

Output-

Configuring /.bashrc file and creating directories in folder-

```
hdp123@stud-OptiPlex-3060: ~/apache-hive-2.3.9-bin/bin
hdp123@stud-OptiPlex-3060: $ sudo nano ~/.bashrc
hdp123@stud-OptiPlex-3060: $ source ~/.bashrc
hdp123@stud-OptiPlex-3060: $ ls
analyzeLogs  apache-hive-2.3.9-bin  Desktop  Downloads  input2002  Pictures  snap  Videos
analyzeLogs  apache-hive-3.1.2-bin.tar.gz  Documents  input2000  Music  Public  Templates
hdp123@stud-OptiPlex-3060: $ cd apache-hive-2.3.9-bin/
hdp123@stud-OptiPlex-3060: ~/apache-hive-2.3.9-bin$ cd ..
hdp123@stud-OptiPlex-3060: $ sudo chown -R hdp123 apache-hive-2.3.9-bin/
hdp123@stud-OptiPlex-3060: $ sudo apache-hive-2.3.9-bin/
sudo: apache-hive-2.3.9-bin/: command not found
hdp123@stud-OptiPlex-3060: $ cd apache-hive-2.3.9-bin/
hdp123@stud-OptiPlex-3060: ~/apache-hive-2.3.9-bin$ ls
bin  binary-package-licenses  conf  examples  hcatalog  jdbc  lib  LICENSE  NOTICE  RELEASE_NOTES.txt  scripts
hdp123@stud-OptiPlex-3060: ~/apache-hive-2.3.9-bin$ cd conf/
hdp123@stud-OptiPlex-3060: ~/apache-hive-2.3.9-bin/conf$ ls
beeline-log4j2.properties.template  hive-exec-log4j2.properties.template  llap-cli-log4j2.properties.template
hive-default.xml  hive-log4j2.properties.template  llap-daemon-log4j2.properties.template
hive-env.sh  ivysettings.xml  parquet-logging.properties
hdp123@stud-OptiPlex-3060: ~/apache-hive-2.3.9-bin/conf$ sudo nano hive-env.sh
hdp123@stud-OptiPlex-3060: ~/apache-hive-2.3.9-bin/conf$ cd ..
hdp123@stud-OptiPlex-3060: ~/apache-hive-2.3.9-bin$ jps
16657 SecondaryNameNode
16244 NameNode
18392 Jps
16809 ResourceManager
17115 NodeManager
16460 DataNode
hdp123@stud-OptiPlex-3060: ~/apache-hive-2.3.9-bin$ hdfs dfs -mkdir -p /user/hive/warehouse
23/05/02 14:29:01 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
hdp123@stud-OptiPlex-3060: ~/apache-hive-2.3.9-bin$ hdfs dfs -mkdir -p /tmp
23/05/02 14:29:09 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
hdp123@stud-OptiPlex-3060: ~/apache-hive-2.3.9-bin/bin
command [genericOptions] [commandOptions]
hdp123@stud-OptiPlex-3060: ~/apache-hive-2.3.9-bin$ hdfs dfs -chmod 777 /user/hive/warehouse
23/05/02 14:30:13 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
hdp123@stud-OptiPlex-3060: ~/apache-hive-2.3.9-bin$ hdfs dfs -chmod 777 /tmp
23/05/02 14:30:21 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
hdp123@stud-OptiPlex-3060: ~/apache-hive-2.3.9-bin$ cd bin
```

