
AWS
re:Invent
NOV. 28 – DEC. 2, 2022 | LAS VEGAS, NV

CMP222

Redefining supercomputing on AWS

Dr. Neil Ashton

Principal Computational Engineering SA
AWS

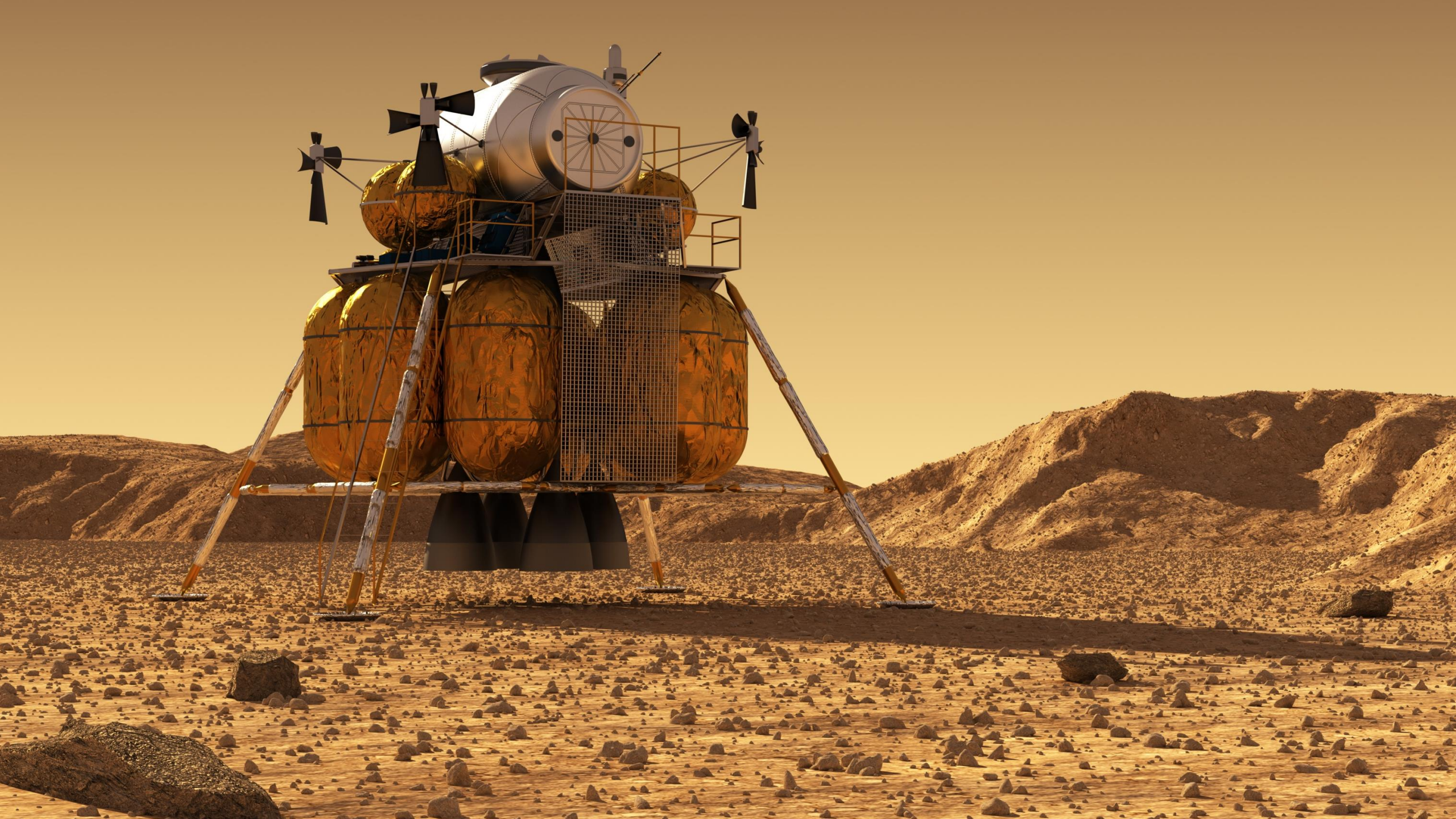
Heidi Poxon

Principal HPC Technologist
AWS

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.









Navier-Stokes Equations (1822)

Claude-Louis Navier & George Gabriel Stokes

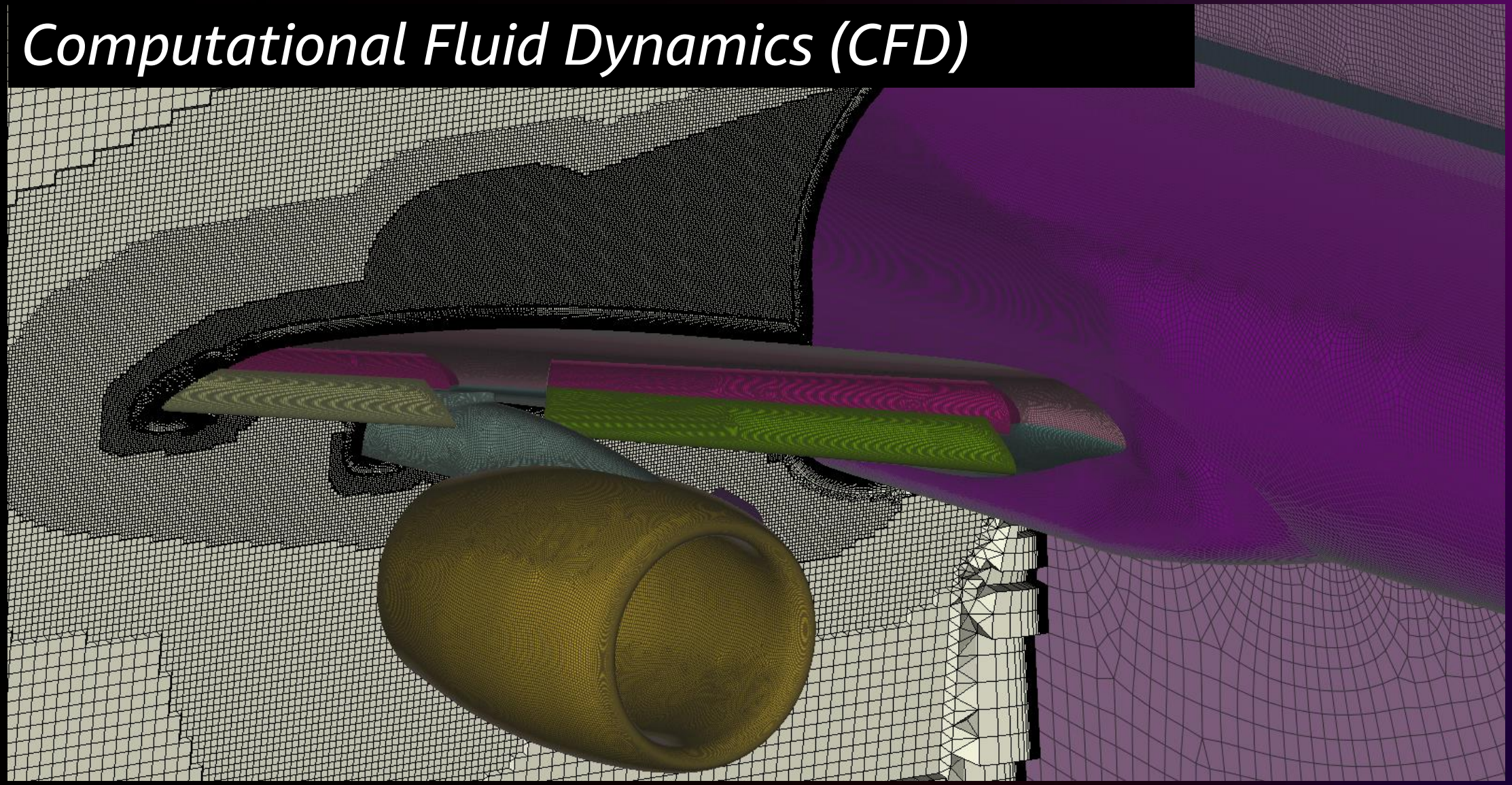
$$\frac{\partial u_i}{\partial x_i} = 0$$

