**Scenario Based questions:**

**Will the reducer work or not if you use “Limit 1” in any HiveQL query?**
Ans. Yes, it will work.

**Suppose I have installed Apache Hive on top of my Hadoop cluster using default metastore configuration. Then, what will happen if we have multiple clients trying to access Hive at the same time?**
Ans. The Throughput of data will be less/slow, as the derby DB is not capable to handle multiple clients.

**Suppose, I create a table that contains details of all the transactions done by the customers: CREATE TABLE transaction_details (cust_id INT, amount FLOAT, month STRING, country STRING) ROW FORMAT DELIMITED FIELDS TERMINATED BY ‘,’ ;
Now, after inserting 50,000 records in this table, I want to know the total revenue generated for each month. But, Hive is taking too much time in processing this query. How will you solve this problem and list the steps that I will be taking in order to do so?**

Ans. You can increase the number of reducers by running below cmd.
set mapreduce.job.reduces=<number>

**How can you add a new partition for the month December in the above partitioned table?**
Ans. CREATE TABLE transaction_details_part ( cust_id INT, amount FLOAT) partition by (month string, Country string);

Insert overwrite transaction_details_part partition(month,country) select cust_id,amount,month,country from transaction_details;

**I am inserting data into a table based on partitions dynamically. But, I received an error – FAILED ERROR IN SEMANTIC ANALYSIS: Dynamic partition strict mode requires at least one static partition column. How will you remove this error?**

Ans. set hive.exec.dunamic.partition.mode = nonstrict;

**Suppose, I have a CSV file – ‘sample.csv’ present in ‘/temp’ directory with the following entries:**
**id first_name last_name email gender ip_address**
**How will you consume this CSV file into the Hive warehouse using built-in SerDe?**
Ans. create table sample_csv
( id int,
first_name string,
last_name string,

email string,
gender string,
Ip_address float)
row format delimited
fields terminated by ‘,’
location ‘/temp/’;

**Suppose, I have a lot of small CSV files present in the input directory in HDFS and I want to create a single Hive table corresponding to these files. The data in these files are in the format: {id, name, e-mail, country}. Now, as we know, Hadoop performance degrades when we use lots of small files.**
**So, how will you solve this problem where we want to create a single Hive table for lots of small files without degrading the performance of the system?**

Ans. We can use below query
Load data local inpath ‘hdfs:///<directory name>’ into table sample;

All the files in that folder will be loaded into the hive table without degrading the performance.

**LOAD DATA LOCAL INPATH ‘Home/country/state/’**
**OVERWRITE INTO TABLE address;**
**The following statement failed to execute. What can be the cause?**

Ans. The query should be Load data local inpath ‘file:///Home/country/state/’ into table address;
Cause: The URI is wrong it should start with file:// if file is in local storage. hdfs:/// if file in Hadoop file system.

**Is it possible to add 100 nodes when we already have 100 nodes in Hive? If yes, how?**

Ans. Yes, we can add a system where hadoop distribution is installed and then configure it as slave and connected it to the master node.

Hive Practical questions:

Hive Join operations

Create a  table named CUSTOMERS(ID | NAME | AGE | ADDRESS  | SALARY)

```
hive> create table CUSTOMERS
    > (
    > ID int,
    > NAME string,
    > AGE int,
    > ADDRESS string,
    > SALARY int);
OK
Time taken: 0.231 seconds
hive>
```

Create a Second  table ORDER(OID | DATE | CUSTOMER_ID | AMOUNT
)

```
hive> create table ORDER
    > (
    > OID int,
    > DATE string,
    > CUSTOMER_ID int,
    > AMOUNT float
    > );
OK
Time taken: 0.202 seconds
hive>
```

Now perform different joins operations on top of these tables
(Inner JOIN, LEFT OUTER JOIN ,RIGHT OUTER JOIN ,FULL OUTER JOIN)

## Inner join

```
hive> select c.*,o.* from customers c inner join orders o on c.id = o.customer_id;
Query ID = cloudera_20220916085858_351a6ec3-c385-4ce6-90c4-dd251b7348ee
Total jobs = 1
Execution log at: /tmp/cloudera/cloudera_20220916085858_351a6ec3-c385-4ce6-90c4-dd251b7348ee.log
2022-09-16 08:58:39     Starting to launch local task to process map join;      maximum memory = 932184064
2022-09-16 08:58:41     Dump the side-table for tag: 0 with group count: 4 into file: file:/tmp/cloudera/07f9fe2c-0ac1-4bf3-bac7-5cc8880f1db2/hive_2022-09-16_08-58-32_520_876221377078971338
9-1/-local-10003/HashTable-Stage-3/MapJoin-mapfile10--.hashtable
2022-09-16 08:58:41     Uploaded 1 File to: file:/tmp/cloudera/07f9fe2c-0ac1-4bf3-bac7-5cc8880f1db2/hive_2022-09-16_08-58-32_520_8762213770789713389-1/-local-10003/HashTable-Stage-3/MapJoin
-mapfile10--.hashtable (420 bytes)
2022-09-16 08:58:41     End of local task; Time Taken: 2.365 sec.
Execution completed successfully
MapredLocal task succeeded
Launching Job 1 out of 1
Number of reduce tasks is set to 0 since there's no reduce operator
Starting Job = job_1663318132024_0025, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1663318132024_0025/
Kill Command = /usr/lib/hadoop/bin/hadoop job  -kill job_1663318132024_0025
Hadoop job information for Stage-3: number of mappers: 1; number of reducers: 0
2022-09-16 08:58:55,962 Stage-3 map = 0%,  reduce = 0%
2022-09-16 08:59:06,148 Stage-3 map = 100%,  reduce = 0%, Cumulative CPU 3.03 sec
MapReduce Total cumulative CPU time: 3 seconds 30 msec
Ended Job = job_1663318132024_0025
MapReduce Jobs Launched:
Stage-Stage-3: Map: 1   Cumulative CPU: 3.03 sec   HDFS Read: 7719 HDFS Write: 402 SUCCESS
Total MapReduce CPU Time Spent: 3 seconds 30 msec
OK
c.id    c.name  c.age   c.address       c.salary        o.oid   o.date  o.customer_id   o.amount
4       'Raju'  26      'Star Garage'   150     1       01/09/2022      4       2000.0
2       'Adani' 65      'PM House'      NULL    2       03/09/2022      2       3000.0
3       'Ambani'        68      'Antilia'       200000000       3       05/09/2022      3       4000.0
1       'Elon'  55      'mars'  NULL    4       01/09/2022      1       10000.0
3       'Ambani'        68      'Antilia'       200000000       5       01/09/2022      3       1234.0
2       'Adani' 65      'PM House'      NULL    6       06/09/2022      2       2340.0
1       'Elon'  55      'mars'  NULL    7       03/09/2022      1       1200.0
4       'Raju'  26      'Star Garage'   150     8       05/09/2022      4       150.0
Time taken: 35.142 seconds, Fetched: 8 row(s)
hive>
```

