



Jenkins User Conference China

中国 · 北京

北京市朝阳区悠唐皇冠假日酒店 2019年7月6日

主办单位:  DevOps时代 |  高效运维社区 | cloudbees | DOIS

Kubernetes在去哪儿的实践



孔晖 运维工程师

2018年加入去哪儿网，目前任职于去哪儿平台数据团队，目前主要是负责公司的容器系统的基础设施建设和elasticsearch集群的开发和维护。

目录

CONTENT

1

容器化现状与发展

2

K8S容器化的实践

3

Elastic search on Kubernetes

4

生产中碰到的一些问题



PART

01



容器化现状与发展

容器规模

1700+

stateful

4700+

stateless

40TB+

logs/day

发展历程



2015

Mesos/Docker 部署到生产环境，提供日志解析作业的容器化，同年 9 月日均处理能力突破 40 亿 / 日。



2016

Elasticsearch on Mesos 部署到生产环境，同时 Spark/Storm on Mesos 也开始支撑业务部门，同年 2 月份处理能力突破 100 亿 / 日。



2017

Flink on Mesos 逐步替代 Spark streaming，进一步增强流计算能力。



2018

调研 Kubernetes。逐步切换数据作业至新的平台。截至到此，Mesos 已累计调度 1783 万容器。



2019

新的容器平台架构，全面迁移



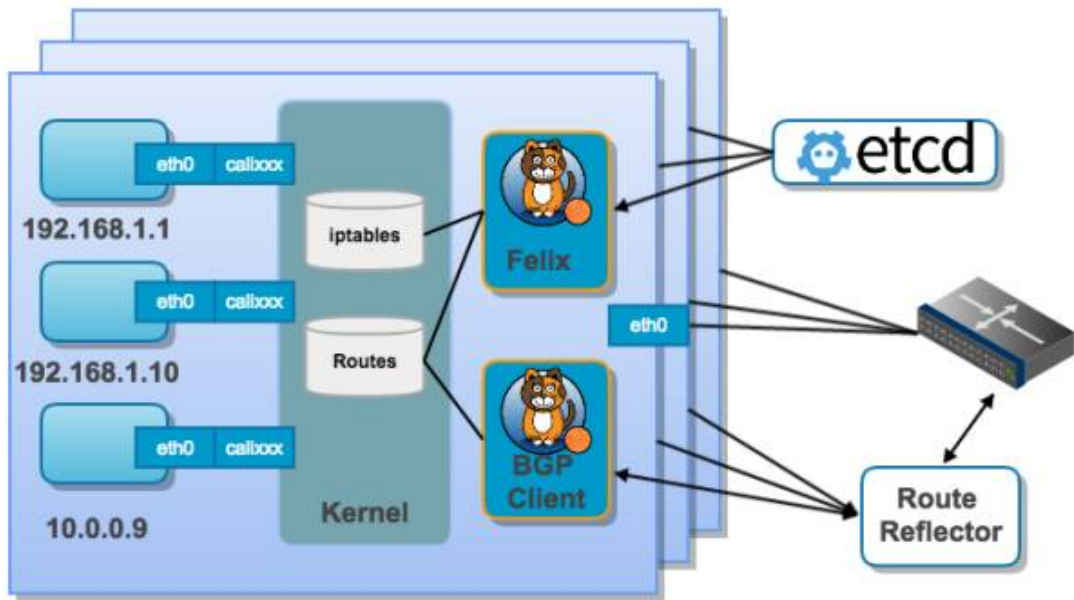
PART

02



K8S容器平台的实践

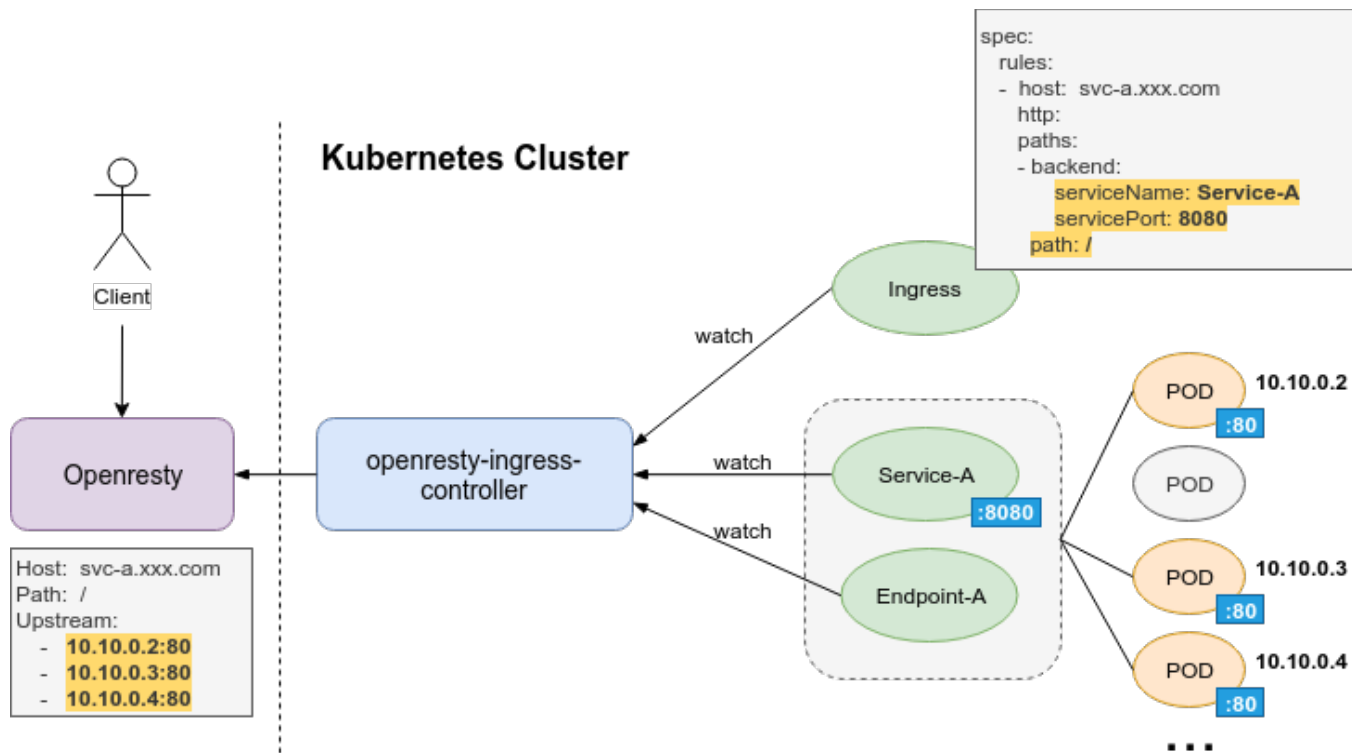
calico网络方案



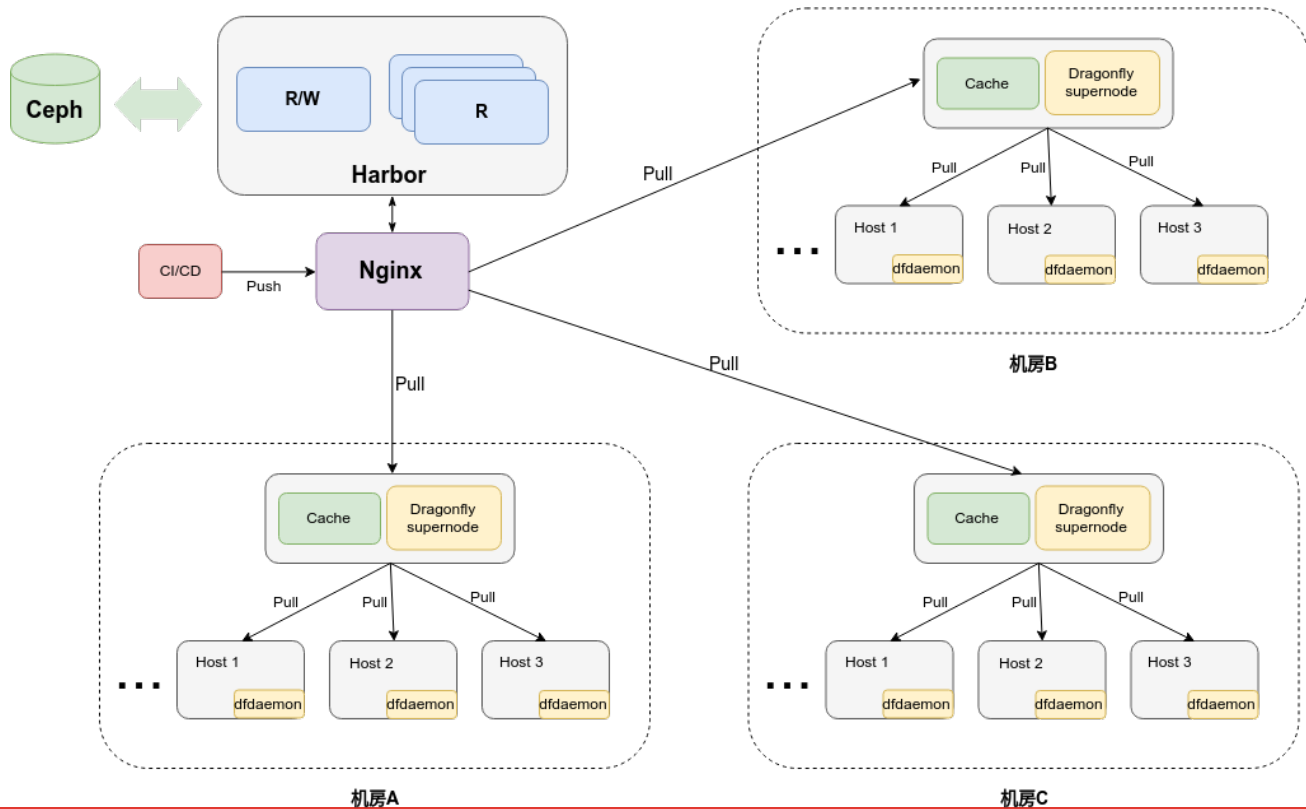
Calico Advertising over BGP + IPVS 4层负载方案




