饿了么

饿了么Elasticsearch自动化运维和监控平台实战



徐胜

2018.3.24

饿了么

个人介绍

2015年，华师大硕士毕业；

饿了么搜索推荐部。

一直做elasticsearch、实时计算，底层架构和源码研究；

Elasticsearch与大数据平台交互；

Elasticsearch与知识图谱研究；

基于LSTM+softmax的用户意图识别；

自然语言处理和语义分析。

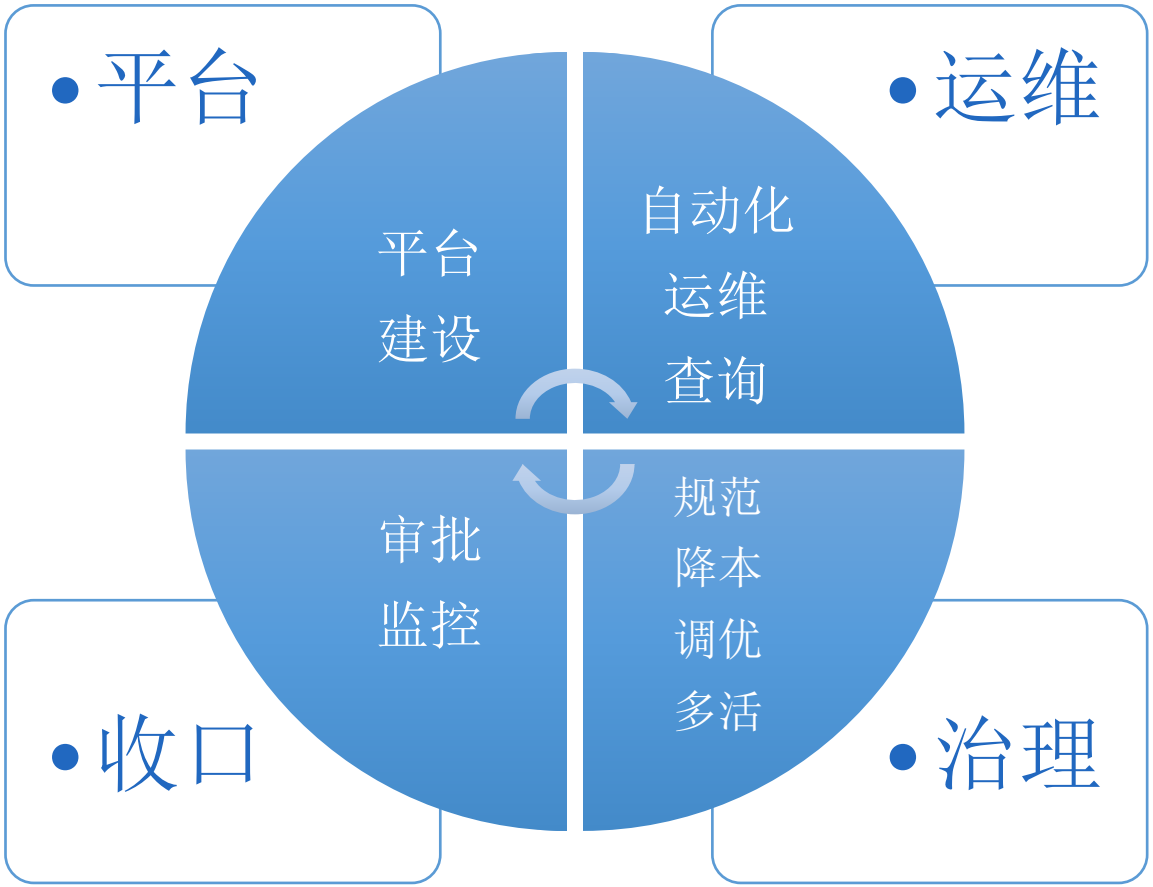
饿了么

CONTENTS

背景介绍

es平台和优化

未来展望



应用场景

核心几个问题

- 业务支撑
  - ✓ 面向C端业务推广
- 优化治理
  - ✓ 优化核心Apollo ES集群
  - ✓ ES业务多活数据
- 自动化运维平台和监控平台
  - ✓ ES-ops自动化运维平台
  - ✓ 基于grafana的监控Dashboard
- 接入收口
  - ✓ 业务接入评估
  - ✓ 监控业务行为

主要问题

ES集群增长规模

2015-2017年

- ✓ 产研ES cluster 12个
- ✓ 机器node 100-200个左右
- ✓ 数据量20TB
- ✓ Apollo每天ES索引量400万单左右
- ✓ 版本提供：2.3.3
- ✓ 业务输出：B端和物流端

2017年底，数据扩大了3倍。

- ✓ 产研部门的ES clusters 32个集群
- ✓ 机器node达300-400个
- ✓ indices索引数量达千万个
- ✓ 数据量60TB级别
- ✓ 业务线：10+多个业务线
- ✓ Apollo每天ES索引量600万单
- ✓ 版本升级：2.3.3和5.4.3
- ✓ C端：帮忙帮送-热门代购、搜索美食推荐、用户交易、订单搜索、新零售、智能客服、流量热区
- ✓ B端：中后台代理商、代金券、营销、特征工程

