**安装impala**

安装说明：本次安装以三节点为例。主机名根据情况修改。

1. 安装impala系统环境依赖包

在三台机器上执行以下命令安装依赖

yum install libevent -y

yum install redhat-lsb -y

yum install python-setuptools -y

yum install cyrus-sasl.x86\_64 -y

1. 安装impala

（1）三台节点安装以下两个rpm包

rpm -ivh bigtop-utils-0.7.0+cdh5.7.0+0-1.cdh5.7.0.p0.78.el6.noarch.rpm

rpm -ivh bigtop-jsvc-0.6.0+cdh5.7.0+804-1.cdh5.7.0.p0.78.el6.x86\_64.rpm

（2）在第一个节点执行

rpm -ivh impala-2.5.0+cdh5.7.0+0-1.cdh5.7.0.p0.147.el6.x86\_64.rpm --nodeps

rpm -ivh impala-state-store-2.5.0+cdh5.7.0+0-1.cdh5.7.0.p0.147.el6.x86\_64.rpm

rpm -ivh impala-catalog-2.5.0+cdh5.7.0+0-1.cdh5.7.0.p0.147.el6.x86\_64.rpm

rpm -ivh impala-udf-devel-2.5.0+cdh5.7.0+0-1.cdh5.7.0.p0.147.el6.x86\_64.rpm

1. 在一二三节点执行

rpm -ivh impala-2.5.0+cdh5.7.0+0-1.cdh5.7.0.p0.147.el6.x86\_64.rpm --nodeps

rpm -ivh impala-shell-2.5.0+cdh5.7.0+0-1.cdh5.7.0.p0.147.el6.x86\_64.rpm

rpm -ivh mpala-udf-devel-2.5.0+cdh5.7.0+0-1.cdh5.7.0.p0.147.el6.x86\_64.rpm

rpm -ivh impala-server-2.5.0+cdh5.7.0+0-1.cdh5.7.0.p0.147.el6.x86\_64.rpm

1. 每个节点都要拷贝core-site，hdfs-site，hive-site到/etc/impala/conf下

**Core-site内容如下**

<configuration>

<!-- 指定hdfs的nameservice为ns1 -->

<property>

<name>fs.defaultFS</name>

<value>hdfs://hzgc/</value>

</property>

<!-- 指定hadoop临时目录 -->

<property>

<name>hadoop.tmp.dir</name>

<value>/usr/local/hadoop-2.7.2/tmp</value>

</property>

<!-- 指定zookeeper地址 -->

<property>

<name>ha.zookeeper.quorum</name>

<value>s114:2181,s115:2181,s116:2181</value>

</property>

<property>

<name>io.compression.codecs</name>

<value>org.apache.hadoop.io.compress.GzipCodec,org.apache.hadoop.io.compress.DefaultCodec,org.apache.hadoop.io.compress.BZip2Codec,org.apache.hadoop.io.compress.SnappyCodec</value>

</property>

</configuration>

**Hdfs-site内容如下**

<configuration>

<!--指定hdfs的nameservice为dksou，需要和core-site.xml中的保持一致 -->

<property>

<name>dfs.nameservices</name>

<value>hzgc</value>

</property>

<!-- dksou下面有两个NameNode，分别是nn1，nn2 -->

<property>

<name>dfs.ha.namenodes.hzgc</name>

<value>nn1,nn2</value>

</property>

<!-- nn1的RPC通信地址 -->

<property>

<name>dfs.namenode.rpc-address.hzgc.nn1</name>

<value>s114:9000</value>

</property>

<!-- nn1的http通信地址 -->

<property>

<name>dfs.namenode.http-address.hzgc.nn1</name>

<value>s114:50070</value>

</property>

<!-- nn2的RPC通信地址 -->

<property>

<name>dfs.namenode.rpc-address.hzgc.nn2</name>

<value>s115:9000</value>

</property>

<!-- nn2的http通信地址 -->

<property>

<name>dfs.namenode.http-address.hzgc.nn2</name>

<value>s115:50070</value>

</property>

<!-- 指定NameNode的元数据在JournalNode上的存放位置 -->

<property>

<name>dfs.namenode.shared.edits.dir</name>

<value>qjournal://s114:8485;s115:8485;s116:8485/hzgc</value>

</property>

<!-- 指定JournalNode在本地磁盘存放数据的位置 -->

<property>

<name>dfs.journalnode.edits.dir</name>

<value>/usr/local/hadoop-2.7.2/data/tmp/journal</value>

</property>

<!-- 开启NameNode失败自动切换 -->

<property>

<name>dfs.ha.automatic-failover.enabled</name>

<value>true</value>

</property>

<!-- 配置失败自动切换实现方式 -->

<property>

<name>dfs.client.failover.proxy.provider.hzgc</name>

<value>org.apache.hadoop.hdfs.server.namenode.ha.ConfiguredFailoverProxyProvider</value>

</property>

<!-- 配置隔离机制方法，多个机制用换行分割，即每个机制暂用一行-->

<property>

<name>dfs.ha.fencing.methods</name>

<value>

sshfence

shell(/bin/true)

