

HPC ルームに導入された PC クラスタの性能評価

川村 暁・浅野 芳樹

Performance Evaluation of the NAS Parallel benchmarks on a PC Cluster at HPC Room

Satoshi KAWAMURA and Yoshiki ASANO

*Dept. of Information Technology and Electronics, Faculty of Science and Engineering,
Ishinomaki Senshu University MIYAGI 986-8580*

Abstract

In this paper, we describe the performance evaluation of the NAS Parallel benchmarks on the PC Cluster built at HPC room and computer systems for comparisons. As the results, performance of PC Cluster depends on architecture of CPUs and characteristics of programs.

1. はじめに

石巻専修大学の情報教育用の諸システムは、数年に一度の頻度で入れ替え（更新）がなされている。ある程度の大規模計算を必要とする教育研究に対応するため、前回導入のシステム（旧システム。平成 17 年 4 月稼働、平成 21 年 3 月まで利用）から、複数台の Linux 計算機を束ねた PC Cluster^{(1)~(3)}を導入した。これは、特殊な計算機環境を導入することは一般的な教育には適さないこと（特定環境に習熟する意味は情報教育としてはあまり意味がない）、分野を限らずに利用してもらうためには汎用的な環境が適していること、導入・構築・運用コストを最適化できることによる。

平成 22 年 4 月に稼働を開始した情報教育用の諸システム（新システム）においても、旧システムの思想を引き継ぎ、汎用的な PC（ただしサーバ機）を束ねた PC Cluster を導入した。新システムの PC Cluster は、システムインテグレータにすべてを任せるのではなく、必要な環境は、ユーザサイドで構築することとした。これは、PC Cluster を実現する手法が複数あり^{(1)~(4)}、一つだけの環境では利用環境として適さないことが最大の理由ではあるが、構築費用及び維持管理費用の低廉化も指向しているためである（大幅なコスト削減を果たした）。

本稿では、新システムで導入された PC Cluster をシステムとして捉え、その基本的な性

能を明らかにする。第 2 章で、構成した PC Cluster の概要（基本要素と成る計算機の詳細仕様を含む）を示し、第 3 章で、PC Cluster を構成する計算機および PC Cluster の性能評価に用いたベンチマーク手法について記す。第 4 章で PC Cluster を構成する要素計算機の性能について触れた後、各 PC Cluster の性能評価結果について考察する。第 5 章はまとめである。

2. PC クラスタの諸元および比較に用いた計算機環境

HPC (High Performance Computing) ルームに導入された PC Cluster 構築用の計算機と、性能の比較のために用いた計算機を表 1 に示す。これらの計算機を束ねて PC Cluster を構築するが、そのための OS 基盤として、Linux OS を用いている。

現在の CPU のトレンドとして、一つのパッケージの中に複数の CPU コアを作り込むマルチコア^{(5)~(7)}がある¹。図 1 に示すように、単一の CPU パッケージ IC の中に CPU コアが一つあるものをシングルコア CPU といい、複数のコアがあるものをマルチコア CPU という（図では CPU コアが 4 つあるクアッドコアを示した）。表 1 で示した PC クラスタを構成する計算機においても、HPC ルームに導入した計算機をはじめ、殆どの計算機がマルチコア（クアッドコア）となっている。さらに、同一 PC に CPU を複数搭載で



図1 シングルコア CPU とマルチ（クアッド）コア CPU の違い。一つの CPU パッケージに、何個の CPU コアがあるか（シングル＝1個、クアッド＝4個）が異なる。

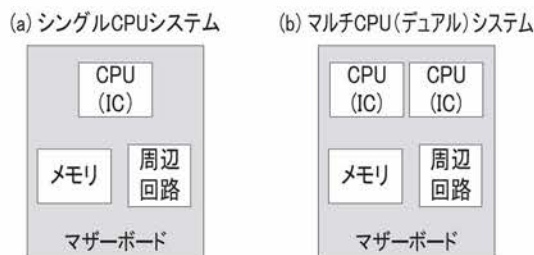


図2 シングル CPU システムとマルチ CPU システムの違い。一つのシステム（マザーボード）上に何個の CPU パッケージを搭載できるかの違いである。

きる場合もある（図2）。CPU を一つだけ搭載できる計算機をシングル CPU システム、CPU を複数搭載できる計算機をマルチ CPU システムという。

表1に示した計算機では3機種がマルチ CPU（デュアル）システムであり、残る2種はシングル CPU システムである。HPC に導入された計算機はシングル CPU システムであるが、マルチコア（クアッドコア）CPU であるので、1台あたり4 CPU 搭載していることと等価である。

