KubeCon

CloudNativeCon

Europe 2018

Kubernetes on Supporting $8 Trillion Card Payments in China

Xin Zhang & Deyuan Deng

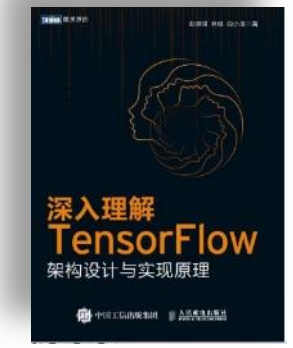
{zhangxin, deyuan}@caicloud.io

caicloud
才云

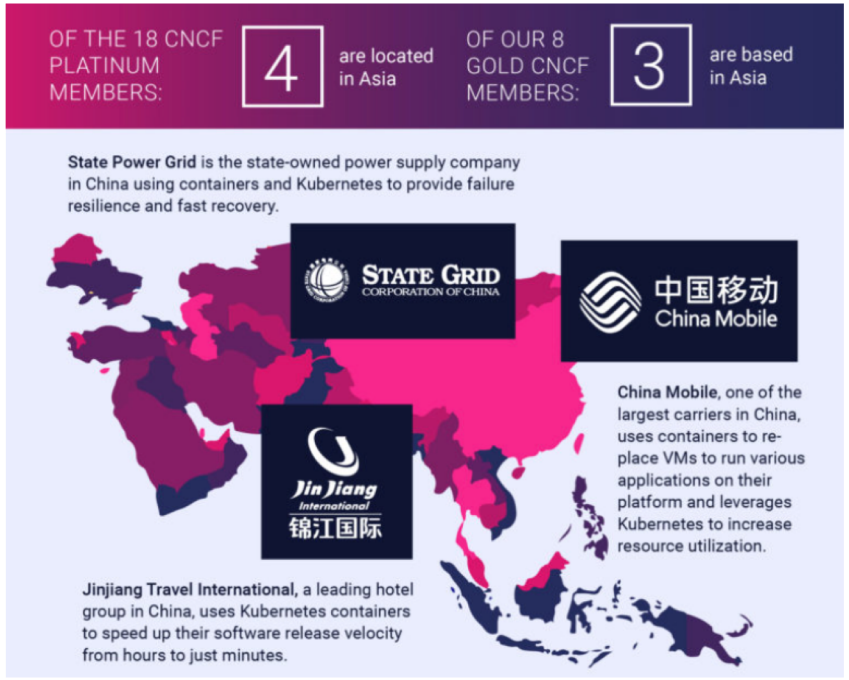
About Us @caicloud.io

KubeCon | CloudNativeCon
Europe 2018

- Technical Contributors
  - From Kubernetes, to TensorFlow, to Kubeflow
    - github.com/kubeflow/kubeflow: Kubernetes-based ML stack



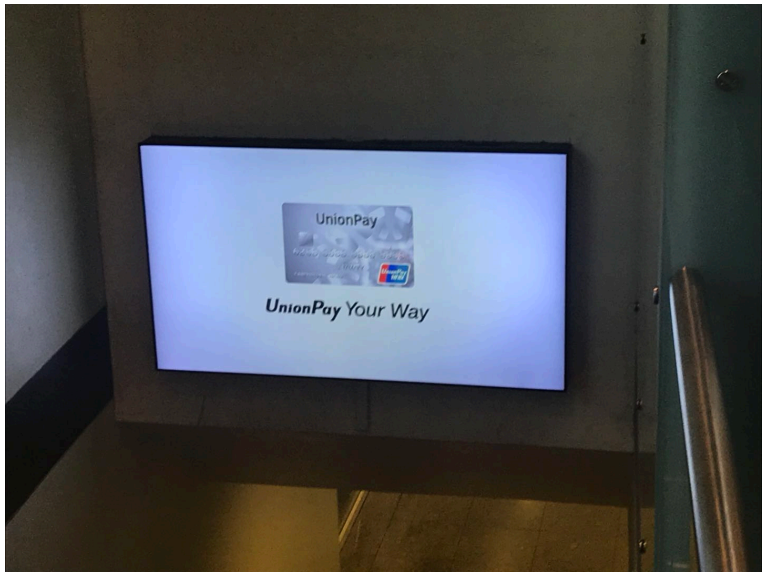
- Chinese Community Organizers
  - Kubernetes Chinese community and i18n project
    - github.com/kubernetes/kubernetes-docs-cn
  - TensorFlow Chinese community operator
    - tensorflowers.cn
  - Kubeflow Chinese community initiator

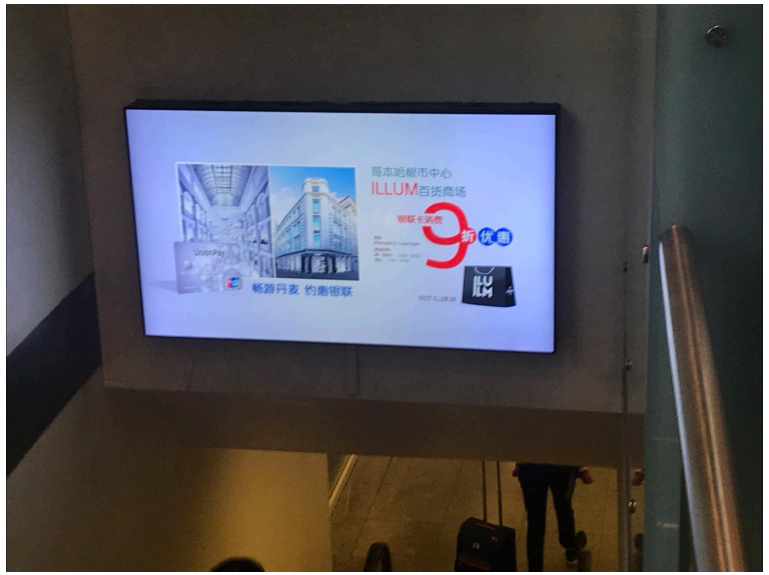


- Enterprise Products and Solutions
  - *Compass*: Kubernetes distro with value add
  - *Clever*: enterprise-grade kubeflow with AI models

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

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

Crossing the Chasm

- 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 ...

- Agility
- Scalability
- Elasticity
- Availability
- Automation

Shout out your answer!

Out-of-Box Panacea?

Multi-cluster/zones

Multi-tenancy

Static IP

Multi-networking-plane

Sophisticated SDN

SSO

OA

Commercial storage solution

IAAS

Restricted Internet

Policy-based LB

CI/CD

HA

Performance

...

Hardened security

The desired banking cloud

Not a chasm,
but a gulp

Namespace

Flat Network

Ceph RBD

GlusterFS

Trident

Kuryr-kubernetes

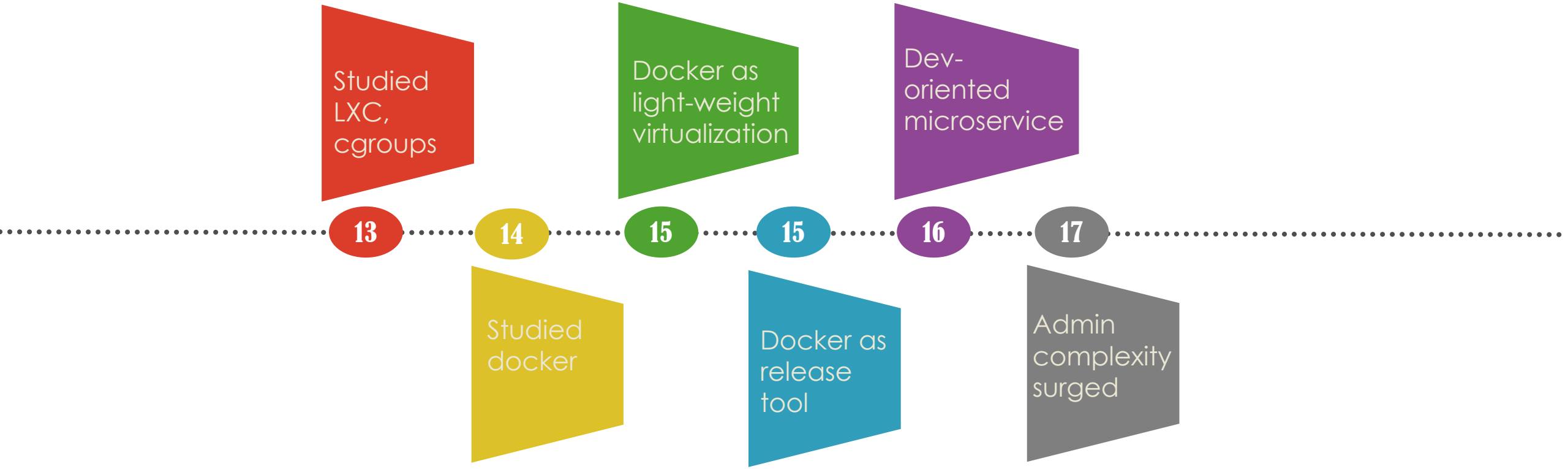
Kube-dns

Service

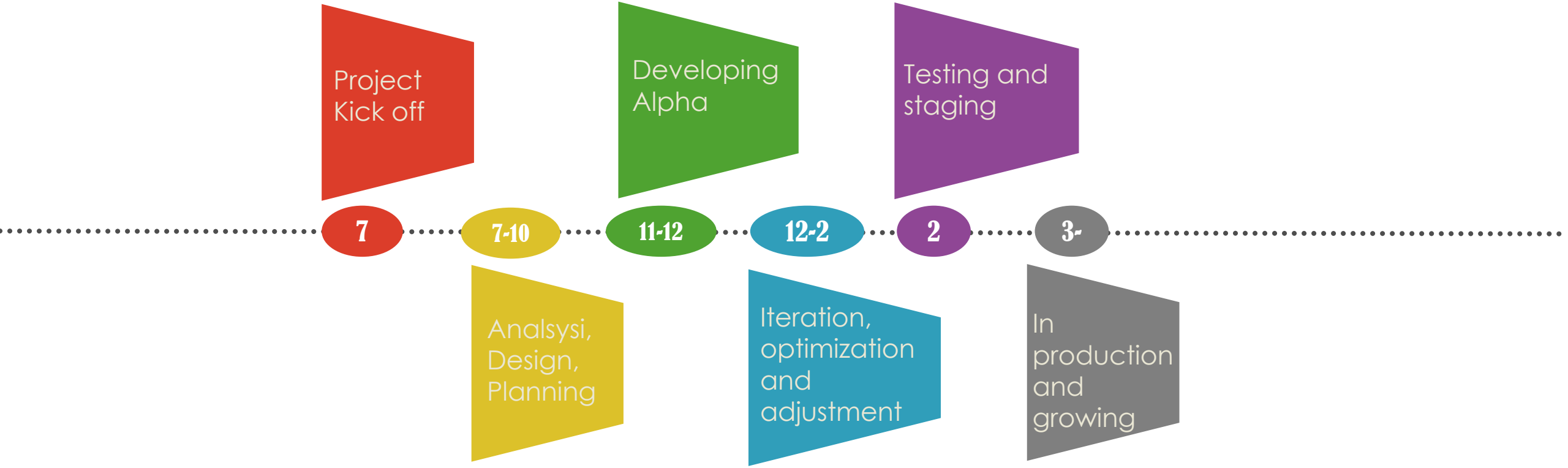
...

