



# Fueling HPC on AWS with GPUs

Don't let your innovation get stuck in a queue !

Chetan Kapoor, Sr Product Manager, AWS

Dr. Srin Chari, Co-founder, Cabot Partners

# Agenda

## Overview of AWS HPC Solutions

### Cabot Partners

- Market Overview – HPC applications using GPUs
- Predominant HPC Use Cases for GPU-based compute
- Summary

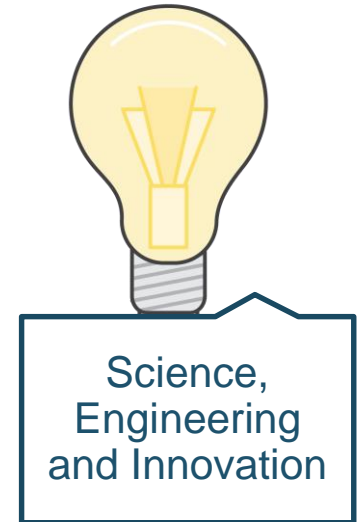
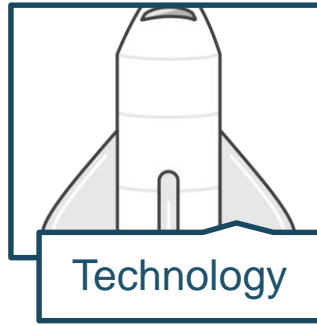
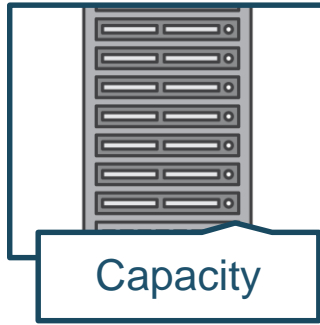
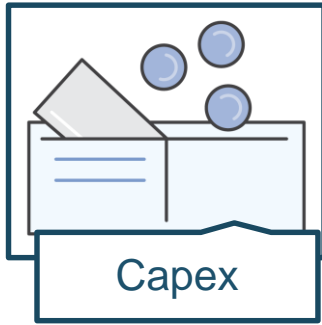
### Wrap up and Additional Resources

# HPC on AWS: Fundamental Rethink of What is Possible

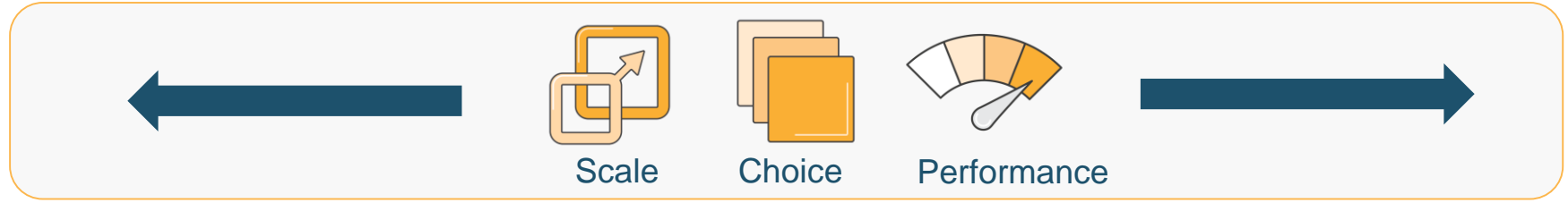
From worrying about

to

Focusing on



# AWS Advantages for HPC Workload Types



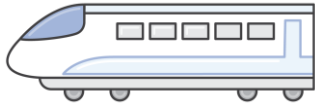
**Tightly Coupled  
Parallel  
Computing**

