

Stream Processing

Lab 7

2018/2019

Docker (<https://www.docker.com/>)

- Docker is an open platform for developers and sysadmins to build, ship, and run distributed applications, whether on laptops, data center VMs, or the cloud.

Docker

- Provides the tools to package applications, including all the required dependencies in a single linux image.
- Applications can be loaded and executed on demand in any hardware that supports the docker engine - eg., linux, windows, macOS.
- The docker engine runs in real hardware or virtual machines/hypervisors.

Docker containers

- A container image is a **lightweight**, stand-alone, executable package of a piece of software that includes everything needed to run it: code, runtime, system tools, system libraries, settings.
- The key advantage is that to run a given application/container it is not necessary to install a full OS...

Docker containers

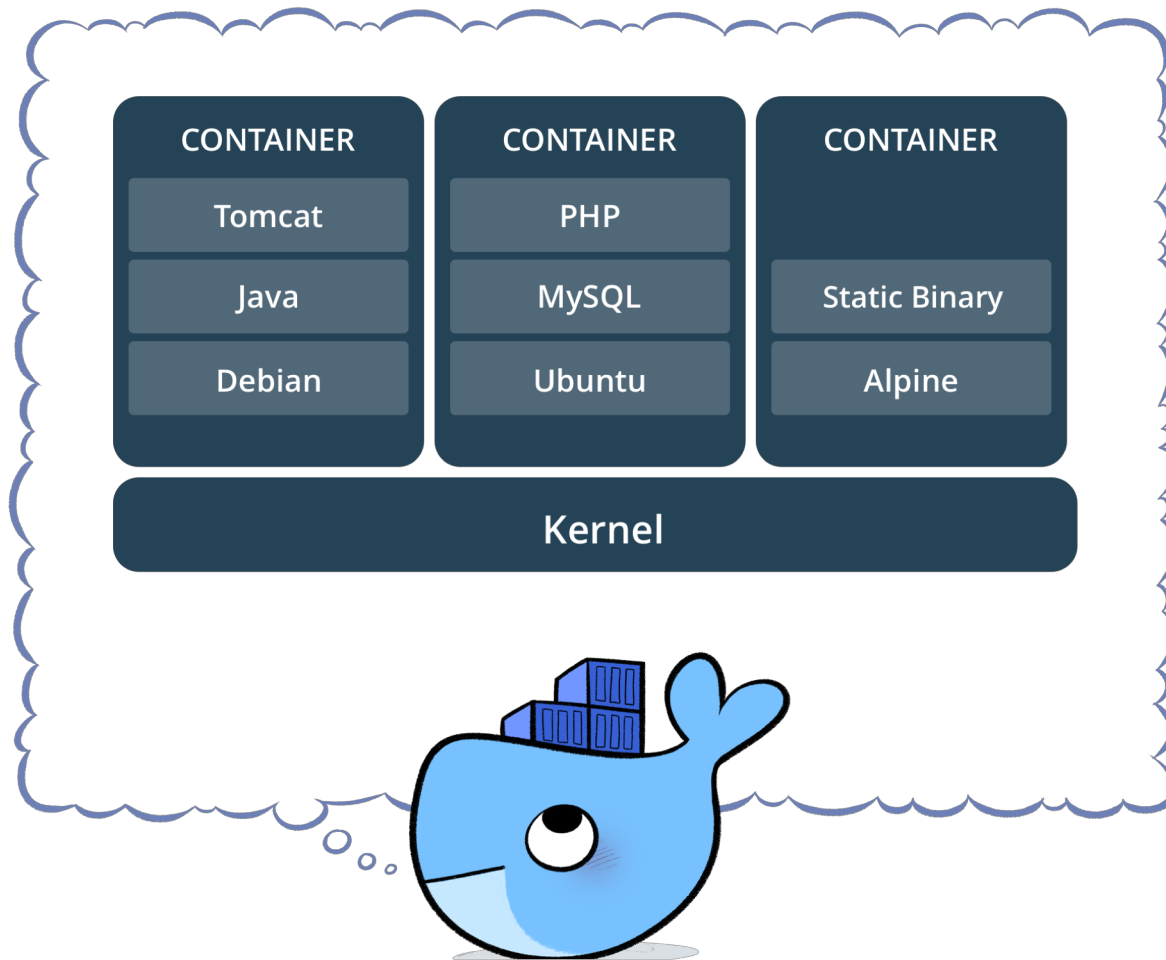
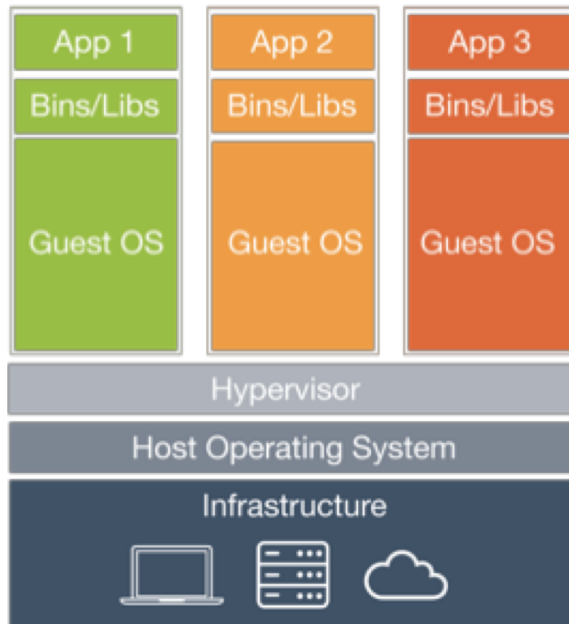