Off-the-shelf Kubernetes

The Journey – Before Us



The Journey – With Us



Things Delivered So Far

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

UnionPay 银联 钱包 Wallet

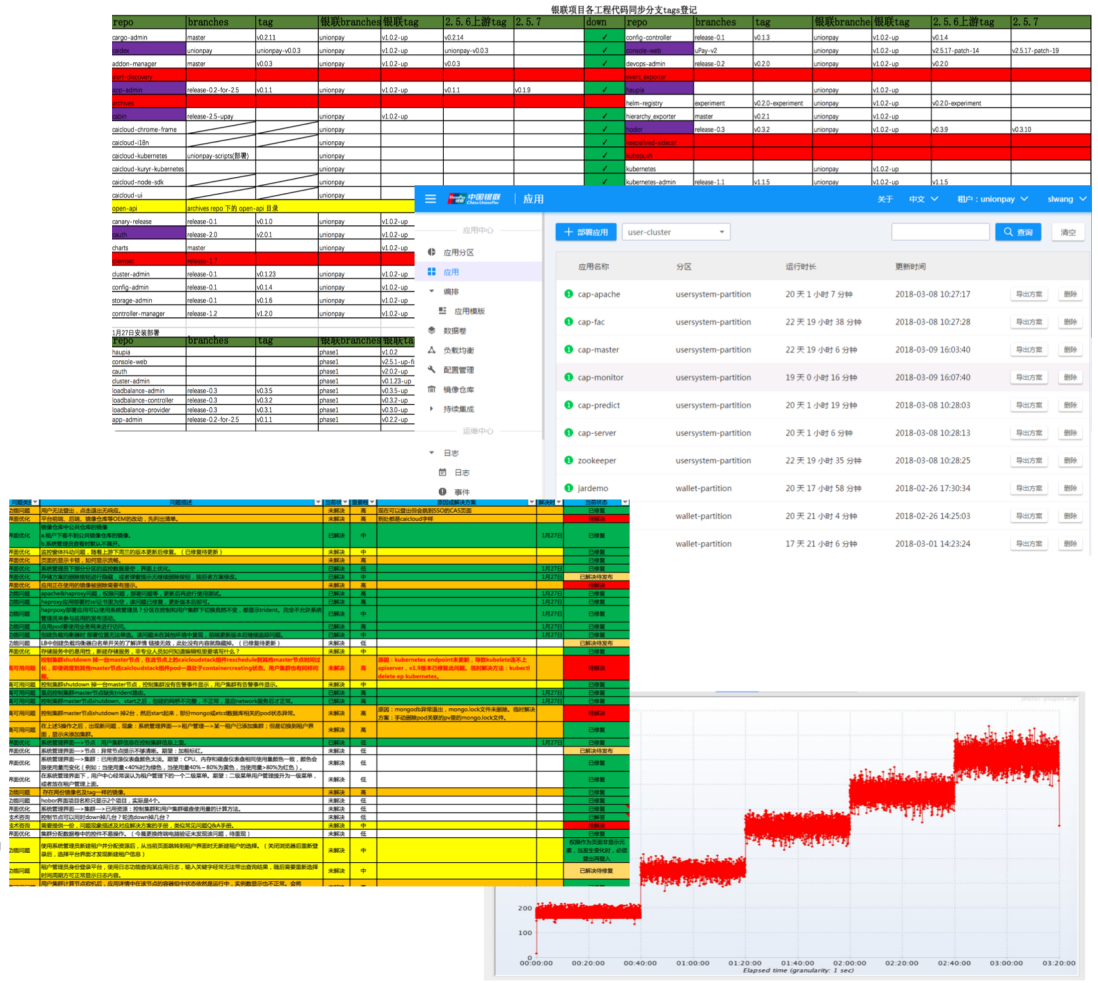
UnionPay 银联 Quick 闪付Pass

Things Delivered So Far

- 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

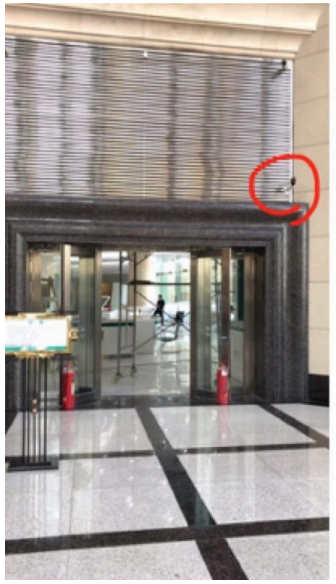
- Artifacts
  - Code review, release, branching, versioning, naming management
  - 54 code repos
  - Syncing and online upgrade process and tools for 7 environments
  - Micro-service consulting and splitting
  - Comprehensive testing reports
  - CI/CD pipeline and practice
  - Various docs: user guide,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

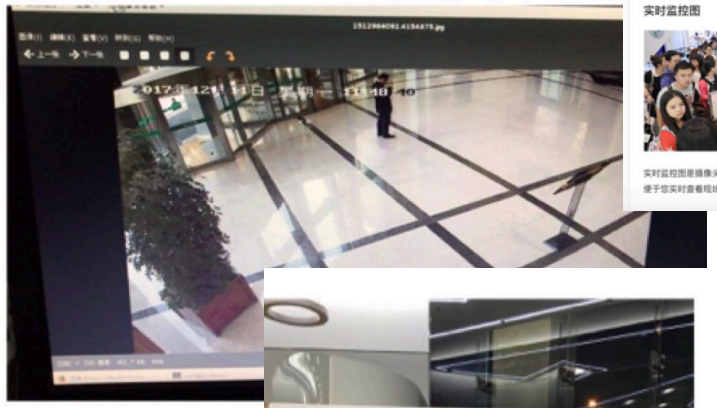


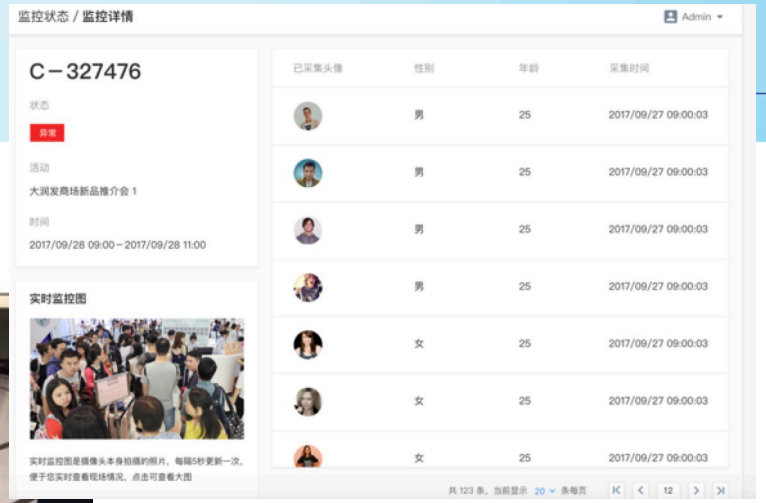
Lessons Learned

- 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!

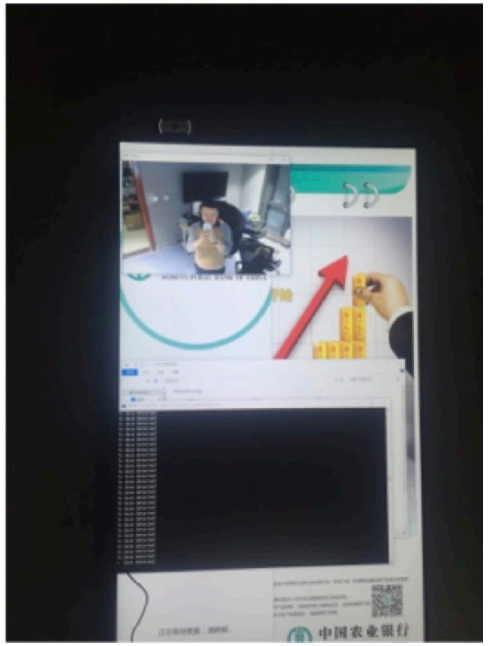
Road Ahead

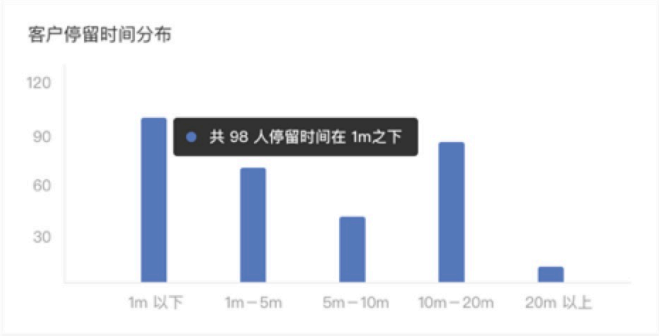





监控状态 / 监控详情
Admin
C－327476
状态
异常
活动
大润发商场新品推介会 1
时间
2017/09/28 09:00－2017/09/28 11:00
实时监控图
已采集头像
性别
年龄
采集时间
男 25 2017/09/27 09:00:03
男 25 2017/09/27 09:00:03
男 25 2017/09/27 09:00:03
男 25 2017/09/27 09:00:03
女 25 2017/09/27 09:00:03
女 25 2017/09/27 09:00:03
女 25 2017/09/27 09:00:03
共 123 条，当前显示 20 条每页
12


中国农业银行
AGRICULTURAL BANK OF CHINA




客户停留时间分布
120
90
60
30
共 98 人停留时间在 1m之下
1m 以下
1m－5m
5m－10m
10m－20m
20m 以上

Now The Technical Meat


KubeCon | CloudNativeCon
Europe 2018

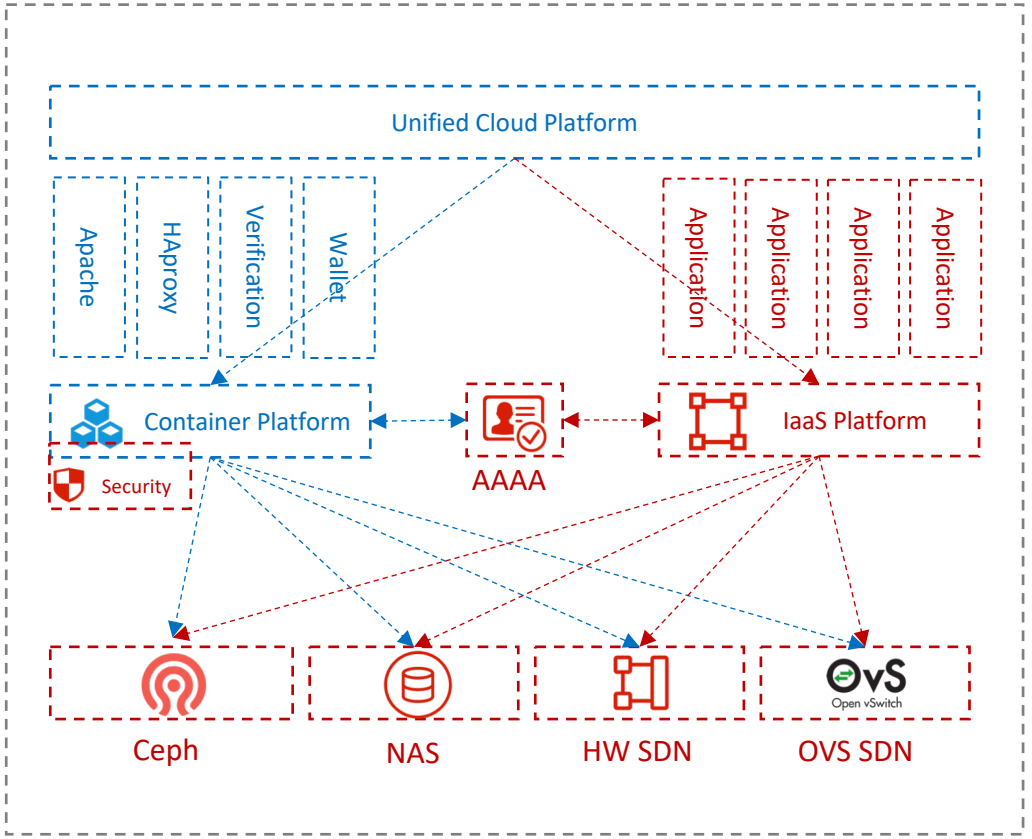
- Special thanks to
  - Jiyuan
  - Xiaojian
  - Libin
  - Qingchuan
  - Zhaole
  - Ang
  - Shanmin
  - And other Caicloud folks

