Tick Tock+: Sustained, Smart, and Steadfast Journey for Faster High Performance Computing

Sponsored by IBM and Intel – March 2010

Srini Chari, Ph.D., MBA.

mailto:chari@cabotpartners.com

Cabot Partners Group, Inc. 100 Woodcrest Lane, Danbury CT 06810, www.cabotpartners.com

Cabot Partners Optimizing Business Value

Introduction – IBM and Intel Enhance HPC Solutions with Improved x86 Architectures

Over the last decade, IBM has *sustained* a leadership position in High Performance Computing (HPC) across a wide range of application and industry segments. The cornerstone of IBM's long-term success has been a *steadfast* strategy of building a rich partner ecosystem and portfolio of *smart* solutions, particularly with its Intel based System x family. Powered with the latest generation of Intel Xeon processor 5600 series - code named Westmere-EP (**Tick**) and the Intel Xeon processor 7500 series - code named Nehalem-EX (**Tock**), IBM has significantly enhanced its System x portfolio for HPC. These economical and faster HPC solutions can substantially increase computing density and memory in a standard enterprise rack or the IBM BladeCenter, or scale-out to 100s of very energy-efficient racks with the IBM System x iDataPlex.

x86 architectures drive sustained HPC cluster growth but data center challenges must be overcome.

Scientists, engineers, and analysts eager to solve challenging problems in engineering, financial services and the life and energy sciences rely on high performance systems. These users are some of the most demanding clients of IT solutions. Their insatiable demand for computational performance continues to drive x86 high performance computing platforms towards increasing performance and system scalability, larger computing density, lower power consumption, efficient system cooling, easier and more reliable system management, and adaptable programming environments to support a large growing portfolio of HPC applications.

Multiple industry research studies indicate that revenue from HPC servers will continue to grow much faster than overall server revenues with x86 clusters being the dominant platform –almost 75 percent of HPC servers. But, these studies also suggest that the costs of server management, power and cooling, facilities management, and people costs in HPC data centers will outpace the costs of buying new servers. This has caused a severe crisis in HPC data centers. IBM and Intel have responded by designing innovative solutions that address these issues while retaining all the attractive attributes of x86 cluster architectures for HPC.

IBM differentiates with smart innovations in packaging, power and cooling, and reliability.

Significant reductions in power consumption, costs, and space requirements can be achieved through the use of the latest Intel multi-core processors that deliver more performance, more cores, and faster memory at the same or lower power. New packaging and cooling methods, traditionally used for large enterprise class systems, are very effective for designing energy-efficient clusters. Better air cooling supplemented with liquid cooling, better placement of components at the board, chassis, and rack levels, and the aggregation of power supplies can significantly boost energy efficiency. This also allows more reliable operation of HPC clusters in a data center while significantly reducing the total cost of ownership (TCO).

IBM and Intel continue steadfast collaborations with HPC partners and application developers.

Worldwide, IBM and Intel continue active collaborations with leading HPC applications developers and partners to migrate and optimize their applications on the System x portfolio to solve challenging problems in the life sciences, financial services, upstream petroleum, engineering, government, and academia. A customer's prior investment in application enablement and skills is completely protected and they can expect a significant boost in their application performance by upgrading to the new IBM System x servers powered by the latest Intel Xeon processors. They also benefit from the flexibility of integrating the best cluster components for storage, interconnect fabrics, cluster management and file system software, applications, and a single point of contact from IBM for problem resolution and support. Customers across several industries and applications will now be able to execute larger HPC workloads faster, at lower costs, with greater simplicity in deployment, and a very rapid payback. IBM plans to provide access to these new systems at several worldwide cloud and innovation centers for testing and benchmarking.

+ "Tick-Tock" is a model adopted by chip manufacturer Intel Corporation since 2007 to follow every microarchitectural change with shrinking of the process technology. Every "tick" is a shrinking of process technology of the previous microarchitecture and every "tock" is a new microarchitecture. Every year, there is expected to be one tick or tock to drive innovation and to continue delivering to Moore's law.

