elasticsearch学习笔记

Elasticsearch是一个基于Apache Lucene(TM)的开源搜索引擎。无论在开源还是专有领域，Lucene可以被认为是迄今为止最先进、性能最好的、功能最全的搜索引擎库。Elasticsearch 不仅仅只是全文搜索，还可以结构化搜索、数据分析、复杂的语言处理、地理位置和对象间关联关系等。官网：<https://www.elastic.co/cn/>

# 1 安装和运行elasticsearch

## 1.1 安装依赖java

|  |  |
| --- | --- |
| 下载依赖java | wget http://sdlc-esd.oracle.com/ESD6/JSCDL/jdk/8u161-b12/2f38c3b165be4555a1fa6e98c45e0808/jre-8u161-linux-x64.tar.gz?GroupName=JSC&FilePath=/ESD6/JSCDL/jdk/8u161-b12/2f38c3b165be4555a1fa6e98c45e0808/jre-8u161-linux-x64.tar.gz&BHost=javadl.sun.com&File=jre-8u161-linux-x64.tar.gz&AuthParam=1523022954\_aa69621a3ecc34577f26e902c1d4e42d&ext=.gz |
| 安装到/usr/local目录下 | tar zxvf jre-8u161-linux-x64.tar.gz  mv jre1.8.0\_161/ jre1.8.0  mv jre1.8.0/ /usr/local/ |
| 把bin目录添加到环境变量 | echo 'export PATH="$PATH:/usr/local/jre1.8.0/bin"' >> /etc/bashrc  使当前终端生效  source /etc/bashrc |

## 1.2 安装Elasticsearch

|  |  |
| --- | --- |
| 下载elasticsearch | wget https://artifacts.elastic.co/downloads/elasticsearch/elasticsearch-6.2.3.tar.gz |
| 安装到/usr/local目录下 | tar zxvf elasticsearch-6.2.3.tar.gz  mv elasticsearch-6.2.3 elasticsearch  mv elasticsearch /usr/local |
| 把bin目录添加到环境变量 | echo 'export PATH="$PATH:/usr/local/elasticsearch/bin"' >> /etc/bashrc  使当前终端生效  source /etc/bashrc |

## 1.3 安装插件X-Pack(可选)

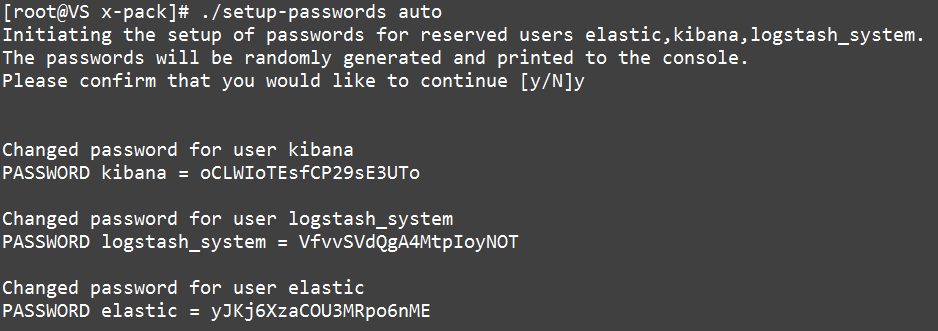
X-Pack monitoring 功能提供了一种有效方法，让您能够密切关注 Elasticsearch、Kibana、Beats 和 Logstash 的性能。它的仪表板集合能够帮助您在各个层级上评估仪表板的状态，同时为您提供所有必要的信息，让您始终最大限度地发挥 Elastic Stack 的作用。

**elasticsearch-plugin install x-pack**

启动elasticsearch后生自动帐号密码

**bin/x-pack/setup-passwords auto**

自动生成的帐号密码，访问Elasticsearch的api必须先登录



|  |
| --- |
| Changed password for user kibana  PASSWORD kibana = oCLWIoTEsfCP29sE3UTo  Changed password for user logstash\_system  PASSWORD logstash\_system = VfvvSVdQgA4MtpIoyNOT  Changed password for user elastic  PASSWORD elastic = yJKj6XzaCOU3MRpo6nME |

