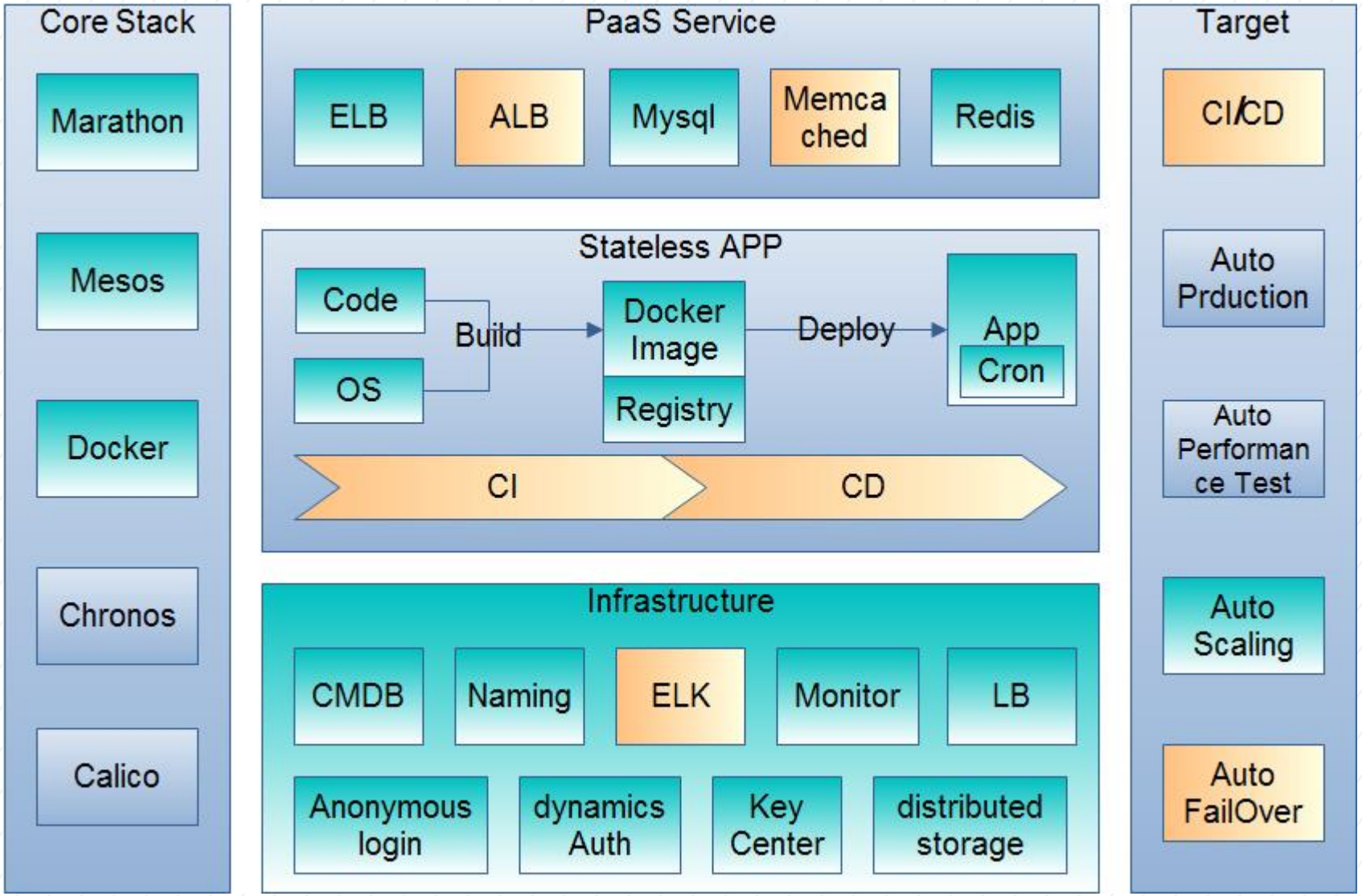
小米弹性调度平台
Ocean

孙寅，现在小米负责运维基础设施、基础平台的构建

1 私有云体系概览

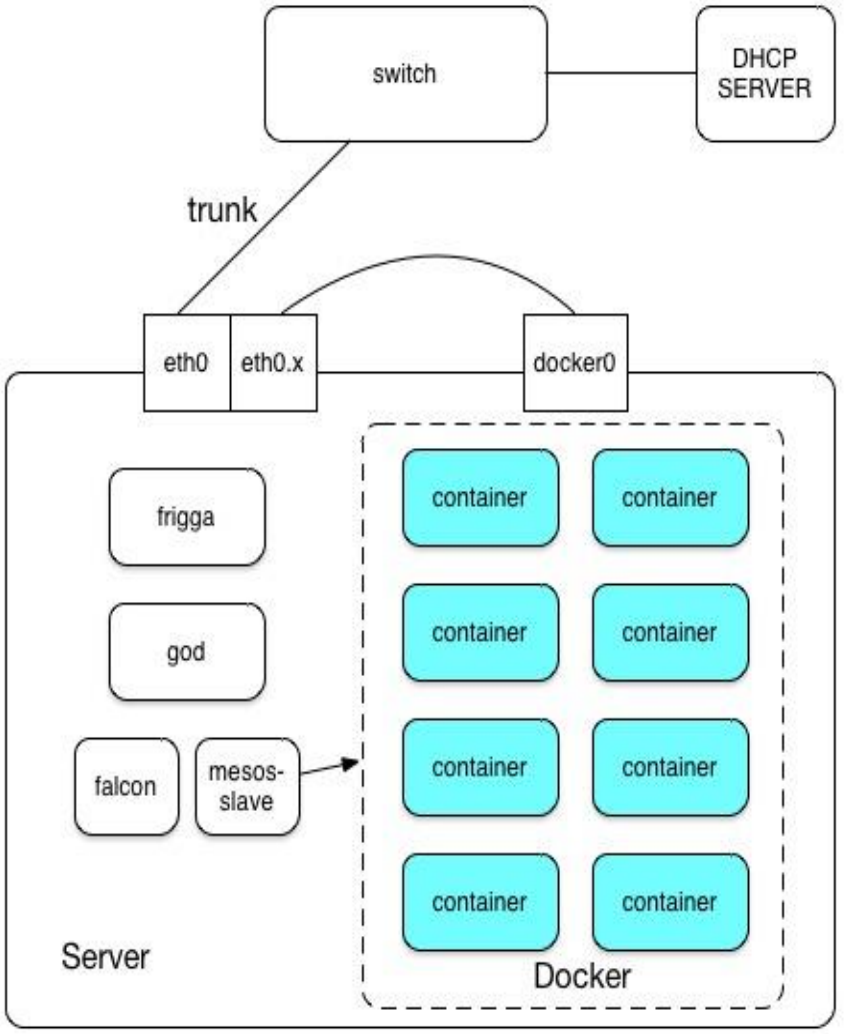

Core Stack
Marathon
Mesos
Docker
Chronos
Calico
PaaS Service
ELB
ALB
Mysql
Memca ched
Redis
Stateless APP
Code
OS
Build
Docker Image
Registry
Deploy
App
Cron
CI
CD
Infrastructure
CMDB
Naming
ELK
Monitor
LB
Anonymous login
dynamics Auth
Key Center
distributed storage
Target
CI/CD
Auto Prduction
Auto Performan ce Test
Auto Scaling
Auto FailOver

