

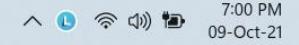
Group E - Notepad

File Edit Format View Help

Data Technology Solutions

Group E

Bibek Shah Shankhar
Davinderjit Singh
Kirandeep Kaur
Kamaldeep kaur Dhami
Anirudh Siddula



```
[hdfs@sandbox-hdp:~]$ vi compensation.csv
[hdfs@sandbox-hdp ~]$ cat compensation.csv
1,Bibek,salary:55000$comm:1000$bonus:100
2,Benjamin,salary:100000$comm:2000$bonus:100
[hdfs@sandbox-hdp ~]$ chmod 777 compensation.csv
[hdfs@sandbox-hdp ~]$ pwd
/home/hdfs
[hdfs@sandbox-hdp ~]$ hdfs dfs -put /home/hdfs/compensation.csv /user/bibek/
[hdfs@sandbox-hdp ~]$ hdfs dfs -ls /user/bibek/
Found 4 items
drwxr-xr-x  - root hdfs          0 2021-10-09 12:26 /user/bibek/bibek_hive_db
-rw-r--r--  1 hdfs hdfs        86 2021-10-09 13:22 /user/bibek/compensation.cs
v
drwxrwxrwx  - root hdfs          0 2021-10-09 12:15 /user/bibek/externalsales
drwxr-xr-x  - hive hdfs         0 2021-10-09 12:05 /user/bibek/salesbypaymentt
ype
[hdfs@sandbox-hdp ~]$ hdfs dfs -chmod 777 /user/bibek/compensation.csv
[hdfs@sandbox-hdp ~]$ hdfs dfs -ls /user/bibek/
Found 4 items
drwxr-xr-x  - root hdfs          0 2021-10-09 12:26 /user/bibek/bibek_hive_db
-rwxrwxrwx  1 hdfs hdfs        86 2021-10-09 13:22 /user/bibek/compensation.cs
v
drwxrwxrwx  - root hdfs          0 2021-10-09 12:15 /user/bibek/externalsales
drwxr-xr-x  - hive hdfs         0 2021-10-09 12:05 /user/bibek/salesbypaymentt
ype
[hdfs@sandbox-hdp ~]$
```

Group E - Notepad

Data Technology Solutions
Group E

Bibek Shah Shankhar
Davinderjit Singh
Kirandeep Kaur
Kamaldeep kaur Dhami
Anirudh Siddula

Ln 4, Col 2 160% Windows (CRLF) UTF-8



```
root@sandbox-hdp:~  
0: jdbc:hive2://sandbox-hdp.hortonworks.com:2> load data inpath '/user/bibek/  
compensation.csv'  
.....> into table employee_data;  
INFO : Compiling command(queryId=hive_20211009132727_3264b24d-33c0-49be-a632  
-53c35ab08f05); load data inpath '/user/bibek/compensation.csv'  
into table employee_data  
INFO : Semantic Analysis Completed (retrial = false)  
INFO : Returning Hive schema: Schema(fieldSchemas:null, properties:null)  
INFO : Completed compiling command(queryId=hive_20211009132727_3264b24d-33c0  
-49be-a632-53c35ab08f05); Time taken: 0.057 seconds  
INFO : Executing command(queryId=hive_20211009132727_3264b24d-33c0-49be-a632  
-53c35ab08f05); load data inpath '/user/bibek/compensation.csv'  
into table employee_data  
INFO : Starting task [Stage-0:MOVE] in serial mode  
INFO : Loading data to table test_db.employee_data from hdfs://sandbox-hdp.h  
ortonworks.com:8020/user/bibek/compensation.csv  
INFO : Starting task [Stage-1:STATS] in serial mode  
INFO : Completed executing command(queryId=hive_20211009132727_3264b24d-33c0  
-49be-a632-53c35ab08f05); Time taken: 0.397 seconds  
INFO : OK  
No rows affected (0.474 seconds)  
0: jdbc:hive2://sandbox-hdp.hortonworks.com:2>
```

Group E - Notepad

Data Technology Solutions
Group E

Bibek Shah Shankhar
Davinderjit Singh
Kirandeep Kaur
Kamaldeep kaur Dhami
Anirudh Siddula

Ln 4, Col 2 160% Windows (CRLF) UTF-8



```
root@sandbox-hdp:~  
0: jdbc:hive2://sandbox-hdp.hortonworks.com:2>  
0: jdbc:hive2://sandbox-hdp.hortonworks.com:2> select * from employee_data;  
INFO : Compiling command(queryId=hive_20211009132825_3147f774-7330-4f91-955e-9b  
780421de05): select * from employee_data  
INFO : Semantic Analysis Completed (retryal = false)  
INFO : Returning Hive schema: Schema(fieldSchemas:[FieldSchema(name:employee_da  
ta.id, type:int, comment:null), FieldSchema(name:employee_data.name, type:string  
, comment:null), FieldSchema(name:employee_data.compensation, type:map<string,in  
t>, comment:null)], properties:null)  
INFO : Completed compiling command(queryId=hive_20211009132825_3147f774-7330-4f  
91-955e-9b780421de05); Time taken: 0.408 seconds  
INFO : Executing command(queryId=hive_20211009132825_3147f774-7330-4f91-955e-9b  
780421de05): select * from employee_data  
INFO : Completed executing command(queryId=hive_20211009132825_3147f774-7330-4f  
91-955e-9b780421de05); Time taken: 0.013 seconds  
INFO : OK  
+-----+-----+-----+  
| employee_data.id | employee_data.name | employee_data.compensation |  
+-----+-----+-----+  
| 1 | Bibek | {"salary":55000,"comm":1000,"bonus":  
100} |  
| 2 | Benjamin | {"salary":100000,"comm":2000,"bonus"  
:100} |  
+-----+-----+-----+  
2 rows selected (0.693 seconds)  
0: jdbc:hive2://sandbox-hdp.hortonworks.com:2>
```

Group E - Notepad

Data Technology Solutions
Group E

Bibek Shah Shankhar
Davinderjit Singh
Kirandeep Kaur
Kamaldeep kaur Dhami
Anirudh Siddula

Ln 4, Col 2 160% Windows (CRLF) UTF-8



7:14 PM
09-Oct-21 1

Group E - Notepad

File Edit Format View Help

Data Technology Solutions

Group E

Bibek Shah Shankhar
Davinderjit Singh
Kirandeep Kaur
Kamaldeep kaur Dhami
Anirudh Siddula



Group E - Notepad

File Edit Format View Help

Data Technology Solutions

Group E

Bibek Shah Shankhar
Davinderjit Singh
Kirandeep Kaur
Kamaldeep kaur Dhami
Anirudh Siddula



Group E - Notepad

File Edit Format View Help

Data Technology Solutions

Group E

Bibek Shah Shankhar
Davinderjit Singh
Kirandeep Kaur
Kamaldeep kaur Dhami
Anirudh Siddula

Ln 4, Col 2 160% Windows (CRLF) UTF-8



```

root@sandbox-hdp:~ 
0: jdbc:hive2://sandbox-hdp.hortonworks.com:2> select * from sales_by_US_view;
INFO : Compiling command(queryId=hive_20211009133507_687214cb-f03b-4a3e-88b9-8e7ceb0328a5): select
* from sales_by_US_view
INFO : Semantic Analysis Completed (retrial = false)
INFO : Returning Hive schema: Schema(fieldSchemas:[FieldSchema(name:sales_by_us_view.productid, type:string, comment:null), FieldSchema(name:sales_by_us_view.salesamount, type:int, comment:null), FieldSchema(name:sales_by_us_view.paymenttype, type:string, comment:null), FieldSchema(name:sales_by_us_view.customername, type:string, comment:null), FieldSchema(name:sales_by_us_view.city, type:string, comment:null), FieldSchema(name:sales_by_us_view.region, type:string, comment:null), FieldSchema(name:sales_by_us_view.country, type:string, comment:null)], properties:null)
INFO : Completed compiling command(queryId=hive_20211009133507_687214cb-f03b-4a3e-88b9-8e7ceb0328a5)
; Time taken: 0.701 seconds
INFO : Executing command(queryId=hive_20211009133507_687214cb-f03b-4a3e-88b9-8e7ceb0328a5): select
* from sales_by_US_view
INFO : Completed executing command(queryId=hive_20211009133507_687214cb-f03b-4a3e-88b9-8e7ceb0328a5)
; Time taken: 0.004 seconds
INFO : OK
+-----+-----+-----+-----+
| sales_by_us_view.productid | sales_by_us_view.salesamount | sales_by_us_view.paymenttype | sales_by_us_view.customername | sales_by_us_view.city | sales_by_us_view.region | sales_by_us_view.country |
+-----+-----+-----+-----+
| Product1                   | 1200                  | Visa                    | Beti                   | Parkville             | MO                     | United States          |
| Product1                   | 1200                  | Mastercard               | Fede                   | Astoria              | OR                     | United States          |
| Product2                   | 3600                  | Visa                    | Gerd                   | Cahaba Heights        | AL                     | United States          |
| Product1                   | 1200                  | Visa                    | LAUR                  | Mickleton             | NJ                     | United States          |
| Product1                   | 1200                  | Mastercard               | Fleu                   | Peoria               | IL                     | United States          |
| Product1                   | 1200                  | Mastercard               | adam                  |                      |                        |                      |

```

Group E - Notepad
File Edit Format View Help
Data Technology Solutions
Group E
Bibek Shah Shankhar
Davinderjit Singh
Kirandeep Kaur
Kamaldeep kaur Dhami
Anirudh Siddula