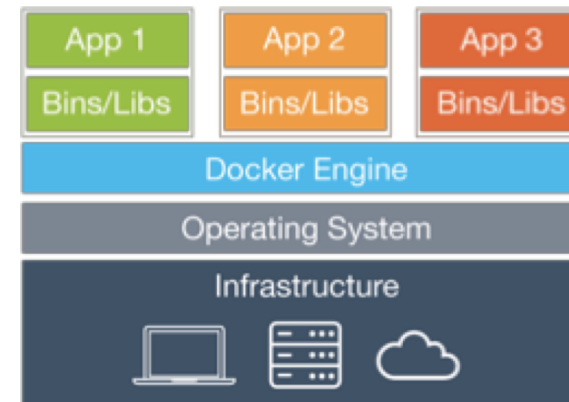
image from <http://www.docker.com>

Docker containers vs Virtual Machine

Virtual Machine



Docker containers



Docker engine

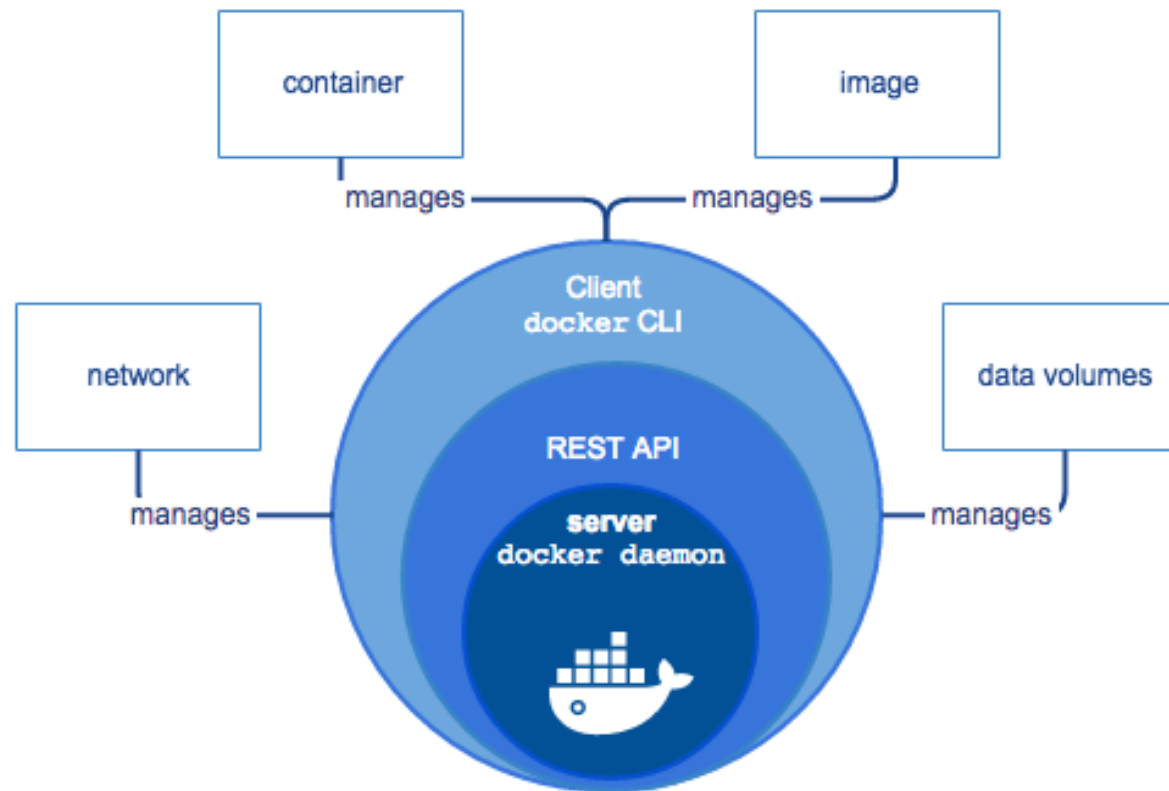


image from <http://www.docker.com>

Docker cli

```
> docker run hello-world
```

```
Unable to find image 'hello-world:latest' locally  
latest: Pulling from library/hello-world
```


Docker cli

Start/run a container

```
> docker run -ti ubuntu
```

List running containers

```
> docker ps -a
```

Attach a shell to a running container (id obtained in docker ps)

```
> docker exec -it <container> bash
```

Remove/delete a container

```
> docker rm -f <container>
```

Docker cli

List local images

```
> docker images
```

Remove image

```
> docker rmi -f <image>
```

Remove all unused containers, networks and images.

```
> docker system prune
```

```
> docker help
```

docker-compose

- A tool for defining and running multi-container Docker applications, for example: anaconda + hadoop.
- Results in an application that can span multiple nodes, emulated as different containers.
- Allows applications to be scaled up and down;
 - for instance, increase the number of hadoop workers from 3 to 5

docker community

(<https://hub.docker.com/>)

- There are many docker images already available provided by the docker community
 - They are downloaded automatically by the docker cli given a container tag.
- > `docker run continuumio/anaconda ...`
- full instructions are usually provided in the description page of the docker image...

Jupyter notebook

- “Open-source web application that allows you to create and share documents that contain live code, equations, visualizations and narrative text. Uses include: data cleaning and transformation, numerical simulation, statistical modeling, data visualization, machine learning, and much more.”

Jupyter notebooks

- Easy way for producing tutorial, where each step of the code can be explained and run.

Let's try it...

- Install docker in your own computer:
- <https://docs.docker.com/install/>

Lab 7

- Docker image with Jupyter, Spark Streaming and Python

➤ `sh run.sh`

- Open Jupyter in a browser, using the URL presented when starting the docker image.
- Notebooks are in directory work.