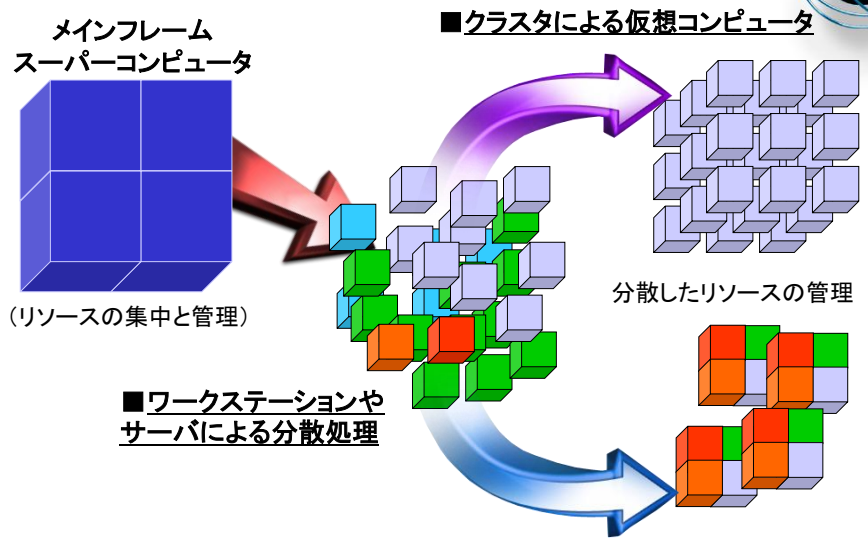
HP²C:High Performance and Productivity

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

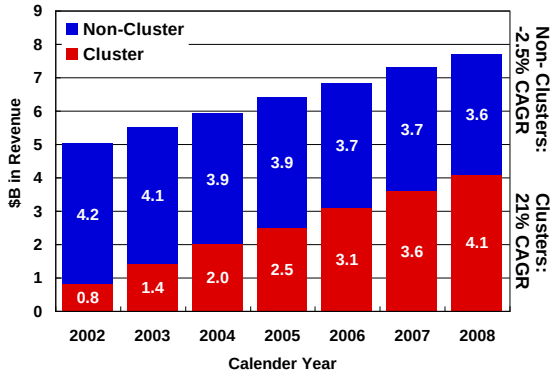
HPCプラットフォームの変遷



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



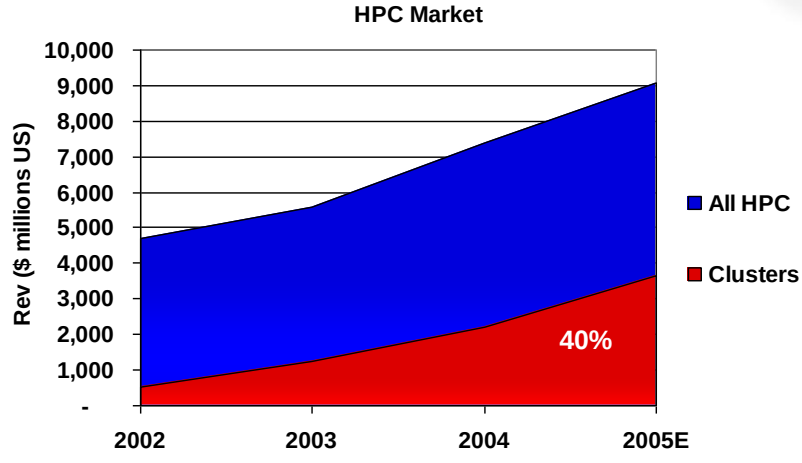
Clusters Accounted for 33% of Revenue in ‘04

