

High Performance Computing in Oracle Cloud Infrastructure

Oracle Cloud Infrastructure (OCI) is a deep and broad platform of cloud services encompassing IaaS, PaaS, and SaaS that enables you to build and run a wide range of applications in a scalable, secure, highly available, fault-tolerant, and high-performance environment. High Performance Computing (HPC) on OCI brings powerful, cost-effective computing capabilities to solve complex mathematical and scientific problems.

HPC on OCI rivals the performance of on-premises solutions with the elasticity and consumption-based costs of the cloud, offering on-demand potential to scale to thousands of cores simultaneously. With HPC on OCI, customers get access to high-frequency processors; fast and dense local storage; high throughput, ultra-low latency cluster networking; and the tools to automate and run jobs seamlessly.

HPC in the cloud is heavily used today for simulation and modeling in diverse fields and applications, including aerodynamics and structural mechanics analysis in product design, seismic data analysis and reservoir simulation in oil and gas, genome sequencing and molecular dynamics in computational biology, and electronic design automation (EDA) for semiconductors, and many other use cases.

Applications that rely on AI and ML workloads benefit from running GPUs as they allow companies to solve complex problems and innovate faster. Because AI workloads need fast and large volumes of storage and computing power, many customers are leveraging Oracle's partnership with NVIDIA to run these workloads on OCI.

Why HPC on OCI?

OCI provides industry-leading compute infrastructure.

- OCI features a performant platform based on bare-metal servers and a high throughput ultra-low latency network.
- Enterprise-level security for sensitive data.
- OCI offers the latest CPUs and GPUs in the market from AMD, NVIDIA, Arm, and Intel.
- OCI's cluster network is very flat with low latency and high bandwidth, providing predictable, scalable performance.
 - RDMA over Converged Ethernet (RoCE V2).
 - ~1.5µs latency.
 - Max 2 hops between nodes.
 - Never over-subscribed, so there is no noisy-neighbor or jitter.

OCI provides scalable, independent storage.

- OCI HPC shapes come with a local NVMe disk that can be used as scratch space for accelerating performance, providing performance at no-charge.
- 3.8TB NVMe
- Up to 32TB Block Storage
- Fastest parallel filesystem performance including reference architectures for Lustre, Spectrum Scale, Weka, and others.

OCI has the lowest cost for compute, storage & network.

- OCI HPC shapes come with a local NVMe disk that can be used as scratch space for accelerating performance, providing performance at no-charge.
- 3.8TB NVMe
- Up to 32TB Block Storage
- Fastest parallel filesystem performance including reference architectures for Lustre, Spectrum Scale, Weka, and others.

Access to Oracle cloud engineers and premier technical services for accelerating workload migrations via Oracle Cloud Lift Services.

- Application architecture.
- Performance analysis.
- Hands on migration.
- Go-live support.
- Training on best practices.
- No additional cost for this service.
- Access to OCI HPC Specialist Team:
 - Subject Matter Experts, deep HPC background.
 - Excellent HPC services.
 - Enablement - the Oracle technical HPC Specialist team is available to support customer projects to move workloads to OCI.



ORACLE

HPC on Oracle Cloud Infrastructure

Secure, high-performance platform for your HPC workloads



Safe harbor statement

The following is intended to outline our general product direction. It is intended for information purposes only and may not be incorporated into any contract. It is not a commitment to deliver any material, code, or functionality, and should not be relied upon in making purchasing decisions.

The development, release, timing, and pricing of any features or functionality described for Oracle's products may change and remains at the sole discretion of Oracle Corporation.

We built our cloud from the ground up to break the rules

11 years after the first generation of cloud, we started with a clean sheet



Off-box virtualization

The way we manage OCI is entirely separate from your resources, maximizing isolation, performance, and security



Nonblocking networks, minimal charges

We optimized our networks so you get guaranteed bandwidth between your resources, with 90% lower costs to access data and 80% lower costs to serve data



Maximum computing density per MW

We pack over 230,000 cores into each megawatt and can deliver an entire cloud region in only 12 racks



