



IFB cloud:

Integration of snakemake workflows in
an appliance designed for ChIP-seq analysis

Claire Rioualen

2015/10/28

French Institute of Bioinformatics (IFB)



- National service infrastructure in bioinformatics
- 32 platforms grouped into regional centers
- French node of the European research infrastructure ELIXIR



French Institute of Bioinformatics (IFB)



- National service infrastructure in bioinformatics
- 32 platforms grouped into regional centers
- French node of the European research infrastructure ELIXIR
- Provides services and infrastructures in bioinformatics for scientists and engineers working in the field of life sciences



<http://www.france-bioinformatique.fr>

IFB cloud

- 200 computing cores
- 2 To of RAM
- 50 To of storage

By 2016

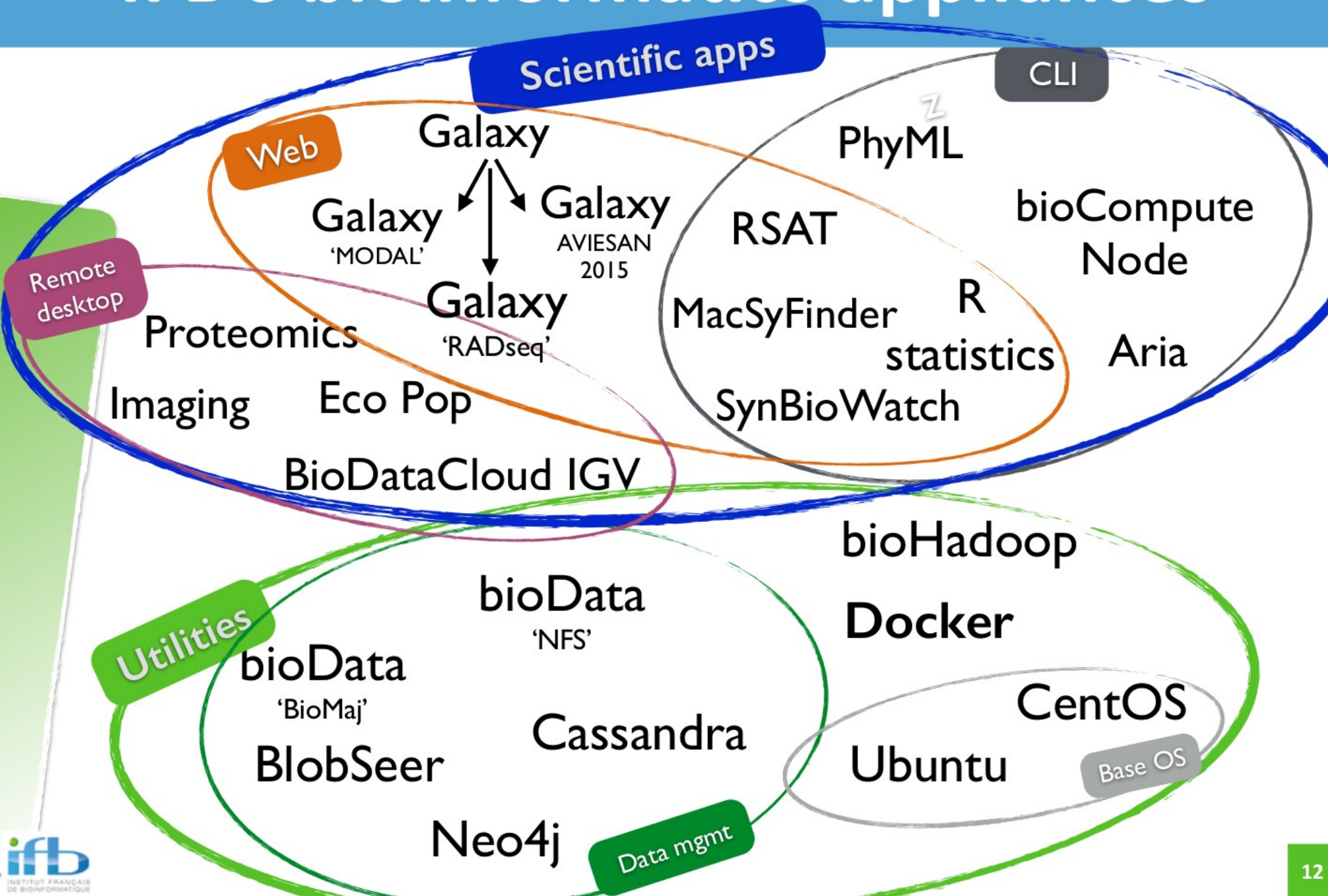
- 3,000 cores
- 500 To of storage

By 2017

- 10,000 cores
- 2 Po of storage



IFB's bioinformatics appliances



A lot of bioinformatics tools



ABYSS 1.3.4
ARIA 2.3
Bioconductor 2.11
biomaj
BLAST+ 2.2.27
Blat 35
Bowtie 0.12.8
Bowtie2 2.0.0-beta7
BWA 0.6.2

BWA 0.7.10
CAP3
CD-HIT 4.6.1
Clustal Omega 1.0.3
CLUSTALW 2.1
Cufflinks 2.0.2
Cutadapt 1.2.1
E-SURGE 1.9.0
Exonerate 2.2.0
eXpress 1.5.1
FastA 3.6
FastQC 0.10.1
Galaxy portal
GATK 2.3.4
HMMer 3.0
ImageJ 1.48
khmer 1.1
M-SURGE 1.8.5
MEME 4.7
MMSEQ 0.11.2a

Mobyle
MODAL
MultAlin 5.4.1
MUSCLE 3.8.31
neo4j
Oases 0.2.08
OMSSA 2.1.9
PeptideShaker 0.18.3
phym1 3.1
PREDATOR 2.1.2
proline
python 2.7
R 2.13
R 3.1.1
R 3.1.2
R-studio
Ray 1.3
RSAT
samtools 0.1.18
Samtools 1.1

SearchGUI 1.10.4
SeqClean
Shiny
Stacks
STAR 2.4.0f1
SuMo v1
TGICL
TopHat 2.0.6
trim_galore 0.3.7
