



KubeCon



CloudNativeCon

Europe 2018

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

Xin Zhang & Deyuan Deng  
{zhangxin, deyuan}@caicloud.io



## • Technical Contributors

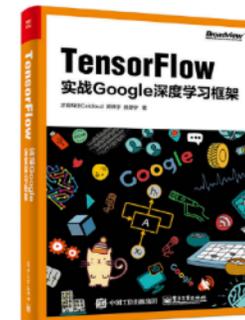
- From Kubernetes, to TensorFlow, to Kubeflow
  - [github.com/kubeflow/kubeflow](https://github.com/kubeflow/kubeflow): Kubernetes-based ML stack

## • 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
  - [tensorflow.cn](https://www.tensorflow.cn)
- Kubeflow Chinese community initiator

## • Enterprise Products and Solutions

- *Compass*: Kubernetes distro with value add
- *Clever*: enterprise-grade kubeflow with AI models



OF THE 18 CNCF PLATINUM MEMBERS: **4** are located in Asia

OF OUR 8 GOLD CNCF MEMBERS: **3** are based in Asia

State Power Grid is the state-owned power supply company in China using containers and Kubernetes to provide failure resilience and fast recovery.

中国移动  
China Mobile

China Mobile, one of the largest carriers in China, uses containers to replace VMs to run various applications on their platform and leverages Kubernetes to increase resource utilization.

Jin Jiang International  
锦江国际

Jinjiang Travel International, a leading hotel group in China, uses Kubernetes containers to speed up their software release velocity from hours to just minutes.

# The Banking Authority Seeking Help



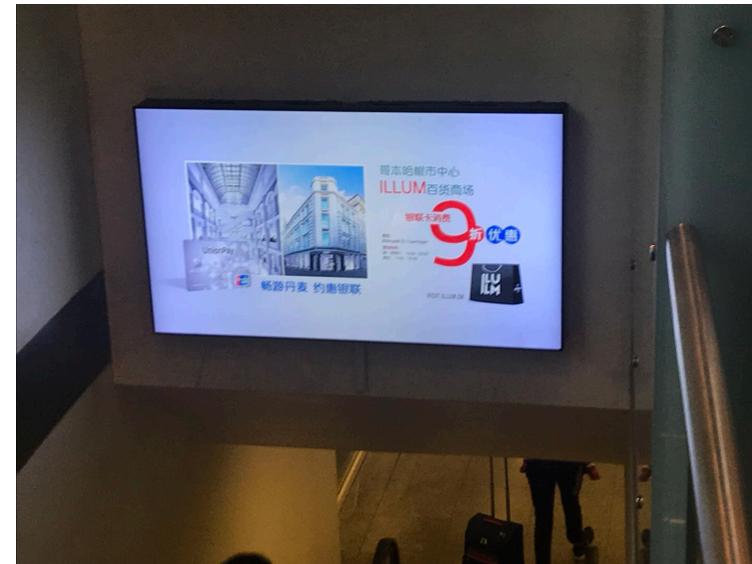
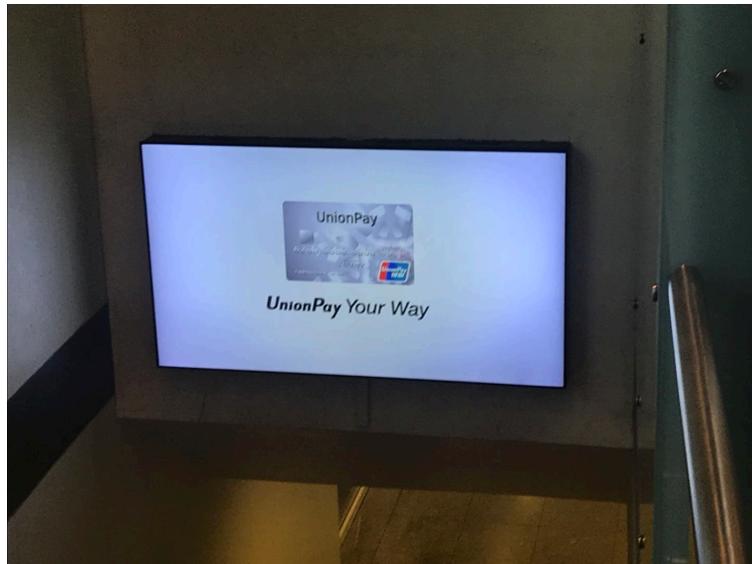
KubeCon



CloudNativeCon

Europe 2018

- 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?

