

●●● 2004 02/20

タンパク質立体構造解析システムFAMS のGRID化事例

日立ソフトウェアエンジニアリング株式会社
ライフサイエンス本部
バイオインフォマティクス開発部
岡山 利次

(C) Hitachi Software Engineering Co., Ltd. 2004 All rights reserved.
本資料中に掲載されている会社名、製品名は、すべて関係各社の商標または登録商標、商品名です。

はじめに

ゲノム ↔ トランスクリプトーム ↔ プロテオーム



タンパク質の立体構造

タンパク質 - リガンド
結合・相互作用 → 基礎から応用まで
生命の根幹

構造決定

実験的手法による決定

→ 膨大なコストと時間

計算機による推定

→ 計算能力と推定速度の
トレードオフ

グリッドコンピュータによる新たなブレークスルー

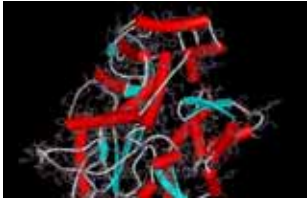
DIGITAL & GLOBAL
日立ソフト

タンパク質立体構造予測プログラムFAMS

DIGITAL & GLOBAL

北里大学と日本SGI
によるFAMSを用いたソリューション

SARSのタンパク質の立体構造を予測
http://www.sgi.co.jp/newsroom/press_releases/2003/may/sars.html



↓

医学・薬学へ情報を提供
応用分野への貢献

(C) Hitachi Software Engineering Co., Ltd. 2004 All rights reserved.

DIGITAL & GLOBAL
日立ソフト

FAMSとは

DIGITAL & GLOBAL

Full Automatic Modeling System

北里大学薬学部梅山研究室チームが2000年に発表した¹、
 計算機による**全自動タンパク質立体構造解析**システム。
 その後日本SGIと共同でシステム面での改良が施された。

アミノ酸配列のみが分かっているタンパク質の立体構造を予測する
 ために、それとアミノ酸配列が類似していて、かつ、立体構造が既知
 のタンパク質を利用する。(ホモロジーモデリング)

