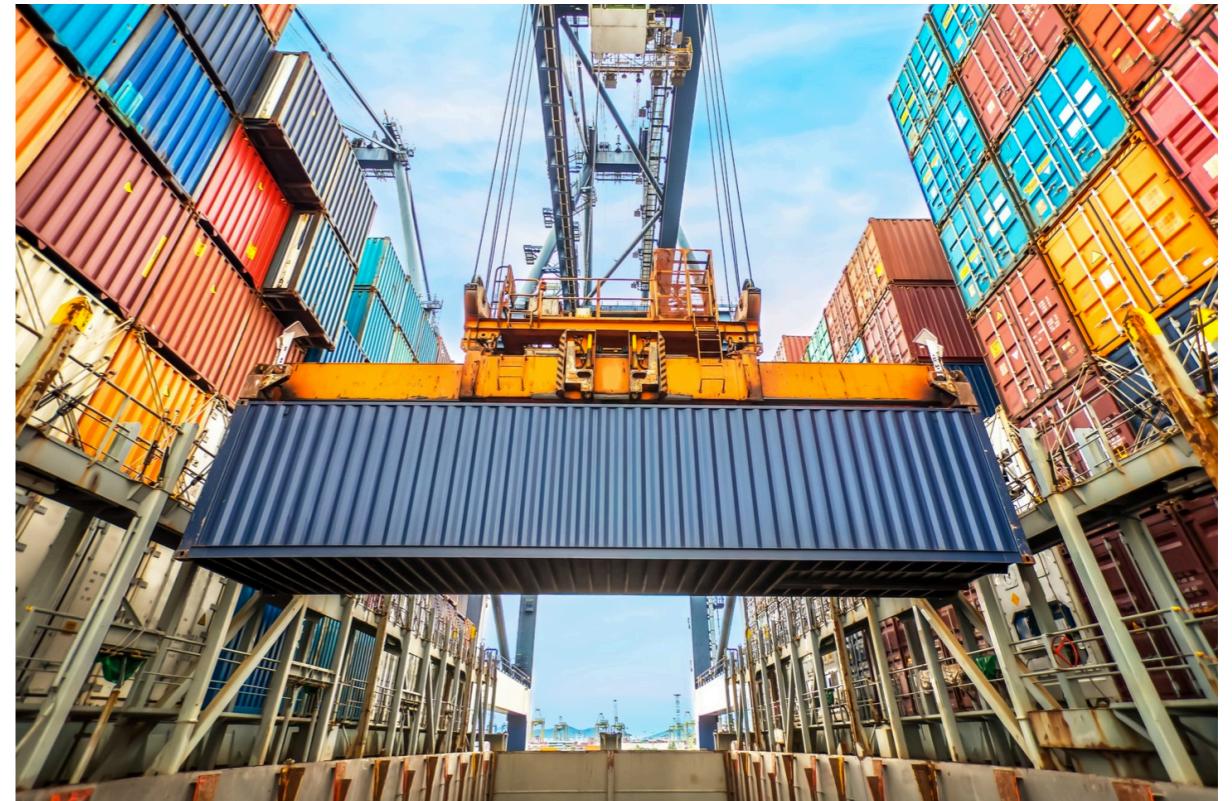


A Beginner's Guide to Docker (for Data Science or otherwise)



Dinesh Shenoy, PhD
Data Scientist, Veritas Technologies LLC

Data Science with Python Meetup
2019 March 22

A Beginner's Guide to Docker (for Data Science or otherwise)

- 1. Why Containers? Why Not Containers?**
- 2. What Docker is *Not***
- 3. What Docker *Is***
- 4. Lingo: Images and Containers**
- 5. Architecture: the Docker daemon**
- 6. Demo: Outside and Inside Containers**
- 7. Demo: Docker for Data Science**
- 8. Security**
- 9. Where to Go Next?**

Why / Why Not Containers?

