



# CYVERSE™

Transforming Science Through Data-driven Discovery

## Singularity Overview

John Fonner

[jfonner@tacc.utexas.edu](mailto:jfonner@tacc.utexas.edu)

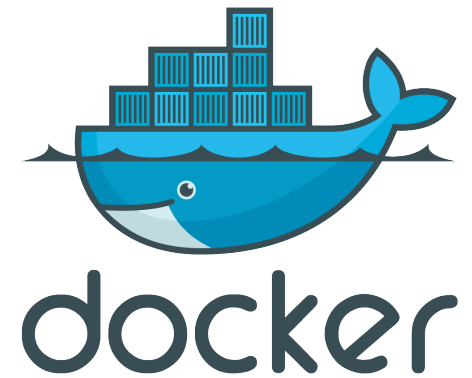


Cold  
Spring  
Harbor  
Laboratory



# Background: Why Containers for Science?

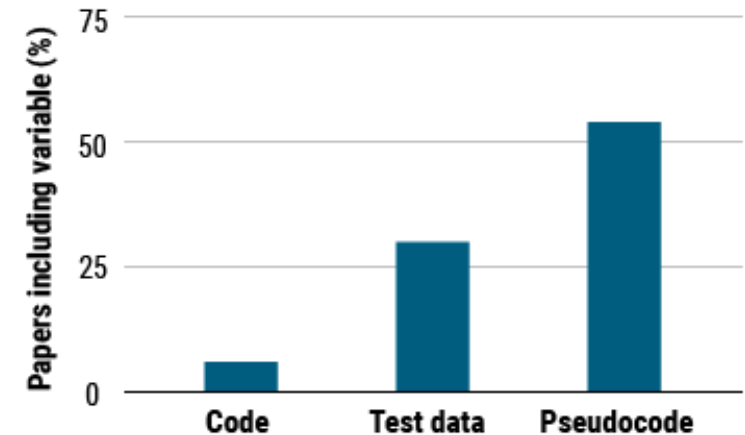
- Web services were the first and primary consumers of containers
  - Replaced VMs as the unit for encapsulating web apps as services
  - Smaller sizes (e.g. Docker's layered image format)
  - Smaller resource footprint (no virtualization layer)
  - Better uptime (from minutes to seconds to reboot)
  - Nice version control / continuous integration support
  - Scalability / portability
- Scientific computing was not the primary use case!



# Background: Why Containers for Science?

- Reproducibility crisis in scientific computing
  - few incentives to optimize software after publication
  - grad students eventually graduate
- Even with code, can you install it?
  - ...and all the dependencies?
- Can you meet the hardware requirements (large memory, GPU, disk storage)?
  - without spending all your grant money on Amazon?

One recent example in machine learning...



Science 16 Feb 2018  
DOI: 10.1126/science.359.6377.725

**YOUR BUILD DOESNT WORK?**



**WORKS FINE ON MY  
MACHINE**

memegenerator.net



# Docker Saves the Day

Right?

- Currently there are scary security vulnerabilities (3/2018)
  - Not a problem when you have root on the VM already
  - Complete non-starter for large shared clusters (which are a large academic resource)
- Requires a recent Linux kernel
  - Not a problem for VMs
  - Large clusters are “time machines”; commonly don’t meet requirements
- GPU support is tricky (nvidia-docker)



# Thus the Need for Other Container Formats

- With containers, the “dependency management” is the container runtime
- Singularity – widespread adoption for non-VM, shared systems
- Shifter – requires a “hosting footprint”, but works well
- CharlieCloud – more minimal user-space Docker runtime
- udocker – userspace Docker. It looks promising.
- and others – rkt, lxc, etc.?



# Singularity for Docker Users

- Singularity can consume Docker images
  - `singularity pull docker://ubuntu`
- No local image registry, just files
  - `docker images -> ls`
  - `singularity exec /path/to/ubuntu.img cat /etc/issue`
- By default, it tries to mount \$PWD and /home
  - same path inside and outside the container
- Networking is essentially a pass through to the host
- Supports GPUs very well; MPI is doable





# Developing with Containers

Docker (review from yesterday)



Develop on your laptop  
(or your favorite server)



Push image to DockerHub



“docker pull” anywhere  
that supports Docker



# Developing with Containers

Singularity



Develop on your laptop  
(or your favorite server)



Push image somewhere...  
SingularityHub or file



“singularity pull” or copy  
anywhere that supports  
Singularity

# Developing with Containers

GitHub + Docker



Develop on your laptop  
(or your favorite server)



Use Github for version  
control



Automatic push to  
DockerHub



“docker pull” anywhere  
that supports Docker

# Developing with Containers

GitHub + Singularity



Develop on your laptop  
(or your favorite server)



Use Github for version  
control



Automatic build from  
SingularityHub



“docker pull” anywhere  
that supports Docker

# Developing with Containers

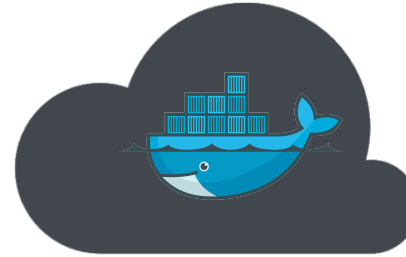
Docker + Singularity + Github



Develop on your laptop  
(or your favorite server)



Use Github for version  
control



Automatic push to  
DockerHub



copy Singularity image to  
SingularityHub or  
directory (optional)



“docker pull” anywhere  
that supports Docker



“singularity pull” or copy  
anywhere that supports  
Singularity

# Summary

- Science has stolen containers from the web developers
- Containers are our current best hope for (more) reproducible computing
- Docker is awesome but not perfect
- The runtime is the dependency
- Developing for containers adds extra steps, but it pays off – let's practice!



CyVerse is supported by the National Science Foundation under Grants No. DBI-0735191 and DBI-1265383.

