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

CMP205

HPC on AWS: Solve complex problems with pay-as-you-go infrastructure

Ian Colle (he/him)

Batch and HPC General Manager
AWS

Chen Su (she/her)

Senior Director, Research IDS
Eli Lilly

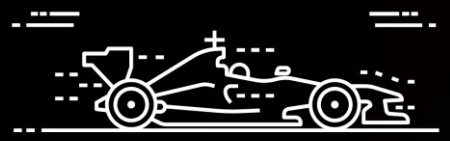
© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Agenda

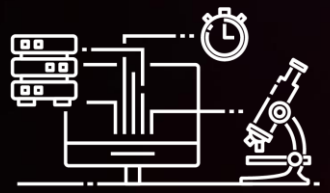
- HPC is all around us
- HPC on AWS
- Eli Lilly drug discovery on AWS
- What will you discover with HPC on AWS?

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

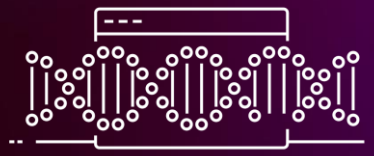
HPC is all around us



Build designs faster with computational fluid dynamics (CFD) simulations



Fast-track drug discovery and structure-based drug design



Advance genomics insights using predictive, real-time, or retrospective data applications



Run geoscientific simulations and seismic processing and iterate models faster



Conduct grid-computing simulations and identify portfolio risks and hedging opportunities



Process complex workloads and analyze massive data pipelines to further research

aws

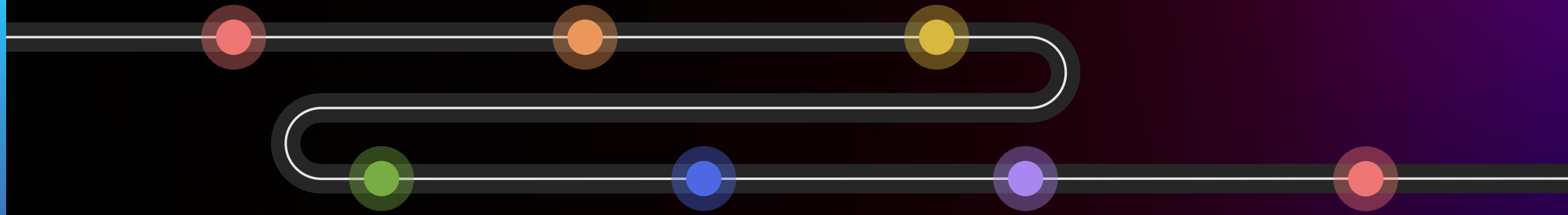
© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

HPC in the cloud

The HPC market is nearly $40B in 2022

Cloud is 14.7% of the overall HPC market

Cloud is growing at a rate greater than 23% in 2022



The on-premises HPC server market is growing ~7–8% a year

HPC cloud usage is expected to grow 17.6% over the next 5 years

HPC cloud revenue is expected to exceed $11B by 2026

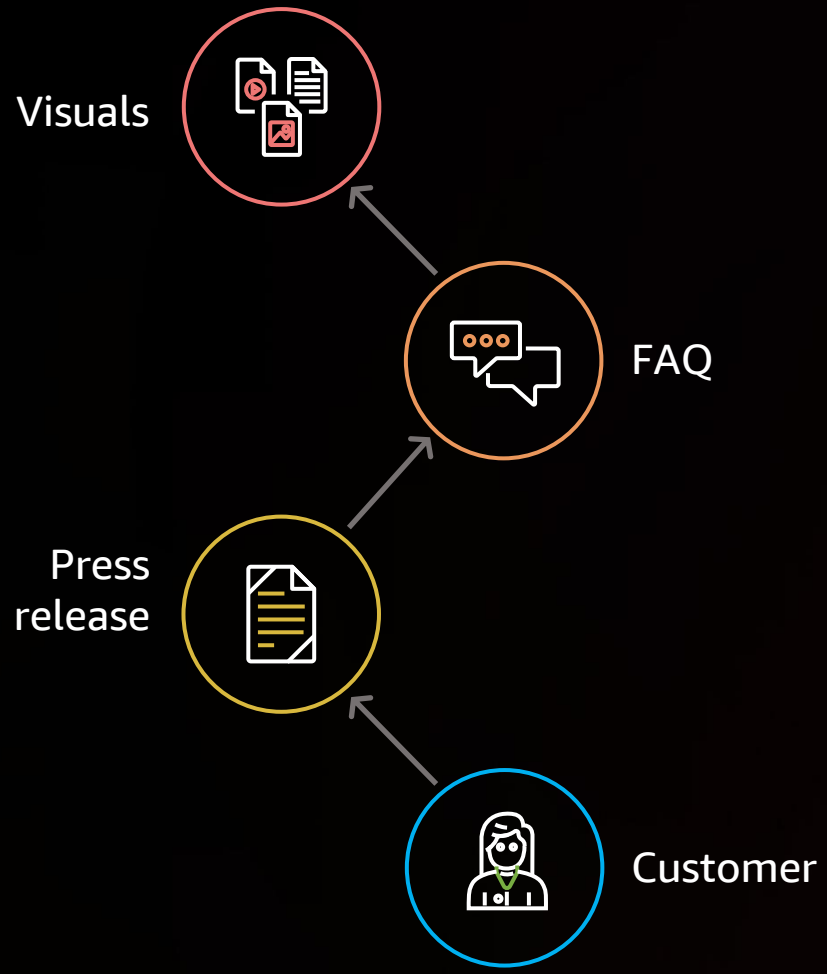
Cloud is used primarily for burst capabilities, yet almost 50% of the users are altering on-premises deployments due to cloud

Source: Hyperion Research 2022

aws

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

How AWS thinks about HPC



The questions we ask before development

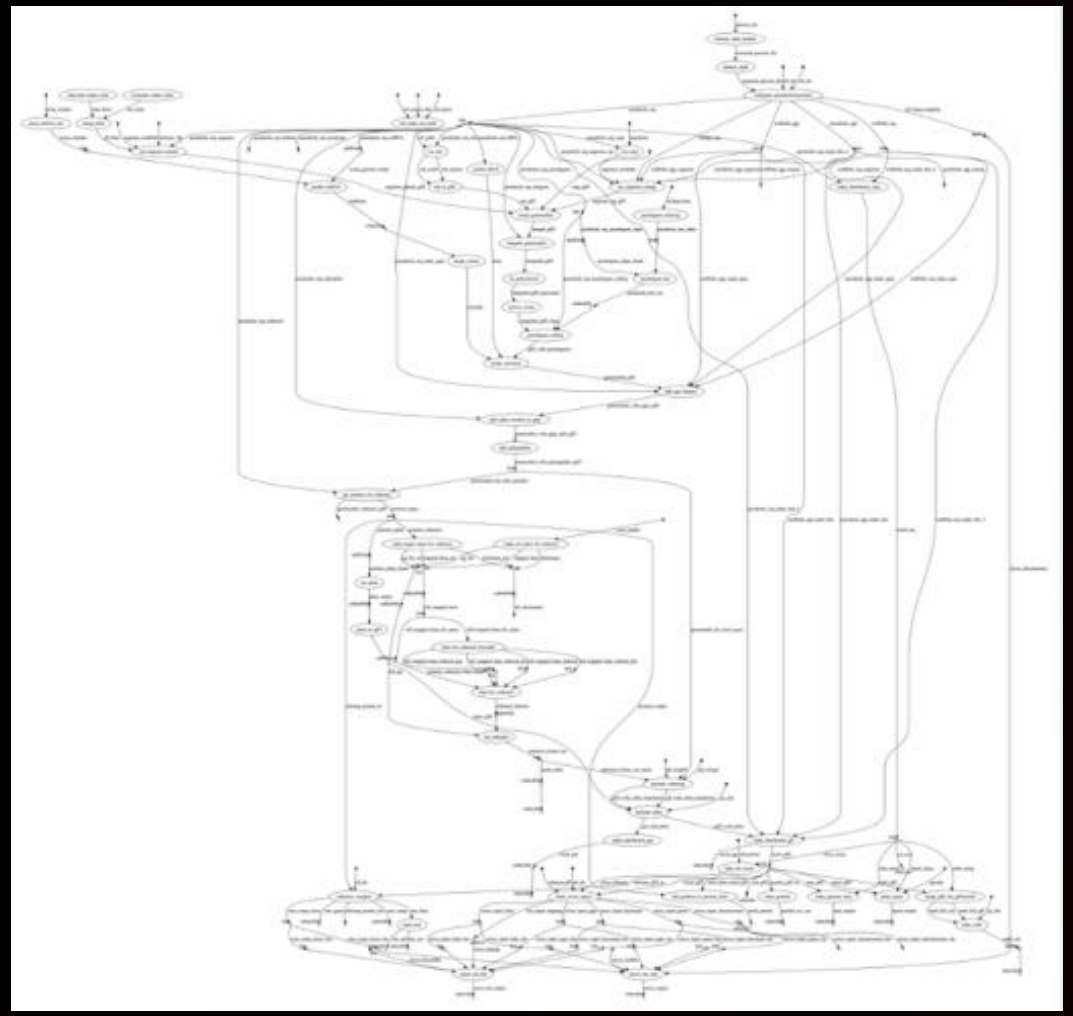
- Who is your customer?
- What is the customer problem or opportunity?
- What is the most important customer benefit?
- How do you know what your customer needs or wants?
- What does the experience look like?

