

# Cryo-EM – Relion benchmarks on Skylake

## Optimized Relion Tests on Dell EMC PowerEdge C6420s with Dell EMC Isilon

### Overview

The purpose of this blog is to validate the optimized [Relion](#) (for **RE**gularised **L**ikelihood **O**ptimizati**N**) on Dell EMC PowerEdge C6420s with Skylake CPUs. Relion was developed from the [Scheres lab at MRC Laboratory of Molecular Biology](#). It uses an empirical Bayesian approach to refining multiple 3D images or 2D class averages for the data generated from **CryoElectron Microscopy** (Cryo-EM). The impressive performance gain from Intel®'s efforts in the collaboration of Relion development team reduced the performance gap between CPUs and GPUs. The CPU/GPU performance comparison results are not shown here; however, the performance gap becomes single digit fold between Skylake CPU systems and Broadwell CPU/Tesla P100 GPU systems.

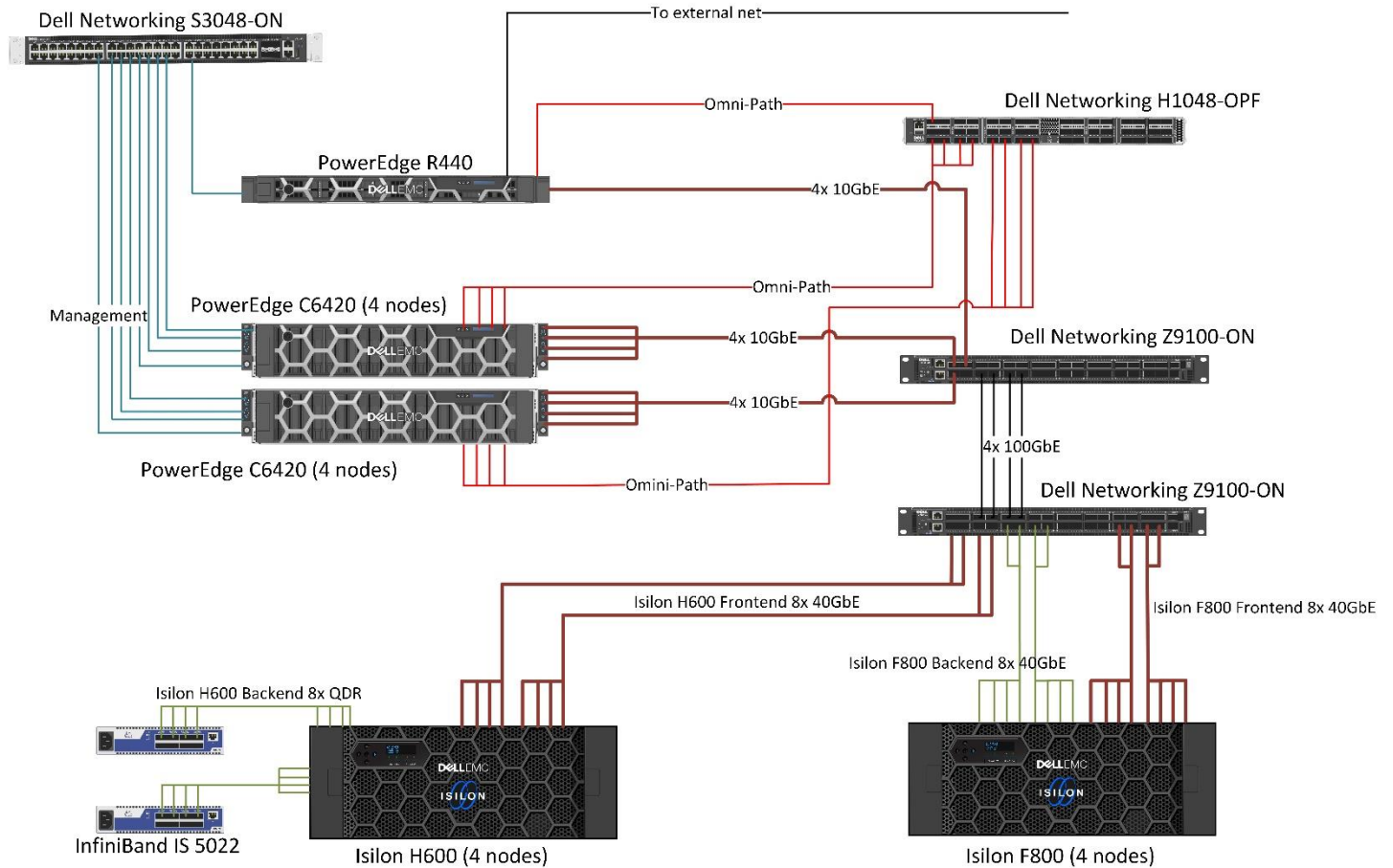
Essentially, Cryo-EM is a type of transmission electron microscopy (TEM) <sup>i</sup> for imaging frozen-hydrated specimens at cryogenic temperatures<sup>ii</sup>. Specimens remain in their native state without the need for dyes or fixatives allowing the study of fine cellular structures, viruses and protein complexes at molecular resolution. A rapid [vitrification](#) at cryogenic temperature is the key step to avoid water molecule crystallization and forming amorphous solid that does almost no damage to the sample structure. Regular electron microscopy requires samples to be prepared in complex ways, and the sample preparations make hard to retaining the original molecular structures. Cryo-EM is not perfect like X-ray crystallography; however, it gains the popularity in the research community fast due to the simple sample preparation steps and flexibility to the sample size, complexity, and non-rigid structure. As the resolution revolution in Cryo-EM progresses due to the 40+ years of dedicated work from the structural biology community, we now have the ability to yield accurate, detailed, 3D models of intricate biological structures at the sub-cellular and molecular scales.

