Galaxy CloudMan

A Gentle Introduction to Data Analysis on the Cloud

Plant & Animal Genome XXII
San Diego, January 15, 2014

Dave Clements
Johns Hopkins University
http://galaxyproject.org/

This talk complements yesterday's workshop →

http://bit.ly/gxypag2014

Galaxy @

Plant and Animal Genome XXII (PAG 2014)

UCSC Genome Browser

Sat 4:00-6:10, California Room
Robert Kuhn

The Banana Genome Hub

Tues 11:50-12:10, Pacific Salon 6-7
Gaëtan Droc, *et al.*

URGI Plant and Fungi Platform
Distributed Resources Through GMOD Tools

Wed 11:10-11:50, GMOD Workshop Golden West
Joelle Amselem, *et al.*

Galaxy for NGS Analysis
A Hands-on Workshop

Tues 4:00-6:10, California Room
Dave Clements, Anushka Brownley

This workshop will introduce the Galaxy platform and walk participants through a multi-step next generation sequencing data analysis, starting with quality control. We will review common choices in NGS data analysis, and demonstrate them within the context of Galaxy, taking advantage of Galaxy's tool set and visualization capabilities.

We will also provide a brief overview of what is needed to set up your own local Galaxy instance. This complements the *Galaxy CloudMan* talk on Wednesday

Galaxy CloudMan
A Gentle Introduction to Data Analysis on the Cloud

Wed 11:50-12:30, GMOD Workshop, Golden West
Dave Clements

Galaxy is open-source and web-based, with over 50 publicly accessible Galaxy servers and hundreds of private installations around the world. Galaxy can also be run on compute clouds using *Galaxy CloudMan*.

This talk will briefly introduce Galaxy, Galaxy CloudMan, and some basic cloud concepts. We'll then show a live demonstration of how to setup a Galaxy server on Amazon Web Services (one of several supported cloud infrastructures) using CloudMan, add a dynamically scalable compute cluster to perform analysis, customize the server by adding new tools, and then shut the server down. All steps can be done through a web browser, without ever using a command line interface.

Poster Sessions

Mon 10:00-11:30

P988: The South Green Bioinformatics Platform, Mathieu Rouard, *et al.*

P1050: Integrative System for Gene Family Gathering and Analysis in a Context of Crops' Stress Response Study, Delphine Lavivière, *et al.*

Mon 3:00-4:30

P135: SNP Genotyping to Accelerate Rice Breeding, Michael Thomson, *et al.*

P1041: RepeatExplorer: Collection of Tools for Mining of Repetitive Elements from NGS Data, Petr Novak, *et al.*

The Galaxy Project

Galaxy is an open source web-based platform for data integration and analysis in life sciences research.

The Galaxy Project is supported by a large and active community.

http://galaxyproject.org

What is Galaxy?

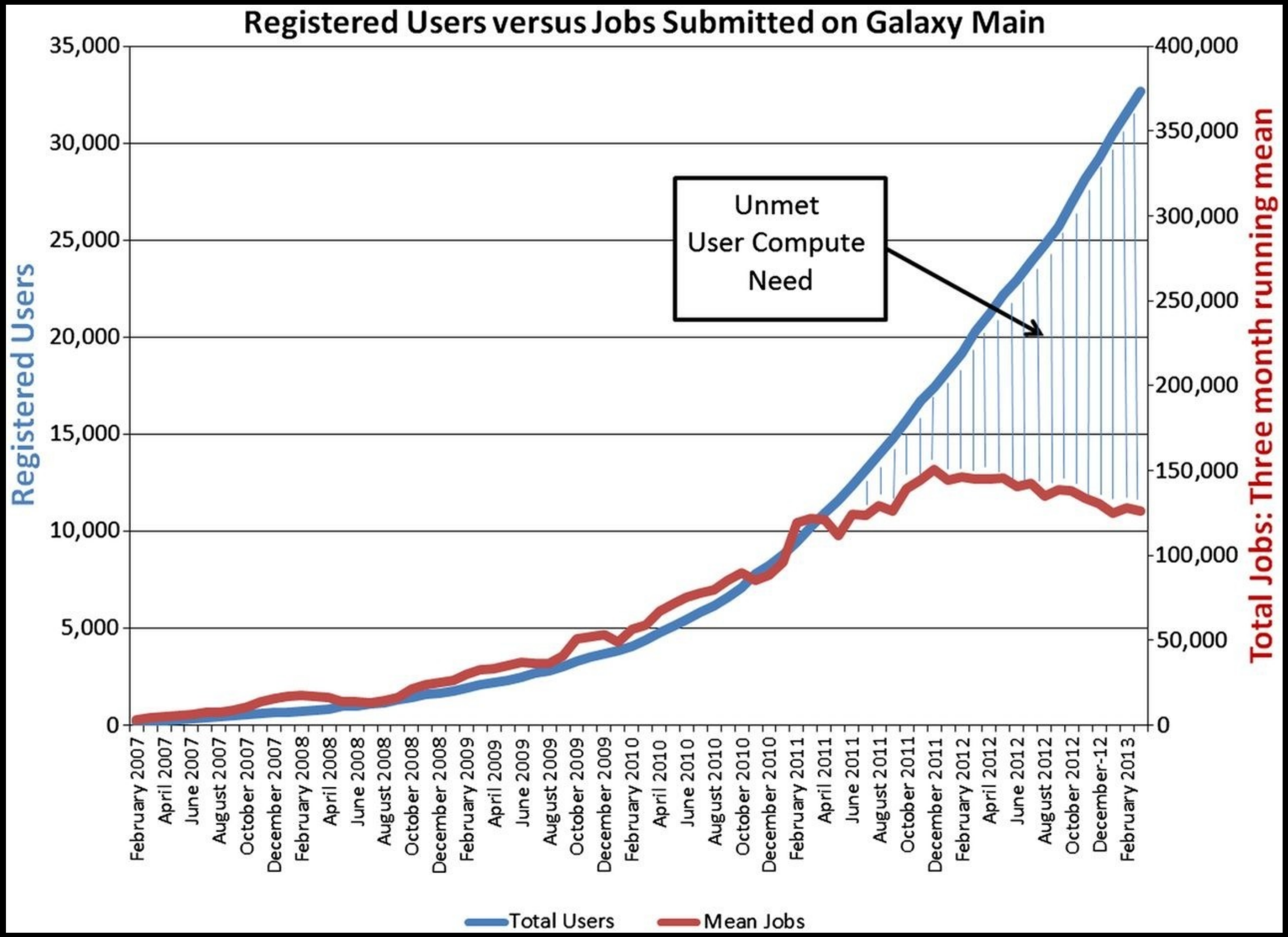
- A web based data integration and analysis framework.
- **Open source software**
- **A free (for everyone) web service**

http://galaxyproject.org

Galaxy is available ...

As a free for everyone web service (http://usegalaxy.org) integrating a wealth of tools, compute resources, terabytes of reference data and permanent storage.

However, ***a centralized solution cannot support the different analysis needs of the entire world.***



Leveraging the national cyberinfrastructure for biomedical research
LeDuc, *et al. J Am Med Inform Assoc doi:10.1136/amiajnl-2013-002059*

Galaxy is available ...

- As a free (for everyone) web service

  http://usegalaxy.org

- **As open source software** that you can install locally

  http://getgalaxy.org

See yesterday's slides

Galaxy is available ...

