



**Hewlett Packard  
Enterprise**

# **HPC and AI in hybrid cloud environments with HPE GreenLake**

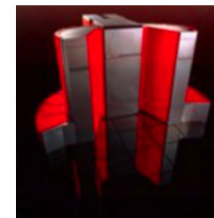
**César Gómez – WW HPC Solutions  
Architect**

Cáceres (Spain), September 14, 2022

**intel**<sup>®</sup>

## Looking at the Gordon Bell Prize

- 1 GFlop/s; 1988; Cray Y-MP; 8 Processors
  - Static finite element analysis
- 1 TFlop/s; 1998; Cray T3E; 1024 Processors
  - Modeling of metallic magnet atoms, using a variation of the locally self-consistent multiple scattering method.
- 1 PFlop/s; 2008; Cray XT5;  $1.5 \times 10^5$  Processors
  - Superconductive materials
- 1 EFlop/s; ~2018; ?;  $1 \times 10^7$  Processors ( $10^9$  threads)

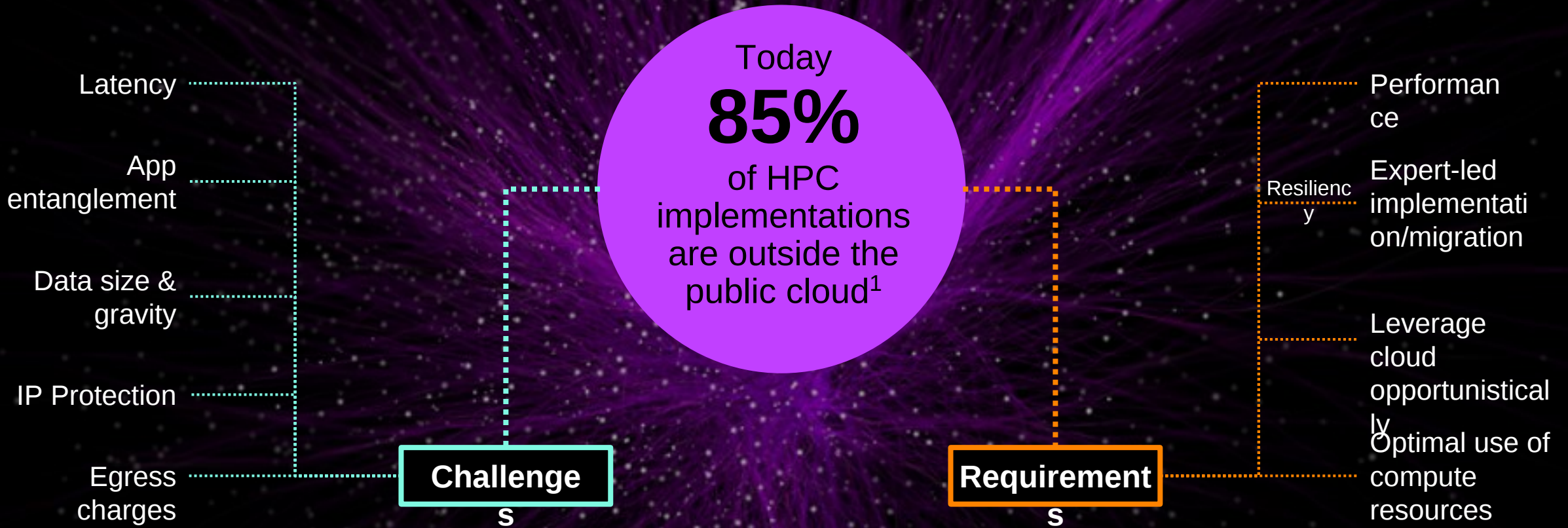


Jack Dongarra

Mérida, 27 Abril, 2010

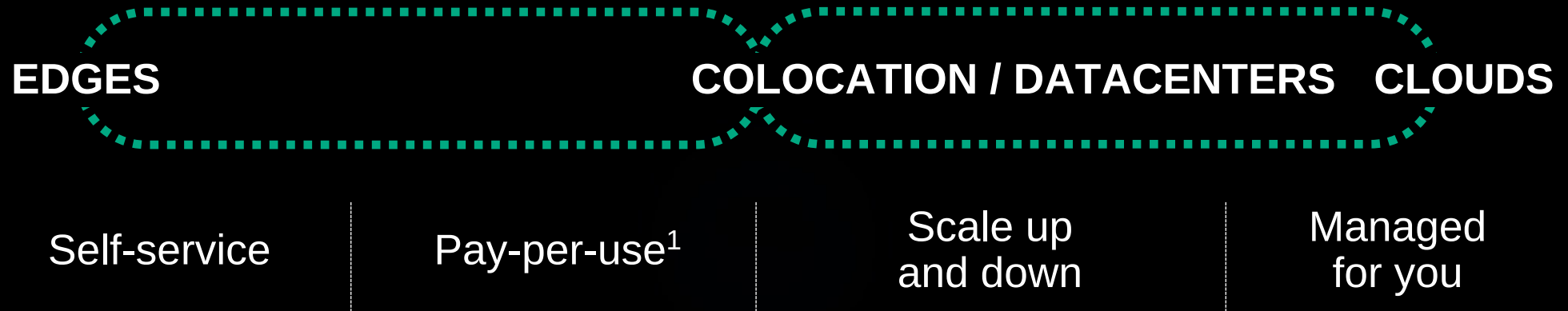
7

