



Elasticsearch 在奇安信的大规模实践

张超

2019/11/24, 资深研发工程师, 奇安信
《Elasticsearch 源码解析与优化实战》作者

主要内容

- Elasticsearch 在奇安信的应用情况
- 内核改进
- 生产环境建议

应用场景

集群规模

总体规模

节点数：6000+

数据总量：50PB+

单集群

最大100+节点

千亿条文档

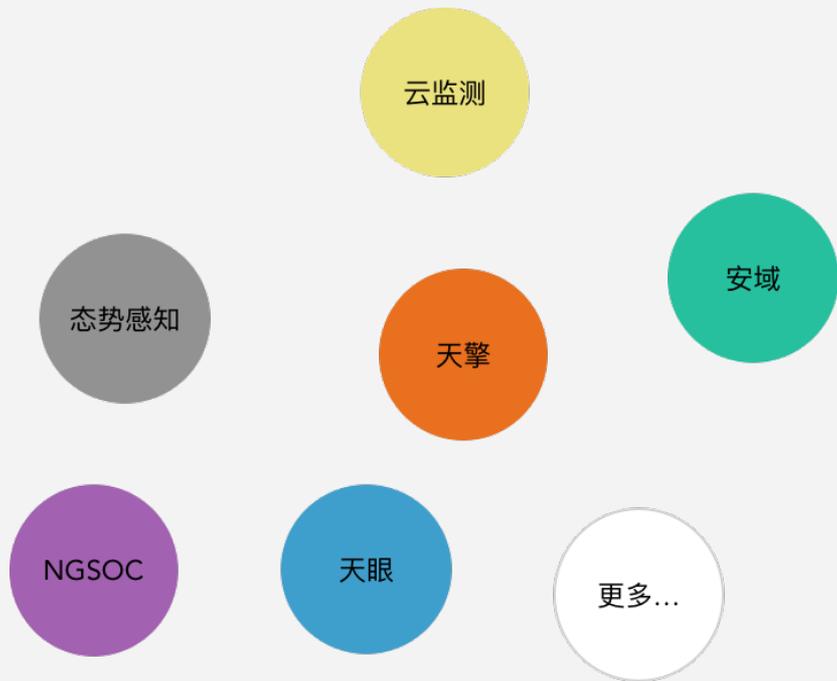
百TB数据总量

典型场景

流量分析

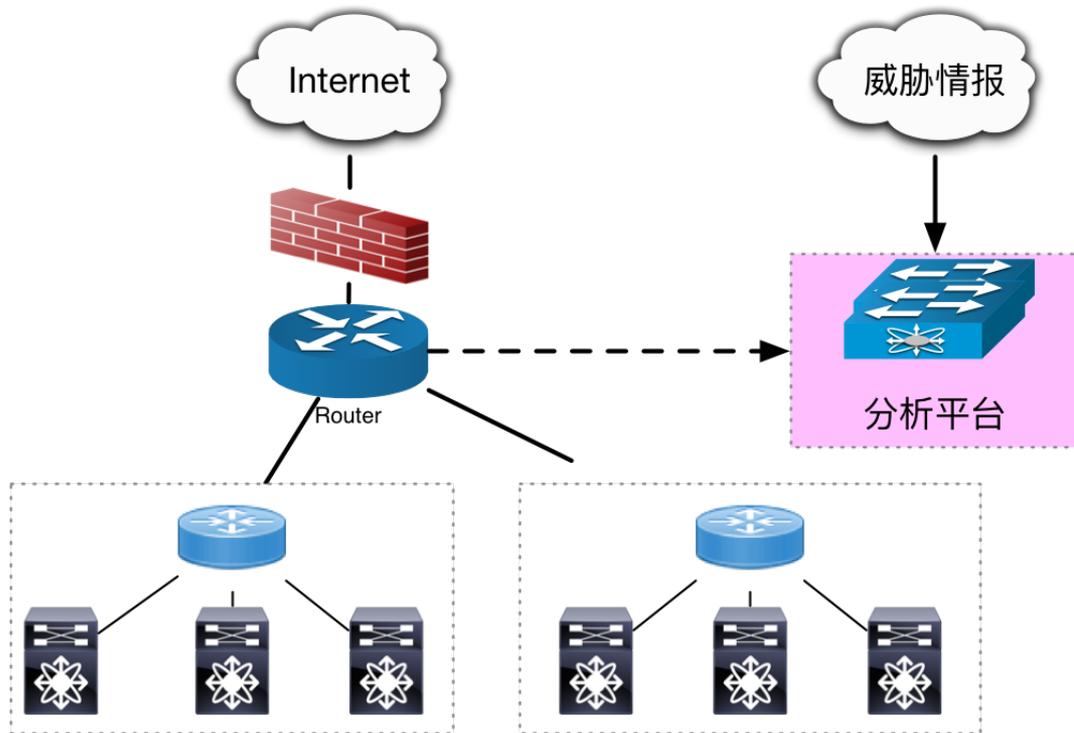
告警信息

应用程序日志



应用场景-流量分析

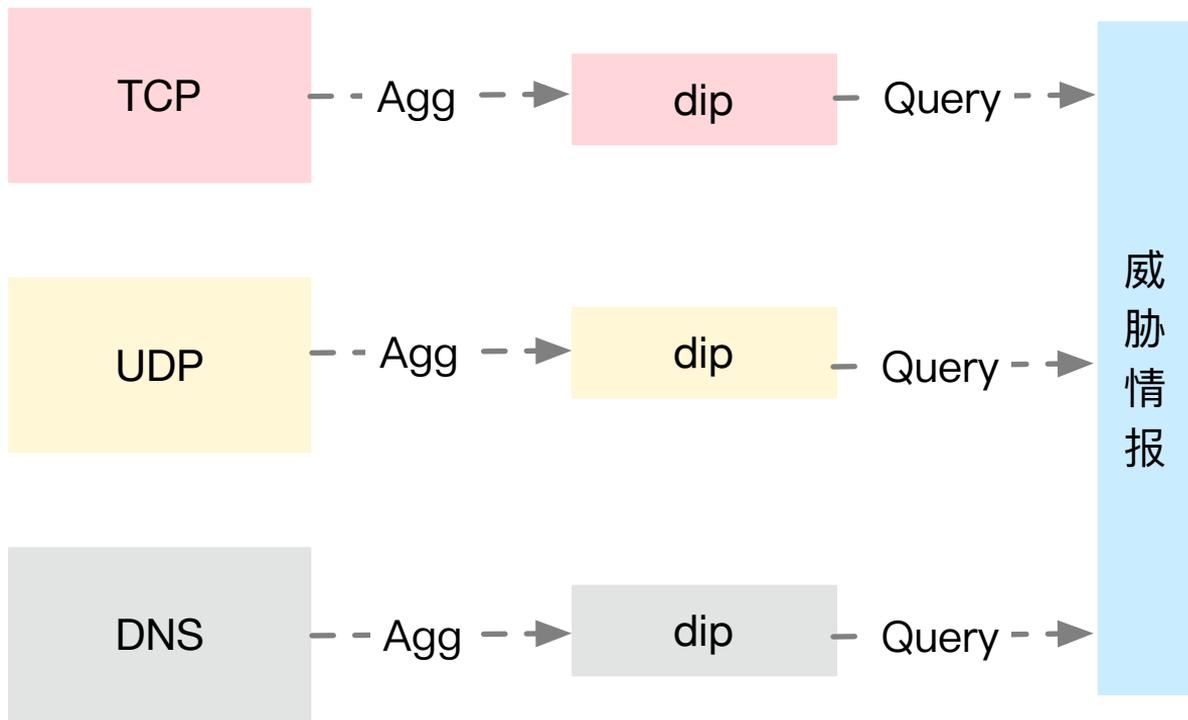
从流量中发现威胁，包括 TCP，UDP，HTTP，DNS，MAIL，LDAP...



