Solr运维管理

www.huawei.com



前言

 本章主要介绍华为FusionInsight HD Solr服务的日常维护、 常见故障处理方法、常见调优方法以及权限说明。



- 学完本课程后,您将能够:
 - 。熟悉Solr安装目录结构;
 - □ 掌握具体的操作方法;
 - □ 掌握日常维护手段;
 - □ 掌握Solr故障处理方法;
 - □ 掌握Solr组件权限管理;



- 1. Solr目录介绍
- 2. Solr服务检查
- 3. 日志介绍
- 4. 常用维护命令或参数
- 5. 常用性能调优命令或参数
- 6. Solr资源权限管理
- 7. 常见问题

Solr安装目录

• Solr安装目录: {\$BIGDATA_HOME}/FusionInsight/FusionInsight-Solr-5.3.1

FusionInsight-Solr-5.3.1的目录结构	主要内容	
/adapter	Solr适配om代码路径	
/bin	Solr适配om脚本	
/cluster	Solr角色目录,有SolrServerAdmin、 SolrServerN(N=1~5)	
/cluster/SolrServerAdmin		
/cluster/SolrServerAdmin/apache-tomcat-7.0.68	该SolrServer实例下的tomcat	
/cluster/SolrServerAdmin/bin	该SolrServer启动、停止、状态检查等脚本	
/cluster/SolrServerAdmin/flag	原始默认配置	
/cluster/SolrServerAdmin/solrhome	该实例IP和端口的标记	
/SolrServer*		
/hbase-indexer	hbase-indexer目录	
/solr-5.3.1	Solr代码	
/solrhaserver	Solr的HA机制代码以及脚本	

Solr数据目录

• 元数据默认目录

/srv/BigData/solr/SolrServerAdmin/collection/core.properties collection表示 collection的replica的名称

/srv/BigData/solr/SolrServerN/collection/core.properties SolrServerN
N=1~5)

提示:/srv/BigData为默认路径,实际的路径以软件安装时规划的数据路径为准。

数据目录

索引数据存储在HDFS上

下载HDFS客户端,采用Hadoop命令查看

hadoop -ls /user/solr/

索引数据存储本地磁盘

默认的数据路径: /srv/BigData/solr



Solr日志目录

• 运行日志目录

默认的日志目录: /var/log/Bigdata/solr

• 审计日志目录

默认审计日志目录: /var/log/Bigdata/audit/solr



- 1. Solr目录介绍
- 2. Solr服务检查
- 3. 日志介绍
- 4. 常用维护命令或参数
- 5. 常用性能调优命令或参数
- 6. Solr资源权限管理
- 7. 常见问题

Solr服务检查

• 服务状态检查

登录FusionInsight Manager,选择服务管理 > Solr,查看Solr概述中健康状态。



Solr服务检查

• 角色实例检查

登录FusionInsight Manager,选择服务管理 > Solr > 实例,查看每个角色实例的健康状态是否为良好。

角色	主机名 ◇	操作状态	健株	េ こ
HBaseIndexer	103-30-0-150	☑已启动	♥良好	♥已同步
HBaseIndexer	103-30-0-151	♥已启动	☑良好	♥日間歩
HBaseIndexer	103-30-0-152	☑已启动	♥良好	♥已同步
SolrServer1	103-30-0-150	♥已启动	♥良好	♥已同步
SolrServer1	103-30-0-151	☑已启动	♥良好	♥已同步



- 1. Solr目录介绍
- 2. Solr服务检查
- 3. 日志介绍
- 4. 常用维护命令或参数
- 5. 常用性能调优命令或参数
- 6. Solr资源权限管理
- 7. 常见问题

日志介绍——安装日志

• Solr安装日志的存储路径为:/var/log/Bigdata/solr/角色名。

日志文件名	分析描述
cleanupDetail.log	Solr服务卸载时候的清理日志。
postinstallDetail.log	Solr服务安装前的准备工作日志。
prestartDetail.log	Solr服务启动前的准备工作日志。
solr-startup.log	Solr启动日志。
solr-stop.log	Solr停止日志。
startDetail.log	HA或者HbaseIndexer启动日志 (根据角色目录来分别)。
stopDetail.log	HA或者HbaseIndexer停止日志 (根据角色目录来分别)。



