

日本HPのHPC向け PCクラスターへの取り組み

根本 雅樹
 テクニカルセールスサポート統括本部
 日本ヒューレット・パッカード株式会社
 2008.7.25
 ～ PC クラスターワークショップ in 京都 ～

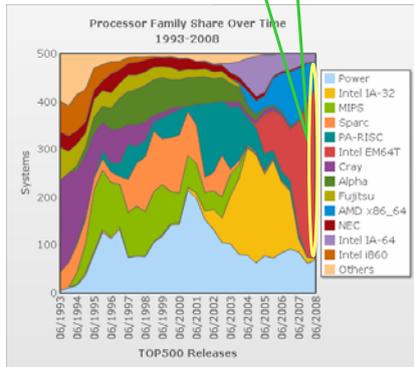
<http://www.hp.com/jp/hpc>



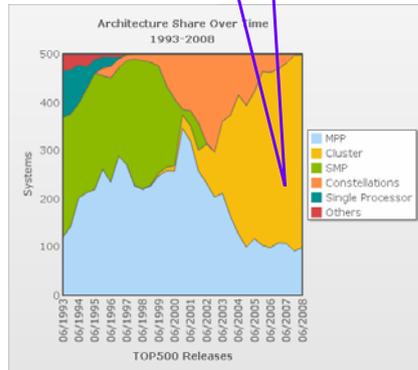
© 2006 Hewlett-Packard Development Company, L.P.
 The information contained herein is subject to change without notice.

TOP500で見るHPCの動向 (TOP500 2008年6月発表)

X86が急伸



クラスターが主流



HPのHPC製品開発方針

1. オープンで最先端の技術を組合せ、コスト対効果、拡張性の高い製品の開発
2. 全世界に製品を供給可能な保守体制の整備
3. 検証済みで高度にエンジニアリングされ、安定稼動、短納期を実現するターンキーシステム
4. 消費電力、設置面積、発熱量に配慮した設計

性能やコストのみならず、安定性、信頼性を重視されるお客様にも安心して導入いただける製品をご提供

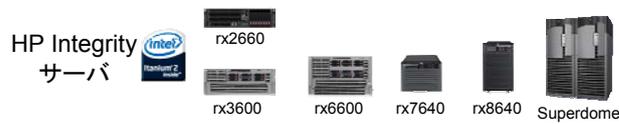


3 2008/8/8

HPC プラットフォーム

～ HPは標準プロセッサを選択：メニー・コアを牽引

- 豊富な選択肢
- 高性能
- 管理容易



4 2008/8/8

HP ProLiant DLサーバ 豊富なラインアップ (2P)

http://h50146.www5.hp.com/products/servers/proliant/dl_lineup2w.html

2P/8C	2P/4C	2P/8C	2P/8C	2P/8C	2P/8C
オンラインで購入	オンラインで購入	オンラインで購入	オンラインで購入	オンラインで購入	オンラインで購入
> DL140 G3	> DL145 G3	> DL160 G5	> DL165 G5	> DL180	> DL180 G5
30人~100人					
<p>コンピューティング処理に必要な機能をクロスアップし、購入コストを抑えたシングルUサーバです。</p> <p>高速プロセッサ搭載、コンピューティング性能にこだわらるHPC環境に最適な1Uサーバです。</p> <p>低コストと大容量ストレージを提供する2Uサーバです。</p> <p>低コストで、最大12HDDのディスク格納容量を持つ、2Uサーバです。</p>					
クアッドコア または デュアルコア インテル® Xeon®	AMD Opteron™ (デュアルコア)	クアッドコア または デュアルコア インテル® Xeon®	クアッドコア または デュアルコア インテル® Xeon®	クアッドコア または デュアルコア インテル® Xeon®	クアッドコア または デュアルコア インテル® Xeon®
32GB		32GB		16GB	16GB
2P/8C	2P/4C	2P/8C	2P/8C	2P/4C	2P/8C
オンラインで購入	オンラインで購入	オンラインで購入	オンラインで購入	オンラインで購入	オンラインで購入
> DL260 G5	> DL265	> DL265 G5	> DL260 G5	> DL295 G2	> DL295 G5
100人~200人			100人~200人		
<p>管理者からの高い評価を得ている、全世界でベストセラーのラインDL360と、DL360をベースにAMD Opteron™を搭載したDL365、1Uサイズにテクノロジーを凝縮し、優れた管理性、メンテナンス性を満喫です。</p> <p>データセンターでのアプリケーション層を支える、2UサーバのベストセラーDL380とAMDプロセッサ搭載のDL385、パワフルな性能に充実した管理機能をプラスし、クラスターサーバのヘッドとしても最適です。</p>					
クアッドコア または デュアルコア インテル® Xeon®	AMD Opteron™ (デュアルコア)	AMD Opteron™ (クアッドコア)	クアッドコア または デュアルコア インテル® Xeon®	AMD Opteron™ (デュアルコア)	AMD Opteron™ (クアッドコア)
32GB			64GB		32GB