## Left Outer Join

```
hive> select c.*,o.* from customers c left outer join orders o on c.id = o.customer_id;
Query ID = cloudera_20220916090202_855fd5de-05e9-493d-88da-00e7533d8d80
Total jobs = 1
Execution log at: /tmp/cloudera/cloudera_20220916090202_855fd5de-05e9-493d-88da-00e7533d8d80.log
2022-09-16 09:03:00     Starting to launch local task to process map join;      maximum memory = 932184064
2022-09-16 09:03:02     Dump the side-table for tag: 1 with group count: 6 into file: file:/tmp/cloudera/07f9fe2c-0ac1-4bf3-bac7-5cc8880f1db2/hive_2022-09-16_09-02-53_216_772096152287903376
5-1/-local-10003/HashTable-Stage-3/MapJoin-mapfile31--.hashtable
2022-09-16 09:03:02     Uploaded 1 File to: file:/tmp/cloudera/07f9fe2c-0ac1-4bf3-bac7-5cc8880f1db2/hive_2022-09-16_09-02-53_216_7720961522879033765-1/-local-10003/HashTable-Stage-3/MapJoin
-mapfile31--.hashtable (555 bytes)
2022-09-16 09:03:02     End of local task; Time Taken: 1.673 sec.
Execution completed successfully
MapredLocal task succeeded
Launching Job 1 out of 1
Number of reduce tasks is set to 0 since there's no reduce operator
Starting Job = job_1663318132024_0028, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1663318132024_0028/
Kill Command = /usr/lib/hadoop/bin/hadoop job  -kill job_1663318132024_0028
Hadoop job information for Stage-3: number of mappers: 1; number of reducers: 0
2022-09-16 09:03:16,552 Stage-3 map = 0%,  reduce = 0%
2022-09-16 09:03:25,514 Stage-3 map = 100%,  reduce = 0%, Cumulative CPU 2.24 sec
MapReduce Total cumulative CPU time: 2 seconds 240 msec
Ended Job = job_1663318132024_0028
MapReduce Jobs Launched:
Stage-Stage-3: Map: 1   Cumulative CPU: 2.24 sec   HDFS Read: 7574 HDFS Write: 436 SUCCESS
Total MapReduce CPU Time Spent: 2 seconds 240 msec
OK
c.id    c.name  c.age   c.address       c.salary        o.oid   o.date  o.customer_id   o.amount
NULL    name    NULL    address NULL    NULL    NULL    NULL    NULL
1       'Elon'  55      'mars'  NULL    4       01/09/2022      1       10000.0
1       'Elon'  55      'mars'  NULL    7       03/09/2022      1       1200.0
2       'Adani' 65      'PM House'      NULL    2       03/09/2022      2       3000.0
2       'Adani' 65      'PM House'      NULL    6       06/09/2022      2       2340.0
3       'Ambani'        68      'Antilia'       200000000       3       05/09/2022      3       4000.0
3       'Ambani'        68      'Antilia'       200000000       5       01/09/2022      3       1234.0
4       'Raju'  26      'Star Garage'   150     1       01/09/2022      4       2000.0
4       'Raju'  26      'Star Garage'   150     8       05/09/2022      4       150.0
Time taken: 33.529 seconds, Fetched: 9 row(s)
hive>
```

