Building a virtual research environment with Galaxy

Galaxy DevOps II : Admins Strike Back

Olivier Inizan, Mikael Loaec, Eric Rasche, Sébastien Bridel and Hadi Quesneville.

INRA URG1,
RD 10 route de Saint Cyr
78026 Versailles Cedex
Mikael Loaec

Olivier Inizan
GCC 2013 : DevOps

- Collaboration
- Infrastructure as a code
- Continuous Delivery
DevOps Collaboration Galaxy
DevOps: the other side

DevOps

- Infrastructure as a code
- Continuous Delivery

Accessible + Reproducible
Infrastructure As A Code

- Is it useful to manage an infrastructure supporting a Galaxy server as a software?
- Tools == software
- Tools accessible with:
- At infrastructure level:
  - Virtualization: the hardware becomes software
  - => Not only tools are accessible … but the whole environment analysis
Continuous Delivery

- Is it useful to be able to build a galaxy server « at any time »?
- At any time, maybe not
- But the ability to build the whole environment analysis ...
- or to rebuild it after the first analysis
- is a contribution to reproducibility
- => Automatic Delivery is more accurate
Concepts rock but ...
Virtual Research Environment

Infrastructure as a code

Reproducibility

Continuous Delivery

VRE

Physical or Virtual Layer

DevOps

Accessibility
Virtual Research Environment

VRE
- Data
  - bam, sam, fasta ...
- Applications + Tools
  - Galaxy ...
- Dependencies
  - Postgres, Apache, Nginx, Sge ...
- Operating System
  - Debian, CentOS, Ubuntu, RedHat ...
- Physical or Virtual Layer
  - KVM, OpenVZ, Docker, VMWare ...
Build it!

VRE

Data

Applications + Tools

Dependencies

Operating System

Physical or Virtual Layer

python
Build it!

Automatic installation and configuration

Provisionning with API

VRE
- Data
- Applications + Tools
- Dependencies
- Operating System

Physical or Virtual Layer
Many solutions...

- **VRE**
  - **Data**
  - **Applications + Tools**
  - **Dependencies**
  - **Operating System**
  - **Physical or Virtual Layer**

- **Physical or Virtual Layer**
  - Cobbler, ProxmoxVE, Openstack, Amazon...

- **Operating System**
  - Debian, CentOS, Ubuntu, RedHat...

- **Dependencies**
  - postgres, apache, nginx, sge...

- **Applications + Tools**
  - Galaxy
  - bam, sam, fasta...

- **Data**
  - bam, sam, fasta...

- **Physical or Virtual Layer**
  - KVM, OpenVZ, Docker, VMWare...

- **Puppet, Salt, Ansible, Chef...**

- **Cobbler, ProxmoxVE, Openstack, Amazon...**
In our Lab

VRE

Data

Applications + Tools

Dependencies

Operating System

Physical or Virtual Layer

VRE

Galaxy

postgres, apache, sge

CentOS

KVM

Cobbler & ProxmoxVE

API

bam, sam, fasta ...

Cobbler & ProxmoxVE

API

Physical or Virtual Layer

VRE

Applications + Tools

Dependencies

Operating System

Galaxy

postgres, apache, sge
What is Puppet?

- A declarative language
- Stand alone or client-server mode
- Automate many repetitive tasks
- Package your code and share it on puppet forge

Welcome to the Puppet Forge

A repository of modules written by our community for Puppet Open Source and Puppet Enterprise IT automation software
And now ...

The story of Galaxy and Puppet
Once upon a time in France ...

- Two guys working on puppet module dedicated to Galaxy
- Installation and configuration of a basic instance
- Publication on the puppet forge

@OlivierInizan
#usegalaxy @inra_urgi A puppet module for a very basic galaxy server (use for development): forge.puppetlabs.com/urgi/galaxy

A module for a basic Galaxy installation. Galaxy is an open, web-based platform for accessible, reproducible, and transparent computational biomedical research.

Version 0.0.2 released Mar 20, 2014 | 65 downloads of this version
Meanwhile in the Texas ...

[...] working on some puppet modules for managing galaxy [...] 
[...] figured it would be useful to the community [...]
And a project is born ...

[GitHub repository link: https://github.com/puppet-galaxy]
Project design

- puppet-galaxy
  - Blocks for building a server

- galaxy-roles-profiles
  - Aggregate building blocks in ready to use classes
  - Profile: a technology layer
  - Role: a complete server configuration
  - A role is composed by one or many profiles
Example: role multicore

Scaling and Load Balancing

Contents
1. Web Server(s)
   2. Job Handler(s)
   3. Remaining configuration options
2. Starting and Stopping
3. Proxy Server
   1. Apache
   2. nginx
4. Notes on legacy configurations
Want to break the wall between dev and ops?

Beta Testers: https://forge.puppetlabs.com/
Developpers: https://github.com/puppet-galaxy/
DevOps

VRE
- Data
- Applications + Tools
- Dependencies
- Operating System

Physical or Virtual Layer

BUILD IT!
DevOps

VRE
- Data
- Applications + Tools
- Dependencies
- Operating System

Physical or Virtual Layer

BUILD IT!

SHARE IT!

UPDATE IT!

?
Thank you
Thank you for your attention!

« May the Force be with you! »

Olivier Inizan
Olivier.Inizan@vesailles.inra.fr
Twitter: @OlivierInizan

Mikael Loaec
Mikael.Loaec@versailles.inra.fr