AWS re:Invent

CMP204

Innovating without infrastructure constraints

High Performance Computing on AWS

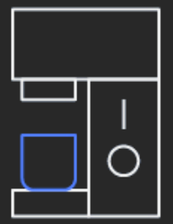
Ian Colle - General Manager, AWS Batch, HPC, and Visualization, AWS
Allison Gorman Nachtigal – Managing Director, Morgan Stanley
Barry Bolding – Director of Business Development, HPC, AWS

AWS re:Invent

aws

HPC impacts your life every day

Your morning coffee



The car you drive

The fuel you use

Knowing the weather

Your retirement portfolio

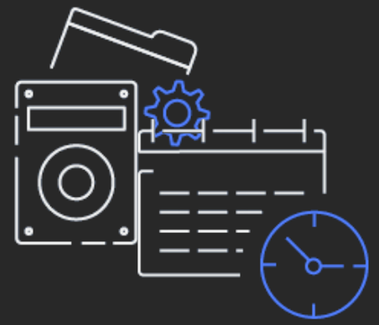
The movies you watch

The medicines you take

What do HPC practitioners and administrators want?

To do what they are really tasked with – science, engineering, research

Collaborate with colleagues. Globally. Securely.



Get results faster. Iterate if needed.

Innovate. Try new things without a cost penalty.

Working backward – the Amazonian way

Translating what our customers wanted

Technology Asks

Enhanced compute

Fast networking

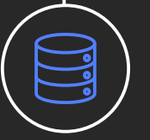
Fast and scalable storage

Operational Asks

Security, data governance

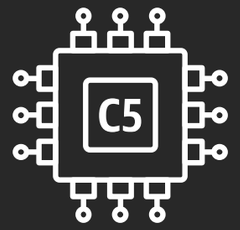
Cost management

Data transfer & management



Taking care of the technical requirements

Updated portfolio of HPC-related services



Amazon EC2
c5n Instances

Elastic Fabric
Adapter

Amazon FSx
for Lustre

AWS
ParallelCluster

Continuous technology innovation

2006 “Instance”

