

Inter-Campus Grids: Topics for Discussion

Brian Bockelman

A Forward

- These slides are meant to start the discussion, not be a final answer.
- There will be a running emphasis throughout to provide something which:
 - Is simple
 - Can be evolved from current deployments.

A Vision

- I want to give users a clear path to “take off” with HTC, with each (small) step building on the previous. The pathway is:

- Desktop
- Cluster
- Campus
- Regional
- Worldwide

It is “Maslow’s hierarchy” for HTC.
Need to achieve at each step to proceed to the next one!

An Implementation

- Desktop: Personal Condor, installed via “yum” as part of their Linux OS.
- Note we can't make all users happy.
- Cluster: Traditional Condor Cluster
- Campus: Flocking and the “campus router” technology (see Derek's slides)
- Regional: Inter-campus clusters.
- Worldwide: OSG

Inter-Campus/Regional

- We are all here to discuss campus grids, which I believe is a common missing step between the cluster and the world grid.
- The next few slides are “starter questions” for topics I want to solve (or start solving).
- Some are existential, some are highly technical.

If there's no discussion, this is going to be a real quick morning!

Who/What are we?

- First question: what organizational umbrella do we fit under?
- Two simple answers: OSG or DiaGrid.
 - I'm familiar with the former... it seems to have more well-defined infrastructure. Willing to hear ideas...
 - More difficult: Something new?

This one doesn't have to be answered now...

Service Discovery

- We know how OSG does this (BDII and MyOSG). How does Purdue accomplish this?
- What improvements are coming from the Condor team?
- I.e., can we build an uber-collector for all sites in 7.5.4?

Service Discovery

- Right now, very rough security list at Nebraska:
 - DENY_WRITE = anonymous@*
 - ALLOW_WRITE = *@unl.edu *@glidein.unl.edu *@claimtobe/*.purdue.edu execute-side@matchsession
 - DENY_ADVERTISE_SCHEDD = *@claimtobe/*.purdue.edu anonymous@*
 - DENY_ADMINISTRATOR = anonymous@*
 - DENY_DAEMON = anonymous@glidein.unl.edu
 - ALLOW_DAEMON = *@unl.edu *@glidein.unl.edu *@claimtobe/*.purdue.edu execute-side@matchsession
 - DENY_NEGOTIATOR = anonymous@glidein.unl.edu
 - ALLOW_NEGOTIATOR = */*.unl.edu, */*.purdue.edu
 - FLOCK_TO = ff-grid.unl.edu, 129.93.227.68, cow.unl.edu, condor-fe00.rcac.purdue.edu, condor-fe01.rcac.purdue.edu, radon-fe00.rcac.purdue.edu, radon-fe01.rcac.purdue.edu, tg-condor.rcac.purdue.edu, tg-steele.rcac.purdue.edu, flamingo.rcac.purdue.edu
- What are we missing for others?
- TODO: Simple list of all resources.
- This afternoon TODO: automated design?

Security

- What are each site's security policy?
- How can sites fit together?
- How do users utilize the OSG securely?
- Can someone draw it up on the board?
- How is cross-campus incident response handled?
- Do we need to sign papers anywhere?

Job Matching

- What common ClassAd variables does each site have?
- Is there an equivalent to gWMS VO frontend environment scripts?
- Is there room for commonality?
 - Environment variables? Software env?
- TODO: list of needs here.

Everyone has a thing they do - can we align them?

What am I missing?

- Is there anything else we've left off for getting job X running on host Y?

“What are we” / Expansion

- UNL and Purdue have a stated goal of cross-campus flocking.
- What do we need to do to include (if they want):
 - OU, Wisconsin, Clemson, Duke, others???
- How do we register/record such a thing?
 - Immediate options: Twiki, web-page, OIM.

Accounting

- My requirements: I want to know how many hours I've gotten and how many I've given away.
- Condor has the Condor accountant; OSG has Gratia. What do we want to use?
- Gratia's data collection/forwarding and viewing are fairly powerful. However, it only instruments the submit node. This feels wrong.
 - Meets the requirements *only if* you instrument all submit nodes. Seems difficult.
- I want accounting sooner rather than later. Nebraska can demo what it looks at:
 - http://rcf-gratia.unl.edu/gratia/xml/glidein_hours_bar_smry
 - <http://t2.unl.edu/gratia/bysite?vo=hcc>

Expansion

- Who are other future collaborators?
- Right now, a collaboration of peers (in technical capability).
- Is there interest to “rope in” new, maybe smaller schools.
- For HCC, South Dakota and Puerto Rico comes to mind!
- Who has effort?

Expansion

- When things go wrong, who do we call?
- I.e., if HCC jobs die at Purdue, who notices? Debugs it?
- Do we have any sort of “minimum size” requirements for our “club”?
- How does Fermigrid fit in here? We have a pre-existing defined interface through OSG... are they different from the universities?

Expansion

- Eventually the user gets big (and hungry) enough to run on the grid.
- The natural way to do this is via flocking to a gWMS frontend node, assuming he has a valid grid certificate.
 - Derek can sketch how this works.
 - How does this compare to JobRouter?
 - Others?

Needs from Outside

- What do we need from:
 - Condor?
 - OSG?
 - Campuses?

Campus Grid Factory

- Derek will be talking about this.