
Pinterest

Accelerating development velocity of production ML systems using docker

Kinnary Jangla

March 7th, 2018

Pinterest is the visual discovery engine.

Mission

Help people discover and do what they love.

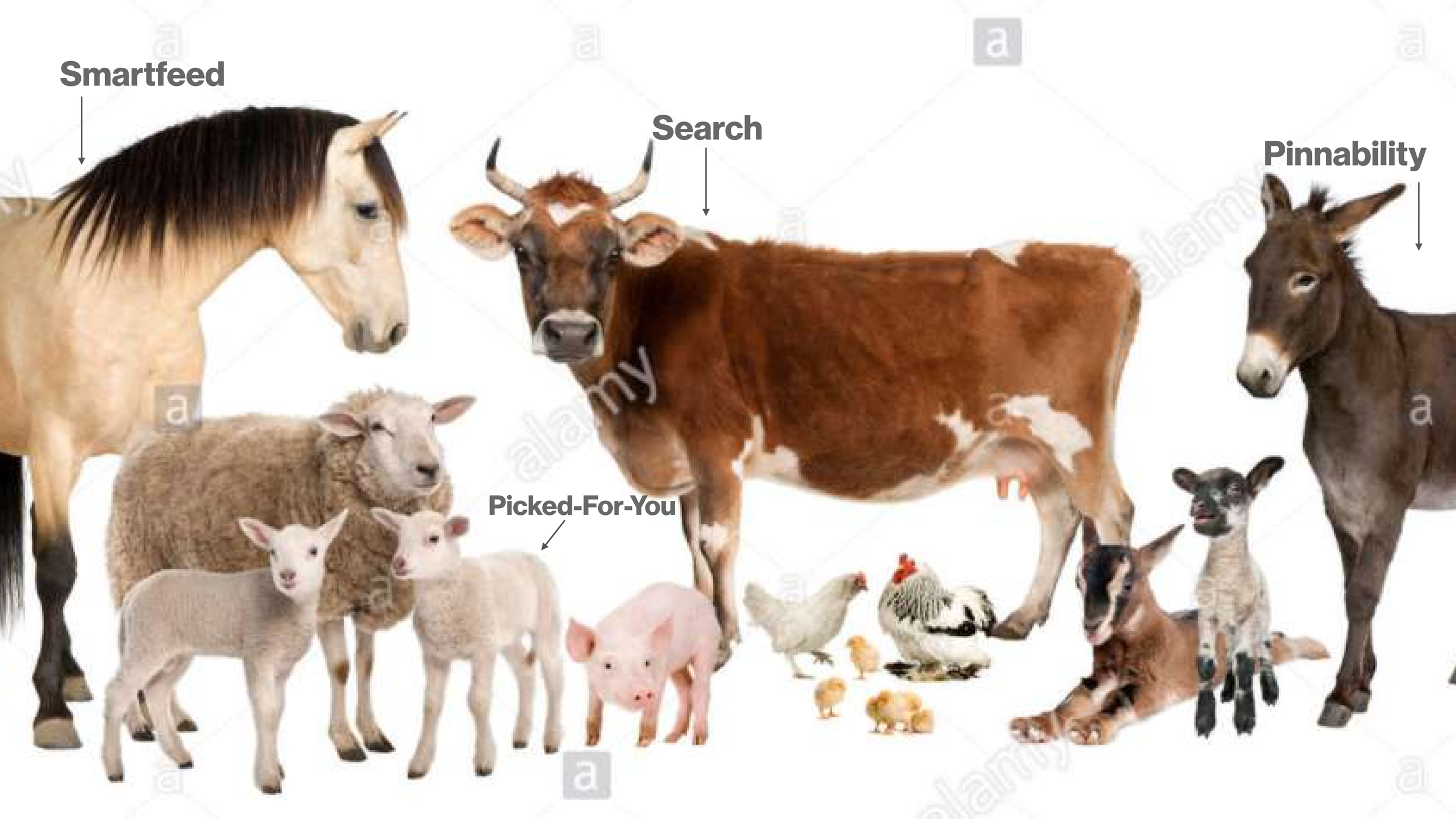
Machine learning @ Pinterest

1 Interest-based feed
2 Recommendations
3 Search
4 Ads
5 Visual Discovery

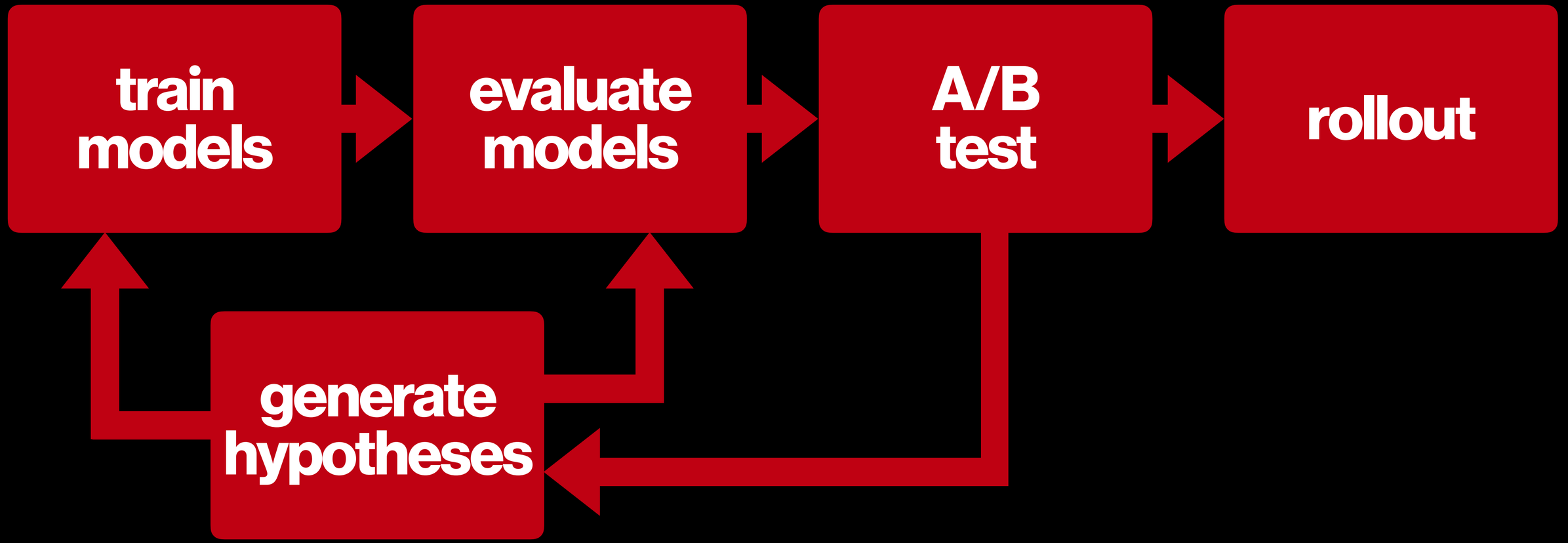
