



Supercomputing 2007 (SC07) HP²C 製品アップデート

この 10 年で大きく変わったコンピューティング・プラットフォームの構成において、現在は、標準化され広く流通しているコンポーネントを効率よく組み合わせることで、製品化されたプラットフォームが、HPC 分野でも主流となっています。もちろん、特別に設計され、特定の目的のために最適化されたシステムの価値は、依然として高いものがあります。しかし、今後 10 年の HPC システムの動向を考えた場合、また今後予想されるワークロードや利用モデルにおけるプラットフォームの進化を考えた場合、標準コンポーネントを利用した HPC プラットフォームが大きな位置を占めることは間違いありません。このような HPC プラットフォームの方向性として、弊社では、「HP²C」というコンセプトを提唱し、そのようなコンセプトを実現するためのプラットフォーム事例を積極的に紹介しています。

米国ネバダ州リノで開催されたハイパフォーマンスコンピューティング、ネットワーキングおよびストレージ関連技術のための年次会議である「Supercomputing 2007 (SC07)」でも、弊社が HP²C システムとしてご紹介している Panasas 社や Ciara/VXTECH はブース展示を行っています。スケーラブルシステムズは、各社のパートナーとして、各ブースでの製品説明なども行っています。

Panasas 社ストレージクラスタ

11 月に開催された SC07 での Panasas 関連のニュースとしては、次のようなものがありました。

1) Panasas の高性能パラレルストレージの Windows HPC への拡張の発表

Panasas 社の高性能パラレルストレージのための DirectFlow クライアントソフトウェアの Windows Cluster Server (Windows CCS) への対応プランが発表されました。これによって、将来、Windows CCS での Panasas ActiveStor パラレルストレージクラスタの利用が可能になります。

従来の Linux プラットフォームに加えて、Windows での HPC システムでの高性能で柔軟なストレージの利用が可能になります。



Creating a Unified HPC and Storage Solution

Panasas ActiveStor Storage Provides Parallel Data Storage for Microsoft® Windows® Compute Cluster Server 2003



Windows® Compute Cluster Server 2003

panasas

Accelerating Time to Results™

Customer Profile

Panasas, Inc. is one of the industry leaders in unified storage clusters for the high-performance computing (HPC) market. Panasas' parallel storage products, ActiveStor 3000 and ActiveStor 5000, can provide unprecedented storage capacity and throughput for applications in the Oil and Gas, Sciences, Financial Services, and Media and Entertainment industries.

The Challenge

Today's businesses, universities, and various government agencies, are leveraging more sophisticated applications to drive new levels of innovation and growth. Whether it's mapping the human genome, simulating a car crash, or imaging the earth's substructure to find new energy reserves, these data-intensive applications require both HPC and high-capacity data storage solutions. However, traditional HPC infrastructures can be difficult to use, often requiring specialized IT skills to set-up, administer, and maintain. Similarly, traditional storage systems require significant administrator involvement - copying data from silos of attached storage for cluster processing and then back again. This type of system is not only time-consuming from an administrative standpoint, it introduces risk into a project lifecycle by impeding the workflow - adding to the potential for higher development costs.

Solution

Panasas ActiveStor 3000 and ActiveStor 5000 combine with Microsoft® Windows® Compute Cluster Server 2003 to create a high-performance storage and compute cluster solution, with fast access to data storage from the supercomputing infrastructure. The combination can provide an efficient, easy to use, and highly scalable solution for client-to-storage access. And by using industry-standard ethernet networks and 64-bit processors, this solution offers an attractive ROI that can help minimize hardware costs and maximize your investment in the existing network infrastructure.

Benefits

Combined, Panasas ActiveStor storage and WCCS can help:

- Simplify HPC and storage administration.
- Provide unsurpassed capacity and performance.
- Improve cross-discipline collaboration.
- Drive faster time to results.

Overview

In today's competitive marketplace, increasing productivity and speeding time to market are essential for reducing the cost of innovation. Implementing an HPC solution to reduce the runtime of data-intensive calculation is essential, but traditional open-source HPC systems are known to be complex, often amassing large IT overhead costs. Windows Compute Cluster Server (WCCS) 2003 can provide an HPC platform that helps to remove cost and complexity barriers from HPC.

Another way to achieve dramatic improvement in the run time for data-intensive application processing is through the storage system. Panasas ActiveStor storage can greatly enhance the performance and productivity of your HPC infrastructure by reducing data duplication, enhancing collaboration, and speeding the overall workflow process.

Panasas ActiveStor Storage Clusters

Panasas ActiveStor storage combines a parallel file system with object-based storage to create a single unified namespace for data storage and retrieval. This can help to dramatically reduce IT overhead and simplify management of the storage infrastructure. With support for both Network File Systems (NFS) and Common Internet File System (CIFS) protocols, ActiveStor storage can integrate seamlessly into existing datacenter environments.

ActiveStor storage high bandwidth and enhanced random I/O performance can provide greatly accelerated throughput and processing speeds. With terabyte to petabyte capacity, ActiveStor storage can grow with your business needs, automatically incorporating storage capacity as it is added.

Windows Compute Cluster Server 2003

WCCS is a high-performance computing platform that is easy to deploy, operate, and integrate with your existing infrastructure. WCCS is deployed on industry-standard 64-bit computers. With its familiar user-interface and cost-effective scale-out capabilities, WCCS is removing the cost and complexity barrier from HPC, leveraging your existing Windows infrastructure and administrator's skill-set for managing the HPC system. WCCS also includes an integrated Job Scheduler, which helps increase your computer utilization rate by managing job queues and scheduling runs according to priority.

The Combination

At the heart of building a high-performance, scalable storage and data processing solution is its ability to scale linearly as disk capacity is added. Panasas ActiveStor storage and WCCS can help maximize scalability and availability, while providing the processing speed and random I/O performance that today's data-intensive applications require to run quickly.