1. K8S Service对象四层负载 (ipvs)
2. 集群内部ClusterIP外网可达
 - Calico advertising over BGP
 - ECMP等价多路路由
3. 外部流量负载策略
 - externalTrafficPolicy: Local
 - externalTrafficPolicy: Cluster

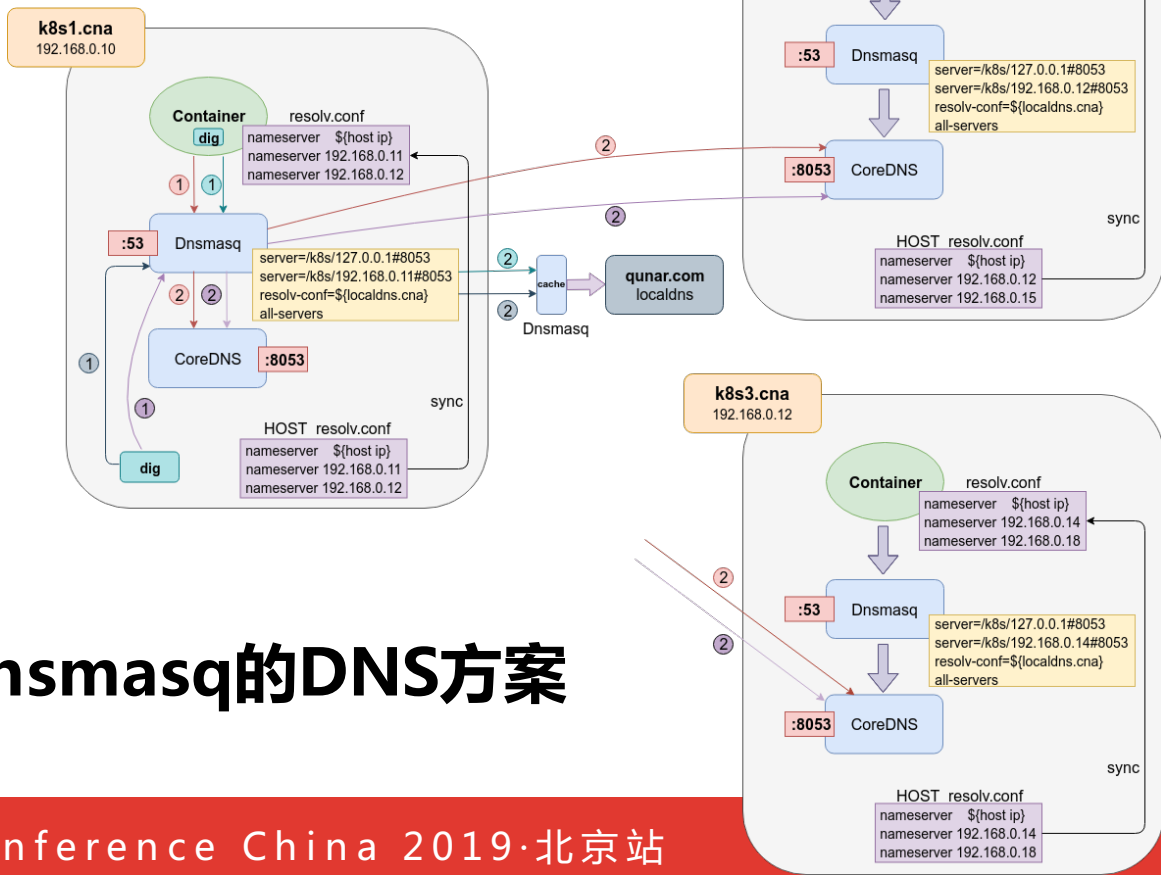
