aws | **NVIDIA.**

AWS for Genomics Featured Partner: NVIDIA

Accelerated genomics analysis, without compromise

Analyze a full 30x human genome in as little as 25 minutes

Reduce time to discovery - from clinical testing when every second matters, to maintaining turnaround times during periods of increased throughput, time is of the essence. Accelerating Genomics analysis is possible with a robust portfolio of purpose-built AWS and NVIDIA Clara Parabricks solutions that enable scientists to process more samples, run more complex analyses, and query at scale.

The need to process larger datasets and run more complex genomic applications highlights limitations of traditional infrastructure, requiring organizations to invest in costly expansions if they're to handle peak volumes and maintain turnaround times.

Challenges faced by genomics organizations

1. **Turnaround time:** in clinical settings when every minute matters, clinicians need to reduce analysis time from days down to hours.

2. **Cost:** traditional infrastructure is becoming more expensive as demands on the system increase.

3. **Accuracy:** speed and cost are secondary to ensuring consistent accuracy and reproducible results, especially when influencing a clinical decision.

4. **Ease of use:** data analysis and manually provisioning tasks are causing barriers in scaling.

5. **Scalability:** the growing scale of genomic data poses challenges in both the transfer of data from the sequencer to the cloud in a secure and timely manner, and the growing costs associated with storage. In many cases, organizations are required to store genomic data for extended periods of time, further increasing storage costs.

6. **Security & compliance:** when handling Protected Health Information (PHI), data security is paramount.

Removing barriers with NVIDIA Clara Parabricks on AWS

To help genomics organizations overcome barriers in data analysis, featured AWS for Health Partner NVIDIA offers its accelerated Clara Parabricks genomics analysis framework on the AWS Marketplace for easy and scalable deployment. Uniting security and scalability with the speed and accuracy of Clara Parabricks, part of the AWS for Genomics curated portfolio of industry tools, organizations around the globe can readily deploy cost-saving analysis tools for genome, exome, and panels for alterations of both germline and somatic nature.

Launch Parabricks on AWS Marketplace in minutes

aws marketplace

Why Clara Parabricks on AWS?

Accelerate turnaround times: using AWS' cloud-based infrastructure to speed up genome analysis.

Enhance interoperability: with access to over 50 publicly available tools to help organizations build their own pipelines or use GATK best-practice pipelines within the Clara Parabricks framework.

Improve accuracy: Clara Parabricks' commonly used genomics tools provide consistency of results that allow researchers the freedom to switch from more traditional infrastructures to a cloud-based environment.

Reduce analysis costs: leverage AWS tiered compute offerings to reduce costs by up to 90%.

Scale globally: deploy Clara Parabricks, currently available across 22 AWS regions globally, and take advantage of AWS' global footprint as the largest cloud provider.

Ensure security and compliance: AWS supports more security standards and compliance certifications than any other offering, helping genomics organizations achieve compliance globally and securely collaborate.

In practice: how to leverage NVIDIA Clara Parabricks on AWS

Population whole genome and exome studies
When you need population wide analyses at low cost, and access to germline and somatic variant callers.

The solution:
G4dn.12xlarge Spot Instance with four NVIDIA T4 Tensor Core GPUs.

The results:
Analyze a whole genome in under 76 minutes and a 200x Tumor/Normal exome pair in under 25 minutes, for less than $3. Run multiple analyses in parallel to dramatically accelerate speed-of-results in large-scale studies, while keeping costs affordable.

Urgent care whole genome and exome analysis
When you need accurate results as quickly as possible, such as in the NICU/PICU.

The solution:
P4d.24xlarge On-Demand Instance with eight NVIDIA A100 Tensor Core GPUs.

The results:
Analyze a full 30x human genome in around 25 minutes, for less than $15.

Large volumes of gene panel samples
When you need high throughput at low cost.

The solution:
G4dn.12xlarge Spot Instance with four NVIDIA T4 Tensor Core GPUs.

The results:
Analyze the full 50x exome in just 10 minutes, for as little as $0.75 (On-Demand) or $0.30 (Spot Instance).

Costs correct as of November 2021, for AWS US East (N.Virginia) region.

> "The management and analysis of genomic data at the scale needed to power large-scale genome studies requires computational and storage infrastructure that exceeds the capacity of most institutions. NVIDIA on AWS provides the resources that allow organizations to focus on the science, and the ability to collaborate with peers in a secure and compliant environment."
>
> Richard Daly, CEO, DNAnexus

Leverage AWS compute tiers to reduce costs and time

### Accelerating analysis with Amazon EC2 P4d instances

AWS offers varying storage tiers and instances to help genomics organizations optimize for what matters most to them.

