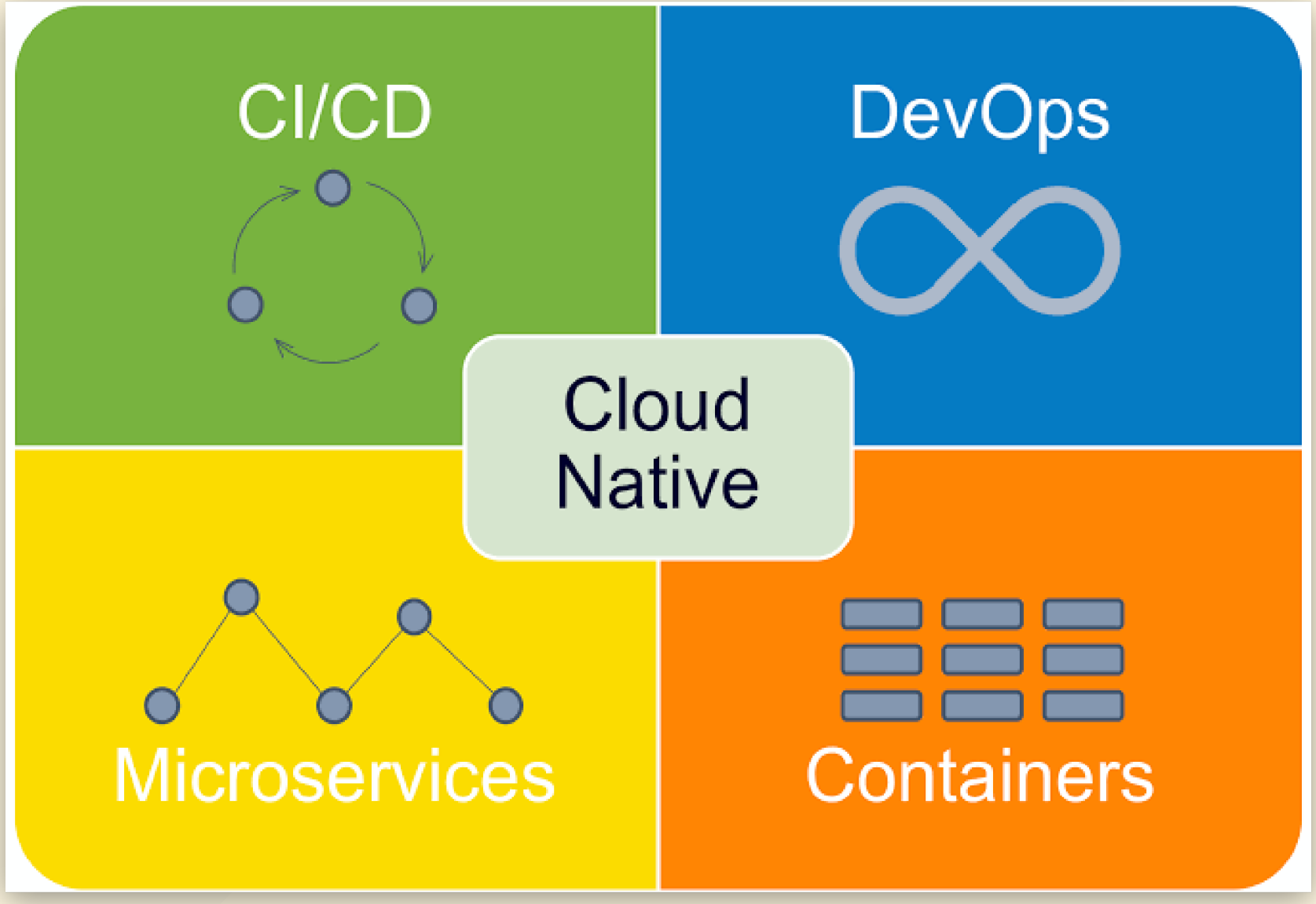
Spark on Kubernetes

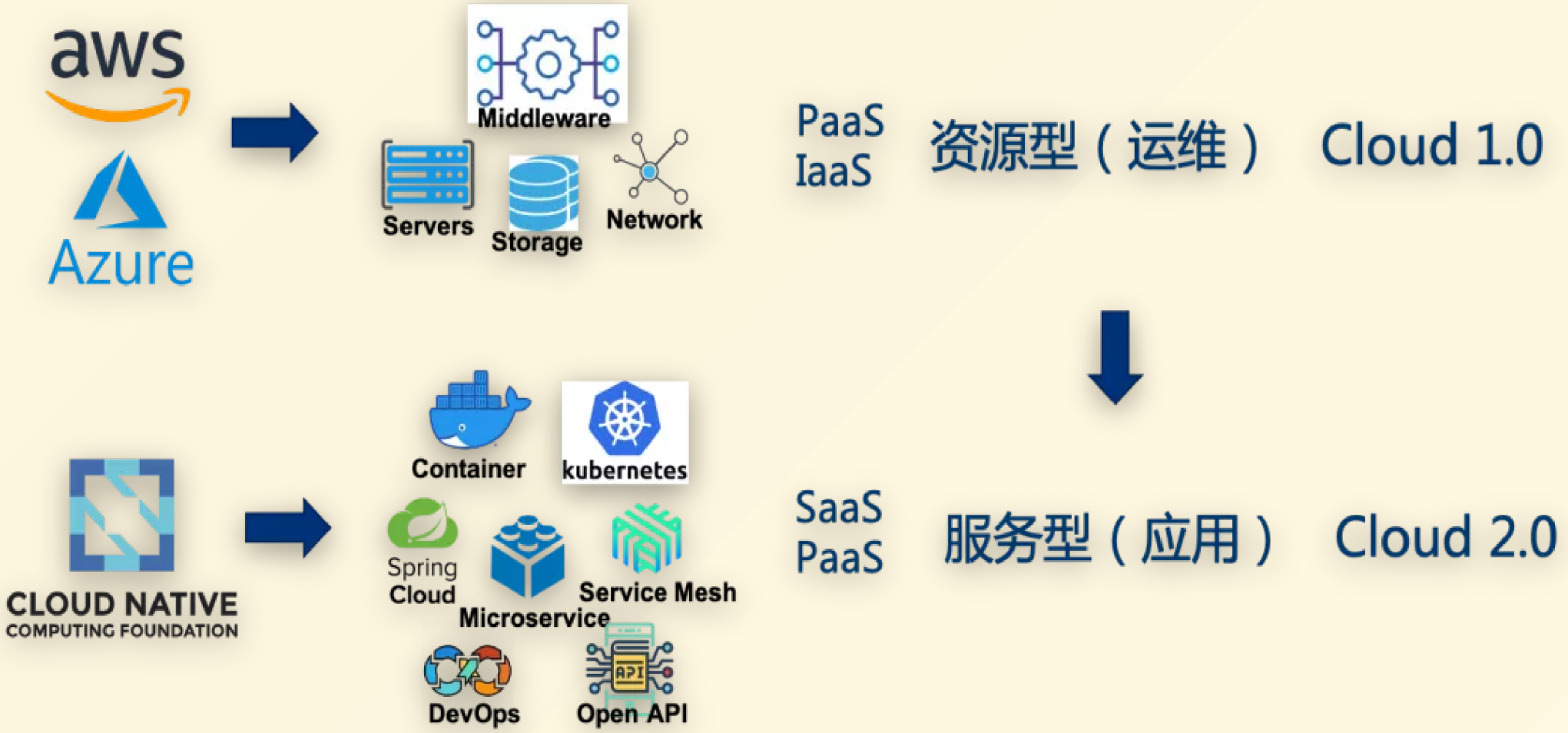
Contents

- Cloud Native Computing
- Docker and Kubernetes
- Spark on Kubernetes

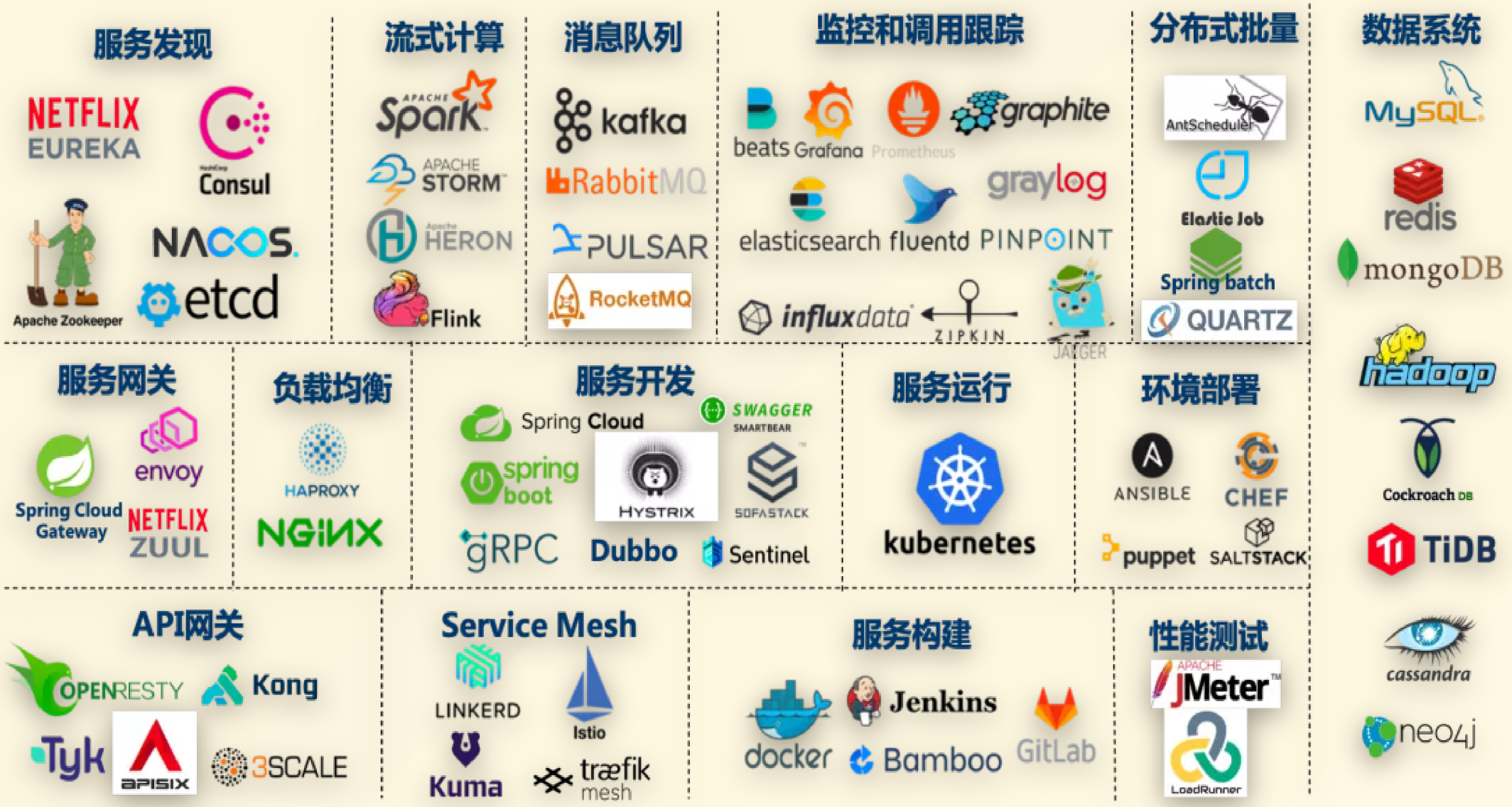
Cloud Native Computing

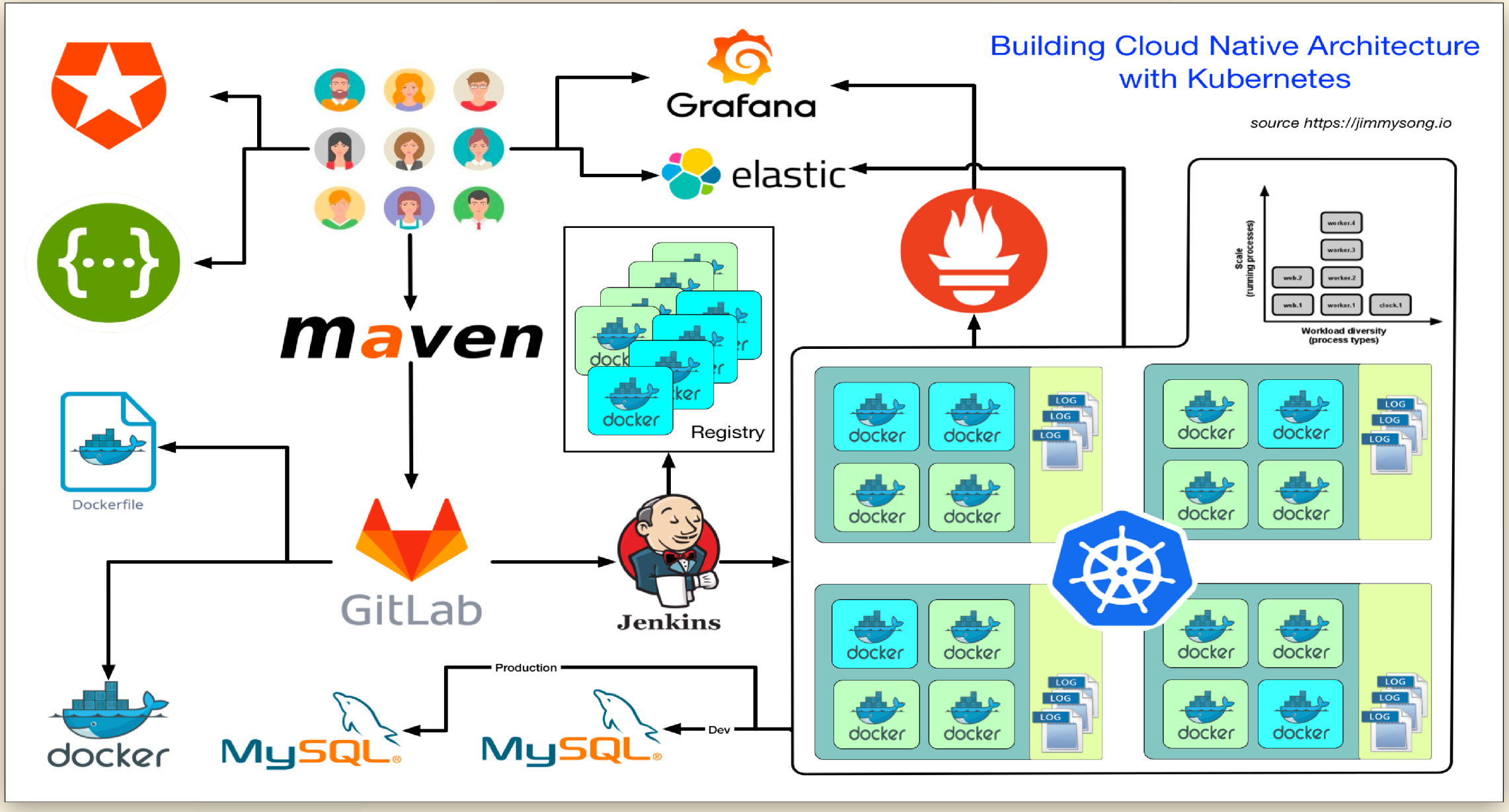
> As defined by the Cloud Native Computing Foundation (CNCF), Cloud native technologies empower organizations to build and run scalable applications in public, private, and hybrid clouds. Features such as **containers**, **service meshes**, **micro-services**, immutable infrastructure, and declarative application programming interfaces (APIs) best illustrate this approach.