Contents of /.bashrc file-

```
hdp123@stud-OptiPlex-3060: ~
GNU nano 6.2 /home/hdp123/.bashrc
# Alias definitions.
# You may want to put all your additions into a separate file like
# ~/.bash_aliases, instead of adding them here directly.
# See /usr/share/doc/bash-doc/examples in the bash-doc package.

if [ -f ~/.bash_aliases ]; then
    . ~/.bash_aliases
fi

# enable programmable completion features (you don't need to enable
# this, if it's already enabled in /etc/bash.bashrc and /etc/profile
# sources /etc/bash.bashrc).
if ! shopt -oq posix; then
    if [ -f /usr/share/bash-completion/bash_completion ]; then
        . /usr/share/bash-completion/bash_completion
    elif [ -f /etc/bash_completion ]; then
        . /etc/bash_completion
    fi
fi

export JAVA_HOME=/usr/lib/jvm/java-8-openjdk-amd64
export HADOOP_HOME=/usr/local/hadoop
export PATH=$PATH:$HADOOP_HOME/bin
export PATH=$PATH:$HADOOP_HOME/sbin
export HADOOP_MAPRED_HOME=$HADOOP_HOME
export HADOOP_COMMON_HOME=$HADOOP_HOME
export HADOOP_HDFS_HOME=$HADOOP_HOME
export YARN_HOME=$HADOOP_HOME
export HADOOP_COMMON_LIB_NATIVE_DIR=$HADOOP_HOME/native
export HADOOP_OPTS="-Djava.library.path=$HADOOP_HOME/native"
export PATH=$PATH:$HIVE_HOME/bin
export HIVE_CONF_DIR=/home/hdp123/apache-hive-2.3.9-bin/conf
export HIVE_HOME=/home/hdp123/apache-hive-2.3.9-bin
```

Setting up the hive environment-

```
hdp123@stud-OptiPlex-3060: ~/apache-hive-2.3.9-bin/bin
- info Show config and schema details
- initSchema Schema initialization
- initSchemaTo <initTo> Schema initialization to a version
- passWord <password> Override config file password
- servers <serverList> a comma-separated list of servers used
in location validation
- upgradeSchema Schema upgrade
- upgradeSchemaFrom <upgradeFrom> Schema upgrade from a version
- userName <user> Override config file user name
- validate Validate the database
- verbose only print SQL statements
hdp123@stud-OptiPlex-3060:~/apache-hive-2.3.9-bin/bin$ ls
beeline ext hive hive-config.sh hiveserver2 hplsql metatool schematool
hdp123@stud-OptiPlex-3060:~/apache-hive-2.3.9-bin/bin$ ./schematool -initSchema -dbType derby
SLF4J: Class path contains multiple SLF4J bindings.
SLF4J: Found binding in [jar:file:/home/hdp123/apache-hive-2.3.9-bin/lib/log4j-slf4j-impl-2.6.2.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: Found binding in [jar:file:/usr/local/hadoop/share/hadoop/common/lib/slf4j-log4j12-1.7.25.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.apache.logging.slf4j.Log4jLoggerFactory]
Metastore connection URL: jdbc:derby::databaseName=metastore_db;create=true
Metastore Connection Driver : org.apache.derby.jdbc.EmbeddedDriver
Metastore connection User: APP
Starting metastore schema initialization to 2.3.0
Initialization script hive-schema-2.3.0.derby.sql
Initialization script completed
schematool completed
```

Executing basic HiveQL commands-

```
hdp123@stud-OptiPlex-3060: ~/apache-hive-2.3.9-bin/bin
schemaTool completed
hdp123@stud-OptiPlex-3060:~/apache-hive-2.3.9-bin/bin$ ./hive
SLF4J: Class path contains multiple SLF4J bindings.
SLF4J: Found binding in [jar:file:/home/hdp123/apache-hive-2.3.9-bin/lib/log4j-slf4j-impl-2.6.2.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: Found binding in [jar:file:/usr/local/hadoop/share/hadoop/common/lib/slf4j-log4j12-1.7.25.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.apache.logging.slf4j.Log4jLoggerFactory]

Logging initialized using configuration in jar:file:/home/hdp123/apache-hive-2.3.9-bin/lib/hive-common-2.3.9.jar!/hive-log4j2.properties 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, t
ez) or using Hive 1.X releases.

hive> show databases;
OK
default
Time taken: 0.192 seconds, Fetched: 1 row(s)
hive> create database tp;
OK
Time taken: 0.104 seconds
```

Creating Emp table-

```
hive> create table emp(ename string,esal int) row format delimited fields terminated by '_' stored as textfile;
OK
Time taken: 0.471 seconds
```

Loading data with sample.txt file-

```
hive> load data local inpath '/home/hdp123/Desktop/sample.txt' into table emp;
Loading data to table default.emp
OK
Time taken: 1.24 seconds
```

Contents of sample.txt file-

```
hdp123@stud-OptiPlex-3060: ~/Desktop
GNU nano 6.2 sample.txt
manudip' 003
'rajat' 005
'nishant' 026
'pranav' 010
'baste' 034
```


