Lab Assignment 1 Finding Facebook mutual friends using Map Reduce, Counting and Summing, Hive and Solr Queries.

Team Id: 4

Member 1: Nagababu Chilukuri

Class Id: 4

Member 2: Grace Sylvia

Class Id: 30

Member 3: PavanKumar Manchalla

Class Id: 16

Introduction.

This lab assignment deals with the implementation of MapReduce algorithm for finding Facebook common friends problem and Counting

and summing problem statement given and running the MapReduce job on Apache Hadoop. The second part deals with running the Hive and Solr queries for the given datasets.

Objective.

1. Implement MapReduce algorithm for finding Facebook common friends problem and run the MapReduce job on Apache Hadoop. Show your implementation through map-reduce diagram

Approach.

Take the input from the given use case A -> B C D,B -> A C D E,C -> A B D E,D -> A B C E,E -> B C D. For each user, iterate the friends and generate the keys of them.In the reducing phase, each intermediate key would get exactly two values. Compare the two values, that is their friend list, and leave the repeated.

WorkFlow.

Mapper.

In Mapper phase from the given input, iterate the friends himself/herself and generate the key,value pairs of each friends.

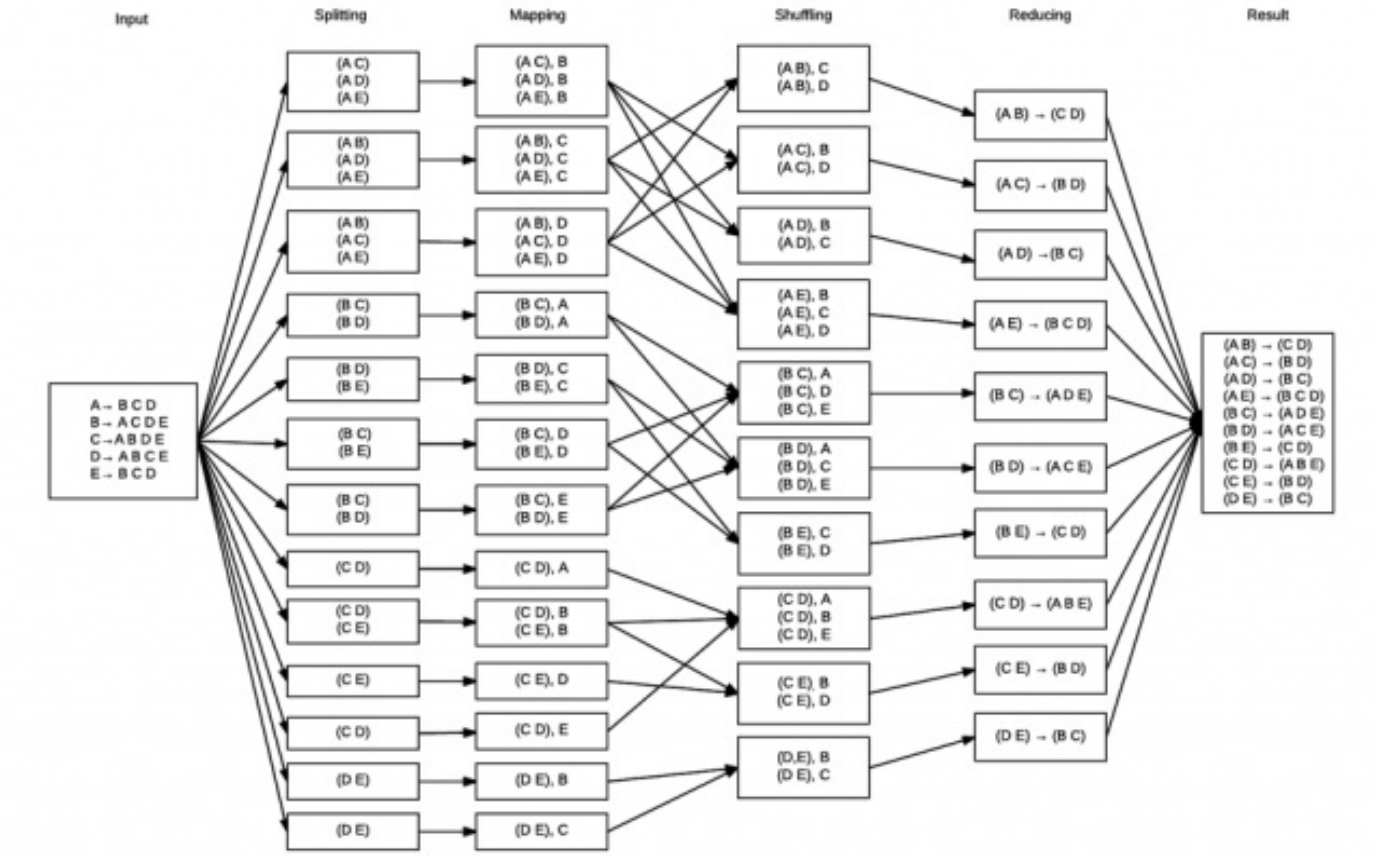
```
public static class Map extends MapReduceBase
        implements Mapper<LongWritable, Text, Text, Text>{
        public void map(LongWritable key, Text value, OutputCollector<Text, Text> output, Reporter reporter)
                throws IOException{
                StringTokenizer tokenizer = new StringTokenizer(value.toString(), "\n");
                String line = null;
                String[] lineArray = null;
                String[] friendArray = null;
                String[] tempArray = null;
                while(tokenizer.hasMoreTokens()){
                        line = tokenizer.nextToken();
                        lineArray = line.split(" : ");
                        friendArray = lineArray[1].split(" ");
                        tempArray = new String[2];
                        for(int i = 0; i < friendArray.length; i++){
                                tempArray[0] = friendArray[i];
                                tempArray[1] = lineArray[0];
                                Arrays.sort(tempArray);
                                output.collect(new Text(tempArray[0] + " " + tempArray[1]), new Text(lineArray[1]));
```

Reducer.

In reducer phase the key,value pairs of friends are grouped based on friends with same keys and then reduced to find the mutual friends.

```
public static class Reduce extends MapReduceBase
        implements Reducer<Text, Text, Text, Text>{
        public void reduce(Text key, Iterator<Text> values,
        OutputCollector<Text, Text> output, Reporter reporter) throws IOException{
                Text[] texts = new Text[2];
                int index = 0;
                while(values.hasNext()){
                        texts[index++] = new Text(values.next());
                }
                String[] list1 = texts[0].toString().split(" ");
                String[] list2 = texts[1].toString().split(" ");
                List<String> list = new LinkedList<String>();
                for(String friend1 : list1){
                        for(String friend2 : list2){
                                if(friend1.equals(friend2)){
                                        list.add(friend1);
                                }
                        }
                }
                StringBuffer sb = new StringBuffer();
                for(int i = 0; i < list.size(); i++){
                        sb.append(list.get(i));
                        if(i != list.size() - 1)
                                sb.append(" ");
                }
                output.collect(key, new Text(sb.toString()));
```

Algorithm



Conclusion

Input.

```
[cloudera@quickstart MutualFriend]$ cat input.txt
A : B C D
B : A C D E
C : A B D E
D : A B C E
E : B C D
[cloudera@quickstart MutualFriend]$
```

Output.

```
                WRONG_MAP=0
                WRONG_REDUCE=0
        File Input Format Counters
                Bytes Read=84
        File Output Format Counters
                Bytes Written=78
A B     C D
A C     B D
A D     B C
B C     A D E
B D     A C E
B E     C D
C D     A B E
C E     B D
D E     B C
[cloudera@quickstart MutualFriend]$ cat input.txt
```

2. Implement MapReduce algorithm for finding Average and count problem and run the MapReduce job on Apache Hadoop. Show your implementation through map-reduce diagram

Workflow

In mapper phase the given input is processed and split into key,value pairs. In reducer phase grouping the occurrence of similar things and counting at the end of reducer phase.

Mapper

In mapper phase the input data is processed with “,” delimiter and gives the first part as keys and second part as values which is send over

through the next phase for summation.

```
@Override
protected void map(LongWritable key, Text value, Context context) throws IOException, InterruptedException {
  String line = value.toString();
  String[] records = line.split(",");
  context.write(new Text(records[0].toString()), new IntWritable(Integer.parseInt(records[1])));

}
```

Combiner

In combiner phase the key value pairs obtained is processed with summation and counting average which is probably reduces the amount of data that need to be processed by reducer.

