**Elasticsearch笔记**

# 简介

Elasticsearch是一个实时的分布式搜索和分析引擎。它可以帮助你用前所未有的速度去处理大规模数据。

# 使用案例

* 维基百科使用Elasticsearch来进行全文搜做并高亮显示关键词，以及提供search-as-you-type、did-you-mean等搜索建议功能。
* 英国卫报使用Elasticsearch来处理访客日志，以便能将公众对不同文章的反应实时地反馈给各位编辑。
* StackOverflow将全文搜索与地理位置和相关信息进行结合，以提供more-like-this相关问题的展现。
* GitHub使用Elasticsearch来检索超过1300亿行代码。
* 每天，Goldman Sachs使用它来处理5TB数据的索引，还有很多投行使用它来分析股票市场的变动。

# 优缺点

优点

* Elasticsearch是分布式的。不需要其他组件，分发是实时的，被叫做”Push replication”。
* Elasticsearch 完全支持 Apache Lucene 的接近实时的搜索。
* 处理多租户（multitenancy）不需要特殊配置，而Solr则需要更多的高级设置。
* Elasticsearch 采用 Gateway 的概念，使得完全备份更加简单。
* 各节点组成对等的网络结构，某些节点出现故障时会自动分配其他节点代替其进行工作。

缺点

* 只有一名开发者（当前Elasticsearch GitHub组织已经不只如此，已经有了相当活跃的维护者）
* 还不够自动（不适合当前新的Index Warmup API）

# Elasticsearch和Solr对比

* 二者安装都很简单；
* Solr 利用 Zookeeper 进行分布式管理，而 Elasticsearch 自身带有分布式协调管理功能;
* Solr 支持更多格式的数据，而 Elasticsearch 仅支持json文件格式；
* Solr 官方提供的功能更多，而 Elasticsearch 本身更注重于核心功能，高级功能多有第三方插件提供；
* Solr 在传统的搜索应用中表现好于 Elasticsearch，但在处理实时搜索应用时效率明显低于 Elasticsearch。

Solr 是传统搜索应用的有力解决方案，但 Elasticsearch 更适用于新兴的实时搜索应用。

# Elasticsearch安装

## 下载地址

elasticsearch.org/download

## 解压运行

将下载的zip包解压,然后运行bin目录下的elasticsearch.bat即可

## 集群和节点

节点是一个运行着的Elasticsearch实例,集群是一组具有相同cluster.name的节点集合

最好指定一个cluster.name来代替默认值,以防止一个新启动的节点加入到相同网络中的另一个同名的集群中.

# 与Elasticsearch交互方式

## Java API客户端

java客户端通过9300端口与服务端进行通信

### 节点客户端

节点客户端以无数据节点身份加入集群,它自己不存储任何数据,但它知道数据在集群中的具体位置,并且能够直接转发请求到相应的节点上.

### 传输客户端

这个更轻量的传输客户端能够发送请求到远程集群.它自己不加入集群,只是简单转发请求给集群中的节点.

## 基于http协议的客户端

除了Java以外的所有程序语言都可以使用RESTful API与Elasticsearch进行通信

基于http协议的客户端通过9200端口与服务端通信

基于http协议的客户端以json为数据交互格式

举个以curl命令与Elasticsearch进行通信的例子

curl -X<VERB> '<PROTOCOL>://<HOST>:<PORT>/<PATH>?<QUERY\_STRING>' -d '<BODY>'

VERB HTTP方法:GET ,POST,PUT,HEAD,DELETE

BODY 一个json格式的请求主体(有些请求不需要该主体)