

EXTERNAL TABLE:

```
[hduser@localhost ~]$ hdfs dfs -mkdir /user/hduser/hivedata2
22/07/01 10:06:26 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
[hduser@localhost ~]$ hdfs dfs -put hivefile2.csv /user/hduser/hivedata2
22/07/01 10:06:35 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
[hduser@localhost ~]$ hdfs dfs -cat hivefile2.csv /user/hduser/hivedata2
22/07/01 10:06:47 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
cat: 'hivefile2.csv': No such file or directory
cat: '/user/hduser/hivedata2': Is a directory
[hduser@localhost ~]$ hdfs dfs -ls /user/hduser/hivedata2
22/07/01 10:07:20 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
Found 1 items
-rw-r--r--  1 hduser supergroup          73 2022-07-01 10:06 /user/hduser/hivedata2/hivefile2.csv
[hduser@localhost ~]$
```

```
hive> show tables;
OK
customers
orders
Time taken: 0.545 seconds, Fetched: 2 row(s)
hive> select * from customers;
OK
1111    Raju    WA
2222    jack   WA
3333    john   CA
4444    Nina   NY
Time taken: 1.687 seconds, Fetched: 4 row(s)
hive> create external table products(id string,title string,cost float)location '/user/hduser/hivedata2';
OK
Time taken: 2.2 seconds
hive> select * from hivedata2;
FAILED: SemanticException [Error 10001]: Line 1:14 Table not found 'hivedata2'
hive> select * from products;
OK
1111,mobile,455.5      NULL  NULL
2222,laptop,30000      NULL  NULL
3333,earphones,5000    NULL  NULL
4444,camera,4000       NULL  NULL
Time taken: 0.519 seconds, Fetched: 4 row(s)
hive>
```

```
hive> drop table products;
OK
Time taken: 2.272 seconds
hive>
```

```
[hduser@localhost ~]$ hdfs dfs -ls /user/hduser/hivedata2
22/07/01 10:12:50 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
Found 1 items
-rw-r--r--  1 hduser supergroup          73 2022-07-01 10:06 /user/hduser/hivedata2/hivefile2.csv
[hduser@localhost ~]$
```

```
hive> create external table if not exists products(id string,title string,cost float) row format delimited fields terminated by ',' stored as textfile location '/user/hduser/hivedata2';
OK
Time taken: 0.346 seconds
hive> select * from products;
OK
1111    mobile  455.5
2222    laptop  30000.0
3333    earphones 5000.0
4444    camera  4000.0
Time taken: 0.266 seconds, Fetched: 4 row(s)
hive>
```

```
hive> describe formatted products;
OK
# col_name           data_type           comment
id                   string
title               string
cost                float

# Detailed Table Information
Database:            hive1
Owner:               hduser
CreateTime:          Fri Jul 01 10:17:12 IST 2022
LastAccessTime:      UNKNOWN
Protect Mode:        None
Retention:           0
Location:             hdfs://localhost:54310/user/hduser/hivedata2
Table Type:          EXTERNAL_TABLE
Table Parameters:
    COLUMN_STATS_ACCURATE    false
    EXTERNAL                  TRUE
    numFiles                  0
    numRows                   -1
    rawDataSize               -1
    totalSize                 0
    transient_lastDdlTime     1656650832
```

