SERVICE MESH MEETUP #6 广州站

基于KUBERNETES的微服务实践

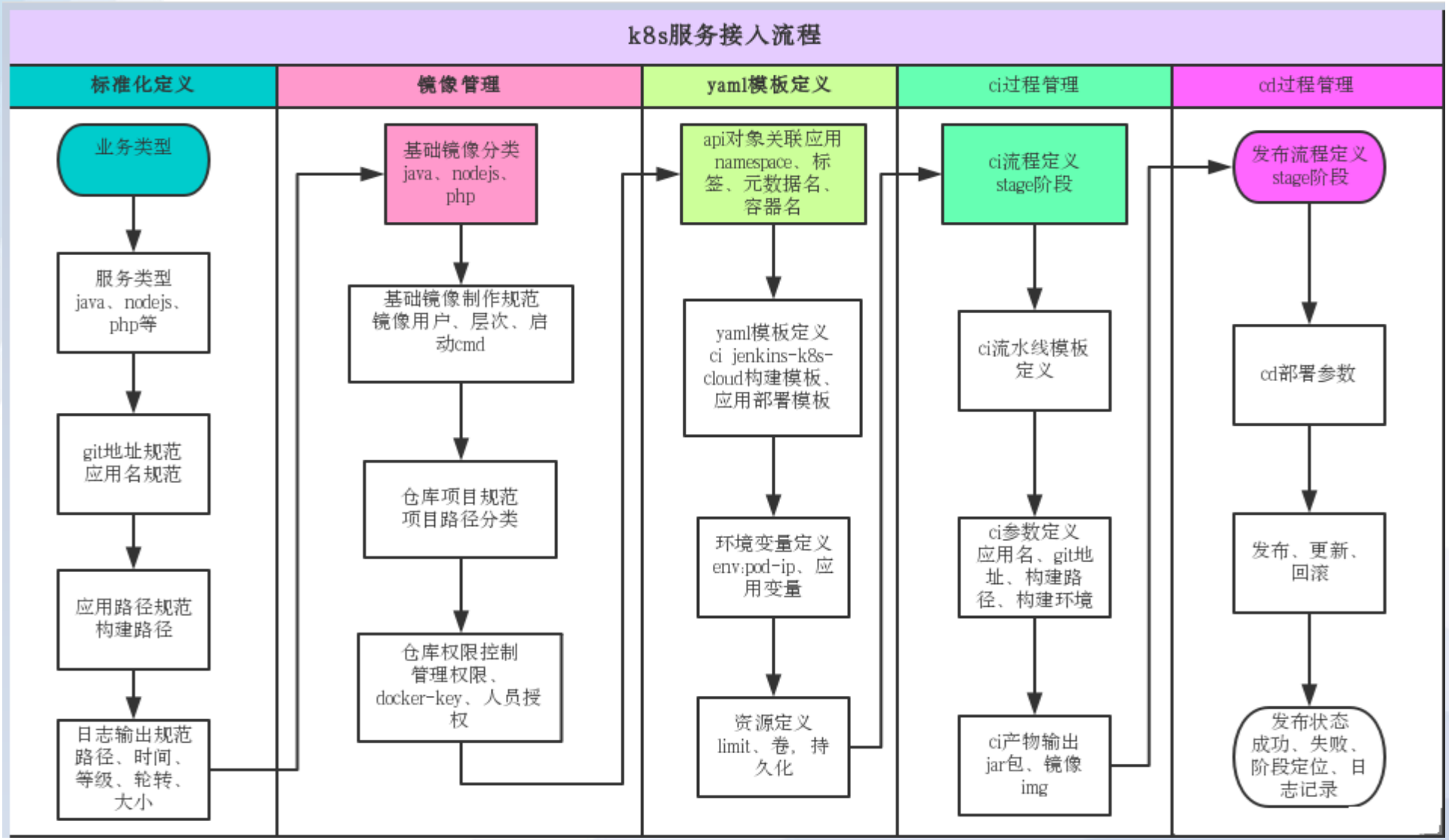
涂小刚

2019.8.11

蚂蚁金服 ANT FINANCIAL ServiceMesher

k8s平台组件

k8s平台接入流程



k8s环境空间和应用名规范

统一规划环境名和业务应用名，适配标准自动化运维。

现有环境名
test
preview
prod

业务线名称
ai
dt
ad

应用名称
ai-dc-server
ai-dc-web
ai-dc-api

k8s-api配置对象	作用
k8s-namespace	通过配置文件关键字dev/test/prod等声明应用所属的环境，隔离不同环境业务，通过特定标识来识别业务线。
k8s-service	k8s-dns注册服务名，通过配置文件关键字关联业务线应用名称，保持应用和k8s之间的关联。
k8s-app-name	容器host应用名称，deployment 名，通过配置文件关键字关联业务线应用名称，保持应用和k8s之间的关联。

规范

业务线名称采用拼音首字母缩写
k8s-namespaces 环境名称定义采用业务线缩写名加环境名组成
k8s-service名称、app名称和应用名称包名保持一致

范例

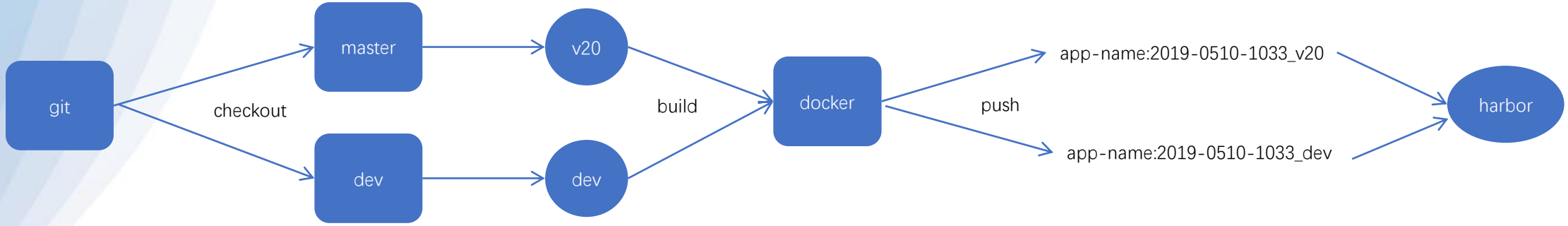
k8s-namespace	k8s-service	k8s-app-name	app-name
ai-test	ai-dc-server	ai-dc-server	ai-dc-servedr
ai-preview	ai-dc-web	ai-dc-web	ai-dc-web
ai-prod	ai-dc-api	ai-dc-api	ai-dc-api

