

Investigation of Docker for data analyses in the context of ASTERICS

Stefan Geißelsöder
September 2016
Friedrich-Alexander-Universität
Erlangen-Nürnberg



ERLANGEN CENTRE
FOR ASTROPARTICLE
PHYSICS



FRIEDRICH-ALEXANDER
UNIVERSITÄT
ERLANGEN-NÜRNBERG

NATURWISSENSCHAFTLICHE
FAKULTÄT



Motivation

Similar to other experiments KM3NeT has to

- keep results and plots reproducible
- reliably keep analyses reproducible
- deal with heterogeneous software and data for many years
- simplify exchange between groups



Investigating a **simple** solution:





What's Docker?

From docker.com:

- “Docker is the world’s leading software containerization platform”
- “Docker containers wrap a piece of software in a complete filesystem that contains everything needed to run: code, runtime, system tools, system libraries (...).”
- “This guarantees that the software will always run the same, regardless of its environment.”



What's Docker to us?

From docker.com:

- “Docker is the world’s leading software containerization platform”
⇒ Support for many systems, easy to exchange.
Linux, Windows, Mac OS. Desktop, grids, clouds.
- “Docker containers wrap a piece of software in a complete filesystem that contains everything needed to run: code, runtime, system tools, system libraries (...).”
⇒ Like a virtual machine. But lightweight and **layered**.
Faster, uses less RAM and **no need to store filesystems or data multiple times** for multiple analyses.
- “This guarantees that the software will always run the same, regardless of its environment.”
⇒ **Analyses are fully reproducible**.
On every¹ machine, regardless of new root, gcc and python versions.
In 10 years, with the same results.

¹x86 compatible



Using Docker for analyses

This is a Dockerfile:

- FROM km3net/km3base:20160909
Derive from a common base

- # Copy your software and inputs to the image
- COPY myAnalysis /myAnalysis
- COPY input /myAnalysis/input
- RUN make -C /myAnalysis

- # Copy the bash/python/whatever script you use to do the analysis
- COPY runMyAnalysis.sh /runMyAnalysis.sh
- RUN chmod 755 runMyAnalysis.sh
Define the script to be the startpoint
- ENTRYPOINT ["/runMyAnalysis.sh"]



Using Docker for analyses

Bild own image:

- `docker build -t myAnalysis:20160914`

Run own image:

- `docker run -v 'pwd'/output/:/myAnalysis/output myAnalysis:20160914`

“-v ‘pwd’/output/:/output” maps the host directory ./output to /myAnalysis/output in the Docker container.

Note: Calibration and filtering of raw data to obtain “official” data sets is not included in this procedure, but this isn’t the intention (at the moment).

Using Docker for analyses

- is currently being investigated for KM3NeT in the context of ASTERICS
- isn't hard
- simplifies taking over an analysis for new students
- provides reproducibility for analyses and results