- As a free (for everyone) web service
- As open source software
- *On the Cloud*

http://wiki.galaxyproject.org/Cloud

AWS in Education Grants Program

http://aws.amazon.com/education

Galaxy is available ...

- As a free (for everyone) web service
- As open source software
- On the Cloud
- *With Commercial Support*

A ready-to-use appliance (BioTeam)

Cloud-based solutions (ABgenomica, AIS, Appistry, GenomeCloud)

Consulting & Customization (Arctix, BioTeam, Deena Bioinformatics)

Galaxy Project: Further reading & Resources

http://galaxyproject.org

http://usegalaxy.org

http://getgalaxy.org

http://wiki.galaxyproject.org/Cloud

http://bit.ly/gxychoices

What is our path?

Today we will:

- Launch our own Galaxy server on Amazon Web Services
- Make the server dynamically scalable in response to demand.
- Run some basic analysis on it.
- Make it go away.

Full Disclosure

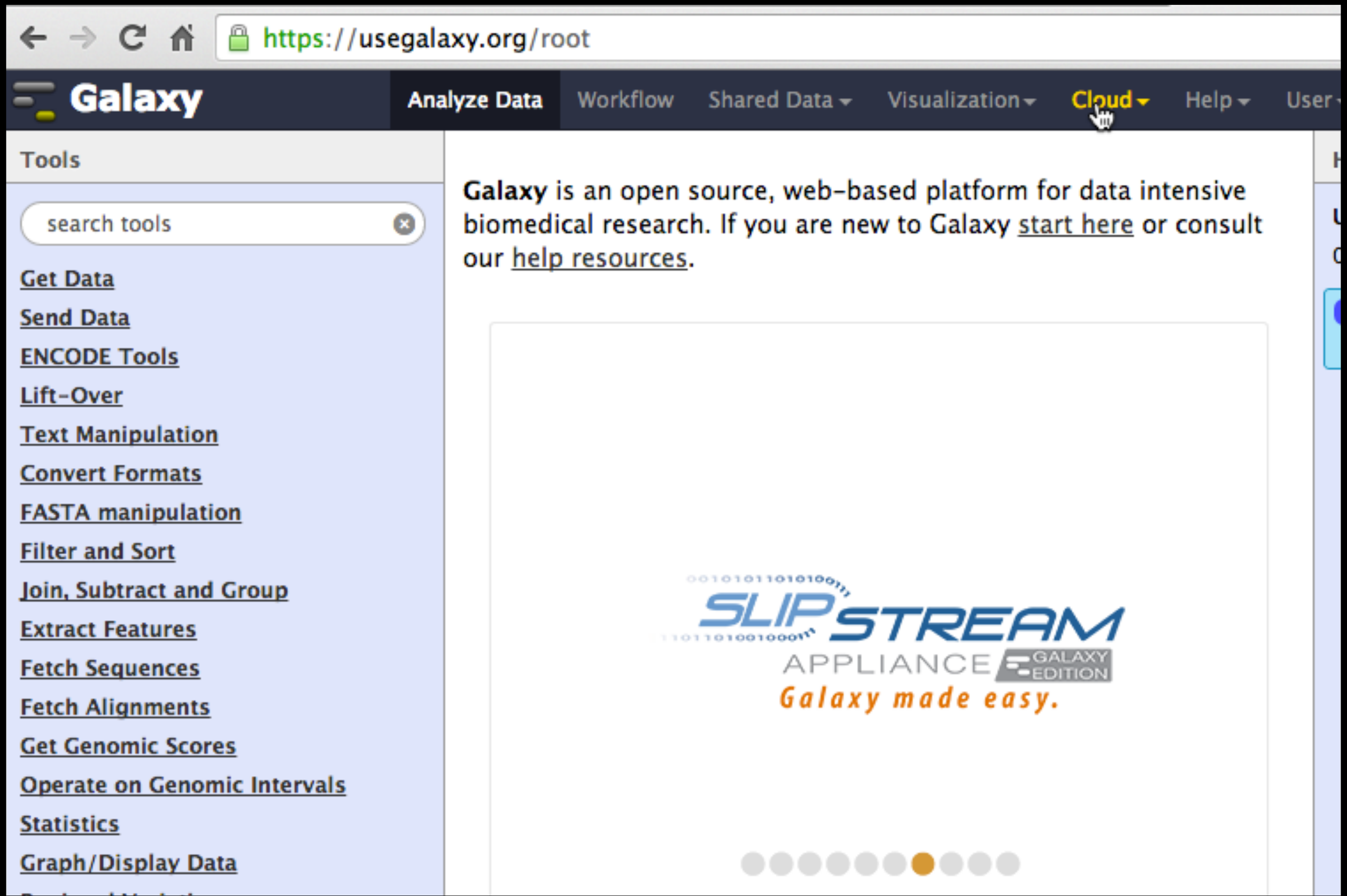
To use AWS you must create an AWS account with a credit card associated with it.

You must also have created a key pair.

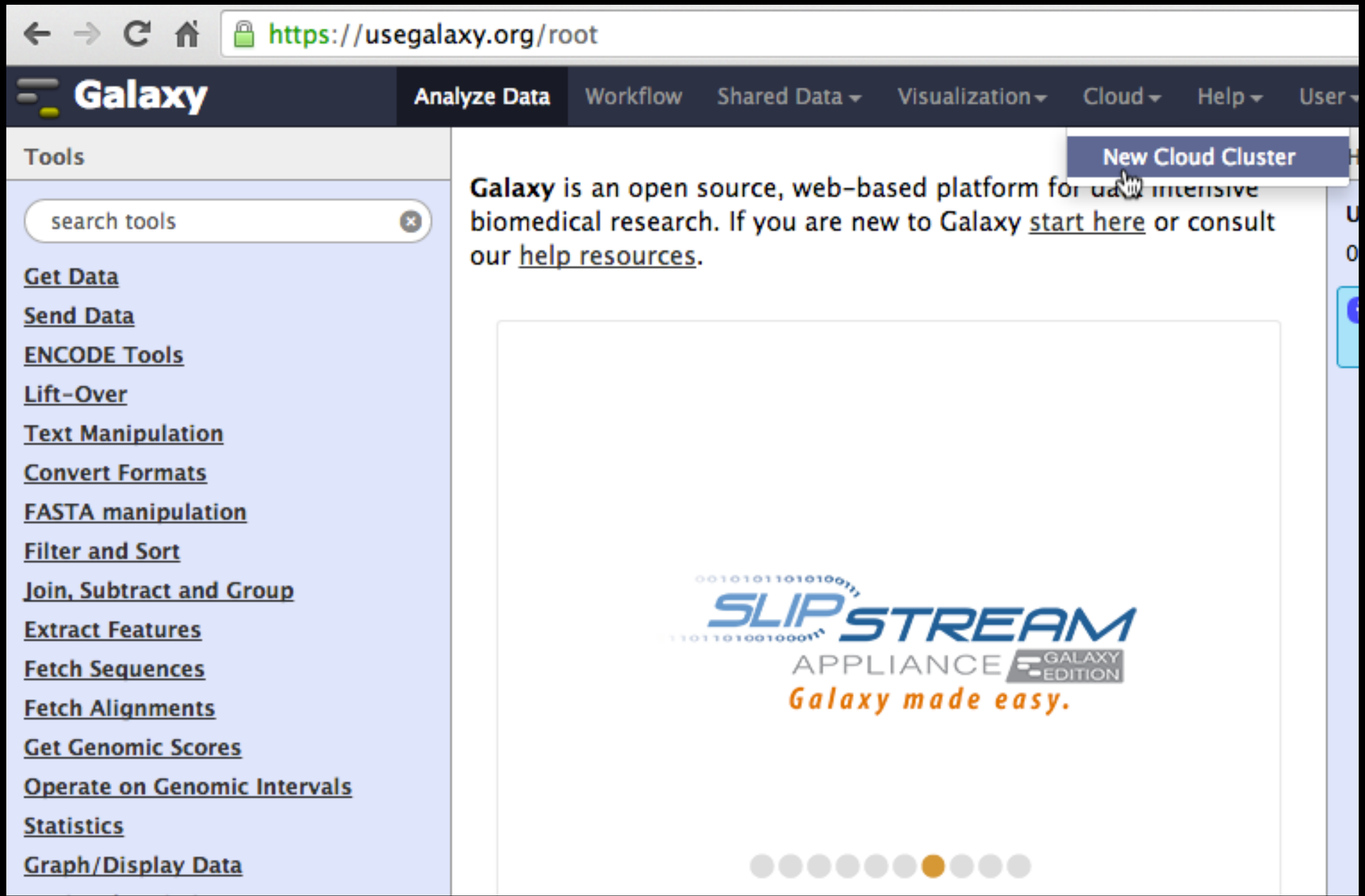
These have been done prior to this talk.

