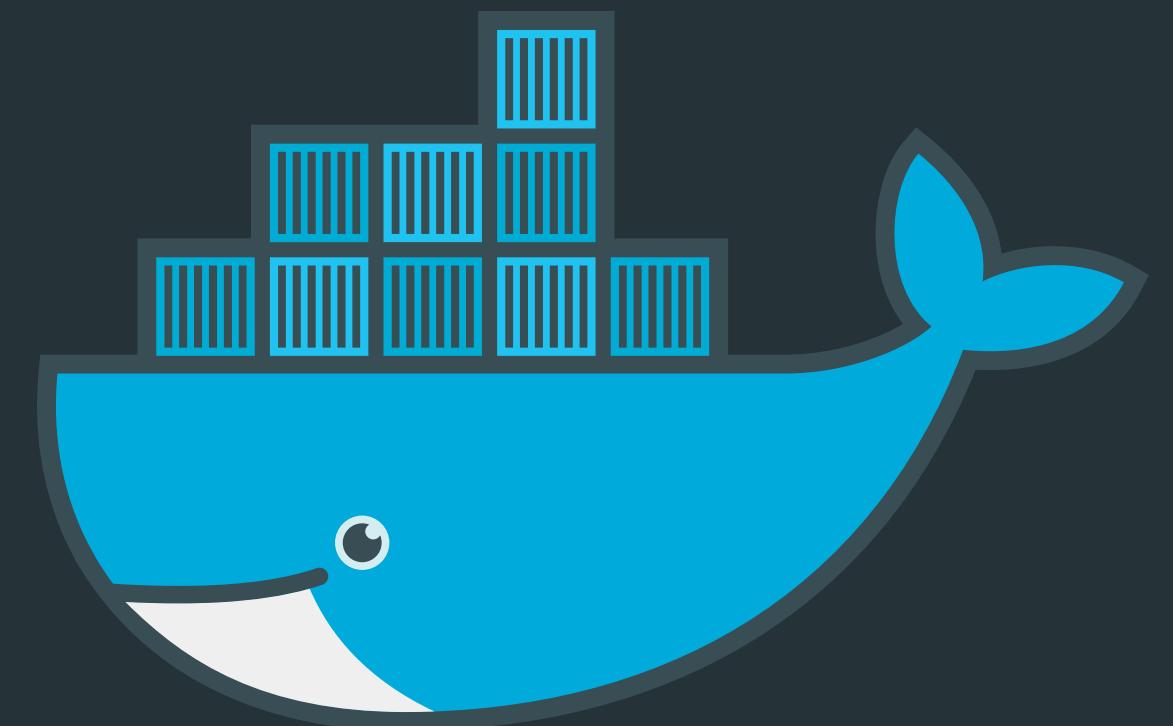


# Docker & radio astronomy



dockercon14 | eu

# Who Am I

Gijs Molenaar

Scientific Software Engineer in NL

PhD student in SA

MSc Artificial Intelligence

@gijzelaerr

<http://pythonic.nl>





# THE PROBLEM

some scientist ***Q*** wants to crunch

some data ***X*** with

some software ***Y*** on

some platform ***Z***

# Scientist Q

Smart person