镜像版本和git版本库规范

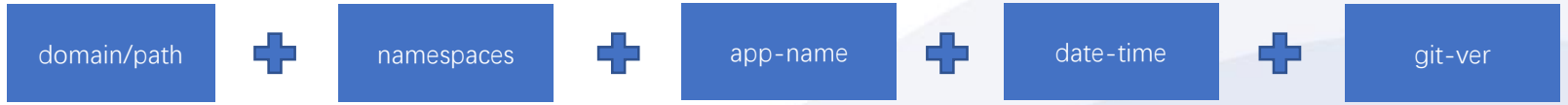
制定git版本规范，开发提交合并master代码，git版本库和业务版本进行关联，出了问题好定位问题。

采用docker容器化之后，ci-cd由运维平台集中控制，git版本和容器镜像必需保持一致关联性，方便问题回溯。

k8s镜像构建过程



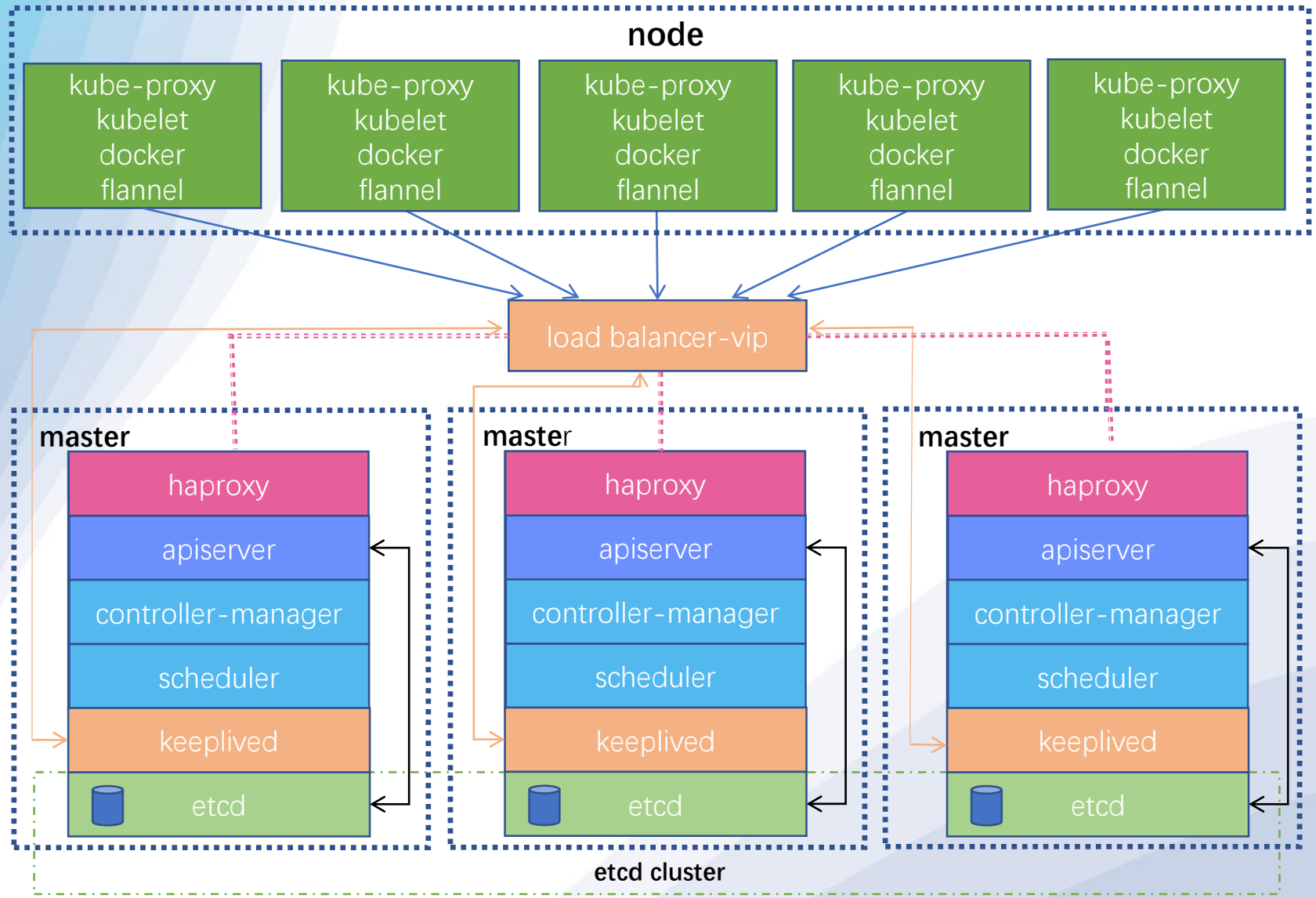
镜像地址组成



镜像地址规范

仓库域名+路径	空间名	应用名称	日期-时间戳	git版本库	镜像完整地址
registry.hz.local/huize	ai-test	ai-dc-web	20190510-1033	v20	registry.hz.local/huize/ai-test_ai-dc-web:20190510-1033_v20

