



High Performance and Productivity HPCシステムの課題と挑戦

スケーラブルシステムズ株式会社

DIRECTION

NORTHEAST EAST SOUTHEAST SOUTH SOUTHWEST WEST

HPCシステムの課題と挑戦

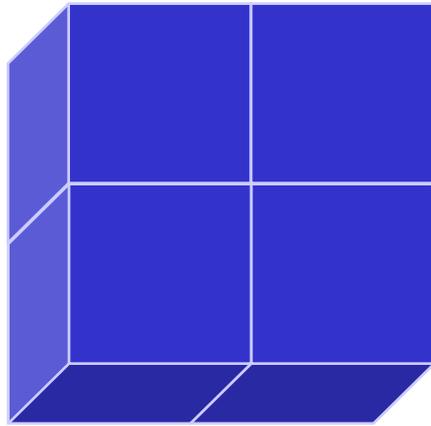


- HPCマーケットの動向とHPCプラットフォームの課題
 - クラスタ .vs. SMPシステム
 - TCOの問題
- HPCシステムの考察
 - ～ 製品事例によるHP²Cシステム提案
 - パーソナルクラスタ
 - スケーラブルx86システム
- まとめとして

HPCプラットフォームの変遷

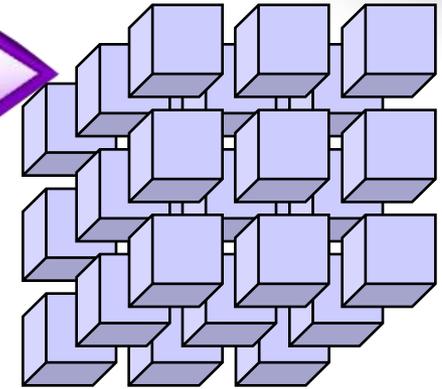


メインフレーム
スーパーコンピュータ



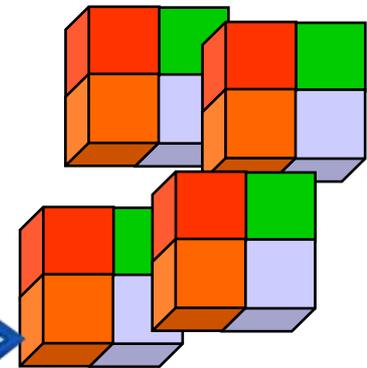
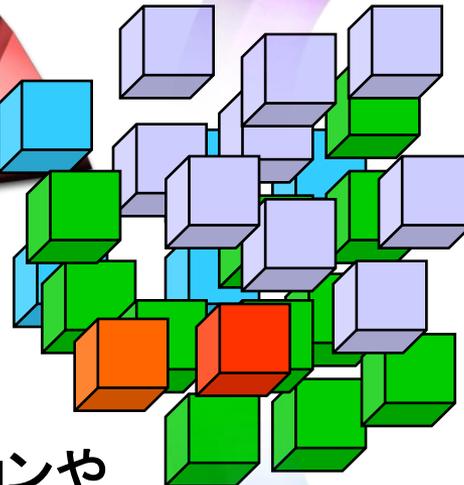
(リソースの集中と管理)

■ クラスタによる仮想コンピュータ



分散したリソースの管理

■ ワークステーションや
サーバによる分散処理



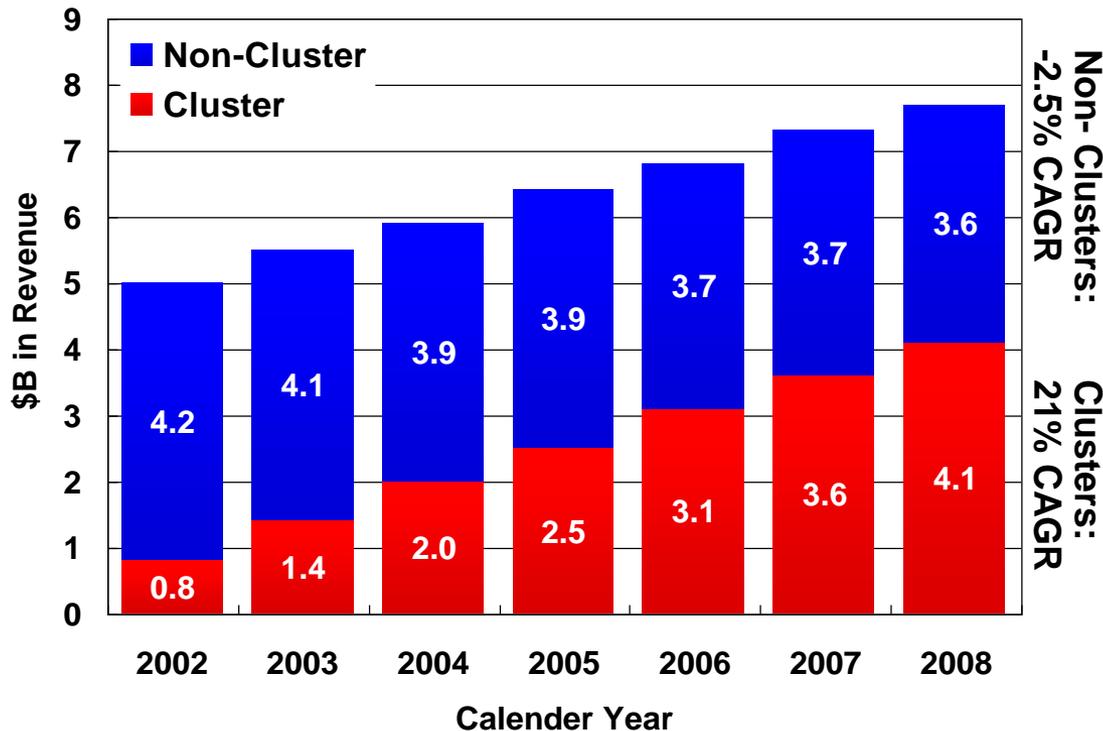
■ 仮想化によるサーバ・コンソリデーション

HPCマーケット



HPCマーケットでのx86サーバの売り上げが、5.9% CAGRであるのに対して、21.6% CAGRの伸びを示している (IDC)

Worldwide High Performance Computing Market



部門向け (Departmental HPC、64ノード以下) クラスタシステムが、クラスタ導入の牽引 (ユニット、売り上げとも)

(補足)
クラスタの出荷数の90%以上は、\$250K以下の価格レンジ
平均のクラスタのプロセッサ数は、8-16

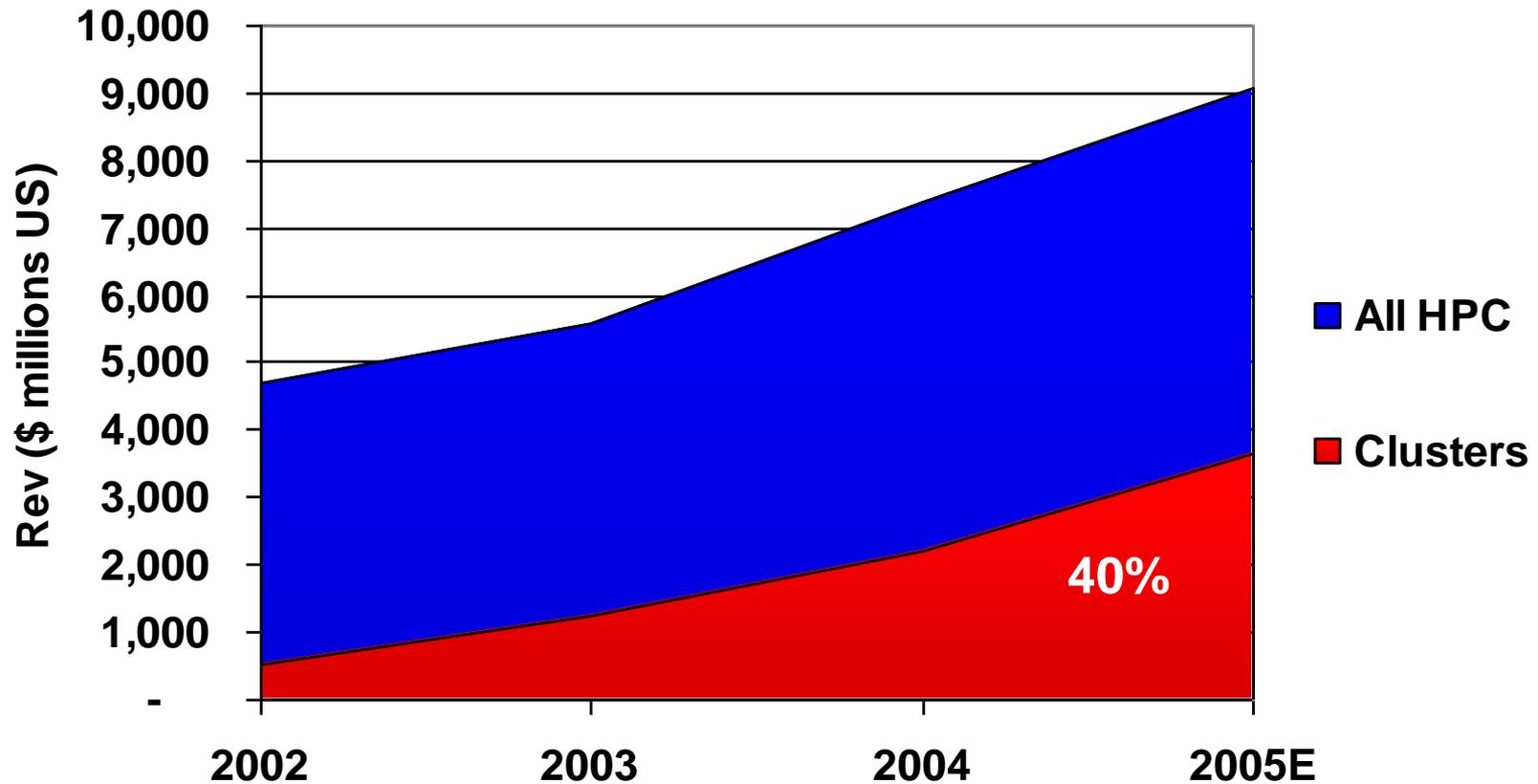
Clusters Accounted for 33% of Revenue in '04

