The Globus Toolkit® 3.0

The Globus Project™
Argonne National Laboratory
USC Information Sciences Institute

http://www.globus.org/
Overview

- Background
  - Grid Computing
  - Globus Toolkit 2
  - OGSA
  - OGSI

- GT3 Overview
  - Hosting Environments
  - Architecture
  - Standards

- GT3 Alpha Environment
  - Development Environment
  - Packaging and Deployment
  - Performance Tuning and Debugging

- GT3 Release Timeline
Grid Scenarios

- A biochemist exploits 10,000 computers to screen 100,000 compounds in an hour
- 1,000 physicists worldwide pool resources for peta-op analyses of petabytes of data
- An emergency response team couples real time data, weather model, population data
- Engineers at a multinational company collaborate on the design of a new product
- An application service provider offloads excess load to a compute cycle provider
Virtual Organizations

- Collaborating individuals and institutions
  - Shared goals
  - Enable sharing of resources
  - Non-locality of participants

- Dynamic in nature
  - VOs come and go
  - Resources/users join and leave VOs
  - Resources change status and fail

- Community-wide goals
The Grid Problem

Resource sharing & coordinated problem solving in dynamic, multi-institutional virtual organizations
Overview

• Background
  – Grid Computing
  – Globus Toolkit 2
  – OGSA
  – OGSI

• GT3 Overview
  – Hosting Environments
  – Architecture
  – Standards

• GT3 Alpha Environment
  – Development Environment
  – Packaging and Deployment
  – Performance Tuning and Debugging

• GT3 Release Timeline

• Migration from GT2
Some Important Definitions

- **Resource**
  - An entity to be shared

- **Network protocol**
  - A formal description of message formats and a set of rules for message exchanges

- **Application Programmer Interface (API)**
  - Defines an interface, but imposes no requirements on the implementation
The Globus Toolkit 2

- Tools enabling resource sharing within VOs
  - GSI (Grid Security Infrastructure)
    > Authentication based on Grid-wide credential
    > Single sign-on, delegation
    > Authorization
  - GRAM (Grid Resource Allocation and Management)
    > Tool for remote job and resource management
  - Data Services
    > GridFTP
    > Replica Management
  - MDS (Monitoring and Discovery Service)
    > Grid-wide information on the state of resources

- Protocols and APIs
The Globus Toolkit v2 in One Slide

- Grid protocols (GSI, GRAM, ...) enable resource sharing within virtual orgs; toolkit provides reference implementation (rectangle = Globus Toolkit services)

- Protocols (and APIs) enable other tools and services for membership, discovery, data mgmt, workflow, ...

Grid protocols (GSI, GRAM, ...) enable resource sharing within virtual orgs; toolkit provides reference implementation (rectangle = Globus Toolkit services). Protocols (and APIs) enable other tools and services for membership, discovery, data mgmt, workflow, ...
The Grid World: Current Status

- Dozens of major Grid projects in scientific & technical computing/research & education
- Considerable consensus on key concepts and technologies
  - Open source Globus Toolkit a de facto standard for major protocols & services
  - Far from complete or perfect, but out there, evolving rapidly, and large tool/user base
- Industrial interest emerging rapidly
- Opportunity: convergence of eScience and eBusiness requirements & technologies
Globus Toolkit 2: Evaluation (+)

- Good technical solutions for key problems, e.g.
  - Authentication and authorization
  - Resource discovery and monitoring
  - Reliable remote service invocation
  - High-performance remote data access

- This & good engineering is enabling progress
  - Good quality reference implementation, multi-language support, interfaces to many systems, large user base, industrial support
  - Growing community code base built on tools
Globus Toolkit 2: Evaluation (−)

- Protocol deficiencies, e.g.
  - Heterogeneous basis: HTTP, LDAP, FTP
  - No standard means of invocation, notification, error propagation, authorization, termination, ...

- Significant missing functionality, e.g.
  - Databases, sensors, instruments, workflow, ...

- Little work on total system properties, e.g.
  - Dependability, end-to-end QoS, ...
  - Reasoning about system properties
Overview

- **Background**
  - Grid Computing
  - Globus Toolkit 2
  - OGSA
  - OGSI

- **GT3 Overview**
  - Hosting Environments
  - Architecture
  - Standards

- **GT3 Alpha Environment**
  - Development Environment
  - Packaging and Deployment
  - Performance Tuning and Debugging

- **GT3 Release Timeline**
Key Publications

- The Grid: Blueprint for a New Computing Infrastructure

- Anatomy of the Grid
  Foster, Kesselman, Tuecke
  - Proposes an abstract architecture in which intergrid protocols enable interoperability among different grids

