TeraGrid Information Services: Building on Globus MDS4

John-Paul “JP” Navarro
TeraGrid Grid Infrastructure Group “GIG” Area Co-Director for Software Integration
University of Chicago, Argonne National Laboratory

GlobusWorld
May 13, 2008
What is the TeraGrid?

1 NSF funded **facility** with 11 resource providers and several other partners

- Resource Provider (RP)
- Software Integration Partner
- Network Hub
What is the TeraGrid?

Operating a coordinated high-performance compute, network, storage, and visualization infrastructure.

Resource Provider (RP)
Software Integration Partner
Network Hub
What is the TeraGrid?

To enable Open Scientific Discover

Resource Provider (RP)

Software Integration Partner

Network Hub
One facility in what sense?

**Common People Interfaces, examples:**
- Request allocations & accounts “POPS”
- Identify available resources
- Learn how to use resources
- Find resource status
- Ask for help
- Community events

Proposal System “POPS”

User Portal

User Documentation & Knowledge Base

Events

Helpdesk

TeraGrid’08
June 9-13, 2008
(Tera)Grid in what sense?

**Coordinated/Standard (User) Software Interfaces**

- Coordinated software
- Coordinated development & runtime Unix environment
- Standardized (Grid) service interfaces

**Grid Services**

- Remote Login (GSI-OpenSSH)
- Data Movement (GridFTP, RFT)
- Data Management (SRB)
- Remote Execution (GRAM)
- Information/Discovery (MDS4, Tomcat, Apache2)

**Coordinated Software**

<table>
<thead>
<tr>
<th>Grid clients &amp; tools</th>
<th>Development tools &amp; languages</th>
<th>Communication, Data, and Math tools &amp; libraries</th>
</tr>
</thead>
</table>
Challenges

Integrating many **more types of** infrastructure **resources**, Including **more services providers**, Providing **more types of services**, With more resources, providers, and services diversity, Supporting specialization.

How can users keep it all straight?
Meeting the Challenge 1
Re-architect CTSS

Took Coordinated TeraGrid Software and Services “CTSS” v3:
- Monolithic coordinated capabilities
- In production since mid 2006

Re-architected into CTSS v4 “capability kits”:
- Small required core integration kit with minimum needed to integrate resources
- Plus ~10 optional user capability kits (at least 1 required)
Information Services Vision:

1) Create a **coordinated way for TeraGrid participants to publish** about the services they offer,

2) Create a **way for the TeraGrid to aggregate and index** the information from TeraGrid participants, and

3) to **publish this information to the public in a form that can easily be used** by user software, user interfaces, and TeraGrid service providers themselves to **discover capabilities and how to access them**.
Information Services Design Goals

Applies Grid concepts to **original** 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 **aggregated** information publishing

1. Aggregation uses standard information services interfaces to retrieve information
2. Publishing aggregated information is done exactly like original information publishing
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 of content
2. Querying can use standard interfaces for original or aggregated information
High-Level Architecture

TeraGrid Wide Information Services

TeraGrid Wide Repositories

Cache

Apache 2.0

WS/REST HTTP GET

Clients

WS/SOAP

Clients

WebMDS

WS MDS4

Service Provider Information Services

Local Info

Adapter

WS MDS4

WS/SOAP

Clients

WS SOAP
Information Services Tooling

**WS/SOAP (Globus 4.0.x MDS4)**
- **Benefits**
  - Indexing, Trigger
  - Registration, Publish, Subscribe
  - Security/Authorization
  - Robust WSRF interface
- **Content**
  - XML

**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)**
- **Benefits**
  - Live MDS4 content access
  - XPath support
  - XSLT transforms
- **Content**
  - Many formats: HTML, XHTML/XML, XML, RSS/Atom
Service Provider vs TG Wide Services

Content:
Locally owned and published information
Can come from existing local systems

Services:
1 general purpose MDS service
1 remote execution MDS services
Service Provider vs TG Wide Services

Content:
- Aggregate/index service provider information
- Plus central information (TeraGrid databases)
- Cached
- Authenticated registrations

Services:
- Several redundant servers (>99.5% availability)
- Each server:
  - Information caching programs
  - MDS4 index services (WS/SOAP)
  - WebMDS/Tomcat, Apache 2.0, ... services (WS/REST)

Services publish in:
- HTML
- XML
- CSV
  - (Atom, JSON, RSS being prototyped)
High-Availability Design

TeraGrid Wide Information Services

Service Provider Information Services

(commercial XEN server)

