**BIGDATA PROGRAMMING Lab 2**

**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.

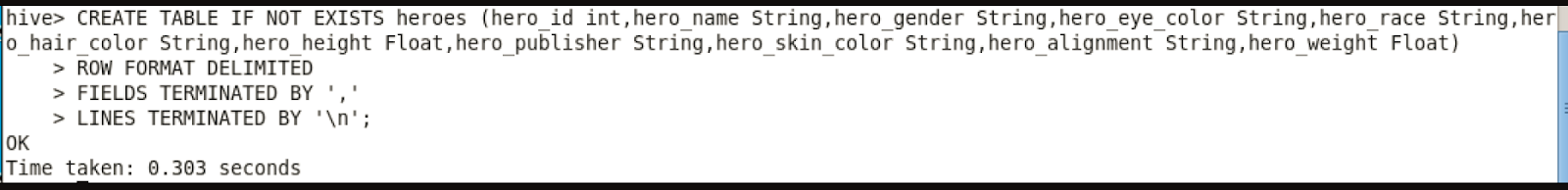
**Datasets:**

**Workflow:**

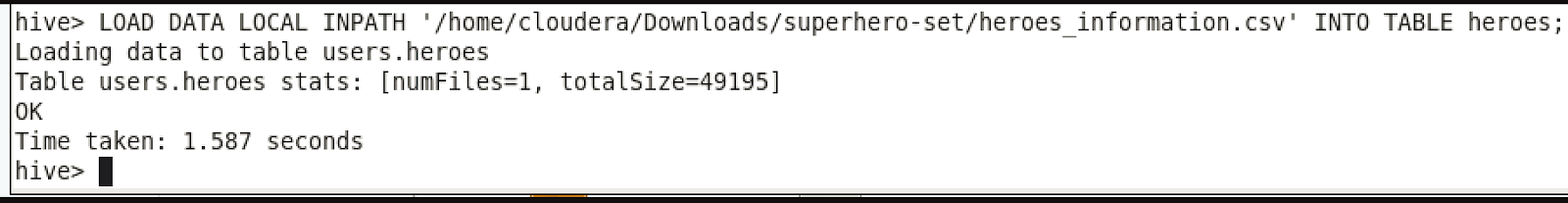
**Hive Queries:**

**We used the super hero dataset for Hive use case.**

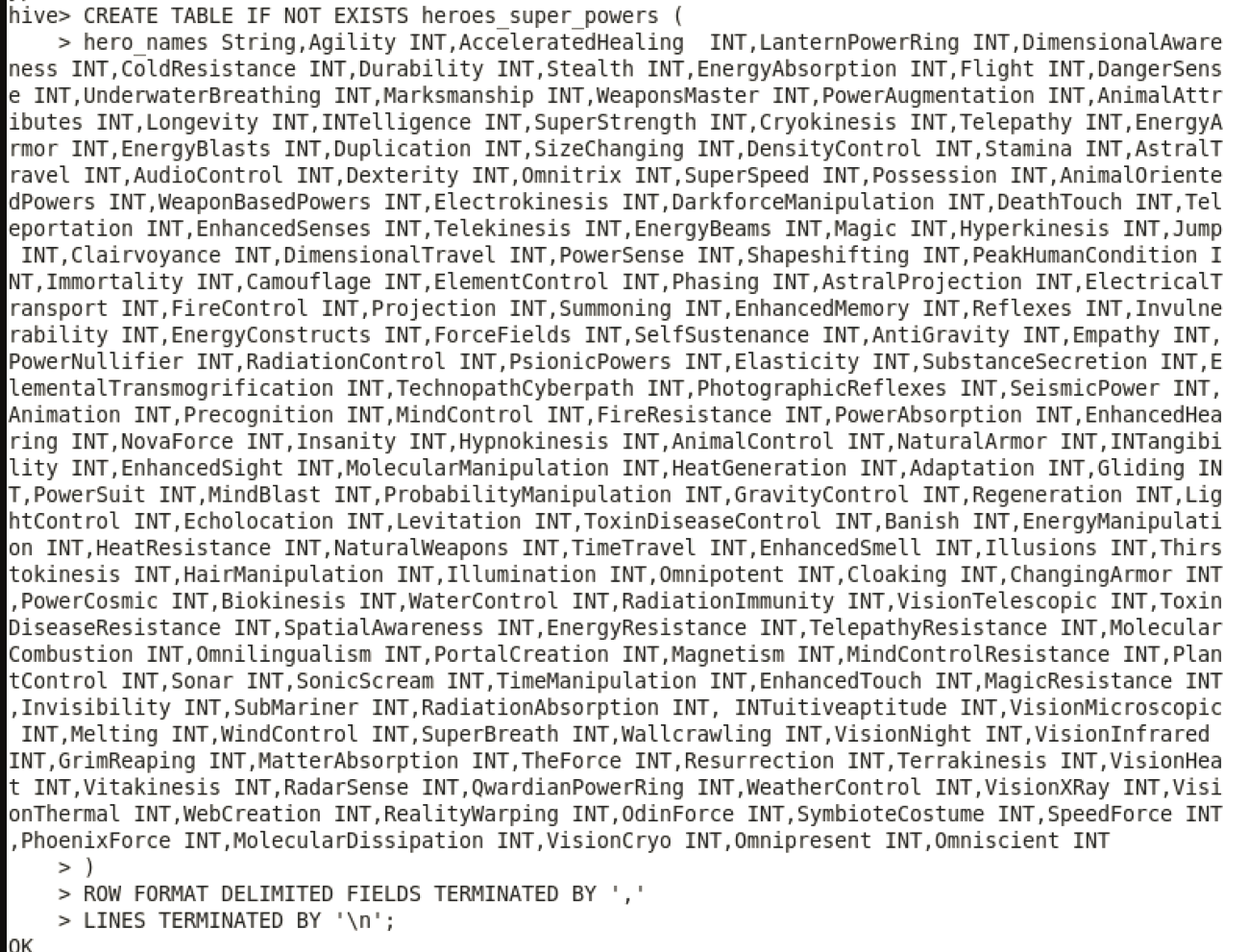
**Create the heroes table.**



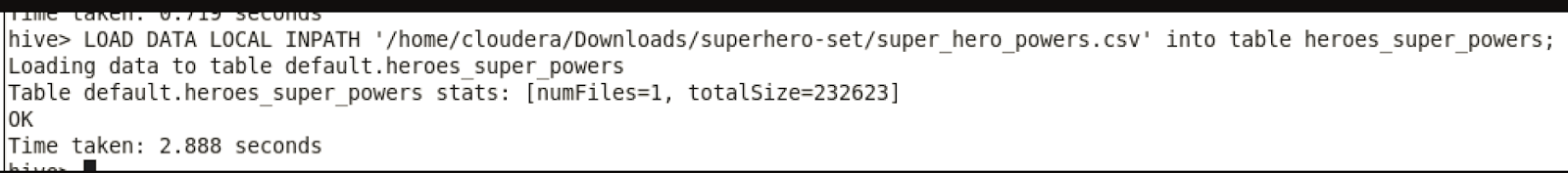
**Load the heroes table using the load statement**



**Created the Heroes Super powers Table.**



**Load heroes Power table.**



**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 +

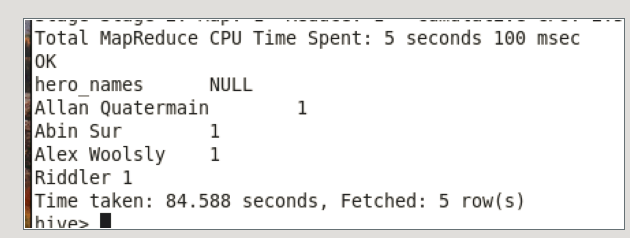
PlantControl +Sonar +SonicScream +TimeManipulation +EnhancedTouch +MagicResistance +Invisibility +SubMariner +RadiationAbsorption + INTuitiveaptitude +VisionMicroscopic +Melting +WindControl +SuperBreath +Wallcrawling +VisionNight +

VisionInfrared +GrimReaping +MatterAbsorption +TheForce +Resurrection +Terrakinesis +VisionHeat +Vitakinesis +RadarSense +QwardianPowerRing +WeatherControl +VisionXRay +VisionThermal +WebCreation +RealityWarping +OdinForce +

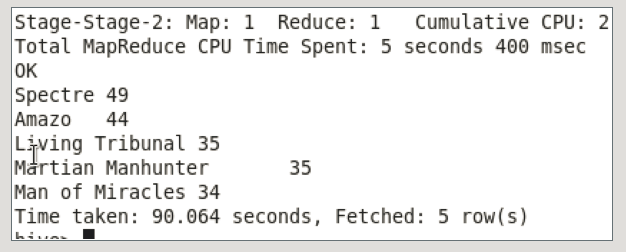
SymbioteCostume +SpeedForce +PhoenixForce +MolecularDissipation +VisionCryo +Omnipresent +Omniscient) from heroes\_super\_powers;

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

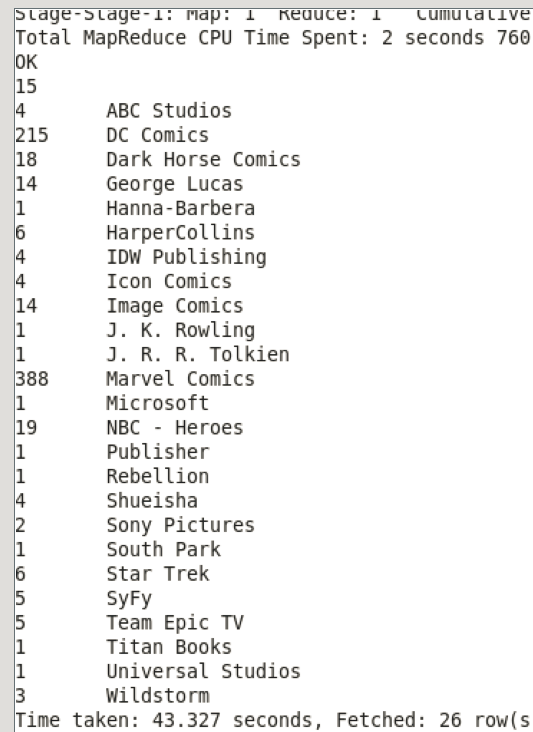
**1.select hero\_names,super\_power\_total from heroes\_super\_power\_total SORT BY super\_power\_total ASC limit 5;**