Creating table flight and loading with flight_data.csv-

```
hdp123@stud-OptiPlex-3060: ~/apache-hive-2.3.9-bin/bin
hive> create table flight(Year int, Month int, DayOfMonth int, DayOfWeek int, DepTime int, CRSDepTime int, ArrTime int, CRSArrTime int, UniqueCarrier string, FlightNum int, TailNum string, ActualElapsedTime int, CRSElapsedTime int, AirTime int, ArrDelay int, DepDelay int, Origin string, Dest string, Distance int, TaxiIn int, TaxiOut int, Cancelled int, CancellationCode string, Diverted string, CarrierDelay int, WeatherDelay int, NASDelay int, SecurityDelay int, LateAircraftDelay int) row format delimited fields terminated by ',';
OK
Time taken: 0.271 seconds
hive> load data local inpath '/home/hdp123/Downloads/flight_data.csv' into table flight;
Loading data to table default.flight
OK
Time taken: 0.246 seconds
hive>
```

For external table flight_ext-

```
hive> create external table flight_ext(Year int, Month int, DayOfMonth int, DayOfWeek int, DepTime int, CRSDepTime int, ArrTime int, CRSArrTime int, UniqueCarrier string, FlightNum int, TailNum string, ActualElapsedTime int, CRSElapsedTime int, AirTime int, ArrDelay int, DepDelay int, Origin string, Dest string, Distance int, TaxiIn int, TaxiOut int, Cancelled int, CancellationCode string, Diverted string, CarrierDelay int, WeatherDelay int, NASDelay int, SecurityDelay int, LateAircraftDelay int) row format delimited fields terminated by ',' location '/user/hive/flight';
OK
Time taken: 0.071 seconds
```

```
hdp123@stud-OptiPlex-3060: ~/apache-hive-2.3.9-bin/bin
hdp123@stud-OptiPlex-3060: ~/apache-hive-2.3.9-bin/bin
stud@stud-OptiPlex-3060: ~$ su hdp123
Password:
hdp123@stud-OptiPlex-3060: /home/stud$ cd
hdp123@stud-OptiPlex-3060: ~$ cd apache-hive-2.3.9-bin/
hdp123@stud-OptiPlex-3060: ~/apache-hive-2.3.9-bin$ cd bin
hdp123@stud-OptiPlex-3060: ~/apache-hive-2.3.9-bin/bin$ hdfs dfs -mkdir /user/hive/flight
23/05/02 15:17:20 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
hdp123@stud-OptiPlex-3060: ~/apache-hive-2.3.9-bin/bin$ hdfs dfs -put /home/stud/Downloads/flight_ext.csv /user/hive/flight
23/05/02 15:17:42 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
put: '/home/stud/Downloads/flight_ext.csv': No such file or directory
hdp123@stud-OptiPlex-3060: ~/apache-hive-2.3.9-bin/bin$ hdfs dfs -put /home/hdp123/Downloads/flight_ext.csv /user/hive/flight
23/05/02 15:19:48 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
hdp123@stud-OptiPlex-3060: ~/apache-hive-2.3.9-bin/bin$ hdfs dfs -ls /user/hive/flight
23/05/02 15:20:02 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 hdp123 supergroup 3842 2023-05-02 15:19 /user/hive/flight/flight_ext.csv
```

Describing table- describe formatter table_name

```
hdp123@stud-OptiPlex-3060: ~/apache-hive-2.3.9-bin/bin
hive> describe formatted emp;
OK
# col_name          data_type          comment
ename               string
esal                 int

# Detailed Table Information
Database:            default
Owner:               hdp123
CreateTime:          Tue May 02 14:58:56 IST 2023
LastAccessTime:      UNKNOWN
Retention:           0
Location:             hdfs://localhost:9000/user/hive/warehouse/emp
Table Type:          MANAGED_TABLE
Table Parameters:
    numFiles          1
    numRows            0
    rawDataSize        0
    totalSize          64
    transient_lastDdlTime 1683020454

# 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: 0.184 seconds, Fetched: 31 row(s)
```

