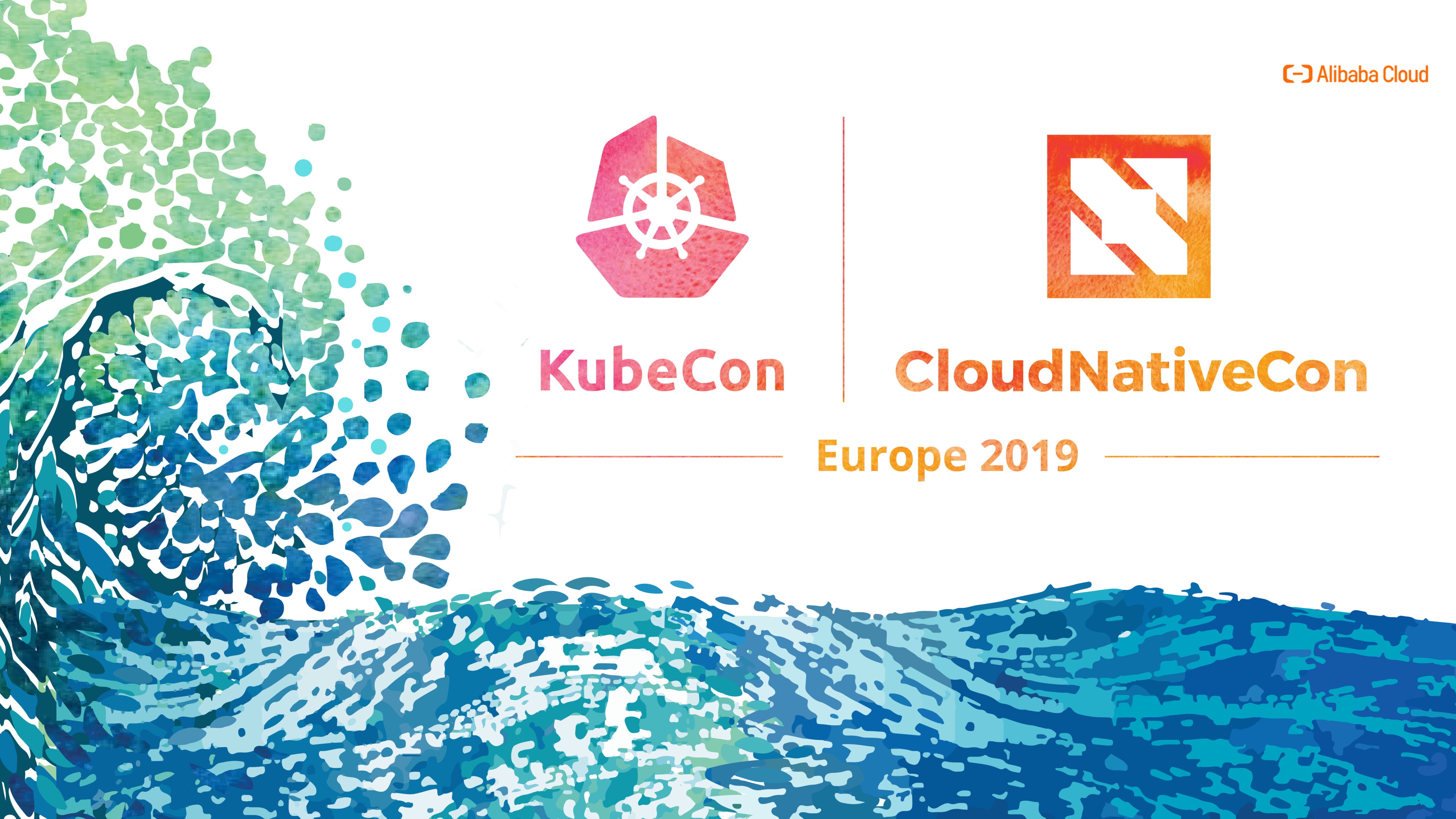
Alibaba Cloud

KubeCon | CloudNativeCon

Europe 2019



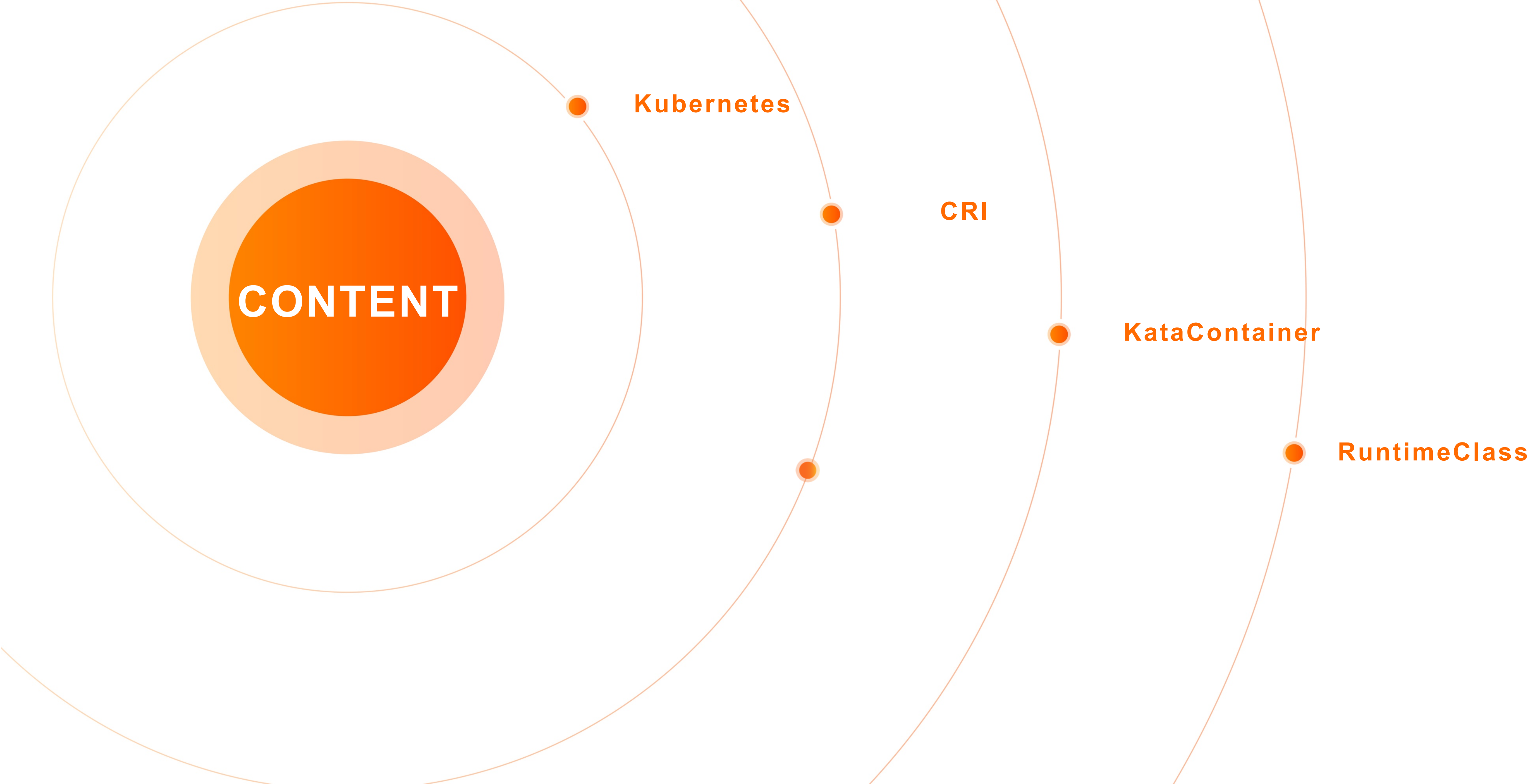
Alibaba Cloud |

Worldwide Cloud Services Partner

Surviving Serverless Battle By Secure Runtime, CRI and RuntimeClass

