Docker pour les ENSAE

Python pour un Data Scientist

Romaric SOKOUNDJOU TAKAM

Data Scientist @VUble

October 2019

About me

- Data Scientist by day @VUble (my opinions are my own, not those of my employer)

- Unix enthusiast, hobbyist mobile (flutter) and web (django) programmer by night

Outline

- What's Docker ?
- Try to make some code work
- Try one more time, with docker
- Some comparisons
- Some commands

Let's run a toy example

- download source code on my github account : git clone https://github.com/rorosan/docker_ensae.git
- try to run it on your pc

Cumbersome, right ?

One more time, with docker ?

the theory

- Download docker
- Add user to docker groups
- Write a dockerfile
- Build your image
- Launch a container
- Start to work !

One more time, with docker ?

some practice now !

- docker build -t image_name:tag_name . (don't forget the full stop !)
- docker run image_name:tag_name

Docker vs virtualEnv

- only python vs other applications
- interoperable
- installing isn't a nightmare
- much more easier to use

Docker vs Virtual Machines

- Quicker to build, test and deploy
- Much, much, much, much more lighter (Gb vs tens of MB)
- More interoperable

But too light ?

- you may need to mount filesystems
- building heavier apps (webservers and database like postgres, ...) is more challenging

Not far from ideal for Data Engineers/Scientists

- Mostly small BATCH ETL Jobs
- Micro services are truly adapted to our workloads
- Containarized algotithms
- Plenty of dockerfiles online !
- Launch a container
- Start to work ![?].

TLDR

- Docker is great !

- Better, Faster, Stronger applications and projects !

What's next ?

- Write dockerfiles, build images and launch containers !

- KUBERNETES, Docker Swarm, Mesos when dealing with production in tech firms :)

That's all Folks !

Bibliographie


Simon G., Vivien P., ingénieurs système, VUble


Documentation officielle Docker.


Will Sargent, Cheat Sheet docker