Why

Portability: share & deploy with confidence

Scalability: 10X , 100X easy

For Python in production: replaces virtualenvs

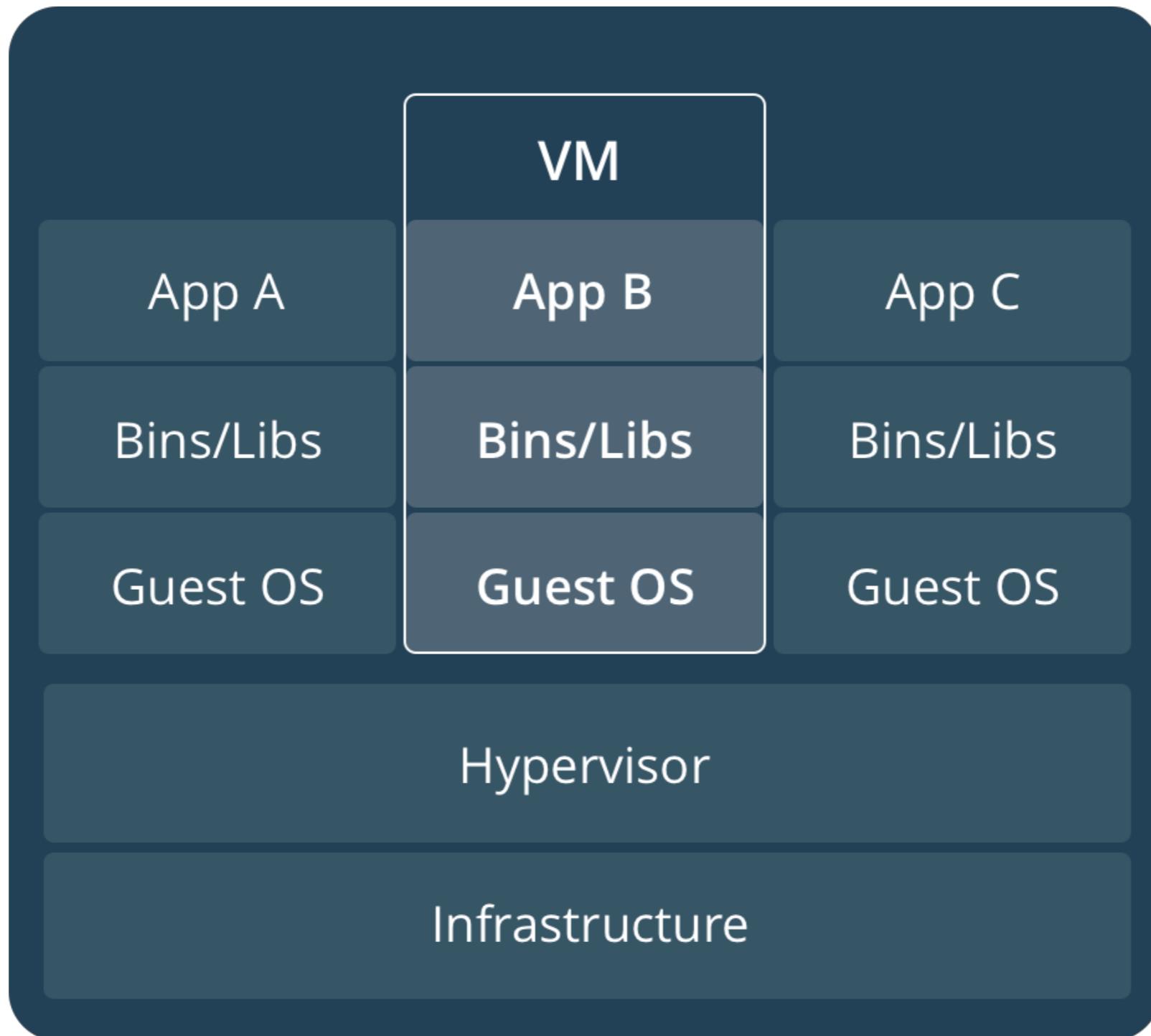
Why Not

Dedicated servers: e.g., databases

Security: Potential risks

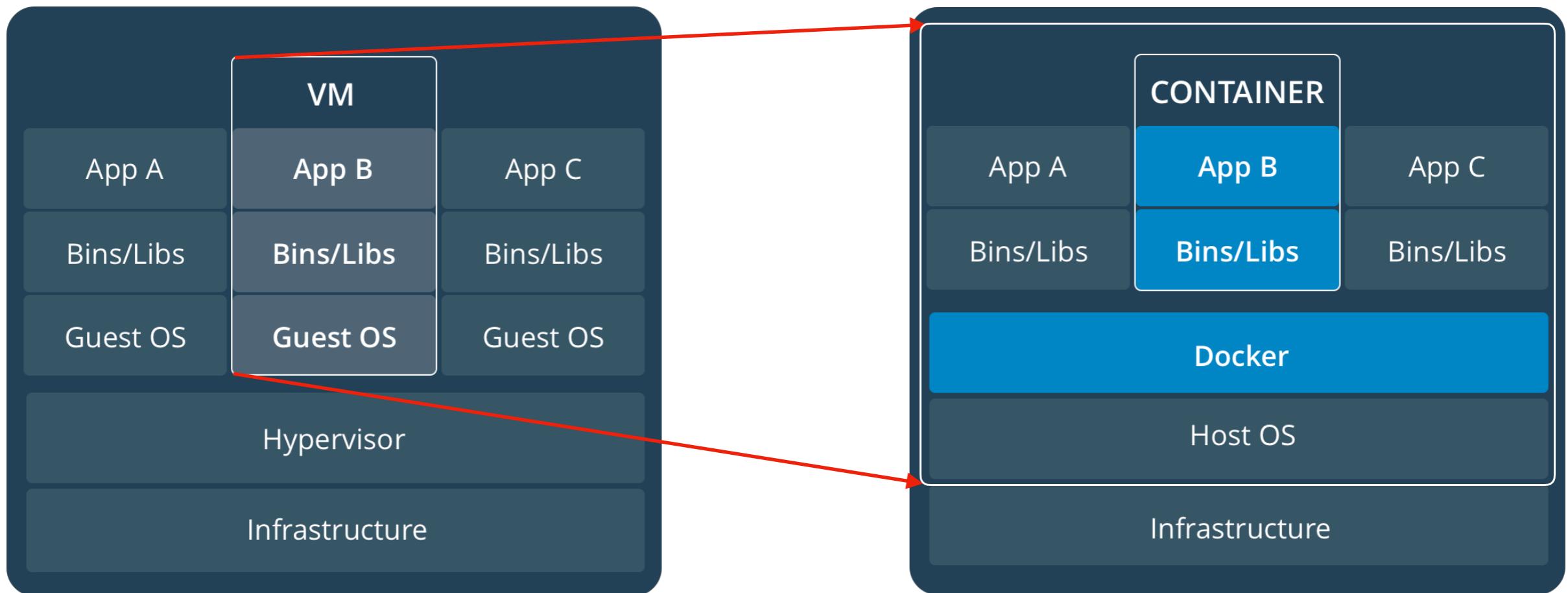
For Python not in production: virtualenvwrapper is still great

