Cloudera-Udacity-Training... × CentOS 64-bit (2)

Cloudera-Udacity-Training YARN.c

Power on this virtual machine

Edit virtual machine settings

Upgrade this virtual machine

Devices

Memory 1 GB

Processors

□ Hard Disk (SCSI) 17 GB

Network Adapter NAT

USB Controller Present

Description

Cloudera/Udacity Training VM



▼ Virtual Machine Details

State: Powered off

Configuration file: C:\Users\balaj\Downloads\Cloudera-Yarn-Udacity-Training

VM\Cloudera-Udacity-Training-VM-4.1.1.c\Cloudera-Training-VM-4.1.1.c\Cloudera-Training-.vmx

Hardware compatibility: Workstation 5.x virtual machine

```
File Edit View Search Terminal Help
training@localhost ~]$ sudo su
root@localhost training]# vi supply-chain.csv
root@localhost training]# rm supply-chain.csv
m: remove regular file 'supply-chain.csv'? y
root@localhost training]# vi supply-chain.csv
root@localhost training]# ls
3dec bk.txt derby.log Documents eclipse
                                                   hive
                                                                                                      workspace
                                                             pig 1406462634650.log
                                                                                    src
4dec data
              Desktop
                         Downloads
                                    employees.csv
                                                   Pictures
                                                             Public
                                                                                    udacity training
root@localhost training]# vi 14dec
root@localhost training]# vi supply-chain
root@localhost training]# vi supply-chain.txt
root@localhost training]# vi supply-chain.txt supply.csv
files to edit
11+ Stopped
                             vi supply-chain.txt supply.csv
root@localhost training]# mv supply-chain.txt supply.csv
root@localhost training]# ls
3dec 14dec bk.txt data derby.log Desktop Documents Downloads eclipse employees.csv hive Pictures pig 14064626346
root@localhost training]# head supply.csv
.4.73, Consumer, 2, Pacific Asia, Indonesia, 1, COMPLETE, 1360, 73, 327, 75, Standard Class, 1, 2018
,4,73,Corporate,2,Pacific Asia,Australia,1,PENDING PAYMENT,1360,73,327.75,Standard Class,1,2018
,4,73,Consumer,2,Pacific Asia,Australia,1,CANCELED,1360,73,327.75,Standard Class,1,2018
,1,73, Home Office,2, Pacific Asia, China, 1, COMPLETE, 1360, 73, 327.75, First Class, 1,2018
,1,73,Corporate,2,Pacific Asia,China,1,PROCESSING,1360,73,327.75,First Class,1,2018
,2,73,Corporate,2,Pacific Asia,China,1,CLOSED,1360,73,327.75,Second Class,1,2018
,1,73,Corporate,2,Pacific Asia,China,1,CLOSED,1360,73,327.75,First Class,1,2018
,2,73,Corporate,2,Pacific Asia,Japón,1,SUSPECTED FRAUD,1360,73,327.75,Second Class,1,2018
,2,73,Consumer,2,Pacific Asia,Indonesia,1,PENDING,1360,73,327.75,Second Class,1,2018
,2,73,Corporate,2,Pacific Asia,Indonesia,1,PENDING,1360,73,327.75,Second Class,1,2018
root@localhost training]#
```

File Edit View Search Terminal Help

[training@localhost ~]\$ sudo su [root@localhost training]# hive Logging initialized using configuration in file:/etc/hive/conf.dist/hive-log4j.p Hive history file=/tmp/root/hive job log root 202401041345 257322594.txt hive> create database supply chain; Database supply chain already exists FAILED: Execution Error, return code 1 from org.apache.hadoop.hive.ql.exec.DDLTa hive> use supply chain; OK Time taken: 0.027 seconds hive> create table supply data(day of shipping int, day of shipment int, categor ategory id int, product price float, shiiping mode string, order month string, order OK Time taken: 1.722 seconds hive> load data local inpath '/home/training/supply.csv' into table supply data; Copying data from file:/home/training/supply.csv Copying file: file:/home/training/supply.csv Loading data to table supply chain.supply data OK. Time taken: 0.942 seconds hive> set hive.exec.dynamic.partition=true; hive> set hive.exec.dynamic.partition.mode=nonstrict; hive> create table supply data dyn(day of shipping int, day of shipment int, cat

