

1.ElasticSearch简介

1.1 ElasticSearch（简称ES）

Elasticsearch是用Java开发并且是当前最流行的开源的企业级搜索引擎。

能够达到实时搜索，稳定，可靠，快速，安装使用方便。

客户端支持Java、.NET（C#）、PHP、Python、Ruby等多种语言。

官方网站: <https://www.elastic.co/>

下载地址: <https://www.elastic.co/cn/start>

创始人:Shay Banon（谢巴农）



应用场景

长沙好玩的地方



百度一下

长沙有哪些好玩的地方推荐? - 知乎



2020年12月22日 空中悬浮体验 ★★★☆☆3. 西湖公园散步,58小镇(我家旁边) ★★★☆☆4. 大学城各种小吃,爬岳麓山观景,体力不行的可以坐缆车上山,玩滑道下山,鸟语林、欣赏湖...

知乎 百度快照

长沙最好玩的几个地方

名称: 长沙青天寨 服务特色: 项目丰富 游玩类型: 农家乐 现价: 90元起

长沙好玩的地方,2020年本人亲身游玩体验写下此篇旅游攻略,纯玩的经验分享,省心又省钱,多种玩法推荐,不走回头路,各种注意事项,自行游需看,交通便利,环境优美,项目丰富有趣...

长沙青天寨生态农业股份 2021-01 广告 保障

长沙有什么好玩的? - 知乎



2016年12月5日 国内出游必玩指南 1,914 人赞同了该回答 长沙并不是一个旅游城市,更多时候长沙被作为革命教育的地方,代表着人民对毛泽东的追思,对长沙人来说长沙有的也是情怀,而...

知乎 百度快照

长沙好玩的地方-2021长沙有什么好玩的地方推荐-去哪儿攻略



去哪儿攻略为您推荐长沙好玩的地方、长沙适合年轻人好玩的地方,2021长沙有什么好玩的地方快来看去哪儿攻略的推荐吧。

travel.qunar.com/p-cs300022-ch... 保障 百度快照

长沙有哪些好玩的地方 - 百度知道



18个回答 - 回答时间: 2020年11月11日

最佳答案: 1、岳麓山 橘子洲和岳麓山,是长沙最值得去的两大景点,因为疫情期间岳麓书院没有开,我还小纠结了一会,要不要在这次行...

searchjd.com/Search?keyword=华为手机&enc=utf-8&wq=华为手机&pvid=e593f2f271264a45a859a6f1067e8443

手机 资料 白起老师 百度一下,你就知道

| 商品名称 | 价格 | 品牌/型号 | 配置 | 评价 | 店铺 |
|---------------------|----------|-------|--|----------|-----------|
| 畅享20 Pro 5G手机 | ¥2599.00 | 华为 | 6.67英寸全高清大屏, 5000mAh大电池, 幻夜黑 (8GB+128GB) | 1000+条评价 | 聚德联盛手机旗舰店 |
| 畅享20 SE手机 | ¥1899.00 | 华为 | 6.67英寸全高清大屏, 5000mAh大电池, 幻夜黑 (8GB+128GB) | 1000+条评价 | 聚德联盛手机旗舰店 |
| 荣耀play4T 全网通手机 | ¥1499.00 | 荣耀 | 6.6英寸全高清大屏, 4000mAh大电池, 标准版 (6GB+128GB) | 3000+条评价 | 聚德联盛手机旗舰店 |
| nova 5i 手机 | ¥2099.00 | 华为 | 6.6英寸全高清大屏, 4000mAh大电池, 标准版 (6GB+128GB) | 3000+条评价 | 聚德联盛手机旗舰店 |
| 荣耀Play3e 智能老人老年备用手机 | ¥958.00 | 荣耀 | 3020mAh长续航, 1300万像素摄像头, 幻夜黑 | 1万+条评价 | 伟德手机专营店 |

1.2 ElasticSearch与Lucene的关系

Lucene可以被认为迄今为止最先进、性能最好的、功能最全的搜索引擎库（框

架)

但是想要使用Lucene，必须使用Java来作为开发语言并将其直接集成到你的应用中，并且Lucene的配置及使用非常复杂，你需要深入了解检索的相关知识来理解它是如何工作的。

Lucene缺点：

- 1) 只能在Java项目中使用,并且要以jar包的方式直接集成项目中.
- 2) 使用非常复杂-创建索引和搜索索引代码繁杂
- 3) 不支持集群环境-索引数据不同步（不支持大型项目）
- 4) 索引数据如果太多就不行，索引库和应用所在同一个服务器,共同占用硬盘,共用空间少.