流量解析后的结果

```
"sport": { "type": "integer"},
"dport": { "type": "integer"},
"sip": { "type": "keyword"},
"dip": { "type": "keyword"},
"stime": { "type": "date", "format": "YYYY-MM-dd HH:mm:ss.SSSSSS"},
"dtime": { "type": "date", "format": "YYYY-MM-dd HH:mm:ss.SSSSSS"},
"proto": { "type": "keyword"},
"uplink_length": { "type": "integer"},
"downlink_length": { "type": "integer"},
"client_os": { "type": "text" },
"server_os": { "type": "text" },
"src_mac": { "type": "keyword"},
"dst_mac": { "type": "keyword"},
"down_payload": { "type": "text" },
"up_payload": { "type": "text" },
"summary": { "type": "keyword" }
```

查询过程实现



流量分析-产生告警

存储检测结果，提供查询和聚合

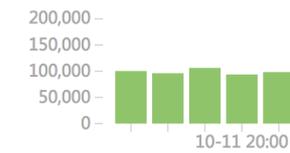


应用场景-用户手工查询

解析日志,原始日志,异常报...

共5059093 条搜索结果(从2019-10-18 00:00:00 - 2019-10-21 00:00:00)

文件传输 Web访问



资产分组

展示字段

已选字段

- ES录入时间
- HTTP-X-Forwarded-For
- HTTP-userAgent
- 源端口

目的端口: " 80 " ×

输入关键字过滤

通用字段

下行字节数 ES录入时间 URI 目的IPv6 源端口 源IP 协议 MIME类型 上行字节数 用户名 源IPv6 采集设备序列号 目的端口 目的IP

流量通用字段

目的IP纬度 源IP所属城市 源IP纬度 源IP经度 源IP所属国家 源IP所属省份 目的IP所属省份 目的IP所属大洲 目的IP经度 目的IP所属国家 源IP所属大洲 目的IP所属城市 厂商 发生时间 设备IP host (完整) host host (完整逆序) Host-MD5 URI (完整逆序) 状态码 URI-MD5 HTTP-Referer URI (完整) 命令行 版本号 返回结果 数据库-类型 上行包数 下行包数 下行payload 源MAC 目的MAC 结束时间 (TCP/UDP) 开始时间 (TCP/UDP) 上行payload