$$\frac{\partial u_i}{\partial t} + u_j \frac{\partial u_i}{\partial x_j} = -\frac{1}{\rho}\frac{\partial p}{\partial x_i} + \frac{\partial \tau_{ij}}{\partial x_j}$$

aws

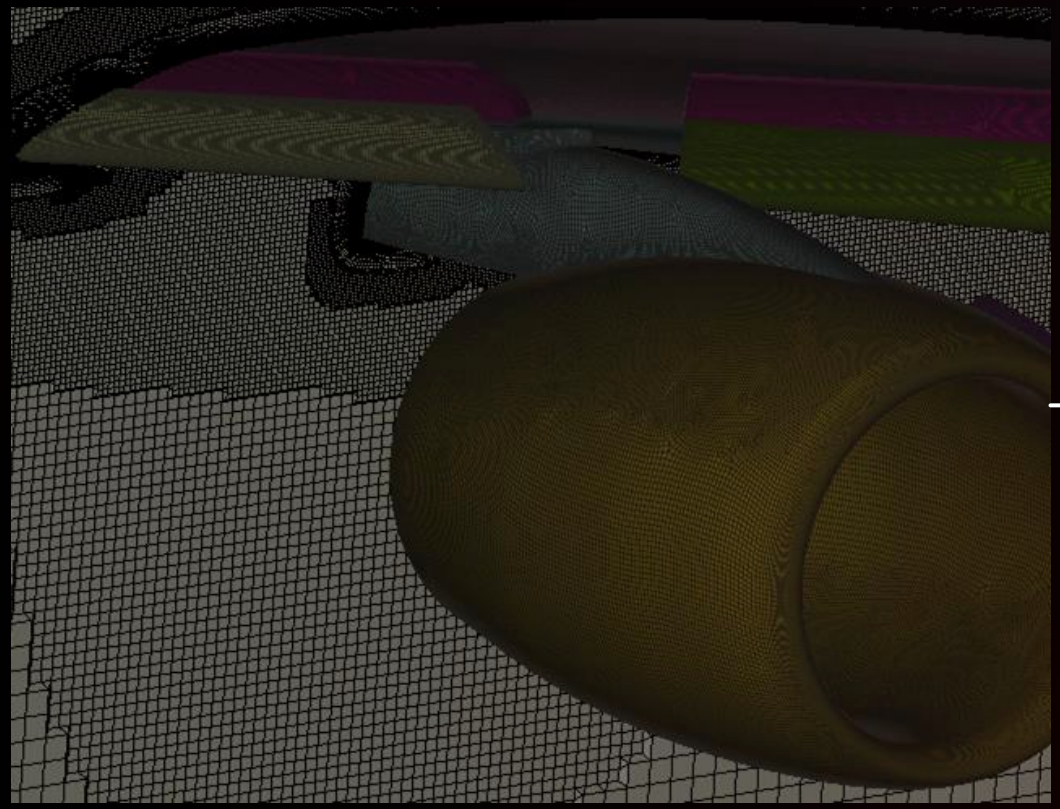
© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

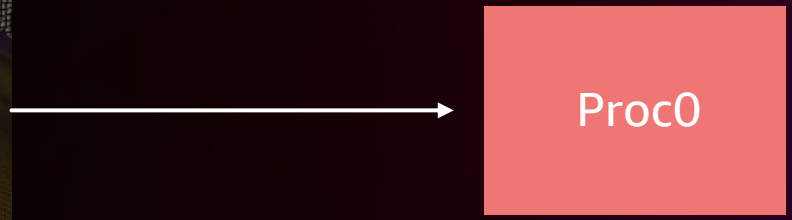
Computational Fluid Dynamics (CFD)



© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

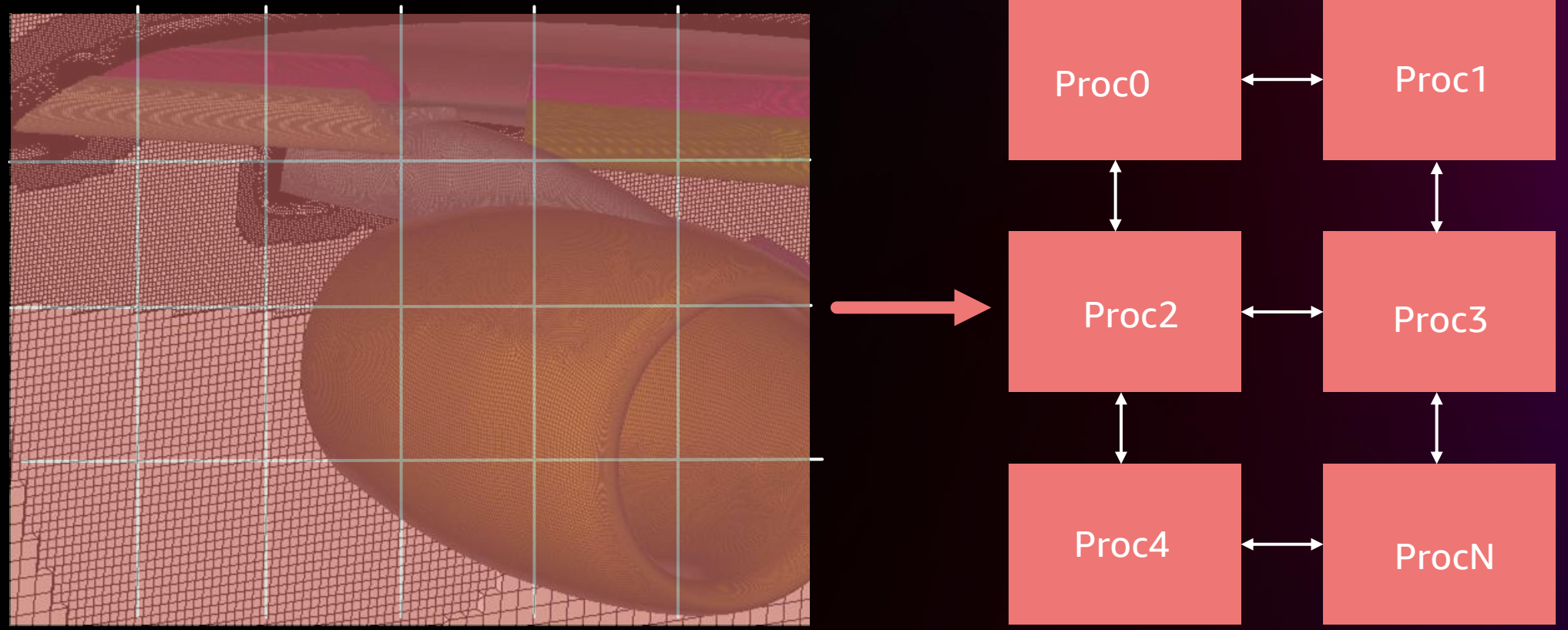
Computing requirements





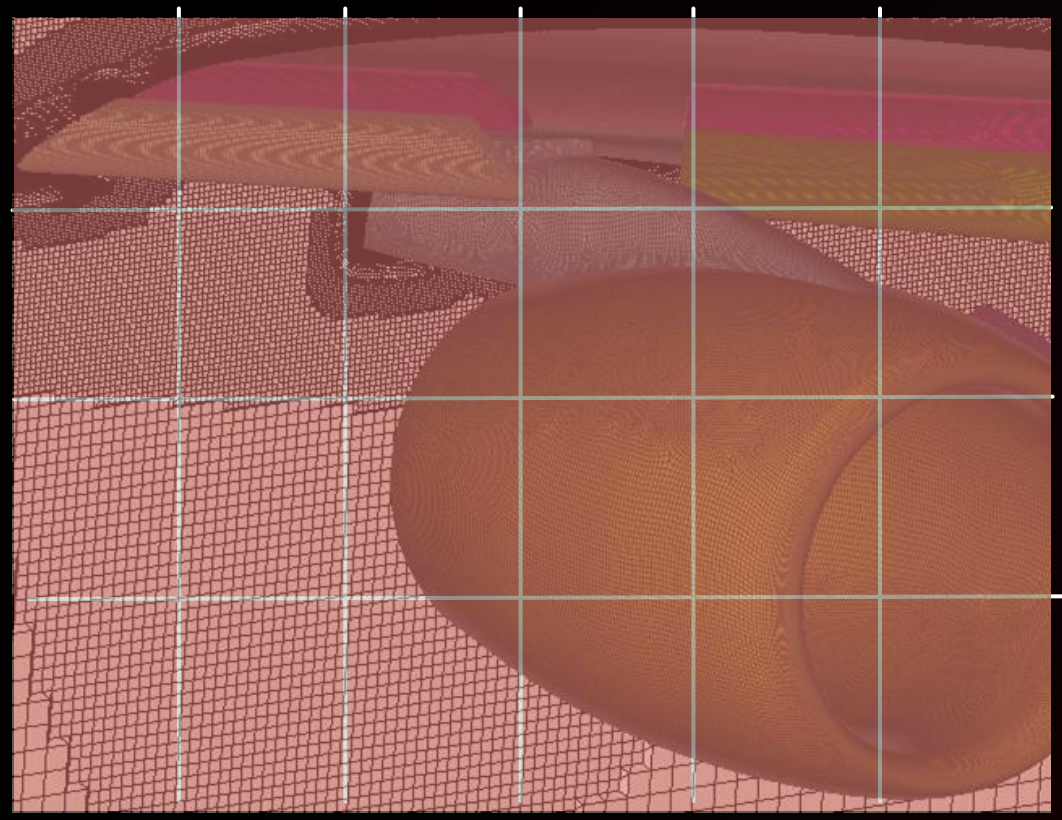
© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

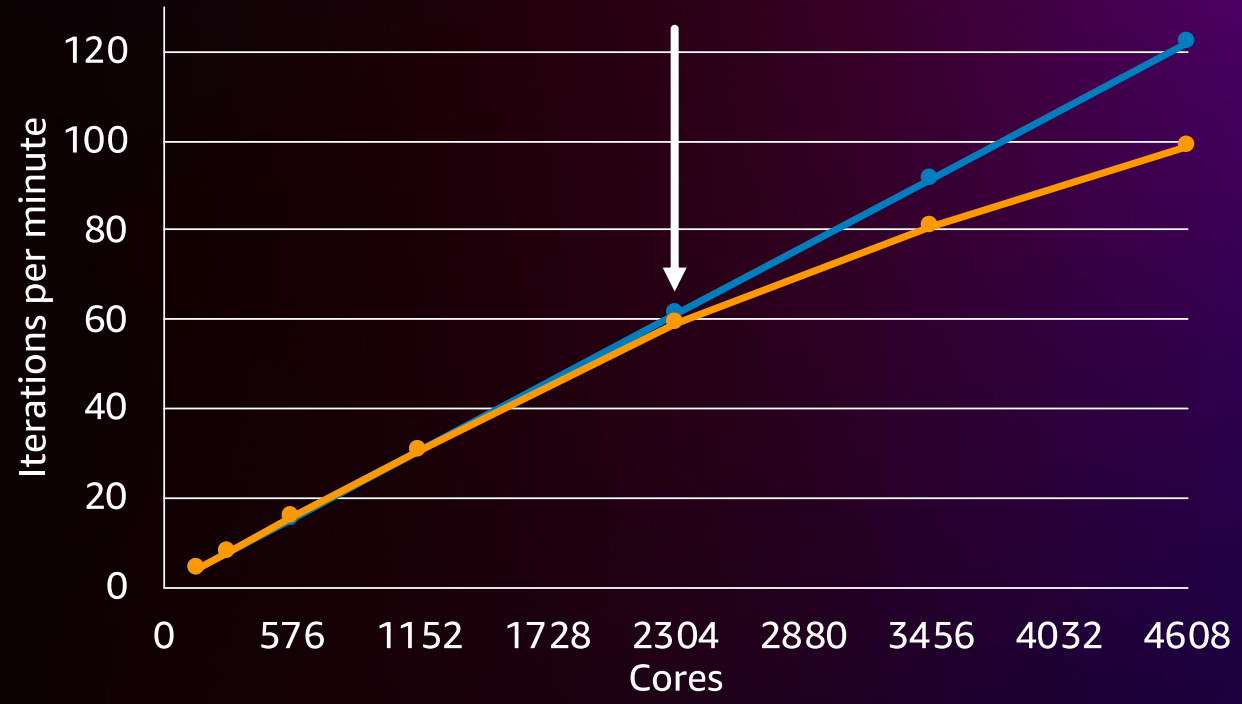
Large-scale tightly-coupled simulation



© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Large-scale tightly-coupled simulation





aws
© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.


What is a supercomputer?


Where is supercomputing technology today?

The convergence of cloud + HPC + AI/ML

Shift towards workflows, ensembles, broader integration

HPC workloads

HPC/cloud convergence

HPC+ML/AI/QC/…

Tightly integrated massive scale computing

Shift towards tightly coupled, high-performance capabilities

Cloud technologies and services

Tightly coupled/Isolated

Loosely coupled/Tightly integrated



CAE
CFD



Seismic reservoir



Financial services risk actuarial



EDA



Drug discovery
Genomics
Cryo-EM



Weather
Climate



Autonomous compute
Digital twins



Quantum computing



Autonomous vehicles



ML frameworks

aws

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Our guiding principles and product vision

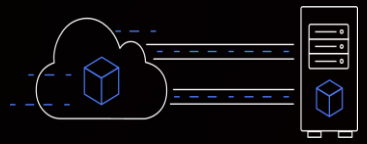
Experience and ease of use



***Think* anyone**

- Reduce friction
- Simplify deployment
- Track, monitor, and manage
- Improve functionality and accessibility

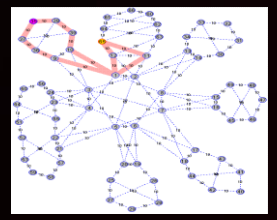
Migration effectiveness



***Think* "One HPC"**

- Harmonize on-prem/cloud
- Compute, data, orchestration and scheduling, performance…

Elasticity and extensibility



***Think* anytime**

- Flexible configuration
- Elastic capacity and scalability
- Grow and shrink as your HPC workloads dictate
- Pay for what you use

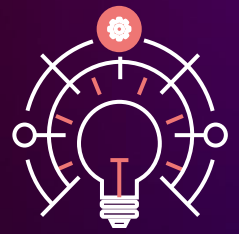
Choice and optionality



***Think* workflows**

- Heterogeneity for optimization
- Support traditional to cloud-centric
- Amplify through adjacent services

Partnerships and co-innovation



***Think* possibilities**

- Pushing boundaries
- New workloads
- Optimization and enablement

aws

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

AWS building blocks for HPC

Versatile compute

Amazon EC2
Compute configurations for every workload

- Processor architectures (AWS, Intel, AMD)
- Processor frequency (up to 4.5 GHz)
- Memory footprint (up to 24 TiB)
- Instance storage (HDD and SSD)
- Accelerated computing (ASIC, GPUs, and FPGA)

More than 600 instance types

Performant and scalable

AWS Nitro System
- ✓ Bare-metal performance
- ✓ OS bypass
- ✓ I/O offload and acceleration
- ✓ Enhanced security

