

The background of the slide is a close-up photograph of numerous water droplets of various sizes on a dark, reflective surface. The droplets are in sharp focus, showing highlights and reflections, creating a textured, organic pattern. The overall color palette is dark with some lighter, iridescent tones from the light reflecting off the water.

Distributed scheduler hell

A scenic landscape featuring a paved road that stretches from the foreground into the distance. The road is flanked by green fields and some trees. In the background, there are large, rugged mountains with significant snow cover under a cloudy sky. The overall tone is somewhat muted, with a blueish-grey overlay.

Story of ..

**How we moved 100(s) of virtual machines, onto
containers**

Take aways

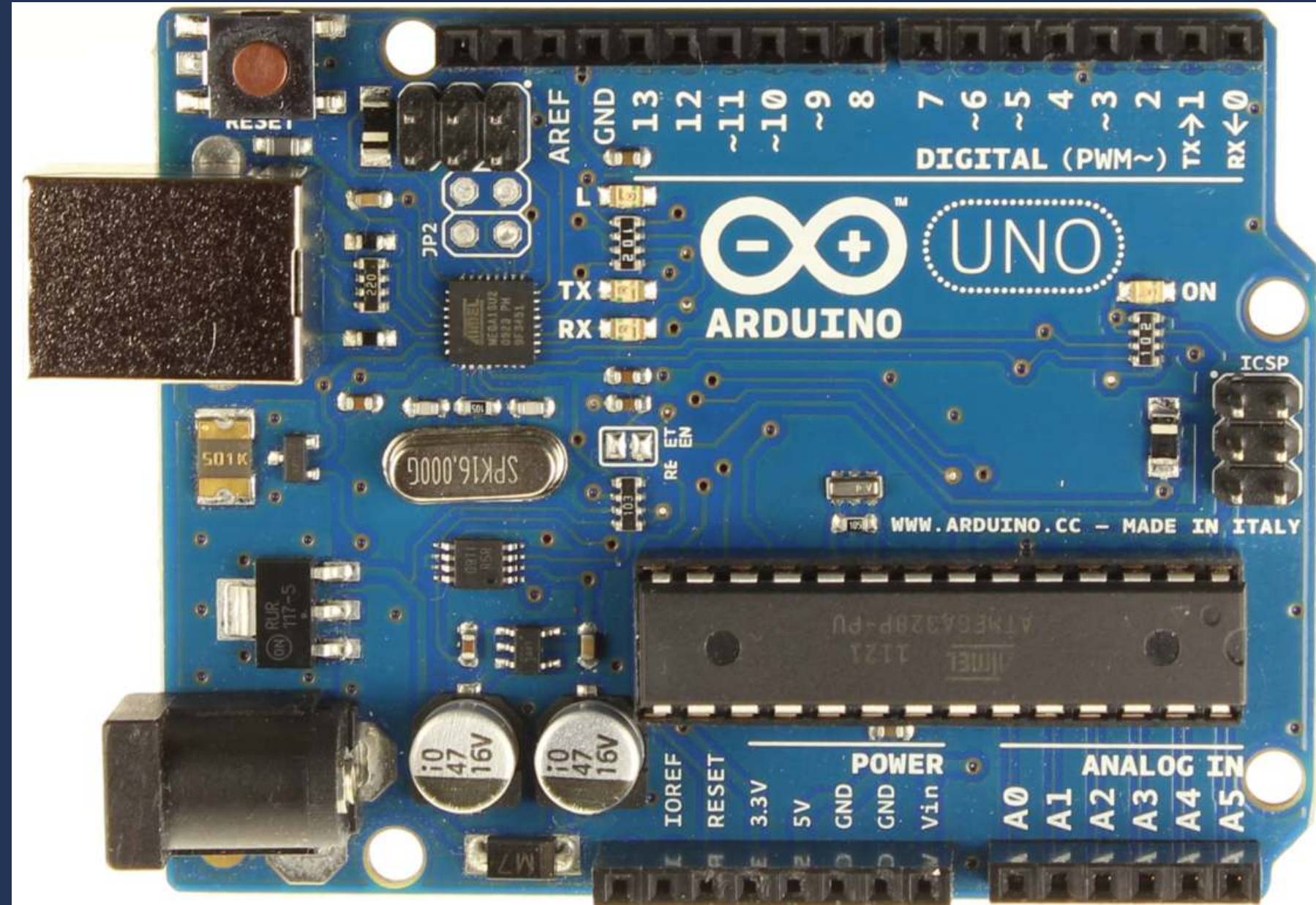
- What do Distributed schedulers do?
- Why?
- Best Practices
- What not to use it for