# CONSIDERATIONS FOR WORKLOAD PLACEMENT



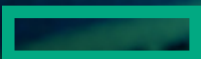
1 Source: Hyperion research 2022

# HPE GREENLAKE EDGE-TO-CLOUD PLATFORM



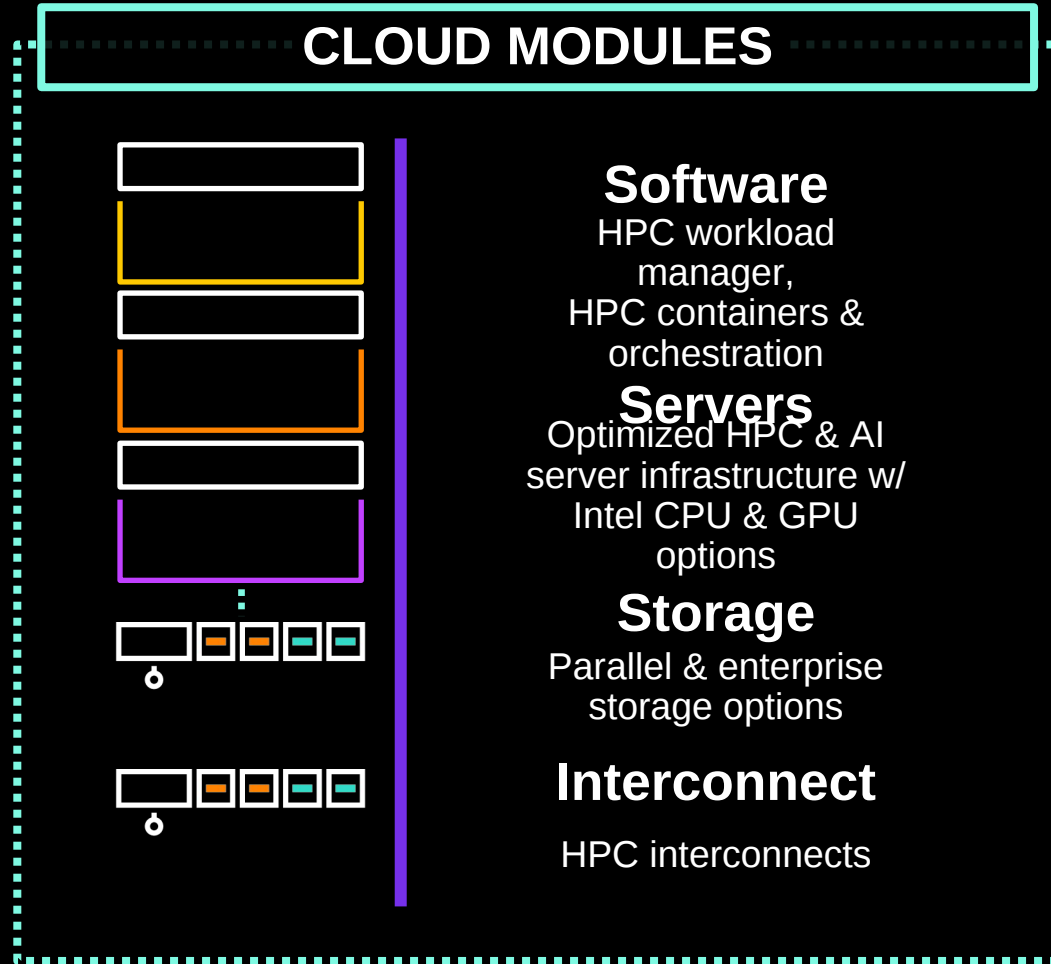
**THE CLOUD THAT COMES TO YOU**

<sup>1</sup> Reserve may apply



# HPE GREENLAKE FOR HPC

## Simplicity and speed with cloud modules



Mix and match compute and storage modules

Validated and optimized for performance

Supports multiple HPC/AI/ML workloads

# NEW HPE GREENLAKE FOR HIGH PERFORMANCE COMPUTING

- HPE Cray and Intel technologies delivering throughput unprecedented in a cloud service

Extensive HPC Partner ecosystem of value-added software and services integrate with HPE GreenLake for HPC by a partner program, with carefully design business interfaces

## HPE Apollo 2000/6500 Systems

More flexibility to open up AI, Machine Learning and more HPC techniques

