



globusWORLD 参加報告

2003 Jan.13-17, San Diego, CA.

中島 研吾

財団法人 高度情報科学技術研究機構

2003年2月26日 第5回GGF調査会

03/02/26, globusWORLD KN



注目点

- 背景
 - 自分はいわゆるアプリケーションサイドの人間
 - GTの利用経験はない
- 主として「TRACK 1: Enterprise Planning for Grids and the GT」に出席した。
 - perfect for those responsible for strategic management
- 実際にどのように使用されているか？
 - 現状, 問題点, 将来

IBM Redbook

Introduction to Grid Computing with Globus

- <http://www.ibm.com/redbooks/>
 - PDF, html version を無料で入手可能
 - ハードコピーを注文することもできる(\$35+送料)
- This redbook covers the following topics:
 - Grid computing fundamentals
 - Architecture and security considerations
 - Introduction of OGSA
 - Description of the components of the Globus Toolkit
 - Globus Toolkit Version 2.2 implementation

3

全体的な印象

- 広い分野で使用されている。
 - 科学技術分野, 研究分野ばかりではない。Industryへの利用。
 - Data Grid, Resource Management
- 問題点
 - 同じことが少しずつ違った言い方で常に言われていた。
 - Security: 技術的, 政治的。
 - 顧客のプライバシー保持とGridのOpen, Shared思想の対立。
 - 専門, 国, 組織, 文化 (Academia vs. Industry) の違い
 - Legacy Application, Legacy Data Style
 - Any Type of H/W, OS etc.

4

出席セッション(1/2)

パネルディスカッション中心

- 1月13日(月)
 - Tutorial 1: Introduction to Grids, GT and OGSA
- 1月14日(火)
 - 0830-0900 Opening
 - 0900-1000 Keynote (I.Foster)
 - 1030-1200 Operational Grids by GT (Tr.1)
 - 1330-1530 Grid Appl. in Industry (Tr.1)
 - 1600-1800 Product Strategies for GT & OGSA (Tr.1)
- 1月15日(水)
 - 0830-0915 Keynote (Cyberinfrastructure)
 - 0900-1000 Keynote (Condor)
 - 1030-1200 Security (Tr.1)

5

出席セッション(2/2)

- 1月15日(水)(続き)
 - 1330-1530 How is GT used ? (Tr.2)
 - 1600-1630 Grid Standards (Tr.1)
 - 1630-1800 Grid Futures (Tr.1)
- 1月16日(木)
 - 0830-0915 Keynote (Ceyba: Optical Equipment Vendor)
 - 0900-1000 Keynote (BIRN: Biomedical Informatics Research Network)
 - 1030-1130 QA with Globus Team
 - 1130-1230 National/International Cyberinfrastructure
 - 1330-1800 Workshop-3: Life Sciences


6



Agenda Overview

- Monday
 - 2 tutorials
 - Opening reception (6:00 – 8:00pm)
 - Poster and sponsor displays, drinks, snacks
- Tuesday – Thursday lunch
 - Conference: 3 parallel tracks
 - Wednesday Evening Forum (7:00-10:30pm)
 - BBQ Dinner, poster and sponsor displays
- Thursday afternoon – Friday
 - 1 tutorial, 3 workshops

7



Conference Tracks

- Track 1: Enterprise Planning for Grids and the Globus Toolkit
 - Target: Strategic management
- Track 2: Architecting Grids with the Globus Toolkit (GT2 & GT3)
 - Target: Project management
- Track 3: Developing & Administering Globus Toolkit v2 (GT2)
 - Target: Developers & Administrators

8



Tutorials

- Introduction to Grids, the Globus Toolkit, and OGSA (Monday)
- Grid Services and Web Services (Monday)

- Globus Toolkit v2 Developers and Administrators Tutorial (Tuesday-Thursday, track 3)
- Globus Toolkit v3 Developers and Administrators Tutorial (Thursday-Friday)



Workshops

(Thursday 12:30 – Friday 1:30)

- Grid Security
 - Organized by: Frank Siebenlist (ANL), Von Welch (U Chicago), Marty Humphrey (U. Virginia)
- Data
 - Organized by: Ann Chervenak (USC/ISI), Ian Foster (ANL & UC)
- Grid in Life Sciences
 - Organized by: Steve Tuecke (ANL), Christine Miller (IBM), Andrew Grimshaw (Avaki)