日志介绍——运行日志

• Solr运行日志的存储路径为: /var/log/Bigdata/solr/角色名

日志文件名	分析描述
solr- <hostname>.log</hostname>	Solr服务运行日志,记录Solr系统 运行时候所产生的大部分日志。
hbaseindexer-omm- <hostname>.log</hostname>	HbaseIndexer系统日志,记录HbaseIndexer 运行时候产生的大部分日志。
solr-process-check.log	Solr进程检查日志。
solr-service-check.log	Solr服务检查日志。
Ha.log	HA运行日志。
rs-floatip.log	浮动IP配置日志。
rs-sendAlarm.log	HA告警日志。
rs-solrhaserver.log	HA健康状态日志。
statusDetail.log	HA进程日志。
ha_monitor.log	HA监控日志。

日志介绍——审计日志

• Solr审计日志的存储路径为:

/var/log/Bigdata/audit/solr

日志文件名	分析描述
solr-audit- <hostname>.log</hostname>	Solr操作审计日志 (例如:添加、删除索引)。

日志介绍——常见错误日志分析

- Solr启动失败:
 - □ 1、界面上查看Solr集群中哪个节点的什么角色的实例启动失败
 - □ 2、进入Solr启动失败节点实例的日志目录:

/var/log/Bigdata/solr/角色/,查看solr-startup.log失败日志。如果启动失败是 SolrServerAdmin实例,查看失败原因是否是因为浮动ip配置失败

- 浮动ip具体失败原因查看如下日志: /var/log/Bigdata/solr/SolrHAServer/ha/script/rs-floatip.log
- 总结:如果启动失败是SolrServerAdmin实例,失败原因往往由于浮动ip已 经被使用



- 1. Solr目录介绍
- 2. Solr服务检查
- 3. 日志介绍
- 4. 常用维护命令或参数
- 5. 常用性能调优命令或参数
- 6. Solr资源权限管理
- 7. 常见问题

常用维护命令—Solr客户端安装

- 登录FusionInsight Manager系统。 在浏览器地址输入访问地址,地址格式 为http://FusionInsight Manager系统的IP地址:8080/web。
- 选择"服务管理 > Solr > 下载客户端 > 完整客户端",勾选"仅保存到服务器如下路径,如果存在客户端文件,会覆盖路径下已有的客户端文件。",
 下载客户端安装包到服务器指定目录。
- 下载完成后,会有窗口提示"客户端生成成功,保存在服务器(主管理节点)如下路径:

/tmp/FusionInsightClient/FusionInsight_V100R002C60SPC200_Solr_Clien t.tar。

使用后请及时删除。"



常用维护命令—Solr客户端安装

- 去上一步指定服务器的指定目录下,解压并安装
 - tar -xvf FusionInsight_V100R002C60SPC200_Solr_Client.tar
 - tar -xvf FusionInsight_V100R002C60SPC200_Solr_ClientConfig.tar
 - cd FusionInsight_V100R002C60SPC200_Solr_ClientConfig
 - □ ./install.sh /opt/client/(该目录为新建空目录)
 - □ 安装完成后,执行source /opt/client/bigdata_env
 - □ 如果是安全集群,则需要Kerberos 登陆,即执行kinit命令
- 完成之后,即可使用solrctl命令行工具
 - solrctl [options] command [command-arg] [command [command-arg]] ...



常用维护命令—Solr配置集操作

命令	solrctl confsetgenerate path [-schemaless]
描述	在本地指定路径生成 Solr 的配置文件集
参数说明	 generate生成配置集指令并放在本地指定路径 path为客户端本地的路径 -schemaless为可选项,生成的配置集为schemaless mode,使用该配置集新建的索引,支持使用Rest API动态地对Collection schema进行修改
示例	solrctl confsetgenerate /home/solrDir

命令	solrctl confsetcreate name path [-force]	
描述	上传本地配置文件集给 Zookeeper, 并指定其名称	
参数说明	 create为创建配置集指令 name是上传给Zookeeper的Solr配置文件集名称 path为客户端本地路径 -force为可选项,表示没有schema.xml和solrconfig.xml依旧可以上传配置文件集。如果配置集为schemaless mode,则需要指定该参数 	
示例	solrctl confsetcreate myconf /home/solrDir/	



