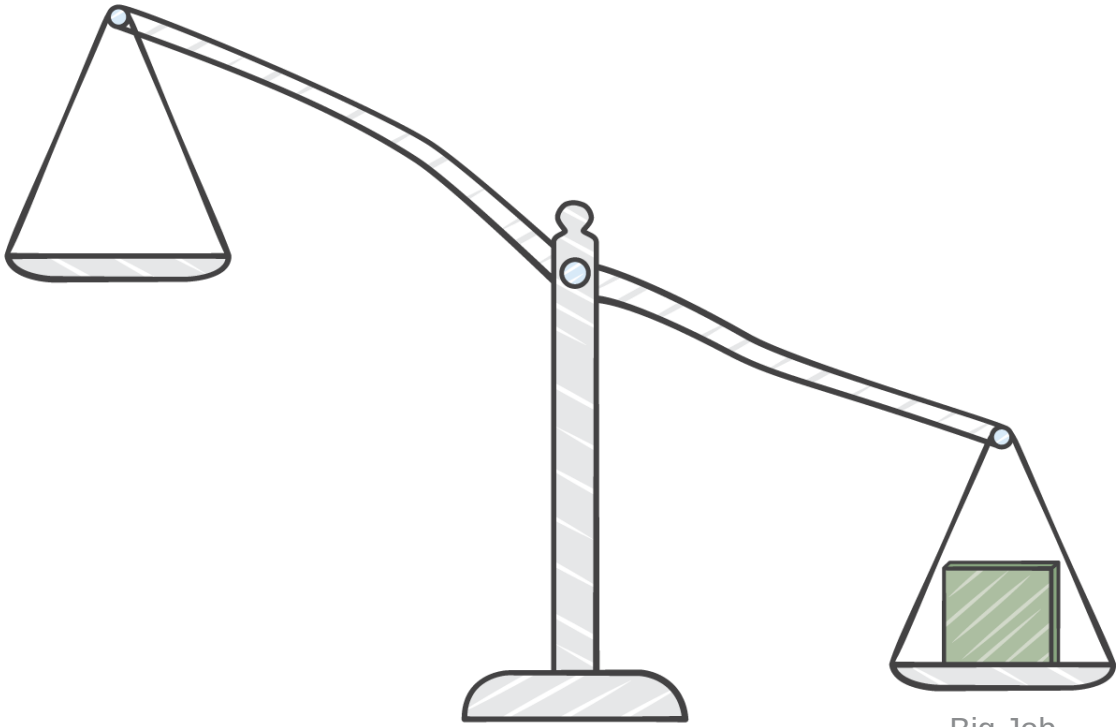
High Performance Computing on AWS

Wen DAI (代闻)
Solutions Architect
Amazon Web Services
wendai@amazon.com

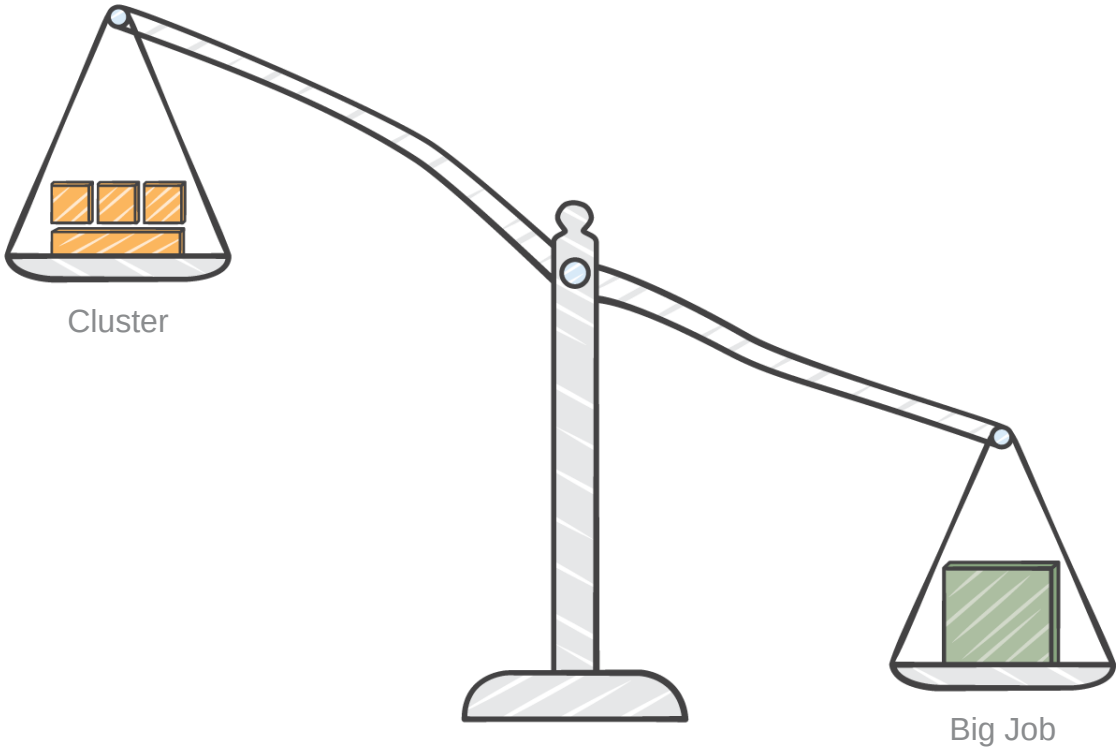
amazon web services

Take a typical big computation task…

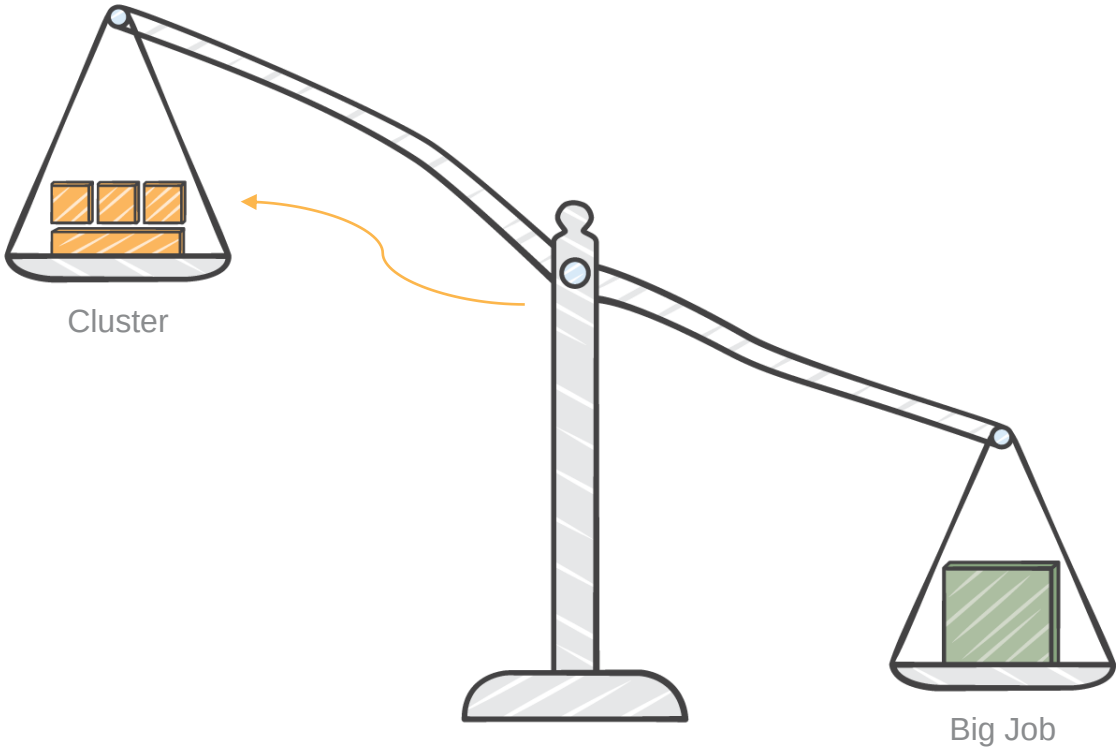


Big Job

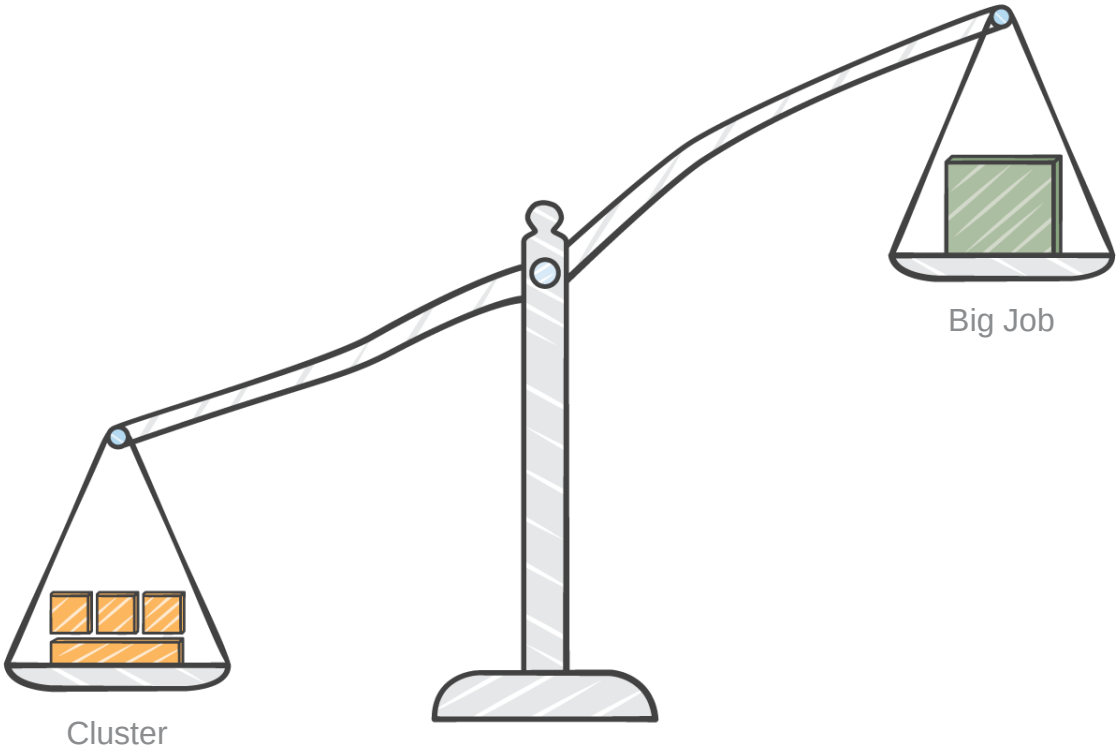
…that an average cluster is too small (or simply takes too long to complete)…



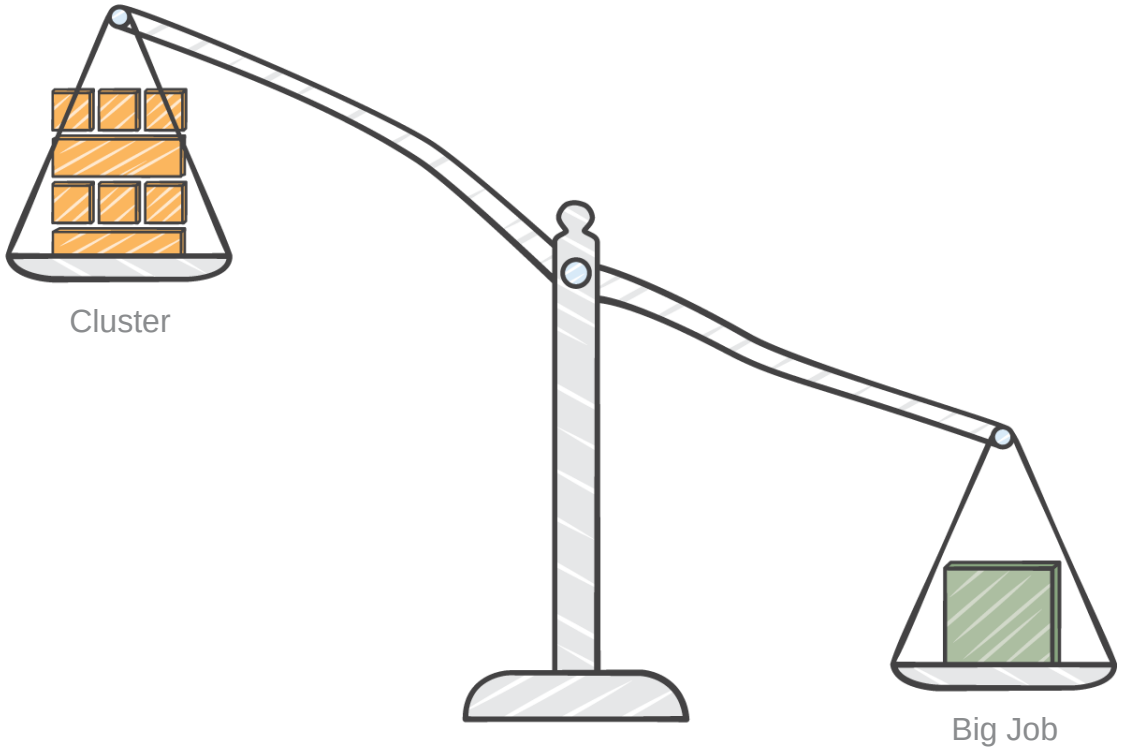
…optimization of algorithms can give some leverage…



