premio

Intro to Intel Scalable System Framework

Premio Inc

intel IoT Solutions Alliance

We make it Possible

Intel Scalable Framework

Businesses exist to solve problems. That is a fundamental truth organizations and professionals around the world over understand. As we move further into the 21st century, the problems businesses are tasked solving with becoming more and more complex and the need to leverage the processing power of "supercomputers" to analyze the mountains of data involved is becoming more and more common. This makes processing power a vital part of helping a modern business achieve their primary goal. The supercomputers analyzing the data are often actually a cluster of servers run in parallel performing what is commonly referred to as high-performance computing. Under the hood, high-performance computing requires much more than just processing power. While high-performance computing is helping solve real-world problems today, it is also plagued by some fundamental problems of its own.

Many high-performance computing systems are imbalanced and plagued by storage, memory, and network bottlenecks that prevent organizations from getting the most out of their supercomputer level processing power. Additionally, a lack of standardization in the world of high-performance computing has led to less than optional applications, barriers to entry, and highly disjointed deployments where one organization may need to implement multiple systems that cannot integrate and communicate efficiently.

Intel has begun an ambitious effort to bring standardization and innovation to the world of high-performance computing with their Scalable System Framework (SSF). Intel's SSF looks to serve as an innovation enabling framework that will help make high-performance computing more accessible and scalable.

In this paper, we will review the topic of high-performance computing, current challenges facing the high-performance computing market, what Intel SSF is and how it addresses the current challenges, what that means for your business, and how Premio can help you architect your next (or first) high-performance computing system.

Intro to HPC (High-Performance Computing)

How prevalent is high-performance computing?

Before we dive into how Intel is changing the face of high-performance computing with their Scalable System Framework, we will discuss exactly what defines high-performance computing and why it matters. Long considered to be the domain of scientists, academics, and government officials almost exclusively, that is no longer the case. As evidenced by insights from trusted industry insiders like Ed Turkel, HPC is becoming more and more prevalent in commercial applications in fields such as “product design, financial analytics, personalized medicine, marketing automation, fraud detection and autonomous vehicles” (Source: CIO). To help give an idea of how common high-performance computing is becoming, the aforementioned CIO article calls out a telling statistic: HPC solutions already account for a quarter of annual server sales. As the Internet of Things creates more and more data points, businesses grow their data warehouses, and competition drives the need for better, faster analytics increases, the demand for high-performance computing will scale even higher. This means it is imperative that businesses understand HPC and how it can keep them competitive as we move into the next generation of computing and analytics

What are the nuts and bolts of high-performance computing?

With an understanding of the growth and importance of HPC, the next question we need to answer is: what exactly quantifies high-performance computing? If you are looking for a definition that pinpoints a given number the ability to perform “1012floating-point operations per second” has been used as a reference (source: TechTarget). However, as Moore’s Law makes most specific benchmarks age quickly, we believe it is more important to understand how HPC works as this helps conceptualize not only the nuts and bolts of high-performance computing but also sets the stage for understanding the current problems facing the HPC community.

At the core of HPC functionality is the concept of parallelism. Parallelism is the use of multiple of the same resource in parallel to complete a set of tasks, as opposed to completing one task after another in series (also known as completing the tasks serially). The use of parallelism that most are probably familiar with is multiple processor cores in a CPU. Having a lower clock speed but more cores and getting better performance is something most IT professionals can understand, so that is a really good place to begin understanding the power of HPC. High-performance computing takes this concept to the next level by allowing not only for processors to operate in parallel but entire servers to operate in parallel in as nodes in a larger high-performance computing system. This approach is what makes high-performance computing systems so powerful, enabling them to solve complex and highly analytical problems exponentially faster than a single system working serially

From an operating system standpoint, at the highest levels, high-performance computing is dominated by *nix operating systems. The current list of the top 500 performing "supercomputers" includes 498 Linux based systems and 2 Unix based systems (Source: Top500). That being said, "Windows shops" are not left out in the cold when it comes to high-performance computing. Microsoft made a foray into the HPC market with their Windows HPC Server 2008 operating system. While that project did not lead to Windows making much headway in the top levels of supercomputing, it wasn't the last of the software giant's HPC focused efforts. At current, it seems Microsoft's focus on HPC is in the cloud through Azure and with HPC Packs that can be run on top of Windows Server 2012 and Windows Server 2016 (for more on HPC Packs, check out this TechNet article).

Beyond the hardware and operating system, what really makes high-performance computing useful in the real world are the applications that produce the output that helps solve problems. One of the bigger challenges of high-performance computing is the difficulty of writing code that maximizes the benefits of parallelism. Coding an application for a single computer and coding an application for a parallel, high-performance computing system are two entirely different paradigms. Some of the popular languages used in high-performance computing are FORTRAN, C and C++, and Julia. FORTRAN has a significant foothold in the HPC community as it was the language of choice in the early days of HPC. C and C++ are known for their efficiency and performance, and are popular for other applications, making them a logical choice for developers working on HPC projects. Julia is an open source programming language that was built with parallelism in mind and is described as “high-level, high-performance dynamic programming language for numerical computing” (Source: Julialang.org). What is unique about Julia is that the language is written with the intent of being scalable and easy to understand and learn, making it more accessible and reducing the human capital required to get a project off the ground.

