メインストリームを目指すHPC

廣安　知之

ハードウエアのコモディティ化と汎用的なソフトウエアが利用されることにより，HPCの利用が，一般のユーザーにまで拡がることが期待されている．本稿では，ハードウエア，ソフトウエアの両面で，HPCを行うスーパーコンピュータもしくはPCクラスタがメインストリームで利用されるに際してどのような課題が存在するかについて解説する．

キーワード：HPC，クラスタ，コモディティ，オープンソース，Windows

1．はじめに

新しい商品やサービス，技術などが登場した際に，早い段階でそれらに飛びつき購入・利用する人もいれば，逆に，多くの人が利用しないと購入・利用意欲がわかない人もいる．マーケティング用語では，前者の層に属する人は「アーリーアダプター」と呼ばれ，ここでの普及率がある程度に達すると，後者の層に属する「マジョリティ」への購入・利用が拡大し，商品やイノベーションは急激に普及・拡大するといわれている．このようにマジョリティに利用される市場が「メインストリーム市場」である．しかしながら，アーリーアダプターは，技術的に特化した部分を求めたり，多くの他のユーザーが利用していると利用意欲が減退するという特性があるので，あまりにアーリーアダプターに特化した対応を行うとなかなかマジョリティへの利用の拡大が進まず，商品やイノベーションがメインストリームに移行しづらいというジレンマがある．そのため，アーリーアダプターとマジョリティの間には「キャズム」と呼ばれる大きな谷間が存在するといわれ，メインストリームに移行するためには，このキャズムを越えていく必要がある[1]．

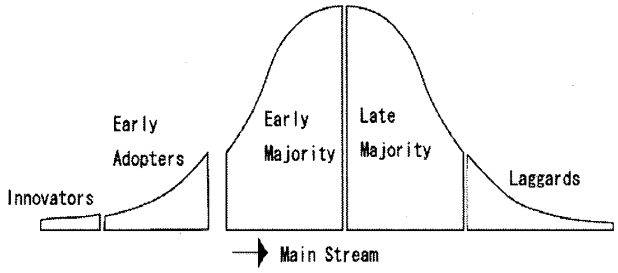

図1　メインストリーム

HPCとはHigh Performance Computingの略で，「自然現象のシミュレートや生物構造の解析など，非常に計算量が多い計算処理のこと．言葉の意味からいえば計算の目的は限定されないが，特に自然科学現象を解明するための計算を指す言葉として用いられる．(「IT用語辞典」[2])」と定義されている．この定義にあるように，計算量が非常に多い処理は10数年前までは，研究者が自然科学現象を解明するために利用するのが主で，一般の人が行うことはまれであった．これは，膨大な計算を行うことが可能な計算機を一般が利用することは難しかったことや，膨大な計算を利用して受けられるコストパフォーマンスに優れたサービスなどが存在しなかったことなどの理由によるであろう．ようするにこれまでHPCはアーリーアダプターが利用する技術であり，メインストリームになかなか移行できない状況であった．

これに対して，一般でも利用可能な計算機の性能の向上が飛躍的に増大した．さらに，オープンソースを中心としたソフトウエアの充実，Windowsなどの多くのユーザーが慣れ親しんだソフトウエアがこの分野へ進出してきた．これにより，いよいよHPCもメインストリームへと移行する下地ができつつある．HPCがメインストリームに移行すると，膨大な計算を行うユーザーが増加するものと予想される．このようにHPCがメインストリームへ移行すれば，本学会で議論されているアルゴリズムや多くの技術がますます利用されるようになり，多くの計算を行うことにより，アルゴリズムが同一でも新たな結果やパラダイムが生まれることが期待される．本稿ではHPCがメイ

ひろやす　ともゆき
同志社大学 工学部
〒610-0321 京田辺市多々羅都谷1-3

表1　第29回 Top 500 リスト（上位5位）

Rank	Site	Computer	Processors	GFlop/s
1	DOE/NNSA/LLNL United States	BlueGene/L IBM	131072	280600
2	Oak Ridge National Lab United States	Jaguar Cray Inc.	23016	101700
3	NNSA/Sandia National Lab United States	Red Storm Cray Inc.	26544	101400
4	IBM Thomas J. Watson Research Center United States	BGW IBM	40960	91290
5	Stony Brook/BNL United States	New York Blue IBM	36864	82161

