

Assignment 3

```
hdp123@stud-OptiPlex-3060: /usr/local/hive/bin
hdp123@stud-OptiPlex-3060:~$ su - hdp123
Password:
hdp123@stud-OptiPlex-3060:~$ start-dfs.sh
23/05/09 15:53:28 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
Starting namenodes on [localhost]
localhost: namenode running as process 12315. Stop it first.
localhost: datanode running as process 12475. Stop it first.
Starting secondary namenodes [0.0.0.0]
0.0.0.0: secondarynamenode running as process 12674. Stop it first.
23/05/09 15:53:30 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
hdp123@stud-OptiPlex-3060:~$ start-yarn.sh
starting yarn daemons
resourcemanager running as process 12826. Stop it first.
localhost: nodemanager running as process 12959. Stop it first.
hdp123@stud-OptiPlex-3060:~$ jps
12674 SecondaryNameNode
15097 Jps
12826 ResourceManager
12475 DataNode
12315 NameNode
12959 NodeManager
hdp123@stud-OptiPlex-3060:~$ cd /usr/local/hive/bin
hdp123@stud-OptiPlex-3060: /usr/local/hive/bin$ ./hive
SLF4J: Class path contains multiple SLF4J bindings.
SLF4J: Found binding in [jar:file:/usr/local/hive/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:/usr/local/hive/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, tez) or using Hive 1.X releases.
hive> show databases;
OK
d1
d157
```

```
hive> show databases;
OK
d1
d157
default
Time taken: 3.144 seconds, Fetched: 3 row(s)
hive> create database demo;
OK
Time taken: 0.195 seconds
```

```
hive> create table emp(ename string,esal int) row format delimited fields terminated by ',' stored as textfile;
FAILED: Execution Error, return code 1 from org.apache.hadoop.hive.q1.exec.DDLTask. AlreadyExistsException(message:Table emp already exists)
hive> create table emp3(ename string,esal int) row format delimited fields terminated by ',' stored as textfile;
OK
Time taken: 0.174 seconds
```

```
hive> create table emp(ename string,esal int) row format delimited fields terminated by ',' stored as textfile;
FAILED: Execution Error, return code 1 from org.apache.hadoop.hive.q1.exec.DDLTask. AlreadyExistsException(message:Table emp already exists)
hive> create table emp3(ename string,esal int) row format delimited fields terminated by ',' stored as textfile;
OK
Time taken: 0.174 seconds
hive> load data local inpath '/home/hdp123/Desktop/sample.txt' into table emp3
```

```
hive> create table flight1(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: 0.089 seconds
hive> load data local inpath '/home/hdp123/Desktop/flight_data.csv' into table flight1;
Loading data to table default.flight1
OK
Time taken: 0.851 seconds
hive>
```

```

hdp123@stud-OptiPlex-3060:~$ start-dfs.sh
23/05/09 17:25:57 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
Starting namenodes on [localhost]
localhost: namenode running as process 12315. Stop it first.
localhost: datanode running as process 12475. Stop it first.
Starting secondary namenodes [0.0.0.0]
0.0.0.0: secondarynamenode running as process 12674. Stop it first.
23/05/09 17:25:59 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
hdp123@stud-OptiPlex-3060:~$ start-yarn.sh
starting yarn daemons
resourcemanager running as process 12826. Stop it first.
localhost: nodemanager running as process 12959. Stop it first.
hdp123@stud-OptiPlex-3060:~$ jps
12674 SecondaryNameNode
18723 Jps
15128 RunJar
12826 ResourceManager
12475 DataNode
12315 NameNode
12959 NodeManager
hdp123@stud-OptiPlex-3060:~$ hdfs dfs -mkdir /usr/hive/flight1
23/05/09 17:26:20 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
mkdir: '/usr/hive/flight1': File exists
hdp123@stud-OptiPlex-3060:~$ hdfs dfs -put /home/hdp123/Documents/Assignment_hive/flight_ext.csv /usr/hive/flight1
23/05/09 17:26:38 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
hdp123@stud-OptiPlex-3060:~$ hdfs dfs -ls /usr/hive/flight1
23/05/09 17:26:48 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 2023-05-09 17:26 /usr/hive/flight1/flight_ext.csv
hdp123@stud-OptiPlex-3060:~$