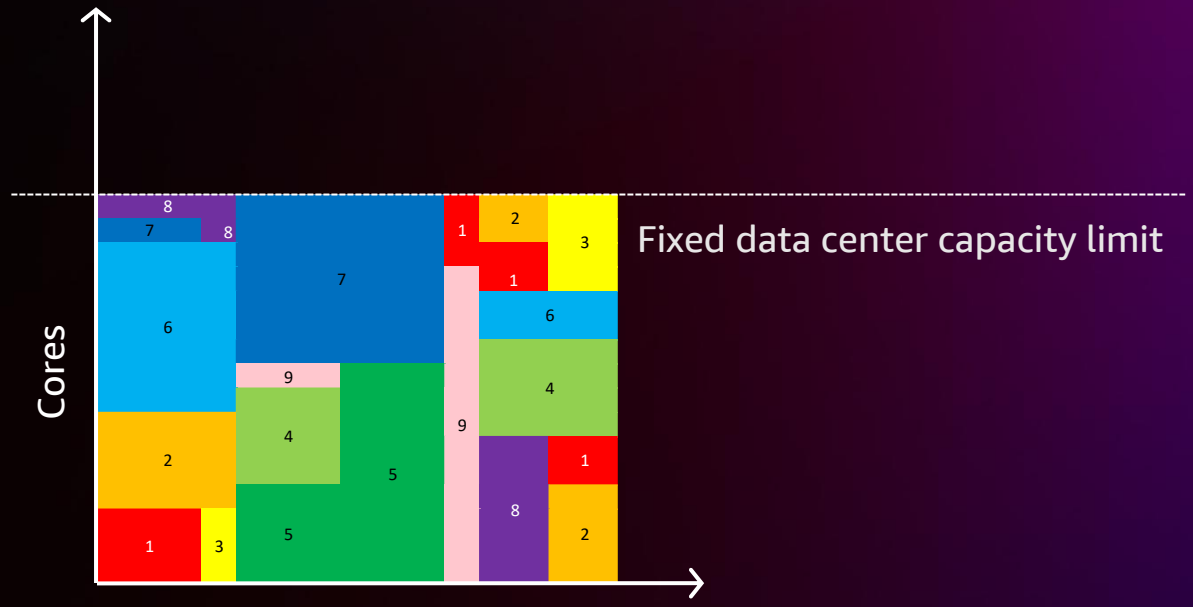
Customers want choice and price performance

- More than 200 AWS services
- Choice of Intel, AMD, and Arm-based processors
- More than 500 different Amazon EC2 instance types to optimize every workload
- Multiple purchase options (including EC2 Spot, On-Demand, Savings Plans)

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

Workflows don’t cope well in fixed-size boxes

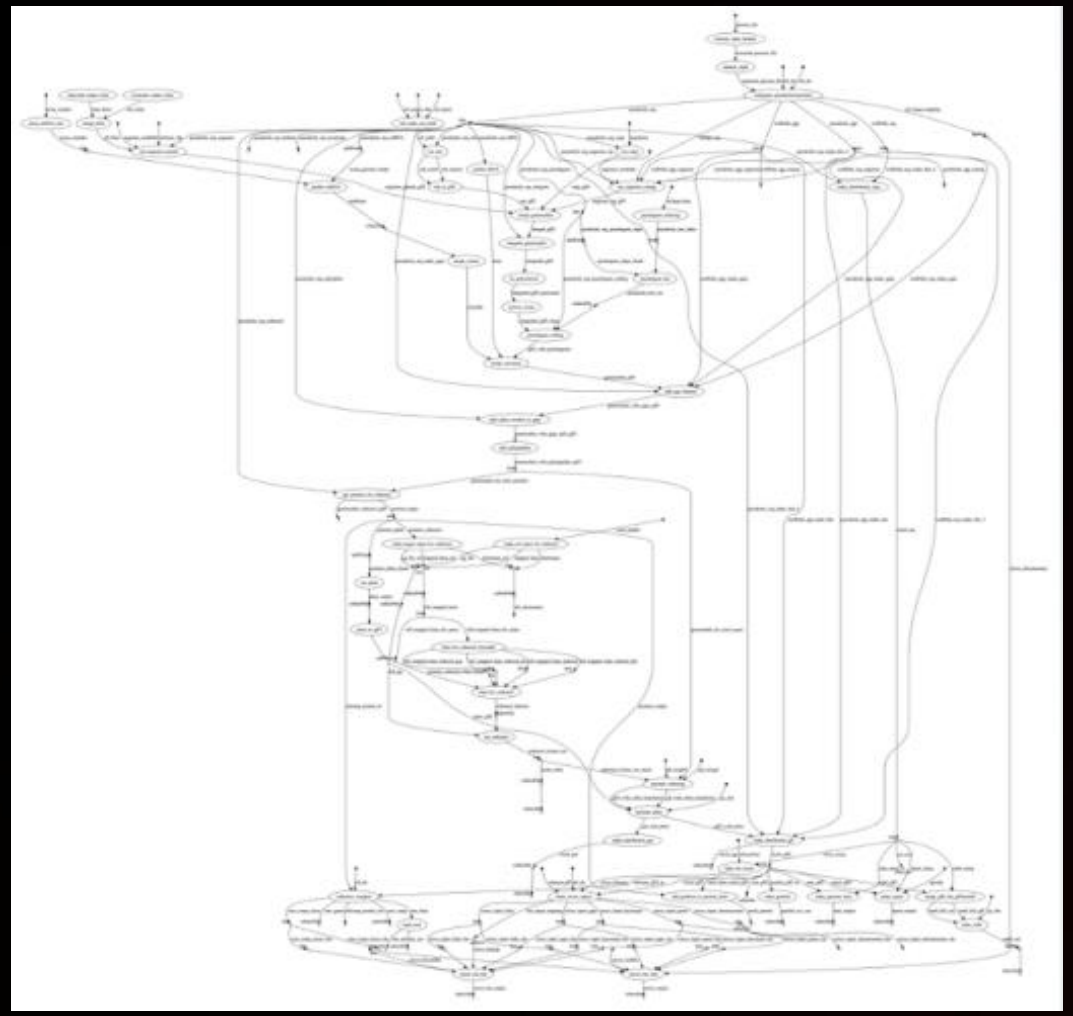


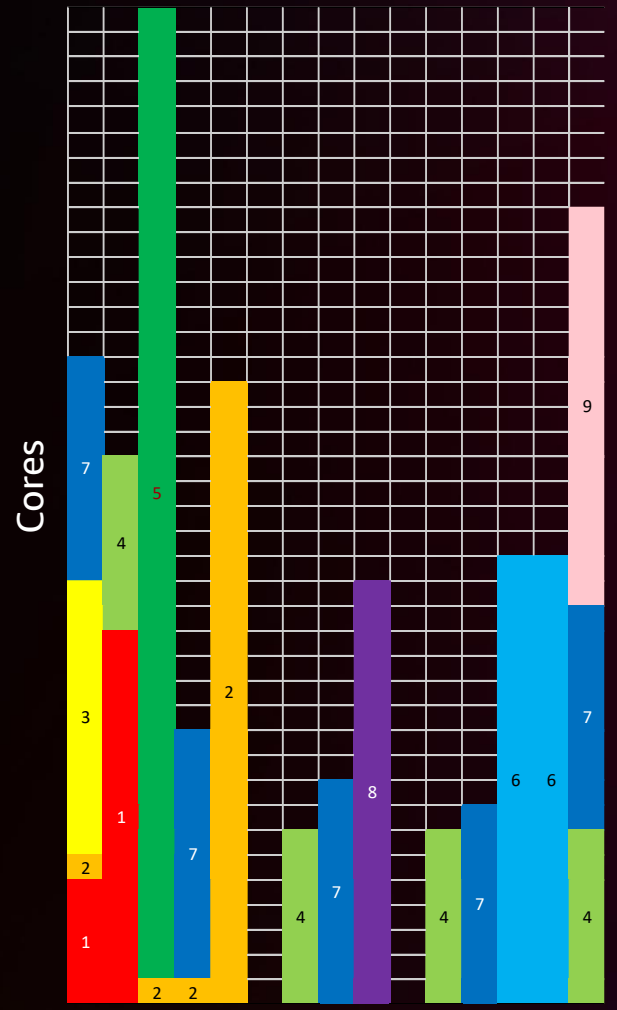


- Limited cores
- 1 or 2 types:
  - 128 x vanilla nodes
  - 4 x “fat” memory nodes
  - Maybe some GPUs