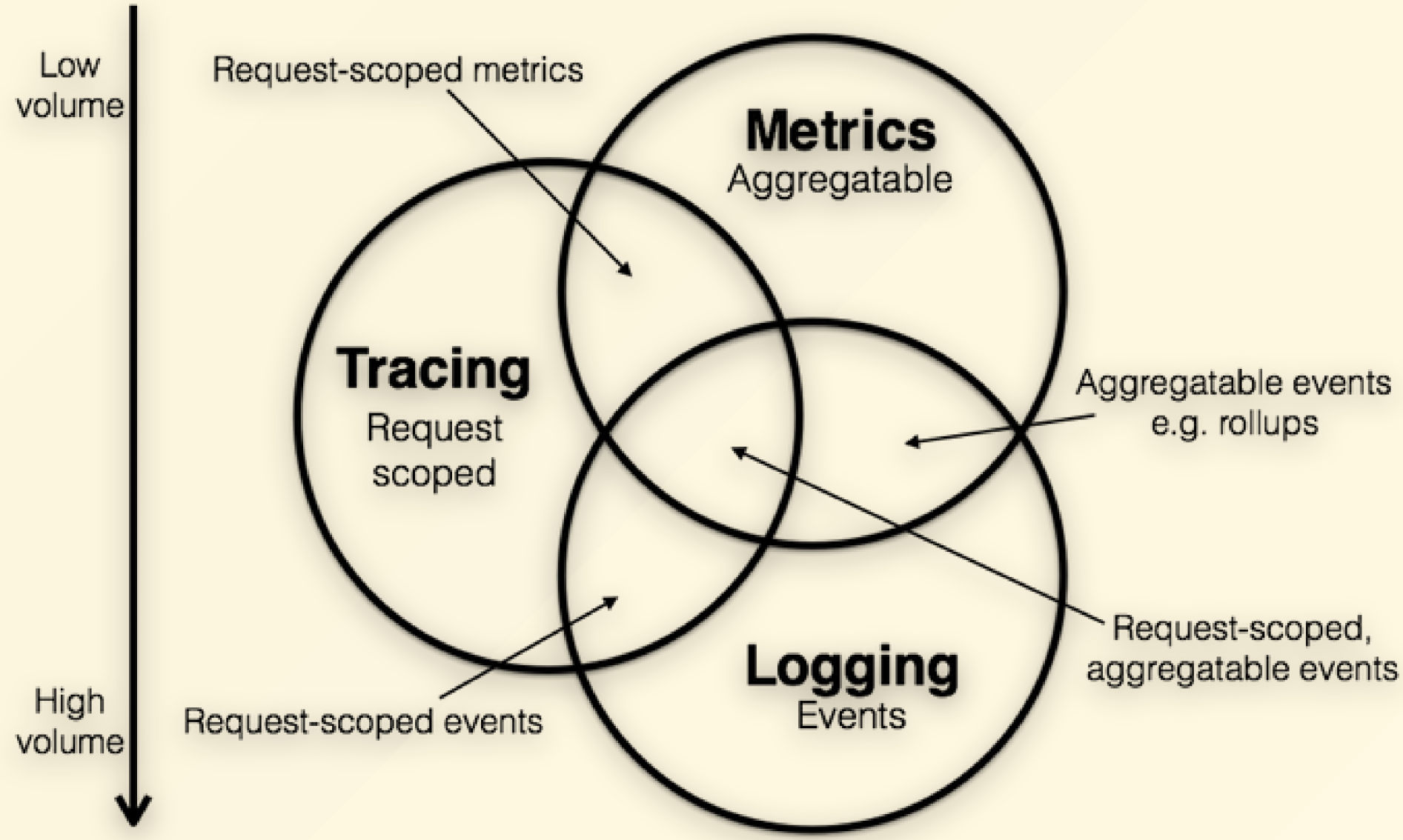

CI/CD
DevOps
Cloud
Native
Microservices
Containers

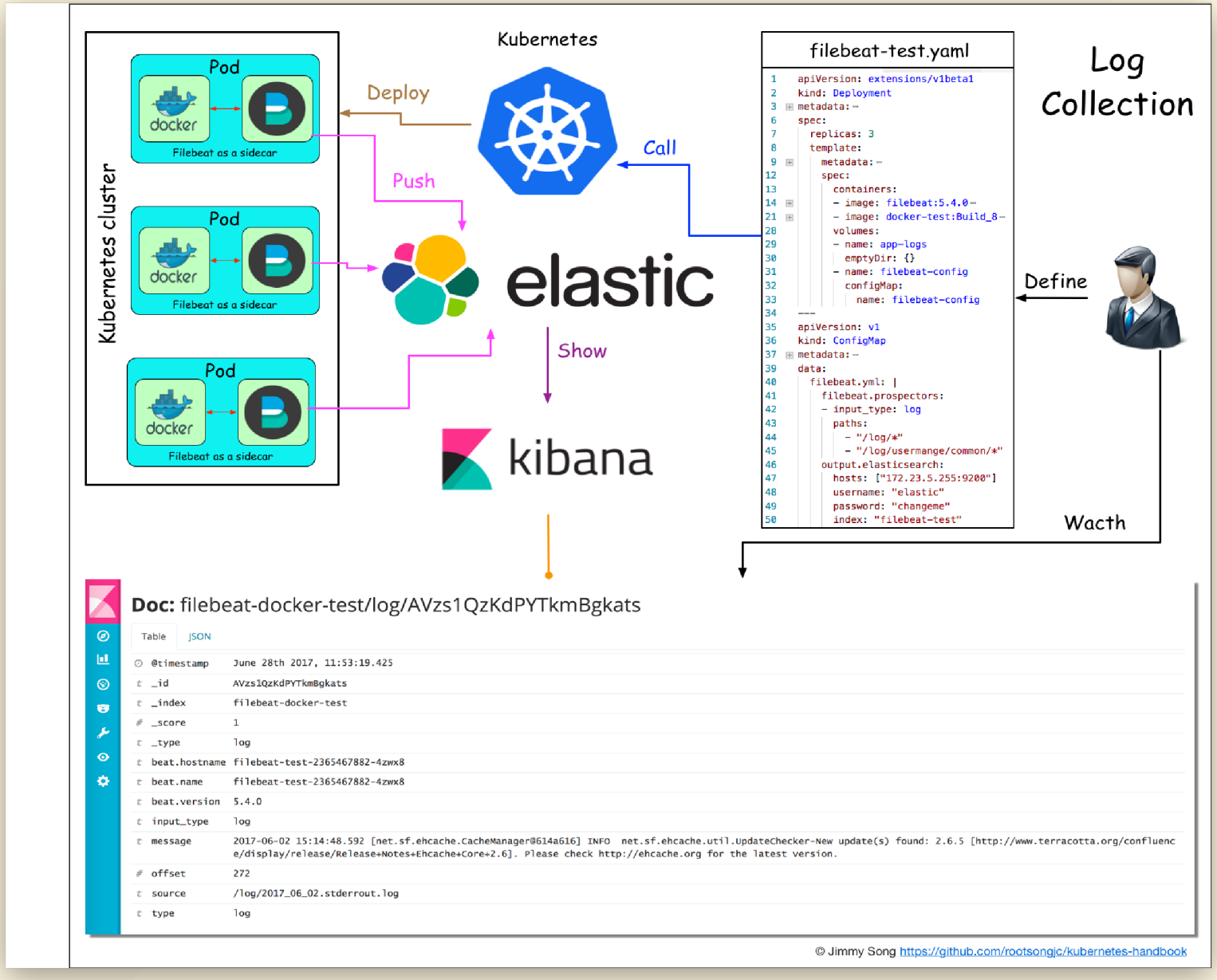


开源技术的蓬勃发展带来与公有云厂商的对标逐渐展开，已经进入cloud 2.0 时代数字经济时代

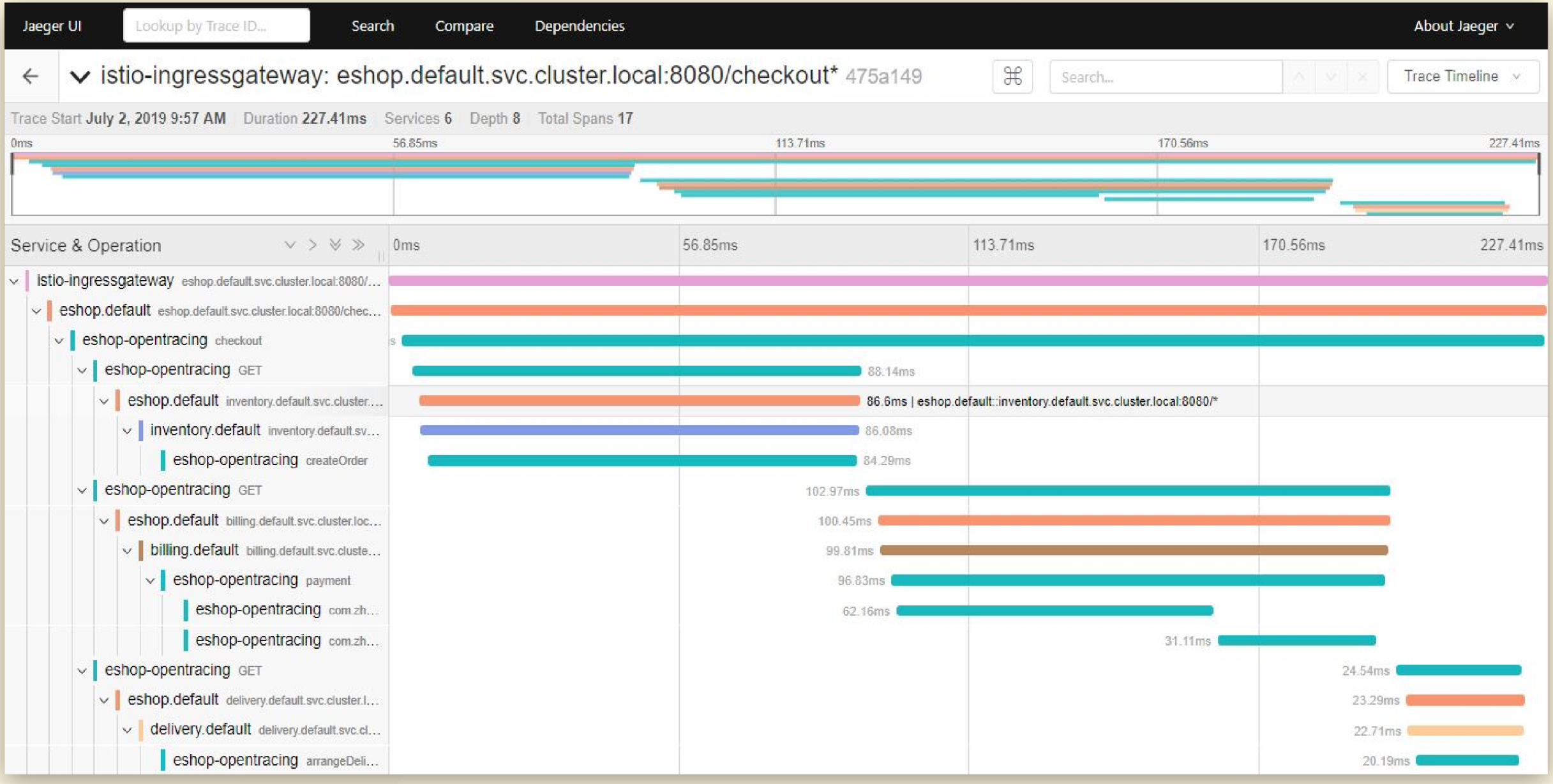

服务发现
NETFLIX EUREKA
Consul
Apache Zookeeper
NACOS.
etcd
流式计算
Spark
APACHE STORM
HERON
Flink
消息队列
kafka
RabbitMQ
PULSAR
RocketMQ
监控和调用跟踪
beats
Grafana
Prometheus
graphite
graylog
elasticsearch
fluentd
PINPOINT
influxdata
ZIPKIN
JAEGER
分布式批量
AntScheduler
Elastic Job
Spring batch
QUARTZ
数据系统
MySQL
redis
mongoDB
hadoop
Cockroach DB
TiDB
cassandra
neo4j
服务网关
envoy
Spring Cloud Gateway
NETFLIX ZUUL
负载均衡
HAPROXY
NGINX
服务开发
Spring Cloud
SWAGGER
SMARTBEAR
spring boot
HYSTRIX
SOFASTACK
gRPC
Dubbo
Sentinel
服务运行
kubernetes
环境部署
ANSIBLE
CHEF
puppet
SALTSTACK
API网关
OPENRESTY
Kong
Tyk
APISIX
3SCALE
Service Mesh
LINKERD
Istio
Kuma
træfik mesh
服务构建
docker
Jenkins
Bamboo
GitLab
性能测试
APACHE JMeter
LoadRunner


Building Cloud Native Architecture
with Kubernetes
source https://jimmysong.io
Grafana
elastic
maven
Registry
Dockerfile
GitLab
Jenkins
docker
MySQL
MySQL
Production
Dev
LOG
Scale
(running processes)
Workload diversity