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ー16

WW HPC Server IDC Forecast

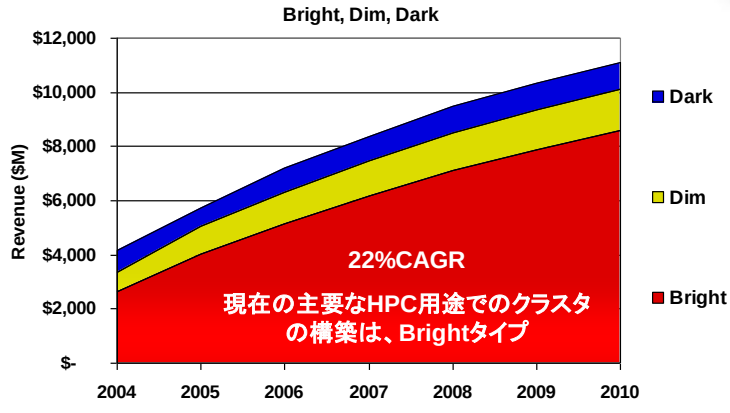
†IDC MCS: The Cluster Revolution in Technical Computing Markets (2006), IDC, Feb 2006, #



→Question:残りの60%は?

WW HPC Server IDC Forecast

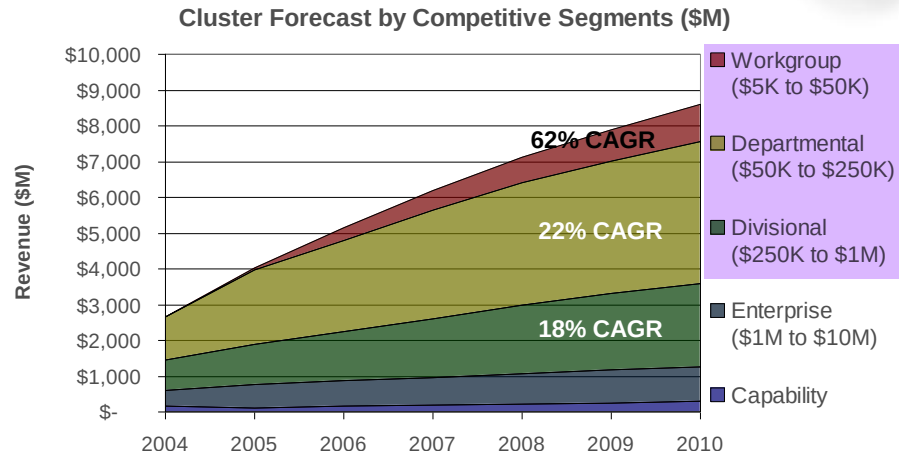
†IDC MCS: The Cluster Revolution in Technical Computing Markets (2006), IDC, Feb 2006, #



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, #



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

将来予測の難しさ

- “I think there is a world market for maybe five computers.”
  - Thomas Watson, chairman of IBM, 1943.
- “There is no reason for any individual to have a computer in their home”
  - Ken Olson, president and founder of digital equipment corporation, 1977.
- “There are only about 100 potential customers worldwide for a Cray-1”
  - Seymour Cray, 1977.
- “640K [of memory] ought to be enough for anybody.”
  - Bill Gates, chairman of Microsoft,1981.

「未来を予測する最良の方法は、それを創造してしまうことである」

"The best way to predict future is to invent it."

Dr. Alan Kay, President of Viewpoints Research Institute, Inc.,

ITマネージメントの課題

- プラットフォームの内部からの保護:
  - ウイルスやワームなど悪意あるソフトウェアからの保護
- 資産管理:
  - 多くの IT 部門では、特定できない資産が問題
- オンラインおよびリモート管理・診断機能:
  - アップグレード、診断、復旧のための作業の効率化
- アプリケーション統合の困難さ:
  - アプリケーションの高度化と複雑化によって、複数のアプリケーションを組み合わせるての動作に問題
- 動的なリソース割り当て:
  - 組織内で未使用のCPUやメモリの活用

