

DELL EMC ISILON F800 AND H600 WHOLE GENOME ANALYSIS PERFORMANCE

ABSTRACT

This white paper provides performance data for a BWA-GATK whole genome analysis pipeline run using Dell EMC Isilon F800 and H600 storage. It is intended for performance-minded administrators of large compute clusters that run genomics pipelines. The paper does not discuss the details of running a variant calling pipeline with BWA and GATK.

January 2018

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 © 2018 Dell Inc. or its subsidiaries. All Rights Reserved. Dell, EMC, and other trademarks are trademarks of Dell Inc. or its subsidiaries. Other trademarks may be the property of their respective owners. Published in the USA. 1/18 White Paper

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

TABLE OF CONTENTS

ABSTRACT.....1

EXECUTIVE SUMMARY4

INTRODUCTION.....4

DELL EMC ISILON4

PERFORMANCE EVALUATION.....4

STORAGE CONFIGURATIONS5

COMPUTE NODES.....5

NETWORK CONNECTIVITY7

WHOLE GENOME SEQUENCE VARIANT ANALYSIS8

SUMMARY 12

REFERENCES..... 12

EXECUTIVE SUMMARY

This Dell EMC technical white paper describes whole genome analysis performance results for Dell EMC Isilon F800 and H600 storage clusters (4 Isilon nodes per cluster). The data is intended to inform administrators on the suitability of Isilon storage clusters for high performance genomic analysis.

INTRODUCTION

The goal of this document is to present an F800 versus an H600 performance comparison, and to compare them to other storage offerings like the Dell HPC Lustre Storage Solution¹ for the processing of genomic pipelines. The same test methodologies and the same test hardware were used where possible to generate the results.

DELL EMC ISILON

Dell EMC Isilon is a proven scale-out network attached storage (NAS) solution that can handle the unstructured data prevalent in many different workflows. The Isilon storage architecture automatically aligns application needs with performance, capacity, and economics. As performance and capacity demands increase, both can be scaled simply and non-disruptively, allowing applications and users to continue working.

The Dell EMC Isilon storage system features:

- A high degree of scalability, with grow-as-you-go flexibility
- High efficiency to reduce costs
- Multi-protocol support such as SMB, NFS, HTTP and HDFS to maximize operational flexibility
- Enterprise data protection and resiliency
- Robust security options

A single Isilon storage cluster can host multiple node types to maximize deployment flexibility. Node types range from the Isilon F (All Flash) to H (Hybrid), and A (Archive) nodes. Each provides a different optimization point for capacity, performance, and cost. Automated processes can be established that automatically migrate data from higher-performance, higher-cost nodes to more cost-effective storage. Nodes can be added “on the fly,” with no disruption of user services. Additional nodes result in increased performance (including network interconnect), capacity and resiliency.

The Dell EMC Isilon OneFS operating system powers all Dell EMC Isilon scale-out NAS storage solutions. OneFS also supports additional services for performance, security, and protection:

- **SmartConnect** is a software module that optimizes performance and availability by enabling intelligent client connection load balancing and failover support. Through a single host name, SmartConnect enables client connection load balancing and dynamic NFS failover and fallback of client connections across storage nodes to provide optimal utilization of the cluster resources.
- **SmartPools** provides rule based movement of data through tiers within an Isilon cluster. Institutions can set up rules keeping the higher performing nodes available for immediate access to data for computational needs and NL and HD series used for all other data. It does all this while keeping data within the same namespace, which can be especially useful in a large shared research environment.
- **SmartFail** and **Auto Balance** ensure that data is protected across the entire cluster. There is no data loss in the event of any failure and no rebuild time necessary. This contrasts favorably with other file systems such as Lustre or GPFS as they have significant rebuild times and procedures in the event of failure with no guarantee of 100% data recovery.
- **SmartQuotas** help control and limit data growth. Evolving data acquisition and analysis modalities coupled with significant movement and turnover of users can lead to significant consumption of space. Institutions without a comprehensive data management plan or practice can rely on SmartQuotas to better manage growth.

Through utilization of common network protocols such as CIFS/SMB, NFS, HDFS, and HTTP, Isilon can be accessed from any number of machines by any number of users leveraging existing authentication services.

PERFORMANCE EVALUATION

The motivation for this performance analysis was to investigate the ability of the F800 and H600 to support human whole genome variant calling analysis. The tested pipeline used the Burrows-Wheeler Aligner (BWA) for the alignment step and Genome Analysis Tool

Kit (GATK) for the variant calling step. These are considered standard tools for aligning and variant calling in whole genome or exome sequencing data analysis.

STORAGE CONFIGURATIONS

Table 1 lists the configuration details of the three storage systems benchmarked. Default OneFS settings, SmartConnect and NFSv3 were used in all the Isilon tests. A development release of OneFS was used on the F800. Upgrading to the same OneFS version as used on the H600 would likely yield slightly better results. The F800 also uses 40GbE as a backend network, compared to the H600 which uses QDR Infiniband. OneFS >= v8.1 has been optimized for use with GbE and performs slightly better than QDR Infiniband. Details on the Dell HPC Lustre Storage Solution used in these tests can be found [here](#)².

