

Домашнее задание

Скачать любой датасет из списка ниже с сайта [Kaggle.com](https://www.kaggle.com) (достаточно большой)

BankChurners.csv

Загрузить этот датасет в HDFS в свою домашнюю папку

```
[student1_13@bigdataanalytics-head-0 ~]$ hdfs dfs -mkdir hw4
[student1_13@bigdataanalytics-head-0 ~]$ hdfs dfs -ls
Found 4 items
drwx----- - student1_13 student1_13      0 2021-02-06 23:02 .staging
drwxr-xr-x - student1_13 student1_13      0 2021-02-09 21:53 hw4
drwxr-xr-x - student1_13 student1_13      0 2021-02-06 23:02 result
drwxr-xr-x - student1_13 student1_13      0 2021-02-06 22:39 test
ekaterina@MacBook-Pro-Ekaterina .ssh % ls
BankChurners.csv      known_hosts
id_rsa                newfile
id_rsa_mac            student1_13@185.241.193.174
id_rsa_mac.pub
ekaterina@MacBook-Pro-Ekaterina .ssh % scp -i ../ssh/id_rsa_mac BankChurners.csv
student1_13@185.241.193.174:.
/etc/profile.d/lang.sh: line 19: warning: setlocale: LC_CTYPE: cannot change locale (UTF-8): No such file or
directory
BankChurners.csv      100% 1475KB 538.8KB/s  00:02
ekaterina@MacBook-Pro-Ekaterina .ssh % ssh -i ../ssh/id_rsa_mac student1_13@185.241.193.174
Last login: Tue Feb  9 21:52:09 2021 from 89-64-64-122.dynamic.chello.pl
-bash: warning: setlocale: LC_CTYPE: cannot change locale (UTF-8): No such file or directory
[student1_13@bigdataanalytics-head-0 ~]$ -ls
-bash: -ls: command not found
[student1_13@bigdataanalytics-head-0 ~]$ ls
BankChurners.csv  folder      hs_err_pid23173.log  mapper.py
f1.txt           hs_err_pid1104.log  hs_err_pid32716.log  reducer.py
f2              hs_err_pid12017.log  hs_err_pid8728.log
[student1_13@bigdataanalytics-head-0 ~]$ hdfs dfs -put BankChurners.csv hw4/
[student1_13@bigdataanalytics-head-0 ~]$ hdfs dfs -ls
Found 4 items
drwx----- - student1_13 student1_13      0 2021-02-06 23:02 .staging
drwxr-xr-x - student1_13 student1_13      0 2021-02-09 22:14 hw4
drwxr-xr-x - student1_13 student1_13      0 2021-02-06 23:02 result
drwxr-xr-x - student1_13 student1_13      0 2021-02-06 22:39 test
[student1_13@bigdataanalytics-head-0 ~]$ hdfs dfs -ls hw4
Found 1 items
-rw-r--r--  2 student1_13 student1_13   1510880 2021-02-09 22:14 hw4/BankChurners.csv
```

Создать собственную базу данных в HIVE

```
[student1_13@bigdataanalytics-head-0 ~]$ hive
SLF4J: Class path contains multiple SLF4J bindings.
SLF4J: Found binding in [jar:file:/usr/hdp/3.1.4.0-315/hive/lib/log4j-slf4j-impl-2.10.0.jar!/org/slf4j/impl/
StaticLoggerBinder.class]
SLF4J: Found binding in [jar:file:/usr/hdp/3.1.4.0-315/hadoop/lib/slf4j-log4j12-1.7.25.jar!/org/slf4j/impl/
StaticLoggerBinder.class]
SLF4J: See http://www.slf4j.org/codes.html#multiple_bindings for an explanation.
SLF4J: Actual binding is of type [org.apache.logging.slf4j.Log4jLoggerFactory]
Connecting to jdbc:hive2://bigdataanalytics-worker-0.novalocal:2181,bigdataanalytics-
worker-1.novalocal:2181,bigdataanalytics-worker-2.novalocal:2181,bigdataanalytics-worker-3.novalocal:2181/
default;password=student1_13;serviceDiscoveryMode=zooKeeper;user=student1_13;zooKeeperNamespace=hive
server2
21/02/09 22:16:07 [main]: INFO jdbc.HiveConnection: Connected to bigdataanalytics-head-0.novalocal:10000
Connected to: Apache Hive (version 3.1.0.3.1.4.0-315)
Driver: Hive JDBC (version 3.1.0.3.1.4.0-315)
Transaction isolation: TRANSACTION_REPEATABLE_READ
Beeline version 3.1.0.3.1.4.0-315 by Apache Hive
0: jdbc:hive2://bigdataanalytics-worker-0.nov> create database bank;
```