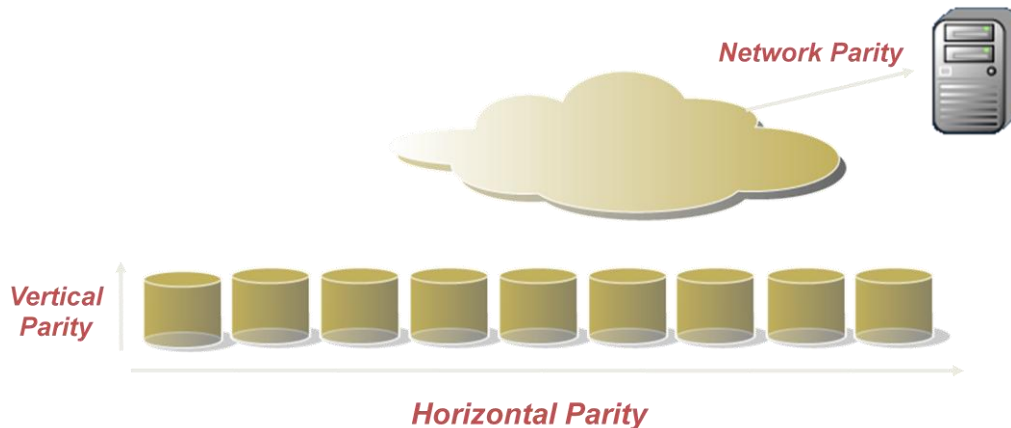
By integrating Panasas ActiveStor storage and WCCS into your organization, you can establish a high-performance solution that requires minimal administrator involvement. It uses industry-standard hardware for cost savings, and helps to capitalize on your existing administrator skill set for managing the system. It helps streamline the workflow process, which can save both time and money, and ultimately enables you to deliver better products to market, faster.

Panasas ActiveStor storage and WCCS can help speed your project development life-cycle, and deliver your products to market ahead of the competition.

Microsoft®

2) Panasas Tiered Parity™ HPCWire Editors' Choice and Readers' Choice 賞の受賞

Panasas 社の革新的なデータプロテクション技術である Tiered Parity™ と ActiveStor パラレルストレージが、HPCWire の年間での編集者賞と読者賞をそれぞれ受賞しました。Tiered Parity™は、従来の RAID システムに対して、高密度が進むディスクのメディアエラーに対する信頼性の向上を図ることが可能なアーキテクチャです。



Panasas Tiered Parity™ アーキテクチャ

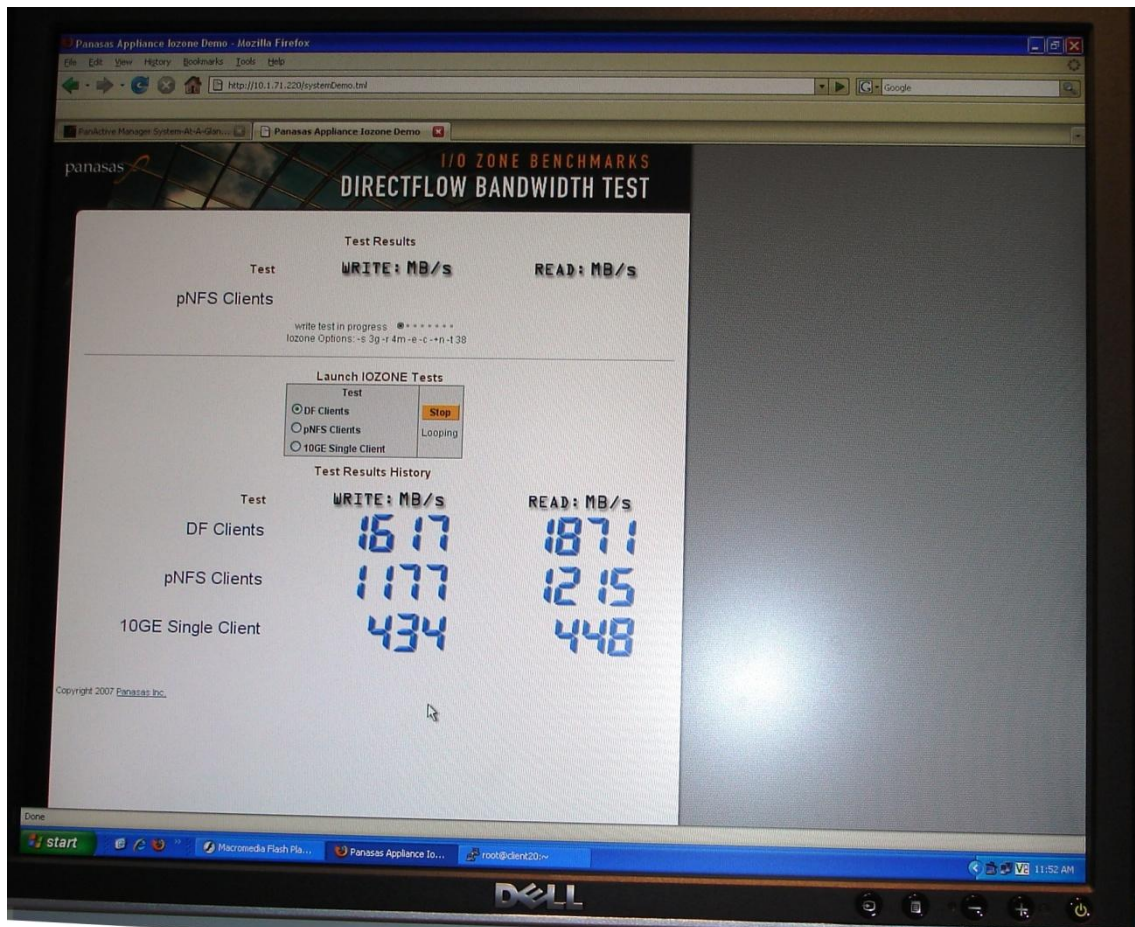
3) Panasas での 10GbE サポート

Panasas の ActiveStor パラレルストレージでの 10GbE のサポートが発表されました。Panasas DirectConnect10 による 10GbE サポートによって、従来の ActiveStor パラレルストレージに対して、50%以上の性能向上が可能となっています。