We will be using an IAM account, a limited AWS account that is useful for teaching with AWS.

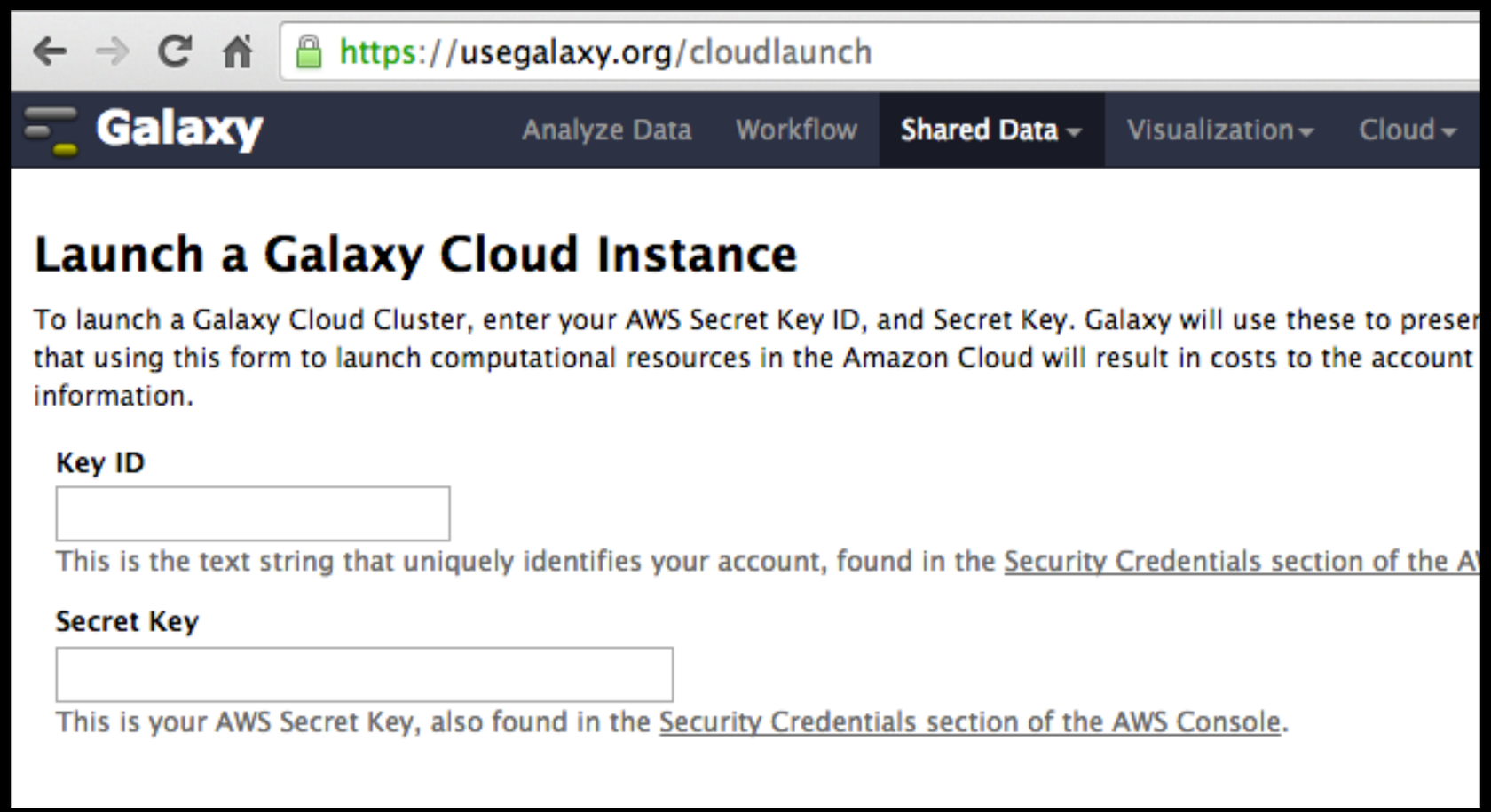
CloudLaunch



CloudLaunch



CloudLaunch



https://usegalaxy.org/cloudlaunch

Galaxy | Analyze Data | Workflow | Shared Data | Visualization | Cloud | Help | User

Launch a Galaxy Cloud Instance

To launch a Galaxy Cloud Cluster, enter your AWS Secret Key ID, and Secret Key. Galaxy will use these to present appropriate options for launching your cluster. Note that using this form to launch computational resources in the Amazon Cloud will result in costs to the account indicated above. See Amazon's pricing for more information.

Key ID

[redacted]

This is the text string that uniquely identifies your account, found in the Security Credentials section of the AWS Console.

Secret Key

[redacted]

This is your AWS Secret Key, also found in the Security Credentials section of the AWS Console.

Instances in your account

New Cluster

Cluster Name

PAG_CLOUD_2

This is the name for your cluster. You'll use this when you want to restart.

