

Lab Assignment_4

NAME: HARISH CHANDRA JYOSHI

CLASS ID: 06

TEAM ID: 04

UMKC EMAIL ID: hjddh@mail.umkc.edu

NAME: ATLURI VENKATA AKHILA KRISHNA

CLASS ID: 01

TEAM ID: 04

UMKC EMAIL ID: vagg2@mail.umkc.edu

TASK 1:

OBJECTIVE 1.HIVE USECASE

- a)Create a Hive Table including Complex Data Types
- b)Use built-in functions in your queries
- c)Perform 10 intuitive questions in Database

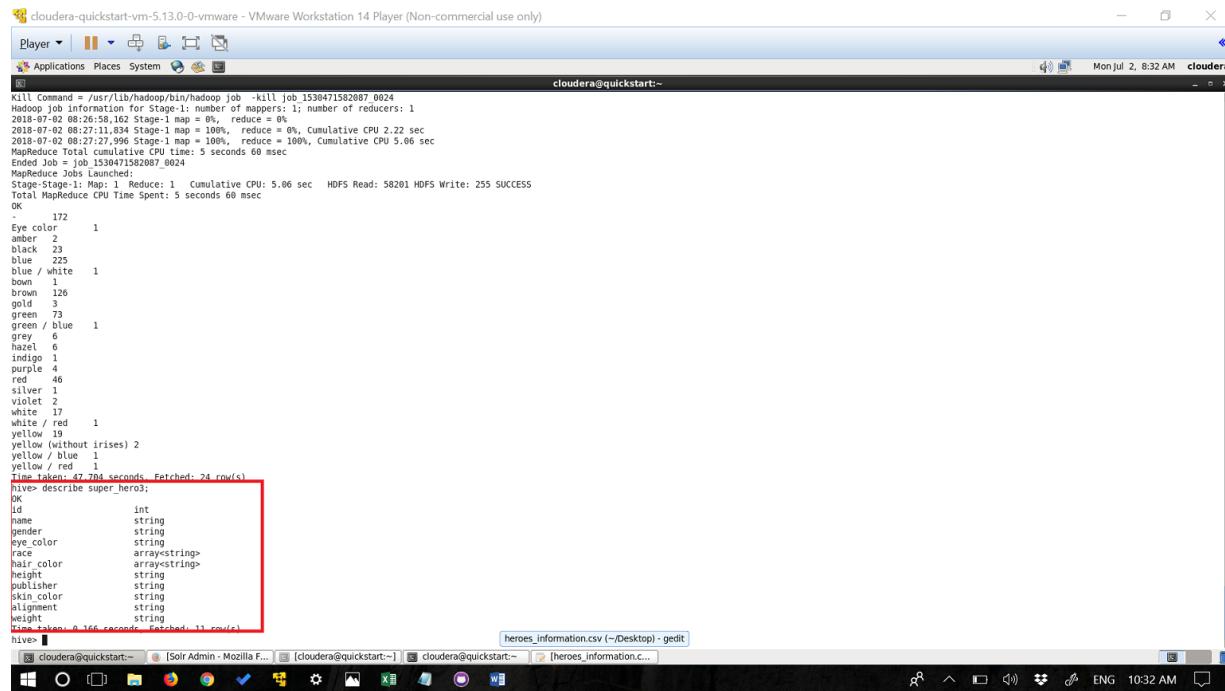
DATASET used: I have used superheroes as my dataset

Queries: I have created a hive table for superheroes dataset with array data type for haircolor and eye_color column.

I performed the following queries.

Lab Assignment_4

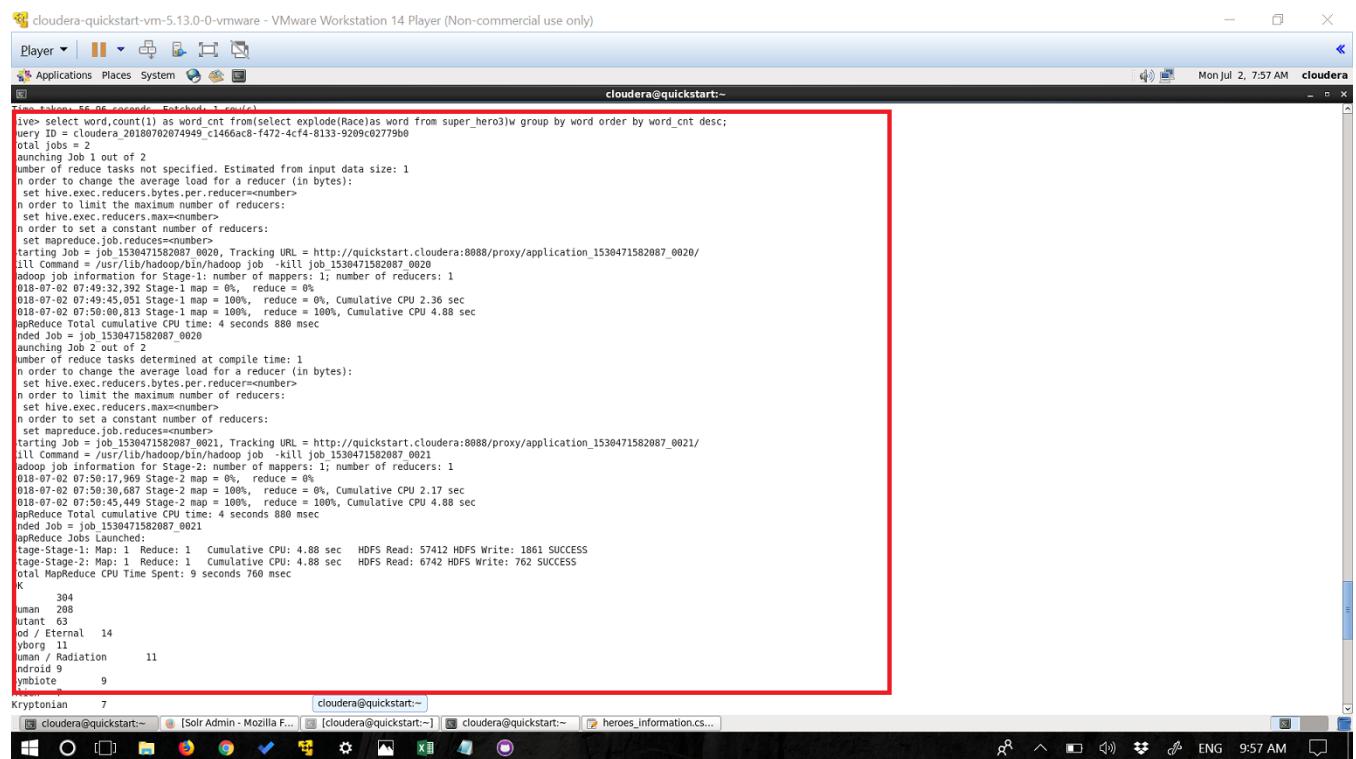
SUPER_HERO TABLE:



```
Kill Command = /usr/lib/hadoop/bin/hadoop job -kill job_1530471582087_0024
Hadoop job Information: Stage-1: number of mappers: 1; number of reducers: 1
2018-07-02 08:27:11,834 Stage-1 map = 0%, reduce = 0%, Cumulative CPU 2.22 sec
2018-07-02 08:27:27,996 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 5.06 sec
MapReduce Total cumulative CPU time: 5 seconds 60 msec
End Job = job_1530471582087_0024
MapReduce Jobs Launched
Stage-Stage-1: Map: 1 Reduce: 1 Cumulative CPU: 5.06 sec HDFS Read: 58201 HDFS Write: 255 SUCCESS
Total MapReduce CPU Time Spent: 5 seconds 60 msec
OK
      172
Eye color      1
amber          2
black          23
blue           25
blue / white   1
brown          126
gold           5
green          73
green / blue   1
grey            6
hazel          6
indigo          1
purple          4
red             46
silver          1
violet          2
white          17
white / red    1
yellow          19
yellow (without irises) 2
yellow / blue   1
yellow / red    1
Time taken: 47.704 seconds. Fetched: 24 row(s)
hive> describe super.hero;
OK
id          int
name        string
gender       string
eye_color   string
race         array<string>
hair_color  array<string>
height       string
publisher    string
skin_color   string
alignment    string
weight       string
Time taken: 0.166 seconds. Fetched: 11 rows
hive> !
```

Query1:

Wordcount



```
Time taken: 56.96 seconds. Fetched: 1 row(s)
hive> select word, count(1) as word_cnt from (select explode(Race) as word from super_hero3) w group by word order by word_cnt desc;
OK
word          word_cnt
total          1
Launching Job 1 out of 2
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_1530471582087_0020, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1530471582087_0020
Kill Command = /usr/lib/hadoop/bin/hadoop job -kill job_1530471582087_0020
adoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2018-07-02 07:49:32,392 Stage-1 map = 0%, reduce = 0%
2018-07-02 07:49:45,001 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 2.36 sec
2018-07-02 07:50:01,813 Stage-2 map = 0%, reduce = 100%, Cumulative CPU 4.88 sec
MapReduce Total cumulative CPU time: 4 seconds 888 msec
ended Job = job_1530471582087_0020
Launching Job 2 out of 2
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_1530471582087_0021, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1530471582087_0021
Kill Command = /usr/lib/hadoop/bin/hadoop job -kill job_1530471582087_0021
adoop job information for Stage-2: number of mappers: 1; number of reducers: 1
2018-07-02 07:50:17,969 Stage-2 map = 0%, reduce = 0%
2018-07-02 07:50:38,687 Stage-2 map = 100%, reduce = 0%, Cumulative CPU 2.17 sec
2018-07-02 07:50:45,449 Stage-2 map = 100%, reduce = 100%, Cumulative CPU 4.88 sec
MapReduce Total cumulative CPU time: 4 seconds 888 msec
ended Job = job_1530471582087_0021
MapReduce Jobs Launched
Stage-Stage-1: Map: 1 Reduce: 1 Cumulative CPU: 4.88 sec HDFS Read: 57412 HDFS Write: 1861 SUCCESS
Stage-Stage-2: Map: 1 Reduce: 1 Cumulative CPU: 4.88 sec HDFS Read: 6742 HDFS Write: 762 SUCCESS
Total MapReduce CPU Time Spent: 9 seconds 760 msec
K
      304
human          208
utant          39
od / Ettmal    14
yborg          11
luman / Radiation  11
ndroid         9
ymbiote        9
Kryptonian    7
Time taken: 1.01 seconds. Fetched: 1 row(s)
hive> !
```

Lab Assignment_4

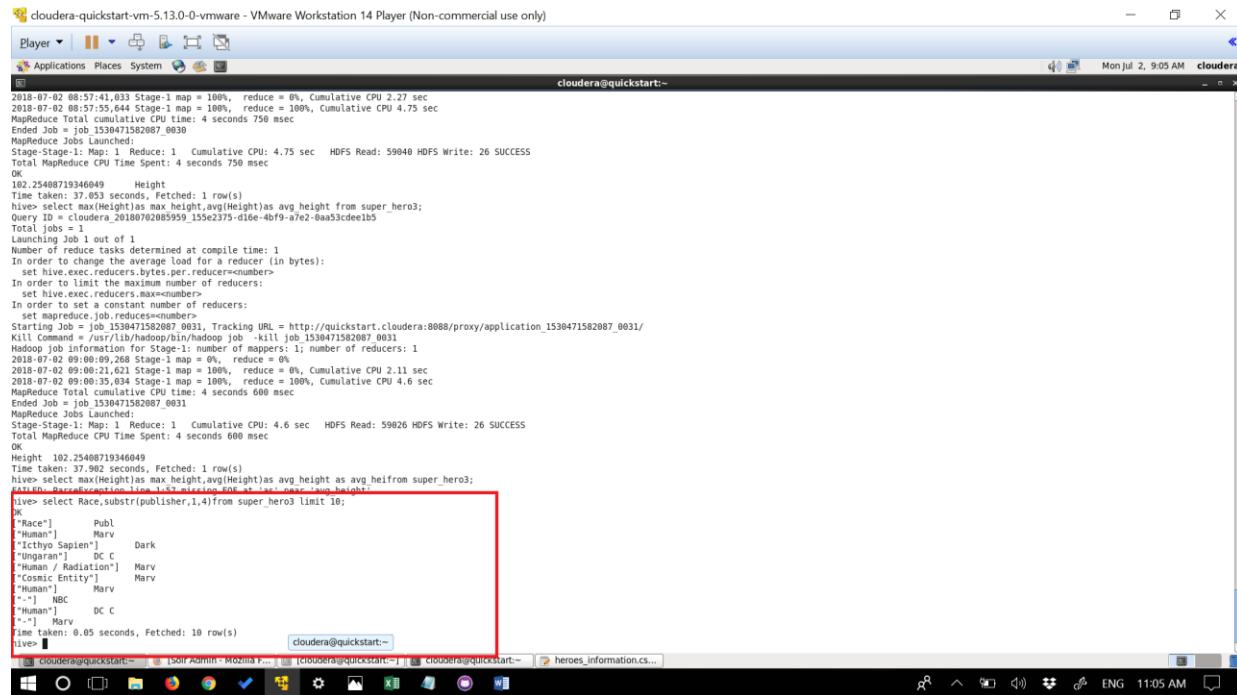
Query2:

Sum of Weights grouped by Publisher

Lab Assignment_4

Query3:

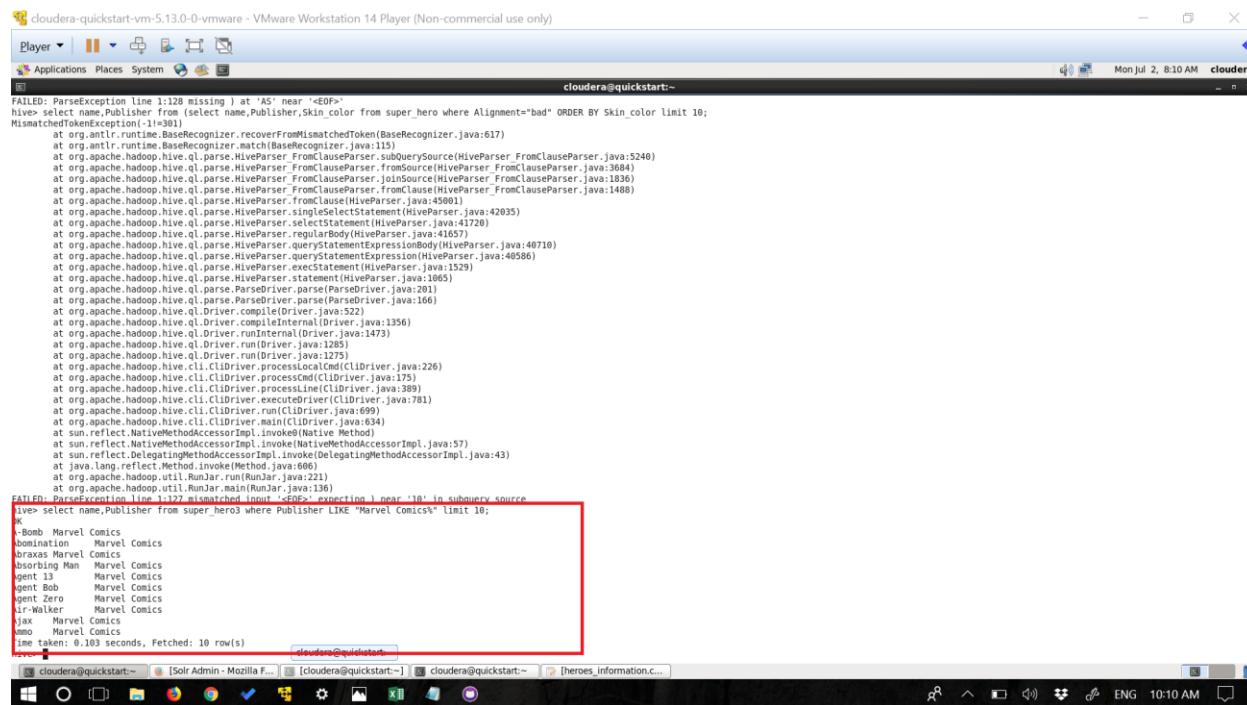
A substring of 1 to 4 characters



```
cloudera-quickstart-vm-5.13.0-0-vmware - VMware Workstation 14 Player (Non-commercial use only)
Player Applications Places System cloudera@quickstart:~ Mon Jul 2, 9:05 AM cloudera
2018-07-02 08:57:41,833 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 2.27 sec
2018-07-02 08:57:55,644 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 4.75 sec
MapReduce Total cumulative CPU time: 4 seconds 750 msec
Ended Job = job_1530471582087_0030
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1 Reduce: 1 Cumulative CPU: 4.75 sec HDFS Read: 59840 HDFS Write: 26 SUCCESS
Total MapReduce Time Spent: 4 seconds 750 msec
OK
102.25480719340649  Height
Time taken: 37.053 seconds, Fetched: 1 row(s)
hive> select maxHeight as max_height, avgHeight as avg_height from super_hero3;
Query ID = cloudera_20180702085959_15e2375-16e4-4bf9-a7e2-0a53cde1e15
Total MapReduce Time Spent: 4 seconds 750 msec
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 number of reducers:
  set hive.exec.reducers.max=<number>
In order to set a constant number of reducers:
  set mapred.job.reducers=<number>
Starting Job = job_1530471582087_0031, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1530471582087_0031/
Kill Command = /usr/bin/hadoop/bin/hadoop job -kill job_1530471582087_0031
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2018-07-02 09:00:09,268 Stage-1 map = 0%, reduce = 0%
2018-07-02 09:00:21,621 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 2.11 sec
2018-07-02 09:00:21,621 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 4.6 sec
MapReduce Total cumulative CPU time: 4 seconds 600 msec
Ended Job = job_1530471582087_0031
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1 Reduce: 1 Cumulative CPU: 4.6 sec HDFS Read: 59826 HDFS Write: 26 SUCCESS
Total MapReduce Time Spent: 4 seconds 600 msec
OK
Height 102.25480719340649
Time taken: 0.052 seconds, Fetched: 1 row(s)
hive> select maxHeight as max_height, avgHeight as avg_height from super_hero3;
+-----+-----+
| max_height | avg_height |
+-----+-----+
| 102.25480719340649 | 102.25480719340649 |
+-----+-----+
Time taken: 0.05 seconds, Fetched: 1 row(s)
hive> select substr(publisher,1,4) from super_hero3 limit 10;
+-----+
| substr(publisher,1,4) |
+-----+
| "Race" | Publ
| "Human" | Marv
| "Ichthy Sapien" | Dark
| "Ungaran" | DC C
| "Human" | Radiation Marv
| "Cosmic Entity" | Marv
| "Human" | Marv
| "-" | NBC
| "Human" | DC C
| "Human" | Marv
+-----+
Time taken: 0.05 seconds, Fetched: 10 row(s)
hive> 
```

Query4:

Pattern Matching



```
cloudera-quickstart-vm-5.13.0-0-vmware - VMware Workstation 14 Player (Non-commercial use only)
Player Applications Places System cloudera@quickstart:~ Mon Jul 2, 8:10 AM cloudera
FAILED: ParseException line 1:128 missing ) at 'AS' near '<EOF>'
hive> select name,Publisher from (select name,Publisher,Skin_color from super_hero where Alignment="bad" ORDER BY Skin_color limit 10;
MismatchedTokenException:11=>301
  at org.antlr.runtime.BaseRecognizer.recoverFromMismatchedToken(BaseRecognizer.java:617)
  at org.antlr.runtime.BaseRecognizer.mismatchedTokenException(BaseRecognizer.java:611)
  at org.apache.hadoop.hive.ql.parse.HiveParser.FromClauseParser_FromClauseParser.java:5240)
  at org.apache.hadoop.hive.ql.parse.HiveParser.FromClauseParser_FromSourceHiveParser_FromClauseParser.java:3684)
  at org.apache.hadoop.hive.ql.parse.HiveParser.FromClauseParser_FromSourceHiveParser_FromClauseParser.java:1836)
  at org.apache.hadoop.hive.ql.parse.HiveParser.FromClauseParser_FromSourceHiveParser_FromClauseHiveParser_FromClauseParser.java:1488)
  at org.apache.hadoop.hive.ql.parse.HiveParser.FromClauseParser_FromSourceHiveParser.java:1481)
  at org.apache.hadoop.hive.ql.parse.HiveParser.singleSelectStatement(HiveParser.java:2035)
  at org.apache.hadoop.hive.ql.parse.HiveParser.selectStatement(HiveParser.java:4172)
  at org.apache.hadoop.hive.ql.parse.HiveParser.regularBody(HiveParser.java:4167)
  at org.apache.hadoop.hive.ql.parse.HiveParser.queryStatementExpression(HiveParser.java:40710)
  at org.apache.hadoop.hive.ql.parse.HiveParser.queryStatement(HiveParser.java:40536)
  at org.apache.hadoop.hive.ql.parse.HiveParser.execStatement(HiveParser.java:1529)
  at org.apache.hadoop.hive.ql.parse.HiveParser.statement(HiveParser.java:1065)
  at org.apache.hadoop.hive.ql.parse.ParserDriver.parse(ParserDriver.java:201)
  at org.apache.hadoop.hive.ql.parse.ParserDriver.run(ParserDriver.java:166)
  at org.apache.hadoop.hive.ql.Driver.compile(Driver.java:522)
  at org.apache.hadoop.hive.ql.Driver.compileInternal(Driver.java:1356)
  at org.apache.hadoop.hive.ql.Driver.runInternal(Driver.java:1473)
  at org.apache.hadoop.hive.ql.Driver.run(Driver.java:1285)
  at org.apache.hadoop.hive.ql.Driver.run(Driver.java:395)
  at org.apache.hadoop.hive.ql.ClDriver.processLocal(ClDriver.java:226)
  at org.apache.hadoop.hive.ql.ClDriver.processCmd(ClDriver.java:175)
  at org.apache.hadoop.hive.ql.ClDriver.processLine(ClDriver.java:389)
  at org.apache.hadoop.hive.ql.ClDriver.run(ClDriver.java:781)
  at org.apache.hadoop.hive.ql.ClDriver.run(ClDriver.java:699)
  at org.apache.hadoop.hive.ql.ClDriver.main(ClDriver.java:634)
  at sun.reflect.NativeMethodAccessorImpl.invoke0(Native Method)
  at sun.reflect.NativeMethodAccessorImpl.invoke(NativeMethodAccessorImpl.java:57)
  at sun.reflect.DelegatingMethodAccessorImpl.invoke(DelegatingMethodAccessorImpl.java:43)
  at java.lang.reflect.Method.invoke(Method.java:606)
  at org.apache.hadoop.util.RunJar.run(RunJar.java:221)
  at org.apache.hadoop.util.RunJar.main(RunJar.java:136)
FAILED: ParseException line 1:127 mismatched input '<EOF>' expecting ) near 'AS' in subquery source
hive> select name,Publisher from super_hero3 where Publisher LIKE 'Marvel Comics%' limit 10;
+-----+-----+
| name | Publisher |
+-----+-----+
| K-Bomb | Marvel Comics
| Bomination | Marvel Comics
| Drax | Marvel Comics
| Absorbing Man | Marvel Comics
| Agent 13 | Marvel Comics
| Agent Bob | Marvel Comics
| Agent Zero | Marvel Comics
| Mr. Wandsworth | Marvel Comics
| Jax | Marvel Comics
| Jimo | Marvel Comics
+-----+-----+
Time taken: 0.103 seconds, Fetched: 10 row(s)
hive> 
```

Lab Assignment_4

Query5:

Counting whose height is greater than 165

```
cloudera-quickstart-vm-5.13.0-0-vmware - VMware Workstation 14 Player (Non-commercial use only)

Player Applications Places System cloudera@quickstar

Vision 1
Vision II 1
Vixen 1
Vulture 1
Walrus 1
War Machine 1
Warbird 1
Warlock 1
Warp 1
Warpath 1
White Queen 1
Winter Soldier 1
Wolfsbane 1
Wonder Girl 1
Wonder Man 1
Wonder Woman 1
Wyatt Wingfoot 1
X-Man 1
Yellow Claw 1
Yellowjacket 1
Yellowjacket II 1
Zoom 1
Zatanna 1
Zoom 1
Time taken: 35.317 seconds. Fetched: 478 row(s).
hive: select count(*) from super_heroes where Height >= 165.8;
Query ID = cloudera_20180702083038_a7c4ae4c-f647-4cdf-883b-2aa24e805627
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_1530471582087_0026, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1530471582087_0026
Kill Command = /usr/lib/hadoop/bin/hadoop job -kill job_1530471582087_0026
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2018-07-02 08:38:12,119 Stage-1 map = 0%,  reduce = 0%
2018-07-02 08:38:25,530 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 2.79 sec
2018-07-02 08:38:41,321 Stage-1 map = 100%,  reduce = 100%, Cumulative CPU 5.56 sec
MapReduce Total cumulative CPU time: 5 seconds 560 msec
Ended Job = job_1530471582087_0026
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1  Reduce: 1  Cumulative CPU: 5.56 sec  HDFS Read: 58741 HDFS Write: 4 SUCCESS
Total MapReduce CPU Time Spent: 5 seconds 560 msec
OK
#84
time taken: 41.461 seconds, Fetched: 1 row(s)
hive: cloudera@quickstart:~
```

Lab Assignment_4

Query6:

Counting different types of eye colors and grouping by eye color

```
cloudera-quickstart-vm-5.13.0-0-vmware - VMware Workstation 14 Player (Non-commercial use only)
Player Applications Places System cloudera@quickstart:~ cloudera@quickstart:~ Mon Jul 2, 8:27 AM cloudera
NULL 1470
Time taken: 44.941 seconds. Fetched: 1 row(s)
hive> select eye_color, count(*) from super_hero3 GROUP BY eye_color;
Query ID: cloudera_20180702082626_106c9c9-9723-410e-8ca0-bab5a7785e5d
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_1530471582087_0024, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1530471582087_0024/
Kill Command = /usr/lib/hadoop/bin/hadoop job -kill job_1530471582087_0024
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2018-07-02 08:26:58,162 Stage-1 map = 0%, reduce = 0%
2018-07-02 08:27:11,834 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 2.22 sec
2018-07-02 08:27:27,998 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 5.06 sec
MapReduce Jobs Launched:
Stage-Stage1: Map: 1 Reduce: 1 Cumulative CPU: 5.06 sec HDFS Read: 58201 HDFS Write: 255 SUCCESS
Total MapReduce CPU Time Spent: 5 seconds 60 msec
OK
+-----+
| eye_color | count |
+-----+
| amber    | 2      |
| black    | 23     |
| blue     | 225    |
| blue / white | 1    |
| brown   | 1      |
| brown   | 126    |
| polo    | 3      |
| green   | 73     |
| green / blue | 1    |
| grey    | 6      |
| hazel   | 6      |
| indigo  | 1      |
| purple  | 4      |
| red     | 46     |
| silver  | 1      |
| violet  | 2      |
| white   | 17     |
| white / red | 1    |
| yellow  | 19     |
| yellow / blue | 2    |
| yellow / red | 1    |
+-----+
Time taken: 47.704 seconds, Fetched: 24 row(s)
hive> [heroes_information.csv (~/Desktop) - edit]
cloudera@quickstart:~ [Sol Admin - Mozilla F... [cloudera@quickstart:~] [cloudera@quickstart:~] [heroes_information.c...]
cloudera@quickstart:~ Mon Jul 2, 8:27 AM cloudera
Windows Taskbar: [cloudera@quickstart:~] [Sol Admin - Mozilla F... [cloudera@quickstart:~] [cloudera@quickstart:~] [heroes_information.c...]
cloudera@quickstart:~ Mon Jul 2, 8:27 AM cloudera

```

Query7:

To get distinct Skin colors in the dataset

```
cloudera-quickstart-vm-5.13.0-0-vmware - VMware Workstation 14 Player (Non-commercial use only)
Player Applications Places System cloudera@quickstart:~ cloudera@quickstart:~ Mon Jul 2, 8:41 AM cloudera
MapReduce Total cumulative CPU Time: 5 seconds 568 msec
Ended Job = job_1530471582087_0026
MapReduce Jobs Launched:
Total MapReduce CPU Time Spent: 5 seconds 568 msec
OK
+-----+
| skin_color |
+-----+
| black     |
| black    |
| blue     |
| blue     |
| blue-white |
| gold     |
| gray     |
| green    |
| gray     |
| orange   |
| orange / white |
| pink    |
| purple   |
| red     |
| red / black |
| silver   |
| white   |
| yellow  |
+-----+
Time taken: 41.461 seconds, Fetched: 1 row(s)
hive> select distinct(skin_color) from super_hero3;
Query ID: cloudera_20180702084040_066c0347-f5f4-4f73-84c8-07d6a003c9f5
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_1530471582087_0027, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1530471582087_0027/
Kill Command = /usr/lib/hadoop/bin/hadoop job -kill job_1530471582087_0027
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2018-07-02 08:40:34,254 Stage-1 map = 0%, reduce = 0%
2018-07-02 08:41:01,288 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 2.08 sec
2018-07-02 08:41:01,288 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 4.56 sec
MapReduce Total cumulative CPU Time: 4 seconds 568 msec
Ended Job = job_1530471582087_0027
MapReduce Jobs Launched:
Stage-Stage1: Map: 1 Reduce: 1 Cumulative CPU: 4.56 sec HDFS Read: 57683 HDFS Write: 126 SUCCESS
Total MapReduce CPU Time Spent: 4 seconds 568 msec
OK
+-----+
| skin_color |
+-----+
| black     |
| blue     |
| blue     |
| blue-white |
| gold     |
| gray     |
| green    |
| gray     |
| orange   |
| orange / white |
| pink    |
| purple   |
| red     |
| red / black |
| silver   |
| white   |
| yellow  |
+-----+
Time taken: 39.122 seconds, Fetched: 18 row(s)
hive> [heroes_information.csv (~/Desktop) - edit]
cloudera@quickstart:~ [Sol Admin - Mozilla F... [cloudera@quickstart:~] [cloudera@quickstart:~] [heroes_information.c...]
cloudera@quickstart:~ Mon Jul 2, 8:41 AM cloudera
Windows Taskbar: [cloudera@quickstart:~] [Sol Admin - Mozilla F... [cloudera@quickstart:~] [cloudera@quickstart:~] [heroes_information.c...]
cloudera@quickstart:~ Mon Jul 2, 8:41 AM cloudera

```

Lab Assignment_4

Query8:

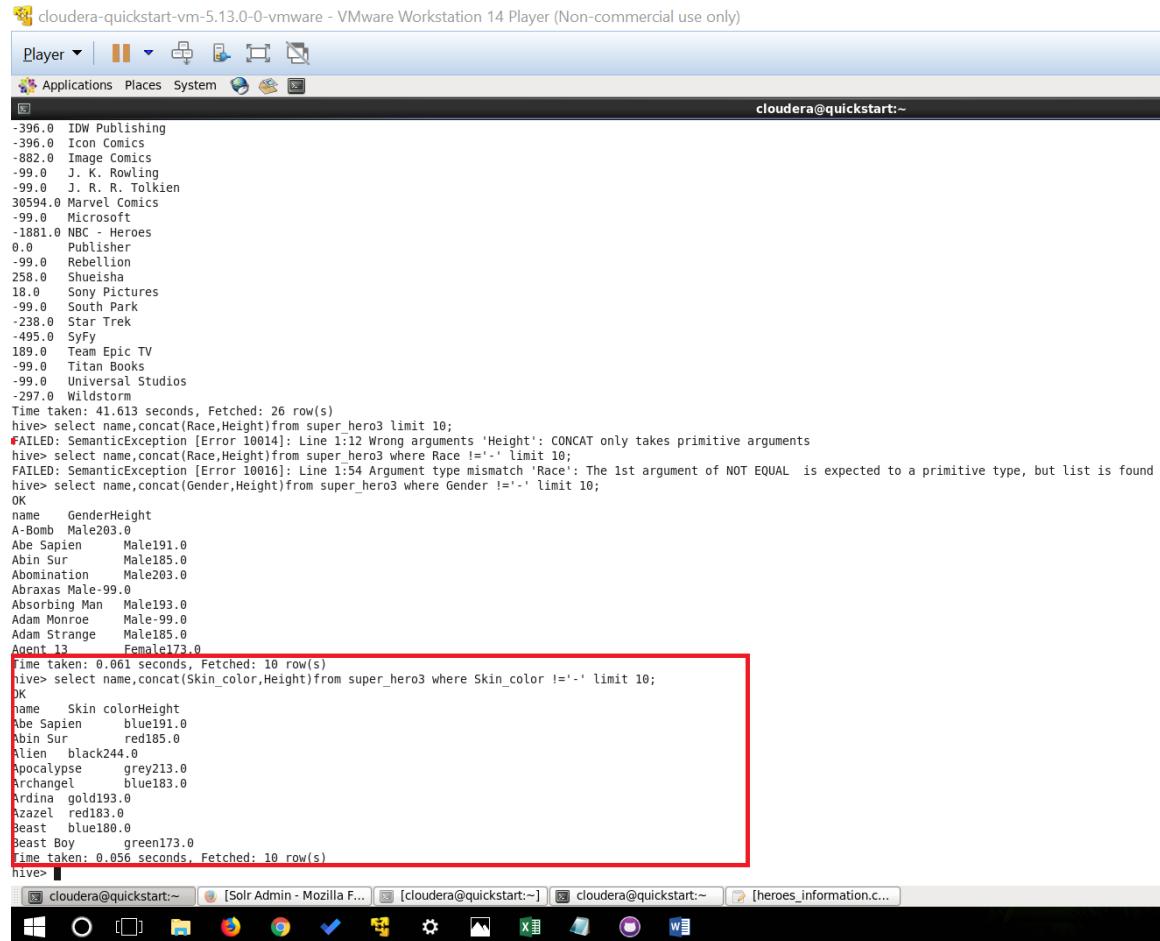
Retrieving average Height in the Dataset

```
cloudera-quickstart-vm-5.13.0-0-vmware - VMware Workstation 14 Player (Non-commercial use only)
Player | Applications Places System Mon
cloudera@quickstart:~$ 
7  Adam Strange  Male  blue  ["Human"]  ["Blond"]  185.0  DC Comics  -  good  88.0
8  Agent 13  Female blue  ["-"]  ["Blond"]  173.0  Marvel Comics  -  good  61.0
Time taken: 0.063 seconds, Fetched: 10 row(s)
hive> create table super_hero3(id INT, name STRING,Gender STRING,Eye_color STRING,Race array<STRING>,Hair_color array<STRING>,Height STRING,Publisher STRING,Skin_color STRING,Alignment STRING,Weight STRING)ROW FORMAT
INATED BY ',' STORED AS TEXTFILE;
OK
Time taken: 0.176 seconds
hive> load data local inpath '/home/cloudera/Desktop/heroes_information.csv' into table super_hero3;
Loading data to table default.super_hero3
Table default.super_hero3 stats: [numFiles=1, totalSize=49197]
OK
Time taken: 0.495 seconds
hive> select *from super_hero3 limit 10;
OK
NULL  name  Gender  Eye color  ["Race"]  ["Hair color"]  Height  Publisher  Skin color  Alignment  Weight
0  A-Bomb  Male  yellow  ["Human"]  ["No Hair"]  203.0  Marvel Comics  -  good  441.0
1  Abe Sapien  Male  blue  ["Ichyo Sapien"]  ["No Hair"]  191.0  Dark Horse Comics  blue  good  65.0
2  Abin Sur  Male  blue  ["Ungaran"]  ["No Hair"]  185.0  DC Comics  red  good  90.0
3  Abomination  Male  green  ["Human / Radiation"]  ["No Hair"]  203.0  Marvel Comics  -  bad  441.0
4  Abraxas  Male  blue  ["Cosmic Entity"]  ["Black"]  -99.0  Marvel Comics  -  bad  -99.0
5  Absorbing Man  Male  blue  ["Human"]  ["No Hair"]  193.0  Marvel Comics  -  bad  122.0
6  Adam Monroe  Male  blue  ["-"]  ["Blond"]  -99.0  NBC - Heroes  good  -99.0
7  Adam Strange  Male  blue  ["Human"]  ["Blond"]  185.0  DC Comics  -  good  88.0
Time taken: 0.077 seconds, Fetched: 10 row(s)
hive> select Publisher,avg(Height) AS avg_height from super_hero3 where Publisher="Marvel Comics" GROUP BY Publisher;
Query ID: cloudera_20180702074343_bdb35761-1791-4f23-bbbb-a140c130c07f
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_1530471582087_0019, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1530471582087_0019/
kill Command = /usr/lib/hadoop/bin/hadoop job -kill job_1530471582087_0019
hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2018-07-02 07:43:55,121 Stage-1 map = 0%,  reduce = 0%
2018-07-02 07:44:12,723 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 3.47 sec
2018-07-02 07:44:28,663 Stage-1 map = 100%,  reduce = 100%, Cumulative CPU 6.3 sec
mapReduce Total cumulative CPU time: 6 seconds 300 msec
ended Job = job_1530471582087_0019
mapReduce Jobs Launched:
Stage-1: Map: 1  Reduce: 1  Cumulative CPU: 6.3 sec  HDFS Read: 59743 HDFS Write: 33 SUCCESS
total MapReduce CPU Time Spent: 6 seconds 300 msec
OK
Marvel Comics  142.75644329896906
Time taken: 56.96 seconds, Fetched: 1 row(s)
hive> 
cloudera@quickstart:~$ [Solr Admin - Mozilla F... [cloudera@quickstart:~] [cloudera@quickstart:~] [cloudera@quickstart:~] heroes_information.csv (~/Desktop) - edit
```