What is a distributed scheduler

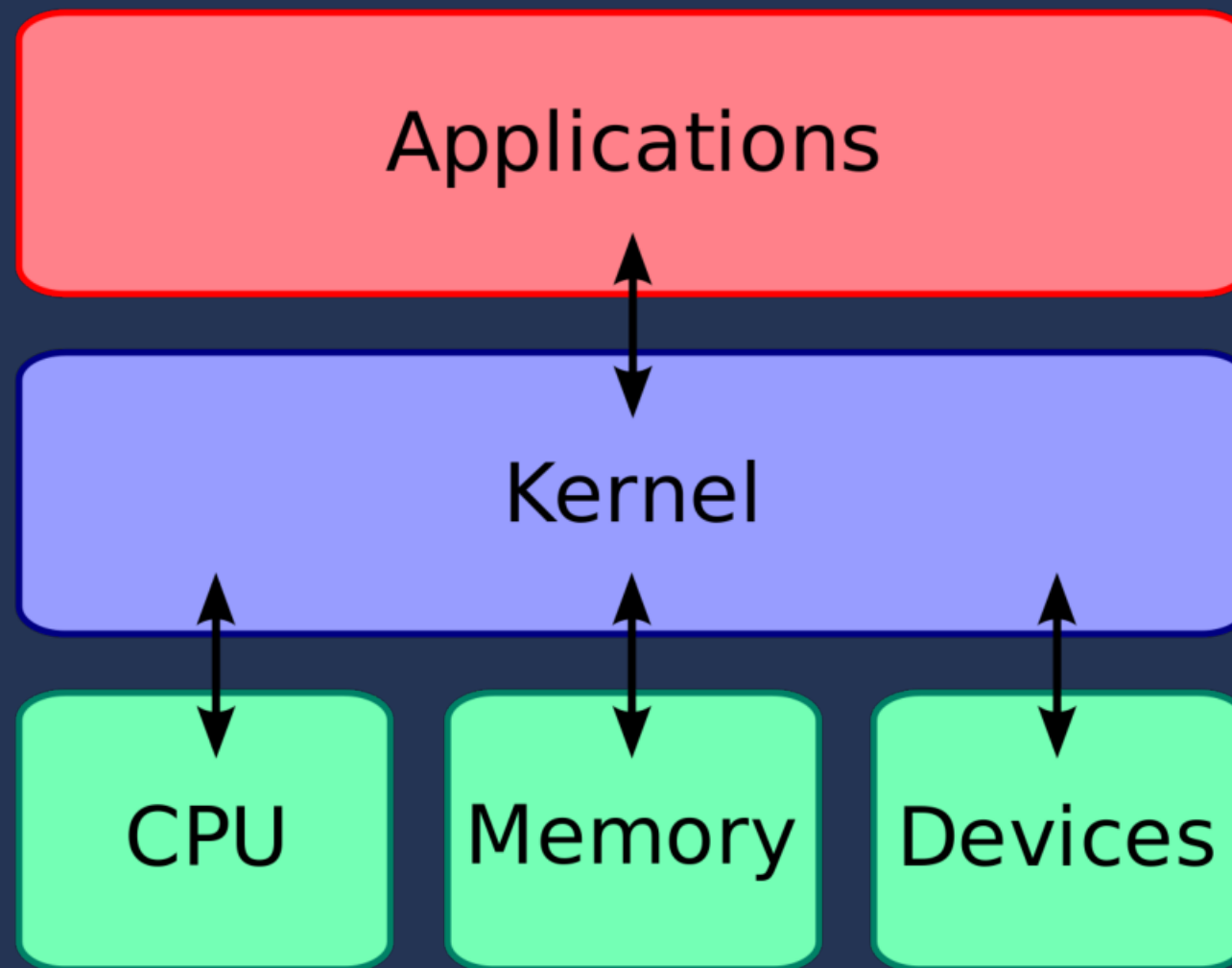
**Your cloud provider: Digital Ocean, AWS, Google.
For containers: Kubernetes, Nomad, Mesos**

Single process



Hypervisor

Linux Scheduler



Kernel Provides

- Virtual Memory
- Process Isolation
- Disk storage
- Networking
- CPU scheduling

Distributed Scheduler Provides

- Container Deploy
- Virtual Machine Deploy
- Memory Quota
- Disk storage
- Networking
- CPU scheduling
- Scaling Instances

FAIL

A man in a pink t-shirt and dark pants is jumping joyfully on a sand dune under a clear blue sky. His arms are raised in a 'V' shape, and his legs are spread wide. A long shadow of him is cast on the sand below. The background is a vast, open desert landscape.

Why?

Microservices...
Deployment...
Lifecycle mangement

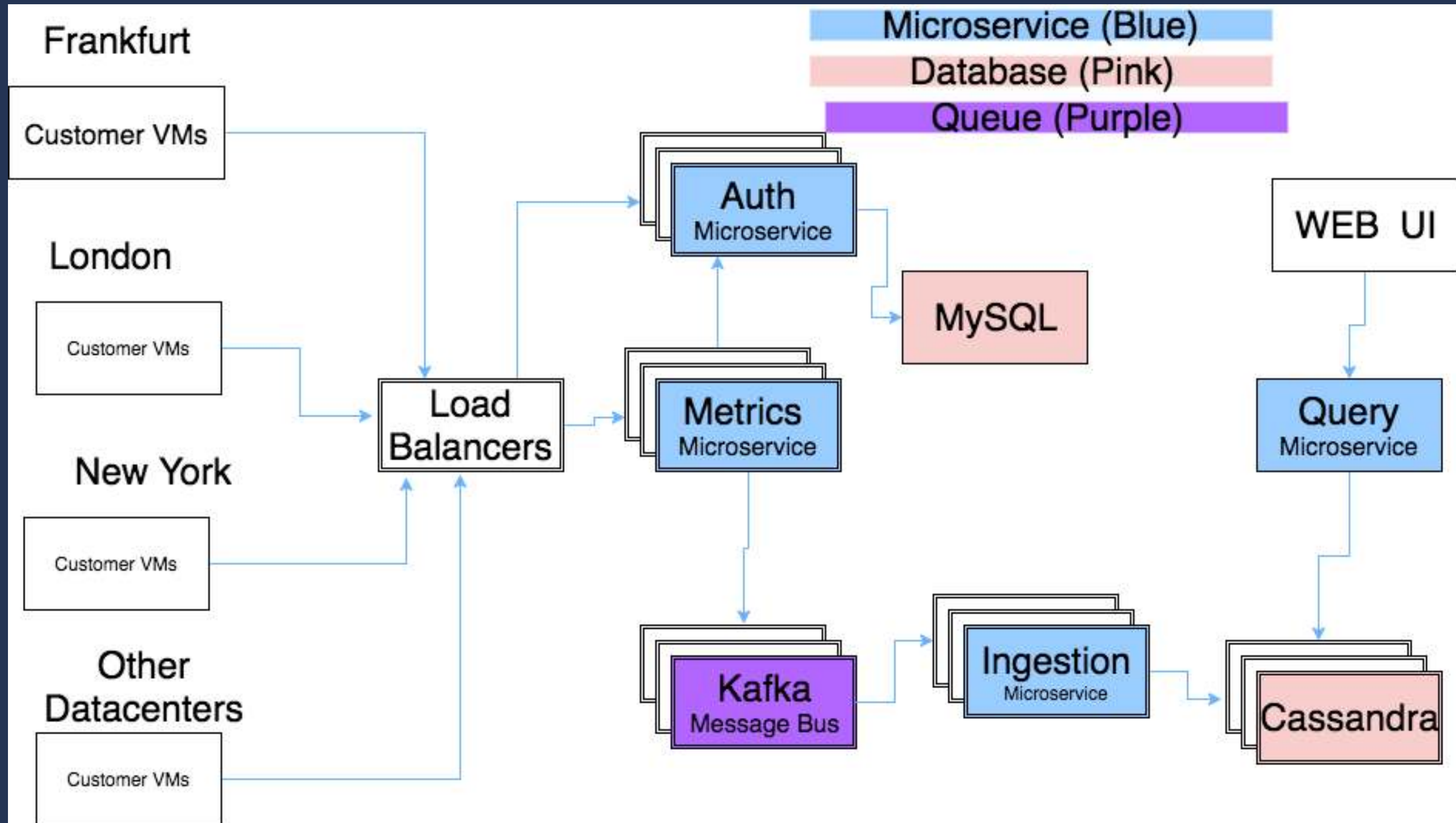
Vulcan : Distributed timeseries database