5 2008/8



HP ProLiant DLサーバ 豊富なラインアップ (4P,8P)

http://h50146.www5.hp.com/products/servers/proliant/dl_lineup4w.html

4P/4C、4P/8C、4P/16C	4P/8C	4P/16C	8P/32C
> DL580 G5	> DL585 G2	> DL585 G5 New!	> DL785 G5 New!
200人以上			
<p>全世界でもっとも売れている4UサーバDL580と、圧倒的なパフォーマンスを誇るAMD Opteron™搭載のDL585。マルチプロセッシングの性能を必要とするWebアプリケーション、大規模メールサーバ、大容量データベースに最適なエンタープライズ向けサーバです。</p> <p>最大8基のクアッドコアAMD Opteron、最大256GBのメモリ、および11個のPCI-Expressをサポートする非常に高い拡張性を備えるDL785 G5は、日々高度化するエンタープライズクラスのデータベース、サーバ統合、仮想化環境に優れた柔軟性、拡張性、およびパフォーマンスを提供します。</p>			
クアッドコア または デュアルコア インテル® Xeon®	AMD Opteron™ (デュアルコア)	AMD Opteron™ (クアッドコア)	AMD Opteron™ (クアッドコア)
256GB	128GB	128GB	256GB



6 2008/8/8



HP Integrity サーバ 豊富なラインアップ

<http://h50146.www5.hp.com/products/servers/integrity/hardware/>



7 2008/8/8



HP BladeSystem c-Class サーバブレード製品群



8 2008/8/8



製品スペック

HP ProLiant BL2x220c Generation 5



サーバノードあたりのスペック

※サーバノードAとBを別スペックで構成可能

プロセッサ	最大2基 <ul style="list-style-type: none">クアッドコア インテル® Xeon® processor 5400ファミリデュアルコア インテル® Xeon® processor 5200ファミリ
メモリ	最大16GB (4スロット)
チップセット	Intel®-5100 chipset
内蔵ドライブ	1ベイ ノンホットプラグSATA 120GB Serial ATA
内蔵NIC	標準1GbE 2ポート
拡張スロット	PCIeメザニンスロット 1個 (x8, Type I)
遠隔管理	Integrated Lights Out 2 Standard Blade Edition
標準保証	1年間のパーツ保証・作業保証・オンサイト サポート

