Joe Griffin
Week 7 Homework
Dec 16, 2021

The ORC and parquet files are much smaller than the CSV file. Over 10x smaller.

```
bitnami@debian:~$ hdfs dfs -du -h /user/hive/warehouse
54.2 M   54.2 M   /user/hive/warehouse/q1_table
1.6 M    1.6 M    /user/hive/warehouse/q1_table_orc
3.1 M    3.1 M    /user/hive/warehouse/q1_table_parquet
```

```
[bitnami@debian:/tmp$ sed '1d' feb.csv > febnh.csv
[bitnami@debian:/tmp$ sed '1d' mar.csv > marnh.csv
[bitnami@debian:/tmp$ head jan.csv
 "YEAR","MONTH","DAY_OF_MONTH","ORIGIN","DEST","DEP_DEL15","ARR_DEL15",
 2019,1,16,"BOS","PHL",0.00,0.00,
 2019,1,17,"BOS","PHL",0.00,0.00,
 2019,1,18,"BOS","PHL",0.00,0.00,
 2019,1,19,"BOS","PHL",0.00,0.00,
 2019,1,20,"BOS","PHL",,,
 2019,1,21,"BOS","PHL",,,
 2019,1,22,"BOS","PHL",,,
 2019,1,23,"BOS","PHL",0.00,0.00,
 2019,1,24,"BOS","PHL",0.00,0.00,
[bitnami@debian:/tmp$ head febnh.csv
 2019,2,24,"BNA","DCA",0.00,0.00,
 2019,2,24,"BNA","DCA",1.00,1.00,
 2019,2,24,"BNA","DCA",1.00,1.00,
 2019,2,24,"BNA","DCA",0.00,0.00,
 2019,2,24,"BNA","DEN",0.00,0.00,
 2019,2,24,"BNA","DEN",0.00,0.00,
 2019,2,24,"BNA","DEN",0.00,0.00,
 2019,2,24,"BNA","DEN",0.00,0.00,
 2019,2,24,"BNA","DTW",0.00,0.00,
 2019,2,24,"BNA","DTW",,,
[bitnami@debian:/tmp$ head marnh.csv
 2019,3,2,"ATL","MCI",0.00,0.00,
 2019,3,2,"ATL","MCI",0.00,0.00,
 2019,3,2,"ATL","MCO",1.00,1.00,
 2019,3,2,"ATL","MCO",0.00,0.00,
 2019,3,2,"ATL","MCO",0.00,0.00,
 2019,3,2,"ATL","MCO",0.00,0.00,
 2019,3,2,"ATL","MCO",,,
 2019,3,2,"ATL","MCO",0.00,0.00,
 2019,3,2,"ATL","MDW",0.00,0.00,
 2019,3,2,"ATL","MDW",1.00,1.00,
[bitnami@debian:/tmp$ cat jan.csv febnh.csv marnh.csv > q1.csv
[bitnami@debian:/tmp$ head q1.csv
 "YEAR","MONTH","DAY_OF_MONTH","ORIGIN","DEST","DEP_DEL15","ARR_DEL15",
 2019,1,16,"BOS","PHL",0.00,0.00,
 2019,1,17,"BOS","PHL",0.00,0.00,
 2019,1,18,"BOS","PHL",0.00,0.00,
 2019,1,19,"BOS","PHL",0.00,0.00,
 2019,1,20,"BOS","PHL",,,
 2019,1,21,"BOS","PHL",,,
 2019,1,22,"BOS","PHL",,,
 2019,1,23,"BOS","PHL",0.00,0.00,
 2019,1,24,"BOS","PHL",0.00,0.00,
[bitnami@debian:/tmp$ tail q1.csv
 2019,3,6,"ORD","CLT",0.00,0.00,
 2019,3,6,"CVG","ORD",0.00,0.00,
 2019,3,6,"DTW","ORD",0.00,0.00,
 2019,3,6,"DTW","EWR",1.00,1.00,
 2019,3,6,"OMA","EWR",1.00,1.00,
 2019,3,6,"DTW","EWR",0.00,0.00,
 2019,3,7,"ATL","ORD",0.00,0.00,
 2019,3,7,"CLT","EWR",0.00,0.00,
 2019,3,7,"ABQ","DEN",0.00,0.00,
 2019,3,7,"CVG","ORD",0.00,0.00,
bitnami@debian:/tmp$
```

```
[0: jdbc:hive2://localhost:10000/> create table q1_data (YEAR int, MONTH int, DAY int, AIRLINE_ID int, ORIGIN string, DEST string, DEL15 st]
ring, ARR15 string, cancelled string) row format delimited fields terminated by ',' stored as textfile;
INFO  : Compiling command(queryId=hadoop_20211216050936_10c76385-ad48-4d3b-a2dc-ea5575f05b09): create table q1_data (YEAR int, MONTH int,
DAY int, AIRLINE_ID int, ORIGIN string, DEST string, DEL15 string, ARR15 string, cancelled string) row format delimited fields terminated
by ',' stored as textfile
INFO  : Concurrency mode is disabled, not creating a lock manager
INFO  : Semantic Analysis Completed (retrial = false)
INFO  : Returning Hive schema: Schema(fieldSchemas:null, properties:null)
INFO  : Completed compiling command(queryId=hadoop_20211216050936_10c76385-ad48-4d3b-a2dc-ea5575f05b09); Time taken: 0.186 seconds
INFO  : Concurrency mode is disabled, not creating a lock manager
INFO  : Executing command(queryId=hadoop_20211216050936_10c76385-ad48-4d3b-a2dc-ea5575f05b09): create table q1_data (YEAR int, MONTH int,
