Grid Computing

Globus Toolkit
Future Directions & Related Software

Paul A. Farrell
Fall 2006

The Globus Commitment to Open Source

• Globus was first established as an open source project in 1996
• The Globus Toolkit is open source to:
  – allow for inspection
    • for consideration in standardization processes
  – encourage adoption
    • in pursuit of ubiquity and interoperability
  – encourage contributions
    • harness the expertise of the community
• The Globus Toolkit is distributed under the (BSD-style) Apache License version 2

The Future: Structure

• NSF Community Driven Improvement of Globus Software (CDIGS) project
  – 5 years of funding for GT enhancement
  – Regular Globus roadmaps outlining plans
• GlobDev http://dev.globus.org
  – Apache-like community development site
  – Community governance of components
  – “Globus Toolkit” & other related software
  – Open for business early 2006
  – “Globus Alliance” = “GlobDev committers”

GlobDev

• The current set of Globus components will be organized into several “Globus Projects”
  – Projects release products
• Each project will have its own group of “Committers”
  – committers are responsible for governance on matters relating to their products
• The “Globus Management Committee” will
  – provide overall guidance and conflict resolution
  – approve the creation of new Globus Projects
The Future: Content

- We now have a solid and extremely powerful Web services base
- Next, we will build an expanded open source Grid infrastructure
  - Virtualization
  - New services for provisioning, data management, security, VO management
  - End-user tools for application development
  - Etc., etc.
- And of course responding to user requests for other short-term needs

Short-Term Priorities: Security

- Improve GSI error reporting & diagnostics
- Secure password, one-time password, Kerberos support for initial log on
- Trust roots, use of GridLogon
- Identity/attribute assertions in GT auth. callouts (e.g., Shib, PERMIS, VOMS, SAML)
- Extend CAS admin & policy support
- Security logging with management control for audit purposes

Short-Term Priorities: Data Management

- Space & bandwidth management in GridFTP
- Concurrency in globus-url-copy
- Priorities in RFT
- Data replication service
- Enhance policy support in data services
- Physical file name creation service
- Scalable & distributed metadata manager

Short-Term Priorities: Execution Management

- Implement GGF JSDL once finalized
- Advance reservation support
- Policy-driven restart of "persistent" jobs
- Improved information collection for jobs
- Improved management of job collections
- Credential refresh
- Development of workspace service
- Integration of virtual machines (Xen, VMware) and associated services
- Windows port of WS GRAM
Short-Term Priorities: Information Services

- Many more information sources, including gateways to other systems
- Automated configuration of monitoring
- Specialized monitoring displays
- Performance optimization of registry
- Archiver service
- Helper tools to streamline integration of new information sources

Short-Term Priorities: WS Core

- Streamlined container configuration
- Remote management interface
- Dynamic service deployment
- Service isolation: multiple service instances
- WS-Notification, subscription performance
- Full functionality in C WS Core
- Optimized WS-ServiceGroup support
- WS-SecureConversation support

The Globus Ecosystem – Related Tools

- Globus components address core issues relating to resource access, monitoring, discovery, security, data movement, etc.
  - GT4 being the latest version
- A larger Globus ecosystem of open source and proprietary components provide complementary components
  - A growing list of components
- These components can be combined to produce solutions to Grid problems
  - We’re building a list of such solutions

Many Tools Build on, or Can Contribute to, GT4-Based Grids

- Condor-G, DAGman
- MPICH-G2
- G RMS
- Nimrod-G
- Ninf-G
- Open Grid Computing Env.
- Commodity Grid Toolkit
- GriPhyN Virtual Data System
- Virtual Data Toolkit
- GridXpert Synergy
- Platform Globus Toolkit
- VOMS
- PERMIS
- GT4IDE
- Sun Grid Engine
- PBS scheduler
- LSF scheduler
- GridBus
- TeraGrid CTSS
- NEES
- IBM Grid Toolbox
- …
Example Solutions

- Portal-based User Reg. System (PURSE)
- VO Management Registration Service
- Service Monitoring Service
- TeraGrid TGCP Tool
- Lightweight Data Replicator
- GriPhyN Virtual Data System

Condor-G

- The Condor Project @ U Wisconsin Madison develops software for high-throughput computing on collections of distributed compute resources
- Condor-G is an interface to GRAM created by the Condor team that allows users to submit jobs to GRAM servers

GridShib

- Allows the use of Shibboleth-transported attributes for authorization in GT4 deployments
  - And, more generally, SAML support
- 2 year project started December 1, 2004
- Participants
  - Von Welch, UIUC/NCSA (PI)
  - Kate Keahey, UChicago/Argonne (PI)
  - Frank Siebenlist, Argonne
  - Tom Barton, UChicago
- Beta software released September 16, 2005
Handle System

- The Handle System from CNRI (http://www.handle.net) is a general-purpose global name service enabling secure name resolution over the internet
- The Handle System-GT Integration Project leverages the Handle System for identifier and resolution services through tight integration with GT4’s Web services protocols

MPICH-G2

- MPICH-G2, developed at Northern Illinois University and Argonne National Lab, is a grid-enabled implementation of the MPI v1.1 standard
- MPICH-G2 is implemented using the pre-WS GRAM component in GT4; integration with GT4 WS GRAM is expected in the near future

Nimrod/G

- Nimrod is a specialized parametric modeling system from Monash University
- Nimrod/G uses a simple declarative parametric modeling language to express parameter sweep experiments. Based on GT4 WS services, Nimrod/G enables the formulation, execution and monitoring of multiple individual parametric experiments

Ninf-G4

- Ninf-G4, from AIST, is a reference implementation of the GGF standard GridRPC API
- Ninf-G4 is provides higher-level programming APIs for the development and execution of parallel applications on the Grid
PERMIS

- PERMIS is an EU-funded Privilege Management service that implements Role-Based Access Control.
- Thanks to the work of the UK Grid Engineering Task Force, services running in a Java WS Core container can use PERMIS via GT4’s SAML authorization callouts.

SRB

- SRB is a package from SDSC providing a uniform interface for connecting to network-based heterogeneous data resources.
- GT4’s GridFTP includes an interface to SRB data sources, and vice versa.

Sun Grid Engine

- Sun Grid Engine is an open source distributed resource management system from Sun Microsystems.
- In a collaboration between the London e-Science Centre, Gridwise and MCNC, the Sun Grid Engine has been integrated with GT4.

Grid Tools & Solutions

- We list links to related projects on the “Related Software” of the Globus Toolkit web
- “Solutions” are documented on the Globus web