HPC on AWS for Life Sciences

Accelerating the pace of discovery

Accelerate time to insight and go beyond the limitations of on-premises solutions by employing HPC on AWS infrastructure designed to satisfy the most highly regulated and demanding workloads of Healthcare and Life Sciences. Run your leading-edge applications on the most mature and reliable cloud platform offering an extensive portfolio of purpose-built HPC tools and services that advance the standards and fuel innovations in Healthcare and Life Sciences.

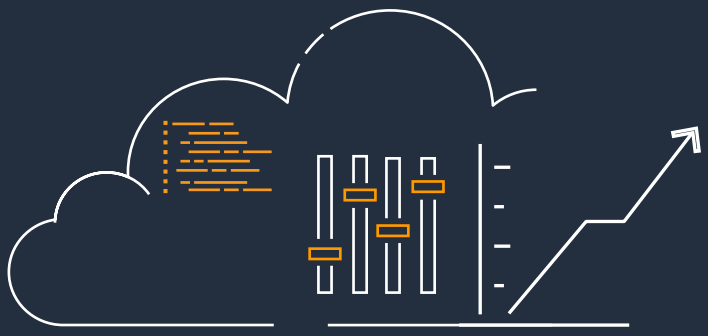
AWS delivers among the broadest and deepest compute, networking, storage, security, cloud orchestration, visualization technologies, and Partner Network to enable customization and performance optimization for a broad range of Healthcare and Life Science workloads. Elevate workloads across Genomics, Computational Chemistry, Structural Biology, Data Science, Modelling and Simulation, and Imaging while mitigating the up-front investments of on-premises HPC solutions.

With AWS, organizations can launch an HPC cluster in minutes, paying for only what they use. Run your own HPC applications or choose from a range of commonly used and verified applications from trusted partners via the AWS Marketplace.

9.1%

Bio-sciences HPC estimated growth by 2023 (from $1.35B to $1.92B)[1]

The AWS advantage



HPC on AWS provides a number of key benefits when compared to an on-premises HPC environment, including:

- **Rapid deployment** – With AWS you can spin up an HPC cluster in minutes allowing you to react quickly to changing business demands.
- **Compute elasticity** – You can have an HPC system with as little as a single compute node stretching all the way to 1 million cores or more, rapidly reducing the time to results.
- **Flexibility of configuration** – Your cloud-based HPC can consist of a mixture of differing compute instances to meet the needs of a variety applications and workloads. This means there's no need to compromise with a one size fits all approach, and instead you can select the best compute profile to meet your HPC workload needs.
- **Purpose-built HPC tools and services** – AWS offers a range of tools and services designed specifically to support HPC workloads such as AWS Batch, AWS ParallelCluster, EFA low latency networking, Amazon FSx for Lustre and NICE DCV, for set up and operation of your HPC cluster in AWS.
- **No need for data centers** – Running HPC workloads in the cloud means no further need for expensive data center facilities, main power circuits and air conditioning systems. It also shifts the burden of hardware procurement, maintenance, refresh cycles and OS software licensing to AWS.
- **Enables productive remote workplace** – Users can gain access to high performance remote virtual desktop computing sessions that can provide similar performance to that of workstation computers used in engineering or design offices. For staff working from home, a virtual desktop infrastructure (VDI) session powered by NICE DCV or Amazon AppStream allows for demanding desktop applications to be used without compromising the performance or security of your assets and intellectual property.

AWS benefits

Accelerated time to insight

AWS brings one of the broadest and deepest HPC services and technologies for Healthcare and Life Science workloads in the cloud. Meet exponential data growth with scalable storage services and leverage secure collaboration platforms such as the AWS Data Exchange (ADX) to accelerate time to insight. Moreover, AWS enables machine learning to analyze large public and private data sets available through the Registry of Open Data (RODA) and AWS Data Exchange, respectively, to help accelerate research and innovation.