1.7 GHz Xeon Processor

1.75 GB of RAM

160 GB of local disk

250 Mbps network bandwidth

2019

4.0 GHz Xeon Processor

z1d instance

24 TiB of RAM

High-memory instances

60 TB of NVMe local storage

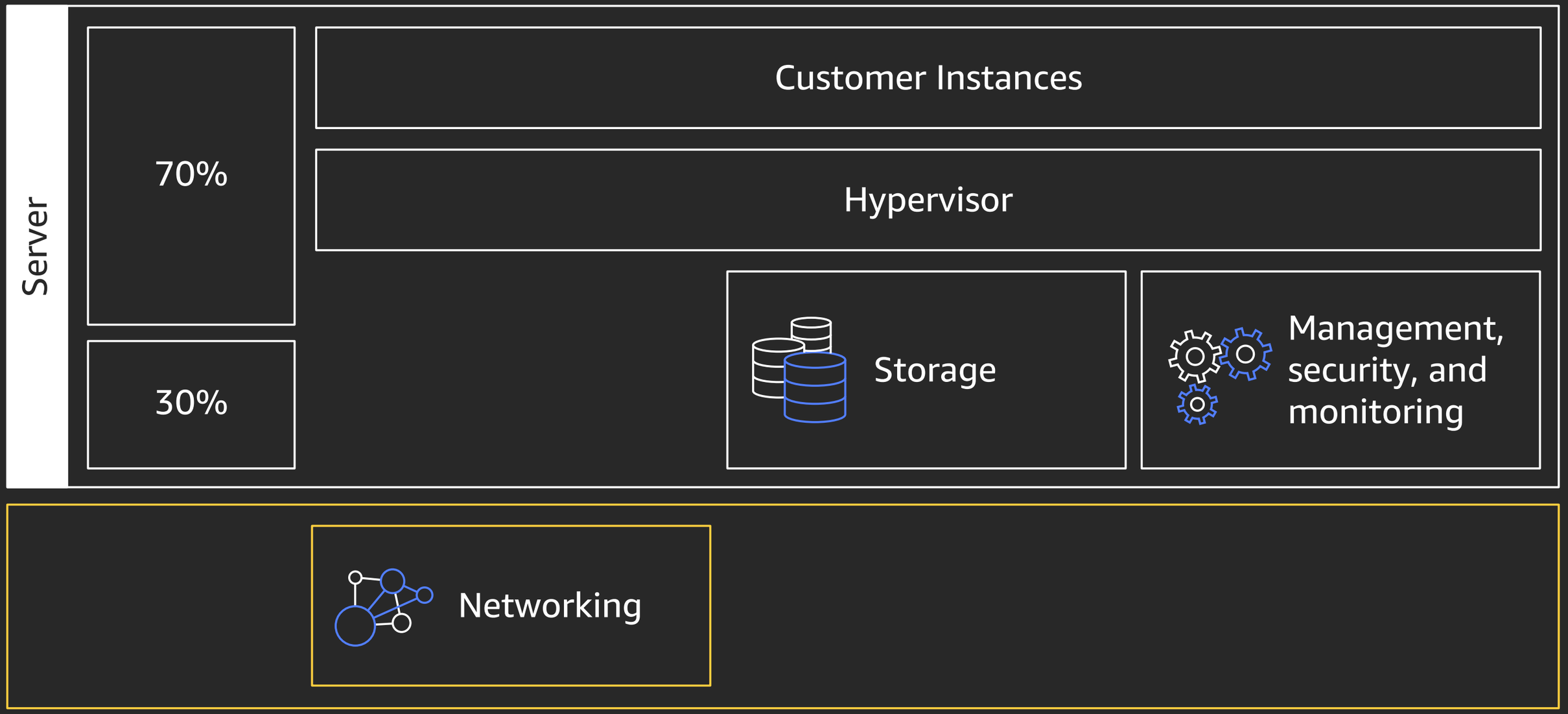
i3en.metal instances

48 TB of local disk

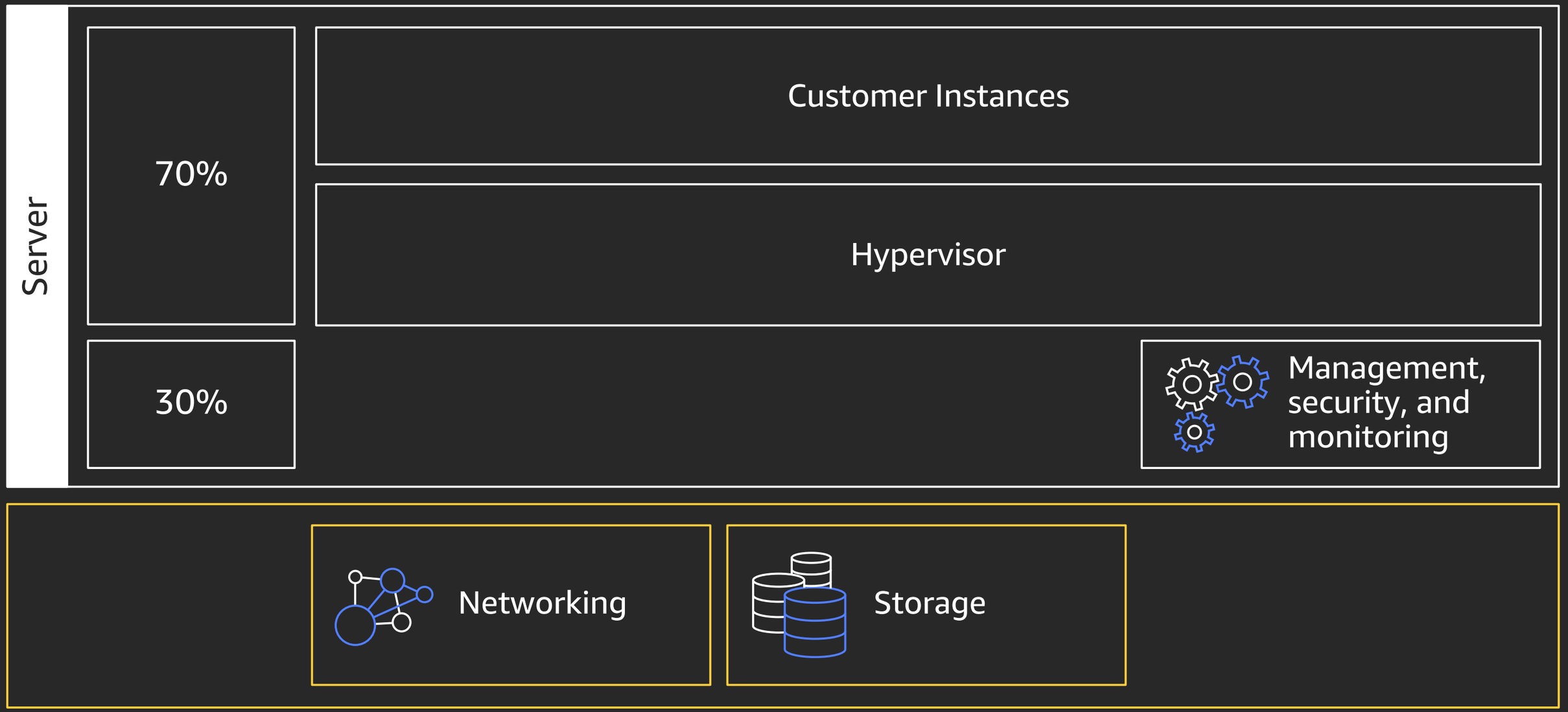
d2.8xlarge

100 Gbps network bandwidth

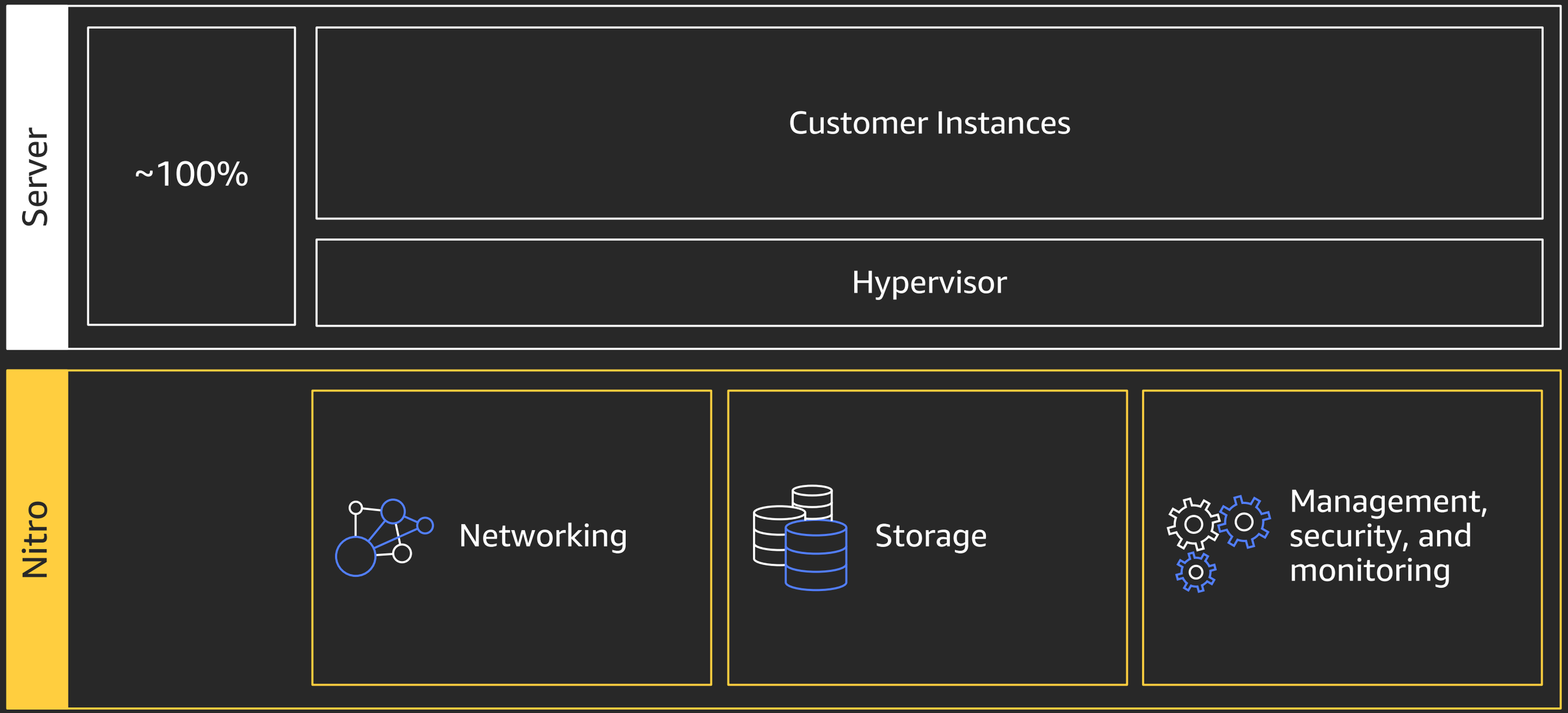
2012 EC2 “Instance” host architecture



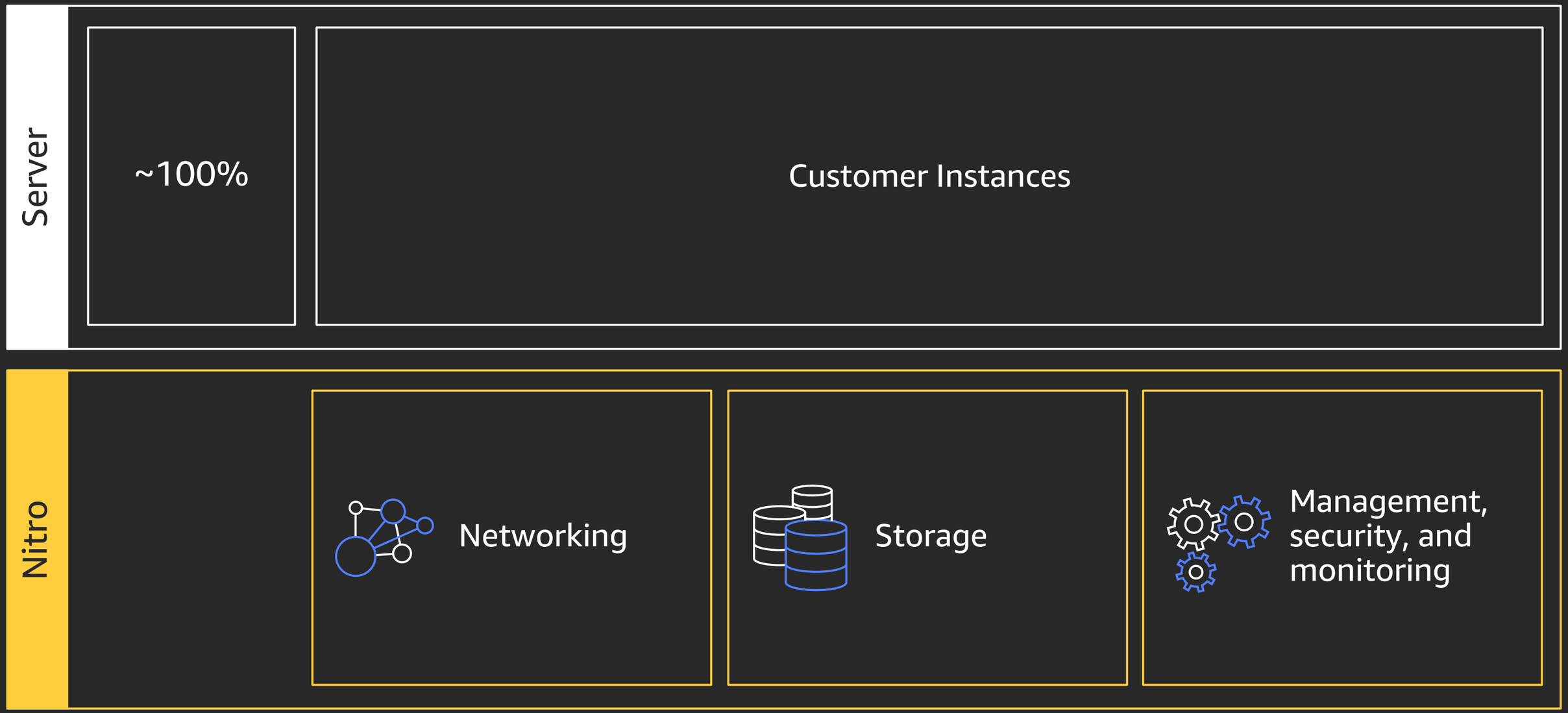
2013 EC2 “Instance” host architecture



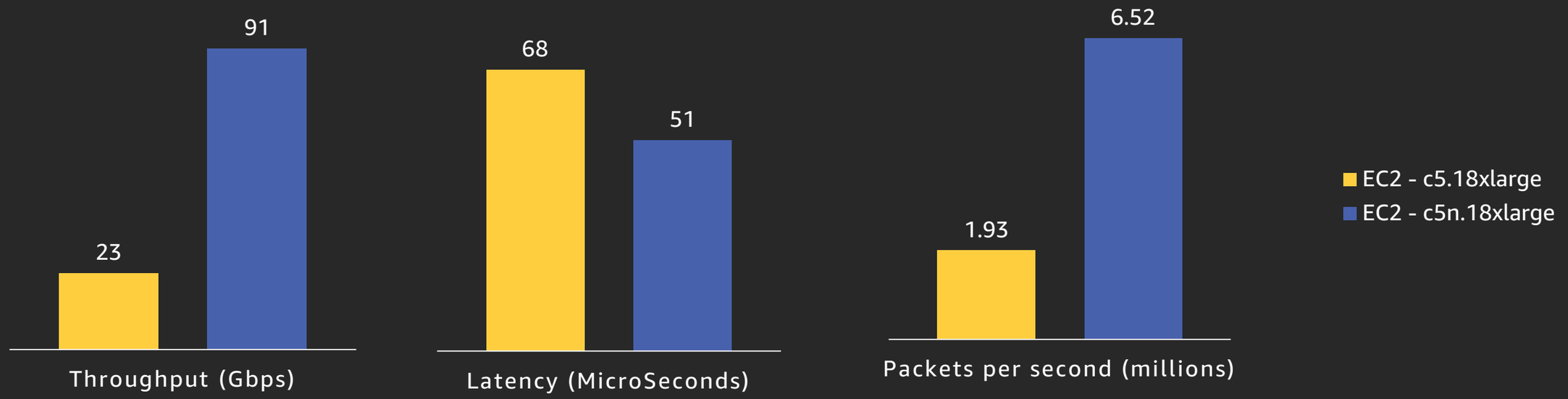
2017 Introducing Nitro Architecture



2018 Nitro enabling Bare Metal Instances