Elastic Fabric Adaptor
- ✓ Accelerated networking for HPC
- ✓ High-bandwidth (to 800 Gb/s)
- ✓ Low latency (< 15 us)
- ✓ OS bypass
- ✓ RDMA and GPUDirect
- ✓ Libfabric core for MPI and NCCL

Fast and flexible storage

Amazon FSx for Lustre
- ✓ HPC parallel filesystem
- ✓ Fully managed
- ✓ Scalable performance (+100 Gb/s)
- ✓ Millions of IOPs
- ✓ Concurrent access for 1000s of nodes
- ✓ Amazon S3 import/export

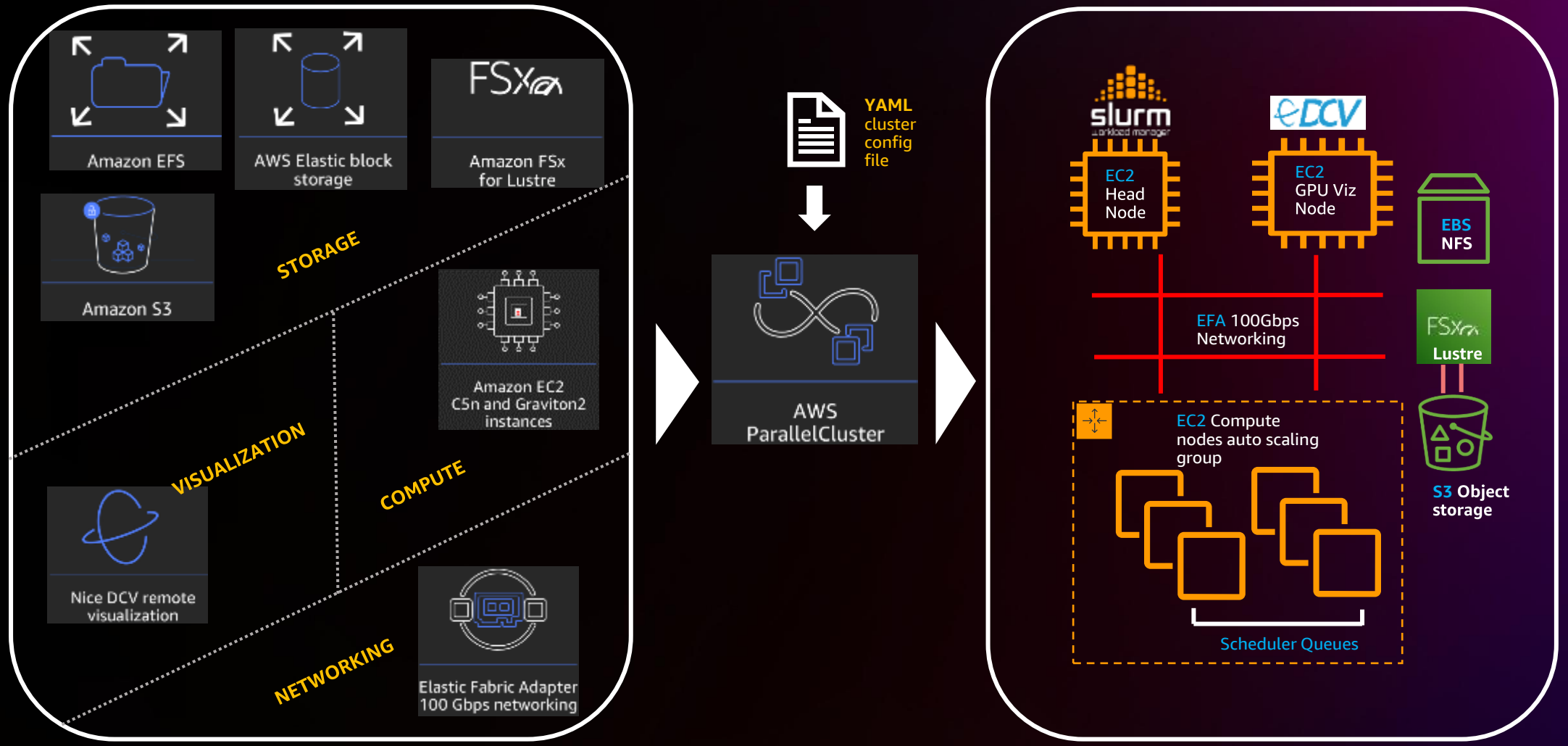
Amazon S3
Durable, low-cost object storage

