



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 toolshed)

The Resources

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

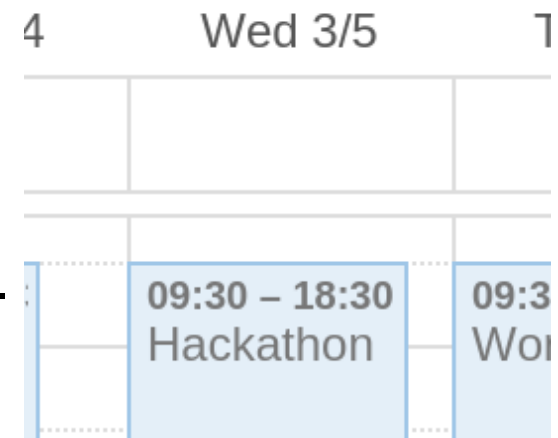
The Plan

Do It Ourselves: open **agile**
deployment

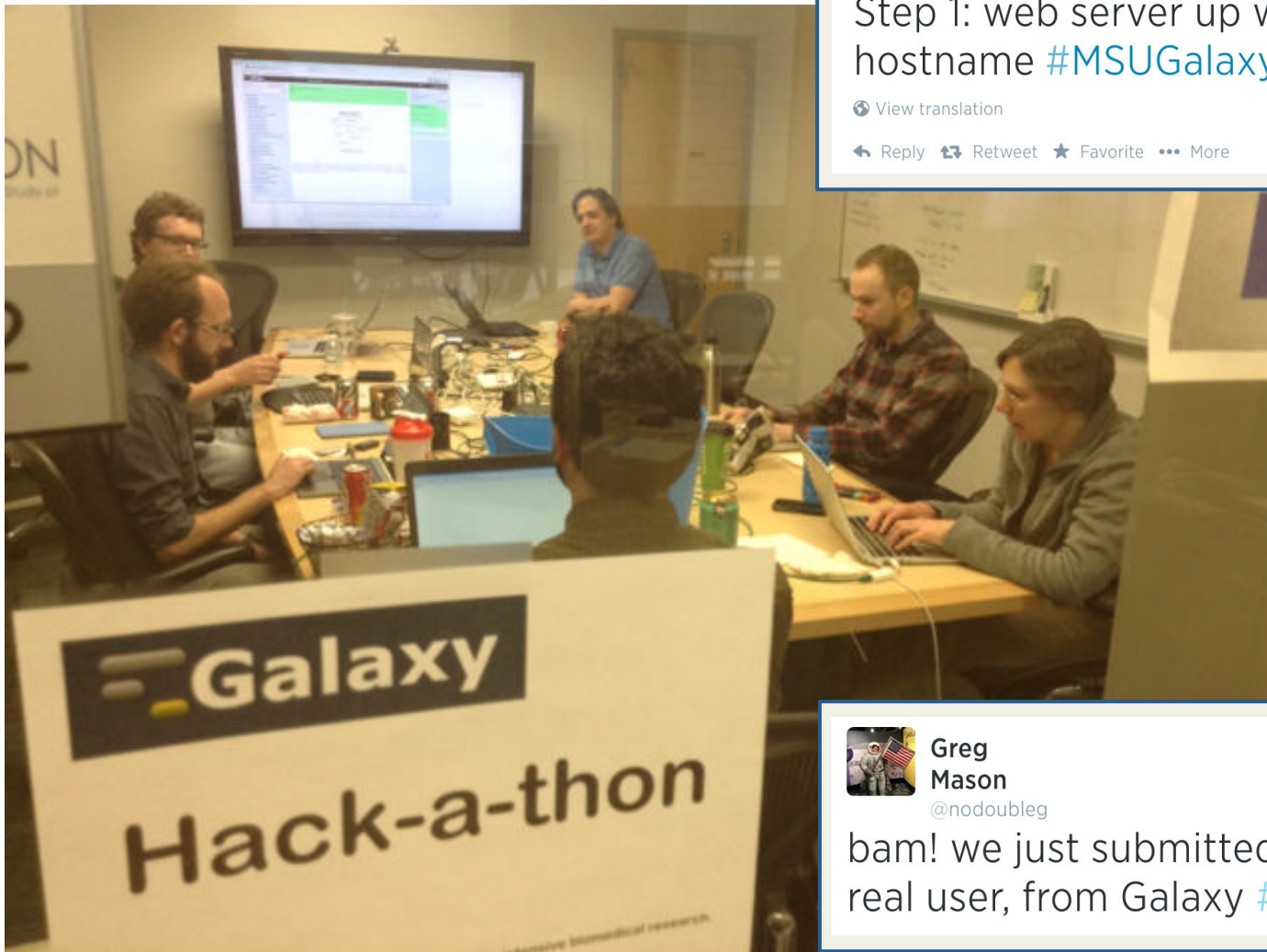
All stakeholders set aside a single work
day to get as much done as possible