Trinity 2.0.4
U-CARE 2.3.2
VCFtools 0.1.11
Velvet 1.2.10
X!tandem
12-10-01-1
XPLOR-NIH 2.30
...

ChIP-seq appliance

- Base Ubuntu 14.04

Including ssh, rsync, git, python 2, python 3...

ChIP-seq appliance

- Base Ubuntu 14.04

Including ssh, rsync, git, python 2, python 3...

- Programming tools

snakemake, graphviz, pandas library...

ChIP-seq appliance

- Base Ubuntu 14.04
Including ssh, rsync, git, python 2, python 3...
- Programming tools
snakemake, graphviz, pandas library...
- NGS tools →




ChIP-seq appliance

- Base Ubuntu 14.04
Including ssh, rsync, git, python 2, python 3...
- Programming tools
snakemake, graphviz, pandas library...
- NGS tools →
- fg-chip-seq git repository
Including workflow examples (py + json)




Using the IFB cloud: request account



 cloud.france-bioinformatique.fr

IFB BIOINFORMATICS CLOUD

WELCOME!
SIGN IN | HELP



SIGN IN

Hosted at  Powered by 






Username

Password


Login

[Lost password](#) | [Request account](#)

IFB acknowledges funding by the call "Infrastructures in Biology and Health"
in the framework of the French "Investments for the Future" initiative




IFB is the French ELIXIR node



Request account



Using the IFB cloud: dashboard

 cloud.france-bioinformatique.fr

Add ssh key

IFB BIOINFORMATICS CLOUD YOU ARE SIGNED IN AS RIOUALEN
NEWS | DASHBOARD | MONITOR | SETTINGS | HELP | SIGN OUT

DASHBOARD

Hosted at  Powered by 

NEWS Shutdown Go Get IPs Rename New Instance New vDisk Show Instances Show vDisks Show Appliances **STORAGE**

ROOM FOR VMS

- c2.large 2 / 2
- c2.small 8 / 8
- c2.xlarge 1 / 1
- c3.large 2 / 2
- c3.medium 4 / 4
- c3.xlarge 1 / 1
- m1.medium 1 / 1

