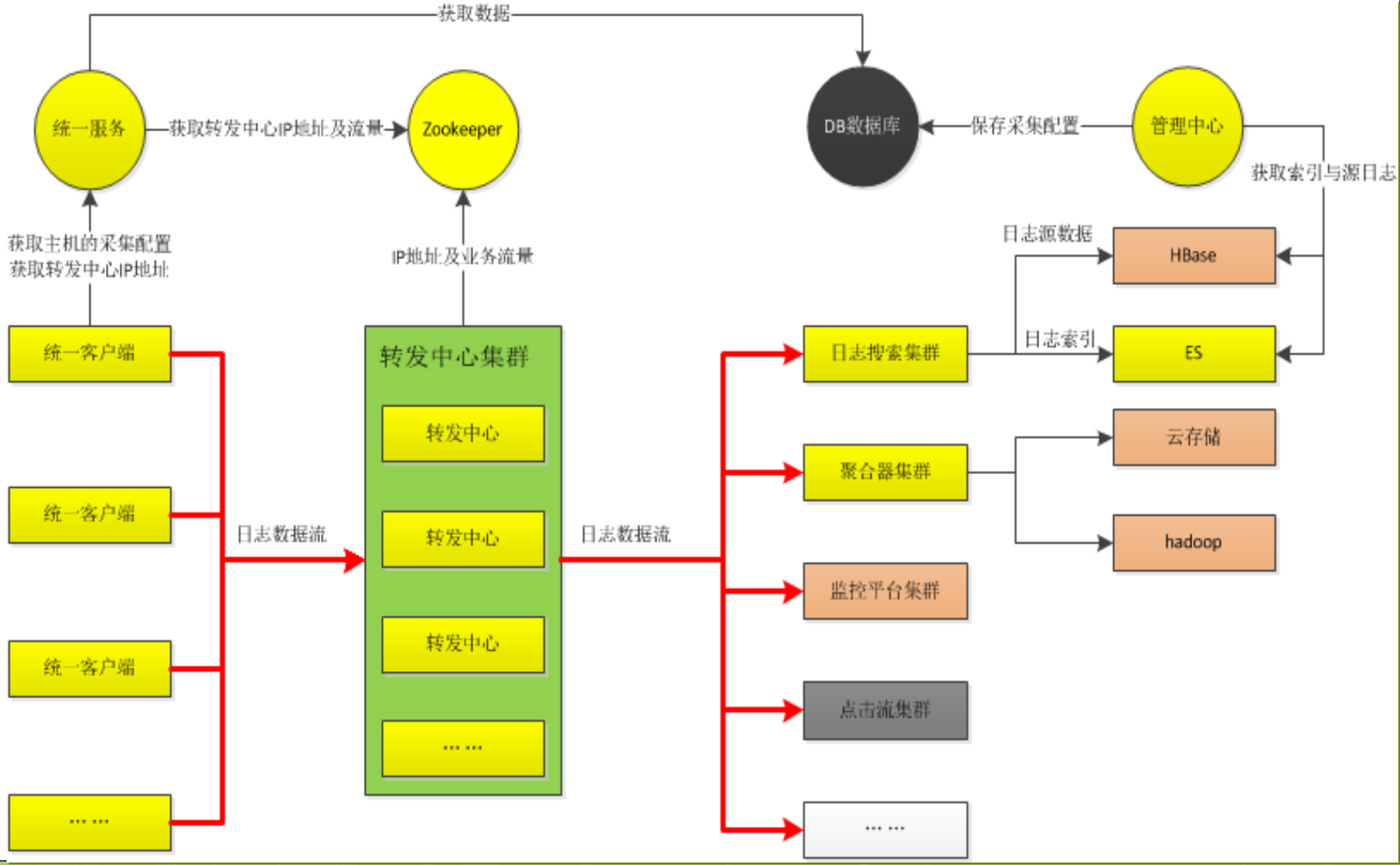
基于es构建实时日志检索平台

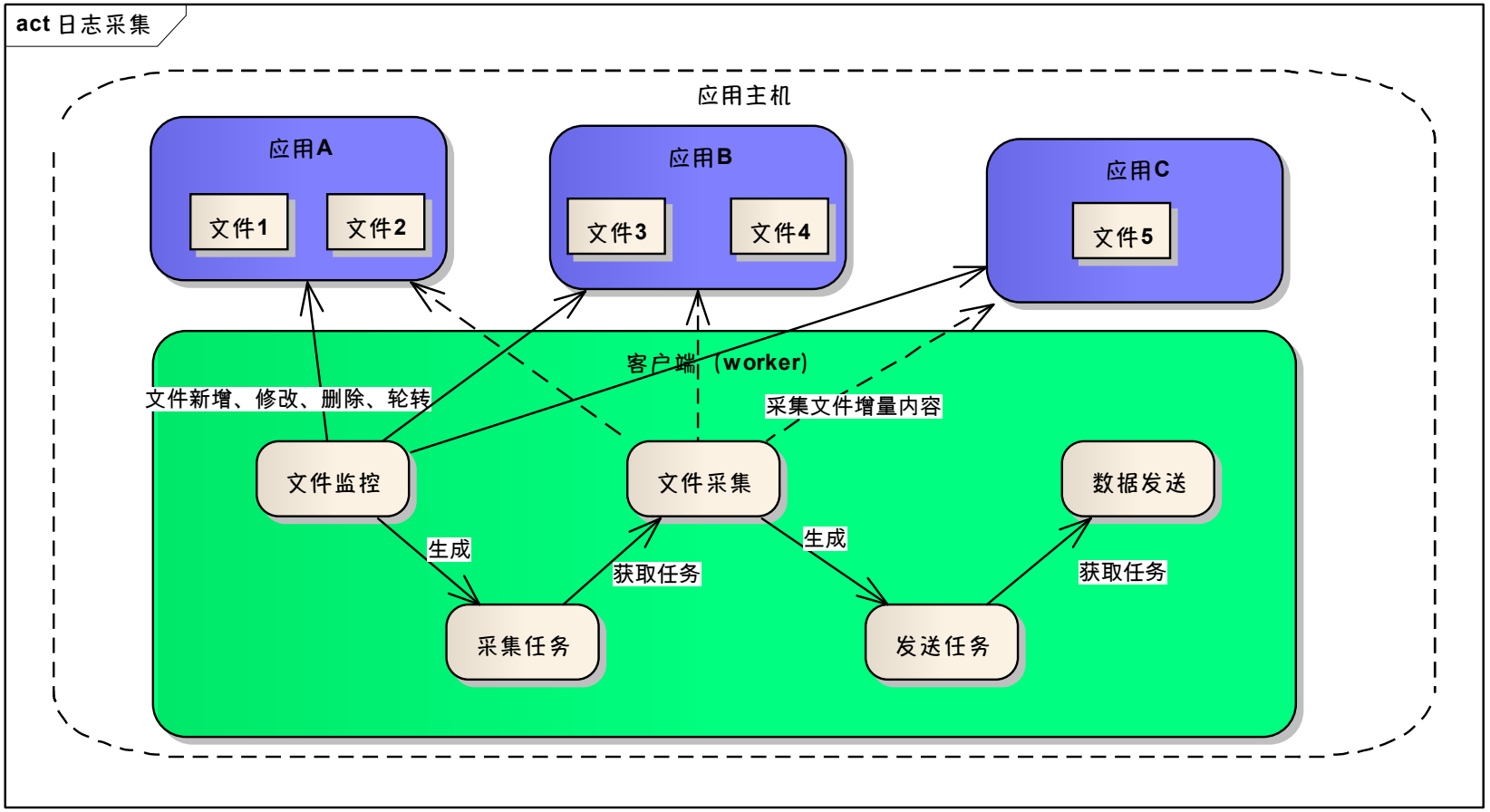
系统架构


获取数据
统一服务
获取转发中心IP地址及流量
Zookeeper
DB数据库
保存采集配置
管理中心
获取索引与源日志
获取主机的采集配置
获取转发中心IP地址
IP地址及业务流量
日志源数据
HBase
统一客户端
统一客户端
统一客户端
… …
日志数据流
转发中心集群
转发中心
转发中心
转发中心
… …
日志数据流
日志搜索集群
日志索引
ES
聚合器集群
云存储
hadoop
监控平台集群
点击流集群
… …

核心技术

- 日志采集
- 日志转发
- 日志搜索
- ES优化

日志采集


act 日志采集
应用主机
应用A
文件1
文件2
应用B
文件3
文件4
应用C
文件5
客户端（worker）
文件新增、修改、删除、轮转
采集文件增量内容
文件监控
文件采集
数据发送
生成
获取任务
生成
获取任务
采集任务
发送任务