## GPU Enhancement

Support for NVLINK, NVIDIA A100, A40, A30 in increments of 2-4-8 accelerators

## HPE Slingshot

Cray networking technology with extremely high speed, tunable Ethernet-based interconnection supercharging performance

## HPE Parallel File System Storage

Scalable, high performance storage system that can match the other components, delivering unprecedented throughput

## Lower entry point

Reduce risk of introducing HPC, test workloads with HPE GreenLake for HPC and scale as needed



# WHY IT MATTERS

---

- Hybrid Options

Flexible Hybrid models for customers, offer elasticity of their HPE GreenLake for HPC service

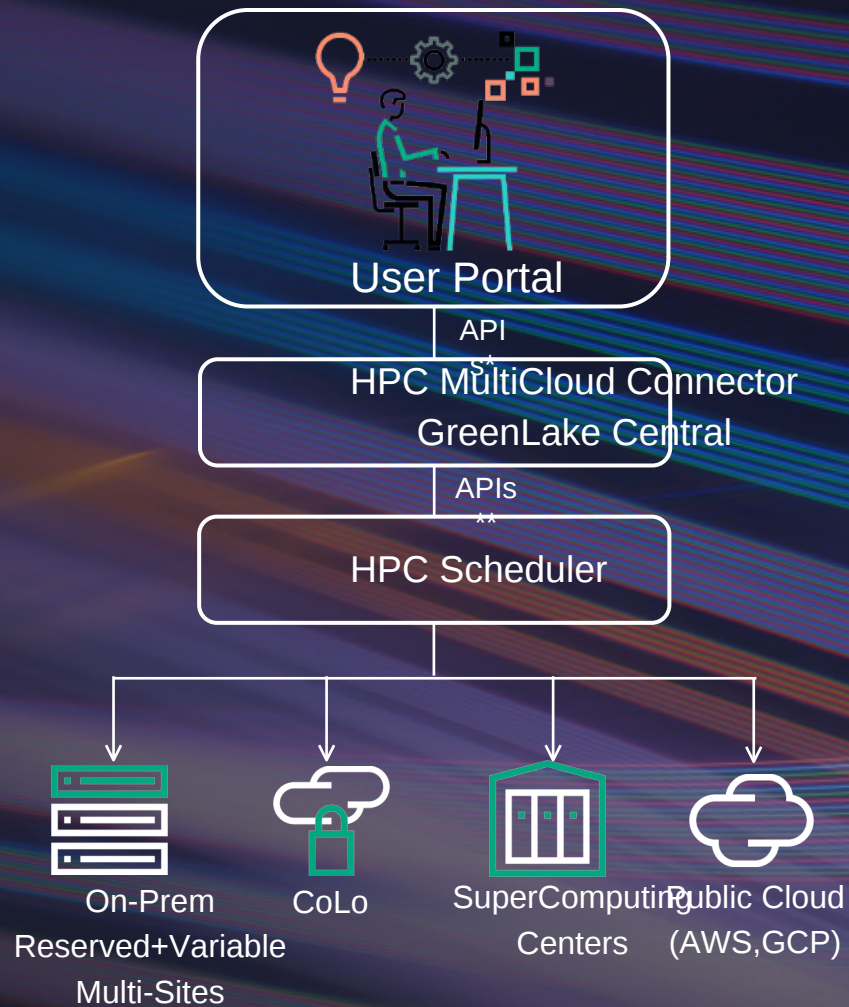
HPE GreenLake for HPC-to-private cloud based on HPE GreenLake for HPC or to-a public cloud