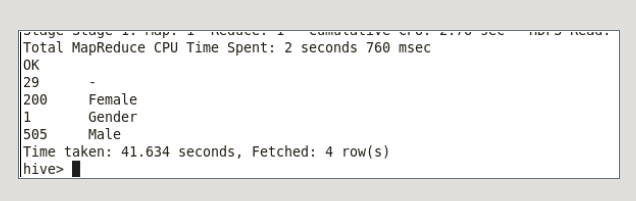
**2.select hero\_names,super\_power\_total from heroes\_super\_power\_total SORT BY super\_power\_total desc limit 5;**



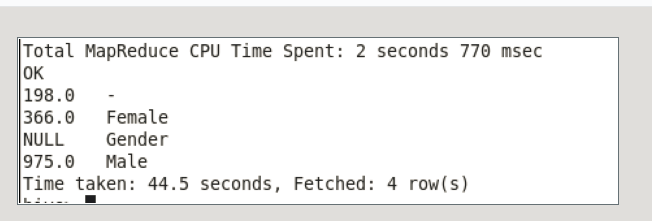
**3.select count(hero\_name) CountHero,hero\_publisher from heroes where hero\_publisher group by(hero\_publisher);**



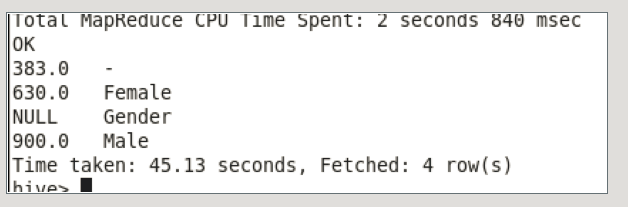
**4. select count(hero\_name) CountHero,hero\_gender from heroes  group by(hero\_gender);**



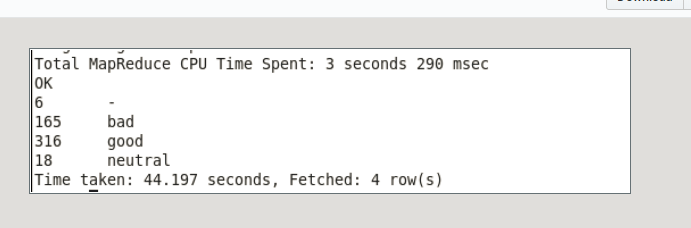
**5. select MAX(hero\_height) MAXHEIGHT,hero\_gender from heroes  group by(hero\_gender);**



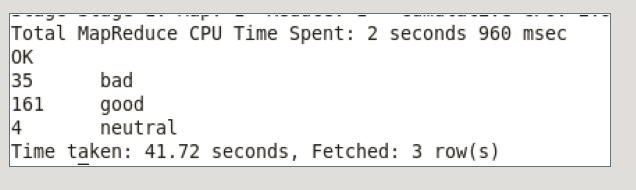
**select MAX(hero\_weight) MAXWEIGHT,hero\_gender from heroes  group by(hero\_gender);**



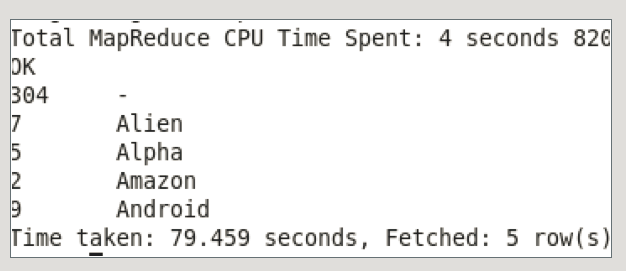
**Select count(hero\_alignment),hero\_alignment from heroes where hero\_gender='Male' group by (hero\_alignment);**