WW HPC Server IDC Forecast



†IDC MCS: The Cluster Revolution in Technical Computing Markets (2006), IDC, Feb 2006, #

HPC Market

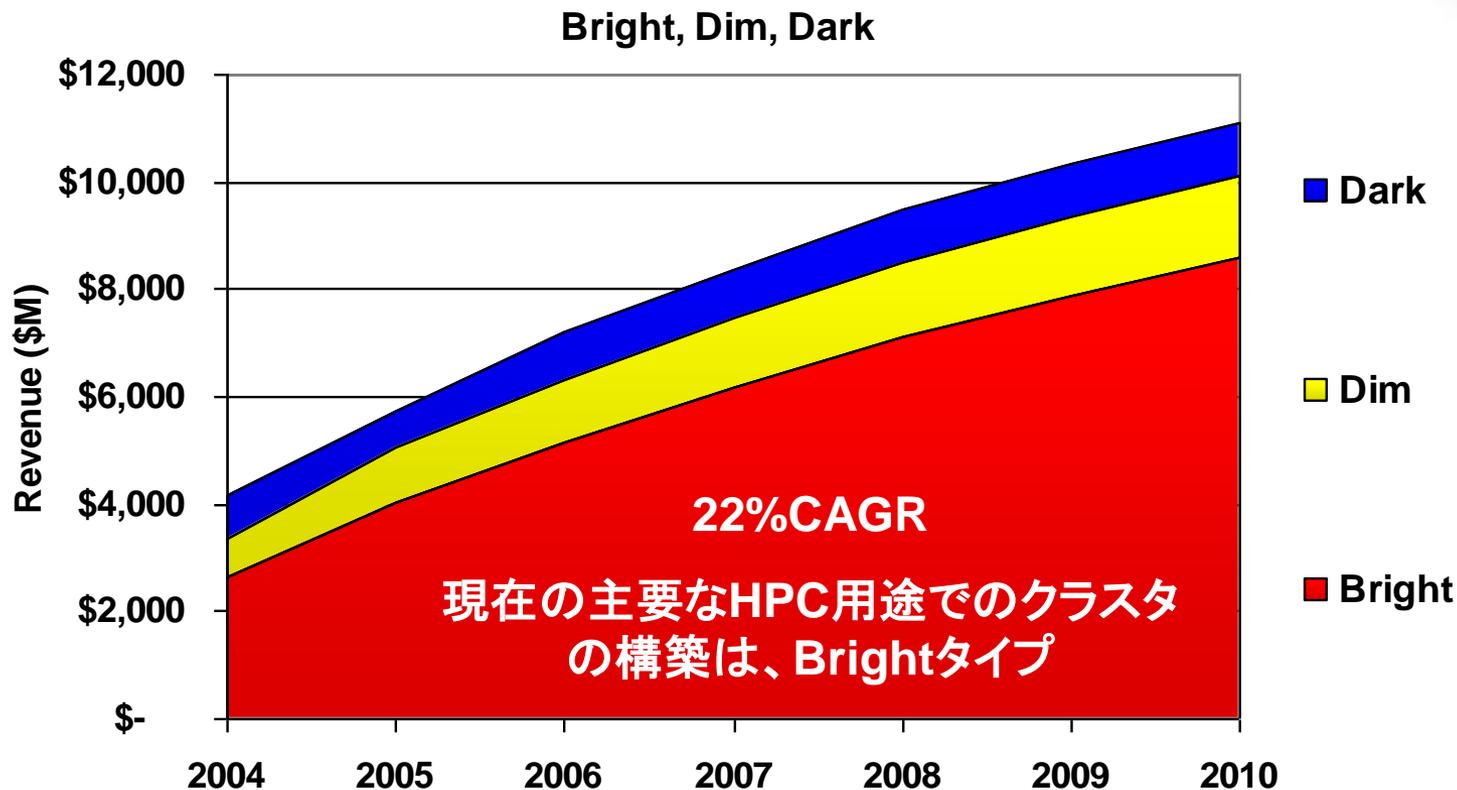


→Question:残りの60%は?

WW HPC Server IDC Forecast



†IDC MCS: The Cluster Revolution in Technical Computing Markets (2006), IDC, Feb 2006, #



Bright Clusters: ベンダーがクラスタを構築して販売し、ノード単位でシステムをカウントするのではなく、トータルなシステムとしてカウントする→究極のBright Clusterは？

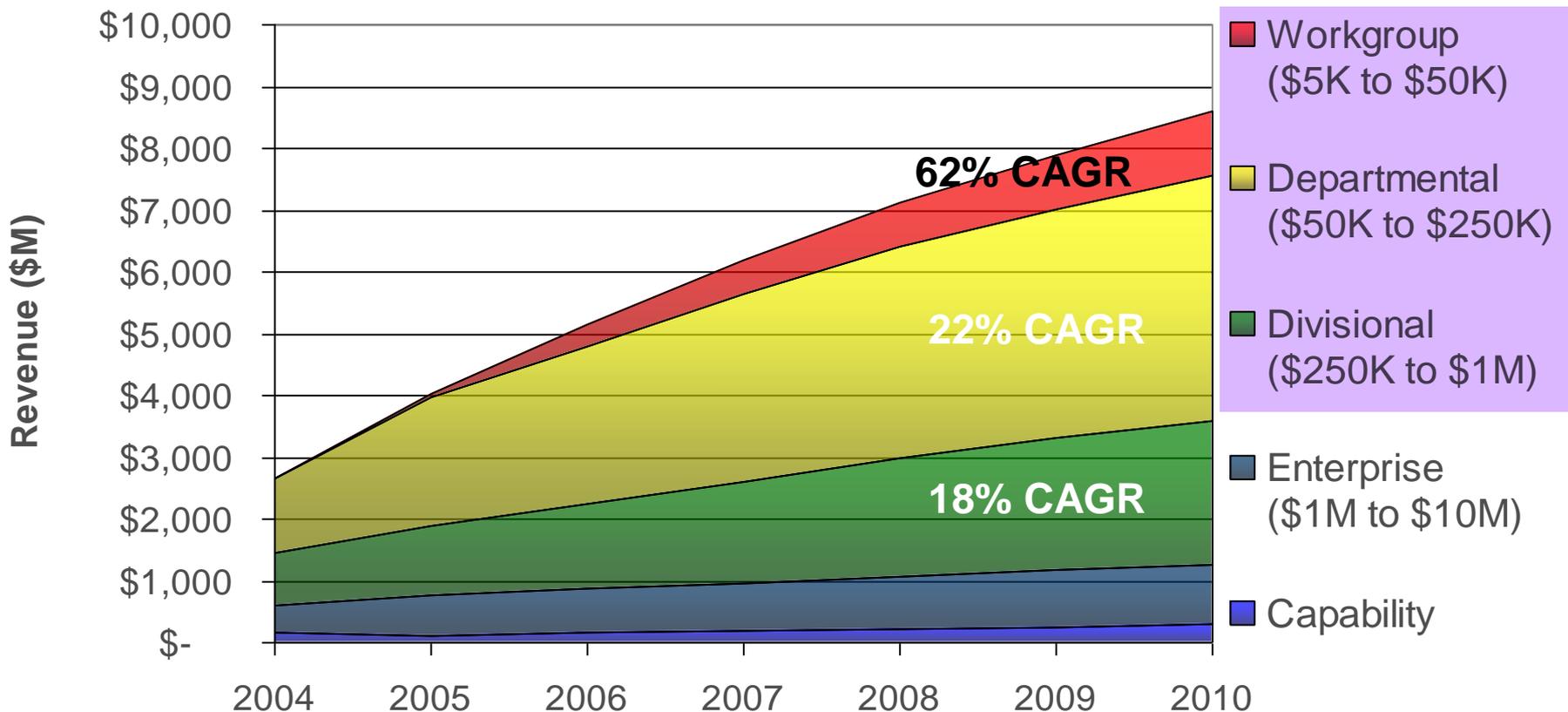
Dim Clusters: ユーザがノードを個別に購入し、クラスタを構築する

WW HPC Server IDC Forecast



†IDC MCS: The Cluster Revolution in Technical Computing Markets (2006), IDC, Feb 2006, #

