HPC2500 でのグリッド環境の構築

名古屋大学情報連携基盤センター
永井　亨

１．はじめに

2005 年 3 月、名古屋大学情報連携基盤センターにスーパーコンピュータシステム HPC2500（動作周波数 2.08Hz、24 ノード、トータルの理論最大性能 13.5Tflops、主記憶 12TB）が導入された。以下で、HPC2500 上に UNICORE ミドルウェアおよび NAREGI ミドルウェアによりそれぞれグリッド環境を構築し、性能測定を行った結果を報告する。

２．UNICORE 環境

2.1　システム構成

24 ノードのうちの 6 ノードをもちいて UNICORE 環境を構築した。システム構成を図 1 に、また、ソフトウェアのバージョンを表 1 にそれぞれ示す。hpc は 128CPU 構成（ただし、UNICORE から実際に使えるのは 64CPU）、残りの 5 ノードは 64CPU 構成で、ノードあたり 512GB の主記憶を搭載している。各ノードはシステムネットワークに GEther で接続されている。また、hpc は学内 LAN にも GEther で接続されている。

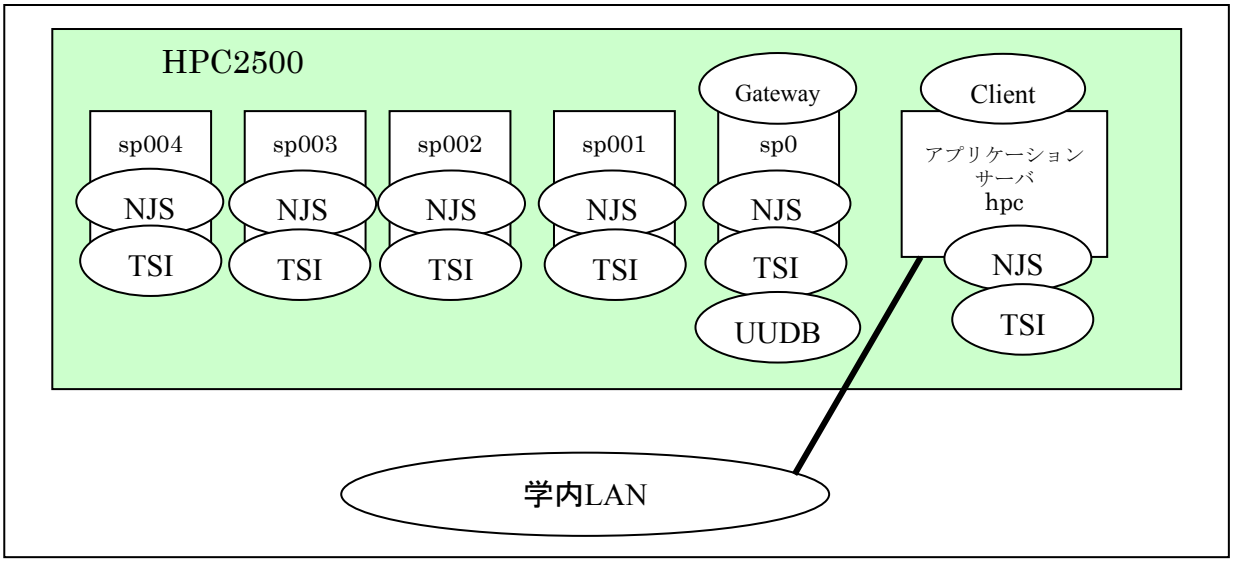

図 1

表 1

OS	Solaris9
UNICORE 環境	Gateway 4.1.0 NJS 4.1.0 TSI 4.1.0 UUDB 1.0.0
Java 環境	J2SE 1.4.2_04

