Combine Hive Screenshots

1. Put Files On HDFS

```
[jc8017@login-2-1 combine]$ hdfs dfs -mkdir tax_final
[jc8017@login-2-1 combine]$ hdfs dfs -mkdir transport_final
[jc8017@login-2-1 combine]$ hdfs dfs -mkdir water_final
[jc8017@login-2-1 combine]$ cd final_input
[jc8017@login-2-1 final_input]$ hdfs dfs -put taxdata.csv tax_final
[jc8017@login-2-1 final_input]$ hdfs dfs -put transportationdata.csv transport_final
[jc8017@login-2-1 final_input]$ hdfs dfs -put waterdata.csv water_final
```

```
[jc8017@login-2-1 final_input]$ hdfs dfs -ls tax_final
Found 1 items
-rw-r-----+   3 jc8017 users      339507 2020-11-28 17:14 tax_final/taxdata.csv
```

```
[jc8017@login-2-1 final_input]$ hdfs dfs -ls transport_final
Found 1 items
-rw-r-----+   3 jc8017 users      250359 2020-11-28 17:14 transport_final/transportationdata
.csv
[jc8017@login-2-1 final input]$
```

```
[jc8017@login-2-1 final_input]$ hdfs dfs -ls water_final
Found 1 items
-rw-r-----+   3 jc8017 users      168372 2020-11-28 17:14 water_final/waterdata.csv
[jc8017@login-2-1 final input]$
```

2. Go onto hive

beeline

```
[jc8017@login-2-1 final_input]$ beeline
Java HotSpot(TM) 64-Bit Server VM warning: ignoring option MaxPermSize=512M; support was removed in 8
.0
Java HotSpot(TM) 64-Bit Server VM warning: Using incremental CMS is deprecated and will likely be rem
oved in a future release
Java HotSpot(TM) 64-Bit Server VM warning: ignoring option MaxPermSize=512M; support was removed in 8
.0
Beeline version 1.1.0-cdh5.15.2 by Apache Hive
```

!connect jdbc:hive2://babar.es.its.nyu.edu:10000/

```
beeline> !connect jdbc:hive2://babar.es.its.nyu.edu:10000/;
scan complete in 1ms
Connecting to jdbc:hive2://babar.es.its.nyu.edu:10000/;
Enter username for jdbc:hive2://babar.es.its.nyu.edu:10000/:: jc8017
Enter password for jdbc:hive2://babar.es.its.nyu.edu:10000/:: ***********
Connected to: Apache Hive (version 1.1.0-cdh5.15.2)
Driver: Hive JDBC (version 1.1.0-cdh5.15.2)
Transaction isolation: TRANSACTION_REPEATABLE_READ
```

Use *net id;*

```
0: jdbc:hive2://babar.es.its.nyu.edu:10000/> use jc8017;
INFO  : Compiling command(queryId=hive_20201128171818_c60c18ec-740c-4f73-8290-8b54b4805f35): use jc80
17
INFO  : Semantic Analysis Completed
INFO  : Returning Hive schema: Schema(fieldSchemas:null, properties:null)
INFO  : Completed compiling command(queryId=hive_20201128171818_c60c18ec-740c-4f73-8290-8b54b4805f35)
; Time taken: 0.285 seconds
INFO  : Executing command(queryId=hive_20201128171818_c60c18ec-740c-4f73-8290-8b54b4805f35): use jc80
17
INFO  : Starting task [Stage-0:DDL] in serial mode
INFO  : Completed executing command(queryId=hive_20201128171818_c60c18ec-740c-4f73-8290-8b54b4805f35)
; Time taken: 0.007 seconds
INFO  : OK
No rows affected (0.342 seconds)
```

