

elasticsearch

加入尚学堂，一起进步！



- Elasticsearch中的核心概念
- cluster
 - 代表一个集群，集群中有多个节点，其中有一个为主节点，这个主节点是可以通过选举产生的，主从节点是对于集群内部来说的。es的一个概念就是去中心化，字面上理解就是无中心节点，这是对于集群外部来说的，因为从外部来看es集群，在逻辑上是个整体，你与任何一个节点的通信和与整个es集群通信是等价的。
 - 主节点的职责是负责管理集群状态，包括管理分片的状态和副本的状态，以及节点的发现和删除。
- 只需要在同一个网段之内启动多个es节点，就可以自动组成一个集群。
- 默认情况下es会自动发现同一网段内的节点，自动组成集群。
- 集群状态查看
 - http://192.168.57.4:9200/_cluster/health?pretty



- Elasticsearch中的核心概念
- shards
 - 代表索引分片，es可以把一个完整的索引分成多个分片，这样的好处是可以把一个大的索引拆分成多个，分布到不同的节点上。构成分布式搜索。分片的数量只能在索引创建前指定，并且索引创建后不能更改。
- 可以在创建索引库的时候指定
 - `curl -XPUT 'localhost:9200/test1/' -d '{"settings":{"number_of_shards":3}}'`
- 默认是一个索引库有5个分片
 - `index.number_of_shards: 5`



- Elasticsearch中的核心概念
- replicas
 - 代表索引副本，es可以给索引设置副本，副本的作用一是提高系统的容错性，当某个节点某个分片损坏或丢失时可以从副本中恢复。二是提高es的查询效率，es会自动对搜索请求进行负载均衡。
- 可以在创建索引库的时候指定
 - `curl -XPUT 'localhost:9200/test2/' -d '{"settings":{"number_of_replicas":2}}'`
- 默认是一个分片有1个副本
 - `index.number_of_replicas: 1`



- Elasticsearch中的核心概念
- recovery
 - 代表数据恢复或叫数据重新分布，es在有节点加入或退出时会根据机器的负载对索引分片进行重新分配，挂掉的节点重新启动时也会进行数据恢复。



- Elasticsearch中的核心概念
- gateway
 - 代表es索引的持久化存储方式，es默认是先把索引存放到内存中，当内存满了时再持久化到硬盘。当这个es集群关闭再重新启动时就会从gateway中读取索引数据。es支持多种类型的gateway，有本地文件系统（默认），分布式文件系统，Hadoop的HDFS和amazon的s3云存储服务。



- Elasticsearch中的核心概念
- gateway
 - 如果需要将数据落地到hadoop的hdfs需要先安装插件elasticsearch/elasticsearch-hadoop，然后再elasticsearch.yml配置
 - gateway:
 - type: hdfs
 - gateway:
 - hdfs:
 - uri: hdfs://localhost:9000



- Elasticsearch中的核心概念
- discovery.zen
 - 代表es的自动发现节点机制，es是一个基于p2p的系统，它先通过广播寻找存在的节点，再通过多播协议来进行节点之间的通信，同时也支持点对点的交互。
- 如果是不同网段的节点如何组成es集群
 - 禁用自动发现机制
 - discovery.zen.ping.multicast.enabled: false
 - 设置新节点被启动时能够发现的主节点列表
 - discovery.zen.ping.unicast.hosts: ["192.168.57.4", "192.168.57.5"]



- Elasticsearch中的核心概念
- Transport
 - 代表es内部节点或集群与客户端的交互方式，默认内部是使用tcp协议进行交互，同时它支持http协议（json格式）、thrift、servlet、memcached、zeroMQ等的传输协议（通过插件方式集成）。



- Elasticsearch中的settings和mappings
- settings修改索引库默认配置
 - 例如：分片数量，副本数量
 - 查看：`curl -XGET http://localhost:9200/bjsxt/_settings?pretty`
 - `curl -XPUT 'localhost:9200/bjsxt/' -d'{"settings":{"number_of_shards":3,"number_of_replicas":2}}'`
- Mapping,就是对索引库中索引的字段名称及其数据类型进行定义，类似于关系数据库中表建立时要定义字段名及其数据类型那样，(和solr中的schme类似)不过es的mapping比数据库灵活很多，它可以动态添加字段。一般不需要指定mapping都可以，因为es会自动根据数据格式定义它的类型，如果你需要对某些字段添加特殊属性（如：定义使用其它分词器、是否分词、是否存储等），就必须手动添加mapping
- 查询索引库的mapping信息
 - `curl -XGET http://localhost:9200/bjsxt/emp/_mapping?pretty`
- mappings修改字段相关属性
 - 例如：字段类型，使用哪种分词工具