设备和系统通用字段

事件ID 原始日志 接收时间 设备IP

2019-10-20 23:59:18.533 Go-http-client/1.1 32993 172.17.0.194 POST

2019-10-11 00:00:00 - 2019-10-21 00:00:00

搜索

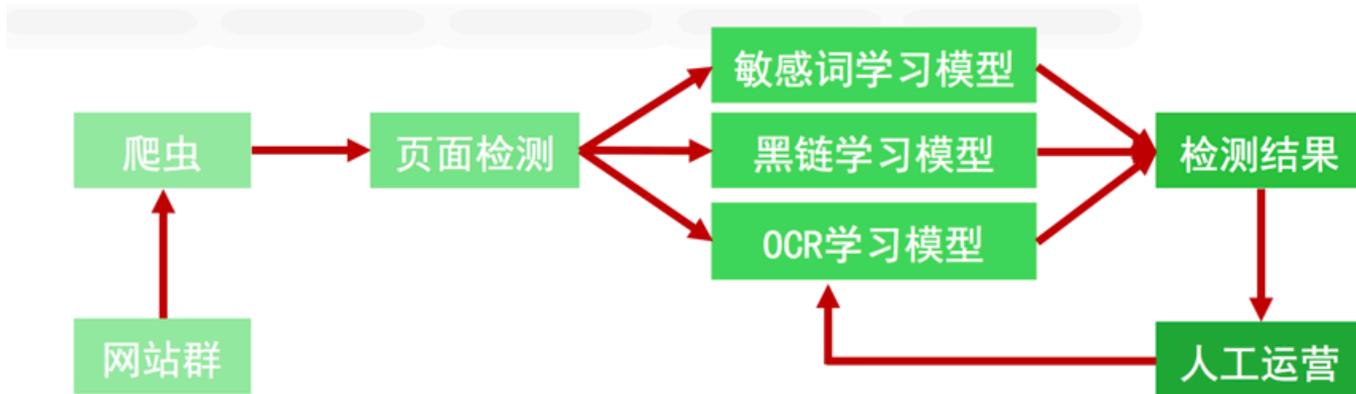
快捷模式

下钻 返回上层 时间轴重置

导出日志

源端口	源IP	HTTP-Method
32993	172.17.0.194	POST

应用场景-网站云监测



结果样本

```
"_source" : {
  "id" : 12492415,
  "type" : "scan",
  "created_at" : "2017-12-27 00:46:05",
  "status" : "0",
  "level" : "3",
  "host_id" : "484742",
  "affected_url" : "http://ccg[REDACTED]v.cn/portal/topicView.do?method=find?typeI
  tarea%3E%3C%2Fscript%3E%20%3Cscript%3Ealert(42873)%3C%2Fscript%3E",
  "vul_md5_id" : "b28c6318ba1a1d1260f56bb02781e9e5",
  "vul_title" : "Cross Site Scripting",
  "vul_common_title" : "跨站脚本攻击漏洞",
  "vul_xml_name" : "Cross_Site_Scripting.xml",
  "vul_origin_url" : "http://ccg[REDACTED]gov.cn/portal/topicView.do?method=find?vie
  "vul_type" : "64"
}
```

