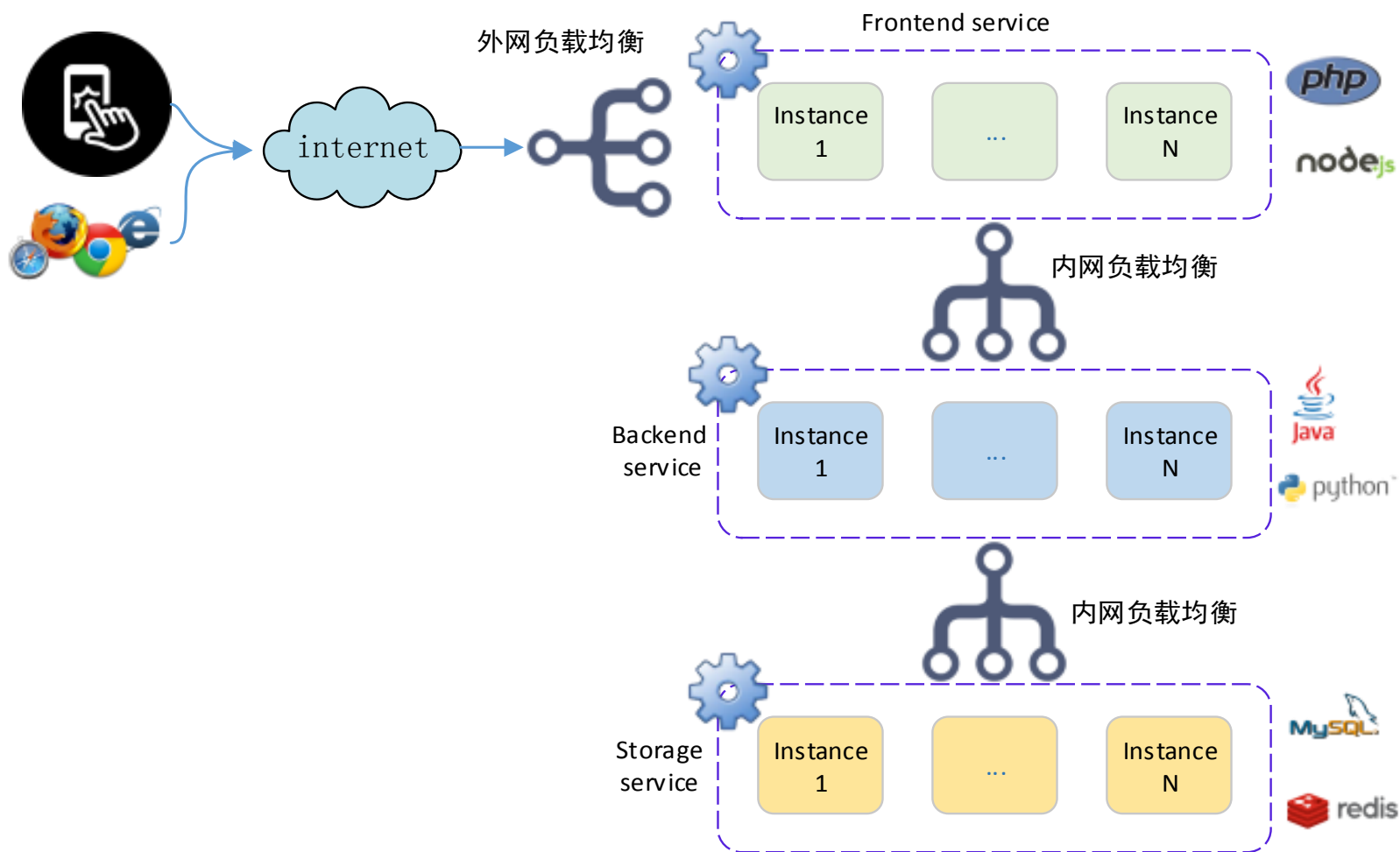


腾讯云容器集群实践

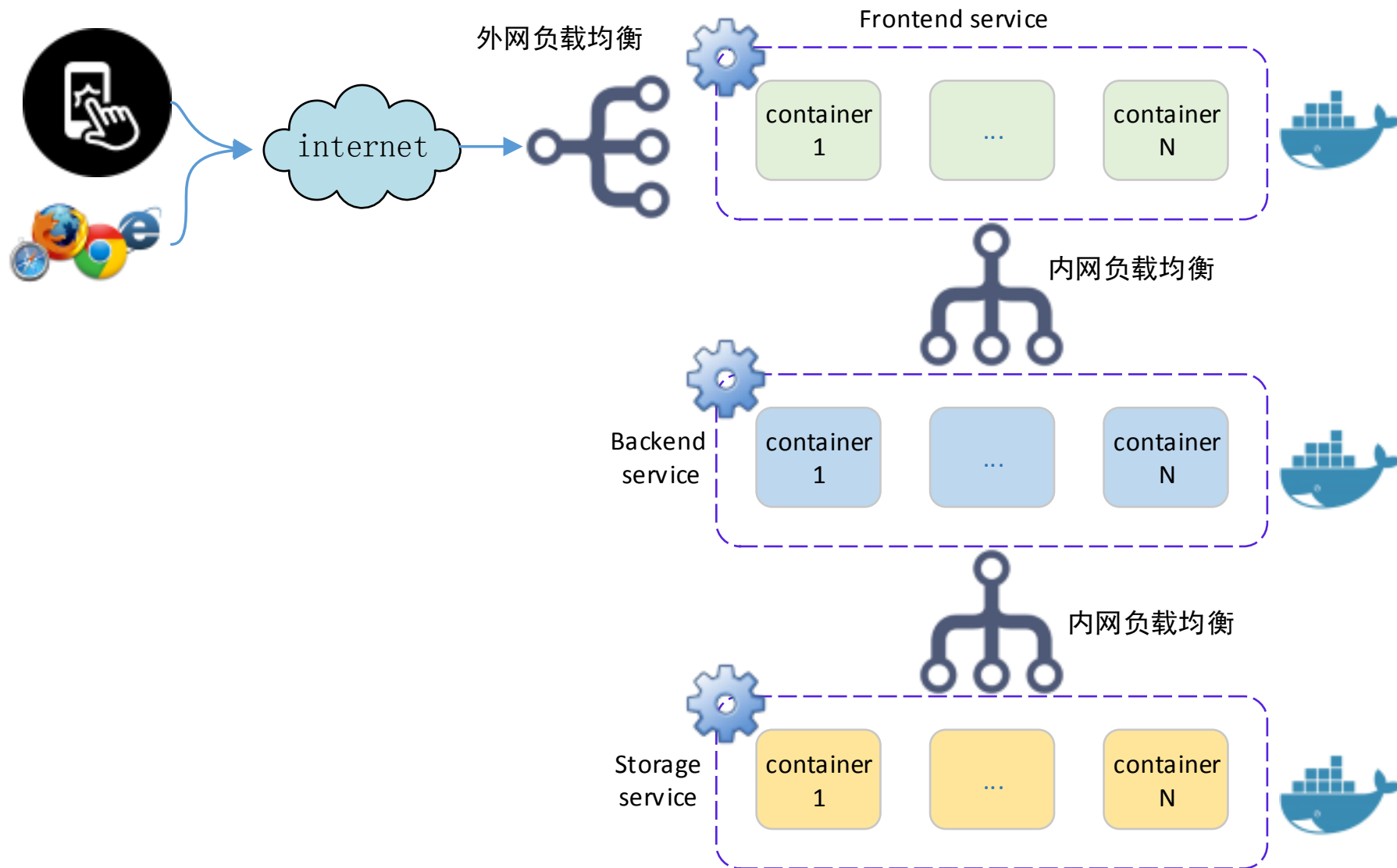
SPEAKER

陈浪交

典型的集群

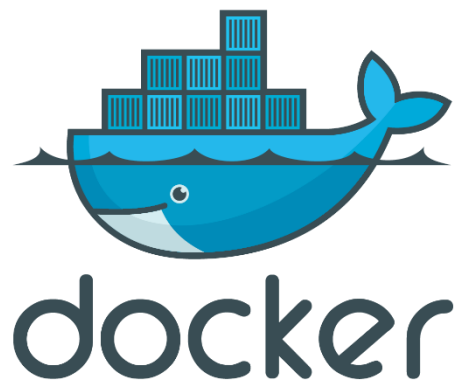


通用的容器集群



通用的容器集群

- ❑ 支持任何linux应用
- ❑ 负载均衡、服务发现、监控、log服务支持
- ❑ 用户更少地关心基础设施，更多地关心自己的业务
- ❑ 服务之间独立，依赖关系由用户自定义
- ❑ 安全：租户隔离、网络隔离、镜像隔离



没有集装箱，就不会有全球化 -- 《经济学人》

QCloud容器集群

- ❑ 用户可以在QCloud上创建基于vpc和cvm的docker集群，集群基于Kubernetes
- ❑ 支持图形界面以及api两种方式来操作集群、节点、服务、容器等集群资源
- ❑ 简单快捷地发布、升级、回滚服务
- ❑ 提供集群、容器级别的可视化监控
- ❑ 支持auto-scaling
- ❑ 智能化容器调度

QCloud容器集群

IaaS层支持


- ❑ 集群主机管理
- ❑ 可规划的私有网络
- ❑ 容器调度
- ❑ 容器分组、丰富的调度策略
- ❑ 容器生命周期管理
- ❑ 容器ip分配、跨主机的容器路由管理
- ❑ 容器动态扩缩容(auto scaling)
- ❑ 容器卷管理
- ❑ 内外网负载均衡
- ❑ 服务发现