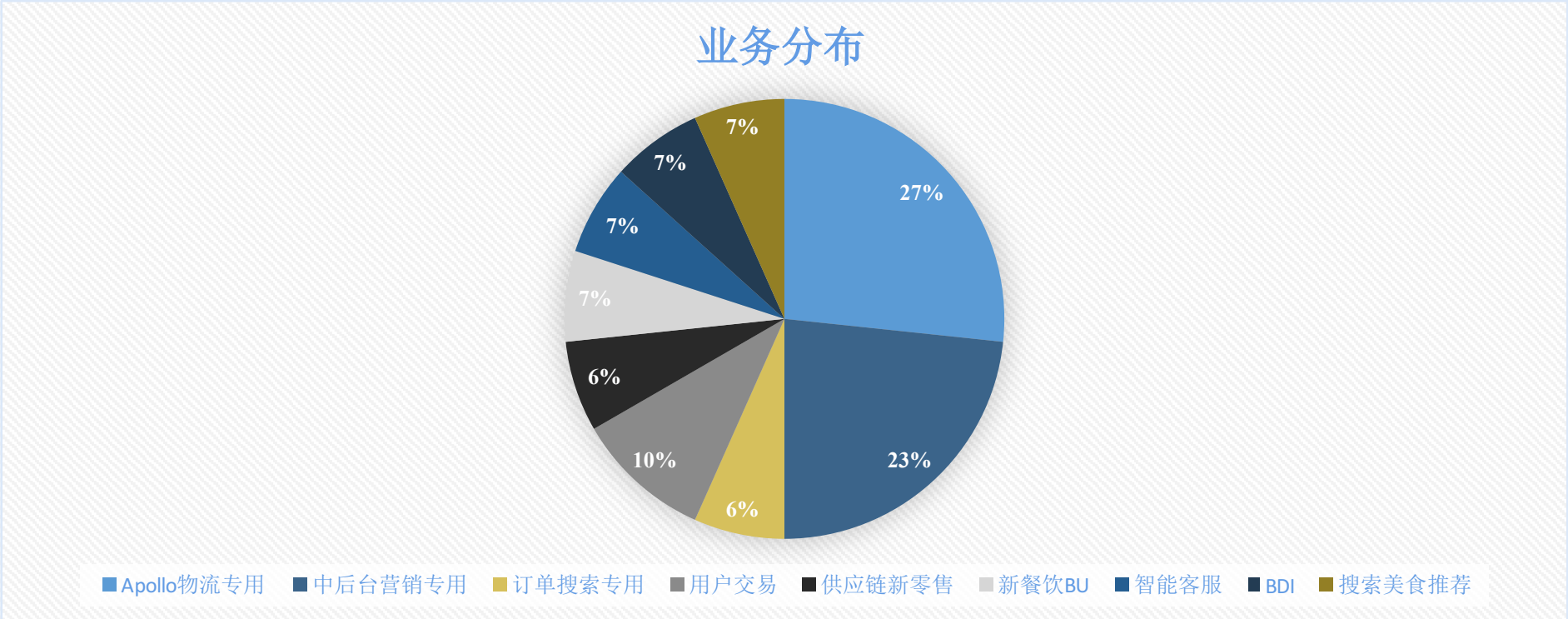
数据规模

ES业务支撑规模

部门	业务类型	集群	多活	数据量	业务对接人
Apollo物流专用	物流	7个	多活	60TB+	
中后台营销专用	B端	7个	多活	20TB+	
订单搜索专用	交易	2个	多活	10TB+	
用户交易	外卖端	3个	多活	10TB+	
供应链新零售	C端	2个	多活	。。	
新餐饮BU	C端	2个	多活	。。	
智能客服	C端	2个	多活	。。	
BDI	大数据	4个	多活		
搜索美食推荐	C端	2个	多活		

