次期超並列型スーパーコンピュータシステムのご紹介

（株）日立製作所

１．はじめに

2008 年 6 月から稼働予定の次期超並列型スーパーコンピュータシステムとして、弊社テクニカルサーバ「HA8000-tc/RS425」952 ノードで構成されたクラスタシステムを納入する予定です。本システムは、筑波大学、東京大学、京都大学の 3 大学で定められた「T2K オープンスパコン仕様」に基づいたクラスタ型のスーパーコンピュータシステムです。

東京大学情報基盤センターで稼働予定の次期超並列型スーパーコンピュータシステムの概要を表 1 と図 1 に示します。本稿では、次期超並列型スーパーコンピュータシステムの特長について、ご紹介します。

表 1　次期超並列型スーパーコンピュータシステムの概要

項目		仕様
システム全体	総理論ピーク演算性能	140.1344TFLOPS
	総主記憶容量	31.25TB
	総ノード数	952（512＋128＋256＋56）
	ノード間ネットワーク性能	5GB/s×双方向（計算ノード群 A） 2.5GB/s×双方向（計算ノード群 B）
	ストレージ装置容量	1PB（RAID6）
プロセッサ	プロセッサ（周波数）	AMD Opteron プロセッサ 8356 （2.3GHz）
	キャッシュメモリ	L2：512kB/コア L3：2MB/プロセッサ
	プロセッサコア理論ピーク演算性能	9.2GFLOPS
ノード	理論ピーク演算性能	147.2GFLOPS
	プロセッサ数（コア数）	4（16）
	主記憶容量	32GB または 128GB
	ローカルディスク容量	250GB（RAID1。OS 領域含む）
	OS	RedHat Enterprise Linux 5

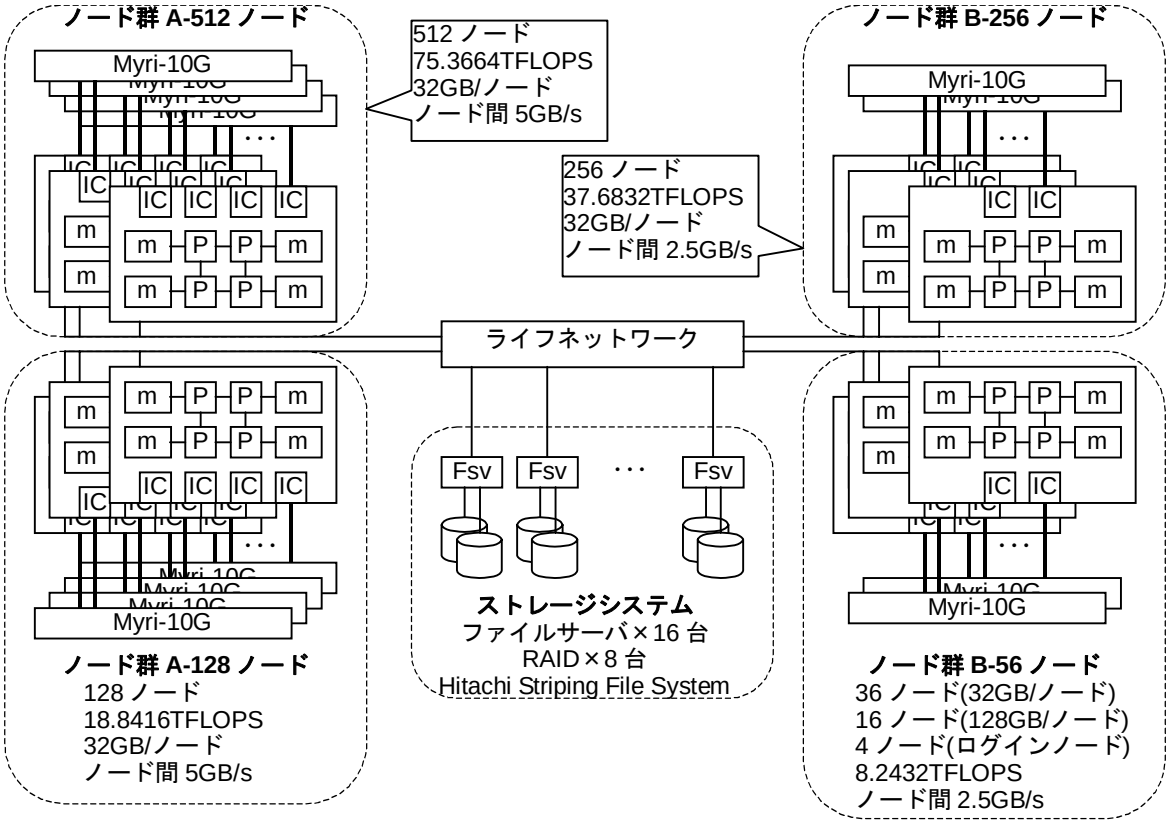

図 1　次期超並列型スーパーコンピュータシステムの構成概略図

２．ハードウェアの特長

テクニカルサーバ「HA8000-tc/RS425」およびクラスタシステムの特長は以下のとおりです。

（1）最新のクアッドコアプロセッサを採用することで優れた性能を発揮

最新のクアッドコア AMD Opteron™プロセッサ 4 個による 16 コア SMP（Symmetric Multi Processor）の並列処理を実現しました。高速多段クロスバネットワークでノード間を接続することによって、ノード間転送の多い大規模科学技術計算を高速に処理します。また、OS はオープンソースである RedHat 社の Linux を採用し、優れたプログラム環境を提供します。