INFO : Compiling command(queryId=hive_20210209221646_3049de34-c210-480a-85ae-1ff18228f360): create database bank
 INFO : Semantic Analysis Completed (retrial = false)
 INFO : Returning Hive schema: Schema(fieldSchemas:null, properties:null)
 INFO : Completed compiling command(queryId=hive_20210209221646_3049de34-c210-480a-85ae-1ff18228f360); Time taken: 0.021 seconds
 INFO : Executing command(queryId=hive_20210209221646_3049de34-c210-480a-85ae-1ff18228f360): create database bank
 INFO : Starting task [Stage-0:DDL] in serial mode
 INFO : Completed executing command(queryId=hive_20210209221646_3049de34-c210-480a-85ae-1ff18228f360); Time taken: 0.051 seconds
 INFO : OK
 No rows affected (0.15 seconds)

Создать таблицы внутри базы данных с использованием всех загруженных файлов. Один файл – одна таблица.

«CLIENTNUM», "Attrition_Flag", "Customer_Age", "Gender", "Dependent_count", "Education_Level", "Marital_Status", "Income_Category", "Card_Category", "Months_on_book", "Total_Relationship_Count", "Months_Inactive_12_mon", "Contacts_Count_12_mon", "Credit_Limit", "Total_Revolving_Bal", "Avg_Open_To_Buy", "Total_Amt_Chng_Q4_Q1",

"Total_Trans_Amt", "Total_Trans_Ct", "Total_Ct_Chng_Q4_Q1", "Avg_Utilization_Ratio", "Naive_Bayes_Classifier_Attrition_Flag_Card_Category_Contacts_Count_12_mon_Dependent_count_Education_Level_Months_Inactive_12_mon_1", "Naive_Bayes_Classifier_Attrition_Flag_Card_Category_Contacts_Count_12_mon_Dependent_count_Education_Level_Months_Inactive_12_mon_2"

```
create external table bank.churners(CLIENTNUM_id string, Attrition_Flag_id string,
Customer_Age string, Gender string, Dependent_count string, Education_Level string,
Marital_Status string, Income_Category string, Card_Category string, Months_on_book
string, Total_Relationship_Count string, Months_Inactive_12_mon string,
Contacts_Count_12_mon string, Credit_Limit string, Total_Revolving_Bal string,
Avg_Open_To_Buy string, Total_Amt_Chng_Q4_Q1 string, Total_Trans_Amt string,
Total_Trans_Ct string, Total_Ct_Chng_Q4_Q1 string, Avg_Utilization_Ratio string,
Naive_Bayes string, Naive_Bayes_2 string) ROW FORMAT serde
"org.apache.hadoop.hive.serde2.lazy.LazySimpleSerDe" WITH SERDEPROPERTIES
(«field.delim" = «,») location "/user/student1_13/hw4" tblproperties
(«skip.header.line.count»=«1»);
```

```
create table bank.churners (CLIENTNUM_id string, Attrition_Flag_id string, Customer_Age string, Gender string,
Dependent_count string, Education_Level string, Marital_Status string, Income_Category string, Card_Category
string, Months_on_book string, Total_Relationship_Count string, Months_Inactive_12_mon string,
Contacts_Count_12_mon string, Credit_Limit string, Total_Revolving_Bal string, Avg_Open_To_Buy string,
```

```
Total_Amt_Chng_Q4_Q1 string, Total_Trans_Amt string, Total_Trans_Ct string, Total_Ct_Chng_Q4_Q1 string,
Avg_Utilization_Ratio string,
Naive_Bayes_Classifier_Attrition_Flag_Card_Category_Contacts_Count_12_mon_Dependent_count_Education_L
evel_Months_Inactive_12_mon_1 string,
Naive_Bayes_Classifier_Attrition_Flag_Card_Category_Contacts_Count_12_mon_Dependent_count_Education_L
evel_Months_Inactive_12_mon_2 string) ROW FORMAT serde
'org.apache.hadoop.hive.serde2.lazy.LazySimpleSerDe' WITH SERDEPROPERTIES ('field.delim' = ',';) location «/
user/student1_13/hw4» tblproperties («skip.header.line.count"="1");
```

```
0: jdbc:hive2://bigdataanalytics-worker-0.nov> show tables;
INFO  : Compiling command(queryId=hive_20210209224132_6c7de5bb-d637-4e77-88a5-
d67a42397eaf): show tables
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_20210209224132_6c7de5bb-
d637-4e77-88a5-d67a42397eaf); Time taken: 0.007 seconds
INFO  : Executing command(queryId=hive_20210209224132_6c7de5bb-d637-4e77-88a5-
d67a42397eaf): show tables
INFO  : Starting task [Stage-0:DDL] in serial mode
INFO  : Completed executing command(queryId=hive_20210209224132_6c7de5bb-
d637-4e77-88a5-d67a42397eaf); Time taken: 0.005 seconds
INFO  : OK
```

```
+-----+
| tab_name |
+-----+
+-----+
```