```

```

hive> create external table flight_ext1(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 '/usr/hive/flight1';
OK
Time taken: 0.071 seconds
hive> describe formatted emp3;
OK
# col_name          data_type          comment
ename               string
esal                int

# Detailed Table Information
Database:            default
Owner:               hdp123
CreateTime:          Tue May 09 16:04:19 IST 2023
LastAccessTime:      UNKNOWN
Retention:           0
Location:             hdfs://localhost:9000/user/hive/warehouse/emp3
Table Type:          MANAGED_TABLE
Table Parameters:
  COLUMN_STATS_ACCURATE {\"BASIC_STATS\": \"true\"}
  numFiles               0
  numRows               0
  rawDataSize           0
  totalSize             0
  transient_lastDdlTime 1683628459

# Storage Information
SerDe Library:       org.apache.hadoop.hive.serde2.lazy.LazySimpleSerDe
InputFormat:         org.apache.hadoop.mapred.TextInputFormat

```

```
hdp123@stud-OptiPlex-3060: /usr/local/hive/bin

# 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: 0.175 seconds, Fetched: 32 row(s)
hive> describe formatted flight_ext1;
OK
# col_name          data_type          comment
year                int
month               int
dayofmonth           int
dayofweek            int
deptime             int
crsdeptime           int
arrtime             int
crsarrrtime         int
uniquecarrier        string
flightnum            int
tailnum             string
actualelapsedtime    int
crselapsedtime       int
airtime             int
arrdelay            int
depdelay            int
origin              string
dest                string
distance            int
```

```
hdp123@stud-OptiPlex-3060: /usr/local/hive/bin
hdp123@stud-OptiPlex-3060: ~

diverted            string
carrierdelay        int
weatherdelay        int
nasdelay            int
securitydelay        int
lateaircraftdelay   int

# Detailed Table Information
Database:            default
Owner:               hdp123
CreateTime:          Tue May 09 17:33:08 IST 2023
LastAccessTime:      UNKNOWN
Retention:           0
Location:            hdfs://localhost:9000/usr/hive/flight1
Table Type:          EXTERNAL_TABLE
Table Parameters:
    EXTERNAL          TRUE
    numFiles          1
    totalSize         3842
    transient_lastDdlTime 1683633788

# 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: 0.04 seconds, Fetched: 57 row(s)
hive> select count(*) from flight_ext1;
```

```

hive> select count(*) from flight_ext1;
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_20230509173445_c318c5eb-b3c2-41ce-9810-6f8424d7f25b
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_1683627453529_0001, Tracking URL = http://stud-OptiPlex-3060:8088/proxy/application_1683627453529_0001/
Kill Command = /usr/local/hadoop/bin/hadoop job -kill job_1683627453529_0001
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2023-05-09 17:34:51,391 Stage-1 map = 0%, reduce = 0%
2023-05-09 17:34:54,476 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 1.24 sec
2023-05-09 17:34:58,538 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 2.61 sec
MapReduce Total cumulative CPU time: 2 seconds 610 msec
Ended Job = job_1683627453529_0001
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1 Reduce: 1 Cumulative CPU: 2.61 sec HDFS Read: 15227 HDFS Write: 102 SUCCESS
Total MapReduce CPU Time Spent: 2 seconds 610 msec
OK
39
Time taken: 14.613 seconds, Fetched: 1 row(s)
hive> create table flight_fare( FlightNumint, Fare int) row format delimited fields terminated by ',';
NoViableAltException(9@[])
at org.apache.hadoop.hiveql.parser.HiveParser.type(HiveParser.java:32847)

```

```

hive> create table flight_fare1( FlightNum int, Fare int) row format delimited fields terminated by ',';
OK
Time taken: 0.068 seconds
hive> load data local inpath '/home/hdp123/Documents/Assignment_hive/flight_fare.csv' into table flight_fare1;
Loading data to table default.flight_fare1
OK
Time taken: 0.619 seconds