表1を俯瞰すると、いくつか興味深い点がある。

- (1) 同じ Opteron 系の CPU で動作周波数と同じでも、Bogomips 値²が異なっている。
- (2) CPU のクロック周波数に応じた比率で BogoMips 値⁽⁸⁾ が計測されていない（同クロックでも値が異なっている）。
- (3) 各レベルのキャッシュ容量が大きく異なる（Opteron 系は3種類が同容量）。
- (4) メモリアクセス速度が大きく異なる。転送速度の高低（高：DDR3 1300～DDR1 333MHz：

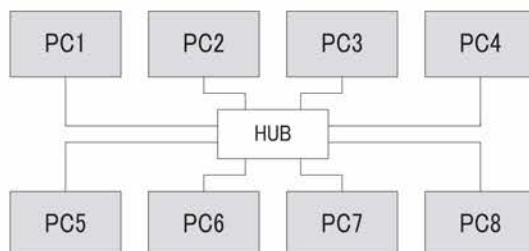


図3 HPC ルームに導入された PC クラスタの概念図。クアッドコア CPU を1基搭載した PC 8 台で PC クラスタを構成する。CPU（コア）数は $8 \times 4 = 32$ となる。

低)だけではなく、CAS レイテンシ（データを要求してからの待ち時間）も大きく異なる³。

(1)および(2)は、CPU アーキテクチャにより x86 命令の処理効率が異なること、キャッシュを含めた効率も世代により異なることなどが原因となる。特に、ビジーウェイトのような短い処理の場合、CPU の L1 キャッシュ（一次キャッシュ）に収まってしまうため、L2（二次）および L3（三次）キャッシュは計測値に殆ど影響しないと考えられる。

(3)および(4)は、CPU の動作速度向上に伴いメモリ転送がボトルネックになってきたため、キャッシュを介在させることで見かけ上のメモリ転送に伴う性能低下を隠蔽するための措置である。即ち、アクセス要求してからデータが CPU に届くまでの待ち時間を縮めるため、比較的高速なキャッシュメモリを搭載する考えを指す。これにより CPU は、（条件がよければ）低速なメモリではなくある程度高速なキャッシュのデータを利用できるため、メモリ転送に伴う性能低下を受けずに済む⁴。

PC Cluster は、図3に示すように、複数の計算機を通信路でまとめて単一の計算機システムとして扱う。このような形態の PC Cluster を構成するためには、複数の計算機を“まとめて一つの計算機システムのように見せるソフトウェア”を用いる必要がある。本稿では、表1に示すように、OS 基盤として用いた CentOS 5.5 で、パッケージ管理システム yum から容易に導入できる mpich2-1.2.1p1 を用いた。プログラムのコンパイルに用いるコンパイラには、gcc 4.1.2 を用い

