

# **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

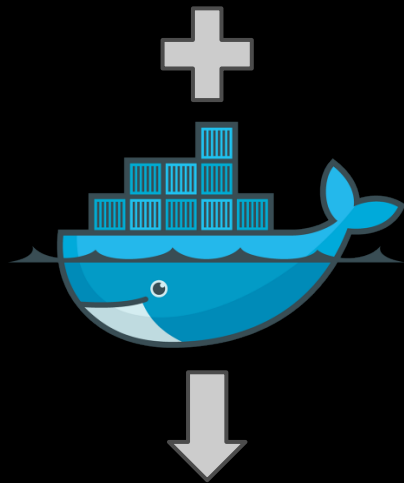
# Main Themes

- 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<sup>Man</sup> 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)



# Tool Shed - setup\_python\_environment

```
<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/khmer-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)

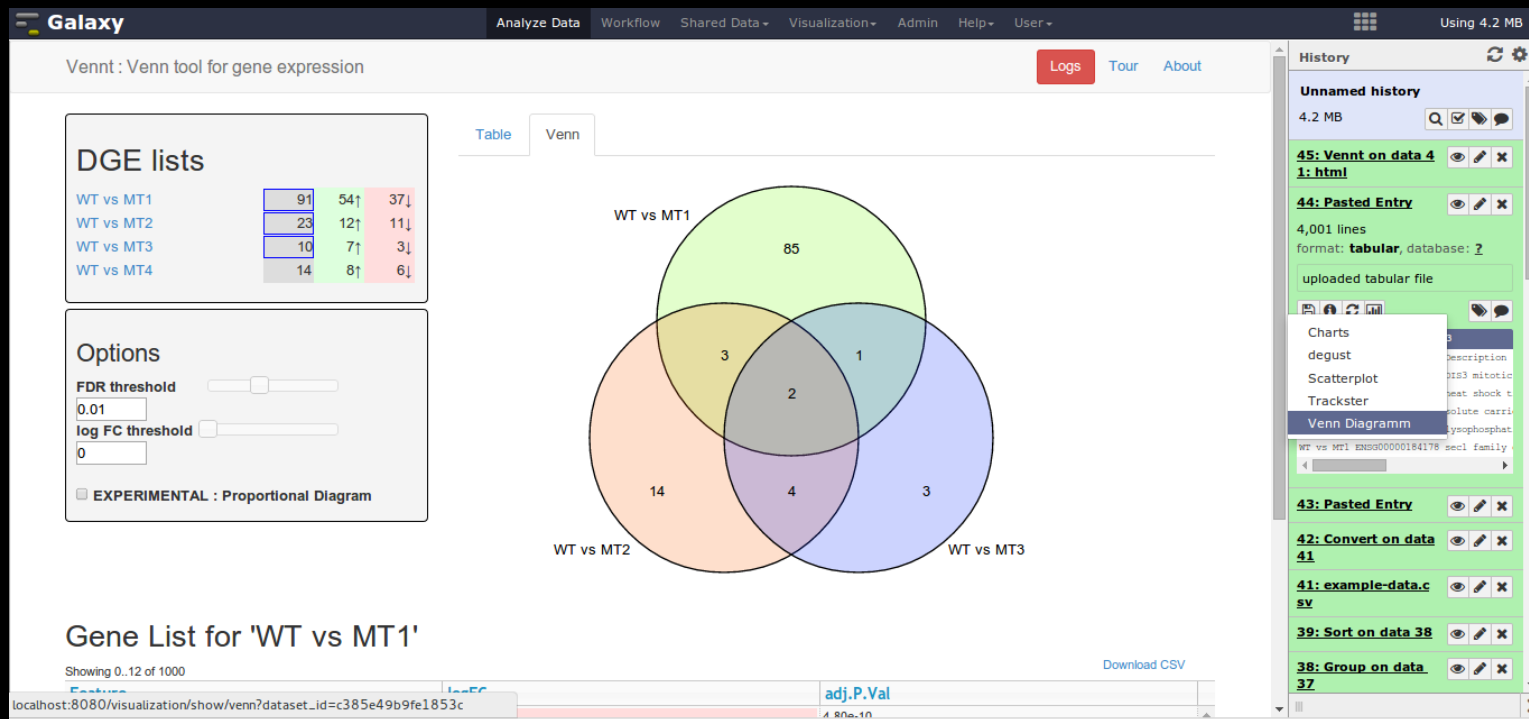


# Visualisation

- 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

The screenshot displays the Galaxy web interface. The top navigation bar includes 'Analyze Data', 'Workflow', 'Shared Data', 'Visualization', 'Admin', 'Help', and 'User'. The 'Tools' panel on the left lists various data sources and analysis tools. The main panel shows a table of protein identifications with columns for Name, Description, Probability, and Num Peptides. The first entry is highlighted, showing a sequence logo for the protein 'Thymidylate synthase; 278197'. The logo displays the sequence '3 G C P S F A Q T E P K D A L M Q T I M D L I Q E Q M T S N K N D G K E M D E A L I K L I K E V T E F F' with a probability scale from 20 to 60. The right panel shows the 'History' section with a list of recent jobs, including '30: squid.psqlite' and '28: mt272.psqlite'.