## Right Outer Join

```
hive> select c.*,o.* from customers c right outer join orders o on c.id = o.customer_id;
Query ID = cloudera_20220916090404_b5d4e7e4-72e7-4cfd-bafa-917430eb00ca
Total jobs = 1
Execution log at: /tmp/cloudera/cloudera_20220916090404_b5d4e7e4-72e7-4cfd-bafa-917430eb00ca.log
2022-09-16 09:04:20     Starting to launch local task to process map join;      maximum memory = 932184064
2022-09-16 09:04:22     Dump the side-table for tag: 0 with group count: 5 into file: file:/tmp/cloudera/07f9fe2c-0ac1-4bf3-bac7-5cc8880f1db2/hive_2022-09-16_09-04-12_626_1989648043542763019-1/-local-10003/HashTable-Stage-3/MapJoin-mapfile40--.hashtable
2022-09-16 09:04:22     Uploaded 1 File to: file:/tmp/cloudera/07f9fe2c-0ac1-4bf3-bac7-5cc8880f1db2/hive_2022-09-16_09-04-12_626_1989648043542763019-1/-local-10003/HashTable-Stage-3/MapJoin-mapfile40--.hashtable (451 bytes)
2022-09-16 09:04:22     End of local task; Time Taken: 1.848 sec.
Execution completed successfully
MapredLocal task succeeded
Launching Job 1 out of 1
Number of reduce tasks is set to 0 since there's no reduce operator
Starting Job = job_1663318132024_0029, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1663318132024_0029/
Kill Command = /usr/lib/hadoop/bin/hadoop job  -kill job_1663318132024_0029
Hadoop job information for Stage-3: number of mappers: 1; number of reducers: 0
2022-09-16 09:04:35,627 Stage-3 map = 0%,  reduce = 0%
2022-09-16 09:04:45,718 Stage-3 map = 100%,  reduce = 0%, Cumulative CPU 3.29 sec
MapReduce Total cumulative CPU time: 3 seconds 290 msec
Ended Job = job_1663318132024_0029
MapReduce Jobs Launched:
Stage-Stage-3: Map: 1   Cumulative CPU: 3.29 sec   HDFS Read: 7751 HDFS Write: 478 SUCCESS
Total MapReduce CPU Time Spent: 3 seconds 290 msec
OK
c.id    c.name  c.age   c.address       c.salary        o.oid   o.date  o.customer_id   o.amount
NULL    NULL    NULL    NULL    NULL    9       5.495109352676118E-5    8       2000.0
NULL    NULL    NULL    NULL    NULL    NULL    date    NULL    NULL
4       'Raju'  26      'Star Garage'   150     1       01/09/2022      4       2000.0
2       'Adani' 65      'PM House'      NULL    2       03/09/2022      2       3000.0
3       'Ambani'        68      'Antilia'       200000000       3       05/09/2022      3       4000.0
1       'Elon'  55      'mars'  NULL    4       01/09/2022      1       10000.0
3       'Ambani'        68      'Antilia'       200000000       5       01/09/2022      3       1234.0
2       'Adani' 65      'PM House'      NULL    6       06/09/2022      2       2340.0
1       'Elon'  55      'mars'  NULL    7       03/09/2022      1       1200.0
4       'Raju'  26      'Star Garage'   150     8       05/09/2022      4       150.0
Time taken: 34.326 seconds, Fetched: 10 row(s)
hive>
```

## Full Outer Join

```
hive> select c.*,o.* from customers c full outer join orders o on c.id = o.customer_id;
Query ID = cloudera_20220916090505_35d36577-50aa-4d19-bdb9-d7fd7a3850b0
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_1663318132024_0030, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1663318132024_0030/
Kill Command = /usr/lib/hadoop/bin/hadoop job  -kill job_1663318132024_0030
Hadoop job information for Stage-1: number of mappers: 2; number of reducers: 1
2022-09-16 09:05:39,114 Stage-1 map = 0%,  reduce = 0%
2022-09-16 09:05:51,293 Stage-1 map = 50%,  reduce = 0%, Cumulative CPU 2.83 sec
2022-09-16 09:05:53,770 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 5.26 sec
2022-09-16 09:06:01,517 Stage-1 map = 100%,  reduce = 100%, Cumulative CPU 8.42 sec
MapReduce Total cumulative CPU time: 8 seconds 420 msec
Ended Job = job_1663318132024_0030
MapReduce Jobs Launched:
Stage-Stage-1: Map: 2  Reduce: 1   Cumulative CPU: 8.42 sec   HDFS Read: 15969 HDFS Write: 512 SUCCESS
Total MapReduce CPU Time Spent: 8 seconds 420 msec
OK
c.id    c.name  c.age   c.address       c.salary        o.oid   o.date  o.customer_id   o.amount
NULL    name    NULL    address NULL    NULL    NULL    NULL
NULL    NULL    NULL    NULL    NULL    NULL    date    NULL    NULL
1       'Elon'  55      'mars'  NULL    7       03/09/2022      1       1200.0
1       'Elon'  55      'mars'  NULL    4       01/09/2022      1       10000.0
2       'Adani' 65      'PM House'      NULL    6       06/09/2022      2       2340.0
2       'Adani' 65      'PM House'      NULL    2       03/09/2022      2       3000.0
3       'Ambani'        68      'Antilia'       200000000       5       01/09/2022      3       1234.0
3       'Ambani'        68      'Antilia'       200000000       3       05/09/2022      3       4000.0
4       'Raju'  26      'Star Garage'   150     8       05/09/2022      4       150.0
4       'Raju'  26      'Star Garage'   150     1       01/09/2022      4       2000.0
NULL    NULL    NULL    NULL    NULL    9       5.495109352676118E-5    8       2000.0
Time taken: 35.844 seconds, Fetched: 11 row(s)
hive>
```

BUILD A DATA PIPELINE WITH HIVE

Download a data from the given location - https://archive.ics.uci.edu/ml/machine-learning-databases/00360/

### 1. Create a hive table as per given schema in your dataset.

```
hive> create table airquality
    > (
    > `Date` string ,
    > `Time` string,
    > `CO(GT)` string,
    > `PT08S1(CO)` int,
    > `NMHC(GT)` int ,
    > `C6H6(GT)` string,
    > `PT08S2(NMHC)` int,
    > `NOx(GT)` int,
    > `PT08S3(NOx)` int,
    > `NO2(GT)` int,
    > `PT08S4(NO2)` int,
    > `PT08S5(O3)` int,
    > `T` string,
    > `RH` string,
    > `AH` string
    > )
    > row format delimited
    > fields terminated by '\;'
    > tblproperties("skip.header.line.count" = "1");
OK
Time taken: 0.092 seconds
```

**2. try to place a data into table location.**

```
hive> load data local inpath 'file:///home/cloudera/Hive_class/AirQualityUCI.csv' into table airquality;
Loading data to table hive_class_b1.airquality
Table hive_class_b1.airquality stats: [numFiles=1, totalSize=785065]
OK
Time taken: 0.846 seconds
```

**3. Perform a select operation .**

```
hive> select * from airquality limit 5;
OK
10/03/2004      18.00.00        2,6     1360    150     11,9    1046    166     1056    113     1692    1268    13,6    48,9    0,7578
10/03/2004      19.00.00        2       1292    112     9,4     955     103     1174    92      1559    972     13,3    47,7    0,7255
10/03/2004      20.00.00        2,2     1402    88      9,0     939     131     1140    114     1555    1074    11,9    54,0    0,7502
10/03/2004      21.00.00        2,2     1376    80      9,2     948     172     1092    122     1584    1203    11,0    60,0    0,7867
10/03/2004      22.00.00        1,6     1272    51      6,5     836     131     1205    116     1490    1110    11,2    59,6    0,7888
Time taken: 0.073 seconds, Fetched: 5 row(s)
hive>
```

