

# Accelerated AI Genomic Discoveries with Parabricks and WekaIO Powered by HPE and NVIDIA

## The Challenges

The Human Genome Project (HGP) is one of the greatest scientific achievements of our time. Completed in 2003, the sequencing of the first human genome took 13 years to accomplish and delivered for the first time in history, the ability to read nature’s complete genetic blueprint for building a human being. Today it’s possible to sequence 3 billion base pairs in under a day, making it feasible to use whole genome sequencing for research and clinical applications.

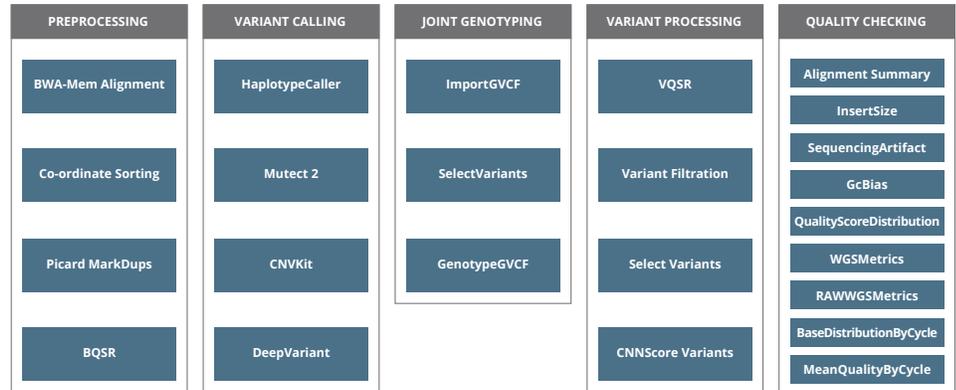
Legacy file systems were architected to support hard disk drives that provide good large-file and sequential access performance, whereas genomic analysis requires both large file sequential access and small-file and random-access performance. Similarly, traditional backup and archive systems were designed for multiple tiers of disk and tape-based systems. This combination adds cost, complexity and considerable management overhead to ensure that the data is accessible and readable when needed.

## The Solution

**A Powerful Integrated Solution from Parabricks, WekaIO, HPE (with NVIDIA GPUs), and Mellanox**

Meet the cost and storage challenges of exponential genomics data with an accelerated integrated solution, built on HPE’s Apollo 6500™ with NVIDIA V100™ GPUs, HPE DL360s™, WekaFS™, Mellanox™ networking and NVIDIA Parabricks™. This AI solution can deliver outcomes more than 30-50 times faster than CPU-only solutions. It can be deployed in minutes/hours rather than weeks/months to enable faster time to discovery, focusing on research rather than software and hardware integration.

### Alignment—BWA-Mem Alignment



### Fast Sequencing at Low Cost

Parabricks is a software suite for performing secondary analysis of next generation sequencing (NGS) DNA data. It is designed to deliver results at significantly fast speeds and low cost. Parabricks can analyze whole human genomes in about 45 minutes, compared to about 30 hours for 30x WGS data. The output results match the commonly used software exactly, making it fairly simple to verify the accuracy of the output.



## Protect Valuable Data

Performance starts with the file system. WekaFS is a distributed, scale-out, POSIX-compliant file system that runs on an standard x86 server platforms and uses off-the-shelf SSDs, greatly improving storage system performance. With active data stored on flash-based storage and catalog data stored on S3 object storage, all within the same namespace, access is near-instantaneous. Valuable data is protected using patented data protection and distribution algorithms that allow the system to sustain up to four simultaneous node or SSD failures.

## World-Class Performance

The HPE Apollo 6500 Gen10 System is powered with industry-leading GPUs, fast GPU interconnect and high bandwidth fabric. The system features flexible configurations to match customer workloads, including up to eight NVIDIA Tesla V100 GPUs per node, NVLink® technology for fast GPU-to-GPU communication, Intel® Xeon® Scalable processor support and up to four high-speed/low-latency fabric adapters.

## Superior Networking

The Mellanox SN2700 switch is an ideal spine and top-of-rack (ToR) solution, allowing maximum flexibility, with port speeds spanning from 10Gb/s to 100Gb/s per port and port density that enables full rack connectivity to any server at any speed.

**For more info please contact:**

855.683.4526

info@groupwaretech.com

groupwaretech.ai



Figure 1

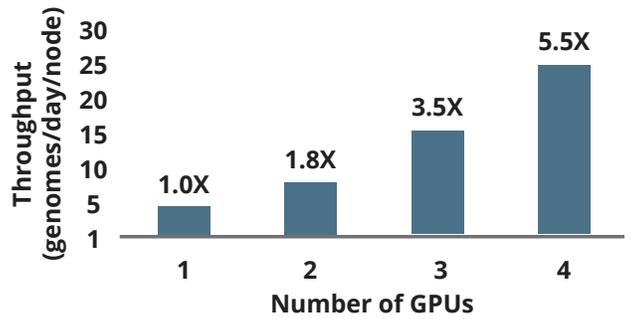


Figure 1 shows the scaling of a single NVIDIA Parabricks genome workflow on increasing numbers of GPUs in a single HPE Apollo 6500 node with WekaIO storage. Performance scales well from one to eight GPUs, similar to using local NVMe SSDs. These results used four FE processes per client node and a single workflow load of up to 0.4 Gbps and 13K OPS. Reducing the number of Weka client FrontEnd (FE) processes from four to two did not significantly affect performance.

Figure 2

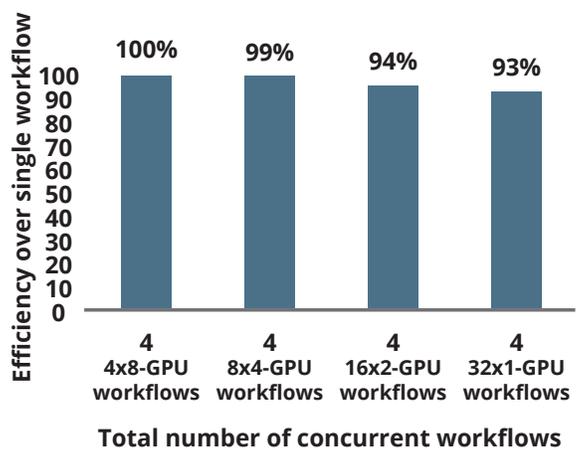
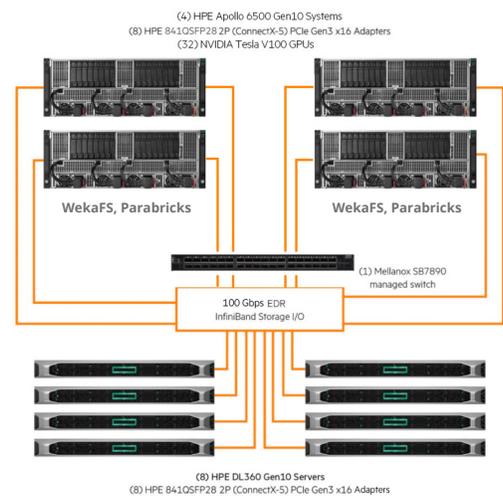


FIGURE 2 shows throughput efficiency of WekaIO running multiple NVIDIA Parabricks genome workflows concurrently on four Apollo 6500 Gen10 Systems



Images and captions courtesy of HPE.

### CONTACT

Groupware Technology, Inc.  
 Email: info@groupwaretech.com  
 Phone: 855.683.4526  
 Web: groupwaretech.ai  
 HQ | Arizona

### HEADQUARTERS

541 Division Street  
 Campbell, CA 95008  
 Phone: 408.540.0090

