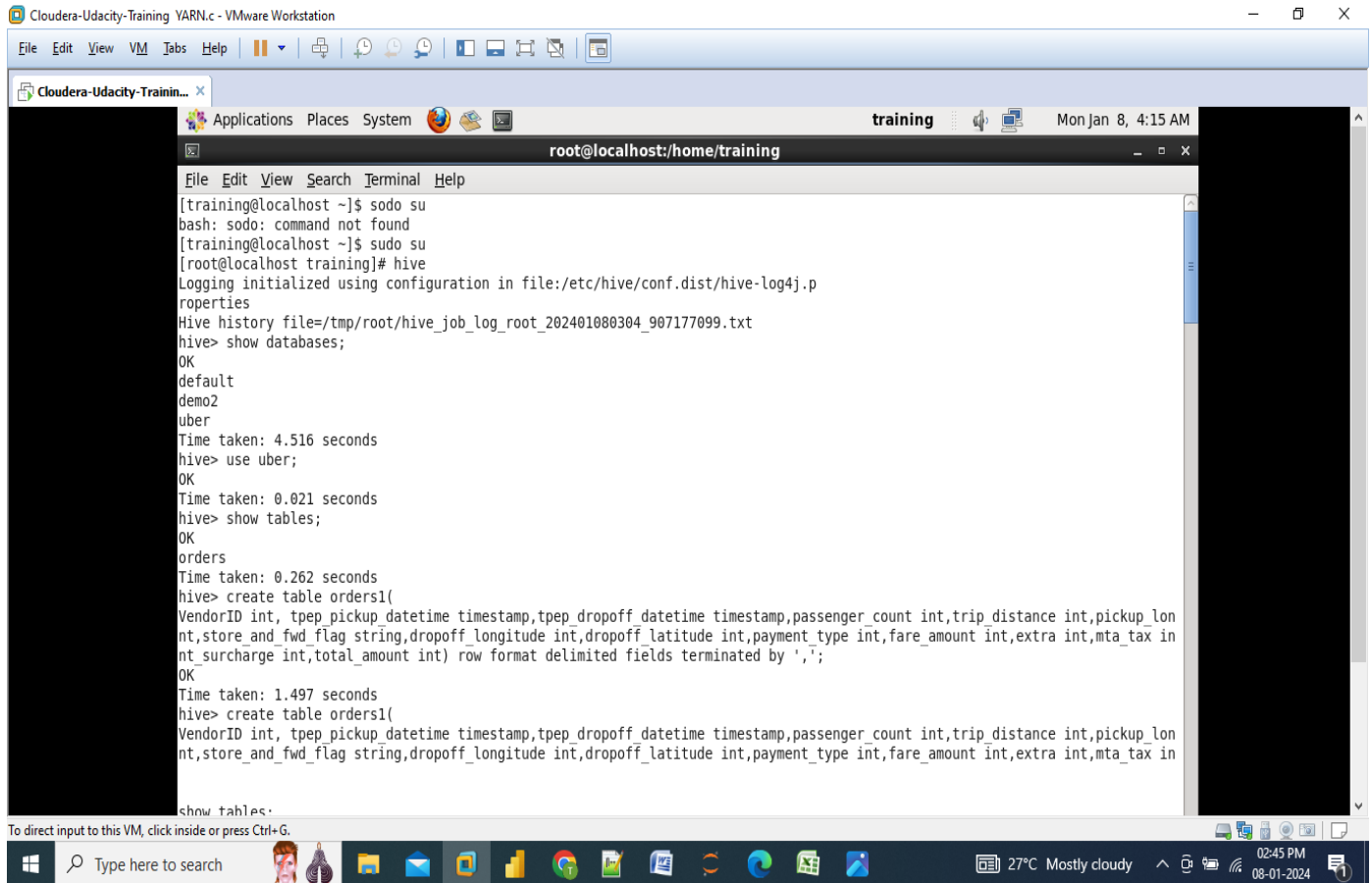


## Assignment No: 5

**Q) Create internal and external table on your project data. Also create static and dynamic partitioning.**

**) Internal Table and External Table:**



The screenshot shows a VMware Workstation window titled "Cloudera-Udacity-Training YARN.c - VMware Workstation". Inside the VM, a terminal window titled "root@localhost:/home/training" is open. The terminal shows the following commands and output:

```
[training@localhost ~]$ sodo su
bash: sodo: command not found
[training@localhost ~]$ sudo su
[root@localhost training]# hive
Logging initialized using configuration in file:/etc/hive/conf.dist/hive-log4j.p
roperties
Hive history file=/tmp/root/hive_job_log_root_202401080304_907177099.txt
hive> show databases;
OK
default
demo2
uber
Time taken: 4.516 seconds
hive> use uber;
OK
Time taken: 0.021 seconds
hive> show tables;
OK
orders
Time taken: 0.262 seconds
hive> create table orders1(
VendorID int, tpep pickup datetime timestamp,tpep dropoff datetime timestamp,passenger count int,trip distance int,pickup_lon
nt,store_and_fwd_flag string,dropoff_longitude int,dropoff_latitude int,payment_type int,fare_amount int,extra int,mta_tax in
nt,surcharge int,total_amount int) row format delimited fields terminated by ',';
OK
Time taken: 1.497 seconds
hive> create table orders1(
VendorID int, tpep pickup datetime timestamp,tpep dropoff datetime timestamp,passenger count int,trip distance int,pickup_lon
nt,store_and_fwd_flag string,dropoff_longitude int,dropoff_latitude int,payment_type int,fare_amount int,extra int,mta_tax in
nt,surcharge int,total_amount int) row format delimited fields terminated by ',';
OK
Time taken: 1.497 seconds
hive> show tables;
```

The terminal window is part of a desktop environment with a taskbar at the bottom showing various application icons and system status (27°C, Mostly cloudy, 02:45 PM, 08-01-2024).

```
Application Cloudera-Udacity-Train... 8, 3:18 AM
root@localhost:/home/training
File Edit View Search Terminal Help
fare_amount int None
extra int None
mta_tax int None
tip_amount int None
tolls_amount int None
improvement_surcharge int None
total_amount int None

# Detailed Table Information
Database: uber
Owner: root
CreateTime: Mon Jan 08 03:11:53 EST 2024
LastAccessTime: UNKNOWN
Protect Mode: None
Retention: 0
Location: hdfs://localhost:8020/user/hive/warehouse/uber.db/orders1
Table Type: MANAGED_TABLE
Table Parameters:
    transient_lastDdlTime 1704701513

# Storage Information
SerDe Library: org.apache.hadoop.hive.serde2.lazy.LazySimpleSerDe
InputFormat: org.apache.hadoop.mapred.TextInputFormat
OutputFormat: org.apache.hadoop.hive.ql.io.HiveIgnoreKeyTextOutputFormat
Compressed: No
Num Buckets: -1