Lab Assignment_4

Query9:

To get Name and Concatenating Skin color and height

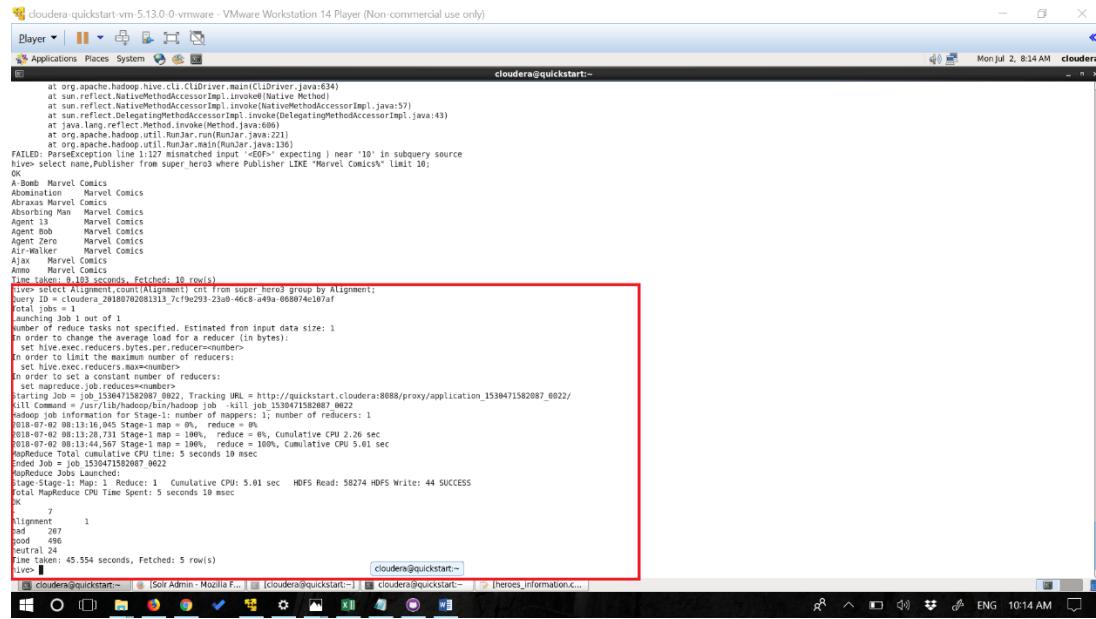


```
cloudera@quickstart-vm-5.13.0-0-vmware - VMware Workstation 14 Player (Non-commercial use only)
Player | ||| | Applications Places System | cloudera@quickstart:~|
cloudera@quickstart:~| cloudera@quickstart:~|
-396.0 IDW Publishing
-396.0 Icon Comics
-882.0 Image Comics
-99.0 J. K. Rowling
-99.0 J. R. R. Tolkien
30594.0 Marvel Comics
-99.0 Microsoft
-1881.0 NBC - Heroes
0.0 Publisher
-99.0 Rebellion
258.0 Shueisha
18.0 Sony Pictures
-99.0 South Park
-238.0 Star Trek
-495.0 SyFy
189.0 Team Epic TV
-99.0 Titan Books
-99.0 Universal Studios
-297.0 Wildstorm
Time taken: 41.613 seconds, Fetched: 26 row(s)
hives> select name,concat(Race,Height)from super_hero3 limit 10;
FAILED: SemanticException [Error 10014]: Line 1:12 Wrong arguments 'Height': CONCAT only takes primitive arguments
hives> select name,concat(Race,Height)from super_hero3 where Race !='-' limit 10;
FAILED: SemanticException [Error 10016]: Line 1:54 Argument type mismatch 'Race': The 1st argument of NOT EQUAL is expected to a primitive type, but list is found
hives> select name,concat(Gender,Height)from super_hero3 where Gender !='-' limit 10;
OK
name  GenderHeight
A-Bomb  Male203.0
Abe Sapien  Male191.0
Abin Sur  Male185.0
Abomination  Male203.0
Abraxas  Male-99.0
Absorbing Man  Male193.0
Adam Monroe  Male-99.0
Adam Strange  Male185.0
Agent 13  Female173.0
Time taken: 0.061 seconds, Fetched: 10 row(s)
hives> select name,concat(Skin_color,Height)from super_hero3 where Skin_color !='-' limit 10;
OK
name  Skin colorheight
Abe Sapien  blue191.0
Abin Sur  red185.0
Alien  black244.0
Apocalypse  grey213.0
Archangel  blue183.0
Ardina  gold193.0
Izazel  red183.0
Beast  blue180.0
Beast Boy  green173.0
Time taken: 0.056 seconds, Fetched: 10 row(s)
hives> | cloudera@quickstart:~ | [Solr Admin - Mozilla F... | [cloudera@quickstart:~] | cloudera@quickstart:~ | [heroes_information.c... |
```

Lab Assignment_4

Query 10:

Counting Different types of alignment in data set.



```
cloudera@quickstart:~$ at org.apache.hadoop.hive.cli.CliDriver.main(CliDriver.java:634)
at sun.reflect.NativeMethodAccessorImpl.invoke0(Native Method)
at sun.reflect.NativeMethodAccessorImpl.invoke(NativeMethodAccessorImpl.java:57)
at sun.reflect.DelegatingMethodAccessorImpl.invoke(DelegatingMethodAccessorImpl.java:43)
at java.lang.reflect.Method.invoke(Method.java:608)
at org.apache.hadoop.util.RunJar.run(RunJar.java:223)
at org.apache.hadoop.util.RunJar.main(RunJar.java:213)
FAILED: ParseException line 1:12 mismatched input '<EOF>' expecting ')' near '10' in subquery source
hive> select name,Publisher from super_heroes where Publisher LIKE 'Marvel Comics' limit 10;
OK
+-----+-----+
| name | Publisher |
+-----+-----+
| Bob | Marvel Comics |
| Abomination | Marvel Comics |
| Abraxas | Marvel Comics |
| Absorbing Man | Marvel Comics |
| Agent 133 | Marvel Comics |
| Agent Bob | Marvel Comics |
| Agent Human | Marvel Comics |
| Al-R'Wahib | Marvel Comics |
| Ajax | Marvel Comics |
| Amra | Marvel Comics |
+-----+-----+
Time taken: 0.100 seconds. Fetched: 10 row(s)
hive> select Alignment,count(Alignment) cnt from super_heroes group by Alignment;
Query ID = cloudera_20180702081313_7cf9e293-23d0-40c8-a49a-688074ec107af
Total MapReduce CPU Time Spent: 0 msec
+-----+-----+
| Alignment | cnt |
+-----+-----+
| 7 | 1 |
| 2d | 267 |
| good | 496 |
| central | 1 |
+-----+-----+
Time taken: 45.554 seconds, Fetched: 5 row(s)
hive> [cloudera@quickstart:~]$
```

Lab Assignment_4

TASK 2:

SOLR USECASE

- a. Create a Solr Collection including our own Field Types

- b. Perform 10 intuitive questions in Dataset .

Dataset used: I have used superheroes as my data set

Queries:

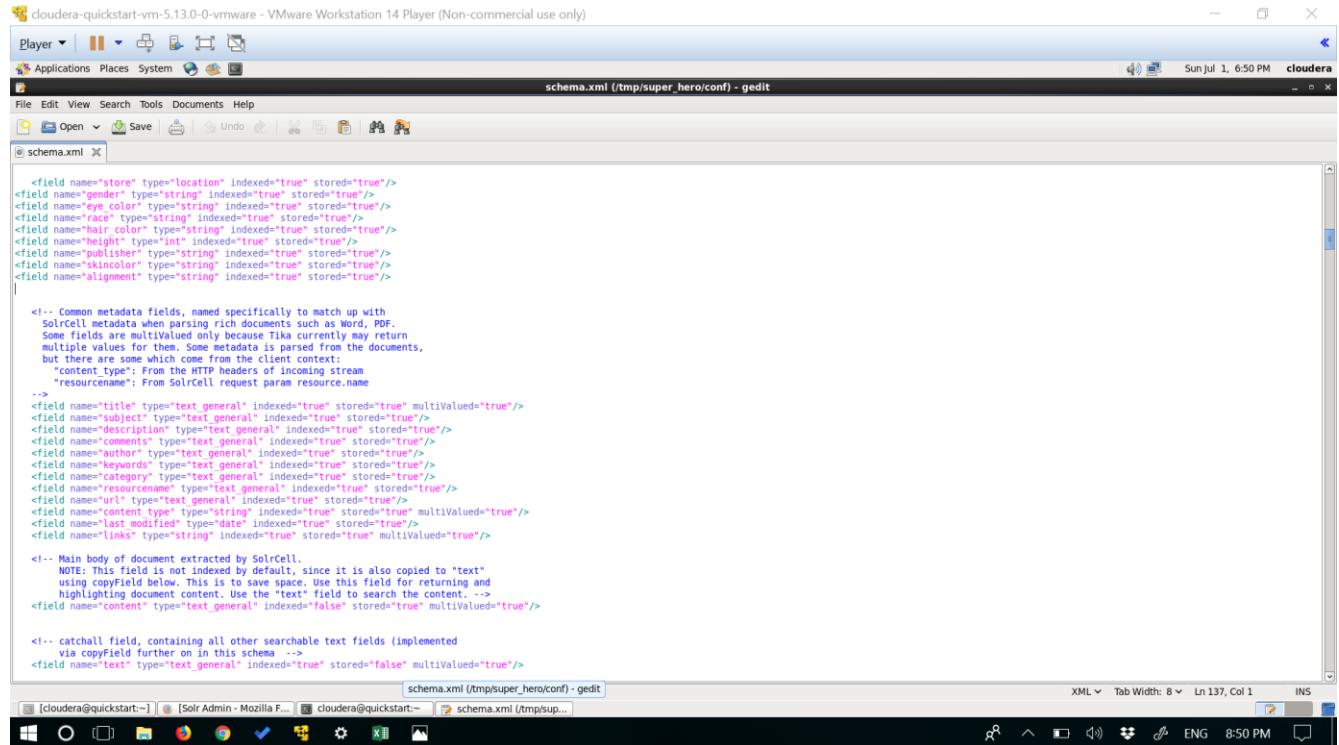
I have performed a various different type of queries. created an instancedir and Solr collection and edited the schema and included various field types.

