Dell EMC HPC Systems - SKY is the limit

Munira Hussain, HPC Innovation Lab, July 2017

This is an announcement about the Dell EMC HPC refresh that introduces support for 14th Generation servers based on the new Intel® Xeon® Processor Scalable Family (microarchitecture also known as "Skylake"). This includes the addition of PowerEdge R740, R740xd, R640, R940 and C6420 servers to the portfolio. The portfolio consists of fully tested, validated, and integrated solution offerings. These provide high speed interconnects, storage, an option for both hardware and cluster level system management, and monitoring software.

On a high level, the new generation Dell EMC Skylake servers for HPC provide greater computation power, which includes support for up to 28 cores and memory speed up to 2667 MT/s; the architecture extends AVX instructions to AVX512. The AVX512 instructions can execute up to 32 DP FLOP per cycle, which is twice the capability of the previous 13th generation servers that used Intel Xeon E5-2600 v4 processors ("Broadwell"). Additionally, the number of core counts per socket is 20% higher per system when compared to the previous generation, which consisted of a maximum 22 cores. It consists of six memory channels per socket; therefore, a minimum of 12 DIMMs are needed for a dual socket server to provide up to full memory bandwidth. The chipset also has 48 PCI-E lanes per socket, up from 40 lanes in the previous generation.

The table below notes the enhancements in the latest PowerEdge servers over the previous generations:

High Level Comparison of the Dell EMC Server Generations for HPC Offering:

<u>PowerEdge R730</u> <u>PowerEdge R740</u> (13th Generation Server) (14th Generation Server) Chipset