Cluster Forecast by Competitive Segments (\$M)



HPC マーケットでのビジネス



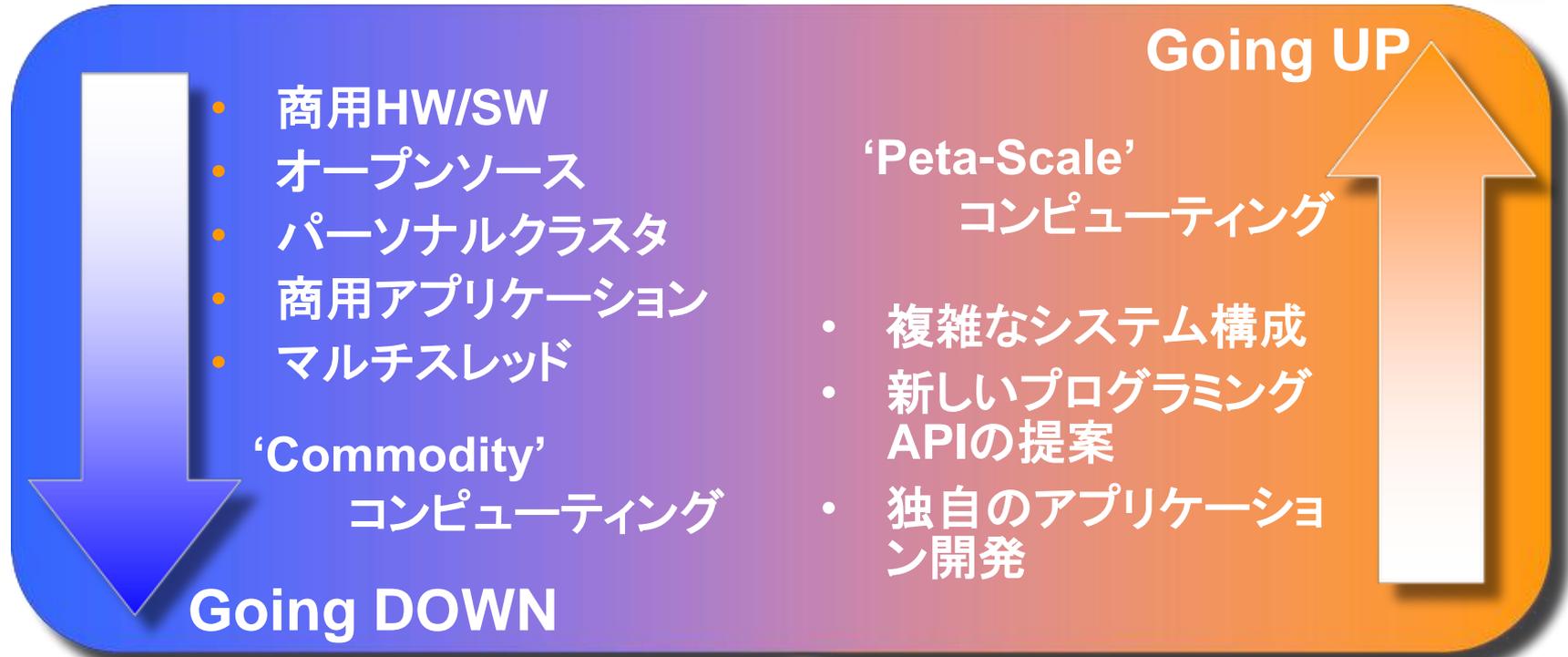
- HPC向けクラスタの伸びは堅調（91.3% CAGR）
- ‘Departmental’と‘Divisional’に分類されるマーケットでは、それぞれ、22%と18%の成長を予測
- ‘Workgroup’は、62% CAGR（2005年末から1010年の間）を予想
- Bright Clusters（OEMがインテグレーションを行い、工場出荷時に既に組みあがっているクラスタ）は、22% CAGR を予想（各社がそのような計画を持つ）

HPCプラットフォームの課題



- HPCシステムとしては、クラスタシステムが一般化しているが、問題も顕在化している
- SMPシステムの利点はOEM及びユーザも理解しているが、また、SMPシステムの開発、販売、導入には問題がある

HPCの二極分化



HPCの二極分化

HPCシステムの課題

- 基盤技術やコアのITテクノロジーの共通化を図りながら、この極端に分極化したHPCシステムへの対応を図ることが必要となる。

Going UP

‘Peta-Scale’
コンピューティング

- 複雑なシステム構成
- 新しいプログラミングAPIの提案
- 独自のアプリケーション開発

- 商用HW/SW
- オープンソース
- パーソナルクラスタ
- 商用アプリケーション
- マルチスレッド

‘Commodity’
コンピューティング

Going DOWN

HPCシステムの問題

- HPCで要求されるシステムの仕様が、大きく分極化し、この双方を一つのシステム・アーキテクチャで実現するのは、技術的に可能だとしても、経済性や生産性の点で問題がある。

HPCの定義

ペタFLOPS級‘

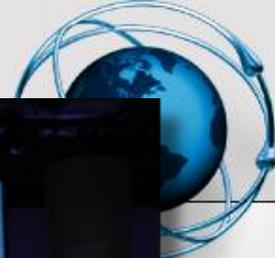
スーパーコンピュータ

- ピーク性能ではなく、アプリケーションの実効性能として、ペタFLOPSを超えるスーパーコンピュータ(ロレンス・リバモア国立研究所のHorst Simon博士の定義)



ハイパフォーマンスコンピューティング

- ハードウェア、ソフトウェア、開発環境など様々な技術を統合して、従来は解析出来なかった問題を十分な経済性をもって、解決すること



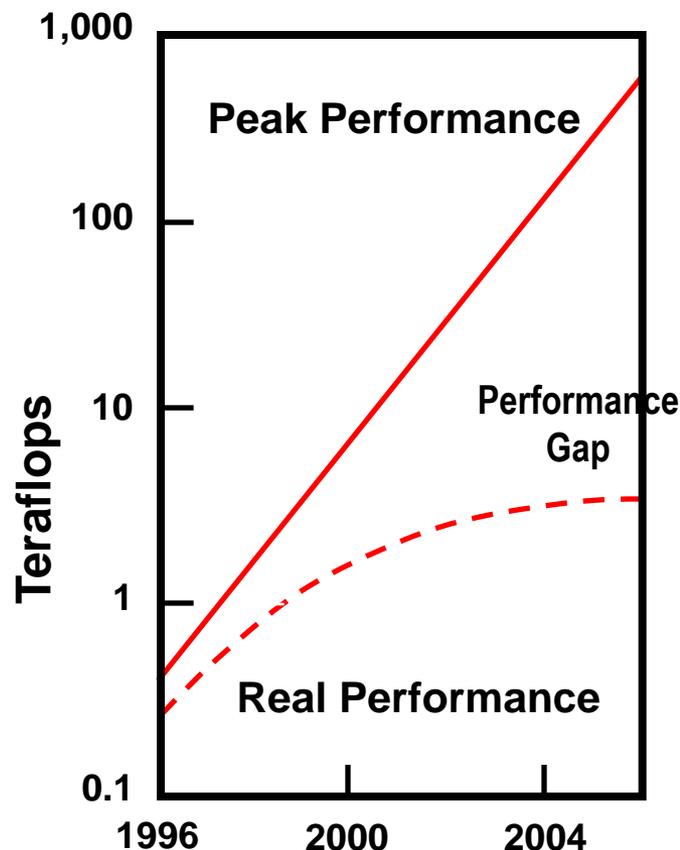
‘Peta-Scale’ コンピューティング

- 複雑なシステム構成
- 新しいプログラミングAPIの提案
- 独自のアプリケーション開発

• ‘Peta-Scale’コンピューティング

- 求められる基本技術と現在のHPCの主要マーケットでの要求はあまりにも差が大きい
- ‘Commodity’のマイクロプロセッサではなく、独自のプロセッサを開発中
- ‘複雑さ’の克服が重要

性能ギャップの拡大



NERSC User Group Meeting June 24-25, 2004
Osni Marques and Tony Drummond
Lawrence Berkeley National Laboratory

- ピーク性能の大幅な向上
 - 1990年代は、性能の向上は、 10^2 のオーダーでしたが、2000年代になると 10^3 のオーダーで性能は向上しています。
- しかし...
 - 多くの科学技術計算用途のアプリケーションのピーク性能に対する実効性能の比率は、5-10%となっています。(1990年代のベクトル計算機は、40-50%の対ピーク性能を示していました。)
- 今、必要なのは
 - より高い実効性能を発揮することが可能な計算アルゴリズムと手法の開発とスケーラビリティの向上
 - プログラミングモデルなども含めて、スケーラブルな計算機環境の構築

ペタスケールシステムの構築



現在のテラ
FLOPS級の問題



‘複雑さ’の壁

ペタスケールシステムの
構築のための兆戦

Source: ORNL

- ソフトウェア（アプリケーション、OS、プログラミングAPIなど）の課題の克服が課題
- システムの複雑さと生産性

例：

Linpack Benchmark

- オリジナルベンチマークプログラム ~100ライン
- HPL ベンチマークプログラム ~10,000ライン (x100より複雑?)

