
关注【金融级分布式架构】公众号，
获取更多云原生相关技术内容
KubeCon
CloudNativeCon
OPEN SOURCE SUMMIT
China 2019

Managing Large-Scale Kubernetes Clusters Effectively and Reliably

Yong Zhang – Ant Financial

About Me

- Yong Zhang
  - Ant Financial - Infra & Data
  - PAAS & Automated Cluster Management System

Agenda

- Background and Motivation
- Design Concept
- Cluster Management Operators
- Q & A

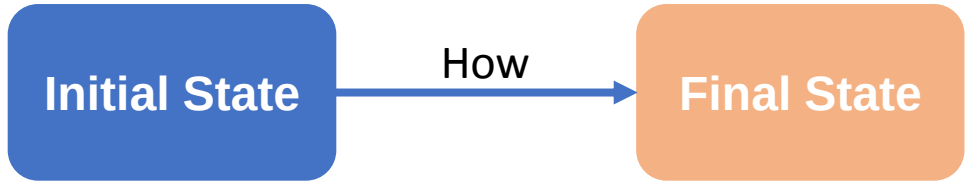
Background: Cluster Scale

- Tens of clusters
  - Tens of thousands of nodes in one cluster
- Hundreds of thousands of pods
  - Tens of thousands of jobs
- Resource cost is huge

Motivation

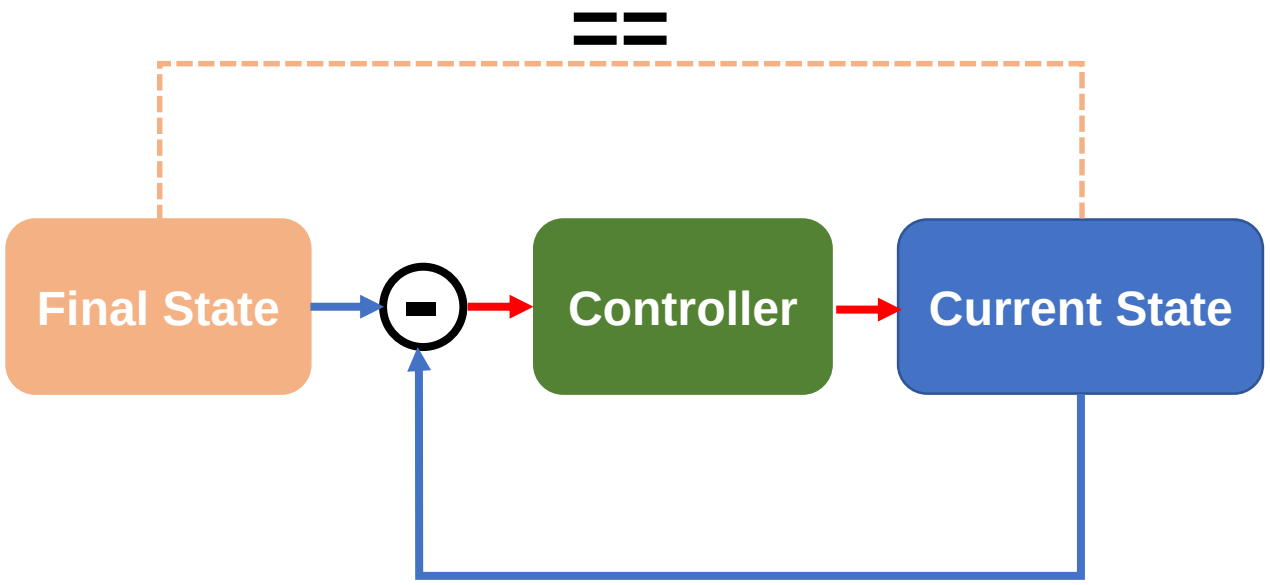
- Cluster Life Cycle Management
  - Create
  - Delete
  - Upgrade
  - …
- Fault self-recovery:
  - hardware failure
  - component service exception
- Change Controllable

