QCon 全球软件开发大会
INTERNATIONAL SOFTWARE DEVELOPMENT CONFERENCE

BEIJING 2017

搜狗商业云平台实践与思考

王宇

主办方 Geekbang. 极客邦科技 InfoQ

王宇

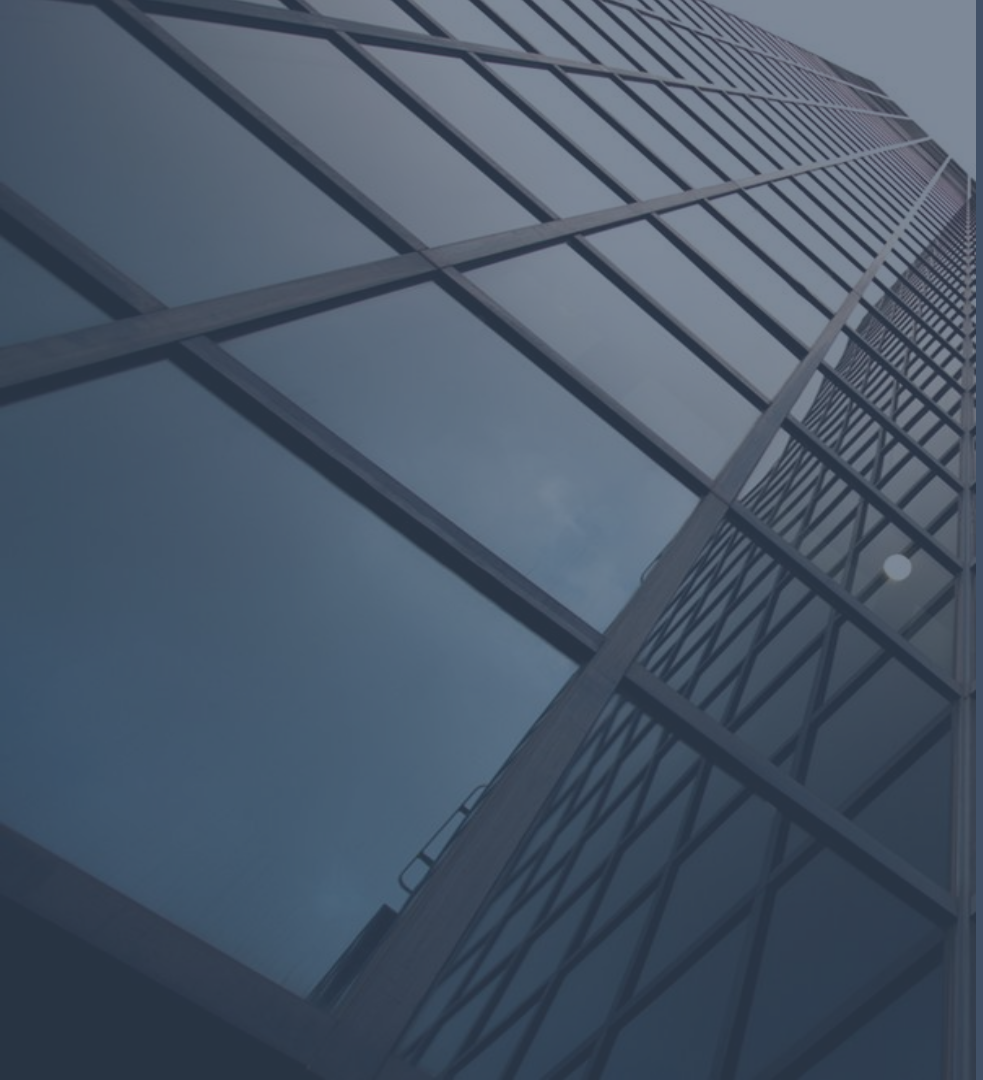
About me

搜狗商业平台研发部　资深开发工程师

主要负责商业平台基础研发工作，包括报文系统、分布式缓存、商业云平台等系统研发工作

目前重点关注分布式、大数据、云计算等相关领域

曾就职于腾讯



CONTENT

搜狗商业平台特点

见多识广

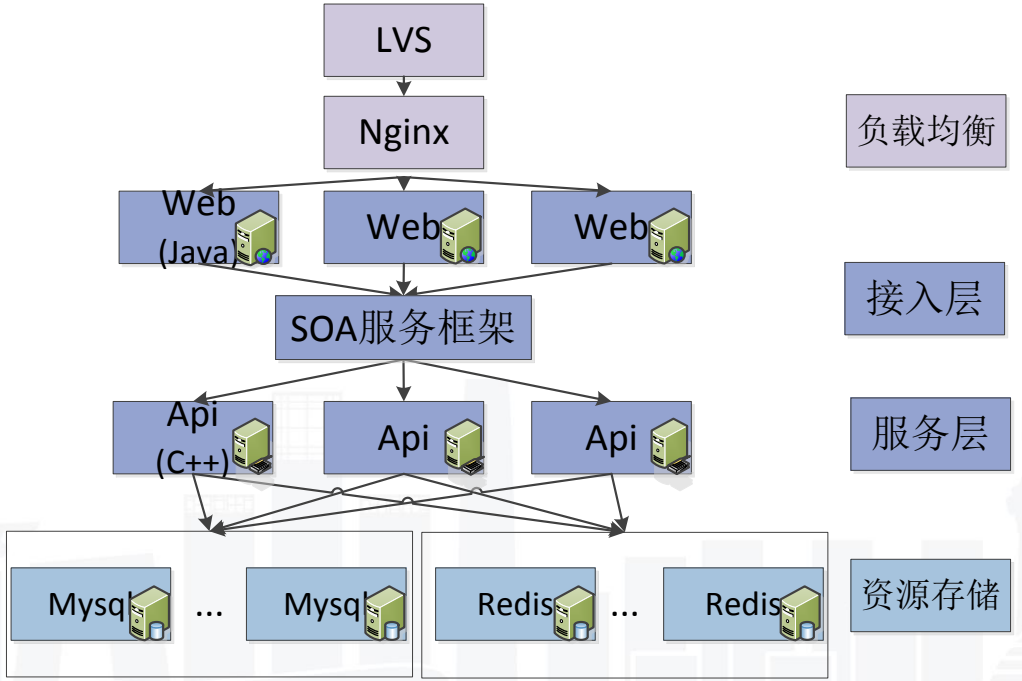
- 多—服务个数多，迭代版本多
- 广—技术体系涵盖广

又快又稳

- 快—性能决定收益
- 稳—稳定压倒一切

典型应用的系统架构

- 分层结构、服务化
- 跨语言、多环境
- 物理隔离



面临的挑战

- 环境复杂性
- 业务低峰期资源利用率低
- 弹性伸缩能力不足

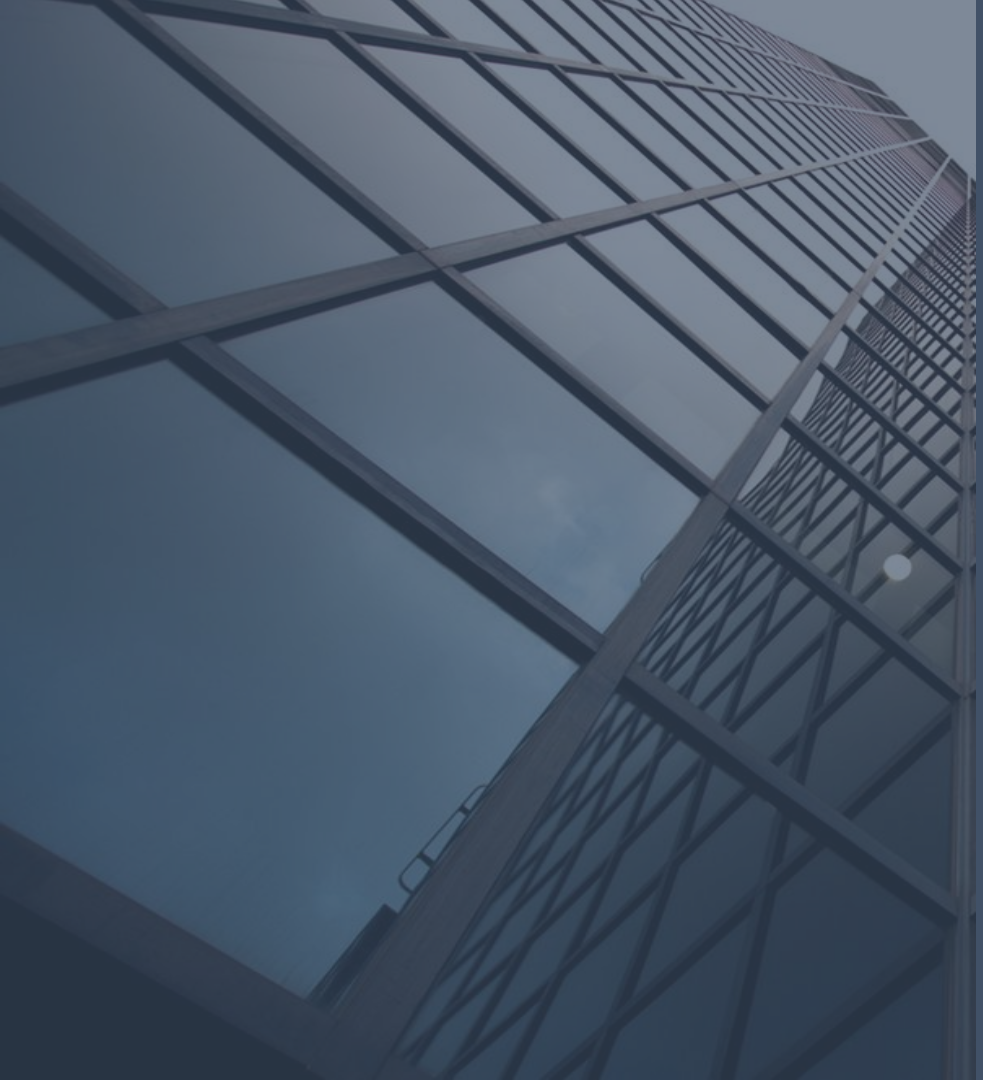
商业云平台

- 标准化
- 线上线下统一

- 资源池化
- 资源调度

- 优化自动化平台
- DevOps

- 弹性伸缩
- 弹性调度



