Hewlett Packard Enterprise

Gen-Z: A new approach to data access

Memory at the speed of innovation

Contents

Technical white paper

Today's enterprises, academic institutions, research facilities, and government entities are tasked with transforming unprecedented volumes of data into actionable business insight. In order for these insights to be of the utmost value, they must be delivered in real time and incorporate data spanning a variety of formats and from a growing number of sources. Challenged by explosive data growth and the increasing need for data-driven insight to accelerate innovation, customers are seeking reliable, functional, and secure computing solutions capable of handling next-generation workloads and applications. These requirements are fundamentally changing the way computing systems access and support data, as well as demanding a complete rethinking of the supporting architecture on which these systems have been built for decades.

Until now, computing systems have been designed around the notion that storage is slow, persistent, and reliable, while data in memory is fast but volatile. However, the emergence of new storage class memory technologies are converging storage and memory attributes, and conventional programmatic and architectural assumptions are no longer optimal to support the next generation of high performance computing (HPC) systems. For example, most traditional interconnects are placed on input/output (I/O) devices like network interface cards (NICs) or host controller adapters (HCAs) that sit on the PCIe bus and are not tightly coupled to applications. These interconnects are struggling to help I/O devices access memory as rapidly as possible, and therefore will be unable to provide the levels of injection bandwidth required to keep future supercomputers balanced. In order to reach the next frontier of HPC, the industry must turn conventional wisdom on its head and work to build a substantially faster network capable of supporting even more data and emerging workloads and technologies.

A few paradigm shifts are driving the industry to seek a new approach to data access. IDC predicts that 37 percent of the 40 zettabytes of data in the digital universe by 2020 will be useful from a Big Data analytics perspective, meaning that the amount of data that requires real-time analysis is expanding exponentially. At the same time, memory capacity and bandwidth are struggling to keep pace with the increase in floating point operations per second (flop/s), sending the trends for memory capacity, memory bandwidth, and interconnect bandwidth vs. flop/s on a downward trajectory even in some of the world's largest and most powerful systems. To resolve these challenges, the industry will require an open architecture and a high-bandwidth, low-latency fabric solution that can seamlessly integrate with existing ecosystems as well as enable a new generation of exaflop systems.

A better approach to data access

Gen-Z is a new data access technology designed to help the computing industry overcome the challenges of existing computer architectures and provide an optimized fabric solution to support both current and future systems. Gen-Z not only enhances memory and data storage solutions, but also provides a framework for both optimized and traditional messaging solutions. Capable of supporting a wide range of new storage-class memory media, hybrid and data-centric computing technologies, memory-centric solution architectures, and a wide range of applications, Gen-Z is a performance-optimized solution stack that is open, efficient, and cost-effective.

In current computing systems, each component is connected via a different type of interconnect. Gen-Z uses a memory-semantic (read and write) protocol that enables multiple types of devices to communicate efficiently using a straightforward and familiar scheme, thus offering a universal interconnect over which all components can communicate. The main goal of Gen-Z is to democratize and dramatically simplify data access, and offer a way for devices to communicate with their own local memory using simple commands. This is the reason why Gen-Z has been dubbed a "memory semantic protocol"—because it allows vast numbers of components to be connected using the same fabric, allowing each to gain equal access to the data.

Gen-Z offers a number of key benefits that will deliver compelling customer value from both technical and business perspectives:

- **High bandwidth**—Support for a wide range of signaling rates and link widths means Gen-Z will be able to support both today's data rates and future optical bandwidths, enabling solutions to scale from tens to several hundred GB/s of bandwidth.
- **Low latency**—A lightweight software interface performs simple memory reads and writes directly to the hardware, which significantly improves performance at the application level and enables sub-100 ns load-to-use memory latency.
- **Support for advanced workloads and technologies**—Gen-Z enables hybrid and data-centric computing, supports real-time analytics, and provides scalable memory pools for in-memory applications.
- **Robust security**—End-to-end secure connectivity is provided through a combination of hardware-enforced isolation techniques and full packet authentication to help protect against cyber threats and destructive intrusions.
- **Mechanical compatibility**—Gen-Z uses a wide range of existing mechanical form factors and cables, so it can be easily incorporated into any solution, from current applications and hardware platforms to HPC architectures of the future.

- **Economical**—By requiring all solutions to support simple read and write operations, Gen-Z simplifies the hardware interface and allows industry-standard components and cables to be reused as computing systems continue to scale, resulting in lower development and support costs and increased solution agility.

