

CS5590 BigData Programming - Lab Assignment 2

Team Id: 3

Member 1: Raju Nekadi
Class Id: 7

Member 2: Sushma Manne
Class Id: 3

Raju's GitHub Link:

https://github.com/rnekadi/CSEE5590_BIGDATA_PROGAMMING_Fall2018/tree/master/Lab2

Sushma's GitHub Link :

https://github.com/sushmamanne/CSEE5590_BIGDATA_PROGRAMMING_FALL2018

Video Link Solr: <https://youtu.be/X9Exh39RCxE>

Video Link Hive Cassandra: <https://youtu.be/wXLF-XWhBbU>

Introduction:

In Lab2, we be using various Big Data frameworks tools like hive, solr and Cassandra along with various datasets.

Objective:

The main aim of this Lab is to any of the given sample dataset and perform various queries using Hive, solr and Cassandra tools.

Approaches:

For this Lab we have uses Super heroes dataset for Hive framework and perform create insert and select queries to get meaningful insight.

For Cassandra we have picked the Coursera use case model and performed various create, insert and select queries to get meaningful data.

For Solr we have used dataset and perform various queries to get answers.

Datasets:

Workflow:

Hive Queries:

We used the super hero dataset for Hive use case.

1. Create the heroes table.

```
hive> CREATE TABLE IF NOT EXISTS heroes (hero_id int,hero_name String,hero_gender String,hero_eye_color String,hero_race String,hero_hair_color String,hero_height Float,hero_publisher String,hero_skin_color String,hero_alignment String,hero_weight Float)
> ROW FORMAT DELIMITED
> FIELDS TERMINATED BY ','
> LINES TERMINATED BY '\n';
OK
Time taken: 0.303 seconds
```

2. Load the heroes table using the load statement

```
hive> LOAD DATA LOCAL INPATH '/home/cloudera/Downloads/superhero-set/heroes_information.csv' INTO TABLE heroes;
Loading data to table users.heroes
Table users.heroes stats: [numFiles=1, totalSize=49195]
OK
Time taken: 1.587 seconds
hive>
```

3. Created the Heroes Super powers Table.

```
hive> CREATE TABLE IF NOT EXISTS heroes_super_powers (
> hero_names String,Agility INT,AcceleratedHealing INT,LanternPowerRing INT,DimensionalAwareness INT,ColdResistance INT,Durability INT,Stealth INT,EnergyAbsorption INT,Flight INT,DangerSense INT,UnderwaterBreathing INT,Marksmanship INT,WeaponsMaster INT,PowerAugmentation INT,AnimalAttributes INT,Longevity INT,Intelligence INT,SuperStrength INT,Cryokinesis INT,Telepathy INT,EnergyArmor INT,EnergyBlasts INT,Duplication INT,SizeChanging INT,DensityControl INT,Stamina INT,AstralTravel INT,AudioControl INT,Dexterity INT,Omnitrix INT,SuperSpeed INT,Possession INT,AnimalOrientation INT,WeaponBasedPowers INT,Electrokinesis INT,DarkForceManipulation INT,DeathTouch INT,Teleportation INT,EnhancedSenses INT,Telekinesis INT,EnergyBeams INT,Magic INT,Hyperkinesis INT,Jump INT,Clairvoyance INT,DimensionalTravel INT,PowerSense INT,Shapeshifting INT,PeakHumanCondition INT,Immortality INT,Camouflage INT,ElementControl INT,Phasing INT,AstralProjection INT,ElectricalTransport INT,FireControl INT,Projection INT,Summoning INT,EnhancedMemory INT,Reflexes INT,Invulnerability INT,EnergyConstructs INT,ForceFields INT,SelfSustenance INT,AntiGravity INT,Empathy INT,PowerNullifier INT,RadiationControl INT,PsionicPowers INT,Elasticity INT,SubstanceSecretion INT,ElementalTransmogrification INT,TechnopathyCyberpath INT,PhotographicReflexes INT,SeismicPower INT,Animation INT,Precognition INT,MindControl INT,FireResistance INT,PowerAbsorption INT,EnhancedHearing INT,NovaForce INT,Insanity INT,Hypnokinesis INT,AnimalControl INT,NaturalArmor INT,Intangibility INT,EnhancedSight INT,MolecularManipulation INT,HeatGeneration INT,Adaptation INT,Gliding INT,PowerSuit INT,MindBlast INT,ProbabilityManipulation INT,GravityControl INT,Regeneration INT,LightControl INT,Echolocation INT,Levitation INT,ToxinDiseaseControl INT,Banish INT,EnergyManipulation INT,HeatResistance INT,NaturalWeapons INT,TimeTravel INT,EnhancedSmell INT,Illusions INT,Thirstokinesis INT,HairManipulation INT,Illumination INT,Omnipotent INT,Cloaking INT,ChangingArmor INT,PowerCosmic INT,Biokinesis INT,WaterControl INT,RadiationImmunity INT,VisionTelescopic INT,ToxinDiseaseResistance INT,SpatialAwareness INT,EnergyResistance INT,TelepathyResistance INT,MolecularCombustion INT,Omnilingualism INT,PortalCreation INT,Magnetism INT,MindControlResistance INT,PlanetaryControl INT,Sonar INT,SonicScream INT,TimeManipulation INT,EnhancedTouch INT,MagicResistance INT,Invisibility INT,SubMariner INT,RadiationAbsorption INT,Intuitiveaptitude INT,VisionMicroscopic INT,Melting INT,WindControl INT,SuperBreath INT,WallCrawling INT,VisionNight INT,VisionInfrared INT,GrimReaping INT,MatterAbsorption INT,TheForce INT,Resurrection INT,Terrakinesis INT,VisionHeat INT,Vitakinesis INT,RadarSense INT,GuardianPowerRing INT,WeatherControl INT,VisionXRay INT,VisionThermal INT,WebCreation INT,RealityWarping INT,OdinForce INT,SymbioteCostume INT,SpeedForce INT,PhoenixForce INT,MolecularDissipation INT,VisionCryo INT,Omnipresent INT,Omniscient INT
> )
> ROW FORMAT DELIMITED FIELDS TERMINATED BY ','
> LINES TERMINATED BY '\n';
OK
```

4. Load heroes Power table.

```
Time taken: 0.719 seconds
hive> LOAD DATA LOCAL INPATH '/home/cloudera/Downloads/superhero-set/super_hero_powers.csv' into table heroes_super_powers;
Loading data to table default.heroes_super_powers
Table default.heroes_super_powers stats: [numFiles=1, totalSize=232623]
OK
Time taken: 2.888 seconds
hive>
```

