|  |
| --- |
| Neusoft |
| ElasticSearch |
| Elastic Search 安装与配置 |

|  |
| --- |
| 成阳阳  2015/3/25 |

# （一）前言

## 下载elasticsearch

一）安装elasticsearch

1）下载elasticsearch-0.90.10，解压，运行\bin\elasticsearch.bat （windwos）

2）进入http://localhost:9200/ 如下图

安装成功！

## 概念

Index：这是ES存储数据的地方，类似于关系数据库的database。

Document type：嗯，类似关系数据库的表，主要功能是将完全不同schema（这个概念以后会讲到，不急）的数据分开，一个index里面可以有若干个Document type。

Document：好吧，这个类似关系数据库的一行，在同一个Document type下面，每一Document都有一个唯一的ID作为区分；

Filed：类似关系数据库的某一列，这是ES数据存储的最小单位。

Cluster和Node：ES可以以单点或者集群方式运行，以一个整体对外提供search服务的所有节点组成cluster，组成这个cluster的各个节点叫做node。

shard：通常叫分片，这是ES提供分布式搜索的基础，其含义为将一个完整的index分成若干部分存储在相同或不同的节点上，这些组成index的部分就叫做shard。

Replica：和replication通常指的都是一回事，即index的冗余备份，可以用于防止数据丢失，或者用来做负载分担。

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。

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。

replicas

代表索引副本，es可以设置多个索引的副本，副本的作用一是提高系统的容错性，当个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。

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。

river

代表es的一个数据源，也是其它存储方式（如：数据库）同步数据到es的一个方法。它是以插件方式存在的一个es服务，通过读取river中的数据并把它索引到es中，官方的river有couchDB的，RabbitMQ的，Twitter的，Wikipedia的，river这个功能将会在后面的文件中重点说到。

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。

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。

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）。

## elasticsearchservicewrapper安装

这个是对elasticsearch执行命令的包装服务，安装之后，方便elasticsearch的启动，停止等等操作。

（1）下载elasticsearchservicewrapper

git clone https://github.com/elasticsearch/elasticsearch-servicewrapper，然后将目录下的service目录拷贝至ES\_HOME/bin目录下。

（2）简单配置jvm的内存

修改ES\_HOME/bin/service/elasticsearch.conf，set.default.ES\_HEAP\_SIZE=1024，该值根据机器的配置可自定义。

（3）安装启动服务

执行命令：ES\_HOME/bin/service/elasticsearch install

（4）启动/停止/重启服务

执行命令：ES\_HOME/bin/service/elasticsearch start/stop/restart

或者

service elasticsearch start/stop/restart

## 插件——head

elasticsearch-head是一个elasticsearch的集群管理工具，它是完全由html5编写的独立网页程序，你可以通过插件把它集成到es。

插件安装方法1：

1.elasticsearch/bin/plugin install mobz/elasticsearch-head --verbose

2.运行es

3.打开http://localhost:9200/\_plugin/head/

插件安装方法2：

1.https://github.com/mobz/elasticsearch-head下载zip 解压

2.建立elasticsearch-1.0.0\plugins\head\\_site文件

3.将解压后的elasticsearch-head-master文件夹下的文件copy到\_site

4.运行es

5.打开<http://localhost:9200/_plugin/head/>

## 插件——bigdesk

bigdesk是elasticsearch的一个集群监控工具，可以通过它来查看es集群的各种状态，如：cpu、内存使用情况，索引数据、搜索情况，http连接数等。

安装命令：bin>./plugin install lukas-vlcek/bigdesk --verbose

进入http://localhost:9200/\_plugin/bigdesk/

## 插件——ik

安装步骤：

1、到github网站下载源代码，网站地址为： https://github.com/medcl/elasticsearch-analysis-ik

右侧下方有一个按钮“Download ZIP"，点击下载源代码elasticsearch-analysis-ik-master.zip。

2、解压文件elasticsearch-analysis-ik-master.zip，进入下载目录，执行命令：

unzip elasticsearch-analysis-ik-master.zip

3、将解压目录文件中config/ik文件夹复制到ES安装目录config文件夹下。

4、因为是源代码，此处需要使用maven打包，进入解压文件夹中，执行命令：

mvn clean package

5、将打包得到的jar文件elasticsearch-analysis-ik-1.2.8-sources.jar复制到ES安装目录的lib目录下。

6、在ES的配置文件config/elasticsearch.yml中增加ik的配置，在最后增加：

|  |
| --- |
| index:  analysis:  analyzer:  ik:  alias: [ik\_analyzer]  type: org.elasticsearch.index.analysis.IkAnalyzerProvider  ik\_max\_word:  type: ik  use\_smart: false  ik\_smart:  type: ik  use\_smart: true |