Faster results

Get instant access to the infrastructure capacity needed for running demanding HPC applications. Deploy an HPC cluster on AWS in minutes, and eliminate the wait times, long job queues, and time it takes to acquire new hardware. The possibility of rapidly deploying resources enables prompt response to changing business demands, especially when compared to the time and effort associated with acquiring and deploying a new HPC cluster inside a data center.

HPC without constraints

AWS offers a broad range of compute, networking, and storage technologies explicitly designed for HPC applications, allowing for much greater flexibility than fixed on-premises systems. Quickly create a new HPC system or burst suitable workloads from on-premises to secure resources that meet user groups' computing needs. AWS is also bringing forward the next generation HPC technologies, including quantum computing, providing access to a complete quantum development, simulators, and quantum hardware to run quantum applications.

Designed for price-performance

Meet changing workload demands and only pay for HPC resources required at a given time, and after completion, decommission them in minutes. Leverage AWS services and HPC clusters to right-size the HPC infrastructure requirement during peak workloads and avoid costly delays to project and production schedules. Drive greater ROI for your organization's total HPC spend by shortening time to results. AWS has helped organizations reduce simulation run times from weeks to hours, thus significantly impacting their time to market.

Tailored tools and services

AWS brings a wide variety of HPC instance types and services tailored to specific scientific requirements. Find solutions for all Healthcare and Life Science workloads with access to a large array of HPC technical and scientific expertise and a rich ecosystem of partners. HPC on AWS ensures Life Science and Healthcare organizations meet and exceed the pace of scientific demands.

Healthcare and Life Sciences Workloads

Genomics

HPC in Genomics:

AWS enables secure analysis, storage and sharing of genomic data through HPC solutions that support each step of the genomics value chain. Use AWS services to build and run data applications that are retrospective, real-time or predictive to accelerate the journey from genomic data to genomic insight. Build incrementally adding use cases and increase scale as you go.

What you need to know:

AWS enables secure analysis, storage and sharing of genomic data through HPC solutions that support each step of the genomics value chain. Use AWS services to build and run scalable and flexible data applications that are retrospective, real-time or predictive to accelerate the journey from genomic data to genomic insight. Reduce data processing times from week to hours and obtain early insights into diseases such as Cancer, Cystic Fibrosis, and Alzheimer's.

AWS Technologies for Genomics:

Amazon FSx for Lustre, AWS Batch, GATK, Cromwell, NextFlow, Dragen F1, Parabricks, Amazon EC2 GPU instances (G4dn, P4d), AWS DataSync, AWS Storage Gateway, AWS Direct Connect, AWS Step Functions

Examples:

- SevenBridges
- Baylor College of Medicine

Imaging

HPC in Imaging:

AWS brings one of the broadest and deepest set of instance types, scalable storage solutions, and advanced machine learning and data integration services to meet imaging workload demands to elevate image analysis and facilitate long-term data retention. Leverage HPC to safely store, transfer and visualize your data and keep up with future demands.

What you need to know:

HPC on AWS supports your growing compute and storage requirements across various imaging workloads spanning basic research, data transfer, instrument integration, diagnostic medical imaging, high content screening, digital pathology, MRI, PET, X-Ray, and CT Scans.

AWS Technologies for Imaging:

Amazon EC2 GPU instances (G4dn, P4d), AWS Batch, Amazon FSx for Lustre, AWS ParallelCluster, AWS DataSync, AWS Storage Gateway, AWS Direct Connect, Amazon SageMaker and NICE DCV

Examples:

- Arterys
- Proscia

Computational Chemistry and Structure Based Drug Design

HPC in Computational Chemistry and Structure Based Drug Design:

AWS HPC offers instant access to almost unlimited computing resources to accelerate Structure Based Drug Design. Combine the rapid advancement of protein structure solutions and faster algorithms for describing and representing flexible 3D molecular models with HPC computing resources to grow and accelerate workloads.

What you need to know:

HPC on AWS improves speed, accuracy and scale across workloads such as Virtual Screening, Molecular Dynamics, Quantum Mechanics and 3-D Structure Solution. Meet your computing demands and avoid the unnecessary costs of on-premises solutions by running virtually unlimited HPC resources and scaling back down to normal loads when the task is finished.

