

高性能 PC クラスターの構築とその性能評価

High performance PC Cluster system and its verification

石原 孟¹⁾ ○ファフックバン²⁾ 藤野 陽三³⁾
Takeshi ISHIHARA¹⁾ Pham Van PHUC²⁾ Yozo FUJINO³⁾

1. はじめに

近年コンピュータ及びネットワーク機器の低価格化により安価な PC クラスターが構築可能になってきた。またネットワーク技術の進歩により分散メモリ型クラスターの性能も共有メモリマシンに遜色はないレベルまで上がっており、並列計算に関する研究も盛んに行われている^{1)~4)}。しかし、PC クラスターを自作するには幾つか解決すべき問題がある。例えば、PC クラスターに使われるネットワークには Fast Ethernet (100B-T), Gigabit Ethernet (1000BASE-T), Myrinet などがあり、また計算ノードには Single CPU, Dual CPU の選択肢もある。PC クラスターを構築する際にはこれらの組み合わせが重要となる。

そこで、本研究では PC クラスターを構築し、流体解析コードを用いてシステムの性能を評価するとともに、ピーク性能とコストパフォーマンスの両面から、安価で高性能な PC クラスターを構築するためのポイントを明らかにする。そして小規模な PC クラスターの性能からより大規模なクラスターの性能を予測する簡易式を提案する。

2. PC クラスターの構築

PC クラスターを構築するに当たり、まず CPU の選定を行った。Single CPU マシンに比べ、Dual CPU マシンはコスト及び占有空間がほぼ半分であることから、計算ノードに Athlon MP 2000+ を搭載した Dual CPU のマシンを採用し、マシン間をネットワークで繋ぐ。ネットワークが PC クラスター

性能に与える影響を調べるため、現在よく用いられている 3 種類の代表的なネットワーク、Myrinet, Gigabit Ethernet, Fast Ethernet を用いた。PC クラスターで用いた計算ノードのハードウェア構成を Table 1 に示す。

Table 1 Cluster node specification

CPU	Athlon MP 2000+ ×2
Motherbord	AMD 760MPX chip set
Memory	1GB DDR SDRAM (PC2100)
NIC 1	Fast Ethernet
NIC 2	Gigabit Ethernet
NIC 3	Myrinet2000
OS	Red Hat Linux 7.2

3. 性能評価

ベンチマークテストには非構造格子対応で領域分割法を採用している汎用 CFD コード FLUENT[®]を用いた。テストの問題には大きさの異なる 2 つの問題を選定した。その詳細を Table 2 にまとめた。

Table 2 Problem details

	Cells	Turbulence model	Solver
Small Class	32,000	k-epsilon	segregated implicit
Medium Class	242,782	k-epsilon	segregated implicit

Figure 1 には 4 計算ノード (8CPU) を有する並列計算機の加速率 (Speed Up) を示す。Myrinet ではミディアムクラスで 6.38 倍、スモールクラスで 5.35 倍のよい性能を示す。一方、Fast Ethernet ではミディアムクラスで 5.35 倍の性能を示すが、スモールクラスで 0.95 倍となり、その性能が急速に悪化している。これは通信時間の割合が計算負荷の減少

1) 東京大学大学院工学系研究科社会基盤工学専攻 助教授
Associate Professor Dept. of Civil Eng., Univ. of Tokyo
2) 東京大学大学院工学系研究科社会基盤工学専攻 大学院生
Graduate Student Dept. of Civil Eng., Univ. of Tokyo

3) 東京大学大学院工学系研究科社会基盤工学専攻 教授
Professor Dept. of Civil Eng., Univ. of Tokyo

に伴い、相対的に大きくなったことによるものである。Gigabit Ethernet ではミディウムクラスで5.67倍、スモールクラスでも3.37倍の性能を得ており、更に Table 3 に示すコストパフォーマンスを考え合わせると、Gigabit Ethernet の採用はよい選択と言える。ただし、ミディウムクラス以上の計算のみを行う場合には Fast Ethernet がよく、またスモールクラスの計算のみを行う場合には Myrinet が良い。