	F800	H600	Lustre
# of nodes	4	4	2 (OSS)
CPU per node	Intel(R) Xeon ® CPU E5-2697A v4 @2.60GHz	Intel(R) Xeon ® CPU E5-2680 v4 @2.40GHz	Dual Intel Xeon ® E5-2630 V4 @ 2.20GHz
Memory per node	256 GB	256 GB	256GB
Storage Capacity	Total usable space: 166.8 TB 41.7 TB per node	Total usable space: 126.8 TB 31.7 TB per node	Total raw space: 480 TB (240 TB per OSS)
SSD L3 cache	NA	2.9 TB per node	NA
Network	Front end network: 40GbE Back end network: 40GbE	Front end network: 40GbE Back end network: IB QDR	Intel OmniPath
OS	Isilon OneFS v8.1.0.DEV.0	Isilon OneFS v8.1.0.0 B_8_1_0_011	Red Hat Enterprise Linux Server release 7.2 (3.10.0-327.el7.x86_64)

Table 1. Storage Configurations

COMPUTE NODES

64 nodes of the Zenith compute cluster³ were used during the tests. Table 2 lists Zenith compute node configuration details. The compute nodes were upgraded from RHEL 7.2 to RHEL 7.3 for the H600 tests.

DELL HPC INNOVATION LABS ZENITH COMPUTE CLUSTER

Compute Clients	64 x PowerEdge C6320s
Processor	CPU: Intel(R) Xeon(R) CPU E5-2697 v4 @ 2.30GHz No. of cores = 18 per processor (36 per node) Processor Base Frequency: 2.3GHz AVX Base: 2.0GHz
Memory	128 GB @ 2400 MHz per node
Operating System	Red Hat Enterprise Linux Server release 7.2 (7.3 for H600 tests)
Kernel	3.10.0-327.13.1.el7.x86_64 (3.10.0-514.el7.x86_64 for H600 tests)
System BIOS Profile	Max Performance <ul style="list-style-type: none">• Turbo mode: Enabled• C-states: disabled• Node interleave: disabled• Logical processor: disabled• Snoop mode: opportunistic snoop broadcast• I/O-Nonposted Prefetch: Disabled
Network	1GbE, 10GbE, and Intel OmniPath

Table 2. Zenith Compute Cluster Node Configuration Details

NETWORK CONNECTIVITY

The Zenith cluster and F800 storage system were connected via 8 x 40GbE links. Figure 1 shows the network topology used in the tests. The H600 was configured in the exact same way as the F800. Figure 2 shows the network configuration of the Dell HPC Lustre Solution. An OmniPath network was used for the Lustre tests.