What Docker is Not



Source: <https://docs.docker.com/get-started/>

What Docker *Is*

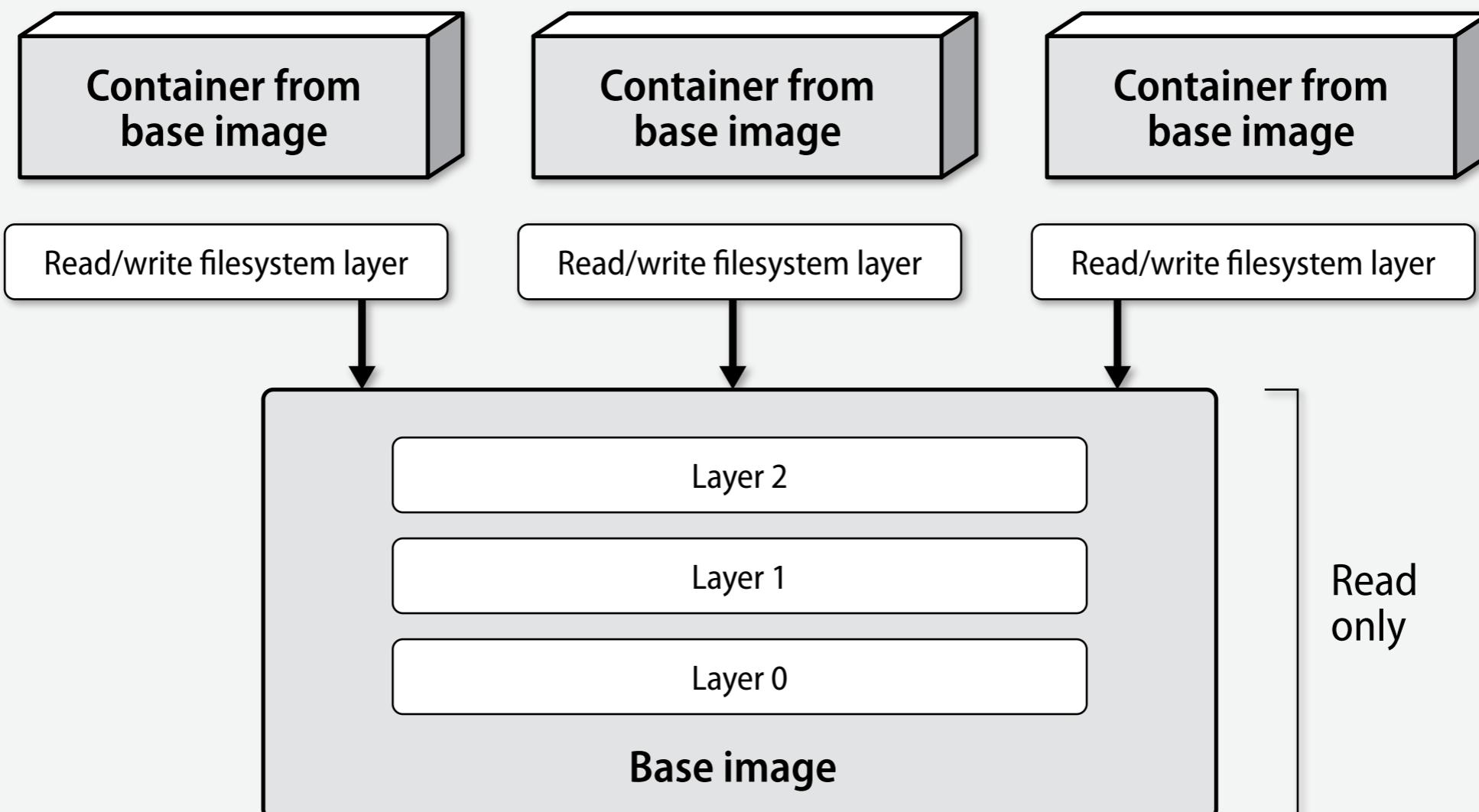


See also:

<https://docs.docker.com/engine/docker-overview/#the-underlying-technology>

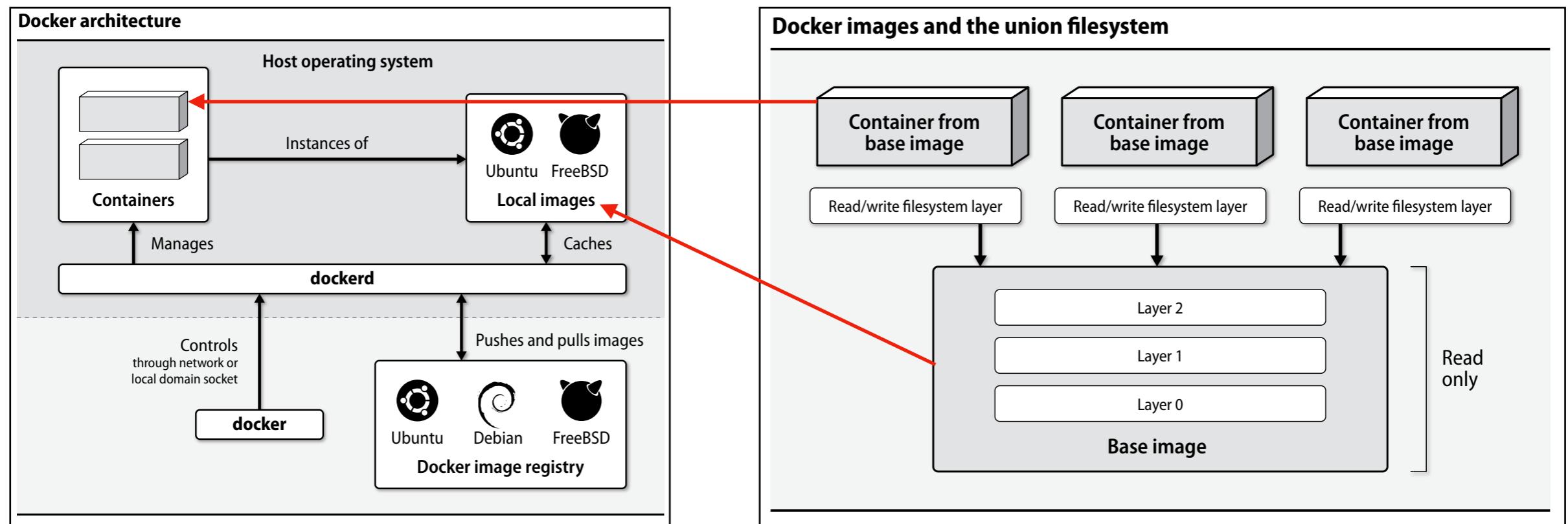
Lingo: Images and Containers

Docker images and the union filesystem



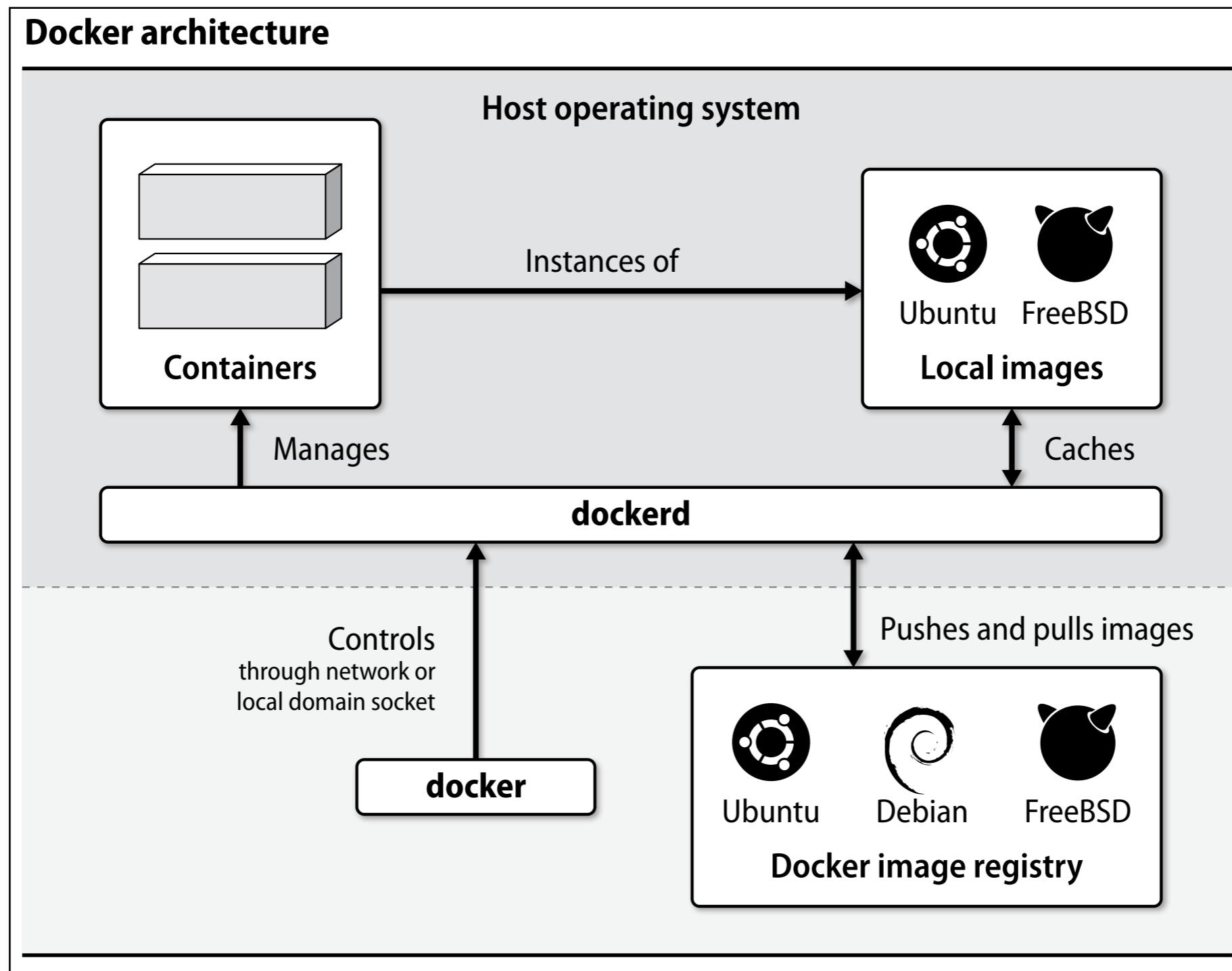
Source: [Unix and Linux System Administration Handbook, 5th ed. \(2017\), Ch. 25 \(a.k.a. my linux bible\)](#)