Docker

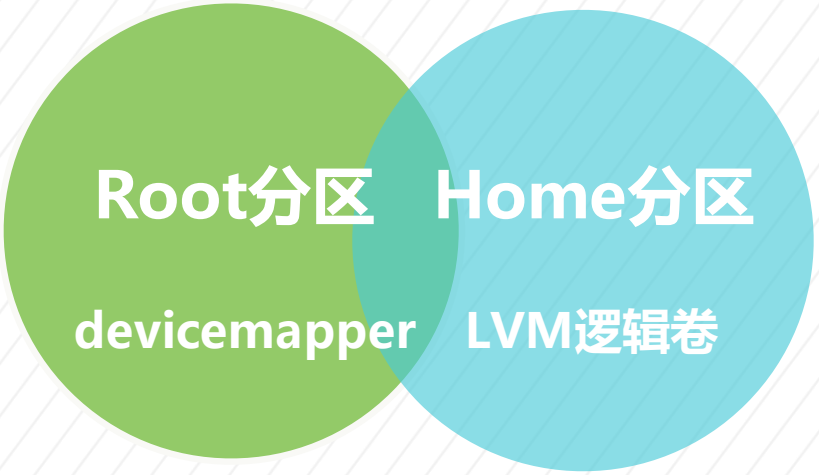
在工业环境落地

1 网络

- 真实内网IP，便于标识和定位
- 与原有物理网络天然互通
- 吞吐、延迟与纯物理网络几无差别
- 大三层网络，无广播风暴风险



2 文件系统



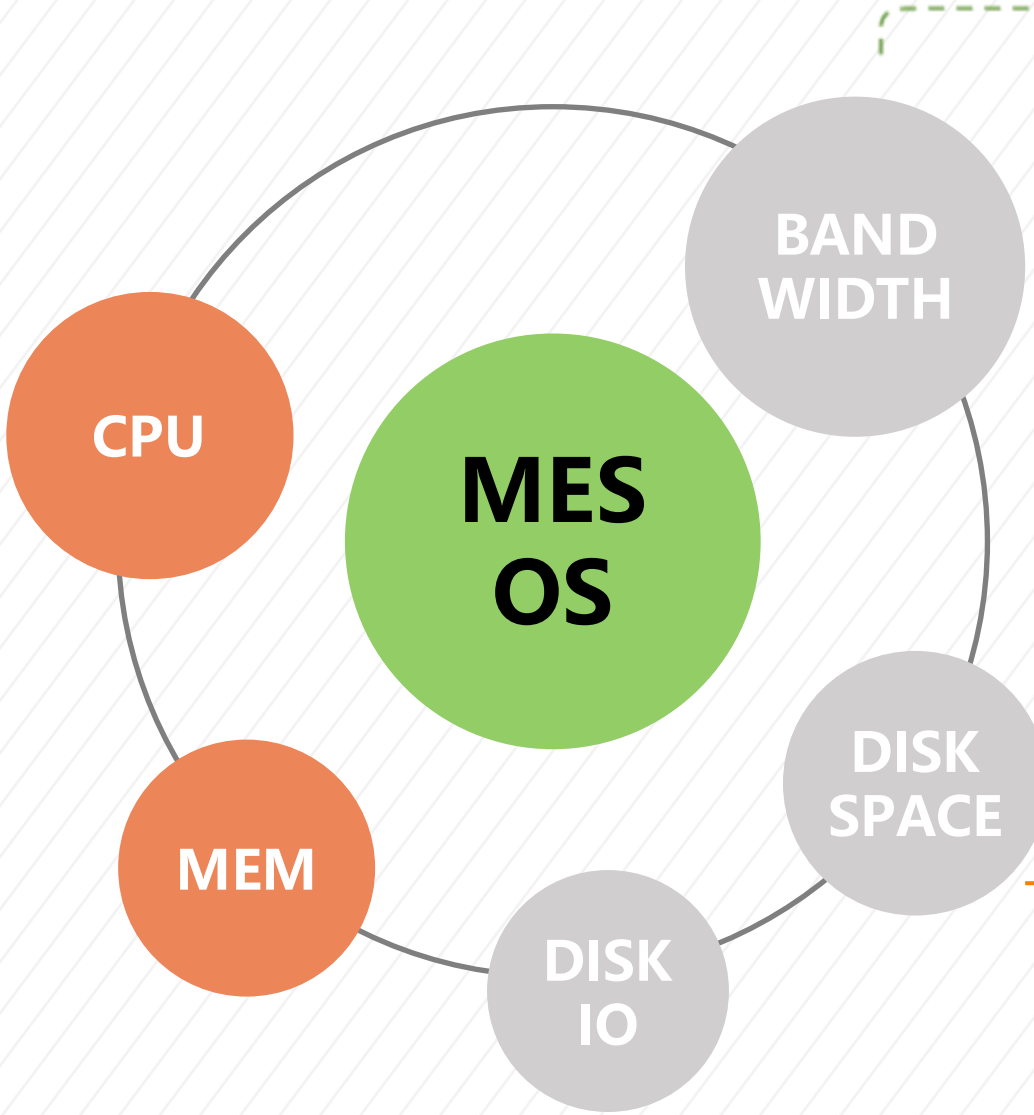
目标

- 容器磁盘空间大小可分配
- 保证磁盘IO性能不下降

技术考量

- aufs、btrfs等等存储方式都会损耗IO性能
- 一个宿主机DM只能设置固定大小
- Home分区引入LVM动态设置空间大小

3 资源隔离资源汇报



- 扩展Mesos Resource
- 环境变量传递资源大小
- Docker中嵌入TC限速模拟网络带宽隔离
- DM+LVM隔离磁盘空间，mesos-slave hook创建和清理LVM

4 build

require:
centos:6.3+resin:4.0.41+scribed:2.0.65
