



链滴

## 【Elasticsearch7.0】之基础介绍

作者: [chaigx](#)

原文链接: <https://ld246.com/article/1557983415904>

来源网站: [链滴](#)

许可协议: [署名-相同方式共享 4.0 国际 \(CC BY-SA 4.0\)](#)

Elasticsearch是非常流行的全文检索框架，接下来让我们来看看Elasticsearch到底是什么东西？

## es的使用场景

Elasticsearch是一个高度可伸缩的全文搜索和分析引擎。es支持存储、搜索、快速分析海量数据。对一些复杂的搜索系统提供底层支持。

可以在以下场景使用：

- 1、大型购物网站，用户可以搜索你的商品。可以把商品相关信息放到es中，可以进行搜索和自动填充。
- 2、分析日志或者交易数据，你可以通过Elasticsearch+Logstash+Kibana来完成这个事情，Elasticsearch提供搜索或者聚合数据功能。
- 3、价格警报平台，如果用户关注的电子产品降价了，需要通知用户，那么可以通过es的反向搜索来配到信息，并推送给用户。
- 4、BI系统，商业智能，es执行数据分析和挖掘。
- 5、站内搜索，如电商、招聘、门户等。
- 6、数据分析
- 7、github、搜索上千亿行代码。
- 8、stack overflow，国外最大的程序异常讨论论坛。

## 基本概念

在Elasticsearch中有一些基本概念，了解了这些概念，对后面的学习会有很大的帮助。

### Near Realtime (NRT)

Near Realtime称为近实时，Elasticsearch是一个近实时的搜索平台。也就是说，es并不是一个实时，一个索引文档可以被搜索到存在一定的延迟，通常情况下大概有1秒的延迟。

### cluster (集群)

集群由1个或者多个节点组成，他们一起保存数据，提供联合索引，跨节点的搜索。每个集群都需要定唯一的名称，默认的名称是：elasticsearch。这个名称很重要，只有名称相同的节点才会被加入到群中。确保不要在不同的环境中使用相同的名称，有可能会加入到一个错误的集群中去。你可以在名后面添加环境，如log-dev，log-test，log-prod。如果集群里面只有一个节点，它也可以运行的很。可以根据不同的名称组建不同的集群。

### node (节点)

节点是一个单独的服务，是一个集群的一部分，用来存储你的数据，参与集群的索引和搜索功能。和群一样，节点也是根据名称来区分的，默认情况下，在启动的时候会给节点分配一个全局的唯一标识(UUID)。当然，你可以手动设置名称。在集群中，节点的名称非常重要，它们通过名称来找到彼此。

可以通过集群名称来配置节点加入一个特定的集群。默认情况下，每个节点都会被加入到elasticsearch集群中。

在单个集群中，可以配置很多节点。如果你之前没有启动任何es节点，如果此时你启动一个节点，那会默认形成一个新的单节点集群，集群名称是：elasticsearch。

## index (索引)

索引是具有相似特性的文档集合。你可以给用户数据创建索引，也可以给订单数据创建索引。每一个索引都必须有唯一的名称，名称必须是小写的。这个名称会在搜索、删除、更新索引的时候使用到。在es中，你可以定义很多索引。索引相当于mysql中的库。

## type (类型)

类型是索引里的逻辑分类或者分区，通过类型可以在索引里存储不同类型的文档。比如一个类型存用信息，另一个类型存博客数据。在最近的版本中不在有type这个概念了，在索引里不能在创建多个类了。6.0.0之后就弃用了。

## document (文档)

文档是建立索引的基本信息单元。例如，你可以对用户，订单，商品创建各自的文档。文档用JSON表示。在索引里可以存储任意数量的文档。

## shards & replicas (分片和副本)

索引可以存大量的数据，在单节点下，可能会超过硬盘的大小限制。例如，你存了10亿文档，占用了1B的容量，这样单个节点就不适用了，因为在搜索的时候可能会变的很慢。

es为了解决这个问题，对索引提供了分片功能。在创建索引的时候，可以定义分片的数量。每个分片是功能全齐并且独立的索引，可以托管在集群任何节点上。

分片为什么重要：

- 1、它支持水平扩容；
- 2、它支持不同分片并行处理，提高性能和吞吐量；

分片如何分布以及聚合搜索结果这些机制都是由es管理的，对使用者来说是透明的。在错综复杂的网环境下，分片或者节点随时可能会出现离线或者消失，因此故障转移机制是非常重要的措施，该机制保证高可用。es允许分片去复制一个或者多个分片，简称：副本。

副本为什么重要：

- 1、它提供高可用，即使有分片或者节点出现故障。前提条件是，分片和该分片副本不要在相同的节上。
- 2、它可以提供吞吐量，因为副本也会参与搜索。

总之，每个索引都可以被分成多个分片。一个索引可以没有副本，也可以创建多个副本。一旦创建索引，那就会有一个主分片（副本的原始分片），一个副本分片（主分片的副本）。

可以在创建索引的时候定义分片和副本的数量。你也可以随时修改已创建索引副本的数量。你也可以用 `_shrink` 和 `_split` 这两个接口来修改索引分片的数量，但是操作起来比较麻烦，所以建议在一开始规划好索引的分片数量。

默认的，es会为每个索引分配一个主分片和一个副本，在集群里至少有两个节点。

每个分片都是Lucene索引，在单个Lucene索引里面文档的数量是有限制的。所以每个分片最多可以存 ,147,483,519个文档。你可以通过 `_cat/shards` 接口来监控分片大小。

也可以关注我的公众号：程序之声

关注公众号，领取更多 [资源](#)。