**Loosely Coupled  
Parallel  
Computing**

**Accelerated  
Computing**

**Visualization and  
Interpretation**

**High Performance  
Data Storage and  
Analytics**



Skip the Queue



EC2 Spot  
Pricing



Early Access to  
Technology

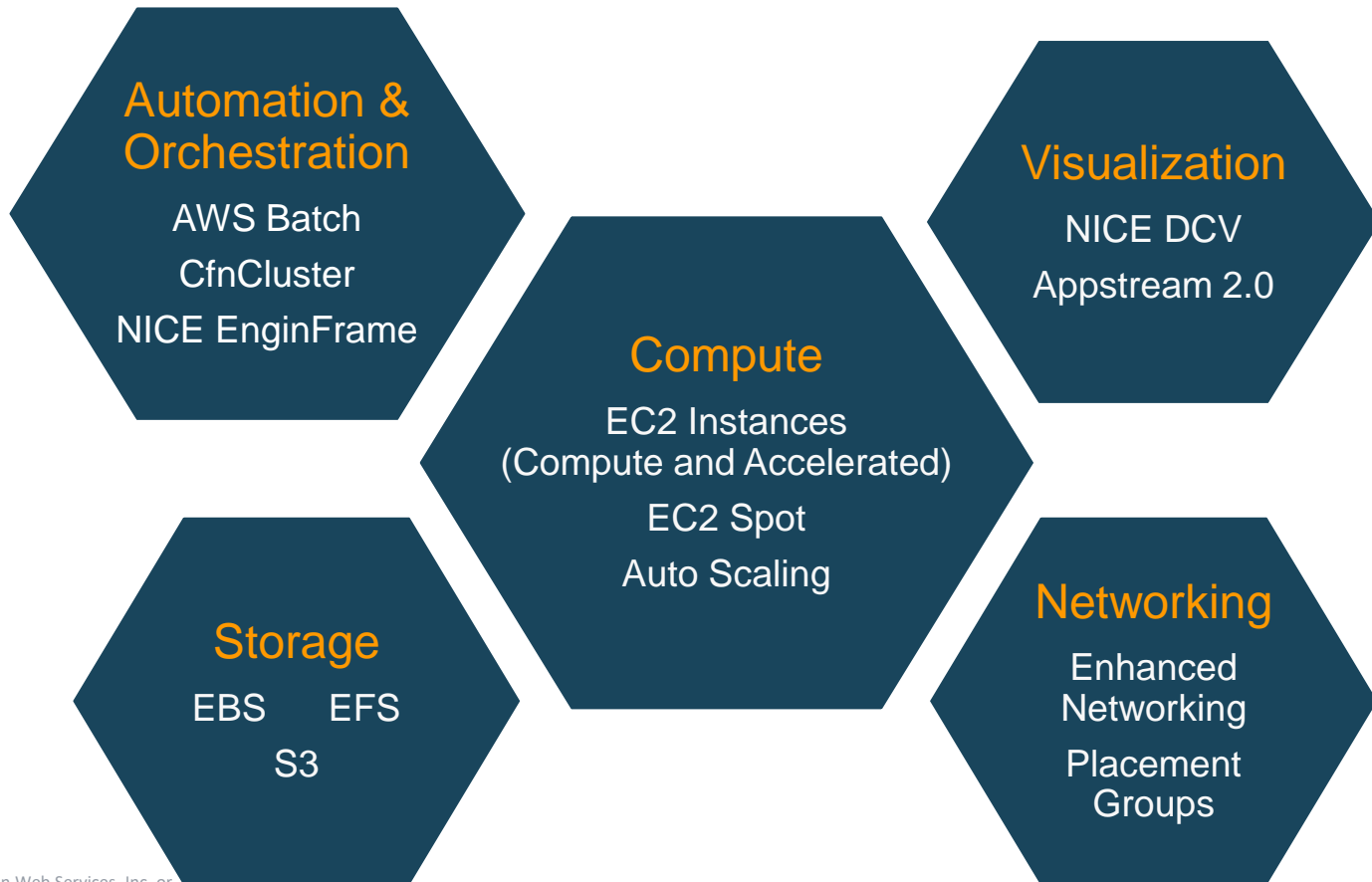


View results  
instantly



Derive unique  
insights with AI/ML

# AWS HPC Solution Components



# High Performance Computing on AWS

- **Innovate faster** with virtually unlimited infrastructure enabling scaling and agility not attainable on-premises
- **Optimize cost** with flexible resource selection and pay per use
- **Increase collaboration** with secure access to clusters around the world



**Faster Time to Results**



**Better ROI**

# Agenda

## Overview of AWS HPC Solutions

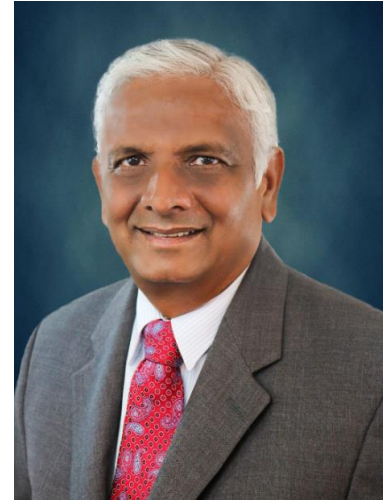
### Cabot Partners

- Market Overview – HPC applications using GPUs
- Predominant HPC Use Cases for GPU-based compute
- Summary

