

高い信頼性と拡張性を実現した Express5800/スケーラブルHAサーバ

那須 康之

要旨

大量のビッグデータを高速に処理するためには、処理対象のデータをできるだけCPUに近い場所に置くことが重要です。「Express5800/スケーラブルHAサーバ」シリーズは最大2TBの大容量メモリを利用可能な、ビッグデータ処理基盤として最適なサーバです。多数のI/Oスロットを備えているため、近年注目されている高速なPCI Express接続型SSDを利用した、高コストパフォーマンスなシステム構築にも適しています。本稿では、その優れた特長とそれらを生かした活用例について紹介します。

キーワード

●サーバ ●メモリ ●SSD ●PCI Express ●インメモリ ●データベース

1. はじめに

増え続ける大量のデータを効率よく高速に分析するために、注目されているテクノロジーがいくつかあります。なかでも、並列分散処理とインメモリ処理の2つは、最近脚光を浴びているテクノロジーです。並列分散処理の代表的なものにはオープンソースのミドルウェア「Hadoop」があり、大量のデータを細かく分割して分散並列処理を行い、その結果を集約することで高速なビッグデータ処理を実現できます。一方、インメモリ処理は、処理対象のデータの一部または全部をレイテンシの短いメインメモリ上に置くことで、超高速でのデータ更新や分析処理を可能にします。メインメモリにすべてのデータを置くことができないほど大量のデータを処理する場合は、PCI Express接続型SSD (Solid State Drive) といった新しい高性能記憶デバイスを活用することで高効率なシステムを構築することが可能です。

2. 「Express5800/スケーラブルHAサーバ」シリーズの概要

「Express5800/スケーラブルHAサーバ」シリーズの最上位モデルであるExpress5800/A1080a (写真1) は、1CPU当たり最大10コアを有するインテル Xeon プロセッサ E7-8800/4800製品ファミリーを最大8CPU搭載し、計160スレッドの同時並列処理可能な高性能サーバです。最大2TBの大容量メモリと、最大14個のPCI Expressスロットを備えており、大量の

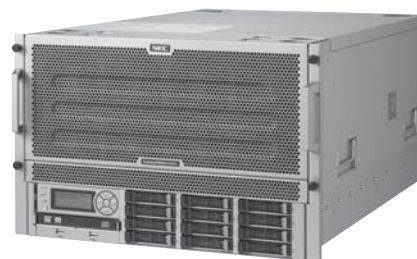


写真1 Express5800/A1080a、Express5800/A1040a外観

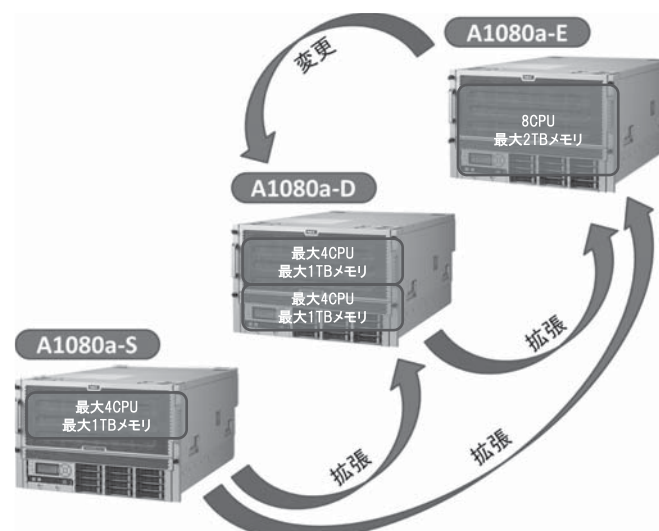


図1 モデルの変更/拡張イメージ



写真2 Express5800/A1020a外観

データを高効率で処理する必要があるビッグデータ処理基盤として最適なサーバです。

また、Express5800/A1080aは、最大4CPUのノードを1つ備えるA1080a-S、最大4CPUのノードを2つ備えるA1080a-D、8CPUのノードを1つ備えるA1080a-Eの3つのバリエーションがあり、**図1**のように導入後に容易に変更/拡張が可能です。例えば、必要な処理能力が少ない場合は、まずA1080a-Sを導入して初期費用を抑え、導入後に必要な処理能力が増えた際にA1080a-Eへ拡張することができます。