No rows selected (0.023 seconds)

```
create external table bank.churners(CLIENTNUM_id string, Attrition_Flag_id string,
Customer_Age string, Gender string, Dependent_count string, Education_Level string,
Marital_Status string, Income_Category string, Card_Category string, Months_on_book
string, Total_Relationship_Count string, Months_Inactive_12_mon string,
Contacts_Count_12_mon string, Credit_Limit string, Total_Revolving_Bal string,
Avg_Open_To_Buy string, Total_Amt_Chng_Q4_Q1 string, Total_Trans_Amt string,
Total_Trans_Ct string, Total_Ct_Chng_Q4_Q1 string, Avg_Utilization_Ratio string,
Naive_Bayes string, Naive_Bayes_2 string) ROW FORMAT serde
"org.apache.hadoop.hive.serde2.lazy.LazySimpleSerDe" WITH SERDEPROPERTIES
("field.delim" = ",") location "/user/student1_13/hw4" tblproperties
("skip.header.line.count"="1");
```

```
INFO  : Compiling command(queryId=hive_20210209231042_25284316-677c-460c-a70e-
d3a2d4f5f23e): create external table bank.churners(CLIENTNUM_id string,
Attrition_Flag_id string, Customer_Age string, Gender string, Dependent_count
string, Education_Level string, Marital_Status string, Income_Category string,
Card_Category string, Months_on_book string, Total_Relationship_Count string,
Months_Inactive_12_mon string, Contacts_Count_12_mon string, Credit_Limit string,
Total_Revolving_Bal string, Avg_Open_To_Buy string, Total_Amt_Chng_Q4_Q1 string,
Total_Trans_Amt string, Total_Trans_Ct string, Total_Ct_Chng_Q4_Q1 string,
Avg_Utilization_Ratio string, Naive_Bayes string, Naive_Bayes_2 string) ROW FORMAT
serde "org.apache.hadoop.hive.serde2.lazy.LazySimpleSerDe" WITH SERDEPROPERTIES
("field.delim" = ",") location "/user/student1_13/hw4" tblproperties
("skip.header.line.count"="1")
```

```
INFO  : Semantic Analysis Completed (retrial = false)
INFO  : Returning Hive schema: Schema(fieldSchemas:null, properties:null)
INFO  : Completed compiling command(queryId=hive_20210209231042_25284316-677c-460c-
a70e-d3a2d4f5f23e); Time taken: 0.017 seconds
INFO  : Executing command(queryId=hive_20210209231042_25284316-677c-460c-a70e-
d3a2d4f5f23e): create external table bank.churners(CLIENTNUM_id string,
Attrition_Flag_id string, Customer_Age string, Gender string, Dependent_count
string, Education_Level string, Marital_Status string, Income_Category string,
Card_Category string, Months_on_book string, Total_Relationship_Count string,
Months_Inactive_12_mon string, Contacts_Count_12_mon string, Credit_Limit string,
Total_Revolving_Bal string, Avg_Open_To_Buy string, Total_Amt_Chng_Q4_Q1 string,
```