Overview

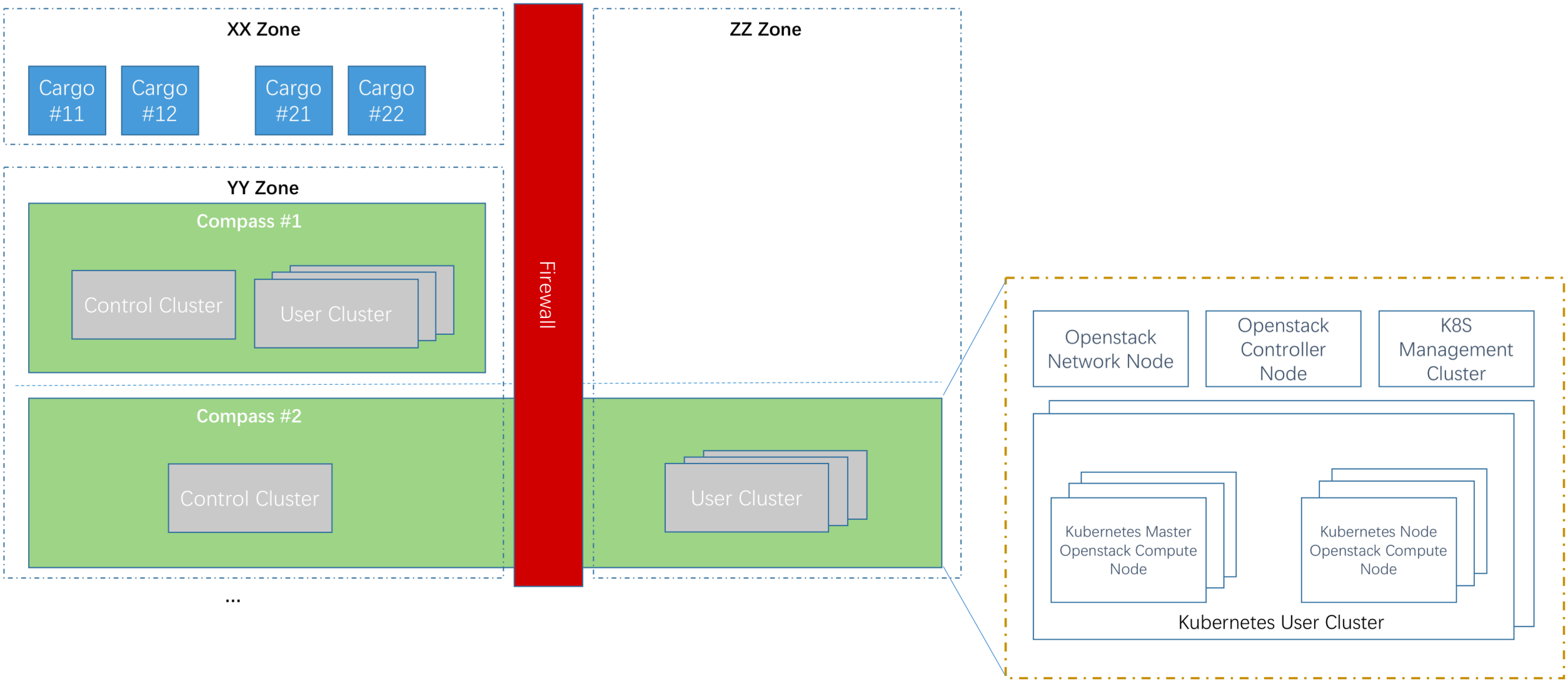
- Applications
  - Frontend Applications: Apache, HAproxy
  - Core Applications: UnionPay Wallet, User Verification System, etc

- 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



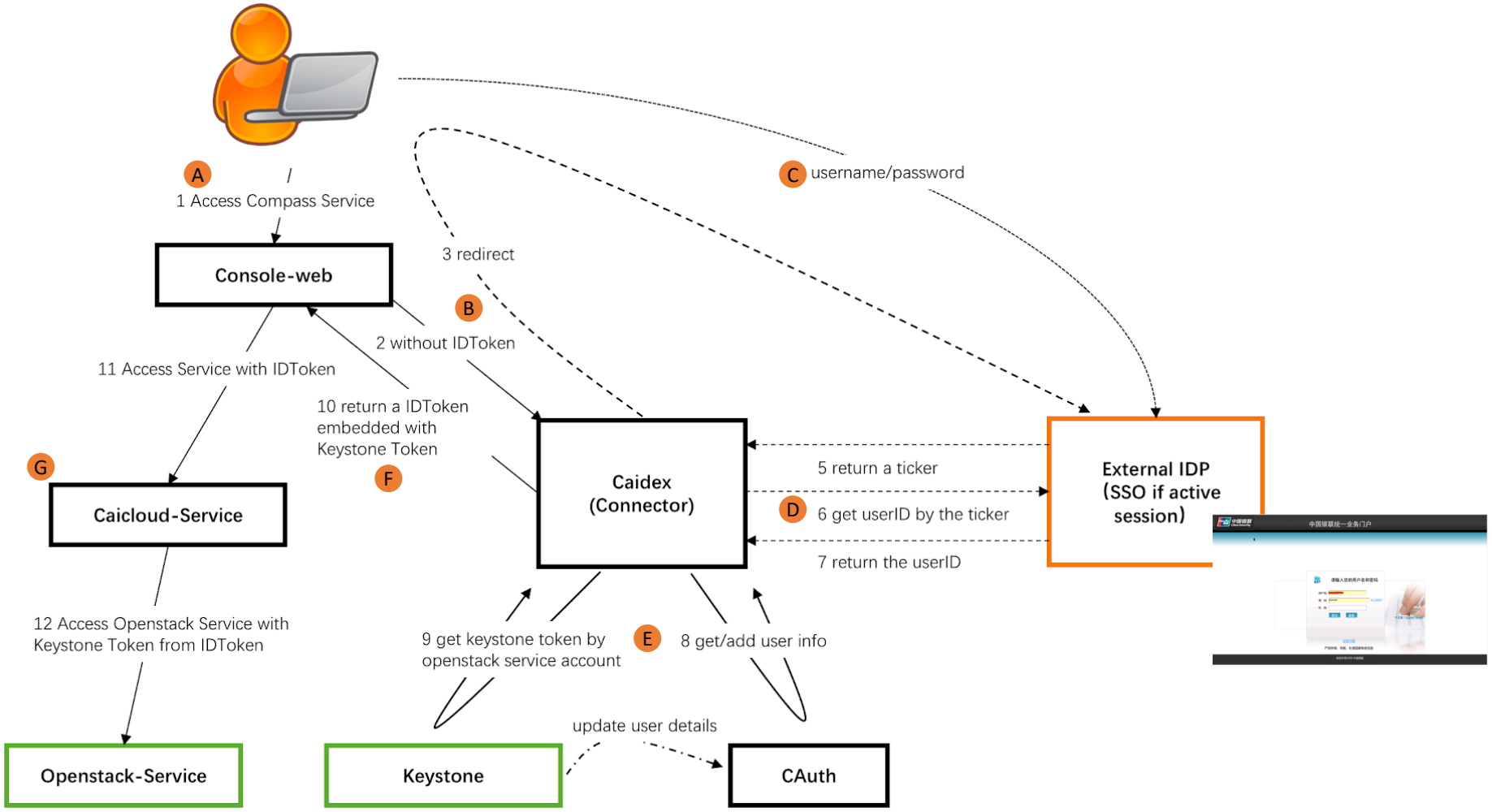
Deployment


XX Zone
Cargo #11
Cargo #12
Cargo #21
Cargo #22
YY Zone
Compass #1
Control Cluster
User Cluster
Compass #2
Control Cluster
...
Firewall
ZZ Zone
User Cluster
Openstack Network Node
Openstack Controller Node
K8S Management Cluster
Kubernetes Master Openstack Compute Node
Kubernetes Node Openstack Compute Node
Kubernetes User Cluster

SSO - Background

- 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)