更に、長期保守サービスを契約することで最長10年間にわたってサーバを安心して利用できる「長期保守対応モデル」も用意しています。サーバの寿命を延ばせることで、サーバの置き換えに伴う多大なシステム更新コストを削減することができます。

「Express5800/スケラブルHAサーバ」シリーズには、この他にも、最大4CPUを搭載し計80スレッドの同時並列処理が可能な最大1TBの大容量メモリを備えるExpress5800/A1040a (写真1) や、1CPU当たり最大8コアを有するインテルXeon プロセッサ E5-2600製品ファミリーを最大2CPU搭載して計32スレッドの同時並列処理が可能な最大384GBのメモリを備えるExpress5800/A1020a (**写真2**) もラインアップし、さまざまな基幹業務へ対応します。

Express5800/A1080aとExpress5800/A1040aは、優れた障害解析能力と診断機能を有するEXPRESSSCOPEエンジンSP2を備えるとともに、長年のメインフレーム開発によって培われた高信頼部品選定基準を使用した設計や厳しい出荷検査などにより、高い信頼性を実現しています。

また「Express5800/スケラブルHAサーバ」シリーズ向けに、ミッションクリティカルな業務に対応可能なLinux基盤の構築をサポートする「Enterprise Linux with Dependable Support」を提供しており、システム障害の検出と原因の究明を可能にする基盤ミドルウェア製品の「MC SCOPE」と、技

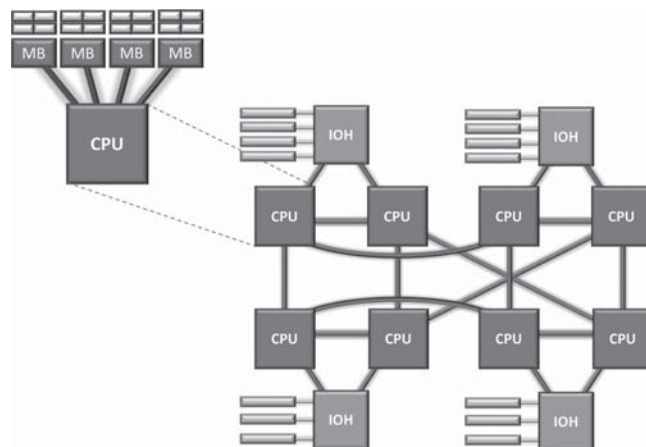


図2 Express5800/A1080a-Eのブロック図

術・運用に関するLinuxサポートサービス「Linuxディペンダブルサポート」によって、Linuxを使った高可用システムを実現します。

2.1 Express5800/A1080aのアーキテクチャ

Express5800/A1080a-Eは、8個のCPUをQPI (QuickPath Interconnect) と呼ばれる高速なインタフェースで接続することで、高い性能と拡張性を実現しています。また、CPUとI/O Hub (IOH) の間もQPIで接続されており、1個のIOHに対して2個のCPUが接続されています (**図2**)。

各CPUには4個のMB (Memory Buffer) を経由して16個のDDR3 DIMMスロットが備えられており、最大128個のDDR3 DIMMを搭載することが可能です。16GBのDIMMを用いて最大2TBの大容量メモリに対応しています。

各IOHには3個または4個のPCI Express 2.0スロットを備えており、最大14枚のPCI Expressカードを搭載できます。この14個のPCI Expressスロットのうち、2個は16レーンに対応しています。

2.2 Capacity Optimization (COPT) 機能

Express5800/A1080aには、予備CPUコアを無効 (利用不可) の状態で導入し、運用開始後の必要なときにCPUコア単位で有効化することが可能な「COPT (Capacity Optimization) 機能」を搭載したモデルをラインアップしています。負荷増

大時への対応だけでなく、CPUコア数分のライセンスが必要なソフトウェアを利用する場合や、メモリ容量を増やすためにCPUを増やしたり、多くのPCI Expressスロットを使用するためにCPUを増やす場合でも、ソフトウェアライセンスは必要なCPUコア数分のみでよいため、無駄な投資を無くすることができます。

2.3 Express5800/A1080aとExpress5800/A1040aのRAS機能

大容量メモリを安心して利用するためには、主記憶メモリのRAS (Reliability Availability and Serviceability) 機能が重要です。

主記憶メモリを構成する2つのDRAMチップが同時に故障しても、システム動作を継続させたままエラーを訂正することができる「DDDC (Double Device Data Correction) 機能」をサポートしています。

