



Open Bioinformatics Grid (OBIGrid)

バイオとインフォマティクスの
真の融合を目指して

小長谷 明彦

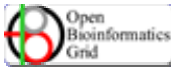
北陸先端科学技術大学院大学

知識科学研究科 教授

理化学研究所ゲノム科学総合研究センター

ゲノム情報科学研究グループ プロジェクトディレクター

Copyright(C) 2002 OBIGrid, All Rights Reserved



内容

- OBIGridとは？
- バイオインフォマティクスからの要求
- OBIGrid設計思想
 - ネットワーク設計
 - 利用環境
- 今後の予定

Copyright(C) 2002 OBIGrid, All Rights Reserved



OBIGridとは？

Copyright(C) 2002 OBIGrid, All Rights Reserved



目的

バイオインフォマティクス研究者のためのグリッド

ネットワークに接続するだけで、研究に必要な

- データベース
- シミュレーション環境
- プログラミング言語
- 実験データ

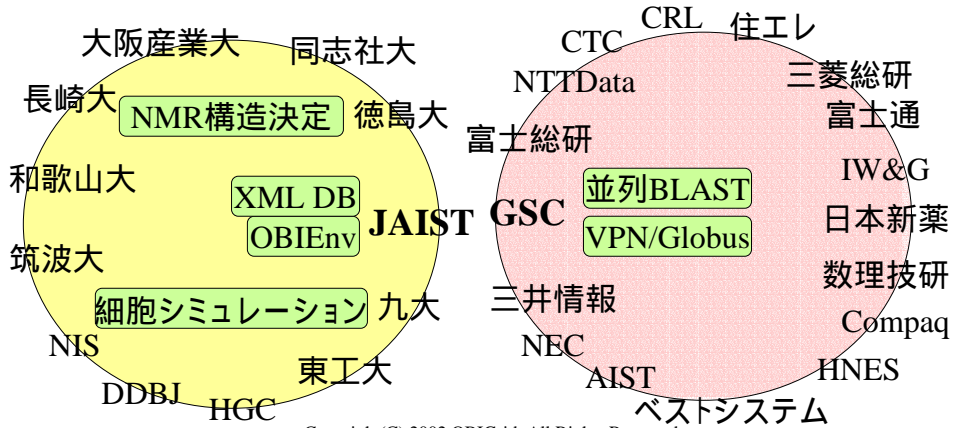
にアクセス可能な利用環境を目指す。

Copyright(C) 2002 OBIGrid, All Rights Reserved

参加メンバー

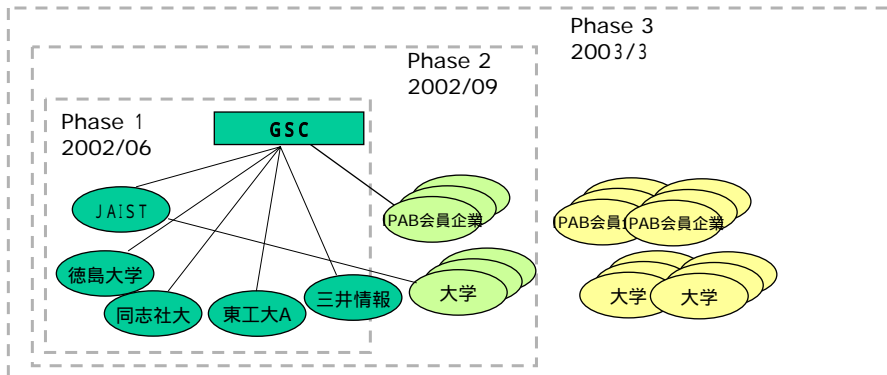
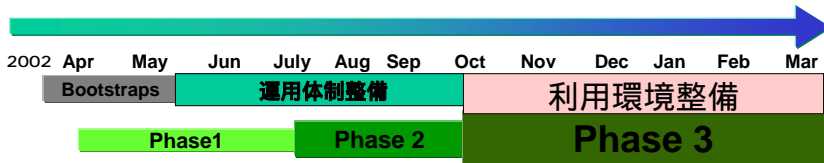
文部科学省
 科研特定領域研究ゲノム情報科学
 ソフトウェア共有化および高速化委員会
 グリッド構築ワーキンググループ