(TeraGrid physical server)

info.teragrid.org

info.dyn.teragrid.org

TeraGrid Dynamic DNS

Server failover propagates globally in 15 minutes

Dynamic paths

Static paths
Goal
Information persistence when service providers are down and when central information services go down

MDS configuration:
Service provider register their existence
Central MDS do not aggregate automatically
Central MDS publishes from local file-system cache

Caching programs:
External and asynchronous to MDS
Query registered
Service provider MDSs, and
Other configured MDSs
Store query results in file-system
Replaces previously cached data only with new data
Not if downstream is down
Or if downstream returns an error
Or if downstream doesn’t return valid data
Primary MDS4 features used

Registration:
- Parallel upstream registration to multiple central servers
- Authenticated/authorized upstream registration

Aggregation:
- Some information is automatically aggregated
  (WS services registration and default GLUE)

Information providers:
- Useful RP custom information providers
  - Scheduling load and queue contents (User Portal)
  - CTSS 4 Capability kit registration
  - TeraGrid ID and description cross-reference (TGCDB)

Secure MDS:
- Authenticated/authorized queries (non-anonymous)
- Multiple index services in same container
  - DefaultIndexService AND SecureIndexService

WebMDS
- XSLT, Xpath
Information Services Users

User Documentation
http://www.teragrid.org/

User Portal
http://portal.teragrid.org/

Inca
Testing Harness

Gateways

Peer Grids

User Applications

info.teragrid.org
## CTSS Capability Kit Availability

View CTSS Kits Available on TeraGrid

### CTSS Resources and Available Kits

(Read the Kit Definitions)

<table>
<thead>
<tr>
<th>Site/Resource</th>
<th>TeraGrid Core Integration Kit</th>
<th>Remote Login Capability Kit</th>
<th>TeraGrid Remote Compute Kit</th>
<th>Application Development and Runtime Support Capability Kit</th>
<th>TeraGrid Data Movement Kit</th>
<th>Parallel Application Capability Kit</th>
<th>Science Workflow Support Kit</th>
<th>TeraGrid Data Management Capability Kit</th>
<th>Data Visualization Support Kit</th>
</tr>
</thead>
<tbody>
<tr>
<td>NCAR: Frost</td>
<td>currently available</td>
<td>currently available</td>
<td>currently available</td>
<td>currently available</td>
<td>currently available</td>
<td>currently available</td>
<td>Not Installed</td>
<td>Not Installed</td>
<td>Not Installed</td>
</tr>
<tr>
<td>PSC: Big Ben</td>
<td>currently available</td>
<td>currently available</td>
<td>currently available</td>
<td>currently available</td>
<td>currently available</td>
<td>Not installed</td>
<td>currently available</td>
<td>Not Installed</td>
<td>Not Installed</td>
</tr>
<tr>
<td>PSC: Rachel</td>
<td>currently available</td>
<td>currently available</td>
<td>currently available</td>
<td>currently available</td>
<td>currently available</td>
<td>Not installed</td>
<td>Not installed</td>
<td>Not Installed</td>
<td>Not Installed</td>
</tr>
<tr>
<td>Purdue: Condor</td>
<td>currently available</td>
<td>currently available</td>
<td>currently available</td>
<td>currently available</td>
<td>Not Installed</td>
<td>currently available</td>
<td>current available</td>
<td>Not Installed</td>
<td>Not Installed</td>
</tr>
<tr>
<td>Purdue: Lear</td>
<td>currently available</td>
<td>currently available</td>
<td>currently available</td>
<td>currently available</td>
<td>currently available</td>
<td>Not Installed</td>
<td>current available</td>
<td>current available</td>
<td>Not Installed</td>
</tr>
<tr>
<td>TACC: Lonestar</td>
<td>currently available</td>
<td>currently available</td>
<td>currently available</td>
<td>currently available</td>
<td>currently available</td>
<td>Not Installed</td>
<td>current available</td>
<td>current available</td>
<td>Not Installed</td>
</tr>
<tr>
<td>TACC: Maverick</td>
<td>currently available</td>
<td>currently available</td>
<td>currently available</td>
<td>currently available</td>
<td>currently available</td>
<td>Not Installed</td>
<td>current available</td>
<td>current available</td>
<td>current available</td>
</tr>
<tr>
<td>NCSA: Cobalt</td>
<td>currently available</td>
<td>currently available</td>
<td>currently available</td>
<td>currently available</td>
<td>currently available</td>
<td>Not Installed</td>
<td>current available</td>
<td>current available</td>
<td>Not Installed</td>
</tr>
<tr>
<td>NCSA: Abe</td>
<td>currently available</td>
<td>currently available</td>
<td>currently available</td>
<td>currently available</td>
<td>currently available</td>
<td>Not Installed</td>
<td>current available</td>
<td>current available</td>
<td>Not Installed</td>
</tr>
<tr>
<td>NCSA: iA &amp; iA Cluster</td>
<td>currently available</td>
<td>currently available</td>
<td>currently available</td>
<td>currently available</td>
<td>currently available</td>
<td>Not Installed</td>
<td>current available</td>
<td>current available</td>
<td>current available</td>
</tr>
</tbody>
</table>
Where are the GridFTP services?