また、主記憶メモリを読み出して訂正可能障害を検出した場合に、エラーを訂正して主記憶メモリへ書き戻す「デマンド・スクラビング機能」や、主記憶メモリのデータ全体を一定時間掛けてくまなく読み出し訂正可能障害を検出した場合には、それが訂正不可能障害に悪化する前に訂正する「パトロール・スクラビング機能」をサポートしています。

更に、Windows Server 2008 R2を利用する場合、「MCA (Machine Check Architecture) リカバリ機能」を利用することができます。MCAリカバリ機能は、キャッシュメモリのデータを主記憶に書き戻す際にハードウェアが検出した訂正不可能障害と、主記憶のパトロール・スクラビングの際にハードウェアが検出した訂正不可能障害に対して、OSと連携して障害個所のデータに関連があるプロセスを修復・再起動するなどしてシステムダウンの可能性を低減する機能です。

3. ビッグデータでの活用事例

「Express5800/スケーラブルHAサーバ」シリーズは、NECが長年培ってきたスーパーコンピュータやメインフレームの設計思想を継承し、高い信頼性と可用性を備えつつ、卓越したパフォーマンスを提供します。その優れた特長により企業の基幹システムに数多く採用されています。また、次に紹介する大量データ処理向けのソフトウェアと組み合わせると、コストパフォーマンスの高いビッグデータ基盤を構築することができます。

3.1 InfoFrame TAM (Table Access Method)

InfoFrame TAM (Table Access Method) は、処理対象のデータをメモリ上に配置することにより、大量のトランザクションをリアルタイムに処理するインメモリ・データベース管理ソフトウェア製品です。COPT機能を備えるExpress5800/A1080aでは、搭載メモリ容量に関係なく必要なCPUコア数分だけInfoFrame TAMのライセンスを購入すればよいため、お客様の投資を最適化することができます。

国内最大の金融情報ベンダである株式会社QUICK様の次世代情報配信基盤では、本ソフトウェアと最大14個の高い拡張性のあるPCI Expressスロットを持つスケーラブルHAサーバ、そして「Enterprise Linux with Dependable Support」の導入によって、大量の情報をミリ秒オーダーで高速配信するミッションクリティカルなサービス基盤を実現しました。

3.2 InfoFrame DataBooster

InfoFrame DataBoosterは、最適なデータ構造と処理アルゴリズムと効率的な内部並列処理によって超高速データ処理を実現するソフトウェア製品です。データをメモリに格納して処理を実行するため、最大2TBの大容量メモリに対応できるスケーラブルHAサーバなら将来のデータ増加にも安心です。メモリ容量を増加させるにはCPUの追加が必要になる場合がありますが、Express5800/A1080aではCOPT機能によってInfoFrame DataBoosterのコアライセンスを追加することなくメモリ容量を増加させることが可能です。

3.3 SAP社製HANA

「SAP HANA (High-Performance Analytic Appliance)」は、SAP社のインメモリ・ソフトウェアです。行と列の両構造のデータベースを圧縮するとともにインメモリで提供することで、高速化を実現しています。ドイツにあるグローバルコンピテンスセンターのNEC技術者が、SAPやSUSEの技術者と協力しながら検証を実施し、Express5800/A1080aとExpress5800/A1040aでSAP HANA用サーバとして認証を取得しました。

SAP HANA向けアプライアンスサーバを商品化するに当たっては、優れた書き込み性能を持つ米国Virident Systems社



写真3 Virident Systems社「FlashMAX」

製フラッシュストレージ「FlashMAX」(写真3)のSLCタイプを採用することによって、SAP HANAの性能を最大限に引き出しました。

「FlashMAX」はPCI Express接続型SSDであり、SASやSATAといったHDDと同じインタフェースで接続されるSSDとは異なり、PCI Expressでサーバ本体に接続することで高い書き込み/読み出しスループットを実現します。

Express5800/A1080aは最大14個のPCI Expressスロットを備えているため、「FlashMAX」を複数枚搭載しても、十分な数のネットワークインタフェースカードやストレージインタフェースカードなどを搭載することができ、ビッグデータ時代のニーズに対応したリアルタイム分析基盤を構築可能です。

4. おわりに

本稿では、ビッグデータ処理基盤として最適な高い拡張性を備えた高信頼サーバ「Express5800/スケーラブルHAサーバ」シリーズについて紹介しました。

「Express5800/スケーラブルHAサーバ」シリーズは大量のデータを処理するビッグデータ処理基盤としてだけでなく、その優れた信頼性と可用性によって、基幹業務やプライベートクラウドを支える基盤としても安心してご利用いただけるサーバとして、これからも進化を続けていきます。