Name	Description	Probability	Num Peptides
lcl comp57626_c0_seq1m.151088	RecName: Thymidylate synthase; 278197 COG0207 Identity: 25.49 EValue:9.4	0.999	3
lcl comp55716_c2_seq1m.99931	RecName: Activating transcription factor 7- interacting protein 1; 9031 NOG82478 Identity: 50.79 EValue:2.0e-30	0.9908	2
lcl comp289902_c0_seq1m.166463	RecName: WD repeat-containing protein 65; 10090 COG2319 Identity: 46.73 EValue:4.0e- 66	0.9862	2
lcl comp50482_c0_seq2m.63494	RecName: UPF0658 Golgi apparatus membrane protein C23H3.04; 284812 NOG68074 Identity: 26.47 EValue:0.08	0.9824	2

(Ira Cooke, James Johnson)

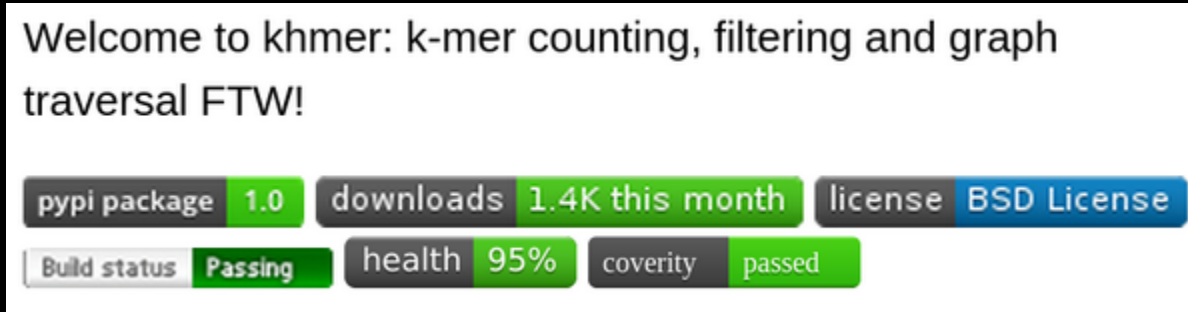
# Galaxy Synapse Interface



- **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



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.

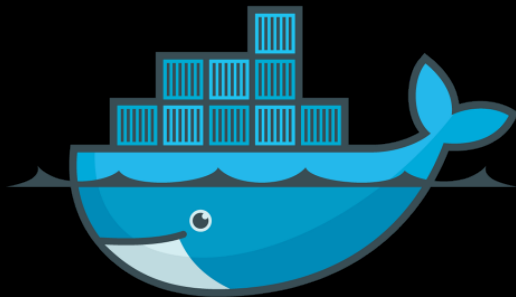
(Docker)

Docker	Docker
Docker	Docker
Docker <sup>1</sup>	Docker <sup>2</sup>