Bucket Columns: []
Sort Columns: []
Storage Desc Params:
    field.delim ,
    serialization.format ,
Time taken: 0.282 seconds
hive> load data local inpath '/uber_data.csv' into table orders1;
Copying data from file:/uber_data.csv
Copying file: file:/uber_data.csv
Loading data to table uber.orders1
OK
Time taken: 1.413 seconds
hive>
```

## 2) Static Partition:

```
Application Cloudera-Udacity-Train... 8, 4:11 AM
root@localhost:/home/training
File Edit View Search Terminal Help
# Storage Information
SerDe Library: org.apache.hadoop.hive.serde2.lazy.LazySimpleSerDe
InputFormat: org.apache.hadoop.mapred.TextInputFormat
OutputFormat: org.apache.hadoop.hive.ql.io.HiveIgnoreKeyTextOutputFormat
Compressed: No
Num Buckets: -1
Bucket Columns: []
Sort Columns: []
Storage Desc Params:
    field.delim ,
    serialization.format ,
Time taken: 0.282 seconds
hive> load data local inpath '/uber_data.csv' into table orders1;
Copying data from file:/uber_data.csv
Copying file: file:/uber_data.csv
Loading data to table uber.orders1
OK
Time taken: 1.413 seconds
hive> create table sat_part(vaid int,VID int, t_p_d timestamp,t_d timestamp,td int,pl int,pla int,RID)partitioned by (pc int)row format delimited fields terminated by ',';
FAILED: ParseException line 1:17 mismatched input 'is' expecting identifier near ',' in column specification

hive> create table sat_part(vaid int,VID int, t_p_d timestamp,t_d timestamp,td int,pl int,pla int,RID int,store string,dl int, dla int,pt int,fa int,extra int,mta int,ta
int,toa int,ins int,total int)partitioned by (pc int)row format delimited fields terminated by ',';
FAILED: SemanticException [Error 10036]: Duplicate column name: vaid
hive> create table sat_part(vaid int,VID int, t_p_d timestamp,t_d timestamp,td int,pl int,pla int,RID int,store string,dl int, dla int,pt int,fa int,extra int,mta int,ta
oa int,ins int,total int)partitioned by (pc int)row format delimited fields terminated by ',';
FAILED: SemanticException [Error 10036]: Duplicate column name: vaid
hive> [root@localhost training]# hive
Logging initialized using configuration in file:/etc/hive/conf.dist/hive-log4j.properties
Hive history file=/tmp/root/hive_job_log_root_202401080344_1656563169.txt
hive> use uber;
OK
Time taken: 3.781 seconds
hive> create table sat_part(vaid int, t_p_d timestamp,t_d timestamp,td int,pl int,pla int,RID int,store string,dl int, dla int,pt int,fa int,extra int,mta int,taoa int,i
ns int,total int)partitioned by (pc int)row format delimited fields terminated by ',';
OK
Time taken: 1.58 seconds
hive> create table sat_part(vaid int, t_p_d timestamp,t_d timestamp,td int,pl int,pla int,RID int,store string,dl int, dla int,pt int,fa int,extra int,mta int,taoa int,i
```

