CRAY

CRAY XD1 DATASHEET

Cray XD1™ Supercomputer

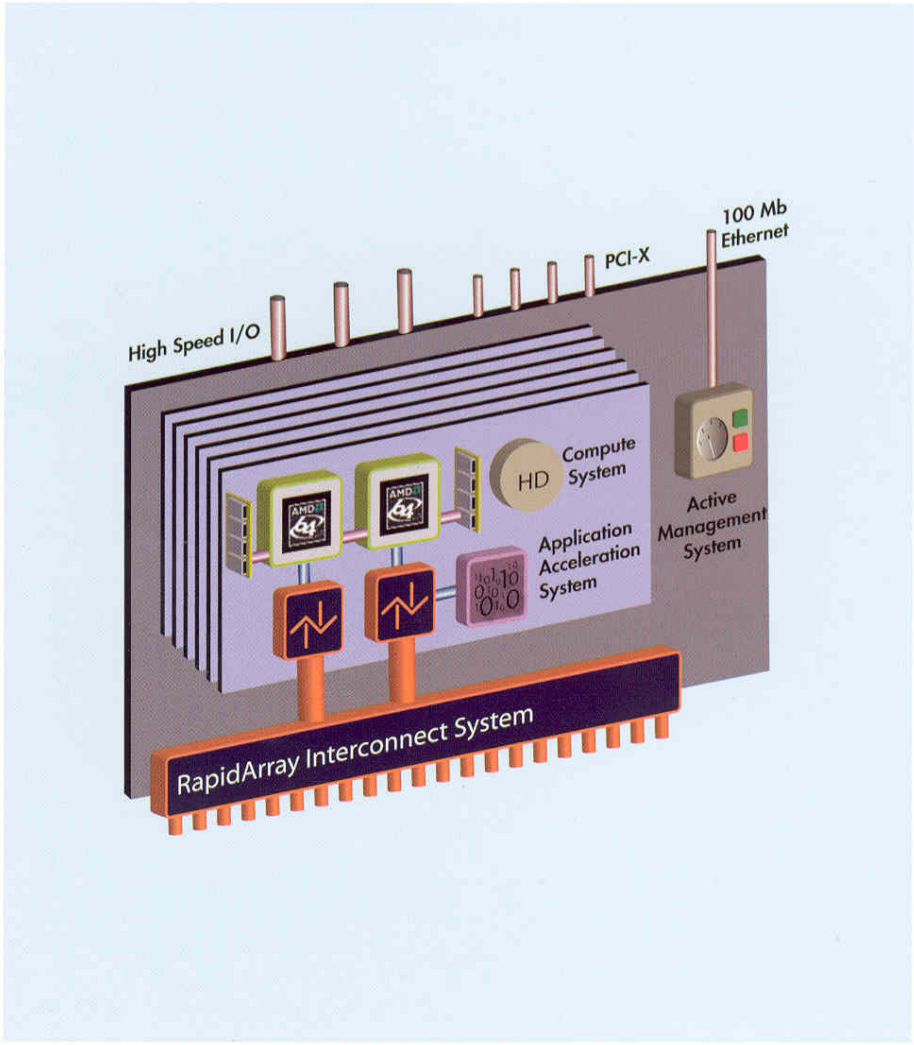
- Purpose-built for HPC — delivers exceptional application performance
- Affordable power — designed for a broad range of HPC workloads and budgets
- Linux, 32 and 64-bit x86 compatible — runs wide variety of ISV applications and open source codes
- Simplified system administration — automates configuration and management functions
- Highly reliable — monitors and maintains system health
- Scalable to hundreds of compute nodes — high bandwidth and low latency let applications scale

The Cray XD1 supercomputer combines breakthrough interconnect, management and reconfigurable computing technologies to meet users' demands for exceptional performance, reliability and usability. Designed to meet the requirements of highly demanding HPC applications in fields ranging from product design to weather prediction to scientific research, the Cray XD1 system is an indispensable tool for engineers and scientists to simulate and analyze faster, solve more complex problems, and bring solutions to market sooner.

