SSTC
High Performance and Productivity
スケーラブルシステムズ株式会社

MCAE シュミレーションと HPC クラスタについて

スケーラブルシステムズ株式会社

MCAE シュミレーションと HPC クラスタについて

目次

■はじめに.....2
■製造業における MCAE シュミレーション.....3
■MCAE シュミレーションにおける HPC システム.....4
■HPC システムにおけるクラスタシステム.....6
■HPC クラスタの構成.....7
■HPC クラスタ上での MCAE アプリケーションの特性 9
■HPC クラスタの課題　～　より少ない TCO でのハイパフォーマンスの実現へ 11
■HPC クラスタと COTS システム～インターコネクトの選択.....12
■まとめとして.....15

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。
ご購入の際は（Tel：03-5875-4718 E-mail：biz@sstc.co.jp ）までお問い合わせ下さい。

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. , 2006. Unauthorized use is strictly forbidden.
無断での引用、転載を禁じます。
2006 年 8 月　改訂

■はじめに

熾烈な国際的競争を強いられる製造業では、製品の品質と市場投入までの時間が企業の成功のための重要な鍵となります。日々高まるグローバルな競争によるプレッシャー、そしてますます増大する顧客ニーズといった問題に直面しています。そのなかで、製造業各社では、顧客満足の向上とコストの削減を実現させるべく、品質向上のための試みを次々と取り入れています。また、「最高の品質を提供できる会社だけがこの競争社会で成功する」という認識が定着し、こうした方向へ企業を駆り立てているのです。この牽引力は品質だけではありません。コストの削減および生産性の向上がほとんどの場合、改善の過程における副産物となっているのです。

自動車をはじめ、様々な製造業で設計品質の向上と設計サイクル期間の短縮、また同時にコストの削減を実現するために、MCAE（メカニカル・コンピュータ支援エンジニアリング：Mechanical Computer Aided Engineering）を、コンピュータによる大規模な数値解析をより高速に実行するために HPC（ハイパフォーマンス・コンピューティング：High Performance Computing）システム上で活用しています。

製造業での設計業務	**IT トレンド**
● **製品ラインの拡大** ● **スタイルと快適性の追求** ● **設計条件や境界条件の複雑化** ● **国が定める基準の変化（より厳しく）** ● **統計的な処理の必要性** ● **統計・確率に基く計算結果の変動予測** ● **組み合わせ最適化の要求**	● **ハードウエア価格の低下** ● **システムの性能の向上** ● **システムの成熟度** ● **システム機能の強化** ● **TCO の増加**

製造業での設計業務の課題と IT トレンド: 製造業の設計業務において、コンピュータによる MCAE は必須のツールとなっています。そのため、製品設計に関して、IT トレンドの影響は大きくなっています。

今日、エンジニアは実際の利用条件下で製品デザインがどのように機能するかを解明するため、有限要素解析(FEA)および計算流体力学(CFD)シミュレーションを利用しています。 このような高度なシミュレーションは、使用可能な計算技術に対する要求レベルをさらに押し上げています。この傾向は、より詳細なモデリングを行い、デザインごとのシミュレーションのケースを増やすことで、設計の不確実性を低減しようと企業のエンジニ アが奮闘している中、ますます顕著なものとなってきています。このような高い要求レベルに応えるため、ハイパフォーマンス・コンピューティング(HPC) システ

ムでは、大規模な共有メモリ、メモリとプロセッサ間の高いバンド幅、ディスクへの高速の I/O、および高いスケーラビリティを実現する低レイテンシ のインターコネクトが不可欠となってきています。MCAE で活用されるアプリケーション・ソフトウエアは、HPC システムの性能を最大限に発揮するために、高度に最適化・並列化されています。また、その最適化・並列化は常に HPC システムの高性能化、高速化のためのシステムの進化と歩調を合わせており、最新の HPC システムによる最適なシュミレーション環境を目指した技術革新が進んでいます。

MCAE 解析に求められるもの	システムに求められるもの
● シュミレーションの回数とモデルの種類の増加 ● シュミレーションモデル自身の複雑さの増大 ● 解析時のデータのスムーズなやり取り ● データのボリュームの増大とその管理	● 堅牢なデータ管理機能 ● データへの容易なアクセス ● 大規模なデータへの効率的なアクセスが可能 ● シュミレーションにおける再現性

システムに求められる高いデータ管理機能:データへの容易なアクセスと大規模なシュミレーション処理の結果をスムーズに処理できる性能が求められます。また、組み合わせ最適化などの処理では、複数のアプリケーション間でのデータ共有も求められます。

■製造業における MCAE シュミレーション

製造業では、MCAE アプリケーションと HPC システムの双方の最新のテクノロジーを組み合わせることで、設計にかかるサイクル時間と費用を削減し、製品の品質向上を図り、また、企業間での競争に打ち勝つ製品の開発を行っています。

