DELL Technologies

Whitepaper

Dell PowerEdge Servers Accelerate Genomics Data Processing with Sentieon Software on 5th Generation AMD EPYC Processors

AMD Sentieon

Table of Contents

Executive Summary

Healthcare and life sciences are advancing at a remarkable pace, driven by breakthroughs in genomic sequencing and the demand for precision medicine. At the center of this transformation are Dell PowerEdge servers—purpose-built platforms engineered to scale performance, improve efficiency, and accelerate discovery.

Deployed with Sentieon® software and 5th Generation AMD EPYC™ processors, PowerEdge servers deliver unmatched results for genomic data analysis. This optimized infrastructure provides the clinical acceleration needed for major precision medicine breakthroughs, scaling multi-sample workflows seamlessly, and outperforming older CPU and GPU-based solutions.

Beyond speed, the PowerEdge advantage extends to efficiency and economics. In preliminary TCO analysis, this solution demonstrates up to a 5x cost reduction compared to an NVIDIA H100 GPU-based system. PowerEdge servers also support diverse sequencing inputs—from Illumina, PacBio, and Oxford Nanopore—while enabling hybrid pipelines that improve accuracy in detecting complex variants.

By cutting analysis times and scaling to meet growing genomic workloads, Dell PowerEdge empowers healthcare and research organizations to accelerate discoveries, streamline clinical decision-making, and deliver personalized treatments. Combining Dell's structural advantage with Sentieon's optimized pipelines and AMD's high-core-count CPUs, this is the foundation for the next era of genomics innovation.

Introduction

The field of next-generation sequencing (NGS) continues its rapid evolution, driven by both technological innovation and increased adoption in clinical and research settings. 2025 has seen the introduction of a new high-throughput platform by Roche Diagnostics,[1] joining established and emerging players like Illumina, MGI/Complete Genomics, Element Biosciences, and Ultima Genomics-as-a-Platform for NGS data generation. This market dynamism, coupled with continued improvements in long-read technologies from PacBio and Oxford Nanopore Technologies (ONT), has made high-quality genomic data more accessible than ever. The cost to sequence a human whole genome to 30x coverage (the average depth at which each base in the genome is sequenced, meaning it is read approximately 30 times for improved accuracy and reliability), a standard for many clinical applications, has fallen well below the $1000 threshold, increasing the demand for robust and efficient downstream analysis.[2]

This growth in data generation underscores the critical importance of scalable computational pipelines. While specialized hardware accelerators offer one path to performance, highly optimized software running on general purpose, high-core-count CPUs presents a powerful and flexible alternative. Sentieon® software is designed to maximize the performance of modern x86 architectures, effectively utilizing multi-threading to scale across a high number of cores. This approach addresses the performance limitations observed in some open-source tools that are not architected to leverage the full parallel processing capabilities of modern servers, thereby challenging the notion that hardware acceleration is a prerequisite for achieving rapid turnaround times in genomics.

In a previous analysis from 2024,[3] we demonstrated the processing of a 30x whole genome from FASTQ to VCF in under 15 minutes at a cloud cost of less than $1.50, using Sentieon software on systems with 4th Generation AMD EPYC processors. Sentieon has continued to innovate and, in 2025, released a new pipeline for analysis of hybrid short and long-read data.[4] The combined pipeline improves variant detection accuracy, especially for structural and complex variants that require multiple technologies. New short-read pipelines designed to leverage advances in the human pangenome are also demonstrating improved accuracy across a range of variation including structural variants, trinucleotide repeats, and large copy number variants.[5] To evaluate the impact of these advancements, we conducted a series of benchmarks using Dell PowerEdge servers and Sentieon software.