Design Concept: Imperative



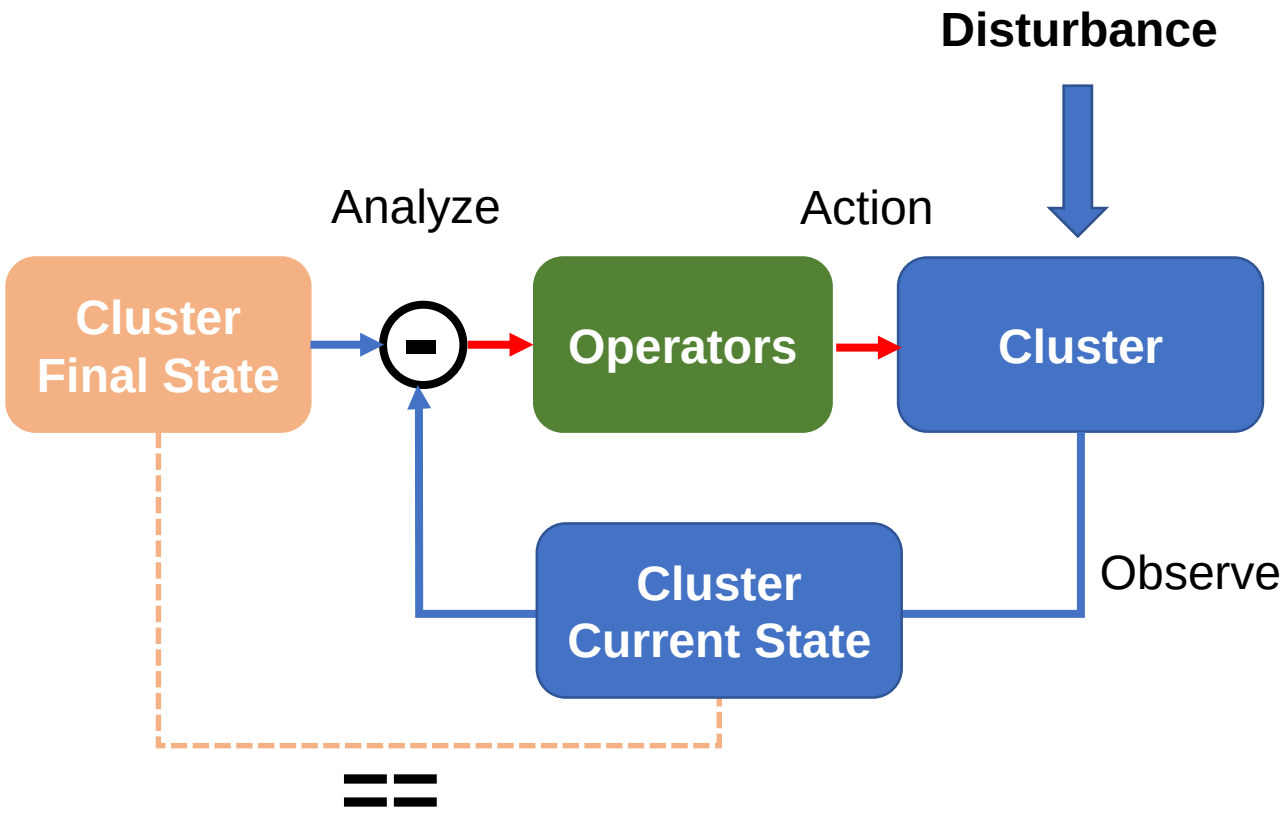
- Imperative
  - Tell What
  - Tell How

Design Concept: Declarative



- Declarative
  - Tell What
  - Not How

Design Concept: Self-recovery



- **Observe**