Showing 0 to 0 of 0 entries

ID	Name	Appliance	CPU%	Mem.	#Storage	Access
No instances available.						
0		0	0	0	0	

Show 25 entries


First Previous Next Last

Create a virtual disk




CPU free (100.00%)

MEMORY

Using the IFB cloud: vDisk

 cloud.france-bioinformatique.fr

IFB BIOINFORMATICS CLOUD YOU ARE SIGNED IN AS **RIQUALEN**
NEWS | **DASHBOARD** | MONITOR | SETTINGS | HELP | SIGN OUT

 **DASHBOARD**  

Operation was successful.

NEWS Shutdown Go Get IPs Rename New Instance New vDisk Show Instances Show vDisks Show Appliances **STORAGE**

Showing 0 to 0 of 0 entries

	ID	Name	Appliance	CPU%	CPU	Mem.	#Storage	Access
No instances available.								

Show 25 entries

Create a virtual Disk

Define Your Parameters

Size ? 10

Name ? ecoli-10G

Create Cancel


ROOM FOR VMS

- c2.large 2 / 2
- c2.small 8 / 8
- c2.xlarge 1 / 1
- c3.large 2 / 2
- c3.medium 4 / 4
- c3.xlarge 1 / 1
- m1.medium 1 / 1



CPU

MEMORY

Using the IFB cloud: create instance

 cloud.france-bioinformatique.fr

IFB BIOINFORMATICS CLOUD YOU ARE SIGNED IN AS RIOUALEN
[NEWS](#) | [DASHBOARD](#) | [MONITOR](#) | [SETTINGS](#) | [HELP](#) | [SIGN OUT](#)

DASHBOARD Hosted at  Powered by 

NEWS Shutdown Go Get IPs Rename New Instance New vDisk Show Instances Show vDisks Show Appliances **STORAGE**

ROOM FOR VMs

	ID
c2.large	2 / 2
c2.small	8 / 8
c2.xlarge	1 / 1
c3.large	2 / 2
c3.medium	4 / 4
c3.xlarge	1 / 1
m1.medium	1 / 1

Showing 0 to 0 of 0 entries

Show 25 entries

Launch a virtual machine

Choose The Appliance

Appliance ? **FG ChIP-Seq 0.2** Choose appliance

Filter by ? **--- THEMATIC FIELDS ---**

--- TOOLS ---

Configure Your Virtual Machine

Name ? **ecoli-workflow** Name

Unique ? ☐

Type ? **c3.large (4 CPU, 16GB RAM)** Up to 8 CPU, 32GB RAM

Number ? **1**

Create appliance ? ☐


Plug Your Additional Storage

Persistent disk ? **ecoli-10G** Plug vDisk

Run **Cancel**



CPU **MEMORY**

Using the IFB cloud: ssh connection

 cloud.france-bioinformatique.fr

IFB BIOINFORMATICS CLOUD YOU ARE SIGNED IN AS RIOUALEN
NEWS | DASHBOARD | MONITOR | SETTINGS | HELP | SIGN OUT

DASHBOARD

Hosted at  Powered by 

NEWS Shutdown Go Get IPs Rename

Showing 1 to 1 of 1 entries

ROOM FOR VMS

- c2.large 1 / 2
- c2.small 4 / 8
- c2.xlarge 0 / 1
- c3.large 1 / 2
- c3.medium 2 / 4
- c3.xlarge 0 / 1
- m1.medium 0 / 1

ID	Name	Appliance	CPU%	CPU	Mem.	#Storage	Access
5598	ecoli-workflow	EG-ChIP-Seq-0.2	0%	4	0	1	ssh http

Search:

First | Prev | 1 | Next | Last

STORAGE

CPU free (50.00%)

MEMORY

Connection Information

Parameters:
host = 192.54.201.111
port = 22
identifiant = root

Command-line connection:
ssh -A -p 22 root@192.54.201.111

Command-line file transfers:
scp -P 22 \${localfile} root@192.54.201.111:
sftp -oPort=22 root@192.54.201.111

