



# 字节跳动的ES智能运维和数据管理中台演进

---

高英举

12/2019, 广告数据平台, 字节跳动



# Agenda

- ES运维和用户服务的痛点
- Datapalace 数据管理中台介绍
- ES智能诊断系统介绍
- 数据管理中台和智能诊断系统的设计与实现
- 未来计划

# ES运维和用户服务的痛点

特定(集群、数据)规模有特定问题

# ES运维和用户服务的痛点

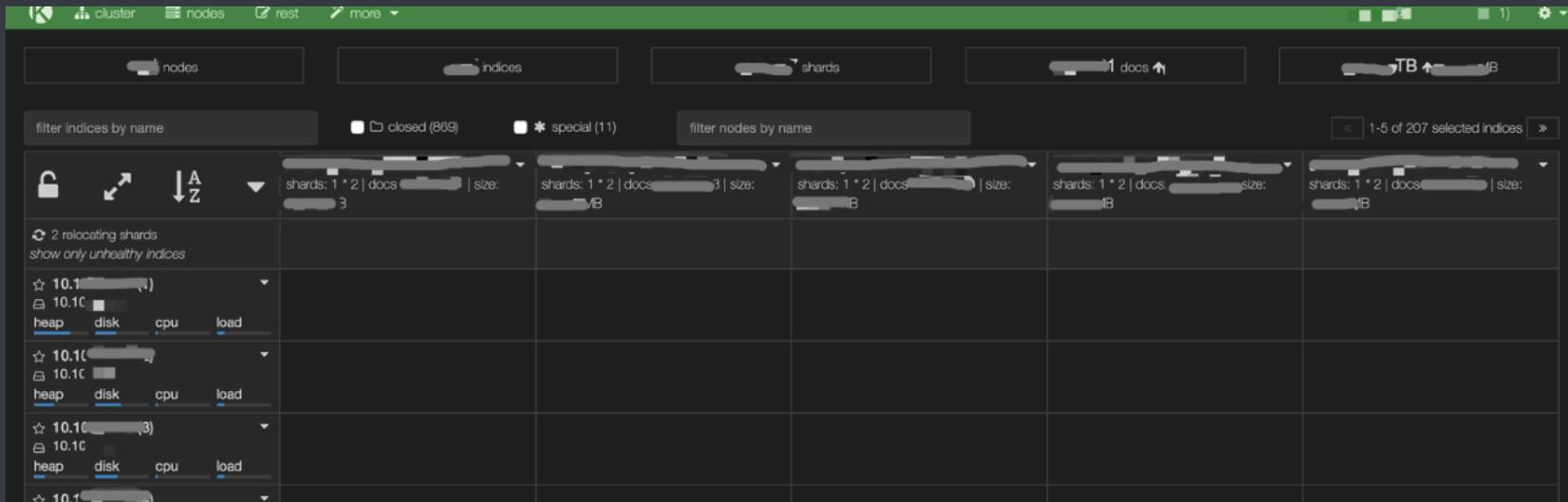
## 痛点列举

- 开源集群、索引管理工具
  - 缺少对公司业务的感知和数据治理能力
  - 缺乏与公司基础设施的联动（如部署、监控、报警）
- ES集群、索引的运维
  - 操作的安全性待优化（加个密码访问就够吗？）
  - 缺乏标准化手段，流程不健全，运维工程师不自觉
- 线上问题排查
  - 较多依赖手动操作和人工经验判断
  - 解决问题的最佳实践没有固化

总结：特定(集群、数据)规模有特定问题，不一定适用所有业务场景

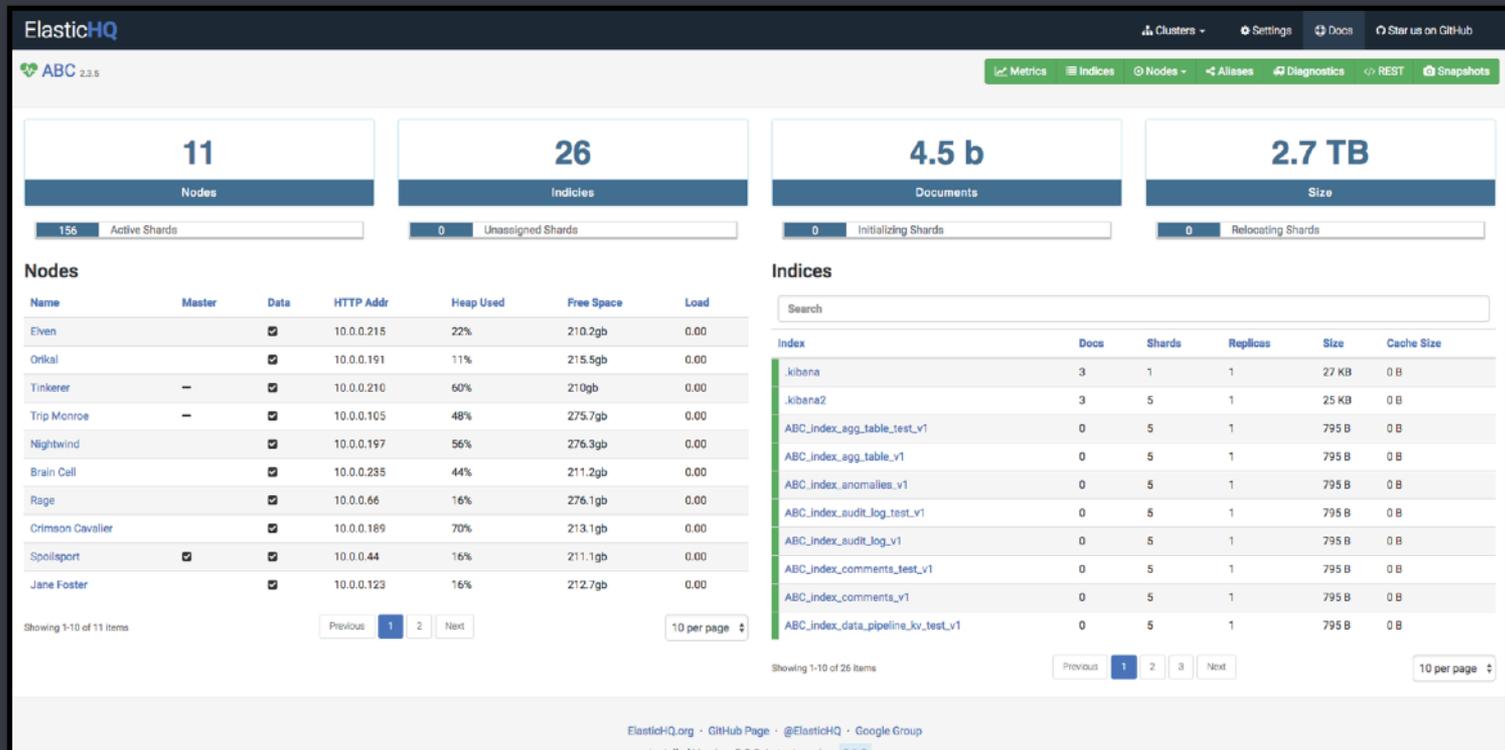
# ES运维和用户服务的痛点

Kopf(Cerebro): <https://github.com/lmenezes/cerebro>



# ES运维和用户服务的痛点

ElasticHQ: <https://github.com/ElasticHQ/elasticsearch-HQ>



# ES运维和用户服务的痛点

## Kibana Full Stack Monitoring

The screenshot displays the Kibana Clusters monitoring interface. At the top, it shows the cluster ID '139c969c3e06434ba3a1725c1029fce0' and the time range 'This month'. Below this, the 'Elasticsearch' section is highlighted, showing a green health status and a warning that the 'Platinum license will expire on April 30, 2020'. The Elasticsearch overview includes three panels: 'Overview' (Version 7.0.0, Uptime 8 days, Jobs 0), 'Nodes: 4' (Disk Available 62.02%, JVM Heap 29.56%), and 'Indices: 67' (Documents 1,369,095,642, Disk Usage 1.4 TB, Primary Shards 103, Replica Shards 100). The 'Kibana' section below shows a green health status and an overview panel with 'Requests 6' and 'Max. Response Time 90 ms'. A second panel for Kibana shows 'Instances: 1' with 'Connections 0' and 'Memory Usage 8.22%'. A vertical navigation sidebar is visible on the left side of the interface.