kubernetes cluster HA

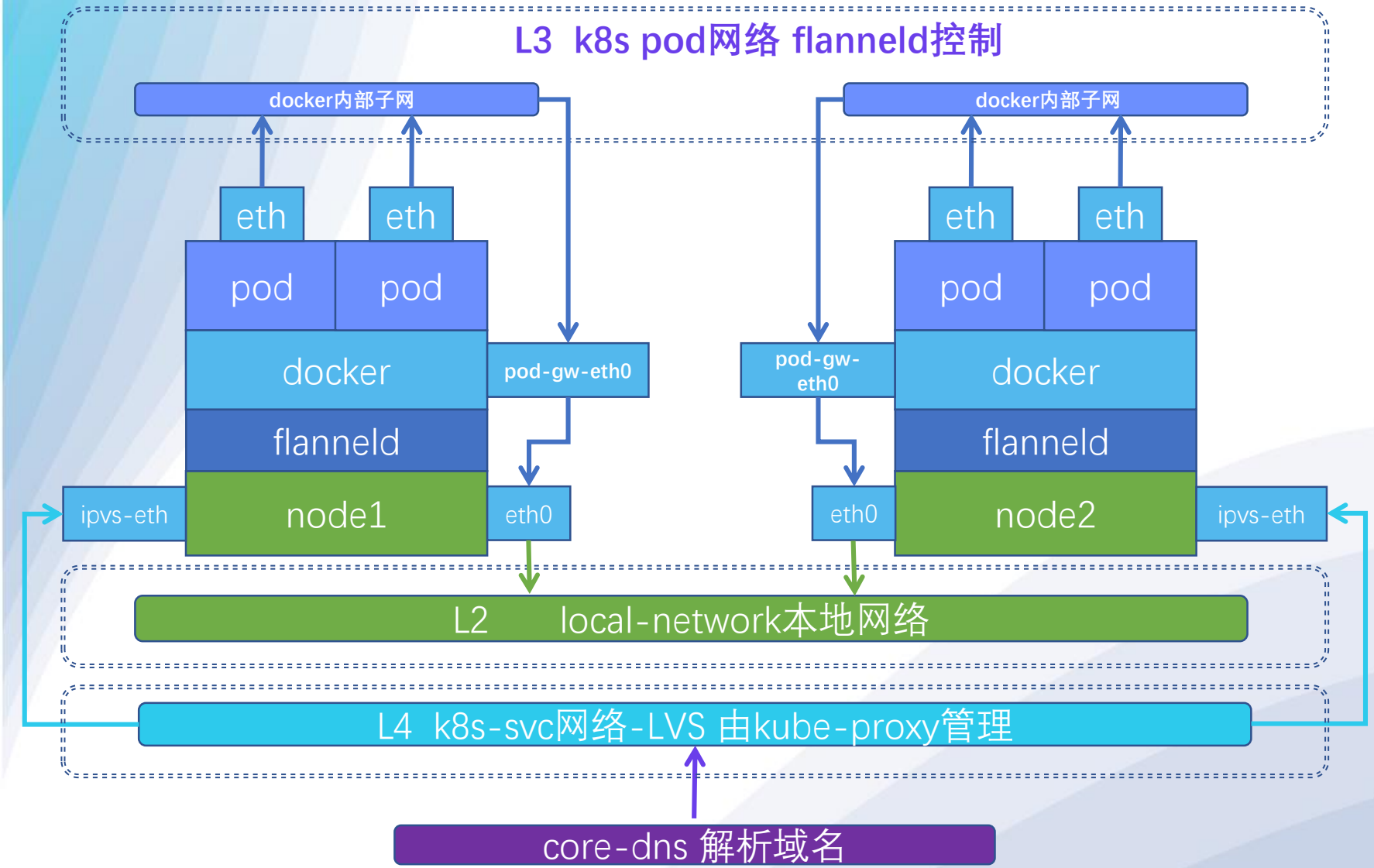


每个控制平面节点运行的一个实例kube-apiserver，kube-scheduler和kube-controller-manager

每个控制平面节点创建一个本地etcd成员，该etcd成员仅与kube-apiserver该节点通信

其中三个控制平台节点运行keeplived和haproxy，node节点和api-server通讯通过vip对接，haproxy将流量转发至apiserver

K8S-flanneld网络分析



工作原理：
Flannel负责在容器集群中的多个节点之间提供第3层IPv4网络。

工作模式：
1.vxlan 通过封装协议解包收发包mtu1450，vxlan可以在分布多个网段的主机间构建2层虚拟网络 。
2.host-gw 通过宿主机路由同步收发包，必需工作在二层。

1.系统启动，flanneld下发docker子网配置，docker启动获取子网配置生成docker0 生成gateway网卡;

2.node1，node的kubelet收到指令创建一个新的pod容器;

3.docker开始创建pod,从flanneld下发子网池生成pod-ip-eth;

4.kube-proxy跟据svc yaml创建ipvs-eth子网卡;

5.flanneld创建同步所有节点docker子网路由表;

访问过程

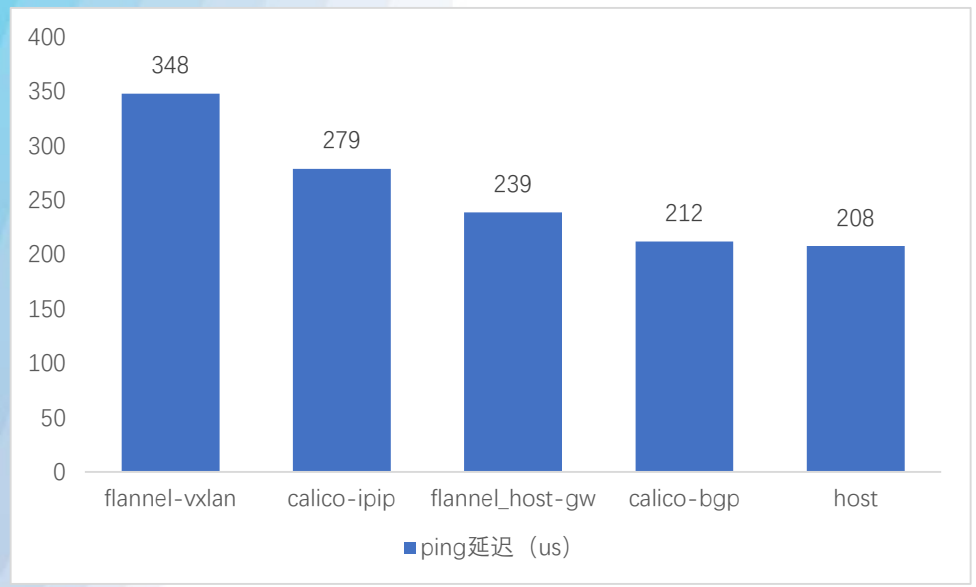
1.core-dns解析来集群内部域名kubernetes.cluster.svc.local;

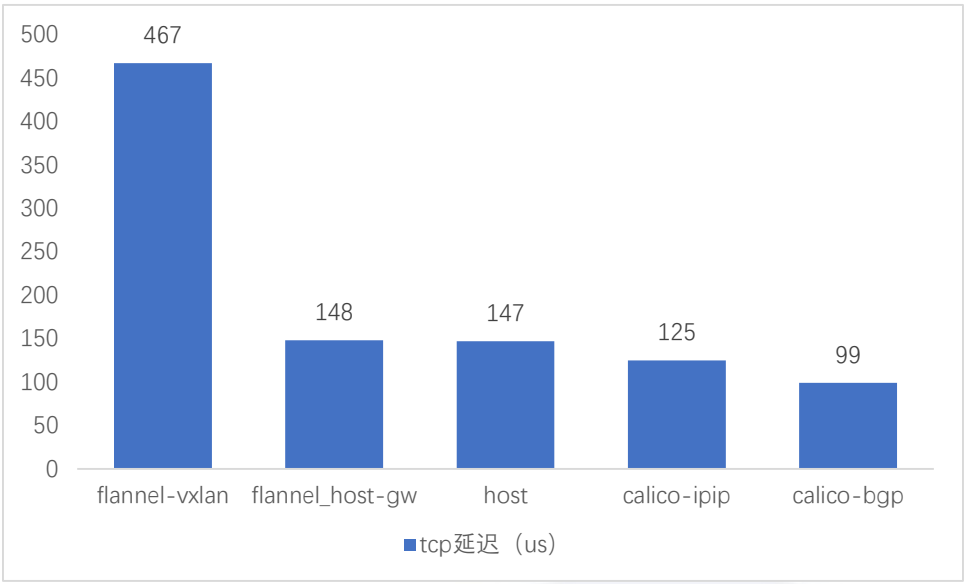
2.流量达到ipvs-eth;
3.lvs负载均衡对流量进行轮寻，寻找目的pod下一跳;

