

Cloudera-Udacity-Training YARN.c

- ▶ Power on this virtual machine
- 🔧 Edit virtual machine settings
- 🔄 Upgrade this virtual machine

▼ Devices

🖥️ Memory	1 GB
🖨️ Processors	1
💾 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

```
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 pig_1406462634650.log src workspace
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

1]+ 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]#
```

```
root@localhost:/home/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,
```

File Edit View Search Terminal Help

```
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,shipping_mode string) partitioned by(order_year string,
OK
Time taken: 0.099 seconds
hive> desc supply_data;
OK
day_of_shipping int
day_of_shipment int
category_id      int
customer_segment string
dept_id int
market string
order_country   string
order_item_quantity int
order_status    string
product_card_id int
product_category_id int
product_price   float
shipping_mode   string
order_month     string
order_year      string
Time taken: 0.255 seconds
hive> desc supply_data_dyn;
```

```
root@localhost:/home/training
File Edit View Search Terminal Help
order_year      string
Time taken: 0.255 seconds
hive> desc supply_data_dyn;
OK
day_of_shipping int
day_of_shipment int
category_id      int
customer_segment string
dept_id int
order_country    string
order_item_quantity int
order_status     string
product_card_id  int
product_category_id int
product_price    float
shipping_mode    string
order_year       string
market           string
order_month      string
Time taken: 0.155 seconds
hive> insert into table supply_data_dyn partition(order_year,market,order_month)
:_category_id,product_price,shipping_mode,order_year,market,order_month from sup
FAILED: SemanticException [Error 10044]: Line 1:18 Cannot insert into target tab
hive> insert into table supply_data_dyn partition(order_year,market,order_month)
```

```
root@localhost:/home/training
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)
OK
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,shipping_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.localdomain:8020/jobtracking/job_1704396276937_0001
Hadoop Command = /usr/lib/hadoop/bin/hadoop job -Dmapred.job.tracker=localhost:9001
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 sec
MapReduce Total cumulative CPU time: 2 seconds 910 msec
Finished Job = job_1704396276937_0001
Finished Job = -280165524, job is filtered out (removed at runtime).
Moving data to: hdfs://localhost:8020/tmp/hive-root/hive_2024-01-04_14-39-44_063
Loading data to table supply_chain.supply_data_dynpart partition (order_year=null)
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
OK
```


NameNode 'localhost.localdomain:8020' (active)

Started: Thu Jan 04 11:56:00 EST 2024
Version: 2.0.0-cdh4.1.1, 581959ba23e4af85afd8db98b7687662fe9c5f20
Compiled: Tue Oct 16 11:07:59 PDT 2012 by jenkins from Unknown
Upgrades: There are no upgrades in progress.
Cluster ID: CID-5920c32c-c3f1-42e5-81a0-c93032e8442a
Block Pool ID: BP-2100437999-127.0.0.1-1355878275444

[Browse the filesystem](#)
[NameNode Logs](#)

Cluster Summary

Security is **OFF**

185 files and directories, 71 blocks = 256 total.

Heap Memory used 11.12 MB is 52% of Committed Heap Memory 21.19 MB. Max Heap Memory is 193.38 MB.
Non Heap Memory used 23.07 MB is 99% of Committed Non Heap Memory 23.19 MB. Max Non Heap Memory is 96 MB.

Configured Capacity	:	14.08 GB
DFS Used	:	25.52 MB
Non DFS Used	:	4.29 GB
DFS Remaining	:	9.76 GB
DFS Used%	:	0.18 %

Applications Places System

training Thu Jan 4, 1:54 PM

Browse and run installed applications

HDFS:/user/hive/warehouse - Mozilla Firefox

File Edit View History Bookmarks Tools Help

HDFS:/user/hive/warehouse

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

Google

The Platform for Bi... Hadoop JobTracker Hadoop NameNode

Contents of directory /user/hive/warehouse

Goto :

[Go to parent directory](#)

Name	Type	Size	Replication	Block Size	Modification Time	Permkssion	Owner	Group
demo.db	dir				2023-12-21 00:37	rwxr-xr-x	root	supergroup
iris.db	dir				2023-12-20 02:31	rwxr-xr-x	training	supergroup
iris_data.db	dir				2023-12-20 02:40	rwxr-xr-x	training	supergroup
supply_chain.db	dir				2024-01-04 13:48	rwxr-xr-x	root	supergroup

[Go back to DFS home](#)

Local logs

[Log directory](#)

[Hadoop](#), 2024.

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

Goto : /user/hive/warehouse/suppl go

[Go to parent directory](#)

Name	Type	Size	Replication	Block Size	Modification Time	Permisslon	Owner	Group
supply_data	dir				2024-01-04 13:46	rwxr-xr-x	root	supergroup
supply_data_dyn	dir				2024-01-04 13:48	rwxr-xr-x	root	supergroup

[Go back to DFS home](#)

Local logs

[Log directory](#)[Hadoop](#), 2024.

ApplicationsPlacesSystem

trainingThu Jan 4, 1:55 PM

Browse and run installed applicationsuser/hive/warehouse/supply_chain.db/supply_data - Mozilla Firefox

FileEditViewHistoryBookmarksToolsHelp

HDFS:/user/hive/warehouse/sup...+

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

Google

The Platform for Bi...Hadoop JobTrackerHadoop NameNode

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

Goto : /user/hive/warehouse/suppl go

[Go to parent directory](#)

Name	Type	Size	Replication	Block Size	Modification Time	Permisslon	Owner	Group
supply.csv	file	4.25 MB	1	64 MB	2024-01-04 13:46	rw-r--r--	root	supergroup

[Go back to DFS home](#)

Local logs

[Log directory](#)

[Hadoop](#), 2024.

Contents of directory [/user/hive/warehouse/supply_chain.db/supply_data_dynpart](#)

Goto :

[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

[Log](#) directory

Hadoop, 2024.