5. Create the super_heroes_total power table from hero's power table.

```
CREATE TABLE IF NOT EXISTS heroes_super_powers_total (hero_names String
+super_power_total INT)
ROW FORMAT DELIMITED FIELDS TERMINATED BY ' +'
LINES TERMINATED BY '\n';
```

```
INSERT INTO heroes_super_power_total
Select hero_names ,(Agility +AcceleratedHealing +LanternPowerRing
+DimensionalAwareness +ColdResistance +Durability +Stealth +EnergyAbsorption
+Flight +
DangerSense +UnderwaterBreathing +Marksmanship +WeaponsMaster
+PowerAugmentation +AnimalAttributes +Longevity +INtelligence +SuperStrength
+Cryokinesis +Telepathy +
EnergyArmor +EnergyBlasts +Duplication +SizeChanging +DensityControl +Stamina
+AstralTravel +AudioControl +Dexterity +Omnitrix +SuperSpeed +Possession
+AnimalOrientedPowers +
WeaponBasedPowers +Electrokinesis +DarkforceManipulation +DeathTouch
+Teleportation +EnhancedSenses +Telekinesis +EnergyBeams +Magic +Hyperkinesis
+Jump +Clairvoyance +DimensionalTravel +
PowerSense +Shapeshifting +PeakHumanCondition +Immortality +Camouflage
+ElementControl +Phasing +AstralProjection +ElectricalTransport +FireControl
+Projection +Summoning +EnhancedMemory +
Reflexes +Invulnerability +EnergyConstructs +ForceFields +SelfSustenance
+AntiGravity +Empathy +PowerNullifier +RadiationControl +PsionicPowers +Elasticity
+SubstanceSecretion +ElementalTransmogrification +
TechnopathCyberpath +PhotographicReflexes +SeismicPower +Animation
+Precognition +MindControl +FireResistance +PowerAbsorption +EnhancedHearing
+NovaForce +Insanity +Hypnokinesis +AnimalControl +NaturalArmor +
INTangibility +EnhancedSight +MolecularManipulation +HeatGeneration +Adaptation
+Gliding +PowerSuit +MindBlast +ProbabilityManipulation +GravityControl
+Regeneration +LightControl +Echolocation +Levitation +
ToxinDiseaseControl +Banish +EnergyManipulation +HeatResistance +NaturalWeapons
+TimeTravel +EnhancedSmell +Illusions +Thirstokinesis +HairManipulation
+Illumination +Omnipotent +Cloaking +ChangingArmor +PowerCosmic +
Biokinesis +WaterControl +RadiationImmunity +VisionTelescopic
+ToxinDiseaseResistance +SpatialAwareness +EnergyResistance +TelepathyResistance
+MolecularCombustion +Omnilingualism +PortalCreation +Magnetism
+MindControlResistance +
PlantControl +Sonar +SonicScream +TimeManipulation +EnhancedTouch
+MagicResistance +Invisibility +SubMariner +RadiationAbsorption + INTuitiveaptitude
+VisionMicroscopic +Melting +WindControl +SuperBreath +Wallcrawling +VisionNight
+
VisionInfrared +GrimReaping +MatterAbsorption +TheForce +Resurrection
+Terrakinesis +VisionHeat +Vitakinesis +RadarSense +QwardianPowerRing
```

+WeatherControl +VisionXRay +VisionThermal +WebCreation +RealityWarping
+OdinForce +
SymbioteCostume +SpeedForce +PhoenixForce +MolecularDissipation +VisionCryo
+Omnipresent +Omniscient) from heroes_super_powers;

6. Select statement from heroes and heroes super powers total table.

**1.select hero_names,super_power_total from heroes_super_power_total SORT BY
super_power_total ASC limit 5;**

```
Total MapReduce CPU Time Spent: 5 seconds 100 msec
OK
hero_names      NULL
Allan Quatermain      1
Abin Sur           1
Alex Woolsly        1
Riddler 1
Time taken: 84.588 seconds, Fetched: 5 row(s)
hive>
```

**2.select hero_names,super_power_total from heroes_super_power_total SORT BY
super_power_total desc limit 5;**

```
Stage-Stage-2: Map: 1 Reduce: 1 Cumulative CPU: 2
Total MapReduce CPU Time Spent: 5 seconds 400 msec
OK
Spectre 49
Amazo 44
Living Tribunal 35
Martian Manhunter 35
Man of Miracles 34
Time taken: 90.064 seconds, Fetched: 5 row(s)
hive>
```

**3.select count(hero_name) CountHero,hero_publisher from heroes where
hero_publisher group by(hero_publisher);**

Stage-Stage-1: map: 1 Reduce: 1 Cumulative
Total MapReduce CPU Time Spent: 2 seconds 760

OK

15

4 ABC Studios

215 DC Comics

18 Dark Horse Comics

14 George Lucas

1 Hanna-Barbera

6 HarperCollins

4 IDW Publishing

4 Icon Comics

14 Image Comics

1 J. K. Rowling

1 J. R. R. Tolkien

388 Marvel Comics

1 Microsoft

19 NBC - Heroes

1 Publisher

1 Rebellion

4 Shueisha

2 Sony Pictures

1 South Park

6 Star Trek

5 SyFy

5 Team Epic TV

1 Titan Books

1 Universal Studios

3 Wildstorm

Time taken: 43.327 seconds, Fetched: 26 row(s)

4. **select count(hero_name) CountHero,hero_gender from heroes group by(hero_gender);**

```
Stage: Stage 1: Map: 1 Reduce: 1 Cumulative CPU: 2.70 sec HDFS Read:
Total MapReduce CPU Time Spent: 2 seconds 760 msec
OK
29      -
200     Female
1       Gender
505     Male
Time taken: 41.634 seconds, Fetched: 4 row(s)
hive> █
```

5. **select MAX(hero_height) MAXHEIGHT,hero_gender from heroes group by(hero_gender);**

```
Total MapReduce CPU Time Spent: 2 seconds 770 msec
OK
198.0   -
366.0   Female
NULL    Gender
975.0   Male
Time taken: 44.5 seconds, Fetched: 4 row(s)
hive> █
```

7. **select MAX(hero_weight) MAXWEIGHT,hero_gender from heroes group by(hero_gender);**

```
Total MapReduce CPU Time Spent: 2 seconds 840 msec
OK
383.0   -
630.0   Female
NULL    Gender
900.0   Male
Time taken: 45.13 seconds, Fetched: 4 row(s)
hive> █
```

8. **Select count(hero_alignment),hero_alignment from heroes where hero_gender='Male' group by (hero_alignment);**

```
Total MapReduce CPU Time Spent: 3 seconds 290 msec
OK
6      -
165    bad
316    good
18     neutral
Time taken: 44.197 seconds, Fetched: 4 row(s)
```

9. **select count(hero_alignment),hero_alignment from heroes where hero_gender='Female' group by (hero_alignment);**

```
Total MapReduce CPU Time Spent: 2 seconds 960 msec
OK
35     bad
161    good
4      neutral
Time taken: 41.72 seconds, Fetched: 3 row(s)
```

9.select count(hero_race),hero_race from heroes group by (hero_race) ORDER BY hero_race asc limit 5;

```
Total MapReduce CPU Time Spent: 4 seconds 826
OK
304      -
7        Alien
5        Alpha
2        Amazon
9        Android
Time taken: 79.459 seconds, Fetched: 5 row(s)
```

10. `select distinct(hero_skin_color) from heroes;`


```
stage-stage-1: Map: 1 Reduce: 1 Cumulative
Total MapReduce CPU Time Spent: 2 seconds 920
OK
-
Skin color
black
blue
blue-white
gold
gray
green
grey
orange
orange / white
pink
purple
red
red / black
silver
white
yellow
Time taken: 42.99 seconds, Fetched: 18 row(s)
```

Cassandra : Coursera Use Case

First start with creating the Keyspace for Coursera usecase .

```
cqlsh> describe KEYSAPACES;

system_schema  system_auth  system  books  system_distributed  system_traces

cqlsh> create keyspace coursera WITH REPLICATION = { 'class':'SimpleStrategy','replication_factor':1};
```