Multi-cloud connector APIs (Hybrid Cloud APIs) that we will publish and drive to become industry standard on how to program submitting HPC jobs to a diverse pool of computing

Ability to orchestrate data-center scale workflows with user-defined policies to determine best computing target where to execute a job

# Multi-Cloud Connector

Performing hybrid capability to submit jobs to an external cluster



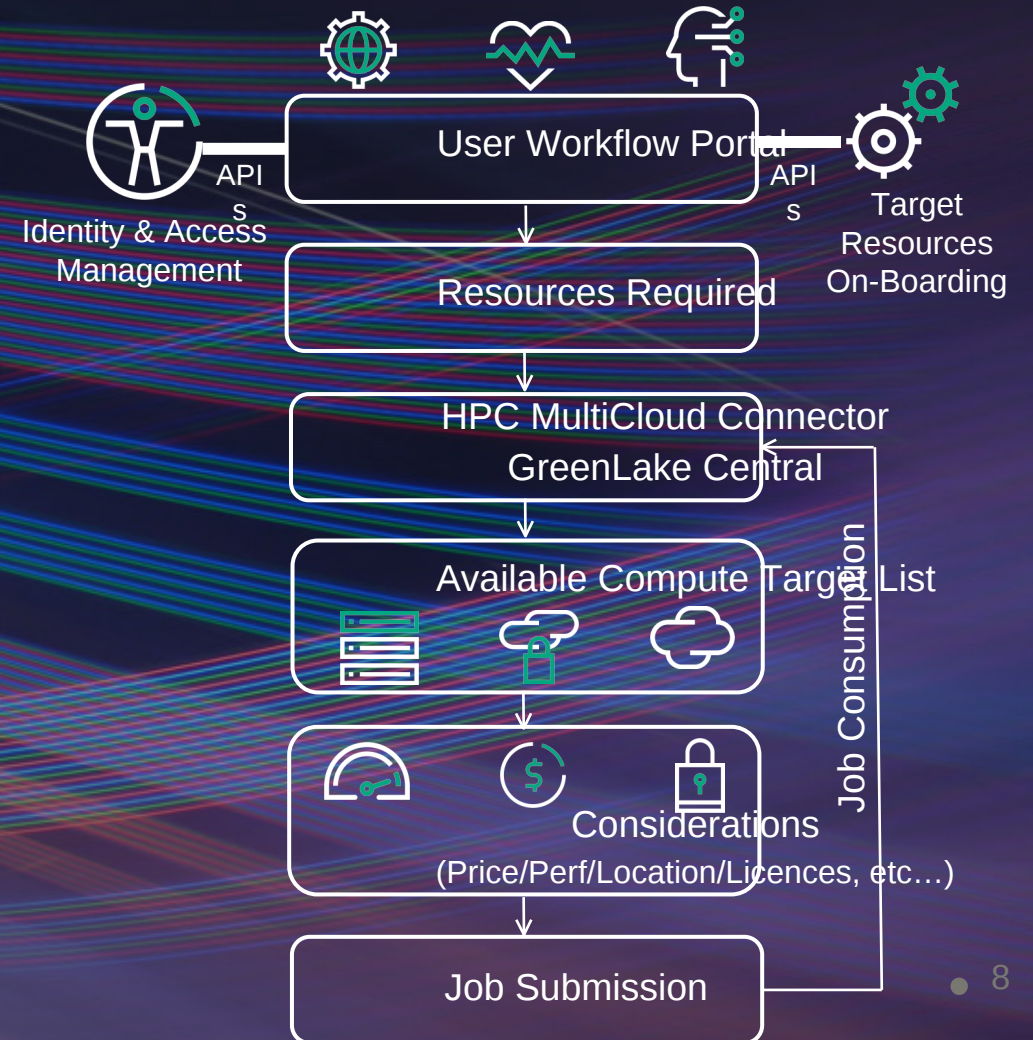
\* APIs: job submission, resource filtering

\*\* APIs: on-boarding, provisioning

# Workflow-aaS

Applied to different verticals:

