



KubeCon



CloudNativeCon

Europe 2018

# Kubernetes on Supporting \$8 Trillion-Card Payments in China

Xin Zhang & Deyuan Deng

{zhangxin, deyuan}@caicloud.io





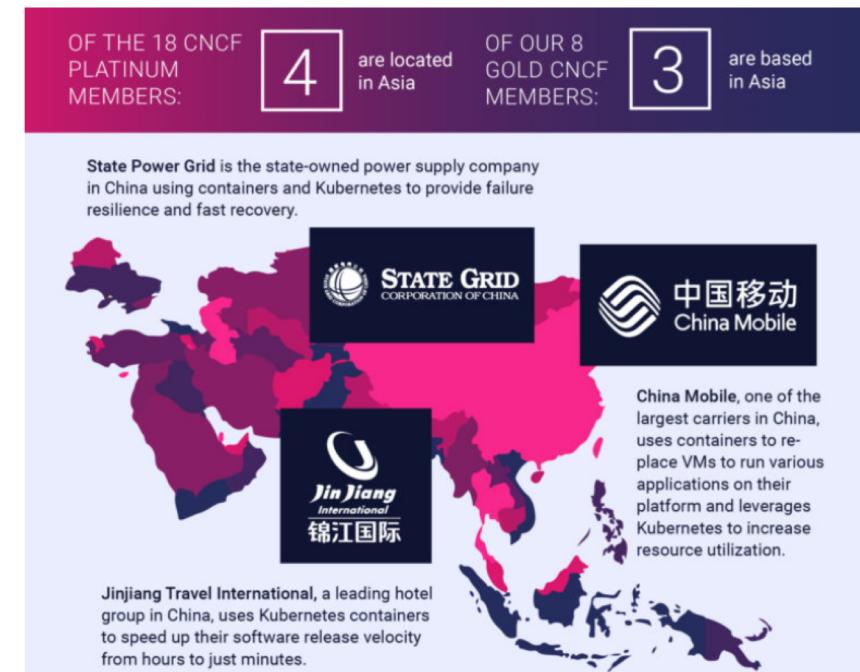
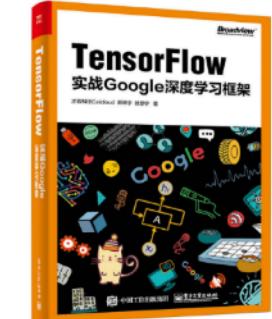
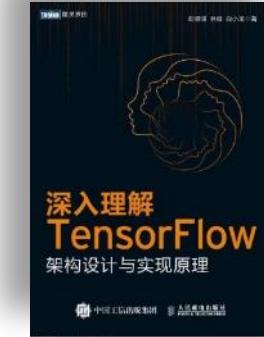
KubeCon



CloudNativeCon

Europe 2018

- Technical Contributors
  - From Kubernetes, to TensorFlow, to Kubeflow
    - [github.com/kubeflow/kubeflow](https://github.com/kubeflow/kubeflow): Kubernetes-based ML stack
- Chinese Community Organizers
  - Kubernetes Chinese community and i18n project
    - [github.com/kubernetes/kubernetes-docs-cn](https://github.com/kubernetes/kubernetes-docs-cn)
  - TensorFlow Chinese community operator
    - [tensorflowers.cn](http://tensorflowers.cn)
  - Kubeflow Chinese community initiator
- Enterprise Products and Solutions
  - *Compass*: Kubernetes distro with value add
  - *Clever*: enterprise-grade kubeflow with AI models





KubeCon

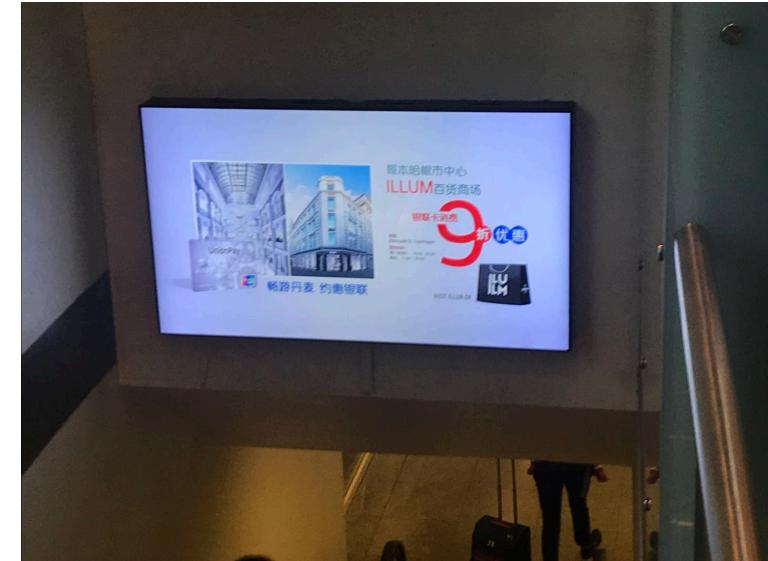
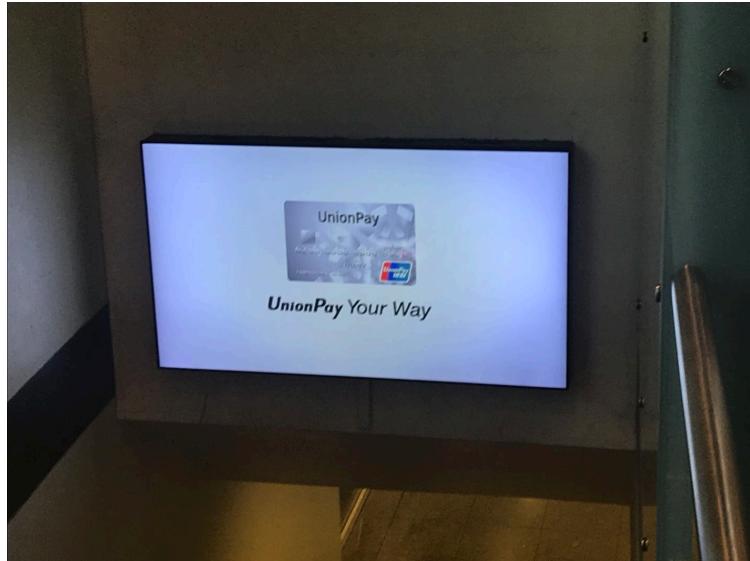


CloudNativeCon

Europe 2018

# The Banking Authority Seeking Help

- The Chinese equivalent of “Visa/Mastercard ” institute
  - Founded in 2002 supervised by Central Bank of China (central government)
  - The only interbank network in China – link all ATMs and banks
- \$14.95 Trillion card payment value in 2017



# Do-or-Die in the Digital Transformation



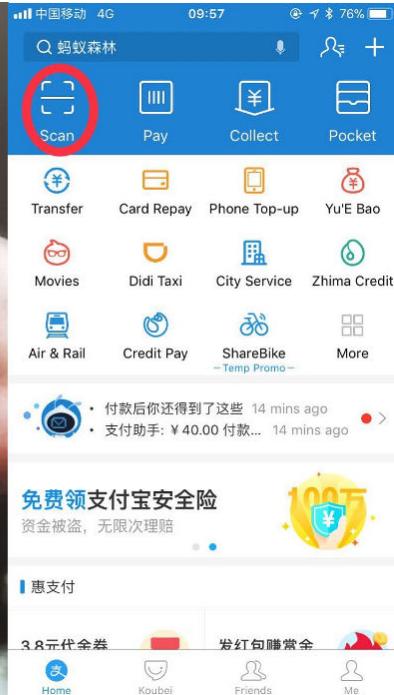
KubeCon



CloudNativeCon

Europe 2018

- When was the last time for a Chinese user using a card to pay for anything?



# Crossing the Chasm



KubeCon



CloudNativeCon

---

Europe 2018

---

- The “Green-field”
  - Agility
  - Scalability
  - Elasticity
  - Availability
  - Automation
- The “Brown-field”
  - VM / OpenStack-based
  - Naked containers with no management layer
  - Human powered