Self educated on the field of software engineer

Gets the job done

Good at duct tape programming



# Data X

Really big data

Datasets of peta, exa bytes

LOFAR - 50 GB/s

SKA - 10 times global Internet traffic



# Software Y

Complex math heavy software

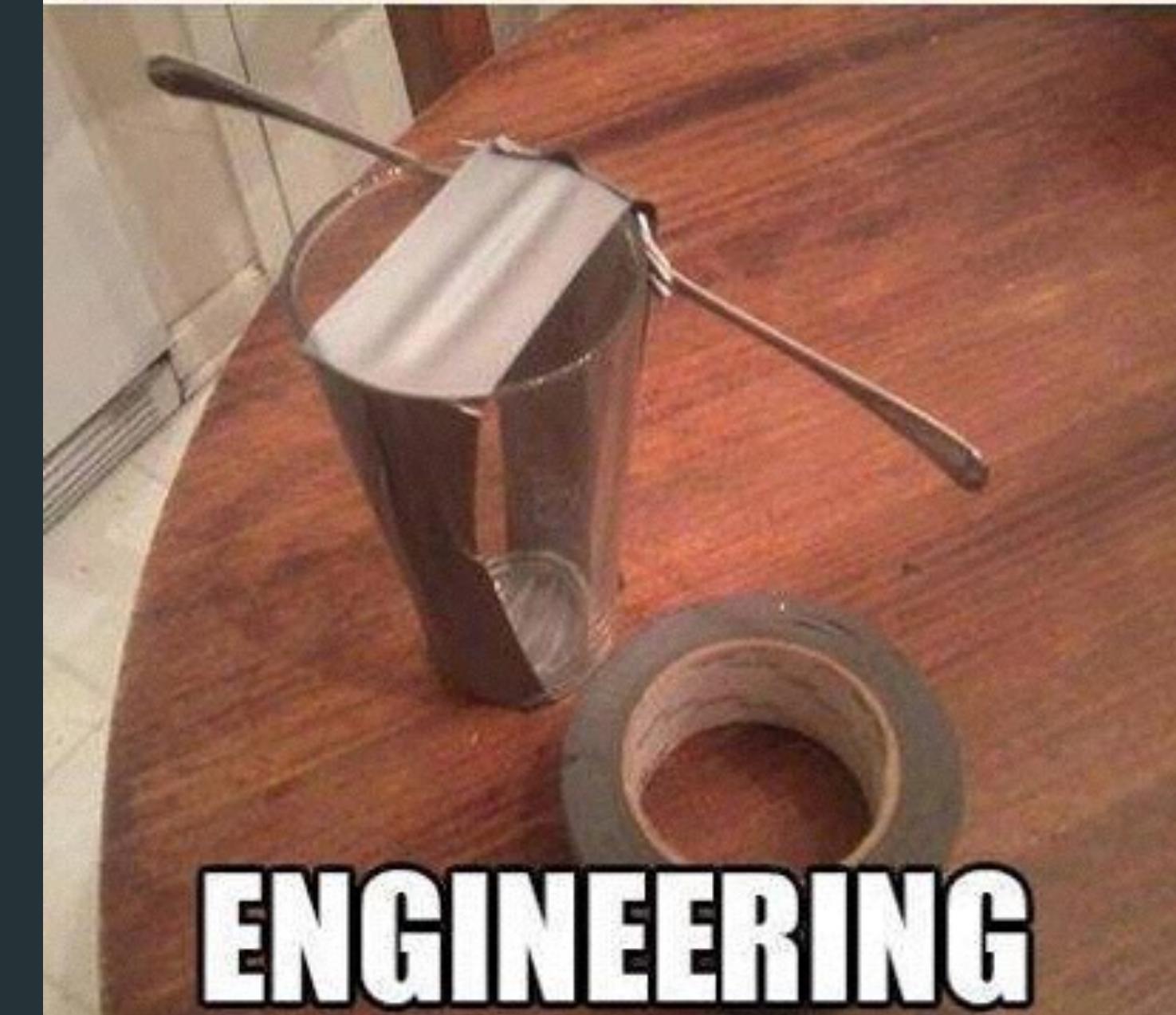
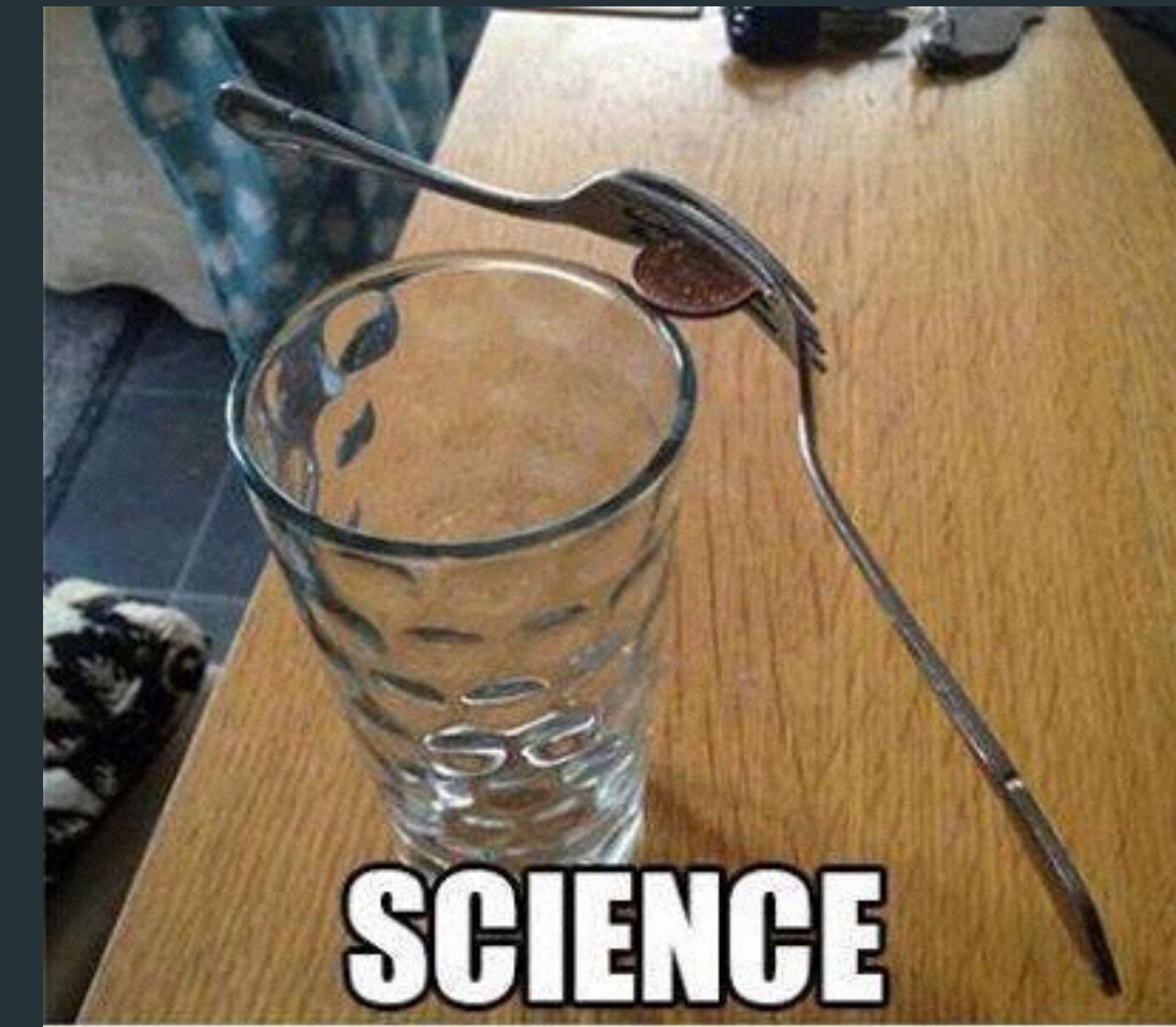
Written during PhD of scientist Q

Often bad coding practices

(Unit) tests?

Version control?

