GOPS2016
Beijing

全球运维大会

2016

DevOps 2.0：重塑运维价值

北京站

会议时间：12月16日 - 12月17日

会议地点：北京国际会议中心

主办单位：高效运维社区 GreatOPS Community　开放运维联盟 OOPSA Open OPS Alliance

Qunar实时流系统实践

吕晓旭 去哪儿网 实时系统负责人

目录

1 我们的实时数据平台-Prism

2 从这里开始

3 架构演进

4 Esaas - Elasticsearch as a Service

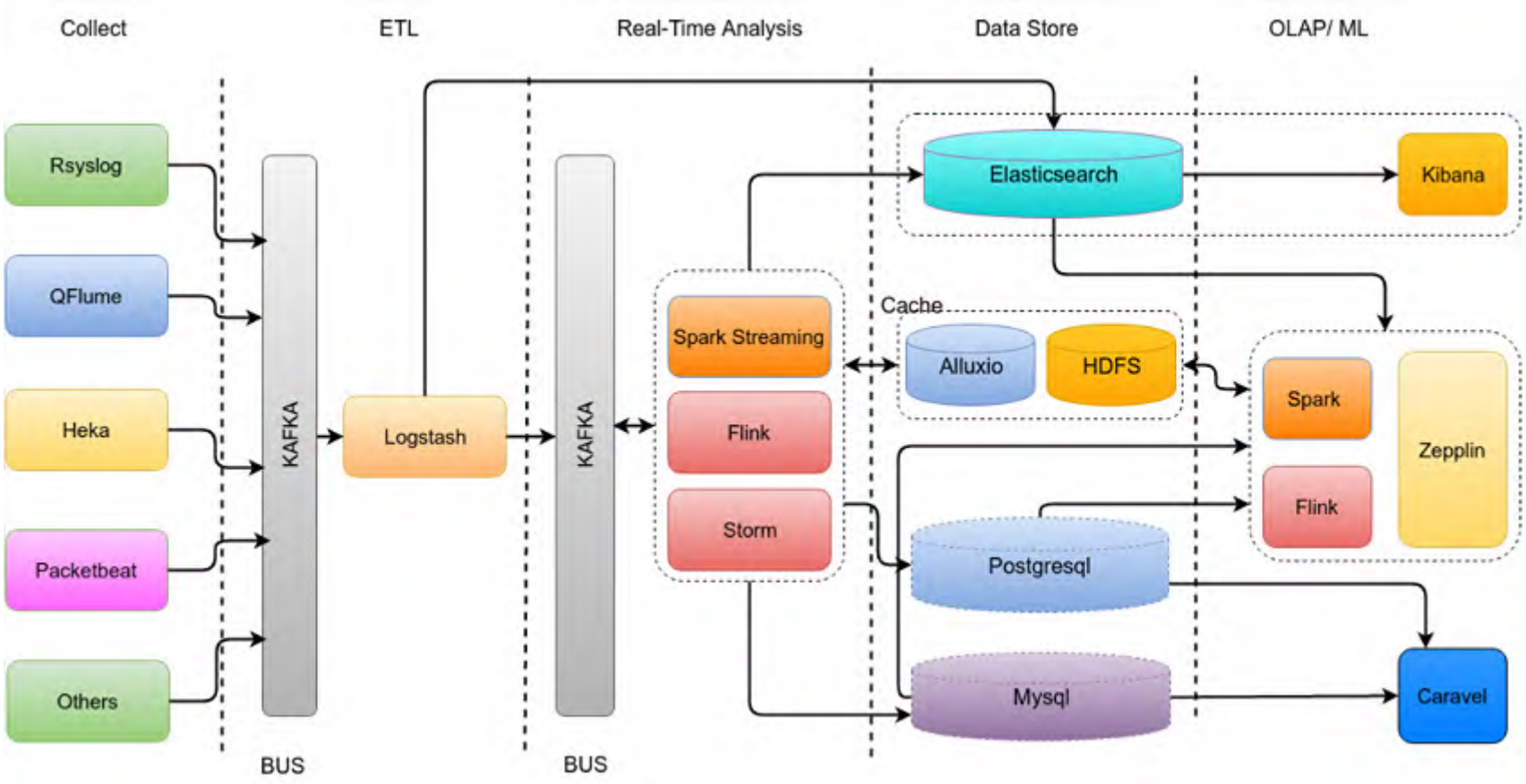
5 监控

6 规模

Prism是什么

1. 宗旨
    1. 以数据可视化为出发点
    2. 以降低数据和数据分析软件获取成本为己任
    3. 的实时数据平台
2. 提供哪些服务
    1. 日志实时监控 - ELK
    2. 数据总线 - Kafka
    3. 数据实时分析 - Spark Streaming/Storm/Flink
    4. 数据存储 - Elasticsearch as a Service
    5. OLAP/试验平台 - Zeppelin+Spark/Flink

Prism数据流图


Collect
ETL
Real-Time Analysis
Data Store
OLAP/ ML
Rsyslog
QFlume
Heka
Packetbeat
Others
KAFKA
BUS
Logstash
KAFKA
BUS
Spark Streaming
Flink
Storm
Elasticsearch
Kibana
Cache
Alluxio
HDFS
Spark
Flink
Zepplin
Postgresql
Mysql
Caravel