PaaS层服务

- ❑ 集群租户隔离
- ❑ 集群用户操作面板
- ❑ 集群资源管理
- ❑ 服务平滑升级
- ❑ 集群API接口
- ❑ 企业级镜像仓库服务



Kubernetes支持

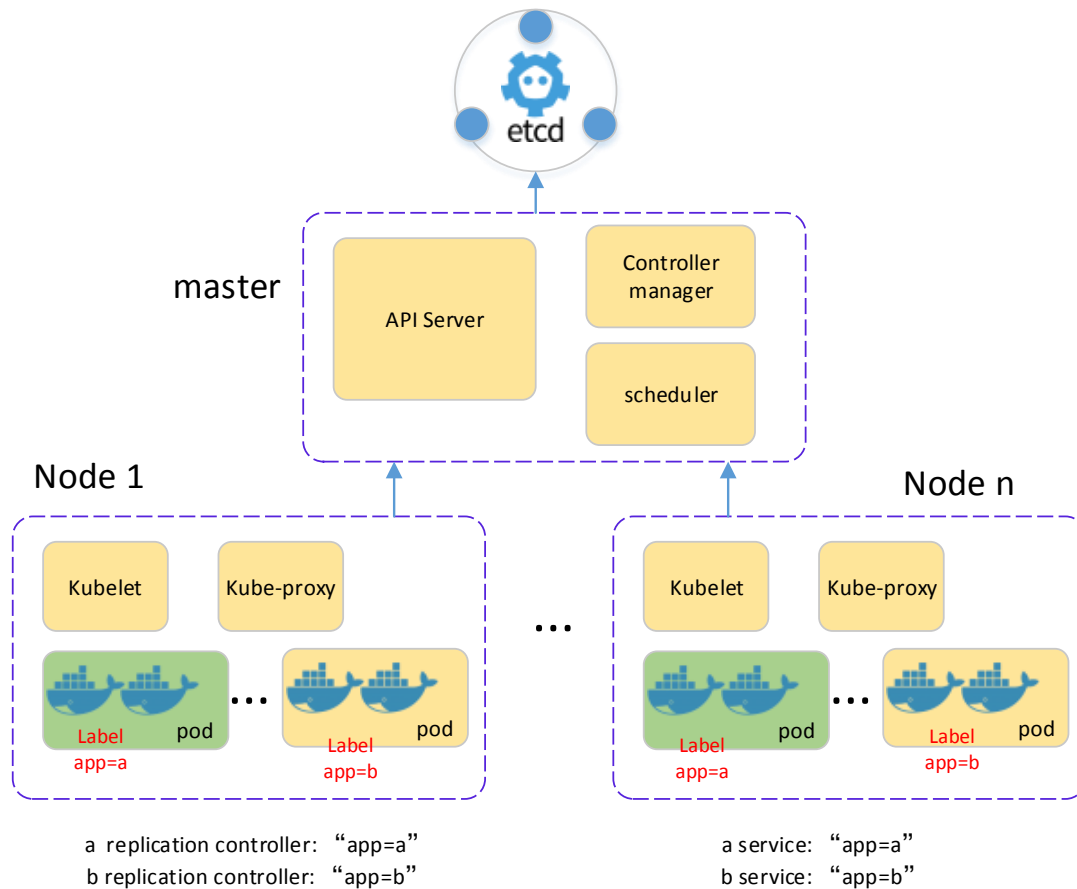


需要适配
Kubernetes来支持



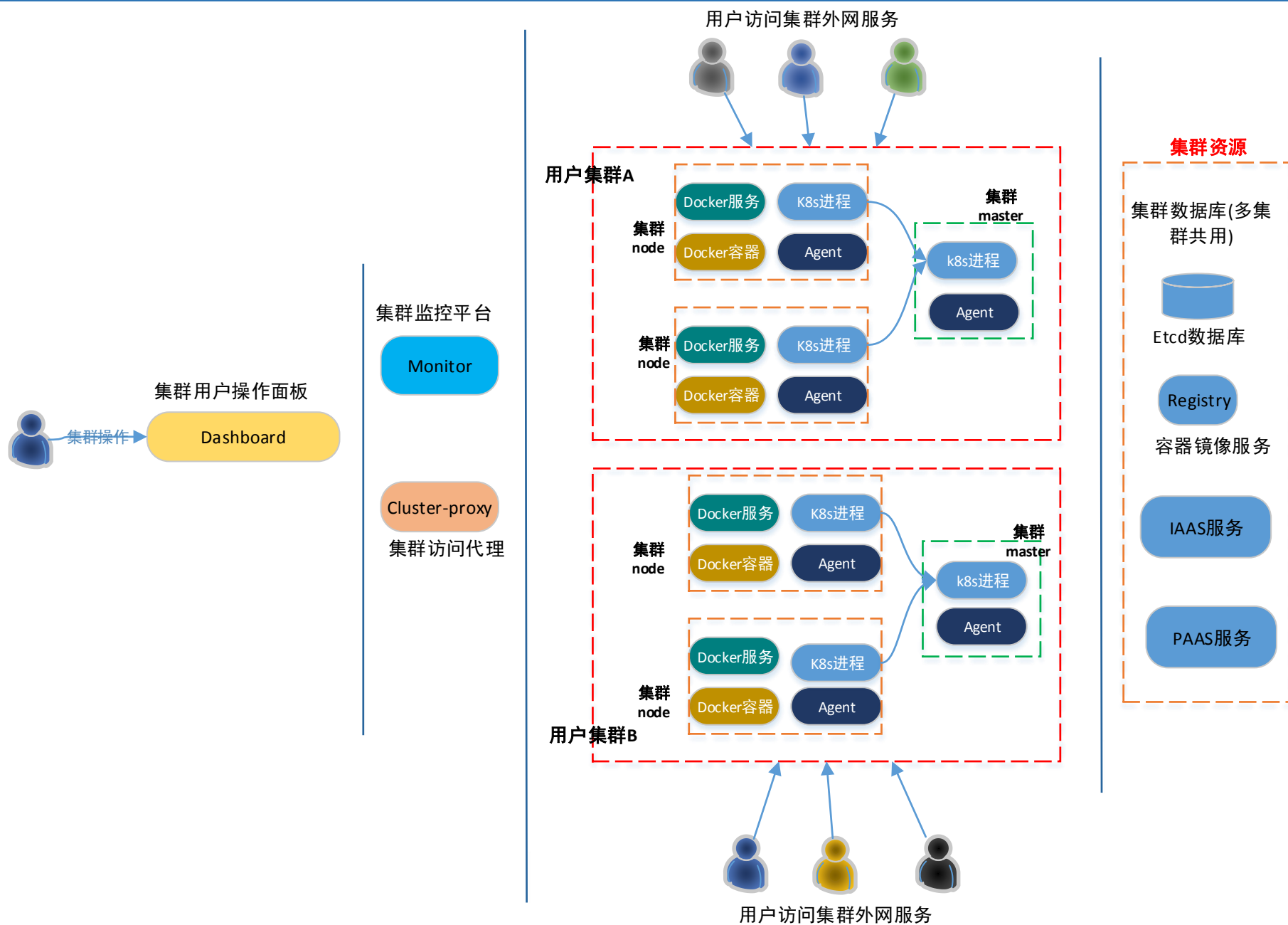
需要我们独立开发

Kubernetes简介



Kubernetes常用名词	说明
label	标签
pod	容器组，Kubernetes的基本资源
replication controller	Pod的副本控制器
service	Pod的流量访问控制器
node	集群节点，运行pod
master	集群master，负责资源的管理、调度
etcd	集群资源持久化db