## Wrap up and Additional Resources

# Cabot Partners Introduction

- IT Analyst Firm Focused on Emerging Computing Technologies
- In Business Since 2006. **Over Twelve Years.**
- Covering High Performance Computing (HPC), AI/Machine Learning, Analytics, Cloud Computing
- Deep Business and IT Expertise with Senior Team.
- Diverse Industry Vertical Knowledge: Spans Financial Services, Manufacturing, Life Sciences, Healthcare, Oil and Gas, Telecommunications and Others
- East and West Coast with Extensive Partner Network



Srin Chari, Ph.D., MBA  
Managing Partner  
Cabot Partners Group



# Outline

## HPC is growing across many industries with a 10% CAGR

- Continuing explosion of data volumes and the need for higher fidelity models
- Lines between Analytics and HPC blurring
- Significant growth of AI and Deep Learning
- GPU adoption increasing as more applications are accelerated by orders of magnitude

## Why is HPC moving to the cloud at over 40% CAGR?

- Many impediments with HPC use on premises
- Traditional barriers of HPC on the cloud collapsing
- Benefits of moving HPC to the AWS cloud
- Many high value industry use cases and benefits with the AWS NVIDIA ecosystem

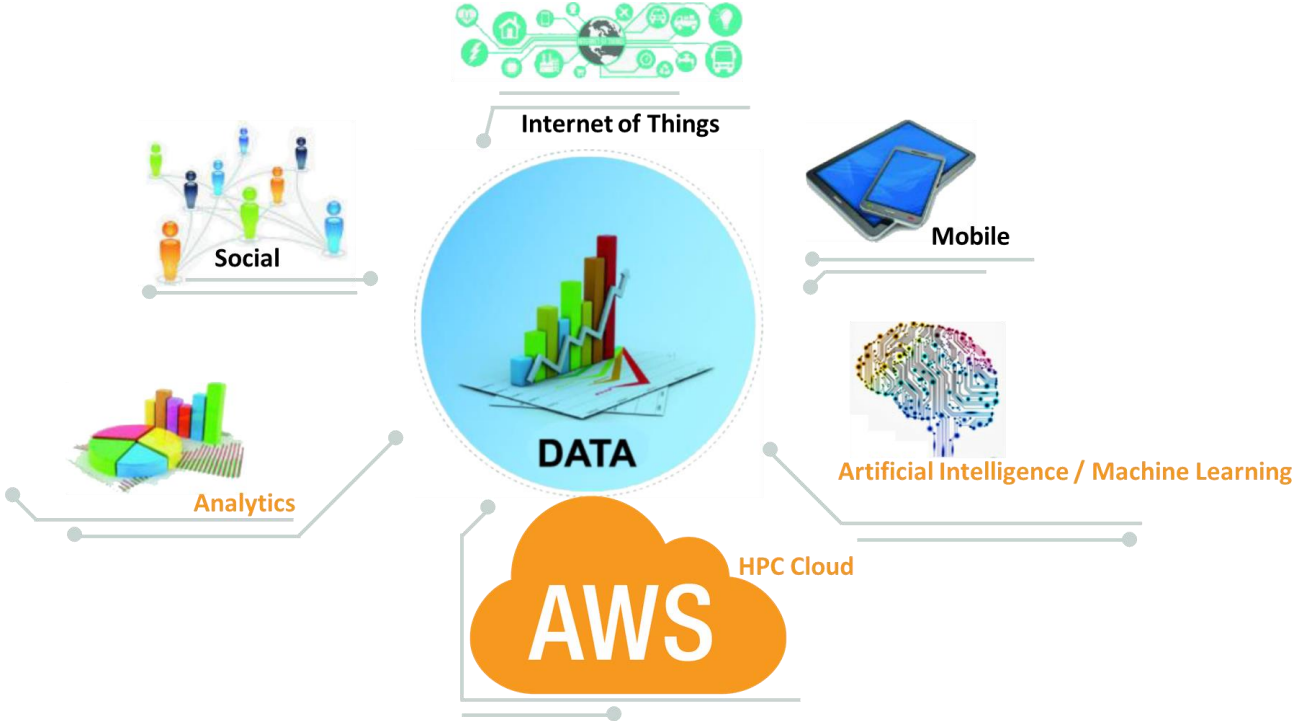
## Compelling reasons for using NVIDIA GPU-based Instances on AWS for HPC

- NVIDIA GPUs are accelerating HPC workflows by orders of magnitude
- Several GPU instances available on AWS cloud with highly innovative features/flexibility