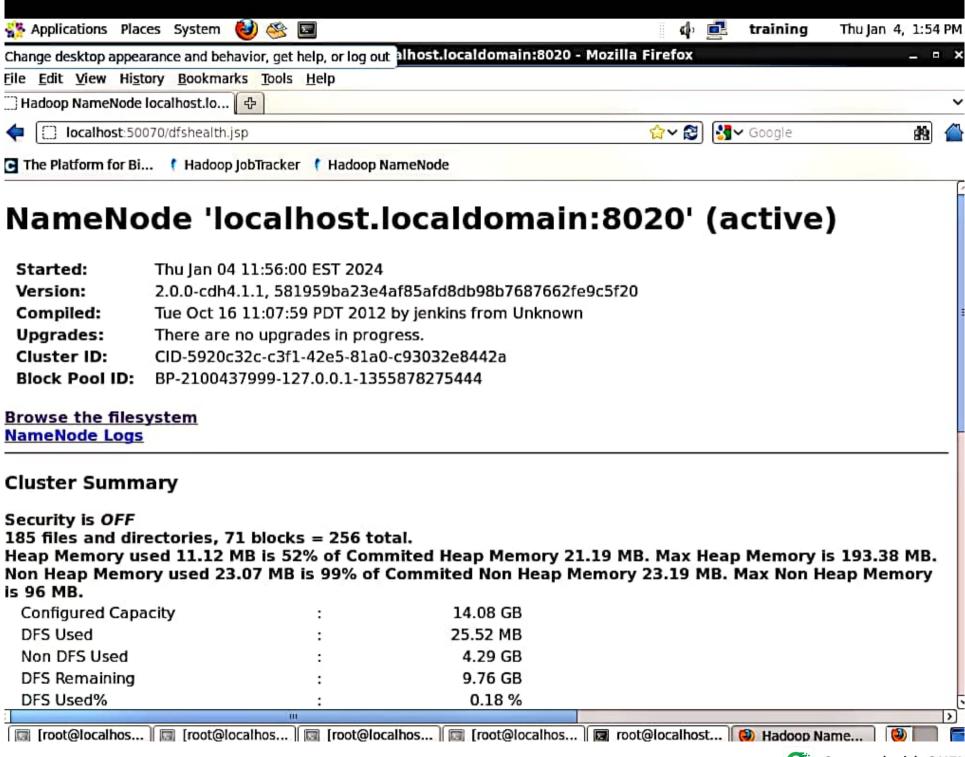
int, product price float, shiiping mode string) partitioned by(order year string, ▽

