High-performance computing with accelerated HP ProLiant servers

Technology brief, 2nd edition

Introduction

An increasing number of research institutions and businesses are using high-performance computing (HPC) technology for faster innovation and competitive advantage. HP brings HPC within the reach of more organizations by offering complete supercomputer clusters of purpose-built ProLiant servers with x86 processors. These supercomputer clusters use high-performance networking and robust cluster management tools to deliver reliable HPC performance affordably and efficiently. Our approach differs from competitors that require long-term commitments to costly, proprietary supercomputers.

We're also raising performance and driving down costs in HPC visualization and computation environments by building industry-standard ProLiant servers that accommodate high-end graphic cards and accelerators. We jointly engineer and test our systems with products from industry-leading accelerator vendors and independent software vendors (ISVs) for a diverse range of HPC environments, including oil and gas, bioscience, financial services, government, automotive, and aerospace.

This paper should be of particular interest to organizations with compute-intensive workloads that can run in an HPC environment and organizations that need to expand their use of HPC. It begins with a brief description of accelerators and graphics cards if you want to begin using HPC products for computation and visualization. Then it describes our efforts to research, qualify, and benchmark leading accelerators and graphics cards in ProLiant servers to make it easier for you to determine which is best for specific application workloads.

Primarily, this paper describes the key design features we build into ProLiant servers to enable HPC products and our collaboration efforts with a growing ecosystem of partners to help ensure these products work well in your HPC environment.

Accelerators and graphics cards

Graphics cards render 2D or 3D images and video to direct-attached displays. Some organizations use graphics cards in clustered workstations or servers to send graphical information to remote systems over a network interconnect. In contrast, accelerators execute general-purpose calculations, as well as image images and video. They may be supported by general math libraries, or have applications specifically written to take advantage of them. Accelerators significantly increase application performance and reduce the power and space that servers need to run them.

Accelerators

A large part of the HPC industry is transitioning from using proprietary supercomputers to using supercomputer clusters of x86 servers with integrated accelerators from vendors such as NVIDIA®, AMD™-ATI, Nallatech™, and others. The most common accelerators are general-purpose graphic processing units (GPGPUs) and field programmable gate arrays (FPGAs). GPGPUs have wide use as powerful coprocessors for floating point computations. FPGAs target more specialized HPC users in areas such as digital signal processing and medical imaging.

GPGPU accelerators

GPGPU accelerators have many cores that replicate integer and floating-point Arithmetic Logic Units to provide massive parallelism. GPGPUs also have onboard, high-bandwidth memory for local data transfer.

GPGPUs are available as expansion cards, integrated modules, or external racked subsystems—typically using PCI Express (PCIe) links to a server. They can boost execution from 2x to 10x (or more) in roughly the same form factor as servers without GPGPUs.

HP was the first company to build an industry-standard server with integrated NVIDIA GPGPUs. We shipped our first systems in 2007. We began shipping our second-generation NVIDIA GPGPU-enabled system, the ProLiant SL390s G7 server, in 2010. The SL390s G7 servers are part of the HP ProLiant SL6500 Scalable System. We also make a number of ProLiant server platforms that accommodate NVIDIA GPGPU accelerator cards.

FPGA accelerators

FPGA accelerators are integrated circuits that trained designers can program to perform complex logical functions. FPGAs contain programmable "logic blocks" and reconfigurable interconnects for wiring these blocks together. Designers can change the functionality of an FPGA and select the appropriate level of parallelism to implement an algorithm. This capability allows a designer to tailor the circuits for a specific task, resulting in higher performance and efficiencies for some applications. But programming an FPGA from scratch can be costly and labor intensive, requiring designers with specific skills.

FPGA-based accelerators are available on PCIe expansion cards or modules that plug into a CPU socket. FPGA vendors include XtremeData™, Nallatech, and others.

Comparing GPGPU and FPGA accelerators

FPGA and GPGPU accelerators achieve better performance than CPUs on certain workloads. There is no definitive way to determine whether GPGPU acceleration or FPGA acceleration is better. The reason is that applications can exhibit different performance characteristics depending on the accelerator design and software coding. Table 1 identifies some advantages of each accelerator.