Component	Health	Overview Metrics	Nodes/Instances	Other Metrics
Elasticsearch	Health is green	Version: 7.0.0 Uptime: 8 days Jobs: 0	Nodes: 4	Disk Available: 62.02% (2.3 TB / 3.8 TB) JVM Heap: 29.56% (19.4 GB / 65.5 GB) Indices: 67 Documents: 1,369,095,642 Disk Usage: 1.4 TB Primary Shards: 103 Replica Shards: 100
Kibana	Health is green	Requests: 6 Max. Response Time: 90 ms	Instances: 1	Connections: 0 Memory Usage: 8.22% (270.1 MB / 3.2 GB)

# Datapalace 数据管理中台

公司级一站式ES集群、索引托管和数据治理平台

# Datapalace 数据管理中台

## 功能介绍

### 1. ES集群功能:

- 创建集群
- 更新集群元数据
- 申请集群运维变更
- 查看集群数据链路
- 查看集群存储容量统计
- 查看节点加入退出记录
- 查看Master切换记录
- 查看集群的诊断信息
- 下线、删除集群
- 自动绑定计费服务树

### 2. ES集群组功能:

- 关联集群组
- 集群组扩容, 缩容
- 集群组上索引的操作

### 3. ES索引功能:

- 接入索引[集群、集群组]
- 索引数据读写的交互式文档
- 更新索引元数据
- 更新索引settings (支持历史索引的更新)
- 更新索引mappings (支持历史索引的更新)
- 历史索引的查看、打开、关闭、删除
- 使用别名管理索引, 别名切换
- 查看索引存储容量统计
- 查看索引数据链路
- 查看索引的诊断信息
- 下线、删除索引 (报警联动)

### 4. 数据导入导出

- 自助式数据导入支持: Spark, Flink, Hive2ES, Logstash
- 自助式数据导出支持: Spark, Flink
- 字节云, 物理机自动部署 Logstash
- 自动绑定计费服务树

# Datapalace 数据管理中台

## 索引接入

① 选择部门和集群      ② 定义数据源      ③ index元数据      ④ 确认信息

选择集群或集群组:

集群业务负责人:

选择集群		
[redacted]	2.3.3	[redacted]
[redacted]	5.2.2	[redacted]
[redacted]	5.2.2	[redacted]
[redacted]	5.4.3	[redacted]

### 索引概念梳理:

- 单集群索引(组)
- 集群组索引(组)
- 索引(组)的物理索引

# Datapalace 数据管理中台

## 索引接入

选择部门 and 集群 ① 定义数据源 ② index元数据 ③ 确认信息 ④

选择数据源:

选择Kafka集群:

Kafka集群Topic:

示例数据: 

```
{
  "value":3,
  "sid": "n4nnci8y53mme"
}
```

脱敏数据

主键字段名称:

Event Time/数据时间字段名:

Event Time字段格式:

定义数据源的目的:

- 自动建立数据链路
- 与数据导入导出工具联动
- 自动推测索引Mappings

# Datapalace 数据管理中台

## 索引接入

选择部门和集群  定义数据源  **index元数据**  确认信

索引名称:

索引创建者&索引业务负责人:

划分方式:  按日  按月  按年  按版本  不划分(不推荐)

保留天数:  天

打开天数:  天

单个划分Doc个数:

单个Doc大小:  Bytes

写入QPS:

查询QPS:

使用别名(Alias)管理Index:  是 (推荐)  不是 (不推荐)

索引存储容量占整个集群Quota:  %

是否支持字段动态增加:  不支持 (推荐)  支持 (不推荐)

Mappings

```
{
  "dynamic":false,
  "properties":{
    "field_one":{
      "type":"date",
      "format":"yyyy-MM-dd HH:mm:ss"
    },
    "field_two":{
      "index":false,
      "type":"long"
    },
    "field_three":{
```

校验mappings

脱敏数据

# Datapalace 数据管理中台

## 索引接入

工单状态

cg\_data\_es集群业务负责人审批( ) 审批通过

日常运维负责人审批 ( ) 审批通过

已完成

工单名称: 创建索引

工单内容:

— 部门和集群 —

集群名称: 集群业务负责人:

工单创建时间: 2019-11-19 15:54:53

— 数据源 —

数据源:

Index元数据

索引名称: 索引创建者&索引业务负责人:

划分方式: 按月

业务负责人、运维负责人参与索引接入的审批，确保业务和运维上的合理性

# Datapalace 数据管理中台

## 物理索引操作

物理索引

批量(open, close)索引   索引别名切换   批量删除索引

索引名称	索引别名	状态	Shared个数, 副本个数	文档数	大小	操作
agent-transaction-log-daily-2010-11-10	agent-transaction-log-daily-2010-11-10	open	1, 1	1	(Bytes)	操作
agent-transaction-log-daily-2010-11-10	agent-transaction-log-daily-2010-11-10	open	1, 1	1	(Bytes)	操作
agent-transaction-log-daily-2010-11-10	agent-transaction-log-daily-2010-11-10	open	1, 1	44		操作
agent-transaction-log-daily-2010-11-10	agent-transaction-log-daily-2010-11-10	open	1, 1	48		操作
agent-transaction-log-daily-2010-11-10	agent-transaction-log-daily-2010-11-10	open	1, 1		(Bytes)	操作
agent-transaction-log-daily-2010-11-10	agent-transaction-log-daily-2010-11-10	open	1, 1	1	(Bytes)	操作

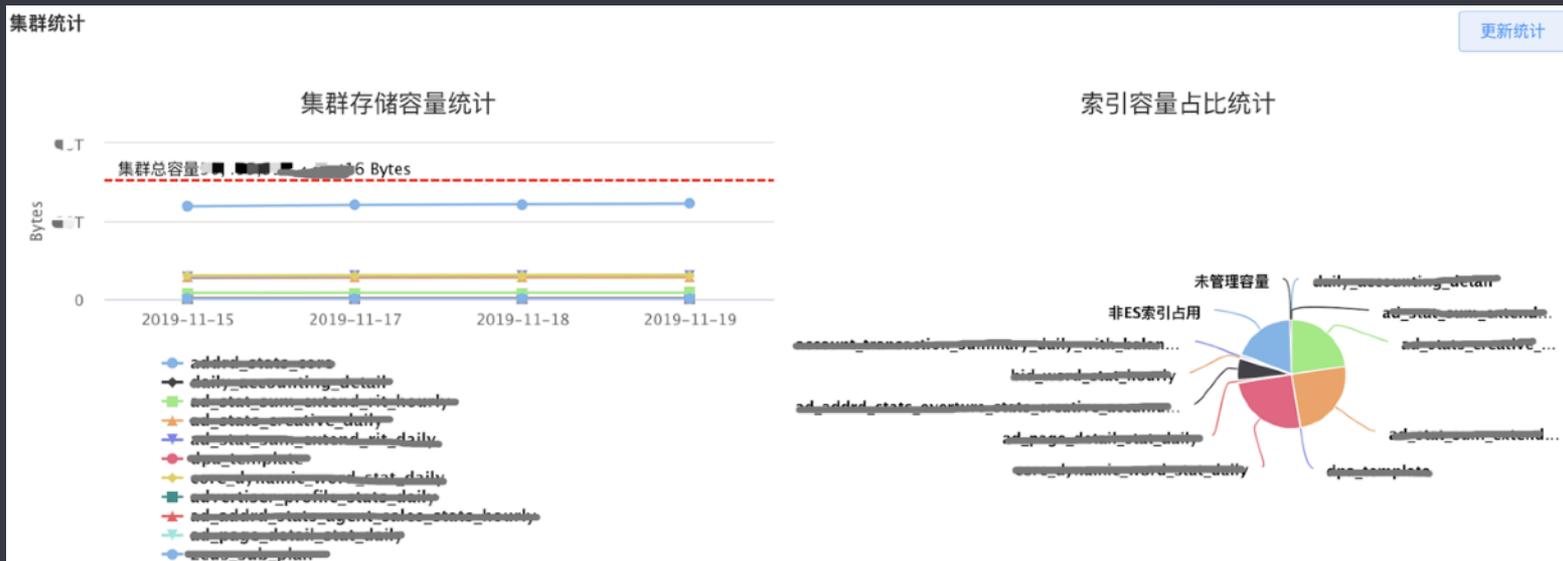
操作菜单:

- 查看Mappings
- 查看Settings
- 查看statsUri
- Alias切换: 第一步创建新版本的物理索引
- Alias切换: 第二步Alias切换

- 批量open, close索引
- 批量删除索引
- 批量切换索引的别名
- 查看索引的别名
- 查看索引的mappings, settings, stats