タンパク質立体構造予測システムの国際的なコンテストである
CAFASPで2000年以来首位を維持。

立体構造の
未知な
アミノ酸配列から

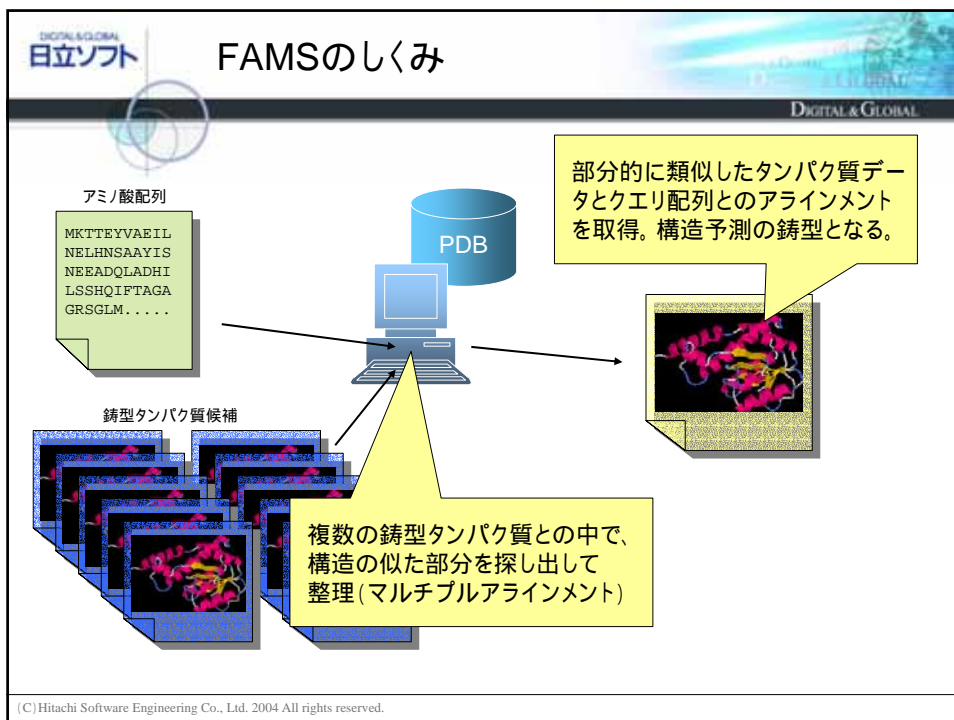
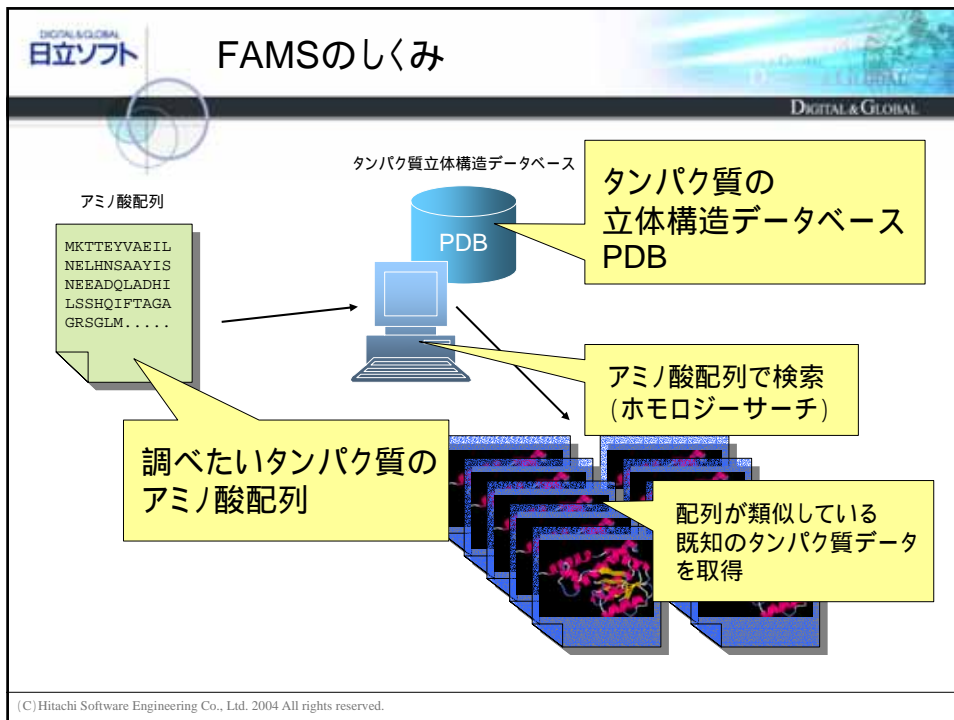
→

