

cutd

博客园 首页 新随笔 联系 订阅 管理

随笔 - 30 文章 - 0 评论 - 4

ElasticSearch性能优化官方建议

公告

ES 手册

如何提高ES的性能

不要返回较大的结果集

ES是设计成一个搜索引擎的，只擅长返回匹配查询较少文档，如果需要返回非常多的文档需要使用Scroll。

避免稀疏

因为ES是基于Lucene来索引和存储数据的，所以对稠密的数据更有效。Lucene能够有效的确定文档是通过一个整数的文档id，无论有没有数据都会花费一个字节的存储空间。稀疏主要影响norms和doc_values，一些可以避免稀疏的推荐：

避免将不相关的数据放到相同的索引中

规范的文档结构

使用相同的字段名来保存同样的数据。

避免类型

不用norms和doc_values在稀疏字段

调整索引速度

使用bulk请求

并且每个请求不超过几十M，因为太大会导致内存使用过大

使用 multiple workers/threads发送数据到ES

多进程或者线程，如果看到 `TOO_MANY_REQUESTS (429)` 和 `EsRejectedExecutionException` 则说明ES跟不上索引的速度，当集群的I/O或者CPU饱和就得到了工作者的数量。

增加刷新间隔

`index.refresh_interval` 默认是1s，可以改成30s以减少合并压力。

在加载大量数据时候可以暂时不用refresh和replicas

`index.refresh_interval` to -1 and `index.number_of_replicas` to 0

禁用swapping

禁用swapping

给文件缓存分配内存

缓存是用来缓存I/O操作的，至少用一般的内存来运行ES文件缓存。

使用更快的硬件

- 使用SSD作为存储设备。
- 使用本地存储，避免使用NFS或者SMB
- 注意使用虚拟存储，比如亚马逊的EBS

索引缓冲大小

`indices.memory.index_buffer_size` 通常是JVM的0.1，确保他足够处理至多512MB的索引。

调整搜索速度

给文件系统缓存大内存

至少给可用内存的一半到文件系统缓存。

使用更快的硬件

- 使用SSD作为存储设备。
- 使用性能更好的CPU，高并发
- 使用本地存储，避免使用NFS或者SMB
- 注意使用虚拟存储，比如亚马逊的EBS

文档建模

避免链接，嵌套会使查询慢几倍，而亲自关系能使查询慢几百倍，所以如果同样的问题可以通过没有链接的非规范回答就可以提升速度。

预索引数据

不知不觉

映射

数值型数据不一定要映射成整形或者长整型

避免scripts

如果实在要使用，就用painless和expressions

强势合并只读索引

<https://www.elastic.co/guide/en/elasticsearch/reference/master/indices-forcemerge.html>

不要强势合并正在写的索引

准备全局顺序

准备文件系统缓存

`index.store.preload`，如果内存不是很大会使搜索变得缓慢。

调整磁盘使用

禁用不需要的功能

- 不需要过滤时可以禁用索引 `"index": false`
- 如果你不需要text字段的score，可以禁用 `"norms": false`
- 如果不需要短语查询可以不索引positions `"index_options": "freqs"`

不用默认的动态字符串匹配

不要使用 `_all`

使用 `best_compression`

使用最小的足够用的数值类型

byte,short,integer,long

half_float,float,double

<https://www.elastic.co/guide/en/elasticsearch/reference/master/indices-create-index.html#mappings>

<https://www.elastic.co/guide/en/elasticsearch/reference/master/index-modules.html#dynamic-index-settings>

<https://www.elastic.co/guide/en/elasticsearch/reference/master/search-request-scroll.html>

posted @ 2016-08-23 20:44 cutd 阅读(...) 评论(...) 编辑 收藏

刷新评论 刷新页面 返回顶部