This white paper, building upon our 2024 white paper, presents an analysis of Sentieon's performance in processing various sequence datasets on two Dell PowerEdge servers with 5th Generation AMD EPYC processors. The tests presented here provide detailed benchmark results across key genomic analysis stages: alignment, preprocessing, and germline variant calling and provide insights in the performance and efficiency of the data processing pipeline. Both servers process a 30x whole genome in less than 10 minutes, faster than the 14 minutes required on with NVIDIA H100 GPUs or the 12.3 minutes reported in our earlier benchmark with 4th Generation AMD EPYC processors.[7] In a new multi-sample benchmark, we achieve amortized per-sample runtimes of less than five minutes on both servers.[8]

[1] https://www.roche.com/investors/updates/inv-update-2025-02-20

[2] Whole genome sequencing in clinical practice, https://link.springer.com/article/10.1186/s12920-024-01795-w

[3] https://www.amd.com/content/dam/amd/en/documents/epyc-technical-docs/white-papers/amd-epyc-9004-wp-hls-sentieon.pdf

[4] https://www.biorxiv.org/content/10.1101/2025.04.15.648987v3

[5] Genotyping structural variants in pangenome graphs using the vg toolkit, https://link.springer.com/article/10.1186/s13059-020-1941-7

[6] Benchmark data https://sentieon.sharepoint.com/:x:/g/EXdm5xZEtytCkht7OMenbFoBJ_s8zzumbvd8-FI5PTbTQA?rtime=sCfmMesF3kg Summary Table 1.

[7] https://www.amd.com/content/dam/amd/en/documents/epyc-technical-docs/white-papers/amd-epyc-9004-wp-hls-sentieon.pdf

[8] Benchmark data https://sentieon.sharepoint.com/:x:/g/EXdm5xZEtytCkht7OMenbFoBJ_s8zzumbvd8-FI5PTbTQA?rtime=sCfmMesF3kg Summary Table 4.

Test setup

This section describes the server configurations used for the tests showcased in this white paper.

Benchmark datasets

The benchmarks described in this white paper use the GIAB (Genome in a Bottle) samples sequenced on the Illumina NovaSeq, PacBio Sequel II (HiFi), and Oxford Nanopore instruments in the Fastq file format:

- Illumina NovaSeq (HG002): These Fastq files were obtained from sequencing the Google Genome in a Bottle* samples and then downsampled to 30x coverage (93 billion bases).
- PacBio Sequel II (HG002): These Fastq files the PacBio Sequel II instrument were obtained from the Genome in a Bottle data repository* and then downsampled to 30x coverage (93 billion bases).
- Oxford Nanopore (HG002): These tests used Fastq files 1-11 from the Human Pangenome Reference Consortium data repository*. These files were selected to obtain a coverage of 29.6x (91.8 billion bases).
- Illumina NovaSeq (multisample): These Fastq files were obtained from the Google Genomics sequencing of Genome-in-a-Bottle samples, without downsampling.

Hardware configuration

This section describes the on-premises and cloud configurations presented in this white paper:

- On-premises: Dell PowerEdge R7725 (dual socket AMD EPYC 5th Generation processors) in the following configurations:
  - Dual AMD EPYC 9755 processor: 128 core, 2.7GHz (256 total cores)
  - Dual AMD EPYC 9965 processor: 192 core, 2.2GHz (384 total cores)
  - Each PowerEdge server was built with dual 1.92TB NVMe drives for data, mirrored NVMe boot drives, and 1,536GB of RAM
- Amazon EC2: The Amazon EC2 Hpc7a instance family provides compute infrastructure suitable for high-performance computing applications at a cost-effective price point. These instances include gp3 EBS storage with 5000 IOPS and 500 MiB/s. These tests also used a ramdisk in the instance memory to hold intermediate files. Sentieon anticipates that the Amazon EC3 C7a instance family would achieve similar results. Sentieon performed the cloud testing described in this white paper.
- Oracle cloud: The Oracle Cloud BM.GPU.H100.8 shape is powered by dual 56-core 4th Generation Intel® Xeon® processors, 8x NVIDIA H100 80GB Tensor Core GPUs, 2 TB of DDR5 memory, 16 x 3.84 TB NVMe local storage, and 8 x 400 Gb/sec cluster networking.