Then create the Courses_By_learner tables

```
cqlsh:coursea> CREATE TABLE courses_by_learner(course_id int,course_name text,learner_id int ,learner_name text,course_reg_date timestamp,course_fee int,
... PRIMARY KEY(course_id, learner_id));
cqlsh:coursea>
```

Then insert the data in Courses_by_learner table

```
cqlsh:coursea> INSERT into courses_by_learner(course_id,course_name,learner_id,learner_name,course_reg_date,course_fee)
... VALUES (1,'DATA SCIENCE',8,'RAJU NEAKDI','2018-08-01',600);
cqlsh:coursea> INSERT into courses_by_learner(course_id,course_name,learner_id,learner_name,course_reg_date,course_fee)
... VALUES (2,'WEB',9,'SWATI SINGH','2018-01-01',800);
cqlsh:coursea> INSERT into courses_by_learner(course_id,course_name,learner_id,learner_name,course_reg_date,course_fee)
... VALUES (3,'SPARK',4,'RUTHVICK P','2017-08-01',900);
cqlsh:coursea> INSERT into courses_by_learner(course_id,course_name,learner_id,learner_name,course_reg_date,course_fee)
... VALUES (4,'BIGDAT',5,'ZEENAT','2016-01-01',1000);
```

Then we have run the three queries as shown below.

1. Select course id, learner id , learner name where course reg date reater than 2017-01-01.

```
cqlsh:coursea> select course_id,learner_id,learner_name from courses_by_learner where course_reg_date>'2017-01-01' ALLOW FILTERING;
```

course_id	learner_id	learner_name
1	8	RAJU NEAKDI
2	9	SWATI SINGH
3	4	RUTHVICK P

(3 rows)

2. Select course fee, course name, for course fee greater than 800.

```
(2 rows)  
[cqlsh:coursera> select course_fee,course_name from courses_by_learner where course_fee>800 ALLOW FILTERING;  
  
  course_fee | course_name  
-----+-----  
      1000 |    BIGDAT  
      900  |    SPARK  
  
(2 rows)
```

3. Select course fee learner name where course name DATA SCIENCE.

```
[cqlsh:coursera> select course_fee,learner_name from courses_by_learner where course_name='DATA SCIENCE' ALLOW FILTERING;  
  
  course_fee | learner_name  
-----+-----  
      600 | RAJU NEAKDI
```

Solr Use Case:

Super Heroes Dataset:

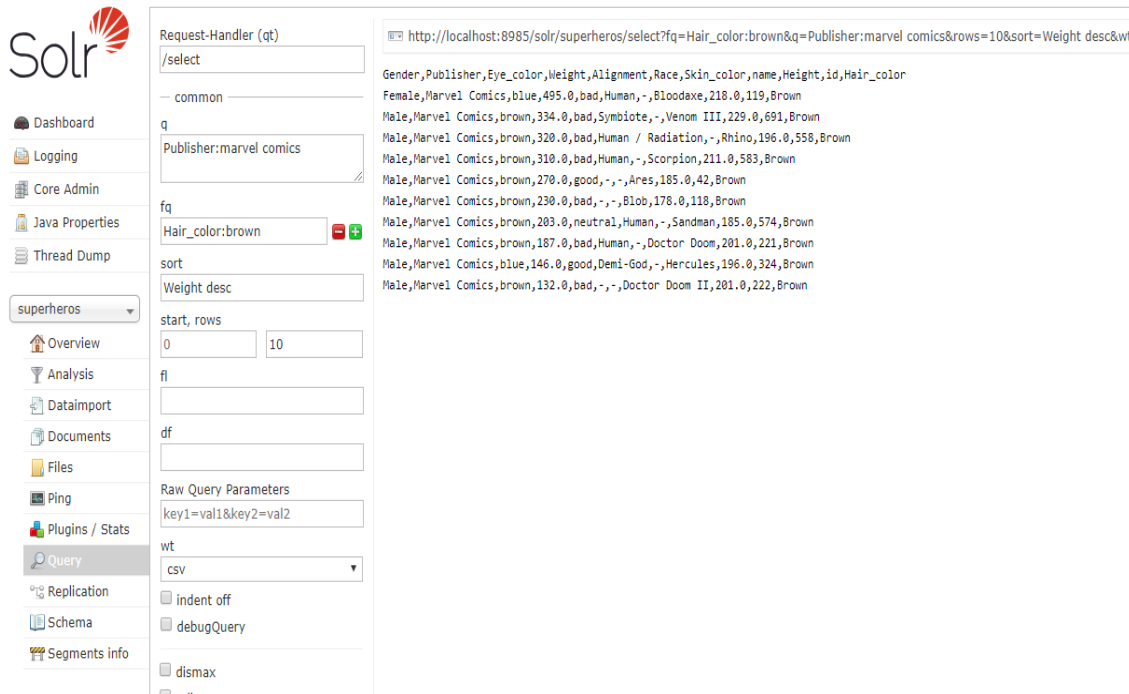
1. Details of superheroes with name SILK

The screenshot shows the Solr Admin interface in a web browser. The address bar indicates the URL is `localhost:8985/solr/#/superheros/query`. The left sidebar contains navigation links: Dashboard, Logging, Core Admin, Java Properties, Thread Dump, a dropdown menu for 'superheros' (with sub-links Overview, Analysis, Dataimport, Documents, Files, Ping, Plugins / Stats, Query, Replication, Schema, and Segments info), and other system links. The main panel is titled 'Request-Handler (qt)' and shows the selected handler as `/select`. The 'q' (query) field contains `name:silk`. The 'fq' (filter query) field is empty. The 'sort' field is empty. The 'start, rows' section shows 'start' as 0 and 'rows' as 10. The 'fl' (fields list) field is empty. The 'df' (default field list) field is empty. The 'Raw Query Parameters' section shows `key1=val1&key2=val2`. The 'wt' (output format) is set to 'csv'. The 'indent off' checkbox is checked. The 'debugQuery' checkbox is checked. The 'diagnose' checkbox is checked. The results pane on the right shows the URL `http://localhost:8985/solr/superheros/select?q=name:silk&wt=csv` and the resulting CSV data:

Gender	Publisher	Eye_color	Weight	Alignment	Race	Skin_color	name	Height	id	Hair_color
Female	Marvel Comics	brown	-99.0	good	Human	-	Silk	-99.0	597	Black
Female	DC Comics	-	-99.0	good	-	-	Silk Spectre	-99.0	598	-
Female	DC Comics	-	-99.0	good	-	-	Silk Spectre II	-99.0	599	-

Query Time : 2msec

2. Details of superheroes whose Publisher is Marvel Comics and has brown Hair_color and sort by Weight in descending order.