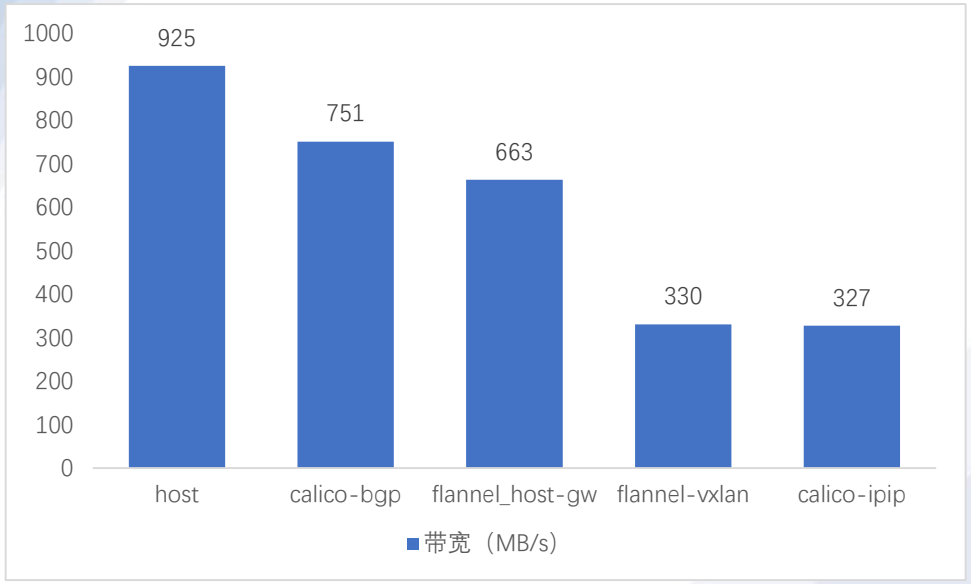
4.流量达到node1,node2的eth0,经过内部路由进行传输;

5.内部路由到达docker-0网关，流量经由flanneld下发的子网到达pod容器;

flannel vs calico







host:指物理机直连网络

calico-bgp:二层bgp模式，自动学习路由

flannel_host-gw:二层直接路由模式

flannel-vxlan:跨三层隧道模式

calico-ipip:跨三层隧道模式

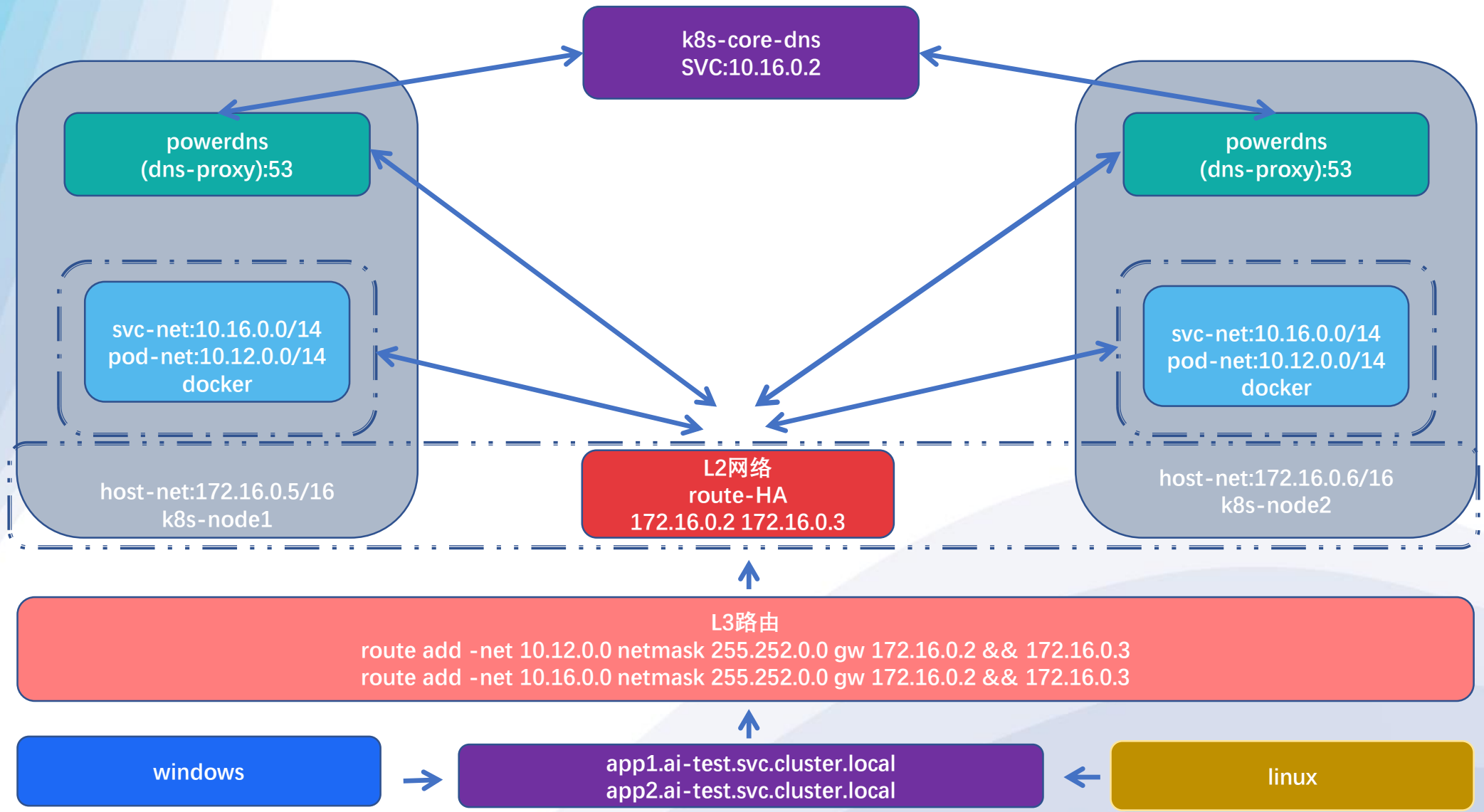
采用万兆网卡的虚拟机，测试方法是不同node节点开启qperf测试

结论：
tcp延迟:calico-bgp<calico<host<flannel_host-gw<flannel-vxlan

ping延迟:flannel-vxlan>calico-ipip>flannel_host-gw>calico-bgp>host

带宽：host>calico-bgp>flannel_host-gw>flannel-vxlan>calico-ipip

网络互通边缘路由



容器平台持续集成交付全流程

Jenkins

jnlp-slave-1 pod

jnlp-slave-2 pod

jnlp-slave-* pod

maven

node

kubectl

helm

java

容器

k8s

node

pull code

push images

harbor

deploy app

pod1 pod2
pod3
pod***

k8s
cluster

ci-cd效果

git scm pull	Build project	Docker-build	Helm-init-yaml	Kubectl-deploy
21s	20s	22s	3s	2s
23s	23s	28s	4s	3s

