

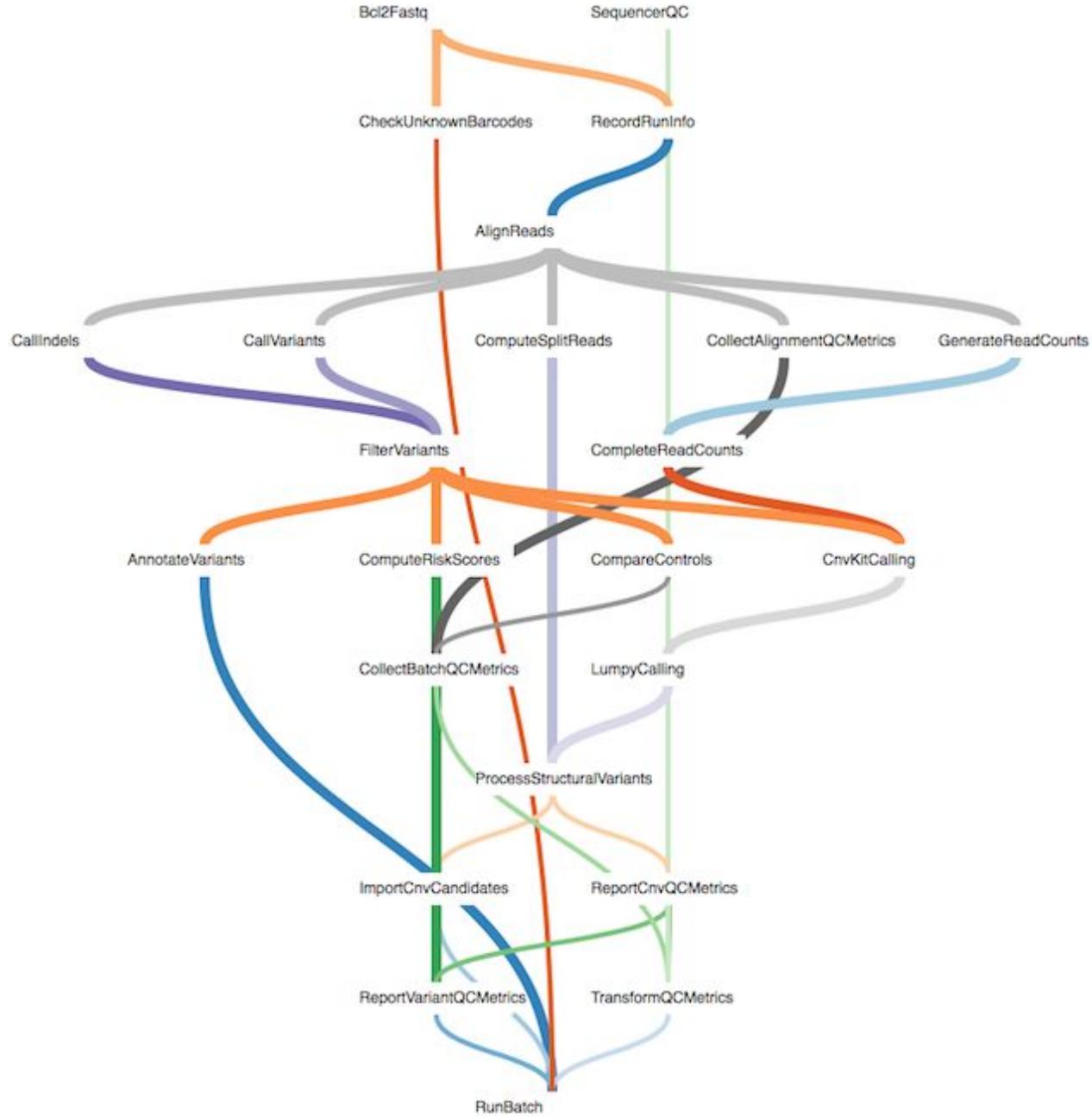
# Containerization of bioinformatics tools

---

Chadi Saad  
Bioinformatics Team  
CHU de Lille

# Preamble

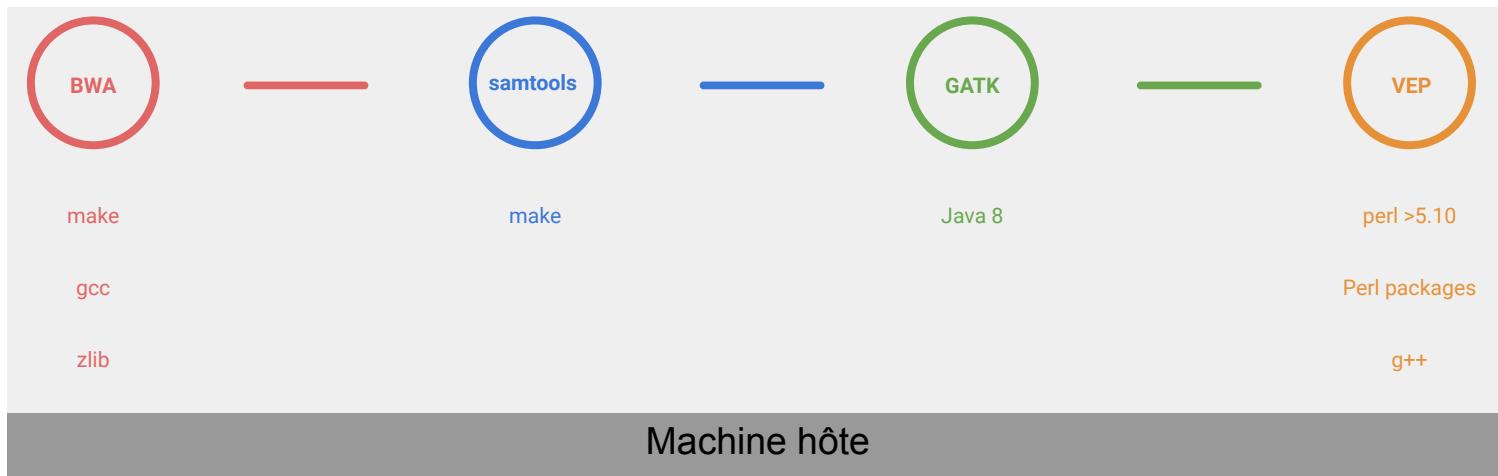
Bioinformatic pipelines are generally based on a combination of several third-party tools.