# Galaxy with Docker



Galaxy in Docker



Docker in Docker



Tools in Docker

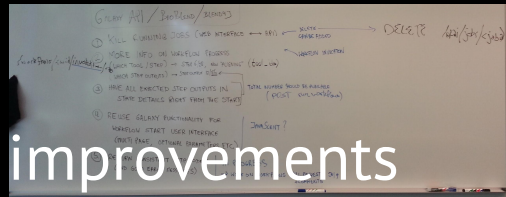
(Aaron Petkau, Nate Olson, Andrew Stewart, Björn Grüning, John Chilton)



# 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

Galaxy

“Get Data”

Database

JSON response

Galaxy

json\_data\_source

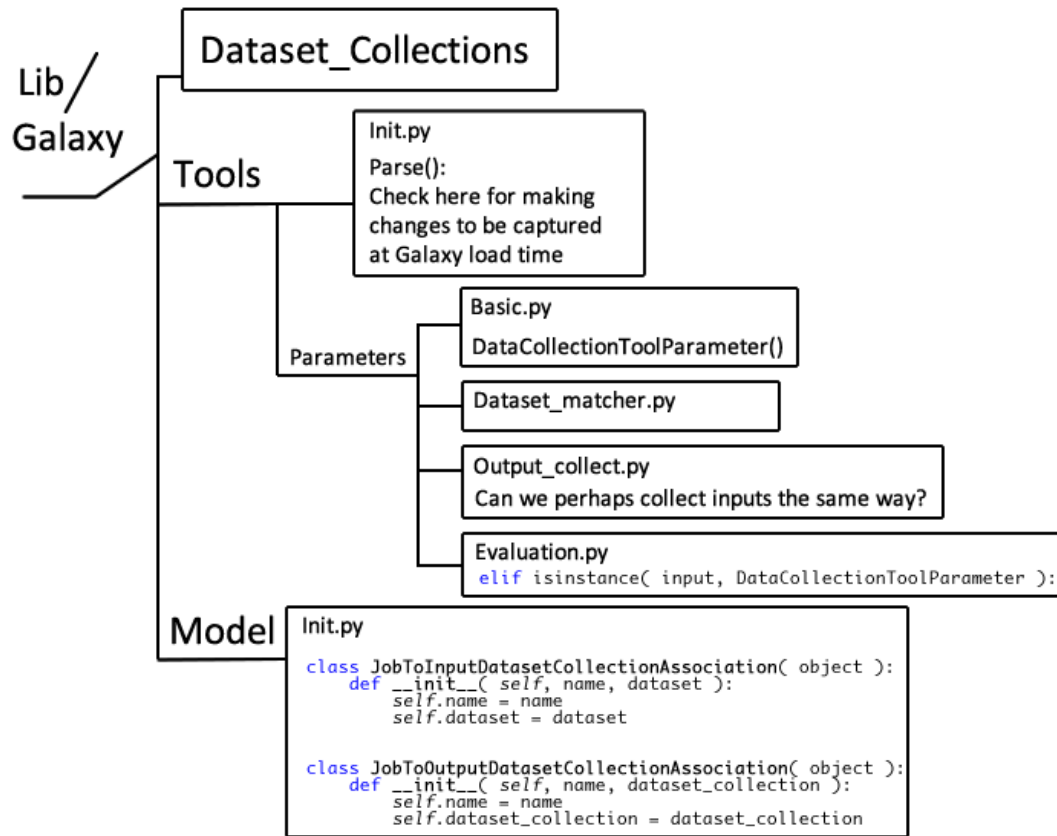
```
[{"url": "http://thecpgdb.org/cpgs_mm9.bed",  
  "name": "CpG islands (mouse)",  
  "extension": "bed",  
  "metadata": {"db_key": "mm9"}  
},  
{"url": "http://thecpgdb.org/cpgs_ce2.bed",  
  "name": "CpG islands (c. elegans)",  
  "extension": "bed",  
  "metadata": {"db_key": "ce2"}  
}]
```

3: CpG islands (mouse) 74,987 regions  
format: bed, database: mm9  
display at UCSC [main](#) [test](#)  
display in IGB [Local](#) [Web](#)  
display at Ensembl [Current](#)

1. Chrom	2. Start	3. End	4. N