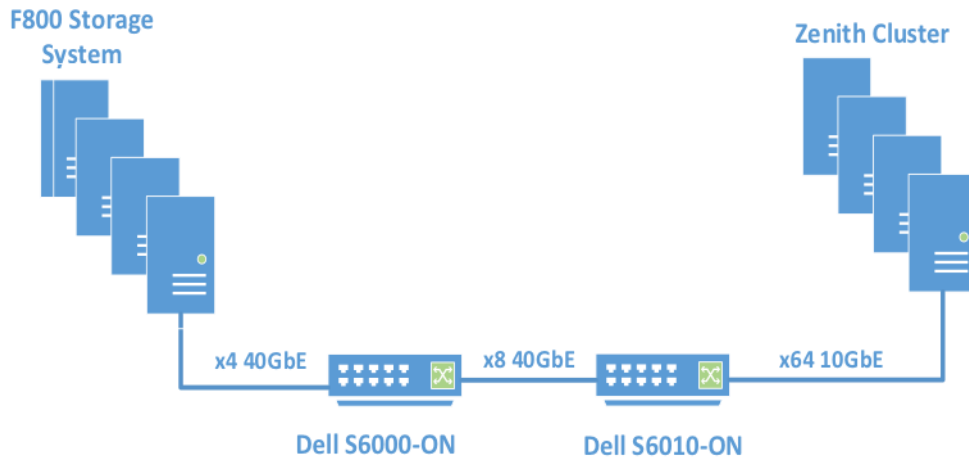


Figure 1. Network Diagram Of The F800 Benchmark Configuration

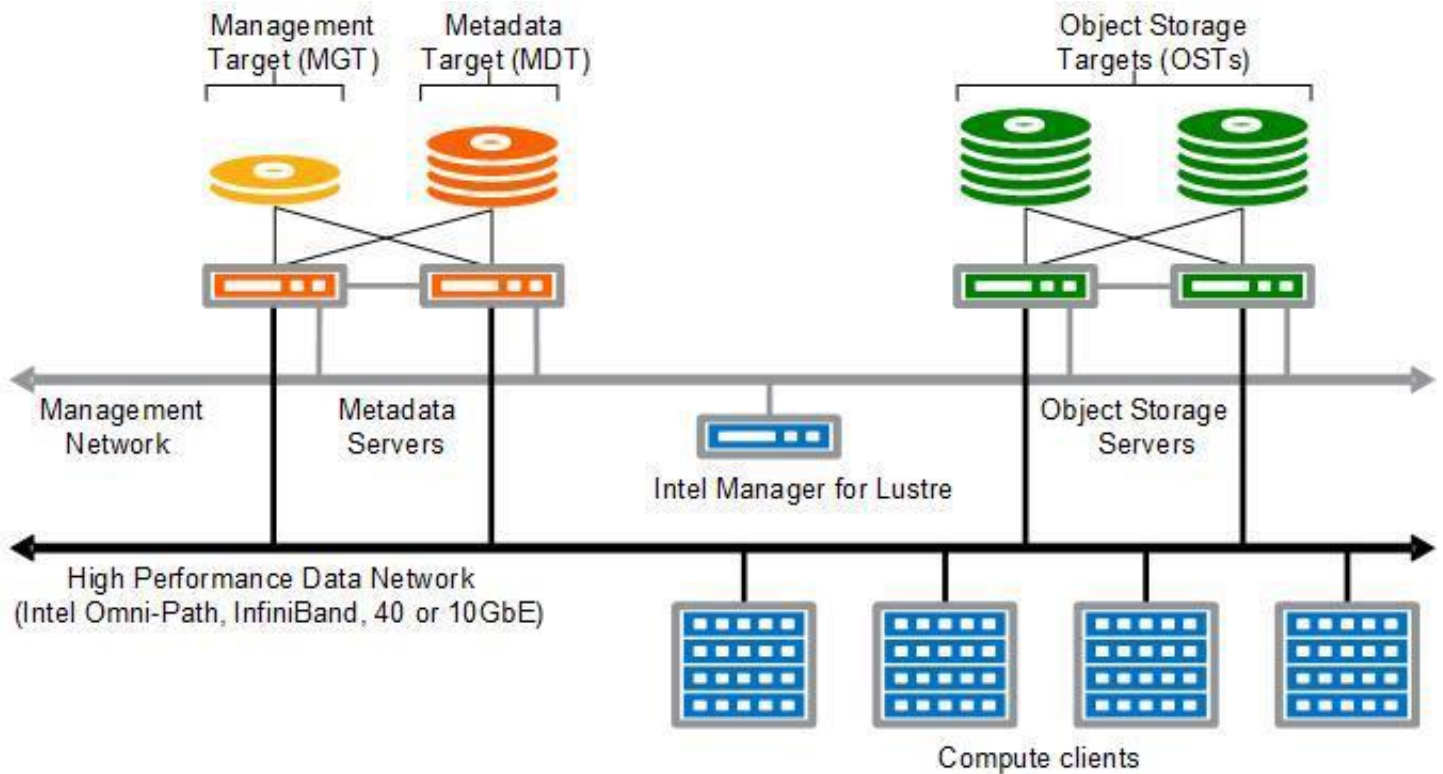
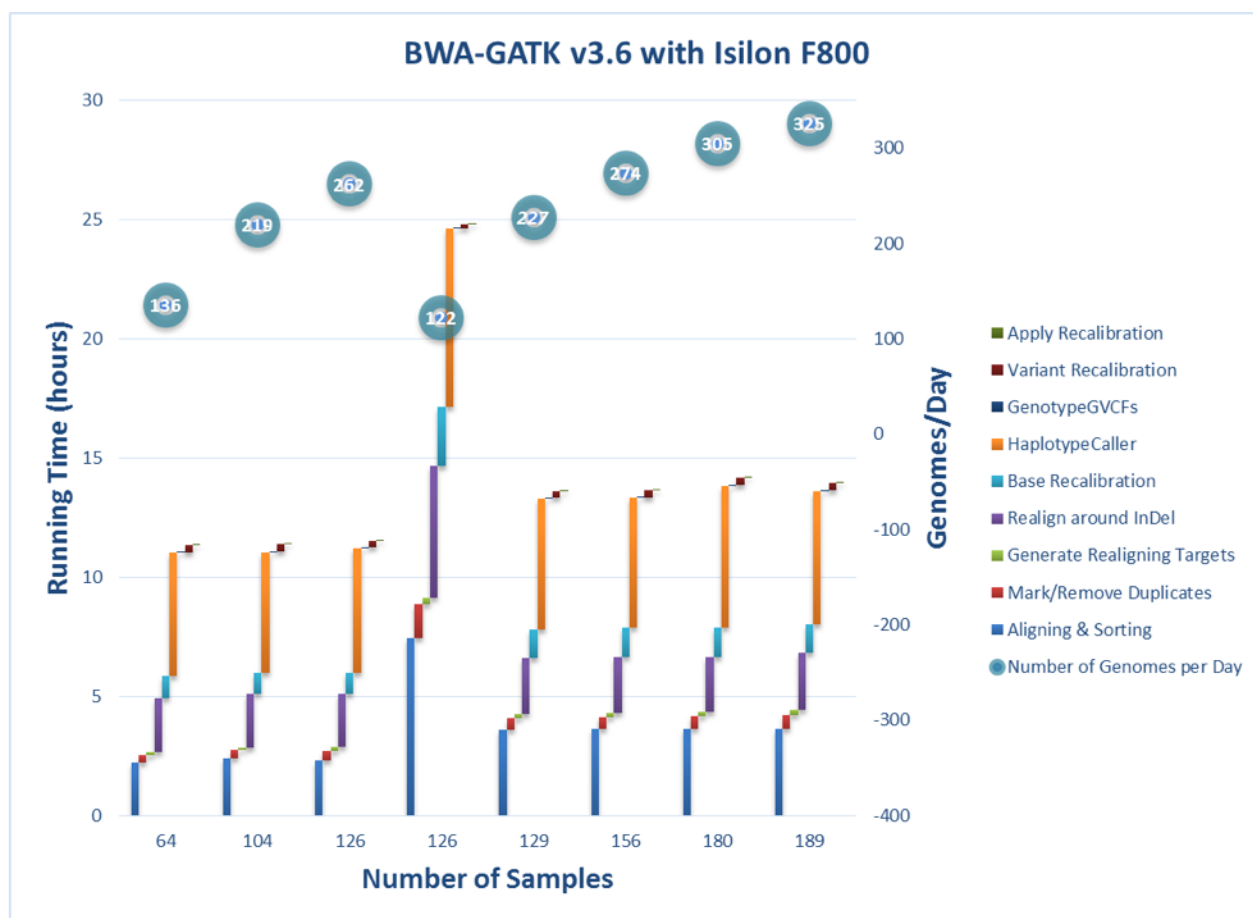


