

0-es调研学习

Created by 杨超, last modified on 2018 Aug 13

- 0 elasticsearch是什么?
- 1 倒排索引
- 2 特性
- 3 接口试用

本身es的语法和用途可谓庞大，短期无法全部涉猎，这里只是简述基本用法

0 elasticsearch是什么?

本身elasticsearch 【以下简称es】是个搜索服务器，个人认为可以简单理解为 **分布式的 支持搜索的 数据库服务器**

1 倒排索引

es也是使用倒排索引作为底层，原理大致如下图所示，不过实际上 的倒排表比这个复杂很多，比如term里可能有词向量 用来进行相似词的匹配等

doc 文档
field 字段
term 词项
token 词条

倒排 inverted index

term	Count	doc
solar	1	<3>
Apache	1	<2>
Elasticsearch	2	<1><2>

2 特性

或者说好处，可以解决之前opensearch里不支持的功能，优化搜索的体验和实现成本

1 match_all 可以支持select all的这种功能，比起opensearch好太多

2 支持多应用搜索 这样方便实现

简单来说，有两个表 一个是人 一个是项目，过去只能一次搜一个，但是es支持两个放在一起进行搜索，可以很好的支持综合搜索这种功能

3 方便客户端自定义字段的权重排序

fields 【title^100, text^1】，如左所示，可以在调用的时候 将权重写在参数里，这样就可以实现不同权重字段排序的动态调节

4 相似搜索更加方便

主要是针对一句比较长的话的搜索，可能在内部工作台 匹配 抓取到的文案 的时候用处比较大

3 接口试用

es内部使用java实现，所以原生支持java Api，不过也提供了Rest API提供了供其他语言使用

个人试用的感觉，restful的接口还是蛮好用的

给个简单的例子，搜索user 关键词为kimchy

```
curl -XGET 'http://localhost:9200/twitter/_search?pretty=true' -H 'Content-Type: application/json' -d '{
  "query" : {
    "match" : { "user": "kimchy" }
  }
}'
```

结果

```
{
  "took" : 143,
  "timed_out" : false,
  "_shards" : {
    "total" : 5,
    "successful" : 5,
    "skipped" : 0,
    "failed" : 0
  },
  "hits" : {
    "total" : 2,
    "max_score" : 0.2876821,
```

```
"hits" : [
  {
    "_index" : "twitter",
    "_type" : "doc",
    "_id" : "2",
    "_score" : 0.2876821,
    "_source" : {
      "user" : "kimchy",
      "post_date" : "2009-11-15T14:12:12",
      "message" : "Another tweet, will it be indexed?"
    }
  },
  {
    "_index" : "twitter",
    "_type" : "doc",
    "_id" : "1",
    "_score" : 0.2876821,
    "_source" : {
      "user" : "kimchy",
      "post_date" : "2009-11-15T13:12:00",
      "message" : "Trying out Elasticsearch, so far so good?"
    }
  }
]
```

有兴趣的可以读下阮一峰的这篇 <http://www.ruanyifeng.com/blog/2017/08/elasticsearch.html>

Like Be the first to like this

No labels

地址：北京市朝阳区建国路86号佳兆业广场北塔6层梦想加空间601室

以太资本由艾普拉斯投资顾问(北京)有限公司运营，提供早期互联网项目的投融资对接服务