- Manufacturing-aaS (d3View)
- HC and LS-aaS (3D structures of biomacromolecules)
- Large scale training-aaS (SmartSim/Determined.ai)





# WHAT IS UNDER THE HOOD?

<b>Management server</b> - GL Gateway 2.0 XL170r - GL Control Plane 2.0 (Keymaker)	<b>Management software</b> - HPCM/CMU - OneView (firmware)	<b>GreenLake Management Services</b> - IT Service Management, System Operations and Administration - Support Services - HPC-Specific Support Services (Data integration, VDI integration, etc.)
<b>Network</b> - Slingshot	<b>Storage</b> - Ezmeral Data Fabric - HPE Parallel File System Storage (IBM Spectrum Scale)	<b>Advisory &amp; Professional Services</b> - Container Adoption Services - Virtual Desktop Infrastructure (VDI) - Customer Portals - Storage and Data Management - Equinix, CyrusOne, Digital Realty - Trials (on-going)
<b>Storage</b> - Local NVMe - Hybrid DL320	<b>Compute</b> - CPU: Apollo 2000 - LH Dense Compute Modules - Multi-GPU module - Standard compute using DLs with Intel - GPU: Apollo 6500 - DL GPU support	<b>Co-location / Data Trials</b>
<b>HARDWARE</b>	<b>SOFTWARE</b>	<b>SERVICES</b>

### Flexible Mix & Match of Compute & Storage Configuration Examples

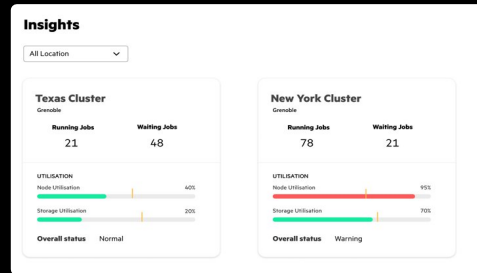
Compute			Storage		
Racks	Nodes	Cores	Racks	Storage nodes	Typical Total Usable Storage in GB
1	16	1024	1 to 2	5	508'000
2	48	3072	2 to 4	8	687'000
3	80	5120	4 to 6	16	1'685'000
4	112	7168	6 to 8	28	2'976'000
5	144	9216			
6	176	11264			
7	208	13312			
8	240	15360			
			NEW GPU		
			Racks	# of GPUs supported	
			1	2-40	
			2	42-80	
			3	82-120	

<b>GreenLake Standard Mandatory Managed &amp; Operated by HPE GreenLake Management Services</b> Complete Care A&PS HPC Cluster Management Service Network integration services	<b>GreenLake Standard Service Options</b> Terms: 3-4-5 years Variable Usage: Metered on core/hour GPU hours and GB services: A&PS Consulting	<b>Hybrid Model Options</b> Multi-Site Shared Buffer CoLo / Private or Specialized CSP (Pilot) Public CSP
<b>Sales Motions</b>		
HPE HPC Sales	GreenLake Sales	HPE Account Sales
		HPE HPC Partner Pilot release

# AN END-TO-END HPC AND AI EXPERIENCE

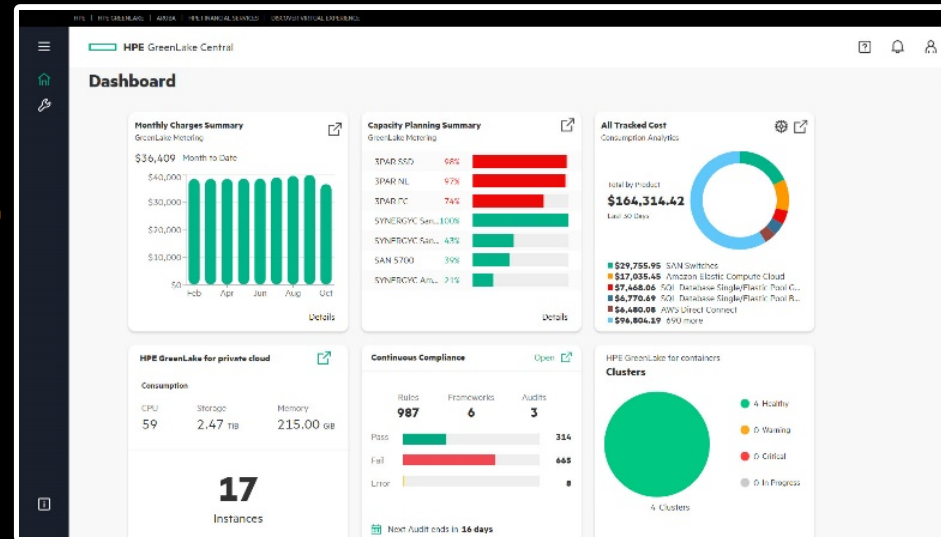
## HPE GreenLake platform changes the experience