表1 構成した PC クラスターの諸元。HPC ルームには、ML115 が 8 台 (CPU コア数は $4 \times 8 = 32$) 導入された。

クラスターの略称	32 bit環境 64 bit環境	ML115-32 ML115-64	Opteron-HE-Q-32 Opteron-HE-Q-64	Opteron-Q-32 Opteron-Q-64	Xeon-Q-32 Xeon-Q-64	Opteron-S-32 Opteron-S-64
機種名等		HP ProLiant ML115	OpteronHE-Q (自作)	Opteron-Q (自作)	Xeon-Q (自作)	HP ProLiant DL145
計算機の数		8	1	1	1	4
合計CPU数(最大)		32 (4core × 1CPU × 8台)	8 (4core × 2CPU × 1台)	8 (4core × 2CPU × 1台)	8 (4core × 2CPU × 1台)	4 (1core × 1CPU × 4台)
型番		Quad-Core AMD Opteron(tm) Processor 1354	Quad-Core AMD Opteron(tm) Processor 2379 HE × 2	Quad-Core AMD Opteron(tm) Processor 2378 × 2	XEON E5506 × 2	AMD Opteron(tm) Processor 244
CPU						
Clock Speed		2400MHz	2400MHz	2400MHz	2130MHz	1800MHz
Bus speed		1000MHz	2000MHz	1000MHz	4.8GT/s(2400MHz)	800MHz
L1 cache (D+I)		64Kbyte + 64Kbyte	64Kbyte + 64Kbyte	64Kbyte + 64Kbyte	32Kbyte + 32Kbyte	64Kbyte + 64Kbyte
L2 cache		1coreあたり 512Kbyte	1coreあたり 512Kbyte	1coreあたり 512Kbyte	1coreあたり 256Kbyte	1Mbyte
L3 cache		共有 2Mbyte	共有 2Mbyte	共有 2Mbyte	共有 4MB	-
cpu family		16	16	16	6	15
model		2	4	4	26	5
stepping		3	2	2	5	8
bogomips		4400	4800	4800	4256	3585
MotherBoard		FOXCONN ORIGINAL	SUPERMICRO H8DI+	Thunder n3600B S2927	Supermicro X8DAi-O	HP ORIGINAL
BIOS		american megatrends v02.61	american megatrends v02.61	american megatrends v02.61	american megatrends v02.61	american megatrends v02.54
Chipset		NVIDIA NFP3400	AMD chipset SP5100	nVidia NFP3600	Intel 5520	AMD-8000
Memory						
型番		HP DDR2 800MHz 1GB × 4	Kingston DDR2 800MHz 1GB × 4	Transcend DDR2 667MHz 512MB × 8	kingston DDR3 1333MHz 2GB × 6	HP DDR 333MHz 1GB × 2
搭載容量		4Gbyte	4Gbyte	4Gbyte	12Gbyte	2Gbyte
規格		DDR2 800MHz	DDR2 800MHz	DDR2 667MHz	DDR3 1333MHz	DDR 333MHz
CAS Latency		5.0	6.0	5.0	9.0	2.5
備考		dual channel	dual channel	dual channel	triple channel	dual channel
name		Seagate ST3160815AS	SEAGATE ST31000520AS	WesternDigital WD10EADS	SEAGATE ST31000520AS	SEAGATE ST340014A
Interface		SATA 3Gb/s	SATA 3Gb/s	SATA 3Gb/s	SATA 3Gb/s	UATA 100MB/s
HardDisk						
cache		8MB	32MB	8MB	32MB	2MB
Capacity		160GB	1TB	1TB	1TB	40GB
Spin Speed		7200RPM	5900RPM	5400rpm	5900RPM	7200rpm
seek time		11.0ms	8.5ms	8.5ms	8.5ms	8.5ms
Graphics		Matrox G200e 2MB	Matrox G200 16MB	XGI Volari Z7 16MB	ASUS EN9400GT 512MB	ATI Rage XL 8MB
Network Interface		Broadcom BCM5722 10/100/1000	Intel 82574L Gigabit Ethernet 10/100/1000	nVidia MCP55 Ethernet 10/100/1000	Intel 82573E Gigabit Ethernet 10/100/1000	Broadcom BCM5704 10/100/1000
OS	32 bit 64 bit	CentOS 5.5 2.6.18-194.8.1.el5 i686 athlon i386 CentOS 5.5 2.6.18-194.8.1.el5 x86_64				
Compiler		gcc 4.1.2				

表2 性能評価に用いた NAS Parallel Benchmarks (NPB) の諸元。問題のクラスは B とした。

略称	説明	特徴	クラス	サイズ	繰り返し数
EP	モンテカルロ法で用いられるような、2次元の乱数列の生成と統計量の計算。	ノード間の転送はほとんど必要としない	B	2147483648	1
MG	ルチグリッド法によるカーネルマルチグリッド法を用いたボアソン方程式の解法	浮動小数点演算性能の最大値を示す 3次元データを用いるため、多くの転送が発生		256 × 256 × 256	20
CG	共役勾配法を用いて正値対称な大規模疎行列の最小固有値の計算 (疎行列-ベクトル)	浮動小数点演算		75000	75
FT	FFTを用いた3次元偏微分方程式の解法	不規則な(長距離)通信 浮動小数点演算		512 × 256 × 256	20
IS	大規模整数ソート(パーティクル法)	(長距離)通信 浮動小数点演算		33554432	10
		整数演算スピードと通信性能の両方のテスト 整数演算(浮動小数点演算を含まない)			

ている。

3. NAS Parallel Benchmark (NPB)

計算機および計算機システムの性能を評価する場合、演算能力に基づく評価をベンチマークソフトウェアを用いて行う場合が多い^{(1)~(4)}。本稿では、性能の計測に用いるベンチマークソフトウェアに、NAS Parallel Benchmarks (NPB)⁽⁹⁾を用いた。

NPB(NAS Parallel Benchmarks)は、5つの Parallel Kernel Benchmarks と 3つの Parallel CFD (Computational Fluid Dynamics)

Application Benchmarks から構成されている。それぞれの問題に対し、問題サイズ・演算に用いる演算 (整数演算だけのものと浮動小数点演算を用いるものがある)・反復回数により規定される問題のクラスが5つ定義されている。各クラスは、小さな問題サイズであれば単一の計算機の性能評価に用いることが出来るが、大規模な問題サイズでは、並列計算機システムのシステムとしての性能を評価することが出来る。

本稿では、クラス B の問題サイズで、Parallel Kernel Benchmarks (EP, MG, CG, FT, IS) の5つを用いて実験及び評価を行った (表2)。