図 2　テクニカルサーバ「HA8000-tc/RS425」外観

（2）高密度実装により、省スペースな高性能システムを実現

2U サイズのコンパクトな筐体にプロセッサやメモリなどを高密度に実装しました。ラック

キャビネットにノードを 16 台搭載したクラスタ環境では単位面積あたりの性能は 3,738GFLOPS/m^2 と省スペースな高性能システムを実現します。

図 3　テクニカルサーバ「HA8000-tc/RS425」内部

（3）高速ノード間ネットワーク

米国 Myricom 社の Myri-10G ネットワークを採用し、10 ギガビットイーサネット級のノード間高速通信を実現しています。特に大量データの高速転送のために、複数のネットワークアダプタを 1 つにまとめて動作させるトランキング機能をサポートしています。

ネットワークは、32 ポートクロスバスイッチ LSI を多段に接続したクロスネットワークを採用しており、任意のノード間データ転送での衝突を最小限にし、システム全体の高いスケーラビリティを実現します。

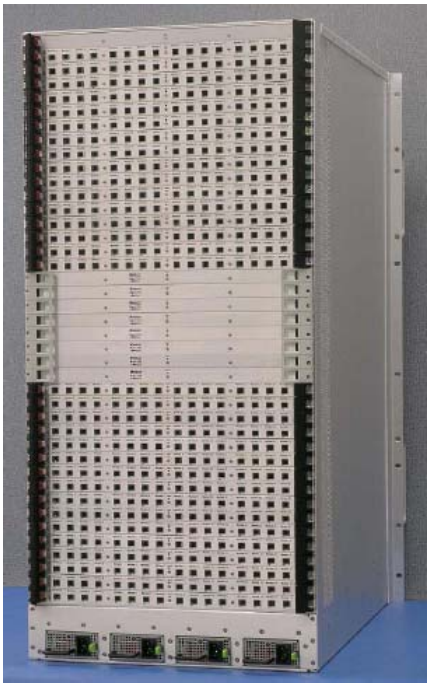

図 4　512 ポートスイッチ

３．ソフトウェアの特長

3.1 RedHat Enterprise Linux 5

HA8000-tc/RS425 クラスタシステムのオペレーティングシステムには RedHat Enterprise Linux 5 を提供します。RedHat Enterprise Linux は、米国 RedHat 社が提供するオープンソースオペレーティングシステムです。

RedHat Enterprise Linux 5 のカーネルバージョンは Linux 2.6.18 カーネルです。GCC 4.1 および glibc 2.4 を提供します。

RedHat Enterprise Linux 5 の詳細につきましては、米国 RedHat 社、またはレッドハット株式会社の Web を参照して下さい。

3.2 Hitachi Striping File System（HSFS）

HA8000-tc/RS425 クラスタシステムのクラスタファイルシステムには、Hitachi Striping File System（HSFS）を提供します。今回のシステムでは、AIX サーバで動作する HSFS サーバ（Hitachi Striping File System for AIX）と Linux ノードで動作する HSFS クライアント（Hitachi Striping File System Client for Linux）が連携して、Linux ノード上にストライピングファイル機構を提供します。

HSFS により、I/O ノード（AIX サーバ）に接続される複数のディスク（サブファイルシステム）を 1 つのファイルシステムに統合します。この統合の方式として、ファイルストライピング方式とブロックストライピング方式の 2 種類を提供します。論理ファイルが作成されるマスタファイルシステムに管理ブロックを格納し、実ファイルのデータをサブファイルシステムに分散するため、マスタファイルシステムは見かけ上、サブファイルシステムの合計容量を持っているようにファイルを作成可能で、ブロックストライピング方式の場合、1 つのサブファイルシステムの容量を超える大容量単一ファイルも作成可能です。

ファイルストライピングは、1 つのファイルを 1 つのサブファイルシステムに格納し、ファイルごとに分散配置します。マスタファイルシステムに対し、複数のサブファイルシステムを定義することで作成される実ファイルを分散配置することができます。

ブロックストライピングでは、1 つのサブファイルシステムの容量を超える大きさのファイルの入出力が可能です。この大規模ファイルへの入出力は、統合したディスクに並列に実行するため、高いファイル入出力性能が実現できます。