Clients must examine their ROI Prudently as HPC Workloads Span a Wide Range

The current economic downturn and the escalating energy and people costs for HPC, will force companies to reevaluate how they can maximize their return on IT. They will need smarter approaches to reduce costs, manage complexity, improve productivity, reduce time to market, and enable innovation. Simply put, clients must carefully examine the business (value and costs) case of HPC investments *for their specific workloads.*

HPC workloads are large, complex, and diverse. HPC clients benefit from prudent workload analysis.

HPC clients have always demanded computing solutions that have the best mix of *functionality, performance, and price/performance.* HPC workloads are limited by several key performance-bottlenecking kernels that stress different aspects of an HPC system, particularly the microprocessor and I/O architectures within a server node. Evaluating systems (and underlying processors) and making purchasing decisions solely on benchmarks such as LINPACK or even gigaflops/watt is a flawed strategy for several reasons. First, to deliver the client's specific business functionality, the client application workload must be available and must perform well on the HPC system. Second, clients must often pay a significant added people cost (beyond IT acquisition cost) to port, optimize, and manage these diverse applications with non-standard architectures. Lastly, these benchmarks are not indicative of the broad diversity of HPC workloads across multiple industries and applications.

Conceptual analysis of HPC application specific workloads mapped to system features.

The following figure depicts expected performance for a range of HPC applications mapped to key system features at the server level: frequency, number of cores/socket, memory capacity, memory bandwidth, and I/O performance. For simplicity and focus, inter-server interconnect features are not considered here.