並列生物情報処理
 イニシアティブ(IPAB)
 グリッド構築ワーキンググループ
<http://www.ipab.org>



Copyright(C) 2002 OBIGrid, All Rights Reserved

2002年度スケジュール



Copyright(C) 2002 OBIGrid, All Rights Reserved



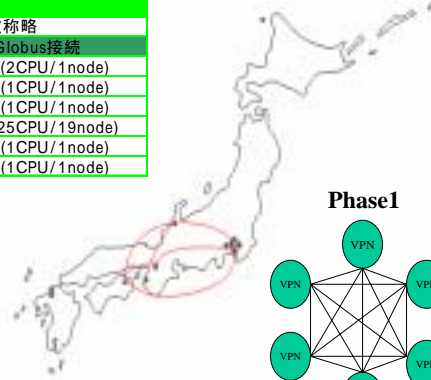
OBIGrid フェーズ1(2002/4月-7月)

•VPN + Linux(IA32)+Globus v1.1.4による
グリッド網構築実験

Phase1の接続状況(2002/7/18現在) * 敬称略		
	VPN接続	Globus接続
GSC	完了	完了(2CPU/1node)
北陸先端大	完了	完了(1CPU/1node)
徳島大	完了	完了(1CPU/1node)
東工大	完了	完了(25CPU/19node)
同志社大	完了	完了(1CPU/1node)
三井情報開発(株)	完了	完了(1CPU/1node)

VPNフルメッシュ接続
専用CA局

<http://www.obigrid.org>

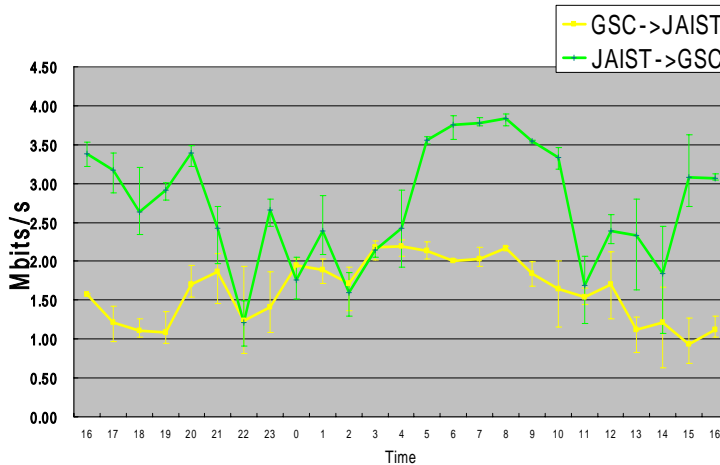


フルメッシュ構成

Copyright(C) 2002 OBIGrid, All Rights Reserved

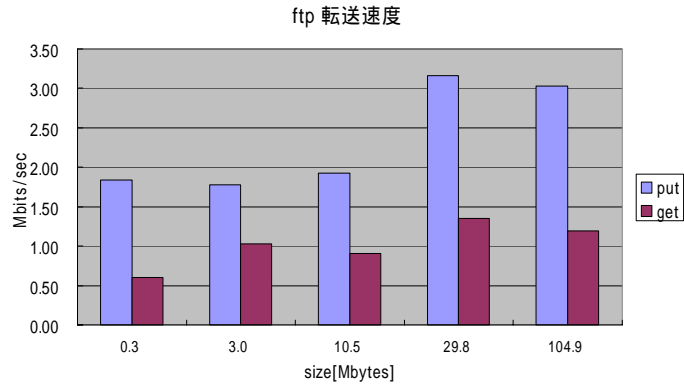


netperf測定結果 (JAIST GSC)