  watch actual state

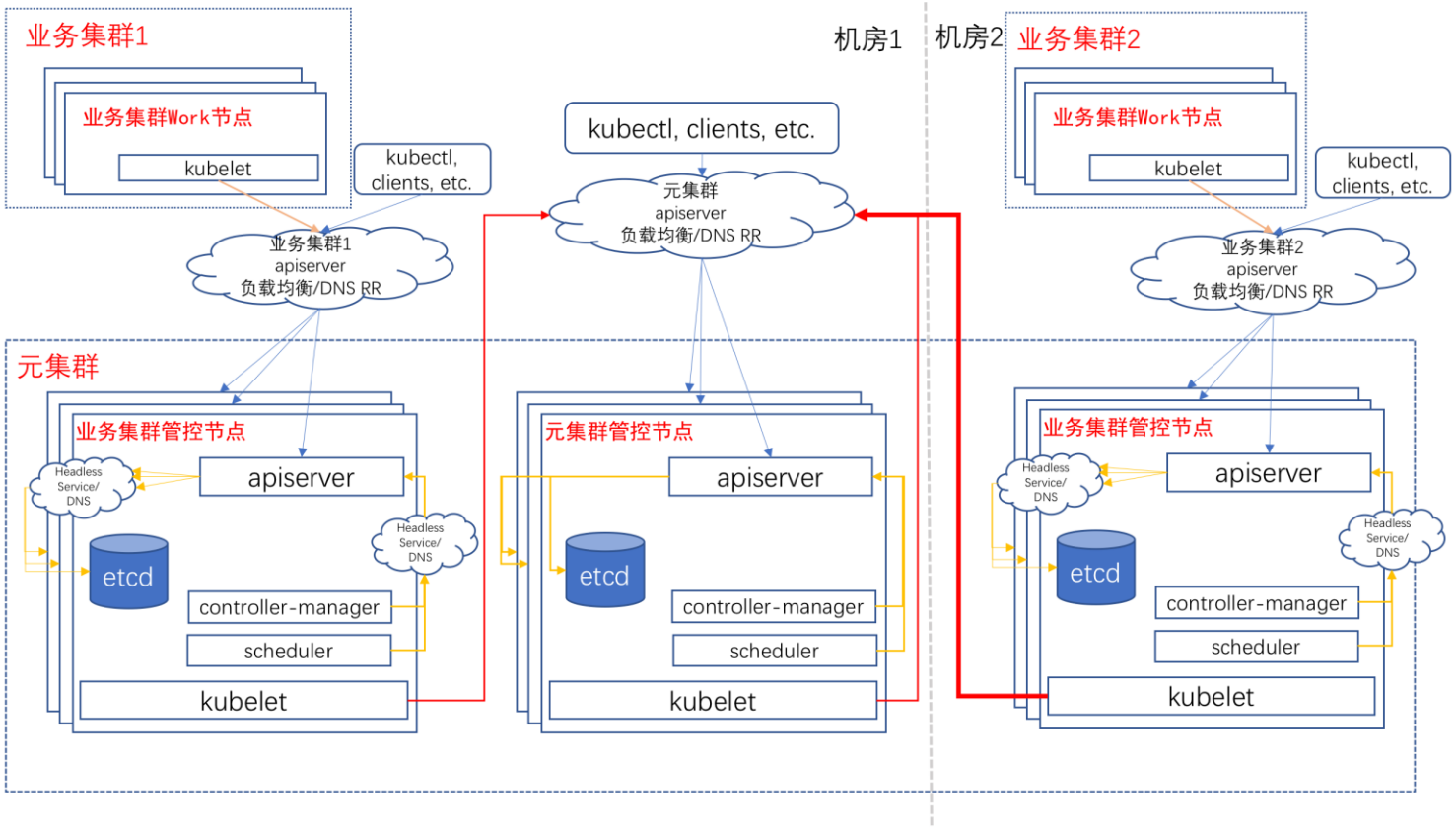
- **Analyze**

  difference from desired and actual state

- **Action**

  change to desired state

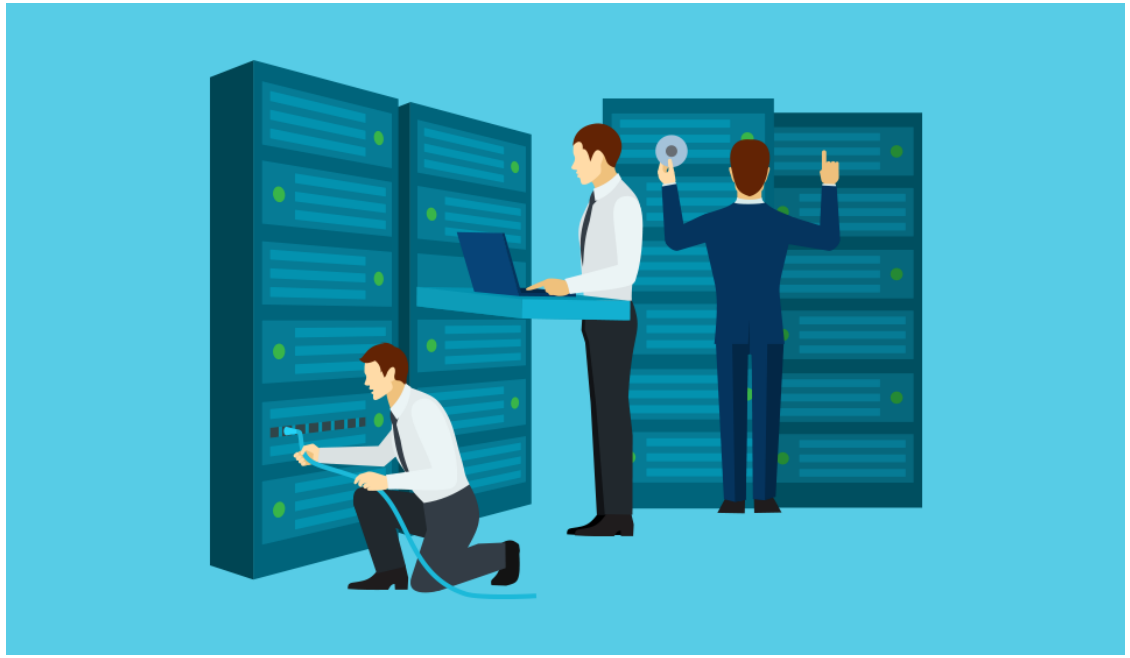
Cluster-Operator



- Cluster CRD
- ClusterPackageVersion