| CPU and Chipset | Grantly with Broadwell – EP | Purley with Skylake | | | | | | |
|---|--|--|--|--|--|--|--|--|
| РСН | Wellsburg DMI2 | Lewisburg DMI3 | | | | | | |
| Front Side Bus | Intel [®] QuickPath | Intel [®] UltraPath Interconnect | | | | | | |
| | Interconnect | | | | | | | |
| Processor and cores | | | | | | | | |
| Cores (max) | 22 cores | 28 cores | | | | | | |
| TDP (max) | 145W | 205W | | | | | | |
| Instruction Set | AVX2 | AVX2/AVX-512 | | | | | | |
| Max DP FLOPS/CLK | 16 per core (w/AVX2) | 32 per core (w/AVX-512) | | | | | | |
| | Memory | | | | | | | |
| Memory Channels per Socket | 4 channels, DDR4 | 6 channels, DDR4 | | | | | | |
| Memory Speed | 1DPC=2400, 2DPC =upto | 1 DPC and 2DPC =2133, 2400, | | | | | | |
| (upto max | 2400, 3DPC =1866 | 2666 | | | | | | |
| speed) | | No 3DPC | | | | | | |
| NVDIMM | N/A | Up to 12 NVDIMM | | | | | | |
| GPGPU | | | | | | | | |
| | | | | | | | | |
| GPGPU x16 PCIE Risers | Up to Two Double width | Up to Three Double width | | | | | | |
| GPGPU x16 PCIE Risers | Up to Two Double width Storage | Up to Three Double width | | | | | | |
| GPGPU x16 PCIE Risers PERC | Up to Two Double width Storage PERC 9 | Up to Three Double width PERC 9/10 | | | | | | |
| GPGPU x16 PCIE Risers PERC HardDrive | Up to Two Double width Storage PERC 9 SATA/SAS/SSD | Up to Three Double width PERC 9/10 SATA/SAS/SSD | | | | | | |
| GPGPU x16 PCIE Risers PERC HardDrive | Up to Two Double width Storage PERC 9 SATA/SAS/SSD | Up to Three Double width PERC 9/10 SATA/SAS/SSD Boot Optimized Storage | | | | | | |
| GPGPU x16 PCIE Risers PERC HardDrive | Up to Two Double width Storage PERC 9 SATA/SAS/SSD | Up to Three Double width PERC 9/10 SATA/SAS/SSD <u>Boot Optimized Storage</u> <u>Subsystem (BOSS)</u> | | | | | | |
| GPGPU x16 PCIE Risers PERC HardDrive | Up to Two Double width Storage PERC 9 SATA/SAS/SSD | Up to Three Double width PERC 9/10 SATA/SAS/SSD <u>Boot Optimized Storage</u> <u>Subsystem (BOSS)</u> Support for NVMe | | | | | | |
| GPGPU x16 PCIE Risers PERC HardDrive | Up to Two Double width Storage PERC 9 SATA/SAS/SSD Dell EMC Hardware System | Up to Three Double width PERC 9/10 SATA/SAS/SSD <u>Boot Optimized Storage</u> <u>Subsystem (BOSS)</u> Support for NVMe Management | | | | | | |
| GPGPU x16 PCIE Risers PERC HardDrive System Management | Up to Two Double width Storage PERC 9 SATA/SAS/SSD Dell EMC Hardware System iDRAC8 | Up to Three Double width PERC 9/10 SATA/SAS/SSD <u>Boot Optimized Storage</u> <u>Subsystem (BOSS)</u> Support for NVMe Management iDRAC9 | | | | | | |
| GPGPU x16 PCIE Risers PERC HardDrive System Management IDRAC IP Source | Up to Two Double width | Up to Three Double width PERC 9/10 SATA/SAS/SSD Boot Optimized Storage Subsystem (BOSS) Support for NVMe Management iDRAC9 DHCP | | | | | | |
| GPGPU x16 PCIE Risers PERC HardDrive System Management IDRAC IP Source (Factory default) | Up to Two Double width Storage PERC 9 SATA/SAS/SSD Dell EMC Hardware System IDRAC8 Static IP | Up to Three Double width PERC 9/10 SATA/SAS/SSD <u>Boot Optimized Storage</u> <u>Subsystem (BOSS)</u> Support for NVMe Management iDRAC9 DHCP | | | | | | |
| GPGPU x16 PCIE Risers PERC HardDrive System Management IDRAC IP Source (Factory default) | Up to Two Double width | Up to Three Double width PERC 9/10 SATA/SAS/SSD Boot Optimized Storage Subsystem (BOSS) Support for NVMe Management iDRAC9 DHCP Random for enhanced security | | | | | | |
| GPGPU x16 PCIE Risers PERC HardDrive System Management IDRAC IP Source (Factory default) IDRAC Login (Factory default) | Up to Two Double width | Up to Three Double width PERC 9/10 SATA/SAS/SSD Boot Optimized Storage Subsystem (BOSS) Support for NVMe Management iDRAC9 DHCP Random for enhanced security | | | | | | |
| GPGPU x16 PCIE Risers PERC HardDrive System Management IDRAC IP Source (Factory default) IDRAC Login (Factory default) | Up to Two Double width | Up to Three Double width PERC 9/10 SATA/SAS/SSD Boot Optimized Storage Subsystem (BOSS) Support for NVMe Management iDRAC9 DHCP Random for enhanced security OpenManage 9.0.1 | | | | | | |
| GPGPU x16 PCIE Risers PERC HardDrive MardDrive System Management IDRAC IP Source (Factory default) IDRAC Login (Factory default) System Management | Up to Two Double width | Up to Three Double width PERC 9/10 SATA/SAS/SSD Boot Optimized Storage Subsystem (BOSS) Support for NVMe Management iDRAC9 DHCP Random for enhanced security OpenManage 9.0.1 DTK 6.0.1 | | | | | | |
| GPGPU x16 PCIE Risers PERC HardDrive MardDrive System Management IDRAC IP Source (Factory default) IDRAC Login (Factory default) System Management tools | Up to Two Double width | Up to Three Double width PERC 9/10 SATA/SAS/SSD Boot Optimized Storage Subsystem (BOSS) Support for NVMe Management iDRAC9 DHCP Random for enhanced security OpenManage 9.0.1 DTK 6.0.1 | | | | | | |
| GPGPU x16 PCIE Risers PERC HardDrive System Management IDRAC IP Source (Factory default) IDRAC Login (Factory default) System Management tools | Up to Two Double width | Up to Three Double width PERC 9/10 SATA/SAS/SSD Boot Optimized Storage Subsystem (BOSS) Support for NVMe Management iDRAC9 DHCP Random for enhanced security OpenManage 9.0.1 DTK 6.0.1 | | | | | | |
| GPGPU x16 PCIE Risers PERC HardDrive HardDrive System Management IDRAC IP Source (Factory default) IDRAC Login (Factory default) System Management tools Default mode | Up to Two Double width Storage PERC 9 SATA/SAS/SSD Dell EMC Hardware System IDRAC8 Static IP Standard OpenManage 8.5 DTK 5.5 BIOS Legacy | Up to Three Double width PERC 9/10 SATA/SAS/SSD <u>Boot Optimized Storage</u> <u>Subsystem (BOSS)</u> Support for NVMe Management iDRAC9 DHCP Random for enhanced security OpenManage 9.0.1 DTK 6.0.1 | | | | | | |

| Max Total System Memory BW (STREAM | 130 (GBps) 1DCP with 2400 MT/s | 215 (Gbps) 2DPC with 2666 MT/s | | | |
|--|-----------------------------------|-----------------------------------|--|--|--|
| TRIAD) | | | | | |
| Theoretical max | 1.4 TFlops | 3.2 Tflops | | | |
| System | | | | | |
| Performance | (Dual Socket – 22 c , 145W, | (Dual Socket – 28 c, 205W, | | | |
| (HPL) | 2400 MT/s) | 2666 MT/s) | | | |
| *This is based on | based on AVX2 | Based on AVX-512 code | | | |
| the max core – | | | | | |
| memory | | | | | |
| configuration | | | | | |

