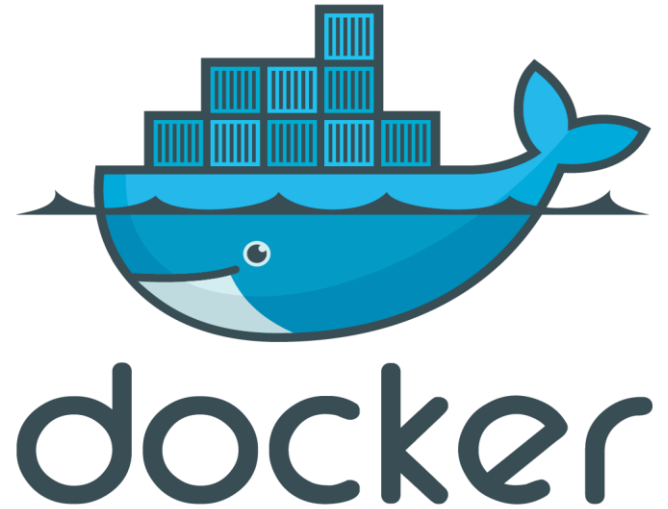


# Docker & Big Data Processing



# Instructor

Fahimeh Ashrafi

Senior System & Network Administrator

M.A Network Information Technology



# Agenda:

- What is docker?
- A History Lesson
- Containers vs Virtual Machines
- Docker Platform Overview and Terminology
  - Docker Engine
  - Images
  - Container
- Docker: Data science environment



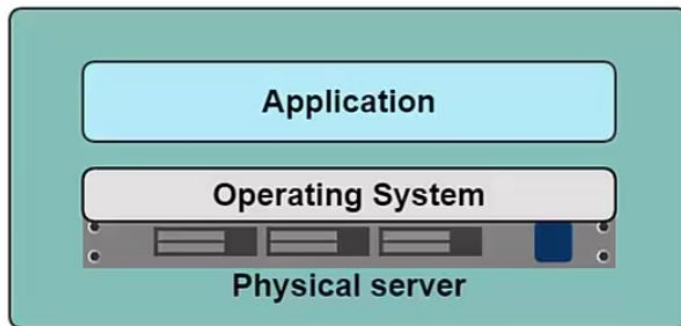
# What is Docker?



