Intro to Docker

...

Luiz Irber
Research Assistant - DIB Lab
CS PhD Student - UC Davis

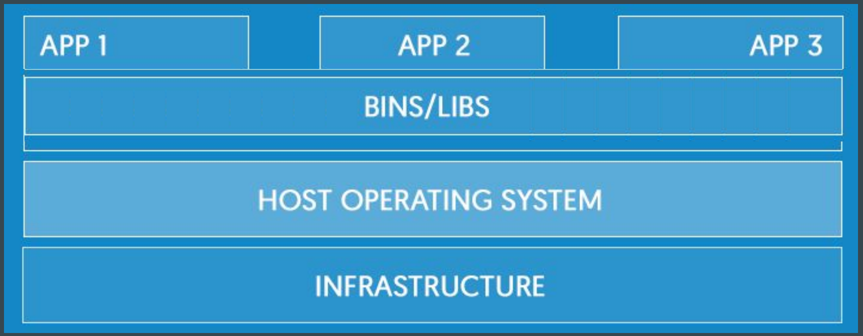
Supercomputing 16

Bioinformatics

- Pipelines
  - (lots of software dependencies)
- Method development
  - How do I share it?
- -> Fairly large initial setup
- Typically ~GB to TB datasets

Evolution of a bioinformatician

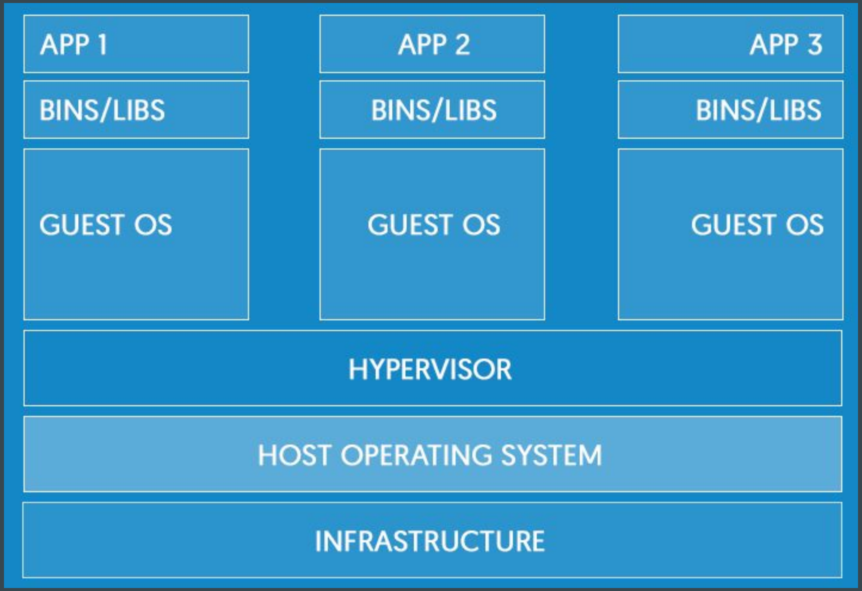
- Prepare analysis in my desktop
- While I'm developing, computer is mostly idle
- Wasted resources while I'm not running anything intensive (or I'm sleeping)
- Need to choose configuration based on peak performance
  - If my analysis need 64 GB of RAM and 32 cores, do I need it all while I'm preparing scripts?
- Hard to move to other system
  - (because I probably didn't keep track of everything I installed)



Based on: https://www.docker.com/what-docker

Evolution of a bioinformatician: the cloud

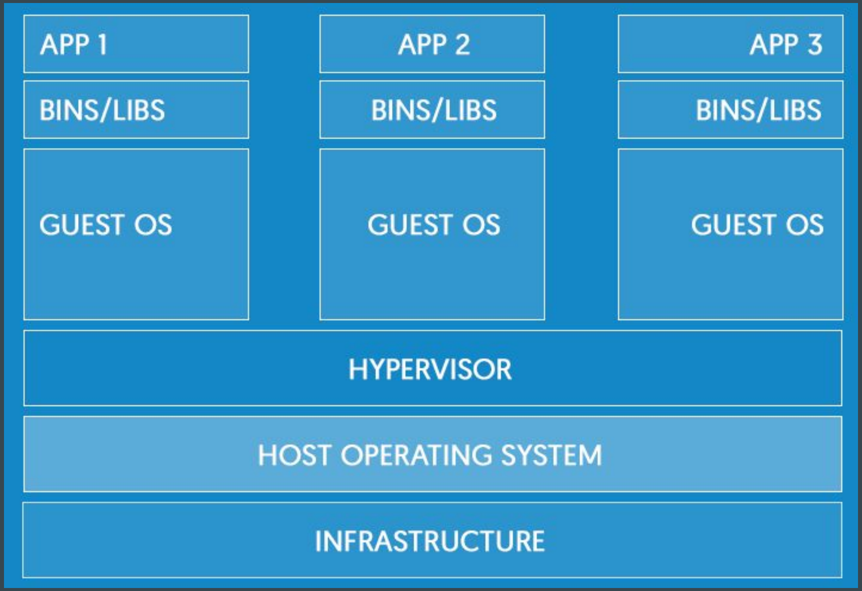
- Move my analysis to a cloud server
- Share a server with other users without conflicts. I can choose:
    - Operating system
    - Compiler
    - specific library versions
    - How many CPUs, RAM and storage I want to use
- Allocation of new instances
- I can save an image and share with others



Source: https://www.docker.com/what-docker

Evolution of a bioinformatician: the cloud (Cons)