The HPC release supporting Dell EMC 14G servers is based on the Red Hat Enterprise Linux 7.3 operating system. It is based on the 3.10.0-514.el7.x86_64 kernel. The release also supports the new version of <u>Bright Cluster Manager 8.0.</u> Bright Cluster Manager (BCM) is integrated with Dell EMC-supported tools, drivers, and third-party software components for the ease of deployment, configuration, and management of the cluster. It includes Dell EMC System Management tools based on OpenManage 9.0.1 and Dell EMC Deployment ToolKit 6.0.1 that help manage, monitor, and administer Dell EMC hardware. Additionally, updated third party drivers and development tools from Mellanox OFED for InfiniBand, Intel IFS for Omni-Path, NVIDIA CUDA for latest Accelerators, and other packages for Machine Learning are also included. Details of the components are as below:

- Based on Red Hat Enterprise Linux 7.3 (kernel 3.10.0-514.el7.x86_64)
- Dell EMC System Management tools from Open Manage 9.0.1 and DTK 6.0.1 for 14G and Open Manage 8.5 and DTK 5.5 for up to 13G Dell EMC servers
- > Updated Dell EMC supported drivers for network and storage deployed during install
 - o megaraid_sas = 7.700.50
 - o igb=5.3.5.7
 - o ixgbe=4.6.3
 - o i40e=1.6.44
 - o tg3=1.137q
 - o bnx2=2.2.5r
 - o bnx2x=1.714.2
- Mellanox OFED 3.4 and 4.0 for InfiniBand
- ▶ Intel IFS 10.3.1 drivers for Omni-Path
- > CUDA 8.0 drivers for NVidia accelerators
- Intel XPPSL 1.5.1 for Intel Xeon Phi processors

Additional Machine Learning packages such as TensorFlow, Caffe, Cudnn, Digits and required dependencies are also supported and available for download

Below are some images of the Bright Cluster Manager 8.0 BrightView:

Figure1: This shows the overview of the Cluster. It displays the total capacity, usage, and job status.



Figure 2: Displays the cascading view of Cluster configuration and respective settings within a group. The settings can be modified and applied from the console.

| - 🏥 I | Bright VIEW Devices > Physical | node | | | | | | ی ح ^ے ک 🕾 • |
|------------|--------------------------------|-------------------------|-------------------|------------------|--------------------|----|-------------------------------|----------------------------|
| 88 | | | | _ ~ ^x | Physical node | | | _ 4 ² |
| | Physical Nodes | TYPE | MAC | OPTIONS | Overview | > | Hostname 🚯 | Tag 🕤 |
| 0 | Switches | Physical node | | Edit 👻 | Settings | > | Hostname | 0000000000000 |
| ā | Power Distribution Units | | | | System Information | \$ | Hostname is required | |
| | Generic Devices | | | | - | ĺ | Mac | |
| Φ | Rack Sensors | | | | Processes | ` | | |
| | GPU Units | Charrie | 00-00-00-00-00 | r du | | | Rack 🕕 | > |
| ø | Metual CMP Modee | Hearl node | FA-16-3E-04/63:43 | Edit - | | | Interfaces () | > |
| | the sets | Gpu unit | 00:00:00:00:00:00 | Edit + | | | 00011-10.1410.9 | |
| ₿ | chassis | Physical node | FA:16:3E:F3:B2:CF | Edit + | | | Category 🕒 | Management network |
| \diamond | MIC Nodes | Physical node | FA:16:3E:2C:38:67 | Edit 👻 | | | derauit | Internainet (derauit) |
| | Nodes Identification | Physical node | FA:16:3E:56:1B:60 | Edit 👻 | | | Enabled Disabled | AUTO (default) |
| = | | Physical node | FA:16:3E:BE:74:05 | Edit 👻 | | | Next install mode () | Data node @ |
| | | Physical node | FA:16:3E:87:F7:9F | Edit 👻 | | | T | Enabled Disabled |
| \$° | | Physical node | 00:00:00:00:00:00 | Edit 👻 | | | PXE Label 🕥 | |
|]ø[| | Physical node | 00:00:00:00:00:00 | Edit 👻 | | | PXE Label | |
| ~ | | Physical node | 00:00:00:00:00:00 | Edit 👻 | | | | |
| <u>•</u> | | Power distribution unit | 00:00:00:00:00 | Edit 👻 | | | Ethernet switch 🕙 | > |
| | | Power distribution unit | 00:00:00:00:00 | Edit 👻 | | | Roles () | > |
| | | | | | | | Kernel modules 🕒 | > |
| | | | | | | | Filesystem mounts () | > |
| | | | | | | | Filesystem exports () | > |
| | | | | | | | Static routes () Reinstall no | de Power) |
| | | | | | | | Sandsroniz | te image Workload |
| | | | | | | | Grab to im | age Software image |
| > | | | Revert Add 🔺 | Delete Save | Back | | | Revert Actions Delete Save |