(process types)
worker.4
worker.3
worker.2
worker.1
web.2
web.1
clock.1


Low volume
High volume
Request-scoped metrics
Metrics
Aggregatable
Tracing
Request scoped
Logging
Events
Aggregatable events e.g. rollups
Request-scoped, aggregatable events
Request-scoped events

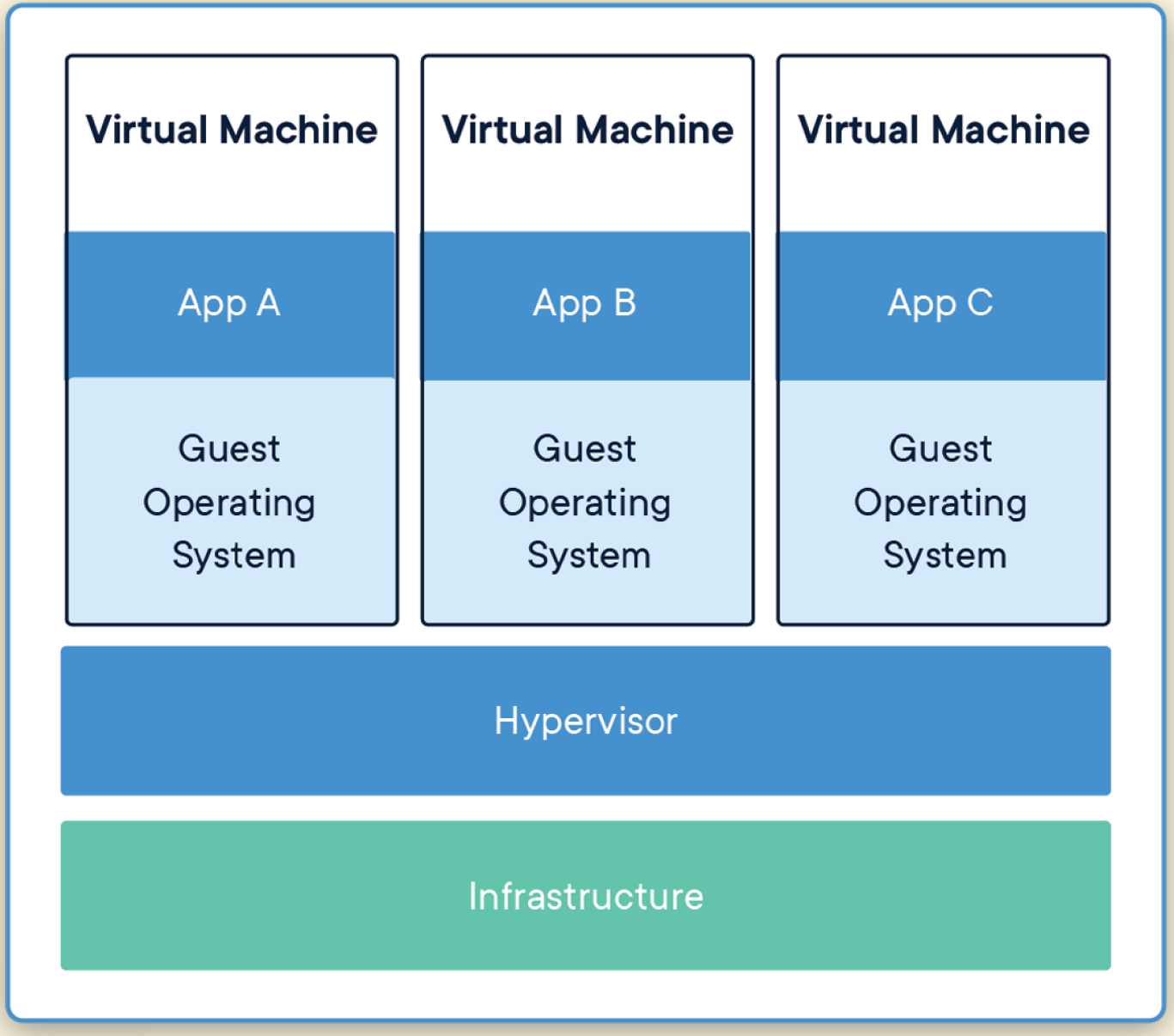

Kubernetes
Kubernetes cluster
Pod
docker
Filebeat as a sidecar
Deploy
Push
Call
elastic
Show
kibana
filebeat-test.yaml
Log Collection
Define
Wacth
Doc: filebeat-docker-test/log/AVzs1QzKdPYTkmBgkats
Table
JSON
@timestamp June 28th 2017, 11:53:19.425
_id AVzs1QzKdPYTkmBgkats
_index filebeat-docker-test
_score 1
_type log
beat.hostname filebeat-test-2365467882-4zwx8
beat.name filebeat-test-2365467882-4zwx8
beat.version 5.4.0
input_type log
message 2017-06-02 15:14:48.592 [net.sf.ehcache.CacheManager@614a616] INFO net.sf.ehcache.util.UpdateChecker-New update(s) found: 2.6.5 [http://www.terracotta.org/confluence/display/release/Release+Notes+Ehcache+Core+2.6]. Please check http://ehcache.org for the latest version.
offset 272
source /log/2017_06_02.stderrout.log
type log
© Jimmy Song https://github.com/rootsongjc/kubernetes-handbook