システムの信頼性



An Overview of High Performance Computing

Jack Dongarra

University of Tennessee and Oak Ridge National Laboratory

HPC Asia 2005



Reliability of Leading-Edge HPC Systems

System	CPUs	Reliability
LANL ASCI Q	8,192	MTBI: 6.5 hours. Leading outage sources: storage, CPU, memory.
LLNL ASCI White	8,192	MTBF: 5.0 hours ('01) and 40 hours ('03). Leading outage sources: storage, CPU, 3 rd -party HW.
Pittsburgh Lemieux	3,016	MTBI: 9.7 hours.

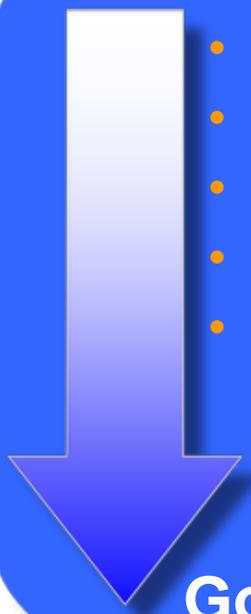
MTBI: mean time between interrupts = wall clock hours / # downtime periods

MTBF: mean time between failures (measured)

HPCの二極分化



- ‘Commodity’コンピューティング
 - ハードウェアは、’ Commodity’なものを利用して、SWの改善、サポート、利用技術のサポート、パッケージ実装などが今後の主要マーケットでの成功の鍵となる

- 
- 商用HW/SW
 - オープンソース
 - パーソナルクラスタ
 - 商用アプリケーション
 - マルチスレッド
- ‘Commodity’
コンピューティング
- Going DOWN**

‘今日の’スーパーコンピュータ



“...現在のスーパーコンピュータ(の性能)は、将来のデスクトップで実現される
....”



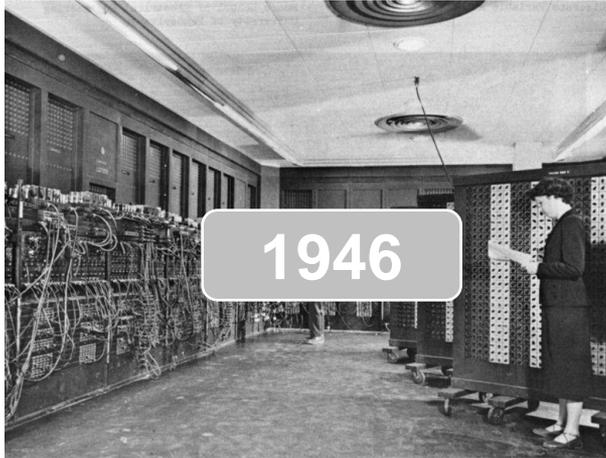
インテル社が主催する開発者向け会議「[Intel Developer Forum \(IDF\) fall 2006](#)」の最終日に米IntelのStephen Pawlowski氏(シニアフェロー、デジタル・エンタープライズ・グループCTO、ジェネラル・マネージャ)による[HPCに関する基調講演](#)より

Yesterday, Today and Tomorrow



ENIAC

20個の変数と300個の定数を記憶するメモリ

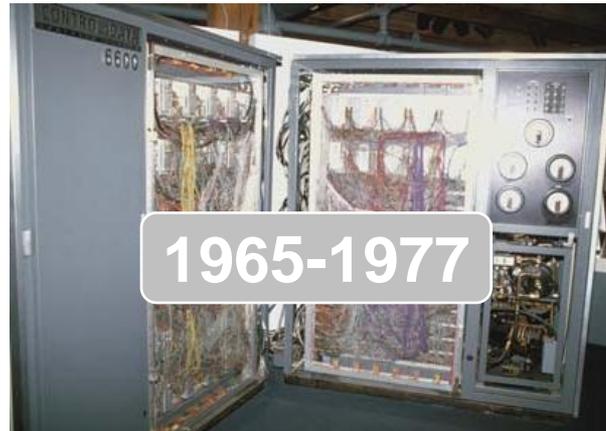


ASCI Red

最初のTFLOPSコンピュータシステム



1965-1977



CDC 6600

最初の商用スーパーコンピュータ

2006



Cluster.....

デュアルコアマイクロプロセッサを搭載

**PetaScale
Platforms**



**Personal
Computing...**

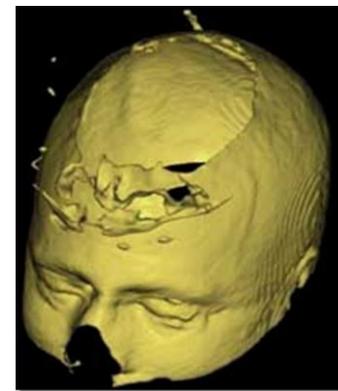
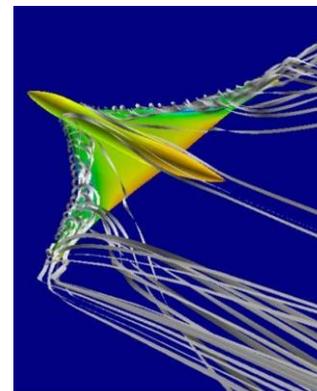
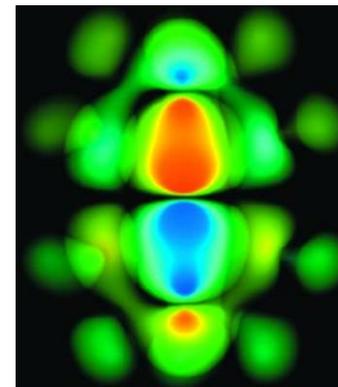
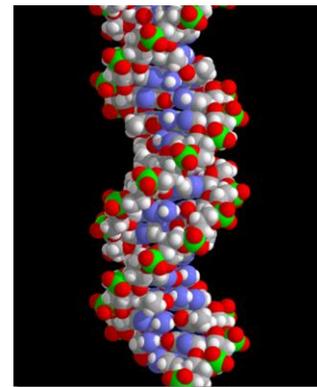
Yesterday, Today and Tomorrow



「...today's supercomputing problem is tomorrow's desktop problem...」

「現在の大規模なスーパーコンピュータを必要とする非常に解析困難な問題も、将来はより強力なデスクトップ・システムによって、解析可能となるだろう...」

Dr. Walter Brooks, NASA



HPCシステムのスケーリング

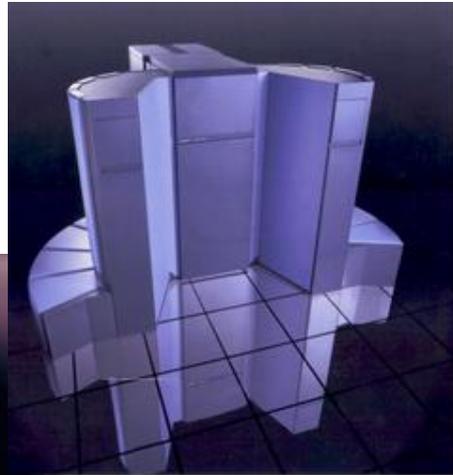


ベクトル処理

共有メモリ並列処理
.....