3. Create tables

a. Create transport table

```
0: jdbc:hive2://babar.es.its.nyu.edu:10000/> create external table transport(hash string, state strin
g, county string, numberOfBridges int, numberOfResidents int, pctOfMediumToFairConditionBridges doubl
e, pctOfPoorConditionBridges double, primaryAndCommercialAirports int, routeMilesOfFreightRailroad do
uble, routeMilesOfPassengerRailroadAndRailTransit double, roadsAcceptable double) row format delimite
d fields terminated by ',' location '/user/jc8017/transport_final';
INFO  : Compiling command(queryId=hive_20201128172222_2df59b2a-d7ea-4f46-80cf-0f2917638d40): create e
xternal table transport(hash string, state string, county string, numberOfBridges int, numberOfReside
nts int, pctOfMediumToFairConditionBridges double, pctOfPoorConditionBridges double, primaryAndCommer
cialAirports int, routeMilesOfFreightRailroad double, routeMilesOfPassengerRailroadAndRailTransit dou
ble, roadsAcceptable double) row format delimited fields terminated by ',' location '/user/jc8017/tra
nsport_final'
INFO  : Semantic Analysis Completed
INFO  : Returning Hive schema: Schema(fieldSchemas:null, properties:null)
INFO  : Completed compiling command(queryId=hive_20201128172222_2df59b2a-d7ea-4f46-80cf-0f2917638d40)
; Time taken: 0.175 seconds
INFO  : Executing command(queryId=hive_20201128172222_2df59b2a-d7ea-4f46-80cf-0f2917638d40): create e
xternal table transport(hash string, state string, county string, numberOfBridges int, numberOfReside
nts int, pctOfMediumToFairConditionBridges double, pctOfPoorConditionBridges double, primaryAndCommer
cialAirports int, routeMilesOfFreightRailroad double, routeMilesOfPassengerRailroadAndRailTransit dou
ble, roadsAcceptable double) row format delimited fields terminated by ',' location '/user/jc8017/tra
nsport_final'
INFO  : Starting task [Stage-0:DDL] in serial mode
INFO  : Completed executing command(queryId=hive_20201128172222_2df59b2a-d7ea-4f46-80cf-0f2917638d40)
; Time taken: 0.65 seconds
INFO  : OK
No rows affected (0.898 seconds)
```

b. Create tax table

```
0: jdbc:hive2://babar.es.its.nyu.edu:10000/> create external table tax(hash string, state string, cou
nty string, numberOfReturns int, adjustGrossIncome int, totalIncomeAmount int, salariesAndWagesAmount
 int, stateAndLocalIncomeTaxAmount int, realEstateTaxes int, totalTaxesPaid int, residentialEnergyTax
CreditAmount int) row format delimited fields terminated by ',' location '/user/jc8017/tax_final';
INFO  : Compiling command(queryId=hive_20201128172323_a27123c7-a46a-447e-b662-c3b81ed20c11): create e
xternal table tax(hash string, state string, county string, numberOfReturns int, adjustGrossIncome in
t, totalIncomeAmount int, salariesAndWagesAmount int, stateAndLocalIncomeTaxAmount int, realEstateTax
es int, totalTaxesPaid int, residentialEnergyTaxCreditAmount int) row format delimited fields termina
ted by ',' location '/user/jc8017/tax_final'
INFO  : Semantic Analysis Completed
INFO  : Returning Hive schema: Schema(fieldSchemas:null, properties:null)
INFO  : Completed compiling command(queryId=hive_20201128172323_a27123c7-a46a-447e-b662-c3b81ed20c11)
; Time taken: 0.195 seconds
INFO  : Executing command(queryId=hive_20201128172323_a27123c7-a46a-447e-b662-c3b81ed20c11): create e
xternal table tax(hash string, state string, county string, numberOfReturns int, adjustGrossIncome in
t, totalIncomeAmount int, salariesAndWagesAmount int, stateAndLocalIncomeTaxAmount int, realEstateTax
es int, totalTaxesPaid int, residentialEnergyTaxCreditAmount int) row format delimited fields termina
ted by ',' location '/user/jc8017/tax_final'
INFO  : Starting task [Stage-0:DDL] in serial mode
INFO  : Completed executing command(queryId=hive_20201128172323_a27123c7-a46a-447e-b662-c3b81ed20c11)
; Time taken: 0.476 seconds
INFO  : OK
No rows affected (0.732 seconds)
```

