```
akansha@akansha-VirtualBox:~$ ssh localhost
Welcome to Ubuntu 18.04.6 LTS (GNU/Linux 5.4.0-148-generic x86_64)

 * Documentation:  https://help.ubuntu.com
 * Management:     https://landscape.canonical.com
 * Support:        https://ubuntu.com/advantage

62 updates can be applied immediately.
21 of these updates are standard security updates.
To see these additional updates run: apt list --upgradable

Ubuntu comes with ABSOLUTELY NO WARRANTY, to the extent permitted by
applicable law.

Failed to connect to https://changelogs.ubuntu.com/meta-release-lts. Check your Internet
connection or proxy settings

Your Hardware Enablement Stack (HWE) is supported until April 2023.
Last login: Sun May 28 00:34:40 2023 from 127.0.0.1
akansha@akansha-VirtualBox:~$ stop-all.sh
WARNING: Stopping all Apache Hadoop daemons as akansha in 10 seconds.
WARNING: Use CTRL-C to abort.
Stopping namenodes on [localhost]
Stopping datanodes
Stopping secondary namenodes [akansha-VirtualBox]
Stopping nodemanagers
Stopping resourcemanager
akansha@akansha-VirtualBox:~$ hadoop namenode -format
WARNING: Use of this script to execute namenode is deprecated.
WARNING: Attempting to execute replacement "hdfs namenode" instead.

2023-05-28 00:57:57,690 INFO namenode.NameNode: STARTUP_MSG:
/************************************************************
STARTUP_MSG: Starting NameNode
STARTUP_MSG:   host = akansha-VirtualBox/127.0.1.1
STARTUP_MSG:   args = [-format]
STARTUP_MSG:   version = 3.2.4
STARTUP_MSG:   classpath =
/home/akansha/hadoop-3.2.4//etc/hadoop:/home/akansha/hadoop-3.2.4//share/hadoop/com
mon/lib/jackson-databind-2.10.5.1.jar:/home/akansha/hadoop-3.2.4//share/hadoop/common/l
ib/snappy-java-1.0.5.jar:/home/akansha/hadoop-3.2.4//share/hadoop/common/lib/jaxb-api-2.
                                        |
@@@@@@@@@@OUTPUT DROPPED@@@@@@@@@@@@@@
                                        |
/tmp/hadoop-akansha/dfs/name has been successfully formatted.
2023-05-28 00:58:03,105 INFO namenode.FSImageFormatProtobuf: Saving image file
/tmp/hadoop-akansha/dfs/name/current/fsimage.ckpt_0000000000000000000 using no
compression
```

```
2023-05-28 00:58:03,262 INFO namenode.FSImageFormatProtobuf: Image file
/tmp/hadoop-akansha/dfs/name/current/fsimage.ckpt_0000000000000000000 of size 399
bytes saved in 0 seconds .
2023-05-28 00:58:03,283 INFO namenode.NNStorageRetentionManager: Going to retain 1
images with txid >= 0
2023-05-28 00:58:03,325 INFO namenode.FSNamesystem: Stopping services started for
active state
2023-05-28 00:58:03,325 INFO namenode.FSNamesystem: Stopping services started for
standby state
2023-05-28 00:58:03,335 INFO namenode.FSImage: FSImageSaver clean checkpoint:
txid=0 when meet shutdown.
2023-05-28 00:58:03,336 INFO namenode.NameNode: SHUTDOWN_MSG:
/************************************************************
SHUTDOWN_MSG: Shutting down NameNode at akansha-VirtualBox/127.0.1.1
************************************************************/
akansha@akansha-VirtualBox:~$ start-all.sh
WARNING: Attempting to start all Apache Hadoop daemons as akansha in 10 seconds.
WARNING: This is not a recommended production deployment configuration.
WARNING: Use CTRL-C to abort.