Architecture: the Docker daemon



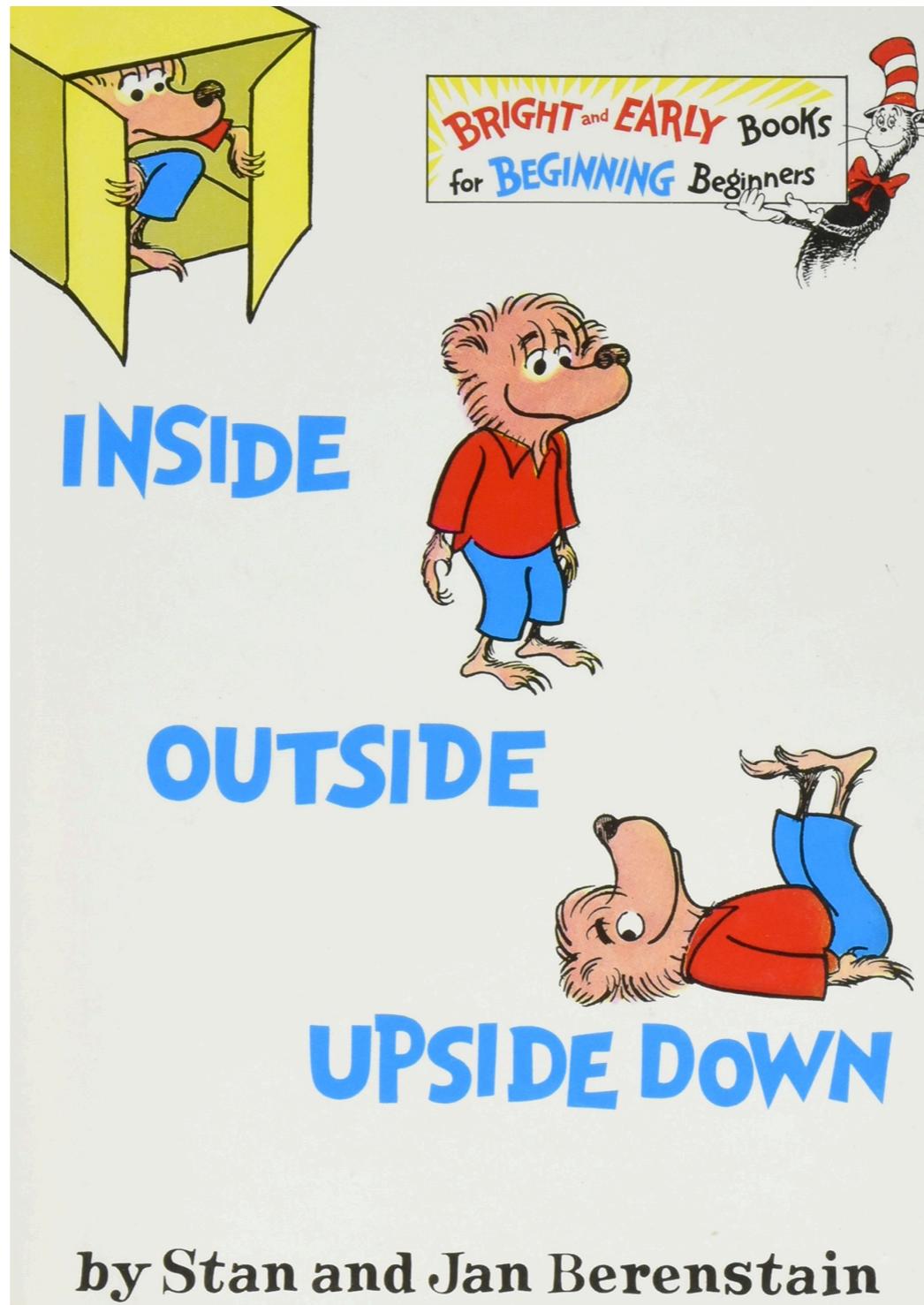
Source: Unix and Linux System Administration Handbook, 5th ed. (2017), Ch. 25
(a.k.a. my linux bible)