QCloud容器集群-总体架构



Kubernetes网络

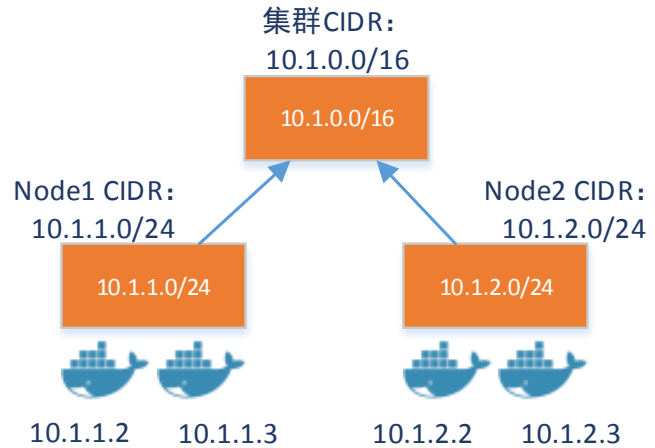
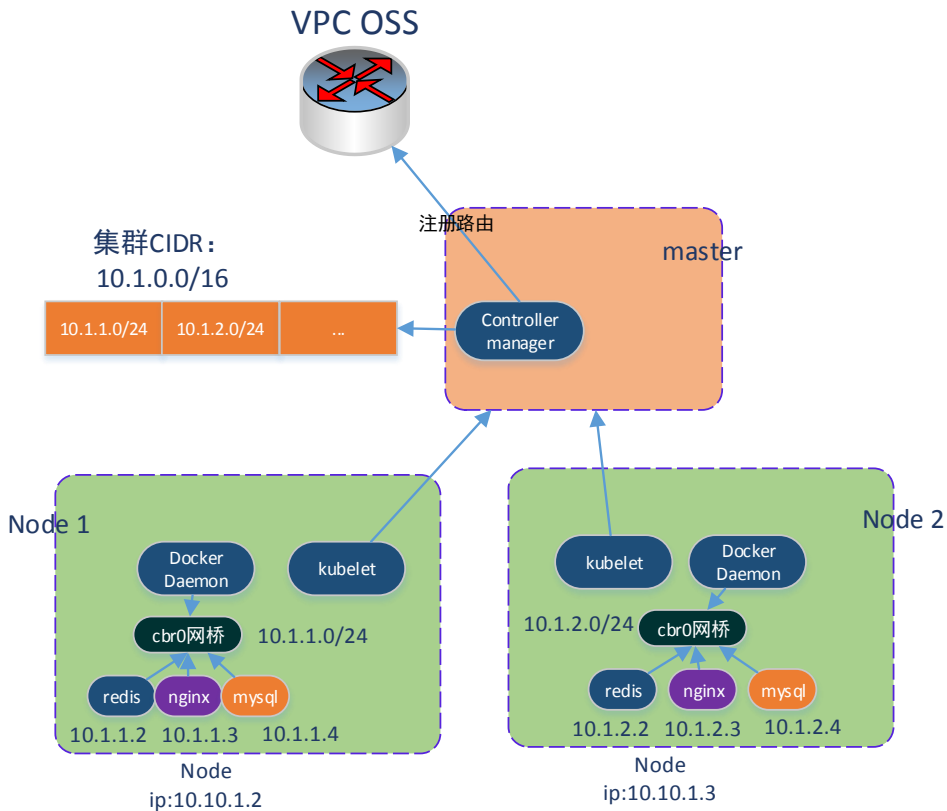
宏观

- ❑ 用户通过互联网访问集群内的外网服务
- ❑ 集群内部服务间互访
- ❑ 集群容器访问集群外甚至外网
- ❑ 集群外网负载均衡
- ❑ 集群内网负载均衡
- ❑ 服务发现

微观

- ❑ 容器ip分配
- ❑ 容器路由
 - 集群内路由
 - 集群外路由
- ❑ iptables

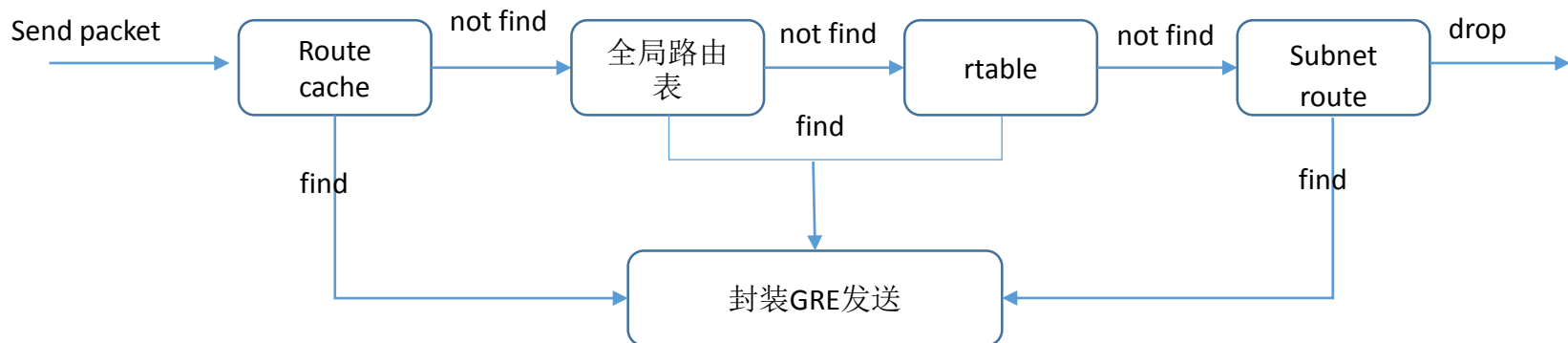
Kubernetes容器网络



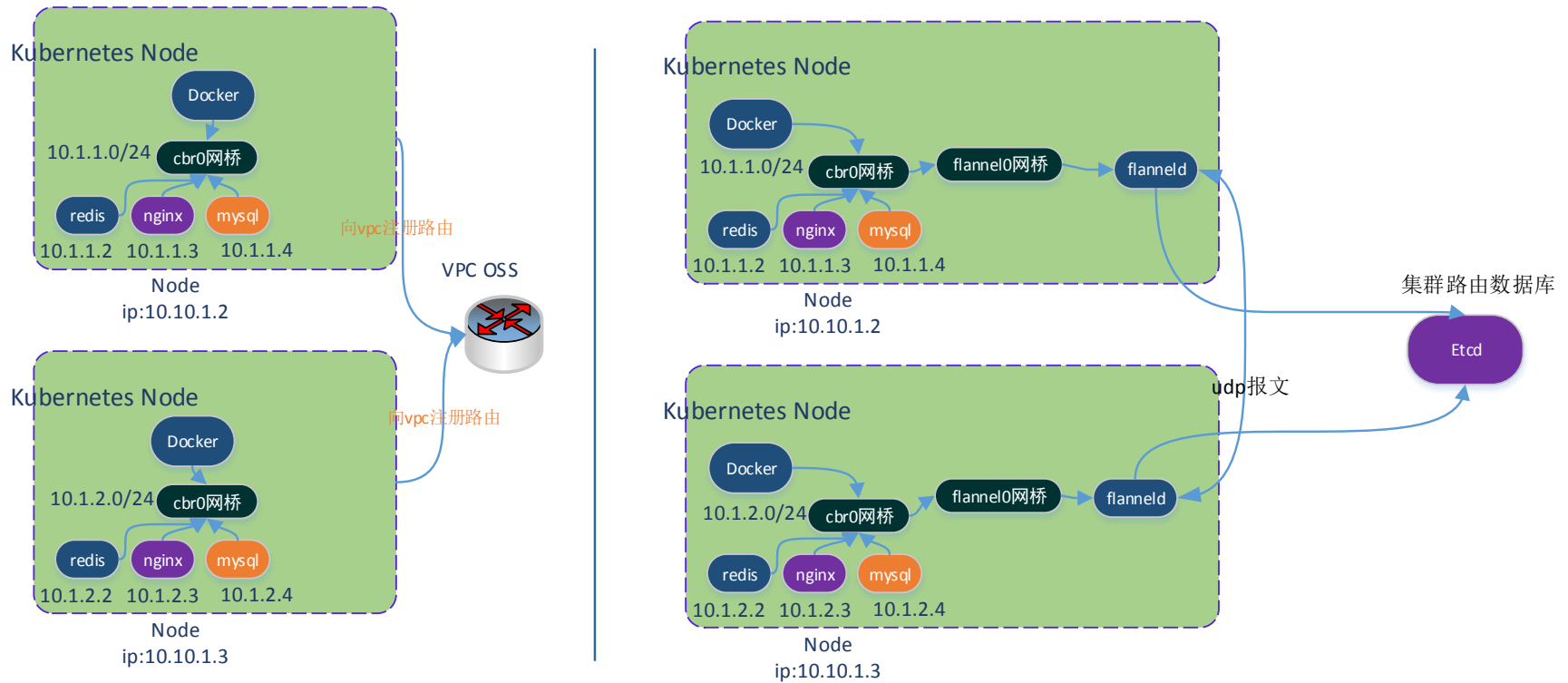
- ❑ 为docker daemon指定网桥
 - --bridge=cbr0
- ❑ Pod的ip由docker daemon分配
- ❑ Pod间跨主机互访不需要NAT
- ❑ 扁平化网络，主机、容器间对等互访
- ❑ Kubernetes提供的网络插件cni
 - 容器路由
 - 网桥
 - veth