c. Create water quality table

```
0: jdbc:hive2://babar.es.its.nyu.edu:10000/> create external table water(hash string, state string, c
ounty string, populationServed int, waterSystems int, citiesServed int, waterSystemsPerCapita double)
 row format delimited fields terminated by ',' location '/user/jc8017/water_final';
INFO  : Compiling command(queryId=hive_20201128172424_5292d823-1fb8-44c4-8491-5783d83a5b0d): create e
xternal table water(hash string, state string, county string, populationServed int, waterSystems int,
 citiesServed int, waterSystemsPerCapita double) row format delimited fields terminated by ',' locati
on '/user/jc8017/water_final'
INFO  : Semantic Analysis Completed
INFO  : Returning Hive schema: Schema(fieldSchemas:null, properties:null)
INFO  : Completed compiling command(queryId=hive_20201128172424_5292d823-1fb8-44c4-8491-5783d83a5b0d)
; Time taken: 0.168 seconds
INFO  : Executing command(queryId=hive_20201128172424_5292d823-1fb8-44c4-8491-5783d83a5b0d): create e
xternal table water(hash string, state string, county string, populationServed int, waterSystems int,
 citiesServed int, waterSystemsPerCapita double) row format delimited fields terminated by ',' locati
on '/user/jc8017/water_final'
INFO  : Starting task [Stage-0:DDL] in serial mode
INFO  : Completed executing command(queryId=hive_20201128172424_5292d823-1fb8-44c4-8491-5783d83a5b0d)
; Time taken: 0.757 seconds
INFO  : OK
No rows affected (0.98 seconds)
```