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

Workflows love elasticity





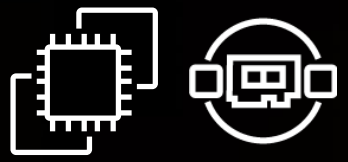
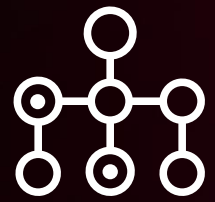
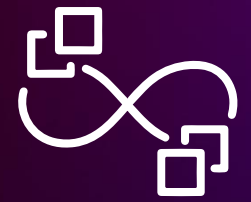
More than 500 instance types

- Multiple CPU and memory architectures
- GPUs and other accelerators (like FPGAs)
- Literally millions of cores
- Fast storage (FSx for Lustre) on demand and cold storage too

aws

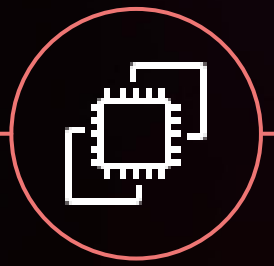
© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

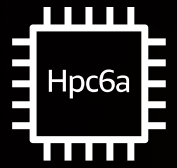
HPC on AWS

Compute & Networking	Storage	Scheduler	Management
			
Amazon EC2 Elastic Fabric Adapter (EFA)	Amazon FSx for Lustre	AWS Batch	AWS ParallelCluster

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

HPC-optimized Amazon EC2 instances

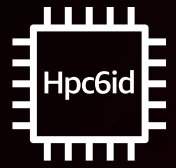




Powered by 3rd generation AMD EPYC processors

Amazon EC2 Hpc6a instances

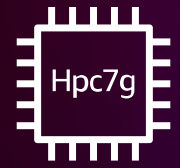
Compute-intensive applications like Computational Fluid Dynamics and Numerical Weather Prediction

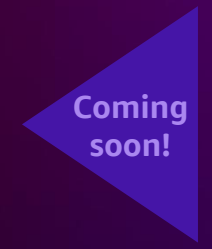


Powered by 3rd generation Intel Xeon Scalable processors

Amazon EC2 Hpc6id instances

Memory-bound and data-intensive workloads like Finite Element Analysis and seismic simulations





Powered by the next generation AWS Nitro System

Amazon EC2 Hpc7g instances

Based on custom AWS Graviton3E processors with low latency, and high network performance for MPI-based applications

aws

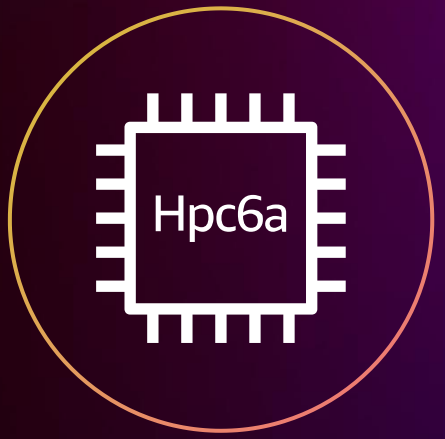
© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Amazon EC2 Hpc6a instances

AMD third-generation EPYC Milan processors:

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

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



AWS Regions:

US East (Ohio)
GovCloud (US-West)
EU (Stockholm)

AMD

aws

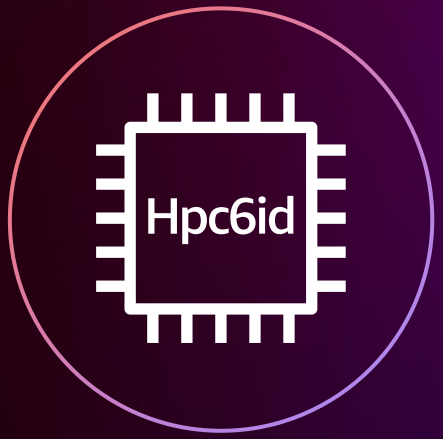
© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Amazon EC2 Hpc6id instances

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 Regions:

US East (Ohio)
GovCloud US West
Frankfurt (EU)

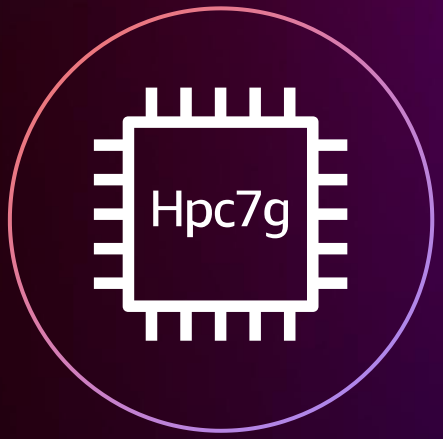
© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Amazon EC2 Hpc7g instances

Based on next generation AWS Graviton3E processors:

- 64 physical cores
- Up to 2.6GHz
- 128 GiB of RAM
- **200 Gbps EFA**

Based on next generation AWS Graviton3E processors that provide up to 35% higher vector instruction performance compared to existing instances based on AWS Graviton3 processors.



Powered by the next generation AWS Nitro System

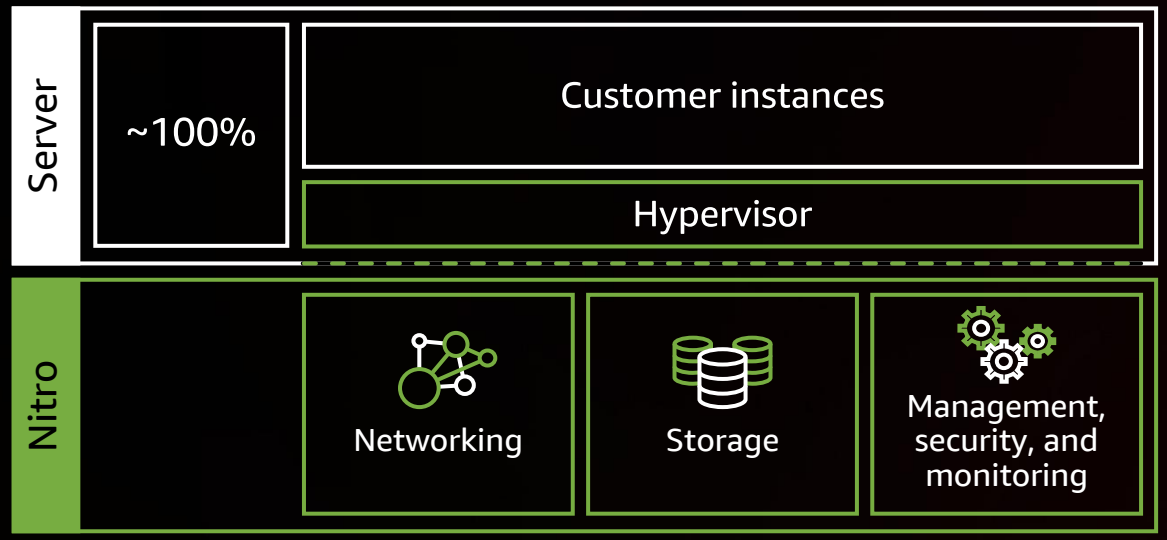
COMING SOON!

