

# 以Mesos/Docker为基础的 Elasticsearch容器化私有云

## 自我介绍



- 姓名： 马文
- 部门： Qunar平台事业部
- 简要介绍： 主要从事Mesos/Docker容器技术的研究，ES搜索技术的研究，ES容器私有云的建设和运维工作

# 目录

- 1 ▶ 背景与现状
- 2 ▶ 技术实现
- 3 ▶ 配置与部署
- 4 ▶ 监控与报警

# 一、背景与现状

# 背景

传统的Elasticsearch交付与维护弊端:

- 传统部署以kvm虚拟机或实体机为节点,需提前申请,过程较慢
- ES运行环境需要人工事先部署,自动化程度不高
- 集群扩容需要人工一遍一遍的部署环境,效率低
- 集群信息不集中,不易于管理,维护成本高
- ES需求量增加之后,更加加大了维护的成本和繁杂性

# 设计目标

针对上述弊端，我们制定的几点设计目标：

- 加快集群构建速度
- 快速扩容和快速迁移能力
- **ES**使用/运维标准化
- 集中的信息展示与良好的用户交互界面

# 设计目标

## 容器化的私有云平台 Elasticsearch as a service (ESAAS)



# 改进

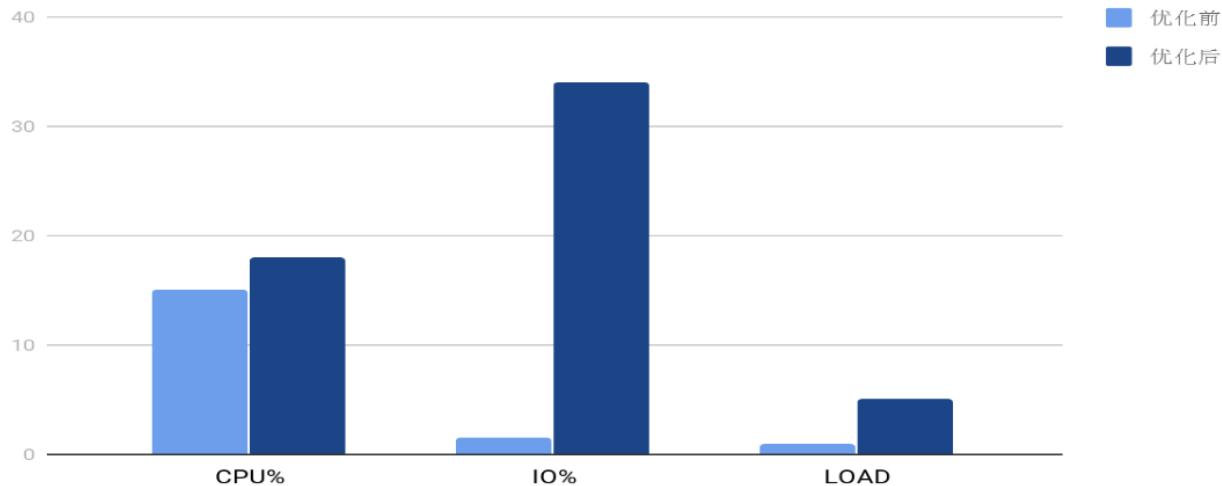
- 成本：
  - 人力成本：大大节省，用户只管使用，我们平台方统一维护
  - 资源成本：均摊下来一个ES集群少于2台机器，且目前仍有大量空闲资源
- 效率：
  - 集群构建：从初始化到任务调度到集群work最快可10min内
  - 集群扩容：扩容节点秒级以内
- 规范：
  - ES使用标准化：appcode表示方式，portal查看申请统一入口
  - 运维标准化：统一集群管理，统一配置管理，配套监控报警