```

```

hive> select flight1.origin, flight1.dest, flight_fare1.fare from flight1 join flight_fare1 on flight1.FlightNum=flight_fare1.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_20230509175348_ef39754e-8e5d-4b59-a01b-a494b1c35c37
Total jobs = 1
SLF4J: Class path contains multiple SLF4J bindings.
SLF4J: Found binding in [jar:file:/usr/local/hive/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]
2023-05-09 17:53:52 Starting to launch local task to process map join; maximum memory = 477626368
2023-05-09 17:53:53 Dump the side-table for tag: 1 with group count: 47 into file: file:/tmp/hdp123/038125d3-a2db-489c-bf84-b33019c667b6/h
ive_2023-05-09_17-53-48_163_5551306870433688156-1/-local-10004/HashTable-Stage-3/MapJoin-mapfile01--.hashtable
2023-05-09 17:53:53 Uploaded 1 File to: file:/tmp/hdp123/038125d3-a2db-489c-bf84-b33019c667b6/hive_2023-05-09_17-53-48_163_555130687043368
8156-1/-local-10004/HashTable-Stage-3/MapJoin-mapfile01--.hashtable (1357 bytes)
2023-05-09 17:53:53 End of local task; Time Taken: 0.675 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_1683627453529_0002, Tracking URL = http://stud-OptiPlex-3060:8088/proxy/application_1683627453529_0002/
Kill Command = /usr/local/hadoop/bin/hadoop job -kill job_1683627453529_0002
Hadoop job information for Stage-3: number of mappers: 1; number of reducers: 0
2023-05-09 17:54:00,036 Stage-3 map = 0%, reduce = 0%
2023-05-09 17:54:03,143 Stage-3 map = 100%, reduce = 0%, Cumulative CPU 1.93 sec
MapReduce Total cumulative CPU time: 1 seconds 930 msec
Ended Job = job_1683627453529_0002
MapReduce Jobs Launched:
Stage-Stage-3: Map: 1 Cumulative CPU: 1.93 sec HDFS Read: 13322 HDFS Write: 1317 SUCCESS
Total MapReduce CPU Time Spent: 1 seconds 930 msec
OK
SEA SJC 5181
SEA PSP 7905
SAN SEA 8781
SEA GEG 7852

```

```

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_1683627453529_0002, Tracking URL = http://stud-OptiPlex-3060:8088/proxy/application_1683627453529_0002/
Kill Command = /usr/local/hadoop/bin/hadoop job -kill job_1683627453529_0002
Hadoop job information for Stage-3: number of mappers: 1; number of reducers: 0
2023-05-09 17:54:00,036 Stage-3 map = 0%, reduce = 0%
2023-05-09 17:54:03,143 Stage-3 map = 100%, reduce = 0%, Cumulative CPU 1.93 sec
MapReduce Total cumulative CPU time: 1 seconds 930 msec
Ended Job = job_1683627453529_0002
MapReduce Jobs Launched:
Stage-Stage-3: Map: 1 Cumulative CPU: 1.93 sec HDFS Read: 13322 HDFS Write: 1317 SUCCESS
Total MapReduce CPU Time Spent: 1 seconds 930 msec
OK
SEA SJC 5181
SEA PSP 7905
SAN SEA 8781
SEA GEG 7852
TUS SEA 9009
LAX SEA 6250
LAX SEA 8535
ANC PDX 7141
LAS SEA 9184
SJC SEA 7997
SEA DEN 9245
ANC SEA 5897
SEA TUS 8321
SEA SAN 8446
SEA SAN 9338
SEA ANC 5785
SEA LAS 5807
SEA LAX 6358
SEA PSP 5434
LAX SEA 9479

```

```

SEA LAX 6358
SEA PSP 5434
LAX SEA 9479
LAX SEA 7158
ANC SEA 6297
DFW SEA 5715
SEA LAX 8082
DEN SEA 6825
SFO SEA 9186
GEG SEA 7127
ANC FAI 6404
SMF SEA 7503
SEA PSP 7247
SAN PDX 6992
LAX SEA 9505
DEN SEA 9582
SFO PSP 7465
SEA SNA 9400
JNU KTN 7573
PDX OAK 9908
KTN SIT 9515
KTN SIT NULL
SNA PDX 8962
SEA OAK NULL
SFO SEA NULL
SNA SEA NULL
SEA LAX NULL
SEA ONT NULL
SEA KTN 9515
SEA KTN NULL
PDX PSP NULL
SEA SAN NULL
JNU SIT NULL
Time taken: 17.058 seconds, Fetched: 50 row(s)
hive>

```