Table 1. Advantages of GPGPUs and FPGAs

GPGPUs	FPGAs
Generally easier to use than FPGAs for creating and modifying acceleration applications	May offer the best performance possible for specific HPC applications that do not require frequent changes
Require no hardware re-programming to run a different acceleration app	Typically requires reprogramming for different applications
Work well with 32-bit and 64-bit floating point computations	Work well on small objects like text or integers (1 to 32 bit)
Tend to use high power	Tend to use less power

Graphics cards

Graphics cards off-load graphic renderings from CPUs and output digital and analog video for high-resolution displays. GPUs have a parallel throughput architecture that simultaneously executes multiple software threads through several processor cores. Some GPUs have hundreds of cores. Graphics cards typically require x16 PCIe 2.0 connectors, but they can run in slots with fewer than 16 electrical lanes (x8 for example).

Graphics cards vary in cost, complexity, and power use. Ultra high-end cards are "double-wide" (two slots wide) and they use up to 225 W. These cards can work in select ProLiant DL series servers. High-end graphics cards occupy a single slot and use less than 150 W. They can fit in a broad range of ProLiant servers.

HP accelerator qualification efforts

At the time of publication, HP has qualified various accelerators and graphics cards to work with 14 ProLiant servers. We design ProLiant servers to meet a variety of conditions, including power, thermals, BIOS, and software. We qualify certain accelerators and graphics cards in select platforms to ensure they work properly. We also benchmark card performance using LINPAK, LAMMPS, GROMAC, ANSYS and others, so you can determine which is more beneficial for your application. The following sections list current HP-qualified accelerators and graphics cards as evidence of our commitment to HPC. Our qualification processes are server-specific and change over time. Go to www.hp.com/go/accelerators for the latest accelerator matrix.

HP-qualified accelerators

We have qualified more accelerators to work in more servers than any other manufacturer. Table 2 lists qualified accelerators as of this publication date.

Table 2. Accelerators qualified in select ProLiant servers

Product	Physical/Interface	Max power use
NVIDIA Tesla S1070	1U external box (4 GPUs)/ x16 PCIe 2.0	800 W (ext. to server)
NVIDIA Tesla S2050	1U external box (4 GPUs)/ x16 PCIe 2.0	900 W (ext. to server)
NVIDIA M2050/2070/2070Q	2-slot (1 GPU)/ x16 PCIe 2.0	225 W
NVIDIA Tesla C2050	2-slot (1 GPU)/ x16 PCIe 2.0	247 W
AMD FirePro V7800	1-slot (1 GPU)/ x16 PCIe 2.0	150 W
Nallatech PCIe-280	1-slot/ x8 PCIe 2.0	88 W typical

HP-qualified graphics controllers

We qualify various graphics cards to work with select ProLiant servers. Table 3 lists the graphics cards that we have qualified as of the publication date.

Table 3. Graphics cards qualified in select ProLiant servers

Product	Physical	Max power use
NVIDIA Quadro FX 3800	1-slot x16 PCIe Gen2 card	108 W
NVIDIA Quadro FX 5800	2-slot x16 PCIe Gen2 card	189 W
NVIDIA Quadro 4000	1-slot x16 PCIe Gen2 card	142 W
NVIDIA Quadro 5000	2-slot x16 PCIe Gen2 card	152 W
NVIDIA Quadro 6000	2-slot x16 PCIe Gen2 card	204 W

Enablement for HPC in ProLiant servers

HP ProLiant servers designed for HPC include the following key design features:

- High I/O bandwidth
- High-speed interconnects
- Highly efficient power and cooling
- Integrated management using the HP Cluster Management Utility

High I/O bandwidth

GPGPU-based accelerators use on-board memory to store and process data sets. But most high-end applications still require moving large data and result sets between system memory and the accelerator. This demands a great deal of I/O bandwidth. To accommodate accelerators, ProLiant servers have both dedicated and shared x16 PCIe 2.0 slots (varies by server). The x16 PCIe slots have a data transfer rate of up to 16 GB/s (8 GB/s in each direction).

