

Solution Brief

Accelerate Genome Sequencing with ONTAP AI and Parabricks

Fast, accurate GPU-accelerated genomic analysis

Key Benefits

Accelerate Genome Sequencing

Speed performance of Parabricks' GPU accelerated genomic sequencing an average of 50x compared to CPU-only solutions.

Maximize Throughput, Minimize Turnaround Time

Perform more genome analysis operations in less time with industry-leading cloud connected all-flash storage from NetApp and NVIDIA DGX servers.

Improve Accuracy and Security

Protect sensitive genomic data while improving test accuracy for AI-powered precision calculations.

Lower Total Cost of Ownership

Lower total cost of ownership with industry-leading data-efficiency technologies and a 25x capacity advantage versus competitive systems.

Simplify Design, Accelerate Return on Investment

Accelerate return on investment with simplified integration, automation, and orchestration of data in clouds and on premises.

The Challenge

Fifteen years ago, the first human genome was sequenced. It took 20 years and cost almost \$3 billion. Today it's possible to sequence 3 billion base pairs in under a day for less than \$1,000, making it feasible to use whole genome sequencing for research and clinical applications.

As the demand for genome sequencing grows, researchers are generating a tidal wave of data that is doubling in size every year. The data generated by whole genome sequencing requires massive amounts of compute power, storage, and data management that can easily become a bottleneck. A single sample can take days to process. With thousands of subjects or patients, it could take years to get through every record. To keep up with the increased demand for sequencing, organizations must be able to complete more sequencing jobs in less time without sacrificing accuracy or security.

The Solution

Parabricks' high-performance GPU-accelerated genomic analysis software speeds processing of genomic pipelines from days to minutes compared to standard CPU-based approaches. Parabricks software in combination with NetApp® ONTAP® AI enables you to deploy a fully integrated, end-to-end AI solution tuned for compute- and data-intensive genomics workloads.

With NetApp ONTAP AI proven architecture, powered by NVIDIA DGX supercomputers and NetApp cloud-connected, all-flash storage, you can fully realize the promise of AI and deep learning (DL) by simplifying, accelerating, and integrating your data pipeline. Streamline the flow of data and speed up training and inference with a NetApp Data Fabric that spans from edge to core to cloud.

Accelerate Sequencing

Parabricks running on ONTAP AI offers dramatic performance improvements over CPU-only solutions. With a single GPU, Parabricks on ONTAP AI processes sequencing pipelines an average of 10x faster than CPU-based solutions. As you add GPUs, the speedup increases incrementally. With 8 GPUs, you can accelerate sequencing by an average of 50x faster than CPU-based solutions.

ONTAP AI is built on NetApp AFF storage systems, which keep data flowing to DL processes with the industry's fastest all-flash storage, validated by SPC-1 benchmark testing. With NVMe performance from the back-end storage to front-end connectivity, the AFF A800 is capable of feeding data to NVIDIA DGX-1 systems 4x to 6x faster than competing solutions.

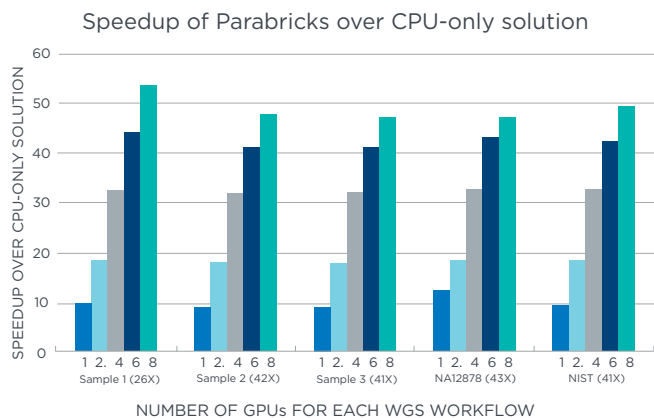


Figure 1) Speedup of Parabricks over CPU-only solution. Testing conditions: ONTAP AI with an AFF A800 storage controller, 8X NVIDIA Tesla V100-SXM2-32GB GPUs, Ubuntu 16.04, and 526GB memory.

