Name : Manav Pahilwani	Class/Roll No. : D16AD/ 37	Grade :

Title of Experiment : Create HIVE Database and Descriptive analytics-based statistics, visualization using Hive/PIG.

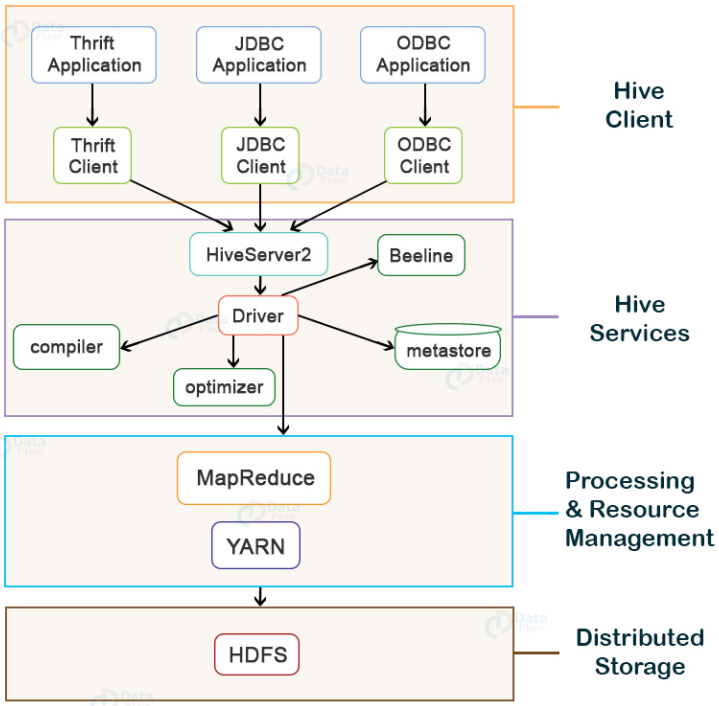
Theory:

Hive is a data warehouse infrastructure tool to process structured data in Hadoop. It resides on top of Hadoop to summarize Big Data, and makes querying and analyzing easy. Initially Hive was developed by Facebook, later the Apache Software Foundation took it up and developed it further as an open source under the name Apache Hive.

Hive is not
- A relational database
- A design for OnLine Transaction Processing (OLTP)
- A language for real-time queries and row-level updates

Features of Hive
- It stores schema in a database and processes data into HDFS.
- It is designed for OLAP.
- It provides SQL type language for querying called HiveQL or HQL.

Architecture of HIVE:



Hive Architecture & Its Components

Program:

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

Upload dataset to cloudera, local machine to hdfs.

```
[cloudera@quickstart ~]$ hdfs dfs -put /home/cloudera/Desktop/forestfires.csv /u
ser/cloudera
[cloudera@quickstart ~]$
```

```
hive> Create External Table forestfire(X int, Y int, Month String, Day String, FFMC Float, DMC Fl
oat, DC Float, ISI Float, Temp Float, RH int, Wind Float, Rain Float, Area Float)
    > Row Format Delimited
    > Fields Terminated by ',';
OK
Time taken: 0.734 seconds
```

```
hive> Load data inpath '/user/cloudera/forestfires.csv' overwrite into table forestfire;
Loading data to table default.forestfire
chgrp: changing ownership of 'hdfs://quickstart.cloudera:8020/user/hive/warehouse/forestfire/fore
stfires.csv': Permission denied. user=root is not the owner of inode=forestfires.csv
chmod: changing permissions of 'hdfs://quickstart.cloudera:8020/user/hive/warehouse/forestfire/fo
restfires.csv': Permission denied. user=root is not the owner of inode=forestfires.csv
Table default.forestfire stats: [numFiles=1, numRows=0, totalSize=25478, rawDataSize=0]
OK
Time taken: 1.532 seconds
hive>
```

Executing Queries:

1. Select * from firesfire limit 10;

```
hive> select * from forestfire limit 10;
OK
NULL    NULL    month   day     NULL    NULL    NULL    NULL    NULL    NULL    NULL    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: 0.222 seconds, Fetched: 10 row(s)
hive>
```

2. Select month and average of ffmc from forestfires table and group by month.

```
hive> select Month, avg(FFMC) as Average from forestfire group by Month;
Query ID = root_20231015064141_684648d2-9b5a-444d-9342-1e49663e3991
Total jobs = 1
Total MapReduce CPU Time Spent: 4 seconds 700 msec
OK
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: 64.686 seconds, Fetched: 13 row(s)
```

3. Maximum humidity in the month september

```
hive> select Month, max(RH) as maximum from forestfire group by Month having Month = 'sep';
Query ID = root_20231015064343_e85cda8a-d613-4096-b872-74ec300bca3d
Total jobs = 1
Total MapReduce CPU Time Spent: 5 seconds 470 msec
OK
sep     86
Time taken: 53.898 seconds, Fetched: 1 row(s)
```

4. Sum of area order by days of the week

```
hive> select Day, Sum(Area) as Area from forestfire group by Day order by day;
Query ID = root_20231015064747_29b68e1c-957f-448d-9c2c-df2f5d5fd7b0
Total jobs = 2
OK
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: 102.536 seconds, Fetched: 8 row(s)
hive>
```

5. Maximum of DC and order it according to the months

```
hive> select Month, max(DC) as maximum from forestfire group by Month order by Month;
Query ID = root_20231015065050_99228dcb-5ef9-4375-bc2f-e0c65389880a
Total jobs = 2
```

```
OK
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: 93.577 seconds, Fetched: 13 row(s)
```

6. List everything where x-coordinate = 7 and y - coordinate = 10.

```
hive> 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: 0.206 seconds, Fetched: 10 row(s)
```

Conclusion:

Successfully, created HIVE database and loaded forest fire data, and executed queries on the table.