Cloudera-Udacity-Training YARN - VMware Workstation

training Mon Jan 8, 4:10 AM

HDFS:/user/hive/warehouse/uber.db - Mozilla Firefox

Contents of directory /user/hive/warehouse/uber.db

Go to: /user/hive/warehouse/uber.db go

[Go to parent directory](#)

Name	Type	Size	Replication	Block Size	Modification Time	Permission	Owner	Group
dai_part	dir				2024-01-08 03:48	rwxf-rf-x	root	supergroup
orders	dir				2024-01-05 00:45	rwxf-rf-x	root	supergroup
orders1	dir				2024-01-08 03:14	rwxf-rf-x	root	supergroup
sat_part	dir				2024-01-08 03:46	rwxf-rf-x	root	supergroup

[Go back to DFS home](#)

Local logs

[Log](#) directory

To direct input to this VM, click inside or press Ctrl+G.

### 3) Dynamic Partition:

root@localhost:/home/training

```
OK
Time taken: 1.58 seconds
hive> create table sat_part(vid int, t_p_d timestamp,t_d timestamp,td int,pl int,pla int,RID int,store string,dl int, dla int,pt int,fa int,extra int,mta int,taoa int,ins int,total int)partitioned by (pc int)row format delimited fields terminated by ',';
FAILED: Error in metadata: AlreadyExistsException(message:Table sat_part already exists)
FAILED: Execution Error, return code 1 from org.apache.hadoop.hive.ql.exec.DDLTask
hive> create table dai_part(vid int, t_p_d timestamp,t_d timestamp,td int,pl int,pla int,RID int,store string,dl int, dla int,pt int,fa int,extra int,mta int,taoa int,ins int,total int)partitioned by (pc int)row format delimited fields terminated by ',';
OK
Time taken: 0.133 seconds
hive> set hive.exec.dynamic.partition=true;
hive> set hive.exec.dynamic.partition.mode=nonstrict;
hive> insert into table dai_part partition(pc int) select create table dai_part(vid int, t_p_d timestamp,t_d timestamp,td int,pl int,pla int,RID int,store string,dl int, dla int,pt int,fa int,extra int,mta int,taoa int,ins int,total int)partitioned by (pc int)row format delimited fields terminated by ',';
FAILED: ParseException line 1:40 mismatched input 'int' expecting ')' near 'pc' in insert clause

hive> insert into table dai_part partition(pc int) select create table dai_part(vid int, t_p_d timestamp,t_d timestamp,td int,pl int,pla int,RID int,store string,dl int, dla int,pt int,fa int,extra int,mta int,taoa int,ins int,total int)partitioned by (pc int)row format delimited fields terminated by ',';
FAILED: ParseException line 1:40 mismatched input 'int' expecting ')' near 'pc' in insert clause

hive> insert into table dai_part partition(pc int) select vid int, t_p_d timestamp,t_d timestamp,td int,pl int,pla int,RID int,store string,dl int, dla int,pt int,fa int,extra int,mta int,taoa int,ins int,total int from orders1;
FAILED: ParseException line 1:29 cannot recognize input near 'partiion' '(' 'pc' in select clause

hive> show tables;
OK
dai_part
orders
orders1
sat_part
Time taken: 0.365 seconds
hive> insert into table dai_part partition(pc int) select vid int, t_p_d timestamp,t_d timestamp,td int,pl int,pla int,RID int,store string,dl int, dla int,pt int,fa int,extra int,mta int,taoa int,ins int,total int from orders1;
FAILED: ParseException line 1:42 mismatched input 'int' expecting ')' near 'pc' in insert clause

hive> insert into table dai_part partition(pc int) select vid int, t_p_d timestamp,t_d timestamp,td int,pl int,pla int,RID int,store string,dl int, dla int,pt int,fa int,extra int,mta int,taoa int,ins int,pc int,total int from orders1;
```

```