DAY int, AIRLINE_ID int, ORIGIN string, DEST string, DEL15 string, ARR15 string, cancelled string) row format delimited fields terminated
by ',' stored as textfile
INFO  : Starting task [Stage-0:DDL] in serial mode
INFO  : Completed executing command(queryId=hadoop_20211216050936_10c76385-ad48-4d3b-a2dc-ea5575f05b09); Time taken: 1.839 seconds
INFO  : OK
INFO  : Concurrency mode is disabled, not creating a lock manager
No rows affected (2.058 seconds)
[0: jdbc:hive2://localhost:10000/> describe q1_data                                                                                      ]
[. . . . . . . . . . . . . . . . > describe q1_data;                                                                                     ]
Error: Error while compiling statement: FAILED: ParseException line 2:0 cannot recognize input near 'q1_data' 'describe' 'q1_data' in tabl
e name (state=42000,code=40000)
[0: jdbc:hive2://localhost:10000/> describe q1_data;                                                                                     ]
INFO  : Compiling command(queryId=hadoop_20211216051121_b140ed61-f7b9-44e6-9217-0673132f06e3): describe q1_data
INFO  : Concurrency mode is disabled, not creating a lock manager
INFO  : Semantic Analysis Completed (retrial = false)
INFO  : Returning Hive schema: Schema(fieldSchemas:[FieldSchema(name:col_name, type:string, comment:from deserializer), FieldSchema(name:d
ata_type, type:string, comment:from deserializer), FieldSchema(name:comment, type:string, comment:from deserializer)], properties:null)
INFO  : Completed compiling command(queryId=hadoop_20211216051121_b140ed61-f7b9-44e6-9217-0673132f06e3); Time taken: 0.83 seconds
INFO  : Concurrency mode is disabled, not creating a lock manager
INFO  : Executing command(queryId=hadoop_20211216051121_b140ed61-f7b9-44e6-9217-0673132f06e3): describe q1_data
INFO  : Starting task [Stage-0:DDL] in serial mode
INFO  : Completed executing command(queryId=hadoop_20211216051121_b140ed61-f7b9-44e6-9217-0673132f06e3); Time taken: 0.134 seconds
INFO  : OK
INFO  : Concurrency mode is disabled, not creating a lock manager
+------------+------------+----------+
|  col_name  | data_type  | comment  |
+------------+------------+----------+
| year       | int        |          |
| month      | int        |          |
| day        | int        |          |
| airline_id | int        |          |
| origin     | string     |          |
| dest       | string     |          |
| del15      | string     |          |
| arr15      | string     |          |
| cancelled  | string     |          |
+------------+------------+----------+
9 rows selected (1.03 seconds)
```

```
[0: jdbc:hive2://localhost:10000/> create table q1_data (YEAR int, MONTH int, DAY int, AIRLINE_ID int, ORIGIN string, DEST string, DEL15 st]
ring, ARR15 string, cancelled string) row format delimited fields terminated by ',' stored as textfile;
INFO  : Compiling command(queryId=hadoop_20211216050936_10c76385-ad48-4d3b-a2dc-ea5575f05b09): create table q1_data (YEAR int, MONTH int,
DAY int, AIRLINE_ID int, ORIGIN string, DEST string, DEL15 string, ARR15 string, cancelled string) row format delimited fields terminated
by ',' stored as textfile
INFO  : Concurrency mode is disabled, not creating a lock manager
INFO  : Semantic Analysis Completed (retrial = false)
INFO  : Returning Hive schema: Schema(fieldSchemas:null, properties:null)
INFO  : Completed compiling command(queryId=hadoop_20211216050936_10c76385-ad48-4d3b-a2dc-ea5575f05b09); Time taken: 0.186 seconds
INFO  : Concurrency mode is disabled, not creating a lock manager
