RUNNING DOCKER ON SURFSARA'S HPC SYSTEMS WITH SINGULARITY
Let's breakdown the title
Running Docker on SURFsara's HPC systems with Singularity
Running Docker on SURFsara's HPC systems with Singularity
Running Docker on SURFsara's HPC systems with Singularity
Running Docker on SURFsara's HPC systems with Singularity

I WILL NOT TALK ABOUT
I WILL TALK ABOUT
ABOUT ME

- Name: Maarten Kooyman
- MSc Bioinformatics + MSc Genetic Epidemiology
- Working for 3 years at SURFsara Distributed Data-processing Group
- Technical lead for ProjectMinE@SURFsara
- 2 children + wife
- 34 years
STRUCTURE OF THIS TALK

- Why Containers?
- Why Singularity?
- A real life use case
- Use Cases Singularity
- Singularity at SURFsara
- Acknowledgements
CONTAINERS
WHY CONTAINERS?
Reproducibility
Portability
Easy to use
WHY SINGULARITY?
Right tool for the right job.
Biggest ecosphere
A REAL LIFE USE CASE
Expansionhunter

- a tool for estimating repeat sizes
- gcc/cmake/boost build dependencies
- made a bioconda package 📦
• New version needs gcc >= 4.9
• conda supports 4.8

Solutions:

• compile gcc 4.9 self inside conda 🐍
• use LVMM
• Use a container
DOCKERFILE

Used Alpine linux: distro made for container

- small and up-to-date
- fun (something else then Debian or CentOS)

FROM alpine:3.6
RUN apk add --no-cache build-base boost boost-dev cmake gcc
RUN git clone https://github.com/Illumina/ExpansionHunter
  cd ExpansionHunter &&
  mkdir build && cd build && cmake .. && make &&
  mv ExpansionHunter /usr/bin/. && cd / && rm -rf ExpansionHunter
RUN apk del build-base cmake gcc git zlib-dev
ENTRYPOINT ["ExpansionHunter"]
CMD ["--help"]
DOCKER BUILD ☕ AND RUN

bash-3.2$
INSTALL SINGULARITY
CONVERT DOCKER TO SINGULARITY (LOCAL)
CONVERT DOCKER TO SINGULARITY (DOCKERHUB)
USE SHELL IN THE IMAGE
USE SHELL IN THE IMAGE (AD-HOC CHANGES)
UPLOAD YOUR IMAGE TO A RESOURCE
Container Creation

- singularity create container.img

Add Content

- singularity import container.img docker:ubuntu
- sudo singularity bootstrap container.img Singularity

Interact and Modify

- sudo singularity shell --entitle container.img

Container Execution

- singularity run container.img
- singularity shell container.img
- singularity exec container.img ...

USER ENDPOINT

SHARED RESOURCE
USE CASES

- Sysadmin
- Developer
- Scientist
- GPU
- Gamer
SYSADMIN

containerize program called hard

```bash
#!/bin/bash

singularity exec -pwd $PWD prog/with/impossible/deps/hard
```

- copy script to `/usr/local/bin/hard`
- user can just call hard without any knowledge of containers
DEVELOPER

- Go crazy with dependencies
- Forget about deb and yum packages
- No more "how do I install" Q's
- Join the "Works on My Machine" Certification Program
**SCIENTIST**

- Reproducibility: international collabs
- Reproducibility: produce same results over a year time
- No need to nag a sysadmin (reduces costs of chocolate/beer/coffee/cake/*)
- More time to focus on science
GPU

Easy to install suites like:

- caffe-gpu
- tensorflow-gpu

Plays also well with MPI
GAMER

- X Window System works out of the box
- Wine already works: http://dolmades.org/
- Fill in Windows Excel only spreadsheets
- scientific programs: MaxQuant?
SINGULARITY AT SURFSARA
ON ALL BATCH SYSTEMS

- Life Science grid
- Gina (Grid@surfsara)
- Lisa
- Grid@nikhef

- Carthesius

- More than 70000 cores, 2 petabyte of memory and 2800 nodes
GENERIC INSTRUCTIONS AND SETUP

- One image to rule them all!
- setup uniform
- Follow "latest greatest" upgrade policy for now: no guarantee for future.
- Information can be found on [https://userinfo.surfsara.nl/](https://userinfo.surfsara.nl/)
BUT WHEN?

Now on GinA and LSG, next week on the other systems
HELPDESK

- In start phase a task force handles questions
- Contact via helpdesk@surfsara.nl
- Improve documentation/FAQ/setup
ACKNOWLEDGEMENTS

- Valeriu Codreanu
- Michel Scheerman
- Luuk Uljee
- Bas van de Vlies
- Dennis van der Dok (Nikhef)
- Maarten van Ingen
- Nuno Ferreira

- Jeroen Schot

- Raymond Oonk and Alex Mechev (LOFAR)
HIDDEN AGENDA

- Less need HPCCloud:
  - less time on system administration
  - More time in science
  - scheduling jobs is resource wise more efficient than cloud*
- Make software independent of infrastructure
- Choose infrastructure based on their hardware requirements, not software requirements
- next step: make data independent of hardware: S3/swift
PROBLEMS?

$PATH$ is outside and inside the container the same

```bash
gcc -march=native
```