- Physiology of the Grid
  Foster, Kesselman, Nick, Tuecke
  - Introduces the concept of an Open Grid Services Architecture
    - Builds on concepts from the Grid and Web Services communities
    - Specifies the way in which grid mechanisms might implement a service-oriented architecture
      - Incorporating grid functionality into a web services framework
Grid Evolution:
Open Grid Services Architecture

- Refactor Globus protocol suite to enable a common base that provides interoperability and expose key capabilities
- Service orientation to virtualize resources and unify resources/services/information
- Embrace key Web services technologies for standard IDL, leverage commercial efforts
- Result: standard interfaces & behaviors for distributed system management: the Grid Service
- Introduced at GGF4 in February 2002
OGSA Structure

- A standard substrate: the Grid Service
  - OGSI = Open Grid Service Infrastructure
  - Standard interfaces and behaviors that address key distributed system issues
  - A refactoring and extension of the Globus Toolkit protocol suite

- ... supports standard service specifications
  - Resource mgt, dbms, workflow, security, ...
  - Target of current & planned GGF efforts

- ... and arbitrary application-specific services based on these & other definitions
Overview

- **Background**
  - Grid Computing
  - Globus Toolkit 2
  - OGSA
  - OGSI

- **GT3 Overview**
  - Hosting Environments
  - Architecture
  - Standards

- **GT3 Alpha Environment**
  - Development Environment
  - Packaging and Deployment
  - Performance Tuning and Debugging

- **GT3 Release Timeline**
OGSI Grid Service Specification

- Defines WSDL conventions and GSDL extensions
  - For describing and structuring services
  - Working with W3C WSDL working group to drive GSDL extensions into WSDL
- Defines fundamental interfaces (using WSDL) and behaviors that define a Grid Service
  - Standard interfaces and behaviors enable interoperability
- Winding its way through the GGF specification process: Version I of the spec to be released this month
Open Grid Services Infrastructure: Transient Service Instances

- “Web services” address discovery & invocation of persistent services
  - Interface to persistent state of entire enterprise

- In Grids, must also support transient service instances, created/destroyed dynamically
  - Interfaces to the states of distributed activities
  - e.g. workflow, video conf., dist. data analysis

- Significant implications for how services are managed, named, discovered, and used
  - In fact, much of OGSA (and Grid) is concerned with the management of service instances

- Transience critical for support of VOs
The Grid Service = Interfaces/Behaviors + Service Data

Required:
- Introspection (service data)
- Explicit destruction
- Soft-state lifetime

Binding properties:
- Authentication
- Reliable invocation
- Transactions
- QoS

Optional:
- Service creation
- Notification
- Registration
- Collections
+ application-specific interfaces

Hosting environment/runtime ("C", J2EE, .NET, ...)
Overview

- **Background**
  - Grid Computing
  - Globus Toolkit 2
  - OGSA
  - OGSI

- **GT3 Overview**
  - Hosting Environments
  - Architecture
  - Standards

- **GT3 Alpha Environment**
  - Development Environment
  - Packaging and Deployment
  - Performance Tuning and Debugging

- **GT3 Release Timeline**
A View from 20000 Metres

What is GT3?

- It is a software toolkit that the Globus Project is building in collaboration with others in the Grid community
- It is open source
- It can be downloaded from the web
- It is available today in alpha form
- It contains
  - Low level blocks with which to build services
  - Higher level services that can be used by applications in support of VOs
  - Development framework to make building services easier
Overview

- Background
  - Grid Computing
  - Globus Toolkit 2
  - OGSA
  - OGSI

- GT3 Overview
  - Hosting Environments
    - Architecture
    - Standards

- GT3 Alpha Environment
  - Development Environment
  - Packaging and Deployment
  - Performance Tuning and Debugging

- GT3 Release Timeline
Standalone/Embedded Hosting Environment

- Light-Weight SOAP/HTTP Server
- Can be embedded in J2SE JVMs
- Used for testing framework and client notification sinks
- Multithreaded, growing and shrinking thread pools
- Everything in one classloader
- HTTP 1.0
Servlet Engine Hosting Environment

- Standard Servlet dispatcher to be deployed in any Servlet API compliant Web Server
- OGSI container sits as one application in a Servlet Engine and has a separate classloader
- Tested with Tomcat 4.0.6 (HTTP 1.1)
- Leverage Web Server HTTP, Multithreading code
Enterprise Bean Hosting Environment

- GridService to EJB proxy generator
- Support for stateful/stateless Session Beans and Entity Beans
- No modification to existing Beans, or EJB Containers necessary, All Grid Service code is in Web Container
- Will run in WebSphere and JBoss
Virtual Hosting Environment