HPC ルームに導入された PC クラスターの性能評価

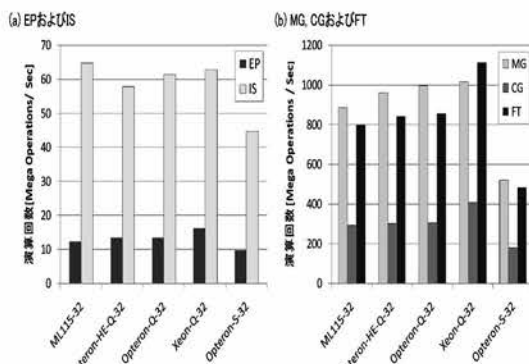


図 4 32 bit OS の場合の、NPB による単体 CPU の性能評価結果。単位時間（秒）あたりに何回問題要素の計算が出来たか（Mega Operations）で示した。グラフの値は 9 回計測した平均値である。

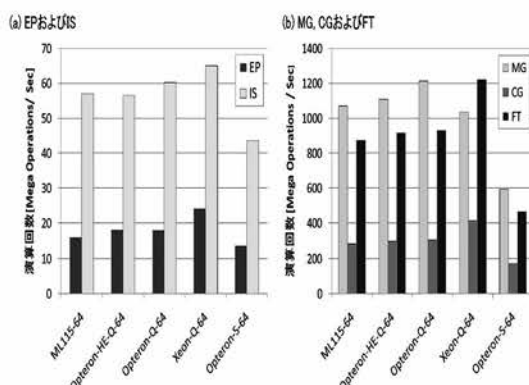
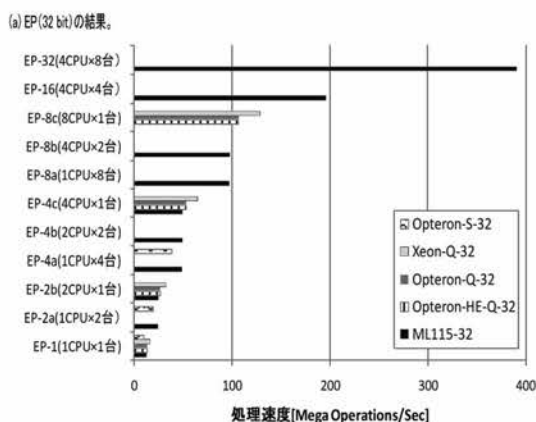


図 5 64 bit OS の場合の、NPB による単体 CPU の性能評価結果。単位時間（秒）あたりに何回問題要素の計算が出来たか（Mega Operations）で示した。グラフの値は 9 回計測した平均値である。

表 2 に示したとおり、問題の並列性により、通信路の性能が結果に大きく影響するものから、問題の並列度が高く（計算要素の間に依存関係が少なく）純粋な演算性能の合計をはかることの出来るベンチマークが提供されている。

4. NPB によるベンチマークの結果と考察

はじめに、PC Cluster を構成する計算機単体かつ単一 CPU 利用時の計算機のベンチマーク結果を示した後、PC Cluster の性能表結果を示し考察する。なお、PC Cluster の性能評価は、現在もっとも利用頻度が高い 32 bit OS 環境下で実施した。



(b) EP (32 bit) の結果、対数グラフで示した。

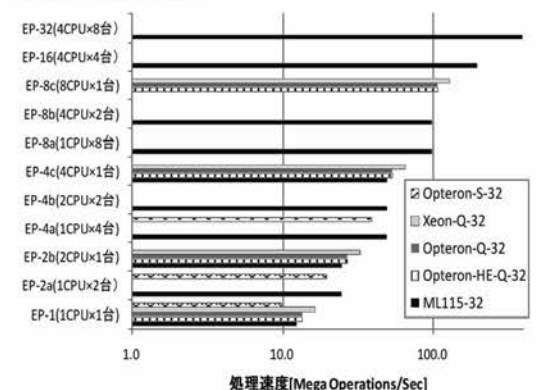


図 6 EP テストの結果（32 bit OS）。単位時間（秒）あたりに何回問題要素の計算が出来たか（Mega Operations/Sec）で示した。グラフの値は 9 回計測した平均値である。利用 CPU 数が増えるに従って処理性能が向上している。

4.1 CPU 当たりの性能の評価

32 bit OS 環境で一つの CPU コアを用いて計測した結果を図 4 に、64 bit OS 環境で一つの CPU コアを計測した結果を図 5 に示す。全体としては、64 bit OS 環境の結果が 32 bit OS 環境の結果より数値が高かった（処理が高速であることを示す）。ただし、結果を個別に見ていくと、IS テストでもっとも高速だったのは、32 bit 環境では ML115 だったのに対し、64 bit 環境では Xeon-Q であった。また、Opteron-Q の MG テストは 32 bit・64 bit を問わず結果がよい（32 bit 環境では僅差の二位、64 bit 環境では一位）が、FT テストおよび CG テストでは Opteron-Q の結果は Xeon-Q に遙かに及ばない。このよう