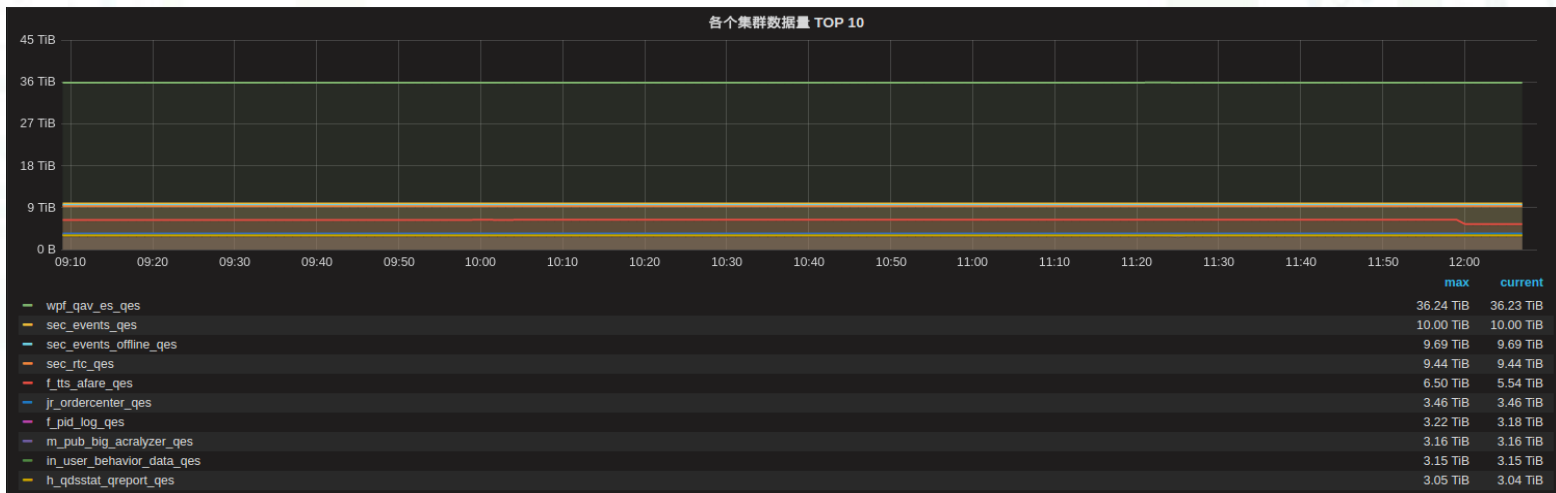
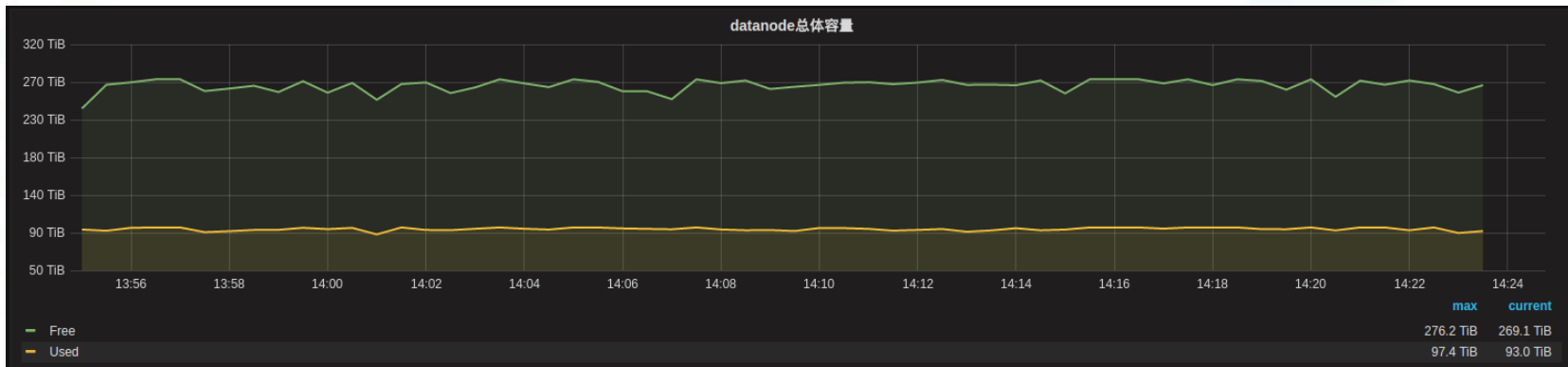
# 改进

## 更高的利用率

平均利用率



# 平台规模

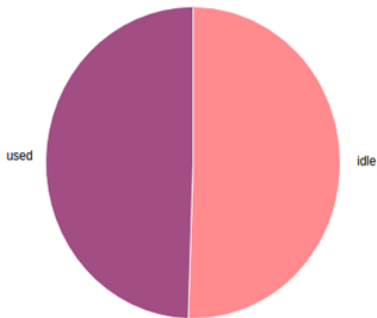


# 资源使用

资源使用未过半，均摊下来一个ES集群资源使用不到2台物理机

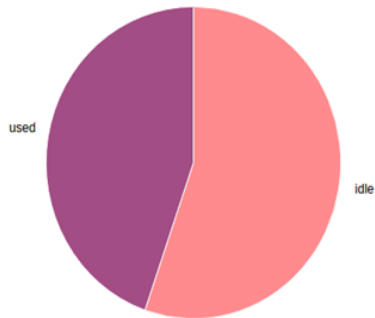
ESAAS 总共的CPU使用占比pie

● idle ● used



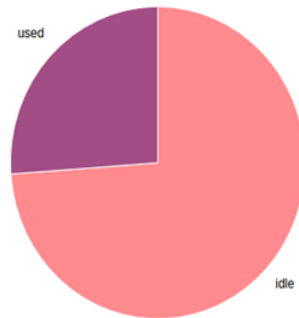
ESAAS 总共的mem使用占比pie

● idle ● used



ESAAS 总共的 disk 使用占比pie

● idle ● used



## 二、技术实现

# 调研选型

- Elastic Cloud
- Amazon Elasticsearch Service
- Elasticsearch Framework on Mesos