k8s运维管理平台-构建

ai应用构建

环境空间: test

搜索应用: ai-api,ai-app

查 询 重 置

当前环境 test

应用列表	本次构建镜像版本	构建状态	构建耗时	流水线发布	是否批量	操作	构建历史
ai-api	ai-api:2019-07-26-13-25_v20	git-pull ✓ code-build ✓ build-docker ✓	45s	是	批量	构建	查看
ai-app	ai-app:2019-07-26-13-25_v20	git-pull ✓ code-build ✓ build-docker ✓	1m	是	批量	构建	查看
ai-web	ai-web:2019-07-26-13-25_v30	git-pull ✓ code-build ✓ build-docker ✓	5m	是	批量	构建	查看
ai-server	ai-server:2019-07-26-13-25_v40	git-pull ✓ code-build ✓ build-docker ✗	10m	是	批量	构建	查看
ai-gateway	ai-gateway:2019-07-26-13-25_v50	git-pull ✓ code-build ✓ build-docker ✓ helm-init-yml ✓ deploy-app ✓	12m	是	批量	构建	查看

批量构建

k8s运维管理平台-容器管理

环境空间: test

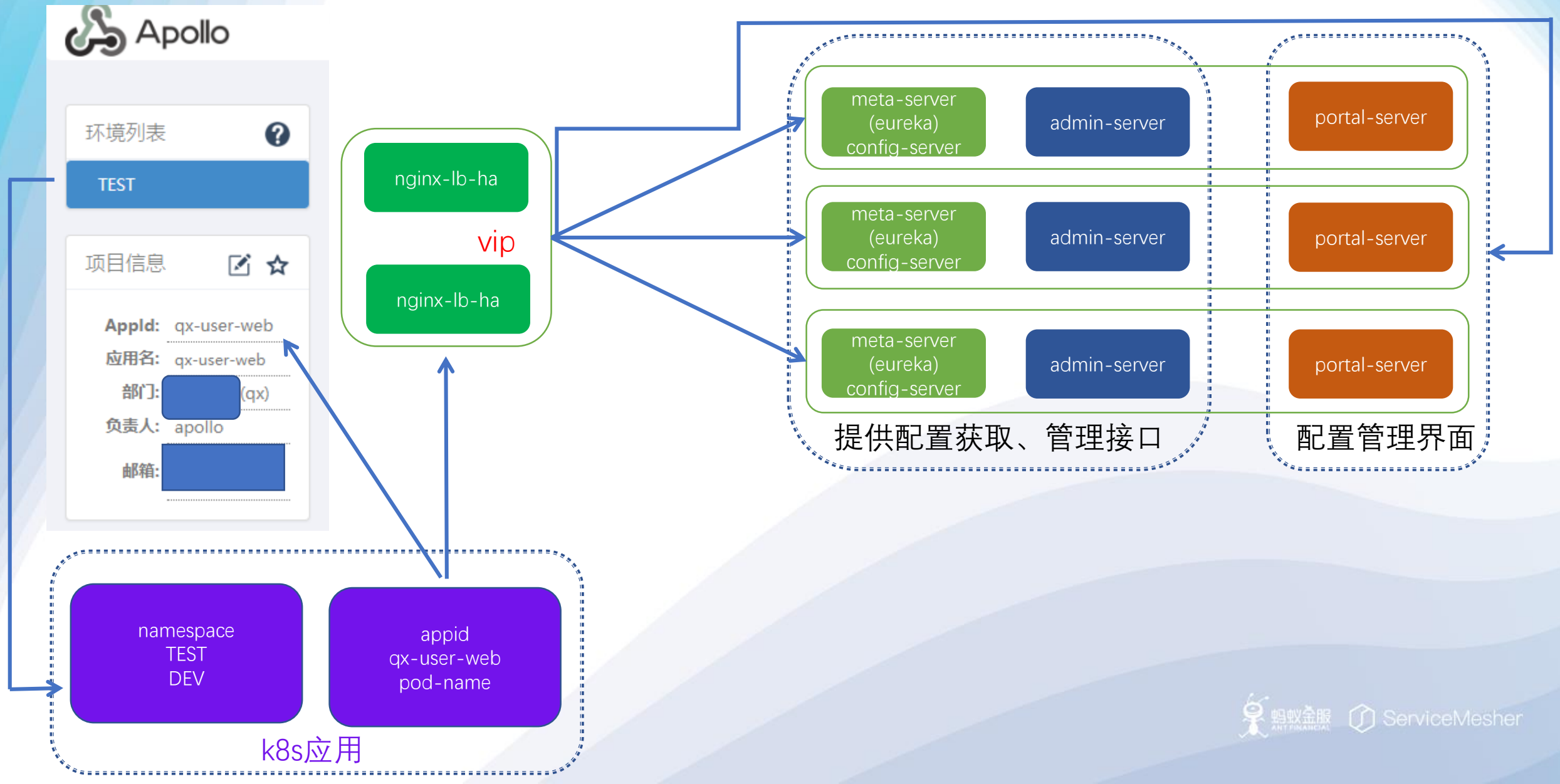
搜索应用: ai-api,ai-app

查 询 重 置