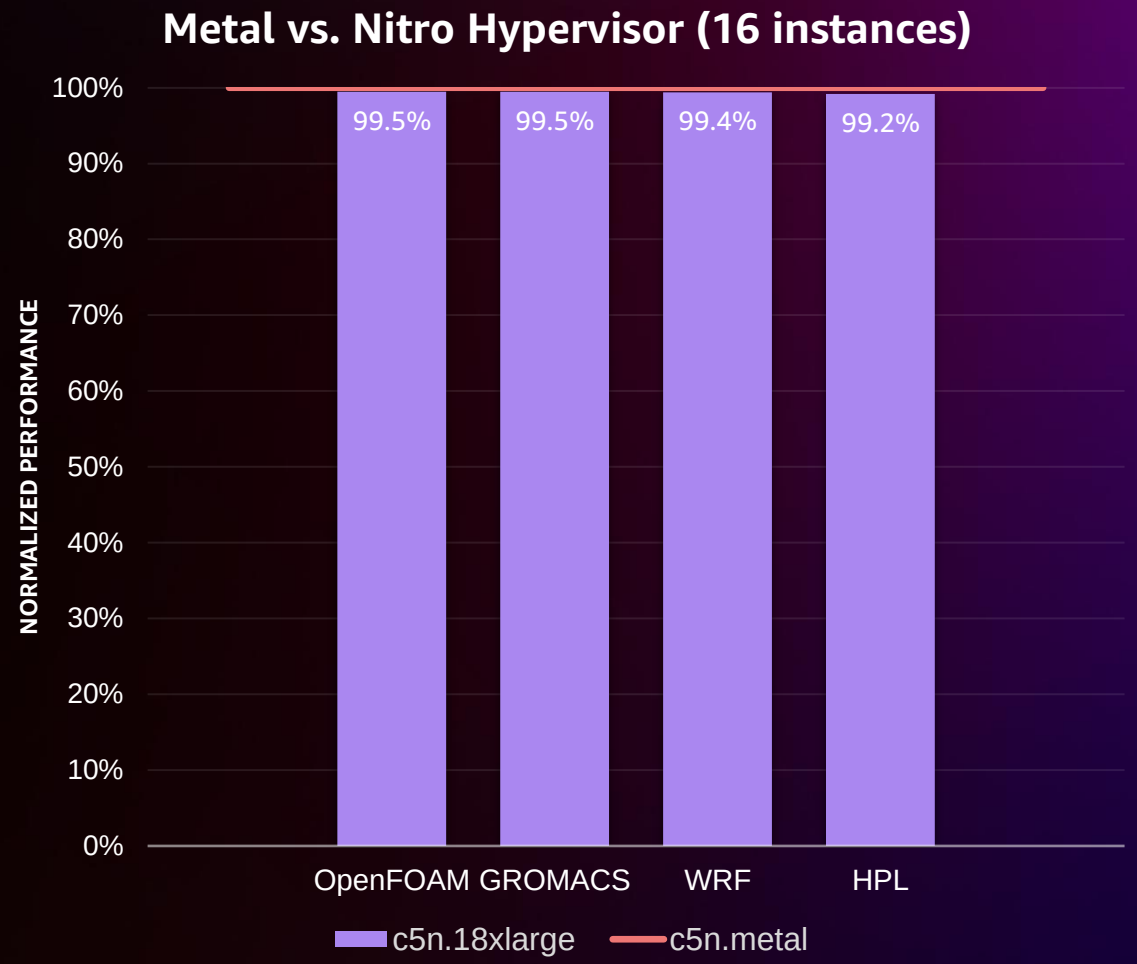
aws

© 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 (EFA)

SRD protocol



Proving myths about latency constraints wrong



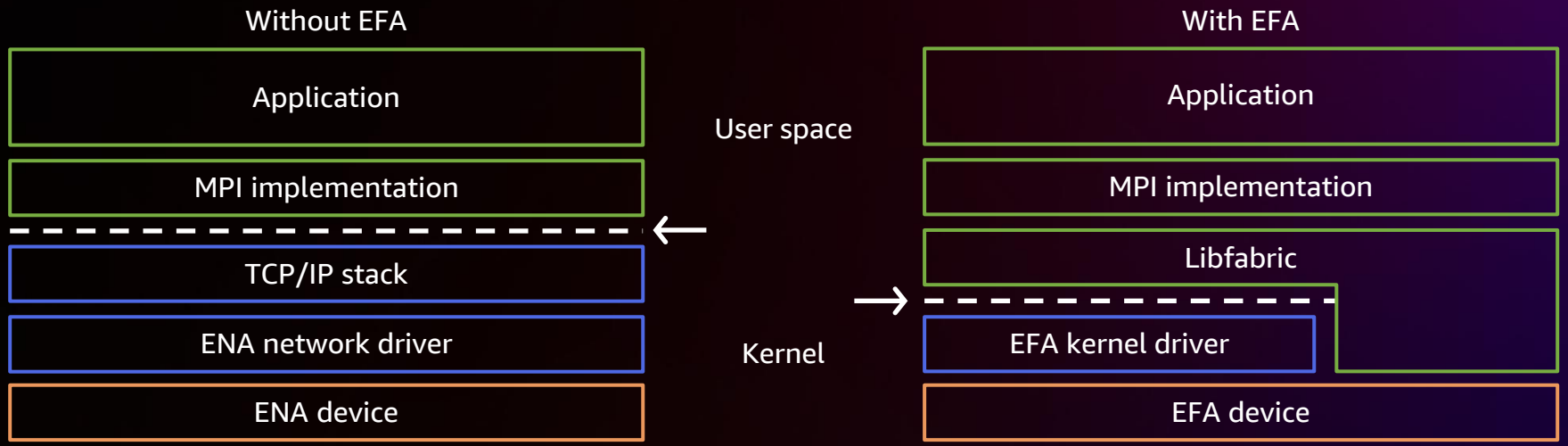
CFD



Seismic



Weather modeling



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

Amazon File Cache

HIGH-SPEED CACHE FOR DATASETS STORED ANYWHERE – ACCELERATE AND SIMPLIFY HYBRID WORKLOADS



- **Fast – access cached data** at sub-millisecond latencies and hundreds of GB/s of throughput
- **Agile – burst compute-intensive** workloads from on premises to compute resources on AWS
- **Simple – unify datasets** across S3 buckets and NFS file systems into a single namespace

---



Media and entertainment

VFX rendering/transcoding



HPC

Financial services, health and life sciences, microprocessor design, manufacturing, weather forecasting, and energy



AI/ML

ML model training and large-scale analytics

Learn more: Amazon File Cache

aws

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

Amazon FSx for Lustre

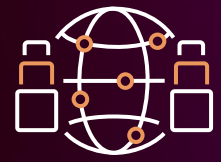
FULLY MANAGED SHARED STORAGE BUILT ON THE WORLD'S MOST POPULAR HIGH-PERFORMANCE FILE SYSTEM



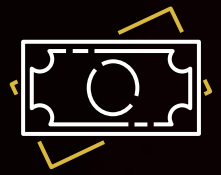
Amazon FSx for Lustre



Sub-ms latencies, **hundreds of GB/s of throughput,** millions of IOPS



Concurrent access for thousands of instances and **100,000s of cores**



Cost-optimized file systems with HDD and SSD storage options



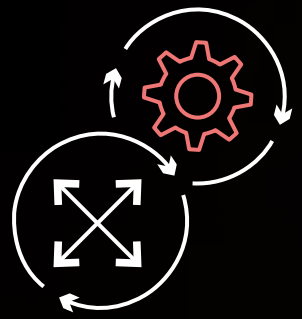
Flexible deployment options for short- and longer-term workloads

Learn more: Amazon FSx for Lustre
https://aws.amazon.com/fsx/lustre/

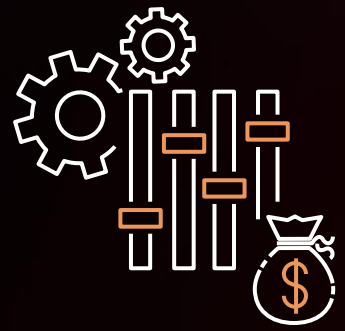
aws

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

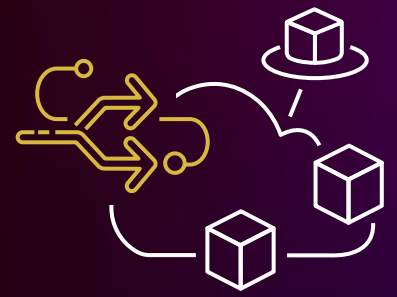
We heard from customers that running high scale, compute-intensive workloads on Kubernetes is challenging