Lei Zhang & Xiaoyu Zhang, Alibaba Group

Content



CONTENT

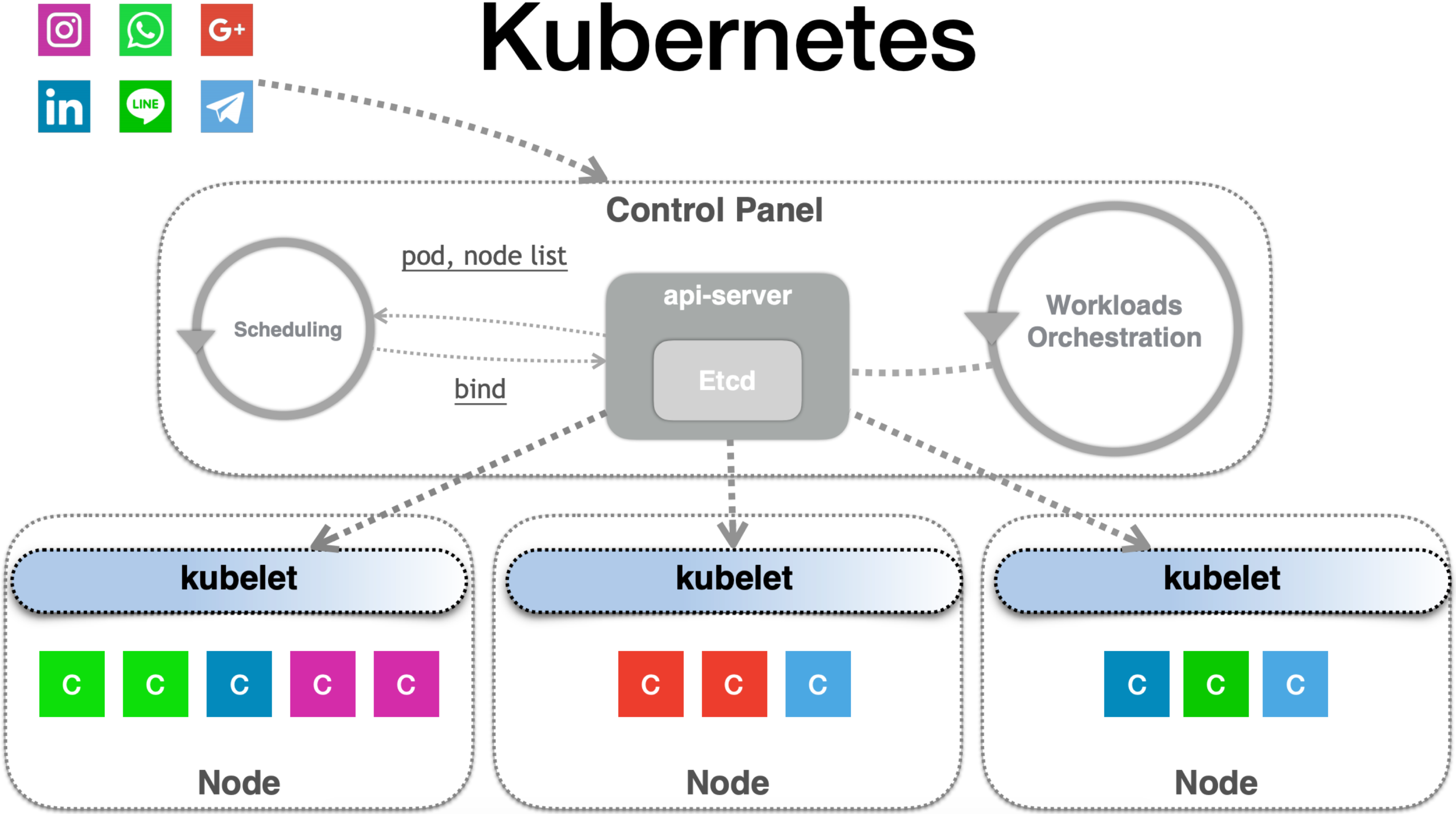
Kubernetes

CRI

KataContainer

RuntimeClass

Kubernetes

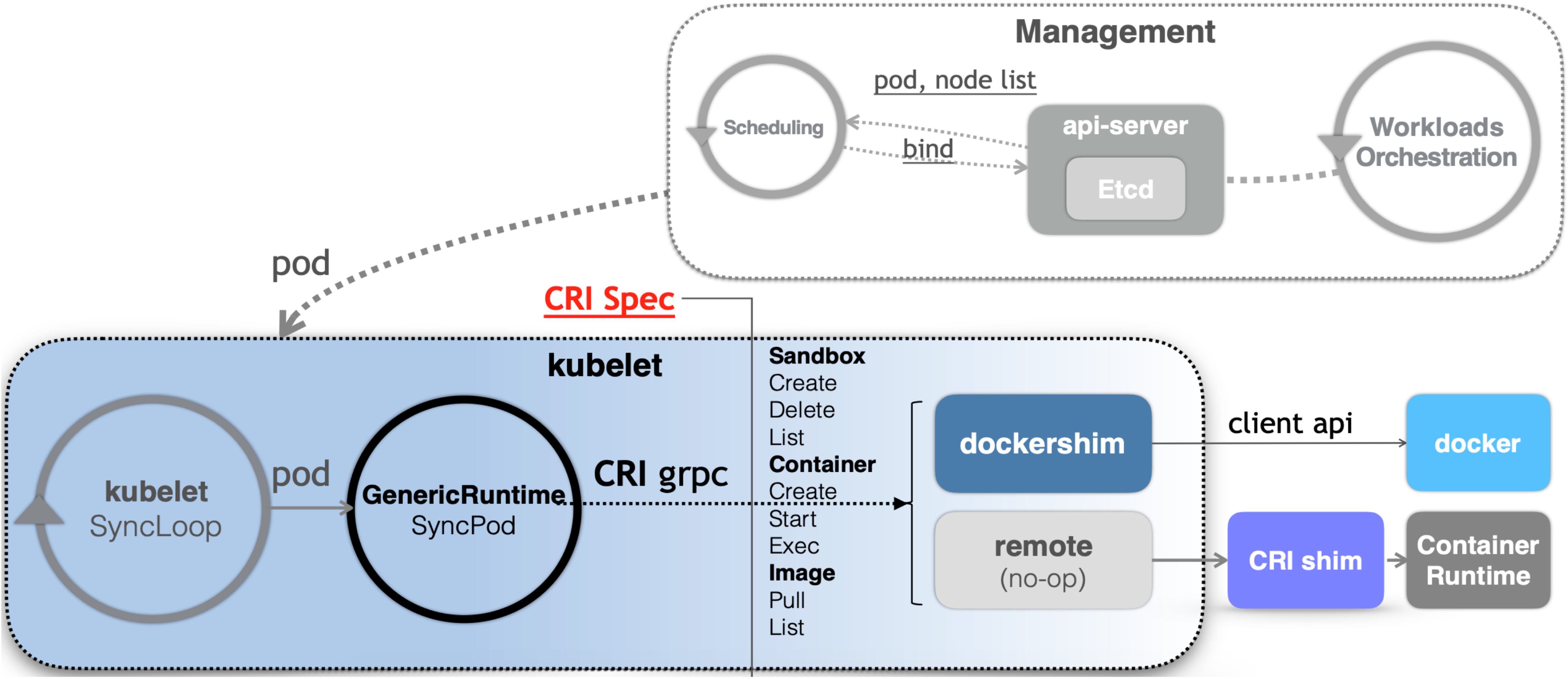

Control Panel
pod, node list
Scheduling
bind
api-server
Etcd
Workloads Orchestration
kubelet
C
C
C
C
C
Node
kubelet
C
C
C
Node
kubelet
C
C
C
Node

Container Runtime Interface (CRI)

- Describe what kubelet expects from container runtimes
- Imperative container-centric interface
  - **why not pod-centric?**
    - Every container runtime implementation needs to understand the concept of pod.
    - Interface has to be changed whenever new pod-level feature is proposed.