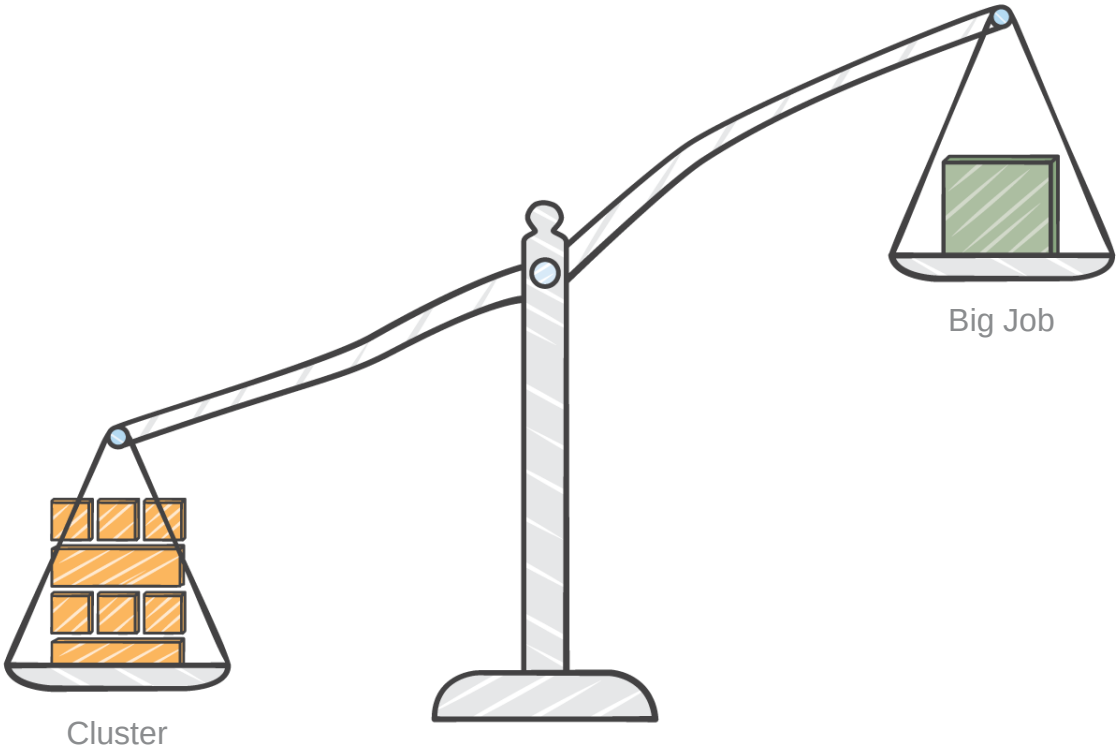
…and complete the task in hand…



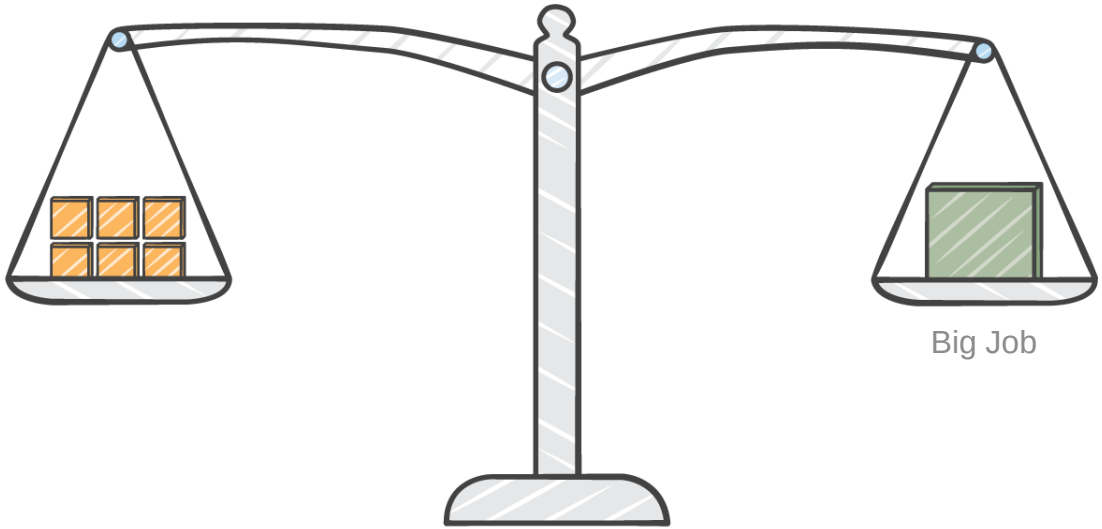
Applying a large cluster…



…can sometimes be overkill and too expensive

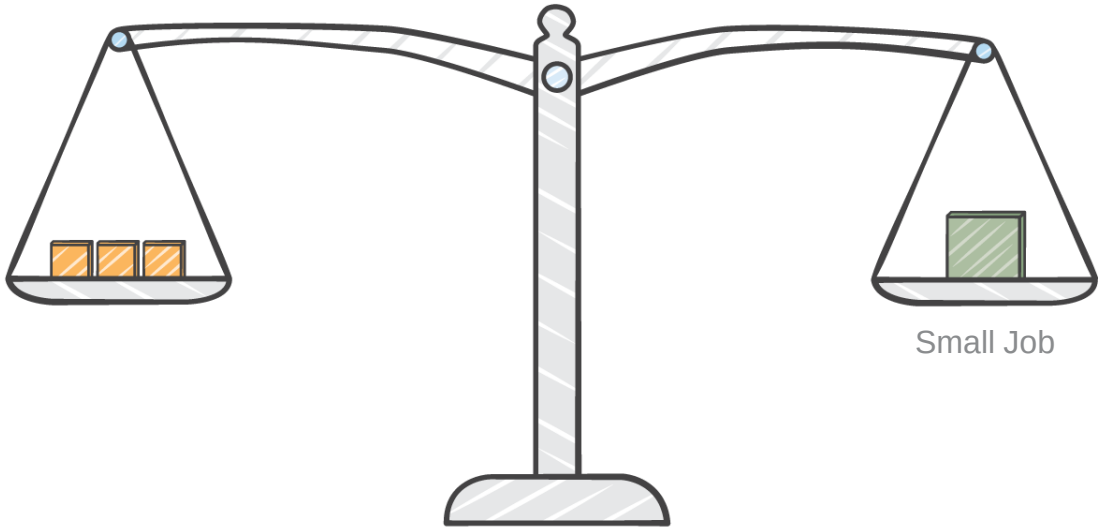


AWS instance clusters can be balanced to the job in hand…

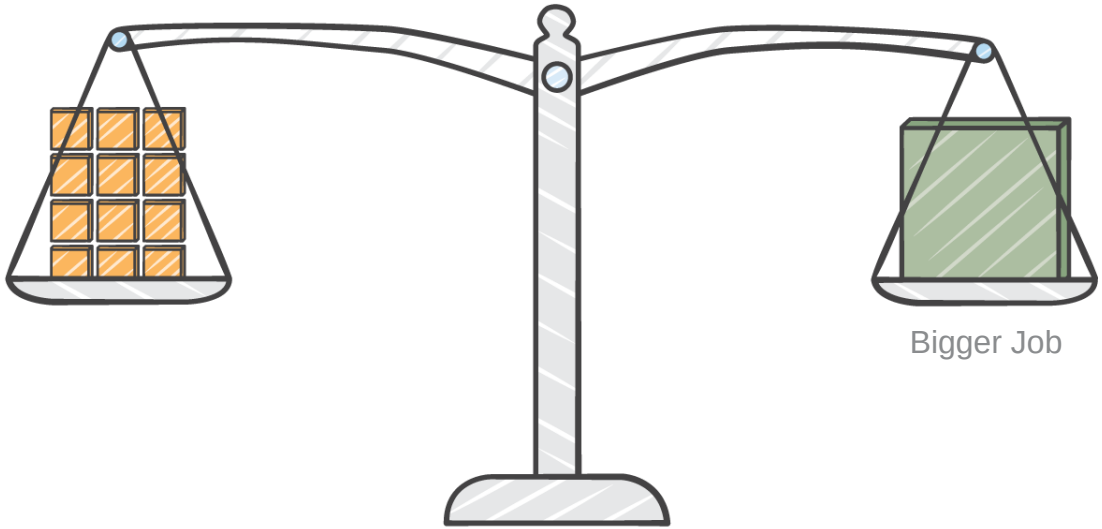


amazon web services

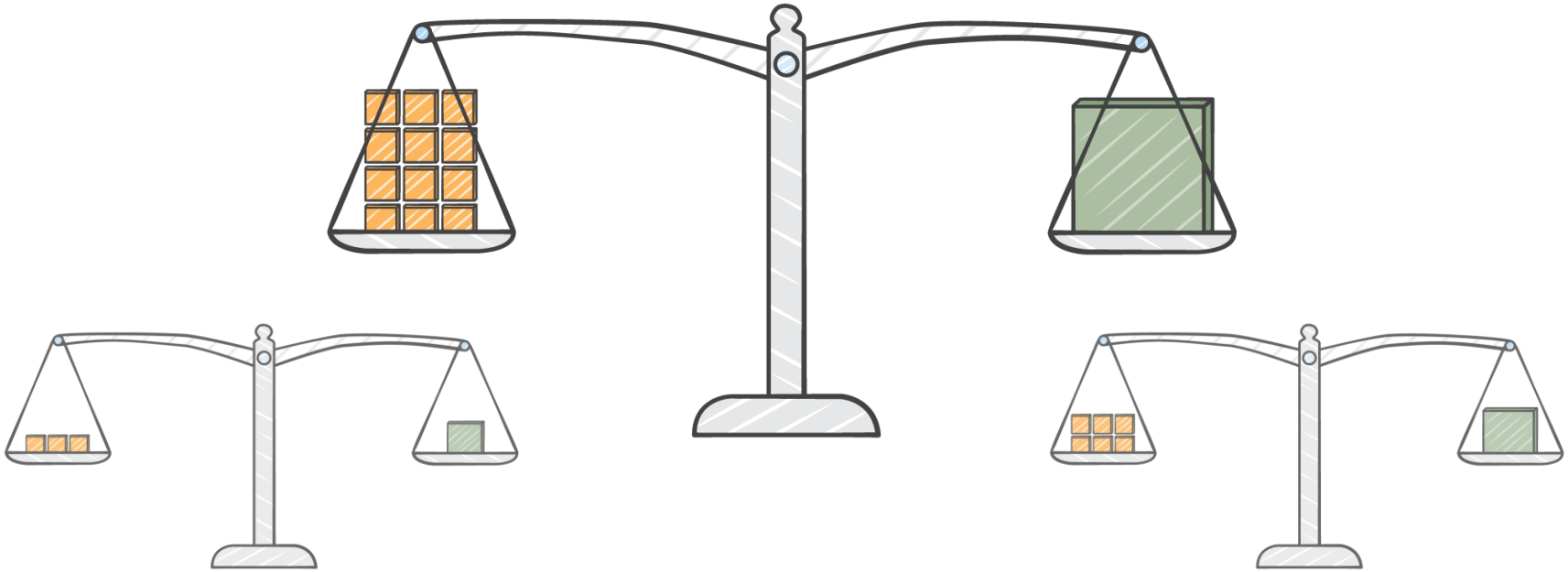
…nor too large…



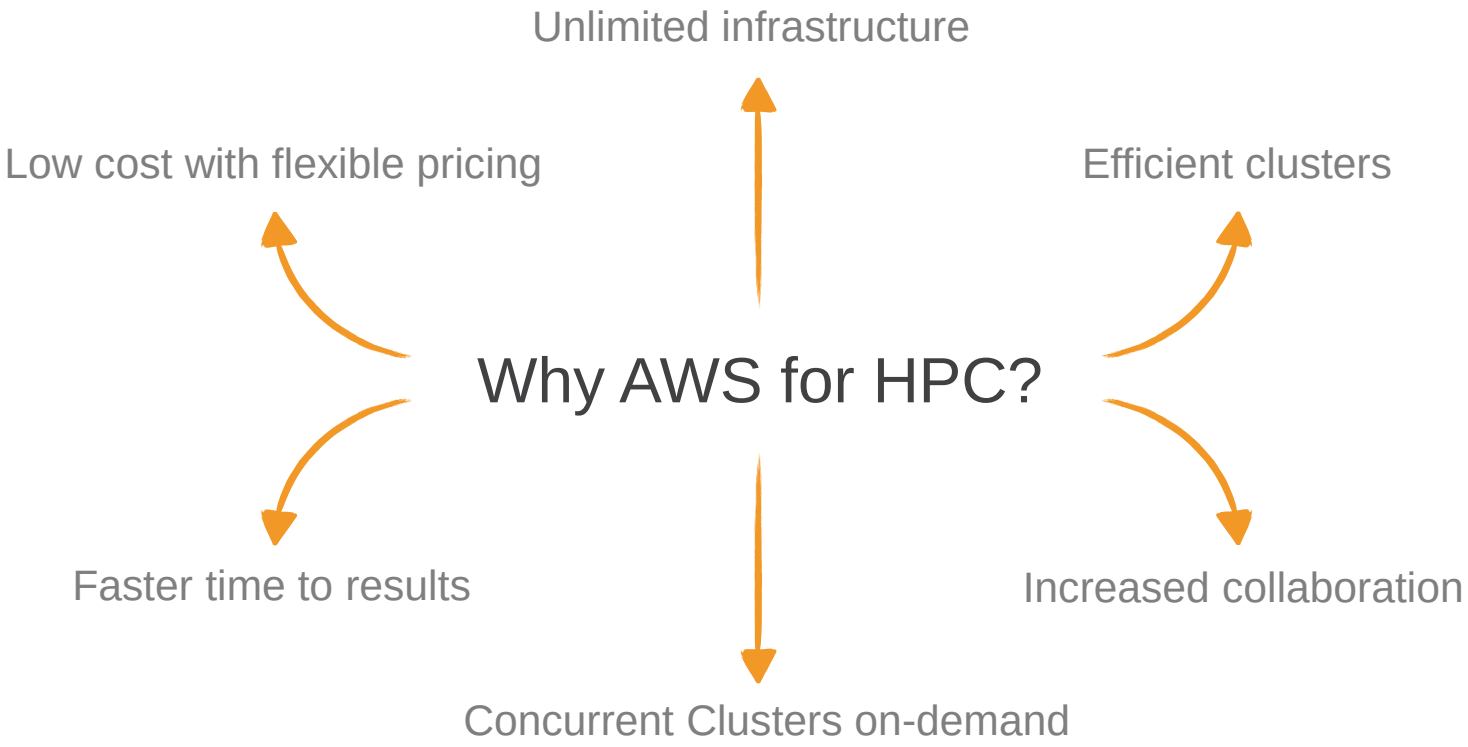
…nor too small…



…with multiple clusters running at the same time



amazon web services



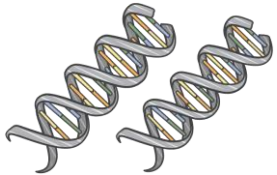
amazon web services

Customers running HPC workloads on AWS

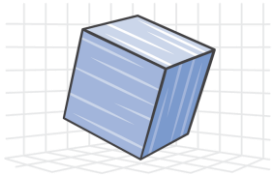
amazon web services

Popular HPC workloads on AWS

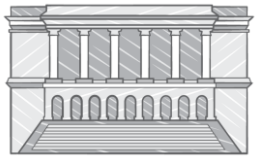
Genome processing



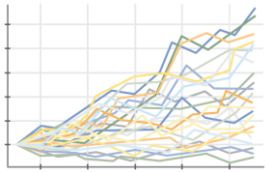
Modeling and Simulation



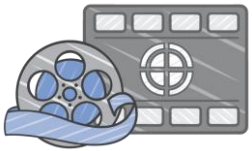
Government and Educational Research



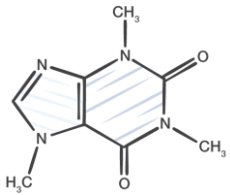
Monte Carlo Simulations



Transcoding and Encoding



Computational Chemistry