<https://github.com/digitalocean/vulcan>

Requirements

- 3 Gbits a second of Network traffic
- 20 TB Storage
- Sub 100ms Read times
- 100k writes a second





dcos-production
172.03.12.1

Dashboard

Services

Datacenter



Mesosphere DCOS v1.8



Dashboard

CPU Allocation

47%

65.1 of 114.2 Shares



Last 60 Seconds ▾

Memory Allocation

32%

24.6GB of 76.8GB



Last 60 Seconds ▾

Task Failure Rate

2%

Current Failure Rate



Last 60 Seconds ▾

Mesos

Service Health

marathon-production Healthy

chronos Healthy

cassandra-dev Sick

marathon-dev Healthy

hdfs-billing Healthy

[View all 64 Services >](#)

Tasks

402
Total Tasks

263
Running Tasks

139
Staged Tasks

Datacenter Health

127
Total Nodes

123
Connected

4
Disconnected

From Static Partitioning to Elastic Sharing

Static Partitioning



Elastic Sharing





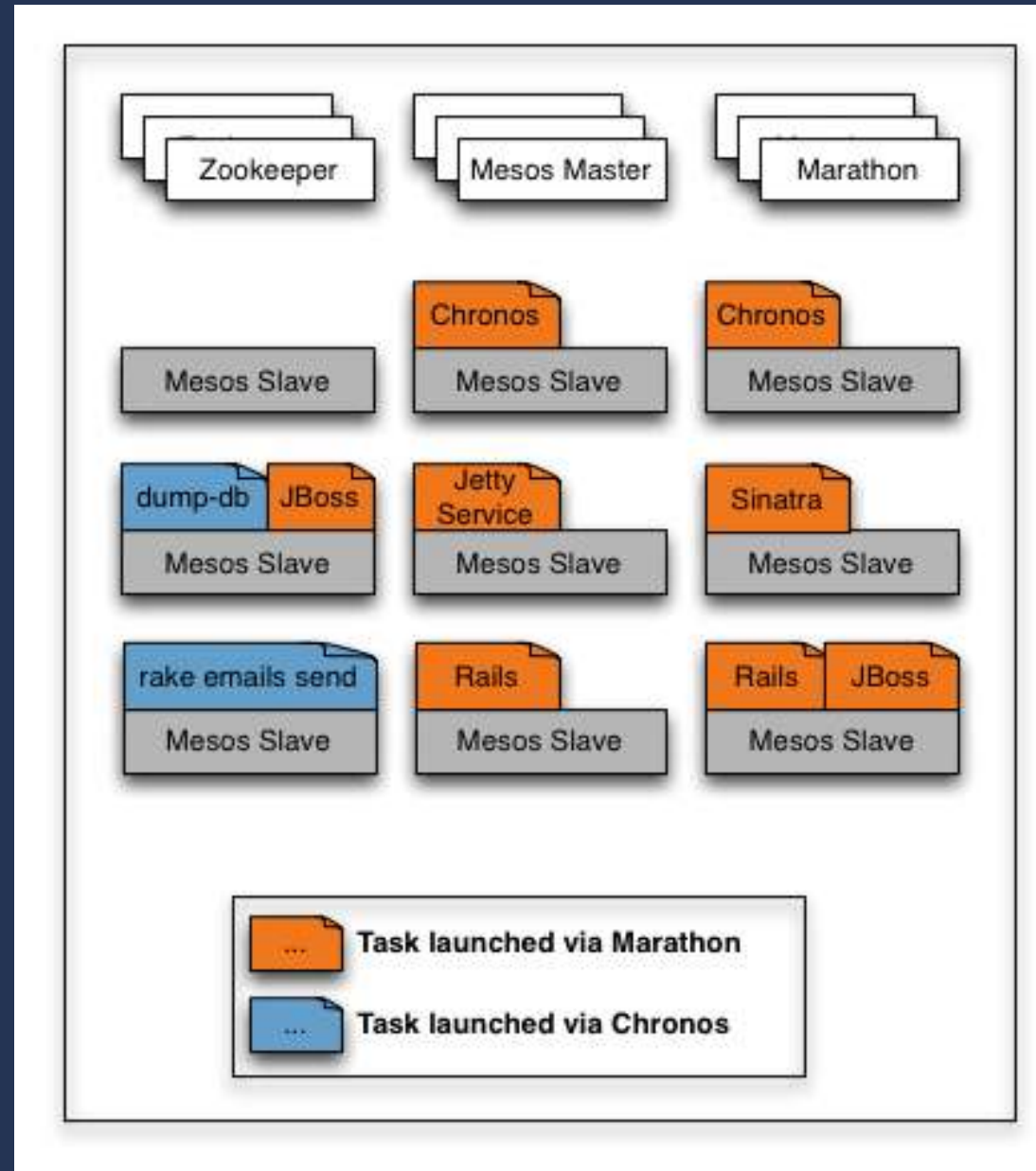


Deploying our App On Mesos

+ New App

ID ▲	Memory (MB)	CPUs	Tasks / Instances	Health	Status
/kibana	512	0.5	0 / 1		Deploying
/mesos-dns	512	0.2	3 / 3		Running
/search	2048	1	2 / 2		Running

Marathon Architecture





Kafka

Custom Scheduler

A background image showing the silhouettes of three people running on a beach towards the right. The sky is a mix of orange, yellow, and blue, suggesting a sunset or sunrise. The water is visible in the foreground, and the beach is a dark line at the bottom.