目录

1 我们的实时数据平台-Prism

2 从这里开始

3 架构演进

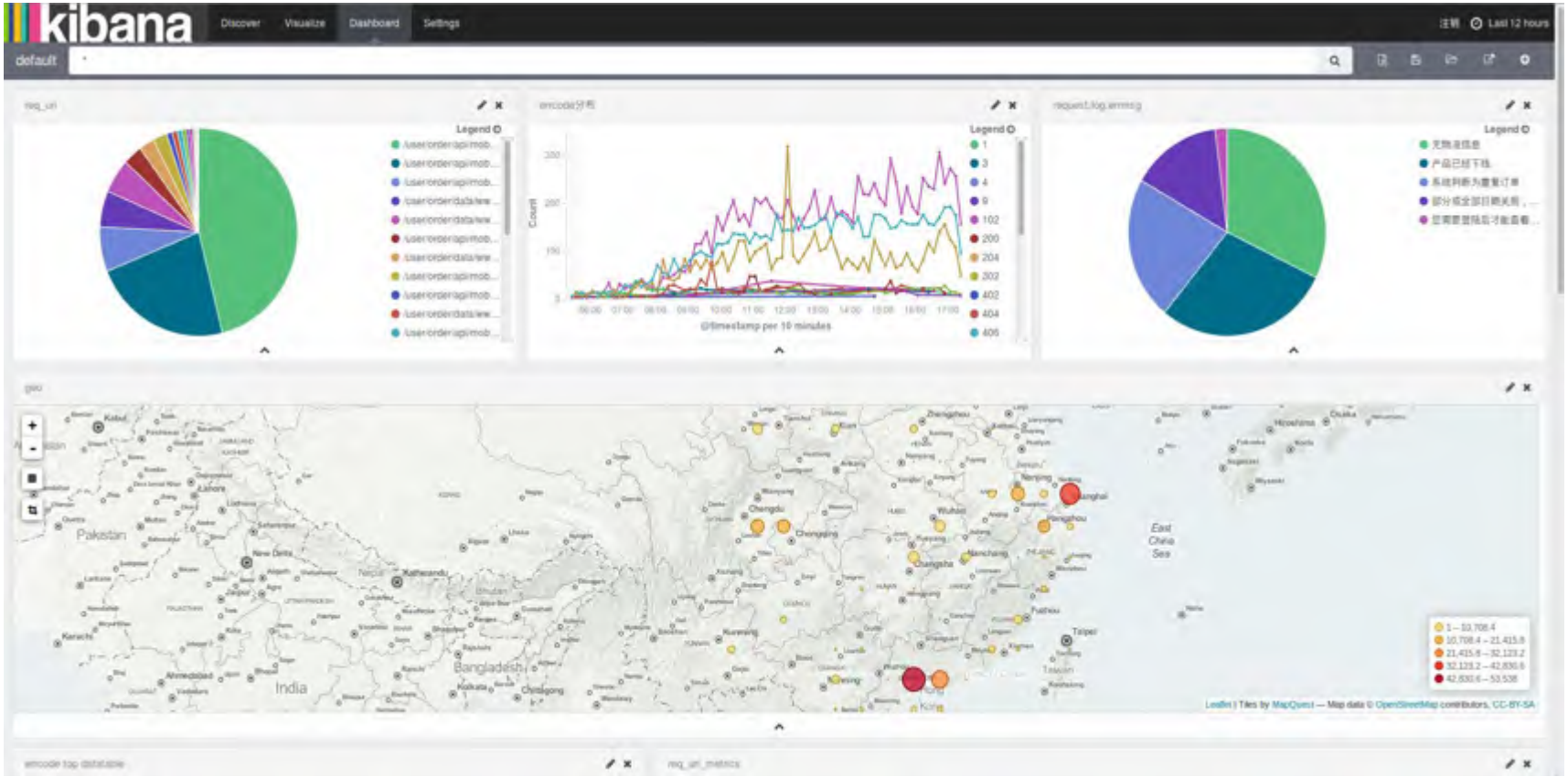
4 Esaas - Elasticsearch as a Service

5 监控

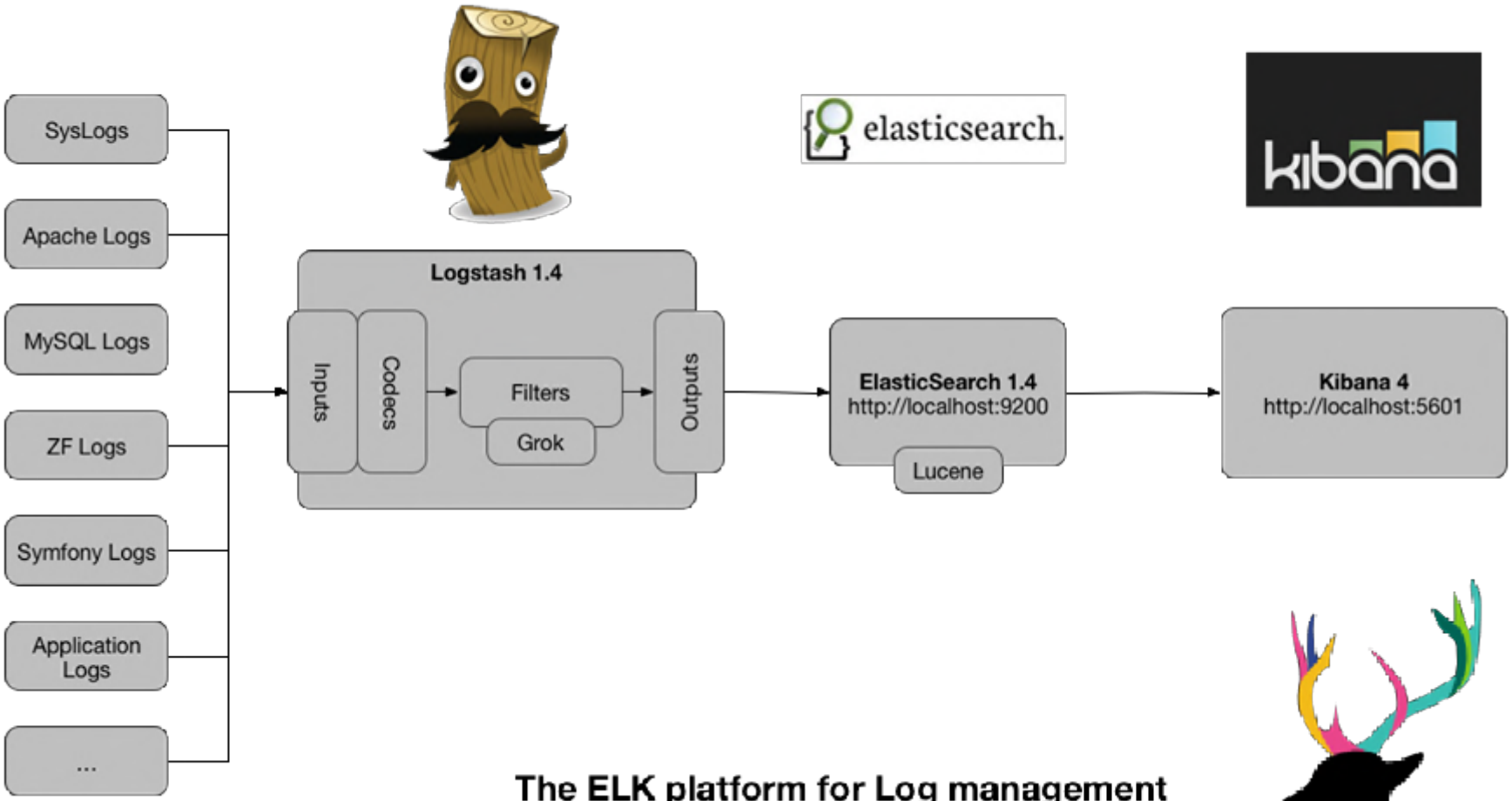
6 规模

Dev成了问题定位瓶颈

ELK


kibana
Discover
Visualize
Dashboard
Settings
default
Legend
@timestamp per 10 minutes
Count
Pakistan
India
Bangladesh
East China Sea

ELK



The ELK platform for Log management

大受欢迎

部署方式和问题

1. 部署方式
    1. 申请虚拟机/添加账号
    2. 使用salt部署
2. 面临的问题
    1. 快速构建业务流
    2. 快速增减容量

目录

1 我们的实时数据平台-Prism

2 从这里开始

3 架构演进

4 Esaas - Elasticsearch as a Service

5 监控

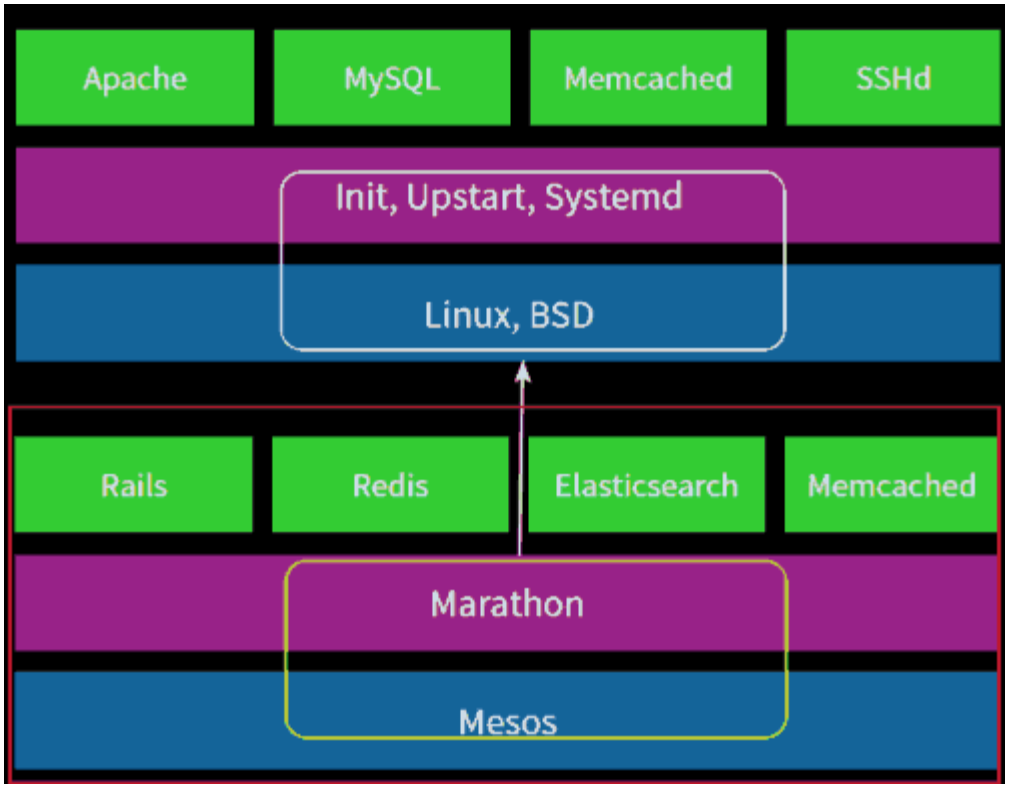
6 规模

发现新大陆

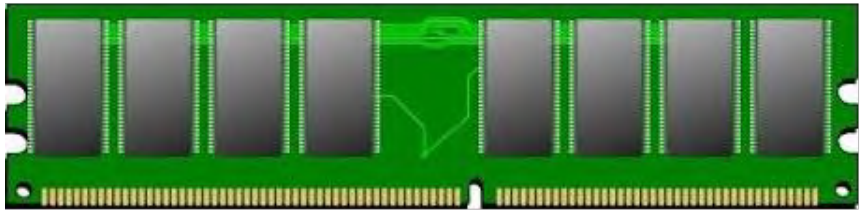
解决了问题

1. 快速增减容量
2. 新工具快速支持
3. 提高硬件资源利用率
4. 降低数据软件的使用成本

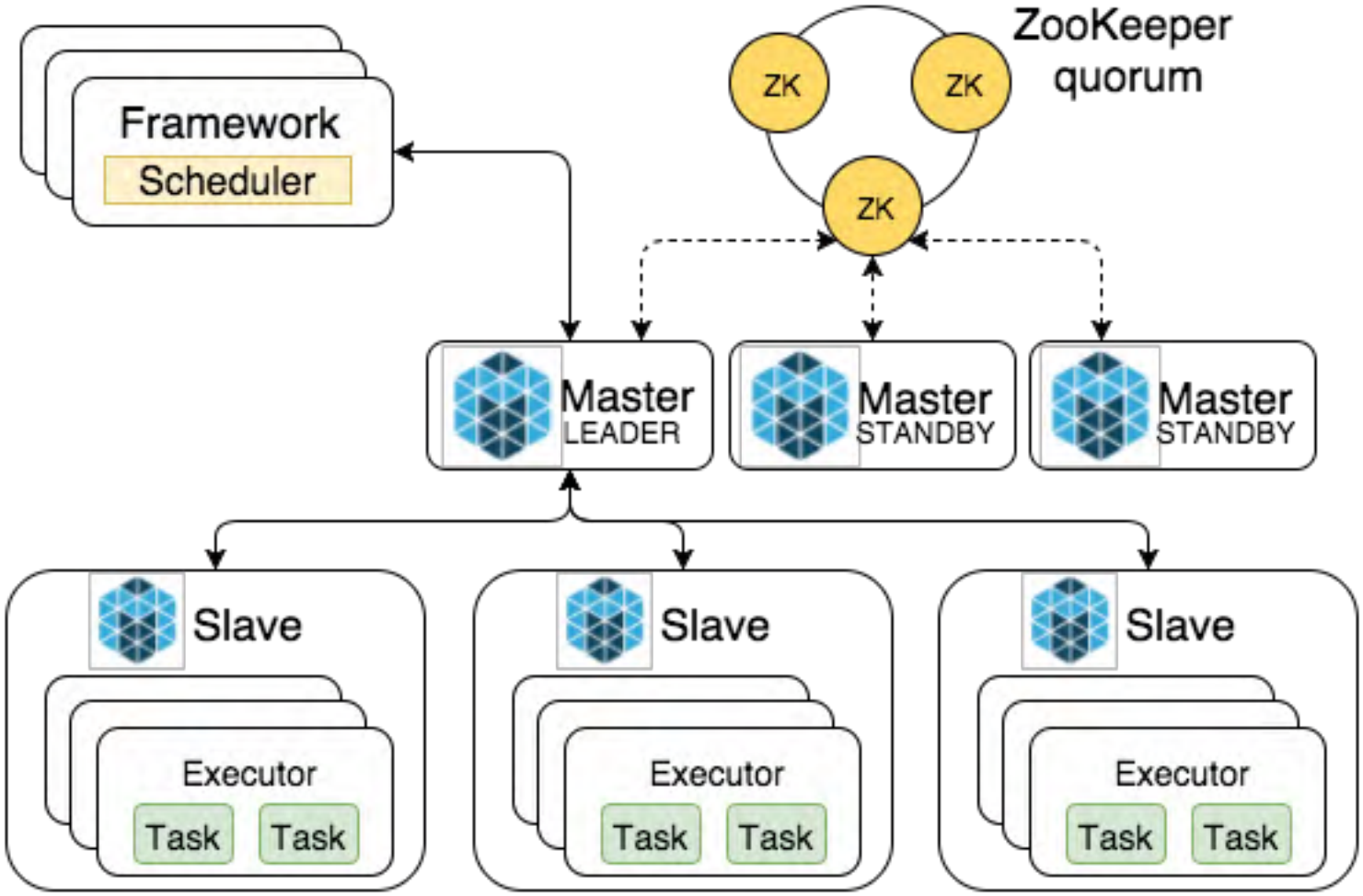
角色


Apache
MySQL
Memcached
SSHd
Init, Upstart, Systemd
Linux, BSD
Rails
Redis
Elasticsearch
Memcached
Marathon
Mesos

