

2018. 星空计划

网易乐得数据库&运维分享会

7月1日 13:30-17:30
网易北京研发中心C1-M6



关注网易乐得DBA

网易乐得Docker平台设计与实践

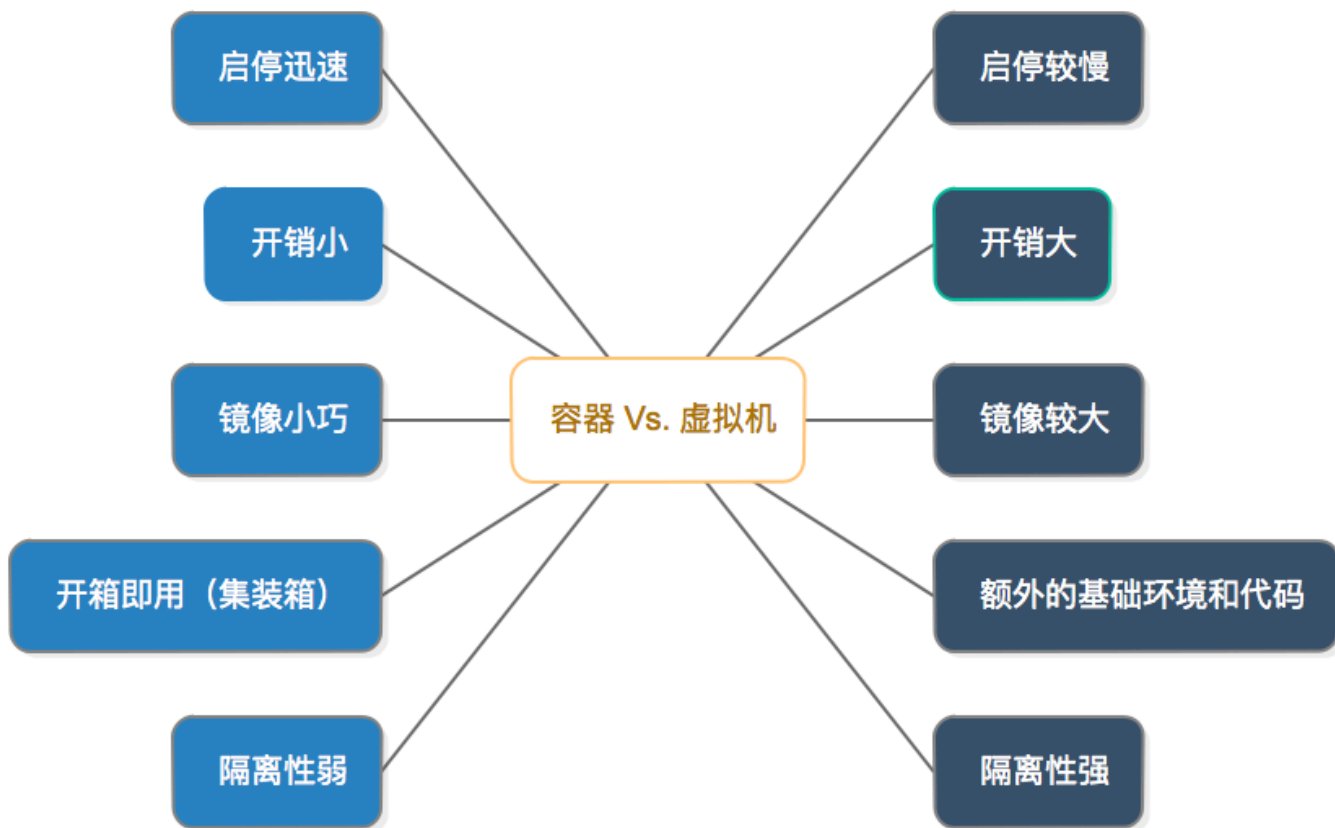
網易 NETEASE
乐得技术团队

目录

- » 为什么使用容器
- » 设计目标
- » 基本架构
- » 功能特性
- » 技术要点

为什么使用容器

为什么使用容器



设计目标

设计目标——容器怎么用？

- » 要不要把容器当做虚拟机？
- » 代码要不要打包到镜像内？
- » 我是否需要一个编排工具？

设计目标——Cloud Native

- » “云原生” 包含一种应用模式，帮助企业快速、持续、可靠、规模化的交付软件
 - 容器化封装：便于代码和组件重用
 - 动态管理：通过集中式编排调度系统来动态管理资源
 - 面向微服务：服务解耦，明确依赖