4. Combine Tables

a. Combine tax and transportation

```
0: jdbc:hive2://babar.es.its.nyu.edu:10000/> create table taxtransport as select transport.*, tax.num
berofreturns, tax.adjustgrossincome, tax.totalincomeamount, tax.salariesandwagesamount, tax.stateandl
ocalincometaxamount, tax.realestatetaxes, tax.totaltaxespaid, tax.residentialenergytaxcreditamount fr
om transport left outer join tax on (transport.hash = tax.hash);
INFO  : Compiling command(queryId=hive_20201128172727_ebd90ce3-0c5d-4bf5-8ff5-fa04e8846acc): create t
able taxtransport as select transport.*, tax.numberofreturns, tax.adjustgrossincome, tax.totalincomea
mount, tax.salariesandwagesamount, tax.stateandlocalincometaxamount, tax.realestatetaxes, tax.totalta
xespaid, tax.residentialenergytaxcreditamount from transport left outer join tax on (transport.hash =
 tax.hash)
INFO  : Semantic Analysis Completed
INFO  : Returning Hive schema: Schema(fieldSchemas:[FieldSchema(name:transport.hash, type:string, com
ment:null), FieldSchema(name:transport.state, type:string, comment:null), FieldSchema(name:transport.
county, type:string, comment:null), FieldSchema(name:transport.numberofbridges, type:int, comment:nul
l), FieldSchema(name:transport.numberofresidents, type:int, comment:null), FieldSchema(name:transport
.pctofmediumtofairconditionbridges, type:double, comment:null), FieldSchema(name:transport.pctofpoorc
onditionbridges, type:double, comment:null), FieldSchema(name:transport.primaryandcommercialairports,
 type:int, comment:null), FieldSchema(name:transport.routemilesoffreightrailroad, type:double, commen
t:null), FieldSchema(name:transport.routemilesofpassengerrailroadandrailtransit, type:double, comment
:null), FieldSchema(name:transport.roadsacceptable, type:double, comment:null), FieldSchema(name:tax.
numberofreturns, type:int, comment:null), FieldSchema(name:tax.adjustgrossincome, type:int, comment:n
ull), FieldSchema(name:tax.totalincomeamount, type:int, comment:null), FieldSchema(name:tax.salariesa
ndwagesamount, type:int, comment:null), FieldSchema(name:tax.stateandlocalincometaxamount, type:int,
comment:null), FieldSchema(name:tax.realestatetaxes, type:int, comment:null), FieldSchema(name:tax.to
taltaxespaid, type:int, comment:null), FieldSchema(name:tax.residentialenergytaxcreditamount, type:in
t, comment:null)], properties:null)
INFO  : Completed compiling command(queryId=hive_20201128172727_ebd90ce3-0c5d-4bf5-8ff5-fa04e8846acc)
; Time taken: 0.314 seconds
INFO  : Executing command(queryId=hive_20201128172727_ebd90ce3-0c5d-4bf5-8ff5-fa04e8846acc): create t
able taxtransport as select transport.*, tax.numberofreturns, tax.adjustgrossincome, tax.totalincomea
mount, tax.salariesandwagesamount, tax.stateandlocalincometaxamount, tax.realestatetaxes, tax.totalta
xespaid, tax.residentialenergytaxcreditamount from transport left outer join tax on (transport.hash =
 tax.hash)
```

```
INFO  : Query ID = hive_20201128172727_ebd90ce3-0c5d-4bf5-8ff5-fa04e8846acc
INFO  : Total jobs = 1
INFO  : Starting task [Stage-5:MAPREDLOCAL] in serial mode
20/11/28 17:27:05 WARN conf.HiveConf: HiveConf of name hive.server2.idle.session.timeout_check_operat
ion does not exist
20/11/28 17:27:05 WARN conf.HiveConf: HiveConf of name hive.entity.capture.input.URI does not exist
Execution log at: /tmp/hive/hive_20201128172727_ebd90ce3-0c5d-4bf5-8ff5-fa04e8846acc.log
2020-11-28 05:27:06     Starting to launch local task to process map join;     maximum memory = 3113
877504
2020-11-28 05:27:06     Uploaded 1 File to: file:/tmp/hive/7c79eb2b-b4c2-4fa1-abb3-67f135ff8d6f/hive_
2020-11-28_17-27-03_681_713332212614613608-539/-local-10003/HashTable-Stage-4/MapJoin-mapfile381--.ha
shtable (160139 bytes)
INFO  : Execution completed successfully
INFO  : MapredLocal task succeeded
INFO  : Launching Job 1 out of 1
INFO  : Starting task [Stage-4:MAPRED] in serial mode
INFO  : Number of reduce tasks is set to 0 since there's no reduce operator
INFO  : number of splits:1
INFO  : Submitting tokens for job: job_1604346392376_5280
INFO  : The url to track the job: http://babar.es.its.nyu.edu:8088/proxy/application_1604346392376_52
80/
INFO  : Starting Job = job_1604346392376_5280, Tracking URL = http://babar.es.its.nyu.edu:8088/proxy/
application_1604346392376_5280/
INFO  : Kill Command = /opt/cloudera/parcels/CDH-5.15.2-1.cdh5.15.2.p0.3/lib/hadoop/bin/hadoop job  -
kill job_1604346392376_5280
INFO  : Hadoop job information for Stage-4: number of mappers: 1; number of reducers: 0
INFO  : 2020-11-28 17:27:21,107 Stage-4 map = 0%,  reduce = 0%
INFO  : 2020-11-28 17:27:28,289 Stage-4 map = 100%,  reduce = 0%, Cumulative CPU 3.44 sec
INFO  : MapReduce Total cumulative CPU time: 3 seconds 440 msec
INFO  : Ended Job = job_1604346392376_5280
INFO  : Starting task [Stage-0:MOVE] in serial mode
INFO  : Moving data to: hdfs://dumbo/user/hive/warehouse/jc8017.db/taxtransport from hdfs://dumbo/use
r/hive/warehouse/jc8017.db/.hive-staging_hive_2020-11-28_17-27-03_681_713332212614613608-539/-ext-100
01
INFO  : Starting task [Stage-6:DDL] in serial mode
INFO  : Starting task [Stage-2:STATS] in serial mode
INFO  : Table jc8017.taxtransport stats: [numFiles=1, numRows=2731, totalSize=379559, rawDataSize=376
828]
INFO  : MapReduce Jobs Launched:
INFO  : Stage-Stage-4: Map: 1   Cumulative CPU: 3.44 sec   HDFS Read: 260962 HDFS Write: 379640 SUCCE
SS
INFO  : Total MapReduce CPU Time Spent: 3 seconds 440 msec
INFO  : Completed executing command(queryId=hive_20201128172727_ebd90ce3-0c5d-4bf5-8ff5-fa04e8846acc)
; Time taken: 27.473 seconds
INFO  : OK
No rows affected (28.03 seconds)
```