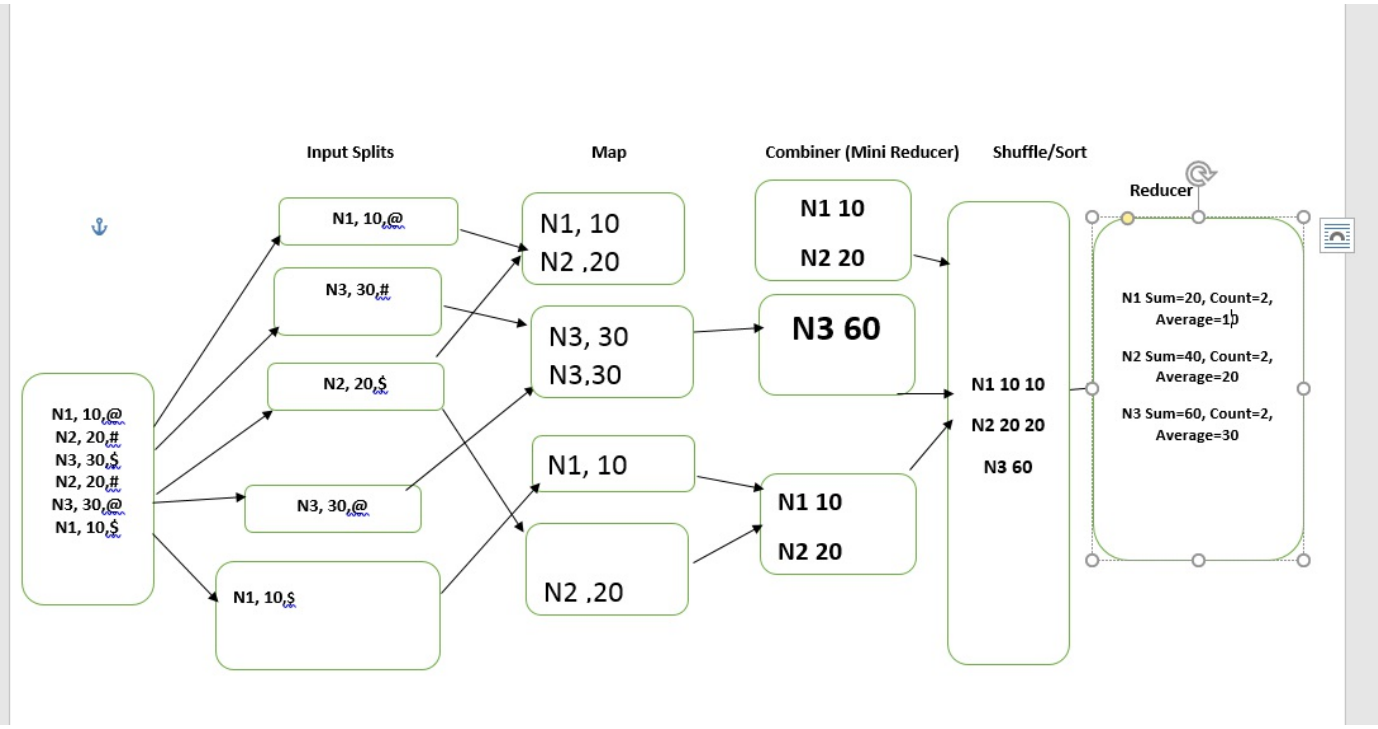
```
protected void reduce(Text key, Iterable<IntWritable> values, Context context) throws IOException, InterruptedException{
  int count = 0;
  int sum = 0;
  int inter_result = 0;
  for (IntWritable value : values){
    sum += value.get();
    count += 1;
  }
  inter_result = sum / count;
  context.write(key, new IntWritable(inter_result));
}
```

Reducer

Finally, in reducer phase the count, average obtained is processed and shows the output.

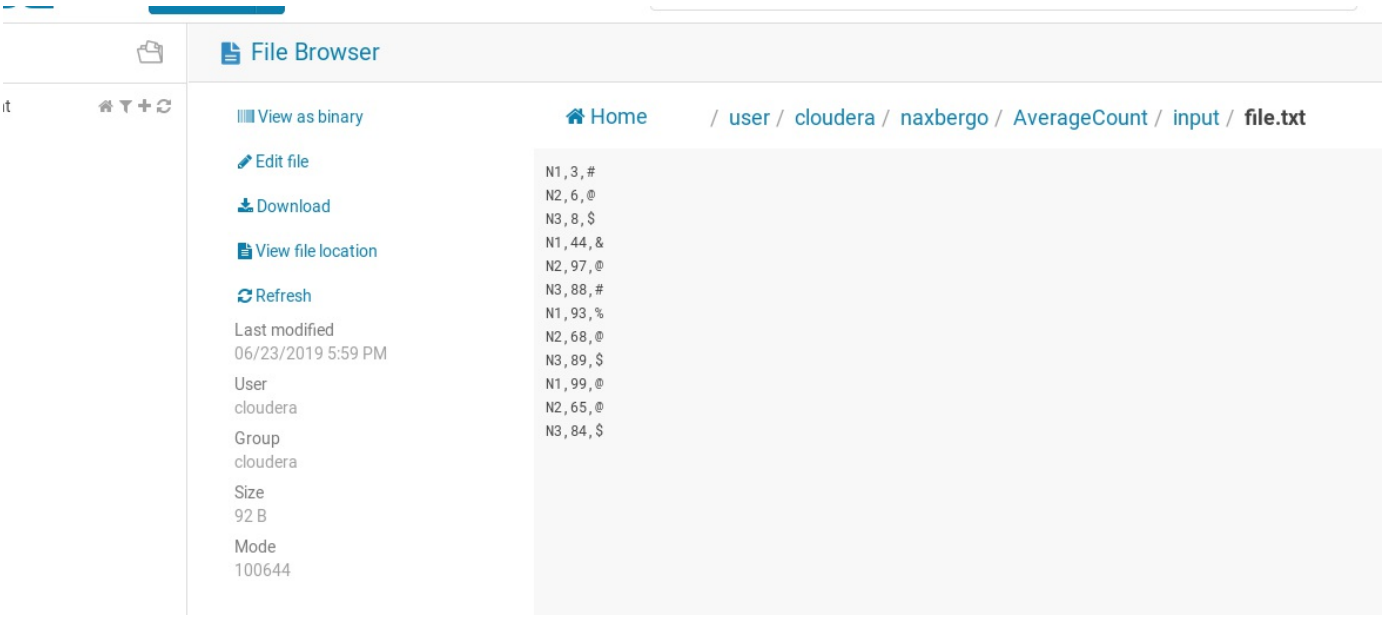
```
int count = 0;
int sum   = 0;
int result = 0;
for(IntWritable value : values){
  sum += value.get();
  count += 1;
}
result = sum / count;
context.write(key, new IntWritable(result));
```

Algorithm

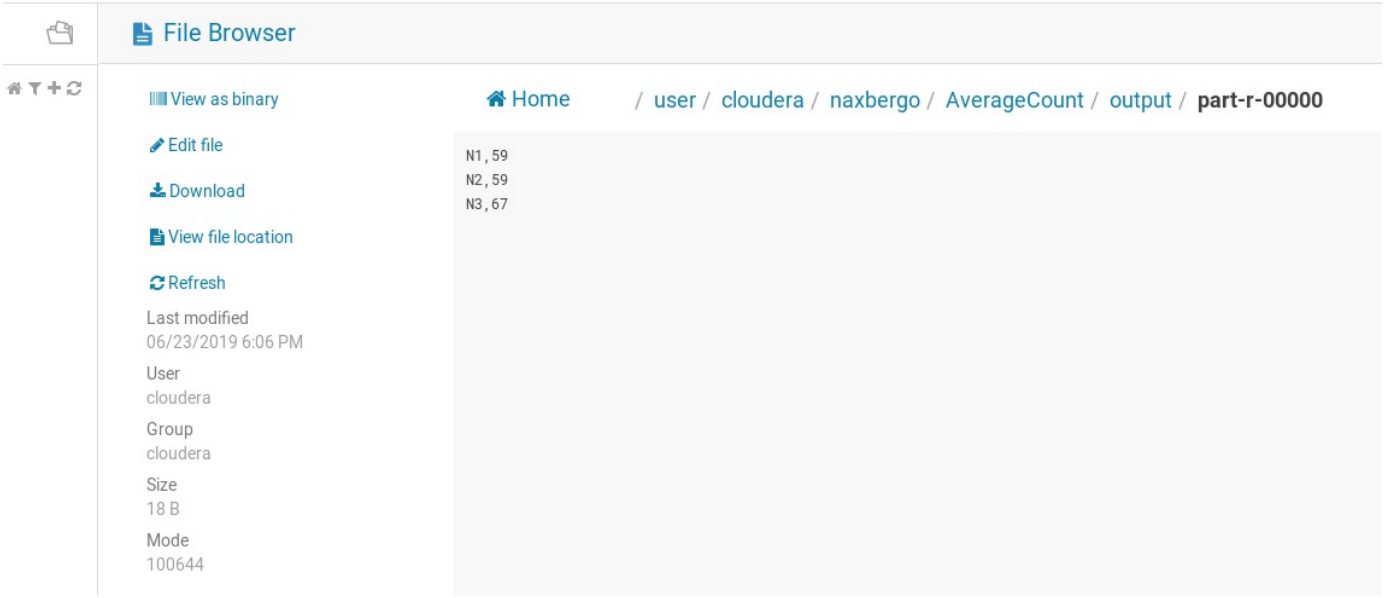


Conclusion

Input.



Output.



3. Create a Hive Table including Complex Data Types:

1. Creating hive table using zomato with complex data types as follows:

```
Time taken: 0.936 seconds
hive>
    >
    > create table zomato (restaurantid double,restaurantname string,country_code INT,city string,address array<string>,locality array<string>,locality_verbose array<string> ,longitude BIGINT,latitude BIGINT,cuisines array<string>,avgcos
t_2 int,currency string,hastable boolean,hasonline boolean,isdelinow boolean,switchtoorder boolean,pricerange int,aggregate int,ratingcolor string,ratingtext string,votes int) ROW FORMAT SERDE 'org.apache.hadoop.hive.serde2.OpenCSVSerde'

    > WITH SERDEPROPERTIES (
    >    "separatorChar" = "\t",
    >    "quoteChar"     = "'",
    >    "escapeChar"    = "\\"
    > )
    > STORED AS TEXTFILE;
OK
Time taken: 0.486 seconds
hive> load data local inpath '/home/cloudera/Downloads/zomato.csv' into table zomato;
Loading data to table default.zomato
Table default.zomato stats: [numFiles=1, totalSize=2257316]
OK
Time taken: 3.87 seconds
hive> select * from zomato limited 10;
NoViableAltException(296@[180:68: ( ( KW_AS )? alias= Identifier )?])
        at org.antlr.runtime.DFA.noViableAlt(DFA.java:158)
        at org.antlr.runtime.DFA.predict(DFA.java:116)
```

Here table is created using comples datatypes with serde properties to handle quotes and comma delimiter in zomato dataset.

