The Open Science Grid / TeraGrid Gateway

Steven C. Timm
Neha Sharma, Daniel R. Yocum, Keith Chadwick

Grid Facilities Department
Fermi National Accelerator Laboratory
Batavia, IL, USA

Supported by the Department of Energy Office of Science SciDAC-2 program from the High Energy Physics, Nuclear Physics and Advanced Software and Computing Research programs, and the National Science Foundation Math and Physical Sciences, Office of CyberInfrastructure and Office of International Science and Engineering Directorates.

Outline

- Acknowledgements
- Introduction
- Comparison—Open Science Grid and TeraGrid
- Motivation—why a job-forwarding gateway
- Security—Authentication and authorization
- Future plans

Acknowledgements

- Coding, developing, and testing of the gateway, as well as liaison with NCSA by Neha Sharma of Fermilab.
- Enabling NCSA site for use as a demo by John Townes, Doru Marcusiu, Jim Basney of NCSA.
- Development of demo TeraGrid Condor collector, Warren Smith, TACC.
- Development of TeraGrid production information systems, John-Paul Navarro, Argonne Natl. Lab / Univ. of Chicago.

Introduction

- The Open Science Grid / TeraGrid gateway is currently a joint demonstration project between OSG and the TeraGrid.
- Goal of the demonstration project is to successfully forward jobs from OSG to one site on the TeraGrid (NCSA, National Center for Supercomputing Applications) at Univ. of Illinois, Urbana-Champaign.
- Working on getting one or two user communities to have a successful deployment before we involve more user communities.
- Disclaimer: This project is partially funded by OSG but information expressed in this talk should not be considered official positions of Open Science Grid or TeraGrid.

Open Science Grid / TeraGrid comparison

- Open Science Grid
- Focuses on high throughput computing
- Funded by US Dept. of Energy and National Science Foundation
- Has 92 registered resources at universities and national laboratories
- Relies on Virtual Organizations for membership

- TeraGrid
- Focuses on high performance parallel computing
- Funded by National Science Foundation
- Consists of universities and major supercomputing centers
- Membership gained by a Principal Investigator receiving an allocation of computing time.

Why a gateway

- Both Open Science Grid and TeraGrid use the same basic Globus software stack, but they diverge slightly in the layout of sites.
- There are a growing number of Open Science Grid users who (a) need as many computing cycles as they can get and (b) have collaborators who also want to work on the TeraGrid.
- Want to develop something that looks as much like an Open Science Grid site as possible, which can eventually figure out which TeraGrid sites the OSG user can run on and route the job automatically.

Technology

- Use jobmanager-cemon (as described in S. Timm's talk yesterday).
- Emulate an OSG site that is running condor, receive the jobs, and forward them to one or more TeraGrid sites.
- In the demo we are hard-wired to grid-hg.ncsa.teragrid.org.

Authentication and Authorization

- Initially all the people participating in the test are entered as individual users of the TeraGrid just like any other user, so they can submit directly to the TeraGrid site as well as coming through our gateway.
- Make use of GT4 Web Services, particularly the credential delegation service.
- User client delegates their globus proxy to the gateway and we in turn delegate it to the final resource. No limited proxies needed as in the FermiGrid gateway.

Resource selection

- TeraGrid has a demo resource selection system which is also based on condor classads.
- Need different matching requirement than number of free slots due to complex allocation strategy. Just because slots are free doesn't mean an OSG user can get them. Use estimated response time.
- TeraGrid is developing a GLUE schema-based information system as well, which can use Globus MDS4 transport but can also be accessed via other technologies such as a web UI.
- http://inca.teragrid.org/inca/html/ctssv4.html

Longer-term plans

- All use of the TeraGrid is based on scientific allocation.
- Right now OSG has a test allocation.
- If gateway is found to be useful, collaborators may apply for larger TeraGrid allocations.
- Two possibilities on how the OSG-TG gateway will work
  - Conventional Teragrid PI: OSG-TG Gateway will use TG information system to find sites where the particular project and principal investigator is allowed to run, and direct the job to the best one via matchmaking
  - Community Accounts: Each VO in OSG would become a “community account” similar to TeraGrid’s Science Gateway community account. There would be one user name for the whole VO and credentials would be stashed on the OSG-TG gateway.

Summary

- The OSG-TG gateway is operational as a demonstration project and has been for almost a year
- We are engaging user communities that are hungry for cycles and have modest I/O needs to attempt to get useful work done on it.
- Eventual value of the gateway will be measured through
  - Decreased startup time for OSG users utilizing TeraGrid cycles
  - Increased cooperation between OSG and TG on information systems
  - OSG-like workload of loosely coupled jobs giving useful backfill to HPC TeraGrid sites
  - Potential future high-throughput computing efforts at TeraGrid sites