或

|  |
| --- |
| index.analysis.analyzer.ik.type : “ik” |

7、重新启动elasticsearch服务，这样就完成配置了，收入命令：

|  |
| --- |
| curl -XPOST "http://localhost:9200/userinfo/\_analyze?analyzer=ik&pretty=true&text=我是中国人" |

测试结果如下：

|  |
| --- |
| {  tokens: [  {  token: text  start\_offset: 2  end\_offset: 6  type: ENGLISH  position: 1  }  {  token: 我  start\_offset: 9  end\_offset: 10  type: CN\_CHAR  position: 2  }  {  token: 中国人  start\_offset: 11  end\_offset: 14  type: CN\_WORD  position: 3  }  {  token: 中国  start\_offset: 11  end\_offset: 13  type: CN\_WORD  position: 4  }  {  token: 国人  start\_offset: 12  end\_offset: 14  type: CN\_WORD  position: 5  }  ]  } |

说明：

1、ES安装插件本来使用使用命令plugin来完成，但是我本机安装ik时一直不成功，所以就使用源代码打包安装了。

2、自定义词库的方式，请参考 https://github.com/medcl/elasticsearch-analysis-ik

## 插件——其他

以下内容为转载：

Elasticsearch扩展性非常好，有很多官方和第三方开发的插件，下面以分词、同步、数据传输、脚本支持、站点、其它这几个类别进行划分。

分词插件

Combo Analysis Plugin (作者 Olivier Favre, Yakaz)

简介：组合分词器，可以把多个分词器的结果组合在一起。

Smart Chinese Analysis Plugin (作者 elasticsearch 团队)

简介：lucene默认的中文分词器

ICU Analysis plugin (作者 elasticsearch 团队)

简介：lucene自带的ICU分词，ICU是一套稳定、成熟、功能强大、轻便易用和跨平台支持Unicode 的开发包。

Stempel (Polish) Analysis plugin (作者 elasticsearch 团队)

简介：法文分词器

IK Analysis Plugin (作者 Medcl)

简介：大名鼎鼎的ik分词，都懂的！

Mmseg Analysis Plugin (作者 Medcl)

简介：mmseg中文分词

Hunspell Analysis Plugin (作者 Jörg Prante)

简介：lucene自带的Hunspell模块

Japanese (Kuromoji) Analysis plugin (作者 elasticsearch 团队).

简介：日文分词器

Japanese Analysis plugin (作者 suguru).

简介：日文分词器

Russian and English Morphological Analysis Plugin (作者 Igor Motov)

简介：俄文英文分词器

Pinyin Analysis Plugin (作者 Medcl)

简介：拼音分词器

String2Integer Analysis Plugin (作者 Medcl)

简介：字符串转整型工具。主要用在facet这个功能上，如果facet的field的值是字符串的话，计算起来比较耗资源。可以把字符串映射成整型，对整型进行facet操作要比对字符串的快很多。

同步插件

CouchDB River Plugin (作者 elasticsearch 团队)

简介：CouchDB和elasticsearch的同步插件

Wikipedia River Plugin (作者 elasticsearch 团队)

简介：wikipedia文件读取插件。wikipedia是维基百科的一个离线库，不定期发布最新数据，是以xml形式发布的。这个river读取这个文件来建索引。

Twitter River Plugin (作者 elasticsearch 团队)

简介：twitter的同步插件，可以同步你twitter上的微博。

RabbitMQ River Plugin (作者 elasticsearch 团队)

简介：rabbitmq同步插件，读取rabbitmq上的队列信息并索引。

RSS River Plugin (作者 David Pilato)

简介：定期索引指定一个或多个RSS源的数据。

MongoDB River Plugin (作者 Richard Louapre)

简介：mongodb同步插件，mongodb必须搭成副本集的模式，因为这个插件的原理是通过定期读取mongodb中的oplog来同步数据。

Open Archives Initiative (OAI) River Plugin (作者 Jörg Prante)

简介：可以索引oai数据提供者提供的数据。

St9 River Plugin (作者 Sunny Gleason)

简介：可以索引索引st9数据（st9是神马？囧！！！）

Sofa River Plugin (作者 adamlofts)

简介：这个插件可以把多个CouchDB的数据库同步到同一个es索引中。

JDBC River Plugin (作者 Jörg Prante)

简介：关系型数据库的同步插件

FileSystem River Plugin (作者 David Pilato)

简介：本地文件系统文件同步插件，使用方法是指定一个本地目录路径，es会定期扫描索引该目录下的文件。

LDAP River Plugin (作者 Tanguy Leroux)