What Are The Current Problems With HPC?

HPC is at the cutting edge of technology in a world that is heavily technology driven, so at the surface, it can be hard to understand how it could be improved beyond waiting for iterations and Moore’s Law to have their incremental impacts in this area of tech like they do in others. However, drill down a layer deeper, you begin to understand that HPC is a number of discrete components in a larger system and these components have evolved at different rates in different ways over time. This has created a number of inefficiencies, inconsistencies, and bottlenecks that Intel’s Scalable System Framework strives to solve. Here we will walk through some of the major problems and challenges facing the high-performance computing community before we discuss how Intel’s Scalable System Framework addresses these problems and challenges.

The Walls and How They Impact Performance In HPC Environments

In their brief titled “Holistic Solution for HPC Infrastructure Needs: Intel® SSF” (Source: Intel) Intel calls out “the walls”, a number of system bottlenecks as some of the key hurdles holding back HPC. Anyone comfortable with the idea that the CPU is almost never the source of a bottleneck in an application can relate to this. Memory, input/output (I/O), network fabric, code that is not optimized, and more can all slow down performance leaving applications powered by even the best CPUs lagging. To help conceptualize the idea of a “wall” as it related to HPC, will touch on a few instances where different components can create a bottleneck and become a wall

While RAM speeds are ever increasing and with DDR5 coming next year, they still cannot keep up with the lightning-fast speeds of multiple processors running in parallel. This can lead to situations where a processor is waiting for memory to catch up, creating a “memory wall”. Similarly, input/output operations per second (IOPS) on hard drive disks (HDDs) and solid-state drives (SSDs) will often lag behind the speeds possible, creating another potential bottleneck. Even if memory, storage, and the processors are firing on all cylinders, network bottlenecks can slow down HPC applications if speeds are not up to par or throughput and bandwidth is insufficient.

Divergent Workloads and Diseconomies of Scope

Another problem significantly impacting the progress of high-performance computing is the challenge of divergent workloads. As mentioned earlier in this article, there are a number of use cases for high-performance computing and the use of high-performance computing is on the rise. Intel makes the point that a number of organizations have begun utilizing discrete systems for applications such as big data, machine learning, and visualization.

What this means is the sharing of data becomes more difficult and resources become stretched across multiple different systems, making solutions less efficient than they could be. A high-level analogy for what is occurring that many IT professionals can relate to is: what is being done with HPC systems is akin to using multiple spreadsheets and Access databases instead of a centralized database and ERP system for resource management. Organizations would be much better served to concentrate their efforts on one holistic solution that solves their HPC problems, but the market has lacked the standardization to make this a reality for some time. This leads to solutions being more expensive and less efficient than they could be if a less disjointed alternative was available.

Complexity as a Barrier

Another challenge Intel discusses that limits the potential of HPC systems is the complexity of the data within a system. There are a myriad of data sources and data can be inconsistent and lack annotation, which sets technical barriers on who can leverage the power of an HPC. To unlock the true potential of HPCs organizations need to be able to make them accessible to professionals who may have a background other than computer science. By democratizing access to HPCs, we can solve more problems and see the sort of exponential advancements we saw as Internet connectivity became ubiquitous. Eliminating the complexity barrier will go a long way in taking HPCs to the next level.

How does Intel's Scalable System Framework Solve The Problems With HPC?

Historically, problems in information technology are best solved by a solution that is based on standardization and scalability. APIs and TCP/IP are great examples of standardization shifting a paradigm and unlocking the technological power that enables businesses and empowers large user bases to become more innovative and productive. If we can solve the challenges facing high-performance computing, we can see similar advancements over the next decade.

As demonstrated by the aforementioned problems facing the high-performance computing community, the current HPC paradigm is one that is imbalanced and lacking holistic, standardized solutions. With their Scalable System Framework, Intel is looking to create a holistic, balanced, scalable framework that can take high-performance computing, and in turn business and science, to the next level. Here, we will dive into some of the key components of Intel's Scalable System Framework and discuss how they can help solve the HPC problems that currently exist

What is Intel's Scalable System Framework?

According to the Intel Scalable System Architecture Specification (Source: Intel), the Intel Scalable System Framework is "is a collection of system building blocks, known as 'Intel SSF elements', and supporting reference architectures". Based on just that description, one can begin to understand how this ambitious framework can serve as the cornerstone for solving many of the problems with HPC systems today. The current paradigm is one that is disjointed and imbalanced, and standardization can go a long way in addressing many of the underlying issues. By taking a balanced, holistic approach to architecting HPC solutions, we can make them more efficient and accessible.

Breaking Through The Walls

How does Intel address the aforementioned bottlenecks referred to as “the walls”? By defining standards and baselines that mitigate them at the beginning of an implementation and leveraging high speed, high-performance technologies throughout all aspects of an HPC system. This approach ensures that the rest of an HPC application is able to keep up with the processing power of today’s high octane processors.