The tests were performed on 8 nodes of Dell PowerEdge C6420s which is a part of Dell EMC Ready Bundle for HPC Life Sciences. A summary of the configuration is listed in **Table 1**.

**Table 1 Test Cluster Configurations**

	Dell EMC PowerEdge C6420
<b>CPU</b>	2x Xeon® Gold 6148F 20c 2.4GHz (Skylake)
<b>RAM</b>	12x 32GB @2666 MHz
<b>OS</b>	RHEL 7.3
<b>Interconnect</b>	Intel® Omni-Path
<b>BIOS System Profile</b>	Performance Optimized
<b>Logical Processor</b>	Enabled
<b>Virtualization Technology</b>	Disabled
<b>Intel Compiler and Libraries</b>	2017.4.196
<b>GLIBC</b>	2.17-196.el7

The test clusters and F800/H600 storage systems were connected via 4 x 100GbE links between two Dell Networking Z9100-ON switches. Each of the compute nodes was connected to the test cluster side Dell Networking Z9100-ON switch via single 10GbE. Four storage nodes in the Dell EMC Isilon F800/H600 were connected to the other switch via 8x 40GbE links. The configuration of the storage is listed in **Table 2**. The detailed network topology is illustrated in **Figure 1**.



**Figure 1** Illustration of Networking Topology

The original tests with an optimized Relion from Intel include Lustre as the main storage; however, we tested the identical optimized Relion binary against Dell EMC Isilon, F800 and H600 due to the limited availability of [Dell EMC Lustre Storage](#) at the time of testing. The detailed running scripts were also obtained from Intel as well to ensure that we reproduce the results.