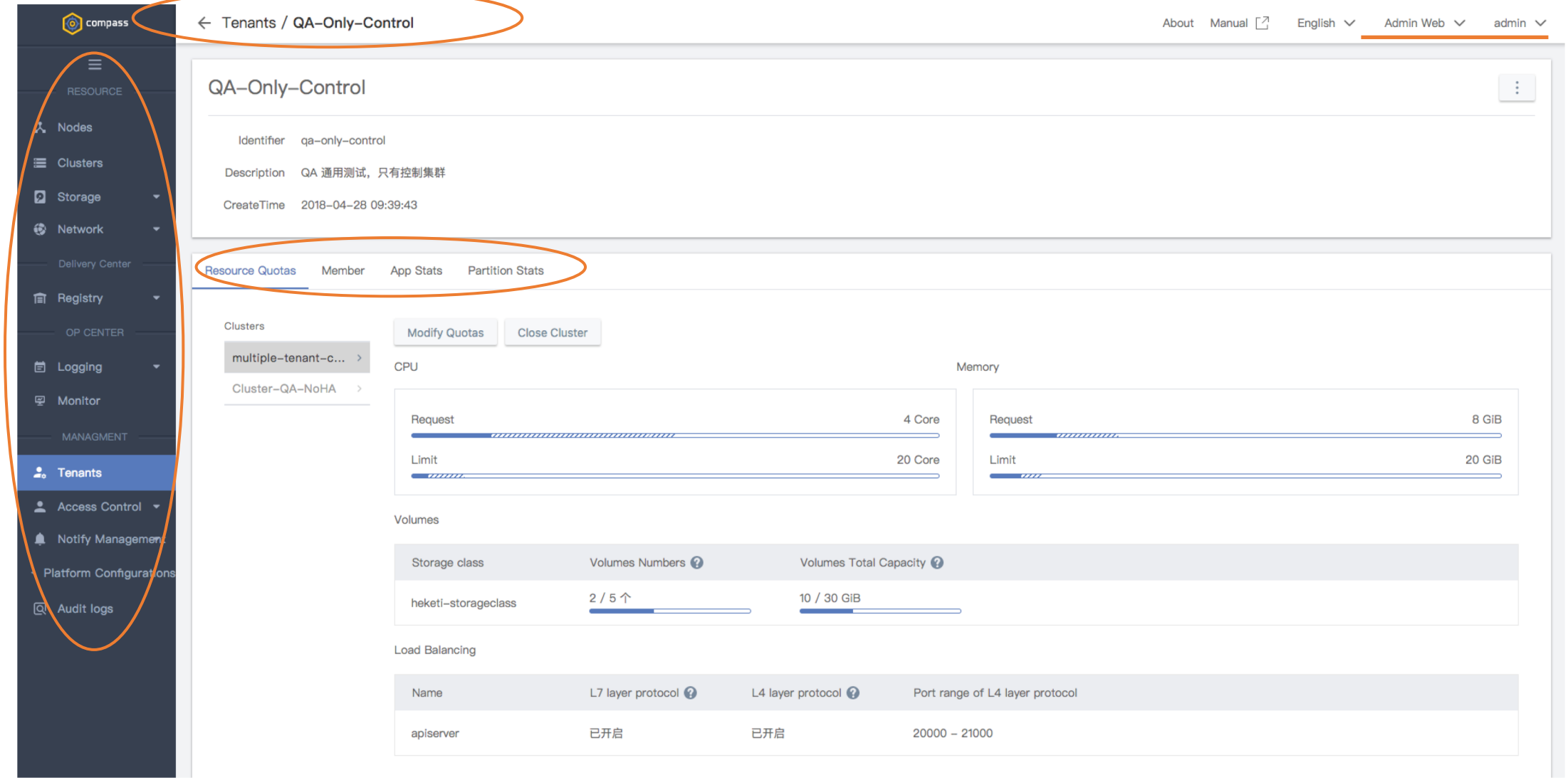
SSO - Workflow



Tenant - Concepts

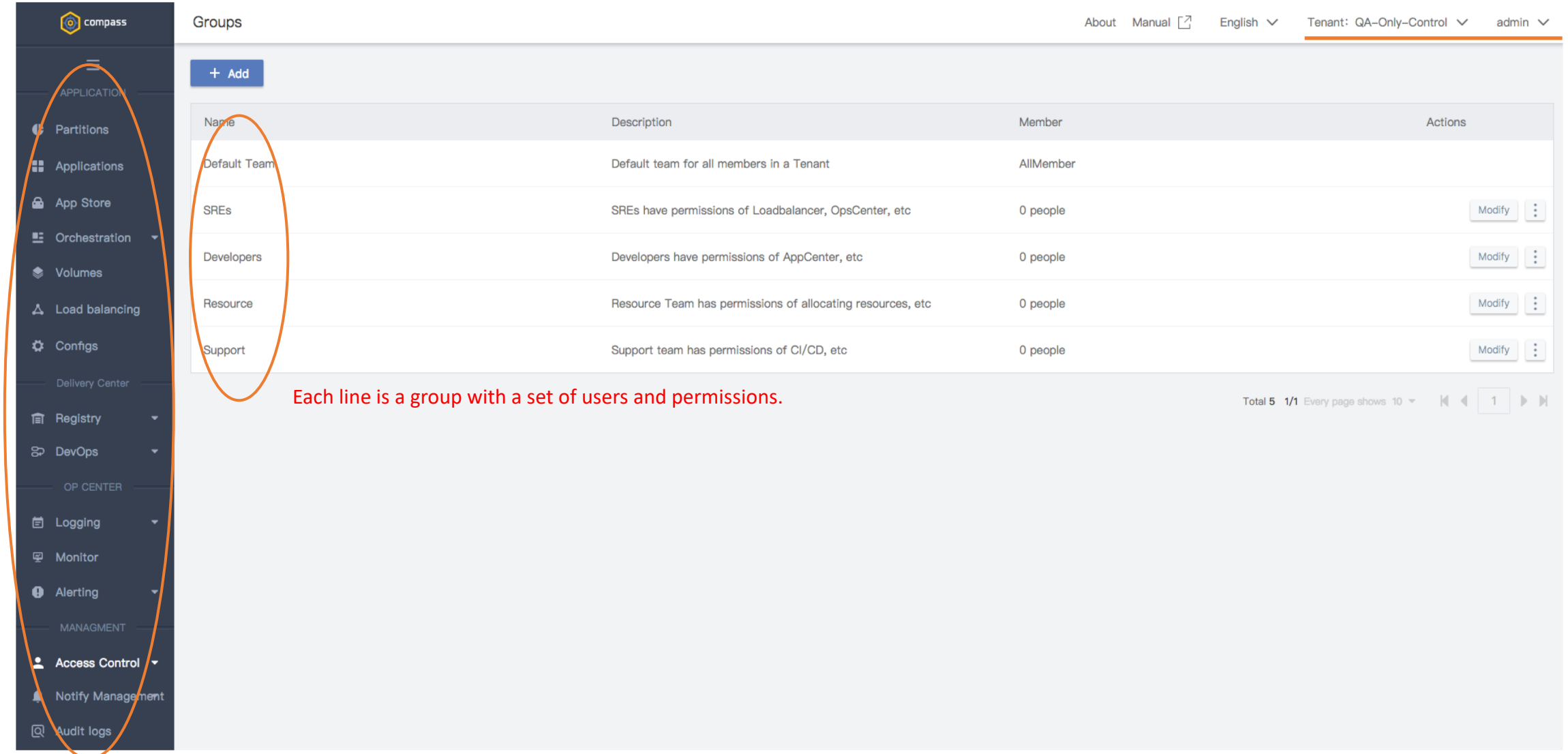
- System Admin
  - manages the whole platform
  - allocate resources (around different clusters) to different tenants
  - assigns roles to different users, e.g. tenant admin, etc
- Tenant Admin
  - manages a single tenant
  - add/delete groups of users
  - allocate resource permissions per group
- Group/team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

Tenant (System Admin)



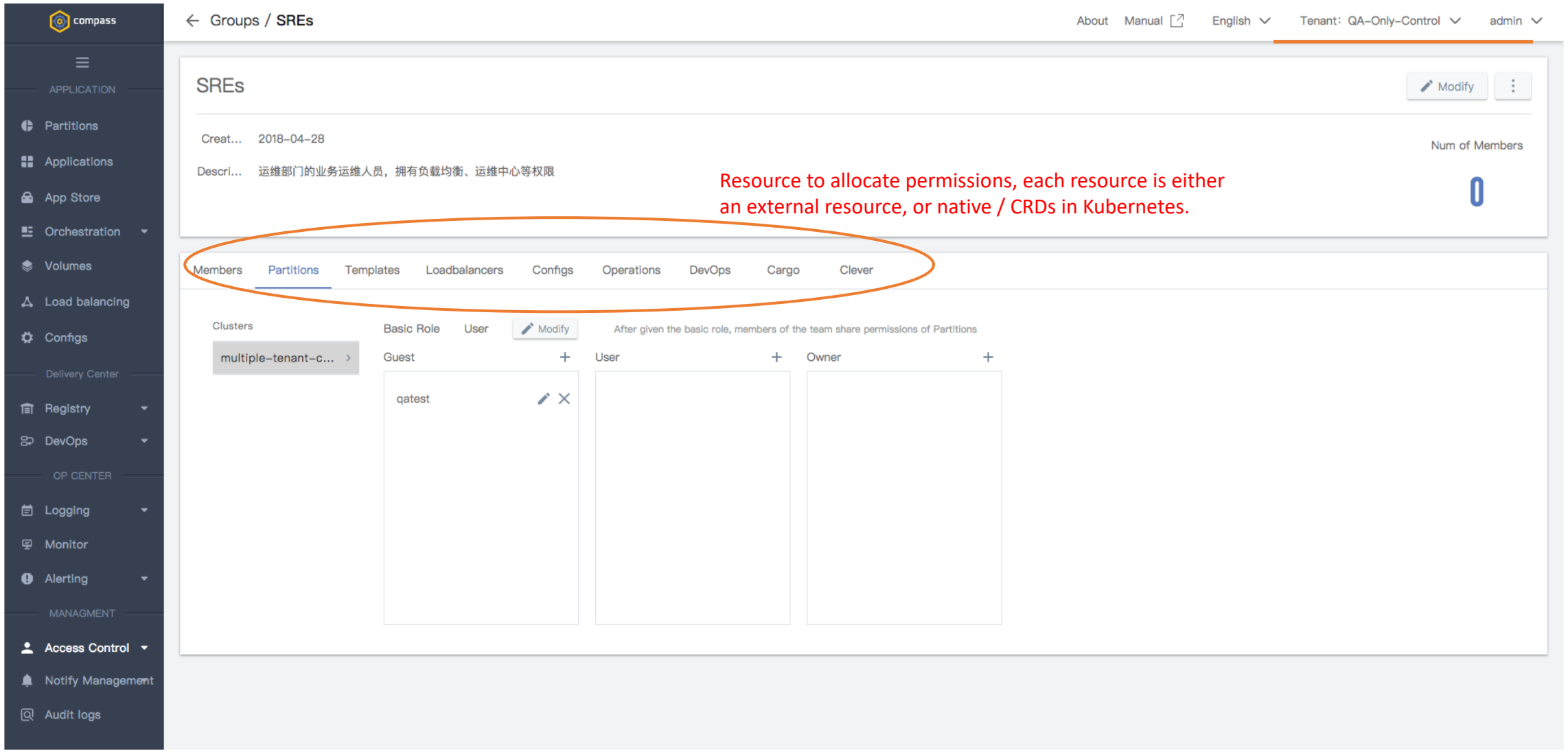
Tenant (Tenant Admins)

KubeCon CloudNativeCon Europe 2018

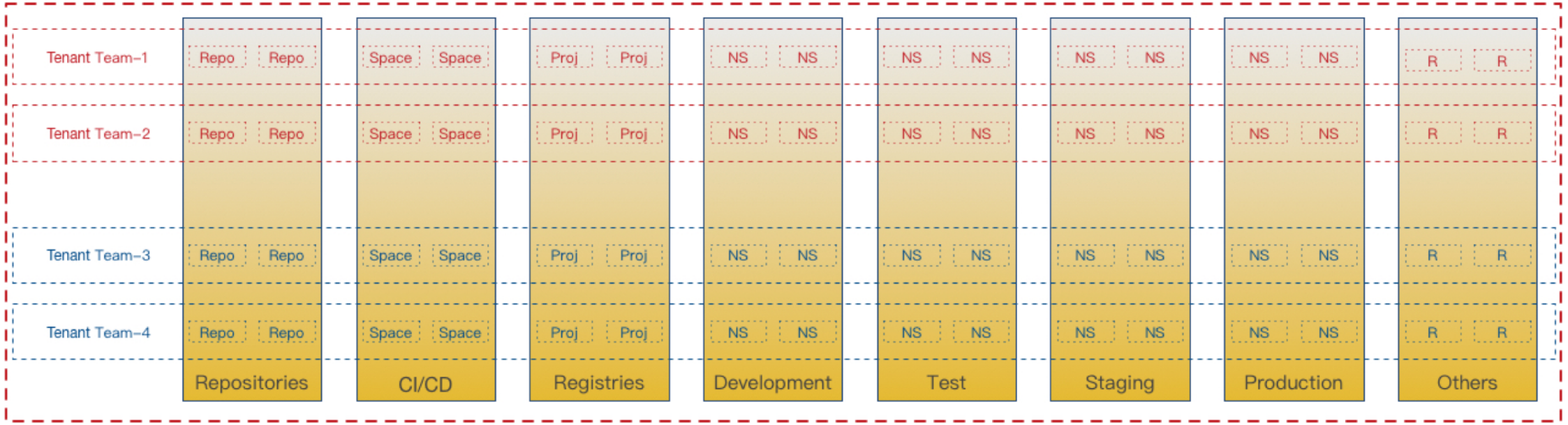


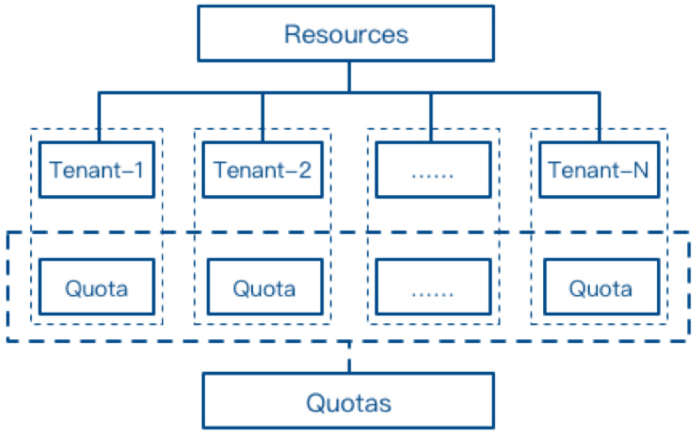
Tenant (Tenant Admins)