例えば、自動車メーカーは、このような MCAE によるコンピュータ・シュミレーションによるエンジニアリング設計により、より安全性の高いクルマの開発を行っています。限られた開発のためのリソースと開発期間内での品質と生産性を高めるために、MCAE は現在では必須のツールとなっています。コンピュータ・シュミレーションには、試作品による実験や試験では得られない利点があります。ここでも、クルマの例になりますが、衝突シュミレーションにおけるダミーモデルも、コンピュータシュミレーションでは、より人体の骨格や組織までも忠実にモデル化することが可能であり、事故の解析と人体の耐性を検証し、クルマの安全性を高めることが可能となります。また、MCAE によるシュミレーションは、設計開発のどの段階でも、高い自由度で利用できるというメリットがあります。試作品の場合、設計が進むにつれて、仕様が固まることで、変更などの自由度が大幅に制限されたり、その試作品の変更に非常にコストがかかるという問題がありますが、数値シュミレーションでは、部品の変更などが、比較的自由に出来ることもあり、調整が遥かに容易です。

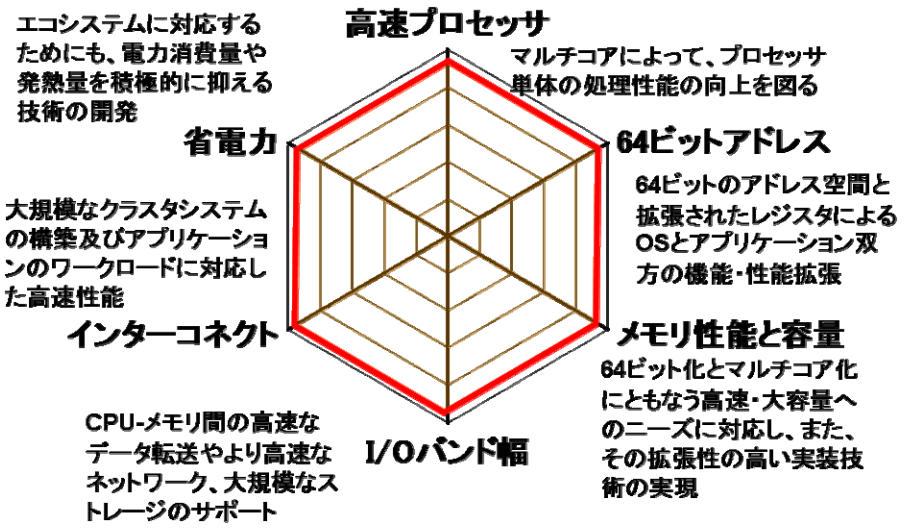
検証のデジタル化と物理的な確証物理的環境のみのパッケージ製品の開発は、費用がかかるだけでなく技術革新の妨げとなります。どの企業でも、いかに画期的な効果があるといえども、入念な検証なしに消費者へ紹介するはありえません。CAE を導入することで、新製品が市場に出るまで数年かけて試行錯誤を繰り返していた、初期工程のプロトタイプ作成をバーチャル・プロトタイプの開発に切り替えることが出来、バーチャル・プロトタイプにより、物理的なモデルを設計する前の段階で製品評価が可能となります。バーチャル・プロトタイプなら、‘もしこうだったら’という想定で何通りも検証することが可能であり、入念な製品評価が可能となります。

■MCAE シュミレーションにおける HPC システム

HPC システムは、大規模な解析を短時間で処理するために利用されますが、そのシステムの構成方法には、幾つかの方法があります。現在の HPC システムは、基本的にはノードと呼ぶ計算機システムを複数接続したものになります。このノードとなる計算機システムとしては、ベクトル計算機、RISC プロセッサを搭載した SMP システム、インテル Xeon などに代表される商用プロセッサを1～2プロセッサなどがあります。ノード間の接続方法は、一般的なネットワークを利用したものから、より高速のデータ転送が可能なインターコネクトを利用したもの、さらに、高い性能を実現するために、特別に開発されたハードウエアやソフトウエアを利用したものなど様々なものが利用されています。

MCAE シュミレーションの性能向上は、このような HPC システムの急激な進化と共に、高度な技術革新によって、現在も進んでいます。特に、HPC システムの技術革新は、特定の企業の研究開発部門の中で進むのではなく、コモディティ製品として、直ちに商用化され、広く利用されます。また、ソフトウエアもオープンソースの活発な動きのなかで機能強化が図られています。現在では、世界のあらゆる場所で行われている技術革新が非常に短時間で利用可能となっています。

HPC システムは、単に高速なプロセッサだけで構築できるものではありません。HPC システムの歴史では、より高速なプロセッサ（現在では、マイクロプロセッサ）の開発を目指してきたことも事実であり、また、過去も現在もそのようなプロセッサが HPC システムの中核を構成しています。ただ、MCAE シュミレーションは、非常に複雑なワークロードを要求するため、高速なプロセッサだけでなく、プラットフォームの様々な面での技術革新とシステムのバランスが重要となります。