HP offers servers in a variety of form factors, with different numbers of CPUs, and a mixture of x16 PCIe Gen 2.0 slots. You can order some ProLiant DL servers with x16 PCIe 2.0 slots for full-length, full-height or half-length, full height cards. Option kits are also available. Select ProLiant DL servers accommodate 2-slot accelerator cards. Refer to the specific server's QuickSpecs for details on the options available.

The SL390s G7 has enhanced expansion slot capabilities for HPC environments. It has three different half-width server trays—1U, 2U, and 4U—with different slot support to serve multiple customer needs (Table 4).

Table 4. ProLiant SL390s server trays

	1U half-width	2U half-width	4U half-width
Standard slot	(1) low-profile x16 PCIe 2.0	(1) low-profile x8 PCIe 2.0	(1) low-profile x8 PCIe 2.0
Internal slots	–	(3) x16 PCIe 2.0 for internal 2-slot GPU (up to 225 W each)	(8) x16 PCIe 2.0 for internal 2-slot GPU (up to 225 W each)

High-speed interconnects

A high-speed interconnect between GPGPU-equipped servers is critical for communication in HPC cluster environments. Interconnect choices include 10 Gigabit Ethernet (10GbE) or InfiniBand™ (Quad Data Rate or Double Data Rate). If a server does not have integrated InfiniBand or 10GbE connectivity, you must install an adaptor card for it. This is expensive and time-consuming in cluster environments containing hundreds or thousands of servers.

The ProLiant SL390s server is purpose-built to provide a scalable infrastructure for HPC. Each ProLiant SL390s G7 server includes integrated InfiniBand and 10GbE capabilities through a LAN on Motherboard (LOM), which is more cost-effective than buying adaptor cards. The LOM lets you immediately connect the server to any InfiniBand or Ethernet switch.

An example of a highly scalable server cluster of SL390s servers is the Tsubame 2.0 system at the Tokyo Institute of Technology. HP, NEC, and Tokyo Tech, with technology partners NVIDIA, Mellanox/Voltaire, Intel and DDN, designed Tsubame 2.0 to support diverse research workloads such as weather, informatics, protein modeling, and more. It consists of 1,408 SL390s G7 servers,

each with three NVIDIA Tesla M2050 GPGPUs, and 34 DL580 G7 servers with two NVIDIA S1070 GPGPUs each. Tsubame 2.0 has a total peak performance of 2.4 PetaFLOPS (floating-point operations per second), making it the fourth-ranked supercomputer in the world on the November 2010 TOP500 list. The Green500 list ranks the Tsubame 2.0 as the second-most efficient system and declares it "The World's Greenest Production Supercomputer.

Power and cooling support

High-performance graphics and accelerator cards can draw much more power than other types of PCIe cards. Many cards use 100 W - 300 W, significantly more than the 75 W of power supported through the standard x16 PCIe connector. The PCIe specification defines a method for supplying additional power to cards using auxiliary power cables and connectors on the motherboard. We added power connectors and optional cable kits for select ProLiant servers to provide a total of 150 W or up to 300 W to accelerator cards.

The additional power use generates more heat and increases the cooling burden on servers and facilities. We understand that this additional operating cost is a concern. To control power use, some ProLiant G6 and G7 servers offer enhancements such as HP Advanced Power Manager for power capping and 94% efficient Platinum Common Slot Power Supplies to help you reclaim power and use it to run more equipment with your current infrastructure. These innovations can result in major savings when you consider that there may be hundreds or thousands of nodes.

We also designed advanced sensor and fan control systems into G6 and G7 ProLiant servers to facilitate cooling of high-performance graphics cards and accelerators. These servers include a "Sea of Sensors" throughout the system and on components such as DDR3 DIMMs and disk drives. The servers use the sensors to construct an accurate view of the thermal profile within the server. The iLO 3 management controller uses a sophisticated control algorithm to optimize the speed for each fan based on the sensor measurements. This reduces power use and minimizes fan noise in non-peak conditions.

