

# HPC on Oracle Cloud Infrastructure

---

Experience on premise-level  
performance and control with  
Oracle Cloud Infrastructure.





Organizations continue to move business applications and databases to the cloud to reduce the cost of purchasing, updating, and maintaining on-premise hardware and software. However, most high-performance computing (HPC) workloads remain on premise—mainly because these jobs have specialized needs that traditional cloud offerings can't handle. HPC applications consistently consume 100 percent of available compute power to achieve a specific outcome or result. They require dedicated network performance, fast storage, high compute capabilities, and significant amounts of memory—resources that are in short supply in the virtualized infrastructure that constitutes today's commodity clouds.

And yet, running HPC jobs in the cloud makes a lot of sense—assuming the cloud service has been architected in an optimum way for high-performance business or technical workloads. These jobs require immense amounts of resources during peak periods—such as when running a simulation, creating a model, or rendering an animated movie—and require little or no resources during the idle time between jobs.

Oracle Cloud Infrastructure offers exceptional performance, security, and control for today's most demanding HPC workloads. From bare metal offerings with the newest and fastest CPUs and GPUs to extreme low-latency flash storage, and lightning-fast RDMA networks, Oracle's cloud-based infrastructure enables customers to solve complex problems fast.

Watch the video on YouTube, [Oracle Unlocks HPC](#).

## Challenges Facing HPC Customers

- Overspending or under utilization: It's difficult to correctly size infrastructure for HPC workloads
- Massive upfront capital cost outlays for on-premise infrastructure: HPC deployments are larger than typical enterprise environments, and may present substantial risk
- Slower time to results in the cloud: Most commodity cloud offerings only use virtual machines that reduce performance
- Meeting the needs of the community: Does your HPC cluster have a long waiting list?

## A Better Solution for HPC Workloads

Oracle's infrastructure as a service (IaaS) brings you consistently leading-edge compute, storage, networking, and software technologies—all at a fraction of the cost of building high-end infrastructure on premise.

From cancer research to drug testing, autonomous driving to reservoir modeling, this infrastructure meets the exacting needs of data scientists, engineers, designers, and researchers.

## Advantages of Oracle Cloud Infrastructure

- Drive faster time to value. Building your on-premise HPC cluster isn't just expensive; it takes a long time
- Put control directly in the hands of the teams that consume HPC infrastructure
- Only pay for the resources that you consume, as you consume them
- Gain the ability to increase capacity to meet spikes in demand
- Focus on your strategic priorities and HPC outcomes, not on managing infrastructure or constantly fighting for capital budget for new hardware



## Compute Scalability and Performance

Oracle Cloud Infrastructure provides extremely fast cluster networking, high-powered bare metal CPU instances, and ultralow-latency NVMe flash storage to help you build powerful HPC clusters. Oracle Cloud Infrastructure now combines its proven HPC instance with a low-latency network that can span up to 20,736 cores. The core HPC building block (*shape* in Oracle Cloud parlance) has 36 cores from two 3.7 GHz Intel® Xeon® Gold 6154 processors, 384 GB RAM, and 6.4 TB NVME local storage.

“By partnering with Intel to deploy Intel Xeon Gold 6154 processors, OCI has been able to leverage the large performance gains from this generation of Intel Xeon Scalable processors and to accelerate the performance of HPC workloads like scientific simulation with the new AVX-512 instructions,” said Jeff Wittich, senior director of cloud service providers at Intel.

Compared to virtual machines (VMs), bare metal compute instances yield more-consistent results that are comparable to an on-premise compute infrastructure. With Oracle Cloud Infrastructure, you don’t need to settle for a virtualized HPC network.

### Scale

When virtualized in your data center, CPU-intensive tasks that require little system interaction, normally experience very little impact or CPU overhead. However, virtualized environments in the cloud include monitoring and management layers, which can add significant overhead.

Because compute overhead caused by virtualization is not synchronized across an entire cluster, it creates problems for MPI jobs which have to wait for the data from slowest node before advancing to the next compute iteration. So if even only one node is being slowed down by virtualization overhead -- or a noisy neighbor, or an oversubscribed network segment -- the entire job suffers.

With Oracle Cloud Infrastructure’s bare metal environment, no hypervisor or monitoring software runs on your compute instance. With limited overhead, HPC applications scale across multiple nodes just as well as they would in your data center.