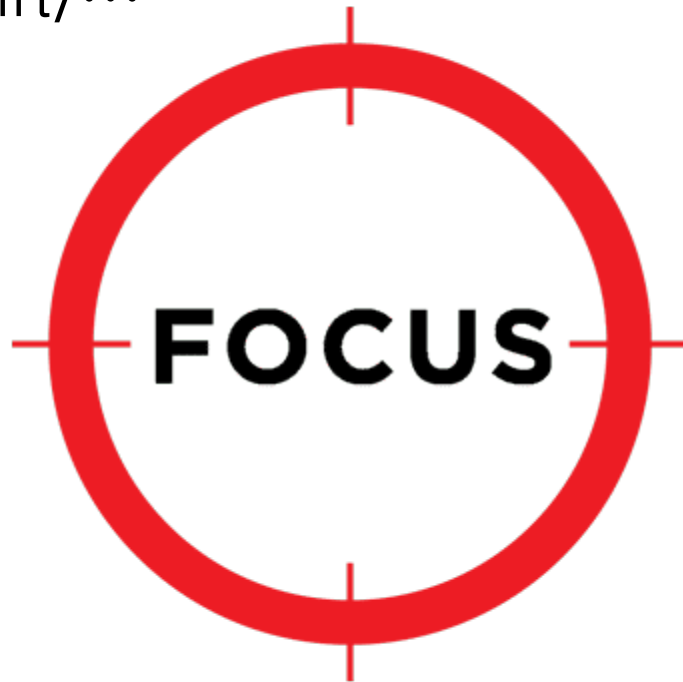
设计目标——PaaS

» 不使用开源PaaS平台

- 如: Rancher/OpenShift/...

我们更实际

更关注用户体验!



