HOKUSAI-GreatWave System

1.1 System Overview

The HOKUSAI-GreatWave system consists of the following key components:

- Massively Parallel Computer
- Application Computing Server including ACS with Large Memory and ACS with GPU
- Front end servers that provide the users with the application interface for the system
- Two types of storages with different purposes, one of which is the Online Storage and the other of which is the Hierarchical Storage.

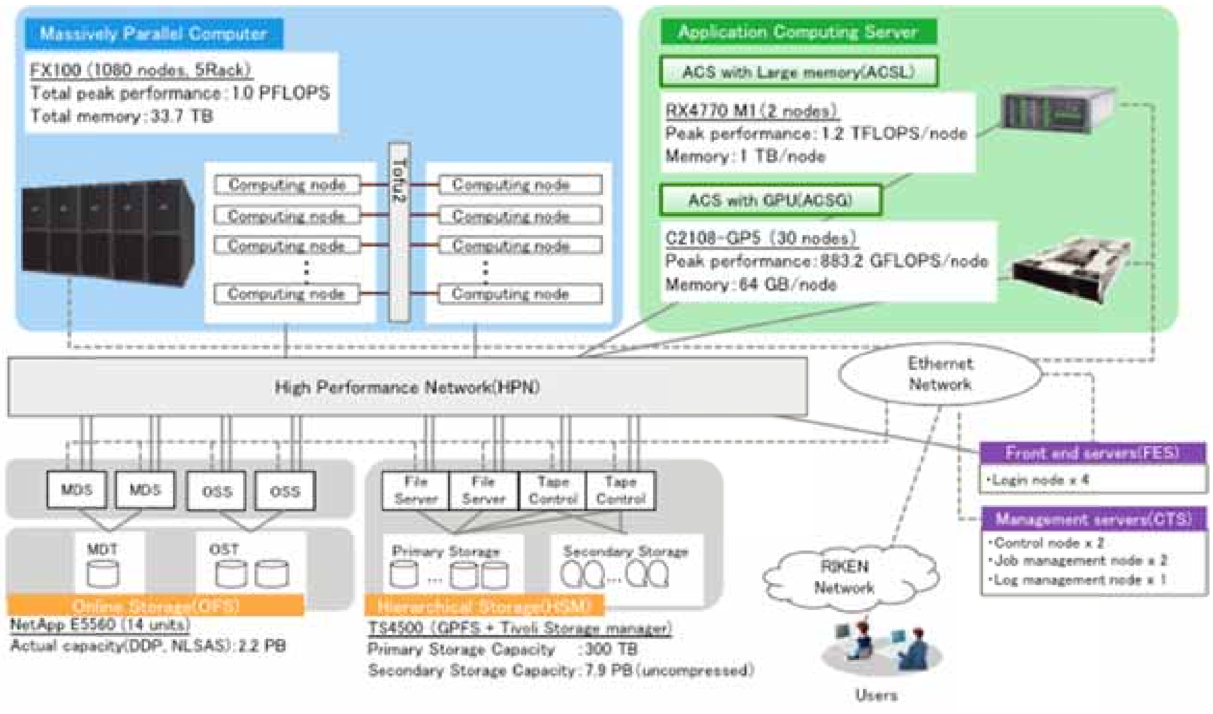

Figure 0-1 System diagram

The Massively Parallel Computer (MPC) comprises FUJITSU Supercomputer PRIMEHPC FX100. FX100, with high performance processors SPARC64 XIfx and high performance memory systems, provides a theoretical peak performance of 1 TFLOPS (double precision) and the high memory bandwidth of 480 GB/s per one node with 32 cores/CPUs. The Massively Parallel Computer of 1,080 nodes provides the a theoretical peak performance of 1 PFLOPS and a total memory capacity of 33.7

TB and uses the 6D mesh/torus network (Torus Fusion Interconnect 2*[1]) to tightly connect each node with 12.5 GB/s high-speed link.

The ACS with Large memory (ACSL) comprises two nodes of PRIMERGY RX4770 M1. Each node provides a theoretical peak performance of 1.2 TFLOPS and a memory capacity of 1 TB. The ACS with GPU (ACSG) consists of thirty nodes of SGI C2108-GP5. Each node provides a theoretical peak performance of 883.2 GFLOPS and a memory capacity of 64 GB. Four NVIDIA Tesla K20X accelerators will be installed on each node of ACS with GPU (by the time of regular operation). The InfiniBand FDR of 6.8 GB/s is used to connect each node to enable high performance communication and file sharing.

