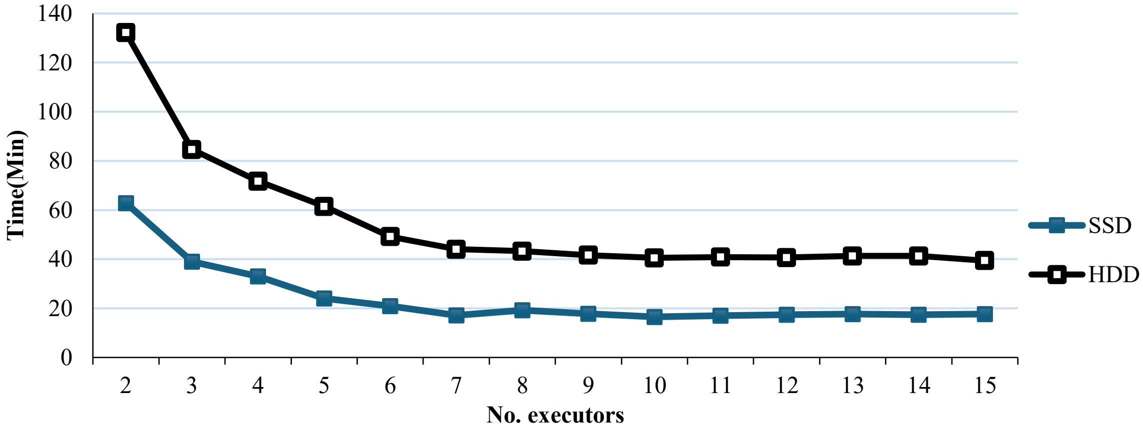


Figure 6. Optimum level of parallelism in SSD and HDD (1 core per VM).

The study concludes that NVMe SSDs do not universally enhance performance for all data processing applications. While database applications can significantly benefit from NVMe SSDs, general Spark applications like K-means may not fully utilize their high I/O throughput [4]. The decision to deploy NVMe SSDs should, therefore, be informed by the specific characteristics and requirements of the application.

This study suggests adjustments in the HDFS mapping scheme for better performance with SSDs and plans to explore big data processing in a containerized environment [8].

H-NVMe: Designing a Hybrid Storage Architecture Using NVMe for Cloud Platforms

In this study, the authors present a groundbreaking framework to enhance the performance of Non-Volatile Memory Express (NVMe) storage systems in cloud computing environments. This study is particularly relevant as cloud data centers increasingly adopt NVMe SSDs over traditional SATA and SAS SSDs due to their higher speed and efficiency [1]. However, the full potential of NVMe SSDs is often not realized in virtualized environments because of the existing VM-hypervisor architectures which introduce significant I/O bottlenecks (Figure 7).

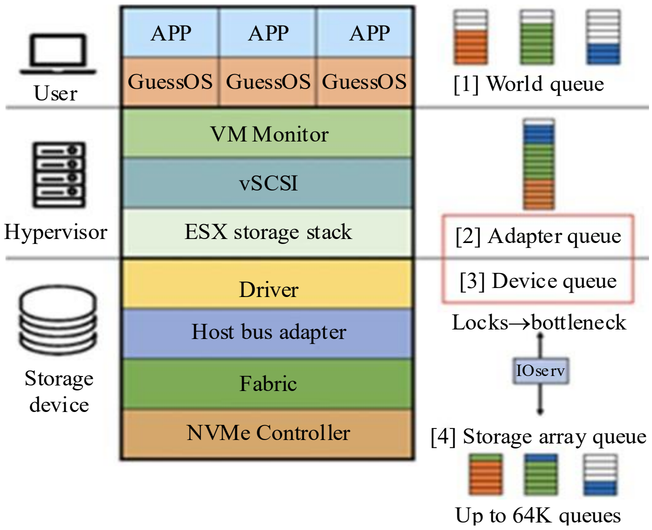


Figure 7. Architecture of current NVMe-based VM Hypervisor.

The proposed H-NVMe framework introduces two innovative operational modes: Direct Access Mode (DAM) and Parallel Queue Mode (PQM). PQM aims to increase parallelism and operational efficiency by integrating lightweight queues within the NVMe driver [1]. This approach enables lock-free operations, significantly improving I/O processing (Figure 8). DAM, in contrast, allows trusted applications direct access to NVMe SSDs, bypassing the hypervisor's I/O stack, thereby providing improved performance isolation and efficiency [9, 10].

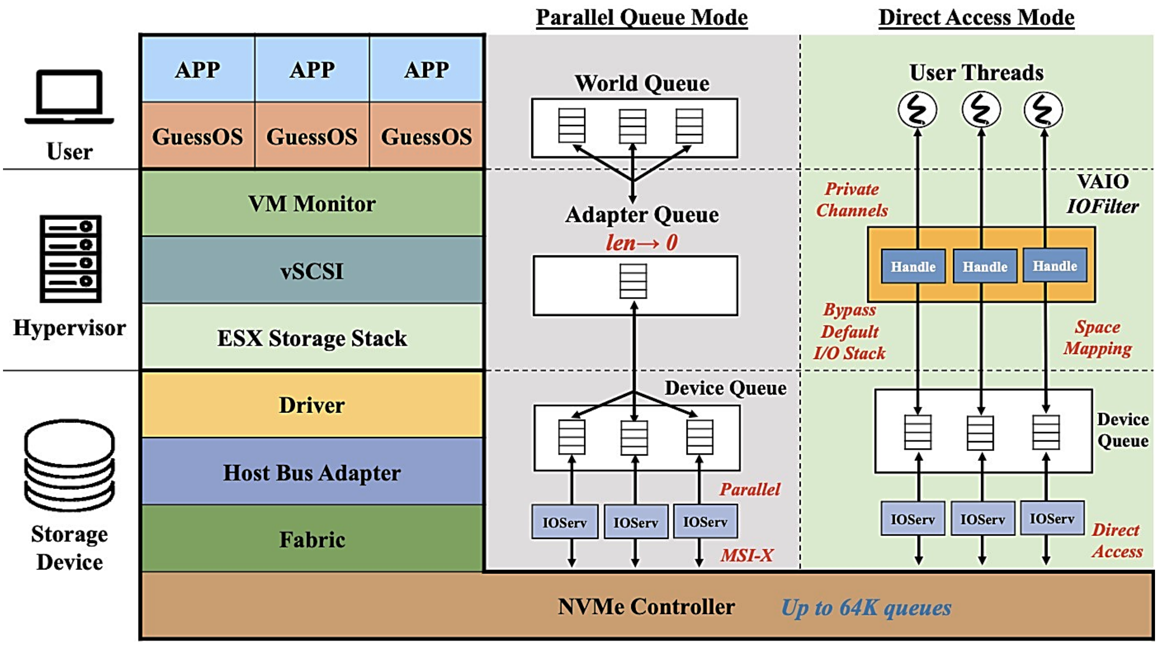
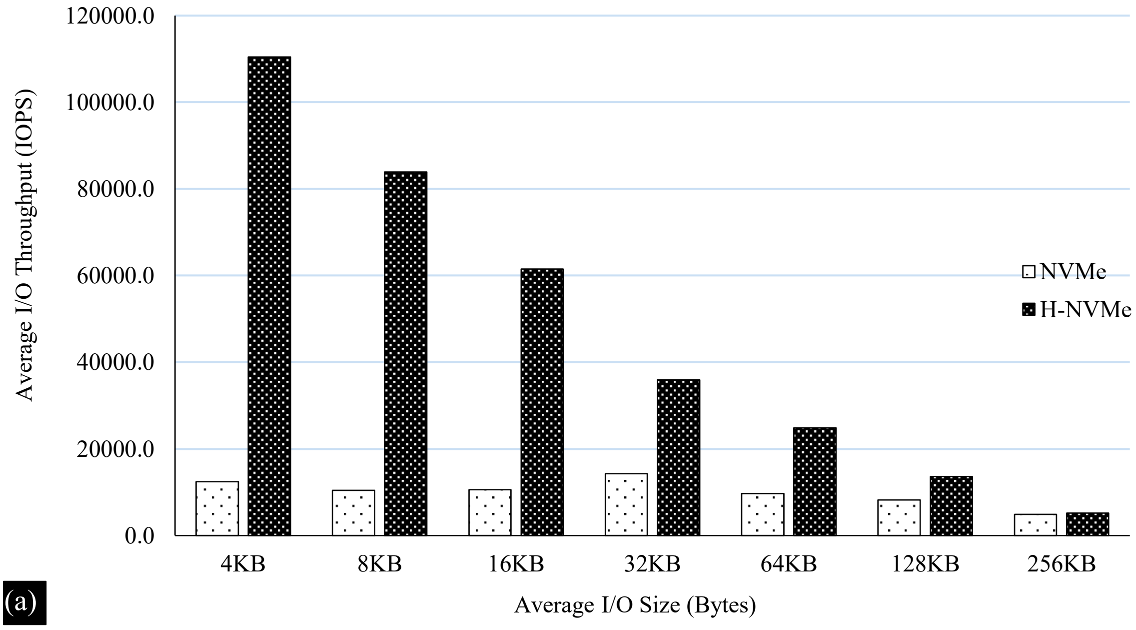
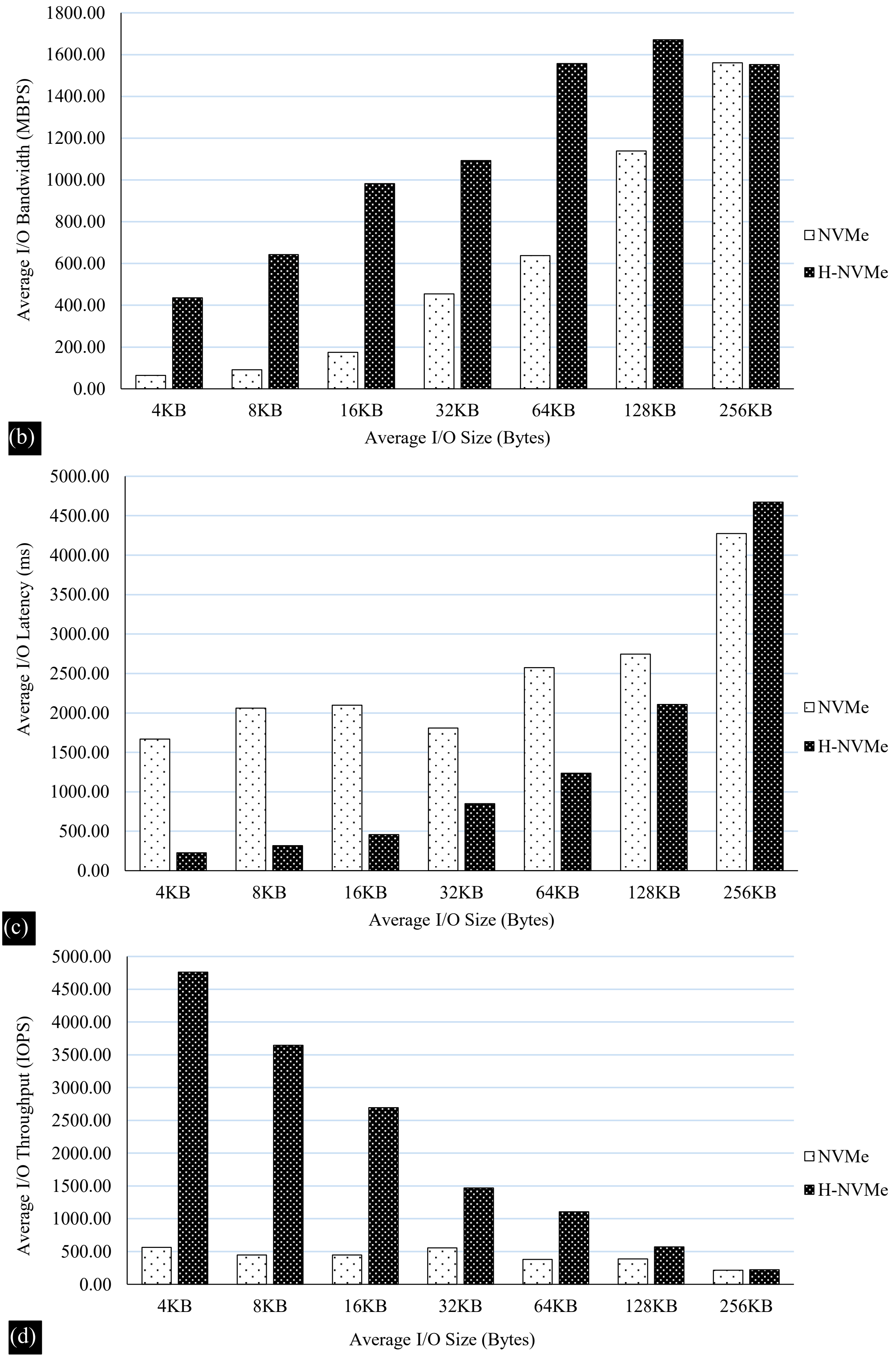