Architecture: the Docker daemon

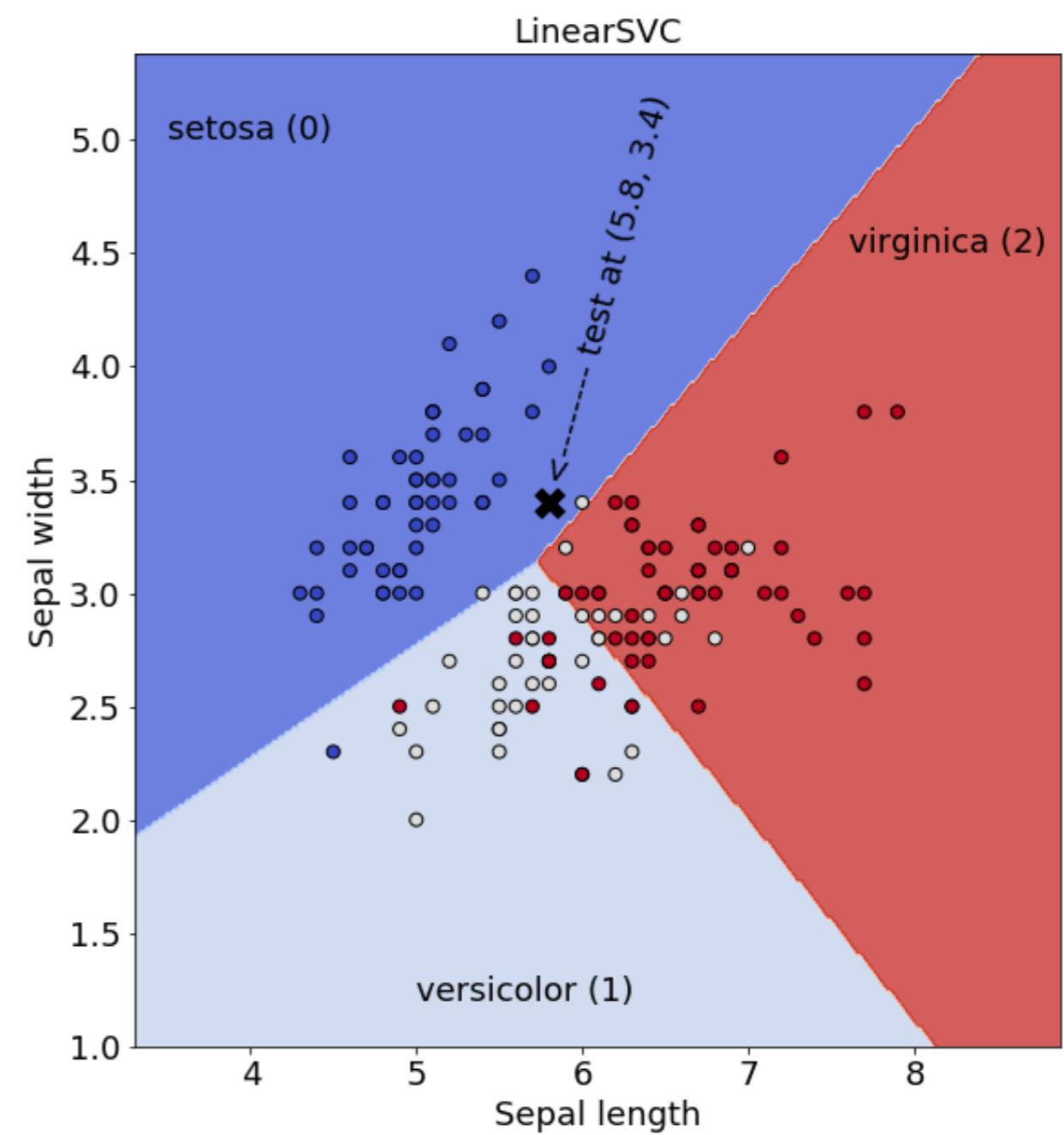