Applications | File Edit View VM Tabs Help | root@localhost/home/training | Cloudera-Udacity-Training ... | shubham | Thu Jan 4, 2:10 PM

File Edit View Search Terminal Help

open string None
high string None
low string None
close string None
a_close string None
volume int None

# Partition Information
# col_name data_type comment
datee string None

# Detailed Table Information
Database: stock_market_partition
Owner: root
CreateTime: Thu Jan 04 13:31:41 EST 2024
LastAccessTime: UNKNOWN
Protect Mode: None
Retention: 0
Location: hdfs://localhost:8020/user/hive/warehouse/stock_market_partition.db/danamic_table
Table Type: MANAGED_TABLE
Table Parameters:
  transient_lastDdlTime 1704393101

# Storage Information
SerDe Library: org.apache.hadoop.hive.serde2.LazySimpleSerDe
InputFormat: org.apache.hadoop.mapred.TextInputFormat
OutputFormat: org.apache.hadoop.hive ql.io.HiveIgnoreKeyTextOutputFormat
Compressed: No
Num Buckets: -1
Bucket Columns: []
Sort Columns: []
Storage Desc Params:
  field.delim
  serialization.format

Time taken: 0.296 seconds
hive> [root@localhost training]# hive
Logging initialized using configuration in file:/etc/hive/conf.dist/hive-log4j.properties
Hive history file=/tmp/root/hive_job_log_root_202401041341_394378730.txt
hive> insert into table danamic table partition(datee) select datee,open,high,low,close,a_close,volume from nor_table;
FAILED: SemanticException [Error 10001]: Line 1:103 Table not found 'nor_table'
hive> use stock_market_partition;
OK
Time taken: 0.1 seconds
hive> insert into table danamic table partition(datee) select datee,open,high,low,close,a_close,volume from nor table;
FAILED: SemanticException [Error 10096]: Dynamic partition strict mode requires at least one static partition column. To turn this off set hive.exec.dynamic.partition.mode=nonstrict
hive> SET hive.exec.dynamic.partition=true;
hive> SET hive.exec.dynamic.partition.mode=nonstrict;
hive> insert into table danamic table partition(datee) select datee,open,high,low,close,a_close,volume from nor_table;
Total MapReduce jobs = 2
Launching Job 1 out of 2
Number of reduce tasks is set to 0 since there's no reduce operator
Starting Job = job_1704389483171_0002, Tracking URL = http://localhost.localdomain:8088/proxy/application_1704389483171_0002/
Kill Command = /usr/lib/hadoop/bin/hadoop job -Dmapred.job.tracker=localhost:9101 -kill job_1704389483171_0002

[2]+ Stopped hive
[root@localhost training]#
```

Cloudera-Udacity-Training YARN.c - VMware Workstation

File Edit View VM Tabs Help | training | Mon Jan 8, 4:27 AM

Cloudera-Udacity-Train...

HDFS:/user/hive/warehouse/uber.db - Mozilla Firefox

File Edit View History Bookmarks Tools Help

HDFS:/user/hive/warehouse/ube...

localhost.localdomain:50075/browseDirectory.jsp?dir=%2Fuser%2Fhive%2Fwarehouse%2Fut...

The Platform for Bi... Hadoop JobTracker Hadoop NameNode

Contents of directory /user/hive/warehouse/uber.db

Goto : /user/hive/warehouse/uber. go

Go to parent directory

Name	Type	Size	Replication	Block Size	Modification Time	Permission	Owner	Group
<a href="#">dal part</a>	dir				2024-01-08 03:48	rwxr-xr-x	root	supergroup
<a href="#">orders</a>	dir				2024-01-05 00:45	rwxr-xr-x	root	supergroup
<a href="#">orders1</a>	dir				2024-01-08 03:14	rwxr-xr-x	root	supergroup
<a href="#">sat part</a>	dir				2024-01-08 03:46	rwxr-xr-x	root	supergroup

Go back to DFS home

Local logs

Log directory

To direct input to this VM, click inside or press Ctrl+G.

Type here to search

27°C 02:57 PM 08-01-2024