Ln 4, Col 2 160% Windows (CRLF) UTF-8



7:20 PM
09-Oct-21 1

Group E - Notepad

File Edit Format View Help

Data Technology Solutions

Group E

Bibek Shah Shankhar
Davinderjit Singh
Kirandeep Kaur
Kamaldeep kaur Dhami
Anirudh Siddula



```
root@sandbox-hdp:~  
INFO : ---  
--  
INFO : Compile Query 0.44s  
INFO : Prepare Plan 9.66s  
INFO : Get Query Coordinator (AM) 0.00s  
INFO : Submit Plan 0.37s  
INFO : Start DAG 2.18s  
INFO : Run DAG 87.61s  
INFO : ---  
--  
INFO :  
INFO : Task Execution Summary  
INFO : ---  
INFO : VERTICES DURATION(ms) CPU_TIME(ms) GC_TIME(ms) INPUT_RECORDS OUTPUT_RECORDS  
INFO : ---  
INFO : Map 1 28390.00 3,720 85 998 4  
INFO : Reducer 2 57201.00 3,350 72 4 0  
INFO : ---  
--  
INFO :  
INFO : org.apache.tez.common.counters.DAGCounter:  
INFO : NUM_SUCCEEDED_TASKS: 3  
INFO : TOTAL_LAUNCHED_TASKS: 3  
INFO : AM_CPU_MILLISECONDS: 4230  
INFO : AM_GC_TIME_MILLIS: 22  
INFO : File System Counters:  
INFO : FILE_BYTES_READ: 361  
INFO : FILE_BYTES_WRITTEN: 159  
INFO : HDFS_BYTES_READ: 65835  
INFO : HDFS_BYTES_WRITTEN: 277  
INFO : HDFS_READ_OPS: 7  
INFO : HDFS_WRITE_OPS: 4  
INFO : HDFS_OP_CREATE: 2  
INFO : HDFS_OP_GET_FILE_STATUS: 6  
INFO : HDFS_OP_OPEN: 1  
INFO : HDFS_OP_RENAME: 2  
INFO : org.apache.tez.common.counters.TaskCounter:  
INFO : REDUCE_INPUT_GROUPS: 4  
INFO : REDUCE_INPUT_RECORDS: 4  
INFO : COMBINE_INPUT_RECORDS: 0
```

```
Group E - Notepad  
File Edit Format View Help  
Data Technology Solutions  
Group E  
  
Bibek Shah Shankhar  
Davinderjit Singh  
Kirandeep Kaur  
Kamaldeep kaur Dhami  
Anirudh Siddula  
  
Ln 4, Col 2 160% Windows (CRLF) UTF-8
```



```
root@sandbox-hdp:~  
INFO : SPILLED_RECORDS: 8  
INFO : NUM_SHUFFLED_INPUTS: 2  
INFO : NUM_SKIPPED_INPUTS: 0  
INFO : NUM_FAILED_SHUFFLE_INPUTS: 0  
INFO : MERGED_MAP_OUTPUTS: 2  
INFO : GC_TIME_MILLIS: 157  
INFO : TASK_DURATION_MILLIS: 46712  
INFO : CPU_MILLISECONDS: 7070  
INFO : PHYSICAL_MEMORY_BYTES: 1549795328  
INFO : VIRTUAL_MEMORY_BYTES: 8055992320  
INFO : COMMITTED_HEAP_BYTES: 1549795328  
INFO : INPUT_RECORDS_PROCESSED: 998  
INFO : INPUT_SPLIT_LENGTH_BYTES: 65835  
INFO : OUTPUT_RECORDS: 4  
INFO : OUTPUT_LARGE_RECORDS: 0  
INFO : OUTPUT_BYTES: 55  
INFO : OUTPUT_BYTES_WITH_OVERHEAD: 75  
INFO : OUTPUT_BYTES_PHYSICAL: 103  
INFO : ADDITIONAL_SPILLS_BYTES_WRITTEN: 0  
INFO : ADDITIONAL_SPILLS_BYTES_READ: 103  
INFO : ADDITIONAL_SPILL_COUNT: 0  
INFO : SHUFFLE_CHUNK_COUNT: 1  
INFO : SHUFFLE_BYTES: 103  
INFO : SHUFFLE_BYTES_DECOMPRESSED: 75  
INFO : SHUFFLE_BYTES_TO_MEM: 0  
INFO : SHUFFLE_BYTES_TO_DISK: 0  
INFO : SHUFFLE_BYTES_DISK_DIRECT: 103  
INFO : NUM_MEM_TO_DISK_MERGES: 0  
INFO : NUM_DISK_TO_DISK_MERGES: 0  
INFO : SHUFFLE_PHASE_TIME: 1146  
INFO : MERGE_PHASE_TIME: 1694  
INFO : FIRST_EVENT_RECEIVED: 1083  
INFO : LAST_EVENT_RECEIVED: 1083  
INFO : HIVE:  
INFO : CREATED_FILES: 2  
INFO : DESERIALIZE_ERRORS: 0  
INFO : RECORDS_IN_Map_1: 998  
INFO : RECORDS_OUT_0: 4  
INFO : RECORDS_OUT_INTERMEDIATE_Map_1: 4  
INFO : RECORDS_OUT_INTERMEDIATE_Reducer_2: 0  
INFO : RECORDS_OUT_OPERATOR_FILE_9: 463  
INFO : RECORDS_OUT_OPERATOR_FS_14: 4
```

Group E - Notepad

Data Technology Solutions
Group E

Bibek Shah Shankhar
Davinderjit Singh
Kirandeep Kaur
Kamaldeep kaur Dhami
Anirudh Siddula

Ln 4, Col 2 160% Windows (CRLF) UTF-8



```
root@sandbox-hdp:~  
INFO : TaskCounter_Reducer_2_INPUT_Map_1:  
INFO : ADDITIONAL_SPILLS_BYTES_READ: 103  
INFO : ADDITIONAL_SPILLS_BYTES_WRITTEN: 0  
INFO : COMBINE_INPUT_RECORDS: 0  
INFO : FIRST_EVENT_RECEIVED: 1083  
INFO : LAST_EVENT_RECEIVED: 1083  
INFO : MERGED_MAP_OUTPUTS: 2  
INFO : MERGE_PHASE_TIME: 1694  
INFO : NUM_DISK_TO_DISK_MERGES: 0  
INFO : NUM_FAILED_SHUFFLE_INPUTS: 0  
INFO : NUM_MEM_TO_DISK_MERGES: 0  
INFO : NUM_SHUFFLED_INPUTS: 2  
INFO : NUM_SKIPPED_INPUTS: 0  
INFO : REDUCE_INPUT_GROUPS: 4  
INFO : REDUCE_INPUT_RECORDS: 4  
INFO : SHUFFLE_BYTES: 103  
INFO : SHUFFLE_BYTES_DECOMPRESSED: 75  
INFO : SHUFFLE_BYTES_DISK_DIRECT: 103  
INFO : SHUFFLE_BYTES_TO_DISK: 0  
INFO : SHUFFLE_BYTES_TO_MEM: 0  
INFO : SHUFFLE_PHASE_TIME: 1146  
INFO : SPILLED_RECORDS: 4  
INFO : TaskCounter_Reducer_2_OUTPUT_out_Reducer_2:  
INFO : OUTPUT_RECORDS: 0  
INFO : org.apache.hadoop.hive.ql.exec.tez.HiveInputCounters:  
INFO : GROUPED_INPUT_SPLITS_Map_1: 1  
INFO : INPUT_DIRECTORIES_Map_1: 1  
INFO : INPUT_FILES_Map_1: 1  
INFO : RAW_INPUT_SPLITS_Map_1: 1  
INFO : Completed executing command(queryId=hive_20211009133621_392fcb58-68c9-42ce-8684-f95d536bf0c4  
); Time taken: 100.016 seconds  
INFO : OK  
+-----+  
| paymenttype | _c1 |  
+-----+  
| Diners | 45600 |  
| Mastercard | 204350 |  
| Visa | 378350 |  
| Amex | 121700 |  
+-----+  
4 rows selected (100.812 seconds)  
0: jdbc:hive2://sandbox-hdp.hortonworks.com:2>
```

```
Group E - Notepad  
File Edit Format View Help  
Data Technology Solutions  
Group E  
  
Bibek Shah Shankhar  
Davinderjit Singh  
Kirandeep Kaur  
Kamaldeep kaur Dhami  
Anirudh Siddula
```



```
[root@sandbox-hdp:~]# vi create_table_script
[root@sandbox-hdp ~]# cat create_table_script
use test_db;
create table SalesJan2009_by_script(
productid string,salesamount int, paymenttype string,customernumber string,city string,region string,
country string)
row format delimited
fields terminated by ','
stored as textfile;
load data inpath '/user/bibek/SalesJan2009.csv' into table SalesJan2009_by_script;

[root@sandbox-hdp ~]# chmod 777 create_table_script
[root@sandbox-hdp ~]# hdfs dfs -put SalesJan2009.csv /user/bibek/
put: '/user/bibek/SalesJan2009.csv': File exists
[root@sandbox-hdp ~]# hive -f create_table_script
```

Group E - Notepad

Data Technology Solutions
Group E

Bibek Shah Shankhar
Davinderjit Singh
Kirandeep Kaur
Kamaldeep kaur Dhami
Anirudh Siddula

Ln 4, Col 2 160% Windows (CRLF) UTF-8



7:37 PM
09-Oct-21 1

Group E - Notepad

File Edit Format View Help

Data Technology Solutions

Group E

Bibek Shah Shankhar
Davinderjit Singh
Kirandeep Kaur
Kamaldeep kaur Dhami
Anirudh Siddula



```
root@sandbox-hdp:~  
row format delimited  
fields terminated by ','  
stored as textfile  
INFO : Semantic Analysis Completed (retrial = false)  
INFO : Returning Hive schema: Schema(fieldSchemas:null, properties:null)  
INFO : Completed compiling command(queryId=hive_20211009135216_fa7982b5-ef06-4350-bclf-c482c3c77a18  
INFO : Executing command(queryId=hive_20211009135216_fa7982b5-ef06-4350-bclf-c482c3c77a18): create  
table SalesJan2009_by_script(  
productid string,salesamount int, paymenttype string,customername string,city string,region string,  
country string)  
row format delimited  
fields terminated by ','  
stored as textfile  
INFO : Starting task [Stage-0:DDL] in serial mode  
INFO : Completed executing command(queryId=hive_20211009135216_fa7982b5-ef06-4350-bclf-c482c3c77a18  
INFO : OK  
No rows affected (0.196 seconds)  
0: jdbc:hive2://sandbox-hdp.hortonworks.com:2> load data inpath '/user/bibek/SalesJan2009.csv' into  
table SalesJan2009_by_script;  
INFO : Compiling command(queryId=hive_20211009135216_6faea3b0-8532-4640-87a4-e4310843b910): load da  
ta inpath '/user/bibek/SalesJan2009.csv' into table SalesJan2009_by_script  
INFO : Semantic Analysis Completed (retrial = false)  
INFO : Returning Hive schema: Schema(fieldSchemas:null, properties:null)  
INFO : Completed compiling command(queryId=hive_20211009135216_6faea3b0-8532-4640-87a4-e4310843b910  
INFO : Executing command(queryId=hive_20211009135216_6faea3b0-8532-4640-87a4-e4310843b910): load da  
ta inpath '/user/bibek/SalesJan2009.csv' into table SalesJan2009_by_script  
INFO : Starting task [Stage-0:MOVE] in serial mode  
INFO : Loading data to table test_db.salesjan2009_by_script from hdfs://sandbox-hdp.hortonworks.com:  
8020/user/bibek/SalesJan2009.csv  
INFO : Starting task [Stage-1:STATS] in serial mode  
INFO : Completed executing command(queryId=hive_20211009135216_6faea3b0-8532-4640-87a4-e4310843b910  
INFO : OK  
No rows affected (0.593 seconds)  
0: jdbc:hive2://sandbox-hdp.hortonworks.com:2>  
0: jdbc:hive2://sandbox-hdp.hortonworks.com:2>  
0: jdbc:hive2://sandbox-hdp.hortonworks.com:2> Closing: 0: jdbc:hive2://sandbox-hdp.hortonworks.com:  
2181/default;password=hive;serviceDiscoveryMode=zooKeeper;user=hive;zooKeeperNamespace=hiveserver2  
[root@sandbox-hdp ~]#
```

Group E - Notepad

Data Technology Solutions
Group E

Bibek Shah Shankhar
Davinderjit Singh
Kirandeep Kaur
Kamaldeep kaur Dhami
Anirudh Siddula