Figure 2. Network Diagram Of The Lustre Benchmark Configuration

WHOLE GENOME SEQUENCE VARIANT ANALYSIS

GATK version 3.5 and BWA version 0.7.2-r1039 were used to benchmark variant calling on the Lustre system, while GATK version 3.6 was used for runs using the F800 and H600. The whole genome workflow was obtained from the workshop, GATK Best Practices⁴, and its implementation is detailed [here](#)⁵ and [here](#)². The publicly available human genome data set used for the tests was ERR091571. ERR091571 is one of Illumina's Platinum Genomes from the NA12878 individual that has been used for benchmarking by many genome analysis developers, and is relatively error free. The data set can be downloaded from the Short Read Archive (SRA) at the European Bioinformatics Institute [here](#)⁶.

To determine the maximum sample throughput possible, an increasing number of genome samples were run on an increasing number of compute nodes with either 2 or 3 samples being run simultaneously on each node. Batches of 64-189 samples were run on 32-63 compute nodes that mounted NFS exported directories from the F800 storage cluster. Figure 3 illustrates the wall-clock time it took for each step in the pipeline (left axis) as well as total run time, while the right axis is a measure of how many genomes per day can be processed utilizing a particular sample size, samples/node ratio and total compute node combination. The samples/node ratio and number of compute nodes used per batch of samples is illustrated beneath the graph. For the 64,104 and 126 sample sizes, 32, 52 and 63 compute nodes were used, respectively, with a sample/node ratio of 2. For the 129,156,180 and 189 sample sizes, 43, 52, 60 and 63 compute nodes were used, respectively, with a sample/node ratio of 3.

