GridNebraska

Brian Bockelman
January 19, 2010

Holland Computing Center

- At the University of Nebraska, all of our high-performance computational resources are centralized in the Holland Computing Center (funded by a generous donation by the Holland family).
- This encompasses users from Lincoln, Omaha, and the medical center.

Holland Computing Center Resources

Omaha

"Firefly" - Linux cluster, ~5000 cores, Infiniband interconnect, 150TB of Panasas storage

Lincoln

"Red" - CMS, 1100 cores

"Prairiefire" - ~500 cores, Infiniband

"Merritt" - SGI Altix, 512GB RAM

The easy part is hardware

- As probably everyone in this room knows, anyone with a sufficient checkbook can build a cluster, run HPL, and make a press release.
- The hard part is building a user base - porting applications, training users about Unix, finding users, teaching them about workflow management, etc.

Building a User Base

- In the last round of funding, we've opened positions I've nicknamed:
  - "go faster": Helping users port application and optimize them (hired 2.0 FTE).
  - "go bigger": Farming out work over multiple clusters and enabling users to access national-level resources (1.0 FTE position to be hired).

GridNebraska

- (Warning: placeholder name bound to change)
- Working text: “GridNebraska aims to enable Nebraska researchers to access computing power across the Holland Computing Center and national computing resources.”

Bullet-point Goals

- Integrating resources: Provide uniform execution environments, global storage across HCC, and grid submission nodes.
- Education: Campus workshops, web FAQs and documentation on how to efficiently use resources and Condor.
- Engagement: One-on-one training with Condor and porting of hacked perl into workflows.
- Research: Investigate new technologies such as MapReduce and cloud computing; work with grad students.

Accomplished

- Web-based user registration and one-password login across the site.
- Uniform execution environment using “module”.
- Ported several applications to OSG using OSG-MM.
- Introductory workshop in December.

TODO

- Global storage system - Panasas?
- Grid Submission nodes.
- OSG/Grid documentation and education.
- Continuously improve uniform execution environment - extend to grid jobs?
- GridNebraska OSG VO?
- Engage, engage, engage!
- More homogeneous system administration.

R&D

- MapReduce / Hadoop-on-Demand
- Cloud Computing
- GlideInWMS
- Data Management

R&D: MapReduce/HOD

- Dovetailing with our investments into HDFS, we have several graduate students using the MapReduce portions of Hadoop.
  - One of our student-run clusters is dedicated to Hadoop.
  - Our SGE cluster has Hadoop-On-Demand installed - you can run MapReduce jobs from within the batch system.
    - Needs to be better maintained, documented, and put into production.

R&D: Cloud Computing

- We can't offer some desired computational environments to our users:
  - "The manufacturer of our gene sequencer has prepackaged software that only runs on RedHat 9."
  - "Our application only runs on Windows."
  - "We need to log in to the app's GUI..."
- We are not staffed to provide these environments.
- Can Cloud Computing help us provision resources for users without having to provide OS support?
  - We'll see...

R&D: GlideInWMS

- We have a GlideInWMS testbed. Compared to OSG-MM, we are eyeing:
  - Greater scalability, potentially better usability, faster turn-around, and even less exposure to grid "warts". All the good glide-in stuff.
  - The product is far less refined, harder to run, documentation is lacking, doesn't work out-of-box. Will take a lot to make usable
    - Need lots of outside expert help.

R&D: Data Management

- Where do we even start with this?
- Library has a digital preservation project.
- Figure out what users need!
- "Global file system + archival storage + let users organize their files" seems like the traditional system with the traditional drawbacks.
    - But we *do not* provide archival storage currently, nor do we have the funding or experience to do this. What to do?

Thoughts and Conclusions

- We have lots of work cut out for us!
  - Some of it is mundane and involves "integration" more than "grid".
  - Keep the main focus on enabling users - usability for single less-expert user is key.
  - Will continue to rely on students to extend our ability to try new things.
  - More worried about storage than computing.