Starting namenodes on [localhost]
Starting datanodes
Starting secondary namenodes [akansha-VirtualBox]
Starting resourcemanager
Starting nodemanagers
akansha@akansha-VirtualBox:~$ hive
Hive Session ID = 8b767977-22e5-4f71-8e40-301af6465778

Logging initialized using configuration in
jar:file:/home/akansha/apache-hive-3.1.2-bin/lib/hive-common-3.1.2.jar!/hive-log4j2.propertie
s Async: true
Hive-on-MR is deprecated in Hive 2 and may not be available in the future versions.
Consider using a different execution engine (i.e. spark, tez) or using Hive 1.X releases.
Hive Session ID = a1531a37-ed59-417c-aad1-d3c55243974f
hive> SHOW DATABASES;
OK
default
flight_info_system
flightinfo
Time taken: 0.411 seconds, Fetched: 3 row(s)
hive> USE flight_info_system;
OK
Time taken: 0.026 seconds
hive> SHOW TABLES;
OK
airports
Time taken: 0.029 seconds, Fetched: 1 row(s)
hive> CREATE TABLE flight_info(
> FL_NUM INT,
```

```
          > ORIGIN_AIRPORT_ID INT,
          > ORIGIN STRING,
          > DEST_AIRPORT_ID INT,
          > DEST STRING,
          > DEP_TIME INT,
          > DEP_DELAY INT,
          > ARR_TIME INT,
          > ARR_DELAY INT
          > )
          > ROW FORMAT DELIMITED FIELDS TERMINATED BY ','
          > tblproperties ("skip.header.line.count"="1");
OK
Time taken: 0.026 seconds
hive>
          > LOAD DATA LOCAL INPATH
          > '/home/akansha/flightdata.csv'
          > INTO TABLE flight_info;
Loading data to table default.flight_info
OK
Time taken: 1.095 seconds
hive> select count(*) from flight_info;
Query ID = akansha_20230528015929_d69d433b-a31a-4403-b996-e9626aaedd7f
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_1685219240726_0001, Tracking URL =
http://akansha-VirtualBox:8088/proxy/application_1685219240726_0001/
Kill Command = /home/akansha/hadoop-3.2.4/bin/mapred job  -kill
job_1685219240726_0001
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2023-05-28 01:59:42,844 Stage-1 map = 0%,  reduce = 0%
2023-05-28 01:59:48,030 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 1.95 sec
2023-05-28 01:59:53,195 Stage-1 map = 100%,  reduce = 100%, Cumulative CPU 3.55 sec
MapReduce Total cumulative CPU time: 3 seconds 550 msec
Ended Job = job_1685219240726_0001
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1  Reduce: 1   Cumulative CPU: 3.55 sec   HDFS Read: 486791 HDFS
Write: 105 SUCCESS
Total MapReduce CPU Time Spent: 3 seconds 550 msec
OK
11231
Time taken: 26.967 seconds, Fetched: 1 row(s)
```

```
hive> ALTER TABLE flight_info RENAME TO flight_info_internal;
OK
Time taken: 0.171 seconds
hive> show tables;
OK
flight_info_internal
Time taken: 0.031 seconds, Fetched: 1 row(s)
hive> DROP TABLE flight_info_internal;
OK
Time taken: 0.405 seconds
hive> show tables;
OK
Time taken: 0.017 seconds
hive>

hive>
            > CREATE EXTERNAL TABLE flight_info(
            > FL_NUM INT,
            > ORIGIN_AIRPORT_ID INT,
            > ORIGIN STRING,
            > DEST_AIRPORT_ID INT,
            > DEST STRING,
            > DEP_TIME INT,
            > DEP_DELAY INT,
            > ARR_TIME INT,
            > ARR_DELAY INT
            > )
            > ROW FORMAT DELIMITED FIELDS TERMINATED BY ','
            > STORED AS TEXTFILE
            > LOCATION '/home/akansha/flightdata.txt'
            > tblproperties ("skip.header.line.count"="1");
OK
Time taken: 0.134 seconds
hive> select count(*) from flight_info;