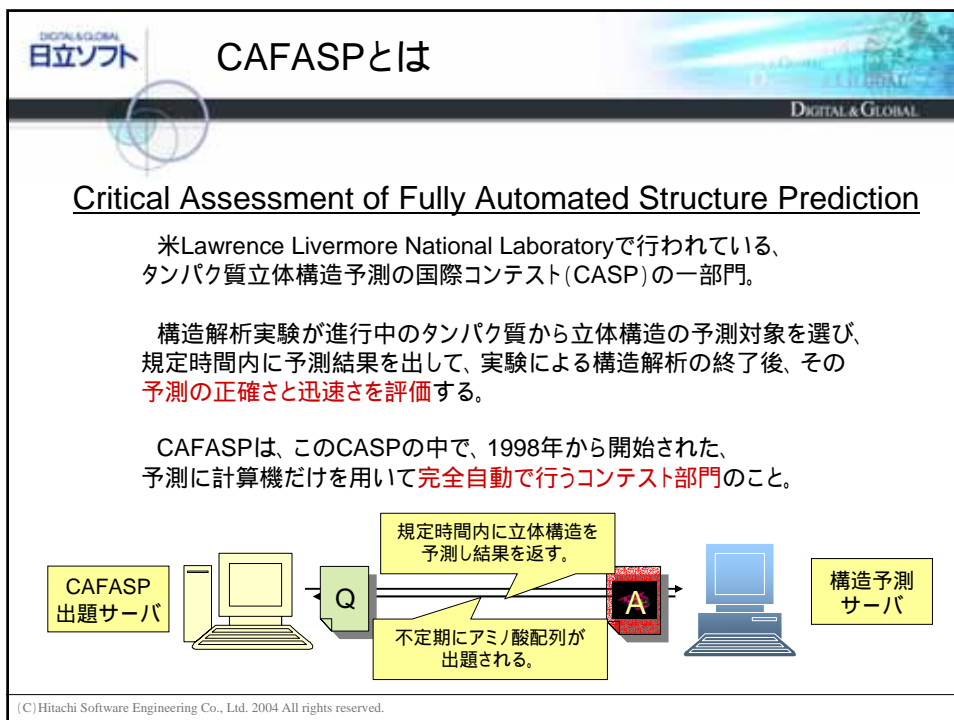
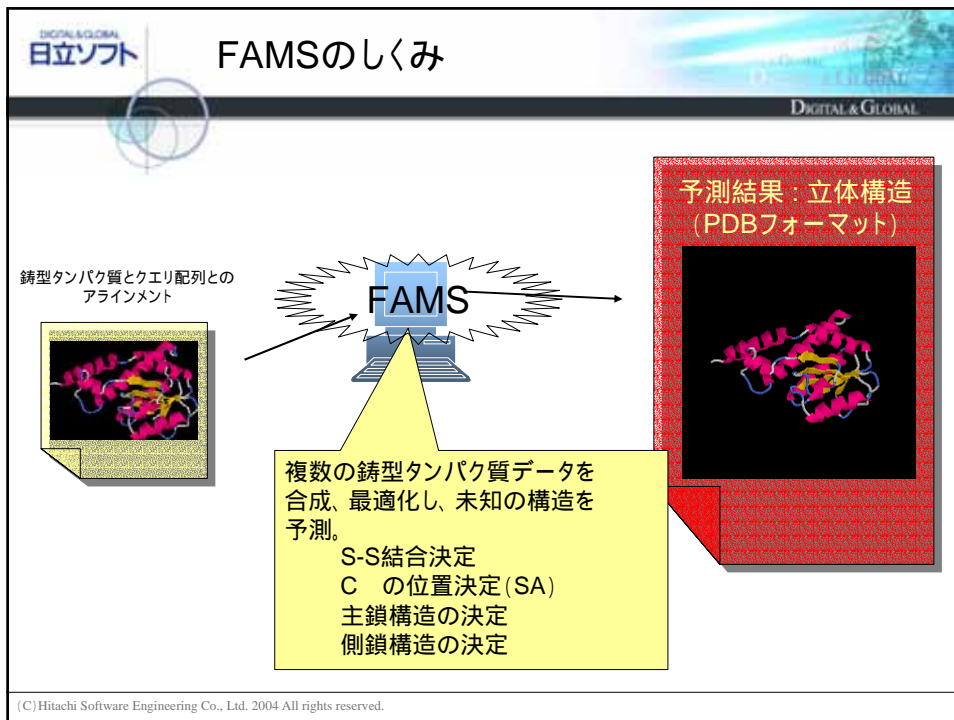
立体構造を予測

FAMS

¹ Ogata, K. and Umeyama, H. (2000) An automatic homology modeling
 method consisting of database searches and simulated annealing. J. Mol. Graphics Mod. 18:258-272.

(C) Hitachi Software Engineering Co., Ltd. 2004 All rights reserved.





DIGITAL & GLOBAL
日立ソフト

CAFASPでの成果

DIGITAL & GLOBAL

CAFASP2 (2000) ¹

ホモロジーモデリング(HM)セクションに参加。
合計49題出題。
1題につき48時間以内に結果を返す。

Dunbrack博士による総合評価

1. FAMS
2. 3D-JIGSAW (英国)
(上位4チーム中)

CAFASP3 (2002) ²

比較モデリング(CM)セクションに参加。
合計77題出題。
1題につき48時間以内に結果を返す。

Dunbrack博士による総合評価

1. FAMS
2. FAMSD (FAMS改良)
3. ALAX (名古屋大のFAMS)
(上位9チーム中)

¹ <http://www.cs.bgu.ac.il/~dfischer/CAFASP2/>
² <http://www.cs.bgu.ac.il/~dfischer/CAFASP3/>

(C) Hitachi Software Engineering Co., Ltd. 2004 All rights reserved.

DIGITAL & GLOBAL
日立ソフト

グリッド化におけるモデリング

DIGITAL & GLOBAL

```

graph TD
    A[FAMSのグリッド化] --> B((1ジョブ当り平均約400部分計算))
    B --> C[データ分割]
    B --> D[プロセス分割]
    C --> E[データベース分割]
    C --> F[問い合わせ分割]
    D --> G[部分計算分割]
    D --> H[試行分割]
    E --- I[BLAST (GT2で実装検証済み)]
    F --- I
    G --- J[依存性が強いとGRIDには不向き]
    H --- K[トランスクリプトーム解析  
・遺伝子組の選択  
・クロスバリデーション  
・ブートストラッピング]
  
```

(C) Hitachi Software Engineering Co., Ltd. 2004 All rights reserved.