CONTENT

行业现状

Container RunTime

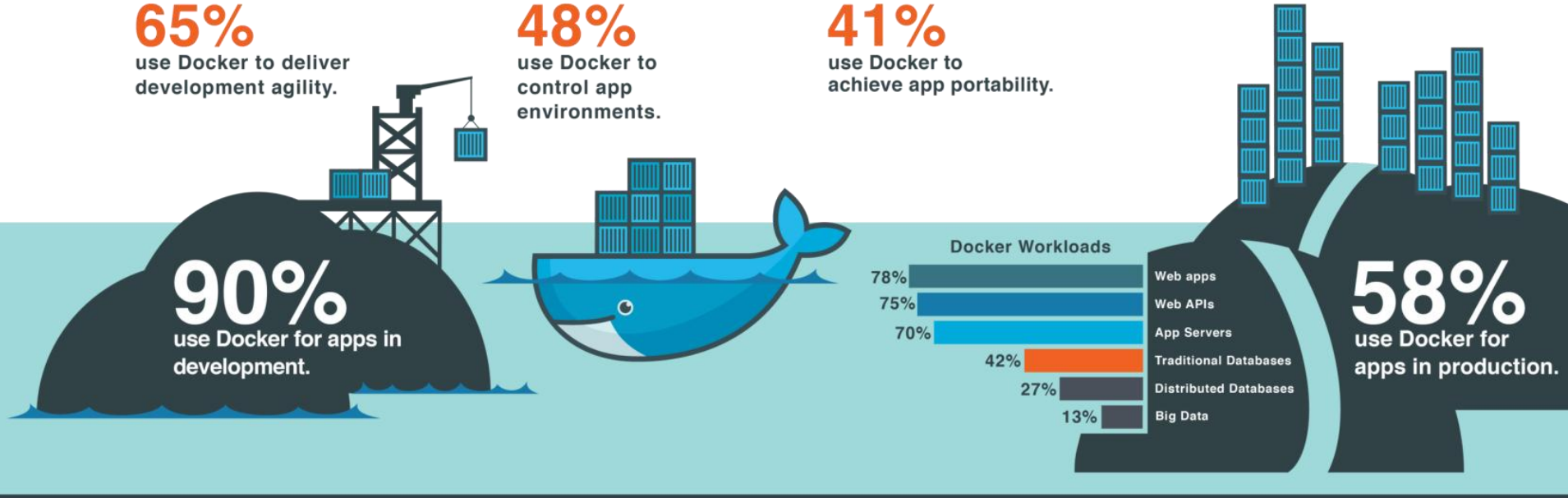
- Docker
- Rkt
- OCI

Scheduling &Orchestration

- Kubernetes
- MESOS
- SWARM

Docker

标准化的构建、交付、运维手段



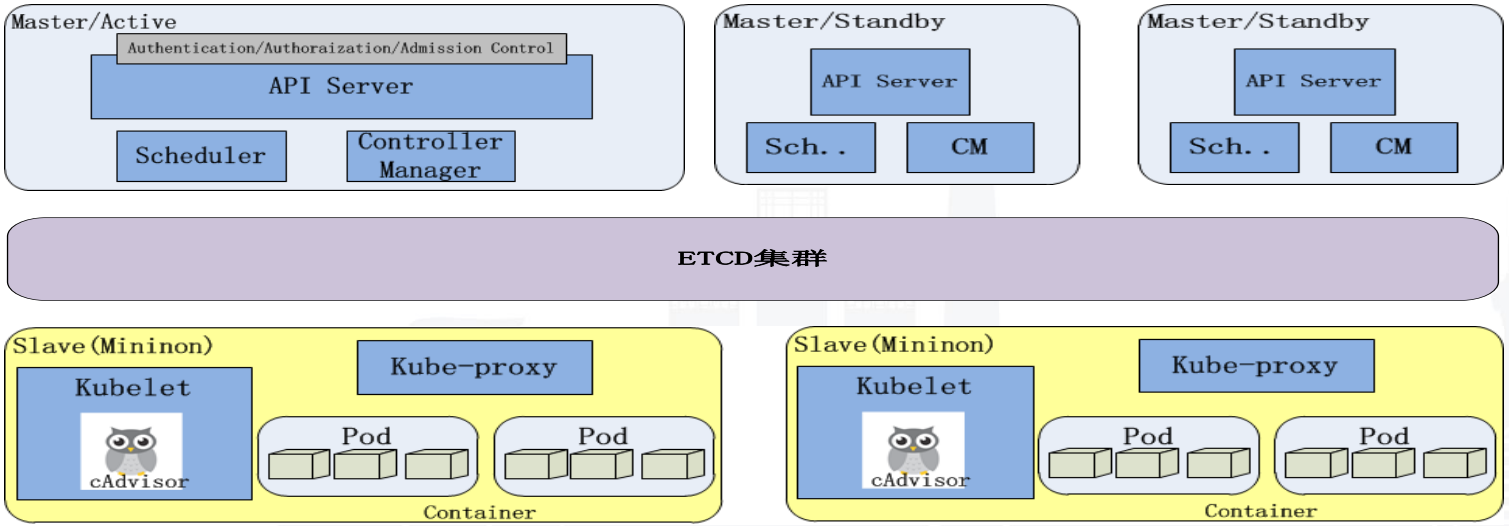
90%
plan dev environments around Docker.

80%
plan DevOps around Docker.

docker

Kubernetes(K8s)

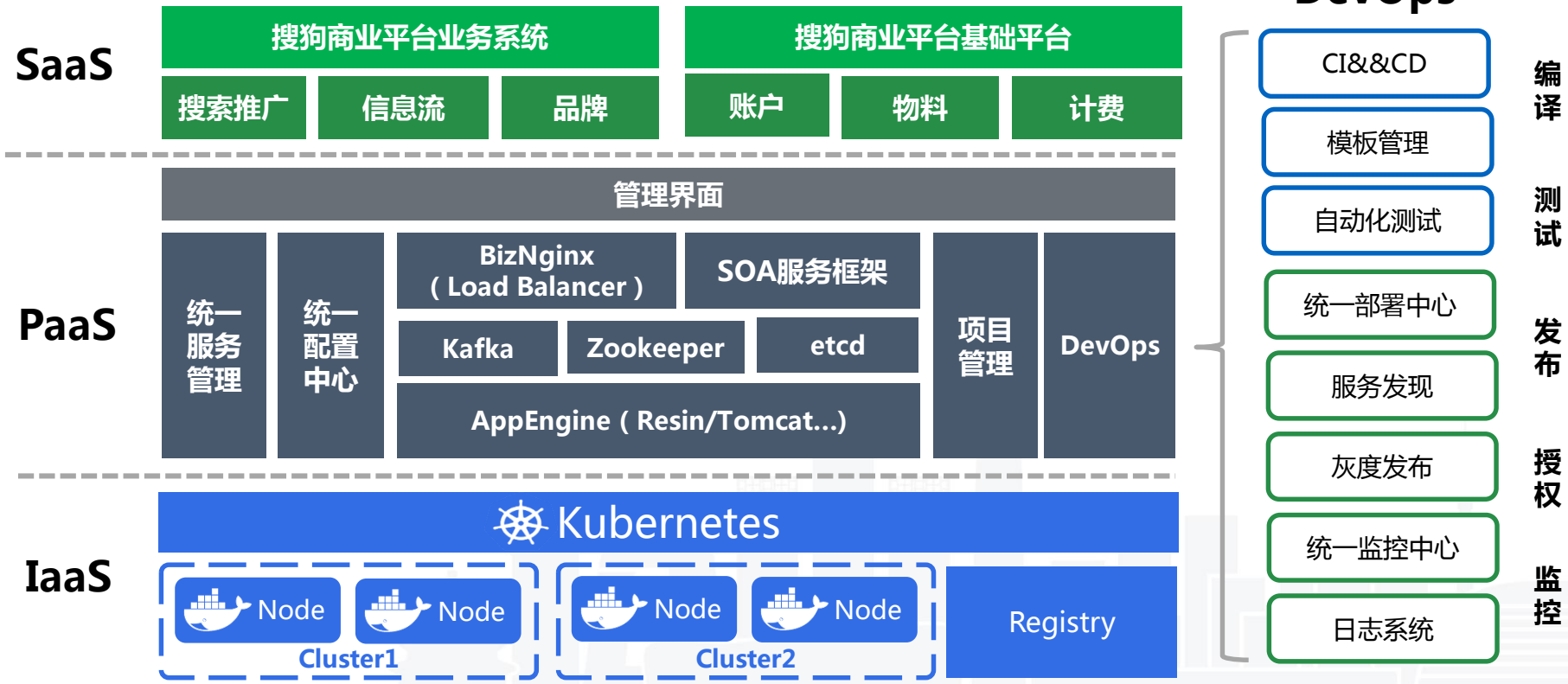
- 蓬勃发展的社区： GitHub、Stackflow
- 优秀的设计理念： Pod、ReplicationController、Deployment、Label 等
- 项目匹配：规模在数千台、需要快速落地

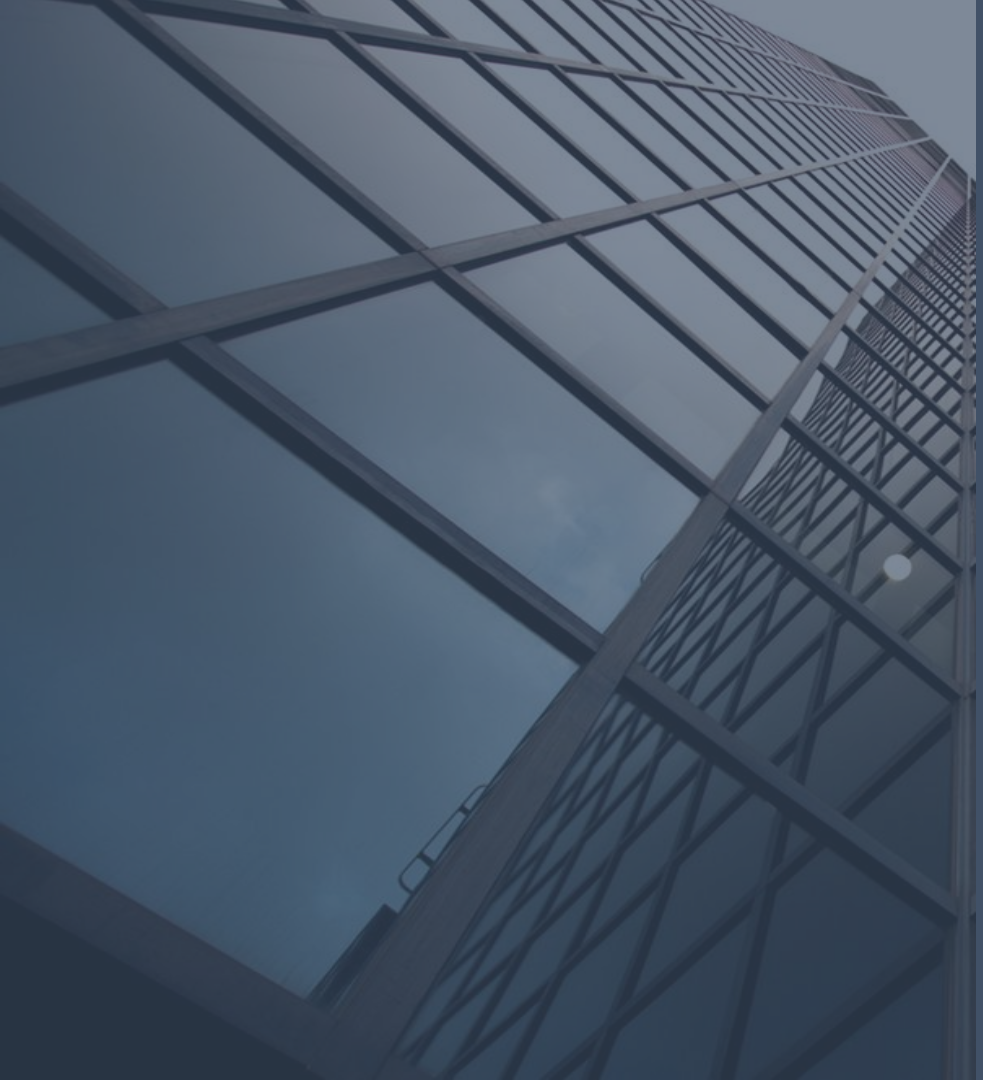


商业云平台

- 云平台=K8s+Docker？
  - 持续集成、服务发现、服务授权？
  - 已有开发、测试、运维流程如何对接？
  - 自动化平台如何实现？
- 自研PaaS