Direct Connected Processor Architecture

The Cray XD1 system is based on the Direct Connected Processor (DCP) architecture, harnessing many processors into a single, unified system to deliver new levels of application performance.

Cray's implementation of the DCP architecture optimizes message-passing applications by directly linking processors to each other through a high performance interconnect fabric, eliminating shared memory contention and PCI bus bottlenecks.



Highly modular, the Cray XD1 base unit is a chassis. Up to 12 chassis can be installed in a rack. Multirack configurations integrate hundreds of processors into a single system.

	Chassis	Each Rack
Compute Processors	12	144
Performance	62 GFLOPs*	749 GFLOPs*
Aggregate Switching Capacity	96 GB/s	1152 GB/s
MPI Interprocessor Latency	1.7 µsec	2.0 µsec
Aggregate Memory Bandwidth	77 GB/s	922 GB/s
Maximum Memory	96 GB	1.0 TB
Maximum Local Disk Storage	1.5 TB	18 TB

** peak theoretical performance with 2.6 GHz AMD Opteron*

Cray XD1 System Highlights

Compute Processors

12 AMD Opteron™ 64-bit processors run Linux and are organized as six 2-way SMPs to deliver 62 GFLOPs* per chassis. Finely tuned memory and I/O performance removes bottlenecks and maximizes processor performance.

RapidArray™ Interconnect

The industry's fastest embedded switching fabric, the RapidArray interconnect uses 12 custom communications processors and a 96 gigabytes (GB) per second nonblocking switching fabric per chassis to deliver 8 GB per second bandwidth between SMPs with 1.7 microsecond MPI latency. Each chassis presents 24 RapidArray links externally with an aggregate 48 GB per second bandwidth between chassis.

Application Acceleration

Six Xilinx Virtex-II Pro™ Field Programmable Gate Arrays (FPGAs) per chassis attach to the RapidArray fabric for massively parallel execution of critical algorithm components, promising orders of magnitude performance improvement for target applications.

Active Management

A management processor in each chassis, a realtime operating system, distributed software and an independent supervisory network operate together to monitor, control, and manage every aspect of the computer. The Cray XD1 system enables single system command and control and provides extensive high availability features.

The Cray XD1 HPC system is based on the DCP architecture. This innovative new computer unifies up to hundreds of processors into a single, balanced computer, delivering exceptional performance, reliability, and usability. The Cray XD1 architecture includes four key subsystems:

- Compute Environment
- RapidArray Interconnect
- Application Acceleration
- Active Management

Compute Environment

The Cray XD1 compute subsystem is composed of a high performance Linux operating system and AMD Opteron 64-bit processors.

Direct Connect Bandwidth

The AMD Opteron processor's integrated memory controller increases memory bandwidth and reduces latencies. Three HyperTransport™ links per processor provide up to 19.2 GB per second I/O bandwidth.

Linux

for a wide variety of HPC codes

Global Process Synchronization

The system's Linux scheduler has been optimized to synchronize processes across the system, improving application efficiency by as much as 50%. By ensuring that an application executes in the same timeslot system-wide, time spent in wait-states can be minimized, substantially improving processor efficiency.

Standards Based

The Linux operating system, combined with AMD Opteron's support for 32 and 64-bit x86 compatible computing, allows users to take advantage of a wide range of commercial and open-source applications.

RapidArray Interconnect

The Cray XD1 RapidArray interconnect directly connects processors over high-speed, low-latency pathways. This DCP architecture eliminates the memory contention and PCI bus bottlenecks.

Each chassis includes:

- 12 custom communications processors
- a 96 GB per second nonblocking embedded switching fabric
- Optimized communications libraries

96 GB/s

nonblocking switching fabric per chassis

RapidArray Communications Processors

Tightly coupled to the AMD Opterons and switching fabric, these processors handle memory-to-memory copies, global memory management, and system-wide process synchronization, freeing the AMD Opteron to perform core compute tasks and enabling concurrent computing and communication.

