The Future of High–Performance Computing (HPC)

Herbert Cornelius
Intel Corporation EMEA, Germany

INTRODUCTION

High-Performance Computing (HPC) is used to address and solve the world's most complex computational problems. For decades, HPC has established itself as an essential tool for discoveries, innovations and new insights in science, research and development, engineering and business across a wide range of application areas in academia and industry. It has become an integral part of the scientific method – the third leg along with theory and experiment.

Today, High-Performance Computing is also well recognized to be of strategic and economic value – HPC matters and is transforming industries (Osseyran & Giles, 2015).

High Performance Computing enables scientists and engineers to solve complex and large science, engineering, and business problems using advanced algorithms and applications that require very high compute capabilities, fast memory and storage, high bandwidth and low latency throughput, high fidelity visualization, and enhanced networking.

Today, the IT industry is being transformed by cloud, big data, social media, artificial intelligence, and "Internet of Things" technologies and business models. All of these trends require advanced computational simulation models and powerful highly scalable systems. Hence, sophisticated HPC capabilities are critical to the organizations and companies that want to establish and enhance leadership positions in their respective areas.

Some industry verticals and application areas where HPC is used are as follows:

- Manufacturing, Computer Aided Engineering (CAE)
- Automotive Industry
- Aerospace Industry
- Weather Forecast and Climate Research
- Energy, Oil & Gas Industry, Geophysics
- Life-Science and Bio-Informatics (Genomics)
- Government Research Laboratories
- Universities (Academics), Machine Intelligence, Machine/Deep Learning, Artificial Intelligence (AI)
- Astrophysics, High-Energy Physics, Computational Chemistry, Material Science
- Financial Services Industry(FSI)
- Digital Content Creation (DCC)
- Defense
- Security and Intelligence

BACKGROUND

After its initial years of proprietary computer systems in the 1970/1980's, HPC has evolved with industry standards that democratized Supercomputing, making advanced computing available to more users and wider application segments.

Today's modern HPC solutions are utilizing high-performance server compute nodes connected with high performance fabrics connected to high-performance storage systems, mainly deployed on distributed cluster architectures running on Linux based operating systems with up to tens of thousands of processors. For specific workloads and applications with the need for large

coherent shared memory capacity (terabytes of data: TB), more specialized solutions and systems are used based on cc:NUMA (Cache-Coherent Non-Uniform Memory Access) architectures. For example, the SGI UV system supports up to 256 CPU sockets and up to 64TB of cache-coherent shared memory in a single system.

While in the past chip designs used to be limited by space and the number of transistors available, now power consumption is becoming the main constraint for High-Performance Computing. With several new emerging technologies there will be multiple opportunities to address some of the ongoing challenges in HPC such as balanced architectures, energy efficiency, density, reliability, resiliency, sustainability, and last but not least ease-of-use.

As stated in a recent article (Bader, 2016), the trend in high- performance computing (HPC) to adopt more accelerators such as graphics processing units (GPUs), field-programmable gate arrays (FPGAs), and coprocessors has led to significant heterogeneity in computation and memory subsystems. Application developers typically employ a hierarchical message passing interface (MPI) programming model across the cluster's compute nodes for internode communication, and an intranode model (normally based on shared memory) such as OpenMP for the CPUs or an accelerator-specific library (usually based on an off-load model) such as compute unified device architecture (CUDA) or open computing language (OpenCL) for accelerator devices within each compute node of a cluster.

The following sections will discuss some of the new emerging technologies that are being developed for all parts of HPC to enable balanced solutions regarding data processing (compute), memory, storage, fabric and I/O, as well as parallel software and programming aspects for HPC.

Processing

A modern HPC solution can be characterized by its functional high level components consisting of the following parts as shown in Figure 1:

- **Compute:** Advanced calculations with data in fast memory performing mathematical and logical operations.
- **Processing:** Including data movements, data packing/unpacking, compression/decompression and encryption/decryption of data.
- **Visualization:** High fidelity visualization of results and data, often in-situ.
- Storage: fast and high capacity permanent data storage.
- **Fabric:** High performance interconnect to connect all parts of a HPC solution together.
- **Networking:** To provide access to other systems and traditional IT infrastructures.

Figure 1. High level view of HPC solution components
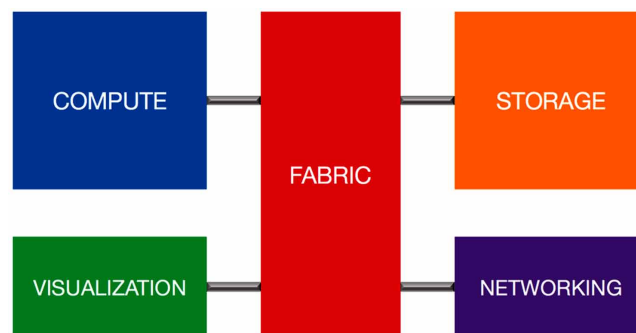
Related Content

AHP-BP-Based Algorithms for Teaching Quality Evaluation of Flipped English Classrooms in the Context of New Media Communication
Xiaofeng Wu (2023). *International Journal of Information Technologies and Systems Approach (pp. 1-12).*
www.irma-international.org/article/ahp-bp-based-algorithms-for-teaching-quality-evaluation-of-flipped-english-classrooms-in-the-context-of-new-media-communication/322096

Grey Wolf-Based Linear Regression Model for Rainfall Prediction
Razeef Mohd, Muheet Ahmed Buttand Majid Zaman Baba (2022). *International Journal of Information Technologies and Systems Approach (pp. 1-18).*
www.irma-international.org/article/grey-wolf-based-linear-regression-model-for-rainfall-prediction/290004

Modified Distance Regularized Level Set Segmentation Based Analysis for Kidney Stone Detection
K. Viswanathand R. Gunasundari (2015). *International Journal of Rough Sets and Data Analysis (pp. 24-41).*
www.irma-international.org/article/modified-distance-regularized-level-set-segmentation-based-analysis-for-kidney-stone-detection/133531

Islamic Medical Ethics in Brunei Darussalam Hospitals
Syazana Fauzi (2021). *Encyclopedia of Information Science and Technology, Fifth Edition (pp. 1860-1871).*
www.irma-international.org/chapter/islamic-medical-ethics-in-brunei-darussalam-hospitals/260313

Hybrid TRS-FA Clustering Approach for Web2.0 Social Tagging System
Hannah Inbarani Hand Selva Kumar S (2015). *International Journal of Rough Sets and Data Analysis (pp. 70-87).*
www.irma-international.org/article/hybrid-trs-fa-clustering-approach-for-web20-social-tagging-system/122780