**select count(hero\_alignment),hero\_alignment from heroes where hero\_gender=‘Female' group by (hero\_alignment);**



**9.select count(hero\_race),hero\_race from heroes  group by (hero\_race) ORDER BY hero\_race asc limit 5;**

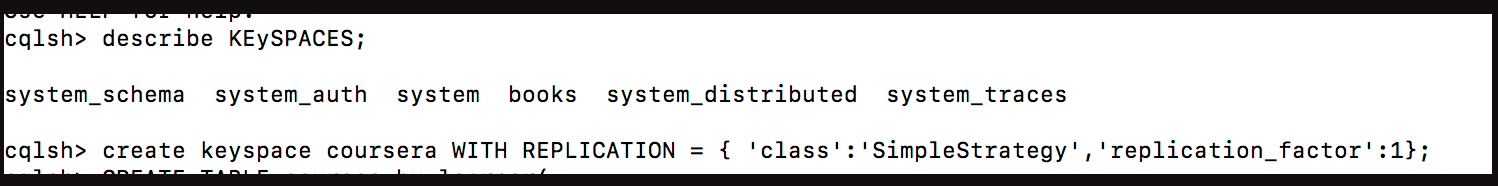


**10. select distinct(hero\_skin\_color) from heroes;**

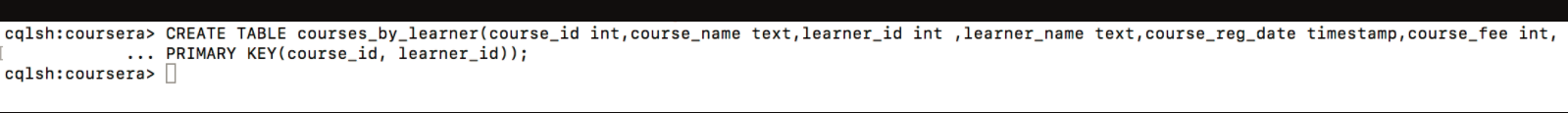


**Cassandra : Coursera Use Case**

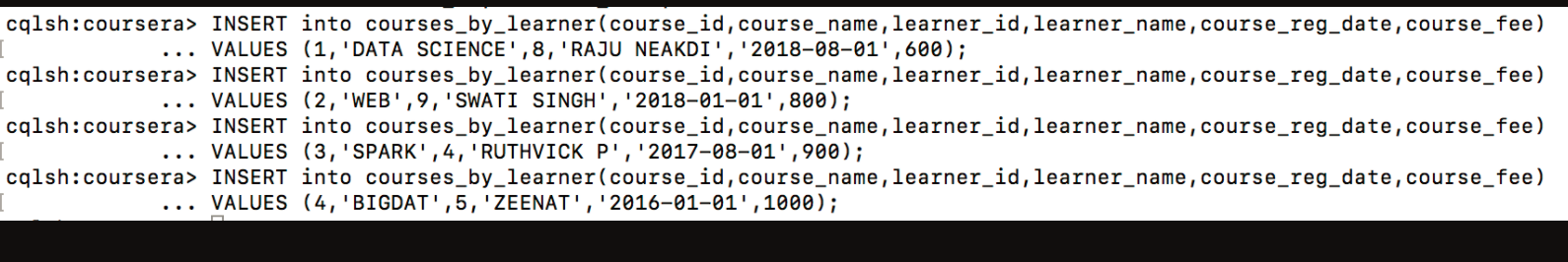
**First start with creating the Keyspace for Coursera usecase .**



