



Open Science Grid and Gluex

Richard Jones

Review

- OSG is a community of scientific collaborations who deploy a common software infrastructure for managing global access to shared storage and computing resources.
- Member experiments include: cms, atlas, ligo, star, alice, cigi, dosar, d0, hypercp, i2u2, icecube, ktev, miniboone, minerva, minos, mipp, numi, nova, and a number of scientific collectives (compbiogrid, nebiogrid, astro, glow, grow, gpn, about a dozen more).
- Gluex joined as a new VO in Sept. 2009.
- First experience with Gluex production work on OSG over the last couple of months.

Present status

- Gluex VO is now in active operation since 12/1/2009.
- So far UConn is only contributor of Gluex VO resources.
 - 384 available job slots
 - 5 TB of available grid storage for Gluex work
 - 200,000 cpu-hr/mo., with ~50% used by local users
 - opportunistic scheduling, local users get first priority
- UConn-OSG is running jobs from sbgrid, ligo, others.

Present status

- *Not a free lunch.* There is no strict accounting of jobs performed by a VO's members vs the resources contributed by a VO, but each site selects the VO's it will accept jobs from – *fair play is expected.*
- The advantage is spreading out your load in time, while gaining high instantaneous turn-around.
- Example: **Blake Leverington et al, pi0 simulation in 12/2009.**
 - needed: one week of data in this channel
 - generate + filter + simulate + strip + analyze → root files
 - altogether about 10 khr of cpu time + 500GB storage
 - all jobs were taken by UConn-OSG
 - broken into ~1000 jobs of 10 hr each
 - completed in ~48 hr

Blake's blog

- Blake wrote a wiki page detailing everything he learned during the process of getting his jobs set up to run on the grid.

[HOWTO get your jobs to run on the Grid](#)

Where are we now?

- Installing the OSG client package in a jlab cue area
 - recently needed some data from Jlab MSS
 - cc.jlab.org documentation says to use srm
 - looked for srm commands – not found (more on this later)
 - installed OSG-Client in ~jonesrt (about 850 MB of quota!)
 - SP volunteered some space in /app for it
 - RJ will install and maintain it
 - accessible to anyone with a cue account
- Provides a streamlined way for someone to get started using the grid without having to install a lot of software

What's next?

- This resource is in place now, available to Gluex students, postdocs, whoever is looking for a place available cycles and fast turn-around for jobs with data-intensive or compute-hungry requirements.
 - run physics simulation and analysis
 - run engineering models with fine-mesh grids
 - publish large data sets with high bandwidth access

What's coming soon?

- Competition for compute/data resources is modest right now, but the day is not so far away when that will change...
- We want to be in a place then to marshal shared resources as a collaboration to carry out a variety of parallel simulation + reconstruction + analysis programs.
- Jlab will provide a set of top-tier services (data warehouse, major mover of raw data for calibration and “cooking”) but we do not want our student’s analysis jobs to sit in the queue behind “mass production” jobs for Gluex, CLAS12 and the other halls.
- We need a plan for moving and storing working data sets off-site, and for sharing global-access compute and storage resources.

Moving data from Jlab off-site?

- This seems not to be well-thought through at present. In general, moving data between on-site and off-site is a complicated dance whose steps change from time to time.
- Users have scripts.
- My scripts that worked last time don't work this time.
- My latest script [runs on ifarml1]
 - check for space on scratch to hold data I want to stage.
 - Gluex space is near the quota limit, write automatic email to Mark Ito.
 - sleep for 6 hours, and repeat.
 - got space, stage data from mss quick before the space disappears.
 - send email to self to run script on jlabs1 pushing data to UConn from scratch
 - run script on jlabs1
 - forget to delete files from scratch (you may need it again soon)

Outlook and summary

- OSG is a very active and growing community, with an extensive network of support (weekly VO conference calls, online “chat” help desk, ticket system, meetings).
- Fermilab is hosting “All Hands” meeting of OSG users in March, RJ invited to chair a session.
- I am ready to help students, others to get started using OSG resources for Gluex.
- Our offline computing requirements that we communicate to the Jlab Computing Group needs to reflect our need to efficiently exchange data with outside institutions.