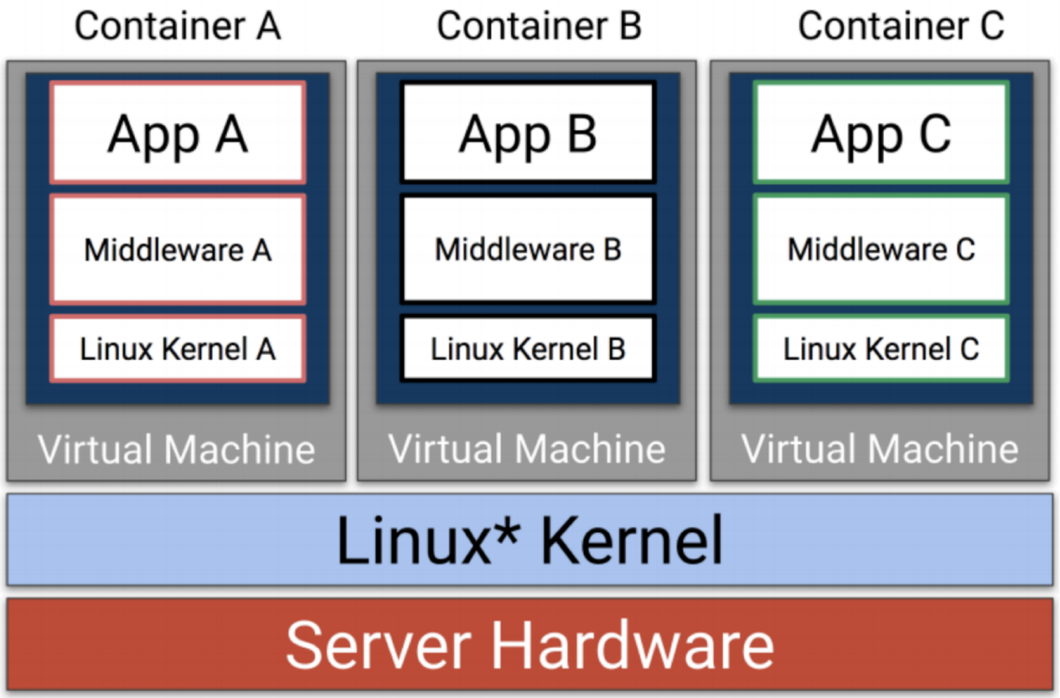
Alibaba Cloud | Worldwide Cloud Services Partner

How CRI Works



KataContainers

- **Container Runtime**
  - Each Pod is hypervisor isolated
    - Independent guest kernel
  - Secure as VM
  - Fast as container
- **Container Image**
  - Same as Linux container