# Datapalace 数据管理中台

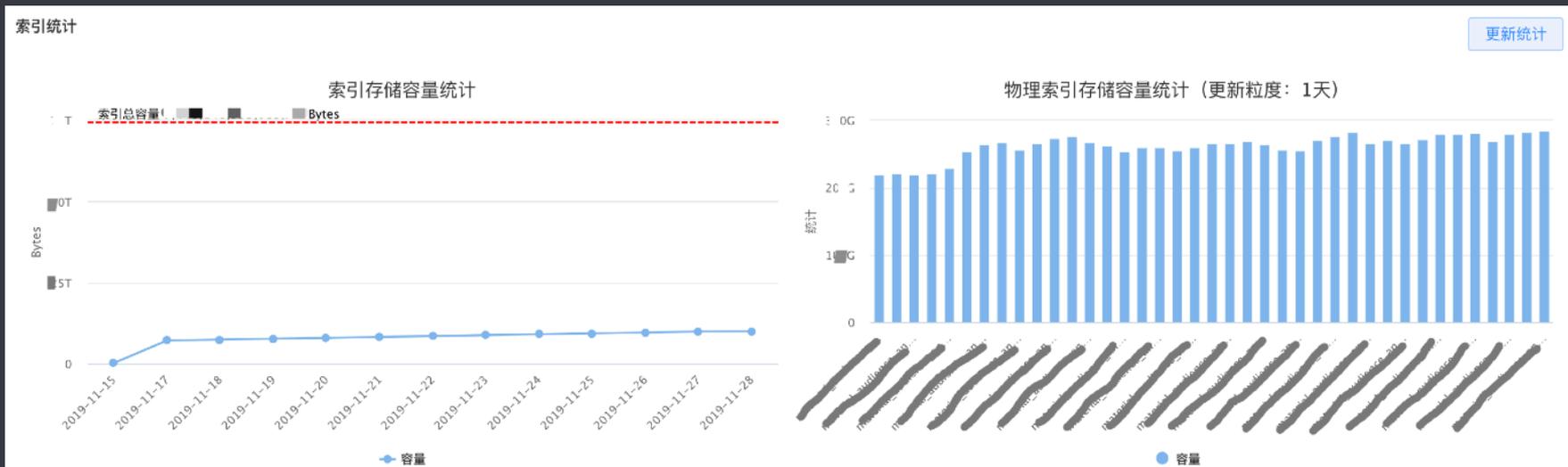
## 集群、索引容量统计：做好容量规划



索引(组)为单位的容量统计

# Datapalace 数据管理中台

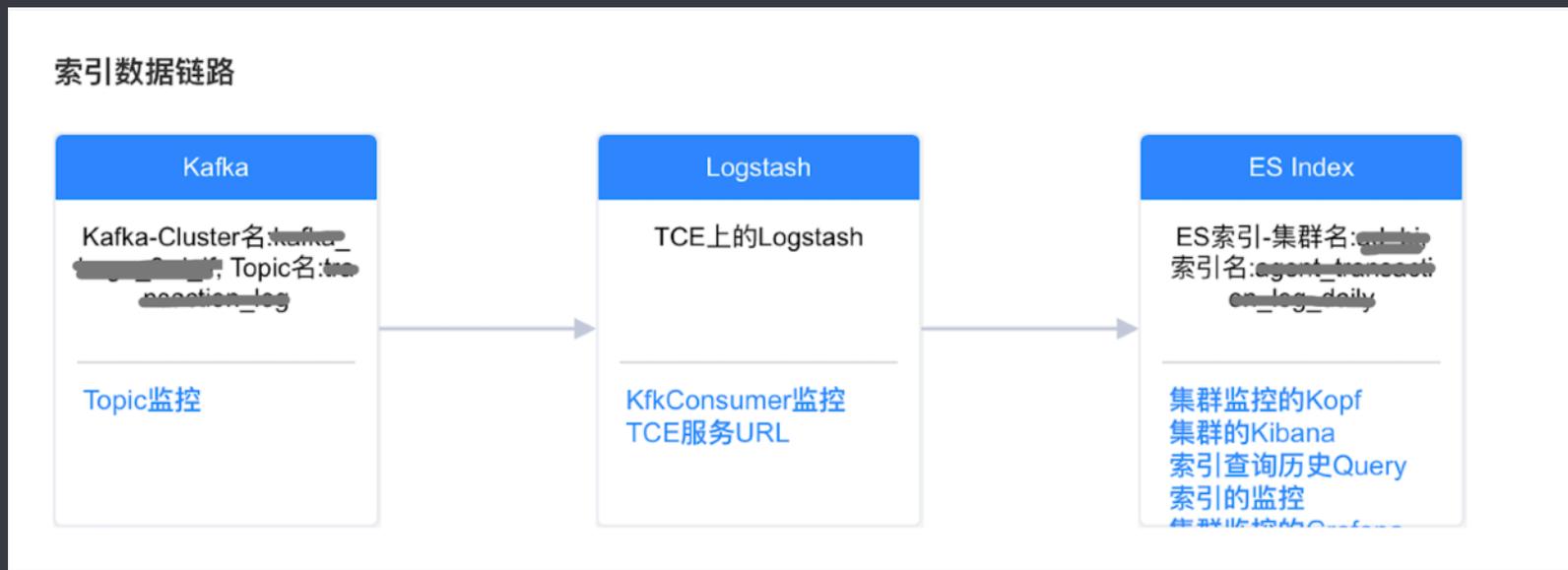
## 集群、索引容量统计：做好容量规划



物理索引为单位的统计

# Datapalace 数据管理中台

索引的数据链路：大大简化了业务和运维负责人管理和学习数据的成本



数据链路中的任意节点提供：监控页面入口，报警管理入口

# Datapalace 数据管理中台

索引的数据链路：大大简化了业务和运维负责人管理和学习数据的成本



多机房容灾场景下的数据链路

# Datapalace 数据管理中台

## ES集群组：多机房容灾的支持

[新建集群组](#)

ID	集群组名称	集群列表	集群组创建者	业务负责人	索引Setting同步	索引Mapping同步	索引数据同步	操作
1	[REDACTED]	[REDACTED] [REDACTED] [REDACTED]	[REDACTED]	[REDACTED]	是	是	否	<a href="#">查看集群组</a>

共 1 条 10条/页 < 1 > 前往 1 页

集群组列表

索引列表

ID	索引名称	集群入口	划分	索引创建者	业务负责人	运维负责人	索引状态	诊断信息	操作
10005	[REDACTED]	[REDACTED]	按日	[REDACTED]	[REDACTED]	[REDACTED]	在线		<a href="#">更新索引元数据</a> <a href="#">更新索引Mapping</a>
10006		[REDACTED]	按日	[REDACTED]	[REDACTED]	[REDACTED]	在线		
10007		[REDACTED]	按日	[REDACTED]	[REDACTED]	[REDACTED]	在线		
10008		[REDACTED]	按日	[REDACTED]	[REDACTED]	[REDACTED]	在线		

集群组中的索引



# Datapalace 数据管理中台

## 集群运维变更工单

变更标题: [redacted] 集群扩容三台机器

运维操作人: [redacted]

影响集群: [redacted]

影响索引: 请选择

变更开始时间: 2019-11-20 12:00:00

变更结束时间: 2019-11-20 14:00:00

变更类型: **缩容扩容** 问题排查 下线问题节点 其他

变更步骤:

1. 申请好三台扩容机器
2. 将三台机器挂服务树
3. 给三台机器打上tag
4. 在elasticsearch\_deploy工程中添加相关配置
5. push elasticsearch\_deploy中的代码, code review后merge, 触发SCM编译

风险: 添加机器后会触发集群的rebalance, 此期间会导致业务读写不稳定

工单状态

① 运维团队负责人审批 ( [redacted] ) 审批通过

② ad\_online\_hi集群业务负责人审批 ( [redacted] )

③ 完成

工单名称: [redacted] 集群扩容三台机器

工单内容:

变更标题: [redacted] 集群扩容三台机器

运维操作人: [redacted]

影响集群: [redacted]

影响索引:

变更开始时间: 2019-11-20 12:00:00

变更结束时间: 2019-11-20 14:00:00

把“线上集群运维变更”这件事程序化、标准化, 以达到运维效率、人力成本、线上集群稳定性的平衡

# Datapalace 数据管理中台

数据导入导出, 多源数据关联分析

数据源:

- Kafka
- Hive
- HDFS
- MySQL
- ES



导入导出工具:

- Dorado[公司平台]
- Waterdrop[开源项目]



数据源:

- Kafka
- Hive
- HDFS
- MySQL
- ES

功能:

- 自助式数据导入支持: Spark, Flink, Logstash
- 自助式数据导出支持: Spark, Flink

数据治理:

- 索引|docs count校验
- 索引|mappings校验
- 数值型字段min, max, avg, count校验
- 业务自定义SQL校验

# Datapalace 数据管理中台

数据导入导出，多源数据关联分析

	Logstash	Spark	Flink
使用和部署成本	低	高	高
开发成本	低	高	高
海量、高吞吐的数据计算	不支持	支持	支持
多源数据关联分析	不支持	支持	支持
使用SQL表达计算逻辑	不支持	支持	支持
有状态的流式计算	不支持	做得不好	非常擅长
表达复杂的计算逻辑	不好做	支持	支持
擅长的领域	小数据量传输，无状态 简单计算	海量数据，无状态 复杂计算	海量数据，有状态 复杂计算

