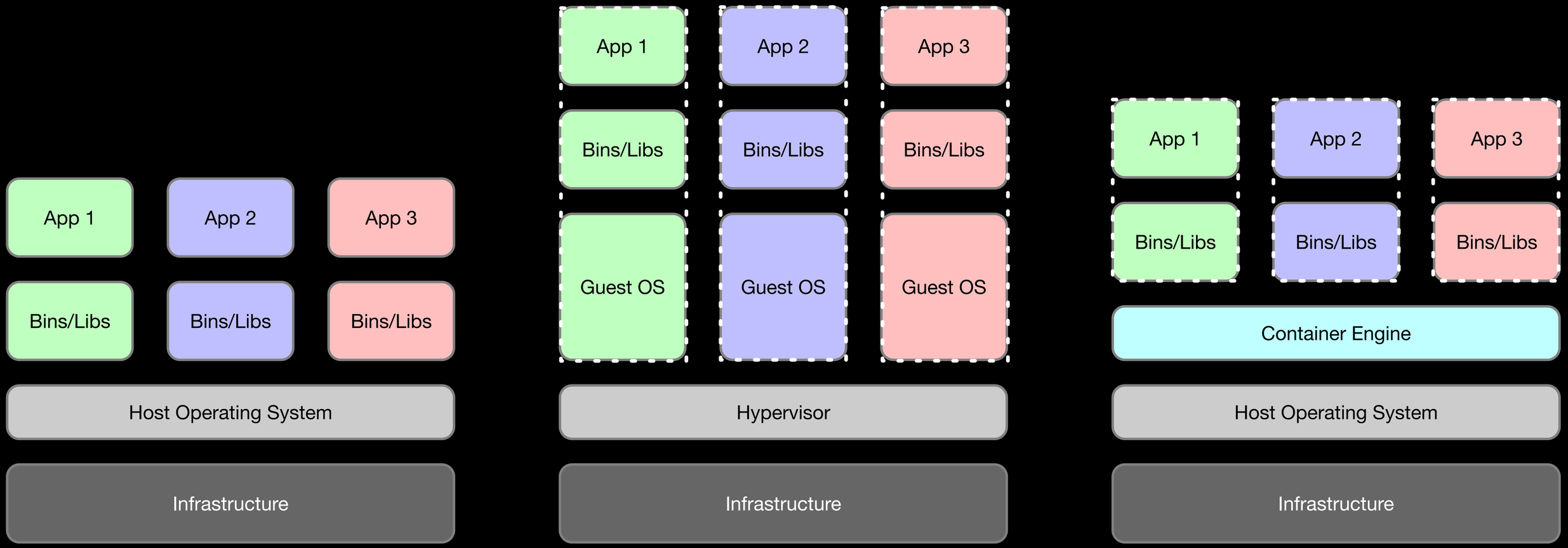
Contain Your Enthusiasm

A 101 Guide to Containers

New Talk
Who dis?

- Jason ‘XenoPhage’ Frisvold
- Twitter: @xenophage
- LinkedIn: https://linkedin.com/in/xenophage
- I blog: https://blog.godshell.com
- Always open to new opportunities

What is containerization?