数据规模

elasticsearch在饿了么业务支撑



业务分布

痛点问题。。。。非常繁杂

- ✓ 资源严重不足

  物流apollo集群问题：资源heap空间占用较大，disk严重不足，查询query复杂，索引结构混乱。

- ✓ 缺少对索引数据可视化查询界面

- ✓ 多活改造需求很大

  ES需要实现多活

- ✓ 数据同步费劲

- ✓ 使用姿势不规范、排障困难

饿了么

针对Apollo的问题提供的改造方案

物流研发部 apollo ES 改造措施

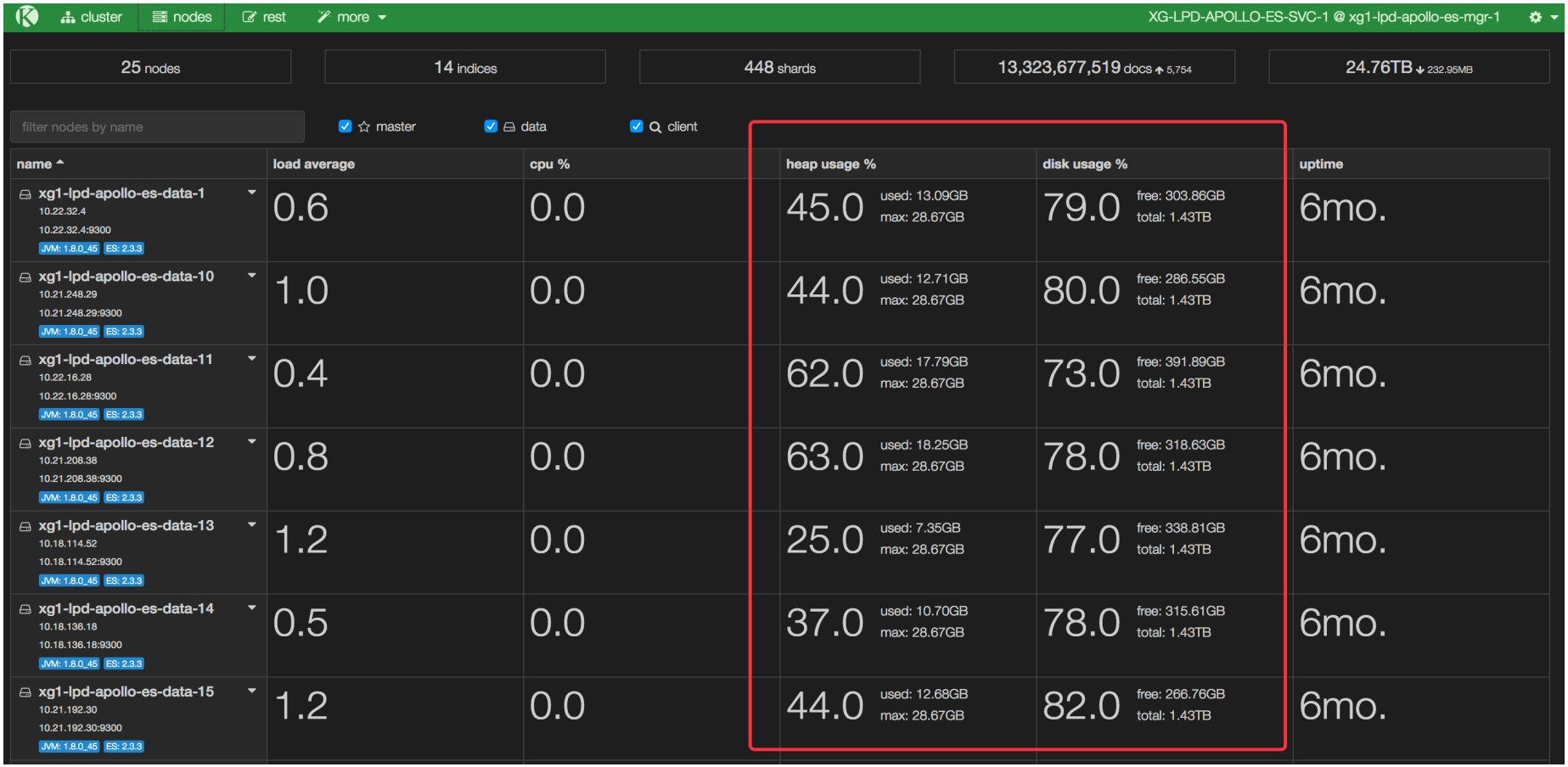
- ✓ 从主流程RMQ中迁移ES，建立索引的队列，避免ES索引慢