日志采集

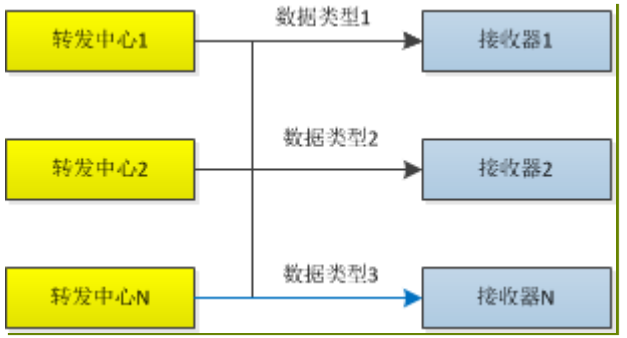
- **文件监控**
  - 监控文件的新增、删除、修改、备份事件，及时响应并生成采集任务
  - **有限资源的最大化利用**：根据文件的变更频率动态调整采集任务的采集频率，如：文件超过1小时未变更，则每分钟执行一次文件状态监控，文件变更频率较高的，每秒钟执行一次文件状态监控并生成采集任务
- **文件采集**
  - **采集内容的准确性**：摘要算法—文件唯一标识，在文件备份事件发生时，保证文件内容的连续性、准确性
- **数据发送**
  - **流量均衡**：通过统一调度中心（统一服务）获取流量最小的转发中心地址，并将数据发送到该转发中心，确保各转发中心的流量均衡。

日志转发

- 转发中心：日志生产方与日志消费方的桥梁
- 四种发送策略

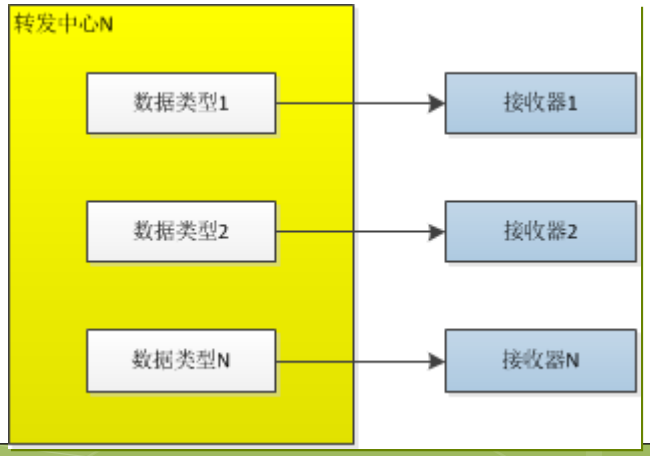
➢ **基于日志分类的流量均衡发送**

- 同一段时间内，转发中心集群所有主机将同一种数据类型发送到接收器集群的同一个IP，比如：HDFS或JSS的接收器



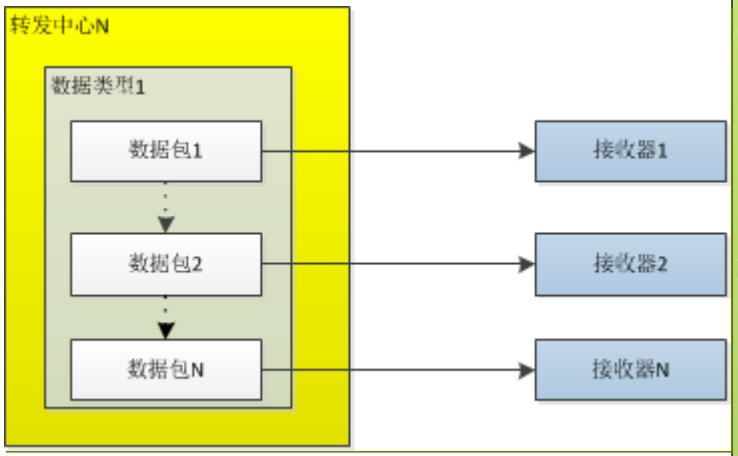
➢ **基于日志分类的轮询发送**

- 同一数据类型在同一转发器上一段时间内只往接收器的一个IP发送，不同转发器可以发送到不同的IP上，比如日志查询接收器

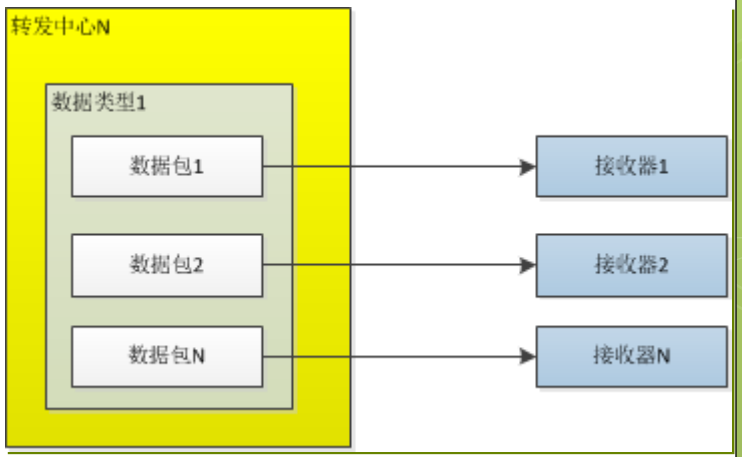


日志转发

- **基于IP地址的轮询发送**
  - 向接收器的所有IP轮询发送数据，如果某IP不可用，将被剔除轮询队列，等到监测OK后，又自动添加到轮询发送数据IP队列中，比如UMP心跳数据