amazon web services

Across several key industry verticals

Utilities

Biopharma

Materials Design

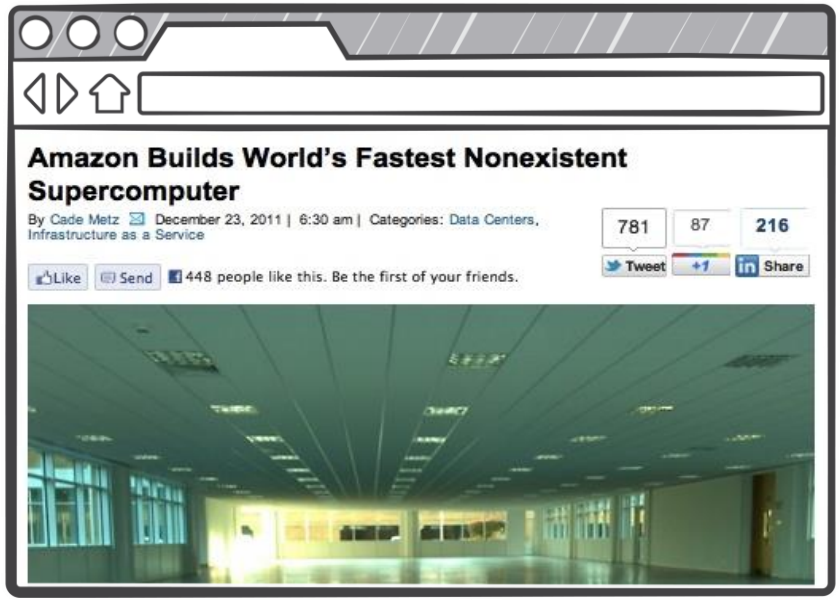
Manufacturing

Academic research

Auto & Aerospace

amazon web services

TOP500: 76th fastest supercomputer on-demand



Jun 2014 Top 500 list

484.2 TFlop/s

26,496 cores in a cluster of EC2 C3 instances

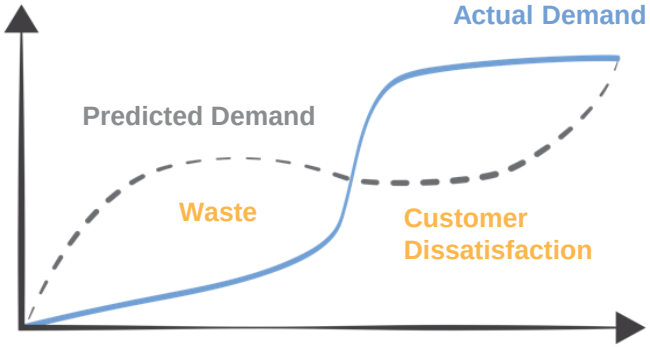
Intel Xeon E5-2680v2 10C 2.800GHzprocessors

LinPack Benchmark

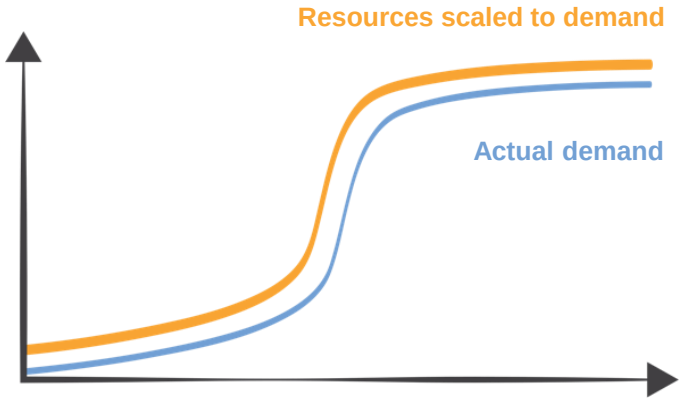
TOP500® SUPERCOMPUTER SITES

amazon web services

Benefits of Agility



Rigid On-Premises Resources



Elastic Cloud-Based Resources

amazon web services

Unilever: augmenting existing HPC capacity

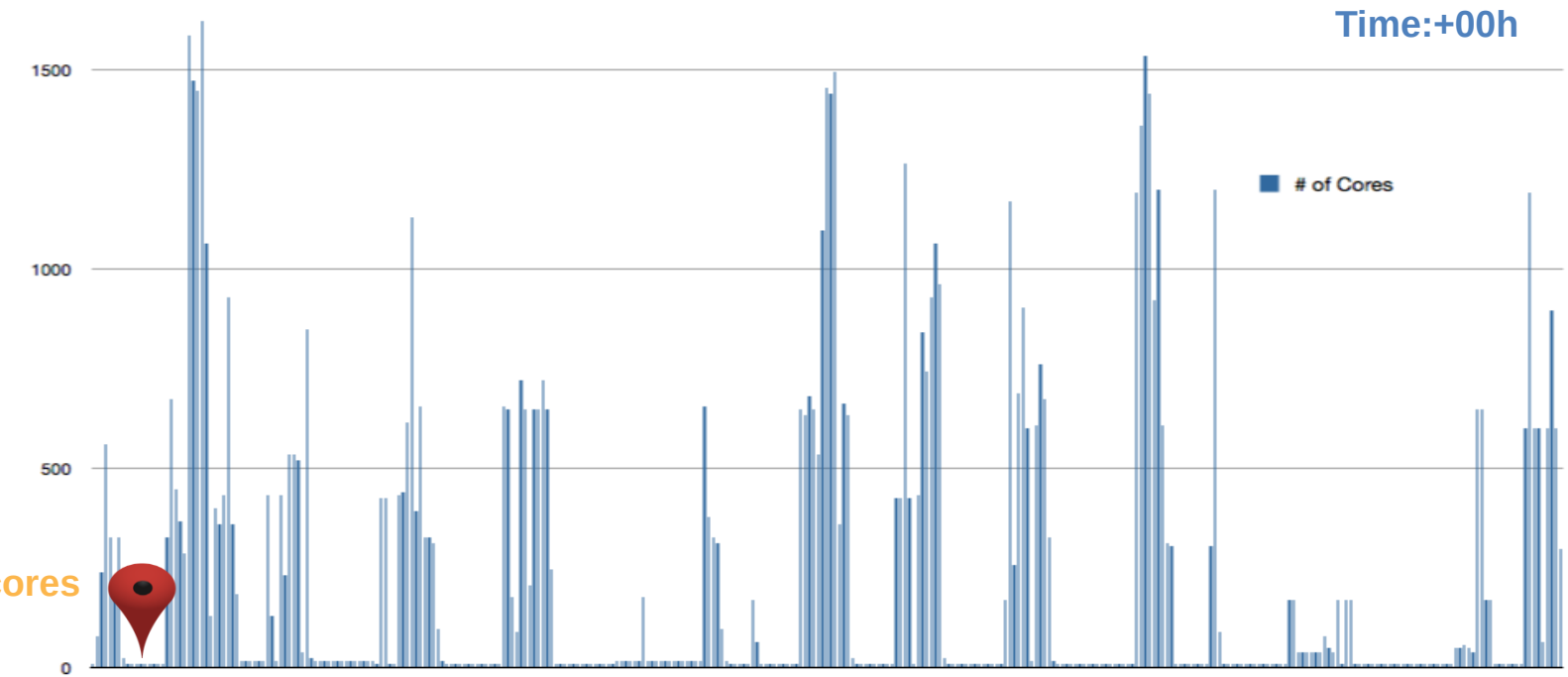
> The key advantage that AWS has over running this workflow on Unilever's existing cluster is the ability to scale up to a much larger number of parallel compute nodes on demand.

Pete Keeley
Unilever Researchs eScience
IT Lead for Cloud Solutions

- Unilever's digital data program now processes genetic sequences twenty times faster