マーケットトレンド

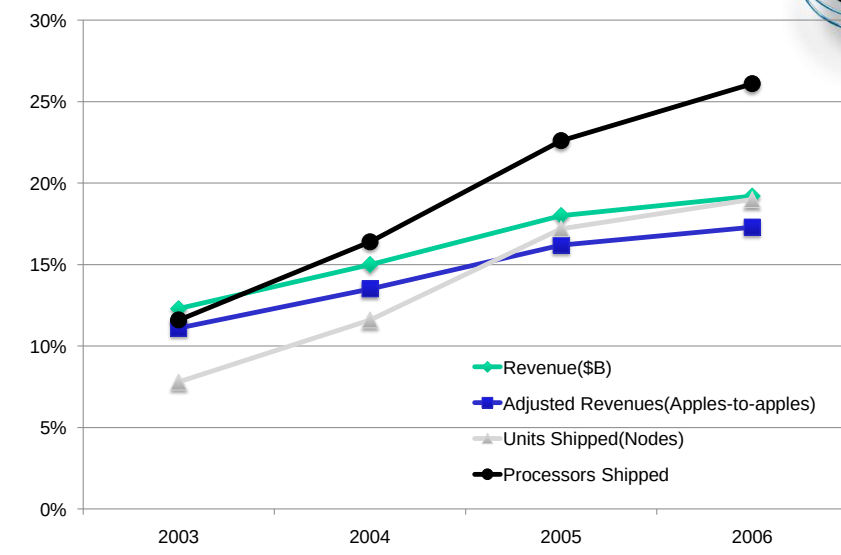
All Servers Worldwide	2003	2004	2005	2006	2003 to 2006 CAGR	2005 to 2006 CAGR
Total Factory Revenue($B)	$46,149	$49,146	$51,268	$52,251	4.2%	1.9%
Units Shipped(same as nodes)	5,278,222	6,307,484	7,050,099	7,472,649	12.3%	6.0%
Processor Dies Shipped	8,662,823	10,134,624	11,712,766	12,779,159	13.8%	9.1%

HPC Technical Servers Worldwide	2003	2004	2005	2006	2003 to 2006 CAGR	2005 to 2006 CAGR
HPC Server Revenue($B)	$5,698	$7,393	$9,208	$10,030	20.7%	8.9%
Adjusted Revenues(To much enterprise)	$5,128	$6,654	$8,287	$9,027	20.7%	8.9%
Node Units Shipped	411,327	734,510	1,215,735	1,419,221	51.1%	16.7%
Processor Elements Shipped	1,002,905	1,657,827	2,681,079	3,351,843	49.5%	25.0%

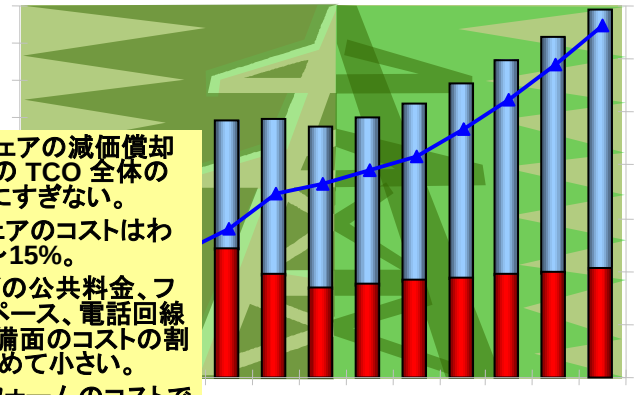
HPC As A Ratio Of All Servers	2003	2004	2005	2006
Revenue($B)	12.3%	15.0%	18.0%	19.2%
Adjusted Revenues(Apples-to-apples)	11.1%	13.5%	16.2%	17.3%
Units Shipped(Nodes)	7.8%	11.6%	17.2%	19.0%
Processors Shipped	11.6%	16.4%	22.6%	26.1%

