An Educational Framework for Galaxy Administration

Martin Čech, Enis Afgan, Nate Coraor, Simon Gladman, Daniel Blankenberg, Björn Grüning, Dave Clements
Outline

- Galaxy admin training
- Events
- Topics
- Infrastructure and Materials
- Takeaways
- Upcoming training
Galaxy Admin Training

Past multi-day Galaxy training events focused on:

- Galaxy users
- Tool developers

Many popular *individual training sessions* at past (and present) GCCs on admin topics but never a focused training event.
Galaxy Admin Training

- Galaxy Admins are big community
- Galaxy is a complex piece of software with many moving parts and interacting pieces
- Setting up a Galaxy server is easy, setting up a fancy Galaxy server is not

“Our biggest unmet training need” - Dave Clements
Admin Training Events
1st Galaxy Admin Training

November 7-11, 2016
Salt Lake City, Utah, USA

- 2 days of **Basic** sessions
- 3 days of **Advanced** sessions

University of Utah Center for High Performance Computing (CHPC)
USTAR Center for Genetic Discovery (UCGD) @ University of Utah
Department of Biomedical Informatics @ University of Utah
Clinical & Translational Science Biomedical Informatics Core (CCTS BMIC) @ University of Utah
SLC: Venues

- SLC Public Municipal Library
  - Basic sessions
  - Network issues
SLC: Venues

- Commander’s house
  - Advanced sessions
  - University of Utah
  - Great network
Salt Lake City: Students and Instructors

33 Students:
70% from USA, plus attendees from Belgium, Canada, France, Germany, Japan, Korea, Norway, South Africa, UK

4 Instructors:
• Dan Blankenberg, Martin Čech, Nate Coraor - Galaxy Team, Penn State University
• Simon Gladman - VLSCI, University of Melbourne, Australia

1 Dave Clements, without whom no events are possible
2nd Galaxy Admin Training

February 6-9, 2017
Melbourne, Australia

- 4 days of Integrated sessions

Victorian Life Sciences Computation Initiative, University of Melbourne (VLSCI)
Carlton Connect Initiative, University of Melbourne
Melbourne: Venue

- Lab 14 seminar space
  - University of Melbourne
  - Close to lots of good food & coffee
Melbourne: Students and Instructors

9 Students:
From Australia

5 Instructors:
- Björn Grüning - University of Freiburg
- Enis Afgan, Nate Coraor - Galaxy Team, Johns Hopkins and Penn State Universities
- Simon Gladman - VLSCI, University of Melbourne, Australia
- Ross Lazarus - Galaxy Team Emeritus
Topics - Basic Sessions

- Get a basic Galaxy server
- Apache vs nginx
- Postgres vs SQLite
- Tools and data managers
- Extensions - mail, ftp, quotas
- Troubleshooting
- Intro to Galaxy Architecture
Topics - Advanced Sessions

- Ansible - get advanced server running
- Monitoring and maintenance
- Control with Systemd and Supervisor
- More on tools
- Authentication/verification
- Heterogeneous compute resources
- Cloud resources
- Advanced troubleshooting
Accelerated Advanced Admin Training

GCC 2017 Training
2.5 Hours
Used materials developed in Salt Lake and Melbourne courses
Admin Training Feedback (from Salt Lake City)

This was fantastic. It was a huge help toward the goals that we're trying to accomplish in our group and we were grateful that the training was offered. Kudos.

Huge THANK YOU

I learned a lot!

This training session was exactly what I needed as far as skills--training

It was very interesting and at a perfect level for me. I learned a lot that I will be able to implement very soon.