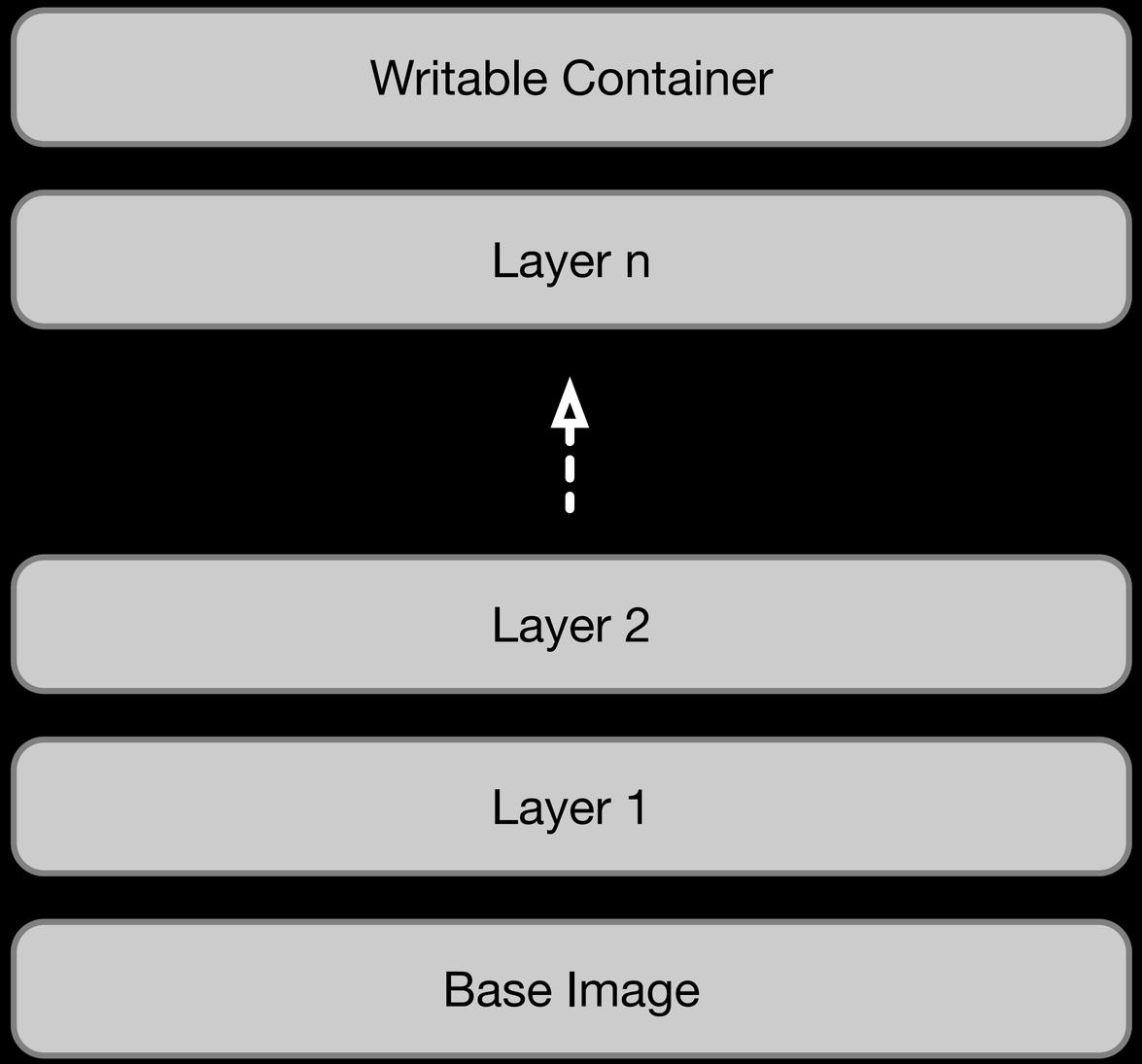
What is a container?

- Layered, Binary Image
- Isolated, or “Sandboxed,” from other containers and the host*
- Portable**
- Cross platform***
- Secure****

* There are, of course, exceptions
** Portable, if everything is in the container
*** Mostly just Linux containers
**** HERE BE DRAGONS