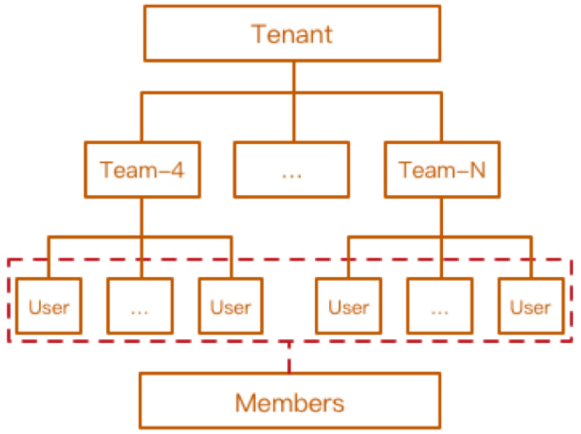
KubeCon CloudNativeCon Europe 2018

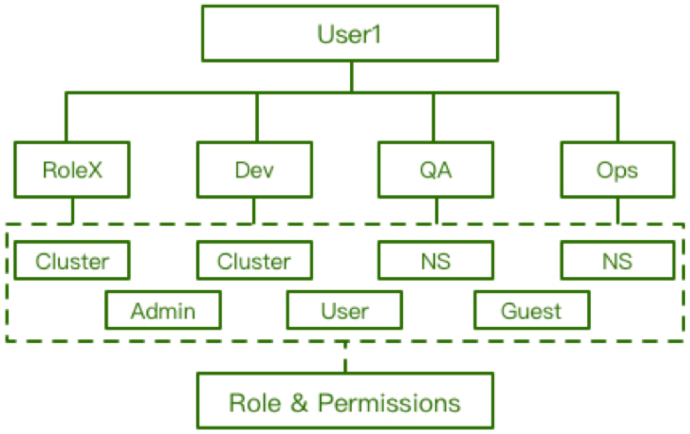


Tenant - Summary



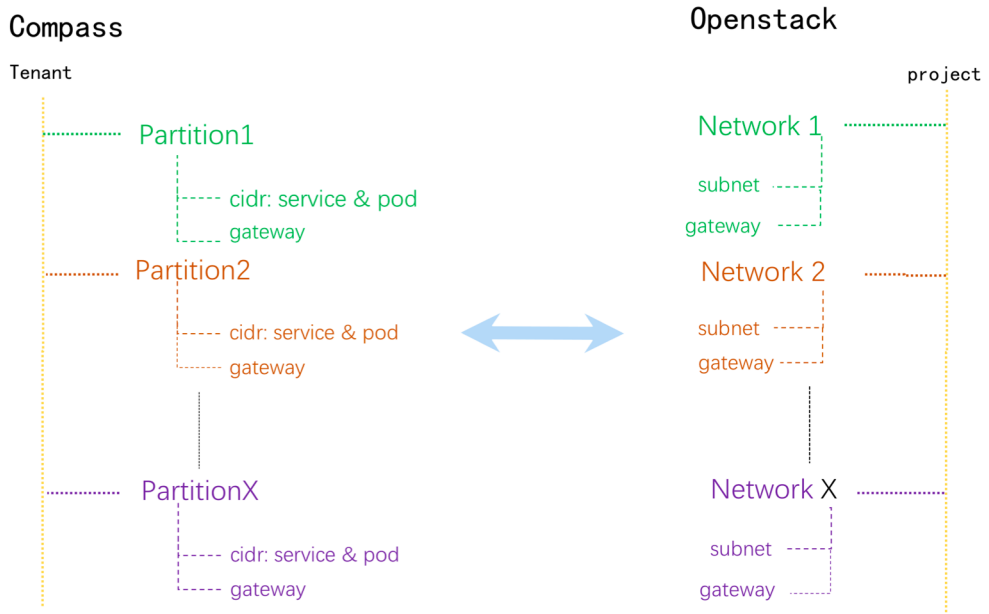




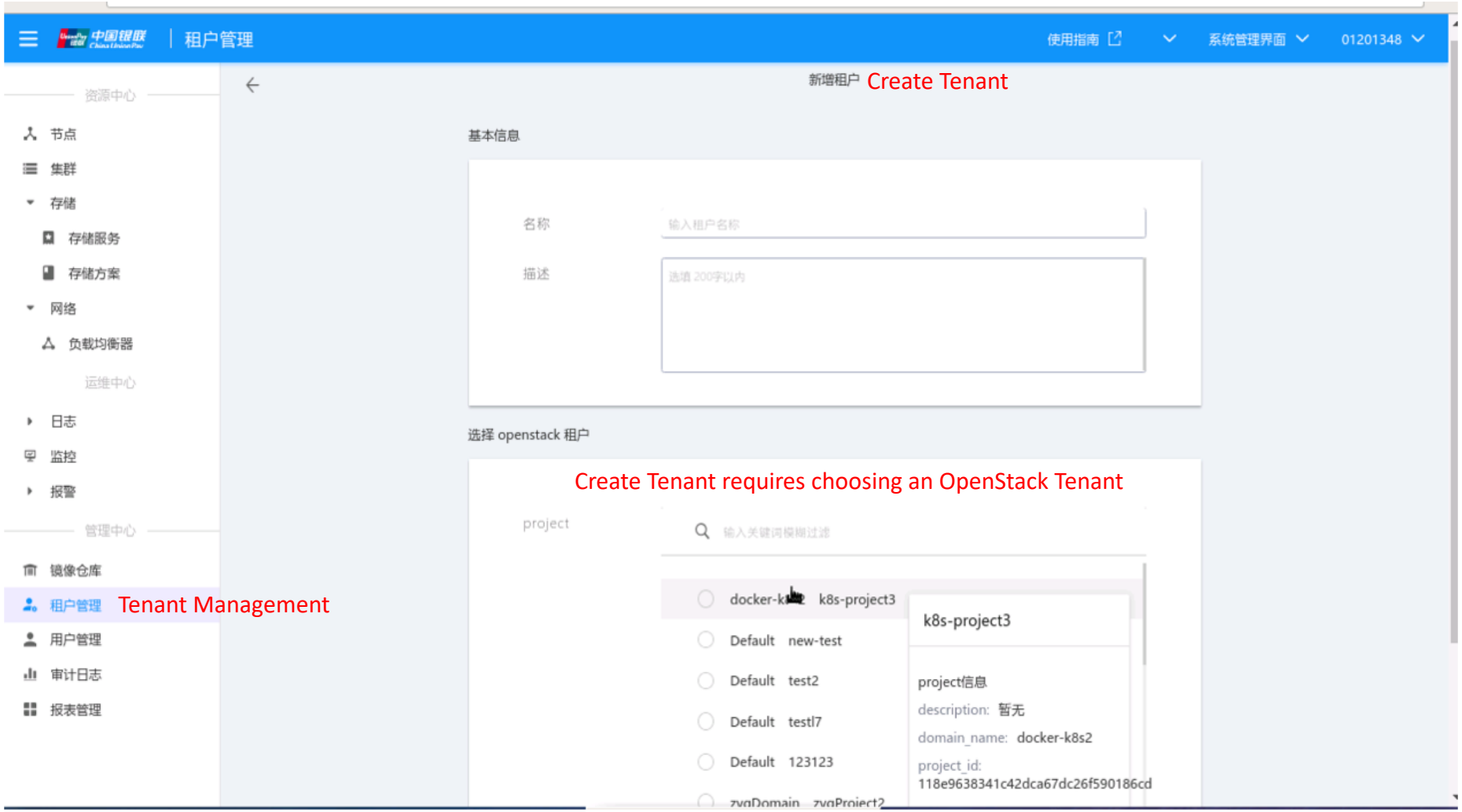


Tenant - OpenStack Mapping

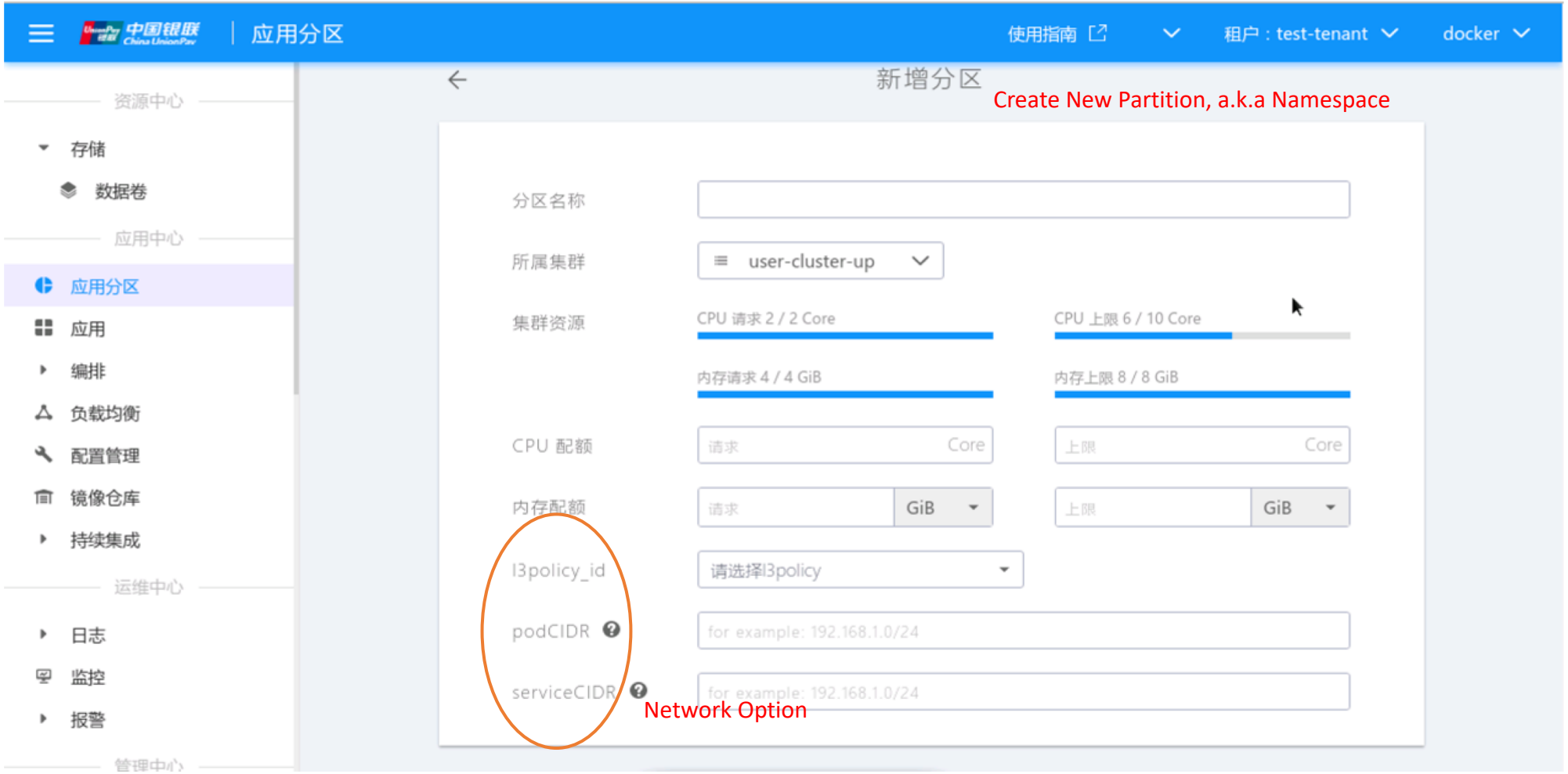
- 1:1 mapping between compass Tenant and openstack Project
- 1:1 mapping between kubernetes Namespace and openstack Network
- 1:1 mapping between kubernetes Pod and openstack Port
- 1:1 mapping between kubernetes Service VIP and openstack LBaaS VIP
- Each namespace has two openstack subnets: PodCIDR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



