概念：

1.节点(node):

一个装有es的服务并且提供故障转移和扩展的服务器，

单独一个es服务可以存储数据及搜索功能

在集群中一个节点的名称必须是唯一的

如果在集群中,通过节点名称进行管理

2.集群(cluster):

集群就是有多个node组织在一起。共同工作分享整个数据具有负载均衡功能的集群

集群里的节点协调工作，共享数据。

集群里通过master节点进行集群管理工作

集群里通过选举选出一个matser节点

一个集群里只有一个matser节点

每一个节点都有成为matser节点的可能

3.索引(Index)

索引就是一个拥有几分相似特征的文档集合

相当于mysql里的database

4.文档(Document)

一个文档是一个可以被索引的基础信息单元

相当于mysql里的行

文档用json格式来表示

5.类型(type)

一个索引中，可以定义一种或者多种数据类型

一个类型是你的索引的一个逻辑上的分类/分区

6.字段(field)

es里的最小单元 相当于数据的某一列

类似于json里一个键

7.分片(shards)

es将索引分成若干份 每个部分是一个shards

每一个shards存在不同的节点上

一个节点上不能存在两个相同的shards

每一个文档通过文档id进行hash来确定放在哪一个节点上

每一个分片都相当于一个独立的Lucene实例

8.副本(replicas)

索引的一份或者多份拷贝

容灾作用，防止主分片丢失后，副本分片会作为新的主分片，保证集群的数据完整性

提供查询性能，query的时候，既可以查询主分片，也可以查询副本分片

架构：

1.Gateway

gateway是ES数据存储的格式

可以使用hdfs，本地，亚马逊的s3等多种存储方式

存储索引信息,集群信息,mapping, 索引碎片信息,以及transaction logs

2.Distributed Lucene Directory

Lucene框架 es就是基于Lucene框架开发的

Lucene框架服发现等

3.Index Module

创建索引的模块

4.Search Module

搜索模块

5.Mapping

相当于mysql里的schema

6.river

从外部获取异构数据 来创建索引

7.Discovery

节点启动后会互相ping 根据在es.yml配置文件里找到对应的端口

进行开始选举，从各个节点任务的master中选，进行id字典排序，选择第一个

如果各个节点上都没有认为的master， 那么就从所有节点中选择

如果就一个节点 那么master就是她自己

ES支持任意数目的集群,通过一个规则，只要所有的节点都遵循同样的规则，得到的信息都是对等的，选出来的主节点肯定是一致的. 但分布式系统的问题就出在信息不对等的情况，这时候很容易出现脑裂（Split-Brain）的问题，大多数解决方案就是设置一个quorum值，要求可用节点必须大于quorum（一般是超过半数节点），才能对外提供服务。而 Elasticsearch 中，这个quorum的配置就是 discovery.zen.minimum\_master\_nodes

8.Scriptsing

脚本执行功能 对查询出来的数据进行处理

支持多种语言

9.3rdplugins

支持安装第三方插件

10. transport

是支持的协议类型 默认使用http进行交互