‘Runs on my computer’



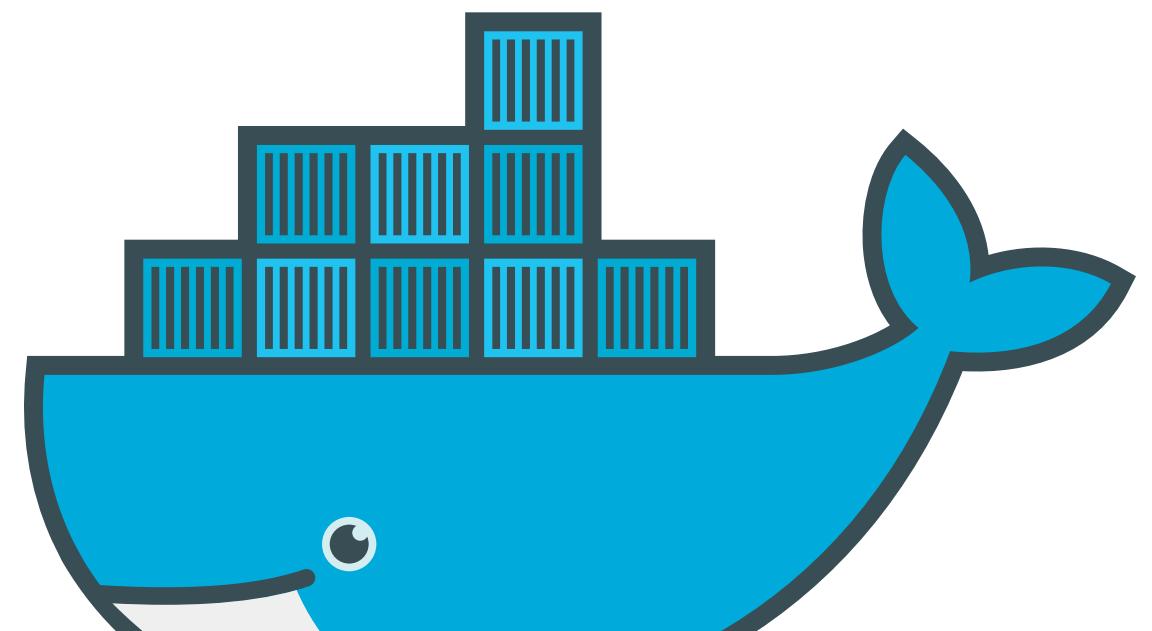


# Google Cloud Platform



 **amazon** web services™ | **EC2**

 **openstack**™  
CLOUD SOFTWARE



# What is docker?

‘Containing’ software

Managing containers

Connecting containers