告警明细

需要密集查询最近结果

告警总数: 5155

去重后网站个数: 95

是否开启自动刷新

处理已选

确认已选

误报已选

处理全部

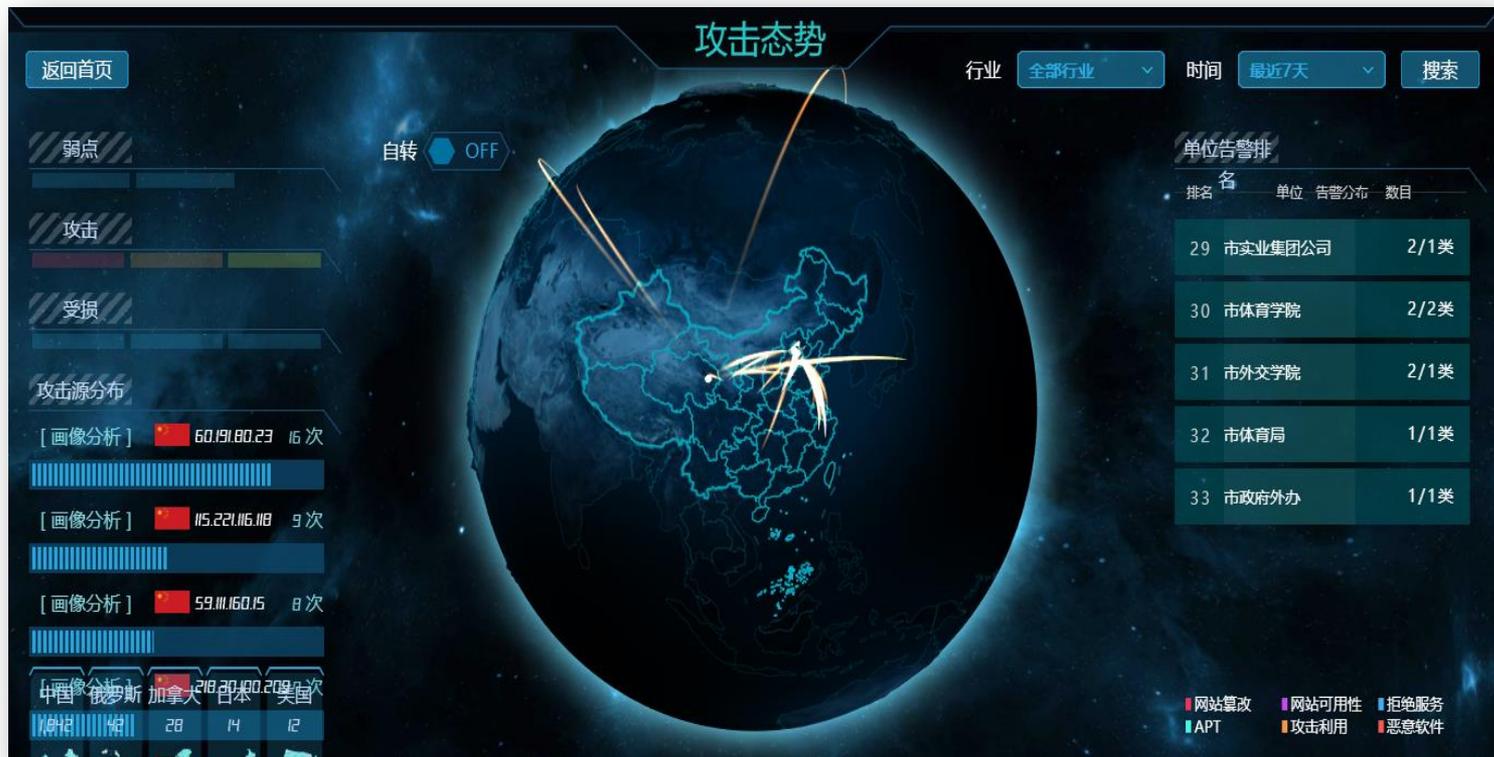
确认全部

误报全部

<input type="checkbox"/>	时间	基本信息	威胁等级	告警类型	告警信息	运维	操作
<input type="checkbox"/>	未读 2019-11-24 17:02:22	网站: http://www.████████.com 网站名: none 账号: test-yunji ██████.cn	中危	违规内容	告警状态: 已确认 词语数量: 1 敏感词语: 狐媚 URL数量: 1 相关资产页面: 查看相关资产页面	误报	通知 处理 详情 调整策略
<input type="checkbox"/>	未读 2019-11-24 17:02:02	网站: http://www.████████.com 网站名: none 账号: test-yunji ██████.cn	中危	违规内容	告警状态: 已确认 词语数量: 1 敏感词语: 色情 URL数量: 4 相关资产页面: 查看相关资产页面	误报	通知 处理 详情 调整策略
<input type="checkbox"/>	2019-11-24 17:00:37	网站: http://1j.████████.pm 网站名: 可视化站47 账号: test-yunjian ██████.cn	高危	可用性	告警状态: 已处理 状态: 已恢复 原因: 总响应时间超过阈值 相关资产页面: 查看相关资产页面		详情 调整策略