9 2008/8/8

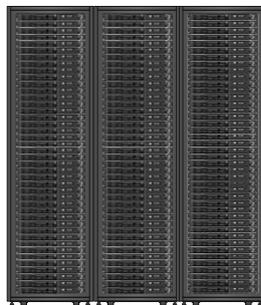


劇的な設置スペース削減

256CPU/1024コアが1ラック: ラックマウントサーバの**3倍密度**

設置スペースの限られたお客様に**最高の集約密度**を提供

HPC市場、Web2.0市場に特化した超高密度ブレード



HP ProLiant
DL360 G5

126台/252CPU/1008コア



HP ProLiant
BL460c

128台/256CPU/1024コア



HP ProLiant
BL2x220c G5

128ノード/256CPU/1024コア

10 2009/8/7

発熱・電力の抑止・コントロール

HP自身が開発したアクティブ冷却ファン

HPで独自開発。20個の特許出願中



高性能

1台のファンで4台の1Uサーバを冷却できる能力

高寿命

同一冷却性能のファンと比較して

単体寿命が約4倍

低消費電力

同一冷却性能のファンと比較して

66%消費電力軽減

低騒音

同一冷却性能のファンと比較して

50%騒音抑止

c-Class ブレードエンクロージャに
最大10個搭載可能



11 2008/8/8

サーバ消費電力の内訳

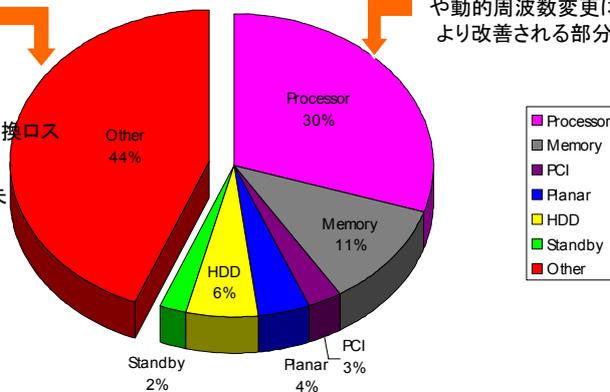
- プロセッサの消費電力が問題視されがちですが、サーバの消費電力を低減させるには、その他の部分を抑える機構が必要です。

HP BladeSystem導入により
改善される部分

低電圧プロセッサ導入
や動的周波数変更により
改善される部分

OTHERとは？

- ・パワーサプライによる変換ロス
- ・冷却ファンの効率性
- ・電力供給ラインでの損失



12 2008/8/8

c7000とc3000

c7000

- サーバ台数が多い(8台以上～数千台)
- I/O本数が沢山必要
- ネットワークなどの高い冗長化を求める
- データセンター、マシンルームに設置



c3000

- サーバ台数が少ない(8台以下)
- AC100Vで利用をしたい
- 手軽にブレードをはじめたい
- I/O本数があまり必要ない

13 2008/8/8



「コンパクト第3世代ブレード」 HP BladeSystem c3000 前面

- ・ 8個のブレードデバイスベイ
- ・ サーバブレードだけでなく、テープブレード・ストレージブレードも挿入可能



工事不要
壁コンセント対応
通常のAC100V
壁コンセントが利用可能



※電源容量はご注意ください



Onboard
Administrator
(管理モジュール)