INFO  : Executing command(queryId=hadoop_20211216050936_10c76385-ad48-4d3b-a2dc-ea5575f05b09): create table q1_data (YEAR int, MONTH int,
DAY int, AIRLINE_ID int, ORIGIN string, DEST string, DEL15 string, ARR15 string, cancelled string) row format delimited fields terminated
by ',' stored as textfile
INFO  : Starting task [Stage-0:DDL] in serial mode
INFO  : Completed executing command(queryId=hadoop_20211216050936_10c76385-ad48-4d3b-a2dc-ea5575f05b09); Time taken: 1.839 seconds
INFO  : OK
INFO  : Concurrency mode is disabled, not creating a lock manager
No rows affected (2.058 seconds)
[0: jdbc:hive2://localhost:10000/> describe q1_data                                                                                      ]
[. . . . . . . . . . . . . . . . > describe q1_data;                                                                                     ]
Error: Error while compiling statement: FAILED: ParseException line 2:0 cannot recognize input near 'q1_data' 'describe' 'q1_data' in tabl
e name (state=42000,code=40000)
[0: jdbc:hive2://localhost:10000/> describe q1_data;                                                                                     ]
INFO  : Compiling command(queryId=hadoop_20211216051121_b140ed61-f7b9-44e6-9217-0673132f06e3): describe q1_data
INFO  : Concurrency mode is disabled, not creating a lock manager
INFO  : Semantic Analysis Completed (retrial = false)
INFO  : Returning Hive schema: Schema(fieldSchemas:[FieldSchema(name:col_name, type:string, comment:from deserializer), FieldSchema(name:d
ata_type, type:string, comment:from deserializer), FieldSchema(name:comment, type:string, comment:from deserializer)], properties:null)
INFO  : Completed compiling command(queryId=hadoop_20211216051121_b140ed61-f7b9-44e6-9217-0673132f06e3); Time taken: 0.83 seconds
INFO  : Concurrency mode is disabled, not creating a lock manager
INFO  : Executing command(queryId=hadoop_20211216051121_b140ed61-f7b9-44e6-9217-0673132f06e3): describe q1_data
INFO  : Starting task [Stage-0:DDL] in serial mode
INFO  : Completed executing command(queryId=hadoop_20211216051121_b140ed61-f7b9-44e6-9217-0673132f06e3); Time taken: 0.134 seconds
INFO  : OK
INFO  : Concurrency mode is disabled, not creating a lock manager
+------------+------------+----------+
|  col_name  | data_type  | comment  |
+------------+------------+----------+
| year       | int        |          |
| month      | int        |          |
| day        | int        |          |
| airline_id | int        |          |
| origin     | string     |          |
| dest       | string     |          |
| del15      | string     |          |
| arr15      | string     |          |
| cancelled  | string     |          |
+------------+------------+----------+
9 rows selected (1.03 seconds)
```

```
0: jdbc:hive2://localhost:10000/> create table q1_table (YEAR int, MONTH int, DAY int, ORIGIN string, DEST string, DEP_DEL15 string) row form
at delimited fields terminated by ',' stored as textfile;
INFO  : Compiling command(queryId=hadoop_20211216053231_526dbe65-9784-4be1-9a62-f8bd5e77be11): create table q1_table (YEAR int, MONTH int, DA
Y int, ORIGIN string, DEST string, DEP_DEL15 string) row format delimited fields terminated by ',' stored as textfile
INFO  : Concurrency mode is disabled, not creating a lock manager
INFO  : Semantic Analysis Completed (retrial = false)
INFO  : Returning Hive schema: Schema(fieldSchemas:null, properties:null)
INFO  : Completed compiling command(queryId=hadoop_20211216053231_526dbe65-9784-4be1-9a62-f8bd5e77be11); Time taken: 0.038 seconds