## Industry use cases illustrating benefits of using AWS and NVIDIA

## Summary and Recommendations

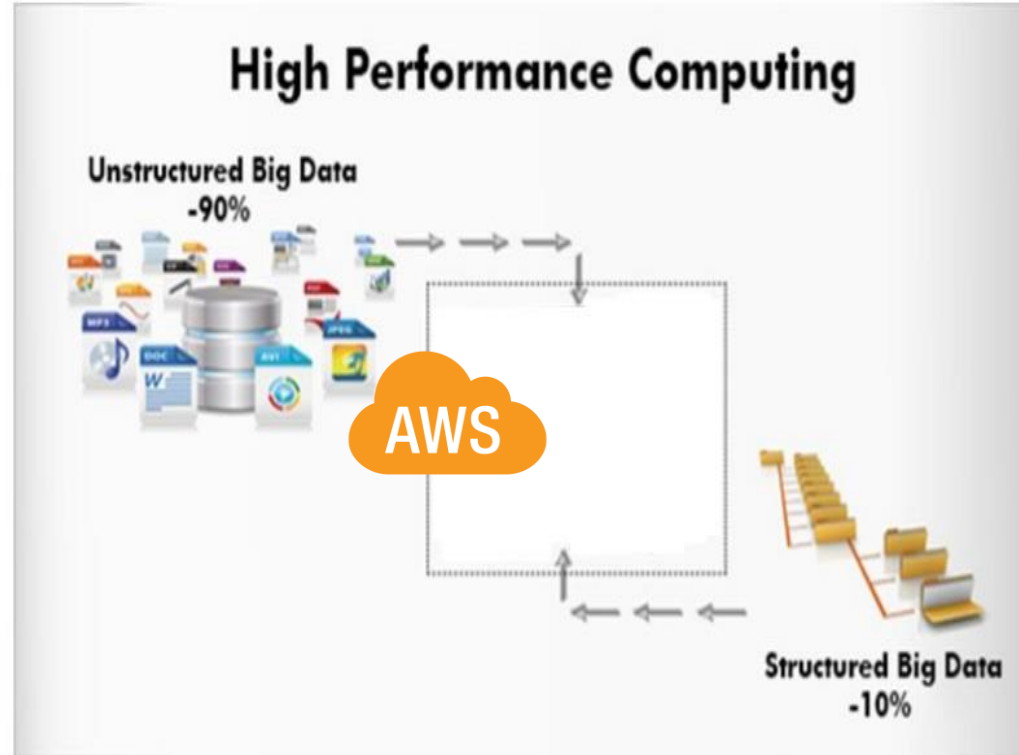
# Intertwined technology trend causing continuing explosion of data volumes



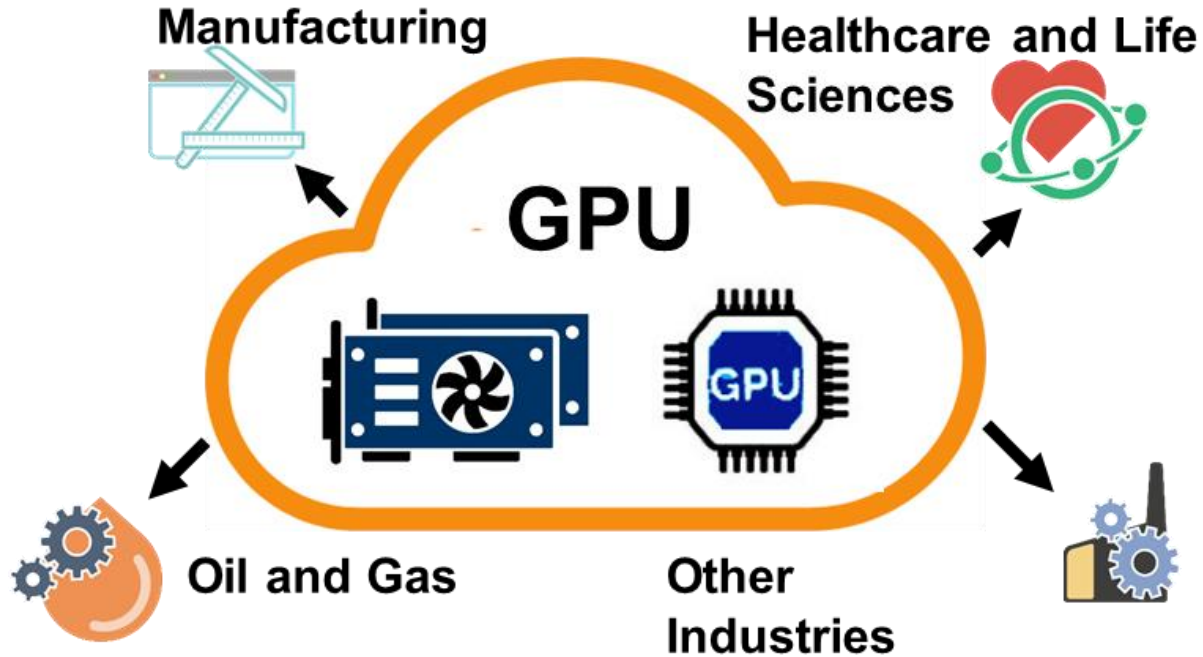
# Blurring of the lines between HPC and Analytics