Monolith to Microservices


Monolith

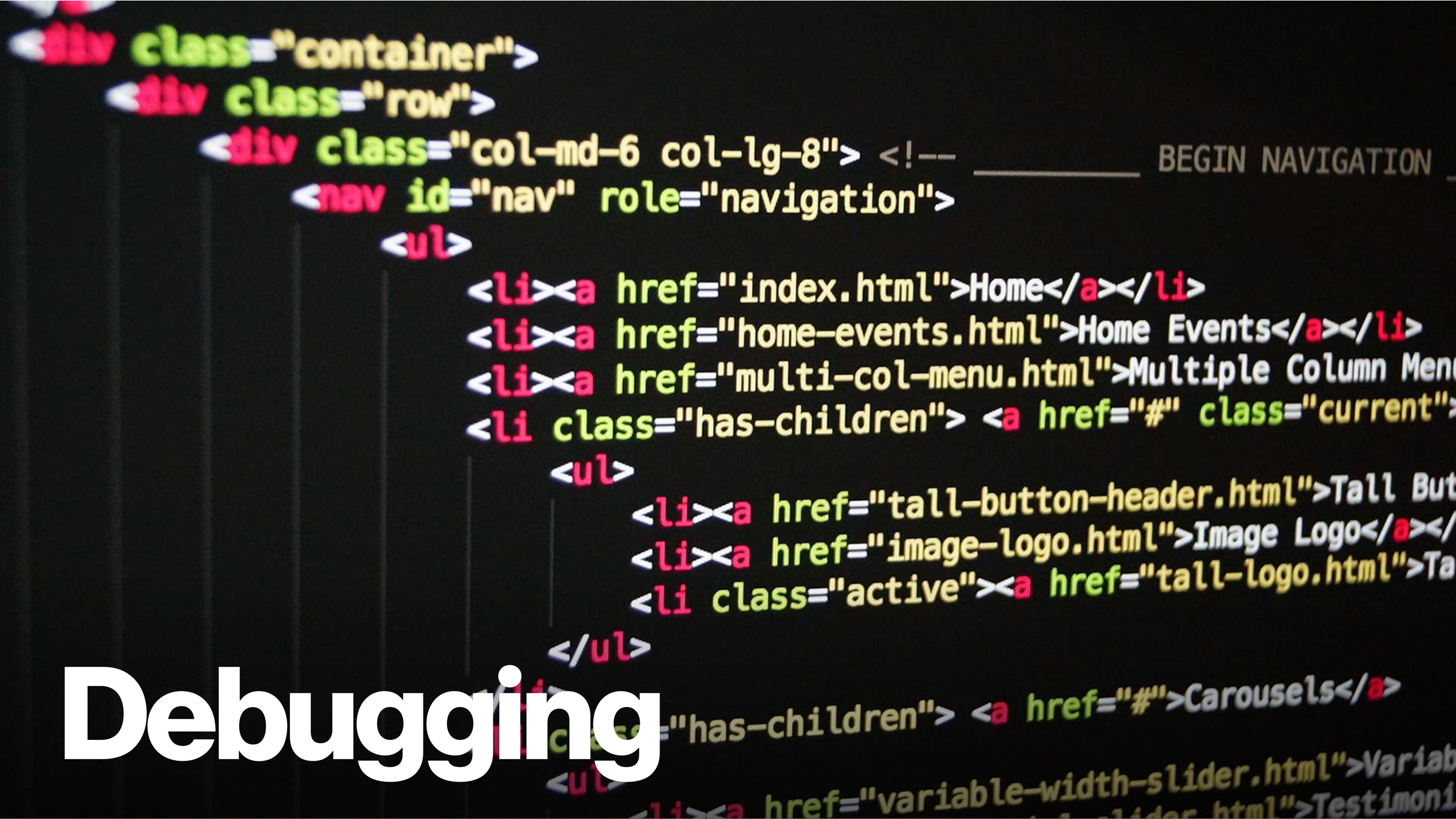
Microservices


Smartfeed
Search
Pinnability
Picked-For-You

Challenges with Microservices



Model training


Debugging


Solutions?