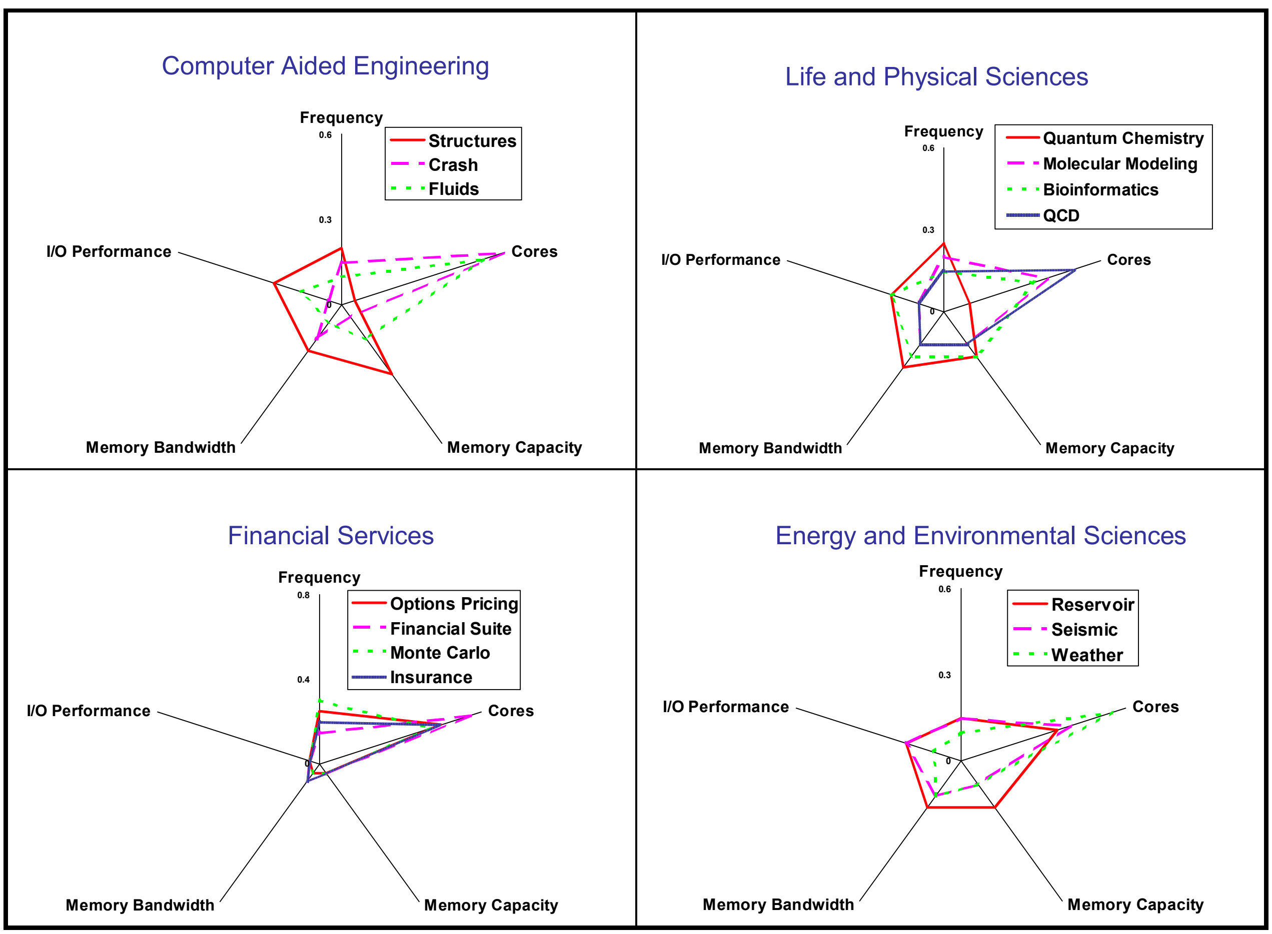


Figure 1: Application specific workload performance analysis driven by key system features

Four major application categories representing today's mainstream use of HPC are considered: Computer Aided Engineering (CAE), Life and Physical Sciences, Financial Services, and Energy and Environmental Sciences. Each category is further divided into several segments. Within each segment, the relative contributions (normalized to one) of system features that typically impact performance are indicated by the colored polygonal lines. We have used actual performance data obtained from Intel and IBM for our high-level conceptual workload analysis described here as follows for each application category:

Computer Aided Engineering (CAE): Industrial CAE environments rarely rely on just one kind of application. Instead, they often run a mix of applications largely provided by a few key providers such as ABAQUS, ANSYS, CD-adapco, ESI, LSTC, and MSC.Software. Most finite element implicit structures applications typically scale only to a few cores/server but speed-up with greater frequency, larger memory capacity and bandwidth, and higher performance I/O. Many crash analysis applications scale with large number of cores and more memory bandwidth but require less memory capacity or I/O. Computational Fluid Dynamics (CFD) applications typically scale well with more cores and also benefit from more memory and I/O.

Life/Physical Sciences: Quantum Chemistry applications such as Gaussian, NWCHEM, and GAMESS benefit more from balanced system features. Molecular Dynamics applications such as Amber, NAMD, and Gromacs typically benefit from more cores but also need other system features to perform well. Some bioinformatics applications like HMMER can benefit greatly with larger number of cores, while sequence assembly applications could benefit from larger memory capacity/bandwidth and frequencies. Quantum Chromodynamics (QCD) applications could benefit greatly with larger number of cores/server.

Financial Services: Adding more cores/server and increasing frequency will benefit almost all financial services applications ranging from options pricing, risk analysis, and actuarial analysis for insurance.

Energy/Environmental Sciences: Many prestack seismic migration applications benefit greatly with more cores/server. Some memory bandwidth sensitive algorithms like reservoir simulation, reverse time and wave equation migration also benefit from added cores/server and more memory capacity and I/O performance.

Beyond workload analysis, a cost-value framework is also needed to assess HPC investments.