## Analytics



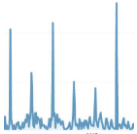
# Rapid adoption of GPU across many industries



# Many Impediments with HPC use on-premises



Expensive to acquire, maintain, and operate on-premises HPC systems and software



Hard to optimize and run applications efficiently (especially spiky workloads) while keeping up with rapid technology refresh cycles to prevent obsolescence



Lack of adequate datacenter space



Implementing security and compliance are challenging or expensive

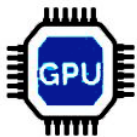


Lack of deep skills to customize HPC deployments and integrate existing workflows.

# Traditional barriers of HPC in the cloud collapsing



Improvements in network bandwidth and latency, security and compliance



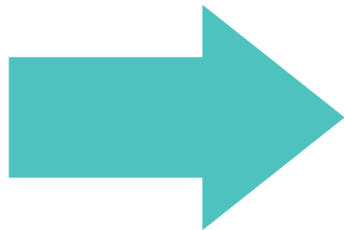
GPU support that can accelerate workloads and allow effective remote visualization



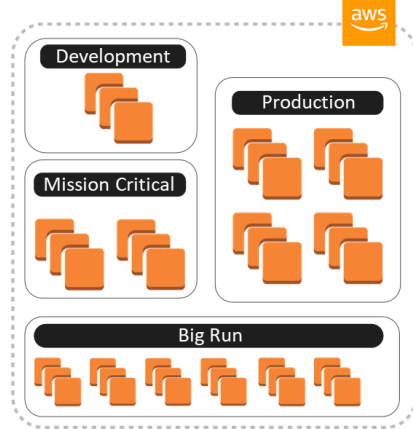
Data replication solutions and container technologies that enable workload portability

# Benefits of Moving On-Premises HPC to AWS Cloud

## On Premises



## HPC in Cloud



## Challenges



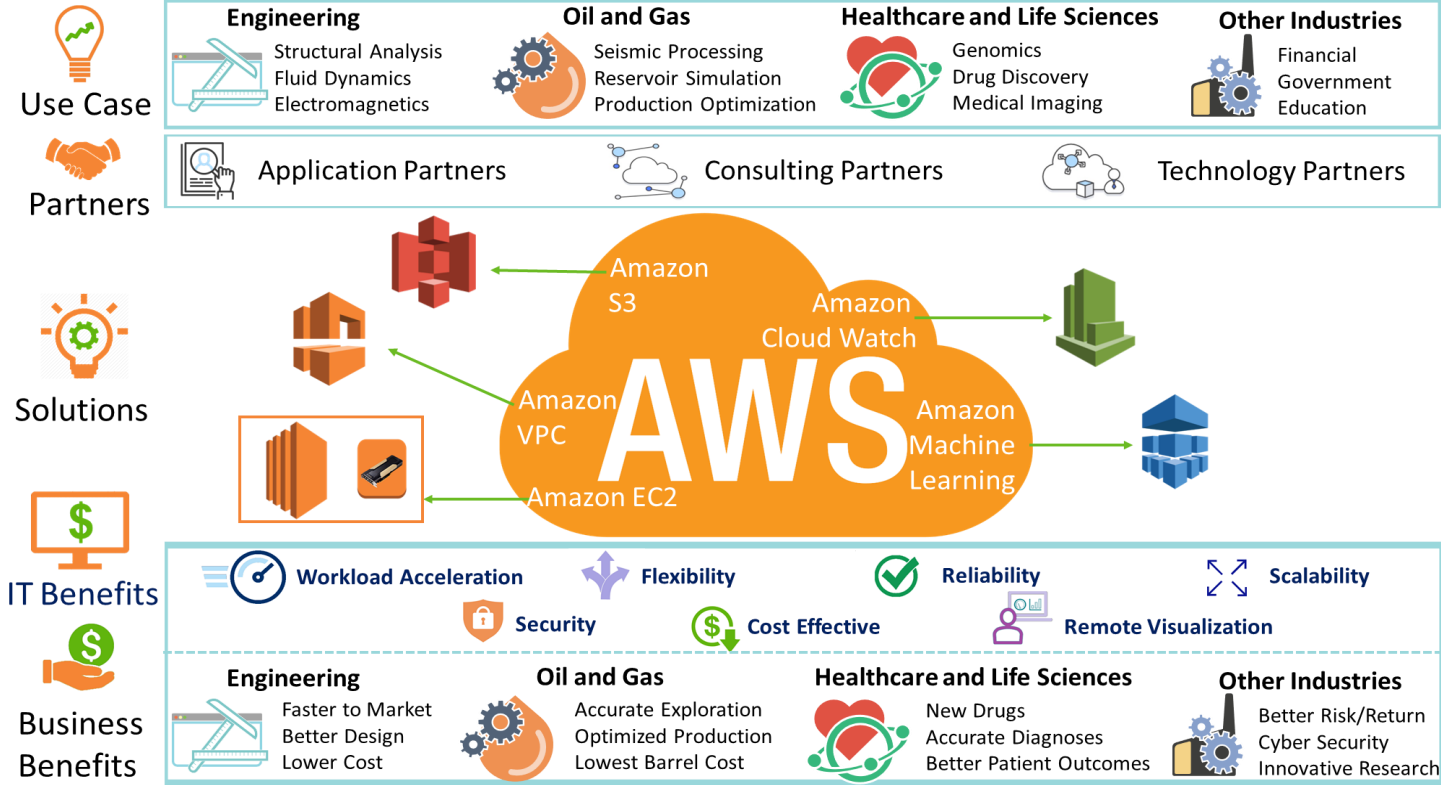
- Expensive to acquire, maintain, operate
- Technology obsolescence
- Lack of datacenter space
- HPC skills shortage

