

1. 索引

索引是ElasticSearch对逻辑数据的逻辑存储，可以分很小的部分。ElasticSearch可以把索引存放在一台机器上，也可以放在多台机器上，每个索引有一或多个分片(shard),每个分片可以有多个副本(replica)

2. 文档

存储在ElasticSearch中主要实体叫做文档(document).用关系型数据库来对比,文档相当于数据库表中的一行记录.注意在ElasticSearch中相同的字段必须有相同的类型.即所有包含title字段的文档,类型都一致,比如说类型为String.

文档由多个字段组成，每个字段可能多次出现在一个文档里，这样的字段叫多值字段(multivalued)。每个字段有类型，如文本、数值、日期等。字段类型也可以是复杂类型，一个字段包含其他子文档或者数组。字段类型在Elasticsearch中很重要，因为它给出了各种操作（如分析或排序）如何被执行的信息。幸好，这可以自动确定，然而，我们仍然建议使用映射。与关系型数据库不同，文档不需要有固定的结构，每个文档可以有不同的字段，此外，在程序开发期间，不必确定有哪些字段。当然，可以用模式强行规定文档结构。从客户端的角度看，文档是一个JSON对象（关于JSON格式的更多内容，参见<http://en.wikipedia.org/wiki/JSON>）。每个文档存储在一个索引中并有一个Elasticsearch自动生成的唯一标识符和文档类型。文档需要有对应文档类型的唯一标识符，这意味着在一个索引中，两个不同类型的文档可以有相同的唯一标识符。

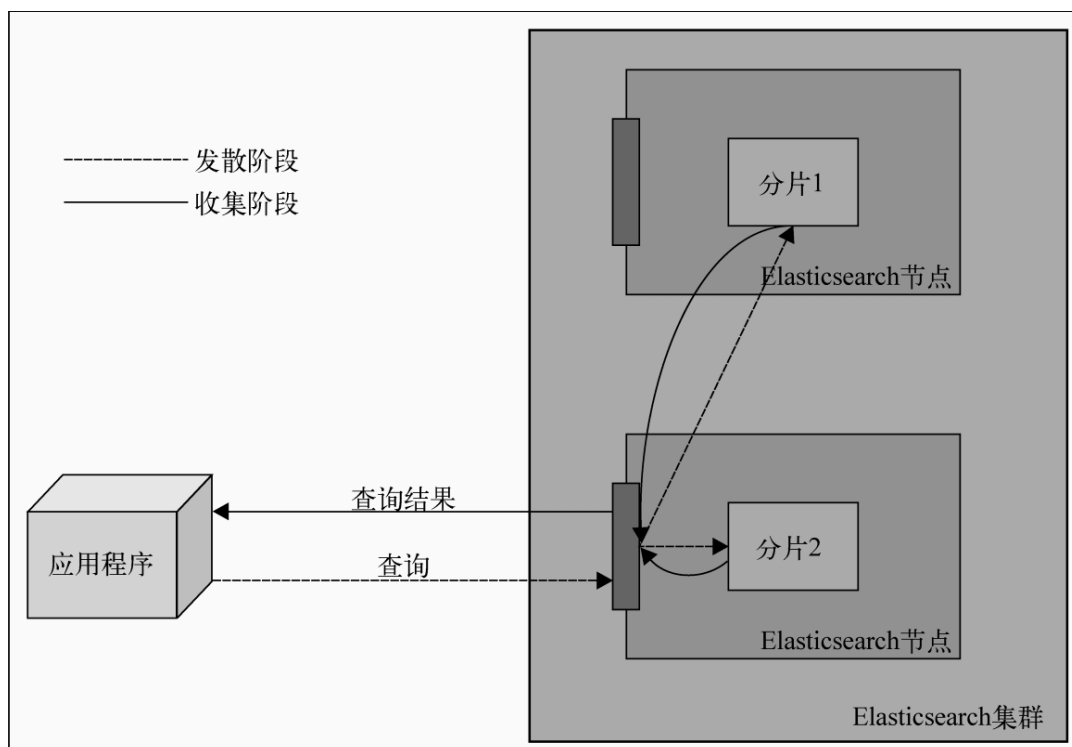
3. 文档类型

在Elasticsearch中，一个索引对象可以存储很多不同用途的对象。例如，一个博客应用程序可以保存文章和评论。文档类型让我们轻易地区分单个索引中的不同对象。每个文档可以有不同的结构，但在实际部署中，将文件按类型区分对数据操作有很大帮助。当然，需要记住一个限制，不同的文档类型不能为相同的属性设置不同的类型。例如，在同一索引中的所有文档类型中，一个叫title的字段必须具有相同的类型。

4. 映射

在有关全文搜索基础知识部分，我们提到了分析的过程：为建索引和搜索准备输入文本。文档中的每个字段都必须根据不同类型做相应的分析。举例来说，对数值字段和从网页抓取的文本字段有不同的分析，比如前者的数字不应该按字母顺序排序，后者的第一步是忽略HTML标签，因为它们是无用的信息噪音。Elasticsearch在映射中存储有关字段的信息。每一个文档类型都有自己的映射，即使我们没有明确定义。

ElasticSearch执行搜索请求图



ElasticSearch交互的接口是基于Http协议和rest的.可以使用简单的web浏览器完成基本的查询和请求.

分片

正如我们之前提到的那样, 集群允许系统存储的数据总量超过单机容量。为了满足这个需求, ElasticSearch 将数据散布到多个物理 Lucene 索引上。这些 Lucene 索引称为分片 (shard), 而散布这些分片的过程叫作分片处理 (sharding)。ElasticSearch 会自动完成分片处理, 并且让这些分片呈现出一个大索引的样子。请记住, 除了 ElasticSearch 本身自动进行分片处理外, 用户为具体的应用进行参数调优也是至关重要的, 因为分片的数量在索引创建时就已经配置好, 而且之后无法改变, 至少对目前的版本是这样的。

副本

分片处理允许用户向 ElasticSearch 集群推送超过单机容量的数据。副本 (replica) 则解决了访问压力过大时单机无法处理所有请求的问题。思路很简单, 即为每个分片创建冗余的副本, 处理查询时可以把这些副本用作最初的主分片 (primary shard)。请记住, 我们并未付出额外的代价。即使某个分片所在的节点宕机, ElasticSearch 也可以使用其副本, 从而不会造成数据丢失, 而且支持在任意时间点添加或移除副本, 所以一旦有需要可随时调整副本的数量。