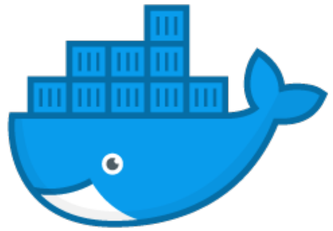
What the heck is Docker?!

Introduction to Big Data & Data Engineering - April 2023

What is Docker?

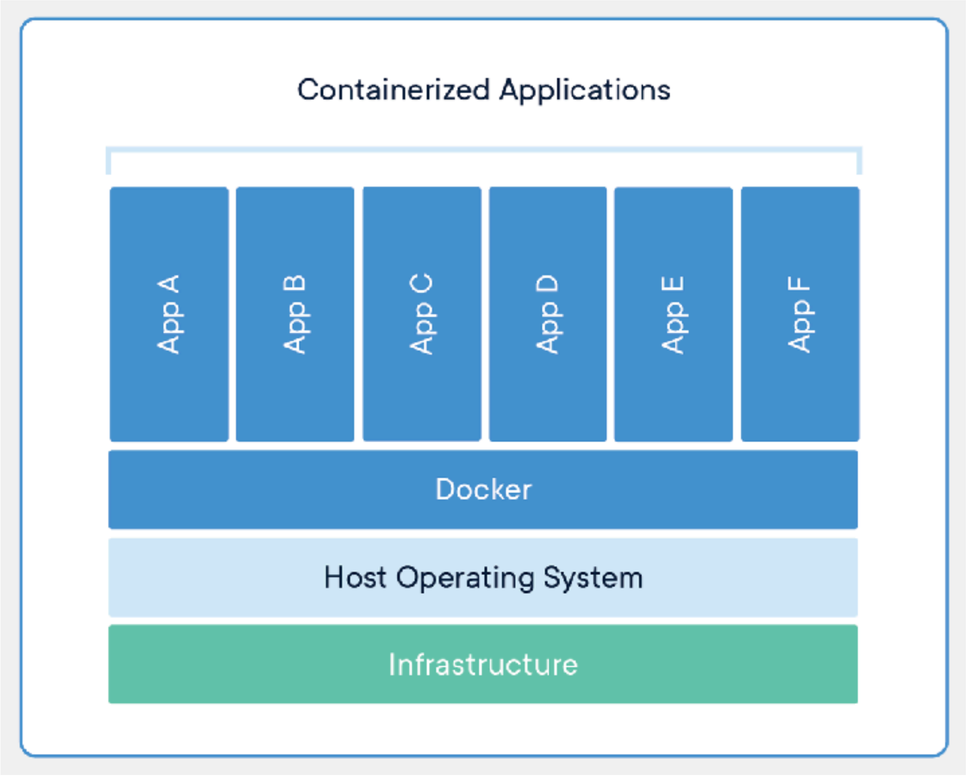


docker

A container is a standard unit of software that packages up code and all its dependencies so the application runs quickly and reliably from one computing environment to another.

A Docker container image is a lightweight, standalone, executable package of software that includes everything needed to run an application: code, runtime, system tools, system libraries and settings.

Containers isolate software from its environment and ensure that it works uniformly despite differences for instance between development and staging.



Software production: Matrix from hell

django web frontend	?	?	?	?	?	?
node.js async API	?	?	?	?	?	?
background workers	?	?	?	?	?	?
SQL database	?	?	?	?	?	?
distributed DB, big data	?	?	?	?	?	?
message queue	?	?	?	?	?	?
	my laptop	your laptop	QA	staging	prod on cloud VM	prod on bare metal

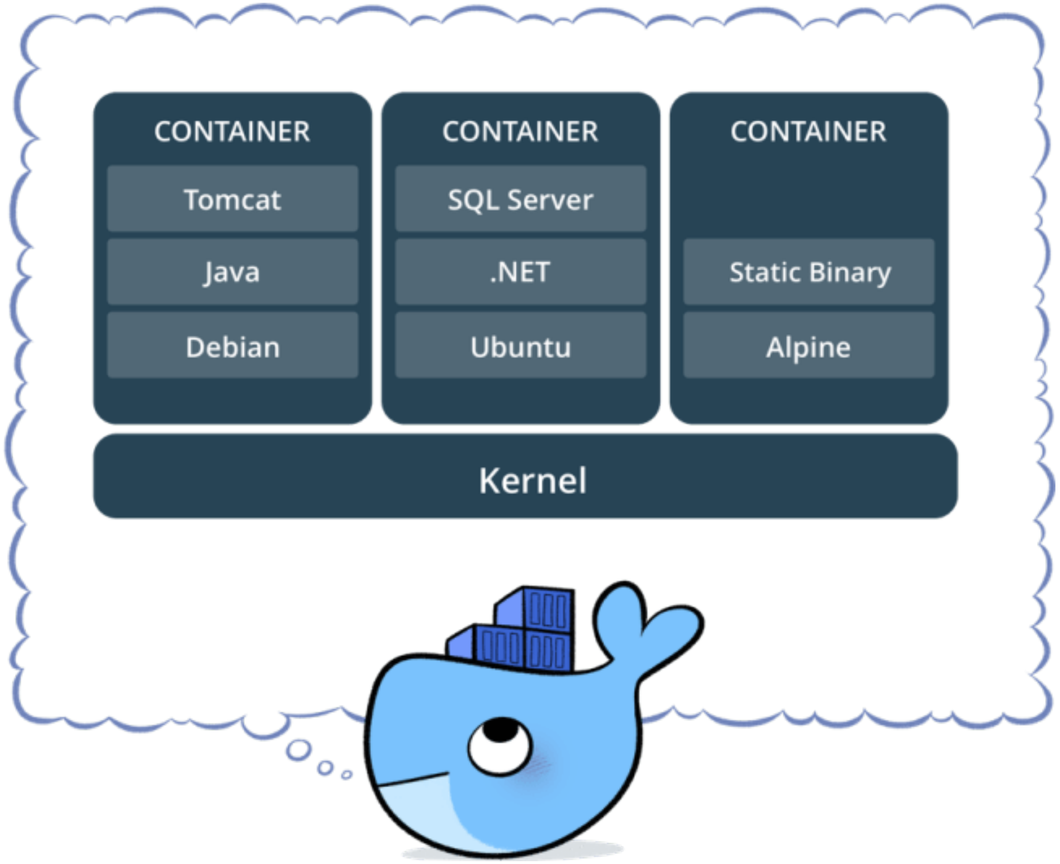
An example from shipping: Matrix from hell

	?	?	?	?	?	?	?
	?	?	?	?	?	?	?
	?	?	?	?	?	?	?
	?	?	?	?	?	?	?
	?	?	?	?	?	?	?
	?	?	?	?	?	?	?
							

The shipping solution

The software solution = Docker

```
FROM node:6-alpine

RUN apk add --update tini \
  && mkdir -p /usr/src/app

WORKDIR /usr/src/app

COPY package.json package.json

RUN npm install \
  && npm cache clean

COPY . .

EXPOSE 3000
# expose these ports on the docker virtual network

CMD ["tini", "--", "node", "./bin/www"]
```



Why might I want to use it?

1. New scientific software
2. Experimenting with other tech
3. Reproducible research
4. Sharing software for use by colleagues
5. Scalable scientific software solutions

Questions?

...