Generated and created a instance dir

```
cloudera-quickstart-vm-5.13.0-0-vmware - VMware Workstation 14 Player (Non-commercial use only)
Player Applications Places System cloudera@quickstart:~ Mon Jul 2, 9:28 AM cloudera
File Edit View Search Terminal Help
admin-extra.menu-bottom.html schema.xml
admin-extra.menu-top.html scripts.conf
clustering solrconfig.xml
currency.xml solrconfig.xml.secure
elevate.xml spellings.txt
lang stopwords.txt
mapping-FoldToASCII.txt synonyms.txt
mapping-ISOLatinAccent.txt update-script.js
pinyin velocity
rest_managed.json xslt
schema analysis stopwords.english.json
[cloudera@quickstart ~]$ gedit /tmp/super_hero/conf/schema.xml
[cloudera@quickstart ~]$ solrctl instancedir --create super_hero /tmp/super_hero
Uploading configs from /tmp/super_hero/conf to quickstart.cloudera:2181/solr. This may take up to a minute.
[cloudera@quickstart ~]$ solrctl collection --create super_hero
<xml version="1.0" encoding="UTF-8"?><response><lst name="responseHeader">
<int name="status">408</int><int name="QTime">69</int></lst><str name="Operation createCollection caused exception:><org.apache.solr.common.SolrException
re><str> org.apache.solr.client.solrj.impl.HttpSolrServer$RemoteSolrException:
Error CREATING SolrCore 'super_hero shard1 replica1': Unable to create core [su
per_hero shard1 replica1] Caused by: [schema.xml] Duplicate field definition for
'weight' [[weight{type=Float,properties=indexed,stored,omitNorms,omitTermFreqA
rrmPositions}]] and [[weight{type=Int,properties=indexed,stored,omitNorms,omitT
ermFreqAndPositions}]]</str></lst></response>
[cloudera@quickstart ~]$ gedit /tmp/super_hero/conf/schema.xml
[cloudera@quickstart ~]$ solrctl instancedir --create super_hero /tmp/super_hero
Uploading configs from /tmp/super_hero/conf to quickstart.cloudera:2181/solr. This may take up to a minute.
[cloudera@quickstart ~]$ solrctl collection --create super_hero
<xml version="1.0" encoding="UTF-8"?><response><lst name="responseHeader">
<int name="status">408</int><int name="QTime">69</int></lst><str name="Operation createCollection caused exception:><org.apache.solr.common.SolrException
re><str> org.apache.solr.common.SolrException: collection already exists: super_hero</str><str name="msg"> collection already exists: super_hero</str><int name="code">
408</int></lst></response>
[cloudera@quickstart ~]$ solrctl collection --create super_hero
<xml version="1.0" encoding="UTF-8"?><response><lst name="responseHeader">
<int name="status">408</int><int name="QTime">0</int><int name="failure">16187</int></lst><str name="failure"><str> org.apache.solr.client.solrj.impl.HttpSolrServer$RemoteSolrExcept
ion:Error CREATING SolrCore 'super_hero shard1 replica1': Unable to create core [super_hero shard1 replica1] caused by: Could not find configName for collection super_hero found:[film, managedTemplateSecure, predefinedTemplateSecure, films, managedTemplate, film2]</str></lst></response>
[cloudera@quickstart ~]$ solrctl collection --create super_hero
<xml version="1.0" encoding="UTF-8"?><response><lst name="responseHeader">
<int name="status">408</int><int name="QTime">69</int></lst><str name="Operation createCollection caused exception:><org.apache.solr.common.SolrException
re><str> org.apache.solr.common.SolrException: collection already exists: super_hero</str><str name="msg"> collection already exists: super_hero</str><int name="code">
408</int></lst></response>
[cloudera@quickstart ~]$ solrctl instancedir --generate /tmp/super_hero
[cloudera@quickstart ~]$ gedit /tmp/super_hero/conf/schema.xml
[cloudera@quickstart ~]$ solrctl instancedir --create super_hero /tmp/super_hero
Uploading configs from /tmp/super_hero/conf to quickstart.cloudera:2181/solr. This may take up to a minute.
[cloudera@quickstart ~]$ solrctl collection --create super_hero
[cloudera@quickstart ~]$
```

Lab Assignment_4

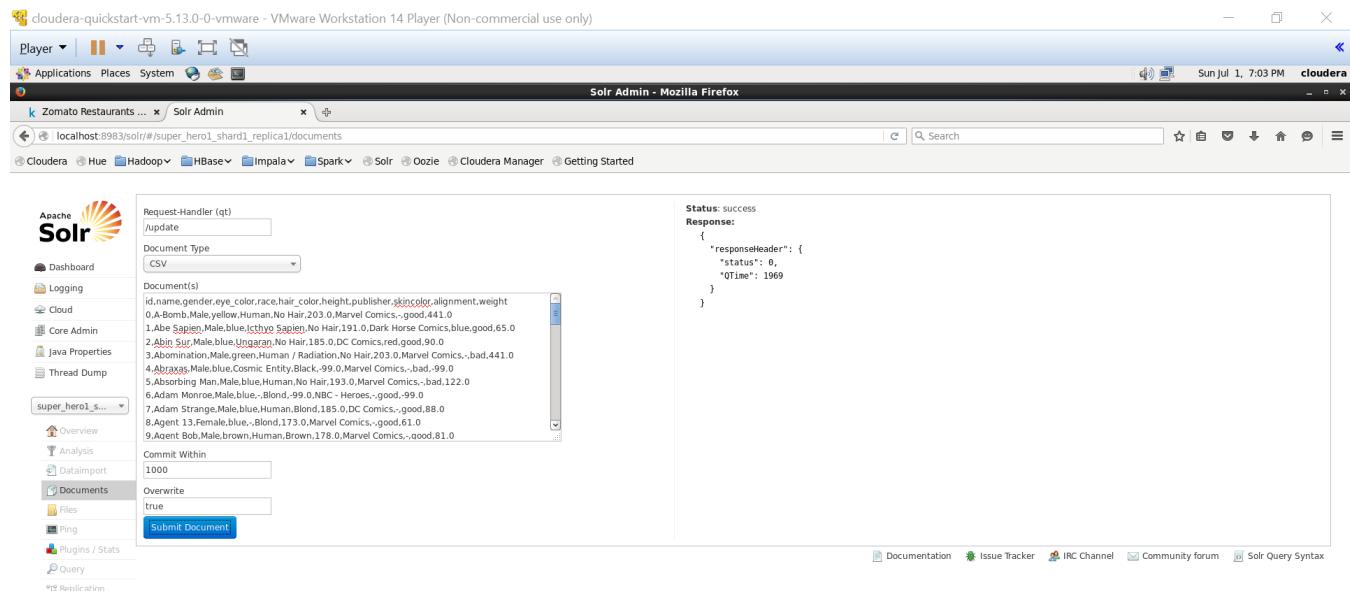
Edited schema and include field types



```
<field name="store" type="location" indexed="true" stored="true"/>
<field name="gender" type="string" indexed="true" stored="true"/>
<field name="eye_color" type="string" indexed="true" stored="true"/>
<field name="race" type="string" indexed="true" stored="true"/>
<field name="name" type="string" indexed="true" stored="true"/>
<field name="height" type="int" indexed="true" stored="true"/>
<field name="publisher" type="string" indexed="true" stored="true"/>
<field name="skincolor" type="string" indexed="true" stored="true"/>
<field name="alignment" type="string" indexed="true" stored="true"/>
|>
<!-- Common metadata fields, named specifically to match up with
     SolrCell metadata when parsing rich documents such as Word, PDF.
     Some fields are multiValue only because Tika currently may return
     multiple values for them. Some metadata is parsed from the documents,
     but there are some which come from the client context:
     -content_type: From the HTTP headers of incoming stream
     -resourceName: From SolrCell request param resource.name
-->
<field name="title" type="text_general" indexed="true" stored="true" multiValued="true"/>
<field name="subject" type="text_general" indexed="true" stored="true"/>
<field name="description" type="text_general" indexed="true" stored="true"/>
<field name="author" type="text_general" indexed="true" stored="true"/>
<field name="keywords" type="text_general" indexed="true" stored="true"/>
<field name="category" type="text_general" indexed="true" stored="true"/>
<field name="resourceName" type="text_general" indexed="true" stored="true"/>
<field name="url" type="text_general" indexed="true" stored="true"/>
<field name="content" type="text_general" indexed="true" stored="true" multiValued="true"/>
<field name="last_modified" type="date" indexed="true" stored="true"/>
<field name="links" type="string" indexed="true" stored="true" multiValued="true"/>
<!-- Main body of document extracted by SolrCell.
     NOTE: This field is not indexed by default, since it is also copied to "text"
     using copyField below. This is to save space. Use this field for returning and
     highlighting document content. Use the "text" field to search the content. -->
<field name="content" type="text_general" indexed="false" stored="true" multiValued="true"/>
<!-- catchall field, containing all other searchable text fields (implemented
     via copyField further on in this schema -->
<field name="text" type="text_general" indexed="true" stored="false" multiValued="true"/>

```

Uploaded document into solr cloud



Request-Handler (qt)
/update
Document Type
CSV
Document(s)
id,name,gender,eye_color,race,hair_color,height,publisher,skincolor,alignment,weight
0.4-Bomb,Male,yellow,Human,No Hair,203.0,Marvel Comics.,good,441.0
1.Abe Sapien,Male,blue,[cthyo Sapien],No Hair,191.0,Dark Horse Comics,blue,good,65.0
2.Abin Sur,Male,blue,Ungaran,,No Hair,185.0,DC Comics,red,good,90.0
3.Abomination,Male,green,Human / Radiation,No Hair,203.0,Marvel Comics.,bad,441.0
4.Abraxas,Male,blue,Entity,Black,-99.0,Marvel Comics.,bad,99.0
5.Absorbing Man,Male,blue,Human,No Hair,193.0,Marvel Comics.,bad,122.0
6.Adam Monroe,Male,blue,,Blond,-99.0,NBC - Heroes.,good,99.0
7.Adam Strange,Male,blue,Human,Blond,185.0,DC Comics.,good,88.0
8.Agent 13,Female,blue,,Blond,173.0,Marvel Comics.,good,61.0
9.Agent Bob,Male,brown,Human,Brown,178.0,Marvel Comics.,good,81.0
Commit Within
1000
Overwrite
true
Submit Document

Status: SUCCESS
Response:
{
 "responseHeader": {
 "status": 0,
 "QTime": 1969
 }
}



Lab Assignment_4

Query1:

Solr Admin - Mozilla Firefox

localhost:8983/solr/#/super_hero1_shard1_replica1/query

Request-Handler (qt)
/select
- common
q
weight:[50.0 TO 90.0]
fq
sort
start, rows
0 10
fl
df
Raw Query Parameters
key1=val1&key2=val2
wt
csv
indent
debugQuery
dismax
edismax
hl
facet
spatial
snarlHack

heroes_information.csv (~/Desktop) - gedit

cloudera@quickstart:~] Solr Admin - Mozilla Fi... [cloudera@quickstart:~] heroes_information.cs...

