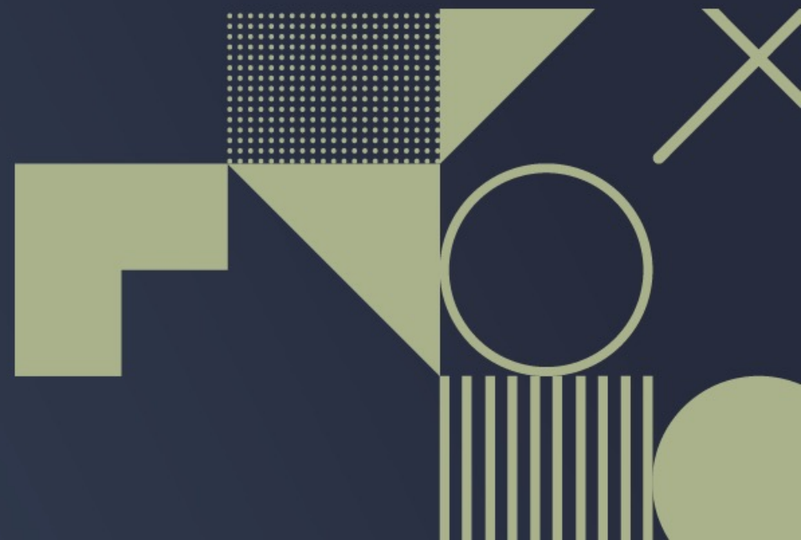
HPC On The Cloud: Opportunities to Redesign the Supercomputer