Ln 4, Col 2 160% Windows (CRLF) UTF-8



```
root@sandbox-hdp:~  
INFO : Compiling command(queryId=hive_20211009135329_b6b6c34e-61f3-4140-b62b-218c9c1e7d68): use tes  
t_db  
INFO : Semantic Analysis Completed (retrial = false)  
INFO : Returning Hive schema: Schema(fieldSchemas:null, properties:null)  
INFO : Completed compiling command(queryId=hive_20211009135329_b6b6c34e-61f3-4140-b62b-218c9c1e7d68  
); Time taken: 0.014 seconds  
INFO : Executing command(queryId=hive_20211009135329_b6b6c34e-61f3-4140-b62b-218c9c1e7d68): use tes  
t_db  
INFO : Starting task [Stage-0:DDL] in serial mode  
INFO : Completed executing command(queryId=hive_20211009135329_b6b6c34e-61f3-4140-b62b-218c9c1e7d68  
); Time taken: 0.007 seconds  
INFO : OK  
No rows affected (0.071 seconds)  
0: jdbc:hive2://sandbox-hdp.hortonworks.com:2> show tables;  
INFO : Compiling command(queryId=hive_20211009135336_0e42ff88-ce59-4ec5-9828-355231cab705): show ta  
bles  
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=hive_20211009135336_0e42ff88-ce59-4ec5-9828-355231cab705  
); Time taken: 0.044 seconds  
INFO : Executing command(queryId=hive_20211009135336_0e42ff88-ce59-4ec5-9828-355231cab705): show ta  
bles  
INFO : Starting task [Stage-0:DDL] in serial mode  
INFO : Completed executing command(queryId=hive_20211009135336_0e42ff88-ce59-4ec5-9828-355231cab705  
); Time taken: 0.013 seconds  
INFO : OK  
+-----+  
| tab_name |  
+-----+  
| employee_data |  
| sales_by_paymenttype |  
| sales_dyn_part |  
| salesjan2009 |  
| salesjan2009_by_script |  
| salesjan2009_ext |  
| salesjan2009_ext_without_loc |  
| salesjan2009_part |  
| salesjan2009_partition |  
+-----+  
9 rows selected (0.097 seconds)  
0: jdbc:hive2://sandbox-hdp.hortonworks.com:2>
```

Group E - Notepad

Data Technology Solutions
Group E

Bibek Shah Shankhar
Davinderjit Singh
Kirandeep Kaur
Kamaldeep kaur Dhami
Anirudh Siddula

Ln 4, Col 2 160% Windows (CRLF) UTF-8



```
root@sandbox-hdp:~  
d, type:string, comment:null), FieldSchema(name:salesjan2009_by_script.salesamount, type:int, comment:null), FieldSchema(name:salesjan2009_by_script.paymenttype, type:string, comment:null), FieldSchema(name:salesjan2009_by_script.customername, type:string, comment:null), FieldSchema(name:salesjan2009_by_script.city, type:string, comment:null), FieldSchema(name:salesjan2009_by_script.region, type:string, comment:null), FieldSchema(name:salesjan2009_by_script.country, type:string, comment:null)], properties:null)  
INFO : Completed compiling command(queryId=hive_20211009135428_59c0faf1-d4aa-487a-8c61-2ba15df85538); Time taken: 0.194 seconds  
INFO : Executing command(queryId=hive_20211009135428_59c0faf1-d4aa-487a-8c61-2ba15df85538): select * from salesjan2009_by_script limit 5  
INFO : Completed executing command(queryId=hive_20211009135428_59c0faf1-d4aa-487a-8c61-2ba15df85538); Time taken: 0.004 seconds  
INFO : OK  
+-----+-----+-----+  
+-----+-----+-----+  
| salesjan2009_by_script.productid | salesjan2009_by_script.salesamount | salesjan2009_by_script.paymenttype | salesjan2009_by_script.customername | salesjan2009_by_script.city | salesjan2009_by_script.region | salesjan2009_by_script.country |  
+-----+-----+-----+  
+-----+-----+-----+  
| Product1 | carolina | 1200 | Basildon | Mastercard | England  
|          | United Kingdom |           |           |           |  
| Product1 | Betina | 1200 | Parkville | Visa | MO  
|          | United States |           |           |           |  
| Product1 | Federica e Andrea | 1200 | Astoria | Mastercard | OR  
|          | United States |           |           |           |  
| Product1 | Gouya | 1200 | Echuca | Visa | Victoria  
|          | Australia |           |           |           |  
| Product2 | Gerd W | 3600 | Cahaba Heights | Visa | AL  
|          | United States |           |           |           |  
+-----+-----+-----+  
+-----+-----+-----+  
5 rows selected (0.23 seconds)  
0: jdbc:hive2://sandbox-hdp.hortonworks.com:2>
```

Group E - Notepad

Data Technology Solutions
Group E

Bibek Shah Shankhar
Davinderjit Singh
Kirandeep Kaur
Kamaldeep kaur Dhami
Anirudh Siddula

Ln 4, Col 2 160% Windows (CRLF) UTF-8



```
[root@sandbox-hdp:~]# hdfs dfs -ls /warehouse/tablespace/managed/hive/test_db.db
Found 7 items
drwxrwxrwx+ - hive hadoop          0 2021-10-09 13:27 /warehouse/tablespace/managed/hive/test_db.db
drwxrwxrwx+ - hive hadoop          0 2021-10-09 12:40 /warehouse/tablespace/managed/hive/test_db.db
drwxrwxrwx+ - hive hadoop          0 2021-10-09 13:05 /warehouse/tablespace/managed/hive/test_db.db
drwxrwxrwx+ - hive hadoop          0 2021-10-09 11:58 /warehouse/tablespace/managed/hive/test_db.db
drwxrwxrwx+ - hive hadoop          0 2021-10-09 13:52 /warehouse/tablespace/managed/hive/test_db.db
drwxrwxrwx+ - hive hadoop          0 2021-10-09 12:51 /warehouse/tablespace/managed/hive/test_db.db
drwxrwxrwx+ - hive hadoop          0 2021-10-09 13:00 /warehouse/tablespace/managed/hive/test_db.db
[salesjan2009_partition
[root@sandbox-hdp ~]# hdfs dfs -ls /warehouse/tablespace/managed/hive/test_db.db/salesjan2009_by_script
Found 1 items
drwxrwxrwx+ - hive hadoop          0 2021-10-09 13:52 /warehouse/tablespace/managed/hive/test_db.db
/salesjan2009_by_script/delta_0000001_0000001_0000
[root@sandbox-hdp ~]#
```

Group E - Notepad

Data Technology Solutions
Group E

Bibek Shah Shankhar
Davinderjit Singh
Kirandeep Kaur
Kamaldeep kaur Dhami
Anirudh Siddula

Ln 4, Col 2 160% Windows (CRLF) UTF-8



```
[root@sandbox-hdp:~]# hdfs dfs -cat /warehouse/tablespace/managed/hive/test_db.db/salesjan2009_by_script/delta_0000001_0000001_0000
cat: `/warehouse/tablespace/managed/hive/test_db.db/salesjan2009_by_script/delta_0000001_0000001_0000
0': Is a directory
[root@sandbox-hdp:~]# clear
[root@sandbox-hdp:~]# hdfs dfs -cat /warehouse/tablespace/managed/hive/test_db.db/salesjan2009_by_script/delta_0000001_0000001_0000/SalesJan2009.csv
Product1,1200,Mastercard,carolina,Basildon,England,United Kingdom
Product1,1200,Visa,Betina,Parkville ,MO,United States
Product1,1200,Mastercard,Federica e Andrea,Astoria ,OR,United States
Product1,1200,Visa,Gouya,Echuca,Victoria,Australia
Product2,3600,Visa,Gerd W ,Cahaba Heights ,AL,United States
Product1,1200,Visa,LAURENCE,Mickleton ,NJ,United States
Product1,1200,Mastercard,Fleur,Peoria ,IL,United States
Product1,1200,Mastercard,adam,Martin ,TN,United States
Product1,1200,Mastercard,Renee Elisabeth,Tel Aviv,Tel Aviv,Israel
Product1,1200,Visa,Aidan,Chatou,Ile-de-France,France
Product1,1200,Diners,Stacy,New York ,NY,United States
Product1,1200,Amex,Heidi,Eindhoven,Noord-Brabant,Netherlands
Product1,1200,Mastercard,Sean ,Shavano Park ,TX,United States
Product1,1200,Visa,Georgia,Eagle ,ID,United States
Product1,1200,Visa,Richard,Riverside ,NJ,United States
Product1,1200,Diners,Leanne,Julianstown,Meath,Ireland
Product1,1200,Visa,Janet,Ottawa,Ontario,Canada
Product1,1200,Diners,barbara,Hyderabad,Andhra Pradesh,India
Product2,3600,Visa,Sabine,London,England,United Kingdom
Product1,1200,Diners,Hani,Salt Lake City ,UT,United States
Product1,1200,Visa,Jeremy,Manchester,England,United Kingdom
Product1,1200,Diners,Janis,Ballynora,Cork,Ireland
Product1,1200,Mastercard,Nicola,Roodepoort,Gauteng,South Africa
Product1,1200,Visa,asuman,Chula Vista ,CA,United States
Product1,1200,Mastercard,Lena,Kuopio,Ita-Suomen Laani,Finland
Product1,1200,Visa,Lisa,Sugar Land ,TX,United States
Product1,1200,Diners,Bryan Kerrene,New York ,NY,United States
Product1,1200,Visa,chris,London,England,United Kingdom
Product1,1200,Visa,Maxine,Morton ,IL,United States
Product1,1200,Visa,Family,Los Gatos ,CA,United States
Product1,1200,Mastercard,Katherine,New York ,NY,United States
Product1,1200,Mastercard,Linda,Miami ,FL,United States
Product1,1200,Mastercard,SYLVIA,Vesenaz,Geneve,Switzerland
Product1,1200,Diners,Sheila,Brooklyn ,NY,United States
Product1,1200,Mastercard,Stephanie,Badhoevedorp,Noord-Holland,Netherlands
```

Group E - Notepad

Data Technology Solutions
Group E

Bibek Shah Shankhar
Davinderjit Singh
Kirandeep Kaur
Kamaldeep kaur Dhami
Anirudh Siddula

Ln 4, Col 2 160% Windows (CRLF) UTF-8



 Group E - Notepad

File Edit Format View Help

Data Technology Solutions
Group E

Bibek Shah Shankhar
Davinderjit Singh
Kirandeep Kaur
Kamaldeep kaur Dhami
Anirudh Siddula