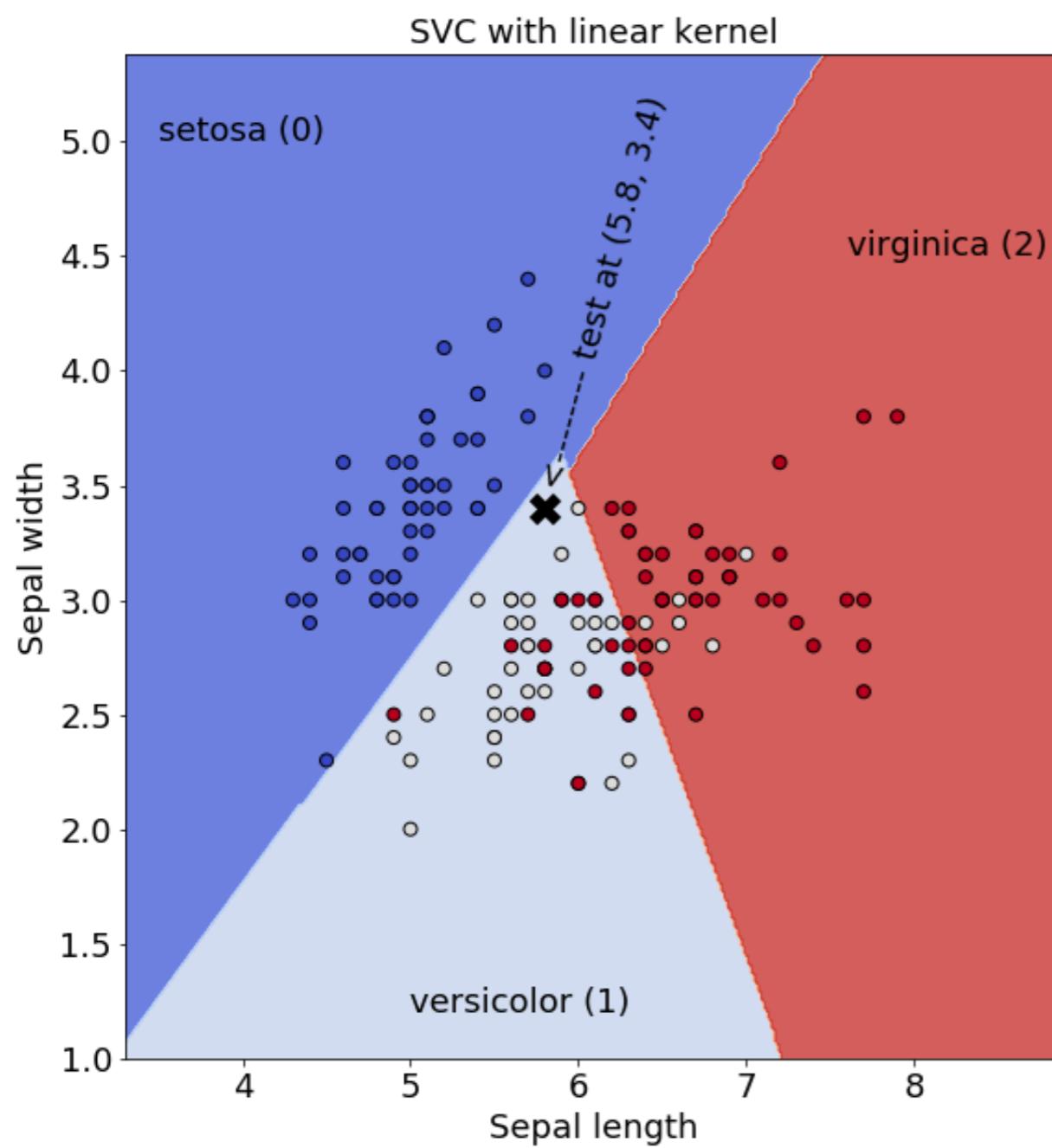