CRAY X-MP/4
1982
800MFLOPS Peak
100MHz



CRAY Y-MP/8
1988
2664MFLOPS Peak
166MHz

6年間



Intel Xeon 5100
2006
24000MFLOPS Peak
3000MHz

Intel Xeon 5300
2006
42560MFLOPS
2666MHz

6ヶ月

60倍



「Fast」「Good」「Cheap」のパズル



**Fast
+ Cheap
Inferior**

高い性能を廉価なシステムで構築することも可能です。ただ、そのようなシステムの場合、システムの構築や利用は、必ずしも容易ではありません。

付加価値の高い、性能の高いシステムは一般には、高価です。その付加価値がユーザにとって、メリットが無ければ、コスト・パフォーマンスの悪いシステムになるだけです。



**Good
+ Fast
Expensive**

**Good
+ Cheap
Slow**

比較的小規模なシステムであれば、廉価で使い勝手の良いものを探すことは可能です。しかし、そのようなシステムでは、拡張性やより大規模なシステム構築が出来ません。



- 代表的なHPCシステムのプラットフォームアーキテクチャ
 - クラスタシステム (1-2pノード)
 - SMPシステム (>2pノード)
 - SMPシステムをベースとしたクラスタシステム
- HPCシステムの利用の現状
 - HPCシステムとしては、クラスタシステムが一般化している
 - SMPシステムの利点はOEM及びユーザも理解しているが、また、SMPシステムの開発、販売、導入には問題がある

HPCシステムの問題



	OEMでの問題	エンドユーザの問題
クラスタ	<ul style="list-style-type: none">• 付加価値• ビジネスでの低マージン&価格競争	<ul style="list-style-type: none">• 運用コストなどを含むTCOは劇的に低下しない• その運用管理には、かなりの経験や知識が必要• システムの利用率及びアプリケーションの実効性能の維持
SMPシステム	<ul style="list-style-type: none">• 開発コスト(ハードウェアとソフトウェア)• SMPシステム(専用システム)とクラスタシステム(一般商用システム)の互換性の問題	<ul style="list-style-type: none">• 導入コスト• スケーラビリティ

HPCシステムのギャップ



SMP (Shared Memory Systems)

ワークステーションやサーバ
PA-RISC, POWER5,
Itaniumなどのプロセッサ
によるSMPサーバ

クラスタシステム

システムの構築には、
高いITスキルが要求される
運用管理コストが高い
複雑なオペレーション環境
複数のOS
クラスタファイルシステム
ソフトウェア、インストールや
アップグレードなど

ワークステーション
サーバ

クラスタ

#Processors

2

4

8

16

32

64

128

クラスタシステムの利点



- ハードウェアコストの劇的な低下
- 非常に高いピーク性能のシステムの導入が可能
- 増設が容易で、必要に応じて、システムの規模の拡大が容易
- 標準コンポーネントの技術革新と性能向上
 - プロセッサの性能向上（‘マルチコア’による省電力での性能向上）
 - 高性能なスケーラブルファイルシステム（オープンソース）
 - 高速な商用インターコネクトスイッチ

HPCシステムへの要求要件



- HPCシステムの増強のニーズは高い
 - より大規模な解析
 - より多くのシミュレーション
 - より短い時間でのシミュレーションの完了
- 同時にシステムに対するコスト・パフォーマンスの要求も厳しい
 - ベンダー間での競合
 - アプリケーションのスケラビリティ
 - より大規模なシステムの導入の希望
- 実質的には、HPCシステムとしては、「コスト・パフォーマンス」に対する要求が強い

Question: SMP(共有メモリ)の利点は？



- Answer:性能、運用管理、プログラミングなどの点で、分散メモリのシステムよりも優れていることは多々あります。
- しかし・・・実際には、共有メモリシステムの構築には、コストがかかります。このコストとそれによって得られる利点を評価した場合、その導入時の評価は非常に難しくなります。
 - コストや性能は定量的なものですが、運用管理の容易さやプログラミングのコストの評価は難しいものとなります。

HPCシステムへの要求要件



- HPCシステムの増強のニーズは高い
 - より大規模な解析
 - より多くのシミュレーション
 - より短い時間でのシミュレーションの完了
- 同時にシステムに対するコスト・パフォーマンスの要求も厳しい
 - ベンダー間での競合
 - アプリケーションのスケラビリティ
 - より大規模なシステムの導入の希望
- 実質的には、HPCシステムとしては、「コスト・パフォーマンス」に対する要求が強い

システム選択の課題

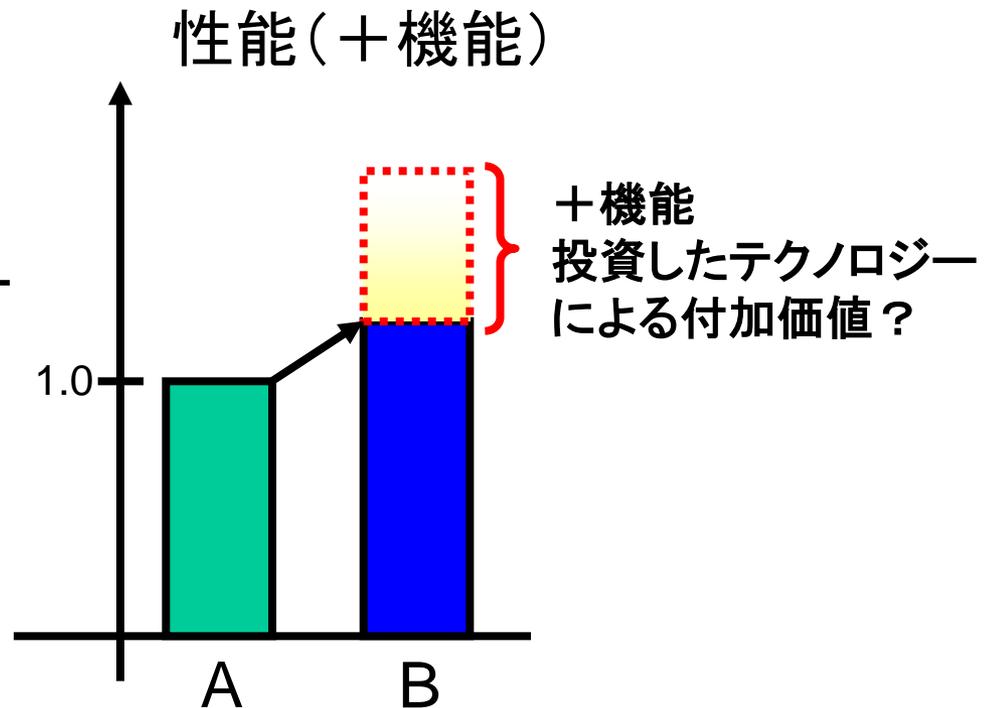
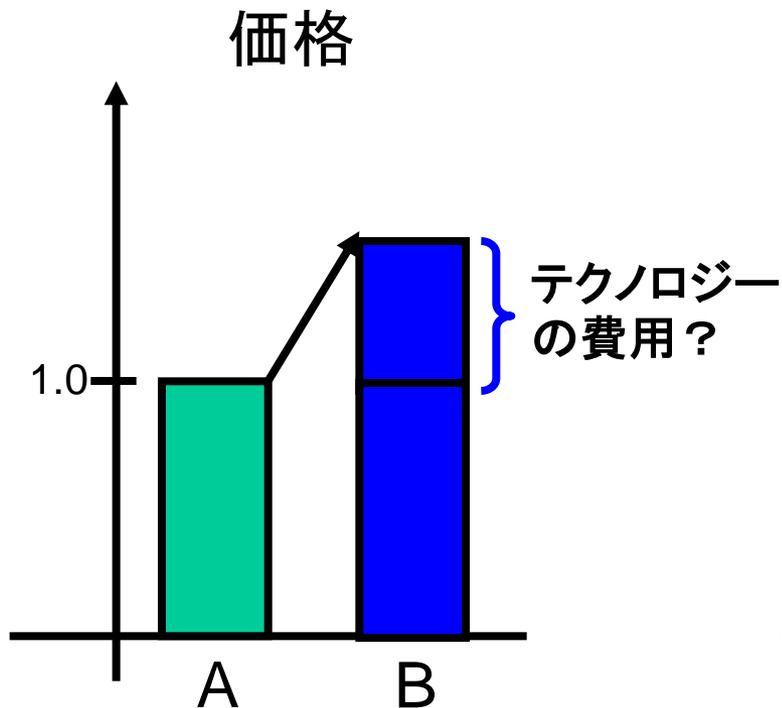


- 構築・運用管理コストの削減（TCO）
- より生産性の高いシステムの構築
 - 複数の技術を効果的に組み合わせることにより解決を図る
 - 提供される機能とその価値の評価
- 生産性の定義は非常に難しい
 - ストレージも含めたトータルな解析システムの提案
 - 運用・管理の容易さ