2.2 性能測定

性能測定には Java で記述された並列計算プログラムをもちいた。Java をもちいたのはプラットフォームに依存しない並列プログラムが記述できるからである。具体的には、浅水方程式を採用し、領域分割法によって解く。テスト問題として収縮している開水路内の流れを考える。支配方程式は非粘性浅水方程式

$$\frac{\partial Q}{\partial t}+\frac{\partial E}{\partial x}+\frac{\partial F}{\partial y}=0$$

$$Q=\begin{bmatrix} h \\ hu \\ hv \end{bmatrix}, E=\begin{bmatrix} hu \\ hu^2+\dfrac{gh^2}{2} \\ huv \end{bmatrix}, F=\begin{bmatrix} hv \\ huv \\ hv^2+\dfrac{gh^2}{2} \end{bmatrix} \tag{1}$$

で与えられる。Q は保存量ベクトル、E、F はそれぞれ x、y 方向の流束ベクトルである。また、h は水面の高さ、u、v はそれぞれ x、y 方向の水深で平均された速度、g は重力加速度を表す。
である。差分解法には 2 次元の MacCormack 法

$$\begin{aligned} Q_{i,j}^{*} &= Q_{i,j}^{n}-\frac{\Delta t}{\Delta x}(E_{i+1,j}^{n}-E_{i,j}^{n})-\frac{\Delta t}{\Delta y}(F_{i,j+1}^{n}-F_{i,j}^{n}) \\ Q_{i,j}^{n+1} &= \frac{1}{2}[Q_{i,j}^{n}+Q_{i,j}^{*}-\frac{\Delta t}{\Delta x}(E_{i,j}^{*}-E_{i-1,j}^{*})-\frac{\Delta t}{\Delta y}(F_{i,j}^{*}-F_{i,j-1}^{*})] \end{aligned} \tag{2}$$

を用いた。このスキームは空間・時間ともに 2 次精度である。全領域における計算格子は図 2 に示すような非直交格子である．図 2 は水路を上から見た図で、赤い線の部分に壁があり、左から右に向かって流れる。水路に収縮した壁がある場合 oblique shock wave（斜め衝撃波）と呼ばれる物理量の不連続が発生することが知られている。計算条件等の詳細はここでは割愛する。比較のため、PC クラスタ上に構築した UNICORE をもちいて 4CPU 測定時の処理時間比を表 3 に示す。
使用した PC クラスタの仕様を表 2 に掲げる。PC クラスタは GEther で接続されている。

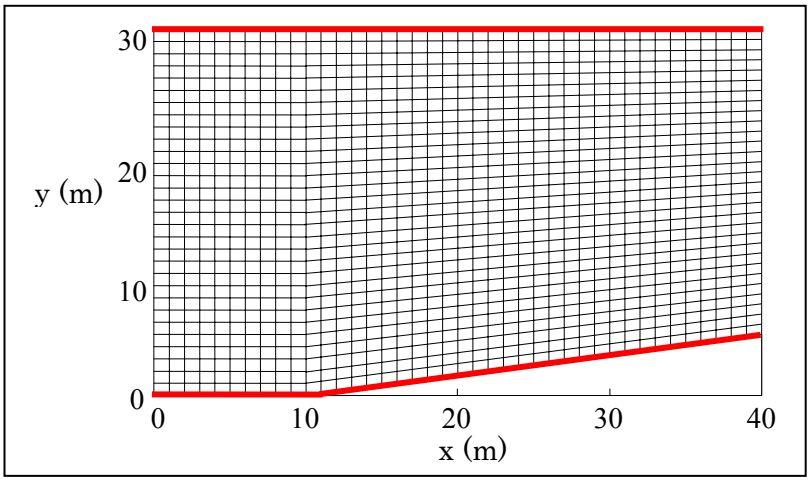

図 2

表 2

機種	DELL PowerEdge 800
CPU	Pentium4 3.6GHz (64bit)
OS	RedHat Enterprise Linux WS 4 (EM64T)
主記憶	2GB
UNICORE 環境	Gateway 4.1.0 NJS 4.2.0 TSI 4.1.0 UUDB 1.0.0
Java 環境	J2SE 1.4.2_08