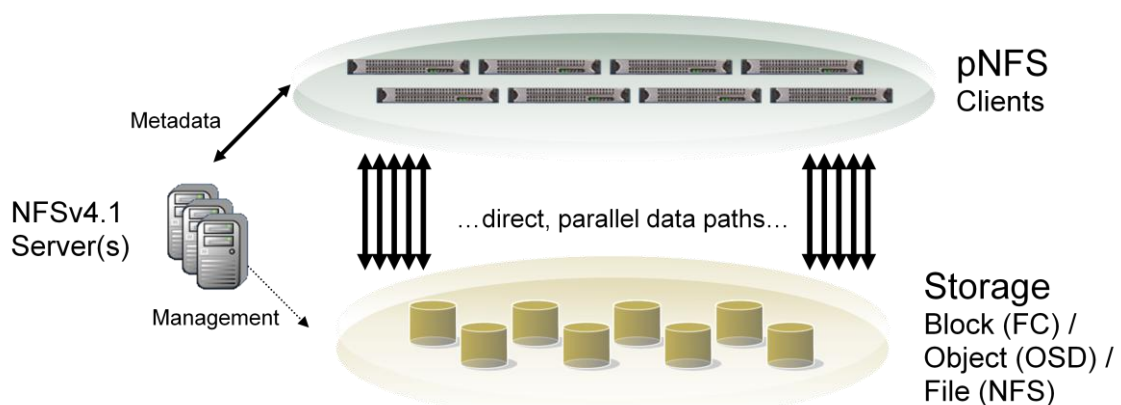
10GbEを利用することで、Panasas の ActiveStor はより高速な処理が可能になります。実機でのデモでは、3 Shelf 構成の ActiveStor に対する I/O ZONE ベンチマークでのデモが行われています。一台のクライアントからの I/O 性能として、~450MB/秒の性能が示されています。

Panasas 社は、現在、NFS 4.1 で採用が検討されている pNFS の開発にも積極的に取り組んでいます。今回のデモでは、この pNFS によるカーネルモジュールを利用したデモも実施されています。(デモの画面) Panasas 社独自のクライアントソフトウェアである DirectFlow (DF Clients) と比較して、この pNFS クライアントは、現時点では性能は劣りますが、今後、ソフトウェアの改良やシステムの設定などによって、より高い性能が得られるようになることが期待出来ます。(この pNFS のカーネルモジュールは、Panasas 社がオープンソースに提供した Linux のカーネルモジュールとなりますが、実際の pNFS としての初めてのデモとして、実機デモが出来るレベルにはなっています。)

【技術解説: http://www.hp2c.biz/doc/HP2C.Biz/hp2c_panasas_pNFS.html 】



I/O ZONE ベンチマークによるバンド幅の計測テスト:Panasas は、3 Shelf 構成で、クライアントは、20 クライアントの構成でのテスト結果。10GbE のテストは、一つのクライアントからの IO バンド幅の計測結果



pNFS の構成図:最初にクライアントは従来の NFS サーバの処理と同じように NFS サーバとデータのやり取りを行います。この後の処理は、従来の NFS とは違います。NFS サーバはクライアントのデータのストレージデバイス上でのマップを確認し、そのデータに対する書き込み/変更/読み

込みの権利に従って処理を行います。従来の NFS では、全てのデータは、NFS サーバを介し、NFS サーバがデータパスの中心に位置していました。pNFS では、データ処理のパスは、クライアントとストレージデバイスが直接データのやり取りを行い、NFS サーバは介在しません。このため、クライアントとストレージデバイス間でのより高速なデータ処理と NFS サーバ自身の処理負担によるオーバーヘッドを大幅に低減します。

NFS の持つ多くの利点をそのままに、ボトルネックを解消し、データへのパラレルアクセスを可能とする pNFS では、より高速なスループットと性能を犠牲にすることなく、容量の増設も可能となります。

Panasas 社の創作者で CTO の Garth Gibson による Panasas 社ブースでの技術説明は、Panasas Tiered Parity™ について、行われています。Panasas Tiered Parity™は、信頼性を損なうことなく性能とスケーラビリティの向上を図ることを可能とします。Panasas Tiered Parity™は、ストレージシステムに対して、そのデータの信頼性の向上を図るソフトウェアによるアプローチです。このアーキテクチャは、メディアエラーに対するユニークなソリューションであり、この技術の導入により将来のディスクメディアのより高密度化に容易に対応することが可能となります。



Panasas Tiered Parity™は、信頼性に関する問題を解決するエラー検知とデータ修正のためのアーキテクチャです。各 Tier オペレーションは、独立したパリティの処理を行うことが可能であり、エラー検知とデータ修正を行います。Panasas の Tiered Parity™が提供する 3 つのパリティ処理は、互いに相互補完を行うことで、ハイパフォーマンスストレージとして、現在、最もスケーラブルで信頼性の高いストレージシステムを構築することを可能とするアーキテクチャです。【技術解説：http://www.hp2c.biz/doc/HP2C.Biz/hp2c_panasas_Tiered_Parity.html】

