HOKUSAI-GreatWave システムの概要

1.1 システム構成

HOKUSAI-GreatWave システムは、超並列演算システム、アプリケーション演算サーバ群(大容量メモリ演算サーバ、GPU 演算サーバ)と、システムの利用入口となるフロントエンドサーバ、用途の異なる 2 つのストレージ(オンライン・ストレージ、階層型ストレージ)から構成されるシステムです。

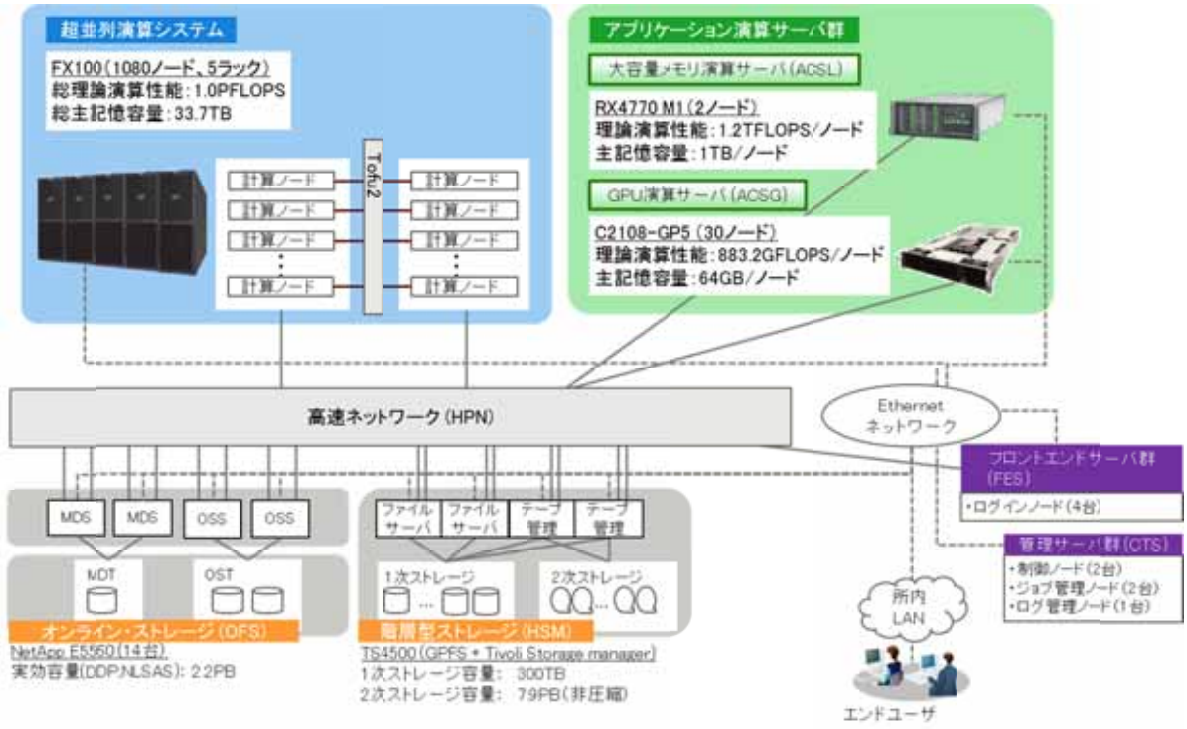

図 0-1 システム構成図

超並列演算システム(MPC)は、FUJITSU Supercomputer PRIMEHPC FX100 で構成します。FX100 は高性能プロセッサ(SPARC64 XIfx)と高速メモリを採用し、1 ノードあたり 32 コア/CPU で 1TFLOPS(倍精度)の理論演算性能と 480GB/s の高いメモリバンド幅を有します。超並列演算システムは 1,080 ノード(総理論演算性能 1PFLOPS、総主記憶容量 33.7TB)で構成され、6 次元メッシュ/トーラスインターコネクト(Torus Fusion インターコネクト 2*[1])により、ノード間は 12.5GB/s×双方向で高速かつ低レイテンシに密結合されます。

*[1] Torus Fusion インターコネクト 2 は、富士通の高速インターコネクトの呼称です。

大容量メモリ演算サーバ(ACSL)は、PRIMERGY RX4770 M1 を 2 ノードで構成します。1 ノードの理論演算性能は 1.2TFLOPS、主記憶容量は 1TB です。GPU 演算サーバ(ACSG)は、SGI C2108-GP5 を 30 ノードで構成します。1 ノードの理論演算性能は 883.2GFLOPS、主記憶容量は 64GB です。GPU 演算サーバ(ACSG)の各ノードにはアクセラレータ(NVIDIA Tesla K20X)を 4 枚搭載します(本運用までに搭載)。各ノードは、InfiniBand FDR(6.8GB/s×双方向)で接続され、高速なノード間通信とファイル共有を実現します。