*Hadoopは、The Apache Software Foundationの登録商標または商標です。

*PCI Expressは、PCI-SIGの登録商標です。

*インテル、Xeonは、米国およびその他の国における Intel Corporation の商標です。

*Linuxは、Linus Torvalds氏の日本およびその他の国における登録商標または商標です。

*Windows Serverは、米国 Microsoft Corporation の米国およびその他の国における登録商標です。

*SAPは、ドイツおよびその他の国々における SAP AG の登録商標または商標です。

*SUSEは、Novell, Inc.の登録商標または商標です。

*FlashMAXは、Virident Systems, Inc.の登録商標または商標です。

参考文献

- 1) NEC プレスリリース：NEC、金融情報ベンダー「QUICK」の次世代情報配信基盤を構築,2011.8.19
<http://www.nec.co.jp/press/ja/1108/1901.html>
- 2) NEC プレスリリース：NEC SAP HANA向けアプライアンスサーバを販売開始,2012.3.9
<http://www.nec.co.jp/press/ja/1203/0901.html>
- 3) SCSK株式会社 プレスリリース：米国Virident Systems社製フラッシュストレージ「FlashMAX」がNECのSAP HANA向けアプライアンスサーバ「NEC High-Performance Appliance for SAP HANA」に採用,2012.3.9
<http://www.scsk.jp/news/2012/press/product/20120309.html>

執筆者プロフィール

那須 康之
ITハードウェア事業本部
技術エキスパート

関連URL

スケーラブルHAサーバ Express5800シリーズ製品情報:
<http://www.nec.co.jp/products/pcserver/scalable/lineup.shtml>
InfoFrame Table Access Method製品情報:
<http://www.nec.co.jp/pfsoft/tam/>
InfoFrame DataBooster製品情報:
<http://www.nec.co.jp/pfsoft/databooster/index.html>
SAPソリューション:
<http://www.nec.co.jp/sap/>

NEC 技報のご案内

NEC 技報の論文をご覧くださいありがとうございます。
ご興味がありましたら、関連する他の論文もご一読ください。

NEC 技報WEBサイトはこちら

NEC 技報(日本語)

NEC Technical Journal(英語)

Vol.65 No.2 ビッグデータ活用を支える 基盤技術・ソリューション特集

ビッグデータ活用を支える基盤技術・ソリューション特集よせて
ビッグデータを価値に変える NEC の IT インフラ

◇ 特集論文

データ管理/処理基盤

超高速データ分析プラットフォーム [InfoFrame DWH Appliance]
SDN 技術で通信フローを制御する [UNIVERGE PF シリーズ]
大量データをリアルタイムに処理する [InfoFrame Table Access Method]
大量データを高速に処理する [InfoFrame DataBooster]
ビッグデータの活用最適なスケールアウト型新データベース [InfoFrame Relational Store]
高い信頼性と拡張性を実現した Express5800/ スケーラブル HA サーバ
大規模データ処理に対する OSS Hadoop の活用
大容量・高信頼グリッドストレージ iStorage HS シリーズ (HYDRAstor)

データ分析基盤

ファイルサーバのデータ整理・活用を支援する [Information Assessment System]
超大規模バイオメトリック認証システムとその実現
WebSAM の分析技術と応用例～インバリエント分析の特長と適用領域～

データ収集基盤

スマートな社会を実現する M2M とビッグデータ
微小な振動を検知する超高感度振動センサ技術開発とその応用

ビッグデータ処理を支える先進技術

多次元範囲検索を可能とするキーバリューストア [MD-HBase]
高倍率・高精細を実現する事例ベースの学習型超解像方式
ビッグデータ活用のためのテキスト分析技術
ビッグデータ時代の最先端データマイニング
ジオタグ付きデータをクラウドでスケラブルに処理するジオフェンシングシステム
柔軟性と高性能を備えたビッグデータ・ストリーム分析プラットフォーム [Blockmon] とその使用事例

◇ 普通論文

地デジ TV を活用した「まちづくりコミュニティ形成支援システム」

◇ NEC Information

NEWS

スケールアウト型新データベース [InfoFrame Relational Store] が 2 つの賞を受賞



Vol.65 No.2
(2012年9月)

特集TOP