- Vertical Parity は、ディスクドライブの信頼性の向上を図ります。これは、メディアエラーの発生に際して、そのデータエラーの排除を修復を可能とします。RAID Array として利用されるディスク単体の信頼性とエラー回復を図ることを可能とします。
- Horizontal Parity は、通常の RAID と同じように複数のディスクドライブ間での RAID グループのデータの信頼性を提供します。Panasas 社の Object RAID は、より高速に、効率よくシステムの再構築を可能とします。
- Network Parity は、ストレージシステムとクライアント間でのデータ統合を行います。ネットワークインフラが引き起こすデータの破損をクライアント自身がデータ検証を行うことで防ぐことができます。

Panasas 関連のイベントでは、Panasas 社主催の朝食会で、「Panasas CAE at SC07」として、Panasas システムの主に製造業での利用についてのアップデートがありました。講演の中では、US ボーイング社からの「Shared Object-Based Storage and the HPC Data Center」としての説明がありました。



Panasas CAE at SC07: Panasas 社のマーケティングディレクターの Stan Posey による Panasas の CAE シミュレーションに対する取組の説明

US ボーイング社は、民間での Panasas のストレージクラスタを大規模に導入している会社の一つですが、実際の運用やシステムの構成などについて、詳しい説明がありました。HPC データセンターとして、数千コアを持つクラスタとベクトル計算機から構成される HPC システムでのストレージシステムの構築事例として、非常に参考となるものです。また、Panasas 社のストレージは、「Production」と「Evaluation」の 2 セットが用意されており、Panasas 社の新しい ActiveScale オペレーティングシステムのアップグレードなどの検証も効率良く行っていると説明されています。また、ボーイング社では、Panasas を含む HSM システムの運用を行っており、その方法も紹介されています。



Engineering, Operations & Technology
Information Technology

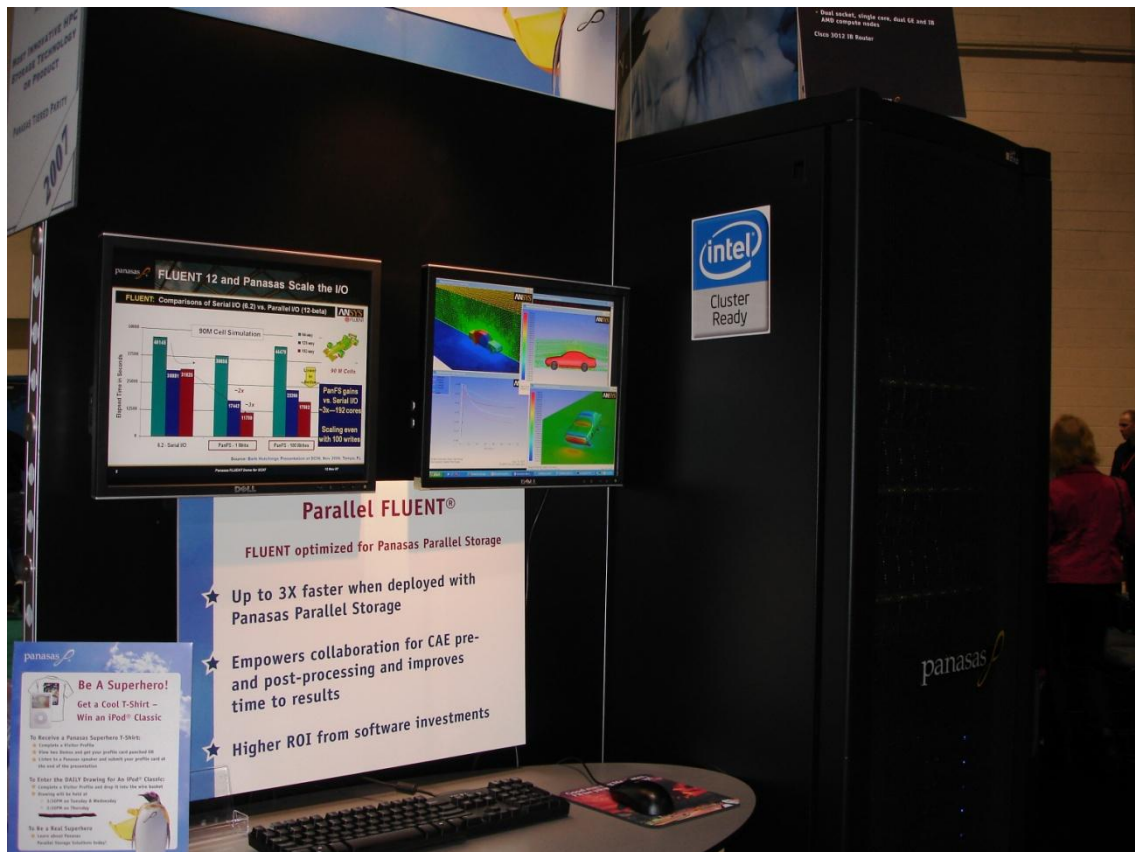
Shared Object-Based Storage and the HPC Data Center

Jim Glidewell/Claude Asher
High Performance Computing
Enterprise Storage and Servers

BOEING is a trademark of Boeing Management Company.
Copyright © 2007 Boeing. All rights reserved.