常用维护命令—Solr配置集操作

命令	solrctl confsetupdate name path [-force]
描述	指定本地配置文件,更新指定配置集
参数说明	 update为更新配置集指令 name是要更新的配置集名称 path为配置集文件本地路径 -force为可选项,当配置集中没有schema.xml和solrconfig.xml时,需要指定该参数。如果配置集为schemaless mode,则需要指定该参数
示例	solrctl confsetupdate myconf /home/solrDir/
命令	solrctl confsetdelete name
描述	删除指定配置集
参数说明	delete为删除配置集指令name是要删除的配置集名称
示例	solrctl confsetdelete myconf
命令	solrctl confsetlist
描述	列出Solr服务中所有配置集名称



常用维护命令—Solr Collection操作

命令	<pre>solrctl collectioncreate name [-s <numshards>] [-a Create collection with autoAddReplicas=true] [-c <collection.configname>] [-r <replicationfactor>] [-m <maxshardspernode>] [-n <createnodeset>]</createnodeset></maxshardspernode></replicationfactor></collection.configname></numshards></pre>
描述	创建collection
参数说明	 create为创建collection指令 name为collection名称 -s numShards为collection中shard数目 -a 表示自动添加replica属性设置true, 当索引存储在HDFS上, 当一个solr shard replica gone时,可自动恢复 -c collection.configName为创建collection指定的配置文件 -r replicationFactor为副本数,为避免单点故障,建议设置值>=2 -m maxShardsPerNode为每个节点最大shard数 -n createNodeSet指定了创建collection的所在节点,格式: solrIP1_port, solrIP2_port
示例	solrctl collectioncreate collectionTest -c myconf -s 3 -r 1 -m 1 -n 192.168.1.1:21104,192.168.1.2:21104,192.168.1.3:21104



常用维护命令—Solr Collection操作

命令	solrctl collectiondelete name
描述	删除指定collection
参数说明	delete为删除collection指令name为需要删除collection名称
示例	solrctl collectiondelete collectionTest
命令	solrctl collectionreload name
描述	重新加载指定 collection ,当该 Collection 对应的配置集更新,可执行该命令使其生效
参数说明	 reload为重新加载collection指令 name为需要重新加载collection名称
示例	solrctl collectionreload collectionTest
命令	solrctl collectionlist
描述	列出Solr服务中所有collection



常用维护命令—HBaseIndexer操作

命令	hbase-indexer add-indexer -n indexer_name -c path/index.xml -cp solr.collection=collection_name -cp solr.zk=hostips:zkport/solr	
描述	创建Indexer	
参数说明	 add-indexer为创建一个新的Indexer命令; indexer_name为创建的indexer名 path/index.xml为配置solr索引与hbase列表映射关系的配置文件的绝对路径 solr.collection为indexer对应的collection名 solr.zk为solr在ZooKeeper中的路径 	
示例	hbase-indexer add-indexer -n userIndexer -c /opt/client/Solr/hbase-indexer/conf/hbase-demo-indexer.xml -cp solr.collection=collection1 -cp solr.zk=192.168.1.1:24002,192.168.1.2:24002,192.168.1.3:24002/solr	
命令	hbase-indexer delete-indexer -n indexer_name	
描述	删除指定的Indexer	
参数说明	 delete-indexer为删除Indexer命令 indexer_name为删除的indexer名 	
示例	hbase-indexer delete-indexer -n userIndexer	



常用维护命令—HBaseIndexer操作

命令	hbase-indexer list-indexers
描述	查看当前建立的Indexers状态



- 1. Solr目录介绍
- 2. Solr服务检查
- 3. 日志介绍
- 4. 常用维护命令或参数
- 5. 常用性能调优命令或参数
- 6. Solr资源权限管理
- 7. 常见问题

常用的性能调优命令或参数

Solr自身配置调优

场景: 优化索引方案, 需要加快索引的速度

□ 调整Solr实例的 "SOLR_GC_OPTS"参数值,加入如下参数设定直接内存大小,优化HDFS 读写速度。

-XX:MaxDirectMemorySize=4g

□ 可以通过修改 "solrconfig.xml"文件内容,来实现调优效果。

调优配置项	修改结果
提高索引速度,加大索引线 程数	<pre><maxindexingthreads>\${solr.maxIndexingThreads :16}</maxindexingthreads></pre>
增大文档索引缓存	<pre><rambuffersizemb>1024</rambuffersizemb></pre>
增大索引段合并因子	<mergefactor>20</mergefactor>
加大索引自动硬提交时间	<maxtime>\${solr.autoCommit.maxTime:60000}</maxtime>
关闭索引的自动软提交	<maxtime>\${solr.autoSoftCommit.maxTime:- 1}</maxtime>