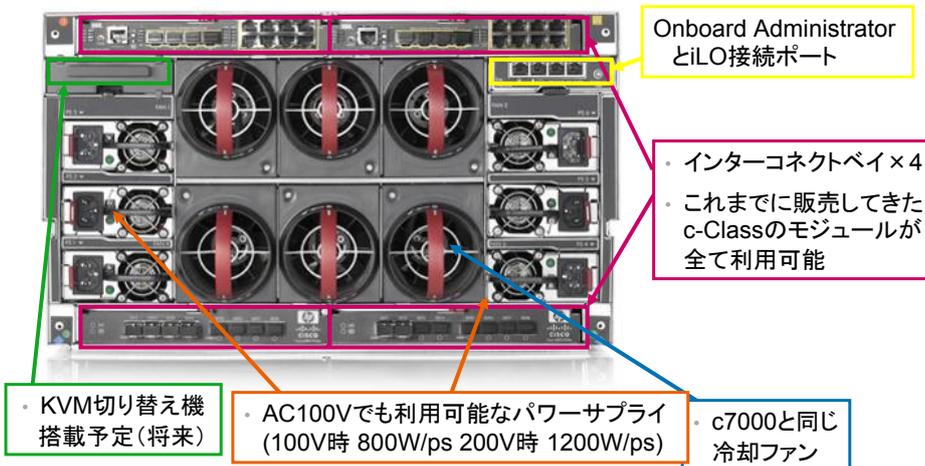
サーバの稼働状況・修理方法
まで教えてくれる
インサイトディスプレイ

複数サーバで同時に利用できる
DVDドライブ

14 2008/8/8



「コンパクト第3世代ブレード」 HP BladeSystem c3000 背面



15 2008/8/8

HP Confidential



Blade-Everything



16 2008/8/8



HP CMU (クラスタ管理ユーティリティ)

ScoreにCMUの管理機能を追加

HPCクラスタシステムの構築・管理を容易にするツール

- HPCクラスタにおいて実績のある管理ツールを提供します
 - ・ 全世界で150システム以上の納入実績



主な機能

演算ノードの一括制御
電源ON/OFF、ブート、シャットダウン、リブート

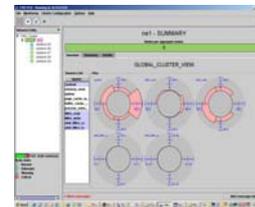
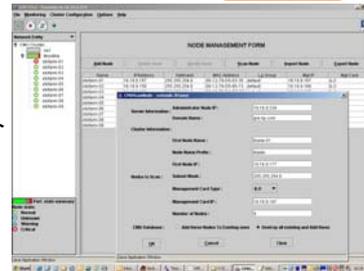
演算ノードのシステム管理をGUIで一元化

1. 論理グループによる管理
2. ノード状態の表示
3. コンソールブロードキャスト機能
 - 全演算ノードへのコマンド一括入力
4. ノード監視とイベント処理

・ クローン機能によるシステムディスクの一括複製

・ CMUモニタ: 演算ノードの負荷状況等をグラフ表示

・ UPS/自動シャットダウン(オプション)



17 2008/8/8

クラスタシステム構築 広範囲な選択肢

クラスタシステム構成要素

プロセッサアーキテクチャ

Xeon, Opteron, Itanium2

オペレーティングシステム

Linux, HP-UX, Windows

設置密度

ラックマウント型、ブレード型

インターコネク

Infiniband, Gigabit Ethernet

MPIライブラリ

HP-MPI, MPICH

システム管理ソフトウェア

商用管理ツール、Free Ware

インテグレーション形態

工場組み立て、カスタマイズ

どの組み合わせが最適か
ノード数が増えても単純な管理ができるか
トラブル時の短時間復旧ができるか
クラスタを運用管理できる人がいない
想定した性能が出るか...

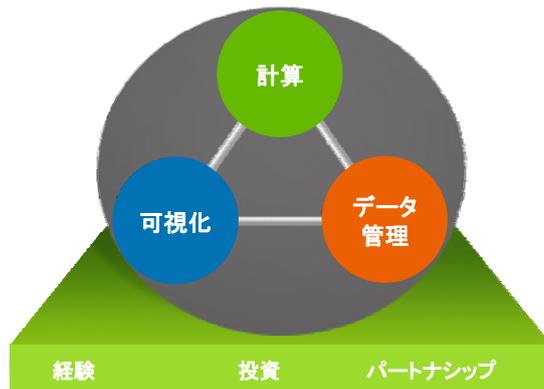


計算需要増大よりHPC向けクラスタの構成ノード数は驚異的な勢いで増加中

18 2008/8/8

完全なHPCソリューションの提供

- 標準技術をベースにした革新性
- HPCソリューションにフォーカスした広範な選択肢
- 手頃で、利用しやすいスーパーコンピューティングクラスの性能



19 2008/8/8



実運用が容易な、HPC用Linuxクラスタソリューション HP XCクラスタ



5ノードから構成可能
大規模クラスタに対応可能

パッケージ製品なので、
HPC用Linuxクラスタ運用の手間が大幅に削減！

- 業務へすばやい展開が可能
- 不具合時に迅速な対応可能
- 計算負荷の変化に伴うシステム構成の変更が容易
- ソフトウェア設定が容易
- ブート時間を短縮
- システムソフトウェア更新が容易

