Hewlett Packard Enterprise

Leading the charge to exascale computing

New thinking and enabling technologies

Contents

Reference guide

The ability to ingest, process, analyze, and simulate vast quantities of information has become forever linked to innovation and technological advancement in this data-driven age. As a result, nations and enterprises around the world are ramping up their investments in high performance computing (HPC) systems in order to maintain competitiveness, drive new discoveries, and solve grand challenges. Particularly for the most data-intensive and mission-critical applications, many HPC technologies are rapidly approaching the end of their usable lifecycles, and there is a realization that the next iteration of advanced computing systems is required to sustain new breakthroughs. This race to the next generation of computing is resulting in an exciting period of fast innovation for the computing community, as it enters the final stretch of research and development (R&D) before the first fully operational exascale system is introduced.

The U.S. Department of Energy (DOE) is spearheading a major R&D effort to address the significant challenges of building an HPC system that is an order of magnitude more powerful than today's largest systems. It is not just about peak performance, but, more importantly, productivity. The DOE has defined a capable exascale system as a supercomputer that can solve science problems 50X faster (or more complex) than today's systems. Floating point operations per second (FLOPS) represent the basic unit of computation, and the fastest DOE production systems today perform approximately $20x10^{15}$ FLOPS (or PetaFLOPS). An exascale machine would then perform above 10^{18} FLOPS, that is, an ExaFLOPS. The DOE has also outlined a delivery timeline for a fully operational exascale system early in the next decade, and certainly within five years.

Exascale computing will undoubtedly be a significant milestone for the scientific, academic, and industrial communities. Access to such unprecedented levels of computational power will enable scientists and researchers to begin solving the grand challenges standing in the way of fundamental advances in a variety of data-intensive areas. The industry is on the cusp of an innovation plateau, and achieving the next significant breakthrough in genomics, cancer research, climate change, cosmology, national security, energy, and more will require exascale computing applications and resources.

However, the enterprise computing community also stands to benefit from these advancements. The technologies that will need to be developed to deliver an exascale machine will have far-reaching implications in smaller-scale deployments and new computationally intensive applications, such as artificial intelligence and deep learning. While not everyone needs or can afford an exascale machine, its petascale building blocks have far-reaching potential for much broader adoption. As efforts are made to drastically lower the energy of computation and pack greater amounts of compute power into smaller form factors, these advancements will have a dramatic impact for not only the scientific elite, but for the mainstream computing industry as a whole.

Maintaining system balance

An evolutionary step of this magnitude presents new system design challenges for the entire HPC ecosystem. Existing HPC systems cannot be simply scaled up to an exascale level without changes, so the development of new computing technologies and a complete rethinking of next-generation computer architecture design will need to come into play. Without question, the chief challenge of exascale is finding ways to minimize the massive energy consumption of data movement and computation at this scale. The memory subsystem, system reliability and recovery, and programming environments will also require major transformations in order to deliver a usable exascale system within the next five years.

Maintaining system balance remains a primary concern of exascale, because balance is key to building a productive system that can deliver beyond peak performance. There are four metrics that generally indicate a system is in balance, all relative to peak FLOPS performance—memory capacity, memory bandwidth, network injection bandwidth, and network bisection bandwidth. These metrics fundamentally drive the way a machine is programmed, and any significant skew can create unexpected bottlenecks and throw the entire machine out of balance. An unbalanced machine ultimately leads to immense pressure on the programmers, and requires heroic programming efforts that prevent exploiting the full computing potential. While achieving peak FLOPS performance is important, next-generation HPC architectures must be designed to also provide a breakthrough in programmer productivity so that it is easier to write effective applications.

The challenge is that these four metrics are increasingly struggling to keep pace with the rapid rise in FLOPS. Even in some of the largest supercomputers in use today, it is apparent that the ratios for memory capacity, memory bandwidth, and interconnect bandwidth vs. FLOPS are all headed in a downward trend, which ultimately results in operating inefficiencies for the user. For example, in a system with too much compute power but not enough memory capacity, the user will be forced to page in and out the amount of data that they need to work with. Up to this point, the computing industry hasn't been able to build an HPC system with memory that is big and fast at the same time, and this is one of the key obstacles of exascale computing.

Achieving technical and commercial usability

Another reality of exascale is that all of this compute capability will mean nothing if we cannot power the machine. With an unlimited supply of energy, a fifty-fold improvement in speed compared to today's supercomputers would be achievable; however, it is extremely challenging to provide more than 20–30 MW of electrical power to a single supercomputer, and that requires a 30X improvement in energy efficiency. Finally, a system of this scale will experience continuous component failures simply because of statistics. Circumventing the many failures and inefficiencies that could occur at any given time in a system of this magnitude remains one of the overarching goals of exascale computing.

Programming an exascale machine is not going to be for the faint of heart. Today's largest supercomputing systems already require million-way concurrency, and exaflop machines are expected to have billion-way concurrency. From a programming standpoint, increasing the parallelism by a factor of a thousand puts enormous pressures on environments, languages, runtimes, and algorithms. Even the most insignificant sequential component is bound to have a dramatic impact on the utilization of an exascale machine, and today's coders are struggling to write programs that focus on eliminating these sequential components.