- Virtual Service Redirecting to remote hosting environments
  - A Grid Service Instance can act as a dispatcher onto other instances
  - Useful for sandboxing, load balancing
- Calls routed through single entry point
  - GSH Rewriting
  - GSR Rewriting
  - HandleResolver Redirection
- This infrastructure is critical for the dynamic assembly of VOs
Overview

- **Background**
  - Grid Computing
  - Globus Toolkit 2
  - OGSA
  - OGSI

- **GT3 Overview**
  - Hosting Environments
  - Architecture
  - Standards

- **GT3 Alpha Environment**
  - Development Environment
  - Packaging and Deployment
  - Performance Tuning and Debugging

- **GT3 Release Timeline**
GT3 Architecture Overview
GT3 Core

- **Implementation of the Grid Service Specification**
  - Designed to fully implement Version 1 of the OGSI specification

- **Common Infrastructure and APIs**
  - Service Data
  - Notification & Queries
  - State Management

- **Grid Service Container Framework**
  - Container enables portability across hosting environments

- **Development and Runtime Environment**
  - To make it easier to build new grid services
GT3 Security

- **Transport Layer Security/Secure Socket Layer (TLS/SSL)**
  - Based around a new protocol called “httpg” to indicate that it is a GSI-enabled http-based protocol

- **SOAP Layer Security**
  - Based on the following standards:
    - WS-Security
    - XML Encryption
    - XML Signature

- **GT3 uses X.509 identity certificates for authentication**

- **It also uses X.509 Proxy certificates to support delegation and single sign-on, updated to conform to latest IETF/GGF draft.**

- **Improved security model reduces the amount of privileged code needed by a service**
  - This makes it easier to set up GT3 in firewall environments
Managed Job Service

- GRAM Architecture rendered in OGSA
- Using Virtual Hosting Environment and Routing Framework

Master User

MMJFS

User 1
MJFS MJS

User 2
MJFS MJS

User 3
MJFS MJS

MMJFS: Master Managed Job Factory Service
MJFS: Master Job Factory Service
MJS: Managed Job Service
Index Service and Service Data Provider Framework

- Collect, Aggregate and Query Service Data
- Create Service Data Dynamically on Demand
- Monitor Data Feeds
  - subscribe
- Poll Data Feeds
  - findServiceData

IS: Index Service
P1..3: Information Providers

Diagram:

- findServiceData
- subscribe
- IS
- P1
- P2
- P3
Reliable File Transfer Service (RFT)

- Reliably performs a third party transfer between two GridFTP servers
- OGSI-compliant service exposing GridFTP control channel functionality
- Recoverable Grid Service
  - Automatically restarts interrupted transfers from the last checkpoint
- Progress and Restart Monitoring

Diagram:

```
RFT --> JDBC --> GridFTP Server 1
       \                                        /      \       
        \                                      / GridFTP Server 2
```
Other “Services” Bundled with GT3

- **GridFTP**
  - Used by RFT

- **Replica Location Service (RLS)**
  - Distributed registry service that records the locations of data copies and allows discovery of replicas
  - Designed and implemented in a collaboration between the Globus and DataGrid projects

The interfaces for these services are not yet OGSI-Compliant
Overview

- Background
  - Grid Computing
  - Globus Toolkit 2
  - OGSA
  - OGSI

- GT3 Overview
  - Hosting Environments
  - Architecture
  - Standards

- GT3 Alpha Environment
  - Development Environment
  - Packaging and Deployment
  - Performance Tuning and Debugging

- GT3 Release Timeline
XML Standards

- **SOAP**
  http://www.w3.org/TR/SOAP/

- **WSDL**
  http://www.w3.org/TR/wsdl

- **XML Schema**
  http://www.w3.org/TR/xmlschema-0/

- **XML Namespaces**
  http://www.w3.org/TR/REC-xml-names/

- **XML-Signature**
  http://www.w3.org/TR/xmldsig-core/

- **XML-Encryption**
  http://www.w3.org/TR/xmlenc-core/
Java Standards

http://www.jcp.org/

- JAX-RPC (JSR-101)
- Java Servlet API (JSR-154)
- J2EE (JSR-151), J2SE
- Enterprise Beans (JSR-153)
- JAXP (JSR-63)
- Java API for WSDL (JSR-110)
Overview

- **Background**
  - Grid Computing
  - Globus Toolkit 2
  - OGSA
  - OGSI

- **GT3 Overview**
  - Hosting Environments
  - Architecture
  - Standards

- **GT3 Alpha Environment**
  - Development Environment
  - Packaging and Deployment
  - Performance Tuning and Debugging