b. Combine tax, transport, and water

```
0: jdbc:hive2://babar.es.its.nyu.edu:10000/> create table final as select taxtransport.*, water.popul
ationserved, water.watersystems, water.citiesserved, water.watersystemspercapita from taxtransport le
ft outer join water on (taxtransport.hash = water.hash);
INFO  : Compiling command(queryId=hive_20201128173333_8f28f132-bda5-456f-a0ec-bf5022c8da3c): create t
able final as select taxtransport.*, water.populationserved, water.watersystems, water.citiesserved,
water.watersystemspercapita from taxtransport left outer join water on (taxtransport.hash = water.has
h)
INFO  : Semantic Analysis Completed
INFO  : Returning Hive schema: Schema(fieldSchemas:[FieldSchema(name:taxtransport.hash, type:string,
comment:null), FieldSchema(name:taxtransport.state, type:string, comment:null), FieldSchema(name:taxt
ransport.county, type:string, comment:null), FieldSchema(name:taxtransport.numberofbridges, type:int,
 comment:null), FieldSchema(name:taxtransport.numberofresidents, type:int, comment:null), FieldSchema
(name:taxtransport.pctofmediumtofairconditionbridges, type:double, comment:null), FieldSchema(name:ta
xtransport.pctofpoorconditionbridges, type:double, comment:null), FieldSchema(name:taxtransport.prima
ryandcommercialairports, type:int, comment:null), FieldSchema(name:taxtransport.routemilesoffreightra
ilroad, type:double, comment:null), FieldSchema(name:taxtransport.routemilesofpassengerrailroadandrai
ltransit, type:double, comment:null), FieldSchema(name:taxtransport.roadsacceptable, type:double, com
ment:null), FieldSchema(name:taxtransport.numberofreturns, type:int, comment:null), FieldSchema(name:
taxtransport.adjustgrossincome, type:int, comment:null), FieldSchema(name:taxtransport.totalincomeamo
unt, type:int, comment:null), FieldSchema(name:taxtransport.salariesandwagesamount, type:int, comment
:null), FieldSchema(name:taxtransport.stateandlocalincometaxamount, type:int, comment:null), FieldSch
ema(name:taxtransport.realestatetaxes, type:int, comment:null), FieldSchema(name:taxtransport.totalta
xespaid, type:int, comment:null), FieldSchema(name:taxtransport.residentialenergytaxcreditamount, typ
e:int, comment:null), FieldSchema(name:water.populationserved, type:int, comment:null), FieldSchema(n
ame:water.watersystems, type:int, comment:null), FieldSchema(name:water.citiesserved, type:int, comme
nt:null), FieldSchema(name:water.watersystemspercapita, type:double, comment:null)], properties:null)
INFO  : Completed compiling command(queryId=hive_20201128173333_8f28f132-bda5-456f-a0ec-bf5022c8da3c)
; Time taken: 0.346 seconds
INFO  : Executing command(queryId=hive_20201128173333_8f28f132-bda5-456f-a0ec-bf5022c8da3c): create t
able final as select taxtransport.*, water.populationserved, water.watersystems, water.citiesserved,
water.watersystemspercapita from taxtransport left outer join water on (taxtransport.hash = water.has