### Configure the cluster



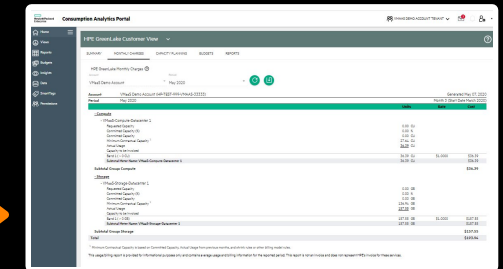
HPC ADMIN

### HPE GreenLake Central

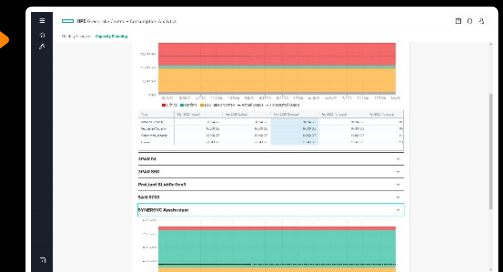


CIO/LOB

### View costs and usage

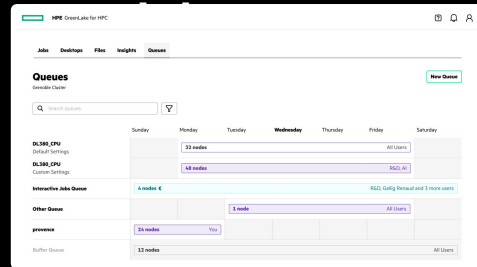


### Forecast capacity



HPC ADMIN

### Launch & monitor HPC

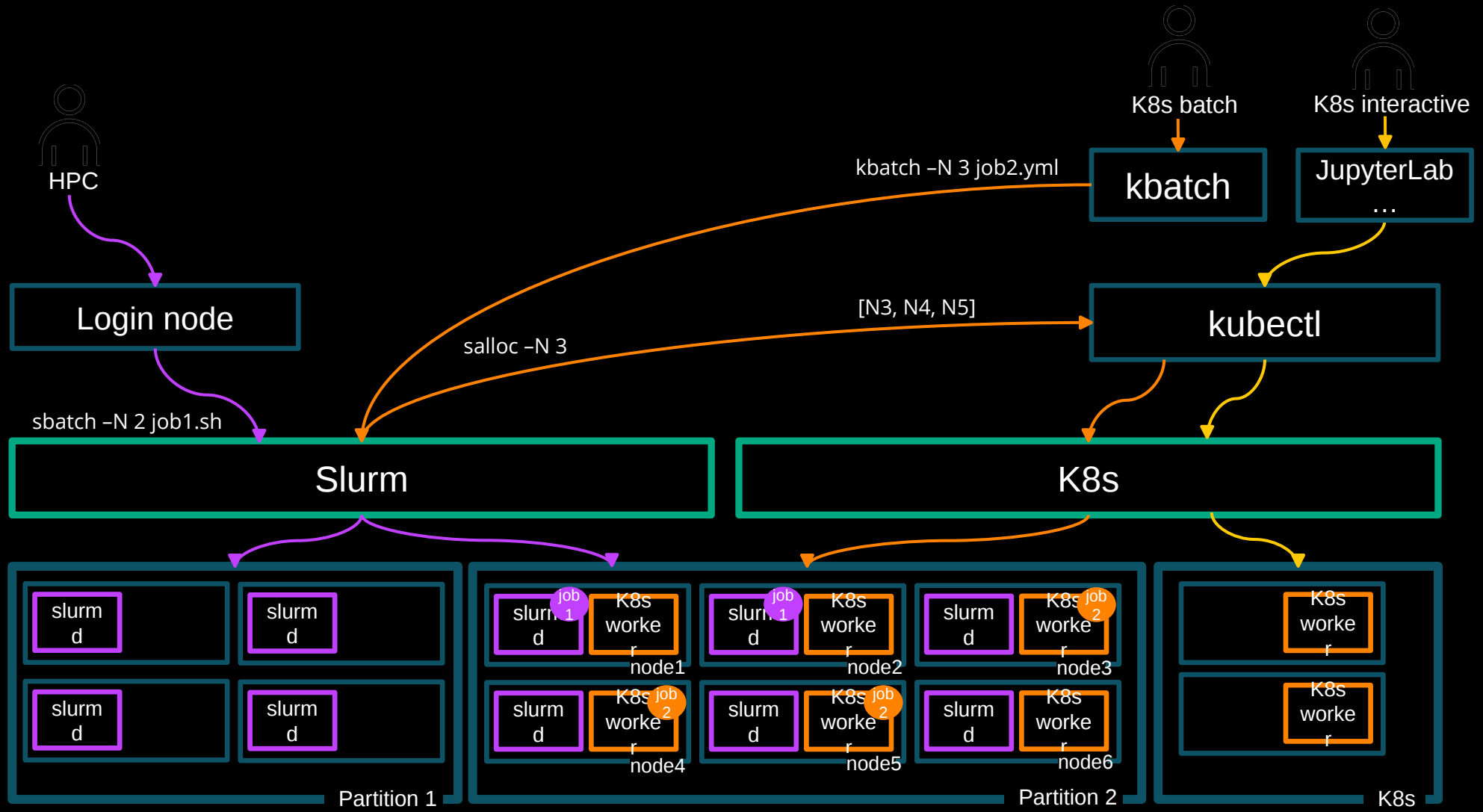


Meter costs and usage

# WHY IS CONVERGENCE OF HPC, AI, AND CLOUD IMPORTANT?

- To take advantage of cloud technologies for HPC/AI-native infrastructure
- To use large scale infrastructure for HPC and AI workloads
- To provide on-demand computing capacity without changing the software stack
- To have tailored or highly configurable software and workload deployments
- To abstract the complexity of running the same application in different HW and SW infrastructures
- To have the ability to run user-defined software stacks using containers

# CONVERGED SLURM/KUBERNETES CLUSTER





# VISION: BASIC TO EXASCALE

	MVP	Expand	Exascale
HPE GreenLake Services capabilities	<ul style="list-style-type: none"> <li>• HPC GreenLake Central integration</li> <li>• Cost analytics</li> <li>• Self Service interface Manage clusters and batch queue</li> <li>• Optional A&amp;PS Services available</li> </ul>	<ul style="list-style-type: none"> <li>• End user Integration to HPE GreenLake Central</li> <li>• Metering enhancements including show back</li> <li>• Simple quoting tool integration</li> <li>• Public, private cloud integration</li> </ul>	<ul