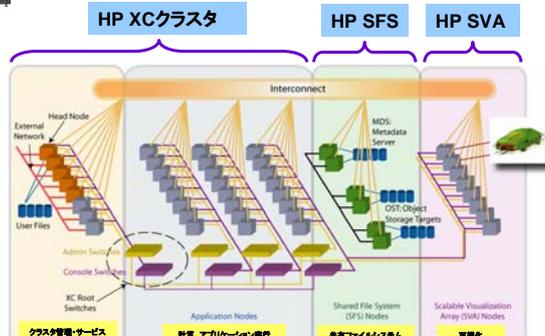
20 2008/8/8

HPがHPC向けの‘ターキー’Linuxクラスタを提供

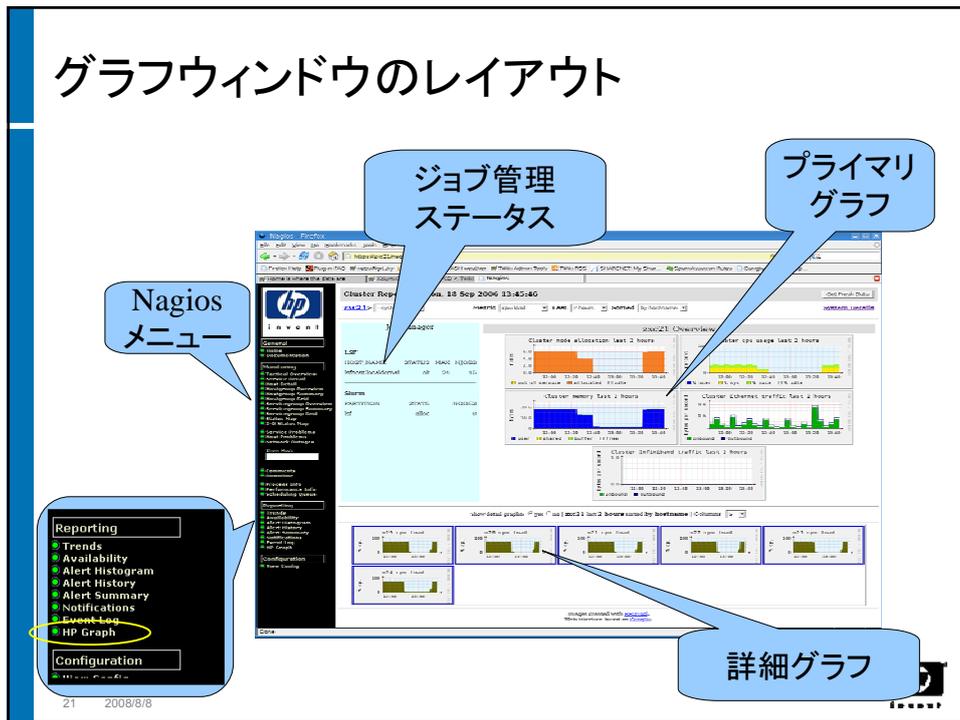
HPCでクラスタシステムを利用するために必要な機能はすべてまとめて提供・サポート

- ノード: Opteron, Xeon, Itanium2
- インターコネクト: InfiniBand, GigE, Quadrics, Myrinet
- OS/管理ツール: XCシステムソフトウェア
 - RedHat EL互換
 - LVS, LSF, SLURM, NAT
 - Kickstart, XC Config, SystemImager, Nagios, SuperMon, pdsh, MySQL
 - HP-MPI

