Intersect360 Research White Paper:
ORACLE CLOUD INFRASTRUCTURE: SHAPING THE CLOUD FOR HPC

EXECUTIVE SUMMARY

The democratization of HPC is bringing greater computing capacity to an ever-broadening range of applications and workloads. Supercomputing levels of performance have become more widely accessible, in short-term or long-term consumption chunks, through advancements in cloud computing. As a result of increased accessibility, high-performance workloads have expanded dramatically to include analytics and machine learning workloads.

Cloud computing has been part of the HPC landscape for over a decade, but early iterations were hampered by a lack of maturity in available solutions. Recently, advancements in security, software license models, and services have allowed the segment to take off, and today, about half of HPC users make some use of cloud computing as an overall part of their extended infrastructure. Cloud computing has seen double-digit growth in five of the past six years in HPC, and with a 22.0% compound annual growth rate (CAGR), cloud is the fastest growing segment of the HPC market, projected to exceed $3.8 billion as a segment by 2024.

Oracle Cloud Infrastructure (OCI) has launched a major initiative to serve high-performance workloads, including HPC and machine learning, in a public cloud setting, and it has already landed major customer wins. A key OCI feature is the availability of bare-metal instances: single-tenant deployments without hypervisors or VMs. These provide dedicated computing that free HPC users from overhead, paving the way to predictable performance and scalability.

OCI bare-metal instances come in different “shapes” to suit various workloads. HPC shapes include the latest Intel Xeon or AMD EPYC high-frequency processor options, as well as machine learning–oriented configurations with the latest NVIDIA GPUs. Beyond the processing layer, scalability in OCI offers high-bandwidth cluster networking and scalable, high-performance storage with NVMe solid-state disks and a choice in parallel file systems.

The real question for HPC is time-to-solution, and OCI is focusing its efforts there. OCI recognizes the need for specialized investment to meet the scalability and performance needs of HPC applications. With its bare-

P.O. Box 60296 | Sunnyvale, CA 94088 | Tel. [888] 256-0124
www.Intersect360.com | info@Intersect360.com

metal offerings, HPC shapes, and key industry partnerships, OCI is positioned to change the conversation about HPC in the cloud.

P.O. Box 60296 | Sunnyvale, CA 94088 | Tel. [888] 256-0124
www.Intersect360.com | info@Intersect360.com

MARKET DYNAMICS

The Democratization of HPC

High Performance Computing (HPC) is often associated with world-leading research organizations: publicly funded supercomputing research labs whose charter is to advance the forefront of scientific understanding. These sites form the visible tip of the iceberg in HPC deployment, and they continue to enable new breakthroughs in both hardware and software at scale. But for years, the volume of HPC usage—and the majority of its growth—has come from a wider swath of commercial and academic implementations that are driven by the day-to-day need for innovation.

This democratization of HPC, bringing greater computing capacity to an ever-broadening range of applications and workloads, is not new, but it is accelerating. Previous generational improvements have allowed supercomputing levels of performance to be reached by clustering together best-of-breed components—which are still familiar to enterprise computing—at the individual server level. HPC systems may have special configurations, but the processing elements, servers, and often interconnects are the same as are available to any buyer.

Over the past decade, this dynamic of democratization has intensified along two separate dimensions. First, supercomputing levels of performance have become even more widely accessible, in short-term or long-term consumption chunks, through advancements in cloud computing. As a result of increased accessibility, high-performance workloads have expanded dramatically to include analytics and machine learning workloads, driven by successive enterprise trends in Big Data and artificial intelligence (AI).

High-performance workloads have expanded dramatically to include analytics and machine learning workloads, driven by successive enterprise trends in Big Data and AI.