INFO  : Concurrency mode is disabled, not creating a lock manager
INFO  : Executing command(queryId=hadoop_20211216053231_526dbe65-9784-4be1-9a62-f8bd5e77be11): create table q1_table (YEAR int, MONTH int, DA
Y int, ORIGIN string, DEST string, DEP_DEL15 string) row format delimited fields terminated by ',' stored as textfile
INFO  : Starting task [Stage-0:DDL] in serial mode
INFO  : Completed executing command(queryId=hadoop_20211216053231_526dbe65-9784-4be1-9a62-f8bd5e77be11); Time taken: 0.169 seconds
INFO  : OK
INFO  : Concurrency mode is disabled, not creating a lock manager
No rows affected (0.226 seconds)
0: jdbc:hive2://localhost:10000/> load data local inpath '/tmp/q1nh.csv' overwrite into table q1_table;
INFO  : Compiling command(queryId=hadoop_20211216053257_6587d656-e405-4d80-87d2-502e636b525a): load data local inpath '/tmp/q1nh.csv' overwri
te into table q1_table
INFO  : Concurrency mode is disabled, not creating a lock manager
INFO  : Semantic Analysis Completed (retrial = false)
INFO  : Returning Hive schema: Schema(fieldSchemas:null, properties:null)
INFO  : Completed compiling command(queryId=hadoop_20211216053257_6587d656-e405-4d80-87d2-502e636b525a); Time taken: 0.095 seconds
INFO  : Concurrency mode is disabled, not creating a lock manager
INFO  : Executing command(queryId=hadoop_20211216053257_6587d656-e405-4d80-87d2-502e636b525a): load data local inpath '/tmp/q1nh.csv' overwri
te into table q1_table
INFO  : Starting task [Stage-0:MOVE] in serial mode
INFO  : Loading data to table default.q1_table from file:/tmp/q1nh.csv
INFO  : Starting task [Stage-1:STATS] in serial mode
INFO  : Completed executing command(queryId=hadoop_20211216053257_6587d656-e405-4d80-87d2-502e636b525a); Time taken: 1.72 seconds
INFO  : OK
INFO  : Concurrency mode is disabled, not creating a lock manager
No rows affected (1.838 seconds)
0: jdbc:hive2://localhost:10000/> select * from q1_table limit 10;
INFO  : Compiling command(queryId=hadoop_20211216053358_8ff3c899-142e-4173-91e9-b1e321e94888): select * from q1_table limit 10
INFO  : Concurrency mode is disabled, not creating a lock manager
INFO  : Semantic Analysis Completed (retrial = false)
INFO  : Returning Hive schema: Schema(fieldSchemas:[FieldSchema(name:q1_table.year, type:int, comment:null), FieldSchema(name:q1_table.month,
 type:int, comment:null), FieldSchema(name:q1_table.day, type:int, comment:null), FieldSchema(name:q1_table.origin, type:string, comment:null
), FieldSchema(name:q1_table.dest, type:string, comment:null), FieldSchema(name:q1_table.dep_del15, type:string, comment:null)], properties:n
ull)
INFO  : Completed compiling command(queryId=hadoop_20211216053358_8ff3c899-142e-4173-91e9-b1e321e94888); Time taken: 0.365 seconds
INFO  : Concurrency mode is disabled, not creating a lock manager
INFO  : Executing command(queryId=hadoop_20211216053358_8ff3c899-142e-4173-91e9-b1e321e94888): select * from q1_table limit 10
INFO  : Completed executing command(queryId=hadoop_20211216053358_8ff3c899-142e-4173-91e9-b1e321e94888); Time taken: 0.008 seconds
INFO  : OK
INFO  : Concurrency mode is disabled, not creating a lock manager
```

q1_table.year	q1_table.month	q1_table.day	q1_table.origin	q1_table.dest	q1_table.dep_del15
2019	1	16	"BOS"	"PHL"	0.00
2019	1	17	"BOS"	"PHL"	0.00
2019	1	18	"BOS"	"PHL"	0.00
2019	1	19	"BOS"	"PHL"	0.00
2019	1	20	"BOS"	"PHL"	
2019	1	21	"BOS"	"PHL"	
2019	1	22	"BOS"	"PHL"	
2019	1	23	"BOS"	"PHL"	0.00
2019	1	24	"BOS"	"PHL"	0.00
2019	1	25	"BOS"	"PHL"	0.00

```
