sqoop import
If you are implementing in emr pls remember these points
sqoop installed in /usr/lib/sqoop/ folder.
So if you want to import data from oracle place all jars at /usr/lib/sqoop/lib

mysql url: jdbc:mysql://mysqldb.ck8xvu8hemdj.ap-south-1.rds.amazonaws.com:3306/mysqldb
oracle url: jdbc:oracle:thin://@oracledb.ck8xvu8hemdj.ap-south-1.rds.amazonaws.com:1521/orcl
mssql url: jdbc:sqlserver://mssql.ck8xvu8hemdj.ap-south-1.rds.amazonaws.com:1433

//import from mysql
sqoop import --connect jdbc:mysql://mysqldb.ck8xvu8hemdj.ap-south-1.rds.amazonaws.com:3306/mysqldb
--username musername --password mpassword --table emp --target-dir=/user/venu/empdata -m 1

//if you are not mention target-dir by default its store in /user/hadoop/ path

sqoop import --connect jdbc:mysql://mysqldb.ck8xvu8hemdj.ap-south-1.rds.amazonaws.com:3306/mysqldb
--username musername --password mpassword --table emp -m 1

//hide password
sqoop import --connect jdbc:mysql://mysqldb.ck8xvu8hemdj.ap-south-1.rds.amazonaws.com:3306/mysqldb
--username musername --password-file s3://apachesparktraining/Training/jars/mpassword.txt --target-dir /testing
--table emp -m 1

//import from oracle

sqoop import --connect jdbc:oracle:thin://@oracledb.ck8xvu8hemdj.ap-south-1.rds.amazonaws.com:1521/orcl
--username ousername --password opassword --table CUSTOMER --target-dir=/user/venu/CUSTOMERdata -m
1
//download jdbc drive and place in sqoop lib folder and warehouse path /usr/lib/sqoop/lib

so download oracle jar ojdbc7.jar from website, with winscp place in emr, than run this command
sudo cp ./*.jar /usr/lib/sqoop/lib
//test these files exists or not
ls /usr/lib/sqoop/lib

//import data based on query ... $CONDITIONS is mandatory its reference as its condition

sqoop import --connect jdbc:oracle:thin://@oracledb.ck8xvu8hemdj.ap-south-1.rds.amazonaws.com:1521/orcl
--username ousername --password opassword --query 'SELECT * FROM OUSERNAME.EMP where sal >2999
and $CONDITIONS' --target-dir=/user/venu/emptarget -m 1

//when u r importing data if already data exists, overwrite it using --delete-target-dir

 sqoop import --connect jdbc:oracle:thin://@oracledb.ck8xvu8hemdj.ap-south-1.rds.amazonaws.com:1521/orcl
--username ousername --password opassword --table EMP --delete-target-dir -m 1

```
// you can join two tables and import the data
sqoop import --connect jdbc:oracle:thin://@oracledb.ck8xvu8hemdj.ap-south-1.rds.amazonaws.com:1521/orcl
--username ousername --password opassword --query 'SELECT b.dname, b.loc, a.ename FROM emp a JOIN
dept b on (a.deptno = b.deptno) WHERE $CONDITIONS'  --target-dir=/user/venu/querybased -m 1


 //check this data is stored in hdfs or not
hdfs dfs -cat /user/venu/emptarget/part*

//if already in hdfs store / append data in hdfs using --append

sqoop import --connect jdbc:oracle:thin://@oracledb.ck8xvu8hemdj.ap-south-1.rds.amazonaws.com:1521/orcl
--username ousername --password opassword --query 'SELECT * FROM OUSERNAME.EMP where sal <2999
and $CONDITIONS' --append --target-dir=/user/venu/emptarget -m 1

//import all tables
sqoop  import-all-tables --connect
jdbc:mysql://mysqldb.ck8xvu8hemdj.ap-south-1.rds.amazonaws.com:3306/mysqldb --username musername
--password mpassword  -m 1


//import all except these tables

sqoop import-all-tables --connect
'jdbc:oracle:thin://@oracledb.ck8xvu8hemdj.ap-south-1.rds.amazonaws.com:1521/ORCL' --username
ousername --password opassword  --warehouse-dir /user/hive/warehouse  --exclude-tables
'AIRLINE_PARQUET,AIRLINE,AIRLEMPINE_PARQUET,CUSTOMERTAB,EMP,MILLIONDATA,ORAT
AB' -m 1
hdfs dfs -ls /user/hadoop/

//overwrite change warehouse path
sqoop  import-all-tables --connect
jdbc:mysql://mysqldb.ck8xvu8hemdj.ap-south-1.rds.amazonaws.com:3306/mysqldb --username musername
--password mpassword  --warehouse-dir /user/oracle -m 1
//warehouse
sqoop import --connect
'jdbc:oracle:thin://@oracledb.ck8xvu8hemdj.ap-south-1.rds.amazonaws.com:1521/ORCL' --username
ousername --password opassword --table EMP --warehouse-dir /user/hive/warehouse   -m 1




//import data from mssql
/////////////////////////
CREATE TABLE tempdb.dbo.Products
  (ProductID int PRIMARY KEY NOT NULL,
   ProductName varchar(25) NOT NULL,
   Price money NULL,
```