The communications processors deliver 8 GB per second bandwidth with 1.7 microsecond MPI latency between SMPs. The communications processors also deliver 1.1 microsecond latency on the randomly ordered ring latency test (part of the HPCC suite). With interconnect bandwidth on par with memory bandwidth, a major system bottleneck is removed for many applications, improving performance and greatly simplifying software development.

RapidArray Embedded Switching Fabric

A 96 GB per second, nonblocking, crossbar switching fabric in each chassis provides four 2 GB per second links to each two-way SMP and twenty-four 2 GB per second interchassis links.

RapidArray Communications Libraries

The RapidArray communications libraries work in conjunction with the RapidArray communications processors to streamline communications protocols, bypassing the Linux kernel where possible and optimizing the compute/communicate interaction. The RapidArray interconnect delivers outstanding performance for highly parallel applications built upon the common MPI, shmem, and Global Arrays standards.

1.7

microsecond MPI latency

System Topologies

The Cray XD1 embedded switching fabric and 24 interchassis links enable a wide variety of network topologies. Topologies initially supported include direct connect and fat tree.

Application Acceleration

The application acceleration subsystem incorporates reconfigurable computing capabilities to deliver superlinear speedup of targeted applications.

Each Cray XD1 chassis can be configured with six application acceleration processors, based on Xilinx Virtex-II Pro FPGAs, that can be programmed to accelerate key algorithms.

Well suited to functions such as all_reduce operations, searching, sorting and signal processing, the application acceleration subsystem acts as a coprocessor to the AMD Opterons, handling the computationally intensive and highly repetitive algorithms that can be significantly accelerated through parallel execution.

The application acceleration processors are tightly integrated with Linux and the AMD Opterons and use standard software programming APIs, removing a major obstacle to application speedup.

Active Management

The active management subsystem delivers outstanding usability and reliability through single system command and control and self-healing capabilities.

Single System Command and Control

The active management system combines partitioning and intelligent self-configuring features to allow administrators to view and manage hundreds of processors as one or more logical computers.

Partitions divide the Cray XD1 system into logical computers. Administrators view and operate on partitions, rather than on individual SMPs. Self-configuration features automatically translate partition details into SMP configurations. Transaction processing techniques ensure configuration integrity.

Single system control is provided over common administrative functions: system configuration, monitoring and fault tolerance, software upgrades, storage management, network management, user management, security, and resource and queue management. These functions may be accessed using the active manager graphical user interface (GUI) or a command line interface (CLI).

Single system command and control eliminates common configuration problems, increases system uptime, and significantly reduces the time and effort necessary to manage a multiprocessor, high performance computer.

200

sensors to monitor system health

Self-Healing

The Cray XD1 system provides extensive fault detection, isolation, and prediction capabilities, coupled with automated proactive and reactive self-healing intelligence.

Fault Detection: In each chassis, a dedicated management processor with its own supervisory network continuously monitors over 200 critical hardware functions, including temperatures, voltages, in-rush currents, parity errors, and component diagnostics.

Proactive Management: Sophisticated proactive controls adjust a broad range of operating parameters to maintain peak performance and optimal operating conditions. The periodic refresh of system software in an SMP helps avoid problems with corrupted software. These proactive measures improve the mean time between failures (MTBF) and ensure system resiliency and job completion.

Recovery from Failures: The Cray XD1's self-healing intelligence facilitates a quick and automated recovery in the event of a hardware failure, reducing outages from hours to minutes.

Redundancy features include "N+1 sparing" and the ability to reallocate resources in the event of a failure, enabling a replacement SMP to assume the persona of a failed SMP and restore full capacity to the affected partition. Jobs are automatically rescheduled from the last checkpoint.

Balance

The Key to Exceptional Application Performance

It takes more than simply adding processors to increase HPC application performance. MPI applications, such as simulation and modeling, are compute- and communicate-intensive. They must perform millions of calculations and then exchange intermediate results before beginning another iteration.

