DELLEMC

Dell EMC Ready Bundle for HPC Digital Manufacturing—Dassault Systèmes' Simulia Abaqus Performance

This Dell EMC technical white paper discusses performance benchmarking results and analysis for Simulia Abaqus on the Dell EMC Ready Bundle for HPC.

Martin Feyereisen
Joshua Weage

Dell EMC HPC Innovation Lab

January 2018

A Dell EMC Technical Whitepaper

Revisions

Date	Description
January 2018	Initial release

Table of contents

1 Introduction

This technical white paper describes the performance of for Dassault Systèmes' Simulia Abaqus on the Dell EMC Ready Bundle for HPC Digital Manufacturing, which was designed and configured specifically for Digital Manufacturing workloads, where Computer Aided Engineering (CAE) applications are critical for virtual product development. The Dell EMC Ready Bundle for HPC Digital Manufacturing uses a flexible building block approach, where individual building blocks can be combined to build HPC systems which are ideal for customer specific work-loads and use cases. The individual building blocks are configured to provide good performance for specific application types and workloads common to this industry.

The architecture of the Dell EMC Ready Bundle for HPC Digital Manufacturing and a description of the building blocks are presented in Section 2. Section 3 describes the system configuration, software and application versions, and the benchmark test cases that were used to measure and analyze the performance of the Dell EMC HPC Ready Bundle for HPC Digital Manufacturing. Section 4 quantifies the capabilities of the system and presents benchmark performance for Simulia Abaqus.

DELLEMC

2   System Building Blocks

The Dell EMC Ready Bundle for HPC Digital Manufacturing is assembled by using preconfigured building blocks. The available building blocks are infrastructure servers, storage, networking, and application specific compute building blocks. These building blocks are preconfigured to provide good performance for typical applications and workloads within the manufacturing domain. The building block architecture allows for a customizable HPC system based on specific end-user requirements, while still making use of standardized, domain-specific building blocks. This section describes the available building blocks along with the rationale of the recommended system configurations.

2.1   Infrastructure Servers

The infrastructure servers are used to administer the system and provide user access. They are not typically involved in computation or storage, but they provide services that are critical to the overall HPC system. Typically these servers are the master nodes and the login nodes. For small sized clusters, a single physical server can provide these functions. The infrastructure server can also be used for storage, by using NFS, in which case it must be configured with additional disk drives or an external storage array. One master node is mandatory and is required to deploy and manage the system. If high-availability (HA) functionality is required, two master nodes are necessary. Login nodes are optional and one login server per 30-100 users is recommended.

A recommended base configuration for infrastructure servers is:

- Dell EMC PowerEdge R640 server
- Dual Intel® Xeon® Bronze 3106 processors
- 192 GB of memory, 12 x 16GB 2667 MT/s DIMMS
- PERC H330 RAID controller
- 1 x 800GB Mixed-use SATA SSD
- Dell EMC iDRAC9 Enterprise
- 2 x 750 W power supply units (PSUs)
- Mellanox EDR InfiniBand™ (optional)

The recommended base configuration for the infrastructure server is described here. The PowerEdge R640 server is suited for this role. A cluster will have only a small number of infrastructure servers; therefore, density is not a concern, but manageability is more important. The Intel Xeon 3106 processor, with 8 cores per socket, is sufficient for this role. 192 GB of memory provided by 12x16 GB DIMMs provides sufficient memory capacity, with minimal cost per GB, while also providing good memory bandwidth. These servers are not expected to perform much I/O, so a single Mixed-use SATA SSD should be sufficient for the operating system. For small systems (four nodes or less), an Ethernet network may provide sufficient performance. For most other systems, EDR InfiniBand is likely to be the data interconnect of choice, which provides a high throughput, low latency fabric for node-node communications, or access to a Dell EMC NFS Storage Solution (NSS) or Dell EMC Intel Enterprise Edition for Lustre (IEEL) storage solution.

2.2   Explicit Building Blocks

Explicit Building Block (EBB) servers are typically used for Computational Fluid Dynamics (CFD) and explicit Finite Element Analysis (FEA) solvers such as Simulia Abaqus Explicit, LSTC LS-DYNA, ANSYS Fluent®, Siemens' STAR-CCM+®, and others. These software applications typically scale well across many processor cores and multiple servers. The memory capacity requirements are typically modest and these solvers perform minimal disk I/O while solving. In most HPC systems used for Digital Manufacturing, the large majority of servers are EBBs.