Cassandra

Pinning to specific Mesos nodes

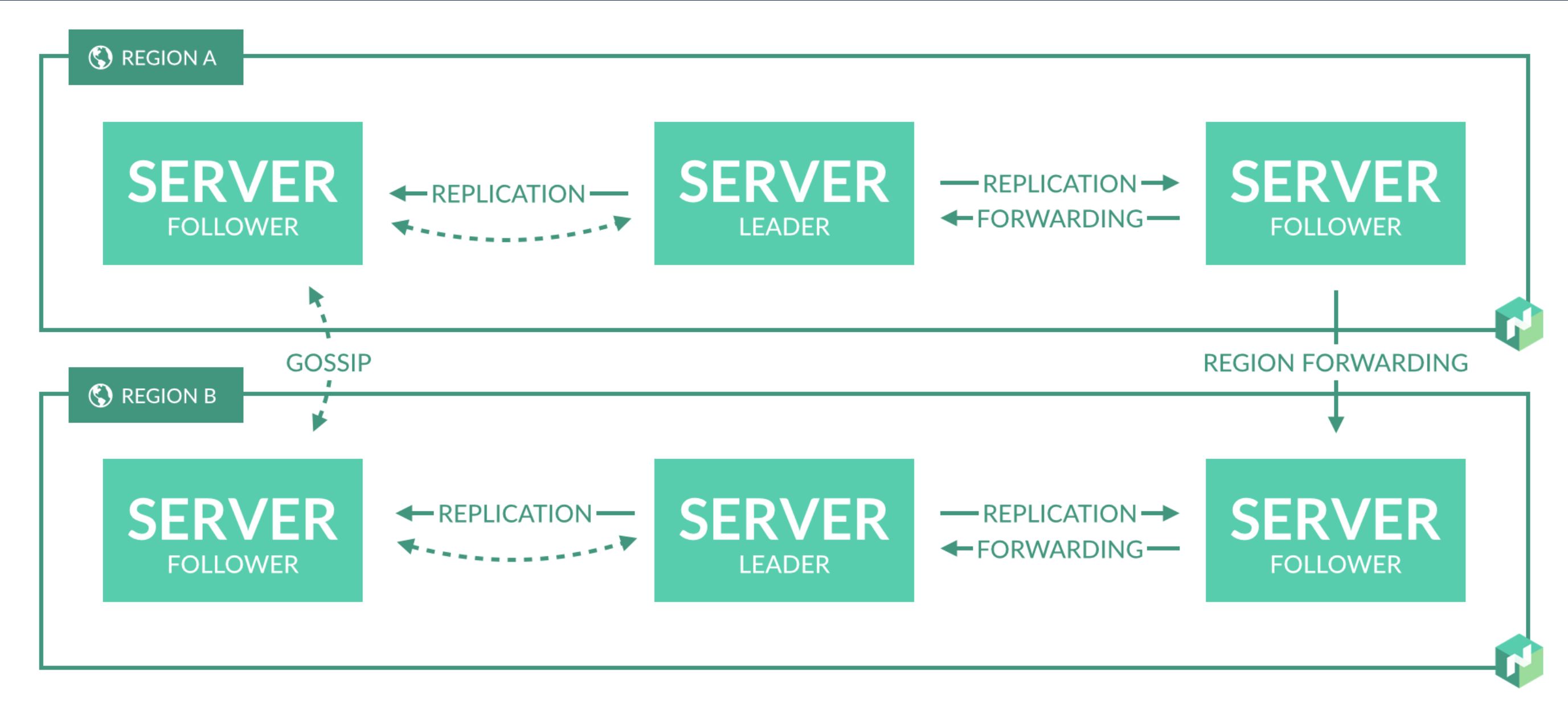


Mesos Failure Modes

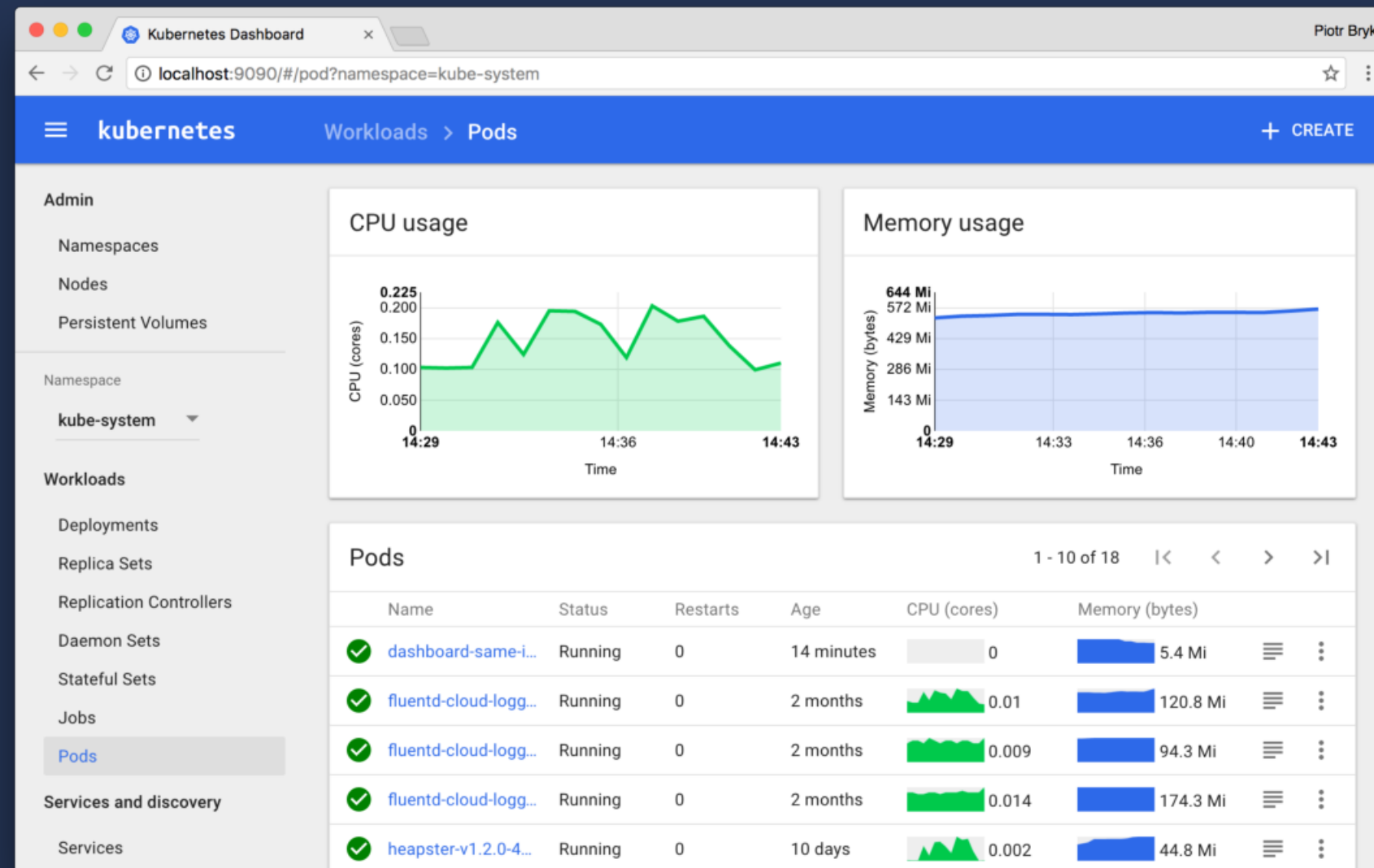
Network Partition FAIL

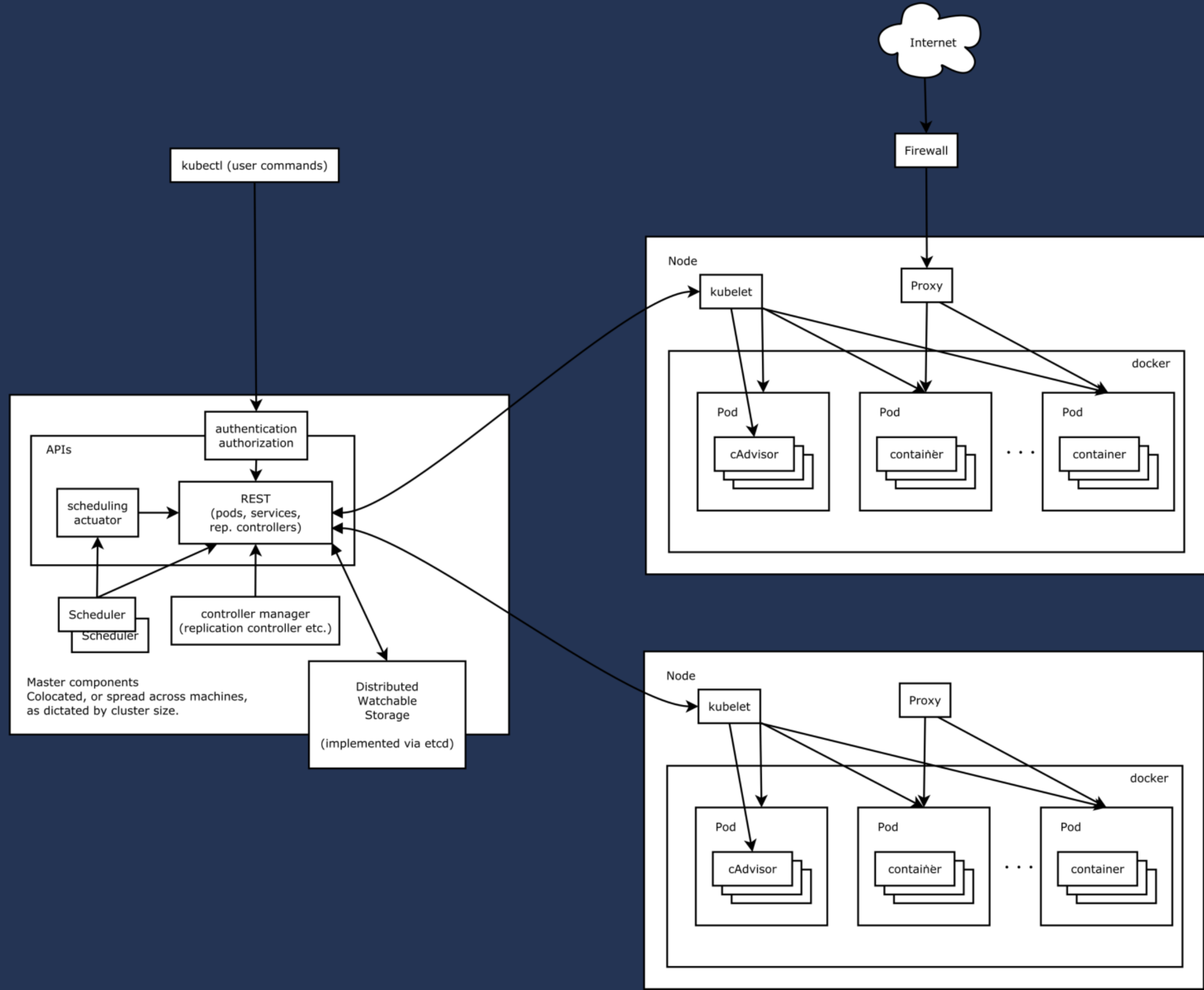
Counter Example: Nomad

Nomad Architecture



Kubernetes





Custom Json Language

```
{
  "application": {
    "name": "timeseries-ingestor",
    "scale": 1,
    "ingresses": {
      "timeseries-ingestor-health": {
        "scheme": "http",
        "container_port": 8001
      }
    }
  },
  "containers": {
    "timeseries-ingestor": {
      "image": "docker.com/timeseries/ingestor",
      "image_tag": "fcb24ca",
      "ports": [
        8001,
        9090
      ],
      "env": {
        "CASSANDRA_ADDRESS_LIST": ""
      },
      "resources": {
        "cpu": "4",
        "memory": "4000"
      }
    }
  },
  "metrics": {
    "port": 8001,
    "path": "/debug/metrics"
  },
  "maintainer": "dummy@digitalocean.com"
}
```

Command Line Deploys

```
docc --contexts dev_env deploy timeseries-microservice1.json
```