Registration Numbers (approximate)

- 450 registrations
 - 150 industry
 - 240 government & academic
 - 35 student
 - 100 non-US
 - 200 for Monday tutorials
 - 220 for Thursday-Friday tutorials and workshops

Operational Grids by GT

- 出席者
 - NEESgrid <http://www.neesgrid.org/>
 - CERN/EU Data Grid <http://eu-datagrid.web.cern.ch/eu-datagrid/>
 - NASA/LBNL <http://www.ipg.nasa.gov>, <http://www.doesciencegrid.org>
 - NorduGrid <http://www.nordugrid.org/>
 - ANL/NSF Tera Grid <http://www.teragrid.org/>
 - Sun
- Globus/Condorについて:F. Gagliardi (CERN/EU Data Grid)
 - Open Source, Modularity, Flexibility, Easy Access, Good Support
 - Issues
 - R&D vs. Industry Quality Support
 - GT2-GT3
 - Different Funding/Goal -> Sharing Software

NEESgrid

<http://www.neesgrid.org/>



- NEESgrid will link earthquake researchers across the U.S. with leading-edge computing resources and research equipment, allowing collaborative teams (including remote participants) to plan, perform, and publish their experiments. NEESgrid, the systems integration component of the NEES project, uses the newest and fastest communications technologies to tie the NEES network together.



NEES

<http://www.nees.org/>



- The George E. Brown, Jr. **Network for Earthquake Engineering Simulation (NEES)** is an ambitious national program to advance the study of earthquake engineering and find new ways to reduce the hazard earthquakes represent to life and property. NEES represents an unprecedented effort, in which research institutions across the country will use simulation— both physical and numerical—to develop increasingly complex, comprehensive, and accurate models of how the infrastructure responds to earthquake loadings. NEES is funded by a Major Research Equipment grant from the National Science Foundation.





Grid Appl's in Industry (1/8)

- IBM <http://www.ibm.com/grid/>
- Butterfly.net <http://www.butterfly.net/>
- GlobeXplorer <http://www.globexplorer.com/>
- Charles Schwab <http://www.schwab.com/>
- Platform <http://www.platform.com/>
- HP <http://www.hp.com/>



Grid Appl's in Industry (2/8)

IBM: On Demand Business

- Government
 - Create large-scale IT infrastructures to drive economic development and/or enable new collaborative government services
 - Taiwan Integrated Grid for Education and Research
- Enterprise Optimization
 - Optimize computing and data assets to improve utilization, efficiency, business continuity and resiliency
 - Butterfly.net
- Research & Development
 - Accelerate and enhance the R&D process by enabling the sharing data and computing power seamlessly for research intensive applications
 - Aventis, Drag Design

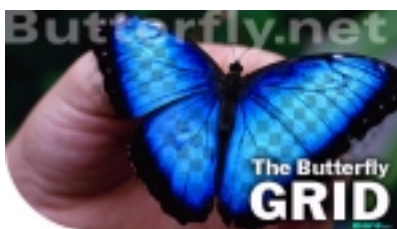
Grid Appl's in Industry (3/8)

IBM: On Demand Business (cont.)

- Engineering & Design
 - Share data and computing power for computing intensive engineering and scientific applications to accelerate product design
 - Microprocessor Design, Benchmarking, Z-Server Design
- Business Analytics
 - Optimize computing and data assets to improve utilization, efficiency, business continuity and resiliency
 - Insurance Firm
- Grid構築のための様々な支援, 教育, 普及活動を実施している。

Grid Appl's in Industry (4/8)

Butterfly.net : Video Game



Grid for MMG
Massively-Multiplayer Games



Grid Appl's in Industry (5/8)