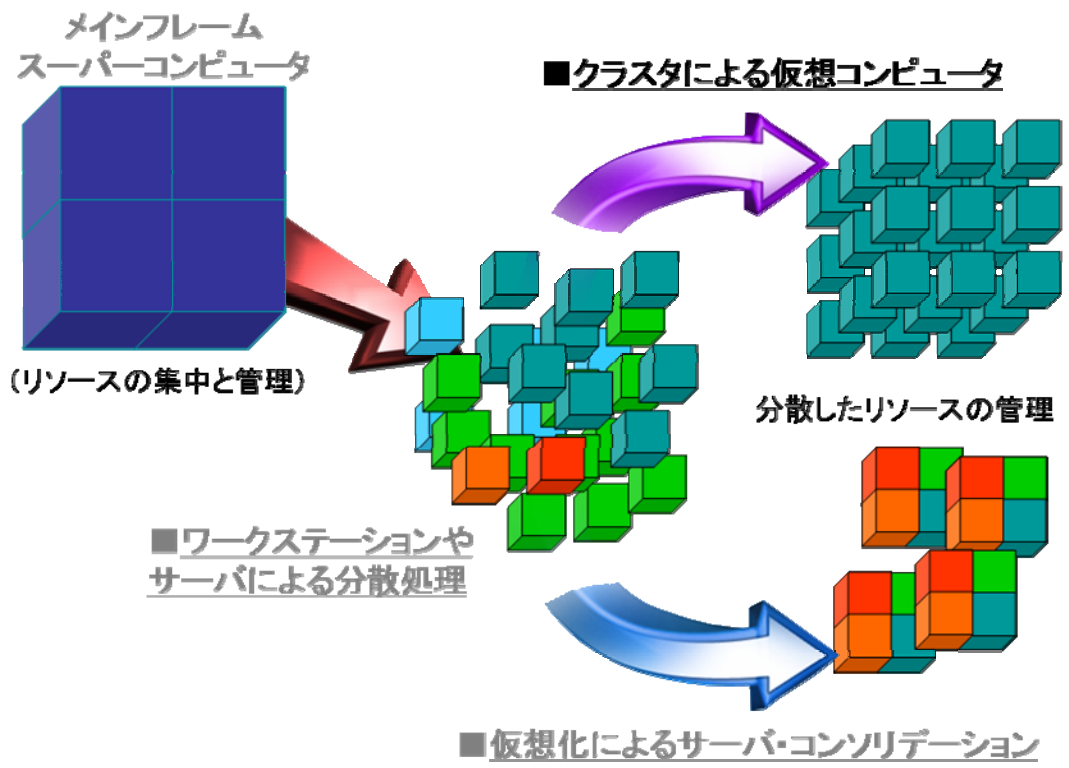
要求されるのはシステムの‘バランス’:このような HPC システムの状況の中で、バランスの取れたシステム構築が非常に重要になっています。

バランスの取れた HPC システムの構築において、従来はコンピュータシステムの開発・販売を行うベンダーが大きな役割を担っていました。しかし、現在では、この HPC システムのバランスは、その多くをマイクロプロセッサに依存しています。現在のマイクロプロセッサは、明らかにプラットフォーム・ソリューション戦略に基く、機能強化と開発が進められています。例えば、インテルは、そのマイクロプロセッサの発表に際して、単にマイクロプロセッサの動作クロックだけを説明するのではなく、マイクロプロセッサの機能強化に基づく、プラットフォーム戦略を明示しています。[1]

[1] *Ts for HPC: HPC プラットフォームにおけるインテル・テクノロジー
http://www.intel.co.jp/jp/business/japan/feature/HPC/scalablereport/index.htm

■HPC システムにおけるクラスタシステム

スーパーコンピュータやメインフレームで行われていた HPC のワークロードは、よりコストパフォーマンスの高いワークステーションやサーバでの分散処理によって処理されるようになりました。この分散処理でのワークステーション（やパーソナルコンピュータ）、サーバの分散処理でのリソースの管理とシステムの一体運用を可能とすることで、クラスタは、現在では HPC のワークロードに対する最も一般的なシュミレーションのためのプラットフォームとなっています。