Figure 8. I/O paths of two proposed modes.

The study's experimental evaluation, conducted on VMware ESXi 6.0.0, demonstrates the substantial performance improvements offered by H-NVMe [1]. The results, as detailed in Figure 9, show that H-NVMe achieves higher throughput, better bandwidth, and lower latency across various workloads compared to the traditional NVMe solutions. These improvements are especially notable for workloads involving small block sizes, where H-NVMe's design maximizes efficiency [1].




Average I/O Bandwidth (MBPS)
0.00
200.00
400.00
600.00
800.00
1000.00
1200.00
1400.00
1600.00
1800.00
4KB
8KB
16KB
32KB
64KB
128KB
256KB
Average I/O Size (Bytes)
□NVMe
▩H-NVMe
(b)
Average I/O Latency (ms)
0.00
500.00
1000.00
1500.00
2000.00
2500.00
3000.00
3500.00
4000.00
4500.00
5000.00
4KB
8KB
16KB
32KB
64KB
128KB
256KB
Average I/O Size (Bytes)
□NVMe
▩H-NVMe
(c)
Average I/O Throughput (IOPS)
0.00
500.00
1000.00
1500.00
2000.00
2500.00
3000.00
3500.00
4000.00
4500.00
5000.00
4KB
8KB
16KB
32KB
64KB
128KB
256KB
Average I/O Size (Bytes)
□NVMe
▩H-NVMe
(d)

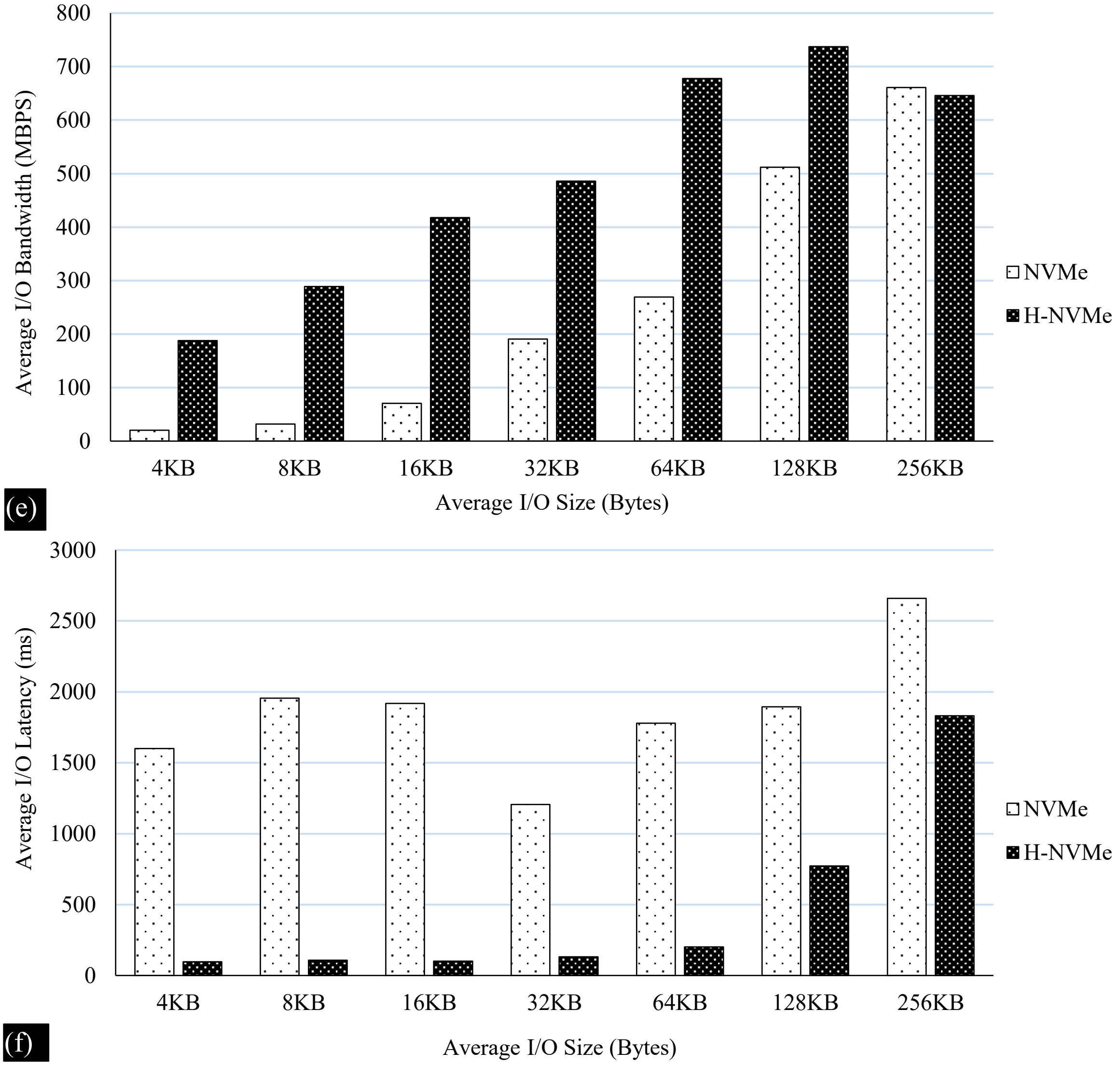
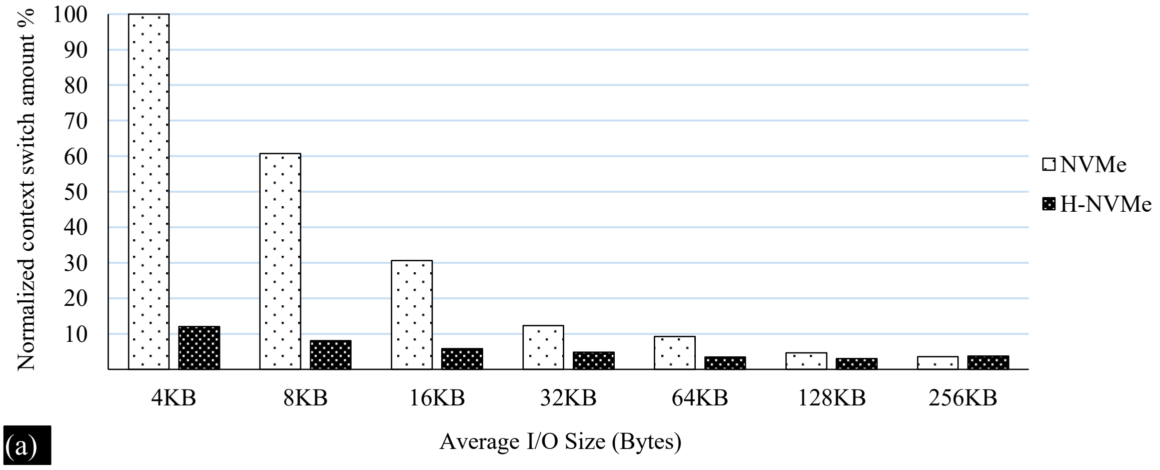