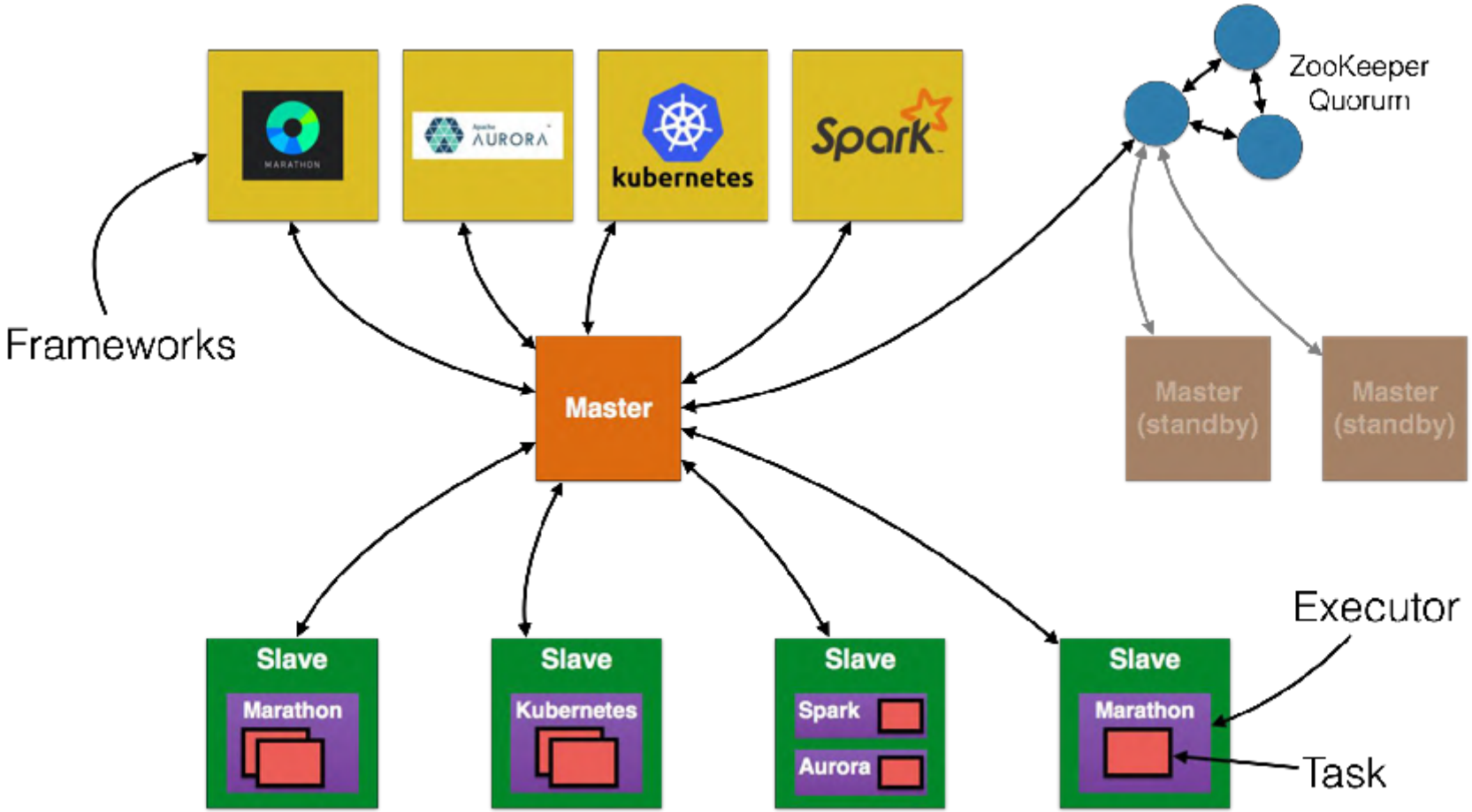
Mesos管理的资源



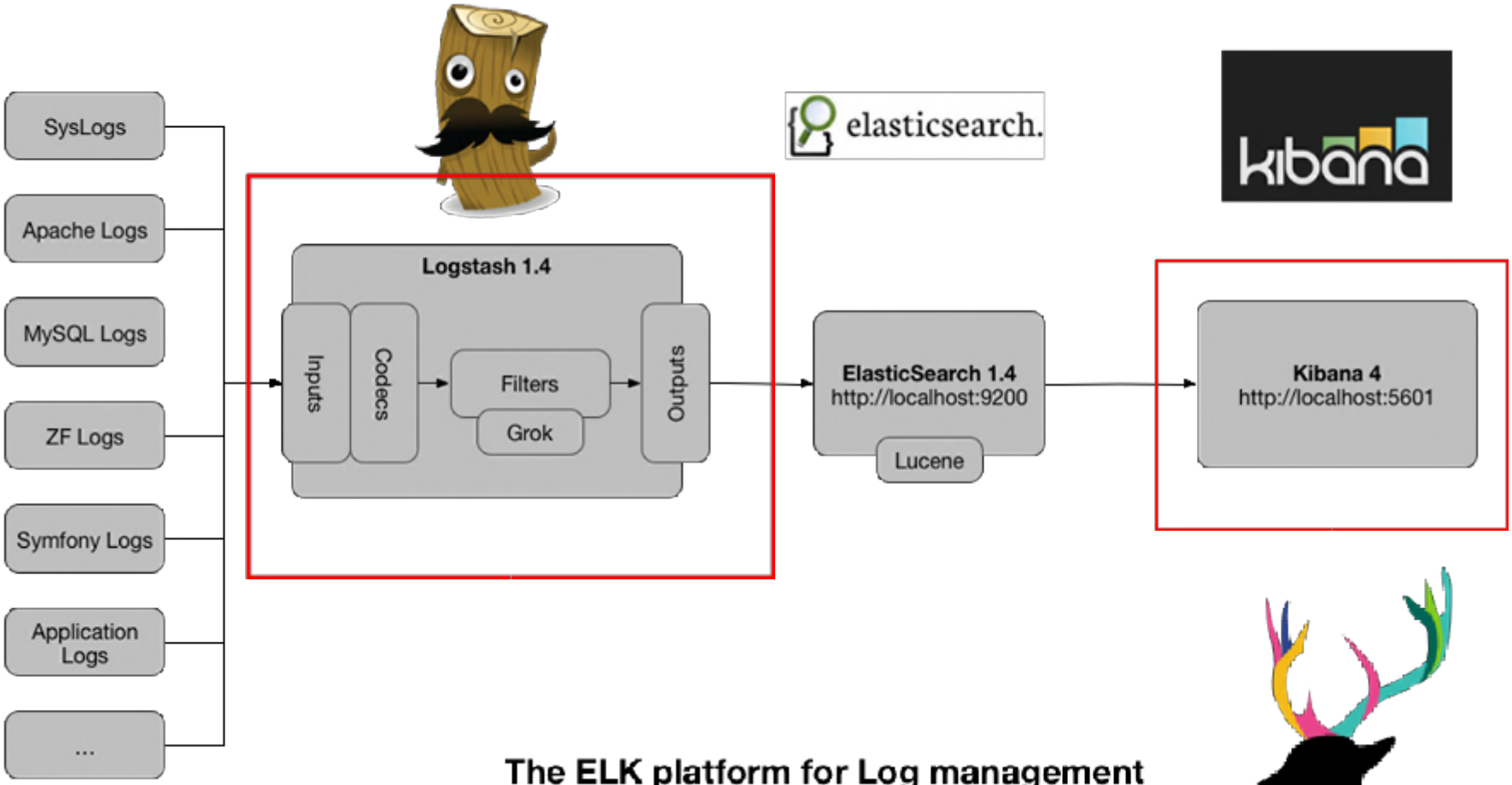
Mesos架构


ZooKeeper quorum
ZK
ZK
ZK
Framework
Scheduler
Master
LEADER
Master
STANDBY
Master
STANDBY
Slave
Executor
Task
Task
Slave
Executor
Task
Task
Slave
Executor
Task
Task