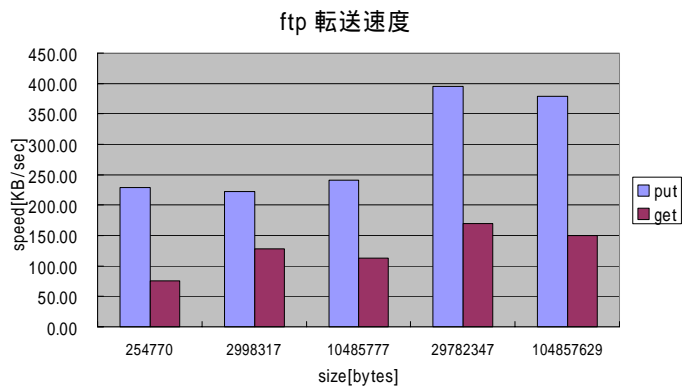
Copyright(C) 2002 OBIGrid, All Rights Reserved

Ftp性能測定結果 (JAIST GSC)



Copyright(C) 2002 OBIGrid, All Rights Reserved

Ftp性能測定結果 (JAIST GSC)



Copyright(C) 2002 OBIGrid, All Rights Reserved

まとめ

- VPN + Globus網で**6サイト**接続を確認
- インターネット経由でのVPN転送性能は
1 - 3 Mbit/秒程度 (**10Mbps** VPN装置使用時)
- Globus網構築のための**運用作業項目**の洗い出し



- 接続サイト数の拡大
(現行6サイト稼動中、年度内 **20 ~ 30サイト**を予定)
- バイオインフォマティクス利用環境(**OBIEnv**)を開発中
11月28日 版リリース予定
- 各種**バイオインフォマティクスアプリケーション**準備中

Copyright(C) 2002 OBIGrid, All Rights Reserved

バイオインフォマティクス からの要求

背景

- ゲノム解析プロジェクトの進展
 - ⇒ **ホモロジー検索**の需要拡大
- 構造ゲノミクスプロジェクトの始動
 - ⇒ **分子シミュレーション**への期待
- 細胞/個体変異解析への関心
 - ⇒ **細胞シミュレーション**の試み

➡ データ、計算資源、人的資源の有機的な連携が不可欠

➡ **グリッド技術**に注力

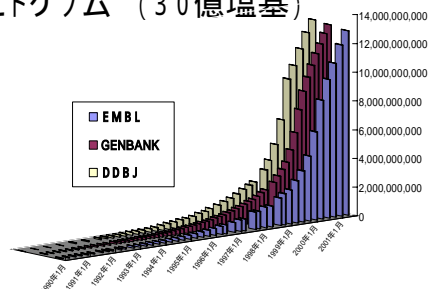
Copyright(C) 2002 OBIGrid, All Rights Reserved

課題1 ホモロジー検索

バイオインフォマティクス研究の計算機資源の7割は
BLAST検索に使われている！

- データ量の増大 (**毎年倍増**、**日更新**、1747万エントリ、50GB)
- パラメータ依存性
ギャップコスト、アミノ酸置換行列
- 組み合わせ的計算量の増大
マウスcDNA (5万配列)* ヒトゲノム (30億塩基)

