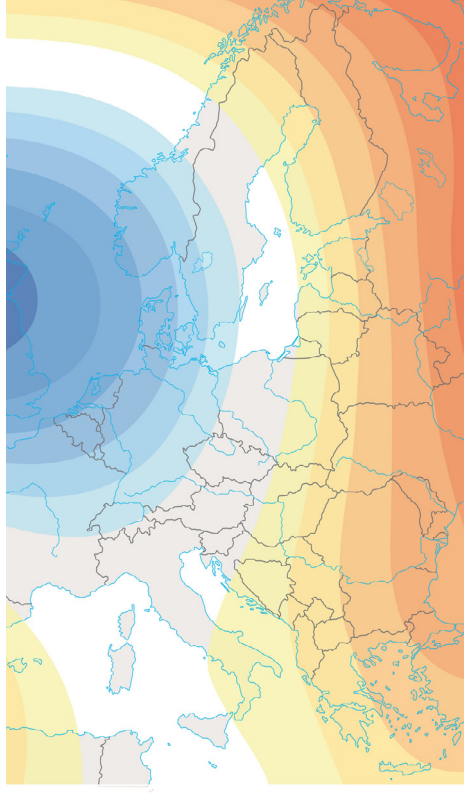aws | NVIDIA.

Accelerate discovery with virtually unlimited HPC infrastructure

Leverage high performance computing (HPC) on AWS powered by NVIDIA GPUs to solve your most complex problems, from scientific discovery to predictive maintenance, while minimizing costs.

HPC use in the cloud is on the rise

Fastest growing markets
Weather (28.1% CAGR), **Government Labs** (24.5% CAGR), and **Computer Aided Engineering** (20.7% CAGR) are the three verticals with the highest projected growth rates for HPC in the cloud.

17.6%
Forecasted growth of cloud usage for HPC over the next 5 years (vs. 7–8% growth of HPC on premises)

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

(Source: Hyperion Research)

Benefits of running GPU-accelerated HPC workloads on AWS

Faster Time to Results
NVIDIA GPU-powered HPC workloads on AWS help you run more of your simulations in parallel to achieve faster time to results.

Access the right tools and services
Use purpose-built and pre-configured tools and services from AWS, powered by NVIDIA, to accelerate workload lifecycle and increase innovation.

Drive efficiency
AWS on-demand infrastructure and services, powered by NVIDIA GPUs combined with software support like the NVIDIA HPC SDK, means teams can run complex workloads with greater efficiency, performance, and accuracy.

No problem is too big
Scalable, agile AWS infrastructure combined with high-performance GPUs give scientists, manufacturers, and designers the freedom to ask and answer questions that weren't possible before.

Shorten time to market
Replace expensive physical model development with virtual models with GPU-optimized AWS instances to move from idea to market faster.
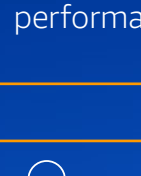
Improve secure collaboration
Your globally distributed teams can collaborate and integrate findings faster while maintaining security.

Meet the future
Adding GPU-accelerated computing prepares you to augment your HPC workloads with artificial intelligence, because standard machine learning tools like TensorFlow and PyTorch can leverage GPU-accelerated computing for faster time to results.

AWS and NVIDIA help overcome cross-industry workload challenges

Workload fluctuations
Sudden changes in workload demand make it difficult to predict infrastructure resources upfront. Limited resources during workload spikes can hinder the timelines of critical computations, compromising time to results.

Length of product development cycles
Managing people, software licenses, limited HPC capacity, and oversubscribed systems can hinder speed to market for products and services. The growing demand for shorter development cycles requires accessible services and tools on demand.

Long-term commitments
Without dynamic provisioning of compute infrastructure, organizations can end up in long-term architecture lock-in with obsolete technology, fall behind market needs, and become impeded when workload requirements change.

Collaboration constraints
In heavily regulated industries such as Financial Services, end users require access to a reliable and easily-accessible infrastructure for seamless global collaboration, while remaining compliant.

Compute capacity limitations
Infrastructure constraints on scale and capacity can affect an organization's ability to address mission-critical questions and make them reluctant to consider new use cases or applications.

High upfront costs
Running compute-intensive workloads on premises often comes with significant capital expenditure and maintenance costs. High upfront costs keep HPC out of reach for smaller businesses and prolong decision-making processes for innovative and strategic investments.

Unpredictable and growing compute needs
With the proliferation of data, innovative and demanding HPC workloads require access to the latest high-performance GPUs combined with the right tools to try new approaches and support continuous innovation.

Did you know?

Lyft's autonomous vehicle (AV) division runs millions of simulations each year on AWS, powered by NVIDIA, to improve the performance and safety of self-driving vehicles.

Siemens Energy built a Digital Twin to support predictive maintenance of their power plants which helped them reduce the frequency of planned shutdowns while maintaining safety.

Multinational engineering company FLSmidth reduced its physics-based simulation time from months to days by using NVIDIA powered HPC on AWS.

Find out how HPC on AWS can accelerate your unique workload ›