Marathon和Spark的位置

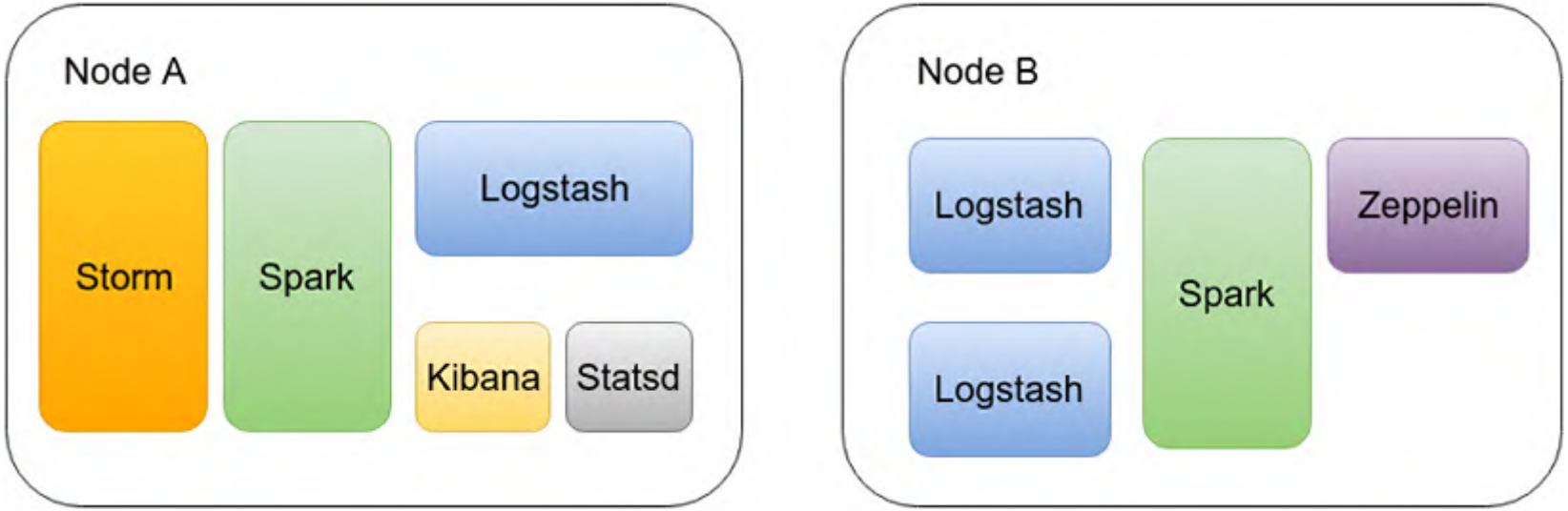


在Mesos上运行无状态服务



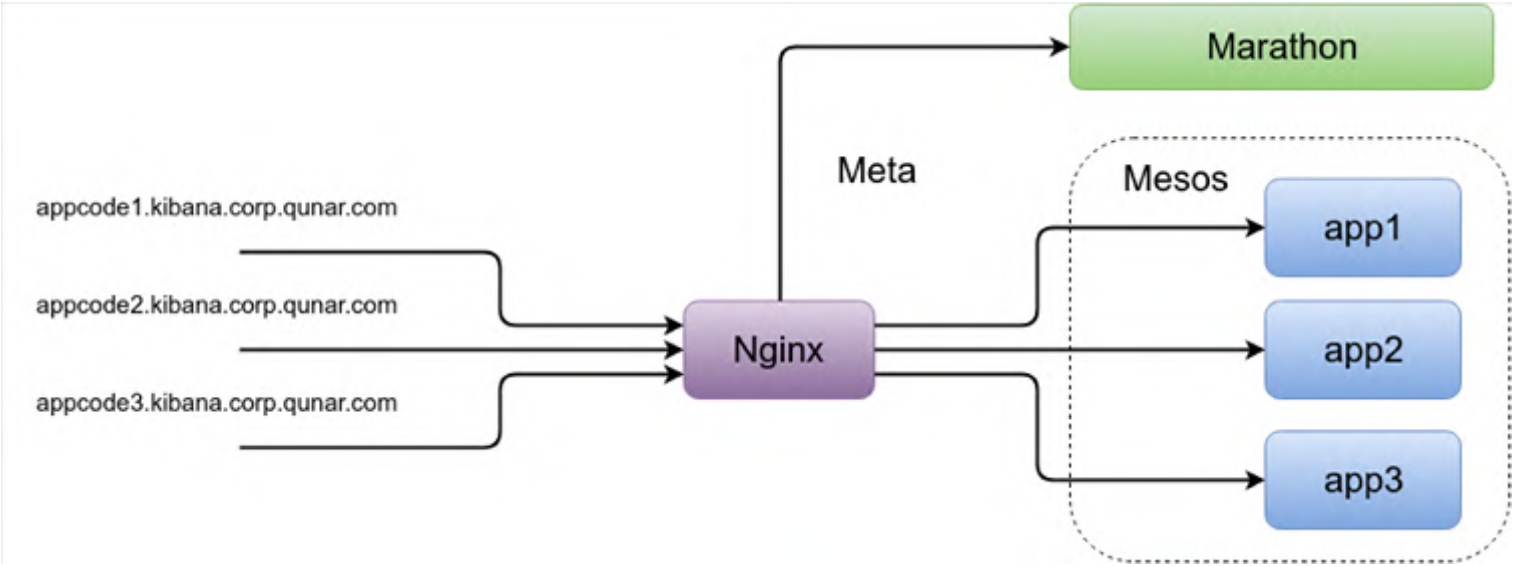
The ELK platform for Log management

节点快照

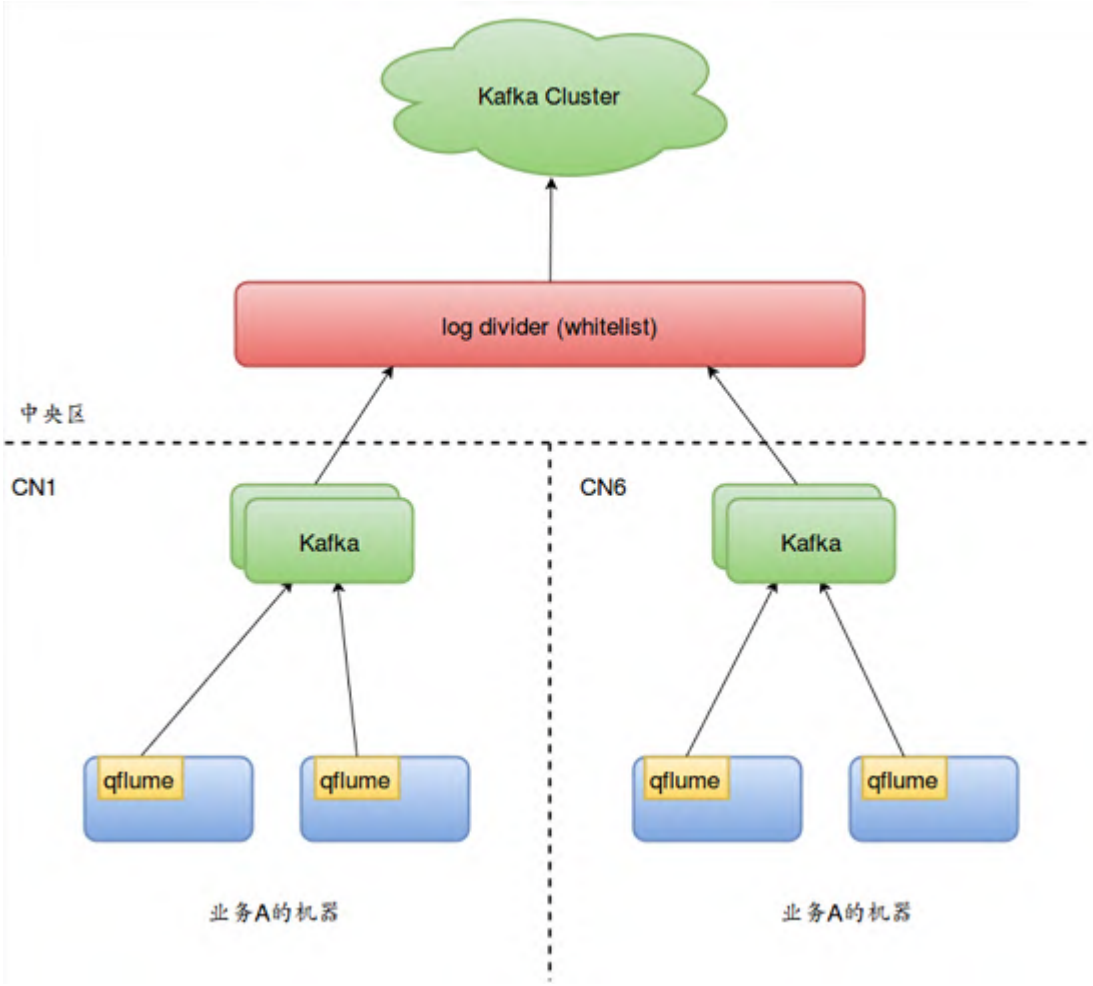

Node A
Storm
Spark
Logstash
Kibana
Statsd
Node B
Logstash
Logstash
Spark
Zeppelin

如何找到Kibana服务

1. 网络方案
    1. --net=host
    2. Calico
    3. CNI (Mesos version >= 1.0)
2. 请求路由/服务发现（HTTP）



多机房日志流汇聚


Kafka Cluster
log divider (whitelist)
中央区
CN1
CN6
Kafka
Kafka
qflume
qflume
qflume
qflume
业务A的机器
业务A的机器

Logstash Debugger

Logstash debugger Life's too short for bad tools

GitLab Wiki xiaoxu.lv 注销

Config 277 Init Debug history

Filter by name

- etl.dujia.m_innovation_pitcher_poster
- etl.hotel.h_qdsstat_qreport
- etl.train.t_train_mwhotdog
- etl.train.t_web_kylin
- etl.ucenter.u_hotdog
- etl.wireless.m_innovation_pitcher_poster
- etl.wireless.m_invocation_flight
- etl.wireless.m_invocation_flight_snk
- etl.wireless.m_invocation_hotdog
- etl.wireless.m_invocation_kylin
- logger.beta.car
- logger.beta.flight
- logger.beta.hotel
- logger.beta.piao
- logger.beta.sp
- logger.beta.ucenter
- logger.beta.wireless
- logger.beta
- logger.car.m_car_awardms

Pattern -> logs.logger.flight.f_av_data Gitlab Watcher Kibana

```
HTTPDATE %{MONTHDAY}/%{MONTH}/%{YEAR}:%{TIME} \+%{INT}
LOGTIME %{YEAR}\-%{MONTHNUM}\-%{MONTHDAY} %{HOUR}:%{MINUTE}:%{SECOND}
ERROR_KEYWORD \b(ERROR|.*?[eE]xception|.*?[Tt]imeout|Trying to connect to|WARN)
CLIENT ((unknown|%{IPV4}),[ ]?)*(%{IPV4}|-)
```

Input

自定义input字段

```
1
```

Filter

```
filter {
#     ruby {
#         code => "event['log_duration'] = (Time.parse(event['@timestamp'].to_s).to_f * 1000).to_i - event['send_time'].to_i"
#     }
#}
```

Debug Commit Deployment

```
scrot -s /tmp/s.png
➜ ~ scrot -s /tmp/s.png
➜ ~ scrot -s /tmp/s.png
```