# Dependencies



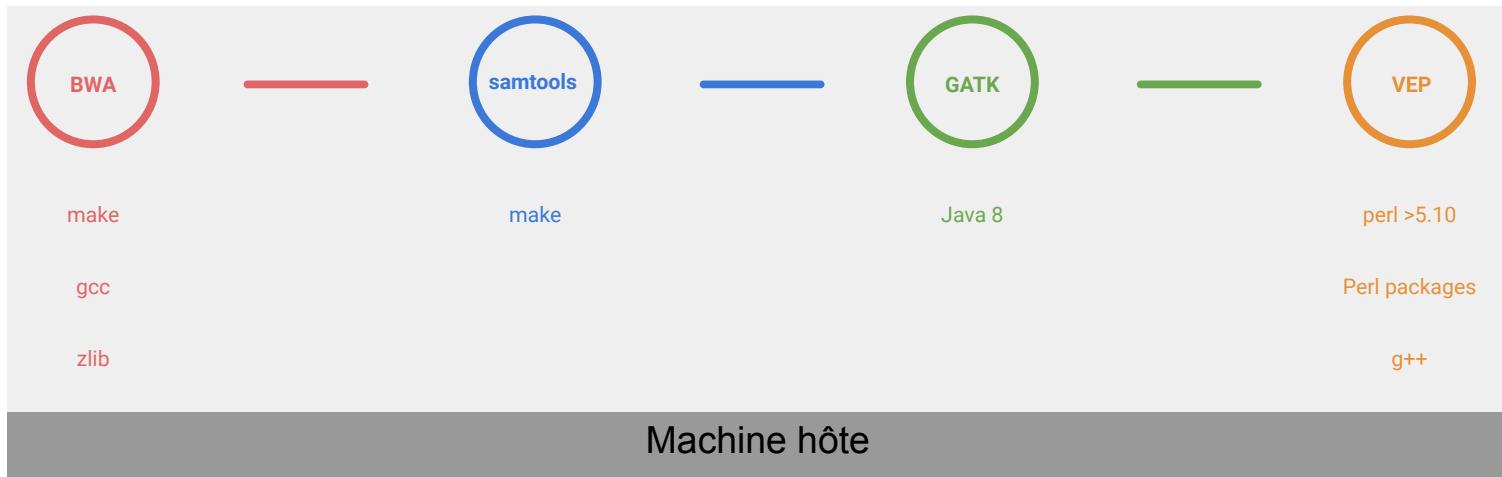
# Dependencies



- Installation/configuration de tous les outils avec leurs **dépendances**
- Maintenances & Mises à jours
- Conflits de paquets (python2.7 vs 3)
- Portabilité entre différents OS (Linux, Windows,...)



# Dependencies

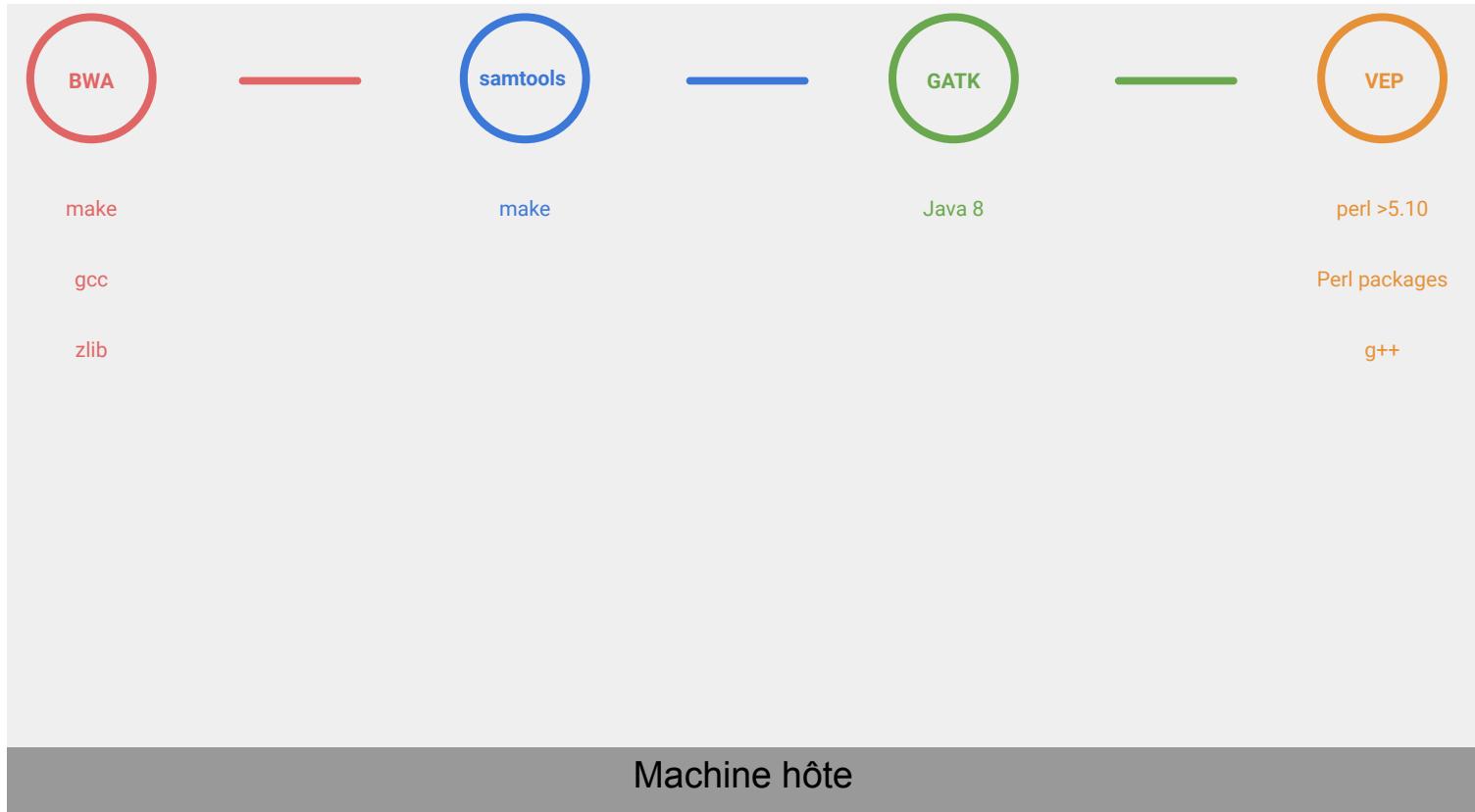


- Installation/configuration de tous les outils avec leurs **dépendances**
- Maintenances & Mises à jours
- Conflits de paquets (python2.7 vs 3)
- Portabilité entre différents OS (Linux, Windows,...)