Flex infrastructure

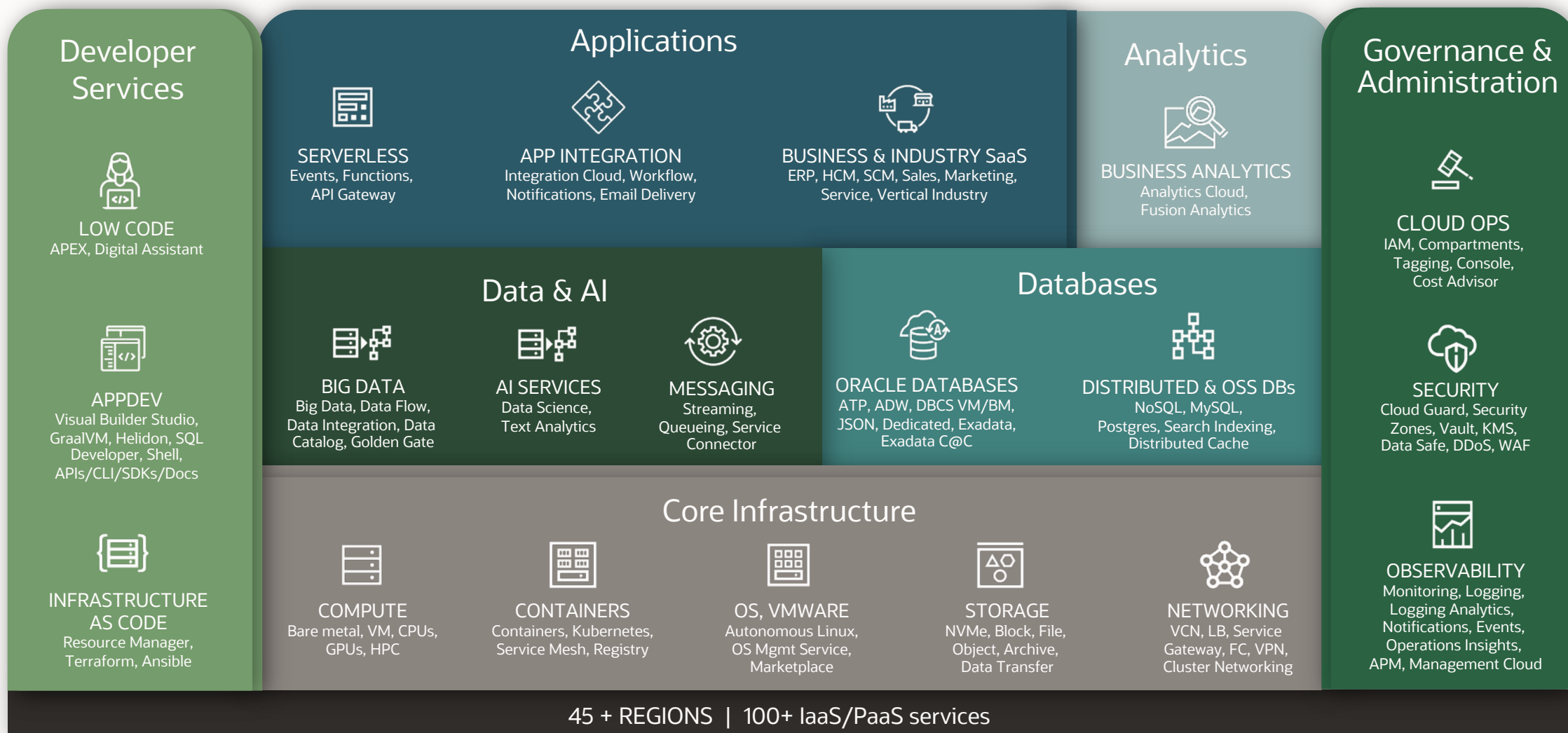
You can choose exactly the amount of cores, memory, and storage performance you need, and pay for exactly that, minimizing waste



Simple, predictable pricing

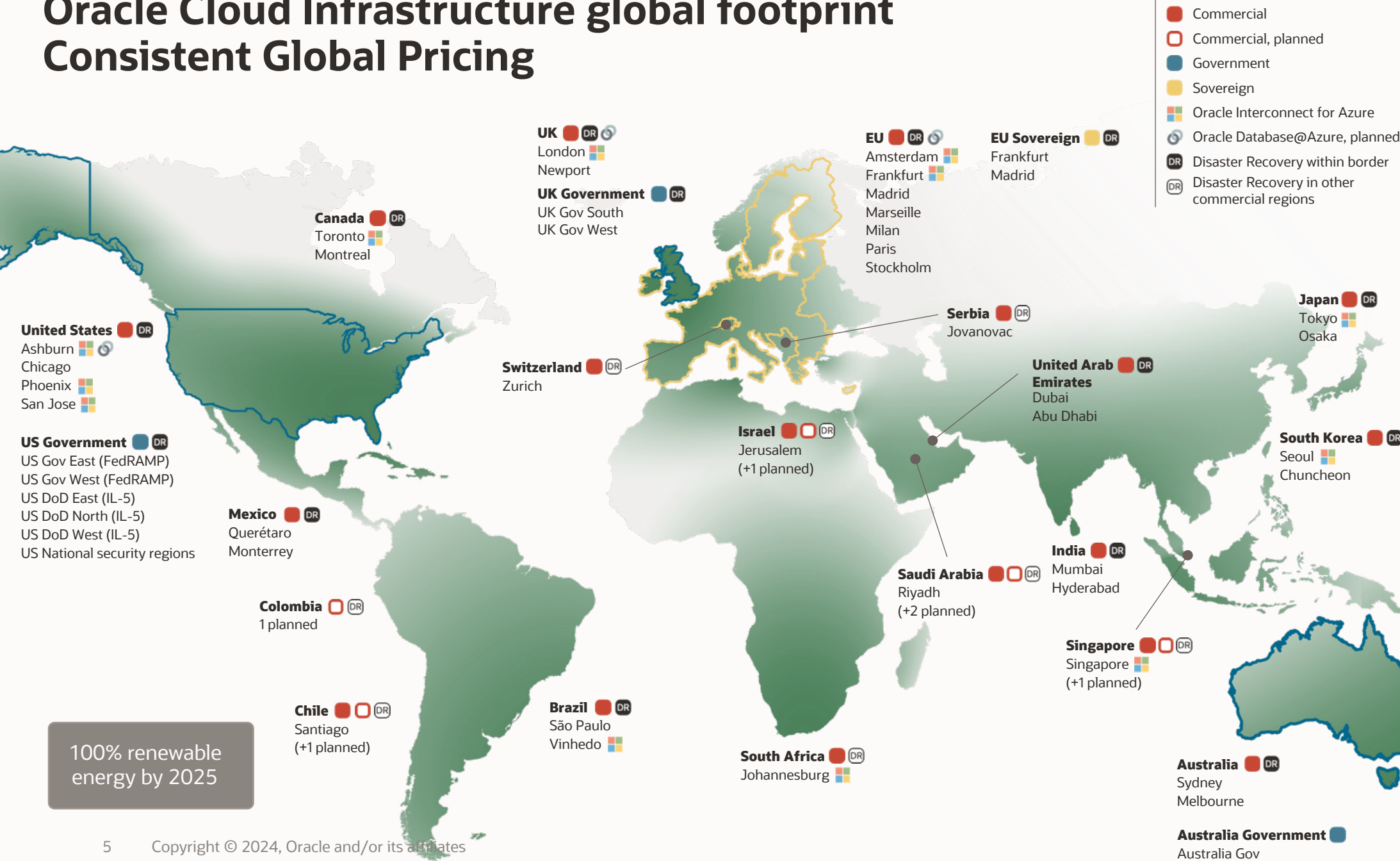
Our pricing is simple to understand, 50-90% lower than other hyperscalers, and consistent worldwide, so you get predictable savings with no surprises

