Sqoop

1. Install Sqoop

   - # yum install -y sqoop mysql libmysql-java

2. Create symbolic link to MySql driver JDBC
   - # ls /usr/share/java/mysql-connector-java.jar
   - # ls /usr/lib/sqoop/lib/mysql-connector-java.jar -l

   - If not exists symbolic link, create it
     # ln -s /usr/share/java/mysql-connector-java.jar /usr/lib/sqoop/lib/

3. Configure MySQL to start on boot
   - # chkconfig mysqld on

4. Start MySQL
   - # service mysqld start

5. Create password for root
   - # /usr/bin/mysqladmin -u root password 'cloudera'

6. Create database
   - $ mysql -u root -p
   - mysql> create database musicbrainz;
   - mysql> grant select,insert,update,delete ON musicbrainz.* TO cloudera@localhost IDENTIFIED BY 'cloudera';

7. Import schema musicbrainz
   - wget ftp://tallergpul/files/musicbrainz.tar.gz
   - tar zxvf musicbrainz.tar.gz
   - $ mysql -u root -p musicbrainz < musicbrainz.sql
8. Show data from tables
   - mysql> select t.trackname, t.position, r.album, r.artist from musicbrainz_tracks t JOIN musicbrainz_releases r ON (t.releaseid=r.id) LIMIT 0,10;
9. Export tables to HDFS

- Table musicbrainz_tracks

  ```
  $ sqoop import --connect jdbc:mysql://localhost/musicbrainz --table musicbrainz_tracks --columns "trackid, releaseid, trackname, position" --fields-terminated-by '\t' --username cloudera --password cloudera
  ```

- Table musicbrainz_releases

  ```
  $ sqoop import --connect jdbc:mysql://localhost/musicbrainz --table musicbrainz_releases --columns "id, album, artist" --fields-terminated-by '\t' --username cloudera --password cloudera
  ```

10. Check paths and files in HDFS

- $ hadoop fs -ls musicbrainz_releases

- `$ hadoop fs -tail musicbrainz_releases/part-m-00003`
- `$ hadoop fs -ls musicbrainz_tracks/`