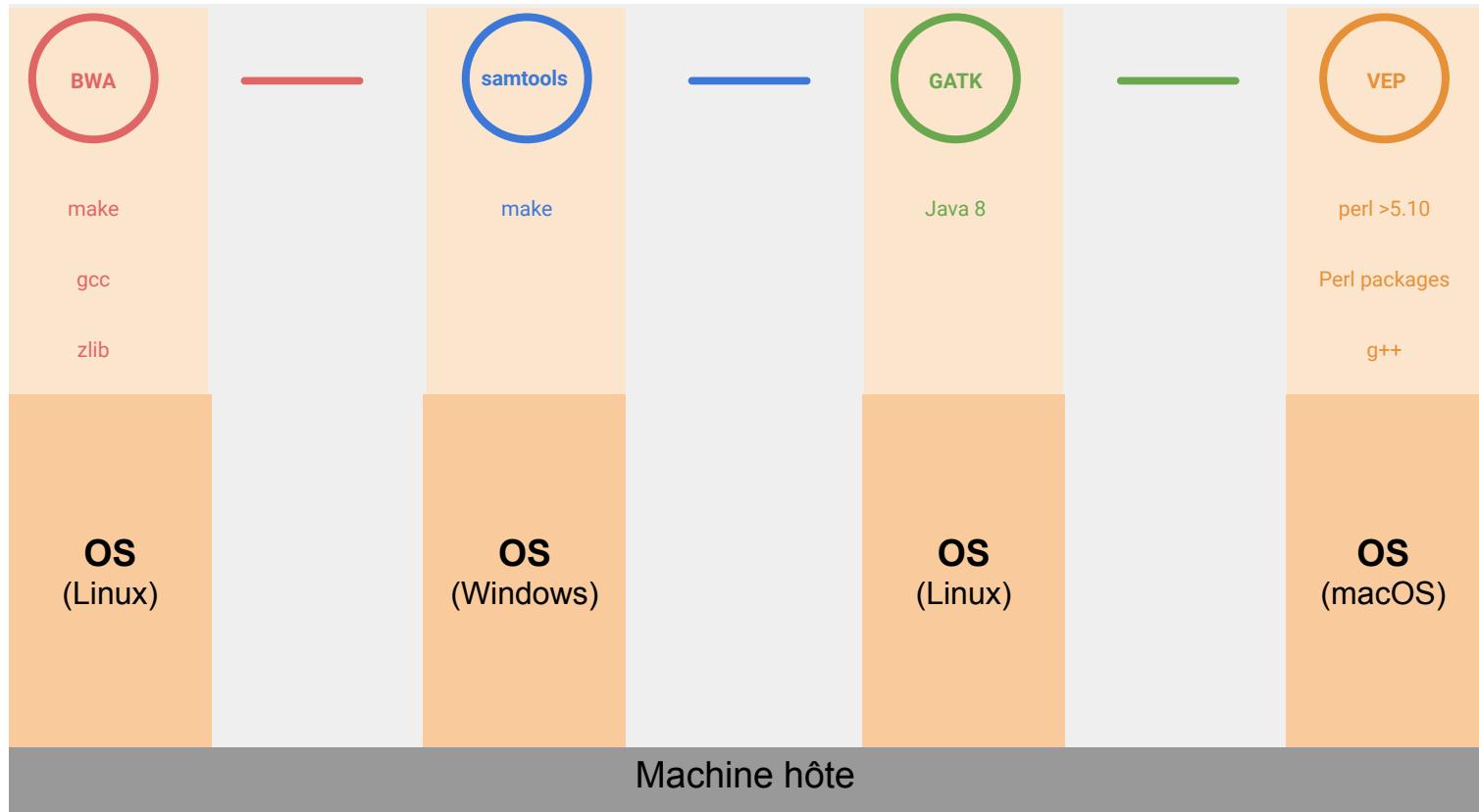
# Virtualization

virtual machines



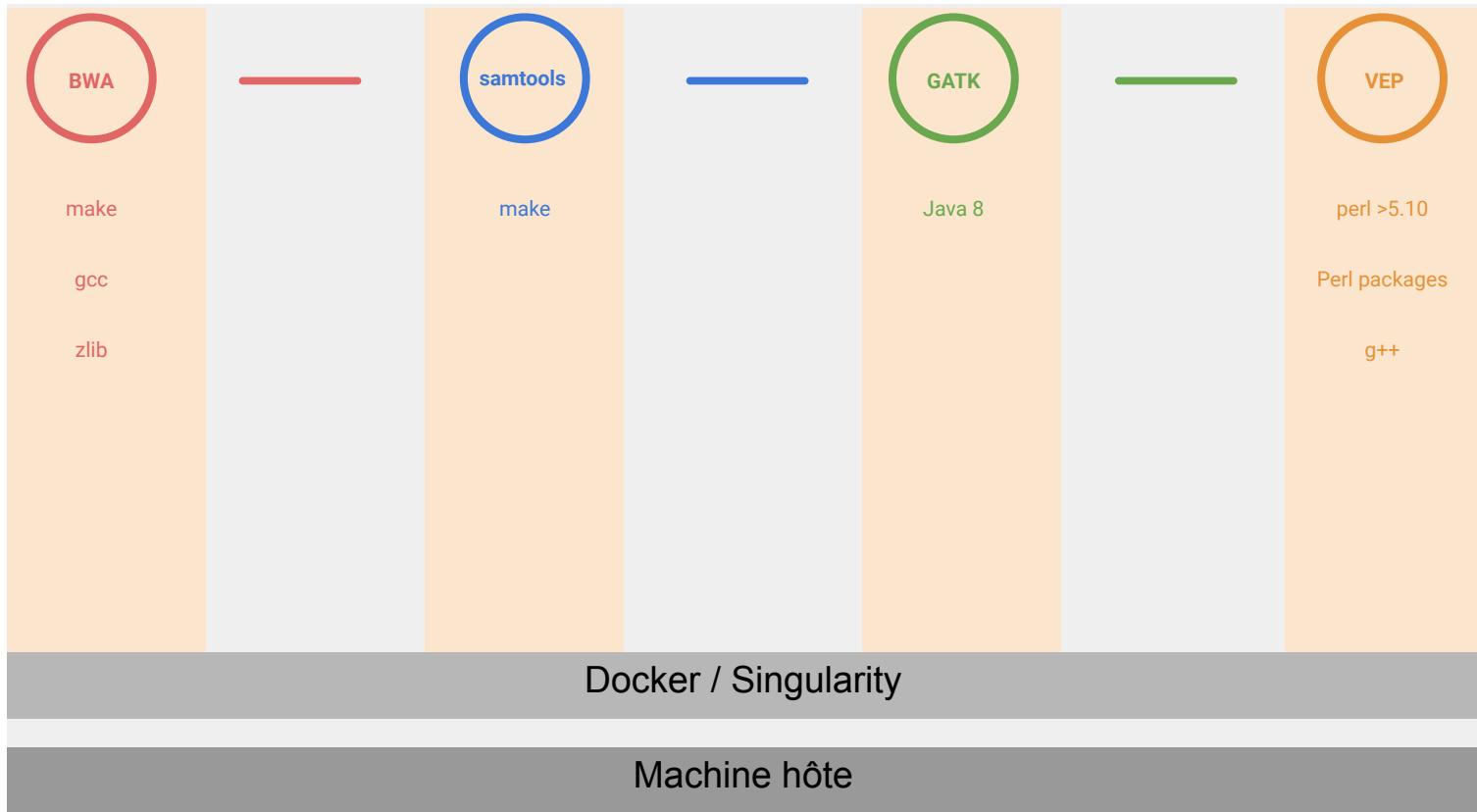
# Virtualization

virtual machines



# Virtualization

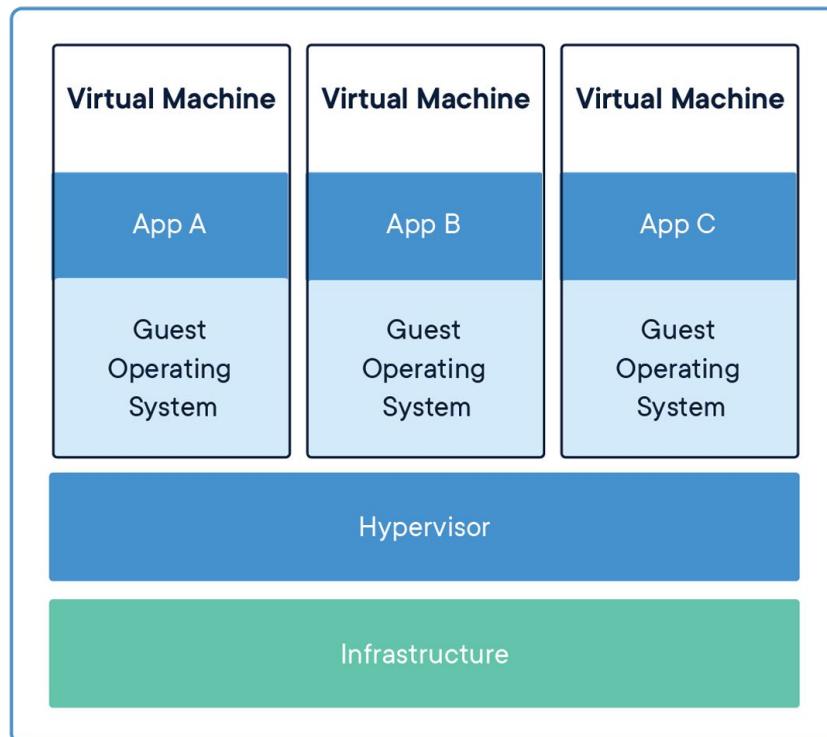
containers



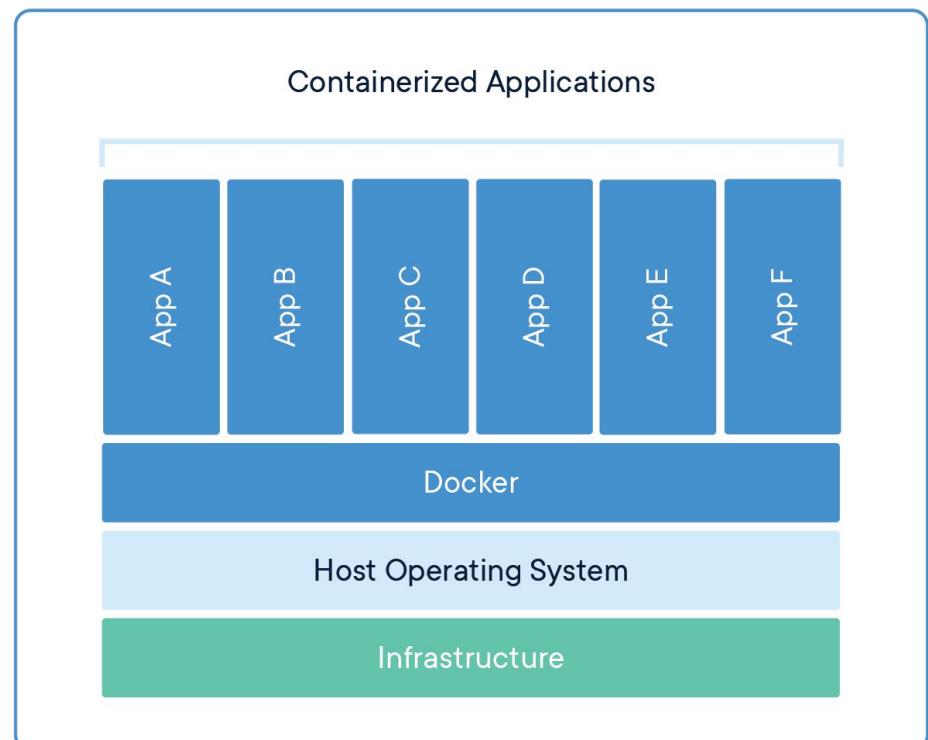
# Virtualization

- A Virtual Machine is the simulation of a computer system that reproduces its actual use
- Abstract all the characteristics of the system and simulate another system to have a **transportable** environment and thus **reproducible** from one machine to another.

Machines virtuelles



Conteneurs