Source: IDC 2007

HPCマーケット（対全サーバマーケット）



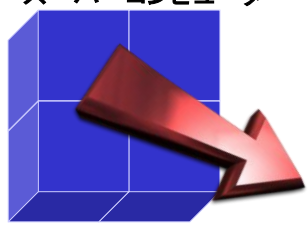
マーケットトレンド



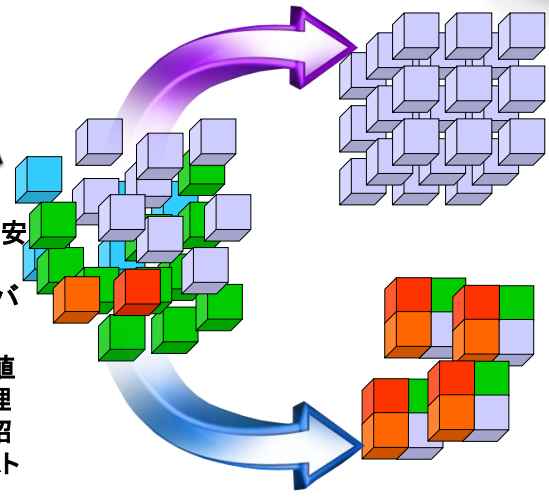
- ハードウェアの減価償却費は IT の TCO 全体の約25% にすぎない。
- ソフトウェアのコストはわずか10～15%。
- 電気などの公共料金、フロア・スペース、電話回線など、設備面のコストの割合もきわめて小さい。
- プラットフォームのコストではなく、TCOの大きな比率を占めるのは人件費となっている。

運用管理コストの低減

メインフレーム
スーパーコンピュータ



クラスタによる仮想コンピュータ



- 「もっと処理性能を」と「より安価に」を実現するために、ワークステーションやサーバでの分散処理の導入
- ユーザや企業に新たな価値をもたらしてはいるが、管理責任とその負担の分散を招き、結果的に運用管理コストを押し上げる

仮想化によるサーバ・コンソリデーション

次世代HPCインフラ

- コアとスレッド
  - より多くのスレッドを効率よく利用可能
  - マルチスレッド向け最適化
- 電力管理
  - 省電力
  - データセンター運用管理機能
- 仮想化
  - 柔軟性と優れた運用管理
  - 仮想的なシステムパーティション
- RAS
  - ハードウエアベースの自己監視/自己管理
  - ファームウエアベースのエラー履歴管理
- システム管理
  - より低いTCOを実現するための一般・標準化されたマネージメント機能

システムの‘バランス’

エコシステムに対応するためにも、電力消費量や発熱量を積極的に抑える技術の開発

高速プロセッサ

マルチコアによって、プロセッサ単体の処理性能の向上を図る

省電力

64ビットアドレス

64ビットのアドレス空間と拡張されたレジスタによるOSとアプリケーション双方の機能・性能拡張

大規模なクラスタシステムの構築及びアプリケーションのワークロードに対応した高速性能

インターコネクト

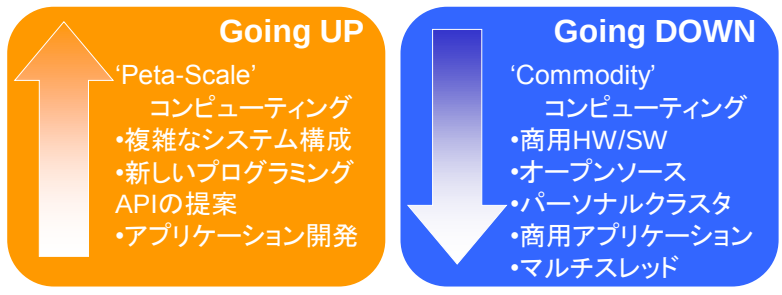
メモリ性能と容量

64ビット化とマルチコア化にともなう高速・大容量へのニーズに対応し、また、その拡張性の高い実装技術の実現

CPU-メモリ間の高速なデータ転送やより高速なネットワーク、大規模なストレージのサポート

I/Oバンド幅

HPCの二極分化



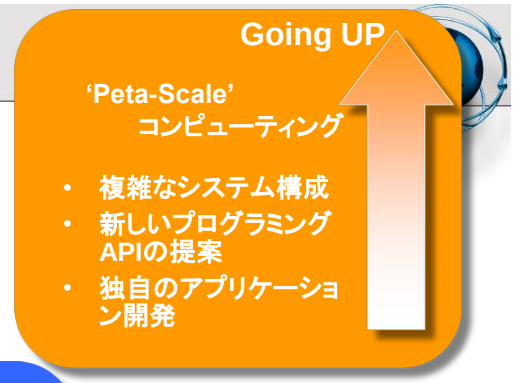
システムとユーザの尺度

システムの尺度		ユーザの尺度
Flop/s	⇔	計算終了までの時間
メモリサイズ(GB)	⇔	モデルのサイズと計算結果
プロセッサ数	⇔	ワークロードでの試行
データ長	⇔	計算精度
システム構成(クラスタ)	⇔	導入コストと運用コスト
スケーラビリティ	⇔	ベンチマーク