The escalating costs of building and operating data centers are being primarily driven by increasing energy, facilities, and other operational costs. Evaluating systems solely on IT acquisition costs and price/performance is no longer a reasonable business practice, since IT acquisition costs as a component of the Total Cost of Ownership (TCO) are declining. The TCO over several years must be assessed in order to make objective cost decisions while evaluating various solution options. But the TCO alone is inadequate. What's needed is a framework of inter-related drivers and associated metrics that examine the total costs incurred and the value delivered by HPC solutions for *client-specific workloads.*

Value

- Business Value: e.g. customer revenues, new business models, compliance regulations, better products, increased business insight, faster time to market, and new breakthrough capability
- Operational Value: e.g. faster time to results, more accurate analyses, more users supported, improved user productivity, better capacity planning
- IT Value: e.g. improved system utilization, manageability, administration, and provisioning, scalability, reduced downtime, access to robust proven technology and expertise.

Costs

- Data Center Capital e.g. new servers, storage, networks, power distribution units, chillers, etc.
- Data Center Facilities e.g. land, buildings, containers, etc.
- Operational Costs: e.g. labor, energy, maintenance, software license, etc.
- Other Costs: e.g. system management, deployment and training, downtime, migration, etc.

Enterprises must evaluate *their workloads* within this broad cost-benefit framework. They must maintain a focus on *maximizing performance and application uptime to reduce "time to results"* – not just for one application but a collection of workloads typical in production HPC environments. This is particularly crucial for large-scale parallel HPC applications that execute over several hours or even days.

In addition to contributing to a substantial increase in TCO, "slower time to results" often translates to a many fold loss in enterprise business value. If users experience consistent, prolonged wait times for job execution, they can not only get frustrated with their HPC infrastructure and support (deteriorating user-experience and hampering innovation), but can also realize lower productivity (engineers may have to wait several days during critical product development cycles), loss of revenues (a one day of delay in bringing out a new drug to market in the pharmaceutical industry could result in a loss of $3M/day[1]), or often be unable to complete their time and mission-critical simulation as is the case for end-of-day processing in financial services.

IBM System x and BladeCenter Portfolio: Optimized for a Wide Range of HPC Workloads

The IBM System x family, with a wide-array of workload optimized systems, provides an unprecedented range of customer choices of economical systems powered by the Intel Xeon processor 5600 series (Westmere-EP, IBM-M3) and the Intel Xeon processor 7500 series (Nehalem-EX, IBM-eX5). These systems can be tailored for specific HPC applications in engineering, life sciences, upstream petroleum, financial services, and scientific research, or can be configured to optimize a collection of HPC workloads. This unprecedented flexibility empowers customers to make more informed decisions for *their workloads.*