The screenshot shows the Solr Admin interface. The left sidebar contains navigation links: Dashboard, Logging, Core Admin, Java Properties, Thread Dump, a dropdown menu with 'superheroes' selected, and links for Overview, Analysis, Dataimport, Documents, Files, Ping, Plugins / Stats, Query (highlighted), Replication, Schema, and Segments info. The main panel is titled 'Request-Handler (qt)' and shows the following configuration:

- Request-Handler (qt): /select
- q: Publisher:marvel comics
- fq: Hair_color:brown
- sort: Weight desc
- start, rows: 0, 10
- Raw Query Parameters: key1=val1&key2=val2
- wt: csv
- debugQuery: ☐
- dismax: ☐

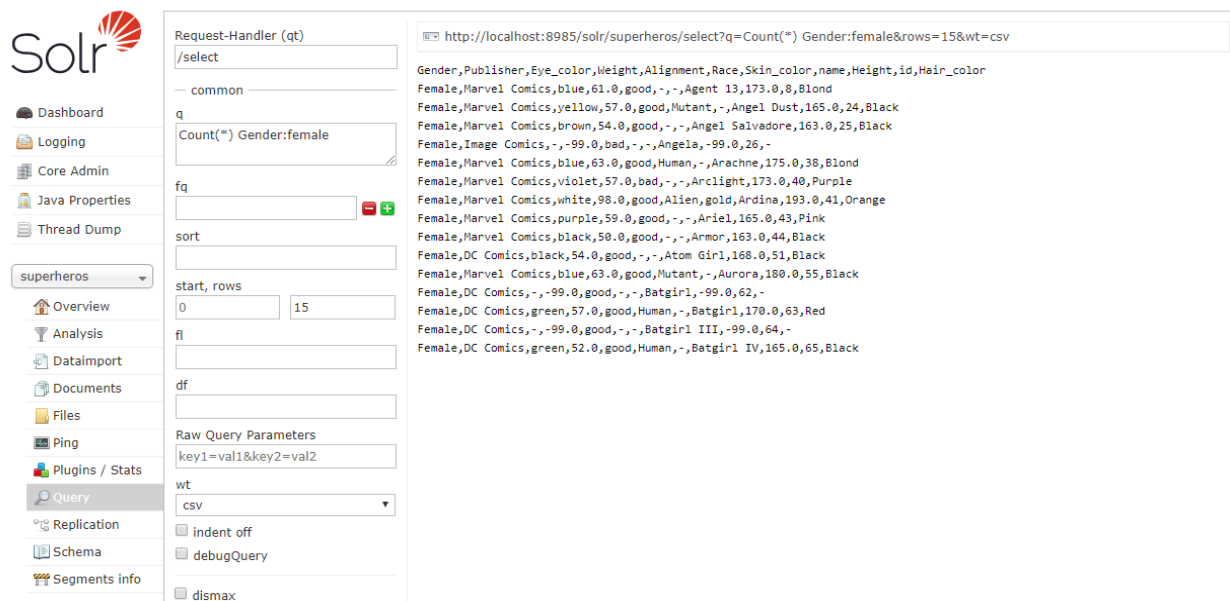
The URL bar shows: <http://localhost:8985/solr/superheroes/select?q=Publisher:marvel comics&rows=10&sort=Weight desc&wt=csv>

The response shows a table of superhero details:

Gender	Publisher	Eye_color	Weight	Alignment	Race	Skin_color	name	Height	id	Hair_color
Female	Marvel Comics	blue	495.0	bad	Human	-	Bloodaxe	218.0	119	Brown
Male	Marvel Comics	brown	334.0	bad	Symbiote	-	Venom III	229.0	691	Brown
Male	Marvel Comics	brown	320.0	bad	Human	/	Radiation	196.0	558	Brown
Male	Marvel Comics	brown	310.0	bad	Human	-	Scorpion	211.0	583	Brown
Male	Marvel Comics	brown	270.0	good	-	-	Ares	185.0	42	Brown
Male	Marvel Comics	brown	230.0	bad	-	-	Blob	178.0	118	Brown
Male	Marvel Comics	brown	203.0	neutral	Human	-	Sandman	185.0	574	Brown
Male	Marvel Comics	brown	187.0	bad	Human	-	Doctor Doom	201.0	221	Brown
Male	Marvel Comics	blue	146.0	good	Demi-God	-	Hercules	196.0	324	Brown
Male	Marvel Comics	brown	132.0	bad	-	-	Doctor Doom II	201.0	222	Brown

Query Time: 1msec

3. Count top 15 Female Superheroes



The screenshot shows the Solr Admin interface. The left sidebar is the same as in the previous screenshot. The main panel is titled 'Request-Handler (qt)' and shows the following configuration:

- Request-Handler (qt): /select
- q: Count(*) Gender:female
- sort: (empty)
- start, rows: 0, 15
- Raw Query Parameters: key1=val1&key2=val2
- wt: csv
- debugQuery: ☐
- dismax: ☐

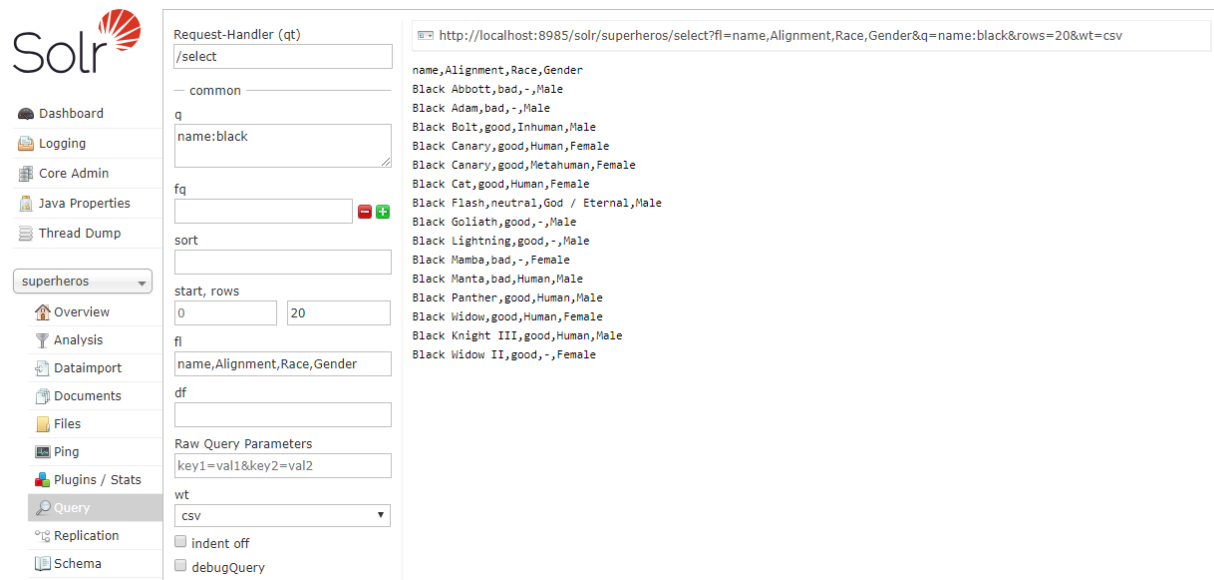
The URL bar shows: [http://localhost:8985/solr/superheroes/select?q=Count\(*\) Gender:female&rows=15&wt=csv](http://localhost:8985/solr/superheroes/select?q=Count(*) Gender:female&rows=15&wt=csv)

The response shows a table of superhero details:

Gender	Publisher	Eye_color	Weight	Alignment	Race	Skin_color	name	Height	id	Hair_color
Female	Marvel Comics	blue	61.0	good	-	-	Agent 13	173.0	8	Blond
Female	Marvel Comics	yellow	57.0	good	Mutant	-	Angel Dust	165.0	24	Black
Female	Marvel Comics	brown	54.0	good	-	-	Angel Salvadore	163.0	25	Black
Female	Image Comics	-	99.0	bad	-	-	Angela	99.0	26	-
Female	Marvel Comics	blue	63.0	good	Human	-	Arachne	175.0	38	Blond
Female	Marvel Comics	violet	57.0	bad	-	-	Arclight	173.0	40	Purple
Female	Marvel Comics	white	98.0	good	Alien	gold	Ardina	193.0	41	Orange
Female	Marvel Comics	purple	59.0	good	-	-	Ariel	165.0	43	Pink
Female	Marvel Comics	black	50.0	good	-	-	Armor	163.0	44	Black
Female	DC Comics	black	54.0	good	-	-	Atom Girl	168.0	51	Black
Female	Marvel Comics	blue	63.0	good	Mutant	-	Aurora	180.0	55	Black
Female	DC Comics	-	99.0	good	-	-	Batgirl	99.0	62	-
Female	DC Comics	green	57.0	good	Human	-	Batgirl	170.0	63	Red
Female	DC Comics	-	99.0	good	-	-	Batgirl III	99.0	64	-
Female	DC Comics	green	52.0	good	Human	-	Batgirl IV	165.0	65	Black

Query Time: 3msec

4. Details (only name, Race, Alignment, Gender) of superheroes whose name has Black



The Solr Admin UI shows a query for superheroes with 'black' in their name. The left sidebar contains navigation links: Dashboard, Logging, Core Admin, Java Properties, Thread Dump, and a dropdown menu for 'superheroes' with sub-links: Overview, Analysis, Dataimport, Documents, Files, Ping, Plugins / Stats, Query (selected), Replication, and Schema. The main panel is titled 'Request-Handler (qt)' and shows the following fields:

- Request-Handler (qt): /select
- common: (empty)
- q: name:black
- fq: (empty)
- sort: (empty)
- start, rows: 0, 20
- fl: name,Alignment,Race,Gender
- df: (empty)
- Raw Query Parameters: key1=val1&key2=val2
- wt: csv
- indent off
- debugQuery

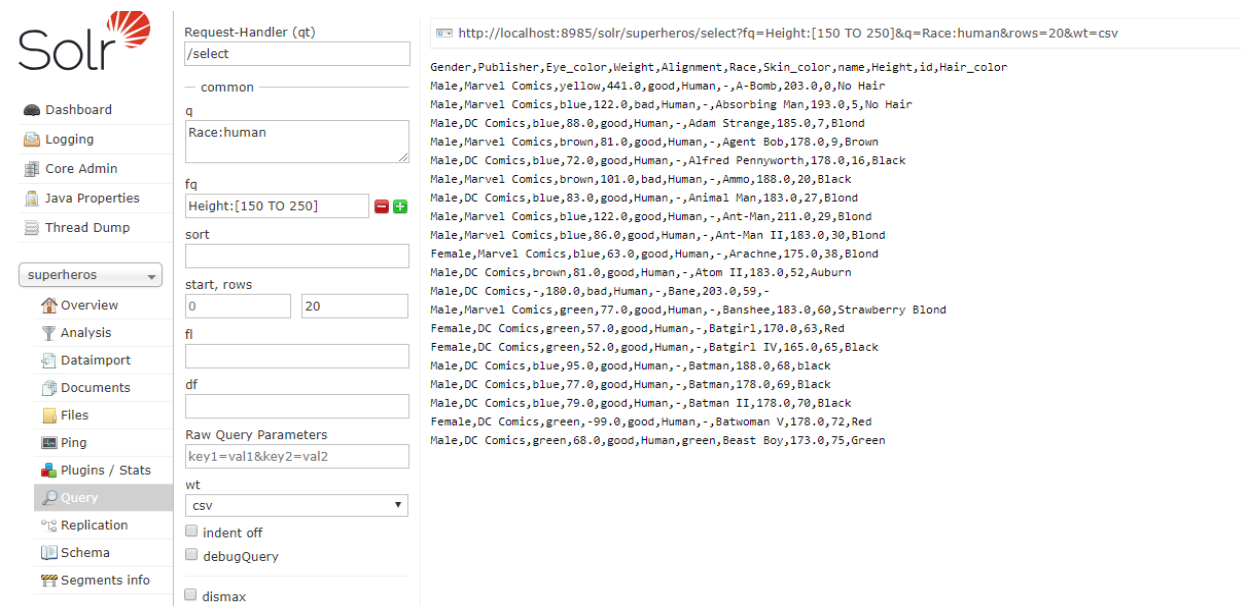
The URL bar shows: <http://localhost:8985/solr/superheroes/select?fl=name,Alignment,Race,Gender&q=name:black&rows=20&wt=csv>

The results are displayed as a list of 20 superheroes, each with their name, Alignment, Race, and Gender:

- Black Abbott,bad,-,Male
- Black Adam,bad,-,Male
- Black Bolt,good,Inhuman,Male
- Black Canary,good,Human,Female
- Black Canary,good,MetaHuman,Female
- Black Cat,good,Human,Female
- Black Flash,neutral,God / Eternal,Male
- Black Goliath,good,-,Male
- Black Lightning,good,-,Male
- Black Mamba,bad,-,Female
- Black Manta,bad,Human,Male
- Black Panther,good,Human,Male
- Black Widow,good,Human,Female
- Black Knight III,good,Human,Male
- Black Widow II,good,-,Female

Query Time : 2msec

5. Top 20 Superheroes whose race is 'Human' and Height range is 150 - 250.



The Solr Admin UI shows a query for superheroes with race 'Human' and height between 150 and 250. The left sidebar contains navigation links: Dashboard, Logging, Core Admin, Java Properties, Thread Dump, and a dropdown menu for 'superheroes' with sub-links: Overview, Analysis, Dataimport, Documents, Files, Ping, Plugins / Stats, Query (selected), Replication, and Schema. The main panel is titled 'Request-Handler (qt)' and shows the following fields:

- Request-Handler (qt): /select
- common: (empty)
- q: Race:human
- fq: Height:[150 TO 250]
- sort: (empty)
- start, rows: 0, 20
- fl: (empty)
- df: (empty)
- Raw Query Parameters: key1=val1&key2=val2
- wt: csv
- indent off
- debugQuery
- dismax

