

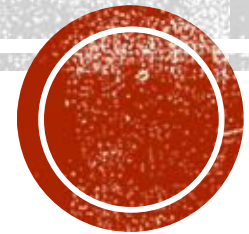
REPRODUCIBLE GALAXY

Improved Development and Administration

Aarif Mohamed Nazeer Batcha, Sebastian Schaaf, Guokun Zhang, Sandra Fischer, Ashok Varadharajan, Ulrich Mansmann

Institute of Medical Informatics, Biometry and Epidemiology

Ludwig-Maximilians-Universität Munich



INTRODUCTION

- Reproducibility

Reproducible research is the distribution of research publication along with data, software source code, and tools required to reproduce the results discussed in the publication. - orgmode.org

- Users :::: Admins ::: Developers



TALK SHOULD HAVE A SUMMARY

History

(Actual) Topic of Interest

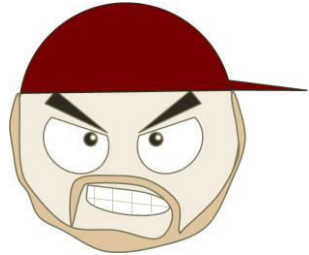
Ansibling Galaxy



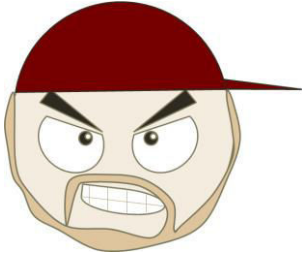
HISTORY - 2011



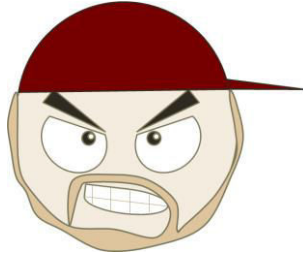
HISTORY - 2011



HISTORY - 2011



HISTORY - 2011



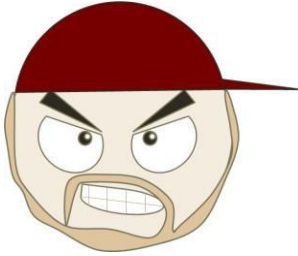
We need
NGS..



HISTORY - 2011



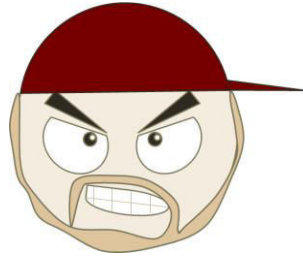
We need
NGS..



I need time..