**4. Fetch the result of the select operation in your local as a csv file .**

```
hive> INSERT OVERWRITE DIRECTORY '/home/cloudera/Hive_class/export' ROW FORMAT DELIMITED FIELDS TERMINATED BY ',' SELECT * FROM airquality;
Query ID = cloudera_20220916030404_fa256e89-4a6f-4284-aeb8-337c0cb8b4f4
Total jobs = 3
Launching Job 1 out of 3
Number of reduce tasks is set to 0 since there's no reduce operator
Starting Job = job_1663318132024_0001, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1663318132024_0001/
Kill Command = /usr/lib/hadoop/bin/hadoop job  -kill job_1663318132024_0001
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 0
2022-09-16 03:04:29,469 Stage-1 map = 0%,  reduce = 0%
2022-09-16 03:04:38,032 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 2.1 sec
MapReduce Total cumulative CPU time: 2 seconds 100 msec
Ended Job = job_1663318132024_0001
Stage-3 is selected by condition resolver.
Stage-2 is filtered out by condition resolver.
Stage-4 is filtered out by condition resolver.
Moving data to: hdfs://quickstart.cloudera:8020/home/cloudera/Hive_class/export/.hive-staging_hive_2022-09-16_03-04-16_996_5178907981834524114-1/-ext-10000
Moving data to: /home/cloudera/Hive_class/export
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1   Cumulative CPU: 2.1 sec   HDFS Read: 790068 HDFS Write: 758350 SUCCESS
Total MapReduce CPU Time Spent: 2 seconds 100 msec
OK
Time taken: 22.217 seconds
hive>
```

## 5. Perform group by operation .

```
hive> select date , sum(T) as T_per_Year from airquality group by date;
Query ID = cloudera_20220916031010_96ac251d-7df4-4292-b50e-40b0b1577a6e
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_1663318132024_0003, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1663318132024_0003/
Kill Command = /usr/lib/hadoop/bin/hadoop job  -kill job_1663318132024_0003
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2022-09-16 03:11:02,277 Stage-1 map = 0%,  reduce = 0%
2022-09-16 03:11:08,547 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 1.64 sec
2022-09-16 03:11:16,094 Stage-1 map = 100%,  reduce = 100%, Cumulative CPU 3.46 sec
MapReduce Total cumulative CPU time: 3 seconds 460 msec
Ended Job = job_1663318132024_0003
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1  Reduce: 1   Cumulative CPU: 3.46 sec   HDFS Read: 794540 HDFS Write: 5986 SUCCESS
Total MapReduce CPU Time Spent: 3 seconds 460 msec
OK
date    t_per_year
        0.0
01/01/2005      0.0
01/02/2005      0.0
01/03/2005      0.0
01/04/2004      -600.0
01/04/2005      0.0
01/05/2004      0.0
01/06/2004      0.0
01/07/2004      0.0
01/08/2004      0.0
01/09/2004      0.0
01/10/2004      -200.0
01/11/2004      0.0
01/12/2004      0.0
02/01/2005      -600.0
02/02/2005      0.0
02/03/2005      0.0
02/04/2004      0.0
02/04/2005      0.0
02/05/2004      0.0
```

## 7. Perform filter operation at least 5 kinds of filter examples .

select * from airquality where Date like '%2004' limit 5;

```
hive> select * from airquality where Date like '%2004' limit 5;
OK
airquality.date airquality.time airquality.co(gt)       airquality.pt08s1(co)   airquality.nmhc(gt)     airquality.c6h6(gt)     airquality.pt08s2(nmhc) airqua
lity.nox(gt)    airquality.pt08s3(nox)  airquality.no2(gt)      airquality.pt08s4(no2)  airquality.pt08s5(o3)   airquality.t    airquality.rh   airquality.ah
10/03/2004      18.00.00        2,6     1360    150     11,9    1046    166     1056    113     1692    1268    13,6    48,9    0,7578
10/03/2004      19.00.00        2       1292    112     9,4     955     103     1174    92      1559    972     13,3    47,7    0,7255
10/03/2004      20.00.00        2,2     1402    88      9,0     939     131     1140    114     1555    1074    11,9    54,0    0,7502
10/03/2004      21.00.00        2,2     1376    80      9,2     948     172     1092    122     1584    1203    11,0    60,0    0,7867
10/03/2004      22.00.00        1,6     1272    51      6,5     836     131     1205    116     1490    1110    11,2    59,6    0,7888
Time taken: 0.134 seconds, Fetched: 5 row(s)
hive>
```

select * from airquality where Date like '%11%' limit 5;

```
hive>  select * from airquality where Date like '%11%' limit 5;
OK
airquality.date airquality.time airquality.co(gt)       airquality.pt08s1(co)   airquality.nmhc(gt)     airquality.c6h6(gt)     airquality.pt08s2(nmhc) airqua
lity.nox(gt)    airquality.pt08s3(nox)  airquality.no2(gt)      airquality.pt08s4(no2)  airquality.pt08s5(o3)   airquality.t    airquality.rh   airquality.ah
11/03/2004      00.00.00        1,2     1185    31      3,6     690     62      1462    77      1333    733     11,3    56,8    0,7603
11/03/2004      01.00.00        1       1136    31      3,3     672     62      1453    76      1333    730     10,7    60,0    0,7702
11/03/2004      02.00.00        0,9     1094    24      2,3     609     45      1579    60      1276    620     10,7    59,7    0,7648
11/03/2004      03.00.00        0,6     1010    19      1,7     561     -200    1705    -200    1235    501     10,3    60,2    0,7517
11/03/2004      04.00.00        -200    1011    14      1,3     527     21      1818    34      1197    445     10,1    60,5    0,7465
Time taken: 0.095 seconds, Fetched: 5 row(s)
hive>
```

