Optimizing LS-DYNA® Productivity in Cluster Environments

Gilad Shainer and Swati Kher
Mellanox Technologies

Abstract

Increasing demand for computing power in scientific and engineering applications has spurred deployment of high-performance computing (HPC) clusters. Finite Element Analysis (FEA) and Computational Fluid Dynamics (CFD) are computational technologies that can take advantage of HPC clusters for increasing engineering design productivity, reduce development cost and faster time to market. The end-user benefits are far more sophisticated, enhanced, safer and robust products. With increase usage of multi-core in HPC clusters, FEA and CFD applications need to be highly parallel and scalable in order to fully utilize cluster computing ability. Moreover, multi-core based clusters impose higher demands on cluster components, in particular cluster interconnect. In this paper we investigate the optimum usage of high-performance clusters for maximum efficiency and productivity, for CAE applications, and for automotive design in particular.

Introduction

High-performance computing is a crucial tool for automotive design and manufacturing. It is used for computer-aided engineering (CAE) from component-level to full vehicle analyses: crash simulations, structure integrity, thermal management, climate control, engine modeling, exhaust, acoustics and much more. HPC helps drive accelerated speed to market, significant cost reductions, and tremendous flexibility. The strength in HPC is the ability to achieve best sustained performance by driving the CPU performance towards its limits.

The motivation for high-performance computing in the automotive industry has long been its tremendous cost savings and product improvements. Total cost of real vehicle crash-tests in order to determine its safety characteristics is in the range of $250,000 or more. On the other hand, the cost of a high-performance compute cluster can be just a fraction of the price of a single crash test, while providing a system that can be used for every test simulation going forward.

HPC is used for many other aspects than just crash simulations. Compute-intensive systems and applications are used to simulate everything from airbag deployment to brake cooling, exhaust systems, thermal comfort and windshield washer nozzles. HPC-based simulations and analyses empower engineers and designers to create vehicles that are more ready and safer for real-life environments.

High-Performance and Scalable Simulations

One of the most demanding applications of automotive design is crash simulations (full-frontal, offset-frontal, angle-frontal, side-impact, rear-impact and more). Crash simulations, while performed very early in the development process, are validated very late in the development process once the vehicle is completely built. The more sophisticated and complex the simulation, the more parts and details can be analyzed. Computer-based analyses provide an early insight into phenomena that are difficult to be gathered experimentally, and if so, only at a later stage and at a substantial cost. Time and money is saved without having to build costly prototypes.

High-performance clusters are scalable performance compute solutions based on commodity hardware on a private system network. The main benefits of clusters are scalability, availability, flexibility and high-performance. A cluster uses the combined compute power of compute sever nodes to form a high-performance solution for cluster-based applications. When more compute power is needed, it can be simply achieved by adding more server nodes to the cluster.

Depending on the design complexity, FEA or CFD simulation can be performed on the entire cluster server nodes, utilizing all of the CPU cores in the cluster, or on a segment of the cluster, enabling multiple parallel simulations. Later, the cluster environment provides a flexible dynamic solution, where each simulation can use a different number of CPU cores.

The way the cluster nodes are connected together has a great influence on the overall application performance, especially when multi-core servers are used. In multi-core environments, it is essential to have I/O that provides the same low latency for each process/core, regardless of the number of cores that operate simultaneously, in order to guarantee linear application scalability. As each of the CPU cores generates and receives data, to and from other server nodes the bandwidth capability needs to be able to handle all the data streams. Furthermore, the interconnect should handle all the communication and offload the CPU cores from I/O related tasks in order to maximize the CPU cycles that are dedicated to the applications.

InfiniBand Architecture (IBA) is an industry standard fabric designed to connect between clustered servers and between servers and storage. InfiniBand is designed to provide highest bandwidth, low-latency computing, scalability for ten thousand nodes and multiple CPU cores per server platform and efficient utilization of compute processing resources. InfiniBand server-to-server and server-to-storage connections deliver up to 40Gb/s of bandwidth. InfiniBand switch-to-switch connections deliver up to 120Gb/s. This high-performance bandwidth is matched with ultra-low application latency of 1µs, and switch latencies under 140ns per hop that enable efficient scale out of compute systems.