Deploying containers



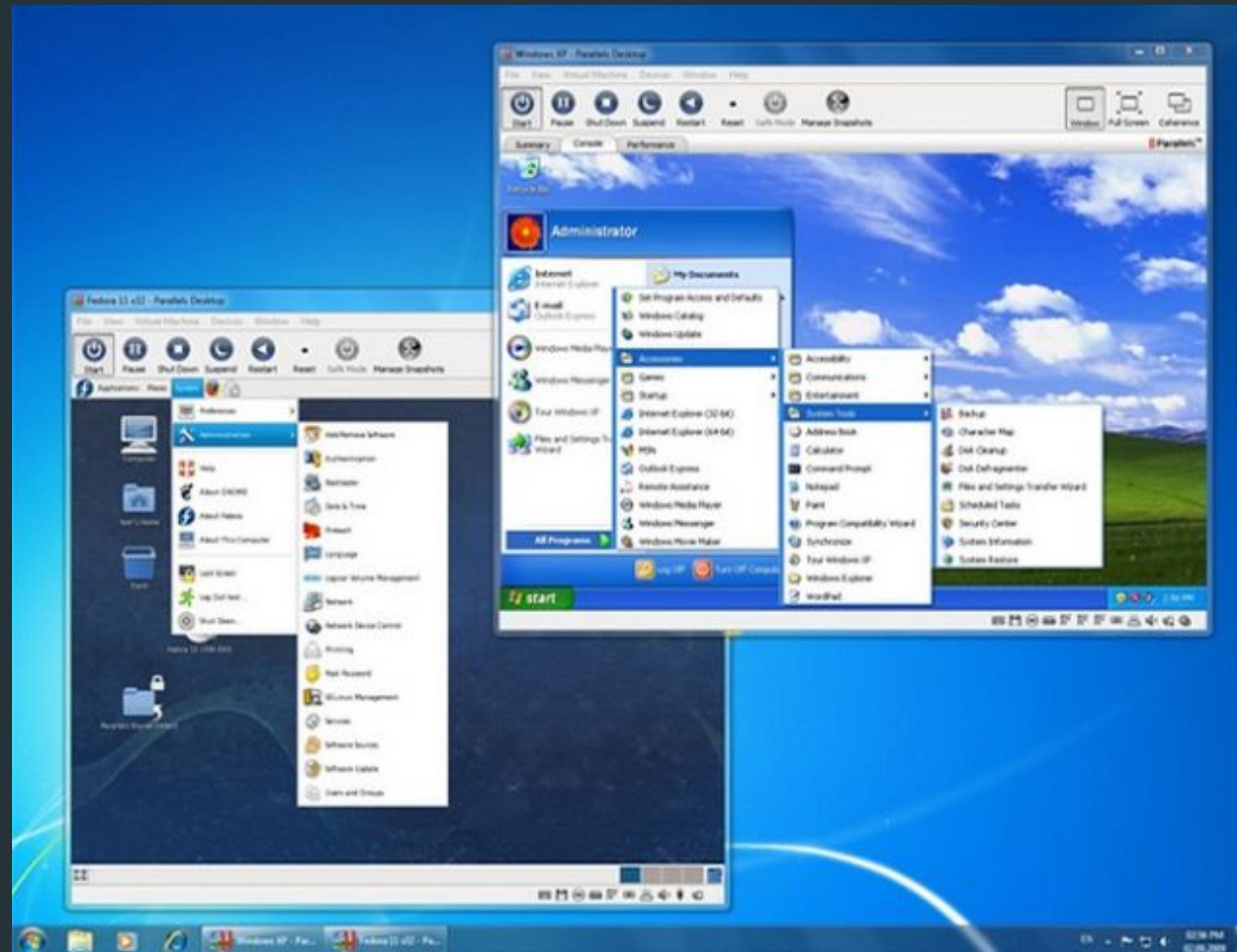
# Like virtual machines

But not

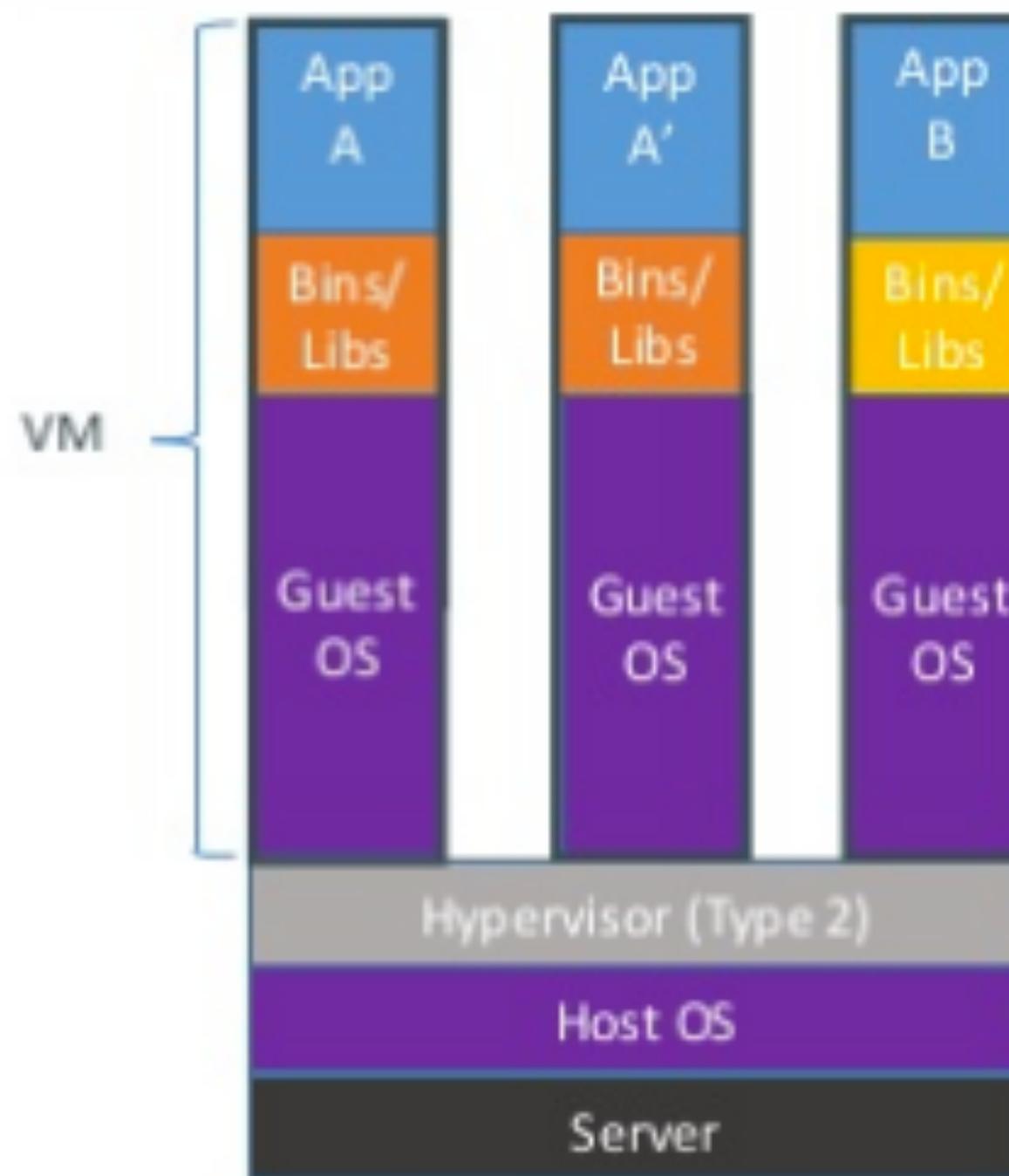
no emulation

Process isolation & virtual networking

More like chroot/jail

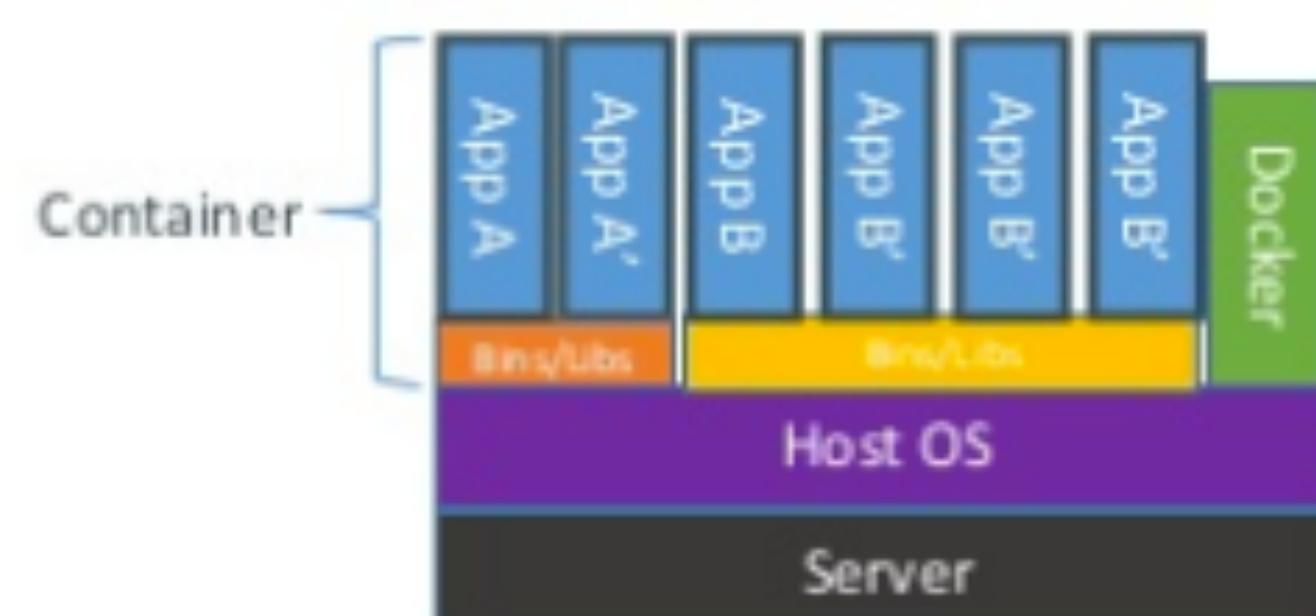