# 调研选型



docker

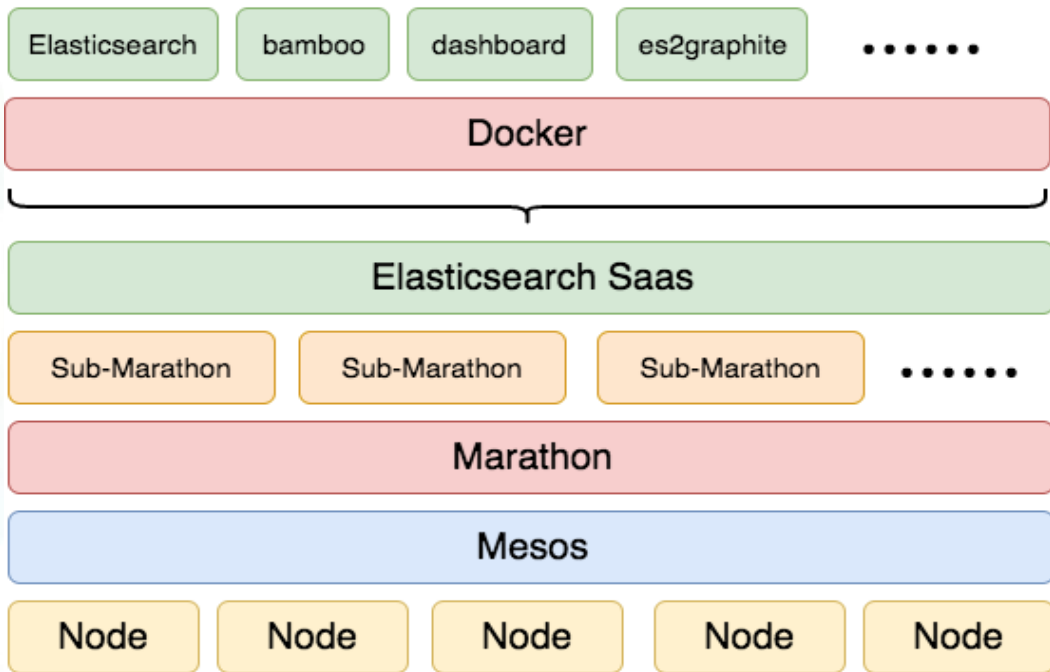


MARATHON

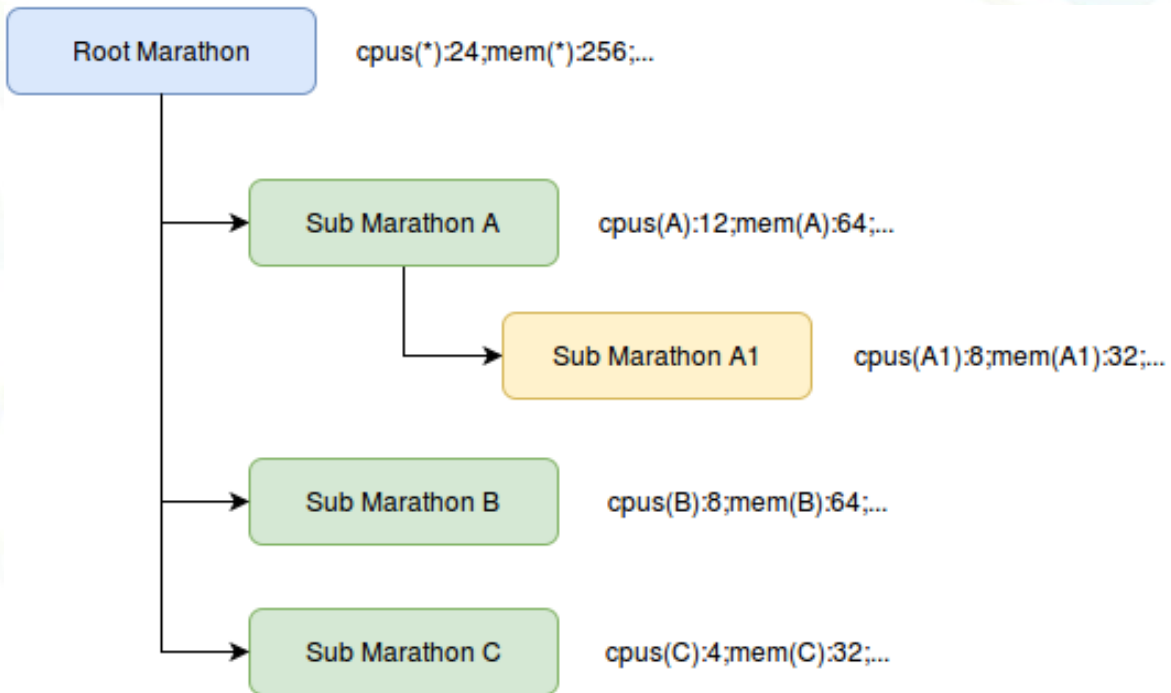


MESOS

# 总体结构

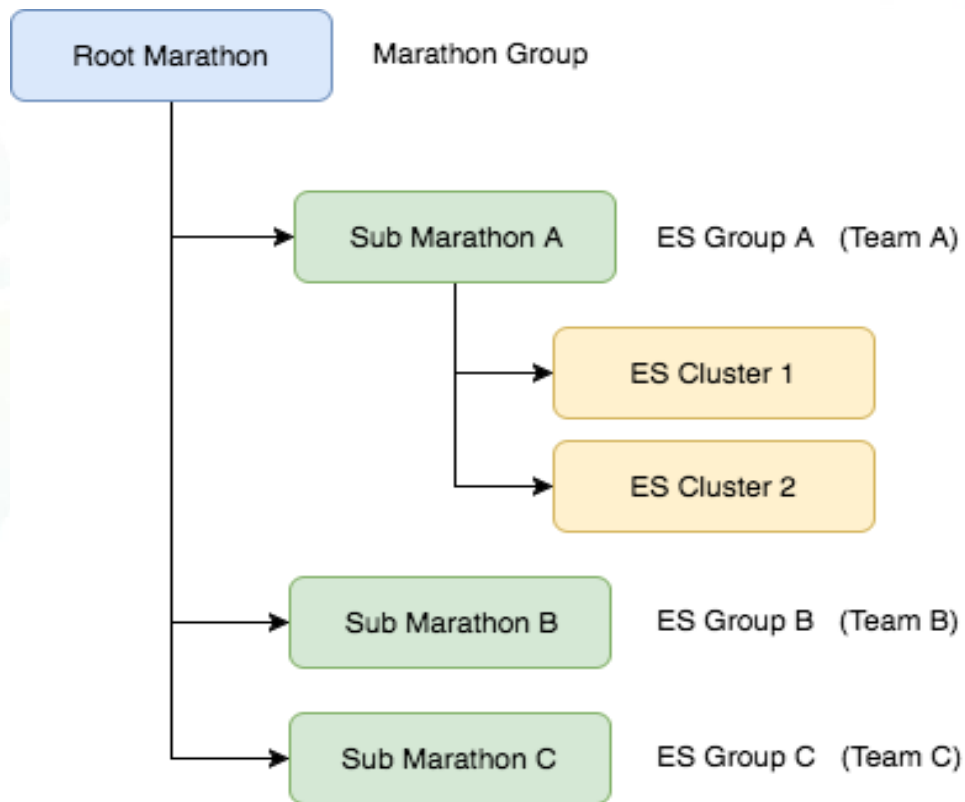