0: jdbc:hive2://localhost:10000/> create table q1_table_orc (YEAR int, MONTH int, DAY int, ORIGIN string, DEST string, DEP_DEL15 string) row
format delimited fields terminated by ',' stored as ORC;
INFO  : Compiling command(queryId=hadoop_20211216053628_7dfe2711-ec14-45cb-8208-c72f52629617): create table q1_table_orc (YEAR int, MONTH int
, DAY int, ORIGIN string, DEST string, DEP_DEL15 string) row format delimited fields terminated by ',' stored as ORC
INFO  : Concurrency mode is disabled, not creating a lock manager
INFO  : Semantic Analysis Completed (retrial = false)
INFO  : Returning Hive schema: Schema(fieldSchemas:null, properties:null)
INFO  : Completed compiling command(queryId=hadoop_20211216053628_7dfe2711-ec14-45cb-8208-c72f52629617); Time taken: 0.045 seconds
INFO  : Concurrency mode is disabled, not creating a lock manager
INFO  : Executing command(queryId=hadoop_20211216053628_7dfe2711-ec14-45cb-8208-c72f52629617): create table q1_table_orc (YEAR int, MONTH int
, DAY int, ORIGIN string, DEST string, DEP_DEL15 string) row format delimited fields terminated by ',' stored as ORC
INFO  : Starting task [Stage-0:DDL] in serial mode
INFO  : Completed executing command(queryId=hadoop_20211216053628_7dfe2711-ec14-45cb-8208-c72f52629617); Time taken: 0.131 seconds
INFO  : OK
INFO  : Concurrency mode is disabled, not creating a lock manager
No rows affected (0.198 seconds)
0: jdbc:hive2://localhost:10000/> create table q1_table_parquet (YEAR int, MONTH int, DAY int, ORIGIN string, DEST string, DEP_DEL15 string)
row format delimited fields terminated by ',' stored as parquet;
INFO  : Compiling command(queryId=hadoop_20211216053714_a96a234c-e745-47b2-9454-32b11aa319bf): create table q1_table_parquet (YEAR int, MONTH
 int, DAY int, ORIGIN string, DEST string, DEP_DEL15 string) row format delimited fields terminated by ',' stored as parquet
INFO  : Concurrency mode is disabled, not creating a lock manager
INFO  : Semantic Analysis Completed (retrial = false)
INFO  : Returning Hive schema: Schema(fieldSchemas:null, properties:null)
INFO  : Completed compiling command(queryId=hadoop_20211216053714_a96a234c-e745-47b2-9454-32b11aa319bf); Time taken: 0.049 seconds
INFO  : Concurrency mode is disabled, not creating a lock manager
INFO  : Executing command(queryId=hadoop_20211216053714_a96a234c-e745-47b2-9454-32b11aa319bf): create table q1_table_parquet (YEAR int, MONTH
 int, DAY int, ORIGIN string, DEST string, DEP_DEL15 string) row format delimited fields terminated by ',' stored as parquet
INFO  : Starting task [Stage-0:DDL] in serial mode
INFO  : Completed executing command(queryId=hadoop_20211216053714_a96a234c-e745-47b2-9454-32b11aa319bf); Time taken: 0.11 seconds
INFO  : OK
INFO  : Concurrency mode is disabled, not creating a lock manager
No rows affected (0.192 seconds)
0: jdbc:hive2://localhost:10000/> show tables;
INFO  : Compiling command(queryId=hadoop_20211216054221_807b9556-de9a-48a7-a421-454c75d46d94): show tables
INFO  : Concurrency mode is disabled, not creating a lock manager
INFO  : Semantic Analysis Completed (retrial = false)
INFO  : Returning Hive schema: Schema(fieldSchemas:[FieldSchema(name:tab_name, type:string, comment:from deserializer)], properties:null)
INFO  : Completed compiling command(queryId=hadoop_20211216054221_807b9556-de9a-48a7-a421-454c75d46d94); Time taken: 0.05 seconds
INFO  : Concurrency mode is disabled, not creating a lock manager
INFO  : Executing command(queryId=hadoop_20211216054221_807b9556-de9a-48a7-a421-454c75d46d94): show tables
INFO  : Starting task [Stage-0:DDL] in serial mode