# A History Lesson

In the Dark Ages

## One Application on one physical server



# Problem In The Past

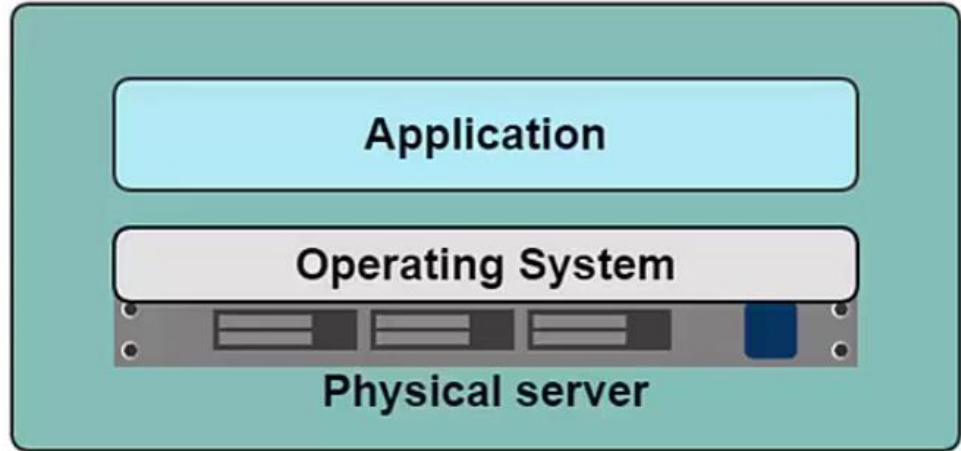
Slow deployment times

Huge costs

Wasted resources

Difficult to scale

Difficult to migrate

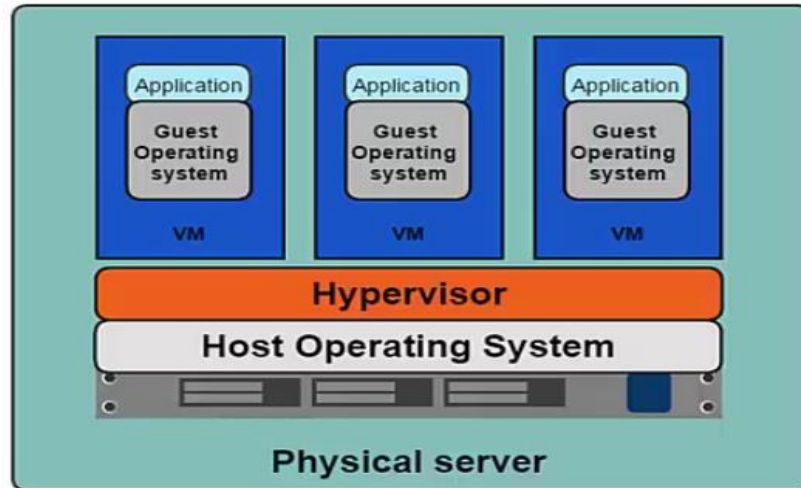


# A History Lesson

Hypervisor-based virtualization

One physical server can contain multiple applications

Each applications runs in a virtual machine



# Benefits of VMs

Better resource pooling

One physical machine divided into multiple virtual machines

Easier to scale

VM's in the cloud

Rapid elasticity

Pay as you go model





# Limitations of VMs

Each VM still requires

CPU allocation

Storage

RAM

An entire guest operating system

I/O time

The more VM's you run, the more resources you need

Guest OS means wasted resources

Application portability not guaranteed



# Introducing Containers

Container based virtualization use the kernel on the host's operating system to run multiple guest instances