- Cluster-Operator

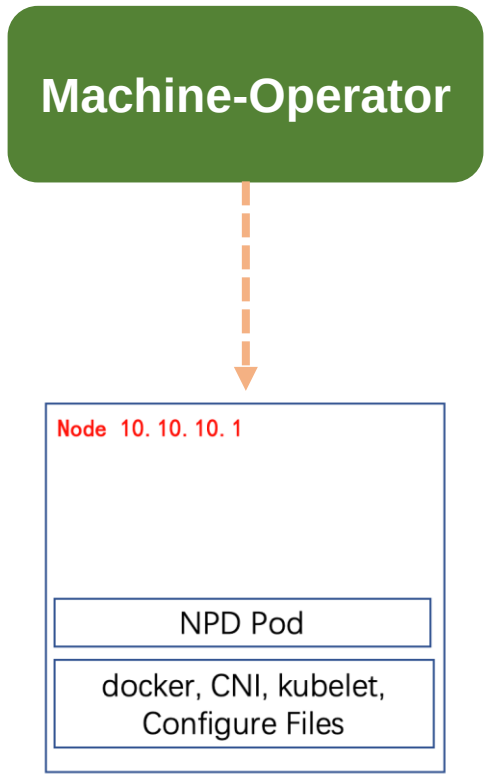
Machine-Operator

- Node Configuration and Kernel Patch Management
- Docker / Kubelet Install、Uninstall、Upgrade
- Node Final-state Management
- Node Fault Self-recovery



