

日本ソフトウェア科学会参加報告

The participation report of Japan Society for Software Science and Technology

岩橋 崇史, 青井 桂子

Takashi IWAHASHI, Keiko AOI

Abstract: Grid is in the tendency for the expectation for the science problem with the data analysis demand of a huge quantity like physics, astronomy and bioinformatics to grow very much. This is the participation report of Japan Society for Software Science and Technology on November 25th, 2002.

1 はじめに

近年, グリッドへの期待が非常に高まっている. グリッドへの期待は, 米国, 欧州, およびアジアの国々で数多くのグリッドプロジェクトが立ち上がりつつあることからわかる. このように期待されるグリッドは, インターネット上に地理的に分散する異質な計算機資源の動的な統合を可能にするものであり, 物理学, 天文学, バイオインフォマティクスのように膨大な量のデータ解析要求を持つ科学問題に対しての応用への期待が非常に高まる傾向にある.

そこで我々はバイオインフォマティクスとグリッドの関連技術と最近の動向を知る目的で, 11月25日に日本ソフトウェア科学会主催「バイオグリッド: グリッドの動向とグリッドのバイオへの応用」に参加した. 本レポートはその参加報告である.

2 チュートリアル

日本ソフトウェア科学会において, グリッドの技術動向や研究動向に興味をもつ技術者や研究者, またグリッドを実際の科学問題へ応用しようとする研究者を対象に, チュートリアル「バイオグリッド: グリッドの動向とグリッドのバイオへの応用」が11月25日に大阪第2ビル6階にて10:30から17:00まで開催された. Fig. 1に, チュートリアルの風景を示す. チュートリアルの概要を次に示す.



Fig. 1 チュートリアルの風景

2.1 グリッドの基礎と環境構築技術 (松岡 聡 氏)

松岡氏は, グリッドの研究を盛んに行われており, ささまざまなグリッドプロジェクトを立ち上げている.

その松岡氏からはグリッドの基本的知識や現在発足されているグリッドプロジェクトと, バイオインフォマティクス分野におけるグリッド環境の構築の説明があった. バイオインフォマティクス分野におけるタンパク質の立体構造予測には膨大な計算量とその媒体となる計算資源が必要とされている. そのためには分散された計算資源を一つに統合し, 大規模計算機を構築するグリッド技術が必要不可欠である. グリッドより構築された大規模計算機上において, ゲノム情報などの実験データを格納する大規模データベース, 解析を行う計算資源が存在する. 現在は, 大規模データベースからリモートの計算資源へのペタバイト級のデータ転送は実現されておらず, できるだけローカルな計算資源で処理されてから転送するといった形になっている.

今後は高速ネットワークとそれによって可能になる大規模計算, 大規模データ処理, 大規模協調環境が必要とされている. Fig. 2にバイオグリッドの概念図を示す.

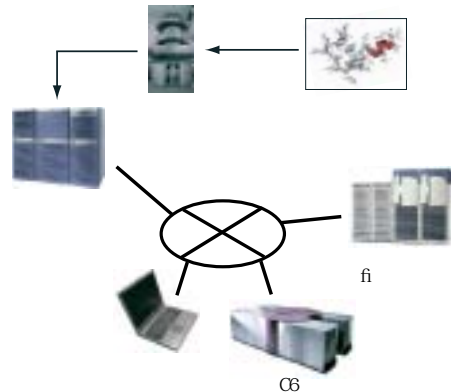


Fig. 2 バイオグリッドの概念図

2.2 グリッドのバイオ分野への応用(下條 真司 氏 伊達 進 氏)

このプログラムは、2部構成になっていた。まず第1部に、下條氏が関わっている BioGrid Project について、バイオグリッドセンターの構想、SC2002でのデモなどの説明があった。バイオグリッドセンターの構想については、超高速ネットワークにより、超高速計算機、大規模データベースを核とするリエゾンセンターを築くことが述べられた。SC2002で行ったデモについては、サンディエゴから大阪大学の超高压電子顕微鏡をリモートで操作し、得られたデータ画像をボルチモアまで伝送するといったものだった。データ転送には IPv6 での広帯域伝送を試みたが、不具合が生じ、実際には IPv4 でのデータ転送を行ったとされていた。Fig. 3 に、SC2002で行われた大阪大学のデモを示す。