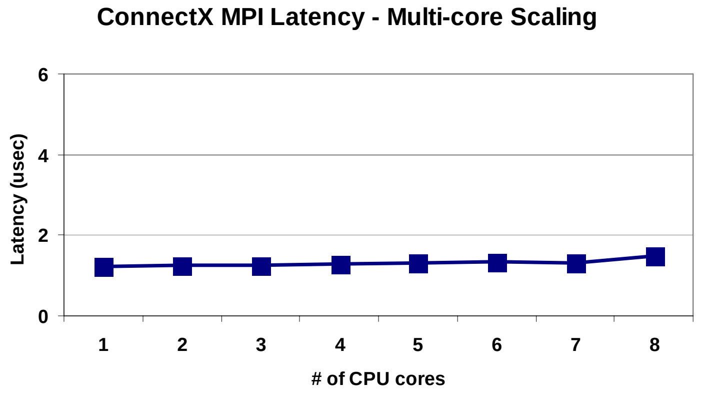

Figure 1 – MPI multi-core latency with Mellanox ConnectX InfiniBand HCA

Figure 1 shows InfiniBand latency scaling on multi-core cluster, where each server includes 8 cores. The latency is being measured using IMB (Pallas) multi-core benchmark for Message Passing Interface (MPI), which is the messaging interface for most of the FEA and CFD applications, including LS-DYNA. MPI is the interface between the applications and the cluster networking. As shown, InfiniBand provides the same low latency regardless on the number of cores that simultaneously carry out the simulations.

Single Server Dilemma – SMP or MPI

A common multi-core environment consists of 8 to 16 CPU cores in a single server. In a typical one server environment, application jobs can be executed in a Shared Memory Processing (SMP) fashion, or with a Message Passing Interface (MPI) protocol. In order to compare between the two options, we have used LS-DYNA neon_refined_revised benchmark.

LS-DYNA is a general purpose structural and fluid analysis simulation software package capable of simulating complex real world problems. It is widely used in the automotive industry for crashworthiness, occupant safety and metal forming and also for aerospace, military and defense and consumer products.

In order to compare between an SMP and MPI approach, a single server was used. The comparison metric was the amount of jobs that can be achieved per 24 hours. According to figure 2, the usage of MPI improves the system's efficiency and parallel scalability, and as more cores were used, the MPI approach performed better versus the traditional SMP.

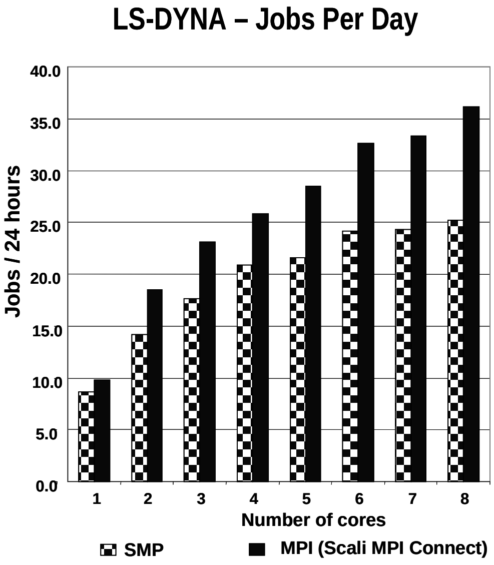

Figure 2: SMP versus MPI (LS-DYNA neon_refined_revised benchmark)

The usage of MPI (in this case, Scali MPI Connect) in a single server not only provides better performance and efficiency, it also enables a smooth integration of a single server into a cluster environment. In addition, when more compute power is needed, no software changes are required.

LS-DYNA Scalability in a multi-core cluster environment: the importance of interconnects

The cluster interconnect is very critical to efficiency and productivity in multi-core platforms. Productivity is being measured by the number of jobs, or simulations that can be performed in a time frame. When more CPU cores are present, the overall cluster productivity is expected to be increased. The productivity and scalability analysis was performed at Mellanox Cluster Center using Neptune cluster cluster. Neptune cluster consist of 16 server nodes, each with dual dual-core AMD Opteron CPUs and 8GB DDR2 memory. The servers were connected with gigabit Ethernet and Mellanox InfiniHost III Ex and ConnectX InfiniBand adapters. LS-DYNA software version mpp971_s_7600 with HP MPI version 2.2.5.1 was used. The InfiniBand drivers were OFED 1.3 from OpenFabrics.