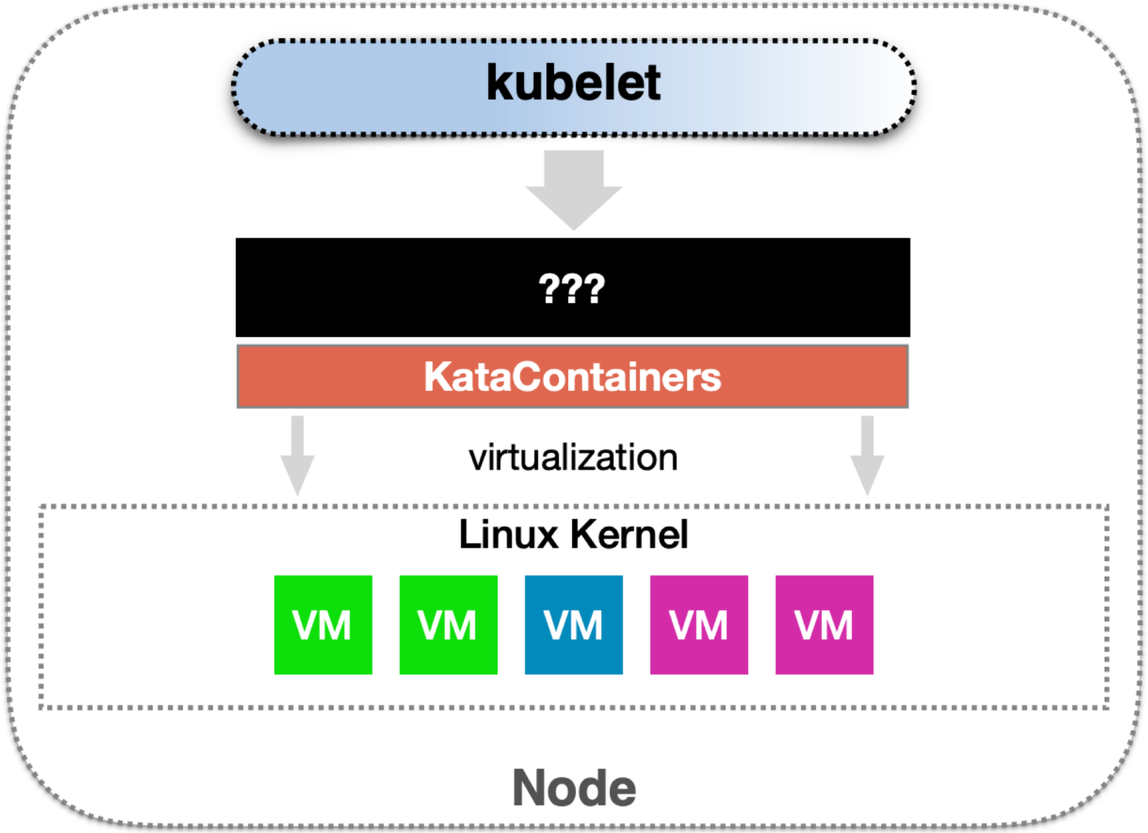


Container Security

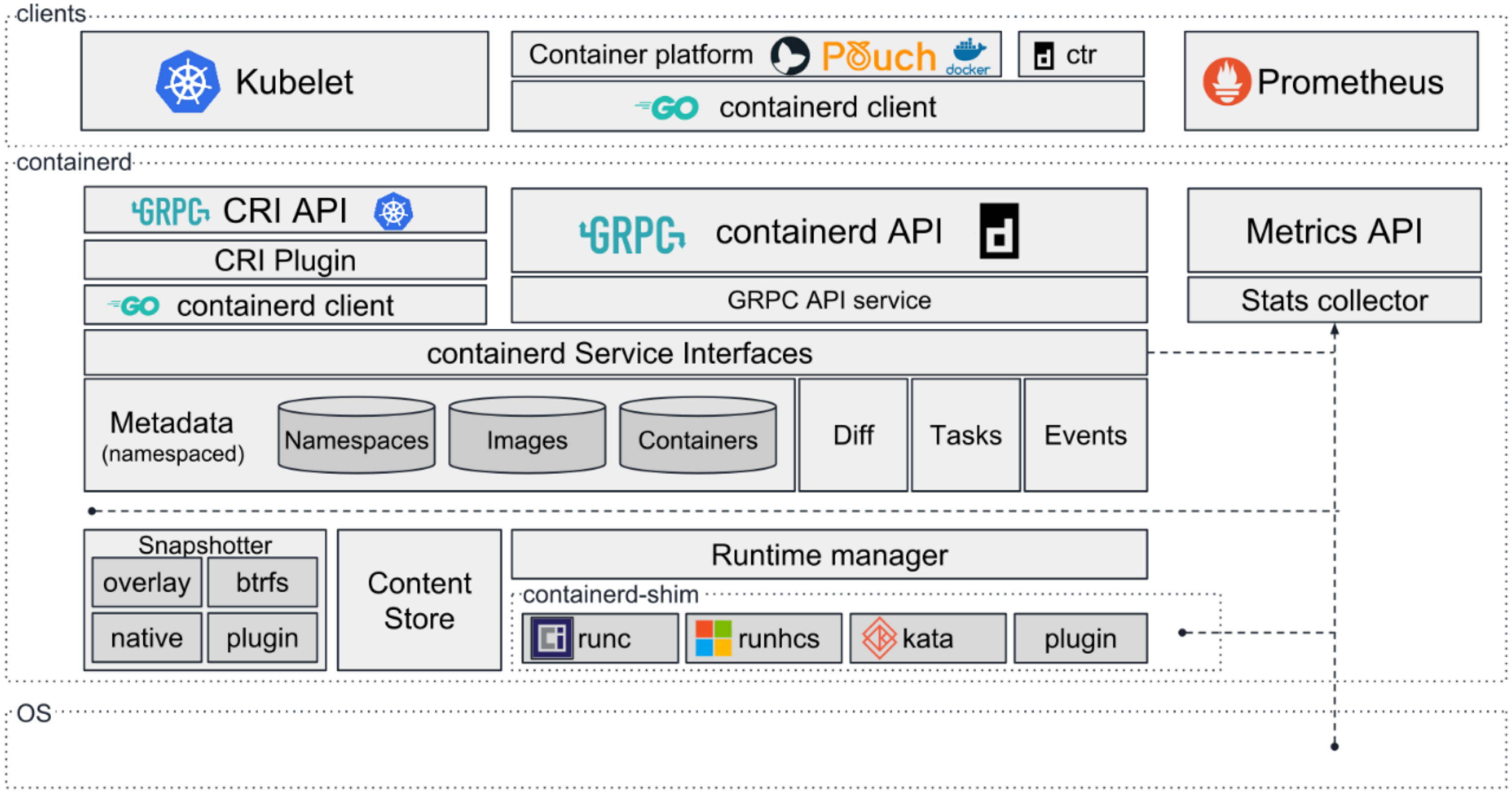
- Linux container
  - Dropping Linux capabilities
  - Read-only mount points
  - Mandatory access controls (MAC)
    - SELinux & AppArmor
  - Dropping syscalls
    - SECCOMP
  - In 99.99% cases
    - wrap containers in VMs

