ElasticSearch基础与应用

1. 基础存储结构

cluster（集群）：由一个或多个ES节点组成的集群

node（节点）：一个运行ES的进程

index（索引）：最顶层的存储单元，相当于数据库中的库，存放着不同的类型数据

type（类型）：相当于数据库中的表，理论上存放着同一类型数据，一个类型也可以存放不同特性的数据，但是那不符合使用的规范，后续也不方便查询。

document（文档）：最底层的单元数据，对应数据库中的每一行，以及用户的每一条数据，ES中的各种条件查询，返回的都是一个或多个文档。

shard（分片）：分片是index的属性，一个index由多个分片组成，多个shard可能存放在不同的节点中，以起到对读写操作的负载均衡作用；shard分为primary和replica，replica是primary的副本，负责容错以及对于读请求的负载均衡。

2. 数据查询方法

ElasticSearch的查询完全遵守restful的风格，通过HTTP的GET、PUT、POST、DELETE等来进行增删改查的操作。

比如：GET <http://localhost:9200/index/type/1>，可以查询index索引，type类型里，id=1的文档。

3. 查询接口\_cat

/\_cat接口可以查询各种与集群相关的状态，比如/\_cat/health可以检查集群中各个节点的状态；/\_cat/nodes可以查询集群中各个节点；/\_cat/plugins可以查询当前ES中安装的插件。

\*直接使用localhost:9200/\_cat接口，可以查看\_cat下的所有查询接口。

4. \_settings

\_settings属性与每个index对应，使用/index/\_settings可以获取到index的settings信息，其中包括这个索引的分片数目、副本数目以及用到的分析器(analyzer)与过滤器(filter)等。

5. analysis分析器

分析器是针对某个索引设定的，作用是分词与规范化文本。

一个分析器由零个或多个字符过滤器与标记过滤器组成，对于用户提交的文本，首先经过字符过滤器，移除一些html标签，转换成大小写等等；接着进行分词操作，并对每个词进行标记；最后经过标记过滤器，去除停用词，或者转换成同义词、词干等。

\*经过了分析器对用户输入文本的分析后，再进行文档的插入，或者对倒排表的查询等，得到最后的查询结果。

\*ES内置了多个预定义的分析器，其中包括standard，simple，whiteapsce分析器等等。标准分析器(Standard Analysis)由标准分词器、标准标记过滤器、小写标记过滤器、停用词标记过滤器组成。

\*使用POST请求，可以显示地调用某个分析器，查看其分析结果，

比如：POST myindex/\_analyze?analyzer=standard -d "After School"，可以显示地调用标准分析器，查看分词与过滤的结果。

6. 查询请求/\_search

使用\_search表示在当前路径下进行查询，比如localhost:9200/\_search就会返回该集群所有索引下的所有文档；localhost:9200/index/type/\_search会返回某索引某类型下的所有文档。

size：分页参数，表示查询结果每一页的数据量

from：页数，表示查询结果显示第几页

pretty：按照标准的JSON格式来显示查询结果，如/\_search?pretty，查询所有文档，并按照JSON格式来显示。

q：查询语句，如/\_search?q=filmName:dds&pretty，查询filmName字段为dds的文档，并按照JSON格式来显示；不写q=表示使用全字段进行匹配。

7. 测试使用分析器/\_analyze