ストレージ環境は、オンライン・ストレージ(OFS)、階層型ストレージ(HSM)で構成します。

オンライン・ストレージ(OFS)は、各ユーザーのホームディレクトリや課題グループ用の共有ディレクトリなど、広帯域でオンライン性のあるファイルシステムであり、超並列演算システム、アプリケーション演算サーバ群およびフロントエンドサーバから参照可能です。利用可能容量は合計 2.2PB です。

階層型ストレージ(HSM)は、長期保存が必要な大容量のデータ・ファイルを格納するファイルシステムであり、1 次ストレージ(キャッシュディスク)300TB、2 次ストレージ(テープライブラリ装置)7.9PB(非圧縮)を用意しています。ユーザーはテープライブラリ装置を操作することなく、データのテープ書込み・読み出し操作が可能となります。

HOKUSAI-GreatWave システムへのアクセスは、ssh/scp によるアクセスと HTTPS アクセス(利用者ポータル、プログラミング支援ツール)が可能です。ユーザーはフロントエンドサーバ上にて、プログラムの編集、コンパイル/リンク、バッチジョブの操作、インタラクティブジョブの実行、チューニング、デバッグ等の作業を行うことが可能です。

1.2 ハードウェア概要

1.2.1 超並列演算システム(MPC)

- 演算性能
  CPU: SPARC64™XIfx (1.975GHz) 1,080 台(1,080CPU, 34,560 コア)
  理論ピーク性能: 1.092PFLOPS (1.975GHz×16 演算×32 コア×1,080CPU)
- メモリ
  メモリ容量: 33.7TB(32GB×1080 台)
  メモリバンド幅: 480GB/s/CPU
  メモリバンド幅/FLOP: 0.47Byte/FLOP
- インターコネクト(Tofu インターコネクト 2)
  6 次元メッシュ/トーラス
  通信性能:ノード間 12.5GB/s×双方向

1.2.2 アプリケーション演算サーバ(ACS)

アプリケーション演算サーバは、大容量メモリ演算サーバ(ACSL)と GPU 演算サーバ(ACSG)で構成されます。

1.2.2.1 大容量メモリ演算サーバ(ACSL)

- 演算性能
  CPU: Intel Xeon E7-4880v2 (2.50GHz) 2 台(8CPU, 120 コア)
  理論ピーク性能: 2.4TFLOPS (2.5GHz×8 演算×15 コア×8CPU)
- メモリ
  メモリ容量: 2TB(1TB × 2 台)
  メモリバンド幅: 85.3GB/s/CPU
  メモリバンド幅/FLOP: 0.28Byte/FLOP
- 内蔵ディスク
  ディスク容量: 3.6TB ((300GB × 2 + 1.2TB) × 2 台)
- インターコネクト
  FDR InfiniBand
  通信性能:ノード間 6.8GB/s×2 本×双方向

1.2.2.2 **GPU 演算サーバ(ACSG)**

- 演算性能
  CPU: Intel Xeon E5-2670 v3 (2.30GHz) 30 台(60CPU, 720 コア)
  理論ピーク性能: 26.4TFLOPS (2.3GHz×16 演算×12 コア×60CPU)
- 主記憶
  メモリ容量: 1.8TB(64GB × 30 台)
  メモリバンド幅: 68.2GB/s/CPU
  メモリバンド幅/FLOP: 0.15Byte/FLOP
- 内蔵ディスク
  ディスク容量: 18TB ((300GB × 2) × 30 台)
- インターコネクト
  FDR InfiniBand
  通信性能:ノード間 6.8GB/s×双方向
- アクセラレータ
  NVIDIA Tesla K20X ×4 枚/ノード

1.3 ソフトウェア構成

HOKUSAI-GreatWave システムで利用可能なソフトウェア一覧を以下に示します。

表 0-1 ソフトウェア一覧

項目	超並列演算システム(MPC)	アプリケーション 演算サーバ群(ACS)	フロントエンドサーバ
OS	XTCOS(FX100 用 OS) (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)
コンパイラ	Technical Computing Language(Fujitsu)	インテル Parallel Studio XE Composer Edition(Intel)	Technical Computing Language(Fujitsu) インテル Parallel Studio XE Composer Edition(Intel)
ライブラリ	Technical Computing Language(Fujitsu) - BLAS, LAPACK, ScaLAPACK, MPI, SSLII, C-SSLII, SSLII/MPI、高速 4 倍精度基本演算ライブラリ	インテル MKL - BLAS, LAPACK, ScaLAPACK, インテル MPI	Technical Computing Language(Fujitsu) インテル MKL インテル MPI IMSL Fortran ライブラリ
アプリケーション	Gaussian	Gaussian, Amber, ADF, ANSYS(multiphysics) GOLD/Hermes, MATLAB, Q-Chem	GaussView, ANSYS(preppost)