- ユーザの尺度での性能(Performance)は、時間当たりにどれだけの仕事を処理出来るか(仕事量 / 時間)
- Flopsでの評価は実際には意味がない。また、問題の規模(small, medium, large) という評価も難しい。
- “スケーラビリティ”は、対象を明確に規定する必要がある

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。

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

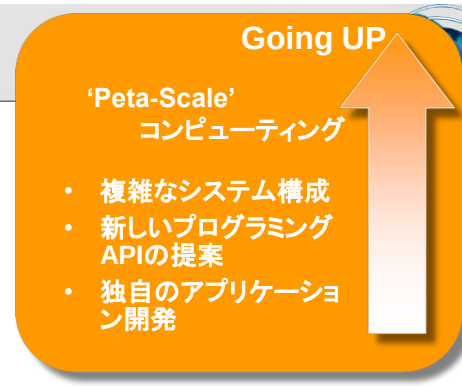
ペタFLOPS級‘
スーパーコンピュータ

- 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（ローレンス・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

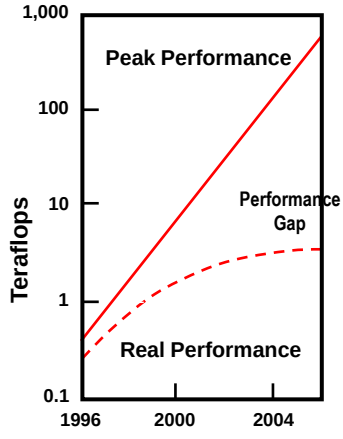
- ハードウエア、ソフトウエア、開発環境など様々な技術を統合して、従来は解析出来なかった問題を十分な経済性をもって、解決すること

HPCの二極分化



- **‘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年台は、性能の向上は、10^2のオーダーでしたが、2000年台になると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

Source: ORNL

- ソフトウエア（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, 3rd-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
  - ハードウエアは、' 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

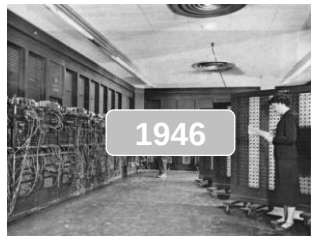
‘今日の’スーパーコンピュータ

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

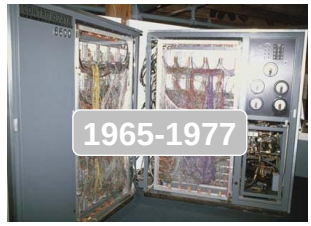
インテル社が主催する開発者向け会議「Intel Developer Forum (IDF) fall 2006」の最終日に米IntelのStephen Pawlowski氏（シニアフェロー，デジタル・エンタープライズ・グループCTO，ジェネラル・マネジャ）によるHPCに関する基調講演 より

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



CDC 6600
最初の商用スーパーコンピュータ

Cluster.....
デュアルコアマイクロプロセッサを搭載

PetaScale Platforms

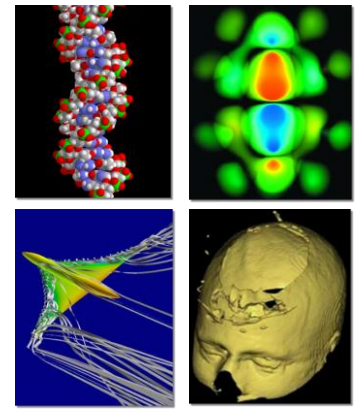
Personal Computing…

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ー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

「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。

HPCシステムの課題

- 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	• 付加価値 • ビジネスでの低マージン&価格競争	• 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	• 開発コスト（ハードウエアとソフトウエア） • 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	• 導入コスト • スケーラビリティ

HPCシステムのギャップ

クラスタシステムの利点

- ハードウエ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
  - より多くのシュミレーション
  - より短い時間でのシュミレーションの完了
- 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ーラビリティ
  - 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