# Containerization

## The advantages of containers:

- Easy to set up and use

```
FROM ubuntu:16.04

RUN apt-get update && \
    apt-get install samtools
```

- Access to isolated and controlled environments
- Ease of searching and sharing Docker containers



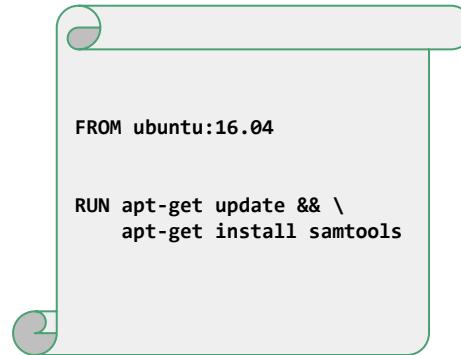
SINGULARITY HUB



# Containerization

## Operating mode

- Recipe / dockerFile



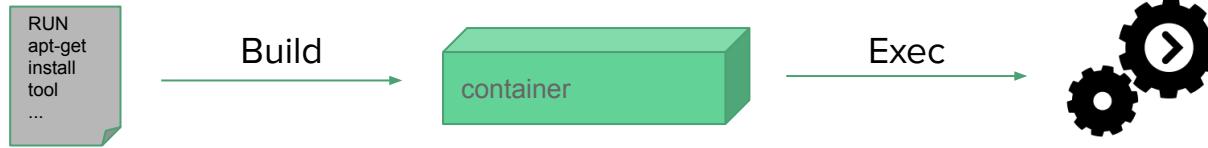
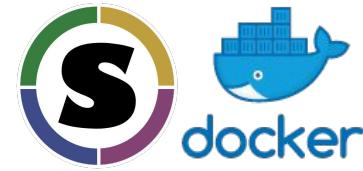
- Construction of the image (container)

```
$ singularity build ... recipe
$ docker build ... dockerfile
```