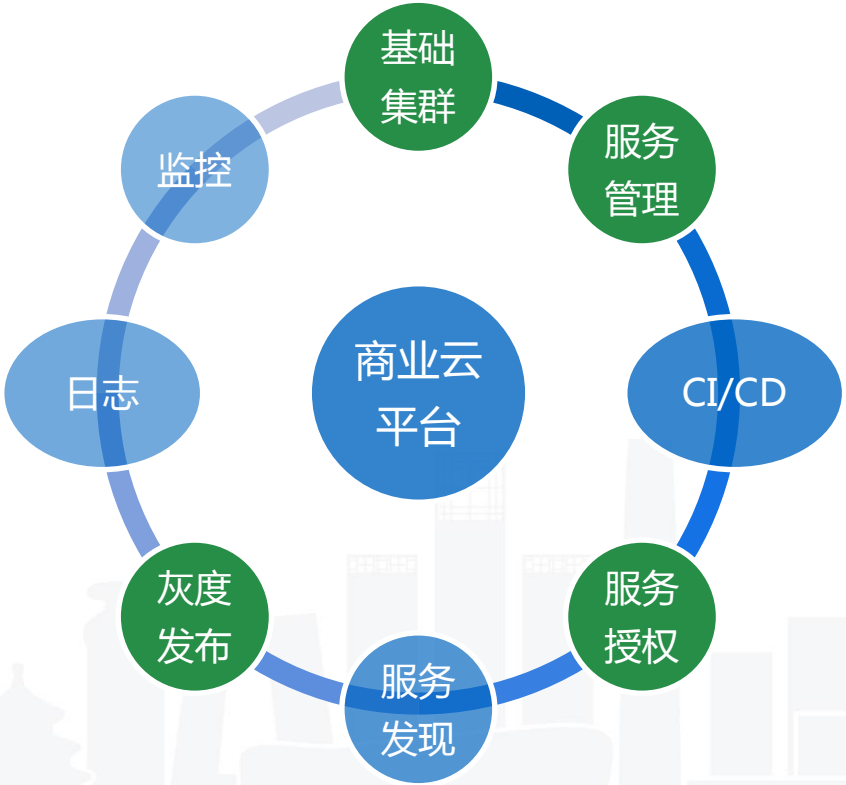
商业云平台


SaaS
搜狗商业平台业务系统
搜索推广
信息流
品牌
搜狗商业平台基础平台
账户
物料
计费
PaaS
管理界面
统一服务管理
统一配置中心
BizNginx (Load Balancer)
SOA服务框架
Kafka
Zookeeper
etcd
AppEngine (Resin/Tomcat...)
项目管理
DevOps
IaaS
Kubernetes
Node
Node
Cluster1
Node
Node
Cluster2
Registry
DevOps
CI&&CD
模板管理
自动化测试
统一部署中心
服务发现
灰度发布
统一监控中心
日志系统
编译
测试
发布
授权
监控



CONTENT

云平台实践



实践

基础集群

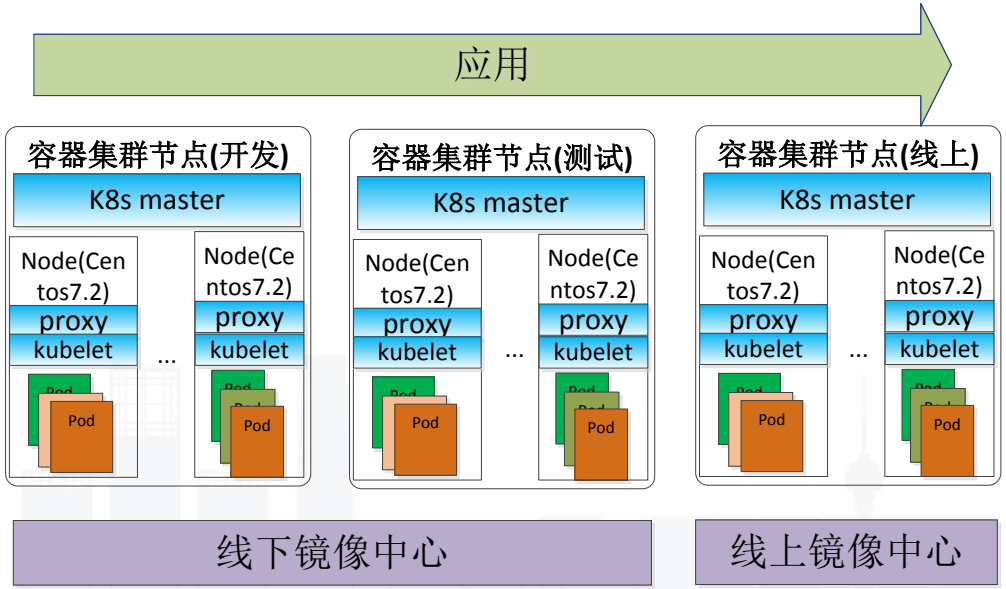
服务管理

服务授权

灰度发布

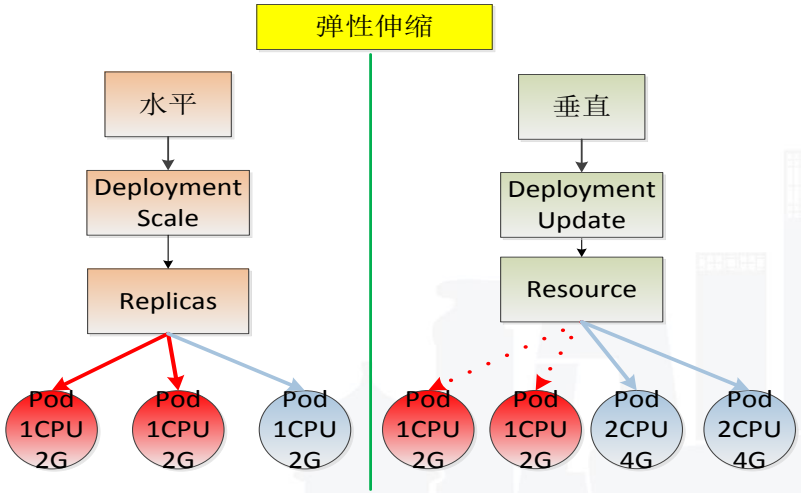
基础集群

- 基础集群：
  - 多环境隔离
  - K8s集群：线上、测试、开发
  - 镜像中心：线上、线下
- 应用
  - 多环境隔离



基础集群

- 自动Failover
- 提升弹性：
  - 水平扩缩容
  - 垂直扩缩容



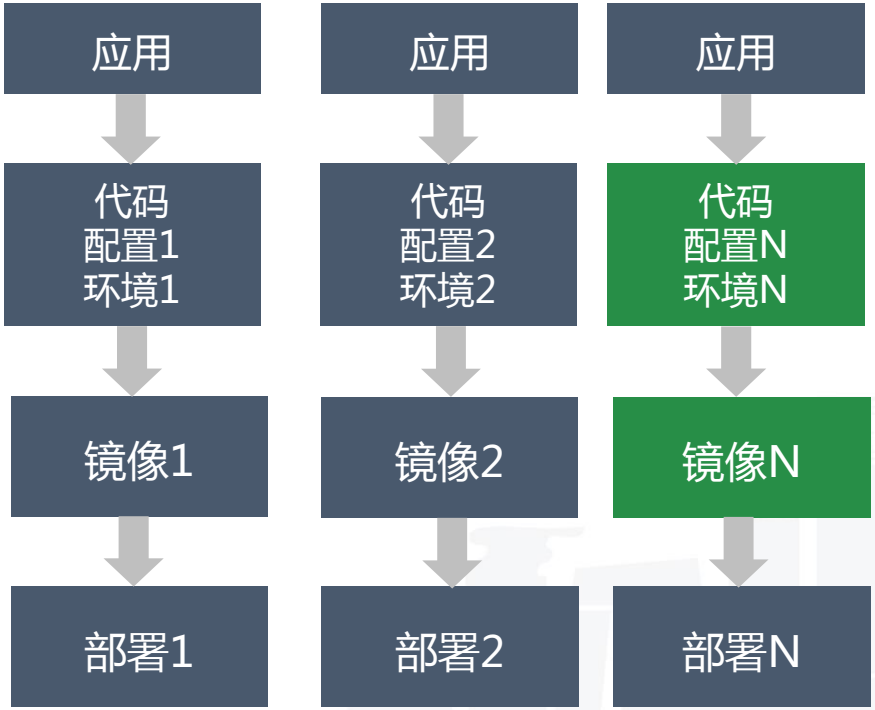
- 提升资源利用率：
  - 任务混布
  - Qos资源定义：limit,request

