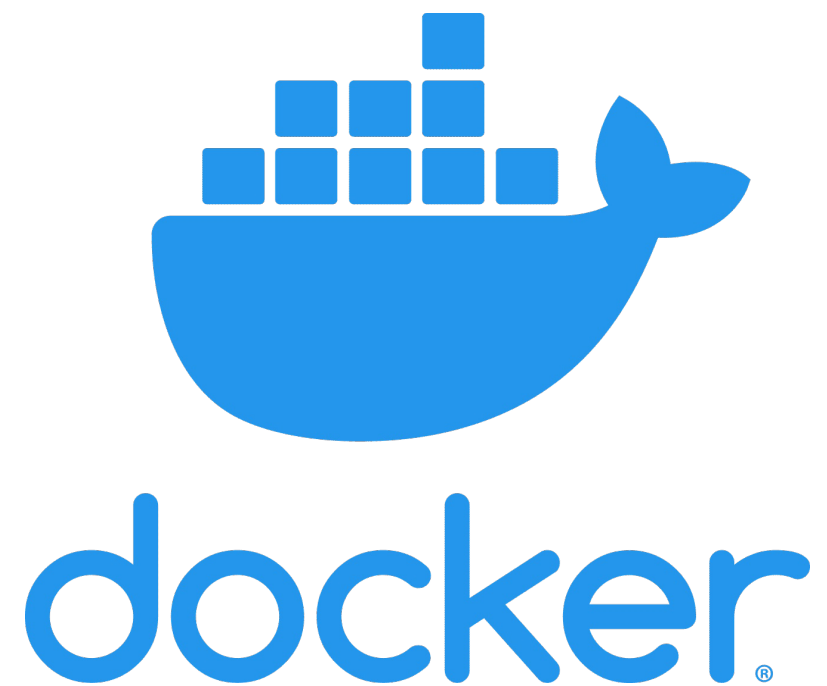


Making self-contained, distributable projects with



What is Docker?

- Standardized packaging for software and dependencies
- Docker lets you create and run applications securely isolated in a container, packaged with all its dependencies and libraries.

Docker in lifescience

Docker containers can be used to allow others to reproduce a complete analysis:

To further support reproducibility, we bundled all tools and dependencies into one Docker container available on DockerHub [19]. `docker run` executes the aforementioned Makefile inside the container.

(click [here](#) to see the Dockerfile)

```
docker run metagenomics/2015-biogas-cebitec
```

Bremges et al GigaScience (2015) 4:33, [doi:10.1186/s13742-015-0073-6](https://doi.org/10.1186/s13742-015-0073-6)

Docker in lifescience

Docker can also be used to define software environments and settings for benchmarking studies

CAMI Challenge: an independent evaluation of several tools in the field of metagenomics

... we defined standards for submitting the software itself, along with parameter settings and required databases and implemented them in Docker container templates...

Sczyrba et al Nature Methods (2017) 14:1063–1071, doi.org/10.1038/nmeth.4458

Docker in lifescience

Docker can also be used to define software environments and settings for benchmarking studies

2nd CAMI Challenge

For reproducibility, participants could submit a Docker container containing the complete workflow, a bioconda script or a software repository with detailed installation instructions...

Meyer et al Nature Methods (2022) 19:429-440, <https://doi.org/10.1038/s41592-022-01431-4>

Docker nomenclature

- A Docker **file** is a recipe used to build a Docker **image**
- A Docker **image** is a standalone executable package of software
- A Docker **container** is a standard unit of software run on the Docker Engine.
- **Docker Hub** is an online service for sharing docker images