2. Creating second table and upload it in the hive as follows:

```
hive>
    > create table crycode (country_code INT,Country string) ROW FORMAT SERDE 'org.apache.hadoop.hive.serde2.OpenCSVSerde'
    > WITH SERDEPROPERTIES (
    >    "separatorChar" = "\t",
    >    "quoteChar"     = "'",
    >    "escapeChar"    = "\\"
    > )
    > STORED AS TEXTFILE;
OK
Time taken: 0.252 seconds
hive> load data local inpath '/home/cloudera/crycode.txt' into table crycode;
Loading data to table default.crycode
Table default.crycode stats: [numFiles=1, totalSize=360]
OK
Time taken: 0.833 seconds
hive> select * from crycode limit 5;
OK
Country_Code    Country
1               India
14              Australia
30              Brazil
37              Canada
Time taken: 0.269 seconds, Fetched: 5 row(s)
hive>
```

Use built-in functions in your queries:

3. Max() and row_number() are the 2 built in functions we have used in hive:

```
hive> select max(avgcost_2) from zomato where avgcost_2 IS NOT NULL;
Query ID = cloudera_20190622120202_ea87a5f5-a498-4c50-b0df-cf8f0c8223c8
Total jobs = 1
Launching Job 1 out of 1
Number of reduce tasks determined at compile time: 1
In order to change the average load for a reducer (in bytes):
  set hive.exec.reducers.bytes.per.reducer=<number>
In order to limit the maximum number of reducers:
  set hive.exec.reducers.max=<number>
In order to set a constant number of reducers:
  set mapreduce.job.reduces=<number>
Starting Job = job_1560982050641_0043, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1560982050641_0043/
Kill Command = /usr/lib/hadoop/bin/hadoop job  -kill job_1560982050641_0043
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2019-06-22 12:03:25,281 Stage-1 map = 0%,  reduce = 0%
2019-06-22 12:04:20,919 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 6.28 sec
2019-06-22 12:05:10,661 Stage-1 map = 100%,  reduce = 100%, Cumulative CPU 13.13 sec
MapReduce Total cumulative CPU time: 13 seconds 130 msec
Ended Job = job_1560982050641_0043
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1  Reduce: 1   Cumulative CPU: 13.13 sec   HDFS Read: 2267008 HDFS Write: 13 SUCCESS
Total MapReduce CPU Time Spent: 13 seconds 130 msec
OK
Gayyolu
Time taken: 160.224 seconds, Fetched: 1 row(s)
```

4. The row_number() function is useful while handling sql query in database:

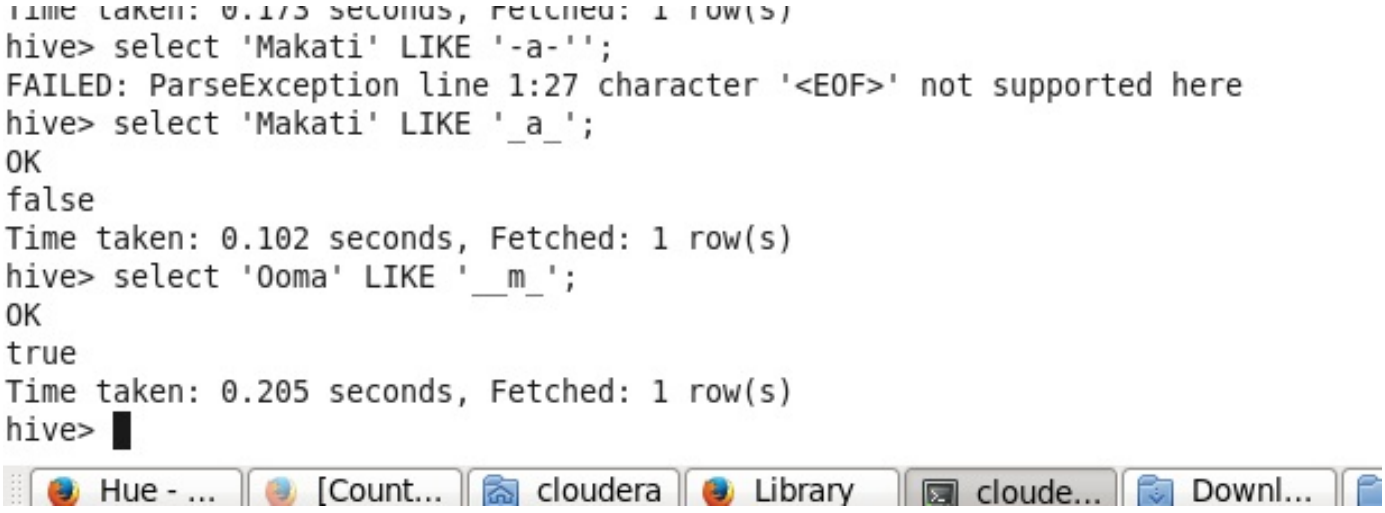
```
hive> select * from zomato limit 5;
OK
6000871 Gokurara Sofrası       208     Ankara  "Emek Mahallesi  Bosna Hersek Caddesi    No 22/C         Yukarı Ankaya         Ankara"         Emek    "Emek    Ankara"         32.81883333     39.91666667     "Kebab   Izgara"         60
sh Lira(TL)     No      No      No      1
18222559        (Niche) - Cafe & Bar    1       New Delhi       "2nd & 3rd Floor         M-16    M Block         Outer Circle    Connaught Place         New Delhi"      Connaught Place "Connaught Place         New Delhi"      77.2
8.6315156       "North Indian    Chinese         Italian         Continental"   1500    Indian Rupees(Rs.)      2
7100938 wagamama        148     Wellington City "33 Customhouse Quay     Wellington Central      Wellington City 6011"  Wellington Central      "Wellington Central      Wellington City"        174.7792237     -41.28303381    "Jap
Asian"  70      NewZealand($)   No      No      No      No      4       3
6401789 tashas  189     Cape Town       "Ground Level    Victoria Wharf  V & A Waterfront        Cape Town"      V & A Waterfront        "V & A Waterfront        Cape Town"      18.421341       -33.902336      "Cafe    Mediterrane
and(R)  No      No      No      No      4
18361747        t Lounge by Dilmah      1       New Delhi       "Flat 44         1st Floor       Khan Market     New Delhi"      Khan Market     "Khan Market     New Delhi"      77.2271332      28.6010869      "Cafe    Tea     Des
00      Indian Rupees(Rs.)      No      No      5
Time taken: 0.242 seconds, Fetched: 5 row(s)
hive>
```

The rowid is column is inserted at the end of the column as follows:

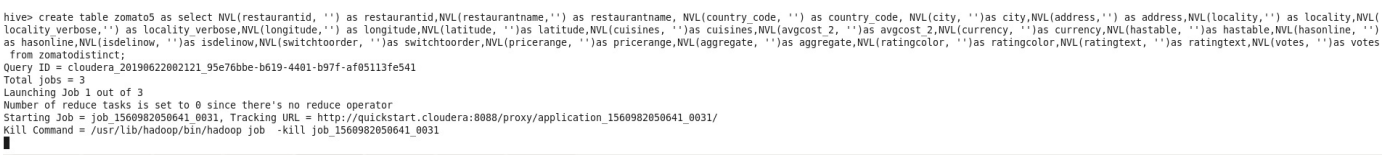
```
FAILED: SemanticException [Error 10011]: Line 1:7 Invalid function 'rowid'
hive> SELECT ROW_NUMBER() OVER () as row_num FROM zomato;
Query ID = cloudera_20190621192626_3118008a-0e29-4cf1-84bc-f2c0e64a1338
Total jobs = 1
Launching Job 1 out of 1
Number of reduce tasks not specified. Estimated from input data size: 1
In order to change the average load for a reducer (in bytes):
  set hive.exec.reducers.bytes.per.reducer=<number>
In order to limit the maximum number of reducers:
  set hive.exec.reducers.max=<number>
In order to set a constant number of reducers:
  set mapreduce.job.reduces=<number>
Starting Job = job_1560982050641_0007, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1560982050641_0007/
Kill Command = /usr/lib/hadoop/bin/hadoop job  -kill job_1560982050641_0007
```

C.Perform 10 intuitive questions in Dataset (e.g.: pattern recognition, topic discussion, most important terms, etc.). Use your innovation to think out of box.

5. Pattern recognition is done as follows:



6. Nullvalue deletion in the hive table:



7. By selecting the notnull values and creating the rownumber we will delete most of the null values inside the table while doing

the table cleaning process:

```
FAILED: ParseException line 2:0 missing EOF at 'FROM' near '10'
hive> SELECT address, COALESCE(aggregate),'Others' AS SO, COUNT(*)
    > FROM zomato limit 10
    > WHERE address IS NOT NULL AND aggregate IS NOT NULL
    > GROUP BY aggregate,address;
FAILED: ParseException line 3:0 missing EOF at 'WHERE' near '10'
hive> SELECT address, COALESCE(aggregate),'Others' AS SO, COUNT(*)
    > FROM zomato limit 10;
FAILED: SemanticException [Error 10025]: Line 1:7 Expression not in GROUP BY key 'address'
hive>
    > SELECT address, COALESCE(aggregate),'Others' AS SO, COUNT(*)
    > FROM zomato
    > WHERE address IS NOT NULL AND aggregate IS NOT NULL
    > GROUP BY aggregate,address limit 5;
Query ID = cloudera_20190622115555_0b89dae4-d48b-47cf-9a06-cfd3ec2adaaa
Total jobs = 1
Launching Job 1 out of 1
Number of reduce tasks not specified. Estimated from input data size: 1
In order to change the average load for a reducer (in bytes):
  set hive.exec.reducers.bytes.per.reducer=<number>
In order to limit the maximum number of reducers:
  set hive.exec.reducers.max=<number>
In order to set a constant number of reducers:
  set mapreduce.job.reduces=<number>
Starting Job = job_1560982050641_0042, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1560982050641_0042/
Kill Command = /usr/lib/hadoop/bin/hadoop job  -kill job_1560982050641_0042
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2019-06-22 11:55:55,043 Stage-1 map = 0%,  reduce = 0%
2019-06-22 11:56:44,750 Stage-1 map = 67%,  reduce = 0%, Cumulative CPU 7.27 sec
2019-06-22 11:56:49,062 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 9.34 sec
2019-06-22 11:57:37,789 Stage-1 map = 100%,  reduce = 67%, Cumulative CPU 14.5 sec
2019-06-22 11:57:44,514 Stage-1 map = 100%,  reduce = 100%, Cumulative CPU 17.28 sec
MapReduce Total cumulative CPU time: 17 seconds 280 msec
Ended Job = job_1560982050641_0042
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1  Reduce: 1   Cumulative CPU: 17.28 sec   HDFS Read: 2268545 HDFS Write: 135 SUCCESS
Total MapReduce CPU Time Spent: 17 seconds 280 msec
OK
"1st Floor       American       Others  2
"2       American       Others  1
"201-202         American       Others  1
"2526    American       Others  1
"2nd Floor       American       Others  2
```