HPC in the Cloud

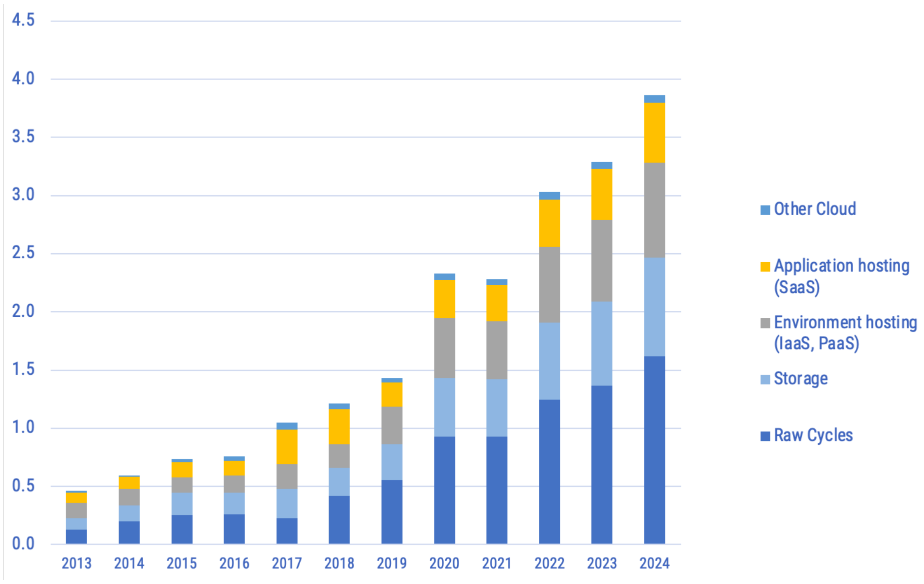
The potential appeal of cloud is simple: by renting rather than buying, HPC becomes more economical for accommodating peak workloads over-and-above baseline capacity, or for newcomers to HPC who wish to take a supercomputer out for a spin without committing the resources to buying one.

Cloud computing has been part of the HPC landscape for over a decade, but early iterations were hampered by a lack of maturity in available solutions. Recently, advancements in security, software license models, and services have allowed the segment to take off, and today, about half of HPC users make some use of cloud computing as an overall part of their extended infrastructure. Cloud computing has seen double-digit growth in five of the past six years in HPC, and with a 22.0% compound annual growth rate (CAGR), cloud is the fastest growing segment of the

HPC market, projected to exceed $3.8 billion as a segment by 2024. (See Chart below.[1])

Consumption of HPC in Public Cloud ($B), 2013 – 2019 Actual, 2020 – 2024 Forecast

Intersect360 Research, 2020



The high variability in this forecast is due to the COVID-19 pandemic. Beyond the lives it has claimed, the novel coronavirus has disrupted markets to an extent not seen before in our lifetimes. While the full effect of the novel coronavirus on our world is far from known, Intersect360 Research has predicted its near-term effects on HPC.

While many vertical markets will see a short-term downturn in capital expenditures in 2020, we expect HPC cloud spending to experience a spike, over and above our previous high-growth forecast. As stated in our forecast guidance:

> *Cloud computing provides flexibility amid uncertainty. It absorbs sudden increases in demand and also bridges the gap when procurements are delayed. Volatility suits cloud computing.* [2]

Cloud computing in HPC has fully emerged from its nascent state. Barriers such as data movement and security hindered early adoption, but these

"Cloud computing provides flexibility amid uncertainty. It absorbs sudden increases in demand and also bridges the gap when procurements are delayed. Volatility suits cloud computing."

---

[1] Intersect360 Research, *Worldwide HPC 2019 Total Market Model and 2020–2024 Forecast: Cloud Categories,* August 2020, plus historical data.

[2] Intersect360 Research, *Worldwide High Performance Computing Market Forecast: Pre-Announce Guidance Related to the Effects of COVID-19*, March 2020.

have been addressed as cloud computing has become a staple of enterprise computing.

Furthermore, virtualization—commonplace in enterprise clouds—has historically had little presence in HPC. HPC systems, once configured, haven't needed the shape-shifting capabilities that virtualization provides, and HPC administrators have become allergic to any overhead that threatens application performance. HPC is increasingly finding a home on single-tenant, "bare metal" clouds that are free of hypervisors or virtual machines (VMs).

As cloud computing has evolved, the value proposition has evolved with it. Sophisticated discussions of cloud go beyond penny-pinching arguments over what is cheapest. With its expanding set of capabilities, cloud computing has the potential to offer premium value over in-house computing, for example, by providing a test bed of new technologies for a broadening set of emerging workloads in HPC.

New HPC Workloads

As noted above, HPC is a long-term stable market, specifically because we are unlikely to run out of science to do in the foreseeable future. These technical computing applications remain as the enduring bedrock of HPC. But beyond these traditional HPC applications, the past ten years have brought about a dramatic expansion of HPC to include additional high-performance workloads.