- Running the tool in the container

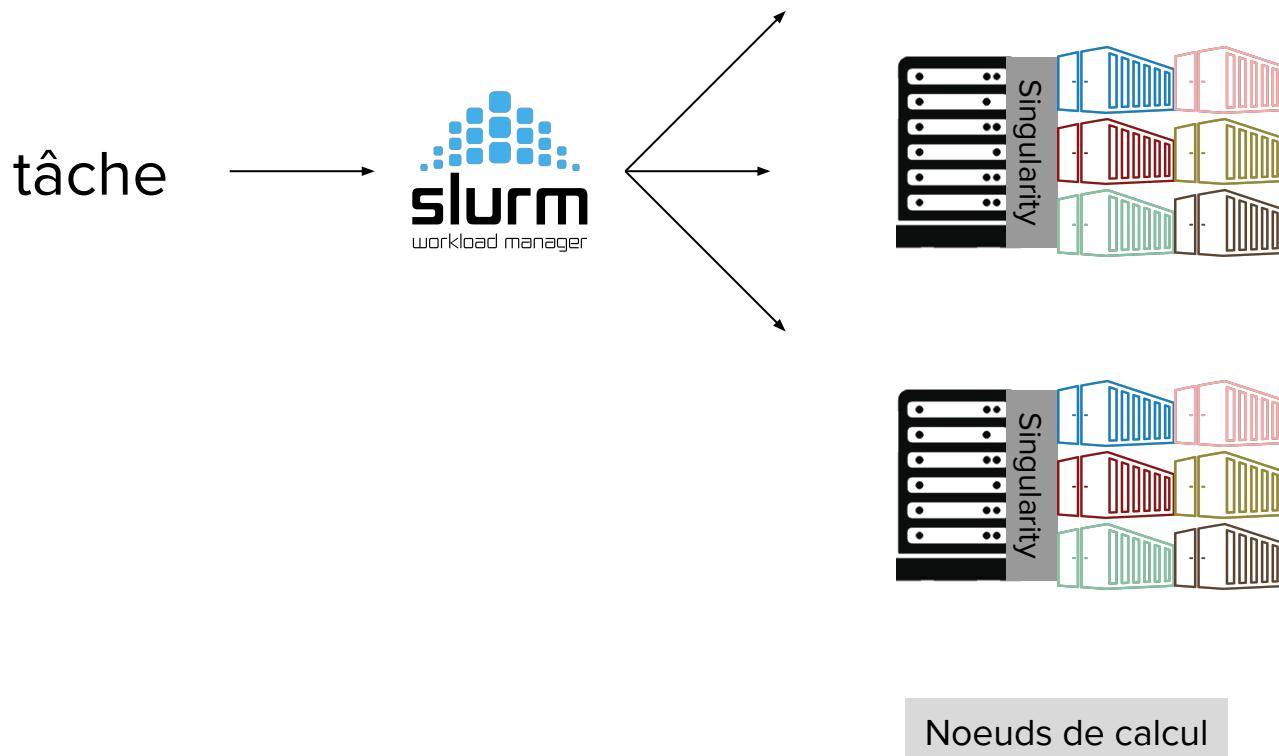
```
$ singularity exec ... samtools.simg samtools index foo.bam
$ docker exec ... samtools_container samtools index foo.bam
```

# Containerization



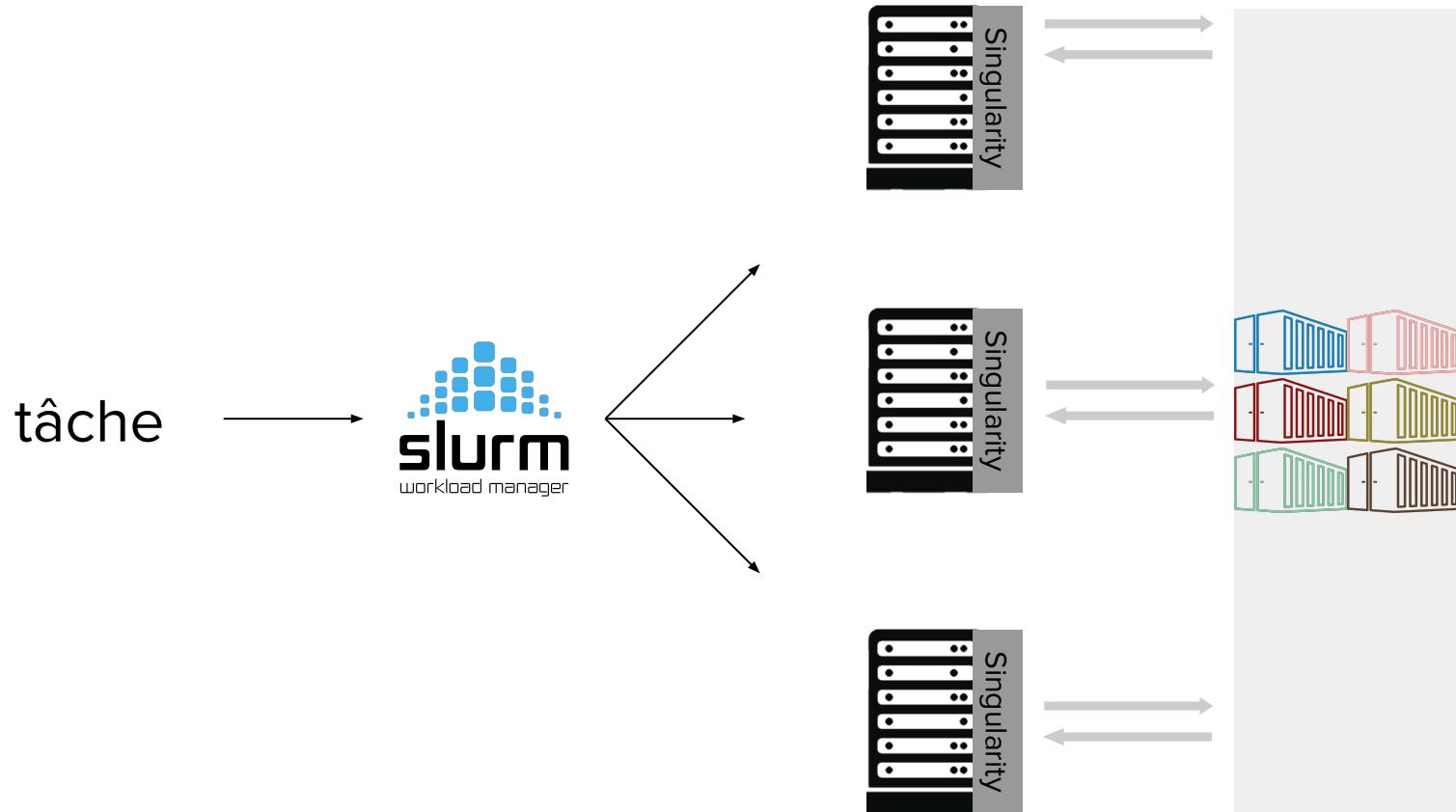
	Docker	Singularity
Security	User running docker commands need to be in special docker group to gain elevated system access	No change in security paradigm. User run singularity image/app without special privileges.
Portability	Docker Hub	Single image file. Can import docker images.
Container activation	Docker daemon running on each host	Image file. No daemon process.
Deployment	Single package install. Daemon service with root privilege	root to install singularity Single .rpm/.deb per host. No scheduler modification needed.
Root escalation	container starts with privileged access	To obtain root in singularity container, it must have been started as root