service:
- port: 8094
add:
- weather-api-v2.xml /home/work/resin/conf/
- scribed.conf /home/work/scribe/conf/

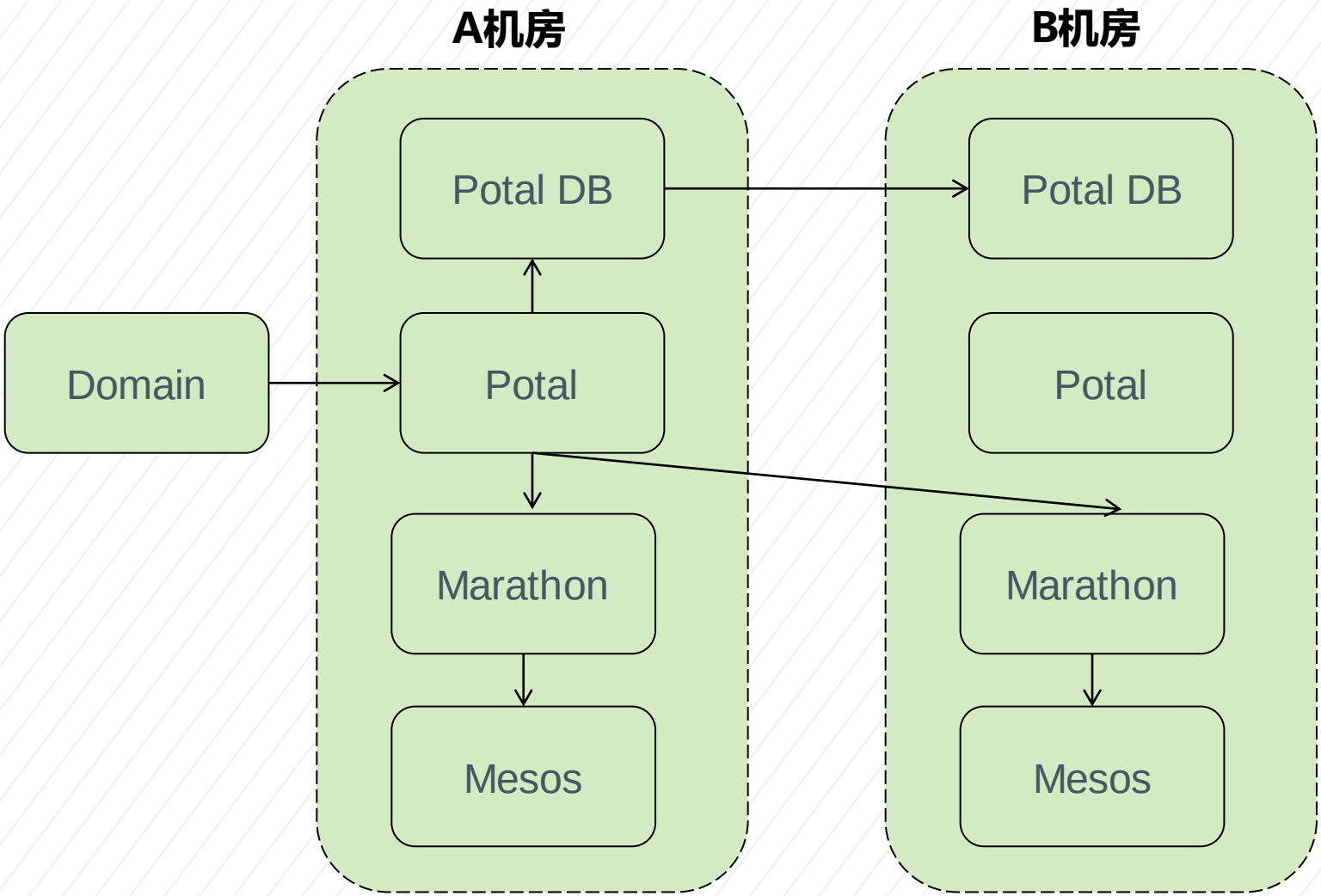
entry:
- /home/work/resin/bin/resin_control -conf /home/work/resin/conf/weather-api-v2.xml -server weather-api-v2 start

4 Docker-Init

- DHCP
- 内置服务管理
- 进程组
- 进程HealthCheck
- 进程启动退出钩子
- 回收僵尸进程
- DEBUG模式
- 与内部基础设施对接
  - Marathon HealthCheck
  - Falcon
  - DoorGod（安全免密登陆）
  - Mysql Auth

