TeraGrid Information Services
Developer Introduction

John-Paul (JP) Navarro
TeraGrid Grid Infrastructure Group (GIG)
Area Co-Director for Software Integration and Information Services
University of Chicago, Argonne National Laboratory

April 2008
Abstract

Grids include interconnected hardware components, coordinated software and (grid) services, and institutions and groups that operate them.

To effectively use grids, users need access to information about the hardware, software, grid services, and the institutions and groups operating them.

The TeraGrid's Information Services vision is to:

1) define a coordinated way for TeraGrid participants to publish what they offer users,

2) define a way for the TeraGrid to aggregate and index the information from all TeraGrid participants, and

3) to publish this information to the public in a form that can easily be used by other software, users, and TeraGrid service providers themselves.

This presentation will detail aspects of TeraGrid's Information Service capability that are of interest to developers and TeraGrid service providers.
High-Level Components

TeraGrid Wide Information Services

- TeraGrid Wide Information
- Cache
- Apache 2.0
- Tomcat
- WebMDS
- WS MDS4
- WS/REST HTTP GET
- Clients

Service Provider Information Services

- Service Provider Information
- WS MDS4
- WS/SOAP
- Clients
- WS/REST HTTP GET
- Clients
High-Availability Design

TeraGrid Dynamic DNS

Server failover propagates globally in 15 minutes
Information Services Motivation

Applies Grid concepts to information publishing

① Publishing is the responsibility of the information owner
② Publishing is done using standard (content) schemas
③ Publishing thru standard interfaces regardless of content and where the data comes from
④ Publishing services should be available globally (subject to authentication/authorization)
⑤ Information owners publish to EVERYONE, not just the TeraGrid
⑥ Publishing is a grid service

Applies Grid concepts to aggregating information

① Publishing aggregated information is done exactly like original information publishing
② Aggregation uses standard information services interfaces to retrieve information
③ This is how a collaboration, such as the TeraGrid, aggregates participant information

Applies Grid concepts to querying information

① Querying can use standard interfaces regardless on content
Very new TeraGrid capability
- Introduced in 2007
- Infrastructure services were part of CTSS 4 rollout

Initial Content
- Queue contents and scheduling load information for user portal
- CTSS 4 capability kit information
- TGCDB institutions and resources (recently added)

Initial tooling and functionality
- Globus 4 MDS, Apache 2, Tomcat
- MDS publish, subscribe
- Custom (cron + script) aggregation and caching
- WebMDS publish
- REST (Apache 2) publish

Driven by new TeraGrid requirements Information Services will:
- Add new content
- Provide new functionality
- Use new tools
Publishing New Content

1) Requirements gathering
   • Identify content
   • Information ownership
   • Information (system) sources
   • Aggregation/refresh/caching
   • Access requirements

2) Content integration
   • Is (some) content in information services
   • How is the content indexed/mapped with other content

3) Development
   • Choose existing schema and/or develop new schema
   • Use existing or develop information providers
   • Use existing or develop aggregation/refresh/caching
   • Use existing or develop access views/applications
New Content Access Views

1) Requirements gathering
   • Identify content
   • Query protocols
   • Query aggregation scope
   • Query reliability
   • Query frequency/performance
   • For users and/or software

2) Development
   • Choose existing access protocols and views
   • Develop new access views
   • Develop new access protocols
New Content Aggregation/Storage

1) Requirements gathering
   • Persistence
   • Versioning
   • Etc

2) Development
   • Extend existing aggregation/storage methods
   • Develop new aggregation/storage methods
Service Provider vs TG Wide Services

Services Provider Information Services

Content:
- Locally owned and maintained information
- Originates anywhere the service provider wishes

Services:
- 1 general purpose MDS service
- 2 scheduling MDS services: authenticated and public (merging)

TeraGrid Wide Information Services

Content:
- Aggregate/index service provider information
- Additional central information (TGCDB, GIG operated services, ...)
- Cached (service providers services can be down)
- Authenticated registrations

Services:
- Several redundant servers (99.5% plus availability)
- Information caching (persistence)
- Several MDS4 services (WS/SOAP)
- WebMDS/Tomcat, Apache 2.0, ... services (WS/REST)
- Content published in: HTML, XHTML/XML, XML, Atom, RSS, ...
WS/* (Tomcat 5.0, Apache 2.0)

Benefits
- Very common web services platform
- Supports several web service interfaces (including simple)
- Supports multiple styles like REST, Web 2.0
- Can be highly scalable

Content
- Many formats: HTML, XHTML/XML, XML, RSS/Atom, ...

WebMDS (Globus 4.0.x/VDT 1.7.1)

Benefits
- Live MDS4 content access
- XPath support
- XSLT transforms

Content
- Many formats: HTML, XHTML/XML, XML, RSS/Atom

WS/SOAP (Globus 4.0.x/VDT 1.x.y MDS4)

Benefits
- Indexing, Trigger
- Registration, Publish, Subscribe
- Security/Authorization
- Robust WSRF interface

Content
- XML
CTSS 4 Capability Kits

For each capability kit on each resource
- Current support level, and target support level
  • Development, Testing, Production
- Support organization and contact
- Inca status URL
- Multiple version of a kit with different support levels
Future looking activities

Expanded content
- Local HPC Software
- Extended GridFTP service information
- (Meta)Scheduling support information

Core Extension
- Information Services Metadata (registration w/o aggregation)

Information Access
- tginfo, universal command line query tool
- WS/REST, Web 2.0 style information access
- Multiple formats: CSV TEXT, RSS/Atom, XML, ...
- GLUE 2.0

Community publishing
- Community supported capabilities
- Community information services registration
More Information

Find out more:
http://info.teragrid.org/ (links to content and documentation)

Request content:
mailto: help@teragrid.org or navarro@mcs.anl.gov

Discuss Information Services content, requirements, and design:
E-mail list tg-cat@teragrid.org

View current Information Services content

User Portal (scheduler load & queue contents):
https://portal.teragrid.org:443/gridsphere/gridsphere?cid=resources

User Documentation (CTSS 4 kits, software, services):
http://www.teragrid.org/userinfo/software/ctss.php

Information Service Main Page:
http://info.teragrid.org/