Total_Trans_Amt string, Total_Trans_Ct string, Total_Ct_Chng_Q4_Q1 string, Avg_Utilization_Ratio string, Naive_Bayes string, Naive_Bayes_2 string) ROW FORMAT serde "org.apache.hadoop.hive.serde2.lazy.LazySimpleSerDe" WITH SERDEPROPERTIES ("field.delim" = ",") location "/user/student1_13/hw4" tblproperties ("skip.header.line.count"="1")

INFO : Starting task [Stage-0:DDL] in serial mode

INFO : Completed executing command(queryId=hive_20210209231042_25284316-677c-460c-a70e-d3a2d4f5f23e); Time taken: 0.057 seconds

INFO : OK

No rows affected (0.197 seconds)

0: jdbc:hive2://bigdataanalytics-worker-0.nov> select * from churners;

INFO : Compiling command(queryId=hive_20210209231128_c7c23324-0624-46e6-a587-8050329e3396): select * from churners

INFO : Semantic Analysis Completed (retrial = false)

INFO : Returning Hive schema: Schema(fieldSchemas: [FieldSchema(name:churners.clientnum_id, type:string, comment:null), FieldSchema(name:churners.attrition_flag_id, type:string, comment:null), FieldSchema(name:churners.customer_age, type:string, comment:null), FieldSchema(name:churners.gender, type:string, comment:null), FieldSchema(name:churners.dependent_count, type:string, comment:null), FieldSchema(name:churners.education_level, type:string, comment:null), FieldSchema(name:churners.marital_status, type:string, comment:null), FieldSchema(name:churners.income_category, type:string, comment:null), FieldSchema(name:churners.card_category, type:string, comment:null), FieldSchema(name:churners.months_on_book, type:string, comment:null), FieldSchema(name:churners.total_relationship_count, type:string, comment:null), FieldSchema(name:churners.months_inactive_12_mon, type:string, comment:null), FieldSchema(name:churners.contacts_count_12_mon, type:string, comment:null), FieldSchema(name:churners.credit_limit, type:string, comment:null), FieldSchema(name:churners.total_revolving_bal, type:string, comment:null), FieldSchema(name:churners.avg_open_to_buy, type:string, comment:null), FieldSchema(name:churners.total_amt_chng_q4_q1, type:string, comment:null), FieldSchema(name:churners.total_trans_amt, type:string, comment:null), FieldSchema(name:churners.total_trans_ct, type:string, comment:null), FieldSchema(name:churners.total_ct_chng_q4_q1, type:string, comment:null), FieldSchema(name:churners.avg_utilization_ratio, type:string, comment:null), FieldSchema(name:churners.naive_bayes, type:string, comment:null), FieldSchema(name:churners.naive_bayes_2, type:string, comment:null)], properties:null)

INFO : Completed compiling command(queryId=hive_20210209231128_c7c23324-0624-46e6-a587-8050329e3396); Time taken: 0.083 seconds

INFO : Executing command(queryId=hive_20210209231128_c7c23324-0624-46e6-a587-8050329e3396): select * from churners