8. By creating the distinct table so that the duplicate values are deleted in the table which is also a cleaning process:

```
hive> create table zomato2 as select distinct restaurantid, restaurantname, country_code, city, address, locality, locality_verbose, longitude, latitude, cuisines, avgcost_2, currency,
range, aggregate, ratingcolor, ratingtext, votes from zomato;
Query ID = cloudera_20190622123333_ccbf6ed4-9af6-4cc7-96c9-4039d1319df0
Total jobs = 1
Launching Job 1 out of 1
Number of reduce tasks not specified. Estimated from input data size: 1
In order to change the average load for a reducer (in bytes):
  set hive.exec.reducers.bytes.per.reducer=<number>
In order to limit the maximum number of reducers:
  set hive.exec.reducers.max=<number>
In order to set a constant number of reducers:
  set mapreduce.job.reduces=<number>
Starting Job = job_1560982050641_0048, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1560982050641_0048/
Kill Command = /usr/lib/hadoop/bin/hadoop job  -kill job_1560982050641_0048
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2019-06-22 12:34:17,516 Stage-1 map = 0%,  reduce = 0%
2019-06-22 12:35:05,892 Stage-1 map = 67%,  reduce = 0%, Cumulative CPU 7.24 sec
2019-06-22 12:35:13,197 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 10.7 sec
2019-06-22 12:36:00,855 Stage-1 map = 100%,  reduce = 68%, Cumulative CPU 16.04 sec
2019-06-22 12:36:06,380 Stage-1 map = 100%,  reduce = 99%, Cumulative CPU 19.98 sec
2019-06-22 12:36:07,745 Stage-1 map = 100%,  reduce = 100%, Cumulative CPU 20.67 sec
MapReduce Total cumulative CPU time: 20 seconds 670 msec
Ended Job = job_1560982050641_0048
Moving data to: hdfs://quickstart.cloudera:8020/user/hive/warehouse/zomato2
Table default.zomato2 stats: [numFiles=1, numRows=10554, totalSize=1887391, rawDataSize=1876837]
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1  Reduce: 1   Cumulative CPU: 20.67 sec   HDFS Read: 4521351 HDFS Write: 1887470 SUCCESS
Total MapReduce CPU Time Spent: 20 seconds 670 msec
OK
Time taken: 162.953 seconds
hive>
```

9. Creating views in a database is always known for hiding certain

personal information so that everyone cannot see the hidden view in a database.

```
Time taken: 141.942 seconds, Fetched: 1 row(s)
hive> create view if not exists zomato_v as select * from zomato where city = 'Agra';
OK
Time taken: 0.533 seconds
hive> select * from zomato_v;
OK
3400025 Jahanpanah      1       Agra    "E 23    Shopping Arcade         Sadar Bazaar    Agra Cantt      Agra"  Agra Cantt      "Agra Cantt      Agra"  78.01154444     27.16166111     "North Indian    Mughlai"       850     Indian Ru
(Rs.)   No      No      No      No
3400341 Rangrezz Restaurant     1       Agra    "E-20    Shopping Arcade         Sadar Bazaar    Agra Cantt      Agra"  Agra Cantt      "Agra Cantt      Agra"  0       0       "North Indian    Mughlai"       700     Indian Rupees(Rs.
o       No      No      No
3400005 Time2Eat - Mama Chicken 1       Agra    "Main Market     Sadar Bazaar    Agra Cantt      Agra"  Agra Cantt      "Agra Cantt      Agra"  78.01160797     27.16083249     North Indian    500     Indian Rupees(Rs.)      No      N
o       No      2       3.6
3400021 Chokho Jeeman Marwari Jain Bhojanalya   1       Agra    "1/48    Delhi Gate      Station Road    Raja Mandi      Civil Lines     Agra"  Civil Lines     "Civil Lines     Agra"  77.99809167     27.19592778     Rajasthani      4
ndian Rupees(Rs.)       No      No      No      No
3400017 Pinch Of Spice  1       Agra    "23/453  Opposite Sanjay Cinema  Wazipura Road   Sanjay Place    Civil Lines     Agra"  Civil Lines     "Civil Lines     Agra"  78.00755278     27.201725       "North Indian    Chinese
lai"    1000    Indian Rupees(Rs.)      No      No
3400325 MoMo Cafe       1       Agra    "Courtyard by Marriott Agra      Phase 2         Fatehabad Road  Tajganj         Agra"  "Courtyard by Marriott Agra      Tajganj"       "Courtyard by Marriott Agra      Tajganj         Agra"  0
"North Indian    European"      2000    Indian Rupees(Rs.)      No      No
3400059 Peshawri - ITC Mughal   1       Agra    "ITC Mughal      Fatehabad Road  Tajganj         Agra"  "ITC Mughal      Tajganj"       "ITC Mughal      Tajganj         Agra"  78.044095       27.160934       "North Indian    Mughlai"
500     Indian Rupees(Rs.)      No      No      No
3400060 Taj Bano - ITC Mughal   1       Agra    "ITC Mughal      Fatehabad Road  Tajganj         Agra"  "ITC Mughal      Tajganj"       "ITC Mughal      Tajganj         Agra"  78.044095       27.160934       Mughlai 2500    Indian Ru
(Rs.)   No      No      No      No
3400348 G Thal  1       Agra    "3/20    KPS Tower       Near Tulsi Talkies      Bypass Road     Khandari        Agra"  Khandari        "Khandari        Agra"  0       0       "Rajasthani      Gujarati        Mughlai"       800     I
n Rupees(Rs.)   No      No
3400072 Dawat-e-Nawab - Radisson Blu    1       Agra    "Radisson Blu    Taj East Gate Road      Tajganj         Agra"  "Radisson Blu    Tajganj"       "Radisson Blu    Tajganj         Agra"  78.057044       27.163303       "North In
Mughlai"        3600    Indian Rupees(Rs.)      No      No      No
3400073 The Latitude - Radisson Blu     1       Agra    "Radisson Blu    Taj East Gate Road      Tajganj         Agra"  "Radisson Blu    Tajganj"       "Radisson Blu    Tajganj         Agra"  78.057044       27.163303       "North In
 Chinese         Continental"    0       Indian Rupees(Rs.)      No      No
3400019 Dasaprakash Restaurant  1       Agra    "Meher Cinema Complex    Gwalior Road    Rakabganj       Agra"  Rakabganj       "Rakabganj       Agra"  78.01505278     27.16510833     "South Indian    Desserts"      550     Indian Ru
(Rs.)   No      No      No      2
3400033 The Charcoal Chimney    1       Agra    "Hotel Samovar   Fatehabad Road  Tajganj         Agra"  Tajganj "Tajganj         Agra"  78.04299        27.159795       "North Indian    Chinese         Continental     Mughlai"       1
ndian Rupees(Rs.)       No      No      No
3400346 Sheroes Hangout 1       Agra    "Opposite The Gateway Hotel      Fatehabad Road  Tajganj         Agra"  Tajganj "Tajganj         Agra"  78.040165       27.16185        "Cafe    North Indian    Chinese"       0       Indian Ru
(Rs.)   No      No      No      No
3400350 Bon Barbecue    1       Agra    "Parador Hotel   3A-3B   Phase 1         Fatehabad Road  Taj Nagri       Tajganj         Agra"  Tajganj "Tajganj         Agra"  0       0       "North Indian    Chinese         Continental"   1
ndian Rupees(Rs.)       No
3400391 Chapter 1 Cafe  1       Agra    "1374 K/1375 K   2nd floor       Dinesh Nagar    Fatehbad Road   Tajganj         Agra"  Tajganj "Tajganj         Agra"  0       0       "Cafe    Italian         Mexican         North Indian
inental"        0       Indian Rupees(Rs.)
3400016 Pind Balluchi   1       Agra    "Opposite Saga Emporium  Fatehabad Road  Tajganj         Agra"  Tajganj "Tajganj         Agra"  78.04725        27.15777222     "North Indian    Mughlai"       900     Indian Rupees(Rs.)      N
o       No      No      3
3400105 Pizza Hut       1       Agra    "8       Handicraft Nagar        Fatehabad Road  Tajganj         Agra"  Tajganj "Tajganj         Agra"  78.03471389     27.16169444     "Italian         Pizza" 700      Indian Rupees(Rs.)      N
o       No      No
3400392 Thaaliwala      1       Agra    "Next To ITC Mughal Hotel        Fatehabad Road  Tajganj         Agra"  Tajganj "Tajganj         Agra"  78.045359       27.158822       "North Indian    Fast Food"     700     Indian Rupees(Rs.
o       No      No      No      2
Time taken: 0.576 seconds, Fetched: 19 row(s)
hive>
```