Brian Barrett

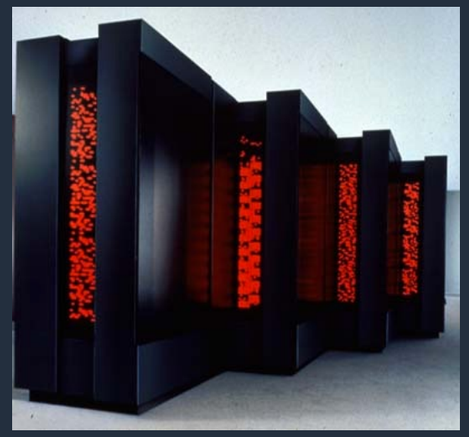
Principal Engineer, Amazon Web Services

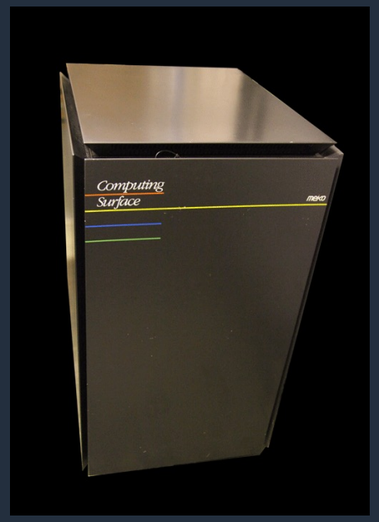
November 15, 2021



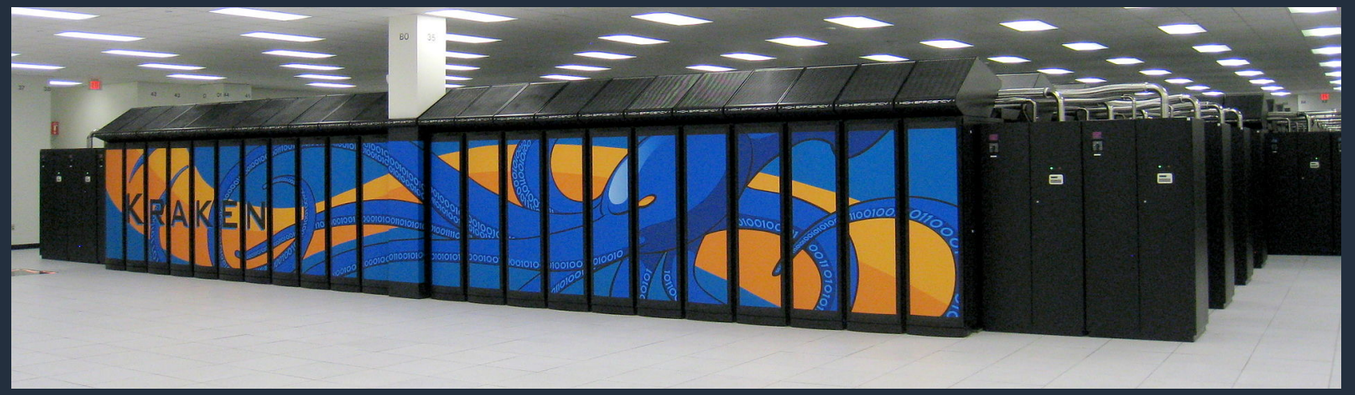
SuperComputing Evolution

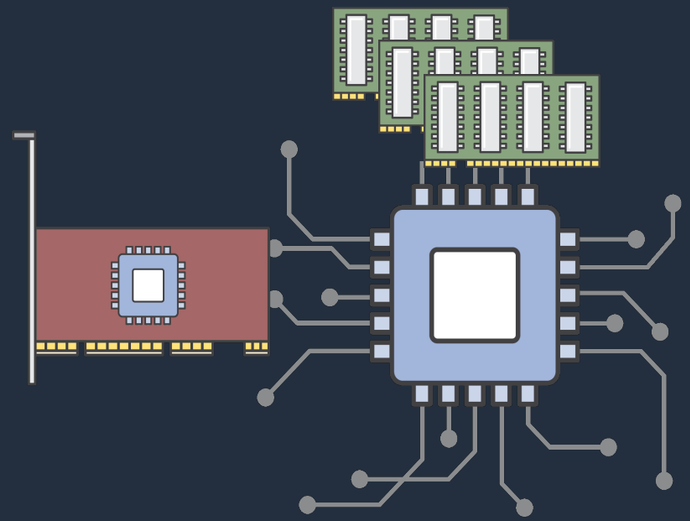
SuperComputers used to be interesting…



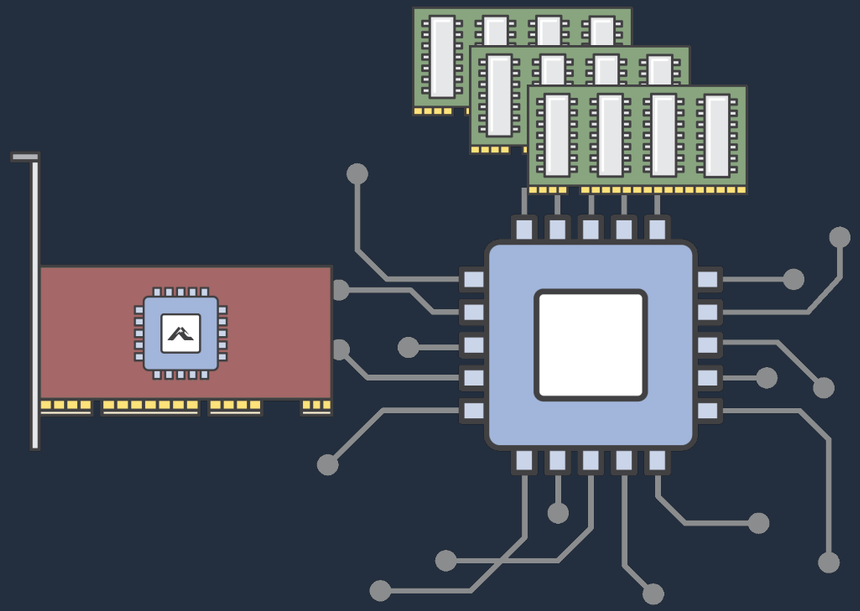


And then they got commodity(-ish)…

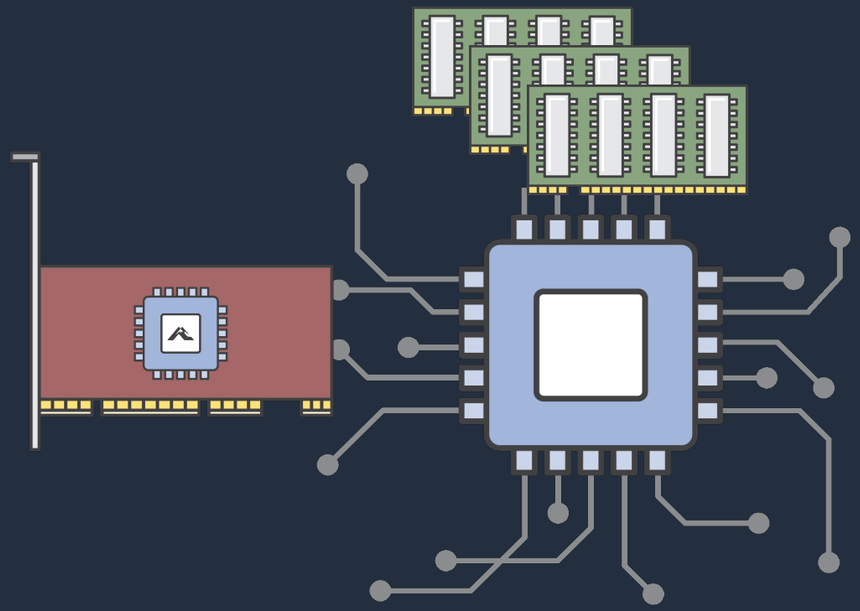


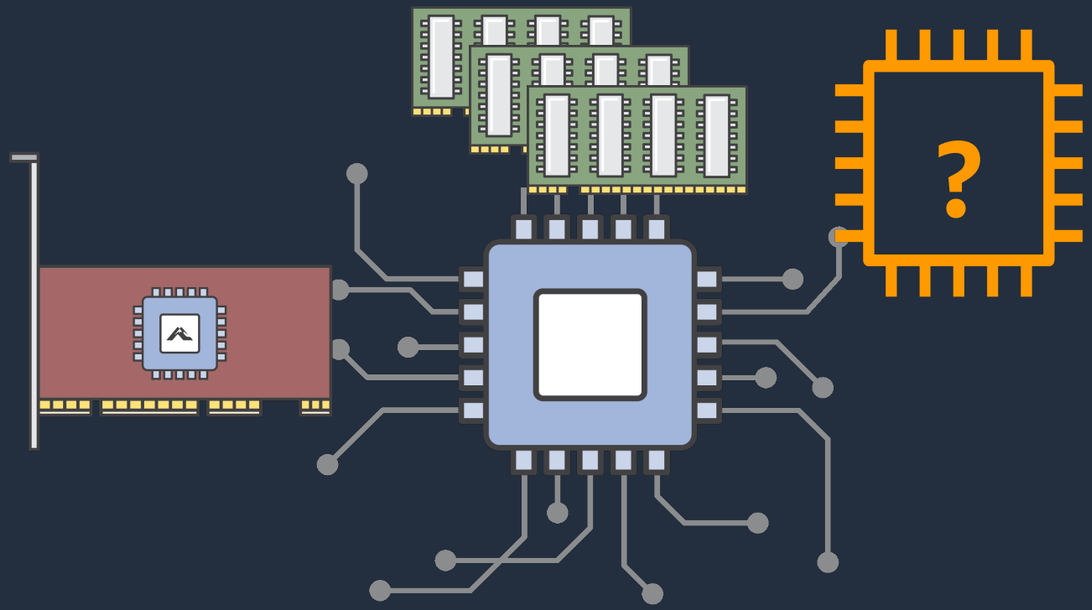


And then came the Cloud…



And then came the Cloud…





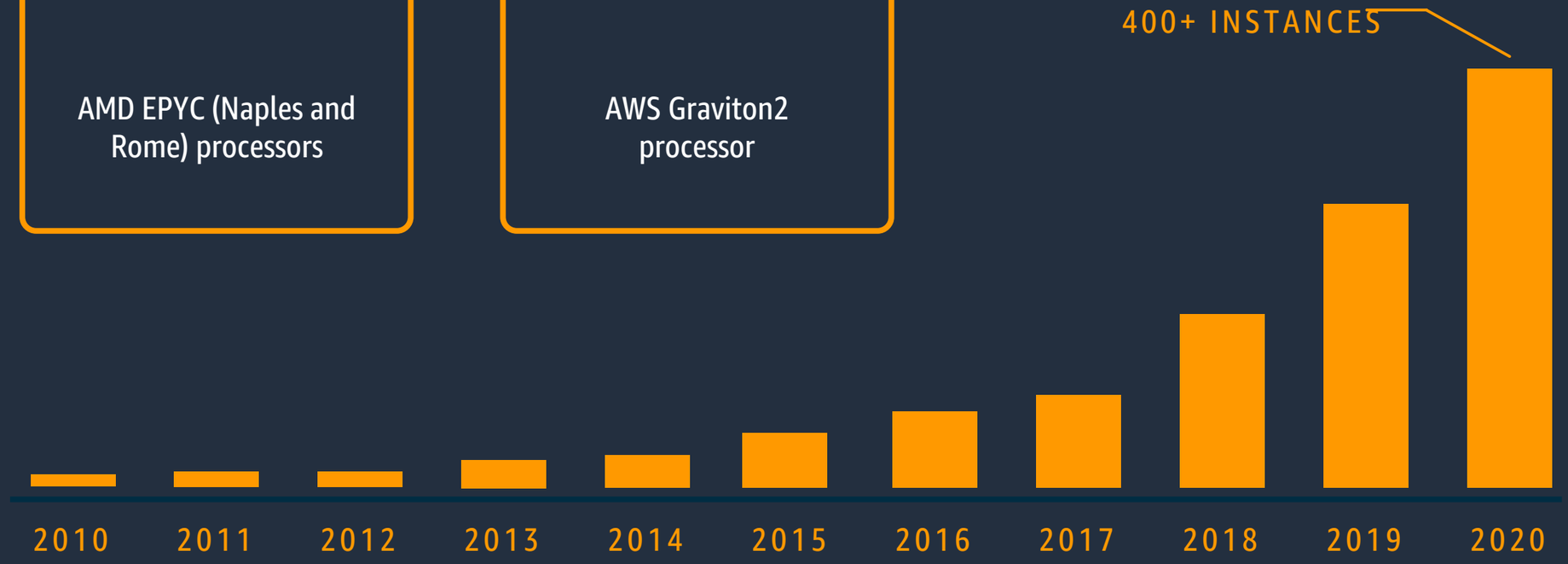
Why Cloud for HPC?

Cloud Advantages: Customization

Intel Xeon Scalable (Skylake and Cascade Lake) processors

AMD EPYC (Naples and Rome) processors

AWS Graviton2 processor



Wide Variety of Instance Types for HPC workloads

Graviton2

C6gn.16xlarge
1S, 64c, 2GB/core
100Gb

Storage-Dense

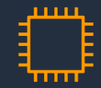
i3en.{24xlarge,metal}
60 TB NVMe
100Gb

x86

Ice Lake
m6i.32xlarge
2Sx 32c, 8GB/core
50Gbps

Cascade Lake
m5zn.24xlarge (4.5GHz)
2Sx 12c, 8GB/core
100Gbps

Cascade Lake
m5n.{24xlarge,metal}
2Sx 24c, 8GB/core
100Gbps



Cascade Lake
r5n.{24xlarge,metal}
2Sx 24c, 8GB/core
100Gbps
+ 'd' variants with NVMe (r5dn, m5dn)