超並列演算システム(SPARC)とアプリケーション演算サーバ群(Intel)は異なる CPU アーキテクチャですが、フロントエンドサーバにて両システムのプログラム開発が可能です。

1.4 RICC ハードウェア概要

PC クラスタは、超並列 PC クラスタ (Massively Parallel Cluster) [486 台(3888 コア)の計算ノード] と多目的 PC クラスタ(Multi-purpose Parallel Cluster)[100 台(800 コア)の計算ノード]で構成されます。

1.4.1 超並列 PC クラスタ

- 演算性能
  Intel Xeon X5570 (2.93GHz) 486 台 (952CPU, 3888 コア)
  理論ピーク性能: 2.93GHz × 4 演算 × 4 コア × 972CPU = 45.6 TFLOPS
- 主記憶容量
  5.8TB(12GB × 486 台)
  メモリバンド幅: 25.58GB/s = 1066MHz (DDR3-1066) × 8Byte × 3channel
  Byte/FLOP: 0.54 (Btye/Flop) = 25.58GB/s ／ (2.93GHz × 4 演算 × 4 コア)
- ディスク容量
  272TB((147GB × 3 + 73GB) × 436 台 + (147GB × 6 + 73GB) × 50 台)
- インターコネクト(DDR InfiniBand)
  486 台の計算ノードに DDR InfiniBand HCA を搭載し、一つの計算用ネットワークとして接続されており、計算用ネットワーク内は、双方向通信可能で片方向 16Gbps の性能が得られるよう構成されています。

1.4.2 多目的 PC クラスタ

- 演算性能
  Intel Xeon X5570 (2.93GHz) 100 台 (200CPU, 800 コア) ＋ NVIDIA Tesla C2075 GPU アクセラレータ 100 台
  理論ピーク性能:2.93GHz × 4 演算 × 4 コア × 100CPU = 9.3 TFLOPS
  1.03 TFLOPS (単精度) × 100 = 103 TFLOPS
- 主記憶容量
  2.3 TB(24GB × 100 台)
  メモリバンド幅: 25.58GB/s = 1066MHz (DDR3-1066) × 8 Byte × 3channel
  Byte/FLOP: 0.54 (Btye/Flop) = 25.58GB/s ／ (2.93GHz × 4 演算 × 4 コア)
- ディスク容量
  25.0 TB (250GB × 100 台)
- インターコネクト(DDR InfiniBand)
  100 台の計算ノードに DDR InfiniBand HCA を搭載し、一つの計算用ネットワークとして接続されており、計算用ネットワーク内は、双方向通信可能で片方向 16Gbps の性能が得られるよう構成されています。

1.4.3 フロントエンド計算機

フロントエンド計算機は、RICC を利用する場合に最初にログインするホストであり、PC クラスタのプログラム開発・実行環境を提供します。

フロントエンド計算機はログインサーバ × 4 台で構成されており、冗長化された高信頼なフロントエンド計算機システムを構成しています。

1.4.4 非並列ジョブ用データ処理 SSD 搭載クラスタ

RICC から利用できる非並列ジョブ用データ処理 SSD 搭載クラスタ（以下、非並列ジョブ用クラスタ）は、主に非並列かつ実行中に高速な I/O が必要なジョブのための環境を提供します。

- ローカルディスク領域
  SSD 360GB (30GB / コア)
- データ転送用インターコネクト
  QDR InfiniBand

1.5 RICC ソフトウェア構成

RICC システムで利用可能なソフトウェア一覧を以下に示します。

表 0-2 ソフトウェア一覧

項目	超並列 PC クラスタ (MPC)	多目的 PC クラスタ (UPC)	非並列ジョブ用クラスタ(SSC)	フロントエンド計算機
OS	Red Hat Enterprise Linux 5 (Linux kernel version 2.6)			
コンパイラ	富士通コンパイラ インテル Parallel Studio XE Composer Edition for Fortran and C++ Linux			
ライブラリ	富士通数学ライブラリ - BLAS, LAPACK, ScaLAPACK, MPI, SSLII, C-SSLII, SSLII/MPI インテル MKL - BLAS, LAPACK, ScaLAPACK			
アプリケーション	GOLD/Hermes	Gaussian, Amber, ADF, Q-Chem	Gaussian, Amber, ADF, Q-Chem, GOLD/Hermes	GaussView