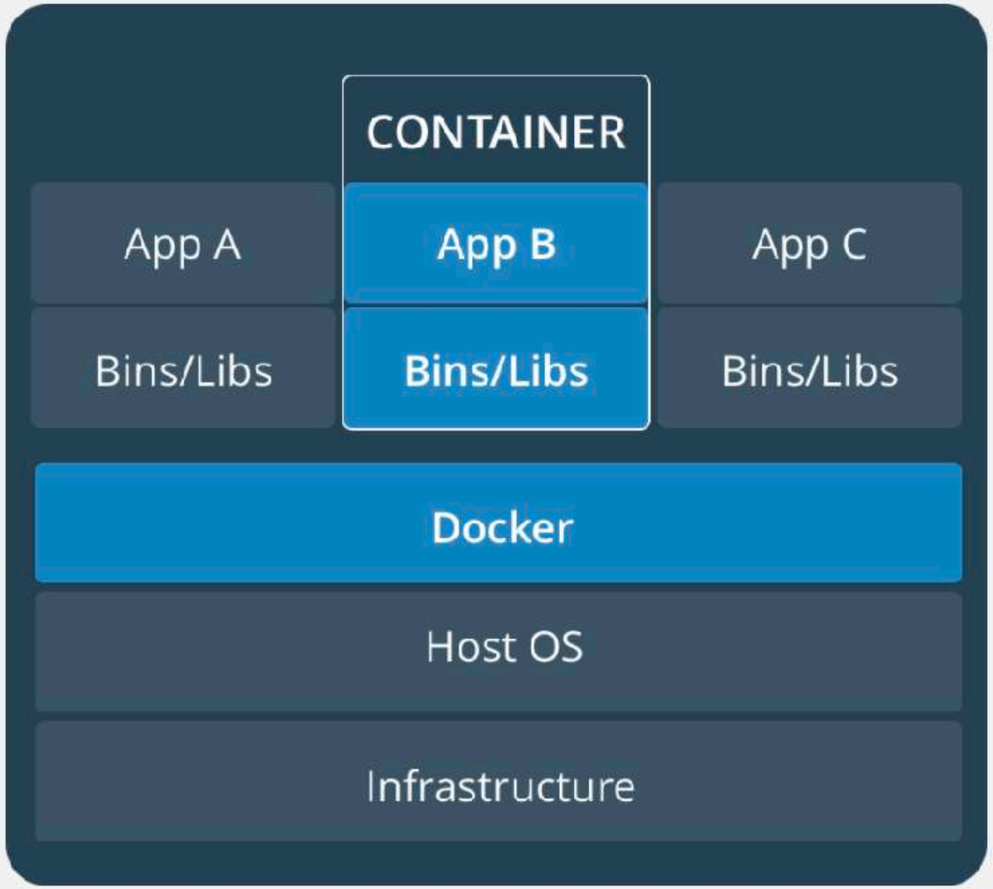
docker

What is...

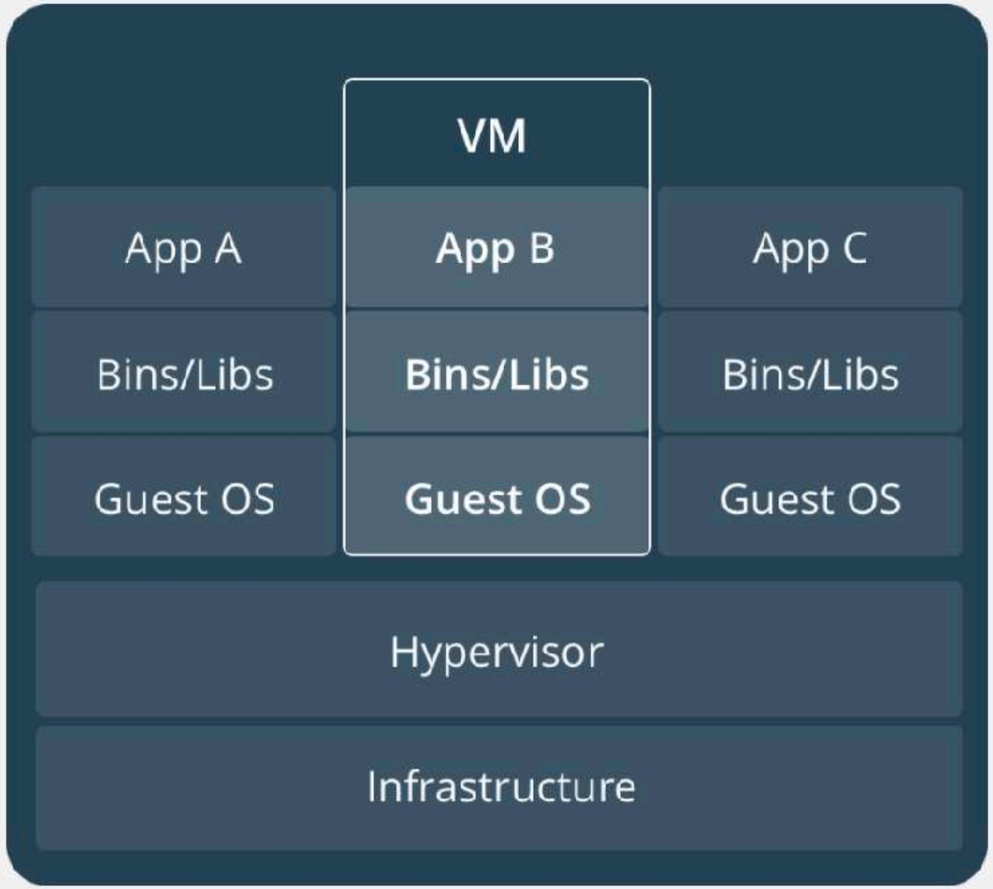
1. Bundle applications with its runtime environment