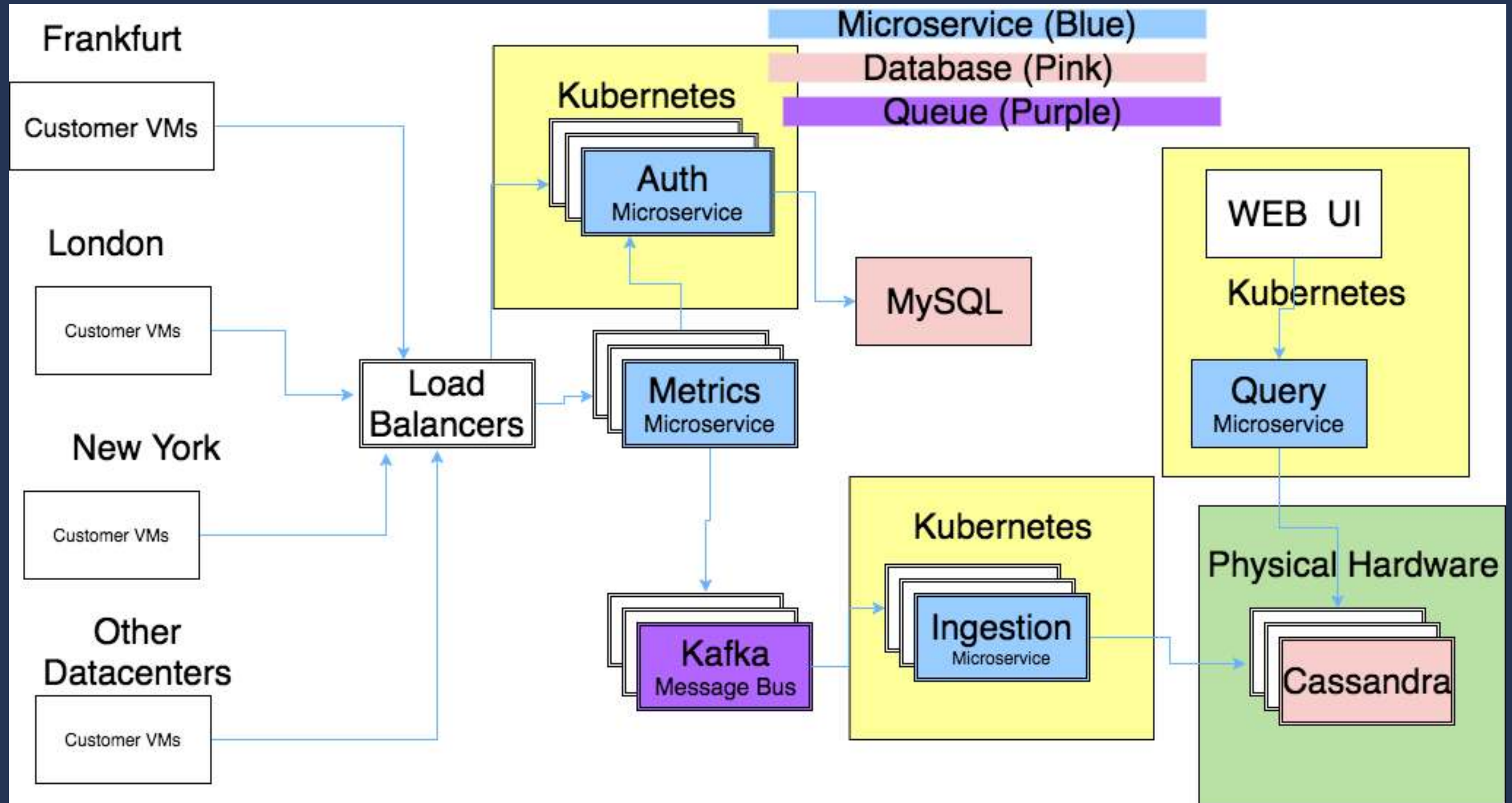
Docc is our internal Kubernetes tool

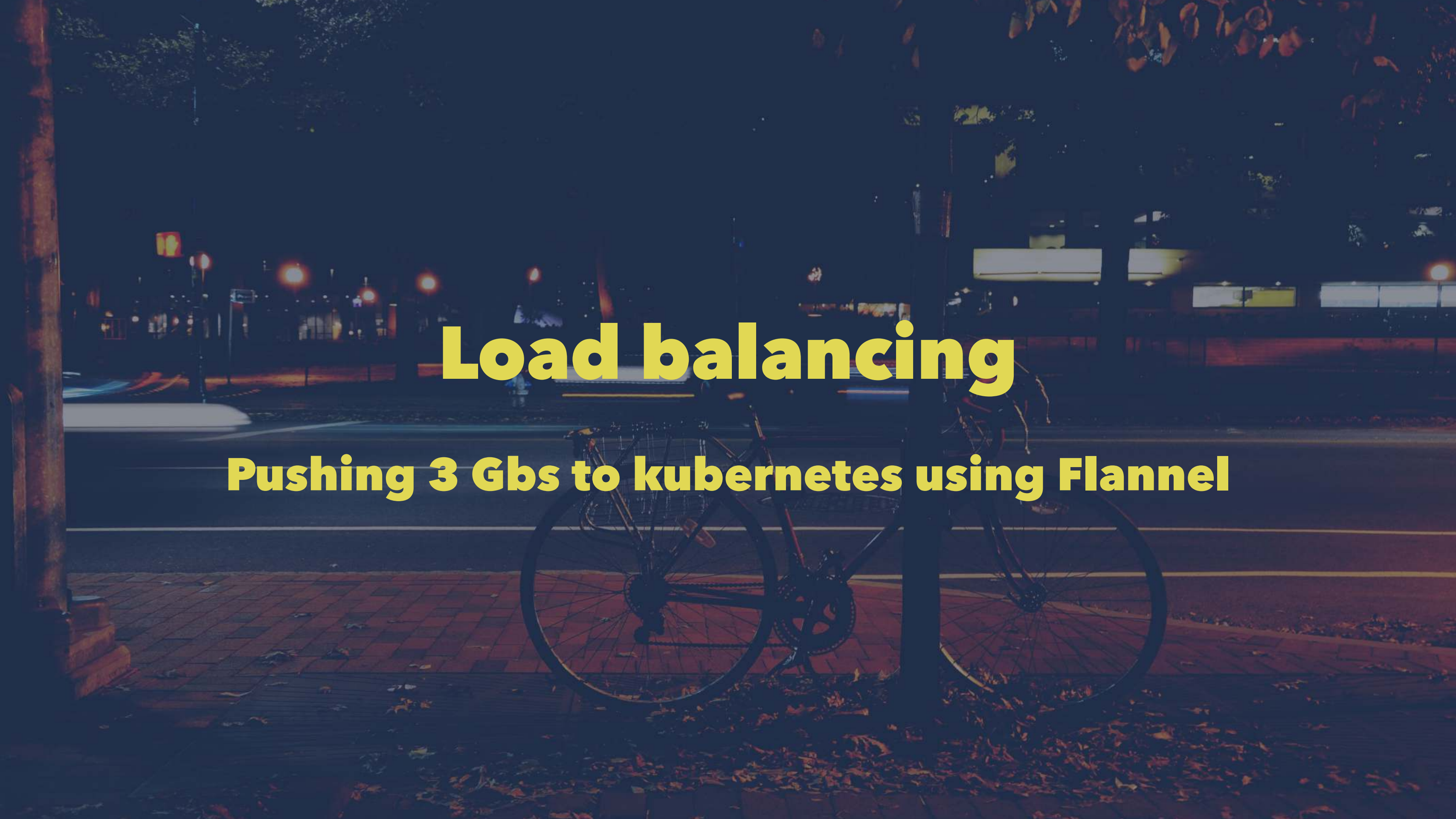
Deployment tool with diffs to Kubernetes

```
      "env": {  
+++      "CASSANDRA_CQL_VERSION": "3.1.7",  
---      "CASSANDRA_KEYSPACE": "staging_vulcan",  
+++      "CASSANDRA_KEYSPACE": "staging_vulcan_123",  
    }  
  }  
}
```