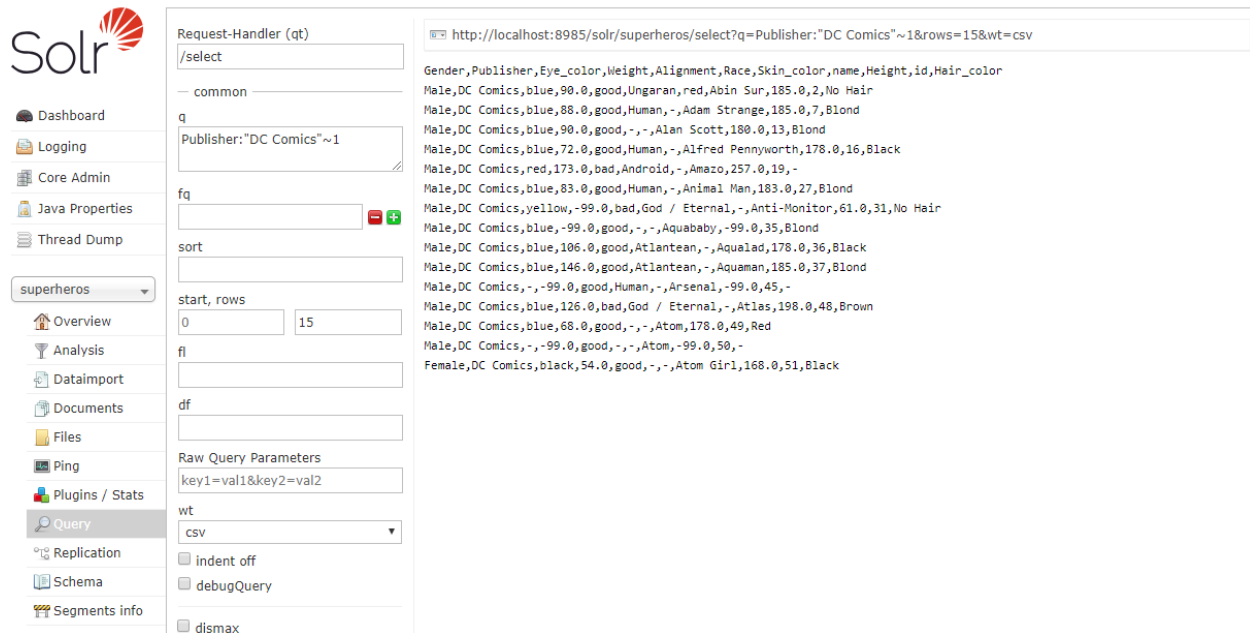
The URL bar shows: [http://localhost:8985/solr/superheroes/select?fq=Height:\[150 TO 250\]&q=Race:human&rows=20&wt=csv](http://localhost:8985/solr/superheroes/select?fq=Height:[150 TO 250]&q=Race:human&rows=20&wt=csv)

The results are displayed as a list of 20 superheroes, each with their Gender, Publisher, Eye_color, Height, Alignment, Race, Skin_color, name, Height, id, and Hair_color:

- Male,Marvel Comics,yellow,441.0,good,Human,-,A-Bomb,203.0,0,No Hair
- Male,Marvel Comics,blue,122.0,bad,Human,-,Absorbing Man,193.0,5,No Hair
- Male,DC Comics,blue,88.0,good,Human,-,Adam Strange,185.0,7,Blond
- Male,Marvel Comics,brown,81.0,good,Human,-,Agent Bob,178.0,9,Brown
- Male,DC Comics,blue,72.0,good,Human,-,Alfred Pennyworth,178.0,16,Black
- Male,Marvel Comics,brown,101.0,bad,Human,-,Ammo,188.0,20,Black
- Male,DC Comics,blue,83.0,good,Human,-,Animal Man,183.0,27,Blond
- Male,Marvel Comics,blue,122.0,good,Human,-,Ant-Man,211.0,29,Blond
- Male,Marvel Comics,blue,86.0,good,Human,-,Ant-Man II,183.0,30,Blond
- Female,Marvel Comics,blue,63.0,good,Human,-,Arachne,175.0,38,Blond
- Male,DC Comics,brown,81.0,good,Human,-,Atom II,183.0,52,Auburn
- Male,DC Comics,-,180.0,bad,Human,-,Bane,203.0,59,-
- Male,Marvel Comics,green,77.0,good,Human,-,Banshee,183.0,60,Strawberry Blond
- Female,DC Comics,green,57.0,good,Human,-,Batgirl I,170.0,63,Red
- Female,DC Comics,green,52.0,good,Human,-,Batgirl IV,165.0,65,Black
- Male,DC Comics,blue,95.0,good,Human,-,Batman,188.0,68,black
- Male,DC Comics,blue,77.0,good,Human,-,Batman,178.0,69,Black
- Male,DC Comics,blue,79.0,good,Human,-,Batman II,178.0,70,Black
- Female,DC Comics,green,-99.0,good,Human,-,Batwoman V,178.0,72,Red
- Male,DC Comics,green,68.0,good,Human,green,Beast Boy,173.0,75,Green

Query Time: 2msec

6. Search for DC Comics within 1 word from each other (Proximity)

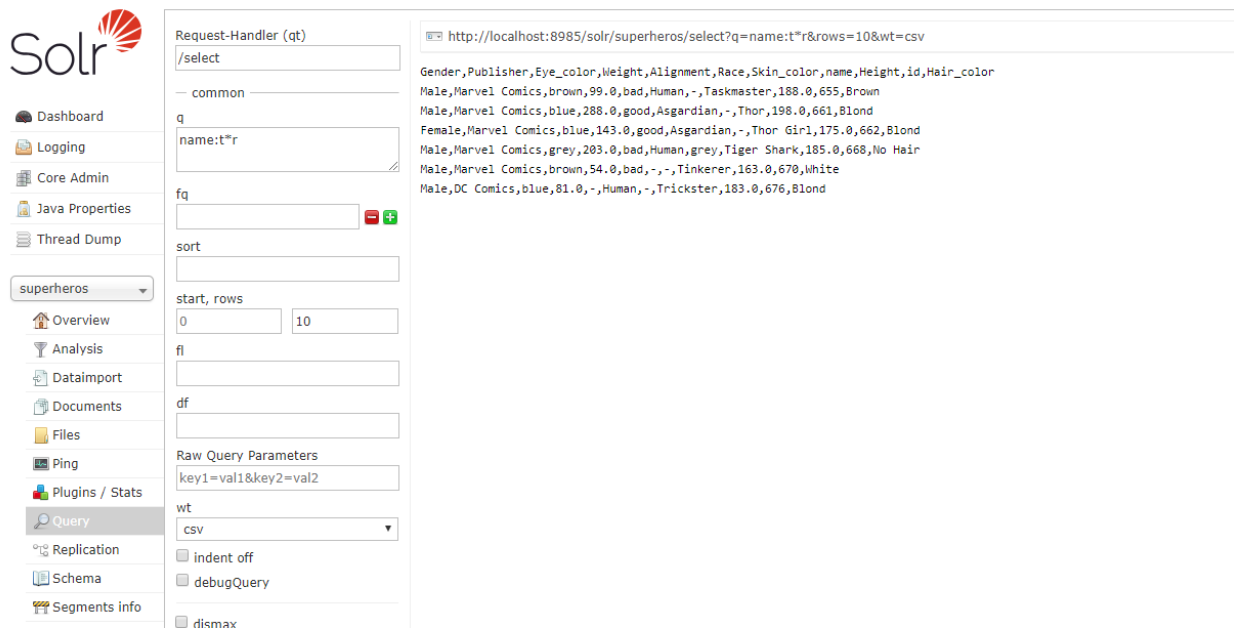


The screenshot shows the Solr Admin UI with the 'Query' tab selected. The 'Request-Handler (qt)' is set to '/select'. The 'q' field contains the query: `Publisher:"DC Comics"~1`. The 'wt' dropdown is set to 'csv'. The 'start, rows' fields are set to 0 and 15 respectively. The 'Raw Query Parameters' field shows `key1=val1&key2=val2`. The 'wt' dropdown is set to 'csv'. The 'indent off' checkbox is checked. The 'debugQuery' checkbox is checked. The 'dismax' checkbox is checked. The URL bar shows: `http://localhost:8985/solr/superheros/select?q=Publisher:"DC Comics"~1&rows=15&wt=csv`. The results are displayed in a table with columns: Gender, Publisher, Eye_color, Weight, Alignment, Race, Skin_color, name, Height, id, Hair_color. The results list various DC Comics characters and their attributes.