ンストリームへ移行するための課題などについて述べる．

2．2極化するスーパーコンピュータ

HPCにおける処理を行う計算機はスーパーコンピュータ（スパコン）と呼ばれてきた．スパコンに要求される性能はその時代，時代によって変わるが，それらの性能の動向を理解するには，Super Computing Sites Top 500[3]のプロジェクトによって作られている世界のトップ500台の計算機のリストを利用するのが良いだろう．これは1年に6月と11月の2回，Linpack Benchmarkと呼ばれるベンチマークでスーパーコンピュータの性能をFlops/sで計測し，リスト化しているものである．表1に現在の上位5台のスーパーコンピュータを示す．詳細は省略するが，そこで使用されているCPUやネットワーク，筐体などは，すべてスーパーコンピュータ用に設計もしくはチューンアップされているものである．すなわち，そのようなハードウエアを利用しなければトップレベルのスーパーコンピュータの構築は難しいのである．また，表中のプロセッサー数に注目してみると，1,000個以上のCPUを利用している．1位のBlueGene/Lにいたっては13万個のCPUを利用している．近年，省電力，コンパクト化が急速に進んだが，それでも非常に大きな電力を必要とし，特別な施設を用意する必要がある．そのため，これらのトップレベルのスーパーコンピュータを構築するためには，CPUやネットワークの設計から始まり，場所の選定，施設の構築などが必要であり，莫大な費用が要求される．これらの理由により，トップレベルのスパコンの構築は，国家レベルのプロジェクトとなるのである．

現在，日本では次世代スーパーコンピュータの計画が行われている．これは，2012年の完成を目指して数十ペタフロップスの性能を有するスーパーコンピュータを構築するプロジェクトである[4]．日立製作所，NEC，富士通の3社が共同して開発し，神戸に設置されることが決定した．

すでに述べたようにこのようなトップレベルのスーパーコンピュータの構築は，国家戦略の一部として実現する必要がある．その際に問題になるのが，多額の税金を投入するために利用目的を明確にしなければならないことである．しかしながら，これまでの歴史を振り返ると，このような人類がまだ接していないレベルのイノベーションは，実際の所，何に使われるかははっきりとしないことが常であろう．パソコンや携帯電話がここまで一般の人に利用されると誰が予想したであろうか．インターネットやウェブが出現したときには，現在のような利用の想定はまったくされていなかったといえよう．地球シミュレータが設計されたときには，これほどのインパクトを世界に与え，民間企業にもスパコンの利用の可能性を示唆するきっかけになるとは多くの人が思わなかったに違いない．数十ペタフロップスの巨大マシンが登場することで，新しい利用方法が開発され，もしくは新しいパラダイムが生まれる可能性も大いに存在する．そのためには，次世代のスパコンにしかできない計算対象，計算アルゴリズムが必要であることは間違いなく，本OR学会に属する研究者のみなさんの成果が大いに期待されるところであろう．このような数十ペタフロップスという性能を有するスパコンをどこの国でも設計・製作する技術を有するわけではない．日本はこれまで，多くの努力によって世界有数のスパコンに関連した技術を培い，有してきた．これらの技術をさらに発展させ，展開させるだけでも大いに意義があると筆者は考える．何よりも世界トップレベルが狙えるイノベーションの創出は国家戦略として必須である．

それに対して，このようなトップレベルのスパコンは広く一般のユーザーが利用可能なものではない．ここで，図2に示したTop 500に掲載されているスパコンの性能の変遷に目を転じてみる．ここには1位のマシン，500位のマシンの性能および500位までの合計の変遷が示されている．

1位のマシンの性能はすでに述べたように特殊な設計が必要であるので，新しいマシンが登場するごとにその性能が急激に向上している．一方で，500位および500位合計の性能は，ログスケールで性能を示すと

