Ansible on the Cloud: A match made in Heaven

ECSS Symposium
02/19/2019
Eric Coulter
Senior XCRI Engineer
What is XCRI?
(XSEDE Cyberinfrastructure Resource Integration)

- Enable seamless transition between local and remote CI Resources
- Software toolkits for local or remote resources
- We also do site visits and remote consultation!
- Spent the last year as an ECSS-GW consultant
The Galaxy Project

- Worked with Enis Afgan and Nate Coraor from the Galaxy team
- Bioinformatics gateway, providing easy access to a whole universe of tools across the US CI Landscape
- Including elastically scaled worker nodes on Jetstream
Galaxy on Jetstream

- https://github.com/galaxyproject/infrastructure-playbook

- Managing all of the Galaxy infrastructure via Ansible!
- Worked to reduce manual intervention in worker nodes
UNAVCO SSARA

- With Scott Baker of UNAVCO
- The Seamless SAR Archive provides seamless access to SAR (Synthetic Aperture Radar) data
- Developed and deployed DC/OS Infrastructure on Jetstream to increase real-time data ingesting and sharing capabilities
UNAVCO SSARA
The 3DQPG Gateway

- With Dr. Murat Maga at Univ. Washington
- Enables processing, analysis, and visualization of 3D images
- Created a Virtual HPC Environment, providing shell environment and an interactive R/shiny server.

3D Quantitative Phenotyping Gateway (3DQPG)

Fig. 1
Visualization of the distances between the atlas surface that was landmarked (p90) and four other surfaces constructed. a 50 % Probability surface (p50); b 70 % Probability surface (p70); c Surface thresholded at grayscale value of 35. d Surface thresholded at grayscale value of 55. RMS: Root mean square error

Ryan Young and A. Murat Maga
https://doi.org/10.1186/s12983-015-0127-8
Jetstream Virtual Clusters

- Hardware-less HPC systems
- Relatively few moving parts
- Build in ~10 minutes with scripts
- Build in half a day by hand (with guidance)
- You are the only group or researcher in the queue!
Customizeable Cyberinfrastructure

- Two levels of custom CI:
  - VM Level
    - Pre-configured VMs
    - User-configured VMs
  - Infrastructure Level
    - Build-your-own networks, security groups, VM selection, etc.
- Powerful not only for science, but also for education
Live Demo:

- Let’s build a virtual cluster!
Hands-on Learning for Admins and Users

- Carrying hardware around is hard! (We’ve tried it)
- Nobody enjoys being lectured at
- HPC Systems are complicated until you have experience with them*
  - *Experience does not nullify complexity, but renders it tractable
Live Demo!

- Back to the headnode!
- Time to install SLURM and create an image for compute nodes
Why Ansible?

- Enables easy use of cloud resources as “cattle” rather than pets
- CLI Client can be scripted, but much more cumbersome to maintain than YAML files
- Easier and safer to back up a playbook than image snapshots alone
- Ansible handles the ugly bits!
Why Ansible?

- YAML is easy to read! (If you generally follow the conventions)
Getting Started with Ansible on Jetstream:

- Basic template available on github: https://github.com/ECoulter/JS_Basic_Template
- Builds the basic Openstack infrastructure (network, router, subnet, security groups, and instances)

```yaml
- name: build instances
  os_server:
    timeout: 300
    state: present
  name: "{{ item.name }}"
  cloud: "{{ cloud_name }}"
  image: "{{ item.image }}"
  key_name: "{{ JS_ssh_keyname }}"
  security_groups: "{{ sec_group_global }}" #,{{ sec_group_internal }}"
  flavor: "{{ item.flavor }}"
  auto_ip: "yes" # for public ip assignment
  network: "{{ network_name }}"
  loop:
    - { name: "example-VM", flavor: "m1.small", image: "{{ default_image }}" }
    - { name: "example-VM2", flavor: "m1.medium", image: "{{ default_image }}" }
```
Live Demo:

• Back to the cluster:
  Time to run some jobs!
Contact us or learn more:

- Basic Ansible template for use with Jetstream: https://github.com/ECoulter/JS_Basic_Template
- Virtual Clusters: https://github.com/XSEDE/CRI_Jetstream_Cluster
- Email: help@xsede.org with XCRI in the subject
In Practice

- Available online: https://github.com/ECoulter/Tutorial_Practice
- Presented at PEARC{17,18}, and Gateways18
- ~50 total students, 4.5/5 satisfaction ratings
- Followup is hard!
  - As is dealing with conference networks…
- Unfortunately, best suited to those with command-line experience – this is a hard problem as well…