# You Know What's Good at ...



KubeCon



CloudNativeCon

---

Europe 2018

---

- Agility
- Scalability
- Elasticity
- Availability
- Automation

**Shout out your answer!**

# Out-of-Box Panacea?

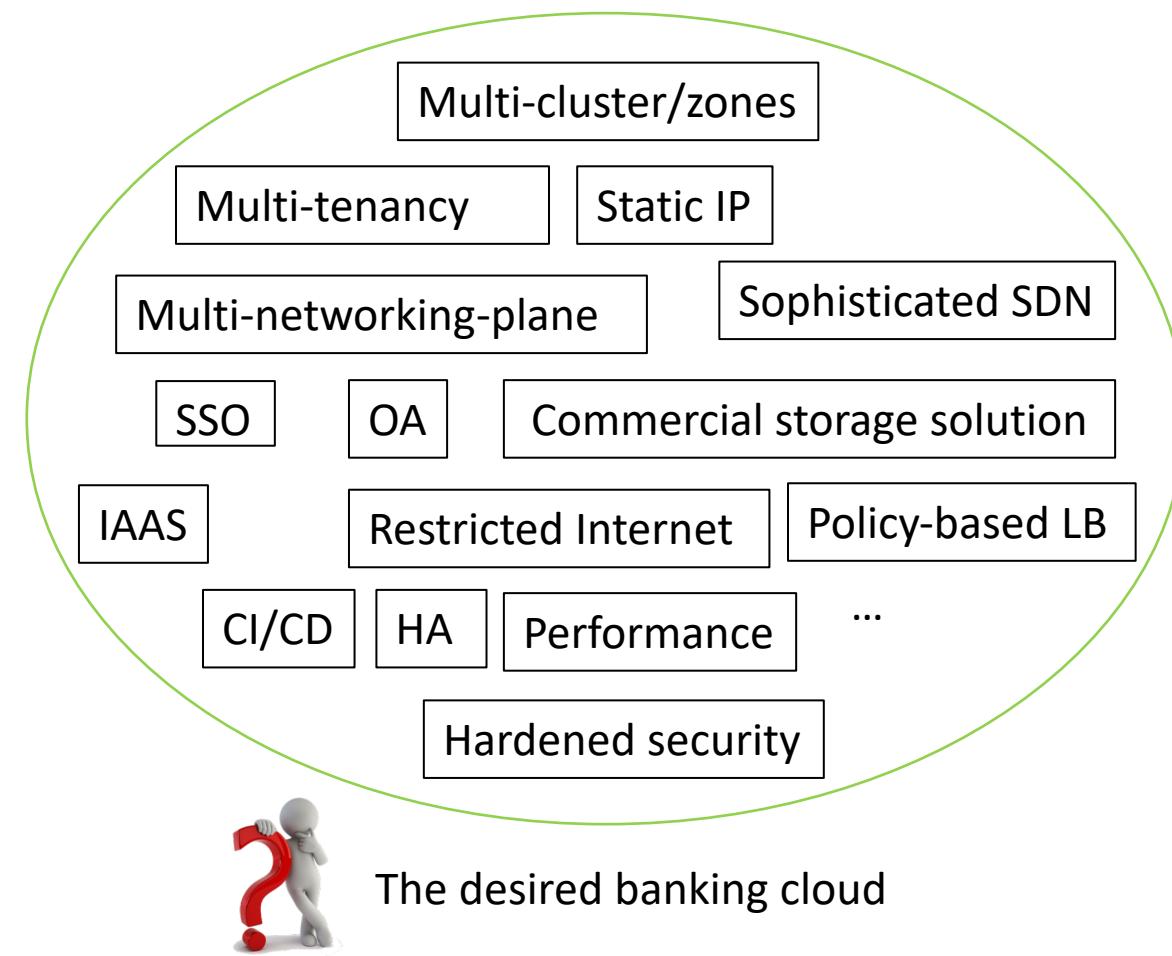


KubeCon

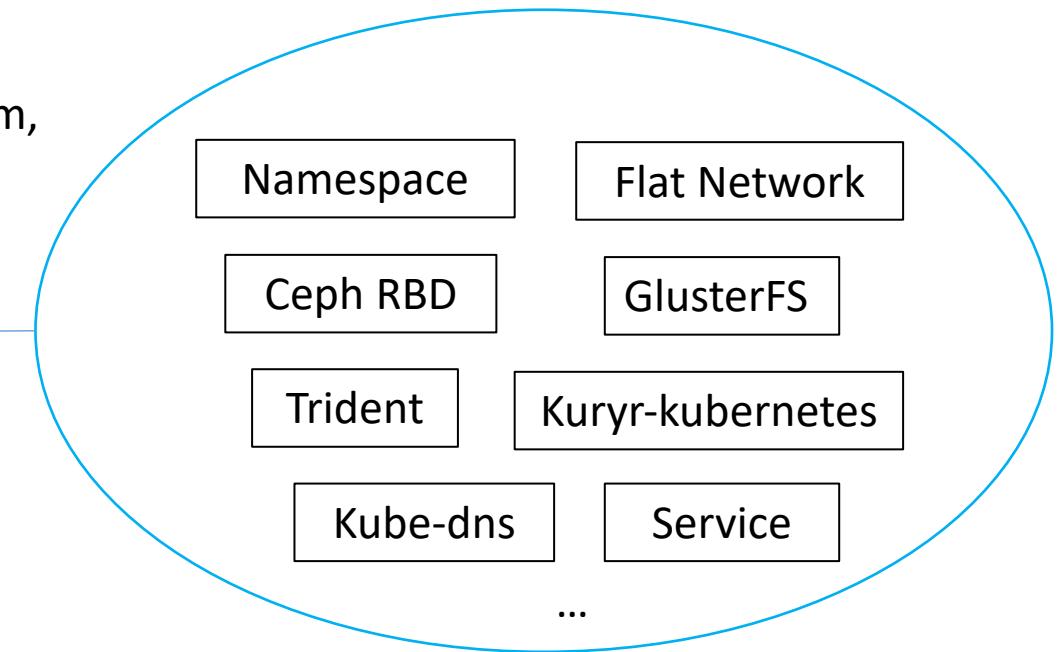


CloudNativeCon

Europe 2018



Not a chasm,  
but a gulp



Off-the-shelf Kubernetes

# The Journey – Before Us

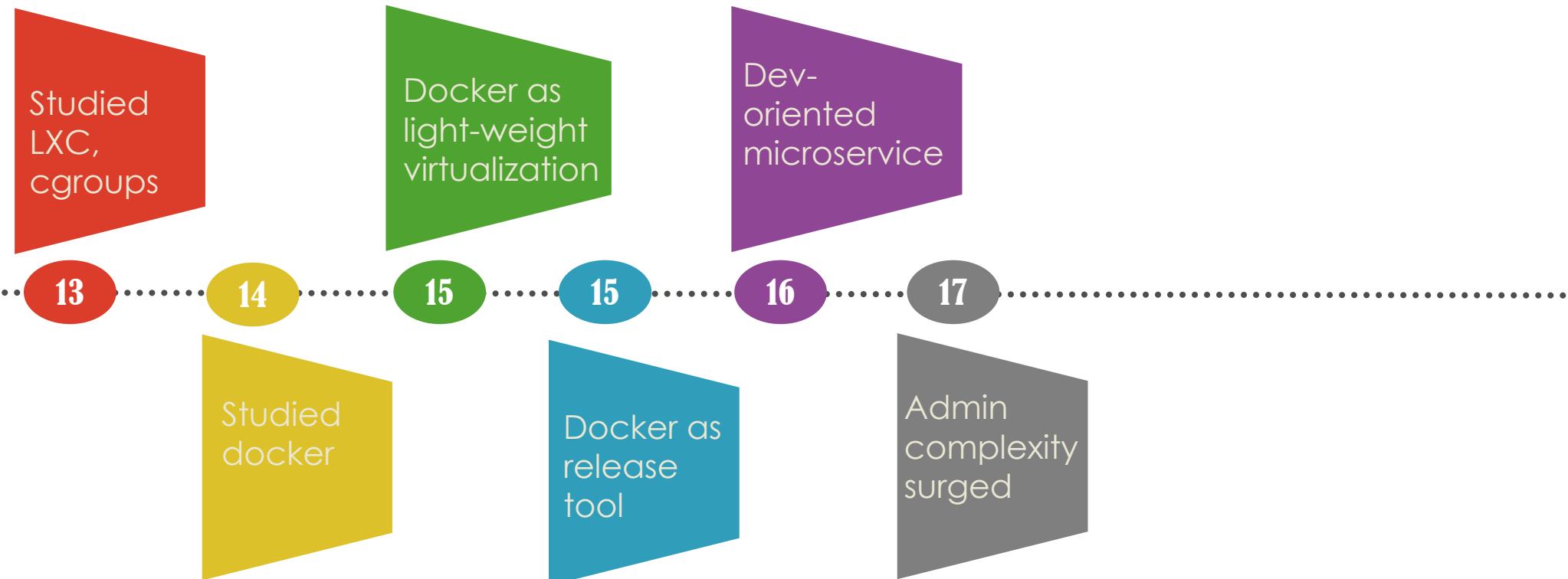


KubeCon



CloudNativeCon

Europe 2018



# The Journey – With Us

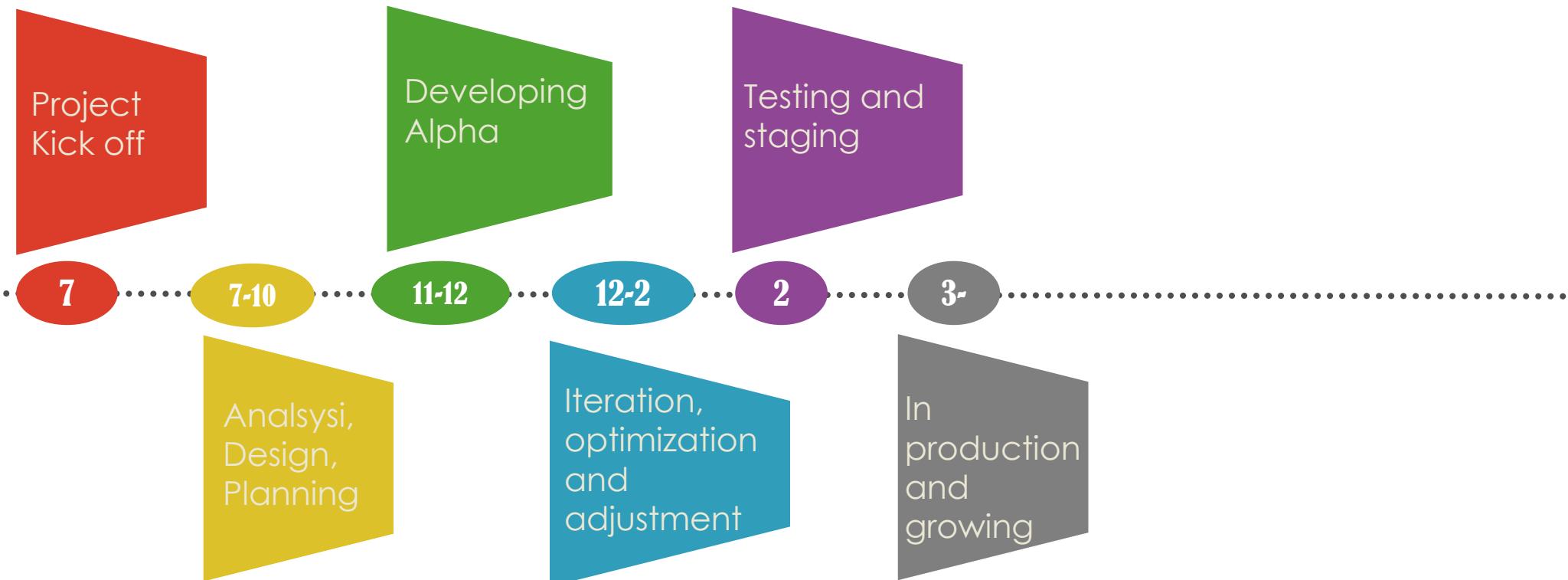


KubeCon



CloudNativeCon

Europe 2018



# Things Delivered So Far



KubeCon



CloudNativeCon

Europe 2018

- Business side
  - Wallet
  - Quick-Pay
  - User Auth



# Things Delivered So Far



KubeCon



CloudNativeCon

Europe 2018

- Technical side
  - SSO with keystone
  - Multi-tenancy
  - HA including multi synchronous image registry
  - multi-networking plane with isolation
  - Configurable static IP
  - network bandwidth and ACL control (white & black listing)
  - Integration with NAS, swift, and other storage solutions
  - Better storage management (online-scaling, disk isolation)
  - Integration with Openstack
  - Richer PaaS functionalities, etc