Cluster Password

••••••••

Cluster Password - Confirmation

••••••••

Key Pair

CloudManKP1

Instance Type

Large

Requesting the instance may take a moment, please be patient. Do not refresh your browser or navigate away from the page

Submit

CloudLaunch

https://usegalaxy.org/cloudlaunch

Galaxy | Analyze Data | Workflow | Shared Data | Visualization | Cloud | Help | User

Launch a Galaxy Cloud Instance

Launch Pending, please be patient.

https://usegalaxy.org/cloudlaunch

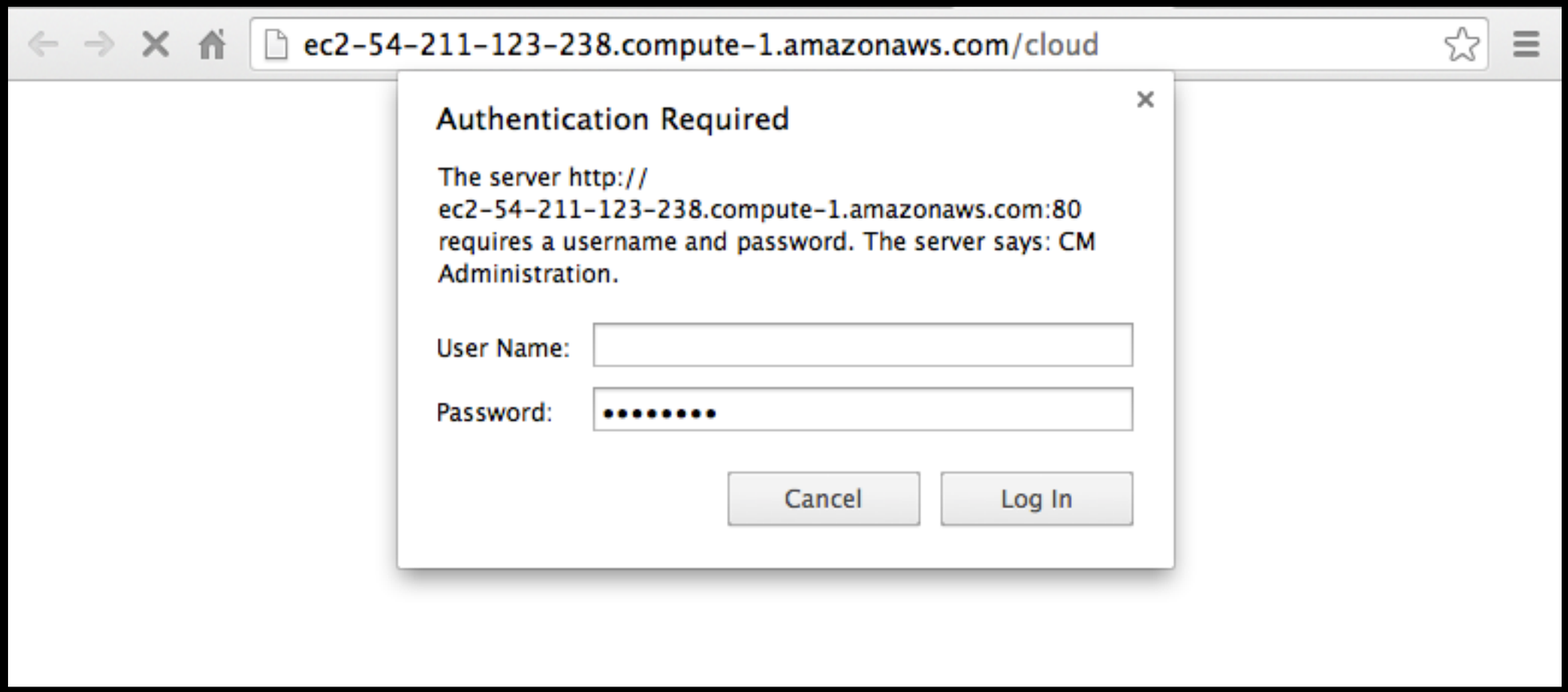
Galaxy | Analyze Data | Workflow | Shared Data | Visualization | Cloud | Help | User

Launch a Galaxy Cloud Instance

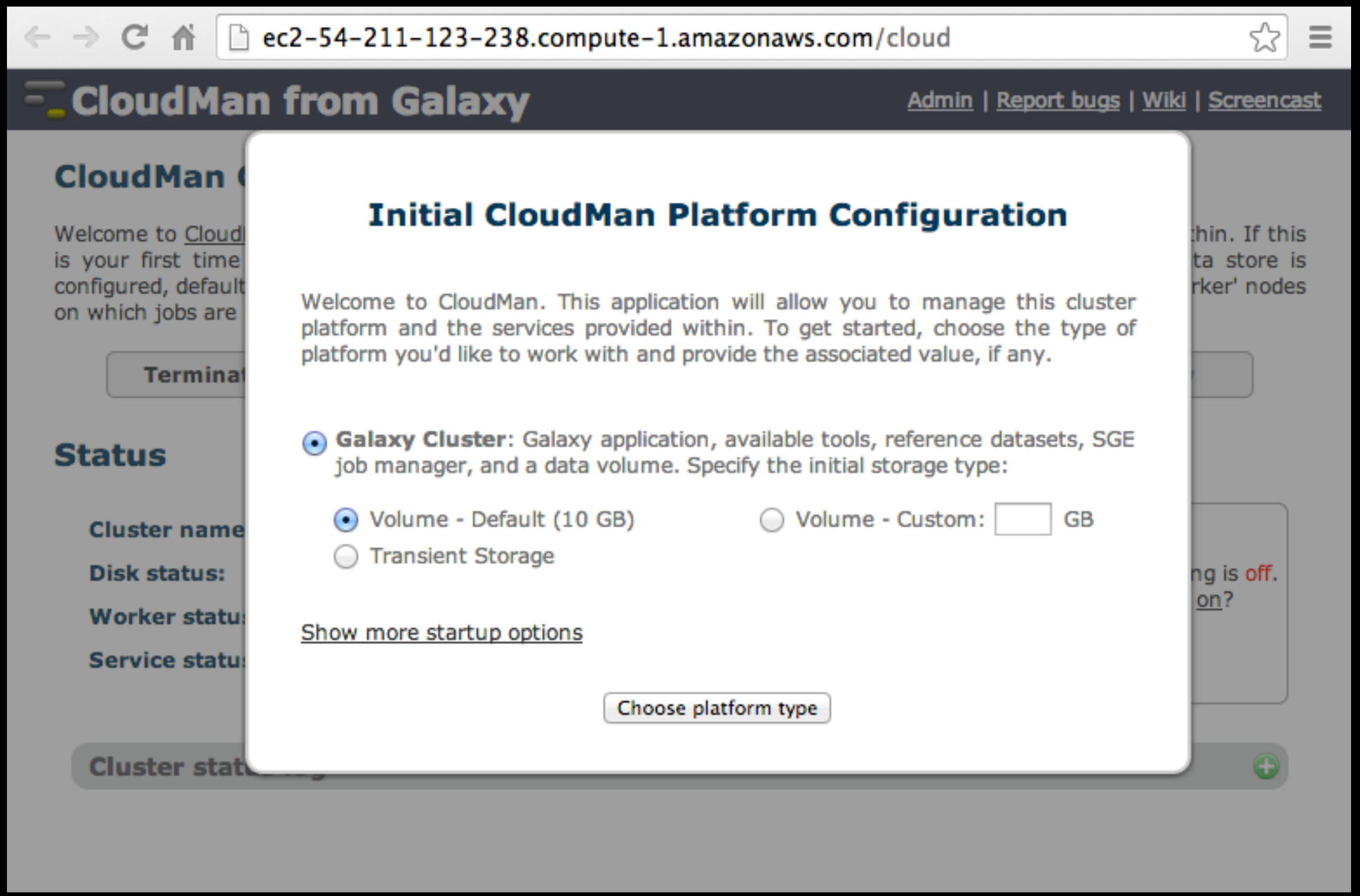
Access Information

- Your instance 'i-94eb82b4' has been successfully launched using the 'ami-a7dbf6ce' AMI.
- While it may take a few moments to boot, you will be able to access the cloud control panel at ec2-54-211-123-238.compute-1.amazonaws.com/cloud.
- SSH access is also available using your private key. From the terminal, you would execute something like:
  `ssh -i CloudManKP1.pem ubuntu@ec2-54-211-123-238.compute-1.amazonaws.com`