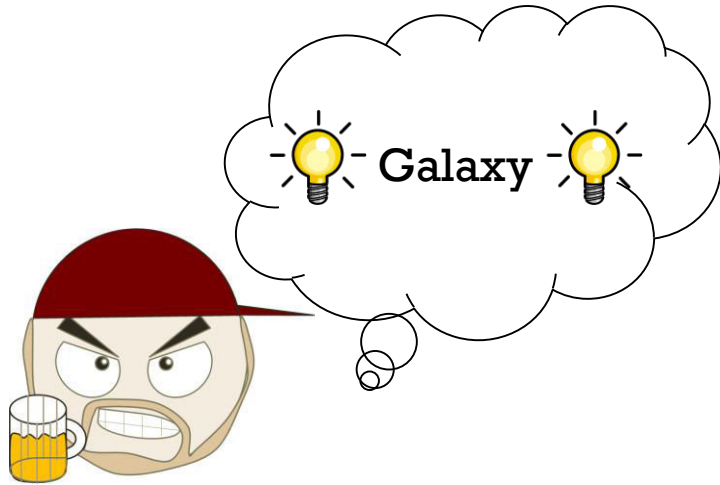


HISTORY - 2011



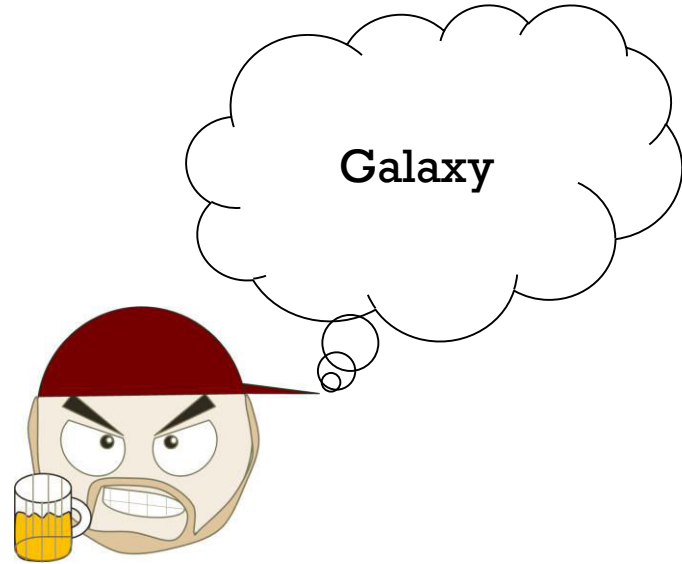
HISTORY - 2011





HISTORY - 2011



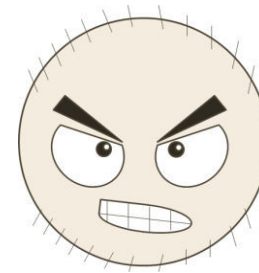


HISTORY - 2012

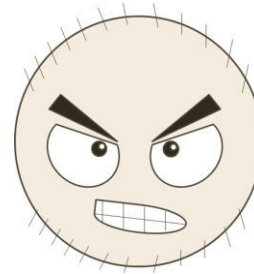




HISTORY - 2013

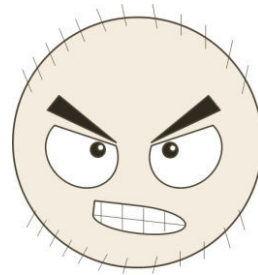


HISTORY - 2013



HISTORY - 2013

Galaxy

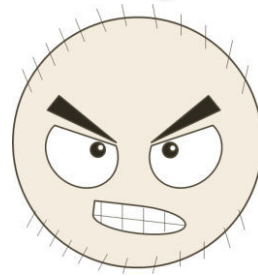


HISTORY - 2013

Galaxy



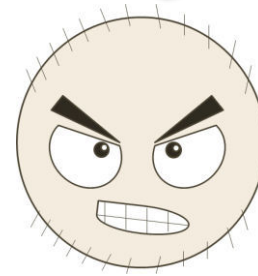
Gaaa...La...Cie...



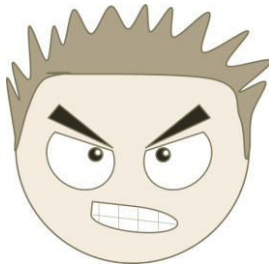
HISTORY - 2014

Galaxy

Gala...xie



Gaa...Lxi

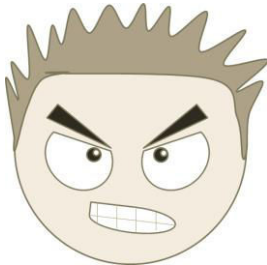
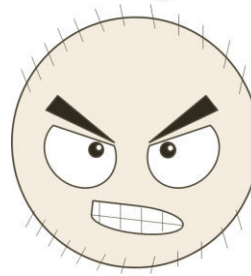


HISTORY - 2014

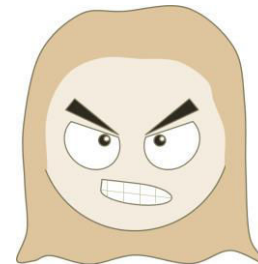
Galaxy

Galaxy

Galaxy



Gaa...Laxi



NOW - 2015

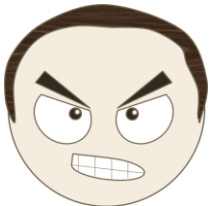
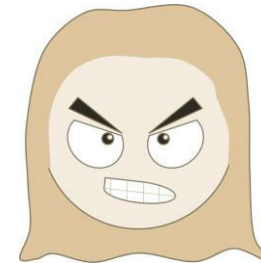
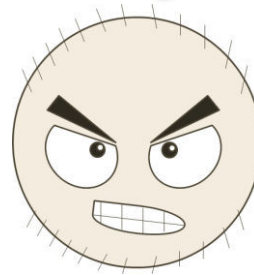
Galaxy