- ✓ 修改索引的mapping，修改索引为每天建立一个索引，方便运维管理、避免短时间的查询扫描一周的大索引。
- ✓ 优化慢查询API接口，优化查询语句，将filter和query结合起来
- ✓ 增加ES request cache的大小，保证第一次查询结果缓存到cache中
- ✓ 设置限流的机制，增加降级开关

优化结果：

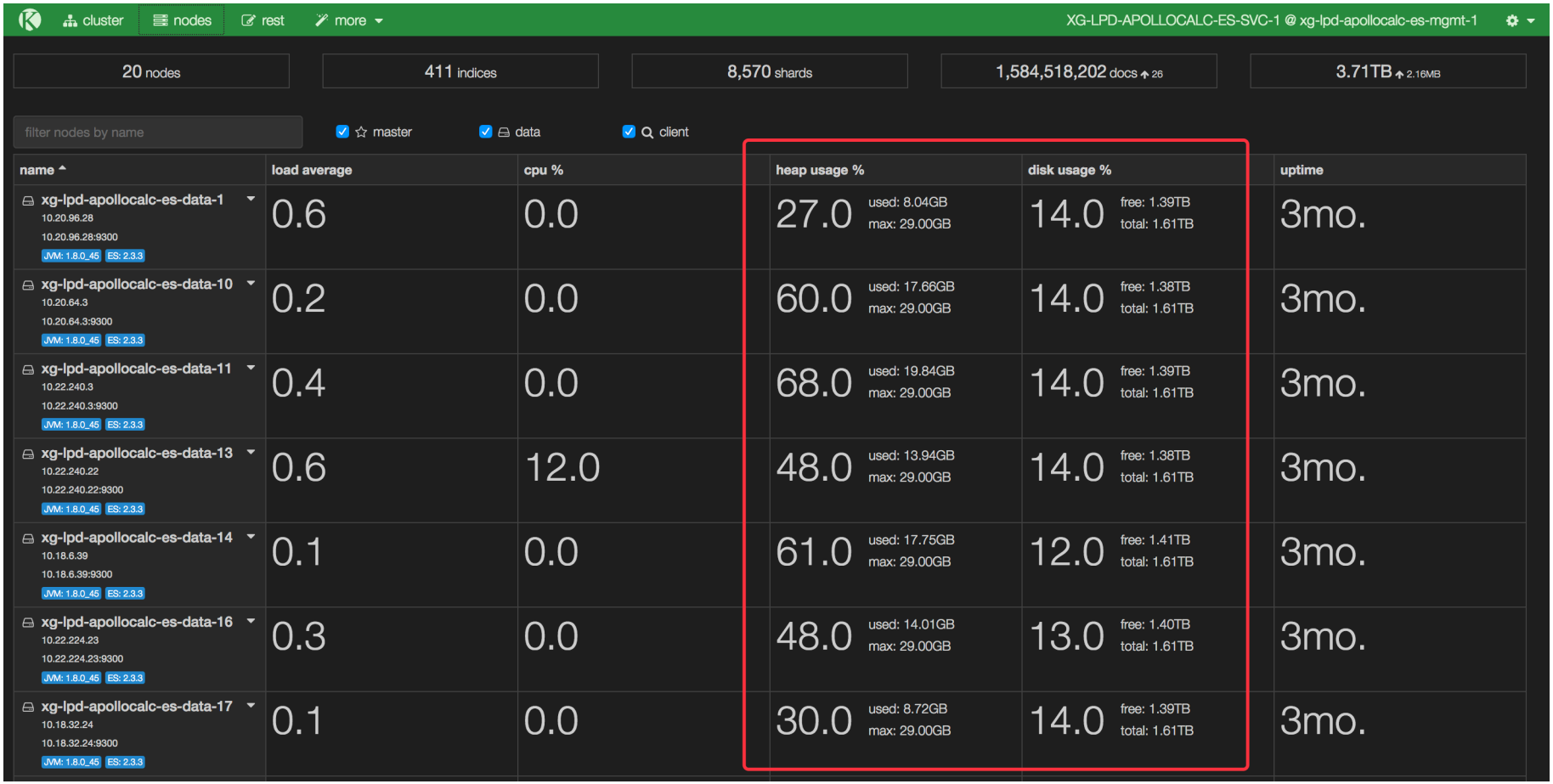
- ✓ 从25台机器（py12）（多活集群50台）降到了20台，Apollo单个集群节省约40万+的机器成本
- ✓ 索引结构减少，从130亿doc到15亿doc
- ✓ 查询效果变快，10ms以后的API接口全部都压到了1s左右