Butterfly.net : Video Game

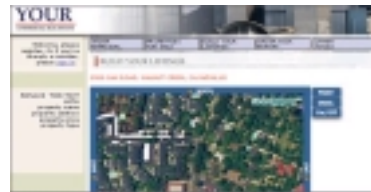
- 人気のあるゲーム, 利用者が集中しているところにリソースを割り当てる:「Applications for Industry」の基本的コンセプト
 - On Demand, IBM
- Butterfly.net is a development studio and technology infrastructure provider for massively-multiplayer online PC, console and mobile games.
- With our unique server grid technology, we've developed the best solution to the challenges of scalability, reliability and performance for persistent-state 3D worlds that must support millions of gamers.
- We've deployed the first commercial server grid for Massively-Multiplayer Games (MMGs), the most exciting and fastest-growing sector of the powerful video game market, in which millions of gamers worldwide face challenges together in real-time, immersive 3D worlds.
- Our MMG platform, dubbed the Butterfly Grid, is available as a fully-managed service, as a packaged software license, or as a complete hardware/software solution.
- Butterfly.net's Grid is the first system to allow MMG game worlds to scale in size and support an unlimited number of players within the same game.
- **Grid does this by allocating communications and computing resources to the most populated areas and most popular games.**

19

Grid Appl's in Industry (6/8)

GlobeXplorer

- 大規模画像配信ビジネス
 - Real Estate, Insurance
 - Travel & Web Portals
 - Mobile & Wiress
- GlobeXplorer saved about 1 million dollars the first year using SUN GridEngine during off-peak hours using our existing assets (production web site and back office machines).
 - Orthorectification
 - Color balancing
 - Mosaicing
 - compression

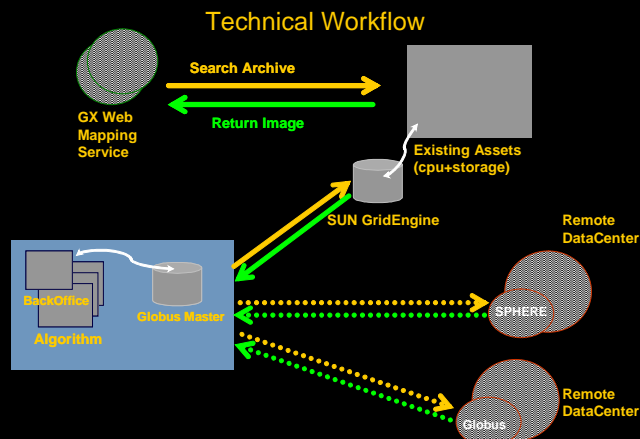


20

Grid Appl's in Industry (7/8)

GlobeXproler:

Remotely Distributed Technology



21

Grid Appl's in Industry (8/8)

Charles Schwab: Grid in Commerce

- Application Usage Profile
 - Jobs not as large
 - Executed over and over, thousands of times
 - Large volumes of users
 - Users typically not just doing one large job
- Failure is not an option
- Database is King
- Programmer Skill
- Administration and Management expectations
- これまでの経験: 大規模な計算をクラスターで試しただけ

22

P.A. Freeman (CISE/NSF)

Keynote: Cyberinfrastructure

- Directorate for Computer & Info. Science & Engineering <http://www.cise.nsf.gov/>
- 様々なプロジェクトの紹介
 - NVO (National Virtual Observatory)
 - Human Genome Sequence
 - NEES
 - EarthScope
 - NEON (National Ecological Observatory Network)
- NSF Cyberinfrastructure について
- 「地球シミュレータ」の悪口

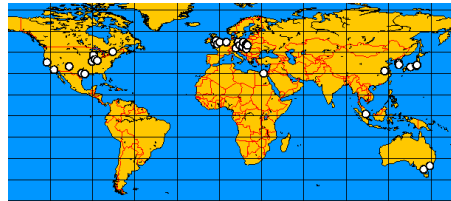
How is GT being used ? (1/3)

- Condor
- ORACLE
- TACC/UT Austin <http://www.tacc.utexas.edu/>
- U.Michigan <http://www.mgrid.umich.edu/>
<http://www.chefproject.org/>
- U.Potsdam/Max-Planck/GridLab
<http://www.gridlab.org/>

How is GT being used ? (2/3)