Source: [Unix and Linux System Administration Handbook, 5th ed. \(2017\), Ch. 25 \(a.k.a. my linux bible\)](#)

Demo: Outside and Inside Containers



Demo: Docker for Data Science



Adapted from: https://scikit-learn.org/stable/auto_examples/svm/plot_iris.html

Security

Privilege Escalation: don't run as root

Untrusted Images: caveat emptor on <https://hub.docker.com/>

Recommended reading:

<https://opensource.com/business/14/7/docker-security-selinux>

<https://www.helpnetsecurity.com/2019/03/12/container-escapes/>

<https://devclass.com/2019/03/04/runc-vuln-still-likely-lingering/>

Where to Go Next?

Learn Docker

The docs: <https://docs.docker.com/get-started/>

ULSAH Chapter 25: www.admin.com

Using Containers in Production

Kubernetes: <https://kubernetes.io/>

Swarm: <https://docs.docker.com/engine/swarm/>

AWS ECS (EC2 Container Service): <https://aws.amazon.com/ecs/>

(Extra Slides, use as needed)

VMs versus Containers

Comparing virtual machines with containers

Virtual machine	Container
A full-fledged OS that shares underlying hardware through a hypervisor	An isolated group of processes managed by a shared kernel
Requires a complete boot procedure to initialize; starts in 1-2 minutes	Processes run directly by the kernel; no boot required; starts in < 1 second
Long-lived	Frequently replaced
Has one or more dedicated virtual disks attached through the hypervisor	Filesystem view is a layered construct defined by the container engine
Images measured in gigabytes	Images measured in megabytes
A few dozen or fewer per physical host	Many per virtual or physical host
Complete isolation among guests	OS kernel and services shared with host
Multiple independent operating systems running side by side	Must run the same kernel as the host (OS distribution may differ)

Source: [Unix and Linux System Administration Handbook, 5th ed. \(2017\), Ch. 24 \(a.k.a. my linux bible\)](#)