### What Is HPC?

HPC (high-performance computing) is a way to aggregate multiple computers and storage devices into a cohesive fabric, architected to solve large problems in science, engineering, and business. In an HPC environment, individual nodes work together to solve a problem larger than any one computer can easily solve. The popularity of this type of infrastructure is growing fast: The global HPC market is expected to increase from US\$32 billion in 2017 to US\$45 billion by 2022. While HPC is often thought of as academic pursuit, many industries use HPC technology in their day-to-day business.

**Genomics** workloads include DNA sequencing and protein analysis, popular for ancestry studies, health testing, and the analysis of drug interactions to reduce time to market for new pharmaceutical products.

**Retailers** run AI models to analyze customer data and offer targeted purchase recommendations. These workloads benefit from GPU chips because they are tightly coupled with accelerated hardware, yielding up to 15× performance improvement.

**Media and entertainment** companies depend on HPC for animation, special-effects rendering, and media transcoding. These jobs are characterized by *bursty* workloads requiring hundreds or even thousands of nodes running in parallel.

**Financial technology (fintech)** companies run HPC jobs for risk analysis, high-frequency trading, and financial modeling. This industry has occasional demands for lots of compute and storage resources, such as when running quarterly reports.

**Automotive** companies run complex simulations throughout the design, manufacture, and testing of new vehicles. HPC workloads support computer-aided engineering (CAE) models for crash testing, simulations, and various types of analyses.

**Oil and gas** companies depend on HPC for geological modeling to predict where to find oil and natural gas resources. These jobs require spatial analysis, seismic analysis, and very large data sets.

**Aerospace** requires HPC infrastructure for computational fluid dynamics, such as simulating airflow over airplane wings. These simulations require lots of nodes, each with lots of CPUs and memory.



## HPC Networking—RDMA Clustering for Latency and Throughput

The Oracle Cloud Infrastructure flat network model is particularly important for HPC applications because with no virtualization or management traffic running through the application network, latency between servers is minimized. All bare metal servers come with at least two 25 Gb/sec Network Interface Cards (NICs) and the HPC shapes offer 125 Gb/sec (one 25 Gb/sec and one 100 Gb/sec). This is great for most enterprise applications, but it isn't always enough for HPC. HPC applications such as computational fluid dynamics (CFD) or automotive crash simulations require extremely low latency (under 2 microseconds) and massive throughput. Oracle Cloud Infrastructure meets this requirement with cluster networking featuring Mellanox RDMA. Oracle offers RDMA over Converged Ethernet (RoCE) v2, an extremely low-latency protocol delivered over (fast) standard Ethernet. Many highly parallel HPC workloads are built around a message-passing interface (MPI) that demands lower latency than can be offered in a typical cloud offering. RDMA cluster networking allows servers to talk to each other directly rather than having to traverse the traditional TCP/IP stack, providing the lowest-latency and highest-throughput networking in the industry.

*“Now is the time for cloud infrastructure architects to look at RDMA-based interconnect to increase their compute, memory and storage performance,” said Motti Beck, senior director of enterprise market development at Mellanox. “Deploying Mellanox’s RoCE v2 solutions in OCI enables Oracle to achieve their SLAs at a lower cost while fulfilling challenging cloud performance and efficiency metrics.”*

### Data Transfer

In addition to data-transfer services for moving large amounts of data, Oracle offers no-cost ingestion for all data and the first 10 TB of data egress for free. Typically, HPC customers opt for a dedicated *Fast Connect* through their ISP. Oracle charges per port for the Fast Connect, which means that you pay once for a port with no additional charge for data egress. This can mean tens or hundreds of thousands of dollars in savings and of course it removes concern about data transfer or vendor lock-in.

*“We benchmark the performance of the latest cloud hardware with HPL, two VASP simulation cases, one GROMACS case and MPI Benchmarks. Our findings demonstrate that Oracle Cloud outperforms other cloud vendors due to the latest generation of the hardware and fast interconnect network.”*

