


Elasticsearch在腾讯的实践

姜国强
腾讯



战略级赞助商  HUAWEI

钻石级赞助商  普翔

白金级赞助商  华夏博格

 神州数码
Digital China

金牌级赞助商  iDataAPI

合作伙伴  开源中国
oschina.net

 掘金

 众语时

 IT大咖说

 otpub

 Broadview
www.broadview.com.cn

 百格活动
bagevent.com

 MAXHUB
高效会议平台

Elasticsearch在腾讯的实践

姜国强 - 腾讯高级工程师

目录

- 发展和现状
- 遇到的挑战
- 优化实践

发展和现状

Tencent 腾讯 | 腾讯云



2015.12

日志分析 ES 1.7



2017.05

监控统计 ES 5.6

2016.10

在线搜索 ES 5.1



2018.06

腾讯云发布ES服务



发展和现状

Tencent 腾讯 | 腾讯云

3年运管
10+地域部署
多版本管理

运管经验

日志：50%
搜索分析：20%
监控统计：20%

场景分布

超大集群
350节点 600w/s写入
在线检索平响10ms级

代表业务

遇到的挑战

可靠性：

- 机房光纤异常
- 误操作
- Full GC、OOM
- 压力不均

性能：

- 查询慢
- 毛刺

成本：

- 存储成本
- 内存



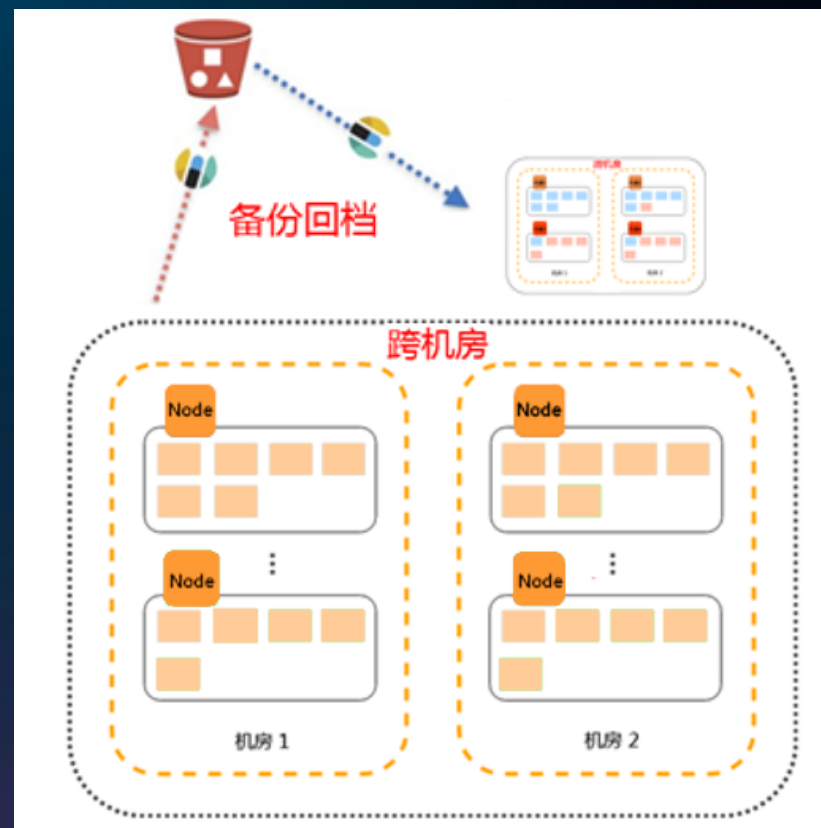
可靠性优化

跨机房容灾

- Shard Allocation Awareness
- Forced Awareness
- 机架位感知

备份回档

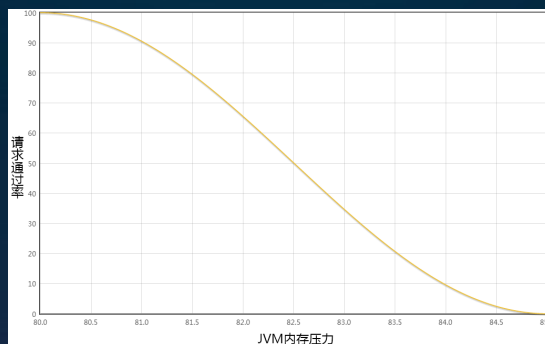
- 对象存储：Repository Plugin
- 周期性调度



可靠性优化

Circuit Breaker

- 默认Breaker配置需调优
- Global Circuit Breaker
 - 避免集群Hang或OOM
 - 集群吞吐稳定
- Better Request Circuit Breaker
 - 避免单个查询拖垮集群



■ Rest前限流 ■ Rest后限流 ■ 原生(OOM)

