International Journal of Computer Trends and Technology
ISSN: 2231–2803 / https://doi.org/10.14445/22312803/IJCTT-V71I11P110

Volume 71 Issue 11, 70-73, November 2023
© 2023 Seventh Sense Research Group®

Original Article

The Hybrid World of HPC and Cloud

Divit Gupta[1], Naresh Kumar Miryala[2]

[1]NACI, Oracle America, TX, USA.
[2]Meta Platforms Inc. CA, USA.

[2]Corresponding Author : nareshm121@gmail.com

Received: 22 September 2023 Revised: 26 October 2023 Accepted: 13 November 2023 Published: 30 November 2023

Abstract *- Oracle Cloud Infrastructure (OCI) stands out as a premier cloud solution, offering the latest technology and diverse hardware options tailored for every workload. Renowned for enterprise-grade high-performance computing (HPC) capabilities and cost-effectiveness, OCI ensures the best price-performance ratio in the public cloud. With a commitment to absolute price-performance excellence, OCI delivers the advantages of on-premises deployments coupled with the flexibility of the public cloud.*

Keywords *- C.2.4.c Distributed databases < C.2.4 Distributed Systems < C.2 Communication/Networking and Information Technology < C Computer, C.1.3.f Heterogeneous (hybrid) systems < C.1.3 Other Architecture Styles < C.1 Processor Architectures < C Computer Systems Orga, C.1.m.b Hybrid systems < C.1.m Miscellaneous < C.1 Processor Architectures < C Computer Systems Organization.*

1. Introduction

Accessing the most advanced technology alongside a diverse array of hardware tailored to specific workloads forms the core proposition of Oracle Cloud Infrastructure (OCI)[1]. Notably, OCI offers enterprise-grade, high performance computing (HPC)[2] capabilities for production, coupled with the most economical pricing within the public cloud domain. Our fundamental principle has consistently been to deliver unparalleled price performance, affording customers the advantages of on premises deployments combined with the flexibility inherent in public cloud solutions.

OCI's performance excellence stems from Oracle's meticulous approach to HPC, delivering a genuine bare a metal experience complemented by the most performant and lowest-latency HPC network among major cloud providers[3]. This unique combination enables seamless scaling of tightly coupled HPC applications, ensuring optimal communication between nodes with minimal delays, thereby facilitating increased core counts without compromising efficiency[4]. This, in turn, translates to significant time and cost savings.

As the premier cloud solution for all HPC requirements, OCI distinguishes itself by offering more speciality hardware locations than any other provider, allowing users to execute workloads at any location and time[5]. With robust disaster protection measures and compliance across all regions, OCI boasts dual cloud regions in six countries and the EU, each comprising up to four availability domains and three fault domains per availability domain.

Oracle Cloud Infrastructure (OCI) introduces high performance computing (HPC) that delivers potent and cost-efficient computing capabilities, addressing intricate mathematical and scientific challenges across diverse industries[6]. OCI's bare metal servers, in conjunction with Oracle's cluster networking grants users the advantage of ultra-low latency RDMA(less than 2 µs latency across clusters of tens of thousands of cores) over converged Ethernet (RoCE) v2[7].

HPC on OCI stands as a formidable counterpart to on premises solutions, combining the elasticity of cloud computing with consumption-based costs. This translates to the ability to scale tens of thousands of cores on demand. HPC on OCI provides access to high-frequency processors, rapid and compact local storage**,** high-throughput, ultra-low latency RDMA cluster networking, and a suite of tools for the seamless automation and execution of jobs[7].

2. Need for HPC

Continued advancements in computer hardware, Increased data accessibility and the development of sophisticated simulation software has empowered scientists and engineers to formulate intricate scenarios to conceptualize potential outcomes. To expedite the discovery process, researchers, scientific communities, and businesses have been dedicating resources to establishing in-house high performance computing (HPC) capabilities[9], entailing the management of data centers.

Undoubtedly, these investments have yielded significant returns. According to Hyperion Research, an investment in HPC translates to an average revenue of $463 for each dollar invested and an average profit of $44 per dollar dedicated to HPC initiatives.

