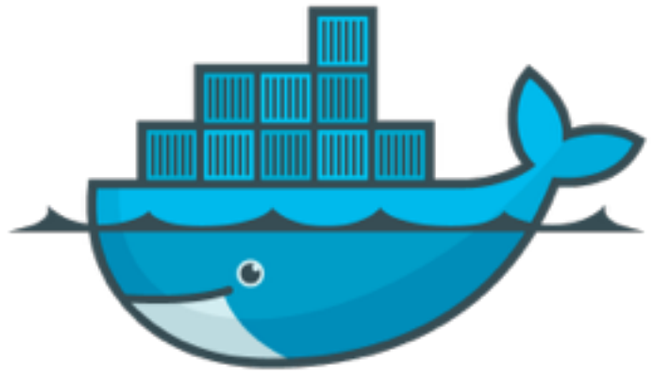


An introduction to Docker for reproducible research

Conference Paper by Carl Boettiger (January 2015)

Presented at TheoSysBio Group Meeting
23rd September 2015

Reproducible research. Why should you care…?

Make collaboration easier

Help raise profile of your work

Condition of funding / publication

Reproducible research. Why should you care…?

The case for open computer programs
Darrel C. Ince, Leslie Hatton & John Graham-Cumming
Nature (22 February 2012)

nature

If a job is worth doing, it is worth doing twice
Researchers and funding agencies need to put a premium on ensuring that results are reproducible, argues Jonathan F. Russell.
Nature (03 April 2013)

Open science decoded
Granting access to publications and data may be a step towards open science, but it's not enough to ensure reproducibility. Making computer code available is also necessary — but the emphasis must be on the quality of the programming
Tony Hey and Mike C. Payne
Nature Physics (May 2015)

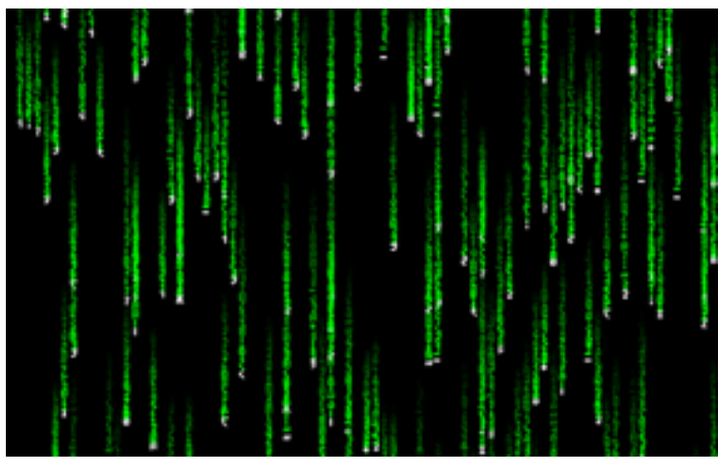
Reproducible research. Why should you care…?

"If the manuscript describes new software tools or the implementation of novel algorithms the software must be freely available to non-commercial users at the time of submission, and appropriate test data should be made available.” - *Oxford Bioinformatics guidelines*

"The source code must be accompanied with documentation on building and installing the software from source, as well as for using the software, including instructions on how a user can test the software on supplied test data.” - *PLoS Comp Bio guidelines*

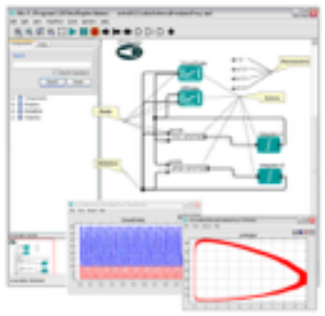
"There is a growing and unstoppable pressure for, and momentum towards greater openness. [...] The pressures embrace not just access, but sharing, re-use, data, open source software, open educational resources"
- March 2015 report commissioned by RCUK