chr	start	end	len
chr10	3001252	3001468	217
chr10	3001990	3002096	107
chr10	3102618	3102897	280
chr10	3103777	3103948	172
chr10	3104651	3105129	479

2: CpG islands (ce2)

(Matt Shirley, Dan Blankenberg)



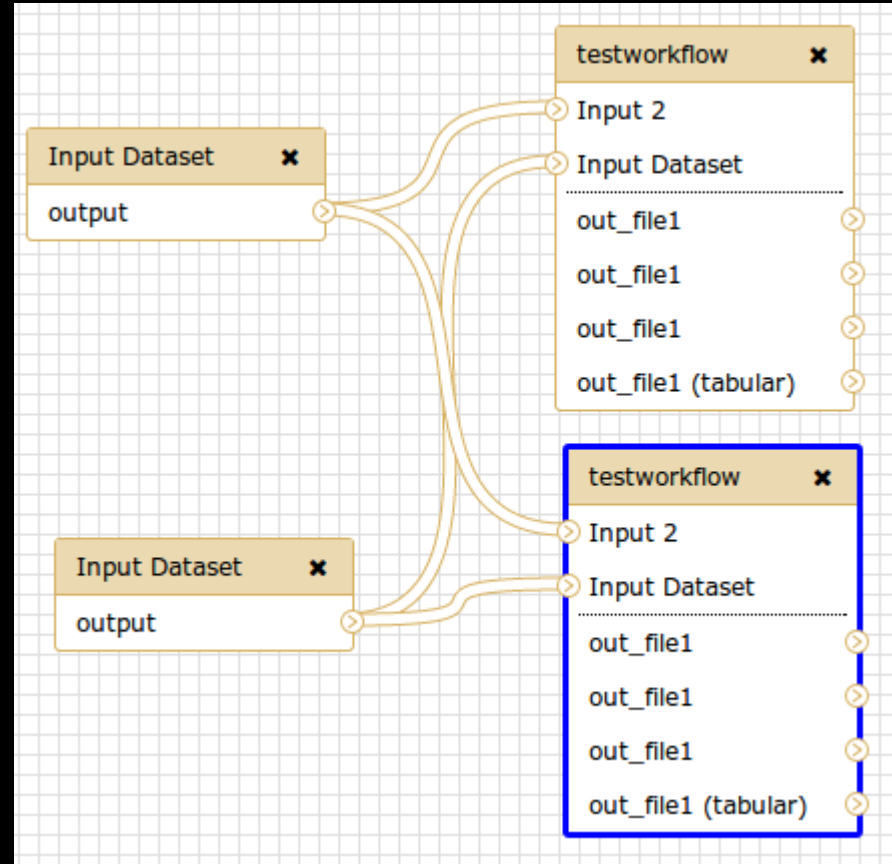
# Code Crawling

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

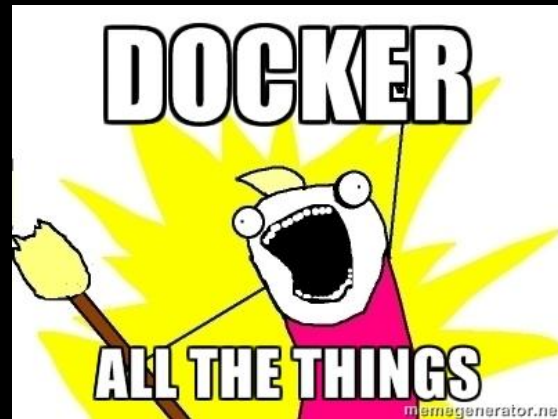
- Initial work on backend Arvados integration
- <http://arvados.org>



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

# Going Forward

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



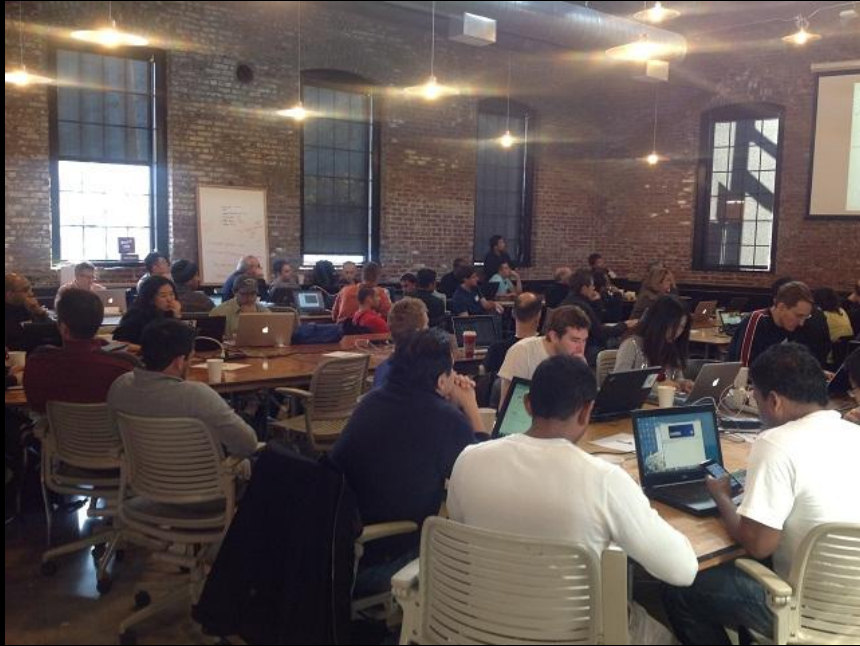
Keep the conversation going:

[chat.freenode.net #galaxyproject](https://chat.freenode.net/#galaxyproject)

[galaxy-dev@lists.bx.psu.edu](mailto: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



Arvados™



amazon  
web services™



# 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)