The recommended configuration for EBBs is:

- Dell EMC PowerEdge C6420 server
- Dual Intel® Xeon® Gold 6142 processors
- 192 GB of memory, 12 x 16GB 2667 MT/s DIMMS
- PERC H330 RAID controller
- 2 x 480GB Mixed-use SATA SSD in RAID 0
- Dell EMC iDRAC9 Enterprise
- 2 x 1600 W power supply units per chassis
- Mellanox EDR InfiniBand™ (optional)

The recommended configuration for the EBB servers is described here. Because the largest percentage of servers in the majority of systems will be EBB servers, a dense solution is important; therefore, the PowerEdge C6420 server is selected. The Intel Xeon Gold 6142 processor is a 16-core CPU with a base frequency of 2.6 GHz and a maximum all-core turbo frequency of 3.3 GHz. 32 cores per server provides a dense compute solution, with good memory bandwidth per core, and a power of two quantity of cores. The maximum all-core turbo frequency is important because EBB applications are typically CPU bound. This CPU model provides the best balance of CPU cores and core speed. 192 GB of memory using 12x16GB DIMMs provides sufficient memory capacity, with minimal cost per GB, while also providing good memory bandwidth. Relevant applications typically perform limited I/O while solving; therefore, the system is configured with two disks in RAID 0 using the PERC H330 RAID controller, which leaves the PCIe slot available for an EDR InfiniBand HCA. The compute nodes do not require extensive out-of-band (OOB) management capabilities; therefore, an iDRAC9 Express is sufficient. For small systems (four nodes or less), an Ethernet network may provide sufficient performance. For most other systems, EDR InfiniBand is likely to be the data interconnect of choice, which provides a high throughput, low latency fabric for node-node communications, or access to an NSS or IEEL storage solution.

2.3   Implicit Building Blocks

Implicit Building Block (IBB) servers are typically used for implicit FEA solvers such as Simulia Abaqus, MSC Nastran, ANSYS Mechanical™, and others. These applications typically have large memory requirements and do not scale to as many cores as the EBB applications. File system I/O performance can also have a significant effect on application performance.

The recommended configuration for IBBs is:

- Dell EMC PowerEdge R640 server
- Dual Intel® Xeon® Gold 6136 processors
- 384 GB of memory, 24 x 16GB 2667 MT/s DIMMS
- PERC H740P RAID controller
- 4 x 480GB Mixed-use SATA SSD in RAID 0
- Dell EMC iDRAC9 Express
- 2 x 750 W power supply units (PSUs)
- Mellanox EDR InfiniBand™ (optional)

The recommended configuration for the IBB servers is described here. Typically, a smaller percentage of the system will be comprised of IBB servers. Because of the memory and drive recommendations explained here, a 1U PowerEdge R640 server is a good choice. The Intel Xeon Gold 6136 processor is a twelve-core CPU with a base frequency of 3.0 GHz and a max all-core turbo frequency of 3.6 GHz. A memory configuration of 24 x 16 GB DIMMs is used to provide the larger memory capacities needed for these applications. While 384GB is typically sufficient for most CAE workloads, customers expecting to handle very large production jobs should consider increasing the memory capacity to 768GB. IBB applications often have large file system I/O requirements and four Mixed-use SATA SSD's in RAID 0 are used to provide fast local I/O. The compute nodes do not require extensive OOB management capabilities; therefore, an iDRAC9 Express is recommended. InfiniBand is not typically necessary for IBBs because most uses cases only require running applications on a single IBB; however, an InfiniBand HCA can be added to enable multi-server analysis or to access an NSS or IEEL storage solution.

2.4   Dell EMC NSS-HA Storage

The Dell EMC NFS Storage Solution (NSS) provides a tuned NFS storage option that can be used as primary storage for user home directories and application data. The current version of NSS is NSS7.0-HA with options of 240 TB or 480 TB raw disk space. NSS is an optional component and a cluster can be configured without NSS.