今回の Panasas 社のブースでは、流体解析アプリケーションの Fluent のデモが行われていました。自動車、航空宇宙をはじめとする製造業では設計サイクルを短縮し、コストを削減することが求められています。一方、安全性や環境に対する配慮は、より高い要求に答えることが必要になります。また、厳しいビジネス上の競争に勝ち抜くためには顧客が求める高い品質と優れたデザインが必要になります。

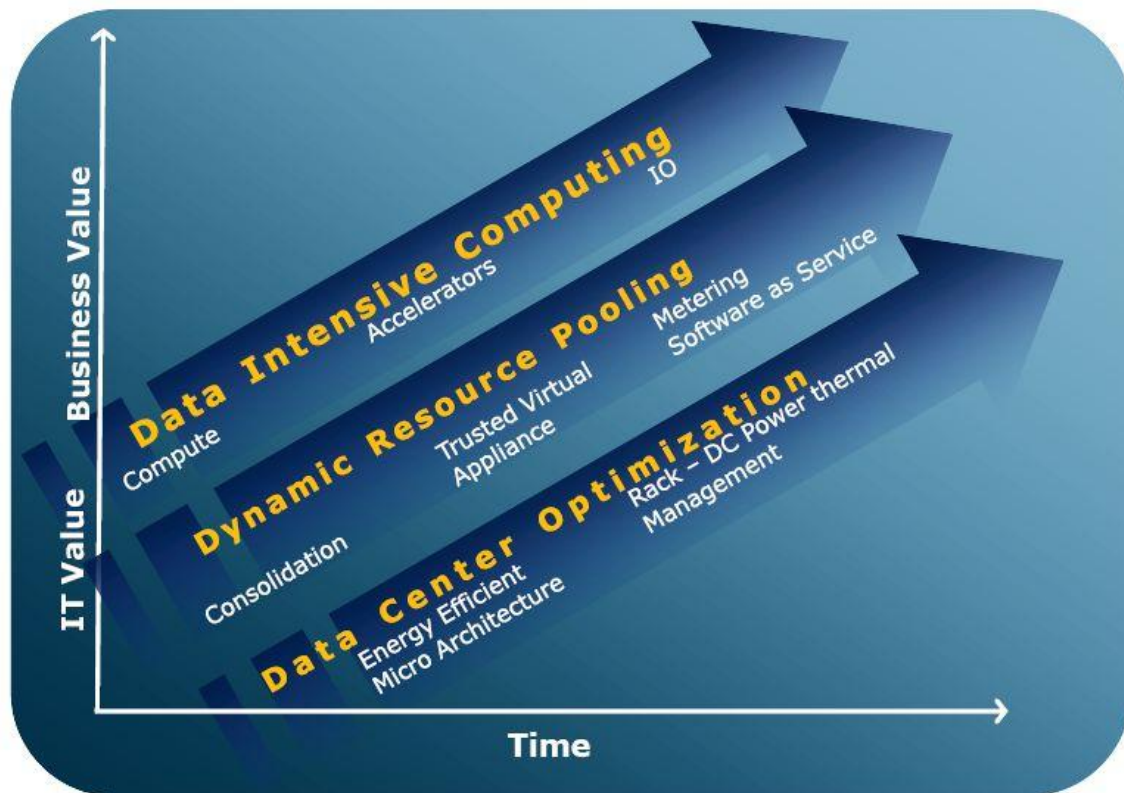
Linux クラスタは、このような製造業が直面する困難な課題に対して、CAE シミュレーションをより高速に、また、低コストで実施するためのソリューションを提供し、活用されています。CAE シミュレーションにおける Linux クラスタは、解析規模や要求されるスケーラビリティが増大するに伴って I/O 処理がボトルネックとなるケースが増大しています。従来から並列処理における I/O 処理は、ボトルネックとして認識されていましたが、現在ではより高いスケーラビリティと絶対的な性能が求められるようになり、従来よりも I/O のボトルネックは大きな問題となっています。Linux クラスタが CAE シミュレーションに革新をもたらしたように、Panasas の ActiveStor ストレージソリューションは、Linux クラスタにおける I/O 処理に革新をもたらし、更に CAE シミュレーションの発展と利用促進を図ることを可能とします。Panasas ActiveStor は、CAE シミュレーションと Linux クラスタの利用効率と速度向上を実現するために次の利点があります。1) アプリケーションの実行に際して、計算と I/O の双方のスケーラビリティの向上を可能とします 2) 大規模な CAE シミュレーションでのポスト処理をより効率的なデータ移動で実現します。



FLUENT でのパラレル IO の成果のプレゼンテーション:

Panasas ActiveStor は、共有ストレージ上でのシングルネームスペースを提供するパラレルストレージソリューションです。CAE シミュレーションとプリ・ポスト処理などの各ステージの CAE ワークフローでのストレージ統合を可能とします。また、Panasas ActiveStor は、最新のオブジェクトベースのストレージアーキテクチャを採用することで、従来のブロックベースの I/O ではなく、I/O 処理を一つのオブジェクトとして処理することで、クラスタの計算ノードからの I/O 要求をパラレルに処理することを可能とします。NFS と NAS(Network Attached Storage)アーキテクチャのボトルネックを解消する新しいストレージアーキテクチャとなります。I/O 負荷の大きな CFD シミュレーション(非定常、LES など)を Linux クラスタで行い場合、この Panasas ActiveStor のパラレル I/O は、アプリケーションのスケラビリティの向上とより高いスループットを提供します。

Fluent の事例では、100 コアを超えるような並列処理での IO 処理によるスケラビリティのボトルネックがどのように解消されたかを示しています。



Source: Intel

ストレージシステムの性能は、HPC データセンターにおいては、現在、非常に重要な IT インフラとなっています。これは、図で示すように時間が経過するに従って、そのビジネス上の価値を高めるものになります。

今回の SC07 発表された Panasas 社の Tiered Parity™や WindowsCCS 対応などについては、今後その動向を紹介していきます。



「Panasas Alliance Partners」として、スケーラブルシステムズ株式会社も紹介されています。

Ciara/VXTECH 社サーバ、クラスタ製品

スケーラブルシステムズが取り扱っている全てのサーバ、クラスタ製品は、インテルの最新クアッドコア、デュアルコアプロセッサをサポートしています。インテルの 45nm プロセス技術を用いて開発されたクアッドコア インテル® Xeon®プロセッサ 5400 番台も搭載可能です。

クアッドコア インテル® Xeon®プロセッサ 5400 番台は、以下のような特徴を持ち、弊社の HP2C 製品は、このプロセッサによりより高性能で高い生産性の提供が可能となります。