Group E - Notepad

File Edit Format View Help

Data Technology Solutions

Group E

Bibek Shah Shankhar
Davinderjit Singh
Kirandeep Kaur
Kamaldeep kaur Dhami
Anirudh Siddula



root@sandbox-hdp:~

```
[root@sandbox-hdp ~]# hdfs dfs -ls /warehouse/tablespace/managed/hive/test_db.db
Found 8 items
drwxrwxrwx+  - hive hadoop          0 2021-10-09 13:27 /warehouse/tablespace/
managed/hive/test_db.db/employee_data
drwxrwxrwx+  - hive hadoop          0 2021-10-09 12:40 /warehouse/tablespace/
managed/hive/test_db.db/sales_by_paymenttype
drwxrwxrwx+  - hive hadoop          0 2021-10-09 13:05 /warehouse/tablespace/
managed/hive/test_db.db/sales_dyn_part
drwxrwxrwx+  - hive hadoop          0 2021-10-09 11:58 /warehouse/tablespace/
managed/hive/test_db.db/salesjan2009
drwxrwxrwx+  - hive hadoop          0 2021-10-09 13:52 /warehouse/tablespace/
managed/hive/test_db.db/salesjan2009_by_script
drwxrwxrwx+  - hive hadoop          0 2021-10-09 14:01 /warehouse/tablespace/
managed/hive/test_db.db/salesjan2009_clustered
drwxrwxrwx+  - hive hadoop          0 2021-10-09 12:51 /warehouse/tablespace/
managed/hive/test_db.db/salesjan2009_part
drwxrwxrwx+  - hive hadoop          0 2021-10-09 13:00 /warehouse/tablespace/
managed/hive/test_db.db/salesjan2009_partition
[root@sandbox-hdp ~]#
```

Group E - Notepad

File Edit Format View Help

Data Technology Solutions Group E

Bibek Shah Shankhar
Davinderjit Singh
Kirandeep Kaur
Kamaldeep kaur Dhami
Anirudh Siddula

Ln 4, Col 2 160% Windows (CRLF) UTF-8



7:47 PM
09-Oct-21 1

Data Technology Solutions
Group E

Bibek Shah Shankhar
Davinderjit Singh
Kirandeep Kaur
Kamaldeep kaur Dhami
Anirudh Siddula



```
root@sandbox-hdp:~  
INFO :     SHUFFLE_BYTES_TO_MEM: 0  
INFO :     SHUFFLE_PHASE_TIME: 533  
INFO : TaskCounter_Reducer_2_OUTPUT_Reducer_3:  
INFO :     ADDITIONAL_SPILLS_BYTES_READ: 0  
INFO :     ADDITIONAL_SPILLS_BYTES_WRITTEN: 0  
INFO :     ADDITIONAL_SPILL_COUNT: 0  
INFO :     OUTPUT_BYTES: 9031  
INFO :     OUTPUT_BYTES_PHYSICAL: 7579  
INFO :     OUTPUT_BYTES_WITH_OVERHEAD: 9131  
INFO :     OUTPUT_LARGE_RECORDS: 0  
INFO :     OUTPUT_RECORDS: 10  
INFO :     SPILLED_RECORDS: 0  
INFO : TaskCounter_Reducer_3_INPUT_Reducer_2:  
INFO :     FIRST_EVENT RECEIVED: 3  
INFO :     INPUT_RECORDS PROCESSED: 10  
INFO :     LAST_EVENT RECEIVED: 6080  
INFO :     NUM FAILED SHUFFLE_INPUTS: 0  
INFO :     NUM SHUFFLED_INPUTS: 10  
INFO :     SHUFFLE_BYTES: 7339  
INFO :     SHUFFLE_BYTES_DECOMPRESSED: 9131  
INFO :     SHUFFLE_BYTES_DISK_DIRECT: 7339  
INFO :     SHUFFLE_BYTES_TO_DISK: 0  
INFO :     SHUFFLE_BYTES_TO_MEM: 0  
INFO :     SHUFFLE_PHASE_TIME: 6083  
INFO : TaskCounter_Reducer_3_OUTPUT_out_Reducer_3:  
INFO :     OUTPUT_RECORDS: 0  
INFO : org.apache.hadoop.hive.ql.exec.tez.HiveInputCounters:  
INFO :     GROUPED_INPUT_SPLITS_Map_1: 1  
INFO :     INPUT_DIRECTORIES_Map_1: 1  
INFO :     INPUT_FILES_Map_1: 1  
INFO :     RAW_INPUT_SPLITS_Map_1: 1  
INFO : Starting task [Stage-2:DEPENDENCY_COLLECTION] in serial mode  
INFO : Starting task [Stage-0:MOVE] in serial mode  
INFO : Loading data to table test_db.salesjan2009_clustered from hdfs://sandbox-hdp.hortonworks.com:8020/warehouse/tablespace/managed/hive/test_db.db/salesjan2009_clustered/.hive-staging_hive_2021-10-09_14-04-11_119_7534238076599760864-2/-ext-10000  
INFO : Starting task [Stage-3:STATS] in serial mode  
INFO : Completed executing command(queryId=hive_20211009140411_6db25946-41e1-456e-b835-3442466ec799); Time taken: 23.328 seconds  
INFO : OK  
No rows affected (23.57 seconds)  
0: jdbc:hive2://sandbox-hdp.hortonworks.com:2>
```

```
Group E - Notepad  
File Edit Format View Help  
Data Technology Solutions  
Group E  
  
Bibek Shah Shankhar  
Davinderjit Singh  
Kirandeep Kaur  
Kamaldeep kaur Dhami  
Anirudh Siddula  
  
Ln 4, Col 2 160% Windows (CRLF) UTF-8
```



```
root@sandbox-hdp:~  
0: jdbc:hive2://sandbox-hdp.hortonworks.com:2>  
0: jdbc:hive2://sandbox-hdp.hortonworks.com:2> describe formatted salesjan2009_clustered;  
INFO : Compiling command(queryId=hive_20211009140508_7d345cbf-a640-479a-9fc3-1605bccb7f55): describe formatted salesjan2009_clustered  
INFO : Semantic Analysis Completed (retrial = false)  
INFO : Returning Hive schema: Schema(fieldSchemas:[FieldSchema(name:col_name, type:string, comment:from deserializer), FieldSchema(name:data_type, type:string, comment:from deserializer), FieldSchema(name:comment, type:string, comment:from deserializer)], properties:null)  
INFO : Completed compiling command(queryId=hive_20211009140508_7d345cbf-a640-479a-9fc3-1605bccb7f55); Time taken: 0.196 seconds  
INFO : Executing command(queryId=hive_20211009140508_7d345cbf-a640-479a-9fc3-1605bccb7f55): describe formatted salesjan2009_clustered  
INFO : Starting task [Stage-0:DDL] in serial mode  
INFO : Completed executing command(queryId=hive_20211009140508_7d345cbf-a640-479a-9fc3-1605bccb7f55); Time taken: 0.022 seconds  
INFO : OK  
+-----+-----+  
| col_name | data_type |  
+-----+-----+  
| # col_name | data_type | comment |  
| productid | string |  
| salesamount | int |  
| paymenttype | string |  
| customername | string |  
| city | string |  
| region | string |  
| country | string |  
| | NULL |  
| # Detailed Table Information | NULL |  
| |
```

Group E - Notepad

Data Technology Solutions
Group E

Bibek Shah Shankhar
Davinderjit Singh
Kirandeep Kaur
Kamaldeep kaur Dhami
Anirudh Siddula

Ln 4, Col 2 160% Windows (CRLF) UTF-8



```
root@sandbox-hdp:~  
| bucketing_version | 2  
| numFiles | 10  
| numRows | 998  
| rawDataSize | 0  
| totalSize | 30920  
| transactional | true  
| transactional_properties | default  
| transient_lastDdlTime | 1633788274  
| NULL | NULL  
| NULL | NULL  
| # Storage Information |  
| SerDe Library: | org.apache.hadoop.hive.ql.io.orc.OrcSerde |  
| InputFormat: | org.apache.hadoop.hive.ql.io.orc.OrcInputFormat |  
| OutputFormat: | org.apache.hadoop.hive.ql.io.orc.OrcOutputFormat |  
| Compressed: | No |  
| Num Buckets: | 10 |  
| Bucket Columns: | [country] |  
| Sort Columns: | [] |  
| Storage Desc Params: |  
| | serialization.format | 1  
+-----+  
-----+  
39 rows selected (0.239 seconds)  
0: jdbc:hive2://sandbox-hdp.hortonworks.com:2>
```

Group E - Notepad

Data Technology Solutions
Group E

Bibek Shah Shankhar
Davinderjit Singh
Kirandeep Kaur
Kamaldeep kaur Dhami
Anirudh Siddula

Ln 4, Col 2 160% Windows (CRLF) UTF-8



7:50 PM
09-Oct-21 1

```
[root@sandbox-hdp ~]# hdfs dfs -ls /warehouse/tablespace/managed/hive/test_db.db/salesjan2009_clustered
Found 1 items
drwxrwxrwx+  - hive hadoop          0 2021-10-09 14:04 /warehouse/tablespace/managed/hive/test_db.db/salesjan2009_clustered/delta_0000001_0000001_0000
[root@sandbox-hdp ~]# hdfs dfs -ls /warehouse/tablespace/managed/hive/test_db.db/salesjan2009_clustered/delta_0000001_0000001_0000
Found 11 items
-rw-rw-rw+  1 hive hadoop          1 2021-10-09 14:04 /warehouse/tablespace/managed/hive/test_db.db/salesjan2009_clustered/delta_0000001_0000001_0000/_orc_acid_version
-rw-rw-rw+  1 hive hadoop         2535 2021-10-09 14:04 /warehouse/tablespace/managed/hive/test_db.db/salesjan2009_clustered/delta_0000001_0000001_0000/bucket_00000
-rw-rw-rw+  1 hive hadoop         3571 2021-10-09 14:04 /warehouse/tablespace/managed/hive/test_db.db/salesjan2009_clustered/delta_0000001_0000001_0000/bucket_00001
-rw-rw-rw+  1 hive hadoop         1765 2021-10-09 14:04 /warehouse/tablespace/managed/hive/test_db.db/salesjan2009_clustered/delta_0000001_0000001_0000/bucket_00002
-rw-rw-rw+  1 hive hadoop         2475 2021-10-09 14:04 /warehouse/tablespace/managed/hive/test_db.db/salesjan2009_clustered/delta_0000001_0000001_0000/bucket_00003
-rw-rw-rw+  1 hive hadoop         8841 2021-10-09 14:04 /warehouse/tablespace/managed/hive/test_db.db/salesjan2009_clustered/delta_0000001_0000001_0000/bucket_00004
-rw-rw-rw+  1 hive hadoop         1784 2021-10-09 14:04 /warehouse/tablespace/managed/hive/test_db.db/salesjan2009_clustered/delta_0000001_0000001_0000/bucket_00005
-rw-rw-rw+  1 hive hadoop         3138 2021-10-09 14:04 /warehouse/tablespace/managed/hive/test_db.db/salesjan2009_clustered/delta_0000001_0000001_0000/bucket_00006
-rw-rw-rw+  1 hive hadoop         2438 2021-10-09 14:04 /warehouse/tablespace/managed/hive/test_db.db/salesjan2009_clustered/delta_0000001_0000001_0000/bucket_00007
-rw-rw-rw+  1 hive hadoop         2135 2021-10-09 14:04 /warehouse/tablespace/managed/hive/test_db.db/salesjan2009_clustered/delta_0000001_0000001_0000/bucket_00008
-rw-rw-rw+  1 hive hadoop         2238 2021-10-09 14:04 /warehouse/tablespace/managed/hive/test_db.db/salesjan2009_clustered/delta_0000001_0000001_0000/bucket_00009
[root@sandbox-hdp ~]#
```