Figure 9. Comparative Analysis of NVMe and H-NVMe Using Seven Workloads: Focus on Throughput, Bandwidth, and Latency, (a) Overall read throughput of each VM, (b) Overall read bandwidth of each VM, (c) Normalized read latency of each VM, (d) Overall write throughput of each VM, (e) Overall write bandwidth of each VM, (f) Normalized write latency of each VM.



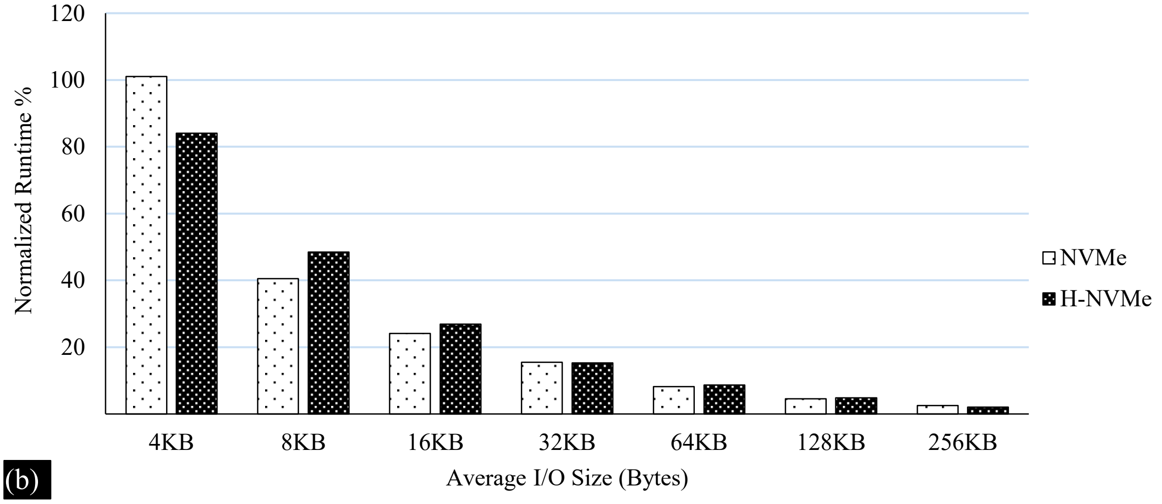


Figure 10. Evaluation of Execution Efficiency and Context Switching: Seven Workloads on NVMe and H-NVMe Platforms, (a) Normalized runtime of each VM, (b) Normalized context switch amount of each VM.

Additionally, Figure 10 highlights the reduced runtime and context switching required in the H-NVMe framework compared to the standard NVMe setup. This not only signifies efficiency in terms of processing speed but also points to better resource utilization within the cloud computing infrastructure [1].

Overall, this study is a significant contribution to the field of cloud computing and data storage [1]. By addressing the inefficiencies in current NVMe-based storage systems in VM-hypervisor environments, H-NVMe emerges as a robust solution that optimizes performance and resource utilization, a critical advancement for large-scale cloud computing data centers.

NVMe-CR: Scalable Runtime Support for Ephemeral Storage in Checkpoint/Restart via NVMe-over-Fabrics

In this study, authors address optimizing I/O performance for checkpoint/restart (C/R) operations in exascale computing environments [2].

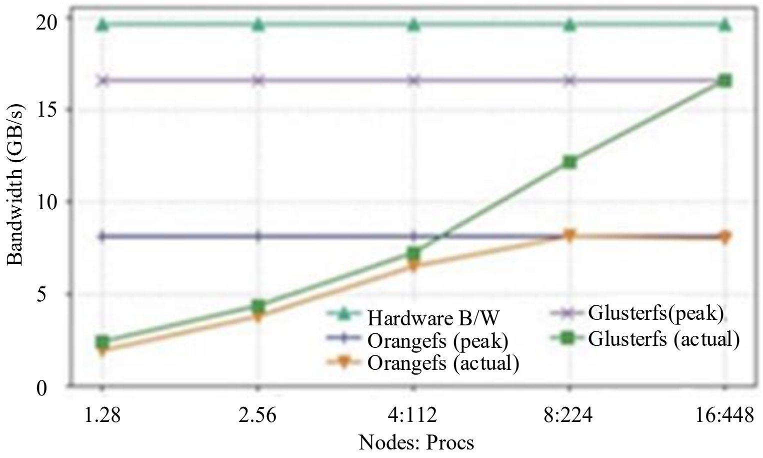


Figure 11. Evaluation of Checkpoint Bandwidth Under Weak Scaling Across Compute Nodes (Refer §IV-A). Notable Difference Between OrangeFS, GlusterFS, and Hardware I/O Limits.

The study identifies limitations in existing parallel filesystems, which struggle with the high concurrency and frequency of large-scale checkpointing, largely due to their reliance on POSIX filesystem semantics and substantial metadata overhead [2]. This issue is visually demonstrated in Figure 11, showing the gap between the checkpoint bandwidth of current filesystems like GlusterFS, and OrangeFS and hardware I/O bandwidth.

In response, the authors introduce NVMe-CR, an ephemeral storage runtime using NVMe-over-Fabrics (NVMf), optimized for C/R operations [2]. Its unique micro's abstraction simplifies the I/O path and minimizes metadata overhead, enabling direct user space access to storage devices. This innovative approach is depicted in Figure 12, highlighting the current approach of application-transparent remote storage access using NVMf [2].

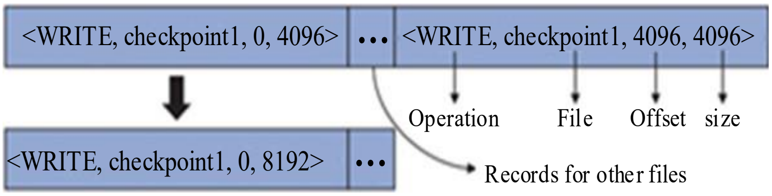


Figure 12. Traditional NVMe-oF-based remote storage access: Fully Kernel-space data plane and No application awareness [2].

The architecture of NVMe-CR's is detailed in Figure 13, which includes a control plane, data plane, and storage balancer, designed to reduce synchronization overhead and enhance scalability [2]. The efficiency of this architecture is further exemplified in Figure 12, which illustrates the log record coalescing implemented in NVMe-CR.