性能差と価格性能比



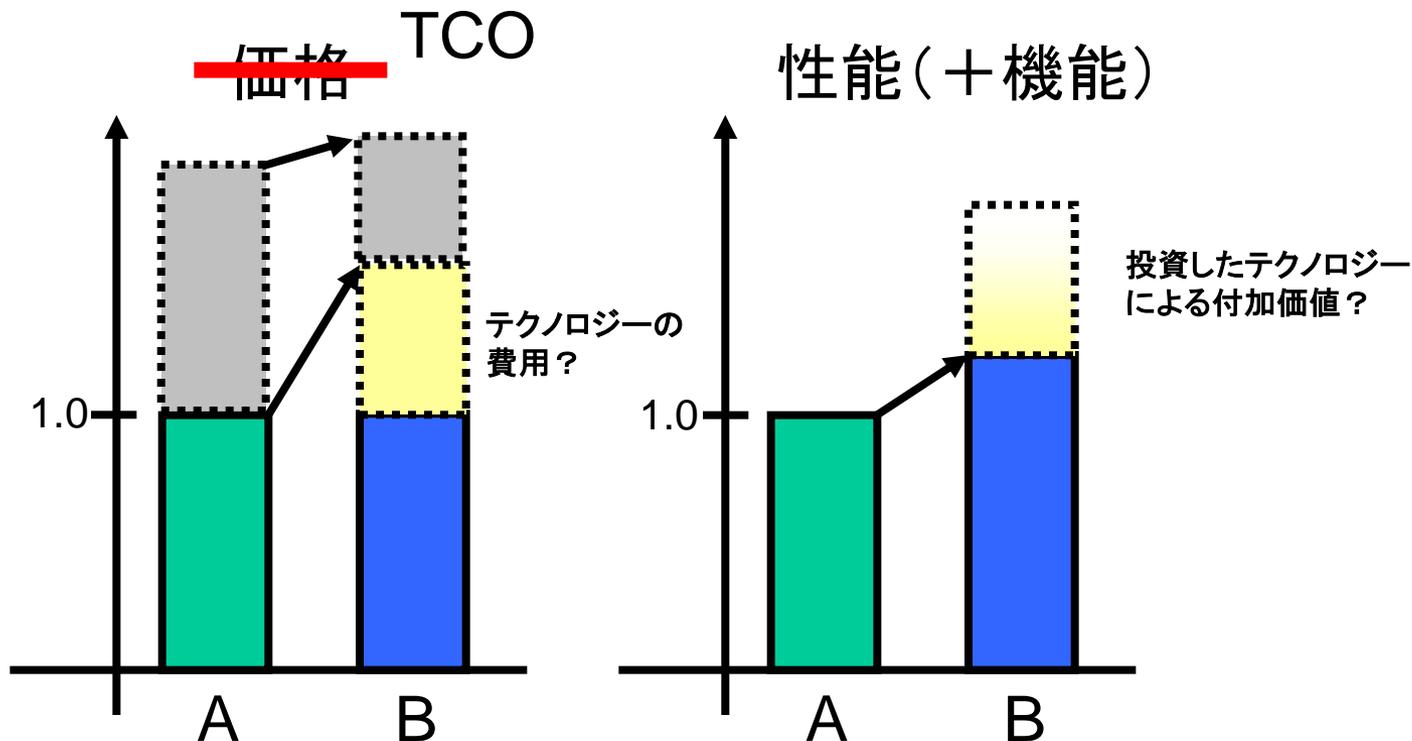
- 性能差は価格差に比例しない
 - 性能差 \neq 価格差 (性能差 $\leq \sqrt{\text{価格差}}$)



(性能＋機能)差とTCOコスト性能比



- (性能＋機能)差はTCOコストに比例する
– (性能＋機能)差 \geq TCOコスト差

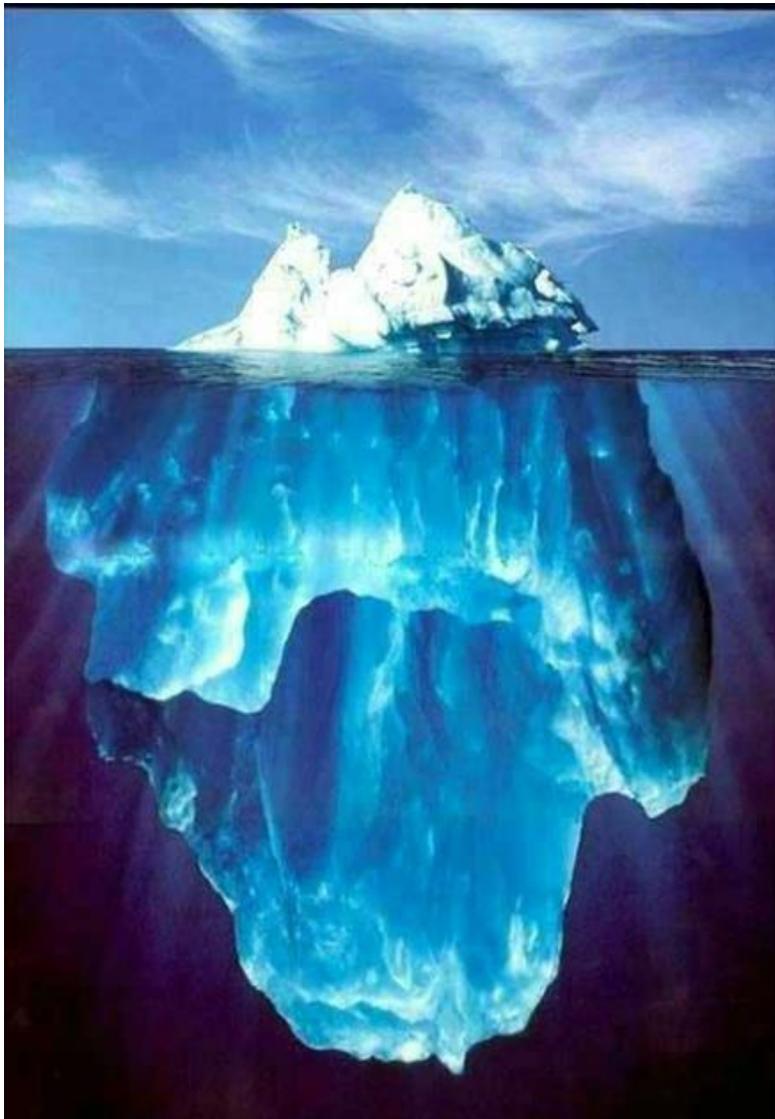


TCOの評価



- ハードウェアだけでなく、全てのコストを考慮したシステムTCOでの評価
- 運用コスト
 - フロアスペース/電力/システム管理
- インストレーションコスト
 - ノード数に大きく依存
- 購入コスト
 - プロセッサ/インターコネクト/メモリ/ソフトウェアコスト

TCO : Total Cost of Ownership



ハードウェアコストは
氷山の一角

ハードウェア導入コスト

▼ ソフトウェア導入コスト

システムサポート

システム運用管理コスト

保守サービス

データマネージメント

アプリケーション開発

アプリケーションライセンス

互換性

.....