QCloud vpc

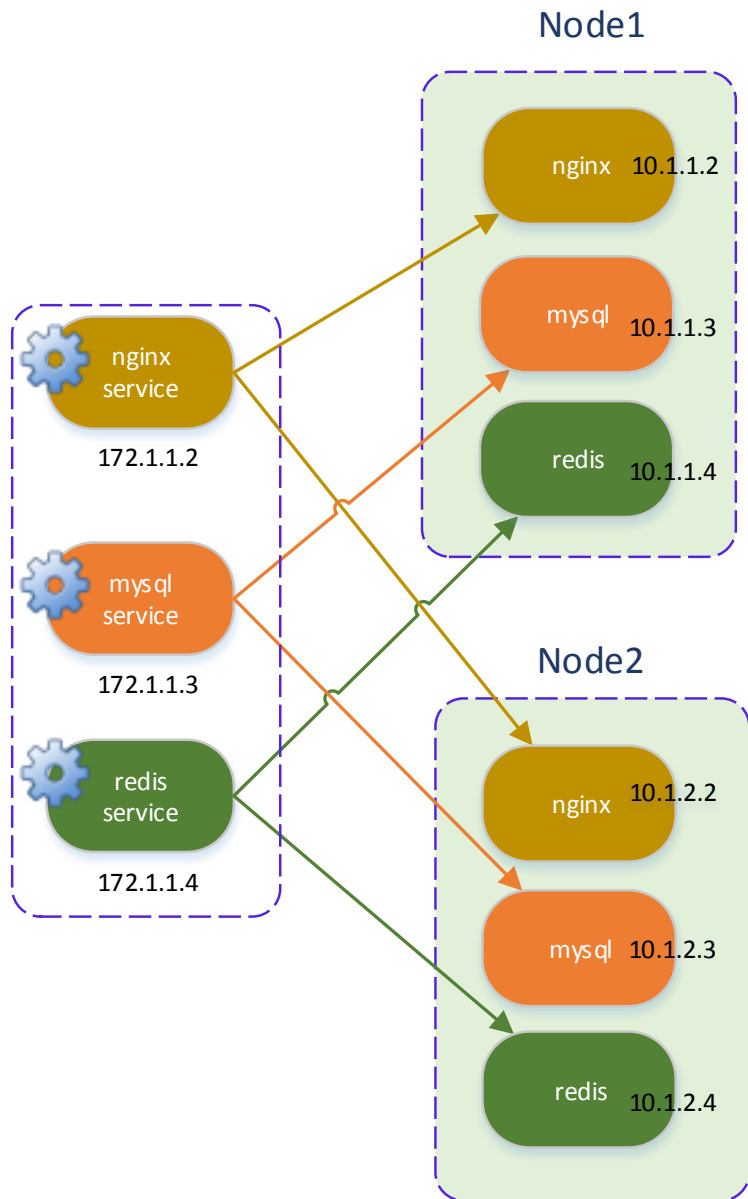
- ❑ 网络隔离
- ❑ 用户可自定义网络、管理子网、路由表和网关
- ❑ 支持 IPsec VPN , SSLVPN
- ❑ 支持vpc之间网络打通 , vpc与基础网络互访
- ❑ 支持vpc内路由动态注册
- ❑ 支持按子网路由



