

REAL IT PLATFORMを支える NX7700iシリーズ

横山 淳・鈴木 健一
鈴木 久美子・河口 進一

要 旨

統合エンタープライズサーバNX7700i/5080H-64は、オリジナル「A³チップセット」により高性能/高信頼/高い柔軟環境を提供し、REAL IT PLATFORMビジョンを具現化します。本稿では、NEC独自開発の「A³チップセット」と、デュアル・コア インテルItanium2プロセッサにより実現する拡張性/信頼性/柔軟性の高いシステムに盛り込んだ技術を紹介します。

キーワード

●統合 ●サーバ ●エンタープライズ ●NX7700i

1. はじめに

統合エンタープライズサーバNX7700iシリーズは情報システムや大規模基幹業務での利用に要求される性能/信頼性/機能を提供する、エンタープライズサーバシリーズです。本稿では、REAL IT PLATFORMビジョンを具現化する基盤ハードウェアプラットフォームとしてNX7700iシリーズの中核モデルNX7700i/5080H-64（写真）を紹介します。

2. ハードウェアアーキテクチャ

2.1 NX7700i/5080H-64アーキテクチャ概要

NX7700i/5080H-64はNECオリジナル開発の「A³チップセット」を搭載、独自の機能を盛り込むことにより優れた拡張性/RAS機能、パーティション機能をサポートし、柔軟・安心・快適なREAL IT PLATFORM環境を提供します。

NX7700i/5080H-64の構成は、CPU/メモリを搭載する最大8個のセル、8個のPCI-Xスロットと最大4台のディスクを格納可能なIOモジュールを最大8個、セル間およびIOモジュールを結ぶクロスバススイッチ、サーバ全体のマネジメントを行うサービスプロセッサなどから構成されます。CPU/メモリを搭載するセルには最大4個のIntel DualCore Itanium2プロセッサと最大128GB（4GB DIMM使用時）のメモリを搭載可能であり、システムとして32CPU、1TBメモリ、64のPCI-Xスロットという大



写真 NX7700i/5080H-64

きな拡張性を提供します。

本サーバはセルとIOモジュールを最小単位として物理的に分割し、異なるサーバとして動作させる物理パーティション機能を有します。完全なハードウェア・ファームウェアによる分割環境の提供により、性能低下がありません。またこの物理パーティション間をまたぐハードウェアはOSからもアクセスできないためセキュアな稼働環境となっています。後述するフローティングIO機能と自律運用環境を提供するマネジメントGlobalMasterにより、柔軟かつセキュアな仮想化プラットフォームを提供します。