The challenges…



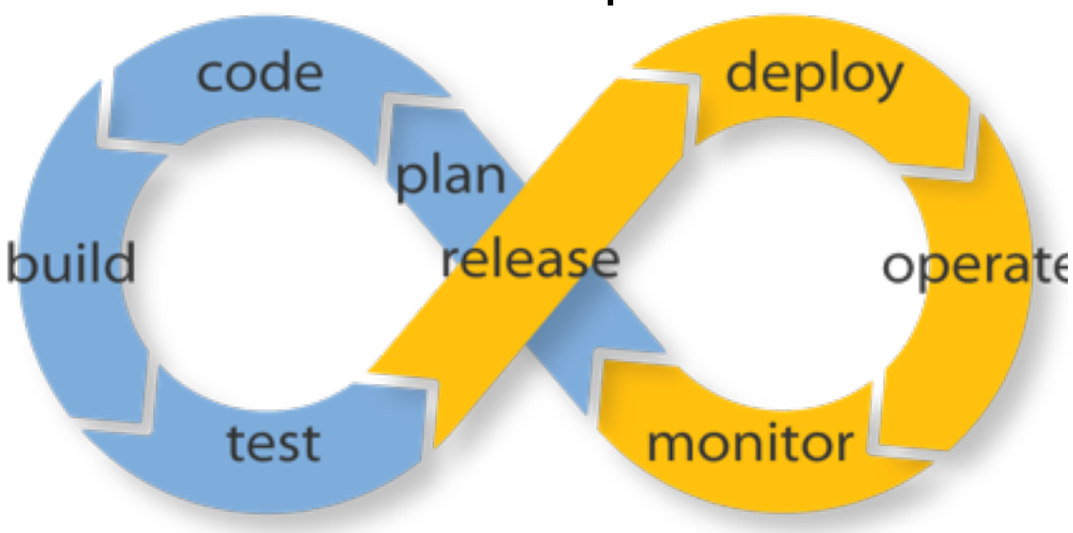
Existing approaches…

Workflow Software

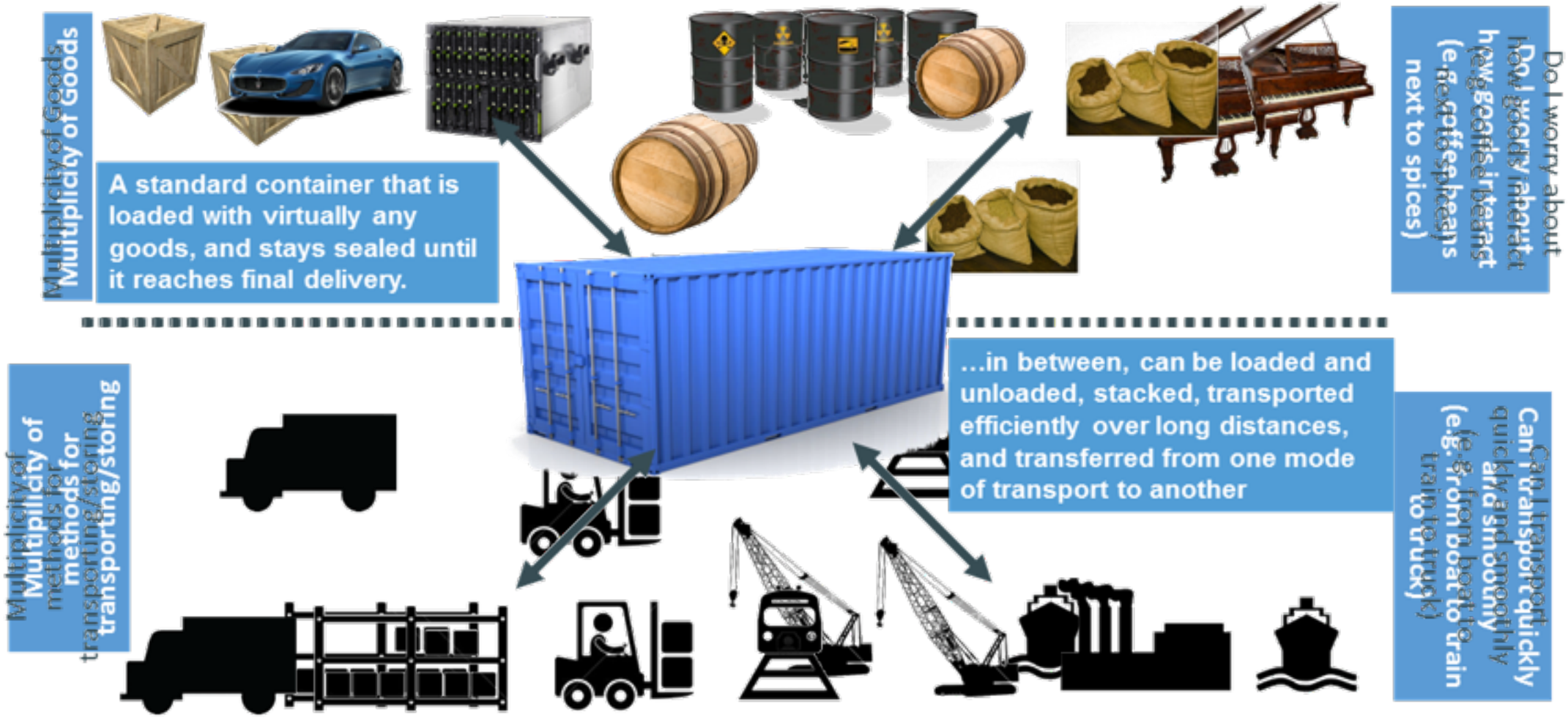


Virtual Machines

DevOps



Shipping analogy…



(borrowed slide…)

What is it…?

