**Connect to Oracle Cloud:**

ssh jliu2@129.150.69.91

**Create a folder called Painpill:**

hdfs dfs -mkdir Painpill

**Create folders in Painpill:**

hdfs dfs -mkdir /user/jliu2/Painpill/top15ms

hdfs dfs -mkdir /user/jliu2/Painpill/productms

**In case for having error message (Error: Error while compiling statement: FAILED: RuntimeException…) in hive/beeline:**

hdfs dfs -mkdir tmp

**Give the permission to beeline for edit files/data:**

hdfs dfs -chmod -R o+w .

**Connect to beeline/hive:**

beeline

!connect jdbc:hive2://bigdai-nov-bdcsce-1:2181,bigdai-nov-bdcsce-2:2181,bigdai-nov-bdcsce-3:2181/;serviceDiscoveryMode=zooKeeper;zooKeeperNamespace=hiveserver2?tez.queue.name=interactive bdcsce\_admin

create database jliu2;

use jliu2;

**Create main table which will include everything from the data:**

create external table if not exists painpilldata (

reporter\_dea\_no string, reporter\_bus\_act string, reporter\_name string, reporter\_addl\_co\_info string, reporter\_address1 string, reporter\_address2 string, reporter\_city string, reporter\_state string, reporter\_zip string, reporter\_county string, buyer\_dea\_no string, buyer\_bus\_act string, buyer\_name string, buyer\_addl\_co\_info string, buyer\_address1 string, buyer\_address2 string, buyer\_city string, buyer\_state string, buyer\_zip string, buyer\_county string, transaction\_code string, drug\_code string, ndc\_no string, drug\_name string, quantity int, unit int, action\_indicator string, order\_form\_no string, correction\_no string, strength string, transaction\_date string, calc\_base\_wt\_in\_gm float, dosage\_unit int, transaction\_id string, product\_name string, ingredient\_name string, measure string, mme\_conversion\_factor string, combined\_labeler\_name string, revised\_company\_name string, reporter\_family string, dos\_str int )

ROW FORMAT DELIMITED

FIELDS TERMINATED BY '\t'

STORED AS TEXTFILE

LOCATION '/user/jliu2/Painpill'

TBLPROPERTIES ("skip.header.line.count"="1");

**Using describe and select from function to check the table is created correct:**

describe painpilldata;

select \* from painpilldata limit 10;

**Create a Market Share by Product table:**

drop table if exists ms\_product;

create table if not exists ms\_product

row format delimited fields terminated by '\t'

stored as textfile location '/user/jliu2/Painpill/productms'

as

select product\_name, sum(quantity) as quantity

from painpilldata group by product\_name order by quantity desc;

**May also use describe and select from function to check the table.**

**Create a table by Top 15 Products’ Market Share:**

create table if not exists top15msp

row format delimited fields terminated by '\t'

stored as textfile location '/user/jliu2/Painpill/top15ms'

as

with top15 as (

select product\_name, quantity from ms\_product order by quantity desc limit 15)

select \* from top15

union all

select "all other" as product\_name, sum(quantity) as quantity

from ms\_product

where product\_name not in (select product\_name from top15)

**Exit hive/beeline or open another terminal connect to Oracle Cloud**

**Check the files and folders in top15ms folder:**

hdfs dfs -ls ./Painpill/top15ms

**Rename/Move the files:**

hdfs dfs -mv ./Painpill/top15ms/1/000000\_0 ./Painpill/top15ms/1/000000\_1

hdfs dfs -mv ./Painpill/top15ms/1/000000\_1 ./Painpill/top15ms/000000\_1

hdfs dfs -mv ./Painpill/top15ms/2/000000\_0 ./Painpill/top15ms/000000\_2

**Combine the files as named “top15productmarketshare” and get it to Oracle Cloud:**

hdfs dfs -cat /user/jliu2/Painpill/top15ms/000000\_\* | hdfs dfs -put - /user/jliu2/Painpill/top15productmarketshare