实践

基础集群　服务管理　服务授权　灰度发布

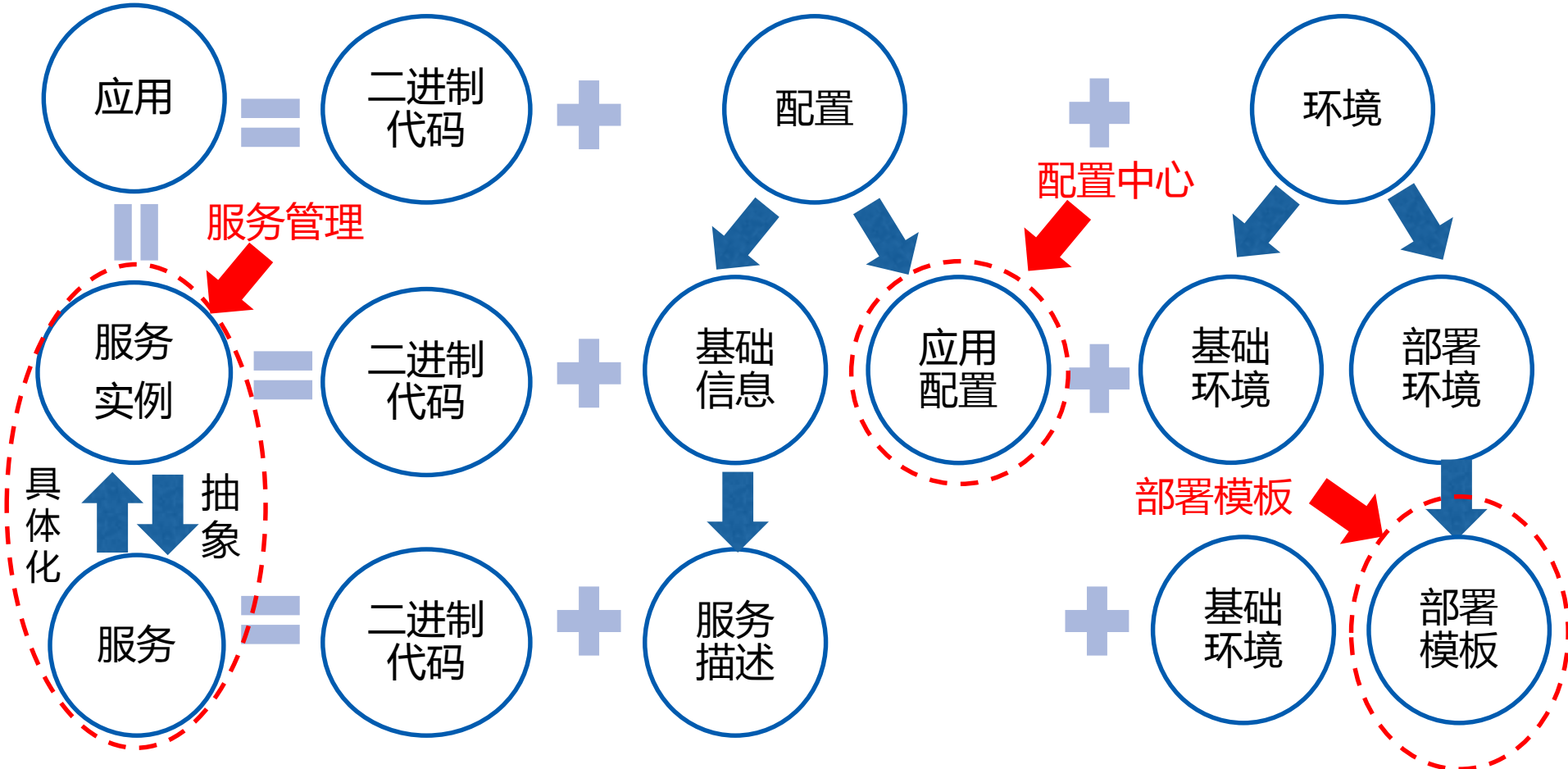



挑战



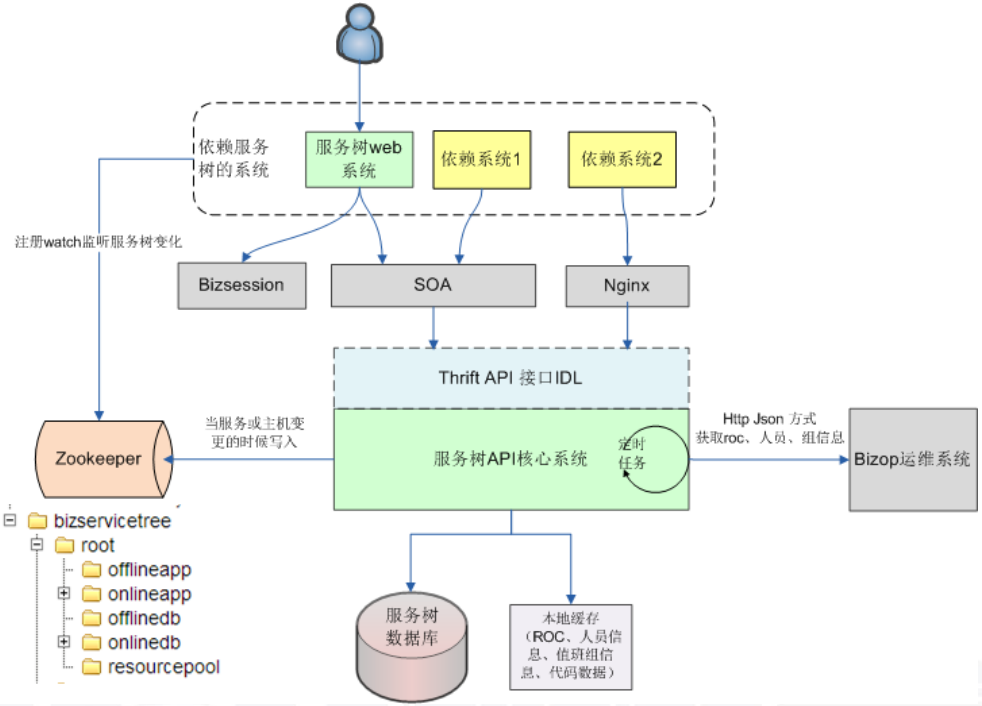
提升可重用性

应用到服务


应用
二进制代码
配置
环境
配置中心
服务管理
服务实例
二进制代码
基础信息
应用配置
基础环境
部署环境
具体化
抽象
部署模板
服务
二进制代码
服务描述
基础环境
部署模板

服务管理平台

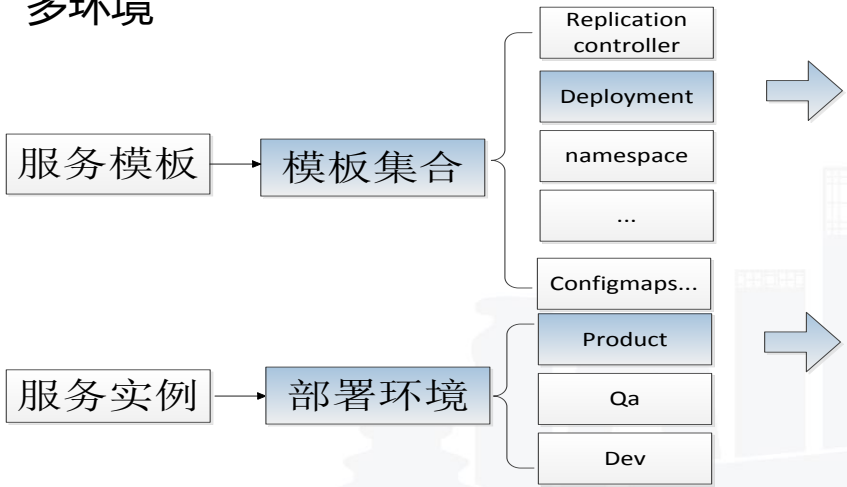
- 服务和服务实例区别：
  - 服务：应用的抽象
  - 服务实例：对应不同环境会有多个服务实例
- 服务实例全局唯一标识：
  - 环境路径+服务名
  - 服务实例名：${ENV_PATH}/${SVC_NAME}
  - 唯一标识贯穿在服务实例的生命周期中



部署模板

部署模板

- Deployment、namespace、Configmaps等
- 多环境



Deployment模板

```
kind: Deployment
apiVersion: extensions/v1beta1
metadata:
   name: {{.ServiceInStanceName}}
   namespace: {{.ServiceName}}-{{.ServiceInstanceId}}
   labels:
      name: {{.ServiceInStanceName}}
spec:
  replicas: {{.Replicas}}
  selector:
    matchLabels:
      app: {{.ServiceInstanceName}}
  template:
    metadata:
      labels:
         app: {{.ServiceInstanceName}}
         version: {{.SrcVersion}}.{{.PubVersion}}
    spec:
      hostNetwork: {{.IsHost}}
      containers:
      - name: {{.ServiceName}}
        image: {{.ImageName}}
        env:
        - name: KUBE_SERVICE_NAME
```

