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

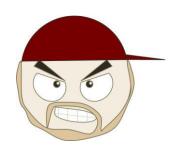
(Actual)Topic of Interest

Ansibling Galaxy



















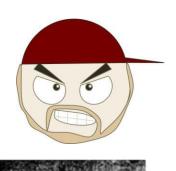


We need NGS..









I need time..





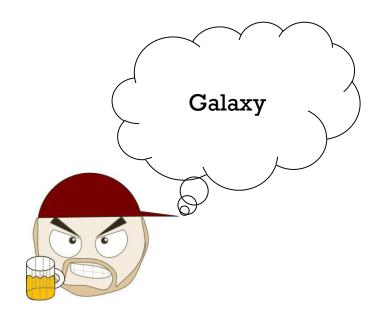






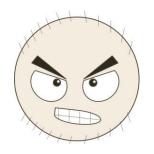






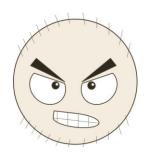








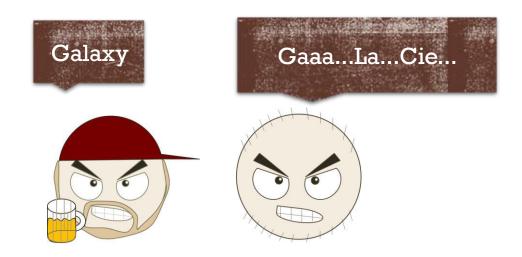








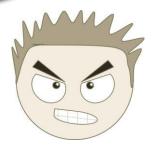




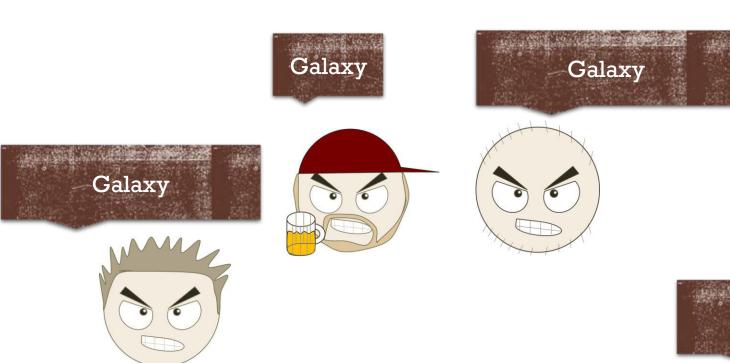










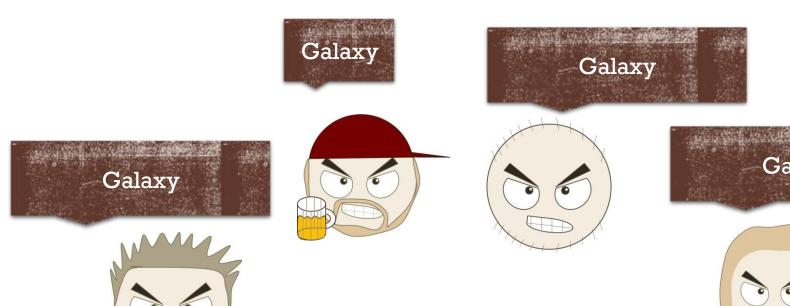








NOW - 2015















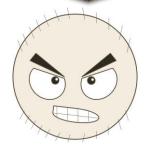
NOW - 2015

Galaxy











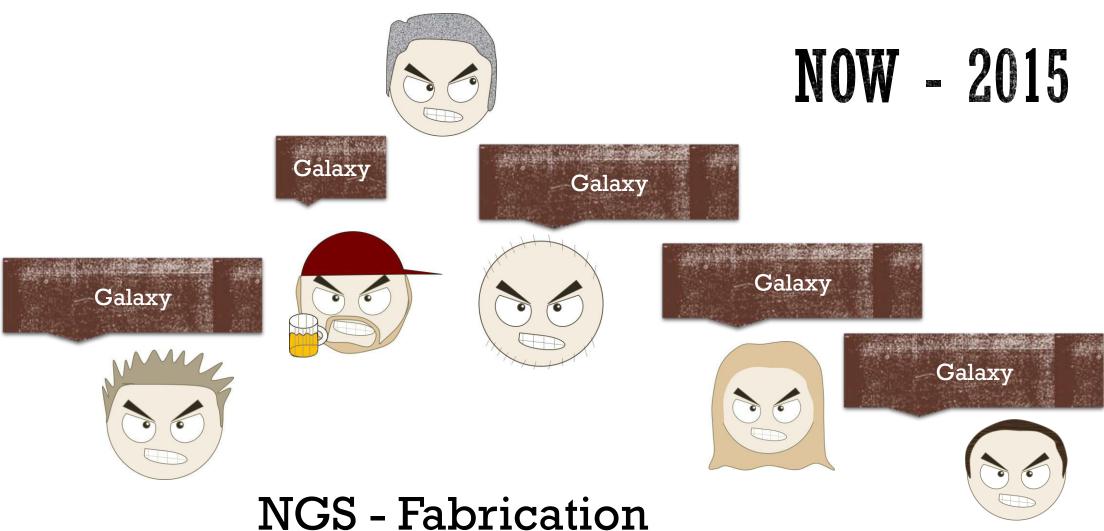








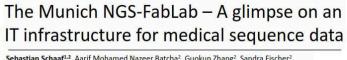




NGS - Fabrication Laboratory



NGS - FABLAB GCC2014





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, ...).

Core Features

Conformity with the hospital's central IT (MIT) infrastructure: usage of SLES, VMware, local Access Directory, integration of the server into their computing center and lots of **conversation**.

Conformity with data security policy: accessible only from the internal network or via virtual desktop (both data security approved); approval of the NGS-FabLab itself is currently pending.

Modularity: Every VM can be sourced out to unload the concept server.

In principle, all connections run via network protocols or NFS. This implies high scalability.