Subqueries in hive:

```
[hduser@localhost ~]$ gedit freshproducts.csv
^C
[hduser@localhost ~]$ hdfs dfs -ls /user/hduser/hivedata2
22/07/01 10:24:36 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
Found 1 items
-rw-r--r--  1 hduser supergroup          73 2022-07-01 10:06 /user/hduser/hivedata2/hivefile2.csv
[hduser@localhost ~]$ hdfs dfs -put freshproducts.csv /user/hduser/hivedata2
22/07/01 10:25:21 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
[hduser@localhost ~]$ hdfs dfs -ls /user/hduser/hivedata2
22/07/01 10:25:33 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
Found 2 items
-rw-r--r--  1 hduser supergroup          60 2022-07-01 10:25 /user/hduser/hivedata2/freshproducts.csv
-rw-r--r--  1 hduser supergroup          73 2022-07-01 10:06 /user/hduser/hivedata2/hivefile2.csv
```

```
hive> create table freshproducts(id string,title string,cost float) row format delimited fields terminated by ',' stored as textfile;
```

```
hive> load data local inpath '/home/hduser/freshproducts.csv' into table freshproducts;
Loading data to table hive1.freshproducts
Table hive1.freshproducts stats: [numFiles=1, totalSize=60]
OK
Time taken: 1.404 seconds
hive> select * from freshproducts;
OK
5555    desktop 70000.0
6666    washing machine 50000.0
7777    ac      80000.0
Time taken: 0.281 seconds, Fetched: 3 row(s)
hive> █
```

```
hive> select * from(select id as product_id from products UNION ALL select id as product_id from freshproducts) t;
Query ID = hduser_20220701103513_f164e109-7926-40c2-aa3d-60cc1317979f
Total jobs = 1
Launching Job 1 out of 1
Number of reduce tasks is set to 0 since there's no reduce operator
Starting Job = job_1656649354731_0001, Tracking URL = http://Inceptez:8088/proxy/application_1656649354731_0001/
Kill Command = /usr/local/hadoop/bin/hadoop job -kill job_1656649354731_0001
Hadoop job information for Stage-1: number of mappers: 2; number of reducers: 0
2022-07-01 10:36:00,550 Stage-1 map = 0%, reduce = 0%
2022-07-01 10:36:22,011 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 3.54 sec
MapReduce Total cumulative CPU time: 3 seconds 540 msec
Ended Job = job_1656649354731_0001
MapReduce Jobs Launched:
Stage-Stage-1: Map: 2 Cumulative CPU: 3.54 sec HDFS Read: 9097 HDFS Write: 50 SUCCESS
Total MapReduce CPU Time Spent: 3 seconds 540 msec
OK
5555
6666
7777
1111
2222
3333
4444
5555
```

```
hive> select distinct (t.product_id) from (select product_id from customers join orders where customers.id=orders.customer_id) t;
Query ID = hduser_20220701104340_420c4e33-907e-42ec-a5fd-533be403803d
Total jobs = 1
SLF4J: Class path contains multiple SLF4J bindings.
SLF4J: Found binding in [jar:file:/usr/local/hadoop/share/hadoop/common/lib/slf4j-log4j12-1.7.10.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: Found binding in [jar:file:/usr/local/hbase/lib/phoenix-4.11.0-HBase-0.98-client.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: Found binding in [jar:file:/usr/local/hbase/lib/slf4j-log4j12-1.6.4.jar!/org/slf4j/impl/StaticLoggerBinder.class]
```

```
broom
cam
phone
t-shirt
Time taken: 74.849 seconds, Fetched: 4 row(s)
hive>
```

```
hive> select * from customers;
OK
1111 Raju WA
2222 jack WA
3333 john CA
4444 Nina NY
Time taken: 0.206 seconds, Fetched: 4 row(s)
hive> show tables;
OK
customers
freshproducts
orders
products
Time taken: 0.305 seconds, Fetched: 4 row(s)
hive> select * from orders
> ;
OK
111111 phone 1111 3 1200.0
111112 cam 1111 1 5200.0
111113 broom 2222 2 20.0
111114 broom 2222 2 20.0
111115 t-shirt 4444 2 66.0
Time taken: 0.227 seconds, Fetched: 5 row(s)
```

```
hive> desc orders;
OK
id bigint
product_id string
customer_id bigint
quantity int
amount double
Time taken: 0.329 seconds, Fetched: 5 row(s)
hive> desc customers;
OK
id bigint
name string
address string
Time taken: 0.348 seconds, Fetched: 3 row(s)
hive>
```

WHERE clause:

1)IN/NOT IN

2)EXISTS/NOT EXISTS

```
hive> select name from customers where customers.id IN (select customer_id from orders);
Query ID = hduser_20220701105820_0ebf4827-80e6-4016-afe7-0c97a4c32c37
Total jobs = 1
SLF4J: Class path contains multiple SLF4J bindings.
SLF4J: Found binding in [jar:file:/usr/local/hadoop/share/hadoop/common/lib/slf4j-log4j12-1.7.10.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: Found binding in [jar:file:/usr/local/hbase/lib/phoenix-4.11.0-HBase-0.98-client.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: Found binding in [jar:file:/usr/local/hbase/lib/slf4j-log4j12-1.6.4.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: See http://www.slf4j.org/codes.html#multiple_bindings for an explanation.
SLF4J: Actual binding is of type [org.slf4j.impl.Log4jLoggerFactory]
22/07/01 10:58:30 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
22/07/01 10:58:31 WARN conf.HiveConf: HiveConf of name hive.mapjoin.optimized.keys does not exist
22/07/01 10:58:31 WARN conf.HiveConf: HiveConf of name hive.optimize.multigroupby.common.distincts does not exist
22/07/01 10:58:31 WARN conf.HiveConf: HiveConf of name hive.mapjoin.lazy.hashtable does not exist
22/07/01 10:58:31 WARN conf.HiveConf: HiveConf of name hive.server2.thrift.http.min.worker.threads does not exist
22/07/01 10:58:31 WARN conf.HiveConf: HiveConf of name hive.server2.thrift.http.max.worker.threads does not exist
22/07/01 10:58:31 WARN conf.HiveConf: HiveConf of name hive.server2.logging.operation.verbose does not exist
Execution log at: /tmp/hduser/hduser_20220701105820_0ebf4827-80e6-4016-afe7-0c97a4c32c37.log
2022-07-01 10:58:32 Starting to launch local task to process map join; maximum memory = 518979584
2022-07-01 10:58:35 Dump the side-table for tag: 1 with group count: 3 into file: file:/tmp/hiveLog/625e219d-0bc1-495a-aa41-e60c90982e8a/hive_2022-07-01_10-58-20_629_7928132592865178676-1/-local-10002/HashTable-Stage-3/MapJoin-mapfile11--.hashtable
2022-07-01 10:58:35 Uploaded 1 File to: file:/tmp/hiveLog/625e219d-0bc1-495a-aa41-e60c90982e8a/hive_2022-07-01_10-58-20_629_7928132592865178676-1/-local-10002/HashTable-Stage-3/MapJoin-mapfile11--.hashtable (320 bytes)
2022-07-01 10:58:35 End of local task; Time Taken: 3.031 sec.
2022-07-01 10:58:35 Dump the side-table for tag: 1 with group count: 3 into file: file:/tmp/hiveLog/625e219d-0bc1-495a-aa41-e60c90982e8a/hive_2022-07-01_10-58-20_629_7928132592865178676-1/-local-10002/HashTable-Stage-3/MapJoin-mapfile11--.hashtable
2022-07-01 10:58:35 Uploaded 1 File to: file:/tmp/hiveLog/625e219d-0bc1-495a-aa41-e60c90982e8a/hive_2022-07-01_10-58-20_629_7928132592865178676-1/-local-10002/HashTable-Stage-3/MapJoin-mapfile11--.hashtable (320 bytes)
2022-07-01 10:58:35 End of local task; Time Taken: 3.031 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_1656649354731_0003, Tracking URL = http://Inceptez:8088/proxy/application_1656649354731_0003/
Kill Command = /usr/local/hadoop/bin/hadoop job -kill job_1656649354731_0003
Hadoop job information for Stage-3: number of mappers: 1; number of reducers: 0
2022-07-01 10:59:13,294 Stage-3 map = 0%, reduce = 0%
2022-07-01 10:59:28,012 Stage-3 map = 100%, reduce = 0%, Cumulative CPU 3.08 sec
MapReduce Total cumulative CPU time: 3 seconds 80 msec
Ended Job = job_1656649354731_0003
MapReduce Jobs Launched:
Stage-Stage-3: Map: 1 Cumulative CPU: 3.08 sec HDFS Read: 6457 HDFS Write: 15 SUCCESS
Total MapReduce CPU Time Spent: 3 seconds 80 msec
OK
Raju
jack
Nina
Time taken: 69.052 seconds, Fetched: 3 row(s)
hive>
```

```
hive> select name from customers where customers.id NOT IN (select customer_id from orders);
john
Time taken: 190.95 seconds, Fetched: 1 row(s)
hive>
```

```
hive> select id from customers where exists (select customer_id from orders where orders.customer_id=customers.id);
OK
1111
2222
4444
Time taken: 55.954 seconds, Fetched: 3 row(s)
```

```
hive> select id from customers where not exists (select customer_id from orders where orders.customer_id=customers.id);
Query ID = hduser_20220701111011_da3adeab-a2b3-480b-929c-c4e6b88f0ce6
Total jobs = 1
SLF4J: Class path contains multiple SLF4J bindings.
SLF4J: Found binding in [jar:file:/usr/local/hadoop/share/hadoop/common/lib/slf4j-log4j12-1.7.10.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: Found binding in [jar:file:/usr/local/hbase/lib/phoenix-4.11.0-HBase-0.98-client.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: Found binding in [jar:file:/usr/local/hbase/lib/slf4j-log4j12-1.6.4.jar!/org/slf4j/impl/StaticLoggerBinder.class]
3333
Time taken: 44.411 seconds, Fetched: 1 row(s)
hive>
```