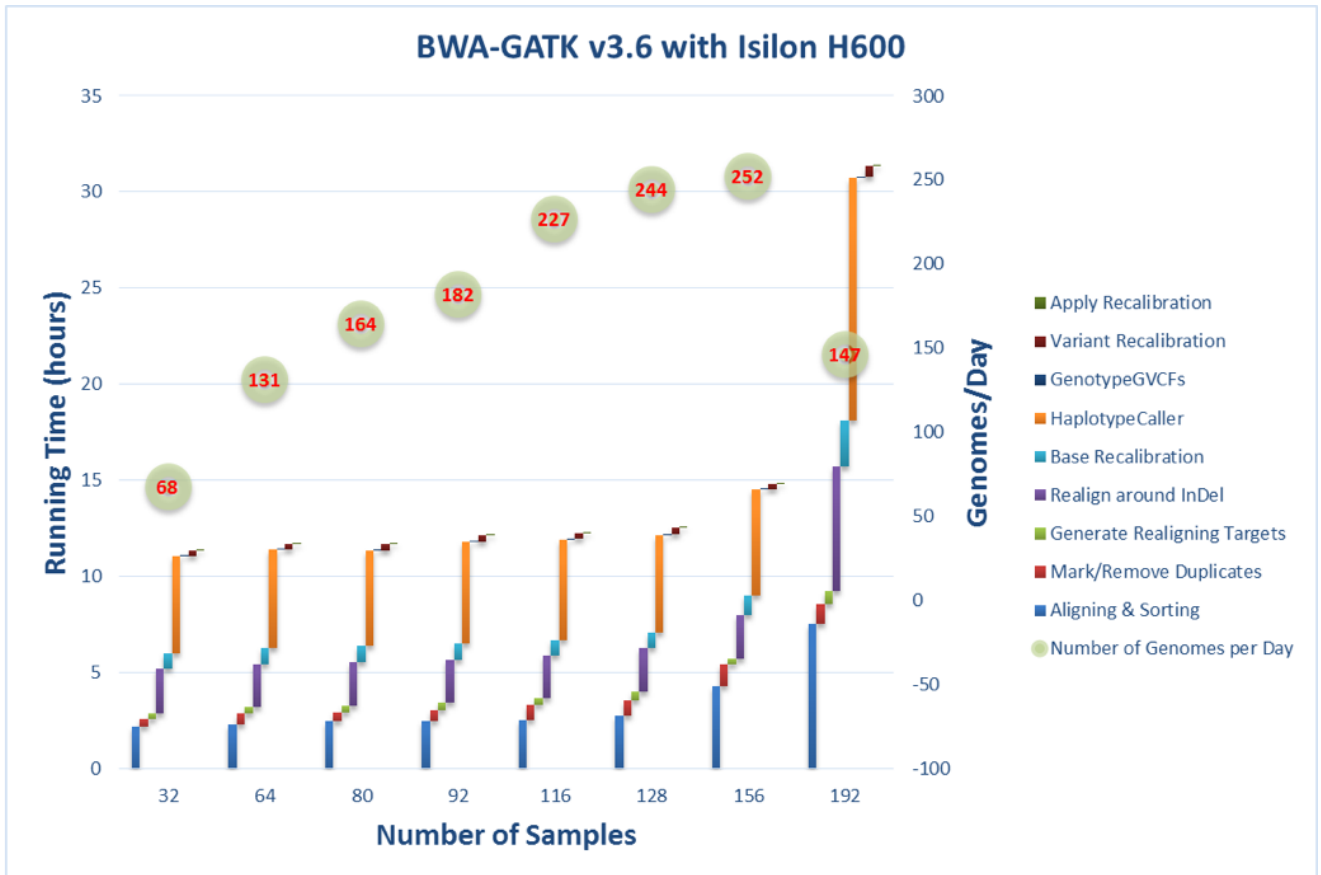


Samples per compute node	2	2	2	2	3	3	3	3
Compute nodes used	32	52	63	63	43	52	60	63

Figure 3. Number of 10x WGS BWA-GATK performance results on F800. The second 126 sample plot is from a test using ~30x genome samples (122 genomes/day).

The benchmark results in Figure 3 illustrate that when running 2 samples/compute node, the total run time is approximately 11.5 hours, while running 3 samples/node yields an approximately 14 hour run time. While the run time is longer when running 3 samples/node the total genomes/day throughput is higher, resulting in 325 genomes/day in the run with 189 samples. Genomes/day is calculated like so: $(24 \text{ hours} / \text{total sample run time (hours)}) \times \text{number of samples} = \text{number of samples that can be processed in a 24-hour period, i.e. genomes/day}$. In the case of the 189 sample run, this equates to $(24 \text{ hours} / 13.94 \text{ hours total run time}) \times 189 = 325.4 \text{ genomes/day}$.

To determine the maximum sample throughput possible while using an H600 for data input and pipeline output, an increasing number of genome samples were run on an increasing number of compute nodes with either 2 or 3 samples being run simultaneously on each node. Batches of 32-192 samples were run on 16-64 compute nodes while using the H600 NFS-mounted to the compute nodes. Figure 4 illustrates the wall-clock time for each step in the pipeline on the left axis as well as total run time, while the right axis is a measure of how many genomes per day can be processed utilizing a particular sample size, samples/node ration and total compute node combination. The samples/node ratio and number of compute nodes used per batch of samples is illustrated beneath the graph. For the 32, 64, 80, 92, 116 and 128 sample sizes, 16, 32, 40, 46, 58 and 64 compute nodes were used, respectively, with a sample/node ratio of 2. For the 156 and 192 sample sizes, 52 and 64 compute nodes were used, respectively, with a sample/node ratio of 3. The best performing samples/node ratio will change depending on the number of cores and memory available on the nodes.



Samples per compute node	2	2	2	2	2	2	3	3
Compute nodes used	16	32	40	46	58	64	52	64

Figure 4. Number 10x WGS BWA-GATK performance results on H600.

The benchmark results in Figure 4 illustrate that when running 2 samples/compute node, the total run time is between 11 and 13 hours, while running 3 samples/node yields an approximately 15 hour run time (156 samples). While the run time is longer when running 3 samples/node the total genomes/day throughput is higher, resulting in 252 genomes/day in the run with 156 samples. While the F800 was able to handle the processing of 189 samples in 14 hours, the H600 could not process 192 samples effectively. The 192 sample total run time was over 30 hours (147 genomes/day), a significant performance decrease compared to the 156 sample run (15 hours). If ≥ 192 genomes/day is required, then an additional H600 should be added to the cluster. Alternatively, as seen in Figure 3, a single F800 can provide that level of performance.