Since the Big Data boom of the early 2010s, data science and advanced analytics have become an ongoing part of enterprise computing. And in the past few years, another even larger trend has exploded onto the landscape: AI—or more specifically, machine learning. While traditional scientific computing applications are mostly deterministic—a set of inputs is calculated on, and the math creates an output—machine learning mimics the way humans learn, experientially. (A child does not learn to catch a frisbee by doing physics calculations to determine its path, but by predicting its path based on learned patterns.) The training aspects of machine learning are compute-intensive and have become a common part of HPC environments.

Many scientific and business applications can benefit from machine learning approaches. Among traditional HPC users, there are several major applications in which machine learning can augment existing methods. Financial institutions can better predict risks of investments. Pharmaceutical companies can identify patterns in potential drug efficacy. Weather forecasters can use experiential data to plot the path of tornadoes. Beyond these areas, several breakthrough applications have revolutionary promise, such as autonomous driving and personalized medicine.

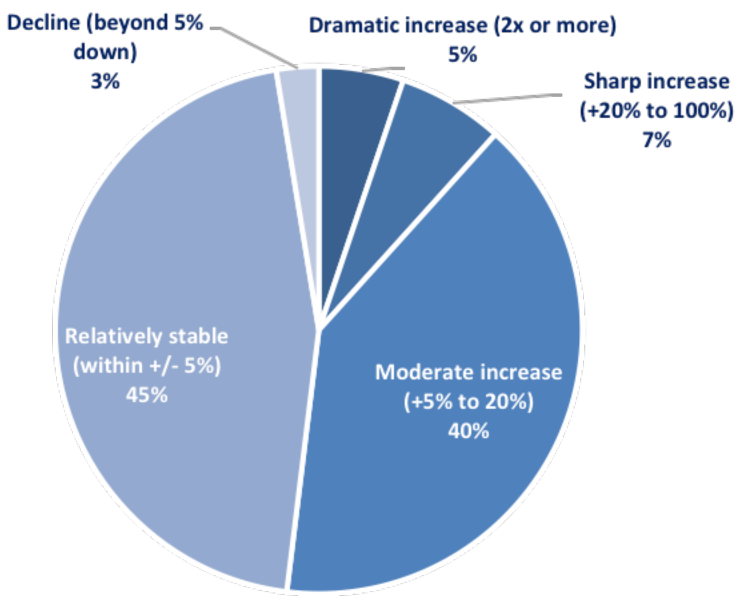
Over 60% of HPC users have begun to implement machine learning.

Among those, there has been a significant increase in budgets for high-performance workloads.

Against this backdrop, Intersect360 Research is seeing an increase in organizations' budgets for high-performance applications. Over 60% of HPC users have begun to implement machine learning, and among those, there has been a significant increase in budgets for high-performance workloads.[3] (See Figure below.)

The result of this expansion is that organizations find they must now support more types of applications within an existing infrastructure. This can present significant challenges. The different applications can have different technical requirements: one needs more memory, one needs more throughput, and another needs specialty processors for faster computation. And even when the workloads can coexist, clashing priorities can cause delays or increased costs. Managing these conflicts without overrunning budgets has become the biggest challenge facing HPC-using organizations.

Effect on High-Performance Workload Budget Related to Incorporation of Machine Learning
Intersect360 Research, 2020



3 Intersect360 Research, HPC User Budget Map survey data, 2020.

INTERSECT360 RESEARCH ANALYSIS

OCI: A New Cloud Infrastructure for the New HPC

Against these ever-escalating dynamics, a new option has emerged to provide high-performance cloud resources for HPC, and it comes from a familiar enterprise leader: Oracle. Oracle Cloud Infrastructure (OCI) has launched a major initiative to serve targeted high-performance workloads, including HPC and machine learning, in a public cloud setting, and it has already landed major customer wins. Whereas other major clouds—such as Amazon Web Services, Microsoft Azure, or Google Cloud—began in other markets that focused on virtualization of resources with fungible, non-discriminated computing assets, OCI has started fresh, building targeted instances with scalable workloads in mind.

A key OCI feature is the availability of bare-metal instances: single-tenant deployments without hypervisors or VMs, and no Oracle software. These provide dedicated computing and free HPC users from overhead, paving the way to predictable performance and scalability. Bare metal caters to HPC users' tendencies to determine the complete stack, from operating environment to application.

Bare metal caters to HPC users' tendencies to determine the complete stack, from operating environment to application.

Additionally, OCI commits not to oversubscribe its computing, networking, or storage resources. This eliminates resource contention within the cloud offerings, so there are consistent response and performance times for users' applications.