5 r

10. AS in the columns in the dataset are similar hence we perform joints as follows:

1.Equijoin:

```
hive>
    > select * from crycode C join zomato2 C1 ON C.country_code = C1.country_code limit 10;
Query ID = cloudera_20190622132121_4df67b8a-d65d-420b-9b5a-b8c5dd22978a
Total jobs = 1
Execution log at: /tmp/cloudera/cloudera_20190622132121_4df67b8a-d65d-420b-9b5a-b8c5dd22978a.log
2019-06-22 01:22:24     Starting to launch local task to process map join;      maximum memory = 1013645312
2019-06-22 01:22:36     Dump the side-table for tag: 0 with group count: 17 into file: file:/tmp/cloudera/3e1945cf-a1e2-4b3b-ba00-25adbe5c8a68/hive_2019-06-22_13-21-49_391_6059897477515869597-1/-local-10003/HashTable-Stage-3/MapJoin-
ile20--.hashtable
2019-06-22 01:22:36     Uploaded 1 File to: file:/tmp/cloudera/3e1945cf-a1e2-4b3b-ba00-25adbe5c8a68/hive_2019-06-22_13-21-49_391_6059897477515869597-1/-local-10003/HashTable-Stage-3/MapJoin-mapfile20--.hashtable (915 bytes)
2019-06-22 01:22:36     End of local task; Time Taken: 12.207 sec.
Execution completed successfully
MapredLocal task succeeded
Launching Job 1 out of 1
Number of reduce tasks is set to 0 since there's no reduce operator
Starting Job = job_1560982050641_0052, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1560982050641_0052/
Kill Command = /usr/lib/hadoop/bin/hadoop job  -kill job_1560982050641_0052
Hadoop job information for Stage-3: number of mappers: 1; number of reducers: 0
2019-06-22 13:23:34,143 Stage-3 map = 0%,  reduce = 0%
2019-06-22 13:24:20,535 Stage-3 map = 100%,  reduce = 0%, Cumulative CPU 5.94 sec
MapReduce Total cumulative CPU time: 5 seconds 940 msec
Ended Job = job_1560982050641_0052
MapReduce Jobs Launched:
Stage-Stage-3: Map: 1   Cumulative CPU: 5.94 sec   HDFS Read: 15080 HDFS Write: 2019 SUCCESS
Total MapReduce CPU Time Spent: 5 seconds 940 msec
OK
1               India   100080  Chokhi Dhani    1       Jaipur  "Chokhi Dhani Village Resort     12 Mile         Tonk Road       Jaipur"         "Chokhi Dhani Village Resort     Tonk Road"     "Chokhi Dhani Village Resort     Tonk Roa
Jaipur" 75.837333       26.766536       Rajasthani      1600    Indian Rupees(Rs.)      No      No      No
1               India   100305  The Forresta Kitchen & Bar      1       Jaipur  "Devraj Niwas    Near Moti Mahal Cinema  Khasa Kothi Crossing    Gopalbari       Jaipur"         "Devraj Niwas    Bani Park"     "Devraj Niwas    Bani Par
Jaipur" 75.7936039      26.9214108      "Continental    Mexican         Beverages       Desserts        North Indian
1               India   100306  Replay  1       Jaipur  "SB 57   5th Floor       Ridhi Tower     Opposite SMS Stadium    Tonk Road       Jaipur"         Tonk Road       "Tonk Road       Jaipur"         75.80689587     26.8923125      '
h Indian        Continental     Chinese         Italian         Mexican"        1500
1               India   101212  Tapri Central   1       Jaipur  "B4 E    3rd Floor       Surana Jewellers        Opposite Central Park   C Scheme        Jaipur"         C Scheme        "C Scheme        Jaipur"         75.81075322     2
518991  "Cafe    Fast Food       Street Food"   750     Indian Rupees(Rs.)      No
1               India   101259  On The House    1       Jaipur  "E 145   Ramesh Marg     Behind Talwalkars       C Scheme        Jaipur"         C Scheme        "C Scheme        Jaipur"         75.7945919      26.9052866      "Italian
Continental      Mexican         Cafe    Bakery"         1300    Indian Rupees(Rs.)
1               India   101311  Taruveda Bistro 1       Jaipur  "1st Floor       Sunraj Villa    2 Mysore House  Jacob Road      Civil Lines     Jaipur"         Civil Lines     "Civil Lines     Jaipur"         75.7890337      26.911377
Cafe     Italian         Japanese        Continental"   1000    Indian Rupees(Rs.)
1               India   1018    Mainland China  1       New Delhi       "Plot 4  Local Shopping Centre   Masjid Moth     Greater Kailash (GK) 2  New Delhi"      Greater Kailash (GK) 2  "Greater Kailash (GK) 2  New Delhi"     77.23867
8.5373346       Chinese 1800    Indian Rupees(Rs.)      Yes     Yes     No      No
1               India   101834  Blackout        1       Jaipur  "Hotel Golden Oak        9th Floor       Ahinsa Circle   Landmark Building       C Scheme        Jaipur"         C Sc                    ur"     75.80570
6.9141424       "North Indian    European        Continental     Finger Food"   1650    Indian Rupees(Rs.)
                101884
                101884
Time taken: 153.67 seconds, Fetched: 10 row(s)
hive>
```

2.Right outer join:

Thus using the dataset of 9558 rows and 22 column including rowid we have cleaned the dataset and got the output for all important queries that is necessary for the corporate culture.

a. Create a Solr Collection including our own Field Types

b. Perform 10 intuitive questions in Dataset (e.g.: pattern recognition, topic discussion, most important terms, etc.). Use your innovation to think out of box.

c. Implement at least 5 nested

queries among the 10.

d. Record the time execution for the queries.