INFO : Completed executing command(queryId=hive_20210209231128_c7c23324-0624-46e6-a587-8050329e3396); Time taken: 0.001 seconds

INFO : OK

churners.clientnum_id	churners.attrition_flag_id	churners.customer_age	
churners.gender	churners.dependent_count	churners.education_level	
churners.marital_status	churners.income_category	churners.card_category	
churners.months_on_book	churners.total_relationship_count		
churners.months_inactive_12_mon	churners.contacts_count_12_mon		
churners.credit_limit	churners.total_revolving_bal	churners.avg_open_to_buy	

churners.total_amt_chng_q4_q1	churners.total_trans_amt	churners.total_trans_ct	churners.total_ct_chng_q4_q1	churners.avg_utilization_ratio	churners.naive_bayes	churners.naive_bayes_2	
768805383	"Existing Customer"	45	"M"	3	"Blue"	"High School"	"Married"
"\$60K - \$80K"		39					5
1	3						12691
777	11914	1.335					
1144	42	1.625					
0.061	9.3448e-05	0.99991					
818770008	"Existing Customer"	49	"F"	5	"Blue"	"Graduate"	"Single"
"Less than \$40K"		44					6
1	2						8256
864	7392	1.541					
1291	33	3.714					
0.105	5.6861e-05	0.99994					
713982108	"Existing Customer"	51	"M"	3	"Blue"	"Graduate"	"Married"
"\$80K - \$120K"		36					4
1	0						3418
0	3418	2.594					
1887	20	2.333					
0	2.1081e-05	0.99998					
769911858	"Existing Customer"	40	"F"	4	"Blue"	"High School"	"Unknown"
"Less than \$40K"		34					3
4	1						3313
2517	796	1.405					
1171	20	2.333					
0.76	0.00013366	0.99987					
709106358	"Existing Customer"	40	"M"	3	"Blue"	"Uneducated"	"Married"
"\$60K - \$80K"		21					5
1	0						4716
0	4716	2.175					
816	28	2.5					
0	2.1676e-05	0.99998					
713061558	"Existing Customer"	44	"M"	2	"Blue"	"Graduate"	"Married"
"\$40K - \$60K"		36					3
1	2						4010
1247	2763	1.376					
1088	24	0.846					
0.311	5.5077e-05	0.99994					

Сделать любой отчет по загруженным данным используя групповые и агрегатные функции.

```
0: jdbc:hive2://bigdataanalytics-worker-0.nov> SELECT clientnum_id, customer_age
FROM churners WHERE customer_age >= 70;
```

INFO : OK

clientnum_id	customer_age
715952883	73
787348608	70

2 rows selected (0.119 seconds)

```
0: jdbc:hive2://bigdataanalytics-worker-0.nov> SELECT education_level,
marital_status, income_category, customer_age, clientnum_id, credit_limit FROM
churners limit 5;
```

INFO : OK

education_level	marital_status	income_category	customer_age	clientnum_id	credit_limit
"High School"	"Married"	"\$60K - \$80K"	45	768805383	12691
"Graduate"	"Single"	"Less than \$40K"	49	818770008	8256
"Graduate"	"Married"	"\$80K - \$120K"	51	713982108	3418
"High School"	"Unknown"	"Less than \$40K"	40	769911858	3313
"Uneducated"	"Married"	"\$60K - \$80K"	40	709106358	4716

5 rows selected (0.12 seconds)

```
0: jdbc:hive2://bigdataanalytics-worker-0.nov> select gender from churners limit 10;
```

INFO : OK

gender
"M"
"F"
"M"
"F"
"M"
"M"
"M"
"M"
"M"
"M"

10 rows selected (0.126 seconds)

```
0: jdbc:hive2://bigdataanalytics-worker-0.nov> SELECT customer_age FROM churners
GROUP BY customer_age;
```

```
0: jdbc:hive2://bigdataanalytics-worker-0.nov> SELECT customer_age, clientnum_id,
credit_limit FROM churners GROUP BY customer_age HAVING (customer_age >= 50);
Error: Error while compiling statement: FAILED: SemanticException [Error 10025]:
Line 1:21 Expression not in GROUP BY key 'clientnum_id' (state=42000,code=10025)
```