Comparing maximum pipeline throughput performance between the F800 (325 genomes/day) and the H600 (252 genomes/day) results in 73 more genomes/day processed using the F800.

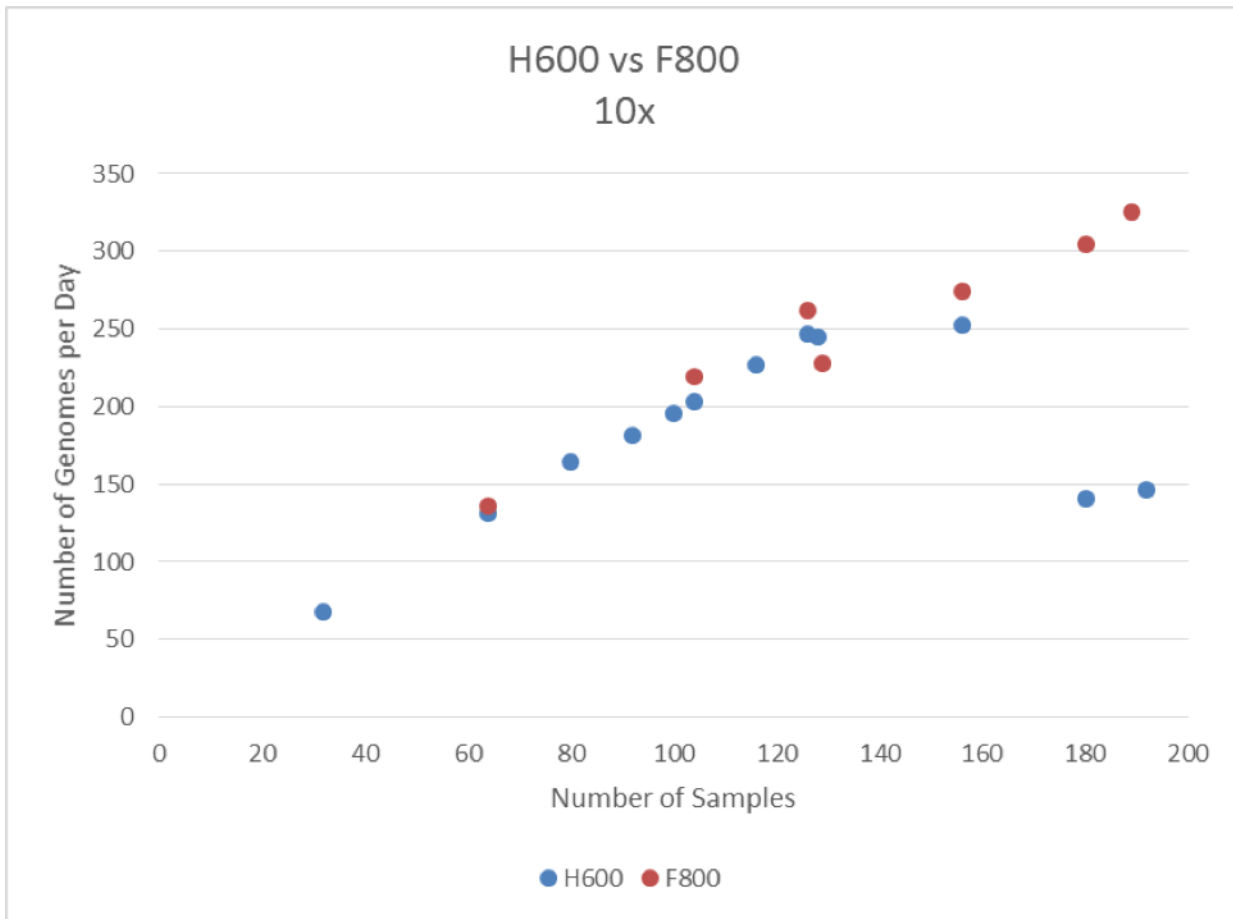


Figure 5. BWA-GATK performance results comparison between F800 and H600.

Plotting genomes/day throughput versus sample size for the F800 and H600 shows that performance on both platforms scales similarly up to 128 samples (Figure 5). Past that, H600 performance levels off and then deteriorates while F800 performance continues to improve. Future tests will utilize more than 64 compute nodes in an attempt to maximize pipeline throughput on the F800.

