

Containers and Some Odds and Ends About Computational Infrastructure

Rob Quick <rquick@iu.edu>

Chief Operations Officer - Open Science Grid
Manager High Throughput Computing

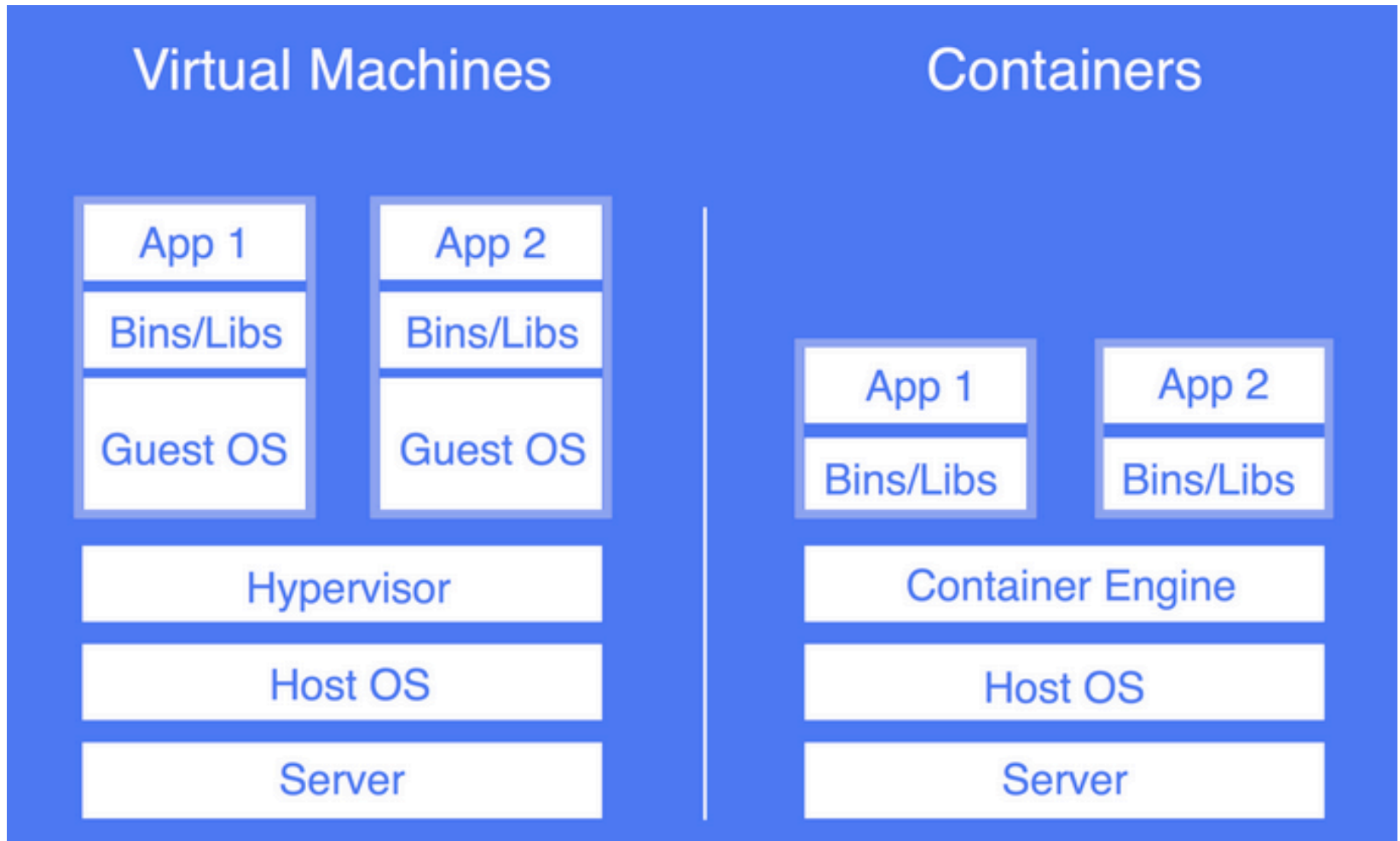
Follow Along at:

- <https://opensciencegrid.github.io/dosar/Materials/TriesteMaterials/>

What are containers?

- Operating System Level Virtualization
 - Lightweight, providing the minimal level of overhead for the application to function properly.
 - Super minimalist VMs
 - No Hypervisor
 - Abstracts away the operating system and hardware
 - Share the OS Kernel with other containers
 - Container size is very small and therefore quick and easy to provision

How do they differ from VMs?



More differences...

- Size
 - Containers are usually 10s of MB
 - VMs can be several GB
- Shared hypervisor vs. shared kernel
- VMs have their own kernels so a deeper level of isolation
- Containers virtualize the OS while VMs virtualize the hardware

Container Advantages

- Size
- Less resource intensive
- Quick provisioning
- Easy allocation of resources
- Quicker development cycles
- Cost effective
- Very good for microservices

Container Disadvantages

- Security – shared kernel with root access
- Less flexibility in OS
- Networking can be tricky
 - Properly configuring sufficient networking resources is challenging

Container Software

- Docker
- Singularity
- LXC, LXD
- Solaris Zones
- RKT
- BSD Jails
- chroot

A quick review



Computing Infrastructures

- Local Laptop/Desktop – Short jobs with small data
- Local Cluster – Larger jobs and larger data but subject to availability
- HPC – Prime performance with parallelized code
- HTC – Sustained computing over a long period for serialized workflows
- Cloud – Need deeper permission on an OS and/or have deeper pockets

Some Examples of Academic CIs Worldwide

- HTC
 - EGI (formally European Grid Initiative)
 - OSG (Open Science Grid)
 - ASGI (Asia Pacific Grid Initiative)
 - NorduGrid
 - Earth System Grid (ESG)
 - Many other regional and national infrastructures

Some Examples of Academic CIs Worldwide

- HPC
 - XSEDE (eXtreme Science and Engineering Discovery Environment)
 - PRACE (Partnership for Advanced Computing in Europe)
 - Compute Canada
 - Greek Research and Technology Network (GRNET)
 - Centre for HPC (South Africa)
 - Many other national infrastructures

Some Examples of Academic CIs Worldwide

- Cloud
 - EGI Federated Cloud and EOSC
 - NeCTaR – National eResearch Collaboration Tools and Resources
 - Jetstream (Part of XSEDE)
 - SwissACC (Swiss Academic Computing Cloud)
 - Many other national cloud infrastructures

Let's take one step at a time

Small

Local



Large

Distributed

- Can you run one job on one computer?
- Can you run one job on another computer?
- Can you run 10 jobs on a set of computers?
- Can you run a multiple job workflow?
- How do we put this all together?

This is the path we'll take



What happens when you go home?

- DOSAR: Distributed Organization for Scientific and Academic Research <http://www.dosar.org/>
- You are welcome to join our bi--weekly video (Vidyo) meetings. Send request to be added to DOSAR email list to Prof. Greenwood: greenw@phys.latech.edu reference you attended the Data Science Summer School in Trieste
- If you want long-term grid access, you can request membership in the DOSAR VO

Questions?

- Questions? Comments?
 - Feel free to ask me questions now or later:
Rob Quick rquick@iu.edu

Exercises start here:

- <https://opensciencegrid.github.io/dosar/Materials/TriesteMaterials/>

Presentations are also available from this URL.