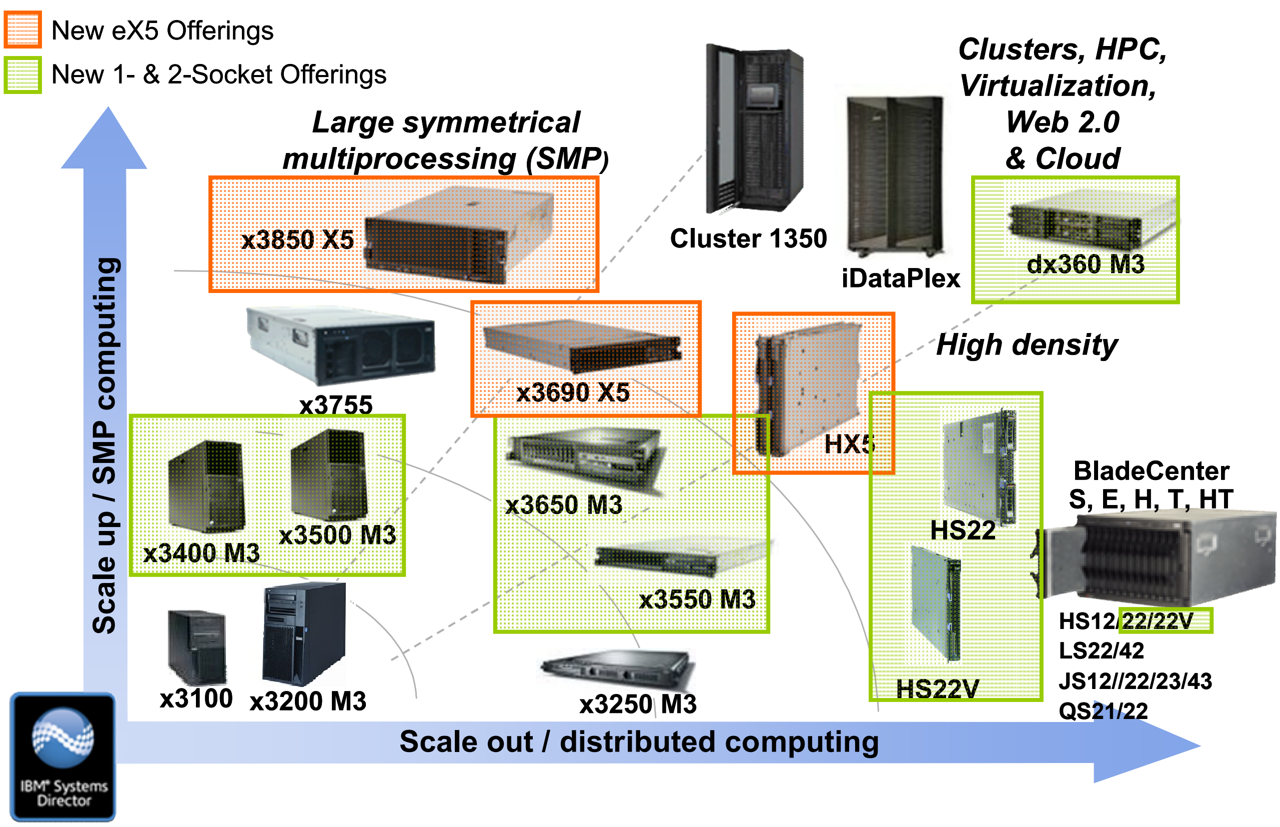


Figure 2: The IBM System x and BladeCenter Portfolio (www.ibm.com/systems/x)

[1] Swami Subramaniam, "Productivity and attrition: key challenges for biotech and pharma", Drug Discovery Today, Volume 8, Issue 12, 15 June 2003, Pages 513-515.

Faster results with the Intel Xeon processor 5600 (Westmere-EP) and 7500 (Nehalem-EX) series.

Both processor architectures are designed to enhance parallel processing performance for speedy HPC. Intel's QuickPath Interconnect (QPI) delivers substantial increase in bandwidth from a scalable shared memory by incorporating an integrated DDR3 memory controller onto the processor die. These processors can run two threads per core simultaneously with Intel Hyper-Threading technology which improves throughput for most HPC workloads. Multi-level shared cache reduces latency to frequently used data thereby improving performance and efficiency significantly.

Intel Turbo Boost Technology increases performance of both multi-threaded and single threaded workloads. This technology is activated when the Operating System requests the highest processor performance state. The maximum frequency of Intel Turbo Boost Technology is dependent on the number of active cores. While the length of time the processor spends in the Intel Turbo Boost Technology state depends on the workload and operating environment, it provides the performance a user needs, when and where the user or application needs it.

For a given workload, the number of active cores, estimated electricity consumption, or processor temperature can set the upper limit of Intel Turbo Boost technology. When the processor is operating below these limits and the user's workload demands additional performance, the processor frequency will dynamically increase by 133 MHz on short and regular intervals until the upper limit is met or the maximum possible upside for the number of active cores is reached. Conversely, when any of the limits are reached or exceeded, the processor frequency will automatically decrease by 133 MHz until the processor is again operating within its limits with the stated frequency as a lower bound. For example, a database query, which doesn't use any of the processor's floating point silicon, can take advantage of that thermal headroom and increase the frequency of all available cores.