Openresty的ingress7层负载方案



Harbor和Dragonfly的镜像仓库方案

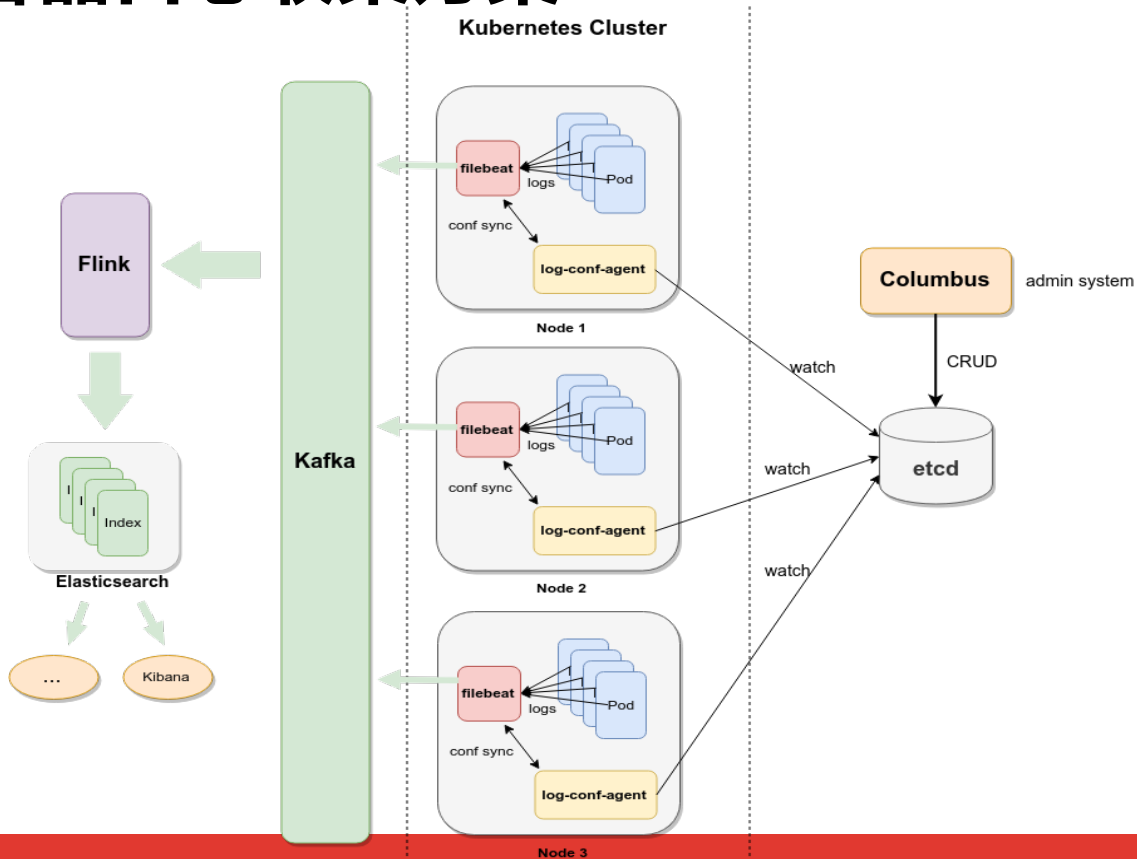


 dig nginx.default.svc.cluster.local
 dig l-xxx.ops.cna
 dig nginx.default.svc.cluster.local
 dig l-xxx.ops.cna