h)
```

```
INFO  : Query ID = hive_20201128173333_8f28f132-bda5-456f-a0ec-bf5022c8da3c
INFO  : Total jobs = 1
INFO  : Starting task [Stage-5:MAPREDLOCAL] in serial mode
20/11/28 17:33:03 WARN conf.HiveConf: HiveConf of name hive.server2.idle.session.timeout_check_operat
ion does not exist
20/11/28 17:33:03 WARN conf.HiveConf: HiveConf of name hive.entity.capture.input.URI does not exist
Execution log at: /tmp/hive/hive_20201128173333_8f28f132-bda5-456f-a0ec-bf5022c8da3c.log
2020-11-28 05:33:04     Dump the side-table for tag: 1 with group count: 2731 into file: file:/tmp/hi
ve/7c79eb2b-b4c2-4fa1-abb3-67f135ff8d6f/hive_2020-11-28_17-33-01_447_8656000507559607658-539/-local-1
0003/HashTable-Stage-4/MapJoin-mapfile391--.hashtable2020-11-28 05:33:04     Uploaded 1 File to: f
ile:/tmp/hive/7c79eb2b-b4c2-4fa1-abb3-67f135ff8d6f/hive_2020-11-28_17-33-01_447_8656000507559607658-5
39/-local-10003/HashTable-Stage-4/MapJoin-mapfile391--.hashtable (127891 bytes)
INFO  : Execution completed successfully
INFO  : MapredLocal task succeeded
INFO  : Launching Job 1 out of 1
INFO  : Starting task [Stage-4:MAPRED] in serial mode
INFO  : Number of reduce tasks is set to 0 since there's no reduce operator
INFO  : number of splits:1
INFO  : Submitting tokens for job: job_1604346392376_5287
INFO  : The url to track the job: http://babar.es.its.nyu.edu:8088/proxy/application_1604346392376_52
87/
INFO  : Starting Job = job_1604346392376_5287, Tracking URL = http://babar.es.its.nyu.edu:8088/proxy/
application_1604346392376_5287/
INFO  : Kill Command = /opt/cloudera/parcels/CDH-5.15.2-1.cdh5.15.2.p0.3/lib/hadoop/bin/hadoop job  -
kill job_1604346392376_5287
INFO  : Hadoop job information for Stage-4: number of mappers: 1; number of reducers: 0
INFO  : 2020-11-28 17:33:18,254 Stage-4 map = 0%,  reduce = 0%
INFO  : 2020-11-28 17:33:25,425 Stage-4 map = 100%,  reduce = 0%, Cumulative CPU 3.35 sec
INFO  : MapReduce Total cumulative CPU time: 3 seconds 350 msec
INFO  : Ended Job = job_1604346392376_5287
INFO  : Starting task [Stage-0:MOVE] in serial mode
INFO  : Moving data to: hdfs://dumbo/user/hive/warehouse/jc8017.db/final from hdfs://dumbo/user/hive/
warehouse/jc8017.db/.hive-staging_hive_2020-11-28_17-33-01_447_8656000507559607658-539/-ext-10001
INFO  : Starting task [Stage-6:DDL] in serial mode
INFO  : Starting task [Stage-2:STATS] in serial mode
INFO  : Table jc8017.final stats: [numFiles=1, numRows=2731, totalSize=437855, rawDataSize=435124]
INFO  : MapReduce Jobs Launched:
INFO  : Stage-Stage-4: Map: 1   Cumulative CPU: 3.35 sec   HDFS Read: 391115 HDFS Write: 437929 SUCCE
SS
INFO  : Total MapReduce CPU Time Spent: 3 seconds 350 msec
INFO  : Completed executing command(queryId=hive_20201128173333_8f28f132-bda5-456f-a0ec-bf5022c8da3c)
; Time taken: 26.765 seconds
INFO  : OK
No rows affected (27.291 seconds)
```