—Exabyte.IO team

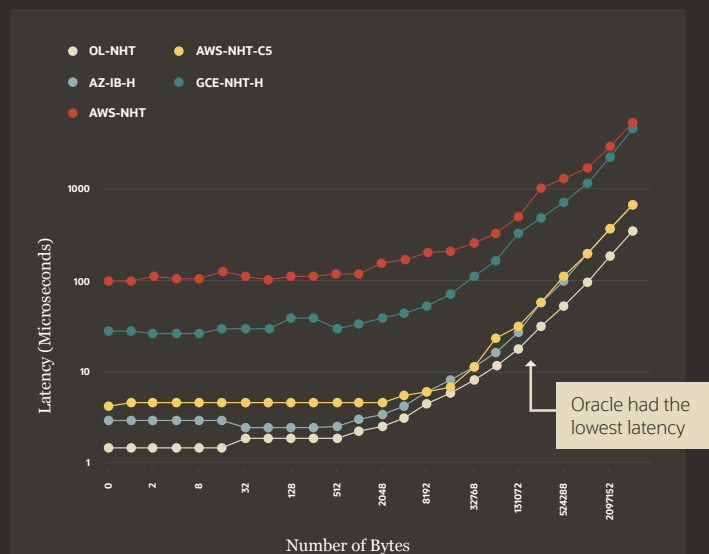


Figure 1: Oracle Demonstrates the Lowest Latency

The graphs shows the result of Intel MPI benchmarks running on Amazon Web Services C5 instances, Azure H-series VMs and Oracle hardware.

[docs.exabyte.io/benchmarks/2018-11-12-comparison/](https://docs.exabyte.io/benchmarks/2018-11-12-comparison/)

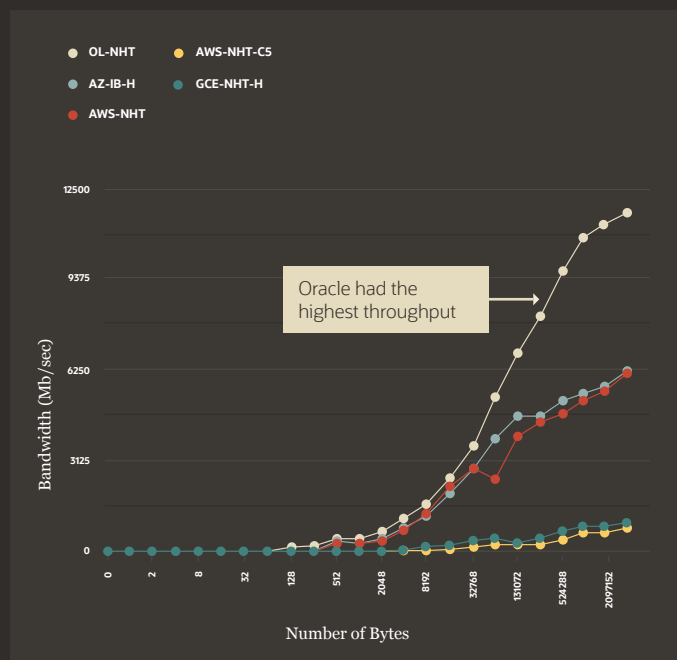


Figure 2: Oracle Demonstrates Highest Throughput

## HPC Storage

Like most cloud providers, Oracle offers a range of basic storage options including various flavors of object, file, and block storage. But Oracle also has offerings that more specifically map to the needs of high-performance computing workloads.

- **Local NVMe SSD:** High-speed local flash storage ideal for large databases, high-performance computing (local scratch and logs), and big data workloads, such as Apache Spark and Hadoop.
- **Block volumes:** Networked block storage services offering up to 60 IOPS per GB, up to a maximum of 25,000 IOPS per volume, backed by Oracle's highest-performance SLA.
- **Parallel file systems:** HPC systems often require larger data sets and higher performance than standard enterprise file servers can provide. Parallel file systems built with Oracle's high-performance bare metal instances can easily achieve over 60 GB/sec throughput at a price of less than five cents per gigabyte per month.
- **The fastest file servers in the cloud:** Oracle and IBM have partnered to provide high-performance file servers at scale on Oracle Cloud Infrastructure. Using IBM Spectrum Scale, you can now build a high-performance computing file server on Oracle Cloud Infrastructure Block Storage for your HPC or big data applications. This file server deploys in minutes, scales easily, costs a few cents per gigabyte per month, and provides incredible performance—ranked in the top 20 file servers worldwide on premise or in the cloud!<sup>1</sup>

## Secure by Design

Security in the cloud that has been recognized as more secure than what is offered on premise by 72 percent of those surveyed in the [2019 Oracle Cloud Threat Report](#). In fact, security is now named as the biggest benefit to the cloud by 66 percent of C-level executives, putting it ahead of both cost and scalability. Oracle Cloud Infrastructure is secure by design. As a platform for Oracle cloud offerings, applications, and databases, it includes architectural security innovations including isolating the network virtualization layer and deploying physical hosts with pristine firmware. Oracle Cloud Infrastructure security provides protection for data and applications by combining several tiers of defense across the network, physical hardware, and web-facing layers with secure and hardened operational procedures. As a result, customers benefit from the following security capabilities:

