Hewlett Packard Enterprise

HPE APOLLO 6500 SYSTEM WITH WEKAIO FILE SYSTEM ADVANCES NEXT-GENERATION SEQUENCING

Key takeaways

- **High scaling and efficiency with the WekaIO File System for genomic analyses**

- **Optimized GPU software for next-generation sequencing**

- **Flexible configuration and use**

Configuration

All tests were run on four HPE Apollo 6500 Gen10 Systems, each comprising an HPE ProLiant XL270d Gen10 Server with Intel® Xeon® Gold 6248 @ 2.50 GHz processors, 2 sockets/40 cores/80 threads, 8 x NVIDIA Tesla V100-SXM2-16GB GPUs, 768 GB memory, and access to 20 TB of WekaFS. The WekaIO storage is a cluster of eight HPE ProLiant DL360 Gen10 Servers running WekaIO v3.5.4, containing a total of 32 670 GB NVMe SSDs, using the Weka client with four FrontEnd (FE) processes per client node. The four HPE Apollo 6500 Systems are connected to this cluster via Mellanox 100 Gbps EDR InfiniBand, as shown in the benchmarking configuration diagram (Figure 3).

All tests used NVIDIA Parabricks Germline Pipeline v2.4.6 running a whole genome sequencing (WGS) workflow with a human genome sample of 43X coverage level. This test case has an input file size of 76 GB, a total of 417 GB of metadata, and an output file size of 1 GB.

Enabling high-performance genomic analyses with HPE, NVIDIA Parabricks, and WekaIO

EXECUTIVE SUMMARY

The HPE Apollo 6500 Gen10 System with a WekaIO (Weka) flash-optimized parallel file system on HPE ProLiant DL360 Gen10 Servers enables the high-throughput execution of secondary genomic analyses using the NVIDIA® Parabricks Germline Pipeline. The NVIDIA Parabricks Germline Pipeline accelerates the Broad Institute's original GATK Best Practices pipeline using NVIDIA GPUs, while the WekaIO File System (WekaFS™) delivers the file I/O performance necessary to keep each GPU fully loaded with genomic data and metadata. This performance scales with high efficiency in order to support large multinode GPU clusters, thereby reducing time to solution and increasing productivity.

CUSTOMER VALUE WITH HPE

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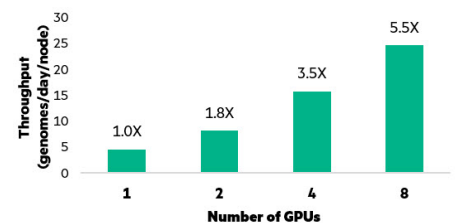

FIGURE 1. Scaling performance of single NVIDIA Parabricks genome workflow on single 8-GPU node with WekaIO

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.[1]
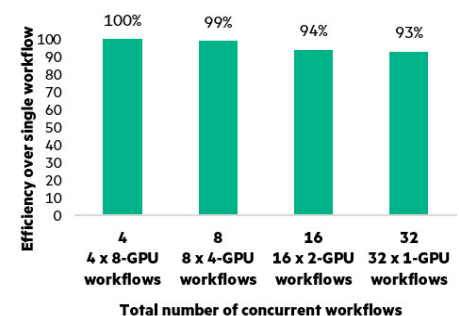

FIGURE 2. Throughput efficiency of WekaIO running multiple NVIDIA Parabricks genome workflows concurrently on four Apollo 6500 Gen10 Systems

---

Figure 2 shows the high efficiency of running multiple NVIDIA Parabricks workflows on four HPE Apollo 6500 Systems, for increasing numbers of workflows running concurrently. The efficiency is defined as the time to process a single workflow divided by the time to process multiple concurrent workflows, which ideally is close to 100%, meaning that the time to solution for running multiple concurrent instances of a job is the same as running a single standalone instance of the same job. Figure 2 demonstrates that WekaFS efficiently services the I/O load of running not only more workflows (up to 3.5 Gbps and 39K OPS for multiple workloads), but also more concurrent workflows per node.

Figure 3 shows the configuration diagram of the benchmarking system used for Figures 1 and 2.

WekaIO storage

The WekaIO File System storage includes the WekaFS flash-optimized parallel file system, which is qualified on HPE Apollo 2000 Gen10 Systems and HPE ProLiant DL360 Gen10 Servers, and utilizes advanced Mellanox interconnect features. The WekaFS software transforms NVMe-based flash storage, HPE qualified servers, and interconnect fabrics into a high-performance, scale-out parallel storage system that is designed to deliver reliable, low latency access to both large and small files in data-intensive workloads.

(4) HPE Apollo 6500 Gen10 Systems
(8) HPE 841QSFP28 2P (ConnectX-5) PCIe Gen3 x16 Adapters
(32) NVIDIA Tesla V100 GPUs

(1) Mellanox SB7890 managed switch

100 Gbps EDR InfiniBand Storage I/O

(8) HPE DL360 Gen10 Servers
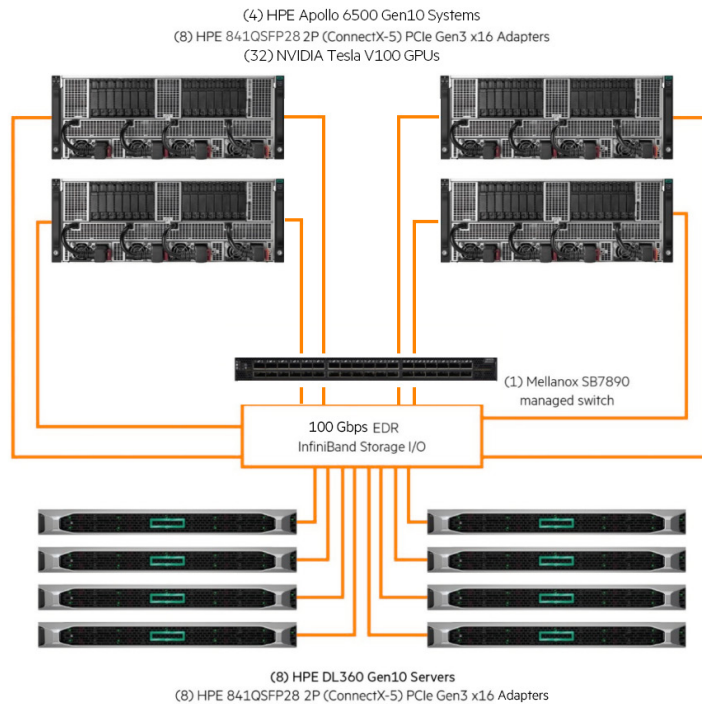(8) HPE 841QSFP28 2P (ConnectX-5) PCIe Gen3 x16 Adapters

FIGURE 3. Benchmarking configuration with four HPE Apollo 6500 Gen10 Systems connected via Mellanox 100 Gbps EDR InfiniBand to eight HPE DL360 Gen10 Servers running WekaIO File System.

[2] HPE white paper, Accelerated throughput of next-generation sequencing workflows, October 2018.

NVIDIA Parabricks software

The NVIDIA Parabricks Germline Pipeline software accelerates the original Broad Institute's GATK Best Practices pipeline using NVIDIA GPUs, with results that are on average 36X faster than an equivalent CPU-only implementation and greater than 99.9% accurate in precision and sensitivity.[2]

Bottom line

Performance benchmarks illustrate the transformative power of HPE servers, through HPE partnerships that aim to increase productivity and fast-forward customer success.

**Make the right purchase decision.
Contact our presales specialists.**

Chat   Email   Call

Get updates

Hewlett Packard Enterprise