Сделать любой отчет по загруженным данным используя JOIN.

```
0: jdbc:hive2://bigdataanalytics-worker-0.nov> SELECT * FROM churners c LEFT JOIN
churners c2 on education_level;
INFO  : Compiling command(queryId=hive_20210210133804_2f55673b-d787-44d6-94a1-
fb94e4fbb918): SELECT * FROM churners c LEFT JOIN churners c2 on education_level
INFO  : Warning: Map Join MAPJOIN[9][bigTable=c] in task 'Map 1' is a cross product
INFO  : Semantic Analysis Completed (retrial = false)
INFO  : Returning Hive schema: Schema(fieldSchemas:[FieldSchema(name:c.clientnum_id,
type:string, comment:null), FieldSchema(name:c.attrition_flag_id, type:string,
comment:null), FieldSchema(name:c.customer_age, type:string, comment:null),
FieldSchema(name:c.gender, type:string, comment:null),
FieldSchema(name:c.dependent_count, type:string, comment:null),
FieldSchema(name:c.education_level, type:string, comment:null),
FieldSchema(name:c.marital_status, type:string, comment:null),
FieldSchema(name:c.income_category, type:string, comment:null),
FieldSchema(name:c.card_category, type:string, comment:null),
FieldSchema(name:c.months_on_book, type:string, comment:null),
FieldSchema(name:c.total_relationship_count, type:string, comment:null),
FieldSchema(name:c.months_inactive_12_mon, type:string, comment:null),
FieldSchema(name:c.contacts_count_12_mon, type:string, comment:null),
FieldSchema(name:c.credit_limit, type:string, comment:null),
FieldSchema(name:c.total_revolving_bal, type:string, comment:null),
FieldSchema(name:c.avg_open_to_buy, type:string, comment:null),
FieldSchema(name:c.total_amt_chng_q4_q1, type:string, comment:null),
FieldSchema(name:c.total_trans_amt, type:string, comment:null),
FieldSchema(name:c.total_trans_ct, type:string, comment:null),
FieldSchema(name:c.total_ct_chng_q4_q1, type:string, comment:null),
FieldSchema(name:c.avg_utilization_ratio, type:string, comment:null),
FieldSchema(name:c.naive_bayes, type:string, comment:null),
FieldSchema(name:c.naive_bayes_2, type:string, comment:null),
FieldSchema(name:c2.clientnum_id, type:string, comment:null),
FieldSchema(name:c2.attrition_flag_id, type:string, comment:null),
FieldSchema(name:c2.customer_age, type:string, comment:null),
FieldSchema(name:c2.gender, type:string, comment:null),
FieldSchema(name:c2.dependent_count, type:string, comment:null),
FieldSchema(name:c2.education_level, type:string, comment:null),
FieldSchema(name:c2.marital_status, type:string, comment:null),
FieldSchema(name:c2.income_category, type:string, comment:null),
FieldSchema(name:c2.card_category, type:string, comment:null),
FieldSchema(name:c2.months_on_book, type:string, comment:null),
FieldSchema(name:c2.total_relationship_count, type:string, comment:null),
FieldSchema(name:c2.months_inactive_12_mon, type:string, comment:null),
FieldSchema(name:c2.contacts_count_12_mon, type:string, comment:null),
FieldSchema(name:c2.credit_limit, type:string, comment:null),
