Vansh Sharma

1NT19IS181

C2 Batch

Date: 23/06/2022

Experiment no. 5

Download the file as .csv file. Using the Sales.csv file, create a database called "salesdb". Create a table under that with the heading "sales". (Adjust the rows by looking at the csv file columns)

1. Insert 5 records using Insert command in HiveQL

2. Import the dataset from the sales.csv file

```
hive> select * from cred;
OK
Product1        1500    Axis     ABC      Bangalore         karnataka        India
Product NULL    Payment_Type     Name     City     State    Country
Product1        1200    Mastercard        carolina          Basildon         England United Kingdom
Product1        1200    Visa     Betina   Parkville         MO       United States
Product1        1200    Mastercard        Federica e Andrea          Astoria OR      United States
Product1        1200    Visa     Gouya    Echuca   Victoria          Australia
Product2        3600    Visa     Gerd W   Cahaba Heights   AL       United States
Product1        1200    Visa     LAURENCE          Mickleton         NJ       United States
Product1        1200    Mastercard        Fleur    Peoria   IL       United States
Product1        1200    Mastercard        adam     Martin   TN       United States
Product1        1200    Mastercard        Renee Elisabeth Tel Aviv           Tel Aviv        Israel
Product1        1200    Visa     Aidan    Chatou   Ile-de-France    France
Product1        1200    Diners   Stacy    New York          NY       United States
Product1        1200    Amex     Meidi    Eindhoven         Noord-Brabant    Netherlands
Product1        1200    Mastercard        Sean     Shavano Park     TX       United States
Product1        1200    Visa     Georgia Eagle    ID       United States
Product1        1200    Visa     Richard Riverside         NJ       United States
Product1        1200    Diners   Leanne   Julianstown       Meath    Ireland
```

3. Count the number of sales done by each country

```
hive> select country,count(*) from cred group by country;
Query ID = hdoop_20220622150858_54785f0a-cb37-4cf0-95e2-34bfe5b8b97f
Total jobs = 1
Launching Job 1 out of 1
Number of reduce tasks not specified. Estimated from input data size: 1
In order to change the average load for a reducer (in bytes):
  set hive.exec.reducers.bytes.per.reducer=<number>
In order to limit the maximum number of reducers:
  set hive.exec.reducers.max=<number>
In order to set a constant number of reducers:
  set mapreduce.job.reduces=<number>
Starting Job = job_1655888206664_0003, Tracking URL = http://admin1-HP-280-G4-MT-Business-PC:8088/proxy/application_1655888206664
_0003/
Kill Command = /home/hdoop/hadoop-3.2.1/bin/mapred job  -kill job_1655888206664_0003
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2022-06-22 15:09:05,186 Stage-1 map = 0%,  reduce = 0%
2022-06-22 15:09:09,281 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 1.32 sec
2022-06-22 15:09:13,429 Stage-1 map = 100%,  reduce = 100%, Cumulative CPU 2.75 sec
MapReduce Total cumulative CPU time: 2 seconds 750 msec
Ended Job = job_1655888206664_0003
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1  Reduce: 1   Cumulative CPU: 2.75 sec   HDFS Read: 70928 HDFS Write: 1444 SUCCESS
Total MapReduce CPU Time Spent: 2 seconds 750 msec
OK
```

```
Japan    2
Jersey   1
Kuwait   1
Latvia   1
Luxembourg       1
Malaysia         1
Malta    2
Mauritius        1
Moldova 1
Monaco  2
Netherlands      22
New Zealand      6
Norway  16
Philippines      2
Poland  2
Romania 1
Russia  1
South Africa     5
South Korea      1
Spain    12
Sweden  13
Switzerland      36
Thailand         2
The Bahamas      2
Turkey  6
Ukraine 1
United Arab Emirates      6
United Kingdom  100
United States    462
Time taken: 16.707 seconds, Fetched: 58 row(s)
hive>
```

```
Argentina        1
Australia        38
Austria 7
Bahrain 1
Belgium 8
Bermuda 1
Brazil  5
Bulgaria         1
CO       1
Canada  76
Cayman Isls      1
China    1
Costa Rica       1
Country 1
Czech Republic   3
Denmark 15
Dominican Republic        1
Finland 2
France  27
Germany 25
Greece  1
Guatemala        1
Hong Kong        1
Hungary 3
Iceland 1
India    3
```