# Containerization & HPC



# Containerization & HPC

singularity



Noeuds de calcul

Stockage central

# Sharing

DockerHub : <https://hub.docker.com>



## biocontainers/samtools ☆

Pulls 100K+

By [biocontainers](#) • Updated 4 months ago

Tools for manipulating next-generation sequencing data

Container

Overview

Tags

Dockerfile

Builds

### Dockerfile

```
#####
# BASE IMAGE #####
#####

FROM biocontainers/biocontainers:latest

#####
# METADATA #####
LABEL base.image="biocontainers:latest"
LABEL version="2"
LABEL software="Samtools"
LABEL software.version="1.3.1"
LABEL about.summary="Tools for manipulating next-generation sequencing data"
LABEL about.home="https://github.com/samtools/samtools"
LABEL about.documentation="https://github.com/samtools/samtools"
LABEL license="https://github.com/samtools/samtools"
LABEL about.tags="Genomics"

#####
# MAINTAINER #####
MAINTAINER Saulo Alves Aflitos <sauloal@gmail.com>

RUN conda install samtools=1.3.1

WORKDIR /data/

CMD ["samtools"]
```

### Source Repository



Github

[BioContainers/containers](#)

### Docker Pull Command

```
docker pull biocontainers/samtools 
```

# Sharing

SingularityHub : <https://singularity-hub.org>

SINGULARITY HUB

Collections ▾

About

User Guide

Get Help



## qbicsoftware/qbic-singularity-samtools

QBIC singularity container for samtools

SUPPLEMENTARY ▾

USAGE

Builds COMMIT

url	Recipe	Status	Tag (Branch)
<a href="#">qbicsoftware/qbic-singularity-samtools:v1.7</a>	Singularity.v1.7	<span>COMPLETE </span>	v1.7 (master)
<a href="#">qbicsoftware/qbic-singularity-samtools:latest</a>	Singularity.latest	<span>COMPLETE</span>	latest (master)
<a href="#">qbicsoftware/qbic-singularity-samtools:v1.6</a>	Singularity.v1.6	<span>COMPLETE </span>	v1.6 (master)

# Sharing

SingularityHub : <https://singularity-hub.org>

The screenshot shows a web interface for a Singularity container on SingularityHub. At the top, there's a navigation bar with links for SINGULARITY HUB, Collections, About, User Guide, Get Help, and a search icon.

The main content area displays a container page for the repository `qbicsoftware/qbic-singularity-samtools`. It includes a brief description: "QBIC singularity container for samtools". Below this are two buttons: "SUPPLEMENTARY" and "USAGE".

The "Builds" section lists two entries:

- url**: `qbicsoftware/qbic-singularity-samtools:v1.7` | **Recipe**: `Singularity.v1.7` (Branch) (master)
- url**: `qbicsoftware/qbic-singularity-samtools:latest` | **Recipe**: `Singularity.latest` (Branch) latest (master)

Below these, another entry is shown:

- url**: `qbicsoftware/qbic-singularity-samtools:v1.6` | **Recipe**: `Singularity.v1.6` (Branch) v1.6 (master)

A large black triangular callout box is positioned over the second build entry, containing the container's recipe script:

```
Bootstrap:docker
From:alpine:3.7

%post
/bin/sh build.sh
cd /build
SAMTOOLS_VERSION=1.7
wget https://github.com/samtools/samtools/releases/download/jxf samtools-$SAMTOOLS_VERSION.tar.bz2
cd samtools-$SAMTOOLS_VERSION
make
make install prefix=/usr

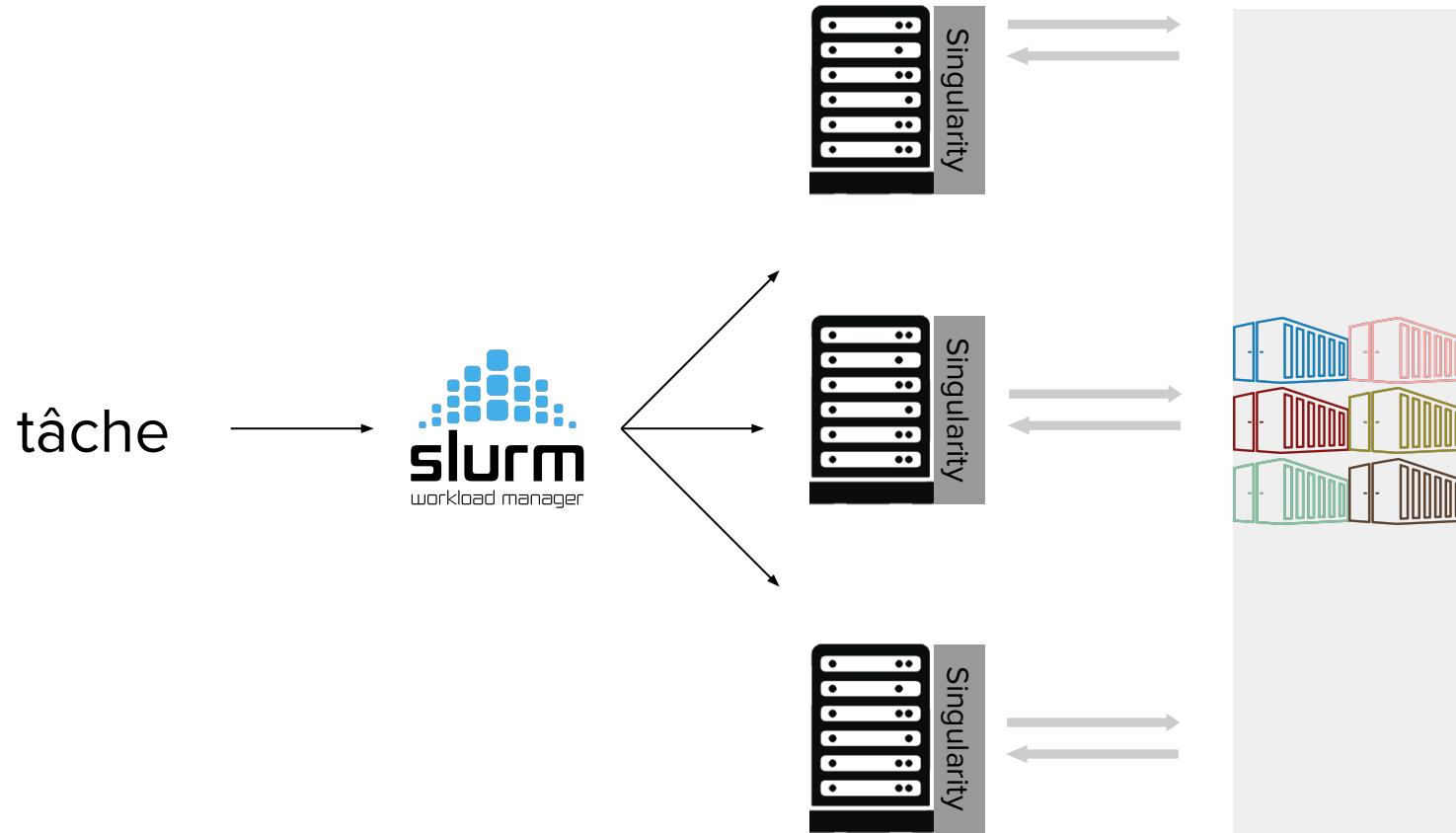
%files
#Installation of Samtools
build.sh

%environment
SAMTOOLS_VERSION=1.7

%labels
Maintainer alexander.peltzer@uni-tuebingen.de
```