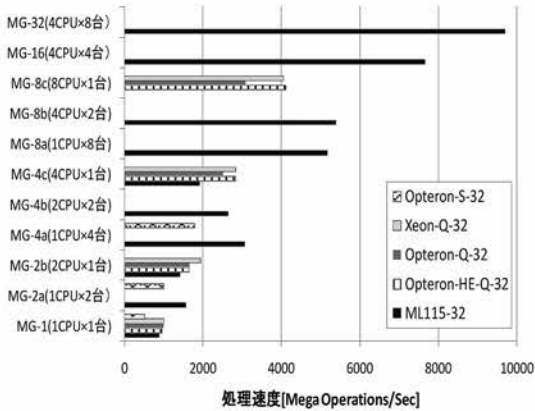


図7 MGテストの結果（32 bit OS）。単位時間（秒）あたりに何回問題要素の計算が出来たか（Mega Operations/Sec）で示した。グラフの値は9回計測した平均値である。図6のEPテストと比較して、利用CPU数が増えた際の処理性能の増加が鈍くなっている。さらに、計算に用いたCPU数が同一でも処理性能に差が見られる。

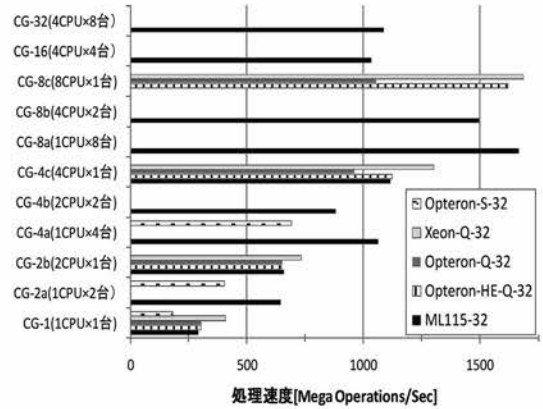


図8 CGテストの結果（32 bit OS）。単位時間（秒）あたりに何回問題要素の計算が出来たか（Mega Operations/Sec）で示した。グラフの値は9回計測した平均値である。通信路を介する形で利用CPU数が増えた場合に顕著であるが、処理性能が向上せず、かえって遅くなる（性能が低下する）場合があることが分かる。

に、テストの種類によって処理速度の順位が異なる結果を得た。これは、第2章（および表1）で触れたとおり、CPUアーキテクチャおよびメモリ周りの性能に起因すると考えられる。

なお、HPC ルームに導入されたML115の性能は、同じOpteronかつ同じ周波数を用いた計算機3機種の中では、EPテストでよい成績を取めた以外は、もっとも低速であった。この理由として、同じCPU family（同じ世代）のOpteronであっても、リビジョンが異なる（modelが2と4である）ことから、CPUアーキテクチャが改善された影響が考えられる。同じCPUだからといって、性能の特徴の出方が同じではないというよい例となっている。

4.2 PC Clusterの性能評価

表1に示すようなPC Clusterと比較のための計算機を用いて、HPC ルームに導入されたPC Clusterの性能評価を行った結果を、テスト毎に図6～図10に示す。なお、これらの結果は、32 bit OS環境下で行った。得られた結果について考察する。

図6のEPテストの結果から、CPU数が増えると単調に処理性能が増えること、性能の増え方が利用したCPU数にだけ依存することが分かる。結果を片対数で示した図6(b)から明らかで

ある。EPテストは、表2に示したとおり、並列計算に分割可能な整数演算を用いているため、純粋にCPU数の優位を生かすことが出来たと考えられる。

図7のMGテストの結果から、EPテスト（図6）の結果とは少し異なっている。計算に用いるCPU数が増えてもEPテストほど演算性能が伸びていないこと、利用するCPU数が同じでも、利用したCPUが同一筐体内にあるかネットワークで結ばれているかによって性能が異なっている。たとえば、4CPUに対するテストでは、同一CPUコア内にある4CPUを用いた場合（MG-4c）が性能がよく、ネットワークで結ばれた4台の計算機にあるCPUを用いた場合（MG-4a、1CPU×4）が性能が悪くなっている。また、同一計算機内にクアドコアCPUが2基搭載されているOpteron-HE-QとOpteron-Qは、4CPUまでの計算ではML115と性能に差がないが、8CPUになると、要素計算機を2台束ねて8CPUとした場合（MG-8b、ML115）だけではなく要素計算機を8台束ねて8CPUとした場合（MG-8a、ML115）にも劣る結果と成っている。この結果から同じ問題でも、問題サイズに適した計算機の形態（組み合わせ方）があること、純粋にCPU数だけでは性能を考えてはならないことが分かる。