FieldSchema(name:c2.total_revolving_bal, type:string, comment:null),
FieldSchema(name:c2.avg_open_to_buy, type:string, comment:null),
FieldSchema(name:c2.total_amt_chng_q4_q1, type:string, comment:null),
FieldSchema(name:c2.total_trans_amt, type:string, comment:null),
FieldSchema(name:c2.total_trans_ct, type:string, comment:null),
FieldSchema(name:c2.total_ct_chng_q4_q1, type:string, comment:null),
FieldSchema(name:c2.avg_utilization_ratio, type:string, comment:null),
FieldSchema(name:c2.naive_bayes, type:string, comment:null),
FieldSchema(name:c2.naive_bayes_2, type:string, comment:null)], properties:null)
INFO  : Completed compiling command(queryId=hive_20210210133804_2f55673b-
d787-44d6-94a1-fb94e4fbb918); Time taken: 1.783 seconds
INFO  : Executing command(queryId=hive_20210210133804_2f55673b-d787-44d6-94a1-
fb94e4fbb918): SELECT * FROM churners c LEFT JOIN churners c2 on education_level
INFO  : Query ID = hive_20210210133804_2f55673b-d787-44d6-94a1-fb94e4fbb918
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_20210210133804_2f55673b-
d787-44d6-94a1-fb94e4fbb918
INFO : Tez session hasn't been created yet. Opening session
ERROR : Failed to execute tez graph.
java.io.IOException: Previous writer likely failed to write hdfs://bigdataanalytics-
head-0.novalocal:8020/user/student1_13/.hiveJars/hive-exec-3.1.0.3.1.4.0-315-
effe339ef81326ce093bc0a1516e86d9d189e11126de97212254c005859b949e.jar. Failing
because I am unlikely to write too.
    at
org.apache.hadoop.hive.ql.exec.tez.DagUtils.localizeResource(DagUtils.java:1296)
~[hive-exec-3.1.0.3.1.4.0-315.jar:3.1.0.3.1.4.0-315]
    at
org.apache.hadoop.hive.ql.exec.tez.TezSessionState.createJarLocalResource(TezSession
State.java:917) ~[hive-exec-3.1.0.3.1.4.0-315.jar:3.1.0.3.1.4.0-315]
    at
org.apache.hadoop.hive.ql.exec.tez.TezSessionState.makeCombinedJarMap(TezSessionStat
e.java:349) ~[hive-exec-3.1.0.3.1.4.0-315.jar:3.1.0.3.1.4.0-315]
    at
org.apache.hadoop.hive.ql.exec.tez.TezSessionState.openInternal(TezSessionState.java
:418) ~[hive-exec-3.1.0.3.1.4.0-315.jar:3.1.0.3.1.4.0-315]
    at
org.apache.hadoop.hive.ql.exec.tez.TezSessionPoolSession.openInternal(TezSessionPool
Session.java:124) ~[hive-exec-3.1.0.3.1.4.0-315.jar:3.1.0.3.1.4.0-315]
    at
org.apache.hadoop.hive.ql.exec.tez.TezSessionState.open(TezSessionState.java:373)
~[hive-exec-3.1.0.3.1.4.0-315.jar:3.1.0.3.1.4.0-315]
    at
org.apache.hadoop.hive.ql.exec.tez.TezTask.ensureSessionHasResources(TezTask.java:37
3) ~[hive-exec-3.1.0.3.1.4.0-315.jar:3.1.0.3.1.4.0-315]
    at org.apache.hadoop.hive.ql.exec.tez.TezTask.execute(TezTask.java:200) ~[hive-
exec-3.1.0.3.1.4.0-315.jar:3.1.0.3.1.4.0-315]
    at org.apache.hadoop.hive.ql.exec.Task.executeTask(Task.java:212) ~[hive-
exec-3.1.0.3.1.4.0-315.jar:3.1.0.3.1.4.0-315]
    at org.apache.hadoop.hive.ql.exec.TaskRunner.runSequential(TaskRunner.java:103)
~[hive-exec-3.1.0.3.1.4.0-315.jar:3.1.0.3.1.4.0-315]
    at org.apache.hadoop.hive.ql.Driver.launchTask(Driver.java:2712) ~[hive-