**Then create the Courses\_By\_learner  tables**

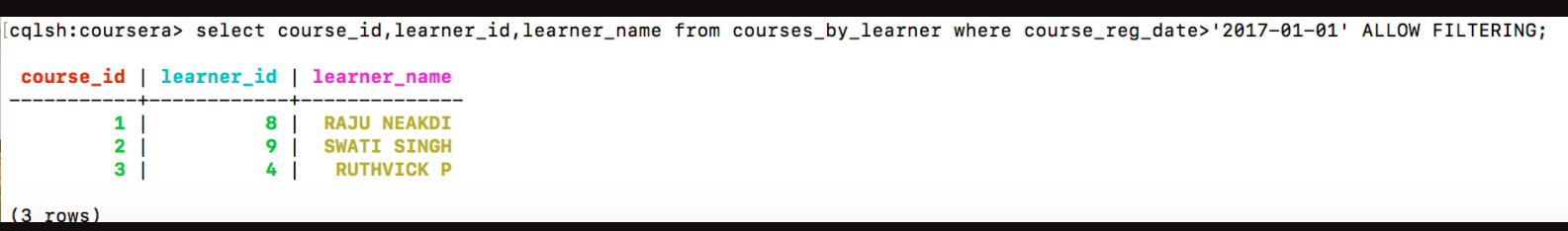


**Then insert the data in Courses\_by\_learner table**

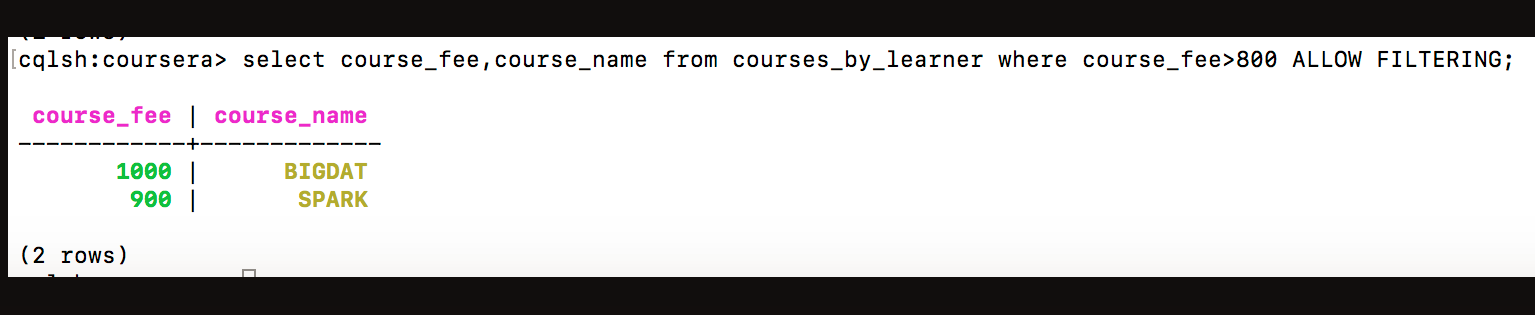


**Then we have run the three queries as shown below.**

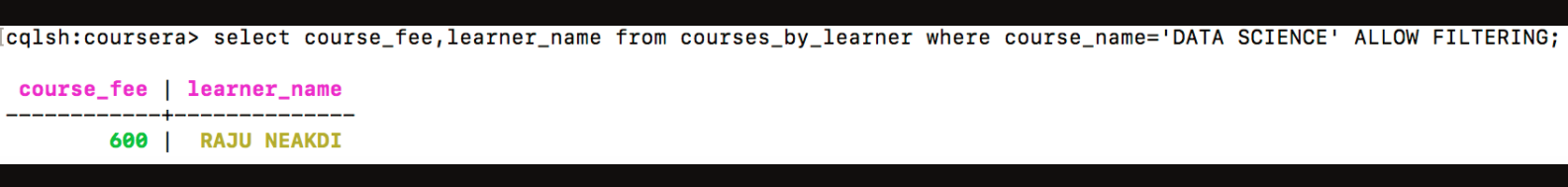
**Select course id, learner id , learner name where course reg date reater than 2017-01-01.**



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



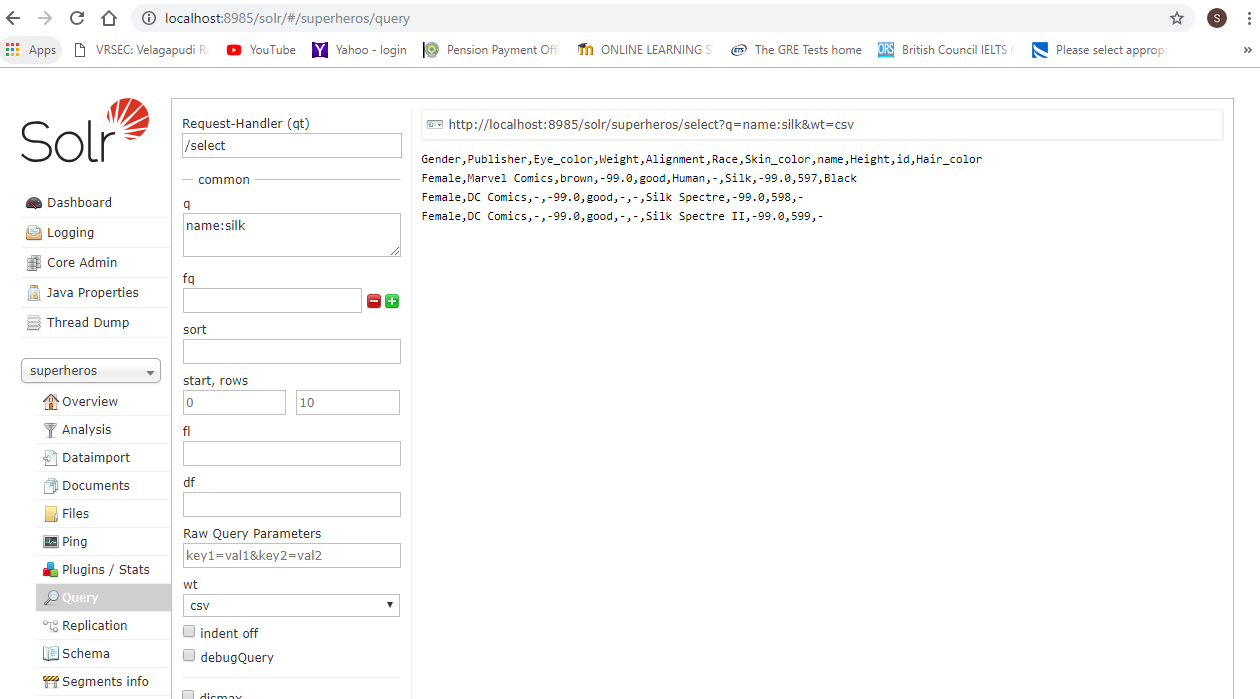
**Select course fee learner name where course name DATA SCIENCE.**