# Containers vs. VMs

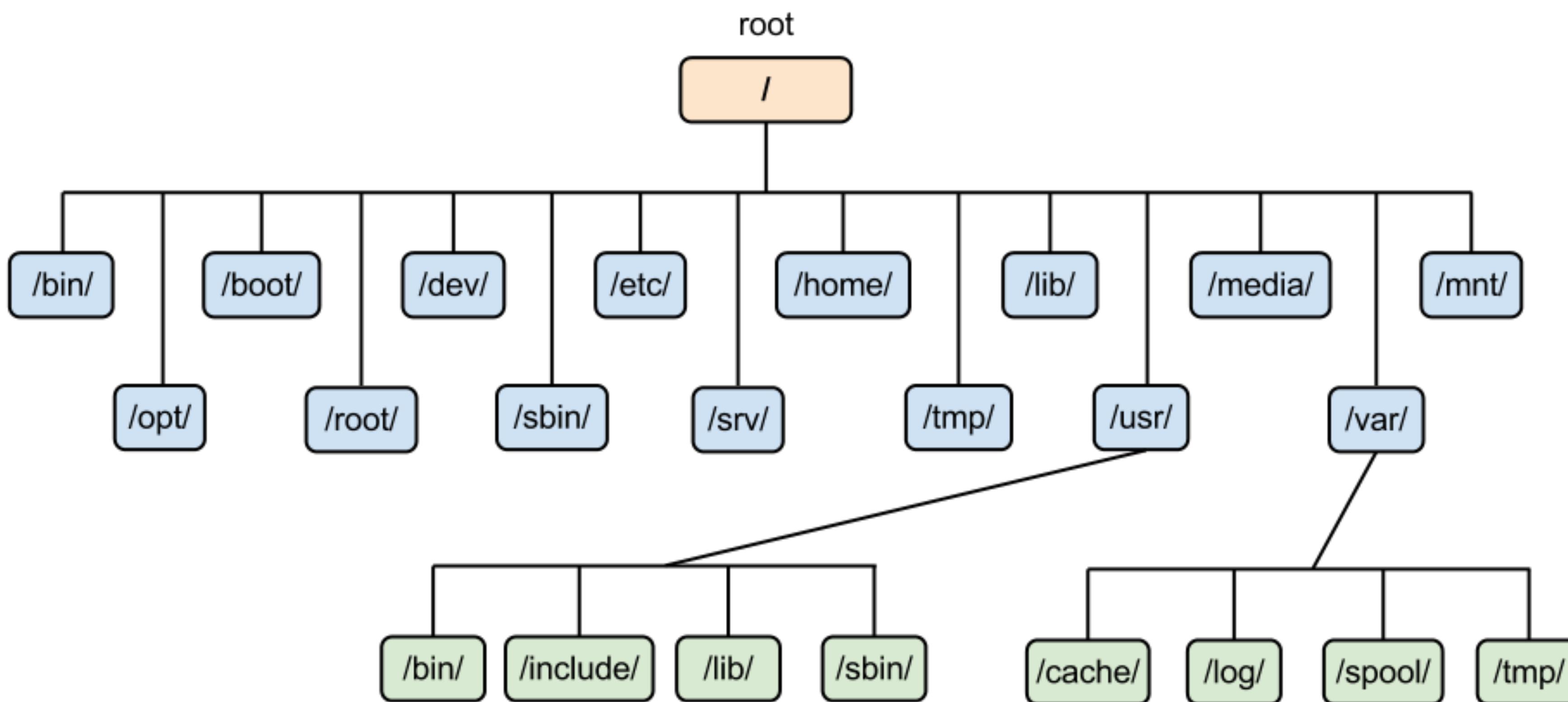


Containers are isolated,  
but share OS and, where  
appropriate, bins/libraries

...result is significantly faster deployment,  
much less overhead, easier migration,  
faster restart