A full-stack portfolio for Government, Defense & Classified Regions



Oracle Cloud Infrastructure global footprint

Consistent Global Pricing



100% renewable energy by 2025

\$18B+

Cloud Applications
Cloud Infrastructure
run rate

66%

cloud infrastructure
revenue growth

46

cloud regions

\$8.7B

CapEx in the
last 12 months



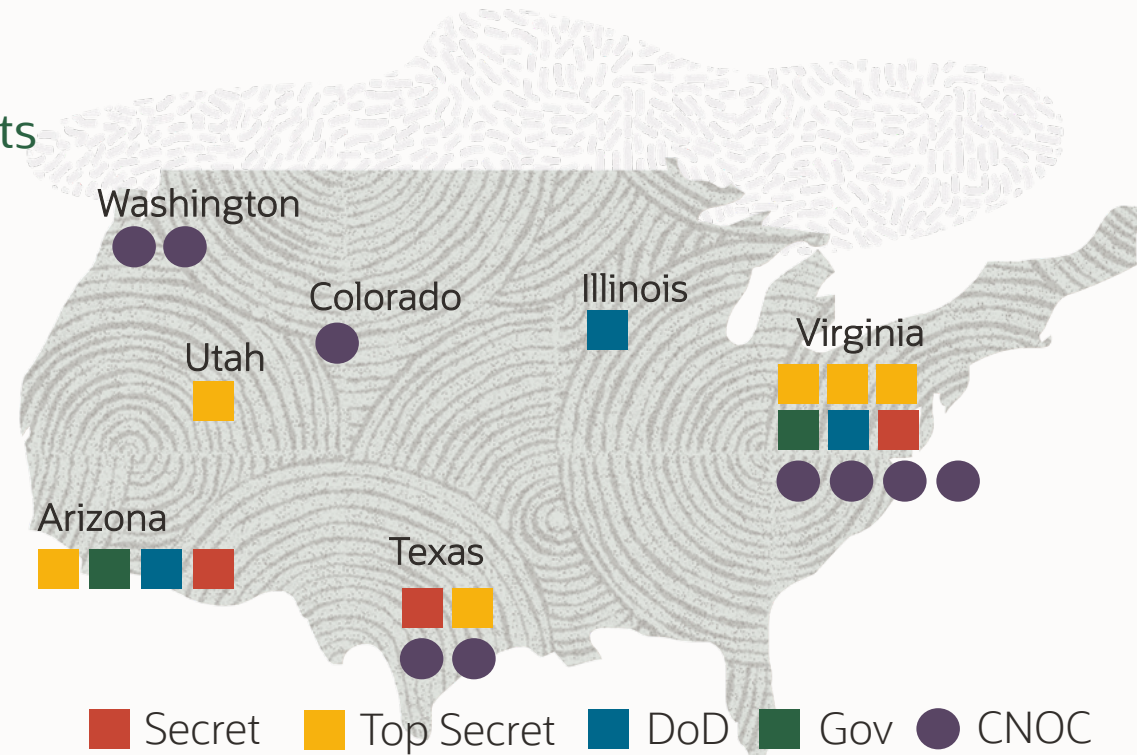
US OCI for Government

Highly secure regions that meet government requirements






Commercial customers can deploy applications in the US Government regions.

AWS and Azure increase prices for government compliance, we do **NOT**. All the government level of security comes with **NO** additional cost.

Our Government Regions have the same level of compliance as AWS and Azure.