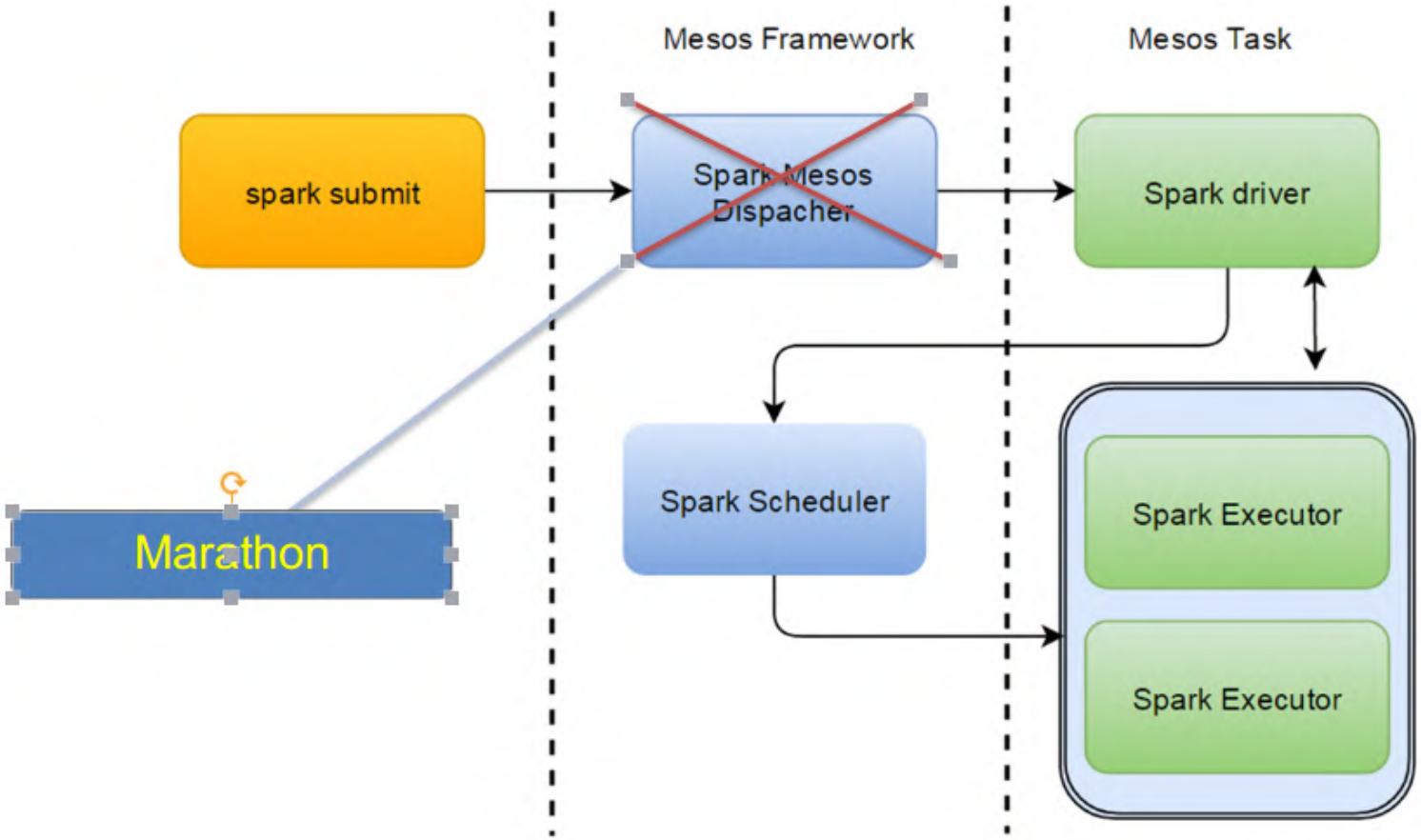
规模

1. 接入模块：300模块
2. 单Kafka最大带宽11G+
3. 每日116亿消息，18T+

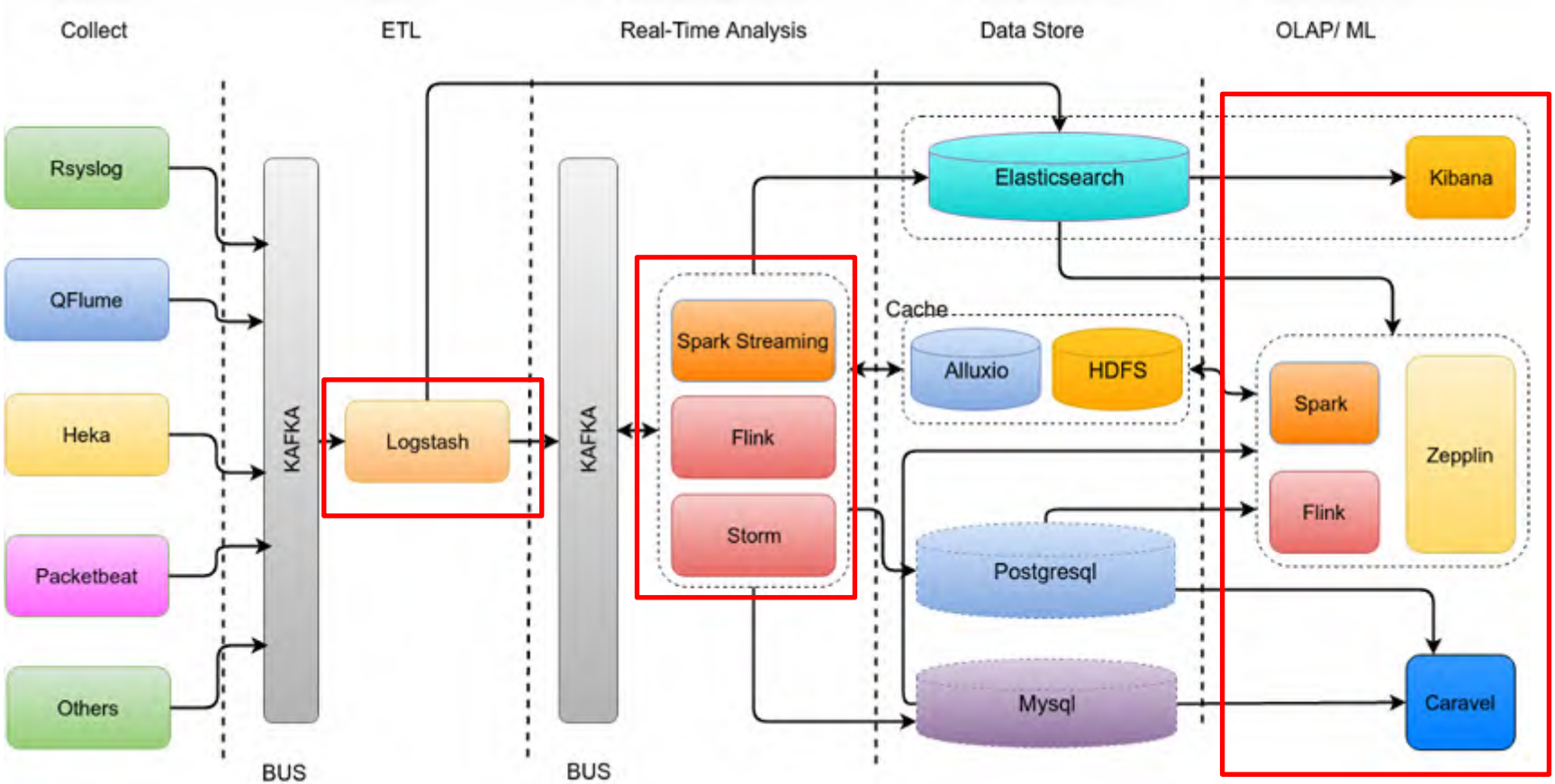
新的需求-更复杂的日志分析

1. 实时推荐
2. 多数据源实时JOIN

Spark on Mesos

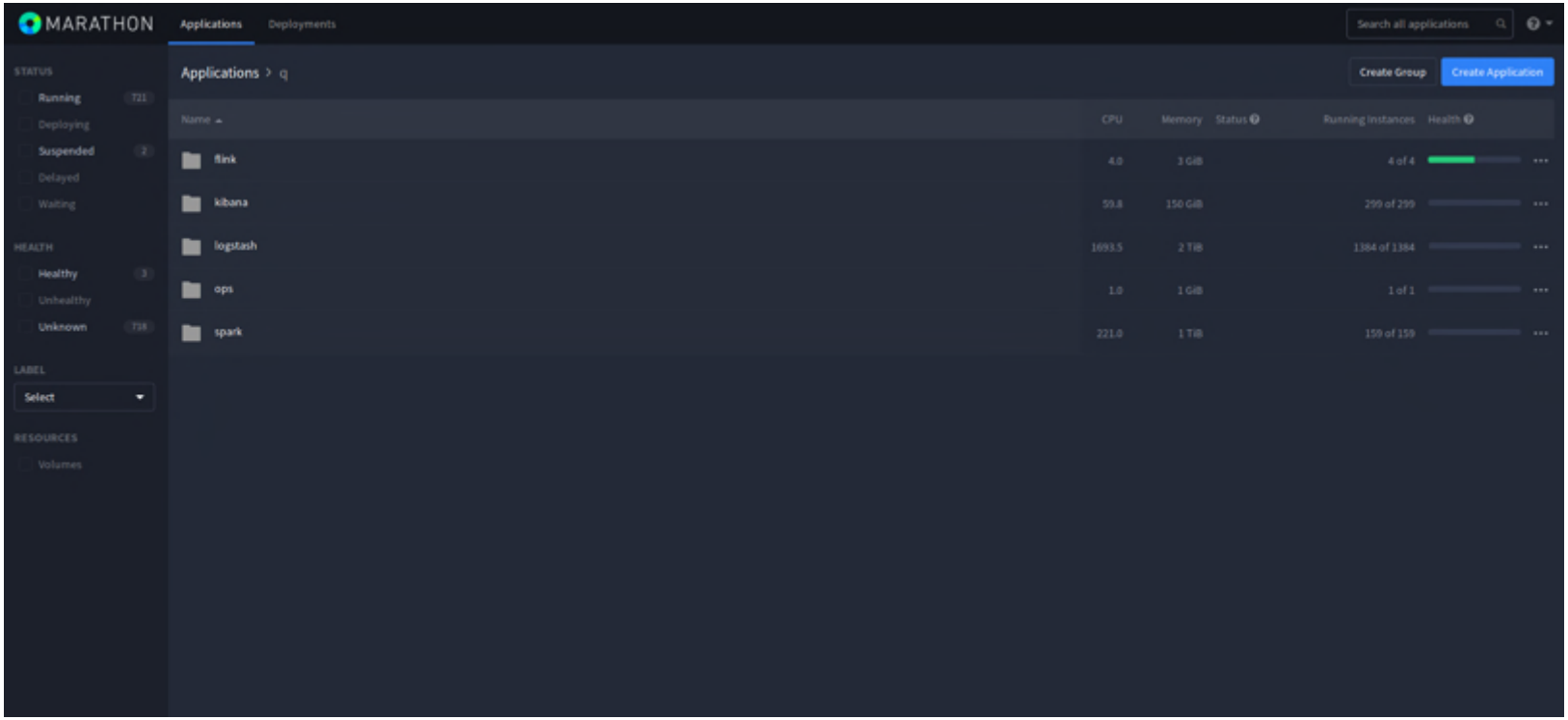


Software on Mesos

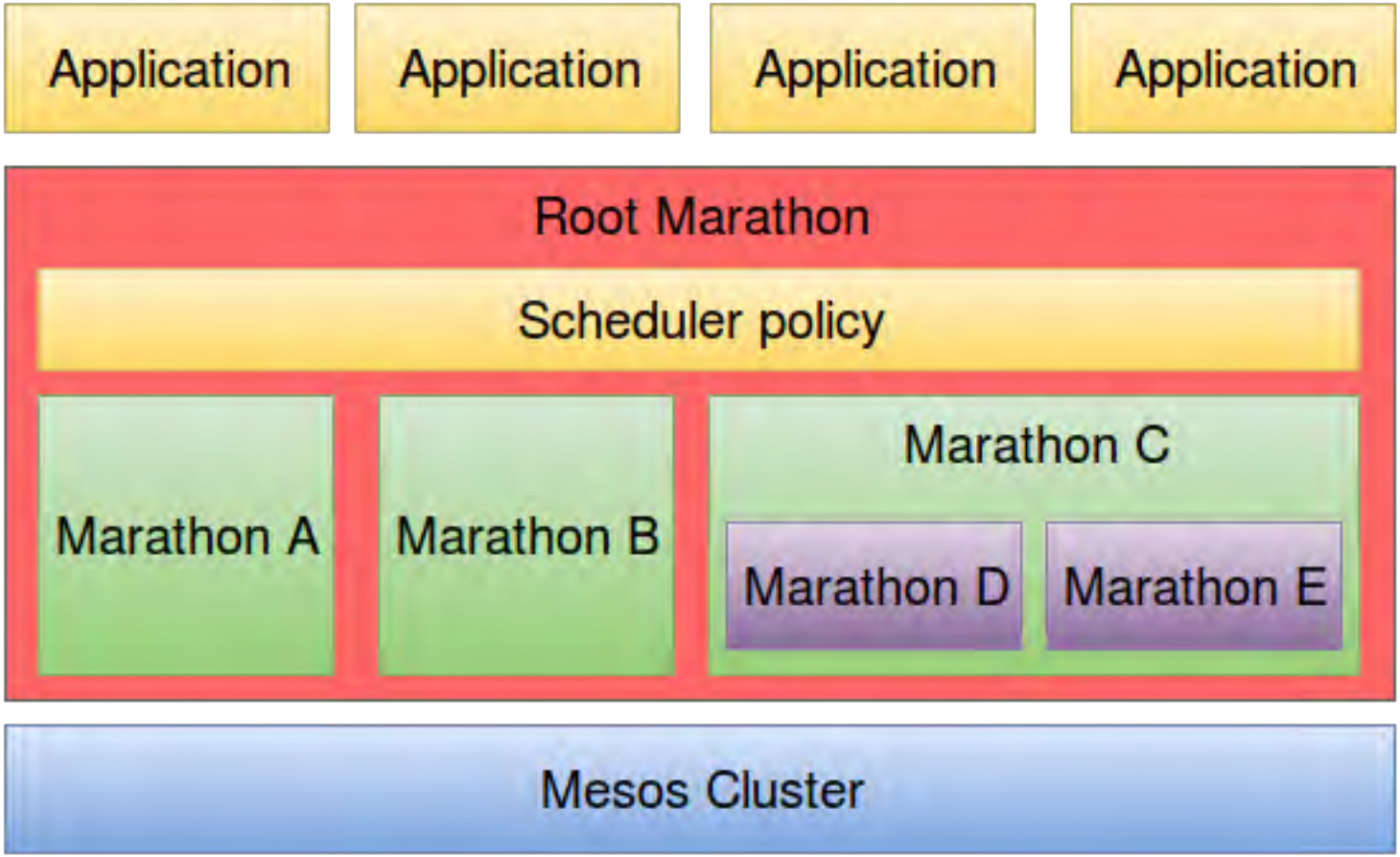