Figure 3 compares between InfiniBand and gigabit Ethernet (GigE) InfiniBand with the new LS-DYNA neon_refined_revised benchmark. The cluster consisted of dual-socket, dual core AMD CPUs server nodes. For any number of cores, InfiniBand shows better efficiency than GigE, enabling up to 39% more LS-DYNA jobs per day with 16 cores, and 1034% higher productivity on 64 cores cluster. Moreover, when scaling beyond 16 cores, GigE failed to provide an increase number of jobs, and even diminishing the overall cluster productivity dramatically. While InfiniBand continued to provide almost linear scalability and high-efficiency, adding any additional CPU cores beyond 16 cores for GigE connected cluster, causes the productivity of the compute cluster to be reduced to levels lower than the one achieved with only 16 cores or even 8 cores.

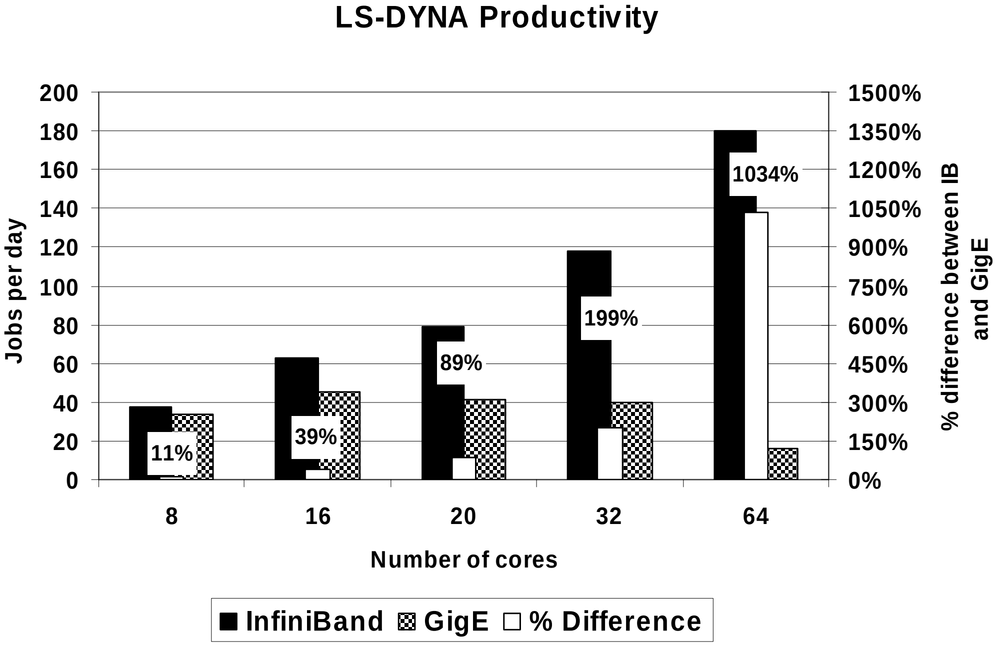

Figure 3: InfiniBand versus GigE (LS-DYNA neon_refined_revised benchmark)

As more jobs can be performed, the quality of the product increases and the time-to-market reduces. The results show that InfiniBand based clusters are required for as low as 16 cores cluster, and is a must for clusters beyond 16 cores. Any additional servers beyond 16 cores connected with GigE only reduce the cluster productivity. Moreover, InfiniBand based cluster with only 16 cores, achieves higher productivity compared to any cluster size with GigE.

Profiling of LS-DYNA Network Activity

In order to understand the I/O requirements of LS-DYNA, we have analyzed the network activity during neon_refined_revised benchmark for two cases- 16 cores and 64 core clusters. The results of the total data that was sent at the corresponded MPI message size (or I/O message size) are shown in figure 4. Need to note that the Y axis is a logarithmic scale.