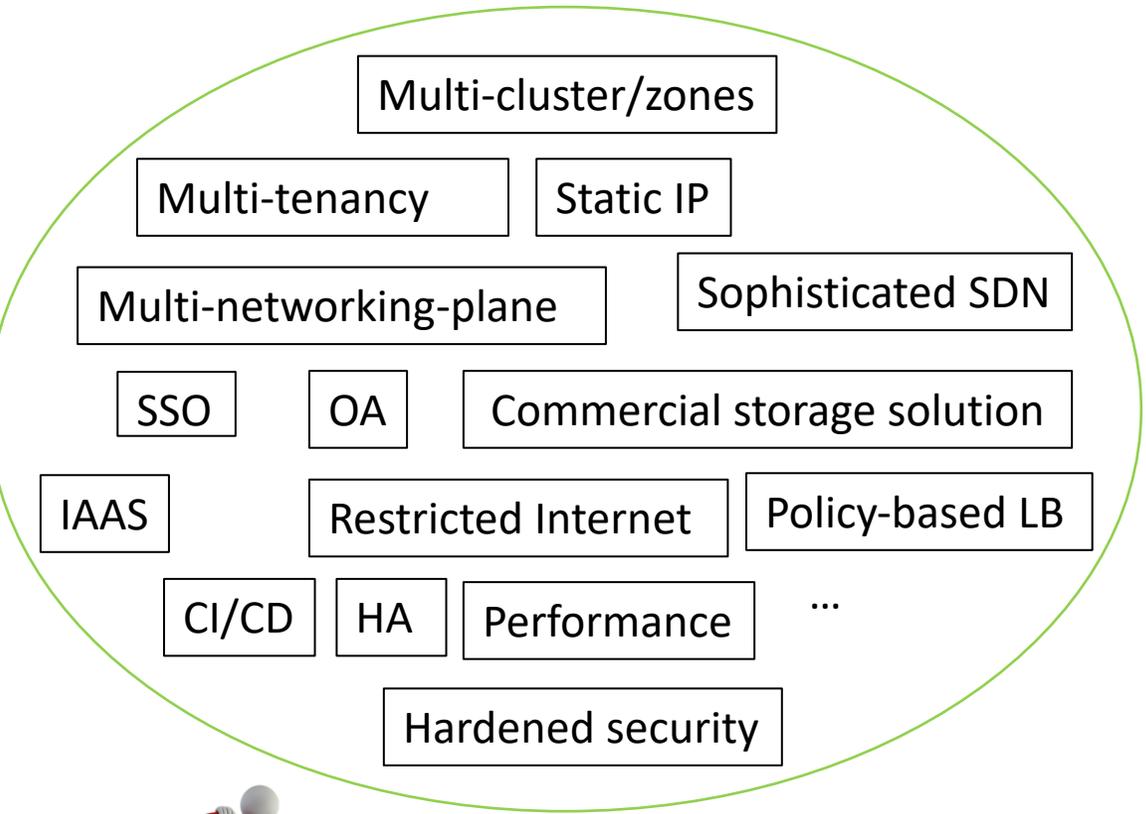


KubeCon



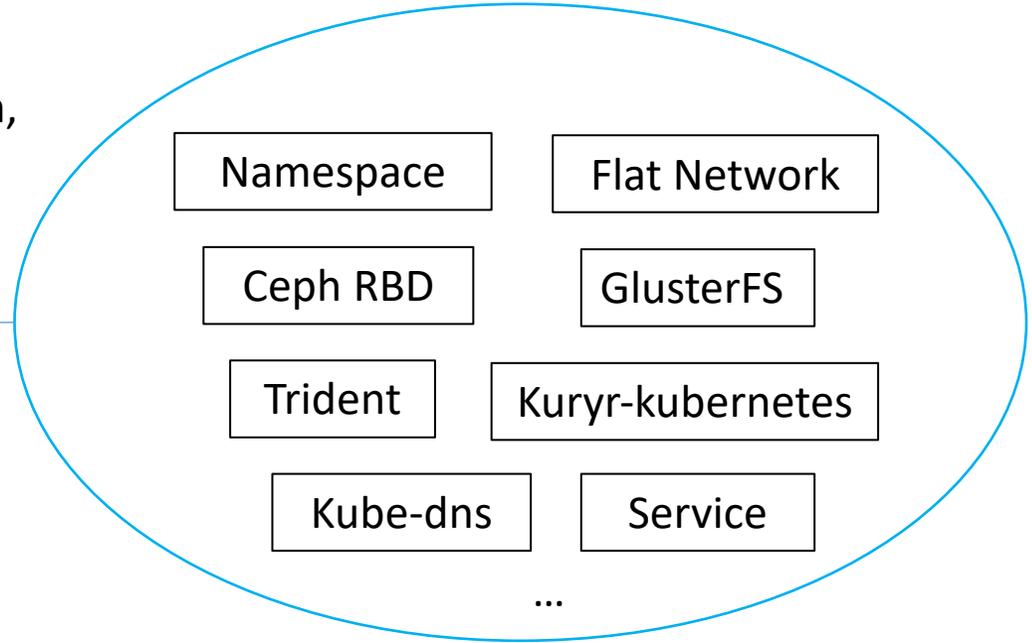
CloudNativeCon

Europe 2018



The desired banking cloud

Not a chasm,  
but a gulp



Off-the-shelf Kubernetes

# The Journey – Before Us



KubeCon



CloudNativeCon

Europe 2018

Studied LXC, cgroups

13

14

Studied docker

Docker as light-weight virtualization

15

Docker as release tool

15

Dev-oriented microservice

16

Admin complexity surged

17

# The Journey – With Us

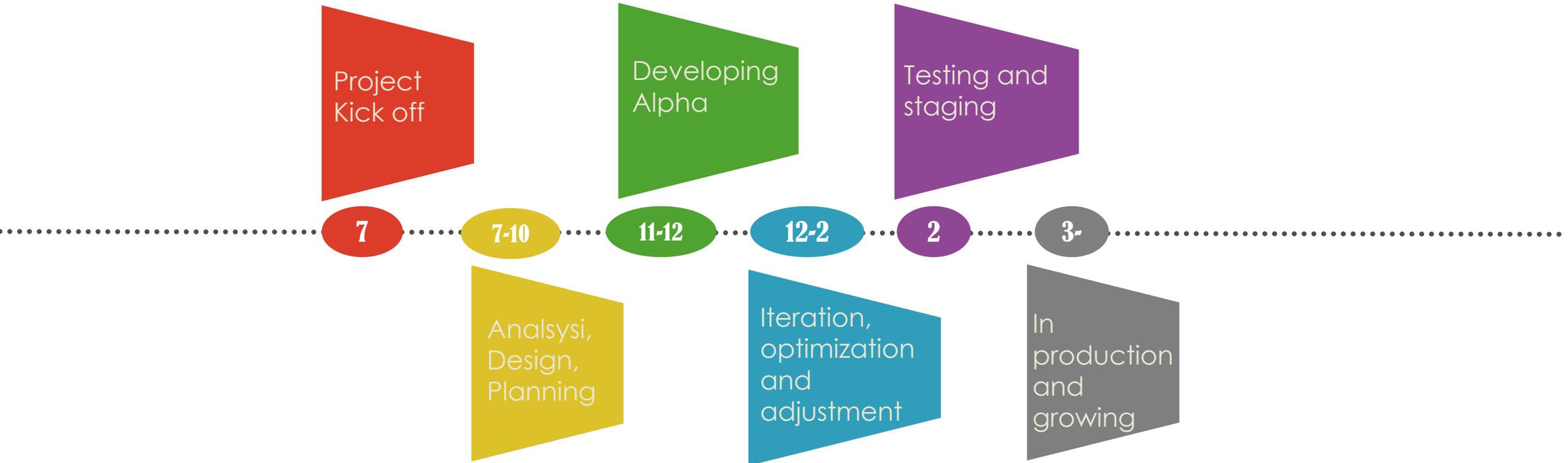


KubeCon



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



## • 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

微服务项目各工程代码同步分支tag记录

repo	branches	tag	最新branch	最新tag	2.5.0上最新tag	2.5.7	dom	repo	branches	tag	最新branch	最新tag	2.5.6上最新tag	2.5.7
cap-admin	master	v0.211	unology	v4.02-up	v0.214		✓	cap-admin	release-01	v0.13	unology	v4.02-up	v0.14	
cap-actuator	unology	unology-v0.03	unology	v4.02-up	unology-v0.03		✓	cap-actuator	release-01	v0.13	unology	v4.02-up	v0.14	
cap-actuator	unology	unology-v0.03	unology	v4.02-up	unology-v0.03		✓	cap-actuator	release-02	v0.20	unology	v4.02-up	v0.20	
cap-actuator	unology	unology-v0.03	unology	v4.02-up	unology-v0.03		✓	cap-actuator	release-03	v0.20	unology	v4.02-up	v0.20	
cap-actuator	unology	unology-v0.03	unology	v4.02-up	unology-v0.03		✓	cap-actuator	release-04	v0.20	unology	v4.02-up	v0.20	
cap-actuator	unology	unology-v0.03	unology	v4.02-up	unology-v0.03		✓	cap-actuator	release-05	v0.20	unology	v4.02-up	v0.20	
cap-actuator	unology	unology-v0.03	unology	v4.02-up	unology-v0.03		✓	cap-actuator	release-06	v0.20	unology	v4.02-up	v0.20	
cap-actuator	unology	unology-v0.03	unology	v4.02-up	unology-v0.03		✓	cap-actuator	release-07	v0.20	unology	v4.02-up	v0.20	
cap-actuator	unology	unology-v0.03	unology	v4.02-up	unology-v0.03		✓	cap-actuator	release-08	v0.20	unology	v4.02-up	v0.20	
cap-actuator	unology	unology-v0.03	unology	v4.02-up	unology-v0.03		✓	cap-actuator	release-09	v0.20	unology	v4.02-up	v0.20	
cap-actuator	unology	unology-v0.03	unology	v4.02-up	unology-v0.03		✓	cap-actuator	release-10	v0.20	unology	v4.02-up	v0.20	
cap-actuator	unology	unology-v0.03	unology	v4.02-up	unology-v0.03		✓	cap-actuator	release-11	v0.20	unology	v4.02-up	v0.20	
cap-actuator	unology	unology-v0.03	unology	v4.02-up	unology-v0.03		✓	cap-actuator	release-12	v0.20	unology	v4.02-up	v0.20	

The screenshot shows a Kubernetes dashboard with a table of application clusters. The table includes columns for application name, namespace, running time, and update time. Applications listed include cap-apache, cap-fac, cap-master, cap-monitor, cap-predict, cap-server, zookeeper, and jenkins.

项目变更记录

日期	变更内容	负责人	状态
2018-03-08	修复 cap-admin 部署失败问题	张三	完成
2018-03-09	新增 cap-monitor 应用	李四	完成
2018-03-10	优化 cap-server 性能	王五	完成
2018-03-11	修复 cap-predict 部署失败问题	赵六	完成
2018-03-12	新增 cap-admin 应用	孙七	完成
2018-03-13	优化 cap-fac 性能	周八	完成
2018-03-14	修复 cap-master 部署失败问题	吴九	完成
2018-03-15	新增 cap-actuator 应用	郑十	完成
2018-03-16	优化 cap-actuator 性能	冯十一	完成
2018-03-17	修复 cap-actuator 部署失败问题	陈十二	完成
2018-03-18	新增 cap-actuator 应用	褚十三	完成
2018-03-19	优化 cap-actuator 性能	褚十四	完成
2018-03-20	修复 cap-actuator 部署失败问题	褚十五	完成
2018-03-21	新增 cap-actuator 应用	褚十六	完成
2018-03-22	优化 cap-actuator 性能	褚十七	完成
2018-03-23	修复 cap-actuator 部署失败问题	褚十八	完成
2018-03-24	新增 cap-actuator 应用	褚十九	完成
2018-03-25	优化 cap-actuator 性能	褚二十	完成