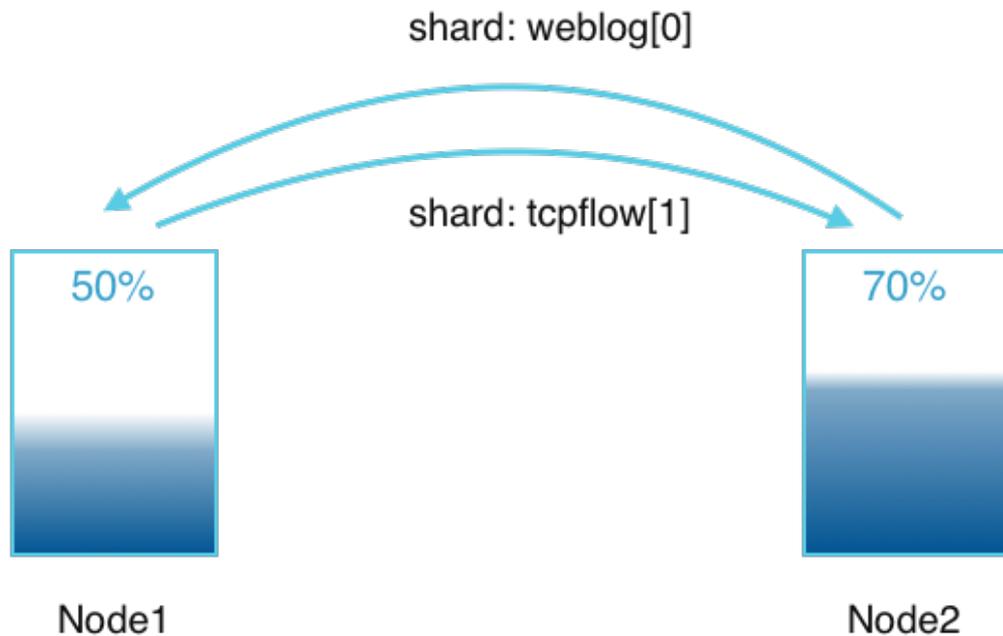
应用场景-态势感知

多维度安全事件监测、预测、回溯等。存储攻击情况，漏洞信息



内核改进

节点间的磁盘平衡



内存优化

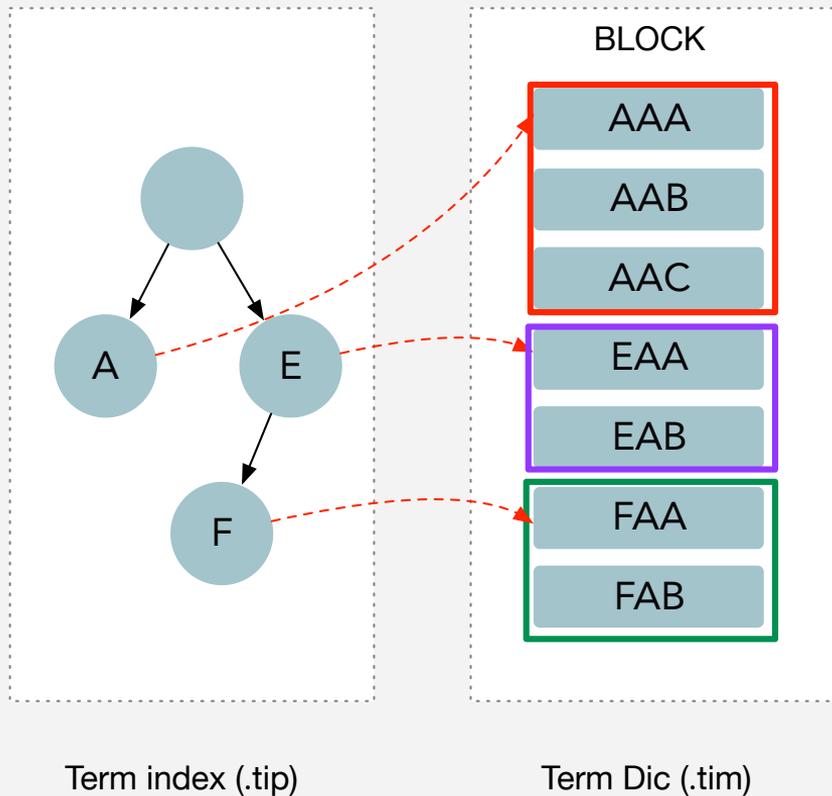
减小 FST 体积

调整 block size

默认值：25-48

修改为：100-198

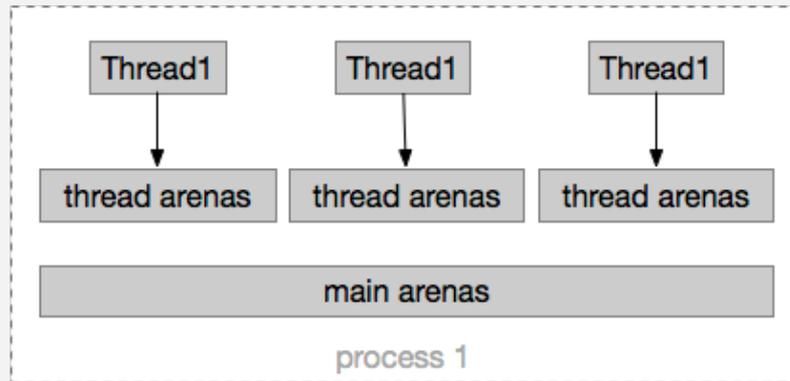
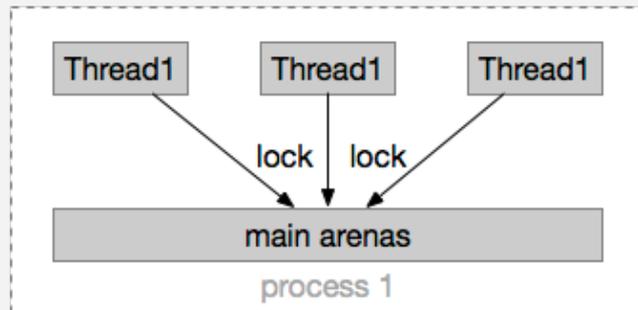
内存占用降低20%+