HPC ルームに導入された PC クラスタの性能評価

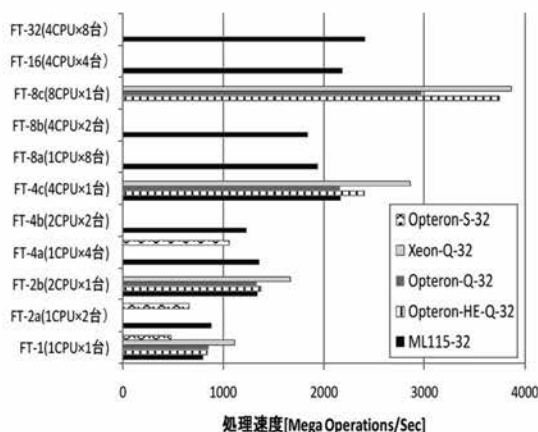


図9 FTテストの結果（32 bit OS）。単位時間（秒）あたりに何回問題要素の計算が出来たか（Mega Operations/Sec）で示した。グラフの値は9回計測した平均値である。通信路を介する形で利用CPU数が増えた場合に顕著であるが、処理性能が向上せず、かえって遅くなる（性能が低下する）場合があることが分かる。

図8のCGテストの結果は、これまでの結果よりおもしろい結果と成っている。もっとも高速だったのは同一筐体に8 CPU コアがある Xeon-Q-32（CG-8C、Xeon-Q）であり、第二位が1 CPU×8 台の計算機の構成を取った ML115（CG-8a、ML115）、第三位が Xeon-Q と同様に同一筐体に8 CPU コアがある Opteron-Q-32（CG-8C、Opteron-HE-Q）となった。計算をするためには通信路を経由するはずの1CPU×8 台の構成の ML115 と同一筐体に8 CPU コアがある Xeon-Q-32 がほぼ同速度と成ったのは興味深い。

図9のFTテストの結果は、これまでの結果以上に並列計算機についての示唆に富んでいる。それは、計算に用いるCPU数を増やしてもCPUの接続形態（同一筐体内にあるのかネットワークを経由するのか）の違いによって、計算速度に大きな影響を与える点である。特に、32 CPU（4 CPU×8）の ML115 の結果（FT-32、ML115）が、同一筐体に8 CPU 搭載している計算機3機種にも劣る結果は好例である。さらに、同じ ML115 を用いた場合でも、同一筐体内の4 CPU を用いた結果（FT-4c、ML115）と16 CPU（4 CPU×4）の結果（FT-16、ML115）がほぼ同値であることは示唆に富んでいる。電力効率で言えば FT-16 の方が大幅に悪い結果であり、問題

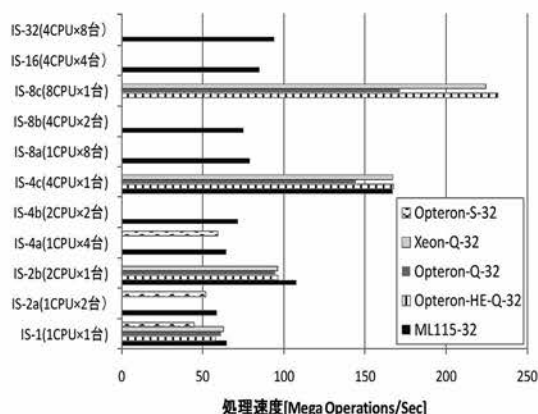


図10 ISテストの結果（32 bit OS）。単位時間（秒）あたりに何回問題要素の計算が出来たか（Mega Operations/Sec）で示した。グラフの値は9回計測した平均値である。通信路を介する形で利用CPU数が増えた場合に顕著であるが、処理性能が向上せず、かえって遅くなる（性能が低下する）場合があることが分かる。

の性質とアルゴリズムの重要性を示唆している（如何に並列計算向きのプログラムを書くことが出来るかが重要である）。

図10のISテストの結果は、並列計算機の特性とCPU世代の影響についてもっともよく表している。ML115で32 CPUを用いた場合の結果（IS-32、ML115）はOpteron-HE・Xeon-Qの約2.6分の1の性能しか出ていない。使用した計算機の数が増えれば性能差が2.6倍であるから、10倍程度効率が悪いことになる。ML115の中で比較しても、ただか2 CPUを用いた場合（IS-2b、ML115）の方が、32 CPUを用いた場合（IS-32、ML115）よりも高速である。ISテストは、ESテストと同様の整数演算に基づくテストであるが、ネットワーク転送が多い点が、結果の違いになって現れたと考えられる。