Group E - Notepad

Data Technology Solutions
Group E

Bibek Shah Shankhar
Davinderjit Singh
Kirandeep Kaur
Kamaldeep kaur Dhami
Anirudh Siddula

Ln 4, Col 2 160% Windows (CRLF) UTF-8



```
root@sandbox-hdp:~# hdfs dfs -cat /warehouse/tablespace/managed/hive/test_db.db/salesjan2009_clustered/delta_0000001_0000001_0000/bucket_00009
ORC
P+
P+
BP>Rbf `` Phh8 "%&-
P+
BP
PD Rbf `` PT (O)M.1 %&03
u P=
"
AmexVisa P@ be!1E%.u\! PT/L
f2 be!-E%.~.
<LT! J] | '/BbKU
(
!
lsaceUpper Normandy PO
%
"
ArgentinMalaysia P
FP LcLQ8 @! AF 3! QFAF ` R
A
0 UBX+V*?rY
3N .H j %
H, / (+Oi M <o $ 2 t5 JQ T v # {2 } S : E Z 9 ds C T ic W ' F YTeu SFuDcQuDV E ] An1
E r
*T U
Nc " o4t ` m . e4w \ )) ) Tnb = tc & 0q O LT L se u F , k < i @
R PNF v M H t 8 } S V G I Gy E ? Q
I r 0w *vA 5 9N
DieYh ' F : L JCJZA Mt ! N

Q
0 ~ E a I 61 b < U f 3 r CC B + A 9
^ b 1 8 éq` [ MGQe 0 ] > & 6 TV1 ! 1
r ; nzY$ < r Y4rr 1B B @ FPFA
```

Group E - Notepad

File Edit Format View Help

Data Technology Solutions
Group E

Bibek Shah Shankhar
Davinderjit Singh
Kirandeep Kaur
Kamaldeep kaur Dhami
Anirudh Siddula

Ln 4, Col 2 160% Windows (CRLF) UTF-8



Group E - Notepad

File Edit Format View Help

Data Technology Solutions

Group E

Bibek Shah Shankhar
Davinderjit Singh
Kirandeep Kaur
Kamaldeep kaur Dhami
Anirudh Siddula



```
root@sandbox-hdp:~  
..... .> region string,  
..... .> country string)  
..... .> row format delimited  
..... .> fields terminated by ','  
..... .> stored as textfile  
..... .> location '/user/bibek/Hive/warehouse';  
INFO : Compiling command(queryId=hive_20211009141234_f46080a8-cfb3-4a93-8007-ac10621caa4b): create  
external table SalesJan2009_external(  
productid string,  
salesamount int,  
paymenttype string,  
customername string,  
city string,  
region string,  
country string)  
row format delimited  
fields terminated by ','  
stored as textfile  
location '/user/bibek/Hive/warehouse'  
INFO : Semantic Analysis Completed (retrial = false)  
INFO : Returning Hive schema: Schema(fieldSchemas:null, properties:null)  
INFO : Completed compiling command(queryId=hive_20211009141234_f46080a8-cfb3-4a93-8007-ac10621caa4b  
); Time taken: 0.024 seconds  
INFO : Executing command(queryId=hive_20211009141234_f46080a8-cfb3-4a93-8007-ac10621caa4b): create  
external table SalesJan2009_external(  
productid string,  
salesamount int,  
paymenttype string,  
customername string,  
city string,  
region string,  
country string)  
row format delimited  
fields terminated by ','  
stored as textfile  
location '/user/bibek/Hive/warehouse'  
INFO : Starting task [Stage-0:DDL] in serial mode  
INFO : Completed executing command(queryId=hive_20211009141234_f46080a8-cfb3-4a93-8007-ac10621caa4b  
); Time taken: 0.06 seconds  
INFO : OK  
No rows affected (0.165 seconds)  
0: jdbc:hive2://sandbox-hdp.hortonworks.com:2>
```

Group E - Notepad

Data Technology Solutions
Group E

Bibek Shah Shankhar
Davinderjit Singh
Kirandeep Kaur
Kamaldeep kaur Dhami
Anirudh Siddula

Ln 4, Col 2 160% Windows (CRLF) UTF-8



```
root@sandbox-hdp:~  
0: jdbc:hive2://sandbox-hdp.hortonworks.com:2> show tables;  
INFO : Compiling command(queryId=hive_20211009141319_4d4c5fdc-b886-4720-a2e9-8d5d7d098528): show ta  
bles  
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=hive_20211009141319_4d4c5fdc-b886-4720-a2e9-8d5d7d098528  
); Time taken: 0.015 seconds  
INFO : Executing command(queryId=hive_20211009141319_4d4c5fdc-b886-4720-a2e9-8d5d7d098528): show ta  
bles  
INFO : Starting task [Stage-0:DDL] in serial mode  
INFO : Completed executing command(queryId=hive_20211009141319_4d4c5fdc-b886-4720-a2e9-8d5d7d098528  
); Time taken: 0.008 seconds  
INFO : OK  
+-----+  
| tab_name |  
+-----+  
| employee_data |  
| sales_by_paymenttype |  
| sales_dyn_part |  
| salesjan2009 |  
| salesjan2009_by_script |  
| salesjan2009_clustered |  
| salesjan2009_ext |  
| salesjan2009_ext without_loc |  
| salesjan2009_external |  
| salesjan2009_part |  
| salesjan2009_partition |  
+-----+  
11 rows selected (0.031 seconds)  
0: jdbc:hive2://sandbox-hdp.hortonworks.com:2>
```

```
Group E - Notepad  
File Edit Format View Help  
Data Technology Solutions  
Group E  
  
Bibek Shah Shankhar  
Davinderjit Singh  
Kirandeep Kaur  
Kamaldeep kaur Dhami  
Anirudh Siddula  
  
Ln 4, Col 2 160% Windows (CRLF) UTF-8
```

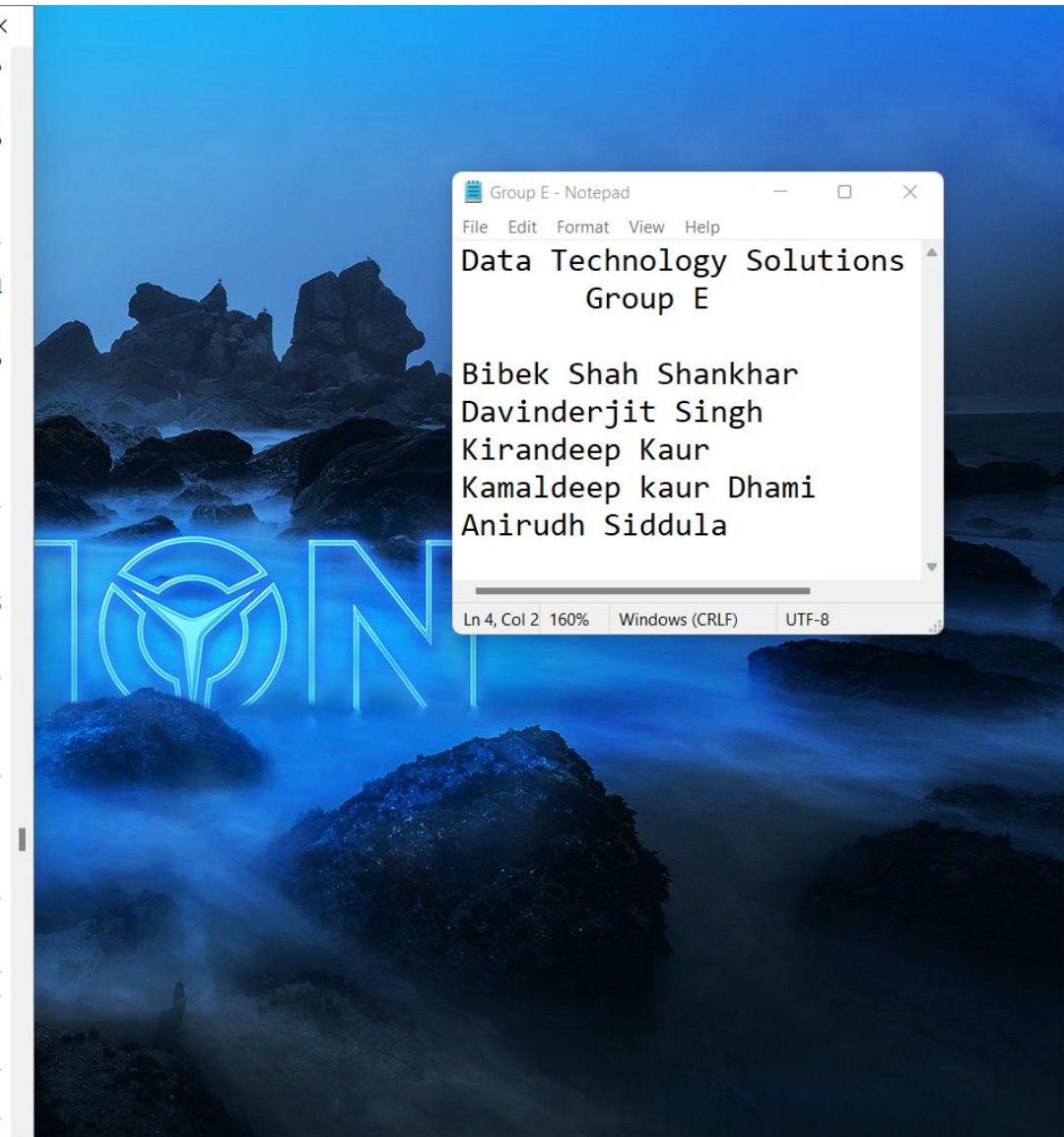


```

root@sandbox-hdp:~ 
0: jdbc:hive2://sandbox-hdp.hortonworks.com:2> select country, count(*) as co
unt from SalesJan2009_ext group by (country) order by count;
INFO : Compiling command(queryId=hive_20211009152709_82d8ee6f-91ad-4cb1-b434
-bb2c5a9a9243): select country, count(*) as count from SalesJan2009_ext group
by (country) order by count
INFO : Semantic Analysis Completed (retrial = false)
INFO : Returning Hive schema: Schema(fieldSchemas:[FieldSchema(name:country,
type:string, comment:null), FieldSchema(name:count, type:bigint, comment:nul
l)], properties:null)
INFO : Completed compiling command(queryId=hive_20211009152709_82d8ee6f-91ad
-4cb1-b434-bb2c5a9a9243); Time taken: 0.236 seconds
INFO : Executing command(queryId=hive_20211009152709_82d8ee6f-91ad-4cb1-b434
-bb2c5a9a9243): select country, count(*) as count from SalesJan2009_ext group
by (country) order by count
INFO : Query ID = hive_20211009152709_82d8ee6f-91ad-4cb1-b434-bb2c5a9a9243
INFO : Total jobs = 1
INFO : Launching Job 1 out of 1
INFO : Starting task [Stage-1:MAPRED] in serial mode
INFO : Subscribed to counters: [] for queryId: hive_20211009152709_82d8ee6f-
91ad-4cb1-b434-bb2c5a9a9243
INFO : Tez session hasn't been created yet. Opening session
INFO : Dag name: select country, count(*) as count fr...count (Stage-1)
INFO : Status: Running (Executing on YARN cluster with App id application_16
33778387453_0010)