OCI bare-metal instances come in different "shapes" to suit various workloads. HPC shapes include the latest Intel Xeon or AMD EPYC high-frequency processor options, as well as machine learning–oriented configurations with the latest NVIDIA GPUs. Users can still select VMs when appropriate, or containers such as Kubernetes, which are often used to provide an extra layer of abstraction or to provide consistency in runtime environments.

Beyond the processing layer, scalability in OCI comes from what Oracle refers to as cluster networking—a 100 Gigabit per second (Gbps) interconnect backend with Mellanox ConnectX-5 network interface cards (NICs), capable of Remote Direct Memory Access (RDMA) over Converged Ethernet (RoCE). A major advantage of RoCE (pronounced "rocky") for HPC is the ability for processors to compute on data referenced by remote nodes in the cluster, passing pointers instead of data, which saves on data movement within the cluster, thereby reducing latency and improving application performance. OCI was among the first to include RDMA, and OCI's virtualized network offers these benefits for HPC applications with no additional charge.

P.O. Box 60296 | Sunnyvale, CA 94088 | Tel. [888] 256-0124
www.Intersect360.com | info@Intersect360.com

The data storage is architected for performance as well. OCI shapes for HPC include on-node NVMe solid-state disks—up to 6.4 terabytes—and high-performance file-, block-, or object-oriented storage. OCI offers a choice in parallel file systems, with optimized configurations on either Lustre, BeeGFS, or IBM Spectrum Scale (previously—and sometimes still, informally—known as GPFS). For data science and analytics applications, OCI offers both Hadoop and Apache Spark, with I/O-intensive configurations.

Using BeeGFS, OCI scored in the top ten worldwide in I/O ranking in 2020, scoring among the world's top supercomputing institutions, with no other cloud provider in the top ten. OCI also ranked #19 using Lustre and #27 with Spectrum Scale.[4] And OCI was recently chosen[5] by Japanese research center RIKEN to be the public cloud provider behind its Fugaku supercomputer, announced in June 2020 as the world's most powerful supercomputer, according to the semiannual TOP500 list.[6]

OCI Partners for the HPC Ecosystems

For HPC users in the midst of migrating to cloud, OCI has signed up Altair as a key strategic partner. Altair is a unique software provider in the HPC industry. Altair is among the top ten providers of middleware and developer tools for the HPC market; its PBS/PBSPro products are among the top five most commonly used job management packages in HPC.[7] But unlike other middleware providers, Altair also offers end-user applications for engineering. Three different Altair applications—HyperWorks, OptiStruct, and Radioss—are among the top ten most commonly used general engineering packages in HPC,[8] and AcuSolve from Altair is among the top ten applications for computational fluid dynamics (CFD).[9] Altair is the only software vendor with top-ten deployments in all three of these HPC categories, and Altair additionally has its HyperMesh CAD software on the engineering desktop.

Because of its presence across engineering workflows, Altair is a critical partner for deploying those workloads to the cloud. The Altair HyperWorks Unlimited Virtual Appliance is integrated to burst workloads directly to OCI, in a supported, high-performance framework.

---

4 IO500 ranking reported by the Virtual Institute for I/O at https://www.vi4io.org/io500/, based on announcements from the ISC High Performance 2020 conference, https://www.isc-hpc.com.

5 https://blogs.oracle.com/cloud-infrastructure/oracle-cloud-infrastructure-behind-worlds-fastest-supercomputer.

6 https://www.top500.org.

7 Intersect360 Research, HPC User Site Census: Middleware and Developer Tools, 2019.

8 Intersect360 Research, HPC User Site Census: Engineering Applications, 2019.

9 Intersect360 Research, HPC User Site Census: CFD Applications, 2019.

Another partner OCI is collaborating with in the HPC space is Rescale. HPC organizations tend to approach any infrastructure decision from the standpoint of their applications and workflow. Rescale brings that sensibility to managed services for cloud computing, with tools to evaluate cloud instances and make them seamlessly available. As such, Rescale has integrated OCI into its HPC platform.

Rescale and Oracle customers can now utilize an HPC environment that is highly optimized for their applications, with administrative features that enable tighter control over hardware expenditures and ISV license utilization. To submit jobs, customers use a web portal to select job criteria, such as number of compute cores and memory. Rescale then provisions an HPC cluster dynamically and starts the job according to customer specifications. When the job is completed, the HPC resources are de-provisioned. The integration with Rescale gives OCI customers access to over 400 applications available on the Rescale platform, bringing their own licenses or with on-demand utility pricing.

