Galaxy Deployments at Indiana University

Carrie Ganote

National Center for Genome Analysis Support

April 16, 2015
About IU and NCGAS, and me

A bit about Indiana University:
Very culturally diverse campuses
Known for our Jacob’s School of Music and Kelley School of Business

A bit about NCGAS:
National Center for Genome Analysis Support
NSF/NIH/IU funded group offering bioinformatics support/Linux help and lab experience to researchers

A bit about me:
NCGAS Staff since Oct 2012
First year PhD student, woohoo!
An unhealthy obsession with turkeys
National Center for Genome Analysis Support: http://ncgas.org
National Center for Genome Analysis Support: [http://ncgas.org](http://ncgas.org)
Fin

Thanks for watching!
Questions and comments:
Email help@ncgas.org
IU Galaxy

• Running since 2012 on Mason cluster at IU
• Deployed on Quarry, Big Red2 and soon to be Karst
• Apache proxy + Kerberos Authentication
• Lustre Filesystem
• Most tools, with emphasis on large memory applications
• University-wide, plus collaborators
• Still using hg

Fancy javascript tool menus!

National Center for Genome Analysis Support: http://ncgas.org
BLAST is an embarrassingly parallel program – it’s our only tool that uses the grid.
The Game Plan

Galaxy Host - VM

Galaxy Framework

Job

Pulsar Runner (localhost)

Job script submits to OSG then submits to Mason

OSG

Highly distributed grid

Splits input files
Creates Condor jobs
Collects all results
Compress and tar

Mason

Large memory cluster

Unpacks results
Merges results in memory
Returns final result

Photo by Soichi Hayashi. National Center for Genome Analysis Support: http://ncgas.org
Trinity Galaxy

National Center for Genome Analysis Support: http://ncgas.org
Trinity Galaxy

National Center for Genome Analysis Support: [http://ncgas.org](http://ncgas.org)
Trinity Galaxy

- One tool/pipeline emphasis
- Ftp server
- Local users

lib
galaxy
  config.py
  managers
    context.py
  tools
    parameters
      grouping.py
  api
    remote_files.py

National Center for Genome Analysis Support: http://ncgas.org
Challenges

Job Runners:
• PBS – still breaks with Torque 4x
  • Seems to have trouble interacting with the dynamic runner
• Drmaa – one destination possible?
• Dev servers are always somehow different than production!

Future Goals

• Trinity needs some more guinea pigs!
  Sign up now!
• IU Galaxy still needs to be moved to Github
• OSG-BLAST is still in beta
Fin (for real)

Thanks for watching!
Questions and comments:
Email help@ncgas.org

National Center for Genome Analysis Support: http://ncgas.org