Galaxy

Galaxy

Galaxy

Galaxy



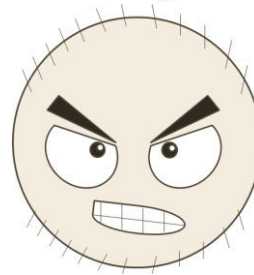
NOW - 2015



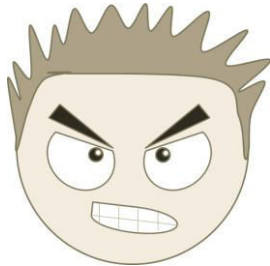
Galaxy

Galaxy

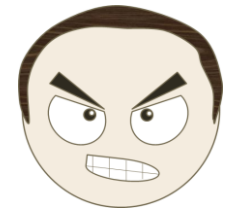
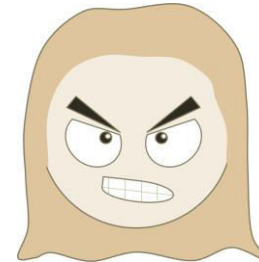
Galaxy



Galaxy



Galaxy



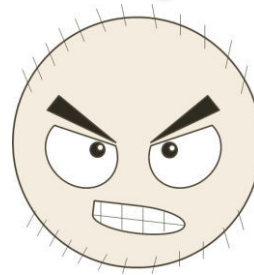
NOW - 2015



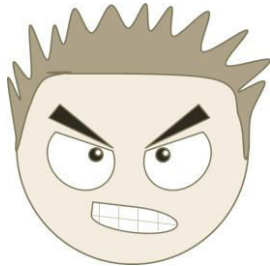
Galaxy

Galaxy

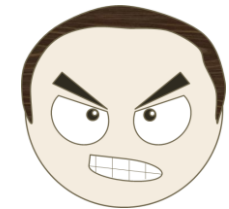
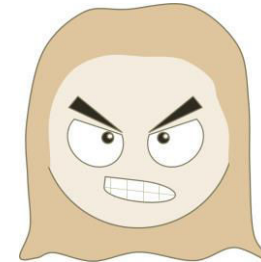
Galaxy



Galaxy



Galaxy



**NGS - Fabrication
Laboratory**



NGS - FABLAB GCC2014

The Munich NGS-FabLab – A glimpse on an IT infrastructure for medical sequence data

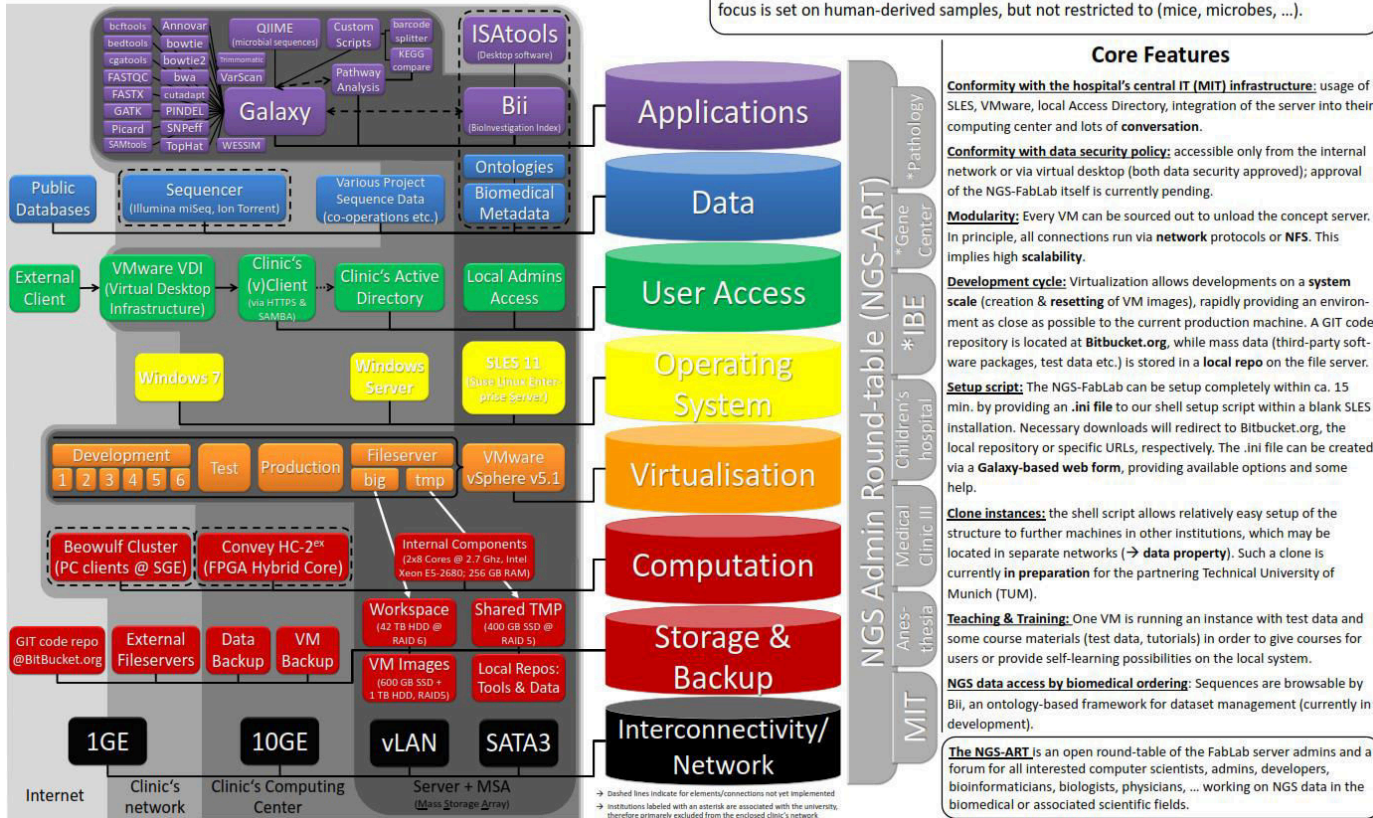
Sebastian Schaaf^{1,2}, Aarif Mohamed Nazeer Batcha², Guokun Zhang², Sandra Fischer²,
Ashok Varadharajan², Ulrich Mansmann^{1,2}