Views in hive:

```
hive> create view customer_purchase as select customer_id,product_id,adress from customers join orders where customers.id=orders.customer_id;
OK
```

```
hive> create table author_details(author_name string,book_names array<string>) row format delimited fields terminated by ',' collection items terminated by ":";
OK
```

Explode and Lateral views:

```
hive> load data local inpath '/home/hduser/books.csv' into table author_details;
Loading data to table hive1.author_details
Table hive1.author_details stats: [numFiles=1, totalSize=89]
OK
Time taken: 1.178 seconds
hive> select * from author_details;
OK
author1 ["book1","book7","book8"]
author1 ["book2","book9"]
author2 ["book3","book10","book11"]
author2 ["book4"]
NULL
Time taken: 0.22 seconds, Fetched: 5 row(s)
hive>
```

```
hive> select explode(book_names) from author_details;
OK
book1
book7
book8
book2
book9
book3
book10
book11
book4
Time taken: 0.211 seconds, Fetched: 9 row(s)
hive> select author_name,b_name from author_details lateral view explode(book_names) book_table as b_name;
OK
author1 book1
author1 book7
author1 book8
author1 book2
author1 book9
author2 book3
author2 book10
author2 book11
author2 book4
Time taken: 0.21 seconds, Fetched: 9 row(s)
```

COMPLEX DATA TYPES:

```
hive> create table mobilephones(id string,title string,cost float,colors array<string>,screen_size array<float>);
OK
Time taken: 0.509 seconds
hive> insert into table mobilephones
> select "redminote7","Redmi Note 7",300,array("white","silver","black"),array(float(4.5))
> union all
> select "motoGplus","Moto G Plus",200,array("black","gold"),array(float(4.5),float(5.5));
```

```
hive> select id,colors from mobilephones;
OK
redminote7      ["white","silver","black"]
motoGplus      ["black","gold"]
Time taken: 0.483 seconds, Fetched: 2 row(s)
hive> select id,colors[0] from mobilephones;
OK
redminote7      white
motoGplus       black
Time taken: 0.363 seconds, Fetched: 2 row(s)
hive>
```

```
hive> select features['camera'] from mobilePhones_new1;
OK
true
true
Time taken: 0.325 seconds, Fetched: 2 row(s)
```

mobilephones.csv

```
samsungj7,Samsung J7,250.0,red#blue#black,5.5,camera:true#dualsim:false,battery:24hrs#camera:2MP
oneplusthree,One Plus Three,450.0,gold#silver,4.5#5.5,autofocus:true,battery:24hrs#camera:null
```

```
hive> create table mobiles(id string,title string,cost float,colors array<string>,screen_size array<float>,features map<string,boolean>,
information struct<battery:string,camera:string>) row format delimited fields terminated by ',' collection items terminated by '#' map key
s terminated by ':';
OK
Time taken: 0.288 seconds
hive> load data local inpath '/home/hduser/mobilephones.csv' into table mobiles;
Loading data to table hive1.mobiles
Table hive1.mobiles stats: [numFiles=1, totalSize=192]
OK
Time taken: 0.813 seconds
hive> select * from mobiles;
OK
samsungj7      Samsung J7      250.0   ["red","blue","black"]  [5.5]   {"camera":true,"dualsim":false} {"battery":"battery:24hrs","came
a":"camera:2MP"}
oneplusthree   One Plus Three  450.0   ["gold","silver"]      [4.5,5.5] {"autofocus":true}      {"battery":"battery:24hrs","came
a":"camera:null"}
Time taken: 0.202 seconds, Fetched: 2 row(s)
hive>
```