select * from airquality where `pt08s1(co)` >1000  limit 5;

```
hive> select * from airquality where `pt08s1(co)` >1000  limit 5;
OK
airquality.date airquality.time airquality.co(gt)       airquality.pt08s1(co)   airquality.nmhc(gt)     airquality.c6h6(gt)     airquality.pt08s2(nmhc) airqua
lity.nox(gt)    airquality.pt08s3(nox)  airquality.no2(gt)      airquality.pt08s4(no2)  airquality.pt08s5(o3)   airquality.t    airquality.rh   airquality.ah
10/03/2004      18.00.00        2,6     1360    150     11,9    1046    166     1056    113     1692    1268    13,6    48,9    0,7578
10/03/2004      19.00.00        2       1292    112     9,4     955     103     1174    92      1559    972     13,3    47,7    0,7255
10/03/2004      20.00.00        2,2     1402    88      9,0     939     131     1140    114     1555    1074    11,9    54,0    0,7502
10/03/2004      21.00.00        2,2     1376    80      9,2     948     172     1092    122     1584    1203    11,0    60,0    0,7867
10/03/2004      22.00.00        1,6     1272    51      6,5     836     131     1205    116     1490    1110    11,2    59,6    0,7888
Time taken: 0.089 seconds, Fetched: 5 row(s)
hive>
```

select * from airquality where `pt08s5(o3)` between 1000 and 1500  limit 5;

```
hive> select * from airquality where `pt08s5(o3)` between 1000 and 1500  limit 5;
OK
airquality.date airquality.time airquality.co(gt)       airquality.pt08s1(co)   airquality.nmhc(gt)     airquality.c6h6(gt)     airquality.pt08s2(nmhc) airqua
lity.nox(gt)    airquality.pt08s3(nox)  airquality.no2(gt)      airquality.pt08s4(no2)  airquality.pt08s5(o3)   airquality.t    airquality.rh   airquality.ah
10/03/2004      18.00.00        2,6     1360    150     11,9    1046    166     1056    113     1692    1268    13,6    48,9    0,7578
10/03/2004      20.00.00        2,2     1402    88      9,0     939     131     1140    114     1555    1074    11,9    54,0    0,7502
10/03/2004      21.00.00        2,2     1376    80      9,2     948     172     1092    122     1584    1203    11,0    60,0    0,7867
10/03/2004      22.00.00        1,6     1272    51      6,5     836     131     1205    116     1490    1110    11,2    59,6    0,7888
11/03/2004      08.00.00        2       1333    64      8,0     900     174     1136    112     1517    1102    10,8    57,4    0,7408
Time taken: 0.087 seconds, Fetched: 5 row(s)
hive>
```

select * from airquality where `pt08s5(o3)` > 1000 and Date like '%2004' limit 5;

```
hive> select * from airquality where `pt08s5(o3)` > 1000 and Date like '%2004' limit 5;
OK
airquality.date airquality.time airquality.co(gt)       airquality.pt08s1(co)   airquality.nmhc(gt)     airquality.c6h6(gt)     airquality.pt08s2(nmhc) airqua
lity.nox(gt)    airquality.pt08s3(nox)  airquality.no2(gt)      airquality.pt08s4(no2)  airquality.pt08s5(o3)   airquality.t    airquality.rh   airquality.ah
10/03/2004      18.00.00        2,6     1360    150     11,9    1046    166     1056    113     1692    1268    13,6    48,9    0,7578
10/03/2004      20.00.00        2,2     1402    88      9,0     939     131     1140    114     1555    1074    11,9    54,0    0,7502
10/03/2004      21.00.00        2,2     1376    80      9,2     948     172     1092    122     1584    1203    11,0    60,0    0,7867
10/03/2004      22.00.00        1,6     1272    51      6,5     836     131     1205    116     1490    1110    11,2    59,6    0,7888
11/03/2004      08.00.00        2       1333    64      8,0     900     174     1136    112     1517    1102    10,8    57,4    0,7408
Time taken: 0.222 seconds, Fetched: 5 row(s)
hive>
```

## 8. show and example of regex operation

```
hive> select * from airquality where date REGEXP '2004$' limit 5;
OK
airquality.date airquality.time airquality.co(gt)       airquality.pt08s1(co)   airquality.nmhc(gt)     airquality.c6h6(gt)     airquality.pt08s2(nmhc) airquality.nox(gt)      airquality.pt
08s3(nox)       airquality.no2(gt)      airquality.pt08s4(no2)  airquality.pt08s5(o3)   airquality.t    airquality.rh   airquality.ah
10/03/2004      18.00.00        2,6     1360    150     11,9    1046    166     1056    113     1692    1268    13,6    48,9    0,7578
10/03/2004      19.00.00        2       1292    112     9,4     955     103     1174    92      1559    972     13,3    47,7    0,7255
10/03/2004      20.00.00        2,2     1402    88      9,0     939     131     1140    114     1555    1074    11,9    54,0    0,7502
10/03/2004      21.00.00        2,2     1376    80      9,2     948     172     1092    122     1584    1203    11,0    60,0    0,7867
10/03/2004      22.00.00        1,6     1272    51      6,5     836     131     1205    116     1490    1110    11,2    59,6    0,7888
Time taken: 0.238 seconds, Fetched: 5 row(s)
hive>
```

## 9. alter table operation

```
hive> describe airquality;
OK
col_name        data_type       comment
date                    string
time                    string
co(gt)                  string
pt08s1(co)              int
nmhc(gt)                int
c6h6(gt)                string
pt08s2(nmhc)            int
nox(gt)                 int
pt08s3(nox)             int
no2(gt)                 int
pt08s4(no2)             int
pt08s5(o3)              int
t                       string
rh                      string
ah                      string
Time taken: 0.119 seconds, Fetched: 15 row(s)
hive> alter table airquality change column date Date_New date;
OK
Time taken: 0.408 seconds
hive> describe airquality;
OK
col_name        data_type       comment
date_new                date
time                    string
co(gt)                  string
pt08s1(co)              int
nmhc(gt)                int
c6h6(gt)                string
pt08s2(nmhc)            int
nox(gt)                 int
pt08s3(nox)             int
no2(gt)                 int
pt08s4(no2)             int
pt08s5(o3)              int
t                       string
rh                      string
ah                      string
Time taken: 0.14 seconds, Fetched: 15 row(s)
hive>
```