With its bare-metal offerings, HPC shapes, and industry partnerships, OCI has already attracted experienced leaders in scientific and technical computing. Two in particular are indicative of the range of applications and organizations that are now finding a home in the cloud.

Manufacturing on OCI: Denso

Manufacturing is a key segment for OCI, and while any HPC segment sees constant innovation, the auto-parts world is currently beset by new categories of development driven by increasing adoption of electronic and hybrid vehicles. This includes both the obvious, primary need for electronic components, such as batteries, but also interesting secondary lines of development. As electric motors run quieter than traditional combustion-driven counterparts, many auto makers have found a demand for quieting other aspects of the drive environment, such as aerodynamics, air conditioning, and other vibrations.

Denso is the second-largest maker of auto parts by worldwide revenue, and they have begun utilizing OCI for a wide range of automotive applications, from fluid dynamics to acoustic analysis.[10] Koji Komura, General Manager of the VE Development Department in the Denso Techno technology group within Denso, explains, “A variety of simulations are utilized in our design such as strength analysis, impact analysis, system analysis, thermal analysis, etc. Especially, the amount of calculation in fluid noise analysis is enormous, and this is an area where HPC's ability is important.” According to Komura, the decision to use OCI

Royal Holloway says its compute time has been reduced from 47 hours to 20 hours per simulation.

Denso has cut calculations for its fluid noise analysis from 15 days down to three days.

These kinds of improvements do lead to lower costs, but this is a secondary effect compared to the top-line gains in productivity.

[10] https://www.oracle.com/au/a/evt/docs/high-performance-computing-on-oracle-cloud-infrastructure.pdf.

was driven by OCI offering "low cost, high performance, and support for the latest technology."

Climate Change Simulation on OCI: Royal Holloway, University of London

"Solving the world's greatest challenges" is one of the traditional catchphrases of HPC and supercomputing, and climate change stands as one of the great challenges of our time. Here there are many subfields of scientific contribution, and thereby also HPC.

One strategy for combating climate change is to explore strategies for atmospheric carbon mitigation. Researchers at Royal Holloway, University of London are using scanned rock data to study tomography at a microscopic level, modeling various minerals' capabilities for capturing and storing carbon dioxide. This 3D microtomography approach is data-intensive and can require massive computational resources. "We can simulate carbon capture sequestration scenarios, address complex environmental problems, and drive meaningful change in the world. Oracle has helped us break the barrier of computational power in the lab," says Saswata Hier-Majumder, Reader in Geophysics from Royal Holloway.[11]

Evolving Clouds to Meet New HPC Needs

Early clouds treated computing as an undifferentiated commodity. Today, OCI recognizes the need for specialized investment to meet the scalability and performance needs of HPC applications. This is important not only in base functionality, but also (and more importantly) in shifting the discussion of utility computing economics.

For too long, too many cloud "value propositions" have relied on an often-flawed notion that clouds are naturally cheaper. For HPC and other applications, they certainly can be, depending on many factors, but regardless, it's not the best start to the conversation. The real question for HPC is time-to-solution, and OCI is focusing its efforts there. Royal Holloway says its compute time has been reduced from 47 hours to 20 hours per simulation. Denso has cut calculations for its fluid noise analysis from 15 days down to three days. These kinds of improvements do lead to lower costs, but this is a secondary effect compared to the top-line gains in productivity.

Innovation won't stop here. COVID-19 has HPC and cloud resources in the spotlight, and every industry will continue to pursue advancements in research and engineering. As they do, OCI is positioned to change the conversation about HPC in the cloud.

---

[11] https://www.oracle.com/a/ocom/docs/royal-holloway-061620.pdf.

P.O. Box 60296 | Sunnyvale, CA 94088 | Tel. [888] 256-0124
www.Intersect360.com | info@Intersect360.com

For more information about Oracle Cloud Infrastructure solutions for HPC, visit https://www.oracle.com/cloud/solutions/hpc.html.

White paper sponsored by Oracle Cloud Infrastructure. Full neutrality statement at https://www.intersect360.com/features/neutrality-statement.
P.O. Box 60296 | Sunnyvale, CA 94088 | Tel. [888] 256-0124
www.Intersect360.com | info@Intersect360.com