We can provide a comparison when running the same genome data set using a Lustre filesystem instead of Isilon. In this case, 80 samples were run using 40 compute nodes with 2 samples/node on the Lustre filesystem described in Table 1^{2.5}. This run configuration was also completed on the H600 (Figure 4). The total run time and genomes/day results were nearly identical (Figure 6). The run finished 4 minutes faster on the Lustre system to give it the smallest advantage in calculating genomes/day; H600 = 164.38, Lustre = 165.37. An 80 sample run was not completed using the F800, but if we average the results from the 64 and 104 samples runs, 136 and 219 genomes/day, respectively, we arrive at 177 genomes/day for 84 samples $((104 + 64)/2)$ on the F800, with an average of 42 nodes used $((32 + 52)/2)$ running 2 samples/node. While not the most scientific of interpolations, this inference makes sense given that the H600 and F800 performance scaled similarly up to approximately 128 samples (Figure 5).

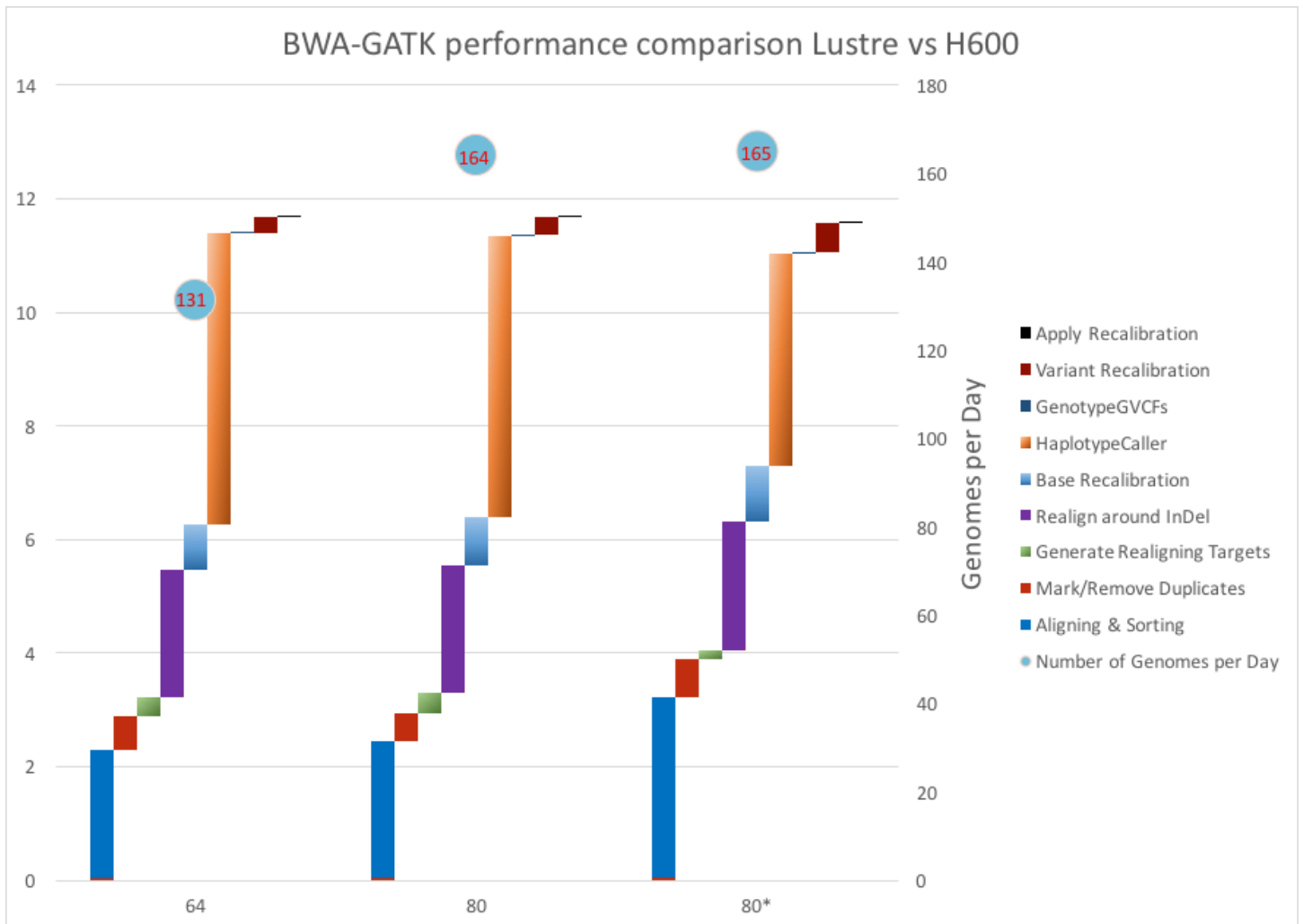


Figure 6. BWA-GATK performance results comparison between H600 and Lustre. The 80* labeled samples were run using a Dell HPC Lustre Storage Solution.

Isilon storage provides an additional performance advantage when running large numbers of genomic analyses that consume large amounts of disk space. As can be seen in Figure 7, genomic analysis performance remains consistent whether the H600 is nearly empty (1% full) or almost completely full (91%). This consistent, scalable performance is critical for projects intent on sequencing thousands or millions of genomes.