amazon web services

Scalability on AWS

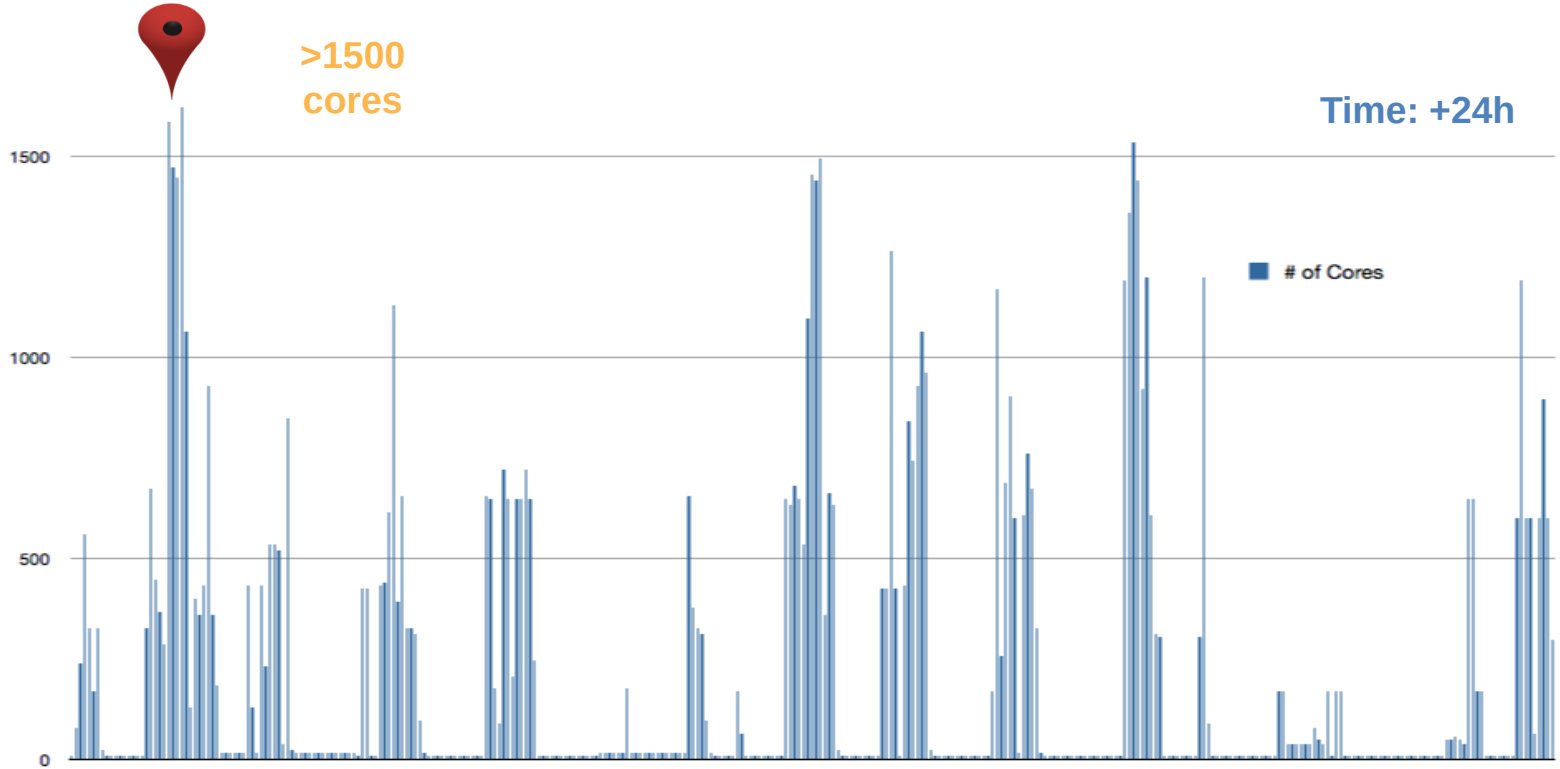


<10 cores

Scale using Elastic Capacity

amazon web services

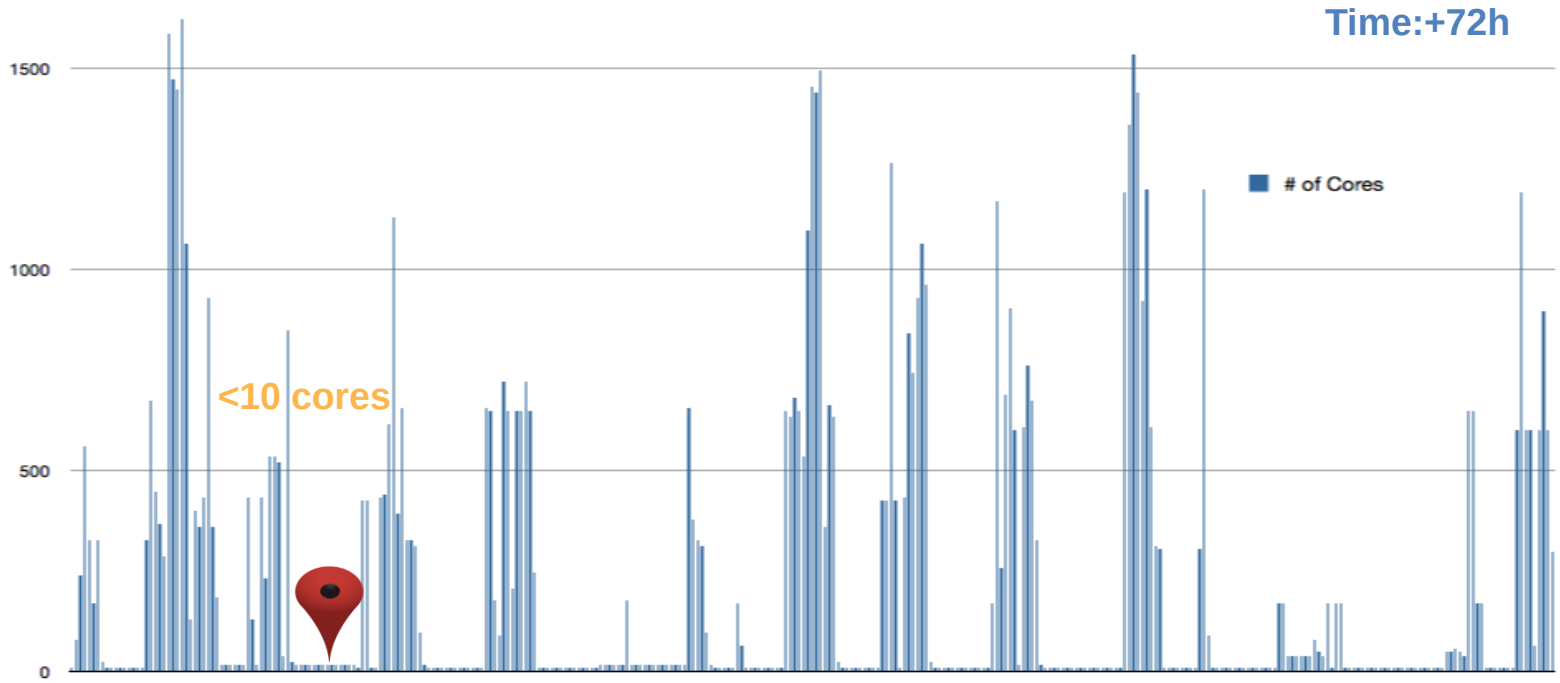
Scalability on AWS



Scale using Elastic Capacity

amazon web services

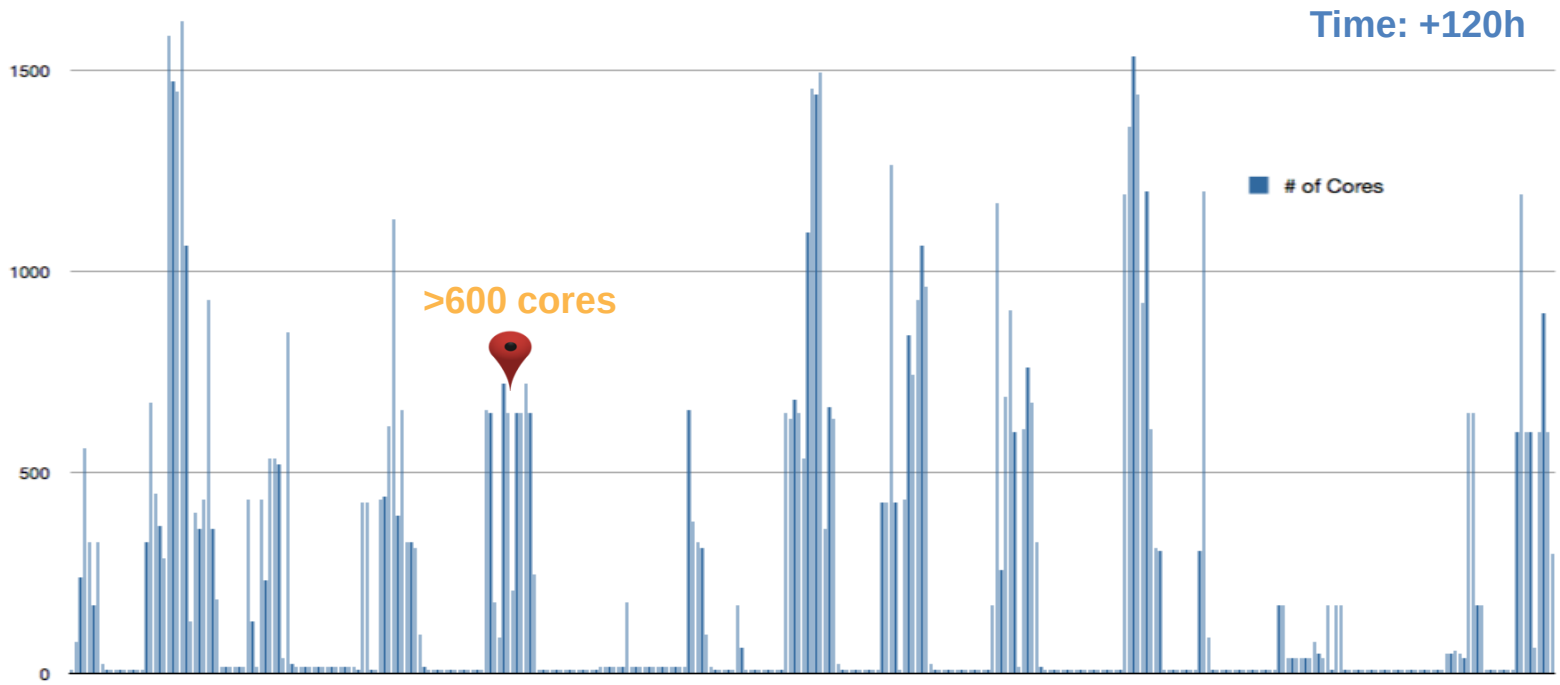
Scalability on AWS



Scale using Elastic Capacity

amazon web services

Scalability on AWS



Scale using Elastic Capacity

amazon web services

Schrodinger & CycleComputing: computational chemistry

Simulation by Mark Thompson of the University of Southern California to see which of 205,000 organic compounds could be used for photovoltaic cells for solar panel material.

Estimated computation time 264 years completed in 18 hours.

SCHRÖDINGER.

CYCLECOMPUTING

- 156,314 core cluster across 8 regions
- 1.21 petaFLOPS (Rpeak)
- $33,000 or 16¢ per molecule

amazon web services

Cost Benefits of HPC in the Cloud

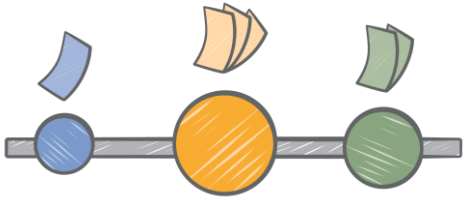
On-Premises

Capital Expense Model

High upfront capital cost

High cost of ongoing support

amazon web services



Pay As You Go Model

Use only what you need

Multiple pricing models

amazon web services

Many pricing models to support different workloads

Free Tier

Get Started on AWS with free usage & no commitment

For POCs and getting started

On-Demand

Pay for compute capacity by the hour with no long-term commitments

For spiky workloads, or to define needs

Reserved

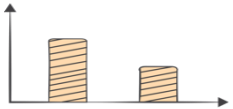
Make a low, one-time payment and receive a significant discount on the hourly charge

For committed utilization

Spot

Bid for unused capacity, charged at a Spot Price which fluctuates based on supply and demand

For time-insensitive or transient workloads



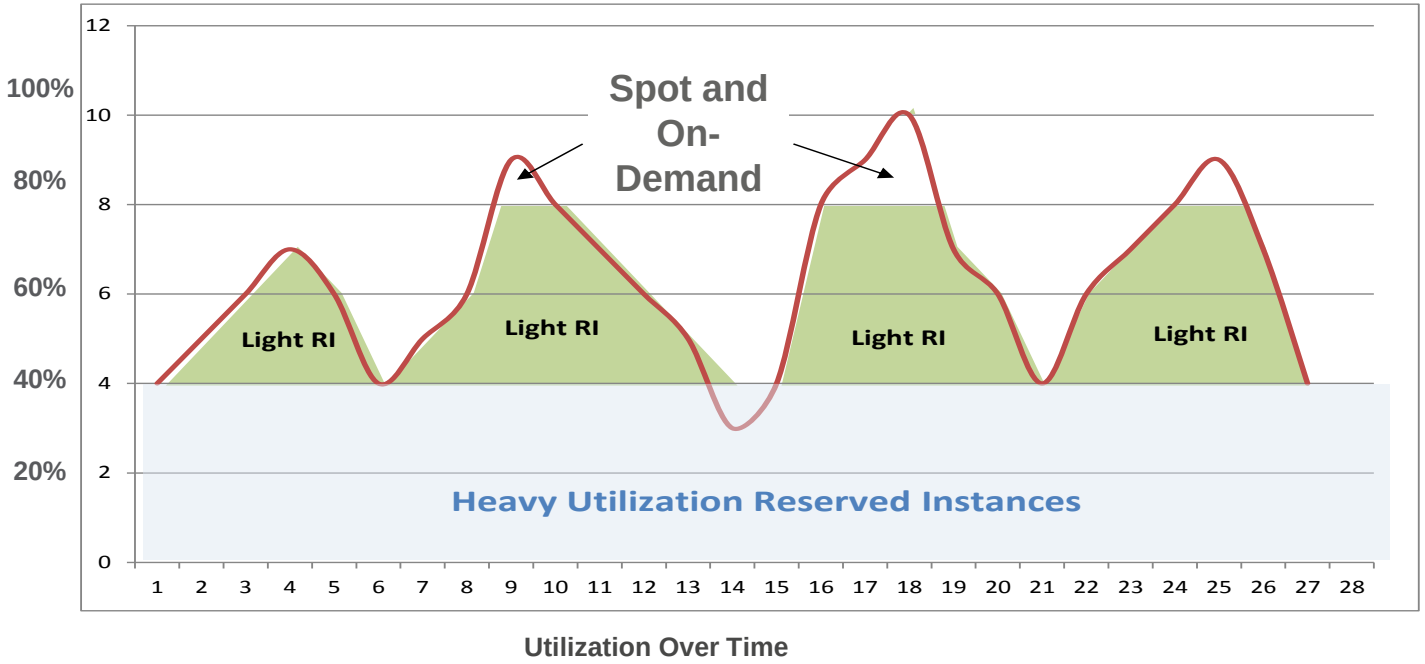
Dedicated

Launch instances within Amazon VPC that run on hardware dedicated to a single customer

For highly sensitive or compliance related workloads

amazon web services

Optimize Cost by using various EC2 instance pricing models



amazon web services

Harvard Medical School: simulation development

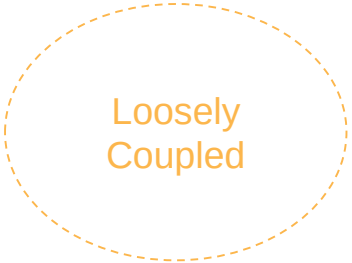
> The combination of our approach to biomedical computing and AWS allowed us to focus our time and energy on simulation development, rather than technology, to get results quickly. Without the benefits of AWS, we certainly would not be as far along as we are.