Simliar to KubeDiff

Final Architecture



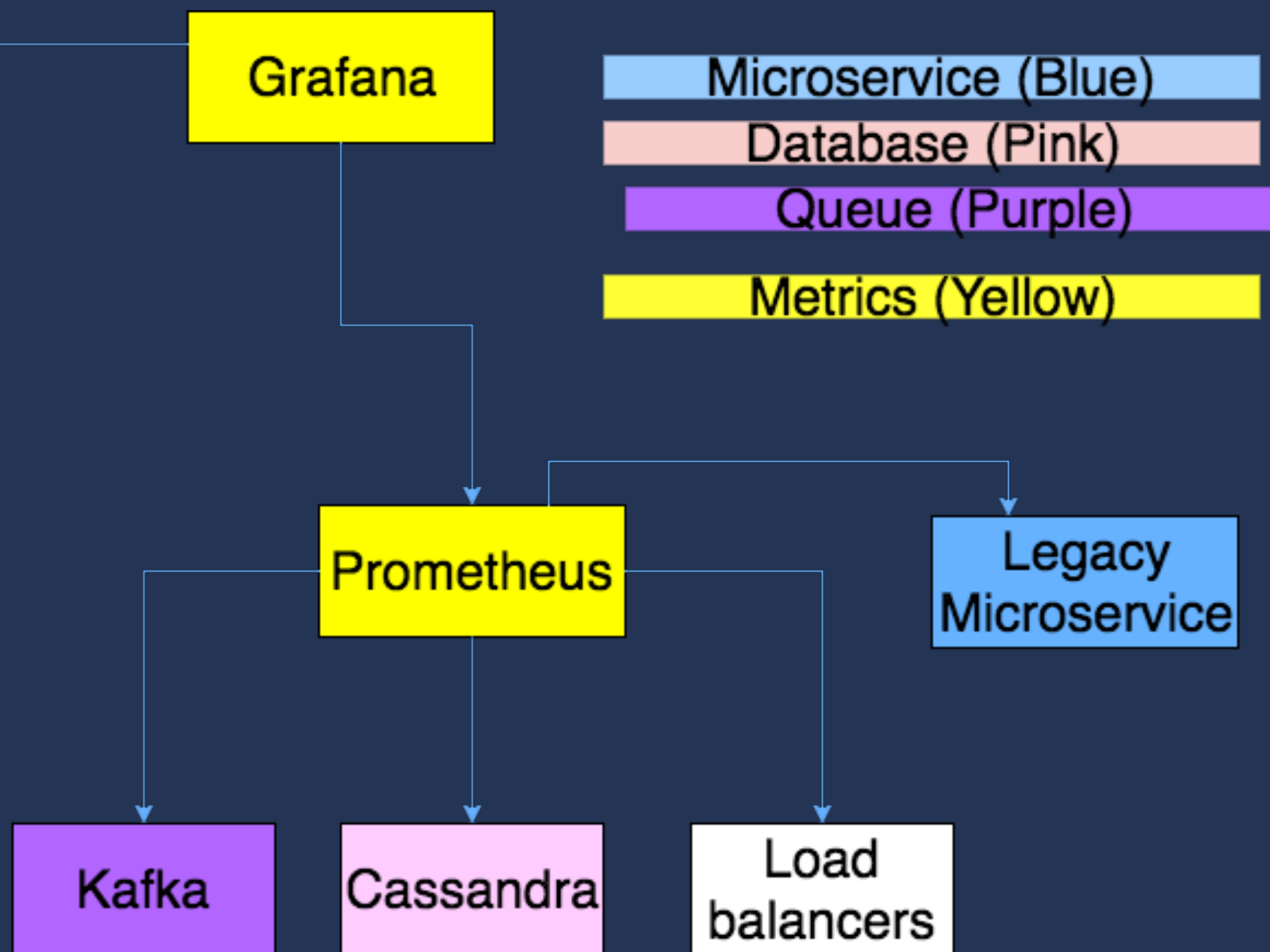
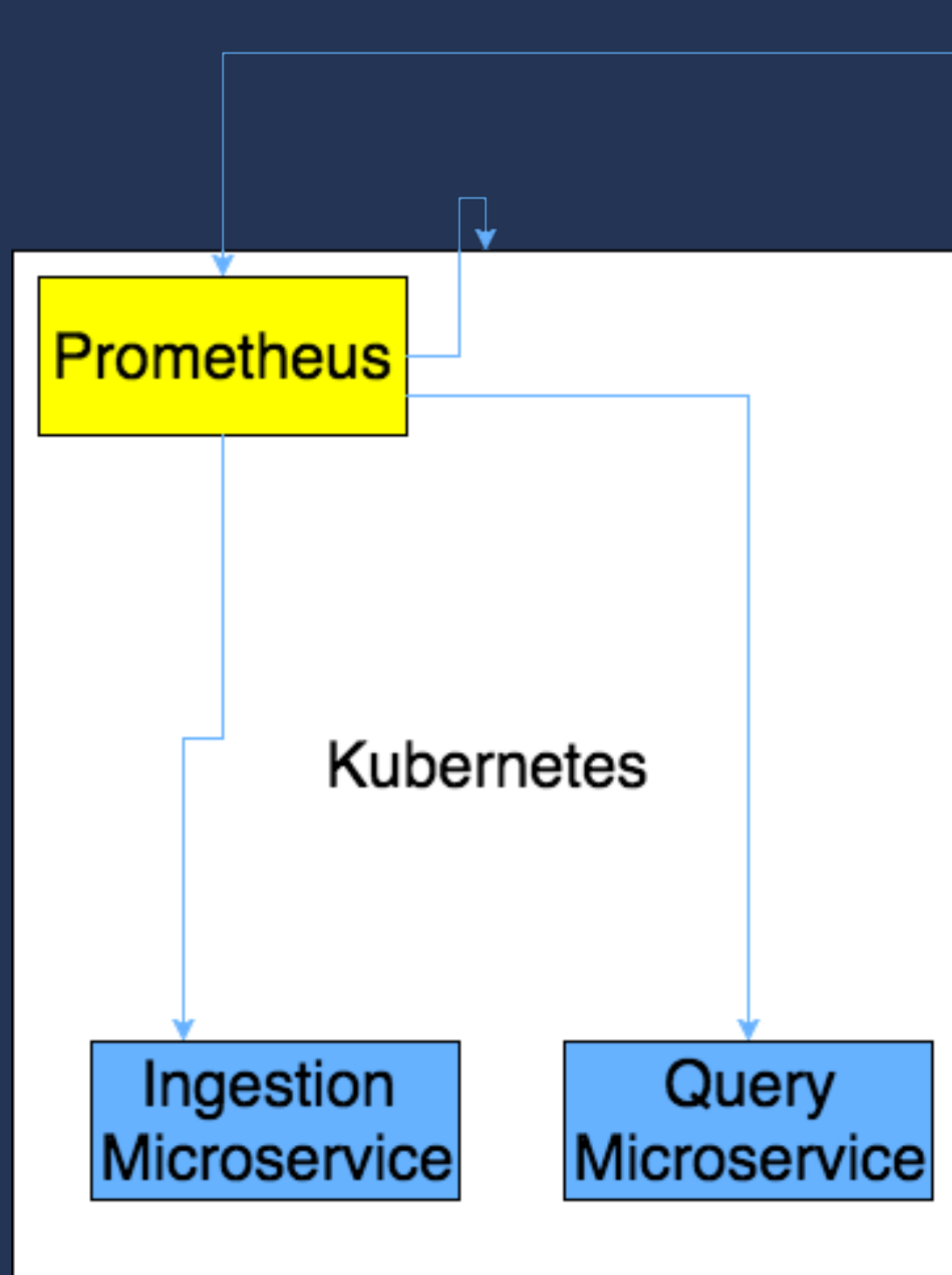
A photograph of a city street at night. In the foreground, a bicycle is leaning against a black metal pole on a brick sidewalk. The ground is covered with fallen leaves. In the background, there is a street with a few cars, streetlights, and buildings with lit windows.

