

Accelerate Genomics Insights and Discovery with High-Performing, Scalable Architecture from Dell and Intel

Author: Rodrigo Escobar (Intel), Esther Baldwin (Intel), Priyanka Sebastian (Intel) and Todd Mottershead (Dell)

Summary

The field of Genomics requires the storage and processing of vast amounts of data. In this brief, Intel and Dell technologists discuss key considerations to successfully deploy BeeGFS based storage for Genomics applications on the latest generation PowerEdge Server portfolio offerings.

Market positioning

The life sciences industry faces intense pressure to speed results and bring in new treatments to market all while lowering costs, especially in genomics. However, life-changing discoveries often depend on processing, storing, and analyzing enormous volumes of genomic sequencing data — more than 20 TB of new data per day by one organization, alone¹, with each modern genome sequencer producing up to 10TB of new data per day. Researchers need high-performing solutions built to handle this volume of data and analytics and artificial intelligence (AI) workloads that are easy to deploy and scale.

Dell and Intel have collaborated on a bill of materials (BoM) that provides life science organizations with a scalable solution for genomics. This solution features high-performance compute and storage building blocks for one of the leading parallel cluster file systems, BeeGFS. The BoM features four Dell PowerEdge rack server nodes powered by 4th Generation Intel® Xeon® Scalable processors, which deliver the performance needed for faster results and time to production.

The BoM can be tailored for each organization's architectural needs. For dense configurations, customers can use the Dell PowerEdge C6600 enclosure with PowerEdge C6620 server nodes instead of standard PowerEdge R660 servers (each PowerEdge C6600 chassis can hold up to four PowerEdge C6620 server nodes). If they already have a storage solution in place using InfiniBand fabric, the nodes can be equipped with an additional Mellanox ConnectX-6 HDR100 InfiniBand adapter.

Key Considerations

Key considerations for deploying genomics solutions on Dell PowerEdge servers include:

- **Core count:** Life sciences organizations often process a whole genome on a cluster, which scales linearly with core count. The Dell PowerEdge solution offers up to 56 cores per CPU to meet performance requirements.
- **Memory requirements:** This BoM provides 512 GB of DRAM to support specific tasks in workloads that have higher memory requirements, such as running Burrows-Wheeler Aligner algorithms.
- **Local and distributed storage:** Input/output (I/O) is a big consideration for genomics workloads because datasets can reach hundreds of gigabytes in size. Dell and Intel recommend 3.2 TB of local storage specifically for commonly used genomics tools that read and write many temporary files.

Available Configurations

Feature	Configuration
Platform	4 x Dell R660 supporting 8 x 2.5" NVMe drives - direct connection
CPU (per server)	2x Intel® Xeon® Platinum 8480+ (56c @ 2.0GHz)
DRAM	512GB (16 x 32GB DDR5-4800MT/s)
Boot device	Dell BOSS-N1 with 2x 480GB M.2 NVMe SSD (RAID1)
Storage	1x 3.2TB Solidigm D7-P5620 NVMe SSD (PCIe Gen4, Mixed-use)
Capacity storage	Dell Ready Solutions for HPC BeeGFS Storage: 500 GB capacity per 30x coverage whole genome sequence (WGS) to be processed; 800 MB/s total (200 MB/s per node).
NIC	Intel® E810-XXV Dual Port 10/25GbE SFP28, OCP NIC 3.0

Software Versions	
Workload	GATK Best Practices for Germline Variant Calling WholeGenomeGermlineSingleSample_v3.1.6
Applications	<ul style="list-style-type: none"> • WARP 3.1.6 • GATK 4.3.0.0 • Picard 3.0.0 • Samtools 1.17 • Burroughs-Wheeler Aligner (BWA) 0.7.17 • VerifyBamID 2.0.1 • MariaDB 10.3.35 • Cromwell 84

Learn more

Contact your [Dell](#) or Intel account team for a customized quote at 1-877-289-3355.

Read about Intel Select Solutions for Genomics Analysis:

<https://www.intel.com/content/dam/www/public/us/en/documents/solution-briefs/select-genomics-analytics.pdf>

Read about Dell HPC Ready Architecture for Genomics: <https://infohub.delltechnologies.com/static/media/6cb85249-c458-4c06-bcec-ef35c1a363ca.pdf?dgc=SM&cid=1117&lid=spr4502976221&linkId=112053582>

Learn more about Dell Ready Solutions for HPC BeeGFS Storage: <https://www.dell.com/support/kbdoc/en-us/000130963/dell-emc-ready-solutions-for-hpc-beegfs-high-performance-storage>

Learn more about Dell Ready Solutions for HPC BeeGFS High Capacity Storage: www.dell.com/support/kbdoc/en-ie/000132681/dell-emc-ready-solutions-for-hpc-beegfs-high-capacitystorage



For more info,
visit the [Dell
Technologies Info
Hub](#)



[Contact us](#) for
feedback and
requests



Follow us for
more news