**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



KubeCon



CloudNativeCon

Europe 2018

监控状态 / 监控详情 Admin

### C-327476

状态: 异常

活动: 大润发商场新品推介会 1

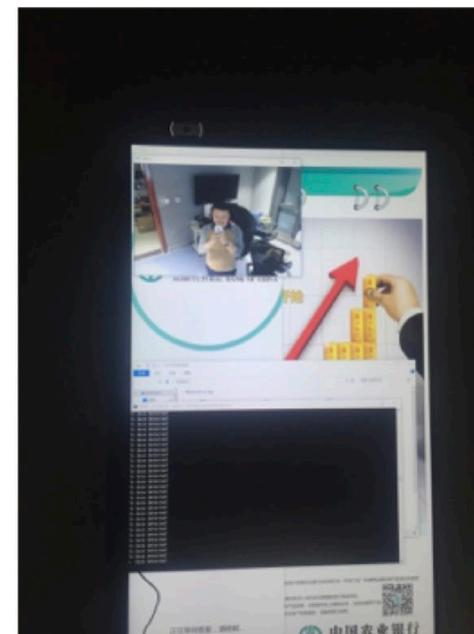
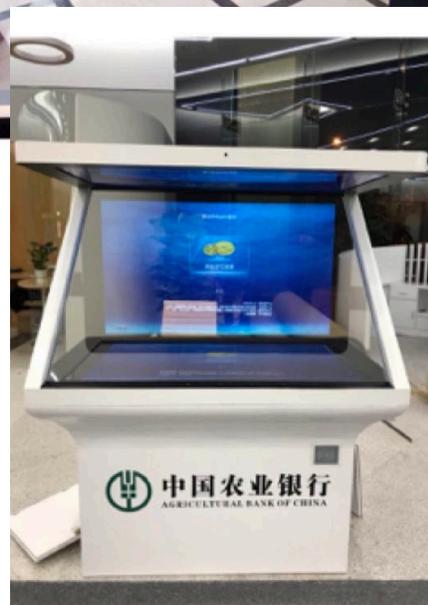
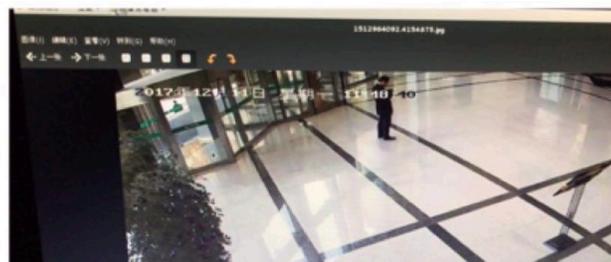
时间: 2017/09/28 09:00 - 2017/09/28 11:00

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
	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

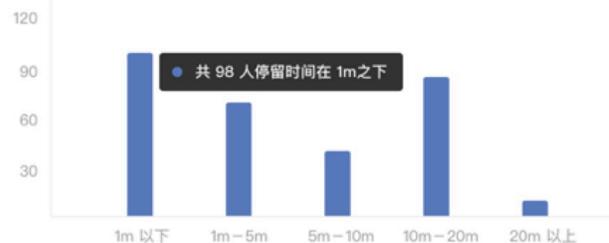
实时监控图

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

共 123 条，当前显示 20 条每页



客户停留时间分布



# 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



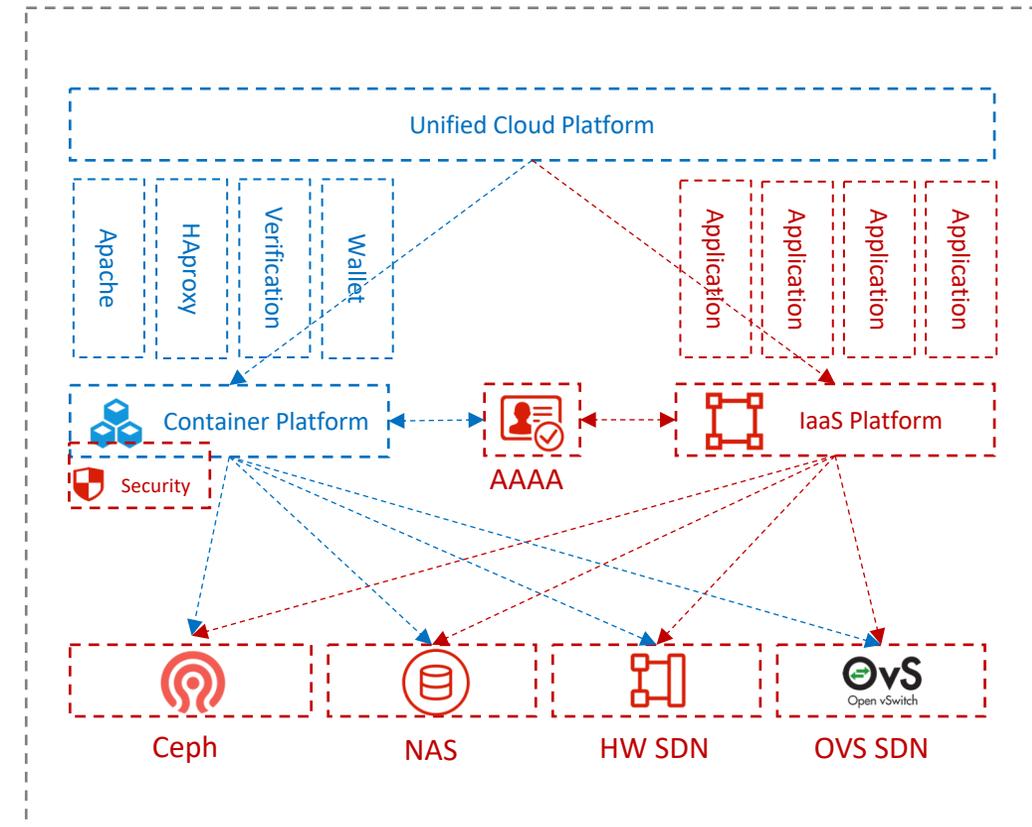
AuthN/Z: Seamlessly integrate with internal security systems