The ProLiant SL6500 Scalable System uses shared fans and power supplies to deliver higher performance per watt than servers packaged with individual fans and power supplies. The SL design, with front cabling and no midplane, allows for unrestricted airflow. The "skinless" SL design reduces heat retention. The SL390s G7 supports accelerator cards with passive cooling. This means that the SL6500 Enclosure's shared fans cool the cards rather than the less efficient fans on the cards themselves.

HP Cluster Management Utility

HPC users want the ability to control complex computing environments. The HP Cluster Management Utility (CMU) makes it easier to manage tens of thousands of compute nodes—both CPUs and GPUs. CMU is an integrated management system with an intuitive graphical interface. It supports all HP Linux-based environments and systems. With CMU, you can control a supercomputer cluster or a simple group of nodes with a single management tool.

With CMU, you can:

- Measure several characteristics of the server environment, including memory utilization and the rate of I/O reads and writes for each server.
- Monitor and set alerts for temperature, fan speeds, and hardware health metrics (including GPU metrics).
- Perform operations on multiple servers, such as starting them up and shutting them down.
- Install the OS on 1 or 1,000 servers, all from scratch, in less than 2 hours.

HP collaborative efforts in the HPC market

Our collaborative efforts with hardware vendors and ISVs help to bring HPC capabilities into the reach of more organizations. We collaborate with accelerator and graphics card vendors to help ensure their products work well in HPC cluster environments, as proven by the large number of accelerators we've qualified (see Tables 2 and 3).

We collaborate with ISVs to help advance development of HPC applications for industry-specific solutions. One example is our collaboration with oil and gas (O&G) industry ISVs to run seismic analysis and reservoir-modeling applications on ProLiant supercomputer clusters with GPGPU accelerators. O&G companies use our supercomputer clusters to render terabytes of 3D seismic data in high-definition. This allows them to better visualize complex geologic formations and discover reserves that were previously impossible to find. O&G companies are also discovering that ProLiant supercomputer clusters with GPGPU accelerators provide similar performance but require less space, power, and capital than larger CPU clusters.

HP HPC partner ecosystem

Our expanding HPC partner ecosystem includes solution and development partners. Solution partners provide hardware and software products you can use to custom-build an HPC system. We have established relationships with a robust portfolio of solution partners in the following industries:

- Computer-aided design
- Electronic design automation
- Finance
- HPC tools and libraries
- Molecular dynamics
- Oil & Gas
- Video and rendering

Development partners make tools such as compilers, debuggers, and libraries that the solution partners use to manufacture their products. Our development partners include Allinea, MATLAB, Rogue Wave, Wolfram, The Portland Group (PGI), and others.

You can find more information on HP HPC solution and development partners and up-to-date data on qualified acceleration products at www.hp.com/go/accelerators.

Conclusion

This paper described how HP is making HPC more affordable and accessible to more organizations by leveraging industry-standard and open technologies. As of the publication date, we provide GPGPUs on 14 different ProLiant servers in multiple configurations. We jointly engineer and test these systems with products from leading acceleration hardware and software partners to ensure reliability in a broad range of HPC environments. Our expanding partner ecosystem can provide HPC solutions to meet the unique needs of your organization.

For more information

Visit the websites listed below if you require data from additional resources.

Resource description	Web address
HP High Performance Computing	www.hp.com/go/hpc
HP SL6500 Scalable System	http://h10010.www1.hp.com/wwpc/us/en/sm/WF04a/15351-15351-3896136-3896139-4236125.html
HP Visualization	www.hp.com/go/visualization
HP Cluster Management Utility	www.hp.com/go/cmu
NVIDIA GPU Computing	www.nvidia.com/object/tesla_computing_solutions.html
ATI Stream Computing	http://ati.amd.com/technology/streamcomputing/
XtremeData Home Page	www.xtremedata.com
Nallatech Accelerated Computing Solutions	www.nallatech.com

Call to action

Send comments about this paper to TechCom@HP.com


Follow us on Twitter: http://twitter.com/ISSGeekatHP

TC1012725, May 2011