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.

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](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
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)
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

Galaxy is an open source, web-based platform for data intensive biomedical research. If you are new to Galaxy [start here](https://usegalaxy.org/root) or consult our [help resources](https://usegalaxy.org/root).

**SLIPSTREAM APPLIANCE**

Galaxy made easy.
CloudLaunch

Galaxy is an open source, web-based platform for data-intensive biomedical research. If you are new to Galaxy start here or consult our help resources.
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 prevent unauthorized access to your account. Note that using this form to launch computational resources in the Amazon Cloud will result in costs to the account holder.

Key ID

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

Secret Key

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

[Redacted]

**Cluster Password – Confirmation**

[Redacted]

**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.
Launch a Galaxy Cloud Instance

Launch Pending, please be patient.

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

Authentication Required

The server http://ec2-54-211-123-238.compute-1.amazonaws.com/cloud requires a username and password. The server says: CM Administration.

User Name: [blank]
Password: [blank]

[Cancel] [Log In]
Welcome to CloudMan. This application will allow you to manage this cluster platform and the services provided within. To get started, choose the type of platform you'd like to work with and provide the associated value, if any.

- **Galaxy Cluster**: Galaxy application, available tools, reference datasets, SGE job manager, and a data volume. Specify the initial storage type:
  - **Volume - Default (10 GB)**
  - **Volume - Custom**
  - **Transient Storage**

**Show more startup options**

Choose platform type
Welcome to CloudMan. This application allows you to manage this cloud cluster and the services provided within. If this is your first time running this cluster, you will need to select an initial data volume size. Once the data store is configured, default services will start and you will be able to add and remove additional services as well as 'worker' nodes on which jobs are run.

**Status**

- **Cluster name:** PAG_CLOUD_2
- **Disk status:** 0 / 0 (0%)
- **Worker status:** Idle: 0 Available: 0 Requested: 0
- **Service status:** Applications • Data •

**Messages**

Initializing 'Galaxy' cluster type. Please wait... (2014-01-15 06:48:34)
Cloud Launched

Messages
Initializing 'Galaxy' cluster type. Please wait... (2014-01-15 06:48:34)
All cluster services started; the cluster is ready for use. (2014-01-15 06:53:24)

is your first time running this cluster, you will need to select an initial data volume size. Once the data store is configured, default services will start and you will be able to add and remove additional services as well as 'worker' nodes on which jobs are run.

Status
Cluster name: PAG_CLOUD_2
Disk status: 3.2G / 10G (32%)
Worker status: Idle: 0 Available: 0 Requested: 0
Service status: Applications ● Data ●

Cluster status log

Autoscaling is off. Turn on?
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


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 ...
- Related feature requests
- Papers using Galaxy for ...
- Documentation on ...
Community can create, vote and comment on issues

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.

**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

**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.
**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@galaxypoint.org.

**Upcoming Events**

<table>
<thead>
<tr>
<th>Date</th>
<th>Topic/Event</th>
<th>Venue/Location</th>
<th>Contact</th>
</tr>
</thead>
<tbody>
<tr>
<td>January 11-15</td>
<td>Galaxy for NGS Data Analysis: A Hands-on Computer Demo</td>
<td>Plant and Animal Genome XXII (PAG 2014), San Diego, California, United States</td>
<td>Dave Clements, Anushka Brownley</td>
</tr>
<tr>
<td></td>
<td>Galaxy Cloudman: A Gentle Introduction to Data Analysis on the Cloud</td>
<td>Part of the GMOD Workshop</td>
<td>Dave Clements, Scott Cain</td>
</tr>
<tr>
<td>January 16-17</td>
<td>2014 GMOD Meeting</td>
<td>San Diego, California, United States</td>
<td>Dave Clements, Scott Cain</td>
</tr>
<tr>
<td>February 5-6</td>
<td>Mosquito Informatics</td>
<td>EBI, Hinxton, United Kingdom</td>
<td>Dan Lawson &lt;lawson AT ebi DOT ac DOT uk&gt;</td>
</tr>
</tbody>
</table>

**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**

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

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

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