The regions support the **highest** Government compliance standards

 FedRAMP High JAB P-ATO & Agency ATO		 DOD IL2, IL4 & IL5	
 BCAP In DoD Realm Only	 CJIS	 HIPAA	

US Government Regions	US Department of Defense (DoD) regions	US National Security Regions
OC2	OC3	OC 6, 7, 11 & 12
FedRAMP High JAB authorized DISA IL2 & IL4 authorized	DISA IL5 authorized Connections to East and West DISA BCAPs	Secret and Top Secret/SCI
Ashburn & Phoenix	Ashburn, Phoenix & Chicago	

Oracle offers a full stack of cybersecurity capabilities

Prevent







Block attacks and malicious traffic

-  **Distributed Denial of Service protection**
Automatic DDoS protection
-  **Web Application Firewall**
Internet-facing endpoint protection

INTERNET & EDGE

Monitor

Log, analyze, and audit activity

-  **Cloud Guard**
Security posture management
-  **Security Zones**
Security policy compliance
-  **Threat Intelligence**
Multi-source, actionable guidance
-  **Threat Detector**
Monitor for known threats
-  **Logging**
Single pane for service logs
-  **Fusion Apps Detector**
Monitor ERP and HCM apps
-  **Vulnerability Scanning**
Patch and port monitoring
-  **Auditing**
OCI API logging

MONITORING & PREVENTION

Mitigate

Isolate communications with secure and reliable networks

-  **Virtual Cloud Network**
Secure, isolated network
-  **Security Lists**
Virtual network firewall rules
-  **Network Firewall**
Advanced firewall service
-  **Bastion**
Time-limited SSH access
-  **Dynamic Routing Gateway**
Virtual router
-  **Fast Connect**
Dedicated, high-speed connection
-  **Virtual Private Network**
Secure connectivity over any network
-  **NAT Gateway**
Protected access to the internet

NETWORK

Protect

Hardware-enabled security built into the architecture

-  **Bare Metal Servers**
Servers with full customer control
-  **Hardware Root of Trust**
Protect from firmware attacks
-  **Signed Firmware**
Prevent rootkit installation
-  **Hardened Disk Images**
OS with expert security settings
-  **Off-box Control Plane**
Isolated admin of compute hardware
-  **Off-box Network Virtualization**
Encapsulated, separated traffic
-  **Oracle Linux & Oracle Enterprise Linux**
Performant, secure, enterprise Linux

COMPUTE

Encrypt

Encrypt and protect all data

-  **Confidential computing**
Encrypt VMs in motion
-  **Data Safe**
Monitor data usage in database
-  **Vault**
Hardware security module protection
-  **Key Management**
Encryption key administration
-  **Secrets Management**
Credential and similar administration
-  **Certificates**
Validation certificate administration

STORAGE & DATABASE

Access

Ensure authentication, authorization, and accounting

-  **Access Governance**
Proactive guidance for user actions
-  **OCI Identity and Access Management**
Control access to cloud resources
-  **Policies**
User access rules
-  **Federation**
Identity provider inter-operation

IDENTITY & OPERATOR ACCESS



Oracle and Palantir Join Forces to Deliver Mission Critical AI Solutions to Governments and Businesses

PR Newswire Thu, Apr 4, 2024, 8:00 AM EDT

Joint cloud, data, and AI solutions aim to accelerate decision-making from HQ to the tactical edge

Palantir selects Oracle Cloud Infrastructure to enhance efficiency, performance and sovereignty for a wide range of organizations





Oracle and Palantir Join Forces to Deliver Mission Critical AI Solutions to Governments and Businesses

Joint cloud, data, and AI solutions aim to accelerate decision-making from HQ to the tactical edge Palantir selects Oracle Cloud Infrastructure to enhance efficiency, performance and sovereignty for a wide range of organizations

AUSTIN, Texas and DENVER, April 4, 2024 /PRNewswire/ -- Oracle and Palantir today announced a partnership to provide secure cloud and AI solutions aiming to power businesses and governments around the world. Oracle's distributed cloud and AI infrastructure, combined with Palantir's leading AI and decision acceleration platforms, will help organizations maximize the value of their data—which will contribute to increasing efficiency, addressing sovereignty requirements, and help them outpace adversaries.

As part of the agreement, Palantir will move Foundry workloads to Oracle Cloud Infrastructure. Palantir will also make its Gotham and AI Platforms deployable across Oracle's distributed cloud: in public cloud regions; Oracle Cloud Infrastructure Dedicated Regions; Oracle Alloy; Oracle EU Sovereign Cloud; Oracle Government Cloud; Oracle Roving Edge, and Oracle's air-gapped regions for defense and intelligence customers. Oracle's vast cloud footprint and sovereign AI capabilities will allow more organizations to use Palantir's leading platforms for data integration and decision-making. Oracle and Palantir are proving there is no trade-off. Organizations can have the best cloud infrastructure, cloud applications, and AI while meeting the highest sovereignty and security standards.

"Oracle is the only hyperscaler capable of delivering its entire AI and cloud suite to any business or government anywhere in the world," said Rand Waldron, vice president, Oracle. "By combining the performance, scalability, and flexibility of Oracle Cloud Infrastructure with Palantir's leading data and AI platforms, we will help our customers win in any industry or environment."