容器网络-对比flannel



kubernetes-内网负载均衡

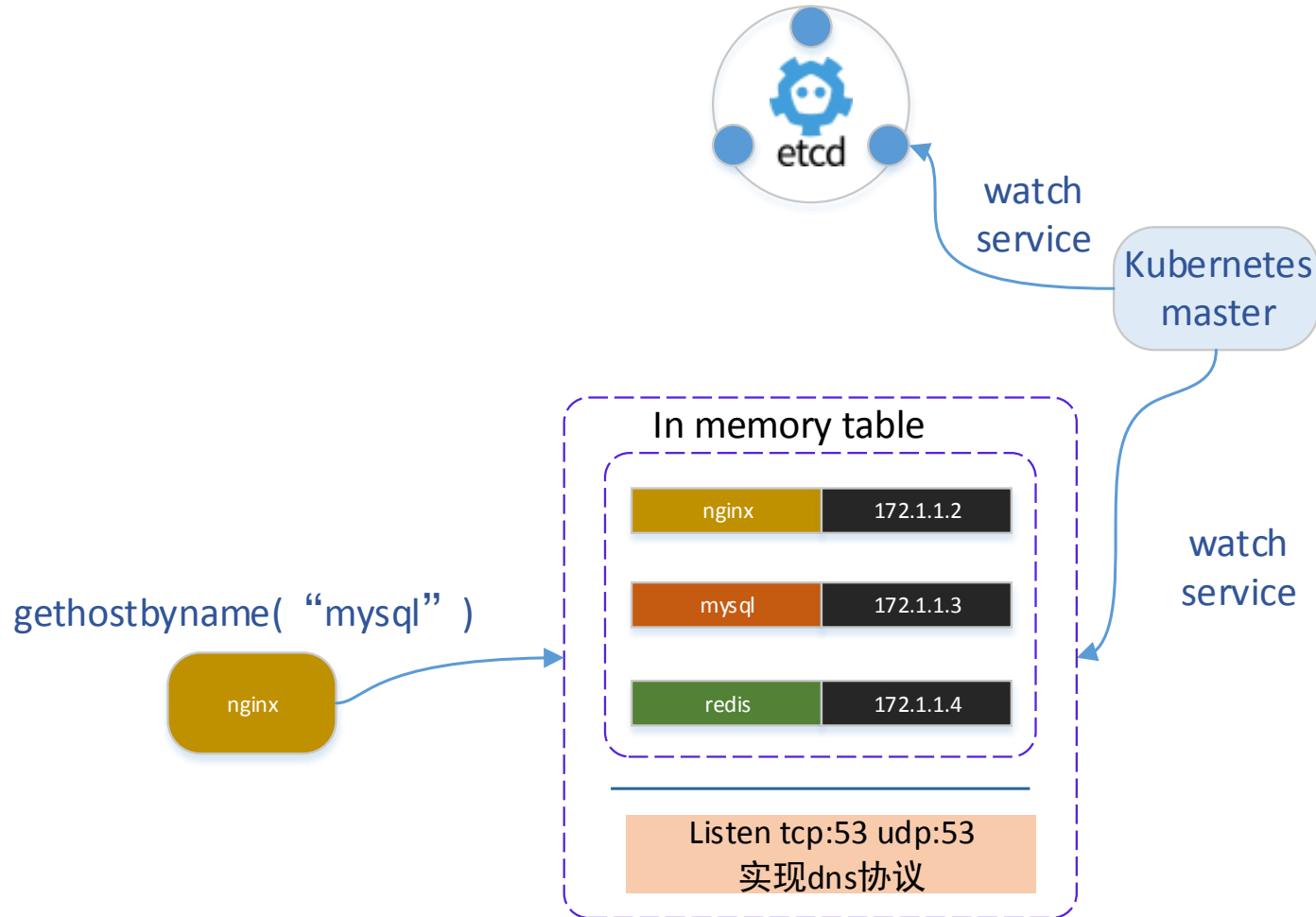


- ❑ Kubernetes为集群内每个service分配一个service virtual ip
 - ❑ Kubernetes为集群内每个node上创建针对每个service virtual ip的iptables规则
 - ❑ 访问某个服务时，只需要指定virtual ip，iptables会做DNAT转化，随机挑选一个该服务下的容器ip
- `sock_conn(“172.1.1.3”, 3306) //nginx容器访问mysql服务，指定mysql virtual ip`
`-> sock_conn(“10.1.1.4”, 3306) //DNAT后，会连接到其中一个mysql容器`

service virtual ip由master运行时分配，编写程序时无法确定

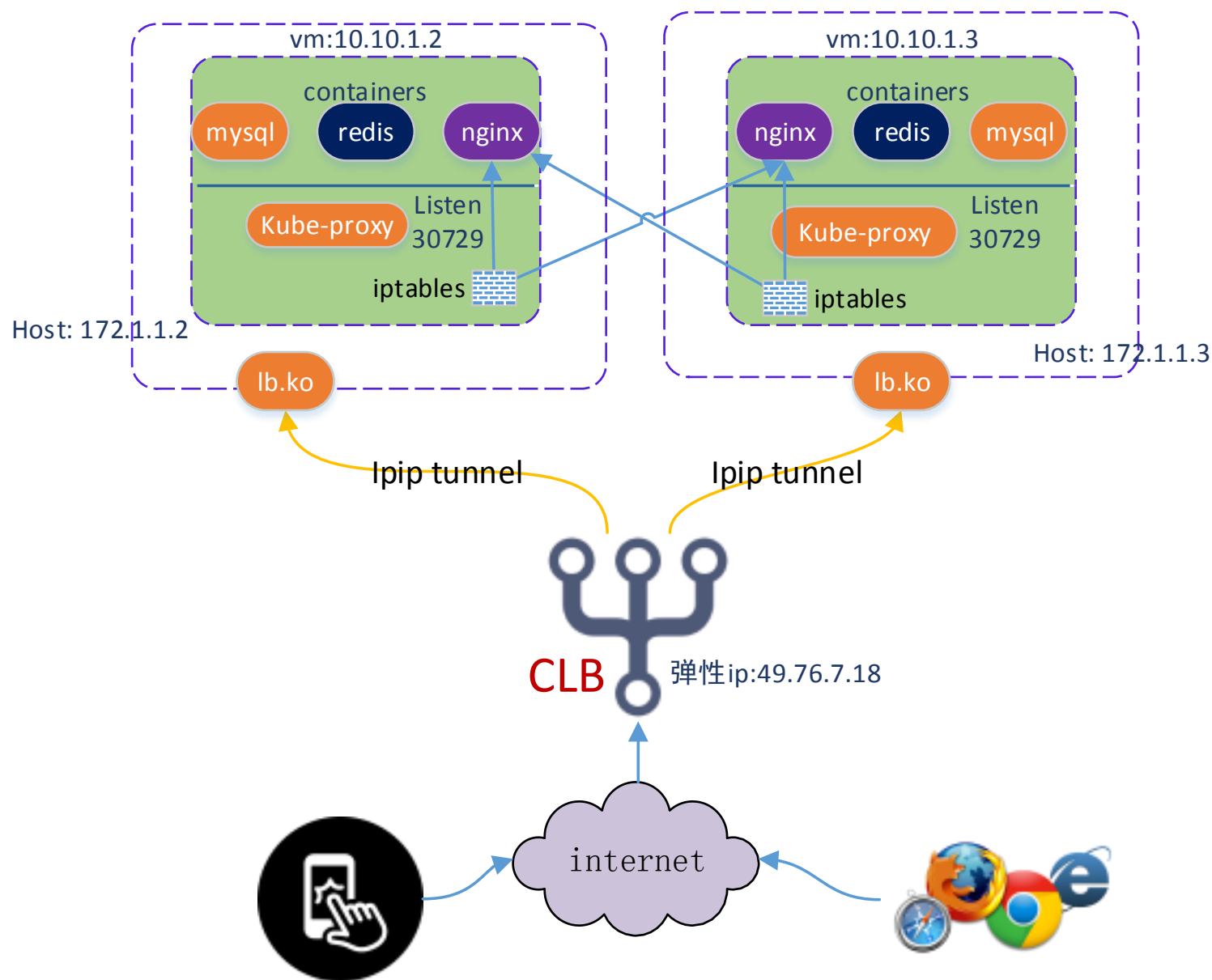
- 环境变量
- 服务发现

Kubernetes-服务发现



```
sock_conn( "mysql" , 3306)  
➔ sock_conn( "172.1.1.3" ,3306)  
➔ sock_conn( "10.1.2.4" ,3306)
```