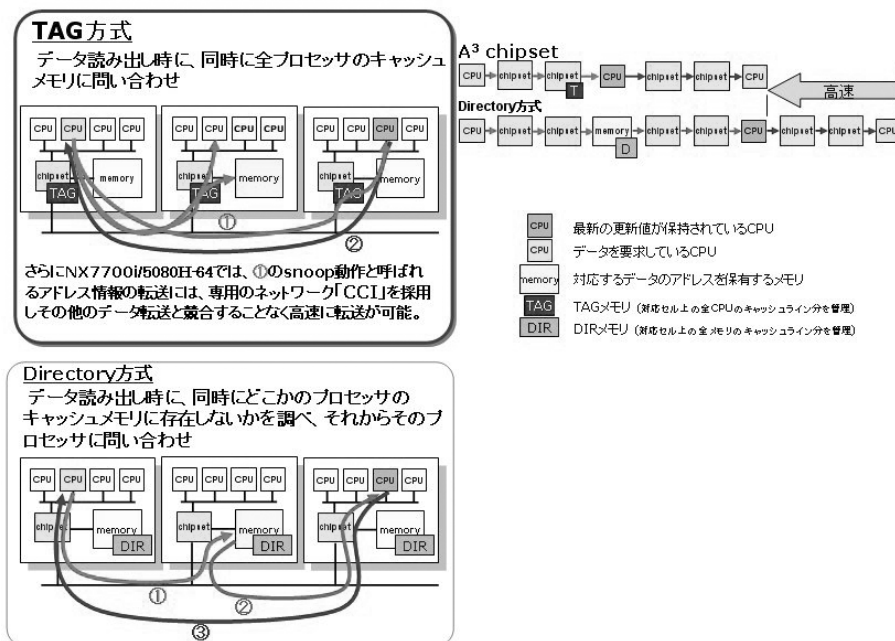


図1 キャッシュ状況情報保持方式比較

2.2 A³チップセット

NX7700i/5080H-64の高い拡張性、信頼性、柔軟性を実現する上での中核を担う独自開発の「A³チップセット」が提供するシステム高速化のための機能を以下に紹介します。

(1)高スループットインタフェース

A³チップセットはセル内に搭載され、CPU/メモリとクロスバススイッチを接続するセルコントローラ、メモリアクセスを制御するメモリコントローラ、セル間・IOを相互接続するクロスバコントローラ、IOコントローラから構成されます。

セル搭載のメモリアクセス性能は25.6GB/s、セルークロスバススイッチ間のインタフェースも25.6GB/sのバンド幅を用意し、セル間をまたぐメモリアクセスにも十分なバンド幅を提供しています。こうした高スループットインタフェースを介して大容量のデータを迅速にハンドリングすることで高性能なItanium2プロセッサの性能をフルに引き出すことが可能となります。

(2)専用キャッシュコヒーレンシインタフェース (CCI)

A³チップセットはまた、データ用インタフェースとは別に

セル間のキャッシュコヒーレンシを制御するための専用インタフェースCCIを用意しています。CCIはすべてのセル間で1対1接続となっており、セル間をまたぐキャッシュ内データのアクセスレイテンシの高速化に寄与します。複数のCPUが稼働するコンピュータにおいて、CPUがキャッシュにアクセスする際、どのCPU内のキャッシュに保持されているデータが最新のものかという状態情報が必要となります。このキャッシュの状態情報をもとに適切なキャッシュにアクセスし、適正なデータを読み出します。このキャッシュ状態情報の保持の代表的な方式として、図1に示すようなTAG方式とDirectory方式とがあります。TAG方式ではチップセット内のTAGが、Directory方式では外部メモリ（DIRメモリ）が状態情報を保有します。TAG方式では、CPUがすべてのTAGに一斉に問合せを行い適正なキャッシュ内データにアクセス可能です。一方でDirectory方式では、データを要求するCPUはまず外部メモリに問合せを行い、それから対応するキャッシュにアクセスします。NX7700i/5080H-64ではキャッシュデータへのアクセス高速化の観点からTAG方式を採用しています。さらにクロスバを介さずすべてのセルどうしをダイレクトにつ

ないだ独自の高速接続方式CCIを採用することで、より高速なデータ転送を実現しています。

3. RAS 機能

3.1 RAS 設計思想

オープンサーバにおいてシステムとしての信頼性・可用性を向上させるためには、一般的にはクラスタ構成をとりますが、NX7700i/5080H-64は単一サーバ構成においても、信頼性・可用性の向上を追求、ディペンダブル・サーバ・テクノロジーによりさらなる安心環境を提供します。「障害時の連続動作」「障害波及の極小化」「故障個所のスムーズな復旧」をテーマに、メモリミラーリングを始め、システムの細部にわたって二重化、冗長化、そしてモジュール化することにより、メインフレームレベルの連続稼働を実現しました。

3.2 メモリミラーリング

メモリ系の訂正不可障害発生時の業務継続とデータ保全性のさらなる向上のために、メモリミラーリング構成をサポートします。メモリミラーリングは、同一のデータを常に2個のメモリブロックへ書き込むという機能で、一方のメモリブロックが障害となった場合にも、残ったブロックによってデータは保全され継続動作が可能となります（図2）。