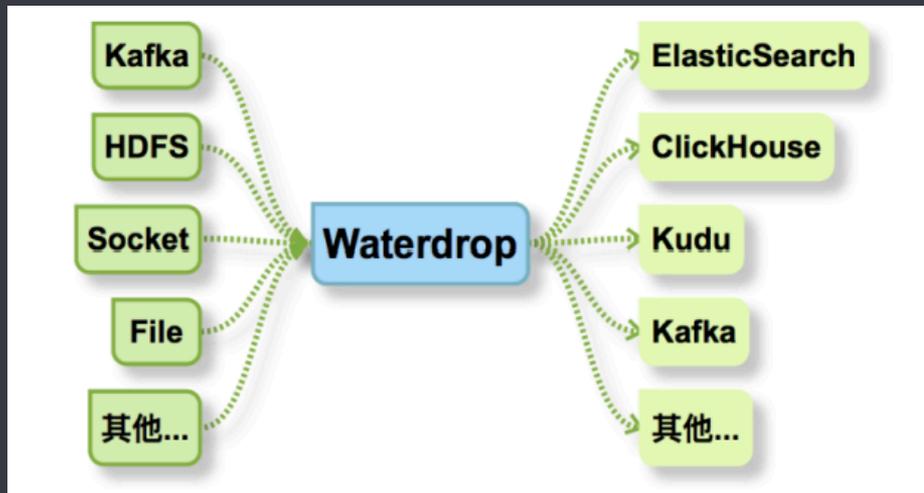
# Datapalace 数据管理中台

## 数据导入导出, 多源数据关联分析工具: Waterdrop

Waterdrop 是一个非常易用, 高性能、支持实时流式和离线批处理的海量数据处理产品, 架构于Apache Spark之上。(支持Flink的版本近期发布)

### Waterdrop 的特性

- 简单易用, 灵活配置, 无需开发
- 实时流式处理
- 离线多源数据分析
- 高性能
- 海量数据处理能力
- 模块化和插件化, 易于扩展
- 支持利用SQL做数据处理和聚合
- 支持Spark Structured Streaming
- 支持Spark 2.x



# ES智能诊断系统

把问题排查手段、实践经验的复用标准化

# ES 智能诊断系统

## 概要介绍

- 支持14种巡检诊断项目
- 支持自动优化
- 诊断结果包含详细的数据披露和明确、可量化的优化建议
- 完全插件化的巡检、诊断、优化的开发
- 支持对接公司MS报警平台实现报警自愈
- 支持Query Profile 采样存储分析
- 使用数学、统计模型来简化问题排查难度

# ES 智能诊断系统

已支持的诊断项列表 (可自动优化比例 = 8/14)

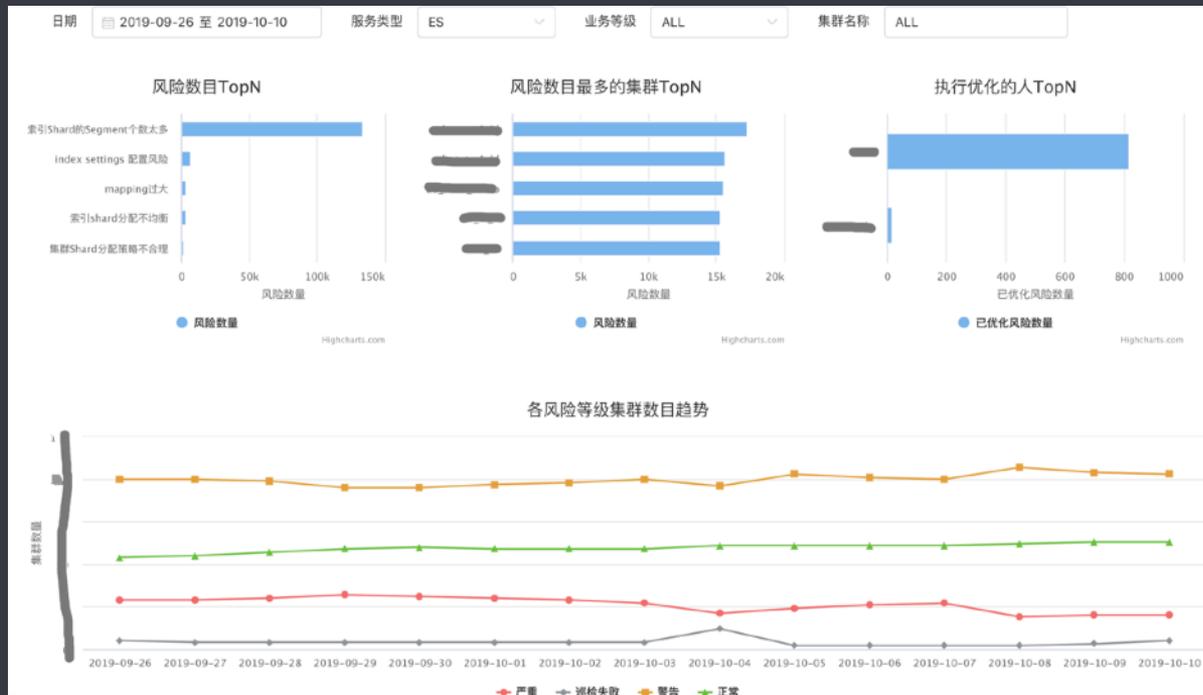
- 集群存储容量不足风险
- 集群Settings配置风险[可自动优化]
- 集群Shard分配策略不合理[可自动优化]
- 集群Shard个数分配不均衡
- 集群节点磁盘占用不均衡
- 集群Index Template Order 校验不合规
- 索引Template 配置风险[可自动优化]
- 索引settings 配置风险[可自动优化]
- 索引Mapping过大
- 索引shard分配不均衡[可自动优化]
- 时序型索引Shard size不合理[可自动优化]
- 非时序型索引Shard size不合理
- 索引Shard的Segment个数太多[可自动优化]
- 索引Shard数据损坏无法配[可自动优化]

# ES 智能诊断系统

## 概要介绍

目标:

- 量化核心KPI
- 明确优化方向
- 快速定位问题



# ES 智能诊断系统

## 概要介绍

特点:

- 诊断项目明确
- 建议有细节, 可量化
- 多种优化方式, 支持批量操作
- 数据披露详尽
- 监控信息入口

Data Palace > 智能诊断 > 诊断详情

### 集群诊断报告

诊断时间: 2019-11-30 17:01:31  
运行状态: 已完成

集群名称: [redacted] / 1  
异常级别: 严重 2  
服务类型: es  
3 [重新巡检]

4 [查看集群信息] [查看Kopf] [查看Grafana]

- 1 时序型索引Shard size不合理 [自动优化] [手动优化] [忽略] >
- 1 集群Shard分配策略不合理 [自动优化] [手动优化] [忽略] >
- 1 索引shard分配不均衡 [自动优化] [手动优化] [忽略] >
- 2 集群存储容量不足风险 [自动优化] [手动优化] [忽略] >

5

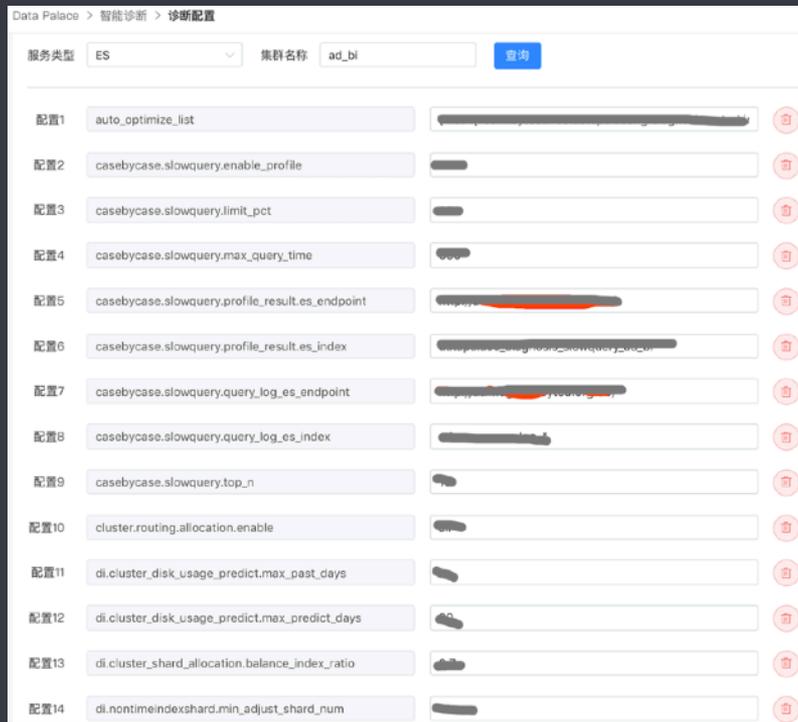
<input type="checkbox"/>	ID	数据源	数据源类型	数据源等级	数据源组	风险等级	数据详情	建议	优化操作
<input type="checkbox"/>	3946471	[redacted]	cluster	[redacted]	[redacted]	严重	[redacted]	查看详情	[自动优化] [手动优化] [忽略]