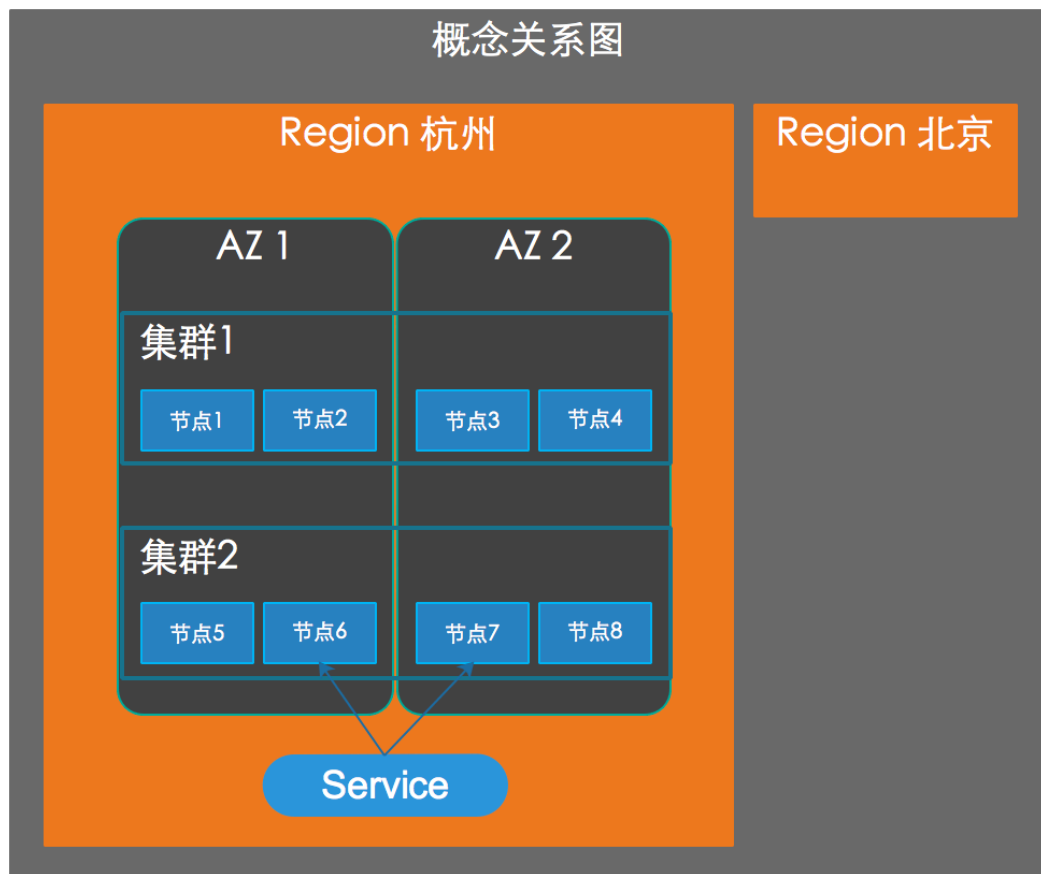
基本架构

基本架构



基本架构——集群结构

- » 租户
- » 区域
- » 可用区
- » 节点
- » 集群
- » 命名空间



功能特性

功能特性——首页

The screenshot displays the OPS dashboard interface. At the top, there are navigation tabs for '租户' (Tenant), '区域' (Region), and '命名空间' (Namespace), with 'SA', '杭州一区', and 'default' selected respectively. The main content area is divided into several sections:

- 资源使用情况 (Resource Usage):** This section contains four bar charts:
 - 计算资源 (Compute Resources):** CPU (8/30核) at 8, 内存 (8/50GB) at 8.
 - 云存储 (Cloud Storage):** 块存储 (10/100GB) at 10, 文件存储 (5/100MB) at 5.
 - 网络IP (Network IP):** eth0 (6/8个) at 6, eth1 (1/1个) at 1.
 - 容器服务 (Container Service):** 服务数量 (6/20个) at 6.
 - 负载均衡器 (Load Balancer):** 负载均衡器 (3/5个) at 3.
 - Redis服务 (Redis Service):** Redis内存 (1/10GB) at 1.
- 账户服务 (Account Service):** Includes a link for 'VPN下载 (可连通当前命名空间下的容器服务)'.
- 操作记录(最近30条) (Operation Log):** A list of recent operations with timestamps and user names, such as '2018-06-22T10:28:43 操作者: [redacted] 重新部署 容器服务 [redacted]'.
- OPS更新日志 (OPS Update Log):** A list of version updates, including '2018-05-16 版本: 3.1.3' with new features like VPN and Kafka, and '2018-04-26 版本: 3.1.2'.

The left sidebar contains a navigation menu with categories like '概览' (Overview), '容器服务' (Container Service), 'Redis服务', 'Zookeeper服务', 'Kafka服务', '负载均衡器', 'RDS服务', '镜像仓库', '云存储', '配置中心', 'CDN配置', 'VPC管理', '弹性IP', 'NAT网关', '权限管理', '安全组', 'DNS域', '日志服务', '安全', 'API管理', '监控', and '基础配置'.

功能特性——SNAT网关

» 问题?