1604842451092111360,65.0,blue,Icthyo Sapien,Dark Horse Comics,1,191.0,No Hair,Abe Sapien,Male,blue
1604842451095257088,90.0,blue,Ungaran,DC Comics,2,185.0,No Hair,Abin Sur,Male,
good,1604842451102597120,88.0,blue,Human,DC Comics,7,185.0,Blond,Adam Strange,Male,
good,1604842451108888576,61.0,blue,,Marvel Comics,8,173.0,Blond,Agent 13,Female,,
good,1604842451109937152,81.0,brown,Human,Marvel Comics,9,178.0,Brown,Agent Bob,Male,,
bad,160484245120422912,81.0,brown,Cyborg,Marvel Comics,12,193.0,Black,Ajax,Male,,
good,1604842451123568640,99.0,blue,,DC Comics,13,180.0,Blond,Alan Scott,Male,,
good,1604842451138248704,72.0,blue,Human,DC Comics,16,178.0,Black,Alfred Pennyworth,Male,,
good,1604842451162365952,68.0,blue,,Marvel Comics,22,183.0,Blond,Angel,Male,,
good,1604842451167608832,57.0,yellow,Mutant,Marvel Comics,24,165.0,Black,Angel Dust,Female, -

Lab Assignment_4

Query2:

Solr Admin - Mozilla Firefox

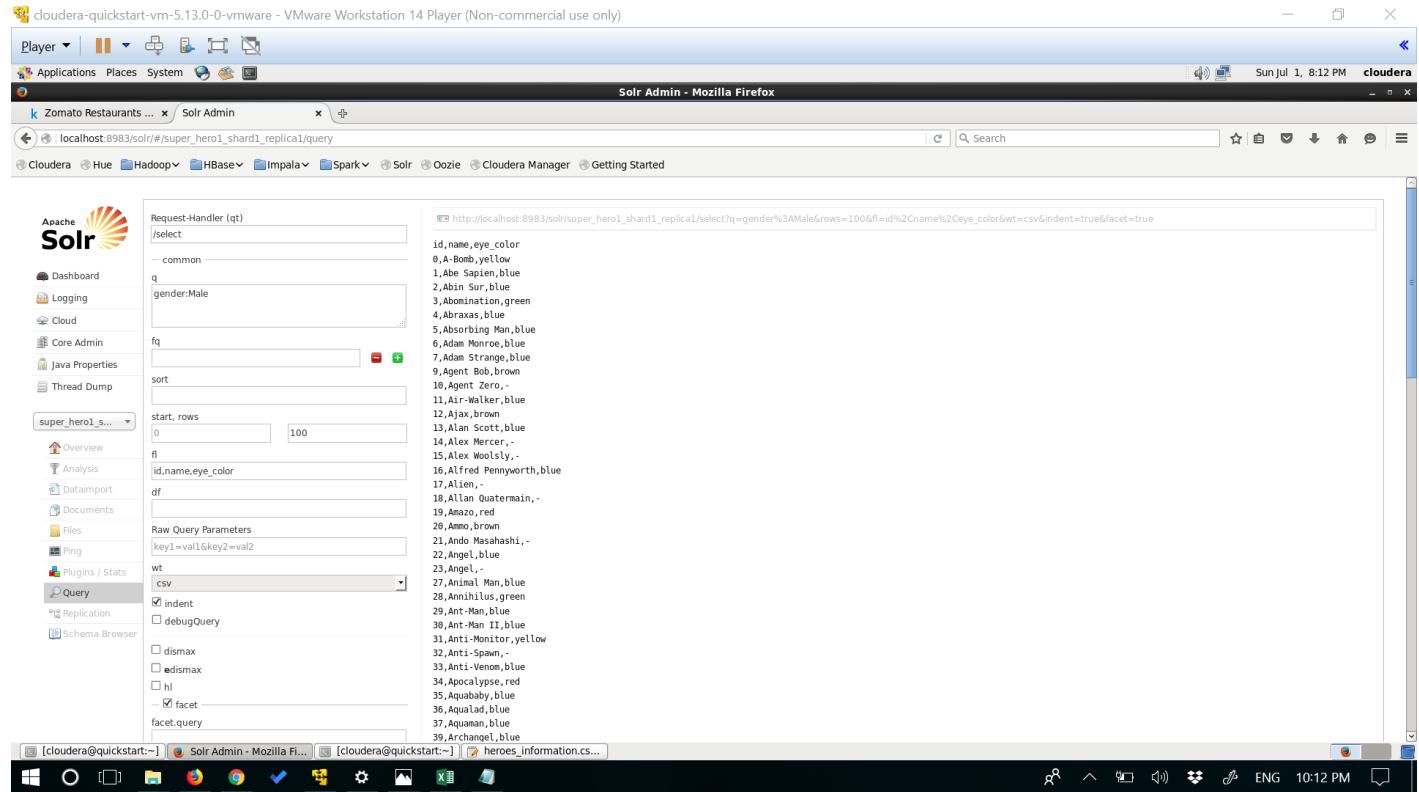
localhost:8983/solr/#/super_hero1_shard1_replica1/query

Cloudera Hue Hadoop HBase Impala Spark Solr Oozie Cloudera Manager Getting Started

Request-Handler (qt)
/select
- common
q
gender:Male
fq
sort
start, rows
0 100
fl
id.name,eye_color
df
Raw Query Parameters
key1=val1&key2=val2
wt
csv
indent
debugQuery
dismax
edismax
hl
facet
facet.query

id,nane,eye_color
0,A-Bomb,yellow
1,Abe Sapien,blue
2,Abin Sur,blue
3,Abomination,green
4,Abraxas,blue
5,Absorbing Man,blue
6,Adam Monroe,blue
7,Adam Strange,blue
9,Agent Bob,brown
10,Agent Zero,-
11,Air-Walker,blue
12,Ajax,brown
13,Alan Scott,blue
14,Alex Mercer,-
15,Alex Woolsty,-
16,Alfred Pennyworth,blue
17,Alien,-
18,Alura Quatermain,-
19,Amazo,red
20,Amo,brown
21,Ando Masahashi,-
22,Angel,blue
23,Angel,-
27,Animal Man,blue
28,Annihilus,green
29,Ant-Man,blue
30,Ant-Man II,blue
31,Ant-Monitor,yellow
32,Ant-Spawn,-
33,Ant-Venom,blue
34,Apocalypse,red
35,Aquabob,blue
36,Aquabob,blue
37,Aquaman,blue
39,Archangel,blue

[cloudera@quickstart-] Solr Admin - Mozilla Fi... [cloudera@quickstart-] heroes_information.cs...



Lab Assignment_4

Query3:

cloudera-quickstart-vm-5.13.0-0-vmware - VMware Workstation 14 Player (Non-commercial use only)

Player | Applications | Places | System | Solr Admin | Cloudera Manager | Getting Started

Solr Admin - Mozilla Firefox

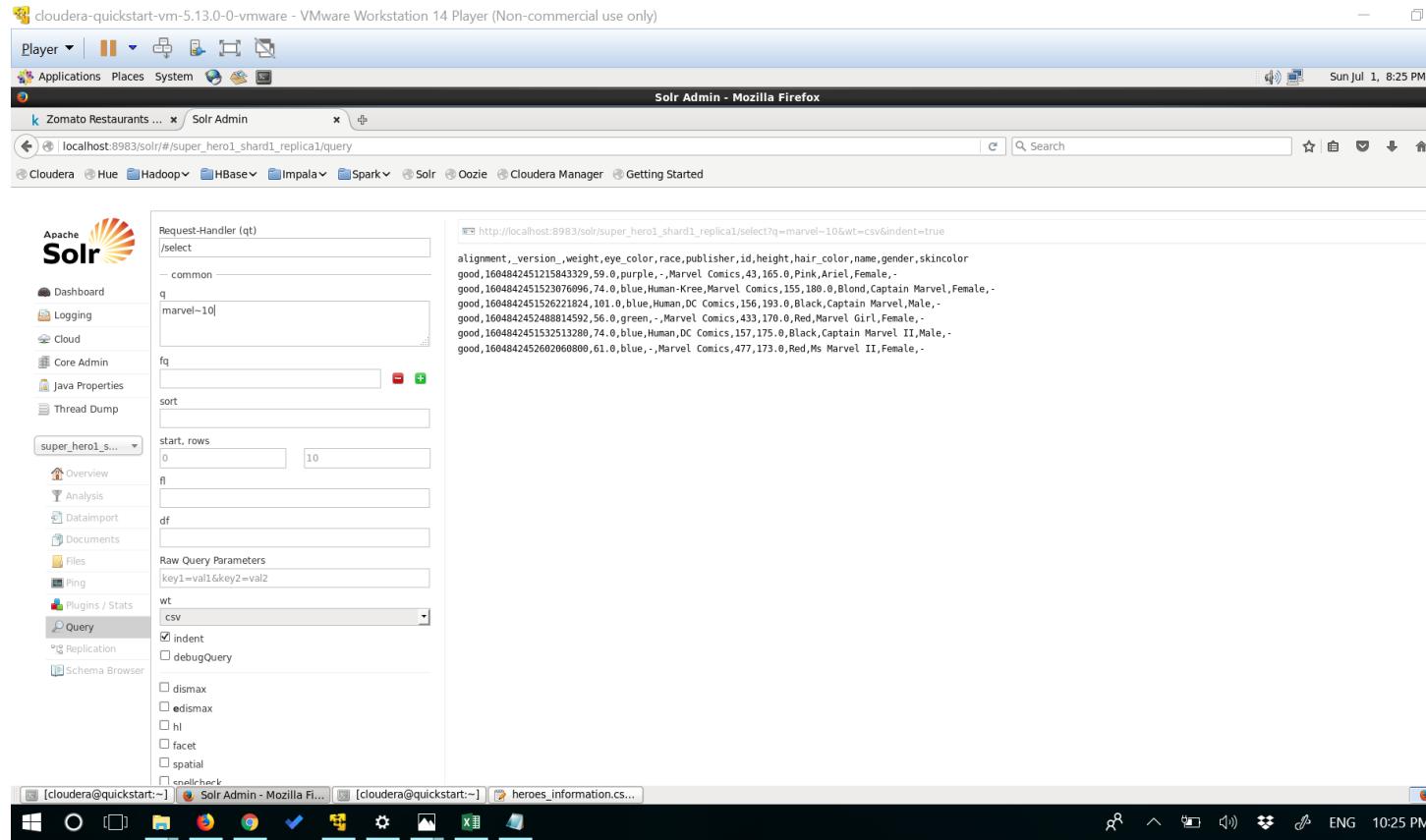
localhost:8983/solr/#/super_hero1_shard1_replica1/query

Request-Handler (qt)
/select
- common
q
marvel~10|
fq
sort
start, rows
0 10
fl
df
Raw Query Parameters
key1=val1&key2=val2
wt
csv
indent
debugQuery
dismax
edismax
hl
facet
spatial
smallCheck

http://localhost:8983/solr/super_hero1_shard1_replica1/select?q=marvel~10&wt=csv&indent=true