# 资源分配结构

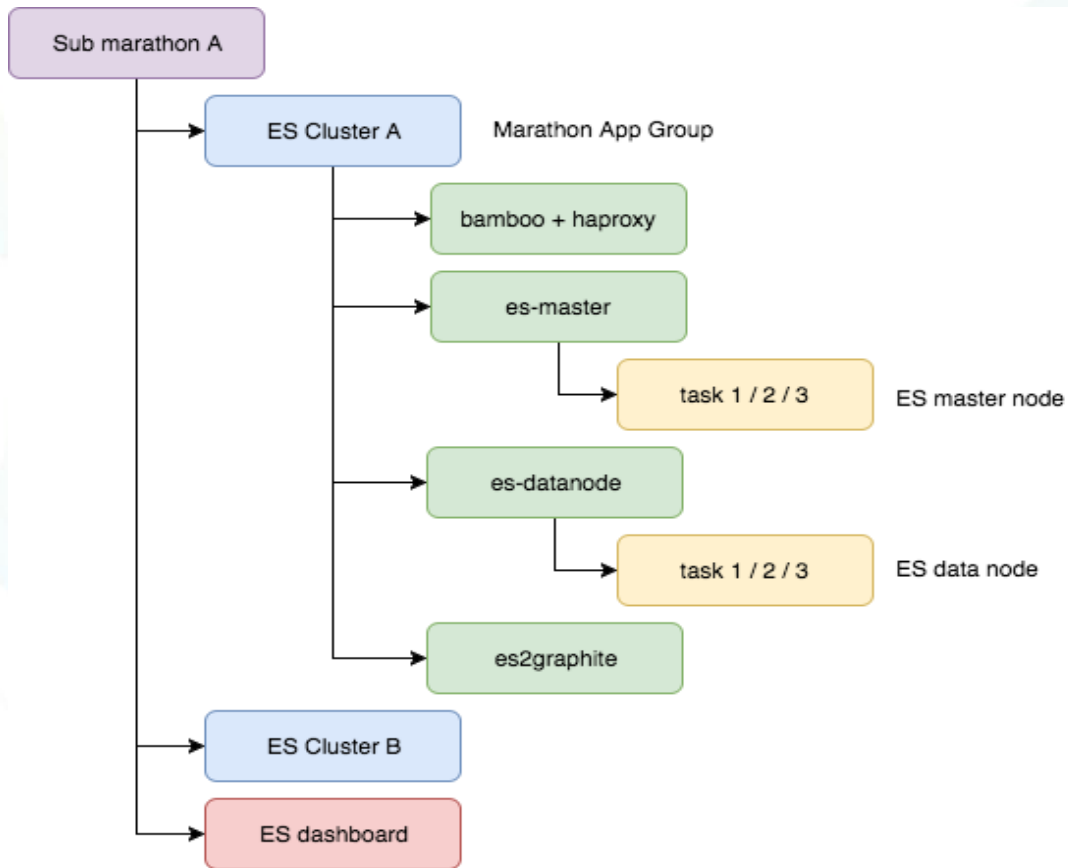




# 集群逻辑隔离

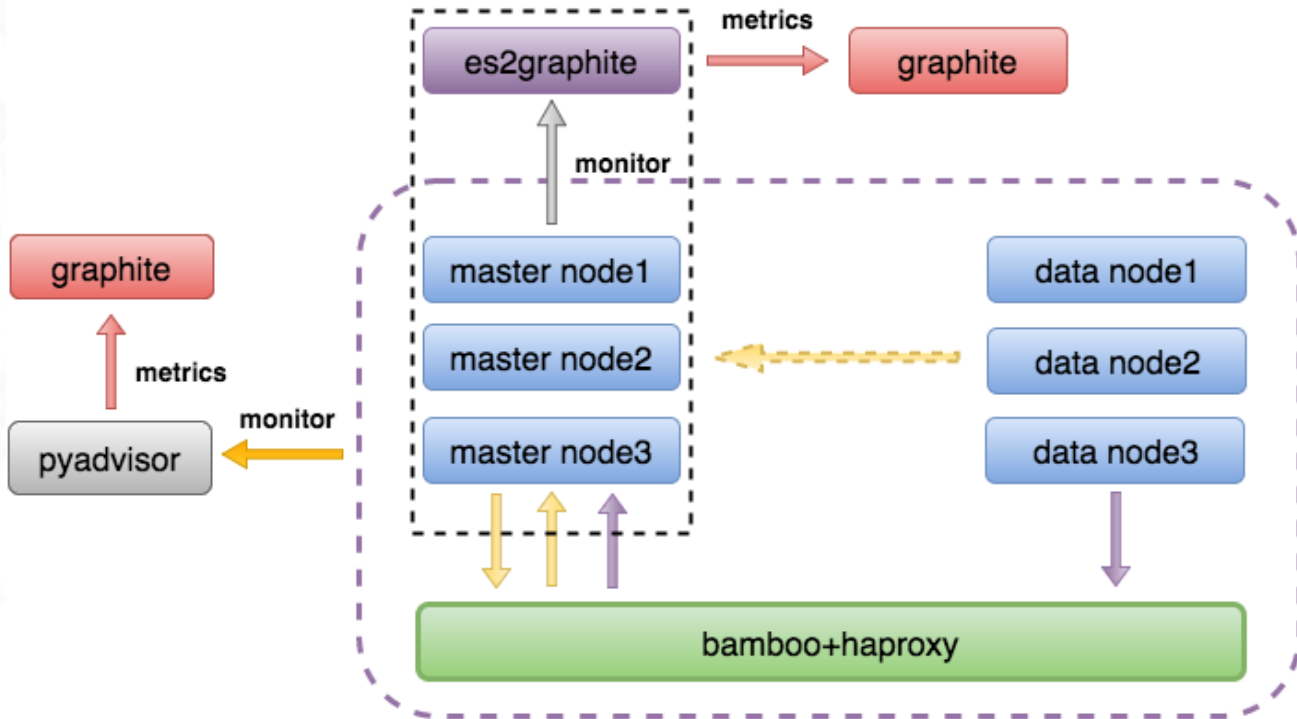


# 集群逻辑隔离

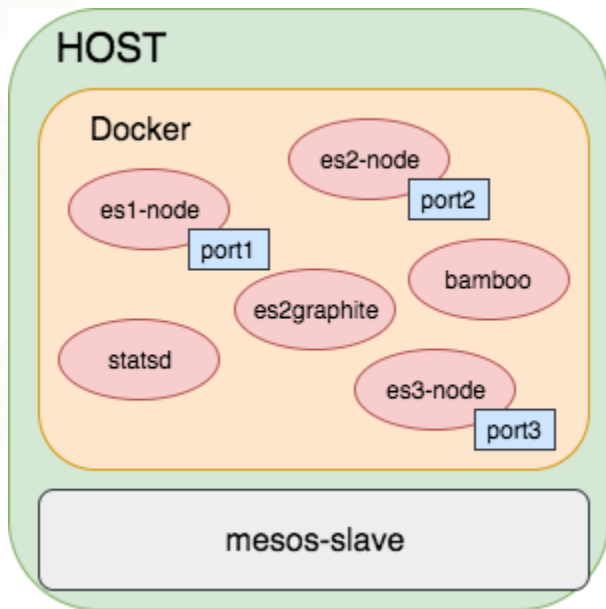


# 一个完整ES集群的逻辑结构

Sub-Marathon