実験結果について考察するため、1 CPUで計算した結果を1.0とした、相対的な評価を試みた結果を表3に示す。表3は、問題の計算に用いたCPUの数を増やしたときに得られる性能が、1 CPUのときに得られる性能の何倍になっているかを表している。なお、図6～図10の結果から、計算に用いるCPU数が同じでも計算結果に差がある場合があることが明らかとなっているため、表2の元と成る数値には、計算に用いたCPU数が同じもののうちもっとも計算結果がよかった場

表3 CPU数を増やした場合の、各テスト毎の処理性能向上の割合。処理性能が、CPU数=1の場合の何倍になっているかで示した。なお、それぞれの数値の導出には、CPU使用数が同一のうちのもっとも処理性能がよかった場合の値を用いている。

(a) EP

	1CPU	2CPU	4CPU	8CPU	16CPU	32CPU
ML115-32	1.00	1.99	4.00	7.95	15.91	31.78
Opteron-HE-Q-32	1.00	2.00	3.99	7.99		
Opteron-Q-32	1.00	1.99	3.99	7.98		
Xeon-Q-32	1.00	1.99	3.98	7.90		
Opteron-S-32	1.00	1.97	3.96			

(b) MG

	1CPU	2CPU	4CPU	8CPU	16CPU	32CPU
ML115-32	1.00	1.77	3.46	6.09	8.66	10.96
Opteron-HE-Q-32	1.00	1.71	2.95	4.28		
Opteron-Q-32	1.00	1.66	1.66	2.52		
Xeon-Q-32	1.00	1.91	2.80	3.99		
Opteron-S-32	1.00	1.94	3.46			

(c) CG

	1CPU	2CPU	4CPU	8CPU	16CPU	32CPU
ML115-32	1.00	2.26	3.82	5.72	3.54	3.73
Opteron-HE-Q-32	1.00	2.15	3.74	5.38		
Opteron-Q-32	1.00	2.13	3.15	3.46		
Xeon-Q-32	1.00	1.79	3.18	4.12		
Opteron-S-32	1.00	2.24	3.85			

(d) FT

	1CPU	2CPU	4CPU	8CPU	16CPU	32CPU
ML115-32	1.00	1.68	2.71	2.42	2.73	3.01
Opteron-HE-Q-32	1.00	1.64	2.85	4.45		
Opteron-Q-32	1.00	1.56	2.52	3.47		
Xeon-Q-32	1.00	1.49	2.57	3.47		
Opteron-S-32	1.00	1.37	2.19			

(e) IS

	1CPU	2CPU	4CPU	8CPU	16CPU	32CPU
ML115-32	1.00	1.66	2.58	1.22	1.31	1.45
Opteron-HE-Q-32	1.00	1.67	2.90	4.01		
Opteron-Q-32	1.00	1.55	2.36	2.80		
Xeon-Q-32	1.00	1.53	2.66	3.57		
Opteron-S-32	1.00	1.16	1.33			

たとえば、8CPUの場合、

(i) 1CPU×8台 (ii) 2CPU×4台

(iii) 4CPU×2台 (iv) 8CPU×1台

の、4つの構成が考えられる(クラスタにより異なる)。よって、数値の導出のための前提条件が異なる場合がある。本表では、もっとも性能が高かった数値を用いている。

合を採用している。

表3(a)が、並列計算においてももっとも理想的な結果である。計算に使用したCPU数の増加に従って、性能が線形に増加していることが分かる。これに対して、表3(b)～表3(f)は、並列計算においては一般的な場合である。即ち、計算に用いたCPU数が増えても性能が向上しないばかりか、かえって性能が低下する場合があることを示している。表3(e)は全体的に、性能の向上割合が低く、並列化しても性能が殆ど向上しないことを示している。

PC Clusterをはじめとする並列計算機で並列に計算を行う場合、問題を、並列計算機を構成する個々の計算機を単位として、分解して計算できる必要がある。このとき、計算要素の間に依存関係がある場合や、大規模行列の演算のようにネットワーク越しに、別の計算機が管理しているメモリにある要素を計算で用いるようなアルゴリズムは、並列計算にはあまりふさわしくない。この様子は、本稿で用いたNPBの説明(表1)からも明らかであろう。表3(a)は並列に分解できる、並列計算に適した問題かつアルゴリズムも並列化されているのに対し、それ以外の問題は、プログラムは並列計算機で実行できるようになってはいないが何らかの問題(要素を相互参照する必要がある

場合・頻繁なデータ転送のためネットワークに負荷がかかる場合)があると示唆される。逆に、NPBは、これらの点を踏まえたベンチマークであることから、並列計算機の性能の評価に適していると考えられる。