**10 . drop table operation**

```
hive> show tables;
OK
tab_name
airquality
airquality_ordered
ordered_airquality
sales_order_csv
sales_order_orc
sorted_airquality
Time taken: 0.03 seconds, Fetched: 6 row(s)
hive> drop table airquality_ordered;
OK
Time taken: 0.194 seconds
hive> show tables;
OK
tab_name
airquality
ordered_airquality
sales_order_csv
sales_order_orc
sorted_airquality
Time taken: 0.024 seconds, Fetched: 5 row(s)
hive>
```

**2 . order by operation .**

```
hive> select * from airquality order by `pt08s5(o3)` desc  limit 5;
Query ID = cloudera_20220916045959_d3e6b300-e866-46f6-9867-2702e791c7b2
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_1663318132024_0010, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1663318132024_0010/
Kill Command = /usr/lib/hadoop/bin/hadoop job  -kill job_1663318132024_0010
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2022-09-16 04:59:29,843 Stage-1 map = 0%,  reduce = 0%
2022-09-16 04:59:40,277 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 3.28 sec
2022-09-16 04:59:50,516 Stage-1 map = 100%,  reduce = 100%, Cumulative CPU 6.52 sec
MapReduce Total cumulative CPU time: 6 seconds 520 msec
Ended Job = job_1663318132024_0010
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1  Reduce: 1   Cumulative CPU: 6.52 sec   HDFS Read: 797109 HDFS Write: 415 SUCCESS
Total MapReduce CPU Time Spent: 6 seconds 520 msec
OK
airquality.date airquality.time airquality.co(gt)       airquality.pt08s1(co)   airquality.nmhc(gt)     airquality.c6h6(gt)     airquality.pt08s2(nmhc) airqua
lity.nox(gt)    airquality.pt08s3(nox)  airquality.no2(gt)      airquality.pt08s4(no2)  airquality.pt08s5(o3)   airquality.t    airquality.rh   airquality.ah
13/12/2004      18.00.00        9,9     1881    -200    50,8    1983    1479    334     269     2271    2523    12,6    55,9    0,8142
22/11/2004      20.00.00        6,5     1706    -200    33,9    1641    1218    381     239     1778    2522    8,9     54,2    0,6183
01/11/2004      01.00.00        3,7     1407    -200    17,8    1235    338     548     77      1780    2519    20,1    71,1    1,6498
22/11/2004      19.00.00        6,9     1766    -200    37,6    1720    1345    361     245     1881    2515    9,1     53,8    0,6216
12/02/2005      23.00.00        4,5     1738    -200    20,4    1309    743     424     231     1543    2494    10,1    59,9    0,7379
Time taken: 36.238 seconds, Fetched: 5 row(s)
hive>
```

## 13 . where clause operations you have to perform .

```
hive> select * from airquality where `pt08s5(o3)` < 1000 limit 5;
OK
airquality.date airquality.time airquality.co(gt)       airquality.pt08s1(co)   airquality.nmhc(gt)     airquality.c6h6(gt)     airquality.pt08s2(nmhc) airqua
lity.nox(gt)    airquality.pt08s3(nox)  airquality.no2(gt)      airquality.pt08s4(no2)  airquality.pt08s5(o3)   airquality.t    airquality.rh   airquality.ah
10/03/2004      19.00.00        2       1292    112     9,4     955     103     1174    92      1559    972     13,3    47,7    0,7255
10/03/2004      23.00.00        1,2     1197    38      4,7     750     89      1337    96      1393    949     11,2    59,2    0,7848
11/03/2004      00.00.00        1,2     1185    31      3,6     690     62      1462    77      1333    733     11,3    56,8    0,7603
11/03/2004      01.00.00        1       1136    31      3,3     672     62      1453    76      1333    730     10,7    60,0    0,7702
11/03/2004      02.00.00        0,9     1094    24      2,3     609     45      1579    60      1276    620     10,7    59,7    0,7648
Time taken: 0.253 seconds, Fetched: 5 row(s)
hive>
```