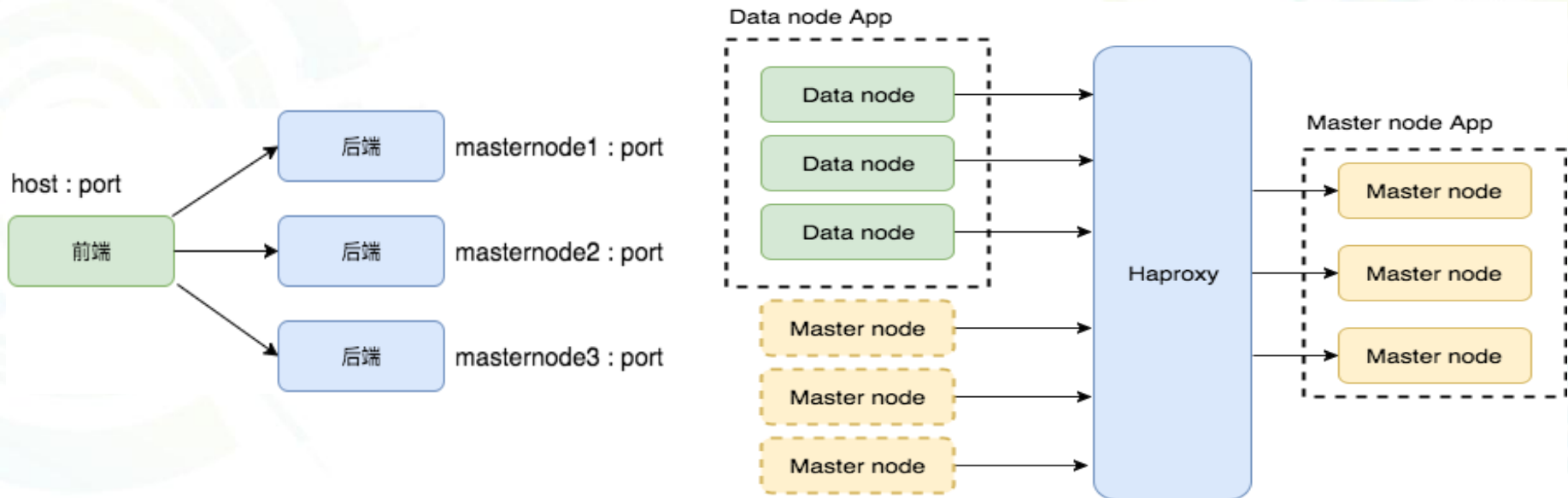
# 单台物理机结构



- 一个机器上可有多个ES 节点
- 使用不同的端口来区别

# 服务发现

discovery.zen.ping.unicast.hosts: ["<haproxy\_ip>:<haproxy:port>"]