By running Clara Parabricks on more powerful instance types like the Amazon EC2 P4d, which use the latest NVIDIA A100 Tensor Core GPUs to deliver industry-leading high throughput and low latency networking, organizations can reduce the time needed for analysis of a 30x genome to as little as 25 minutes.

The AWS Cloud enables genomics research labs to use only what they need, when they need it. Amazon EC2 Auto Scaling helps you maintain application availability and allows you to automatically add or remove EC2 instances according to conditions you define. You can also use the dynamic and predictive scaling features of EC2 Auto Scaling to add or remove EC2 instances. Dynamic scaling responds to changing demand and predictive scaling automatically schedules the right number of EC2 instances based on predicted demand. Dynamic scaling and predictive scaling can be used together to scale faster. AWS offers storage and compute offerings at a variety of price points. Services such as Amazon EC2 Spot Instances offer up to a 90 percent discount compared to on-demand compute prices.
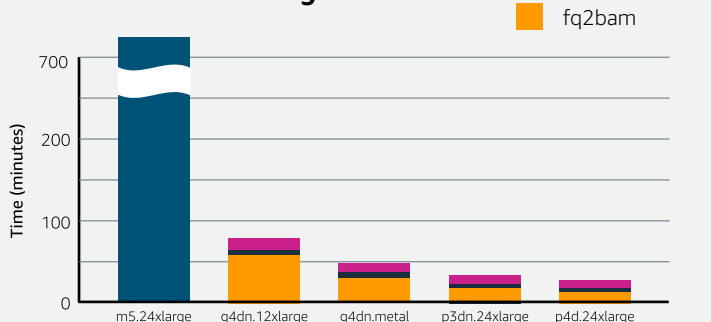
Comparison of execution time for running NVIDIA Clara Parabricks germline pipeline on a sample dataset on G4, P3 and P4 Amazon Elastic Compute (EC2) Cloud GPU-enabled instances, as well as a baseline CPU instance. Times are provided starting with FASTQ through Variant Call Format (VCF) generation.

Example costs for NVIDIA Clara Parabricks on AWS

Organizations looking to reduce costs can leverage Amazon EC2 Spot Instances to run fault-tolerant implementations of Parabricks for up to 90% off. For example, organizations can analyze a 100GB full human genome at 30x coverage for as little as $2.58.

Software and AWS instance type	GATK germline baseline pipeline on a CPU-based Amazon EC2 (m5.24xlarge)		Clara Parabricks on an NVIDIA T4 GPU on Amazon EC2 (g4dn.12xlarge)		Clara Parabricks on an NVIDIA A100 GPU on Amazon EC2 (p4d.24xlarge)	
Time taken	30 hours		76 minutes		25 minutes	
Infrastructure cost	$139 EC2 On-Demand	$52 EC2 Spot	$5.87 EC2 On-Demand	$2.13 EC2 Spot	$13.66 EC2 On-Demand	$4.10 EC2 Spot
NVIDIA Clara Parabricks software cost ($0.30/hour)	$9		$0.45		$0.13	
Combined cost	**$148**	**$61**	**$6.32**	**$2.58**	**$13.79**	**$4.23**

typical costs to analyze a 100GB human genome at 30x coverage, using NVIDIA Clara Parabricks germline pipeline.

3

GPU-Accelerated bioinformatics tools

Alignment	Preprocessing	Variant calling			Variant processing	Quality checking
BWA-MEM	MergeBam	HaplotypeCaller	Muse	Smoove	VQSR	Alignment Summary
STAR	Co-ordinate Sorting	DeepVariant	Strelka	Manta	Variant Filtration	InsertSize
						SequencingArtifact
	Picard MarkDups	STAR-Fusion	Mutect2	CNVKit	Select Variants	GcBias
						QualityScoreDistribution
	BQSR	GLnexus	Somatic Sniper		CNNScore Variants	WGSMetrics
						RawWGSMetrics
	mpileup	BCF Tools Call	LoFreq		VCF Merging	BaseDistributionByCycle
						MeanQualityByCycle

NOT ALL TOOLS SHOWN. A selection of the tools available in the NVIDIA Clara Parabricks software suite

- ■ Structural variant caller
- ■ Somatic variant caller (Strelka included)
- ■ Germline variant caller
- ■ Alignment and QC Tools

GET STARTED TODAY

NVIDIA Clara Parabricks is available on AWS Marketplace as an Amazon Machine Image (AMI), enabling organizations to get started in minutes. Organizations can use genomics data they already store in AWS, or one of the open genomics datasets available on AWS.

1 Go to AWS Marketplace

2 Search for 'NVIDIA Clara Parabricks Pipelines'

3 Subscribe

4 Launch and configure an Amazon Elastic Compute Cloud (Amazon EC2) instance

5 Download sample data, or your organization's data stored on AWS

6 Run the first NVIDIA Clara Parabricks analysis

aws | NVIDIA

Learn more about AWS for Genomics ›