Query ID = akansha_20230528021330_34732551-1e8f-437f-bcdc-dfe7697ee18d
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_1685219240726_0002, Tracking URL =
http://akansha-VirtualBox:8088/proxy/application_1685219240726_0002/
Kill Command = /home/akansha/hadoop-3.2.4/bin/mapred job  -kill
job_1685219240726_0002
```

```
Hadoop job information for Stage-1: number of mappers: 0; number of reducers: 1
2023-05-28 02:13:38,038 Stage-1 map = 0%,  reduce = 0%
2023-05-28 02:13:44,220 Stage-1 map = 0%,  reduce = 100%, Cumulative CPU 2.08 sec
MapReduce Total cumulative CPU time: 2 seconds 80 msec
Ended Job = job_1685219240726_0002
MapReduce Jobs Launched:
Stage-Stage-1: Reduce: 1   Cumulative CPU: 2.08 sec   HDFS Read: 6683 HDFS Write: 101
SUCCESS
Total MapReduce CPU Time Spent: 2 seconds 80 msec
OK
0
Time taken: 14.891 seconds, Fetched: 1 row(s)

/*
LOCAL PATH DIDN'T WORK IN EXTERNAL TABLE
*/

hive> show tables;
OK
flight_info
Time taken: 0.02 seconds, Fetched: 1 row(s)
hive> LOAD DATA LOCAL INPATH
          > '/home/akansha/flightdata.csv'
          > INTO TABLE flight_info;
Loading data to table default.flight_info
OK
Time taken: 0.616 seconds
hive> show tables;
OK
flight_info
Time taken: 0.029 seconds, Fetched: 1 row(s)
hive> select count(*) from flight_info;
Query ID = akansha_20230528021451_47aa1d6b-d321-469e-8d09-3bf492bed2dd
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_1685219240726_0003, Tracking URL =
http://akansha-VirtualBox:8088/proxy/application_1685219240726_0003/
Kill Command = /home/akansha/hadoop-3.2.4/bin/mapred job  -kill
job_1685219240726_0003
```

```
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2023-05-28 02:14:57,540 Stage-1 map = 0%,  reduce = 0%
2023-05-28 02:15:01,669 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 1.44 sec
2023-05-28 02:15:06,822 Stage-1 map = 100%,  reduce = 100%, Cumulative CPU 3.05 sec
MapReduce Total cumulative CPU time: 3 seconds 50 msec
Ended Job = job_1685219240726_0003
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1  Reduce: 1   Cumulative CPU: 3.05 sec   HDFS Read: 486784 HDFS
Write: 105 SUCCESS
Total MapReduce CPU Time Spent: 3 seconds 50 msec
OK
11231
Time taken: 17.031 seconds, Fetched: 1 row(s)

hive> DROP TABLE flight_info;
OK
Time taken: 0.12 seconds

/*
IN NEW TERMINAL, COPY flightdata.txt FILE INTO HDFS USING
COMMAND:
akansha@akansha-VirtualBox:~$ hdfs dfs -copyFromLocal flightdata.txt /data1/
*/

hive>
hive> CREATE EXTERNAL TABLE flight_info(
          > FL_NUM INT,
          > ORIGIN_AIRPORT_ID INT,
          > ORIGIN STRING,
          > DEST_AIRPORT_ID INT,
          > DEST STRING,
          > DEP_TIME INT,
          > DEP_DELAY INT,
          > ARR_TIME INT,
          > ARR_DELAY INT
          > )
          > ROW FORMAT DELIMITED FIELDS TERMINATED BY ','
          > STORED AS TEXTFILE
          > LOCATION '/data1'
          > tblproperties ("skip.header.line.count"="1");
OK
Time taken: 0.063 seconds
hive> select count(*) from flight_info;
Query ID = akansha_20230528022054_c5756f67-d906-4755-b587-16a859f65a93
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_1685219240726_0004, Tracking URL =
http://akansha-VirtualBox:8088/proxy/application_1685219240726_0004/
Kill Command = /home/akansha/hadoop-3.2.4/bin/mapred job  -kill
job_1685219240726_0004
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2023-05-28 02:21:00,625 Stage-1 map = 0%,  reduce = 0%
2023-05-28 02:21:04,751 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 1.74 sec