Nitro enables performance improvements



AWS network benchmarking data, March 2019

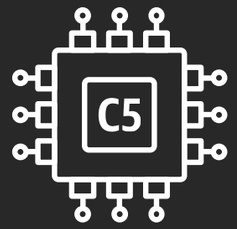
Broadest and deepest platform choice

Categories	Capabilities	Options
General purpose	Choice of processor (AWS, Intel, AMD)	Elastic Block Store
Burstable	Fast processors (up to 4.0 GHz)	Elastic Inference
Compute intensive	High memory footprint (up to 12 TiB)	
Memory intensive	Instance storage (HDD and NVMe)	
Storage (high I/O)	Accelerated computing (GPUs and FPGA)	
Dense storage	Networking (up to 100 Gbps)	
GPU compute	Bare Metal	
Graphics intensive	Size (Nano to 32xlarge)	

270+

instance types

for virtually every workload and business need



Amazon EC2 c5n Instances

Two key HPC related features

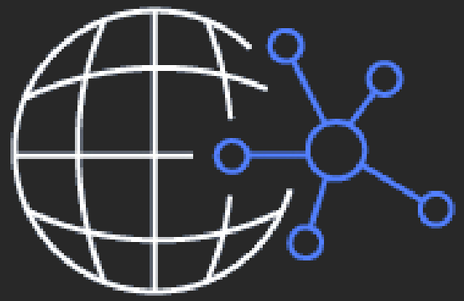
- More memory bandwidth
- 100G network throughput

Other 100G instances powered by AWS Nitro System

p3dn, i3en, m5n, r5n, g4dn

Elastic Fabric Adapter

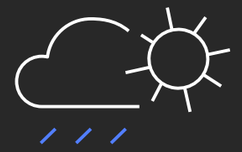
SRD protocol



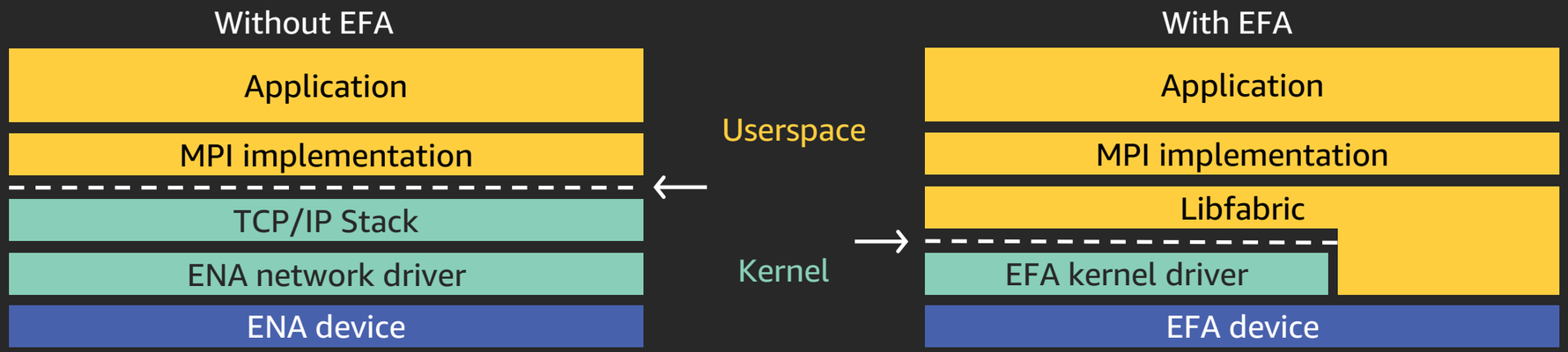
Proving myths about latency constraints wrong

CFD

Seismic



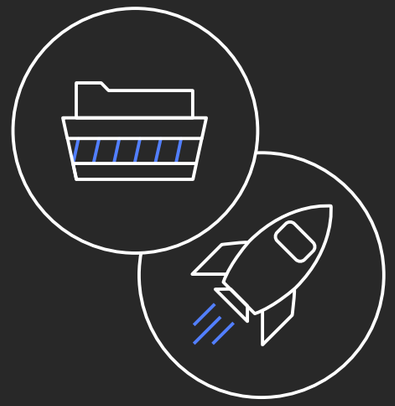
Weather modeling



MAXAR

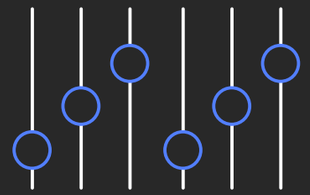
Scale tightly coupled HPC applications on AWS