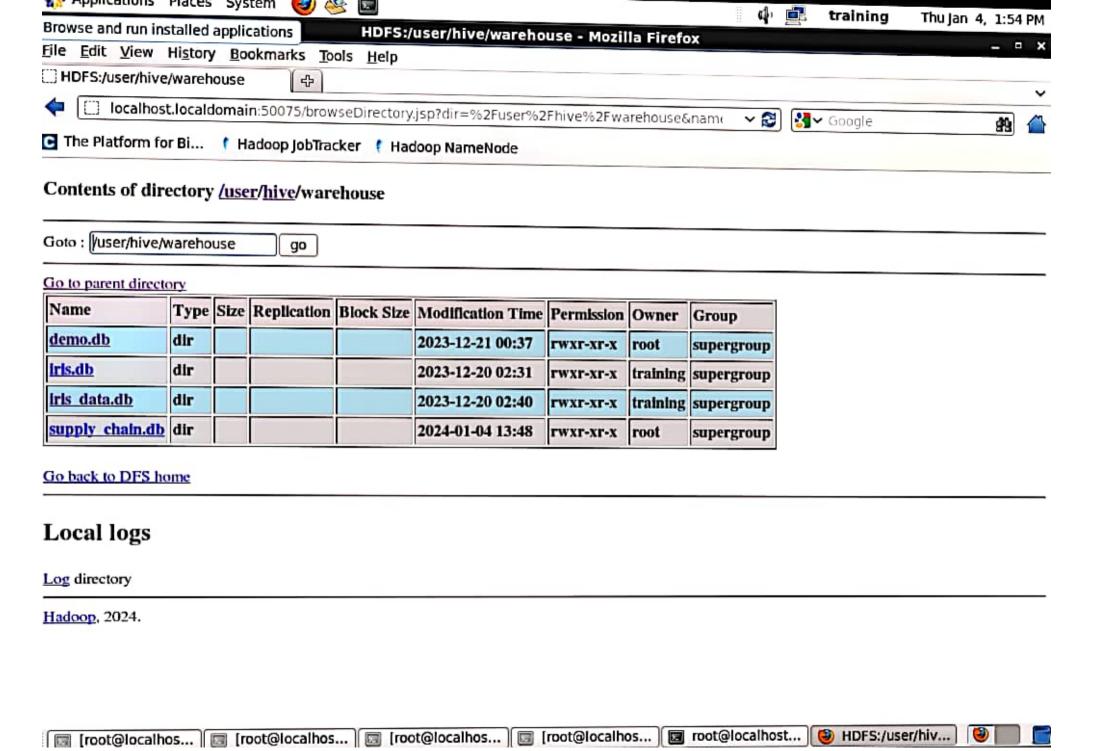
```
7.
File Edit View Search Terminal Help
nive> set hive.exec.dynamic.partition.mode=nonstrict;
rive> create table supply data_dyn(day of shipping int, day of shipment int, cat
int, product price float, shiiping mode string) partitioned by (order year string,
)K
ime taken: 0.099 seconds
nive> desc supply data;
)K
lay of shipping int
lay of shipment int
ategory id
               int
:ustomer segment
                       string
lept id int
narket string
order country string
order item quantity
                       int
order status string
product card id int
roduct category id
                       int
product price float
hiiping mode string
order month string
order year string
ime taken: 0.255 seconds
nive> desc supply data dyn;
```