主なテクノロジー

- ☐ 45nm プロセス技術に基づくクアッドコア/デュアルコア インテル® Xeon® プロセッサ
- ☐ 50% 大容量化された L2 キャッシュ
- ☐ 20% 高速化された FSB(1600MHz)
- ☐ PCI Express* 2.0 によって 2 倍に拡大した帯域幅
- ☐ 大容量メモリのサポート (最大 128GB)
- ☐ 40 レーンの構成可能な PCI Express* 1.0 および 2.0

主な機能

- ☐ より高いメモリバンドとメモリ領域を必要とするアプリケーション
- ☐ ハイパフォーマンスクラスタ環境
- ☐ マルチタスク・ユーザー環境

今回の SC07 の期間中インテルはこの新しいデュアル、クアッドコアを発表しました。各社のブースでもこの新しいプロセッサのデモや展示が数多く見られました。

FUSION 1200 アップデート

スケーラブル SMP サーバである FUSION 1200 は、3 台の FUSION 1200 を接続した共有メモリ構成のデモが行われています。

デモンシステムの構成は以下のようになっています。

プロセッサ数: 32

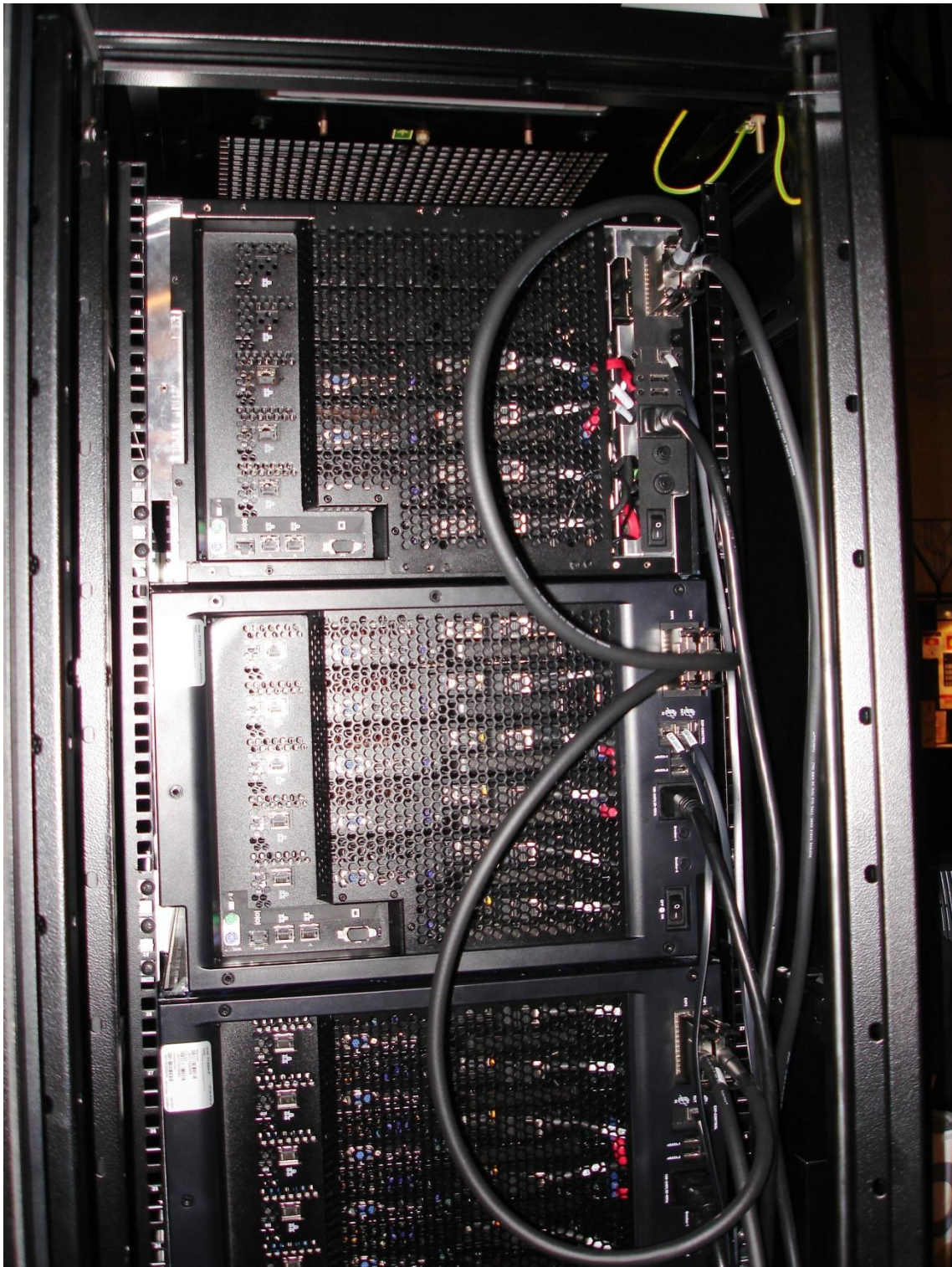
インテルクアッドコアプロセッサ: 128 コア SMP 構成

メモリ: 512GB (共有メモリ)