ES优化治理

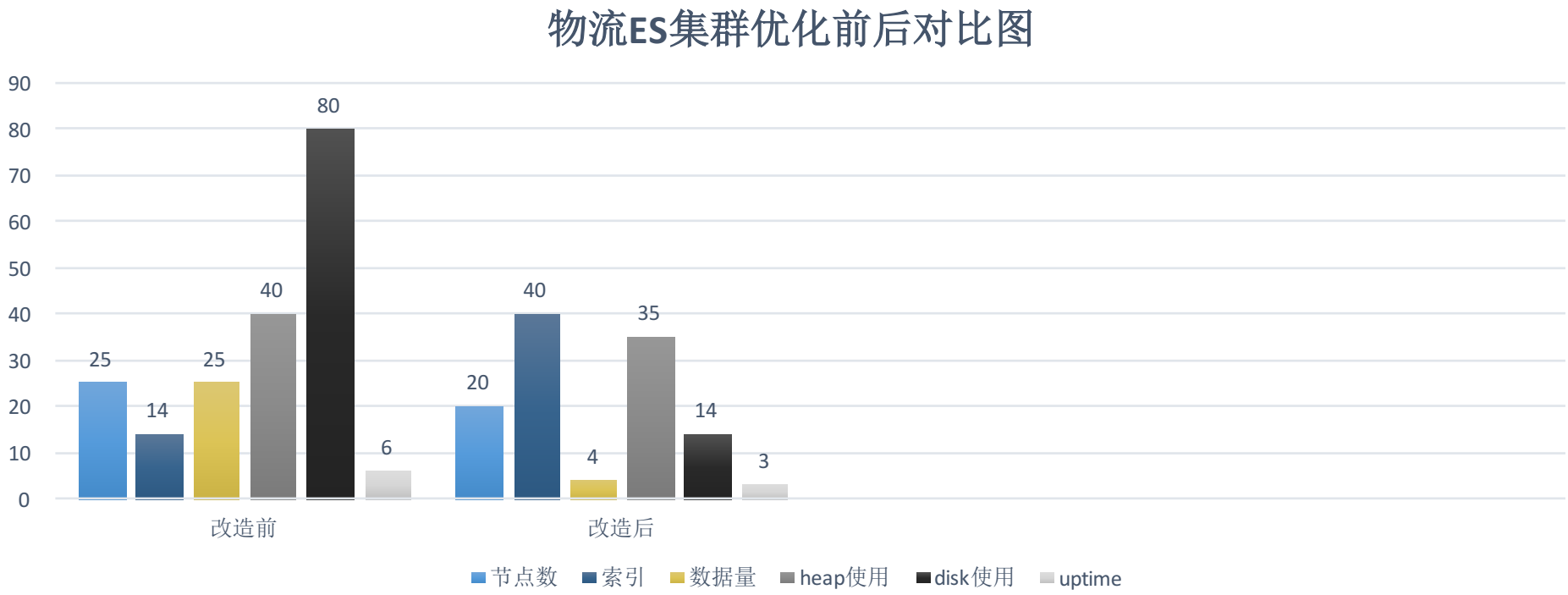
Before


cluster
nodes
rest
more
XG-LPD-APOLLO-ES-SVC-1 @ xg1-lpd-apollo-es-mgr-1
25 nodes
14 indices
448 shards
13,323,677,519 docs 5,754
24.76TB 232.95MB
filter nodes by name
master
data
client
name
load average
cpu %
heap usage %
disk usage %
uptime
xg1-lpd-apollo-es-data-1
10.22.32.4
10.22.32.4:9300
JVM: 1.8.0_45 ES: 2.3.3
0.6
0.0
45.0 used: 13.09GB max: 28.67GB
79.0 free: 303.86GB total: 1.43TB
6mo.
xg1-lpd-apollo-es-data-10
10.21.248.29
10.21.248.29:9300
JVM: 1.8.0_45 ES: 2.3.3
1.0
0.0
44.0 used: 12.71GB max: 28.67GB
80.0 free: 286.55GB total: 1.43TB
6mo.
xg1-lpd-apollo-es-data-11
10.22.16.28
10.22.16.28:9300
JVM: 1.8.0_45 ES: 2.3.3
0.4
0.0
62.0 used: 17.79GB max: 28.67GB
73.0 free: 391.89GB total: 1.43TB
6mo.
xg1-lpd-apollo-es-data-12
10.21.208.38
10.21.208.38:9300
JVM: 1.8.0_45 ES: 2.3.3
0.8
0.0
63.0 used: 18.25GB max: 28.67GB
78.0 free: 318.63GB total: 1.43TB
6mo.
xg1-lpd-apollo-es-data-13
10.18.114.52
10.18.114.52:9300
JVM: 1.8.0_45 ES: 2.3.3
1.2
0.0
25.0 used: 7.35GB max: 28.67GB
77.0 free: 338.81GB total: 1.43TB
6mo.
xg1-lpd-apollo-es-data-14
10.18.136.18
10.18.136.18:9300
JVM: 1.8.0_45 ES: 2.3.3
0.5
0.0
37.0 used: 10.70GB max: 28.67GB
78.0 free: 315.61GB total: 1.43TB
6mo.
xg1-lpd-apollo-es-data-15
10.21.192.30
10.21.192.30:9300
JVM: 1.8.0_45 ES: 2.3.3
1.2
0.0
44.0 used: 12.68GB max: 28.67GB
82.0 free: 266.76GB total: 1.43TB
6mo.

After


cluster
nodes
rest
more
XG-LPD-APOLLOCALC-ES-SVC-1 @ xg-lpd-apollocalc-es-mgmt-1
20 nodes
411 indices
8,570 shards
1,584,518,202 docs ↑ 26
3.71TB ↑ 2.16MB
filter nodes by name
master
data
client
name
load average
cpu %
heap usage %
disk usage %
uptime
xg-lpd-apollocalc-es-data-1
10.20.96.28
10.20.96.28:9300
JVM: 1.8.0_45 ES: 2.3.3
0.6
0.0
27.0 used: 8.04GB max: 29.00GB
14.0 free: 1.39TB total: 1.61TB
3mo.
xg-lpd-apollocalc-es-data-10
10.20.64.3
10.20.64.3:9300
JVM: 1.8.0_45 ES: 2.3.3
0.2
0.0
60.0 used: 17.66GB max: 29.00GB
14.0 free: 1.38TB total: 1.61TB
3mo.
xg-lpd-apollocalc-es-data-11
10.22.240.3
10.22.240.3:9300
JVM: 1.8.0_45 ES: 2.3.3
0.4
0.0
68.0 used: 19.84GB max: 29.00GB
14.0 free: 1.39TB total: 1.61TB
3mo.
xg-lpd-apollocalc-es-data-13
10.22.240.22
10.22.240.22:9300
JVM: 1.8.0_45 ES: 2.3.3
0.6
12.0
48.0 used: 13.94GB max: 29.00GB
14.0 free: 1.38TB total: 1.61TB
3mo.
xg-lpd-apollocalc-es-data-14
10.18.6.39
10.18.6.39:9300
JVM: 1.8.0_45 ES: 2.3.3
0.1
0.0
61.0 used: 17.75GB max: 29.00GB
12.0 free: 1.41TB total: 1.61TB
3mo.
xg-lpd-apollocalc-es-data-16
10.22.224.23
10.22.224.23:9300
JVM: 1.8.0_45 ES: 2.3.3
0.3
0.0
48.0 used: 14.01GB max: 29.00GB
13.0 free: 1.40TB total: 1.61TB
3mo.
xg-lpd-apollocalc-es-data-17
10.18.32.24
10.18.32.24:9300
JVM: 1.8.0_45 ES: 2.3.3
0.1
0.0
30.0 used: 8.72GB max: 29.00GB
14.0 free: 1.39TB total: 1.61TB
3mo.