Software configuration

This section describes the on-premises and cloud configurations presented in this white paper:

- Sentieon software in the latest benchmark:
  - Sentieon: v202503.01
  - CLI: The sentieon-cli v1.2.3 was used for germline variant calling of the long-read datasets.
- DNAscope model bundles used during data processing:
  - Illumina WGS: v2.2
  - PacBio HiFi: v2.1
  - ONT: v2.1

Germline alignment and variant calling for short reads

Sentieon software processed the Illumina NovaSeq whole genome 30x sample in 9.5 minutes with the Dell PowerEdge R7725 with AMD EPYC 9755 processor and in 10.0 minutes with the Dell PowerEdge R7725 with AMD EPYC 9965 processor. These runtimes are about 28% faster than the previously published NVIDIA Parabricks pipeline runtimes[9] running with eight NVIDIA H100 GPUs, or about 18% faster than our previous benchmark using the older generation of CPUs.[10]

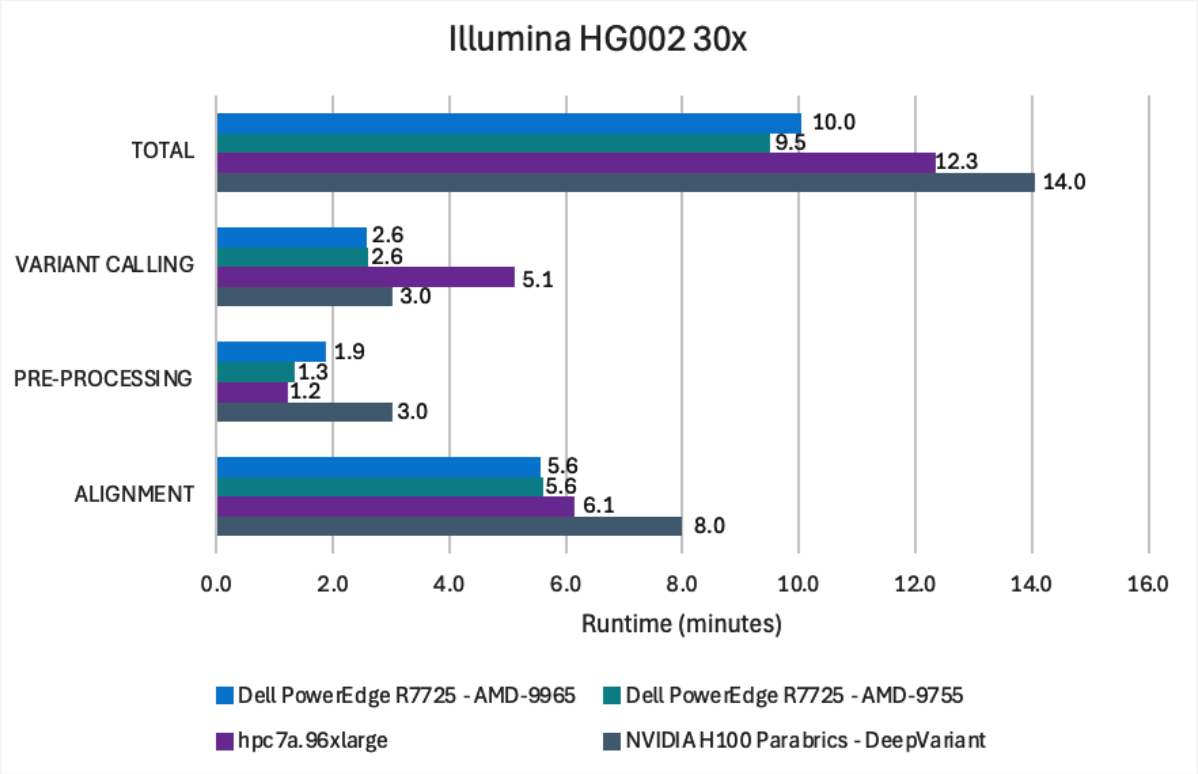

Figure 1: Sentieon DNAscope and NVIDIA runtime performance[11]

Figure 1 shows the on-premises and cloud runtime performance of the Sentieon DNAscope and NVIDIA Parabricks pipelines. NVIDIA reports the NVIDIA Parabricks numbers using a configuration with 8x NVIDIA H100 GPUs. The faster runtime in the latest benchmark is due to both hardware and software improvements.

[9] https://developer.nvidia.com/blog/accelerate-genomic-analysis-for-any-sequencer-with-parabricks-v4-2/

[10] https://www.amd.com/content/dam/amd/en/documents/epyc-technical-docs/white-papers/amd-epyc-9004-wp-hls-sentieon.pdf

[11] Benchmark data https://sentieon.sharepoint.com/:x:/g/EXdm5xZEtytCkht7OMenbFoBJ_s8zzumbvd8-FI5PTbTQA?rtime=sCfmMesF3kg Summary Table 1.