Each guest instance is called a container

Each container has its own

Root filesystem

Processes

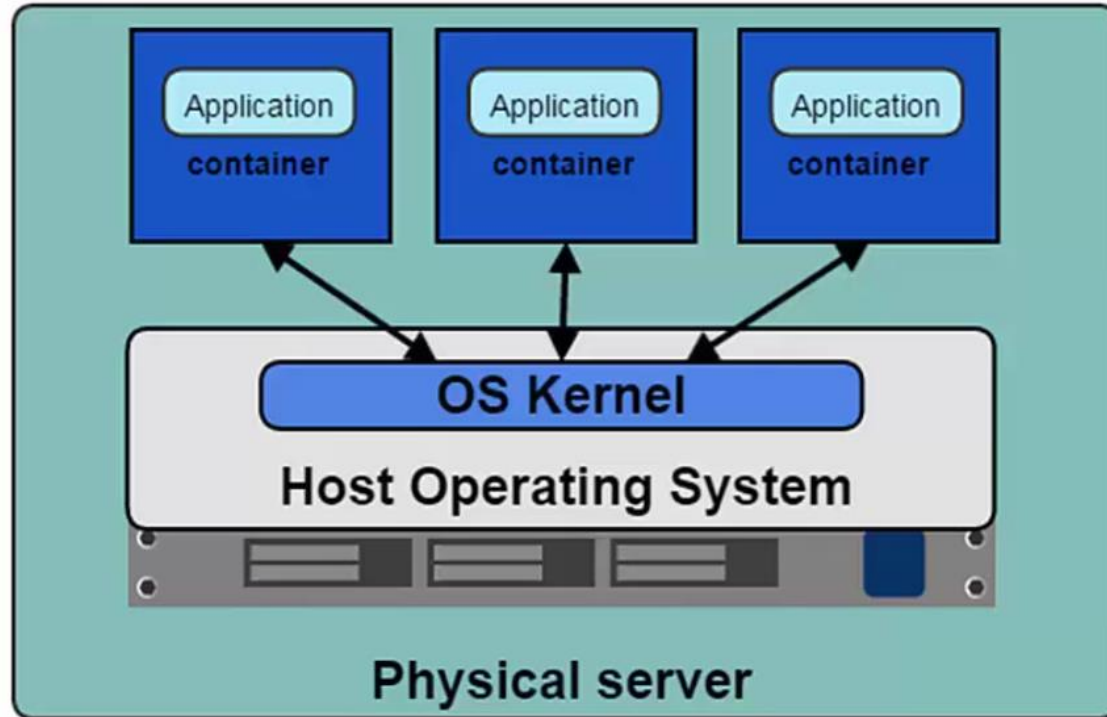
Memory

Devices

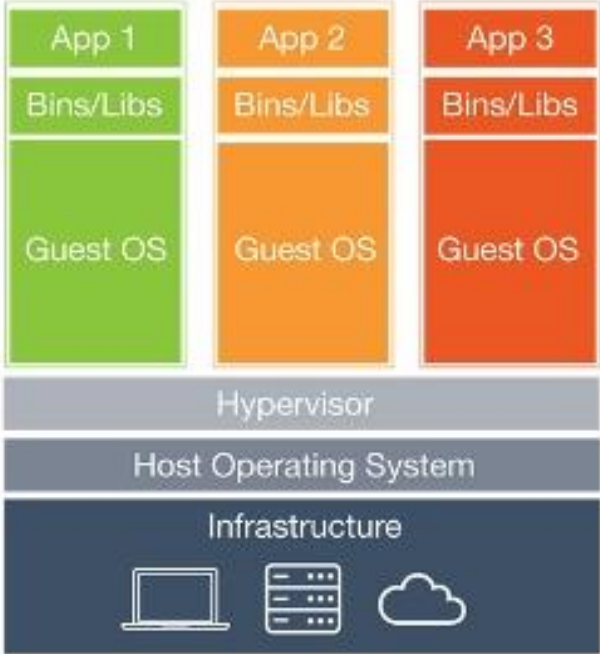
Network ports



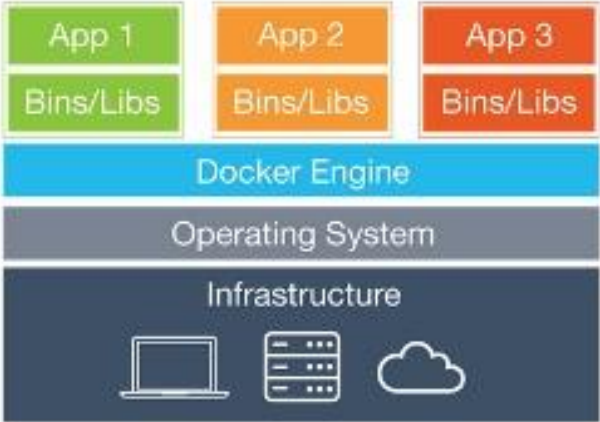