hdfs dfs -get /user/jliu2/Painpill/top15productmarketshare

**Open a new terminal and Download the file to local pc:**

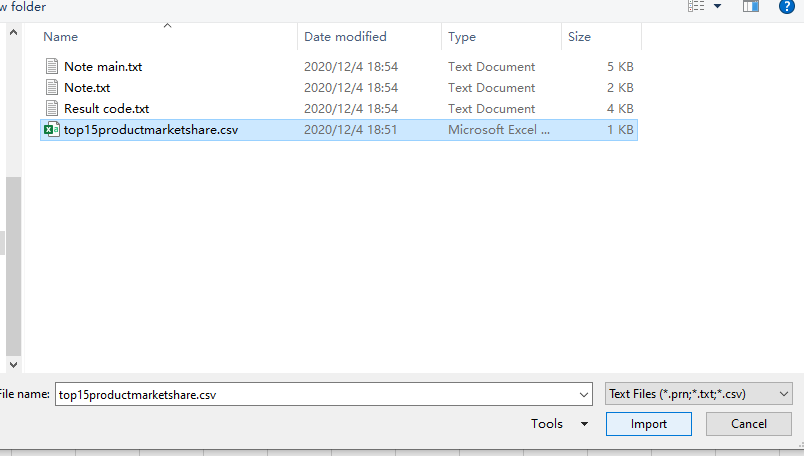
scp jliu2@129.150.69.91:/home/jliu2/top15productmarketshare top15productmarketshare.csv

**Open Excel and load the data/.csv file to excel file as following:**

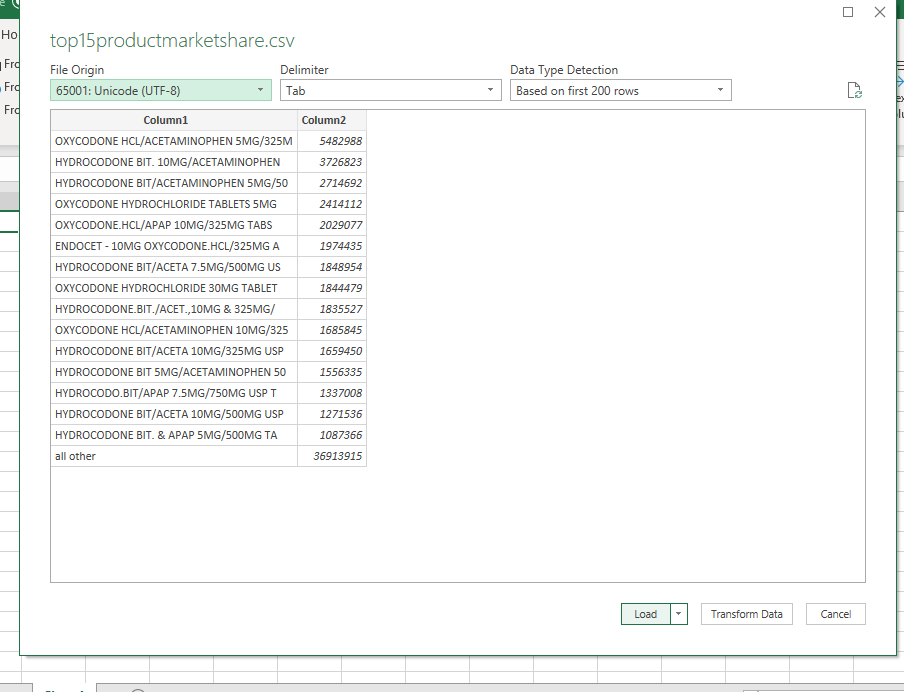
**Data🡪 From Text/CSV -> Import**

**Graphical user interface, application, table

Description automatically generated**

****

**Leave the default setting or you may change the file origin to Unicode (UTF-8) if you want and Load the data:**

****

**Changing Column1 and Column2 to “Product Name” and “Quantity”:**

**Graphical user interface, text, application, table

Description automatically generated**

**Select the cell from A2 to B17 and click Insert -> Charts -> 2-D Pie -> Pie:**

**Graphical user interface, application, table

Description automatically generated**

**Right Click the Chart area and select Move Chart -> New sheet**

**And Now you can change with the Chart Design and Format option:**

**Chart, pie chart

Description automatically generated**