第2部では、伊達氏による Globus を用いたグリッド環境構築と題して、Globus2.0 のインストール方法やジョブスケジューラーの PBS、グリッド環境上で MPI プログラムを実行可能にする MPICH-G2 を Globus に組み込む方法が述べられた。

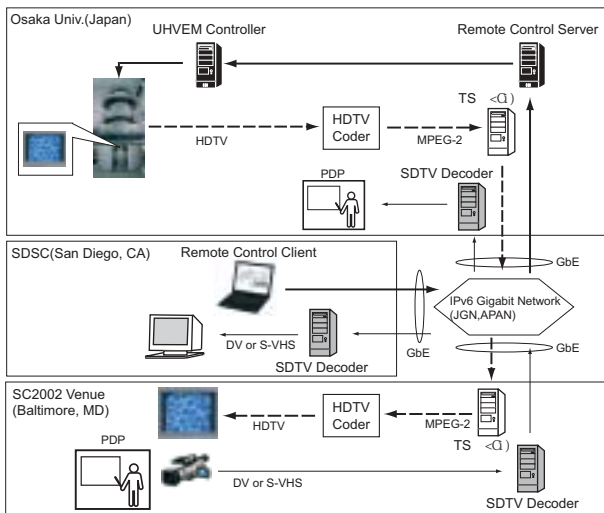


Fig. 3 SC2002 のデモ

2.3 ゲノム情報解析とそのためのグリッド(松田 秀雄 氏)

Fig. 4 に NCBI²⁾ で示されていたゲノムデータの登録数の増加を示す。ここで示したように、ゲノム情報解析では、大量のデータが発生し、データとその解析結果の体系的利用が必要であるとされている。そのためには、大量のデータ、大量の計算処理、生物学・医学・薬学・情報科学などの連携が必要であり、本質的にはそれらの機関が分散協調処理を行う基盤すなわちグリッド技術が必要とされている。現在のグリッドミドルウェアは

API が低レベルなために、分散協調処理が可能な高度 Web サービスを指向したミドルウェア開発が重要とされている。

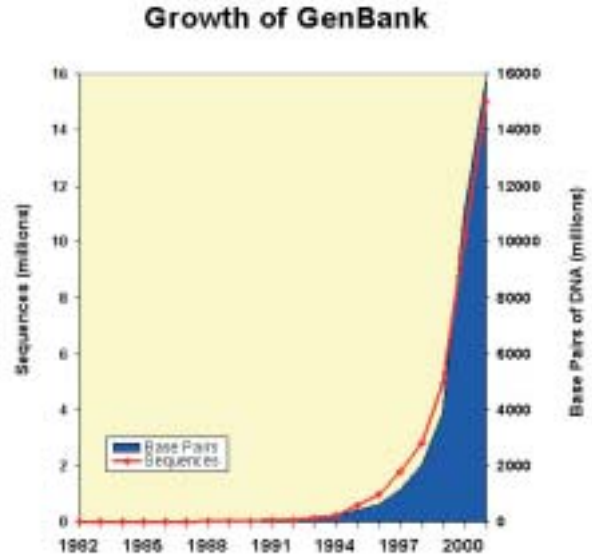


Fig. 4 ゲノムデータの登録数の増加

3 感想

● 岩橋

バイオインフォマティクスへのグリッドの適用はまだまだ課題が山積みであると実感した。キーポイントになるのが、解析や立体構造予測で得られた大量の実験データをデータベースに格納し、それをどのようにして他の遠隔の計算資源へ伝送するのかということである。

● 青井

現在、バイオグリッドの分野で進められているプロジェクトを知ることができた。これらのプロジェクトは、実際のタンパク質を用いて解析を行い、このデータ処理(フーリエ変換やスペクトル解析など)にグリッドを用いるものであった。また、タンパク質グループの行っているシミュレーションによる第一原理からの立体構造予測は日本のバイオの研究分野においてはごくまれであると感じた。

参考文献

- 1) 大阪大学 超高压電子顕微鏡センター.
<http://www.uhvm.osaka-u.ac.jp/index-j.html>
- 2) National Center for Biotechnology Information.
<http://www.ncbi.nlm.nih.gov/Genbank/genbankstats.html>