exec-3.1.0.3.1.4.0-315.jar:3.1.0.3.1.4.0-315]
    at org.apache.hadoop.hive.ql.Driver.execute(Driver.java:2383) ~[hive-
exec-3.1.0.3.1.4.0-315.jar:3.1.0.3.1.4.0-315]
    at org.apache.hadoop.hive.ql.Driver.runInternal(Driver.java:2055) ~[hive-
exec-3.1.0.3.1.4.0-315.jar:3.1.0.3.1.4.0-315]
    at org.apache.hadoop.hive.ql.Driver.run(Driver.java:1753) ~[hive-
exec-3.1.0.3.1.4.0-315.jar:3.1.0.3.1.4.0-315]
    at org.apache.hadoop.hive.ql.Driver.run(Driver.java:1747) ~[hive-
exec-3.1.0.3.1.4.0-315.jar:3.1.0.3.1.4.0-315]
    at org.apache.hadoop.hive.ql.reexec.ReExecDriver.run(ReExecDriver.java:157)
~[hive-exec-3.1.0.3.1.4.0-315.jar:3.1.0.3.1.4.0-315]
    at
org.apache.hive.service.cli.operation.SQLOperation.runQuery(SQLOperation.java:226)
~[hive-service-3.1.0.3.1.4.0-315.jar:3.1.0.3.1.4.0-315]
    at
org.apache.hive.service.cli.operation.SQLOperation.access$700(SQLOperation.java:87)
~[hive-service-3.1.0.3.1.4.0-315.jar:3.1.0.3.1.4.0-315]
    at
org.apache.hive.service.cli.operation.SQLOperation$BackgroundWork$1.run(SQLOperation
.java:324) ~[hive-service-3.1.0.3.1.4.0-315.jar:3.1.0.3.1.4.0-315]
    at java.security.AccessController.doPrivileged(Native Method) ~[?:1.8.0_191]
    at javax.security.auth.Subject.doAs(Subject.java:422) ~[?:1.8.0_191]
    at
org.apache.hadoop.security.UserGroupInformation.doAs(UserGroupInformation.java:1730)
~[hadoop-common-3.1.1.3.1.4.0-315.jar:~]
```



```

    at
org.apache.hive.service.cli.operation.SQLOperation$BackgroundWork.run(SQLOperation.j
ava:342) ~[hive-service-3.1.0.3.1.4.0-315.jar:3.1.0.3.1.4.0-315]
    at java.util.concurrent.Executors$RunnableAdapter.call(Executors.java:511)
~[?:1.8.0_191]
    at java.util.concurrent.FutureTask.run(FutureTask.java:266) ~[?:1.8.0_191]
    at java.util.concurrent.Executors$RunnableAdapter.call(Executors.java:511)
~[?:1.8.0_191]
    at java.util.concurrent.FutureTask.run(FutureTask.java:266) ~[?:1.8.0_191]
    at
java.util.concurrent.ThreadPoolExecutor.runWorker(ThreadPoolExecutor.java:1149)
~[?:1.8.0_191]
    at
java.util.concurrent.ThreadPoolExecutor$Worker.run(ThreadPoolExecutor.java:624)
~[?:1.8.0_191]
    at java.lang.Thread.run(Thread.java:748) [?:1.8.0_191]
ERROR : FAILED: Execution Error, return code 1 from
org.apache.hadoop.hive.ql.exec.tez.TezTask
INFO  : Completed executing command(queryId=hive_20210210133804_2f55673b-
d787-44d6-94a1-fb94e4fbb918); Time taken: 25.138 seconds
Error: Error while processing statement: FAILED: Execution Error, return code 1 from
org.apache.hadoop.hive.ql.exec.tez.TezTask (state=08S01,code=1)

0: jdbc:hive2://bigdataanalytics-worker-0.nov> SELECT clientnum_id, customer_age,
credit_limit FROM churners c JOIN churners co ON c.customer_age = 60 GROUP BY
clientnum_id HAVING (credit_limit >= 1000);
Error: Error while compiling statement: FAILED: SemanticException Column
clientnum_id Found in more than One Tables/Subqueries (state=42000,code=40000)

```