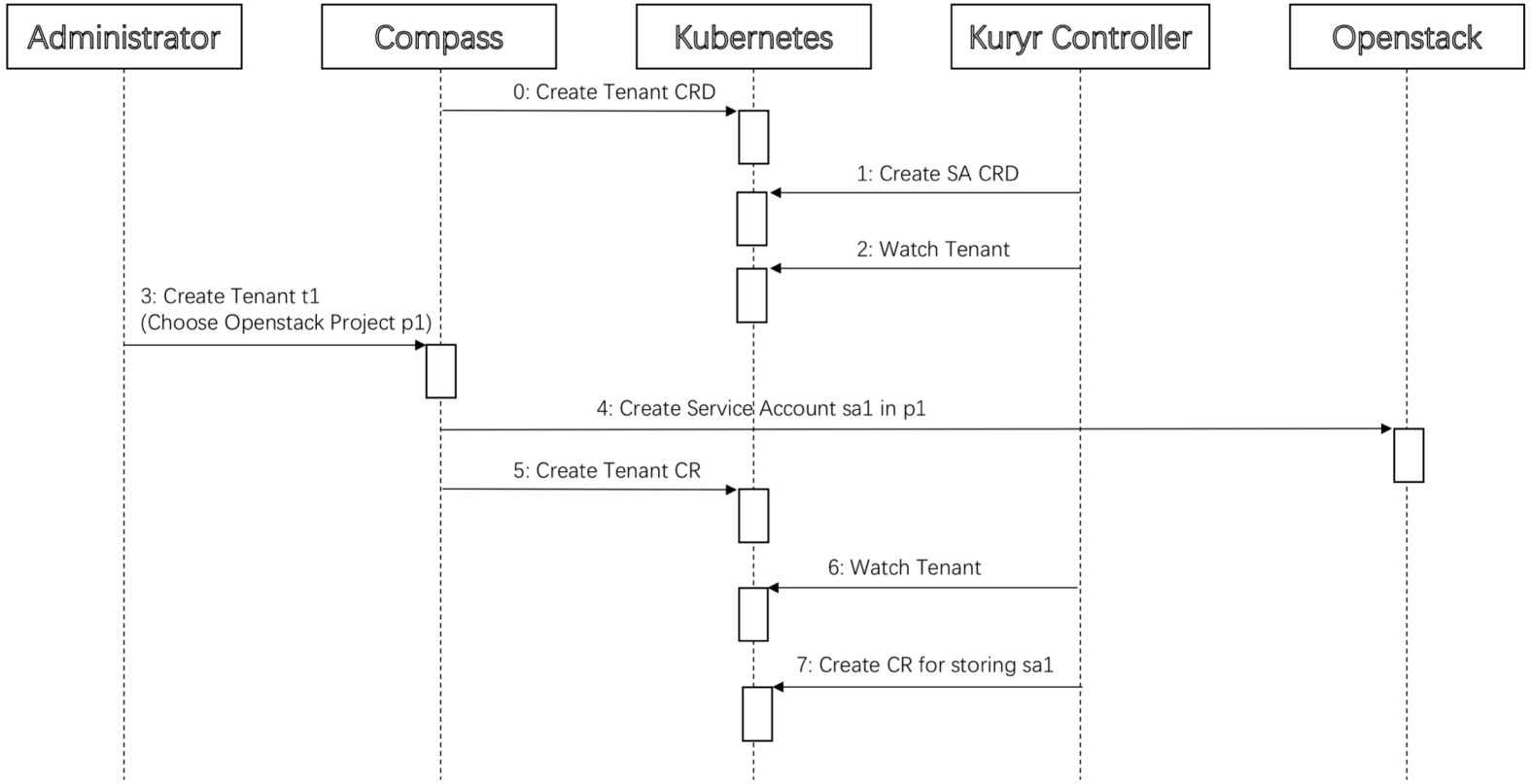
Tenant - with OpenStack



Tenant - with OpenStack



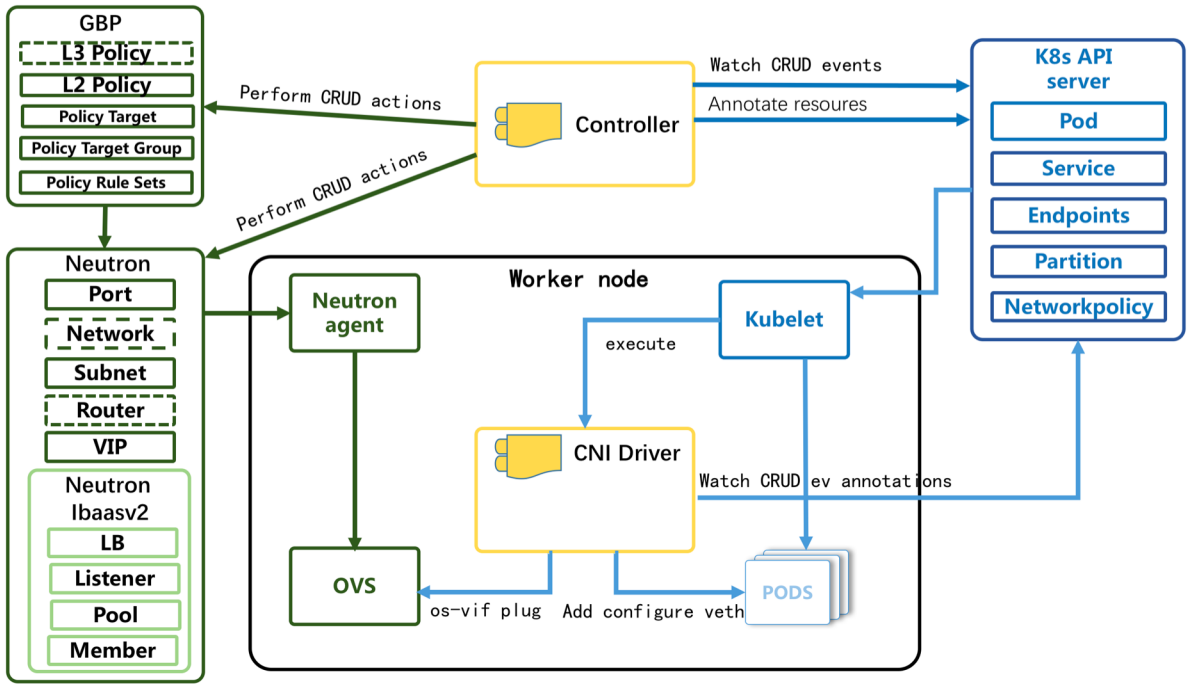
Tenant - with OpenStack



- 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

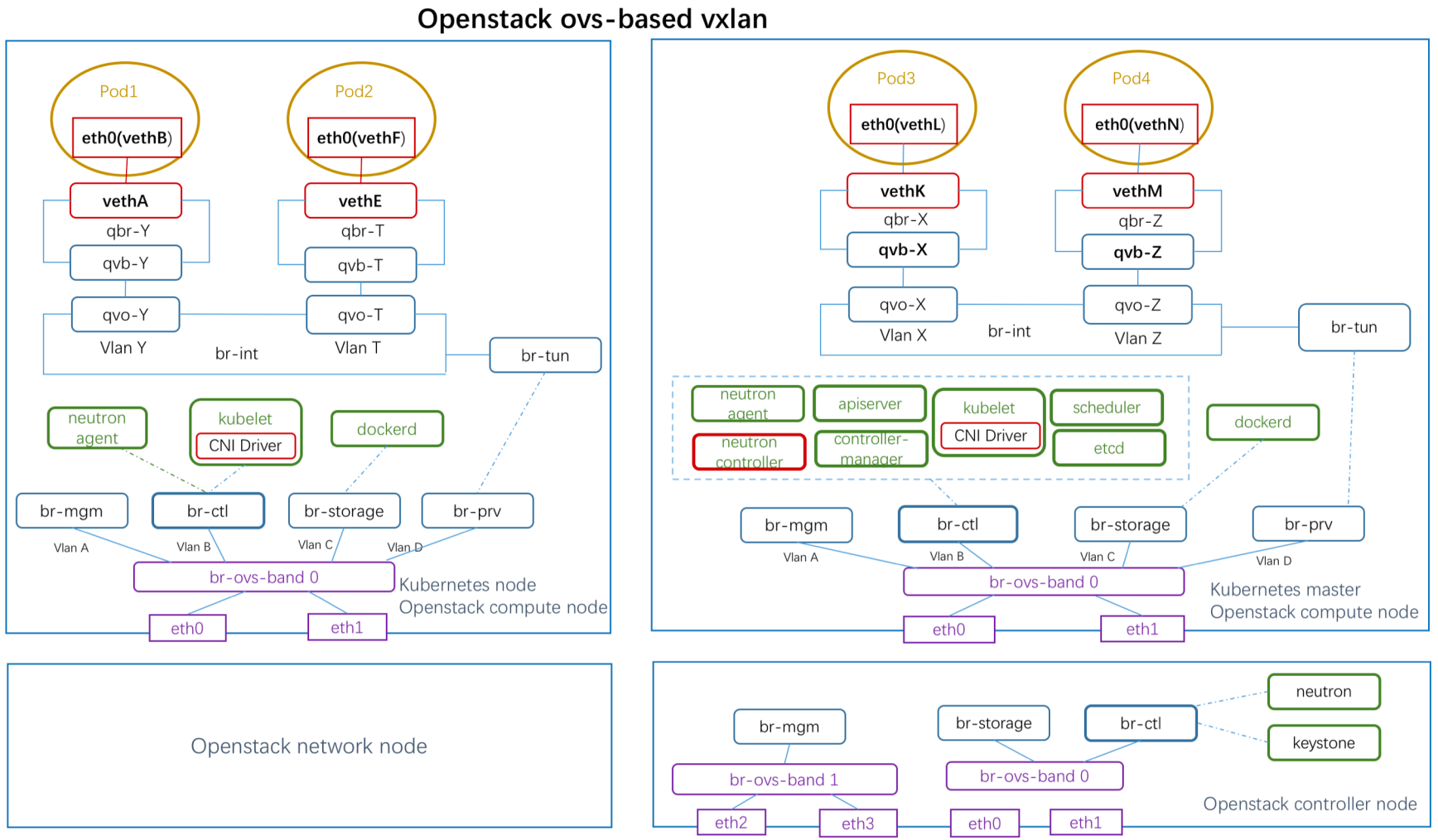
- Requirements
  - Pod should have ‘real’ IP, no NAT
  - Configurable traffic shaping and isolation across tenant
  - Multiple network planes (management, control, business, storage)
- Based on kuryr-kubernetes, with modification:
  - Support multi-tenant
  - Support GBP
- Components
  - Kubernetes Controller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