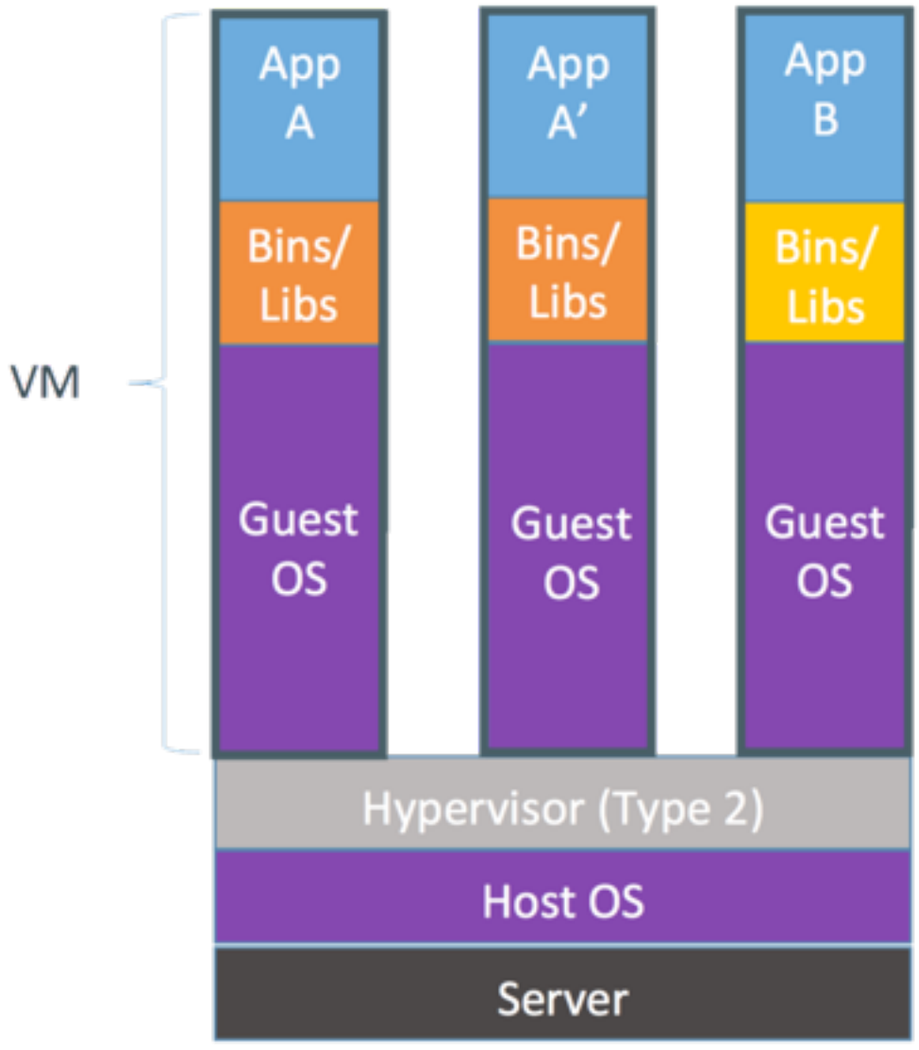
Open source

Mature technologies

Easy to install

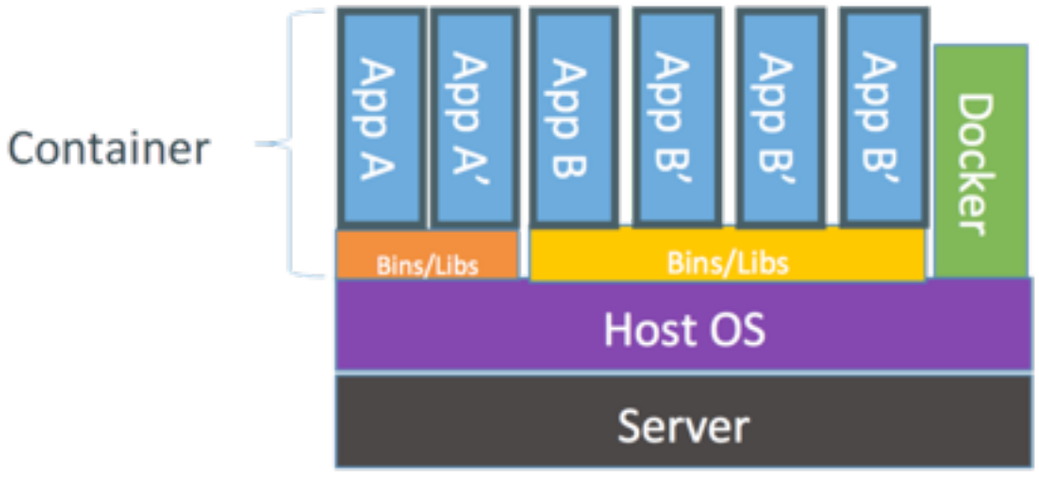
Well supported

VMs & Containers…