- 安全性高的服务，需要对IP授权

NAT网关
创建/查看/修改NAT网关

总览 展开 ▾

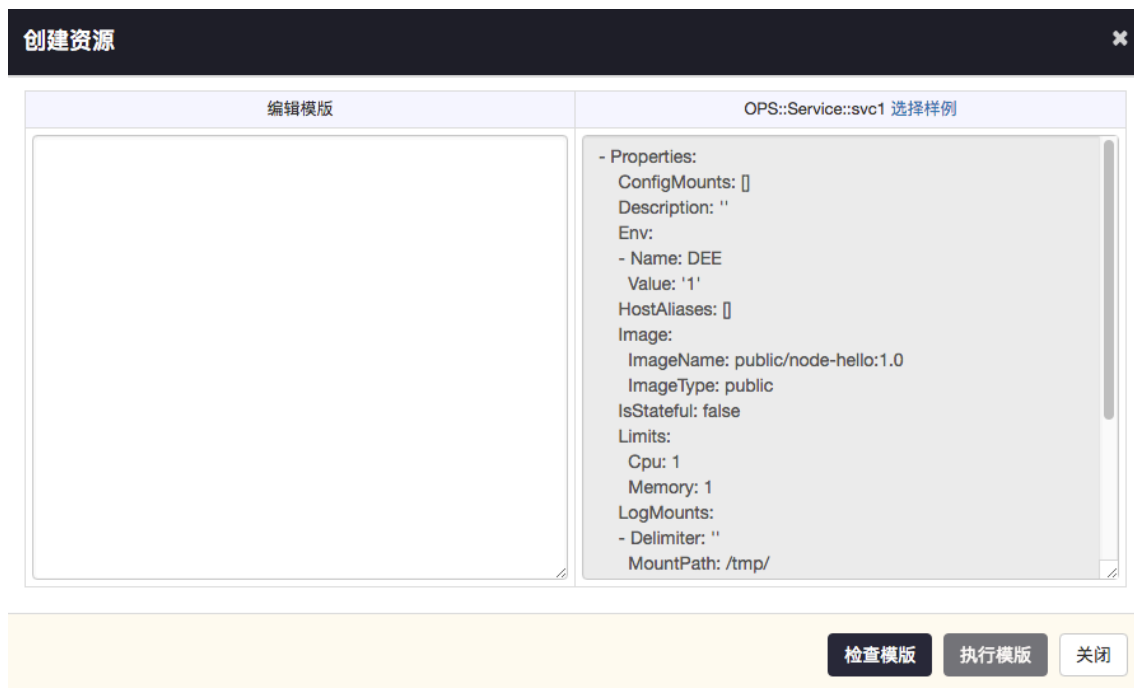
↔
⊕
创建
删除
修改

出口ip	目的网段	集群	租户
XXXXXXXXXX	10.0.0.0/8	k8s_cluster_online	内部系统
XXXXXXXXXX	110.0.0.0/8	k8s_cluster_online	内部系统

功能特性——服务编排

» 问题?

- 环境部署数量多，相似度高



功能特性——服务编排

» 快速导入、批量部署

```
- Properties:
  ConfigMounts: []
  Description: "
  Env:
  - Name: DEE
    Value: '1'
  HostAliases: []
  Image:
    ImageName: public/node-hello:1.0
    ImageType: public
  IsStateful: false
  Limits:
    Cpu: 1
    Memory: 1
  LogMounts:
  - Delimiter: "
    MountPath: /tmp/
  Ports:
  - ContainerPort: 8080
    HostPort: 8080
    Protocol: TCP
  Replicas: 2
  VolumeMounts: []
ResourceId: OPS::Service::svc1
```



```
- Properties:
  ConfigMounts: []
  Description: "
  Env:
  - Name: DEE
    Value: '1'
  HostAliases: []
  Image:
    ImageName: public/node-hello:1.0
    ImageType: public
  IsStateful: false
  Limits:
    Cpu: 1
    Memory: 1
  LogMounts:
  - Delimiter: "
    MountPath: /tmp/
  Ports:
  - ContainerPort: 8080
    HostPort: 8080
    Protocol: TCP
  Replicas: 2
  VolumeMounts: []
ResourceId: OPS::Service::svc2
```

功能特性——自定义DNS

» 问题?

- 频繁的内部DNS调整

DNS域管理

管理DNS域

总览

创建 删除 管理DNS记录

名称	种
internal.lede.com.	...

DNS记录管理

管理DNS记录

总览 展开 v

创建 删除

名称	类型	值	TTL(秒)
fa-common1.hangzhou-1.kafka	A	...	600
fa-common2.hangzhou-1.kafka	A	...	600
fa-common3.hangzhou-1.kafka	A	...	600
fa-common1.hangzhou-1.zookeeper	A	...	600

功能特性——配置中心

» 问题?

