COMMON ERRORS WHILE RUNNING THE SQOOP

q)Job tracker in safe mode
A) Run the below command from the HOME directory to leave the safe mode.
hadoop dfsadmin -safemode leave

q)

****
Installing SQOOP(1.4.4) in UBUNTU(14.04) for HADOOP (1.2.1)

Download link for sqoop
http://archive.apache.org/dist/sqoop/1.4.4/

Select the hadoop-1.0.0.tar,gz file for download.

Unzip and rename the folder to sqoop and move it to $HOME

$HOME is the preffered sqoop installation directory.

Setup MySQL database on ubuntu system running hadoop

#sudo apt-get update
#sudo apt-get upgrade
#sudo apt-get install mysql-server
Run the above commands from the home directory

Check if all the packages have installed by executing this command

```
dpkg -l | grep
mysql
```

This will list all the additional packages that would be installed with sql server like client , client core, server core etc

Note : During the installation phase, the system will prompt for password for the "root" user. (Username is root here)

#mysql -u root – p

The above command will ask for password to take us to the sql shell
sql >
sql > show databases ;

Will list the available databases

Download MySQL JDBC driver

https://dev.mysql.com/downloads/file.php?id=13597

We have to download, unzip and copy the .jar file from the the unzipped folder and place it in sqoop/lib

Unzip the ZIP file using the following command (only if its a zip file, else use tar -xzvf filename if its a tarball)
unzip file.zip -d

Note: At times we need to configure the sqoop-env.sh in the sqoop/conf directory
1)copy the sqoop-env-template.sh to sqoop-env.sh file using the below command
cp sqoop-env-template.sh sqoop-env.sh

Add the following lines in the above file (vi sqoop-env.sh)
#Set path to where bin/hadoop is available
export HADOOP_COMMON_HOME=/home/hduser/hadoop

#Set path to where hadoop-*-core.jar is available
export HADOOP_MAPRED_HOME=/home/hduser/hadoop

To check if sqoop is configured with the right kind of connectors, we need to check using the following commands.
#sqoop help
#sqoop version

We have finished installing SQOOP and configuring it.
We have finished installing MySQL and we have to create databases, tables and load the tables with data.

MY SQL commands to create a table and load it with data

create database sample ;

use sample ;

create table department (dept_no char(2), dept_name varchar(20), primary key (dept_no));

insert into department values ('01','HR')

Note: Insert some 10 rows into the table
We shall create another table called employees

create table employee (empid int(3), empname varchar(30), empage int(3), primary key (emp age));

Load the above table with some data

insert into employee values ('001','Ram','30')

SQOOP commands

Latest sqoop link
http://www.eu.apache.org/dist//sqoop/1.99.6/

```
sudo chown hduser:hadoop -R /usr/local/sqoop/
```

NOTE: (before executing any sqoop commands, ensure the priovilages are granted to the user) in the sql> prompt execute the below command
GRANT ALL PRIVILEGES ON * . * TO 'root'@'localhost';

To list the databases

sqoop list-databases --connect jdbc:mysql://localhost/ --username root --password hadoop

sqoop **import** --connect jdbc:mysql://localhost/sample --username root --password hadoop --table department

Importing a table without private keys (setting the mapper to 1)

sqoop **import** --connect jdbc:mysql://localhost/sample --username root --password hadoop --table department --m 1

BY USING alt column and (default 4 reducers)

sqoop import --connect jdbc:mysql://localhost/sample --username root –password hadoop --table company **--split-by cid**

sqoop import --connect jdbc:mysql://localhost/sample --username root –password hadoop --table company **--split-by cid**

sqoop import --connect jdbc:mysql://localhost/sample --username root –password hadoop --table company --split-by cid --incremental append --check-column cid --last-value 10

Note : for every new row that is fetched, it will go into a new file in the folder (folder name and table being imported will have same name)

sqoop import --connect jdbc:mysql://localhost/sample --username root –password hadoop --table company --split-by cid --incremental lastmodified --check-column cname --last-value "IBM"

NOTE: The above lastmodified will work only if the column is date or time field !

IMPORTING A SUBSET OF A TABLE !

sqoop import --connect jdbc:mysql://localhost/sample --username root --password hadoop --table company --where "cid> 3“ --target-dir /home/hduser/part_company

q)How to create a sqoop job

sqoop job --create import-company -- import --connect jdbc:mysql://localhost/sample --username root --password hadoop --table company –m 1 --target-dir company2

Note: "import-company" is the name of the job !

This job will be stored in the sqoop metastore. We can list all the jobs using the below command
sqoop job --list

sqoop job --show <name of the job>

sqoop job --delete <name of the job>

Q)How to execute a saved job
sqoop job --exec <name of the job>

Password options ! (command line and a file)

sqoop import --connect jdbc:mysql://localhost/sample --username root -P --table company –m 1 –target-dir /user/hduser/company1

**********some issue with password file option**********
sqoop import --connect jdbc:mysql://localhost/sample --username root --password-file /user/hduser/mysql_pwd/sqoop.password --table company –m 1 --target-dir /user/hduser/company4

OPTIONS FILE and its contents !
-----name of the file could be------ (my_options_file)
import --connect jdbc:mysql://localhost/sample --username root –password hadoop

How to execute the sqoop command with the help of options-file !

sqoop --options-file < path of the options file > --table <table name> --target-dir

EXPORTING TABLES FROM HDFS TO rdbms USING SQOOP

sqoop export --connect jdbc:mysql://localhost/sample --username root --password hadoop --table company_new --export-dir <hdfs path >

Note : the file in HDFS must be a CSV and a new empty table must be already created in MySQL by the name (company_new)

Updating a table in MySQL using SQOOP and the data in HDFS

sqoop export --connect jdbc:mysql://localhost/employes --username root --password hadoop --table departments_new --export-dir <directory path> --update-key dept_no

the dept_no is the column name in the database table to which the new data from the HDFS will be updated to. We have to ensure that the hdfs data has the updated set of data in the same schema as that of the RDBMS table !

Update old rows and insert new rows ...

sqoop export --connect jdbc:mysql://localhost/employes --username root --password hadoop --table departments_new --export-dir <directory path> --update-key dept_no --update-mode allowinsert

Delimiter options

sqoop export --connect jdbc:mysql://localhost/sample --username root --password hadoop --table Ptemp_emp --export-dir sample_data/temp_emp.noncsv --input-fields-terminated-by \t