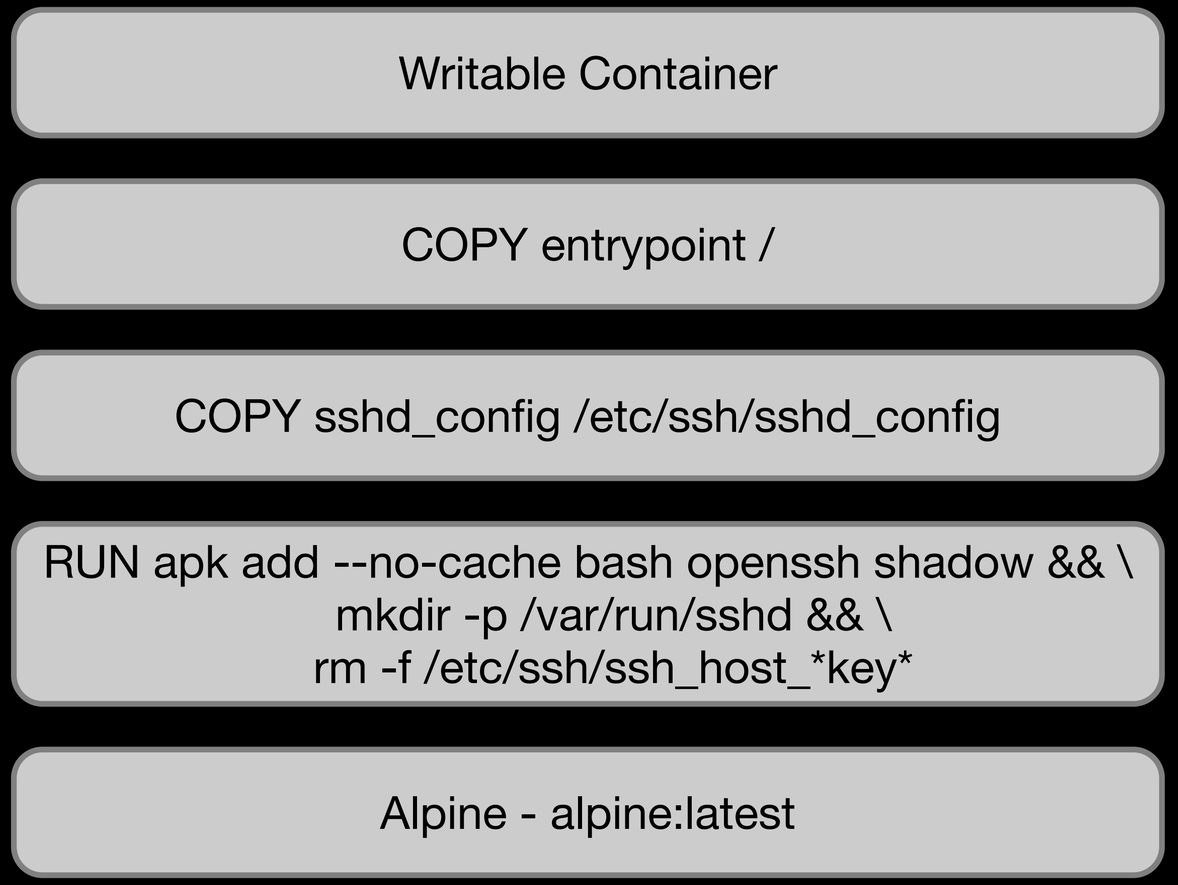
Layers?

- Each “instruction” creates a layer
- Layers are “chained” together
- Layers are cached
- Don’t download the same layer twice
- Layers are “immutable”
- Last layer is the only writable layer
- All combined together using a Union File System



Example Container

```
FROM alpine:latest

ARG GIT_HASH="UNSET"

MAINTAINER Jason Frisvold <friz@godshell.com>
LABEL Description="SFTP Server"
LABEL GitHash=$GIT_HASH

RUN apk add --no-cache bash openssh shadow && \
    mkdir -p /var/run/sshd && \
    rm -f /etc/ssh/ssh_host_*key*

COPY sshd_config /etc/ssh/sshd_config
COPY entrypoint /

EXPOSE 22

ENTRYPOINT ["/entrypoint"]
```