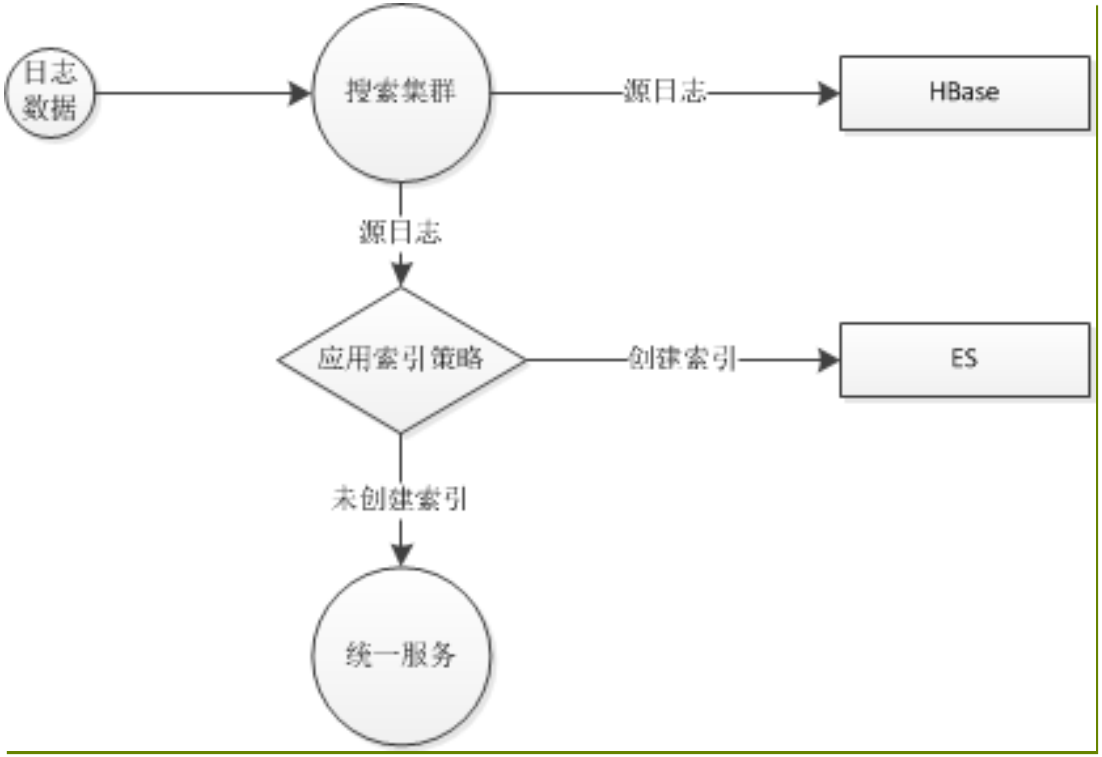


- **基于IP地址的并发发送**
  - 同一种数据并发发送到指定接收器，比如：UMP的JVM数据



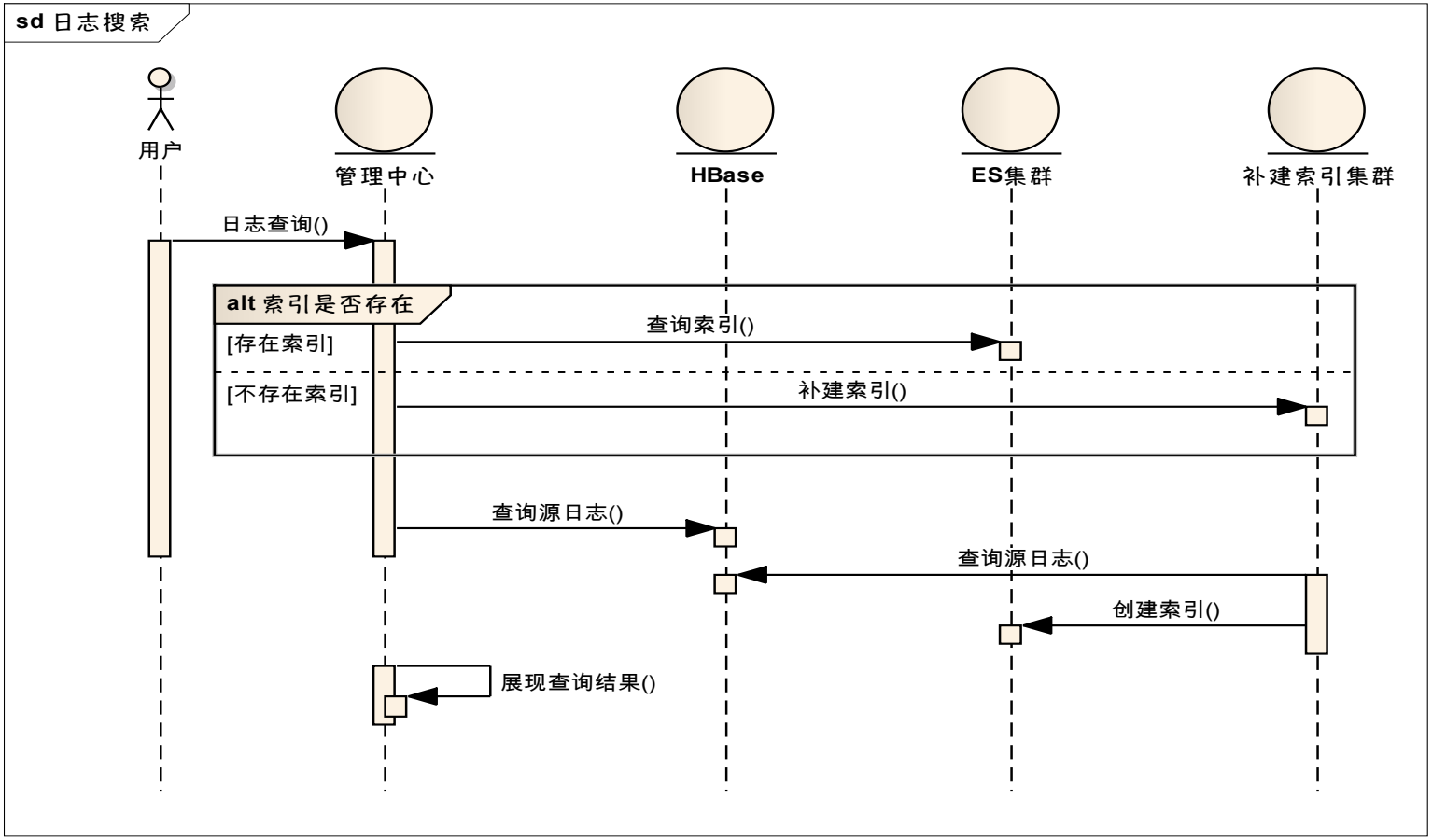
支持将一份数据同时转发到多个消费方（接收器）

日志搜索



应用索引策略：减少日志创建索引压力的有效手段
按应用所属的系统级别定制索引策略

日志搜索


sd 日志搜索
用户
管理中心
HBase
ES集群
补建索引集群
日志查询()
alt 索引是否存在
[存在索引]
查询索引()
[不存在索引]
补建索引()
查询源日志()
查询源日志()
创建索引()
展现查询结果()

ES优化

- **内存优化**
- 有限的机器资源第一个跳出来出问题是内存，lucene查询与建索引的模块jvm 分配16G内存，经常out of memory。
- 优化方向：
- 1.lucene cache使用方式调整为filter cache
- 2.gc 优化

ES优化

- ***lucene merge*的困扰**
- 有限的机器资源第二个跳出来出问题是cpu，24 core 机器 cpu load达到200，最高500。
- 问题根源：
- 1.Lucene 索引同时只能由一个线程执行写操作
- 2.lucene要根据条件进行索引段合并(merge)，以提高查询效率
- 当如此大的数据量一旦触发merge滚雪球效应的时候，可能会持续几分钟。导致后续发送过来的数据等待写入，同时数据也会挤压在内存，也会导致内存问题。

ES优化

- ***lucene merge*的困扰**
- 优化方向：

1.将一次长merge合并操作尽量分散在多次merge合并操作中调整lucene merge操作相关参数

2.将日志量大的应用分布到多个lucene 索引中，同时避免不同大应用分布到相同lucene 索引上。

调整参数	说明	调整前值	调整后值
mergeFactor	当大小几乎相当的段的数量达到此值的时候，开始合并	20	10
maxMergeSize	当一个段的大小大于此值的时候，就不再参与合并	4G	3G

谢谢