3.3 チップセットの障害時にマルチパーティションダウンを回避

物理パーティションにより、複数のサーバに分割している構成をとった場合、クロスバススイッチを共有しているためクロスバススイッチの障害が複数パーティションに波及し、複数の業務が停止に追い込まれるという問題があります。NX7700i/5080H-64では、このマルチパーティションダウンを極力回避するためにチップセットに部分縮退機能を持たせています。クロスバコントローラの内部は、パーティション分割時に論理的に分離可能な複数のサブユニットから構成されます。あるとき1個のサブユニットで訂正不可の障害が発生した場合、障害を伝播させることなく、一部のサブユニットに障害を閉塞させ、同サブユニットが属するパーティションのみを強制的にダウンさせた後、障害にかかわる一部のサブユ

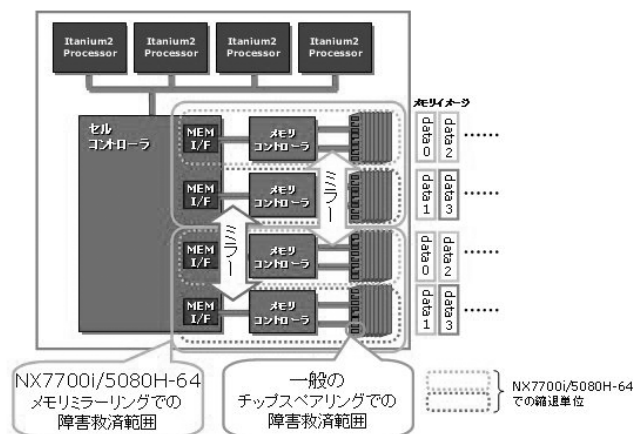


図2 メモリミラーリングによる救済箇所

ニットのみを縮退させます。これにより、他のパーティションへの障害波及を阻止し、複数パーティションが同時に停止することを回避できます。さらに、ダウンしたパーティションも障害経路のサブブロックを縮退後にリブートすることで業務を再開することができます。

3.4 高可用センタープレーン

一般にはセンタープレーンに搭載されることが多いクロスバススイッチを、NX7700i/5080H-64ではモジュール化して独立させています。センタープレーン上には、そのほかの電気的な能動部品も一切排除し、故障確率を極小化しています。また、クロスバススイッチもモジュール化すると同時にOSに意識させることなく、動的に交換可能な機能を持たせています。これにより万一障害が発生した場合でも、前述のとおり故障にかかわるノードのみいったんリブートが発生しますが、その後の保守交換の際には一切業務は停止せず、該当クロスバススイッチのモジュールを交換しシステムの完全復旧が可能となります（図3）。

3.5 クロックカード

クロックカードをモジュール化するとともに、二重化することによってクロック障害の影響範囲を極小化しています。クロックを二重化する際に、多くの場合はクロックの発信器

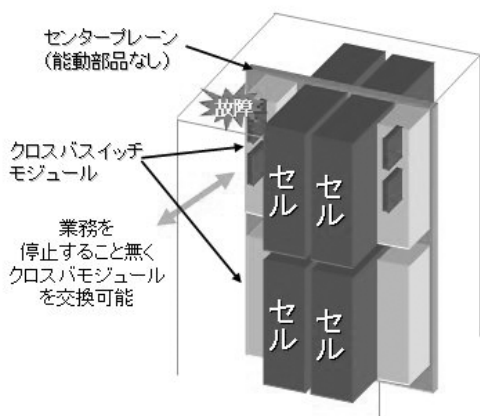


図3 高可用センタープレーン

(オシレータ) は二重化しているにもかかわらず、クロックドライバや増幅器からなるクロック分配部は単一としています。しかし、こうした方法では、クロック分配部あるいは分配パスの障害の場合に対処できません。NX7700i/5080H-64ではクロックの分配部、オシレータと一緒にモジュール化した上で、このモジュールを二重化しています。よって、一方のモジュール内のクロック障害時はリブート後、健全なモジュール側に切り替えを行うことができます。これにより、障害モジュールの交換を待たずに継続動作を可能とし、ダウンタイムを極小化します。また、二重化されているクロックモジュールのそれぞれから筐体内の半分ずつにクロックを供給することが可能です。このモードを利用することで、片方のクロックに障害が発生してもすべてのパーティションがダウンすることを防ぐ構成をとることができます。