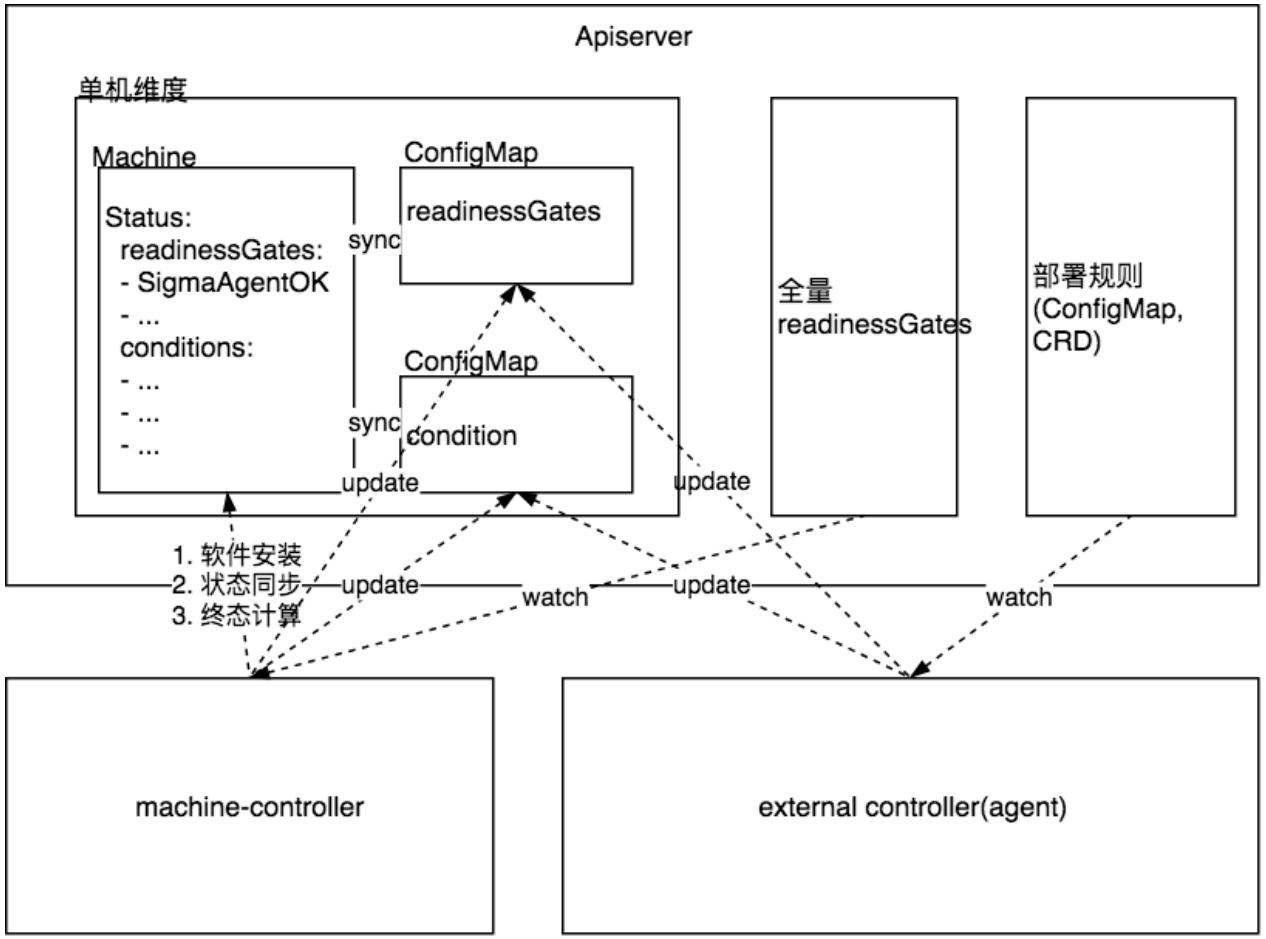
Machine-Operator


```
kind: Machine
spec:
  idc: xxx
  ip: 10.10.10.1
  versions:
    pouch: 1.0
    kubelet: 1.1
status:
  phase: Running
  readinessGates:
    - conditionType: PouchOK
    - conditionType: OSOK
  versions:
    pouch: 1.0
    kubelet: 1.0
```




```
kind: MachinePackageVersion
metadata:
  name: pouch-1.0
spec:
  packageName: pouch
  config:
    rpm: http://pouch-1.0.rpm
  configMaps:
    - name: pouch-config
      value: v1
```