**Solr Use Case:**

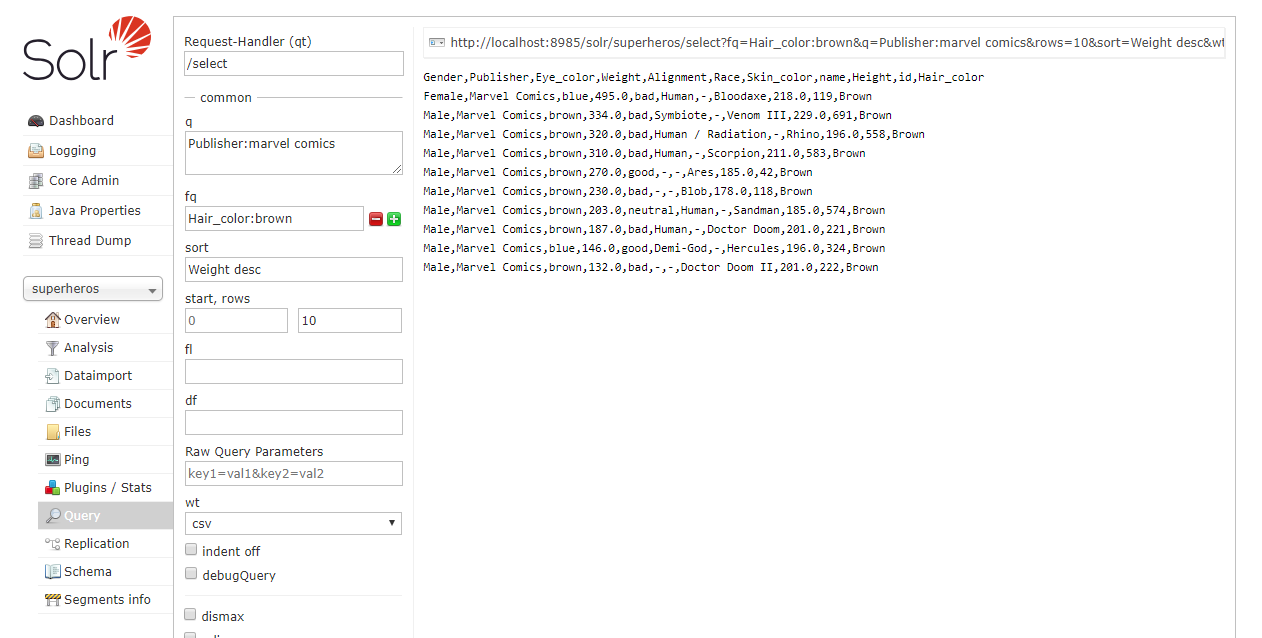
**Super Heroes Dataset:**

**Details of  superheroes with name SILK**

****

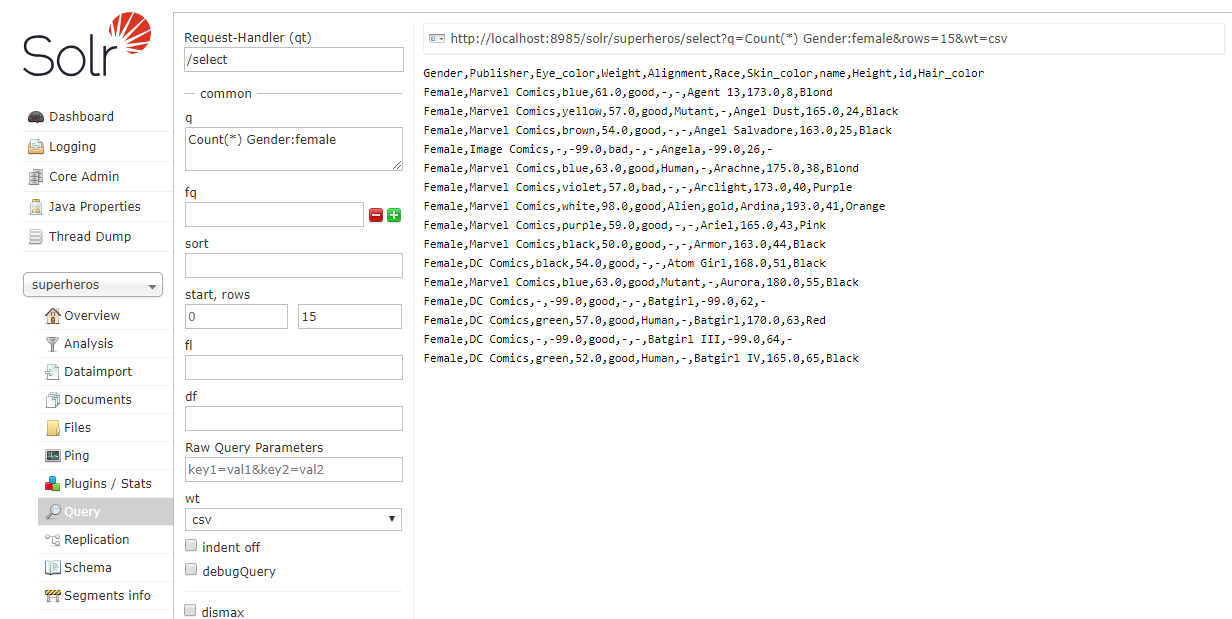
**Query Time : 2msec**

**2. Details of superheroes whose Publisher is Marvel Comics and  has brown Hair\_color and sort by Weight in descending order.**