— 相同业务的不同环境配置不同

创建服务
✕

选择镜像
基础配置
动态配置
端口映射
状态跟踪

【启动命令】 启动容器后执行的CMD命令,会覆盖镜像原有的CMD命令

【启动参数】 启动容器后执行命令的参数,不会覆盖镜像原有的CMD命令

启动命令

启动参数

【环境变量】 部署容器时, 通过启动参数-e "key=value"传递环境变量

+ -

【自定义配置项】 添加和挂载自定义的配置项, 可在 [配置中心] 中查看和管理

选择
创建
+ -

【注入hosts】 部署容器时, 加入自定义hosts, 多个hostnames以逗号分隔

IP	hostnames
<input style="width: 100%;" type="text"/>	<input style="width: 100%;" type="text"/>

创建容器
关闭

创建配置
✕

选择标签*

配置命名*

配置路径*

默认权限*

配置内容*

导入模版+

```
[production]
arg1=1
arg2=2
```

保存配置
关闭

功能特性——公有云接入

» 问题?

- 不同基础环境的部署需求

编辑区域

×

基本信息 系统配置

【基础系统】

外部系统	选择类型	AWS
	AWS区域	中国(北京)
	密钥ID	
	密钥Key	
	SSH端口	22

技术要点

技术要点

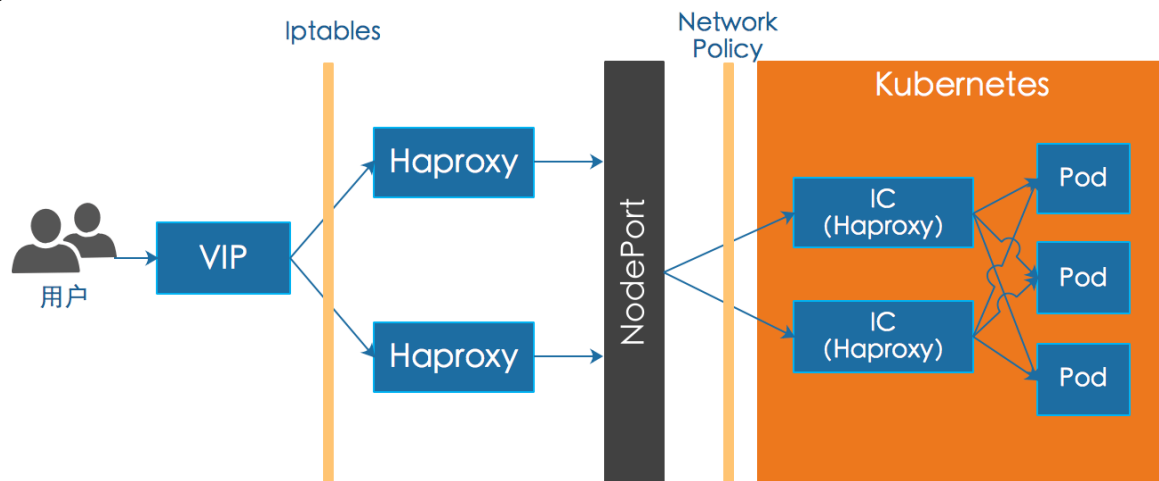
» Kubernetes资源封装

- 无状态服务: Deployment
- 有状态服务: StatefulSet
- 配置中心: ConfigMap
- 云存储: PersistentVolume
- 安全组: NetworkPolicy
- 负载均衡器: Ingress
- 自动扩容: Horizontal Pod Autoscaler (HPA)