"Palantir and Oracle are both dedicated to defending western interests and institutions around the world," said Josh Harris, executive vice-president, Palantir. "Oracle Cloud Infrastructure's unique ability to help customers meet their regulatory, performance, and security needs will increase our impact and help our global clients gain the full benefits of cloud and AI."

In any region, including commercial, sovereign, and government air-gapped environments, Oracle Cloud provides more than 100 cloud services and applications, including the latest innovation in generative AI, running on a blazing fast AI infrastructure. Oracle Cloud Infrastructure services and pricing are consistent across deployment types to simplify planning, portability, and management.

Palantir Artificial Intelligence Platform (AIP) enables organizations to bring the power of large language models to their enterprise networks, private data, and core operations – with maximum security and trust. The platform is designed to bring together disconnected data sources, logic assets, and systems of action into a single common operating picture. Oracle's AI strategy, including its wide range of partners, provides generative AI services and infrastructure that extend Palantir's AI capabilities to help customers accelerate decision-making. Oracle's long history in defense and intelligence provides a depth of experience and technology critical to the success of high-stakes missions. Together, Oracle and Palantir will bring powerful new capabilities to the defense industry.

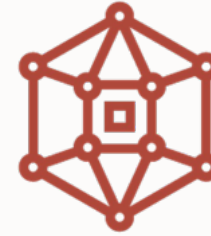
Oracle and Palantir will jointly sell and support cloud and AI services across government and commercial industries.

Four reasons why Engineering Simulation is better on Oracle Cloud ...



1. Instances Designed for HPC

- **Off-box Virtualization**
- **Bare Metal shapes** that provide performance comparable to on-prem with advantages of cloud
- **VM shapes** an option for flexible use cases



2. Flat & fast RDMA network

- RDMA over Converged Ethernet (RoCE V2)
- ~1.5 μ s latency
- Max 2 hops between nodes
- Never over-subscribed



3. Scalable, Independent Storage

- 3.8TB NVMe local disk, performance boost
- NFS Service: FSS
- Fast parallel filesystem performance: Lustre, Weka, Spectrum Scale....
- Object & Archive tiers



4. Lowest cost for Compute, Storage & Network

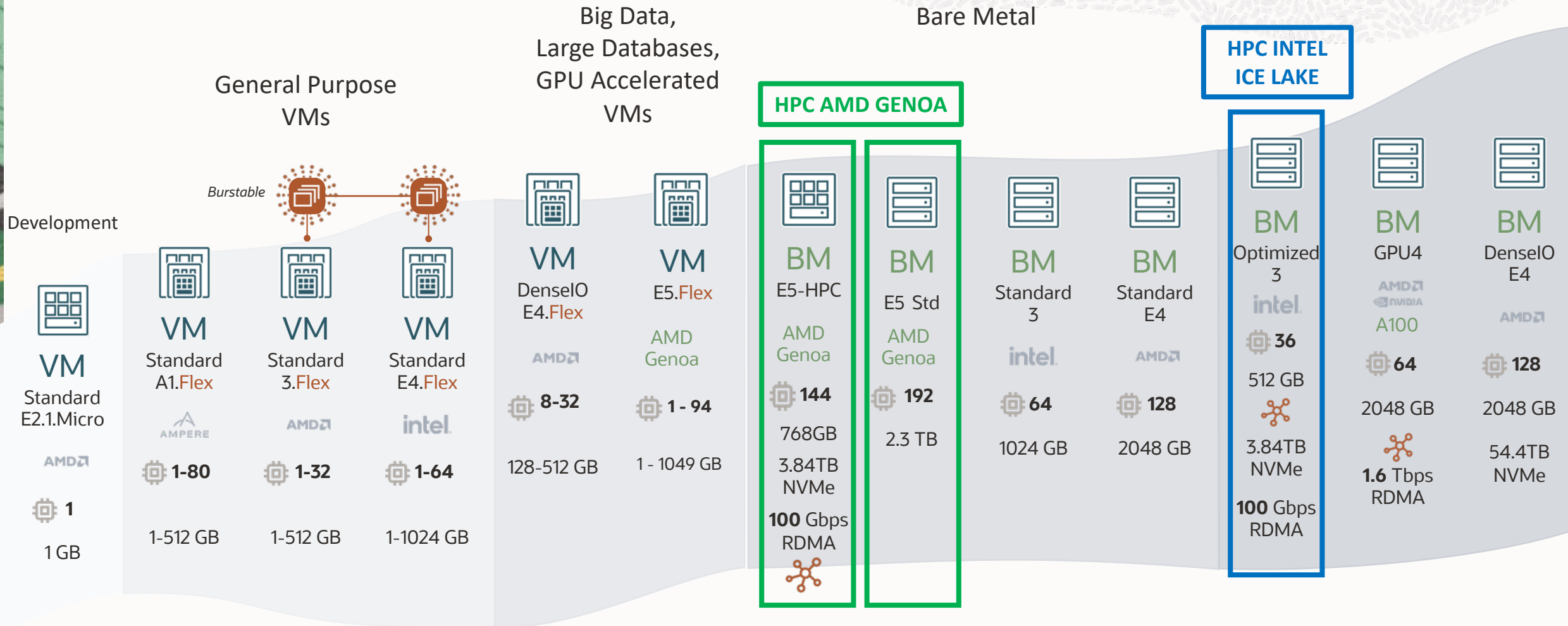
- **Overarching goal: maximize your ISV investment!**
- Latest CPU and GPU processors
- UCC – flexible billing to match usage requirements
- No hidden costs such as egress fees (FastConnect)