The storage environment consists of the Online Storage (OFS) and the Hierarchical Storage (HSM).

The Online Storage (OFS) is a high bandwidth online file system used for the users' home directories, the shared directories for projects and so on, and can be accessed from the Massively Parallel Computer, the Application Computing Server, and the front end servers. The total capacity is 2.2 PB.

The Hierarchical Storage (HSM) consists of the primary storage (cache disks) of 300 TB and the secondary storage (tape library devices) of 7.9 PB (uncompressed) and is the file system used to store large volumes of data files that should be retained for a long term. The users can read or write data to the tapes without manipulating the tape library devices.

You can access the HOKUSAI-GreatWave system using ssh/scp for login/file transfer, or using HTTPS for the User Portal and FUJITSU Software Development Tools (Development Tools). On the front end servers, you can mainly do the following:

- create and edit programs
- compile and link programs
- manage batch jobs and launch interactive jobs
- tune and debug programs

*[1] Tofu Fusion Interconnect 2 is Fujitsu's proprietary high speed interconnect.

1.2 Hardware Overview

1.2.1 Massively Parallel Computer (MPC)

- Computing performance
  CPU: SPARC64™XIfx (1.975 GHz) 1,080 units (1,080 CPUs, 34,560 cores)
  Theoretical peak performance: 1.092PFLOPS (1.975 GHz x 16 floating-point operations x 32 cores x 1,080 CPUs)
- Memory
  Memory capacity: 33.7 TB (32 GB x 1,080 units)
  Memory bandwidth: 480 GB/s/CPU
  Memory bandwidth/FLOP: 0.47 Byte/FLOP
- Interconnect (Tofu Interconnect 2)
  6D mesh/torus
  Theoretical link throughput: 12.5 GB/s x 2 (bidirectional)

1.2.2 Application Computing Server (ACS)

The Application Computing Server (ACS) consists of the ACS with Large memory (ACSL) and the ACS with GPU (ACSG).

1.2.2.1 ACS with Large Memory (ACSL)

- Computing performance
  CPU: Intel Xeon E7-4880v2 (2.50 GHz) 2units (8 CPUs, 120 cores)
  Theoretical peak performance: 2.4 TFLOPS (2.5 GHz x 8 floating-point operations x 15 cores x 8 CPUs)
- Memory
  Memory capacity: 2 TB (1TB x 2 units)
  Memory bandwidth: 85.3 GB/s/CPU
  Memory bandwidth/FLOP: 0.28 Byte/FLOP
- Local disk
  Disk capacity: 3.6 TB ((300 GB x 2 + 1.2 TB) x 2 units)
- Interconnect
  FDR InfiniBand
  Theoretical link throughput: 6.8 GB/s x 2 paths x 2 (bidirectional)

1.2.2.2 **ACS with GPU (ACSG)**

- Computing performance

  CPU: Intel Xeon E5-2670 v3 (2.30GHz) 30 units (60 CPUs, 720 cores)

  Theoretical peak performance: 26.4 TFLOPS (2.3 GHz x 16 floating-point operations x 12 cores x 60 CPUs)
- Memory

  Memory capacity: 1.8 TB (64 GB x 30 units)

  Memory bandwidth: 68.2 GB/s/CPU

  Memory bandwidth/FLOP: 0.15 Byte/FLOP
- Local disk

  Disk capacity: 18 TB ((300 GB x 2) x 30 units)
- Interconnect

  FDR InfiniBand

  Theoretical link throughput: 6.8 GB/s x 2 (bidirectional)
- Accelerator

  NVIDIA Tesla K20X x 4 devices/node

1.3 Software Overview

The software available on the HOKUSAI-GreatWave system are listed as follows:

