

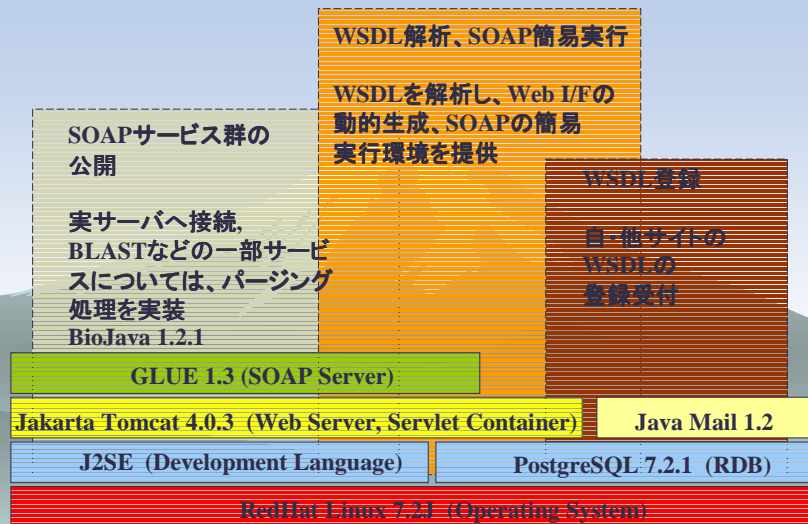
X-DDBJにおける 巨大オブジェクトと長時間ジョブ への対応

国立遺伝学研究所
生命情報・DDBJ研究センター
菅原秀明、宮崎 智
富士通株式会社
ライフサイエンスシステム事業部
重元康昌、山口政仁

生命情報・DDBJ研究センター

- ◆塩基配列データおよび付随データの検索・解析
サービス向上の一環としてWebサービスを開始
- ◆GRIDにおける大規模進化機構解析

X-DDBJの組み立て(サーバー)



2003/06/18

国立遺伝学研究所生命情報・DDBJ研究センター

3

X-DDBJの組み立て(クライアント)

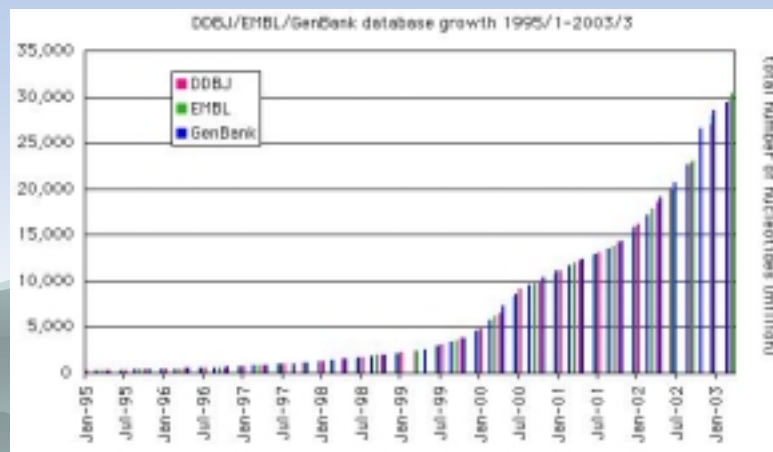
- ◆ Perl
 - SOAP-Lite
 - libwww-perl
 - MIME-Base64
 - URI
 - XML-Parser
- ◆ Java
 - GLUE

2003/06/18

国立遺伝学研究所生命情報・DDBJ研究センター

4

DBの動向:塩基配列DBの膨張



Release 53(March 2003) 29,711,299,332塩基 (23,250,813件)
四半期増加率10% → 年1.5倍

2003/06/18

国立遺伝学研究所生命情報・DDBJ研究センター

5

DBの動向:配列の制限長撤廃

従来:20年来の歴史的経緯において、350Kbpを
上限としてきた。

今後:1995年以後の自立した生物のゲノム配列
決定の隆盛と計算機資源の充実から、
この制限長を撤廃。

現象:メガbps単位でのデータを受付公開する
ことになる。

2003/06/18

国立遺伝学研究所生命情報・DDBJ研究センター

6

2003/06/18 国立遺伝学研究所生命情報・DDBJ研究センター 7

DBの動向: データ量増大の影響

キーワード検索、相同性検索、多重整列、遺伝子(候補)配列からのタンパク質立体構造予測などのサービスに対するprogrammable interfaceを提供する。