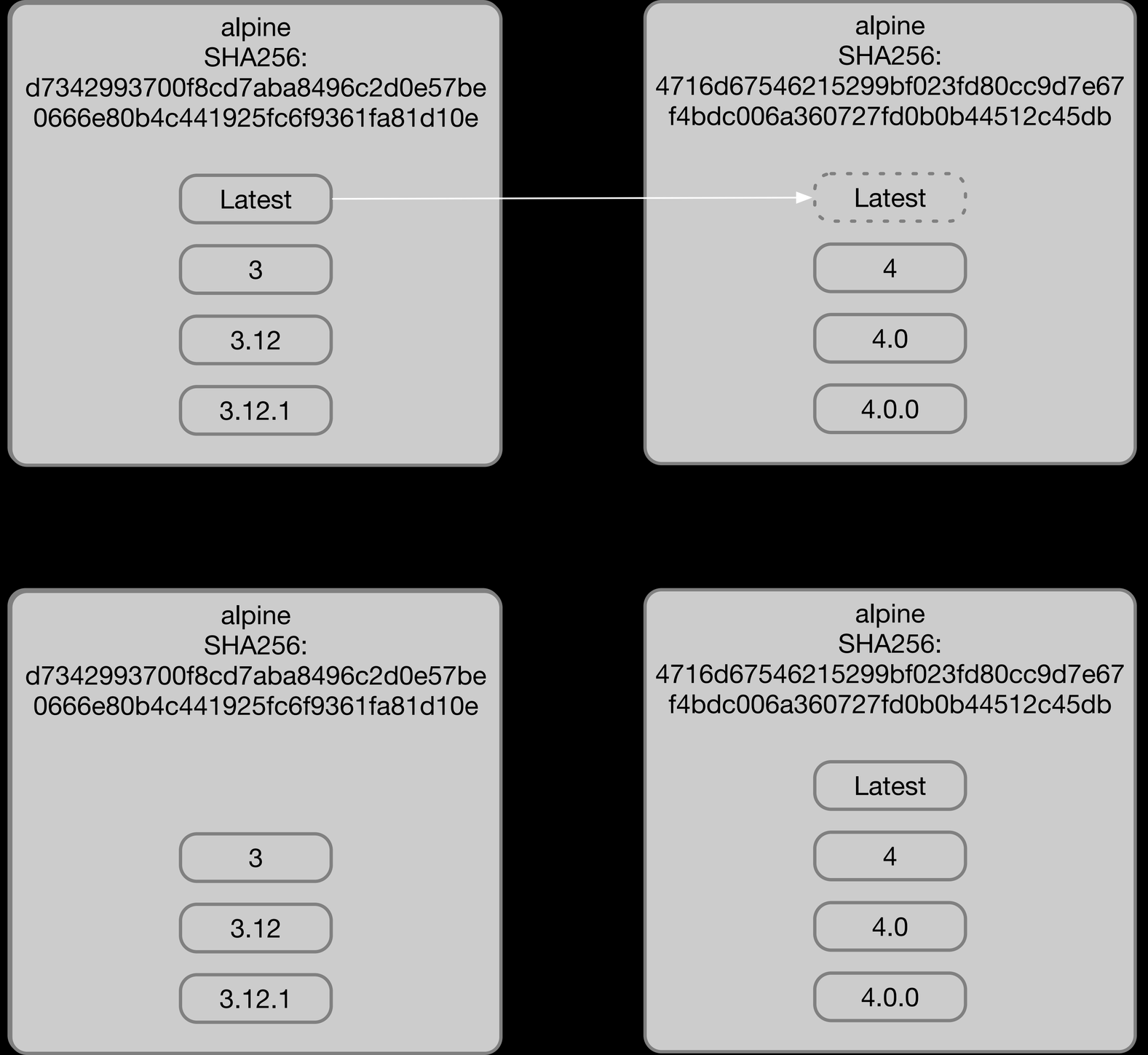


Example Cont.

```
[root@odin docker]# docker history sftp-server:latest
IMAGE               CREATED              CREATED BY                                      SIZE                COMMENT
7ede58024dd1        54 seconds ago       /bin/sh -c #(nop)  ENTRYPOINT ["/entrypoint"]   0B
8ddea9f71594        54 seconds ago       /bin/sh -c #(nop)  EXPOSE 22                    0B
8838108eab8f        55 seconds ago       /bin/sh -c #(nop) COPY file:ca6c99d8d5579ec3…   3.94kB
8991fafb0586        55 seconds ago       /bin/sh -c #(nop) COPY file:1a8e39058dad3bfe…   477B
4e7dc88f863d        55 seconds ago       |1 GIT_HASH=6f30caba17d359b8abe577a4413283b7…   9.43MB
0c74e16b1642        59 seconds ago       /bin/sh -c #(nop)  LABEL GitHash=6f30caba17d…   0B
d74bd8f9bb0a        59 seconds ago       /bin/sh -c #(nop)  LABEL Description=SFTP Se…   0B
01a7fc814700        About a minute ago   /bin/sh -c #(nop)  MAINTAINER Jason Frisvold…   0B
195acebab481        About a minute ago   /bin/sh -c #(nop)  ARG GIT_HASH=UNSET           0B
4d90542f0623        17 months ago        /bin/sh -c #(nop)  CMD ["/bin/sh"]              0B
<missing>           17 months ago        /bin/sh -c #(nop) ADD file:fef3b00b3ae63967d…   5.58MB
```