Figure 3: Dell EMC Settings Tab shows the parsed info on hardware configuration and the required BIOS level settings.

| - 🔅 E | Bright VIEW Devices > node001 > Dell Settings | | | | | | | | ⇔ <i>≃</i> ≡ ▷ ≗· |
|---------------|---|-------------------------|------------------------------------|---|---|-----|---|-----------------------|-------------------|
| | Head Nodes | C Physical node node001 | ysical node node001 k ² | | | | Contraction Contractica Con | _ * ² | |
| .m. | Physical Nodes | Overview | > | Hostname 🕤 | Tag 🖯 | | applyimmediately () | modelName 🖯 | |
| 0 | | Settings | > | node001 | 000000000000000000000000000000000000000 | - 1 | Enabled Disabled | modelName | |
| с П | | System Information | > | Mac 🖯 | | - 1 | serviceTag 🖯 | nodeInterleaving 🖯 | |
| 8 | | Bencorror | | C8:1F:06:88:24:98 | | | serviceTag | Enabled Disabled | |
| Ψ | | FILLESSES | 1 | Rack () | ; | > | snoopMode HomeSnoop | Enabled Disabled | |
| | | | | Interfaces () | , | , | numberDfCores () | BCache () | |
| 2 | | | | BOOTIF - 10.141.0.1, drac0 - 10.148.0.1 | | _ | numberOfCores | Enabled Disabled | |
| | | | | Category () | Management network | . I | turboBoost () | monitor/MWait () | |
| 6 | | | | default | | | Enabled Disabled | Enabled Disabled | |
| \bigcirc | | | | Enabled Disabled | Install mode ALITO (default) | • | biosBootOrder 🖯 | serialCommunication 🖯 | |
| <u>.</u> | | | | Next install mode () | Data mede () | | biosBootOrder | Off | |
| ā | | | | • | Enabled Disabled | | serialPort () | failsafeBaudRate 🕘 | |
| 00 | | | | PXE Label () | | | senari com i senari 2 | 115200 | |
| Tel | | | | PXE Label | | | vt100vt220 · | Enabled Disabled | |
| | | | | Ethemat suitch (3) | | | | | |
| ~ | | | | Ethemet switch () | · · · · · · · · · · · · · · · · · · · | _ | devicefirmware 🖯 | > | |
| * | | | | Roles () | ; | > | nics () | > | |
| | | | | Kernel modules 🕕 | ; | > | svsprofile () | procPwrPerf () | |
| | | | | Filesystem mounts () | ; | > | Custom | MaxPerf • | |
| | | | | Filesystem exports () | ; | > | mmioAbove4Gb () | memFrequency () | |
| | | | | Static routes 🛞 | 2 | > | Enabled Disabled | MaxPerf • | U |
| | | | | Services () | , | > | procTurboMode () | energyEfficientTurbo | |
| | | | | nsica, munge, siurmd, cgconfig, cgred | | _ | | | |
| | | | | GPU Settings 🖯 | ; | > | Enabled Disabled | Enabled Disabled | |
| | | BMC Settings () | | | | > | collaborativeCpuPerfC | ioNonPostedPrefetch | |
| | | | | Dell Settings () | \$ | > | Enabled Disabled | Enabled Disabled | |
| \rightarrow | | Back | | | Revert Actions A Delete Sa | ave | Back | Revert Delete | |

Dell EMC HPC Systems based on the 14th Generation servers expand HPC computation capacity and demands. They are fully balanced and architected solutions that are validated and verified for the customers, and the configurations are scalable. Please stay tuned as follow-on blogs will cover performance and application study; these will be posted here: http://en.community.dell.com/techcenter/high-performance-computing/

Dell - Internal Use - Confidential