210,000

146,000

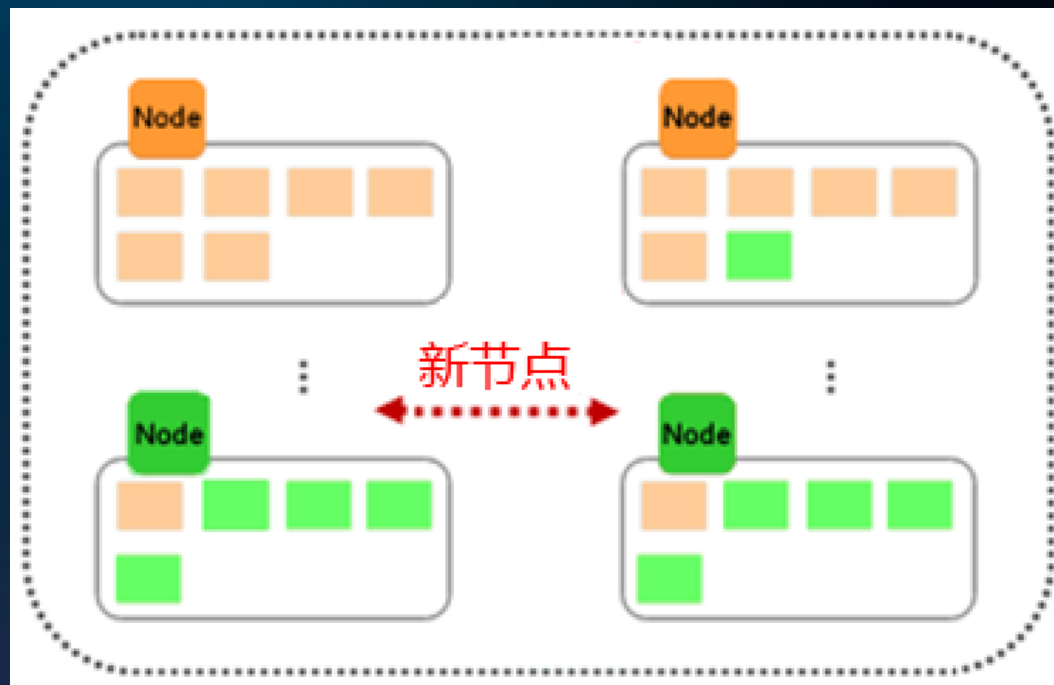
0

高并发写入压测

可靠性优化

集群均衡策略

- 新节点分片不均
 - 已有集群扩容节点
 - 故障机器恢复后加入
- 创建索引慢
 - 已有近万分片，创建几十秒

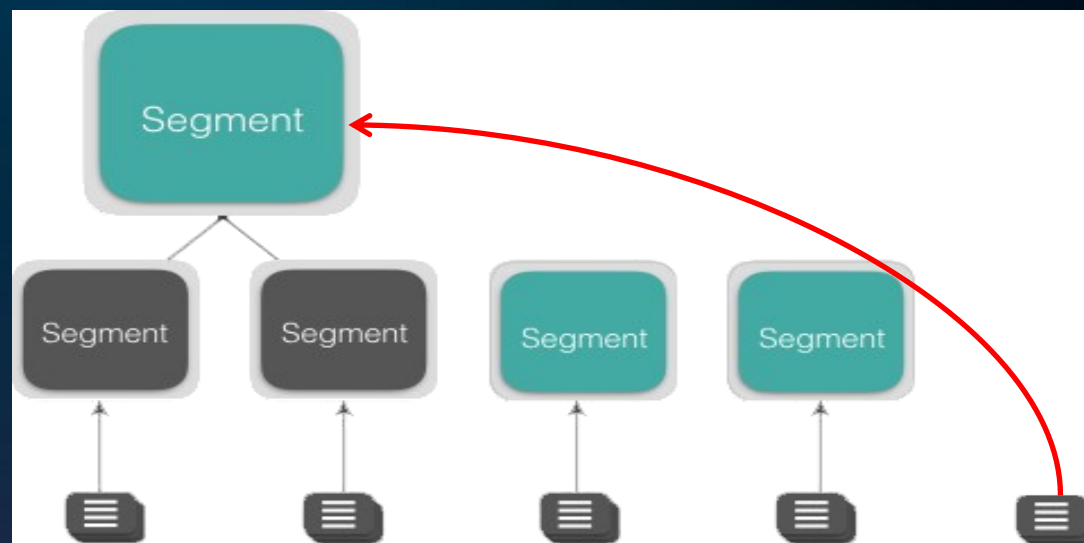


- TCP半连接队列打满
- Master任务去重 (6.4.0)

性能优化

Merge策略优化

- 时间序Merge
 - 加重随机io
 - 不利于Segment裁剪
- 冷索引收敛Merge
 - 冷索引Segment数量：30~
 - Inactive后自动Merge
- 提升：50% ~ 100%