Referencing the aforementioned Intel brief, some of the specific hardware that can be leveraged within the Intel’s Scalable System Framework include: high performance processors like the Intel Xeon E5-2600 v4 line or Intel Xeon Phi line to ensure processing speed is at HPC levels, Intel Optane Technology-based SSDs for NVMe (Non-Volatile Memory Express) to minimize storage bottlenecks, Intel Omni-Path Fabric, and 10/40Gb Ethernet networking gateways to break through network “walls”

Balance and Standardization Across the Board

Intel’s Scalable System Framework will have standards in place for operating systems, the Linux Kernel, access controls, programming interfaces, runtime environments, storage, file systems, and more. Additionally, Intel’s Scalable System Framework is compatible with common HPC programming models and can leverage the benefits of Intel Architecture (IA) based code. This allows for compatibility in areas that may have been previously disjointed and helps solve the diseconomies of scope problem.

To dive a little deeper on items standardized in Intel's SSF, they include:

- The Linux Standard Base (LSB) command system used
- The ANSI Standard C/C++ runtimes
- The minimum amount of RAM on nodes
- The minimum amount of accessible storage on nodes
- The Python, Perl, and Tcl scripting languages used
- The architecture of the Intel processor used
- The type of cluster file systems used

Simply by skimming through that abbreviated list, one can begin to understand how being able to trust that all those components can be known can enable HPC application developers to build more scalable, dependable, business-enabling applications and further drive innovation.

Leverage a Common Infrastructure and Broadly Available Hardware and Software Components

Intel's Scalable System Framework allows high-performance computing systems within an organization to leverage a common, purpose-built infrastructure. This holistic design is more balanced and efficient than the current paradigm, further addressing the diseconomies of scope problem. Additionally, the hardware and software components that make up Intel's Scalable System Framework is widely available and more familiar to users, not only allowing for a competitive, innovation driving environment but also helping to solve the complexity barrier problem mentioned earlier.

What Does This All Mean For Your Business?

While the ability of the Intel's Scalable System Framework to address these high-level HPC problems may be impressive or at least ambitious, does it mean anything to an IT professional in the average enterprise? Yes, it means that you can not only implement a high-performance computing solution to solve complex business challenges, but you can do so in a way that is cost-effective, scalable, and accessible to less technical users.

There are a few specific benefits Intel calls out in their brief for users at different places on the high-performance computing vertical, and we will summarize some of them below as well as add some additional perspective

- **OEMs and ODMs**- Original equipment manufacturers (OEMs) and original design manufacturers (ODMs) will see costs and risks driven down as pre-validated hardware solutions are made available. This will allow manufacturers to focus more intensely on their areas of specialty and drive innovation.
- **Software vendors**- Similar to the benefits of standard APIs and the innovation that has been made possible through technologies like REST, Intel's SSF will reduce costs for software vendors and increase innovation by providing a single, authoritative standard by which vendors can build around.
- **End users**- Much like what the TCP/IP stack did for networking and telecommunications, Intel's SSF can do for high-performance computing. By creating an environment where vendors build to an inter-operable specification and leverage standard technologies, SSF will help make HPC more accessible to more end users and allows for more balanced, modular, scalable, extensible solutions to be architected. In short, standardization can help end users better leverage HPC to solve problems.

In a world where analytics, machine learning, and big data are becoming more and more important to more and more industries, a scalable and affordable HPC solution can be the difference between being on the cutting edge and being the one struggling to keep up.

How Does Premio Fit In?

At the core of Intel's Scalable System Framework are various "nodes" that carry out the compute functionality in a high-performance computing system. Moving from concept to practice, these "nodes" will often be servers that need to contain high-performance, robust, enterprise-grade components, and as an innovator and market leader in server and storage solutions, Premio is uniquely equipped to provide the balanced, robust server technologies that will enable Intel's Scalable System Framework.

For example, some of our ScaleStreams High-Density Storage Servers support the aforementioned Intel Xeon E5-2600 v4 processors that are called out in Intel's brief on their Scalable System Framework as well as dual 10Gb Ethernet adapters to allow for high-speed network connectivity and avoid bottlenecks in your HPC system. Alternatively, for applications with even greater performance demands, the FlacheSAN1N4C-D4 offers support for Intel Xeon E5-2600 v4 processors, NVMe storage options, and Dual Gb Intel i210 Optional 56Gb FDR Infiniband QSFP+ or Dual 40Gb Ethernet adapters.

In addition to a wide variety of HPC ready servers that can fill multiple roles within a HPC system (beyond storage, we also have options for servers that can run as nodes that enable the compute portion of an HPC solution), we also have extensive experience and expertise when it comes to architecting balanced solutions that mitigate bottlenecks, break through "walls", and help you achieve the optimized speed and performance for your investment. Contact us today to learn how Premio can help optimize your data center and enable your business.

https://premioinc.com/