多容器集群管理

4 多集群管理


A机房
B机房
Potal DB
Potal DB
Domain
Potal
Potal
Marathon
Marathon
Mesos
Mesos

服务发现

2 服务发现

RPC框架

Zookeeper

Nginx Module

Mesos-DNS

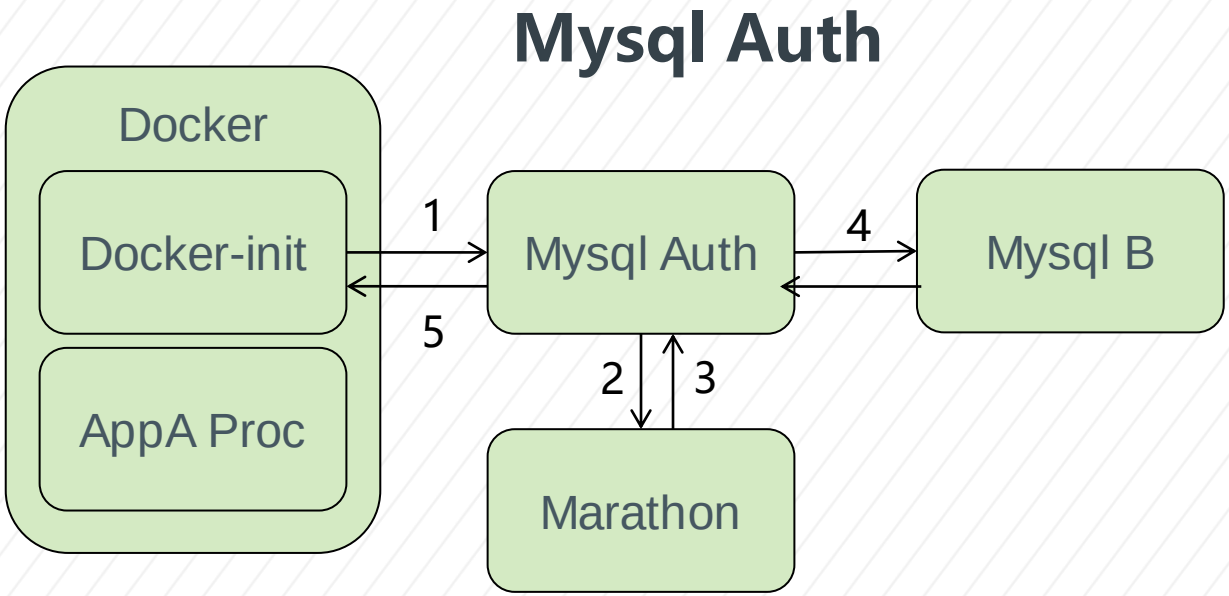
Mysql Proxy

Redis Cluster(Gossip)

Other

动态安全

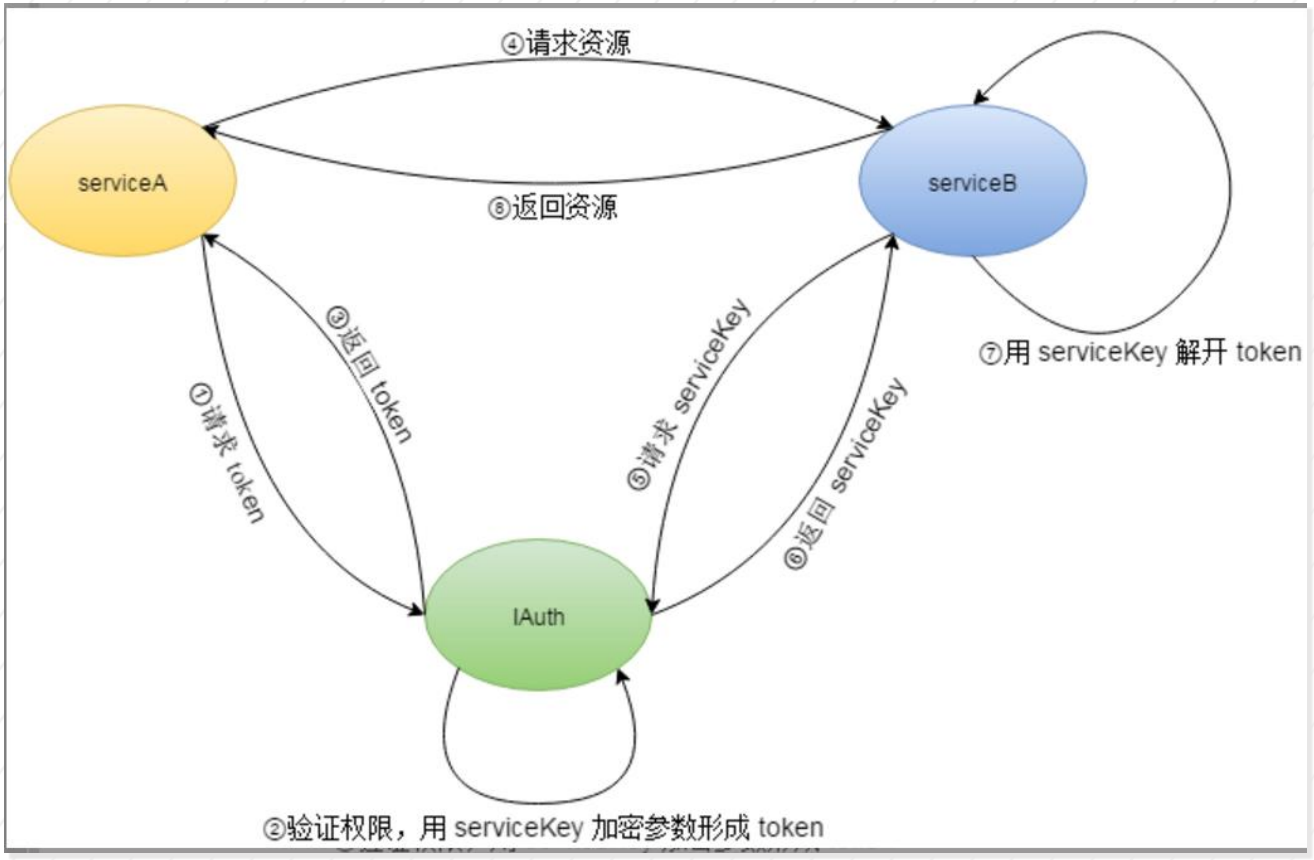
3 动态安全



1. 声明属于AppA，申请对MysqlB的认证授权
2. 询问第三方Marathon，源IP是否属于AppA
3. 返回AppA的instance信息
4. 如果认证成功，且Mysql Auth中有MysqlB对AppA的授权，则去MysqlB去做实际授权操作
5. 同步返回Docker-init，任何失败都会让Docker-init退出，即容器退出