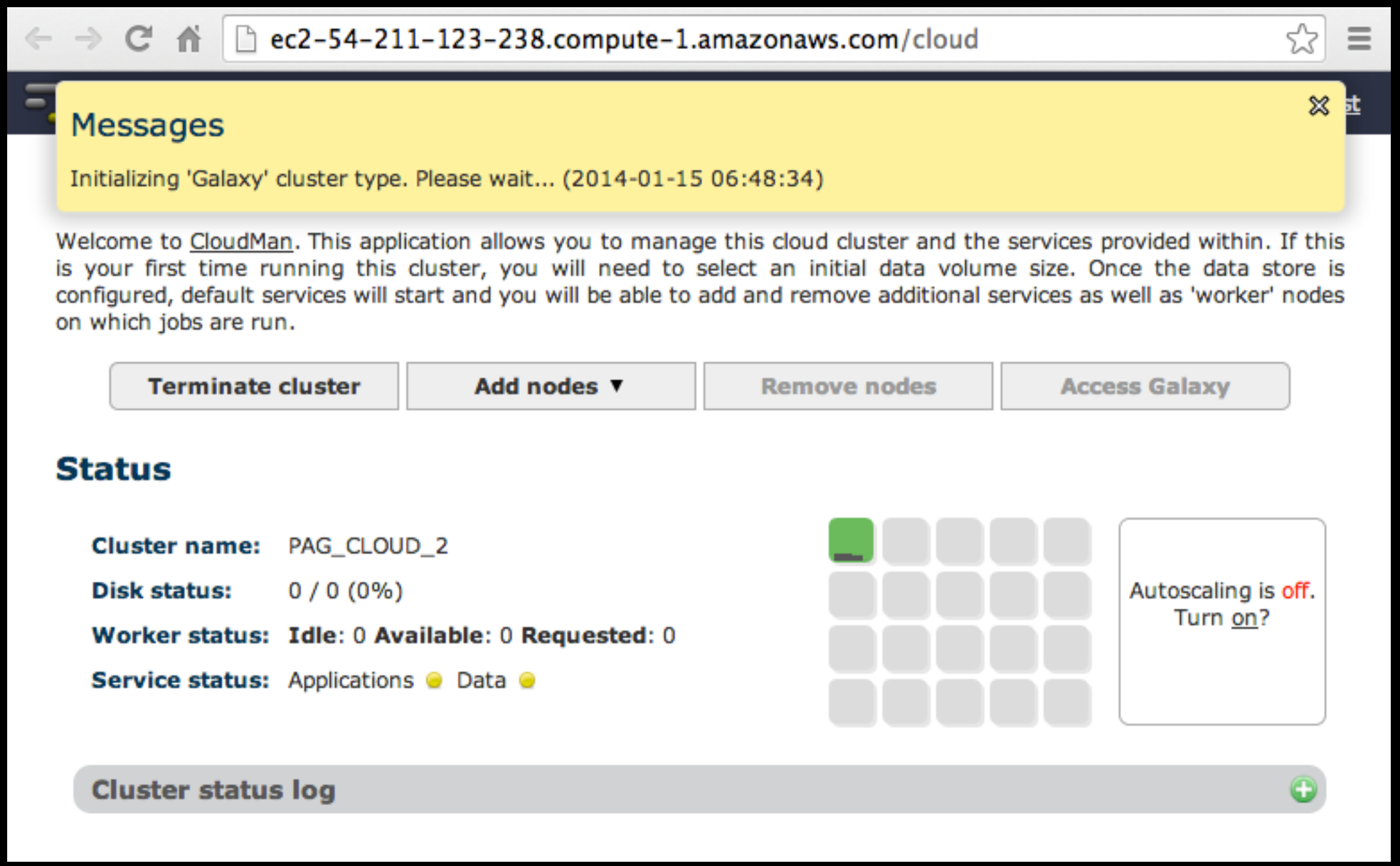
CloudLaunch



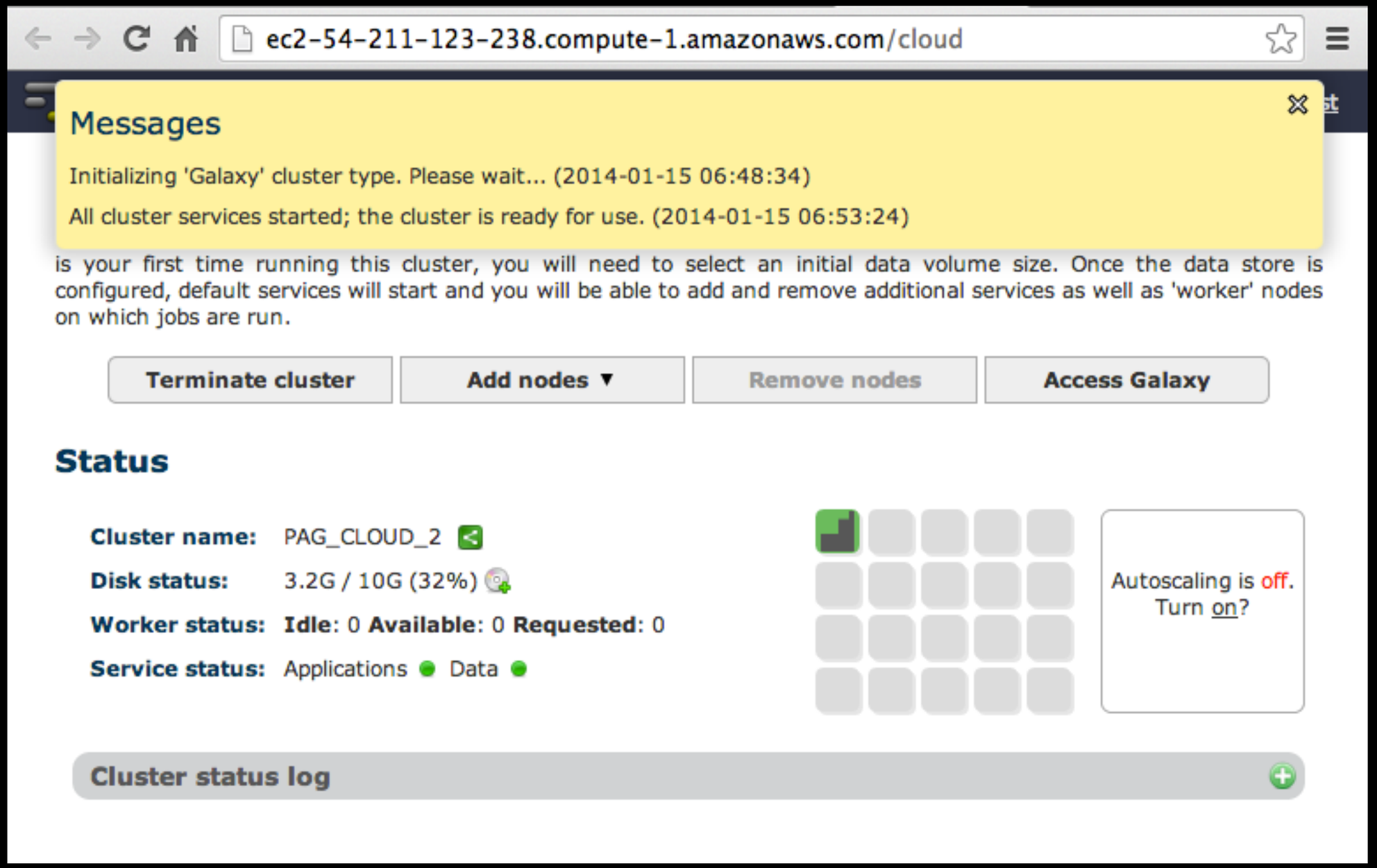
CloudLaunch



CloudLaunch



Cloud Launched



Cool things to do

- Create a login
- Become an admin
- Set up autoscaling
- Run Galaxy 101 (with Chicken chr1!)
- Shut it down