Images and containers

Example: Use a different OS

Check OS on local machine

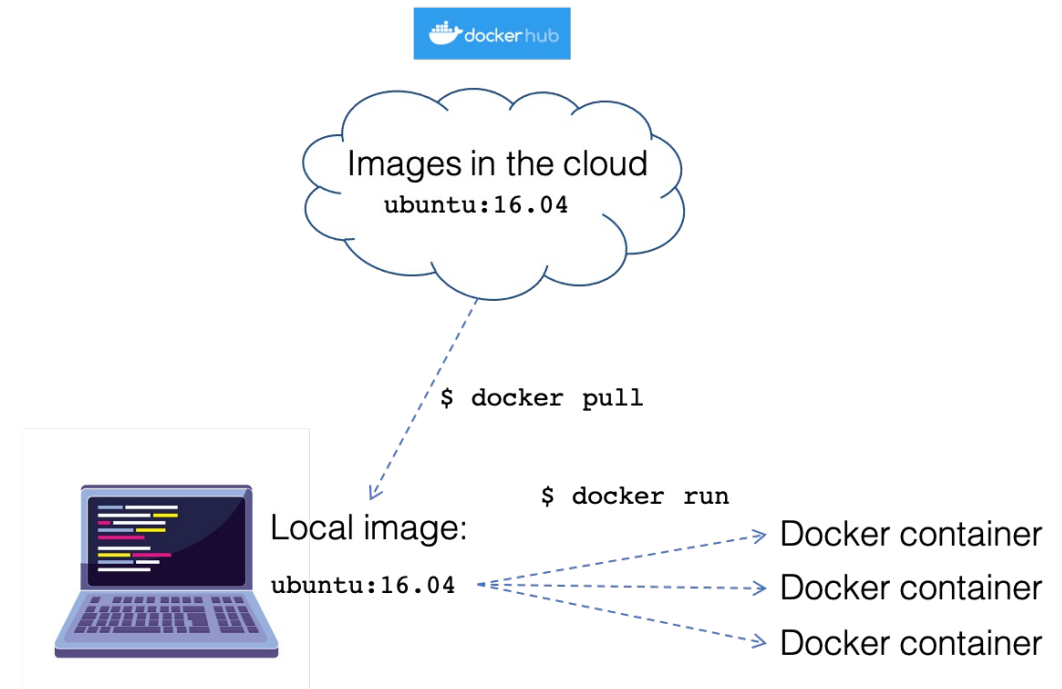
```
$ uname -a
Darwin johnsmbp.local 19.6.0 \
Darwin Kernel Version 19.6.0: \
Tue Jan 12 22:13:05 PST 2021; \
root:xnu-6153.141.16~1/RELEASE_X86_64 x86_64
```

Pull Ubuntu (Linux) image

```
$ docker pull ubuntu:16.04
16.04: Pulling from library/ubuntu
22dc81ace0ea: Pull complete
...
Digest: sha256:e348fbbea0e0a0e73ab0370de151e7800684445c509d46195aef73e090:
Status: Downloaded newer image for ubuntu:16.04
```

Run a container and check OS version

```
$ docker run -it ubuntu:16.04
root@407b0fd13fe5:/# uname -a
Linux 407b0fd13fe5 4.9.60-linuxkit-aufs #1 SMP Mon Nov 6 16:00:12 UTC
2017 x86_64 x86_64 x86_64 GNU/Linux
```



Mounting volumes

```
$ docker run -it -v $PWD/data:/home/data ubuntu:16.04
```

Local project directory:

```
$ ls project/
```

```
| - doc/
|
| - data/
|   | - raw_external/
|   | - raw_internal/
|   | - meta/
|
| - code/
| - notebooks/
|
| - intermediate/
| - scratch/
| - logs/
|
| - results/
|   | - figures/
|   | - tables/
|   | - reports/
|
| - Snakefile
| - config.yml
| - environment.yml
| - Dockerfile
```

Docker image file system

```
$ ls /
```

```
| - bin/
| - boot/
| - dev/
| - etc/
| - home/
|   | - data/ # data folder mounted under home/
|     | - raw_external/
|     | - raw_internal/
|     | - meta/
| - lib/
| - lib64/
| - media/
| - opt/
| - proc/
| - root/
| - run/
| - sys/
| - tmp/
| - usr/
| - var/
```


What can I use Docker for?

- As an advanced **environment** manager

```
docker run -it -v $PWD:/home my_image /bin/bash
```

- To package your **code** with the **environment** it needs

```
docker run \  
-v $PWD/data:/home/data \  
-v $PWD/results:/home/results \  
my_image snakemake report.pdf
```

- To package a whole workflow with **environment**, **code** and **data** (e.g. to accompany a manuscript).

```
docker run \  
-v $PWD/results:/home/results \  
my_image snakemake report.pdf
```

- and much more...

What is Singularity?



SINGULARITYCE

an open source container platform suitable for HPC clusters

<https://sylabs.io/guides/latest/user-guide/>

Singularity was created to run complex applications on HPC clusters in a simple, portable, and reproducible way.

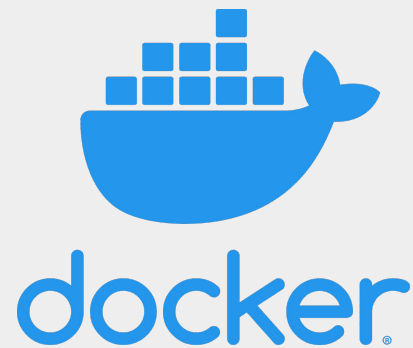
What is Apptainer?



In 2021 Singularity joined the Linux Foundation and became Apptainer.

The company Sylabs still maintains Singularity

Docker vs. Singularity



runs as a daemon process with superuser privileges

runs as regular user

images stored centrally

image files that you can move around. No layers!

isolates the host and container file system by default

containers have access to host file system

well supported on Mac/Linux/Windows

limited support on Mac/Windows

Questions?