```

	VERTICES FAILED	MODE	STATUS	TOTAL	COMPLETED	RUNNING	PENDING
	KILLED						
Map 1	0	container	RUNNING	1	0	1	0
Reducer 2	0	container	INITED	2	0	0	2
		---	INITED	1	0	0	1
	VERTICES FAILED	MODE	STATUS	TOTAL	COMPLETED	RUNNING	PENDING
	KILLED	---	---	---	---	---	---
Map 1	0	[>>-----]	0%	ELAPSED TIME: 3.30 s			
Reducer 2	0	container	RUNNING	1	0	1	0
		---	INITED	2	0	0	2



root@sandbox-hdp:~

```
INFO :     RAW_INPUT_SPLITS_Map_1: 1
INFO : Completed executing command(queryId=hive_20211009152709_82d8ee6f-91ad
-4cb1-b434-bb2c5a9a9243); Time taken: 95.555 seconds
INFO : OK
+-----+-----+
|   country    | count |
+-----+-----+
| Kuwait      | 1
| Bahrain     | 1
| Bermuda     | 1
| Moldova     | 1
| Bulgaria    | 1
| Cayman Isls | 1
| Mauritius   | 1
| Dominican Republic | 1
| Hong Kong   | 1
| Russia      | 1
| Israel       | 1
| Argentina   | 1
| China        | 1
| Costa Rica  | 1
| Greece       | 1
| Guatemala   | 1
| Iceland     | 1
| Jersey       | 1
| Latvia       | 1
| Luxembourg  | 1
| Malaysia    | 1
| Romania     | 1
| South Korea | 1
| Ukraine     | 1
| Malta        | 2
| Poland       | 2
| Philippines | 2
| Finland     | 2
| The Bahamas | 2
| India        | 2
| Japan        | 2
| Monaco       | 2
| Thailand     | 2
| Czech Republic | 3
| Hungary     | 3
| South Africa | 5
| Brazil       | 5
```

Group E - Notepad

File Edit Format View Help

Data Technology Solutions Group E

Bibek Shah Shankhar
Davinderjit Singh
Kirandeep Kaur
Kamaldeep kaur Dhami
Anirudh Siddula

Ln 4, Col 2 160% Windows (CRLF) UTF-8



9:14 PM
09-Oct-21

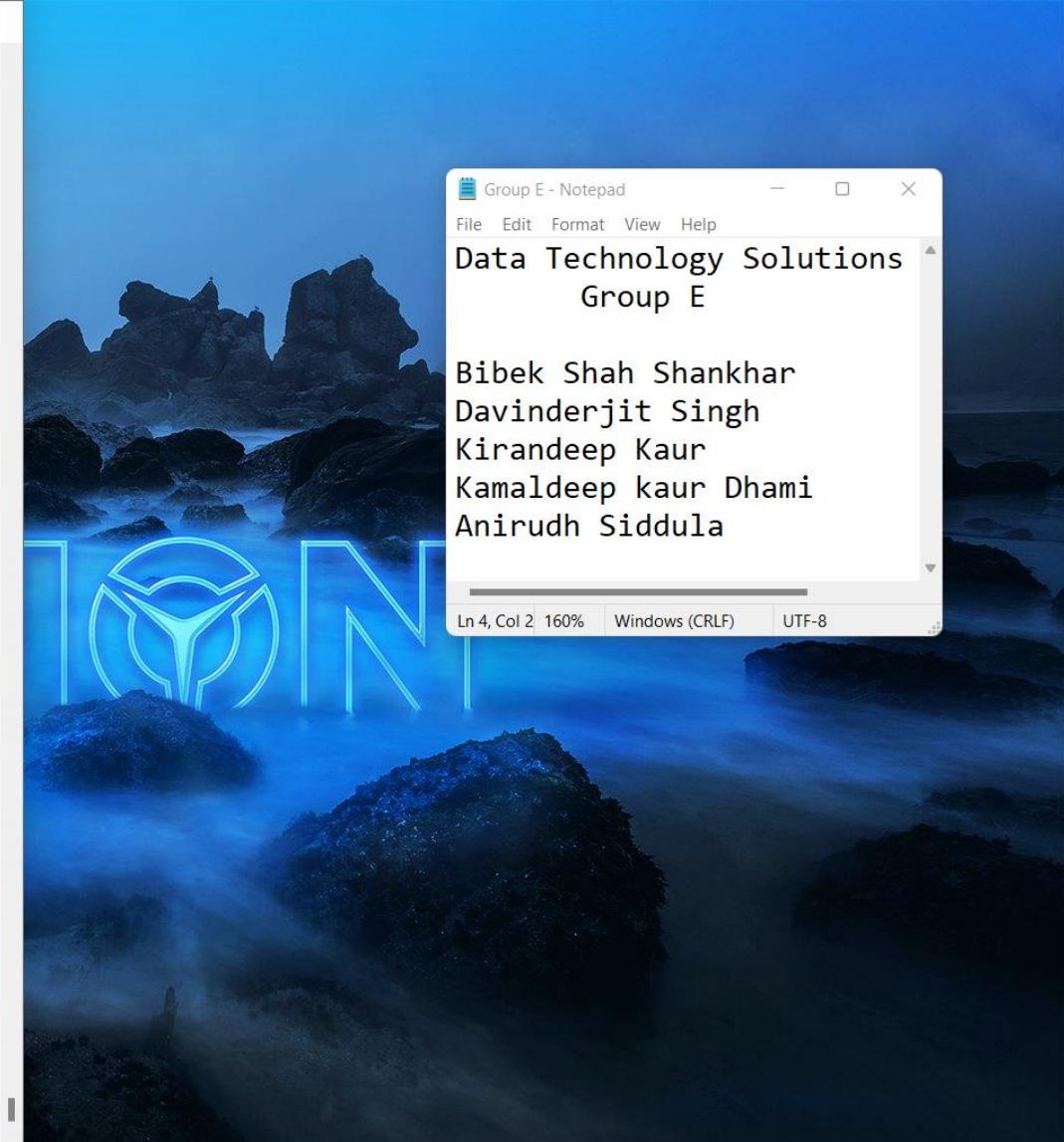
```
root@sandbox-hdp:~
```

Guatemala	1
Iceland	1
Jersey	1
Latvia	1
Luxembourg	1
Malaysia	1
Romania	1
South Korea	1
Ukraine	1
Malta	2
Poland	2
Philippines	2
Finland	2
The Bahamas	2
India	2
Japan	2
Monaco	2
Thailand	2
Czech Republic	3
Hungary	3
South Africa	5
Brazil	5
Turkey	6
New Zealand	6
United Arab Emirates	6
Austria	7
Belgium	8
Spain	12
Sweden	13
Italy	15
Denmark	15
Norway	16
Netherlands	22
Germany	25
France	27
Switzerland	36
Australia	38
Ireland	49
Canada	76
United Kingdom	100
United States	463

56 rows selected (96.127 seconds)

0: jdbc:hive2://sandbox-hdp.hortonworks.com:2>

```
- □ ×
```



Group E - Notepad

File Edit Format View Help

Data Technology Solutions Group E

Bibek Shah Shankhar
Davinderjit Singh
Kirandeep Kaur
Kamaldeep kaur Dhami
Anirudh Siddula

Ln 4, Col 2 160% Windows (CRLF) UTF-8



9:14 PM
09-Oct-21

Group E - Notepad

File Edit Format View Help

Data Technology Solutions

Group E

Bibek Shah Shankhar
Davinderjit Singh
Kirandeep Kaur
Kamaldeep kaur Dhami
Anirudh Siddula

Ln 4, Col 2 160% Windows (CRLF) UTF-8



```
root@sandbox-hdp:~  
.....> country string)  
.....> skewed by (country) on ('United States','United Kingdom')  
.....> stored as directories  
.....> row format delimited  
.....> fields terminated by ',';  
INFO : Compiling command(queryId=hive_20211009153329_62295e49-ffd6-41c5-8cb2  
-albb735ea3le); create table salesjan2009_skewed(  
productid string,  
salesamount int,  
paymenttype string,  
customername string,  
city string,  
region string,  
country string)  
skewed by (country) on ('United States','United Kingdom')  
stored as directories  
row format delimited  
fields terminated by ','  
INFO : Semantic Analysis Completed (retrial = false)  
INFO : Returning Hive schema: Schema(fieldSchemas:null, properties:null)  
INFO : Completed compiling command(queryId=hive_20211009153329_62295e49-ffd6  
-41c5-8cb2-albb735ea3le); Time taken: 0.221 seconds  
INFO : Executing command(queryId=hive_20211009153329_62295e49-ffd6-41c5-8cb2  
-albb735ea3le); create table salesjan2009_skewed(  
productid string,  
salesamount int,  
paymenttype string,  
customername string,  
city string,  
region string,  
country string)  
skewed by (country) on ('United States','United Kingdom')  
stored as directories  
row format delimited  
fields terminated by ','  
INFO : Starting task [Stage-0:DDL] in serial mode  
INFO : Completed executing command(queryId=hive_20211009153329_62295e49-ffd6  
-41c5-8cb2-albb735ea3le); Time taken: 0.821 seconds  
INFO : OK  
No rows affected (1.235 seconds)  
0: jdbc:hive2://sandbox-hdp.hortonworks.com:2>
```

Group E - Notepad

Data Technology Solutions
Group E

Bibek Shah Shankhar
Davinderjit Singh
Kirandeep Kaur
Kamaldeep kaur Dhami
Anirudh Siddula

Ln 4, Col 2 160% Windows (CRLF) UTF-8



9:19 PM
09-Oct-21

```
| salesjan2009_part |  
| salesjan2009_partition |  
| salesjan2009_skewed |  
+-----+  
12 rows selected (0.046 seconds)  
0: jdbc:hive2://sandbox-hdp.hortonworks.com:2> describe formatted salesjan2009_skewed  
.....>;  
INFO : Compiling command(queryId=hive_20211009153724_48d866b3-7e8f-444e-933d-b870177021ba): describe form  
atted salesjan2009_skewed  
INFO : Semantic Analysis Completed (retrial = false)  
INFO : Returning Hive schema: Schema(fieldSchemas:[FieldSchema(name:col_name, type:string, comment:from d  
eserializer), FieldSchema(name:data_type, type:string, comment:from deserializer), FieldSchema(name:commen  
t, type:string, comment:from deserializer)], properties:null)  
INFO : Completed compiling command(queryId=hive_20211009153724_48d866b3-7e8f-444e-933d-b870177021ba); Tim  
e taken: 0.024 seconds  
INFO : Executing command(queryId=hive_20211009153724_48d866b3-7e8f-444e-933d-b870177021ba): describe form  
atted salesjan2009_skewed  
INFO : Starting task [Stage-0:DDL] in serial mode  
INFO : Completed executing command(queryId=hive_20211009153724_48d866b3-7e8f-444e-933d-b870177021ba); Tim  
e taken: 0.452 seconds  
INFO : OK  
+-----+-----+  
|       col_name |           data_type |  
| comment        |  
+-----+-----+  
| # col_name      | data_type          | comment  
| productid       | string             |  
| salesamount     | int                |  
| paymenttype    | string             |  
| customername   | string             |  
| city            | string             |  
| region          | string             |  
| country         | string             |
```