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

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

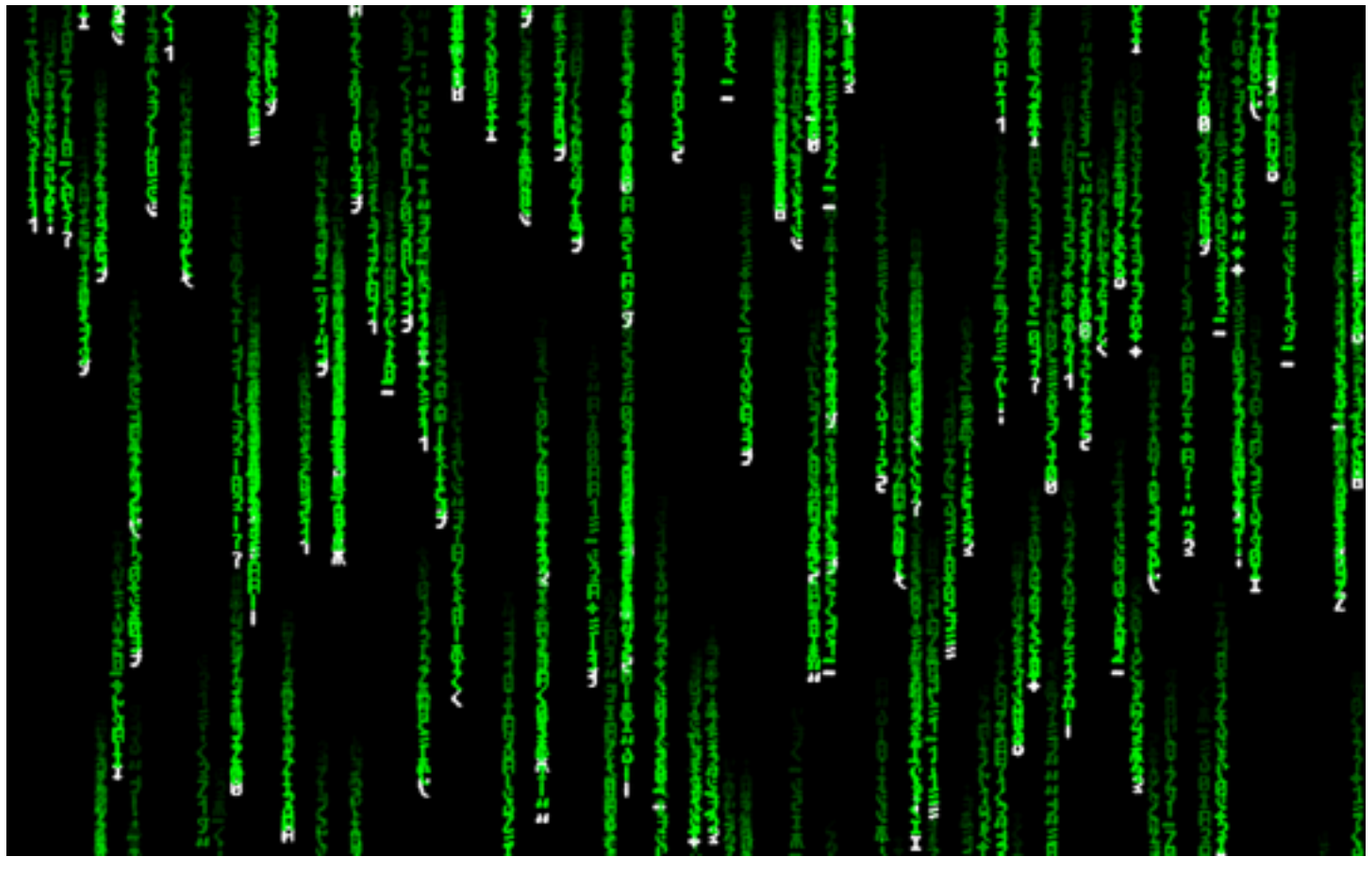


(borrowed slide…)