4. Count the number of sales done by each state

```
hive> select state,count(*) from cred group by state;
Query ID = hdoop_20220622151300_b74a958b-cd3b-4aa4-9e2c-c3936e728bb1
Total jobs = 1
Launching Job 1 out of 1
Number of reduce tasks not specified. Estimated from input data size: 1
In order to change the average load for a reducer (in bytes):
  set hive.exec.reducers.bytes.per.reducer=<number>
In order to limit the maximum number of reducers:
  set hive.exec.reducers.max=<number>
In order to set a constant number of reducers:
  set mapreduce.job.reduces=<number>
Starting Job = job_1655888206664_0004, Tracking URL = http://admin1-MP-280-G4-MT-Business-PC:8088/proxy/application_1655888206664
_0004/
Kill Command = /home/hdoop/hadoop-3.2.1/bin/mapred job  -kill job_1655888206664_0004
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2022-06-22 15:13:05,231 Stage-1 map = 0%,  reduce = 0%
2022-06-22 15:13:09,395 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 1.32 sec
2022-06-22 15:13:13,527 Stage-1 map = 100%,  reduce = 100%, Cumulative CPU 2.77 sec
MapReduce Total cumulative CPU time: 2 seconds 770 msec
Ended Job = job_1655888206664_0004
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1  Reduce: 1   Cumulative CPU: 2.77 sec   HDFS Read: 71010 HDFS Write: 4814 SUCCESS
Total MapReduce CPU Time Spent: 2 seconds 770 msec
OK
```

```
Bern     1                          Georgia 1
Black River      1                  Groningen        2
Bohol   1                           Guangdong        1
British Columbia         18         Guatemala        1
Brittany        2                   HI      9
Brussels (Bruxelles)     6          Hainaut 1
Bucuresti       1                   Hamburg 1
Budapest        1                   Hamilton         1
Buenos Aires    1                   Haryana 1
Buskerud        1                   Heredia 1
CA      66                          Hessen  2
CO      10                          Nordaland        2
CT      9                           IA      3
Catalonia       3                   ID      2
Ceara   1                           IL      16
Centennial      1                   IN      2
Central Bohemia 1                   Ile-de-France    8
Chisinau        1                   Istanbul        5
Cork    10                          Ita-Suomen Laani         1
DC      3                           Izmir   1
DE      1                           KS      2
Distrito Nacional        1          KY      2
Dubayy  5                           Kerry   1
Dublin  22                          Kiev    1
East Bohemia    1                   Kilkenny         1
England 86                          Kobenhavn        7
Etela-Suomen Laani       1          Krung Thep       1
FL      29                          Kuala Lumpur     1
Franche-Comte   2                   KwaZulu-Natal    2
Frederiksborg   3                   LA      2
Freeport        1                   Lazio   1
```

5. Display (Product, name) grouped by product

```
hive> select prod_id,count(uname) from cred group by prod_id;
Query ID = hdoop_20220622152000_6477cbcf-183c-499f-9dc3-bed1d2a0c428
Total jobs = 1
Launching Job 1 out of 1
Number of reduce tasks not specified. Estimated from input data size: 1
In order to change the average load for a reducer (in bytes):
  set hive.exec.reducers.bytes.per.reducer=<number>
In order to limit the maximum number of reducers:
  set hive.exec.reducers.max=<number>
In order to set a constant number of reducers:
  set mapreduce.job.reduces=<number>
Starting Job = job_1655888206664_0005, Tracking URL = http://admin1-HP-280-G4-MT-Business-PC:8088/proxy/application_1655888206664
_0005/
Kill Command = /home/hdoop/hadoop-3.2.1/bin/mapred job -kill job_1655888206664_0005
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2022-06-22 15:20:06,711 Stage-1 map = 0%,  reduce = 0%
2022-06-22 15:20:10,802 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 1.31 sec
2022-06-22 15:20:14,896 Stage-1 map = 100%,  reduce = 100%, Cumulative CPU 2.52 sec
MapReduce Total cumulative CPU time: 2 seconds 520 msec
Ended Job = job_1655888206664_0005
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1  Reduce: 1   Cumulative CPU: 2.52 sec   HDFS Read: 71267 HDFS Write: 183 SUCCESS
Total MapReduce CPU Time Spent: 2 seconds 520 msec
OK
```

```
OK
Product 1
Product1         848
Product2         136
Product3          15
Time taken: 16.848 seconds, Fetched: 4 row(s)
```