5. Get hive tables

INSERT OVERWRITE DIRECTORY '/output_final' ROW FORMAT DELIMITED FIELDS
TERMINATED BY ',' LINES TERMINATED BY "\n" SELECT * FROM final;

```
0: jdbc:hive2://babar.es.its.nyu.edu:10000/> INSERT OVERWRITE DIRECTORY '/output
_final' ROW FORMAT DELIMITED FIELDS TERMINATED BY ',' LINES TERMINATED BY "\n" S
ELECT * FROM final;
Error: Error while compiling statement: FAILED: SemanticException Error creating
 temporary folder on: /output_final (state=42000,code=40000)
0: jdbc:hive2://babar.es.its.nyu.edu:10000/> INSERT OVERWRITE DIRECTORY '/user/j
c8017/output_final' ROW FORMAT DELIMITED FIELDS TERMINATED BY ',' LINES TERMINAT
ED BY "\n" SELECT * FROM final;
INFO  : Compiling command(queryId=hive_20201128182525_aa09b505-662d-4df1-9a77-78
3da037aef1): INSERT OVERWRITE DIRECTORY '/user/jc8017/output_final' ROW FORMAT D
ELIMITED FIELDS TERMINATED BY ',' LINES TERMINATED BY "\n" SELECT * FROM final
INFO  : Semantic Analysis Completed
INFO  : Returning Hive schema: Schema(fieldSchemas:[FieldSchema(name:final.hash,
 type:string, comment:null), FieldSchema(name:final.state, type:string, comment:
null), FieldSchema(name:final.county, type:string, comment:null), FieldSchema(na
me:final.numberofbridges, type:int, comment:null), FieldSchema(name:final.number
ofresidents, type:int, comment:null), FieldSchema(name:final.pctofmediumtofairco
nditionbridges, type:double, comment:null), FieldSchema(name:final.pctofpoorcond
itionbridges, type:double, comment:null), FieldSchema(name:final.primaryandcomme
rcialairports, type:int, comment:null), FieldSchema(name:final.routemilesoffreig
htrailroad, type:double, comment:null), FieldSchema(name:final.routemilesofpasse
ngerrailroadandrailtransit, type:double, comment:null), FieldSchema(name:final.r
oadsacceptable, type:double, comment:null), FieldSchema(name:final.numberofretur
ns, type:int, comment:null), FieldSchema(name:final.adjustgrossincome, type:int,
 comment:null), FieldSchema(name:final.totalincomeamount, type:int, comment:null
), FieldSchema(name:final.salariesandwagesamount, type:int, comment:null), Field
Schema(name:final.stateandlocalincometaxamount, type:int, comment:null), FieldSc
hema(name:final.realestatetaxes, type:int, comment:null), FieldSchema(name:final
.totaltaxespaid, type:int, comment:null), FieldSchema(name:final.residentialener
gytaxcreditamount, type:int, comment:null), FieldSchema(name:final.populationser
ved, type:int, comment:null), FieldSchema(name:final.watersystems, type:int, com
ment:null), FieldSchema(name:final.citiesserved, type:int, comment:null), FieldS
chema(name:final.watersystemspercapita, type:double, comment:null)], properties:
null)
```

```
INFO  : Completed compiling command(queryId=hive_20201128182525_aa09b505-662d-4d
f1-9a77-783da037aef1); Time taken: 0.31 seconds
INFO  : Executing command(queryId=hive_20201128182525_aa09b505-662d-4df1-9a77-78
3da037aef1): INSERT OVERWRITE DIRECTORY '/user/jc8017/output_final' ROW FORMAT D
ELIMITED FIELDS TERMINATED BY ',' LINES TERMINATED BY "\n" SELECT * FROM final
INFO  : Query ID = hive_20201128182525_aa09b505-662d-4df1-9a77-783da037aef1
INFO  : Total jobs = 3
INFO  : Launching Job 1 out of 3
INFO  : Starting task [Stage-1:MAPRED] in serial mode
INFO  : Number of reduce tasks is set to 0 since there's no reduce operator
INFO  : number of splits:1
INFO  : Submitting tokens for job: job_1604346392376_5349
INFO  : The url to track the job: http://babar.es.its.nyu.edu:8088/proxy/applica
tion_1604346392376_5349/
INFO  : Starting Job = job_1604346392376_5349, Tracking URL = http://babar.es.it
s.nyu.edu:8088/proxy/application_1604346392376_5349/
INFO  : Kill Command = /opt/cloudera/parcels/CDH-5.15.2-1.cdh5.15.2.p0.3/lib/had
oop/bin/hadoop job  -kill job_1604346392376_5349
INFO  : Hadoop job information for Stage-1: number of mappers: 1; number of redu
cers: 0
INFO  : 2020-11-28 18:26:02,713 Stage-1 map = 0%,  reduce = 0%
INFO  : 2020-11-28 18:26:08,861 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU
 3.12 sec
INFO  : MapReduce Total cumulative CPU time: 3 seconds 120 msec
INFO  : Ended Job = job_1604346392376_5349
INFO  : Starting task [Stage-6:CONDITIONAL] in serial mode
INFO  : Stage-3 is selected by condition resolver.
INFO  : Stage-2 is filtered out by condition resolver.
INFO  : Stage-4 is filtered out by condition resolver.
INFO  : Starting task [Stage-3:MOVE] in serial mode
INFO  : Moving data to: hdfs://dumbo/user/jc8017/output_final/.hive-staging_hive
_2020-11-28_18-25-47_788_3744802541958706659-587/-ext-10000 from hdfs://dumbo/us
er/jc8017/output_final/.hive-staging_hive_2020-11-28_18-25-47_788_37448025419587
06659-587/-ext-10002
INFO  : Starting task [Stage-0:MOVE] in serial mode
INFO  : Moving data to: /user/jc8017/output_final from hdfs://dumbo/user/jc8017/
output_final/.hive-staging_hive_2020-11-28_18-25-47_788_3744802541958706659-587/
-ext-10000
INFO  : MapReduce Jobs Launched:
INFO  : Stage-Stage-1: Map: 1   Cumulative CPU: 3.12 sec   HDFS Read: 444969 HDF
S Write: 437855 SUCCESS
INFO  : Total MapReduce CPU Time Spent: 3 seconds 120 msec
INFO  : Completed executing command(queryId=hive_20201128182525_aa09b505-662d-4d
f1-9a77-783da037aef1); Time taken: 23.641 seconds
INFO  : OK
No rows affected (24.62 seconds)
```