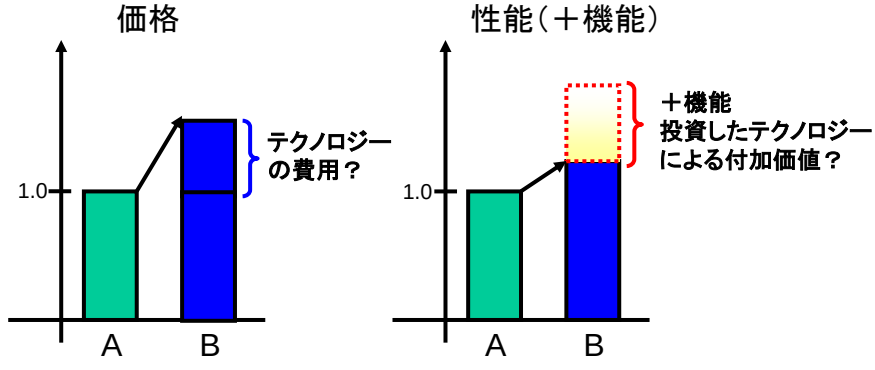
- HPCシステムの増強のニーズは高い
  - より大規模な解析
  - より多くのシュミレーション
  - より短い時間でのシュミレーションの完了
- 同時にシステムに対するコスト・パフォーマンスの要求も厳しい
  - ベンダー間での競合
  - アプリケーションのスケーラビリティ
  - より大規模なシステムの導入の希望
- 実質的には、HPCシステムとしては、「コスト・パフォーマンス」に対する要求が強い

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

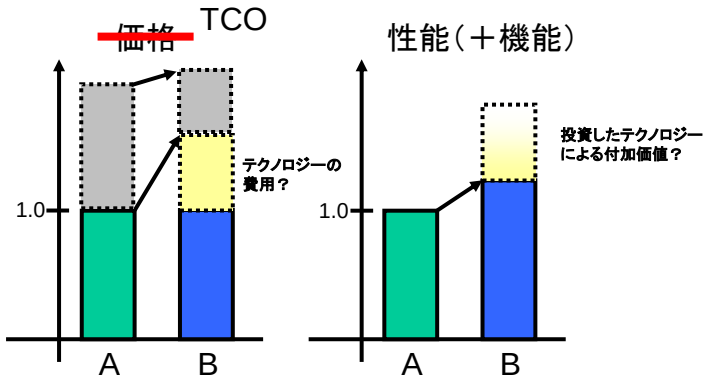
性能差と価格性能比

- 性能差は価格差に比例しない
  - 性能差 ≠ 価格差（性能差 ≦ √価格差?）



（性能＋機能）差とTCOコスト性能比

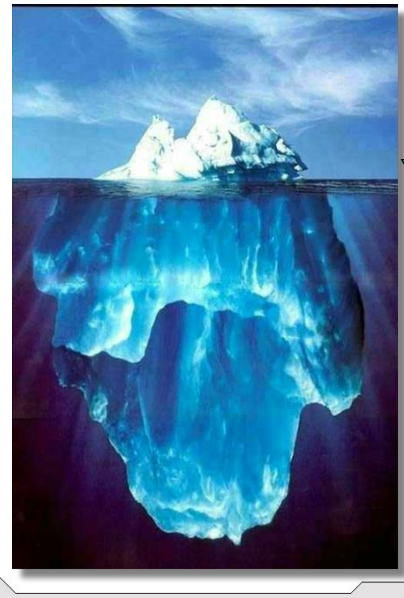
- （性能＋機能）差はTCOコストに比例する
  - （性能＋機能）差　≧　TCOコスト差