AWS Technologies for Computational Chemistry and Structure based Drug Design:

Amazon EC2, Amazon EC2 Spot, AWS Batch, AWS ParallelCluster, Amazon FSx for Lustre, Amazon EMR, NICE DCV

Examples:

- OpenEye Scientific
- InhibOx

Modeling and Simulation

HPC in Modeling and Simulation:

Get instant access to almost unlimited computing and storage resources on AWS to meet Modeling and Simulation workload demands. This helps you gain deeper insights into drug candidates' properties and effects and use them to inform clinical trial design and predict trial outcomes. Leverage on-demand HPC resources to facilitate emerging initiatives such as Digital Twins that analyze and predict future medical states at the atomic, cellular, and biological level.

What you need to know:

AWS enables you to satisfy computational requirements and facilitate data standardization and interoperability across Modeling and Simulation workloads such as Computational Fluid Dynamics (CFD), Pharmacodynamics (PD), Pharmacokinetics (PK), Clinical Trial, and Systems Biology Simulations.

AWS Technologies for Modeling and Simulation:

Amazon Sagemaker Notebooks, AWS ParallelCluster, AWS Deep Learning AMI, Amazon EMR, Amazon EC2 GPU instances (G4dn, P4d), Amazon Comprehend, Amazon Comprehend Medical and AWS Batch

Examples:

- Bristol Meyers Squibb
- Numerate

Data Science

HPC in Data Science:

Enable Data Scientists to shorten time to insight uncovered from structured and unstructured data and gain a competitive edge by mining Gigabytes to Petabytes of diverse scientific, medical, and healthcare data securely.

What you need to know:

Derive insight from Data Science workloads by accessing the broadest and deepest set of tools supporting integrated data processing, wrangling, visualization, model deployment, model management, and explainable AI.

AWS Technologies for Data Science:

Amazon EC2 GPU instances (G4dn, P4d), AWS Deep Learning AMI, Amazon SageMaker, Amazon SageMaker Ground Truth, Amazon Comprehend Medical, AWS Lake Formation, AWS Step Functions, AWS Glue, AWS Batch

Examples:

- Core Informatics
- Bluedot

Data Lakes

HPC in Data Lakes:

Organizations are focusing on digital transformation, machine learning and data-driven drug discovery strategies to accelerate time to insight. Consolidating data silo's by standardizing data management policies and combining data and information across multiple scientific disciplines to create Data Lakes and Data Houses.

What you need to know:

AWS offers you all the instruments you need to take advantage from data lakes analytics. Enabling you to harness more data, from more sources, in less time. Empowering you to collaborate and analyze data in different ways leads to better, faster decision making.

AWS Technologies for Data Lakes:

Amazon S3, AWS Glue, Amazon Athena, Amazon Redshift, Amazon DynamoDB, Amazon Elasticsearch Service, Amazon CloudWatch, AWS Lake Formation

Examples:

- Philips
- Biogen

Key products and services

A High Performance Computing solution is configured based on a number of core technology components which can be tuned based on the customers' specific workload needs.

Compute

AWS provides a broad range of compute services and instance types to suit the customers' needs and supports all major technology vendors including Intel, Nvidia, Arm and AMD. Given this broad range customers can choose the instances which is best suited to their workload (e.g., accelerated instances are often used by Genomics/ Modelling and Simulation customers).

- **Compute Optimized** – C5/C5n/M5zn (Intel), C6g (Arm/ Graviton2) instance options for applications that require increased CPU performance.
- **Memory Optimized** – R5/R5n (Intel), R5a (AMD), X1/ X1e High Memory (Intel 4 socket) instances for customers that require larger memory requirements.
- **Accelerated Optimized** – P2/P3/P4d (Nvidia GPU), G3, G4 (Nvidia), Inf1 (Inferentia CPU), F1 (FPGA) instance options for customers that require GPU instances for GPU workloads and remote visualization, Inf1 and F1 Instances for customers who have applications that can be programmed to run explicitly on programmable instance types.
- **Storage Optimized** – I3/I3en/D2/H1 (Intel) various options for high performance local IO. Storage optimized instances are tuned for customers with large IO needs.
- **AWS Nitro System** – The Nitro System offloads virtualization layer to a dedicated Nitro card which enables AWS to provide near bare metal performance with the benefits of virtualization.
- **Quantum Computing Services** – AWS Bracket provides a fully managed quantum computing service including developer framework, tools, simulators and access to physical quantum computers on a pay as you go basis to allow development and deployment of next generation quantum applications.