What’s this tag thing?

- Images have tags
- A tad easier than remembering the sha-256 hash
- Tags are UNIQUE per image
- BUT, you can “MOVE” them
- The “latest” standard


alpine
SHA256:
d7342993700f8cd7aba8496c2d0e57be
0666e80b4c441925fc6f9361fa81d10e
Latest
3
3.12
3.12.1
alpine
SHA256:
4716d67546215299bf023fd80cc9d7e67
f4bdc006a360727fd0b0b44512c45db
Latest
4
4.0
4.0.0
alpine
SHA256:
d7342993700f8cd7aba8496c2d0e57be
0666e80b4c441925fc6f9361fa81d10e
3
3.12
3.12.1
alpine
SHA256:
4716d67546215299bf023fd80cc9d7e67
f4bdc006a360727fd0b0b44512c45db
Latest
4
4.0
4.0.0

Container Registry

So I have a container, but where is it?

- Stored locally after build
  - Local storage is basically a registry, but not remotely accessible
- Can push to a remote registry
- Registries can be authenticated
- Registries store layers and manifests
  - A manifest is a list of layers, sha-256 hashes, tags, and other info

Isolation, Part 1

- namespaces
  - net - Network Interfaces
  - mnt - Mounts
  - ipc - Inter Process Comms
  - uts - hostname/nis domain
  - user/pid - uid/gid, processes

What is a namespace?