简介：索引LDAP目录下的文件数据。

Dropbox River Plugin (作者 David Pilato)

简介：索引dropbox网盘上的文件。通过oauth协议来调用dropbox上的api建索引。

ActiveMQ River Plugin (作者 Dominik Dorn)

简介：activemq队列的同步插件，和之前rabbitmq的类似

Solr River Plugin (作者 Luca Cavanna)

简介：solr同步插件，可以把solr里面的索引同步到es

CSV River Plugin (作者 Martin Bednar)

简介：通过指定目录地址来索引csv文件。

数据传输插件

Servlet transport (作者 elasticsearch 团队)

简介：Servlet rest插件，通过servlet来封装rest接口。

Memcached transport plugin (作者 elasticsearch 团队)

简介：本插件可以通过memcached协议进行rest接口的调用。注意：这里不是使用memcache作为es的缓存。

Thrift Transport (作者 elasticsearch 团队)

简介：使用thrift进行数据传输。

ZeroMQ transport layer plugin (作者 Tanguy Leroux)

简介：使用zeromq进rest接口的调用。

Jetty HTTP transport plugin (作者 Sonian Inc.)

简介：使用jetty来提供http rest接口。默认是使用netty。这个插件的好处是可以对http接口进行一些权限的设置。

脚本插件

Python language Plugin (作者 elasticsearch 团队)

简介：python脚本支持

JavaScript language Plugin (作者 elasticsearch 团队)

简介：javascript脚本支持

Groovy lang Plugin (作者 elasticsearch 团队)

简介：groovy脚本支持

Clojure Language Plugin (作者 Kevin Downey)

简介：clojure脚本支持

站点插件（以网页形式展现）

BigDesk Plugin (作者 Lukáš Vlček)

简介：监控es状态的插件，推荐！

Elasticsearch Head Plugin (作者 Ben Birch)

简介：很方便对es进行各种操作的客户端。

Paramedic Plugin (作者 Karel Minařík)

简介：es监控插件

SegmentSpy Plugin (作者 Zachary Tong)

简介：查看es索引segment状态的插件

Inquisitor Plugin (作者 Zachary Tong)

简介：这个插件主要用来调试你的查询。

其它插件

Mapper Attachments Type plugin (作者 elasticsearch 团队)

简介：附件类型插件，通过tika库把各种类型的文件格式解析成字符串。

Hadoop Plugin (作者 elasticsearch team)

简介：hadoop和elasticsearch的集成插件，可以通过hadoop的mapreduce算法来并行建立索引，同时支持cascading，hive和pig等框架。

AWS Cloud Plugin (作者 elasticsearch 团队)

简介：elasticsearch与amazon web services的集成。

ElasticSearch Mock Solr Plugin (作者 Matt Weber)

简介：elasticsearch的solr api接口。用了这个插件可以使用solr的api来调用es，直接用solrj就可以调用es。比较适用于从solr转es时暂时过度。

Suggester Plugin (作者 Alexander Reelsen)

简介：es 搜索提示功能插件，不过es0.9版本后自带了这个功能，

ElasticSearch PartialUpdate Plugin (作者 Medcl)

简介：elasticsearch的部分更新插件。

ZooKeeper Discovery Plugin (作者 Sonian Inc.)

简介：通过zookeeper管理集群的插件。通过这个插件，es的分布式架构和solrcloud相似。

ElasticSearch Changes Plugin (作者 Thomas Peuss)

简介：elasticsearch索引操作记录插件。通过这个插件可以查看用户对索引的增删改操作。

ElasticSearch View Plugin (作者 Tanguy Leroux)

简介：这个插件可以把es的文档以html，xml或text的方式显示出来，它也可以通过查询生成web页面。

ElasticSearch New Relic Plugin (作者 Vinicius Carvalho)

简介：elasticsearch和newrelic的集成插件。newrelica是一个性能监控工具。这个插件会把节点的状态数据传到newrelic的账号上。

# 集群环境配置

## 环境说明

服务器1: 192.168.24.100

服务器2:192.168.24.101

## 配置

配置es。这里只做最简单的配置，修改ES\_HOME/config/elasticsearch.yml文件

1、 服务器1

将node.name的值设置为“test-node1”，表示当前这个es服务节点名字为test-node1。

2、服务器2

将node.name的值设置为“test-node1”，表示当前这个es服务节点名字为test-node1。

两台es服务同时起来，因为配置文件中均默认cluster.name=elasticsearch，所以这两台机器自动构建成一个集群，集群名字为elasticsearch。

# 创建索引

# 分词

# 集群同步