<table>
<thead>
<tr>
<th>Service selection: globus-gridftp-server</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>NCAR Blue Gene/L (Frost)</strong></td>
</tr>
<tr>
<td>Type: gridftp</td>
</tr>
<tr>
<td>Version: 4.0.3</td>
</tr>
<tr>
<td>Endpoint: gsiftp://gridftp.frost.ncar.teragrid.org:2811/</td>
</tr>
<tr>
<td>Name: globus-gridftp-server</td>
</tr>
<tr>
<td><strong>PSC XT3 (BigBen)</strong></td>
</tr>
<tr>
<td>Type: gridftp</td>
</tr>
<tr>
<td>Version: 4.0.3</td>
</tr>
<tr>
<td>Endpoint: gsiftp://gridftp.bigben.psc.teragrid.org:2811/</td>
</tr>
<tr>
<td>Name: globus-gridftp-server</td>
</tr>
<tr>
<td><strong>PSC Alpha EV7 (Rachel)</strong></td>
</tr>
<tr>
<td>Type: gridftp</td>
</tr>
<tr>
<td>Version: 4.0.3</td>
</tr>
<tr>
<td>Endpoint: gsiftp://gridftp.psc.teragrid.org:2811/</td>
</tr>
<tr>
<td>Name: globus-gridftp-server</td>
</tr>
<tr>
<td><strong>Purdue Condor Pool</strong></td>
</tr>
<tr>
<td>Type: gridftp</td>
</tr>
<tr>
<td>Version: 4.0.3</td>
</tr>
<tr>
<td>Name: globus-gridftp-server</td>
</tr>
<tr>
<td><strong>Purdue EM64T Cluster (Lear)</strong></td>
</tr>
<tr>
<td>Type: gridftp</td>
</tr>
<tr>
<td>Version: 4.0.3</td>
</tr>
<tr>
<td>Name: globus-gridftp-server</td>
</tr>
</tbody>
</table>
Queue Contents in User Portal

TeraGrid User Portal

Systems Monitor

Job summary for login-abe.ncsa.teragrid.org:
47 Running Jobs
68 Queued Jobs
50 Other Jobs