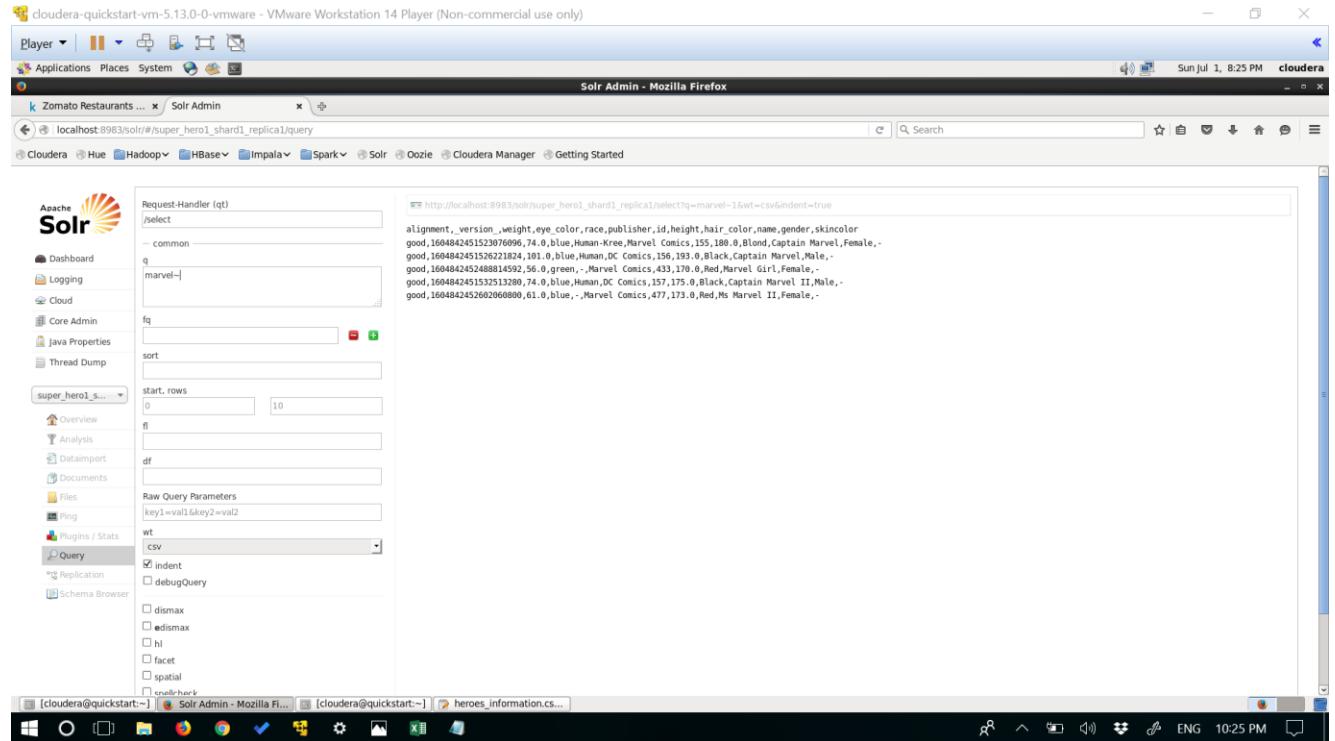
```
alignment,_version_,weight,eye_color,race,publisher,id,height,hair_color,name,gender,skincolor
good,1604842451215843329,59.0,purple,.,Marvel Comics,43,165.0,Pink,Ariel,Female,.
good,1604842451523076696,74.0,blue,Human-Kree,Marvel Comics,155,180.0,Blond,Captain Marvel,Female,.
good,1604842451526221824,101.0,blue,Human,DC Comics,156,193.0,Black,Captain Marvel,Male,.
good,160484245248814592,56.0,green,.,Marvel Comics,433,170.0,Red,Marvel Girl,Female,.
good,1604842451532513280,74.0,blue,Human,DC Comics,157,175.0,Black,Captain Marvel II,Male,.
good,1604842452602060800,61.0,blue,.,Marvel Comics,477,173.0,Red,Ms Marvel II,Female,.
```

[cloudera@quickstart:~] Solr Admin - Mozilla Fi... [cloudera@quickstart:~] heroes_information.cs...

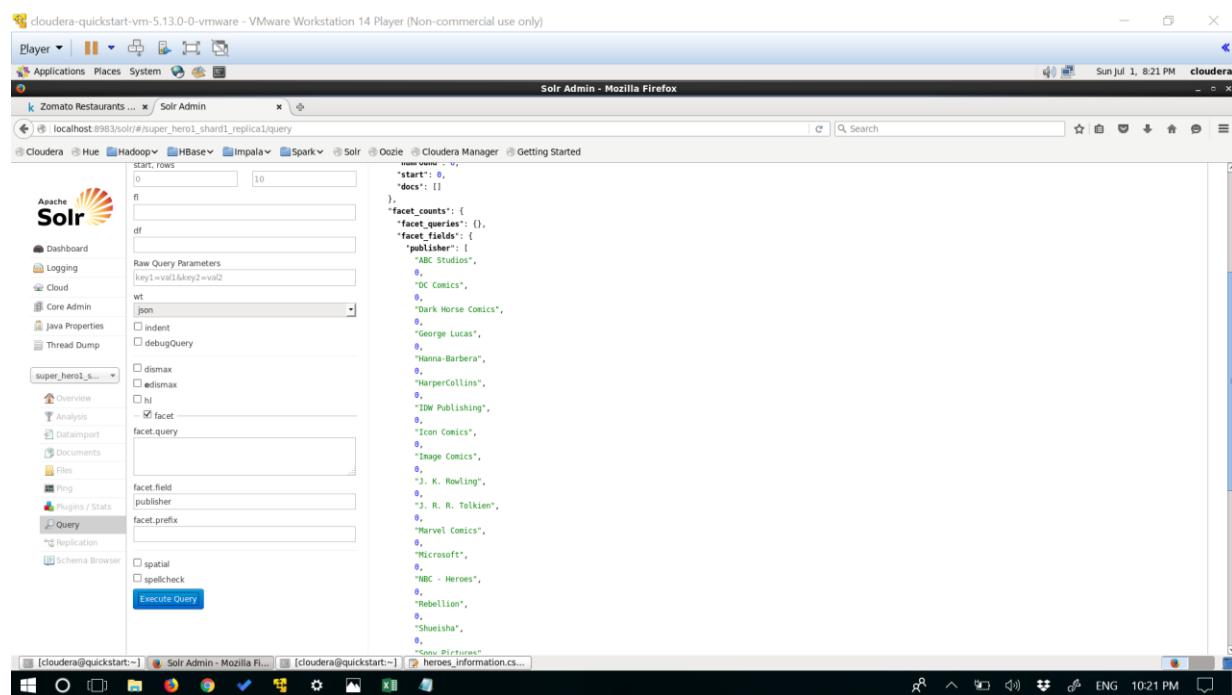


Lab Assignment_4

Query4:



Query5:



Lab Assignment_4

Nested_Query1:

cloudera-quickstart-vm-5.13.0-0-vmware - VMware Workstation 14 Player (Non-commercial use only)

Player | Applications Places System

Solr Admin - Mozilla Firefox

k Zomato Restaurants ... x Solr Admin

localhost:8983/solr/#/super_hero1_shard1_replica1/query

Cloudera Hue Hadoop HBase Impala Solr Oozie Cloudera Manager Getting Started

Request-Handler (qt)
/select
common
q alignment:good AND eye_color:brown
fq
sort
start_rows
0 10
fl
df
Raw Query Parameters
key1=val1&key2=val2
wt
csv
 indent
 debugQuery
 dismax
 edismax
 hl
 facet
 spatial
 cndllr:check

http://localhost:8983/solr/super_hero1_shard1_replica1/select?qt=alignment%3Agood+AND+eye_color%3Abrown&wt=csv&indent=true

alignment,_version_,weight,eye_color,race,publisher.id,height,hair_color,name,gender,skincolor
good,1604842451109937152,81.0,brown,Human,Marvel Comics,9,178.0,Brown,Agent Bob,Male,-
good,1604842451168657468,54.0,brown,-,Marvel Comics,25,163.0,Black,Angel Salvadore,Female,-
good,1604842451215843328,270.0,brown,-,Marvel Comics,42,185.0,Brown,Ares,Male,-
good,1604842451231571968,-99.0,brown,,-,46,-99.0,Black,Astro Boy,Male,-
good,1604842451232620544,101.0,brown,Mutant,Marvel Comics,47,183.0,Red,Atlas,Male,-
good,1604842451250446336,81.0,brown,Human,DC Comics,52,183.0,Auburn,Atom II,Male,-
good,1604842451254640640,72.0,brown,-,DC Comics,54,-99.0,Black,Atom IV,Male,-
good,1604842451263029248,-99.0,brown,Human,DC Comics,57,-99.0,Black,Azrael,Male,-
good,1604842451279369281,54.0,brown,-,Marvel Comics,61,165.0,Black,Bantam,Male,-
good,1604842451300777984,133.0,brown,-,Marvel Comics,71,198.0,Black,Battlestar,Male,-

[cloudera@quickstart:~] Solr Admin - Mozilla Firefox [cloudera@quickstart:~] heroes_information.cs...

Windows Taskbar: cloudera@quickstart:~ [cloudera@quickstart:~] Solr Admin - Mozilla Firefox [cloudera@quickstart:~] heroes_information.cs...

Lab Assignment_4

Nested_Query2:

cloudera-quickstart-vm-5.13.0-0-vmware - VMware Workstation 14 Player (Non-commercial use only)

Player | Applications Places System

Solr Admin - Mozilla Firefox

k Zomato Restaurants ... x Solr Admin x

localhost:8983/solr/#/super_hero1_shard1_replica1/query

Cloudera Hue Hadoop HBase Impala Spark Solr Oozie Cloudera Manager Getting Started

Apache Solr

Request-Handler (qt)

/select

common

q publisher:"Marvel Comics" AND eye_color:brown NOT alignment:good

fq

sort

start, rows

0 10

fl

df

Raw Query Parameters

key1=val1&key2=val2

wt csv

indent

debugQuery

dismax

edismax

hl

facet

spatial

enableQueryParser

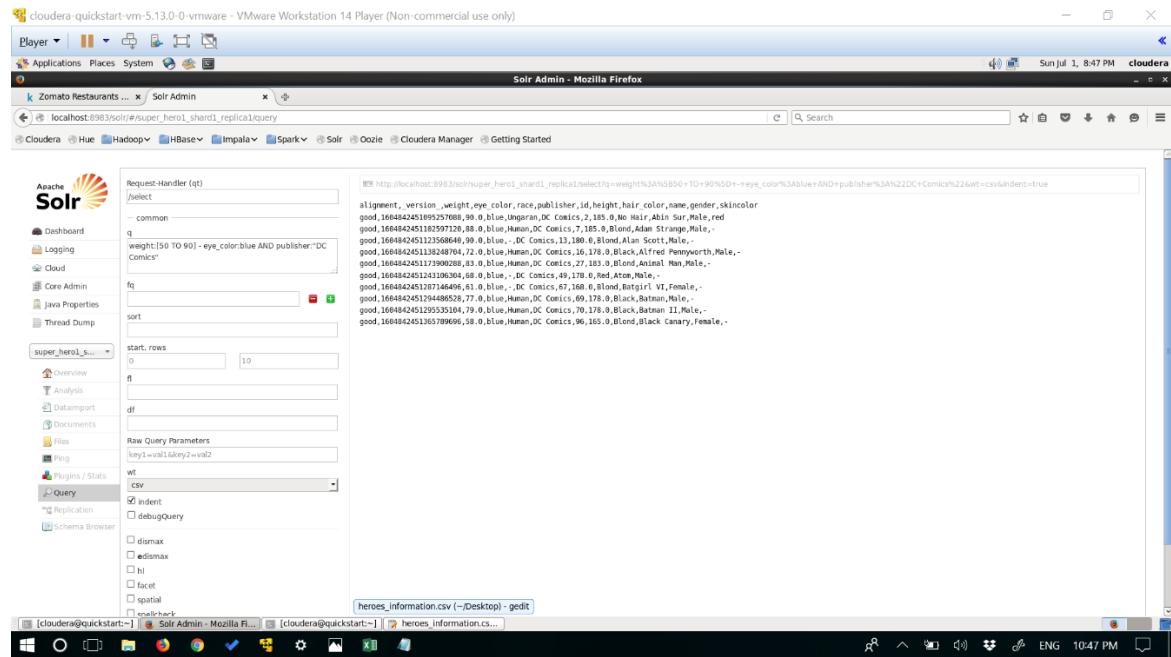
http://localhost:8983/solr/super_hero1_shard1_replica1/select?q=publisher%3A%22Marvel+Comics%22+AND+eye_color%3Abrown+NOT+alignment%3Agood&wt=csv

alignment, version, weight, eye_color, race, publisher_id, height, hair_color, name, gender, skin_color
bad, 1604842451120422912, 90, 0, brown, Cyborg, Marvel Comics, 12, 193, 0, Black, Ajax, Male, -
bad, 1604842451158171648, 101, 0, brown, Human, Marvel Comics, 20, 188, 0, Black, Ammo, Male, -
bad, 1604842451424509952, 77, 0, brown, Human, Marvel Comics, 117, 175, 0, Brown, Blizzard II, Male, -
bad, 1604842451424509953, 230, 0, brown, -, Marvel Comics, 118, 178, 0, Brown, Blob, Male, -
bad, 1604842451563970560, 81, 0, brown, -, Marvel Comics, 170, 180, 0, Black, Changeling, Male, -
bad, 1604842451613253632, 99, 0, brown, Human, Marvel Comics, 187, 183, 0, Black, Cottonmouth, Male, -
neutral, 1604842451679313920, 95, 0, brown, Mutant, Marvel Comics, 212, 188, 0, No Hair, Deadpool, Male, -
bad, 1604842451696091136, 90, 0, brown, Human, Marvel Comics, 218, 193, 0, Black, Diamondback, Male, -
bad, 1604842451705528320, 187, 0, brown, Human, Marvel Comics, 221, 201, 0, Brown, Doctor Doom, Male, -
bad, 1604842451705528321, 132, 0, brown, -, Marvel Comics, 222, 201, 0, Brown, Doctor Doom II, Male, -

[cloudera@quickstart:~] Solr Admin - Mozilla Fi... [cloudera@quickstart:~] heroes_information.cs...