Control Plane: A unified control plane across organization



Security: Conform to financial security regulation



# Deployment

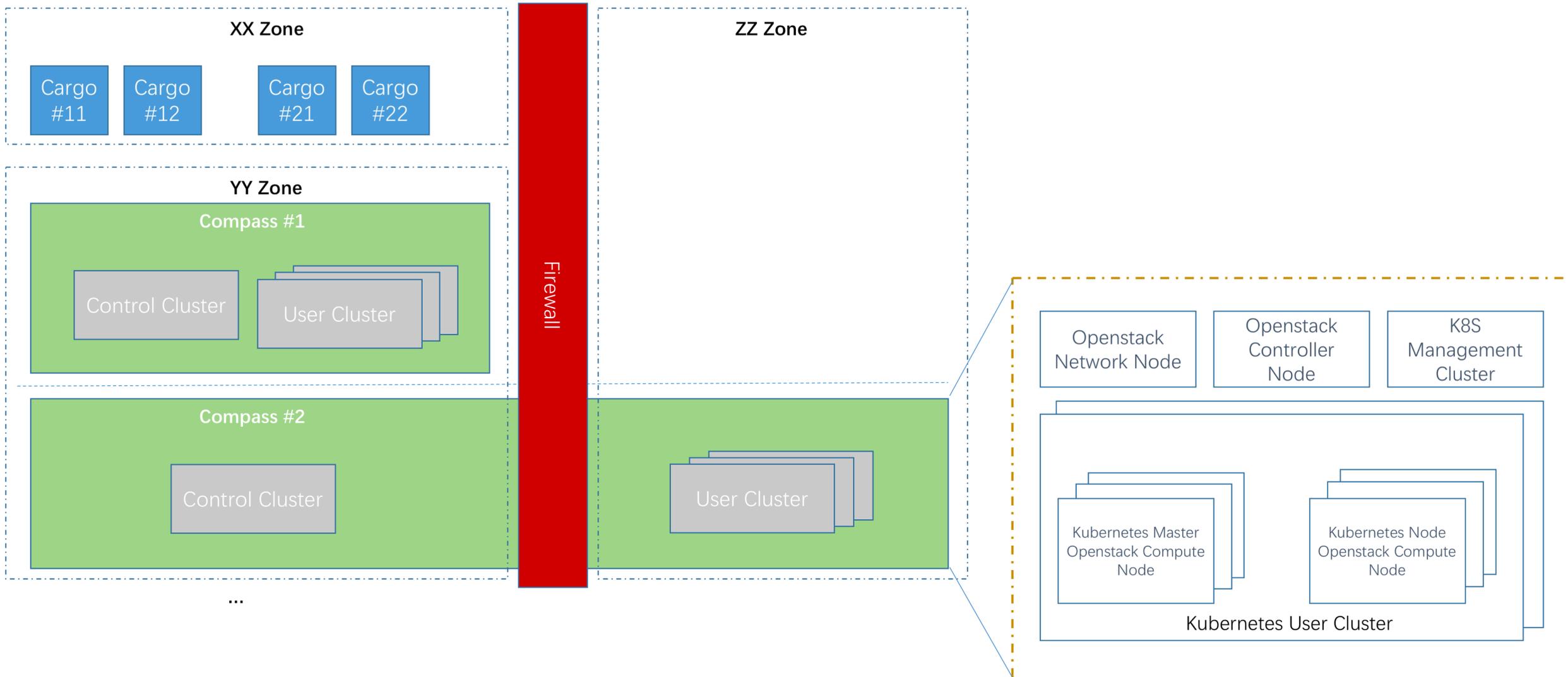


KubeCon



CloudNativeCon

Europe 2018



# SSO - Background



KubeCon



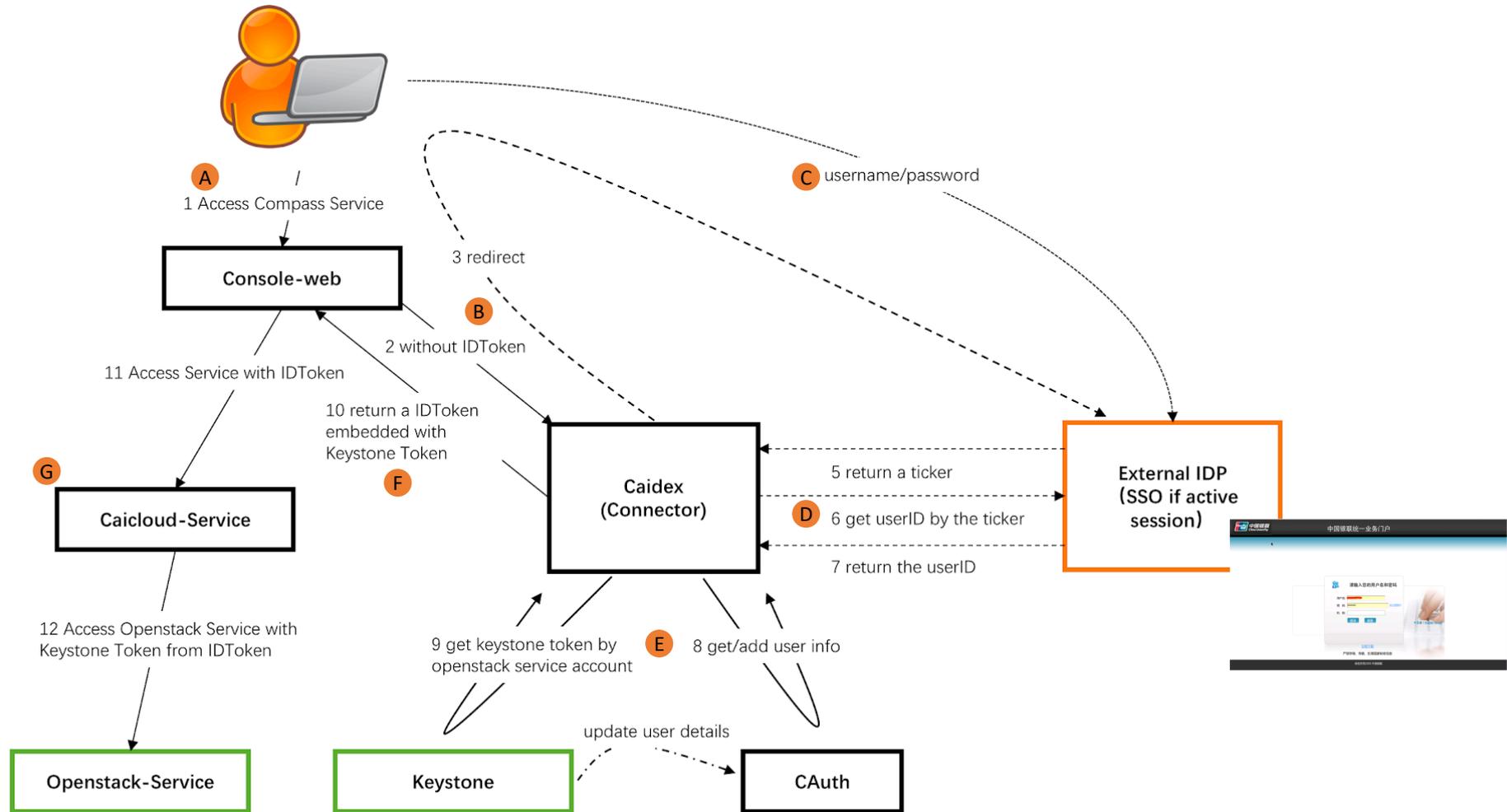
CloudNativeCon

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



X Synchronize keystone information to cauth with a long running daemon

# 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 displays the Compass UI for a tenant named 'QA-Only-Control'. The breadcrumb navigation at the top left is '← Tenants / QA-Only-Control'. The main content area shows the cluster details and resource quotas.

**Cluster Details:**

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

**Resource Quotas:**

The 'Resource Quotas' tab is selected, showing the following details:

- CPU:** Request 4 Core, Limit 20 Core
- Memory:** Request 8 GiB, Limit 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

# Tenant (Tenant Admins)



KubeCon



CloudNativeCon

Europe 2018

The screenshot shows the Compass Groups management interface. The left sidebar is circled in orange and contains the following navigation items: Partitions, Applications, App Store, Orchestration, Volumes, Load balancing, Configs, Delivery Center, Registry, DevOps, OP CENTER, Logging, Monitor, Alerting, and MANAGEMENT. The main content area is titled 'Groups' and features a '+ Add' button. Below it is a table with the following data:

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

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

Page navigation: Total 5 1/1 Every page shows 10

# Tenant (Tenant Admins)



KubeCon



CloudNativeCon

Europe 2018

compass

← Groups / SREs

About Manual English Tenant: QA-Only-Control admin

## SREs

Modify

2018-04-28

Num of Members

0

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

Members Partitions Templates Loadbalancers Configs Operations DevOps Cargo Clever