1cpu 100日
10cpu 10日
100cpu 1日



Copyright(C) 2002 OBIGrid, All Rights Reserved

課題2 分子シミュレーション

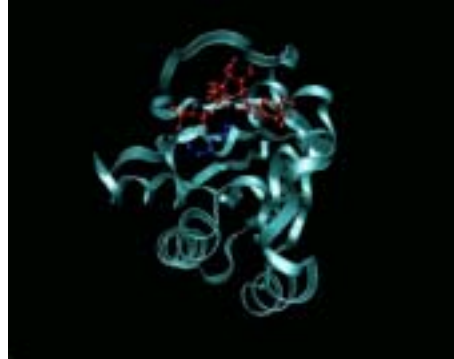
10万原子、ナノ秒レベルでのシミュレーションができる
よくなると応用範囲急拡大

- 変異構造安定性
- リガンド結合
- タンパク間相互作用
- タンパク複合体安定性

of residue: 189
Molecular weight: 21kD
Oncogene Variant
Gly12 Val
5ns

1000h/32Gflops

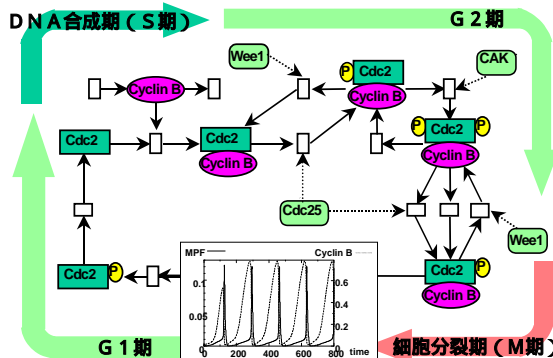
Ras p21 G Molecular Dynamics Simulation



Copyright(C) 2002 OBIGrid, All Rights Reserved

課題3 細胞シミュレーション

生命現象を分子の言葉で説明する。



相互作用シミュレーション

未知パラメタ推定

実験データ収集(相互作用、速度パラメタ)

54パラメタ推定

4時間/50cpu

Copyright(C) 2002 OBIGrid, All Rights Reserved

まとめ

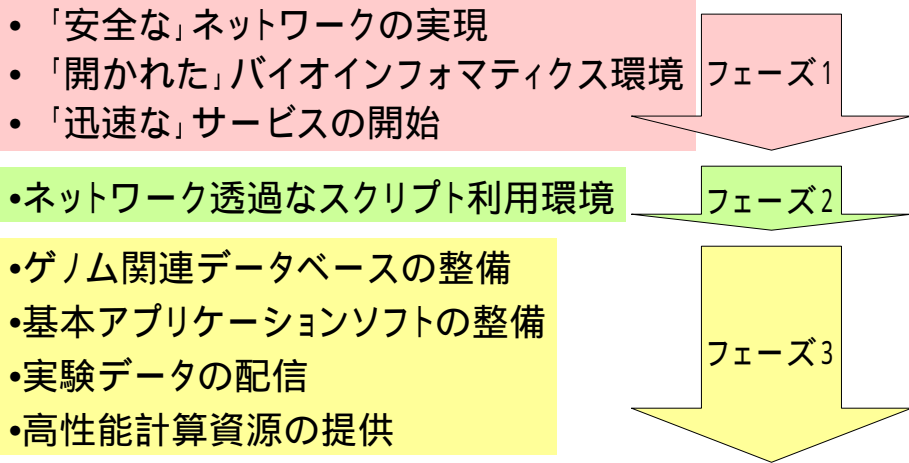
- 増大する配列データ
→ 管理オーバーヘッドの削減、並列検索
- シミュレーションへの関心
→ 高性能計算機の利用
- 遺伝子機能解析へのシフト
→ 実験データ、知識の共有

Copyright(C) 2002 OBIGrid, All Rights Reserved

OBIGrid設計思想

Copyright(C) 2002 OBIGrid, All Rights Reserved

設計目標

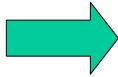


Copyright(C) 2002 OBIGrid, All Rights Reserved

ネットワーク設計

安全なネットワークの実現

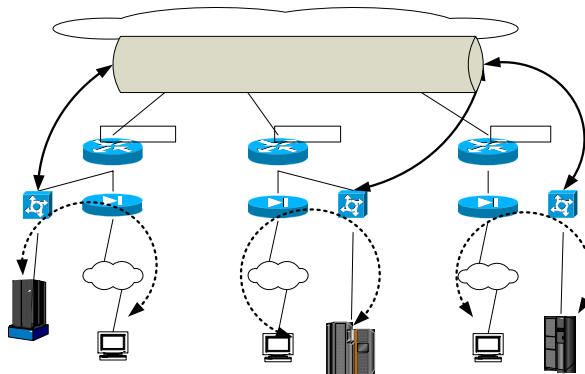
- エンドツーエンドでのアプリケーションおよびデータのセキュリティの確保
- グリッドの持つパワーを利用した、グリッド外への攻撃の回避