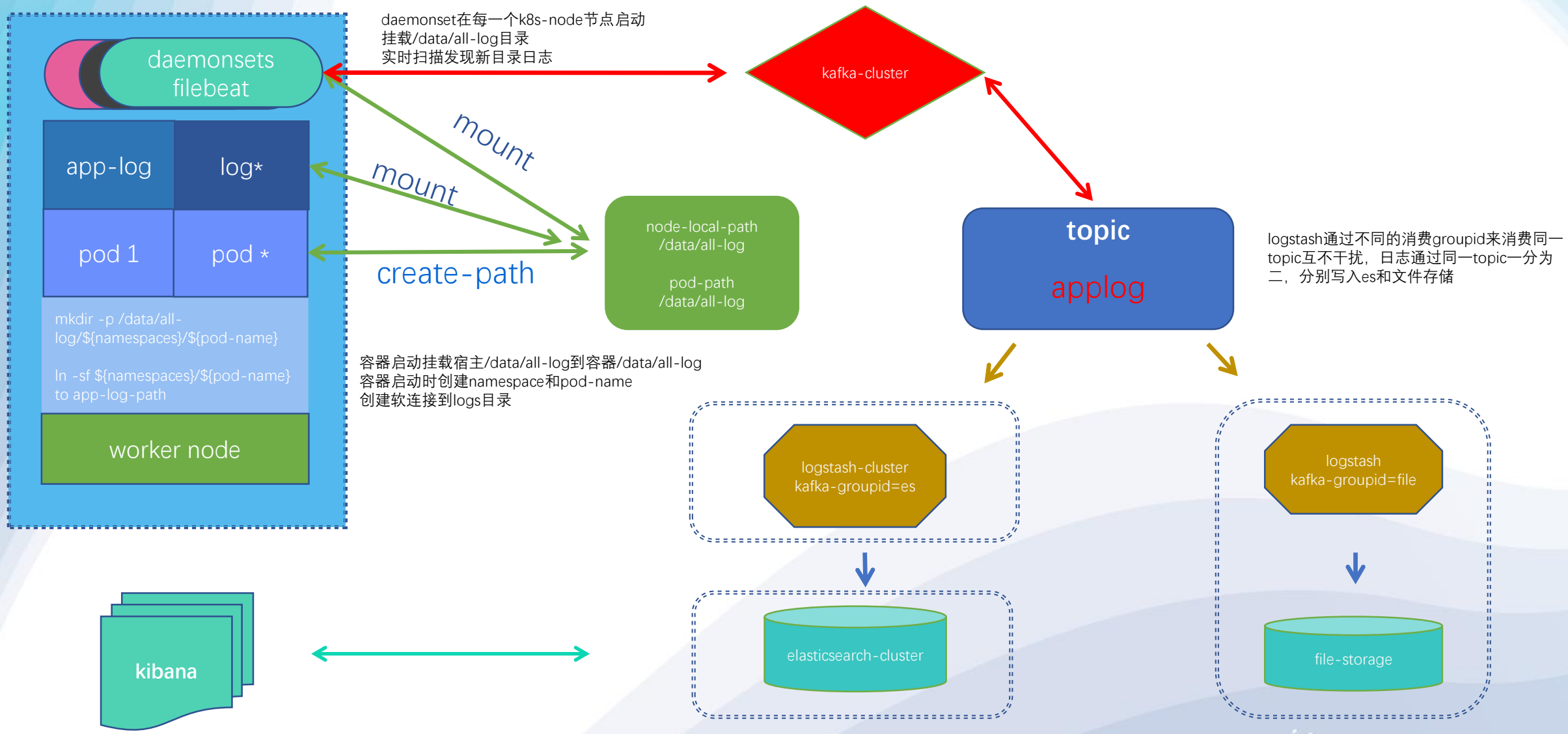
已选择 pod管理 当前环境 ai- (test)

pod名	状态	已重启	pod-ip	k8s-node-ip	svc名	svc-ip-port	已创建	cpu核	cpu限制	内存	内存限制	操作	命令行	启动日志	更多
ai-api-8566595ffb-gwpha	RUNING	0	10.12.0.3	k8s-test-node-0-121	ai-api	10.16.0.3:9090	1 m	0.5	4	1G	3G	重启	登入	查看	_
ai-app-8566595ffb-gwphf	RUNING	2	10.12.0.4	k8s-test-node-0-122	ai-app	10.16.0.4:9090	1h	1	4	2G	4G	重启	登入	查看	_
ai-web-8566595ffb-gwphe	RUNING	0	10.12.0.5	k8s-test-node-0-123	ai-web	10.16.0.5:9090	1d	2	4	2.9G	3G	重启	登入	查看	_
ai-server-8566595ffb-gwphc	Unknown	4	10.12.0.6	k8s-test-node-0-124	ai-server	10.16.0.6:9090	2d	3	4	3.8G	4G	重启	登入	查看	_
ai-gateway-8566595ffb-gwphd	Fai-led	5	10.12.0.7	k8s-test-node-0-109	ai-gateway	10.16.0.7:9090	5d	4	5	4G	5G	重启	登入	查看	_