Dr. Peter Tonellato,
LPM, Center for Biomedical Informatics, Harvard Medical School

- Leveraged EC2 spot instances in workflows
- 1 day worth of effort resulted in 50% in cost savings

amazon web services

Characterizing HPC



Embarrassingly parallel

Elastic

Batch workloads

Interconnected jobs

Network sensitivity

Job specific algorithms

Characterizing HPC

Embarrassingly parallel
Elastic
Batch workloads

Data management
Task distribution
Workflow management

Interconnected jobs
Network sensitivity
Job specific algorithms

Characterizing HPC

Embarrassingly parallel
Elastic
Batch workloads

Data management
Task distribution
Workflow management

Interconnected jobs
Network sensitivity
Job specific algorithms

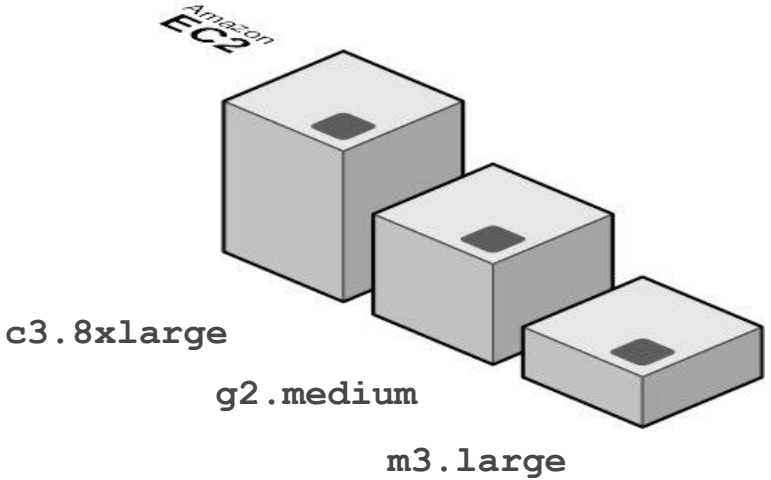
Compute Services

Elastic Compute Cloud (EC2)

Basic unit of compute capacity,virtual machines

Range of CPU, memory & local disk options

Choice of instance types, from micro to cluster compute

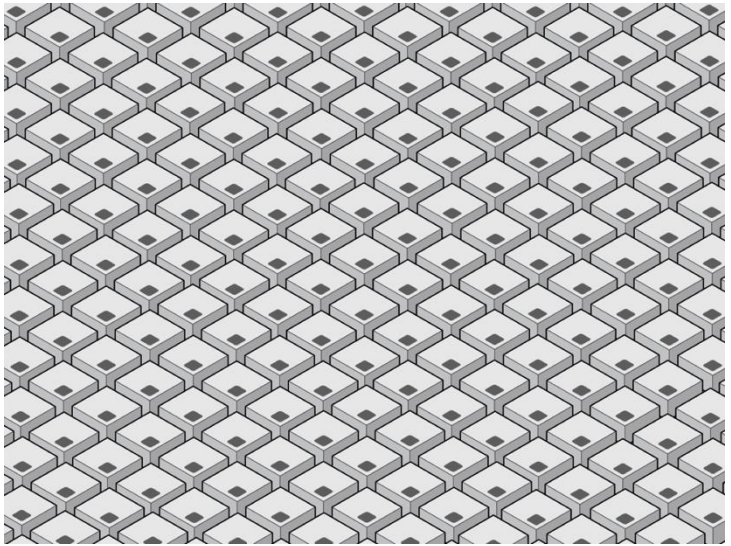


amazon web services

Automation & Control

CLI, API and Console

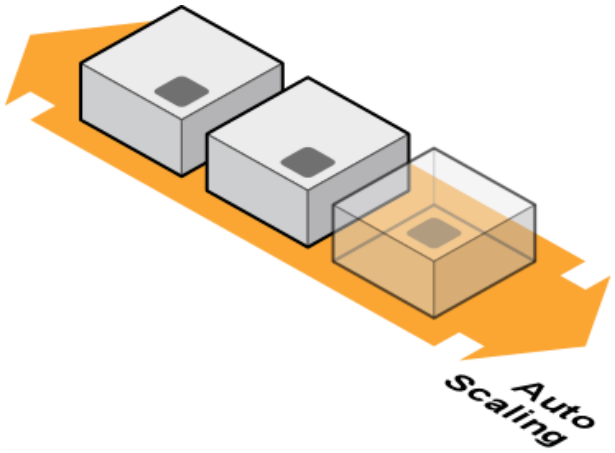
Scripted configurations



amazon web services

Auto Scaling

Automatic re-sizing of compute clusters based upon demand and policies



amazon web services

Characterizing HPC

Embarrassingly parallel
Elastic
Batch workloads

Data management
Task distribution
Workflow management

Interconnected jobs
Network sensitivity
Job specific algorithms

What if you need to:

Implement MPI?
Code for GPUs?

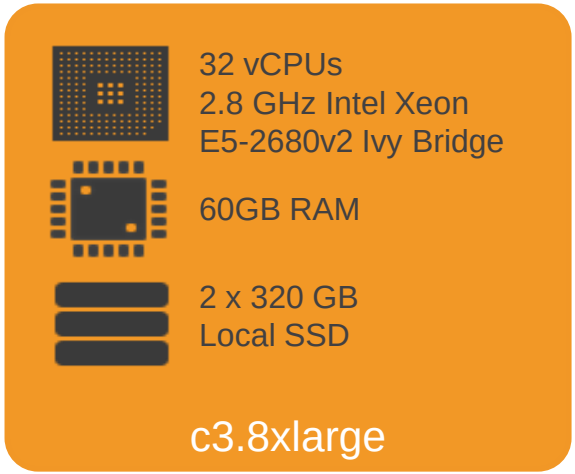
Tightly coupled

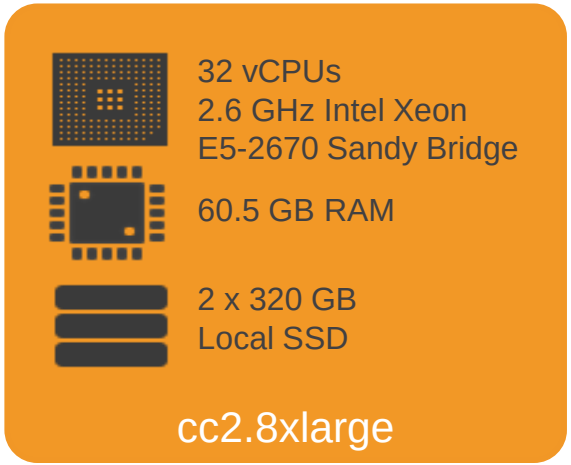
Cluster compute instances

Implement HVM process execution

Intel® Xeon® processors

10 Gigabit Ethernet –c3 has Enhanced networking, SR-IOV

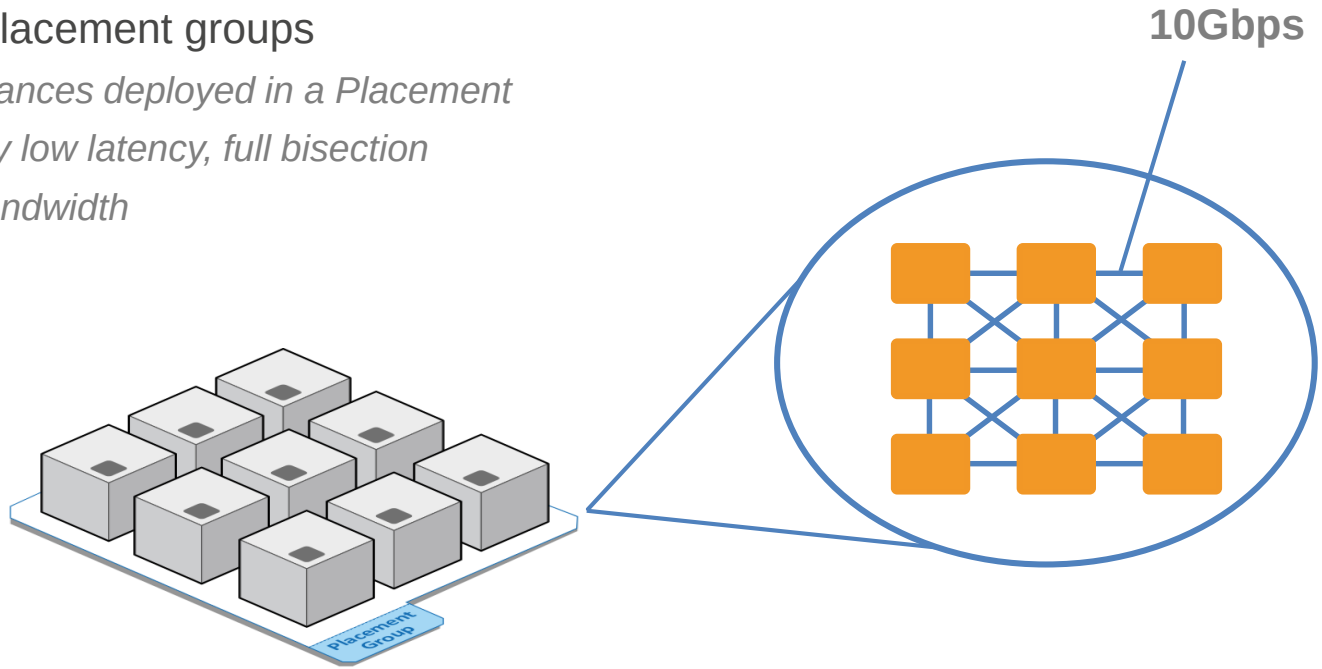




Tightly coupled

Network placement groups

Cluster instances deployed in a Placement Group enjoy low latency, full bisection 10 Gbps bandwidth



amazon web services

Tightly coupled

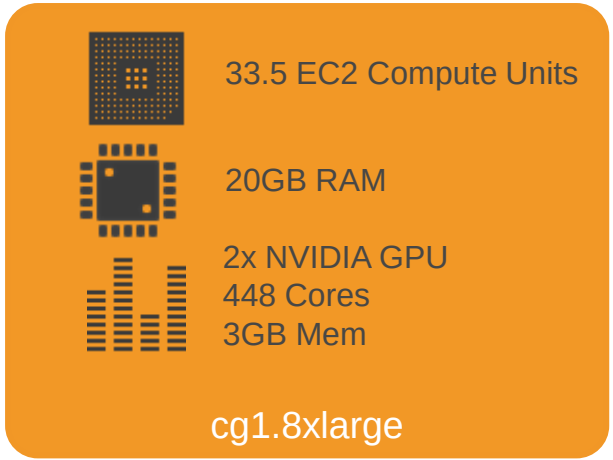
GPU compute instances

CG1 instances

Intel® Xeon® X5570 processors

2 x NVIDIA Tesla "Fermi" M2050 GPUs

I/O Performance: Very High (10 Gigabit Ethernet)

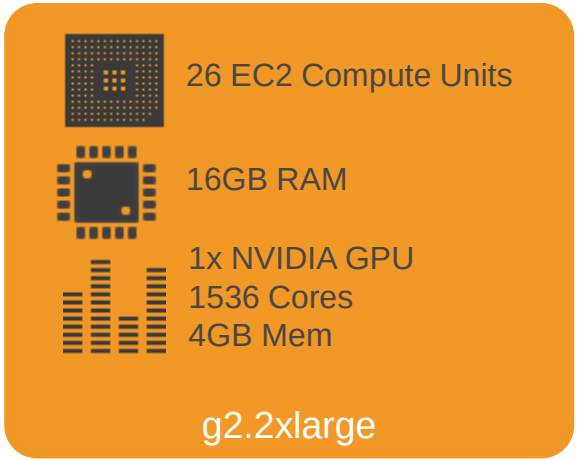


G2 instances

Intel® Intel Xeon E5-2670

1 NVIDIA Kepler GK104 GPU

I/O Performance: Very High (10 Gigabit Ethernet)



amazon web services

National Taiwan University: shortest vector problem

> Our purpose is to break the record of solving the shortest vector problem (SVP) in Euclidean lattices…the vectors we found are considered the hardest SVP anyone has solved so far.