Skylake
c5n.{18xl,metal}
2Sx 18c, ~4GB/core
100Gbps

Accelerator

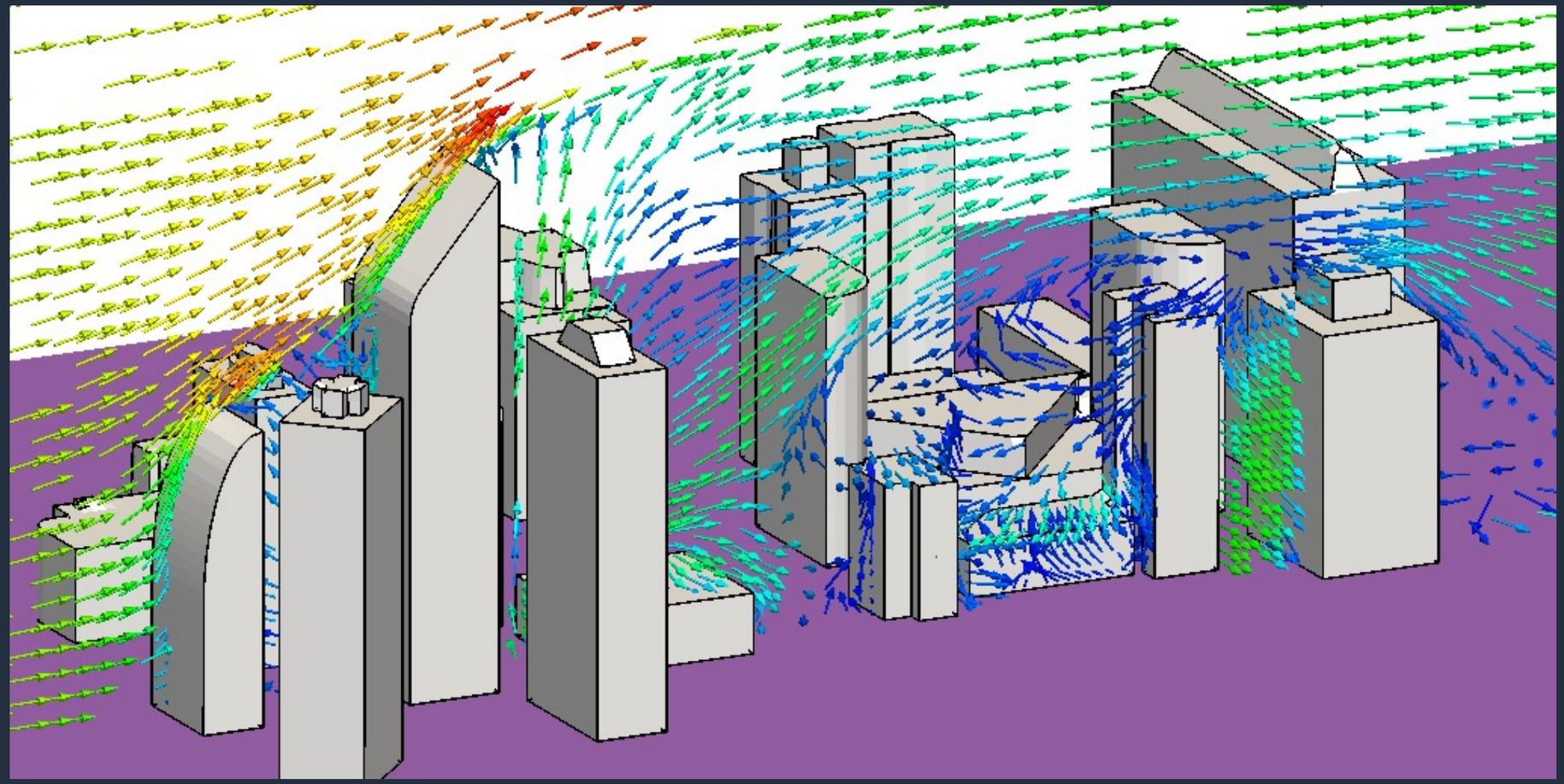
p4d.24xlarge
8x A100 GPUs
400Gb

p3dn.24xlarge
8x V100 GPUs
100Gb

inf1.24xlarge
16 Inferentia chips
100Gb

G4dn.{16xlarge, metal}
8x T4 GPUs
100Gb

Complex problems made tractable



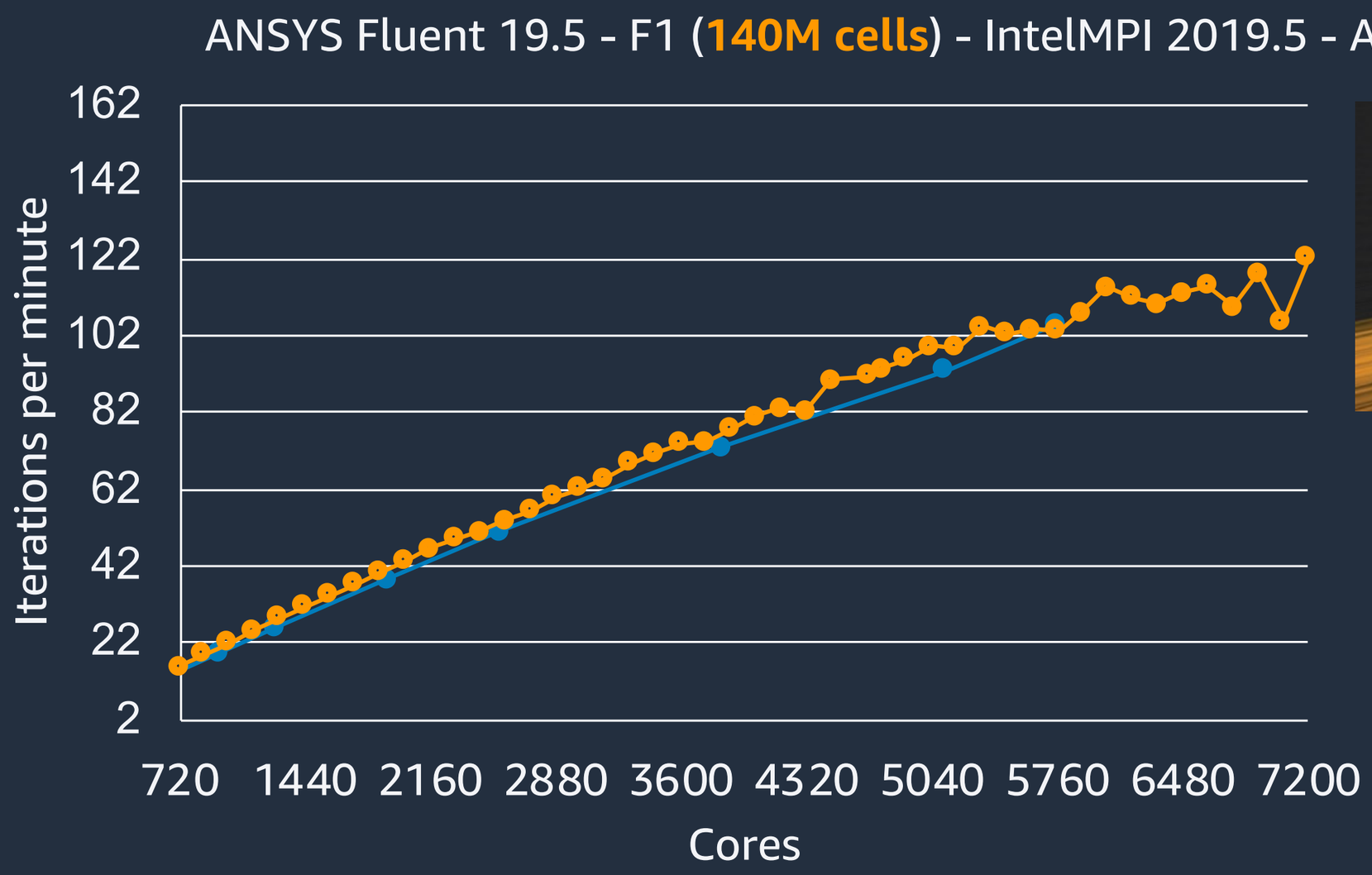
Credit: CFD.direct, thanks to Chris Greensheilds