Node Final-state Management



- ReadinessGates
  - Node Schedulable Conditions
- External controller
  - DaemonSet
  - ……
- Condition ConfigMap
  - External Conditions

Fault Self-recovery

Hardware

- Power
- Disk
- Memory
- Motherboard
- ……

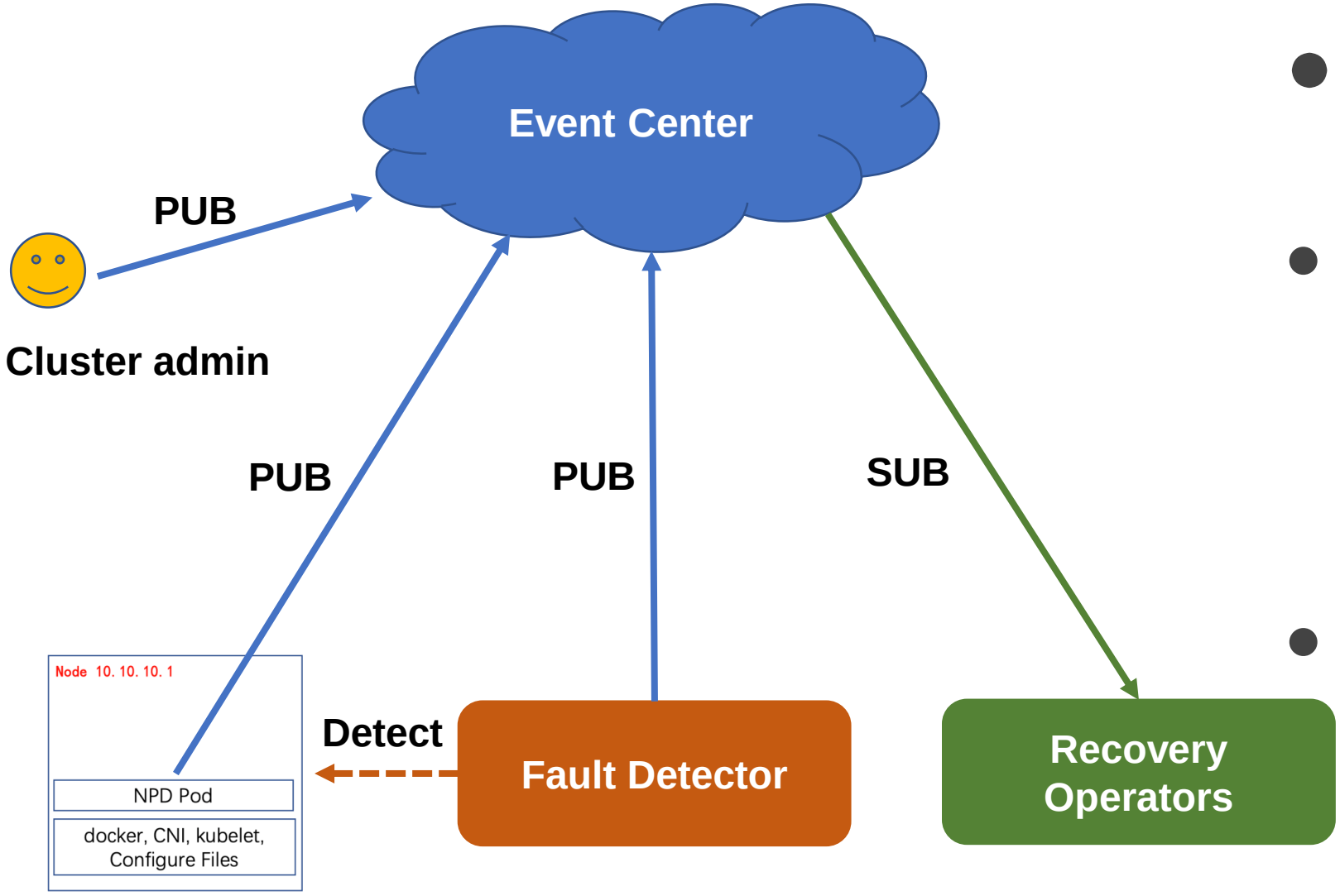