规模

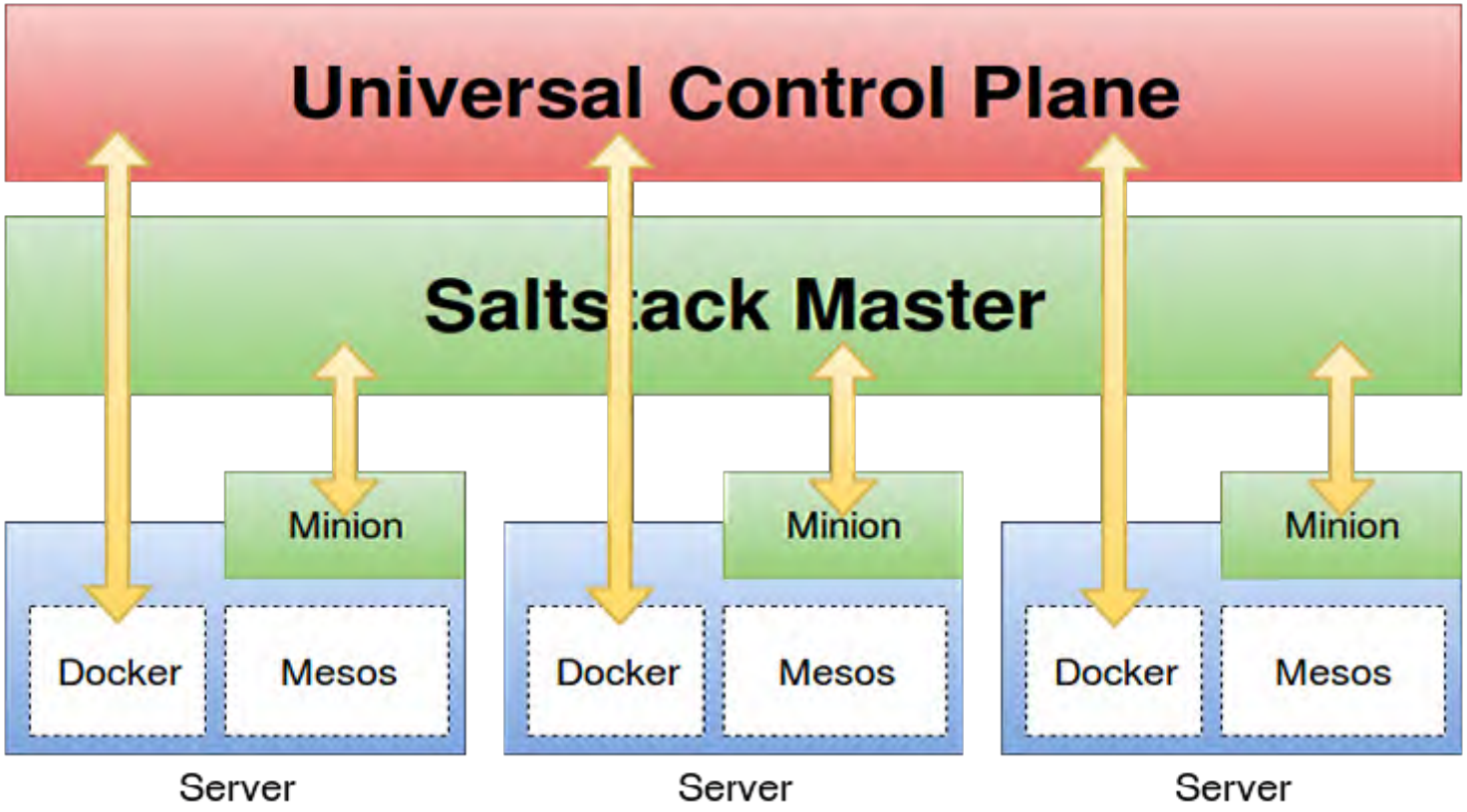
1. Spark Streaming任务：50个
2. Storm集群：5个
3. Flink集群：2个



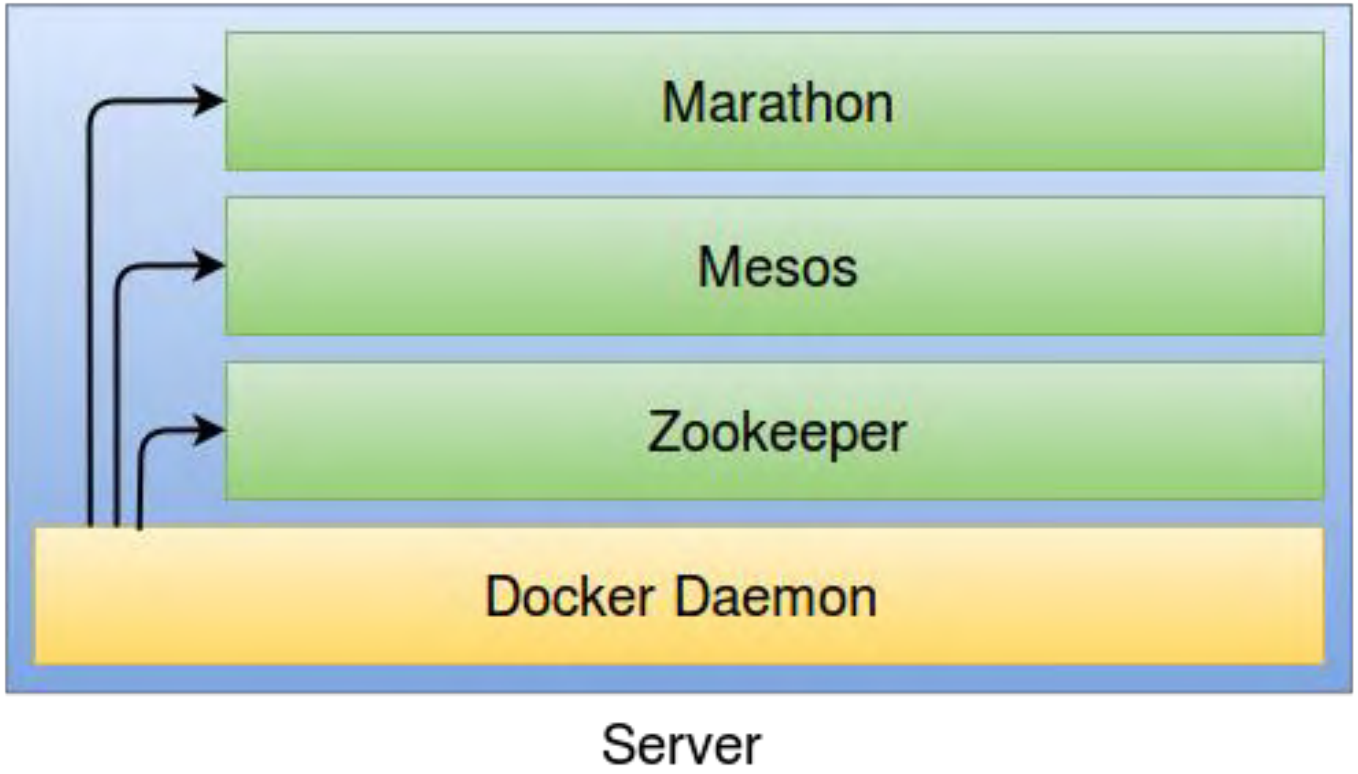
Quota


Application
Application
Application
Application
Root Marathon
Scheduler policy
Marathon A
Marathon B
Marathon C
Marathon D
Marathon E
Mesos Cluster