Extremely happy with the knowledge transfer.
Training Infrastructure
Training Infrastructure

- Utah: Virtual machines from the University of Utah
- Melbourne: Cloud instances from the Nectar Cloud
- Montpellier: Cloud instances from:
  - CLIMB: Cloud Infrastructure for Microbial Bioinformatics in UK
  - Nectar: Research cloud in Australia
  - Jetstream: NSF research cloud in the US
## Virtual Machines

<table>
<thead>
<tr>
<th>Name</th>
<th>Machine</th>
<th>Image</th>
<th>IP Address</th>
<th>Size</th>
<th>Location</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nate</td>
<td>GAT-1</td>
<td>Ubuntu 16.04 LTS</td>
<td>115.146.91.55</td>
<td>m2.medium</td>
<td>melbourne-qh2</td>
<td>SSH into the machine</td>
</tr>
<tr>
<td>Björn</td>
<td>GAT-2</td>
<td>Ubuntu 16.04 LTS</td>
<td>115.146.91.6</td>
<td>m2.medium</td>
<td>melbourne-qh2</td>
<td></td>
</tr>
<tr>
<td>Enis</td>
<td>GAT-3</td>
<td>Ubuntu 16.04 LTS</td>
<td>115.146.91.62</td>
<td>m2.medium</td>
<td>melbourne-qh2</td>
<td>With: ssh ubuntu@&lt;IP-address&gt;</td>
</tr>
<tr>
<td>Ross</td>
<td>GAT-4</td>
<td>Ubuntu 16.04 LTS</td>
<td>115.146.89.9</td>
<td>m2.medium</td>
<td>melbourne-qh2</td>
<td>Password is on the whiteboard!</td>
</tr>
<tr>
<td>Simon</td>
<td>GAT-5</td>
<td>Ubuntu 16.04 LTS</td>
<td>130.56.254.135</td>
<td>m2.medium</td>
<td>NCI</td>
<td></td>
</tr>
<tr>
<td></td>
<td>GAT-6</td>
<td>Ubuntu 16.04 LTS</td>
<td>115.146.89.94</td>
<td>m2.medium</td>
<td>melbourne-qh2</td>
<td></td>
</tr>
<tr>
<td></td>
<td>GAT-7</td>
<td>Ubuntu 16.04 LTS</td>
<td>115.146.90.0</td>
<td>m2.medium</td>
<td>melbourne-qh2</td>
<td></td>
</tr>
<tr>
<td></td>
<td>GAT-8</td>
<td>Ubuntu 16.04 LTS</td>
<td>115.146.90.124</td>
<td>m2.medium</td>
<td>melbourne-qh2</td>
<td></td>
</tr>
<tr>
<td></td>
<td>GAT-9</td>
<td>Ubuntu 16.04 LTS</td>
<td>115.146.90.128</td>
<td>m2.medium</td>
<td>melbourne-qh2</td>
<td></td>
</tr>
<tr>
<td></td>
<td>GAT-10</td>
<td>Ubuntu 16.04 LTS</td>
<td>115.146.90.138</td>
<td>m2.medium</td>
<td>melbourne-qh2</td>
<td></td>
</tr>
<tr>
<td></td>
<td>GAT-11</td>
<td>Ubuntu 16.04 LTS</td>
<td>115.146.90.22</td>
<td>m2.medium</td>
<td>melbourne-qh2</td>
<td></td>
</tr>
<tr>
<td></td>
<td>GAT-12</td>
<td>Ubuntu 16.04 LTS</td>
<td>115.146.90.234</td>
<td>m2.medium</td>
<td>melbourne-qh2</td>
<td></td>
</tr>
<tr>
<td></td>
<td>GAT-13</td>
<td>Ubuntu 16.04 LTS</td>
<td>115.146.90.237</td>
<td>m2.medium</td>
<td>melbourne-qh2</td>
<td></td>
</tr>
<tr>
<td></td>
<td>GAT-14</td>
<td>Ubuntu 16.04 LTS</td>
<td>115.146.90.28</td>
<td>m2.medium</td>
<td>melbourne-qh2</td>
<td></td>
</tr>
<tr>
<td></td>
<td>GAT-15</td>
<td>Ubuntu 16.04 LTS</td>
<td>115.146.90.37</td>
<td>m2.medium</td>
<td>melbourne-qh2</td>
<td></td>
</tr>
</tbody>
</table>
Training Materials
Training Materials

Created *dagobah-training* repository in GitHub
Course details and logistics

Galaxy Administrators Course

dagobah - The Solar System for Galaxy Training

"The planet shown in Dagobah, in the Sluis sector, is a world of murky swamps, stinking rivers, and a thick, heavy atmosphere."