Storage

AWS provides a range of storage solutions to help customers design the IO fabric to suit the needs of their application including local scratch/temp, central filesystems and long-term data repository (R&D Data Lake).

- **EBS/Local IO** – Elastic Block Storage and ephemeral storage options (instance dependent) provide required storage capacity and performance needs for applications.
- **FSx for Lustre** – Provides a high performance parallel file system and enables multiple compute instances to share a performant storage cache for IO demanding workloads.
- **Amazon S3** – S3 can be used as a foundational data lake to store research data. S3 also includes services such as Amazon Glacier and Amazon Glacier Deep Archive, for long term retention of very large data sets.
- **Amazon Elastic File Systems (EFS)** – Provides a simple, scalable, fully managed elastic NFS file system and is typically used as a standard performance file system for performant shared storage needs.

Networking

Low latency, high-bandwidth networking is typically required to run HPC workloads, AWS has developed a number of technologies to improve performance and enable customers to run latency sensitive workloads on AWS.

- **Elastic Fabric Adaptor (EFA) –** Is a high bandwidth low latency interconnect supporting MPI protocols used by distributed applications.
- **Placement Groups –** Ensures compute resources are deployed physically close to reduce network hops and therefore latency.
- **Enhanced Networking –** Enables you to configure your HPC environment with best practices for HPC deployments including increased packets-per-second performance and up to 400Gb/s throughput.
- **Amazon VPC –** Enables you to provision locally isolated sections of the AWS Cloud so you can launch AWS resources in a virtual network that you define.
- **AWS Direct Connect** – Provides customers with a dedicated network connection performance tuned to enable data transfer required for HPC workloads.

Automation and Orchestration

AWS provides a range of automation and orchestration tools to simplify the process of running HPC workloads.

- **AWS Batch –** Is a managed service to run batch jobs automatically for clients.
- **AWS ParallelCluster –** Is an AWS supported open-source cluster management that makes it easy for you to deploy and manage HPC clusters on AWS.
- **NICE EnginFrame –** Is an advanced web front-end for accessing and managing HPC clusters and deploying applications.
- **Amazon EKS –** A container management framework for deploying and managing AI/ML environments based on containers.
- **AWS Step Functions** – Enables complex HPC workflows to be managed and orchestrated.
- **AWS Lambda** – Is code execution as a service, providing serverless compute.

Frameworks

- **Scale-Out Computing (SOCA)** – Provides a ready-to-use simple to deploy multi-user HPC environment that can be tuned for specific industry workflows.
- **Biotech Blueprint Quick Starts** – Helps building a preclinical, cloud-based infrastructure on the AWS Cloud.

Visualization

AWS has two tools to support remote visualization which enables customers to manage an end-to-end HPC experience in the cloud without having to transfer data back to local workstations.

- **NICE DCV** – High performance remote visualization protocol.
- **Amazon AppStream 2.0** – Stream applications from AWS to any computer, including Chromebooks, Macs, and PCs.

Management

Using foundational AWS services, you can manage an HPC environment with granularity to enable assessment, audit, governance, compliance and monitoring of resources.

- **AWS CloudTrail** – API usage tracker for governance, compliance, operational auditing and risk.
- **Amazon CloudWatch** – Monitoring and observability service.
- **AWS Config** – Enables you to assess, audit and evaluate the configurations of your AWS resources.
- **AWS Cost Management** – A range of tools to help you manage your run cost of HPC environments and workloads.
- **AWS Budgets** – Improve planning and cost control with flexible budgeting and forecasting.
- **AWS Identity and Access Management** – Securely controls access to AWS services, and enables you to centrally manage users, security credentials, and permissions.