****

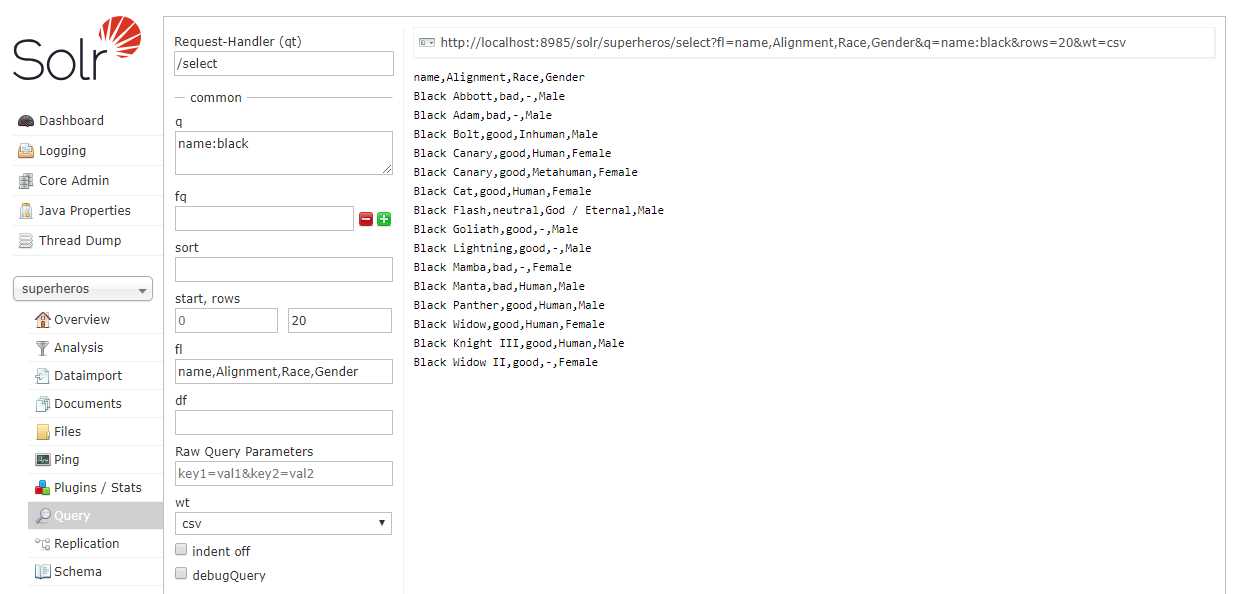
**Query Time: 1msec**

**3. Count top 15 Female Superheroes**

****

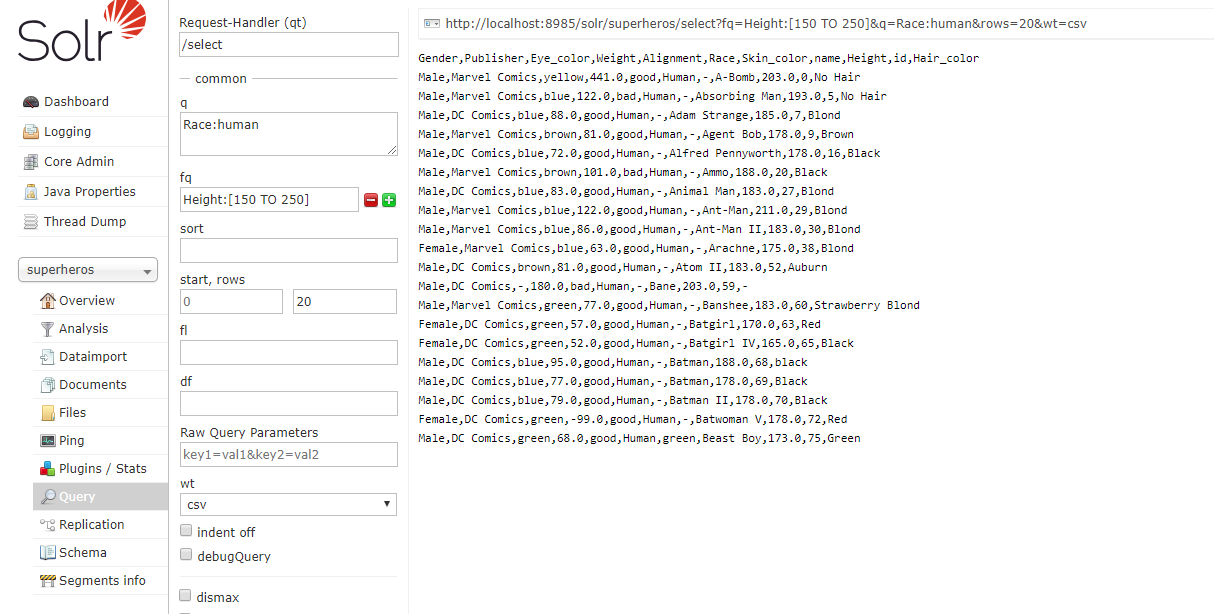
**Query Time: 3msec**

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

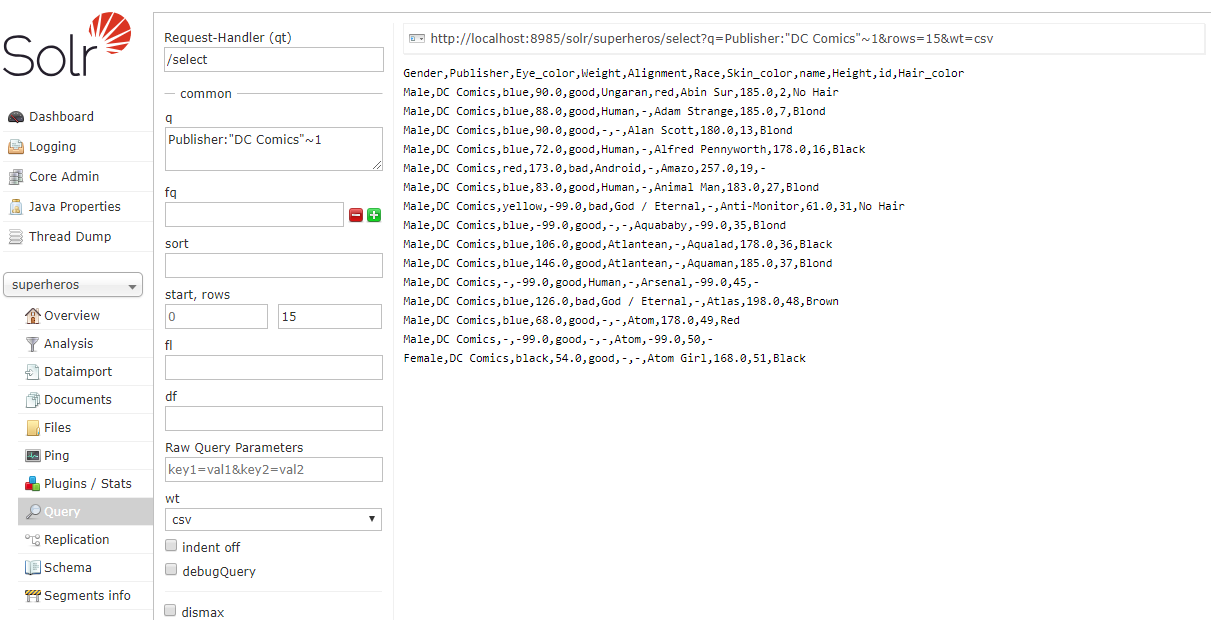