¹ German Cancer Consortium (DKTK), Heidelberg, Germany

² Department of Medical Informatics, Biometry and Epidemiology (IBE), Ludwig Maximilians University (LMU) Munich, Germany



The **NGS-FabLab** is the first multi-institutional IT infrastructure for research at the university hospital and medical faculty of the LMU (which are in fact two juristic persons) in Munich, supporting molecular biomedical and clinical researchers from basic research up to experimental diagnostics and translational studies. Also due to data security laws a focus is set on human-derived samples, but not restricted to (mice, microbes, ...).





- versioning also here difficult
- custom indices: download vs. self-built

→ read the docs carefully

TOPIC OF INTEREST

- Ansible
 - SSH
 - Hostname / ip address
- Ansible Playbooks
 - YML scripts



ANSIBLE

- host inventory
- Eg:

```
$ cat hosts.sh  
[webservers]  
vm1  
vm2  
[examplegroup]  
vm2  
vm4
```



ANSIBLE - PLAYBOOKS

```
---  
- name: install and start apache  
  hosts: webservers  
  user: root  
  
  tasks:  
  
    - name: install httpd  
      yum: name=httpd state=present  
  
    - name: start httpd  
      service: name=httpd state=running
```



ANSIBLE - PLAYBOOKS

```
---
- name: install and start apache
  hosts: webservers
  user: root

  tasks:

    - name: install httpd
      yum: name=httpd state=present

    - name: start httpd
      service: name=httpd state=running
```

```
PLAY [install and start apache] *****

GATHERING FACTS *****
ok: [vm3]
ok: [vm2]

TASK: [install httpd] *****
changed: [vm3]
changed: [vm2]

TASK: [start httpd] *****
changed: [vm2]
changed: [vm3]

PLAY RECAP *****
vm2                : ok=3    changed=2    unreachable=0    failed=0
vm3                : ok=3    changed=2    unreachable=0    failed=0
```