The majority of high-performance computing (HPC) workloads are executed on-site to attain superior performance, low latency, and heightened throughput. Organizations aim to maximize their return on investment, making the optimization of HPC workloads crucial due to the substantial cost associated with these powerful machines[10][19]. Nevertheless, running HPC workloads on-site presents several challenges[6], including:

2.1. Capacity Planning

Among the primary hurdles faced by HPC customers is the challenge of effective capacity planning. Organizations often find themselves in a dilemma of either overprovisioning or under-provisioning capacity, resulting in prolonged queuing times and a subsequent impact on productivity.

2.2. Outdated Hardware

Infrastructure updates generally occur every 3–5 years, Necessitating system users to patiently await access to the latest HPC infrastructure. This delay acts as a hindrance to both innovation and productivity[11].

2.3. Data Center Management Costs

This encompasses expenses related to power, cooling, networking, storage, hardware, and software. For the majority of organizations running HPC workloads, the management of facilities and hardware is not their core competency[12].

2.4. Security and Compliance Requirements

Legacy systems necessitate regular maintenance to ensure ongoing adherence to security and compliance standards[6].

3. Cloud-based High-Performance Computing

Integrating high-performance computing (HPC) workloads into the cloud is becoming more prevalent[13]. As per a 2020 Intersect360 Research report, the cloud sector has experienced consistent double-digit growth over the past 5–6 years, boasting a robust 22.0% compound annual growth rate (CAGR). Projections indicate that the cloud segment is expected to surpass $3.8 billion by 2024."

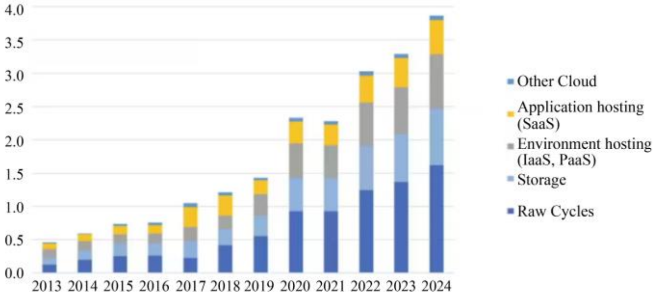

Fig. 1 Consumption of HPC in public cloud ($B), Intersect360 Report.

Transitioning your high-performance computing (HPC) workloads to the cloud provide a range of advantages:

3.1. Enhanced Performance

Access the latest generation of infrastructure and on-demand capacity to meet your business requirements. HPC customers can deploy bare metal virtual machines (VMs), GPUs, and Compute shapes as necessary[20].

3.2. Cost Optimization

The cloud eliminates concerns about upfront capital costs. Utilize the pay-as-you-use pricing model and run without interruptions[12].

3.3. Scalability Flexibility

The cloud grants organizations the flexibility to scale their IT requirements up or down to align with the business's dynamic needs[14].

3.4. Core Business Focus

With HPC in the cloud, the burden of building and managing data centers is lifted, enabling you to concentrate on your strategic business priorities.

4. Future of HPC

The forefront of high-performance computing (HPC) has arrived with the introduction of high-performance computing on Oracle Cloud Infrastructure (OCI)[15][20]. HPC solutions play a pivotal role in tackling some of the world's most challenging problems, such as powering molecular modeling for disease combat and simulating car crashes to enhance vehicle safety and reduce fatalities. Leveraging the cloud for intricate simulations and mathematical model analyses make high-performance computing more accessible to a broader audience of scientists, engineers, and analysts, given the significantly reduced costs compared to on-premises server setups[16].

5. The Toyota Story

Toyota migrates high-performance workloads to the Oracle Cloud.

The largest automotive manufacturer globally transfers. High-performance computing workloads to Oracle Cloud. Infrastructure to enhance efficiency in car design and development.

5.1. Toyota's Business Challenges

In pursuit of enhancing its car quality, Toyota is executing the "Toyota New Global Architecture," a structural innovation spanning its global car manufacturing operations. This initiative aims to bring significant enhancements to the fundamental performance of Toyota vehicles. The pivotal strategy involves elevating the the efficiency of automotive design and development through computational tests and simulations, persistently striving for advancements that result in cars exhibiting outstanding driving performance and environmental sustainability.

Historically, Toyota has managed high-performance computing workloads within an on-premises environment. However, while maintaining the utilization of existing on-premises resources, the company embarked on exploring cloud services. This strategic shift enables Toyota to address user requests dynamically, such as rapid resource scaling and exploring emerging technologies.