# Things Delivered So Far



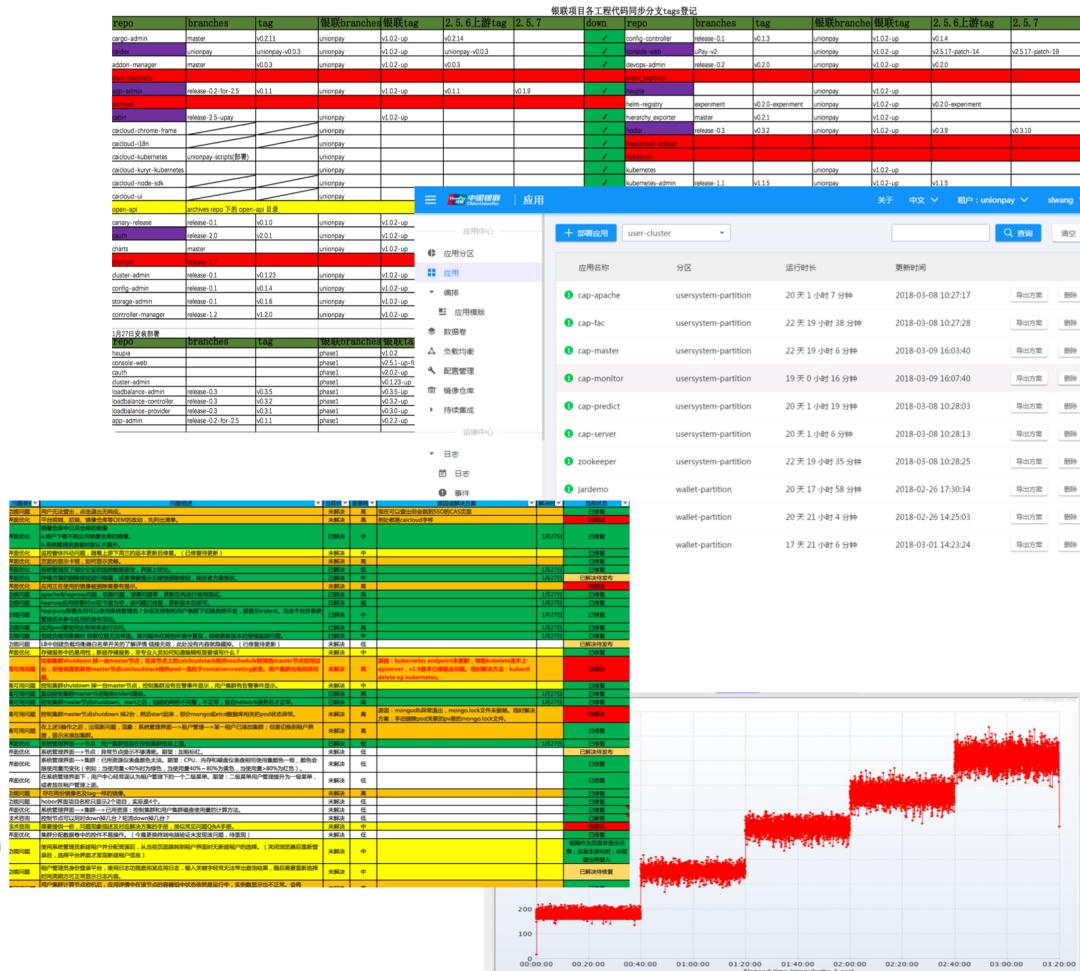
KubeCon



CloudNativeCon

Europe 2018

- Artifacts
  - Code review, release, branching, versioning, naming management
  - 54 code repos
  - Syncing and online upgrade process and tools for 7 environments
  - Micro-service consulting and splitting
  - Comprehensive testing reports
  - CI/CD pipeline and practice
  - Various docs: user guide, admin guide, incident playbook, reports, etc



## 压测结论

- 在 50 个并发时，响应时间就在 200ms ~ 240ms 之间
- 所以，TPS 随着并发数的提高而提高，最开始 TPS 200 左右，是因为请求数比较少。
- 当请求线程数增加时，TPS 最多可以到 750 ~ 850 之间

结论：压测过程中没有失败，和响应时间特别长的请求，性能也比较好。

# Lessons Learned



KubeCon

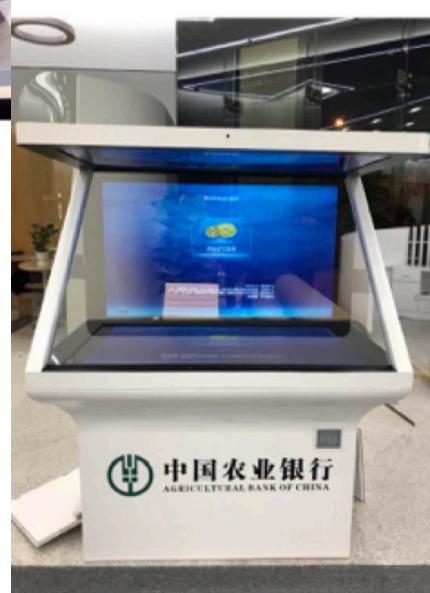
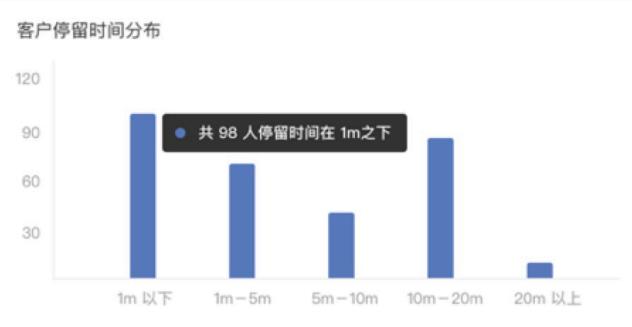
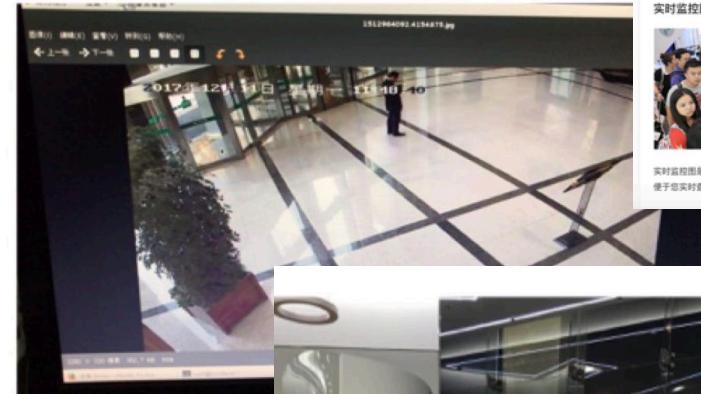


CloudNativeCon

Europe 2018

- Enterprises are complex; welcome to the real-world with noises!
  - Physical constraints, Internet-accessibility constraints, existing IaaS environments
- Decisions are not always driven by technical merits
- Agile iteration with an open mind
- Process and docs are as important as code
- Don't get burned out from the all-nighters!

# Road Ahead



监控状态 / 监控详情

C - 327476

状态  
异常

活动  
大润发商场新品推介会 1

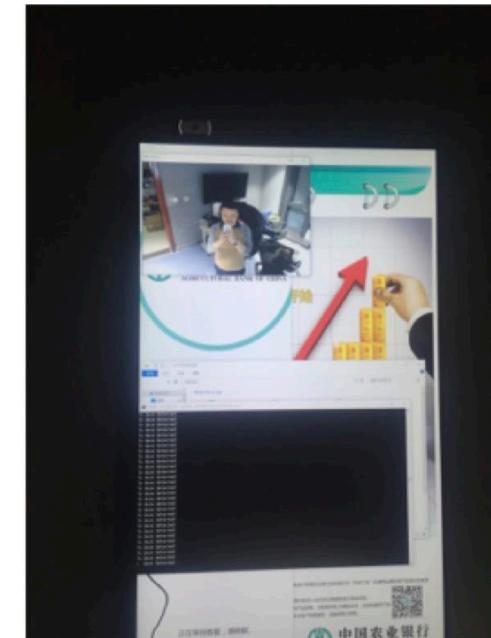
时间  
2017/09/28 09:00 - 2017/09/28 11:00

实时监控图

实时监控图是摄像头本身拍摄的照片，每隔5秒更新一次，便于您实时查看现场情况，点击可查看大图。

已采集头像	性别	年龄	采集时间
	男	25	2017/09/27 09:00:03
	男	25	2017/09/27 09:00:03
	男	25	2017/09/27 09:00:03
	男	25	2017/09/27 09:00:03
	女	25	2017/09/27 09:00:03
	女	25	2017/09/27 09:00:03
	女	25	2017/09/27 09:00:03

共 123 条, 当前显示 20 条每页 | < | < 12 | > | > |



KubeCon



CloudNativeCon

Europe 2018

# Now The Technical Meat



KubeCon



CloudNativeCon

---

Europe 2018

---

- Special thanks to
  - Jiyuan
  - Xiaojian
  - Libin
  - Qingchuan
  - Zhaole
  - Ang
  - Shanmin
  - And other Caicloud folks

# Overview



KubeCon



CloudNativeCon

Europe 2018

## Applications

- Frontend Applications: Apache, HAProxy
- Core Applications: UnionPay Wallet, User Verification System, etc