Germline alignment and variant calling for long reads

Although most NGS datasets are generated using short-read sequencers, high quality long-read data are a compelling alternative for whole-genome sequencing or other applications that can benefit from longer read lengths. Sentieon processed high-quality PacBio HiFi and Oxford Nanopore Technologies (ONT) duplex sequencing datasets at approximately 30x coverage using the Sentieon DNAscope LongRead pipeline in these tests and in the previous benchmark with an Amazon EC2 hpc7a.96xlarge instance.

The results presented here show a substantial performance improvement relative to the earlier benchmarks on the Amazon EC2 hpc7a.96xlarge instance. Processing of the PacBio HiFi sample completed in less than 10.2 minutes compared to 16.9 minutes on the hpc7a.96xlarge instance. Processing of the ONT sample completed in 17.4 minutes compared to 39.7 minutes on the hpc7a.96xlarge instance. The runtime improvements are the result of both hardware and software improvements in the new benchmark.

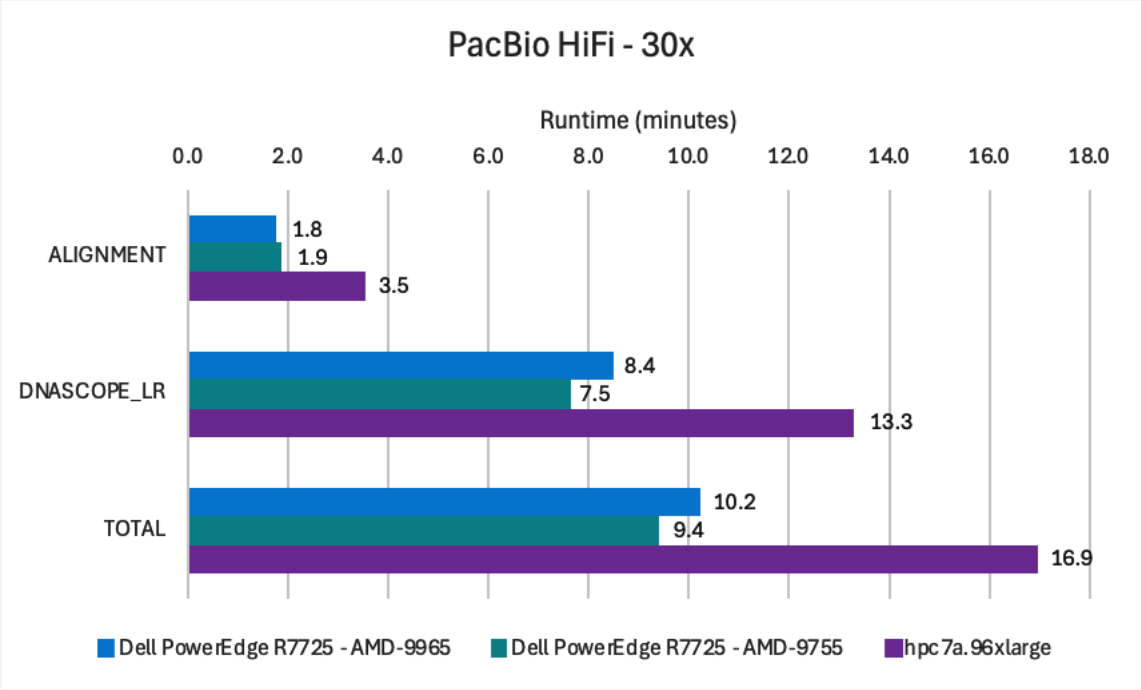

Figure 2: Sentieon DNAscope LongRead performance with PacBio HiFi data[12]