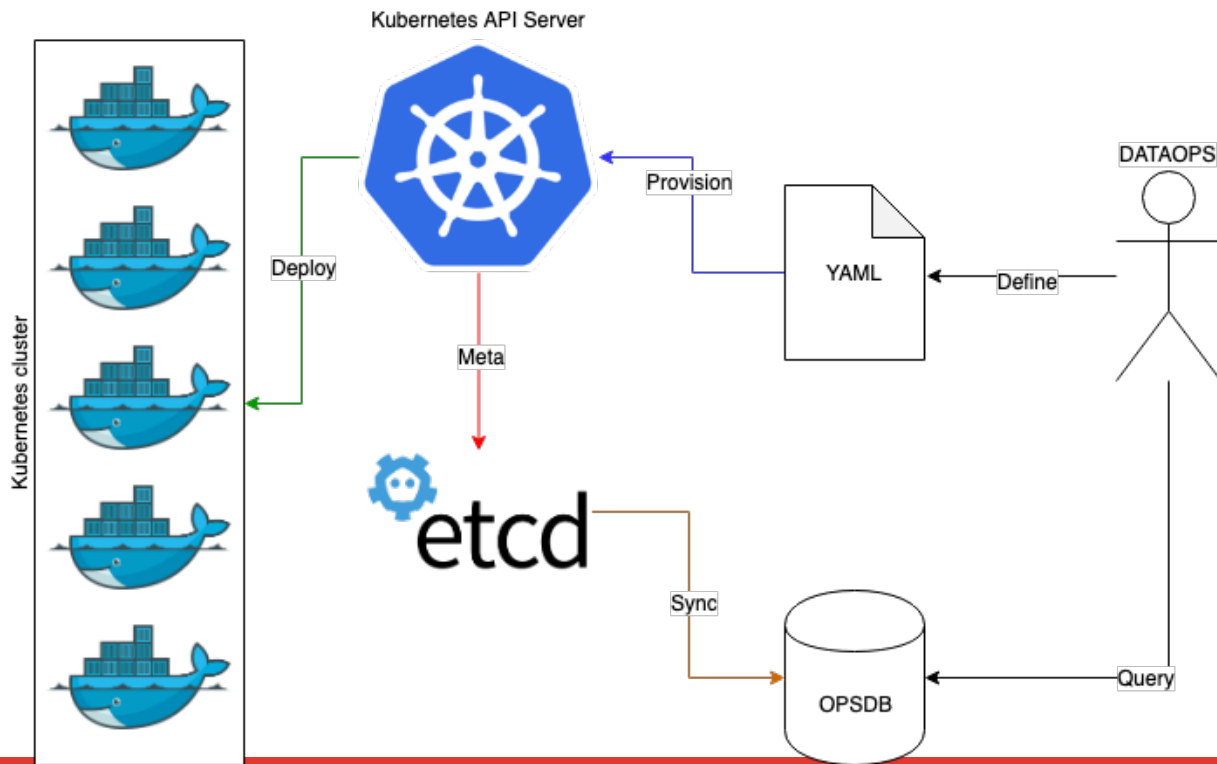


CoreDNS和Dnsmasq的DNS方案

Filebeat容器日志收集方案



运维数据源同步方案



运维标准化

1. 上线前规范检查

- 必要字段检查
- 命名规范检查

2. 环境变量注入

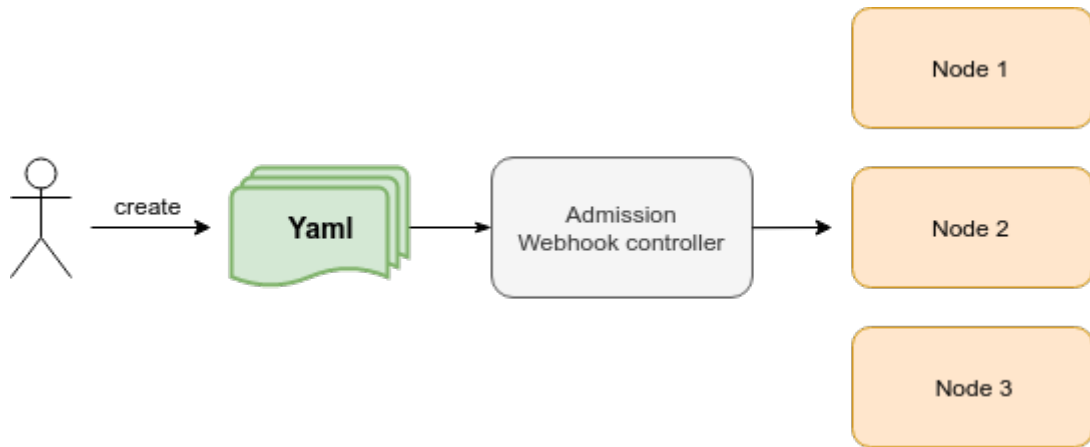
- JVM环境变量注入