ProductDescription text NULL)

INSERT tempdb.dbo.Products (ProductID, ProductName, Price, ProductDescription)
    VALUES (1, 'Clamp', 12.48, 'Workbench clamp')

 INSERT tempdb.dbo.Products (ProductName, ProductID, Price, ProductDescription)
    VALUES ('Screwdriver', 50, 3.17, 'Flat head')

 INSERT tempdb.dbo.Products
    VALUES (75, 'Tire Bar', NULL, 'Tool for changing tires.')

 INSERT Products (ProductID, ProductName, Price)
    VALUES (3000, '3mm Bracket', .52)

 UPDATE tempdb.dbo.Products
    SET ProductName = 'Flat Head Screwdriver'
    WHERE ProductID = 50

 https://docs.microsoft.com/en-us/sql/t-sql/lesson-1-creating-database-objects

 sqoop import --connect
'jdbc:sqlserver://mssql.ck8xvu8hemdj.ap-south-1.rds.amazonaws.com:1433;database=tempdb' --username
msusername --password mspassword --driver com.microsoft.sqlserver.jdbc.SQLServerDriver --table
"dbo.Products" --target-dir=/user/venu/mssqlimport -m 1

//like oracle, place jdbc driver in sqoop folder

//if you want to import data from mssql or teradata or redshift table, ist mandatory use --driver

//import data to hive

sqoop import --connect jdbc:oracle:thin://@oracledb.ck8xvu8hemdj.ap-south-1.rds.amazonaws.com:1521/orcl
--username ousername --password opassword --table EMP --hive-import -m 1
//now when u r importing data from oracle table name emp, so in hive also tablename is emp if you want to
overwrite tablename use --hive-table

sqoop import --connect jdbc:oracle:thin://@oracledb.ck8xvu8hemdj.ap-south-1.rds.amazonaws.com:1521/orcl
--username ousername --password opassword --table EMP --hive-import --hive-table newemp -m 1

//if you have already table overwrite its
 sqoop import --connect
jdbc:oracle:thin://@oracledb.ck8xvu8hemdj.ap-south-1.rds.amazonaws.com:1521/ORCL --username ousername
--password opassword --table EMP --hive-overwrite -m 1

if you want to import hive data in particular target folder target-dir doesn't support alternatively use
warehouse-dir
sqoop import --connect jdbc:oracle:thin://@oracledb.ck8xvu8hemdj.ap-south-1.rds.amazonaws.com:1521/orcl

--username ousername --password opassword --table EMP --hive-import --target-dir=/user/hive/hiveimp -m 1

sqoop import --connect jdbc:oracle:thin://@oracledb.ck8xvu8hemdj.ap-south-1.rds.amazonaws.com:1521/orcl --username ousername --password opassword --table EMP --hive-import --warehouse-dir=/user/hive/ -m 1