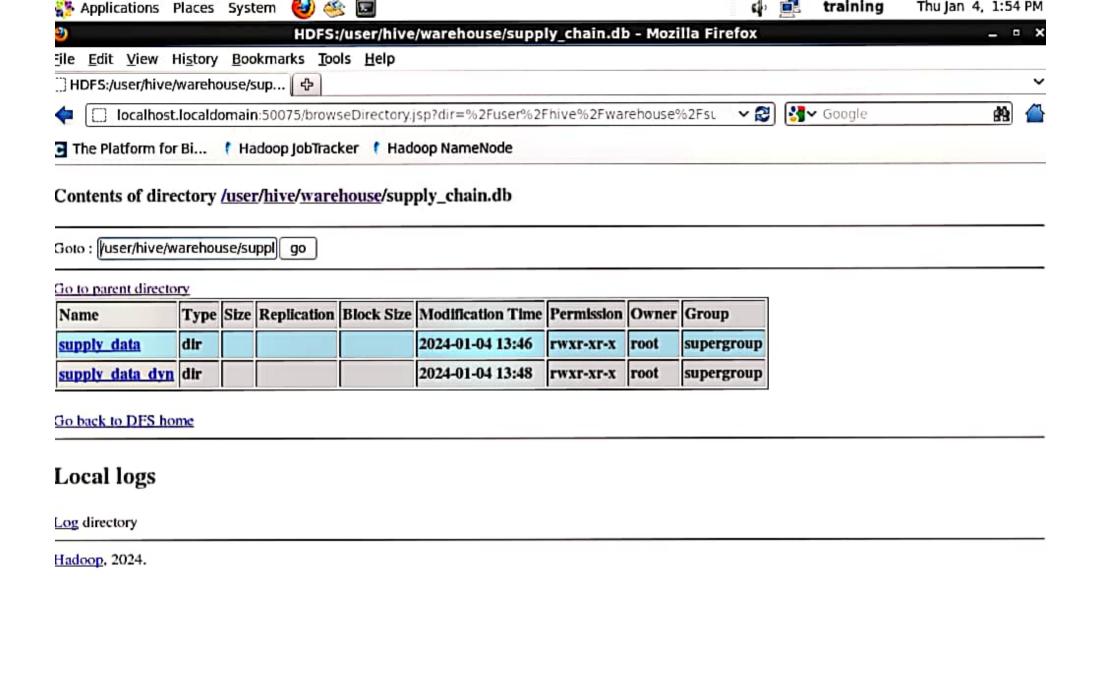
```
root@localhost:/home/training
File Edit View Search Terminal Help
          string
order year
Time taken: 0.255 seconds
nive> desc supply data dyn;
)K
lay of shipping int
lay of shipment int
:ategory id
               int
:ustomer segment
                       string
fept id int
order country string
order item quantity
                       int
order status
               string
product card id int
product category id
                       int
product price float
shiiping mode string
order year
           string
market string
order month
               string
Time taken: 0.155 seconds
nive> insert into table supply data dyn partition(order year,market,order month)
: category id, product price, shiiping mode, order_year, market, order_month from sup
FAILED: SemanticException [Error 10044]: Line 1:18 Cannot insert into target tab
nive> insert into table supply data dyn partition(order year,market,order_month)
```

root@localhost:/home/training 73 File Edit View Search Terminal Help order year string Time taken: 0.137 seconds hive> create table supply data dynpart(day of shipping int,day of shipment int,c ct category id int, product price float, shiiping mode string, order month string) Time taken: 0.25 seconds hive> insert into table supply data dynpart partition?(order year) select day of product price, shiiping mode , order month, order year from supply data; FAILED: ParseException line 1:47 mismatched input '?' expecting (near 'partitio hive> insert into table supply data dynpart partition(order year) select day of roduct price, shiiping mode , order month, order year from supply data; FAILED: ParseException line 1:145 cannot recognize input near ',' 'order country hive> insert into table supply data dynpart partition(order year) select day of oduct price, shiiping mode , order month, order year from supply data; 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 1704396276937 0001, Tracking URL = http://localhost.localdoma Kill Command = /usr/lib/hadoop/bin/hadoop job -Dmapred.job.tracker=localhost:91 Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 0 2024-01-04 14:40:03,905 Stage-1 map = 0%, reduce = 0% 2024-01-04 14:40:13,259 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 2.91 se ✓

