

Elasticsearch入门篇（一、基本概念）

注意：所有es入门篇博客都是根据es官网6.2版本的文档翻译而来

近实时（NRT）

Elasticsearch是一个接近实时的搜索平台。这意味着从索引文档的时间到可搜索的时间之间存在轻微的延迟（通常为一秒）。

集群(cluster)

群集是一个或多个节点（服务器）的集合，它们一起保存整个数据，并提供跨所有节点的联合索引和搜索功能。集群由默认名称为“elasticsearch”的唯一名称标识。此名称很重要，因为如果节点设置为通过名称加入群集，则该节点只能成为一个集群的一部分。

确保不要在不同的环境中重复使用相同的集群名称，否则可能会导致节点加入错误的集群。例如，您可以使用logging-dev，logging-stage以及logging-prod作为 开发环境、试运行环境和生产环境的集群名称。

请注意，有一个只有一个节点的集群是完全正确的。此外，您可能拥有多个独立的集群，每个群集都有自己的唯一集群名称。

节点(node)

节点是属于集群一部分的单个服务器，存储数据并参与集群的索引和搜索功能。就像一个集群一样，一个节点由一个名称来标识，默认情况下该名称是在启动时分配给节点的随机通用唯一标识符（UUID）。如果您不需要默认值，您可以定义任何您想要的节点名称。此名称对于管理目的非常重要，您需要确定网络中的哪些服务器与Elasticsearch集群中的哪些节点相对应。

可以将节点配置为按集群名称加入特定集群。默认情况下，每个节点都设置为加入一个名为“elasticsearch”的集群，这意味着如果启动网络中的多

个节点并假设它们可以互相发现，它们将自动形成并加入一个名为elasticsearch的群集。

在单个群集中，您可以拥有任意数量的节点。此外，如果网络上当前没有其他Elasticsearch节点正在运行，则默认情况下启动单个节点将形成名为elasticsearch的新单节点集群。

索引(index)

索引是一些具有相似特征的文档集合。例如，您可以拥有客户数据的索引，产品目录的另一个索引以及订单数据的另一个索引。索引由名称（必须全部为小写）标识，并且此名称用于在对其中的文档执行索引，搜索，更新和删除操作时使用。

在单个群集中，您可以根据需要定义多个索引。

类型 (type)

注意：在6.0.0版本中被弃用

一种类型曾经是您的索引的逻辑类别或逻辑分区，允许您在同一索引中存储不同类型的文档，例如一种用户类型的文档，另一种是博客文章类型的文档。不再可能在索引中创建多个类型，并且在更高版本中将删除整个类型的概念。有关更多信息，请参阅移除映射类型介绍篇。

文档 (document)

文档是可被索引的基本信息单元。例如，您可以为单个客户创建一个文档，为单个产品创建另一个文档，为单个订单创建另一个文档。本文档以JSON（JavaScriptObject Notation）表示，它是一种很普遍的互联网数据交换格式。

在索引或类型中，您可以根据需要存储任意数量的文档。请注意，尽管文档实际上驻留在索引中，但实际上文档必须被索引或被分配给索引内的类型。

分片和副本 (Shards & Replicas)

索引可能潜在地存储大量数据，这些数据可能会超出单个节点的硬件限制。例如，占用1TB磁盘空间的十亿份文档的单个索引可能不适合单个节点的磁盘，或者可能太慢而无法单独向单个节点提供搜索请求。

为了解决这个问题，Elasticsearch提供了将索引细分为多个碎片的能力。当您创建索引时，您可以简单地定义所需的碎片数量。每个分片本身都是一个功能齐全且独立的“索引”，可以在集群中的任何节点上进行托管。

分片很重要，主要有两个原因：

- 它允许您水平分割或者扩展内容量
- 它允许您跨越分片（可能在多个节点上）分发和并行化操作，从而提高性能或者吞吐量

分片如何分布的机制以及其文档如何聚合回搜索请求完全由Elasticsearch管理，并且对用户都是透明的。

在任何时候都可能出现故障的网络/云环境中，非常有用并且强烈建议有一个故障切换机制，以防碎片/节点以某种方式宕机或因任何原因而消失。为此，Elasticsearch允许您将索引碎片的一个或多个备份制作为简称为副本碎片或副本。

复制很重要，主要有两个原因：

- 它在碎片/节点失败的情况下提供高可用性。由于这个原因，需要注意的是，副本分片永远不会分配到与原始或主分片相同的节点上。
- 它允许您扩展搜索量或者吞吐量，因为搜索可以在所有副本上并行执行。

总而言之，每个索引可以分成多个分片。索引也可以被复制为零（意味着没有副本）或更多次。一旦复制，每个索引将具有主分片（从中复制的原始分片）和副本分片（主分片的副本）。在创建索引时，可以为每个索引定义分片和副本的数量。在创建索引之后，您可以随时更改动态副本的数量，但您无法在事后更改碎片的数量。

默认情况下，Elasticsearch中的每个索引都分配了5个主分片和1个副本，

这意味着如果群集中至少有两个节点，则索引将包含5个主分片和另外5个副本分片（1个完整副本），总共每个索引10个碎片。

注意：

每个Elasticsearch分片都是一个Lucene索引。您可以在单个Lucene索引中创建的文档数量是有限制的，截至LUCENE-5843的上限是2,147,483,519（=Integer.MAX_VALUE - 128）个文件。您可以使用_cat/shardsAPI监控分片大小。

接下来，让我们开始有趣的部分...