- VPNによるネットワークの分離
- 外部ネットワークとの通信制限

Copyright(C) 2002 OBIGrid, All Rights Reserved

基本ネットワークアーキテクチャ



グリッドへ参加する計算機群を、仮想プライベートネットワーク(VPN)で囲い、さらに、各計算機群を組織内部のネットワークから隔離することで、エンドツーエンドのセキュリティを確保。

Copyright(C) 2002 OBIGrid, All Rights Reserved

Network & Security Policy

- OBIグリッドに参加する各サイト間を、IPsecを使用しVPN接続する。
- VPN装置は、**各サイトのFirewall外に設置し**、Global IPアドレスを付与する。
- 各サイトのVPN装置の内側に設置されるネットワーク/計算機は、各サイトのローカルネットワークとは隔離されていること(VPN装置のみが入り口となり、**Dual Homing禁止**)。
- 各サイトに設置されるVPN装置は、事務局側から一括管理をおこなう。
- ユーザは、各自のPC上にインストールされた**VPNクライアント**(IPsec)を使用して、インターネット経由でVPN網内へアクセスする。(または、直接計算機の前へ行って使用)
- **インターネットから、VPN網内へのアクセス**は、メールやDNSなど最低限の通信を除き、基本的に**禁止**する。
- **VPN網内から、インターネットへのアクセス**は、メールやDNSなど最低限の通信を除き、基本的に**禁止**する。
- 接続する各計算機の物理セキュリティは、各サイトの責任で確保する。
- 接続する各計算機のソフトウェアセキュリティは、各サイトの責任で確保する。
- 各ユーザの身元は、各サイトが保証する。



ニーズに応じて、今後緩和

Copyright(C) 2002 OBIGrid, All Rights Reserved

Why VPN?

	概要	評価				総合評価
		セキュリティ	拡張性	Gridへの参加しやすさ	グリッド外部への影響排除	
案1 VPN接続	Gridに接続されるサーバ間をVPNトンネル接続する					
案2 Firewall越え	Gridに接続されるサーバ間を、既存のFirewallを越えて接続する				×	
案3 直接インターネット上に配置	Gridに接続されるサーバを全て、データセンタへ配置する	×		×	×	×

- グリッドの周囲をVPNで完全に囲ってしまう案1のみが、設計条件を満足することが可能と判断
- Globus Toolkit(1.1.4)は、ランダムに双方向のポートを複数使用するなど、Firewallとの親和性が低く、サイトの増加にあわせて、Firewallの設定を変更する必要があり、拡張性に乏しい
- 直接インターネット上に設置するのは、あまりにもセキュリティ的に危険



VPNを必要としないセキュアなグリッド技術の開発が急務!!!

Copyright(C) 2002 OBIGrid, All Rights Reserved

開かれたバイオインフォマティクス環境

- 手軽なエンドユーザー環境
 - インターネット環境からのアクセス
 - PCサーバーによる参加
- スケーラビリティ
 - 接続サイト数 > 10 (30 ~ 50)
 - 利用人数 > 100 (200 ~ 300)



- **Publicユーザー**と**developerユーザー**の分離
- ローカルサイトによる**publicユーザー**認証
- 「**代表アカウント**」によるリモート計算資源の利用

Copyright(C) 2002 OBIGrid, All Rights Reserved

Account Policy

- ユーザとしては、public user, developer userの2種類を想定
- **Public user**は、あるマシンにローカルアカウントを所有するそのサイトに所属するユーザで、**OBIEnv上でのみアプリケーションを実行可能**。リモートマシン上では、Globusの**代表アカウント/代表ローカルアカウント**にて処理を実行。当該**サイト管理者の責任でユーザ管理**。
- **Developer user**は、専用のglobus アカウントとマシン上に専用のローカルアカウントを所有し、**任意の処理を実行することが可能**。サイト管理者の承認の元でOBIグリッド事務局で**Globus Account**の管理を行い、マシンを利用する場合には、該当する**サイト管理者にローカルアカウントの発行を依頼**。
- Developer Userを受け入れるかどうかは、各サイトの判断(Policy)である。
- OBIEnvが稼動するマシンには原則としてdeveloper userを受け入れない。