//import data in avro format

sqoop import --connect jdbc:oracle:thin://@oracledb.ck8xvu8hemdj.ap-south-1.rds.amazonaws.com:1521/ORCL --username ousername --password opassword --table EMP --as-avrodatafile -m 1

hdfs dfs -cat /user/hadoop/EMP/part*

//import to in avro format

sqoop import --connect jdbc:oracle:thin://@oracledb.ck8xvu8hemdj.ap-south-1.rds.amazonaws.com:1521/ORCL --username ousername --password opassword --table EMP --as-avrodatafile -m 1

//import in parquet format
sqoop import --connect jdbc:oracle:thin://@oracledb.ck8xvu8hemdj.ap-south-1.rds.amazonaws.com:1521/ORCL --username ousername --password opassword --table EMP --as-parquetfile -m 1

//Sequencefile

sqoop import --connect jdbc:oracle:thin://@oracledb.ck8xvu8hemdj.ap-south-1.rds.amazonaws.com:1521/ORCL --username ousername --password opassword --table EMP --as-sequencefile -m 1

//when you are importing data by default row separated by , but if you want to separated by tab \t use --fields-terminated-by '\t'

sqoop import --connect 'jdbc:oracle:thin://@oracledb.ck8xvu8hemdj.ap-south-1.rds.amazonaws.com:1521/ORCL' --username ousername --password opassword --table ASL --target-dir=/tabdatatarget/ --fields-terminated-by '\t' -m 1

--incremental import
/////////////////////////
its useful in streaming data.

sqoop import --connect 'jdbc:oracle:thin://@oracledb.ck8xvu8hemdj.ap-south-1.rds.amazonaws.com:1521/ORCL' --username ousername --password opassword --table EMP --check-column empno --incremental append --last-value 7934 --warehouse-dir /user/hive/warehouse -m 1

sqoop job --create myjob import --connect 'jdbc:oracle:thin://@oracledb.ck8xvu8hemdj.ap-south-1.rds.amazonaws.com:1521/ORCL' --username ousername --password opassword --table EMP --check-column empno --incremental append --last-value 7934 --warehouse-dir /user/hive/warehouse -m 1
sqoop job --list

sqoop job --exec myjob

//export/////////very head ache

first create a table in oracle
put data in hdfs folder only
hdfs dfs -mkdir /asldata/

than export the database

sqoop export --connect
'jdbc:oracle:thin://@oracledb.ck8xvu8hemdj.ap-south-1.rds.amazonaws.com:1521/ORCL' --username
ousername --password opassword --table ASL --export-dir=/asldata/   -m 1

sqoop export --connect
'jdbc:oracle:thin://@oracledb.ck8xvu8hemdj.ap-south-1.rds.amazonaws.com:1521/ORCL' --username
ousername --password opassword --table ASL --export-dir=/samp/   -m 1
//its append
//now if you want to update data in oracle follow this command.
sqoop export --connect
'jdbc:oracle:thin://@oracledb.ck8xvu8hemdj.ap-south-1.rds.amazonaws.com:1521/ORCL' --username
ousername --password opassword --table ASL --update-key NAME --update-mode allowinsert
--export-dir=/tabdata --input-fields-terminated-by ',' -m 2

https://stackoverflow.com/questions/25887086/sqoop-export-using-update-key

//limits if you want skip header in sqoop not possible
similarly when u r importing its not possible to import schema.
especially mal records, very headache in sqoop

sqoop export --connect
'jdbc:oracle:thin://@oracledb.ck8xvu8hemdj.ap-south-1.rds.amazonaws.com:1521/ORCL' --username
ousername --password opassword --table USADATAONLY --export-dir=/USDATA/   -m 1

//WHEN YOU are export data, bydefault fields terminated by ',' so instead of , if you want tab \t separated use
--input-fields-terminated-by '\t';
means if you have each data separated by tab use it
let eg:
cat >tabdata.txt
venkat 88 hyd
priya 44 blr
munna 74 ogl
susila 99 blr
ctrl+d
hdfs dfs -mkdir /tabdata

```
hdfs dfs -put tabdata.txt /tabdata
 sqoop export --connect
'jdbc:oracle:thin://@oracledb.ck8xvu8hemdj.ap-south-1.rds.amazonaws.com:1521/ORCL' --username
ousername --password opassword --table ASL --export-dir=/tabdata/ --input-fields-terminated-by '\t'  -m 1
```