## Goal of Container Platform

- Migrate applications to containers
- DevOps, Microservices, etc

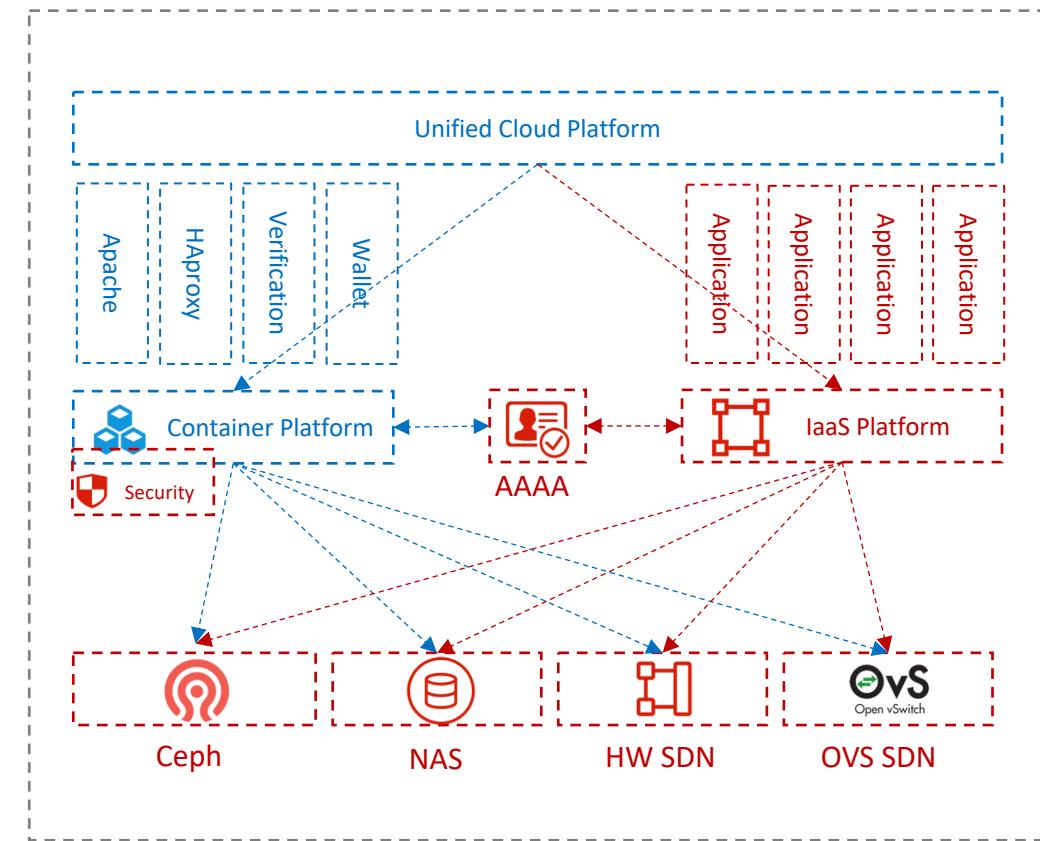
## Network: Integrate with IaaS SDN

## Storage: Integrate with IaaS storage solutions

## AuthN/Z: Seamlessly integrate with internal security systems

## Control Plane: A unified control plane across organization

## Security: Conform to financial security regulation



# Deployment

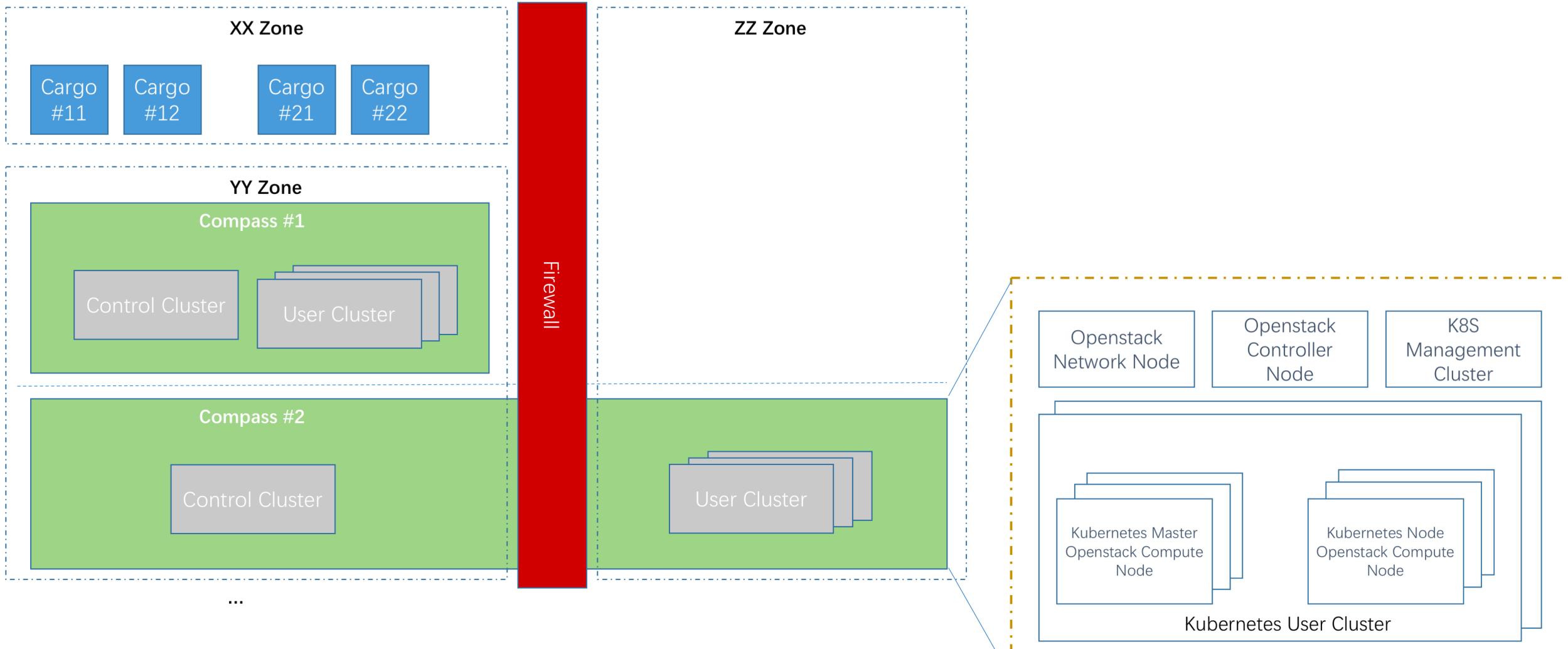


KubeCon



CloudNativeCon

Europe 2018



# SSO - Background



KubeCon



CloudNativeCon

Europe 2018

- Requirement
  - Login once and then everywhere (between kubernetes and openstack)
- Existing environment
  - OA system: organization and employee information
  - Identification service (SSO server): employee ID and password
  - Keystone: integrate with ID service and pulls information from OA daily (midnight)



# SSO - Workflow

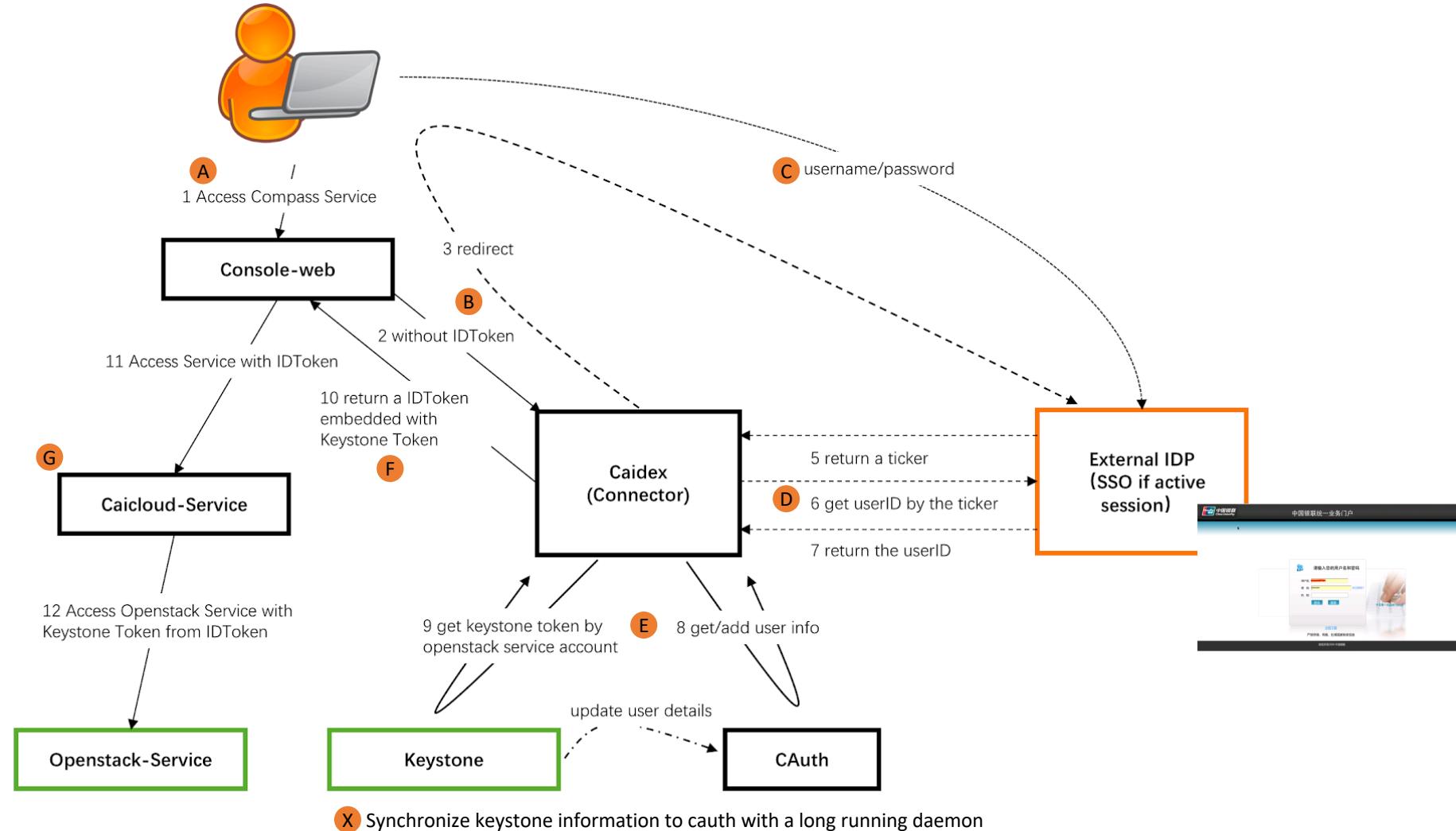


KubeCon



CloudNativeCon

Europe 2018



# Tenant - Concepts



KubeCon



CloudNativeCon

Europe 2018

- System Admin
  - manages the whole platform
  - allocate resources (around different clusters) to different tenants
  - assigns roles to different users, e.g. tenant admin, etc
- Tenant Admin
  - manages a single tenant
  - add/delete groups of users
  - allocate resource permissions per group
- Group/team & User
  - group/team holds permissions of resources
  - users inherit permissions from the groups they are in

# Tenant (System Admin)



KubeCon



CloudNativeCon

Europe 2018

The screenshot shows a web-based tenant management interface. On the left is a sidebar with a dark theme containing the following navigation items:

- compass (highlighted with an orange circle)
- RESOURCE
- Nodes
- Clusters
- Storage
- Network
- Delivery Center
- Registry
- OP CENTER
- Logging
- Monitor
- MANAGEMENT
- Tenants (highlighted with an orange circle)
- Access Control
- Notify Management
- Platform Configurations
- Audit logs

The main content area is titled "QA-Only-Control". It displays the following information:

- Identifier:** qa-only-control
- Description:** QA 通用测试, 只有控制集群
- CreateTime:** 2018-04-28 09:39:43

Below this, there are tabs for "Resource Quotas" (highlighted with an orange circle), "Member", "App Stats", and "Partition Stats". Under "Resource Quotas", there are sections for "Clusters", "CPU", and "Memory".