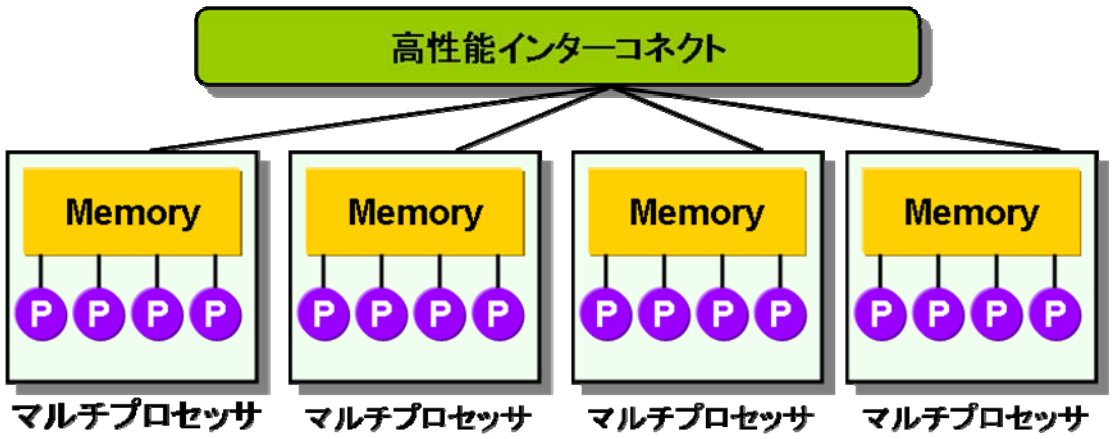


分散処理の広まりとその問題点：「もっと処理性能を」と「より安価に」を実現するために、ワークステーションやサーバでの分散処理の導入が進み、ユーザや企業に新たな価値をもたらしてはいますが、管理責任とその負担の分散を招き、結果的に運用管理コストを押し上げることになります。

このようなクラスタシステムの利用は、今後も HPC 分野における利用が進み、更に様々な用途で利用されることになります。

■HPC クラスタの構成

クラスタは、購入するものではなく、構築するものであるとよく言われてきました。しかし、現在では、各社が HPC クラスタとして製品展開を図っています。これは、HPC クラスタを構築するコンポーネントが、標準化され、クラスタ構築の技術も進歩することで、最小のカスタマイズだけでも、求められるワークロードに対応できるようになってきたためです。

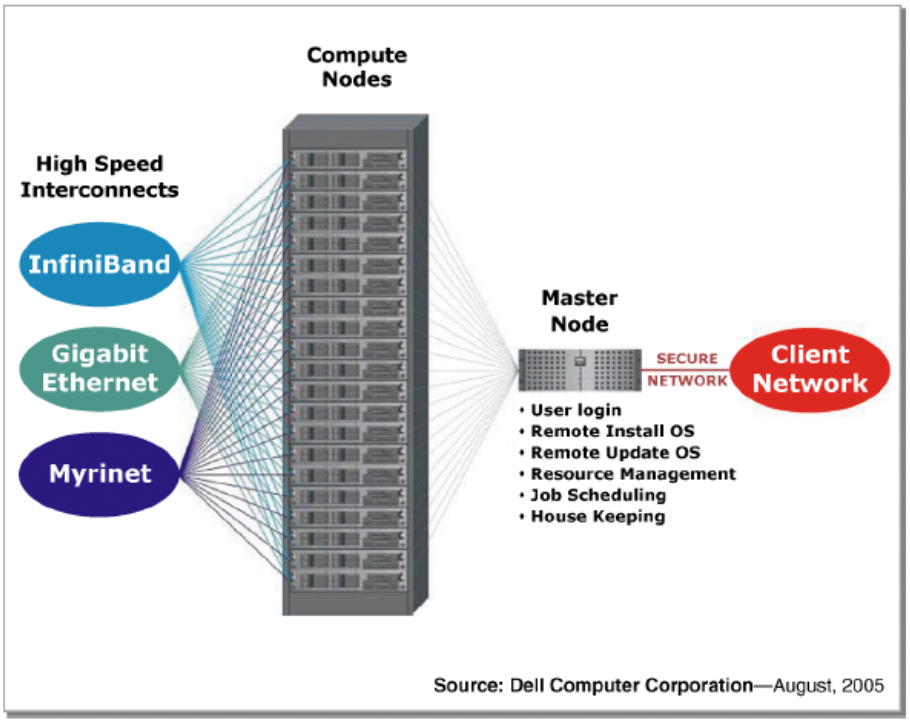


HPC クラスタの構成： クラスタ内の各ノードは高速インターコネクト・テクノロジで接続されます。InfiniBand や PCI Express テクノロジが登場する前は、独自規格に基づいた高性能で高価なテクノロジと、標準規格に基づいた低コストでやや性能の低いテクノロジのいずれかを選択する必要がありました。コスト制約の厳しいクラスタの場合は、ネットワーク接続用の Ethernet テクノロジが広く利用されていますが、これは並列アプリケーションのようにノード間の緊密な連携が要求される環境ではボトルネックとなります。InfiniBand ベースのインターコネクトを導入すれば、このようなトレードオフは解消されます。

また、HPC クラスタは一般的には、数多くのアプリケーションを実行するよりも、特定の少数のアプリケーション専用に利用される場合が多く、利用するアプリケーションの特性がシステムの選定時に明らかになっており、その特性に最適な構成をシステムの導入検討時に把握できることで、既にカスタマイズされたクラスタを購入して利用することを可能としています。