**Table 2 Storage Configurations**

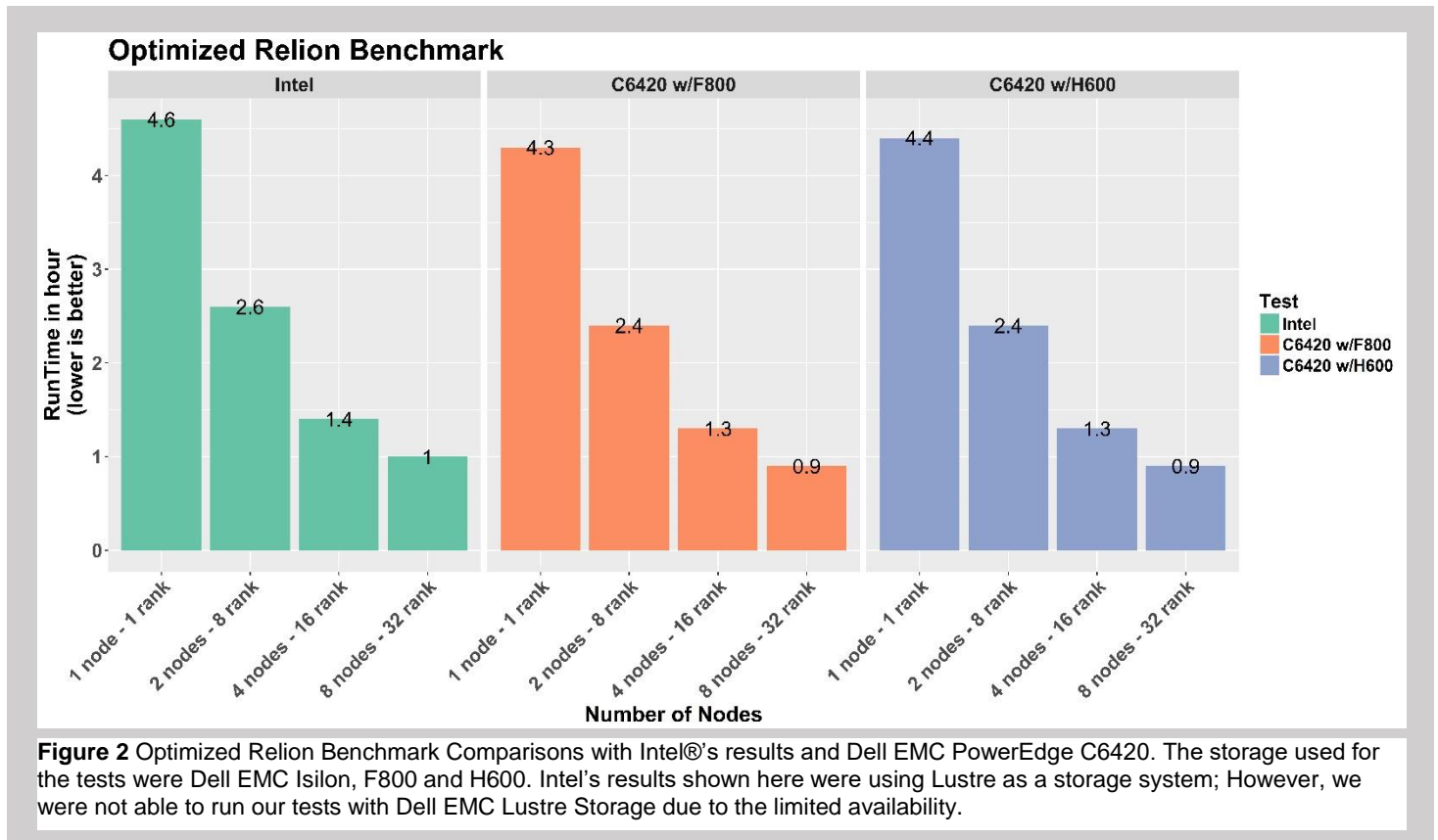
	<b>Dell EMC Isilon F800</b>	<b>Dell EMC Isilon H600</b>
<b>Number of nodes</b>	4	4
<b>CPU per node</b>	Intel® Xeon™ CPU E5-2697A v4 @2.60 GHz	Intel® Xeon™ CPU E5-2680 v4 @2.40GHz
<b>Memory per node</b>	256GB	256GB
<b>Storage Capacity</b>	Total usable space: 166.8 TB, 41.7 TB per node	Total usable space: 126.8 TB, 31.7 TB per node
<b>SSD L3 Cache</b>	N/A	2.9 TB per node
<b>Network</b>	Front end network: 40GbE Back end network: 40GbE	Front end network: 40GbE Back end network: IB QDR
<b>OS</b>	Isilon OneFS v8.1.0 DEV.0	Isilon OneFS v8.1.0.0 B_8_1_0_011

The test data is the recommend standard benchmark data which is presented in [Wong et al., 2014](#). This Plasmodium ribosome data set is download from the [ftp site](#).

