DOCKER + SPARK

A glance at the data world, throughout a container

Daniel Restrepo Hincapié

Big Data Engineer Senior @ SoftServe.

softserve

Luis Fernando Vásquez
Data Architect Senior - Software Designer

softserve

Workshop Contents

SOURCE & ARCHITECTURE

DATA SOURCE

Open GPX Tracker
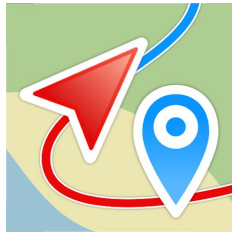
Open GPX Tracker supports multiple map tile servers

Apple Maps | OpenStreetMap | Mapquest | OpenCycleMap | CartoDB

```xml
<?xml version="1.0" encoding="UTF-8"?>
<gpx xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance" xmlns="http://www.topografix.com/GPX/1/1"
    xsi:schemaLocation="http://www.topografix.com/GPX/1/1 http://www.topografix.com/GPX/1/1/gpx.xsd"
    version="1.1"
    creator="Open GPX Tracker for iOS">
  <trk>
    <trkseg>
      <trkpt lat="6.297475984325909" lon="-75.5781921186257">
        <ele>1668.879306793213</ele>
        <time>2022-03-01T20:33:48Z</time>
      </trkpt>
      <trkpt lat="6.297476068144941" lon="-75.57814610197728">
        <ele>1668.386142730713</ele>
        <time>2022-03-01T20:33:
      </trkpt>
    </trkseg>
  </trk>
</gpx>
```
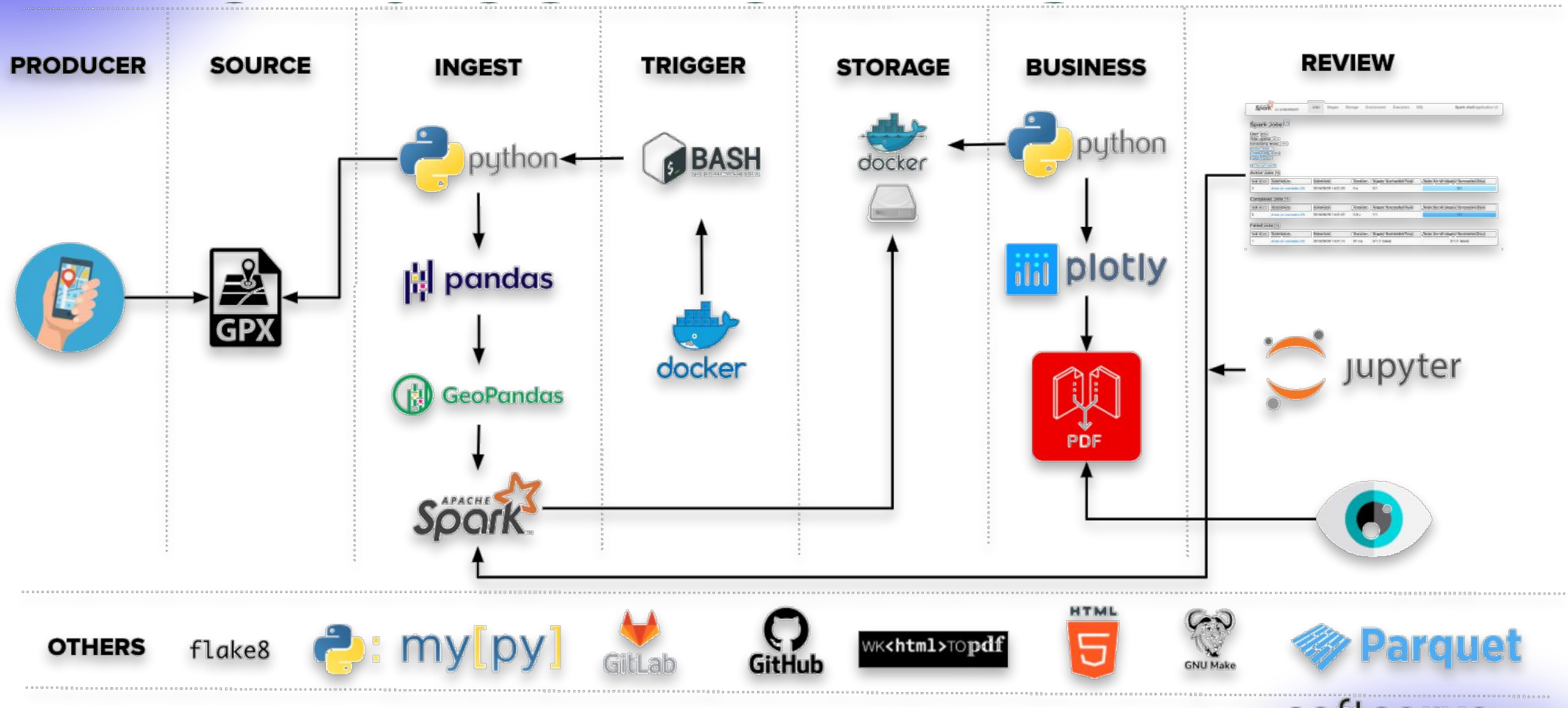
gpx ..
- @xmlns:xsi: http://www.w3.org/2001/XMLSchema-instance
- @xmlns: http://www.topografix.com/GPX/1/1
- @xsi:schemaLocation: http://www.topografix.com/GPX/1/1 http://www.topografix.com/GPX/1/1/gpx.xsd
- @version: 1.1
- @creator: Open GPX Tracker for iOS
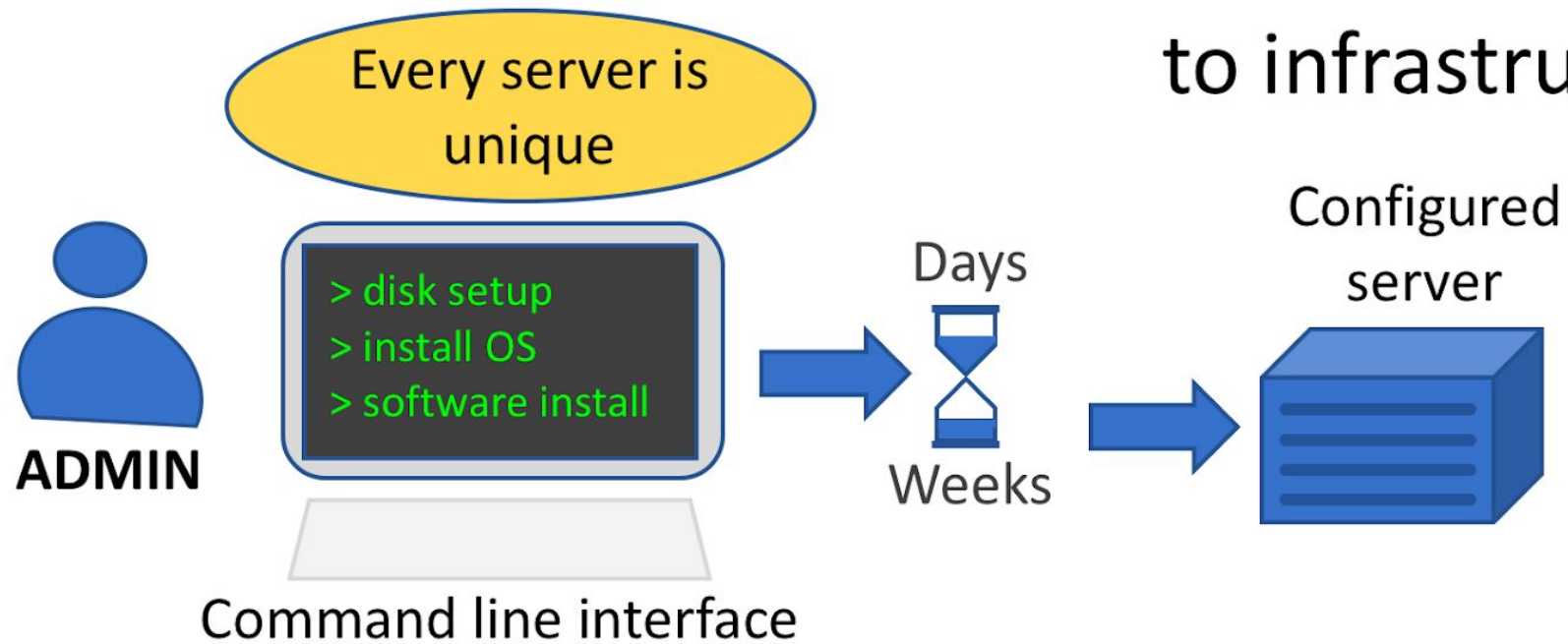
trk ..
  trkseg ..
    trkpt ..
    - @lat: 6.297475984325909
    - @lon: -75.5781921186257
    - ele 1668.879306793213
    - time 2022-03-01T20:33:48Z