2. Similar to a virtual machine

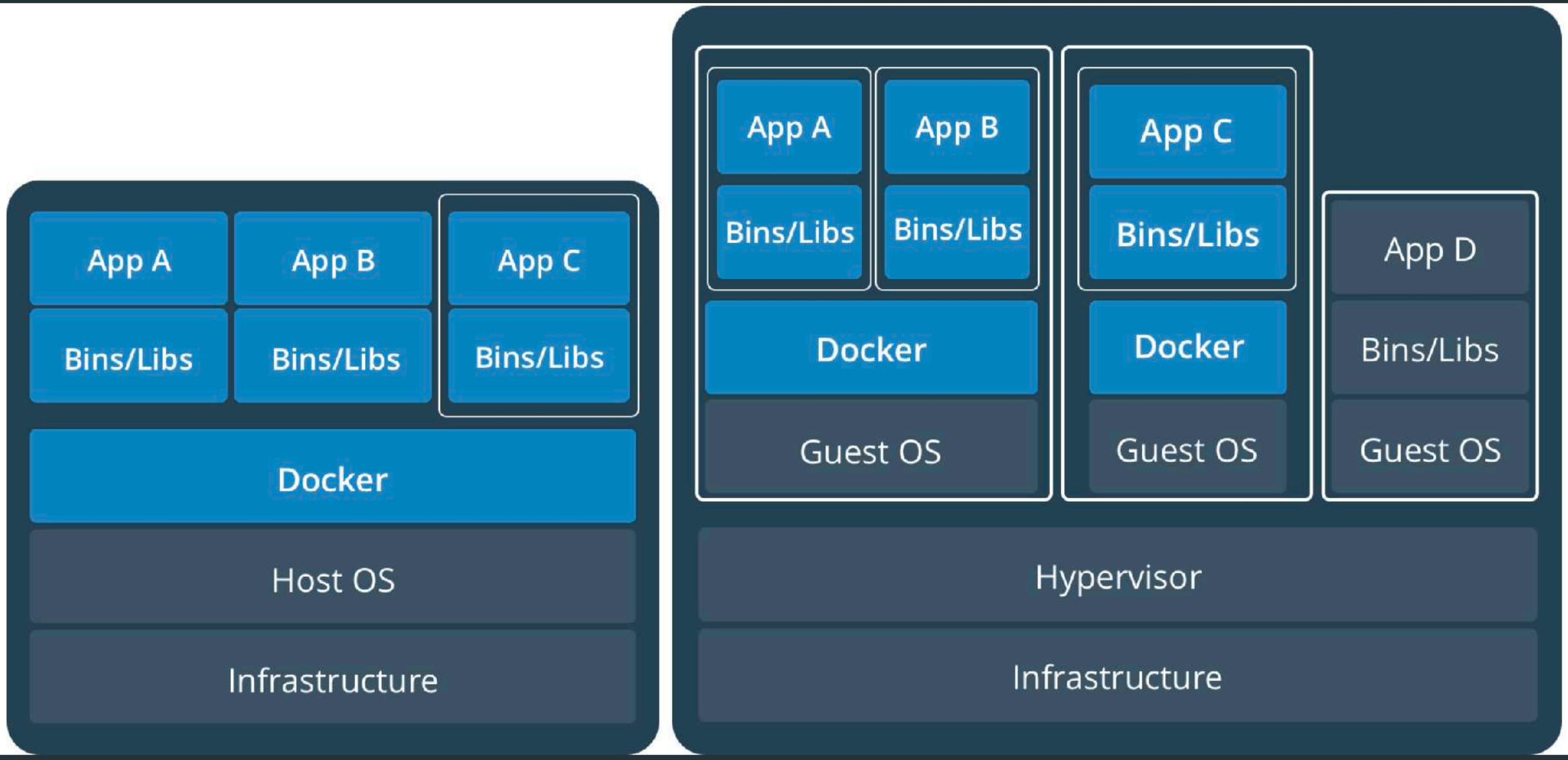
3. Runs on it own network interface



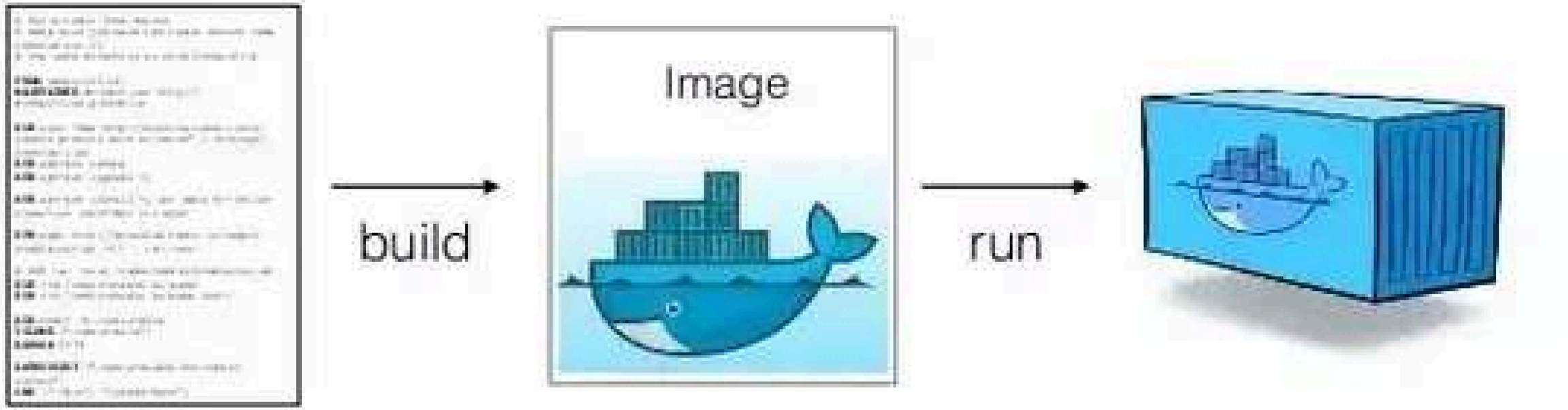
CONTAINERS



VIRTUAL MACHINES


App A
App B
App C
Bins/Libs
Bins/Libs
Bins/Libs
Docker
Host OS
Infrastructure
App A
App B
Bins/Libs
Bins/Libs
Docker
Guest OS
App C
Bins/Libs
Docker
Guest OS
App D
Bins/Libs
Guest OS
Hypervisor
Infrastructure