Amazon FSx for Lustre



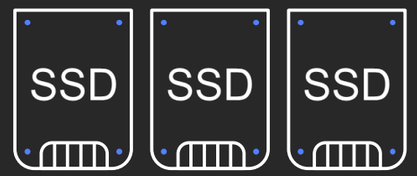
High and scalable performance

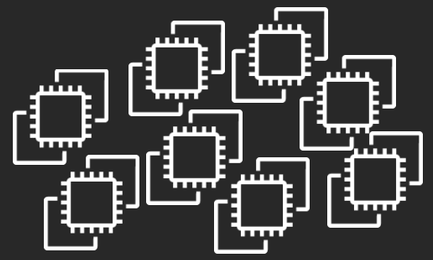
Parallel File System



100+ GiB/s throughput
Millions of IOPS
Consistent sub-millisecond latencies

SSD-based



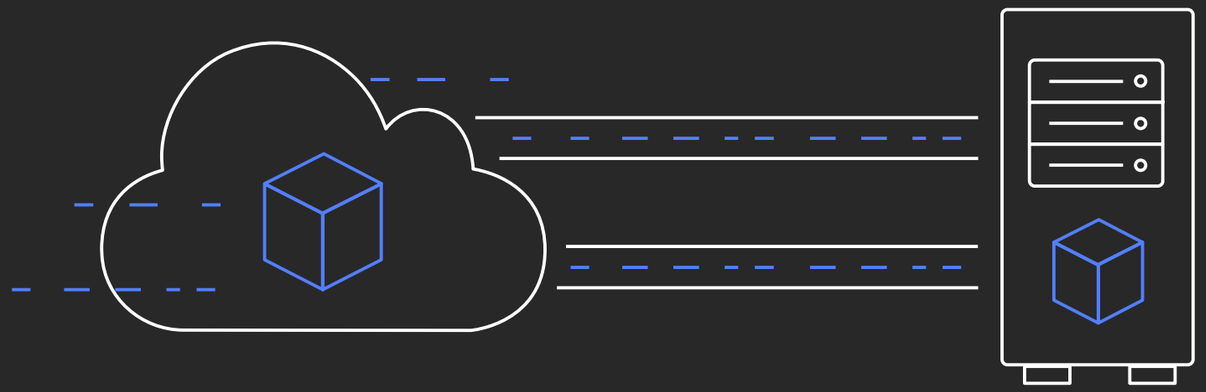


Supports concurrent access from hundreds of thousands of cores

CONDUCTOR

Conductor Technologies accelerates rendering workloads by up to 4X using Amazon FSx for Lustre

AWS ParallelCluster



One-stop shop to set up your HPC Cluster

Easy integration with AWS services

- Amazon FSx for Lustre
- Amazon EC2 Instances
- Elastic Fabric Adapter
- NICE DCV
- AWS Batch

Scale and agility: Biggest advantage in moving your HPC workloads to AWS

Scale

1000+ core workloads every day in production

Agility

Western Digital®

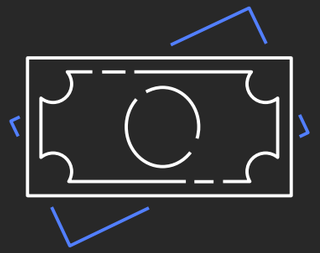
1,000,000+ core clusters commissioned on demand within hours

Taking care of the operational requirements

Native AWS Services to address operational challenges

Data security and data governance

Data transfer



Cost and cost management

Data security and data governance

Control where your data is stored and who can access it

- AWS Key Management System
- Amazon Macie

Fine-grain identity and access control so resources have the right access

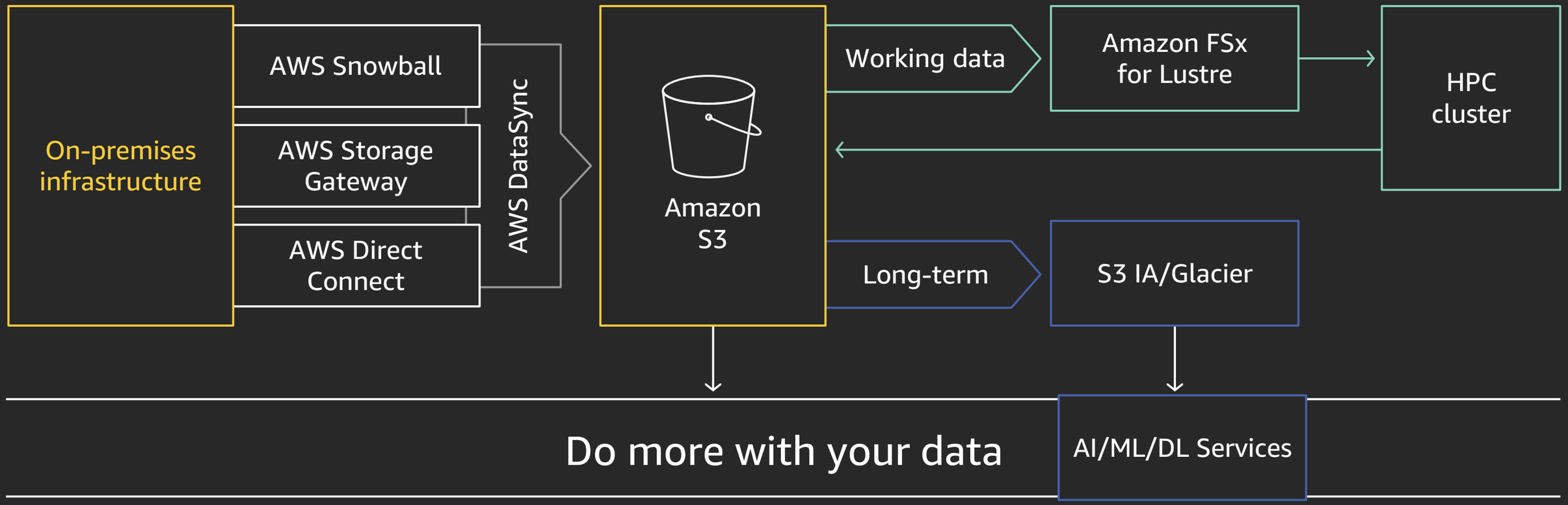
- AWS Identity and Access Manager

Reduce risk via security automation and continuous monitoring

- Amazon Inspector
- Amazon Guard Duty
- AWS Cloud Watch

Data transfer

Multiple ways to get data into the cloud and manage the data



HPC workloads with different compute and throughput characteristics

VOLKSWAGEN GROUP

Tightly coupled workloads

illumina®

Loosely coupled workloads

SCHRÖDINGER.

