Date: 26-04-2022

NAME – KAVEEN GANDHI

EMPLOYEE ID: 2135539

TASK [hdfs, Sqoop and Hive]

1. Create hdfs directory by name “mydata” under “/user/ubh01”
   Steps :
   - Step 1: Start both the services i.e, Hadoop and Yarn
     $HADOOP_HOME/sbin/start-dfs.sh
     $HADOOP_HOME/sbin/start-yarn.sh
   - Step 2: Check whether all nodes are running, so we use **sudo jps** command
   - Step 3: **hdfs dfs -mkdir /user/ubh01/mydata**
   - **Hdfs dfs -ls**
   - From step 3, the command will create a directory mydata in ubh01

2. Import table called “device” under devsh_loudacre database from Mysql into hdfs under /user/ubh01/mydata location.

   Steps:
   1. Start sqoop, To see all commands for sqoop TYPE sqoop help,
      Check the database using command:
      Sudo list-databases --connect jdbc:mysql://ubh01 --username sqoop --password password .
      List of databases is displayed.
   2. Check whether the Device table is available in devsh_loudacre databases using command,
      Sudo list-tables --connect jdbc:mysql://ubh01/devsh_loudacre --username sqoop --password password
   3. If the device table is present in the desh_loudacre database run the below command,
      Sudo import --connect jdbc:mysql://ubh01/devsh_loudacre --table device --username sqoop --password password --target-dir /user/ubh01/mydata –fields-terminated-by ‘\t’

      The above command import device table from devsh_loudacre fr0m mysql to hdfs under the directory /user/ubh01/mydata in part files (part-m-00000)

4. To check the files and directory is present use the below command:
   Hdfs dfs -ls (list of all files)
   Hdfs dfs -ls /user/ubh01/mydata (list of files in mydata)

3. Check from namenode WebUI how many blocks got created.

   Step 1: Open firefox browser
   Step 2: Use the port number for namenode webui is 50070
   Step 3: http://ubh01:50070 (search for namenode webui)

4. Create hive table called “mydevices” on hive based on below sample data of the devices table imported under /user/ubh01/mydata.

   Note: Table must not be a managed table.

device_num	release_dt	device_name	device_type
1	2008-10-21 00:00:00	Sorrento F00L	phone
2	2010-04-19 00:00:00	Titanic 2100	phone
3	2011-02-18 00:00:00	MeeToo 3.0	phone
4	2011-09-21 00:00:00	MeeToo 3.1	phone
5	2008-10-21 00:00:00	iFruit 1	phone

   1. Start HADOOP and YARN services, check whether all nodes are running or not by using this command **sudo jps**
   2. Type **hive** (to start hive)
   3. **Show databases; (**check list of databases)
   4. **Use default;** (default databases is present already so we are using default)
   5. **Show tables;** (lists of tables)
   6. **Create external table mydevice(device_num int, release_dt string, device string, device _type string)**
      row format delimited
      fields terminated by ‘,’
      location ‘/user/ubh01/mydata’;
   7. The above tables is in external which means the table can be droped from metadata
   8. The output of this command is **OK, the table is created**
   9. **Describe mydevice;** (describe the table mydevice i.e, datatype of the column in the table)

5. Filter the records from mydevices with device_name as “Sorrento…”
   Select * from mydevice where device_name like ‘Sorrento%’;

6. Count the number of each device_name.
   Select count(device_name) from mydevice group by device_name;

د