Customer Services/Assistance

AWS offers a range of support plans that provide access to tools and expertise that support the successful deployment and ongoing operational health of your AWS solutions. All AWS support plans provide 24/7 access to customer service, documentation, technical papers, and support forums. You can choose a support plan that best aligns with your AWS use case, providing you resources to plan, deploy, and improve your AWS environment.

Partner Ecosystem

The AWS Partner Network (APN) is a global community of partners with skills and experience in designing, deploying, migrating and supporting solutions running in the AWS Cloud. Partners can provide additional tools, services, support and training in order to maximize your effectiveness and performance when running workloads in AWS.

HPC Application ISVs:
AWS Partners that provide application software for workloads such as Computational Fluid Dynamics (CFD), Molecular Modeling, Reservoir Modeling, and Weather Modeling.

Partners: Ansys, Siemens, Altair, Cadence, Synopsys, OpenEye Scientific, CFD Direct, OnScale, S-Cube

HPC Management:
AWS Partners that provide a fully managed cloud HPC environment and Partners that provide solutions such as end-to-end cluster provisioning, deployment, management, and support for customers to deploy their HPC workloads on AWS.

Partners: Rescale, Altair, IBM, SchedMD, Zenotech, Ronin, TotalCAE, Core Scientific

Foundational Technology:
AWS Partners that provide enabling technologies such as processors, accelerators, and operating systems for customers to run their HPC workloads on AWS.

Partners: Intel, Nvidia, AMD, ARM

Consulting partners:
System integrators and strategic consultancies, that help customers of all types and sizes accelerate their HPC journey to the cloud.

Partners: Six Nines IT, Ronin

Case Studies:

- University of Sydney's Wildlife Genomic group protects wildlife
- Core Informatics
- Numerate Discovers Drug Candidates Five Times Faster by Running on AWS
- Arterys
- Fred Hutch Microbiome Researchers Use AWS to Perform Seven Years of Compute Time in Seven Days
- Numerate Discovers Drug Candidates Five Times Faster by Running on AWS
- Relay Therapeutics Uses AWS to Accelerate Drug Discovery
- Biogen's Genomics Solution Uses AWS to Advance Neurological Disease Research
- Quantitative Biology Center Drives New Genomics Research Faster Using AWS
- Gigadocking™ in Orion™ Molecular Design Platform Rapidly Identifies Novel Chemical Entities for GPCR Targets

Whitepapers, eBooks, and Infographics:

- Whitepaper: Lowering Time-to-Results with Elastic Fabric Adaptor
- Whitepaper: The Cloud Steps up to Tightly Coupled HPC Codes
- Whitepaper: Challenging Barriers to High Performance Computing in the Cloud
- Analyst Spotlight: Computational Evaluation of Commercial Cloud HPC with a Global Atmospheric Model
- Whitepaper: Democratizing High Performance Computing
- Achieving optimal price/performance for your HPC workloads on AWS
- Hyperion Research Technology Spotlight: Smart Orchestration Speeds HPC Workflows in the Cloud
- Whitepaper: HPC on AWS Redefines what is possible
- Whitepaper: What a TCO Analysis won't tell you

Getting Started Projects:

- Create an Elastic HPC Cluster using AWS ParallelCluster
- Create an End-to-End HPC Environment
- Getting Started with Amazon FSx for Lustre
- Getting started with Elastic Fabric Adapter(EFA)

Start your AWS journey for running HPC workloads with the following technical resources, contact your AWS account team or use the "Contact Us" page to reach us directly.

https://aws.amazon.com/hpc/hcls/

1 US Drug approvals hit a 21 year high in 2018, with 105 NDAs. The EU approved 92 new drugs, and China will be speeding up approvals in the world's second largest market. However, only 35% of the US approvals were for big pharma (a continually decreasing trend), and R&D Internal Rate of Return (IRR) for the top 12 drug makers was at an 8-year low of 1.8% in 2019 (from 10.1% in 2010)