As the technology behind Gen-Z evolves, so do its implications for both the achievement of future computing systems and for the entire HPC community. Gen-Z will enable customers to rapidly innovate and deploy new capabilities and services whenever they are ready, without waiting for the industry to move in lock-step. For example, Gen-Z abstracts memory media from the memory controller so a wide range of new media types or various generations of media can be deployed and transparently supported in any solution. A highly efficient and flexible protocol that simplifies every operation to a basic read/write level also reduces complexity in hardware and software designs, streamlines deployment, and reduces development and support costs. And Gen-Z's interoperability enables the architecture to simultaneously and efficiently transport standard and customized communications between components, allowing customers to grow and expand their clusters as they are ready.

Driving the industry's innovation engine

Many IT technologies are governed by standards that the industry finds useful; however, the HPC community is increasingly adopting "open" standards that aren't tied to a specific proprietary architecture or controlled by a particular vendor/group. Open standards benefit the entire HPC community in a number of ways. First, they help foster a more open, competitive ecosystem where every industry player can contribute their own differentiated value for the betterment of the entire HPC industry. In an open ecosystem, companies aren't limited to falling into lock-step with a dictated pace of evolution; instead, every partner is free to accelerate the cadence of their own innovation, and contribute that value whenever they feel it's fully matured. Open standards also help new technologies achieve widespread adoption across the entire HPC industry, broadening the business opportunity for participating companies and decreasing the odds that a new technology will fall into a specific vertical or niche.

Gen-Z is being designed as a fully open industry standard, enhancing the ability for everyone to have a voice in the future of interconnects. In an effort to promote widespread adoption and distribution, the Gen-Z specifications are already freely available, and there will be no restraints on who may use or implement the standard. A core goal of Gen-Z is to design a standard that can be freely and easily accessed by the entire computing community, helping to promote widespread adoption and a collaborative environment where all industry players can work together to improve upon the standard as it matures.

Establishing the Gen-Z Consortium was a critical step in fostering a truly open ecosystem where customers and technology leaders can directly engage with one another to create these new specifications. Currently comprised of 34 leading technology companies, the Gen-Z Consortium is dedicated to creating and commercializing the Gen-Z technology as well as delivering a suite of specifications and IP enablement technology that will enable Gen-Z to be integrated into any solution. The Gen-Z Consortium was developed as an open, non-proprietary standards body that could foster an atmosphere where adoption, differentiation, and innovation are promoted as the new industry standards.

The Gen-Z Consortium is a not-for-profit organization that allows any interested member to join. Twelve member companies currently hold Board of Director seats and provide the overall governance for the organization; however, any member of the consortium is free to participate in any workgroup to develop Gen-Z specifications and other materials. Together, these industry leaders are helping to develop a memory-semantic interconnect that will not only bridge existing solutions, but also drive innovation all the way to the future generation of HPC systems.

A dedication to the open standards community

As one of the founding members of the Gen-Z Consortium, Hewlett Packard Enterprise (HPE) is committed to supporting and strengthening open standards projects and working collaboratively with the communities that develop them. HPE is helping to foster healthy open standards communities by joining forces with other industry leaders in the Gen-Z Consortium, whose members span virtually every sector of HPC hardware and software. HPE is a strong believer in the notion that open communities will almost always triumph over closed ones, and have adopted an enterprise-wide open methodology of diversity, teamwork, and collaboration. By tapping into the innovation cycle of the Gen-Z Consortium, HPE is part of a critical mass that will help the entire HPC industry not only develop a new data access interconnect for today and tomorrow, but also simply build better products.

The Gen-Z technology is expected to help set the stage for true memory-centric computing, which is a key tenet of HPE's The Machine. Gen-Z is a critical enabler of the Memory-Driven Computing architecture on which The Machine is built, and HPE is using technologies very similar to Gen-Z in the fabric of The Machine as well as our other high-level HPC platforms that are currently in development.

Conclusion

As data expansion continues to drive the demand for faster, more powerful computing systems, existing hardware solutions are beginning to hold the industry back from achieving the next level of HPC. The Gen-Z standard aims to resolve the complexities associated with processing and analyzing tremendous volumes of data in real time and in a reliable, scalable, and repeatable fashion. HPE is working alongside industry-leading partners as part of the Gen-Z Consortium, and helping to develop a new universal interconnect that will ultimately usher in a new set of simpler, yet more powerful computing architectures. As this game-changing technology is widely adopted and improved upon by the HPC community, it will help to ignite a broad wave of innovation across the entire IT industry.

Learn more at

genzconsortium.org/

a00009348ENW, May 2017