Group E - Notepad

Edit Format View Help

ta Technology Solutions

Group E

bek Shah Shankhar

vinderjit Singh

randeep Kaur

maldeep kaur Dhami

irudh Siddula



```
root@sandbox-hdp:~  
| Table Parameters: | NULL | NULL  
| | COLUMN_STATS_ACCURATE | | {"BASIC_STATS":\"  
true\", \"COLUMN_STATS\":{\"city\":\"true\", \"country\":\"true\", \"customername\":\"true\", \"paymenttype\":  
\"true\", \"productid\":\"true\", \"region\":\"true\", \"salesamount\":\"true\"}} } |  
| | bucketing_version | 2  
| | numFiles | 0  
| | numRows | 0  
| | rawDataSize | 0  
| | totalSize | 0  
| | transactional | true  
| | transactional_properties | default  
| | transient_lastDdlTime | 1633793609  
| | NULL | NULL  
| # Storage Information | NULL | NULL  
| SerDe Library: | org.apache.hadoop.hive.ql.io.orc.OrcSerde | NULL  
| InputFormat: | org.apache.hadoop.hive.ql.io.orc.OrcInputFormat | NULL  
| OutputFormat: | org.apache.hadoop.hive.ql.io.orc.OrcOutputFormat | NULL  
| Compressed: | No | NULL  
| Num Buckets: | -1 | NULL  
| Bucket Columns: | [] | NULL  
| Sort Columns: | [] | NULL  
| Stored As SubDirectories: | Yes | NULL  
| Skewed Columns: | [country] | NULL
```

Group E - Notepad

Edit Format View Help

ta Technology Solutions
Group E

bek Shah Shankhar
vinderjit Singh
randeep Kaur
maldeep kaur Dhami
irudh Siddula

Col 2 160% Windows (CRLF) UTF-8



9:22 PM
09-Oct-21

```
root@sandbox-hdp:~  
| totalSize | 0  
| transactional | true  
| transactional_properties | default  
| transient_lastDdlTime | 1633793609  
| NULL | NULL  
| # Storage Information |  
| Storage Library: |  
| InputFormat: org.apache.hadoop.hive.ql.io.orc.OrcSerde | NULL  
| OutputFormat: org.apache.hadoop.hive.ql.io.orc.OrcInputFormat | NULL  
| Compressed: org.apache.hadoop.hive.ql.io.orc.OrcOutputFormat | NULL  
| No | NULL  
| Num Buckets: -1 | NULL  
| Bucket Columns: [] | NULL  
| Sort Columns: [] | NULL  
| Stored As SubDirectories: Yes | NULL  
| Skewed Columns: [country] | NULL  
| Skewed Values: [[united kingdom], [united states]] | NULL  
| Storage Desc Params: NULL | NULL  
| field.delim | ,  
| serialization.format | ,  
+-----+  
43 rows selected (0.509 seconds)  
0: jdbc:hive2://sandbox-hdp.hortonworks.com:2>
```

Group E - Notepad

ta Technology Solutions
Group E

bek Shah Shankhar
vinderjit Singh
randeep Kaur
maldeep kaur Dhami
irudh Siddula

Col 2 160% Windows (CRLF) UTF-8



9:22 PM
09-Oct-21

Group E - Notepad

Data Technology Solutions
Group E

Bibek Shah Shankhar
Davinderjit Singh
Kirandeep Kaur
Kamaldeep kaur Dhami
Anirudh Siddula



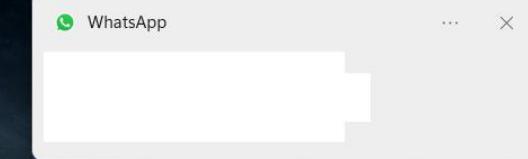
```
root@sandbox-hdp:~  
INFO :     INPUT_RECORDS_PROCESSED: 998  
INFO :     INPUT_SPLIT_LENGTH_BYTES: 65835  
INFO : TaskCounter_Map_1_OUTPUT Reducer_2:  
INFO :     ADDITIONAL_SPILLS_BYTES_READ: 0  
INFO :     ADDITIONAL_SPILLS_BYTES_WRITTEN: 0  
INFO :     ADDITIONAL_SPILL_COUNT: 0  
INFO :     OUTPUT_BYTES: 3269  
INFO :     OUTPUT_BYTES_PHYSICAL: 1795  
INFO :     OUTPUT_BYTES_WITH_OVERHEAD: 3279  
INFO :     OUTPUT_LARGE_RECORDS: 0  
INFO :     OUTPUT_RECORDS: 1  
INFO :     SPILLED_RECORDS: 0  
INFO : TaskCounter_Reducer_2_INPUT_Map_1:  
INFO :     FIRST_EVENT RECEIVED: 25  
INFO :     INPUT_RECORDS_PROCESSED: 1  
INFO :     LAST_EVENT RECEIVED: 25  
INFO :     NUM FAILED SHUFFLE INPUTS: 0  
INFO :     NUM SHUFFLED_INPUTS: 1  
INFO :     SHUFFLE_BYTES: 1771  
INFO :     SHUFFLE_BYTES_DECOMPRESSED: 3279  
INFO :     SHUFFLE_BYTES_DISK_DIRECT: 1771  
INFO :     SHUFFLE_BYTES_TO_DISK: 0  
INFO :     SHUFFLE_BYTES_TO_MEM: 0  
INFO :     SHUFFLE_PHASE_TIME: 33  
INFO : TaskCounter_Reducer_2_OUTPUT_out_Reducer_2:  
INFO :     OUTPUT_RECORDS: 0  
INFO : org.apache.hadoop.hive.ql.exec.tez.HiveInputCounters:  
INFO : GROUPED_INPUT_SPLITS_Map_1: 1  
INFO : INPUT_DIRECTORIES_Map_1: 1  
INFO : INPUT_FILES_Map_1: 1  
INFO : RAW_INPUT_SPLITS_Map_1: 1  
INFO : Starting task [Stage-2:DEPENDENCY_COLLECTION] in serial mode  
INFO : Starting task [Stage-0:MOVE] in serial mode  
INFO : Loading data to table test_db.salesjan2009 skewed from hdfs://sandbox-hdp.hortonworks.com:8020/war  
ehouse/tablespace/managed/hive/test_db.db/salesjan2009_skewed/.hive-staging_hive_2021-10-09_15-39-58_766_6  
580656727018752906-2/-ext-10000  
INFO : Starting task [Stage-3:STATS] in serial mode  
INFO : Completed executing command(queryId=hive_20211009153958_7a429712-3495-4f45-b2a7-4a2c8f096bde); Tim  
e taken: 12.461 seconds  
INFO : OK  
No rows affected (13.232 seconds)  
0: jdbc:hive2://sandbox-hdp.hortonworks.com:2>
```

Group E - Notepad

Data Technology Solutions
Group E

Bibek Shah Shankhar
Davinderjit Singh
Kirandeep Kaur
Kamaldeep kaur Dhami
Anirudh Siddula

Ln 4, Col 2 160% Windows (CRLF) UTF-8



9:26 PM
09-Oct-21

Data Technology Solutions
Group E

Bibek Shah Shankhar
Davinderjit Singh
Kirandeep Kaur
Kamaldeep kaur Dhami
Anirudh Siddula



```
root@sandbox-hdp:~  
0: jdbc:hive2://sandbox-hdp.hortonworks.com:2>  
0: jdbc:hive2://sandbox-hdp.hortonworks.com:2> describe formatted salesjan2009_parquet;  
INFO : Compiling command(queryId=hive_20211009154759_f0a2f856-2982-448f-91b2-8f98a13bc313): describe form  
atted salesjan2009_parquet  
INFO : Semantic Analysis Completed (retrial = false)  
INFO : Returning Hive schema: Schema(fieldSchemas:[FieldSchema(name:col_name, type:string, comment:from d  
eserializer), FieldSchema(name:data_type, type:string, comment:from deserializer), FieldSchema(name:commen  
t, type:string, comment:from deserializer)], properties:null)  
INFO : Completed compiling command(queryId=hive_20211009154759_f0a2f856-2982-448f-91b2-8f98a13bc313); Tim  
e taken: 0.323 seconds  
INFO : Executing command(queryId=hive_20211009154759_f0a2f856-2982-448f-91b2-8f98a13bc313): describe form  
atted salesjan2009_parquet  
INFO : Starting task [Stage-0:DDL] in serial mode  
INFO : Completed executing command(queryId=hive_20211009154759_f0a2f856-2982-448f-91b2-8f98a13bc313); Tim  
e taken: 0.075 seconds  
INFO : OK  
+-----+-----+-----+  
| col_name | data_type | comment |  
+-----+-----+-----+  
| # col_name | data_type | comment |  
| productid | string | |  
| salesamount | int | |  
| paymenttype | string | |  
| customername | string | |  
| city | string | |  
| region | string | |  
| country | string | |  
| | NULL | NULL |  
| # Detailed Table Information | NULL | NULL |
```

Group E - Notepad

Group E

Data Technology Solutions

Group E

Ibuk Shah Shankhar
Avinderjit Singh
Irandeep Kaur
Amaldeep Kaur Dhami
Nirudh Siddula



9:33 PM
09-Oct-21

```

root@sandbox-hdp:~|
| Database:          | test_db
| OwnerType:        | USER
| Owner:            | hive
| CreateTime:       | Sat Oct 09 15:46:24 UTC 2021
| LastAccessTime:   | UNKNOWN
| Retention:        | 0
| Location:         | hdfs://sandbox-hdp.hortonworks.com:8020/warehouse/tablespace/managed/hive/test_db.db/salesjan2009_parquet
| Table Type:       | MANAGED_TABLE
| Table Parameters: | NULL
|                   | COLUMN_STATS_ACCURATE
| true", "COLUMN_STATS": { "city": "true", "country": "true", "customername": "true", "paymenttype": "true", "productid": "true", "region": "true", "salesamount": "true" } }
|                   | bucketing_version
|                   | 2
| numFiles          | 0
| numRows           | 0
| parquet.compress | snappy
| rawDataSize       | 0
| totalSize         | 0
| transactional    | true
| transactional_properties | insert_only
| transient_lastDdlTime | 1633794384
| NULL              | NULL
| # Storage Information | NULL

```

Group E - Notepad
File Edit Format View Help
Data Technology Solutions
Group E
Mibek Shah Shankhar
Savinderjit Singh
Irandeep Kaur
Amaldeep kaur Dhami
Nirudh Siddula



9:33 PM
09-Oct-21

```
|          numFiles                                | 0
|          numRows                               | 0
|          parquet.compress                      | snappy
|          rawDataSize                           | 0
|          totalSize                             | 0
|          transactional                         | true
|          transactional_properties              | insert_only
|          transient_lastDdlTime                | 1633794384
|          NULL                                  | NULL
| # Storage Information                         | NULL
| SerDe Library:                            | org.apache.hadoop.hive.ql.io.parquet.serde.ParquetHiveSerDe | NULL
| InputFormat:                             | org.apache.hadoop.hive.ql.io.parquet.MapredParquetInputFormat | NULL
| OutputFormat:                            | org.apache.hadoop.hive.ql.io.parquet.MapredParquetOutputFormat | NULL
| Compressed:                             | No                                         | NULL
| Num Buckets:                            | -1                                         | NULL
| Bucket Columns:                          | []                                         | NULL
| Sort Columns:                            | []                                         | NULL
| Storage Desc Params:                   | NULL                                       | NULL
|                                         serialization.format | 1
+-----+
40 rows selected (0.421 seconds)
0: jdbc:hive2://sandbox-hdp.hortonworks.com:2>
```