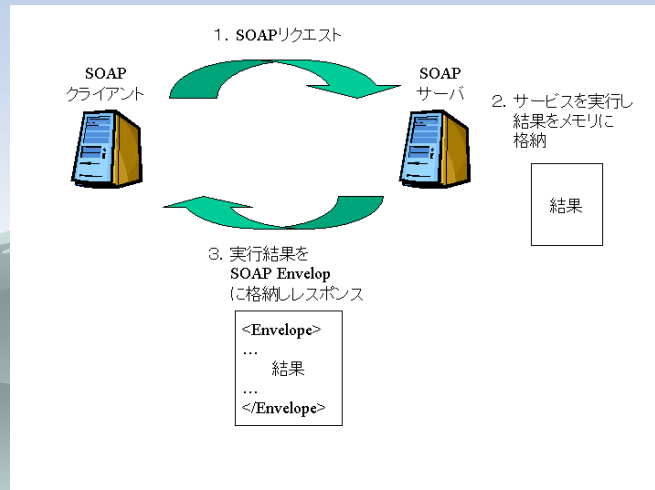
データ総量が増える
エントリーあたりのデータ量も桁違いに増える

転送するデータ量が大きくなる
処理時間が長くなる

2003/06/18 国立遺伝学研究所生命情報・DDBJ研究センター 8

巨大オブジェクトへの対応

一般のSOAPサービスにおけるデータ通信



2003/06/18

国立遺伝学研究所生命情報・DDBJ研究センター

9

巨大オブジェクトへの対応

- ◆ ステップ2のサービス実行結果が数キロバイト程度であった場合はレスポンスに問題はなかった。
- ◆ 数十メガバイト程度に大きくなるとシステム負荷が非常に高くなった。
- ◆ ステップ3実行時にXMLをパーシングする時間が長時間になりレスポンスが非常に悪くなった。

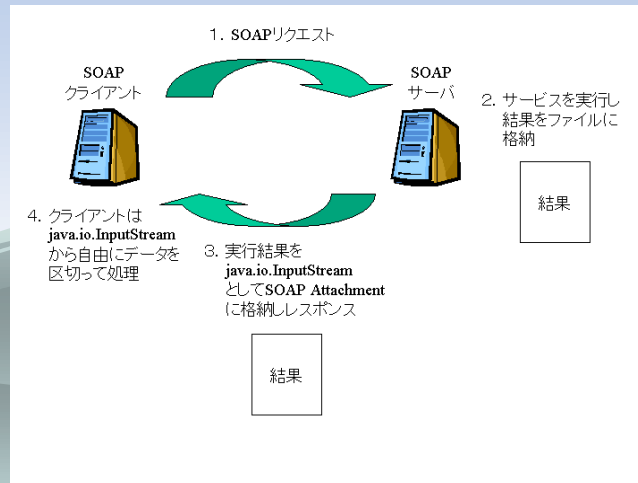
2003/06/18

国立遺伝学研究所生命情報・DDBJ研究センター

10

巨大オブジェクトへの対応

SOAP Attachmentとjava.io.InputStream利用



2003/06/18

国立遺伝学研究所生命情報・DDBJ研究センター

11

巨大オブジェクトへの対応

- ◆ テキストデータもAttachment経由で通信
 - ❖ → XMLパーシングの処理を省略
 - ◆ → SOAPクライアントでの処理が軽減
- ◆ java.io.InputStream技術
 - ❖ SOAPクライアントでは、結果を少しずつ受け取り必要な処理を実行
 - ❖ サービスの結果(例 相同性検索の結果10Mバイト)を、順次、送り出す量だけ(数キロバイト程度)メモリーを消費
- ◆ 数ギガバイト程度の転送が実用になることを確認

2003/06/18

国立遺伝学研究所生命情報・DDBJ研究センター

12

長時間ジョブへの対応

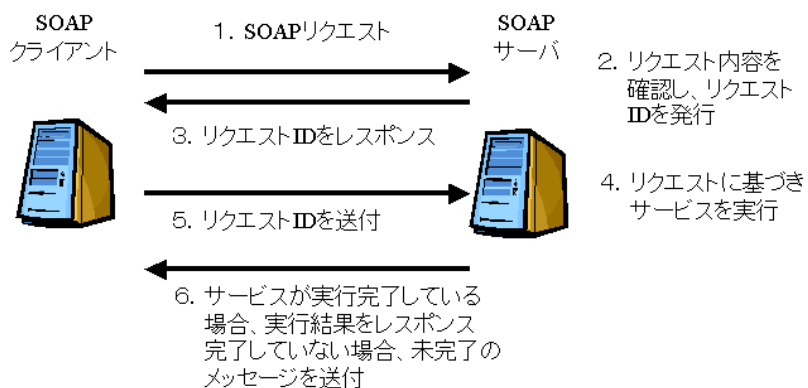
- ◆ 「XML Central of DDBJ」でこれまで提供してきたサービスは同期通信
- ◆ したがって、長時間のジョブについては、要求元は待ち状態になり、接続し続ける必要があった。
- ◆ そこで、結果取得を別のタイミングで取得するような非同期での通信処理を開発した。

2003/06/18

国立遺伝学研究所生命情報・DDBJ研究センター

13

長時間ジョブへの対応



2003/06/18

国立遺伝学研究所生命情報・DDBJ研究センター

14

期待される応用例

- ◆ 新規ゲノム配列一本をQUERYとする相同性検索
- ◆ ゲノム対ゲノムの多重整列
- ◆ 全生物種を対象とする系統解析
- ◆ 特定の生物種の遺伝子配列の網羅的検索
- ◆ 特定の遺伝子配列の網羅的検索

2003/06/18

国立遺伝学研究所生命情報・DDBJ研究センター

15



2003/06/18

国立遺伝学研究所生命情報・DDBJ研究センター

16