    trkpt
    - @lat: 6.297476068144941
    - @lon: -75.57814610197728
    - ele 1668.386142730713
    - time 2022-03-01T20:33:49Z

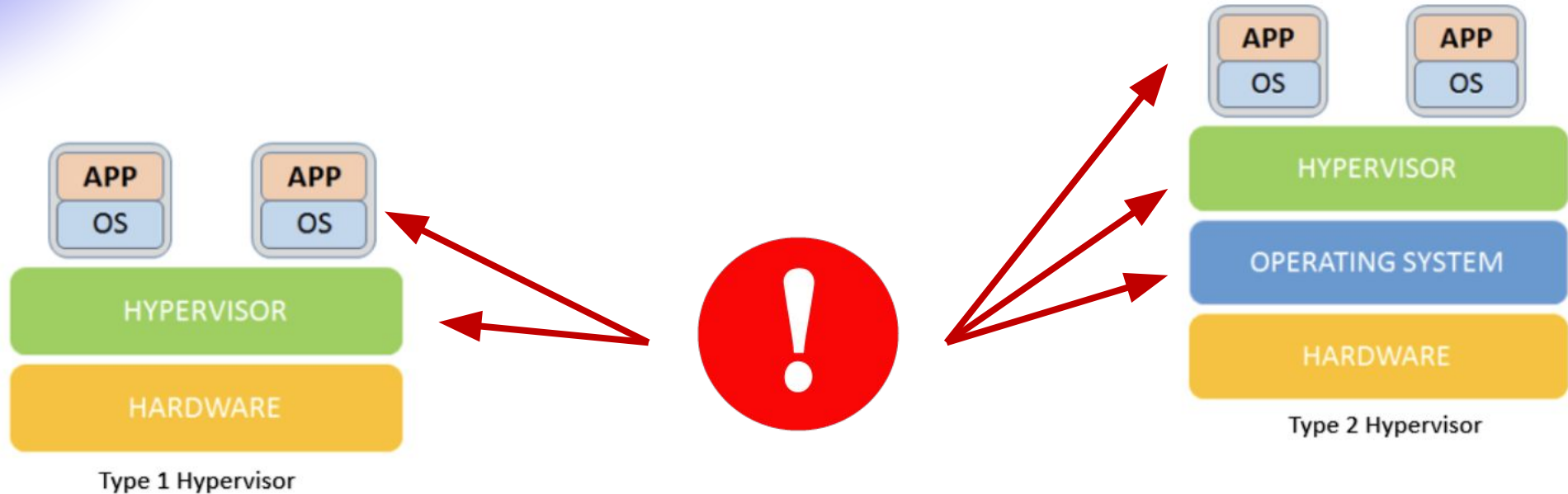# FLOW & COMPONENTS

CORE CONCEPTS

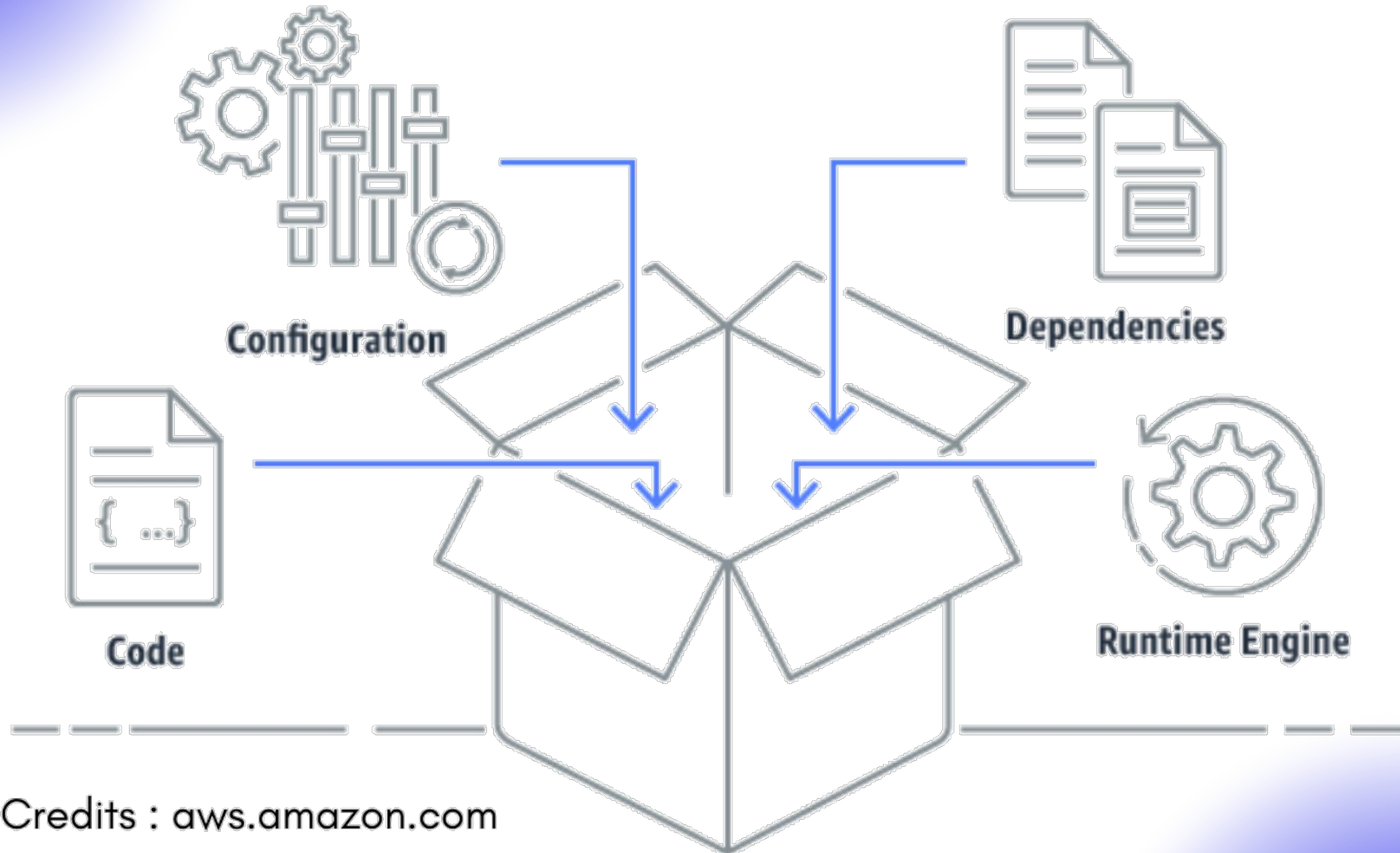# UNDERSTANDING THE TRADITIONAL IT INFRASTRUCTURE

Traditional approach to infrastructure

Every server is unique
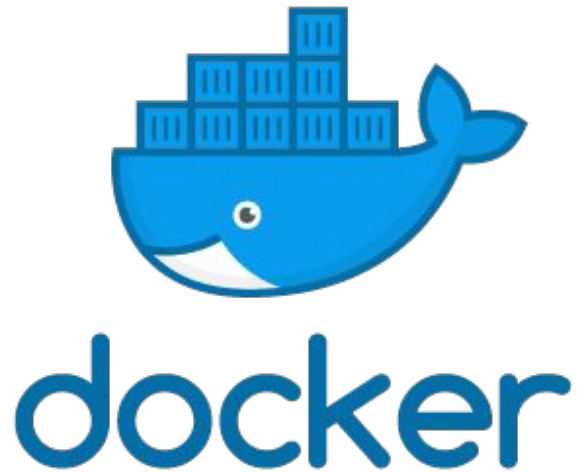
> disk setup
> install OS
> software install

ADMIN

Command line interface

Days

Weeks

Configured server

UNDERSTANDING THE HYPERVISOR

UNDERSTANDING THE CONTAINER

Configuration

Dependencies

Code

Runtime Engine

WHAT IS DOCKER?