ANSIBLING GALAXY

<https://bitbucket.org/ibe/galaxy-setup-scripts/src>

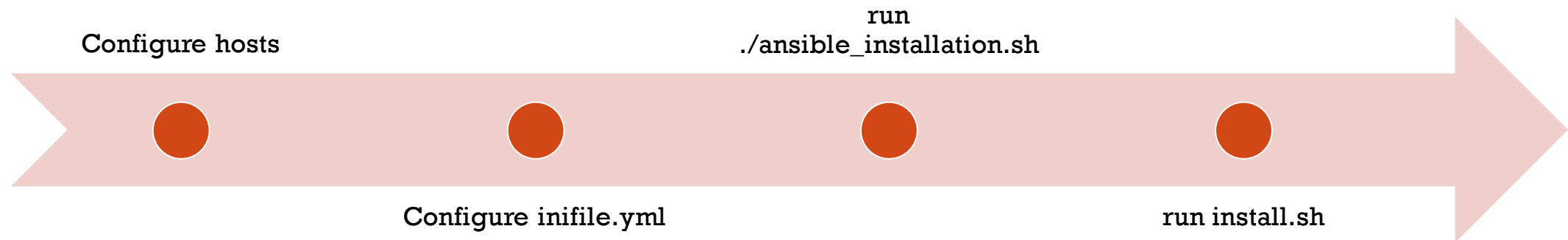
<https://bitbucket.org/ibe/galaxy-setup-scripts.git>



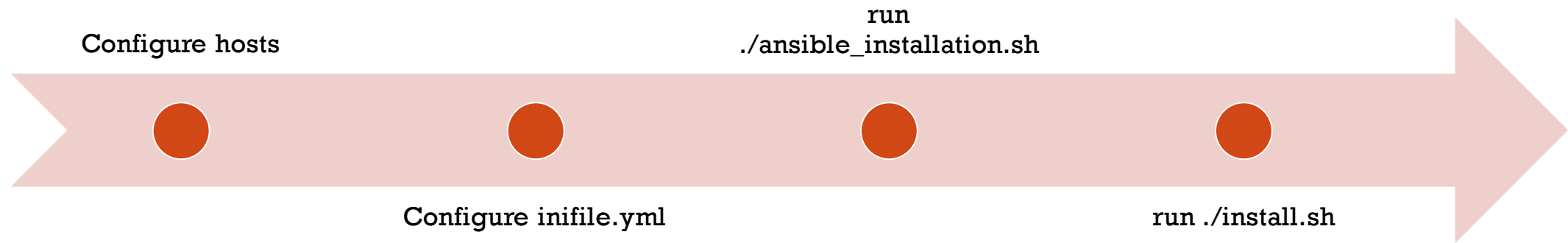
ANSIBLING GALAXY

<https://bitbucket.org/ibe/galaxy-setup-scripts/src>

<https://bitbucket.org/ibe/galaxy-setup-scripts.git>



ANSIBLING GALAXY



```
~/dev_home/BITBUCKET/ngs/playbooks> ./install.sh
1) galaxy-apache2-postgresql-nfs_server_setup-client_connection_setup-grid_engine_setup-galaxy_grid_configuration
2) galaxy-apache2-postgresql-nfs_server_setup-client_connection_setup
3) galaxy-apache2-postgresql-client_connection_setup
4) galaxy-apache2-postgresql
5) grid_engine_setup-galaxy_grid_configuration
6) galaxy
7) apache2
8) postgresql
9) nfs_server_setup
10) client_connection_setup
11) galaxy_grid_configuration
12) exit
#?
```



ANSIBLING GALAXY

Pre-requisites

- Hosts must be accessible via SSH
- Hosts must have identical root password
- Internet connection for downloads



Preliminary Inquiry

- Select installation procedure
- Enter root password for the hosts

Host ENV



- Create root bashrc variables
- Create users and groups



NFS server setup

- NFS kernel installation
- Create mount points for galaxy/grid



PostgreSQL

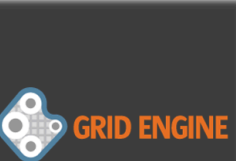
- Install postgres and dependencies
- Configure postgresql for access
- databases and users for the hosts



- Install apache2
- Add modules and configure SSL



- Create galaxy .bashrc variables
- Install Galaxy from GitHub
- Configure galaxy.ini file



Grid engine setup (service and connections)

- Install tool dependencies
- Configure grid engine
- Install master, execution and submission hosts