7

据此集群过去 7 天的日写入数据量预估每日写入量为 [redacted] GB/天。预计集群磁盘会在 3 天后容量不足, 无法分配索引Shard, 请尽快 (1)扩容 (2)减少数据写入 (3) 删除历史数据。(当前数据: 磁盘总大小 [redacted] GB, 已使用 [redacted] GB, 占比 66 %, Watermark: 90 %)

# ES 智能诊断系统

## 概要介绍



集群、索引级别的诊断配置

# ES 智能诊断系统

诊断项：索引Shard的Segment个数太多[可自动优化]

### 集群诊断报告

重新巡检

诊断时间：2019-11-29 09:01:06      集群名称：[REDACTED]      服务类型：es

运行状态：已完成      异常级别：警告

[查看集群信息](#)   [查看Kopf](#)   [查看Grafana](#)

---

索引Shard的Segment个数太多 自动优化 手动优化 忽略

ID	数据源	数据源类型	数据源等级	数据源组	风险等级	数据详情	建议	优化操作
3903866	[REDACTED]	index	Unknown	[REDACTED]	警告	查看详情	此索引Segment个数太多，平均Shard Size = [REDACTED]MB，Segment最多的Shard中有5个Segment。预计merge以后单个Shard中Segment个数 = 2。建议通过force merge(参考： <a href="https://www.elastic.co/guide/en/elasticsearch/reference/current/indexes-forcemerge.html">https://www.elastic.co/guide/en/elasticsearch/reference/current/indexes-forcemerge.html</a> )的方式做merge。注意：还在写数据的索引，不要执行 force merge。	自动优化 手动优化 忽略
3903869	[REDACTED]	index	Unknown	[REDACTED]	警告	查看详情	此索引Segment个数太多，平均Shard Size = [REDACTED]MB，Segment最多的Shard中有4个Segment。预计merge以后单个Shard中Segment个数 = 2。建议通过force merge(参考： <a href="https://www.elastic.co/guide/en/elasticsearch/reference/current/indexes-forcemerge.html">https://www.elastic.co/guide/en/elasticsearch/reference/current/indexes-forcemerge.html</a> )的方式做merge。注意：还在写数据的索引，不要执行 force merge。	自动优化 手动优化 忽略

集群诊断报告

# ES 智能诊断系统

诊断项：索引Shard的Segment个数太多[可自动优化]

某个Shard优化前（单Shard 102个文件）：

```
ts8-ws3teg71Plg/0/index$ pwd
ts8-ws3teg71Plg/0/index$ ls
4a.dii      4a_Lucene54_0.dvd  5z.fnm      5z.si        6i_Lucene50_0.tip  7n.fdx      7n.nvm     9k.cfs      9o.fnm      9o.si      9r.si      9u.si
4a.dim     4a_Lucene54_0.dvm  5z_Lucene50_0.doc  6i.dii      6i_Lucene54_0.dvd  7n.fnm      7n.si     9k.si     9o_Lucene50_0.doc  9p.cfe    9s.cfe    segments_f
4a.fdt     4a.nvd            5z_Lucene50_0.pos  6i.dim      6i_Lucene54_0.dvm  7n_Lucene50_0.doc  9f.cfe    9n.cfe    9o_Lucene50_0.pos  9p.cfs    9s.cfs    write.lock
4a.fdx     4a.nvm            5z_Lucene50_0.tim  6i.fdt      6i.nvd            7n_Lucene50_0.pos  9f.cfs    9n.cfs    9o_Lucene50_0.tim  9p.si     9s.si
4a.fnm     4a.si            5z_Lucene50_0.tip  6i.fdx      6i.nvm            7n_Lucene50_0.tim  9f.si     9n.si     9o_Lucene50_0.tip  9q.cfe    9t.cfe
4a_Lucene50_0.doc  5z.dii          5z_Lucene54_0.dvd  6i.fnm      6i.si            7n_Lucene50_0.tip  9g.cfe    9o.dii    9o_Lucene54_0.dvd  9q.cfs    9t.cfs
4a_Lucene50_0.pos  5z.dim          5z_Lucene50_0.doc  6i_Lucene50_0.doc  7n.dii          7n_Lucene54_0.dvd  9g.cfs    9o.dim    9o_Lucene54_0.dvm  9q.si     9t.si
4a_Lucene50_0.tim  5z.fdt          5z.nvd            6i_Lucene50_0.pos  7n.dim          7n_Lucene54_0.dvm  9g.si     9o.fdt    9o.nvd            9r.cfe    9u.cfe
4a_Lucene50_0.tip  5z.fdx          5z.nvm            6i_Lucene50_0.tim  7n.fdt          7n.nvd            9k.cfe    9o.fdx    9o.nvm            9r.cfs    9u.cfs
```

优化后（单Shard 29个文件）：

```
ts8-ws3teg71Plg/0/index$ pwd
ts8-ws3teg71Plg/0/index$ ls
4a.dii      4a.fdx      4a_Lucene50_0.pos  4a_Lucene54_0.dvd  4a.nvm     9v.dim     9v.fnm     9v_Lucene50_0.tim  9v_Lucene54_0.dvm  9v.si
4a.dim     4a.fnm     4a_Lucene50_0.tim  4a_Lucene54_0.dvm  4a.si     9v.fdt     9v_Lucene50_0.doc  9v_Lucene50_0.tip  9v.nvd     segments_g
4a.fdt     4a_Lucene50_0.doc  4a_Lucene50_0.tip  4a.nvd          9v.dii     9v.fdx     9v_Lucene50_0.pos  9v_Lucene54_0.dvd  9v.nvm     write.lock
```

# ES 智能诊断系统

## 诊断项：集群存储容量不足风险

### 集群诊断报告

诊断时间: 2019-10-04 17:00:38      集群名称: [REDACTED]      服务类型: es

运行状态: 已完成      异常级别: 严重

[查看集群信息](#)   [查看Kopf](#)   [查看Grafana](#)

[重新巡检](#)

---

🚨 索引Shard的Segment个数太多 自动优化 手动优化 忽略 >

🚨 集群存储容量不足风险 自动优化 手动优化 忽略 v

<input type="checkbox"/>	ID	数据源	数据源类型	数据源等级	数据源组	风险等级	数据详情	建议	优化操作
<input type="checkbox"/>	2096227	[REDACTED]	cluster	Unknown	[REDACTED]	严重	<a href="#">查看详情</a>	据此集群过去 7 天的日写入数据量预估每日写入量为 [REDACTED] GB/天。预计集群磁盘会在 8 天后容量不足, 无法分配索引Shard。请尽快 (1)扩容 (2)减少数据写入 (3) 删除历史数据。(当前数据: 磁盘总大小 [REDACTED] GB, 已使用 [REDACTED] GB, 占比 80 %, Watermark: 95 %)	<a href="#">自动优化</a> <a href="#">手动优化</a> <a href="#">忽略</a>

共 1 条    10条/页    < 1 >    前往 1 页

优化建议:

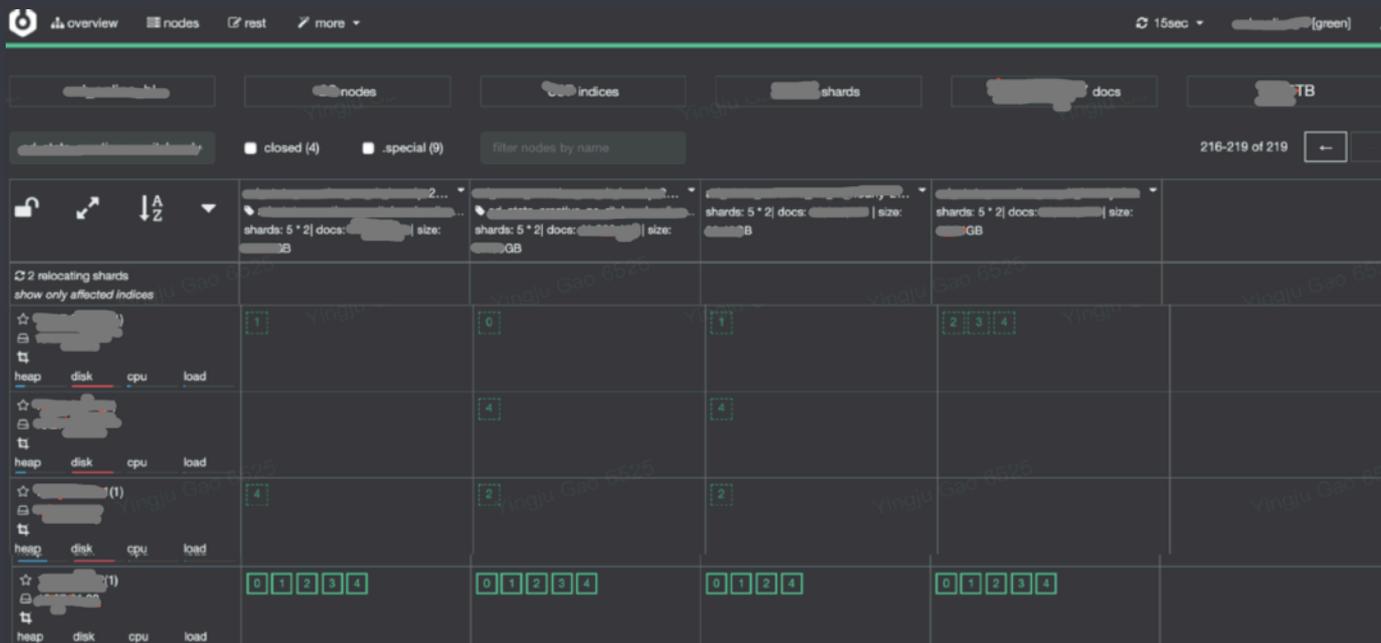
据此集群过去 7 天的日写入数据量预估每日写入量为 ? GB/天。预计集群磁盘会在 8 天后容量不足, 无法分配索引Shard。请尽快 (1)扩容 (2)减少数据写入 (3) 删除历史数据。(当前数据: 磁盘总大小 ? GB, 已使用 ? GB, 占比 80 %, Watermark: 95 %)

优化效果:

此诊断项上线后, 容量不足从事后报警转变为提前预测, 由容量不足引发的集群 RED(Primary Shard无法分配)的电话报警从一周最多4次减少到0次。

# ES 智能诊断系统

诊断项：索引shard分配不均衡[可自动优化]



问题现象

# ES 智能诊断系统

诊断项：索引shard分配不均衡[可自动优化]

### 集群诊断报告

诊断时间: 2019-08-07 10:55:53      集群名称: [REDACTED]      服务类型: es

运行状态: 已完成      异常级别: 警告

[重新巡检](#)

[查看集群信息](#)   [查看Kopf](#)   [查看Grafana](#)

---

① 索引shard分配不均衡 自动优化 手动优化 忽略

ID	数据源	数据类型	数据源等级	数据源组	风险等级	数据详情	建议	优化操作
1377616	[REDACTED]	index	P4	[REDACTED]	警告	<a href="#">查看详情</a>	此索引在各个节点或着在同一节点的不同分区上的shard分配不均衡, 这个索引的总shard个数为10.0, 允许分配的节点个数为18, 需要迁移的shard个数为5, 需要迁移的数据总大小为 [REDACTED] GB, 迁移比例为0.50, 存在有多个shard分配到同一磁盘分区的情况有 $\$(data.conflictOnDiskPartitionNum)$ 个, 建议手动Reroute(参考 <a href="https://www.elastic.co/guide/en/elasticsearch/reference/current/cluster-reroute.html">https://www.elastic.co/guide/en/elasticsearch/reference/current/cluster-reroute.html</a> )此索引的shard。	<a href="#">自动优化</a> <a href="#">手动优化</a> <a href="#">忽略</a>

优化建议:

此索引在各个节点或着在同一节点的不同分区上的shard分配不均衡。这个索引的总shard个数为10, 允许分配的节点个数为18。

**需要迁移的shard个数为5,需要迁移的数据总大小为30.60/GB, 迁移比例为0.50, 存在有多个shard分配到同一磁盘分区的情况有 0 个。**

建议手动Reroute(参考 <https://www.elastic.co/guide/en/elasticsearch/reference/current/cluster-reroute.html>)此索引的shard。

# ES 智能诊断系统

## 诊断项：时序型索引Shard size不合理[可自动优化]

### 集群诊断报告

诊断时间: 2019-10-04 17:00:36      集群名称: [REDACTED]      服务类型: es

运行状态: 已完成      异常级别: 严重

[重新检测](#)      [查看集群信息](#)      [查看Kopf](#)      [查看Grafana](#)

---

● 时序型索引Shard size不合理 自动优化 手动优化 忽略

ID	数据源	数据源类型	数据源等级	数据源组	风险等级	数据详情	建议	优化操作
2095294	[REDACTED]	index	Unknown	[REDACTED]	严重	<a href="#">查看详情</a>	此索引为时序型索引(按日、月分割的索引), Shard size不合理。索引总大小 [REDACTED] MB, shard 个数: 26。一般每个shard合理的大小范围是2000 ~ 60000 MB。根据此索引近期的实际shard size分析并结合该索引可分配节点的磁盘分区总数: 83 * 90%, 建议调整此索引对应index template 中 number_of_shards = 11, 来使新生成的索引shard大小更合理。	<a href="#">自动优化</a> <a href="#">手动优化</a> <a href="#">忽略</a>

共 1 条    10条/页    < 1 >    前往 1 页

问题：目前准备优化的Shard Size阈值是Google上查到的 2G ~60G的范围，可能不合理

优化建议：

此索引为时序型索引(按日、月分割的索引), Shard size不合理, 索引总大小 [REDACTED] MB, shard 个数: 26。一般每个shard合理的大小范围是2000 ~ 60000 MB。根据此索引近期的实际shard size分析并结合该索引可分配节点的磁盘分区总数: 83 \* 90%, 建议调整此索引对应index template 中 number\_of\_shards = 11, 来使新生成的索引shard大小更合理。

优化目标：

- 自动调整不合理的索引Shard个数

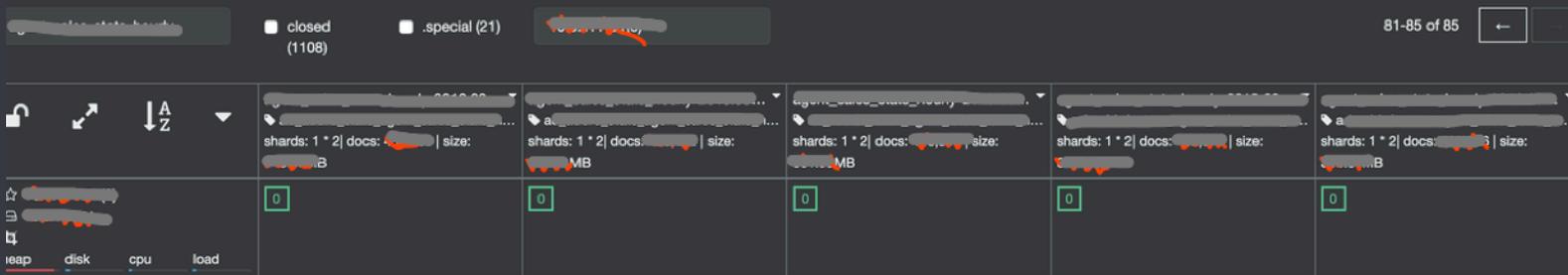
典型案例：

- 一个只有几MB的索引, 有100+个Shard
- 一个几TB的索引, 只有2个Shard
- 按日划分的索引, 其大小会随着时间变化, 如果shard个数始终不变是不合理的。

# ES 智能诊断系统

## 诊断项：集群Shard个数分配不均衡

举例，ad\_bi集群，现象：部分分日写入的索引，shard集中分配到了同一个节点上。



# ES 智能诊断系统

## 诊断项：集群Shard个数分配不均衡

### 集群诊断报告

诊断时间: Invalid date      集群名称: [REDACTED]      服务类型: es

运行状态: 运行中      异常级别: 未获取

[查看集群信息](#)   [查看Kopf](#)   [查看Grafana](#)

#### 集群shard个数分配不均衡

自动优化 手动优化 忽略

ID	数据源	数据源类型	数据源等级	数据源组	风险等级	数据详情	建议	优化操作
1402389	[REDACTED]	cluster	Unknown	[REDACTED]	警告	查看详情	集群中各个节点的shard个数分配不均，归一化后的最大方差反映了偏离程度为0.02，对应的for_index的值为NO NE，设有该值的节点中平均每个节点应该分配136.97个shard，建议手动Reroute(参考 <a href="https://www.elastic.co/guide/en/elasticsearch/reference/current/cluster-reroute.html">https://www.elastic.co/guide/en/elasticsearch/reference/current/cluster-reroute.html</a> ) 集群中的shard。	自动优化 手动优化 忽略