- KataContainers
  - Hardware virtualization
  - Independent Linux instance per Pod
    - e.g. run Linux 3.16 container on a Linux 4.0 host

Kubernetes + KataContainers



K8s + CRI + Containerd + KataContainer



Copied from https://containerd.io/

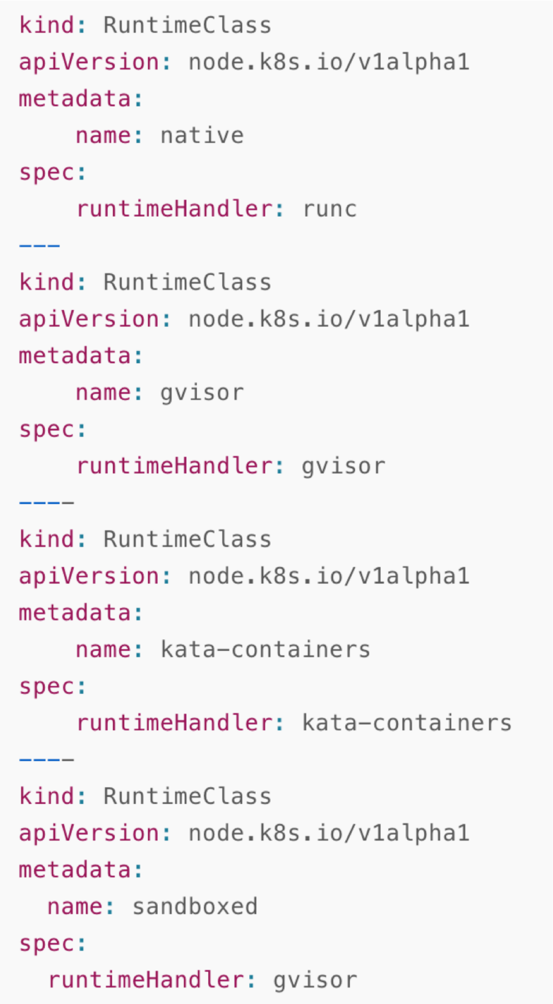
RuntimeClass

- How can we use different runtimes in a single Kubernetes cluster?
  - RuntimeClass
- What can RuntimeClass do for us?
  - More choices
  - Better abstract
- Why RuntimeClass is significant important to CNCF and the whole OpenSource ecosystem?
  - Provide a mechanism for surfacing container runtime properties to the control plane
  - Support multiple runtimes per-cluster, and provide a mechanism for users to select the desired runtime

