

GPU

Appro HyperPower Start-up Cluster Promo

Based on an open and flexible GPU cluster computing, Appro offers the option of creating scalable clusters using NVIDIA Tesla GPU to solve advancing high performance computing with a massively multi-threaded architecture. This solution combines 8x 1U GPU system powered by the NVIDIA® Tesla™ S1070 together with 8x four-processor quad-core high density server powered by Appro including space for 4U switches in a 24U half size rack configuration.

With 128-processor computing core per GPU, C-language development environment for the GPU, a suite of developer tools as well as the world's largest GPU computing ISV development community, Appro HyperPower start up cluster promotion allows scientific and technical professionals the opportunity to test and experiment their ability to develop applications faster and to deploy them across multiple generations of processors. In addition, it gives customers a choice of server, networking, interconnects and software that can easily be configured, tested and pre-integrated as a part of a complete package to include HPC professional services and support.

Appro HyperPower™ GPU Start-up Cluster Promo

Customization options are available

Configurations	Up to 8x Appro 1424x Quad Socket CPU Server Up to 8x S1070 1U Tesla Quad Socket GPGPU Server Up to 32 Quad Core processors/ 128 cores Up to 32 GPGPU processors/ 7,680 cores One Appro 2U master node One Infiniband and one Ethernet switch
Rack Cabinet	24U - ½ size Rack
Performance	Up to 750GF peak X86 Quad Core Performance Up to 32.88TF peak GPU performance (single precision) Up to 2.76TF peak GPU performance (double precision)
Memory	Up to 768GB of DDR3 ECC memory for x86 servers Up to 304GB of GDDR3 memory for Tesla servers
Storage	Up to 16x 3.5" SATA HDDs or 32x 2.5" SATA HDDs, equal to 16TB of local storage (2TB per server)
IPC Interconnect	DDR or QDR Infiniband (options)
External I/O Interfaces	Gigabit, 10 Gigabit, DDR or QDR Infiniband (options)
Power/Cooling	8kW to 13kW, depending on configurations
Options	Cluster Management Professional Services



Benefits of HyperPower™ GPU Start-up Cluster



- Offers the combination of GPU and CPU processing cores
- Delivers best performance at lower cost and fewer systems than standard CPU-only clusters
- Supports 128 CPU/7,680 GPU cores in an 24U half rack configuration
- Delivers performance and scalability in a dense cluster computing
- Offers a pre-tested configuration that is ready to be used for scientific research development environments
- Choice of server, networking and software with a variety of configuration options
- Tested and pre-integrated solution deployed as a complete package – 20% discount for Universities
- HPC professional services and support available

Benefits of using 1U NVIDIA Tesla

▪Massively Multi-Threaded Computing Architecture - Executes thousands of concurrent processing threads for high throughput parallel processing of mathematically intensive problems.

▪NVIDIA GPU Computing Drivers - Management of the GPU resources and an extensive runtime library for enhanced data management and program execution. Offers a high speed data transfer path and streamlined driver for computing, independent of the graphics driver.

▪Multi-GPU Computing - Multiple Tesla GPUs can be controlled by a single CPU via the GPU computing driver, delivering incredible throughput on computing applications. The power of the GPU to solve large-scale problems can be multiplied by splitting the problem across multiple GPUs.

Major Cluster Components		
Apro Compute Server	- Apro 1424X Server (two motherboards per chassis) Below configuration is per system: - Two Quad-Core Intel Xeon processors. - Up to 48GB of memory, 12 DIMM slots, DDR3 1333MHz - Up to 4x 150GB 2.5" 10k RPM hot-swap SATA Drives - Integrated single DDR/QDR IB port - Integrated dual Gigabit Ethernet ports - Integrated IPMI for remote management - 1x PCI-E 16x Gen2 expansion slot - 1200W cold-swappable PS	
NVIDIA S1070	- Four NVIDIA Tesla GPUs with 960 computing cores (240 cores per processor) - 1.296 to 1.44 GHz frequency processing cores - IEEE 754 single/ double floating point precision performance - 16GB (4GB dedicated memory per GPU) - 408GB/sec (102GB/s per GPU to local memory bandwidth) - Connects to PCI-E x16 Gen 2 card w/ Extender - 2 GPU per PCI-E connector - PCI-E switch internal - Max Power consumption: 800W - Cuda Technology - CUDA™ C programming environment	
NVIDIA CUDA C Programming Technology	- Simplifies many-core programming and enhances performance by off-loading computationally-intensive activities from the CPU to the GPU. - Enables developers to utilize NVIDIA GPUs to solve the most complex computation-intensive challenges. - The world's only C-language development environment for the GPU, the NVIDIA CUDA software development kit includes a standard C compiler, hardware debugger tools, and a performance profiler for simplified application development.	