Table 0-1 Software overview

Category	Massively Parallel Computer (MPC)	Application Computing Server (ACS)	Front End Servers
OS	XTCOS (OS for FX100) (Linux kernel version 2.6)	Red Hat Enterprise Linux 6 (Linux kernel version 2.6)	Red Hat Enterprise Linux 6 (Linux kernel version 2.6)
Compiler	Technical Computing Language (Fujitsu)	Intel Parallel Studio XE Composer Edition	Technical Computing Language (Fujitsu) Intel Parallel Studio XE Composer Edition
Library	Technical Computing Language (Fujitsu) - BLAS, LAPACK, ScaLAPACK, MPI, SSLII, C-SSLII, SSLII/MPI, Fast Basic Operations Library for Quadruple Precision	Intel MKL - BLAS, LAPACK, ScaLAPACK, Intel MPI	Technical Computing Language (Fujitsu) Intel MKL Intel MPI IMSL Fortran Numerical Library
Application	Gaussian	Gaussian, Amber, ADF, ANSYS (multiphysics) GOLD/Hermes, MATLAB, Q-Chem	GaussView, ANSYS (preppost)

You can develop programs on the front end servers for both the Massively Parallel Computer (SPARC) and the Application Computing Server (Intel) although these two systems have different architectures.

1.4 RICC Hardware outline

PC Clusters consist of Massively Parallel Cluster [486 nodes (3888 cores)] and Multi-purpose Parallel Cluster [100 nodes (800 cores)].

1.4.1 Massively Parallel Cluster

- Computation performance

  Intel Xeon X5570 (2.93GHz) 1048 nodes (952 CPUs, 3888 cores)

  Total peak performance: 2.93 GHz x 4 calculations x 4 cores x 972 CPUs = 45.6 TFLOPS
- Memory

  12.5TB (12GB x 1048 nodes)

  Memory bandwidth: 25.58GB/s = 1066MHz (DDR3-1066) x 8Byte x 3channel)

  Byte/FLOP: 0.54 (Byte/Flop) = 25.58GB/s / (2.93GHz x 4calculations x 4cores)
- HDD

  272TB((147GB × 3 + 73GB) × 436 + (147GB × 6 + 73GB) × 50)
- Interconnect (DDR InfiniBand)

All 486 nodes with DDR InfiniBand HCA are configured as a computer network of two-way communication with performance of 16 Gbps per way.

1.4.2 Multi-purpose Parallel Cluster

- Computation performance

  Intel Xeon X5570 (2.93GHz) 100 nodes (200 CPUs, 800 cores) + NVIDIA Tesla C2075 GPU type accelerator x 100

  Total peak perfomance: 2.93GHz x 4 calculations x 4 cores x 100 CPUs = 9.3 TFLOPS

  1.03 TFLOPS (single precision) x 100 = 103 TFLOPS
- Memory

  2.3 TB (24GB x 100 nodes)

  Memory bandwidth: 25.58GB/s = 1066MHz (DDR3-1066) x 8Byte x 3channel)

  Byte/FLOP: 0.54 (Byte/Flop) = 25.58GB/s / (2.93GHz x 4calculations x 4cores)
- HDD

  25.0TB (250GB x 100 nodes)
- Interconnect (DDR InfiniBand)

  All 100 nodes with DDR InfiniBand HCA are configured as a computer network of two-way communication with performance of 16 Gbps per way.

1.4.3 Frontend system

Frontend system is the first host to login to access RICC. Also it provides environment for program development and execution for PC Clusters, MDGRAPE-3 Cluster and Large Memory Capacity Server.

Frontend system has 4 Login Servers. The Login Servers are connected to 2 load-balancers for redundancy and high availability.

1.4.4 Cluster for single jobs using SSD

This cluster can be used via a RICC, provides an environment for jobs that require high-speed I/O and non-parallel.

- Local disk area
  SSD 360GB (30GB / core)
- Interconnect for data transfer
  QDR InfiniBand

1.5 RICC Software Overview

The software available on the RICC system are listed as follows:

Table 0-2 Software overview

Category	Massively Parallel Cluster (MPC)	Multi-purpose Parallel Cluster(UPC)	Cluster for single jobs using SSD(SSC)	Front End system
OS	Red Hat Enterprise Linux 5 (Linux kernel version 2.6)			
Compiler	Fujitsu compiler Intel Parallel Studio XE Composer Edition for Fortran and C++ Linux			
Library	Fujitsu's math libraries - BLAS, LAPACK, ScaLAPACK, MPI, SSLII, C-SSLII, SSLII/MPI Intel MKL - BLAS, LAPACK, ScaLAPACK			
Application	GOLD/Hermes	Gaussian, Amber, ADF, Q-Chem	Gaussian, Amber, ADF, Q-Chem, GOLD/Hermes	GaussView