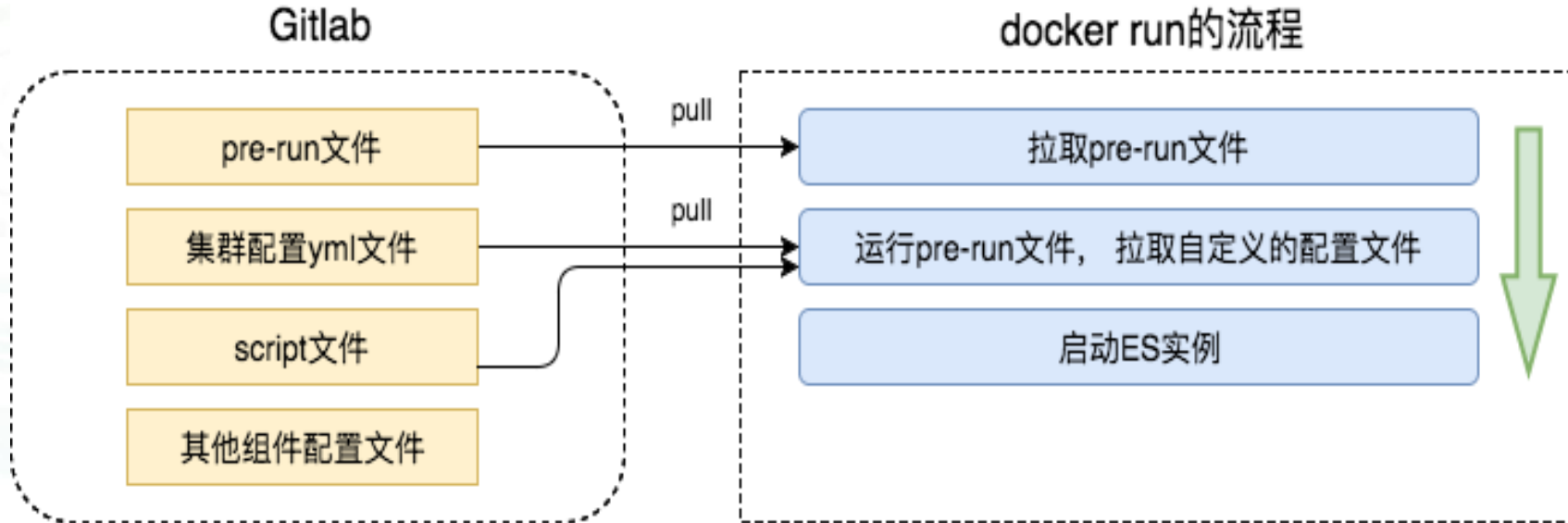
## 可靠性 & 持久化

- Marathon 持久化卷功能, 保证ES节点重启之后能在原机器节点之上 restart
- `index.number_of_replicas >= 1`
- 默认配备hdfs插件
- replica 数量大于ES实例数 (防止主备分片被分配在同一台机器上面)


### 三、配置和部署

# 自动化配置

- ES配置文件存在gitlab上，自助修改
- 添加pre-run文件，在ES实例启动之前做自定义的事





wen.ma

ops\_gss\_test\_qes >

ops\_test\_v5\_qes >

ops\_event\_log\_qes >

- 集群概况
- 集群配置
- 索引清理
- 操作日志
- ES集群日志

ops\_watcher\_events\_qes >

状态: Health

### 概况

集群名称	ops_event_log_qes
集群版本	2.3.3
http端口	1: L
transport端口	1: ( )
node数量	4
shards总量	176
kopf访问地址	http://l-qesaa:cn2.qunar.com:12101
marathon地址	http://marathon:cn2.qunar.com
配置地址	http://gitlab:cn2.qunar.com/nfcarbon/ops_event_log_qes
watcher监控	http://watcher:cn2.qunar.com/team/ops_event_log_qes

上月已出账单:

CPU	126.37 ¥
Memory	606.58 ¥
Disk	0.01 ¥
<b>合计:</b>	<b>732.96 ¥</b>

本月未出账单:

CPU	86.93 ¥
Memory	417.29 ¥
Disk	0.01 ¥
<b>合计:</b>	<b>504.23 ¥</b>

### masternode

l-qesaa:cn2.qunar.com

l-qesaa:cn2.qunar.com

更多 >

### datanode

l-qesaa:cn2.qunar.com

l-qesaa:cn2.qunar.com

更多 >

### Plugins

kopf

head

# 自助配置



ops\_qss\_test\_qes >

ops\_test\_v5\_qes >

ops\_event\_log\_qes ▾

≡ 集群概况

⚙️ 集群配置

🗑️ 索引清理

📅 操作日志

📄 ES集群日志

ops\_watcher\_events\_qes >

状态: Health

配置 [打开](#) [关闭](#)

## 插件安装

cerebro

安装

## Kibana

4.5.4

安装

## Replica

该设置为全局设置，集群所有的index都会被修改，谨慎操作！线上备份数线上设置应该大于等于1(replicas>=1)