Salt Lake City - 7th to 11th November 2016

Event Logistics | Admin Training Home Page

Jump to Monday | Tuesday | Wednesday | Thursday | Friday

Built slides have an index.

Instructors

- Niate Coraor - Galaxy Project, BMB, Penn State University, USA
- Simon Gladman - VLSCI, University of Melbourne, Australia
- Dan Blankenberg - Galaxy Project, BMB, Penn State University, USA
- Martin Čech - Galaxy Project, BMB, Penn State University, USA
- Dave Clements - Galaxy Project, Department of Biology, Johns Hopkins University

Timetable

<table>
<thead>
<tr>
<th>Time</th>
<th>Topic</th>
<th>Links</th>
<th>Instructor</th>
</tr>
</thead>
<tbody>
<tr>
<td>09:15</td>
<td>Check-in</td>
<td></td>
<td>All</td>
</tr>
<tr>
<td>09:30</td>
<td>Welcome and Introduction</td>
<td>Slides</td>
<td>All</td>
</tr>
<tr>
<td>09:45</td>
<td>Deployment and Platform Options</td>
<td>Slides</td>
<td>(N)</td>
</tr>
<tr>
<td>10:15</td>
<td>Get a Basic Galaxy Server Up and Running</td>
<td>Slides</td>
<td>(N + M)</td>
</tr>
<tr>
<td>11:00</td>
<td>Morning Break</td>
<td></td>
<td>(N + M)</td>
</tr>
<tr>
<td>11:20</td>
<td>Database choices and configuration. Introduction to PostgreSQL</td>
<td>Slides, Exercise</td>
<td>(N + M)</td>
</tr>
<tr>
<td>12:20</td>
<td>Lunch, catered</td>
<td></td>
<td></td>
</tr>
<tr>
<td>13:20</td>
<td>Web server choices and configuration. Introduction to Apache and NGINX</td>
<td>Slides, Exercise 1, Exercise 2</td>
<td>(N)</td>
</tr>
<tr>
<td>14:20</td>
<td>Introduction to the Galaxy Tool Shed: Identifying and installing well-defined tools</td>
<td>Slides (Shed), Slides (Tools), Slides (Dependencies), Exercise</td>
<td>(M)</td>
</tr>
<tr>
<td>15:30</td>
<td>Afternoon Break</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Benefits

- Quick collaborative editing for instructors
- Participants can contribute to materials via pull request
- Easy reuse for subsequent training
- Slides and Exercises in Markdown
Materials Reuse

Subsequent trainings are *branches* in dagobah
Takeaways
Problems Encountered

Network @ SLC Public Library

- No outbound SSH connections allowed
- IT staff not on site until day 1 starting time
- Tried to add an exception but were unsuccessful
Problems Encountered

Keeping people caught up over a 5 day course

- Ansible playbook to get instances up to day 2 final state

```yaml
become_user: root
  with_items:
    - ppa:webupd8team/java
      name: Update repos
      apt: update_cache=yes
      become: yes
      become_user: root
      - name: Install system packages
        become: yes
        apt: name="{{ item }}" state=latest
        with_items:
```
Problems Encountered

docker pull bgruening/docker-jupyter-notebook

- Ground VMs to a near halt
- Total time to pull for ~30 VMs: over an hour
Lessons Learned

- Have a simple method for getting people caught up
- Pre-pull large images or have participants pull them before they’re needed
- Test everything!
- Use flexible infrastructure
- Have a backup plan
Upcoming Events

Oslo January 2018

A week long training course covering basic and advanced topics

Keep your eyes peeled for more details!