迅速なサービスの開始

- 利用技術の取捨選択
- 運用負荷の軽減
- 事務作業の簡素化



- 既存技術の組合せ(VPN+Globus)による運用
- グリッド技術(Globus)と利用環境(OBIEnv)とのインターフェースの最小化
- 試験運用に問題ない範囲でのCA局の運用
- 負荷の少ないグリッド運用ポリシーの策定

Copyright(C) 2002 OBIGrid, All Rights Reserved



グリッドの運用ポリシー

- 基本方針
 - 各サイトは、最低1名のサイト管理者を任命
 - **計算機そのものに関わる運用作業は、各サイトで実施**
(計算機の障害監視や、Globusのmapファイルへの登録など)
 - **OBIGrid事務局においては、グリッド全体に関わる運用作業を実施**
- 事務局の業務
 - OBIGrid CA局の運用 (Host/User証明書の発行)
 - コアアカウントの登録・変更・削除
 - グリッドユーザからの計算機利用申請(mapファイル登録処理)の各サイト管理者への取次ぎ
 - VPN網の運用管理
 - 各サイトに対する技術支援

Copyright(C) 2002 OBIGrid, All Rights Reserved

まとめ

- 現状のGlobus Toolkit (v1.1.4/v2)による実インターネット上でのグリッド展開は、セキュリティ上のリスクを伴う。
- グリッドの持つ計算パワーを、グリッド外に行使されるのを防ぐための仕組みが必要。
- グリッドのネットワーク設計においては、各サイトのネットワークセキュリティポリシーに合致するように、仮想組織としてのポリシー決定が必要。



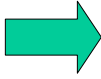
OGSAをベースとするGlobus3.0にて、インターネットを越える環境への実装が容易となることを期待

Copyright(C) 2002 OBIGrid, All Rights Reserved

利用環境

ネットワーク透明なスクリプト利用環境

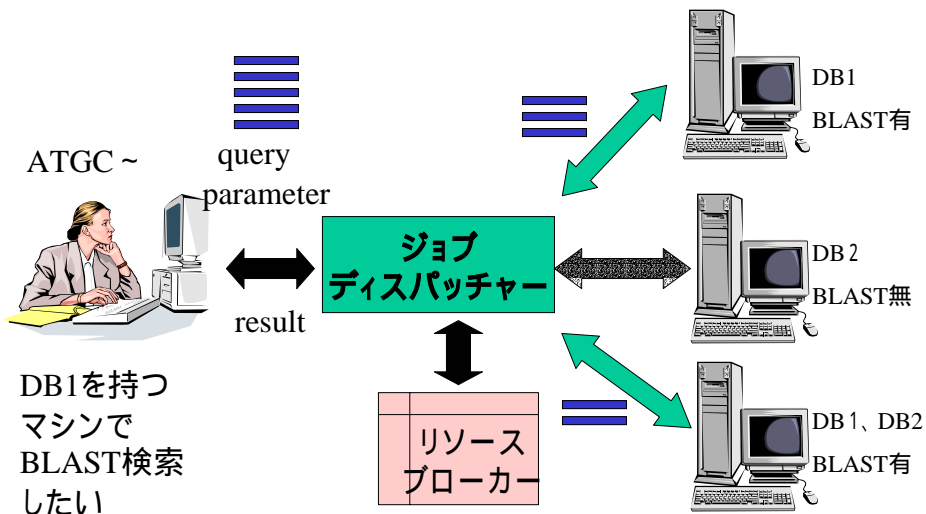
- スクリプトによる**遠隔大量データ処理**の実現
- 条件に適したマシン群の検索とジョブの投入
CPU、プログラム、データ