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

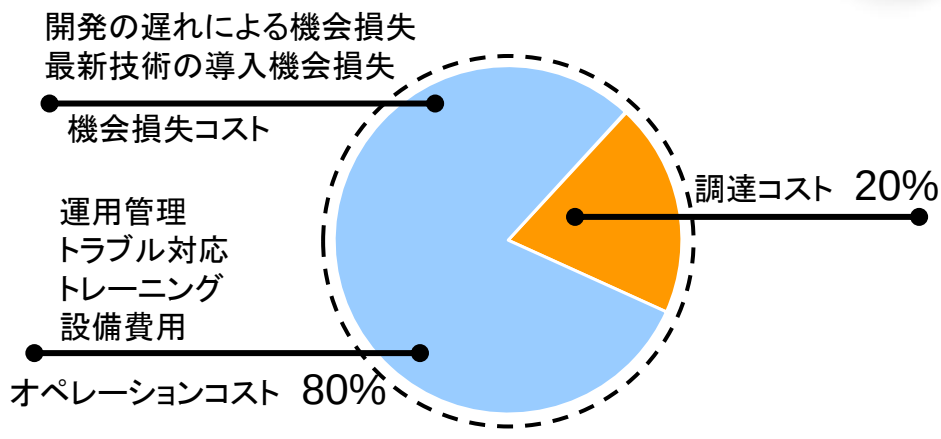


ハードウエアコストは
氷山の一角

ハードウエア導入コスト
ソフトウエア導入コスト

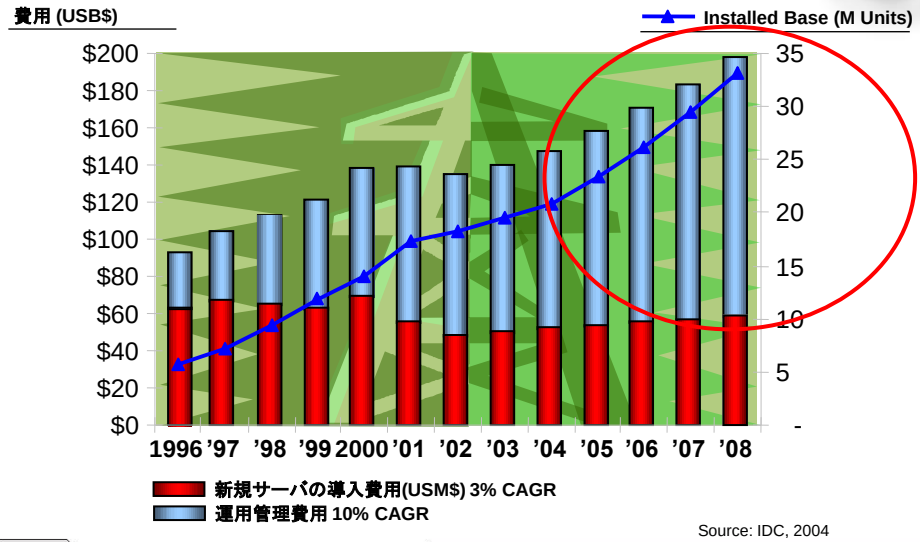
システムサポート
システム運用管理コスト
保守サービス
データマネージメント
アプリケーション開発
アプリケーションライセンス
互換性
・・・・・・・