表 3

― 4CPU での処理時間比 ―

HPC2500 : PC クラスタ
2 : 1

【参考】表 3 の補足

同一プログラムの非 UNICORE 環境での測定結果

富士通株式会社 2006/3/28 現在

<JINI プログラム>

VM / OS	1.4.2_04		1.4.2_08	
	ServerVM	ClientVM	ServerVM	ClientVM
RHEL 4.0 for EM64T	12375	13972	12307	13677
RHEL 4.0 for x86	12504	13601	12371	13787
RHEL 3.0 for x86	10604	11672	10272	11634
Solaris 9	56047	60582	55593	60017

(単位:ms)

【マシン環境～RHEL 系】

CPU:Xeon 3.6GHz×2, MEMORY:6GB)

RHEL-AS4 update 2(EM64T)

RHEL-AS4 update 2(x86)

RHEL-AS3 update 6(x86)

【マシン環境～Ｓｏｌａｒｉｓ】

CPU：SPARC64 IV 603MHz×4、MEMORY：4.5GB

Solaris 9

VM / OS	1.4.2_04		1.4.2_08	
	ServerVM	ClientVM	ServerVM	ClientVM
RHEL 4.0 for EM64T	0.22	0.23	0.22	0.23
RHEL 4.0 for x86	0.22	0.22	0.22	0.23
RHEL 3.0 for x86	0.19	0.19	0.18	0.19
Solaris 9	1.00	1.00	1.00	1.00

CPU クロック比だけに着目すると、この結果は SPARC の遅さは妥当である感じを受けるが、実行プロセス数と CPU 数を考慮に入れると、Java 単体アプリでの差が本アプリでの現状の Java 浮動小数点演算の差と考えるべきだ。

Java 単体アプリでは、起動プロセスは 1 個。

JINI プログラムは、8 個のプロセス(javaSpaces:6 個、Worker:1 個、Master:1 個)が起動する。

TOP コマンドで観察したところ、常に CPU を使用しているのは、JavaSpaces の 2 個、Worker の 1 個、Master の 1 個 の計 4 個。

<JINI-Linux solaris>

【Solaris 9】

ステップ数	ClientVM14204			ServerVM14204			ClientVM14208			ServerVM14208		
	1回目	2回目	3回目	1回目	2回目	3回目	1回目	2回目	3回目	1回目	2回目	3回目
10	61419	61508	64449	67643	66886	65712	61245	60237	60951	64449	65414	64122
20	61251	60993	61107	56090	56404	56348	60618	60727	60066	56575	56324	55690
30	61352	60939	61351	55869	56075	55616	60860	59666	60399	56036	55866	55579
40	59933	60204	60389	54611	54561	54383	59831	59296	59640	54080	54593	53820
50	60245	60310	60072	54863	54210	54851	60210	60494	59671	54337	54682	53964
60	60887	60485	60730	55182	54826	54843	60119	59312	60006	54464	54610	54126
70	59518	59430	59585	54528	53448	53922	59108	60533	59593	53476	53591	53311
80	59863	59957	60163	54953	54238	54221	60335	59208	58760	54759	54256	53263
90	60545	60688	60681	53268	55168	55655	60369	59826	60185	54790	54892	54151
100	59509	59935	59957	54060	54339	54636	59701	59484	60054	54191	54376	54005
10個の平均値	60452	60445	60848	56107	56016	56019	60240	59878	59933	55716	55860	55203
3回分の平均値			60582			56047			60017			55593