饿了么 | 改造vs

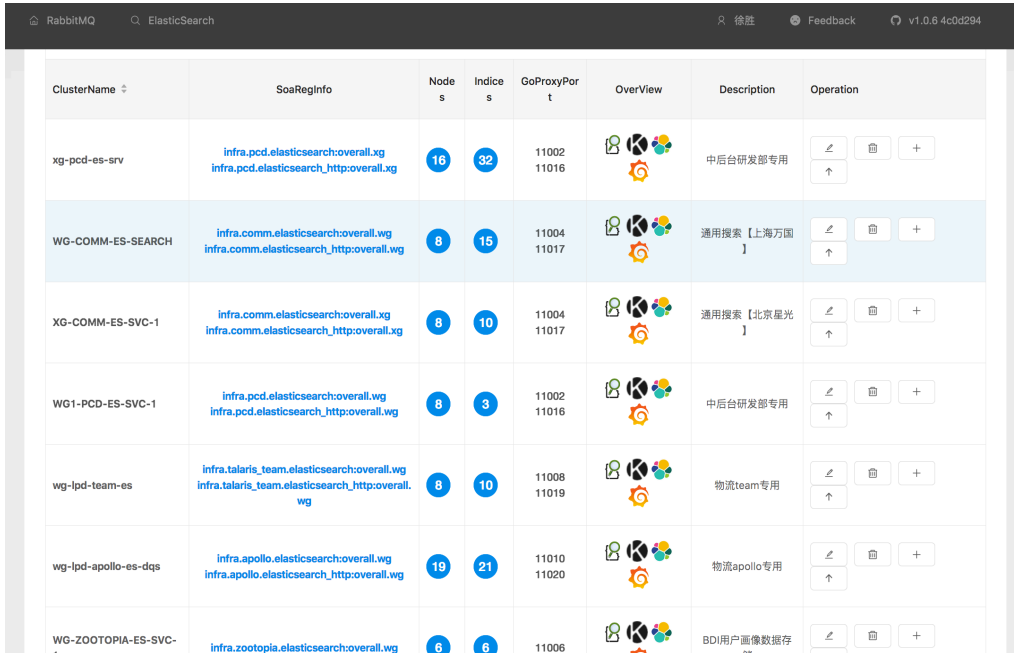


物流优化情况

ES自动化运维平台和监控平台建设

实现的功能特点：

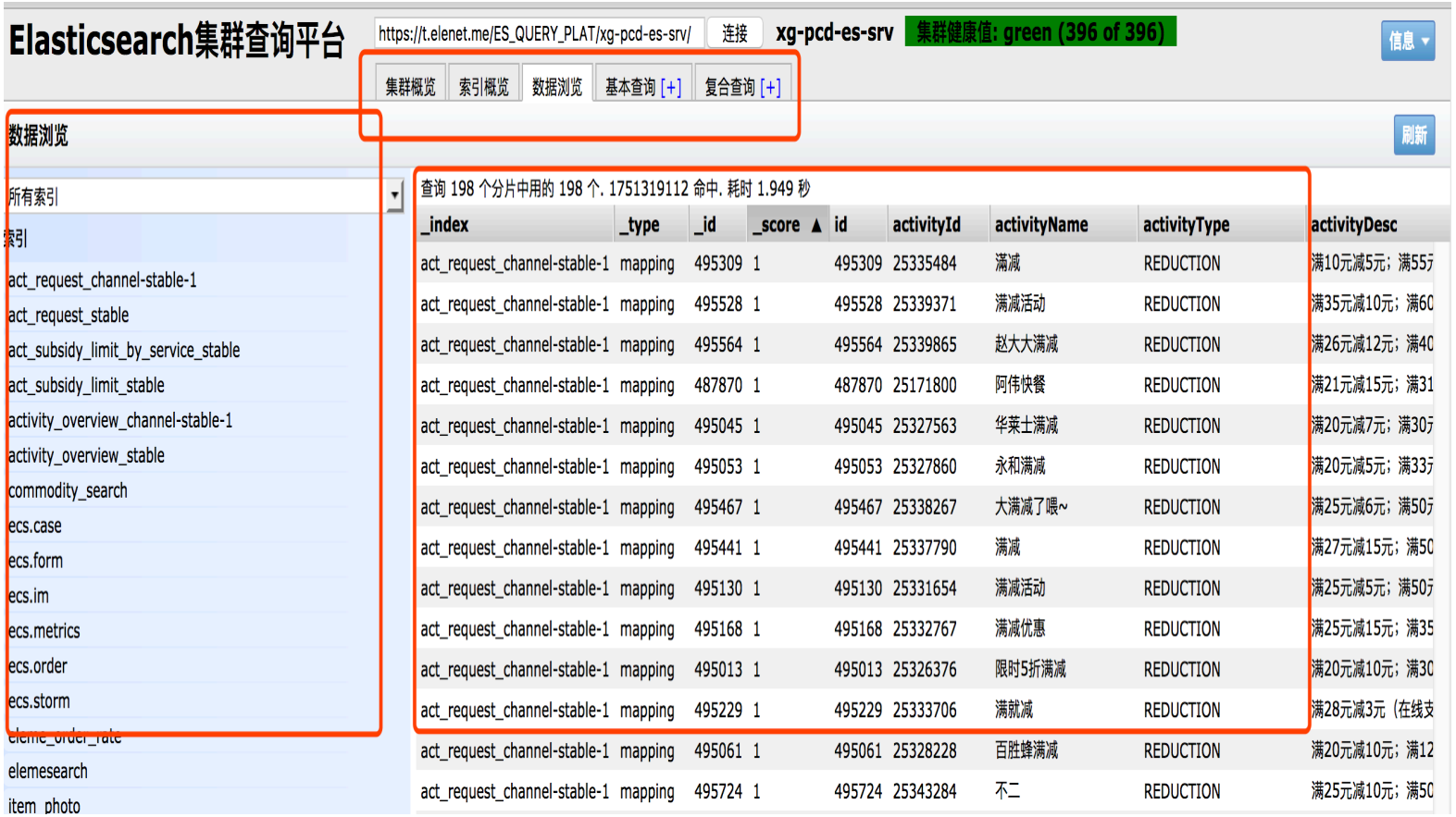
- 自动化扩容、缩容，增加节点整合到一个平台
- 增加查询平台插件
- 查看集群的监控信息
- 集群节点查看
- 集群端口连接数据