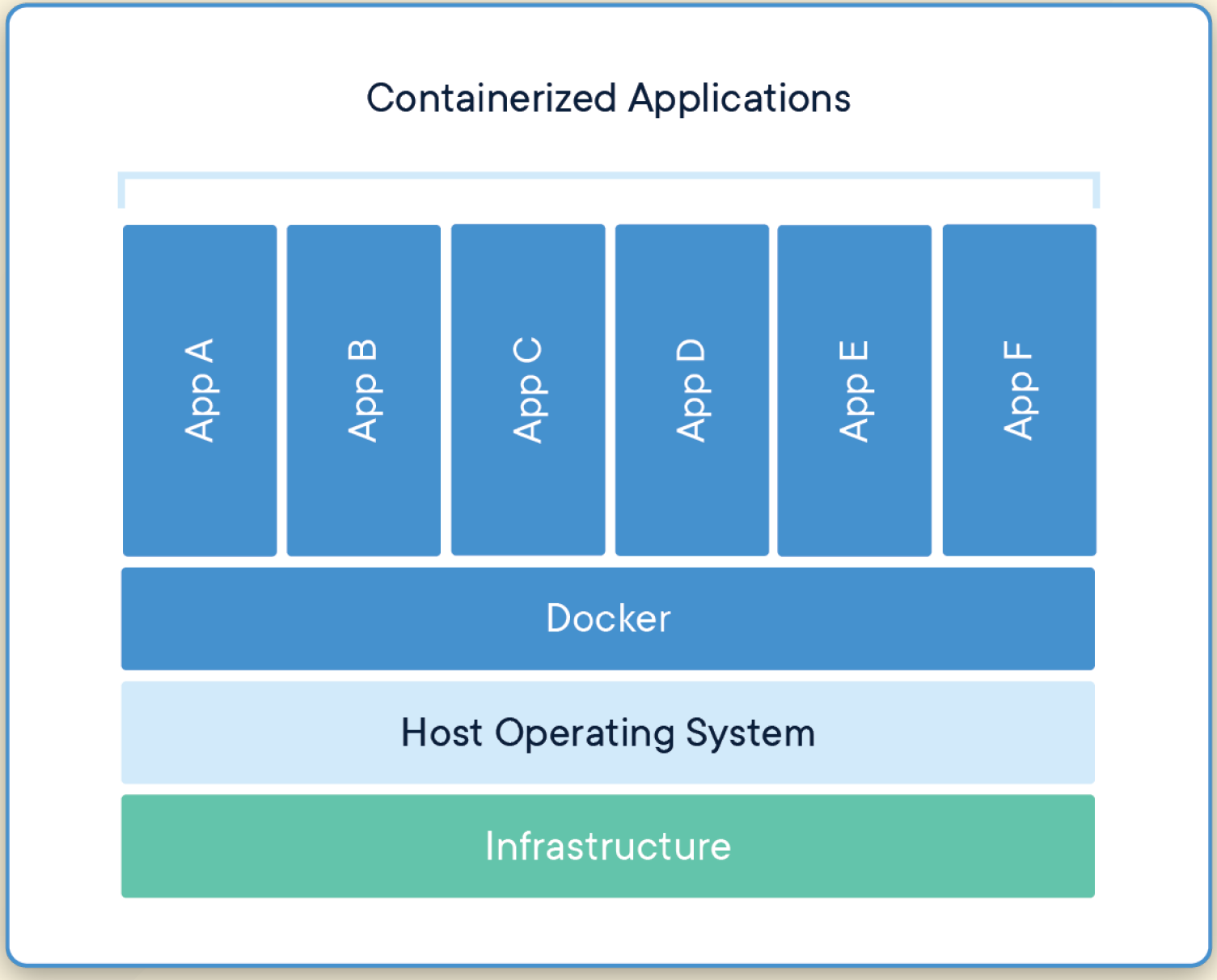

Stats Overview
Last 1 hour
Memory / CPU
memory
cpu
logins
timeshift -1h
logins
logins (-1 hour)
Memory
80 B
Sign ups
275
Logouts
161
Sign outs
279
server requests
web_server_01
web_server_02
web_server_03
web_server_04
client side full page load
70 GB
60 GB
60 GB
30 GB
40 GB
83 GB
44 GB
47 GB
94 GB
4 GB
73 GB
44 GB
19 GB
22 GB
20 GB
81 GB
sda1
sda2
sda3
sda4
sda5
sda6
sda7
sda8
sda9
sda10
sda11
sda12
sda13
sda14
sda15
sda16


Jaeger UI
Lookup by Trace ID...
Search
Compare
Dependencies
About Jaeger
istio-ingressgateway: eshop.default.svc.cluster.local:8080/checkout* 475a149
Search...
Trace Timeline
Trace Start July 2, 2019 9:57 AM
Duration 227.41ms
Services 6
Depth 8
Total Spans 17
0ms
56.85ms
113.71ms
170.56ms
227.41ms
Service & Operation
istio-ingressgateway eshop.default.svc.cluster.local:8080/...
eshop.default eshop.default.svc.cluster.local:8080/chec...
eshop-opentracing checkout
eshop-opentracing GET
88.14ms
eshop.default inventory.default.svc.cluster....
86.6ms | eshop.default::inventory.default.svc.cluster.local:8080/*
inventory.default inventory.default.sv...
86.08ms
eshop-opentracing createOrder
84.29ms
eshop-opentracing GET
102.97ms
eshop.default billing.default.svc.cluster.loc...
100.45ms
billing.default billing.default.svc.cluste...
99.81ms
eshop-opentracing payment
96.83ms
eshop-opentracing com.zh...
62.16ms
eshop-opentracing com.zh...
31.11ms
eshop-opentracing GET
24.54ms
eshop.default delivery.default.svc.cluster.l...
23.29ms
delivery.default delivery.default.svc.cl...
22.71ms
eshop-opentracing arrangeDeli...
20.19ms

Docker and Kubernetes

> **Docker** is a set of platform as a service (PaaS) products that use OS-level virtualization to deliver software in packages called containers.
> **Kubernetes** is an open-source container-orchestration system for automating computer application deployment, scaling, and management.

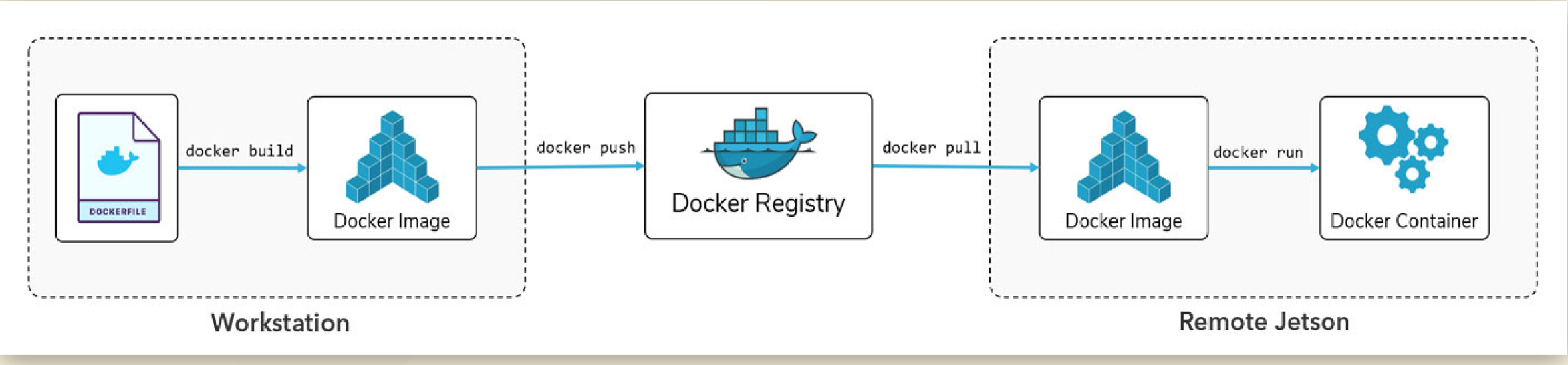

Virtual Machine
Virtual Machine
Virtual Machine
App A
App B
App C
Guest Operating System
Guest Operating System
Guest Operating System
Hypervisor
Infrastructure


Containerized Applications
App A
App B
App C
App D
App E
App F
Docker
Host Operating System
Infrastructure

特性	VM	Docker
启动速度	分钟级	秒级
CI/CD	😢	测试、生产环境一致
性能	损耗大	近似物理机
体积	较大(GB)	极小(MB)
拓展	😢	跨平台

Docker Workflow

- Image: 跨平台、可移植的程序+环境包, 基于Dockerfile制作
- Register: 镜像的存储位置，
  官方镜像仓库地址（https://hub.docker.com/）
- Container: 进行了资源隔离的镜像运行时环境



