



InfiniBand Technology to Speed up Dawning 5000A High-Performance Computing System

Introduction

Mellanox ConnectX, the leading InfiniBand technology in high-performance computing environments, has been selected as the interconnect of choice for the Dawning 5000A, China's fastest supercomputer. ConnectX technology was the best choice because of its ability to meet the most stringent I/O performance and reliability requirements while keeping multi-core CPU overhead and energy requirements to a minimum.

The Challenge of Scale

High-performance computing clusters are widely used in information security, oil exploration, weather forecasting, bio-pharmaceutical, engineering simulation, scientific computing, business computing and other fields. Countries all over the world have invested huge amounts of manpower and resources to build compute clusters capable of carrying out this research.

There are several challenges involved in building a supercomputer system capable of 100TFlops or higher performance. The operating system, software compilers, systems and applications must be optimized to function in a parallel computing environment that supports 10,000 or more CPUs. There are also huge network interconnection issues, because the network needed to connect such a massive system must be highly efficient in order to minimize delays in the transfer of information. At the same time, the network must be reliable and energy-efficient.

Scalability Issues

Gigabit Ethernet, 10 Gigabit Ethernet, and InfiniBand can all be used as the basic technology for interconnecting servers in a supercomputer network. In a fat-tree topology, the peak theoretical speed of the system is 230T Flops. The Dawning 5000A uses ConnectX DDR InfiniBand interconnect to achieve a



Linpack test result of 180.6 TFlops, or more than 77 percent efficiency. At present, the Dawning 5000A is the fastest high-performance computer system in China.

Reliability issues

A 100T Flops supercomputer system will have many CPUs as well as hundreds of TB of memory. With so many CPUs working on large-scale problems, it is vital to protect the reliability of the system hardware and system software. The interconnect must be designed with built-in fault isolation and recovery mechanisms so as to effectively reduce the system MTBF from the current level without relying on servers to perform such error-handling.

Solution

High-Performance

The Dawning 5000A's high-performance computers use AMD's latest low-power, quad-core Barcelona processor. Based on the structure of the blade architecture, the system incorporates a total of about

30,000 computing cores, system memory of more than 120TB, and a 700TB SAN data storage capacity. The 20Gb/s network connectivity employs non-blocking InfiniBand CLOS.

The Dawning 5000A's built-in ConnectX DDR InfiniBand uses optical transmission technology to decrease the number of InfiniBand connections by 50 percent while ensuring high-speed, long-distance signal transmission. With this non-blocking fat-tree design, the interconnect provides a 20Gb/s transfer rate with a measured communication latency of 1.6 microseconds.

High-Productivity

Dawning 5000A uses 4-way blade nodes and 8-way fat nodes, a substantial increase in the parallel system and the degree of practical application. The CPUs use direct memory access technology in the node to achieve non-blocking memory access bandwidth and delay. A single server blade directly addresses 64GB of memory, while the single SMP node directly addresses 128GB to meet the special needs of challenging industry applications.

High Density

For the first time, Dawning 5000A uses a 4 of 4 cores blade server design, in which 7U chassis can be deployed within 40 CPU to achieve 160-core computing density. Dawning 5000A can be deployed inside a single 7U high chassis, which means that the 5000A can deliver a 200-CPU, 800-core, ultra-high density computing environment in a single computing cabinet that delivers more than 6T Flops.

Dawning 5000A uses the built-in blade server ConnectX DDR InfiniBand Switch Module to complete the high-performance computing unit.

High Reliability

Dawning 5000A uses the entire blade server design redundancy, so that the system has no single point of failure.

The Results

Mellanox's ConnectX 20Gb/s InfiniBand adapters and InfiniScale III InfiniBand switch chips offer the best low-latency, high-performance interconnect for Dawning's innovative and highly scalable computing platform. With the Dawning 5000A, China has built the world's most powerful supercomputer. The high-performance computer system achieves maximum performance of 233 trillion floating-point operations. Mellanox's ConnectX adapters provide maximum data throughput and the lowest latency in high-performance computing (HPC) environments to speed up applications and data transmissions.

Summary

Mellanox ConnectX 20Gb/s InfiniBand adapters and InfiniScale III InfiniBand switch chips offer the best low-latency, high-performance interconnect for the Dawning 5000A computing platform.



350 Oakmead Parkway, Sunnyvale, CA 94085
Tel: 408-970-3400 • Fax: 408-970-3403
www.mellanox.com

© Copyright 2009. Mellanox Technologies. All rights reserved. Preliminary information.
Subject to change without notice.

Mellanox and ConnectX are registered trademarks of Mellanox Technologies, Inc. and BridgeX, InfiniBlast, InfiniBridge, InfiniHost, InfiniRISC, InfiniScale, and InfiniPCI are trademarks of Mellanox Technologies, Inc.