# Docker Image



# Image vs container

## **Image**

complete filesystem

set of ‘layers’

read-only

stateless

build with Dockerfile

or commit container

or just tarball

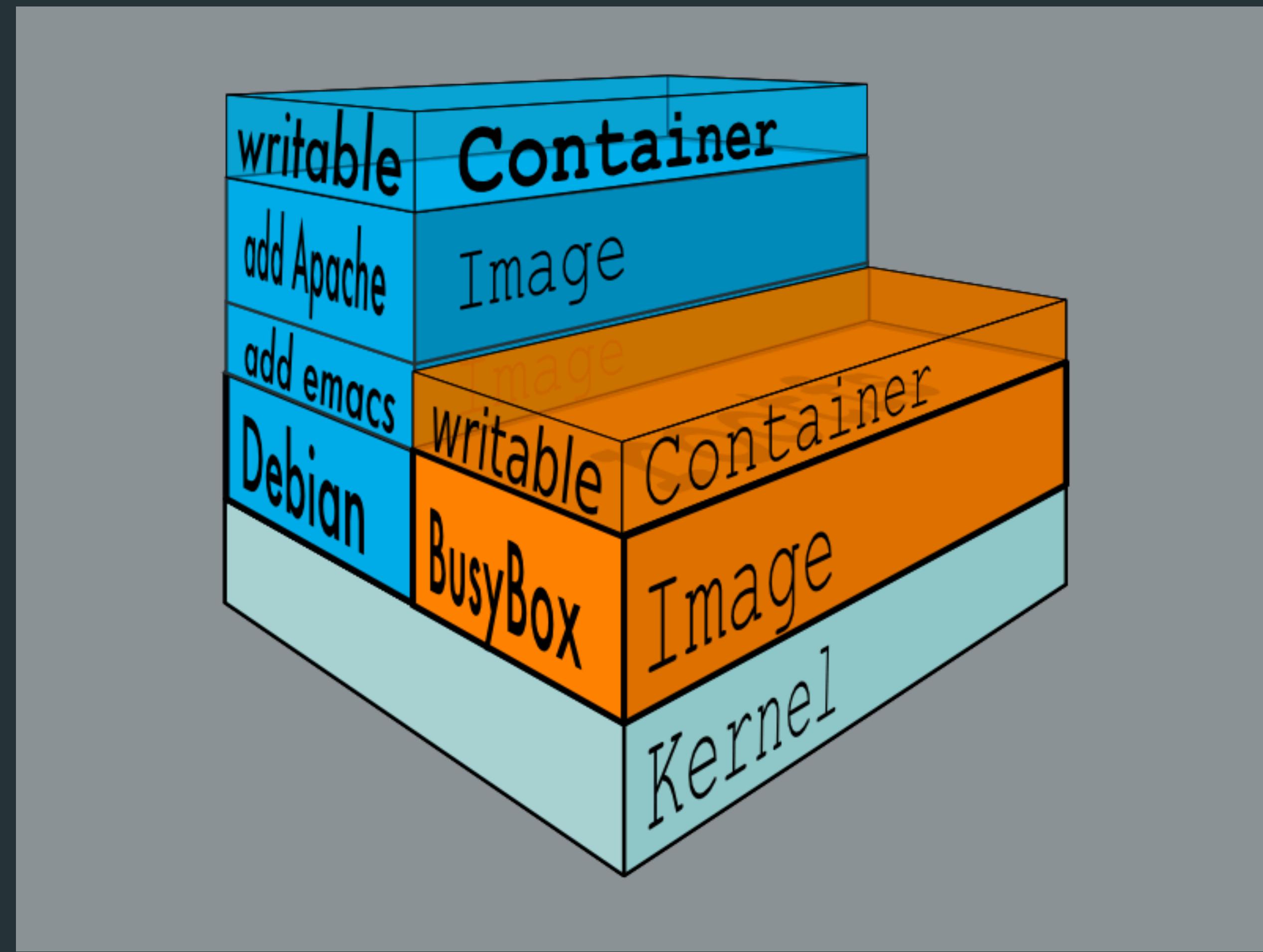
## **container**

instantiation of image

stateful

image + container layer

# Layered filesystem (unionFS)



# Dockerfile

```
FROM ubuntu:14.04
MAINTAINER gijs@pythonic.nl

ENV DEBIAN_FRONTEND noninteractive

# enable universe, multiverse, restricted with world wide mirrors
ADD apt.sources.list /etc/apt/sources.list
RUN apt-get update
RUN apt-get upgrade -y
RUN apt-get install -y software-properties-common

## add radio astro PPA (with source packages)
RUN add-apt-repository -ys ppa:radio-astro/main
RUN apt-get update
```

<https://github.com/radio-astro/docker-images>



# Dockerfile 2

```
FROM radioastro/base

ADD debian_packages /
ADD python_packages /

# install debian packages
RUN apt-get update && cat /debian_packages | xargs apt-get install -y

# install all python modules
RUN pip install -r /python_packages

## Expose the ipython notebook port
EXPOSE 8888

VOLUME /notebooks

## Run ipython notebook
CMD /usr/bin/python /usr/local/bin/ipython notebook \
-ip=*--notebook-dir=/notebooks --pylab inline
```



# Fig

db:

```
image: postgres:9.3
```

broker:

```
image: dockerfile/rabbitmq
```

worker:

```
build: django_kat
```

```
command: celery worker -A django_kat
```

links:

- broker
- db

django:

```
build: django_kat
```

```
command: uwsgi--module django_kat.wsgi
```

links:

- db
- broker

volumes:

- /code

volumes\_from:

- worker

viewer:

```
image: gijzelaerr/cyberska_viewer
```

volumes\_from:

- worker



# Docker hub

<https://registry.hub.docker.com/repos/radioastro/>

upload docker image

automated build from github repo



# 3 use cases for docker (for us)

software delivery

build software solutions

reproducible science



OSX



# packaging



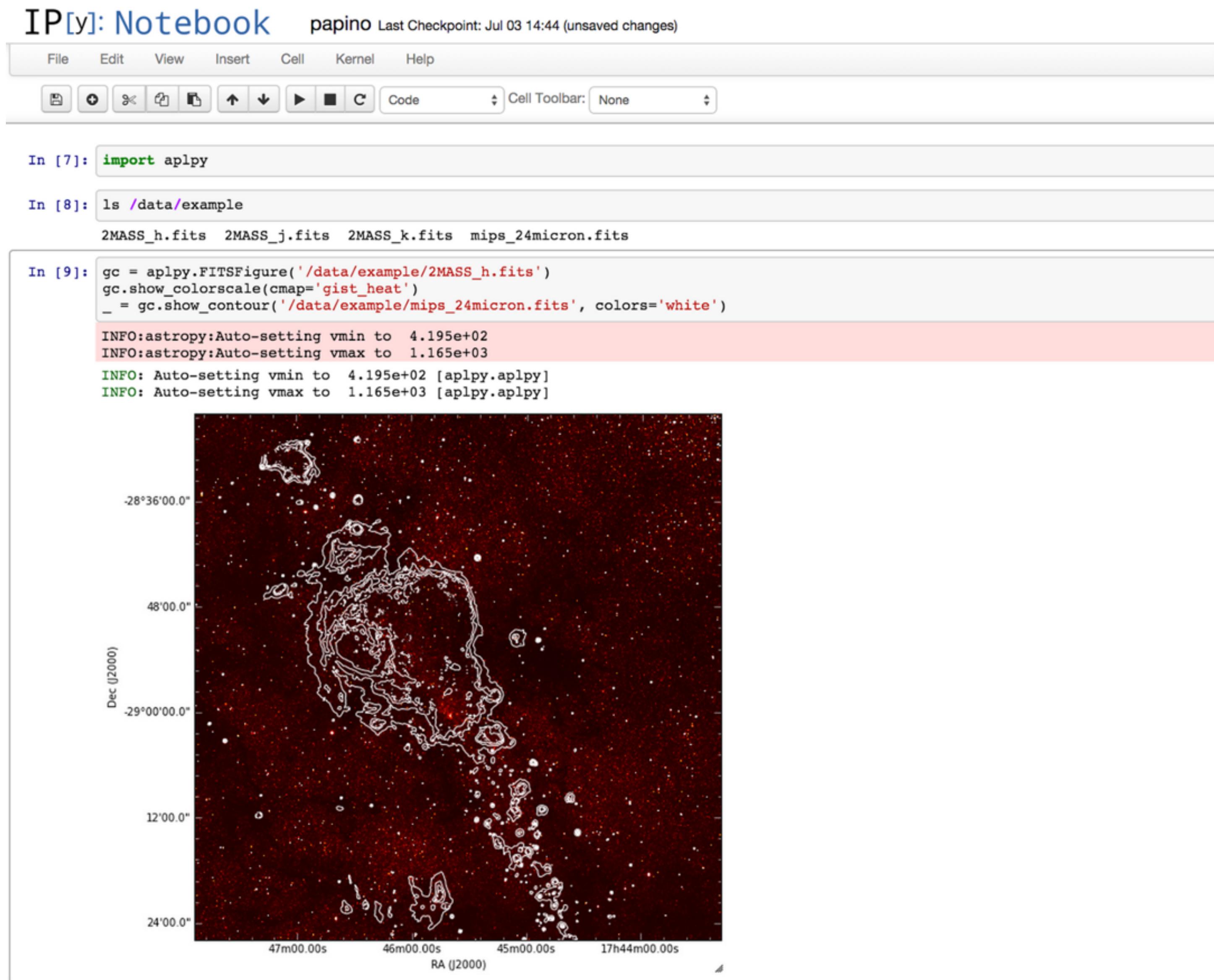
ubuntu



launchpad

<http://github.com/radio-astro/docker-images>  
<http://github.com/radio-astro/packaging>