Clusters

multiple-tenant-c... >

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

Basic Role	User	+	User	+	Owner	+
qatest						

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

# Tenant - Summary

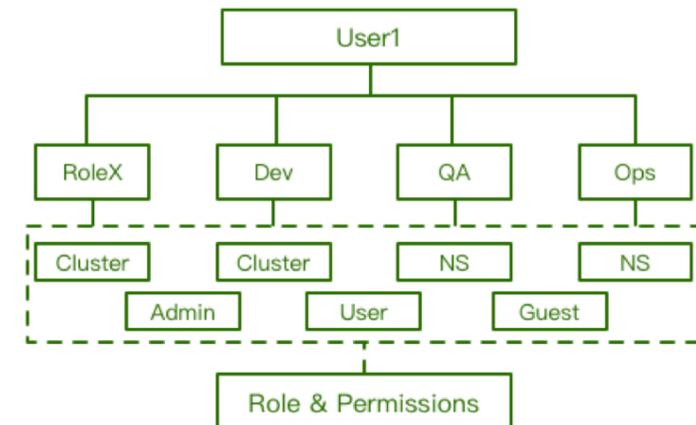
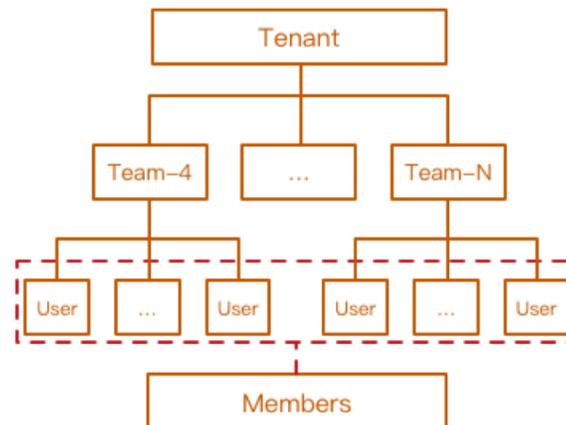
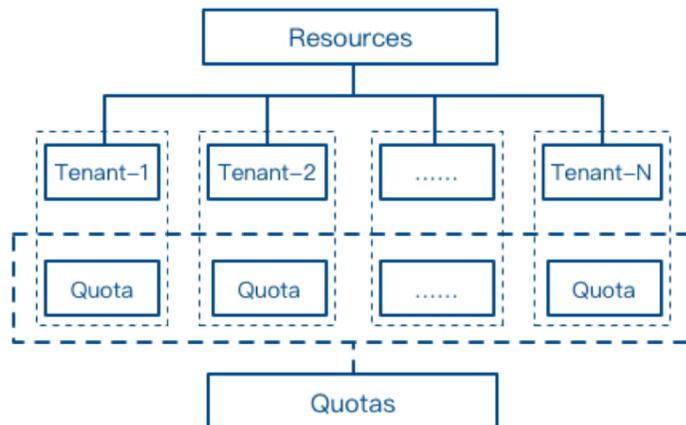
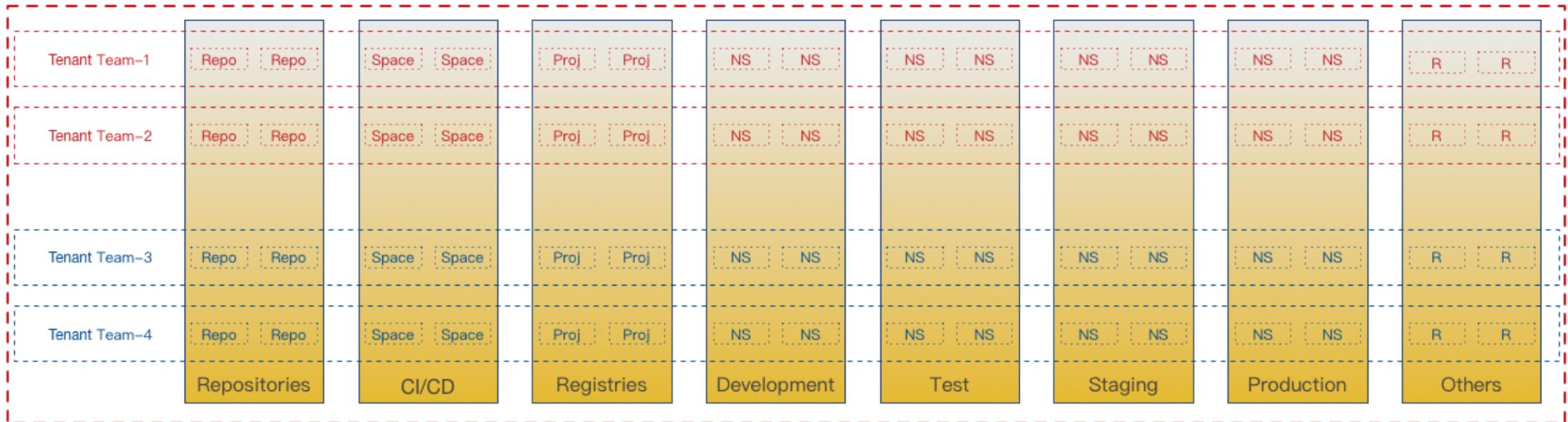


KubeCon



CloudNativeCon

Europe 2018



# Tenant - OpenStack Mapping



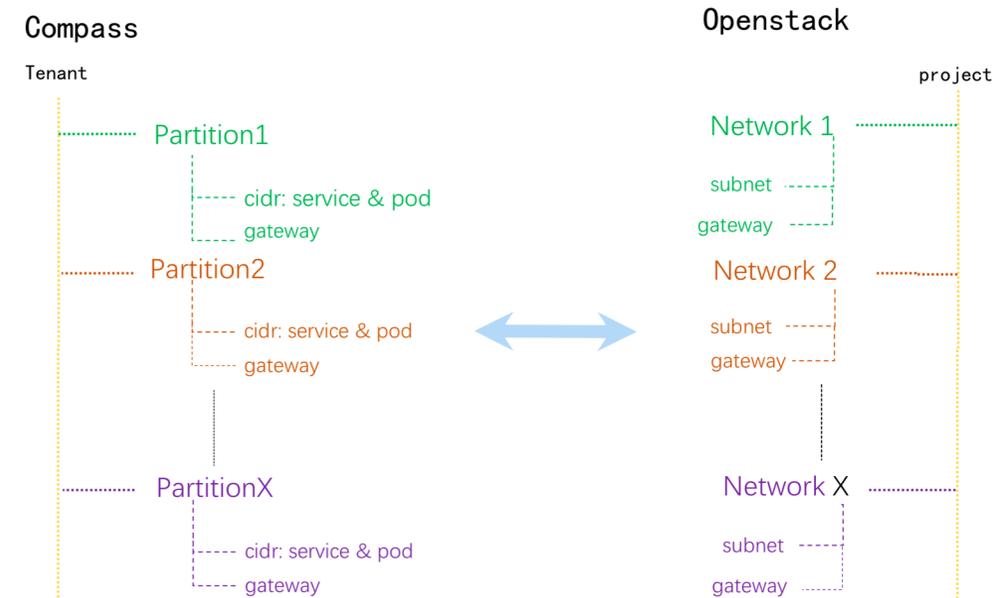
KubeCon



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 the '租户管理' (Tenant Management) interface. The main content area is titled '新增租户 Create Tenant'. It contains a '基本信息' (Basic Information) section with two input fields: '名称' (Name) with a placeholder '输入租户名称' and '描述' (Description) with a placeholder '选项 200字以内'. Below this is a section titled '选择 openstack 租户' (Select OpenStack Tenant) with a search bar and a dropdown menu. The dropdown menu is open, showing a list of projects. The first item, 'docker-k8s k8s-project3', is selected. A tooltip for this item shows the following details:

project信息	
description:	暂无
domain_name:	docker-k8s2
project_id:	118e9638341c42dca67dc26f590186cd

Red annotations on the screenshot include 'Create Tenant requires choosing an OpenStack Tenant' pointing to the dropdown menu, and '租户管理 Tenant Management' pointing to the selected menu item in the left sidebar.

