

HPC and AI in hybrid cloud environments with HPE GreenLake

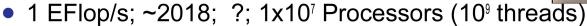
César Gómez – WW HPC Solutions Architect

Cáceres (Spain), September 14, 2022



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; 1.5x10⁵ Processors
 - Superconductive materials





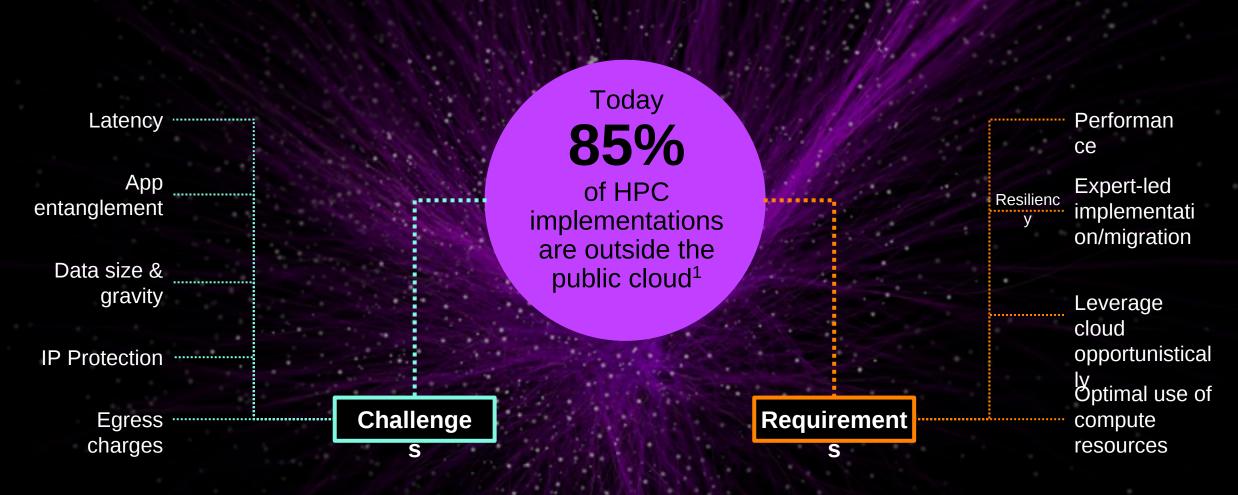








CONSIDERATIONS FOR WORKLOAD PLACEMENT

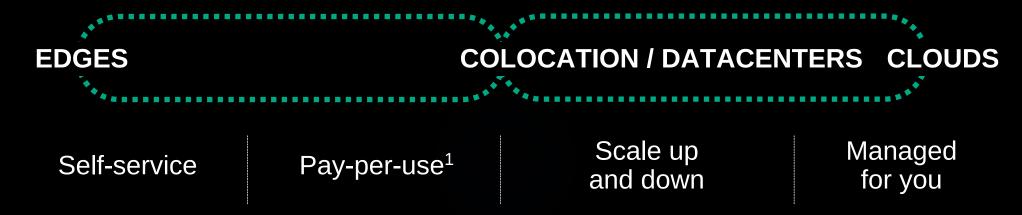


1 Source: Hyperion research 2022



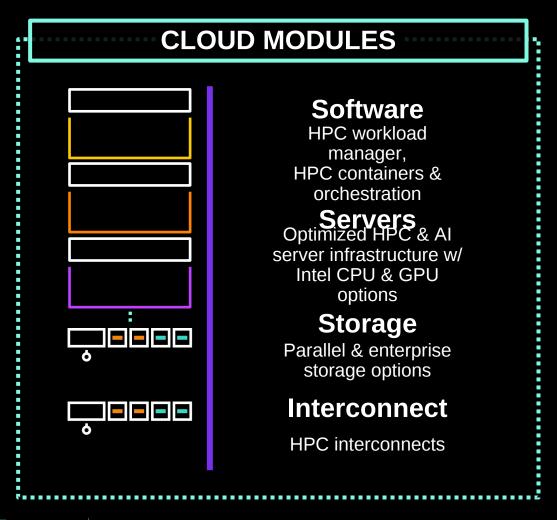


HPE GREENLAKE EDGE-TO-CLOUD PLATFORM



THE CLOUD THAT COMES TO YOU

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

More flexibility to open up AI, Machine Learning and **HPE Apollo 2000/6500** more HPC techniques **Systems** Support for NVLINK, NVIDIA A100, A40, A30 in **GPU** increments of **Enhancement** 2-4-8 accelerators Cray networking technology with extremely high speed, **HPE Slingshot** 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

Reduce risk of introducing HPC, test workloads with Lower entry point 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