6. Create separate views for VISA and Mastercard

```
hive> create view visa as (
    > select * from sales where payment_type='Visa');
OK
Time taken: 1.437 seconds
hive> select * from visa;
OK
Product1      1200    Visa    Betina  Parkville         MO       United States
Product1      1200    Visa    Gouya   Echuca  Victoria          Australia
Product2      3600    Visa    Gerd W  Cahaba Heights  AL       United States
Product1      1200    Visa    LAURENCE        Mickleton         NJ       United
States
Product1      1200    Visa    Aidan   Chatou  Ile-de-France    France
Product1      1200    Visa    Georgia Eagle   ID      United States
Product1      1200    Visa    Richard Riverside         NJ       United States
Product1      1200    Visa    Janet   Ottawa  Ontario Canada
Product2      3600    Visa    Sabine  London  England United Kingdom
Product1      1200    Visa    Jeremy  Manchester      England United Kingdom
Product1      1200    Visa    asuman  Chula Vista     CA       United States
Product1      1200    Visa    Lisa    Sugar Land      TX       United States
Product1      1200    Visa    chris   London  England United Kingdom
```

```
hive> create view mastercard as (
    > select * from sales where payment_type='Mastercard');
OK
Time taken: 0.12 seconds
```

```
hive> select * from mastercard;
OK
Product1       1200    Mastercard      carolina        Basildon        England
United Kingdom
Product1       1200    Mastercard      Federica e Andrea       Astoria OR    U
nited States
Product1       1200    Mastercard      Fleur   Peoria  IL      United States
Product1       1200    Mastercard      adam    Martin  TN      United States
Product1       1200    Mastercard      Renee Elisabeth Tel Aviv        Tel Avi
v       Israel
Product1       1200    Mastercard      Sean    Shavano Park    TX      United
States
Product1       1200    Mastercard      Nicola  Roodepoort      Gauteng South A
frica
Product1       1200    Mastercard      Lena    Kuopio  Ita-Suomen Laani       F
inland
Product1       1200    Mastercard      Katherine       New York        NY     U
nited States
Product1       1200    Mastercard      Linda   Miami   FL      United States
Product1       1200    Mastercard      SYLVIA  Vesenaz Geneve  Switzerland
Product1       1200    Mastercard      Stephanie       Badhoevedorp    Noord-H
olland  Netherlands
Product1       1200    Mastercard      Anneli  Houston TX      United States
```

7. Show all the transactions done in Seattle

```
hive> select * from cred where city="Seattle";
OK
Product1        1200    Visa    Vicki and Bill  Seattle WA      United States
Product1        1200    Visa    Craig   Seattle WA      United States
Product1        1200    Visa    Kim     Seattle WA      United States
Product2        3600    Visa    cristina        Seattle WA      United States
Product1        1200    Amex    Justine Seattle WA      United States
Product1        1200    Visa    Reto    Seattle WA      United States
Time taken: 0.103 seconds, Fetched: 6 row(s)
hive>
```

8. Find the max number of transactions done within the state of
Ontario

```
hive> select payment_type,max(cnt) from (select payment_type,count(*) as cnt fr
om sales where state='Ontario' group by payment_type) as t;
FAILED: SemanticException [Error 10025]: Line 1:7 Expression not in GROUP BY ke
y 'payment_type'
hive> (select payment_type,count(*) as cnt from sales where state='Ontario' gro
up by payment_type order by cnt LIMIT 1);
Query ID = hdoop_20220622152136_ddb069b4-c015-452f-8e8d-d4c80854f2d2
Total jobs = 2
Launching Job 1 out of 2
Number of reduce tasks not specified. Estimated from input data size: 1
In order to change the average load for a reducer (in bytes):
  set hive.exec.reducers.bytes.per.reducer=<number>
In order to limit the maximum number of reducers:
  set hive.exec.reducers.max=<number>
In order to set a constant number of reducers:
  set mapreduce.job.reduces=<number>
Starting Job = job_1655888018522_0011, Tracking URL = http://admin1-HP-280-G4-M
T-Business-PC:8088/proxy/application_1655888018522_0011/
Kill Command = /home/hdoop/hadoop-3.2.1/bin/mapred job  -kill job_1655888018522
_0011
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
```

```
In order to set a constant number of reducers:
  set mapreduce.job.reduces=<number>
Starting Job = job_1655888018522_0014, Tracking URL = http://admin1-HP-280-G4-M
T-Business-PC:8088/proxy/application_1655888018522_0014/
Kill Command = /home/hdoop/hadoop-3.2.1/bin/mapred job  -kill job_1655888018522
_0014
Hadoop job information for Stage-2: number of mappers: 1; number of reducers: 1
2022-06-22 15:22:51,659 Stage-2 map = 0%,   reduce = 0%
2022-06-22 15:22:55,789 Stage-2 map = 100%,   reduce = 0%, Cumulative CPU 1.06 s
ec
2022-06-22 15:22:59,900 Stage-2 map = 100%,   reduce = 100%, Cumulative CPU 2.34
 sec
MapReduce Total cumulative CPU time: 2 seconds 340 msec
Ended Job = job_1655888018522_0014
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1  Reduce: 1   Cumulative CPU: 3.52 sec   HDFS Read: 71343
HDFS Write: 196 SUCCESS
Stage-Stage-2: Map: 1  Reduce: 1   Cumulative CPU: 2.34 sec   HDFS Read: 7790 H
DFS Write: 107 SUCCESS
Total MapReduce CPU Time Spent: 5 seconds 860 msec
OK
Visa    20
Time taken: 36.4 seconds, Fetched: 1 row(s)
```