# Containers



# Containers VS VM's



Virtual Machines



Containers



# Docker and the Linux Kernel

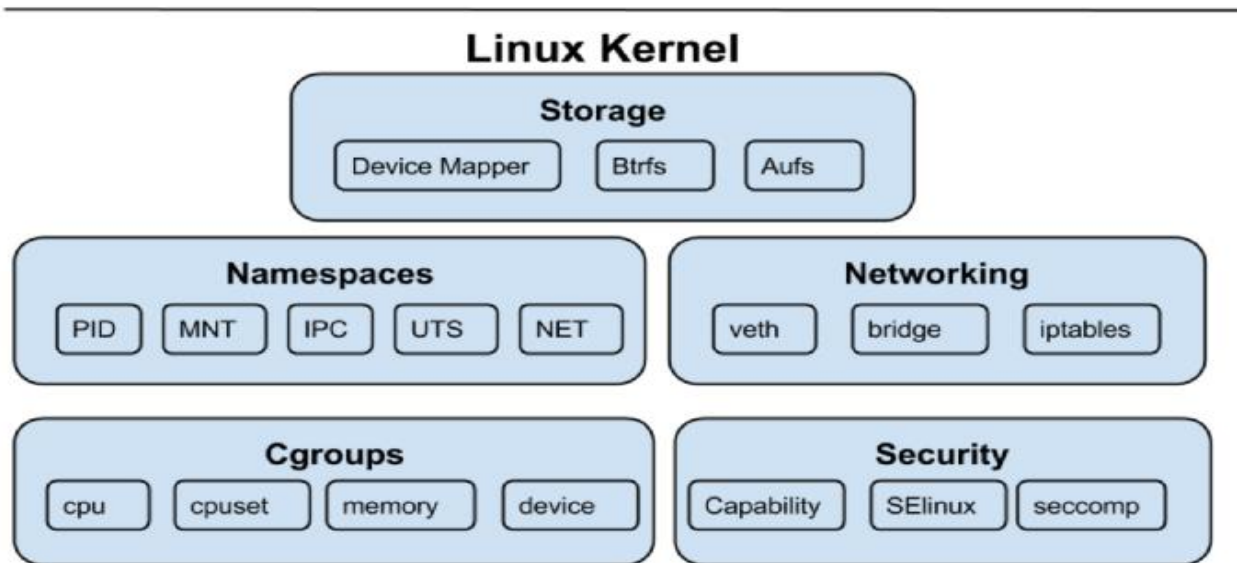
Docker Engine is the program that enables containers to be built, shipped and run.