3 动态安全

IAuth

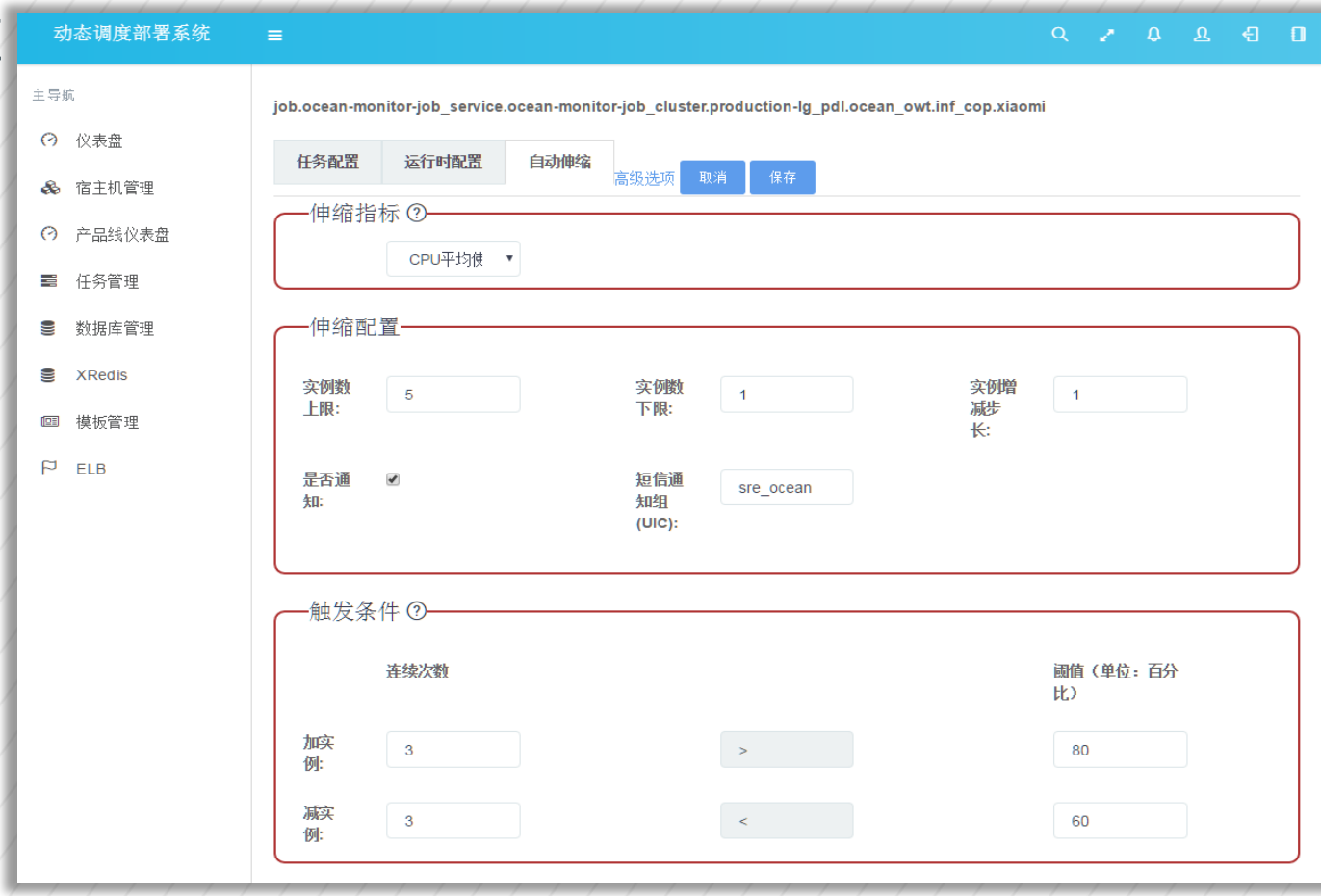


1. serviceA和B均需要植入SDK
2. 可以通过scope对函数集粒度进行授权
3. 动态换token

自动伸缩

1 自动伸缩

- Falcon自动采集并监控容器的
  - CPU IDLE
  - MEM FREE
  - PROC QPS
  - PROC DELAY
- Hook回调Marathon API
- 最短10s触发伸缩