Prof. Chen-Mou Cheng
Principle Investigator of Fast Crypto Lab

- $2,300 for using 100x Tesla M2050 for ten hours

Position	Dimension	Euclidean norm	Seed	Contestant	Solution	Algorithm	Subm. Date
1	120	2851	0	Po-Chun Kuo, Michael Schneider	vec	ENUM,BKZ	2011-04-6
2	116	2825	0	Po-Chun Kuo, Michael Schneider	vec	ENUM,BKZ	2011-04-1
3	114	2778	0	Po-Chun Kuo, Michael Schneider	vec	ENUM,BKZ	2011-03-21

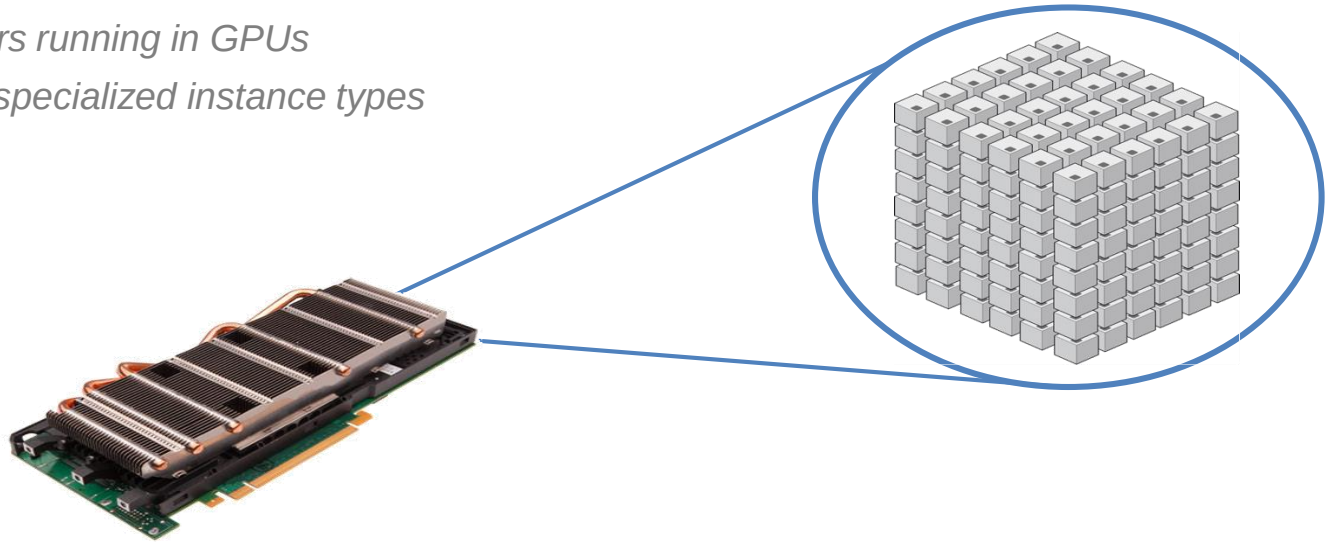
amazon web services

Tightly coupled

CUDA & OpenCL

Massive parallel clusters running in GPUs

NVIDIA Tesla cards in specialized instance types



amazon web services

Characterizing HPC

Embarrassingly parallel
Elastic
Batch workloads

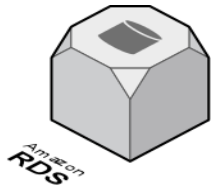
Data management
Task distribution
Workflow management

Interconnected jobs
Network sensitivity
Job specific algorithms

Supporting Services

Data management

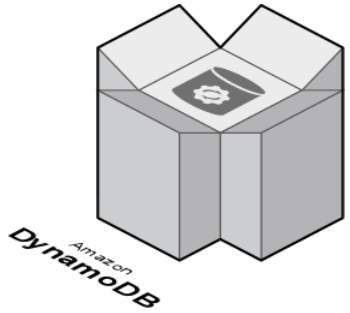
Fully-managed SQL, NoSQL, and object storage



Relational Database Service

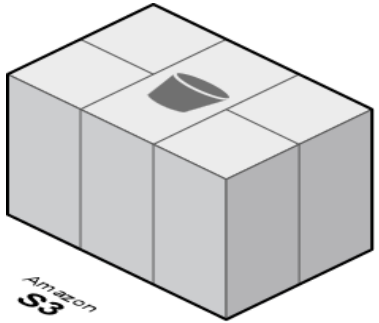
Fully-managed database

(MySQL, Oracle, MSSQL, PostgreSQL)



DynamoDB

NoSQL, Schemaless, Provisioned throughput database



S3

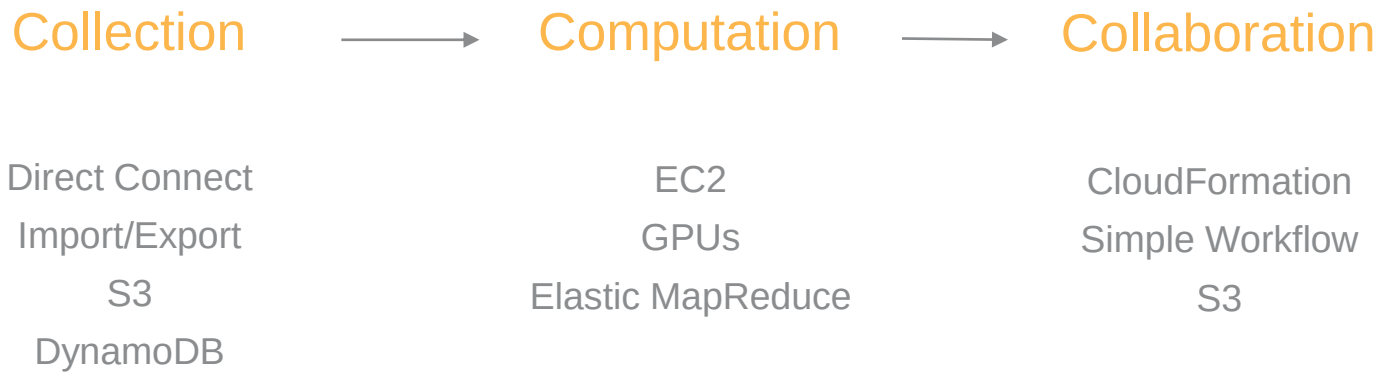
Object datastore up to 5TB per object

Internet accessibility

amazon web services

Moving compute closer to the data

“Big Data” changes dynamic of computation and data sharing



TRADERWORX: Market Information Data Analytics System

> For the growing team of quant types now employed at the SEC, MIDAS is becoming the world's greatest data sandbox. And the staff is planning to use it to make the SEC a leader in its use of market data
>
> **Elisse B. Walter,**
> Chairman of the SEC
> Tradeworx

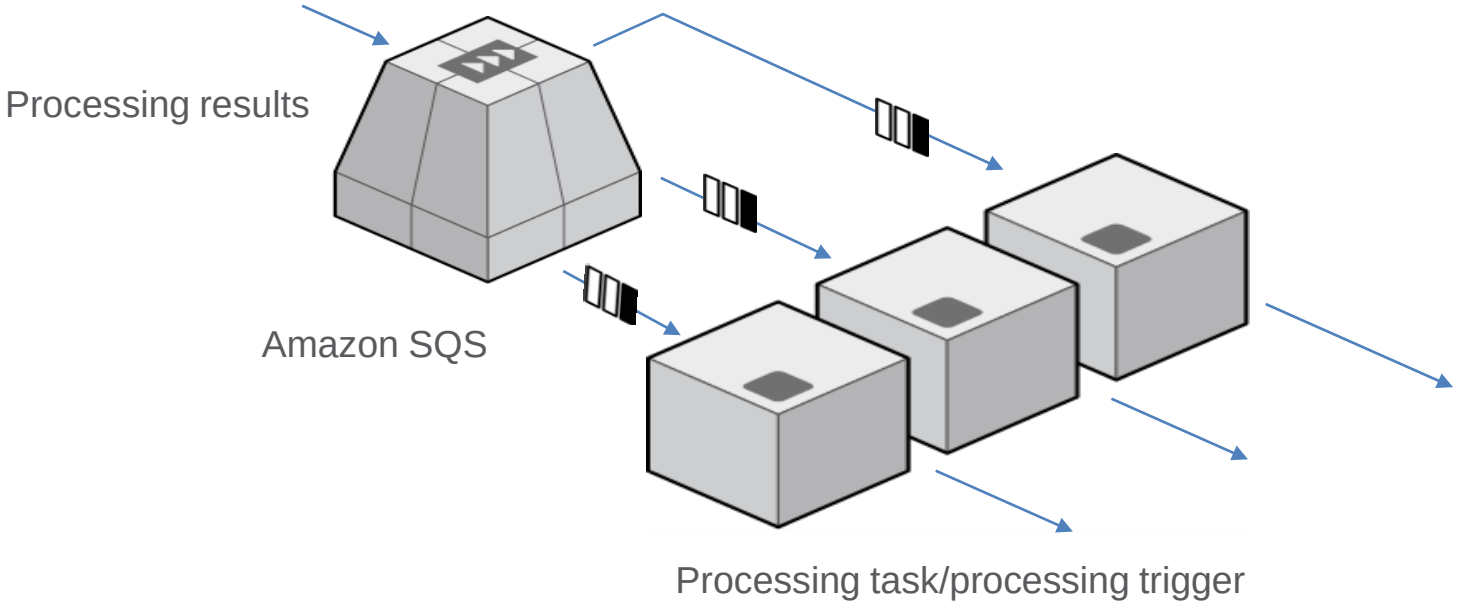
- Powerful AWS-based system for market analytics
- 2M transaction messages/*sec*; 20B records and 1TB/*day*