3.6 障害検出回路自体のチェック機能

A³チップセットには、主要なデータパスにECCが設けられており、1ビットエラーを検出した際にはハードウェアによるデータ訂正が実行されます。また、チップセットのインタフェース間には多ビットエラー検出、エラーデータの再送機能をサポートしています。A³チップセットが持つこれら優れたRAS機能によってNX7700iシリーズのデータインテグリティが強化されているわけですが、さらに信頼性を高めるために、チップセットの障害検出回路自体をチェックする機能が装備

されています。この機能は、システムのブート時に毎回、すべての障害検出回路を事前にチェックするというもので、業務の稼働中にエラー検出できなくなるという事態を未然に回避します。

3.7 サービスプロセッサ

NX7700i/5080H-64に搭載しているサービスプロセッサは、サーバマネジメント機能や障害処理機能を備えており、RAS機能の中核ともいえる部分です。サービスプロセッサには障害発生時にチップセットから採取したエラーログ情報をもとに組み込み診断 (Built-In Diagnosis : BID) を備えています。BIDは自動的に故障した交換可能ユニット (Field Replaceable Unit : FRU) をピンポイントで特定できるので、ユニットの交換が容易に行え、復旧時間も短縮することができます。また、障害が発生した際に、ハードウェア詳細ログを自動的に通報する機能も備えています。これにより、保守時間のさらなる短縮を実現するとともに、内部パスのデータ通過履歴などのログ採取により、ソフトウェアレベルの障害解析性を高めています。加えて、診断エージェントによる予兆診断機能も有し致命的な障害の発生前に計画的な保守をすることで突然のダウンを未然に防ぐことも可能となります。

4. 優れた柔軟性・運用性 —フローティングI/Oによるリソースの仮想化—

アーキテクチャの項目に記載したように、CPU/メモリを搭載したセルは、クロスバススイッチを介してI/Oモジュールと接続されています。このセルとI/Oモジュールの接続関係は任意にコントロールできるため、物理パーティション機能と組み合わせることでCPU/メモリリソースを仮想化していることと同様な効果を得ることができます。たとえばある処理に必要なCPU/メモリリソースが一時的に不足するような場合にも別の業務 (開発系など) からリソースを融通するといったことが可能になります。さらに、プロセッサメモリなどの致命的な障害が発生した場合にも、予備リソースや他の業務を実行していたリソースの一時転用により速やかに業務を再開することも可能となります。こうしたハードウェアの割り当てやいざという場合の代替などを自律的に行う環境をマネジメントソフトウェア GlobalMasterとの連携によって提供します。

REAL IT PLATFORMを支えるNX7700iシリーズ

5. むすび

以上、REAL IT PLATFORMビジョンを具現化し、柔軟・安心・快適な環境を提供するNX7700i/5080H-64を紹介しました。今後とも、本稿で紹介した高速化技術、高信頼性技術、柔軟性をさらに発展させ、魅力のある製品を提供し続けていく予定です。

*Intel、Itaniumは米国Intel社の商標です。

参考文献

- 1) 高橋ほか：「VALUMOプラットフォームItanium2 32wayサーバNX7700/i9510」、NEC技報、Vol.56、No.7、pp.18-21
- 2) 千田ほか：「Itanium2 32wayサーバシステムアーキテクチャ」、NEC技報、Vol.56、No.1、pp.26-29

執筆者プロフィール

横山 淳
第一コンピュータ事業本部
コンピュータ事業部製品技術部
技術エキスパート

鈴木 健一
第一コンピュータ事業本部
コンピュータ事業部製品技術部
主任

鈴木 久美子
第一コンピュータ事業本部
コンピュータ事業部製品技術部
主任

河口 進一
NECコンピュータテクノ
コンピュータ第一技術部
技術エキスパート

●本論文に関する詳細は下記をご覧ください。

関連URL

<http://www.sw.nec.co.jp/products/nx7700i>