Qcloud容器集群-使用CLB构建集群外网负载均衡



Kubernetes网络-深入理解外网负载均衡

❑ 两层负载均衡

- client -> node
- node -> pod

❑ L3负载均衡

- service type为LoadBalancer
- 实现Kubernetes loadbalancer插件，插件创建loadbalancer时必须返回loadbalancer的vip
- node上做基于loadbalancer vip的DNAT

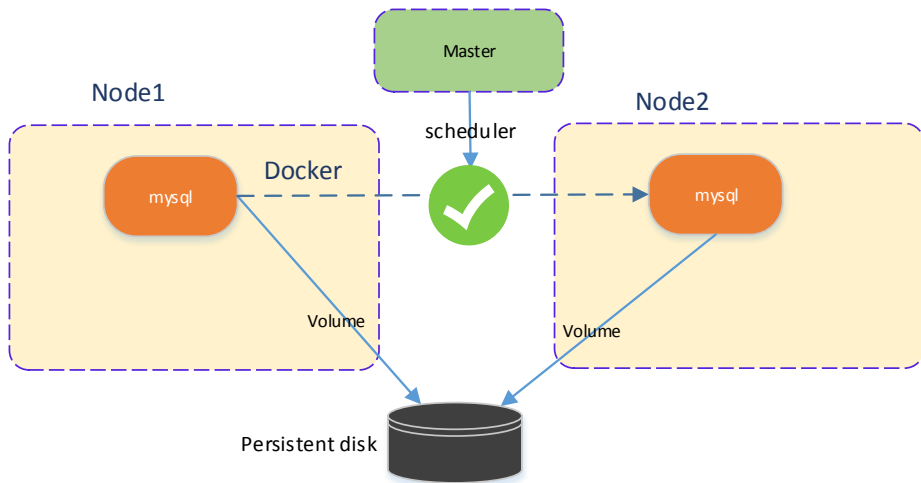
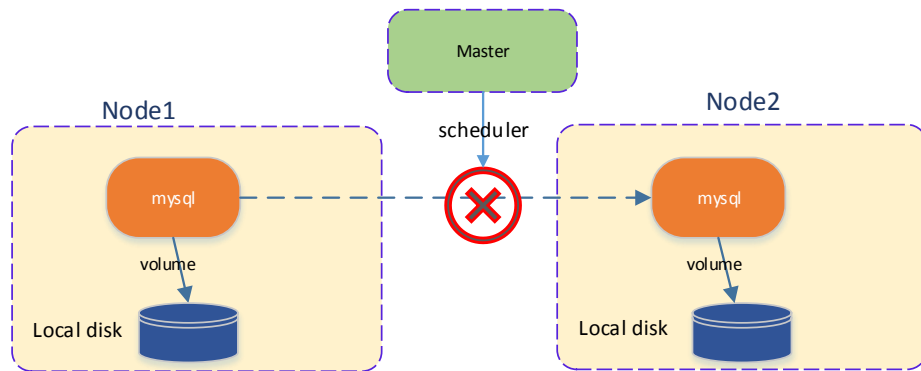
❑ 无负载均衡也要提供外网服务

- service type为NodePort
- node上基于NodePort的DNAT

❑ L4负载均衡

- service type为LoadBalancer
- service的port mapping必须指定NodePort
- 实现Kubernetes loadbalancer插件，插件创建loadbalancer时，不必返回loadbalancer的vip
- node上基于NodePort的DNAT

Kubernetes volume



- ❑ 创建pod时可以指定volume
- ❑ 支持本地盘
- ❑ 支持nfs、ceph , glusterfs等开源网盘
- ❑ 支持gcePersistentDisk,awsElasticBlockStore等云服务商网盘
- ❑ Kubernetes提供volume插件，实现其驱动后可让Kubernetes支持指定的网盘