amazon web services

Supporting Services

Feeding workloads

Using highly available Simple Queue Service to feed EC2 nodes

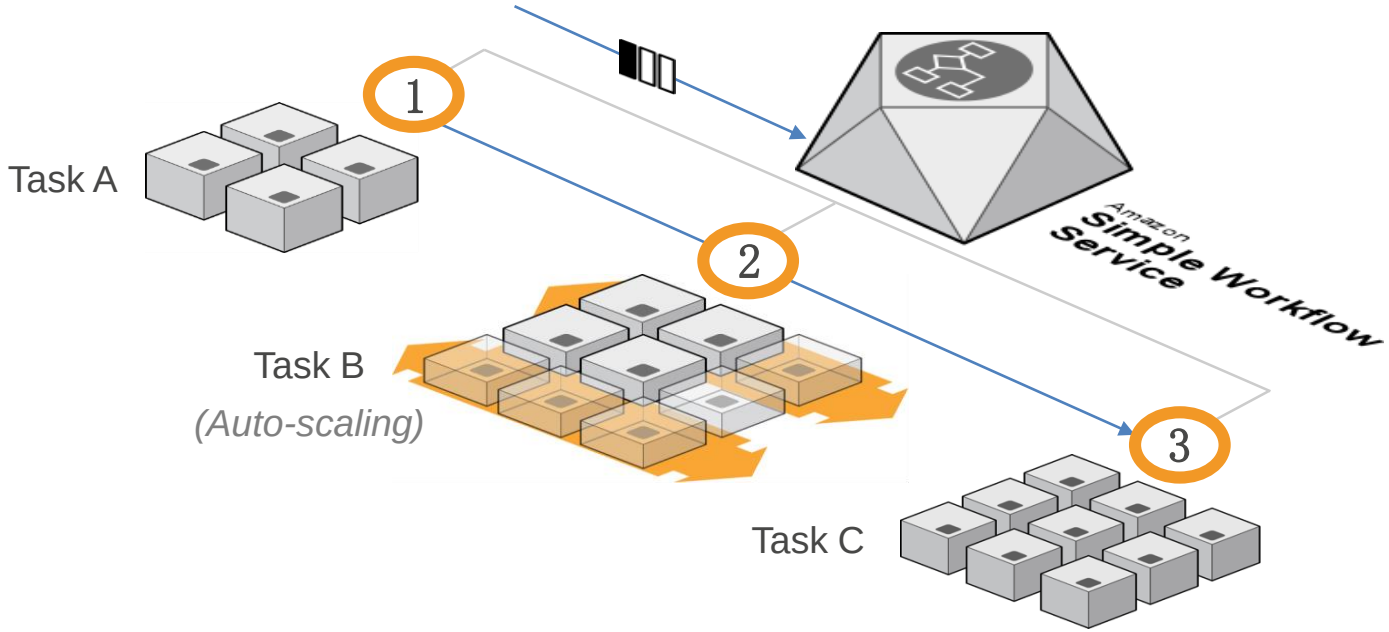


amazon web services

Supporting Services

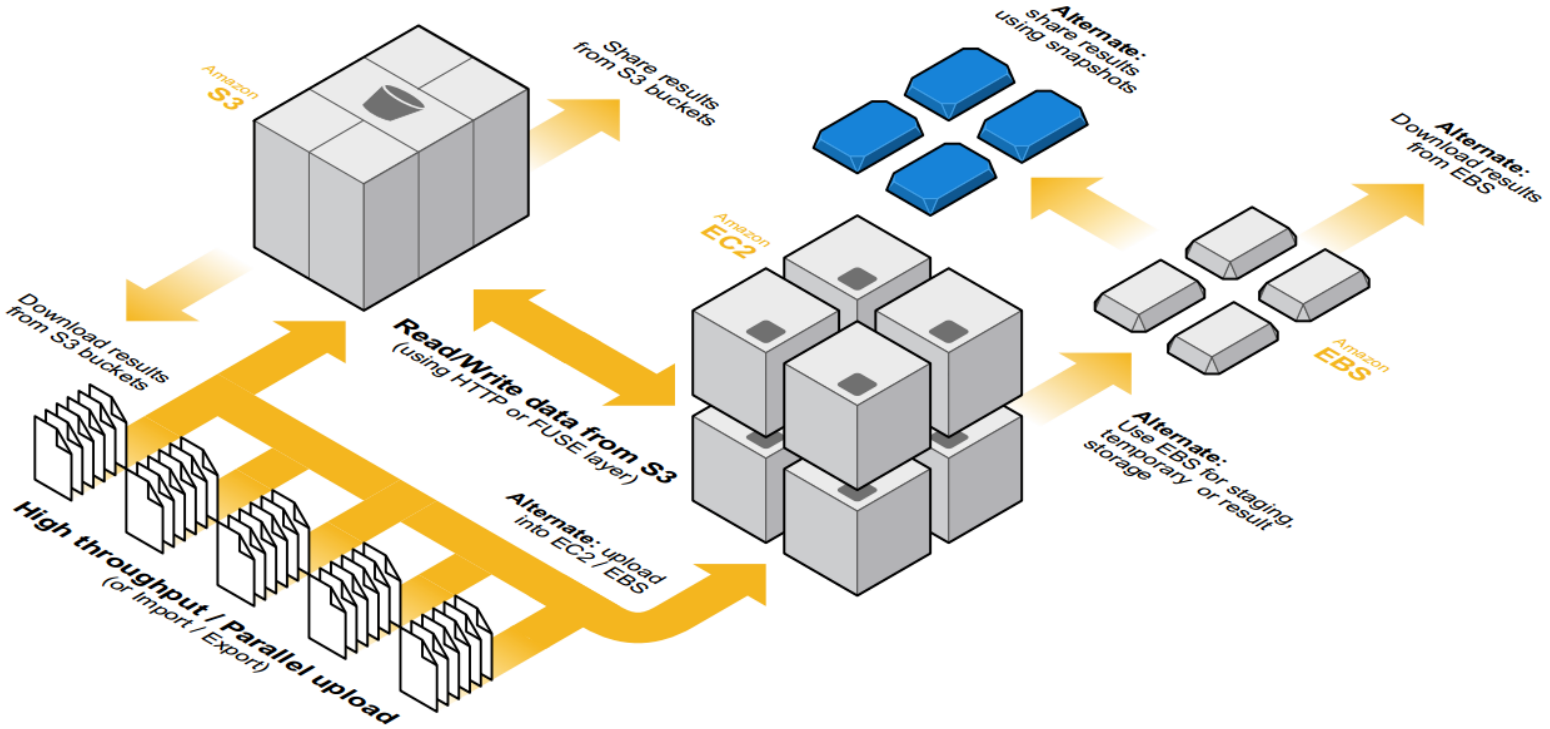
Coordinating workloads & task clusters

Handle long running processes across many nodes and task steps with Simple Workflow



amazon web services

Architecture of large scale computing and huge data sets



amazon web services

NYU School of Medicine: Transferring large data sets

> Transferring data is a large bottleneck; our datasets are extremely large, and it often takes more time to move the data than to generate it. Since our collaborators are all over the world, if we can't move it they can't use it.

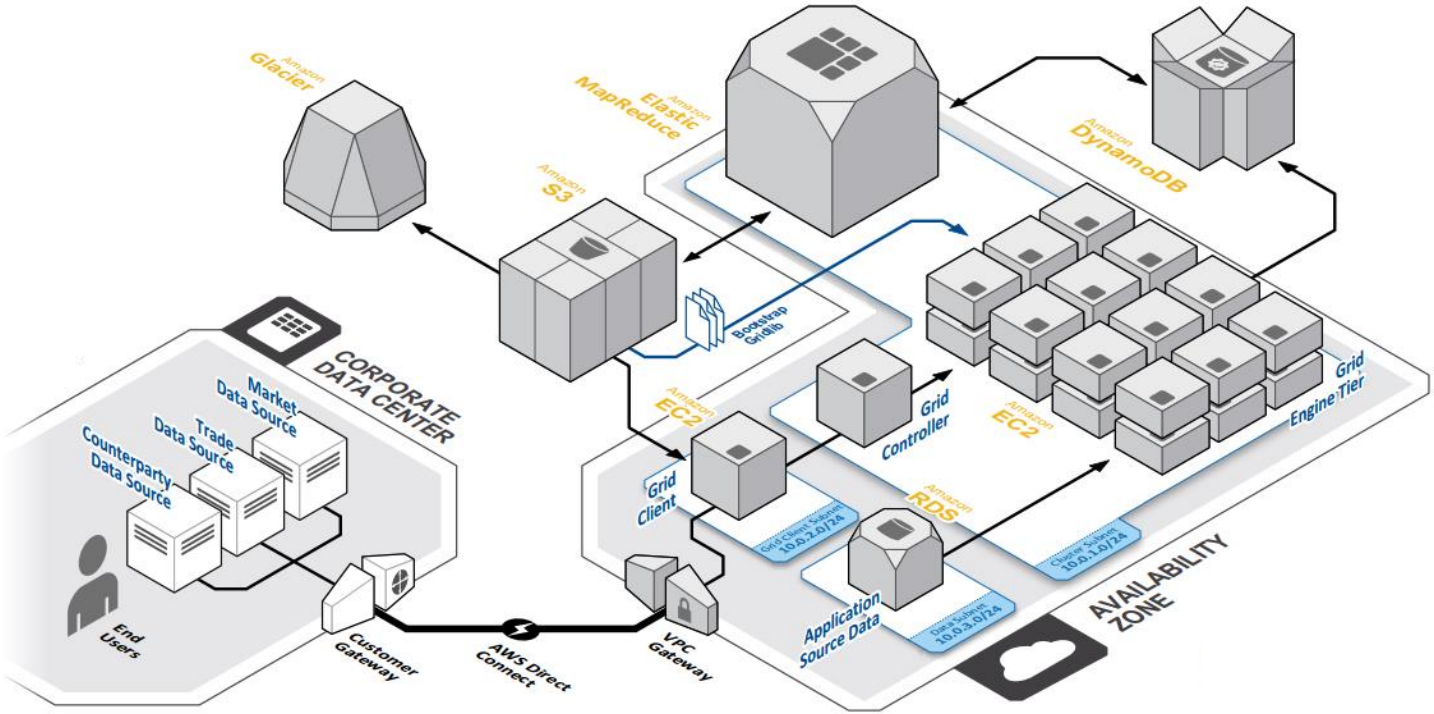
Dr. Stratos Efstathiadis
Technical Director of the
HPC facility, NYU

NYU School of Medicine
NYU LANGONE MEDICAL CENTER

- Uses Globus Online
- Data transfer speeds of up to 50MB/s

amazon web services

Architecture of a financial services grid computing



amazon web services

Bankinter: credit-risk simulation

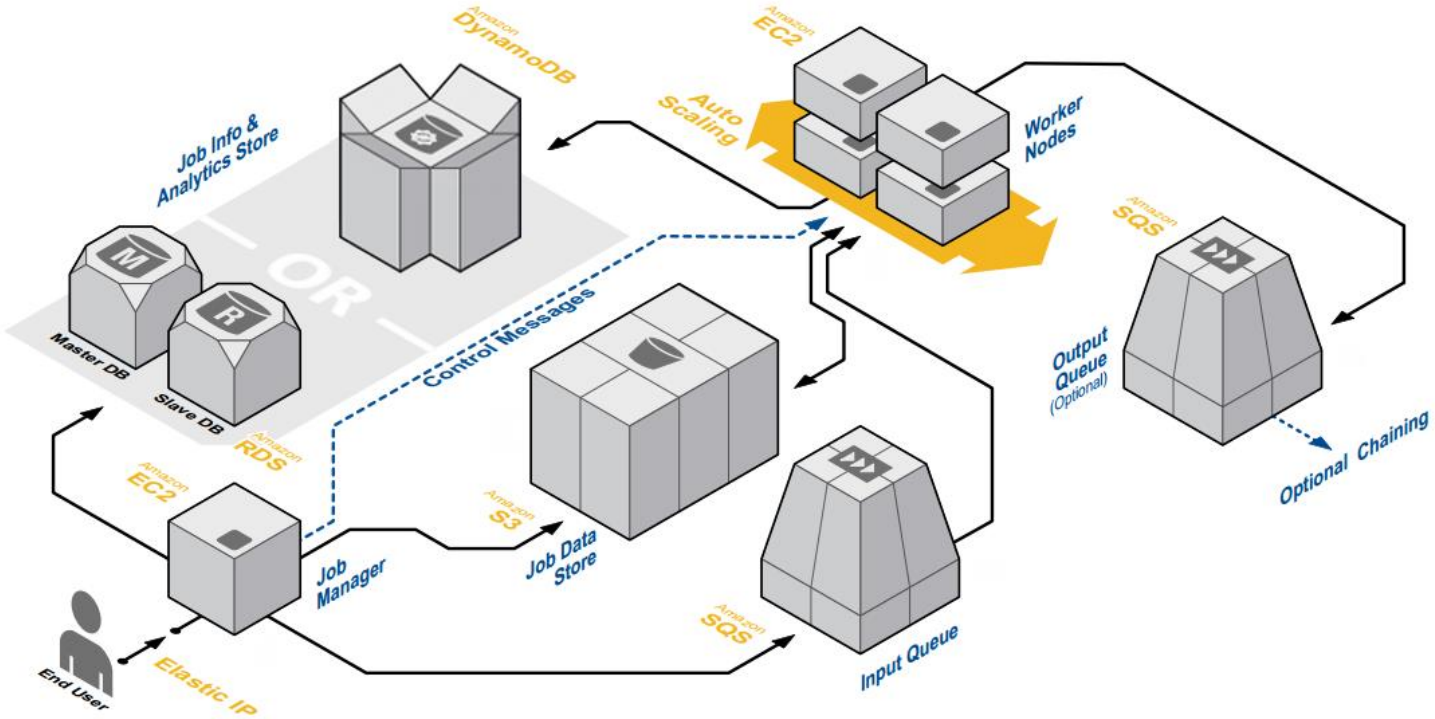
> With AWS, we now have the power to decide how fast we want to obtain simulation results. More important, we have the ability to run simulations that were not possible before due to the large amount of infrastructure required.
>
> **Javier Roldán**
> Director of Technological Innovation, Bankinter

bankinter

- Reduced processing time of 5,000,000 simulations from 23 hours to 20 minutes

amazon web services

Architecture of queue-based batch processing



amazon web services

When to consider running HPC workloads on AWS

Improvement

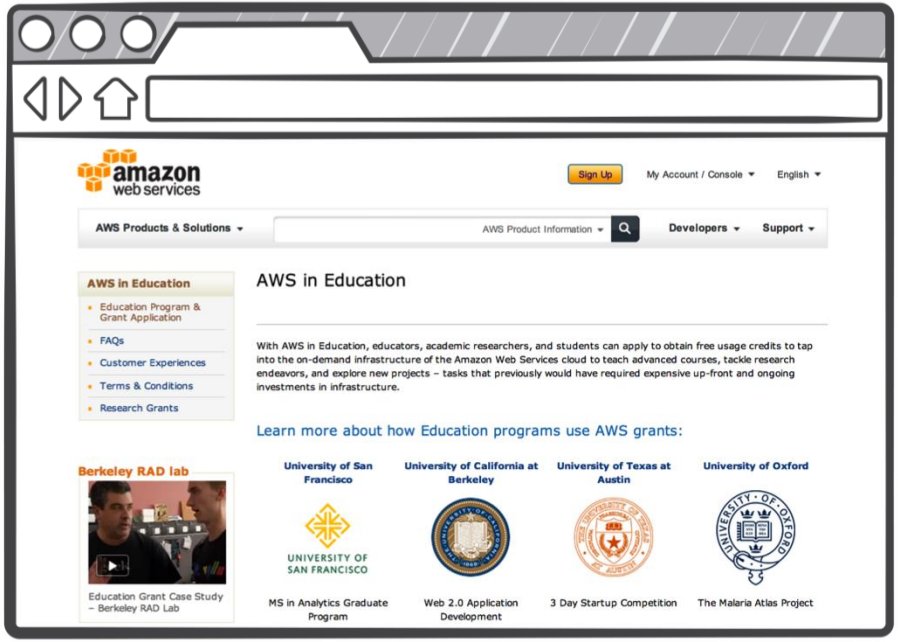
Remove the queue
Hardware refresh cycle
Reduce costs
Collaboration of results
Increase innovation speed
Reduce time to results