Architects simulate fluid dynamics around proposed buildings to estimate the impact on “Pedestrian Comfort” (which windy sidewalks, due to turbulence impact badly).

Previously, this was the domain of $10M supercomputers, expensive software, and sophisticated users.

Today, it’s available for $1.75/hr on AWS, enabling better decisions before concrete is poured and mistakes become permanent.

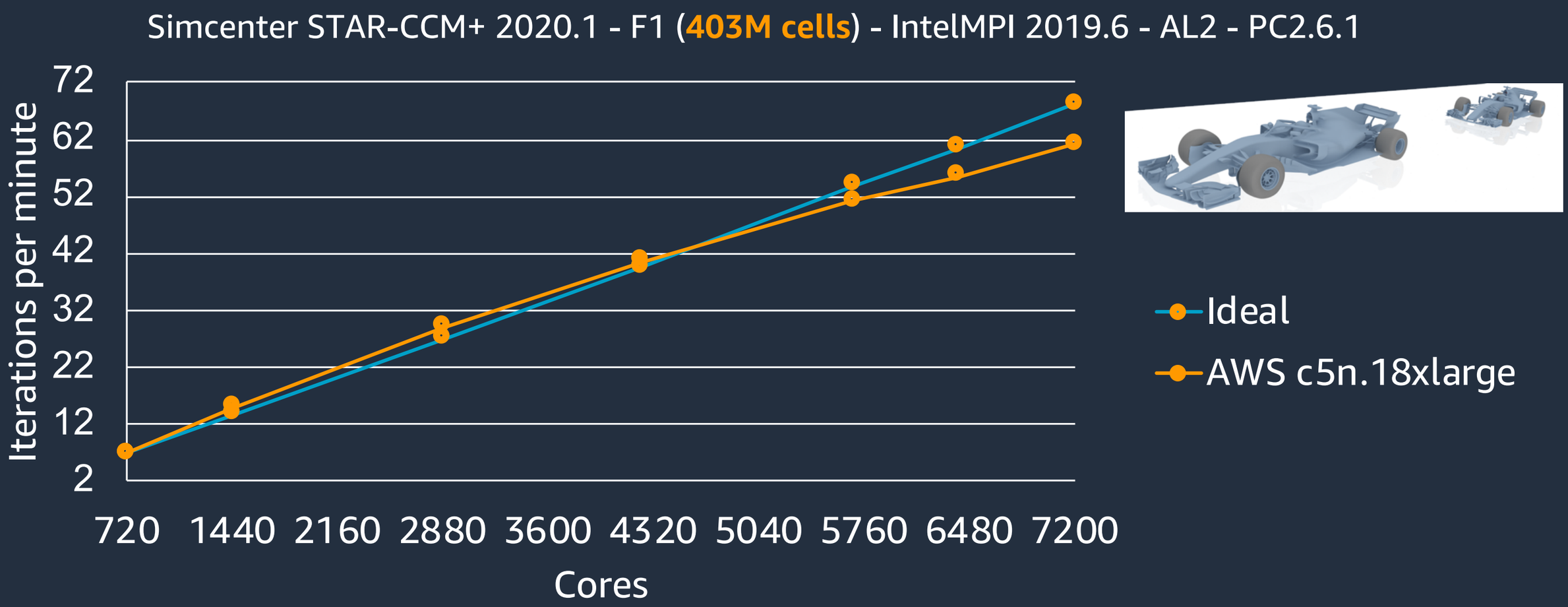
Scaling on AWS - ANSYS Fluent



Cray XC50

AWS c5n.18xlarge

Scaling on AWS – STAR-CCM+



June 2021 TOP500: #40 Position Using C/M/R Instances

9.95 petaflops (Rmax) of HPL performance at 65.9% HPL efficiency using an all-CPU cluster with 172,692 Intel cores (4,096 C5/M5/R5 instances)