Network - Multi-planes

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

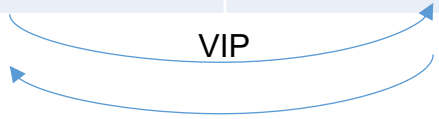
Network - Multi-planes



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	--



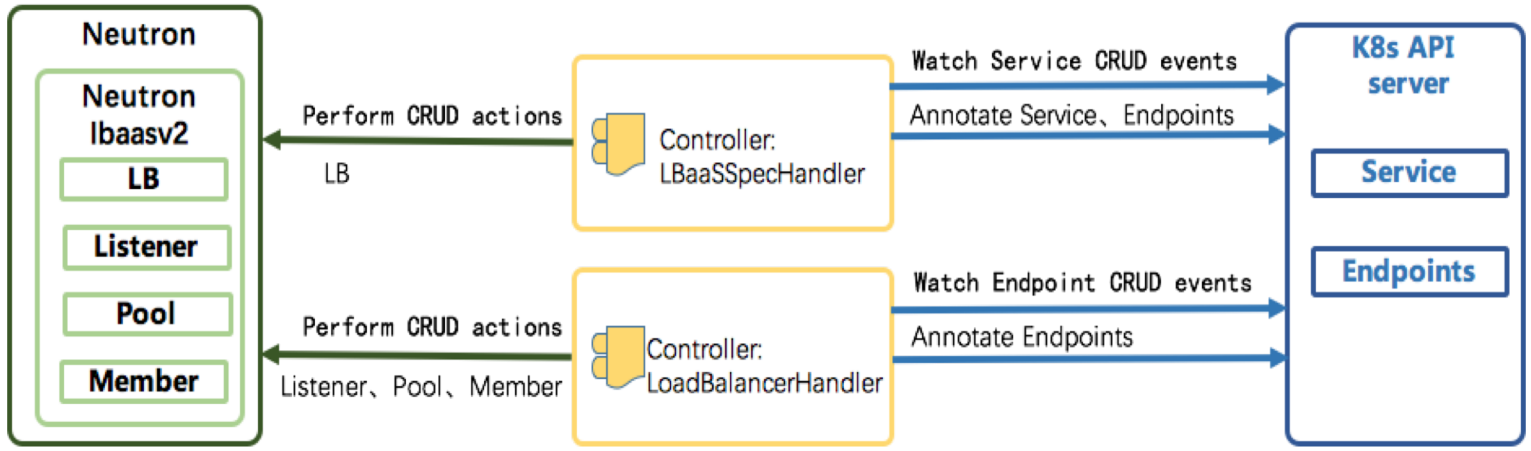
Protocol, Port, Members (Endpoints)