Cluster Bootstrap



All in Docker



目录

1 我们的实时数据平台-Prism

2 从这里开始

3 架构演进

4 Esaas - Elasticsearch as a Service

5 监控

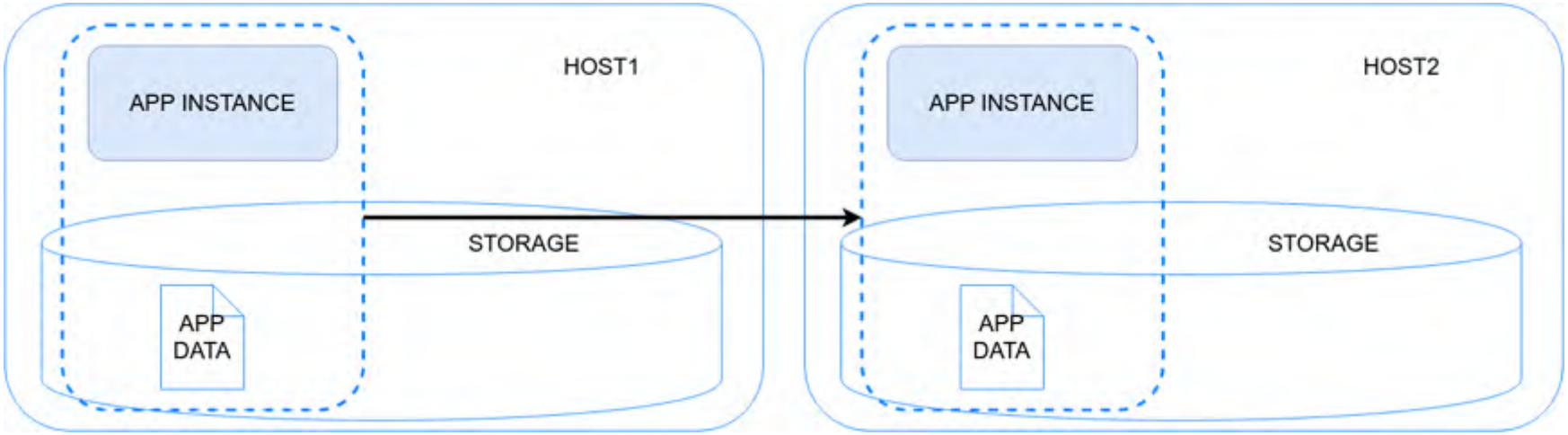
6 规模

有状态服务

1. 本地存储数据：Elasticsearch
2. 服务发现：Flink，HAProxy

带状态服务运行于Mesos要解决的问题

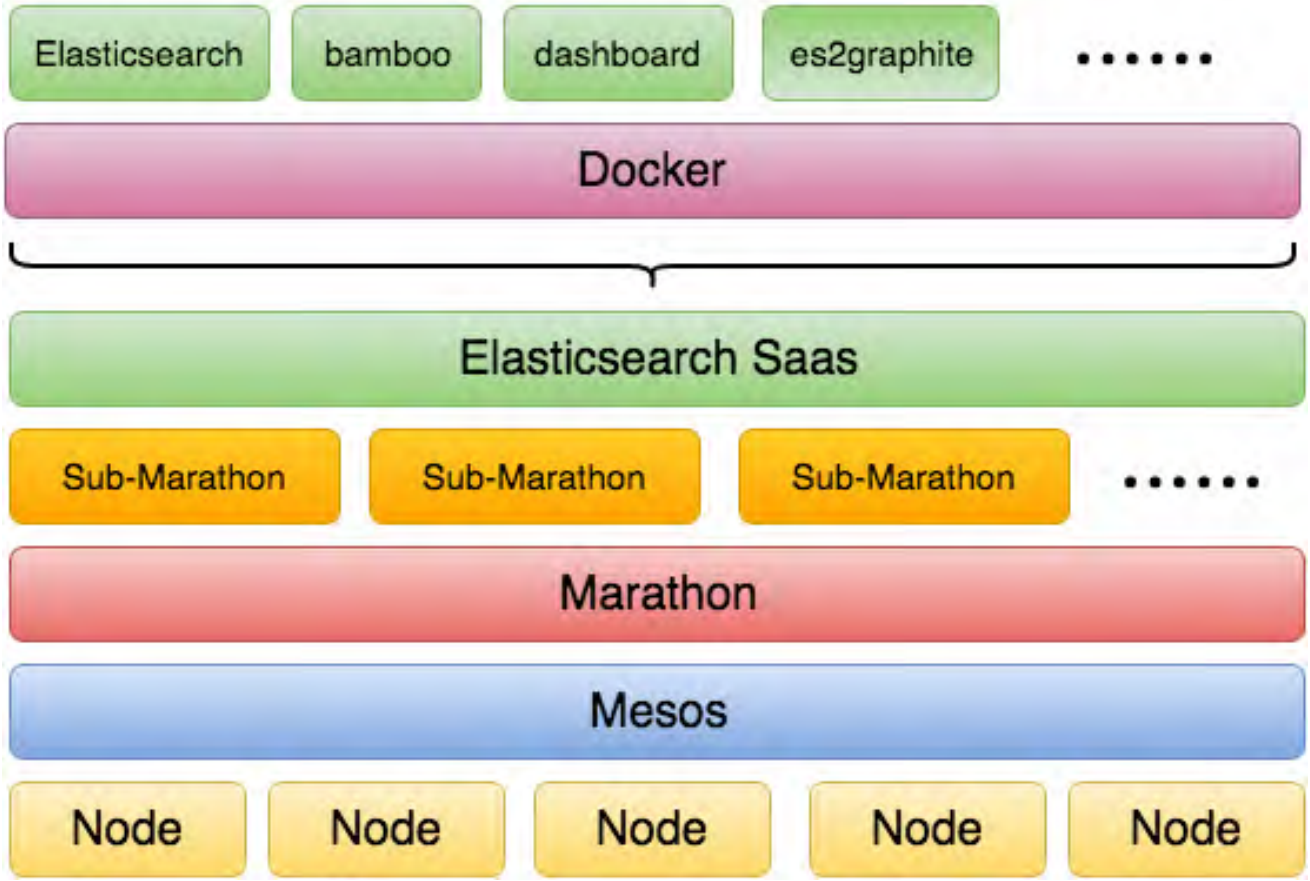
1. 本地存储数据：Elasticsearch
2. 服务发现：Flink，HAProxy



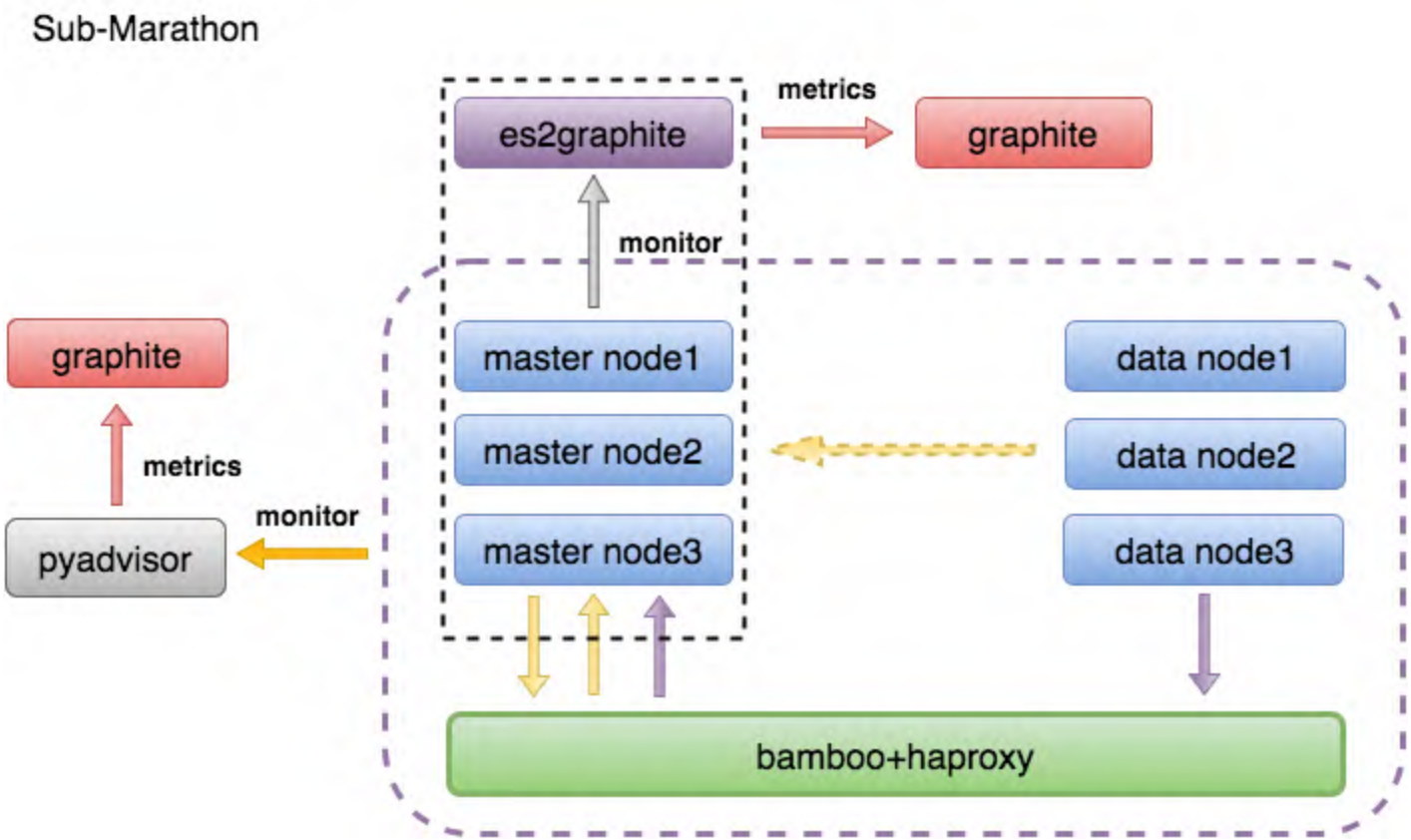
Mesos版本要求

- 持久化卷
- 动态预留
- Mesos > 0.23
- Marathon > 1.0

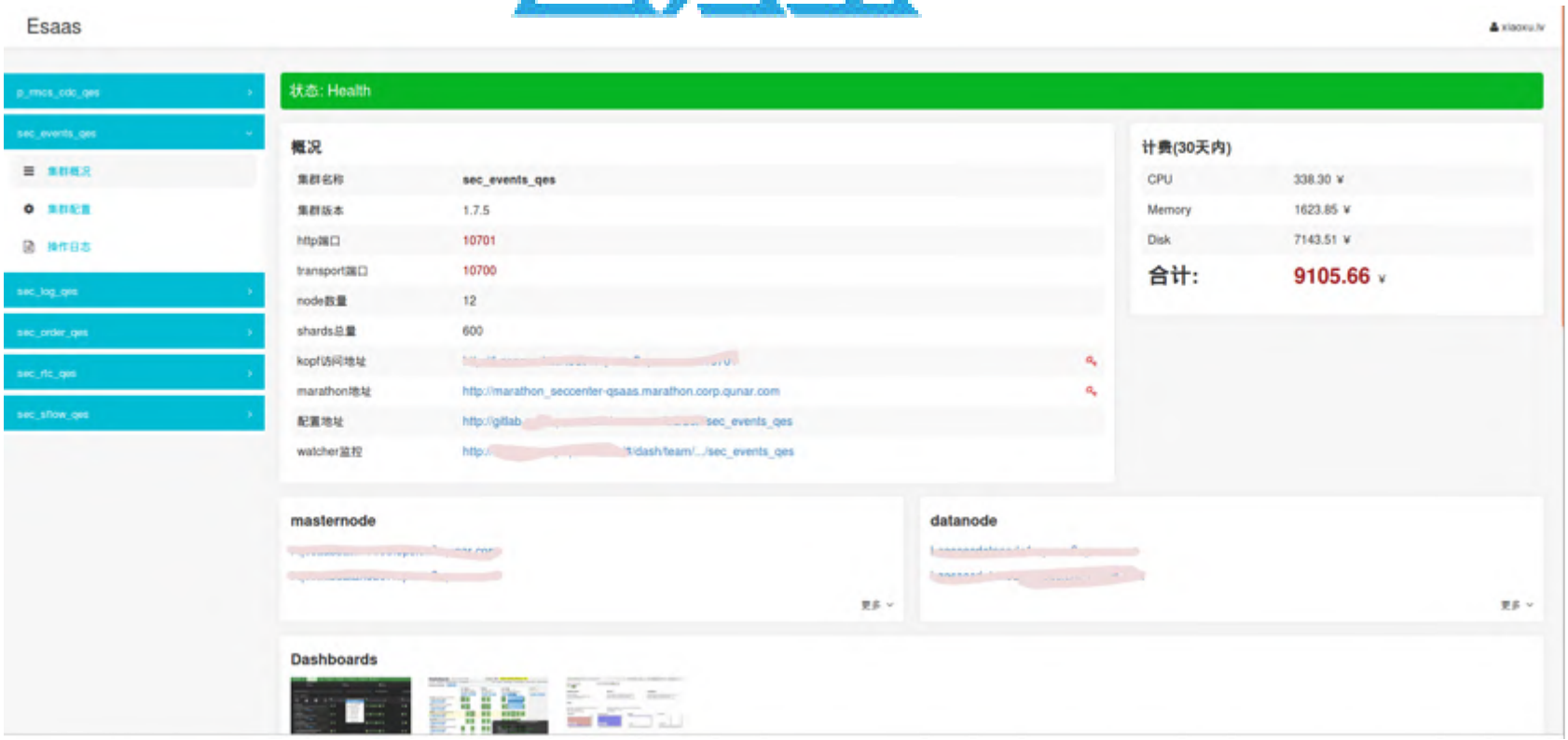
Esaas结构


Elasticsearch
bamboo
dashboard
es2graphite
.....
Docker
Elasticsearch Saas
Sub-Marathon
Sub-Marathon
Sub-Marathon
.....
Marathon
Mesos
Node
Node
Node
Node
Node

