**HƯỚNG DẪN CÀI APACHE HIVE**

**I.Cài Java 8**

Do java 14 cài Apache Hive không được nên chúng ta sẽ cài Java 8 bằng câu lệnh sau:

*sudo apt-get install openjdk-8-jre*

Sau đó thay đổi JAVA\_HOME trong file .bashrc, thực hiện như sau:

*nano ~/.bashrc*

Tìm đến dòng export JAVA\_HOME và thay đổi như sau:

*export JAVA\_HOME=/usr/lib/jvm/java-8-openjdk-amd64*

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

Sau đó lưu lại cấu hình bằng câu lệnh:

*source ~/.bashrc*

Đồng thời thay đổi đường dẫn trong file hadoop-env.sh

*cd hadoop/etc/hadoop*

*nano hadoop-env.sh*

Tìm đến dòng export JAVA\_HOME và thay đổi như sau:

*export JAVA\_HOME=/usr/lib/jvm/java-8-openjdk-amd64*

Nhấn tổ hợp Ctrl+X, và nhấn Y để lưu.

**II. Cài Apache Hive 3.1.2**

Bước 1: Tải Apache Hive

*cd /home/hadoopuser*

*wget https://downloads.apache.org/hive/hive-3.1.2/apache-hive-3.1.2-bin.tar.gz*

Bước 2: Giải nén và đổi tên thư mục

*tar -xzf apache-hive-3.1.2-bin.tar.gz*

*mv apache-hive-3.1.2-bin hive*

Bước 3: Cấu hình

* Thêm đường dẫn HIVE\_HOME vào file .bashrc như sau:

*export HIVE\_HOME=/home/hadoopuser/hive*

*export HIVE\_CONF\_DIR=$HIVE\_HOME/conf*

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

*export CLASSPATH=$CLASSPATH:/home/hadoopuser/hadoop/lib/\*:.*

*export CLASSPATH=$CLASSPATH:/home/hadoopuser/hive/lib/\*:.*

* Lưu lại cấu hình:

*source ~/.bashrc*

Bước 4: Kiểm tra version:

*hive –version*

Bước 5: Tạo nơi chứa data trên HDFS

*hdfs dfs -mkdir -p /user/hadoopuser/hive/warehouse*

*hdfs dfs -mkdir /tmp*

Bước 6: Cấp quyền cho các thư mục vừa tạo ở bước 5:

*hdfs dfs -chmod g+w /user/hadoopuser/hive/warehouse*

*hdfs dfs -chmod g+w /tmp*

Bước 7: Thiết lập Hadoop trong file hive-env.sh

Đang ở thư mục hadoopuser

*cd hive/conf*

*cp hive-env.sh.template hive-env.sh*

*nano hive-env.sh*

Thêm các dòng sau vào file hive-env.sh:

*export HADOOP\_HOME=/home/hadoopuser/hadoop*

*export HADOOP\_HEAPSIZE=512*

*export HIVE\_CONF\_DIR=/home/hadoopuser/hive/conf*

Nhấn Ctrl+X, gõ Y để lưu lại thay đổi.

Sau đó thêm properties vào file hive-site.xml, các bước thực hiện như sau:

*cd $HIVE\_HOME/conf*

*cp hive-default.xml.template hive-site.xml*

*nano hive-site.xml*

thêm các dòng sau vào <configuration></configuration> trong file:

<property>

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

<value>

jdbc:derby:;databaseName=/home/hadoopuser/hive/metastore\_db;create=true

</value>

<description>

JDBC connect string for a JDBC metastore.

To use SSL to encrypt/authenticate the connection, provide database-specific SSL flag in the connection URL.

For example, jdbc:postgresql://myhost/db?ssl=true for postgres database.

</description>

</property>

**Thay đổi các dòng sau trong file hive-site.xml:**

* <name>hive.metastore.warehouse.dir</name>:

+ Nhấn Ctrl+W và nhập vào: hive.metastore.warehouse.dir và enter

+ thay đổi <value>/user/hive/warehouse</value> thành <value>/user/hadoopuser/hive/warehouse</value>

* <name>hive.exec.local.scratchdir</name>:

+ Nhấn Ctrl+W và nhập vào: hive.exec.local.scratchdir và enter

+ Thay đổi <value>…</value>:

<!--<value>…</value> -- >

* <name>hive.download.resource.dir</name>:

+ Nhấn Ctrl+W và nhập vào: hive.download.resource.dir và enter

+ Thay đổi <value>…</value>:

<!--<value>…</value> -- >

**III. Cài Apache Derby 10.15.2.0**

Bước 1: Tải Derby

*cd /home/hadoopuser*

*wget http://archive.apache.org/dist/db/derby/db-derby-10.15.2.0/db-derby-10.15.2.0-bin.tar.gz*

Bước 2: Giải nén

*tar –xzf db-derby-10.15.2.0-bin.tar.gz*

*mv db-derby-10.15.2.0-bin derby*

Bước 3: Thiết lập biến môi trường

*nano ~/.bashrc*

Thêm các dòng sau vào cuối file .bashrc

*export DERBY\_HOME=/home/hadoopuser/derby*

*export PATH=$PATH:$DERBY\_HOME/bin*

*export CLASSPATH=$CLASSPATH:$DERBY\_HOME/libderby.jar:$DERBY\_HOME/lib/derbytools.jar*

Sau đó lưu cấu hình: *source ~/.bashrc*

Tạo nơi chứa Metastore data:

*mkdir $DERBY\_HOME/data*

*cd hive/conf*

*touch jpox.properties*

*nano jpox.properties*

Thêm các dòng sau vào file jpox.properties

*javax.jdo.PersistenceManagerFactoryClass =org.jpox.PersistenceManagerFactoryImpl  
org.jpox.autoCreateSchema = false  
org.jpox.validateTables = false  
org.jpox.validateColumns = false  
org.jpox.validateConstraints = false  
org.jpox.storeManagerType = rdbms  
org.jpox.autoCreateSchema = true  
org.jpox.autoStartMechanismMode = checked  
org.jpox.transactionIsolation = read\_committed  
javax.jdo.option.DetachAllOnCommit = true  
javax.jdo.option.NontransactionalRead = truejavax.jdo.option.ConnectionDriverName = org.apache.derby.jdbc.ClientDriverjavax.jdo.option.ConnectionURL = jdbc:derby://hadoop1:1527/metastore\_db;create = truejavax.jdo.option.ConnectionUserName = APPjavax.jdo.option.ConnectionPassword = mine*

Nhấn Ctrl+X, nhập Y để lưu file.

Bước 4: Cấp quyền Hive sử dụng metastore schema của derby như sau:

cd $HIVE\_HOME

bin/schematool -initSchema -dbType derby

Sau đó chạy lệnh hive để kiểm tra cài đặt Apache Hive.