## Performance Evaluation

### Performance Comparisons of the Optimized Relion on Dell EMC PowerEdge C6420

Dell EMC PowerEdge C6420s were configured as listed in **Table 1**, and two Dell EMC Isilons, F800 and H600, were used for the tests. Although Intel's results shown here in **Figure 2** from the tests with a Lustre storage, we were not able to test with Dell EMC Lustre Storage due to the limited availability. As shown in **Figure 2**, the running times with various numbers of compute nodes are consistently slightly faster than the Intel's results. This is small variations in our configurations such as Xeon® Gold 6148F instead of non-F series CPUs Intel used and the performance differences in Lustre and Isilon storages.



## Performance Comparisons of Relion without the optimization

**Figure 3** shows Relion performances without the code optimization. The bars labeled with 'Intel E5-2697 v4' are from the Intel's publication. In the CPU test results, Relion v2.0.3 took 47.6 hours on Intel E5-2697 v4 (Broadwell) whereas the same version of Relion took 34.5 hours on Dell EMC PowerEdge C6420 and the latest version of Relion (Relion2-beta-v2.0) took 32.8 hours on Dell EMC PowerEdge R740xd.

Unfortunately, a GPU currently available, NVIDIA Tesla V100, is not compatible with Relion v2.0.3. Hence, the test was performed with Relion2-beta-v2.0. Therefore, the results for GPU tests are not fair comparisons.

NVIDIA Tesla V100 with 16GB memory performs similar to the result from 2 nodes – 8 rank test in **Figure 2**.

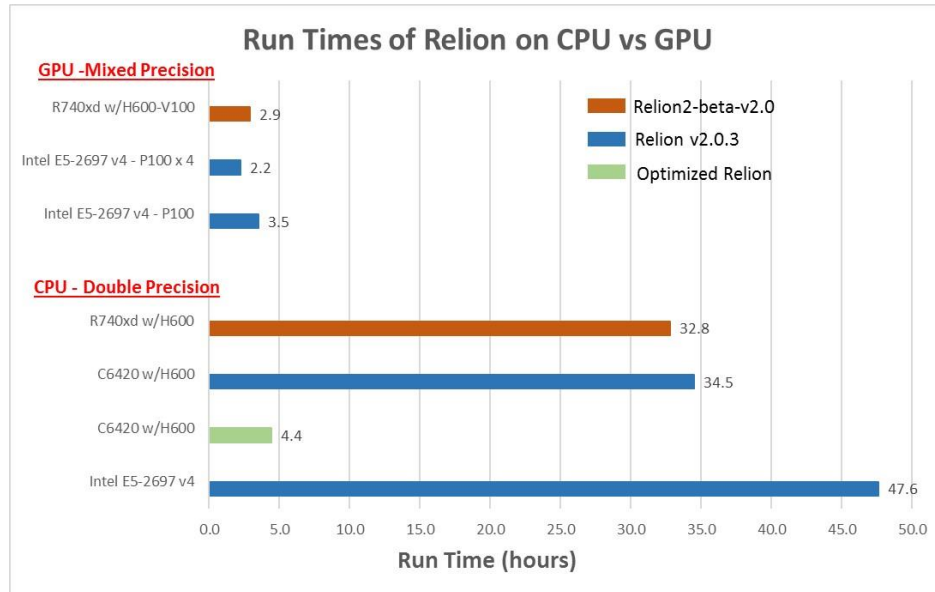


Figure 3 Relion performance comparisons with CPU and GPU

## Conclusion

Dell EMC PowerEdge C6420 shows that it is an ideal compute platform for the Optimized Relion. It scales well over various number of compute nodes with Plasmodium ribosome data. In the future study, we plan to use a larger protein data and more compute nodes to accomplish more comprehensive scaling tests.

## Resources

Internal web page

External web page

### Contacts

Americas

Kihoon Yoon

Sr. Principal Systems Dev Eng

[Kihoon.Yoon@dell.com](mailto:Kihoon.Yoon@dell.com)

+1 512 728 4191

<sup>i</sup> Transmission Electron Microscopy refers to a technique where the image of specimen is formed by directing a high energy electron beam at a thin sample.

<sup>ii</sup> It is not well defined; however, cryogenic temperatures is temperatures below around 123 K, which equals -150°C or -238°F.