**Clusters:** multiple-tenant-c... > Cluster-QA-NoHA >

**CPU:**

Request	Limit
4 Core	20 Core

**Memory:**

Request	Limit
8 GiB	20 GiB

**Volumes:**

Storage class	Volumes Numbers	Volumes Total Capacity
heketi-storageclass	2 / 5 个	10 / 30 GiB

**Load Balancing:**

Name	L7 layer protocol	L4 layer protocol	Port range of L4 layer protocol
apiserver	已开启	已开启	20000 – 21000



KubeCon



CloudNativeCon

Europe 2018

# Tenant (Tenant Admins)

Groups

About Manual English Tenant: QA-Only-Control admin

Name	Description	Member	Actions
Default Team	Default team for all members in a Tenant	AllMember	<button>Modify</button> <button>⋮</button>
SREs	SREs have permissions of Loadbalancer, OpsCenter, etc	0 people	<button>Modify</button> <button>⋮</button>
Developers	Developers have permissions of AppCenter, etc	0 people	<button>Modify</button> <button>⋮</button>
Resource	Resource Team has permissions of allocating resources, etc	0 people	<button>Modify</button> <button>⋮</button>
Support	Support team has permissions of CI/CD, etc	0 people	<button>Modify</button> <button>⋮</button>

Each line is a group with a set of users and permissions.

Total 5 1/1 Every page shows 10 ▶ 1 ▷

APPLICATION

- Partitions
- Applications
- App Store
- Orchestration
- Volumes
- Load balancing
- Configs
- Delivery Center
- Registry
- DevOps
- OP CENTER
- Logging
- Monitor
- Alerting
- MANAGEMENT
- Access Control
- Notify Management
- Audit logs



KubeCon



CloudNativeCon

Europe 2018

# Tenant (Tenant Admins)

Groups / SREs

About Manual English Tenant: QA-Only-Control admin

SREs

Create... 2018-04-28

Descri... 运维部门的业务运维人员，拥有负载均衡、运维中心等权限

Num of Members 0

Resource to allocate permissions, each resource is either an external resource, or native / CRDs in Kubernetes.

Members Partitions Templates Loadbalancers Configs Operations DevOps Cargo Clever

Clusters Basic Role User Modify After given the basic role, members of the team share permissions of Partitions

multiple-tenant-c... > Guest + User + Owner +

qatest

Basic Role	User	Owner
Guest	qatest	



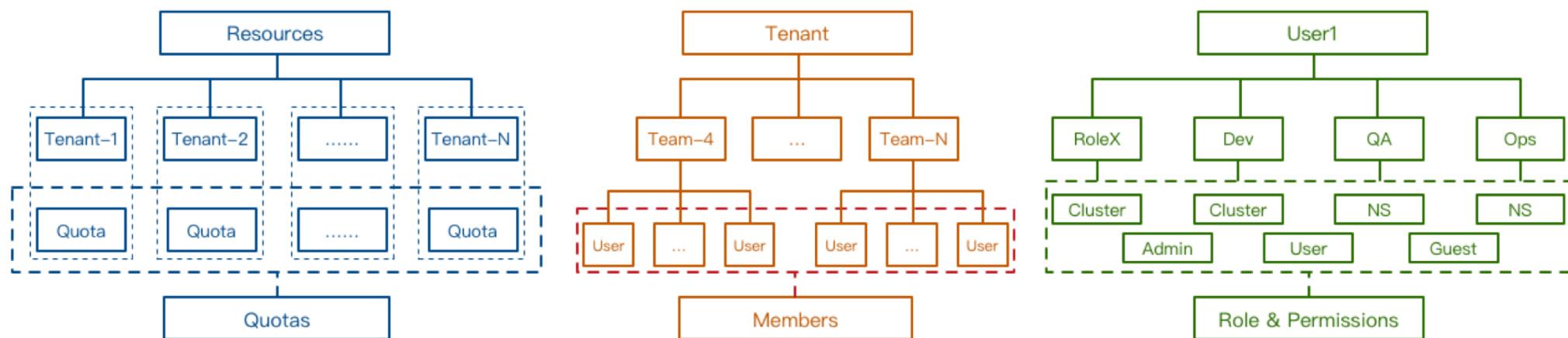
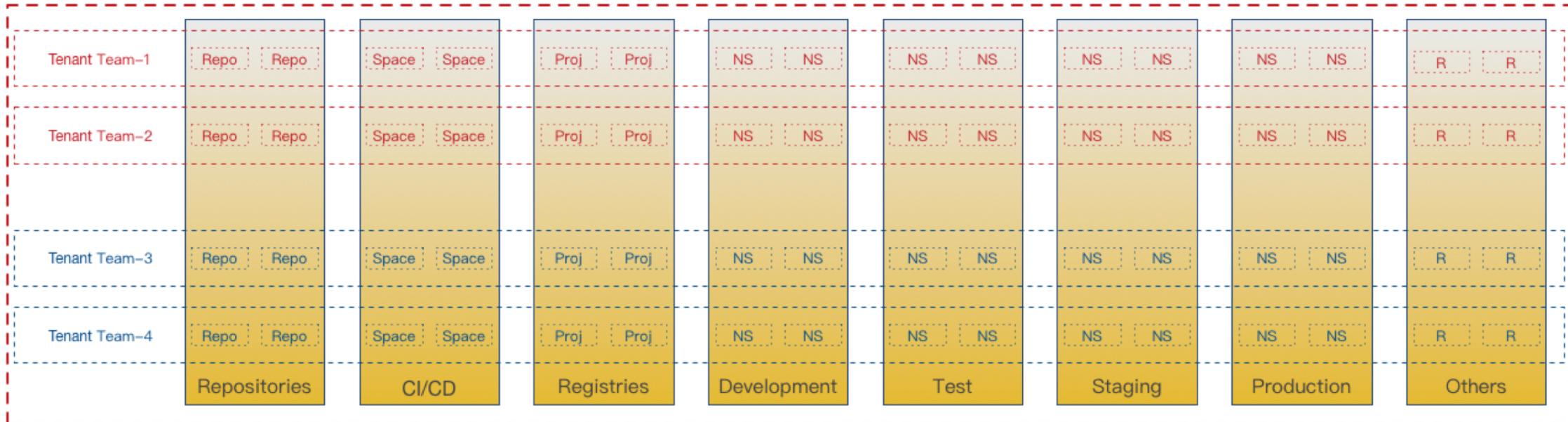
KubeCon



CloudNativeCon

Europe 2018

# Tenant - Summary



# Tenant - OpenStack Mapping



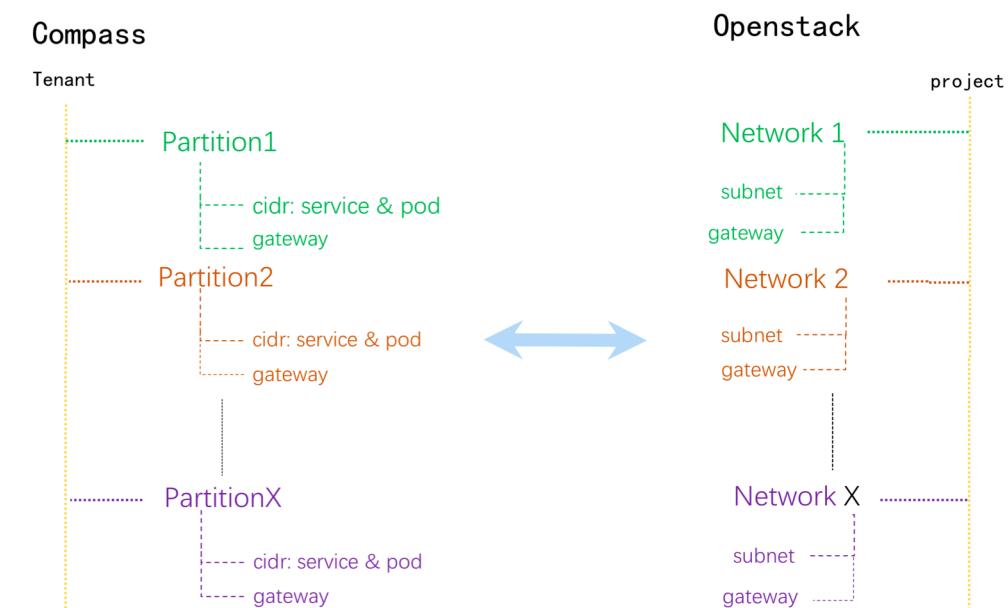
KubeCon



CloudNativeCon

Europe 2018

- 1:1 mapping between compass Tenant and openstack Project
- 1:1 mapping between kubernetes Namespace and openstack Network
- 1:1 mapping between kubernetes Pod and openstack Port
- 1:1 mapping between kubernetes Service VIP and openstack LBaaS VIP
- Each namespace has two openstack subnets: PodCIDR and ServiceCIDR
- Each namespace has an openstack gateway



# Tenant - with OpenStack



KubeCon



CloudNativeCon

Europe 2018

The screenshot shows a cloud management interface for China UnionPay. The left sidebar includes sections for 资源中心 (Resource Center) and 管理中心 (Management Center). In the Management Center, Tenant Management is selected. The main content area displays a "Create Tenant" form with fields for Name and Description. Below it, a modal window titled "Create Tenant requires choosing an OpenStack Tenant" lists several OpenStack projects:

project	description
docker-k8s-project3	k8s-project3
Default	new-test
Default	test2
Default	testl7
Default	123123
zvnDomain	zvnProject2

A tooltip for the "k8s-project3" row provides detailed information:

**k8s-project3**  
project信息  
description: 暂无  
domain\_name: docker-k8s2  
project\_id:  
118e9638341c42dca67dc26f590186cd

# Tenant - with OpenStack



KubeCon



CloudNativeCon

Europe 2018

新增分区  
Create New Partition, a.k.a Namespace

分区名称:

所属集群: user-cluster-up

集群资源:

CPU 请求 2 / 2 Core

CPU 上限 6 / 10 Core

内存请求 4 / 4 GiB

内存上限 8 / 8 GiB

CPU 配额:

请求: Core

上限: Core

内存配额:

请求: GiB

上限: GiB

I3policy\_id: 请选择I3policy

podCIDR: for example: 192.168.1.0/24

serviceCIDR: for example: 192.168.1.0/24

Network Option

# Tenant - with OpenStack

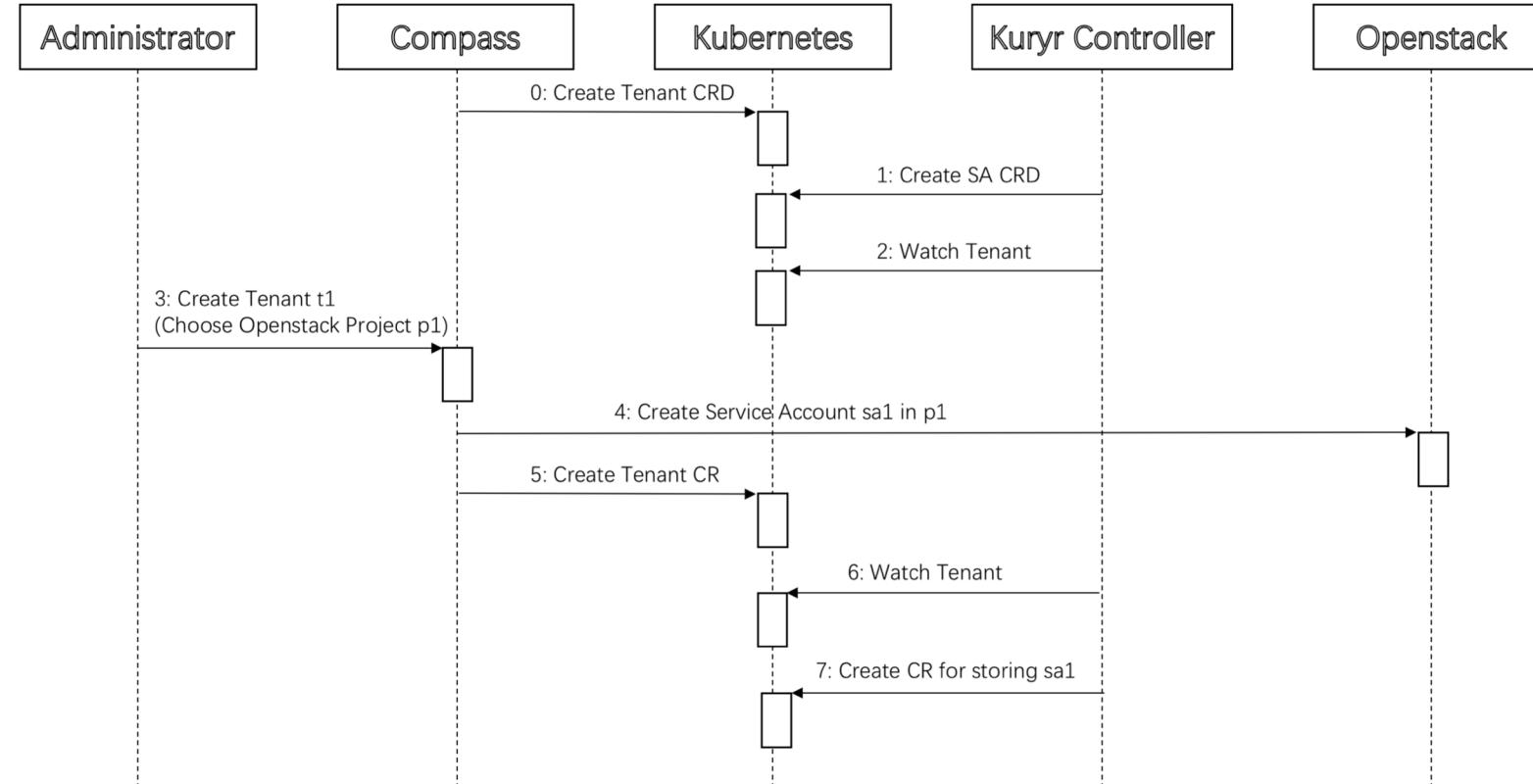


KubeCon



CloudNativeCon

Europe 2018



- OpenStack Service Account CRD allows tenant in Kubernetes to access OpenStack Resources
- Tenant and quota is represented as CRD to natively integrates with Kubernetes



KubeCon

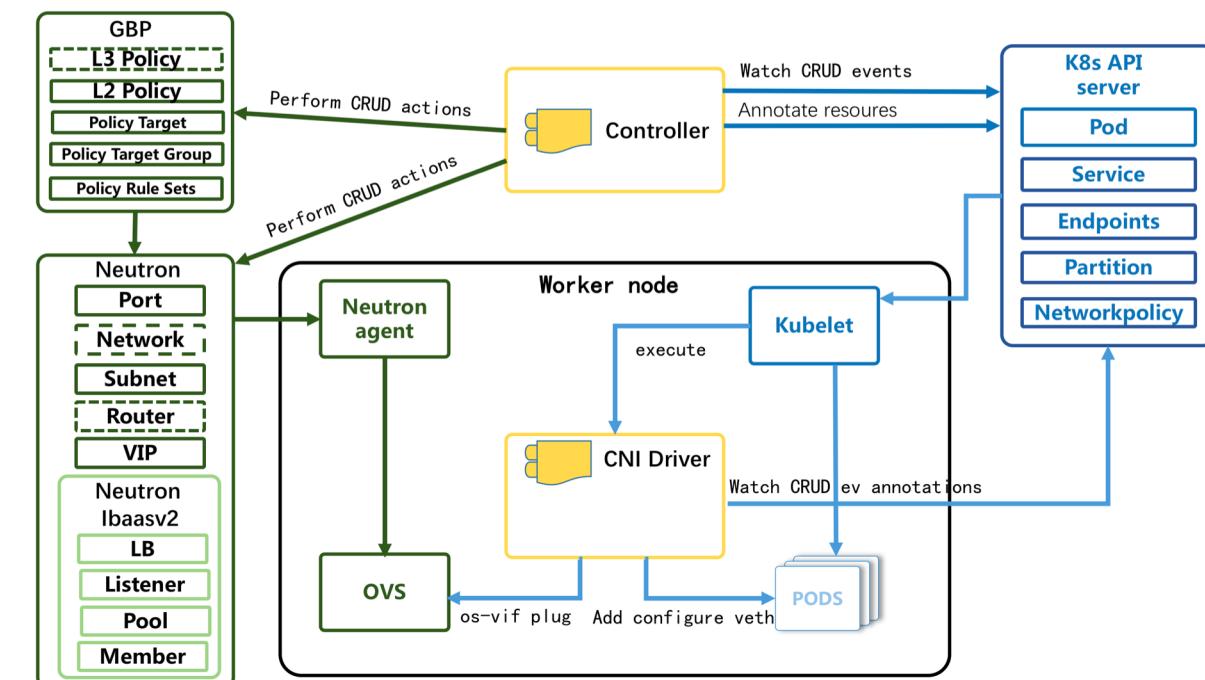


CloudNativeCon

Europe 2018

# Network - Overview

- Requirements
  - Pod should have ‘real’ IP, no NAT
  - Configurable traffic shaping and isolation across tenant
  - Multiple network planes (management, control, business, storage)
- Based on **kuryr-kubernetes**, with modification:
  - Support multi-tenant
  - Support GBP
- Components
  - Kubernetes Controller
    - Translate to OpenStack Neutron Model
    - Pass information to CNI Driver via Annotation
  - CNI Driver
    - Bind Kubernetes Pod with Neutron Port



# Network - Multi-planes



KubeCon



CloudNativeCon

Europe 2018

Network Plane	Description
Management (br-mgm)	IaaS and container platform portal
Control (br-ctl)	Container platform system components
Storage (br-storage)	Container Image Pull/Push
Business (br-prv)	All network traffic for applications

# Network - Multi-planes



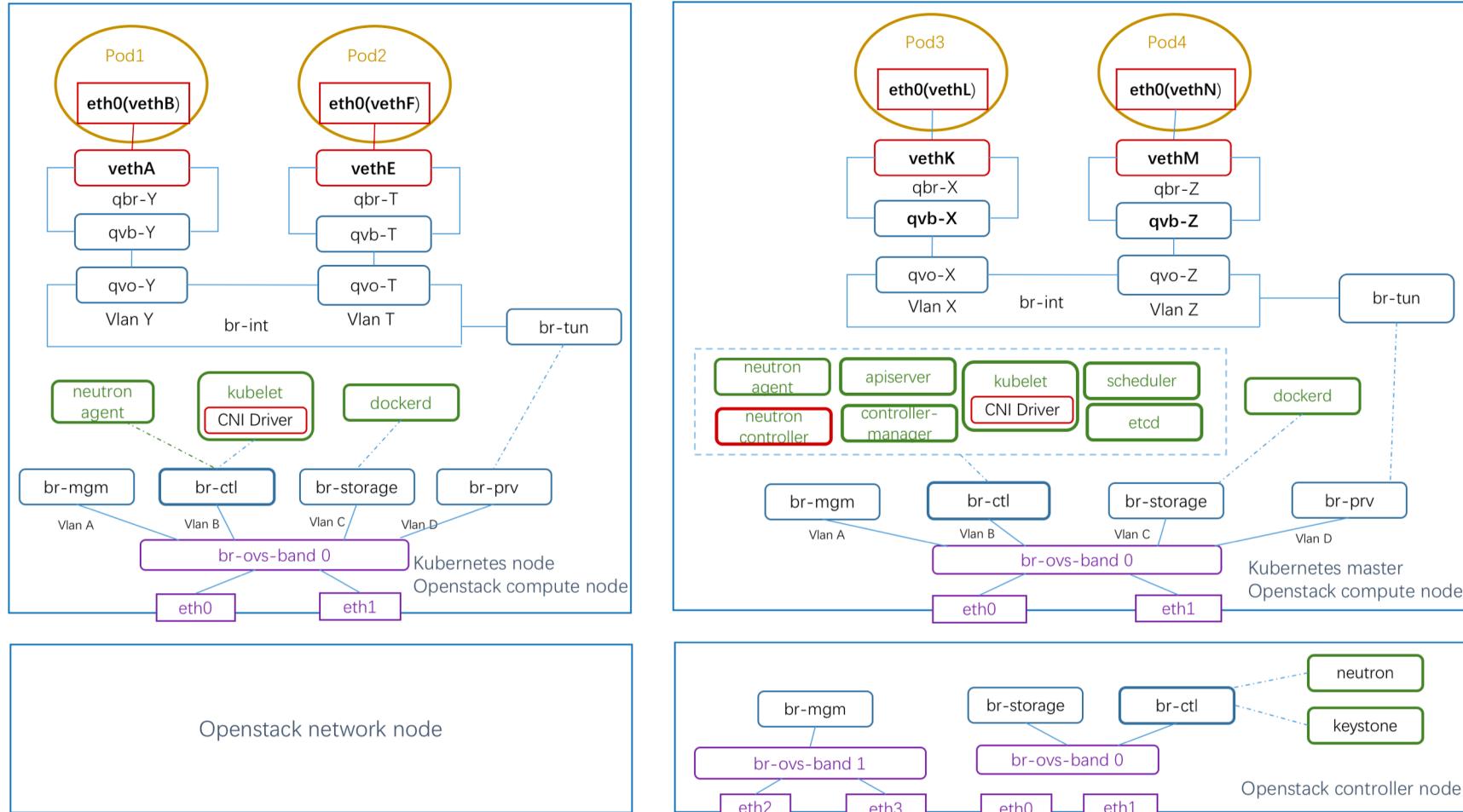
KubeCon



CloudNativeCon

Europe 2018

Openstack ovs-based vxlan





KubeCon



CloudNativeCon

Europe 2018

# Network - Service

- A new type of service: Tenant Service
- Each tenant has its own tenant service CIDR
- Tenant service virtual IP comes from LBaaS (Virtual IP)