NSS-HA is a high performance computing network file system (NFS) storage solution from Dell EMC, optimized for performance, availability, resilience, and data reliability. The best practices used to implement this solution result in better throughput compared to non-optimized systems. A high availability (HA) setup, with an active-passive pair of servers, provides a reliable and available storage service to the compute nodes. The HA unit consists of a pair of Dell EMC PowerEdge R730 servers. A Dell EMC PowerVault MD3460 dense storage enclosure provides 240 TB of storage for the file system with 60 x 4 TB, 7.2K near-line SAS drives. This unit can be extended with a PowerVault MD3060e to provide an additional 240 TB of disk space for the 480 TB solution. Each of the PowerVault arrays is configured with 6 virtual disks (VDs). Each VD consists of 10 hard drives configured in RAID6 (8+2).
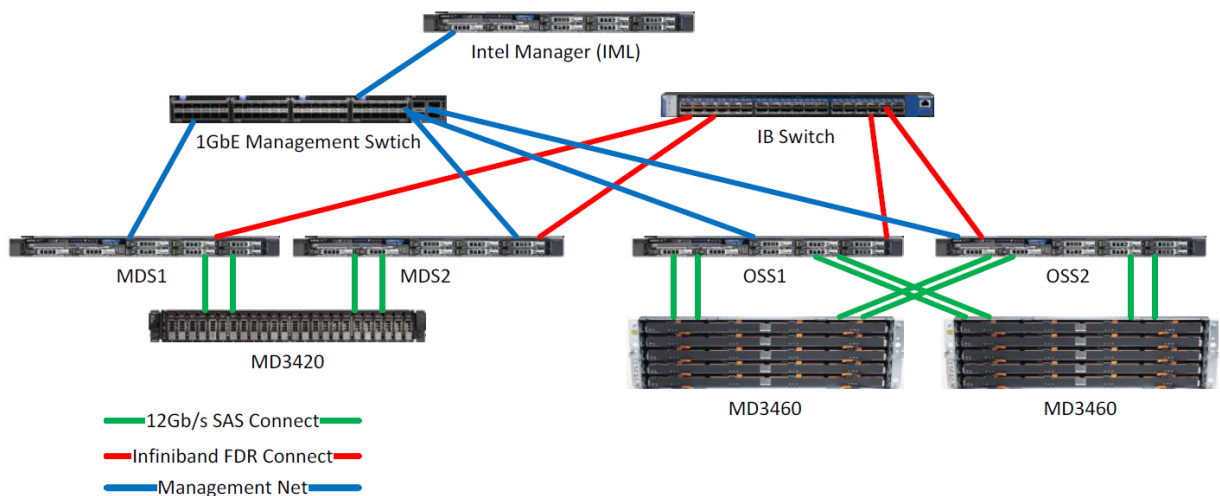
The NFS server nodes are directly attached to the dense storage enclosures by using 12 Gbps SAS connections. NSS7.0-HA provides two network connectivity options for the compute cluster: EDR InfiniBand and 10 Gigabit Ethernet. The active and passive NFS servers run Red Hat Enterprise Linux (RHEL) 7.2 with Red Hat's Scalable File System (XFS) and Red Hat Cluster Suite to implement the HA feature.

2.5   Dell EMC IEEL Storage

Dell EMC IEEL storage is an Intel Enterprise Edition for Lustre (IEEL) based storage solution consisting of a management station, Lustre metadata servers, Lustre object storage servers, and the associated backend storage. The management station provides end-to-end management and monitoring for the entire Lustre storage system.

The Dell EMC IEEL storage solution provides a parallel file system with options of 480 TB or 960 TB raw storage disk space. This solution is typically used for scratch space for larger clusters.

Figure 1   **Overview of the Dell EMC IEEL Components and Connectivity**

2.6   System Networks

Most HPC systems are configured with two networks—an administration network and a high-speed/low-latency switched fabric. The administration network is typically Gigabit Ethernet that connects to the onboard LOM/NDC of every server in the cluster. This network is used for provisioning, management and administration. On the EBB and IBB servers, this network will also be used for IPMI hardware management. For infrastructure and storage servers, the iDRAC Enterprise ports may be connected to this network for OOB server management. The heartbeat ports for NSS-HA and IEEL Ethernet management ports may also be connected to this network. The management network typically uses the Dell Networking S3048-ON Ethernet switch. If there is more than one switch in the system, multiple switches will be stacked with 10 Gigabit Ethernet stacking cables.

A high-speed/low-latency fabric is recommended for clusters with more than four servers. The current recommendation is an EDR InfiniBand fabric. The fabric will typically be assembled using Mellanox SB7890 36-port EDR InfiniBand switches. The number of switches required depends on the size of the cluster and the blocking ratio of the fabric.

DELLEMC

2.7   Cluster Software

The Cluster Software is used to install and monitor the system's compute servers. Bright Cluster Manager (BCM) is the recommended cluster software.

2.8   Services and Support

The Dell EMC Ready Bundle for HPC Digital Manufacturing is available with full hardware support and deployment services, including NSS-HA and IEEL deployment services.

3 Reference System

The reference system was assembled in the Dell EMC HPC Innovation Lab using the building blocks described in Section 2. The building blocks used for the reference system are listed in Table 1.