⇒HPが標準製品としてのサポートを提供



グラフウィンドウのレイアウト

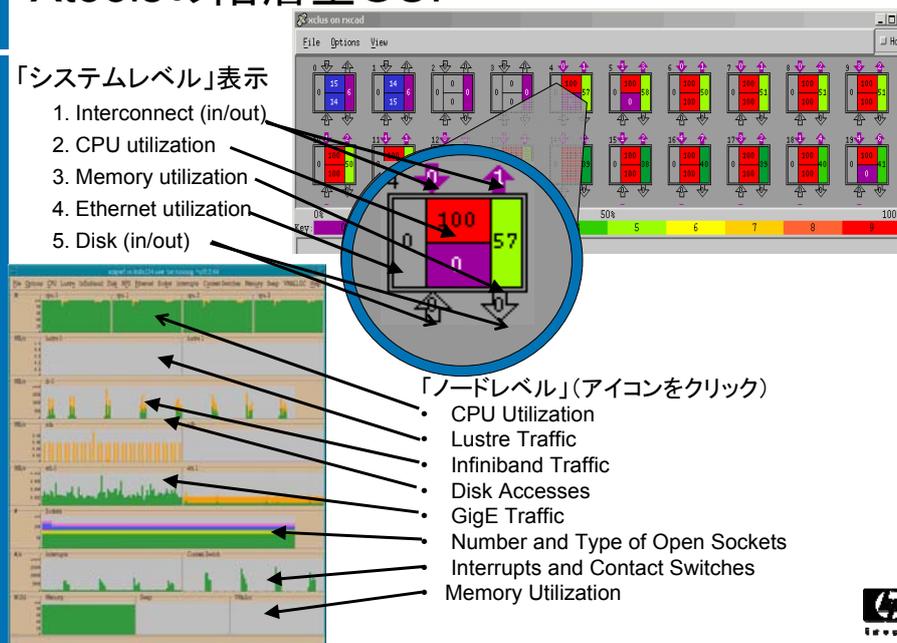


21 2008/8/8

Xtoolsの階層型GUI

「システムレベル」表示

1. Interconnect (in/out)
2. CPU utilization
3. Memory utilization
4. Ethernet utilization
5. Disk (in/out)



「ノードレベル」(アイコンをクリック)

- CPU Utilization
- Luster Traffic
- Infiniband Traffic
- Disk Accesses
- GigE Traffic
- Number and Type of Open Sockets
- Interrupts and Contact Switches
- Memory Utilization

HP XCクラスタ



256 ノード構成 (さらに大規模構成へ)
 - TATA(インド)で約1800ノード構成の実績 -

HP XC3000 Xeonベース	HP XC4000 Opteronベース	HP XC6000 Itanium2ベース
<ul style="list-style-type: none"> C-Classブレード DL160 (2CPU) InfiniBand, GigE, Myrinet 	<ul style="list-style-type: none"> C-Classブレード DL165 (2CPU) DL585 (4CPU) InfiniBand, GigE, Quadrics, Myrinet 	<ul style="list-style-type: none"> C-Classブレード rx2660 (2CPU) Quadrics, InfiniBand, GigE

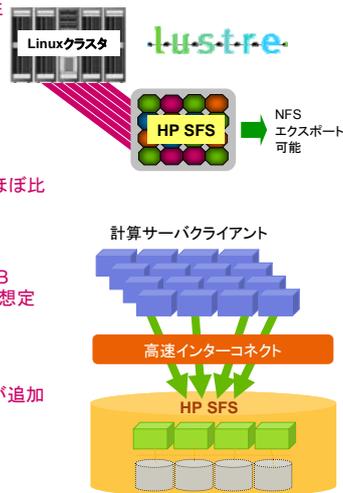
23 2008/8/8



HP StorageWorks Scalable File Share (HP SFS)