日立ソフト DIGITAL & GLOBAL

従来のFAMSシステム

ジョブを投入

PCクラスタ

host 1

host 2

...

host n

NFS

FAMSを実行

環境を拡充、改変するには
工数とコストが不可欠。

←→ データの流れ
→ プログラムの実行

計算機資源のひっ迫

PCクラスタでのシステム

- マシンは全てLinuxで、専用で共通の環境構築が必要。
- データ、プログラムの共有方法が別々に必要。

(C) Hitachi Software Engineering Co., Ltd. 2004 All rights reserved.

日立ソフト DIGITAL & GLOBAL

FAMSシステムのグリッド化

ジョブを投入

グリッド環境

host a

host b

host c

host d

host x

host y

FAMSを実行

←→ データの流れ
→ プログラムの実行

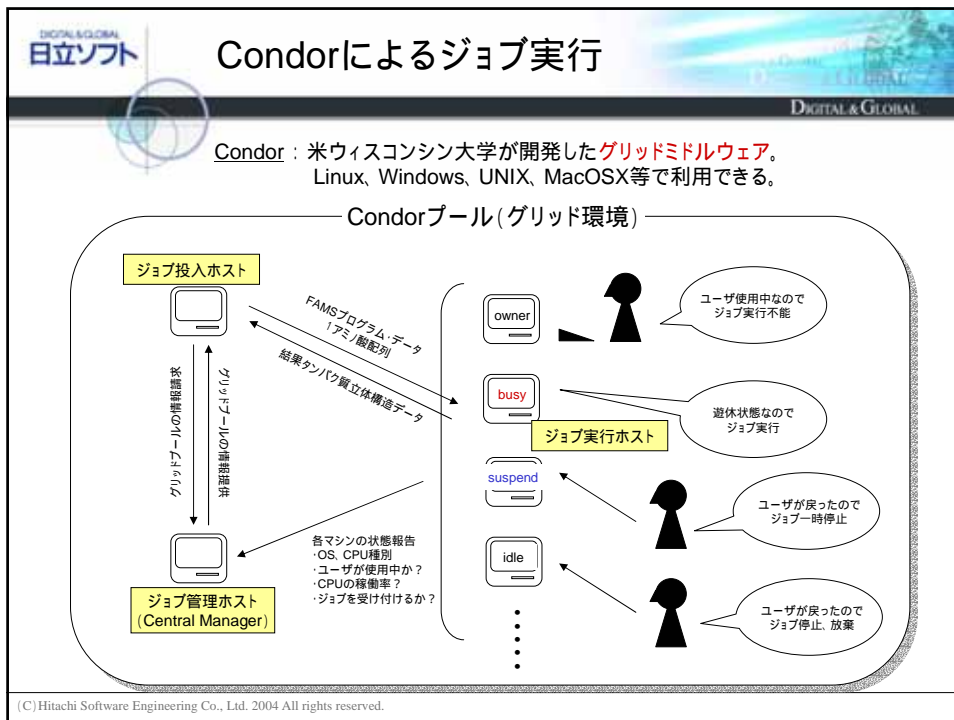
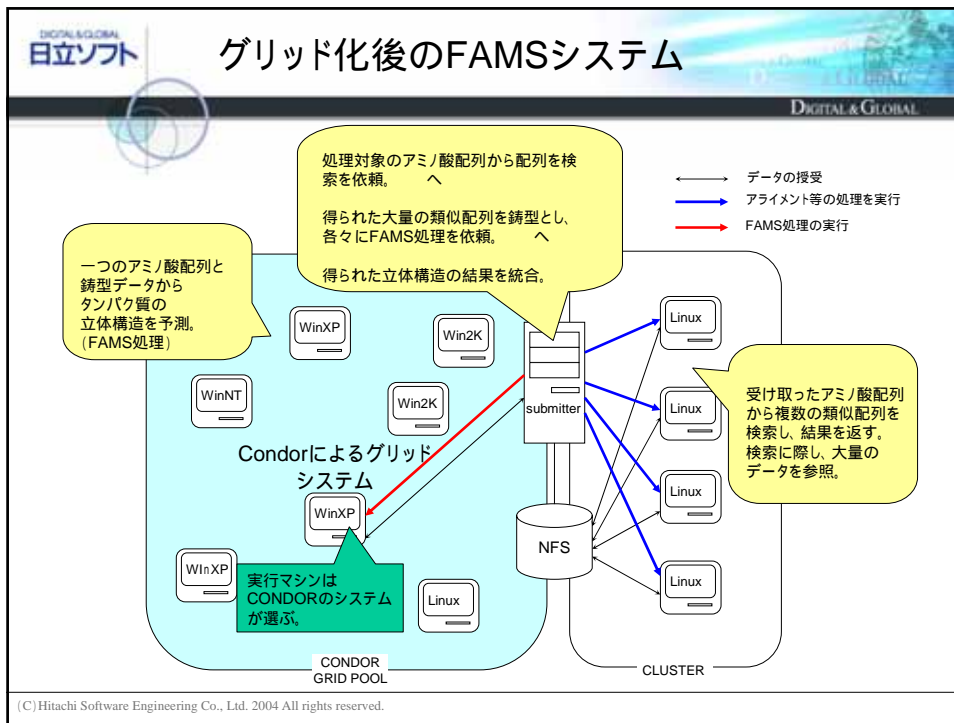
Condorミドルウェア¹でのグリッド構築

