PRODUCT DATASHEET

NextIO vCORE™ Extreme GPU Consolidation Appliance

High Performance
Delivers over 16 TeraFLOPS of performance

Reconfigurable
Allows GPUs to be switched between hosts on-the-fly as cluster workloads change

Standards Based
PCIe host connections and GPUs from industry leading vendor NVIDIA

Scalable
4 to 16 GPUs in 3U - the ability to grow as needed

Management
Ease of deployment and monitoring

Serviceable
Easy to remove carriers and advanced management

NextIO's vCORE™ Extreme appliance brings data center capabilities to GPU computing by delivering serviceability, manageability, and flexibility to today's GPU farms. The vCORE Extreme increases uptime and reduces complexity to deliver on the promises of GPU computing.

Scalable GPU Compute

The NextIO vCORE™ Extreme delivers 16 NVIDIA M-Class GPUs in a 3U chassis with a fully serviceable and redundant design that allows for over 16 TeraFLOPS of GPU computing power to be delivered to as many as 8 servers.

Flexibile and Dynamic GPU Configurations

The NextIO vCORE™ Extreme's flexible chassis design reduces GPU configuration complexity. The system allows dynamic changes in server to GPU ratios driven by software instead of physical cabling changes. GPU hot-swap capability is built into the chassis providing quick repair or upgrade of the GPUs, power supplies, and fans without opening servers to make the change. The vCORE product family offers the best in GPU computing for either compute intensive or data visualization needs.

Form Factor	3U 19" Chassis
GPU Modules	10 Front Hot-Plug Slots 6 Rear Hot-Plug Slots
Server Interface Cards	PCIE x16 or PCIe x8
PCI Express Output	8 PCIe x16 server ports
PCI Fan Out Options	x16 to 1, 2, 3, or 4 slots
GPU's Supported	NVIDIA M2070 NVIDIA M2090 NVIDIA M2070Q
Thermals	High-Efficiency 92mm Fans N+1 Fan Redundancy
Management	On-Board BMC IPMI 2.0 Dedicated Management Port
Power Supplies	4 x 1400W, 208V Hot-Plug, High-Efficiency PSU's N+1 Power Redundancy

Product Ordering Information

SKU	Components	Description
C220-XN2090-4	vCORE Extreme - 4 x 2090	3U vCORE Extreme Chassis + 4 Nvidia M2090 GPU
C220-XN2090-8	vCORE Extreme - 8 x 2090	3U vCORE Extreme Chassis + 8 Nvidia M2090 GPU
C220-XN2090-12	vCORE Extreme - 12 x 2090	3U vCORE Extreme Chassis + 12 Nvidia M2090 GPU
C220-XN2090-16	vCORE Extreme - 16 x 2090	3U vCORE Extreme Chassis + 16 Nvidia M2090 GPU
C220-XN2070-4	vCORE Extreme - 4 x 2070	3U vCORE Extreme Chassis + 4 Nvidia M2070 GPU
C220-XN2070-8	vCORE Extreme - 8 x 2070	3U vCORE Extreme Chassis + 8 Nvidia M2070 GPU
C220-XN2070-12	vCORE Extreme - 12 x 2070	3U vCORE Extreme Chassis + 12 Nvidia M2070 GPU
C220-XN2070-16	vCORE Extreme - 16 x 2070	3U vCORE Extreme Chassis + 16 Nvidia M2070 GPU
C220-XN2070Q-4	vCORE Extreme - 4 x 2070Q	3U vCORE Extreme Chassis + 4 Nvidia M2070Q GPU
C220-XN2070Q-8	vCORE Extreme - 8 x 2070Q	3U vCORE Extreme Chassis + 8 Nvidia M2070Q GPU
C220-XN2070Q-12	vCORE Extreme - 12 x 2070Q	3U vCORE Extreme Chassis + 12 Nvidia M2070Q GPU
C220-XN2070Q-16	vCORE Extreme - 16 x 2070Q	3U vCORE Extreme Chassis + 16 Nvidia M2070Q GPU

For assistance in determining the correct configuration for your organization please contact NextIO at 1-877-7-NEXTIO or via email at info@nextio.com or your authorized NextIO reseller.

NextIO vCORE™ Features and Benefits

Optimized for Scale-Out or Scale-Up Parallel Processing	Scalable from 2 GPUs per server to 8 GPUs per server connection. Supports 8 GPUs and up to 8 server connections in 3U
Ease of GPU Management and Deployment	A centralized pool of GPU resources providing non-disruptive GPU replacements and upgrades
Designed for Dense Server Configurations	Supports dense, space-limited, 1U multi-core servers which would not otherwise host internal GPUs
Ideal for High GPU and High Job Counts	Provides the ability to run more jobs faster than traditional clustered GPU solutions
GPU's Powered by the NVIDIA Tesla Fermi Architecture	Delivers cluster performance at 1/10th the cost and 1/20th the power of CPU-only systems based on the latest multi-core CPUs
8192 CUDA Cores (512/2090 GPU) 7168 CUDA Cores (448/2070 GPU)	Delivers up to 665 GigaFLOPS of double-precision peak performance in each GPU, enabling 10.6 TeraFLOPS of double precision performance in a 3U of space. Single precision peak performance is over a TeraFLOPS per GPU over 21.3 TeraFLOPS for a full system (Using M2090 GPUs)
ECC Memory	Meets a critical requirement for mission critical applications with uncompromised computing accuracy and reliability. Offers protection of data in memory to enhance data integrity and reliability for applications. Registers, L1/L2 caches, shared memory, and DRAM all are ECC protected.
System Monitoring Features	Simplifies management and remote monitoring post-installation via IPMI. Status lights on the front and rear of the unit ensures IT staff can see the status whether they are on the either side of the rack.
Up to 96GB of GDDR5 Memory	Maximizes performance and reduces data transfers by keeping larger data sets in local memory that is attached directly to the GPU
CUDA Programming Environment with Broad Support of Programming Lanaguages and API's	Choose C, C++, OpenCL, DirectCompute, or Fortran to express application parallelism and take advantage of the "Fermi" GPUs innovative architecture
High Speed, PCIe Gen 2.0 Data Transfer	Maximizes bandwidth between the host system and the Tesla processors. Enables Tesla systems to work with virtually any PCIe-compliant host system with an open PCI-E slot (x8 or x16)