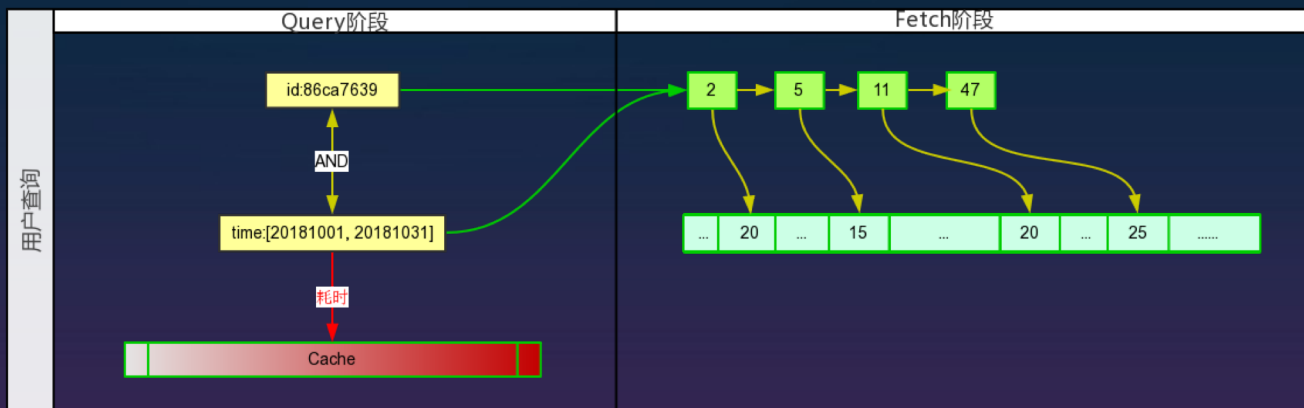


性能优化

查询计划优化

- Query阶段优化
 - 优化缓存：毛刺 (50ms -> 750ms)
 - Index Sorting
- Fetch阶段优化
 - 顺序Fetch
 - 避免Fetch空抓取

```
1 GET host_monitor_60/_search?routing=b3528e3a
2 {
3   "size": 65536,
4   "stored_fields": "_none_",
5   "docvalue_fields": ["cpu_usage"],
6   "sort": {"timestamp": {"order": "asc"}},
7   "query": {
8     "bool": {
9       "filter": [
10        {
11          "range": {1. 7天超10亿条数据, 缓存耗时间、空间
12            "timestamp": {
13              "gte": "2018-11-1 00:00:00",
14              "lte": "2018-11-8 00:00:00",
15              "format": "yyyy-MM-dd HH:mm:ss",
16              "time_zone": "+08:00"
17            }
18          }
19        },
20        {
21          "term": {2. host数据分布分散, 随机io高
22            "host_uuid": "84c0d3a9-dbdf-4e02-9460-2a86c05afdfc"
23          }
24        }
25      ]
26    }
27  }
28 }
```



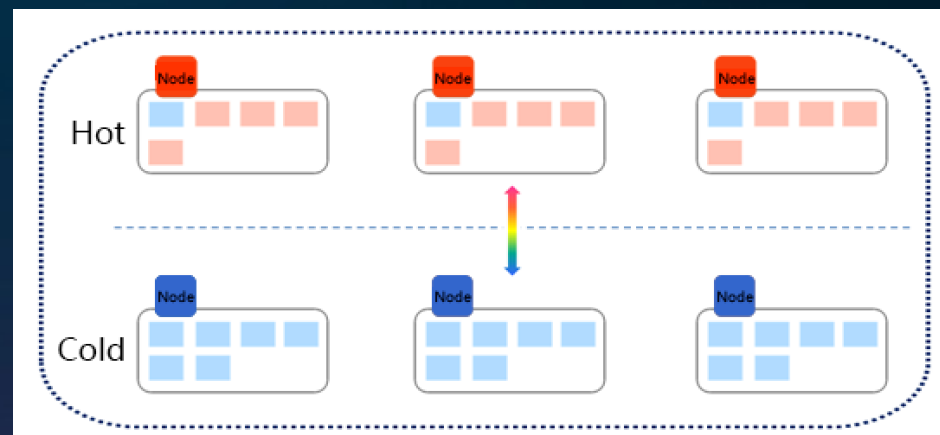
成本优化

针对业务调优

- 日志、结构化数据分析、全文检索

冷热分离

- SSD & HDD
- Shard Allocation Filtering
- 冷热迁移调度



成本优化

预聚合 (Rollup)