- 複数の環境 (WindowsやLinux) で実行可能。オフィスデスクトップでも計算機資源になる。
- データ、プログラムの共有方法は不要。新たに追加するホストは最小限の設定でOK。

低コストで動的な実行環境を実現。

¹ <http://www.cs.wisc.edu/condor/>

(C) Hitachi Software Engineering Co., Ltd. 2004 All rights reserved.



(1) FAMSをWindows、Linux双方で実行可能なように移植、改変。



Linux専用にかかれたFAMSのプログラムを、Windows用に改変、実行ファイル作成。Cygwin環境下で実行可能に。

(2) FAMSをCondorジョブとして実行可能なように改変。

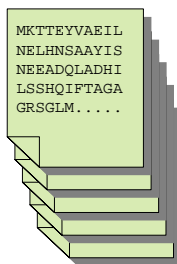


Linux、Windows用のFAMSプログラムを、グリッドのジョブとして実行できるように、Condorの仕様にあわせて改変。

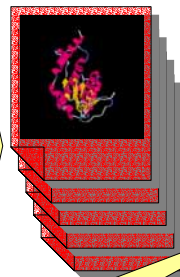
(3) 日立ソフト内業務用PC100台規模を用いてグリッド環境を作成。



社内で様々な業務・開発に用いているLinux、Windows2000、WindowsXPの各マシンにCondorミドルウェアとCygwinを導入。



複数ジョブを投入



稼働状況
2003年11月21日14:15:22 現在

| Host | Job | SubmissionTime | State | ExitCode | ExitCode | Max. Mem. Usage | WallTime |
|---------|--------|---------------------|---------|----------|----------|-----------------|----------|
| node01 | job1 | 2003-11-21 14:15:22 | Running | 0 | 0 | 1024 | 00:01:00 |
| node02 | job2 | 2003-11-21 14:15:22 | Running | 0 | 0 | 2048 | 00:02:00 |
| node03 | job3 | 2003-11-21 14:15:22 | Running | 0 | 0 | 3072 | 00:03:00 |
| node04 | job4 | 2003-11-21 14:15:22 | Running | 0 | 0 | 4096 | 00:04:00 |
| node05 | job5 | 2003-11-21 14:15:22 | Running | 0 | 0 | 5120 | 00:05:00 |
| node06 | job6 | 2003-11-21 14:15:22 | Running | 0 | 0 | 6144 | 00:06:00 |
| node07 | job7 | 2003-11-21 14:15:22 | Running | 0 | 0 | 7168 | 00:07:00 |
| node08 | job8 | 2003-11-21 14:15:22 | Running | 0 | 0 | 8192 | 00:08:00 |
| node09 | job9 | 2003-11-21 14:15:22 | Running | 0 | 0 | 9216 | 00:09:00 |
| node10 | job10 | 2003-11-21 14:15:22 | Running | 0 | 0 | 10240 | 00:10:00 |
| node11 | job11 | 2003-11-21 14:15:22 | Running | 0 | 0 | 11264 | 00:11:00 |
| node12 | job12 | 2003-11-21 14:15:22 | Running | 0 | 0 | 12288 | 00:12:00 |
| node13 | job13 | 2003-11-21 14:15:22 | Running | 0 | 0 | 13312 | 00:13:00 |
| node14 | job14 | 2003-11-21 14:15:22 | Running | 0 | 0 | 14336 | 00:14:00 |
| node15 | job15 | 2003-11-21 14:15:22 | Running | 0 | 0 | 15360 | 00:15:00 |
| node16 | job16 | 2003-11-21 14:15:22 | Running | 0 | 0 | 16384 | 00:16:00 |
| node17 | job17 | 2003-11-21 14:15:22 | Running | 0 | 0 | 17408 | 00:17:00 |
| node18 | job18 | 2003-11-21 14:15:22 | Running | 0 | 0 | 18432 | 00:18:00 |
| node19 | job19 | 2003-11-21 14:15:22 | Running | 0 | 0 | 19456 | 00:19:00 |
| node20 | job20 | 2003-11-21 14:15:22 | Running | 0 | 0 | 20480 | 00:20:00 |
| node21 | job21 | 2003-11-21 14:15:22 | Running | 0 | 0 | 21504 | 00:21:00 |
| node22 | job22 | 2003-11-21 14:15:22 | Running | 0 | 0 | 22528 | 00:22:00 |
| node23 | job23 | 2003-11-21 14:15:22 | Running | 0 | 0 | 23552 | 00:23:00 |
| node24 | job24 | 2003-11-21 14:15:22 | Running | 0 | 0 | 24576 | 00:24:00 |
| node25 | job25 | 2003-11-21 14:15:22 | Running | 0 | 0 | 25600 | 00:25:00 |
| node26 | job26 | 2003-11-21 14:15:22 | Running | 0 | 0 | 26624 | 00:26:00 |