共 1 条    10条/页    < 1 >    前往 1 页

### 数据详情

数据披露

```
"10.8.10.133": 1101,
"10.10.10.1": 0,
"10.8.10.144": 162,
"10.10.10.151": 151,
"10.11.10.159": 109,
"10.11.10.139": 139,
"10.8.10.110": 110,
"10.8.10.117": 117,
"10.8.10.112": 112,
"10.10.10.121": 121,
"10.11.10.154": 154,
"10.10.10.129": 129,
"10.8.10.128": 128,
"10.8.10.207": 207,
"10.10.10.111": 111,
"10.10.10.130": 130,
"10.8.10.133": 133,
"10.10.10.1": 1,
"10.8.10.109": 109,
"10.10.10.1": 0,
"10.8.10.107": 107,
"10.8.10.123": 123,
"10.11.10.202": 202,
"10.8.10.133": 1093,
```

诊断结果显示：有两个节点分配的shard个数过多，达到1100+，是其他节点的6倍以上。

优化步骤：

1. 下线业务
2. 调整c.r.a.balance.index, c.r.a.balance.shard两个参数
3. 等待集群完成自动relocating
4. 上线业务

# ES 智能诊断系统

## 诊断项: Index Template 校验不合规

问题:

Cluster Health	Node 1	Node 2
shards: 15 * 3   docs: 1000000   size: 1GB	shards: 15 * 3   docs: 1000000   size: 1GB	shards: 15 * 3   docs: 1000000   size: 1GB
60 unassigned shards show all indices	0 0 1 1 2 2 3 3 4 4 5 5 6 6 7 7 8 8 9 9 10 10 11 11 12 12 13 13 14 14	0 0 1 1 2 2 3 3 4 4 5 5 6 6 7 7 8 8 9 9 10 10 11 11 12 12 13 13 14 14
☆  (1)	2	4
🗑️  (1)		
🔧		
heap disk cpu load		
☆  (1)	6	8

诊断结果:

数据源组	风险等级	数据详情	建议	优化操作
	严重	查看详情	index template order顺序重复, 导致无法确定template如何生成index中的setting和mapping	自动优化 手动优化 忽略

# ES 智能诊断系统

诊断项: Index Template 校验不合规

原因 & 解决:

数据详情

```
数据披露:
  ]
}
},
"riskTemplate": {
  "0: *": {
    "base-template",
    "template",
    "template",
    "template"
  ]
}
}
```

```
update template base-template
base-template
1 {
2   "order": 0,
3   "template": "*",
4   "settings": {
5     "index": {
6       "number_of_shards": "5",
7       "routing": {
8         "allocation": {
9           "total_shards_per_node": "1"
10        }
11      }
12    }
13  },
14  "mappings": {},
15  "aliases": {}
16 }
```

```
update template
"base-template"
1 {
2   "order": 0,
3   "template": "*",
4   "settings": {
5     "index": {
6       "number_of_shards": "15",
7       "routing": {
8         "allocation": {
9           "total_shards_per_node": "3"
10        }
11      }
12    }
13  },
14  "mappings": {
```

更新为 order: 10

# ES 智能诊断系统

## 诊断项：索引Template 配置风险[可自动优化]

目标：能够自动适应集群节点数量变化

- 集群扩容， 缩容
- 异常下线节点后
- 设置node attribute

Index template 配置风险										自动优化 手动优化 忽略
<input type="checkbox"/>	ID	数据源	数据源类型	数据源等级	数据源组	风险等级	数据详情	建议	优化操作	
<input type="checkbox"/>	2096953	[REDACTED]	index	Unknown	[REDACTED]	严重	<a href="#">查看详情</a>	此索引对应的index template([REDACTED]-template)中配置的total_shards_per_node = 2 大小不合适。该模板下的索引总shard数为 $22 * (1 + 1)$ ，shard数除以shard_per_node 大于集群节点总数12，将会导致shard无法完全分配。根据此索引近期的实际shard size分析来看，建议调整 <u>total_shards_per_node = 4</u>	<a href="#">自动优化</a> <a href="#">手动优化</a> <a href="#">忽略</a>	

# ES 智能诊断系统

诊断项：索引Template 配置风险[可自动优化]

优化前：

```
update template [redacted]-template
[redacted]-template

1 {
2   "order": 2,
3   "template": "[redacted]*",
4   "settings": {
5     "index": {
6       "number_of_shards": "22",
7       "routing": {
8         "allocation": {
9           "total_shards_per_node": "2"
10        }
11      }
12    }
13  },
14  "mappings": {
```

自动优化后：

```
update template [redacted]-template
[redacted]-template

1 {
2   "order": 2,
3   "template": "[redacted]*",
4   "settings": {
5     "index": {
6       "number_of_shards": "22",
7       "routing": {
8         "allocation": {
9           "total_shards_per_node": "4"
10        }
11      }
12    }
13  },
14  "mappings": {
15    "default": {
```

集群扩容、缩容，索引冷热节点迁移都需要重新计算 total\_shards\_per\_nodes

# ES 智能诊断系统

## CaseByCase

### 1. 自动保存慢查询Profile

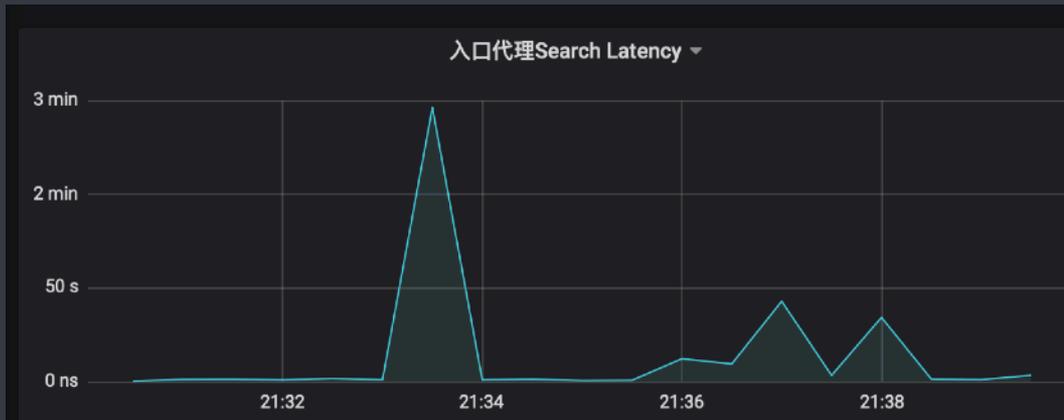
- 自动根据Search Latency PCT99 检测慢查询。
- 及时执行Profile，方便慢查询复现。
- 避免侵入业务。
- 兜底限速，防止影响正常业务。

### 2. 监控数据关联分析（皮尔逊系数）

- 分析监控数据的关联性（使用皮尔逊系数寻找监控时序数据的相关性）。
- 快速缩小排查问题的范围。
- 为工程师定位根因节约时间。

# ES 智能诊断系统

## CaseByCase: Search PCT99 慢查询关联指标分析



从3000+指标中找到线上查询时延毛刺的可能原因

指标数值:

- `jvm.mem.heap_used`
- `bulk_queue`
- `search_queue`
- `pending_tasks`
- `load.1min`
- `cpu.busy`
- `cpu.iowait`
- `disk.io.write_bytes/device=+`
- `disk.io.read_bytes/device=+`
- ...

指标纬度:

- 集群
- 节点
- 索引
- 宿主机
- 磁盘分区
- 网卡
- ...