- 降低数据存储成本
 - 近期保留细粒度，历史保留聚合结果
- 查询更高效
 - 预计算
 - 数据量下降
- 异步调度：限制速度、并发

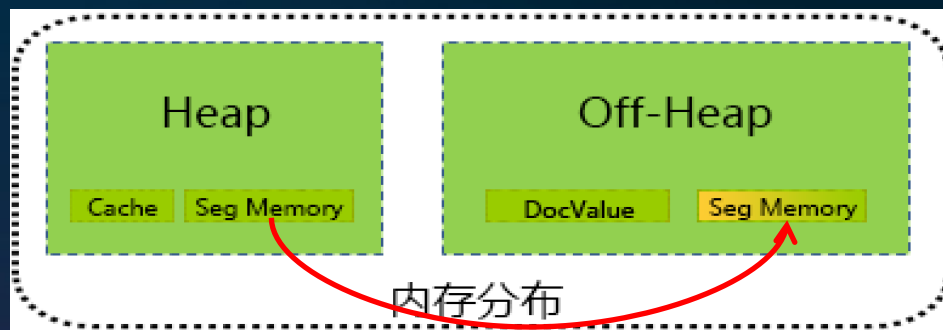
1	region	host	time	cpu_usage
2	gz	10.11.12.13	2017-07-01 10:00:00	20
3	bj	14.15.16.17	2017-07-01 10:00:00	20
4	gz	10.11.12.13	2017-07-01 10:00:10	30
5	bj	14.15.16.17	2017-07-01 10:00:10	40
6	gz	10.11.12.13	2017-07-01 10:00:20	25

1	region	host	time	cpu_usage
2	gz	10.11.12.13	2017-07-01 10:00:00	25
3	bj	14.15.16.17	2017-07-01 10:00:00	30

成本优化

内存优化

- 常见内存分布
 - SM、Cache、Index Buffer ...
- 内存不足
 - 1TB 数据 -> 2~5GB+ SM
 - 关闭索引：使用成本增加
 - 扩大内存：COOP (32GB)、更长GC
- LRU & Off-Heap
 - 方便，节省堆内存
 - GC更低频、短暂



管控平台

服务托管

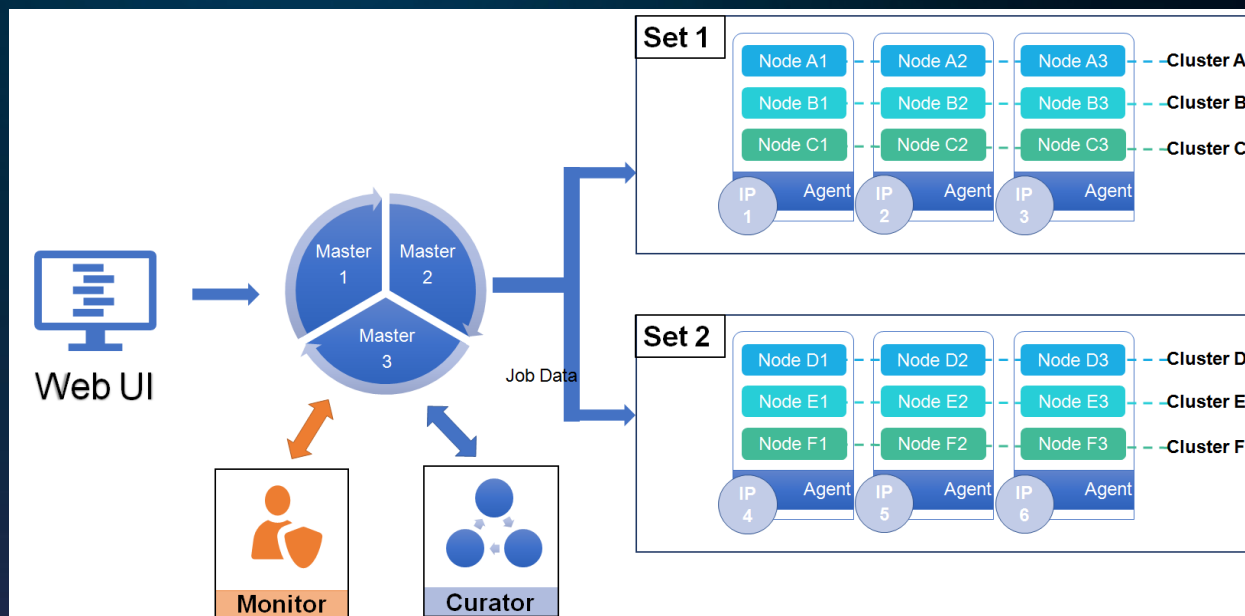
- 集群管理
- 监报告警
- Curator管理

常用插件

- SQL插件
- 分词插件

安全

- VPC隔离
- 权限系统



Thank You



腾讯云
ES 介绍



腾讯云
云+社区 ES 专栏



elastic 中文社区

专业、垂直、纯粹的 Elastic 开源技术交流社区

<https://elasticsearch.cn/>