==========

Import Oracle Data into Hadoop using Sqoop

==========

Step 1: Download oracle 12c Express Edition from oracle website

https://www.oracle.com/database/technologies/xe-downloads.html

Note: *Needs to register for download. Please register it by using your email id*

Download oracle jdbc jar from the below url
https://repo1.maven.org/maven2/com/oracle/database/jdbc/ojdbc8/21.1.0.0/ojdbc8-21.1.0.0.jar

Step 2: Install dependencies, database and configure DB

cd /etc/yum.repos.d

sudo wget http://yum.oracle.com/public-yum-ol7.repo

sudo sed -i 's/enabled=0/enabled=1/g' /etc/yum.repos.d/public-yum-ol7.repo

sudo sed -i 's/gpgcheck=1/gpgcheck=0/g' /etc/yum.repos.d/public-yum-ol7.repo

sudo yum -y install oracle-database-preinstall-21c

cd ~/Downloads

sudo yum -y localinstall oracle-database-xe-21c-1.0-1.ol7.x86_64.rpm

sudo /etc/init.d/oracle-xe-21c configure

Step 3: Add environment variables

vi ~/.bashrc

#add below lines at the end of .bashrc file

export ORACLE_SID=XE

export ORACLE_BASE=/opt/oracle/oradata

export ORACLE_HOME=/opt/oracle/product/21c/dbhomeXE

export PATH=$PATH:/opt/oracle/product/21c/dbhomeXE/bin

#Save and exit

source ~/.bashrc

Step 5: Set flag to start an Oracle Database automatically

sudo vi /etc/oratab

XE:/opt/oracle/product/21c/dbhomeXE:Y

#Save and exit

Step 6: Configure Linux to automatically start Oracle Database when the system starts

sudo vi /lib/systemd/system/dbora.service

#Copy paste below part (between START and END) to dbora.service file

===START===

[Unit]
Description=The Oracle Database Service
After=syslog.target network.target

[Service]
Type=forking
RemainAfterExit=yes
KillMode=none
User=oracle
ExecStart=/opt/oracle/product/21c/dbhomeXE/bin/dbstart
/opt/oracle/product/21c/dbhomeXE
ExecStop=/opt/oracle/product/21c/dbhomeXE/bin/dbstop
/opt/oracle/product/21c/dbhomeXE

[Install]
WantedBy=multi-user.target
===END===

#Save and exit

Step 7: Start the service

#Enable service

sudo systemctl enable dbora.service

sudo systemctl daemon-reload

sudo systemctl start dbora.service

sudo systemctl stop firewalld

sudo systemctl disable firewalld

Step 8: Run below command in Oracle SQLPlus to create user

sqlplus
Enter username: system
Enter password: hduser

SQL> alter session set "_ORACLE_SCRIPT"=true;
SQL> CREATE USER hduser IDENTIFIED BY hduser;
SQL> grant connect to hduser;
SQL> grant ALL PRIVILEGES TO hduser;
SQL> grant unlimited tablespace to hduser;
SQL> quit;

Step 9: Login again to oracle sqlplus as hduser, create table and insert data

sqlplus

Enter username: hduser
Enter password: hduser

SQL> alter session set "_ORACLE_SCRIPT"=true;

SQL> create table customer(custid int,firstname varchar(100),lastname varchar(100),age int,profession varchar(200));
insert into customer values(400001,'Kristina','Chung',55,'Pilot');
insert into customer values(400002,'Paige','Chen',74,'Teacher');
insert into customer values(400003,'Sherri','Melton',34,'Lawyer');
insert into customer values(400004,'Gretchen','Hill',66,'Carpenter');
insert into customer values(400005,'Karen','Puckett',74,'Doctor');

SQL> create table tutorials (id int, tutorial_name varchar(100));
insert into tutorials values(101,'Hadoop Learning');
insert into tutorials values(102,'Sqoop Tutorials');
insert into tutorials values(103,'Hive Workouts');
insert into tutorials values(104,'Spark Internals');
insert into tutorials values(105,'Python Tutorials');

SQL> commit;
*SQL> select * from customer;*

--To get the table details
SQL> describe customer;

--To list out the tables
SQL> SELECT table_name FROM user_tables;

SQL> quit;

Step 10: Install Hadoop and Sqoop

#Run the script to install hadoop and sqoop
bash ~/Downloads/Hadoop_Sqoop_Install.sh

Step 11: Sqoop import from oracle to hdfs

sudo cp ~/Downloads/ojdbc8-21.1.0.0.jar /usr/local/sqoop/

***sqoop import --connect jdbc:oracle:thin:@localhost:1521:XE ***
--username hduser --password hduser --table "CUSTOMER" -m 1

***sqoop import --connect jdbc:oracle:thin:@localhost:1521:XE ***
***--username hduser --password hduser --table "TUTORIALS" -m 1 ***
--target-dir /user/hduser/tutorials