クラスタ構成では、要求されるインターコネクトの性能に応じて、インターコネクトの選択をすることが一般的です。小規模クラスタやアプリケーションの特性によっては、組み込まれている Gigabit イー

サネットで十分な場合もあります。アプリケーションの起動やリソース管理などは、Master Node として、特定のノードがその処理を行います。

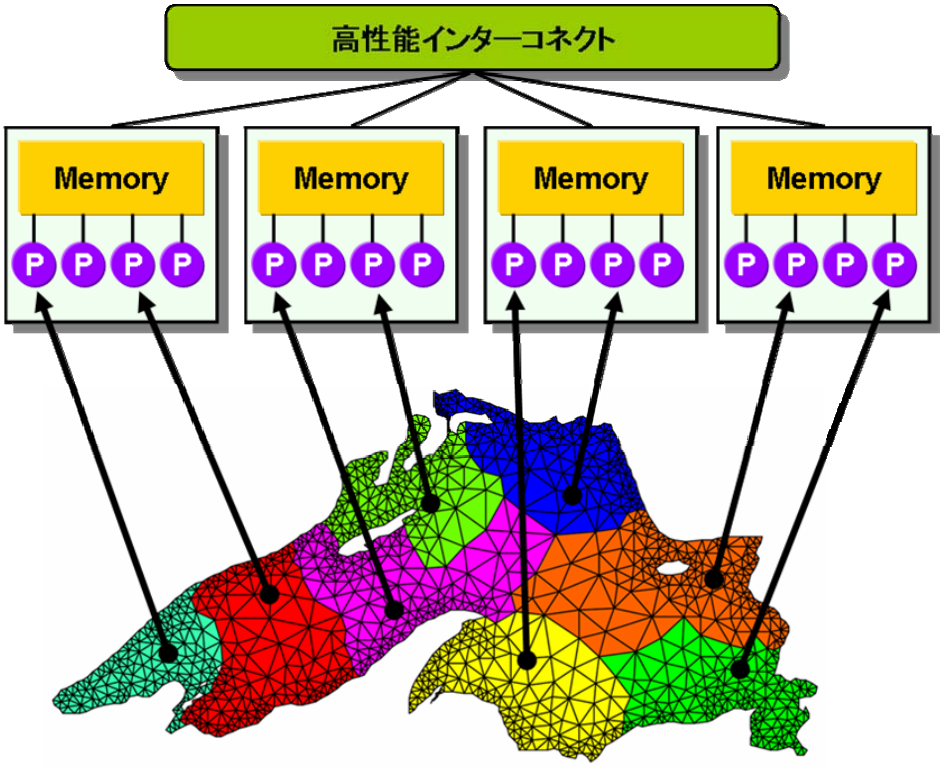


高い生産性の提供が求められる計算機リソースの導入の選択に際しては、単にシステムのピーク性能や一般的なベンチマークテストでのシステム評価では、不十分であり考慮すべき点も多々あります。解析対象に対するより詳細なモデル化や複雑な物理モデルの組み合わせによって、計算処理はますます複雑になり、かつ、それらはデータ集約型になります。

■HPC クラスタ上での MCAE アプリケーションの特性

HPC クラスタの選択において、並列アプリケーションの特性やその利用方法に関して、検討することが必要となります。アプリケーションの特性によって、クラスタの構成を決定することが重要になります。並列アプリケーションをクラスタ上で実行する場合、クラスタ上のノード間でのデータ転送が必要になります。このデータ転送をどのように行うかは、個々のアプリケーションによって、大きく異なります。クラスタ間で、頻繁にデータの送受信を行うのであれば、より高速のインターコネクトの選択が必要となります。