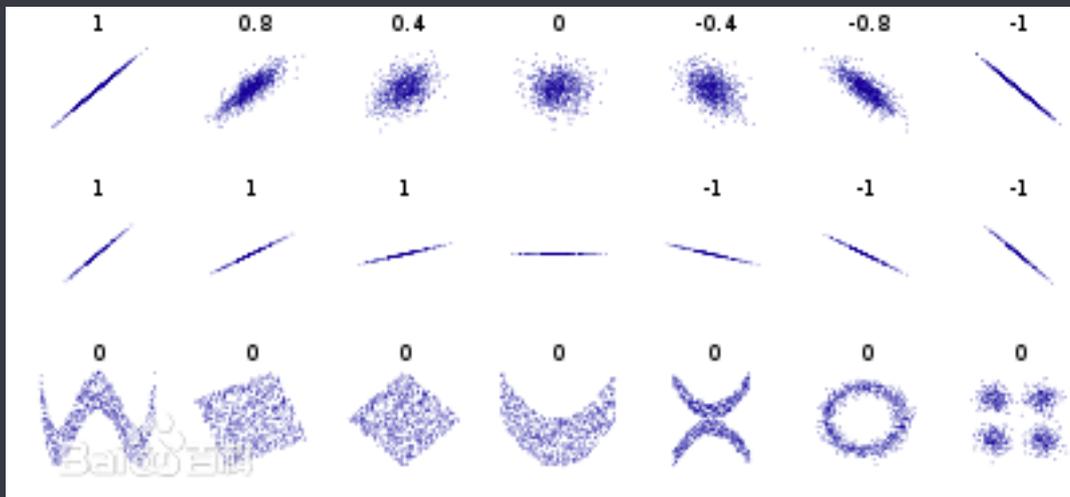
# ES 智能诊断系统

## CaseByCase: Search PCT99 慢查询关联指标分析

皮尔逊相关系数：

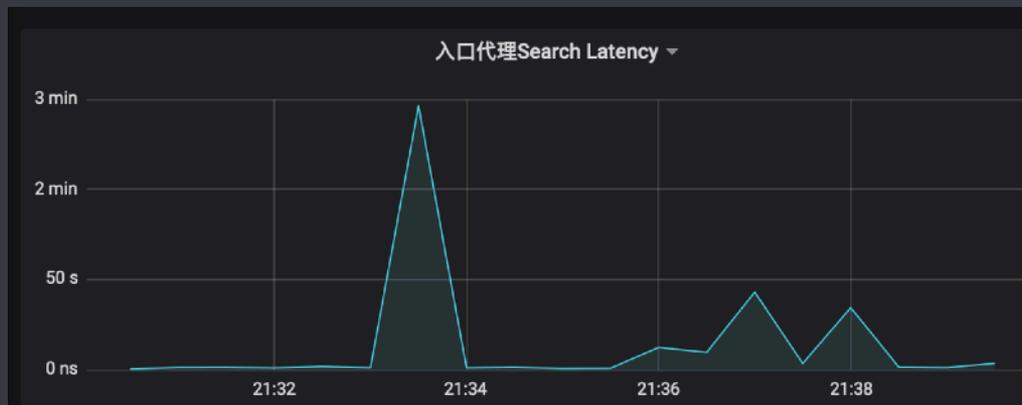
- 皮尔逊相关系数可用于度量两个变量的相关性。
- 值处于-1到1之间，1表示正相关，-1表示负相关，0表示无关。

$$\rho_{X,Y} = \frac{\text{cov}(X, Y)}{\sigma_X \sigma_Y} = \frac{E[(X - \mu_X)(Y - \mu_Y)]}{\sigma_X \sigma_Y},$$



# ES 智能诊断系统

## CaseByCase: Search PCT99 慢查询关联指标分析



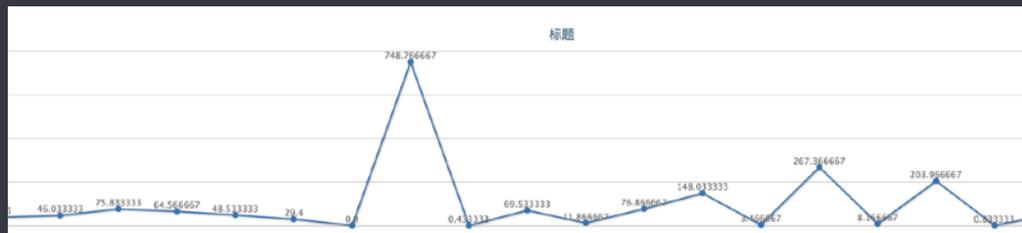
Search Latency Pct99趋势图

结论:

- 进行慢查询定位分析, 发现关联度最高的是 10.10.xx.yy 下的 `disk.io.read_requests/device=sdl` 指标关联度最高。
- 数据计算时间 < 10s, 远远小于人工定位时间

问题:

- 会遇到集群80%节点的磁盘读写都与毛刺关联度较高的情况, 关联分析没有意义。

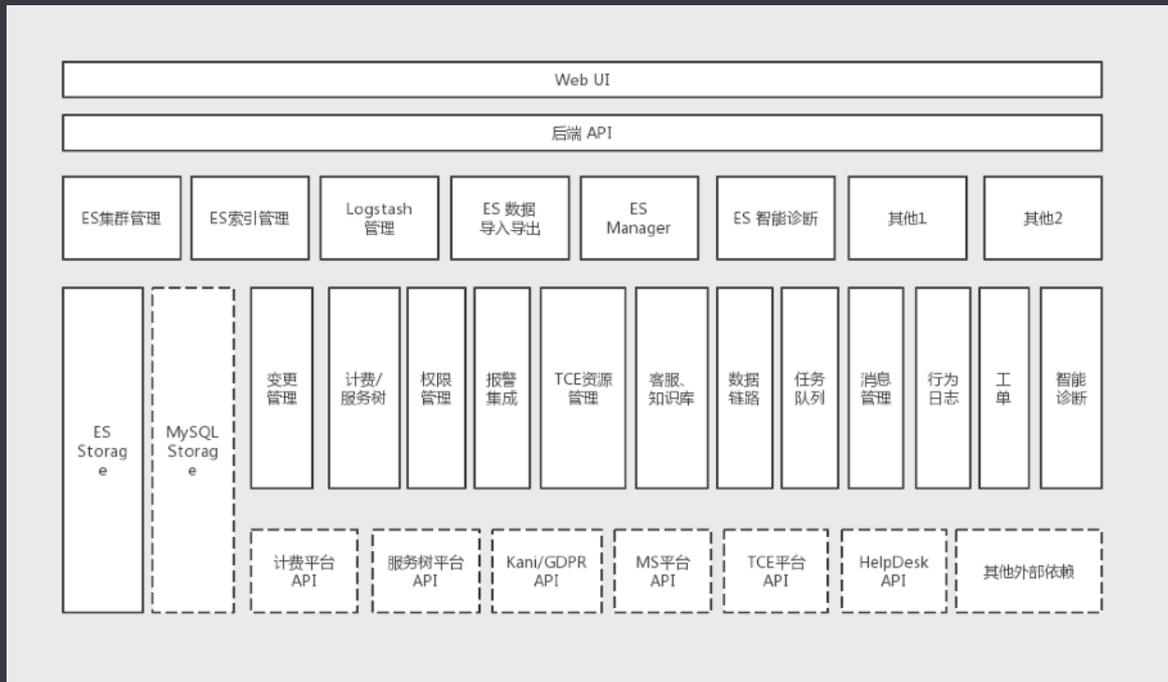


10.10.xx.yy节点的`disk.io.read_requests/device=sdl`的趋势图

# Datapalace 数据管理中台与ES智能诊断系统的设计与实现

# Datapalace 数据管理中台的技术设计与实现

## 分层、分模块架构



分层、分模块设计原则：

- 学习业务、理解业务后，从业务模型到软件架构的映射。
- 整体划分为基础通用功能、业务功能两层。
- 多个单一职责的模块，便于多人高效协作。
- 流程、业务分离；流程统一实现，业务插件化接入。
- 模块内高内聚，模块间低耦合。
- 部分适用：上层不依赖底层，两者依赖抽象(DIP)。

DIP(Dependence Inversion Principle) 依赖倒置/控制反转

# Datapalace 数据管理中台的技术设计与实现

## 核心模块的设计与实现：数据链路



核心实体：

- Graph: 数据链路
- Vertex: 链路上的节点
- Edge: 节点的连接关系

```
public class Graph {  
    private long id;  
    private String name;  
    private String description;  
    private List<Vertex> vertices;  
    private List<Edge> edges;  
}
```

```
/**  
 * 创建数据链路  
 */  
public List<Vertex> createGraph(Graph graph);  
  
/**  
 * 连接节点，要求至少有一个vertex已存在于graph中  
 */  
public List<Vertex> connect(Vertex source, Vertex target);  
  
/**  
 * 删除链路节点  
 */  
public Vertex deleteVertex(Long vertexId);  
  
/**  
 * 通过链路中的任一节点，查询链路  
 */  
public Graph queryGraphByVertexId(long vertexId);
```

# 未来计划

# 未来计划

- 提高ES写入速度（很迫切）
  - Spark直写Lucene，再Restore到ES中
  - 修改Lucene源码
  - 对于不需要倒排索引的场景，尝试用RocksDB替换Lucene
- 提高ES查询性能
  - PCT99 慢查询，节点OOM根因定位
  - 查询流控
  - 修改查询源码，实现类似Spark的推测执行（多副本同时执行，看谁先执行完）
- 改进Master单点性能问题
  - 升级ES 版本  $\geq$  6.8.0
  - 修改Master源码
- AIOps + ES
  - Query执行时长，资源消耗预测(神经网络)
  - 慢查询根因定位(决策树，贝叶斯，神经网络)

# Thanks



个人微信

Contributors: 张海雷 高英举 汤明 伏开宇 郭峰 宋红 范文棋 袁宇豪 文炫达 卜衡 方品 龚凌磊 胡东林