Why Toyota Motor chose Oracle, The company benchmarked a number of cloud service providers as part of its HPC multicloud strategy, looking into performance, cost, computational accuracy, flexibility, stability, and other requirements. After considering these factors, Toyota decided to move the foundation of its computational simulations to Oracle Cloud Infrastructure (OCI) while also using existing on-premises systems. Oracle Cloud Infrastructure offers the industry’s first and only public cloud with bare metal HPC computing. It has a low latency of less than 2 microseconds and 100 Gbps bandwidth, achieved with a Remote Direct Memory Access network. This protocol transfers data from the memory of a local computer to the memory of a separate, remote computer. OCI’s unique HPC solution allows Toyota to run large and complex computational simulations, which require a massive amount of computing power, all in the cloud and without any performance compromises.

5.2. Why Toyota Motor Opted for Oracle?

In pursuit of its HPC multicloud strategy, the company conducted a thorough benchmarking of several cloud service providers, evaluating factors such as performance, cost, computational accuracy, flexibility, stability, and more[21]. After carefully considering these elements, Toyota strategically decided to transition the core of its computational simulations to Oracle Cloud Infrastructure (OCI) while concurrently leveraging its existing on-premises systems.

Oracle Cloud Infrastructure stands out as the industry's initial and sole public cloud, featuring bare metal HPC computing. Boasting a latency of less than 2 microseconds and a bandwidth of 100 Gbps, achieved through a Remote Direct Memory Access network, OCI offers a distinctive HPC solution. This protocol facilitates the seamless transfer of data from the memory of a local computer to that of a distinct, remote computer. OCI's unparalleled HPC capabilities empower Toyota to execute extensive and intricate computational simulations, demanding substantial computing power, all within the cloud, without compromising performance.

6. Results

The execution of high-performance workloads for computational simulations on Oracle Cloud Infrastructure has enabled Toyota to accelerate the pace and efficiency of car design and development, all while optimizing costs[23].

Furthermore, Toyota has significantly reduced the lead time in procuring computing resources, a process that previously took over six months in the on-premises environment. Now, with OCI, this procurement is accomplished in just a few days, provided there is sufficient physical space available.

OCI grants Toyota the flexibility to conduct tests on new technologies seamlessly, a task that posed challenges in the on-premises setup. The adoption of OCI has resulted in a commendable level of cost performance for Toyota, streamlining the time required for computations through enhancements in computational capability.

7. Conclusion

The article discusses the utilization of Oracle Cloud Infrastructure (OCI) for high-performance computing (HPC) workloads, emphasizing the advantages of accessing advanced technology and tailored hardware for specific tasks[21]. OCI offers enterprise-grade HPC capabilities with cost-effective pricing, delivering performance excellence through bare metal servers and a low-latency HPC network[3][13]. The article highlights Toyota's migration of HPC workloads to OCI, driven by the need for efficient car design and development. The decision to choose OCI is attributed to its unique bare metal HPC computing, low latency, and high bandwidth. The transition has enabled Toyota to enhance efficiency, reduce lead times, and achieve cost savings. The broader context discusses the growing trend of integrating HPC workloads into the cloud, emphasizing the benefits of enhanced performance, cost optimization, scalability flexibility, and the ability to focus on core business priorities[20][21][23]. The article concludes with Toyota's successful outcomes, emphasizing OCI's role in accelerating computational processes and improving overall cost performance.

References

[1] Charles Bell, “Oracle Cloud Infrastructure,” *MySQL Database Service Revealed*, pp. 17-75, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[2] Maicon Ança dos Santos, and Gerson Geraldo H. Cavalheiro, “Cloud Infrastructure for HPC Investment Analysis,” *Revista de Informática Teórica e Aplicada*, vol. 27, no. 4, pp. 45-62, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[3] Vojtech Uhlir, Ondrej Tomanek, and Lukas Kencl, “Latency-Based Benchmarking of Cloud Service Providers,” *Proceedings of the 9th International Conference on Utility and Cloud Computing*, pp. 263-268, 2016. [CrossRef] [Google Scholar] [Publisher Link]

[4] Hang Cao et al., “An Efficient Cloud-Based Elastic RDMA Protocol for HPC Applications,” *CCF Transactions on High Performance Computing*, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[5] Abel Souza et al., “Hybrid Resource Management for HPC and Data Intensive Workloads,” *2019 19th IEEE/ACM International Symposium on Cluster, Cloud and Grid Computing (CCGRID)*, pp. 399-409, 2019. [CrossRef] [Google Scholar] [Publisher Link]

[6] Sara Koleini, “HPC System Security: Threats and Vulnerabilities, Challenges and Solutions,” 2023.