INFO  : Completed executing command(queryId=hadoop_20211216054221_807b9556-de9a-48a7-a421-454c75d46d94); Time taken: 0.029 seconds
INFO  : OK
INFO  : Concurrency mode is disabled, not creating a lock manager
+-------------------+
|     tab_name      |
+-------------------+
| q1_table          |
| q1_table_orc      |
| q1_table_parquet  |
+-------------------+
```

```
        > insert into table q1_table_orc  select * from q1_table;
Automatically selecting local only mode for query
Query ID = root_20211216055517_7cf849f8-0266-4a0d-a344-d1f6a45d4be9
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>
Job running in-process (local Hadoop)
2021-12-16 05:55:25,911 Stage-1 map = 0%,  reduce = 0%
2021-12-16 05:55:41,318 Stage-1 map = 100%,  reduce = 100%
Ended Job = job_local1623538176_0001
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:8020/user/hive/warehouse/q1_table_orc/.h
ive-staging_hive_2021-12-16_05-55-17_943_5102688540634637548-1/-ext-10000
Loading data to table default.q1_table_orc
MapReduce Jobs Launched:
Stage-Stage-1:  HDFS Read: 113682032 HDFS Write: 3343457 SUCCESS
Total MapReduce CPU Time Spent: 0 msec
OK
Time taken: 25.899 seconds
hive> select * from q1_table_orc limit 10;
OK
2019    1       16      "BOS"   "PHL"   0.00
2019    1       17      "BOS"   "PHL"   0.00
2019    1       18      "BOS"   "PHL"   0.00
2019    1       19      "BOS"   "PHL"   0.00
2019    1       20      "BOS"   "PHL"
2019    1       21      "BOS"   "PHL"
2019    1       22      "BOS"   "PHL"
2019    1       23      "BOS"   "PHL"   0.00
2019    1       24      "BOS"   "PHL"   0.00
2019    1       25      "BOS"   "PHL"   0.00
Time taken: 0.401 seconds, Fetched: 10 row(s)
hive> insert into table q1_table_parquet  select * from q1_table;
Automatically selecting local only mode for query
Query ID = root_20211216055710_25418f2d-5d5f-4705-9596-6432d7e203eb
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>
Job running in-process (local Hadoop)
2021-12-16 05:57:13,951 Stage-1 map = 0%,  reduce = 0%
2021-12-16 05:57:30,172 Stage-1 map = 100%,  reduce = 100%
Ended Job = job_local50877962_0002
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:8020/user/hive/warehouse/q1_table_parque
t/.hive-staging_hive_2021-12-16_05-57-10_503_6664825315255136668-1/-ext-10000
Loading data to table default.q1_table_parquet
MapReduce Jobs Launched:
Stage-Stage-1:  HDFS Read: 230758294 HDFS Write: 9948353 SUCCESS
Total MapReduce CPU Time Spent: 0 msec
OK
Time taken: 20.849 seconds
hive> select * from q1_table_parquet limit 10;
OK
2019    1       16      "BOS"   "PHL"   0.00
2019    1       17      "BOS"   "PHL"   0.00
2019    1       18      "BOS"   "PHL"   0.00
2019    1       19      "BOS"   "PHL"   0.00
2019    1       20      "BOS"   "PHL"
2019    1       21      "BOS"   "PHL"
2019    1       22      "BOS"   "PHL"
2019    1       23      "BOS"   "PHL"   0.00
2019    1       24      "BOS"   "PHL"   0.00
2019    1       25      "BOS"   "PHL"   0.00
Time taken: 0.412 seconds, Fetched: 10 row(s)
hive> select * from q1_table limit 10;
OK
2019    1       16      "BOS"   "PHL"   0.00
2019    1       17      "BOS"   "PHL"   0.00
2019    1       18      "BOS"   "PHL"   0.00
2019    1       19      "BOS"   "PHL"   0.00
2019    1       20      "BOS"   "PHL"
2019    1       21      "BOS"   "PHL"
2019    1       22      "BOS"   "PHL"
2019    1       23      "BOS"   "PHL"   0.00
2019    1       24      "BOS"   "PHL"   0.00
2019    1       25      "BOS"   "PHL"   0.00
Time taken: 0.304 seconds, Fetched: 10 row(s)
```