# Containerization & HPC

singularity

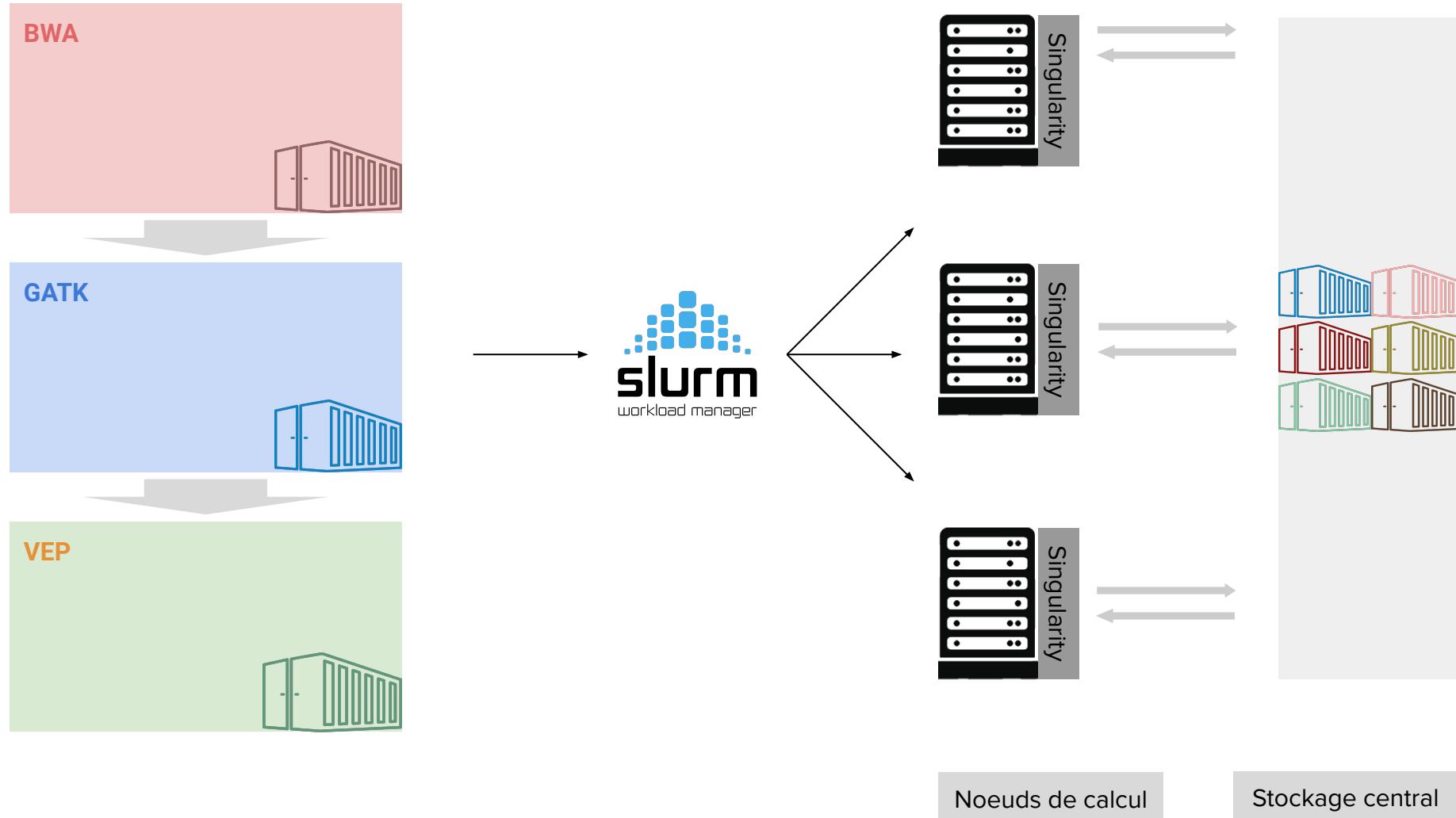


Noeuds de calcul

Stockage central

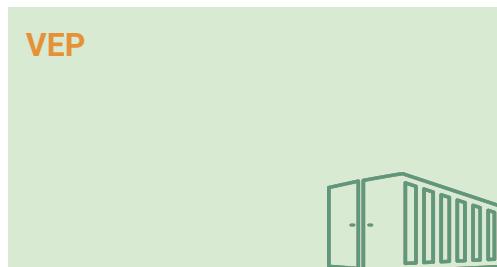
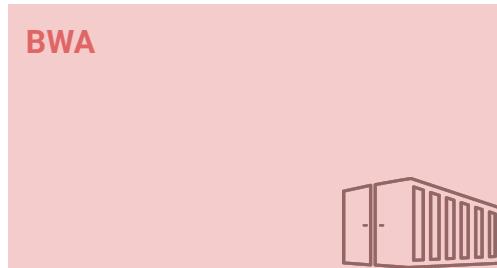
# Containerization & HPC : Pipelines