## Benefits



- Minimize wait time
- Reduction in CAPEX, Conversion to OPEX
- On-demand usage-based pricing
- Flexibility and elasticity of resources
- Leverage state of the art GPU/CPU infrastructure
- Benefits from AWS investments in AI/ML/DL

# High Value Use Cases/Benefits of HPC Cloud





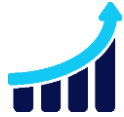
# Compelling Reasons to use AWS and NVIDIA for HPC



Faster Procurement and Provisioning



Scalability and Agility



Improving Efficiency of HPC



Global Collaboration and Remote Visualization

# NVIDIA GPU are accelerating HPC Workflows

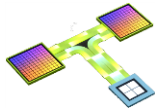


GPU use is rapidly growing in HPC



NVIDIA continues to innovate and invest heavily in the growth of GPU adoption for HPC

NVIDIA provides several innovative capabilities for HPC:



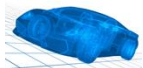
NVLink interconnect fabric



Parallel programming APIs



Acceleration of over 70% of the most popular HPC applications



NVIDIA IndeX

# Amazon Elastic Computing Cloud (EC2) with GPU instances



[Amazon EC2 P3 Instances](#) have up to 8 NVIDIA Tesla V100 GPUs

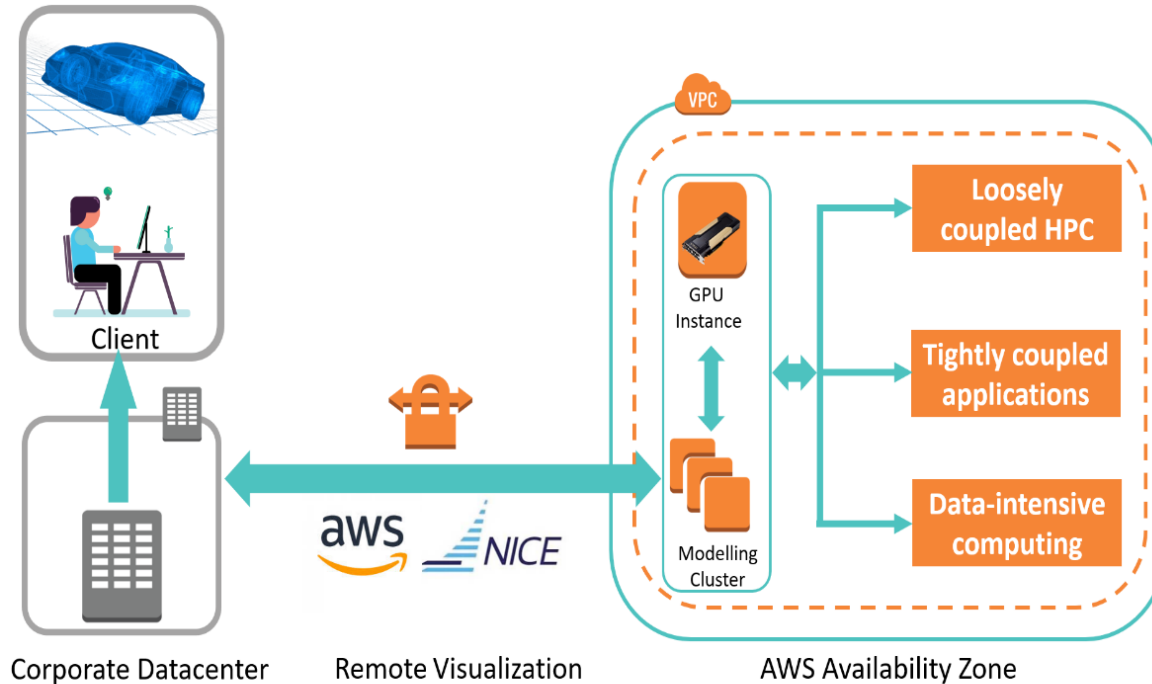


[Amazon EC2 P2 Instances](#) have up to 16 NVIDIA K80 GPUs.



[Amazon EC2 G3 Instances](#) have up to 4 NVIDIA Tesla M60 GPUs

# Key Differentiated Features of the AWS HPC Offering



**Loosely coupled HPC:**

**Tightly coupled applications**

**Data intensive computing**

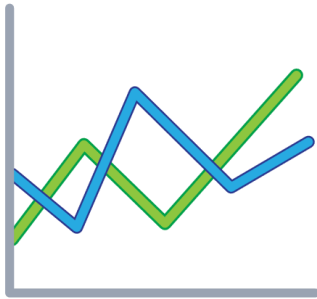
**Advanced remote visualization**

**Container Services**

# Amazon EC2 Flexible Consumption Based Pricing Options

## On Demand

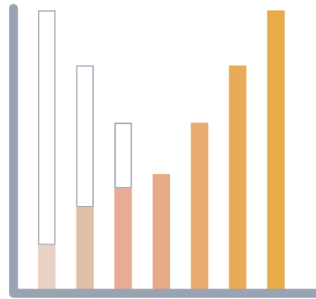