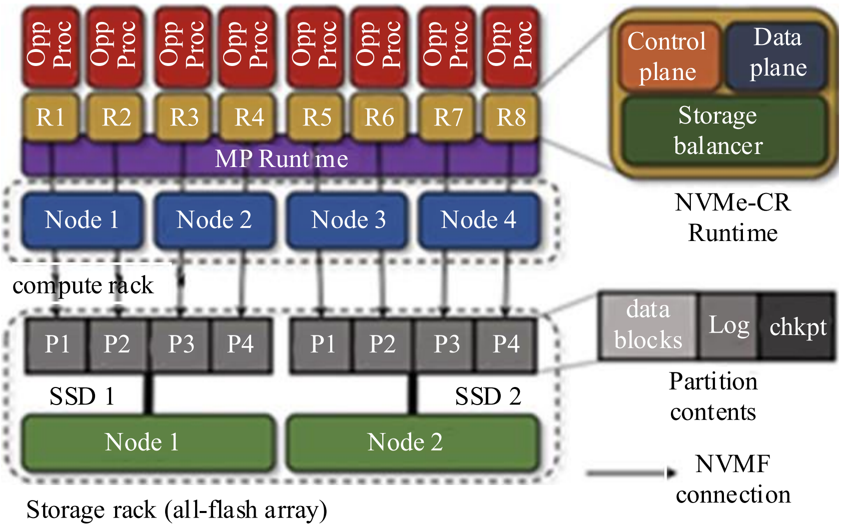


Figure 13. The NVMe-CR runtime architecture enables each instance to independently access its associated remote SSD via NVMe-over-Fabrics (NVMf).

The drilldown evaluation of NVMe-CR, using the ECP CoMD application, shows us its efficiency and scalability in check-pointing compared to existing storage systems. Figure 14 provide a view of NVMe-CR's performance, including checkpoint times for different huge block sizes and full subscription performance on local and remote SSDs. Furthermore, Table 1 compares the metadata overhead with CoMD across different storage systems, underlining NVMe-CR's efficiency [2].

Table 1. Metadata overhead with CoMD in MB.

OrangeFS*	GlusterFS*	NVMe-CR*
2686.25	3.5	445.25

**Per storage node # per runtime*

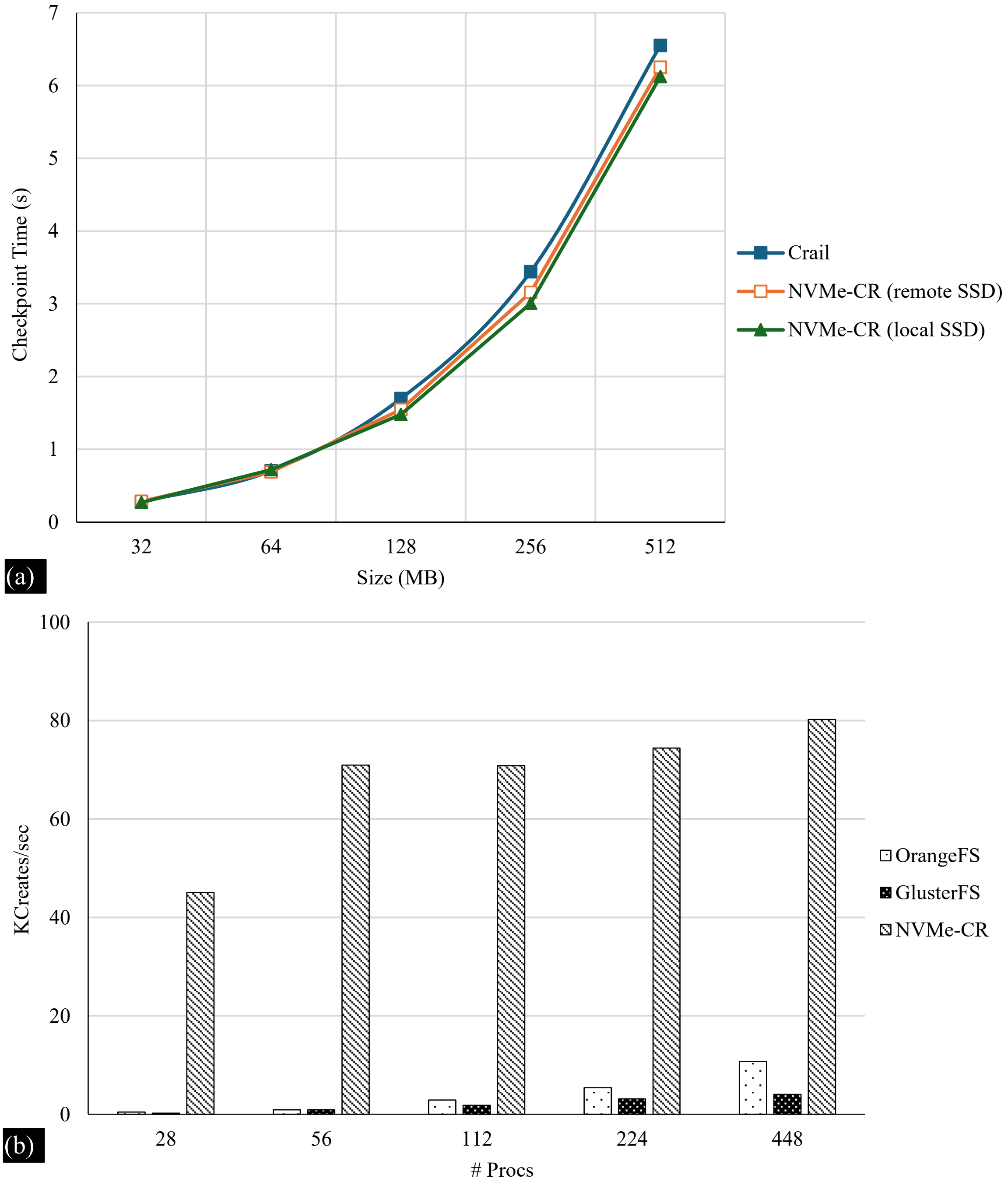


Figure 14. a) Evaluation of NVMe-CR's full subscription performance on local and remote SSDs, and b) File creation performance comparison between NVMe-CR, GlusterFS, and OrangeFS.

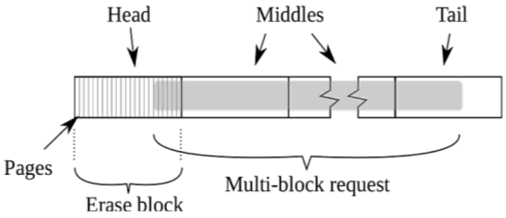
In conclusion, NVMe-CR is a breakthrough in addressing the I/O bottlenecks in exascale computing [2]. By leveraging modern NVMe storage technologies, it offers a scalable, efficient solution for C/R operations, as demonstrated by the comprehensive figures and tables in the study. This makes NVMe-CR a valuable tool for the future of HPC.

Modeling the Performance of Flash-Based SSDs Under Real-World Workloads

In this study, the authors address the challenge of performance modeling for NAND flash-based solid-state drives (SSDs) in real-world applications. They aim to bridge the gap between the theoretical capabilities of NAND flash SSDs and their actual performance under varying workloads [5]. The focus is on developing analytical models for two commonly used Flash Translation Layer (FTL) algorithms, along with a methodology for applying these models to real operational workloads [5].

The models are validated through simulations and further enhanced with measurement-based approximations for scenarios where detailed device parameters are unavailable. This validation against real SSD hardware ensures the practical applicability of the proposed models [5].

The study begins by highlighting the inherent complexity of modeling SSD performance, which stems from diverse workload characteristics and the behavior of different FTL algorithms. It introduces a workload analysis framework that categorizes request patterns into head, middle, and tail fragments, as illustrated in Figure 15 [5]. This breakdown allows for a clearer understanding of how various types of data requests impact SSD performance.



Device parameters:

V_P	number of physical blocks
V_L	number of logical blocks
f	free space ratio ($\frac{V_P-V_L}{V_L}$)
B	erase block size in pages

Workload parameters:

w	mean write length, in pages
T	total number of writes in pages

Other variables:

E	number of erasures
M_s, M_p, M_f	number of switch, partial, full merges
F_h, F_m, F_t	number of head, middle, tail fragments
H_l	mean head length in pages
L_f	total number of fragments per log (FAST)
L_v	number of valid fragments per log (FAST)

Figure 15. Division of a request into head, middle, and tail fragments.

In order to demonstrate the model's accuracy in predicting the performance of SSDs, the authors validate their analytic model against the simulation results for a single write (Figure 16). In addition, they extend their methods for predicting performance of arbitrary workload and show that the model can be used in real-world scenarios [5].