Table 3 Cost performances for several networks

Network	Network / Machine	Speed Up	
		Small Case	Medium Case
Fast Ethernet	1 %	0.95	5.35
Gigabit Ethernet	30 %	3.37	5.67
Myrinet	133 %	5.35	6.38

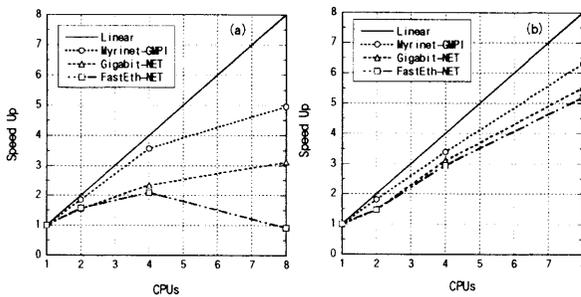


Figure 1 Performance of Network Interface Card

a) Small Class; b) Medium Class.

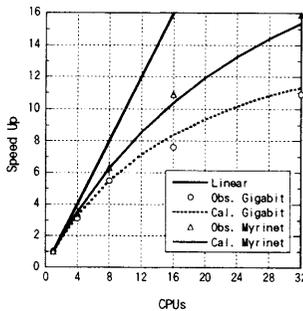


Figure 2 Comparison of estimated and measured Speed Up

並列計算の効率、数値解析効率、並列効率、負荷バランス効率に依存する⁶⁾。本研究では、同じ性能の計算ノードを用い、また領域は等分割ができたため、負荷バランス効率が1となる。更にベンチマークテストに用いた汎用 CFD コードの数値解析効率がほぼ1であるため、n 個の CPU を用いた並列計算の加速率 S_n は並列効率のみに依存し、次式のよ

うになる。

$$S_n = \frac{T_s}{T_n} = \frac{T_s}{T_{com} + T_s/n} = \frac{1}{T_{com}/T_s + 1/n} \quad (1)$$

ここで、 T_s は 1CPU の演算時間、 T_n は n 個の CPU での演算時間、 T_{com} は通信時間を表す。通信時間と 1CPU の演算時間との比は小規模な PC クラスタ (N 個の CPU) の加速率の実測値から次式のように求められる。

$$\frac{T_{com}}{T_s} = \frac{N - S_N}{N \cdot S_N} = C_s \quad (2)$$

更に通信時間が領域の分割数によらず、ほぼ一定となることを考え、式(2)を式(1)に代入することにより、任意 CPU を有する PC クラスタの加速率を求める式が得られる。

$$S_n = \frac{1}{C_s + 1/n} \quad (3)$$

Figure 2 には式(3)により予測された加速率と実測値との比較を示す。ここで、 C_s の値は 4 計算ノードを有する小規模な PC クラスタの実測値から求めた値である。この図から、式(3)により予測された加速率は実測値とよく一致することが分かる。

4. まとめ

本研究では PC クラスタを構築し、流体解析コードを用いて性能評価を行い、以下の結論を得た。

- 1) PC クラスタの加速率はネットワーク性能に大きく依存し、格子数の少ないケースではその影響が顕著である。
- 2) コストパフォーマンスの面では Gigabit Ethernet を採用したシステムが優れている。
- 3) PC クラスタの加速率は通信時間と演算時間との比から簡易式により予測できる。

参考文献

- 1) 須江: 第 13 回数値流体力学講演論文集, D04-3, 1999.
- 2) 足立, 福田, 太呉: 第 13 回数値流体力学講演論文集, D04-2, 1999.
- 3) 黒川 他: 第 14 回数値流体力学講演論文集, E08-2, 2000.
- 4) 檜山, 玉井, 土木学会論文集: No.668, pp.43-53, 2001.
- 5) FLUENT6.0: <http://www.fluent.com/>
- 6) Schreck, E., Peric, M.: Int. J. Numer. Methods Fluids, Vol.16, pp.303-327, 1993.