2 定时触发伸缩

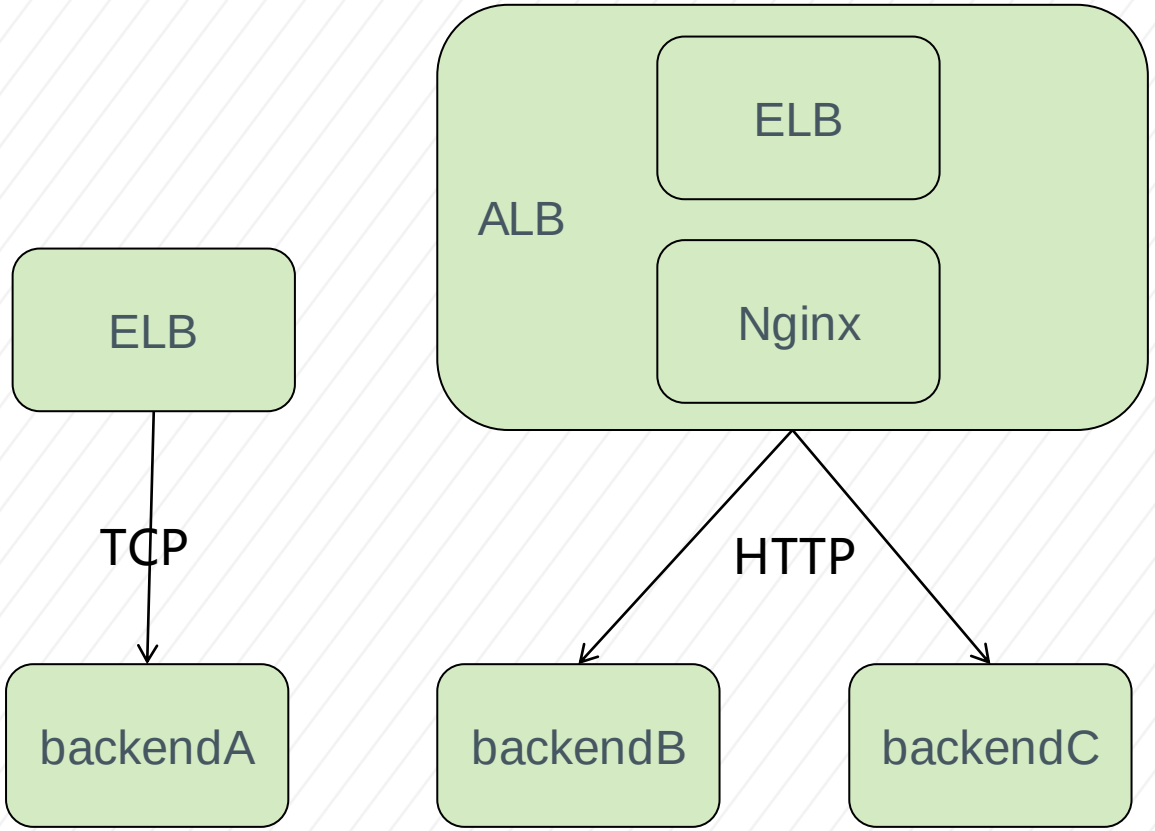
- 利用Chronos调用Marathon API

服务化组件

3 ELB

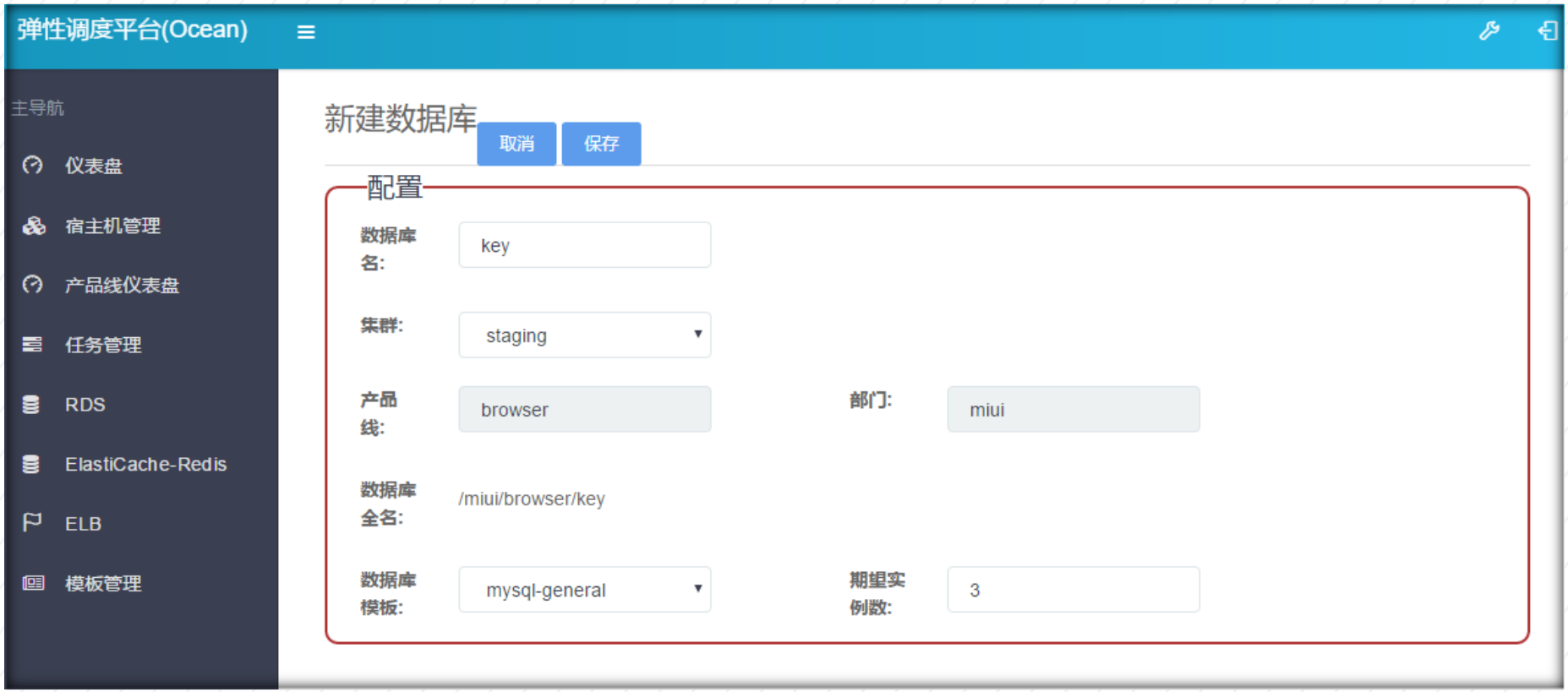
- 自动配置内网域名，按运营商自动划分线路
- docker-init和旁路模块都会动态更新LVS配置，可重入