# Tenant - with OpenStack



KubeCon



CloudNativeCon

Europe 2018

中国银联 China UnionPay | 应用分区

使用指南 租户: test-tenant docker

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

分区名称

所属集群

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  Core  Core

内存配额

**l3policy\_id**

**podCIDR**

**serviceCIDR**

Network Option

# Tenant - with OpenStack

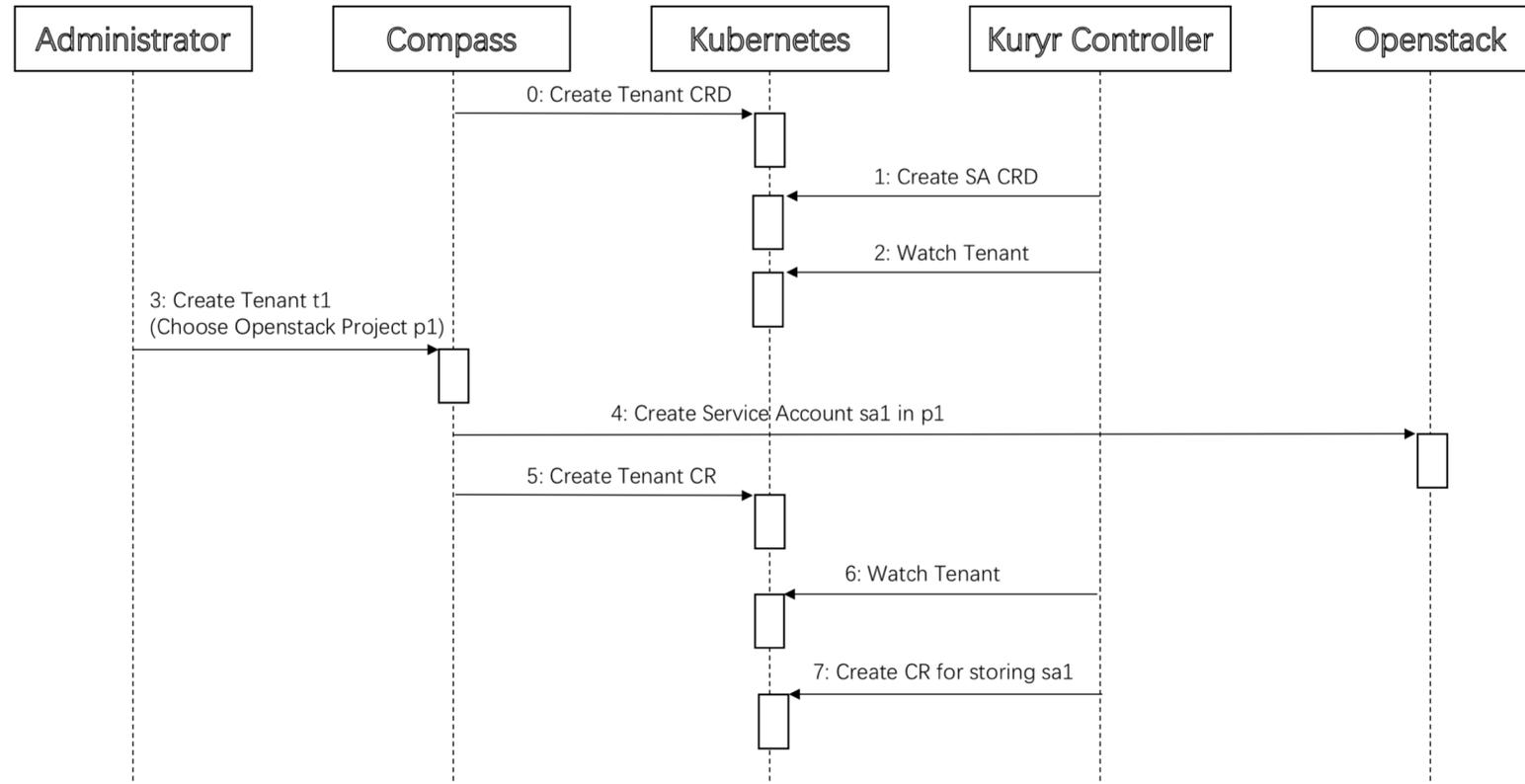


KubeCon



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

# Network - Overview



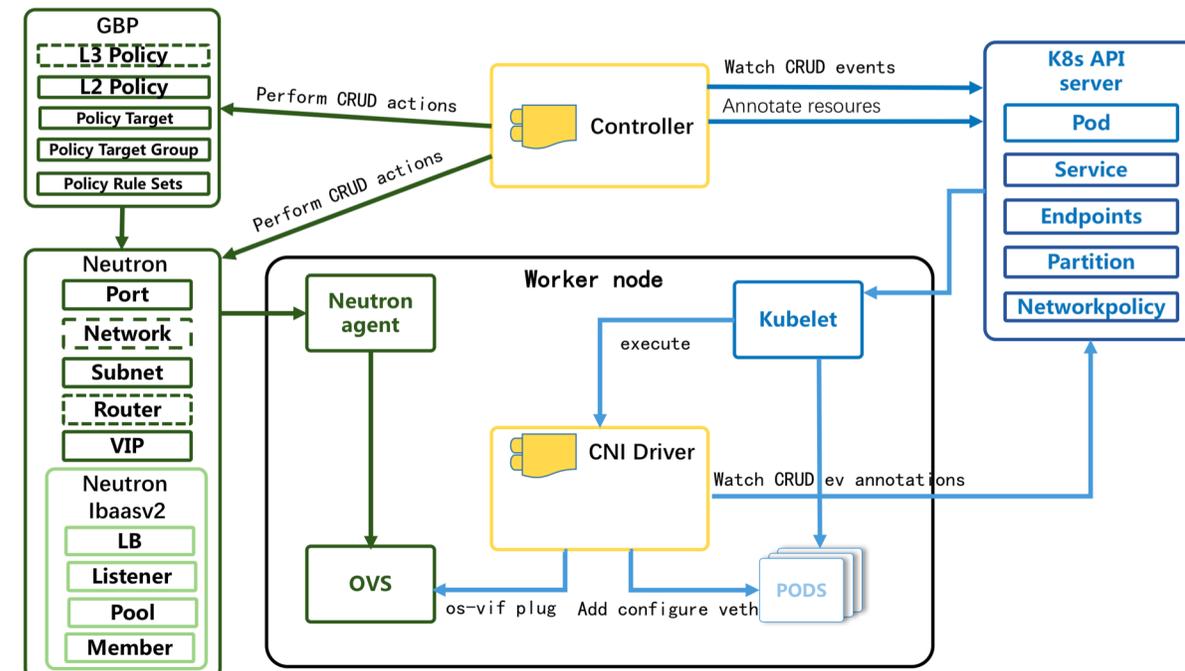
KubeCon



CloudNativeCon

Europe 2018

- 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



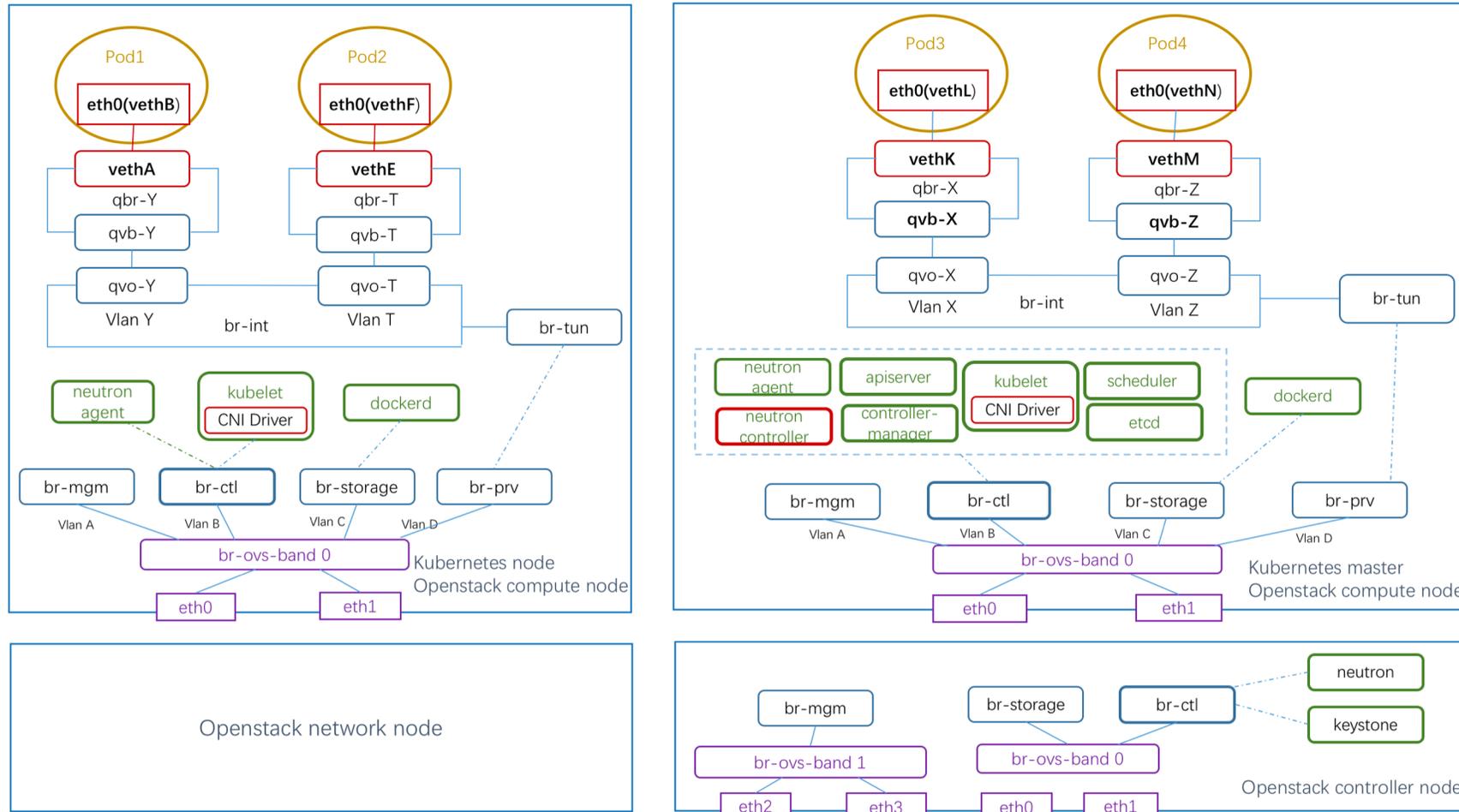
KubeCon



CloudNativeCon

Europe 2018

## Openstack ovs-based vxlan



# Network - Service



KubeCon



CloudNativeCon

Europe 2018

- 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	--



Protocol, Port, Members (Endpoints)