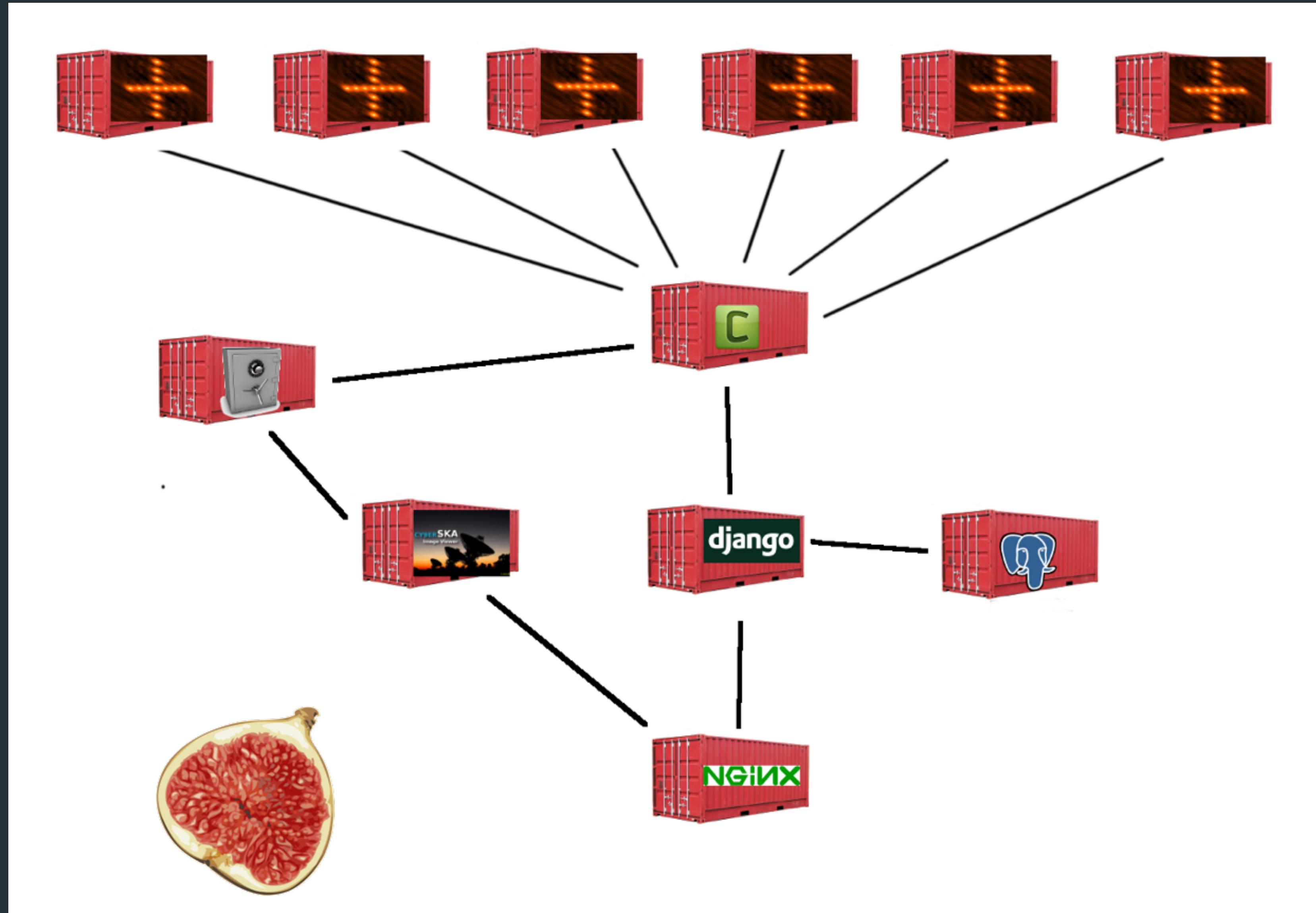
```
$ docker run -dP radioastro/notebook
```



# R.O.D.R.I.G.U.E.S.

**R**ATT  
**O**nline  
**D**econvolved  
**R**adio  
**I**mage  
**G**eneration  
**U**sing  
**E**clectic / **E**citing / **E**xperimental /  
**E**xotic /**E**xtraordinary / **E**xistential /  
**E**phemeral / **E**soteric  
**S**oftware

# Rodrigues





# Problems

security - trusted users only

<https://github.com/docker/docker/issues/6324>

container size

# Upcoming features

<https://github.com/docker/machine>

<https://github.com/docker/swarm>

# Tips & tricks

keep containers small

Don't build VM's

Make single purpose containers and link them together (fig)

Use environment variables to configure containers

Or mount config files in them

**docker exec** is really useful

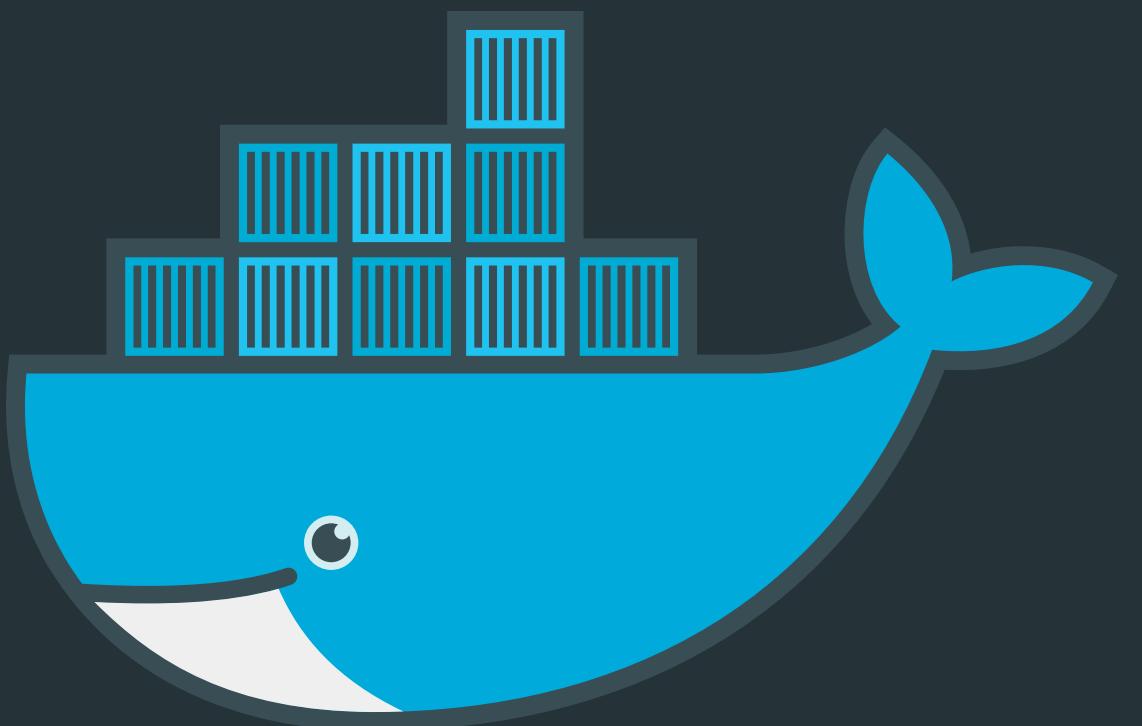


# future work

- Use containerisation as base for scientific pipelines
- Thinking about pipeline specific DSL using this graph based computing based higher order language
- Pyxis 2?
- Collaborate with other fields?

# Thank You.

@gijzelaerr - <http://pythionic.nl>



dockercon14 | eu