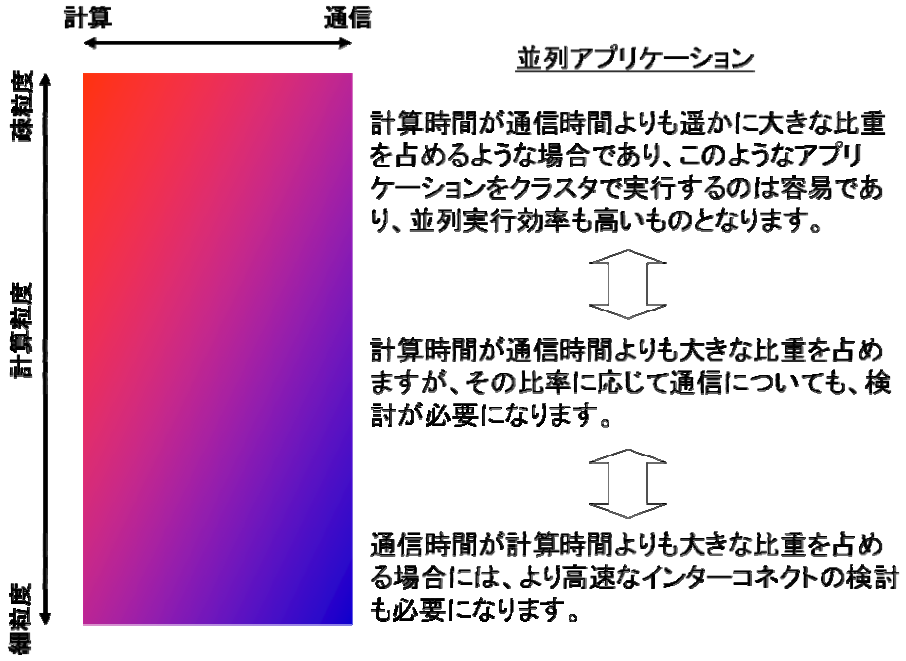
アプリケーションをクラスタで実行する場合、その並列実行の方法は様々なものが考えられます。計算対象を分割して、各領域を各ノードで同時に処理し、その計算結果をまとめるような領域分割計算や計算の入力パラメータを変更して様々な計算を行うようなケースもあります。



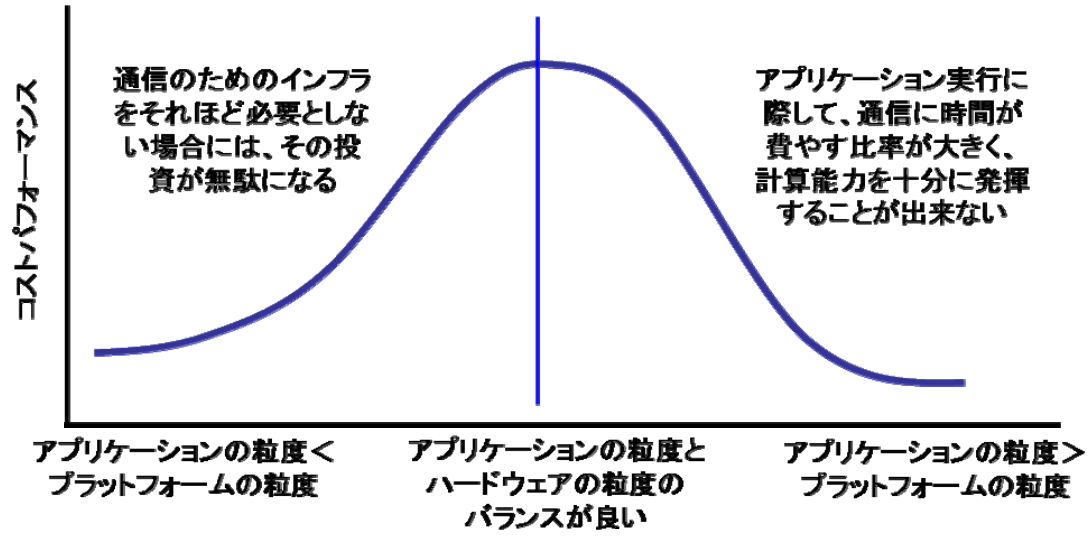
領域分割による並列処理： MCAE シュミレーションで利用されるアプリケーションは、計算モデルを分割し、それぞれの領域をプロセッサが分担して計算を行います。領域の分割、負荷の分散の均等化、高速での領域間でのデータ交換などにおいて、技術革新が続いており、計算精度、計算速度の双方ともに、著しい進歩があります。

実際には、同じアプリケーションでも、その粒度は、対象とする計算の規模や実行するプロセッサ数によって、変わってきます。より多くのプロセッサを利用して、解析を行う場合には、個々のプロセッサが担当する計算量が相対的に小さくなり、プロセッサ数の増加によって、プロセッサ間での通信が増えることで、計算と通信の比重は変わってきます。

このようなアプリケーションの粒度とプラットフォームの粒度（プロセッサの計算能力とプロセッサ間での通信性能の比率）のバランスが最適なシステムの構成には重要です。特に、システムの対費用効果（コスト・パフォーマンス）を考えた場合、このアプリケーションの粒度とプラットフォームの粒度を最適に選択すうことが必要となります。



アプリケーションの分類： クラスタ上でのアプリケーションの実行を理解するには、アプリケーション実行時の計算と通信の比重が重要になります。一般には、この計算と通信の比率によりアプリケーションの分類が出来ます。