【RHEL4.0 for x86】

ステップ数	ClientVM14204			ServerVM14204			ClientVM14208			ServerVM14208		
	1回目	2回目	3回目	1回目	2回目	3回目	1回目	2回目	3回目	1回目	2回目	3回目
10	0 14382	13572	14308	14620	18459	18200	14705	14311	14401	18231	17316	17681
20	14254	13211	13597	12059	13243	12760	13907	13652	14068	12206	12590	11205
30	14231	14161	13718	12944	12236	12618	14278	13564	14274	12525	11277	12911
40	13534	13394	13180	11141	12216	12092	13711	13652	13532	11758	11894	11245
50	13724	13651	13461	11881	11806	11998	13734	13288	14061	11932	11214	12077
60	13648	13239	14266	11494	12039	12111	13795	14365	13955	12127	12025	11468
70	14042	12596	12991	11604	11400	12148	13695	13316	13413	12228	11456	12002
80	13653	13819	14066	11438	12126	11615	14178	13558	14155	12186	11975	11581
90	13472	13197	13264	12350	11339	12277	13369	13333	13182	11835	11129	11426
100	13468	13449	13074	11435	11671	11805	13648	13085	13422	10871	11815	10935
10個の平均値	13781	13429	13593	12097	12654	12762	13902	13612	13846	12590	12269	12253
3回分の平均値			13601			12504			13787			12371

（単位:ms）

【RHEL4.0 for EM64T】

ステップ数	ClientVM14204			ServerVM14204			ClientVM14208			ServerVM14208		
	1回目	2回目	3回目	1回目	2回目	3回目	1回目	2回目	3回目	1回目	2回目	3回目
10	14437	15018	15003	16572	15437	13487	13600	14415	14552	15430	15414	18032
20	13805	15462	14226	12496	12221	12509	14028	13392	14207	12677	11984	12577
30	13924	13895	13469	11949	11774	12378	14099	12771	13757	12119	12404	12066
40	13007	13463	14625	11714	12182	11542	13760	13842	13785	11460	11124	11784
50	13629	15425	14085	11745	11415	12025	14111	12999	13673	12028	11570	11775
60	13758	13683	13705	11952	11611	13437	13641	13607	13822	11872	11712	12008
70	13269	13854	14430	11607	11844	12364	13185	13783	13612	11967	11655	12006
80	13644	14831	13526	11918	11777	13067	14035	13205	13836	12130	11949	11943
90	13388	13699	13207	11482	12081	13452	13580	12782	13097	12041	11399	11517
100	13017	13227	14453	11643	11046	12517	13643	13839	13641	11779	11297	11493
10個の平均値	13588	14256	14073	12308	12139	12678	13768	13464	13798	12350	12061	12520
3回分の平均値			13972			12375			13677			12307

（単位:ms）

【RHEL3.0 for x86】

ステップ数	ClientVM14204			ServerVM14204			ClientVM14208			ServerVM14208		
	1回目	2回目	3回目	1回目	2回目	3回目	1回目	2回目	3回目	1回目	2回目	3回目
10	12339	12488	12421	15817	13774	16561	12361	12242	11350	15253	13371	14548
20	11610	11851	11669	10421	11328	10587	11486	12155	11756	10413	10716	10692
30	11502	11896	11799	11302	11586	11572	11774	12137	11325	9985	9954	9971
40	11681	11600	10987	9269	9093	9152	11291	11754	11257	10270	9659	9474
50	11243	11163	11541	10093	9731	9802	11303	12204	11903	9515	9165	9624
60	11585	11647	12066	10249	10143	10065	11255	11669	11319	9846	9771	9928
70	12063	11597	11718	9789	9920	9864	11434	11416	10927	9543	9440	9563
80	11503	11777	11642	10038	9931	10006	11846	11949	11752	9928	9686	9677
90	11245	11511	11567	9886	9692	9723	11510	11675	11304	9613	9660	9925
100	11710	11640	11091	9432	9818	9475	11279	11717	11759	9830	9646	9491
10個の平均値	11648	11717	11650	10630	10502	10681	11554	11892	11465	10420	10107	10289
3回分の平均値			11672			10604			11637			10272