Running and scaling batch workloads on Kubernetes is difficult and requires significant investment



Operational and cost optimizations are different for batch workloads



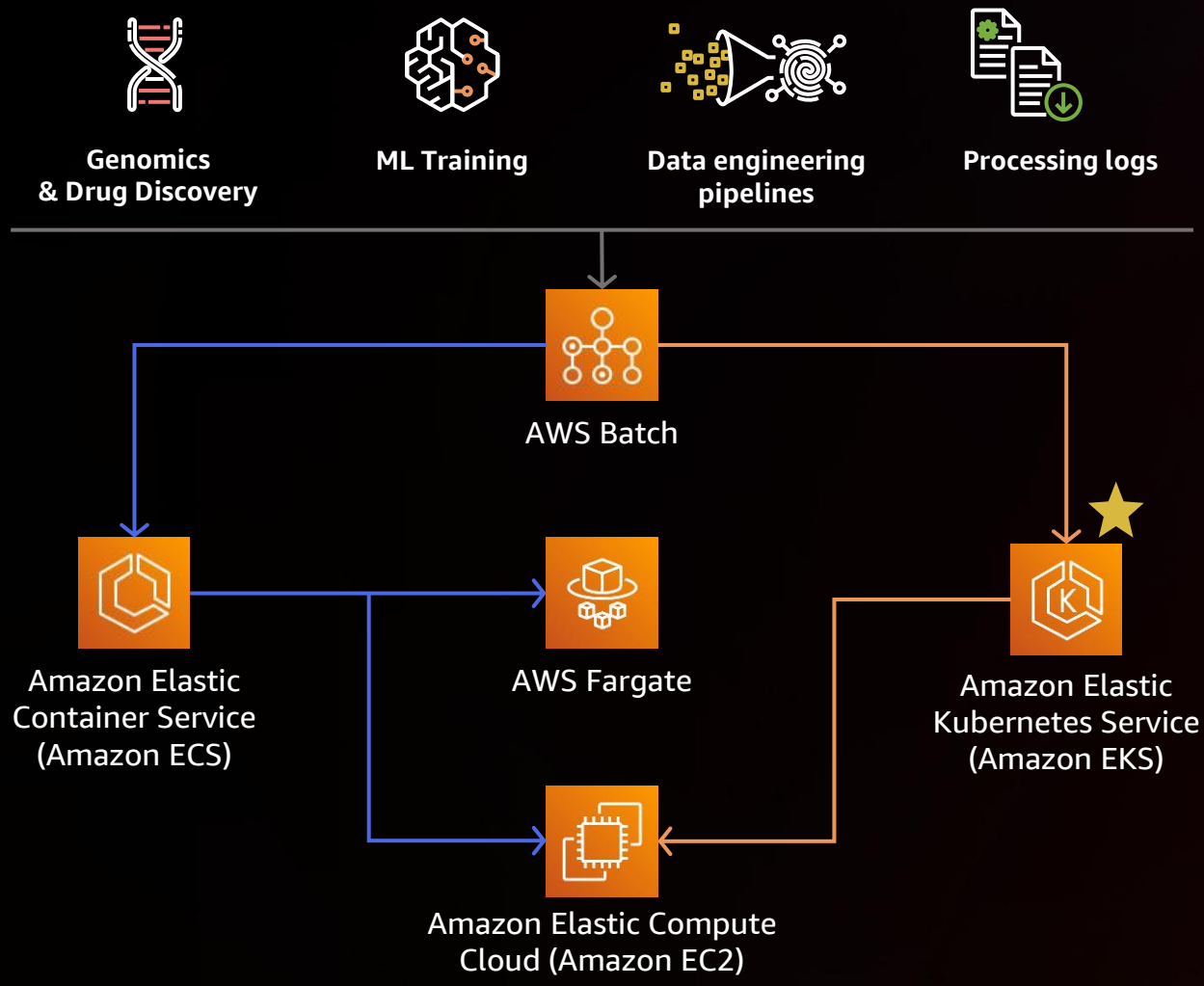
Scheduling pods quickly and efficiently is a challenge for the k8s native scheduler

All this extra work was leading to a lot of undifferentiated heavy lifting.

aws

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

AWS Batch – now with support for Amazon EKS



- **Fully-managed** service for batch workloads
- Optimized for **throughput, scalability, and cost**
- **Workload aware** cluster scaling, job scheduling and full-lifecycle management
- Works with your **existing EKS clusters**
  - Does not conflict with your other EKS workloads
  - Monitoring and governance with **daemon sets**

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

AWS ParallelCluster

One-stop shop to set up your HPC cluster



Integrated with AWS services you need

FSx for Lustre

Amazon EC2 instances

EFA

NICE DCV

AWS Batch

EnginFrame

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

HPC-scale cloud-native platform for drug discovery research

A case story of cryo-electron microscopy on AWS

Dr. Chen Su

Senior Director, Research IDS
Eli Lilly and Company

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

AWS services currently used at Lilly





Jump start of our cloud journey through **the joint ALIGN workshop** provided by AWS ProServe

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

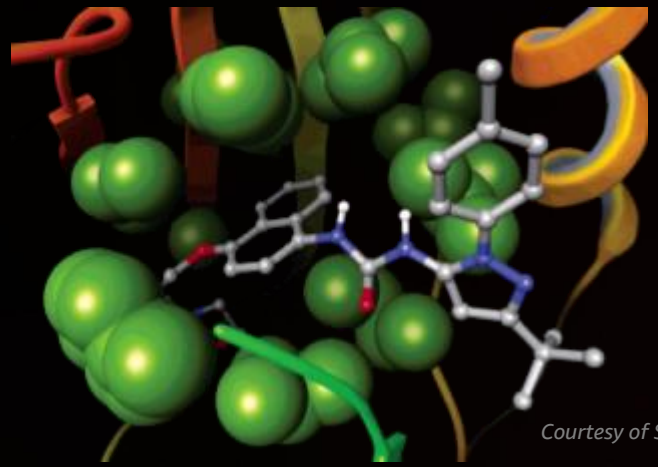
3-D protein structure is vitally important for drug designs

Target identification and selection
- Design of function
- Druggability assessment

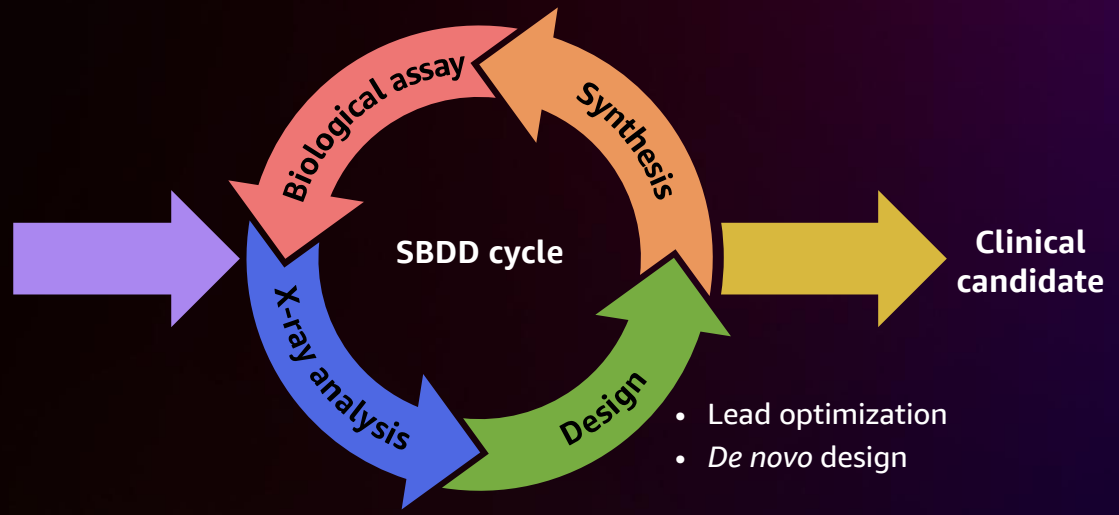
Assay design and development
- Design of constructs
- Design of mutants