アプリケーションとプラットフォームの粒度： 実際には、この粒度の問題は非常に複雑なものとなりますが、双方の粒度の特性を理解することは重要です。

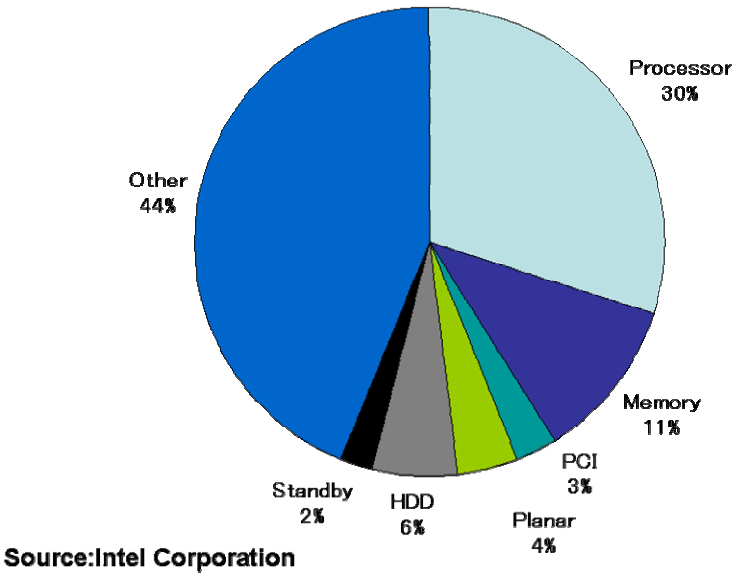
■HPC クラスタの課題 ～ より少ない TCO でのハイパフォーマンスの実現へ

クラスタシステムは、多くの計算資源を一箇所で集中管理し、これらの計算資源を同時に利用することで、大幅な処理性能の向上を図ることを可能としています。ただ、このリソースの集約は、その運用コストに於いて、電力とシステムの冷却という問題をより顕在化させることになります。また、継続的な処理能力と処理データの増大は導入されるコンピュータリソースの加速度的な増加をもたらしています。このため、より高密度でのシステムの実装が求められることになります。コンピュータを増設するために新しい建物を増設したりすることは、一般には不可能であり、コンピュータは既設の施設への設置が求められます。

このためには、従来以上のエネルギー効率と高密度の実装がシステムに求められています。コンピュータシステムにおいて、エネルギー効率の改善と消費電力や発熱量の削減において、最も重要なコンポーネントは、もちろんマイクロプロセッサです。マイクロプロセッサの開発において、そのエネルギー効率の改善に対する各社の取り組みは様々ですが、その設計方針の基盤のひとつになっていることは間違いありません。

エネルギー効率を追求するだけで、HPC システムの TCO を削減し、より高い性能を実現することは困難です。エネルギー効率の追求は、もちろん、重要なポイントですが、TCO 削減のためには、プラ

ットフォームとしても改善が必要です。



Source:Intel Corporation

サーバプラットフォームにおける消費電力： サーバシステムの消費電力に占めるマイクロプロセッサの比重は、必ずしもその大部分を示すものではありません。コンピュータの様々なコンポーネントとプラットフォームとしてエネルギー効率の改善を図る必要があります。
この図[2]で、“Other”で示す部分の比重が非常に大きくなっていますが、この部分には、プラットフォームレベルでの電力分配の非効率さと冷却のためのファンの電力消費が含まれています。プロセッサの消費電力当たりの性能向上を図るのと同時にプロセッサを含むシステムの発熱量の低減とプラットフォームとしての電力利用の効率化を図ることが必要となります。

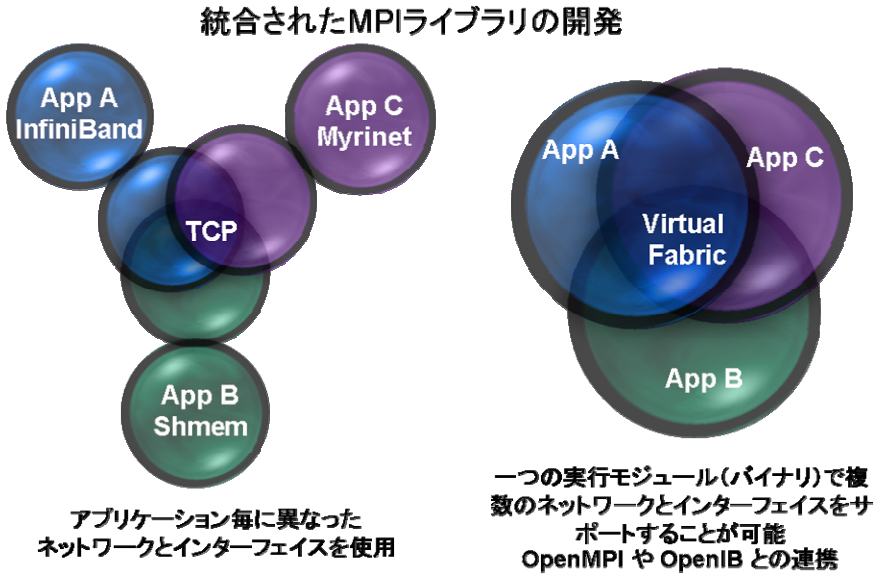
■HPC クラスタと COTS システム～インターコネクトの選択

標準コンポーネントによるHPC クラスタの構築では、非常に広範囲な選択肢があります。これは、良い点でもあり、また、問題でもあります。インターコネクトについても、一般のギガビット・イーサネットから Myrinet や InfiniBand といった高速のインターコネクトまで、その選択肢はいくつもあります。また、オペレーティングシステムにしても、オープンソースの幾つかのディストリビューションの選択、ク

