L'ngagement Program

- Mission
 - Help new user communities from diverse scientific domains adapt their research computing to leverage OSG
 - Facilitate University Campus CI deployment, and interconnect it with the national organizations
 - Drive new requirements and important feedback to infrastructure developers and providers
- Methodology: EIE-4CI
 - Embedded Immersive Engagement for Cyberinfrastructure

Brian Blanton: Coastal



Modeling "Afl, Kevin, Mats, and I set up a test of running the coarse/fast ADCIRC system on 50,000 Monte Carlo simulated tracks that impact NC. The tracks are from Peter's track generation methods. Kevin and Mats set up the 50K runs to run on the NSF/DOE Open Science Grid. There are 2 images attached. The first shows the jobs running on OSG, and the big, sustained blip of jobs is the submission and delegation of the 50K runs onto available compute resources. From the graph, it looks like the 50K runs took about 7 hours.

The second figure is the max elevation (in meters) at the 489 coastal nodes for all of the 50K tracks. I haven't looked in detail at the results, although they look reasonable.

The main point is that this was fairly easy to do, and this will allow us to explore sensitivities to track selections for the Flood Plain Mapping simulations." Brian

y Blake: Information and Library Science

- "Claim Jumping through Scientific Literature"
 - a collaborative research project with Dr. Catherine Blake, Assistant Professor in the School of Information and Library Science at UNC-CH
- investigating approaches to multi-document summarization of scientific literature across disciplines.
- natural language parsing (NLP) of a large sample set (162,000) of biomedical research papers from the TREC (Text Retrieval Conference - NIST) Genomics Track document collection.
- "Using the OSG for this task has reduced NLP analysis time for the TREC collection from weeks to only a few days. The dramatic reduction in running time has allowed us to experiment and to fix problems iteratively in the text preprocessing and NLP that would not have been possible on a multi-week time scale."
- $\bullet \qquad http://www.opensciencegrid.org/About/What_We\%27 re_Doing/Research_Highlights/RENCI_Research_$

Initial interactions with new users

- User describes executable, needed inputs and example on how to run the model
- Every user is different, but in general, the Engagement team creates:
 - submit tool (creates jobs / dags)
 - -job-wrapper (wraps model remotely)
 - -job-success-check (checks stdout)

The Matchmaker

- Matchmaker matches compatible classads
 - Matchmaker ensures that Server and Customer Constraints are satisfied
 - Server and Customer Preferences are honored (within bounds of framework)
- Relevant parties notified when successfully matched
- Matched parties claim each other



www.cs.wisc.edu/condor

OSG: Resource Discovery

CE advertises capabilities and state (GIP & CEMon)

- ReSS Resource Selection Service
 - Condor ClassAd format
- BDII Berkeley Database Information Index
 - LDIF format





ReSS

- Collects data from compute elements (CE), storage elements (SE), and software entities
- Publishes the data in Condor ClassAd format
- One ClassAd per Cluster, Subcluster, CE, SE, VO
 - Cardinality of CE*Cluster*Subcluster*VO*SE*VO
 - Currently about 15,000 ads





Information in ReSS

- OS name / version
- LRM information
 - Total number of job slots
 - Assigned slots
 - Open job slots
- Memory / CPU / Disk
- Network setup
- Storage configuration

Validity of ClassAds

• Each ad is augmented with validity tests in the form of classad attributes

MyType = "Machine"	
GlueSubClusterLogicalCPUs = 2	
GlueCEPolicyAssignedJobSlots = 0	
GlueCEInfoHostName = "antaeus.hpcc.ttu.edu"	
GlueHostNetworkAdapterOutboundIP = TRUE	
GlueHostArchitectureSMPSize = 2	
<pre>DSGMM_Software_Rosetta_v3 = TRUE</pre>	
$DSGMM_MemPerCPU = 1010460$	
GlueSubClusterWNTmpDir = "/state/partition1"	
OSGMM_OSGAPPWriteWorkNode = TRUE	
<pre>GlueCEInfoContactString = "antaeus.hpcc.ttu.edu:2119/jobmanager-lss</pre>	f"
<pre>GlueHostOperatingSystemName = "CentOS"</pre>	



OSGMM – OSG Match Maker

- Simple match maker for Condor-G jobs
 - Based on *"Matchmaking in the Grid Universe"* in the Condor manual
- Open Source

- http://osgmm.sourceforge.net/

• Installs on top of the OSG Client software stack

Match Making against CEs• CE as a black box



- Opportunistic cycles
- Drop some jobs in and see how it

goes

- Keep some history of success /
 - performance
- Adjust Rank / Requirements

OSGMM – How does it work?

- Retrieve base ClassAds from ReSS
- Validate/maintain the sites with probe jobs
- Determine the current state of the system by looking at current job states and success rates (continuous system feedback)
- Merge the information, and insert into local Condor system
- Let Condor manage the jobs



dditional Job Requirements for Resource Selection

- Job fails...
- Job is in the queue for too long result another too long
- Job is running for too long...

resubmit to another site

• When submitting to another site, do not submit to a site which we have already failed on

Questions?

OSG Engagement VO https://twiki.grid.iu.edu/twiki/bin/view/Engagement/WebHome

engageteam@opensciencegrid.org

http://www.cs.wisc.edu/condor/CondorWee