QCloud容器集群-支持cbs volume

cbs – cloud block service

❑ 多副本容灾

- 三副本

❑ 快照

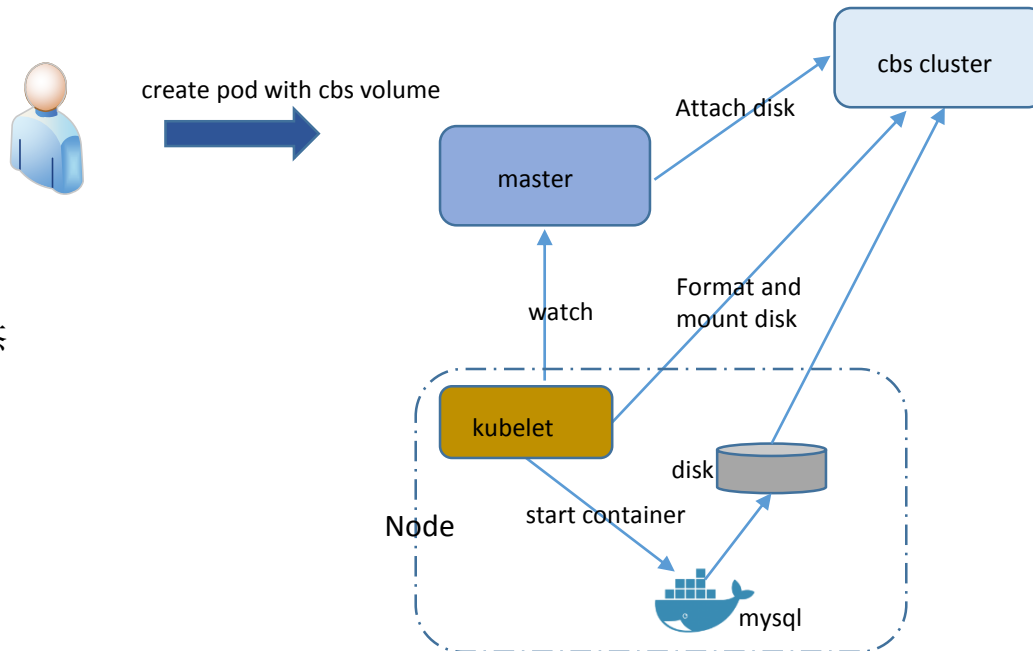
- 瞬间回滚至任意快照点状态
- 可按需会滚到7天内任一秒数据状态

❑ 高效

- 250+MB/s
- 24000+IOPS

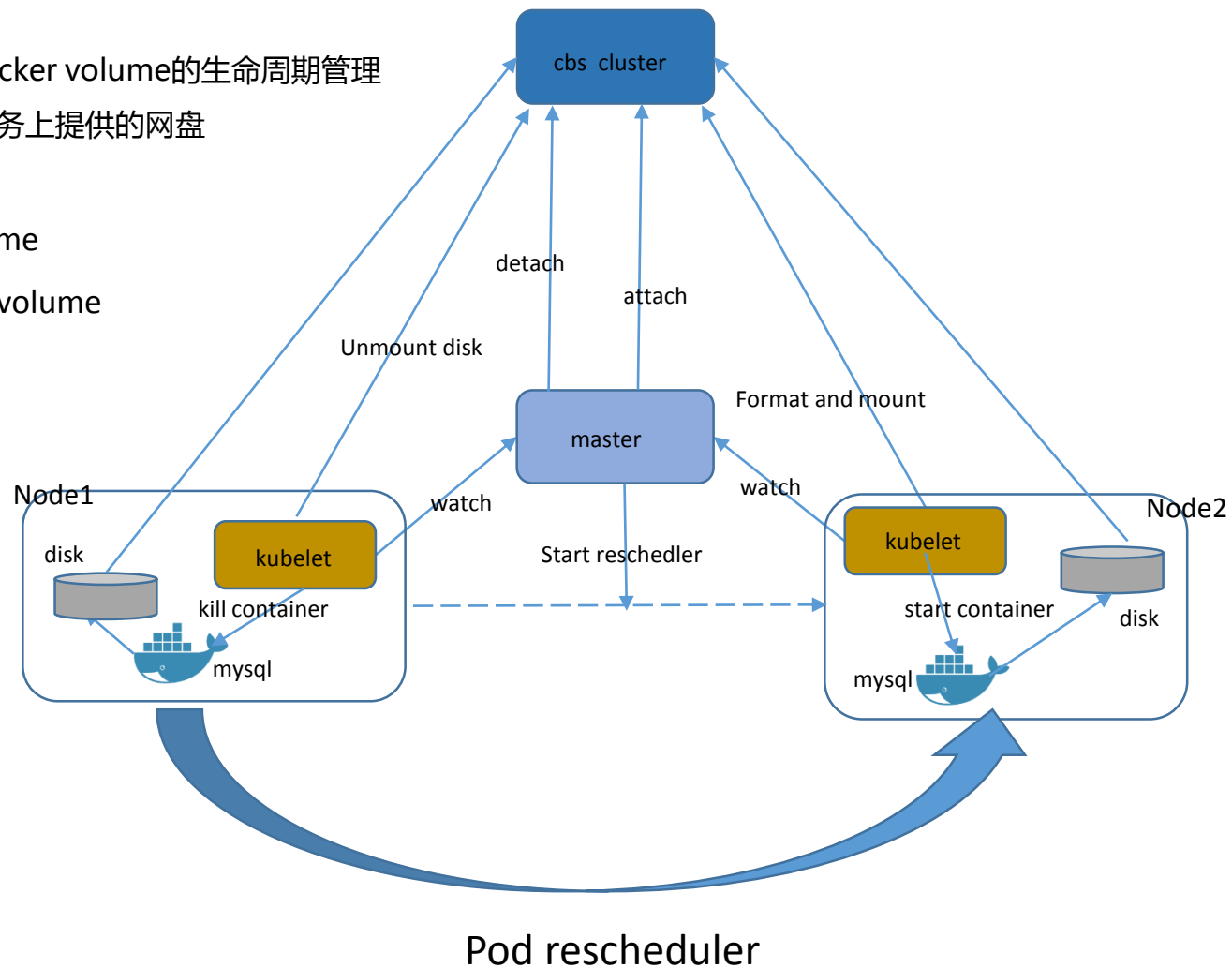
❑ 灵活

- 弹性容量
- 快速迁移



QCloud容器集群-支持cbs卷

- ❑ Kubernetes volume组件负责docker volume的生命周期管理
- ❑ Kubernetes通过插件来支持云服务上提供的网盘
- ❑ cbs卷支持一写多读
 - 一个pod定义rw模式volume
 - 多个pod可以定义ro模式volume



QCloud中使用的Kubernetes 插件总结

❑ CNI

- 创建cbr0网桥
- 配置网桥路由
- 为容器分配veth

❑ Route

- 容器路由管理-路由托管给vpc

❑ Loadbalancer

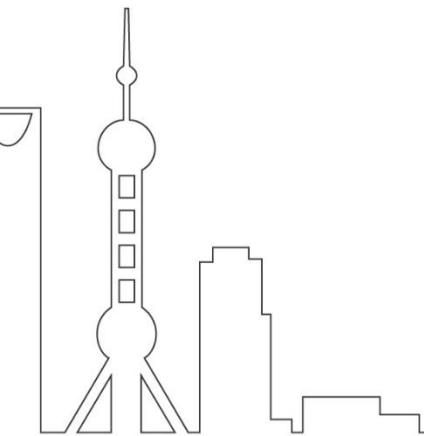
- cloud loadbalancer托管给Kubernetes

❑ Zone/instance

- 周期性检查集群instance状态，KCM以此作为node状态的依据
- 在Kubernetes集群中展示QCloud instance相关的信息

❑ Persistent volume

- 使用QCloud cbs 网盘
- attach&mount
- detach&unmount



Thanks!

International Software Development Conference