As we hinted before, resiliency and failure recovery in an exascale machine is also a major issue. As the number of computing components increase one hundred-fold, it's inevitable that the system will begin to experience more frequent failures. Exascale machines, which are expected to have about 100,000 processors, will be likely to experience a hardware failure every few hours. Two things need to happen. First, enough redundancy has to be architected so that not all the failures translate to application interruptions. Then, in order to tolerate the applications-visible failures and productively continue computing, the machine must be able to save and restore its entire state very rapidly, almost on an hourly basis. When the likelihood of failure becomes so imminent that a system spends more time checkpointing than actually computing, the usability of that system declines to an unacceptable level.

Enabling technologies of exascale

The difficult truth to accept is that there is no "silver bullet" to exascale, and the only way to build a system of this magnitude is to meticulously optimize and balance every single component. Many of the HPC technologies in widespread use today have limited applicability to exascale in the sense that they would quickly exceed the DOE's power envelope of 20–30 MW, or become ineffective from a usability standpoint. Research groups inside Hewlett Packard Enterprise have been hard at work developing and prototyping enabling technologies of exascale for over a decade, and these technologies promise to deliver new, more efficient ways to compute at an unprecedented level.

On the networking side, there needs to be a determined push toward more energy-efficient data communication technologies, such as silicon photonics. As higher amounts of bandwidth are shifted closer to the compute node (to preserve the system balances we have already discussed), driving this level of bandwidth through electrical interconnects quickly becomes cost-prohibitive and inefficient. Silicon photonics, which uses photons (light), and not electrons, for data transmission, has emerged as the most promising option for reaching the desired target of five picojoules per bit, which would be necessary for a balanced exascale system's fabric within the 20–30 MW range. Hewlett Packard Enterprise has significantly invested in several optical technology R&D programs that are geared to driving a much-needed breakthrough in this area, and making silicon photonics a reality for the next generation of computing systems.

On the memory front, non-volatile memory (NVM) will need to be used in a far more aggressive way than it is being used today, primarily for the purpose of checkpointing and caching the first tier of persistent data for exascale programs. As previously mentioned, a key challenge of exascale is increasing the speed at which the machine's entire state can be saved and restored. A balanced exascale machine will have between 10 and 20 petabytes of DRAM, which needs to be continuously saved and restored to fast storage. And, to keep multiple checkpoints and cached data, this translates to 50 and 100 petabytes of fast storage. There are a number of NVM technologies that can be used in this context to achieve the speed that's necessary to reach this level of bandwidth performance. In addition, the industry will need to begin using co-packaged memory on a more pervasive level. Many of today's many-core processors and high-end GPUs have begun to pull the high-bandwidth memory tier into the compute package. This co-packaging of the fastest memory tier with the processor package offers the opportunity to boost memory performance and achieve the bandwidth necessary to move data at an exascale level.

To address these challenges, Hewlett Packard Enterprise is embracing a memory-driven computing vision and has prototyped it in The Machine first test vehicle system developed by Hewlett Packard Labs in 2016. This vision for the future of computing puts memory—not processing—at the center of the system architecture. Memory-driven computing combines massive pools of NVM, fast memory fabric, and task-specific processing to build a machine where everything looks, behaves, and is accessed like memory. This approach to computing allows the system to provide an access mechanism that is extremely lightweight and tightly coupled with simple memory semantics (load-and-store) operations directly driven by the processor instruction set.

In a memory-driven computing model, open protocols such as Gen-Z become extremely important. This new scalable computing interconnect and protocol will provide a path to easily access large volumes of data at lower costs and faster speeds than today's interconnects. Protocols such as Gen-Z move memory semantics operations to the large-scale fabric to enable data access across the entire system, providing communication at the speed of memory: high bandwidth, low latency, lower costs, and reduced software overheads. By tightly coupling memory and communication through protocols such as Gen-Z, everything begins to act like memory and there is a possibility to move greater levels of bandwidth closer to the processor.

Why Hewlett Packard Enterprise?

Hewlett Packard Enterprise's approach to exascale is geared to taking a full system view, which allows us to incorporate the industry's best-of-breed components and new technologies as they mature. This open-standards view helps avoid lock-in to any one specific proprietary technology and lets us selectively integrate the most promising HPC technologies from a variety of partners in the HPC ecosystem.

Many of the foundational technologies for exascale have been in the works in our labs for well over a decade, and we have built a solid foundation of knowledge as well as strong partner network around these technologies. In addition to understanding the strengths and limitations of these new technologies, Hewlett Packard Enterprise's leadership in enterprise computing also provides the means to build a healthy supply chain around them. The combined strength of our many years driving new technological advances along with a long track record of success in building high-end systems provide a unique perspective into solving some of exascale's largest challenges.

Hewlett Packard Enterprise also continues to make significant investments in the future of HPC. The recent closing of our acquisition of SGI, a global leader in high performance solutions for compute, data analytics, and data management, combines SGI's strengths in HPC with Hewlett Packard Enterprise's global reach in the commercial computing markets. This merger will help us extend a leadership position in the global HPC market, and expand our presence and participation into the TOP500 supercomputing sites in the world.

Conclusion

Explosive data growth is fundamentally changing the way computers must operate, and a paradigm shift is now needed in order to solve the grand challenges of today and tomorrow. Exascale computing promises to help drive the breakthroughs that are so desperately needed in the realms of science, business, and engineering. Hewlett Packard Enterprise is leading the charge to exascale by completely rethinking the way computer architectures are built, and developing new foundational technologies that will help make the next generation of computing a not-so-distant reality.

Learn more at

labs.hpe.com/research/themachine

Sign up for updates

Hewlett Packard Enterprise

a00006294ENW, May 2017