TCO : Total Cost of Ownership



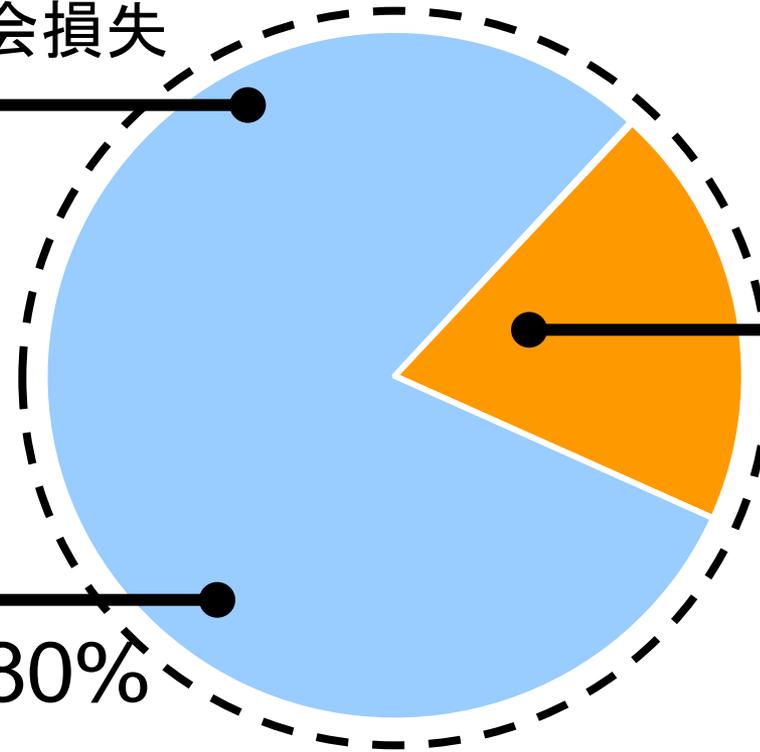
開発の遅れによる機会損失
最新技術の導入機会損失

機会損失コスト

運用管理
トラブル対応
トレーニング
設備費用

調達コスト 20%

オペレーションコスト 80%



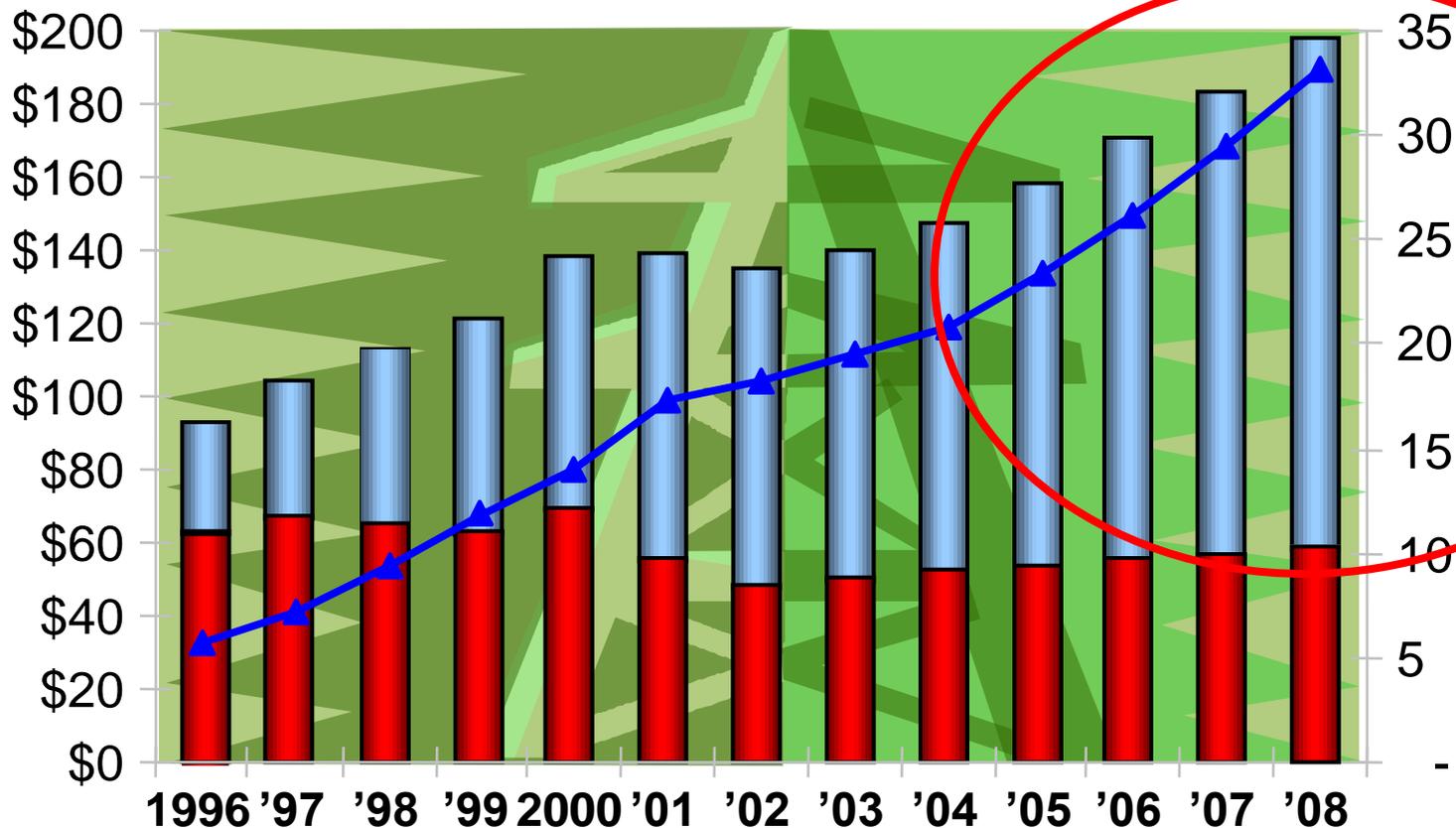
Source: Gartner Group 2005

IDCによるサーバビジネスの予想



費用 (USB\$)

▲ Installed Base (M Units)



- 新規サーバの導入費用(USM\$) 3% CAGR
- 運用管理費用 10% CAGR

Source: IDC, 2004

結果として……

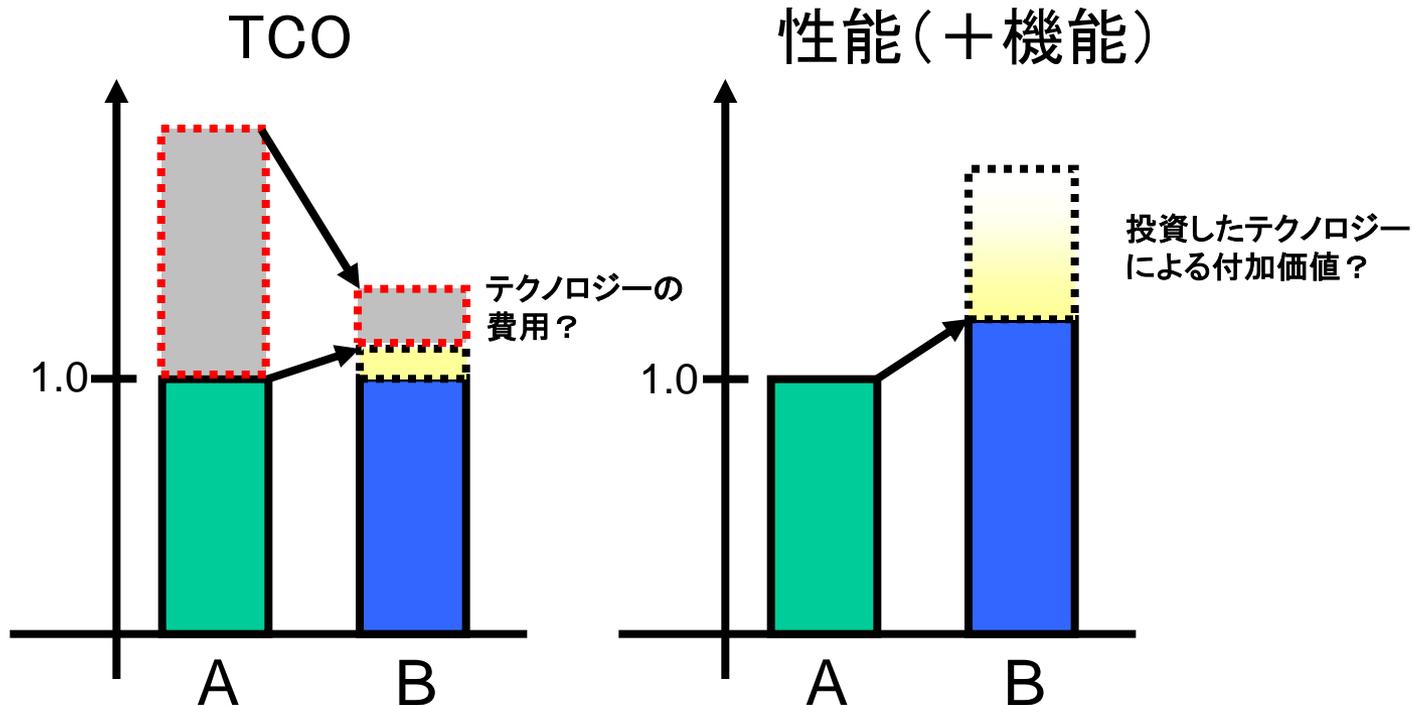


- 性能差は価格差に比例しない
 - 性能差 \neq 価格差
 - 従って、単純な価格性能比(価格/性能)の比較では、廉価なシステムが有利
- (性能+機能)差はTCOコスト差に比例する
 - (性能+機能)差 \geq TCOコスト差
 - TCOの向上を図るための機能強化(コスト)
 - 従って、TCOコスト/(性能+機能)の比較では、単純に廉価なシステムは有利とはならない

