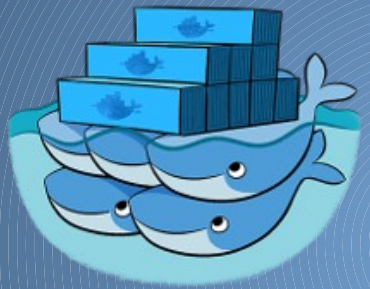
Docker for Data Science

R-Ladies Cape Town
Apr 17 2018

Dr Monika Obrocka
@MonikaObrocka

Docker Components



pinsleepe.github.io

Grand Vision

pinsleepe.github.io

Outline

- Why use Docker?
  - Terminology
- Docker Hub
  - Useful Commands
- Usecases ← interesting bit
- Your usecase?

pinsleepe.github.io

My story with Docker

Joined new team

Given half finished project

Spaghetti code

Config files? Virtual environment?

Never finished it....

pinsleepe.github.io

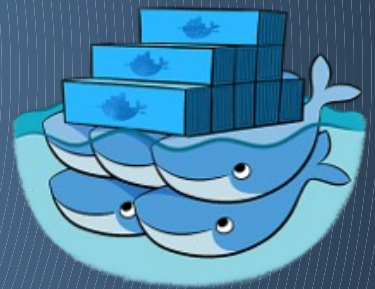
Why use Docker?

It provides everything an application needs to run, like file system and software applications.

Low barrier to entry. Yes.

Ecosystem with multiple external tools.

Not only for DevOps



Terminology

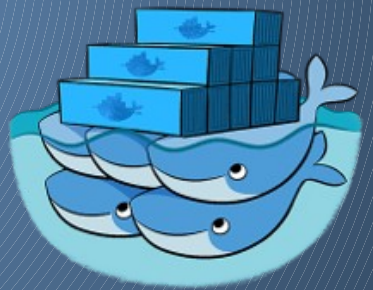
Dockerfile — a text file with the build instructions

↓

Image — essentially a disk image.

↓

Container — a running image.



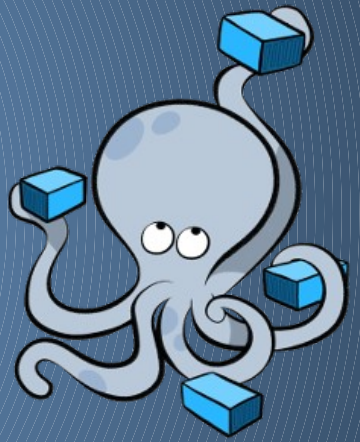
Useful Commands

docker search and **docker pull**

docker run and **docker stop/start**

docker rm and **docker rmi**

docker ps and **docker images**



Docker Hub

App store for Docker images

No Dockerfile necessary - only Docker Hub account

Major features:

- Image repositories
- Automated builds
- Organizations with work groups
- GitHub and Bitbucket integration

Go to Hub

Usecases + live demo

- Rstudio
- TensorFlow with Jupyter
- Python
- TimescaleDB

pinsleepe.github.io

RStudio

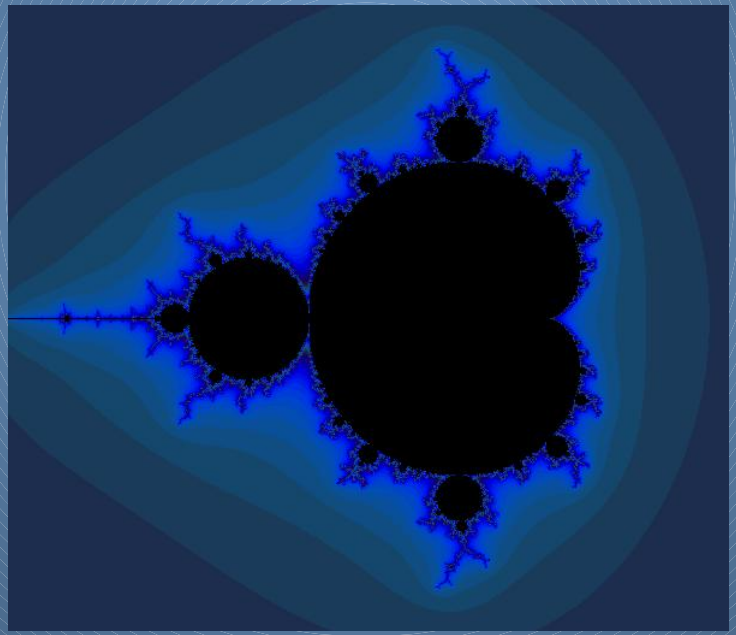
rocker a collaboration between Carl Boettiger and Dirk Eddelbuettel

Demo...

TensorFlow with Jupyter

Jupyter notebooks are a popular option for sharing data science workflows

Demo...



pinsleepe.github.io

Python

python

Demo...

pinsleepe.github.io

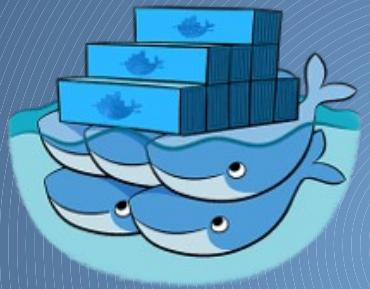
TimescaleDB

timescale

Demo...

pinsleepe.github.io

Next time!



pinsleepe.github.io