（単位:ms）

- プログラムの条件

DIVISION=15

VERTICAL_GRID=32

VIZTIMING=10

-Java オプション

- JavaSpaces

-noverify -Xss256k -Xms512m -Xmx512m -Xbatch

- Master, Worker

-Xss256k -Xms512m -Xmx512m -Xbatch

<java 単体アプリ>

VM / OS	1.4.2_04		1.4.2_08	
	ServerVM	ClientVM	ServerVM	ClientVM
RHEL 4.0 for EM64T	6127	5767	6120	5756
RHEL 4.0 for x86	6609	6423	6608	6528
RHEL 3.0 for x86	6453	6571	6410	6478
Solaris 9	56004	63435	55951	63386

(単位:ms)

【マシン環境～RHEL 系】

CPU:Xeon 3.6GHz×2, MEMORY:6GB)

RHEL-AS4 update 2(EM64T)

RHEL-AS4 update 2(x86)

RHEL-AS3 update 6(x86)

【マシン環境～Ｓｏｌａｒｉｓ】

CPU：SPARC64 IV 603MHz×4、MEMORY：4.5GB

Solaris 9

<単体-Solaris-Linux>

			1 回目	2 回目	3 回目	平均
RHEL-AS4(EM64T)	J2SE1.4.2_04	ClientVM	5775	5726	5800	5767
		ServerVM	6150	6115	6116	6127
	J2SE1.4.2_08	ClientVM	5773	5720	5775	5756
		ServerVM	6140	6129	6091	6120
RHEL-AS4(x86)	J2SE1.4.2_04	ClientVM	6388	6450	6431	6423
		ServerVM	6639	6673	6694	6669
	J2SE1.4.2_08	ClientVM	6495	6537	6553	6528
		ServerVM	6496	6660	6668	6608
RHEL-AS3(x86)	J2SE1.4.2_04	ClientVM	6609	6549	6554	6571
		ServerVM	6480	6469	6409	6453
	J2SE1.4.2_08	ClientVM	6498	6463	6472	6478
		ServerVM	6408	6422	6401	6410
Solaris 9	J2SE1.4.2_04	ClientVM	63097	63449	63758	63435
		ServerVM	56191	55827	55994	56004
	J2SE1.4.2_08	ClientVM	63589	63630	62939	63386
		ServerVM	55825	55724	56304	55951

(単位:ms)

【マシン環境～RHEL 系】
CPU:Xeon 3.6GHz×2, MEMORY:6GB)
RHEL-AS4 update 2(EM64T)
RHEL-AS4 update 2(x86)
RHEL-AS3 update 6(x86)

【マシン環境～Ｓｏｌａｒｉｓ】
CPU：SPARC64 IV 603MHz×4、MEMORY：4.5GB
Solaris 9

３．NAREGI ミドルウェアα版のインストール
これまでの進捗状況を以下に示す。

2005 年 12 月末　グリッド研究開発推進拠点（国立情報学研究所）にてサーバ系構築作業（2 泊 3 日で完了せず）

2006 年 1 月下旬　外部へのネットワーク接続を遮断し、NAREGI 拠点での環境を再現して動作確認終了

2 月初旬　NAREGI-CA による証明書の作成

2 月中旬　本センターの環境でサンプルプログラムをもちいて動作確認終了（ただし、実行ファイルのアクセス権を 700 から 705 に変更する必要があった）

2 月下旬　計算ノード（PC）を追加し、構築作業に入る。

α版が Solaris9 で動作しないことが判明した。このため、現在は計算ノードとして PC を使用している。現状ではサービスマシンである HPC2500 上に構築するのは困難であると判断したため、今後、スパコン予算とは別に（グリッド用補正予算）調達した HPC2500（1.3GHz、32CPU 構成が 2 ノード）上にβ版をインストールし計算ノードとして使用する予定。Solaris9 で動作するかは不明。

以上