Tenant	Neutron	Service	API Server
(System Tenant)	--	Normal Service	--service-cluster-ip-range 10.0.0.0/16
Tenant1 Network	Service CIDR 10.20.0.0/16	Tenant Service	--
Tenant2 Network	Service CIDR 10.20.0.0/16	Tenant Service	--
Tenant3 Network	Service CIDR 10.40.0.0/16	Tenant Service	--

VIP

Protocol, Port, Members (Endpoints)

# Network - Service



KubeCon



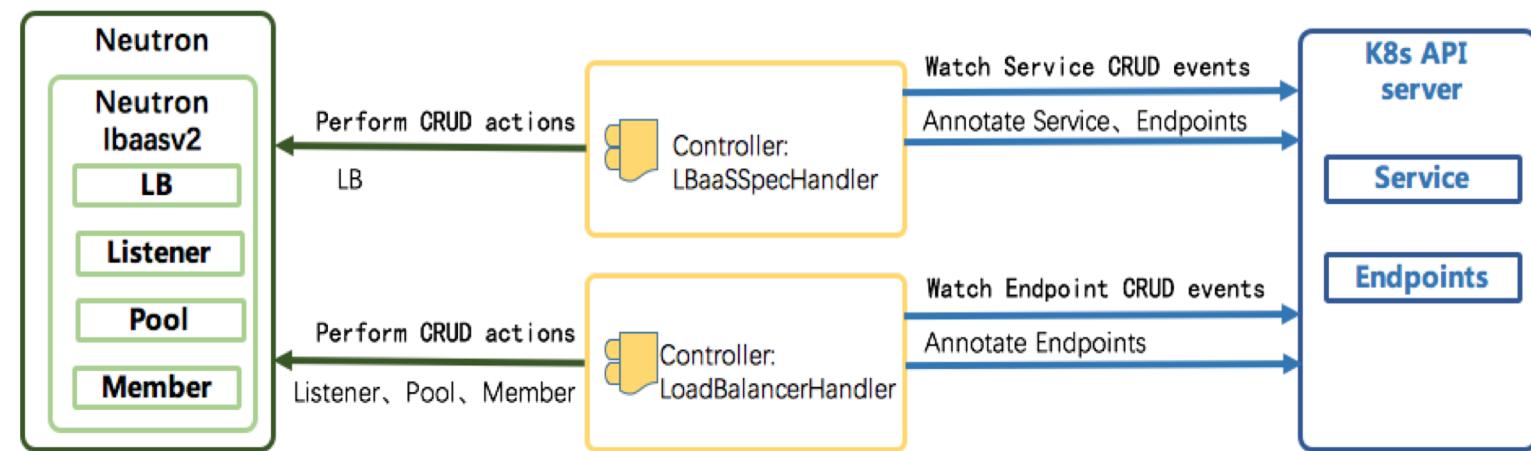
CloudNativeCon

Europe 2018

Neutron LBaaS	Kubernetes
Loadbalancer VIP	Service Cluster IP
Loadbalancer FIP	Service External IP
Protocol and Port of Listener	Protocol and Port of Service
Loadbalance Method of Member Pool	Loadbalance Method of Service/Endpoints
Members (IP, Port)	Endpoints (IP, Port)

Mapping from Neutron Concepts to  
Kubernetes Concepts

Implement the mapping using two  
Controllers



# Network - DNS



KubeCon



CloudNativeCon

Europe 2018

- Existing Kubernetes setup runs DNS as a regular Pod, this doesn't work:
  - DNS Pod needs to access API server, which runs in different plane
    - br-prv vs br-ctl
  - but if we run DNS Pod in control plane, normal Pods can't access
    - br-ctl vs br-prv
  - There is no 'per-tenant DNS' in UnionPay DNS

# Network - DNS



KubeCon

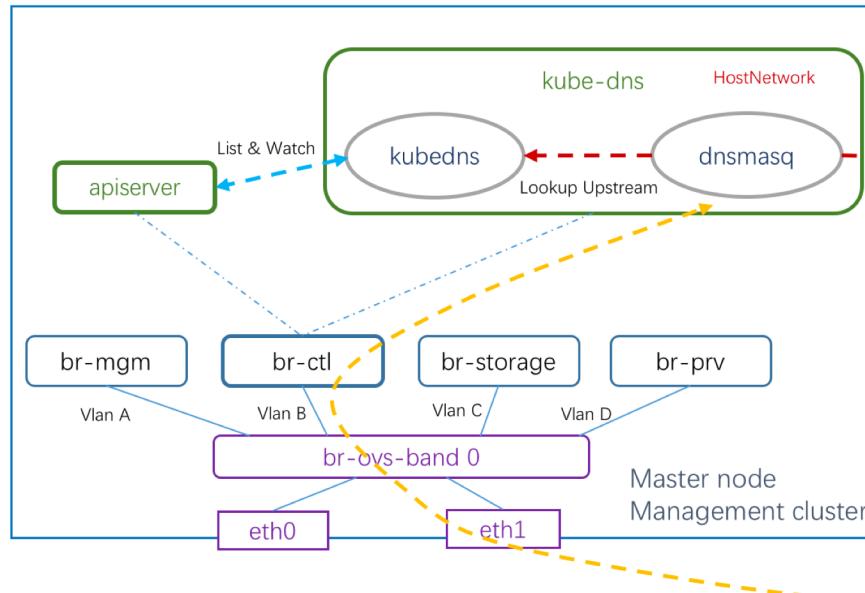


CloudNativeCon

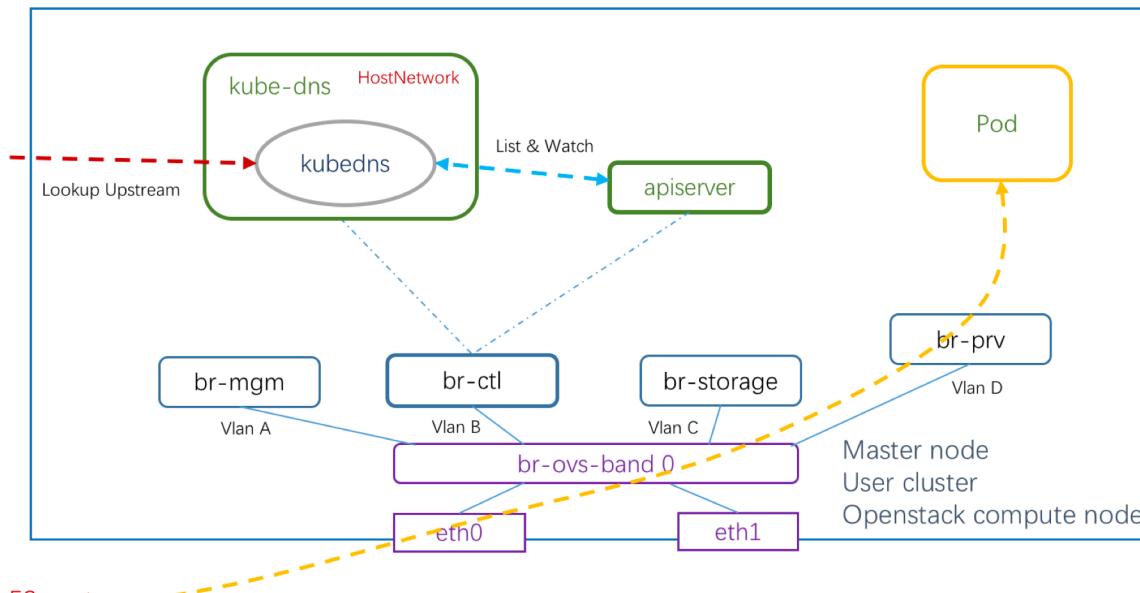
Europe 2018

- For management cluster, run ‘kube-dns’ in HostNetwork, with container: kubedns, dnsmasq and healthz
- For user cluster, run ‘kube-dns’ in HostNetwork, with container: kubedns and healthz
- dnsmasq is used for DNS query