数量

1

配置

## cache清理

/\_cache/clear

清理

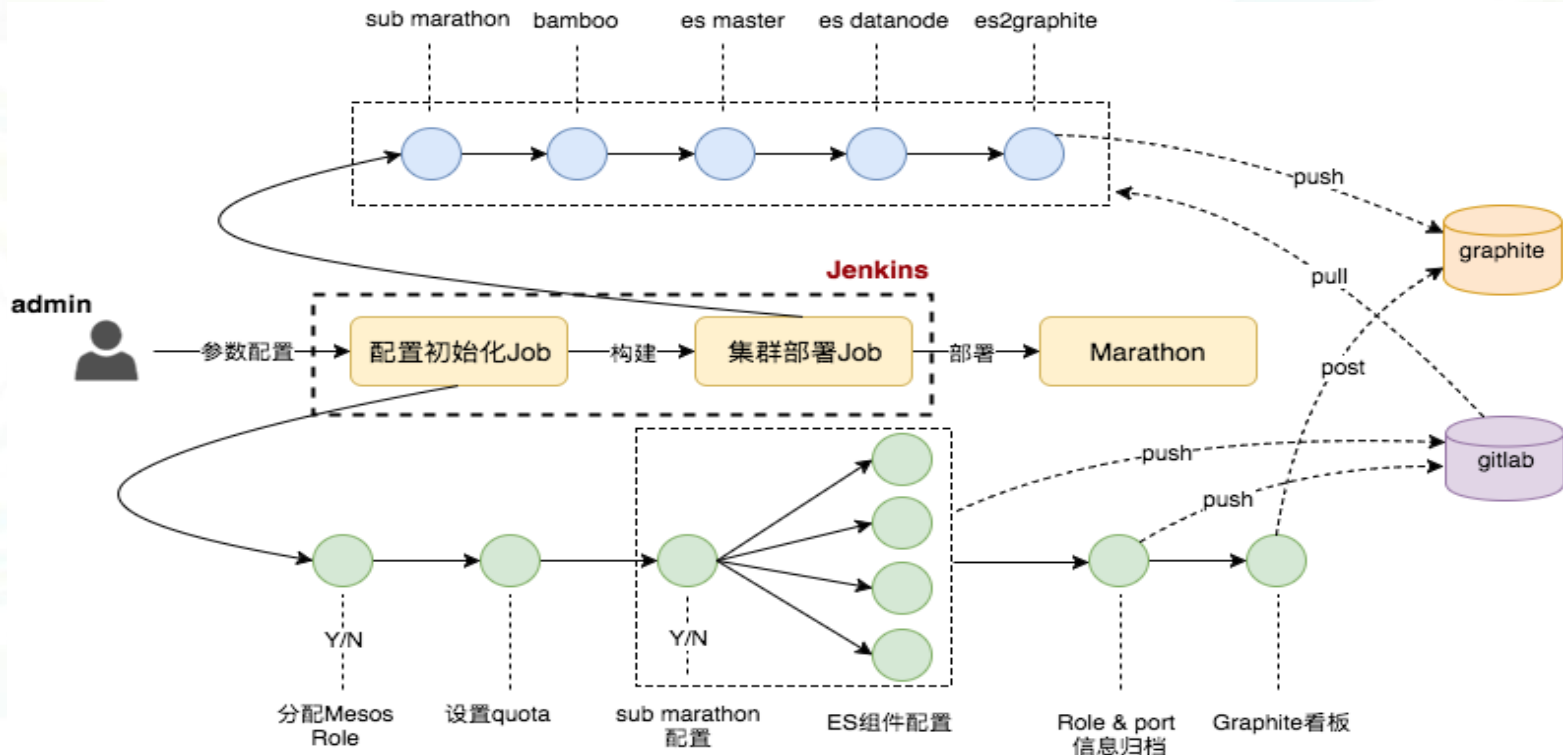
## 节点重启

l-qesaas ■ ■ ■ ■ ■ ■ ■ ■ ■ ■

重启

## Log级别设置

# 自动化部署

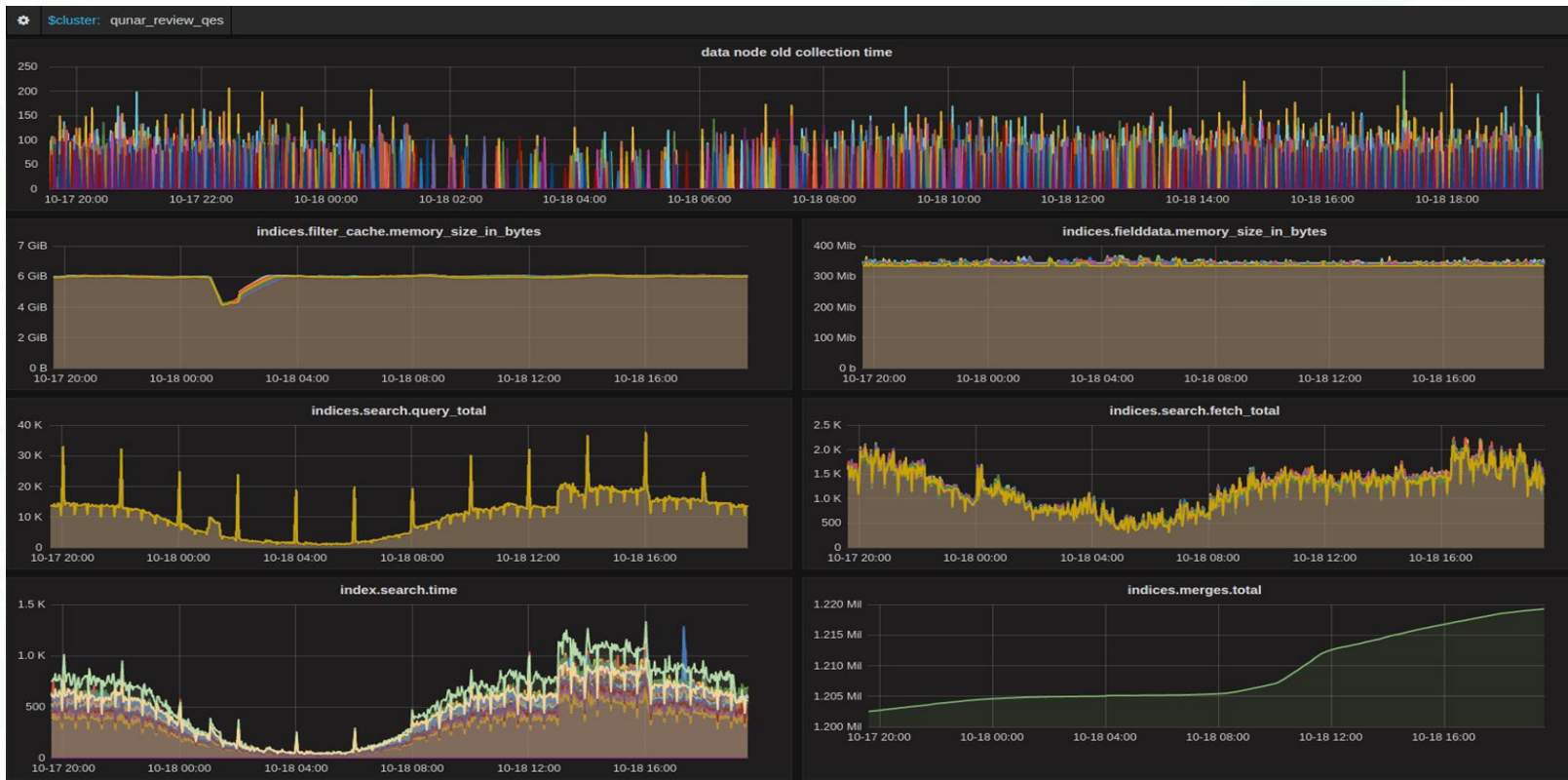


## 四、监控与报警

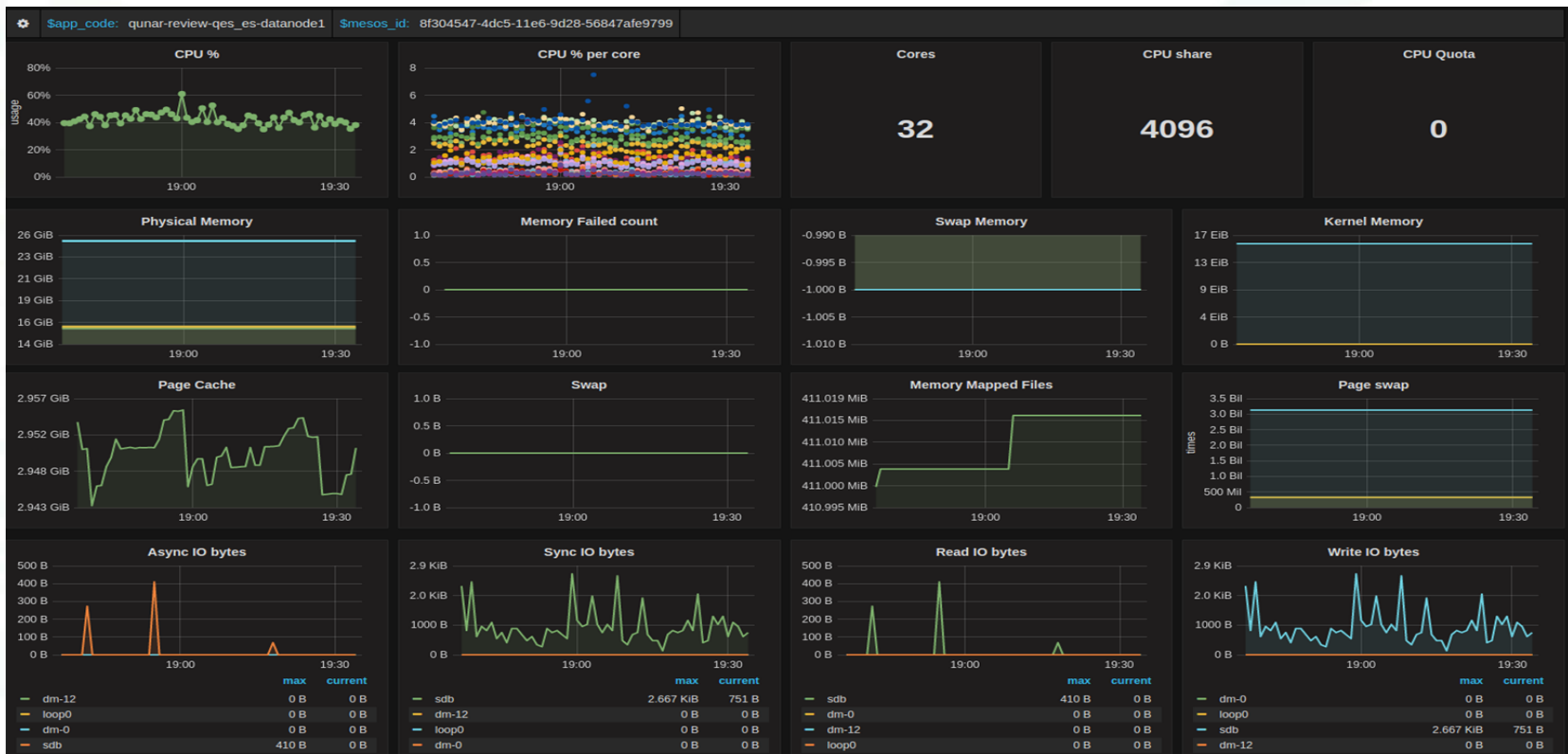