style="list-style-type: none"> <li>• HPC Customer portal integrations</li> <li>• Specific ISV platforms</li> <li>• New UoMs for HPC clusters. More granularity (core, memory, IO, bandwidth, latency and capability-based) UoM</li> <li>• SLA/QoS-based billing</li> </ul>
	Initial platform	Expanded HPC reach	Exascale platform
Platform releases	<ul style="list-style-type: none"> <li>• HPE Apollo</li> <li>• Standard storage</li> <li>• Aruba interconnect</li> </ul>	<p>Cloud modules</p> <ul style="list-style-type: none"> <li>• HPE Parallel file system storage, Slingshot and GPUs</li> <li>• HPE Ezmeral Container Platform— integrated support for Slurm and Singularity</li> <li>• MLOps, AI integrated</li> </ul>	<ul style="list-style-type: none"> <li>• Cray compute, storage options</li> <li>• Autoscaling of resources with any combination of standard and premium offerings</li> <li>• Flexible storage layers (HPE Ezmeral DF, HPE Parallel filesystem, Intel DAOS)</li> </ul>
Key use cases	<ul style="list-style-type: none"> <li>• Ansys/CAE</li> <li>• Others opportunistically</li> </ul>	<ul style="list-style-type: none"> <li>• Visualization use cases; workflow</li> <li>• Additional workload focus EDA, CFD, FSI, Biomedical, molecular dynamics</li> </ul>	<ul style="list-style-type: none"> <li>• Additional workload focus Seismic, Weather forecast, high content screening, FSI risk management</li> </ul>

# Thank you

---

César Gómez – [cesar.gomez@hpe.com](mailto:cesar.gomez@hpe.com)