Network - Service

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

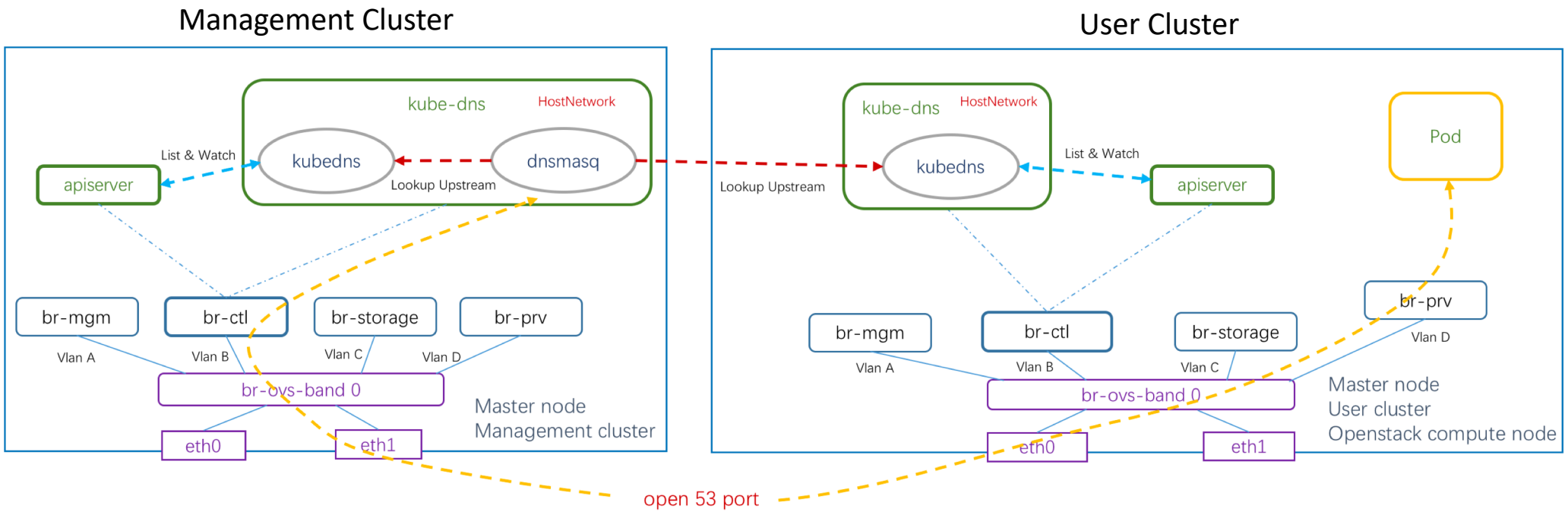


Network - DNS

- 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

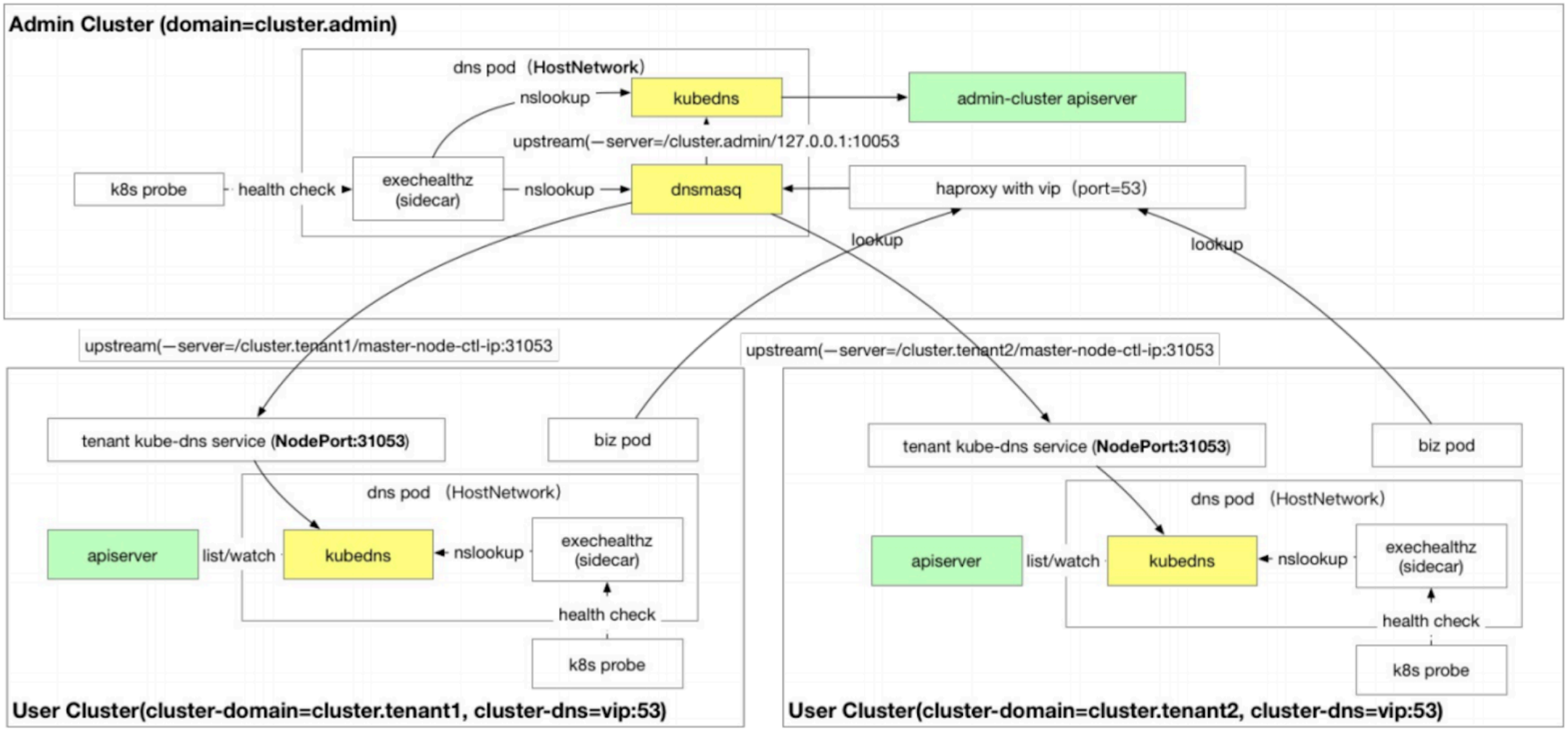
Network - DNS

- 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

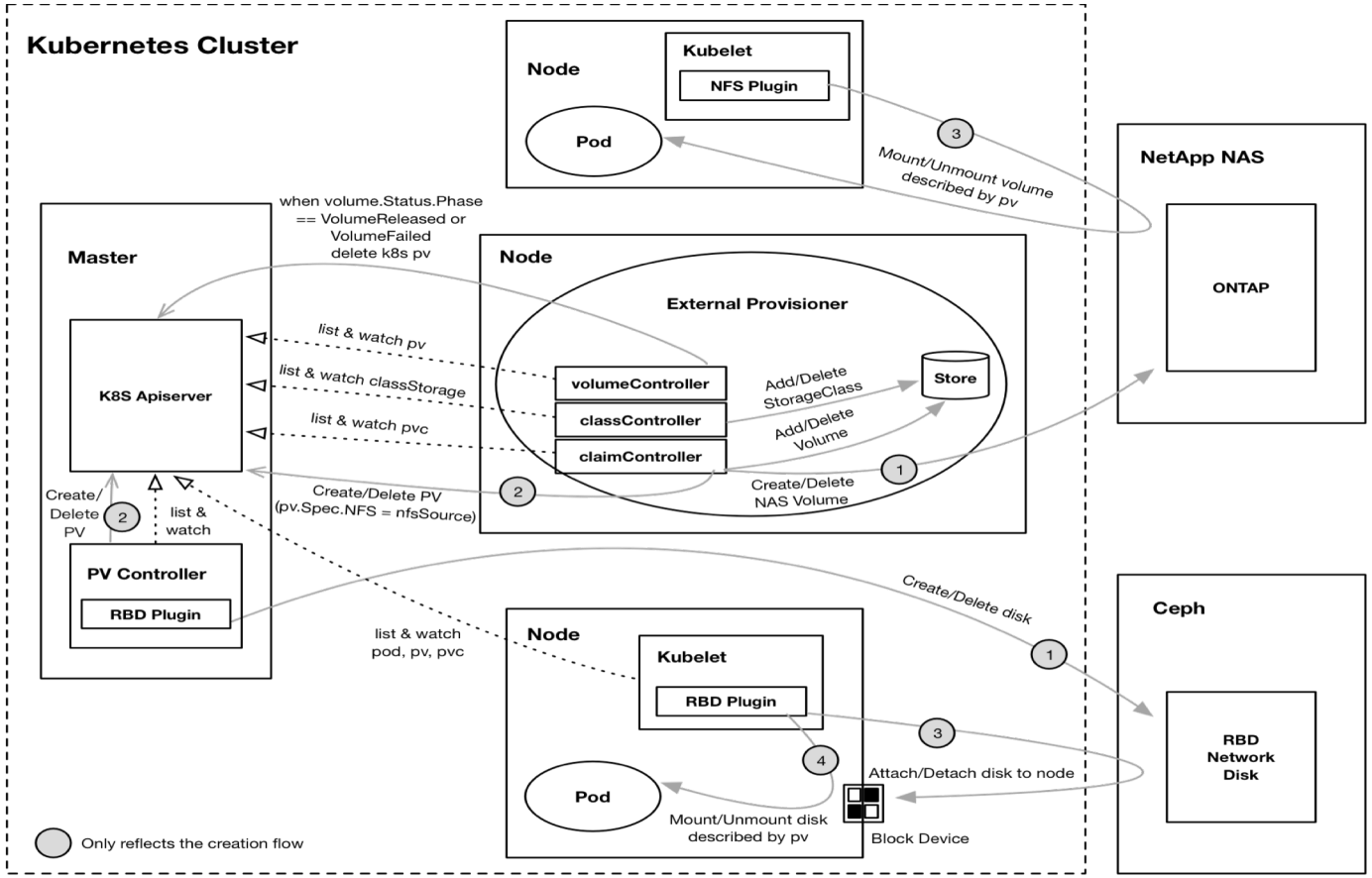
KubeCon CloudNativeCon Europe 2018



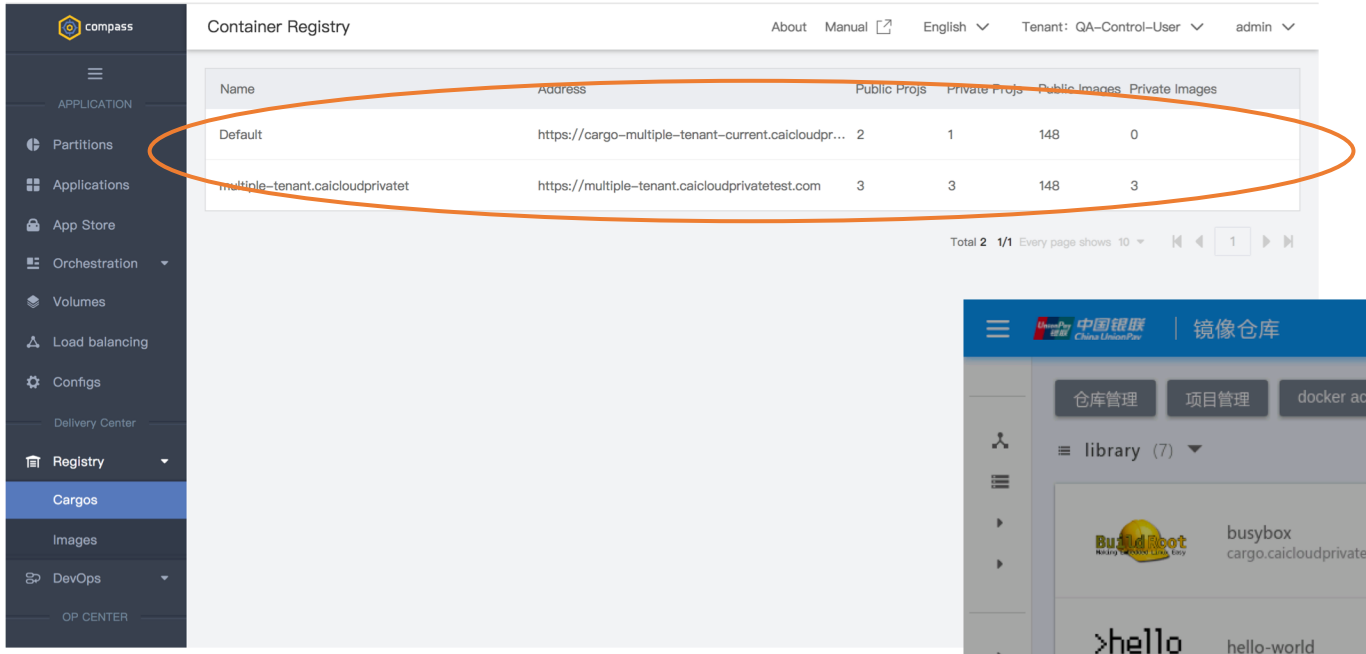
Storage - NAS

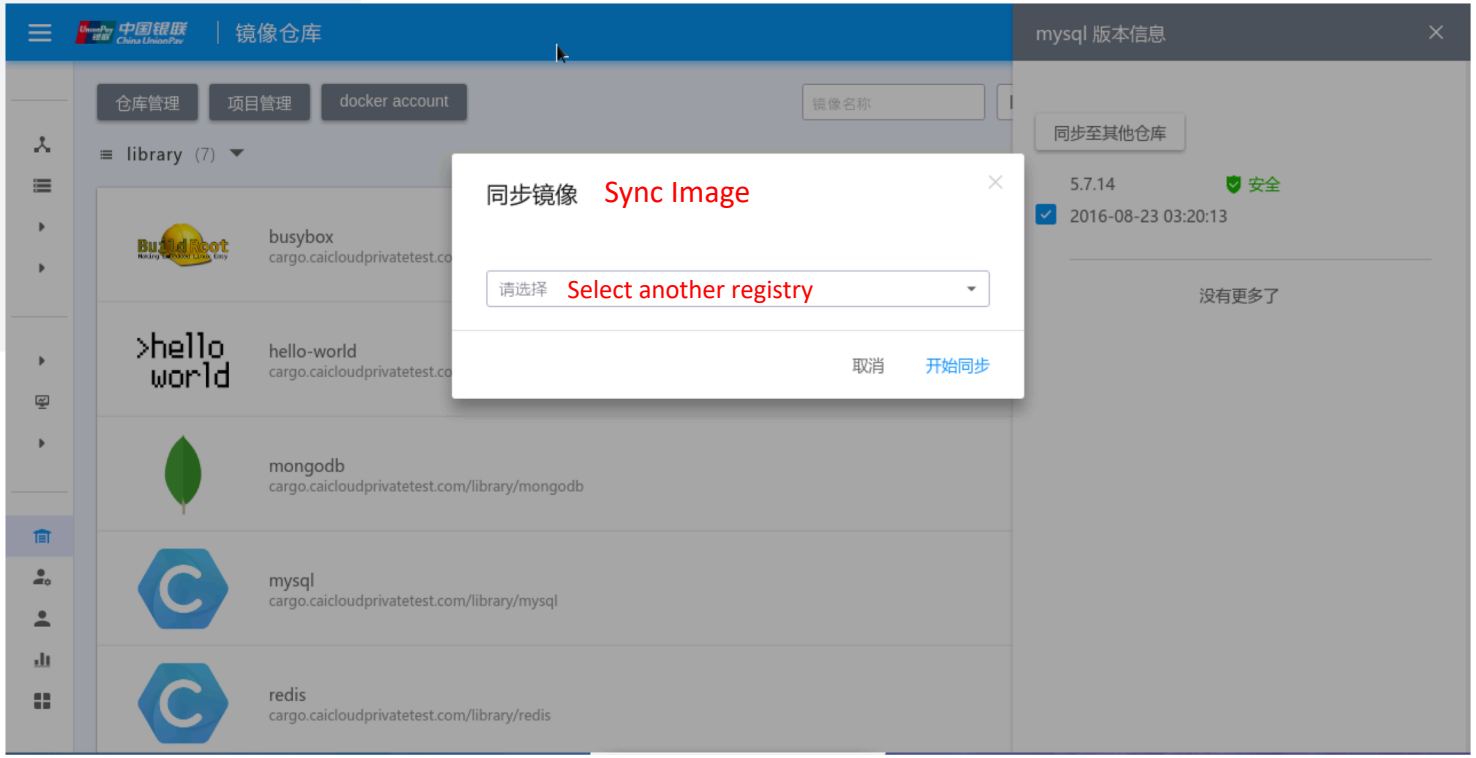
- No out-of-box solution in Kubernetes
- but enterprise users care about data

- Depends on existing environment
  - NetApp NAS: use trident
  - Ceph



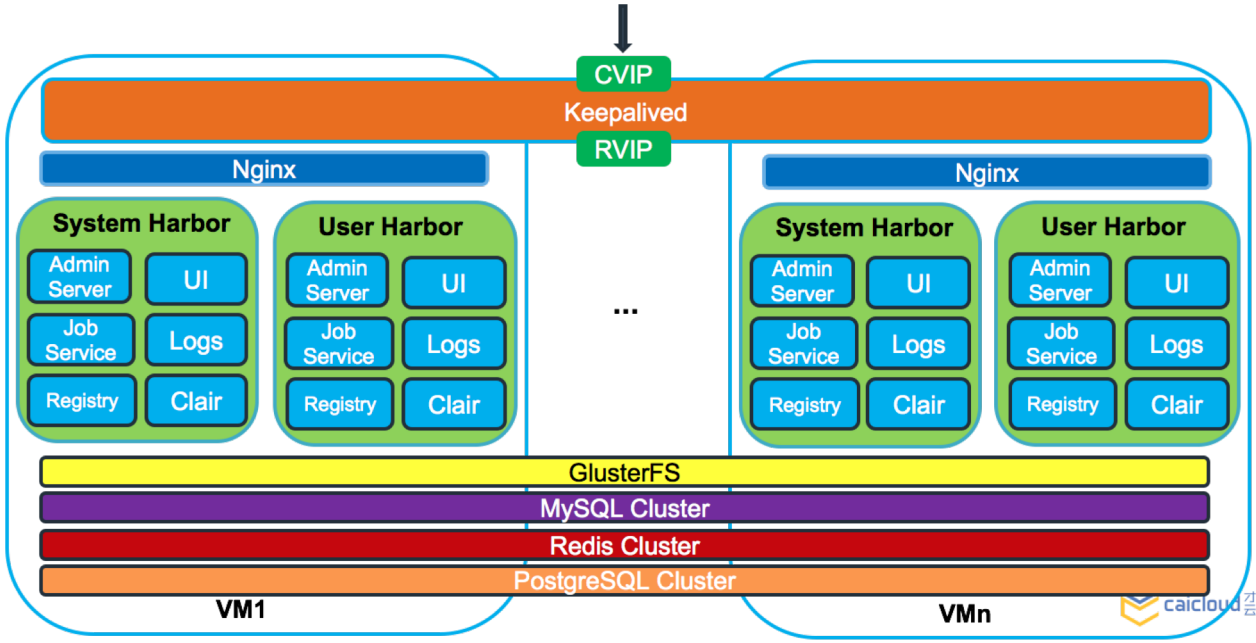
Multi-registries





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