章节二 elasticSearch 6.2版本基础讲解到阿里云部署实战

2、搜索引擎知识介绍和相关框架

简介：介绍搜索的基本概念，市面上主流的搜索框架elasticSearch和solr等对比

什么是搜索：在海量信息中获取我们想要的信息

传统做法：

1、文档中使用系统的Find查找

2、mysql中使用like模糊查询

问题：

1、海量数据中不能及时响应,少量数据可以通过传统的MySql建立索引解决

2、一些无用词不能进行过滤，没法分词

3、数据量大的话难以拓展

4、相同的数据难以进行相似度最高的进行排序

搜索引擎:

1、存储非结构化的数据

2、快速检索和响应我们需要的信息，快-准

3、进行相关性的排序，过滤等

4、可以去掉停用词(没有特殊含义的词，比如英文的a,is等，中文： 这，的，是等)，框架一般支持可以自定义停用词

常用框架：

1、Lucene

Apache下面的一个开源项目，高性能的、可扩展的工具库，提供搜索的基本架构；

如果开发人员需用使用的话，需用自己进行开发,成本比较大，但是性能高

2、solr

Solr基于Lucene的全文搜索框架，提供了比Lucene更为丰富的功能，

同时实现了可配置、可扩展并对查询性能进行了优化

建立索引时，搜索效率下降，实时索引搜索效率不高

数据量的增加，Solr的搜索效率会变得更低,适合小的搜索应用，对应java客户端的是solrj

3、elasticSearch

基于Lucene的搜索框架, 它提供了一个分布式多用户能力的全文搜索引擎，基于RESTful web接口

上手容易，拓展节点方便，可用于存储和检索海量数据，接近实时搜索，海量数据量增加，搜索响应性能几乎不受影响；

分布式搜索框架，自动发现节点，副本机制，保障可用性