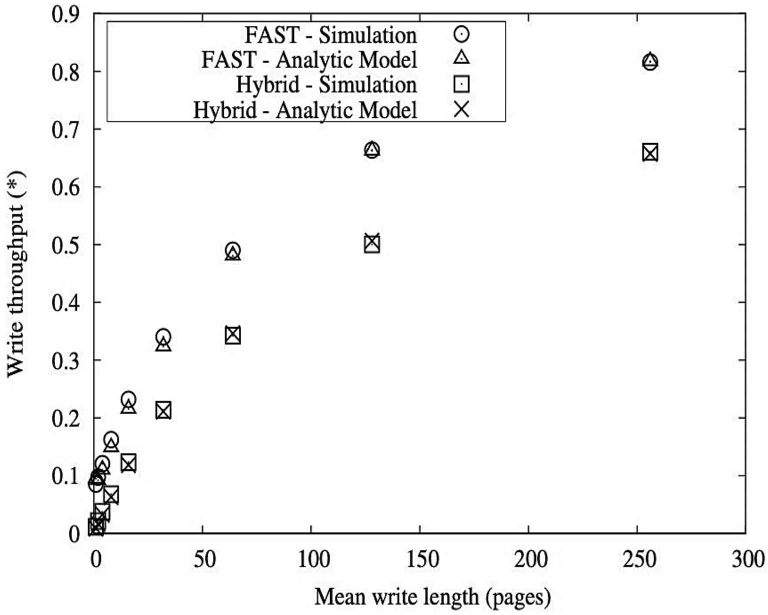


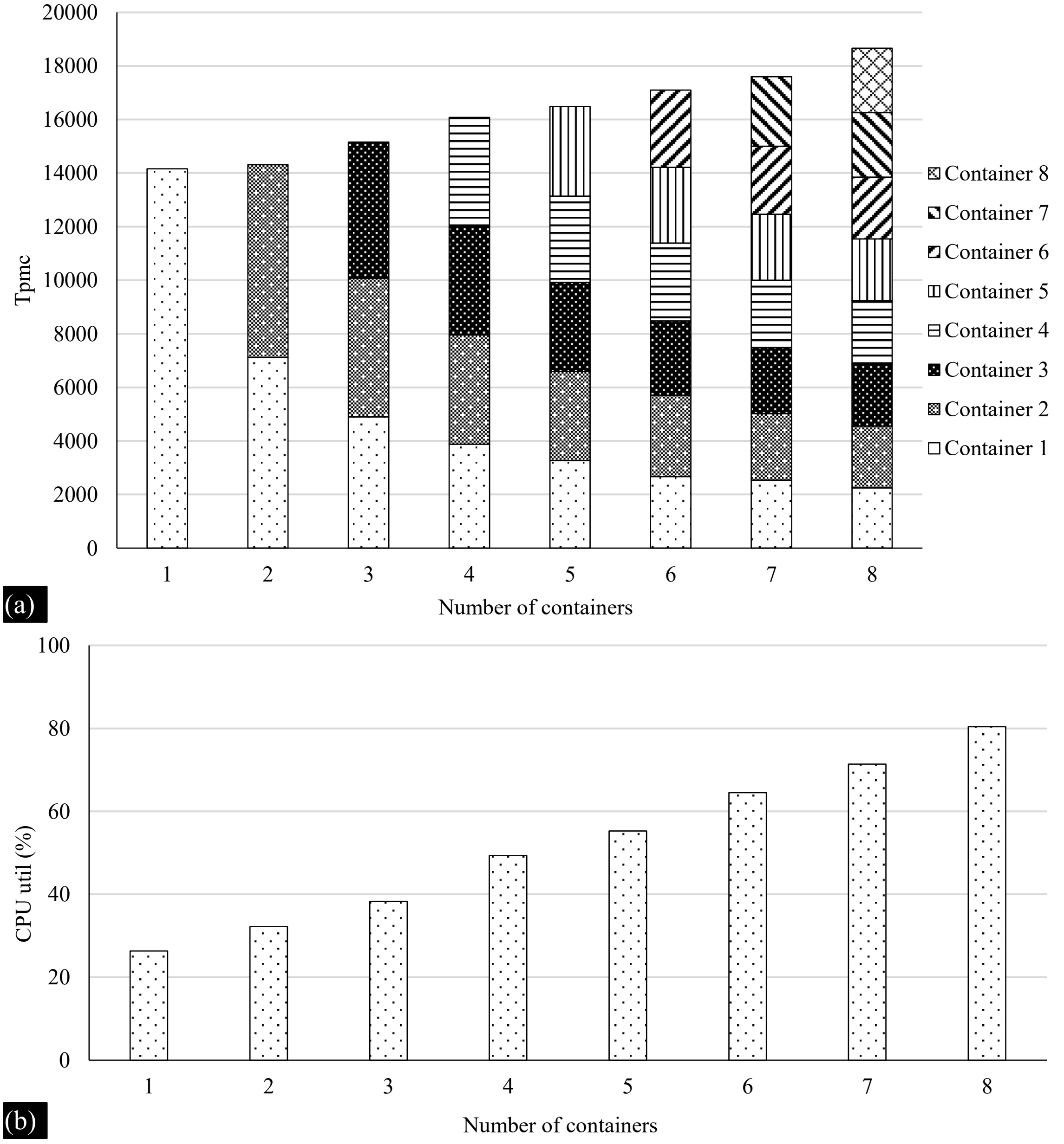
Figure 16. Write throughput analysis using analytical model validated against simulation results: performance tied to erasure efficiency.

In addition, the study had shown us a good correlation between the predicted throughput and the measured throughput in most cases, showing a comparison of analytical based predictions of real workload [5]. It is crucial that, in practical and real world situations, this aspect of the study supports the effectiveness of the model.

Finally, this study has made important contributions to understanding and predicting the performance of flash Based SSDs in realistic applications [5]. It offers valuable insights into the behavior of SSDs under various workloads and presents a massive methodology for performance prediction, making it a valuable resource for researchers and practitioners in the field of data storage and memory technology.

Performance Analysis of I/O-Intensive Containerized Workloads on NVMe SSDs

In this study, the authors suggest the increasing adoption of Docker due to its scalability and efficiency in data centers [3]. It then addresses the performance challenges in multi-tenant Docker environments, stating the importance of understanding and managing performance bottlenecks for application fairness and resource efficiency.



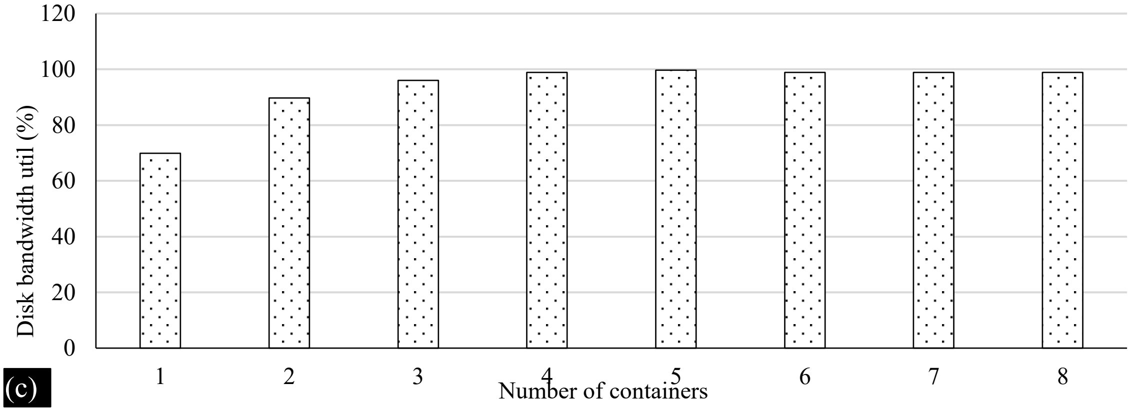
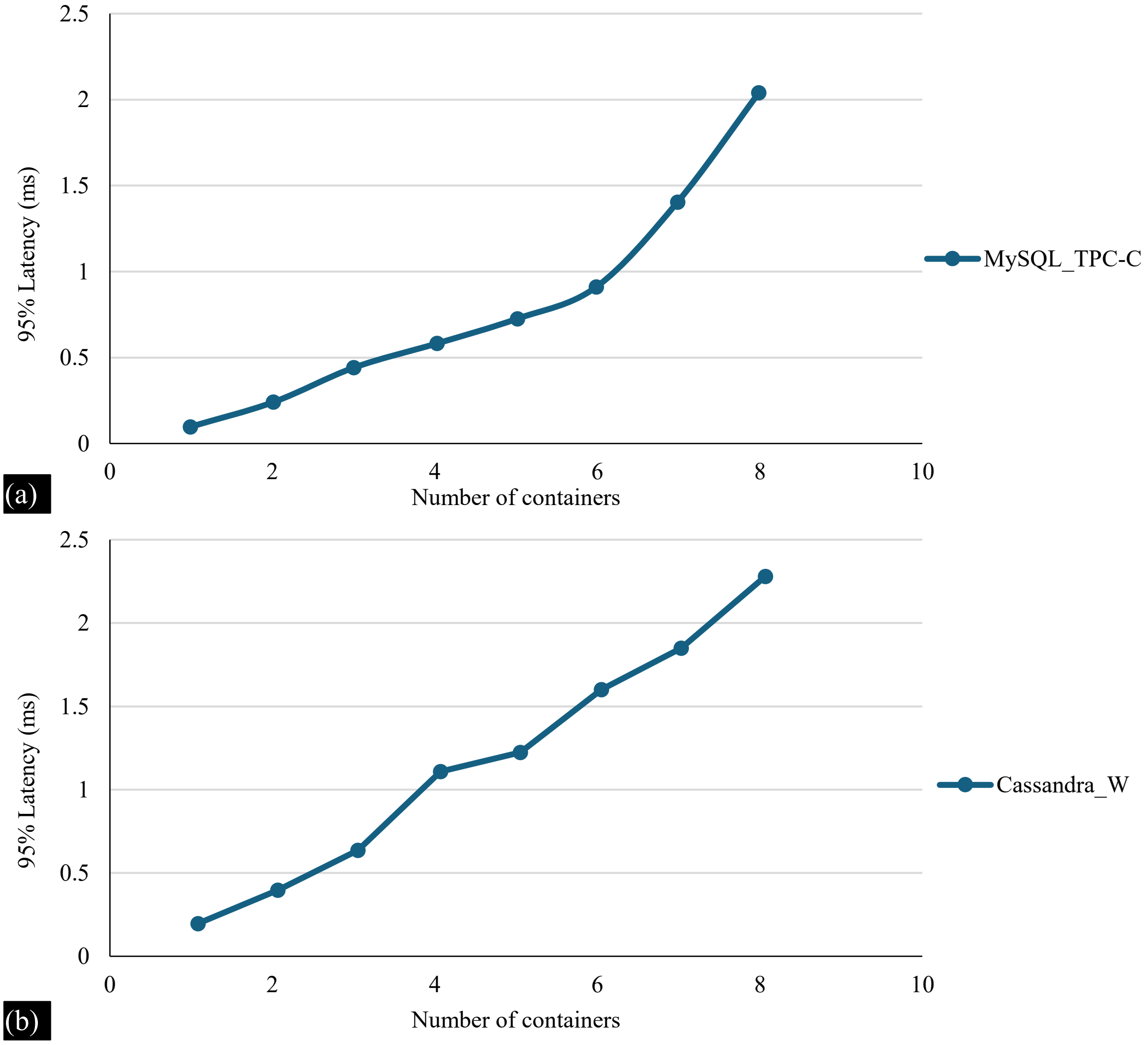