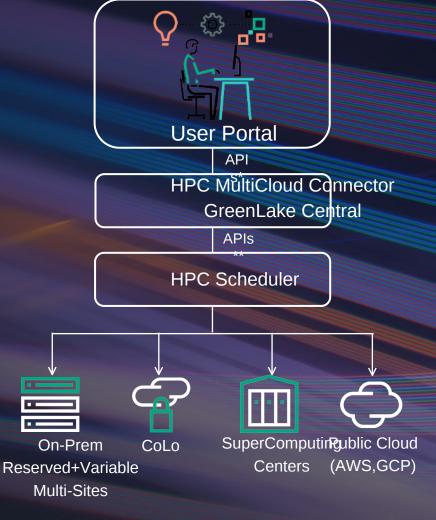
Multi-cloud connector APIs (Hybrid Cloud APIs) that we will publish and drive to become industry standard on how to program submitting HPC jobs to a diverse pool of computing

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



Multi-Cloud Connector

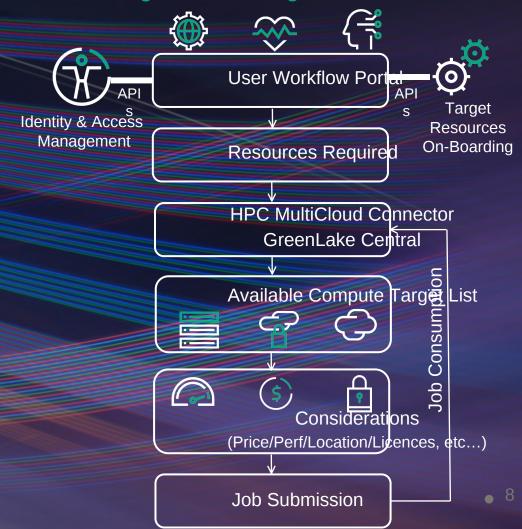
Performing hybrid capability to submit jobs to an external cluster



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?

Management server - GL Gateway 2.0 XL170r - GL Control NEW Plane 2.0 (Keymaker) Network - Slingshot Storage - Local NVMe Hybrid DL320 Compute - CPU: Apollo 2000 -LH Dense Compute Modules -Multi-GPU module

-Standard compute

-GPU: Apollo 6500

-DL GPU support

HARDWAR

using DLs with

Management software

- HPCM/CMU - OneView (firmware)
 - **Storage**
- Ezmeral Data **Fabric** HPE Parallel File System Storage (IBM Spectrum Scale)

Compute

- Singularity/Podma
 - HPCM/CMU
- SLURM
- SLES 15SP2

SOFTWAR Е

GreenLake Management Services

- IT Service Management, System Operations and Administration
- Support Services
- HPC-Specific **Support Services** (Data integration, VDI integration,
- $_{
 m e}$ Professional Services
- Container Adoption Services
- Virtual Desktop Infrastructure (VDI)
- Customer

Perta Socation /

trials - Equinix.

CyrusOne, Digital NEW Realty

SERVICES

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

Compute			Storage			
Racks	Nodes	Cores	Racks		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	NEW	GPU	GPUs	
6	176	11264	Racks	#	# of GPUs	
7	208	13312	Rack	SI SI	supported	
8	240	15360	1		2-40	
			2		42-80	
			3		82-120	

GreenLak eStandar d **Mandator**

Managed & Operated by HPE GreenLake

Management Services Complete

A&PS HPC Cluster Management

Service

Care

Network integration services

GreenLak Standard Service **Options**

Terms: 3-4-5 years

> Variable Usage:

core/hour **GPU** hours epoli GFBI services: A&PS

Metered on

Public CSP

Hybrid

Model

Options

Multi-Site

Shared

Private or

Specialized

CSP (Pilot)

Suffer CoLo /

Sales Motions

Consulting

HPE	GreenLa	HPE
HPC	ke	Account
Sales	Sales	Sales

Е intel.