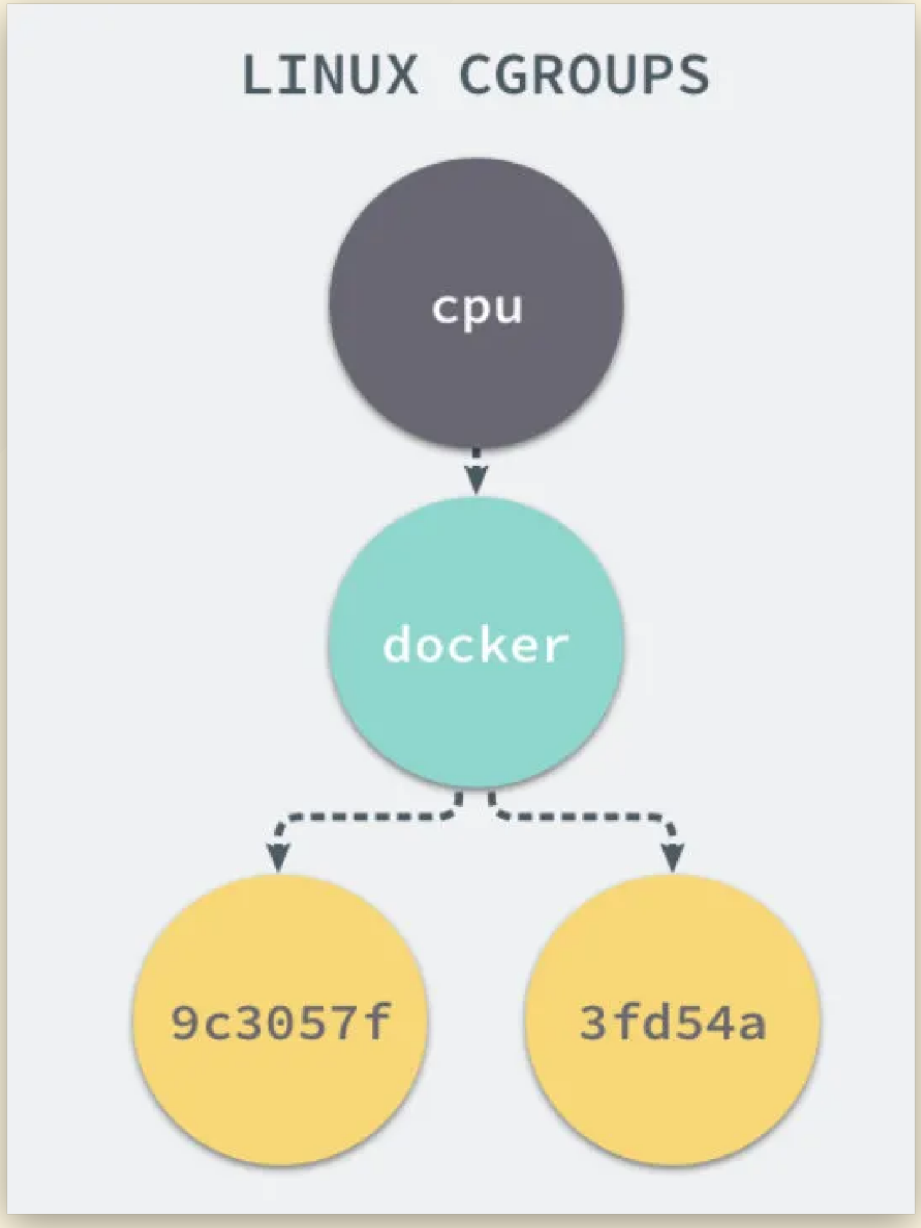
Docker 实现原理

- Linux **Namespaces**:
  - CLONE_NEWNS (mount)
  - CLONE_NEWUTS (Unix Time Sharing)
  - CLONE_NEWIPC (InterProcess Communication)
  - CLONE_NEWPID (Process ID)
  - CLONE_NEWNET
  - CLONE_NEWUSER

Clone(2)

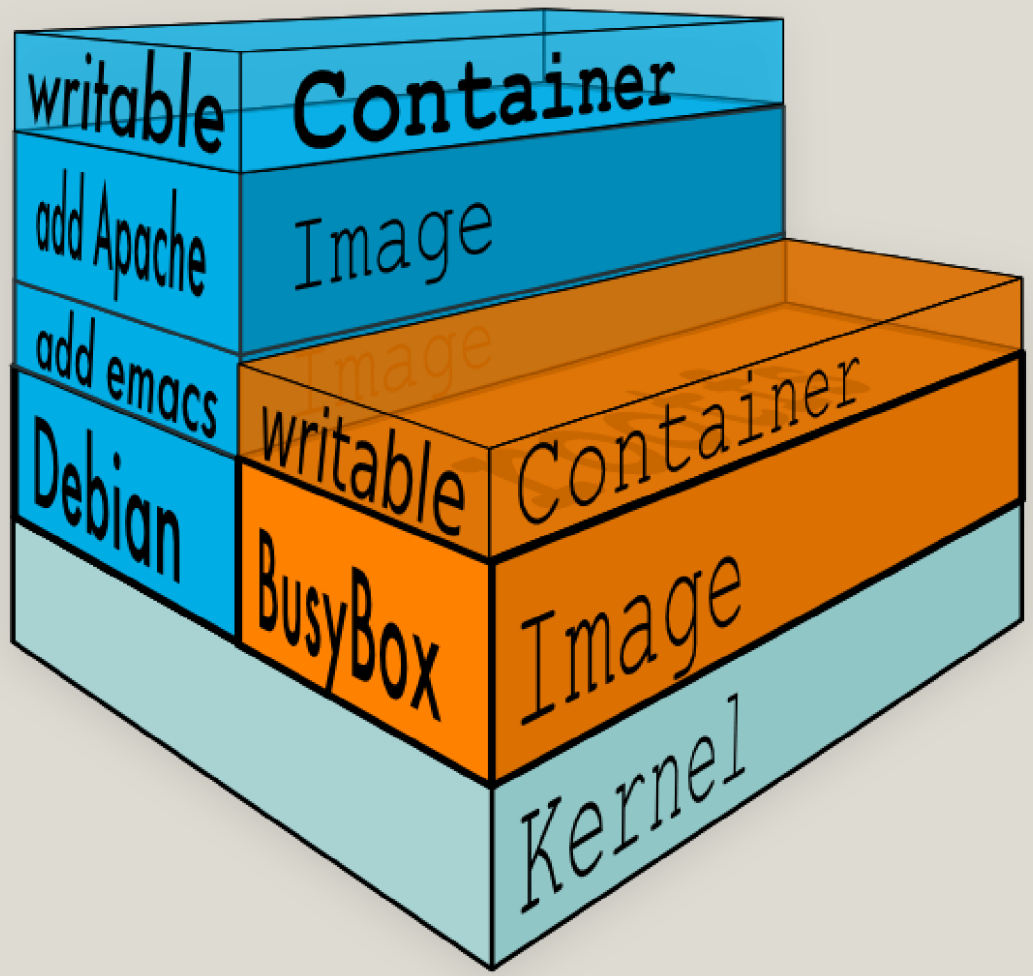
```
int clone(int (*fn)(void*), void *stack, int flags, void *arg);
```

- **CGroups** (Control Groups)
  - Limit_Cpu
  - Limit_Memory
  - Limit_IO



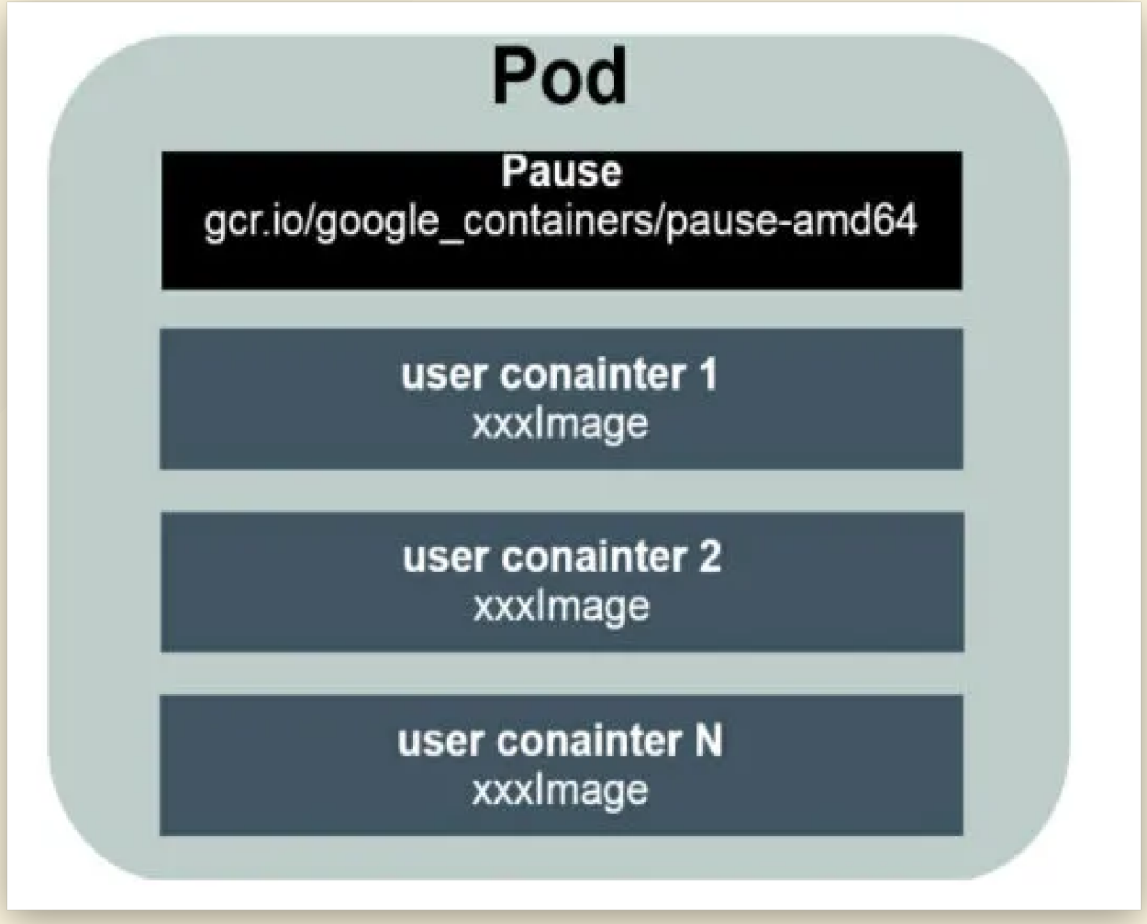
- **UnionFS** (Union File System)