上述Lucene框架中的缺点,ES全部都能解决.

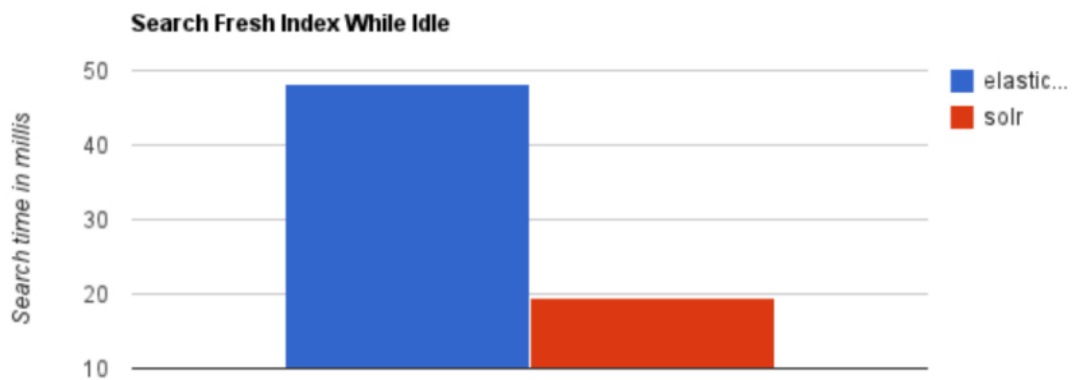
1.3 哪些公司在使用Elasticsearch

- 1 1. 京东
- 2 2. 携程
- 3 3. 去哪儿
- 4 4. 58同城
- 5 5. 滴滴
- 6 6. 今日头条
- 7 7. 小米
- 8 8. 哔哩哔哩
- 9 9. 联想
- 10 10. GitHub
- 11 11. 微软
- 12 12. Facebook
- 13 等等...

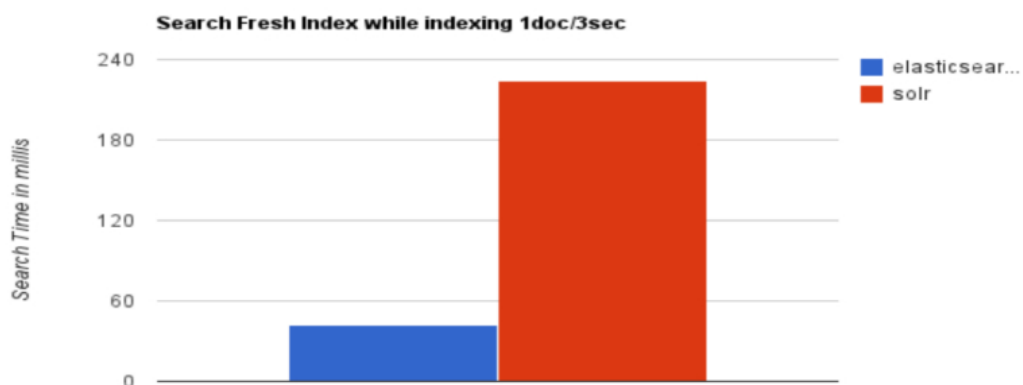
1.4 ES vs Solr比较

1.4.1 ES vs Solr 检索速度

当单纯的对已有数据进行搜索时，Solr更快。



当实时建立索引时, Solr会产生io阻塞, 查询性能较差, Elasticsearch具有明显的优势。



大型互联网公司, 实际生产环境测试, 将搜索引擎从Solr转到 Elasticsearch以后的平均查询速度有了50倍的提升。



总结:

二者安装都很简单。

1、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。

2、Solr 支持更多格式的数据，比如JSON、XML、CSV，而 Elasticsearch 仅支持 json文件格式。

3、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。

4、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。

1.4.2 ES vs 关系型数据库

| | | | | |
|----------------------|--------------------------|---------------------|-------------------------|----------------------|
| 关系型数据库 | Database (数据库) | Table (表) | Row (行) | Column (列) |
| ElasticSearch | Index (索引库) | Type (类型) | Document (文档) | Field (字段) |