Accelerated computing

milk

Visualization

F1™

AI/ML

MAXAR

High-volume data analytics

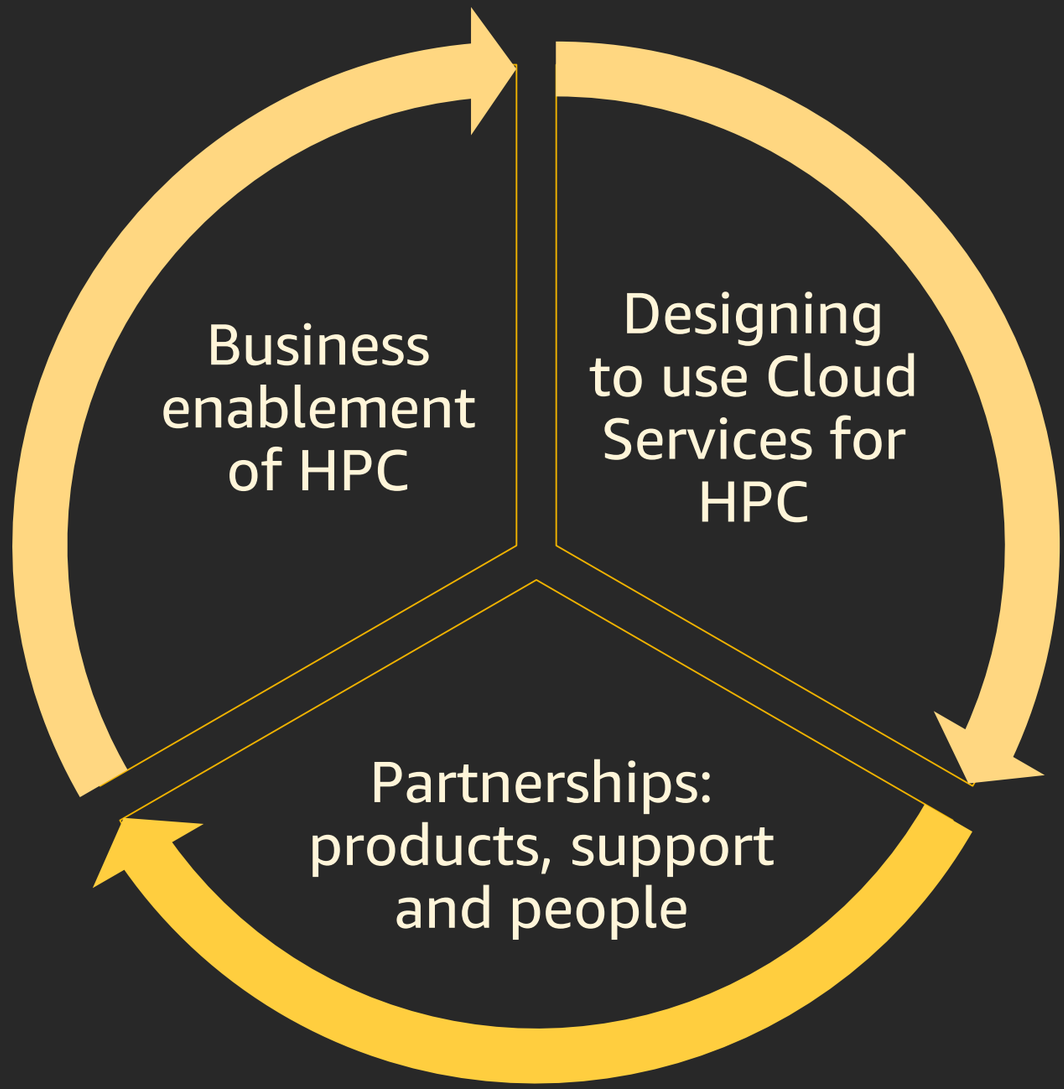
Migration of HPC to the cloud as a business enabler