- 其他通用的环境变量

3. 集群资源管控

- 集群超卖控制

4. 兼容性控制

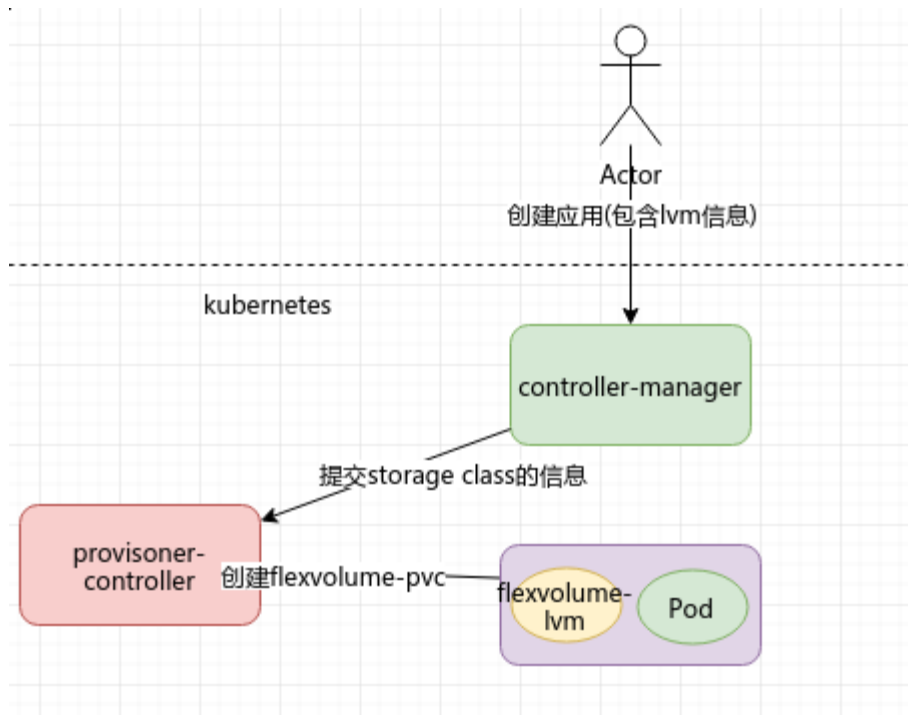
- 屏蔽集群由于K8S版本导致的配置不一致的问题



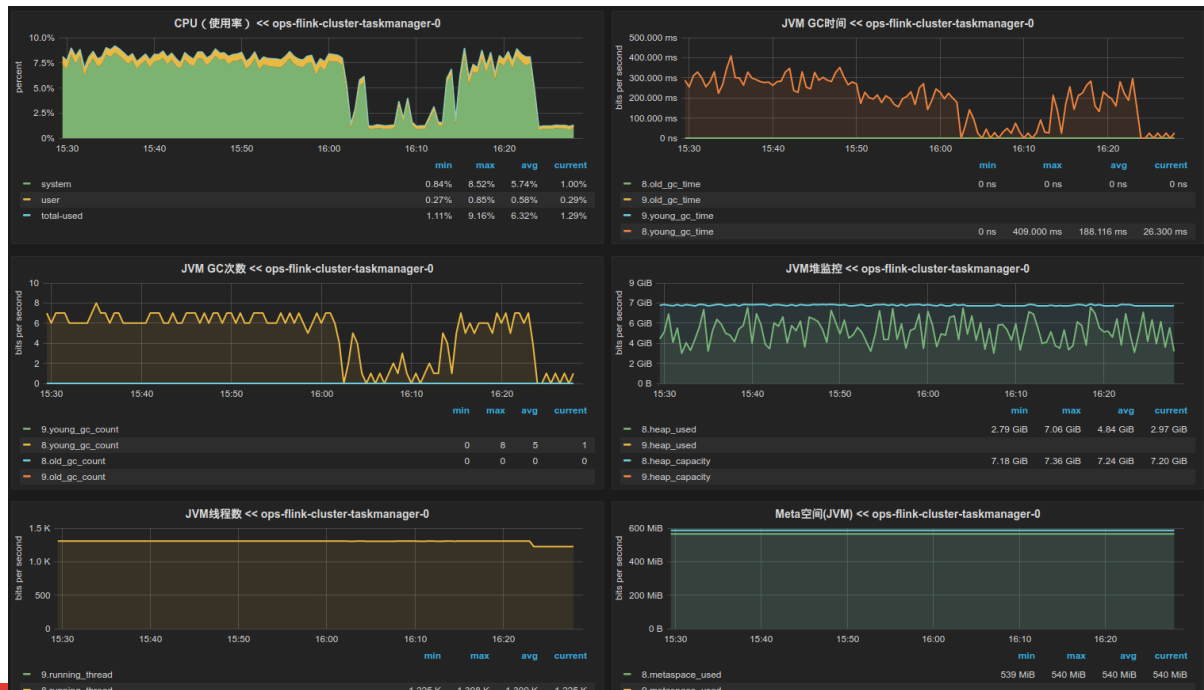
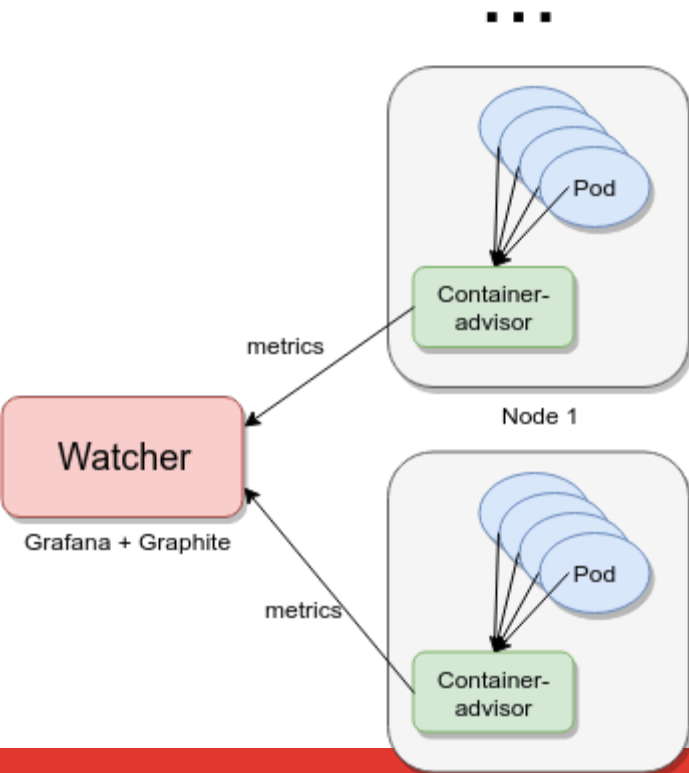
存储方案