Docker vs Dependency Hell

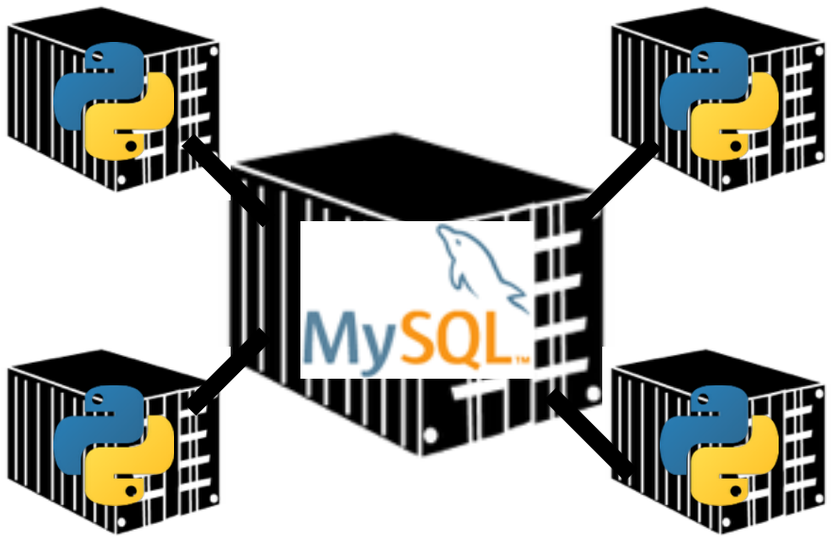
Docker vs Documentation

Docker vs Code Rot



Docker vs Barriers

How could *we* use it…?



