DNA Club

Using CAW with Singularity containers on UPPMAX clusters

Maxime Garcia

2017-09-20

What is Singularity?

http://singularity.lbl.gov/
https://singularity-hub.org/

- Docker-like containers technology
- Specific for HPC environnment

What is Singularity?

http://singularity.lbl.gov/
https://singularity-hub.org/

- Docker-like containers technology
- Specific for HPC environnment
- Without the root user security problem

What is Singularity?

http://singularity.lbl.gov/
https://singularity-hub.org/

- Docker-like containers technology
- Specific for HPC environnment
- Without the root user security problem
- Singularity hub offers a Docker hub like repository
- Singularity containers can be pulled from Docker hub

What is Singularity?

http://singularity.lbl.gov/
https://singularity-hub.org/

- Docker-like containers technology
- Specific for HPC environnment
- Without the root user security problem
- Singularity hub offers a Docker hub like repository
- Singularity containers can be pulled from Docker hub
- Supported by Nextflow

Why would we want to use that?

- For a better control

Why would we want to use that?

- For a better control
  - of every tool used
  - of every version used

Why would we want to use that?

- For a better control
    - of every tool used
    - of every version used
- For a better reproducibility

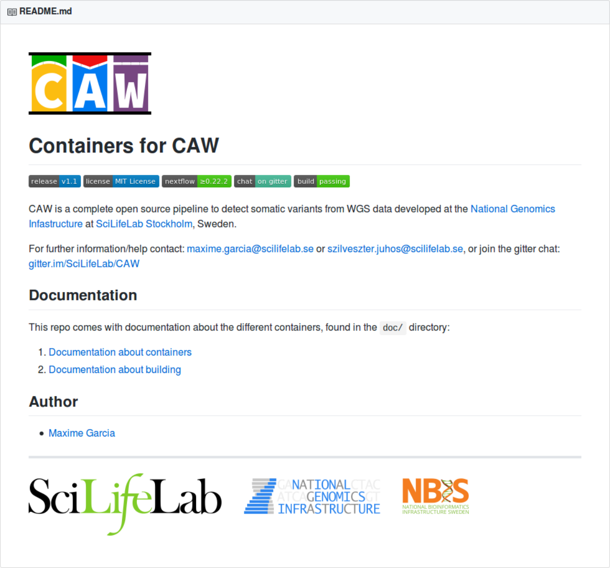
Why would we want to use that?

- For a better control
    - of every tool used
    - of every version used
- For a better reproducibility

Containerization is not only portable but also reproducible.

We can do everything!

The CAW-containers repository



Containers for CAW

CAW is a complete open source pipeline to detect somatic variants from WGS data developed at the National Genomics Infastructure at SciLifeLab Stockholm, Sweden.

For further information/help contact: maxime.garcia@scilifelab.se or szilveszter.juhos@scilifelab.se, or join the gitter chat: gitter.im/SciLifeLab/CAW

Documentation

This repo comes with documentation about the different containers, found in the `doc/` directory:

1. Documentation about containers
2. Documentation about building

Author

- Maxime Garcia

How many containers?

How many containers?

- One container for everything

How many containers?

- One container for everything
- Several containers (one for each process)

The One Script

`SciLifeLab/CAW-containers`

- 24 different Docker containers for all the different processes

The One Script

`SciLifeLab/CAW-containers`

- 24 different Docker containers for all the different processes
- One Script to build them all,

The One Script

`SciLifeLab/CAW-containers`

- 24 different Docker containers for all the different processes
- One Script to build them all,

One Script to push them

The One Script

`SciLifeLab/CAW-containers`

- 24 different Docker containers for all the different processes
- One Script to build them all,

One Script to push them

- One Script to pull them all,

The One Script

`SciLifeLab/CAW-containers`

- 24 different Docker containers for all the different processes
- One Script to build them all,

  One Script to push them

- One Script to pull them all,

  and with Singularity run them

Inside the script

`SciLifeLab/CAW-containers/main.nf`

```
docker build -t $repository/$container:$tag \
  $baseDir/containers/$container/.

docker push $repository/$container:$tag

singularity pull --name $container-${tag}.img \
  docker://$repository/$container:$tag
```

Nextflow and Singularity for the win

- Nextflow natively support Singularity

Nextflow and Singularity for the win

nextflow

- Nextflow natively support Singularity
- Easy configuration

Nextflow and Singularity for the win

- Nextflow natively support Singularity
- Easy configuration
- Automatic pull of the containers