服务实例名

服务名

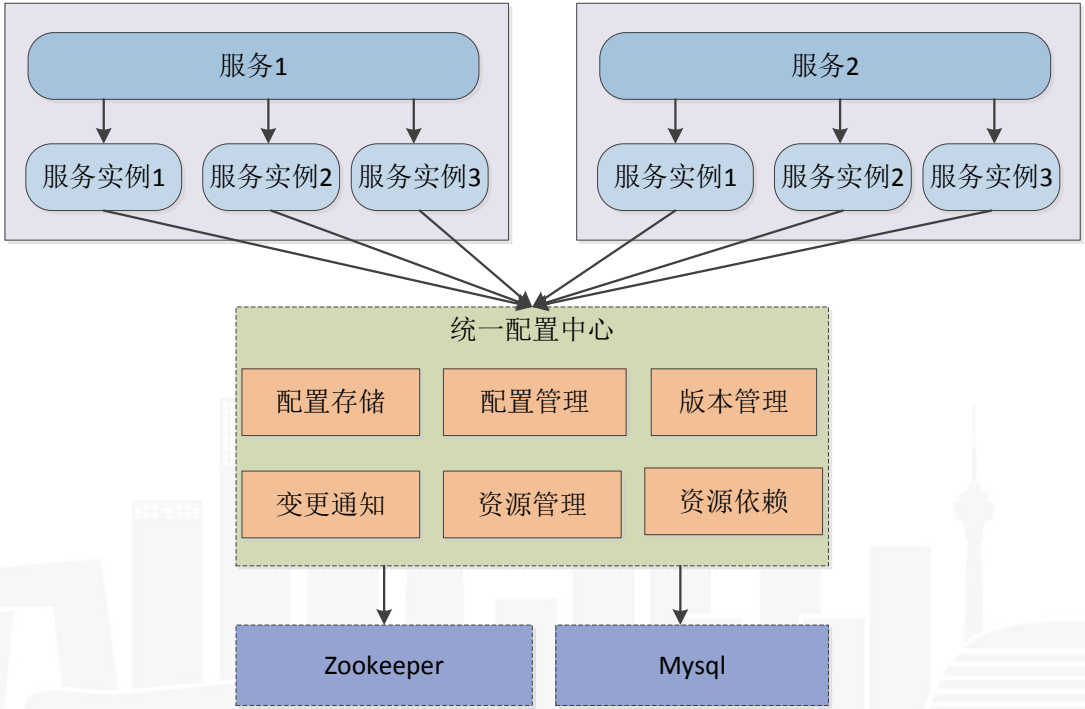
副本数

版本号

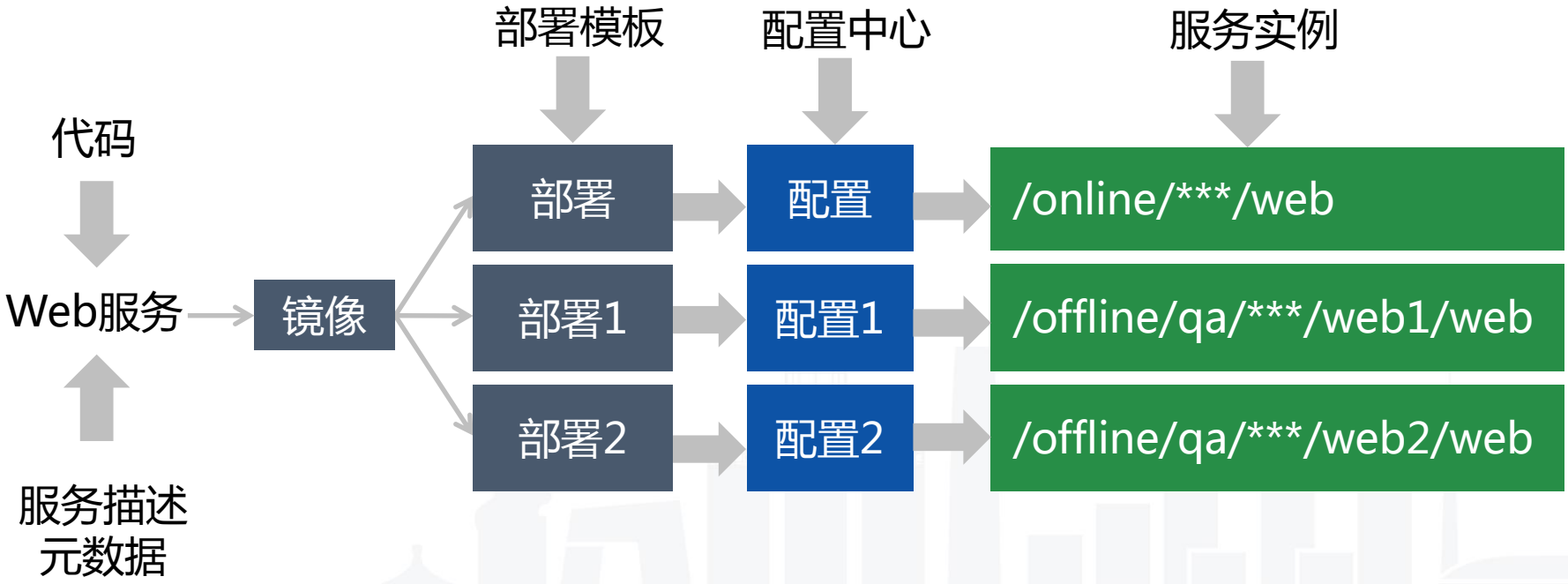
镜像名

统一配置中心

- 多环境支持：
- 服务实例上传应用配置
- 应用配置管理：
  - 支持配置备份存储
  - 支持配置变更热加载



服务管理示例


部署模板
配置中心
服务实例
代码
Web服务
镜像
服务描述
元数据
部署
部署1
部署2
配置
配置1
配置2
/online/***/web
/offline/qa/***/web1/web
/offline/qa/***/web2/web

实践

基础集群

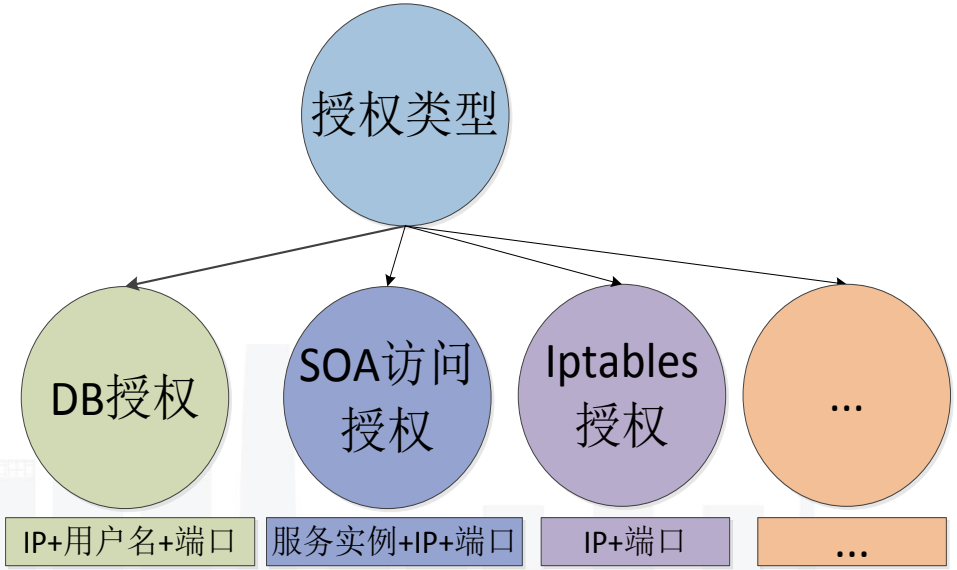
服务管理

服务授权

灰度发布

服务授权

- 服务授权：
  - 授权的重要性
    - 防止测试流量打到线上
    - 防止恶意访问等
  - 授权是云平台的核心功能
- 服务授权的挑战
  - 服务依赖关系
  - 在云平台下，容器IP会随时变动
  - 多种授权类型：粒度不同



服务授权

Who

- 如何自动计算给哪些服务授权

When

- 什么时候进行授权

How

- 怎么样进行授权

Who: 配置中心

配置文件

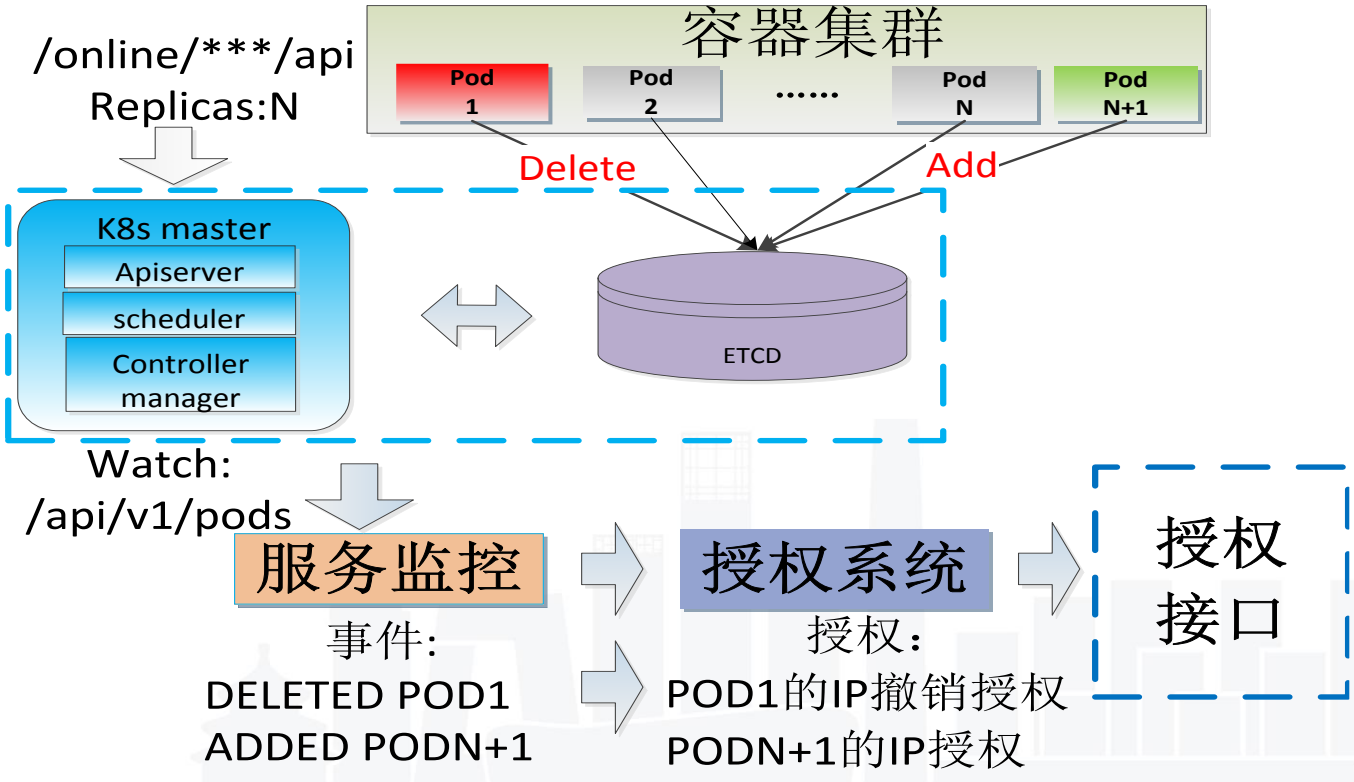
```
├──/online/***/dubhe-api   ⇐ 服务实例
│  ├──properties
│  │  ├──redis.properties
│  │  ├──memcache.properties        } 公共资源
│  │  ├──soa_registry.properties    ⟶ SOA资源
│  │  ├──sequence-server.properties ⟶ 域名资源
│  │  └──db.properties              ⟶ 数据库资源
```

```
# Shard DB common Configuration #
datasource.shard.driverClassName=com.mysql.jdbc.Driver
datasource.shard.username=***
datasource.shard.password=***
# Shard DB Configuration #
datasource.shard01=jdbc:mysql://***.dbname01.mysql:3306/
dbname01?
datasource.shard01.s1=jdbc:mysql://***.s1dbname01.mysql:3306/
dbname01?
datasource.shard01.s2=jdbc:mysql://***.s2dbname01.mysql:3306/
dbname01?
...
```