常用的性能调优命令或参数

Solr自身配置调优

场景: 优化查询方案, 需要加快查询索引的速度

缓存在Solr中充当了一个非常重要的角色,Solr中主要包括以下3种缓存:

这里假设:用户每次查询字段数为3,实例并发数为10,不同查询和不同排序个数为500

缓存类型	修改方案
过滤器缓存	<filtercache autowarmcount="50" class="solr.FastLRUCache" initialsize="200" size="200"></filtercache> "size"和 "initialSize"值为过滤器的数量 "autowarmCount"为 "initialSiz"值的1/4
查询结果缓存	<queryresultcache autowarmcount="750" class="solr.FastLRUCache" initialsize="3000" size="3000"></queryresultcache> "size"和 "initialSize"的值 = 不同查询和不同排序的个数 × 每次查询字段数 × 2 "autowarmCount"值为 "initialSize"值的1/4
文档缓存	<pre><documentcache class="solr.FastLRUCache" initialsize="1000" size="1000"></documentcache> "size"和 "initialSize"的值 = 一次查询返回最大的文档数量× 实例查询的并发数</pre>



常用的性能调优命令或参数

• Solr自身配置调优

场景: Solr索引数据存放在HDFS上, 进行调优

启用HDFS的 "HDFS Short Circuit Local Reads"功能。

参数(name)	参数值(value)
dfs.client.read.shortcircuit	true
dfs.client.read.shortcircuit.streams.cache.size	1000
dfs.client.read.shortcircuit.streams.cache.expiry.ms	30000

□ 可以通过修改 "solrconfig.xml"文件内容,来实现调优效果。

修改参数 "solr.hdfs.blockcache.slab.count"的值

"<intname="solr.hdfs.blockcache.slab.count">\${solr.hdfs.blockcache.slab.count:1}</int>

其中每个slab的大小为128MB ,18个slab,大约占用2.3GB内存,这是每个分片Shard的配置,如果一个台主机6个shard,那么共计占用13.8GB内存。



Solr对HDFS数据索引调优建议

- 硬件组网及系统环境调优
 - □ 磁盘划分的时候要注意,zookeeper单独占用一个磁盘或一个磁盘分区,否则当数据量过大,如果和HDFS共磁盘,频繁的数据访问和配置集访问,会导致ZK挂死。
 - HDFS采用多块单盘或者多个RAIDO分别挂载多个数据目录,MR任务处理数据量很大时磁盘IO会成为瓶颈。当前测试磁盘划分仅供参考(使用命令df -h 查询,Hadoop数据盘划分如下图所示):

/dev/sdcl	1.8T	279G	1.6T	16% /srv/BigData/hadoop/datal
/dev/sddl	1.8T	268G	1.6T	15% /srv/BigData/hadoop/data2
/dev/sdel	1.8T	261G	1.6T	15% /srv/BigData/hadoop/data3
/dev/sdfl	1.8T	265G	1.6T	15% /srv/BigData/hadoop/data4
/dev/sdgl	1.8T	272G	1.6T	15% /srv/BigData/hadoop/data5
/dev/sdhl	1.8T	270G	1.6T	15% /srv/BigData/hadoop/data6
/dev/sdil	1.8T	271G	1.6T	15% /srv/BigData/hadoop/data7
/dev/sdjl	1.8T	271G	1.6T	15% /srv/BigData/hadoop/data8

Solr对HDFS数据索引调优建议

■ Solr在执行读取HDFS时,会生成很多个临时端口(连接本地DataNode),出现TIME_WAIT,最终导致任务失败。为避免此种情况,需要在每个数据节点(包括SolrServerAdmin节点)配置/etc/sysctl.conf,添加如下配置项,保存退出后,执行sysctl -p:

```
net.ipv4.tcp_syncookies = 1
net.ipv4.tcp_tw_reuse = 1
net.ipv4.tcp_tw_recycle = 1
net.ipv4.tcp_fin_timeout = 30
net.ipv4.tcp_timestamps = 1
net.ipv4.tcp_tw_recycle = 1
```

Solr对HDFS数据索引调优建议

MR任务参数调优

设置Yarn服务的配置参数,然后重启Yarn服务:

- mapreduce.task.timeout: 1800000 (原始值为60000, 处理大数据量时可以适当调大)。
- parn.nodemanager.resource.cpu-vcores: 24 (原始值为8, 该值在处理大数量时可设置为当前节点总的CPU个数的1~2倍)。
- yarn.nodemanager.resource.memory-mb: 该参数最好对每个nodemanaegr 分别配置,查看host界面每个节点的内存使用率,每个nodemanager该参数配 置为:空闲的内存数减去8G。

Solr对HBase数据索引调优建议

• 批量索引调优建议

批量索引的原理是通过创建MR任务的方式,将HBase中的数据同步到Solr当中,因此需要对影响MR任务性能相关的服务配置进行优化。

□ Yarn服务的配置优化(修改完成后需要重启服务):

参数(name)	参数值(value)
mapreduce.reduce.memory.mb	增加至8192或更大
yarn.resourcemanager,scheduler.class	修改为 org.apache.hadoop.yarn.server.resource manager.scheduler.fair.FairScheduler

Solr对HBase数据索引调优建议

- HBaseIndexer实例的配置优化,进入Solr客户端hbaseindexer的配置目录,修改以下配置 文件,参数值可以根据环境情况进行调整
- hbase-indexer-site.xml配置文件修改:

参数 (name)	参数值(value)
solr.record.writer.batch.size	500
solr.record.writer.num.threads	5
solr.record.writer.max.queues.size	3
solr.record.writer.maxSegments	5

yarn-site.xml配置文件修改:

参数 (name)	参数值(value)
mapreduce.map.speculative	false
mapreduce.reduce.speculative	false



- 1. Solr目录介绍
- 2. Solr服务检查
- 3. 日志介绍
- 4. 常用维护命令或参数
- 5. 常用性能调优命令或参数
- 6. Solr资源权限管理
- 7. 常见问题

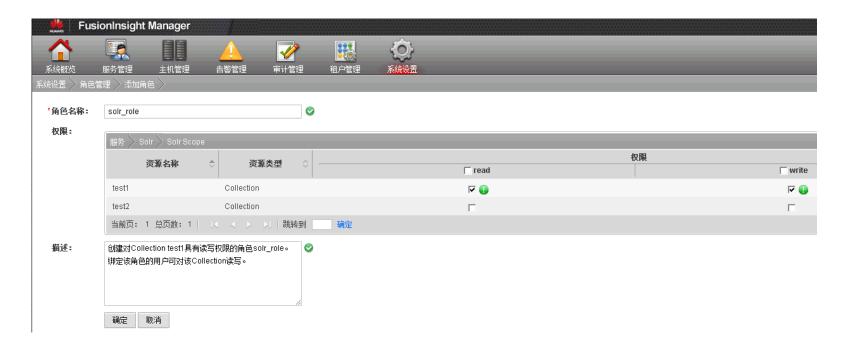
创建Solr admin角色

- 创建Solr admin role
 - 登录FusionInsight Manager,选择系统设置>角色管理,点击添加角色按钮,勾选
 SUPER_USER_GROUP,点击确定,完成Solr admin role角色创建。



创建Solr角色

- 创建对某Collection具有读写权限的Solr role
 - 登录FusionInsight Manager,选择系统设置>角色管理,点击添加角色按钮,选择Solr
 Scope,勾选资源名称为test1的read、write权限,点击确定,完成Solr role角色创建。



Solr资源权限管理

操作	所需权限
Collection Create	Solr group Solr admin role
Collection Delete	Collection owner、Solr admin role
Collection Read/Write	Collection owner、Solr Role、Solr admin role
Create/Split Shard	Collection owner、Solr admin role
Replica Add/Delete	Collection owner、Solr admin role
Solr Admin UI	Solr group、Super group、Solr admin role

业务操作权限管理

操作	所需权限	
Shell solrctl操 作	solr、hue或admin用户	
Hue Search	admin用户、hue用户	
对 HBase 数据 索引	用户属于Solr group和HBase group,且有对要访问的资源有如下权限: 要操作的HBase表:读写权限,如涉及到表创建则需要创建、管理权限 要操作的Collection:读写权限,如涉及到索引创建,则需要有创建权限 	
对 HDFS 数据 索引	用户属于Solr group和HDFS group,且有对要访问的资源有如下权限: 要操作的HDFS文件:读写权限,如涉及到文件创建删除则需要创建、管理权限 要操作的Collection:读写权限,如涉及到索引创建,则需要有创建权限	



