

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

In collaboration with:



### Tech Note by

Todd Mottershead

[todd.mottershead@dell.com](mailto:todd.mottershead@dell.com)

Krzysztof Cieplucha

[krzysztof.cieplucha@intel.com](mailto:krzysztof.cieplucha@intel.com)

### Summary

The field of Genomics requires the storage and processing of vast amount of data. In this brief Intel and Dell technologists discuss key considerations to successful deploy BeeGFS based storage for Genomics applications on the most recent 15th Generation PowerEdge Server portfolio offerings.

The life sciences industry faces intense pressure to speed results and bring new treatments to market while lowering costs, especially in the area of genomics. But life-changing discoveries are often dependent on processing, storing and analyzing enormous volumes of genomic sequencing data—more than 20 TB of new data per day by one organization alone.<sup>1</sup> Researchers need high-performing solutions built to handle this volume of data, in addition to demanding analytics and artificial intelligence (AI) workloads, and that are also easy to deploy and scale.

Dell and Intel have collaborated on a bill of materials (BoM) that provides life sciences 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 EMC™ PowerEdge™ rack server nodes powered by 3rd 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 EMC PowerEdge C6400 enclosure with PowerEdge C6520 server nodes instead of standard PowerEdge R650 servers (each PowerEdge C6400 chassis can hold up to four PowerEdge C6520 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 EMC PowerEdge servers include:

- **Core count:** Life sciences organizations often process a whole genome on a cluster, which scales linearly with core count. The Dell EMC PowerEdge solution offers up to 28 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 Burrow-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.

The information in this publication is provided as is. Dell Inc. makes no representations or warranties of any kind with respect to the information in this publication, and specifically disclaims implied warranties of merchantability or fitness for a particular purpose.

Use, copying, and distribution of any software described in this publication requires an applicable software license.

Copyright © 2021 Dell Inc. or its subsidiaries. All Rights Reserved. Dell, EMC, PowerEdge and other trademarks are trademarks of Dell Inc. or its subsidiaries. Other trademarks may be the property of their respective owners.

Dell Inc. believes the information in this document is accurate as of its publication date. The information is subject to change without notice.

## Available Configurations

	Configuration
<b>Platform</b>	4 x Dell EMC™ PowerEdge™ R650 server
<b>CPU (per server)</b>	2 x Intel® Xeon® Gold 6348 processor (28 cores at 2.6 GHz)
<b>DRAM</b>	512 GB (16 x 32 GB DDR4-3200 MHz)
<b>Boot device</b>	Dell EMC™ Boot Optimized Server Storage (BOSS)-S2 with 2 x 480 GB Intel® SSD S4510 M.2 Serial ATA (SATA®) (RAID1)
<b>Local storage</b>	1 x 3.2 TB P5600 NVM Express® (NVMe®) solid-state drive (SSDs)
<b>Capacity storage</b>	Dell EMC™ 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)
<b>Network interface controller (NIC)</b>	Intel® Ethernet Network Adapter E810-XXV for OCP3 (dual-port 25 Gb)

### Learn More

Contact your dedicated Dell or Intel account team for a customized quote. 1-877-289-3355

Read about Intel Select Solutions for Genomics Analysis:

[www.intel.com/content/www/us/en/products/docs/select-solutions/select-solutions-for-genomics-analytics-brief-v2.html](http://www.intel.com/content/www/us/en/products/docs/select-solutions/select-solutions-for-genomics-analytics-brief-v2.html)

Read about Dell EMC™ HPC Ready Architecture for Genomics: [www.delltechnologies.com/asset/en-us/products/ready-solutions/industry-market/dell-ra-genomics.pdf](http://www.delltechnologies.com/asset/en-us/products/ready-solutions/industry-market/dell-ra-genomics.pdf)

Learn more about Dell EMC Ready Solutions for HPC BeeGFS Storage: [www.dell.com/support/kbdoc/en-ie/000130963/dell-emc-ready-solutions-for-hpc-beegfs-high-performance-storage](http://www.dell.com/support/kbdoc/en-ie/000130963/dell-emc-ready-solutions-for-hpc-beegfs-high-performance-storage)

Learn more about Dell EMC Ready Solutions for HPC BeeGFS High Capacity Storage: [www.dell.com/support/kbdoc/en-ie/000132681/dell-emc-ready-solutions-for-hpc-beegfs-high-capacity-storage](http://www.dell.com/support/kbdoc/en-ie/000132681/dell-emc-ready-solutions-for-hpc-beegfs-high-capacity-storage)

<sup>i</sup> Broad Institute. "Sharing Data and Tools to Enable Discovery." [www.broadinstitute.org/sharing-data-and-tools/cloud-computing#top](http://www.broadinstitute.org/sharing-data-and-tools/cloud-computing#top).