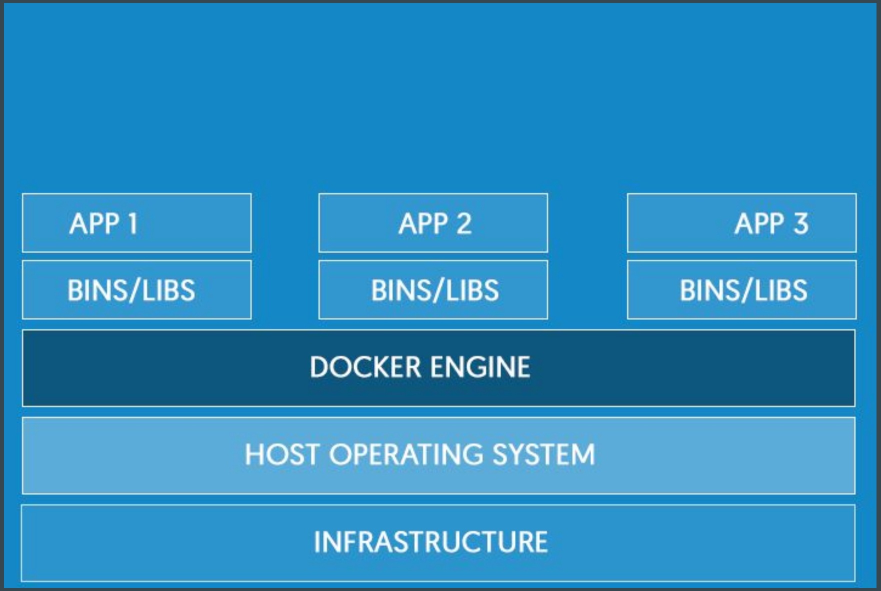
- Image still is a 'binary blob'
  - What was installed?
  - Which compiler options or features were used?
  - Saves the whole guest OS
  - Possible solution: configuration management (Ansible, Chef, Puppet, Salt)
- Opaque resource allocation
  - Once I allocate resources, other VMs can't use it (even if my VM is idle)
- Guest OS overhead



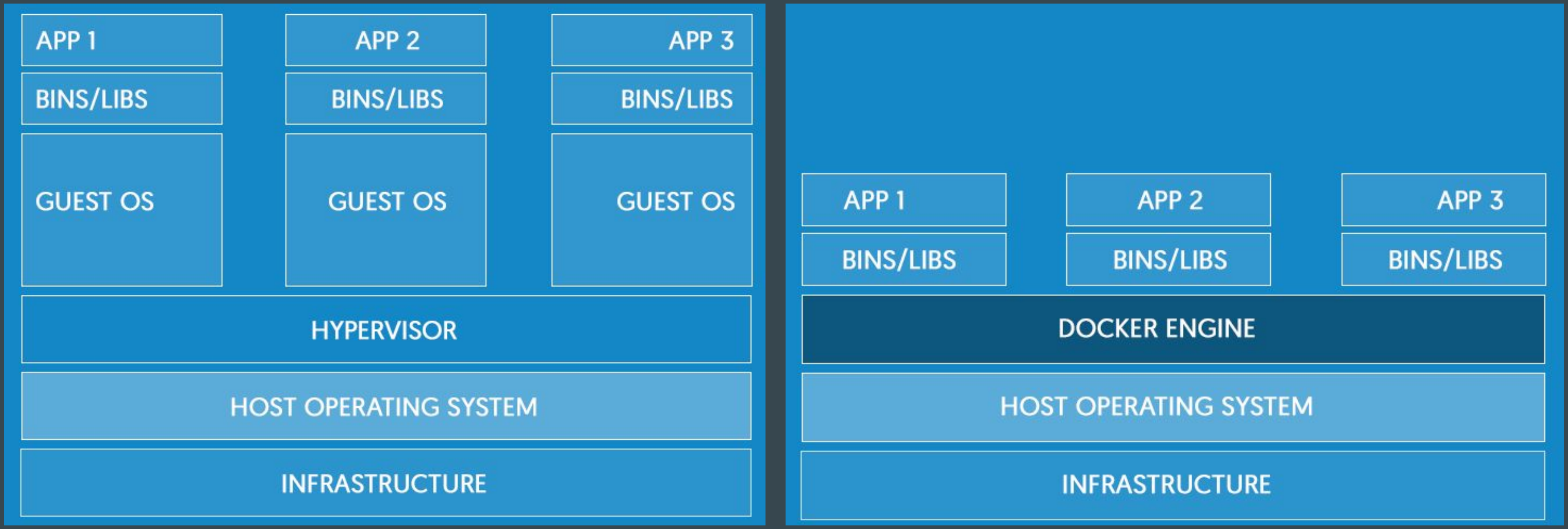
Source: https://www.docker.com/what-docker

Evolution of a bioinformatician: containers

- uses the kernel on the host's OS to run multiple guest instances
- Container: Each guest instance
- Each container has its own
  - Root filesystem
  - Processes
  - Memory
  - Devices
  - Network ports
- Isolated application platform



Source: https://www.docker.com/what-docker



APP 1 | APP 2 | APP 3
BINS/LIBS | BINS/LIBS | BINS/LIBS
GUEST OS | GUEST OS | GUEST OS
HYPERVISOR
HOST OPERATING SYSTEM
INFRASTRUCTURE

APP 1 | APP 2 | APP 3
BINS/LIBS | BINS/LIBS | BINS/LIBS
DOCKER ENGINE
HOST OPERATING SYSTEM
INFRASTRUCTURE

Containers vs VMs

- Containers are more lightweight
- No need to install guest OS
- Less CPU, RAM, storage space required
- More containers per machine than VMs
- Greater portability

Source: https://www.docker.com/what-docker

Happy bioinformatician

- Easier to share pipelines
- Easier to share analysis
- Standardized tools to build upon
- Reproducibility

COMMON
WORKFLOW
LANGUAGE

binder