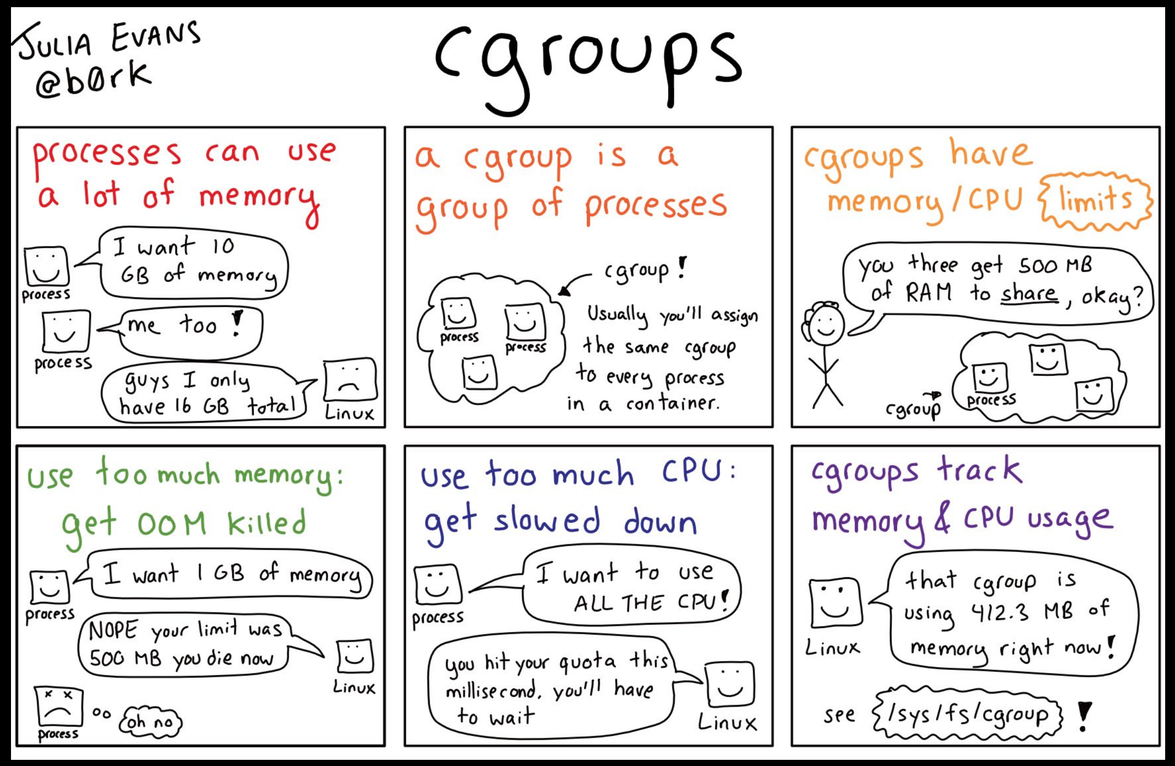
- A mapping of an object from one view to another

```
[root@dockercontainer ~]# ps -ef
UID        PID  PPID  C STIME TTY          TIME CMD
root         1     0  0 Nov27 ?        00:00:12 apache2 -DFOREGROUND
www-data    18     1  0 Nov27 ?        00:00:56 apache2 -DFOREGROUND
www-data    20     1  0 Nov27 ?        00:00:24 apache2 -DFOREGROUND
www-data    21     1  0 Nov27 ?        00:00:22 apache2 -DFOREGROUND
root       559     0  0 14:30 ?        00:00:00 ps -ef
```

```
[root@dockerhost ~]# ps -ef
UID        PID  PPID  C STIME TTY          TIME CMD
root         1     0  0 Oct15 ?        00:02:40 /usr/lib/systemd/systemd
--switched-root --system --deserialize 21
root         2     0  0 Oct15 ?        00:00:03 [kthreadd]
root         3     2  0 Oct15 ?        00:03:44 [ksoftirqd/0]
...
root      1514     1  0 Oct15 ?        04:28:40 /usr/bin/dockerd-current
--add-runtime docker-runc=/usr/libexec/docker/docker-runc-current --
default-runtime=docker-runc --exec-opt nat
root      1673  1514  0 Oct15 ?        01:27:08 /usr/bin/docker-
containerd-current -l unix:///var/run/docker/libcontainerd/docker-
containerd.sock --metrics-interval=0 --start-timeout
root      4035  1673  0 Oct31 ?        00:00:07 /usr/bin/docker-
containerd-shim-current
d548c5b83fa61d8e3bd86ad42a7ffea9b7c86e3f9d8095c1577d3e1270bb9420 /var/
run/docker/libcontainerd/
root      4054  4035  0 Oct31 ?        00:01:24 apache2 -DFOREGROUND
33        6281  4054  0 Nov13 ?        00:00:07 apache2 -DFOREGROUND
33        8526  4054  0 Nov16 ?        00:00:03 apache2 -DFOREGROUND
33       24333  4054  0 04:13 ?        00:00:00 apache2 -DFOREGROUND
root     28489  1514  0 Oct31 ?        00:00:01 /usr/libexec/docker/
docker-proxy-current -proto tcp -host-ip 0.0.0.0 -host-port 443
-container-ip 172.22.0.3 -container-port 443
root     28502  1514  0 Oct31 ?        00:00:01 /usr/libexec/docker/
docker-proxy-current -proto tcp -host-ip 0.0.0.0 -host-port 80
-container-ip 172.22.0.3 -container-port 80
33       19216  4054  0 Nov13 ?        00:00:08 apache2 -DFOREGROUND
```

Isolation, part deux

- cgroups
  - Limit cpu
  - Limit memory
- By default, full throttle



Portability

- Build once, deploy many
- Replicas
- Automatic restarts
- Great for difficult to install/ configure apps
  - Application X requires special version of Library Y

Portability Gotchas

- Stateful Containers
  - Local disk
  - State in memory
- Environmental differences
  - Hardware access
  - Insufficient Resources

Deployment
A quick side journey

- Deploy new versions “live”
  - Round robin deployment
  - Fail back to previous version
- Quick re-deploy of previous version if necessary
- Can “pre-load” but takes scripting