- A smaller attack footprint to prevent threats from gaining access to customer data and spreading to other customer tenants
- Granular customer isolation with a separate network virtualization layer and physical hosts deployed with pristine firmware
- Multiple layers of defense with built-in firewalls, DDoS, and encryption from the physical hardware to the web layer
- Infrastructure monitoring and identity enforcement to remediate risks and ensure the right people get access to the right information
- Expedited compliance for all regions through consistent architecture
- A secure cloud platform for data and applications running on the cloud and on premise

## The Best Guarantees in the Industry

Enterprises demand more than just availability from their cloud infrastructure. Mission-critical workloads also require consistent performance, and the ability to manage, monitor, and modify resources running in the cloud at any time. Only Oracle offers end-to-end SLAs covering performance, availability, and manageability of services. It's not enough for your IaaS resources to be merely accessible; they should consistently perform the way you expect them to. Oracle is the first cloud vendor to guarantee performance, so you can rely on your infrastructure for enterprise applications. As an example, if Oracle cannot provide at least 90 percent of published network performance at least 99 percent of the time, customers can claim up to 25 percent credit on affected services. Our SLAs aren't just targets; they are commitments.

In addition to network and storage performance SLAs, Oracle also offers management and availability SLAs. Availability SLAs cover compute, block volumes, object storage and FastConnect. Manageability SLAs cover compute, block volumes, and database management systems.

*“Oracle offers bare metal and virtual compute instances as well as NVMe versions with up to 25 TB of NVMe storage for performance unlike anything else seen in the cloud.”*

...

*“The bottom line is that forward-thinking businesses that can derive measurable value from high-performance cloud should definitely be evaluating what Oracle has going on.”*

—Storage Review Enterprise Lab, 2018

[Read the report](#) from *Storage Review*



## HPC ISV Ecosystem

Application vendors use Oracle Cloud Infrastructure to expand their reach, bringing more capability to more customers.

3D rendering and animation

Neural networks and deep learning

Computational fluid dynamics

Computer aided engineering

Scientific computing



View the webcast, [From Artificial Intelligence to Engineering Simulations in the Cloud](#).

## A New Philosophy for IaaS

Oracle developed its IaaS offering based the strategy of taking the best attributes of on-premise computing environments and making them available in the cloud. For business applications, that was Exadata. For high-performance applications, it's the HPC *shapes* with bare metal CPUs, and flat networking. Thus, every element of Oracle Cloud Infrastructure, from its best-in-class server hardware to its fully switched networks to Oracle's service level agreements (SLAs), has been chosen and executed with this objective in mind. Oracle's flexible pricing works well for the iterative nature of HPC applications, since you can spin up plenty of capacity for big jobs, then spin it back down when you're done. Or spin it back up with a completely different configuration to meet the needs of a different job.

- Oracle doesn't oversubscribe its compute, network, or storage resources. This ensures unbeatable performance and consistency across enterprise, cloud native, and HPC workloads.
- Oracle bare metal servers are some of the fastest in the industry, opening the door for enterprises to run HPC simulations requiring fast and scalable local storage in the cloud.
- Each bare metal server is isolated from all other tenants, giving you complete control over the environment.

By contrast, the most popular offerings from Amazon Web Services (AWS) and Microsoft Azure are focused on rapid, flexible deployments of nonmission-critical workloads. These first-generation cloud offerings simply weren't designed to handle performance-intensive workloads because they are virtualized and oversubscribed. While Amazon and others are starting to address the high-performance market with higher-performance offerings, Oracle still leads the way. Oracle has the fastest cluster networking, the fastest file server, the latest CPUs and GPUs, and a continuing commitment to the HPC market.

Furthermore, Oracle IaaS pricing is straightforward and predictable. While AWS has hundreds of SKUs for its most basic services, and has the potential for unexpected charges, Oracle has no hidden costs associated with using its cloud (which customers often encounter with Amazon's inter region and inter zone data-transfer costs).



## Case in Point: Fluid Dynamics for Science, Engineering, and Research

**Organization:** Altair provides software and cloud solutions for product development, high-performance computing, and data intelligence. Its software applications optimize design performance across multiple disciplines encompassing structures, motion, fluids, thermal management, electromagnetics, system modeling, and embedded systems.

**The challenge:** Altair's customers run complex physics and flow simulations to study the aerodynamic properties of buildings, vehicles, and environmental problems. For example, Altair's nanoFluidX application predicts the flow in complex geometries with complex motion, such as oiling in powertrain systems with rotating gears and shafts. High-end GPUs provide the ideal computing platform to handle these strenuous workloads, but they aren't always available for Altair's customers.