## 14 . sorting operation you have to perform .

```
hive> select * from airquality sort by `pt08s5(o3)` desc  limit 5;
Query ID = cloudera_20220916050101_c6cfe60f-d84f-4aaa-94ce-a2f7e0d6a290
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_1663318132024_0011, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1663318132024_0011/
Kill Command = /usr/lib/hadoop/bin/hadoop job  -kill job_1663318132024_0011
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2022-09-16 05:01:17,989 Stage-1 map = 0%,  reduce = 0%
2022-09-16 05:01:29,556 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 3.77 sec
2022-09-16 05:01:39,674 Stage-1 map = 100%,  reduce = 100%, Cumulative CPU 7.2 sec
MapReduce Total cumulative CPU time: 7 seconds 200 msec
Ended Job = job_1663318132024_0011
Launching Job 2 out of 2
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_1663318132024_0012, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1663318132024_0012/
Kill Command = /usr/lib/hadoop/bin/hadoop job  -kill job_1663318132024_0012
Hadoop job information for Stage-2: number of mappers: 1; number of reducers: 1
2022-09-16 05:01:56,385 Stage-2 map = 0%,  reduce = 0%
2022-09-16 05:02:05,124 Stage-2 map = 100%,  reduce = 0%, Cumulative CPU 1.87 sec
2022-09-16 05:02:16,077 Stage-2 map = 100%,  reduce = 100%, Cumulative CPU 5.09 sec
MapReduce Total cumulative CPU time: 5 seconds 90 msec
Ended Job = job_1663318132024_0012
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1  Reduce: 1   Cumulative CPU: 7.2 sec   HDFS Read: 795995 HDFS Write: 524 SUCCESS
Stage-Stage-2: Map: 1  Reduce: 1   Cumulative CPU: 5.09 sec   HDFS Read: 9149 HDFS Write: 415 SUCCESS
Total MapReduce CPU Time Spent: 12 seconds 290 msec
OK
airquality.date airquality.time airquality.co(gt)       airquality.pt08s1(co)   airquality.nmhc(gt)     airquality.c6h6(gt)     airquality.pt08s2(nmhc) airqua
lity.nox(gt)    airquality.pt08s3(nox)  airquality.no2(gt)      airquality.pt08s4(no2)  airquality.pt08s5(o3)   airquality.t    airquality.rh   airquality.ah
13/12/2004      18.00.00        9,9     1881    -200    50,8    1983    1479    334     269     2271    2523    12,6    55,9    0,8142
22/11/2004      20.00.00        6,5     1706    -200    33,9    1641    1218    381     239     1778    2522    8,9     54,2    0,6183
01/11/2004      01.00.00        3,7     1407    -200    17,8    1235    338     548     77      1780    2519    20,1    71,1    1,6498
22/11/2004      19.00.00        6,9     1766    -200    37,6    1720    1345    361     245     1881    2515    9,1     53,8    0,6216
12/02/2005      23.00.00        4,5     1738    -200    20,4    1309    743     424     231     1543    2494    10,1    59,9    0,7379
Time taken: 70.813 seconds, Fetched: 5 row(s)
hive>
```

## 15 . distinct operation you have to perform .

```
hive> select distinct date from airquality limit 5;
Query ID = cloudera_20220916045353_82850268-b11c-4fd7-b669-2f34978d23c3
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_1663318132024_0005, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1663318132024_0005/
Kill Command = /usr/lib/hadoop/bin/hadoop job  -kill job_1663318132024_0005
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2022-09-16 04:53:34,079 Stage-1 map = 0%,  reduce = 0%
2022-09-16 04:53:43,186 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 2.33 sec
2022-09-16 04:53:54,293 Stage-1 map = 100%,  reduce = 100%, Cumulative CPU 4.83 sec
MapReduce Total cumulative CPU time: 4 seconds 830 msec
Ended Job = job_1663318132024_0005
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1  Reduce: 1   Cumulative CPU: 4.83 sec   HDFS Read: 793681 HDFS Write: 45 SUCCESS
Total MapReduce CPU Time Spent: 4 seconds 830 msec
OK
date

01/01/2005
01/02/2005
01/03/2005
01/04/2004
Time taken: 35.291 seconds, Fetched: 5 row(s)
hive>
```

## 16 . like an operation you have to perform .

```
hive> select * from airquality where Date like '%2004' limit 5;
OK
airquality.date airquality.time airquality.co(gt)       airquality.pt08s1(co)   airquality.nmhc(gt)     airquality.c6h6(gt)     airquality.pt08s2(nmhc) airqua
lity.nox(gt)    airquality.pt08s3(nox)  airquality.no2(gt)      airquality.pt08s4(no2)  airquality.pt08s5(o3)   airquality.t    airquality.rh   airquality.ah
10/03/2004      18.00.00        2,6     1360    150     11,9    1046    166     1056    113     1692    1268    13,6    48,9    0,7578
10/03/2004      19.00.00        2       1292    112     9,4     955     103     1174    92      1559    972     13,3    47,7    0,7255
10/03/2004      20.00.00        2,2     1402    88      9,0     939     131     1140    114     1555    1074    11,9    54,0    0,7502
10/03/2004      21.00.00        2,2     1376    80      9,2     948     172     1092    122     1584    1203    11,0    60,0    0,7867
10/03/2004      22.00.00        1,6     1272    51      6,5     836     131     1205    116     1490    1110    11,2    59,6    0,7888
Time taken: 0.186 seconds, Fetched: 5 row(s)
hive>
```