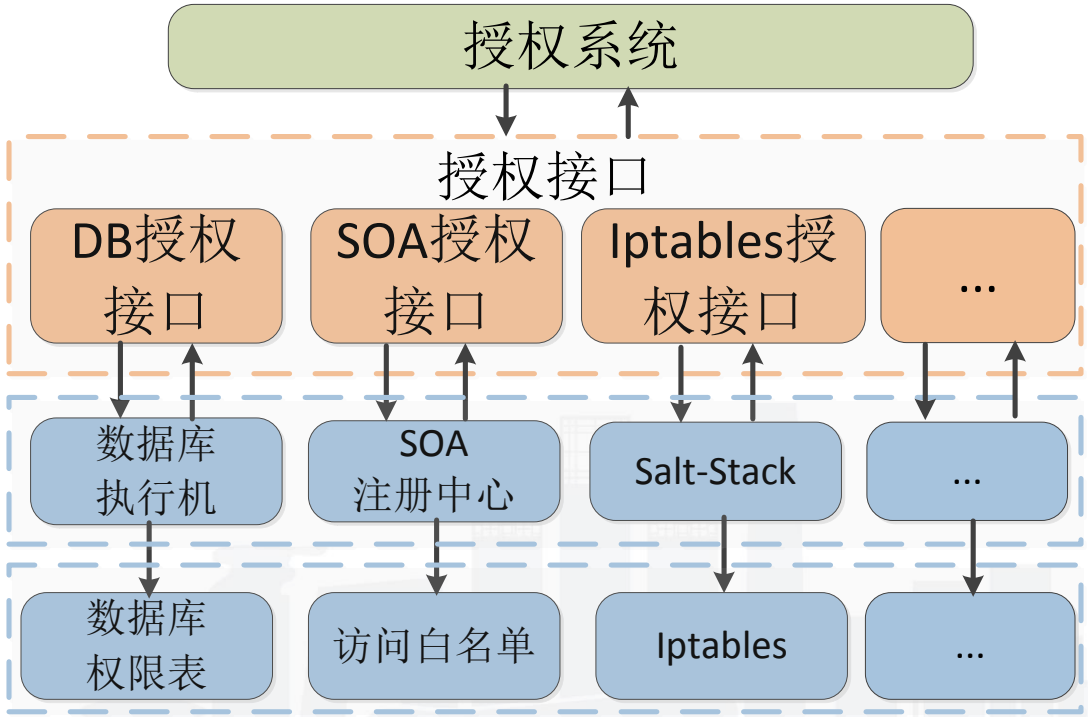
唯一标识

配置中心依赖视图

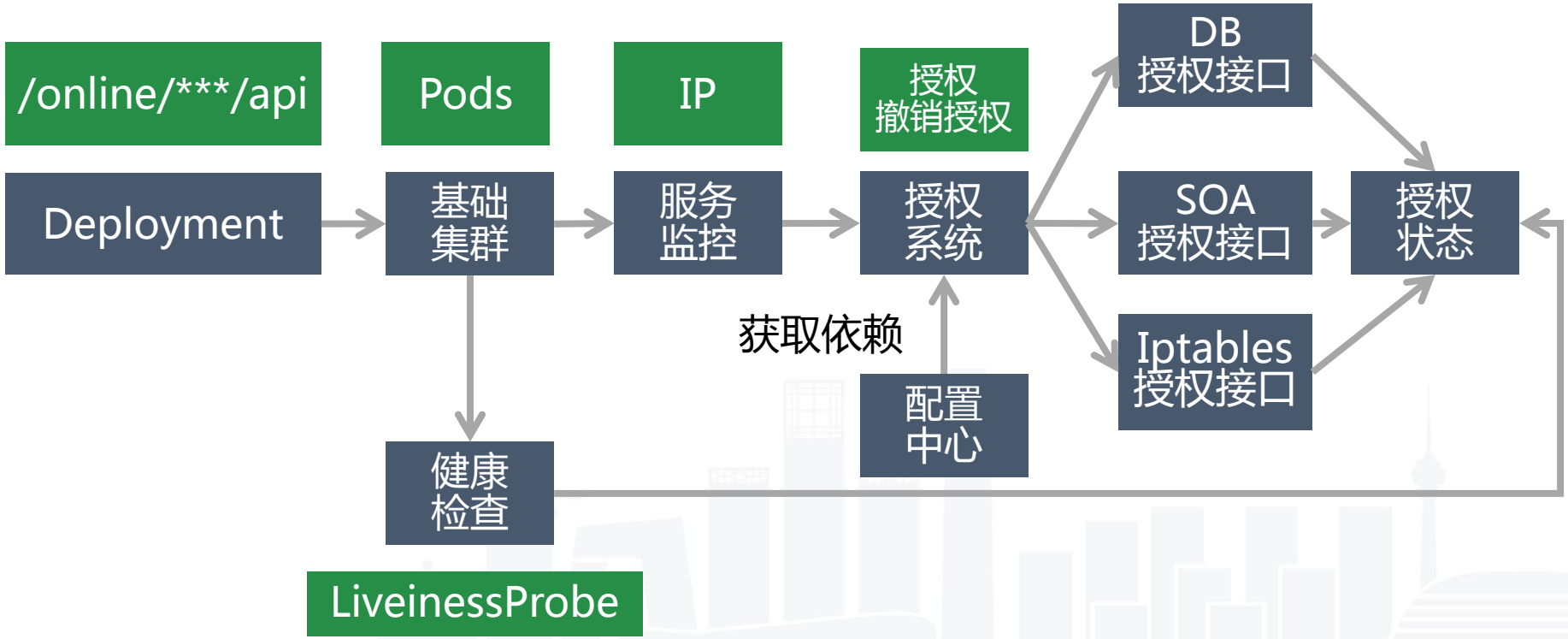
When: 容器变更


/online/***/api
Replicas:N
容器集群
Pod 1
Pod 2
.....
Pod N
Pod N+1
Delete
Add
K8s master
Apiserver
scheduler
Controller manager
ETCD
Watch:
/api/v1/pods
服务监控
授权系统
授权
接口
事件:
DELETED POD1
ADDED PODN+1
授权:
POD1的IP撤销授权
PODN+1的IP授权

How: 授权接口


授权系统
授权接口
DB授权接口
SOA授权接口
Iptables授权接口
...
数据库执行机
SOA注册中心
Salt-Stack
...
数据库权限表
访问白名单
Iptables
...

服务授权示例


/online/***/api
Pods
IP
授权
撤销授权
DB
授权接口
Deployment
基础
集群
服务
监控
授权
系统
SOA
授权接口
授权
状态
获取依赖
Iptables
授权接口
配置
中心
健康
检查
LiveinessProbe