Load balancing

Pushing 3 Gbs to kubernetes using Flannel

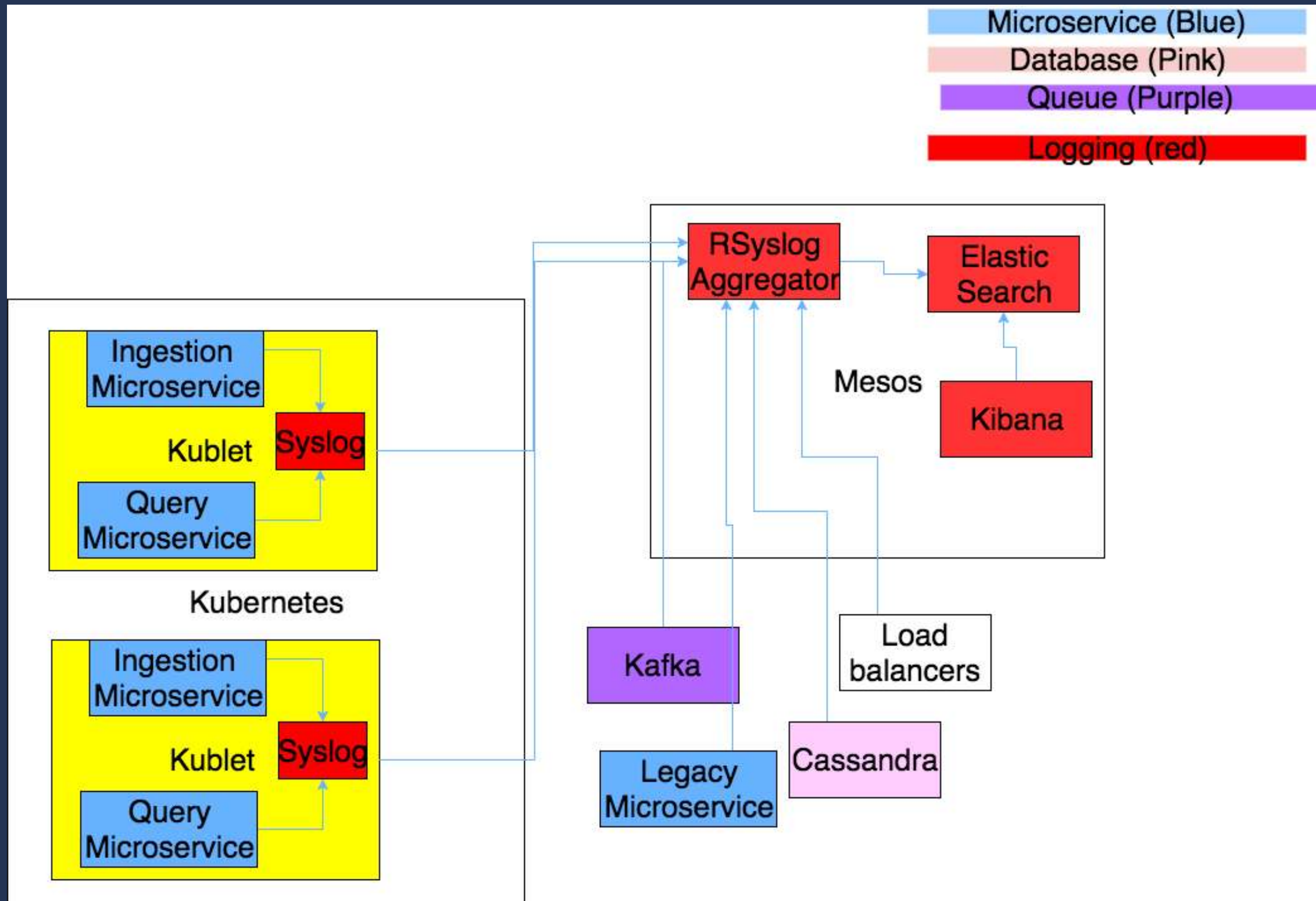


Metrics



The background of the image is a close-up of water droplets on a glass surface. The droplets are of various sizes and are scattered across the frame. The background has a color gradient, transitioning from a deep blue on the left to a warm orange on the right. The word "Logging" is centered in the middle of the image in a bold, yellow, sans-serif font.

Logging



A photograph of a sunset over the ocean, taken from the deck of a boat. The sun is low on the horizon, casting a warm glow across the sky and water. The boat's railing and some equipment are visible in the foreground. The text "Upsides to Distributed Schedulers" is overlaid in a bold, yellow font.

Upsides to Distributed Schedulers



How to choice your abstraction



Questions ?