Hit/lead generation
- *In silico* screening
- Library design
- *De novo* design

Lead optimization
- Analysis of SAR
- Potency
- Selectivity/ broad spectrum

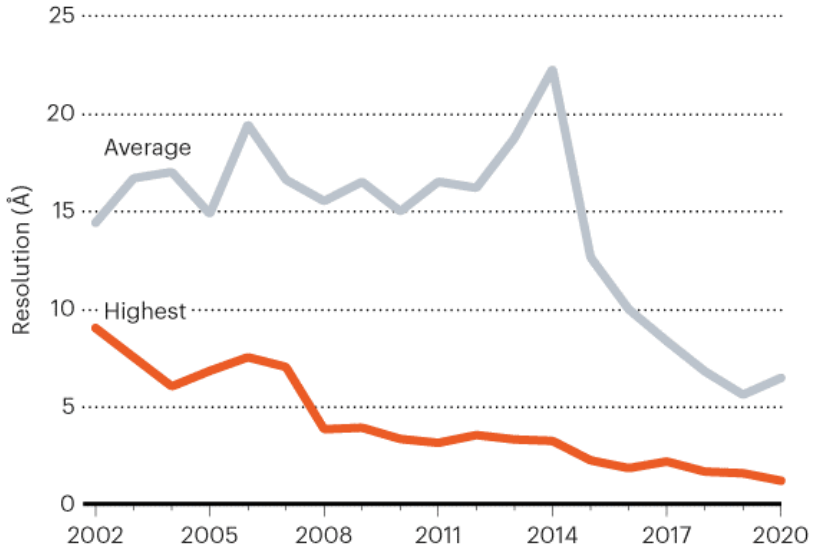


Courtesy of Schrodinger

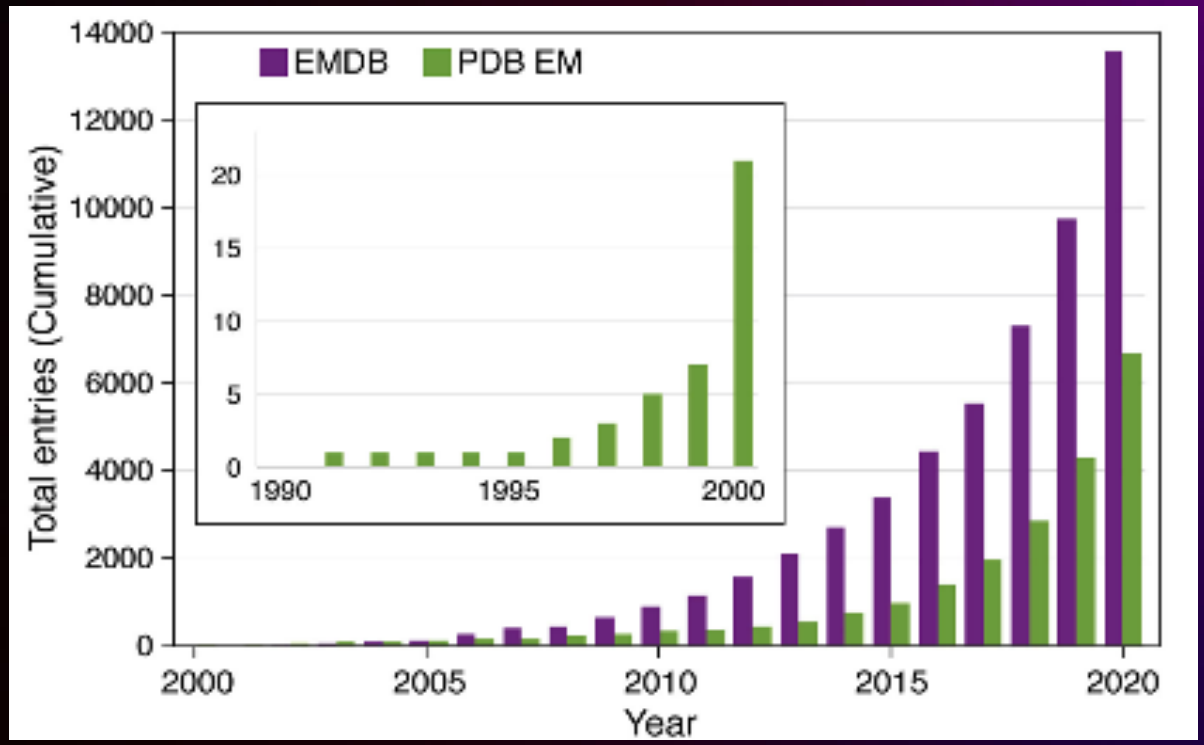


© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Cryo-EM is powerful for 3-D structure determination compared to X-ray crystallography



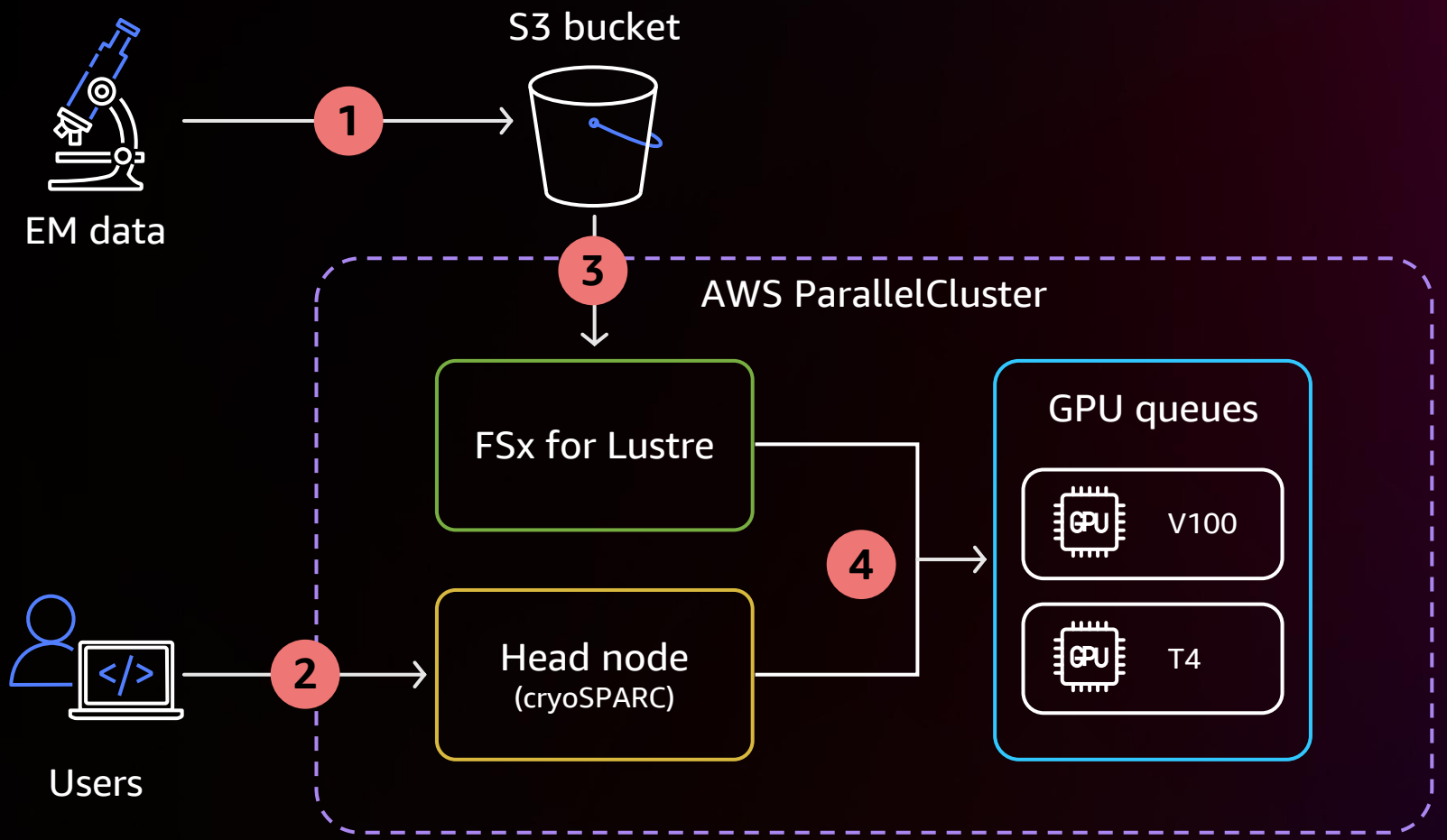
The average resolution of a cryo-EM structure has improved from 15 Å to ~6 Å and it is increasingly common to deliver protein structures in the range of 3–4 Å



