

ElasticSearch串讲问题:

- 1、地理位置支持，地理位置是指？
- 2、在说到“Term Index的倒排序索引以及Document的字段数据”时，这里的字段数据有哪些？
- 3、中文检索，是根据词还是单个字？
- 4、位置信息存在哪？
- 5、文档ID自动生成或手动指定。
- 6、单个字段的value值是否有限制？文档内容过多（大约10万行）时，是否可以建索引？如何存储和读取？

1.地理位置支持，地理位置是指？

Elasticsearch提供了丰富的搜索和分析功能，地理位置功能可以让我们把基于地理位置的搜索、聚合、计算与全文搜索、结构化搜索和分析结合到一起

2.在说到“Term Index的倒排序索引以及Document的字段数据”时，这里的字段数据有哪些？

这里的字段数据是由用户指定的

3.中文检索，是根据词还是单个字？

官方默认的分词器支持中文采用的方法为单字切分。会将词汇单元转换成小写形式，并去除停用词和标点符号。

可安装中文分词插件elasticsearch-analysis-ik。

es-ik 带有两个分词器

ik_max_word: 会将文本做最细粒度的拆分；尽可能多的拆分出词语

ik_smart: 会做最粗粒度的拆分；已被分出的词语将不会再次被其它词语占有

可在创建索引指定mapping 信息来配置使用哪个分词器，例如

```
mapping = {
  'properties': {
    'content': {
      'type': 'text',
      'analyzer': 'ik_max_word',
      'search_analyzer': 'ik_max_word'
    }
  }
}
```

4、位置信息存在哪？

ES建立索引的过程会先将生成的segments保存在内存中，然后和translog一起写入磁盘中

默认是保存在es_data = /usr/lib/elasticsearch这个文件夹下面

segments_xxx文件记录了lucene包下面的segment文件数量

每个segment会包含如下的文件

Name	Extension	Brief Description
Segment Info	.si	segment的元数据文件
Compound File	.cfs, .cfe	一个segment包含了如下表的各个文件，为减少打开文件的数量，在segment小的时候，segment的所有文件内容都保存在cfs文件中，cfe文件保存了lucene各文件在cfs文件的位置信息
Term Dictionary	.tim	倒排索引的元数据信息
Term Index	.tip	倒排索引文件，存储了所有的倒排索引数据

5、文档ID自动生成或手动指定。

文档ID 可以手动指定，如果没有手动指定那么ES 为自动生成一个ID

6、单个字段的value值是否有限制？文档内容过多（大约10万行）时，是否可以建索引？如何存储和读取？

ES 中对于str字段有两个类型 Keyword和Text类型

text类型：支持分词、全文检索，不支持聚合、排序操作。适合大字段存储，如：文章详情、content字段等；

keyword类型：支持精确匹配，支持聚合、排序操作。

适合精准字段匹配，如：url、name、title等字段。

其中text字段没有大小限制

keyword大小限制默认为32766个UTF-8类型的字符