Easy configuration

`SciLifeLab/CAW/configuration/singularity.config`

```
singularity {
  enabled = true
}
process {
  $RunFastQC.container        = 'docker://maxulysse/fastqc:1.1'
  $RunFreeBayes.container     = 'docker://maxulysse/freebayes:1.1'
  $RunGenotypeGVCFs.container = 'docker://maxulysse/gatk:1.1'
  $RunManta.container         = 'docker://maxulysse/runmanta:1.1'
  $RunMultiQC.container       = 'docker://maxulysse/multiqc:1.1'
  $RunMutect1.container       = 'docker://maxulysse/mutect1:1.1'
}
```

Easy configuration

```
SciLifeLab/CAW/configuration/singularity.config

singularity {
  enabled = true
}
process {
  $RunFastQC.container         = 'docker://maxulysse/fastqc:1.1'
  $RunFreeBayes.container      = 'docker://maxulysse/freebayes:1.1'
  $RunGenotypeGVCFs.container = 'docker://maxulysse/gatk:1.1'
  $RunManta.container          = 'docker://maxulysse/runmanta:1.1'
  $RunMultiQC.container        = 'docker://maxulysse/multiqc:1.1'
  $RunMutect1.container        = 'docker://maxulysse/mutect1:1.1'
}

process.container = 'shub://MaxUlysse/shubcontainer'

export NXF_SINGULARITY_CACHEDIR=$HOME/.singularity
```

What to do on a secure cluster?

- How about running CAW with Singularity on Bianca or Irma?

What to do on a secure cluster?

- How about running CAW with Singularity on Bianca or Irma?
- No automatic pull of the containers

What to do on a secure cluster?

- How about running CAW with Singularity on Bianca or Irma?
- No automatic pull of the containers
- The configuration is still easy

Still an easy configuration

```
SciLifeLab/CAW/configuration/singularity-download.config

singularity {
  enabled = true
}
process {
  $RunFastQC.container        = 'containers/fastqc-1.1.img'
  $RunFreeBayes.container     = 'containers/freebayes-1.1.img'
  $RunGenotypeGVCFs.container = 'containers/gatk-1.1.img'
  $RunManta.container         = 'containers/runmanta-1.1.img'
  $RunMultiQC.container       = 'containers/multiqc-1.1.img'
  $RunMutect1.container       = 'containers/mutect1-1.1.img'
}

export NXF_SINGULARITY_CACHEDIR=$HOME/.singularity

process.container = '$NXF_SINGULARITY_CACHEDIR/container.img'
```

How to get the containers?

- Use the script that does the thing

How to get the containers?

- Use the script that does the thing
- To pull all containers from Docker hub into Singularity containers

How to get the containers?

- Use the script that does the thing
- To pull all containers from Docker hub into Singularity containers
- Transfer the containers to the secure cluster

One last detail

- Do not forget to create the UPPMAX specific directories in the containers

One last detail

- Do not forget to create the UPPMAX specific directories in the containers
- `/pica`
- `/proj`
- `/sw`

Let the magic happen

`SciLifeLab/CAW/main.nf`

```
nextflow run ~/CAW/main.nf --project project --step preprocessing \
  --genome GRCh38 --sample sample.tsv -profile singularityLocal
```