Amazon EBS
Elastic, performant block storage

Amazon EFS
Peta-scale shared file storage

aws
© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

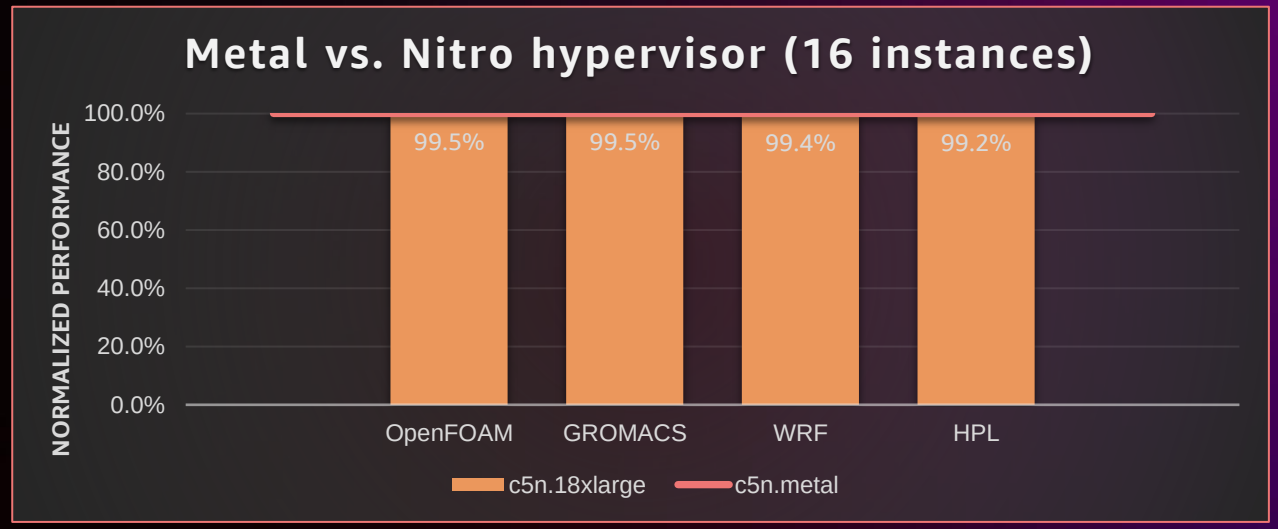
Composable infrastructure and services

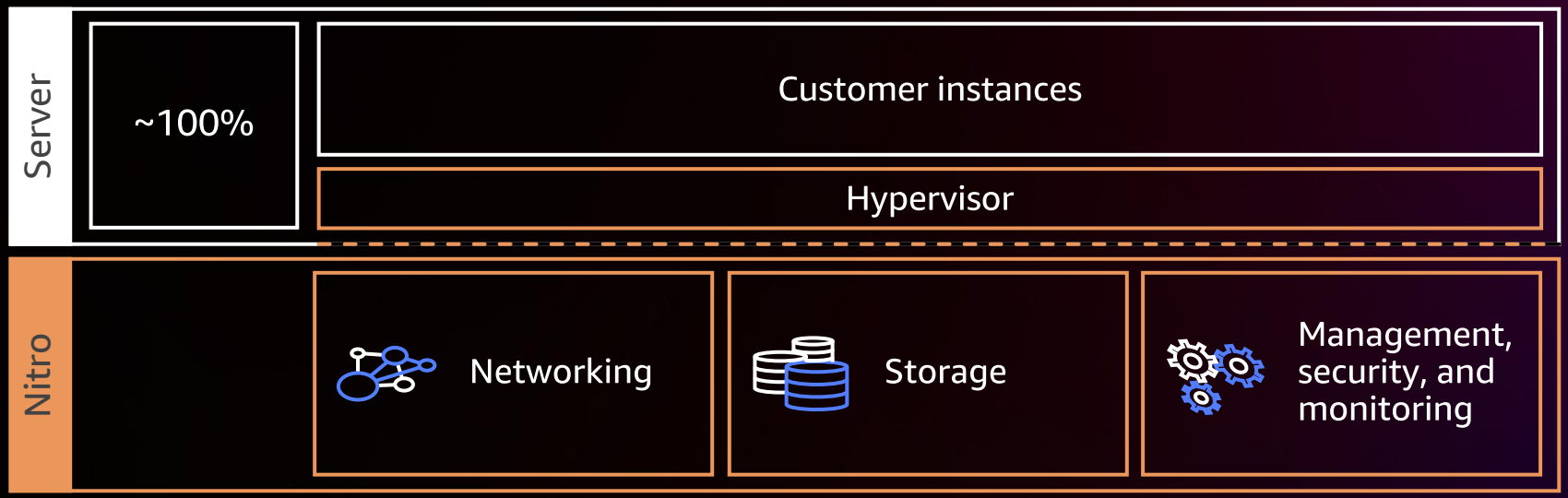


© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

The AWS Nitro System

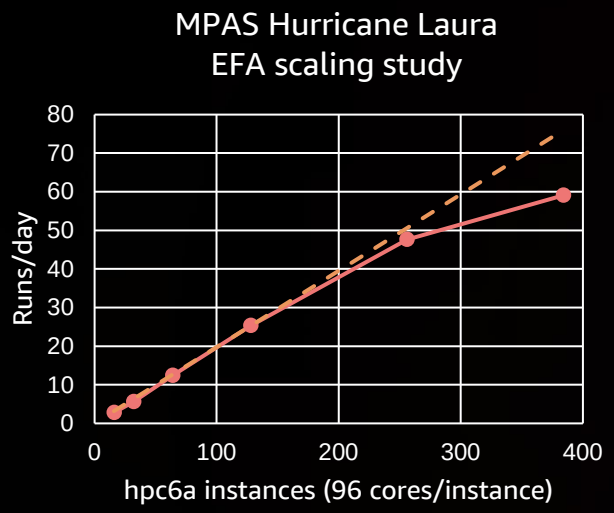
Engineered with a **hardware-based root of trust** using the Nitro Security Chip, allowing for the system to be continuously measured and validated





© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

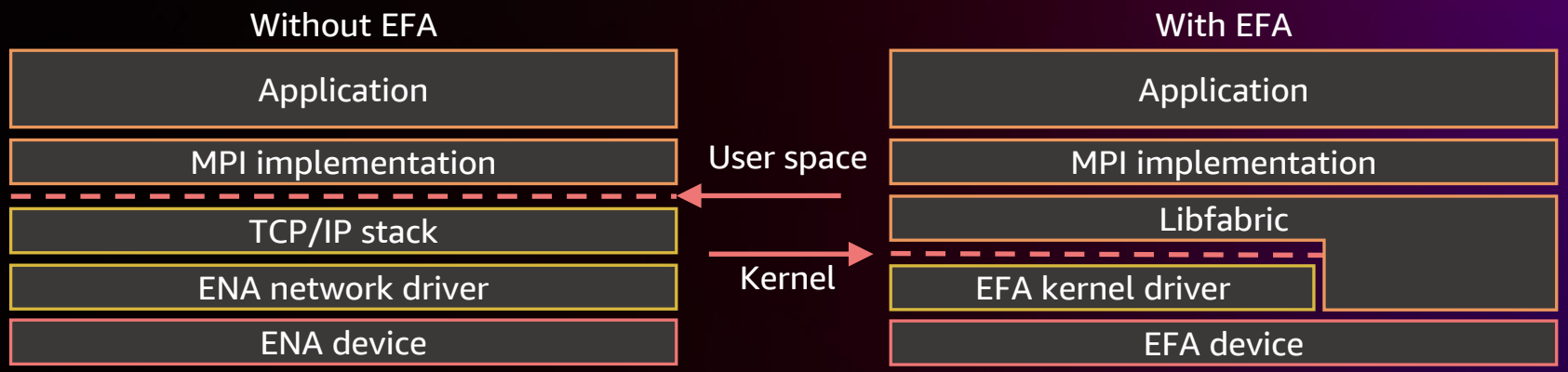
Elastic Fabric Adapter: Network built to scale



Scaling for tightly coupled workloads

- ✓ OS bypass
- ✓ GPUdirect and RDMA
- ✓ Libfabric core supports wide array of MPIs and NCCL

- ➢ High-speed/low-latency
- ➢ Cloud-scale congestion control
- ➢ Up to 800 Gb/s bandwidth

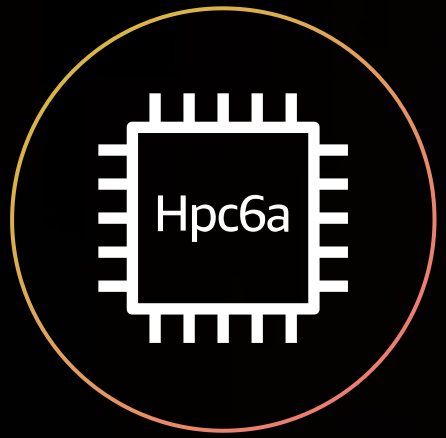


"We can run CFD workflow jobs 15 times faster on AWS thanks to the computing power and inter-node network performance we get using the Amazon EC2 C5n.18xlarge instances and EFA."
– Guillaume Martinat, Lead Aerodynamics Engineer, Flying Whales

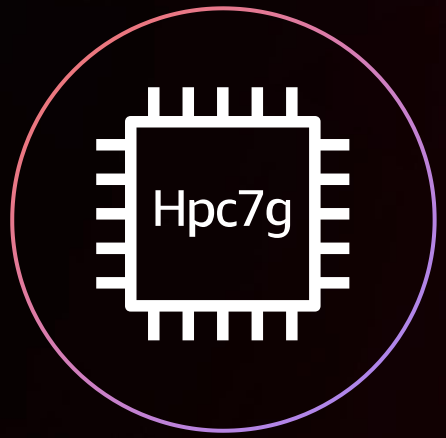
aws
© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

HPC optimized instances

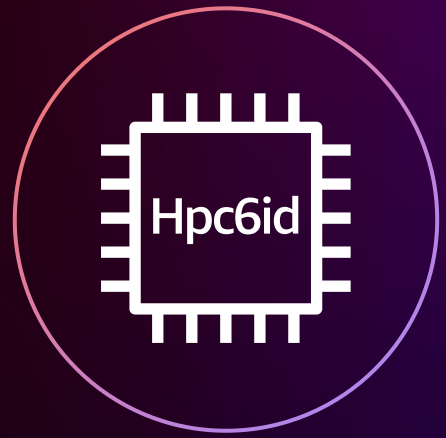
Compute intensive



Compute & network intensive



Data and memory intensive



aws

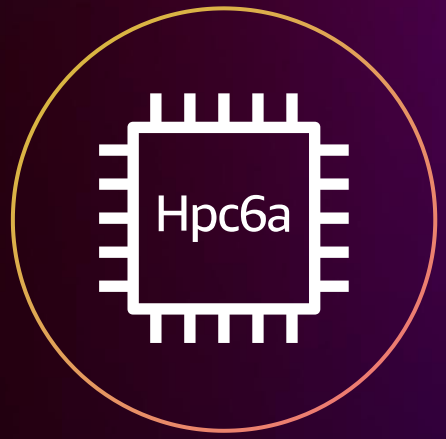
© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Amazon EC2 Hpc6a instances