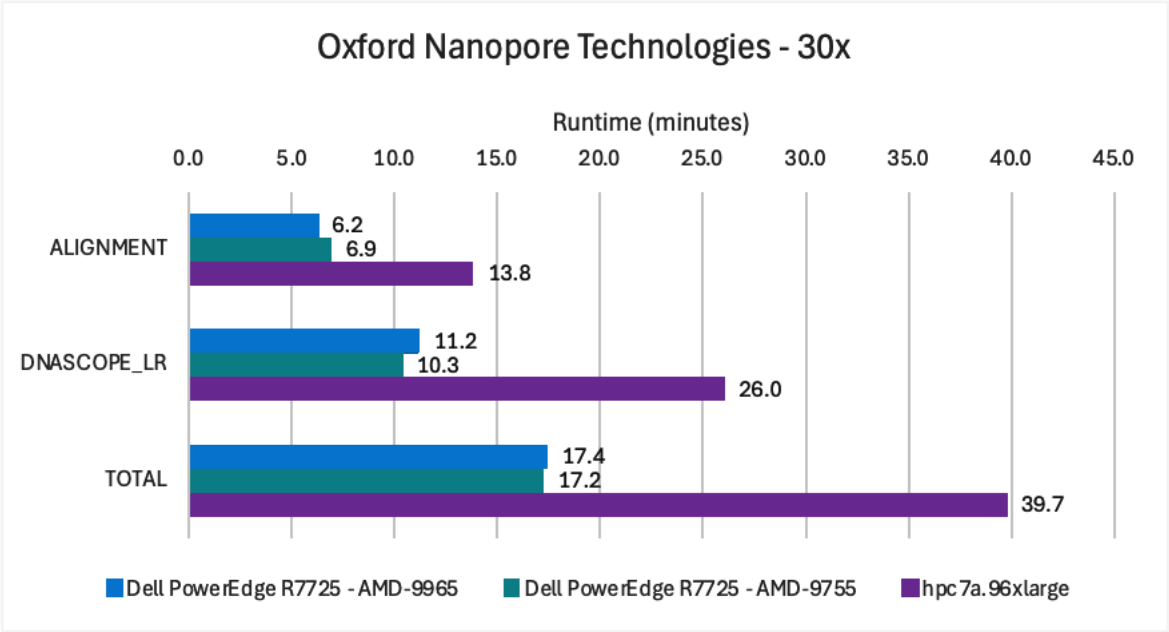

Figure 3: Sentieon DNAscope LongRead performance with ONT data[13]

[12]Benchmark data https://sentieon.sharepoint.com/:x:/g/EXdm5xZEtytCkht7OMenbFoBJ_s8zzumbvd8-FI5PTbTQA?rtime=sCfmMesF3kg Summary Table 2.

[13]Benchmark data https://sentieon.sharepoint.com/:x:/g/EXdm5xZEtytCkht7OMenbFoBJ_s8zzumbvd8-FI5PTbTQA?rtime=sCfmMesF3kg Summary Table 3.

Germline short read alignment and variant calling with multiple samples

Single-sample benchmarks provide compelling processing speeds for individual samples with fast turnaround times on powerful servers like the Dell PowerEdge R7725 with AMD EPYC 5th Generation processors. To evaluate the performance of our pipeline with a larger workload, Sentieon concurrently processed eight NovaSeq 30x samples, roughly corresponding to one lane of the largest NovaSeq X 25B flowcell, on a single server.

Sentieon processed eight 30x samples with an amortized per-sample runtime of less than 4.29 minutes, with all eight samples finishing in under 40 minutes on both servers. These results demonstrate that processing multiple samples in parallel can provide substantial improvements in throughput relative to serial processing of single samples.