实践

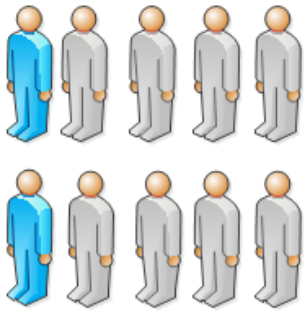
基础集群

服务管理

服务授权

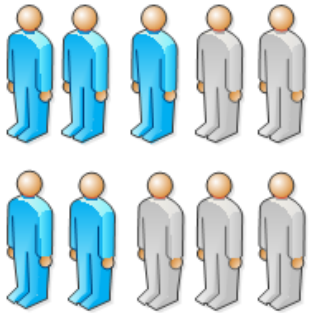
灰度发布

灰度发布



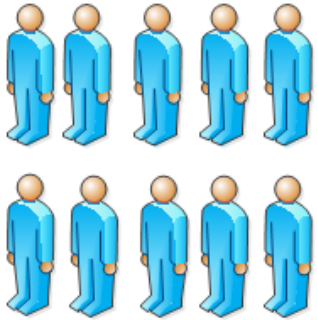
阶段一

推送20%



阶段二

推送50%



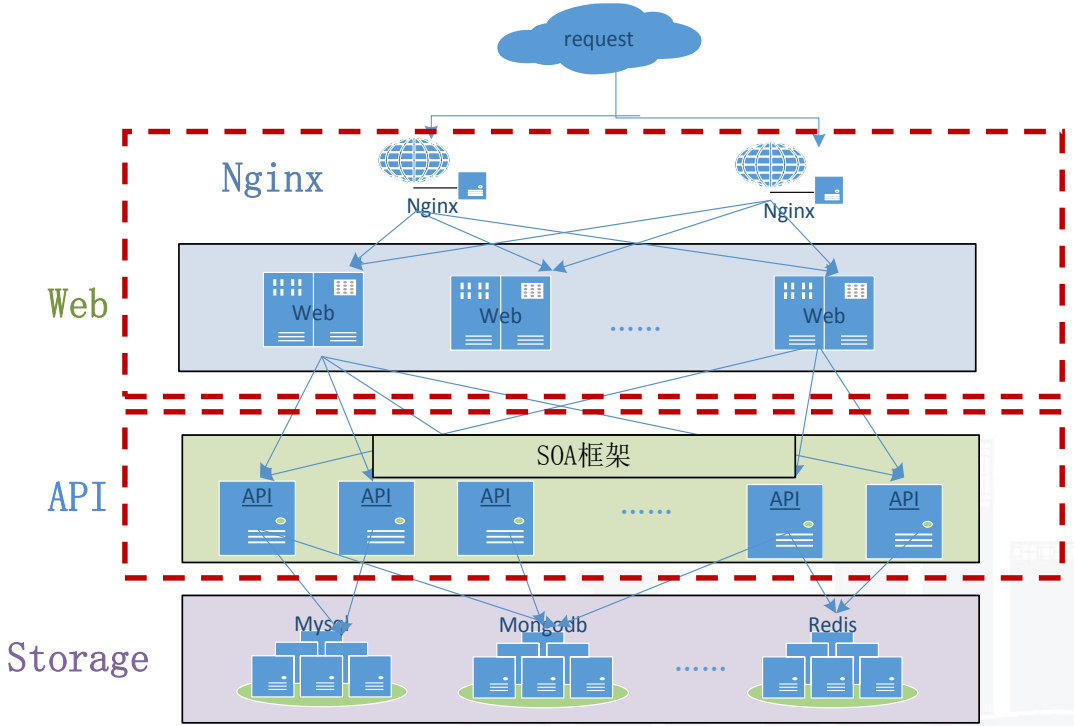
阶段三

推送100%

作用

- 降低风险
- 提升上线效率
- 收集使用反馈

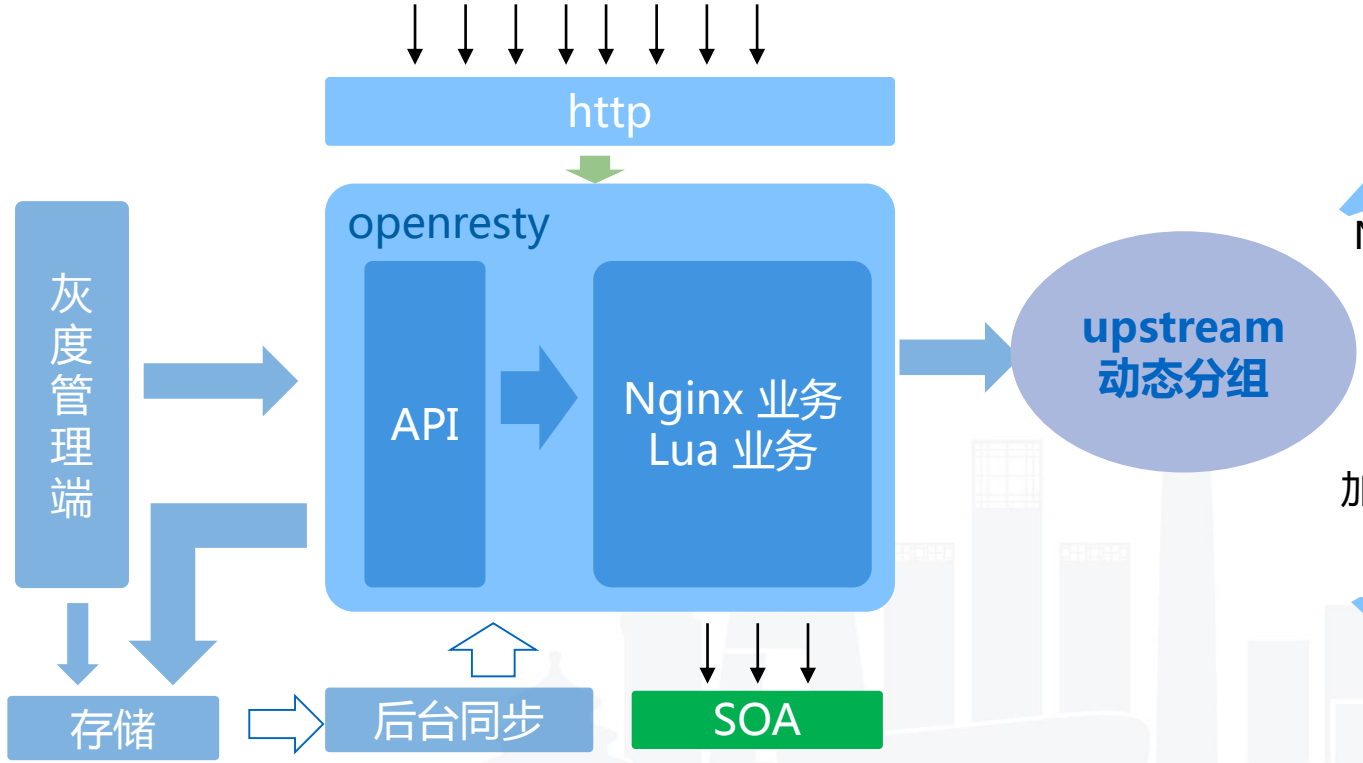
灰度发布



挑战：

- 需要在用户接入层和SOA层统筹考虑灰度机制
- 实现容器类灰度发布功能
  - 滚动升级！=灰度发布
  - 服务发现
  - 非容器兼容

接入层灰度



ngx_dynamic_upstream

Nginx开源模块，通过http api方式实现动态upstream 功能

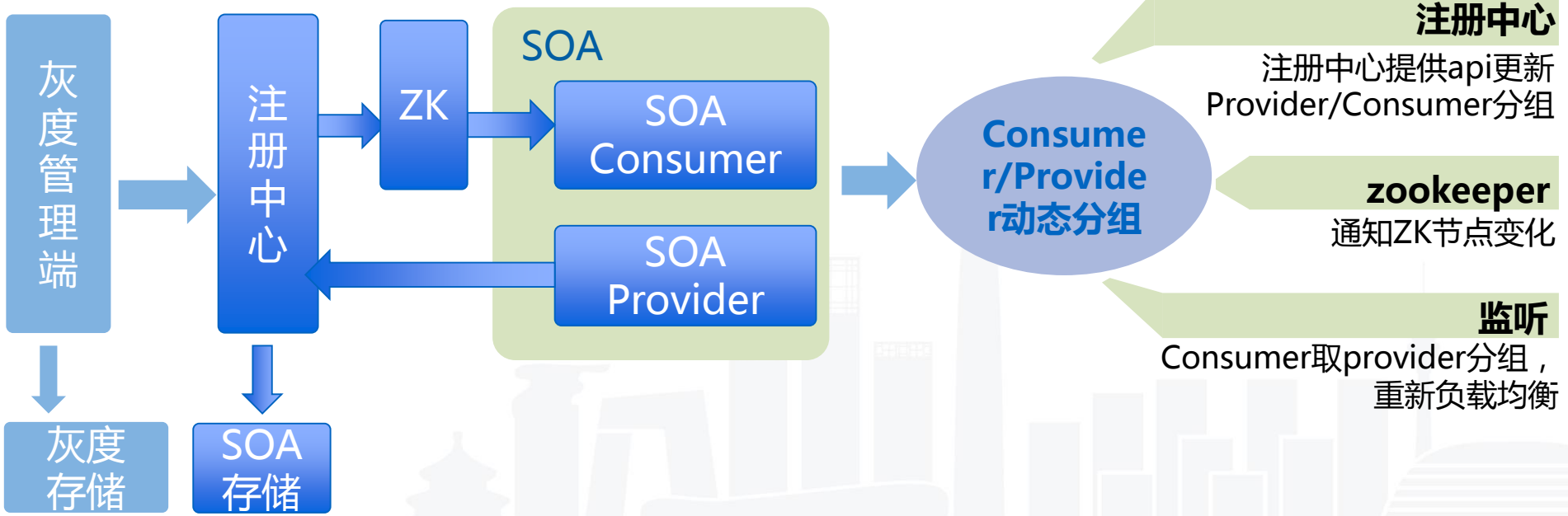
Openresty lua 处理逻辑

加载upstream信息到lua缓存，提高openresty处理性能

管理端

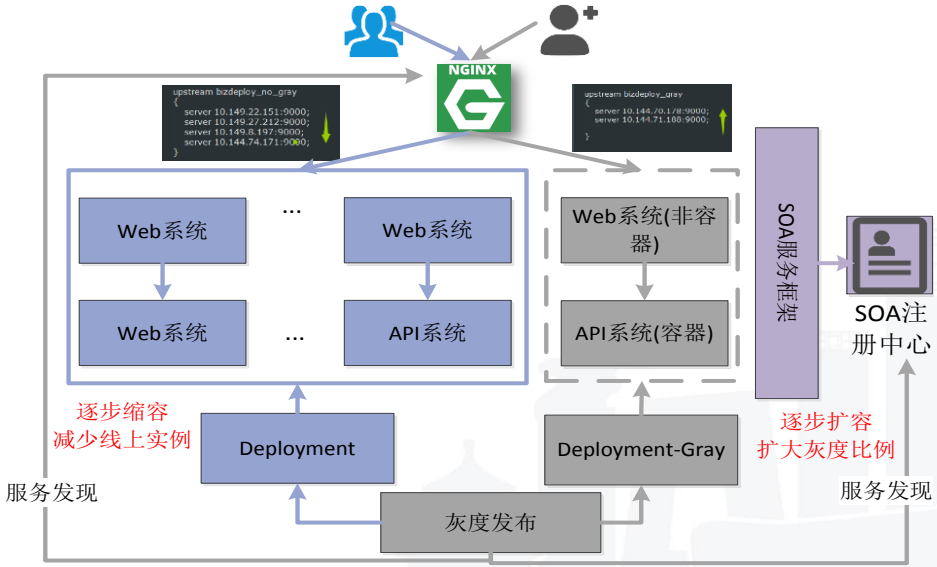
提供外界操作入口

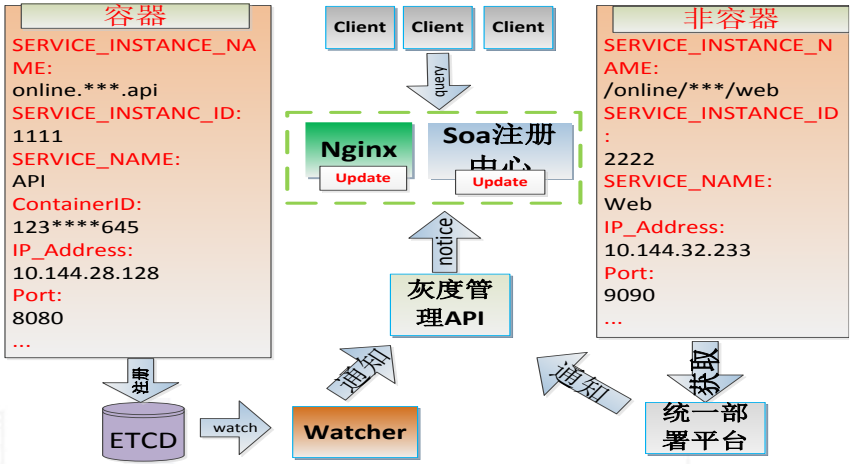
SOA层灰度


灰度管理端
注册中心
ZK
SOA
SOA Consumer
SOA Provider
灰度存储
SOA存储
Consumer/Provider动态分组
注册中心
注册中心提供api更新Provider/Consumer分组
zookeeper
通知ZK节点变化
监听
Consumer取provider分组，重新负载均衡

灰度发布

- 灰度deployment
- 服务分组
- 动态调节两个部署实例数
- 服务发现





服务发现：
- 实例注册ETCD
- Watch变更
- Nginx动态update
- Soa注册中心update

灰度发布流程

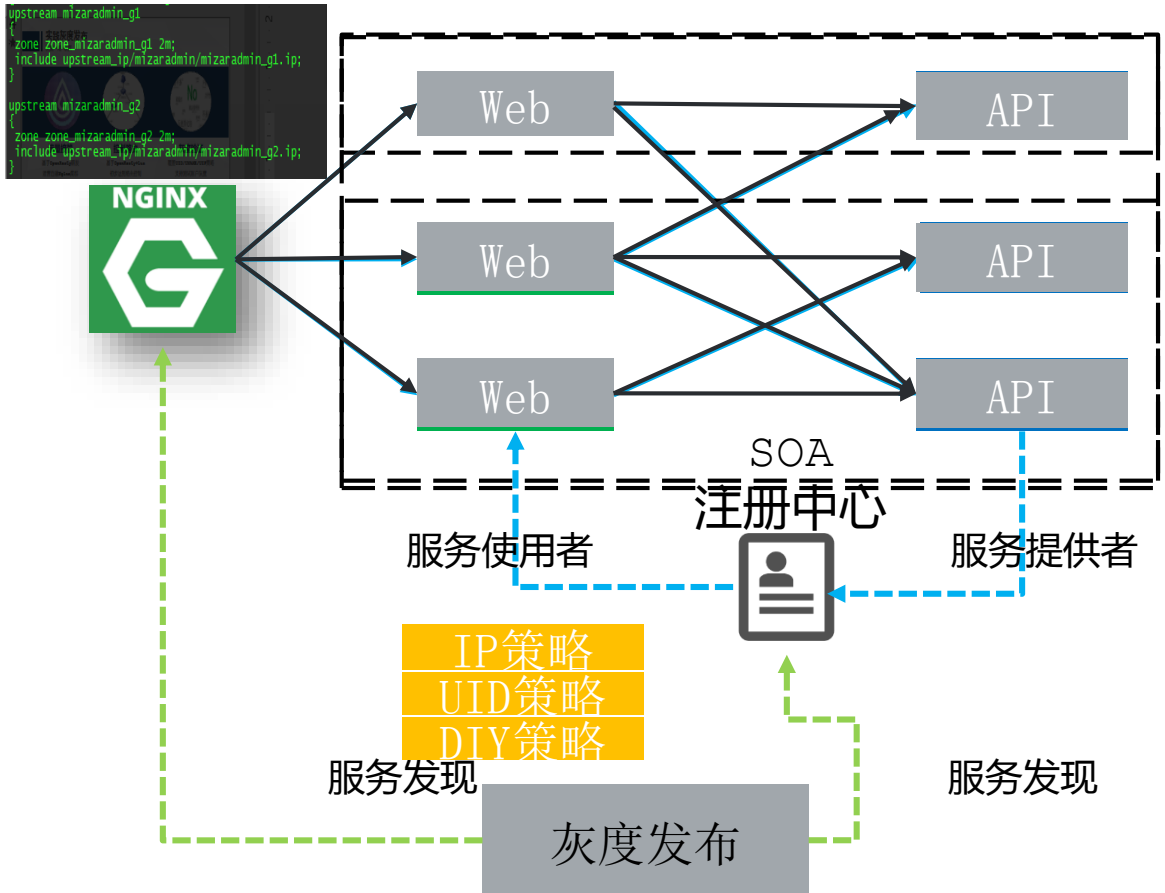
灰度发布示例(一组服务)