### Spiky Workloads



Pay per use with no long term commitment

## Reserved

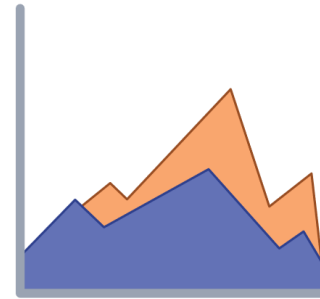
### Committed Steady State Usage



1 or 3 year commitment at significant discount

## Spot

### Time Flexible Stateless Workloads



Spare capacity up to 90% discount

# Case Studies



**Oil & Gas**



**Life Sciences and Healthcare**



**Manufacturing**

# Oil & Gas



## Challenges

- Industry subject to severe ups and downs (boom or bust)
- Finding new sources of hydrocarbon prohibitively expensive
- Severe constraints on IT capital

## Solution

- AWS elasticity lends itself to satisfying cyclical demand
- Run models faster and cheaper, optimize extraction, collaborate with ease, improve safety, improve execution efficiencies by:
  - Reducing CAPEX with many usage-based pricing options (Up to 60% savings in running RTM simulation)
  - Leveraging the GPU/CPU infrastructure on AWS instead of in-house. Insurance against technological obsolescence.

## Tools/Customer Success

- Reduction of co-located datacenter size and power usage by 50%, reduction of physical server inventory by 400 high-end servers and 3200 GPUs (major oil company – Hess)
- HPC/Clustered graphics on AWS, “Pre-Stack Pro” is pre-stack seismic analysis software that combines pre-stack visualization, processing, and interpretation in one (Sharp Reflections)
- Next-gen reservoir simulator running on CPUs and GPUs on AWS: tNavigator (Rock Flow Dynamics), 6X (Ridgeway Kite)

