

Practical introduction to Singularity containers

November 11, 2021

Tom Harrop

tharrop@unimelb.edu.au

melbournebioinformatics.org.au

 [@MelBioInf](https://twitter.com/MelBioInf)

Reproducibility

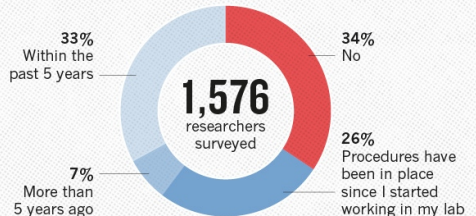
IS THERE A REPRODUCIBILITY CRISIS?



©nature

HAVE YOU ESTABLISHED PROCEDURES FOR REPRODUCIBILITY?

Among the most popular strategies was having different lab members redo experiments.



©nature

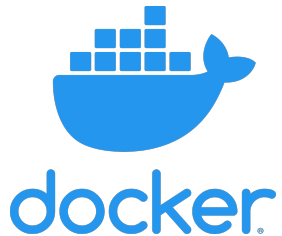
Analysis practices

Guidelines:

1. Don't modify raw data
2. Record the steps
(with version control)
3. Capture the computing environment

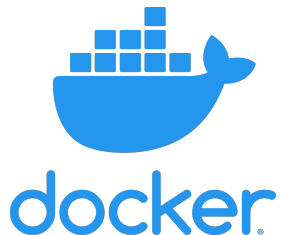
- ◆ Data management:
 - see Gayle Philip's slides
- ◆ Project management, source control:
 - Jessica Chung's slides
- ◆ Workflow management:
 - Bobbie Shaban's slides
- ◆ Reproducible computing:
 - use containers

Why use containers?



Motivation: Use software without installing it

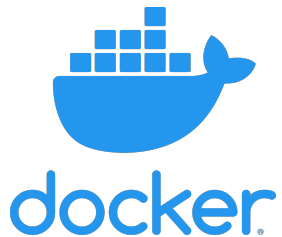
Why use containers?



Motivation: Use software without installing it

Benefits: Portability, reproducibility

Why use containers?



Motivation: Use software without installing it

Benefits: Portability, reproducibility

Drawbacks: Effort

Use software without installing it

On my computer:

```
# not provided for my OS  
guppy_basecaller --version
```

guppy_basecaller: command not found

Using a container:

```
singularity exec guppy_5.0.16.sif \  
guppy_basecaller --version
```

: Guppy Basecalling Software, (C) Oxford Nanopore

Get the version you need

On my computer:

```
# installed by my OS package manager  
samtools --version
```

```
samtools 1.11  
Using htlib 1.11-4  
Copyright (C) 2020 Genome Research Ltd.
```

Using a container:

```
singularity exec samtools_1.12.sif \  
samtools --version
```

```
samtools 1.12  
Using htlib 1.12  
Copyright (C) 2021 Genome Research Ltd.
```


Avoid tricky software installations

Dependencies

Funannotate has a lot of dependencies. However, it also comes with a few tools to help you get everything installed. The first is that of `funannotate check`. You'll see in the output below that the `fastx` tool is missing, which is Bill Pearson's `fastx` a dependency of the PAGA pipeline. Also the `STAR` and `SRANIMATE` variables are not set, that is because on this particular machine they are not installed. I.e. Funannotate will alert you at runtime if it is missing a dependency.

[illegible]

- lots of dependencies



🧩 not all dependencies are listed

Avoid tricky software installations

Dependencies

Farmanote has a lot of dependencies. However, it also comes with a few tools to help you get everything installed. The first is that of `farmanote check`. You'll see in the output below that the `twine` tool is missing, which is Bill Pearson's `twine` a dependency of the RASA pipeline. Also the `SHASOME` and `FRONTTRIM` variables are not set, that is because on this particular machine they are not installed. I.e. Farmanote will alert you at runtime if it is missing a dependency.

[illegible]

- lots of dependencies
- not all dependencies are listed
- need specific versions
- package managers don't always help

Avoid tricky software installations

Dependencies

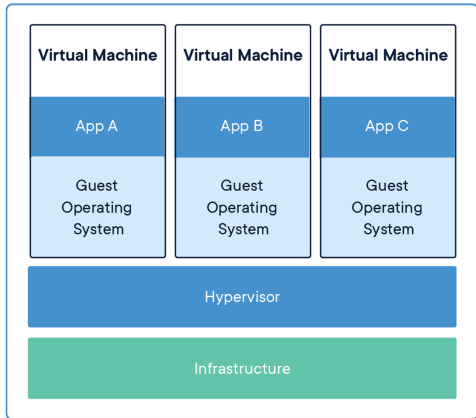
Farmanote has a lot of dependencies. However, it also comes with a few tools to help you get everything installed. The first is that of `funannotate check`. You'll see in the output below that the `twisto` tool is missing, which is Bill Pearson's `twisto` a dependency of the PASA pipeline. Also the `SHORE` and `SRKATrime` variables are not set, that is because on this particular machine they are not installed. I.e. Farmanote will alert you at runtime if it is missing a dependency.

