修订日志：

|  |  |  |
| --- | --- | --- |
| 修改日期 | 修改人 | 备注 |
| 11.17 | 王文祥 | thrift搭建 |
| 11.18 | 李玫 | jdk、es、elasticHD、kibana、ik安装 |
|  |  |  |
|  |  |  |
|  |  |  |

**thrift**

因为本系统采用了hbase来存储评论信息，而hbase只提供了原生的Java接口，所以要与python进行交互的话就需要使用其他方式。而thrift服务器是hbase中的一种服务，其主要作用是用于hbase对于多语言API的支持，因此我们在安装了hbase数据库的服务器上安装并启动了thrift服务。

thrift可以直接通过指令“yum install thrift”在服务器上安装，然后在命令行中使用“hbase thrift start &”便可以启动thrift服务。

# 搜索部分

## elasticsearch7.4

es7不需要依赖java的环境了，其中已经整合了java环境，但是应用过程中还是碰到了一些插件需要依赖java环境的情况。

用yum安装jdk：安装了默认的jdk1.8.0版本

yum install java

同样通过yum获取了elasticsearch的7.4版本，从官方获取了elasticsearch.repo信息：

[elasticsearch-7.x]name=Elasticsearch repository for 7.x packagesbaseurl=https://artifacts.elastic.co/packages/7.x/yumgpgcheck=1gpgkey=https://artifacts.elastic.co/GPG-KEY-elasticsearchenabled=1autorefresh=1type=rpm-md

然后创建相关文件并通过yum下载

sudo yum install elasticsearch

配置和启动ES服务器进程

# /sbin/chkconfig --add elasticsearch

# systemctl start elasticsearch

最后通过curl -X GET localhost:9200 进行验证，出现内容就说明装好了

## 集群搭建

本次项目中，总共将三台服务器用于es的部署，内网ip分别是：172.26.240.210（master）、172.26.240.223、172.26.240.220

将三台服务器作为一个集群，一个是master节点另外两个是slave节点。

需要通过配置elasticsearch.yml文件让这三个服务器相互发现。用的是阿里云服务器，一开始因为一些端口没有设置对外开放的问题走了很多弯路，最后采用内网ip成功进行相互连接。

master节点配置如下

cluster.name: movies #cluster名字三个服务器要一样

node.name: node-1 #节点名称

node.master: true #作为主节点

node.data: true #同时存放数据

path.data: /var/lib/elasticsearch #数据路径

path.logs: /var/log/elasticsearch #日志路径

network.host: 0.0.0.0 #允许所有ip访问

http.port: 9200

transport.tcp.port: 9300 #不同节点的通信端口

#向两个slave服务器 单播

discovery.zen.ping.unicast.hosts: ["172.26.240.223:9300","172.26.240.220:9300"]

对于slave服务器2：

cluster.name: movies

node.name: node-3

node.master: false

node.data: true

path.data: /var/lib/elasticsearch

path.logs: /var/log/elasticsearch

network.host: 0.0.0.0

http.port: 9200

transport.tcp.port: 9300

discovery.zen.ping.unicast.hosts: ["172.26.240.210:9300"]

另外一个与之类似

* 为了防止节点脑裂问题，指明主节点（在另外两台服务器中指明node.master=false）
* host ip设置为0.0.0.0允许外网访问。
* 通过向两个slave服务器发送单播请求来互相识别，其中cluster-name必须是一致的

最后通过curl -X GET localhost:9200验证集群搭建成功与否。

## 分词与可视化

之前采用elasticHD插件进行可视化，

首先下载zip压缩包：

https://github.com/360EntSecGroup-Skylar/ElasticHD/releases/

修改权限：chmod -R 777 ElasticHD

运行: ./ElasticHD -p 127.0.0.1:9800

后来发现es官方有出综合可视化和数据分析的工具Kibana。安装部署在master服务器上，并开放9301端口用于节点和索引的状态查看和管理。将kibana下载并解压后，进入kibana.yml配置文件配置相关信息，允许任何人访问，且端口为9301：

server.host: "0.0.0.0"

server.port: 9301

由于es自带的四种analyzer中不包含对中文的分词处理，对于中文的处理仅仅是简单的按字分词，最终搜索效果不佳，因此选用了流行的IK分词器针对中文进行分词。IK插件需要在三台服务器上分别部署。下载的版本与es的版本需要完全一致。可以使用es的plugin install方法安装：

/usr/share/elasticsearch/bin/elasticsearch-plugin install <https://github.com/medcl/elasticsearch-analysis-ik/releases/download/v6..0/elasticsearch-analysis-ik-7.4.2.zip>

速度实在太慢，放弃了。到ik的github网页，下载对应的版本release并解压。在es的plugin文件夹下新建ik的文件夹并将解压内容拷贝进来。

# 