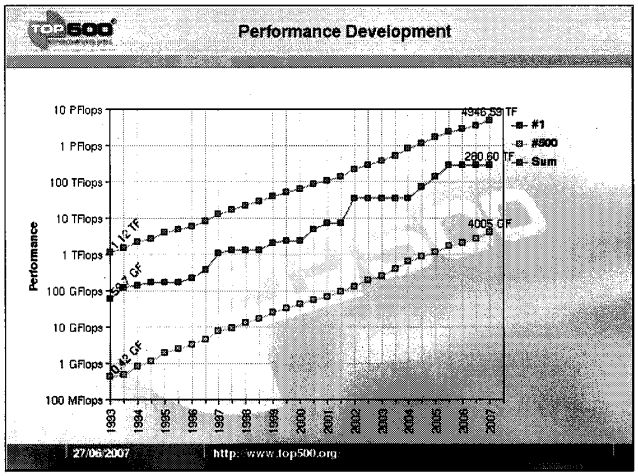

図2　スパコン性能の変遷（Top 500 リストより抜粋）

それらの性能は線形に伸びていることがわかる．これは時代，時代によって性能の向上を阻害する要因は異なるがそれに対して問題に対応してきた結果であり，今後もこの傾向は続くものと考えられる．このように今後もその性能の向上が期待される Top 500 のリストの後半に掲載されているマシンであるが，そこでは，我々の PC で利用されている汎用的な CPU やメモリ，ネットワークが利用され，コストパフォーマンスを高めている．このように，一般的な商品として手に入る商品・部品のことを「コモディティハードウエア」と呼んでいる．このコモディティハードウエアを利用することで，コストパフォーマンスの高いスパコンが構築できるのである．さらに，このコモディティハードウエアの性能は極めて高い．例えば，近年よく利用されている Intel 社の Core 2 Duo CPU の浮動小数点演算は 2 コア 4 演算が可能なので，クロック当り 8 演算が可能であり，3 GHz の CPU の場合，24 GFlop/s の理論性能を有することになる．理論性能通りに実行性能が得られることは難しいものの，1994 年までは 500 位のマシンの性能が 1 GFlop/s を切っていたことからもわかるように，一時代前のスパコンの性能を一つの CPU で有しているといっても過言ではない．また，昨年登場した新しいゲーム機で利用されている CPU も高い計算性能を有している．

このように，スパコンは，国家戦略に基づくトップレベルのスパコンとコモディティハードウエアを利用している下位レベルのスパコンとの 2 極化が進展している状態であるといえる．本稿で注目するのは，コモディティハードウエアを利用しているスパコンである．これらの利用がさらに浸透すれば HPC のメインストリーム化が達成され，これまで HPC を利用してこなかったユーザの利用も行われ，これまで見られなかった HPC の利用の展開が期待される．このような状況になれば，そこから新たな結果やパラダイム，イノベーションが生まれてくる可能性も大いにある．では，メインストリーム化するためには何が必要であろうか．それには，さらに手軽に利用可能なハードウエアの登場と使い勝手のよいソフトウエアが必要であると考える．

次節からはメインストリーム化するための鍵となるこれらのハードウエアおよびソフトウエアについて述べていきたい．

3．パーソナルクラスタ

現在の TOP 500 のうち 373 台がクラスタとよばれるアーキテクチャである．クラスタとはネットワーク接続された複数の計算機を仮想的に 1 台の並列コンピュータとして利用する形式のものである．近年，汎用的に利用されている PC の性能が飛躍的に向上したため，これらの PC を複数台ネットワーク接続することによりコストパフォーマンスにすぐれた並列計算機を構築することが可能となった．これが PC クラスタである．図 3 は代表的な PC クラスタの構成を示している．ユーザーはネットワークを通じて，ユーザノードと呼ばれるノードにアクセスする．ユーザは簡単な編集やコンパイル，デバッグをここで行い並列実行可能なファイルを生成する．一般にジョブスケジューラと呼ばれる仕事を管理するツールを通じて仕事が計算ノードにて処理される．PC クラスタにおいて最もノード数が多いのが計算ノードである．これら複数の計算ノード間でデータのやりとりを行いながら並列もしくは平行的に仕事の処理を行う．複数のユーザが存在するため，それらのユーザ情報の管理が必要になるが，それを行うのが管理ノードである．さらに先に述べたジョブスケジューラ，時間の管理を行うツールなど多くの管理ツールが必要となるがそれを行っているのが管理ノードである．そして，多くのデータの置き場所として分散ファイルシステムとしてのリポジトリノードが存在する．これらが基本的な PC クラスタの構成である．これらの複数のノードが存在するだけでなく，近年では，さらに CPU 内に複数のコアが存在するマルチプルコアが普及し始めている．CPU 内で発生するリーク電流などに起因する消費電力や発熱の問題が性能向上の大きな問題となり，かつ，従来から行われ

ている微細加工技術だけでは，クロック周波数を高めるのが非常に困難になっていることなどが大きな理由である．これらの問題に対処するために，最新のCPUでは，複数のコアを搭載したマルチプルコアの構成をとり，CPUのクロック周波数をあげることなく総合的にCPUの性能を向上させ，さらに消費電力あたりのCPU性能を向上させることができるようになってきているのである．