Partner Ecosystem



Fast and scalable VM, Bare Metal, and GPU compute

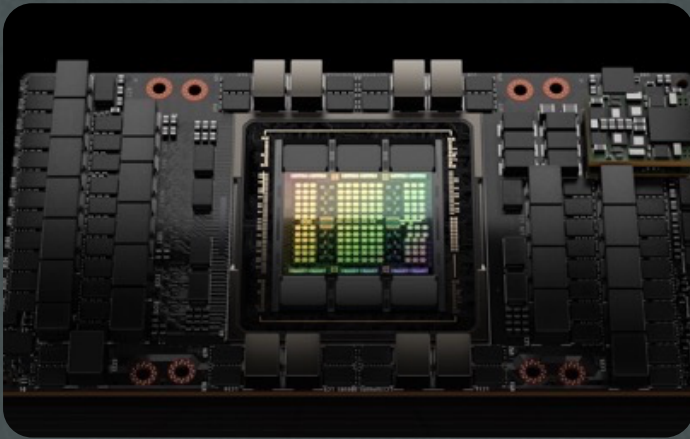
HPC, AI/ML,
Compute Clustering
Bare Metal



Represents CPU cores (not threads or vCPUs)

Inter-node network with RDMA running on RoCE (< 1.5 μs latency)

GPUs for AI and Accelerated Computing



OCI Compute with NVIDIA H100 and A100

Supports superclusters that can scale above 10K GPUs

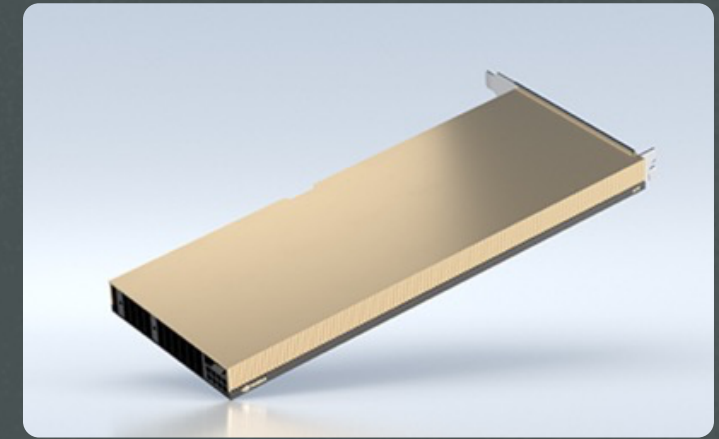
Bare metal servers with industry-leading 3.2 Tbps cluster network