4 七层ALB

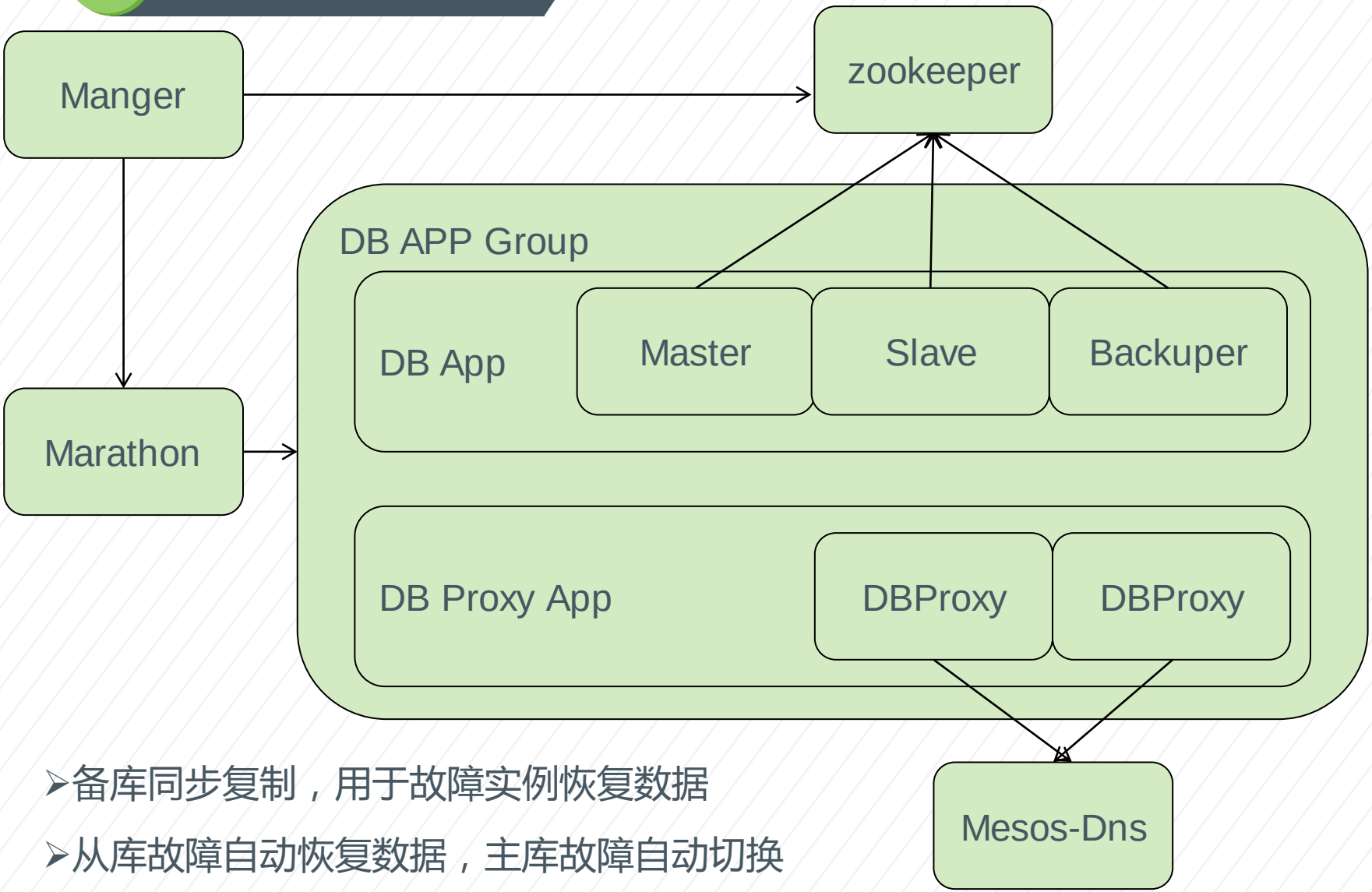


- 自动配置
- 上传SSL证书
- 动态更新upstream
- 根据qps自动伸缩

5 RDS

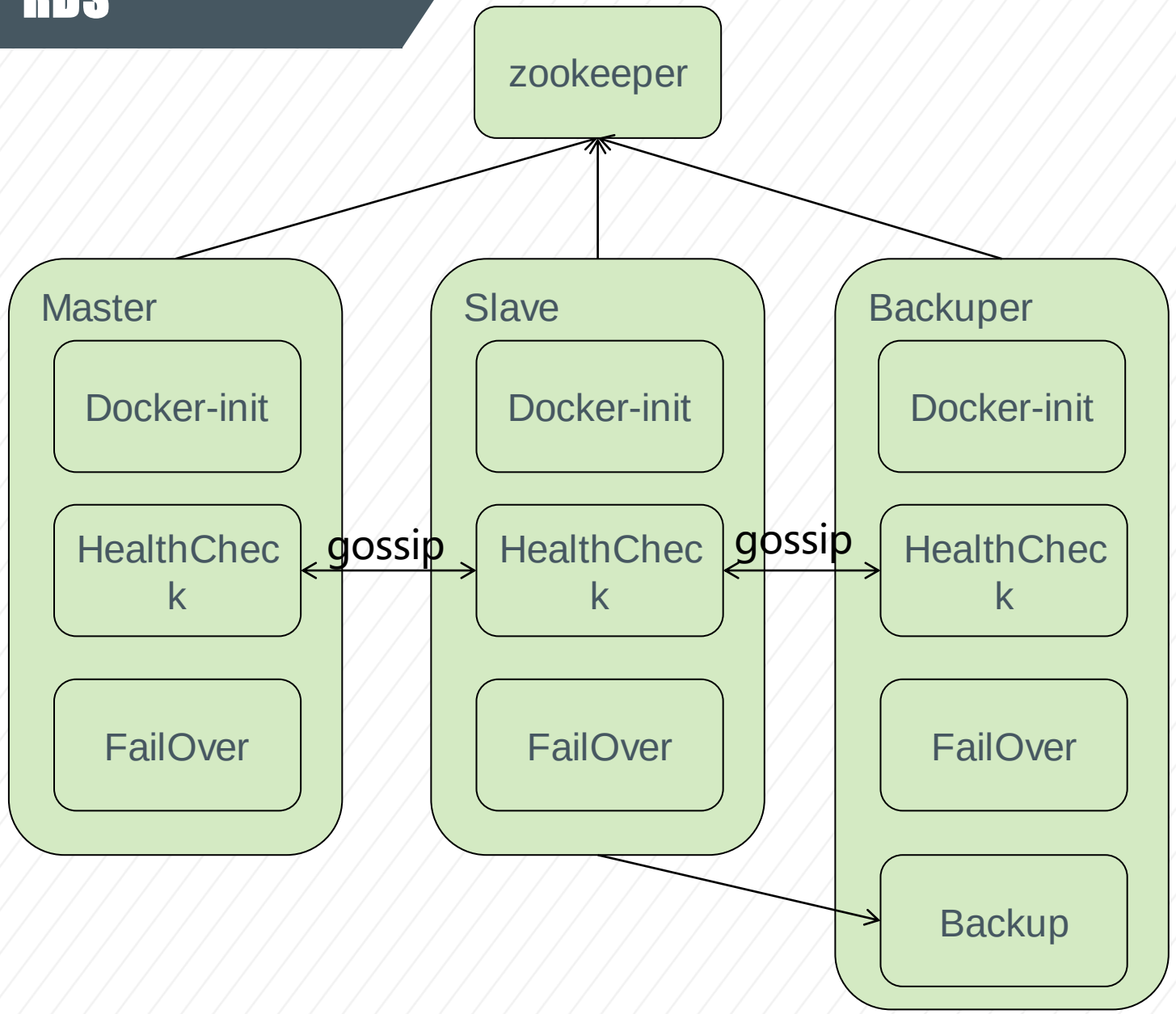

弹性调度平台(Ocean)
主导航
仪表盘
宿主机管理
产品线仪表盘
任务管理
RDS
ElastiCache-Redis
ELB
模板管理
新建数据库
取消
保存
配置
数据库名:
key
集群:
staging
产品线:
browser
部门:
miui
数据库全名:
/miui/browser/key
数据库模板:
mysql-general
期望实例数:
3

5 RDS



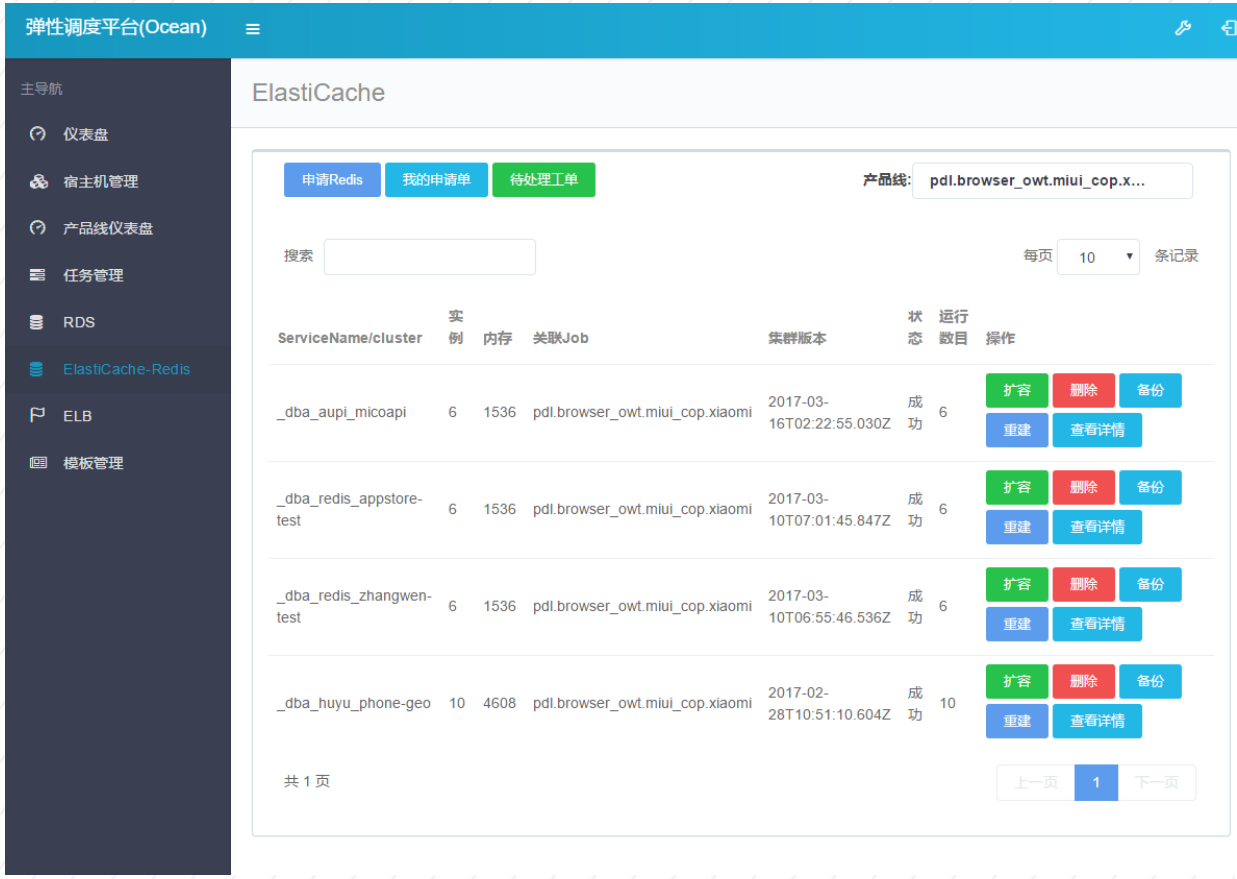
- 备库同步复制，用于故障实例恢复数据
- 从库故障自动恢复数据，主库故障自动切换

5 RDS


zookeeper
Master
Docker-init
HealthChec
k
FailOver
gossip
Slave
Docker-init
HealthChec
k
FailOver
gossip
Backuper
Docker-init
HealthChec
k
FailOver
Backup

6 ElastiCache

- Redis和Memcached
- Redis cluster
- 集成对应监控插件



CI/CD

CI/CD

Pipeline

Step	Condition	Action	Object
Step1	Branch staging change	Action1: build & ut test	Staging job
		Action2: deploy	Staging job
		Action3: bvt test	Staging job
		Action4: merge	Preview branch
Step2	Branch preview change	Action1: deploy	Preview job
		Action2: merge	Production branch
Step3	Branch production change	Action1: deploy	Production C1 job
		Action2: bvt test	Production C1 job
		Action3: deploy	Production C2 job
		Action4: bvt test	Production C2 job

未来规划

未来规划

补充完善有状态组件——zookeeper、Kafka、Hbase等

App间的依赖链管理和运用

强化故障自愈和自动容灾能力

网络再设计——SDN

FAQ

WECHAT