Gender	Publisher	Eye_color	Weight	Alignment	Race	Skin_color	name	Height	id	Hair_color
Male	DC Comics	blue	90.0	good	Ungaran	red	Abin Sur	185.0	2	No Hair
Male	DC Comics	blue	88.0	good	Human	-	Adam Strange	185.0	7	Blond
Male	DC Comics	blue	90.0	good	-	-	Alan Scott	180.0	13	Blond
Male	DC Comics	blue	72.0	good	Human	-	Alfred Pennyworth	178.0	16	Black
Male	DC Comics	red	173.0	bad	Android	-	Amazo	257.0	19	-
Male	DC Comics	blue	83.0	good	Human	-	Animal Man	183.0	27	Blond
Male	DC Comics	yellow	99.0	bad	God / Eternal	-	Anti-Monitor	61.0	31	No Hair
Male	DC Comics	blue	99.0	good	-	-	Aquababy	99.0	35	Blond
Male	DC Comics	blue	106.0	good	Atlantean	-	Aqualad	178.0	36	Black
Male	DC Comics	blue	146.0	good	Atlantean	-	Aquaman	185.0	37	Blond
Male	DC Comics	-	99.0	good	Human	-	Arsenal	99.0	45	-
Male	DC Comics	blue	126.0	bad	God / Eternal	-	Atlas	198.0	48	Brown
Male	DC Comics	blue	68.0	good	-	-	Atom	178.0	49	Red
Male	DC Comics	-	99.0	good	-	-	Atom	99.0	50	-
Female	DC Comics	black	54.0	good	-	-	Atom Girl	168.0	51	Black

Query Time: 61msec

7. Details of Superheroes whose name starts with T and ends with R

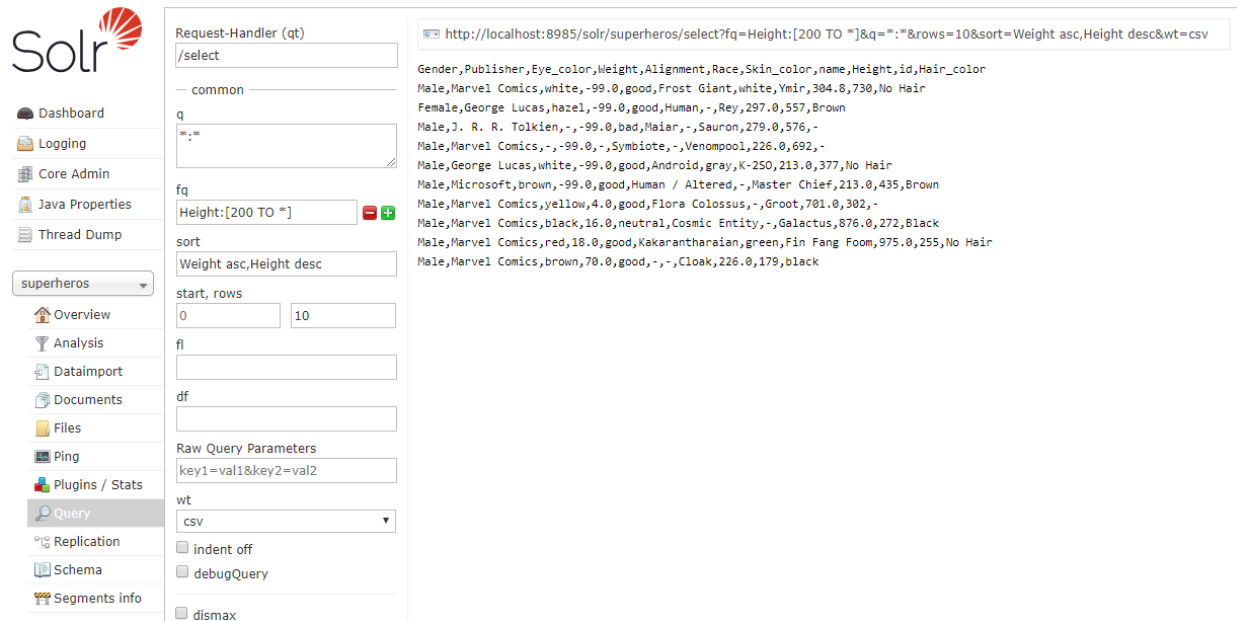


The screenshot shows the Solr Admin UI with the 'Query' tab selected. The 'Request-Handler (qt)' is set to '/select'. The 'q' field contains the query: `name:t*r`. The 'wt' dropdown is set to 'csv'. The 'start, rows' fields are set to 0 and 10 respectively. The 'Raw Query Parameters' field shows `key1=val1&key2=val2`. The 'wt' dropdown is set to 'csv'. The 'indent off' checkbox is checked. The 'debugQuery' checkbox is checked. The 'dismax' checkbox is checked. The URL bar shows: `http://localhost:8985/solr/superheros/select?q=name:t*r&rows=10&wt=csv`. The results are displayed in a table with columns: Gender, Publisher, Eye_color, Weight, Alignment, Race, Skin_color, name, Height, id, Hair_color. The results list various superheroes whose names start with 'T' and end with 'R'.

Gender	Publisher	Eye_color	Weight	Alignment	Race	Skin_color	name	Height	id	Hair_color
Male	Marvel Comics	brown	99.0	bad	Human	-	Taskmaster	188.0	655	Brown
Male	Marvel Comics	blue	288.0	good	Asgardian	-	Thor	198.0	661	Blond
Female	Marvel Comics	blue	143.0	good	Asgardian	-	Thor Girl	175.0	662	Blond
Male	Marvel Comics	grey	203.0	bad	Human	grey	Tiger Shark	185.0	668	No Hair
Male	Marvel Comics	brown	54.0	bad	-	-	Tinkerer	163.0	670	White
Male	DC Comics	blue	81.0	-	Human	-	Trickster	183.0	676	Blond

Query Time: 5msec

8. Details of Superheroes whose Height is greater than 150 and sort by Height in descending and Weight in ascending order.

