Artificial Intelligence and Data Science Department

BDA / Odd Sem 2023-23 / Experiment 5

Name: Akshat Tiwari	Class/Roll No: D16AD /ʹ	Grade:

Program:

Download dataset from: https://archive.ics.uci.edu/ml/datasets/forest+fires

1. *Upload Dataset into Cloudera:*

```
[cloudera@quickstart ~]$ ls
cloudera-manager  doc.txt    enterprise-deployment.json  Music      Public
cm_api.py         Documents  express-deployment.json     name.txt   Templates
data.txt          Downloads  kerberos                    parcels    Videos
Desktop           eclipse    lib                         Pictures   workspace
[cloudera@quickstart ~]$ cd Desktop
[cloudera@quickstart Desktop]$ cd Data
[cloudera@quickstart Data]$ hdfs dfs -put forestfires.csv /user/cloudera
```

2. *Opening Hive Shell & Creating ForestFire Table:*

```
[cloudera@quickstart ~]$ sudo hive

Logging initialized using configuration in file:/etc/hive/conf.dist/hive-log4j.p
roperties
WARNING: Hive CLI is deprecated and migration to Beeline is recommended.
hive> Create External Table forestfire (X INT,Y INT, Month STRING, Day STRING, F
FMC FLOAT, DMC FLOAT, Dc FLOAT, ISI FLOAT, Temp FLOAT, RH INT, Wing FLOAT, Rain
FLOAT, AREA FLOAT)
    > Row FORMAT DELIMITED
    > FIELDS TERMINATED BY ',';
OK
Time taken: 2.376 seconds
```

3. *Loading Data from Dataset into ForestFire Table:*

```
    > LOAD DATA INPATH '/user/cloudera/forestfires.csv' OVERWRITE INTO TABLE for
estfire;
Loading data to table default.forestfire
chgrp: changing ownership of 'hdfs://quickstart.cloudera:8020/user/hive/warehous
e/forestfire/forestfires.csv': Permission denied. user=root is not the owner of
inode=forestfires.csv
chmod: changing permissions of 'hdfs://quickstart.cloudera:8020/user/hive/wareho
use/forestfire/forestfires.csv': Permission denied. user=root is not the owner o
f inode=forestfires.csv
Table default.forestfire stats: [numFiles=1, numRows=0, totalSize=25478, rawData
Size=0]
OK
Time taken: 0.537 seconds
```

4. *Executing Queries:*

Query 1 – select * from forestfire limit 10;

```
    > select * from forestfire limit 10;
OK
NULL    NULL    month   day     NULL    NULL    NULL    NULL    NULL    NULL    N
ULL     NULL    NULL
7       5       mar     fri     86.2    26.2    94.3    5.1     8.2     51      6
.7      0.0     0.0
7       4       oct     tue     90.6    35.4    669.1   6.7     18.0    33      0
.9      0.0     0.0
7       4       oct     sat     90.6    43.7    686.9   6.7     14.6    33      1
.3      0.0     0.0
8       6       mar     fri     91.7    33.3    77.5    9.0     8.3     97      4
.0      0.2     0.0
8       6       mar     sun     89.3    51.3    102.2   9.6     11.4    99      1
.8      0.0     0.0
8       6       aug     sun     92.3    85.3    488.0   14.7    22.2    29      5
.4      0.0     0.0
8       6       aug     mon     92.3    88.9    495.6   8.5     24.1    27      3
.1      0.0     0.0
8       6       aug     mon     91.5    145.4   608.2   10.7    8.0     86      2
.2      0.0     0.0
8       6       sep     tue     91.0    129.5   692.6   7.0     13.1    63      5
.4      0.0     0.0
Time taken: 0.294 seconds, Fetched: 10 row(s)
```

Query 2 – select * from forestfire where x = 7 and y = 4 limit 10;

```
    > select * from forestfire where X=7 and Y=4 limit 10;
OK
7      4      oct    tue    90.6   35.4    669.1   6.7    18.0   33    0.9    0.0    0.0
7      4      oct    sat    90.6   43.7    686.9   6.7    14.6   33    1.3    0.0    0.0
7      4      jun    sun    94.3   96.3    200.0   56.1   21.0   44    4.5    0.0    0.0
7      4      aug    sat    90.2   110.9   537.4   6.2    19.5   43    5.8    0.0    0.0
7      4      aug    sat    93.5   139.4   594.2   20.3   23.7   32    5.8    0.0    0.0
7      4      aug    sun    91.4   142.4   601.4   10.6   16.3   60    5.4    0.0    0.0
7      4      sep    fri    92.4   117.9   668.0   12.2   19.0   34    5.8    0.0    0.0
7      4      sep    mon    90.9   126.5   686.5   7.0    19.4   48    1.3    0.0    0.0
7      4      oct    fri    90.0   41.5    682.6   8.7    11.3   60    5.4    0.0    0.0
7      4      aug    sun    94.8   108.3   647.1   17.0   16.4   47    1.3    0.0    1.56
Time taken: 0.2 seconds, Fetched: 10 row(s)
```

Query 3 – select MONTH, avg (FFMC) as Average from forestfire group by MONTH;

```
apr     85.7888895670573
aug     92.33695594124173
dec     84.96666717529297
feb     82.90499916076661
jan     50.39999961853027
jul     91.32812428474426
jun     89.42941194422104
mar     89.44444345544886
may     87.3499984741211
month   NULL
nov     79.5
oct     90.45333251953124
sep     91.24302336227062
Time taken: 29.623 seconds, Fetched: 13 row(s)
```
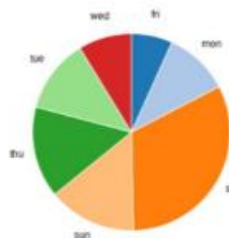
Query 4 – SELECT MONTH, MAX (RH) AS MAXIMUM FROM forestfire GROUP BY MONTH HAVING MONTH = 'sep';

```
OK
sep     86
Time taken: 26.654 seconds. Fetched: 1 row(s)
```

Query 5 **-** select DAY, SUM (AREA) AS AREA from forestfire group by DAY ORDER BY DAY;

```
day     NULL
fri     447.24000039696693
mon     706.5299995839596
sat     2144.8599796295166
sun     959.9299972057343
thu     997.1000298261642
tue     807.79000864923
wed     578.5999903082848
Time taken: 45.033 seconds, Fetched: 8 row(s)
```

<u>Query 6</u> **-** SELECT MONTH, MAX (DC) AS MAXIMUM FROM forestfire GROUP BY MONTH ORDER BY MONTH;

```
apr      97.1
aug      819.1
dec      354.6
feb      353.5
jan      171.4
jul      795.9
jun      433.3
mar      103.8
may      113.8
month    NULL
nov      106.7
oct      696.1
sep      860.6
Time taken: 50.182 seconds, Fetched: 13 row(s)
```