9. Find the number of transactions whose price is in between 1500-3600

```
hive> select count(*) from sales where price>1500 and price<3600;
Query ID = hdoop_20220622152647_38f55533-4ad2-472b-97c2-a98792740b23
Total jobs = 1
Launching Job 1 out of 1
Number of reduce tasks determined at compile time: 1
In order to change the average load for a reducer (in bytes):
  set hive.exec.reducers.bytes.per.reducer=<number>
In order to limit the maximum number of reducers:
  set hive.exec.reducers.max=<number>
In order to set a constant number of reducers:
  set mapreduce.job.reduces=<number>
Starting Job = job_1655888018522_0015, Tracking URL = http://admin1-HP-280-G4-M
T-Business-PC:8088/proxy/application_1655888018522_0015/
Kill Command = /home/hdoop/hadoop-3.2.1/bin/mapred job  -kill job_1655888018522
_0015
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2022-06-22 15:26:53,331 Stage-1 map = 0%,   reduce = 0%
```

```
In order to limit the maximum number of reducers:
  set hive.exec.reducers.max=<number>
In order to set a constant number of reducers:
  set mapreduce.job.reduces=<number>
Starting Job = job_1655888018522_0015, Tracking URL = http://admin1-HP-280-G4-M
T-Business-PC:8088/proxy/application_1655888018522_0015/
Kill Command = /home/hdoop/hadoop-3.2.1/bin/mapred job  -kill job_1655888018522
_0015
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2022-06-22 15:26:53,331 Stage-1 map = 0%,   reduce = 0%
2022-06-22 15:26:57,455 Stage-1 map = 100%,   reduce = 0%, Cumulative CPU 2.3 se
c
2022-06-22 15:27:03,601 Stage-1 map = 100%,   reduce = 100%, Cumulative CPU 4.02
 sec
MapReduce Total cumulative CPU time: 4 seconds 20 msec
Ended Job = job_1655888018522_0015
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1  Reduce: 1   Cumulative CPU: 4.02 sec   HDFS Read: 72450
HDFS Write: 101 SUCCESS
Total MapReduce CPU Time Spent: 4 seconds 20 msec
OK
2
Time taken: 17.39 seconds, Fetched: 1 row(s)
hive>
```

10. List all the transactions done in the United States using Mastercard

```
hive> select * from sales where country="United States" and payment_type="Maste
rcard";
OK
Product1        1200    Mastercard      Federica e Andrea       Astoria OR      U
nited States
Product1        1200    Mastercard      Fleur   Peoria  IL      United States
Product1        1200    Mastercard      adam    Martin  TN      United States
Product1        1200    Mastercard      Sean    Shavano Park    TX      United
States
Product1        1200    Mastercard      Katherine       New York        NY      U
nited States
Product1        1200    Mastercard      Linda   Miami   FL      United States
Product1        1200    Mastercard      Anneli  Houston TX      United States
Product1        1200    Mastercard      Marie   Ball Ground     GA      United
States
Product1        1200    Mastercard      Nikki   New Rochelle    NY      United
States
Product1        1200    Mastercard      June    Beachwood       OH      United
States
Product1        1200    Mastercard      chrissy W Lebanon       NH      United
States
Product2        3600    Mastercard      Valda   Irvine  CA      United States
```