To maximize processor performance, this exchange of data must be fast enough to prevent the compute processor from sitting idle, waiting for data. Thus, the faster the processors, the greater the bandwidth required and the lower the latency that can be tolerated.

A well-balanced system matches processing power with memory, interprocessor and I/O bandwidth, ensuring that applications are free to communicate results without experiencing delays.

CRAY

CRAY

CRAY XD1 SUPERCOMPUTER
Technical data

CPU	64-bit AMD Opteron 200 series processors, 12 per chassis
Cache	64K L1 instruction cache, 64K L1 data cache, 1 MB L2 cache per processor
FLOPS	62 GFLOPs theoretical peak performance (@ 2.6 GHz)
SMP	Six 2-way SMPs per chassis
Main Memory	12–48 GB PC3200 (DDR400) Registered ECC SDRAM per chassis or 96 GB PC2700 (DDR333) Registered EEC SDRAM per chassis (1- 8 GB per processor)
Memory Bandwidth	12.8 GB/s per SMP
Interconnect	2 or 4 Cray RapidArray links per SMP (4 or 8 GB/s per SMP)
	Fully nonblocking Cray RapidArray switch fabric (48 GB/s or 96 GB/s)
	12 or 24 external Cray RapidArray interchassis links – 24 or 48 GB/s aggregate
	1.7 µs MPI latency between SMPs
	Direct Connect or Fat Tree Topology
Application Acceleration (FPGA)	6 Xilinx Virtex-II Pro FPGAs, XC2VP50-7, 16 MB QDR RAM, 3.2 GB/s interconnect
External I/O	4 PCI-X bus slots
	Dual Port Gigabit Ethernet PCI-X card (up to 8 Gig-E ports per chassis)
	Dual Port Fibre Channel HBA (up to 8 Fibre Channel ports per chassis)
Disk	Up to six 3.5 inch Serial ATA drives (74GB 10K RPM or 250 GB 7200 RPM)
System Administration	Graphical and command line system administration tool sets for partition management, fault management, configuration, security, software updating, telemetry and provisioning across all chassis in a system
	Partitioning of system into multiple logical computers
	Administration of entire partition as a single entity (single system command and control)
	Transaction processing used to ensure configuration consistency
	Workload management
	Automatic interconnect topology verification and auto-configuration of L2 and IP networking
	Automatic response to component failures: isolation of hard failures, re-initialization on soft failures, switching around redundant components.
	Automatic re-start of jobs from last checkpoint following system failure
Reliability	Management processor, network, over 200 measurement points on each Cray XD1 chassis
	Independent 100 Mb/s management fabric within and between Cray XD1 chassis
	Thermal stability maintained through temperature monitoring and regulation of fan speed
	Proactive detection of impending component failure (CPUs, fans, power supplies, memory, interconnect) and automatic isolation of failed components
Operating System	Cray HPC enhanced Linux, Kernel version 2.6.5
File Systems	EXT2/3, NFS v2/3, ReiserFS, Lustre Global Parallel File System
Parallel Processing	MPI 1.2
Shared Memory Access	Shmem, OpenMP, Global Arrays
Compilers	Fortran 77, 90, 95, HPF; C/C++; Java
Power	2200 W typical. 3 phase: Circuit requirement: 20A, 208 V per chassis Single phase: Circuit requirement: 30A, 230 V per chassis
Dimensions	3 VU (5.25") x 23" W x 32" D per chassis (13.3 cm high x 58.4 cm wide x 81.3 cm deep), 12 chassis per rack

For the most up-to-date information please visit www.cray.com

The Supercomputer Company

Global Headquarters:

Cray Inc.
411 First Avenue S., Suite 600
Seattle, WA 98104-2860 USA

tel (206) 701 2000
fax (206) 701 2500

Sales Inquiries:

North America: 1 (877) CRAY INC
Worldwide: 1 (651) 605 8817
sales@cray.com

www.cray.com

0305