[2] **Solving Power and Cooling** Challenges for High Performance Computing Intel Technical resource / http://www.intel.com/business/bss/products/server/hpc/

ラスタ構築のためのミドルウエアの活用の検討なども可能です。これらは、幅広い選択肢があるという利点がある反面、どのようなものを選択し、それを実装するのかを検討する必要があります。

また、クラスタでのアプリケーションの並列実行のための標準 API である MPI についても、実際には多くの実装があり、また、それぞれの MPI もネットワーク毎に異なったインターフェイスを用意しています。



システムの共通化の促進:統合された MPI ライブラリの開発

これらの多くの実装については、現在では、統一した実装によって、より広範囲なネットワークインターフェイスをサポートするような方向での検討も始まっています。例えば、MPI についても、OpenMPI として、様々な MPI の実装を統一することも検討されています。ネットワークインターフェイスについても、商用のインターコネクトでも各社が個別に開発していたドライバなどの統一を図る動きがあります。HPC クラスタのインターコネクトとして、注目されている InfiniBand でも、「OpenIB Alliance」が結成され、各種ネットワーキングプロトコルの実装に加え、InfiniBand ハードウェアドライバなどの開発での協力の動きがあります。標準コンポーネントでは、競合するところでは、競合するが、一方では、協力できる部分での協業によって、より良い製品を提供することも可能となります。

このように、標準プラットフォームでは、オープンソースなどの動きを見ながら、統一された規格や基準を最大限に活用したインフラの整備が進んで、今後、さらに高性能で、コスト・パフォーマンスに優れた製品の利用が可能となります。従来、クラスタは COTS (Commercial off-the-shelf 一般商用) コンポーネントで構成されるというような説明がなされています。現在では、この COTS の意味として

は、(Common off-the-shelf 共通)コンポーネント構成という位置づけが適当です。標準コンポーネントを最大限に活用することで、COTS の定義も変わってきています。

COTS システムの利点として、様々な技術を容易に組み合わせることが可能なことがあります。先に示したようにアプリケーションの実行特性に応じて、計算粒度は大きく異なります。また、同じアプリケーションでも、計算の規模や計算に利用するプロセッサ(コア)数の違いによって、十分な性能を発揮するために必要なインターコネクトの構成などは変わります。このように、様々な要因の影響を受けるアプリケーションの実行に際して、様々な技術を組み合わせることで、最も、コスト・パフォーマンスの高いシステムの構築が可能となります。

例えば、クラスタシステムを構築する場合、複数の並列ジョブを実行する場合、それぞれのジョブの実行を行う単位のノード内だけ、高速のインターコネクトで接続し、クラスタ全体を一般には、コストの高いこのような高速のインターコネクト用のハードウエアで接続すること無しでシステムを構築することで、システム全体の TCO の削減を図ることも可能となります。

■まとめとして

標準化されたコンポーネントによる HPC クラスタは、現在では小規模な部門クラスタからスーパーコンピュータクラスのシステムまで様々な構成と規模で構成されています。スーパーコンピュータクラスのシステムは、従来は特別なインターコネクトとシステムと従来は特別なインターコネクトとソフトウエアによって、システムが構築される場合が主流でしたが、現在では標準コンポーネントと標準の Linux オペレーティングシステムで構築される場合も増えており、今後も様々な用途での導入が進むと思われます。

高い生産性の提供が求められる計算機リソースの導入の選択に際しては、単にシステムのピーク性能や一般的なベンチマークテストでのシステム評価では、不十分であり考慮すべき点も多々あります。解析対象に対するより詳細なモデル化や複雑な物理モデルの組み合わせによって、計算処理はますます複雑になり、かつ、それらはデータ集約型になります。今後は、マイクロプロセッサの‘マルチコア化’によって、クラスタのノードはより強力になり、また、次世代の標準メモリテクノロジーなどにより、より大容量メモリの実装が可能となります。

現在、サーバシステムは次世代インフラへの進化が進んでいます。64 ビット、マルチコア、仮想化、省電力技術といった「テクノロジー」市場に投入されるなかで、個々の技術についてではなく、システムとしての次世代インフラを考える必要があります。HPC クラスタシステムの構築では、ある意味、ユーザが抱える本質的な問題や課題を解決するための IT インフラとして、HPC クラスタを考えると、標準化（COTS）されたコンポーネントでのシステム構築の価値は、今後一層大きくなり、それらをシステムとして、ユーザに提供することが今、求められています。