# 指标收集

- pyadvisor 收集docker 容器的指标
- es2graphite 收集ES集群的指标

# 聚合汇总



# 聚合汇总



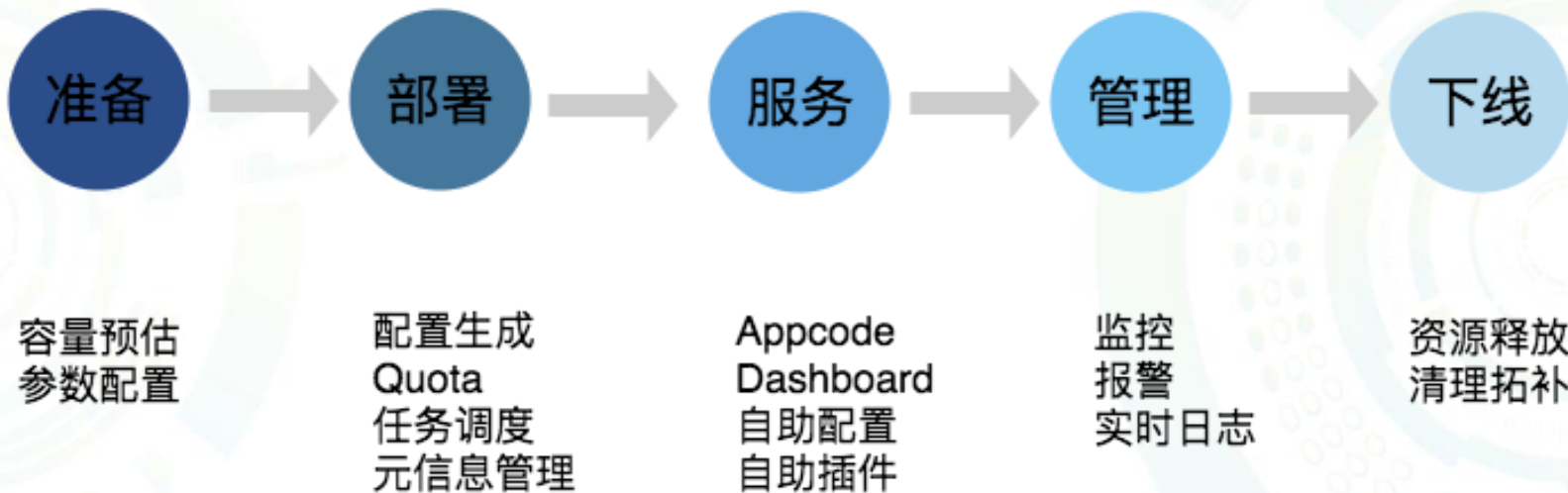
# 报警

- ES集群
  - ES集群状态（非green状态）
  - ES节点GC时间
  - 节点数量
- ESAAS平台
  - ZK相关
  - 节点load
  - 节点磁盘容量





# 总结



# Q&A