U.Potsdam/Max-Placnk/GridLab

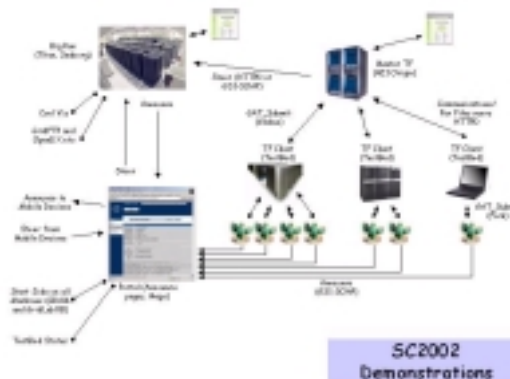
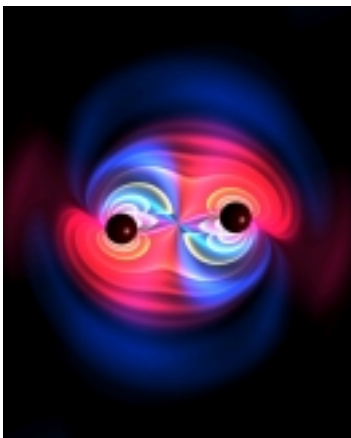
- ブラックホールのシミュレーション
- Dynamic Adaptive Distributed Computation by Cactus/MPI-G2: Gordon Bell in SC2001
- Distributed Computation in SC2002 using Grid Appl. Toolkit (GAT) in GridLab.
 - 70 systems all over the world including Mac/OSX
 - Winner of Bandwidth Challenge
- GridLab.
 - 「Legacy Code on Grid」のためのツールキット
 - Cactusベース, C & Fortran



25

How is GT being used ? (3/3)

U.Potsdam/Max-Placnk/GridLab (cont.)



26

Grid Futures (1/5)



- 基本的にかなり皆さん勝手なことを言っていた。
- Francine Berman (NPACI/UCSD): **他の人が引用**
 - “The promise of the grid has not been oversold, but the difficulty of developing the requisite Grid infrastructure has been underestimated”
- Carl Kesselman (USC)
 - Cost感覚は常に 必要
 - ITが社会で行なわれていることをいかにサポートするかが重要
 - 基本的な考え方: VO, Collaboration: ビジネスの世界でも可能: IBMとMS
 - スタンダードにとられすぎると失敗する: e.g. CORBA
- Rick Stevens (ANL/U.Chicago)
 - PCだけが計算機ではない: PDA, PS2(なぜPS2か?)
 - Embedded Processor Based Products, Grids for those
 - User-Centered Grid, Personal Grids, Grids-on-a-chip ...

27

Grid Futures (2/5)

Grid Technology Partner

www.gridpartners.com

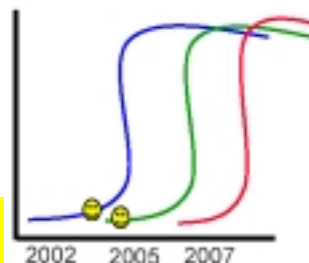
Grid Business Opportunity & Applications

Three Technology S curves:

R&D: starting now peaking in 2005

IT: starting in 2003 peaking around 2007

Consumer: starting in 2007 peaking later in the decade.



まだまだ実生活に入ってくるのには時間がかかるということか？

Grid Technology Partners

Preparation
Globus World - Annual Address 2

28

Grid Futures (3/5)

Grid Technology Partners

www.gridpartners.com

"Gains so large that they are just not believable"

- Monsanto accelerates rice genome discovery by 6 years
-Kathy Stenson, Project Director, Monsanto
- Fluent offers Computational Fluid Dynamics software at US \$ 15 per CPU hour. Per seat license cost was US \$28,000 annually
-Paul Senis, VP Fluent
- American Diabetes Association leverages 250 PCs at its headquarters to reduce Archimedes simulation time from 8 hours to 8 minutes
- Dr. Richard Kahn, Chief Scientist ADA
- UC Berkeley saves US \$800,000 on a research server for MathCAD and Mathematica applications by leveraging PCs in MBA computer lab.
- Dr. Alex Karapetian, Researcher, UC Berkeley

 Grid Technology Partners

Proprietary
GLOBUS WORLD - MITRE/ASDC 4

Grid Futures (4/5)

Grid Technology Partners

www.gridpartners.com

Stumbling Blocks

4. Security
3. Organizational Inertia
2. Standards
1. Applications, Applications & Applications

 Grid Technology Partners

Proprietary
GLOBUS WORLD - MITRE/ASDC 4

Grid Futures (5/5)

Grid Technology Partners

www.gridpartners.com



But if We are Successful...