OCI Compute with NVIDIA L40S

Supports clusters with up to 1K GPUs

Bare metal servers with industry-leading storage and memory capacity



OCI Compute with NVIDIA A10 GPUs

VM instances with one or two A10 GPUs

Bare metal instances with four A10 GPUs

Ideal for AI inference, graphics, and virtual workstations

OCI GPU Supercluster



Compute

Baremetal instances with
NVIDIA A100 & H100 GPUs



Non-blocking RDMA network

Ultra low latency : $\sim 2\mu\text{s}$
High bandwidth

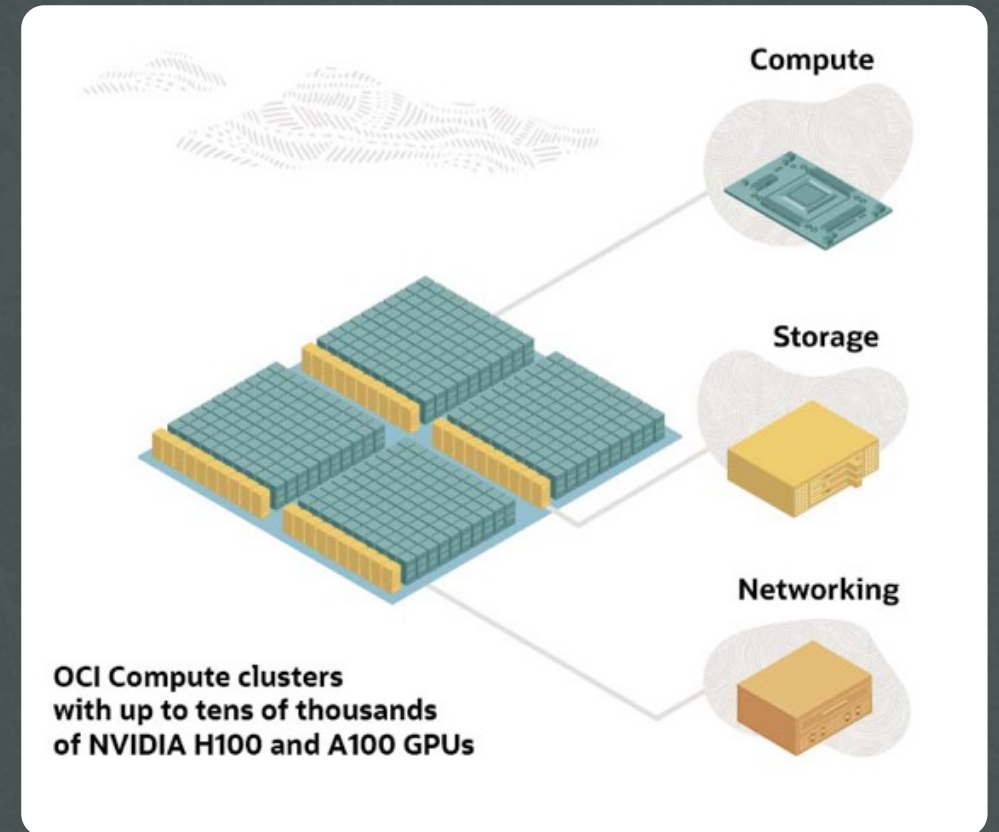
- H100: 3.2Tbps
- A100: 1.6Tbps



Storage

Block, File, Object
Local checkpointing

- H100: 61.4TB/node
- A100: 27.2 TB/node



OCI HPC Shape performance

■ HPC Software

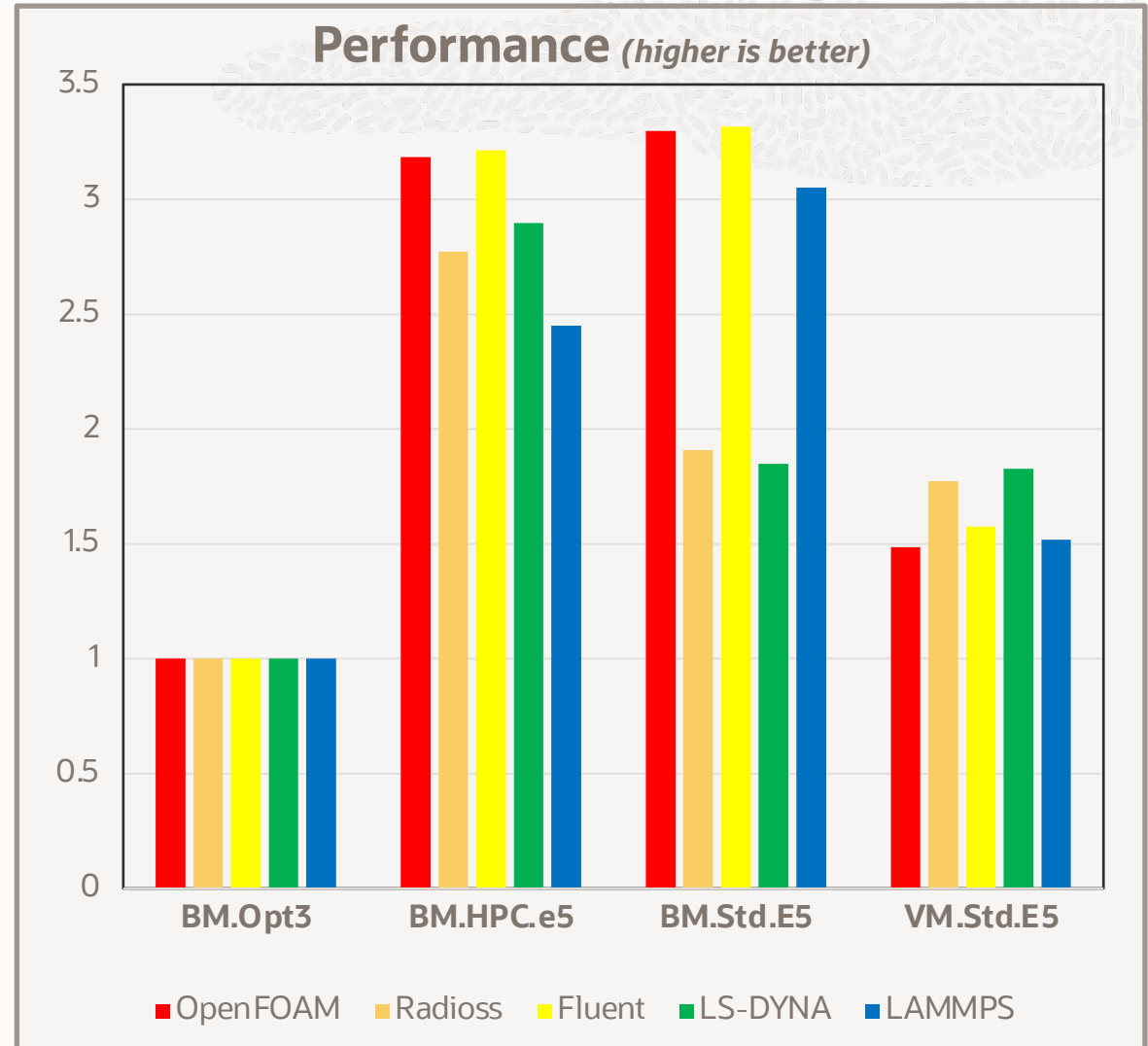
- OpenFOAM v2206
- Altair Radioss™ 2023
- Ansys Fluent® 2023R1
- Ansys LS-DYNA R14.0.0
- LAMMPS 2Aug23