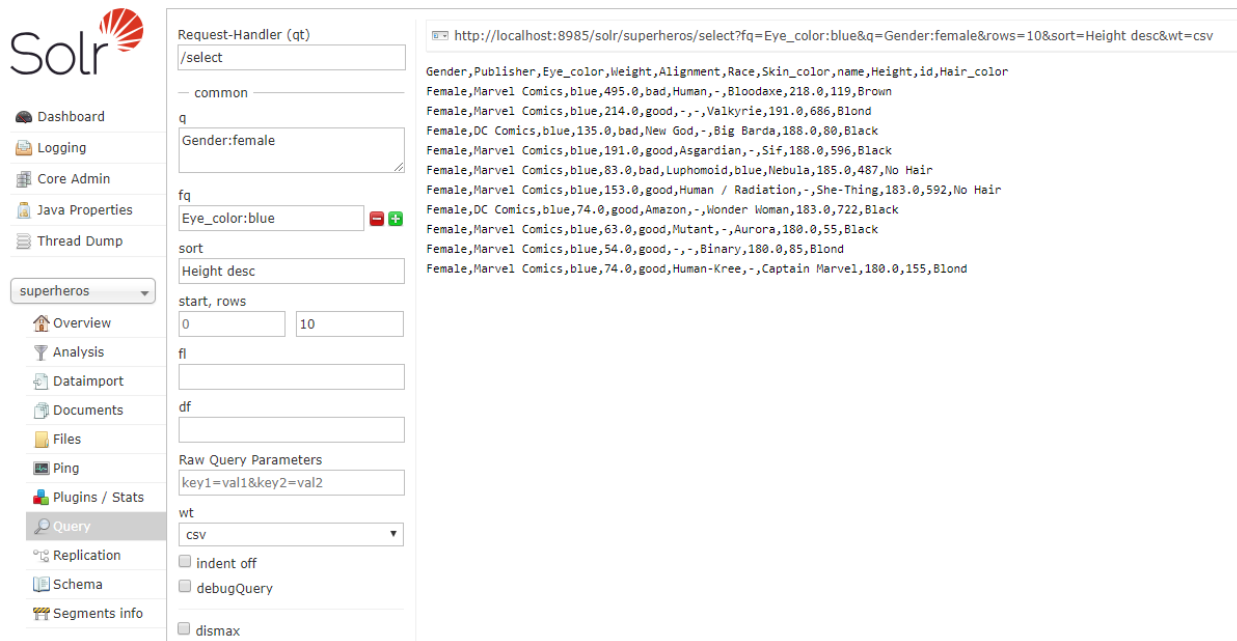


The screenshot shows the Solr Admin interface with the 'superheroes' collection selected. The 'Query' tab is active. The request handler is set to '/select'. The query is: `Height:[200 TO *]`. The sort order is 'Weight asc, Height desc'. The start row is 0 and the number of rows is 10. The raw query parameters are 'key1=val1&key2=val2'. The weight format is 'csv'. The query results are displayed in a table with columns: Gender, Publisher, Eye_color, Weight, Alignment, Race, Skin_color, name, Height, id, Hair_color. The results show 10 superheroes with heights greater than or equal to 200, sorted by weight in ascending order and then height in descending order.

Gender	Publisher	Eye_color	Weight	Alignment	Race	Skin_color	name	Height	id	Hair_color		
Male	Marvel Comics	white	-99.0	good	Frost Giant	white	Ymir	304.8	730	No Hair		
Female	George Lucas	hazel	-99.0	good	Human	-	Rey	297.0	557	Brown		
Male	J. R. R. Tolkien	-	-99.0	bad	Maia	-	Sauron	279.0	576	-		
Male	Marvel Comics	-	-99.0	-	Symbiote	-	Venompool	226.0	692	-		
Male	George Lucas	white	-99.0	good	Android	gray	K-250	213.0	377	No Hair		
Male	Microsoft	brown	-99.0	good	Human	-	Altered	-	Master Chief	213.0	435	Brown
Male	Marvel Comics	yellow	4.0	good	Flora Colossus	-	Groot	701.0	302	-		
Male	Marvel Comics	black	16.0	neutral	Cosmic Entity	-	Galactus	876.0	272	Black		
Male	Marvel Comics	red	18.0	good	Kakarantharaian	green	Fin Fang Foom	975.0	255	No Hair		
Male	Marvel Comics	brown	70.0	good	-	-	Cloak	226.0	179	black		

Query Time: 1msec

9. Details of Female Superheroes with blue eyes and sort by Height in descending order.



The screenshot shows the Solr Admin interface with the 'superheroes' collection selected. The 'Query' tab is active. The request handler is set to '/select'. The query is: `Gender:female&Eye_color:blue`. The sort order is 'Height desc'. The start row is 0 and the number of rows is 10. The raw query parameters are 'key1=val1&key2=val2'. The weight format is 'csv'. The query results are displayed in a table with columns: Gender, Publisher, Eye_color, Weight, Alignment, Race, Skin_color, name, Height, id, Hair_color. The results show 10 female superheroes with blue eyes, sorted by height in descending order.

Gender	Publisher	Eye_color	Weight	Alignment	Race	Skin_color	name	Height	id	Hair_color		
Female	Marvel Comics	blue	495.0	bad	Human	-	Bloodaxe	218.0	119	Brown		
Female	Marvel Comics	blue	214.0	good	-	-	Valkyrie	191.0	686	Blond		
Female	DC Comics	blue	135.0	bad	New God	-	Big Barda	188.0	80	Black		
Female	Marvel Comics	blue	191.0	good	Asgardian	-	Sif	188.0	596	Black		
Female	Marvel Comics	blue	83.0	bad	Luphomoid	blue	Nebula	185.0	487	No Hair		
Female	Marvel Comics	blue	153.0	good	Human	-	Radiation	-	She-Thing	183.0	592	No Hair
Female	DC Comics	blue	74.0	good	Amazon	-	Wonder Woman	183.0	722	Black		
Female	Marvel Comics	blue	63.0	good	Mutant	-	Aurora	180.0	55	Black		
Female	Marvel Comics	blue	54.0	good	-	-	Binary	180.0	85	Blond		
Female	Marvel Comics	blue	74.0	good	Human-Kree	-	Captain Marvel	180.0	155	Blond		

Query Time: 3msec

10. Details of Superheroes with either White Hair colour or Grey Skin colour.

The screenshot displays the Solr Admin interface. On the left is a sidebar with navigation links: Dashboard, Logging, Core Admin, Java Properties, Thread Dump, a dropdown menu with 'superheroes' selected, Overview, Analysis, Dataimport, Documents, Files, Ping, Plugins / Stats, Query (highlighted), Replication, Schema, and Segments Info. The main panel is titled 'Request-Handler (qt)' and shows a query configuration for the '/select' handler. The 'q' field contains the query: 'Hair_color:white Skin_color:grey'. The 'fq' field is empty. The 'sort' field is empty. The 'start, rows' field shows '0' and '10'. The 'wt' field is set to 'csv'. The 'Raw Query Parameters' field shows 'key1=val1&key2=val2'. The 'wt' dropdown is set to 'csv'. The 'indent off' checkbox is checked. The 'debugQuery' checkbox is checked. The 'dismax' checkbox is checked. The right panel shows the URL: 'http://localhost:8985/solr/superheroes/select?q=Hair_color:white Skin_color:grey&rows=10&wt=csv'. Below the URL is a table of results with columns: Gender, Publisher, Eye_color, Weight, Alignment, Race, Skin_color, name, Height, id, Hair_color. The results are as follows:

Gender	Publisher	Eye_color	Weight	Alignment	Race	Skin_color	name	Height	id	Hair_color
Male	Marvel Comics	white	97.0	good	Alien	grey	Century	201.0	166	White
Male	Marvel Comics	blue	188.0	bad	-	-	Air-Walker	188.0	11	White
Male	Marvel Comics	black	63.0	good	-	-	Beak	175.0	73	White
Male	Marvel Comics	red	184.0	bad	Demon	white	Blackout	191.0	108	White
Male	Marvel Comics	red	88.0	-	Alien	-	Blackwulf	188.0	110	White
Male	Marvel Comics	blue	158.0	good	Mutant	-	Cable	203.0	144	White
-	Marvel Comics	-	-99.0	good	-	-	Clea	-99.0	178	White
Female	Marvel Comics	red	67.0	neutral	Mutant	blue	Copycat	183.0	185	White
Male	DC Comics	blue	101.0	neutral	Human	-	Deathstroke	193.0	215	White
Male	DC Comics	red	412.0	bad	Alien	-	Doomsday	244.0	229	White

Query Time: 2msec

Evaluation:

We can see that Cassandra, Solr and hive framework tools can handle with any type of data and we can get meaningful insights from it.

Conclusion: Cassandra, Solr and Hive can work with Complex type of dataset very efficiently.