1. 本地存储

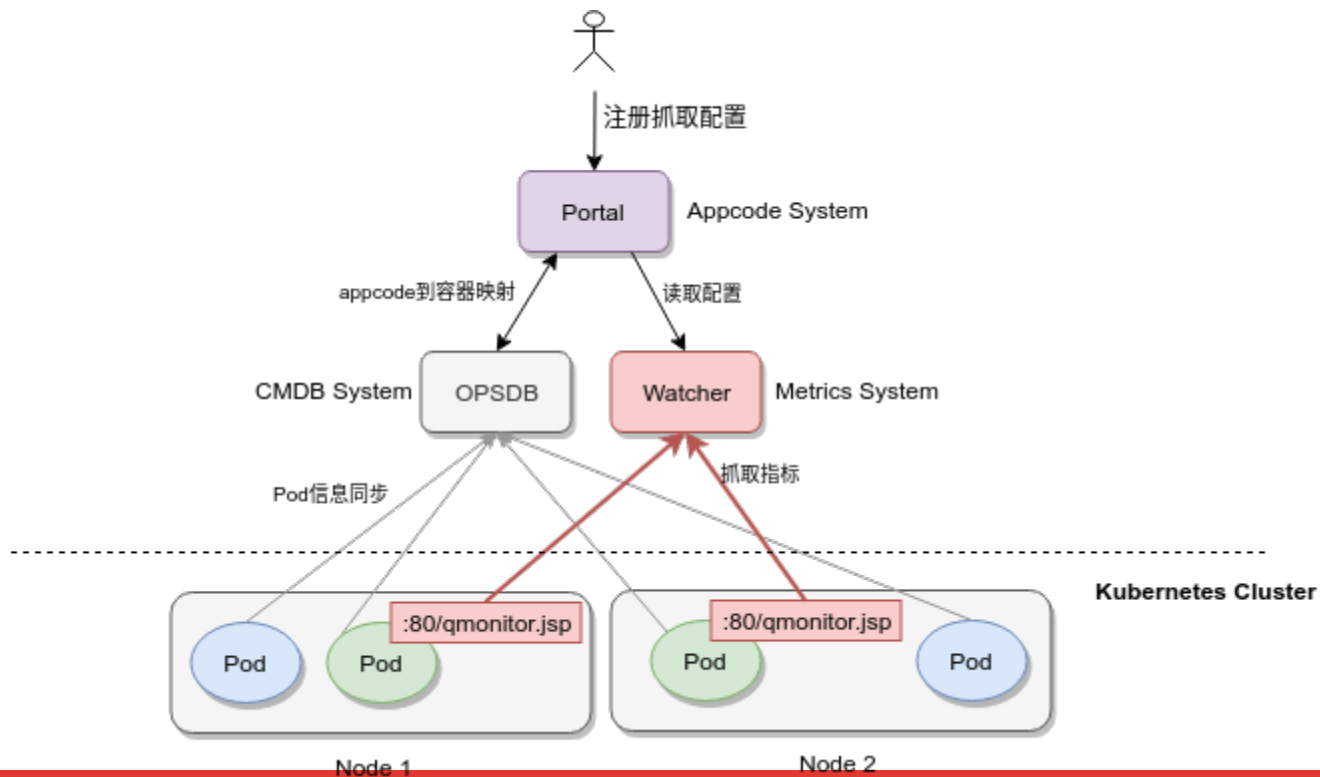
- 日志卷：lvm
 - Deployment多Pod，flexvolume以subPath的方式挂载
- 数据卷：lvm Flexvolume provision by PVC
 - Statefulset下使用PVC，每个pod都有惟一独立的卷



容器基础指标监控



容器业务监控指标



容量管理

1. 容量展示

- 展示集群的资源使用情况

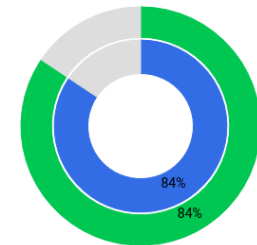
2. 容量计算

- 计算当前的集群可以最多扩容多少机器

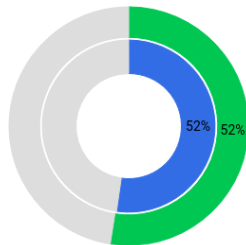
3. 容量明细

- namespace级别容量计算
- 资源类型级别的容量展示
- 报告下载

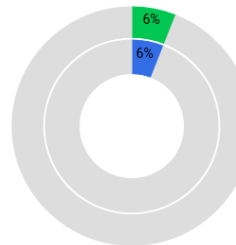
Global Allocation



CPU allocation (cores)
 ■ Requests 1438
 ■ Limits 1438



Memory allocation (bytes)
 ■ Requests 3.92 Ti
 ■ Limits 3.91 Ti



Disk allocation
 ■ Requests 15.50 Ti
 ■ Limits 15.50 Ti

Allocated PerNamespace

Namespace	CPU Request ↑	CPU Limit	Memory Request	Memory Limit	Disk Request	Disk Limit
d-dba-prod-collect	2核	2核	2.95 Gi	2.95 Gi	1,024.00 Mi	1,024.00 Mi
ops-opsapp	2核	2核	3.00 Gi	3.00 Gi	0.00	0.00
pf-coverage-web	2核	2核	3.00 Gi	3.00 Gi	0.00	0.00
h-supplier-order-adapter-intl	2核	2核	2.95 Gi	2.95 Gi	1,024.00 Mi	1,024.00 Mi
default	2核	2核	2.00 Gi	2.00 Gi	0.00	0.00