# Life Sciences and Healthcare



<b>Challenge</b>	<ul style="list-style-type: none"><li>• Sorting and analyzing massive amounts of life sciences data to make it actionable information is complicated</li><li>• Siloed legacy systems impede effective collaboration</li><li>• Pressure to develop value-based personal care/therapies</li></ul>
<b>Solution</b>	<ul style="list-style-type: none"><li>• Utilize the current in-house infrastructure and burst into AWS cloud as needed for additional capacity</li><li>• Use the GPU/CPU infrastructure on AWS instead of in-house; providing insurance against technological obsolescence and enhancing collaboration.</li><li>• Leverage AWS/NVIDIA investments in DL frameworks and devices (e.g., AWS Sagemaker and DeepLens) to personalize care/therapies</li></ul>
<b>Tools/Customer Success</b>	<ul style="list-style-type: none"><li>• Ability to solve extreme-scale intractable problems:<ul style="list-style-type: none"><li>○ Massive DNA sequencing using AWS GPU infrastructure of over 270 Billion data points/year (Illumina)</li><li>○ Ability to bring lifesaving drugs to market more quickly by allowing scientists to perform four times as many simulations of new compounds with AWS P3 (high performance GPU) instances (Schrodinger).</li><li>○ Nephele, a platform that allows researchers to perform large-scale analysis of microbial genetics data (NIH)</li><li>○ Largest database of human genomes along with related phenotype and clinical data for genomic research (Human Longevity Inc.)</li></ul></li><li>• Early cancer detection using Deep Learning (Matrix Analytics)</li></ul>



# Manufacturing



## Challenge

- Need for higher accuracy and interdisciplinary analyses are resulting in larger and more complex simulation models
- Hard to visualize large result datasets with high fidelity remotely
- Smaller suppliers do not have the resources, skills, and business justification to deploy expensive on-premises HPC clusters

## Solution

- AWS offers virtually unlimited compute capacity, so companies can bypass on-premises capacity constraints to perform complex simulations
- Leverage the GPU/CPU infrastructure on AWS instead of re-creating in-house
- Advanced remote visualization tools provide users additional avenues to interactively view the dataset in high fidelity

## Tools/Customer Success

- AWS has partnered with Rescale to provide cloud-based high-performance computing simulation platform. Rescale with its deep HPC skills and expertise allows users to instantly scale out complex simulations and analytics
  - Boeing (150 lb. weight reduction translates to \$180 Mil. Savings)
  - Boom (Testing reduced from months in wind tunnel to hours of CFD)
- TLG, an aerospace company saw 75% cost reduction per CFD simulation (STAR-CCM+ application) using Amazon EC2 instances.
- FEM, CFD, electromagnetics simulation software (ANSYS) and many other CAE apps leverage AWS and GPUs for better performance and scale extremely well.

# Summary and Recommendations

- HPC has become indispensable across many industries
- Lines between HPC, analytics, and deep learning continue to blur
- Users are migrating to the cloud at almost four times the growth rate of HPC
- HPC workflows accelerated with GPUs by orders of magnitude
- AWS and NVIDIA both pioneers and leaders ***fueling the HPC clouds with GPUs***

## **Organizations should actively consider AWS/NVIDIA HPC solutions to:**

- Augment their existing on-prem infrastructure with bursting
- Accelerate and scale workloads economically and reduce CAPEX
- Leverage the GPU/CPU infrastructure on AWS instead of re-creating in-house
- Run their entire workflow from data ingest to compute to visualization
- Leverage deep HPC skills at AWS, NVIDIA, and the growing, vibrant partner ecosystem
- Benefit from AWS and NVIDIA investments in machine learning and deep learning

# Agenda

## Overview of AWS HPC Solutions

### Cabot Partners

- Market Overview – HPC applications using GPUs
- Predominant HPC Use Cases for GPU-based compute
- Summary

# Additional Resources

Visit:

<https://aws.amazon.com/hpc/>

## High Performance Computing (HPC)

Imagine the problems you can solve with virtually unlimited infrastructure

[Create a Free Account](#)  
[Contact Sales](#)

[Overview](#) [Getting Started](#) [Resources](#)

High Performance Computing (HPC) allows scientists and engineers to solve complex, compute-intensive problems. HPC applications often require high network performance, fast storage, large amounts of memory, very high compute capabilities, or all of these. AWS enables you to increase the speed of research and reduce time-to-results by running HPC in the cloud and scaling to larger numbers of parallel tasks than would be practical in most on-premises environments. AWS helps to reduce costs by providing CPU, GPU, and FPGA servers on-demand, optimized for specific applications, and without the need for large capital investments.

**Video: The benefits of running HPC on AWS**

### Benefits of Running HPC on AWS

- Enable quick discoveries**  
Instantly launch or scale up High Performance Computing clusters on AWS. By eliminating job queue times and scaling your cluster as high as needed, when needed, you can reduce the time to market or publication.
- Increase productivity**  
Focus on applications and research output over infrastructure maintenance and upgrades. When AWS upgrades hardware, you can gain access instantaneously — simply rewrite your cluster configuration file and reboot to move to the latest hardware.
- Leverage flexible architectures**  
Let your research dictate infrastructure, not the other way around. With the flexible configuration options AWS provides, you can start with your hypothesis and create HPC clusters that are optimized for your unique application requirements — GPU today, CPU tomorrow.

# Thank you