PCクラスタの高密度化，消費電力の問題への対応の結果，近年，「パーソナルクラスタ」と呼ばれる従来のサーバー1台よりも少し大きめの筐体に例えば4コアを有するCPUが2台搭載されたノードが6ノード配置されたPCクラスタが出現している．これだけで4×2×6＝48コアも存在し，立派な多ノードのPCクラスタであるといえる．Ciara Technologies社製のNexxusやTyan社のTyphoon，ビジュアルテクノロジ社のVT 64 Personal Clusterなどがそれに該当する．パーソナルクラスタの特徴は，先に述べたようにコンパクトな筐体に多くのコアが存在すること，高速ネットワークなども有したオールインワンタイプのものであること，静穏対策なども行われており個人のデスクサイドに置いて利用が可能であること，ユーザが必要なソフトウエアやアプリケーションなどが導入時にプリインストール済みであり，導入直後に利用可能であることなどが挙げられる．

このようなパーソナルクラスタの利用がメインストリームの市場で行われれば，現在のPCのように，部署ごとに，もしくは個人ごとにクラスタを有するもしくは利用することが可能となる．そのときに切っても切り離せないのが次節で述べるコストパフォーマンスに優れたユーザフレンドリなソフトウエアの存在であろう．

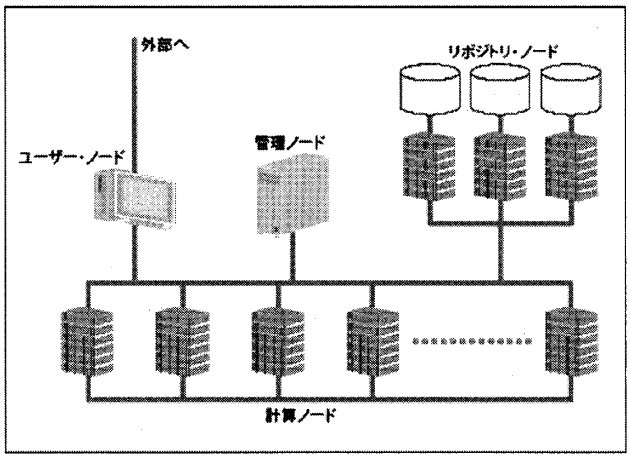

図3　PCクラスタ

4.　Linuxクラスタと Windowsクラスタ

HPCがメインストリームで利用されるためのもう一つのキーがソフトウエアである．前節で述べたPCクラスタを構築するためには，非常に多くのソフトウエアを用意する必要がある．これらのソフトウエアは大きくわけて，クラスタを構築する際，もしくはアップグレードする際に必要な管理ツールと，ユーザーが計算を行う際にシステムが有していなければならないツールが存在する．システムのインストーラやアップグレードするシステム，大規模なレポジトリを提供するファイルシステムは前者にあたる．また，PCクラスタのような分散メモリ型のマシンにおいて並列プログラムを構築するためにはメッセージパッシングとよばれる手法によるプログラムが必要となるが，現在標準で利用されているメッセージパッシング用のライブラリであるMPIや，複数人が複数の仕事を効率よくシステムを利用できるようにするためのスケジューラなどは後者にあたる．これらのソフトウエアをすべて用意しなければ実際にはシステムとして利用することができない．