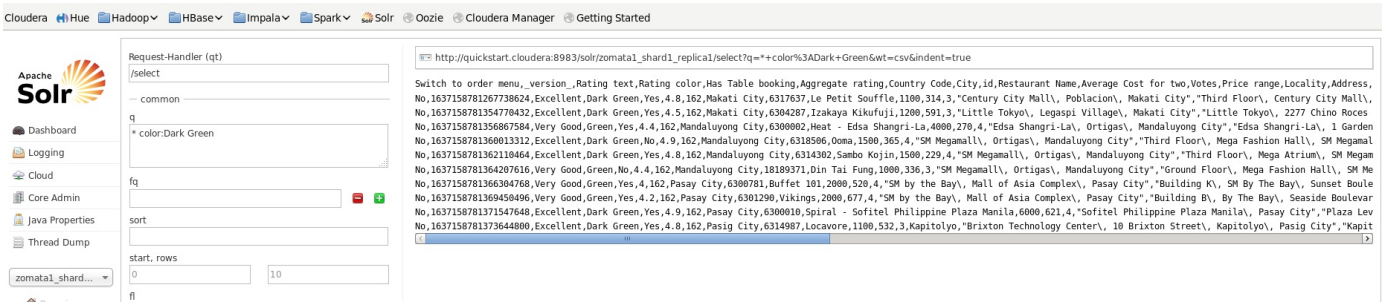
Dataset

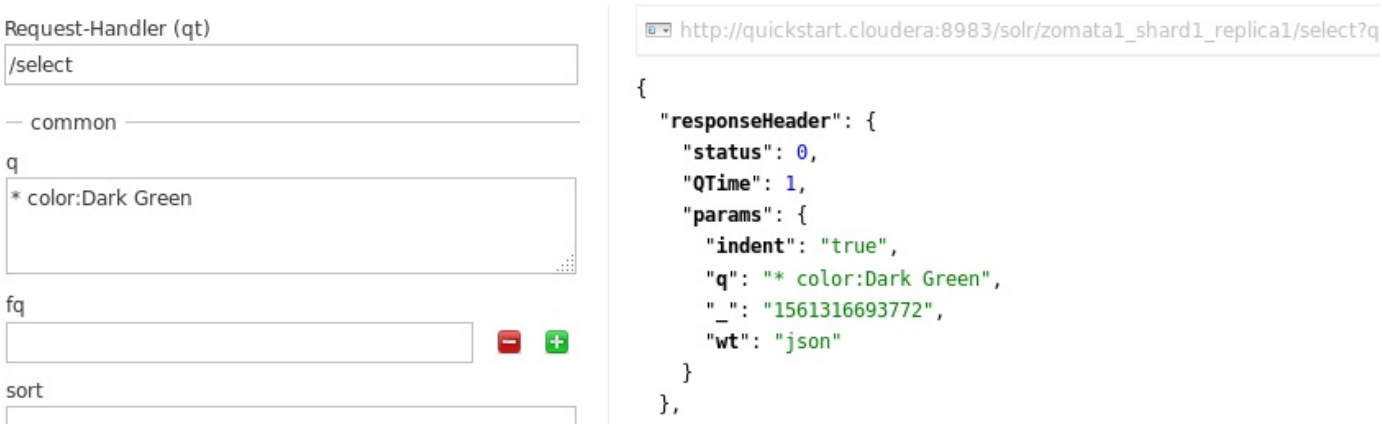
A. Zomato Restaurants Data

https://www.kaggle.com/shrutimehta/zomato-restaurants-data

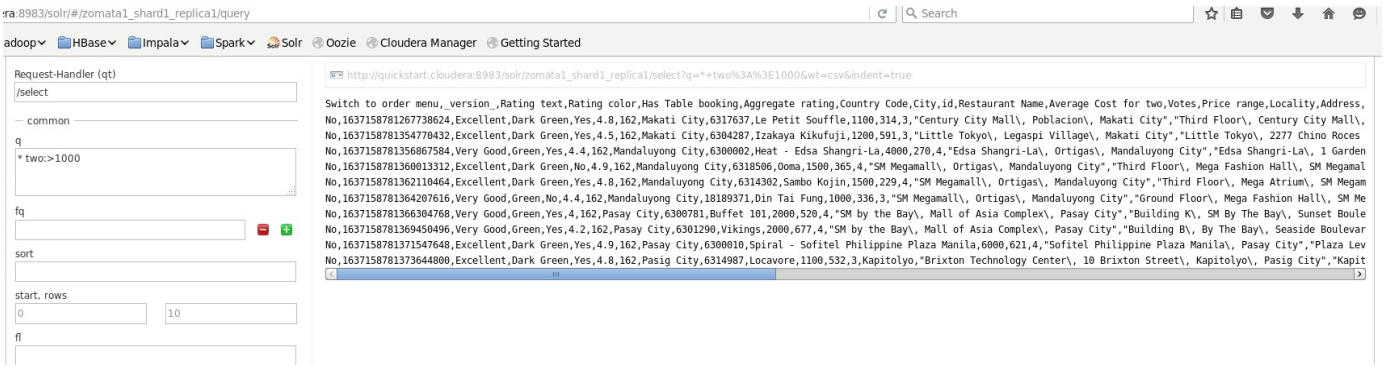
Queries

1. Number of rows with data color Green.



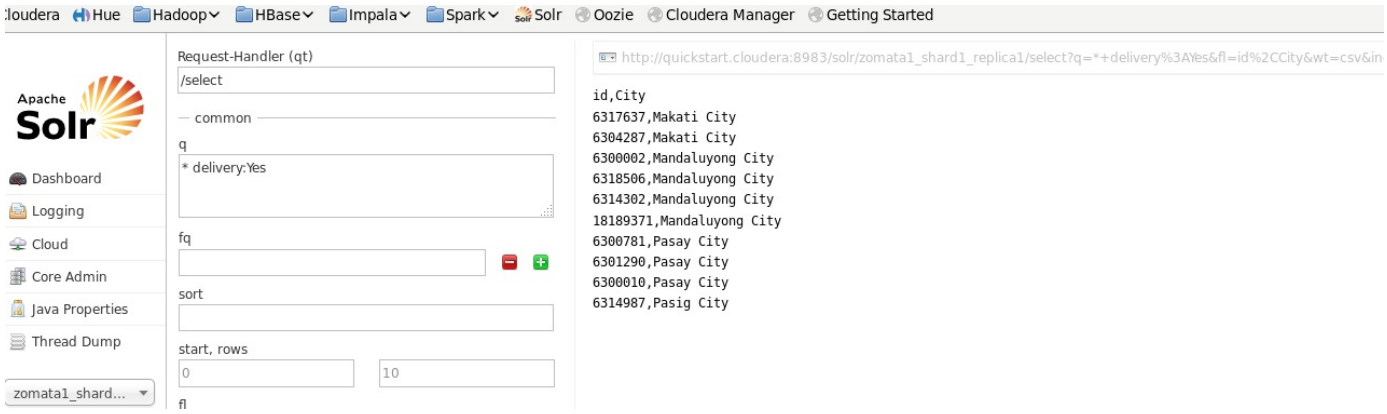


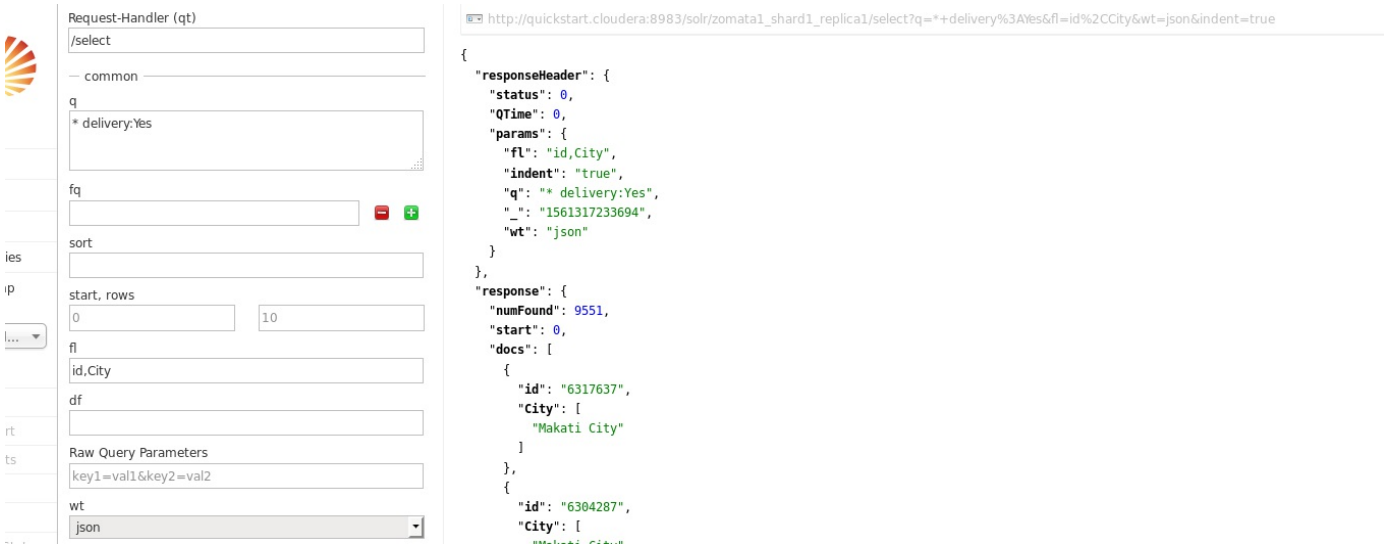
2. Average cost for two people greater than 1000.



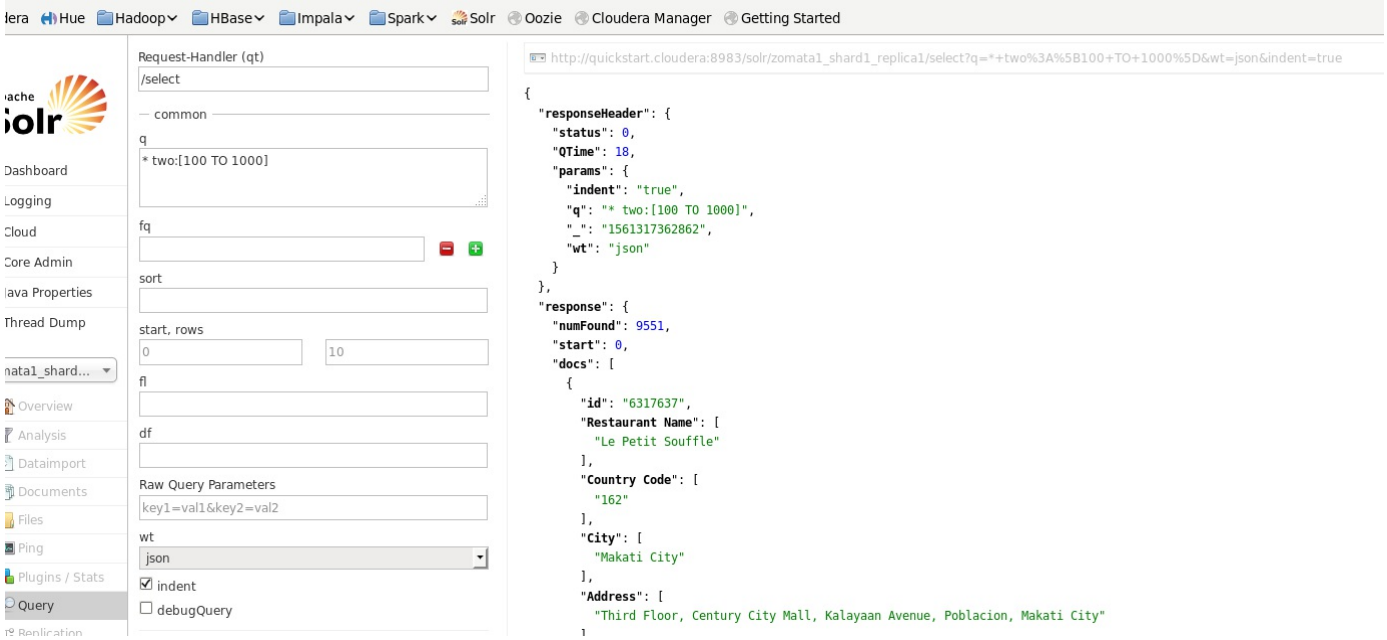
•

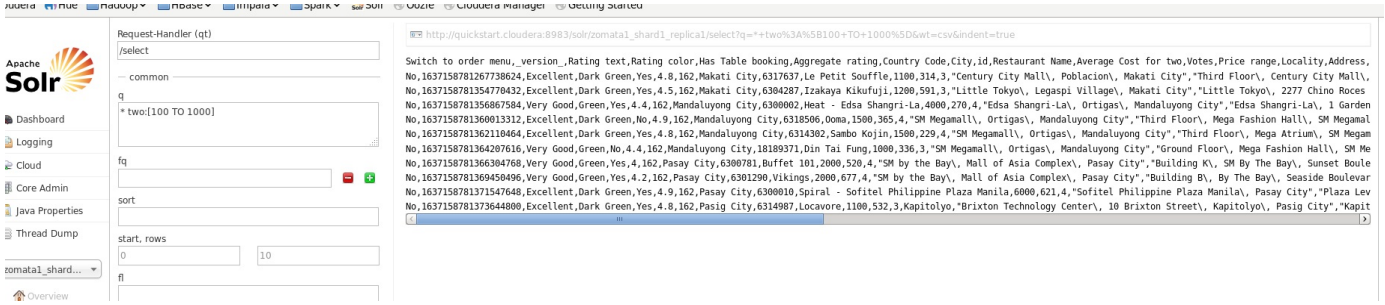
3. Number of rows with delivery yes.