**The solution:** Thanks to a partnership between Oracle and Altair, customers can run Altair's computational fluid dynamics solvers in a scalable, high-performance cloud environment. Altair's advanced CFD solvers yield rapid simulation results for even the most complex cases. Oracle offers cutting-edge GPU instances and accompanying storage capabilities that preclude the need to move large amounts of data.

**The results:** With this on-demand solution from Oracle and Altair, engineers don't have to wait in job queues or endure long HPC hardware procurement cycles. Altair provides their product as a service on Oracle Cloud Infrastructure, making it easy for users to submit and monitor jobs and visualize results in the cloud. Oracle's state-of-the-art NVIDIA GPUs provide a 4x speed increase over CPU-based solutions. Customers obtain exceptional computational performance immediately, without the cost or build-out time required for an on-premise solution.

Watch the video, [Altair Selects Oracle Cloud Infrastructure for CFD](#).



## Case in Point: Special Effects, Animation, and 3D Rendering

**Organization:** Snowball Studios produces top-quality 3D animation and visual effects for clients in the film, television, computer games, advertising, and mobile industries. Positioned at the heart of the Israeli hi-tech scene, and with offices in London and Toronto, Snowball's digital creations are aired all over the world.

**The challenge:** Large brands such as Disney, Mattel, and many other industry leaders depend on Snowball to run commercial animation projects that involve rendering thousands of frames. These projects place huge demands on the infrastructure during peak periods—such as when creating a new commercial or rendering a full-length film. It often takes several hours of rendering time to create each individual frame. Acquiring and maintaining the necessary hardware and software to run these resource-intensive jobs was a mounting challenge.

**The solution:** Snowball relies on Oracle to supply advanced compute, storage, networking, and software technologies on demand. Oracle Cloud Infrastructure can be scaled instantly to handle complex processing tasks, then shut down between projects.

*“Oracle uses superior servers with more memory, stronger processors, and more cores. When we considered what we would get for the same amount of money from other cloud providers, we saw that we were getting more from Oracle.”*

—Yoni Cohen, founder and chief creative officer, Snowball Studios

**The results:** Snowball has grown to be the leading 3D animation and visual effects facility in its region, combining the best international creative talent with cutting-edge HPC technology. The company does not need to acquire, manage, update, and maintain a massive technology infrastructure to do this work efficiently.

Watch the video, [Snowball Studios](#).



# Get Started Today

HPC is used for compute-intensive, complex problems in manufacturing, media and entertainment, seismic research, AI, fluid dynamics, and many other computationally intensive domains. These applications often require high network performance, fast storage, high compute capabilities, and significant memory.

You can now run any HPC workload on Oracle Cloud Infrastructure with the same robust, scalable, and predictable performance that you would expect from a high-end on-premise HPC infrastructure. It's easy to get started: just add your data—no configuration is required.

## Why Move HPC Workloads to Oracle Cloud Infrastructure?

- Pay for compute-intensive workloads as predictable operational expenses, and avoid massive capital expenditures
- Gain performance that matches and even exceeds on-premise deployments
- Scale infrastructure up and down quickly, and pay only for the resources that you use
- Complete simulations, renderings, and AI training fast with the highest-performing IaaS among public clouds
- Focus on your simulations and renderings—not on keeping up with the latest hardware and software

You can try out Oracle Cloud Infrastructure for free at [oracle.com/cloud/free/](https://oracle.com/cloud/free/)

You can sign up for a cloud account today to receive always-free access to compute virtual machines, object storage, data egress, and other essential building blocks developers need to create applications on top of Oracle Autonomous Database. Plus get US\$300 of free credits for 30 days to use even more services to prototype your own applications, run machine learning models in notebooks, or try software from Oracle Cloud Marketplace.

Copyright © 2020, Oracle and/or its affiliates. All rights reserved. This document is provided for information purposes only, and the contents hereof are subject to change without notice. This document is not warranted to be error-free, nor subject to any other warranties or conditions, whether expressed orally or implied in law, including implied warranties and conditions of merchantability or fitness for a particular purpose. We specifically disclaim any liability with respect to this document, and no contractual obligations are formed either directly or indirectly by this document. This document may not be reproduced or transmitted in any form or by any means, electronic or mechanical, for any purpose, without our prior written permission. Oracle and Java are registered trademarks of Oracle and/or its affiliates. Other names may be trademarks of their respective owners. **VDL51082 202001**