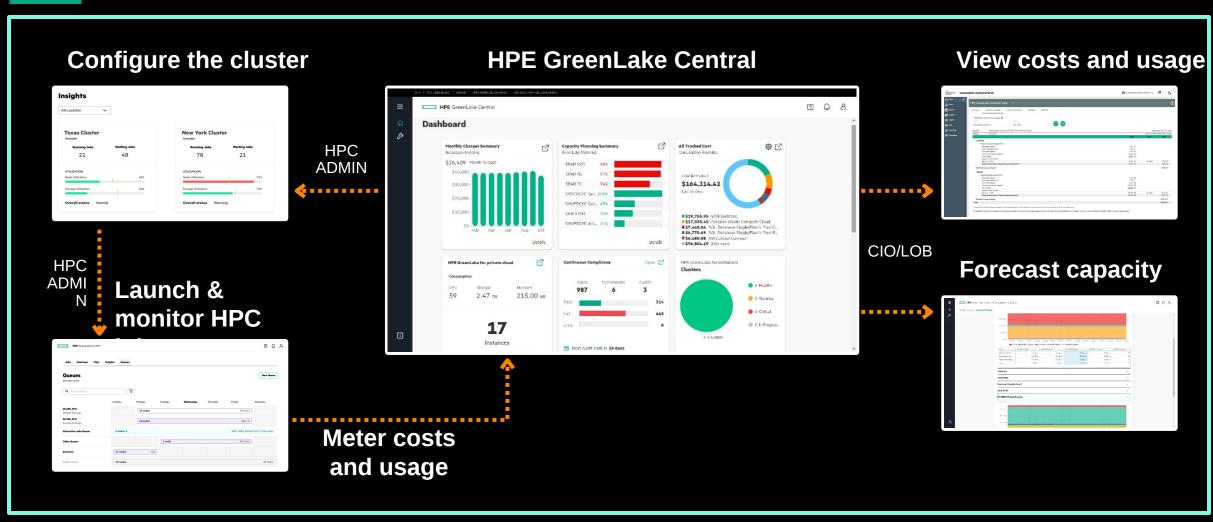
Hewlett Packard Enterprise

• 9

HPC Partner

Pilot release

AN END-TO-END HPC AND AI EXPERIENCE HPE GreenLake platform changes the experience



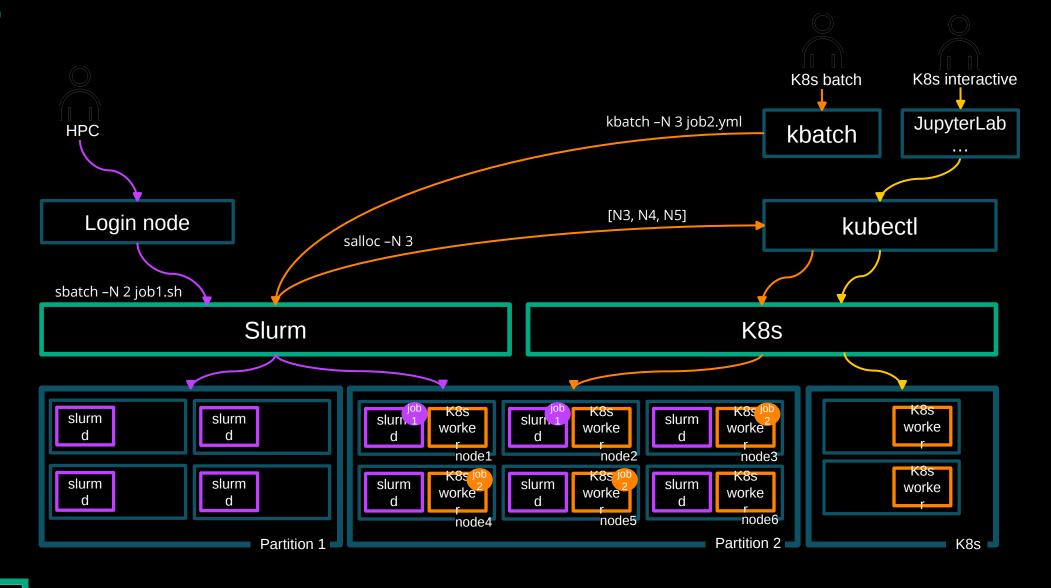




WHY IS CONVERGENCE OF HPC, AI, AND CLOUD IMPORTANT?

- 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 <u>highly configurable software</u> and workload <u>deployments</u>
- To <u>abstract the complexity</u> of running the same application in <u>different HW and SW</u> infrastructures
- To have the ability to run <u>user-defined software stacks using containers</u>

CONVERGED SLURM/KUBERNETES CLUSTER



VISION: BASIC TO EXASCALE

HPE GreenLake Services capabilities

MVP

Expand

Exascale

- HPC GreenLake Central integration
- Cost analytics

ollogA 39H

Standard storage

Aruba interconnect

- Self Service interface Manage clusters and batch queue
- Optional A&PS Services available
- End user Integration to HPE GreenLake Central
- Metering enhancements including show back
- Simple quoting tool integration
- Public, private cloud integration

- HPC Customer portal integrations
- Specific ISV platforms
- New UoMs for HPC clusters.
 More granularity (core, memory, IO, bandwidth, latency and capability-based) UoM
- SLA/QoS-based billing

Platform releases

Initial platform

Cloud modules

 HPE Parallel file system storage, Slingshot and GPUs

Expanded HPC reach

- HPE Ezmeral Container Platform— integrated support for Slurm and Singularity
- MLOps, Al integrated

Exascale platform

- Cray compute, storage options
 Autopooling of recourses with a
- Autoscaling of resources with any combination of standard and premium offerings
- Flexible storage layers (HPE Ezmeral DF, HPE Parallel filesystem, Intel DAOS)

Key use cases

- Ansys/CAE
- Others opportunistically

- Visualization use cases; workflow
- Additional workload focus EDA, CFD, FSI, Biomedical, molecular dynamics
- Additional workload focus
 Seismic, Weather forecast, high
 content screening, FSI risk
 management





Thank you

César Gómez – cesar.gomez@hpe.com



