Hive

1. Install Hive and MetaStore

- $ sudo yum install -y hive hive-metastore

2. Add driver Mysql to Hive path

- $ sudo ln -s /usr/share/java/mysql-connector-java.jar /usr/lib/hive/lib

3. Create database MetaStore

- $ mysql -u root -p
- mysql> CREATE DATABASE metastore;
- mysql> USE metastore;
- mysql> SOURCE /usr/lib/hive/scripts/metastore/upgrade/mysql/hive-schema-0.10.0.mysql.sql;

4. Create Mysql user for Hive configuration

- mysql> CREATE USER 'hive'@'localhost' IDENTIFIED BY 'password';
- ...
- mysql> REVOKE ALL PRIVILEGES, GRANT OPTION FROM 'hive'@'localhost';
- mysql> GRANT SELECT,INSERT,UPDATE,DELETE,LOCK TABLES,EXECUTE ON metastore.* TO 'hive'@'localhost';
- mysql> FLUSH PRIVILEGES;
- mysql> quit;

5. Configure Hive

- $ sudo vi /etc/hive/conf/hive-site.xml

```
<?xml version="1.0" encoding="UTF-8"?>
<configuration>
<property>
        <name>javax.jdo.option.ConnectionURL</name>
        <value>jdbc:mysql://localhost/metastore</value>
        <description>the URL of the MySQL database</description>
</property>
<property>
        <name>javax.jdo.option.ConnectionDriverName</name>
        <value>com.mysql.jdbc.Driver</value>
</property>
<property>
        <name>javax.jdo.option.ConnectionUserName</name>
        <value>hive</value>
</property>
<property>
        <name>javax.jdo.option.ConnectionPassword</name>
        <value>password</value>
<description>Indicar la contraseña de la bd</description>
</property>
<property>
        <name>datanucleus.autoCreateSchema</name>
        <value>false</value>
</property>
<property>
        <name>datanucleus.fixedDatastore</name>
        <value>true</value>
</property>
```

```
<property>
        <name>datanucleus.autoStartMechanism</name>
        <value>SchemaTable</value>
</property>
<property>
        <name>hive.metastore.uris</name>
        <value>thrift://127.0.0.1:9083</value>
        <description>IP address (or fully-qualified domain name) and port of the
metastore host</description>
</property>
</configuration>
```

6. Copy configuration to all nodes

- $ for i in **tiger horse monkey**; do sshpass -p 'cloudera' scp /etc/hive/conf/hive-site.xml root@$i:/etc/hive/conf/hive-site.xml; done

7. Create path on HFDS for Hive

- sudo -u hdfs hdfs dfs -mkdir /tmp
- sudo -u hdfs hdfs dfs -chmod 1777 /tmp
- sudo -u hdfs hdfs dfs -mkdir /user/hive/warehouse
- sudo -u hdfs hdfs dfs -chmod 1777 /user/hive/warehouse
- sudo -u hdfs hdfs dfs -ls /user/hive

8. Start MetaStore

   $ sudo service hive-metastore restart

- Start on boot
  $ sudo chkconfig hive-metastore on

9. Connect to client Hive

- $ hive
- hive> SHOW TABLES;
- hive> quit;

10. Create metadata from HDFS files

- $ hive
- hive> CREATE EXTERNAL TABLE musicbrainz_tracks (trackid INT, releaseid INT, trackname STRING, position INT) ROW FORMAT DELIMITED FIELDS TERMINATED BY '\t' LOCATION '/user/cloudera/musicbrainz_tracks';
- hive> CREATE EXTERNAL TABLE musicbrainz_releases (id INT, album STRING, artist STRING) ROW FORMAT DELIMITED FIELDS TERMINATED BY '\t' LOCATION '/user/cloudera/musicbrainz_releases';

11. Count records from musicbrainz_tracks

- hive> SELECT COUNT(*) FROM musicbrainz_tracks;

12. Go to job id given in terminal, for example:

- http://horse:50030/jobdetails.jsp?jobid=job_201410072245_0004

13. Try a JOIN query

- select t.trackname, t.position, r.album, r.artist from musicbrainz_tracks t JOIN musicbrainz_releases r ON (t.releaseid=r.id) **LIMIT 10**;