<u>The Intel Xeon processor 5600 (Westmere-EP) series</u> provides an extremely balanced computing architecture that dramatically improves total application performance by automatically and intelligently adjusting performance according to application and environmental needs to deliver maximum performance when needed. Several performance technologies combine to increase HPC density delivering more performance per square meter than ever before. Based on 32nm technology, this series has up to six cores/socket, three memory channels, up to 12MB of last-level cache, and several power states.

<u>The Intel Xeon processor 7500 (Nehalem-EX) series</u> provides HPC clients with access to a "super node", delivering more integrated memory capacity and easily scaling up to 8 sockets per server. This series delivers the necessary compute, memory and memory bandwidth performance to solve large HPC problems faster. Based on 45nm technology, this series has up to eight cores/socket, four memory channels, up to 24MB of last-level cache, and several power states for extreme HPC performance for problems limited by memory capacity and memory bandwidth.

Early benchmarks on the newest Intel Xeon processor 7500 series for many HPC applications *record a mean of about two times improvement* for the "time to results" over the previous generation Intel architecture. Applications sensitive to memory capacity, memory bandwidth, and number of cores benefit the most. This is particularly beneficial for many production HPC workloads that have reliable and well-tested computational algorithms that could be difficult to implement on specialized accelerators such as Graphics Processing Units (GPUs).

The IBM eX5 servers: maximizing memory, enhancing reliability, and reducing storage costs.

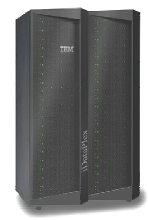
IBM recently introduced an entirely new portfolio of eX5-based servers that incorporate the latest Intel Xeon processor series. The new IBM server portfolio leverages a US$1 billion investment by IBM in enterprise X-Architecture design and the System x product line, as well as a decade of collaboration with Intel and other technology leaders. The eX5 server line is designed to maximize memory, minimize costs, speed processing, and simplify deployment for HPC clients. These eX5 servers not only deliver increased

HPC density with the new Intel processors, but also enable users to add a full drawer of extra memory, a feature available exclusively from IBM. In fact, so much processing performance and memory are available that many large-memory intensive HPC applications can run on just a few of the eX5 servers. The eX5 systems can drive twice the memory per four-socket server—up to 64 DIMMs standard and 96 DIMMs with MAX5 memory expansion. This opens the door to a streamlined HPC architecture for applications that can take advantage of non-clustered systems with extra memory capacity. The eX5 systems also include enhanced reliability with FlexNode - a feature which allows a single system to dynamically become two distinct systems or single back again. Integrated solid-state disks with eXFlash enhance I/O performance or reduce storage costs by up to 97% for the same performance.

The IBM System x iDataPlex: providing massive scale-out for HPC and clouds.

The System x iDataPlex internet-scale computing server solution from IBM is uniquely positioned to help enterprise clients overcome compute density and energy efficiency constraints. The iDataPlex supports massive scale-out data centers and high performance computing solutions.

The iDataPlex is a half-depth server solution, optimized both mechanically and component-wise for maximum efficiency with power and cooling. It is an industry-standard based server platform designed to minimize utilization of data center floor space (doubles the compute density/square foot compared to a 1U rack), and power and cooling infrastructure (20% less cooling and 63% less fan power). It is configurable for customer-specific compute, storage, or I/O needs and delivered pre-configured for rapid deployment. Customers have seen up to 40% energy efficiency improvements with four times the performance of a previous solution using iDataPlex for their HPC requirements.



The IBM BladeCenter: delivering tightly integrated energy-efficient dense packaging.

By integrating servers, storage and networking, IBM BladeCenter helps HPC clients sweep aside complexity. Its wide application solution blades, truly efficient chassis, and open design, are packed into an answer to today’s data center challenge of straggling racks and overheated server rooms.