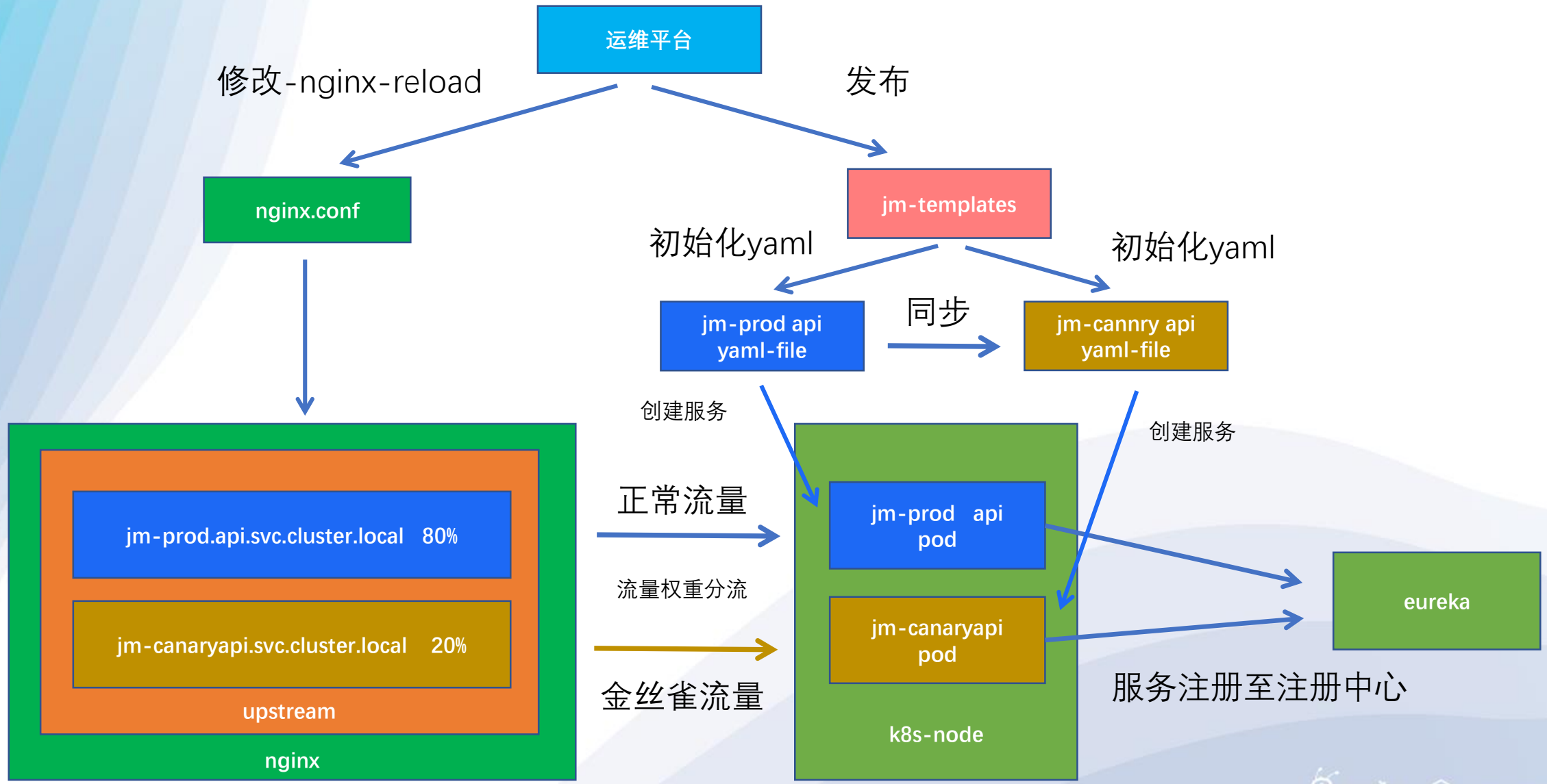
k8s应用对接阿波罗



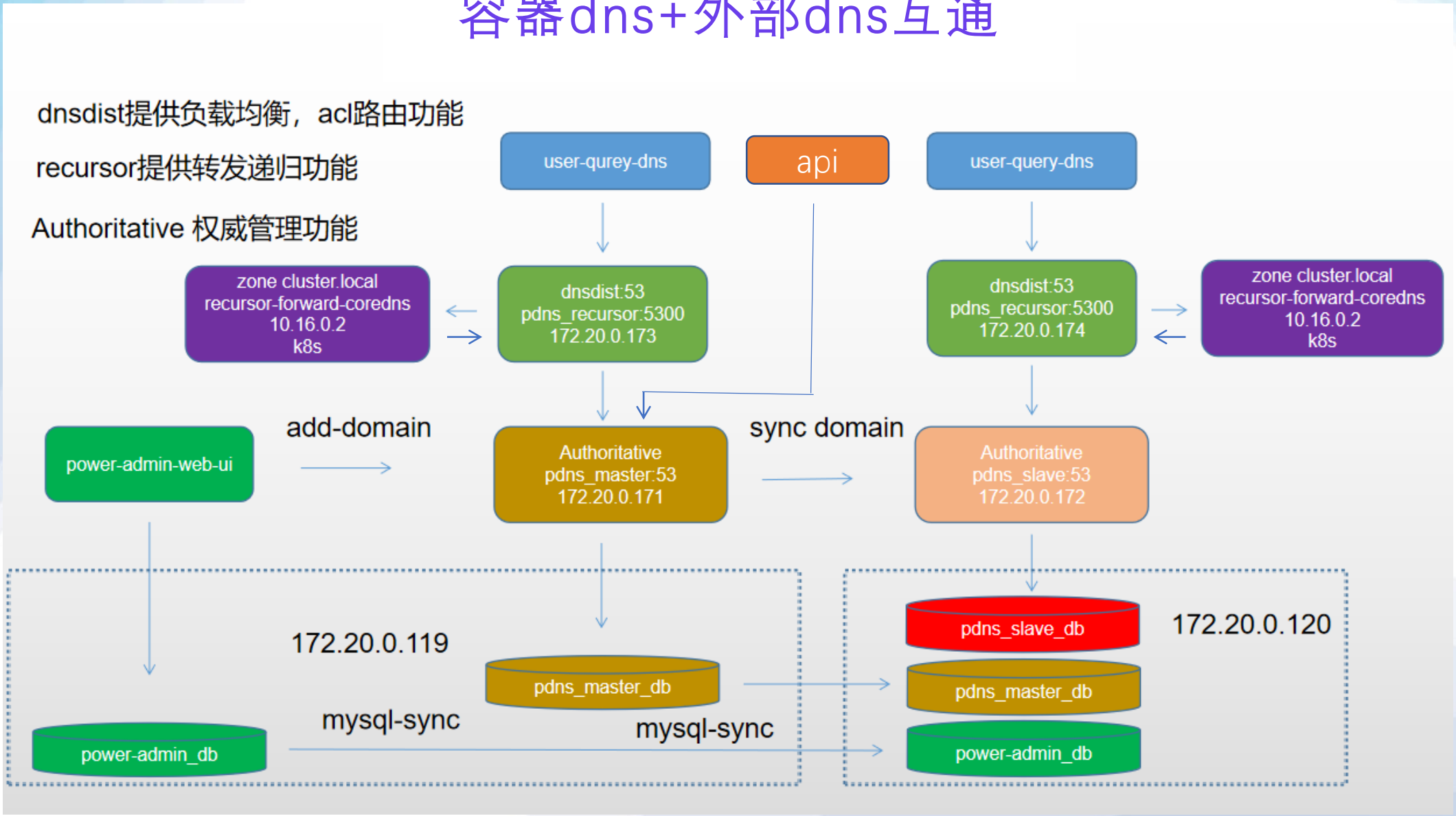
容器平台日志解决方案



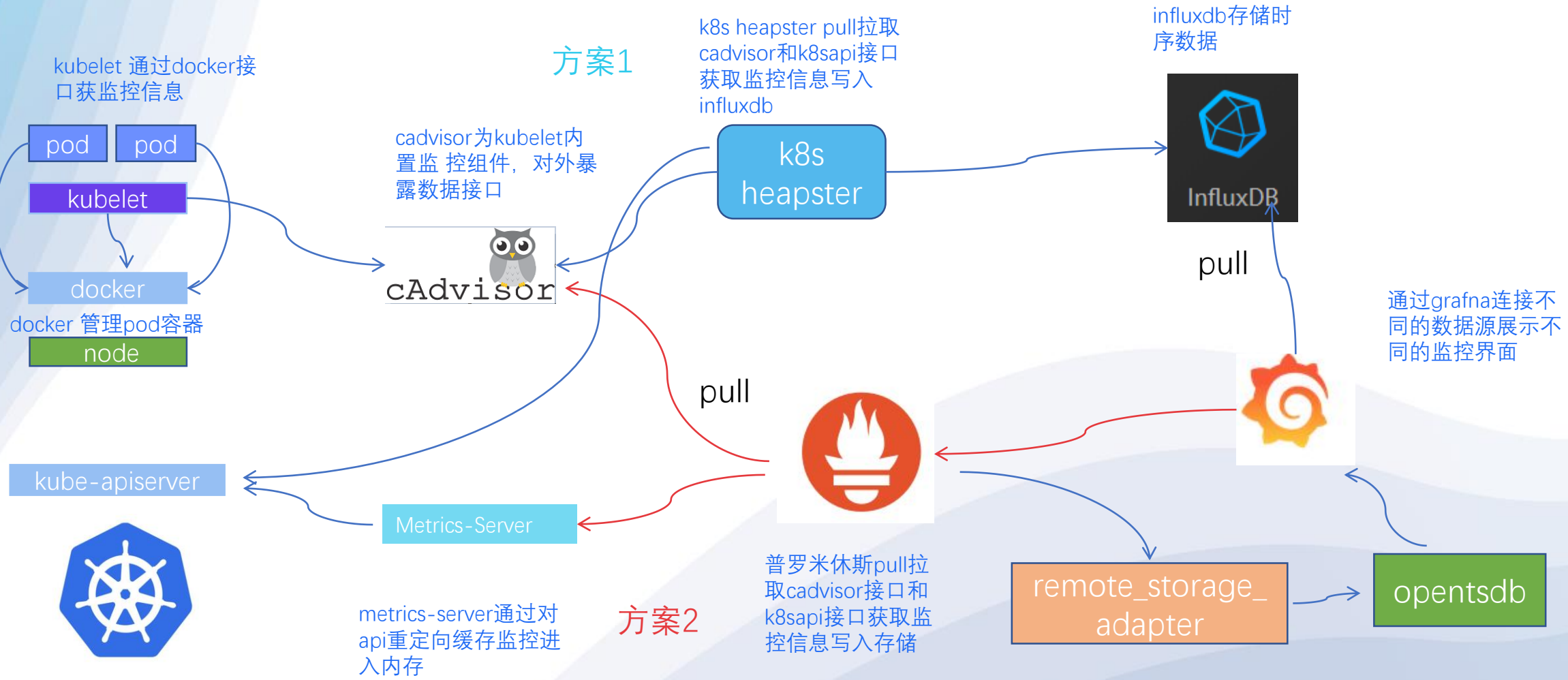
金丝雀灰度发布



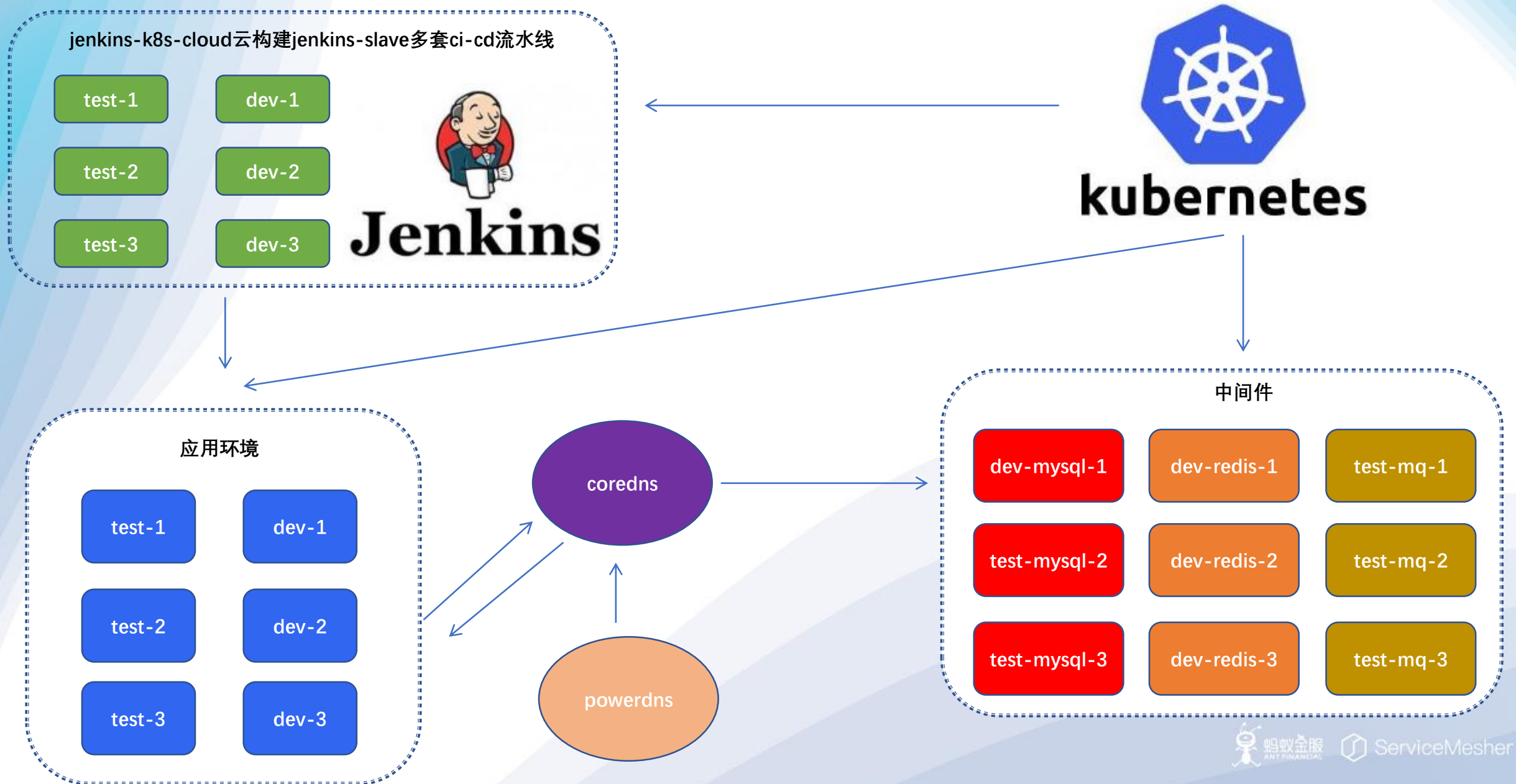
容器dns+外部dns互通



容器监控



多套环境快速交付



关注 **ServiceMesher** 微信公众号
获取社区最新信息

关注 金融级分布式架构 微信公众号
获取 **SOFAStack** 最新信息

ServiceMesher 社区是由一群拥有相同价值观和理念的志愿者们共同发起，
于 2018 年 4 月正式成立，致力于成为 Service Mesh 技术在中国的布道者和领航者。
社区官网：https://www.servicemesher.com

蚂蚁金服 ANT FINANCIAL　ServiceMesher