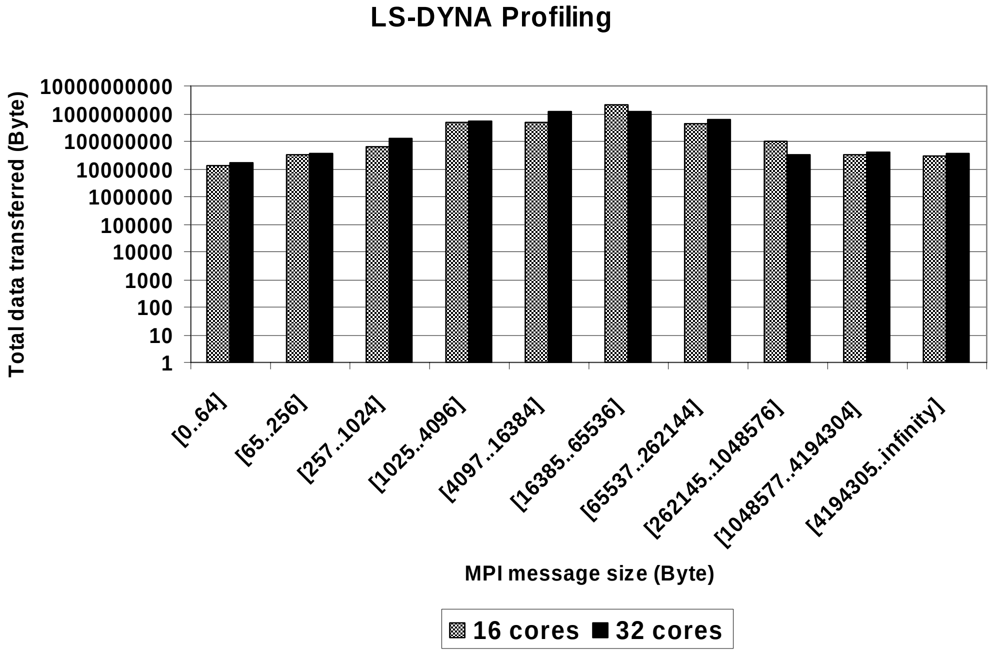

Figure 4 – neon_refined_revised network profiling

There peaks of I/O throughput throughout neon_refined_revised execution are in the range of 1KB to 256KB message sizes, which reflect the need for high bandwidth interconnect solution. Moreover, LS-DYNA uses very large messages, of a 1MB and higher in size which explain why GigE is not adequate for car crash simulations.

As the number of cores (and cluster server nodes) increases from 16 servers to 32 cores, higher number of small messages (2B-64B) was seen, this is a results of the need to synchronize more servers. In regards to the data messages, the increase of the number of core and servers resulted in the increase of the total data send via the 4KB-16KB and 64KB-256KB messages. As clusters increase in size, there are more control messages (small size) and more large data messages.

The network profiling results show that with the increase in cluster size there is a higher number of control messages (small size), hence the need for lowest latency, and larger data messages, hence the need for highest throughput. We can conclude that higher throughput (such as InfiniBand QDR 40Gb/s) and lower latency interconnects will continue to improve LS-DYNA performance and productivity.

Conclusions

From concept to engineering and from design to test and manufacturing; engineering relies on powerful virtual development solutions. Finite Element Analysis (FEA) and Computational Fluid Dynamics (CFD) are used in an effort to secure quality and speed up the development process. Cluster solutions maximize the total value of ownership for FEA and CFD environments and extend innovation in virtual product development.

Multi-core cluster environments impose high demands for cluster connectivity throughput, low-latency, low CPU overhead, network flexibility and high-efficiency in order to maintain a balanced system and to achieve high application performance and scaling. Low-performance interconnect solutions, or lack of interconnect hardware capabilities will result in degraded system and application performance.

Livermore Software Technology Corporation (LSTC) LS-DYNA software and the new LS-DYNA benchmark case was investigated. In all InfiniBand based cases, LS-DYNA had demonstrated high parallelism and scalability, which enable it to take full advantage of multi-core HPC clusters. Moreover, according to the results, a low-speed interconnect, such as GigE is ineffective with any cluster size, starting from 16 cores cluster size (4 nodes) and above.

We have profile the networking usage of LS-DYNA software to determine LS-DYNA requirements from the interconnect, which have lead to the pervious conclusions. The results indicted that large data messages are highly used and the amount of the data send via the large message sizes increase with cluster size. We also evidenced the large amount of small messages, which mostly used to synchronize between the cluster server nodes. From those results we have concluded that a combination of very high bandwidth and extremely low latency, with low CPU overhead, interconnect is required in order to maintain and sustain high scalability and efficiency.

From all of the above results, it is clear that GigE can no longer be used as a cluster interconnect for FEA or CFD applications such as LSTC LS-DYNA.

In order to achieve the desired virtual design, CAE users rely on HPC clusters to gain the needed compute ability. InfiniBand based multi-core cluster environments provide the needed high-performance compute system for the CAE simulations, and the CAE software is ready to take full advantage of the hardware setting.