PART

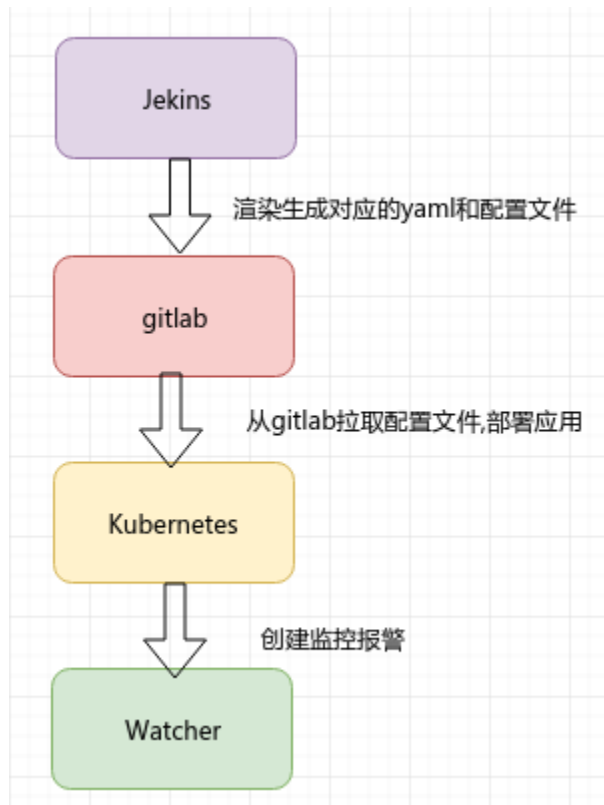
03



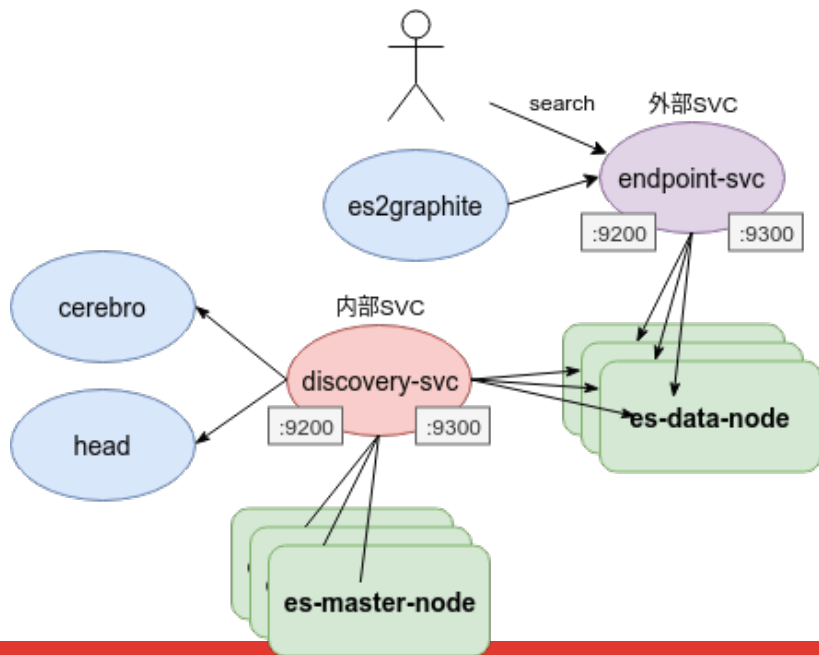
ElasticSearch on Kubernetes

集群的发布过程

1. 生产集群所需的资源和配置
2. 根据生成后的配置创建集群
3. 创建集群对应的监控和报警信息



工作原理



qunar.corp.ops.ops_prism_qes 🚩

概况

集群名称	ops_prism_qes
集群版本	6.7.2
http端口	9200
transport端口	9300
node数量	8
shards总量	10
cerebro访问地址	http://ops-prism-qes.cerebro.k8service
集群连接地址	ops-prism-qes.ops-prism-qes.svc.z2.k8s.
配置地址	http://gitlab.corp/esaasconfcarbon/ops_prism_qes
watcher监控	http://watcher.corp/#/dash/team/.../ops_prism_qes



PART

04



一些生产中碰到的问题

一些生产中碰到的问题

1. kubernetes集群无法正常的续签证书

- 原因：由于客户端过期之后没有正常启动cert-manager导致的
- 解决方案：
 - 在kubelet.config文件里面不仅要enable bootstrap的feature还需要额外添加serverTLSBootstrap: true这个配置

2. Extend resource 被清0的问题

- 原因：在kubelet每次重启之后会自动清空集群通过patch方式加入的资源标识
- 解决方案：6月已经有人通过pr方式以及改掉了这个问题，到时候会提供一个参数供大家选择
 - 在每次kubelet启动之后，自动执行脚本进行patch

我是一个标题

1. 我是第一个大论点

- 我是一个小论点
- 我是另一个小论点

2. 我是第二个大论点

- 我是小论点
- 我也是小论点



Thanks

DevOps 时代社区荣誉出品

想第一时间看到高效运维社区
的新动态吗？