# Network - Service



KubeCon



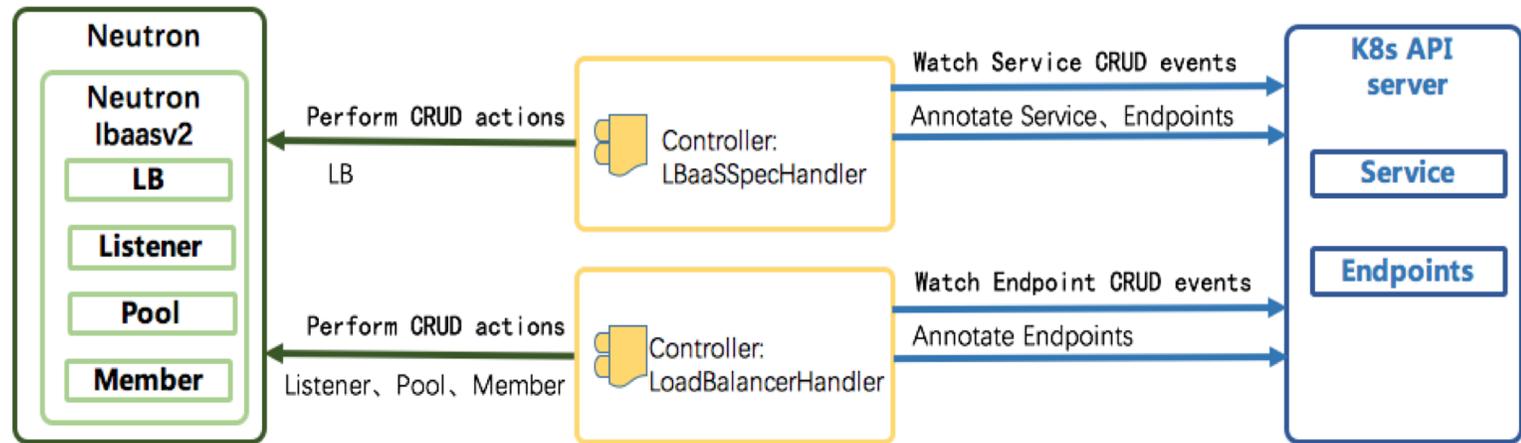
CloudNativeCon

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



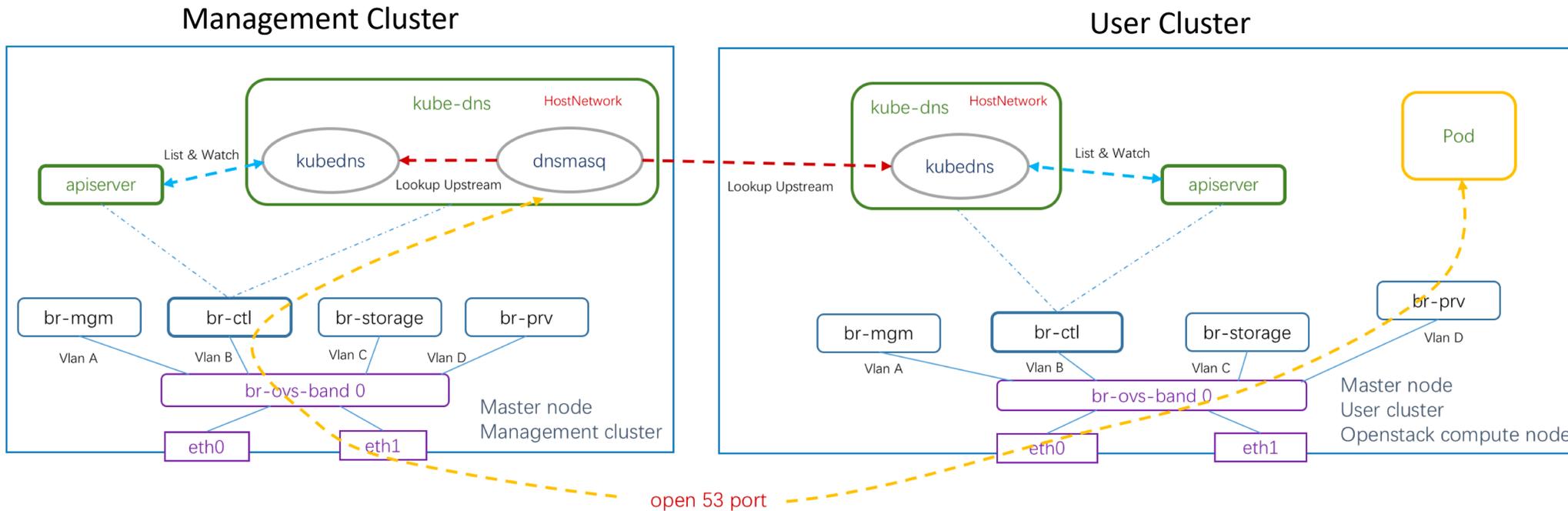
CloudNativeCon

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

- 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



# Network - DNS Deployment

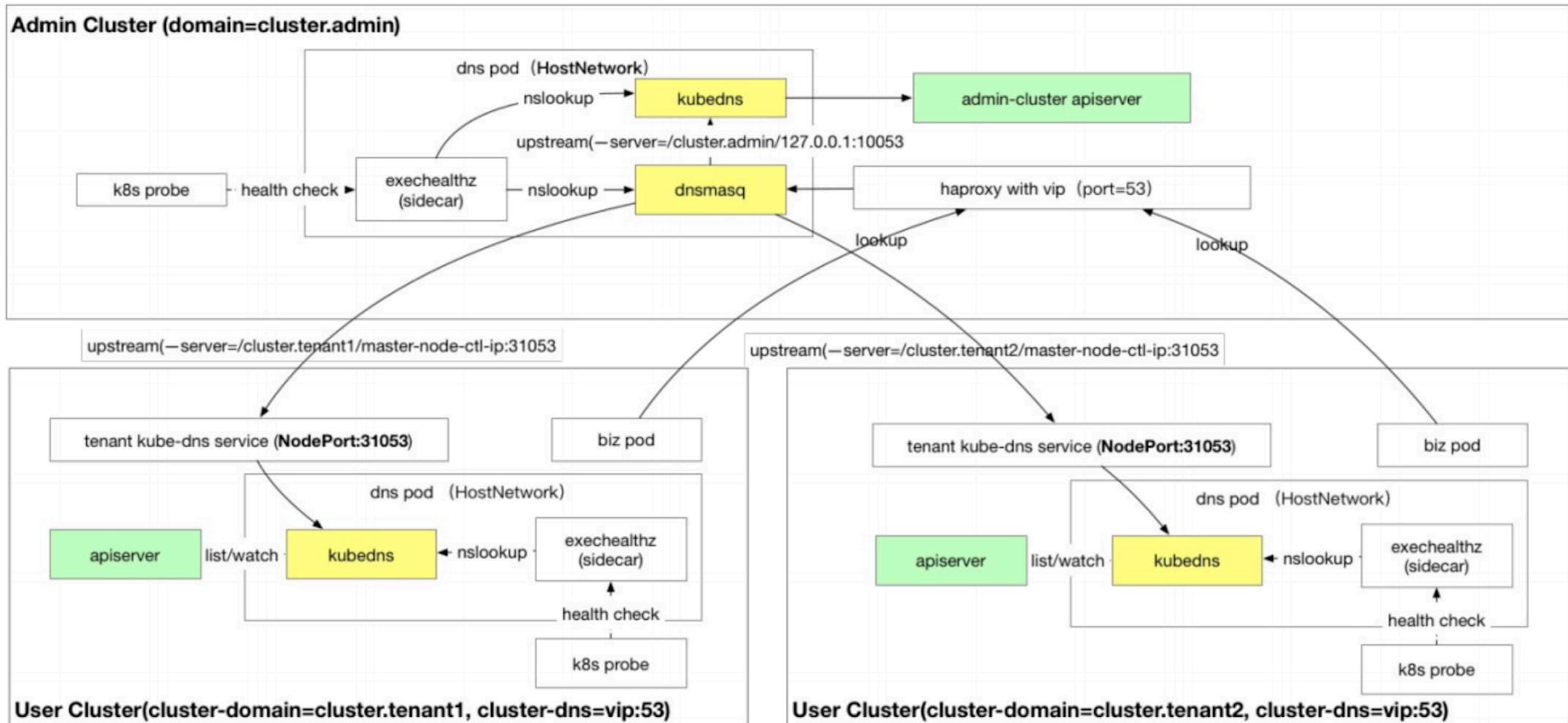


KubeCon



CloudNativeCon

Europe 2018



# Storage - NAS



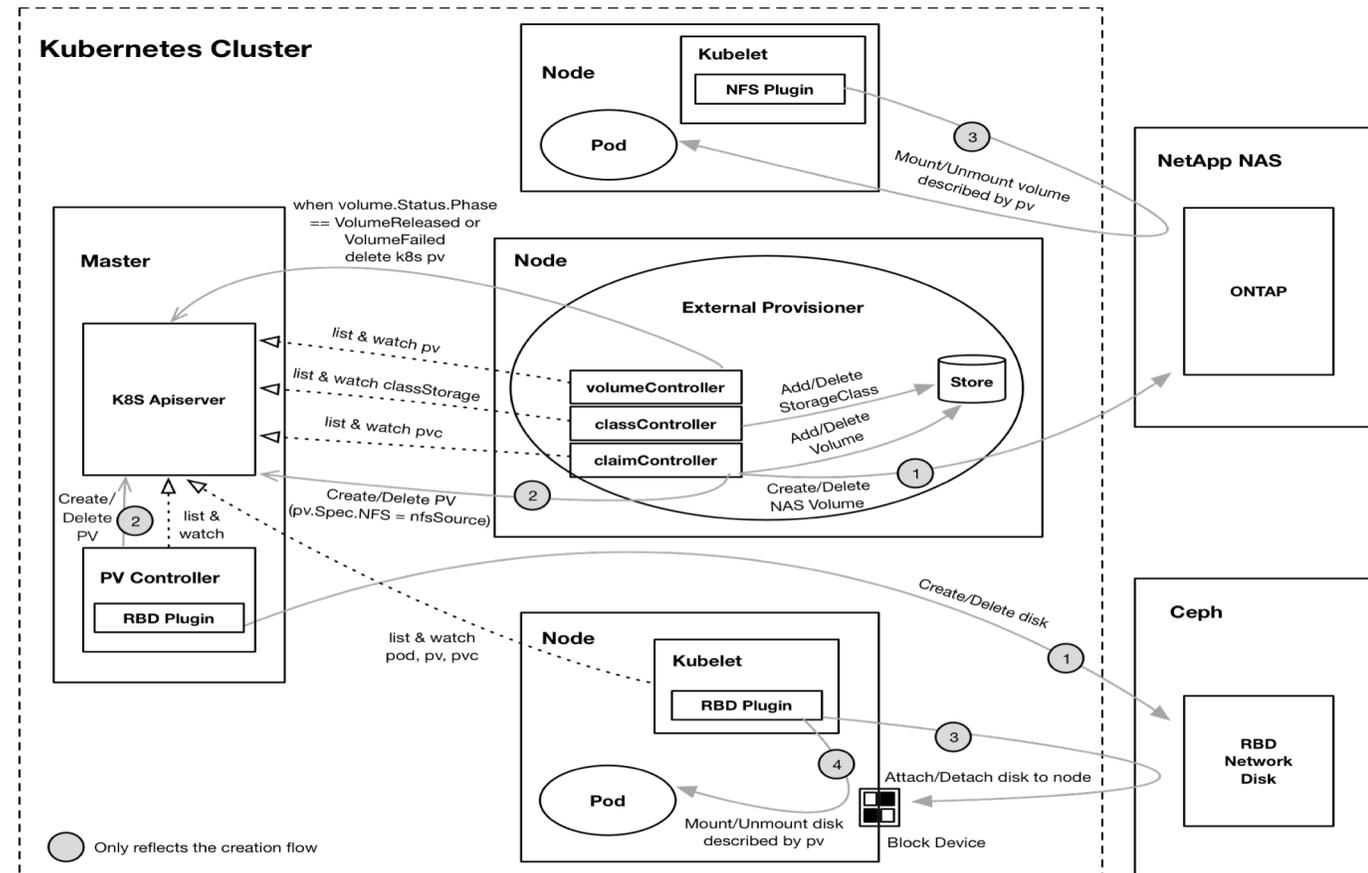
KubeCon



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