Community Resources

Galaxy Resources and Community: Mailing Lists

http://wiki.galaxyproject.org/MailingLists

Galaxy-Announce

Project announcements, low volume, moderated

Low volume (47 posts in 2013, 3400+ members)

Galaxy-User

Questions about using Galaxy and usegalaxy.org

High volume (1328 posts in 2013, 2600+ members)

Galaxy-Dev

Questions about developing for and deploying Galaxy

High volume (5200 posts in 2013, 900+ members)

Community: Public Galaxy Instances

http://bit.ly/gxyServers

ChIP-chip and ChIP-seq?

✓ Cistrome

Statistical Analysis?

✓ Genomic Hyperbrowser

Protein synthesis?

✓ GWIPS-viz

de novo assembly?

✓ CBIIT Galaxy

Reasoning with ontologies?

✓ OPPL Galaxy

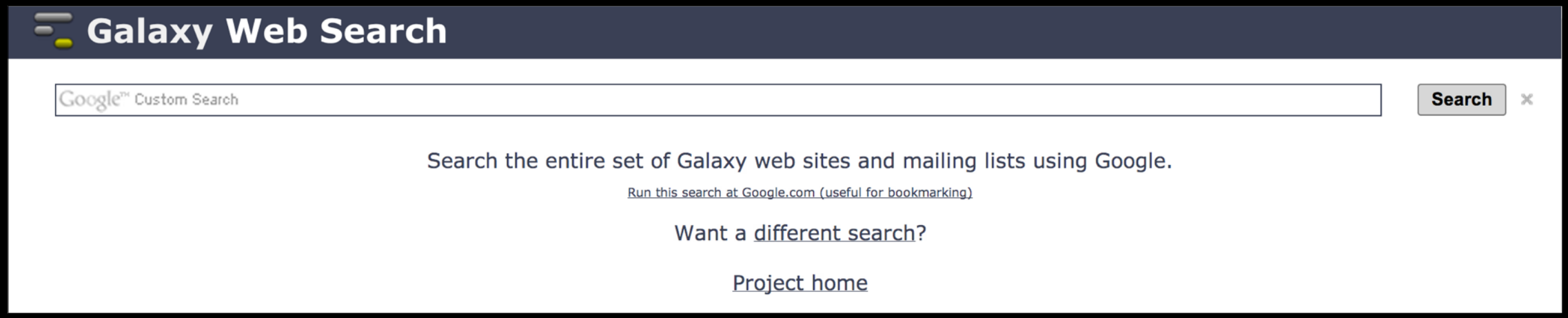
Repeats!

✓ RepeatExplorer

Everything?

✓ Andromeda

Unified Search: http://galaxyproject.org/search



Find

Everything on …

Tools for …

Email about …

Source code for …

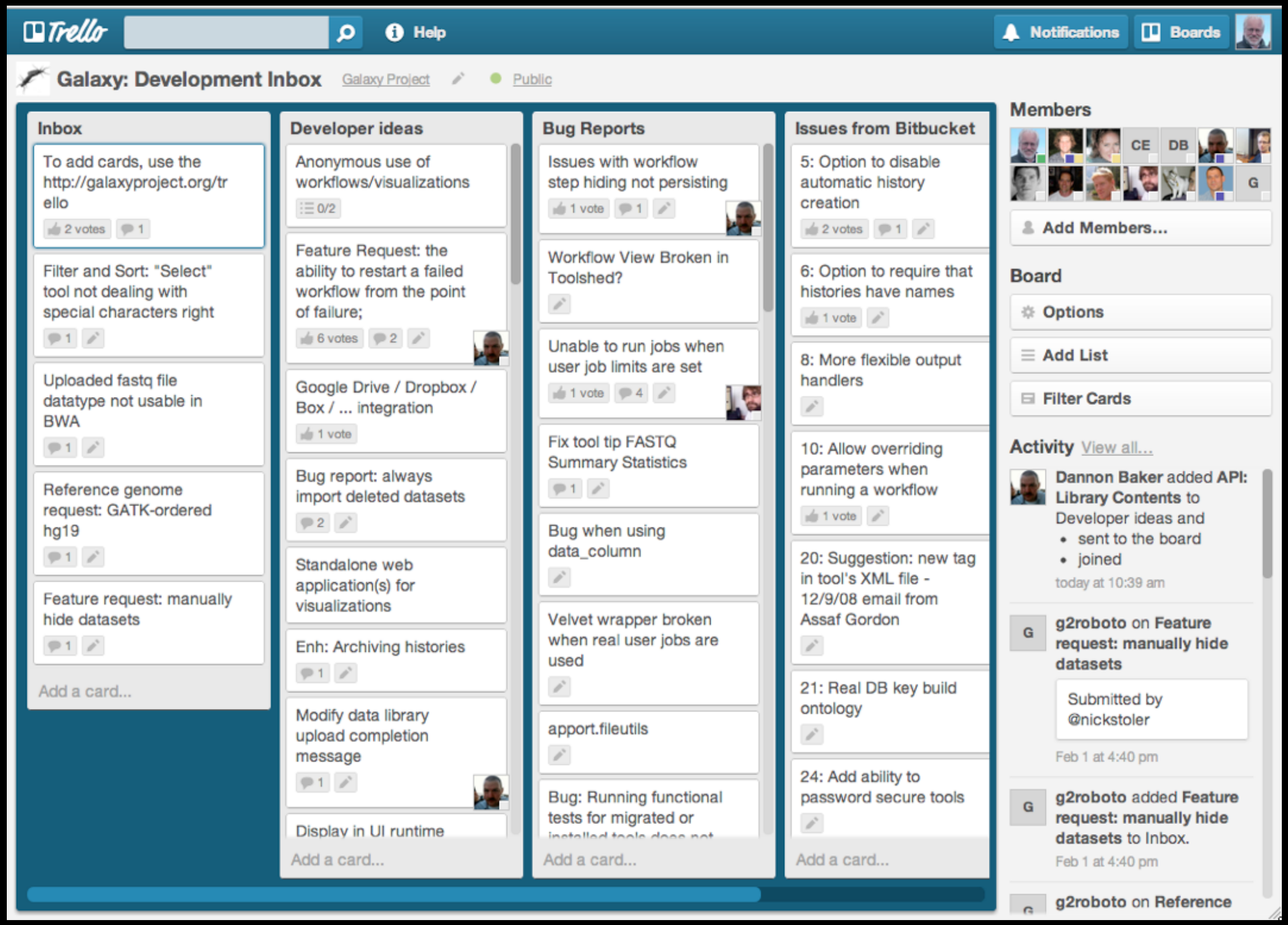
Published Histories, Pages, Workflows, about …

Documentation on …

Papers using Galaxy for …

Related feature requests

Community can create, vote and comment on issues



http://bit.ly/gxyissues

http://wiki.galaxyproject.org

Galaxy Wiki

Login | Search: Titles Text

FrontPage

Locked History Actions

Galaxy

Galaxy is an open, web-based platform for *accessible*, *reproducible*, and *transparent* computational biomedical research.