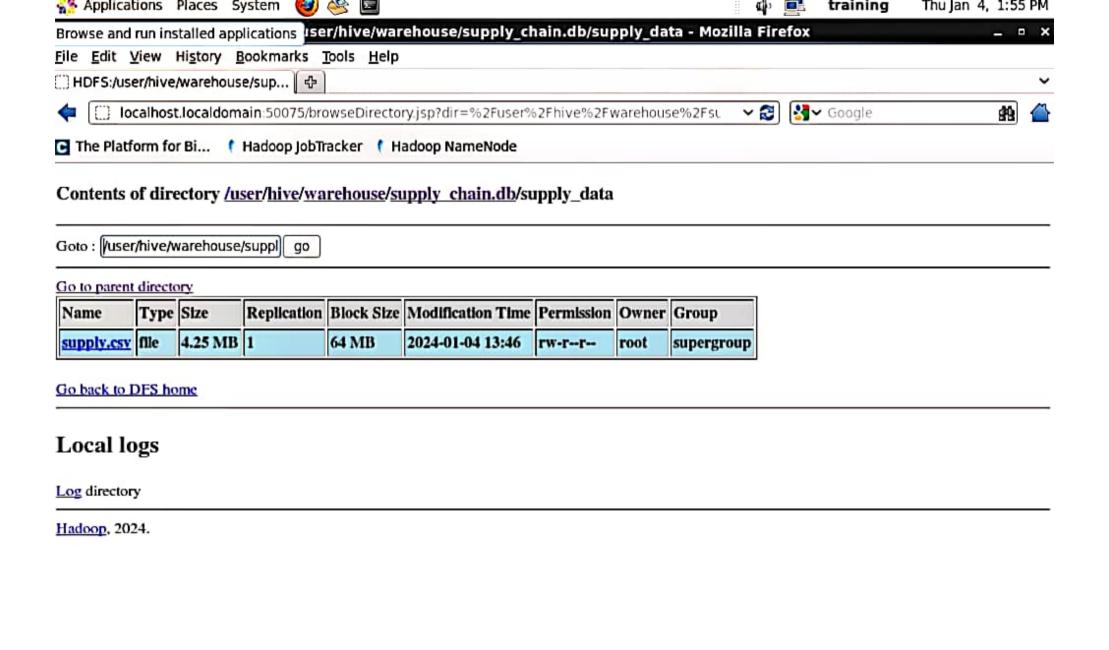
```
File Edit View Search Terminal Help
duct price, shiiping mode , order month, order year from supply data;
otal MapReduce jobs = 2
aunching Job 1 out of 2
lumber of reduce tasks is set to 0 since there's no reduce operator
tarting Job = job 1704396276937 0001, Tracking URL = http://localhost.localdoma
ill Command = /usr/lib/hadoop/bin/hadoop job -Dmapred.job.tracker=localhost:91
ladoop job information for Stage-1: number of mappers: 1; number of reducers: 0
:024-01-04 14:40:03,905 Stage-1 map = 0%, reduce = 0%
'024-01-04 14:40:13,259 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 2.91 se
TapReduce Total cumulative CPU time: 2 seconds 910 msec
inded Job = job 1704396276937 0001
inded Job = -280165524, job is filtered out (removed at runtime).
loving data to: hdfs://localhost:8020/tmp/hive-root/hive 2024-01-04 14-39-44 063
oading data to table supply chain supply data dynpart partition (order year=nul
       Loading partition {order year=2015}
       Loading partition {order year=2018}
       Loading partition {order year= HIVE DEFAULT PARTITION }
       Loading partition {order year=2017}
       Loading partition {order year=2016}
50999 Rows loaded to supply data dynpart
MapReduce Jobs Launched:
Job 0: Map: 1 Cumulative CPU: 2.91 sec HDFS Read: 4459305 HDFS Write: 398312
Total MapReduce CPU Time Spent: 2 seconds 910 msec
DK.
```







[[root@localhos...] [[root@localhos...]



☐ [root@localhos...] ☐ [root@localhos...] ☐ [root@localhos...] ☐ [root@localhos...] ☐ root@localhos...] ☐ [root@localhos...] ☐ [root@l

The Platform for Bi... 🕴 Hadoop JobTracker 🕴 Hadoop NameNode

Contents of directory /user/hive/warehouse/supply_chain.db/supply_data_dynpart

Goto : //user/hive/warehouse/suppl go

Go to parent directory

Name	Type	Size	Replication	Block Size	Modification Time	Permission	Owner	Group
order_year=2015	dir				2024-01-04 14:40	rwxr-xr-x	root	supergroup
order year=2016	dir				2024-01-04 14:40	rwxr-xr-x	root	supergroup
order_year=2017	dir				2024-01-04 14:40	rwxr-xr-x	root	supergroup
order year=2018	dir				2024-01-04 14:40	rwxr-xr-x	root	supergroup
order year= HIVE DEFAULT PARTITION	dir				2024-01-04 14:40	rwxr-xr-x	root	supergroup

Go back to DFS home

Local logs

og directory

Hadoop, 2024.