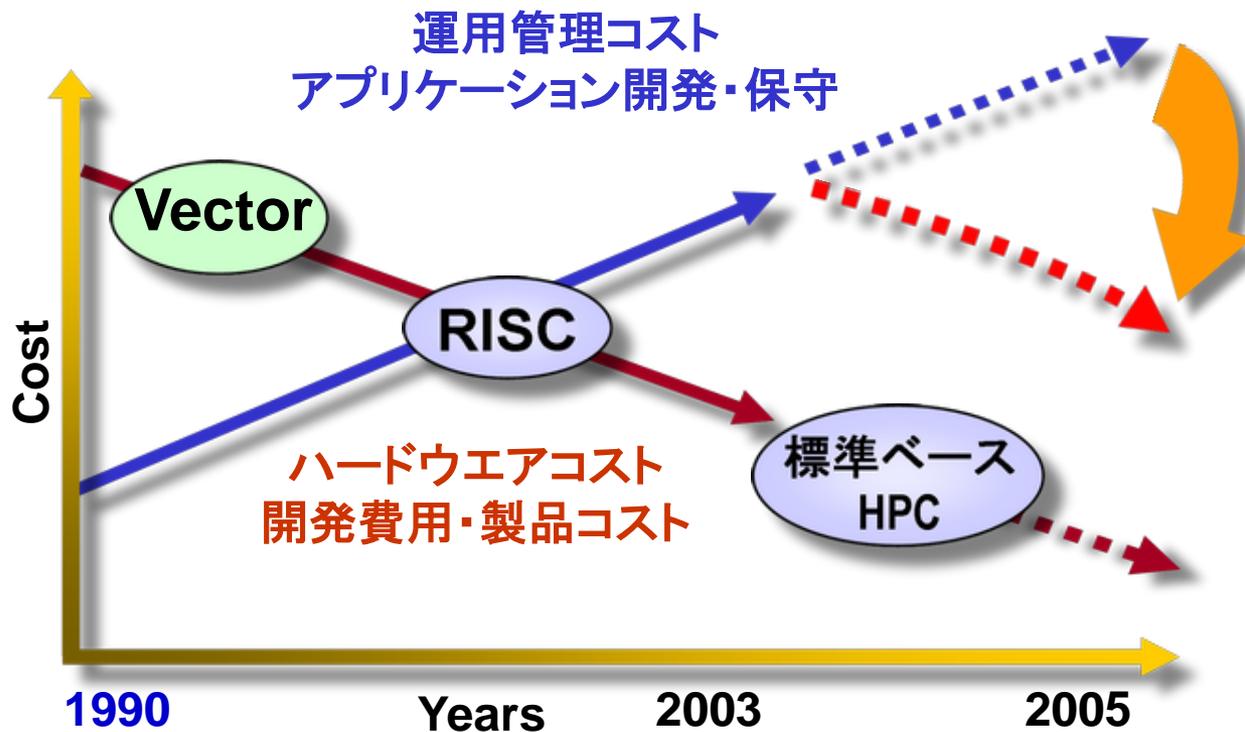
新しい取り組み



- 以下にコストを低減したテクノロジーの革新を実現するのか?が現在の課題となっている
- TCOの低減を大幅に図り、性能と機能の強化を同時に図るアプローチが必要では?



HPCシステムでのTCO



生産性の高いプログラミング
OpenMP
クラスタOpenMP
統合クラスタ向けOS
Windows CCS
新しいシステムコンセプト
パーソナルクラスタ
スケーラブルx86システム

TCO低減の実現のための
ソリューション

SMPとクラスタのギャップを埋める



SMP (Shared Memory Systems)

ワークステーションやサーバ
PA-RISC, POWER5,
Itaniumなどのプロセッサ
によるSMPサーバ

クラスタシステム

システムの構築には、
高いITスキルが要求される
運用管理コストが高い
複雑なオペレーション環境
複数のOS
クラスタファイルシステム
ソフトウェア、インストールや
アップグレードなど

ワークステーション
サーバ

クラスタ

#Processors

2

4

8

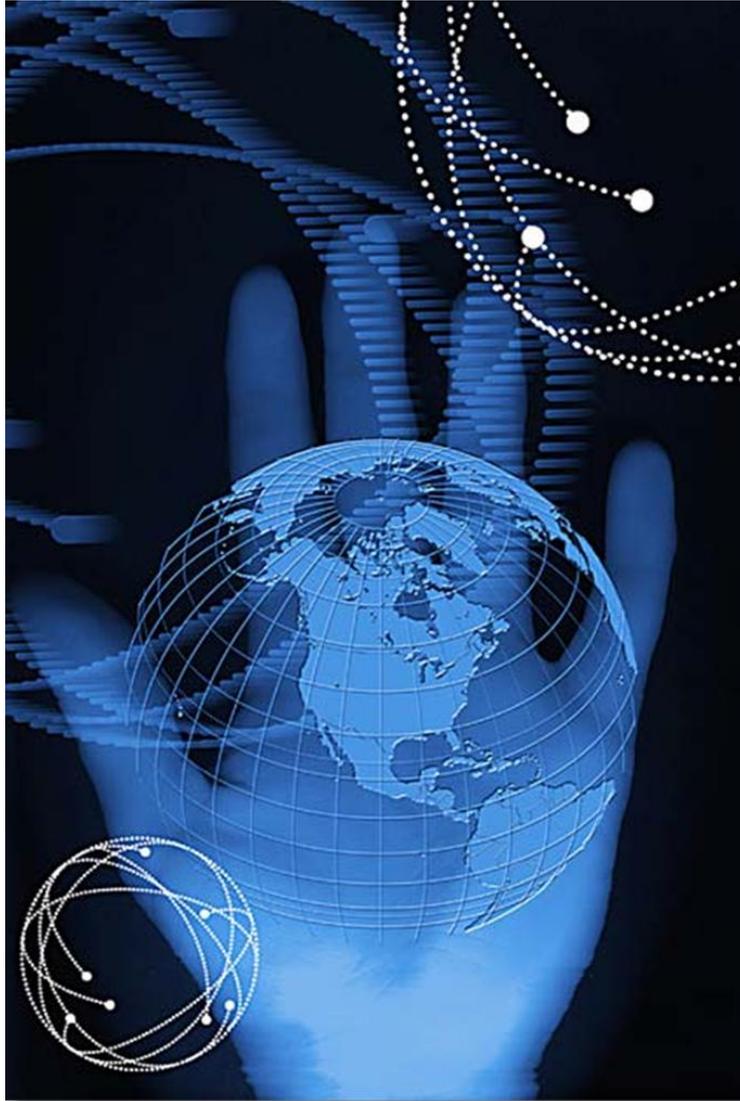
16

32

64

128

この資料について



ここに掲載した資料は、弊社の調査と見解に基くものであり、資料の中で示されている製品やサービスを提供している各社の公式な見解でも、また、マーケティング戦略に基くものではありません。あくまで、弊社としての意見だということにご注意ください。これらの資料の無断での引用、転載を禁じます。

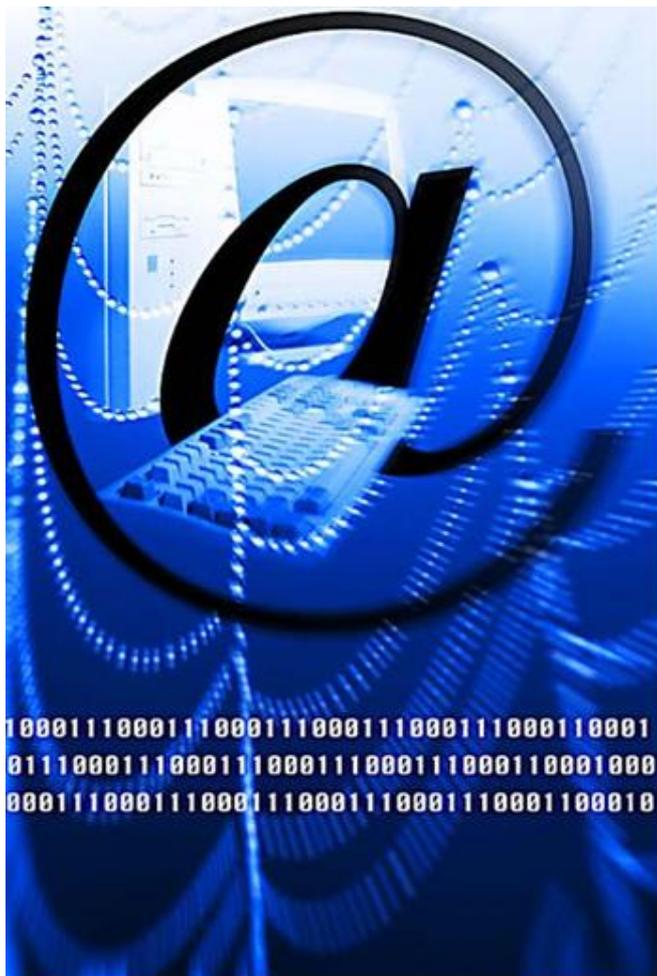
社名、製品名などは、一般に各社の商標または登録商標です。なお、本文中では、特に®、TMマークは明記しておりません。

In general, the name of the company and the product name, etc. are the trademarks or, registered trademarks of each company.

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

2007年1月

さらに詳しい情報や最新情報は.....



ホームページにて公開しています。
ホームページには、お問い合わせ窓口も開設してありますので、ご利用ください。

コンサルテーション

<http://www.sstc.co.jp>

製品技術

<http://www.hp2c.biz>

2007年1月