Table 1.   Reference System Configuration

Building Block	Quantity
Infrastructure Server	1
Explicit Building Block with EDR InfiniBand	8
Implicit Building Block	1
Dell Networking S3048-ON Ethernet Switch	1
Mellanox SB7790 EDR InfiniBand Switch	1

The BIOS configuration options used for the reference system are listed in Table 2.

Table 2.   BIOS Configuration Options

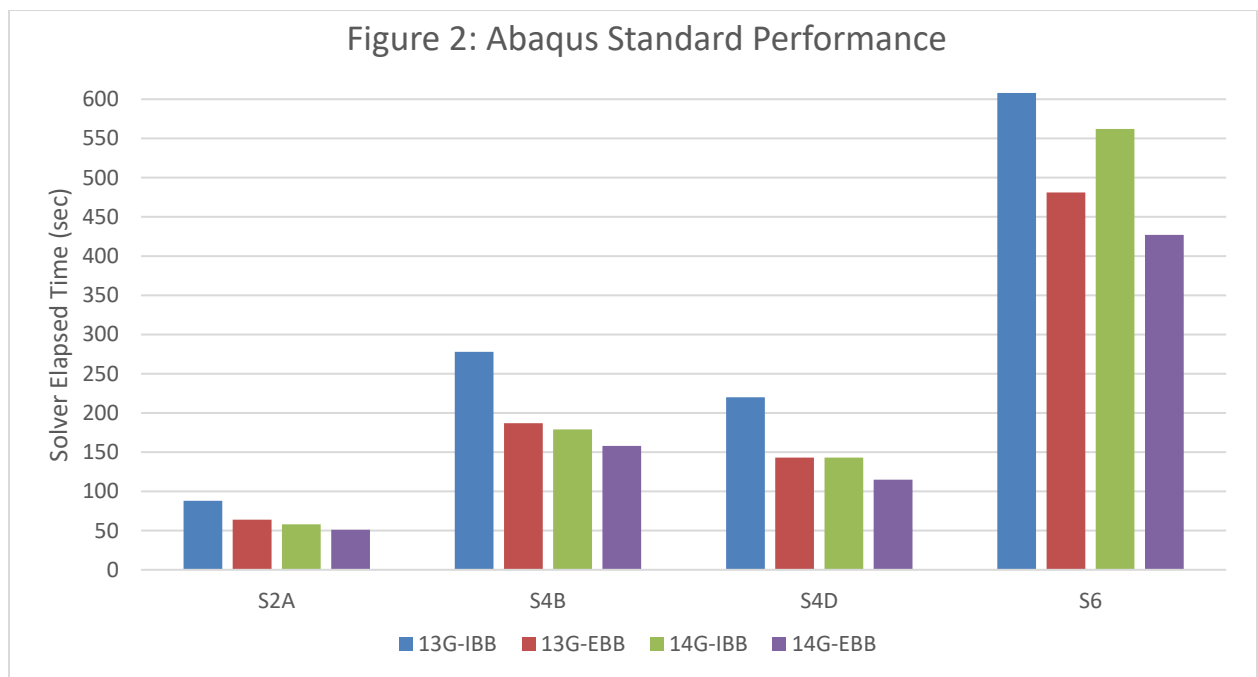| BIOS Option	Setting
Logical Processor	Disabled
Virtualization Technology	Disabled
System Profile	Performance Optimized