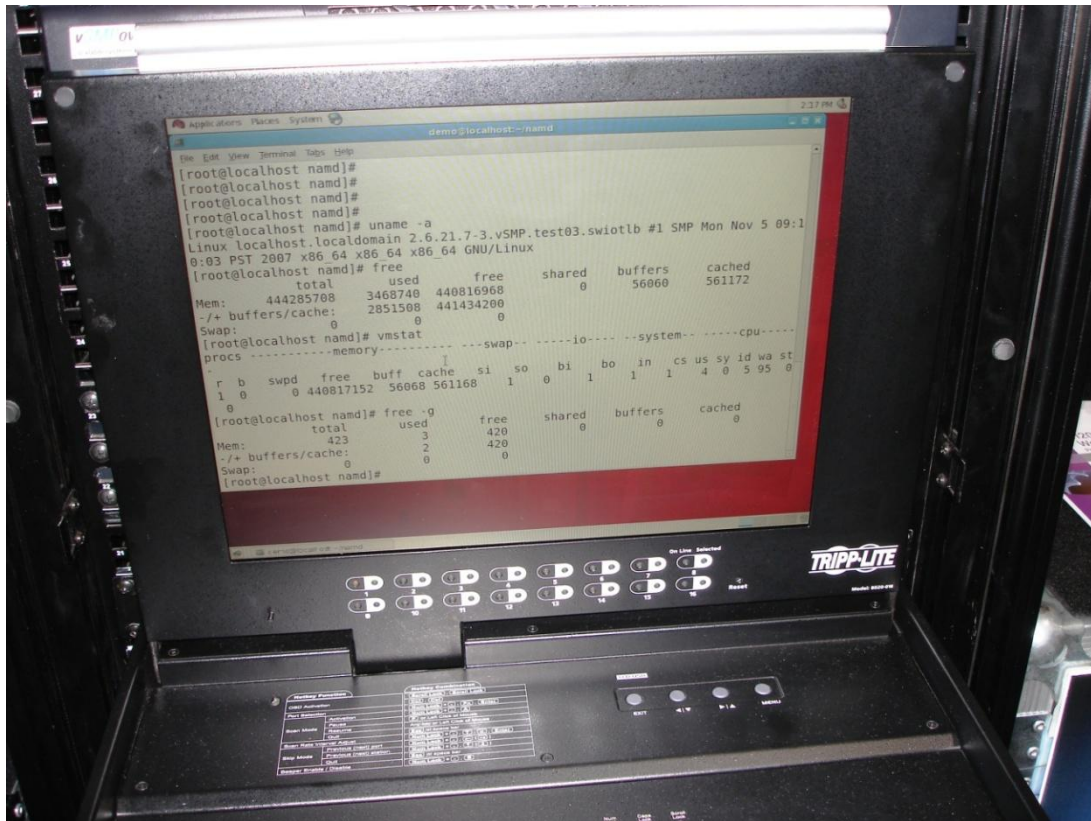
ScaleMP vSMP Foundation



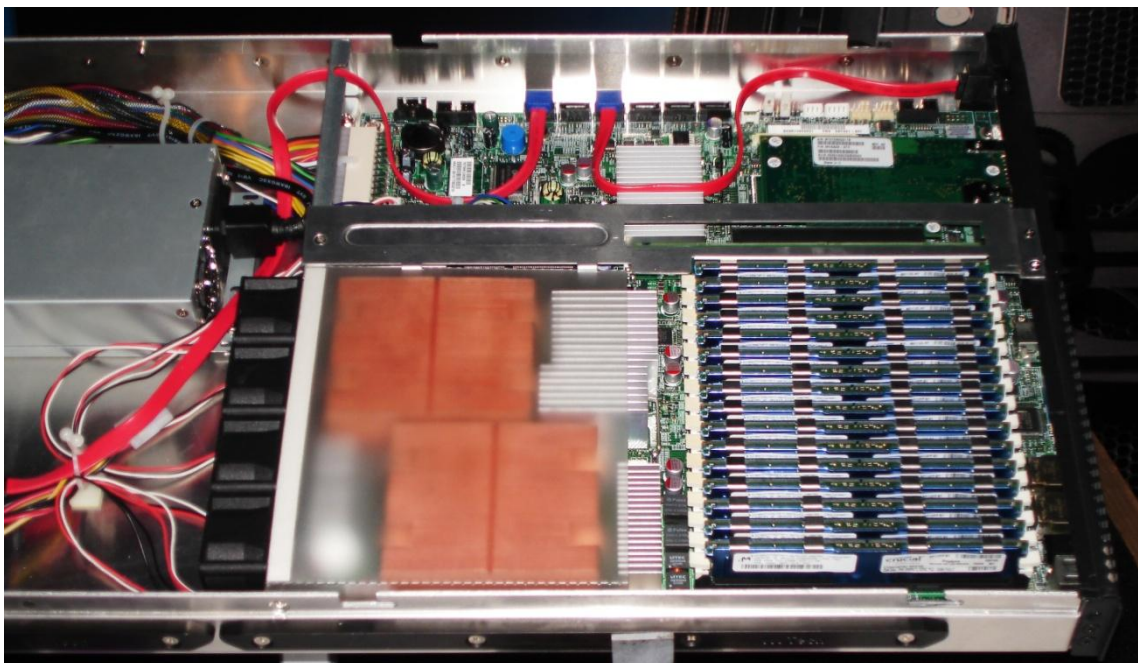
3 台の FUSION 1200 を結合して、128 コアの共有メモリスシステムを構築: デモでは、128 並列でのアプリケーション実行のデモを行っていました。OS のモニターツールなどは、これだけ多くのプロセッサを想定していないため、専用のツールでシステムの負荷状況を表示していました。



ラック背面:FUSION 1200 の複数筐体の接続を示します。非常にシンプルな構成で、クラスタシステムのように、ケーブルがノード間で煩雑に交差することはありません。