Group E - Notepad

File Edit Format View Help

Data Technology Solutions

Group E

ibek Shah Shankhar
avinderjit Singh
irandeep Kaur
amaldeep kaur Dhami
nirudh Siddula

VERTICES	MODE	STATUS	TOTAL	COMPLETED	RUNNING	PENDING	FAILED	KILLED
Map 1	container	SUCCEEDED	1	1	0	0	0	0
Reducer 2	container	SUCCEEDED	1	1	0	0	0	0

VERTICES: 02/02 [=====>>] 100% ELAPSED TIME: 4.27 s

Group E - Notepad

File Edit Format View Help

Data Technology Solutions

Group E

Bibek Shah Shankhar

Davinderjit Singh

Kirandeep Kaur

Kamaldeep kaur Dhami

Anirudh Siddula



```
root@sandbox-hdp:~  
INFO : Status: DAG finished successfully in 3.57 seconds  
INFO :  
INFO : Query Execution Summary  
INFO :  
INFO : -----  
INFO : OPERATION DURATION  
INFO : -----  
INFO : Compile Query 0.20s  
INFO : Prepare Plan 1.51s  
INFO : Get Query Coordinator (AM) 0.00s  
INFO : Submit Plan 0.03s  
INFO : Start DAG 0.57s  
INFO : Run DAG 3.56s  
INFO :  
INFO :  
INFO : Task Execution Summary  
INFO :  
INFO : VERTICES DURATION(ms) CPU_TIME(ms) GC_TIME(ms) INPUT_RECORDS OUTPUT_RECORDS  
INFO : -----  
INFO : Map 1 2011.00 3,320 90 998 1  
INFO : Reducer 2 0.00 440 6 1 0  
INFO :  
INFO :  
INFO : org.apache.tez.common.counters.DAGCounter:  
INFO : NUM_SUCCEEDED_TASKS: 2  
INFO : TOTAL_LAUNCHED_TASKS: 2  
INFO : AM_CPU_MILLISECONDS: 310  
INFO : AM_GC_TIME_MILLIS: 0  
INFO : File System Counters:  
INFO : FILE_BYTES_READ: 1803  
INFO : FILE_BYTES_WRITTEN: 1803  
INFO : HDFS_BYTES_READ: 65835  
INFO : HDFS_BYTES_WRITTEN: 38651  
INFO : HDFS_READ_OPS: 7  
INFO : HDFS_WRITE_OPS: 6  
INFO : HDFS_OP_CREATE: 4  
INFO : HDFS_OP_GET_FILE_STATUS: 6  
INFO : HDFS_OP_MKDIRS: 1  
INFO : HDFS_OP_OPEN: 1  
INFO : HDFS_OP_RENAME: 1  
INFO : org.apache.tez.common.counters.TaskCounter:  
INFO : SPILLED_RECORDS: 0  
INFO : NUM_SHUFFLED_INPUTS: 1
```

```
Group E - Notepad  
File Edit Format View Help  
Data Technology Solutions  
Group E  
  
Bibek Shah Shankhar  
Davinderjit Singh  
Kirandeep Kaur  
Kamaldeep kaur Dhami  
Anirudh Siddula  
  
Ln 4, Col 2 160% Windows (CRLF) UTF-8
```



9:36 PM
09-Oct-21

```
root@sandbox-hdp:~  
INFO : TaskCounter_Map_1_INPUT_salesjan2009_ext:  
INFO : INPUT_RECORDS_PROCESSED: 998  
INFO : INPUT_SPLIT_LENGTH_BYTES: 65835  
INFO : TaskCounter_Map_1_OUTPUT Reducer_2:  
INFO : ADDITIONAL_SPILLS_BYTES_READ: 0  
INFO : ADDITIONAL_SPILLS_BYTES_WRITTEN: 0  
INFO : ADDITIONAL_SPILL_COUNT: 0  
INFO : OUTPUT_BYTES: 3269  
INFO : OUTPUT_BYTES_PHYSICAL: 1795  
INFO : OUTPUT_BYTES_WITH_OVERHEAD: 3279  
INFO : OUTPUT_LARGE_RECORDS: 0  
INFO : OUTPUT_RECORDS: 1  
INFO : SPILLED_RECORDS: 0  
INFO : TaskCounter_Reducer_2_INPUT_Map_1:  
INFO : FIRST_EVENT RECEIVED: 24  
INFO : INPUT_RECORDS_PROCESSED: 1  
INFO : LAST_EVENT RECEIVED: 24  
INFO : NUM FAILED_SHUFFLE_INPUTS: 0  
INFO : NUM_SHUFFLED_INPUTS: 1  
INFO : SHUFFLE_BYTES: 1771  
INFO : SHUFFLE_BYTES_DECOMPRESSED: 3279  
INFO : SHUFFLE_BYTES_DISK_DIRECT: 1771  
INFO : SHUFFLE_BYTES_TO_DISK: 0  
INFO : SHUFFLE_BYTES_TO_MEM: 0  
INFO : SHUFFLE_PHASE_TIME: 32  
INFO : TaskCounter_Reducer_2_OUTPUT_out_Reducer_2:  
INFO : OUTPUT_RECORDS: 0  
INFO : org.apache.hadoop.hive.ql.exec.tez.HiveInputCounters:  
INFO : GROUPED_INPUT_SPLITS_Map_1: 1  
INFO : INPUT_DIRECTORIES_Map_1: 1  
INFO : INPUT_FILES_Map_1: 1  
INFO : RAW_INPUT_SPLITS_Map_1: 1  
INFO : Starting task [Stage-2:DEPENDENCY_COLLECTION] in serial mode  
INFO : Starting task [Stage-0:MOVE] in serial mode  
INFO : Loading data to table test_db.salesjan2009_parquet from hdfs://sandbox-hdp.hortonworks.com:8020/wa  
rehouse/tables/managed/hive/test_db.db/salesjan2009_parquet  
INFO : Starting task [Stage-3:STATS] in serial mode  
INFO : Completed executing command(queryId=hive_20211009155007_e1a6feea-0266-4a36-b56a-e67f861c519a); Tim  
e taken: 6.288 seconds  
INFO : OK  
No rows affected (6.59 seconds)  
0: jdbc:hive2://sandbox-hdp.hortonworks.com:2>
```

```
Group E - Notepad  
File Edit Format View Help  
Data Technology Solutions  
Group E  
  
Bibek Shah Shankhar  
Davinderjit Singh  
Kirandeep Kaur  
Kamaldeep kaur Dhami  
Anirudh Siddula  
  
Ln 4, Col 2 160% Windows (CRLF) UTF-8
```



9:36 PM
09-Oct-21

```
root@sandbox-hdp:~ - X
login as: root
root@192.168.239.133's password:
Last login: Sat Oct  9 15:20:07 2021 from 172.18.0.3
[root@sandbox-hdp ~]# hdfs dfs -ls /warehouse/tablespace/managed/hive/test_db.db/salesjan2009_parquet
Found 1 items
drwxrwxrwx+ - hive hadoop          0 2021-10-09 15:50 /warehouse/tablespace/managed/hive/test_db.db/salesjan2009_parquet/delta_0000001_0000001_0000
[root@sandbox-hdp ~]# hdfs dfs -ls /warehouse/tablespace/managed/hive/test_db.db/salesjan2009_parquet/delta_0000001_0000001_0000
Found 1 items
-rw-rw-rw+ 1 hive hadoop      33933 2021-10-09 15:50 /warehouse/tablespace/managed/hive/test_db.db/salesjan2009_parquet/delta_0000001_0000001_0000/000000_0
[root@sandbox-hdp ~]# hdfs dfs -ls /warehouse/tablespace/managed/hive/test_db.db/salesjan2009_parquet/delta_0000001_0000001_0000/000000_0
-rw-rw-rw+ 1 hive hadoop      33933 2021-10-09 15:50 /warehouse/tablespace/managed/hive/test_db.db/salesjan2009_parquet/delta_0000001_0000001_0000/000000_0
[root@sandbox-hdp ~]# hdfs dfs -ls /warehouse/tablespace/managed/hive/test_db.db/salesjan2009_parquet/delta_0000001_0000001_0000/000000_0
-rw-rw-rw+ 1 hive hadoop      33933 2021-10-09 15:50 /warehouse/tablespace/managed/hive/test_db.db/salesjan2009_parquet/delta_0000001_0000001_0000/000000_0
[root@sandbox-hdp ~]# hdfs dfs -ls /warehouse/tablespace/managed/hive/test_db.db/salesjan2009_parquet/delta_0000001_0000001_0000/000000_0
-rw-rw-rw+ 1 hive hadoop      33933 2021-10-09 15:50 /warehouse/tablespace/managed/hive/test_db.db/salesjan2009_parquet/delta_0000001_0000001_0000/000000_0
[root@sandbox-hdp ~]# hdfs dfs -ls /warehouse/tablespace/managed/hive/test_db.db/salesjan2009_parquet/delta_0000001_0000001_0000/000000_0
PAR1bProductProductProduct3    Product3 [REDACTED] ,6( Product3Product1[REDACTED]          AA@ p!TA
`0@@@QAPD(8    P@CPP@Dd@$DHHL[L [REDACTED],2[REDACTED],2[REDACTED] :&00&P8
$(          00$PPL
MastercardVisaDinersAmex[REDACTED],6(VisaAmexD筠 Id[REDACTED]S
UPMP[REDACTED]U!IpEIVEPeTt]A[TBKUT[REDACTED]Bt[REDACTED]`A<]I[eA([REDACTED]@UAAQ0]&UR]uM[REDACTED]CWWUP[REDACTED]U!EX[P[REDACTED]EUDTUT[REDACTED]uEuP[REDACTED]q
EUD[REDACTED]FQ[$Uy]uFUEQG[REDACTED]GWUWTUE[REDACTED]\Etm PUFWO@dT[REDACTED]
fG[REDACTED]DIDdREE[REDACTED]!|L[REDACTED]
carolinaBetinaFederica e AndreaGouyaGerd WLAURENCEfleuradamRenee ElisabethAidanStacyHeid
Bryan KerrenechrisMaxineFamilytbKatherineLindaSYLVIASheilasNicolStephanieKellyJamejenniferAnneliRitzSylvia
Marie
Mehmet FatihAnabelaNicole
Christiane SarisimoneVanessaAnupamKarinaFrankAngelaDarrenNikkiAnushkaJune Baybars
```

Group E - Notepad

File Edit Format View Help

Data Technology Solutions

Group E

Bibek Shah Shankhar

Davinderjit Singh

Kirandeep Kaur

Kamaldeep kaur Dhami

Anirudh Siddula