2023-05-28 02:21:09,915 Stage-1 map = 100%,  reduce = 100%, Cumulative CPU 3.16 sec
MapReduce Total cumulative CPU time: 3 seconds 160 msec
Ended Job = job_1685219240726_0004
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1  Reduce: 1   Cumulative CPU: 3.16 sec   HDFS Read: 464238 HDFS
Write: 105 SUCCESS
Total MapReduce CPU Time Spent: 3 seconds 160 msec
OK
11231
Time taken: 16.974 seconds, Fetched: 1 row(s)
hive>


hive> INSERT INTO TABLE flight_info
VALUES(1872,10397,"ATL",14757,"SEA",1111,-2,1316,-24);
Query ID = akansha_20230528023519_b62e39c2-ee02-4580-aa65-8718eab89ea6
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_1685219240726_0005, Tracking URL =
http://akansha-VirtualBox:8088/proxy/application_1685219240726_0005/
Kill Command = /home/akansha/hadoop-3.2.4/bin/mapred job  -kill
job_1685219240726_0005
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2023-05-28 02:35:26,025 Stage-1 map = 0%,  reduce = 0%
2023-05-28 02:35:32,252 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 2.42 sec
2023-05-28 02:35:36,366 Stage-1 map = 100%,  reduce = 100%, Cumulative CPU 3.82 sec
```

```
MapReduce Total cumulative CPU time: 3 seconds 820 msec
Ended Job = job_1685219240726_0005
Stage-4 is selected by condition resolver.
Stage-3 is filtered out by condition resolver.
Stage-5 is filtered out by condition resolver.
Moving data to directory
hdfs://localhost:9000/data1/.hive-staging_hive_2023-05-28_02-35-19_811_64121228598362
85794-1/-ext-10000
Loading data to table default.flight_info
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1  Reduce: 1   Cumulative CPU: 3.82 sec   HDFS Read: 23100 HDFS
Write: 536 SUCCESS
Total MapReduce CPU Time Spent: 3 seconds 820 msec
OK
Time taken: 18.927 seconds


hive> ALTER TABLE flight_info
    > ADD COLUMNS (cancelled BOOLEAN);
OK
Time taken: 0.088 seconds


hive> describe flight_info;
OK
fl_num                  int
origin_airport_id       int
origin                  string
dest_airport_id                 int
dest                    string
dep_time                int
dep_delay               int
arr_time                int
arr_delay               int
cancelled               boolean
Time taken: 0.029 seconds, Fetched: 10 row(s)

hive>
hive> ALTER TABLE flight_info
    > SET TBLPROPERTIES ('default.cancelled'='False');
OK
Time taken: 0.052 seconds
hive>
hive> SELECT AVG(DEP_DELAY) AS average
    > FROM flight_info
    > WHERE DEP_TIME=2008;
Query ID = akansha_20230528030342_405c8a49-5c0b-4d13-bc79-bf95eeba20ee
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_1685219240726_0006, Tracking URL =
http://akansha-VirtualBox:8088/proxy/application_1685219240726_0006/
Kill Command = /home/akansha/hadoop-3.2.4/bin/mapred job  -kill
job_1685219240726_0006
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2023-05-28 03:03:48,923 Stage-1 map = 0%,  reduce = 0%
2023-05-28 03:03:54,055 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 2.49 sec
2023-05-28 03:03:59,204 Stage-1 map = 100%,  reduce = 100%, Cumulative CPU 4.69 sec
MapReduce Total cumulative CPU time: 4 seconds 690 msec
Ended Job = job_1685219240726_0006
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1  Reduce: 1   Cumulative CPU: 4.69 sec   HDFS Read: 468316 HDFS
Write: 118 SUCCESS
Total MapReduce CPU Time Spent: 4 seconds 690 msec
OK
12.571428571428571
Time taken: 17.591 seconds, Fetched: 1 row(s)
hive>
```