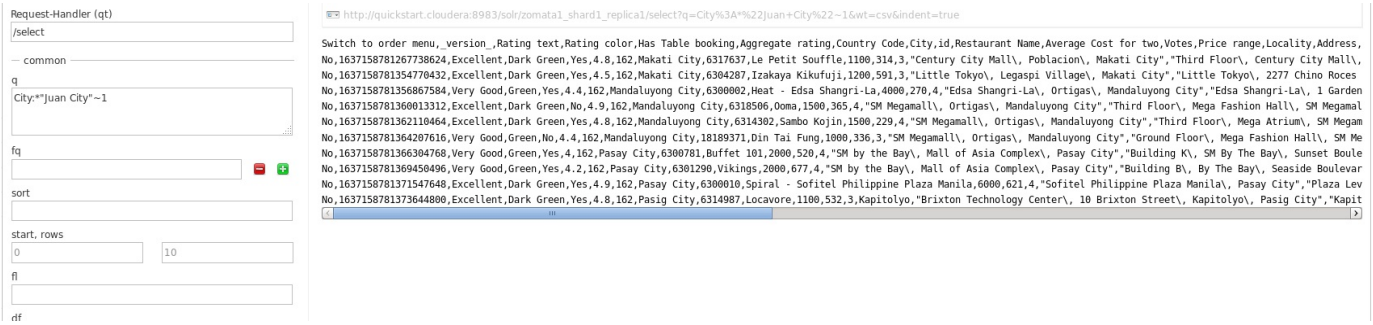


4. Number of rows with range two people 100 to 1000.

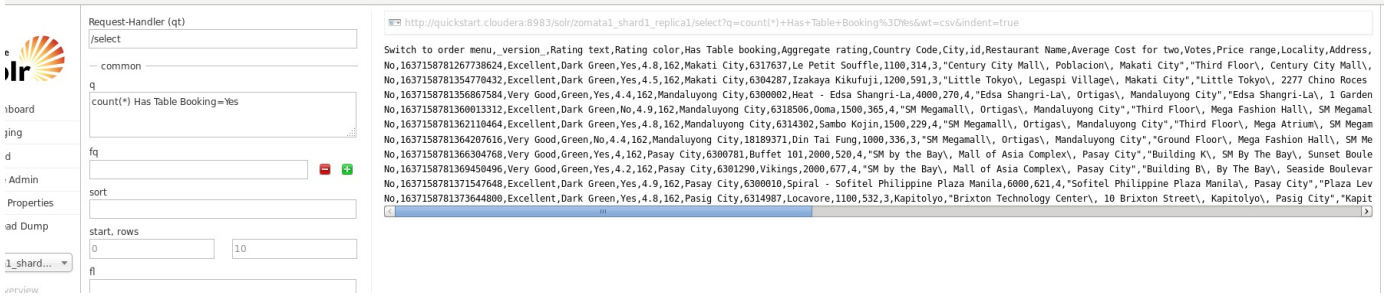


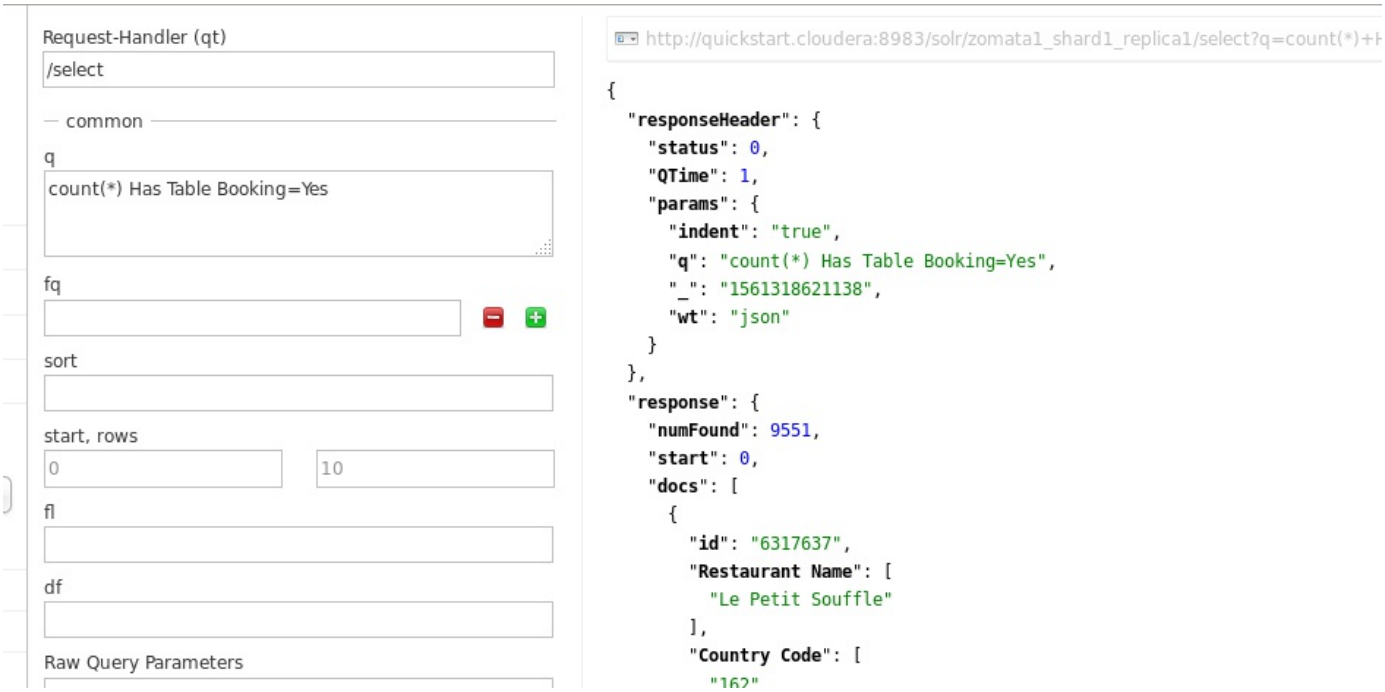


5. Number of rows with Juan City.

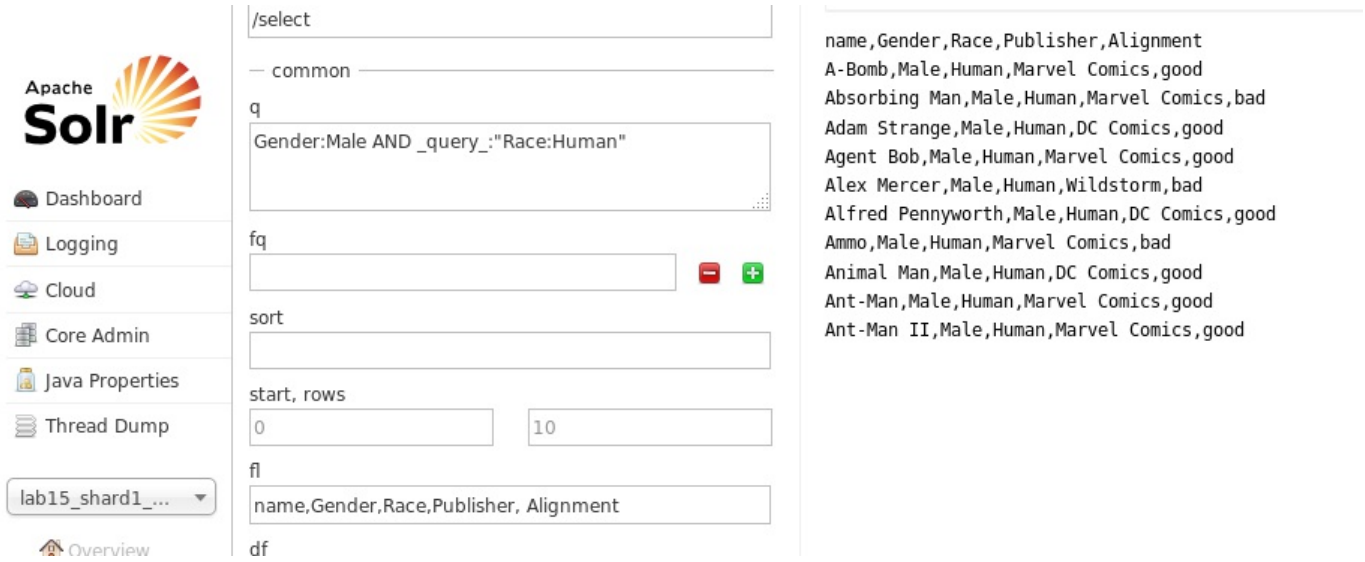


6. Count the Number of rows with delivery yes.





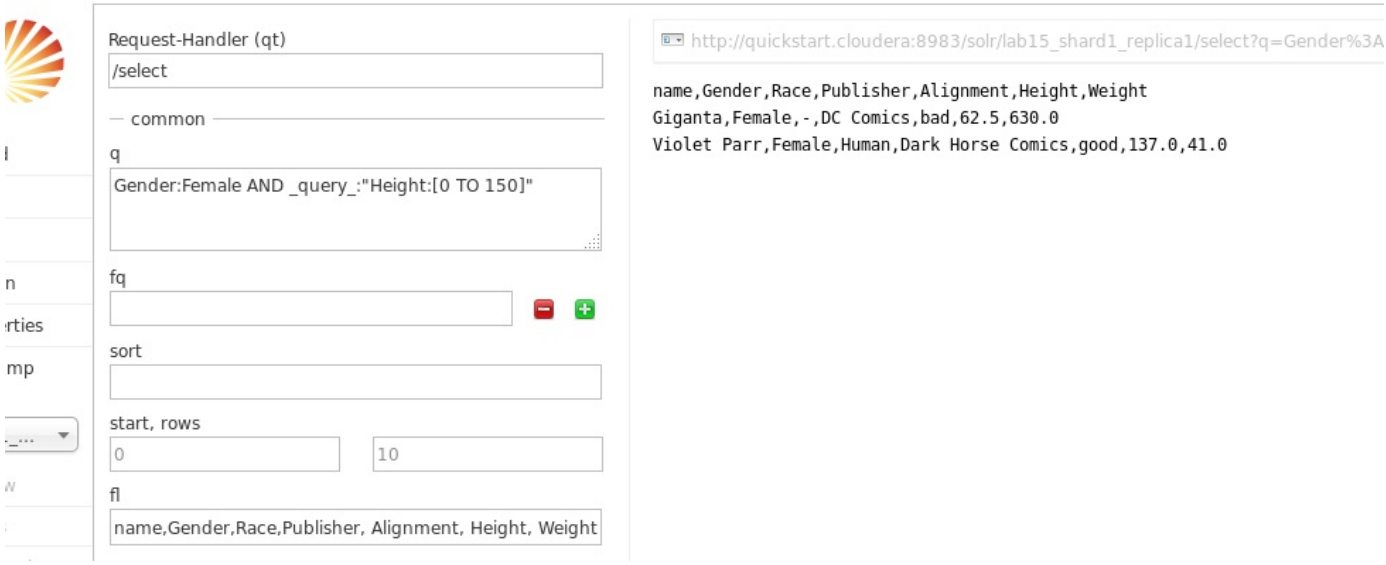
7. Selecting the rows with Human raced Male super heroes.



Elapsed Time: 3msec.

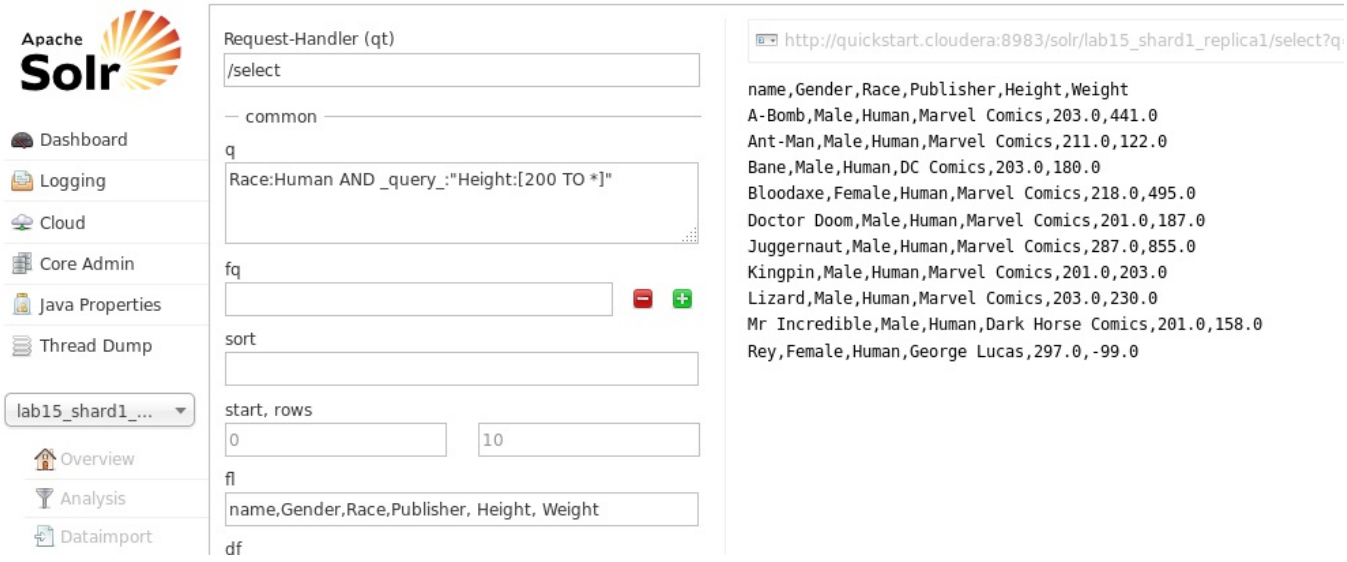
8. Selecting the rows with 0<height<150 female super

heroes.



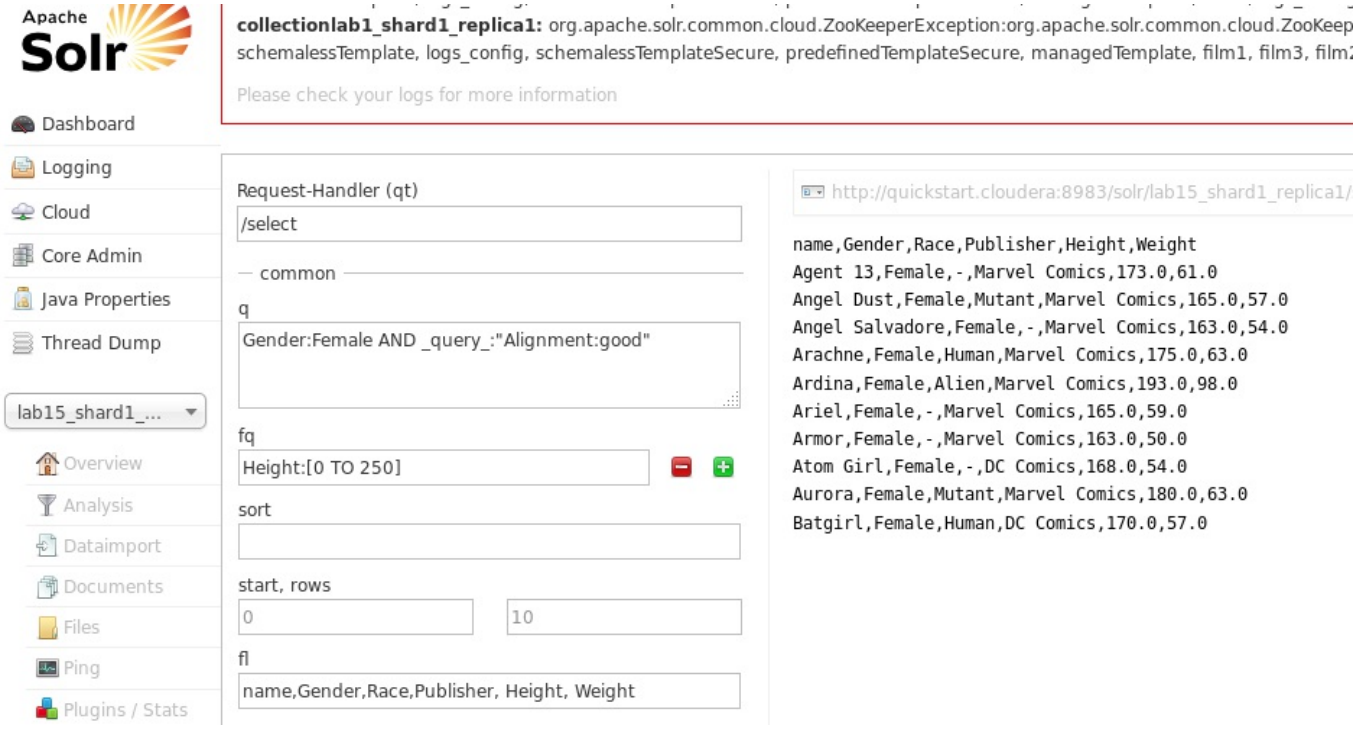
Elapsed Time: 5msec.

9. Selecting the no of rows with Height>200 Human race super heroes



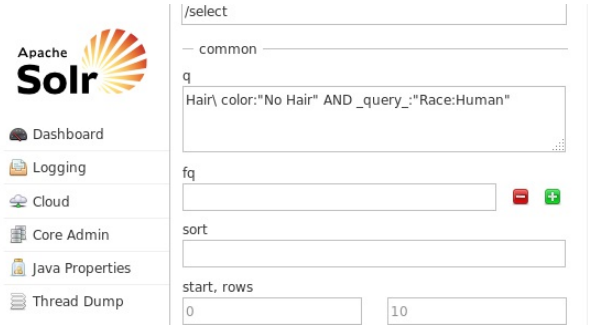
Elapsed Time: 10msec.

10. Selecting the no of rows with the super heroes who are females and having good



Elapsed Time: 6msec.

11. Selecting the rows with the superheroes who are not having hair and from human race




```