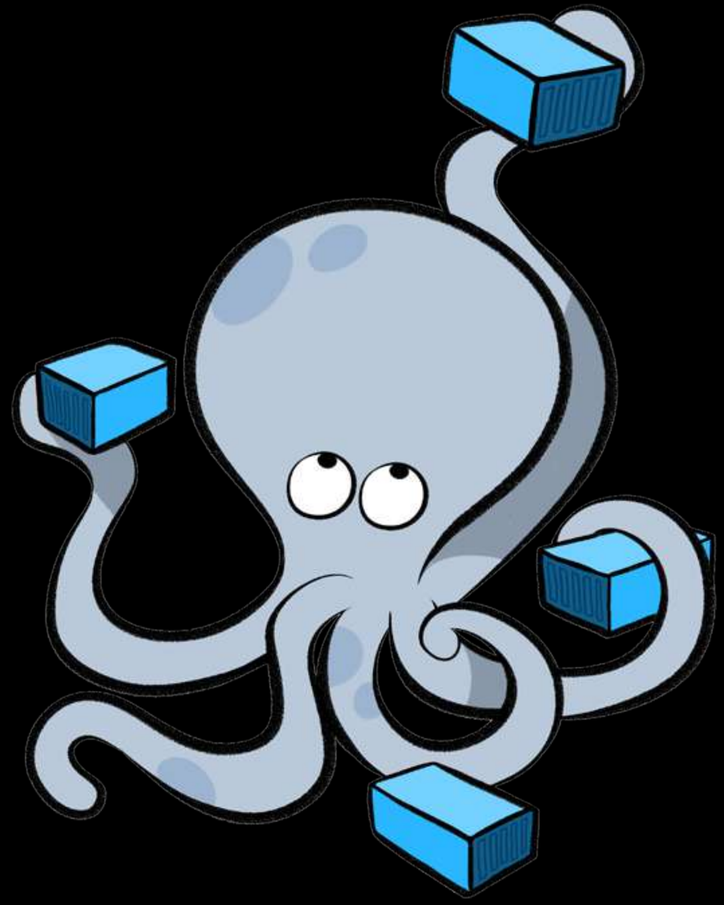
Single Container Docker Workflow



```
# Refer https://docs.docker.com/engine/userguide/eng-image/dockerfile_best-practices/
# for best practices maintaining this file.
# This file is used to build docker images for development purposes only.
# This file cannot be used to build docker images on production.

# Base image from https://phabricator.pinadmin.com/diffusion/BDI
FROM 998131032990.dkr.ecr.us-east-1.amazonaws.com/oracle-java8:latest

# Default environment variables required to run service, can be overridden by docker run
ENV CONFIG_FILE=config/instantpfy.test.properties \
    HEAP_SIZE=4G \
    LOG4J_CONFIG_FILE=config/log4j_instantpfy.xml \
    NEW_SIZE=2G

# Create and set current directory
WORKDIR /opt/fantasio

# Add the build artifact under /opt, can be overridden by docker build
ARG ARTIFACT_PATH=target/fantasio-0.1-SNAPSHOT-bin.tar.gz
ADD $ARTIFACT_PATH /opt/fantasio

# Default command to run service, do not override it in docker run unless have a good reason
# Use "docker logs ID" to view stdout and stderr
CMD ["scripts/run_in_container.sh"]
~
~
~
~
~
~
~
~
~
~
~
~
~
~
~
~
```

Dockerfile

Connecting services



Docker Compose

```
version: '2'
services:
  smartfeed_service:
    build:
      context: ./smartfeed
      dockerfile: Dockerfile
    ports:
      - "5002:9191"
    command: service
    container_name: smartfeed_service
    volumes:
      - "/home/{{USER}}/code/optimus/smartfeed:/var/src/smartfeed"
      - "/var/serverset:/var/serverset"
      - "/var/config:/var/config"
    environment:
      - LOG4J_CONFIG_FILE_SERVICE=config/log4j_smartfeed_service.xml
      - CONFIG_FILE_SERVICE=config/config.smartfeed_service.test.properties
      - HEAP_SIZE=4G
      - NEW_SIZE=2G

  pinacle2:
    build:
      context: ./pinacle2
      dockerfile: Dockerfile-dev
    ports:
      - "5001:8887"
    command: "bash scripts/local_test_server.sh"
    container_name: pinacle2
    volumes:
      - "/home/{{USER}}/code/optimus/pinacle2:/var/src/pinacle2"
      - "/var/serverset:/var/serverset"
      - "/var/config:/var/config"
    environment:
      - LOG4J_CONFIG_FILE=config/log4j.dev.properties
      - CONFIG_FILE=config/config_test.yaml
      - HEAP_SIZE=4G
      - NEW_SIZE=2G
      -  =/opt/pinacle2/pin e2-0.1-SNAPSHOT

  fantas :
    bui
      co ./fa
      dockerfile: Dockerfile-dev
    ports:
      - "5004:9020"
```

Docker-compose file

Demo

Anything that can go wrong will go wrong , Edward Murphy

Docker commands

1. Inspect
2. Attach
3. Exec
4. Pause, Stop, Remove
5. Top

Software Engineering Daily Podcast

Docker for micro services, Apress

Q&A

kinnary@pinterest.com