TCO : Total Cost of Ownership



Source: Gartner Group 2005

IDCによるサーバビジネスの予想



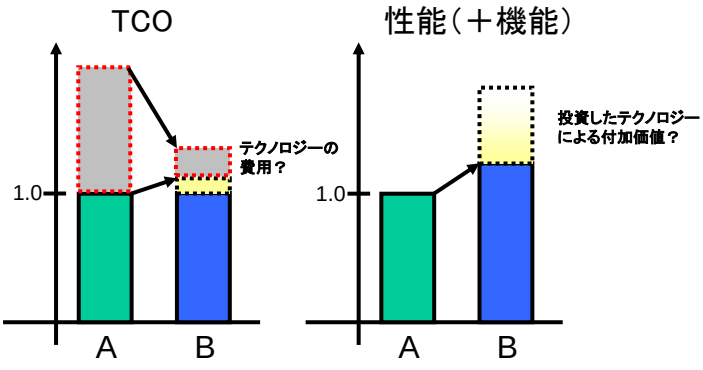
Source: IDC, 2004

結果として・・・・・

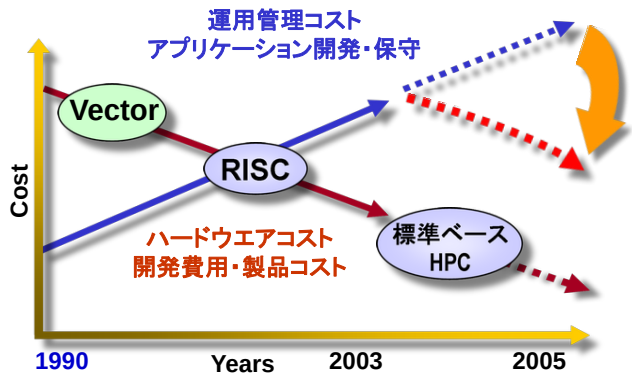
- 性能差は価格差に比例しない
  - 性能差 ≠ 価格差
  - 従って、単純な価格性能比（価格/性能）の比較では、廉価なシステムが有利
- （性能＋機能）差はTCOコスト差に比例する
  - （性能＋機能）差 ≧ TCOコスト差
  - TCOの向上を図るための機能強化（コスト）
  - 従って、TCOコスト/(性能+機能）の比較では、単純に廉価なシステムは有利とはならない

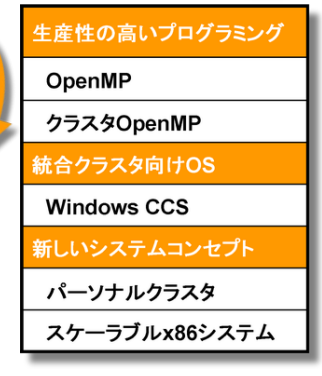
新しい取り組み

- 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？



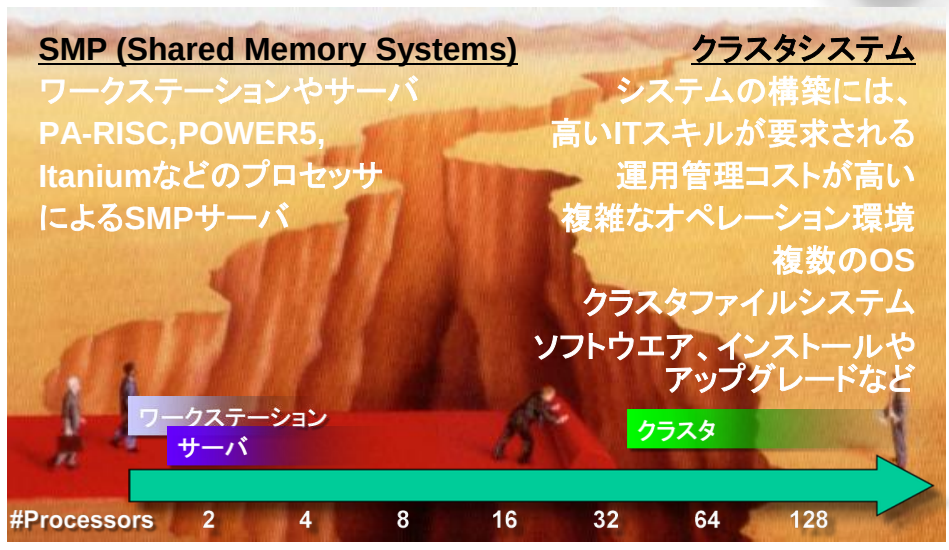
HPCシステムでのTCO





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



この資料について

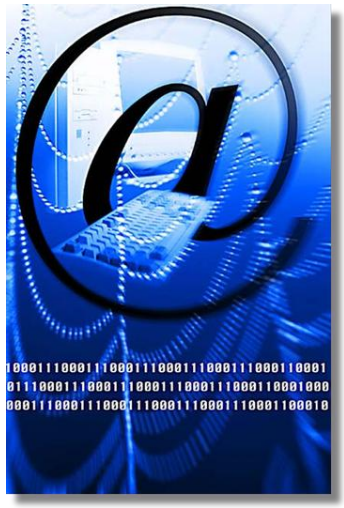
ここに掲載した資料は、弊社の調査と見解に基くものであり、資料の中で示されている製品やサービスを提供している各社の公式な見解でも、また、マーケティング戦略に基くものではありません。あくまで、弊社としての意見だということにご注意ください。これらの資料の無断での引用、転載を禁じます。

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.

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