47 Running Jobs on login-abe.ncsa.teragrid.org

<table>
<thead>
<tr>
<th>Status</th>
<th>Job Id</th>
<th>Name</th>
<th>Owner</th>
<th>Queue</th>
<th>Submission Time</th>
<th>Processors</th>
</tr>
</thead>
<tbody>
<tr>
<td>Running</td>
<td>35721.abem5.ncsa.uiu</td>
<td>d_1-40</td>
<td>petefred</td>
<td>normal</td>
<td></td>
<td>8</td>
</tr>
<tr>
<td>Running</td>
<td>35730.abem5.ncsa.uiu</td>
<td>l_1-39</td>
<td>petefred</td>
<td>normal</td>
<td></td>
<td>8</td>
</tr>
<tr>
<td>Running</td>
<td>36071.abem5.ncsa.uiu</td>
<td>rturb</td>
<td>pakshing</td>
<td>normal</td>
<td></td>
<td>16</td>
</tr>
<tr>
<td>Running</td>
<td>36518.abem5.ncsa.uiu</td>
<td>A2Q_8-8</td>
<td>seabra</td>
<td>normal</td>
<td></td>
<td>4</td>
</tr>
<tr>
<td>Running</td>
<td>36521.abem5.ncsa.uiu</td>
<td>A2Q_8-12</td>
<td>seabra</td>
<td>normal</td>
<td></td>
<td>4</td>
</tr>
<tr>
<td>Running</td>
<td>36523.abem5.ncsa.uiu</td>
<td>A2Q_12-12</td>
<td>seabra</td>
<td>normal</td>
<td></td>
<td>4</td>
</tr>
<tr>
<td>Running</td>
<td>36563.abem5.ncsa.uiu</td>
<td>AirNoFSTur</td>
<td>azun</td>
<td>normal</td>
<td></td>
<td>30</td>
</tr>
<tr>
<td>Running</td>
<td>36584.abem5.ncsa.uiu</td>
<td>DI14</td>
<td>jhsin</td>
<td>normal</td>
<td></td>
<td>128</td>
</tr>
<tr>
<td>Running</td>
<td>36627.abem5.ncsa.uiu</td>
<td>rturb</td>
<td>pakshing</td>
<td>normal</td>
<td></td>
<td>16</td>
</tr>
<tr>
<td>Running</td>
<td>36647.abem5.ncsa.uiu</td>
<td>P3</td>
<td>amyshih</td>
<td>normal</td>
<td></td>
<td>24</td>
</tr>
<tr>
<td>Running</td>
<td>36690.abem5.ncsa.uiu</td>
<td>rturb</td>
<td>pakshing</td>
<td>normal</td>
<td></td>
<td>16</td>
</tr>
<tr>
<td>Running</td>
<td>36718.abem5.ncsa.uiu</td>
<td>cr-323-02</td>
<td>kjjin</td>
<td>normal</td>
<td></td>
<td>4</td>
</tr>
<tr>
<td>Running</td>
<td>36828.abem5.ncsa.uiu</td>
<td>m0r8_RC0s</td>
<td>moo</td>
<td>normal</td>
<td></td>
<td>16</td>
</tr>
<tr>
<td>Running</td>
<td>36833.abem5.ncsa.uiu</td>
<td>Estr_Prod</td>
<td>seabra</td>
<td>normal</td>
<td></td>
<td>16</td>
</tr>
<tr>
<td>Running</td>
<td>36842.abem5.ncsa.uiu</td>
<td>Script.abe</td>
<td>dcollins</td>
<td>normal</td>
<td></td>
<td>16</td>
</tr>
<tr>
<td>Running</td>
<td>36900.abem5.ncsa.uiu</td>
<td>L10_N64_41</td>
<td>gbryan</td>
<td>normal</td>
<td></td>
<td>16</td>
</tr>
<tr>
<td>Running</td>
<td>36903.abem5.ncsa.uiu</td>
<td>frthall</td>
<td>zetienne</td>
<td>normal</td>
<td></td>
<td>150</td>
</tr>
</tbody>
</table>
## Inca Capability Kit Testing