<u>Development cycle</u>: Virtualization allows developments on a **system** scale (creation & **resetting** of VM images), rapidly providing an environment as close as possible to the current production machine. A GIT code repository is located at **Bitbucket.org**, while mass data (third-party software packages, test data etc.) is stored in a **local repo** on the file server.

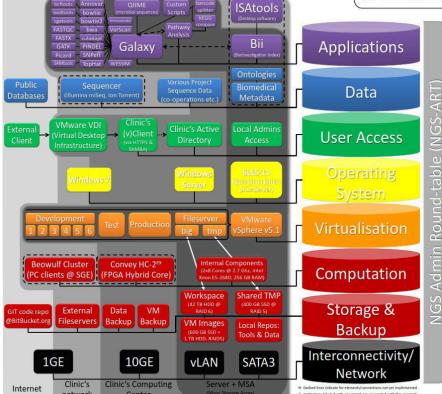
Setup script: The NGS-FabLab can be setup completely within ca. 15 mln. by providing an .ini file to our shell setup script within a blank SLES installation. Necessary downloads will redirect to Bitbucket.org, the local repository or specific URLs, respectively. The .ini file can be created via a Galaxy-based web form, providing available options and some help.

Clone instances: the shell script allows relatively easy setup of the structure to further machines in other institutions, which may be located in separate networks (→ data property). Such a clone is currently in preparation for the partnering Technical University of Munich (TUM).

<u>Teaching & Training:</u> One VM is running an instance with test data and some course materials (test data, tutorials) in order to give courses for users or provide self-learning possibilities on the local system.

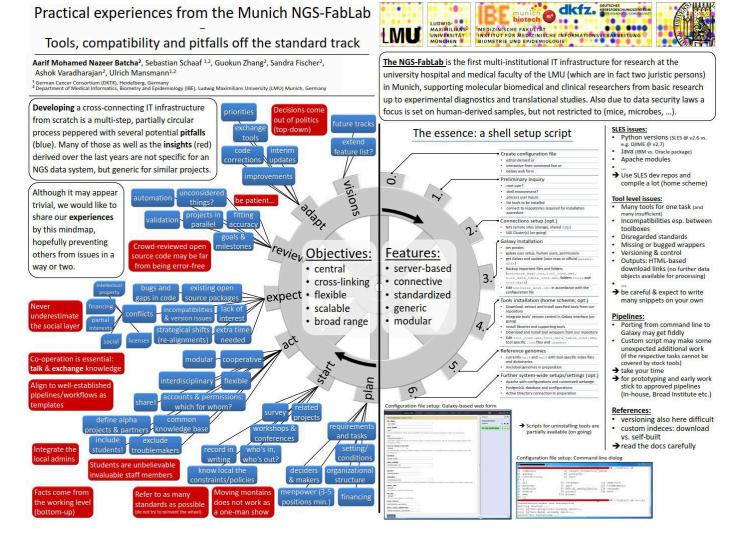
NGS data access by biomedical ordering: Sequences are browsable by Bii, an ontology-based framework for dataset management (currently in development).

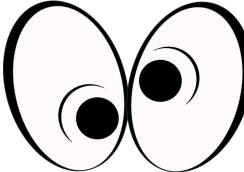
The NGS-ART is an open round-table of the FabLab server admins and a forum for all interested computer scientists, admins, developers, bioinformaticians, biologists, physicians, ... working on NGS data in the biomedical or associated scientific fields.