- Internet Democratizes Information ... Grid Democratizes Computation
 - Gateway Grid at 10 cent per CPU hour
- Little Men escape to become Free Professionals (C. Wright Mills © 1953)
- Demographics + Grid changes Global Competitive Picture

Grid Technology Partners

Proprietary
Globus World - Asterix

F. Bermann (NPACI/SDSC) Cyberinfrastructure Panel



- TeraGrid by friends is already difficult.
- Cyberinfrastructure: more challenges
 - Programming Model
 - Programming Environment
 - Logistical, Legal, Ideological Challenges
- International Issues
 - Resources
 - Policy
 - Time Sharing: 9-to-5

Distributed Computing開発指針 Open Forumでの質疑

- 質問したのはSCEC (Southern California Earthquake Center)の人
- 非専門家向けのガイドラインを作成中である。
 - develop
 - debug
 - test, evaluation
 - 特にリモート環境からのテスト方法

Life Sciences Workshop (1/4)

- 1月16日(木)午後, 17日(金)午前
 - 16日午後のみ出席
 - 40人程度の出席者(16日午後の段階)
- プログラム
 - Keynote(1件, 50分: P.Afflard, Novartis Pharma AG)
 - Vendor Presentation(3件, 15分ea.), Talk(9件, 25分ea.)
 - Roundtables(70分)
 - このうち, Vendor 2件, Talk 6件に出席

Life Sciences Workshop (2/4)

発表内容: 概要

- Bioinformatics分野のData Grid関連発表が多かった。
 - AVAKI Data Grid <http://www.avaki.com/>
 - 大阪大学: BioGrid <http://www.biogrid.jp/>
 - Indiana Univ.: IUBio/BioGrid <http://iubio.bio.indiana.edu/biogrid/>
 - Virginia Tech: PathPort (pathogen portal)
<http://www.vbi.vt.edu/~pathport/>
- シミュレーションは2件
 - The Scripps Institute
 - IBM Almaden: OptimalGrid
 - <http://www.almaden.ibm.com/software/ds/OptimalGrid/index.shtml>

35

Life Sciences Workshop (3/4)

発表内容: Data Grid

- AVAKI
 - 20種類以上の分野に利用されているが, Life Science分野が最も重要な顧客。
 - OGSi/OGSA, GT3を100%サポートしている。1月13日のGT3発表と同時に提携。
- BioGrid, IUBio, PathPort
 - Biological Dataの特徴
 - Distributed, Various Formats, Various Modes of Access ...
 - データベース構築, データ解析ツール, 可視化ツール
 - XMLによる共通データ構造, Parserの開発
 - 共通化に関して最終日に議論があったと思われる。GGFIにRG₉₆



Life Sciences Workshop (4/4)

発表内容: Simulation

- IBM Almaden
 - なぜか、ボクセル法による並列FEM関連Middlewareの話でLife Scienceとは直接関係なかった(人間の頭部のメッシュを見せてくれたが)。
 - 「Perfect Parallel Computing」:このようなAppl.では当然のことであるが。
- Scripps Institute
 - Protein FoldingのシミュレーションをMetacomputingにより実施
 - U.VirginiaのクラスターとASCI Blue Horizon (NPACI)を「Legion」でつなぐ

37



The Globus Toolkit®

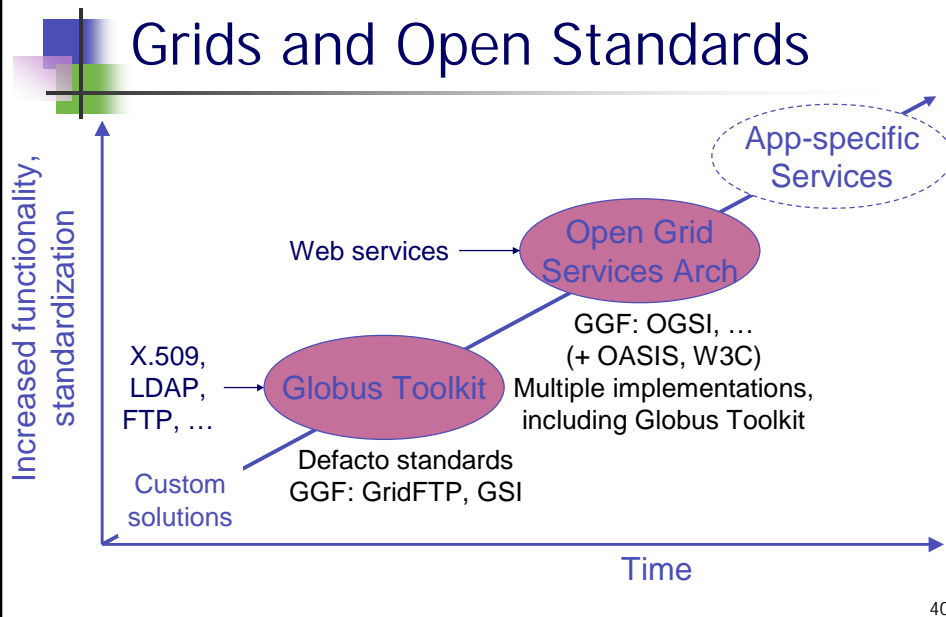
- A software system addressing key technical problems in the development of Grid-enabled tools, services, and applications
 - Offer a modular set of orthogonal services
 - Enable *incremental* development of Grid-enabled tools and applications
 - Implement standard Grid protocols and APIs
 - Available under liberal open source license
 - Large community of developers & users
 - Multiple commercial support providers