HSFS 上のファイルに対する mmap()システムコールは読み込み操作のみのサポートですので、ご注意下さい。

3.3 最適化 Fortran

最適化 Fortran は、ISO 国際規格 ISO1539-1;1997、米国標準規格 ANSI X3.198-1992 及び、JIS X3001-1:1998（Fortran95）規格に準拠した Fortran コンパイラです。

最適化 Fortran は、スーパーコンピュータで実現した強力な最適化機能に加えて、ハードウェアの性能を最大限に引き出すために、各種最適化機能を提供します。プロセッサ間高速同

期処理とコンパイラの自動並列化技術により、高い実効性能を実現します。また、x64 アーキテクチャ単体プロセッサ性能を引き出すためのストリーミング SIMD 拡張命令（SSE、Streaming SIMD Extensions）の発行も合わせて行います。

最適化 Fortran は以下の特長を有し、さらに OpenMP 2.0 仕様をフルサポートします。

○ SMP 並列化機能
①高度な自動並列化機能
②高精度なプログラム解析能力
③自動プライベート化、リダクション並列化
④ループ構造変換による並列化、パイプライン並列化等の先進的並列化
⑤自動並列化支援のための各種指示文を用意
⑥高速並列処理方式と同期削減最適化により、高い並列化効率を実現

○ 最適化機能
①最適化レベル（レベル 0、レベル 3、レベル 4）にもとづき最適化を実施
②ハードウェア性能を最大限引き出すための豊富な最適化
③ループ構造変換最適化
④命令レベル最適化
⑤広域自動手続き自動インライン機能
⑥その他、一般的最適化の殆どすべてを実装
⑦各種最適化指示文を用意

3.4 最適化 C と最適化標準 C++

最適化 C は、国際標準規格 ISO/IEC 9899:1990 及び米国標準規格 ANSI X3.159-1989 に準拠する C コンパイラです。また、コンパイルオプションの指定にしたがって旧言語仕様（K&R 仕様）に対応した互換仕様を利用できます。

最適化標準 C++は、国際標準規格 ISO/IEC 14882:1998 に準拠する C++コンパイラです。また、コンパイルオプションの指定にしたがって旧言語仕様（ARM 仕様）に対応した互換仕様を利用できます。

最適化 C および最適化標準 C++は、スーパーコンピュータで実現した強力な最適化機能に加えて、ハードウェアの性能を最大限に引き出す自動ノード内並列化機能を提供します。また、x86 アーキテクチャ単体プロセッサ性能を引き出すためのストリーミング SIMD 拡張命令（SSE、Streaming SIMD Extensions）の発行も合わせて行います。最適化 C および最適化標準 C++は、OpenMP 2.0 仕様をフルサポートします。

3.5 数値計算副プログラムライブラリ MSL2

MSL2（Mathematical Subprogram Library 2）は、数値計算をする上で必要となる代表的な数値計算上の手法を提供します。利用者が作成した Fortran 言語または C 言語で作成された主プログラムから呼び出すことによって手軽に利用でき、SR11000 で提供している MSL2 と同一インターフェイスを有します。MSL2 には、次の特長があります。

○ 信頼性の高い手法を集めています。また、同一の目的に対して数種の手法による副プログラムを用意し問題への適応性を高めています。
○ 通常のデータだけでなく、性質の悪いデータについても配慮して設計しています。引数の内容については、特に厳重にエラーチェックをしています。
○ 引数の名称および並び順を統一し、使いやすくしています。
○ エラーコードは、ライブラリ全体で統一されており、副プログラム実行後、このコードを調べることによって、利用者のプログラム内で適切な処置ができるようになっています。また、エラーのレベルに応じてメッセージのリスト出力を制御できます。

3.6 行列計算副プログラムライブラリ MATRIX/MPP

MATRIX/MPP（MATRIX calculation subprogram library / Massively Parallel Processors）は、連立 1 次方程式や固有値、高速フーリエ変換、擬似乱数生成といった技術計算の分野でよく使われる機能を提供している副プログラムライブラリです。Fortran 言語および C 言語で作成したプログラムから利用できます。主な機能は以下の通りです。共有メモリ型並列に加え、分散メモリ型並列に対応したインターフェイスを装備しています。SR11000 で提供している MATRIX/MPP と同一インターフェイスを有しています。

○ 基本配列演算
○ 連立 1 次方程式直接解法
○ 連立 1 次方程式反復解法
○ 逆行列
○ 固有値、固有ベクトル
○ 高速フーリエ変換
○ 擬似乱数

3.7 行列計算副プログラムライブラリ MATRIX/MPP/SSS

MATRIX/MPP/SSS（MATRIX calculation subprogram library/MPP/Skyline Sparse matrix Solver）は、構造解析の分野で扱う大次元疎行列に対するライブラリです。Fortran 言語および C 言語で作成したプログラムから利用できます。主な機能は以下の通りです。SR11000 で提供している MATRIX/MPP/SSS と同一インターフェイスを有しています。

○ 基本配列演算
○ スカイライン法
○ オーダリング
○ スパースソルバ
○ 連立 1 次方程式反復解法
○ 固有値、固有ベクトル

以上