- Recent and ongoing advances in cryo-Electron Microscopy have made it a powerful tool for Structure Based Drug Discovery (SBDD)
- Membrane proteins and other challenging targets, that are not readily amenable for X-ray crystallography, are being tackled by cryo-EM

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

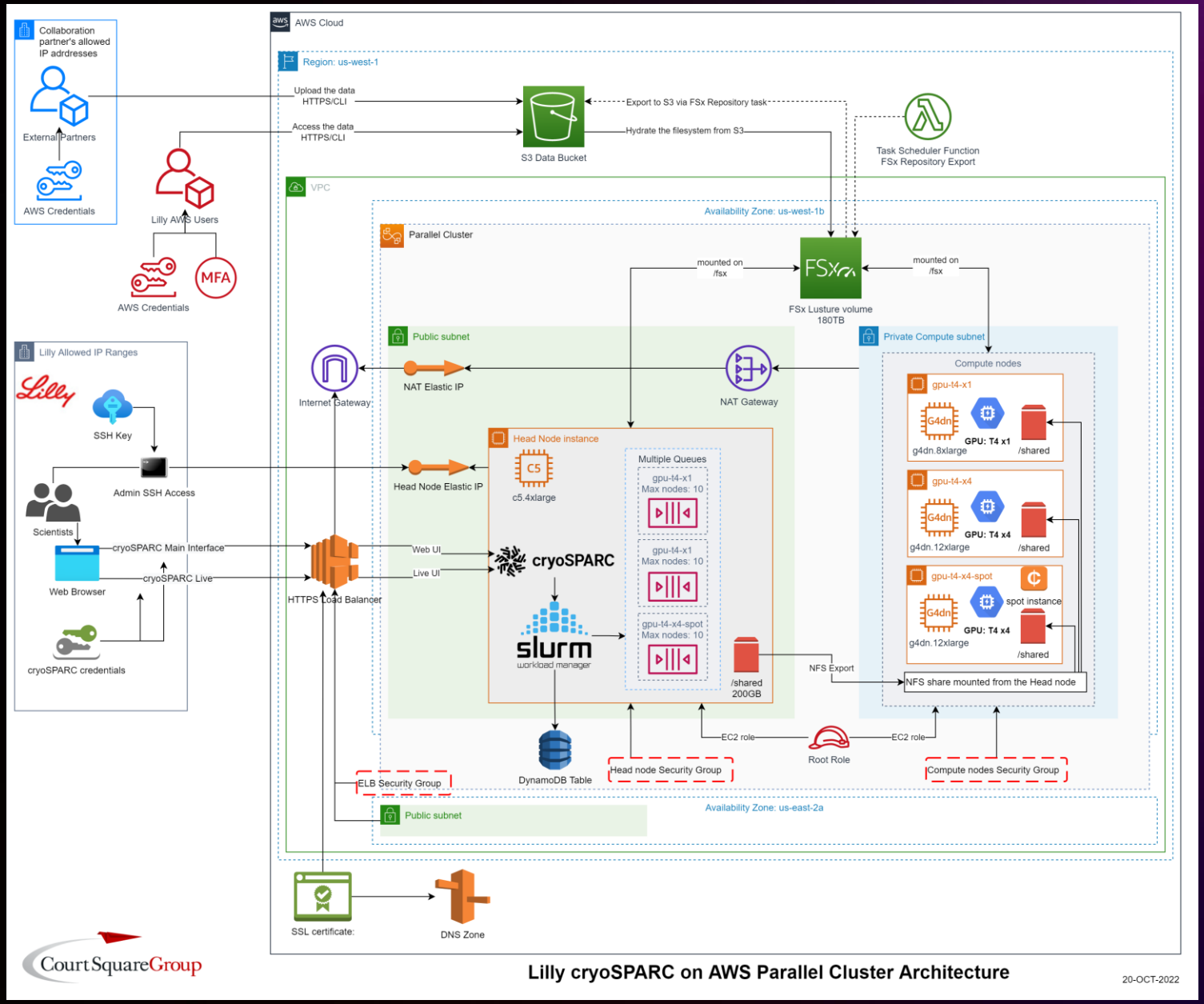
A simplified workflow of cryo-EM data processing



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

Courtesy of ThermoFisher Scientific white paper

Cryo-EM on AWS architecture



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

Use case summary

For the cloud-native cryo-EM Platform:

- Amazon S3 facilities large-scale data transfer and storage for both internal and external collaborators
- Amazon FSx for Lustre offers fast scratch space for high performance computation

A variety of Amazon EC2 GPU resources:

- High-end GPUs for better performance
- Lower-end GPUs for cost saving but with reasonable performance
- Scalability with “unlimited” GPU resources in the cloud
- Flexibility for high performance space as well as archive in Amazon S3 Glacier (yes, there are costs associated)
- Such HPC-scale cloud native platforms will be more popular in the future due to a variety of reasons

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

DTN doubles weather forecasting performance on AWS

Challenges

Global data analytics company DTN needed to efficiently increase the frequency and accuracy of its weather forecasting models to provide more-timely insights to its customers in weather-dependent industries

Solutions

The company is using a suite of AWS services, including Amazon EC2 Hpc6a instances, to run its high-resolution global forecast modeling workloads in the cloud and increase performance

Results

- Increased high-resolution model frequency from two to four runs per day
- Rendered 1 hour of forecast data in under 1 minute in test scenario
- Supports faster results to customers

> “Working on AWS brings agility to HPC. We can go from idea to production rapidly and scale in a way that’s beneficial to us and our customers.”

Brent Shaw
Chief Weather Architect and Director of Core Content Services, DTN

CUSTOMER PROFILE

DTN

INDUSTRY
Software and internet

REGION
United States and the Netherlands

DTN is a global data, analytics, and technology company that delivers unparalleled operational intelligence to help businesses prosper and organizations improve service delivery in the agriculture, energy, and other weather-dependent industries

aws

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

Dallara uses HPC on AWS to offload peak CFD workloads for race car simulations

Challenges

Italian race car manufacturer Dallara Automobili (Dallara) needed more high performance computing (HPC) for simulation and testing than what was available in its on-premises environment

Solutions

Dallara landed on Amazon Web Services (AWS) for the HPC that it needed. Using AWS, Dallara built an HPC system that met its benchmarks for performance and cost.

Results

- Built a stable infrastructure within 5 months
- Increased HPC capacity more than 3x from on premises only
- Scaled the AWS cluster by 2x
- Met a 6-month burst in customer demand

> AWS ParallelCluster is a smart, flexible tool. It helps manage the HPC, so our information technology team is not dedicated to hardware problems. We can scale on more nodes than we thought possible."

Elisa Serioli
CFD Methodology Team Leader, Dallara Automobili

CUSTOMER PROFILE

dallara

INDUSTRY
Automotive

REGION
United States and the Netherlands

aws

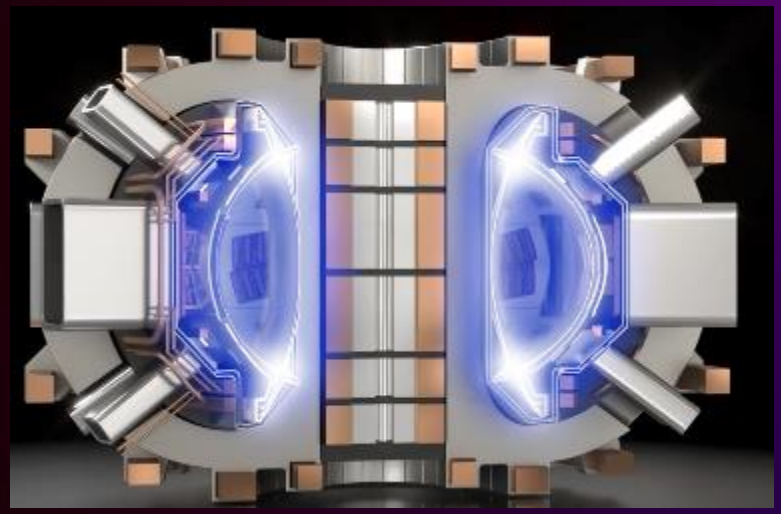
© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Commonwealth Fusion Systems (CFS)