Figure 17. MySQL Performance Under TPC-C in a Homogeneous (a) Transaction Rate (TpmC), (b) CPU Utilization, and (c) Disk Usage.

A central aspect of the study is the performance characterization of Dockerized applications on persistent storage, specifically NVMe SSDs. The study presents an in-depth analysis of the impact of different numbers of Docker containers on various workloads [3]. This includes an exploration of the performance dynamics under both homogeneous and heterogeneous environments, as depicted in Figures 17–19.



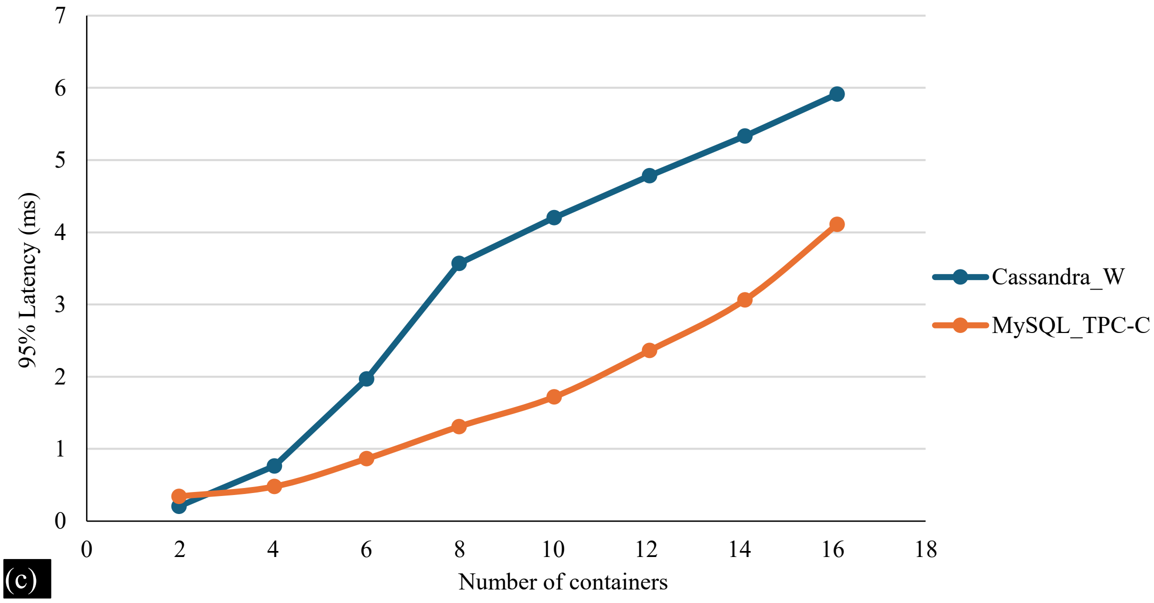
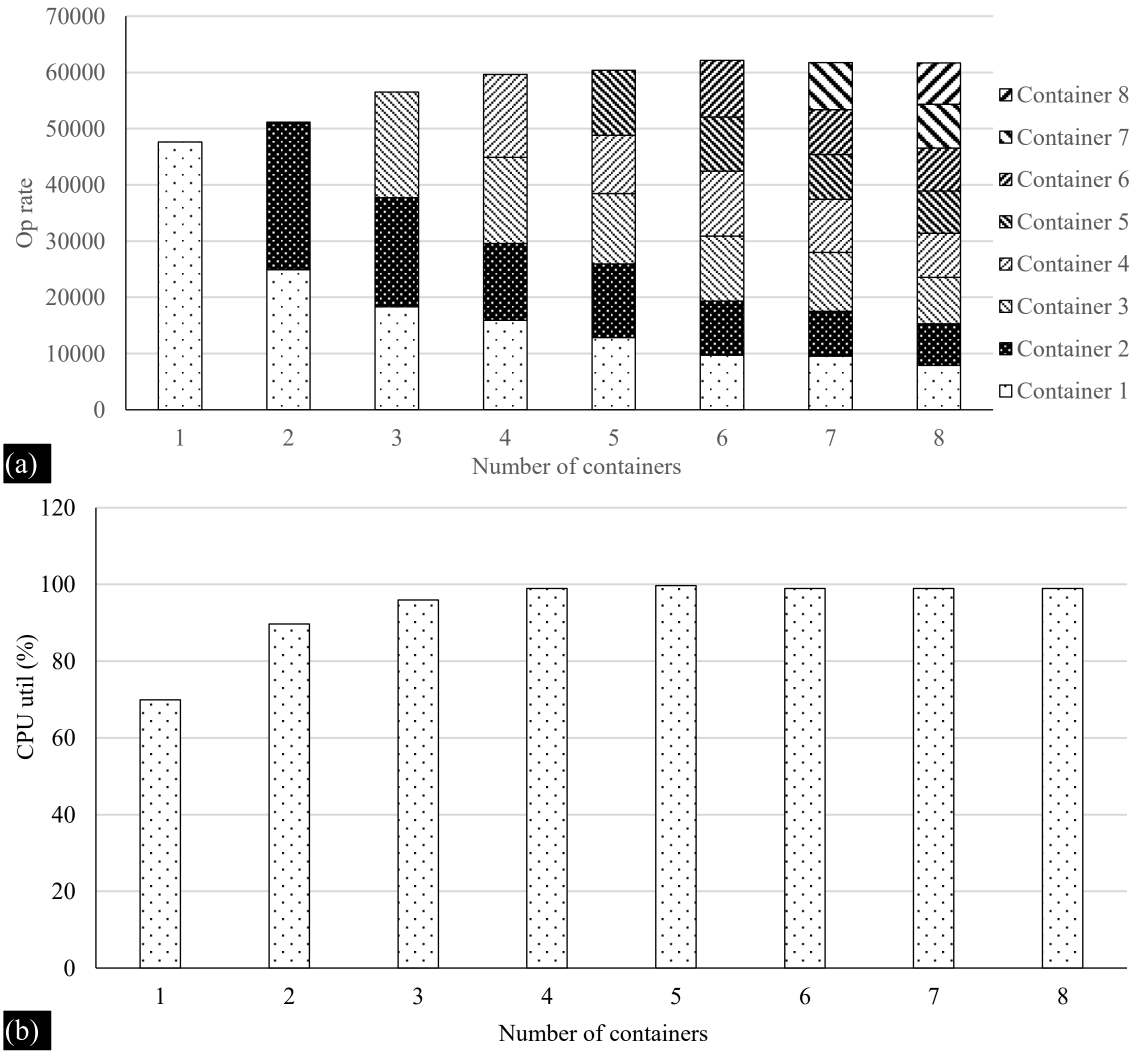


Figure 18. Measured average latency for all concurrently running containers, (a) homogeneous MySQL, (b) homogeneous Cassandra MySQL+ Cassandra W (c) heterogeneous.