The software versions used for the reference system are listed in Table 3.
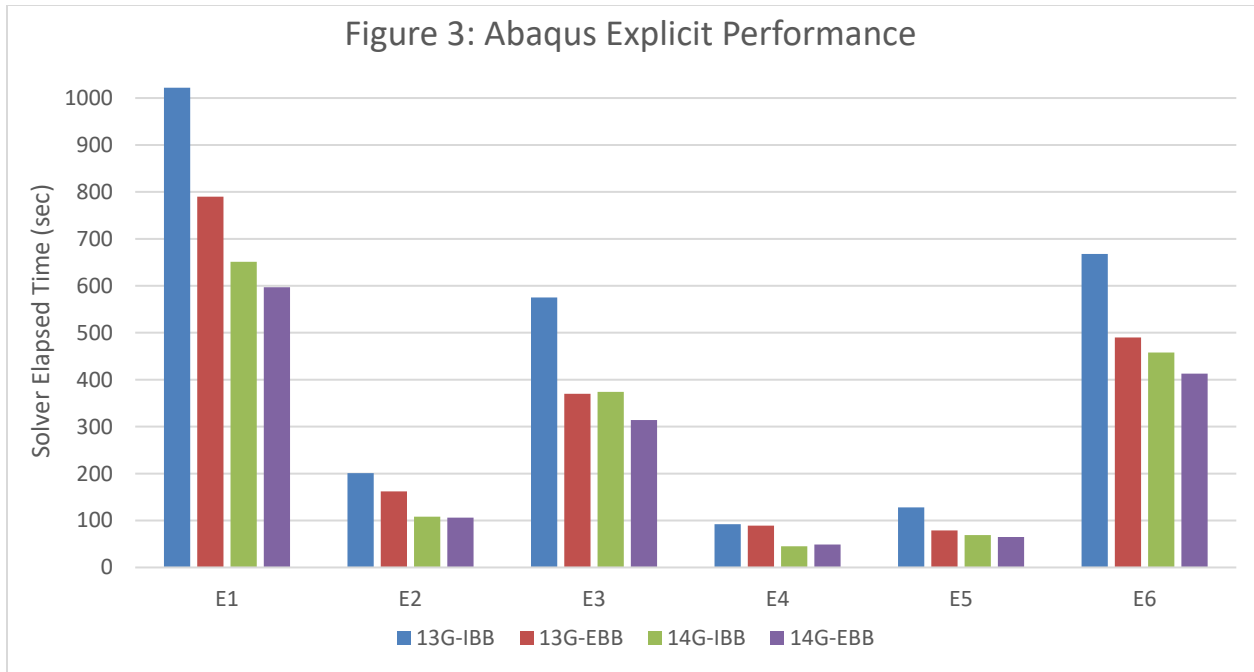
Table 3.   Software Versions

Component	Version
Operating System	RHEL 7.3
Kernel	3.10.0-514.el7.x86_64
OFED	Mellanox 3.4-2.0.0.0
Bright Cluster Manager	7.3 with RHEL 7.3 (Dell version)
Simulia Abaqus	2017

4   Simulia Abaqus Performance

Abaqus is a multi-physics Finite Element Analysis (FEA) software commonly used in multiple engineering disciplines. Depending on the specific problem types, FEA codes may or may not scale well across multiple processor cores and servers. Implicit FEA problems often place large demands on the memory and disk I/O sub-systems. Abaqus contains several solver options, both implicit and explicit. As such it is difficult to summarize the overall performance potential of Abaqus with a few benchmarks. Each Abaqus release distribution does contain some standard benchmarks, both for the implicit solver (Sxx or standard) and the explicit solver (Ex for explicit). These benchmarks are useful to get an indication of the relative performance potential for different systems, so should be viewed more qualitatively than quantitatively. Figure 2 shows a single server performance comparison for four standard Abaqus benchmarks when using all processors cores on the server, where the value for each benchmark is the solver wall time based on the output at the bottom of the .msg file.
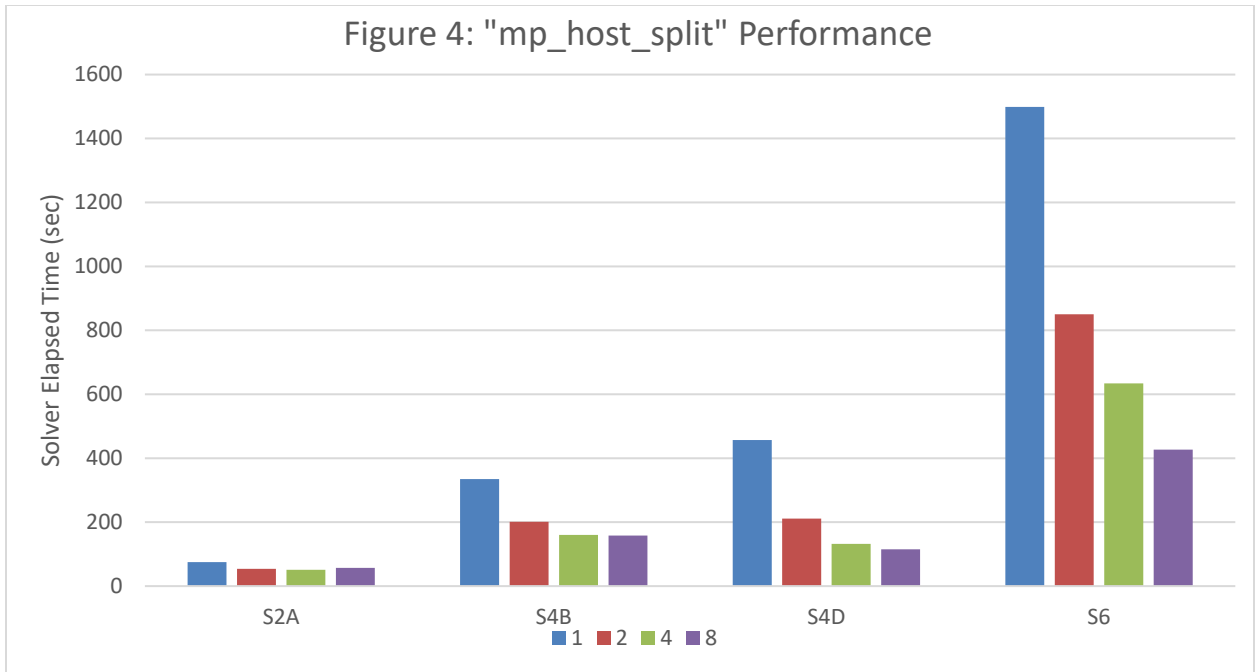