全体を俯瞰してみると、第2章で示した様々な要因(CPUアーキテクチャの違い、メモリの違い、キャッシュの違い)と、並列計算機の一つであるPC Clusterに由来する要因(ネットワークで複数の計算機を結合している)、および、プログラムの要因(並列計算向きの問題か向いていない問題か、問題を書き下したアルゴリズムが十分に並列化されているかどうか)により、得られる性能に大きく影響することが明らかとなった。

5. まとめ

HPCルームに導入されたPC Clusterについて、並列計算機システムとしてのベンチマークを、NAS Parallel Benchmark (NPB)を用いて行った。PC Clusterは、導入経費および維持経費削減のため、筆者らが構築を行った。NPBによるベンチマークの結果、計算に利用するCPU数を増やしても、1CPUの場合の演算性能を線形にスケールアップした結果が得られるわけではないこと、同じCPUアーキテクチャでも性能の

傾向が異なることが示された。

並列計算機である PC Cluster の性能を十全に発揮させるには、並列計算に向けたアルゴリズムを用いたプログラムを用いる必要性が、強く示された。

文献

- (1) Thomas L. Sterling ら著、北野 宏明ら翻訳 (2001) PC クラスタ構築法 Linux によるベオウルフ・システム、産業図書
- (2) 石川裕ら著 (2002) Linux で並列処理をしよう - SCore で作るスーパーコンピュータ、共立出版
- (3) 小畑正貴 著 (2001) たのしくできる並列処理コンピュータ、東京電機大学出版局
- (4) William Gropp ら著、畑崎 隆雄 (翻訳) (2002) 実践 MPI-2 メッセージパッシング・インターフェースの上級者向け機能、ピアソンエデュケーション
- (5) 中森 章 (2010) マイクロプロセッサ変遷史 / 1990 年代～2000 年代、p.44～p.63、Interface、2010 年 10 月号、CQ 出版
- (6) David A. Patterson、John L. Hennessy 著、成田 光彰 訳 (2006) コンピュータの構成と設計～ハードウェアとソフトウェアのインタフェース 第 3 版 (上)、日経 BP 社
- (7) David A. Patterson、John L. Hennessy 著、成田 光彰 訳 (2006) コンピュータの構成と設計～ハードウェアとソフトウェアのインタフェース 第 3 版 (下)、日経 BP 社
- (8) BogoMips mini-Howto, Linux JFURL, <http://www.linux.or.jp/JF/JFdocs/BogoMips/>
- (9) NAS Parallel Benchmarks (URL), <http://www.nas.nasa.gov/Resources/Software/npb.html>

註

- 1 単一の CPU コアで性能を向上させることが、プロ

セス世代が進む毎に増大するリーク (漏れ) 電流と素子密度増大に伴う発熱量の急増 (単位面積あたりの発熱量はホットプレートを超えている)、動作時の電力消費及び発熱を抑えるために低下している動作電圧、等のため、非常に難しくなっている。このため、そこそこの性能を持つ CPU コアを複数個同一ダイ上に載せることによって、処理性能を稼ぐ方向にシフトしつつある。

2 BogoMips とは、Linux kernel がブート時に、ビジーウェイト (kernel 2.2.14 から変更はない) を用いて、負荷が殆どかかっていないときの CPU の大まかなループ速度を計測した値である。Bogo についているように、単なるビジーウェイトとほぼ等価であるため、CPU の実処理速度の正確な指標とはなりえない。しかしながら、CPU キャッシュ等の初期化処理後の計測であるため、CPU や CPU キャッシュが正常に動作しているかどうか判定することは可能である。

3 たとえば、計算機 A (DDR2 667MHz・CAS Latency 5.0) と 計算機 B (DDR1 333MHz・CAS Latency=2.5) の場合、メモリ上にあるデータを CPU が要求してから待ち時間は、どちらも同値である (計算機 B の CAS Latency を計算機 A の条件に換算すると、 $333\text{MHz} : 667\text{MHz} = 2.5 : x \therefore x \approx 5.0$ となる)。ただし、ある番地のデータから連続したデータを読み出す場合にはバースト転送となるため、単位時間当たりでは計算機 A が計算機 B のおよそ 2 倍のデータ転送能力があることになる。

4 高速に動作するキャッシュを実装するのはコストがかかるため、複数段のキャッシュを用いる手法が用いられている。低容量だが非常に高速な L1 (一次) キャッシュ、若干速度は落ちるが容量の多い L2 (二次) キャッシュ、比較的低速 (メモリよりは高速) で比較的容量の多い (メモリよりは小さい) L3 (三次) キャッシュである。容量と速度には、以下の関係が成り立つ。

容量 : $L1 < L2 < L3$, 速度 : $L1 > L2 > L3$