AMD 3rd generation EPYC Milan processors:

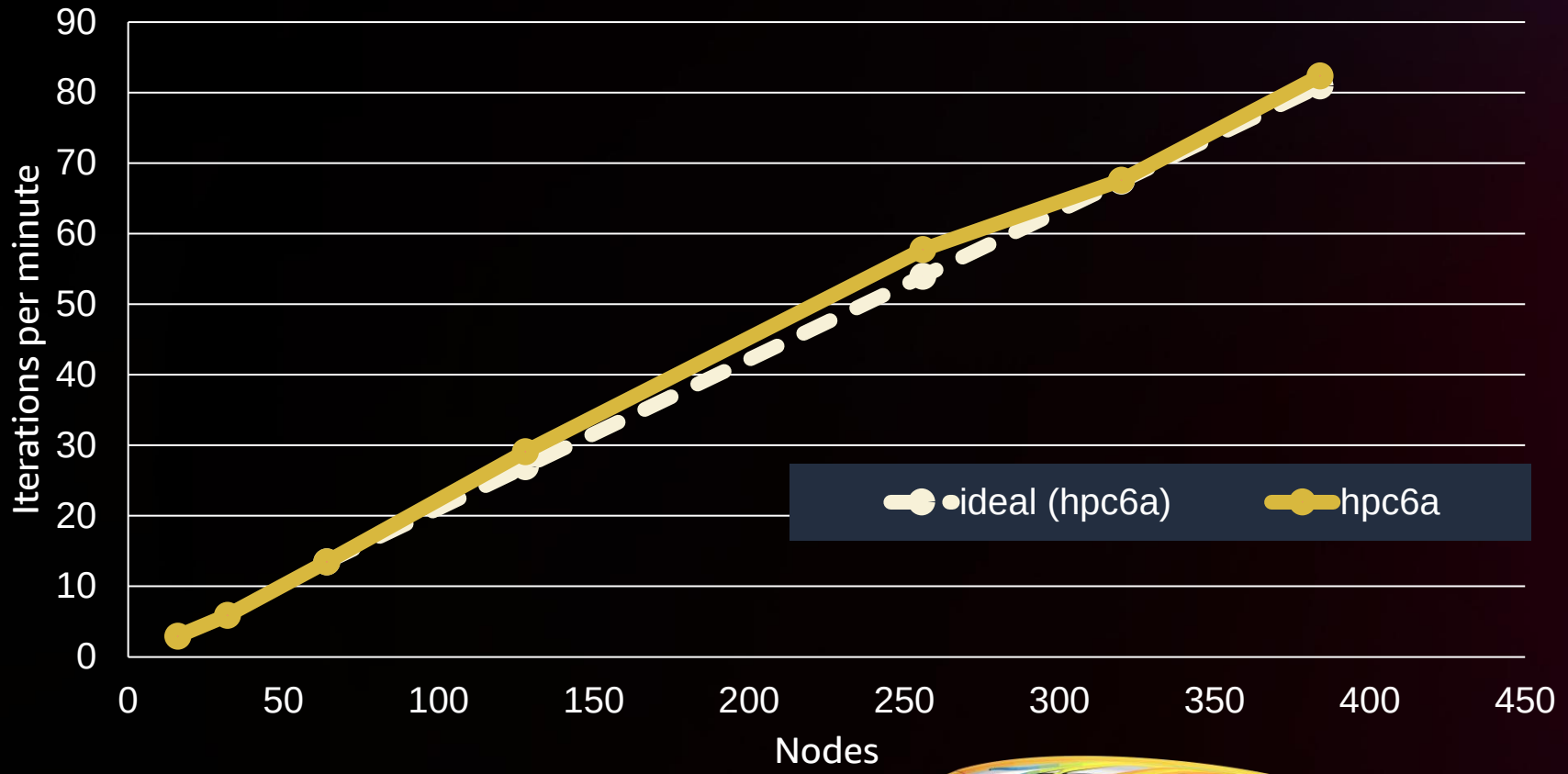
- One size: Hpc6a.48xl
- 96 physical cores (no simultaneous multithreading)
- 3.6 GHz all-core turbo
- 384 GB of RAM (4GB/core)
- **100 Gbps EFA**

Up to 65% better price performance over comparable Amazon EC2 x-86 based, compute-intensive instances



© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Amazon EC2 Hpc6a instances



26 days
(300 cores)

5 hrs
(40k cores)



© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

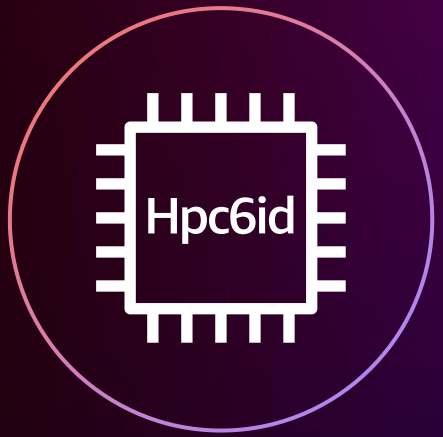
Amazon EC2 Hpc6id instances

New!

Intel 3rd generation Xeon Scalable processors:

- One size: Hpc6id.32xlarge
- 64 physical cores (no hyperthreading)
- Up to 3.5GHz
- 5 GB/s memory bandwidth per vCPU
- 1 TB of RAM
- Up to 15.2 TB local storage
- **200 Gbps EFA**

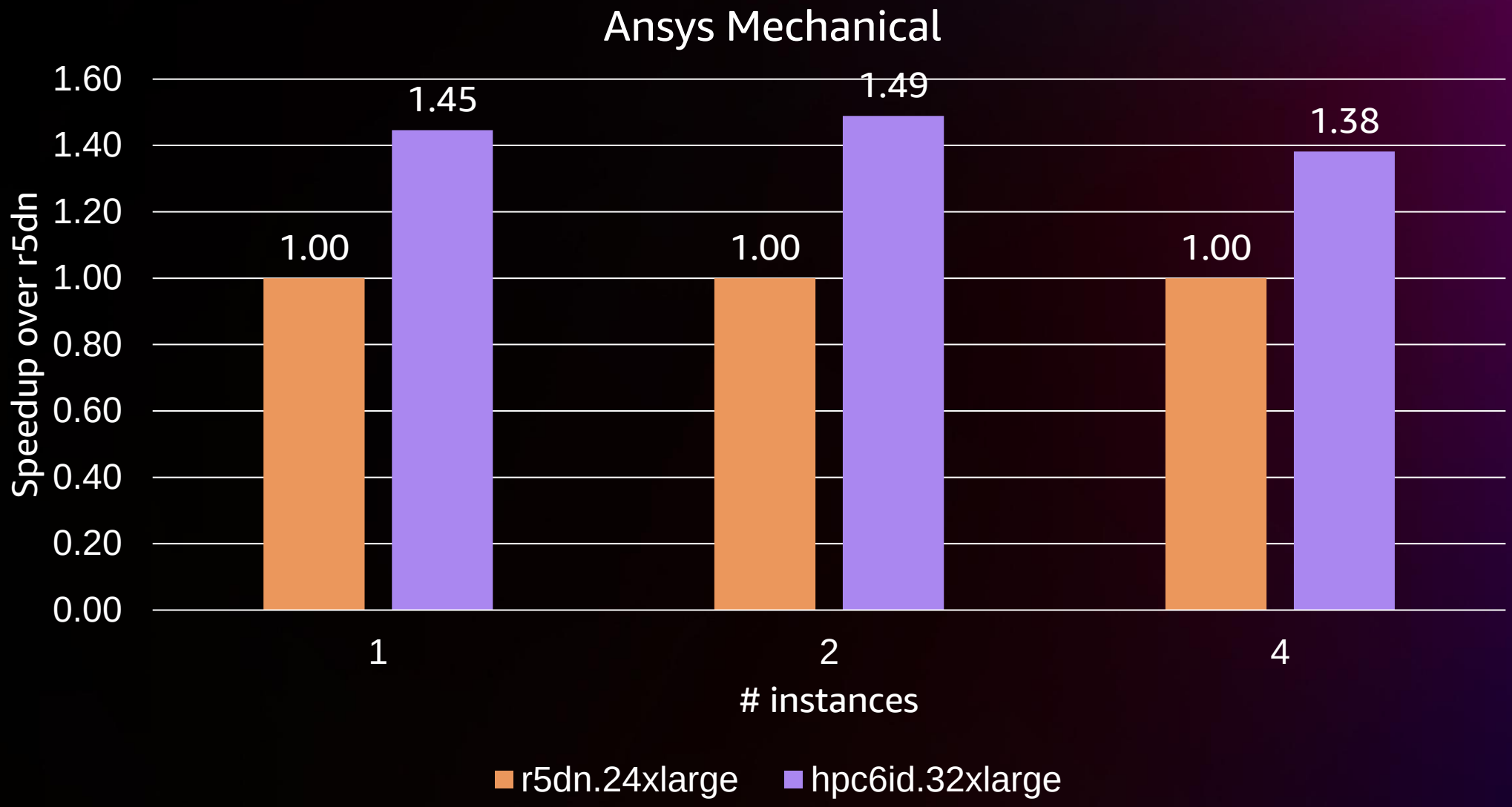
The best price performance for memory and data intensive HPC workloads in Amazon EC2