</value>

</property>

<!-- 使用sshfence隔离机制时需要ssh免登陆 -->

<property>

<name>dfs.ha.fencing.ssh.private-key-files</name>

<value>/root/.ssh/id\_rsa</value>

</property>

<!-- 配置sshfence隔离机制超时时间 -->

<property>

<name>dfs.ha.fencing.ssh.connect-timeout</name>

<value>30000</value>

</property>

<property>

<name>dfs.permissions</name>

<value>false</value>

</property>

<property>

<name>dfs.replication</name>

<value>2</value>

</property>

<property>

<name>dfs.client.read.shortcircuit</name>

<value>true</value>

</property>

<property>

<name>dfs.domain.socket.path</name>

<value>/var/run/hadoop-hdfs/dn.\_PORT</value>

</property>

<property>

<name>dfs.client.file-block-storage-locations.timeout.millis</name>

<value>10000</value>

</property>

<property>

<name>dfs.client.read.shortcircuit.skip.checksum</name>

<value>false</value>

</property>

<property>

<name>dfs.datanode.hdfs-blocks-metadata.enabled</name>

<value>true</value>

</property>

<property>

<name>dfs.client.use.legacy.blockreader.local</name>

<value>false</value>

</property>

<property>

<name>dfs.datanode.data.dir.perm</name>

<value>750</value>

</property>

<property>

<name>dfs.block.local-path-access.user</name>

<value>e3base</value>

</property>

<property>

<name>hadoop.native.lib</name>

<value>true</value>

<description>Should nativehadoop libraries, if present, be used.</description>

</property>

</configuration>

**Hive-site内容如下**

<configuration>

<property>

<name>javax.jdo.option.ConnectionURL</name>

<value>jdbc:mysql://s114:3306/hive?createDatabaseIfNotExist=true&amp;useUnicode=true&amp;characterEncoding=latin1</value>

<description>JDBC connect string for a JDBC metastore</description>

</property>

<property>

<name>javax.jdo.option.ConnectionDriverName</name>

<value>com.mysql.jdbc.Driver</value>

<description>Driver class name for a JDBC metastore</description>

</property>

<property>

<name>javax.jdo.option.ConnectionUserName</name>

<value>root</value>

<description>username to use against metastore database</description>

</property>

<property>

<name>javax.jdo.option.ConnectionPassword</name>

<value>123456</value>

<description>password to use against metastore database</description>

</property>

<property>

<name>hive.server2.thrift.port</name>

<value>10000</value>

</property>

<property>

<name>hive.server2.thrift.bind.host</name>

<value>s115</value>

</property>

<property>

<name>hive.metastore.uris</name>

<value>thrift://s115:9083</value>

</property>

<property>

<name>hive.metastore.client.socket.timeout</name>

<value>3600</value>

<description>MetaStore Client sockettimeout in seconds</description>

</property>

</configuration>

1. 三个节点执行以下命令

adduser impala

chmod u+w /etc/sudoers

echo impala ' ALL=(ALL) NOPASSWD: NOPASSWD: ALL' >> /etc/sudoers

chmod u-w /etc/sudoers

1. 三个结点都要修改/etc/default/impala（三个节点内容一致）

IMPALA\_CATALOG\_SERVICE\_HOST=s114

IMPALA\_STATE\_STORE\_HOST=s114

IMPALA\_STATE\_STORE\_PORT=24000

IMPALA\_BACKEND\_PORT=22000

IMPALA\_LOG\_DIR=/var/log/impala

IMPALA\_CATALOG\_ARGS=" -log\_dir=${IMPALA\_LOG\_DIR} "

IMPALA\_STATE\_STORE\_ARGS=" -log\_dir=${IMPALA\_LOG\_DIR} -state\_store\_port=${IMPALA\_STATE\_STORE\_PORT}"

IMPALA\_SERVER\_ARGS=" \

-log\_dir=${IMPALA\_LOG\_DIR} \

-catalog\_service\_host=${IMPALA\_CATALOG\_SERVICE\_HOST} \

-state\_store\_port=${IMPALA\_STATE\_STORE\_PORT} \

-use\_statestore \

-state\_store\_host=${IMPALA\_STATE\_STORE\_HOST} \

-be\_port=${IMPALA\_BACKEND\_PORT}"

ENABLE\_CORE\_DUMPS=false

# LIBHDFS\_OPTS=-Djava.library.path=/usr/lib/impala/lib

# MYSQL\_CONNECTOR\_JAR=/usr/share/java/mysql-connector-java.jar

# IMPALA\_BIN=/usr/lib/impala/sbin

# IMPALA\_HOME=/usr/lib/impala

# HIVE\_HOME=/usr/lib/hive

# HBASE\_HOME=/usr/lib/hbase

# IMPALA\_CONF\_DIR=/etc/impala/conf

# HADOOP\_CONF\_DIR=/etc/impala/conf

# HIVE\_CONF\_DIR=/etc/impala/conf

# HBASE\_CONF\_DIR=/etc/impala/conf

HIVE\_HOME=/usr/local/apache-hive-1.1.0-bin

HBASE\_HOME=/usr/local/hbase-1.2.0

HADOOP\_HOME=/usr/local/hadoop-2.7.2

1. 三个节点都要拷贝mysql-connector-java.jar到/usr/share/java/
2. 三个节点都要rm -rf /usr/lib/impala/lib/\*
3. 三个结点都要拷贝lib包到/usr/lib/impala/（自己提供）
4. 三个节点都要mkdir /var/run/hadoop-hdfs
5. 三个节点都要在/etc/default/bigtop-utils添加以下内容

export JAVA\_HOME=/opt/tool/jdk

export CLASSPATH=./:$JAVA\_HOME/lib:$JAVA\_HOME/lib/tools.jar

export PATH=$JAVA\_HOME/bin:$HIVE\_HOME/bin:$PATH