- リソースマネージャーによる資源管理
- リソースブローカーによるジョブ配分
- 推奨ソフトウェア環境によるスクリプト動作保証
(必ずしも、同一のソフトウェア環境は要求しない)

Copyright(C) 2002 OBIGrid, All Rights Reserved

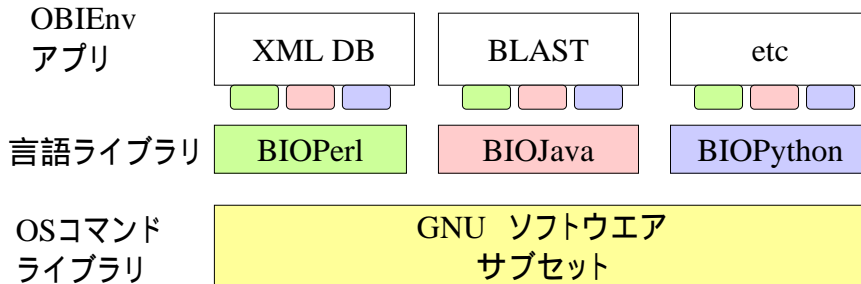
OBIEnvでの実行イメージ



Copyright(C) 2002 OBIGrid, All Rights Reserved



Grid接続マシンのソフトウェア推奨環境



Copyright(C) 2002 OBIGrid, All Rights Reserved



GNU Software(1)

- fileutils(Tools for file operations)
ls, cp, ...
- Sed(A stream-oriented non-interactive text editor)
- gawk(String manipulation language)
- Grep(Finds lines that match entered patterns)
grep, egrep, fgrep
- gzip(Compresses and decompresses files)
gzip, gunzip

Copyright(C) 2002 OBIGrid, All Rights Reserved

GNU Software(2)

- GNU tar(Creates tar archives)
- Diffutils(FINDs differences between and among files)
diff, cmp, sdiff, diff3
- Textutils(Text utilities)
cat, cut, ...
- bash(Shell of the GNU operating system)
- ash, bsh, tcsh, csh

Copyright(C) 2002 OBIGrid, All Rights Reserved

まとめ

- スクリプト利用を指向した環境構築
- リソース情報収集、リモートマシンでのプログラム実行に必要な仕組み
- 推奨ソフトウェア環境の定義



まずは、動く環境の実現を優先し、
スクラップ&ビルドで仕様を決定



今後の予定

Copyright(C) 2002 OBIGrid, All Rights Reserved

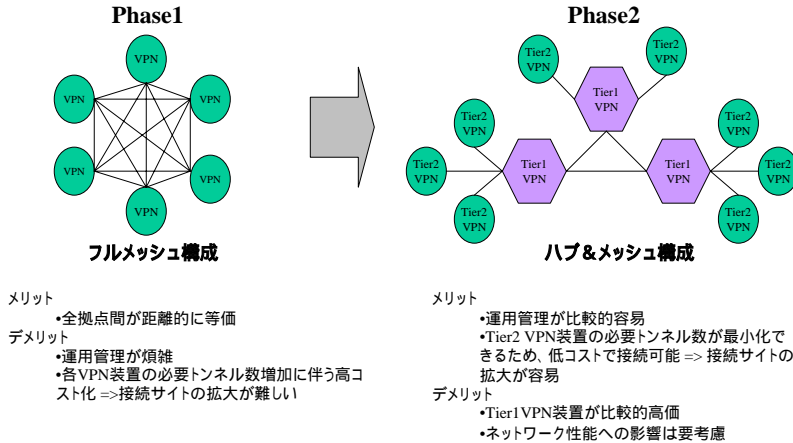


OBIGridフェーズ2 (7月-9月)

- 接続サイトの拡大 (*Net2*)
 - 6サイトから、**20サイト以上**を予定。VPN網を再設計。
- Globus ToolkitのV2化 (*Globus2*)
 - **インストレーションサマーキャンプ**を実施(9/5,6 at JAIST)
- Globus Toolkitのパッケージ化
 - Globus Toolkit V2 + MPCHI- G + Ninf- G等