## 1.4 启动elasticsearch

elasticsearch默认是禁止在root用户启动，使用非root启动

在终端运行：**elasticsearch**

以守护进程运行：**elasticsearch -d**

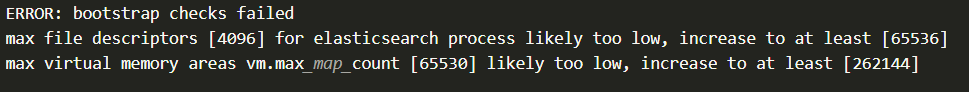
启动前配置，默认只有本机访问，修改绑定的IP地址，使其他电脑可以访问。

**vim /usr/local/elasticsearch/config/elasticsearch.yml**

添加network.host: 0.0.0.0



启动时如果出现下面错误

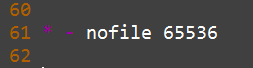


修改linux系统配置

(1) 修改进程的最大打开文件数

**vim /etc/security/limits.conf**

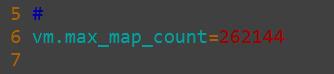
修改nofile 后面数字为65536



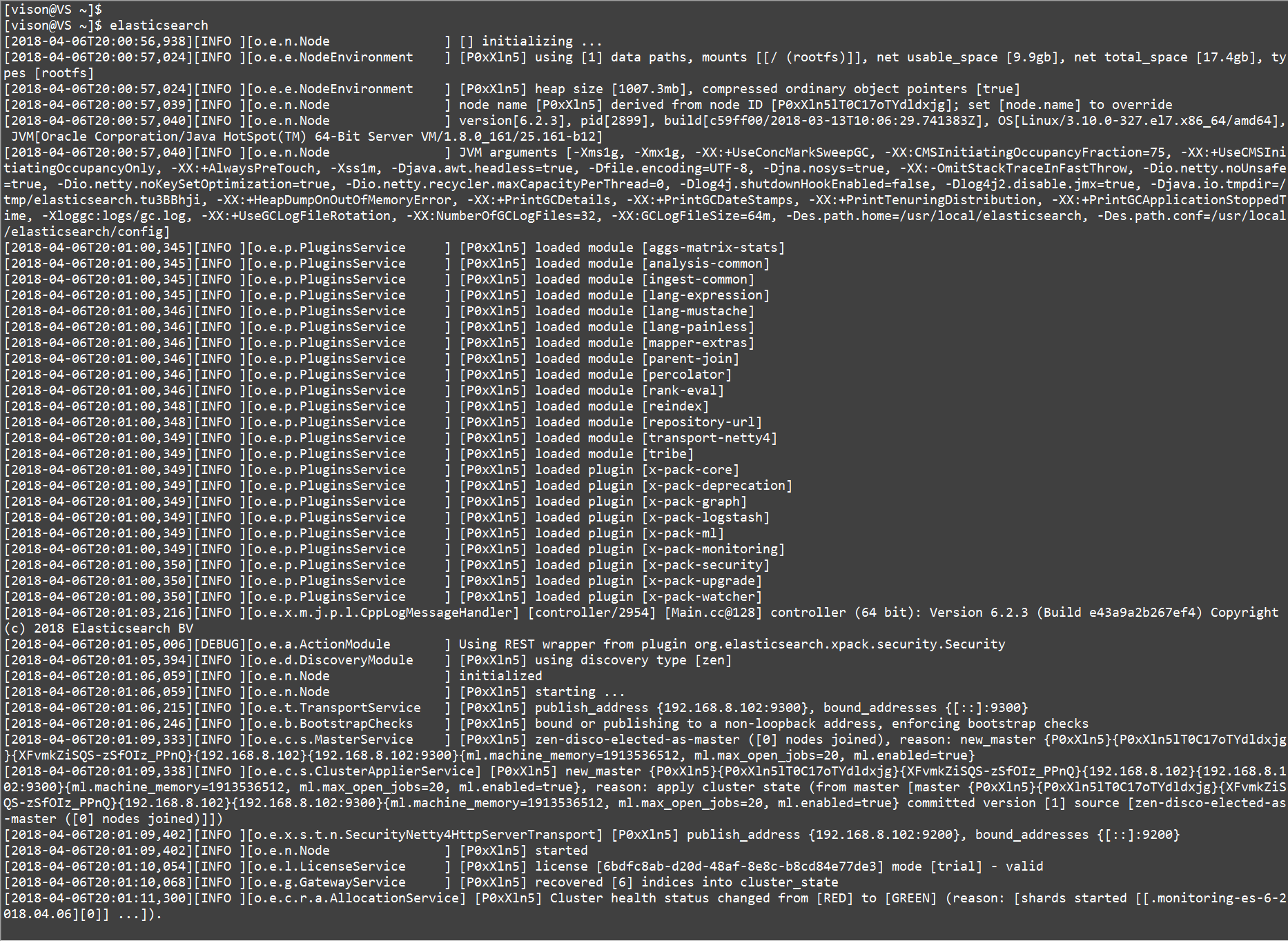
(2) 修改最大使用虚拟内存

**vim /etc/sysctl.conf**

在文件最后添加vm.max\_map\_count=262144



修改系统配置后重启linux，然后启动elasticsearch。



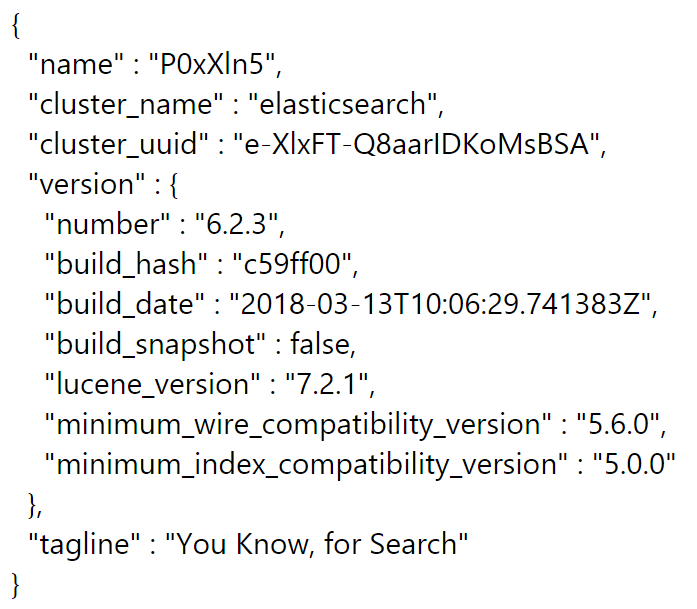
## 1.5 查看elasticsearch是否运行正常

在浏览器输入http://192.168.8.200:9200

如果使用x-pack自动生成帐号密码，需要先登录。

输入在x-pack生成的帐号：elastic 密码：yJKj6XzaCOU3MRpo6nME

返回下面结果说明运行正常。



# 2 安装中文分词器

## 2.1 安装analysis-ik

Elasticsearch 自带的分词器会粗暴地把每个汉字直接分开，没有根据词库来分词。为了处理中文搜索，还需要安装中文分词插件。我使用的是 [elasticsearch-analysis-ik](https://github.com/medcl/elasticsearch-analysis-ik/)，支持自定义词库。进入<https://github.com/medcl/elasticsearch-analysis-ik/releases>

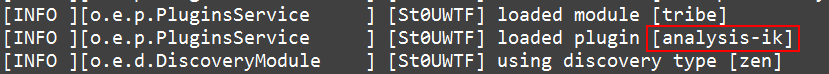
下载已经编译完成的压缩包(.zip)，注意下载对应elasticsearch版本

|  |
| --- |
| 如果下载的源码包(.tar.gz)，需要编译自行打包  yum install maven  cd elasticsearch-analysis-ik-1.9.0  mvn package  如果一切顺利，在 target/releases/ 目录下可以找到编好的文件 |

解压和把打包好的文件移动到elasticsearch/plugins的插件目录下