[illegible]

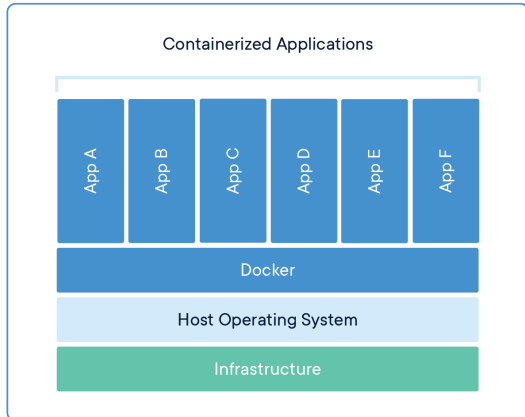
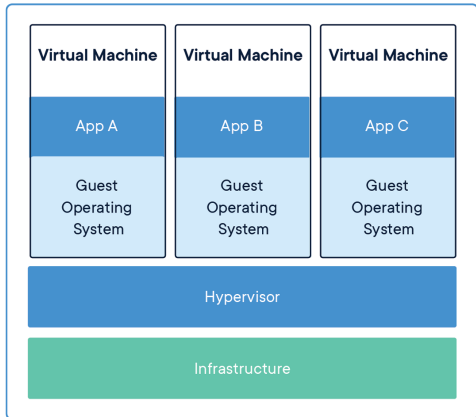
- lots of dependencies
- not all dependencies are listed
- need specific versions
- package managers don't always help

```
singularity run docker://nextgenusfs/funannotate
```

What is a container?



What is a container?



Docker, or Singularity?

Docker

- popular on cloud platforms
- effectively requires root to run*
- doesn't respect job schedulers*

* may have been fixed

Docker, or Singularity?

📦 Docker

- popular on cloud platforms
- effectively requires root to run*
- doesn't respect job schedulers*

* may have been fixed

📦 Singularity

- runs on HPC
- popular in research / edu?
- **supports Docker containers**

Docker, or Singularity?

📦 Docker

- popular on cloud platforms
- effectively requires root to run*
- doesn't respect job schedulers*

📦 Singularity

- runs on HPC
- popular in research / edu?
- **supports Docker containers**

* may have been fixed

Suggested workflow: build (or obtain) your containers in Docker format and use Singularity to run them

How to use Singularity

1. Install Singularity

- 💡 not too hard to install locally
- 💡 HPCs often have it installed
- 💡 you can ask the admin to install it

e.g. on Spartan:

```
$ module load singularity/3.7.3  
$ singularity --version  
singularity version 3.7.3
```

How to use Singularity

1. Install Singularity

- ✦ not too hard to install locally
- ✦ HPCs often have it installed
- ✦ you can ask the admin to install it

e.g. on Spartan:

```
$ module load singularity/3.7.3
$ singularity --version
singularity version 3.7.3
```

2. Run a container

- ✦ detailed instructions at singularity.hpcng.org
- ✦ basic commands is
`singularity exec {container URL} \`
`{command} {args}`
- ✦ usually need some options, e.g.
 - `-B` binds paths to make files available in the container
 - `--nv` enables GPU support

Workflow managers support containers

Snakefile:

```
rule trim_adaptors:
    input:          'data/raw_reads.fastq'
    output:         'output/trimmed.fastq'
    container:      'docker://my_repos/trim_adaptors:1.7'
    shell:          'trim_adaptors --raw_reads={input} > {output}'

rule run_assembly:
    input:          'output/trimmed.fastq'
    output:         'output/assembly.fasta'
    container:      'docker://my_repos/tidy_assembler:2.9'
    shell:          'tidy_assembler --reads={input} > {output}'
```