NGS - FABLAB GCC2014







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] *******
ok: [vm3]
ok: [vm2]
TASK: [install httpd]
changed: [vm3]
changed: [vm2]
TASK: [start httpd]
changed: [vm2]
changed: [vm3]
                                                                     failed=0
                            : ok=3
                                      changed=2
                                                   unreachable=0
                                      changed=2
                                                   unreachable=0
                                                                     failed=0
```



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

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



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

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

Configure hosts

./ansible_installation.sh

Configure inifile.yml

run

run

./ansible_installation.sh



Configure hosts

./ansible_installation.sh

Configure inifile.yml

run

//ansible_installation.sh

```
:~/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
```

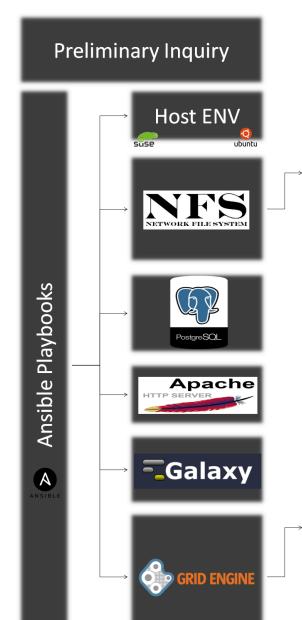
12) exit #?



Pre-requisites

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





- Select installation procedure
- Enter root password for the hosts
- Create root bashrc variables
- Create users and groups

NFS server setup

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

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.

Big Host (Storage) (Shared virtual TMP, PostgreSQL) File/Database Servers

Grid engine setup (service and connections)

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

For Galaxy submission host:

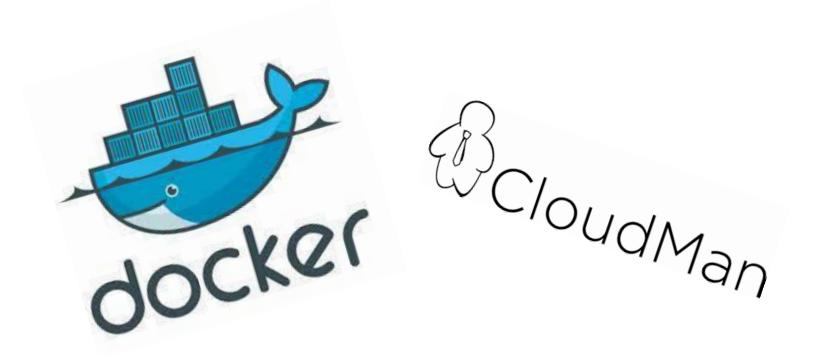
- Create job configuration file
- Edit galaxy.ini file for grid configuration

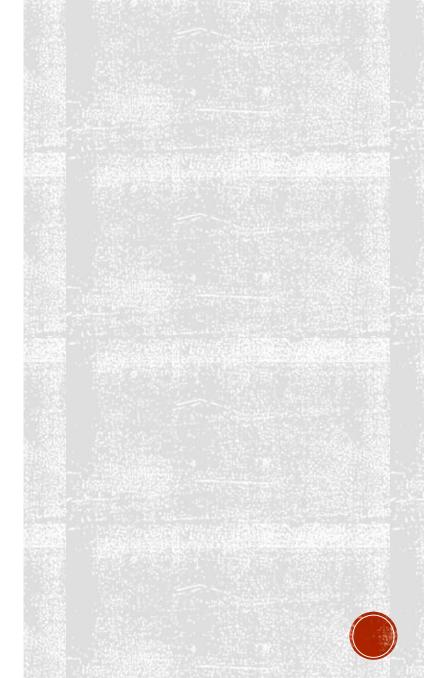


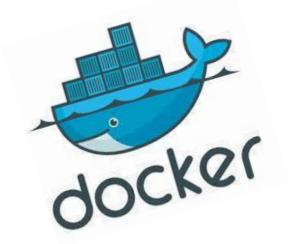
TO BE HONEST

- 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



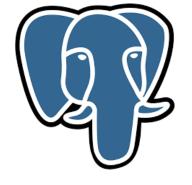
















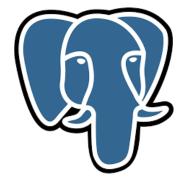


















R.T.F.M.

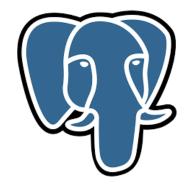
The Way To Success

















R.T.F.M.

The Way To Success

Thank you

