Experiment 4: Sqoop and Hive

1. jps
2. open mysql

create table empsal (eid int, esal int);

insert into empsal values (1,1000), (2,2000), (3,3000), (4,4000), (5,5000);

select \* from empsal;

1. Import data into hdfs using sqoop

after jps command, type:

sqoop import --connect jdbc:mysql://localhost:3306/emp --table empsal --username hadoop --password 123456 --target-dir /sqooptranser/emp -m 1

<press enter key>

1. Load recently imported hdfs data into hive and perform analysis:

type command in terminal

hive

in hive prompt, type

use test;

create external table emp (eid int, esal int) ROW FORMAT DELIMITED FIELDS TERMINATED BY ‘,’ LINES TERMINATED BY ‘\n’ LOCATION ‘/sqooptranser/emp1’;

select \* from emp;

select \* from emp where esal>3000;

**if mysql does not work**

**ssh-keygen**

**ssh-copy-id -i.ssh/id\_rsa.pub hadoop@localhost**

**sudo mysql -u root**

**mysql> create user 'hadoop'@'localhost' identified by '123456'** ;

**grant all privileges on \*.\* to 'hadoop'@'localhost';**

**flush privileges;**

**If NameNode is not coming in jps**

**killall java**

**rn -rf hdfsdrive/\***

**hadoop namenode -format**

**./start-hadoop.sh**

**jps**

**SQOOP EXPERIMENT**

**./start-hadoop.sh**

**jps**

**mysql -u hadoop -p123456**

**mysql> create database demo;**

**use demo;**

**create table emp(eid int, esal int);**

**insert into emp values(1,1000),(2,2000),(3,3000),(4,4000),(5,5000),(6,6000);**

**To load data from mysql to a file in hdfs**

**sqoop import --connect jdbc:mysql://localhost:3306/demo --username hadoop --password 123456 --table emp --target-dir /sqooptrf/emp -m 1**

**Enter in browser localhost:50070**

**Go to utilities->browse the file system**

**Click on sqooptrf**

**Click on emp**

**Click on part-m-00000**

**Click on download**

**Open from downloads** part-m-00000 **file. This file has the records from table in database**

**How to fetch data from hdfs file to a table in hive**

**hive**

**hive> create database test1;**

**Create external table. Source is** /sqooptrf/emp i.e. where it was in website

**hive> create external table emp (eid int, esal int) ROW FORMAT DELIMITED FIELDS TERMINATED BY ',' LINES TERMINATED BY '\n' LOCATION '/sqooptrf/emp';**

**select \* from emp;**