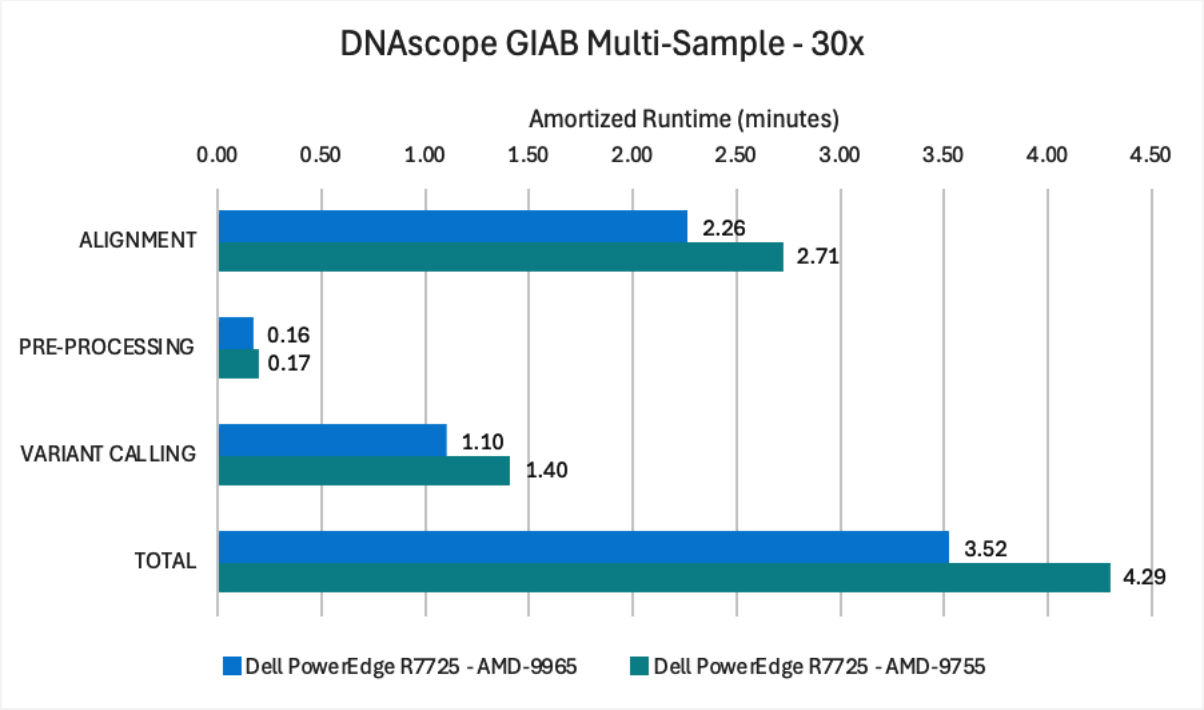

Figure 4: Sentieon DNAscope performance with multiple samples[14]

Preliminary cost analysis

In evaluating high-performance computing options for healthcare applications—such as AI-driven diagnostics or large-scale data processing—we conducted a preliminary cost analysis focused on compute units. This compared Amazon Web Services (AWS) EC2 instances with on-premises Dell PowerEdge servers with AMD processors, and similar servers equipped with 8 NVIDIA H100 GPUs (graphics processing units optimized for AI workloads). Key findings, based on AWS's three-year reserved instance pricing (which offers discounts over on-demand rates):

- An AWS EC2 hp7a.96xlarge instance is approximately 4 times more expensive than a comparable AMD-powered PowerEdge server.
- A PowerEdge server with 8 NVIDIA H100 GPUs is approximately 5 times more expensive than the same server with AMD processors.

These estimates highlight potential cost savings from on-premises AMD-based solutions, though factors like maintenance, scalability, and energy efficiency should also be considered in a full total cost of ownership (TCO) assessment.

[14]Benchmark data https://sentieon.sharepoint.com/:x:/g/EXdm5xZEtytCkht7OMenbFoBJ_s8zzumbvd8-FI5PTbTQA?rtime=sCfmMesF3kg Summary Table 4.

Conclusion

Sentieon software, running on Dell PowerEdge servers with 5th Generation AMD EPYC processors, delivers a groundbreaking combination of speed, flexibility, scalability, and cost-efficiency for genomic data processing. By achieving processing speeds 28% faster (compared to Oracle Cloud BM.GPU.H100.8 instances), and reducing costs by a factor of five, this solution empowers researchers and bioinformaticians to analyze complex genomic data faster and more affordably. Whether tackling single-sample workflows or processing large multi-sample datasets, the Sentieon-Dell-AMD combination offers flexible and reliable performance that meets the demands of high-throughput genomic research.

This powerful infrastructure not only accelerates data insights but also removes the complexities of scalability, giving researchers the tools they need to fuel discovery and innovation in genomics. By reducing obstacles to efficient data processing, it enables breakthroughs in key areas like personalized medicine and advanced variant detection, driving progress in clinical and research applications.

These advancements empower researchers to process genomic data faster and more cost-effectively, enabling breakthroughs in clinical and research settings. Discover how Sentieon software on Dell PowerEdge servers with AMD EPYC processors can accelerate your genomic research. Please email info@sentieon.com* or visit Sentieon today to get started with Sentieon software. For questions about AMD, please contact your AMD sales representative or visit www.amd.com. For questions about Dell Technologies, please contact your Dell sales representative or visit www.dell.com.