2. Lucene全文检索框架

2.1 什么是全文检索

全文检索是指：

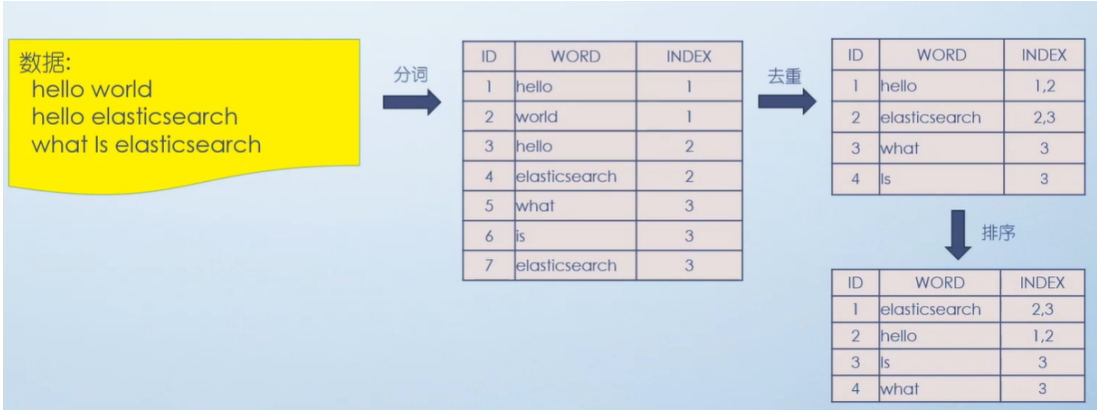
- 通过一个程序扫描文本中的每一个单词，针对单词建立索引，并保存该单词在文本中的位置、以及出现的次数
- 用户查询时，通过之前建立好的索引来查询，将索引中单词对应的文本位置、出现的次数返回给用户，因为有了具体文本的位置，所以就可以将具体内容读取出来了

hello what world ==> hello

what

world

2.2 分词原理之倒排索引



| 文档编号 | 文档内容 |
|------|-----------------------------|
| 1 | 谷歌地图之父跳槽Facebook |
| 2 | 谷歌地图之父加盟Facebook |
| 3 | 谷歌地图创始人拉斯离开谷歌加盟Facebook |
| 4 | 谷歌地图之父跳槽Facebook与Wave项目取消有关 |
| 5 | 谷歌地图之父拉斯加盟社交网站Facebook |

| 单词ID | 单词 | 倒排列表 |
|------|----------|-------------------------------|
| 1 | 谷歌 | (1;1),(2;1),(3;2),(4;1),(5;1) |
| 2 | 地图 | (1;1),(2;1),(3;1),(4;1),(5;1) |
| 3 | 之父 | (1;1),(2;1),(4;1),(5;1) |
| 4 | 跳槽 | (1;1),(4;1) |
| 5 | Facebook | (1;1),(2;1),(3;1),(4;1),(5;1) |
| 6 | 加盟 | (2;1),(3;1),(5;1) |
| 7 | 创始人 | (3;1) |
| 8 | 拉斯 | (3;1),(5;1) |
| 9 | 离开 | (3;1) |
| 10 | 与 | (4;1) |
| 11 | Wave | (4;1) |
| 12 | 项目 | (4;1) |
| 13 | 取消 | (4;1) |
| 14 | 有关 | (4;1) |
| 15 | 社交 | (5;1) |
| 16 | 网站 | (5;1) |

倒排索引总结：

索引就类似于目录，平时我们使用的都是索引，都是通过主键定位到某条数据，那么倒排索引呢，刚好相反，数据对应到主键。这里以一个博客文章的内容为例：

1.索引

| 文章ID | 文章标题 | 文章内容 |
|------|-------------|------------------------------|
| 1 | 浅析JAVA设计模式 | JAVA设计模式是每一个J 员都应该掌握的进阶知识 |
| 2 | JAVA多线程设计模式 | JAVA多线程与设计模式结 |