aws

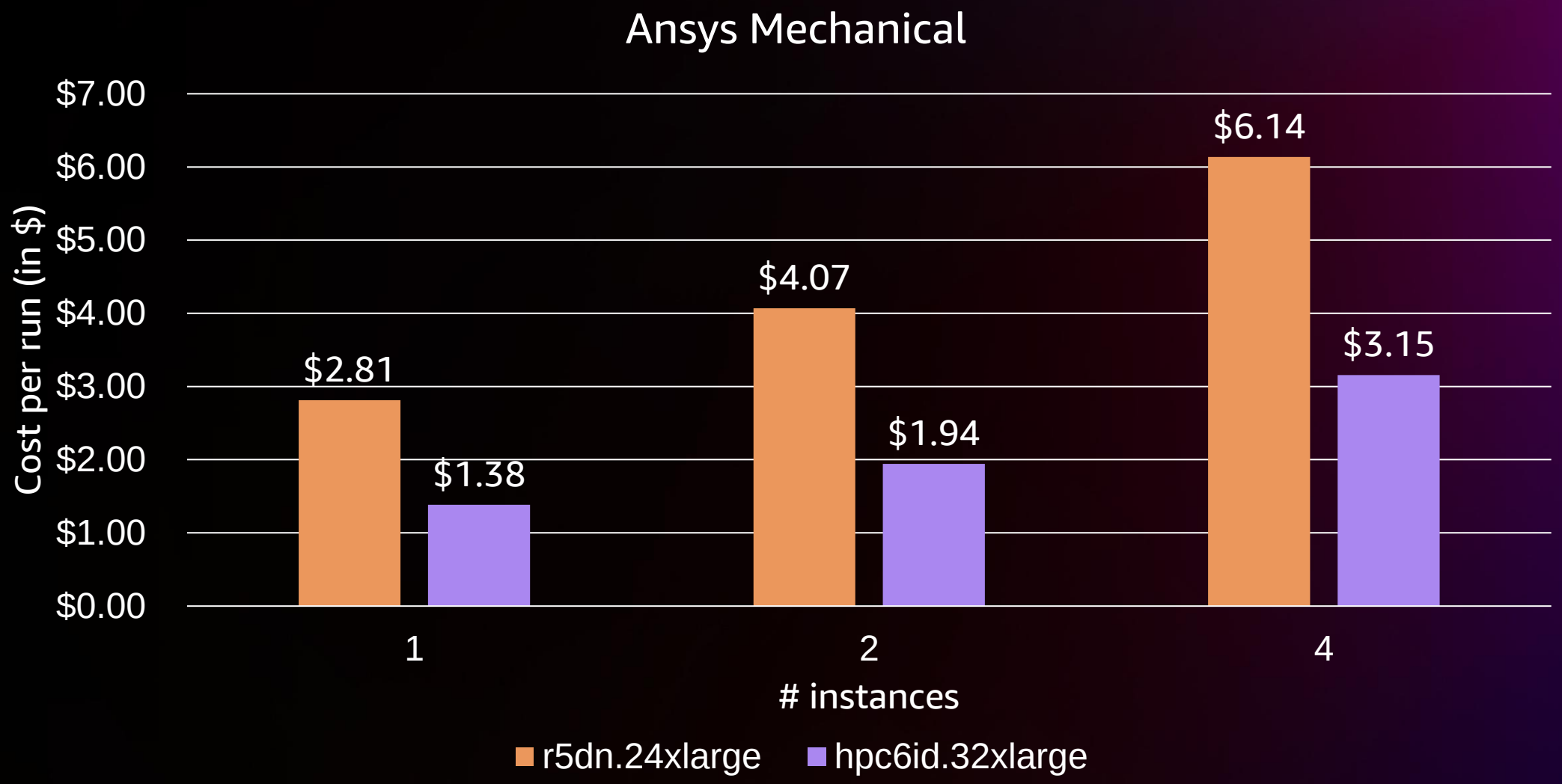
© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Example performance



© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Example cost reduction



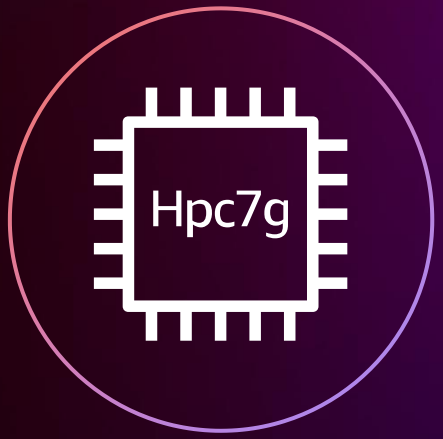
© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Amazon EC2 Hpc7g instances

Next generation AWS Graviton3E processors:

- 64 physical cores
- Up to 2.6 GHz
- 128 GB of RAM
- **200 Gbps with 2nd generation EFA**

Provides up to 35% higher vector instruction performance compared to existing instances based on AWS Graviton3 processors



