



Deploying new applications on the OSG

Greg Thain
Computer Sciences Department
University of Wisconsin-Madison
gthain@cs.wisc.edu

Lessons Learned

Molrep, a structural bio app, ran on OSG

Took a couple of months

Joint effort with Scientists & OSG Engage

Several surprises along the way

These are some of our lessons

Overview

1: Learn how to run **at all** on the grid:

- Discover file dependencies
 - You might be surprised...
- Discover OS/CPU dependencies
 - You will be surprised...

2: Learn how to run efficiently on the grid

- Premature optimization: root of all evil

Running (at all) on a grid

- > Local cluster good place to start
 - **strace** is helpful tool to find file dependencies
 - Shared library dependencies often happen
 - Static linking is usually required
- > Goal: get a working submit file

Talk to Engage

- After running locally, then one Grid site
 - Which one? How do I get a cert? Etc. etc.
- This is where the Engage team helps
 - We will save you days!
 - But we need to learn about your app.

Running on one(!) OSG site

- > Ignore performance at first
- > Bring everything with you:
 - You might discover new dependencies!
- > Lack of direct logins can be trying!
 - Local clusters are useful for debugging

I/O, I/O, off to the Grid we go

- I/O issues are the hardest to deal with
- Techniques the “big guys” use (srm, dcache), not necessarily best for new users
 - lots of admin overhead

Simple I/O strategy worked for us!

- > GridFTP to prestage to \$OSG_DATA
 - Careful filename choices (use serial numbers)
 - Useful for dozens of GBs updated monthly

Final Phase: Hardening

- From “Worked Once” to “Always Works”
 - As much work as from zero to “Worked Once”
 - Critical for production work
- Dagman is helpful for managing workflow
- Site selection still tricky
 - OSG MM good choice
 - <http://osgmm.sourceforge.net> (Mats Rynge, RENCI)
 - Also condor_job_router, glide-in
 - But none are currently available as part of OSG stack
 - Engage can help with these choices

Summary

- > Run local first
 - Finds lots of surprises
- > Understand your CPU and I/O needs
- > Get help from Engage Team!