Workflow managers support containers

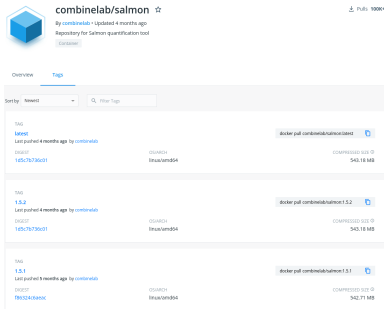
Snakefile:

```
rule trim_adaptors:
    input:          'data/raw_reads.fastq'
    output:         'output/trimmed.fastq'
    container:      'docker://my_repos/trim_adaptors:1.7'
    shell:          'trim_adaptors --raw_reads={input} > {output}'
rule run_assembly:
    input:          'output/trimmed.fastq'
    output:         'output/assembly.fasta'
    container:      'docker://my_repos/tidy_assembler:2.9'
    shell:          'tidy_assembler --reads={input} > {output}'
```

Run the workflow:

```
snakemake --use-singularity \
    --singularity-args ...
```

Where to find containers



combinelab/salmon ☆

By combinelab • Updated 4 months ago
Repository for Salmon quantification tool

1000+

Overview Tags

Sort by: Newest

Filter tags


TAG	DOCKER pull	COMPRESSED SIZE
latest Last pushed 4 months ago by combinelab Digest: 1d5c7b739cd1 OS/ARCH: linux/amd64	docker pull combinelab/salmon:latest	543.18 MB
1.5.2 Last pushed 4 months ago by combinelab Digest: 1d5c7b739cd1 OS/ARCH: linux/amd64	docker pull combinelab/salmon:1.5.2	543.18 MB
1.5.1 Last pushed 5 months ago by combinelab Digest: f8b32418a00c OS/ARCH: linux/amd64	docker pull combinelab/salmon:1.5.1	543.18 MB



Some developers provide docker containers

```
singularity pull \  
docker://combinelab/salmon:1.1.0
```

You can search for images on hub.docker.com

 dockerhub

[Explore](#) [Pricing](#) [Sign In](#) [Sign Up](#)

 Docker Containers Plugins**Filters**1 - 16 of 16 results for **trinityrnaseq**. [Clear search](#)

Best Match

☐ **Verified Publisher** ⓘ

☐ **Official Images** ⓘ
Official Images Published By Docker

Categories ⓘ☐ Analytics☐ Application Frameworks☐ Application Infrastructure☐ Application Services☐ Base Images **trinityrnaseq/trinityrnaseq-wdl** **50K+** **0**
Downloads Stars
By [trinityrnaseq](#) • Updated a year ago
Container Linux x86-64 **trinityrnaseq/trinityrnaseq** **10K+** **4**
Downloads Stars
By [trinityrnaseq](#) • Updated 2 months ago
Trinity Docker
Container Linux x86-64

13 / 22

Check with the software developer

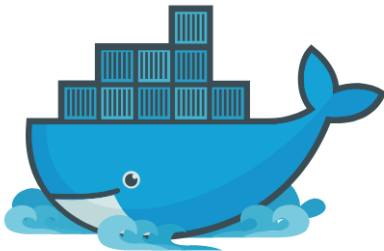
trinityrnaseq / trinityrnaseq Public

<> Code 29 Issues Pull requests 3 Actions Projects Wiki Security Insights

Trinity in Docker

Brian Haas edited this page on 1 Apr 2020 · 13 revisions

Run Trinity Using Docker



If you have [Docker](#) installed, you can pull [our image from DockerHub](#), which contains Trinity and all software used for downstream analyses supported within the larger Trinity framework.

Build them yourself!

TomHarrop / container-flye Public

generated from [deardenlab/container-build-template](#)

Unwatch 1

Star 0

Fork 0

<> Code

Issues

Pull requests

Actions

Projects

Wiki

Security

Insights

Settings

main 1 branch 2 tags

Go to file Add file Code

TomHarrop Create README.md ✓ d8caa3a 16 days ago 11 commits

.github/workflows	Initial commit	16 days ago
Dockerfile.2.9	certificates	16 days ago
README.md	Create README.md	16 days ago
VERSION	2.9	16 days ago

README.md

Singularity and Docker containers for [Flye](#).

About

No description, website, or topics provided.

Readme

Releases 2

Release v2.9 Latest 16 days ago

+ 1 release


Packages 1

container-flye

Languages

Roff 100.0%


Build them yourself!


 **TomHarrop** / **container-flye** Public

generated from [deardenlab/container-build-template](#)

Unwatch 1 Star 0 Fork 0

[Code](#) [Issues](#) [Pull requests](#) [Actions](#) [Projects](#) [Wiki](#) [Security](#) [Insights](#) [Settings](#)

 **container-flye**

 Install from the command line:

```
$ docker pull ghcr.io/tomharrop/container-flye:2.9
```

[Learn more](#)

Recent tagged image versions

2.9 ↓ 6


Published 16 days ago · Digest ...