Descartes Labs

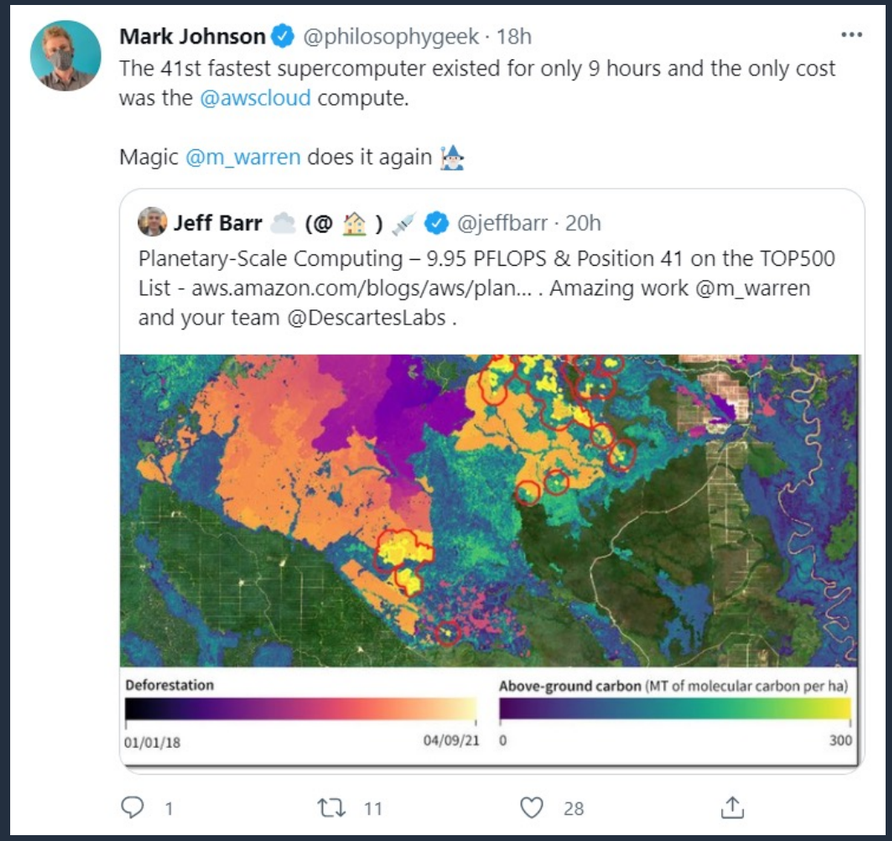
+

aws

Descartes Labs Achieves #40 in TOP500 with Cloud-based Supercomputing Demonstration Powered by AWS, Signaling New Era for Geospatial Data Analysis at Scale

Descartes Labs uses Amazon Web Services, Inc. (AWS) to run a 9.95 petaflops High-Performance LINPACK (HPL) benchmark, placing it #40 in the June 2021 TOP500 ranking. The company improves on its previous AWS-based 2019 TOP500 submission (1.926 petaflops, #136) by 417% in HPL performance and 96 ranking spots in only two years.

"Mike has worked for decades to prove to the world that mass-produced, commodity hardware and software can be used to build a supercomputer, and the results more than speak for themselves." – *Jeff Barr, VP & Chief Evangelist, AWS*



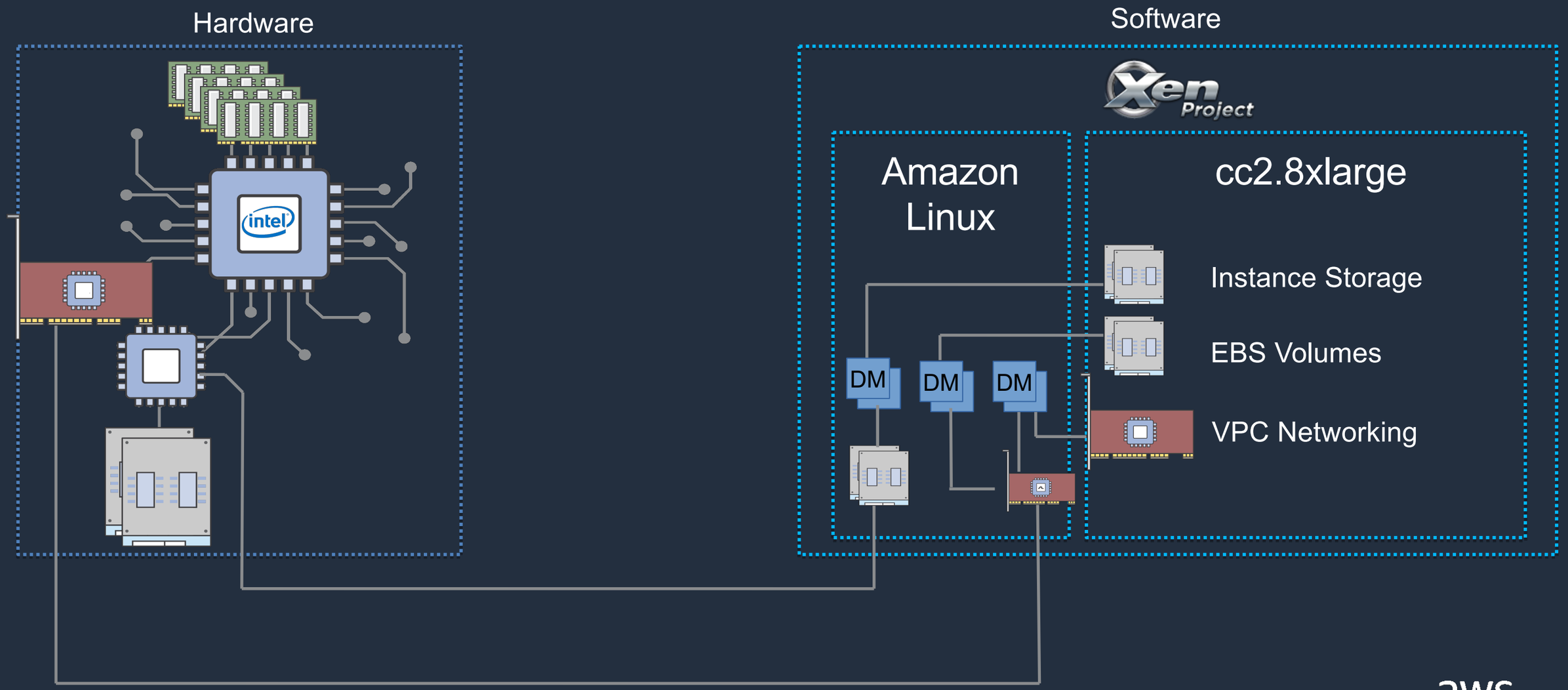
Jeff Barr Blog- https://aws.amazon.com/blogs/aws/planetary-scale-computing-9-95-pflops-position-41-on-the-top500-list/