服务发现-TCP


Sub-Marathon
es2graphite
metrics
graphite
monitor
graphite
metrics
monitor
pyadvisor
master node1
master node2
master node3
data node1
data node2
data node3
bamboo+haproxy

DashBoard


Esaas
状态: Health
sec_events_qes
集群概况
集群配置
操作日志
sec_log_qes
sec_order_qes
sec_rfc_qes
sec_sflow_qes
概况
集群名称 sec_events_qes
集群版本 1.7.5
http端口 10701
transport端口 10700
node数量 12
shards总量 600
kopf访问地址
marathon地址 http://marathon_seccenter-qsaas.marathon.corp.qunar.com
配置地址 http://gitlab sec_events_qes
watcher监控 http:/ t/dash/team/.../sec_events_qes
计费(30天内)
CPU 338.30 ¥
Memory 1623.85 ¥
Disk 7143.51 ¥
合计: 9105.66 ¥
masternode
datanode
Dashboards

目录

1 我们的实时数据平台-Prism

2 从这里开始

3 架构演进

4 Esaas - Elasticsearch as a Service

5 监控

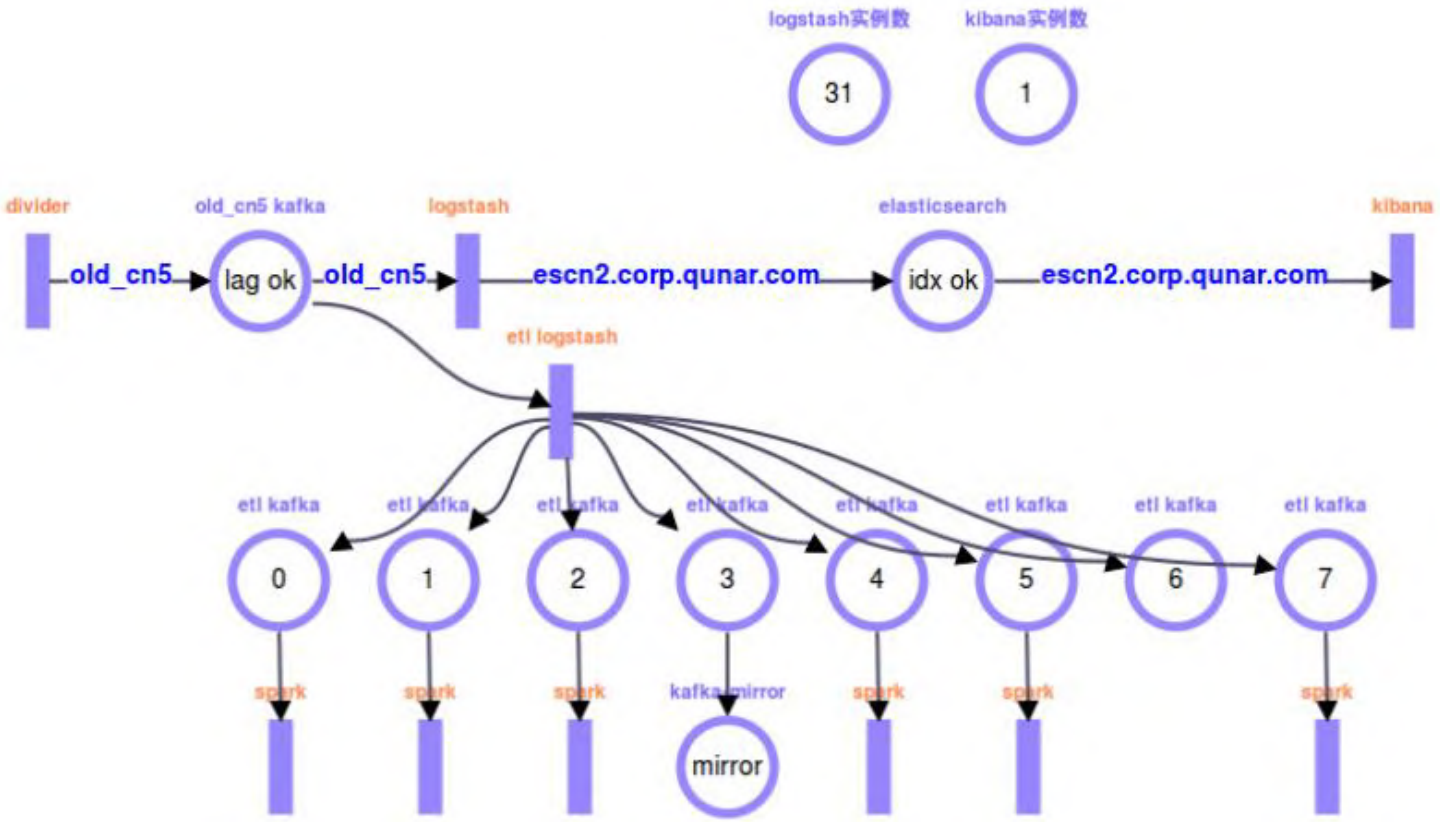
6 规模

监控

1. 数据处理模块拓扑监控
2. 业务监控
    1. 队列堆积：Kafka Topic Lag
    2. 流量：Search Count/Message Count
    3. 错误：Reject/Exception
3. 基础监控/容量监控
    1. IO使用率
    2. CPU使用率
    3. 内存使用率
    4. JVM/GC等
    5. 计算资源使用量监控

数据处理模块拓扑监控

logs_wireless_m_pub_web_hotdog的网络

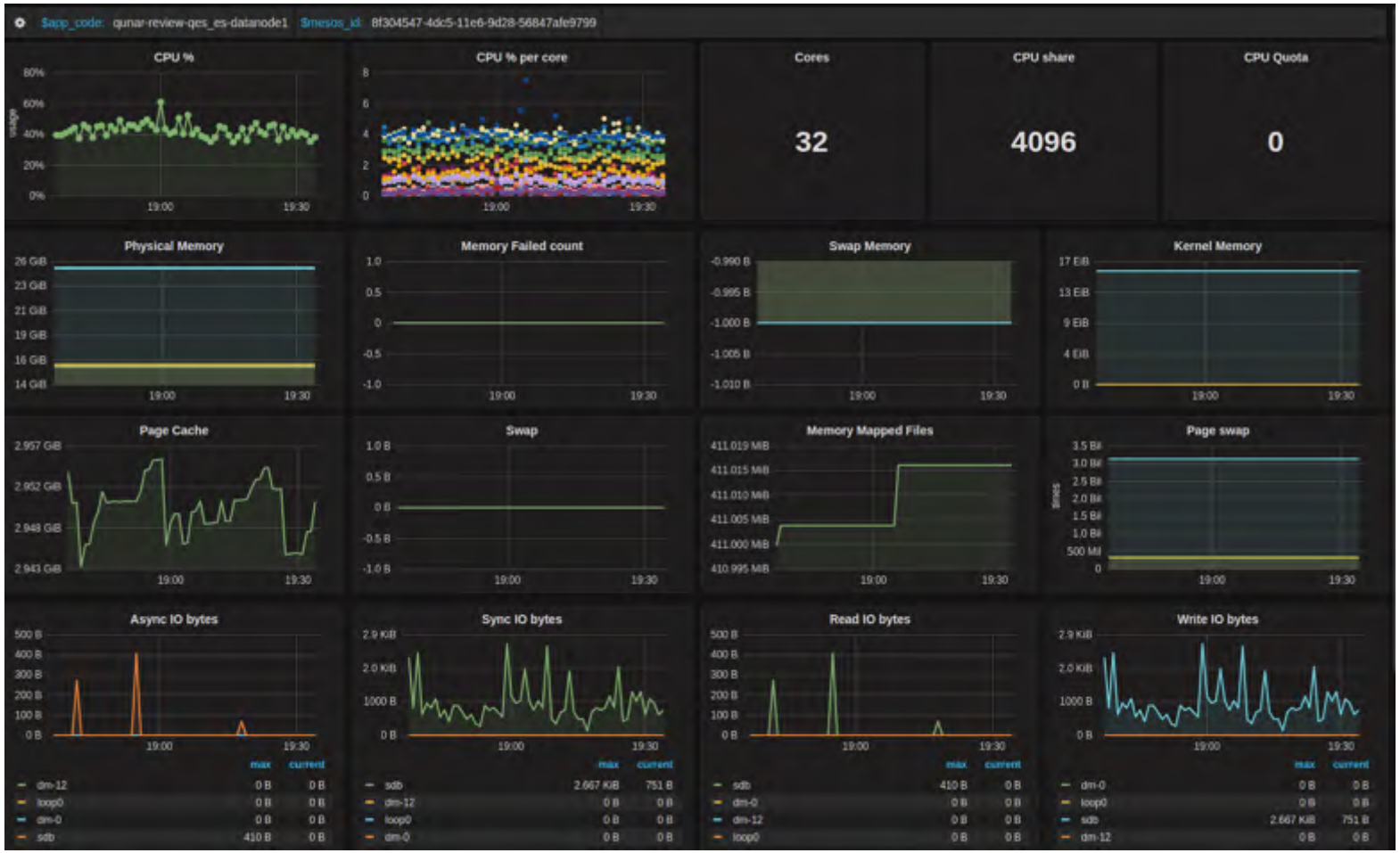


实时流监控

[flight] [cn1] log count
[flight] [cn1] which topic is sinner
[flight] [cn1] partition lag
2016-12-04 16:34:00
current
total
1.767 K
416.823 K

Esaas某ES集群监控

基础监控


$app_code: qunar-review-qes_es-datanode1 $mesos_id: 8f304547-4dc5-11e6-9d28-56847afe9799
CPU %
CPU % per core
Cores
32
CPU share
4096
CPU Quota
0
Physical Memory
Memory Failed count
Swap Memory
Kernel Memory
Page Cache
Swap
Memory Mapped Files
Page swap
Async IO bytes
Sync IO bytes
Read IO bytes
Write IO bytes

pyadvisor

https://github.com/QunarOPS/pyadvisor

目录

1 我们的实时数据平台-Prism

2 从这里开始

3 架构演进

4 Esaas - Elasticsearch as a Service

5 监控

6 规模

规模

1. 计算集群120+； 2600+ 容器
2. Esaas 50+； 47 ES集群； 600+ 容器
3. Cpu Usage: Openstack – 14%; Mesos – 28%

总结

1. 我们做的事儿
   - 解决数据软件的部署的门槛
   - 解决Mesos环境部署的门槛
2. 仍存在问题
   - 负载不均衡
   - 数据异常定位速度慢
3. 下一步计划
   - 解决以存在的问题
   - 接入新软件
   - GPU计算平台建设

DevOpsDays 即将首次登陆中国

DevOps 之父 Patrick Debois 与您相约
DevOpsDays 北京站 2017年3月18日

门票早鸟价仅限前100名，请从速哟

http://2017-beijing.devopsdayschina.org/

想第一时间看到

高效运维社区公众号

的好文章吗？

请打开高效运维社区公众号，点击右上角小人，如右侧所示设置就好

Thanks

高效运维社区

开放运维联盟

荣誉出品