DOCKER

Virtual Machines

App 1	App 2	App 3
Bins/Lib	Bins/Lib	Bins/Lib
Guest OS	Guest OS	Guest OS

Hypervisor

Infrastructure

Containers

App 1	App 2	App 3
Bins/Lib	Bins/Lib	Bins/Lib

Container Engine

Operating System

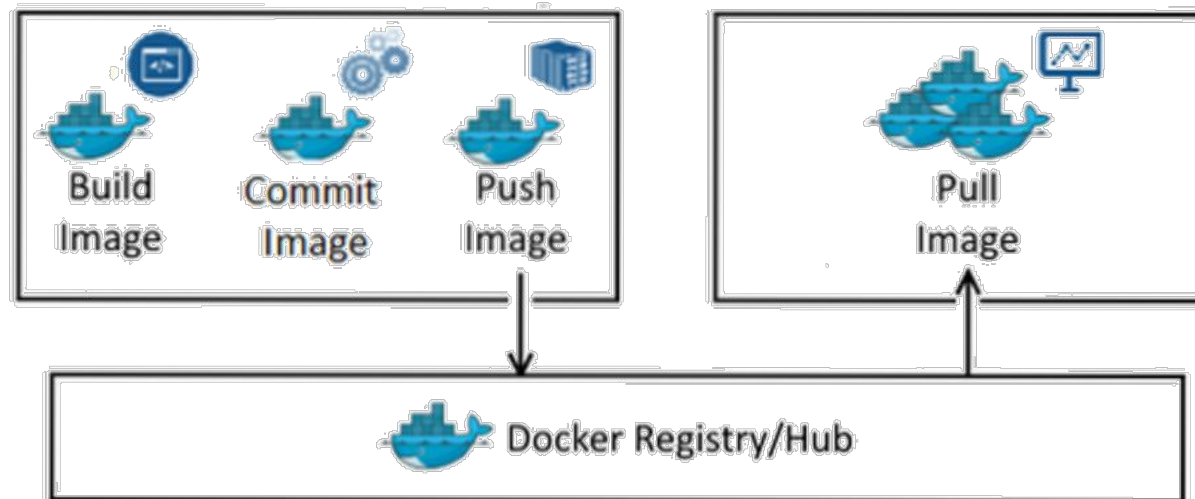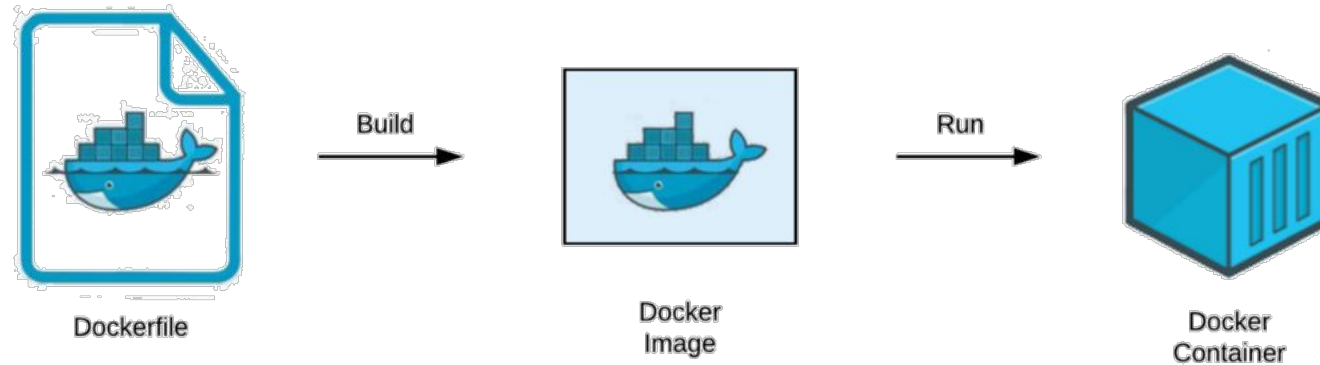**Infrastructure**
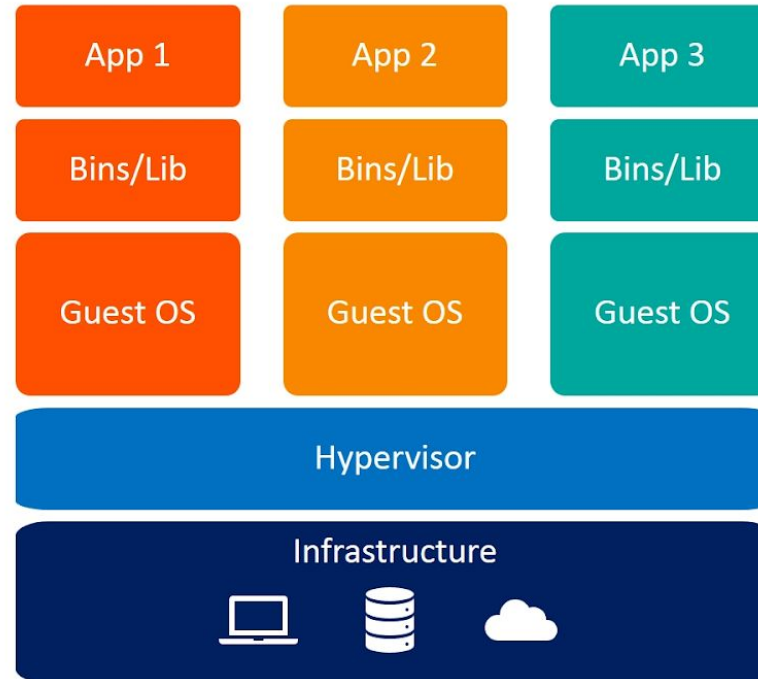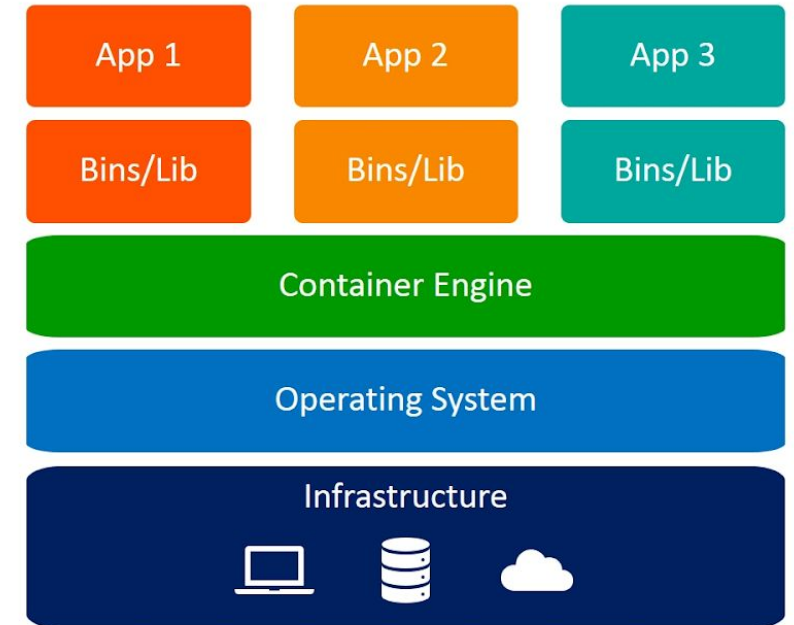
CORE FEATURES

DOCKER USE CASES IN THE WORLD

- ⚙ **SIMPLIFIED CONFIGURATION**
- **SERVER CONSOLIDATION**
- **PRODUCTIVITY**
- **PIPELINES**
- **AUTOMATION**
- **EASY CONFIGURATION**
- **APP ISOLATION**
- **CODE VALIDATION**
- **DEBUG CAPABILITIES**

13 million + developers

7 million + applications

13 billion + monthly image downloads

CREATING OUR DOCKER IMAGE

Walkthrough the Dockerfile

We're going to defining the context of our Docker container step by step.

Let's build the Image

After defining the context within the Dockerfile, let's buid the Docker Image.

Let's build the Container

Right after the Image has built, let's start the cointainer and play with it.

DEMO ETL

Detailed example of a data pipeline

We've buit a local represantation of how a data pipeline would work from end to end.

Spark Code + UI + Results Review

Once the data pipeline has finished, let's review its results and also let's dive into the Spark code and the History of operation performed thoroughout the UI.

Break Section

01:00pm ~ 01:30pm

QUESTIONS ?

THANKS !