K8s + RuntimeClass + X

Alibaba Cloud | Worldwide Cloud Services Partner

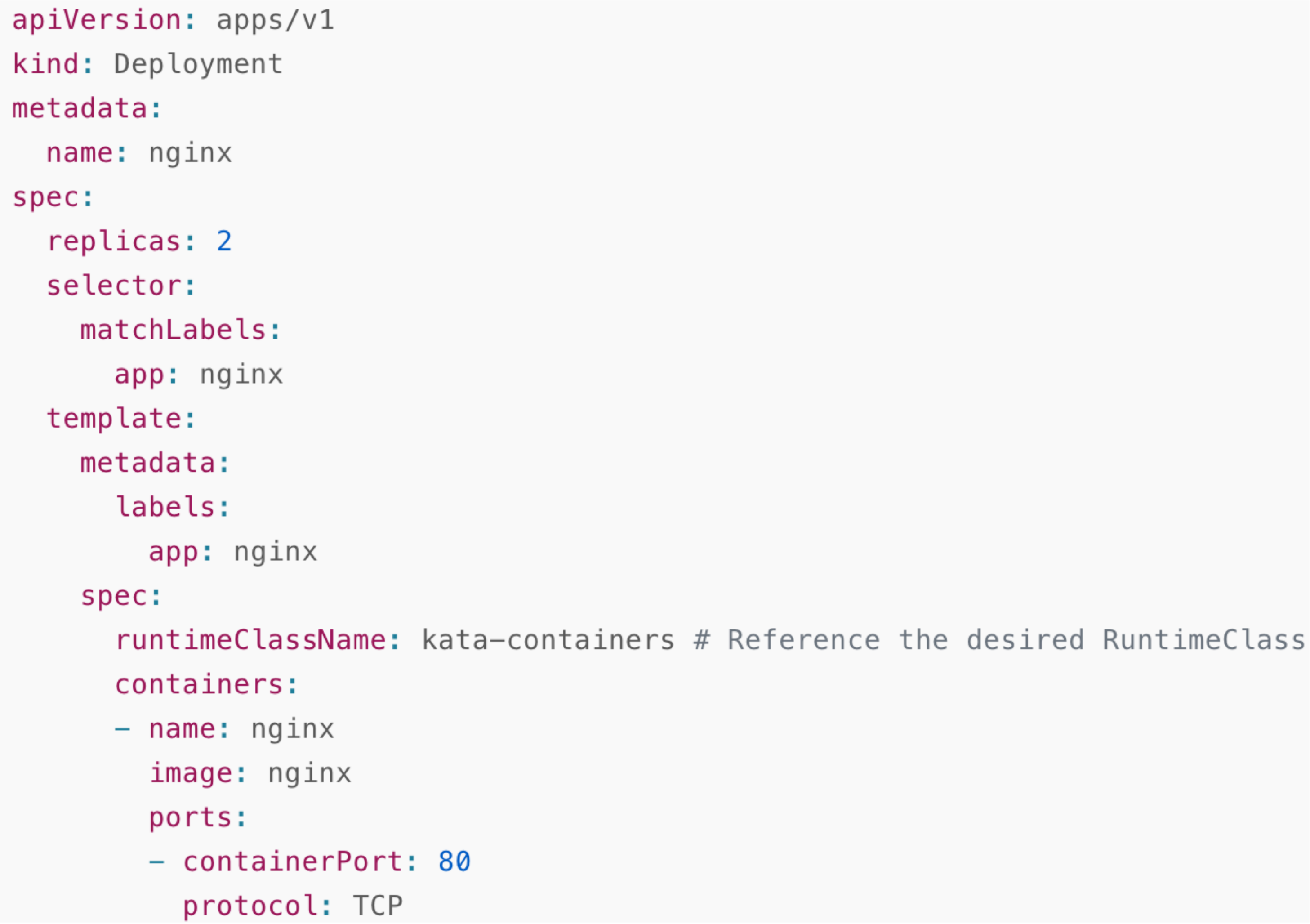
- RuntimeClass Config



```
kind: RuntimeClass
apiVersion: node.k8s.io/v1alpha1
metadata:
    name: native
spec:
    runtimeHandler: runc
---
kind: RuntimeClass
apiVersion: node.k8s.io/v1alpha1
metadata:
    name: gvisor
spec:
    runtimeHandler: gvisor
----
kind: RuntimeClass
apiVersion: node.k8s.io/v1alpha1
metadata:
    name: kata-containers
spec:
    runtimeHandler: kata-containers
----
kind: RuntimeClass
apiVersion: node.k8s.io/v1alpha1
metadata:
  name: sandboxed
spec:
  runtimeHandler: gvisor
```

K8s + RuntimeClass + X

- RuntimeClass Use Case



```yaml
apiVersion: apps/v1
kind: Deployment
metadata:
  name: nginx
spec:
  replicas: 2
  selector:
    matchLabels:
      app: nginx
  template:
    metadata:
      labels:
        app: nginx
    spec:
      runtimeClassName: kata-containers # Reference the desired RuntimeClass
      containers:
      - name: nginx
        image: nginx
        ports:
        - containerPort: 80
          protocol: TCP
```

Secure Runtime, CRI and RuntimeClass make Serverless Possible

Alibaba Cloud

Worldwide Cloud Services Partner

WWW.ALIBABA CLOUD.COM