Copyright(C) 2002 OBIGrid, All Rights Reserved

OBIGRID Network



Copyright(C) 2002 OBIGrid, All Rights Reserved

Phase2接続サイトの募集

- 参加条件/規約 <http://www.obigrid.org/tech/>
 - 募集サイト数 特に制限なし
 - グリッド接続時期 9月中旬以降順次開始
 - VPN装置 共同購入にてTier2で約12万円
- お問い合わせは、 info@obigrid.org まで

Copyright(C) 2002 OBIGrid, All Rights Reserved

会合スケジュール

- Globus 2.0 インストールサマーキャンプ
(9月6、7日 於 JAIST)
- 第3回オープンバイオインフォマティクスフォーラム
(9月20日 於 鶴見GSC)
- IPABシンポジウム
(11月29日 於 品川プリンスホテル)
OBIEEnv リリース予定
- 第4回オープンバイオインフォマティクスフォーラム
(1月17日 於 鶴見GSC)
- 第5回オープンバイオインフォマティクスフォーラム
(3月20日 於 JAIST)

Copyright(C) 2002 OBIGrid, All Rights Reserved

グリッド協議会への期待

- グリッドイメージの具体化
何ができて、何ができないのか？
- Public Grid Hosting Service
個別にグリッドを立ち上げるには、大きな努力が必要。
多くのVOを容易に、かつSecureにHostingできるような、
グリッドインフラ整備が不可欠
- 新グリッド技術の開発と普及



今後のOBIGridの活動は**グリッド応用システム**
開発に注力

Copyright(C) 2002 OBIGrid, All Rights Reserved



謝辞

敬称略

VPN/Globus網構築 梅田 (IBM)、小西 (GSC)

OBIEnv構築 佐藤、山本、Defago (JAIST)
辻、中島 (HNES)

事務担当 蛭川 (JAIST)、佐竹、遠藤 (GSC)

科研特定領域研究ゲノム情報科学
ソフトウェア共有化および高速化委員会
グリッド構築ワーキンググループ

岡本 (九大)、小野 (徳島)、小坂 (大阪産業大)、染谷 (長崎大)、
田中 (東工大)、中田 (AIST、東工大)、中西 (九大)、廣安 (同志社大)、
福田 (九大)、松岡 (東工大)、丸山 (筑波大)、山村 (東工大)

並列生物情報処理イニシアティブ (IPAB)
グリッド構築ワーキンググループ

株式会社 NTTデータ
コグニティブリサーチラボ 株式会社
コンパックコンピュータ 株式会社
独立行政法人 産業技術総合研究所
株式会社 数理技研
住商エレクトロニクス 株式会社

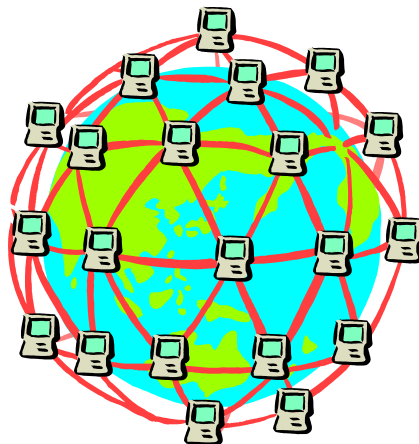
日本アイ・ピー・エム 株式会社
日本新薬 株式会社
日本電気 株式会社
株式会社 富士総合研究所
富士通 株式会社
株式会社 ベストシステムズ

株式会社 三菱総合研究所 北陸日本電気ソフトウェア 株式会社
理化学研究所 横浜研究所 (GSC)
株式会社 理経
有限会社 エスカ
インテック・ウェア・アソシエーツ・ゲノム・インフォマティクス(株)
CTCラボラトリーシステムズ 株式会社

Copyright(C) 2002 OBIGrid, All Rights Reserved



Open Bioinformatics Grid



End

Copyright(C) 2002 OBIGrid, All Rights Reserved