技术要点

» 负载均衡器

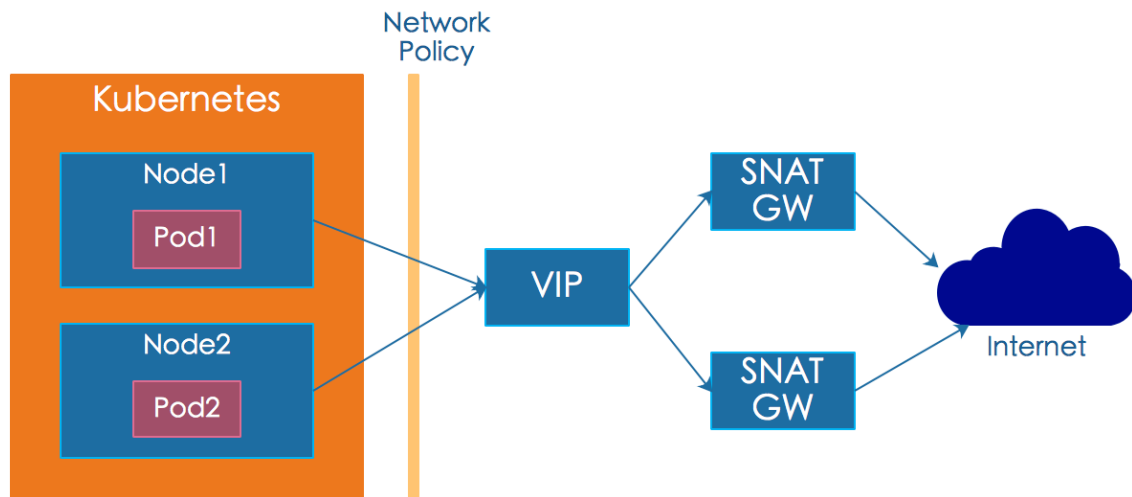
- Keepalived + Haproxy + IngressController
- 防火墙、安全组



技术要点

» SNAT网关

- Node配置策略路由转发到网关
- 通过ipset判断租户归属，做SNAT



技术要点

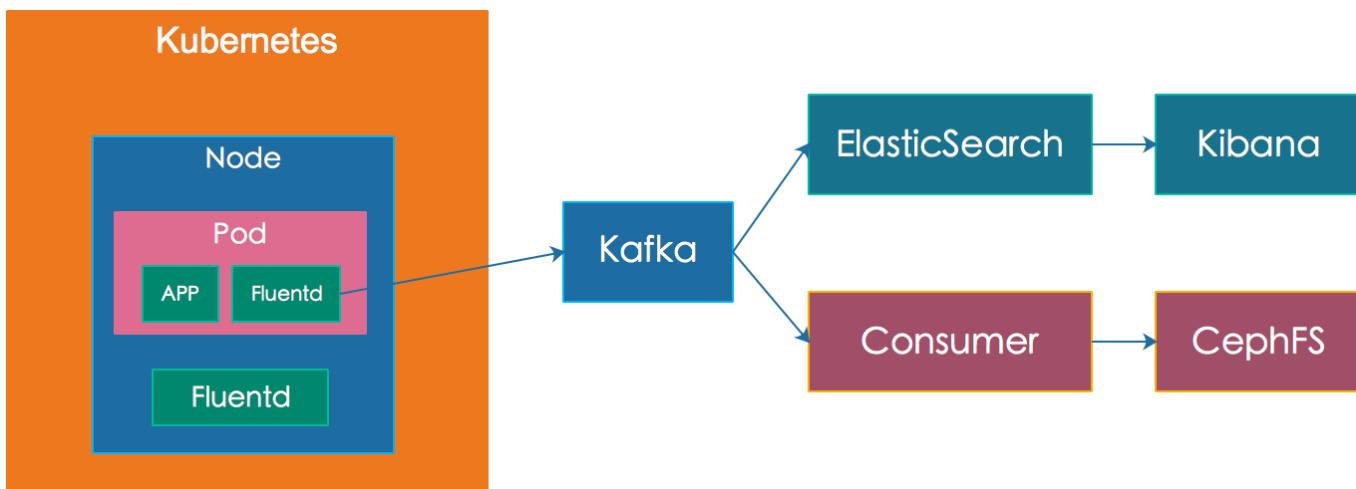
» 存储

- 网络存储 (Ceph)
 - RBD、CephFS
 - 提供云硬盘服务和离线备份 (SSD+SAS)
- 本地存储
 - 需要高速缓存的场景 (如: Redis持久化)
 - 本机SSD

技术要点

» 日志收集

- Fluentd + ElasticSearch + Kibana (EFK)
- 容器内日志/标准输出



技术要点

» 监控告警

- Prometheus
 - 容器资源监控
 - 接口监控
 - 日志监控 (ES)
- Zabbix
 - 物理节点

近期将在社区开源，欢迎关注

谢谢！