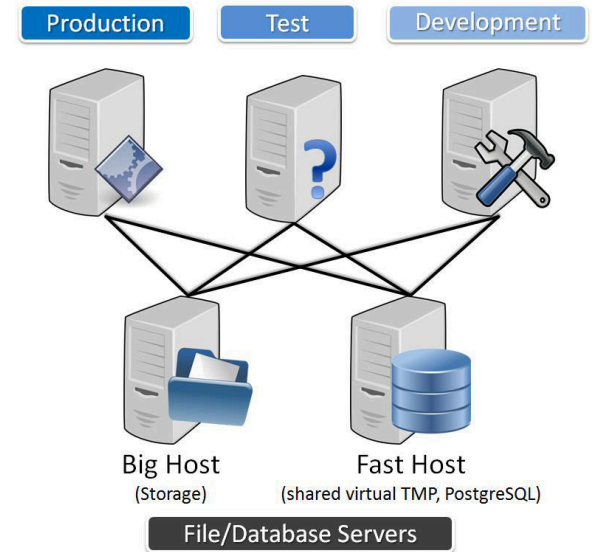
ANSIBLING GALAXY

NFS client connections

- Create directory tree on client side
- Mount necessary folders to remote servers

```
-- home
-- galaxy-user
--   -- downloads
--   -- external_resources
--       -- hostname
--       --   dataset_files
--       --   job_working_directory
--       --   new_file_path
--       -- genomes
--       -- ftp_upload_dir
--   -- external_tools
--   -- galaxy
--   -- homescheme_installation
--   -- shed_tools
--   -- tool_libraries
```

Fig. 3 – Default directory tree (client side); directories underlayed in yellow are configurable as optional NFS mounts.

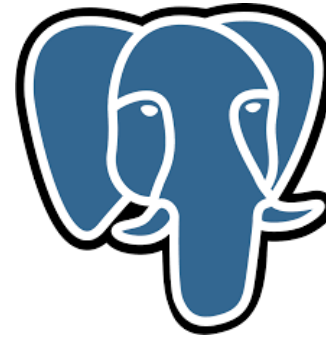


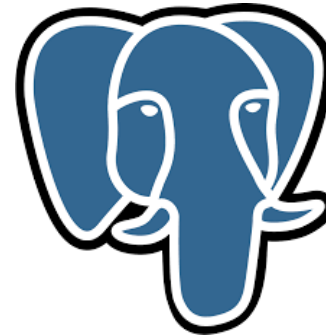
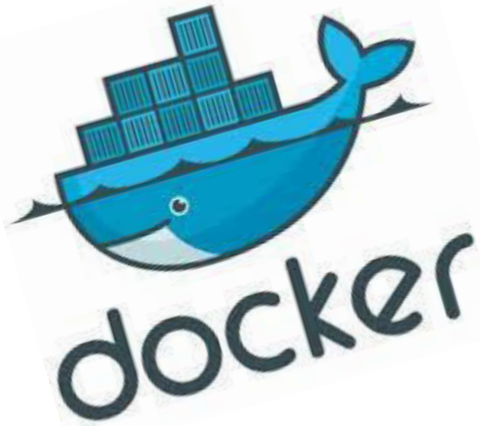
TO BE HONEST

- Minor Bugs ::::: Like every other scripts
- Tested on Ubuntu 14.04 and SLES 11 SP3
- Tried to run it on docker container with Ubuntu (docker pull ubuntu)
 - Couldn't start postgres
 - Rest of the installation worked fine
- Tried to run it on one of the GCC2015 vm images (training monday)
 - Worked fine









Apache

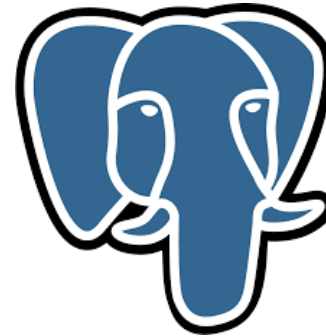
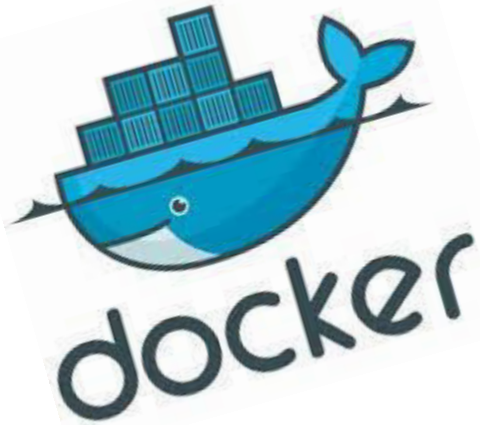


R.T.F.M.

The Way To Success

<https://en.wikipedia.org/wiki/RTFM>





Apache



R.T.F.M.

The Way To Success

Thank you