大規模Linuxクラスタのための共有ファイルシステムLustreの実装製品

- 全ノードから一貫性のあるファイルシステムアクセス
 - Lustreファイルシステム: POSIX準拠のデータの一貫性、整合性
 - すべてのノードから同じファイルを同じパス名で直接参照可能
 - ファイルシステムを単一イメージで提供
 - MPIジョブからの並列I/Oに伴う特殊なプログラム変更が不要
- スケーラブルなI/Oバンド幅
 - 200MB/s ~ 35GB/s さらに拡大可能
 - 総合I/Oスループット(実際に利用者が享受する性能)
 - I/O用ノードやストレージ装置の増設によるスケーラビリティ
 - 合計性能だけでなく、単一ファイルの処理性能もサーバー数にほぼ比例して増加
 - 高速インターコネクを経由のデータ転送
- スケーラブルな容量: 最大数百TB
 - I/O用ノード、ストレージ装置の増設でSFS一式あたり最大512TB
 - スケーラブルなアクセス: 数千クライアントからの同時アクセスを想定
- サービスの継続性
 - 障害に強い構成、縮退運転機能(フェールオーバー機能)
 - RAID5、RAID5+1、RAID6、RAID6+1サポート
 - 自動アラート機能、診断機能など安定運用のための機能をHPが追加
- 用途に合わせたストレージ構成
 - 価格性能重視: StorageWorks SFS20
 - 高信頼性重視: StorageWorks EVA
- HP製品としての保守体制
 - オープンソースLustreファイルシステムを製品としてHPが保守

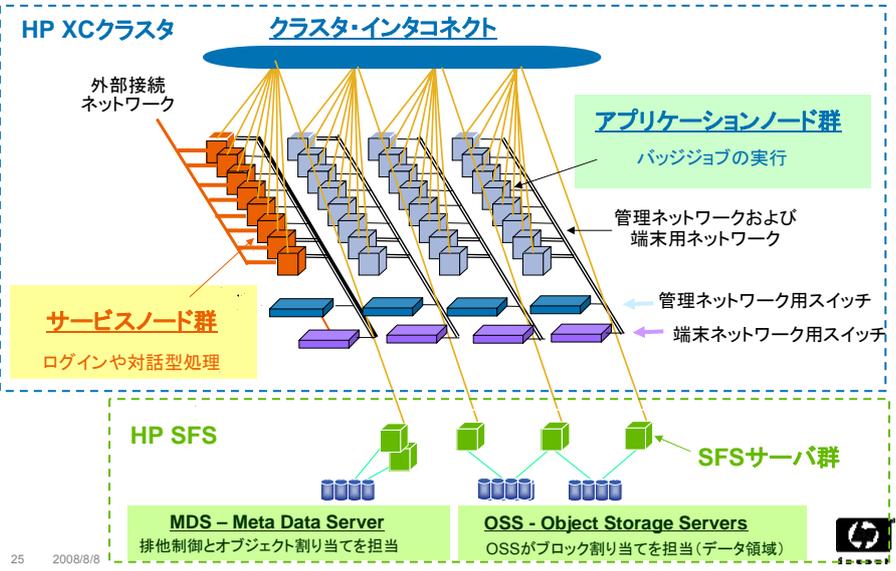


24 2008/8/8



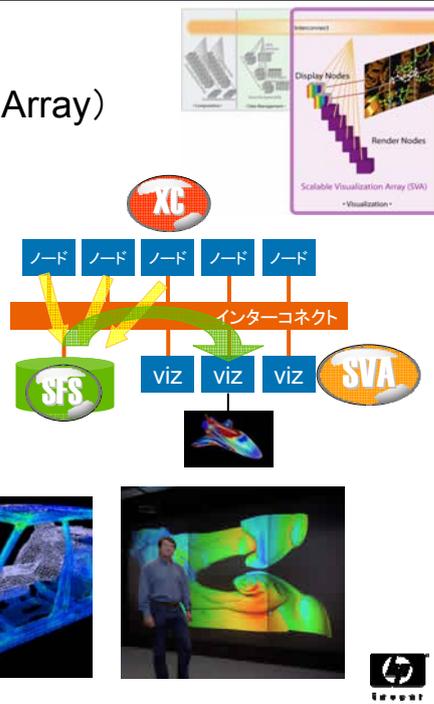
XCクラスタとSFS

XCクラスタの共有ストレージには、高性能・大容量・高信頼性のSFSファイルサーバが最適！



HP SVA (HP Scalable Visualization Array)