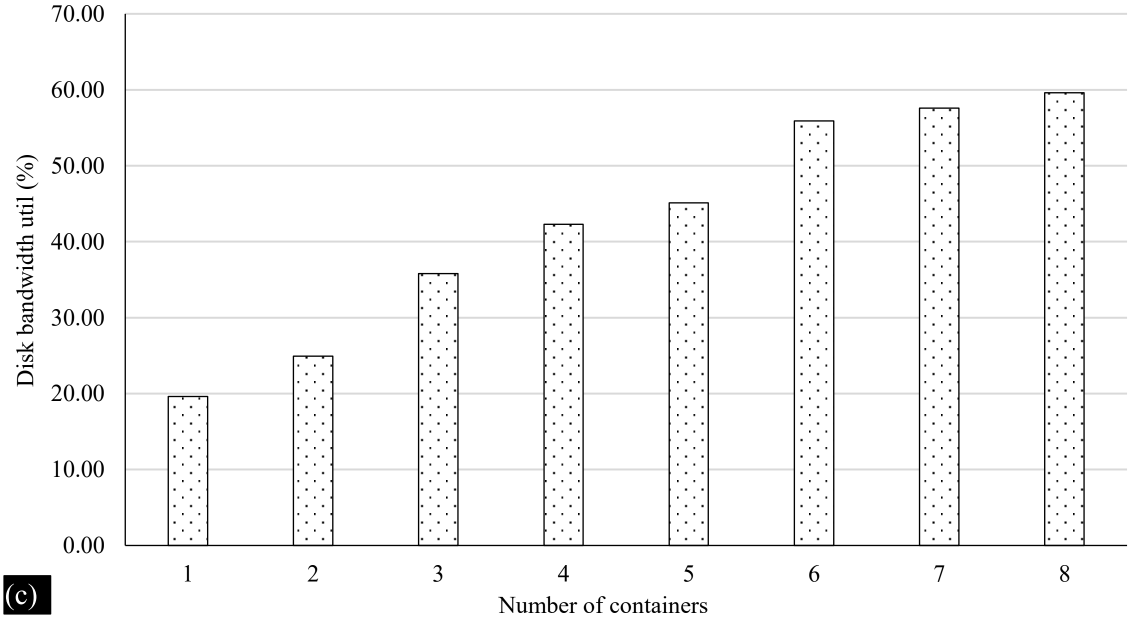


Figure 19. Homogeneous with Cassandra (Cassandra utilization W workload) (a) Cassandra throughput, (b) CPU utilization, (c) Disk band width.

Figure 20 illustrates the containerized system architecture used in the study. It features an array of three enterprise-grade NVMe drives configured to deliver higher disk bandwidth for Dockerized database applications [3].

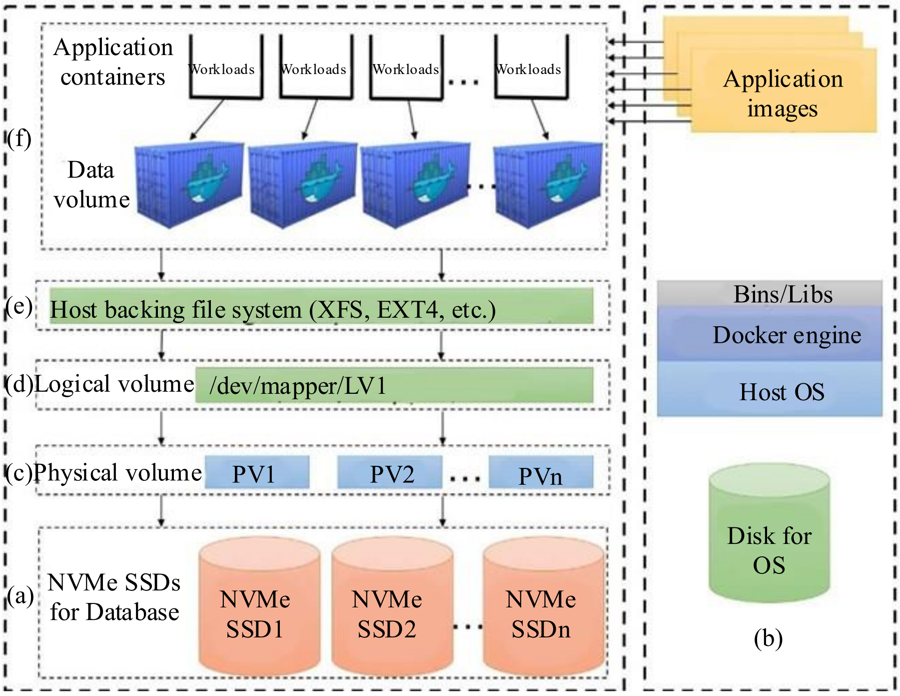


Figure 20. Containerized system on flash volume of SSDs.

The experimental analysis extends to heterogeneous setups, such as simultaneous operations of Cassandra W with FIO random writes, as shown in Figure 21. This figure demonstrates an unfair throughput throttle where FIO random write containers sacrifice more throughput compared to Cassandra in such mixed workload environments [3].

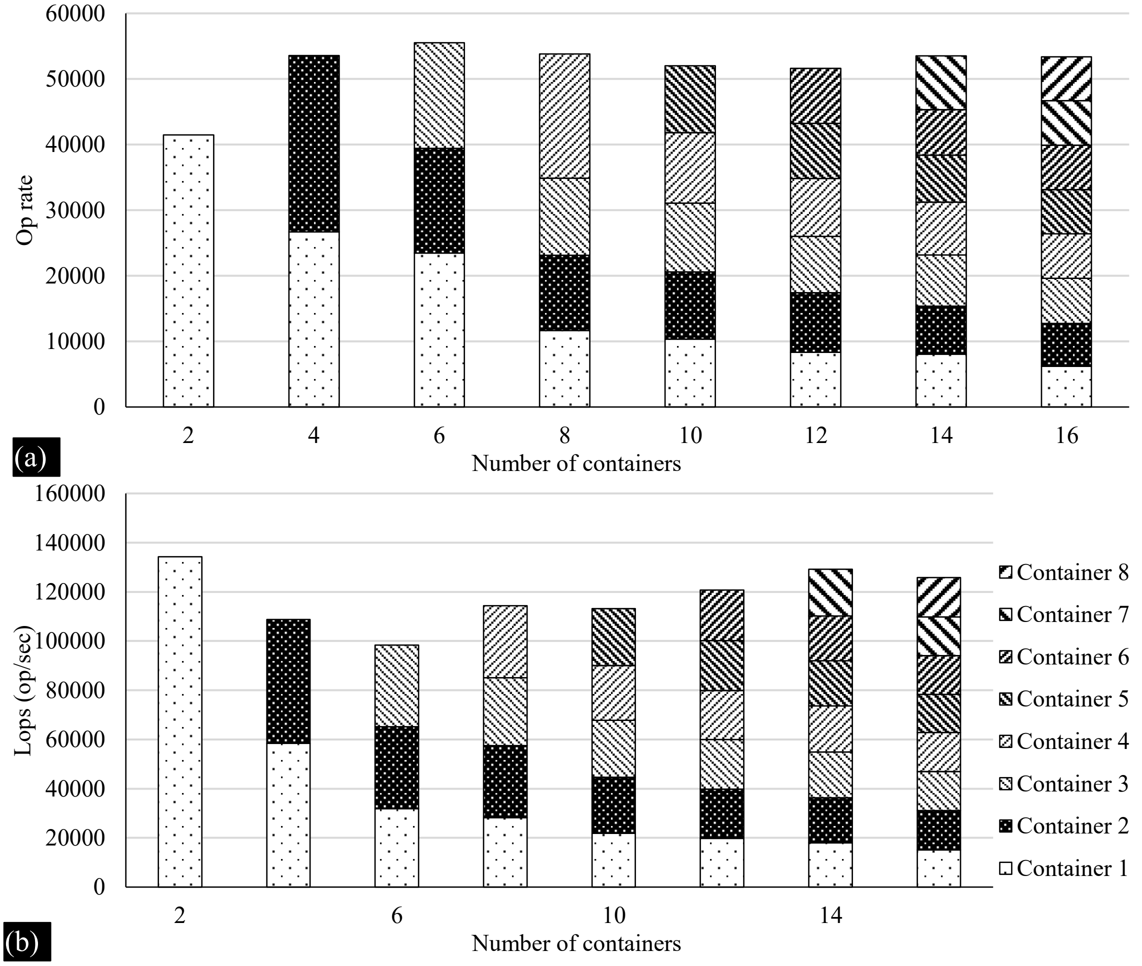


Figure 21. Mixed Workload Scenario: (a) Cassandra Throughput Under Native Load, (b) FIO Random Write Throughput (4 kb).