Container Registry

About Manual English Tenant: QA-Control-User admin

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

Total 2 1/1 Every page shows 10

镜像仓库

仓库管理 项目管理 docker account

library (7)

- 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

同步镜像 Sync Image

请选择 Select another registry

取消 开始同步

mysql 版本信息

同步至其他仓库

5.7.14 安全

2016-08-23 03:20:13

没有更多了

# Registry HA



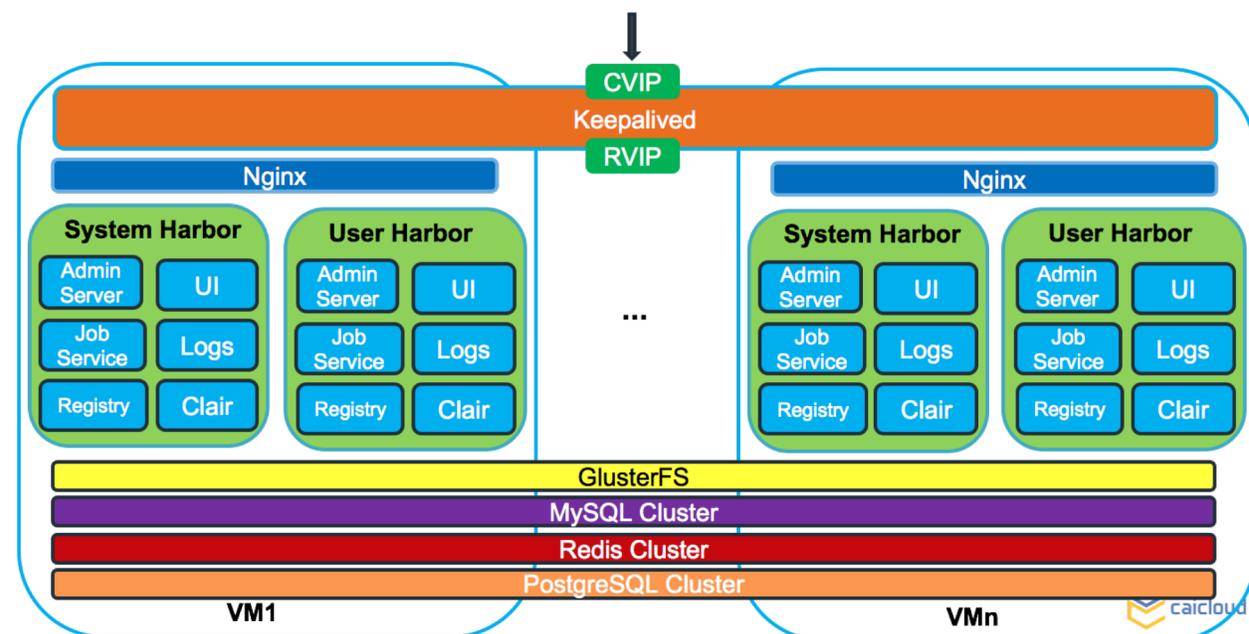
KubeCon



CloudNativeCon

Europe 2018

- 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