制定策略

预发布

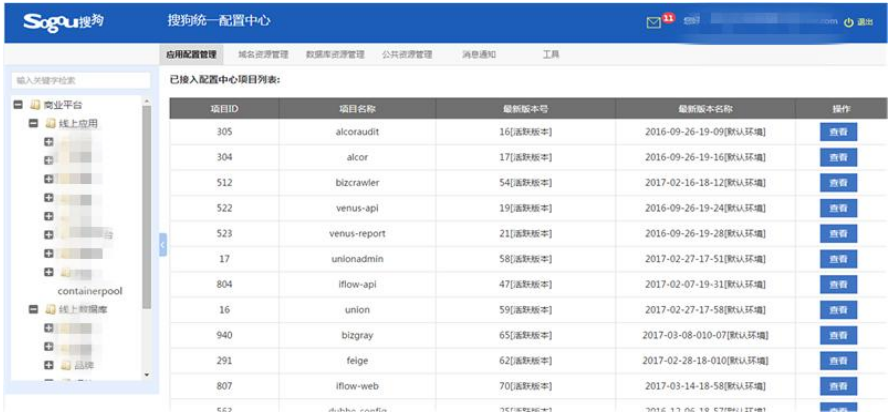
灰度分流

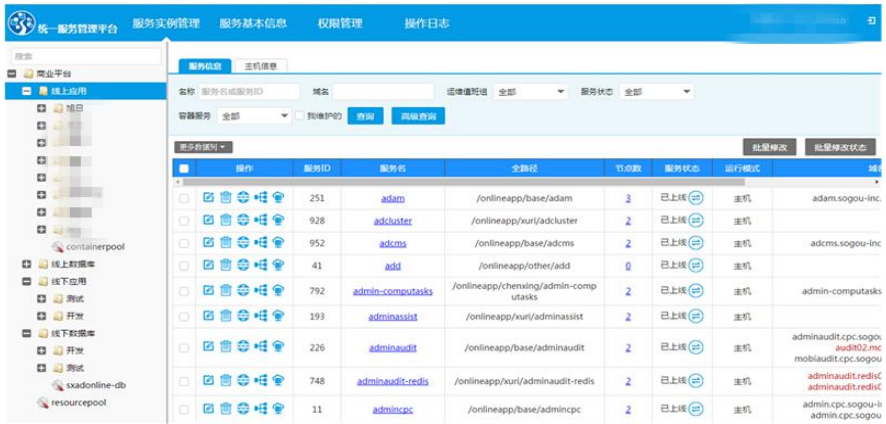
灰度收尾

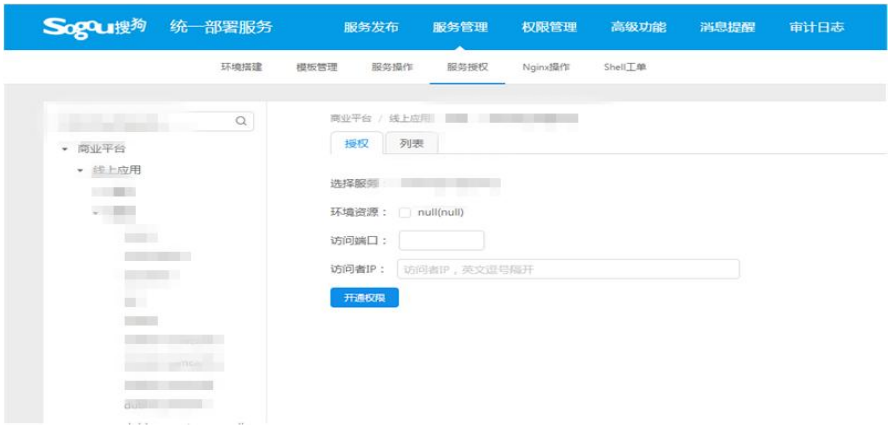


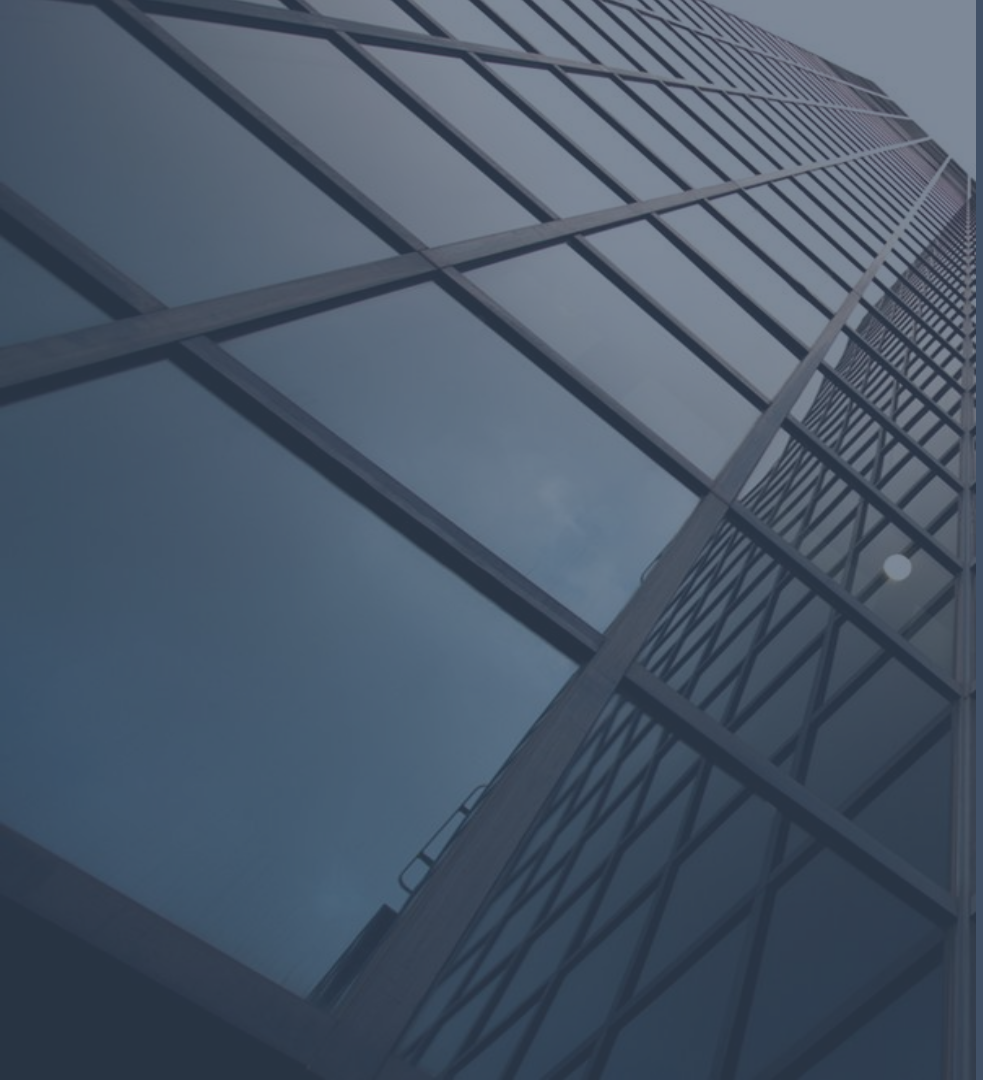
云平台截图


Sogou搜狗 商业云平台
首页 服务应用 集群 监控 镜像仓库 日志
Search
商业平台
容器管理 YAML管理 配置管理
弹性扩容 配置升级 服务级别调整 重启 暂停
服务状态：待部署
序号 容器ID 容器名 Pod名 镜像名 宿主机IP 容器状态 操作
暂无数据


Sogou搜狗 搜狗统一配置中心
应用配置管理 域名资源管理 数据库资源管理 公共资源管理 消息通知 工具
已接入配置中心项目列表：
项目ID 项目名称 最新版本号 最新版本名称 操作
305 alcoraudit 16[活跃版本] 2016-09-26-19-09[默认环境] 查看
304 alcor 17[活跃版本] 2016-09-26-19-16[默认环境] 查看
512 bizcrawler 54[活跃版本] 2017-02-16-18-12[默认环境] 查看
522 venus-api 19[活跃版本] 2016-09-26-19-24[默认环境] 查看
523 venus-report 21[活跃版本] 2016-09-26-19-28[默认环境] 查看
17 unionadmin 58[活跃版本] 2017-02-27-17-51[默认环境] 查看
804 iflow-api 47[活跃版本] 2017-02-07-19-31[默认环境] 查看
16 union 59[活跃版本] 2017-02-27-17-58[默认环境] 查看
940 bizgray 65[活跃版本] 2017-03-08-010-07[默认环境] 查看
291 feige 62[活跃版本] 2017-02-28-18-010[默认环境] 查看
807 iflow-web 70[活跃版本] 2017-03-14-18-58[默认环境] 查看
containerpool


统一服务管理平台
服务实例管理 服务基本信息 权限管理 操作日志
服务信息 主机信息
批量修改 批量修改状态
操作 服务ID 服务名 全路径 节点数 服务状态 运行模式
containerpool
resourcepool


Sogou搜狗 统一部署服务
服务发布 服务管理 权限管理 高级功能 消息提醒 审计日志
环境搭建 模板管理 服务操作 服务授权 Nginx操作 Shell工单
商业平台
线上应用
授权 列表
选择服务
环境资源：
null(null)
访问端口：
访问者IP：
访问者IP，英文逗号隔开
开通权限



CONTENT

总结

- 分享了：
  - 研发商业云平台的背景和目标
  - 我们的技术选型
  - 我们的实践之路：基础集群、服务管理、服务授权、灰度发布

总结

- 效果：
  - 降低了整体成本，缩短项目周期
  - 资源率利用率有效提升
  - 弹性计算能力有效提升
- 经验
  - 需求驱动：前期需求调研的重要性
  - 快速落地：不追求最完美的设计，追求尽快落地，逐步完善
  - 平滑迁移：兼容已有流程、用户习惯
  - 持续优化：优化无止境

展望

- 基础资源容器化
- 联邦集群
- 更智能的调度

Thanks

简历砸来：
bizjob@sogou-inc.com