38

Globus Toolkit: A Story of Evolution

- Definition of Grid problem has been stable since original Globus Project proposal in 1995
 - Although we've got better at articulating it
- But our approach to its solution has evolved
 - from APIs and custom protocols ...
 - to standard protocols ...
 - to Grid services (OGSA)
- Driven by experience implementing GT, deploying GT, & building real applications

Grids and Open Standards



Grid, VO, Grid Architecture

Grid

"Grid" computing has emerged as an important new field, distinguished from conventional distributed computing by its focus on **large-scale resource sharing**, innovative applications, and, in some cases, high-performance orientation.

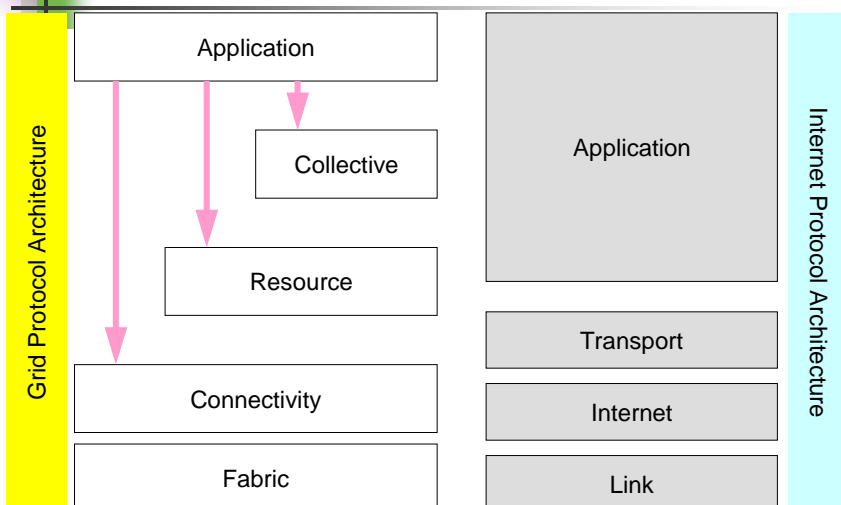
Virtual Organizations (VO)

...First, we review the "Grid problem," which we define as flexible, secure, coordinated resource sharing among dynamic collections of individuals, institutions, and resources—what we refer to as **virtual organizations**. In such settings, we encounter unique authentication, authorization, resource access, resource discovery, and other challenges. It is this class of problem that is addressed by Grid technologies.