| node27 | job27 | 2003-11-21 14:15:22 | Running | 0 | 0 | 27648 | 00:27:00 |
| node28 | job28 | 2003-11-21 14:15:22 | Running | 0 | 0 | 28672 | 00:28:00 |
| node29 | job29 | 2003-11-21 14:15:22 | Running | 0 | 0 | 29696 | 00:29:00 |
| node30 | job30 | 2003-11-21 14:15:22 | Running | 0 | 0 | 30720 | 00:30:00 |
| node31 | job31 | 2003-11-21 14:15:22 | Running | 0 | 0 | 31744 | 00:31:00 |
| node32 | job32 | 2003-11-21 14:15:22 | Running | 0 | 0 | 32768 | 00:32:00 |
| node33 | job33 | 2003-11-21 14:15:22 | Running | 0 | 0 | 33792 | 00:33:00 |
| node34 | job34 | 2003-11-21 14:15:22 | Running | 0 | 0 | 34816 | 00:34:00 |
| node35 | job35 | 2003-11-21 14:15:22 | Running | 0 | 0 | 35840 | 00:35:00 |
| node36 | job36 | 2003-11-21 14:15:22 | Running | 0 | 0 | 36864 | 00:36:00 |
| node37 | job37 | 2003-11-21 14:15:22 | Running | 0 | 0 | 37888 | 00:37:00 |
| node38 | job38 | 2003-11-21 14:15:22 | Running | 0 | 0 | 38912 | 00:38:00 |
| node39 | job39 | 2003-11-21 14:15:22 | Running | 0 | 0 | 39936 | 00:39:00 |
| node40 | job40 | 2003-11-21 14:15:22 | Running | 0 | 0 | 40960 | 00:40:00 |
| node41 | job41 | 2003-11-21 14:15:22 | Running | 0 | 0 | 41984 | 00:41:00 |
| node42 | job42 | 2003-11-21 14:15:22 | Running | 0 | 0 | 43008 | 00:42:00 |
| node43 | job43 | 2003-11-21 14:15:22 | Running | 0 | 0 | 44032 | 00:43:00 |
| node44 | job44 | 2003-11-21 14:15:22 | Running | 0 | 0 | 45056 | 00:44:00 |
| node45 | job45 | 2003-11-21 14:15:22 | Running | 0 | 0 | 46080 | 00:45:00 |
| node46 | job46 | 2003-11-21 14:15:22 | Running | 0 | 0 | 47104 | 00:46:00 |
| node47 | job47 | 2003-11-21 14:15:22 | Running | 0 | 0 | 48128 | 00:47:00 |
| node48 | job48 | 2003-11-21 14:15:22 | Running | 0 | 0 | 49152 | 00:48:00 |
| node49 | job49 | 2003-11-21 14:15:22 | Running | 0 | 0 | 50176 | 00:49:00 |
| node50 | job50 | 2003-11-21 14:15:22 | Running | 0 | 0 | 51200 | 00:50:00 |
| node51 | job51 | 2003-11-21 14:15:22 | Running | 0 | 0 | 52224 | 00:51:00 |
| node52 | job52 | 2003-11-21 14:15:22 | Running | 0 | 0 | 53248 | 00:52:00 |
| node53 | job53 | 2003-11-21 14:15:22 | Running | 0 | 0 | 54272 | 00:53:00 |
| node54 | job54 | 2003-11-21 14:15:22 | Running | 0 | 0 | 55296 | 00:54:00 |
| node55 | job55 | 2003-11-21 14:15:22 | Running | 0 | 0 | 56320 | 00:55:00 |
| node56 | job56 | 2003-11-21 14:15:22 | Running | 0 | 0 | 57344 | 00:56:00 |
| node57 | job57 | 2003-11-21 14:15:22 | Running | 0 | 0 | 58368 | 00:57:00 |
| node58 | job58 | 2003-11-21 14:15:22 | Running | 0 | 0 | 59392 | 00:58:00 |
| node59 | job59 | 2003-11-21 14:15:22 | Running | 0 | 0 | 60416 | 00:59:00 |
| node60 | job60 | 2003-11-21 14:15:22 | Running | 0 | 0 | 61440 | 01:00:00 |
| node61 | job61 | 2003-11-21 14:15:22 | Running | 0 | 0 | 62464 | 01:01:00 |
| node62 | job62 | 2003-11-21 14:15:22 | Running | 0 | 0 | 63488 | 01:02:00 |
| node63 | job63 | 2003-11-21 14:15:22 | Running | 0 | 0 | 64512 | 01:03:00 |
