ES安装操作手册

1. 系统参数配置

安装ES版本前需要对系统参数进行修改配置，否则ES无法正常启动，所有系统参数的增加和修改都在root用户下进行操作。

1.[root@rhino001]# vim /etc/security/limits.conf

在文件最下方添加如下内容：

|  |
| --- |
| \* soft nofile 65536  \* hard nofile 65536  \* soft nproc 65536  \* hard nproc 65536  \* soft memlock -1  \* hard memlock -1 |

2.[root@rhino001]#vi /etc/profile

添加如下两行

ulimit -SHn 65536

ulimit -SHu 65536

1. [root@rhino001]#vi /etc/sysctl.conf

|  |
| --- |
| vm.max\_map\_count=262144  vm.swappiness = 0  vm.min\_free\_kbytes = 2097152 |

然后执行sysctl -p使配置生效

1. ES版本的参数配置
2. 修改config目录下的elasticsearch.yml文件，如果是单机版，则修改以下几项配置即可。

path.data:/data01/esdata,/data02/esdata,/data03/esdata,/data04/esdata,/data05/esdata

--配置ES数据的存储路径

network.host:192.168.10.145 --配置本机ip或host名称

bootstrap.memory\_lock: true --设置为true来锁住内存不进行swapping。因为当jvm开始swapping时es的效率会降低。

如果是集群则还需添加以下两个配置：

1. discovery.zen.minimum\_master\_nodes --按照（总结点数/2 + 1）的规则来配置，如有3台服务器，则该值配为2，如果只有两台服务器，则一台做master，一台不做master，只做data。
2. discovery.zen.ping.unicast.hosts:["host1","host2"] --设置集群中master节点的初始列表，可以通过这些节点来自动发现新加入集群的节点。
3. 修改config目录下的jvm.options文件，将java虚拟机所占内存配置修改为30g，修改成-Xms30g，-Xmx30g，官方建议该值大小不超过32g。同时需要修改GC参数如下：

-XX:+UseCompressedOops

-XX:+UseG1GC

-XX:+UnlockExperimentalVMOptions

-XX:MaxGCPauseMillis=200

-XX:G1NewSizePercent=1

-XX:InitiatingHeapOccupancyPercent=65

-XX:+ParallelRefProcEnabled

-XX:ConcGCThreads=4

-XX:ParallelGCThreads=12

-XX:MaxTenuringThreshold=1

-XX:G1HeapRegionSize=32m

-XX:G1MixedGCCountTarget=32

-XX:G1OldCSetRegionThresholdPercent=2

-XX:SurvivorRatio=4

-XX:+UnlockDiagnosticVMOptions

-XX:+G1SummarizeConcMark

1. 如果ES版本部署在ARM服务器上，则需要在elasticsearch.yml文件中添加xpack.ml.enabled: false
2. 如果需要安装head插件进行集群状态以及索引信息的查看，则需要在elasticsearch.yml文件中添加：

(1).http.cors.enabled: true

(2).http.cors.allow-origin: "\*"

5.在bin目录下执行./elasticsearch -d,后台启动ES，然后通过curl ip:9200命令进行查看，如果能查看到ES的版本号等信息说明ES已正常启动。如果是集群部署，则通过curl ip:9200/\_cat/nodes查看所有节点是否都已加入集群。

1. ES常用命令
2. 创建一个名为test的索引

curl -XPUT ip:9200/test

1. 删除名为test的索引

curl -XDELETE ip:9200/test

1. 查看所有索引

curl ip:9200/\_cat/indices

1. 查看索引test对应的字段结构信息

curl ip:9200/test/\_mapping?pretty

1. 创建一个index为test，type为test，且包含字段content，age，title，类型分别为text(分词),long,keyword(不分词).

curl -XPUT 'ip:9200/test?pretty' -H 'Content-Type: application/json' -d'{

"mappings" : {

"test" : {

"properties" : {

"content": {

"type": "text"

},

"age": {

"type": "long"

},

"title": {

"type": "keyword"

}

}

}

}

}'

1. 往index为test，type为test的索引中手工插入几条数据：

(1).curl -XPOST ip:9200/test/test?pretty -H 'Content-Type: application/json' -d'{"age":29,"title":"中国人","content":"南京欢迎您"}'

(2).curl -XPOST ip:9200/test/test?pretty -H 'Content-Type: application/json' -d'{"age":30,"title":"世界杯","content":"全心全意为人民服务"}'

1. ES数据导入

大数据量往ES中进行数据导入时，目前一般采用两种方式进行导入。

1. .采用官方提供的bulk批量入库接口，语法如下：

curl -XPOST 'localhost:9200/test/test/\_bulk?pretty' -H 'Content-Type: application/json' --data-binary @1.json

其中1.json格式为：

{"index":{"\_index":"test","\_type":"test"}}

{"age":29,"title":"中国人","content":"南京欢迎您"}

{"index":{"\_index":"test","\_type":"test"}}

{"age":38,"title":"足球赛","content":"2018年世界杯在俄罗斯举行"}

{"index":{"\_index":"test","\_type":"test"}}

{"age":29,"title":"首都","content":"北京是大城市"}

{"index":{"\_index":"test","\_type":"test"}}

{"age":29,"title":"篮球","content":""}

因为官方建议执行bulk对应的json文件大小最好不要超过100M，否则会影响性能，所以采用该方法进行批量导入时需要对数据进行切分，然后通过shell脚本去执行各个json文件中数据的导入。

1. 采用etl数据导入工具进行，目前对ES5.1.1版本进行数据导入时用的工具为rhinoetl-5.1.1.tar.gz，对于ES6.X版本已开发出新的入库工具，本次重点介绍现场用到的rhinoetl-5.1.1.tar.gz。

首先对工具进行解压，配置conf目录的rhino.properties文件，cluster.name配置集群名称，默认为elasticsearch；cluster.server配置需要导入数据的节点ip和端口，端口为9300；index和type填写需要导入数据的索引名称；key.file配置key.txt的路径，其中key.txt里填写索引的字段名称；data.file为需要导入的数据路径；其他选项按默认的无需修改。

选项修改后进入bin目录执行sh start-etl.sh进行入库操作，由于入库数据量比较大，所以一般采用后台进行操作: nohup sh start-etl.sh &

1. .后台通过spark算子将数据导入到ES中
2. ES数据查询

1.REST接口查询

(1).match匹配查询

curl -XGET ip:9200/test/\_search?pretty' -d'{"query": {"match" : {"content" : "珍惜和平"}}}'

(2).term匹配查询

curl -XGET ip:9200/test/\_search?pretty' -d'{"query": {"term" : { "content" : "来之不易" } }}'

(3).查询索引test的记录数

curl ip:9200/test/\_count?pretty

1. SQL语句查询

在bin目录下执行./elasticsearch-sql-cli进入sql界面，然后执行sql查询

(1).一般查询：select \* from test where match(content,'江苏');

(2).相关度： select score() ,\* from test where match(content,'南京') ORDER BY SCORE() DESC;

(3).正则：select \* from test where title like 'k\_l%'; ---%代表多个字符，\_代表一个字符

(4).多字段查询：select \* from rhinotest where match('content,title','江苏');

(5).多条件查询：select \* from rhinotest where query('地方,学习');