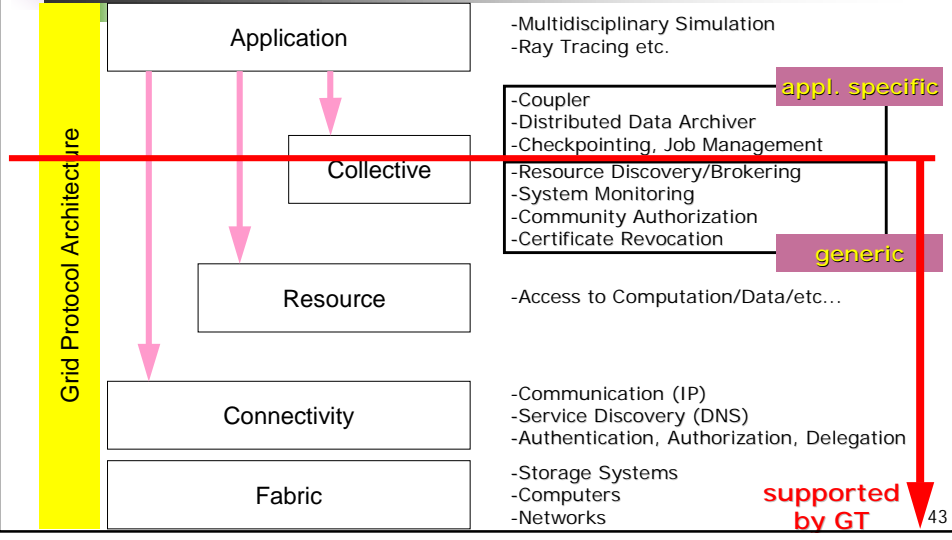
Grid Architecture

...Next, we present an extensible and open **Grid architecture**, in which protocols, services, application programming interfaces, and software development kits are categorized according to their roles in enabling resource sharing.

Grid Architectureの階層性



Grid Architectureの階層性



Relationships with Other Tech's

World Wide Web

The ubiquity of Web technologies (i.e., IETF and W3C standard protocols—TCP/IP, HTTP, SOAP, etc.—and languages, such as HTML and XML) makes them attractive as a platform for constructing VO systems and applications. **However, while these technologies do an excellent job of supporting the browser-client-to-web-server interactions that are the foundation of today's Web, they lack features required for the richer interaction models that occur in VOs.**

APS/SSP

Internet/P2P Computing

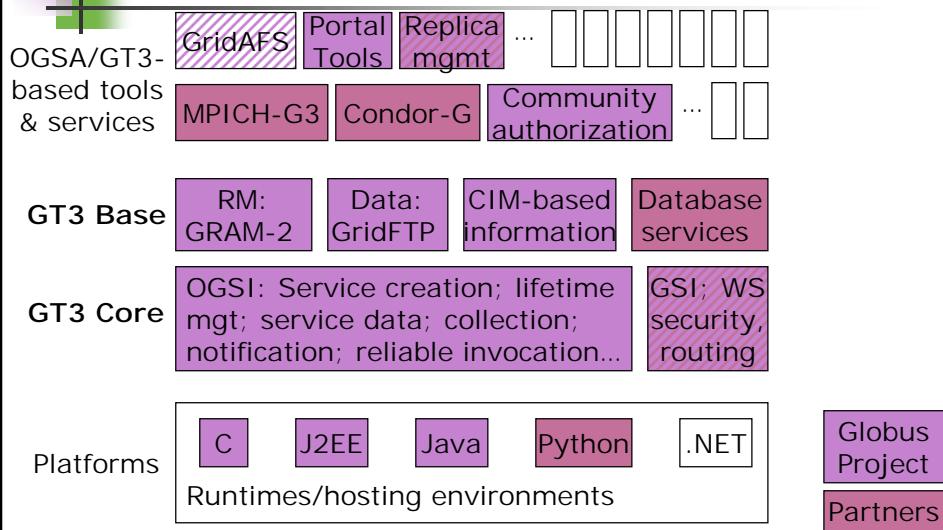
Enterprise Computing

Enterprise development technologies such as CORBA, Enterprise Java Beans, Java 2 Enterprise Edition, and DCOM are all systems designed to enable the construction of distributed applications. They provide standard resource interfaces, remote invocation mechanisms, and trading services for discovery and hence make it easy to share resources within a single organization. However, these mechanisms address none of the specific VO requirements listed above. Sharing arrangements are typically relatively static and restricted to occur within a single organization. The primary form of interaction is client-server, rather than the coordinated use of multiple resources.

Globus Toolkit v3 (GT3) Open Source OGSA Technology

- Implement core OGSI interfaces
 - Reference implementation of GS Spec
- Support primary GT2 interfaces
 - High degree of backward compatibility
- Target multiple platforms & hosting envs
 - J2EE, Java, C, .NET
- New services
 - SLA negotiation (GRAM-2), registry, replica location, community authorization, ...
- Growing number of external contributions
 - Database services (UK DAIS), Python hosting (LBNL), ...
- External adoption
 - E.g., IBM, Avaki, science projects

GT3: Open Source OGSA Technology



Architecture Framework by OGSA Integrated Functionality