- 1. Solr目录介绍
- 2. Solr服务检查
- 3. 日志分析
- 4. 常用维护命令或参数
- 5. 常用性能调优命令或参数
- 6. Solr资源权限管理
- 7. 常见问题



• 故障: Solr服务不可用

• 现象描述

在FusionInsight Manager 界面有Solr服务不可用告警,Solr服务的健康状态为故障

0

可能原因

- □ Solr服务所依赖的服务故障。
- □ 网络连接中断导致Solr与依赖的服务无法通讯。
- 定位思路
 - 查看ZooKeeper与HDFS服务是否正常。
 - 查看两个SolrServerAdmin的IP地址是否相同。

- 处理步骤
- 1.登录FusionInsight Manager系统。
- 2.在服务列表中查看Solr服务所依赖的ZooKeeper和HDFS服务健康状态是否为"良好"。
 - 是,执行步骤4。
 - 否,执行步骤3。
- 3.选择健康状态不是良好的服务,查看故障是否恢复。
 - 是,操作结束。
 - 否,执行步骤4。
- 4.使用PuTTY工具,以root用户登录任意一个SolrServerAdmin实例的服务器,执行ping 另一个SolrServerAdmin实例IP地址命令检查另一个SolrServerAdmin实例的服务器是否可达。
 - 是,执行步骤6。
 - 否,执行步骤5



• 处理步骤

- 5.联系系统管理员处理网络故障。若故障恢复,则操作结束,否则执行步骤6。
- 6.选择"系统设置>日志下载",获取相应日志。

在"服务"下拉框中勾选"Solr",单击"确定"。

设置日志的起始收集时间"开始时间"和"结束时间",单击"下载"。

7.联系技术支持,并发送已收集的故障日志信息。

- 故障: SolrServerAdmin主备节点间心跳中断
- 现象描述

在FusionInsight Manager 界面有SolrServerAdmin主备节点间心跳中断告警。

- 可能原因
 - SolrServerAdmin的主备实例中,有一个实例处于停止状态。
 - 主备SolrServerAdmin节点间链路异常。
- 定位思路
 - 查看是否有停止的SolrServerAdmin实例。
 - 查看两个SolrServerAdmin的IP是否相同。



• 处理步骤

- **1.**登录FusionInsight Manager界面,通过"服务管理 > Solr > 实例"查看SolrServerAdmin实例的状态是否都为已启动状态。
 - □ 是,执行步骤3。
 - □ 否,执行步骤 2。
- **2.**将停止的**SolrServerAdmin**实例选中,选择"更多操作 > 启动实例",在告警列表中查看该告警是否清除。
 - □ 是,操作结束。
 - □ 否,执行步骤 **3**。
- 3.使用PuTTY工具,以root用户登录任意一个SolrServerAdmin实例的服务器,执行ping 另一个SolrServerAdmin实例IP地址命令检查另一个SolrServerAdmin实例的服务器是否可达。
 - □ 是,执行步骤5。
 - □ 否,执行步骤 4。
- 4.联系系统管理员处理网络故障。若故障恢复,则操作结束,否则执行步骤5。



- 处理步骤
- 5.选择"系统设置 > 日志下载",获取相应日志。
 - □ 在"服务"下拉框中勾选"Solr",单击"确定"。
 - □ 设置日志的起始收集时间"开始时间"和"结束时间",单击"下载"。
 - 6.联系技术支持,并发送已收集的故障日志信息。



本章主要介绍了FusionInsight HD Solr服务的日常维护、常用的维护命令、故障处理方法、性能调优方法以及权限说明

0

习题

- 1. 【多选】下面哪些选项是查询时会用到的缓存? ()。
- A. Filter Cache

 B. Document Cache
- C. Query Result Cache D. Index Cache

- 2. 【多选】下面哪些用户可以对Collection进行更新? ()。
- A. Collection的Owner B. 属于Super group
- C. 属于Solr admin role
- D. 属于具有对该Collection具有Read/Write操作权限的角色

②思考题

1. Solr的默认日志目录是什么?当出现**Solr**服务的某个实例启动失败时,看哪个日志?

Thank you

www.huawei.com