Community support solicited via galaxy-
dev@ and Twitter

Public chat room to document our work



March 5th, 2014



Tracy Teal
@tracykteal



Following

Step 1: web server up with galaxy.msu
hostname [#MSUGalaxy](#) [#usegalaxy](#)

[View translation](#)

[Reply](#) [Retweet](#) [Favorite](#) [More](#)



Greg Mason
@nodoubleg



Follow

bam! we just submitted a job as a normal,
real user, from Galaxy [#msugalaxy](#)

Community Assistance

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

MICHIGAN STATE UNIVERSITY

NOTE: In order to use this application your browser must accept cookies.

Enter your MSU ID and password

MSU ID

(MSU NetID or MSU Community ID)

Password

Authenticator

MSU Net ▾

Login

For Authorized Use Only

The Result

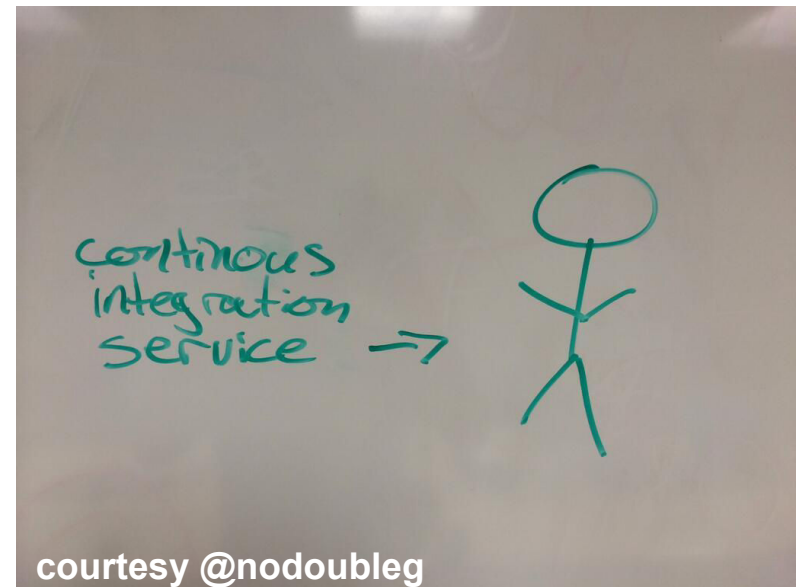
The screenshot displays the Galaxy web interface. At the top, the navigation bar includes 'Galaxy', 'Analyze Data', 'Workflow', 'Shared Data', 'Visualization', 'Admin', 'Help', and 'User'. The main content area features a dark green banner with the text 'MICHIGAN STATE UNIVERSITY | Institute for Cyber Enabled Research' and the ICER logo. Below the banner is a paragraph of text: 'Galaxy is an open, web-based platform for data intensive biomedical research. The Galaxy team is a part of BX at Penn State, and the Biology and Mathematics and Computer Science departments at Emory University. The Galaxy Project is supported in part by NHGRI, NSE, The Huck Institutes of the Life Sciences, The Institute for CyberScience at Penn State, and Emory University.'

On the left side, the 'Tools' panel is visible, containing a search bar and a list of tool categories: 'Get Data', 'Trimming and QC', 'NGS: QC and manipulation', 'NGS: Mapping', 'NGS: RNA Analysis', 'NGS: SAM Tools', 'NGS: GATK Tools (beta)', 'Phenotype Association', 'SNP Detection', 'Krona', 'visualization', 'NGS: Assembly', 'Metagenomics', 'Community Analysis', 'SNP', and 'Workflows'. A yellow callout box with a blue border and an arrow pointing to the 'NGS: GATK Tools (beta)' category contains the text 'Tools using already installed software'.

On the right side, the 'History' panel shows 'Unnamed history' with '0 bytes' and a message: 'Your history is empty. Click 'Get Data' on the left pane to start'.

The Future

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



Credits

Dirk Colbry¹, **Michael R. Crusoe**², Andy Keen¹, Greg Mason¹, Jason Muffett¹, Matthew Scholz¹,
Tracy K. Teal²¹ Michigan State University, Institute for Cyber-Enabled Research
² Michigan State University, Department of Microbiology and Molecular Genetics