- **Accessible:** Users without programming experience can easily specify parameters and run tools and workflows.
- **Reproducible:** Galaxy captures information so that any user can repeat and understand a complete computational analysis.
- **Transparent:** Users share and publish analyses via the web and create Pages, interactive, web-based documents that describe a complete analysis.

This is the Galaxy Community Wiki. It describes all things Galaxy.

Use Galaxy

Galaxy's public service web site makes analysis tools, genomic data, tutorial demonstrations, persistent workspaces, and publication services available to any scientist. Extensive user documentation (applicable to any public or local Galaxy instance) is available on this wiki and elsewhere.

usegalaxy.org

Deploy Galaxy

Galaxy is open source for all organizations. Local Galaxy servers can be set up by downloading and customizing the Galaxy application.

- Admin
- Cloud
- Galaxy Appliance

getgalaxy.org

Community & Project

Galaxy has a large and active user community and many ways to Get Involved.

- Community
- News
- Events
- Support

Contribute

- **Users:** Share your histories, workflows, visualizations, data libraries, and Galaxy Pages, enabling others to use and learn from them.
- **Deployers and Developers:** Contribute tool definitions to the Galaxy Tool Shed (making it easy for others to use those tools on their installations), and code to the core release.

Galaxy @ PAG/GMOD

GALAXY COMMUNITY CONFERENCE
BALTIMORE, MD | JUNE 30 - JULY 2, 2014

Training Day voting closes Jan 17

Use Galaxy

Servers • Learn
Main • Share • Search

Communicate

Support • News
Events • Twitter
Mailing Lists (search)

Deploy Galaxy

Get Galaxy • Cloud
Admin • Tool Config
Tool Shed • Search

SLIPSTREAM APPLIANCE Galaxy made easy.

Contribute

Tool Shed • Share
Issues & Requests

Events

Galaxy Wiki Login | Search:

Events

Galaxy Event Horizon

Events with Galaxy-related content are listed here.

Also see the Galaxy Events Google Calendar for a listing of events and deadlines that are relevant to the Galaxy Community. This is also available as an RSS feed.

If you know of any event that should be added to this page and/or to the Galaxy Event Calendar, please add it here or send it to outreach@glaxyproject.org .

Contents

1. Upcoming Eve
2. Other Calenda
3. Past Events
    1. 2014
    2. Archive

Upcoming Events

Date	Topic/Event	Venue/Location	Contact
January 11-15	*Galaxy for NGS Data Analysis: A Hands-on Computer Demo*	**Plant and Animal Genome XXII (PAG 2014)**, San Diego, California, United States	Dave Clements, Anushka Brownley
	Galaxy Cloudman: A Gentle Introduction to Data Analysis on the Cloud Part of the GMOD Workshop		Dave Clements, Scott Cain
	Plus 3 more talks and 4 posters		See list
January 16-17	2014 GMOD Meeting	San Diego, California, United States	Dave Clements, Scott Cain
February 5-6	*Mosquito Informatics*	EBI, Hinxton, United Kingdom	Dan Lawson <lawson AT ebi DOT ac DOT uk>

News

Galaxy Wiki Login | Search:

News

News

Announcements of interest to the Galaxy Community. These can include items from the Galaxy Team or the Galaxy community and can address anything that is of wide interest to the community.

The Galaxy News is also available as an RSS feed.

See Add a News Item below for how to get an item on this page, and the RSS feed. Older news items are available in the Galaxy News Archive.

See also

- Galaxy News Briefs
- Galaxy Updates
- Galaxy on Twitter
- Events
- Learn
- Support
- About the Galaxy Project

News Items

January 2014 CloudMan Release
GCC2014 Training Day Topics: Vote!
January 2014 Galaxy Update
2013 Galaxy Day Report
Galaxy Community Log Board
Galaxy Deployment Catalog
Nominate 2014 Training Day Topics
December 2013 Galaxy Update
Nov 04, 2013 Galaxy Distribution
November 2013 Galaxy Update
December Bioinformatics Boot Camps
GCC2014: Save These Dates!
Galaxy Day, 4 décembre à Paris

News Archive

News Items

January 2014 CloudMan Release

We just released an update to Galaxy CloudMan. CloudMan offers an easy way to get a personal and completely functional instance of Galaxy in the cloud in just a few minutes, without any manual configuration.

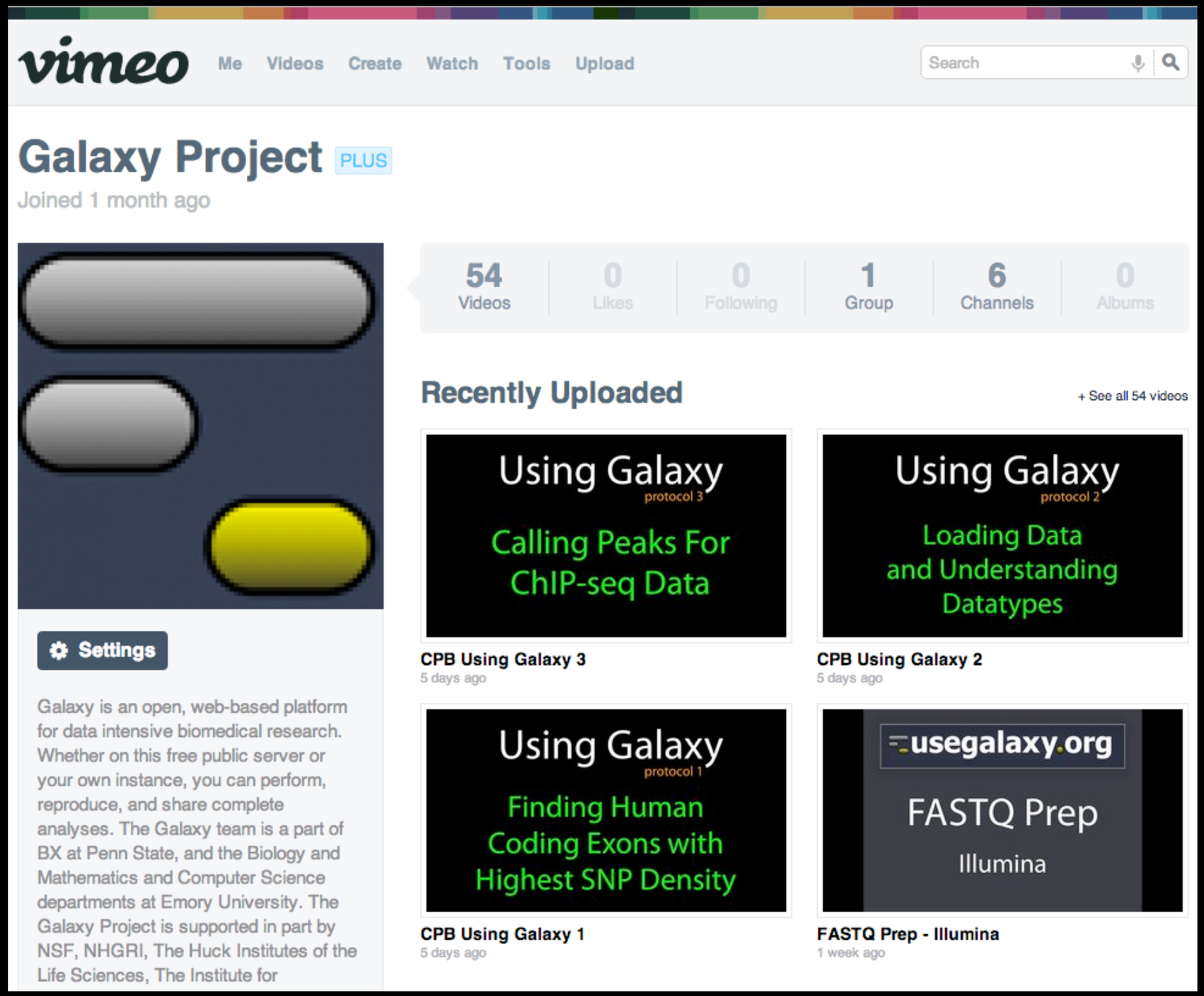
This update brings a large number of updates and new features, the most prominent ones being:

GALAXY COMMUNITY CONFERENCE

BALTIMORE, MD | JUNE 30 - JULY 2, 2014

http://bit.ly/gcc2014

Galaxy Resources & Community: Videos

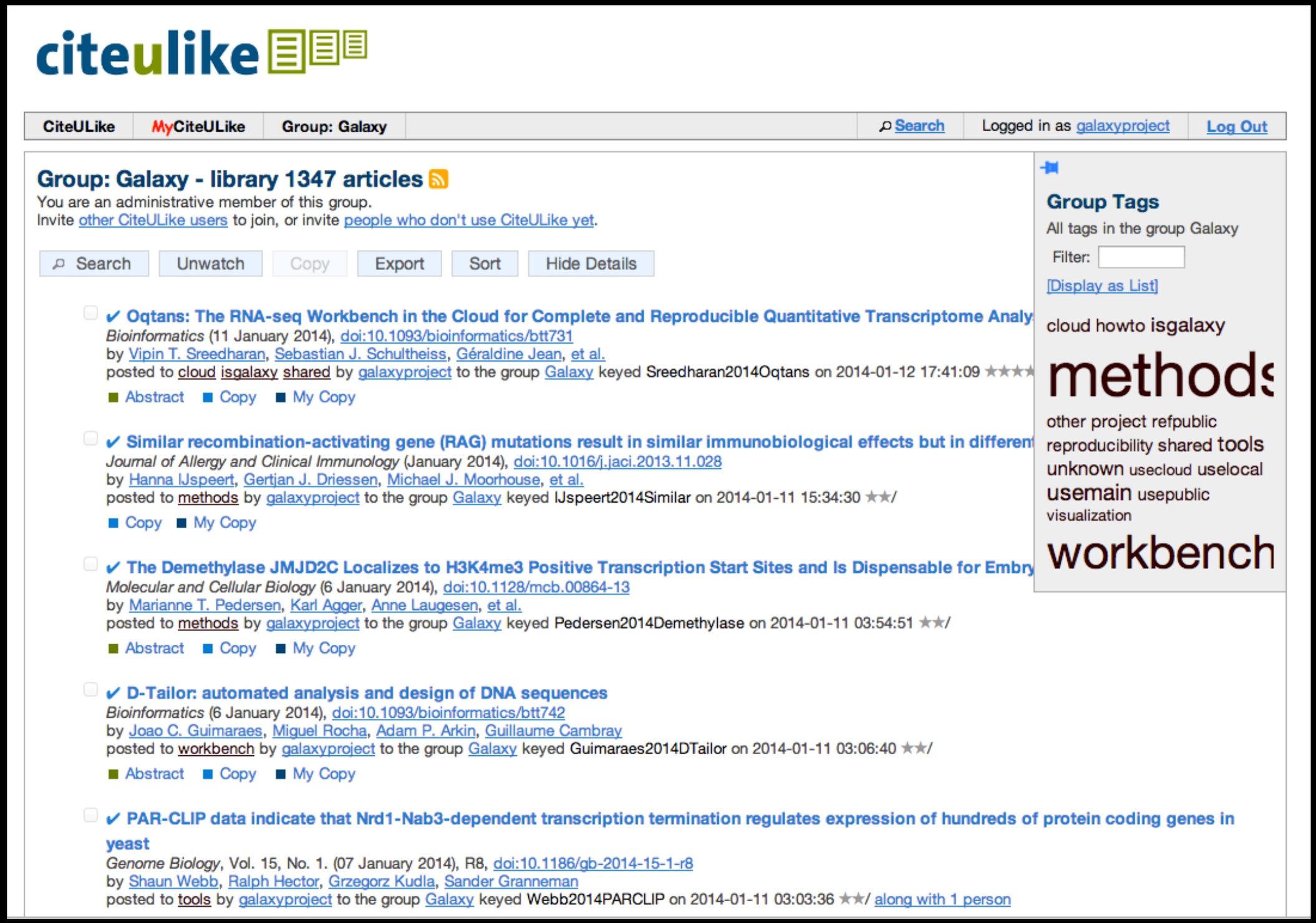


"How to" screencasts on using and deploying Galaxy

Talks from previous meetings.

http://vimeo.com/galaxyproject

Galaxy Resources & Community: CiteULike Group



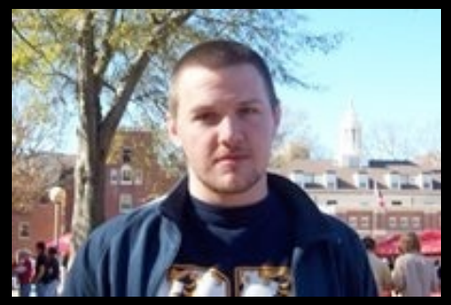
Over 1300 papers

17 different tags

http://bit.ly/gxycul

The Galaxy Team

Enis Afgan



Dannon Baker

Dan Blankenberg

Dave Bouvier

Marten Cech

Dave Clements

Nate Coraor

Carl Eberhard

Dorine Francheteau

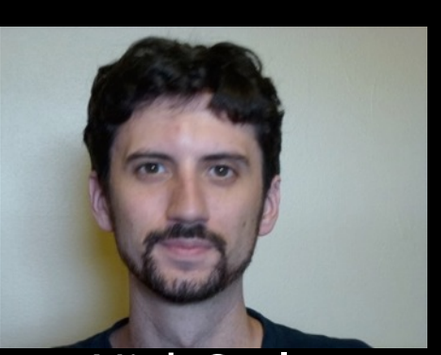
Jeremy Goecks

Sam Guerler

Jen Jackson

Greg von Kuster

Ross Lazarus



Nick Stoler

Anton Nekrutenko

James Taylor

http://wiki.galaxyproject.org/GalaxyTeam

Galaxy is hiring post-docs and software engineers

Please help.

http://wiki.galaxyproject.org/GalaxyIsHiring

Thanks

Dave Clements
Galaxy Project
Johns Hopkins University
clements@galaxyproject.org