Allison Nachtigal
Managing Director, Morgan Stanley

AWS re:Invent

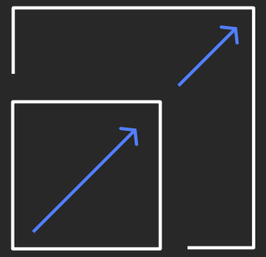
aws

Customer Spotlight: Morgan Stanley's use of HPC



Taking advantage of the cloud for HPC applications

Infrastructure choice



Scale

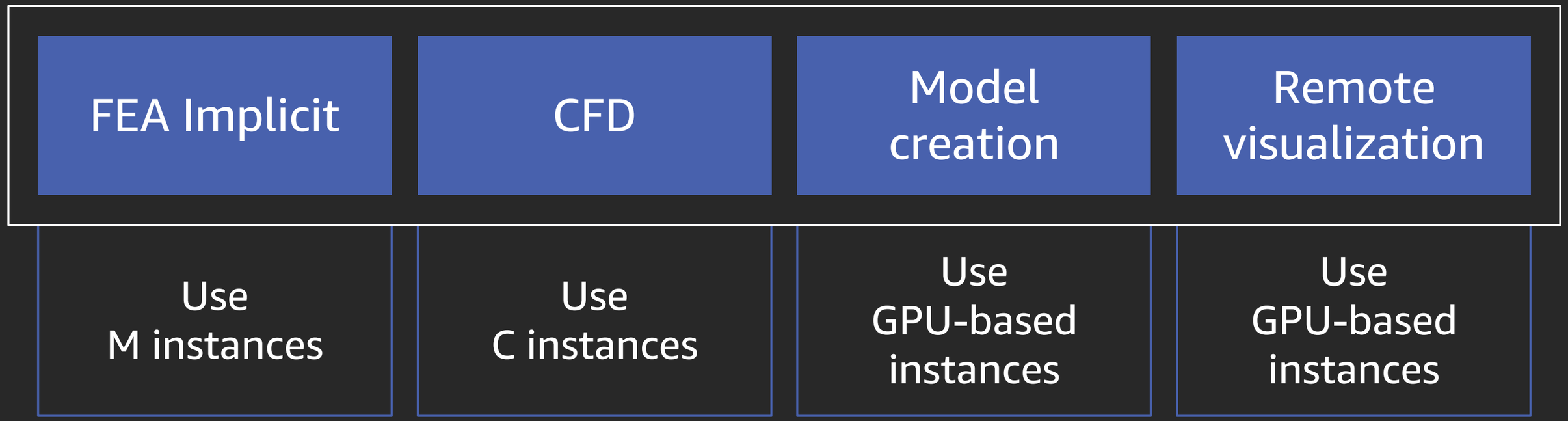
Pricing models

Agility

Rethink your workflow architecture

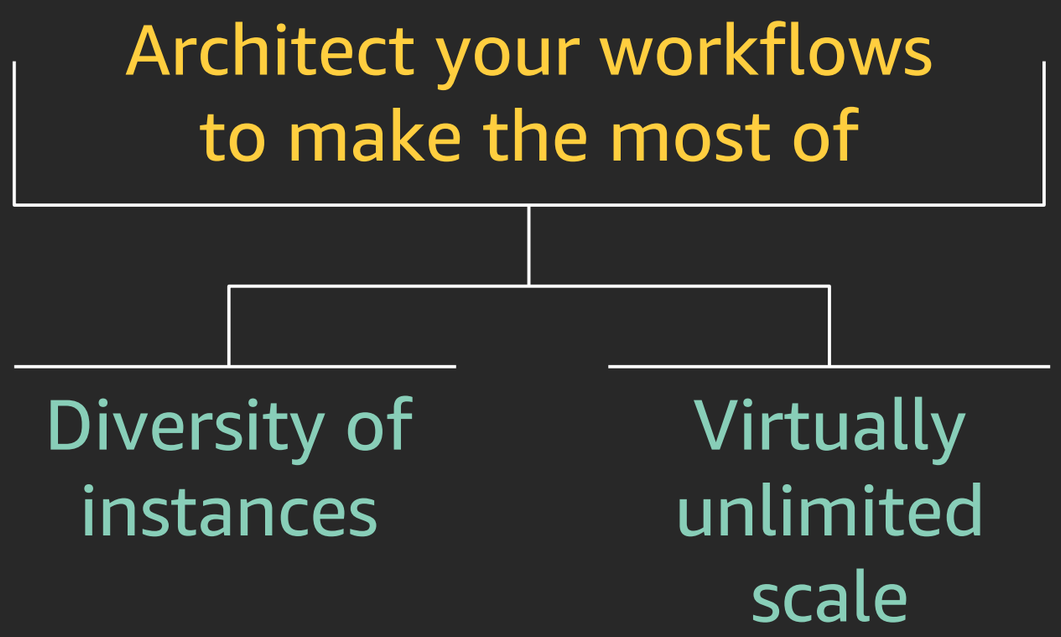
Infrastructure choice

Take advantage of the infrastructure choice available on AWS



Fit infrastructure to your application, not the other way around

Take advantage of the scale available



Western Digital®

Over 2.3 million simulation jobs on a single HPC cluster of 1 million vCPUs
— built using Amazon EC2 Spot Instances

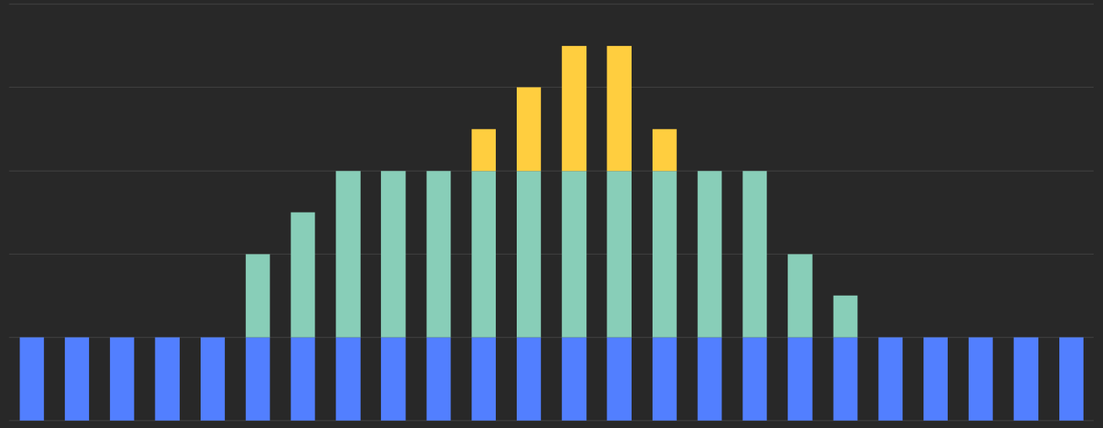
Time to results: 20 days → 8 hours

Pricing models

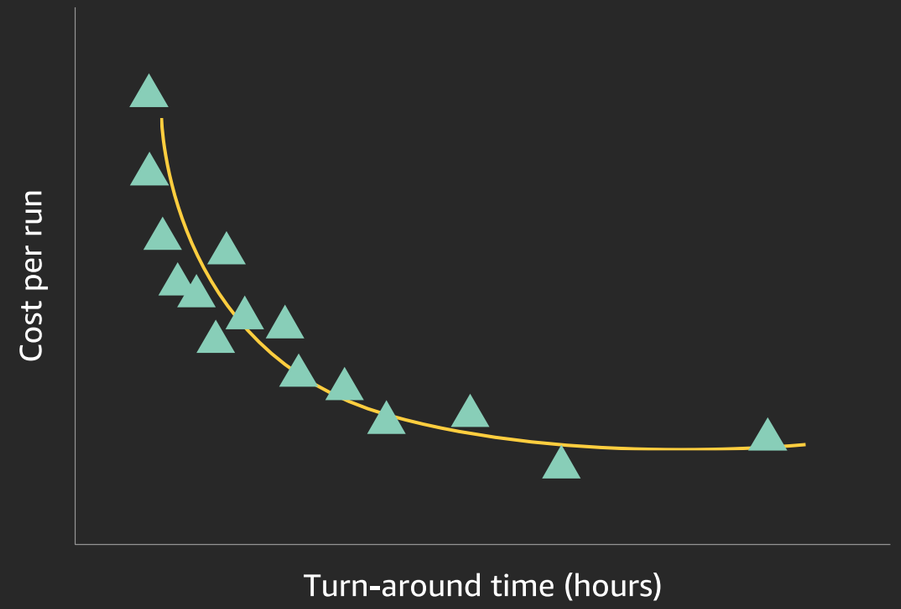
Take advantage of the different pricing models