- **GT3 Release Timeline**
Development Environment

- **Apache Axis - SOAP Engine**

- **Jakarta Tomcat - Servlet Engine**

- **Jakarta Ant – Build System**

- **JUnit - Test Framework**
  [http://www.junit.org/index.htm](http://www.junit.org/index.htm)

- **Jakarta Commons & Log4J -Debugging**

- **Java Cog Kit – Security, GRAM Tooling**
Service Browser Framework

- Extensible GUI to browse and test services
- Showcase Samples
- Test Queries/Notifications
- Look at WSDL and Service Data in XML viewer
WSDL Tools

- **GSDL2Java**
  - GridLocator generation as extension to JAX-RPC

- **WSDL Binding Generator**
  - Only PortType needs to be declared. SOAP/HTTP binding is generated

- **WSDL Binding Decorator**
  - Used to decorate existing WSDL definitions to support standards OGSI port types like the mandated GridServicePortType
Ant Tools

- Deployment (EJB/Servlet/WebStart)
  - build-services.xml
- Admin (Start/Stop containers)
  - build-services.xml
- Packaging (Gar construction)
  - build-packages.xml
- Building (Stub Generation, WSDL Generation)
  - build-services.xml
- Samples
  - build-samples.xml
  - guide/build.xml
Overview

- Background
  - Grid Computing
  - Globus Toolkit 2
  - OGSA
  - OGSI

- GT3 Overview
  - Hosting Environments
  - Architecture
  - Standards

- GT3 Alpha Environment
  - Development Environment
  - Packaging and Deployment
  - Performance Tuning and Debugging

- GT3 Release Timeline
Grid Archive

- .gar (zip/jar format)
- Jars
- Stubs
- WSDL and XML Schemas
- Documentation
- Deployment Descriptor
- Configuration Files
Overview

- **Background**
  - Grid Computing
  - Globus Toolkit 2
  - OGSA
  - OGSI

- **GT3 Overview**
  - Hosting Environments
  - Architecture
  - Standards

- **GT3 Alpha Environment**
  - Development Environment
  - Packaging and
  - Performance Tuning and Debugging

- **GT3 Release Timeline**
Performance Tuning

- Lightweight
- Easy to use
- Not a full instrumentation system, but offers a look at your Grid service performance
  - On a service level
  - On a code block level
Debugging

- Log4J Logging through Apache Commons logging API
- Filters configured in log4j.properties
- Typically class level filtering: FATAL, ERROR, WARNING, INFO, DEBUG
- SOAP Tracing: Axis tcpmon, MS SOAP Toolkit or .NET Trace Utility
- Provide a message logging handler
Overview

- **Background**
  - Grid Computing
  - Globus Toolkit 2
  - OGSA
  - OGSI

- **GT3 Overview**
  - Hosting Environments
  - Architecture
  - Standards

- **GT3 Alpha Environment**
  - Development Environment
  - Packaging and Deployment
  - Performance Tuning and Debugging

**GT3 Release Timeline**
2003 GT 3.0 Release Timeline

(First prototype Grid Service implementation was in January 2002)
Conclusion

- GT3 provides online negotiation of access to services in a standard way
- GT3 enables the creation of dynamic, extensible systems
- GT3 embraces state-of-the-art protocols and leverages community standards
- GT3 is evolutionary, not revolutionary
  - Not changing higher-level functionality, changing protocols
Acknowledgments

- **Grid computing, Globus Project, and OGSA**
  - Ian Foster, Steve Tuecke @ ANL
  - Carl Kesselman @ USC/ISI
  - The talented team of scientists and engineers at ANL, USC/ISI, elsewhere (see http://www.globus.org)

- **Open Grid Services Architecture (OGSA)**
  - Karl Czajkowski @ USC/ISI, Jeff Nick, Steve Graham, Jeff Frey @ IBM, www.globus.org/ogsa

- **Support from DOE, NASA, NSF, IBM, Microsoft**
Further Information

- **My email:** Lisa Childers <childers@mcs.anl.gov>
- **GT3**
  - General information at [http://www.globus.org/gt3](http://www.globus.org/gt3)
  - Technical discussion list: ogsa-alpha@globus.org. To subscribe, sendemail to majordomo@globus.org with a message body containing: subscribe ogsa-alpha <email address>
- **Grids and the Globus Project**
  - General information: [http://www.globus.org](http://www.globus.org)
  - Global Grid Forum: [http://www.gridforum.org](http://www.gridforum.org)
  - Technical Papers: [http://www.globus.org/research/papers.html](http://www.globus.org/research/papers.html)
  - The Grid: Blueprint for a New Computing Infrastructure, I. Foster, C. Kesselman, Morgan-Kaufmann, 1999
- **Web Services**