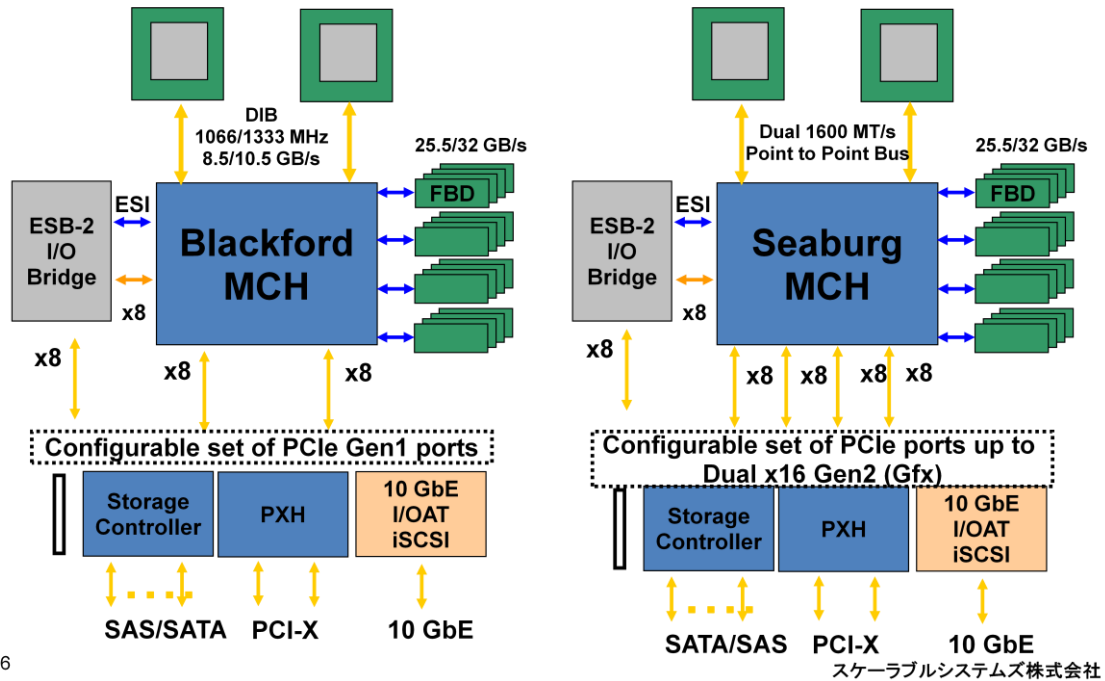


コンソールでのメモリサイズの確認を行うと 440GB のメモリ領域を確認出来ます。FUSION 1200 では、実メモリの 10%をvSMP 用の領域として確保するため、実際に搭載されたメモリの全領域を OS は利用できません。このvSMP が確保した領域を最大限に活用して、スケーラブルな共有メモリシステムを作っています。



FUSION 1200 では、従来の Shelf (通常のクラスタのノードに相当)に対して、2 倍の容量と FSB を高速化したものを

発表しています。この Shelf では、最大 128GB のメモリを搭載できるため、FUSION1200 の一筐体で、768GB のメモリまで拡張出来ます。(従来の Shelf は、32GB が最大容量)



36

また、現在は 32 プロセッサ (128 コア) が最大構成となりますが、これは 48 プロセッサ (192 コア) に来年には拡張される予定です。また、FUSION 1200 でのハイエンドグラフィックスのサポートも計画されており、将来は、大規模な共有メモリ空間を利用して、大規模な並列化や可視化処理も可能となる予定です。

スケーラブルシステムズ株式会社は、この Ciara/VXTECH のブースでもパートナーとして、会社が紹介されています。

NEXXUS 4000 アップデート

NEXXUS については、幾つかの新機能の追加が紹介されています。

- クアッドコア インテル® Xeon® プロセッサ 5400 番台のサポート
- ハイエンドグラフィックス搭載ブレードのサポート
- X38ML のサポート

特に、ハイエンドグラフィックスのサポートは、「パーソナルクラスタ」の機能拡張として、要望が多かったのですが、このハイエンドグラフィックスもサポートされました。

また、現在、WindowsCCS のバンドルもプロモーションされており、ブースでは実際に WindowsCCS を搭載した 32 コアモデルの価格が紹介されています。

NEXXUSについては、NEXXUS G2と呼ぶ第 2 世代の「スケーラブルサーバ」のプランも NDA(守秘義務契約)ベースでは紹介されています。「パーソナルクラスタ」として、確かな位置を築いた NEXXUS は、NEXXUS G2 で新しい高性能システムとして、今後紹介できることになります。



また、インテルの高密度 HPC 向けマザーボード X38ML もサポートされました。



- Wolfdale / Yorkfield / Conroe / Kentsfield processor (35W, 65W, 130W)をサポート
- X38 chipset
 - 1333 MHz Frontside Bus
 - FSB Overclock のサポート(1600 又は 2000MHz)
- Single PCI-Express x16 connector
- ハイエンドグラフィックスもサポート
- 4 DIMMs (最大 16GB) DDR2 667/800メモリ
- Integrated 4 port SATA 3.0Gb/s with RAID 0 and 1
- Integrated graphics
- Integrated BMC supporting IPMI 2.0
- Integrated Dual Gigabit Ethernet (Zoar)
- Custom 5.9" x13" form factor

X38ML などについては、【技術動向レポート】[High Density HPC Solutions](#) をご参照ください。

【技術動向レポート】[High Density HPC Solutions](#)



多くの HPC サイトで問題になっている消費電力とシステムの冷却の問題を含めて、IT インフラとしての高密度実装が可能な HPC プラットフォームについて、その性能も含めて検討した資料です。HPC プラットフォームは、単に高密度実装が可能なだけでなく、性能も非常に重要になりますが、性能に関する情報も含めて、これらのプラットフォームの動向を示しています。



〒102-0083 東京都千代田区麹町 3-5-2 BUREX 麹町 8F

電話:03-5875-4718 FAX:03-3237-7612 www.sstc.co.jp

スケーラブルシステム株式会社では、IT 技術と HPC システムに関する様々な調査レポートを発行しています。
社名、製品名などは、一般に各社の商標または登録商標です。

Copyright Scalable Systems Co., Ltd. , 2007. Unauthorized use is strictly forbidden.

無断での引用、転載を禁じます。 平成 19 年 11 月 26 日