Docker Engine uses Linux Kernel namespaces and control groups

Namespaces give us isolated workspace



# Docker and the Linux Kernel



# Docker Containers and Images

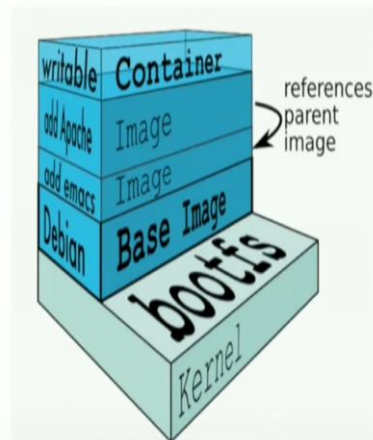
## Images

Read only templates(Dockerfile) used to create containers

Build by you or other Docker users

Stored in the Docker Hub or your local registry

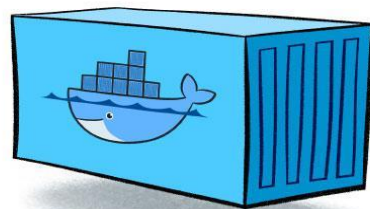
Image layers



## Containers

Isolated application platform

Contains everything needed to run your application



# Benefits of Docker

Separation of concerns

Developer focus on building their apps

System admin focus on deployment

Fast development cycle

Application portability

Build in one environment, ship to another

Scalability

Easily spin up new containers if needed

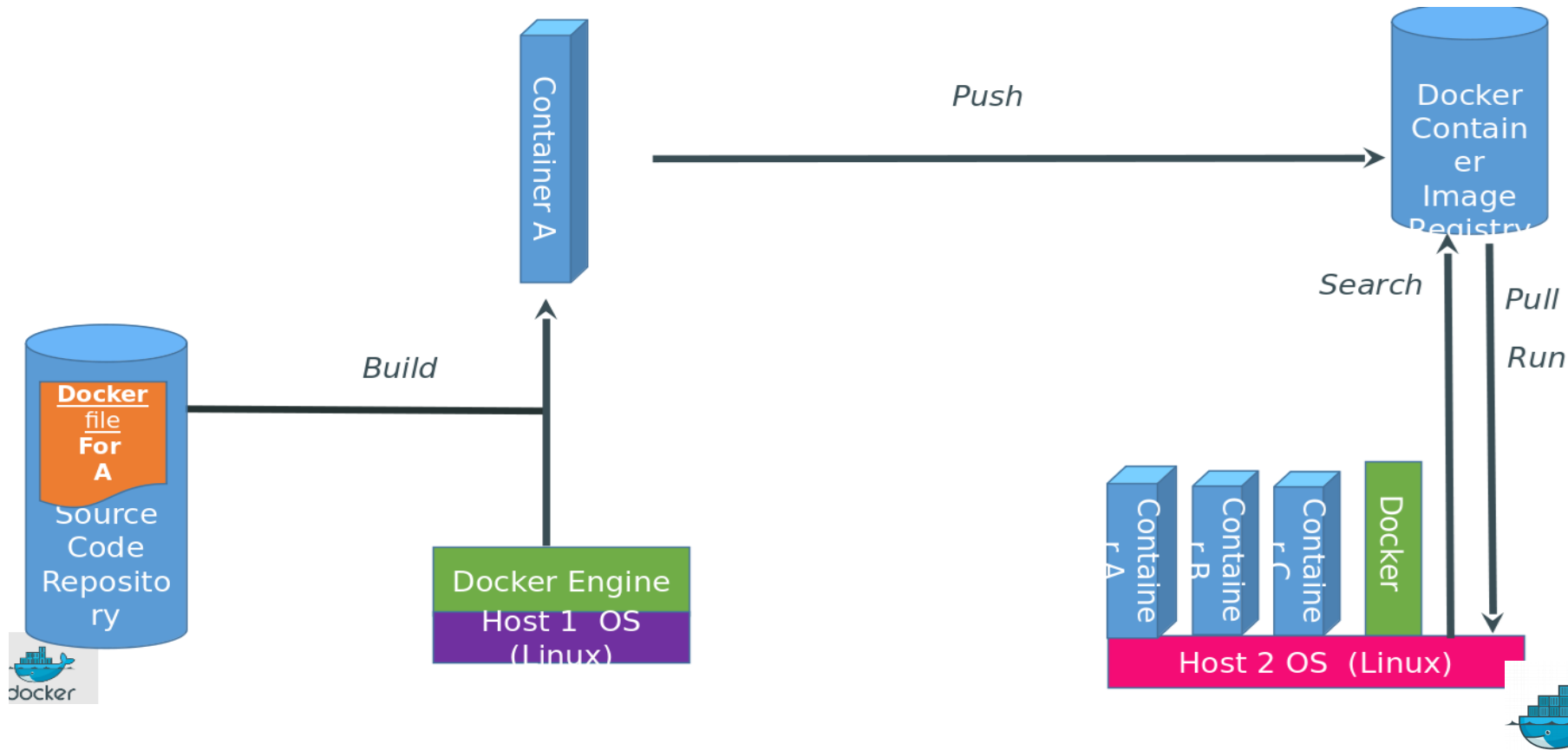
Run more apps on one host machine

Version control and component reuse

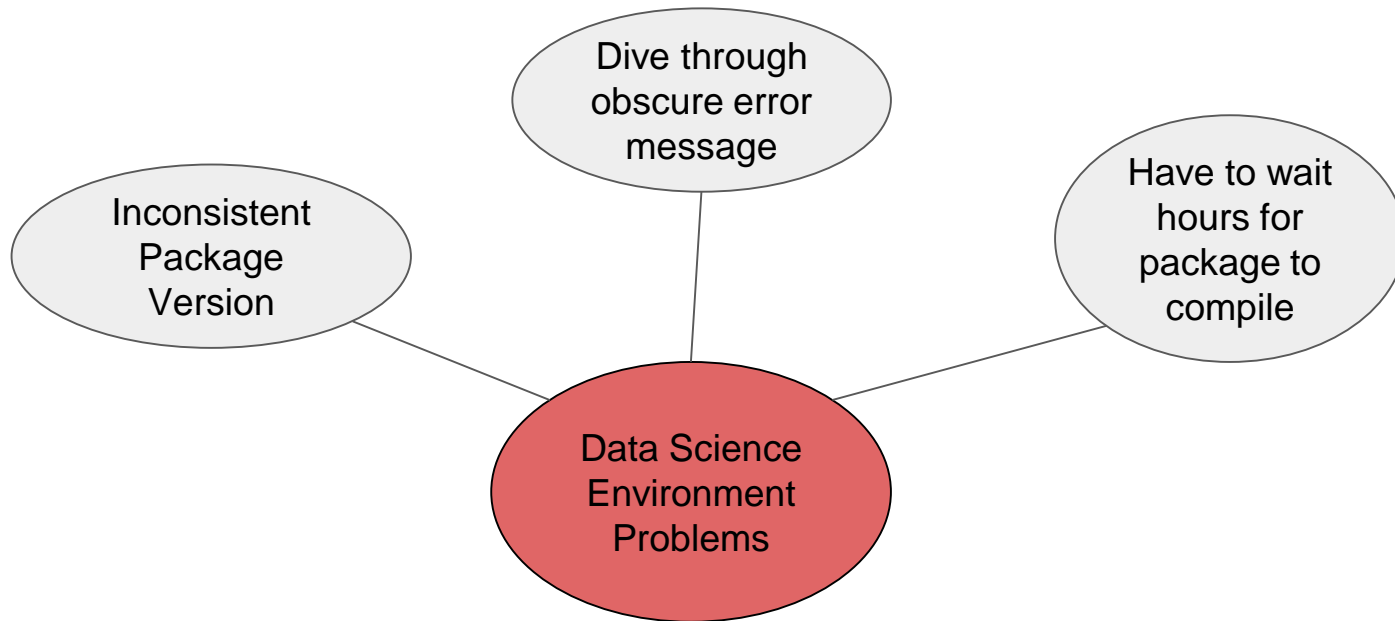




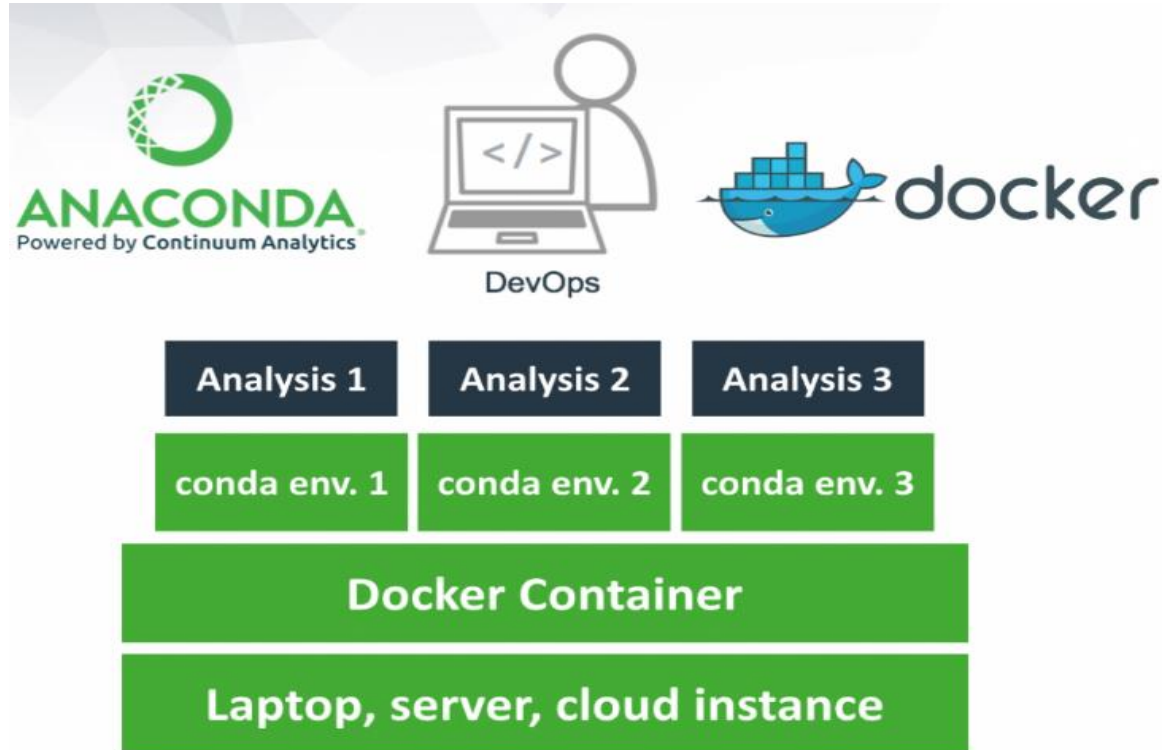
# What are the basics of the Docker system?



# Docker: Data Science Environment



# Docker: Data Science Environment



**Thanks for your Attention**

**Any question?**