- ✓ クラスタでの高速・大規模・並列可視化が容易
- ✓ HP SFSによる高速・大容量データ共有
 - ✓ シミュレーション計算 ⇒ 可視化
 - ✓ レンダリングノード ⇒ ディスプレイノード
- ✓ 主要な可視化アプリケーションに対応
 - ✓ EnSight
 - ✓ AVS 等
 - ✓ 可視化アプリケーションのランチャー提供
- ✓ HP XCクラスタによる運用管理
 - ✓ Nagiosによる稼働状況監視
 - ✓ SLURMによる資源管理
 - ✓ LSFジョブスケジューラ 等
- ✓ ベンダ製品としての保守体制



HP-MPI: 性能、移植互換性に優れ、多くのISV様がサポート表明しています

**XCクラスタに
標準装備**

- ・MPI 2.0準拠
- ・HPが研究開発、販売、保守
- ・ASCI-Q, PNNL 等の世界最大級のクラスタ計算機環境での稼働実績
- ・TCP/IP サポート
- ・PGI, Intel コンパイラ サポート
- ・豊富なプロファイリング機能
- ・TotalViewとの連動
- ・HP-UX、Linuxサポート
- ・Windowsサポート
 - CCS, XP



27 2008/8/8

HP-MPIは主要なISVのアプリケーションソフトウェアに採用されています



Note: end users of these ISV applications on industry-standard clusters may receive HP-MPI transparently from these ISVs as part of the ISV's application distribution. "Transparent" means no cost and no FlexLM licensing requirement. The first point of contact for support is the ISV.
All ISVs listed may distribute HP-MPI on Linux & HP-UX. Those with

Example ISV web sites listing platforms supported with HP-MPI:
www.abaqus.com/support/v66/v66_mpi.html
www.ansys.com/services/ss-interconnects.htm

* have also signed up for Windows CCS.
28 2008/8/8



HPCソリューションセンター

- ・ HPCスペシャリストがお客様の検証作業をご支援致します。
 - 新製品、新技術の検証やベンチマーク
 - ポーティング・チューニング
 - お客様の課題や問題の解決

弊社HPCポータルサイトからご予約いただけます。

<http://www.hp.com/jp/hpc/>

その他、お問い合わせは

Mailto: hp-scs-collabo-center@hp.com



設備:

HP XCクラスタ、HP SFS
HP BladeSystem
HP ProLiant、HP Integrity
各種ストレージ
ワークステーション 等々



29 2008/8/8

まとめ

- ・ HPCソルバーのプラットフォームとして、クラスタが主流
 - HPC用途クラスタは性能のみならず、安定稼働が重要
- ・ 業界標準技術の活用
 - IntelやAMDの性能ロードマップ及びHPCシステムアーキテクチャ
 - ISVアプリケーションと業界標準技術のマッチング
- ・ HP製品で稼働するCAE・EDAアプリケーションの完全なポートフォリオ
 - 多くのCAE・EDA ISV様と技術やビジネスについて深いパートナーシップ
 - 業界標準クラスタについてISV様のアプリにHP-MPIを提供し、最適化
- ・ HPCに対する統合ソリューション提供
 - 計算、IO、可視化
 - XCクラスタや、統合環境、パートナー様のソリューションを提供
 - ベンチマーク、導入、保守サポートなどを提供
- ・ HPCへのブレードの適用
 - 性能、設置場所、電力、冷却
 - XCクラスタとしても提供可能
- ・ HPCにおけるお客様のROIを最適化

- ・ HPCに関する情報
 - www.hp.com/jp/HPC



30 2008/8/8