ES查询平台建设

解决业务的痛点

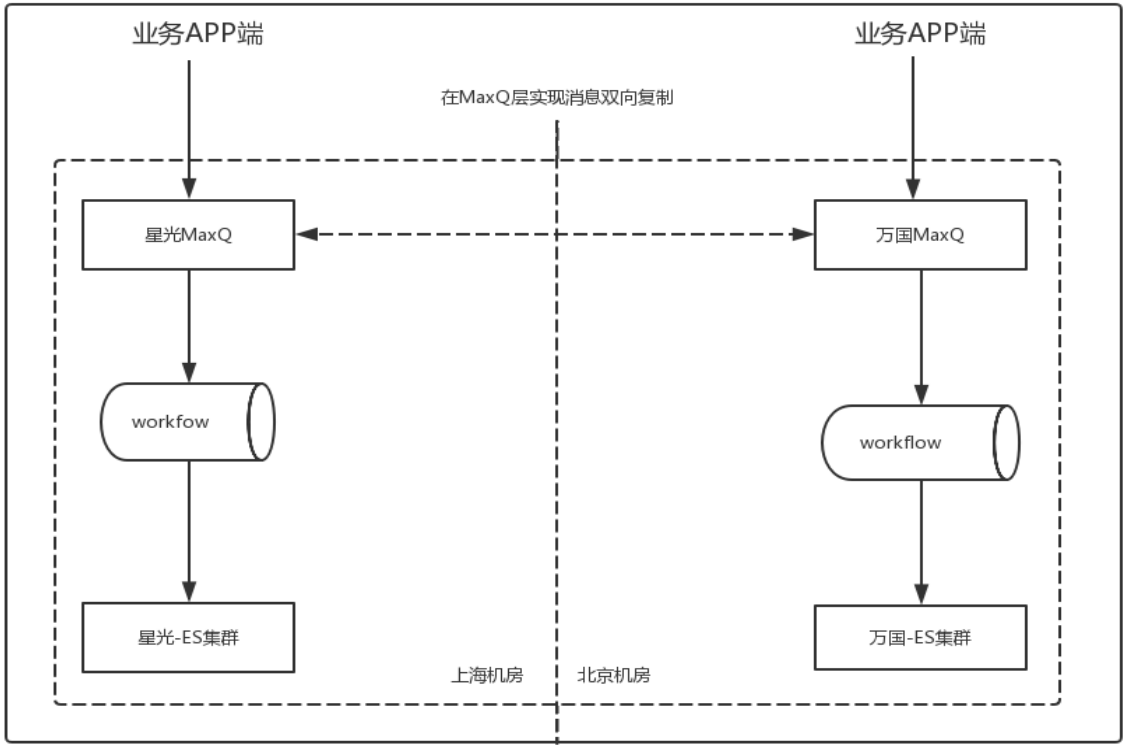
- 集群浏览
- 索引概览
- 数据浏览
- 基本查询
- 复合查询



平台建设

提供多活的设计方案：通过MQ消息复制

- ES多活方案
- 协调各方资源，进行ES workFlow、MQ的整合

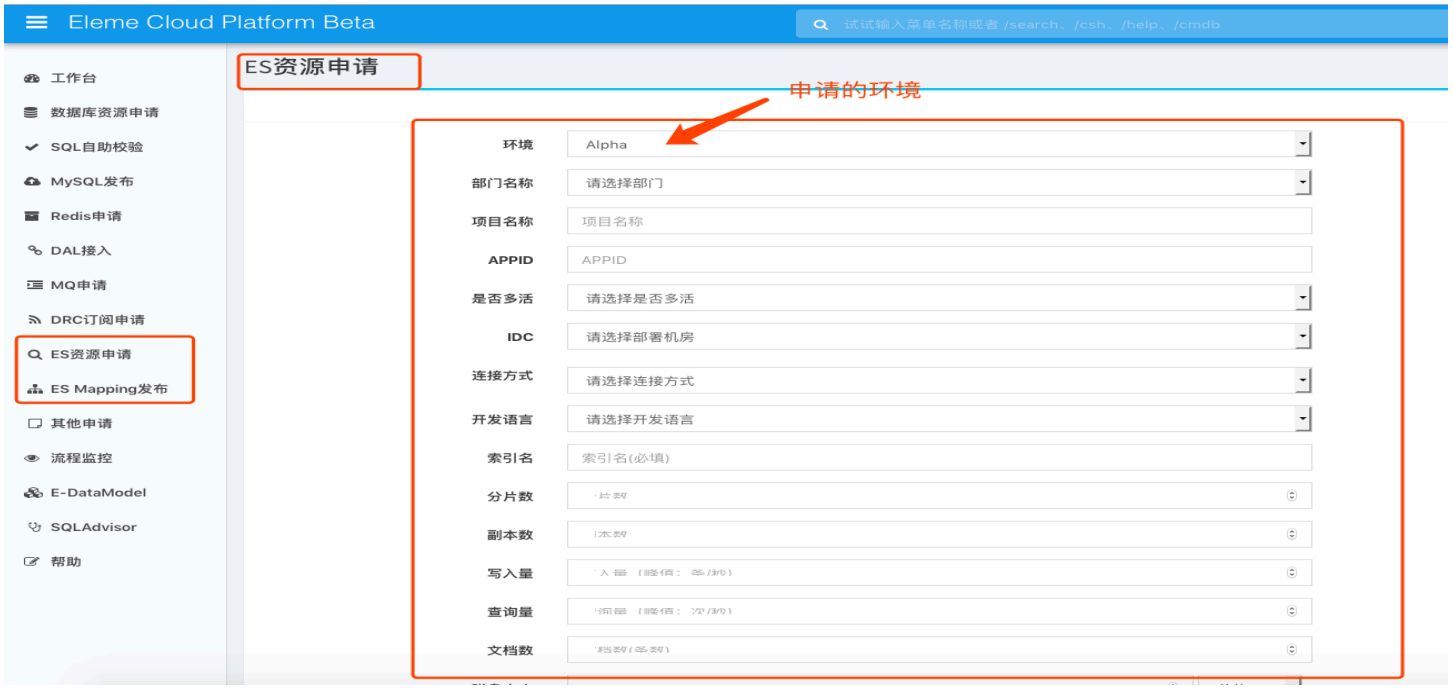


平台建设

Eden资源申请规范化

推进申请规范化

- 增加ES资源申请
- 增加Mapping审核

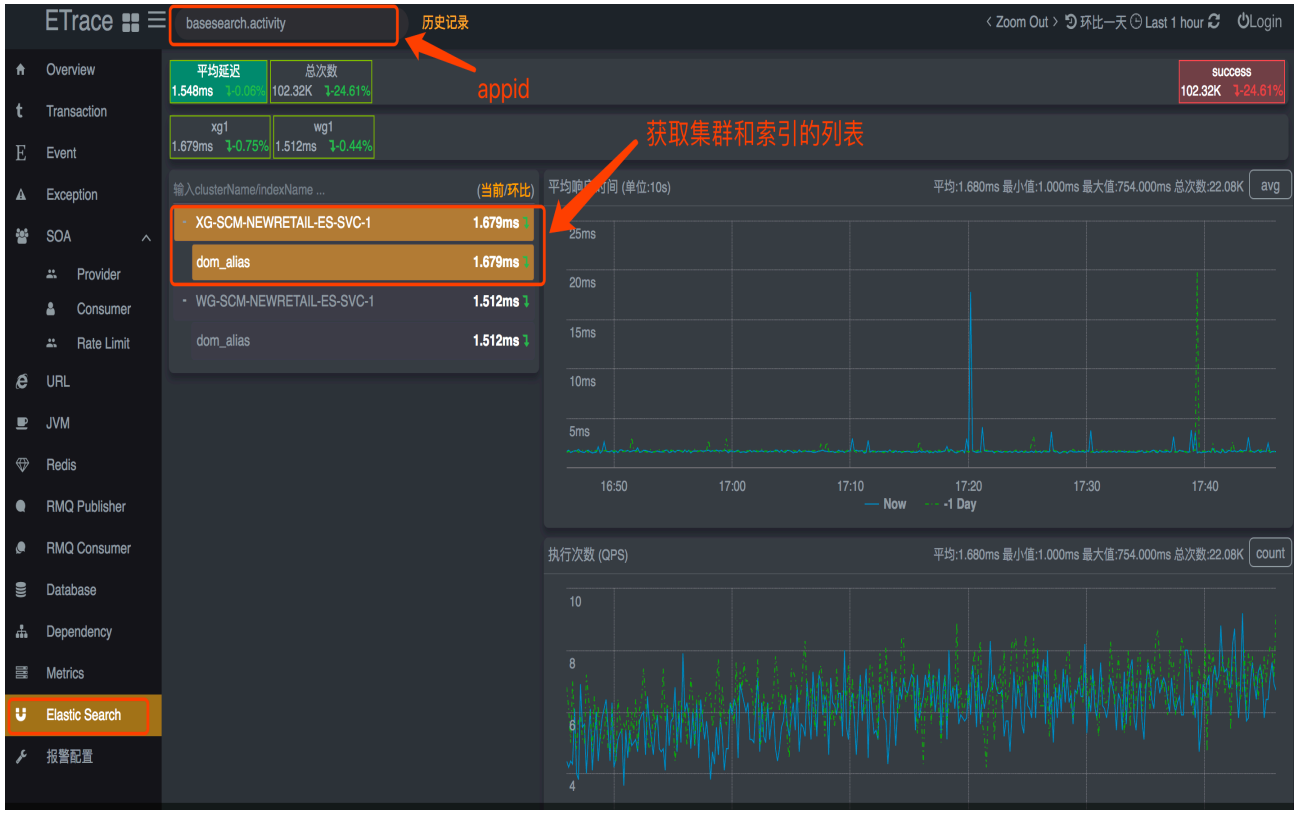


平台建设

提供通过SDK搜索服务，集成到Trace中

解决的痛点

- Elasticsearch集成到Trace中
- 提供了一个通用的SOA服务
- 帮助业务进行增删改查
- 能够定位到集群、索引、Query信息



ES作为基础设施已经越来越重要了，**2018年，面向C端、B端、物流端、商户端**，计划从三个方面来规划：

- ✓ 自动化运维部署平台
- ✓ 通过优化索引和集群配置来降低成本
- ✓ 提供通用SOA服务，解决业务部门的共性问题

2018年年度规划

欢迎关注饿科技

饿科技



We are hiring

饿了么

感谢聆听



徐胜

2018.2.26