IBM System x Cluster 1350: pre-integrated for rapid deployment and quicker payback.

IBM System x Cluster 1350 reduces deployment time for Linux clusters and Windows clusters, benefits from power and cooling through IBM Systems innovation, and offers integrated global hardware support. Clients can speed up installation of an HPC cluster, simplify its management and support, and reduce mean time to payback.

Innovative software environment: increasing HPC productivity.

Together, IBM and Intel drive toward simplicity in the software design to take advantage of system features that deliver high performance at consistent reliability and security. The programming and administration environment is based on familiar programming languages, libraries, job management tools, and parallel file systems. HPC applications developers greatly benefit from these innovative software components without facing a steep learning curve.

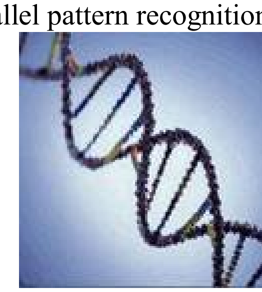
IBM System x clusters are highly optimized to reduce the deployment time for Linux clusters and Windows clusters. A Linux HPC Cluster is equipped with computing power at low-cost, collective intelligence of open standards, portability, flexibility and high availability. Likewise, the Windows HPC Server 2008 on IBM System x clusters and the BladeCenter servers provide reliable, cost-effective HPC cluster solutions. Windows HPC Server 2008 is designed to simplify the deployment, administration and management of the client's entire system with scalable performance, an easy-to-use scheduler and a new management interface that facilitates a familiar Windows environment so that supercomputing is more accessible.

Additional HPC cluster software available on IBM clusters includes the General Parallel File System (GPFS) and Tivoli Workload Scheduler LoadLeveler for Linux. GPFS is the top performing cluster-wide file system providing superior scalability and high reliability. Tivoli Workload Scheduler LoadLeveler is a job scheduler designed to maximize resource utilization and job throughput to get the most out of the available resources. IBM clusters also support other workload management solutions available from partners. These integrated software tools sustain highly productive environments with thousands of server nodes running large parallel HPC workloads.

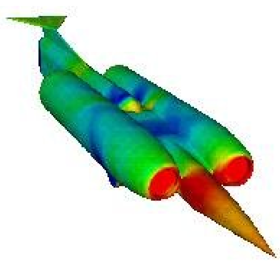
The IBM-Intel HPC Advantage: Delivering Value for Multiple Industries and Applications

Over the last decade, in conjunction with application developers, IBM and Intel have made substantial investments to optimize and tune HPC applications to take advantage of the IBM System x portfolio. Substantial performance improvements have been achieved by careful load balancing, maximizing single socket performance, maintaining data locality, minimizing cache misses, improving I/O performance, and maximizing the computation-to-communication ratio. Application performance can be further enhanced by the use of IBM developed high-performance I/O solutions like GPFS, and MIO-memory mapped I/O. All these investments translate readily to the new IBM System x family.

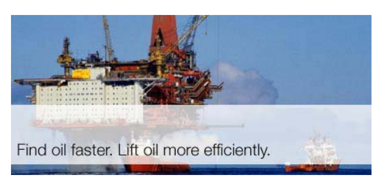
Life sciences researchers frequently use bioinformatics applications that use parallel pattern recognition, search, and data mining algorithms based on data-decomposition and distributed query techniques on scalable parallel systems. These along with imaging analysis - high content analysis, storage, and retrieval of voluminous and unstructured image data for both molecular imaging and diagnostic imaging - are some of the challenging applications that could perform very economically on the IBM System x portfolio. Researchers would be able to rapidly glean valuable insights that are crucial for drug discovery and development from a wide array of exploding and diverse data sources.