Lab Assignment_4

Nested_Query3:



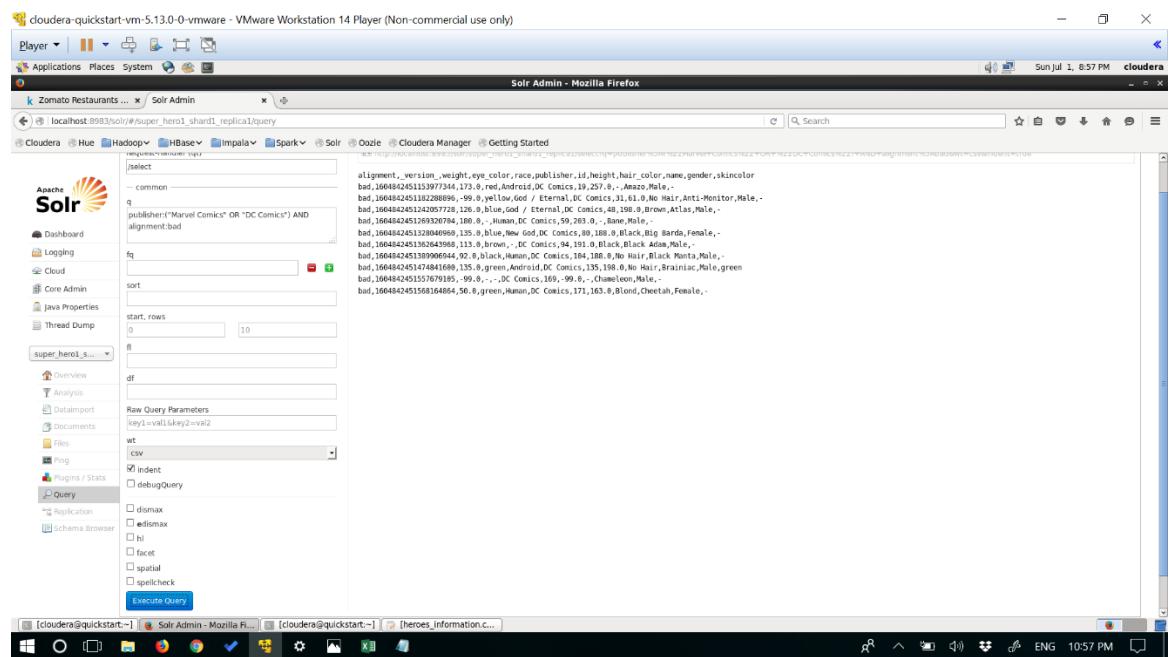
The screenshot shows the Solr Admin interface in a Mozilla Firefox browser. The URL is `localhost:8983/solr/super_hero1_shard1_replica1/query`. The query is:

```
q=weight:[50 TO 90] -eye_color:blue AND publisher:"DC Comics"
```

The results list several entries, including:

- good,16684245109525768,98.0,blue,Human,DC Comics,2,185.0,No hair,Akin,Sur,Male,red
- good,166842451182597126,98.0,blue,Human,DC Comics,7,185.0,Blond,Adam Strange,Male,-
- good,16684245123568649,98.0,blue,Human,DC Comics,13,180.0,Blond,Alan Scott,Male,-
- good,16684245138248704,72.0,blue,Human,DC Comics,16,178.0,Black,Alfred Pennyworth,Male,-
- good,166842451421386844,98.0,blue,Human,DC Comics,45,178.0,Blond,Barry Allen,Male,-
- good,166842451421386844,98.0,blue,Human,DC Comics,45,178.0,Red,Atom,Male,-
- good,166842451387146466,81.0,blue,Human,DC Comics,67,186.0,Blond,Batgirl vt,Female,-
- good,16684245179488528,77.0,blue,Human,DC Comics,69,178.0,Black,Batman,Male,-
- good,166842451295535104,79.0,blue,Human,DC Comics,70,178.0,Black,Batman II,Male,-
- good,166842451365789696,98.0,blue,Human,DC Comics,96,165.0,Blond,Black Canary,Female,-

Nested_query4:



The screenshot shows the Solr Admin interface in a Mozilla Firefox browser. The URL is `localhost:8983/solr/super_hero1_shard1_replica1/query`. The query is:

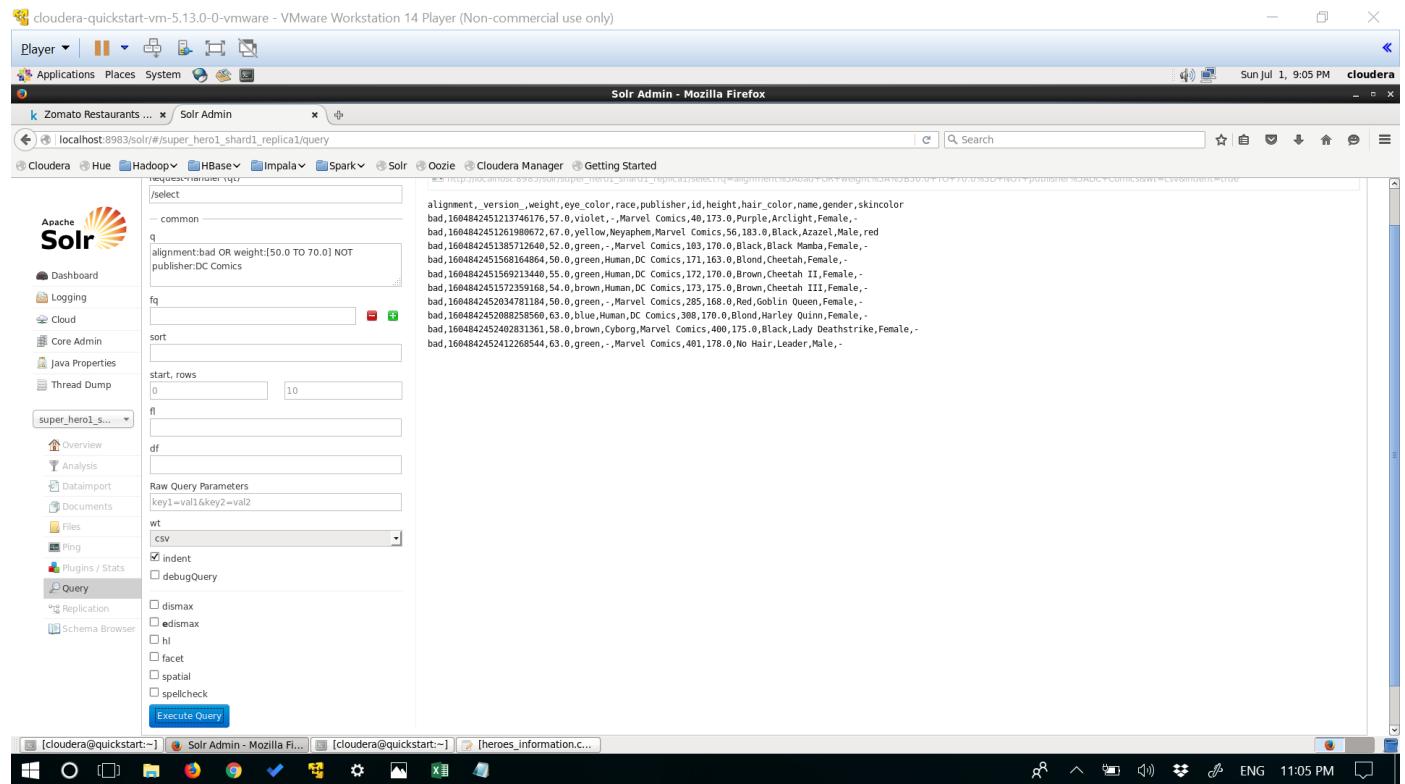
```
q=publisher:"Marvel Comics" OR "DC Comics" AND alignment:bad
```

The results list several entries, including:

- alignment,_version_,weight,eye_color,race,publisher.id,height,hair_color,name,gender,skincolor,bad,16684245115397344,173.0,red,Android,DC Comics,19,257.0,,Amazo,Male,-
- bad,16684245115397344,198.0,yellow,Android,DC Comics,31,181.0,No Hair,Anti-Monitor,Male,-
- bad,16684245132077224,100.0,blue,Human,DC Comics,10,170.0,Blond,Atlas,Male,-
- bad,1668424513263327674,100.0,blue,Human,DC Comics,59,283.0,,None,Male,-
- bad,166842451327049969,135.0,blue,New God,DC Comics,89,188.0,Black,Big Barda,Female,-
- bad,1668424513678439588,135.0,brown,,DC Comics,94,191.0,Black,Black Adam,Male,-
- bad,16684245139996644,92.0,black,Human,DC Comics,184,188.0,No Hair,Black Manta,Male,-
- bad,16684245147484169,135.0,green,Android,DC Comics,135,188.0,No Hair,Brainiac,Male,green
- bad,16684245155707939,99.0,-,DC Comics,189,199.0,,Chameleon,Male,-
- bad,16684245156810464,90.0,green,Human,DC Comics,171,185.0,Blond,Cheetah,Female,-

Lab Assignment_4

Nested_query5:



The screenshot shows the Solr Admin interface in a Mozilla Firefox browser. The URL is `localhost:8983/solr/#/super_hero1_shard1_replica1/query`. The query entered is `alignment:bad OR weight:[50.0 TO 70.0] NOT publisher:DC Comics`. The results table displays the following data:

| id | name | publisher | height | weight | eye color | race | hair color | gender | skin color | alignment |
|--------------------------|------|-----------|---------------|---------------|-----------|---------|--------------|------------------|------------|-----------|
| bad_1604842451213746176 | 57.0 | violet | Marvel Comics | 49 | 172.0 | Purple | Arclight | Female | red | bad |
| bad_1604842451261988672 | 67.0 | yellow | Neophom | Marvel Comics | 56 | 183.0 | Black | Azzel | Male | red |
| bad_1604842451385712648 | 52.0 | green | Marvel Comics | 183 | 170.0 | Black | Black Mamba | Female | black | bad |
| bad_1604842451568164864 | 50.0 | green | Human | DC Comics | 171 | 163.0 | Blond | Cheetah | Female | black |
| bad_1604842451599213448 | 55.0 | green | Human | DC Comics | 172 | 170.0 | Brown | Cheetah II | Female | black |
| bad_1604842451572359168 | 54.0 | brown | Human | DC Comics | 173 | 175.0 | Brown | Cheetah III | Female | black |
| bad_1604842452034781194 | 50.0 | green | Marvel Comics | 285 | 160.0 | Red | Goblin Queen | Female | black | bad |
| bad_1604842452088258560 | 63.0 | blue | Human | DC Comics | 308 | 170.0 | Blond | Harley Quinn | Female | black |
| bad_16048424524082831361 | 58.0 | brown | Cyborg | Marvel Comics | 400 | 175.0 | Black | Lady Deathstrike | Female | black |
| bad_1604842452412266544 | 63.0 | green | Marvel Comics | 401 | 170.0 | No Hair | Leader | Male | black | bad |

References:

1. https://www.tutorialspoint.com/apache_solr/apache_solr_querying_data.htm
2. <https://community.modeanalytics.com/sql/tutorial/sql-subqueries/>
3. https://lucene.apache.org/solr/guide/6_6/the-standard-query-parser.html
4. <https://acadgild.com/blog/working-with-hive-complex-data-types>
5. <https://www.kaggle.com/claudiodavi/superhero-set/data>