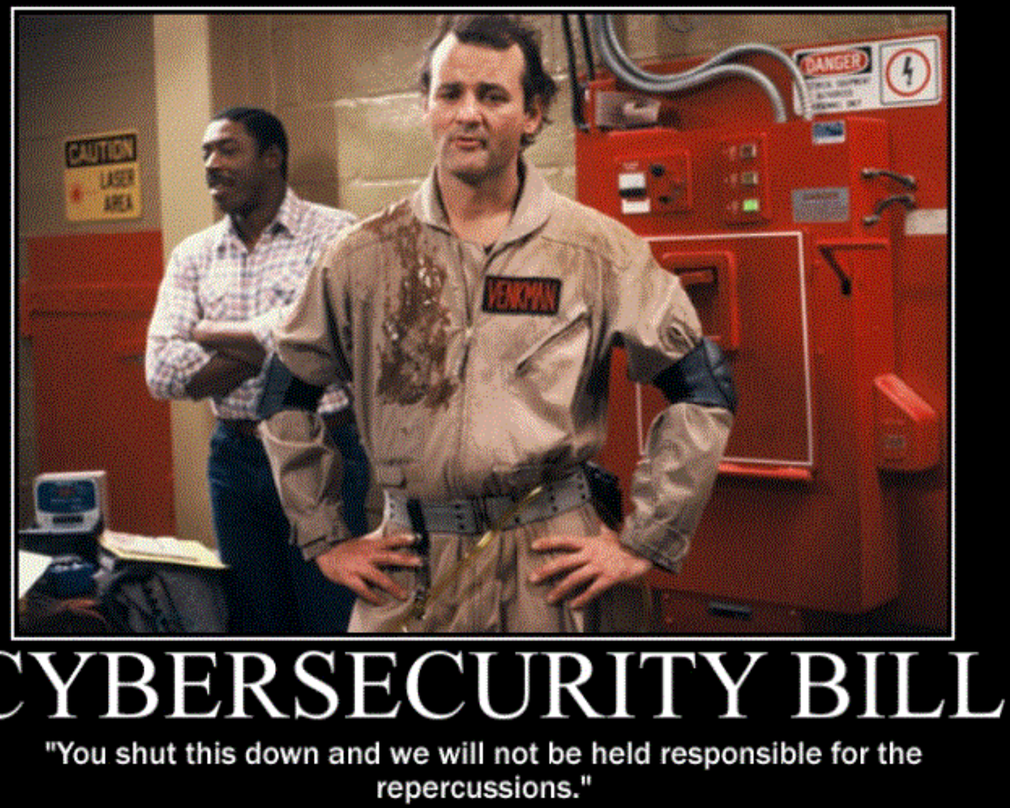
Cross Platform

- Linux container
  - No real problem, can run on Linux/Mac/Windows
- Windows container
  - Can only run on Windows hosts
  - Different base windows versions across different hosts versions gets complicated
    - Process isolation vs hyper-v isolation
      - hyper-v isolation requires a hyper-v capable machine
    - “Old” containers can run on newer Windows hosts with hyper-v isolation
    - “New” containers can NOT run on older Windows hosts
- macOS container
  - Yep, it’s sort of possible. It’s just a VM in a container, though…
  - https://github.com/sickcodes/Docker-OSX

Security

I mean, this is a security conference, right?

- selinux
  - Layered on top, not explicitly required
  - Provides the protections you expect
- Minimized containers
  - Install only what you need
  - Can be VERY stripped down
  - There are tools to help
- Container network isolation
- “Immutable” means you can “restart" a container and be back to a pristine environment



Making it all work

- Easy to spin up a handful of containers
- At scale we’re talking cattle, not pets
- How do we manage this madness?

Orchestration Tools

- Docker Swarm
  - The “original”
- Kubernetes (k8s)
  - Based on Google Borg
- Nomad
- Marathon
  - Based on Apache Mesos

Summary

- Small footprint
- Rapid deployment
- Pristine environments
- Immutability
- Security
- Portability
- CICD friendly

Questions?

I'll be in the Discord, just @ me…