GCC2014 Galaxy Hackathon
(duplicate the sample slide and have at it)

Dannon Baker, Brad Chapman, John Chilton, Kyle Ellrott
(and about 40 other people)
Making Connections, Providing Footholds

- Developer Community Building
- Bandwidth is important
- Great opportunity for ‘fun’ projects
By the Numbers

- 40 people
- 1 lab
- 3 days (Training Day Overlap)
- 2 sponsors
- 17 active Trello Cards (8 marked complete)
- 33 Members on Trello Board
- 6 new pull requests for galaxy-central
- 2 new pull requests for cloudman
- NaN boxes of coffee, soda, snacks
Main Themes

- Cloud Deployment
- Tools and Improving the Toolshed
- Docker
- Improving workflows
- Improving the API
http://forge.puppetlabs.com/urgi/

```
docker build -t="galaxy" .
```

```
docker run -p 8000:8000 galaxy
```

http://localhost:8000/

(Mikael Loaec*, Olivier Inizan)
Deployment Automation with Ansible

- Deployed supervisord via Ansible at CloudMan boot, easily extensible for additional Ansible tasks
- Worked toward a unified Galaxy deployment and management system (have a production server in one command) http://j.mp/gxansible

(Peter van Heusden, Nate Coraor)
CloudMan Improvements

- CloudMan with Amazon VPC
- Architectural Planning for in EC2
- libcloud integration to extend cloud infrastructures support (e.g. VMware vCloud, OpenNebula) - work in progress

(Angel Pizarro, David van Enckevort, Enis Afgan, Dannon Baker)
<actions>
  <action type="setup_python_environment">
    <repository name="package_python_2_7" owner="iuc">
      <package name="python" version="2.7" />
    </repository>
    <!-- allow downloading and installing an Python package from https://pypi.org/ -->
    <package>https://pypi.python.org/kgmer-1.0.tar.gz#md5#b60639a8b2938df757</package>
  </action>
  <action type="set_environment">
    <environment_variable action="set_to" name="VAR">$/INSTALL_DIR</environment_variable>
  </action>
</actions>

(Björn Grüning)
Visualisation

- using Carl’s visualisation framework
- Two visualisation tools for RNA-Seq DGE
  - vennt (https://github.com/drpowell/vennt)
  - degust (http://www.vicbioinformatics.com/degust)

(Simon Gladman, Björn Grüning)
Visualisation

(Simon Gladman, Björn Grüning)
Visualisation

(Simon Gladman, Björn Grüning)
Visualisation

(Ira Cooke, James Johnson)
Galaxy Synapse Interface

- **Synapse: a platform for scientific data:**
  - data, source code, text and graphics
  - versioning, provenance tracking and access control
- **Common goals with Galaxy:**
  - open science, open data
  - reproducibility
- **Goal: share data, provenance, annotations, etc.**

(Christopher Bare*, Kyle Ellrott)
khmer project integration &
discover_datasets testing

Welcome to khmer: k-mer counting, filtering and graph traversal FTW!

Will be upgraded to dataset collections output when that is ready. 1 bug found by MRC & fixed by JC; 4+ Trello cards filed
(Michael R. Crusoe*, John Chilton, Dave B.)
Docker Docker Docker Docker Docker Docker Docker

- Docker
- Docker

Docker docker docker docker, docker, docker docker docker docker docker docker. Docker! Docker docker, docker.

(Docker)

<table>
<thead>
<tr>
<th>Docker</th>
<th>Docker</th>
</tr>
</thead>
<tbody>
<tr>
<td>Docker</td>
<td>Docker</td>
</tr>
<tr>
<td>Docker(^1)</td>
<td>Docker(^2)</td>
</tr>
</tbody>
</table>
Galaxy API, BioBlend, and blend4j

- **API wish list**
  - Ilya, Freek + John discussed 5 potential improvements
  - 1: fixed by Ilya, 2: to be done, 3: should work, we’ll check, 4: very difficult & 5: is already work in progress

- **Extending blend4j**
  - JJ and Freek worked on similar extensions on metadata (i.e. tool parameters)
  - we’ve compared code, will add it in coming months
    Also: multiprocess Galaxy + MySQL with MyISAM = failed workflow steps

(Ilya Sytchev, JJ Johnson, Freek de Bruijn, John Chilton)
Send multiple datasets to Galaxy

- **Problem:** can only send one dataset at a time
- **Solution:** JSON schema for query response

```
[{
  "url": "http://thecpgdb.org/cpgs_mm9.bed",
  "name": "CpG islands (mouse)",
  "extension": "bed",
  "metadata": {"db_key": "mm9"}
},
{
  "url": "http://thecpgdb.org/cpgs_ce2.bed",
  "name": "CpG islands (c. elegans)",
  "extension": "bed",
  "metadata": {"db_key": "ce2"}
}]
```

(Matt Shirley, Dan Blankenberg)
Code Crawling

Targets for implementing input data collections
Workflow as a step in another workflow

Include reference to workflow in another workflow, at runtime call steps of nested workflow. (Peter van Heusden, ideas from Saket Choudhary)

http://j.mp/gxwfinwf
Provenance and Scaling

- Initial work on backend Arvados integration
- http://arvados.org

Framework for data distribution, Docker integrated and provenance storage (PostgreSQL).
(Brad Chapman, Jonathan Sheffii)
Going Forward

- Docker
- Work distribution/federation
- Dataset Collections
- Workflow engine

Keep the conversation going:

chat.freenode.net #galaxyproject
galaxy-dev@lists.bx.psu.edu
http://galaxyproject.org/trello
Join us next week @ BOSC Codefest

http://goo.gl/NAFwA6

July 9th and 10th, 2014 (the Wednesday and Thursday before BOSC 2014) in Cambridge, Massachusetts at hack/reduce
Thanks!

Saravanaraj Ayyampalayam
Dannon Baker
Christopher Bare
Daniel Blankenberg
Carlos Borroto
Emil Bouvier
Michael Cariaso
Martin Cech
Brad Chapman
Alistair Chilcott
John Chilton
Dave Clements
Ira Cooke
Nathan Coraor
Michael Crusoe
Freek de Bruijn
Kyle Ellrott
Carrie Ganote
Nuwan Goonasekera
Björn Grüning
Olivier Inizan
James Johnson
Dawei Lin
Mikael Loaec
Yizhen Lu
Mark Mammel
Nathan Olson
Simon Gladman
Aaron Petkau
Angel Pizarro
Jonathan Sheffi
Matt Shirley
Andrew Stewart
Nick Stoler
Ilya Sytchev
Vipin T Sreedharan
Michael Ta
Frederick Tan
David van Enckevort
Peter van Heusden
Xijun Zhang
Enis Afgan
Hackathon Template Slide (copy me)

- Bullets
- Orange

Feel free to include screenshots/pictures. Please do include the names of people who hacked with you on the project. Pictures and logos too, if you have them!

(contributor 1*, contributor 2, contributor 3, docker)