Management Cluster



User Cluster



open 53 port



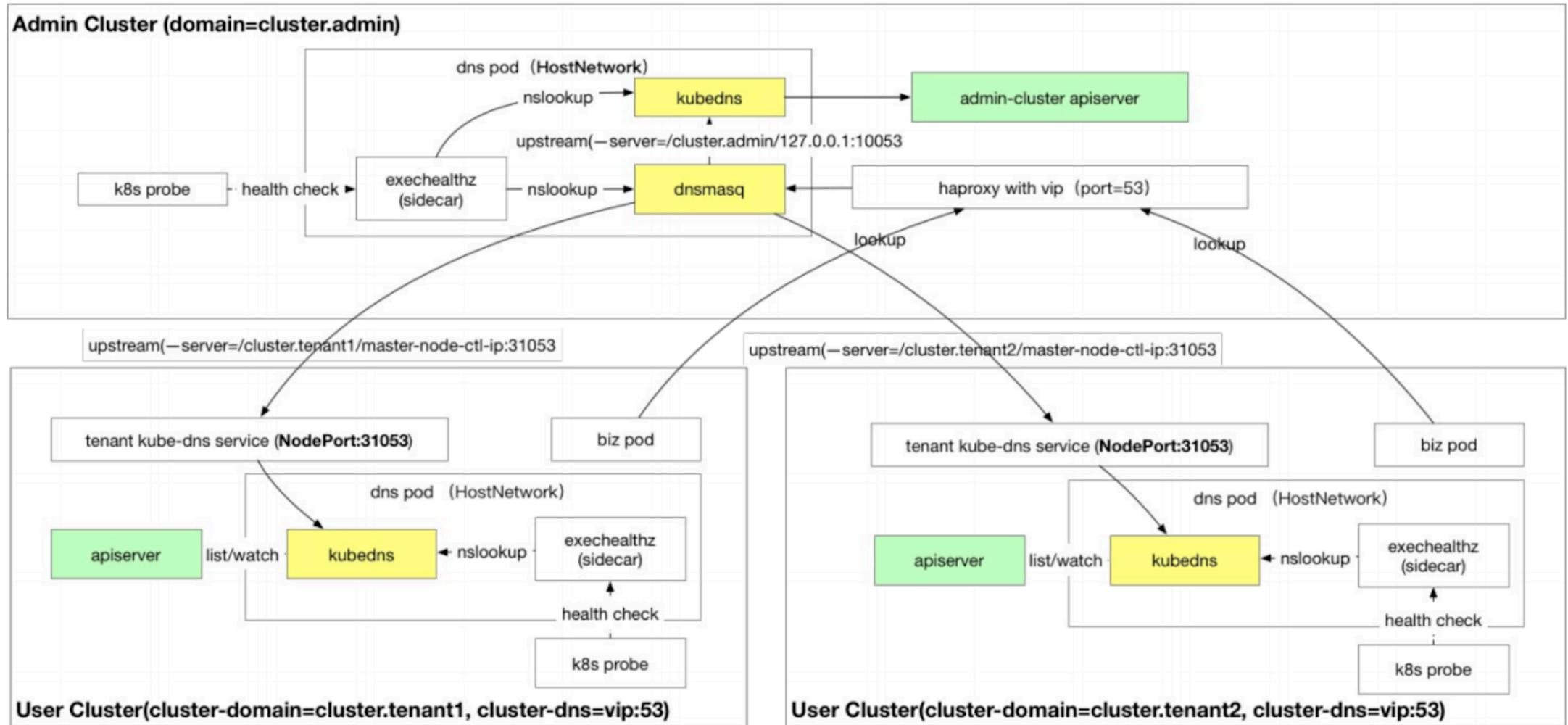
KubeCon



CloudNativeCon

Europe 2018

# Network - DNS Deployment



# Storage - NAS



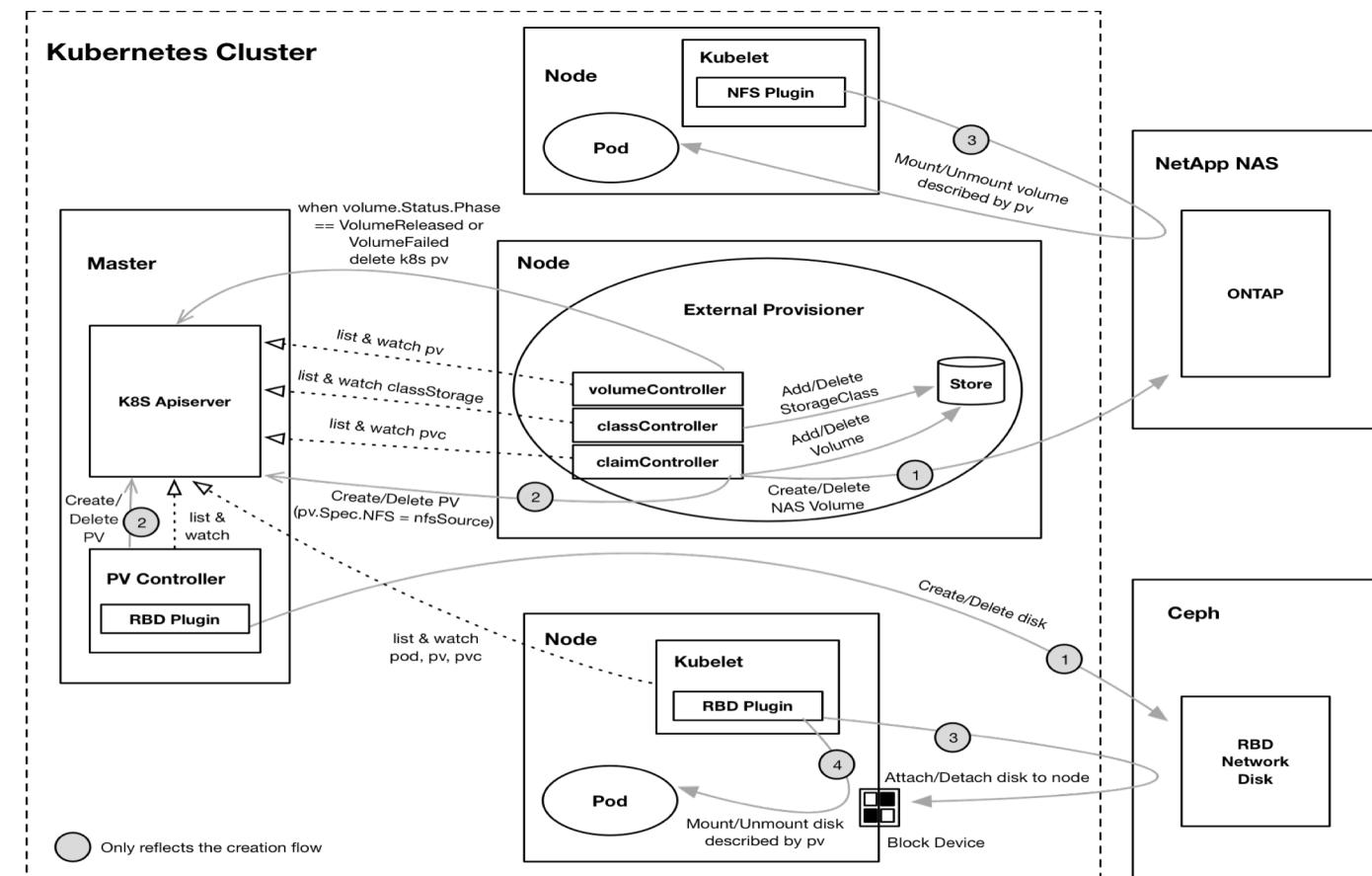
KubeCon



CloudNativeCon

Europe 2018

- No out-of-box solution in Kubernetes
- but enterprise users care about data
- Depends on existing environment
  - NetApp NAS: use trident
  - Ceph



# Multi-registries



KubeCon



CloudNativeCon

Europe 2018

The screenshot shows a 'Container Registry' interface with a sidebar menu. The 'Registry' option is selected in the sidebar. The main area displays a table with two rows:

Name	Address	Public Projs	Private Projs	Public Images	Private Images
Default	https://cargo-multiple-tenant-current.caicloudpr...	2	1	148	0
multiple-tenant.caicloudprivatetest	https://multiple-tenant.caicloudprivatetest.com	3	3	148	3

An orange oval highlights the table area.

The screenshot shows a '镜像仓库' (Image Repository) interface with a 'Sync Image' dialog box. The dialog box has the title '同步镜像 Sync Image' and a dropdown menu with the placeholder '请选择 Select another registry'. In the background, there is a list of images:

- busybox cargo.caicloudprivatetest.com
- hello-world cargo.caicloudprivatetest.com
- mongodb cargo.caicloudprivatetest.com/library/mongodb
- mysql cargo.caicloudprivatetest.com/library/mysql
- redis cargo.caicloudprivatetest.com/library/redis

A modal window titled 'mysql 版本信息' is also visible on the right side.



KubeCon

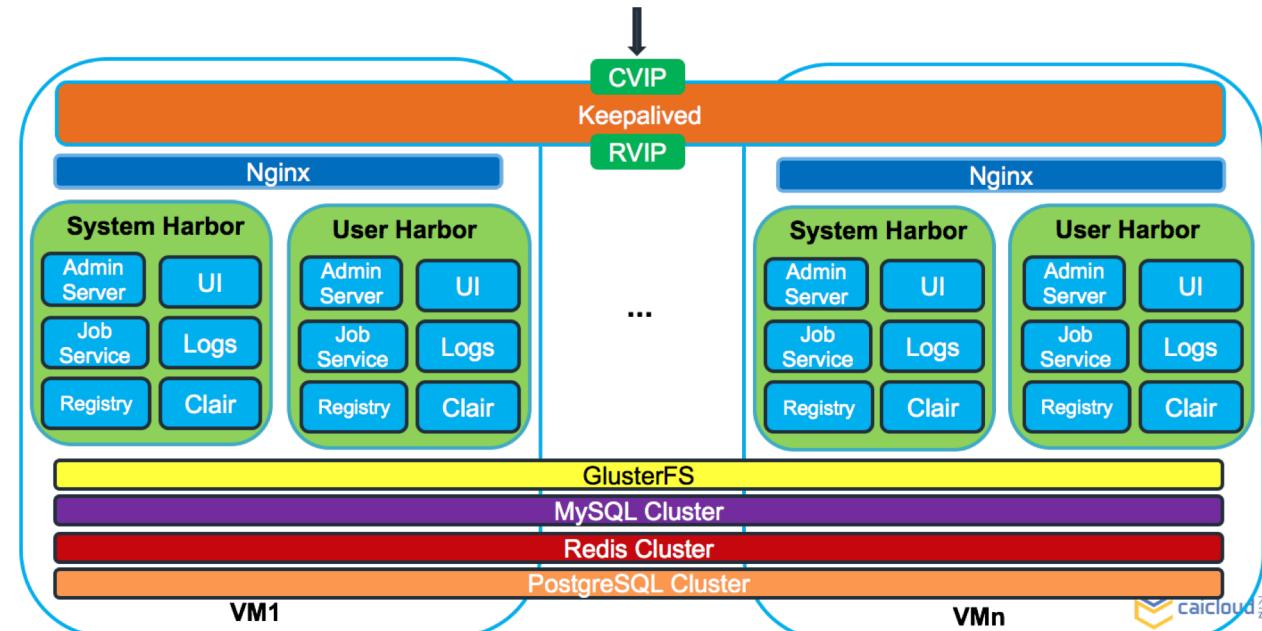


CloudNativeCon

Europe 2018

# Registry HA

- A HA registry runs on multiple VMs
- A VM can run multiple registry instance
  - traffic goes through different NIC
  - efficient use of resources
- Multiple Harbors
  - System Harbor vs User Harbor
  - System Harbor: for system component
  - System Harbor: for user applications





KubeCon



CloudNativeCon

Europe 2018

# Thanks !

Xin Zhang & Deyuan Deng  
[{zhangxin, deyuan}@caicloud.io](mailto:{zhangxin,deyuan}@caicloud.io)