| node64 | job64 | 2003-11-21 14:15:22 | Running | 0 | 0 | 65536 | 01:04:00 |
| node65 | job65 | 2003-11-21 14:15:22 | Running | 0 | 0 | 66560 | 01:05:00 |
| node66 | job66 | 2003-11-21 14:15:22 | Running | 0 | 0 | 67584 | 01:06:00 |
| node67 | job67 | 2003-11-21 14:15:22 | Running | 0 | 0 | 68608 | 01:07:00 |
| node68 | job68 | 2003-11-21 14:15:22 | Running | 0 | 0 | 69632 | 01:08:00 |
| node69 | job69 | 2003-11-21 14:15:22 | Running | 0 | 0 | 70656 | 01:09:00 |
| node70 | job70 | 2003-11-21 14:15:22 | Running | 0 | 0 | 71680 | 01:10:00 |
| node71 | job71 | 2003-11-21 14:15:22 | Running | 0 | 0 | 72704 | 01:11:00 |
| node72 | job72 | 2003-11-21 14:15:22 | Running | 0 | 0 | 73728 | 01:12:00 |
| node73 | job73 | 2003-11-21 14:15:22 | Running | 0 | 0 | 74752 | 01:13:00 |
| node74 | job74 | 2003-11-21 14:15:22 | Running | 0 | 0 | 75776 | 01:14:00 |
| node75 | job75 | 2003-11-21 14:15:22 | Running | 0 | 0 | 76800 | 01:15:00 |
| node76 | job76 | 2003-11-21 14:15:22 | Running | 0 | 0 | 77824 | 01:16:00 |
| node77 | job77 | 2003-11-21 14:15:22 | Running | 0 | 0 | 78848 | 01:17:00 |
| node78 | job78 | 2003-11-21 14:15:22 | Running | 0 | 0 | 79872 | 01:18:00 |
| node79 | job79 | 2003-11-21 14:15:22 | Running | 0 | 0 | 80896 | 01:19:00 |
| node80 | job80 | 2003-11-21 14:15:22 | Running | 0 | 0 | 81920 | 01:20:00 |
| node81 | job81 | 2003-11-21 14:15:22 | Running | 0 | 0 | 82944 | 01:21:00 |
| node82 | job82 | 2003-11-21 14:15:22 | Running | 0 | 0 | 83968 | 01:22:00 |
| node83 | job83 | 2003-11-21 14:15:22 | Running | 0 | 0 | 84992 | 01:23:00 |
| node84 | job84 | 2003-11-21 14:15:22 | Running | 0 | 0 | 86016 | 01:24:00 |
| node85 | job85 | 2003-11-21 14:15:22 | Running | 0 | 0 | 87040 | 01:25:00 |
| node86 | job86 | 2003-11-21 14:15:22 | Running | 0 | 0 | 88064 | 01:26:00 |
| node87 | job87 | 2003-11-21 14:15:22 | Running | 0 | 0 | 89088 | 01:27:00 |
| node88 | job88 | 2003-11-21 14:15:22 | Running | 0 | 0 | 90112 | 01:28:00 |
| node89 | job89 | 2003-11-21 14:15:22 | Running | 0 | 0 | 91136 | 01:29:00 |
| node90 | job90 | 2003-11-21 14:15:22 | Running | 0 | 0 | 92160 | 01:30:00 |
| node91 | job91 | 2003-11-21 14:15:22 | Running | 0 | 0 | 93184 | 01:31:00 |
| node92 | job92 | 2003-11-21 14:15:22 | Running | 0 | 0 | 94208 | 01:32:00 |
| node93 | job93 | 2003-11-21 14:15:22 | Running | 0 | 0 | 95232 | 01:33:00 |
| node94 | job94 | 2003-11-21 14:15:22 | Running | 0 | 0 | 96256 | 01:34:00 |
| node95 | job95 | 2003-11-21 14:15:22 | Running | 0 | 0 | 97280 | 01:35:00 |
| node96 | job96 | 2003-11-21 14:15:22 | Running | 0 | 0 | 98304 | 01:36:00 |
| node97 | job97 | 2003-11-21 14:15:22 | Running | 0 | 0 | 99328 | 01:37:00 |
| node98 | job98 | 2003-11-21 14:15:22 | Running | 0 | 0 | 100352 | 01:38:00 |
| node99 | job99 | 2003-11-21 14:15:22 | Running | 0 | 0 | 101376 | 01:39:00 |
| node100 | job100 | 2003-11-21 14:15:22 | Running | 0 | 0 | 102400 | 01:40:00 |