### Remote Compute 3.0.0

<table>
<thead>
<tr>
<th>Test</th>
<th>ANL Grid</th>
<th>ANL Ia64</th>
<th>ANL Viz</th>
<th>Indiana BigRed</th>
<th>NCSA Coral</th>
<th>NCSA Grid-Abe</th>
<th>NCSA Grid-HG</th>
<th>NCSA Tungsten</th>
<th>OML Login</th>
<th>PSC BigBen</th>
<th>PSC Rachel</th>
<th>Purdue Condor</th>
<th>Purdue Steele</th>
<th>TACC Lonestar</th>
<th>TACC Ranger</th>
<th>TACC Viz</th>
</tr>
</thead>
<tbody>
<tr>
<td>ctss-remote-compute-registration</td>
<td>n/a</td>
<td>n/a</td>
<td>PASS</td>
<td>PASS</td>
<td>PASS</td>
<td>PASS</td>
<td>PASS</td>
<td>PASS</td>
<td>PASS</td>
<td>PASS</td>
<td>ERROR</td>
<td>ERROR</td>
<td>ERROR</td>
<td>n/a</td>
<td>n/a</td>
<td>t/a</td>
</tr>
<tr>
<td>globus-wsrf</td>
<td>n/a</td>
<td>n/a</td>
<td>PASS</td>
<td>PASS</td>
<td>PASS</td>
<td>PASS</td>
<td>PASS</td>
<td>PASS</td>
<td>ERROR</td>
<td>ERROR</td>
<td>PASS</td>
<td>PASS</td>
<td>n/a</td>
<td>n/a</td>
<td>PASS</td>
<td>n/a</td>
</tr>
<tr>
<td>wsrfr-query-mds</td>
<td>n/a</td>
<td>PASS</td>
<td>PASS</td>
<td>PASS</td>
<td>PASS</td>
<td>PASS</td>
<td>PASS</td>
<td>PASS</td>
<td>ERROR</td>
<td>ERROR</td>
<td>PASS</td>
<td>PASS</td>
<td>n/a</td>
<td>n/a</td>
<td>PASS</td>
<td>n/a</td>
</tr>
<tr>
<td>wsrfr-query-mds-auth-jobsRIP</td>
<td>n/a</td>
<td>PASS</td>
<td>PASS</td>
<td>PASS</td>
<td>PASS</td>
<td>PASS</td>
<td>PASS</td>
<td>PASS</td>
<td>ERROR</td>
<td>ERROR</td>
<td>PASS</td>
<td>PASS</td>
<td>n/a</td>
<td>n/a</td>
<td>PASS</td>
<td>n/a</td>
</tr>
<tr>
<td>wsrfr-query-mds-auth-loadRIP</td>
<td>n/a</td>
<td>PASS</td>
<td>PASS</td>
<td>PASS</td>
<td>PASS</td>
<td>PASS</td>
<td>PASS</td>
<td>PASS</td>
<td>ERROR</td>
<td>ERROR</td>
<td>PASS</td>
<td>PASS</td>
<td>n/a</td>
<td>n/a</td>
<td>PASS</td>
<td>n/a</td>
</tr>
<tr>
<td>prews-gram-netstat</td>
<td>n/a</td>
<td>n/a</td>
<td>PASS</td>
<td>PASS</td>
<td>PASS</td>
<td>PASS</td>
<td>PASS</td>
<td>PASS</td>
<td>PASS</td>
<td>PASS</td>
<td>PASS</td>
<td>PASS</td>
<td>n/a</td>
<td>n/a</td>
<td>PASS</td>
<td>n/a</td>
</tr>
<tr>
<td>prews-gram-gatekeeper-pong</td>
<td>n/a</td>
<td>n/a</td>
<td>PASS</td>
<td>PASS</td>
<td>PASS</td>
<td>PASS</td>
<td>PASS</td>
<td>PASS</td>
<td>PASS</td>
<td>PASS</td>
<td>PASS</td>
<td>PASS</td>
<td>n/a</td>
<td>n/a</td>
<td>PASS</td>
<td>n/a</td>
</tr>
<tr>
<td>prews-gram-fork</td>
<td>n/a</td>
<td>n/a</td>
<td>PASS</td>
<td>PASS</td>
<td>PASS</td>
<td>PASS</td>
<td>PASS</td>
<td>PASS</td>
<td>PASS</td>
<td>PASS</td>
<td>PASS</td>
<td>PASS</td>
<td>n/a</td>
<td>n/a</td>
<td>PASS</td>
<td>n/a</td>
</tr>
<tr>
<td>prews-gram-batch</td>
<td>n/a</td>
<td>n/a</td>
<td>PASS</td>
<td>PASS</td>
<td>PASS</td>
<td>PASS</td>
<td>PASS</td>
<td>PASS</td>
<td>PASS</td>
<td>PASS</td>
<td>PASS</td>
<td>PASS</td>
<td>n/a</td>
<td>n/a</td>
<td>PASS</td>
<td>n/a</td>
</tr>
<tr>
<td>tg-audit-config</td>
<td>n/a</td>
<td>n/a</td>
<td>PASS</td>
<td>PASS</td>
<td>PASS</td>
<td>PASS</td>
<td>PASS</td>
<td>PASS</td>
<td>PASS</td>
<td>PASS</td>
<td>PASS</td>
<td>PASS</td>
<td>n/a</td>
<td>n/a</td>
<td>PASS</td>
<td>n/a</td>
</tr>
</tbody>
</table>
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
CTSS 4 Capability Kit Software

For each kit software component on each resource
- Name, version, how to access it
- Multiple versions of a single component
CTSS 4 Capability Kit Services

For each kit service on each resource

- Name, type, version, and Endpoint (contact location)
- GSI OpenSSH, GridFTP, SRB servers, PreWS & WS GRAM, MDS4
- Multiple services of the same type
What’s in Development?

Expanded content
- TeraGrid Gateways list
- Extended GridFTP service information (striping, bandwidth, etc)
- Local HPC Software
- (Meta)Scheduling support information
- Information Services Metadata

Enhanced caching, aggregation, and storage

Improved information access
- tginfo, universal command line query tool
- WS/REST, Web 2.0 style information access
- Multiple formats: CSV TEXT, XML, JSON, RSS/Atom, ...
- GLUE 2.0
- Usage metrics
The Team

Information Services design, development, and support team
- Eric Blau
- Jason Brechin
- Lee Liming
- JP Navarro
- Laura Pearlman

Information Services downstream developers and consumers
- End-to-end data transfer optimization team @ PSC (Derek Simmel, others)
- INCA (Kate Ericson)
- Operations (Jason Brechin, Tony Rimovsky)
- User Documentation (Mike Dwyer, Diana Diehl)
- User Portal (Maytal Dahan)
- Other lurkers...
More Information

Information Services main page:
http://info.teragrid.org/
(links to content and documentation)

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

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

Inca Monitoring Framework
http://www.teragrid.org/userinfo/software/ctss.php

E-mail:
<navarro@mcs.anl.gov>