Powered by the next generation AWS Nitro System

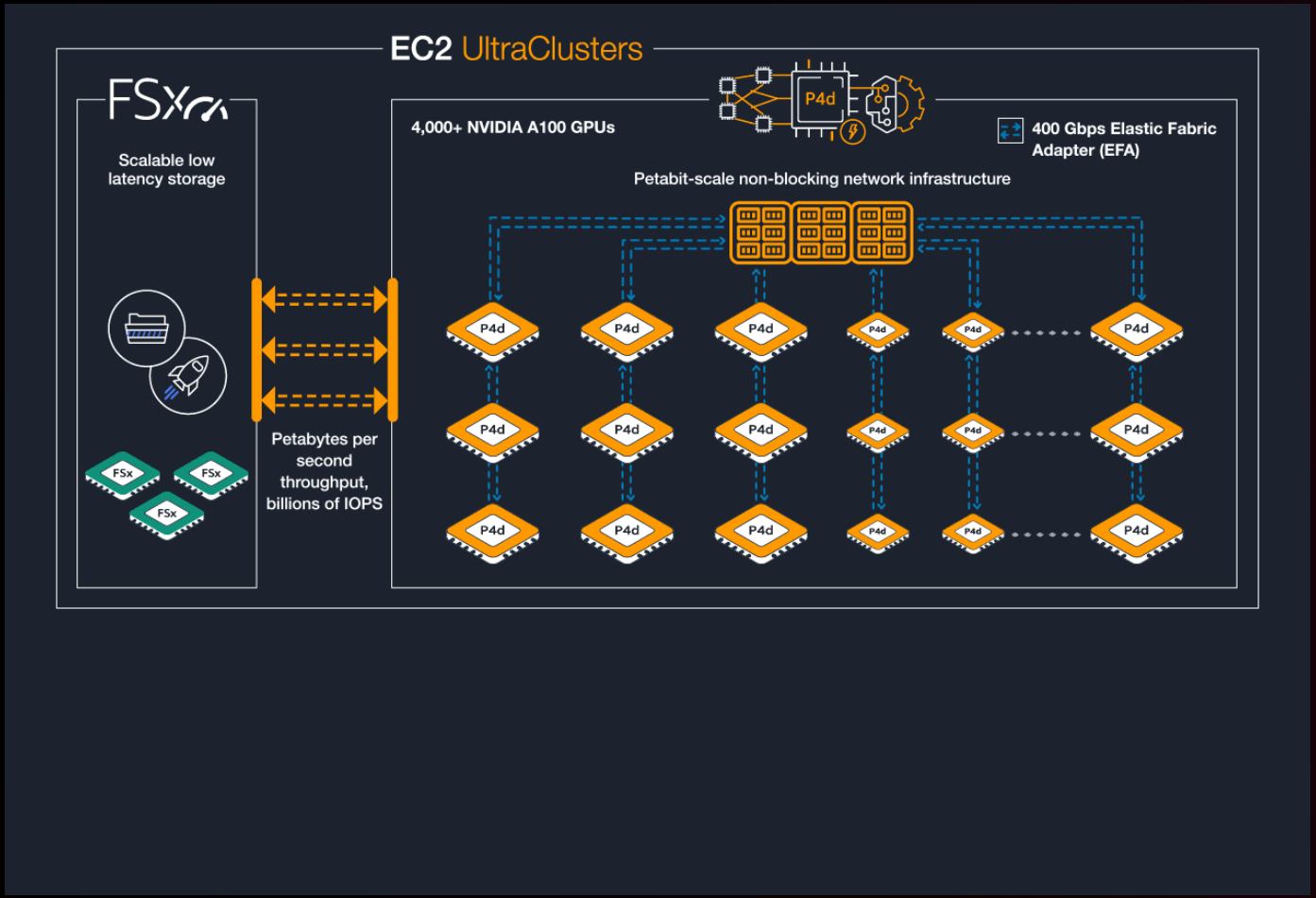
COMING SOON!

Go to https://day1hpc.com/ to learn more about EFA

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Amazon EC2 UltraClusters: P4d instances with NVIDIA GPUs

SUPERCOMPUTING-CLASS PERFORMANCE TARGETING LARGE-SCALE DISTRIBUTED TRAINING

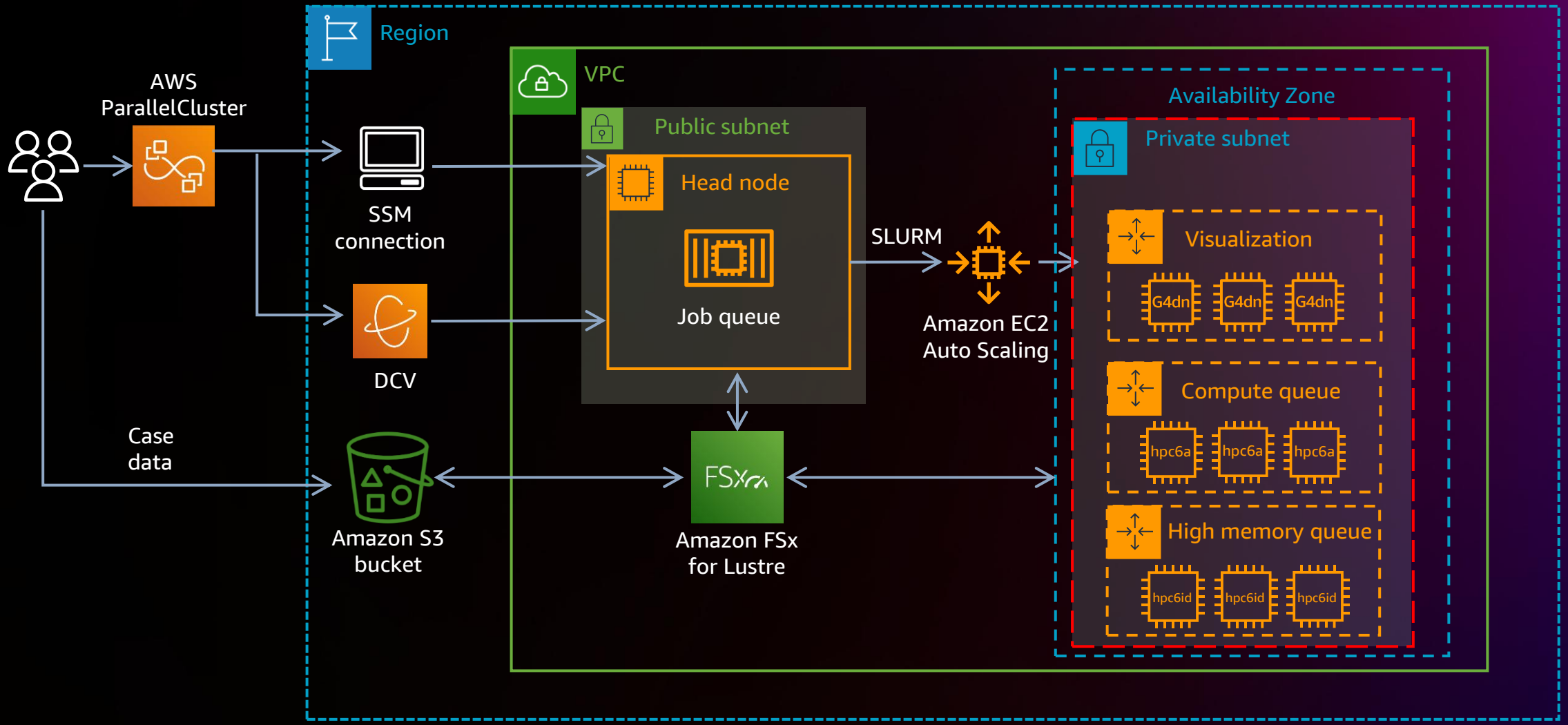


- 1–10s of thousands of NVIDIA A100 GPUs
- Fully non-blocking petabit-scale network infrastructure
- High throughput, low-latency storage from FSx for Lustre
- Current largest UltraCluster is capable of 7.5 exaflops (assuming 3k P4d, 24k GPUs, 312 FP16 TFLOPS per GPU)

aws

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Example architecture for cloud-centric CFD



© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Case study: Joby Aviation



- Novel aircraft meant limited prior knowledge; CFD was key to the design
- Conduct **thousands** of simulations on **hundreds** of cores using Amazon EC2 instances
- **AWS ParallelCluster** to easily deploy an HPC environment
- Amazon S3 Glacier to **store hundreds of terabytes of data**

"Using AWS helps us get results from our CFD work faster. In some cases, we get results in 24 hours that may have previously taken a week."

Alex Stoll
Aeromechanics Lead, Joby Aviation

aws
© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Growing together with the HPC community

The more we learn together, the more opportunity we have for stronger solutions.

There is no compression algorithm for experience.

– Andy Jassy, CEO, AWS

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Thank you!

Dr. Neil Ashton
neashton@amazon.com

Heidi Poxon
hpoxon@amazon.com



Please complete the session survey in the **mobile app**

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.