UnionFS可以把多个目录内容联合挂载到同一个目录下，而目录的物理位置是分开的。UnionFS 可以把只读和可读写文件系统合并在一起，具有写时复制功能，允许只读文件系统的修改可以保存到可写文件系统当中。



Kubernetes Concepts

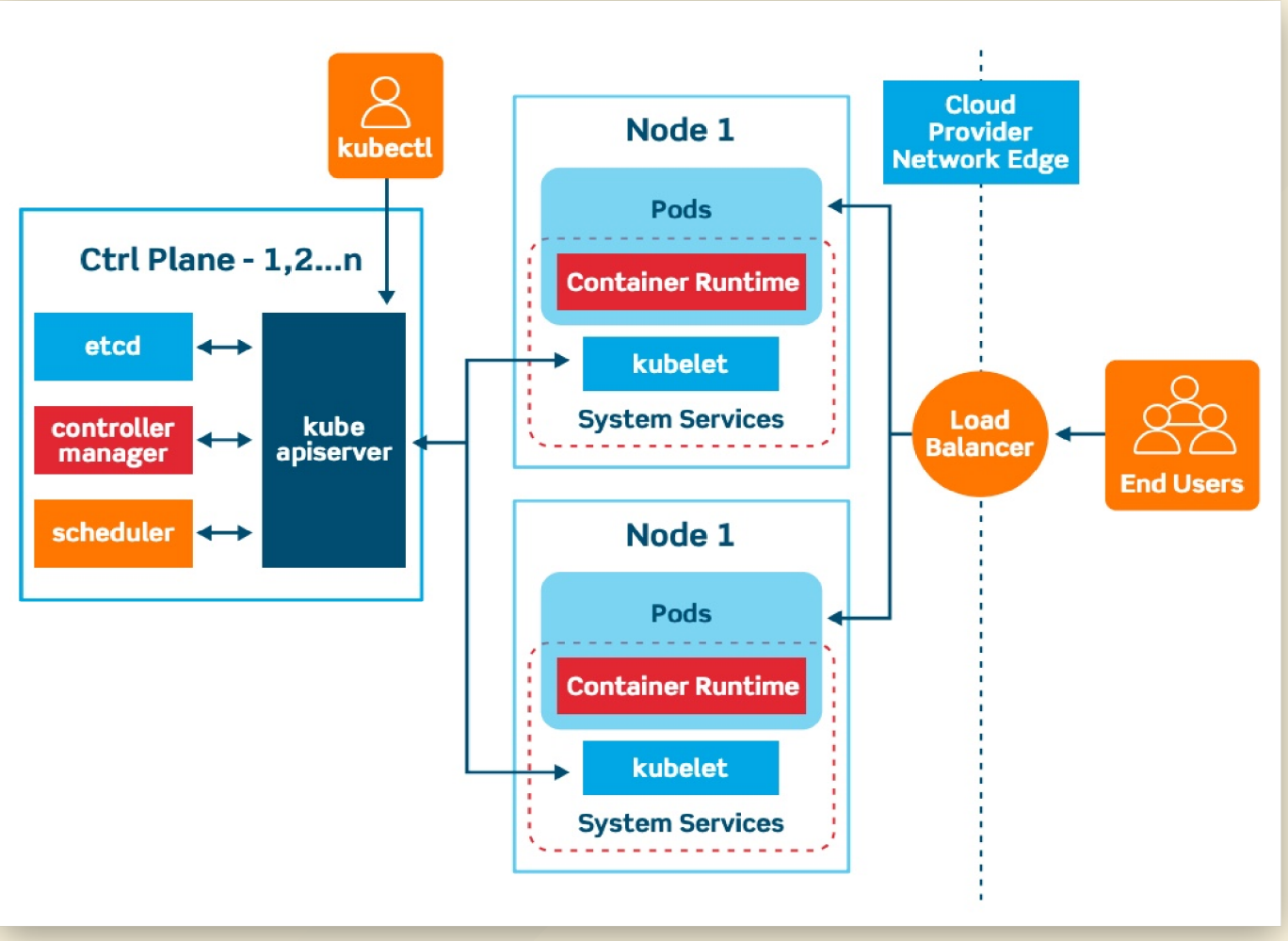
- Pod
  Pods are the smallest deployable units of computing that you can create and manage in Kubernetes.
- Namespaces
  Kubernetes supports multiple virtual clusters backed by the same physical cluster. These virtual clusters are called namespaces.



Kubernetes Features

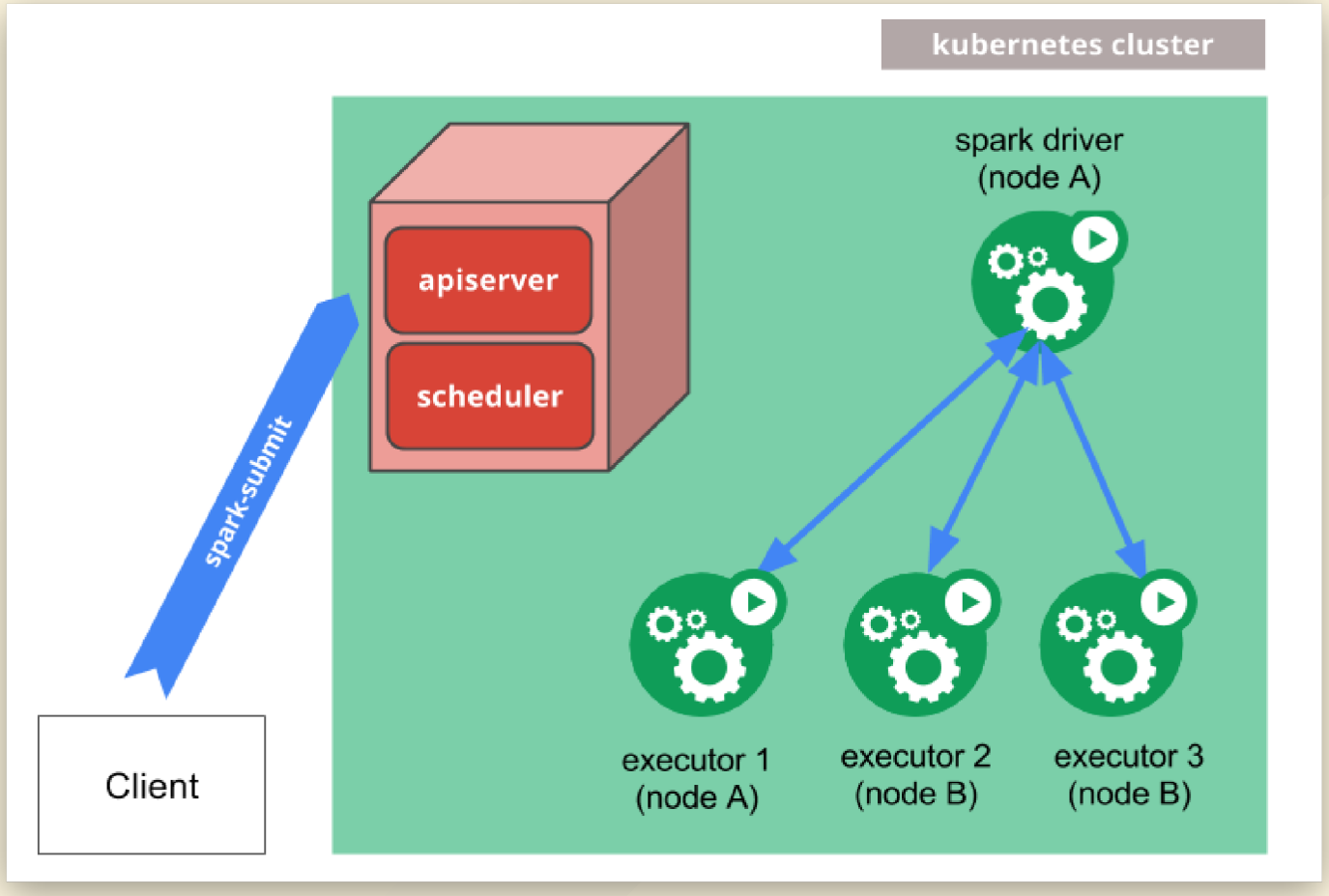
- Service discovery and load balancing
- Automated rollouts and rollbacks
- Automatic bin packing
- Self-healing
- Secret and configuration management