Engineering users engaged in Computational Fluid Dynamics (CFD) can economically leverage System x clusters for several, concurrent CFD simulations typical in parametric or design optimization studies. Additionally, these engineers would be able to solve larger and more complex interdisciplinary problems involving combustion, turbulence, and aircraft flutter. Likewise, in structural analysis, full vehicle dynamic response and crash analysis with occupant models are used in iterative design optimization studies. Crash analysis incorporates very sophisticated material models with very small explicit time steps to capture the rapidly changing physics. These simulations are very computationally intensive. Fortunately, most crash codes work well in parallel mode on cluster architectures with high-speed interconnects and should scale very economically on IBM clusters.



Upstream petroleum exploration and production problems often require large-scale and complex computations for multidimensional prestack time and depth seismic migration studies. Modern seismic surveys that are vaster and deeper and basin-wide reservoir optimization problems with multiphysics could be very suitable for the enhanced IBM System x portfolio.



Financial Services applications for risk management, complex asset valuation, Monte Carlo simulations, fraud detection and anti-money laundering are some of the long-running calculations that could benefit from the economical scale-out of IBM clusters. For ultra low-latency financial trading execution environments, eX5 systems such as the x3850 X5 have enough processor cores, adequate main memory, and substantial internal storage and IO bandwidth to support all major US market level 1 direct feeds in a single box. The data center crisis is most acute in this sector. The unique energy and space-efficient data center value proposition of the IBM System x portfolio should resonate well with large investment banks and insurance companies as they struggle to compete globally.

Some emerging workloads are being considered for GPUs. Today, most GPUs are being used to accelerate specific niche computation-bound kernels such as LINPACK, MATLAB, reverse time migration for seismic analysis, certain financial analysis algorithms, and bioinformatics. We expect, in the future, that some of the more traditional HPC ISV applications could become available in production on a GPU based system, as these ISVs make the investments needed to develop and support a GPU-optimized code base.

HPC has always pushed the envelope in performance as measured by peak flops. Today, GPUs are theoretically capable of delivering the most peak flops per processor and also the most flops/watt. Some clients at research centers, the energy sector, and financial services are using GPU capability integrated with an x86 platform for some workloads that require an order of magnitude boost in performance. And many IT vendors are providing GPU capability integrated with x86 servers for HPC.

Today, clients must invest additional resources and tooling to enable, tune, and optimize their workloads to reach a substantial fraction of this peak using custom codes. Most clients would probably already have x86 systems. It may be prudent for them to extract additional performance and value from current x86 systems while making the resource investments on GPUs. As some customers begin to use GPUs in production for some HPC workloads, they would probably still need an x86 system for other workloads, and managing and supporting these hybrid environments could increase people costs and complexity as the GPU application development environment matures. But the benefits could justify these costs for certain workloads.

The Future of Faster HPC: Still Sustained, Steadfast, and Smart

While Intel based x86 systems are the incumbent leader in HPC with a large install base, HPC users traditionally have low switching costs and vendor loyalty compared with the rest of the IT industry. Yet, we believe x86 and the IBM-Intel leadership will continue for the foreseeable future for the following reasons:

1. As the use of HPC continues to mainstream across many industries, customers will demand well-tested, reliable, integrated application solutions that deliver performance and value across a range of workloads.
2. The ISV ecosystem for x86 clusters is vast as a result of *sustained* and *steadfast* investments by Intel and IBM over the last decade. Others, including systems with GPUs, will have to make similar investments. This takes time, especially for the broad class of mainstream HPC workloads.
3. Powered with the latest generation of Intel Xeon processor 5600 series and the Intel Xeon processor 7500 series, IBM has significantly enhanced its System x portfolio for HPC.
4. Differentiated with *smart* innovations in packaging, power and cooling, and reliability, these economical and faster HPC solutions can substantially increase computing density and memory in a standard enterprise rack or the IBM BladeCenter, or scale-out to 100s of energy-efficient racks with the iDataPlex.
5. Lastly, the eX5 servers outflank the competition and are designed to maximize memory, minimize costs, speed processing, enhance reliability, and simplify deployment for HPC clients.