- Ansys Fluent, EM/Maxwell, mechanical
- **10,000+** AWS compute cores to run detailed simulations
- Almost **50%** reduction in runtime
- Reduce compute costs by more than **50%**
- Using Amazon EC2 Hpc6a.48xlarge instances to massively scale up simulations

> CFS has benefited greatly from high performance computing. The new Amazon EC2 Hpc6a has been a game changer…. We've been able to increase the speed of many simulation tasks, cut runtimes approximately in half, and reduce our computing costs by over 50%. As CFS works to bring clean, limitless commercial fusion energy to the grid, we're excited to work with AWS and their HPC team."
>
> **Nate O'Farrell**
> Head of IT Infrastructure, Commonwealth Fusion Systems

Commonwealth Fusion Systems



aws

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

AWS Best HPC Cloud Platform 5 years running



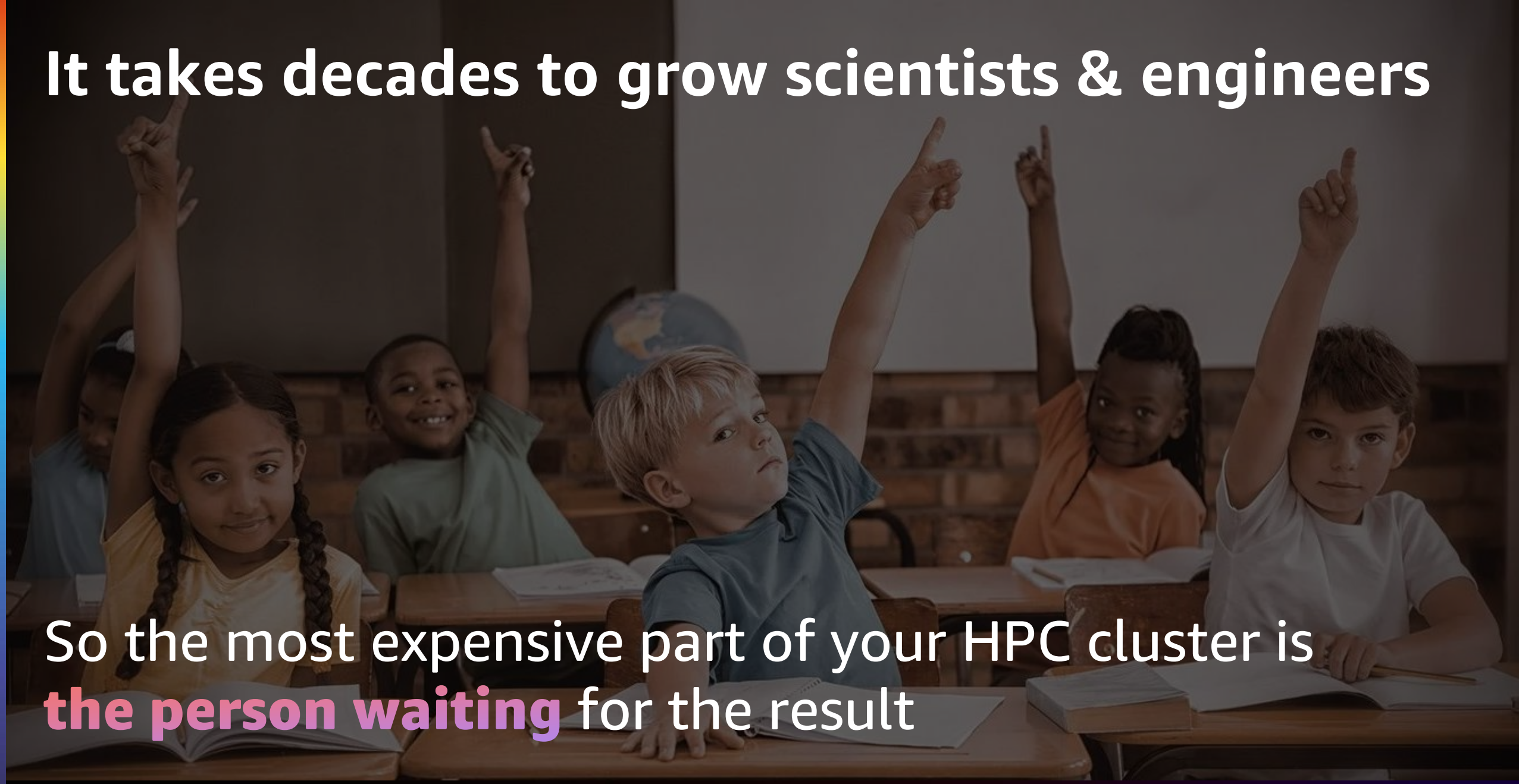






© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

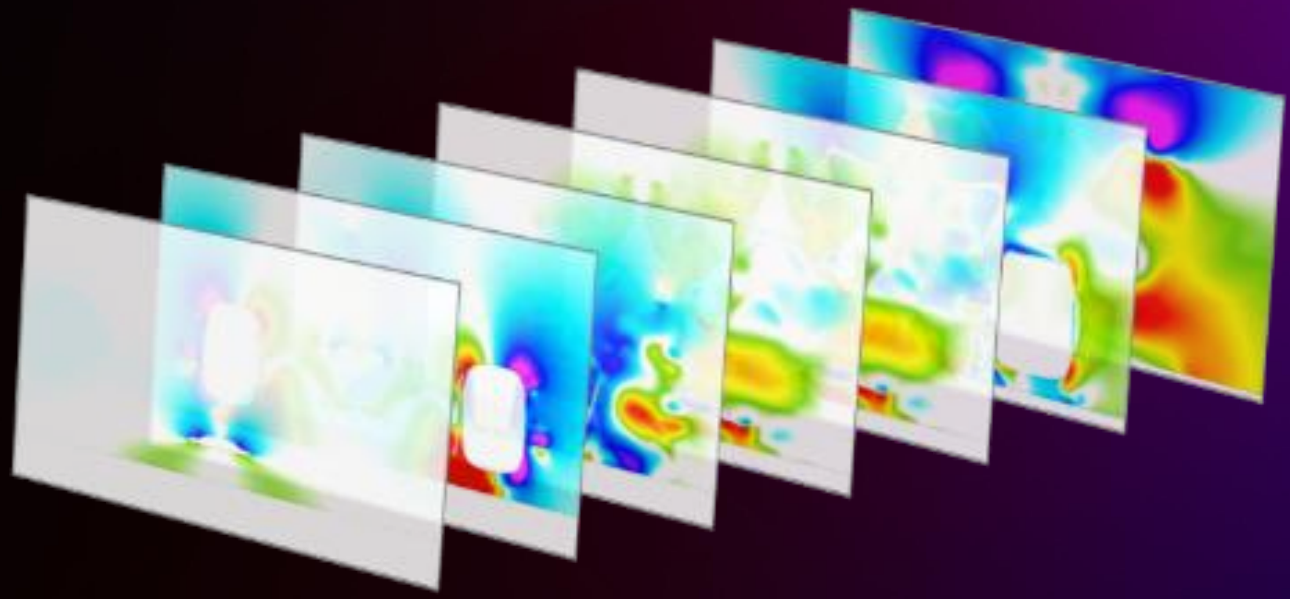
It takes decades to grow scientists & engineers



So the most expensive part of your HPC cluster is **the person waiting** for the result

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

What will you discover with HPC on AWS?



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

Looking for more HPC sessions?

Breakout

CMP205 | CMP205-OF
Solve complex problems with pay-as-you-go infrastructure

CMP220 | CMP220-OF
How AWS and aerospace and geospatial companies are rethinking HPC

CMP222
Redefining Supercomputing on AWS

Chalk Talk

CMP316
Solving problems in TV broadcasting and production with HPC and ML

CMP316
Coding for multiple CPU architectures using lessons learned in HPC

CMP326
The convergence of HPC and ML on AWS

CMP329-R | CMP329-R1
Deep dive on using AWS Batch with container-based clusters

CMP402
A day in the life of a cloud-empowered aerodynamics engineer

Workshop

CMP305
Best practices for high performance computing in the cloud

CMP335-R | CMP335-R1
Data analysis with Amazon EKS and AWS Batch

Demo

CMP003
AWS Batch on Amazon EKS: Full lifecycle job management for batch workloads

Builders

CMP407
Elastic Fabric Adapter advanced topics for AI/ML and HPC

CMP317-R | CMP317-R1
AWS Batch ask me anything

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

Thank you!



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

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