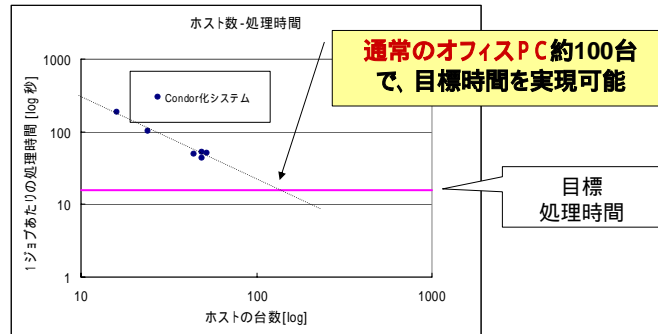
グリッド環境内の各ホストのマシンのプロフィール
負荷状態
ジョブ実行状況
をリアルタイムで表示。

パフォーマンス測定実験

グリッド環境内のLinuxマシンから、退勤時間ごろ、約420ジョブを投入し、全ジョブ処理時間を計測。1ジョブあたりの処理時間を算出。
グリッド環境への参加ホスト台数は、日々変動。

グリッド環境構成例

WindowsXP : 29台
Windows2000 : 12台
Redhat Linux7.3 : 11台



(C) Hitachi Software Engineering Co., Ltd. 2004 All rights reserved.

(1) グリッド環境を1000台規模への拡張



グリッドとしてのスケーラビリティ、コストパフォーマンスを実証。



自社開発セキュリティシステム (秘文®) との連携強化検証。

(2) グリッドビジネスへの展開



バイオインフォマティクス分野でのグリッドソリューションメニューの拡充。(計算グリッド/データグリッド)
当社ASPサービス (DNASIS GeneIndex®) との連携



デファクトスタンダードへの対応と
ユビキタス時代へ向けたグリッド開発ノウハウの活用。

(C) Hitachi Software Engineering Co., Ltd. 2004 All rights reserved.

謝辞

北里大学薬学部
生物分子設計学教室(梅山研究室)の皆様

御協力企業

日本SGI株式会社
株式会社インシリコサイエンス

- 以上 -