Running query- SELECT COUNT(*) FROM FLIGHT_EXT-

```
hdp123@stud-OptiPlex-3060: ~/apache-hive-2.3.9-bin/bin
hive> select count(*) from flight_ext;
WARNING: 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.
Query ID = hdp123_20230502152237_b5dee5e8-cb68-4093-a42a-5f607f770824
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_1683014159692_0001, Tracking URL = http://stud-OptiPlex-3060:8088/proxy/application_1683014159692_0001/
Kill Command = /usr/local/hadoop/bin/hadoop job -kill job_1683014159692_0001
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2023-05-02 15:22:44,317 Stage-1 map = 0%, reduce = 0%
2023-05-02 15:22:49,521 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 1.54 sec
2023-05-02 15:22:54,653 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 3.24 sec
MapReduce Total cumulative CPU time: 3 seconds 240 msec
Ended Job = job_1683014159692_0001
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1 Reduce: 1 Cumulative CPU: 3.24 sec HDFS Read: 15222 HDFS Write: 102 SUCCESS
Total MapReduce CPU Time Spent: 3 seconds 240 msec
OK
39
Time taken: 18.062 seconds, Fetched: 1 row(s)
```

Creating another table flight_fare and loading data-

```
hdp123@stud-OptiPlex-3060: ~/apache-hive-2.3.9-bin/bin
hive> create table flight_fare( FlightNum int, Fare int) row format delimited fields terminated by ',';
OK
Time taken: 0.059 seconds
hive> load data local inpath '/home/hdp123/Downloads/flight_fare.csv' into table flight_fare;
Loading data to table default.flight_fare
```

Joining tables-

```
hdp123@stud-OptiPlex-3060: ~/apache-hive-2.3.9-bin/bin
FAILED: SemanticException [Error 10001]: Line 1:65 Table not found 'flight_csv'
hive> select flight_ext.Origin, flight_ext.Dest, flight_fare.fare from flight_ext join flight_fare on flight_ext.FlightNum=flight_fare.FlightNum;
WARNING: 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.
Query ID = hdp123_20230502152551_60957f40-2f6e-4197-a82b-eb00f07ae6a1
Total jobs = 1
SLF4J: Class path contains multiple SLF4J bindings.
SLF4J: Found binding in [jar:file:/home/hdp123/apache-hive-2.3.9-bin/lib/log4j-slf4j-impl-2.6.2.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: Found binding in [jar:file:/usr/local/hadoop/share/hadoop/common/lib/slf4j-log4j12-1.7.25.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.apache.logging.slf4j.Log4jLoggerFactory]
2023-05-02 15:25:56 Starting to launch local task to process map join; maximum memory = 477626368
2023-05-02 15:25:56 Dump the side-table for tag: 1 with group count: 47 into file: file:/tmp/hdp123/7274126d-939f-4f07-b10b-5dbd4478c531/h
ive_2023-05-02_15-25-51_788_996744776622609264-1/-local-10004/HashTable-Stage-3/MapJoin-mapfile01--.hashtable
2023-05-02 15:25:56 Uploaded 1 File to: file:/tmp/hdp123/7274126d-939f-4f07-b10b-5dbd4478c531/hive_2023-05-02_15-25-51_788_996744776622609
264-1/-local-10004/HashTable-Stage-3/MapJoin-mapfile01--.hashtable (1357 bytes)
2023-05-02 15:25:56 End of local task; Time Taken: 0.705 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_1683014159692_0002, Tracking URL = http://stud-OptiPlex-3060:8088/proxy/application_1683014159692_0002/
Kill Command = /usr/local/hadoop/bin/hadoop job -kill job_1683014159692_0002
Hadoop job information for Stage-3: number of mappers: 1; number of reducers: 0
2023-05-02 15:26:02,921 Stage-3 map = 0%, reduce = 0%
2023-05-02 15:26:07,004 Stage-3 map = 100%, reduce = 0%, Cumulative CPU 1.92 sec
MapReduce Total cumulative CPU time: 1 seconds 920 msec
Ended Job = job_1683014159692_0002
MapReduce Jobs Launched:
Stage-Stage-3: Map: 1 Cumulative CPU: 1.92 sec HDFS Read: 12233 HDFS Write: 110 SUCCESS
Total MapReduce CPU Time Spent: 1 seconds 920 msec
OK
SEA SMF NULL
Time taken: 16.274 seconds, Fetched: 1 row(s)
hive>
```