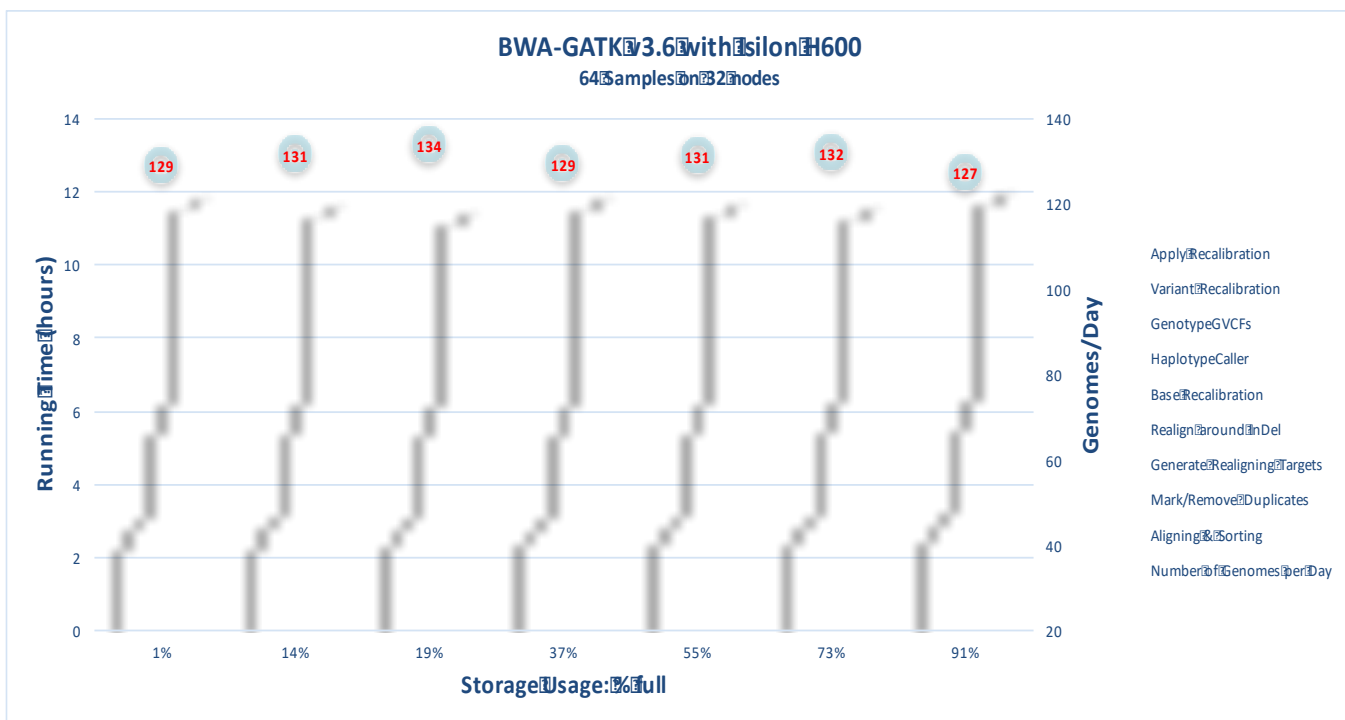


Figure 7. BWA-GATK performance results on H600 as storage usage increases.

SUMMARY

The results in this paper demonstrate that running whole genome analyses on the F800 and H600 platforms scales predictably and that both are capable of supporting hundreds of simultaneous whole genome analyses in a single day. The F800 performs slightly better with low sample workloads and much better with higher sample workloads than the H600, so the F800 is recommended for the highest throughput environments.

If the HPC environment is exclusively for processing genomic analyses, then both Isilon and Lustre are good choices, but Isilon is a better choice if features like backup, snapshots and multiple protocol (SMB/NFS/HDFS) support are required. If Isilon is chosen, then given the results in this paper, a rough genomes/day calculation can be made in order to choose between the F800 or H600. However, if the HPC workload is mixed and includes MPI-based or other applications that require low latency interconnects (Infiniband or OmniPath) in order to scale well, then Lustre is the better choice.

REFERENCES

1. Dell HPC Lustre Storage Solution: [Dell HPC Lustre Storage with IEEL 3.0](#)
2. [Dell EMC HPC System for Life Sciences v1.1 \(January 2017\)](#)
3. All benchmark tests were run on the Zenith cluster in the Dell HPC Innovation Lab in Round Rock, TX. Zenith ranked #292 in the Top 500 ranking as of November 2017: <https://top500.org/list/2017/11/?page=3>
4. **GATK Best Practices** (<https://software.broadinstitute.org/gatk/best-practices/>; Accessed 1/24/2018)
5. Variant Calling Benchmark – Not Only Human (<http://en.community.dell.com/techcenter/high-performance-computing/b/genomics/archive/2016/05/27/variant-calling-benchmark-not-only-human>; Accessed 1/24/2018)
6. [Human genome 10x coverage data set ERR091571](ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR091/ERR091571/) (<ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR091/ERR091571/>; Accessed 1/24/2018)