Drawbacks…

relies on
host kernel

VM on
non-Linux

security?

rate of
adoption?

huge
datasets?

needs
64-bit

Convincing…?

Something needed to improve reuse/replicability.

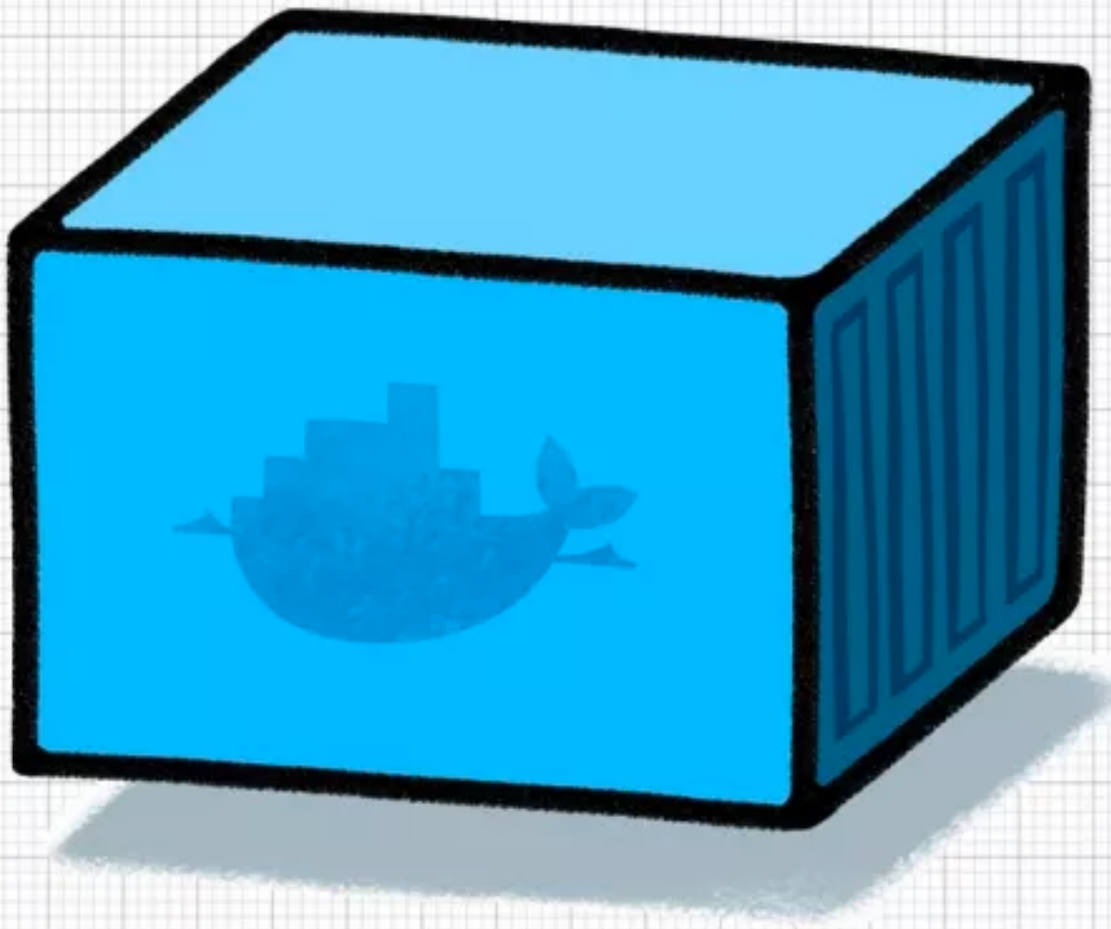
Docker can help tackle a lot of the challenges

Will it become the standard in science?

Relatively small investment needed to try, so low risk.

Demo…

The real value of Docker is not technology



It’s getting people to agree on something