Less talking, more doing

Crowd-sourcing the integration of Galaxy with a high-performance computing cluster
The Goal

Enable users of the Michigan State University Genomics Core to perform their own analysis using their High Performance Computing Cluster infrastructure

Via:

1. Integrated institutional login
2. Import/export data from/to cluster storage while respecting permissions
3. Utilize existing node allocations and quotas; jobs must run as a HPCC user not a generic Galaxy user
4. Use the existing installed bioinformatics tools (no installs from the tool-shed)
Institute for Cyber-Enabled Research

- $10 million for developing collaborative, interdisciplinary computational projects through a faculty scholars program and post-doctoral fellowships
- Home of Michigan State University’s HPCC

High Performance Computing Center

- 8, 16, 32, or 64 cores
- 8GiB - 2TiB of memory/node
- Advanced GPU and Intel PHI capabilities also available
- > 7000 cores in main cluster incl 800 core HTCondor system
- 339 TB scratch storage, 192TB user storage
The Plan

Do It Ourselves: open agile deployment

All stakeholders set aside a single work day to get as much done as possible

Community support solicited via galaxy-dev@ and Twitter

Public chat room to document our work
March 5th, 2014

Step 1: web server up with galaxy.msu hostname #MSUGalaxy #usegalaxy

bam! we just submitted a job as a normal, real user, from Galaxy #msugalaxy
6 people joined our chat room to provide encouragement and very useful advice

Thanks to Marten Martenson, Alper Kucukural, Dannon Baker, Lauren M and Nate Coraor!
Zero to Success in 8 Hours

- No code changes needed
- Only minimal prep beforehand
- Login using existing Shibboleth infrastructure (no new accounts or passwords)
- Jobs running as the user’s account with quota control on the existing compute cluster
- Frontend + database running on a VMware ESXi 5.1 virtual machine (4 cores, shared, NetApp NFS backed)
- Deployed using Puppet
- Will be migrating to the community’s Puppet configuration
The Result

Tools using already installed software
The Future

- Filesystem permissions automation (each homedir is own filesystem & needs the SHARENFS property managed)
- Galaxy upgrade procedure & testing
- More user outreach

courtesy @nodoubleg
Dirk Colbry\textsuperscript{1}, \textbf{Michael R. Crusoe}\textsuperscript{2}, Andy Keen\textsuperscript{1}, Greg Mason\textsuperscript{1}, Jason Muffett\textsuperscript{1}, Matthew Scholz\textsuperscript{1}, Tracy K. Teal\textsuperscript{2}

1 Michigan State University, Institute for Cyber-Enabled Research
2 Michigan State University, Department of Microbiology and Molecular Genetics

Nicholas Beckloff, Genomics Core Director