New ideas

New HPC project
Proof of concept
New application features
Training
Benchmarking algorithms

amazon
web services

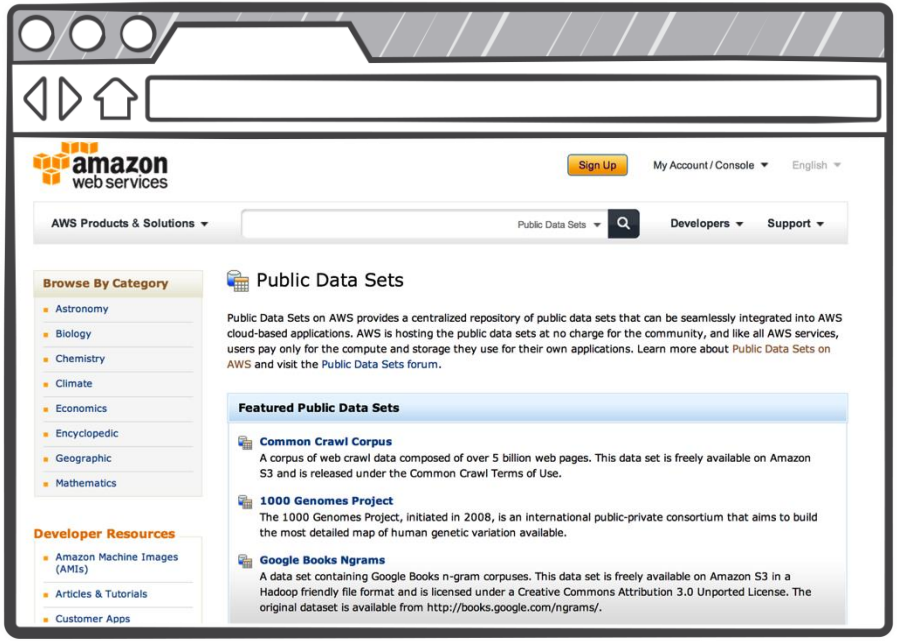
AWS Grants Program



aws.amazon.com/grants

amazon web services

AWS Public Data Sets

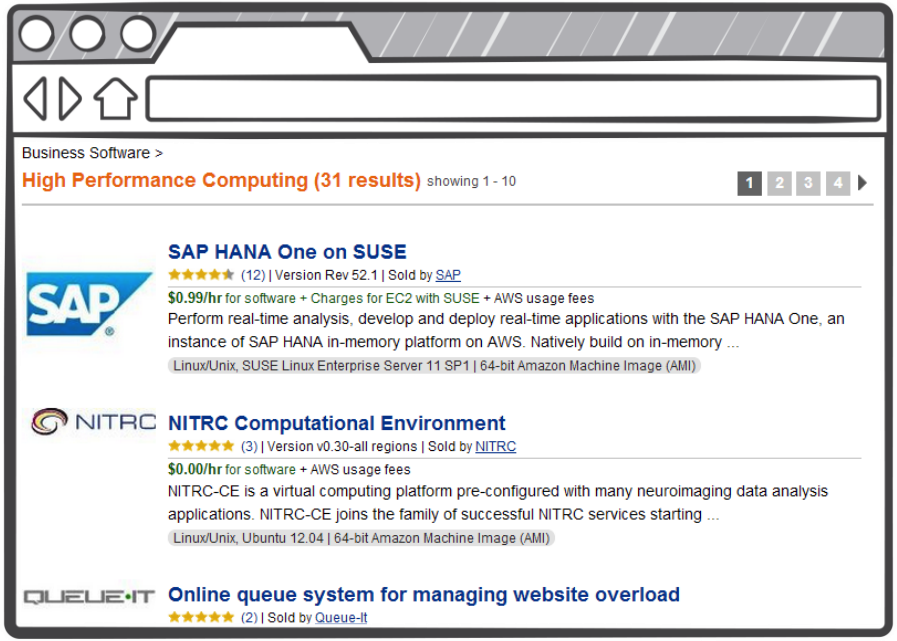


aws.amazon.com/datasets

free for everyone

amazon web services

AWS Marketplace – HPC category

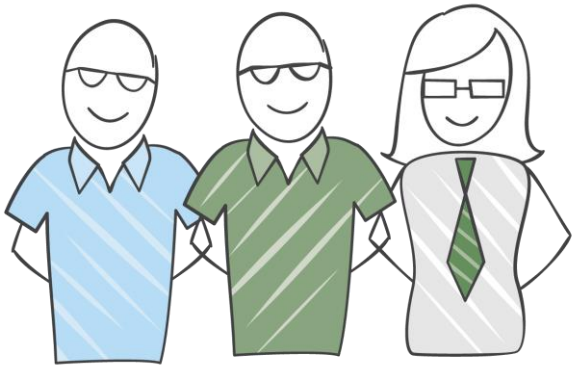


aws.amazon.com/marketplace

awsmarketplace

amazon web services

Getting Started with HPC on AWS



Sales and Solutions Architects

Enterprise Support

Trusted Advisor

Professional Services

aws.amazon.com/hpc

contact us, we are here to help

amazon web services

Try out our HPC CloudFormation-based demo

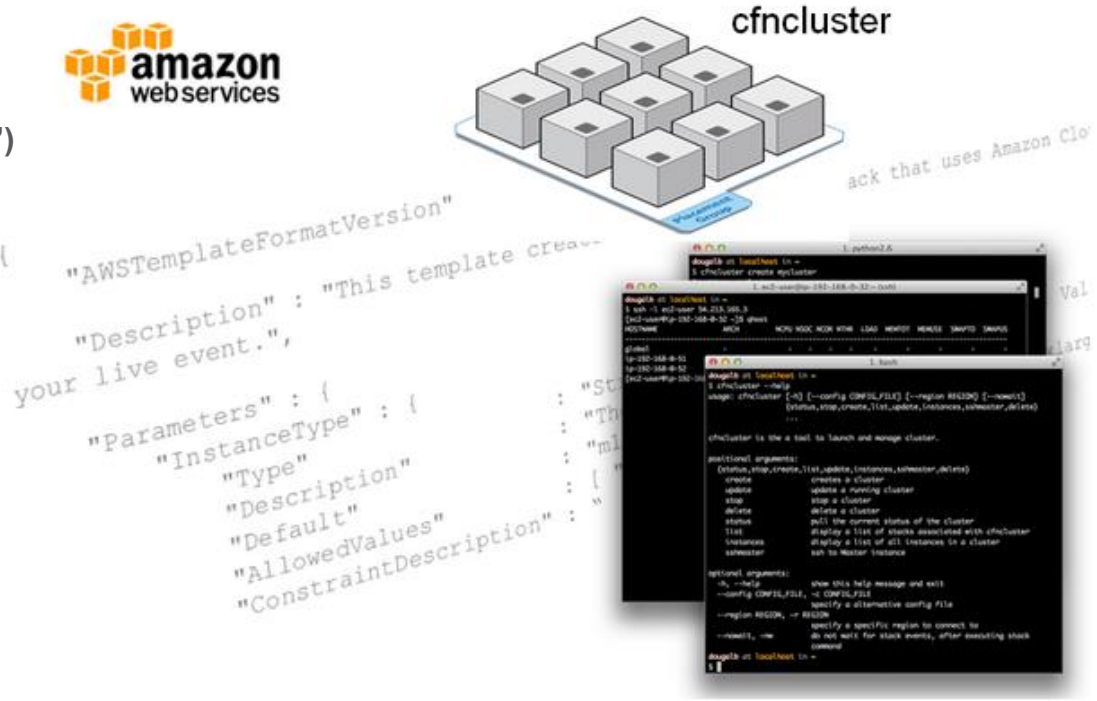
cfncluster (“CloudFormation cluster”)

Command Line Interface Tool

Deploy and demo an HPC cluster

For more info:

aws.amazon.com/hpc/resources



HPC Partners and Apps

Costomers are using AWS for more and more HPC workloads

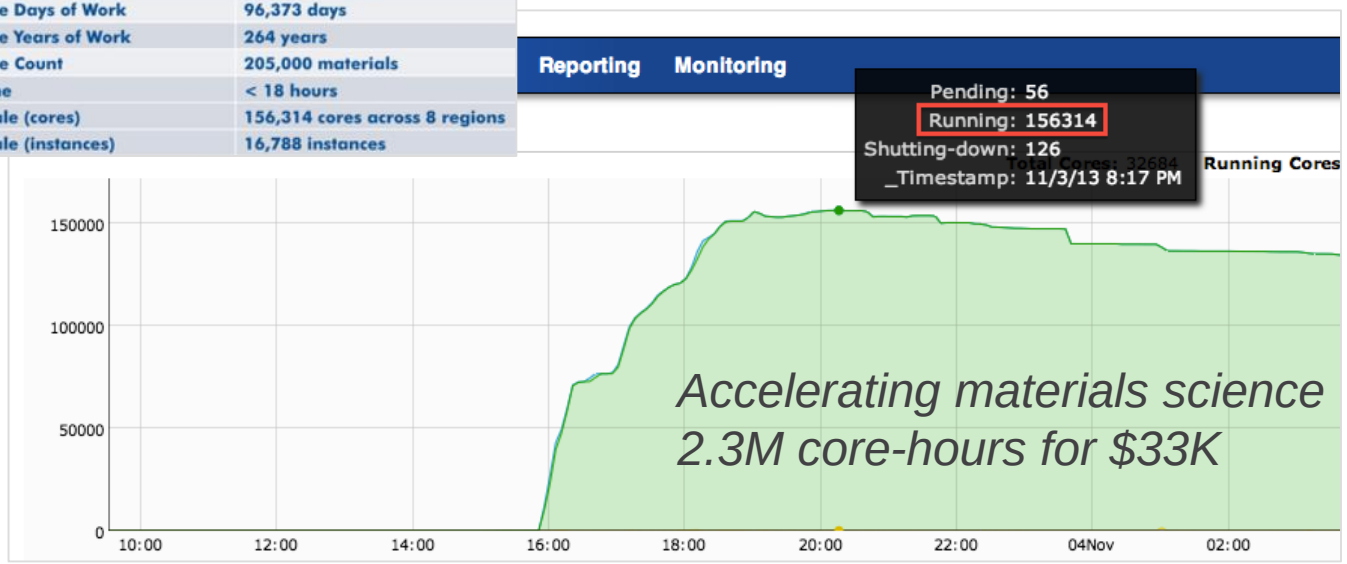
Oil and Gas	Manufacturing & Engineering	Life Sciences	Media & Entertainment	Scientific Computing	Finaancial
Seismic Data Processing	Computational Fluid Dynamics (CFD)	Genome Analysis	Transcoding and Encoding	Computational Chemistry	Monte Carlo Simulations
Reservoir Simulations, Modeling	Finite Element Analysis (FEA)	Molecular Modeling	DRM, Encryption	High Energy Physics	Wealth Management Simulations
Geospatial applications	Wind Simulation	Protein Docking	Rendering	Stochastic Modeling	Portfolio, Credit Risk Analytics
Predictive Maintenance				Quantum Analysis	High Frequency Trading Analytics
				Climate Models	

amazon web services

So What Does Scale Mean on AWS?

CYCLECOMPUTING
LEADER IN CONDOR GRID COMPUTING SOLUTIONS

Metric	Count
Compute Hours of Work	2,312,959 hours
Compute Days of Work	96,373 days
Compute Years of Work	264 years
Molecule Count	205,000 materials
Run Time	< 18 hours
Max Scale (cores)	156,314 cores across 8 regions
Max Scale (instances)	16,788 instances



amazon web services

Cyclopic energy: computational fluid dynamics

> AWS makes it possible for us to deliver state-of-the-art technologies to clients within timeframes that allow us to be dynamic, without having to make large investments in physical hardware.
>
> **Rick Morgans**
> Technical Director (CTO),
> cyclopic energy

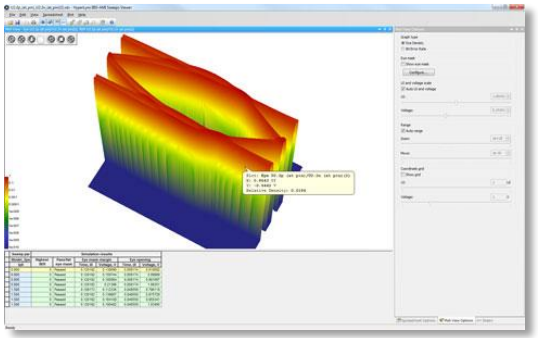
- Two months worth of simulations finished in two days

amazon web services

Mentor Graphics: virtual lab for design and simulation

> Thanks to AWS, the Mentor Graphics customer experience is now fast, fluid, and simple.
>
> **Ron Fuller**
> Senior Director of Engineering, Mentor

- Developed a virtual lab for ASIC design and simulation for product evaluation and training



AeroDynamic Solutions: turbine engine simulation

> We’re delighted to be working closely with the U.S. Air Force and AWS to make time accurate simulation a reality for designers large and small.

George Fan
CEO, AeroDynamic Solutions

AeroDynamic Solutions

- Time accurate simulation was turned around in 72 hours with infrastructure costs well below $1,000

amazon web services

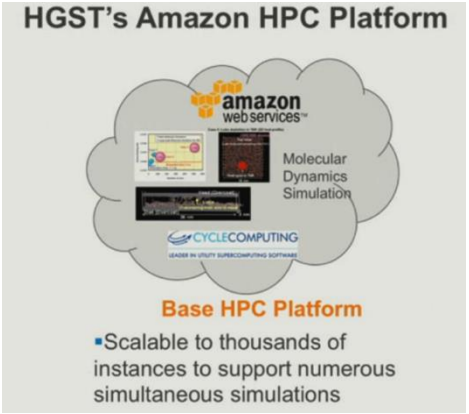
HGST: molecular dynamics simulation

HGST is using AWS for a higher performance, lower cost, faster deployed solution vs buying a huge on-site cluster.

Steve Philpott
CIO, HGST

HGST
a Western Digital company

- Uses HPC on AWS for CAD, CFD, and CDA



amazon web services

Pfizer: large-scale data analytics and modeling

AWS enables Pfizer's Worldwide Research and Development to explore specific difficult or deep scientific questions in a timely, scalable manner and helps Pfizer make better decisions more quickly.

Dr. Michael Miller
Head of HPC for R&D,
Pfizer

- Pfizer avoiding having to procure new HPC hardware by being able to use AWS for peak work loads.

Thank you!

Please visit **aws.amazon.com/hpc** for more info

amazon web services