Kubernetes Architecture



Spark on Kubernetes

How it works

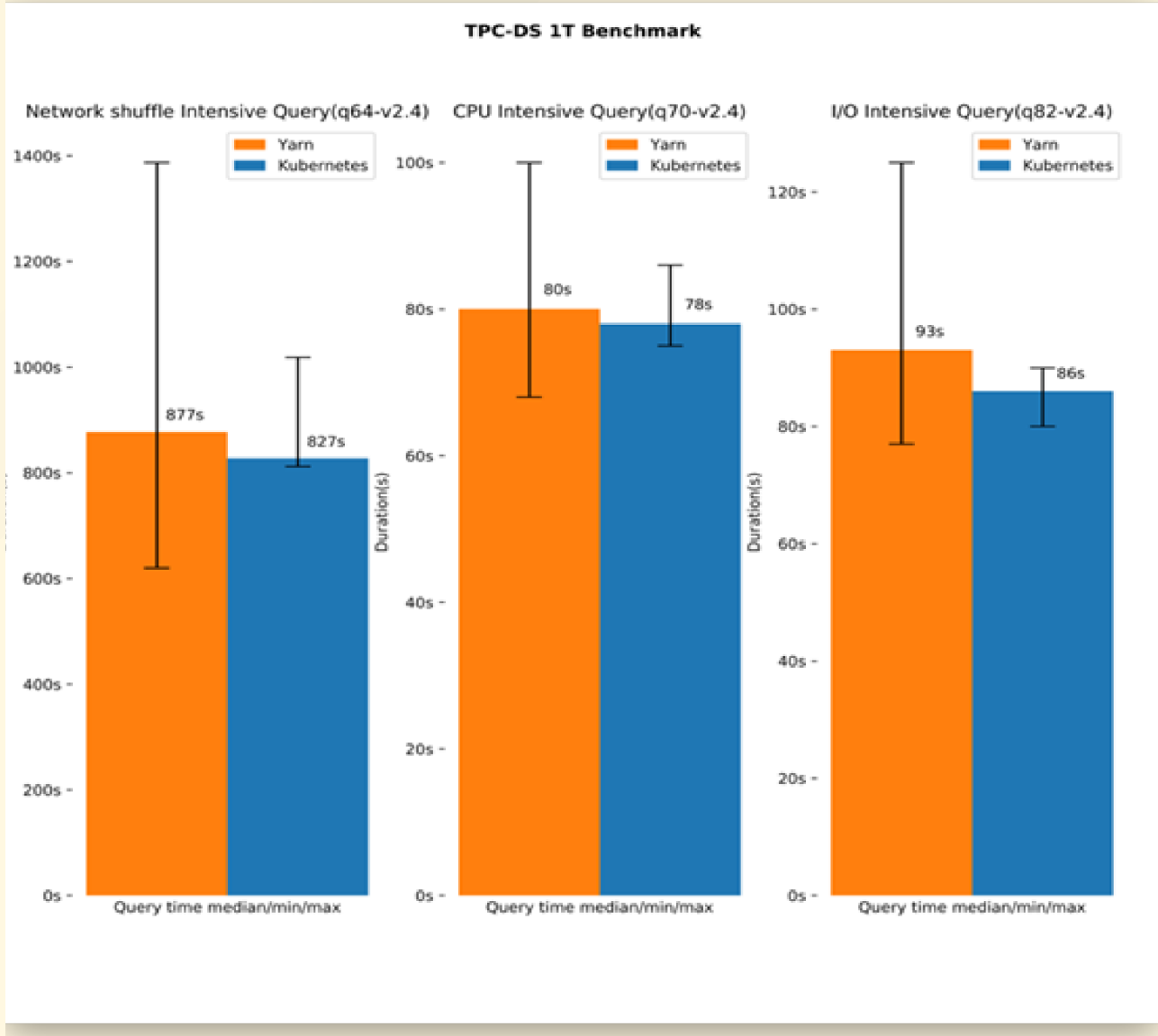

kubernetes cluster
spark driver
(node A)
apiserver
scheduler
spark-submit
Client
executor 1
(node A)
executor 2
(node B)
executor 3
(node B)

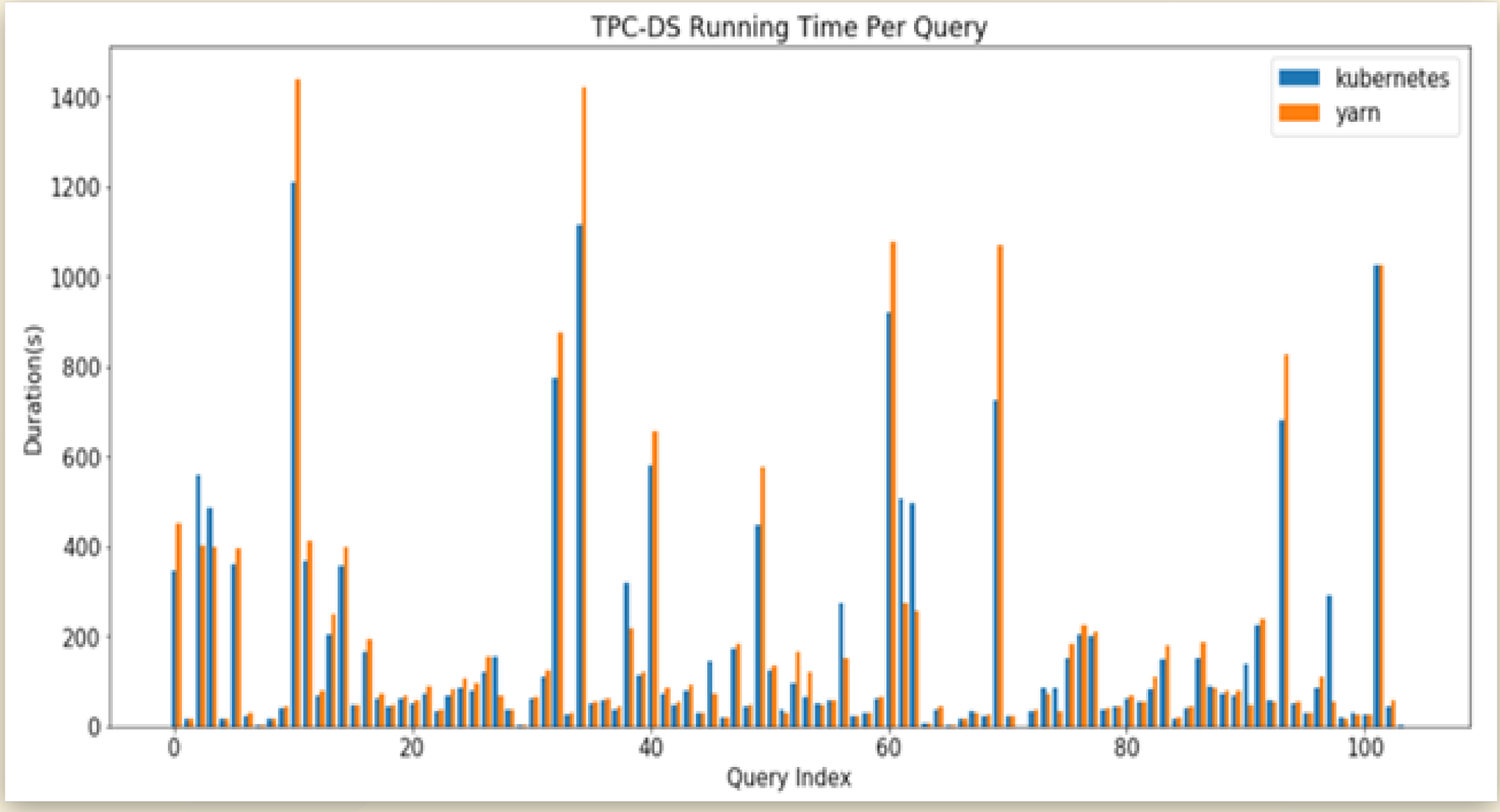
Kubernetes vs Yarn

- Version and dependency control
- Isolating jobs
- Kubernetes ecosystem

TPC-DS is the de-facto standard benchmark for measuring the performance of decision support solutions.

> [AWS]Optimizing Spark performance on Kubernetes ”




TPC-DS Running Time Per Query
kubernetes
yarn
Duration(s)
0
200
400
600
800
1000
1200
1400
0
20
40
60
80
100
Query Index

Demo Show

Thank You

For Your Attention