Using parralelization to do multiple tests

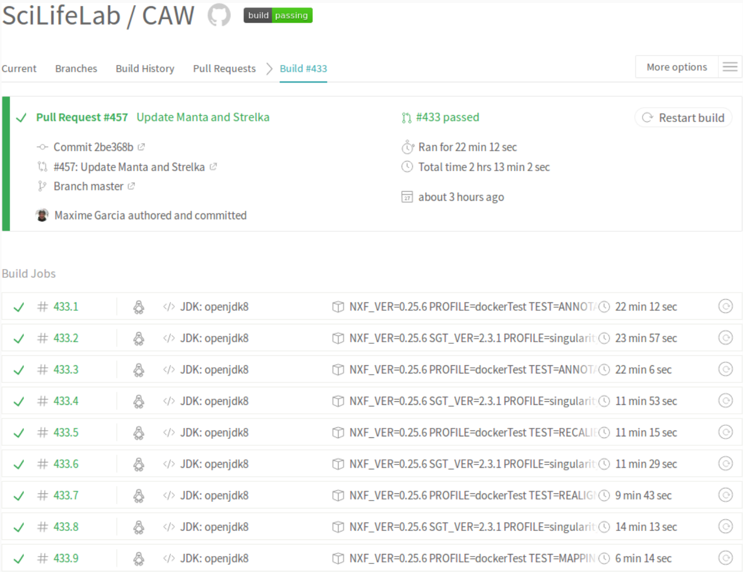
SciLifeLab/CAW/.travis.yml

```
^^Isudo: required
^^Ilanguage: java
^^Ijdk: openjdk8
^^Iservices:
^^I^^I- docker
^^Ienv:
^^I^^I- NXF_VER=0.25.6 PROFILE=dockerTest TEST=ANNOTATEVEP TOOL_INSTALL=nextflow
^^I^^I- NXF_VER=0.25.6 SGT_VER=2.3.1 PROFILE=singularityTest TEST=ANNOTATESNPEFF TOOL_INSTALL=all
^^I^^I- NXF_VER=0.25.6 PROFILE=dockerTest TEST=ANNOTATESNPEFF TOOL_INSTALL=nextflow
^^I^^I- NXF_VER=0.25.6 SGT_VER=2.3.1 PROFILE=singularityTest TEST=RECALIBRATE TOOL_INSTALL=all
^^I^^I- NXF_VER=0.25.6 PROFILE=dockerTest TEST=RECALIBRATE TOOL_INSTALL=nextflow
^^I^^I- NXF_VER=0.25.6 SGT_VER=2.3.1 PROFILE=singularityTest TEST=REALIGN TOOL_INSTALL=all
^^I^^I- NXF_VER=0.25.6 PROFILE=dockerTest TEST=REALIGN TOOL_INSTALL=nextflow
^^I^^I- NXF_VER=0.25.6 SGT_VER=2.3.1 PROFILE=singularityTest TEST=MAPPING TOOL_INSTALL=all
^^I^^I- NXF_VER=0.25.6 PROFILE=dockerTest TEST=MAPPING TOOL_INSTALL=nextflow

^^Iinstall: # Install Nextflow (and Singularity if needed)
^^I^^I- "./scripts/install.sh --tool $TOOL_INSTALL"

^^Iscript:
^^I^^I- "./scripts/test.sh --profile $PROFILE --test $TEST"
```

All your test are belong to us.



✓ **Pull Request #457** Update Manta and Strelka

Commit 2be368b

#457: Update Manta and Strelka

Branch master

Maxime Garcia authored and committed

#433 passed

Ran for 22 min 12 sec

Total time 2 hrs 13 min 2 sec

about 3 hours ago

Build Jobs

✓	# 433.1	JDK: openjdk8	NXF_VER=0.25.6 PROFILE=dockerTest TEST=ANNOT	22 min 12 sec
✓	# 433.2	JDK: openjdk8	NXF_VER=0.25.6 SGT_VER=2.3.1 PROFILE=singularit	23 min 57 sec
✓	# 433.3	JDK: openjdk8	NXF_VER=0.25.6 PROFILE=dockerTest TEST=ANNOT	22 min 6 sec
✓	# 433.4	JDK: openjdk8	NXF_VER=0.25.6 SGT_VER=2.3.1 PROFILE=singularit	11 min 53 sec
✓	# 433.5	JDK: openjdk8	NXF_VER=0.25.6 PROFILE=dockerTest TEST=RECALI	11 min 15 sec
✓	# 433.6	JDK: openjdk8	NXF_VER=0.25.6 SGT_VER=2.3.1 PROFILE=singularit	11 min 29 sec
✓	# 433.7	JDK: openjdk8	NXF_VER=0.25.6 PROFILE=dockerTest TEST=REALIG	9 min 43 sec
✓	# 433.8	JDK: openjdk8	NXF_VER=0.25.6 SGT_VER=2.3.1 PROFILE=singularit	14 min 13 sec
✓	# 433.9	JDK: openjdk8	NXF_VER=0.25.6 PROFILE=dockerTest TEST=MAPPIN	6 min 14 sec

The List of People Involved

Sebastian DiLorenzo
Jesper Eisfeldt
Maxime Garcia
Szilveszter Juhos
Max Käller
Malin Larsson
Marcel Martin
Markus Mayrhofer
Monica Nistèr
Björn Nystedt
Pall Olason
Markus Ringnér
Pelin Sahlén
Johanna Sandgren
Teresita Díaz De Ståhl

Where to find us?

- We are on the SciLifeLab Slack
  `#cancer-pipeline`

Where to find us?

- We are on the SciLifeLab Slack
  `#cancer-pipeline`
- We have a gitter channel
  `https://gitter.im/SciLifeLab/CAW`

Where to find us?

- We are on the SciLifeLab Slack
  `#cancer-pipeline`
- We have a gitter channel
  `https://gitter.im/SciLifeLab/CAW`
- Our code is hosted on Github
  `https://github.com/SciLifeLab/CAW`

Any questions?

Any questions?