Close

ssh connection

Using the IFB cloud: data

```
root@vm0080:~# ls -R /root/mydisk
/root/mydisk:
data  genomes  results

/root/mydisk/data:
design.tab  GSM000001  GSM000002  samples.tab

/root/mydisk/data/GSM000001:
SRR12345.fastq

/root/mydisk/data/GSM000002:
SRR12345.fastq

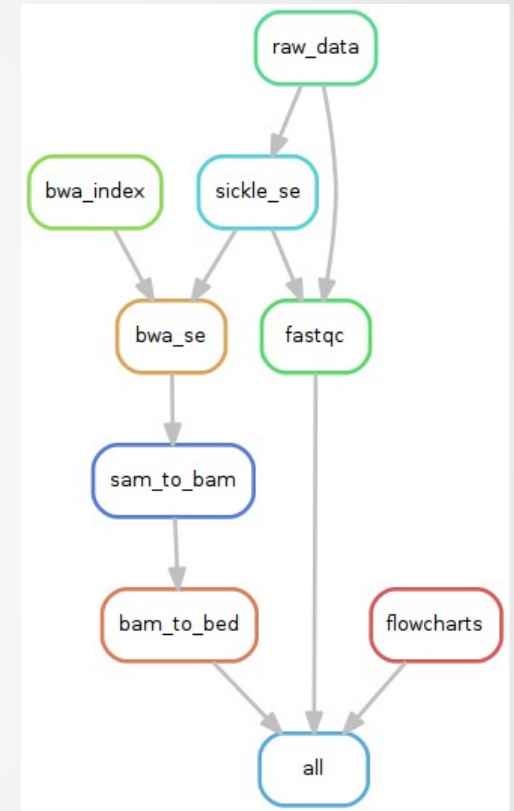
/root/mydisk/genomes:
ecoli

/root/mydisk/genomes/ecoli:
ecoli_K12.fasta

/root/mydisk/results:
root@vm0080:~#
```

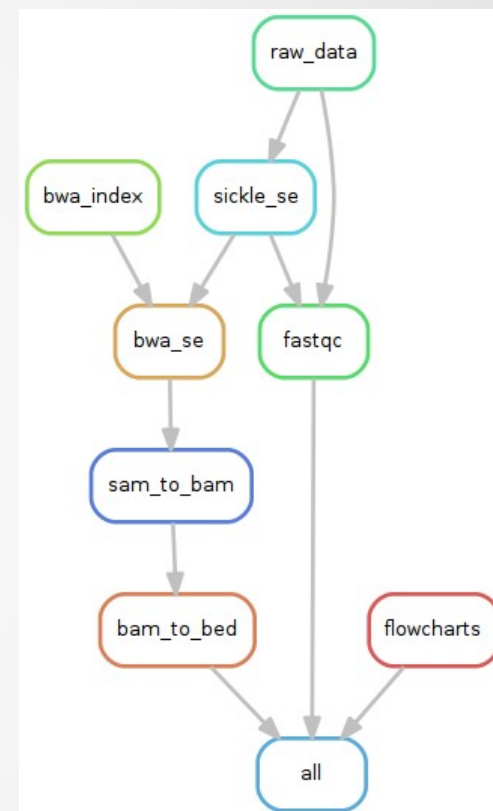

Using the IFB cloud: execute workflow

- Several example workflows already available & adaptable in the git repository
- 1 config file (json) + 1 workflow file (py)
 - Paths to data, genome & results directories
 - Parameters



Using the IFB cloud: execute workflow

- Several example workflows already available & adaptable in the git repository
- 1 config file (json) + 1 workflow file (py)
 - Paths to data, genome & results directories
 - Parameters
- Results are saved in the vDisk



```
root@vm0080:~# su rg
rg@vm0080:/root$ cd ~/workspace/fg-chip-seq/
rg@vm0080:~/workspace/fg-chip-seq$ ls
doc  README.md  scripts
rg@vm0080:~/workspace/fg-chip-seq$ snakemake -s scripts/snakefiles/workflows/ecoli.py
```



Thank you!