Resource

- Load
- Memory Pressure
- Disk Pressure

Component

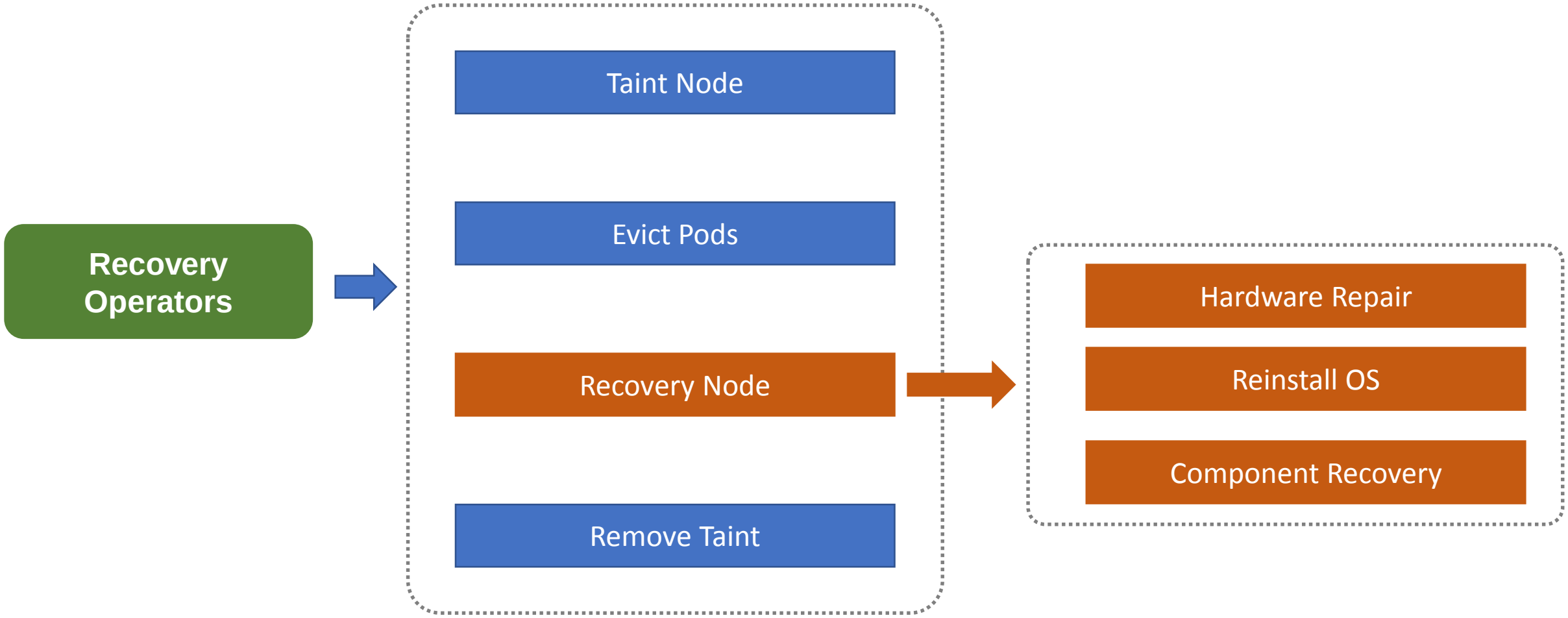
- Component Crash
- Configuration Error

Fault Self-recovery

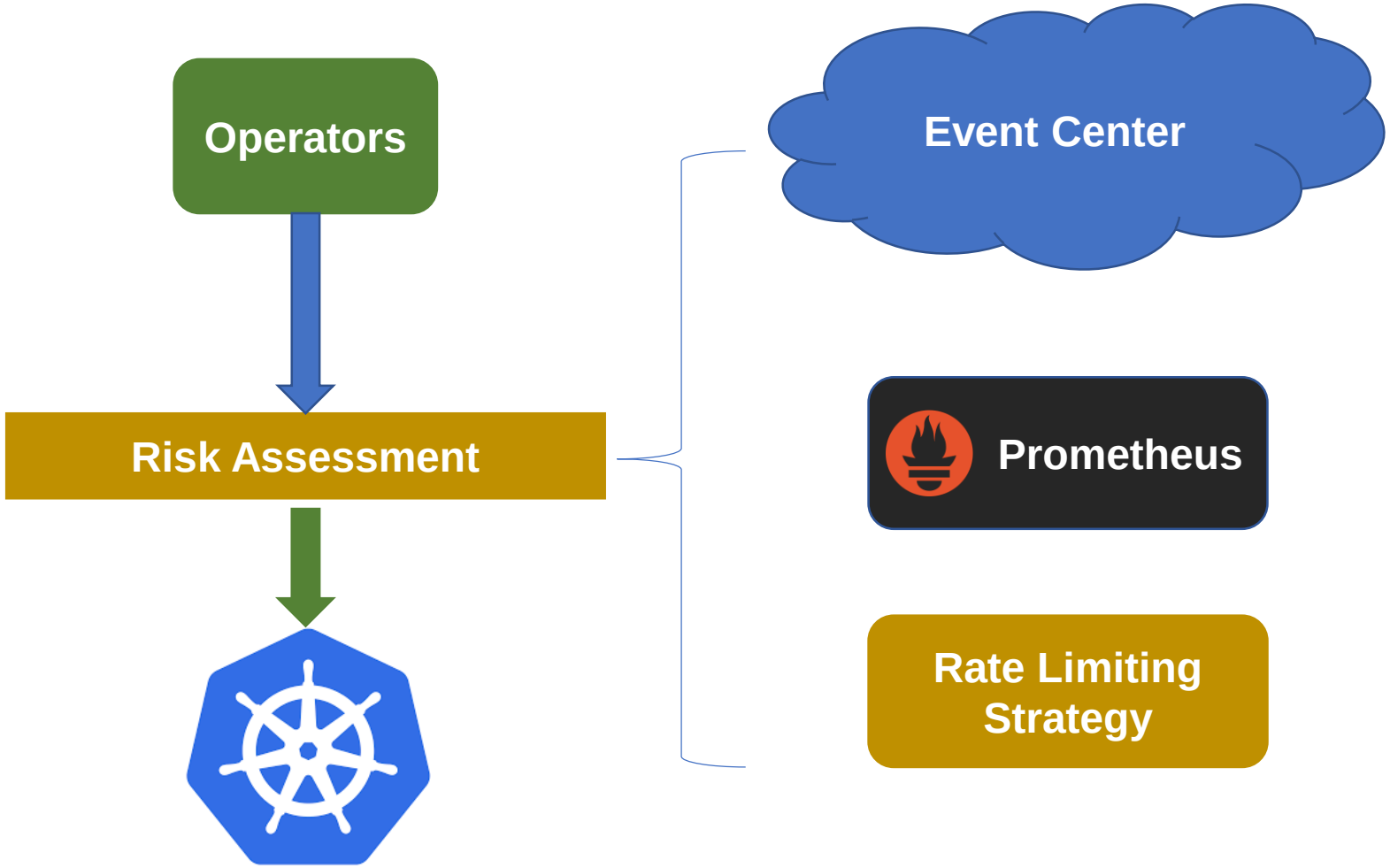


- Event Center
  - Publish & Subscribe Event
- Fault Detection
  - Cluster admin
  - NPD
  - Fault Detector
- Fault Recovery
  - Recovery Operators

Fault Self-recovery



Risk Prevention



- Circuit-Breaker
- Rate Limit

Q & A

KubeCon CloudNativeCon OPEN SOURCE SUMMIT China 2019

Thanks !

扫一扫上面的二维码图案，加我微信