Publisher,id,Eye color,_version_,Weight,Gender,name,Hair color,Race,Height,Alignment,Skin color
Marvel Comics,0,yellow,1637178915704274944,441.0,Male,A-Bomb,No Hair,Human,203.0,good,-
Marvel Comics,5,blue,1637178915731537920,122.0,Male,Absorbing Man,No Hair,Human,193.0,bad,-
DC Comics,104,black,1637178915844784128,92.0,Male,Black Manta,No Hair,Human,188.0,bad,-
DC Comics,319,blue,1637178916141531136,81.0,Male,Heat Wave,No Hair,Human,180.0,bad,-
Marvel Comics,391,blue,1637178916301963264,203.0,Male,Kingpin,No Hair,Human,201.0,bad,-
Marvel Comics,392,red,1637178916301963265,97.0,Male,Klaw,No Hair,Human,188.0,bad,red
DC Comics,405,green,1637178916328177665,95.0,Male,Lex Luthor,No Hair,Human,188.0,bad,-
Marvel Comics,412,red,1637178916336566272,230.0,Male,Lizard,No Hair,Human,203.0,bad,-
Marvel Comics,479,brown,1637178916459249664,79.0,Male,Mysterio,No Hair,Human,180.0,bad,-
Shueisha,502,-,1637178916492804096,69.0,Male,One Punch Man,No Hair,Human,175.0,good,-
```

Elapsed Time: 15msec.