Figure 2: Abaqus Standard Performance

For comparison, corresponding performance with the 13G IBB and EBB systems were also included in the figure, where the 13G IBB was a Dell EMC PowerEdge R630 with dual 8-core Intel Xeon E5-2667v4 processors, and the 13G EBB was the Dell EMC PowerEdge C6320 with dual 16-core Intel Xeon E5-2697Av4 processors. Figure 3 displays the same information for the six explicit benchmarks.

DELLEMC

Figure 3: Abaqus Explicit Performance

These results substantiate that the performance for the 14G servers is noticeably superior to the 13G servers.
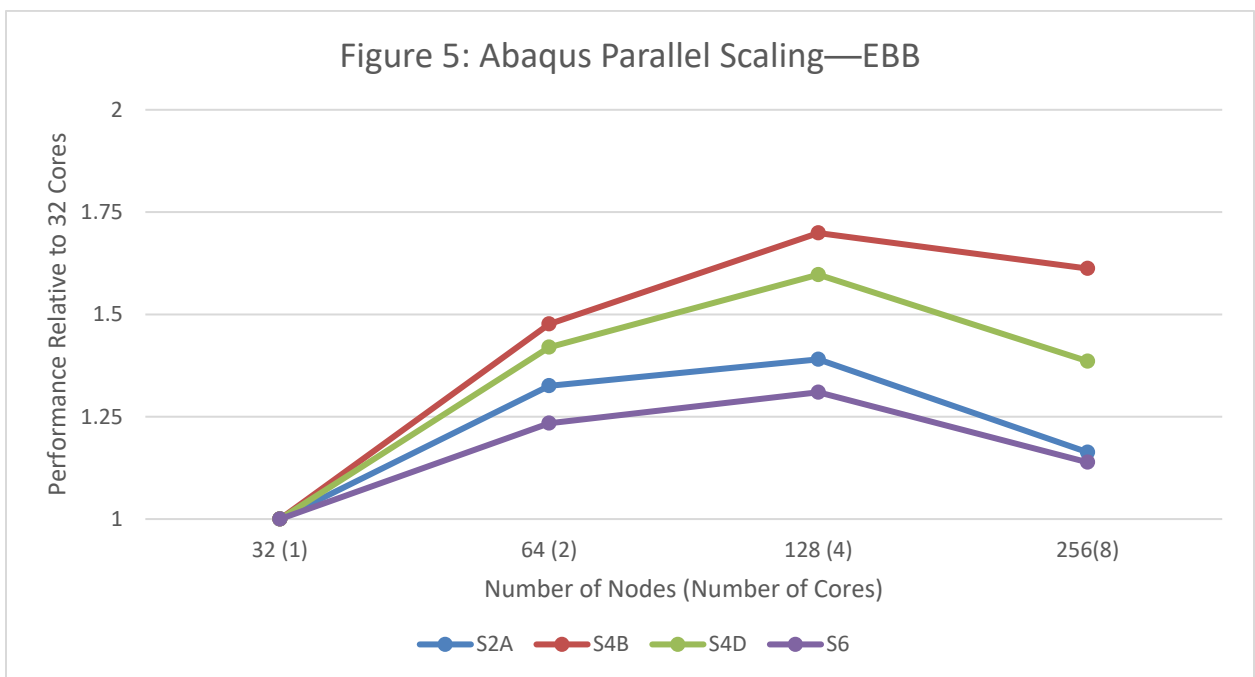
While the explicit solver in Abaqus is a straight forward MPI parallel implementation, the typical standard solver employs a hybrid parallel algorithm using both shared memory parallel threads and MPI domain parallelism. The default run mode for the standard solver is to use a simple MPI domain per server, with parallel threads for each available core on the server. However, the parallel efficiency of the thread parallelism tends to drop off depending on the model size and features after 5-10 threads. Abaqus enables the user to carry out simulations by placing more than a single MPI domain on a server to reduce the number of shared memory parallel threads per domain to increase overall program efficiency. This can be easily activated with the command line argument "mp_host_split=xx" argument. There is no absolute method to *a priori* determine the optimal number of MPI domains per server. Typically an even number is preferred, since it would allow MPI processor binding to be enabled to further improve performance. Typically there is a slight increase in the total memory requirement per server when more MPI domains are placed on each server, and one needs to be careful to avoid "out-of-core" solutions, causing potentially significant I/O activity, decreasing the overall performance. The user can examine the domain memory requirements to minimize I/O in the .dat file to make sure this does not occur.