Scale using Spot, On-Demand, or both

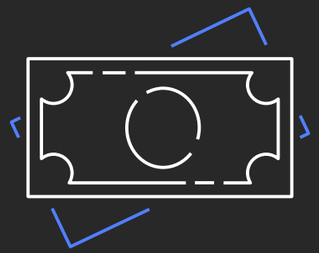
Use Reserved Instances or your Savings Plan for known/steady-state workloads



Evaluate the trade-off of time to solution vs. cost for scaling



Cost and cost management



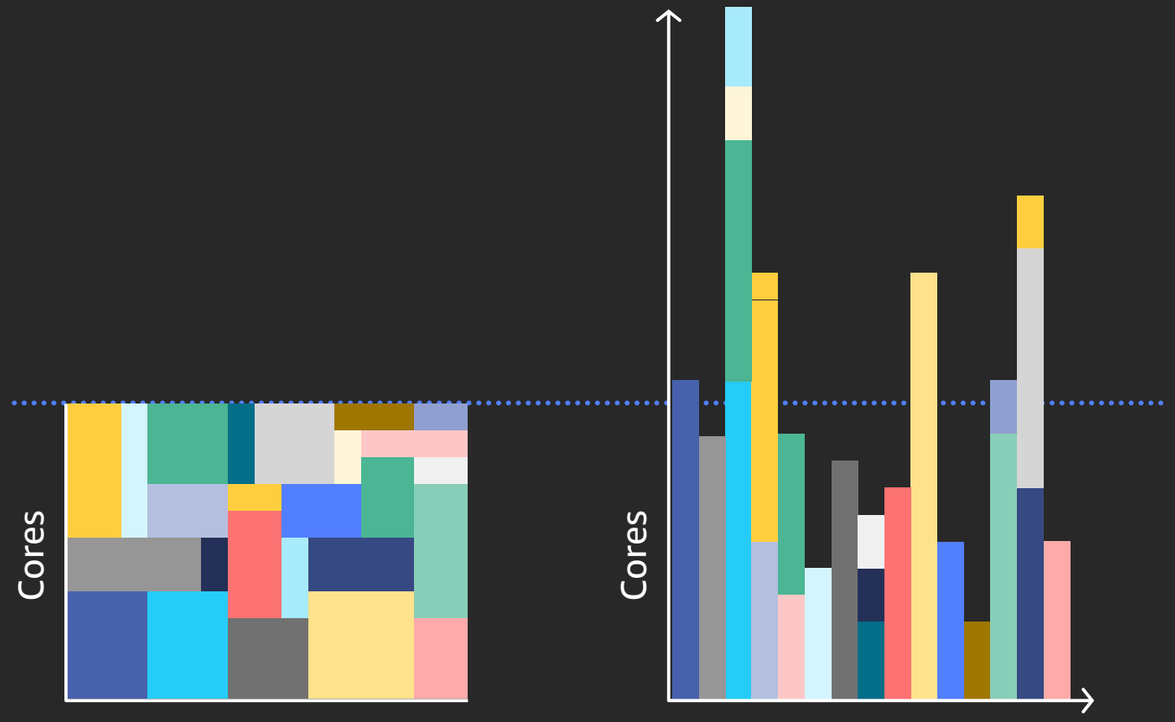
Cost

- **Compute**
  - Choose from a range of instances and three pricing options
- **Storage**
  - Multiple storage options (shared, short- term, long-term)

Cost Management

- **AWS Budgets**
  - Use SNS and SES services to receive alerts before budget limits are hit
- **AWS Cost Explorer**
  - Use tagging option in AWS ParallelCluster to allocate costs and identify spend

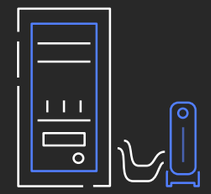
Optimize for business agility



Finite capacity, usually with long queues to wait in

Massive capacity when needed to speed up time to results, and agile environment when additional hardware and software experimentation is needed

72.8% of organizations that use HPC reported delayed or cancelled HPC jobs*



Outdated technology

Almost 20% of the useful life of new technology/hardware lost in the procurement process

HPC workloads across industries

FRED HUTCH
CURES START HERE™

Life Sciences

FINRA

Financial Services

Oil & Gas

Design & Engineering

MAXAR

Climate & Earth Sciences

drive.ai

Autonomous Vehicles

AWS expediting cures for cancer

FRED HUTCH
CURES START HERE™

Fred Hutch decreased the time it took to analyze 10,000 biological samples from 7 years to 7 days

By accelerating our research processes on AWS, we can get closer to developing therapeutics to fight cancer.

— Sam Minot, PhD
Staff Scientist

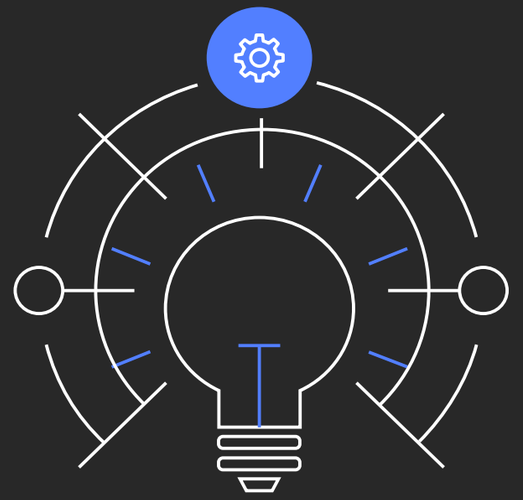
HPC Wire: Best HPC Cloud Platform

High Performance Computing on AWS

From worrying about

Capex Capacity Technology

to



Focusing on innovation

A fundamental rethink of what is possible

Thank you!

AWS re:Invent

aws

Please complete the session survey in the mobile app.