|  |
| --- |
| unzip elasticsearch-analysis-ik-6.2.3.zip  mv elasticsearch ik  mv ik /usr/local/elasticsearch/plugins/ |

重启elasticsearch 服务，启动信息如果有[analysis-ik]，说明elasticsearch的中文分词插件安装完成，如果下图所示：



在/usr/local/elasticsearch/plugins/ik/config/目录下有很多词表，直接用文本编辑器打开就可以修改。

## 2.2 配置同义词

Elasticsearch 自带一个名为 synonym 的同义词 filter。为了能让 IK 和 synonym 同时工作，我们需要定义新的 analyzer，用 IK 做 tokenizer，synonym 做 filter。听上去很复杂，实际上要做的只是加一段配置。

打开elasticsearch配置文件

**vim /usr/local/elasticsearch/config/elasticsearch.yml**

添加下面内容：

|  |
| --- |
| index:  analysis:  analyzer:  ik\_syno:  type: custom  tokenizer: ik\_max\_word  filter: [my\_synonym\_filter]  ik\_syno\_smart:  type: custom  tokenizer: ik\_smart  filter: [my\_synonym\_filter]  filter:  my\_synonym\_filter:  type: synonym  synonyms\_path: analysis/synonym.txt |

以上配置定义了 ik\_syno 和 ik\_syno\_smart 这两个新的 analyzer，分别对应 IK 的 ik\_max\_word 和 ik\_smart 两种分词策略。根据 IK 的文档，二者区别如下：

ik\_max\_word：会将文本做最细粒度的拆分，例如「中华人民共和国国歌」会被拆分为「中华人民共和国、中华人民、中华、华人、人民共和国、人民、人、民、共和国、共和、和、国国、国歌」，会穷尽各种可能的组合；

ik\_smart：会将文本做最粗粒度的拆分，例如「中华人民共和国国歌」会被拆分为「中华人民共和国、国歌」；

ik\_syno 和 ik\_syno\_smart 都会使用 synonym filter 实现同义词转换。为了方便后续测试，建议创建synonym.txt文件，

**vim /usr/local/elasticsearch/config/analysis/synonym.txt**

输入一些同义词并存为 utf-8 格式，例如：

|  |
| --- |
| ua,user-agent,userAgent  js,javascript  谷歌=>google |