Note: Descartes Labs moved up from #41 to #40 after official rankings were announced due to the removal of the #33 entry

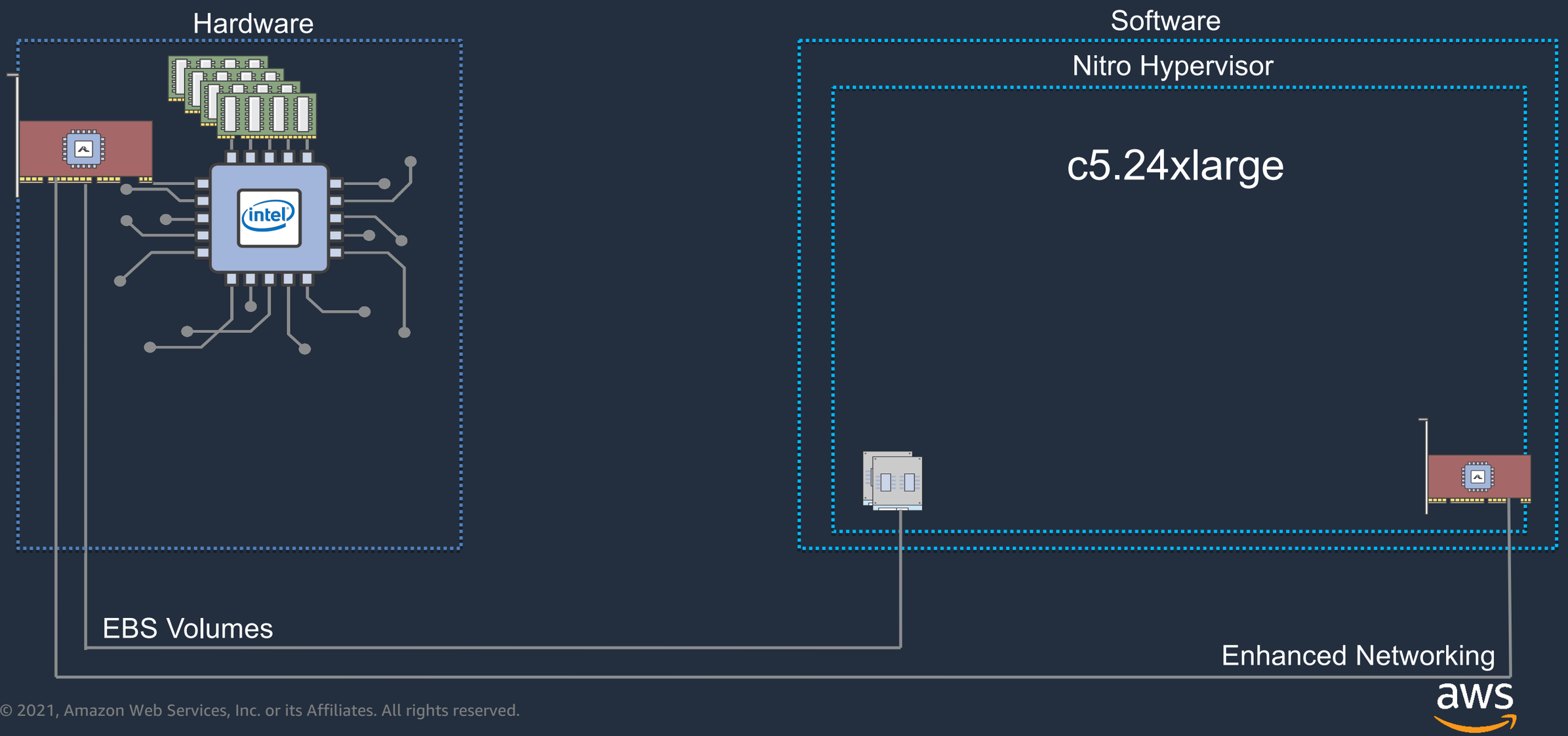
aws

Cloud HPC Technologies

Traditional Cloud Hypervisor Model



Nitro Hypervisor Model



Hypervisor Costs

HPCG (2 node)

24xlarge	46.6565 GFLOPS
Metal	48.5497 GFLOPS

NAS Parallel Benchmarks (2 node)

Benchmark	24xlarge	Metal	% improvement
BT	306.06 s	309.22 s	-1%
CG	156.82 s	156.25 s	0.3%
EP	29.39 s	29.43 s	-1.4%
FT	126.25 s	127.40 s	-1%
LU	154.42 s	148.59	3.9%

aws

GPCNet Benchmark Results

Instance	Avg (us)	P99 (us)	P99.9 (us)
C5n.18xlarge	34.1	45.0	56.2
C6gn.16xlarge	30.08	41.8	48.18
Cloud InfiniBand	3.27	24.93	46.1

4 instance results, 36 processes per instance

System Software Challenges

All the questions of on-prem, plus

- What instance should I use?
- What OS should I use?
- What compiler should I use?
- What MPI should I use?