In conclusion, this study is an important step in enhancing our understanding of I/O-intensive containerized applications, particularly in NVMe SSD contexts. It offers critical insights and practical guidelines for optimizing these systems, proving invaluable for system architects and engineers in cloud computing and data center domains [3]. The inclusion of detailed figures and tables from the study adds depth to the understanding of the complex interactions between containerized applications and their underlying storage systems.

RESULTS

The five papers collectively contribute to a deeper understanding and enhancement of data processing and storage in various computing environments, with a specific emphasis on the use of NVMe SSDs. At their core, each study explores different facets of how NVMe technology can revolutionize data handling, offering insights into performance improvements, scalability, and optimization strategies across diverse scenarios.

Paper by Yang *et al*. investigates their use in HPC environments [1]. The research introduces a novel storage run-time optimized for checkpoint/restart operations, using NVMe-over-Fabrics. This approach not only amplifies the performance capabilities of NVMe SSDs in handling IO-intensive tasks but also illustrates the technology's scalability and efficiency in demanding HPC scenarios.

Building on this understanding, the study by Gugnani *et al*. intrudes into the cloud computing realm [2]. It proposes an innovative framework designed to optimize NVMe SSDs' performance in virtualized cloud environments, addressing some of the unique challenges presented by these settings. This study shows the adaptability of NVMe technology in different environments, showing its potential to enhance cloud storage solutions.

The paper "Understanding Performance of I/O Intensive Containerized Applications for NVMe SSDs," by Bhimani *et al.* addresses the increasingly popular domain of containerized applications [3]. It explores how NVMe SSDs impact the performance of I/O-intensive applications in containerized setups like Docker. This research is pivotal in understanding the role of NVMe technology in modern, flexible computing environments, highlighting how these drives can be optimized for containerized applications.

Paper by Bayati *et al.* lays the groundwork by examining how NVMe SSDs can significantly boost big data processing capabilities in enterprise data centers, especially when compared to traditional HDDs [4]. This study highlights the nuanced relationship between application characteristics and the performance benefits offered by NVMe SSDs, offering a foundational perspective on how these drives can be effectively utilized in large-scale data processing environments.

Finally, the paper by Boboila and Desnoyers takes a more analytical approach [5]. It develops performance models for flash-based SSDs, focusing on real-world workload conditions. This study bridges the gap between theoretical performance and practical application, providing valuable insights for optimizing systems and applications for better SSD performance.

Collectively, these papers present a comprehensive overview of the critical role NVMe SSDs play in advancing data processing and storage across diverse computing environments. From big data analytics in enterprise data centers to cloud computing, high-performance computing, and modern containerized applications, NVMe SSDs emerge as a transformative technology. They deliver significant gains in performance, scalability, and efficiency, key attributes essential to meeting the demands of today’s rapidly evolving computing landscape.

DISCUSSION

Technology Limitations of NVMe SSDs

- Big Data Processing within Enterprise-Scale Data Centers
  - *I/O Bottlenecks:* Even though NVMe SSDs offer higher throughput, they can still encounter I/O bottlenecks, especially in applications that do not fully utilize their I/O capabilities.
  - *Application-Specific Dependencies:* The performance improvement varies greatly based on the application’s nature and design, meaning that NVMe SSDs are not universally beneficial for all types of applications.
- Integration of NVMe-Based Storage in Cloud Computing Environments
  - *Virtualization Overhead: In virtualized environments, NVMe SSD performance can be affected by challenges such as queue synchronization locks and interference between virtual machines (VMs), leading to potential performance degradation.*
  - *Resource Sharing Issues:* The efficient sharing and allocation of NVMe resources in a cloud environment are complex and can lead to performance inconsistencies.
- NVMe-CR in HPC Environments
  - *Hardware Resource Limitations:* There is a limit to the scalability offered by NVMe SSDs, particularly when dealing with extremely high volumes of data common in HPC environments.
  - *Dependency on Fabric Technologies:* NVMe-over-Fabrics is essential for NVMe-CR, and any limitations in fabric technology could impact overall system performance.
- Performance Models for Real Workloads
  - *Write Amplification:* This issue can significantly affect the lifespan and performance efficiency of NVMe SSDs, especially under certain workloads.

  - *Complexity in Performance Prediction:* Accurately predicting NVMe SSD performance under varying real-world conditions remains complex and sometimes imprecise.
- Containerized Applications for NVMe SSDs
  - *Resource Saturation:* As the number of containers increases, the hardware resources can become saturated, leading to performance degradation.
  - *Optimization Challenges:* Tuning and optimizing containerized environments for NVMe SSDs can be challenging due to the dynamic nature of container workloads.

CONCLUSION

This review has systematically examined the evolving landscape of NVMe SSD technology across various computing environments, highlighting its transformative impact on data processing and storage efficiency NVMe SSDs have emerged as a cornerstone technology, driving advancements across a range of computing domains, from enhancing big data processing in enterprise data centers and optimizing storage systems in cloud environments, to improving I/O-intensive application performance in containerized setups and supporting high-performance computing workloads.

The papers reviewed underscore the significant performance gains NVMe SSDs offer over traditional storage solutions, particularly in terms of throughput, latency, and I/O efficiency. However, the studies also reveal certain limitations and challenges, such as I/O bottlenecks, application-specific dependencies, and complexities in virtualized and containerized environments. These challenges necessitate ongoing research and development to fully harness the potential of NVMe technology.

Looking ahead, the continuous evolution of NVMe SSDs promises to further reshape the computing landscape. As we address the identified research gaps, such as the need for broader application scope studies, long-term durability and reliability analysis, and more comprehensive cost-benefit evaluations, the full potential of NVMe SSDs will become increasingly evident. Moreover, advancements in related technologies, such as fabric networks in HPC environments and software optimizations in cloud and containerized infrastructures, will complement the strengths of NVMe SSDs.

In conclusion, NVMe SSD technology stands at the forefront of a storage revolution, offering unprecedented opportunities for enhancing data processing and storage capabilities. Its ongoing evolution will undoubtedly play a pivotal role in meeting the burgeoning demands of modern computing infrastructures, driving efficiencies, and enabling more sophisticated and data-intensive applications.

REFERENCES

1. Yang Z, Hoseinzadeh M, Wong P, Artoux J, Mayers C, Evans DT, Bolt RT, Bhimani J, Mi N, Swanson S. H-NVMe: A hybrid framework of NVMe-based storage system in cloud computing environment. In 2017 IEEE 36th International Performance Computing and Communications Conference (IPCCC). 2017 Dec 10; 1–8.
2. Gugnani S, Li T, Lu X. Nvme-cr: A scalable ephemeral storage runtime for checkpoint/restart with nvme-over-fabrics. In 2021 IEEE International Parallel and Distributed Processing Symposium (IPDPS). 2021 May 17; 172–181.
3. Bhimani J, Yang J, Yang Z, Mi N, Xu Q, Awasthi M, Pandurangan R, Balakrishnan V. Understanding performance of I/O intensive containerized applications for NVMe SSDs. In 2016 IEEE 35th International Performance Computing and Communications Conference (IPCCC). 2016 Dec 9; 1–8.
4. Bayati M, Bhimani J, Lee R, Mi N. Exploring benefits of nvme ssds for bigdata processing in enterprise data centers. In 2019 IEEE 5th International Conference on Big Data Computing and Communications (BIGCOM). 2019 Aug 9; 98–106.
5. Boboila S, Desnoyers P. Performance models of flash-based solid-state drives for real workloads. In 2011 IEEE 27th Symposium on Mass Storage Systems and Technologies (MSST). 2011 May 23; 1–6.

6. Kim B, Kim J, Noh SH. Managing array of {SSDs} when the storage device is no longer the performance bottleneck. In 9th USENIX Workshop on Hot Topics in Storage and File Systems (HotStorage 17). 2017.
7. Guz Z, Li H, Shayesteh A, Balakrishnan V. NVMe-over-fabrics performance characterization and the path to low-overhead flash disaggregation. In Proceedings of the 10th ACM International Systems and Storage Conference. 2017 May 22; 1–9.
8. Kim HJ, Lee YS, Kim JS. {NVMeDirect}: A User-space {I/O} Framework for Application-specific Optimization on {NVMe}{SSDs}. In 8th USENIX Workshop on Hot Topics in Storage and File Systems (HotStorage 16). 2016.
9. Jung M, Choi W, Shalf J, Kandemir MT. Triple-A: A non-SSD based autonomic all-flash array for high performance storage systems. ACM SIGARCH Computer Architecture News. 2014 Feb 24; 42(1): 441–54.
10. Mao B, Wu S, Duan L. Improving the SSD performance by exploiting request characteristics and internal parallelism. IEEE Trans Comput Aided Des Integr Circuits Syst. 2017;37(2):472-84.