2.倒排索引

假如，我们有一个站内搜索的功能，通过某个关键词来搜索相关的文章，那么这个关键词可能出现在标题中，也可能出现在文章内容中，那我们将会在创建或修改文章的时候，建立一个关键词与文章的对应关系表，这种，我们可以称之为倒排索引，因此倒排索引，也可称之为反向索引。如：

| 关键词 | 文章ID |
|------|------|
| JAVA | 1 |
| 设计模式 | 1,2 |
| 多线程 | 2 |

注：这里涉及中文分词的问题

3. Elasticsearch中的核心概念

3.1 索引 index

一个索引就是一个拥有几分相似特征的文档的集合。比如说，可以有一个客户数据的索引，另一个产品目录的索引，还有一个订单数据的索引

一个索引由一个名字来标识（必须全部是小写字母的），并且当我们要对对应于这个索引中的文档进行索引、搜索、更新和删除的时候，都要使用到这个名字

3.2 映射 mapping

ElasticSearch中的映射（Mapping）用来定义一个文档

mapping是处理数据的方式和规则方面做一些限制，如某个字段的数据类型、默认值、分词器、是否被索引等等，这些都是映射里面可以设置的

3.3 字段Field

相当于是数据表的字段 | 列

3.4 字段类型 Type

每一个字段都应该有一个对应的类型，例如：Text、Keyword、Byte等

3.5 文档 document

一个文档是一个可被索引的基础信息单元，类似一条记录。文档以JSON（Javascript Object Notation）格式来表示；

3.6 集群 cluster

一个集群就是由一个或多个节点组织在一起，它们共同持有整个的数据，并一起提供索引和搜索功能

3.7 节点 node

一个节点是集群中的一个服务器，作为集群的一部分，它存储数据，参与集群的索引和搜索功能

一个节点可以通过配置集群名称的方式来加入一个指定的集群。默认情况下，每个节点都会被安排加入到一个叫做“elasticsearch”的集群中

这意味着，如果在网络中启动了若干个节点，并假定它们能够相互发现彼此，它们将会自动地形成并加入到一个叫做“elasticsearch”的集群中

在一个集群里，可以拥有任意多个节点。而且，如果当前网络中没有运行任何Elasticsearch节点，这时启动一个节点，会默认创建并加入一个叫做“elasticsearch”的集群。

3.8 分片和副本 shards&replicas

3.8.1 分片

- 一个索引可以存储超出单个结点硬件限制的大量数据。比如，一个具有10亿文档的索引占据1TB的磁盘空间，而任一节点都没有这样大的磁盘空间；或者单个节点处理搜索请求，响应太慢
- 为了解决这个问题，Elasticsearch提供了将索引划分成多份的能力，这些份就叫做分片
- 当创建一个索引的时候，可以指定你想要的分片的数量
- 每个分片本身也是一个功能完善并且独立的“索引”，这个“索引”可以被放置到集群中的任何节点上
- 分片很重要，主要有两方面的原因
 - 允许水平分割/扩展你的内容容量
 - 允许在分片之上进行分布式的、并行的操作，进而提高性能/吞吐量
- 至于一个分片怎样分布，它的文档怎样聚合回搜索请求，是完全由Elasticsearch管理的，对于作为用户来说，这些都是透明的

3.8.2 副本

- 在一个网络/云的环境里，失败随时都可能发生，在某个分片/节点不知怎么的就处于离线状态，或者由于任何原因消失了，这种情况下，有一个故障转移机制是非常有用并且是强烈推荐的。为此目的，Elasticsearch允许你创建分片的一份或多份拷贝，这些拷贝叫做副本分片，或者直接叫副本

- 副本之所以重要，有两个主要原因

1) 在分片/节点失败的情况下，提供了高可用性。

注意到复制分片从不与原/主要（original/primary）分片置于同一节点上是非常重要的

2) 扩展搜索量/吞吐量，因为搜索可以在所有的副本上并行运行

每个索引可以被分成多个分片。一个索引有0个或者多个副本

一旦设置了副本，每个索引就有了主分片和副本分片，分片和副本的数量可以在索引

创建的时候指定

在索引创建之后，可以在任何时候动态地改变副本的数量，但是不能改变分片的数量