Figure 4 shows the effect of modifying the number of MPI domains for the four standard benchmarks on the 14G EBB.
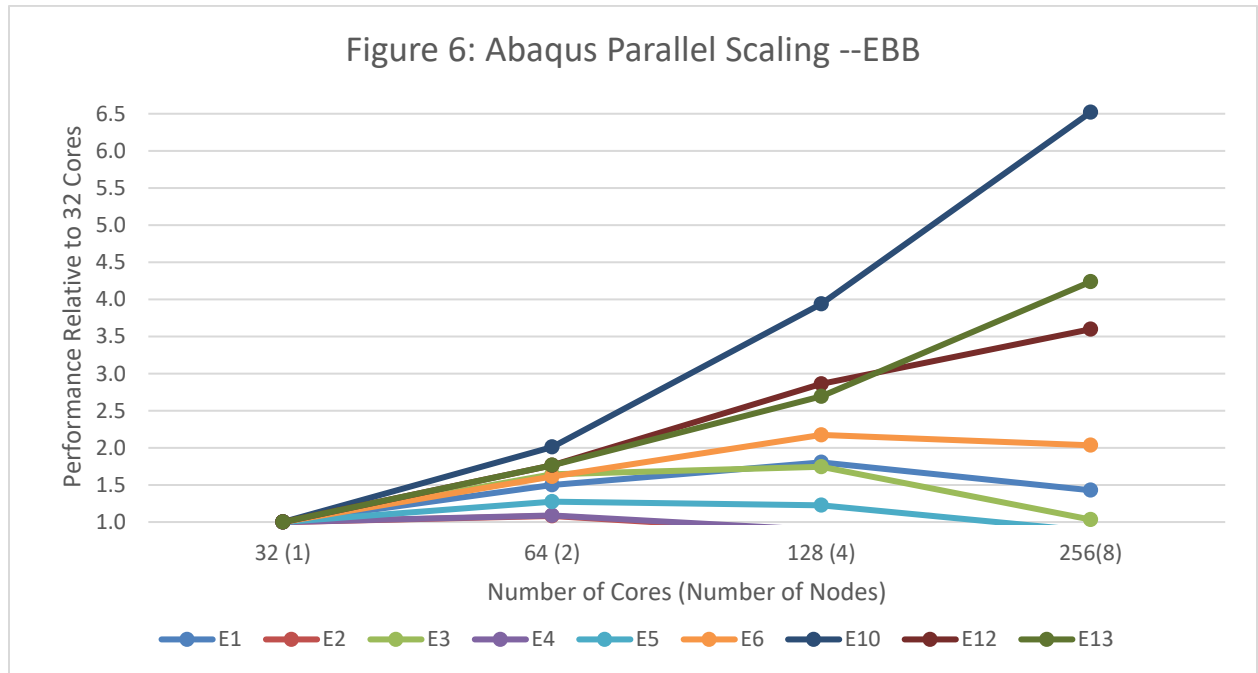
Figure 4: "mp_host_split" Performance

Apart from the very small 'S2A' model, using 8 domains (4 threads per domain) delivers the optimal performance. Users are encouraged to examine this option with their models to determine the optimal value.

Figure 5 shows the measured performance of the test system, on one to eight EBBs, using 32 to 256 cores for the standard benchmarks. Each data point on the graphs reports the performance relative to the single node 32-way parallel result.

