



《手写OS操作系统》小班二期招生，全程直播授课，大牛带你掌握硬核技术！

点此查看

从所有教程的词条中查询...

全部开发者教程

第23周 protoc插件开发、cobra命令行

1. protoc调试源码

2. protoc自定义gin插件

第24周 log日志包设计

日志源码

第25周 ast代码生成工具开发

错误码

第26周 三层代码结构

通用app项目启动



bobby · 更新于 2022-11-16

6. userop-web... 2. elasticsearch... 下一节

1. mysql搜索面临的问题

- 1. 性能低下
- 2. 没有相关性排名 - 刚需
- 3. 无法全文搜索
- 4. 搜索不准确 - 没有分词

2. 什么是全文搜索

我想找简历：只要出现java、go 原本的需求就是只要简历中出现了 go 分布式 elasticsearch 都是我

我们生活中的数据总体分为两种：结构化数据和非结构化数据。

- **结构化数据**：指具有固定格式或有限长度的数据，如数据库，元数据等。
- **非结构化数据**：指不定长或无固定格式的数据，如邮件，word文档等。

非结构化数据又一种叫法叫全文数据。

按照数据的分类，搜索也分为两种：

- 对结构化数据的搜索：如对数据库的搜索，用SQL语句。再如对元数据的搜索，如利用windows搜索对文件名，类型，修改时间进行搜索等。
- 对非结构化数据的搜索：如利用windows的搜索也可以搜索文件内容，Linux下的grep命令，再如用Google和百度可以搜索大量内容数据。

对非结构化数据也即对全文数据的搜索主要有两种方法：

一种是顺序扫描法(Serial Scanning)：所谓顺序扫描，比如要找内容包含某一个字符串的文件，就是一个文档一个文档的看，对于每一个文档，从头看到尾，如果此文档包含此字符串，则此文档为我们要找的文件，接着看下一个文件，直到扫描完所有的文件。如利用windows的搜索也可以搜索文件内容，只是相当的慢。假如有一个80G硬盘，如果想在上面找到一个内容包含某字符串的文件，可能需要几个小时的时间。Linux下的grep命令也是这一种方式。这是一种比较原始的方法，但对于小数据量的文件，这种方法还是最直接，最方便的。但是对于大量的文件，这种方法的速度就很慢。

另一种是全文检索(Full-text Search)：即先建立索引，再对索引进行搜索。索引是从非结构化数据中提取出之后重新组织的信息。

3. 什么是elasticsearch

意见反馈

收藏教程

标记书签



Elasticsearch 是一个分布式可扩展的实时搜索和分析引擎,一个建立在全文搜索引擎 Apache Lucene™ 基础上的搜索引擎.当然 Elasticsearch 并不仅仅是 Lucene 那么简单，它不仅包括了全文搜索功能，还可以进行以下工作:

- 分布式实时文件存储，并将每一个字段都编入索引，使其可以被搜索。
- 实时分析的分布式搜索引擎。
- 可以扩展到上百台服务器，处理PB级别的结构化或非结构化数据。

ES的适用场景

- 维基百科
- The Guardian、新闻
- Stack Overflow
- Github
- 电商网站、检索商品
- 日志数据分析、logstash采集日志、ES进行复杂的数据分析(ELK)
- 商品价格监控网站、用户设定价格阈值
- BI系统、商业智能、ES执行数据分析和挖掘

ES特点

1. 可以作为一个大型的分布式集群(数百台服务器)技术，处理PB级数据，服务大公司，可以运行在单机上，服务小公司。
2. ES不是什么新技术，主要是将全文检索、数据分析以及分布式技术合并在一起，才形成了独一无二的ES.lucene（全文检索）、商用的数据分析软件、分布式数据库 (mycat)
3. 对用户而言，是开箱即用，非常简单，作为中小型的应用，直接3分钟部署ES，就可以作为生产环境的系统使用，数据量不大，操作不是很复杂。
4. 数据库的功能面对很多领域是不够用的(事务，还有各种联机事务的操作)：特殊的功能，比如全文检索、同义词处理、相关度排名、复杂数据分析、海量数据近实时处理；ES作为传统数据库的一个补充，提供了数据库所不能提供的很多功能。