[View and manage all versions](#)


README.md

Singularity and Docker containers for [Flye](#).

Details

 TomHarrop

 container-flye

 Readme

Last published

16 days ago

Issues


0

Total downloads

9

Collaborators

1

 TomHarrop Tom Harrop

[Open an issue](#)

[Package settings](#)

Some barriers to container usage

- ❖ Building containers yourself can be painful
- ❖ Duplication of effort
- ❖ Some software **shouldn't** go in a container because of “unfortunate licensing issues”
 - DTU software e.g. rnammer, tmhmm
 - GATech: GeneMark
 - GIRInst's RepBase
- ❖ Getting Singularity installed

When to use containers

Modules:

Conda, venv, *etc.*:

Containers:

When to use containers

Modules:

```
module load \  
  samtools/1.12
```

Conda, venv, etc.:

Containers:

- 📦 installed by the sysadmin
- 📦 properly configured
- 📦 easy

When to use containers

Modules:

```
module load \  
samtools/1.12
```

- ❖ installed by the sysadmin
- ❖ properly configured
- ❖ easy

Conda, venv, etc.:

```
conda install -c \  
bioconda samtools
```

- ❖ latest packages
- ❖ testing in a specific environment

Containers:

When to use containers

Modules:

```
module load \  
samtools/1.12
```

- ❖ installed by the sysadmin
- ❖ properly configured
- ❖ easy

Conda, venv, etc.:

```
conda install -c \  
bioconda samtools
```

- ❖ latest packages
- ❖ testing in a specific environment

Containers:

```
singularity run \  
samtools.sif
```

- ❖ reproducibility
- ❖ collaboration
- ❖ portability
- ❖ performance?

Get started with containers

◆ How I use containers:

1. **Write Docker recipes** (one tool, one recipe)
2. Remotely build Docker image store it permanently
3. Use with Singularity via a workflow manager

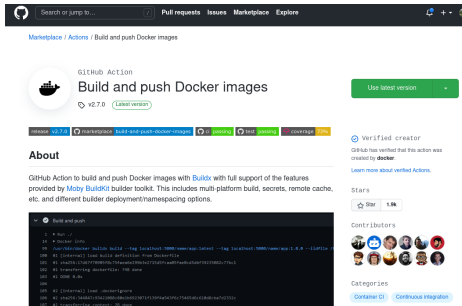
◆ Singularity has good documentation: [quick start](#)

◆ Docker documentation is ... extensive: [Best practices for writing Dockerfiles](#)

◆ Integrating with workflow managers is easy

Building containers on GitHub

TomHarrop/container-build-template



The screenshot shows the GitHub Actions interface for the repository 'TomHarrop/container-build-template'. The main heading is 'Build and push Docker images' with a 'v2.7.0' label indicating the latest version. A green button labeled 'Use latest version' is present. Below the heading, there's a section titled 'About' which states: 'GitHub Action to build and push Docker images with **Buildx** with full support of the features provided by **Moby BuildKit** builder toolkit. This includes multi-platform build, secrets, remote cache, etc. and different builder deployment/namespacing options.' A code block shows a snippet of the workflow file:

```
name: Build and push
on:
  push:
    branches: [ main ]
  pull_request:
    branches: [ main ]
jobs:
  build:
    runs-on: ubuntu-latest
    steps:
      - name: Checkout repository
        uses: actions/checkout@v2
      - name: Set up Docker Buildx
        uses: docker/setup-buildx-action@v1
      - name: Login to Docker Hub
        uses: docker/login-action@v1
        with:
          username: ${{ secrets.DOCKER_USERNAME }}
          password: ${{ secrets.DOCKER_PASSWORD }}
      - name: Build and push
        uses: docker/build-push-action@v2
        with:
          context: .
          push: true
          tags: ${{ secrets.DOCKER_USERNAME }}/container-build-template:latest
```

On the right side, there's a 'Verified creator' badge for 'docker', a 'Stars' section showing 1.5k stars, and a 'Contributors' section with avatars of several people. At the bottom, there are 'Categories' listed as 'Container CI' and 'Continuous Integration'.

Docker and Singularity example

TomHarrop/container-flye

- 💡 generally use the Docker container in Singularity
- 💡 some old versions can't login to GHCR

Is this useful?

In your experience do developers appreciate pull requests so that everyone has access to the container from the official page - or do containers take a bit of time to maintain so developers would be less keen?

- question from Alistair Legione

- ◆ do enough people use these setups?
- ◆ is Conda preferred?
- ◆ PRs back to the software developer?