## 17 . union operation you have to perform .

```
hive> select * from airquality limit 5
    > union all
    > select * from airquality_ordered limit 5;
Query ID = cloudera_20220916052121_470d722c-77e3-48a7-a5cd-04d1b39320b6
Total jobs = 3
Launching Job 1 out of 3
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_1663318132024_0018, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1663318132024_0018/
Kill Command = /usr/lib/hadoop/bin/hadoop job  -kill job_1663318132024_0018
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2022-09-16 05:22:08,143 Stage-1 map = 0%,  reduce = 0%
2022-09-16 05:22:17,060 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 1.98 sec
2022-09-16 05:22:27,362 Stage-1 map = 100%,  reduce = 100%, Cumulative CPU 4.58 sec
MapReduce Total cumulative CPU time: 4 seconds 580 msec
Ended Job = job_1663318132024_0018
Launching Job 2 out of 3
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_1663318132024_0019, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1663318132024_0019/
Kill Command = /usr/lib/hadoop/bin/hadoop job  -kill job_1663318132024_0019
Hadoop job information for Stage-3: number of mappers: 1; number of reducers: 1
2022-09-16 05:22:42,847 Stage-3 map = 0%,  reduce = 0%
2022-09-16 05:22:51,978 Stage-3 map = 100%,  reduce = 0%, Cumulative CPU 2.26 sec
2022-09-16 05:23:02,323 Stage-3 map = 100%,  reduce = 100%, Cumulative CPU 4.89 sec
MapReduce Total cumulative CPU time: 4 seconds 890 msec
Ended Job = job_1663318132024_0019
Launching Job 3 out of 3
Number of reduce tasks is set to 0 since there's no reduce operator
Starting Job = job_1663318132024_0020, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1663318132024_0020/
Kill Command = /usr/lib/hadoop/bin/hadoop job  -kill job_1663318132024_0020
Hadoop job information for Stage-2: number of mappers: 2; number of reducers: 0
2022-09-16 05:23:19,077 Stage-2 map = 0%,  reduce = 0%
2022-09-16 05:23:30,150 Stage-2 map = 100%,  reduce = 0%, Cumulative CPU 4.04 sec
MapReduce Total cumulative CPU time: 4 seconds 40 msec
Ended Job = job_1663318132024_0020
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1  Reduce: 1   Cumulative CPU: 4.58 sec   HDFS Read: 15128 HDFS Write: 475 SUCCESS
Stage-Stage-3: Map: 1  Reduce: 1   Cumulative CPU: 4.89 sec   HDFS Read: 11535 HDFS Write: 524 SUCCESS
Stage-Stage-2: Map: 2   Cumulative CPU: 4.04 sec   HDFS Read: 9765 HDFS Write: 793 SUCCESS
```

```
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1  Reduce: 1   Cumulative CPU: 4.58 sec   HDFS Read: 15128 HDFS Write: 475 SUCCESS
Stage-Stage-3: Map: 1  Reduce: 1   Cumulative CPU: 4.89 sec   HDFS Read: 11535 HDFS Write: 524 SUCCESS
Stage-Stage-2: Map: 2   Cumulative CPU: 4.04 sec   HDFS Read: 9765 HDFS Write: 793 SUCCESS
Total MapReduce CPU Time Spent: 13 seconds 510 msec
OK
_u1.date        _u1.time        _u1.co(gt)      _u1.pt08s1(co)  _u1.nmhc(gt)    _u1.c6h6(gt)    _u1.pt08s2(nmhc)        _u1.nox(gt)     _u1.pt08s3(nox) _u1.no2(gt)     _u1.pt08s4(no2) _u1.p
t08s5(o3)       _u1.t   _u1.rh  _u1.ah
12/02/2005      23.00.00        4,5     1738    -200    20,4    1309    743     424     231     1543    2494    10,1    59,9    0,7379
22/11/2004      19.00.00        6,9     1766    -200    37,6    1720    1345    361     245     1881    2515    9,1     53,8    0,6216
01/11/2004      01.00.00        3,7     1407    -200    17,8    1235    338     548     77      1780    2519    20,1    71,1    1,6498
22/11/2004      20.00.00        6,5     1706    -200    33,9    1641    1218    381     239     1778    2522    8,9     54,2    0,6183
13/12/2004      18.00.00        9,9     1881    -200    50,8    1983    1479    334     269     2271    2523    12,6    55,9    0,8142
10/03/2004      21.00.00        2,2     1376    80      9,2     948     172     1092    122     1584    1203    11,0    60,0    0,7867
10/03/2004      20.00.00        2,2     1402    88      9,0     939     131     1140    114     1555    1074    11,9    54,0    0,7502
10/03/2004      19.00.00        2       1292    112     9,4     955     103     1174    92      1559    972     13,3    47,7    0,7255
10/03/2004      18.00.00        2,6     1360    150     11,9    1046    166     1056    113     1692    1268    13,6    48,9    0,7578
Date    Time    CO(GT)  NULL    NULL    C6H6(GT)        NULL    NULL    NULL    NULL    NULL    NULL    T       RH      AH
Time taken: 96.577 seconds, Fetched: 10 row(s)
hive>
```

## 18 . table view operation you have to perform .

```
hive> create view ordered_airquality as
    > select * from airquality order by `pt08s5(o3)` desc limit 5;
OK
date    time    co(gt)  pt08s1(co)      nmhc(gt)        c6h6(gt)        pt08s2(nmhc)    nox(gt) pt08s3(nox)     no2(gt) pt08s4(no2)     pt08s5(o3)      t       rh      ah
Time taken: 0.255 seconds
hive> select * from ordered_airquality;
Query ID = cloudera_20220916050909_2aa8748a-4762-46bd-a927-ba3ab0a7022f
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_1663318132024_0015, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1663318132024_0015/
Kill Command = /usr/lib/hadoop/bin/hadoop job  -kill job_1663318132024_0015
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2022-09-16 05:10:05,157 Stage-1 map = 0%,  reduce = 0%
2022-09-16 05:10:15,281 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 3.48 sec
2022-09-16 05:10:30,782 Stage-1 map = 100%,  reduce = 100%, Cumulative CPU 9.92 sec
MapReduce Total cumulative CPU time: 9 seconds 920 msec
Ended Job = job_1663318132024_0015
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1  Reduce: 1   Cumulative CPU: 9.92 sec   HDFS Read: 797614 HDFS Write: 415 SUCCESS
Total MapReduce CPU Time Spent: 9 seconds 920 msec
OK
ordered_airquality.date ordered_airquality.time ordered_airquality.co(gt)       ordered_airquality.pt08s1(co)   ordered_airquality.nmhc(gt)     ordered_airquality.c6h6(gt)     ordered_airqu
ality.pt08s2(nmhc)      ordered_airquality.nox(gt)      ordered_airquality.pt08s3(nox)  ordered_airquality.no2(gt)      ordered_airquality.pt08s4(no2)  ordered_airquality.pt08s5(o3)   order
ed_airquality.t ordered_airquality.rh   ordered_airquality.ah
13/12/2004      18.00.00        9,9     1881    -200    50,8    1983    1479    334     269     2271    2523    12,6    55,9    0,8142
22/11/2004      20.00.00        6,5     1706    -200    33,9    1641    1218    381     239     1778    2522    8,9     54,2    0,6183
01/11/2004      01.00.00        3,7     1407    -200    17,8    1235    338     548     77      1780    2519    20,1    71,1    1,6498
22/11/2004      19.00.00        6,9     1766    -200    37,6    1720    1345    361     245     1881    2515    9,1     53,8    0,6216
12/02/2005      23.00.00        4,5     1738    -200    20,4    1309    743     424     231     1543    2494    10,1    59,9    0,7379
Time taken: 42.695 seconds, Fetched: 5 row(s)
hive>
```