singularity



# Containerization & HPC : Pipelines

NextFlow



```
// Input
input_channel = Channel.fromFilePairs(params.reads)

# Align
process alignment {
    container /path/to/containers/BWA.simg

    input:
        val (sample), file(fq1), file(fq2) from input_channel
    output:
        val (sample), file(bam)           into bam_files_channel
    script:
        bwa mem ref.fa fq1 fq2 > ${sample}.bam
}

// Call variants
process variant_calling {
    container /path/to/containers/GATK.simg

    input:
        val (sample), file(bam)   from bam_files_channel
    output:
        val (sample), file(vcf)   into vcf_files_channel
    script:
        GATK -T HaplotypeCaller -R ref.fa -I ${sample}.bam -o ${sample}.vcf
}

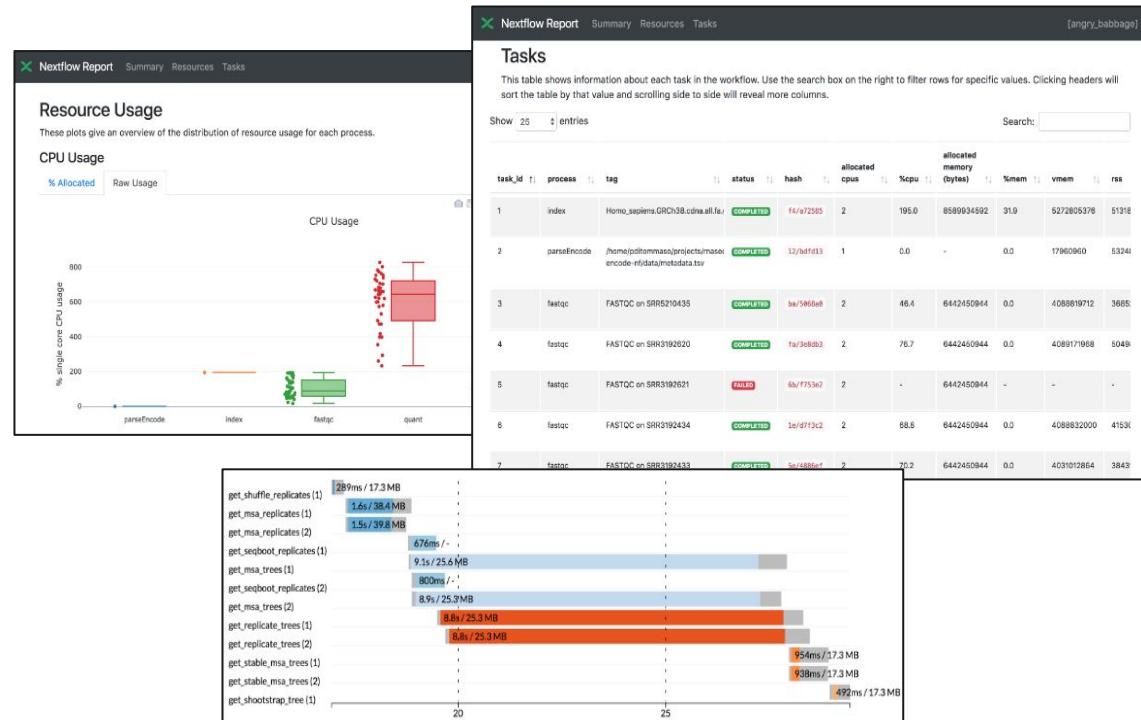
// Annotate Variants
/* ... */
```

**nextflow**

# Nextflow

- Parallelization management : *dataflow programming model*
- Resource Management (RAM, CPU)
- Easy to use with a cluster (slurm, aws, google cloud, ...)
- Easy to use with containers
- Automatic error/logs management : traceability !
- Groovy language

- ...



nf-core

# nf-core



A community effort to collect a curated set of analysis pipelines built using Nextflow.

VIEW PIPELINES

## Used by groups all over the world

The nf-core community is spread all over the globe and includes a large number of contributing users. [See all »](#)



NATIONAL  
GENOMICS  
INFRASTRUCTURE



QBIC



CRG<sup>R</sup>  
Centre for Genomic  
Regulation



A\*STAR  
Genome Institute  
of Singapore



THE  
FRANCIS CRICK  
INSTITUTE



welcome  
**sanger**  
institute



CANCER  
RESEARCH  
UK



BEATSON  
INSTITUTE



International Agency for Research on Cancer



IMP  
Institute of  
Molecular Pathology



lifebit



SYSUCC  
SINCE 1964



中山大學  
肿瘤防治中心  
SUN YAT-SEN UNIVERSITY CANCER CENTER



University  
of Colorado  
Boulder



GMI  
GREGOR MENDEL INSTITUTE  
OF MOLECULAR PLANT BIOLOGY

# nf-core

Search keywords

Filter:

Released 7

Under development 13

Sort:

Last Release

Alphabetical

Status

Stars

## nf-core/mhcquant ✓

☆ 2

mass-spectrometry peptides

Identify and quantify peptides from mass spectrometry raw data

Version 1.2.2

Published 23 hours ago

## nf-core/methylseq ✓

☆ 21

bisulfite-sequencing dna-methylation methyl-seq

Methylation (Bisulfite-Sequencing) analysis pipeline using Bismark or bwa-meth + MethylDackel

Version 1.2

Published 4 weeks ago

## nf-core/rnaseq ✓

☆ 73

rna rna-seq

RNA sequencing analysis pipeline using STAR or HISAT2, with gene counts and quality control

Version 1.2

Published 2 months ago

## nf-core/deepvariant ✓

☆ 12

deep-variant dna google variant-calling

Google's DeepVariant variant caller as a Nextflow pipeline

Version 1.0

Published 2 months ago

## nf-core/eager ✓

☆ 9

adna ancientdna pathogen-genomics population-genetics

A fully reproducible and state of the art ancient DNA analysis pipeline.

Version 2.0.4

Published 3 weeks ago

## nf-core/ampliseq ✓

☆ 19

16s amplicon-sequencing docker singularity

16S rRNA amplicon sequencing analysis workflow using QIIME2

Version 1.0.0

Published 2 months ago

## nf-core/hlatyping ✓

☆ 1

dna hla hla-typing immunology optotype personalized-medicine rna

Precision HLA typing from next-generation sequencing data

Version 1.1.2

Published 2 months ago

## nf-core/smrnaseq !

☆ 12

small-rna smrna-seq

Analysis pipeline for small-RNA sequencing data.

No releases yet

# Sources

- <https://hub.docker.com>
- <https://bioinfo-fr.net/virtualisez-plus-de-reproductibilite>
- <https://singularity-hub.org>
- [https://tin6150.github.io/psg/blogger\\_container\\_hpc.html](https://tin6150.github.io/psg/blogger_container_hpc.html)
- <https://nf-co.re/>