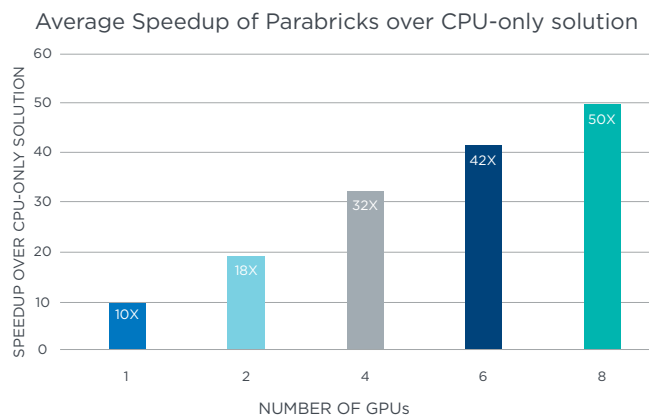


Figure 2) Average speedup of Parabricks over CPU-only solution. Testing conditions: ONTAP AI with an AFF A800 storage controller, 8X NVIDIA Tesla V100-SXM2-32GB GPUs, Ubuntu 16.04, and 526GB memory.

As shown in Figure 2, the average GPU-enabled performance improvement for each number of GPUs ranges from an average speed of 10x with only 1 GPU to 50x with 8 GPUs.

Minimize Turnaround Time, Maximize Throughput

With Parabricks on ONTAP AI, you can optimize throughput to complete more genomics jobs in less time. This integrated solution gives you four ways to run the GATK pipeline on an 8-GPU node, so you can flexibly scale to accommodate more genomic tasks and meet increased customer demand.

Improve Accuracy and Security

Parabricks was designed to dramatically improve performance without sacrificing accuracy. Protect valuable research and patient data with integrated data protection and compliance. NetApp enables you to integrate, protect, and secure your data from edge to core to cloud. With data encryption at rest and in flight, your data remains protected no matter where it resides or where it moves to.

Lower Total Cost of Ownership

DL training routines demand massive amounts of compute power. Faster image training can cut overall compute costs while accelerating AI innovation and productivity. Just one DGX-1 server provides over 1 PFLOPS of AI computing power, the equivalent of an entire data center of traditional CPU-based servers.

With NetApp AFF storage, you can achieve an unparalleled 22:1 overall data-reduction ratio and up to 54% lower TCO compared to direct-attached storage. Simplify data management with a single set of tools across your hybrid cloud.

Smart features, such as intelligent data movers, auto data tiering, autoprovisioning, and predictive analytics, enable you to optimize your mix of cloud and on-premises resources.

Simplify Deployment, Accelerate Return on Investment

The rapid pace of AI innovation makes designing an effective AI infrastructure challenging. But with Parabricks on ONTAP AI, you can eliminate guesswork and get started faster with a validated reference architecture that detangles design complexity. Simplified integration, automation, and orchestration of data in clouds and on premises enables you to accelerate your return on investment.

About Parabricks

Parabricks provides high performance GPU-based software solutions for analyses of next-generation sequencing data, resulting in high throughput at reduced costs compared to other available solutions. Processing the GATK4 pipeline is reduced from days to less than one hour, dropping the cost by up to 75%—all while generating fully equivalent and accurate output.

[Learn more.](#)

About NetApp

NetApp is the data authority for hybrid cloud. We provide a full range of hybrid cloud data services that simplify management of applications and data across cloud and on-premises environments to accelerate digital transformation. Together with our partners, we empower global organizations to unleash the full potential of their data to expand customer touchpoints, foster greater innovation, and optimize their operations. For more information, visit www.netapp.com. #DataDriven