内存优化

降低堆外内存使用量

```
00002afb2c021000 65404 0 0 --- [ anon ]
00002afb28021000 65404 0 0 --- [ anon ]
00002afb24021000 65404 0 0 --- [ anon ]
00002afb20021000 65404 0 0 --- [ anon ]
00002afb1c021000 65404 0 0 --- [ anon ]
00002afb18021000 65404 0 0 --- [ anon ]
00002af96c021000 65404 0 0 --- [ anon ]
00002af958021000 65404 0 0 --- [ anon ]
00002af954021000 65404 0 0 --- [ anon ]
00002af950021000 65404 0 0 --- [ anon ]
00002af94c021000 65404 0 0 --- [ anon ]
00002af934021000 65404 0 0 --- [ anon ]
00002af930021000 65404 0 0 --- [ anon ]
```



glibc >= 2.10

64位系统: cpu cores*8*64M ~ 12GB

export MALLOC_ARENA_MAX=1

其他小改进

节点内磁盘均衡

- 以剩余可用空间作为权重，加权轮询决定分片放置到哪个磁盘

分批查询

- `search.shard_count.limit`
- `max_concurrent_shard_requests`

提升主节点效率

- 并发写索引和分片 state
- 达到 `recovery` 并发上限时跳过分片分配

生产环境建议

生产环境建议

使用方式要合理

- 尽量分离角色部署
- 尽量避免使用通配查询
- 尽量避免业务直接读写 ES
- scroll 并发别太大
-

调整默认配置

- `node_left.delayed_timeout`
- `routing.allocation.enable`
- `total_shards_per_node`
- `destructive_requires_name`

版本选择

- 坚持持续升级
- 生产环境延迟升级

重点监控

Request QPS

indexing/search

ThreadPool

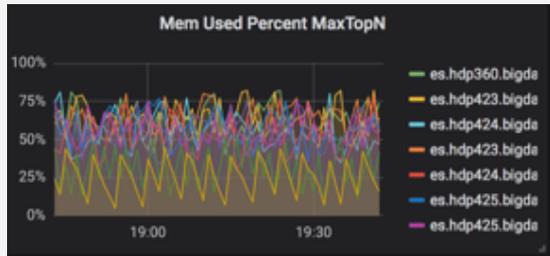
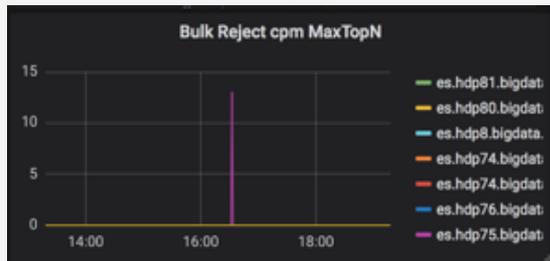
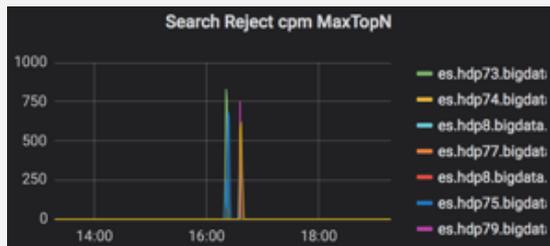
active/queue/reject

SlowLog

indexing/query/fetch

Memory

used percent/segments memory



Thank you!