一般のマジョリティのユーザーに利用してもらうためには，できるだけコストパフォーマンスに優れており，かつ，汎用的に利用可能なソフトウエアが必要である．

これまでPCクラスタにおいては，オープンソースの利用が中心的に行われてきた．特にOSにおいては，

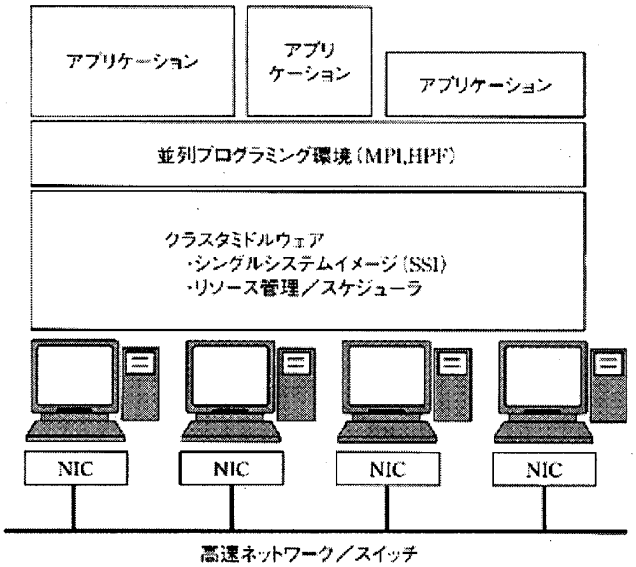

図4　PCクラスタに必要なソフトウエア

Linuxが非常に多く使われている．逆にいえば，このオープンソースの利用によって，高性能なクラスタがコストパフォーマンス良く構築できるようになったことで，TOP 500の半分以上を占めるようなPCクラスタの利用の普及が行われたものと考えられる．Linuxをベースとする PCクラスタ構築ツールは多々存在するが，その主なものには以下のようなものがある．NPACI (National Partnership for Advanced Computational Infrastructure) Rocks[5]は San Diego Supercomputing Centerを中心に開発されている非常に優れたPCクラスタ管理ツールである．高速にインストールが可能で，多くのジョブスケジューラに対応し，単純で簡潔なインストールを実現している．SCore Cluster System Software[6]は日本で開発されている非常に高性能なPCクラスタ用のプログラミング環境である．単一システムイメージを実現し，高速通信ライブラリであるPMを実装するなど高性能なPCクラスタを構築するのに適したツールである．筆者らもLinuxのディストリビューションの一つであるDebian GNU/Linuxで動作するDCAST (Dynamic Cluster Auto Setup Tool)[7]を開発している．

さらに，現在の企業や研究所，大学などで広く使われているOSの一つにWindowsがある．これはオープンソースではないが，一般の家庭にまで普及されている非常になじみのあるOSである．このWindowsでHPCを行うことが可能であれば，さらにHPCのメインストリーム化は進むであろう．

このような背景によりWindowsを開発しているマイクロソフト社では2006年10月にWindows Compute Cluster Server 2003 (Windows CCS 2003)[8]の発売を開始した．これはWindows Server 2003をベースにしたPCクラスタの管理・運営を行うためのソフトウエアである．クラスタの基本OSおよびドライバをインストールするDVDと並列処理に必要なライブラリやツールを提供するCD-ROMから構成されている．先に述べたLinuxをOSとするクラスタとWindows CCSをOSとするクラスタとの違いを述べるならば，WindowsはXPやVista, serverなどの多くのユーザーが慣れ親しんだGUIと同一のGUIを有することや，エクセルを始めとするアプリケーションの利用が可能であることがあげられるが，その利用方法が大きく異なる．

Linuxクラスタでは，ユーザーはユーザーノード(ヘッドノード)にて並列実行ファイルとデータを用意し，ジョブスケジューラを通じて計算ノードに対してジョブの投入を行う．それに対してWindowsクラスタでは，マイクロソフトが進めているサーバー利用の延長上にクラスタも存在すると考えられる．ネットワーク接続されたプリンタを例にすると考えやすいであろう．すなわち，管理者は，新しいプリンタがオフィスに導入されると，ネットワーク上でプリンタなどの資源を管理するアクティブディレクトリ (AD) に情報を登録し設定を行う．ユーザーが利用するワークステーション・PC側では，ネットワーク上の資源を利用する際にはADを参照するように設定をしておく．こうすることにより，新しいプリンタが導入されるごとにユーザー側で設定を行うことなくメニューからその資源を利用することが可能となる．ワードなどで文章を作り，プリントしたいときにプリントメニューからプリンターを選択して印刷を開始するのである．Windowsクラスタの基本的な利用方法もこれと同様である．Windowsクラスタがオフィスに導入されると管理者はオフィスのネットワークにクラスタを接続し，クラスタの情報をADに追加する．ユーザーは例えばVisual Studioなどの開発環境を利用してユーザー側のローカルなワークステーションもしくはPCで開発を行う．Linuxクラスタであれば，ソースやデータをヘッドノード（ユーザーノード）に移行してコンパイル実行するわけだが，Windowsクラスタでは，クライアント側で実行ファイルを生成し，そのままローカル側からプリントを行うイメージと同様に仕事の投入を行う．すなわち，メニューから仕事の投入を選択し，適切なクラスタを決定することによって仕事を開始するのである．オフィスに統一的なファイルシステムなどを導入することによりさらに連続した仕事がシームレスにつながっていく．例えば，CADでモデルを作成し，そのモデルをWindowsクラスタにてFEM解析することを考えてみる．CADでモデルを作成する際にWindows上で処理を行っているユーザも多くいらっしゃることであろう．そのモデルをネットワーク上のファイルシステムに保存する．メッシュ生成など時間のかかる処理を行う場合にはメニューからクラスタでの処理を選択し実行する．実際には，クラスタで処理が行われるがユーザは意識することなく処理が行われる．つづいて，FEMの設定画面にてファイルを開き，設定を行う．これらの処理もローカルで行われる．仕事の投入は再びメニューからクラスタ