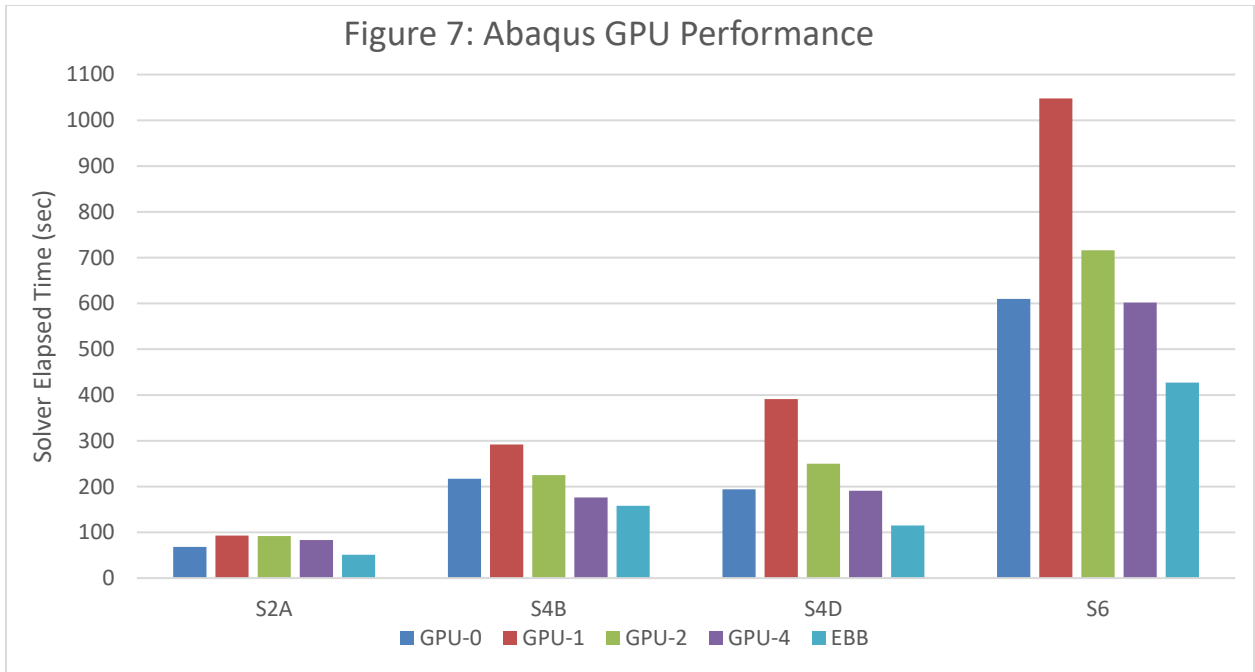Figure 5: Abaqus Parallel Scaling—EBB

DELLEMC

The parallel speedup when running jobs across more than a single node is somewhat mixed. These datasets are rather small by current production standards, and do not represent typical production sized simulations. However, these models do show significant speedup across two nodes, and modest speedups up to four nodes.

Figure 6 shows the parallel speedup for the explicit benchmark cases on the EBBs up to 8 nodes. Each data point on the graphs reports the performance relative to the single node 32-way parallel result.

Figure 6: Abaqus Parallel Scaling --EBB

The smaller data sets (E1-E6) show only modest parallel speedups above two nodes, while the newer and larger benchmarks (E10, E12, E13) display almost linear speedup up to 8 nodes.

Many FEA applications such as Abaqus, have been modified to allow for GPU acceleration with appropriate GPU enabled servers. Figure 7 contains performance information for the four Abaqus standard benchmarks models described above run on a Dell EMC PowerEdge C4130 server, equipped with dual Intel Xeon E5-2680v4 processors and four NVIDIA P100 GPUs, as well as the results of the 14G EBB for comparison. For each benchmark, the wall clock time (in sec) is shown. For the GPU enabled runs, all 26 Xeon cores were used, and runs were made utilizing 0, 1, 2, and 4 GPUs as well.

Figure 7: Abaqus GPU Performance

These results show that host GPUs do not improve performance over using all of the existing Xeon processors alone when all of the Xeon cores are utilized. It may be the case that for certain simulations, GPUs may offer significant performance advantages over the existing host processors, but this is not common, so care should be taken when choosing systems with GPUs.

5   Conclusion

This technical white paper presents a validated architecture for the Dell EMC Ready Bundle for HPC Digital Manufacturing. The detailed analysis of the building block configurations demonstrate that the system is architected for a specific purpose—to provide a comprehensive HPC solution for the manufacturing domain. The design takes into account computation, storage, networking and software requirements and provides a solution that is easy to install, configure and manage, with installation services and support readily available. The performance benchmarking bears out the system design, providing actual measured results on the system for Simulia Abaqus.

References

- [Bright Computing](#), Bright Cluster Manager.
- [Dassault Systèmes' Simulia Abaqus](#).
- [Dell EMC Ready Solutions for High Performance Computing](#).
- [Dell EMC Intel Enterprise Edition for Lustre Storage Solution](#).
- [Dell EMC NFS Storage Solution](#).
- [Dell HPC System for Manufacturing—System Architecture and Application Performance](#), Technical White Paper, July 2016.