[7] Dingyu Yan et al., “A Survey of RoCEv2 Congestion Control,” *The 7th International Conference on Information Science, Communication and Computing*, pp. 42-56, vol. 350, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[8] Amanda Bienz et al., *High Performance Computing. ISC High Performance 2022 International Workshops*, 1st ed., Lecture Notes in Computer Science, vol. 13387, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[9] Rodrigo Da Rosa Righi et al., “Designing Cloud-Friendly HPC Applications,” *High Performance Computing in Clouds*, pp. 99-126, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[10] Patrik Omland et al., “HPC Hardware Design Reliability Benchmarking with HDFIT,” *IEEE Transactions on Parallel and Distributed Systems*, vol. 34, no. 3, pp. 995-1006, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[11] Jorge Lozoya Arandia et al., “Green Energy HPC Data Centers to Improve Processing Cost Efficiency,” *Latin American High Performance Computing Conference*, vol. 1540, pp. 91-105, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[12] Rodrigo Da Rosa Righi et al., “Towards Cloud-Based Asynchronous Elasticity for Iterative HPC Applications,” *Journal of Physics: Conference Series*, vol. 649, pp. 1-20, 2015. [CrossRef] [Google Scholar] [Publisher Link]

[13] Justin Shi et al., “Program Scalability Analysis for HPC Cloud: Applying Amdahl’s Law to NAS Benchmarks,” *2012 SC Companion: High Performance Computing, Networking Storage and Analysis*, pp. 1215-1225, 2012. [CrossRef] [Google Scholar] [Publisher Link]

[14] Nicolas Dube et al., “Future of HPC: The Internet of Workflows,” *IEEE Internet Computing*, vol. 25, no. 5, pp. 26-34, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[15] Herbert Cornelius, *The Future of High-Performance Computing (HPC)*, 4th ed., Encyclopedia of Information Science and Technology, pp. 1-14, 2018. [CrossRef] [Google Scholar] [Publisher Link]

[16] Vanderlei Munhoz, and Márcio Castro, “HPC@Cloud: A Provider-Agnostic Software Framework for Enabling HPC in Public Cloud Platforms,” *Anais do XXIII Simpósio em Sistemas Computacionais de Alto Desempenho*, pp. 157-168, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[17] Iacopo Colonnelli et al., “Federated Learning Meets HPC and Cloud,” *Machine Learning for Astrophysics*, vol. 60, pp. 193-199, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[18] Aman Verma, “Cloud Platform Optimization for HPC,” *Asian Conference on Supercomputing Frontiers*, pp. 55-64, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[19] Vitaly Antonenko et al., “On HPC and Cloud Environments Integration,” *Performance Evaluation Models for Distributed Service Networks*, pp. 159-185, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[20] Chaoqun Sha et al., “Facilitating HPC Operation and Administration Via Cloud,” *Supercomputing Frontiers and Innovations*, vol. 6, no. 1, pp. 23-35, 2019. [CrossRef] [Google Scholar] [Publisher Link]

[21] Abhishek Gupta, and Dejan Milojicic, “Evaluation of HPC Applications on Cloud,” *2011 Sixth Open Cirrus Summit*, pp. 22-26, 2011. [CrossRef] [Google Scholar] [Publisher Link]

[22] Dhabaleswar K. Panda, and Xiaoyi Lu, “HPC Meets Cloud: Building Efficient Clouds for HPC, Big Data and Deep Learning Middleware and Applications,” *Proceedings of the 10th International Conference on Utility and Cloud Computing*, pp. 189-190, 2017. [CrossRef] [Google Scholar] [Publisher Link]

[23] Feng Li et al., “ElasticBroker: Combining HPC with Cloud to Provide Realtime Insights into Simulations,” *ArXiv*, pp. 1-15, 2020. [CrossRef] [Google Scholar] [Publisher Link]