での処理を選択し実行する．再度，結果はネットワーク上のファイルシステムに保存され，解析結果はクライアント側のアプリケーションにて行われる．このような一連の流れは，先に述べたネットワーク上のプリンタの利用と同一であることを理解していただけたことであろう．このようなシームレスな利用を行うためには，CADやFEMその他のアプリケーションが，Windowsクラスタにて稼動することだけでなく，クライアント側で利用するアプリケーションも開発し，メニューからプルダウンで利用することができるようにすることが必要であり，これがHPCのメインストリーム化を行う一つの鍵となる．初期段階ではすべてのアプリケーションをすぐに対応させるのは難しいかもしれない．その場合，キラーアプリとなりうるのはエクセルである．新しいエクセルのバージョンでは，Windowsクラスタに仕事を投入する機能が盛り込まれている．多くのクラスタを必要とするアプリケーションでは，パラメータの設定が主な入力要素となる場合が多い．そのため，クライアント側の設定アプリケーションとしてエクセルのファイルを用意しておき，そこからアプリケーションへの仕事の投入を上記のイメージで投入することが可能なのである．

5. HPCの今後の利用とまとめ

本稿では，HPCのメインストリーム化の流れとそのために必要なハードウエアおよびソフトウエアについて述べた．HPCがメインストリームにおいてさらに利用されるようになると，これまで見られなかったパラダイムやイノベーションが出現する可能性が大いにある．そこではOR学会などで議論されているアルゴリズムやアプリケーションがますます重要となることであろう．例えば，大規模なデータを蓄積して解析し利用することもますます容易になる．そのために，学習や最適化，データマイニングの分野での利用はHPCのメインストリーム化により進展するであろう．遺伝的アルゴリズムといったようなこれまで計算を膨大に必要とする手法も使い勝手が良くなるはずである．小野らはグリッドにまで拡張した遺伝的アルゴリズムのモデルの提案を行っている[9]．筆者らも多目的遺伝的アルゴリズムにおいて同様の提案を行っている[10]．これらの研究は並列処理を本稿で説明した環境で実行する際に参考になろう．

本稿で説明したとおり，HPCはメインストリームに進みつつある．これらの進展の草の根レベルで啓蒙活動を行うために，筆者らのグループでは超並列計算研究会[11]を定期的に開催し，自由な研究活動や講習会を行ってきた．さらに昨年度から同志社大学の寄付教育研究プロジェクトの一つとしてWindows HPCコンソーシアム[12]を立ち上げ，HPCのさらなる普及の活動を行っている．ご興味があればウェブページなどを参考にしていただきたい．

参考文献

[1] ジェフリー・ムーア：“キャズム,”翔泳社, (2002).

[2] IT用語辞典e-words, http://e-words.jp/w/HPC-1.html

[3] Top 500 Supercomputing Sites, http://e-words.jp/w/HPC-1.html

[4] 次世代スーパーコンピュータプロジェクト：http://www.nsc.riken.jp/p4.html

[5] NPACI Rocks, http://www.rocksclusters.org/wordpress/

[6] Score, http://www.pccluster.org/

[7] DCAST, http://dcast.sourceforge.jp/

[8] Windows CCS, http://www.microsoft.com/japan/windowsserver2003/ccs/default.aspx

[9] 小野功ほか：“Ninf-1/Ninf-Gを用いたNMR蛋白質立体構造決定のための遺伝アルゴリズムのグリッド化,”Vol. 46, No. SIG 12 ACS 11, 情報処理学会 (2005).

[10] 吉井健吾, 廣安知之ほか：“多目的遺伝的アルゴリズムにおける近傍交叉の効果,”Vol. 48, No. SIG 2 TOM 16, 情報処理学会 (2006).

[11] 超並列計算研究会：http://www.is.doshisha.ac.jp/smpp/

[12] Windows HPCコンソーシアム：http://www.is.doshisha.ac.jp/WindowsHPC/