AWS HPC Orchestration Options

AWS Batch

AWS ParallelCluster

ALINUX 2 | CENTOS 7 | UBUNTU 18/20

DCV | EFA | OPEN MPI | INTEL MPI | NCCL

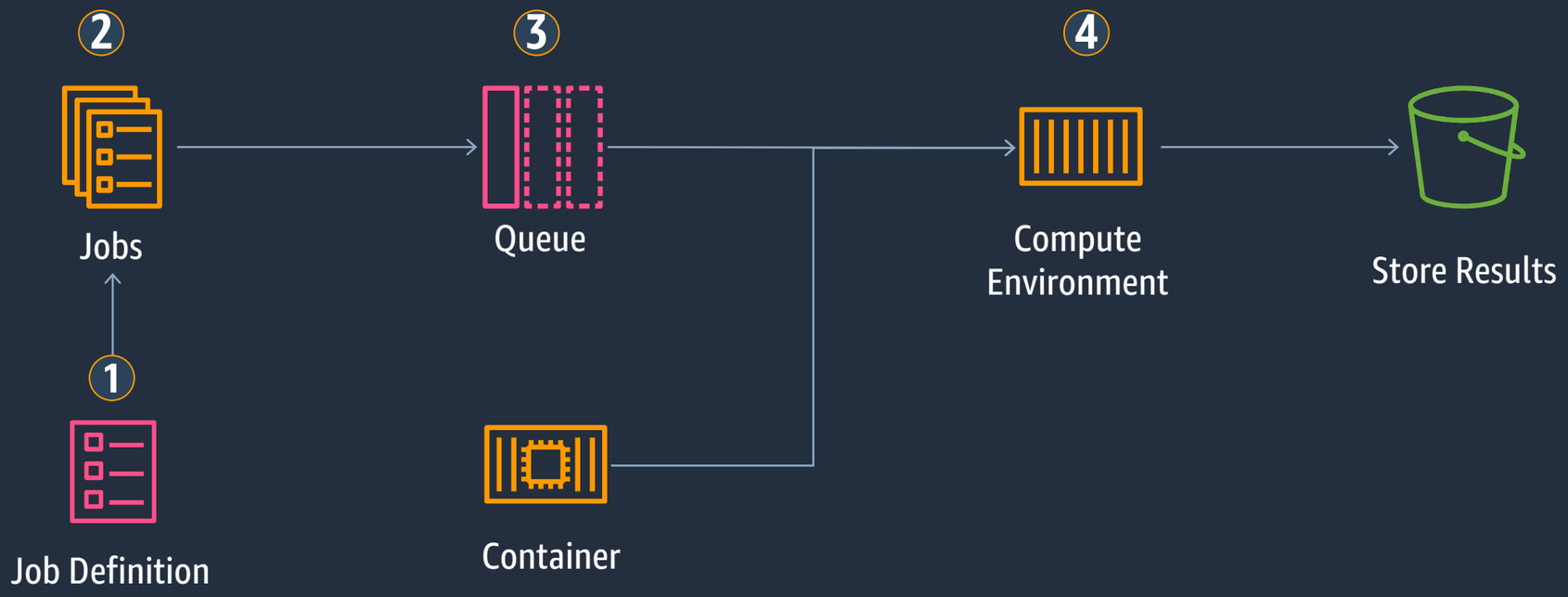
SLURM | SGE | TORQUE | AWS BATCH

FSX | EFS | S3 | EBS | RAID

ON-DEMAND | SPOT

VPC & SUBNETS

AWS Batch Architecture



AWS Batch in Numbers

1,243,000 vCPUs - Largest Cluster

500,000 – The record number of simultaneous jobs (and container images)

---

1 Billion jobs run in August across all regions

---

500 Tasks per second – Fastest dispatch rate in a production environment

We have thousands of concurrent batch customers, some of whom run some of the largest workloads in the world.

Oh, and it's **free**.

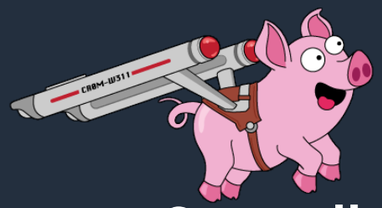
aws

AWS Batch: A Natural Fit for Workflow Managers

AWS Native

AWS Step Functions

3rd Party
Open Source



Cromwell

Airflow

METAFLOW

Nextflow

Luigi

Flyte

New Opportunities…

Operating Systems

- Choice of OS can be pushed to the job level
- Noise is still important – Return of the lightweight kernels?
- We haven’t solved many of the OS challenges for HPC
    - Shared memory & MPI mismatch
    - Page allocation & memory pinning
    - Upgrades and reproducibility

Containers

- Importing/injecting libraries from base OS deeply dissatisfying
    - We’ve resigned ourselves to having a base OS
    - Do we need better composability?
- Relying on (slow) portable libraries deeply dissatisfying
- Container runtime space still evolving

aws

Run-times

- The accelerators are coming; how are we going to manage them?
- Removing OS from customer worries appealing; how build a portable, secure, scalable run-time interface?
- How far can we push latency hiding techniques?

High Performance Computing on AWS

Innovate fast. Innovate securely. Innovate within budget.

aws