■ OCI Hardware shapes

- BM.Optimized3.36
 - 36-core, 512GB mem, RDMA
- BM.HPC.E5.144
 - 144-core, 768GB mem, RDMA
- BM.Standard.E5.192
 - 192-core, 2306GB mem, non-RDMA
- VM.Standard.E5.Flex
 - 92-core, 384GB mem, non-RDMA

■ Observations

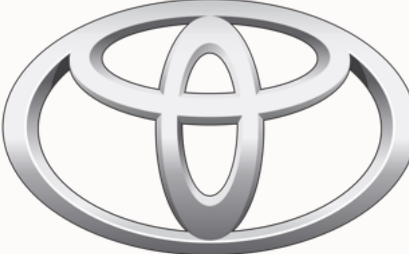
- New shapes offer significantly more single node performance
- Nodes with RDMA capable network are crucial for largest multi-node parallel jobs
- Some workloads cannot effectively use all cores on BM.Standard.E5.192 due to memory bandwidth limitations per core



Oracle Cloud Public HPC (Engineering and R&D) Customers



SUBARU®



TOYOTA



mazda



DENSO





Only a few elite mechanics can make this car run at peak performance. That's who Montis is for you and your customers with HPC workloads.



Software

install and configure Linux
Set up provisioning system: (Warewulf, SMC, Bright)
Set up workload managers (PBS, SLURM, & others) for end-to-end job submission
Set up monitoring of all system resources
Maintain CSP SAAS offering and/or custom software stack

Configure and Maintain Clusters

Install applications
Update software
Make resources available for end users
Configure, manage, optimize, and have proactive support to resolve administrative issues
Special projects as needed (IT, user, stack, etc.)

Monitoring

Provide relevant reports and tracking to ensure system capabilities are functional as desired
System utilization
Storage metrics
Testing and tuning
Bottleneck identification
Application benchmarking
Memory and interconnect performance

Additional Custom Services

US-based standard support beyond normal business hours
International support during normal business hours
Various turnaround times (4, 12, 24 hours)
L1 support (L2/L3 support is standard)
Weekend, 24x5, and 24x7 support
Various service levels depending on business needs and the number of users and applications
Special projects beyond adding users, installing software, debugging system problems, etc.

System Admin Tasks and User Support

Adding/disabling users
Troubleshooting user issues
Resolving network and connectivity issues, etc.
Improvements on end-user workflows
Research failed jobs for system failures
Maintaining custom hooks/job submission scripts for administrative use and end user support
Upgrading system software/firmware as requested in maintenance windows
Working with third party vendors to support engineers' work and maintain system availability/performance
Configuring workload manager scheduling policies
Installing patches to software (Operating system or applications)
License server administration
Can meet regularly with management for status reporting, system & software change planning, etc.
Storage:

- Monitor space usage and enforce policy and quotas
- Manage dedicated devices including firmware/software update

Linux HPC system experts from 5 hours per week on up with monthly payments and annual true-up

Client Hardware and Cloud Services

Maintain, install and or validate physical hardware (including GPUs)
Help with POCs or benchmarks
Auto-scaling
Perfect cloud images
Failover
Backup
Data recovery
Setup and configuration of storage
Adjust workload management for maximum efficiency and cost performance (including AI/ML, LLM, CFD, etc.)
Firewalls and license servers
Configure networking (IP, switches)
Optimize high speed network (InfiniBand, Omnipath, 100g ethernet)
Tune bios for HPC workloads

Meet the Montis Leaders

Your **Trusted** HPC Experts



Nate Fuhriman
Founder, CTO
24 years in HPC
Has fitted/put in production
the 3rd largest system in the
world on down

nate@montis.tech



Victor Wright
Founder, CEO
34 years in IT, 12 in HPC
People-first mentality
Honesty and trust are
deep in company DNA

victor@montis.tech

+1 801-360-1564



Mark Wardrop
Director of Support
30+ years in HPC
Applications expert
HPC analyst 2003-2022 on
ConocoPhillips Seismic and
Reservoir HPC team

mark@montis.tech



Devin Jensen
Business partner
30+ years in IT, 24 in HPC
Champions customers
Listener

devin@montis.tech

+1 801-653-2300



We create long-term relationships through a joint discovery process

- Listen carefully to understand what your customers or prospects believe their issues are
- Suggest possible solutions that we review and agree on together
- Identify salient action items with follow up reporting discussions
- Rinse and repeat as often as needed

Key operating tenets:

- Immediately support any issue deemed critical by client
- Honestly admit when we don't know the answer or results disappoint
- Say what we'll do and do what we said (return and report)
- Selfless with best practices/happy to train
- Under promise and over deliver
- Treat others like we should be treated