****

**Query Time : 2msec**

**5. Top 20  Superheroes whose race is ‘Human’ and Height range is 150 - 250.**

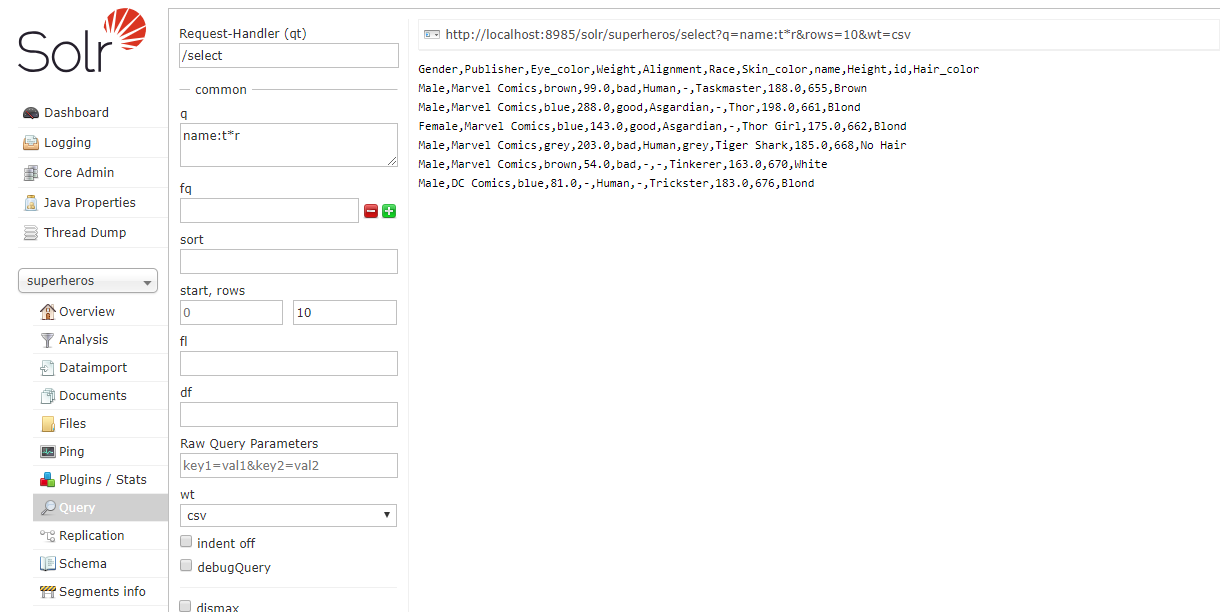
**Query Time: 2msec**

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

****

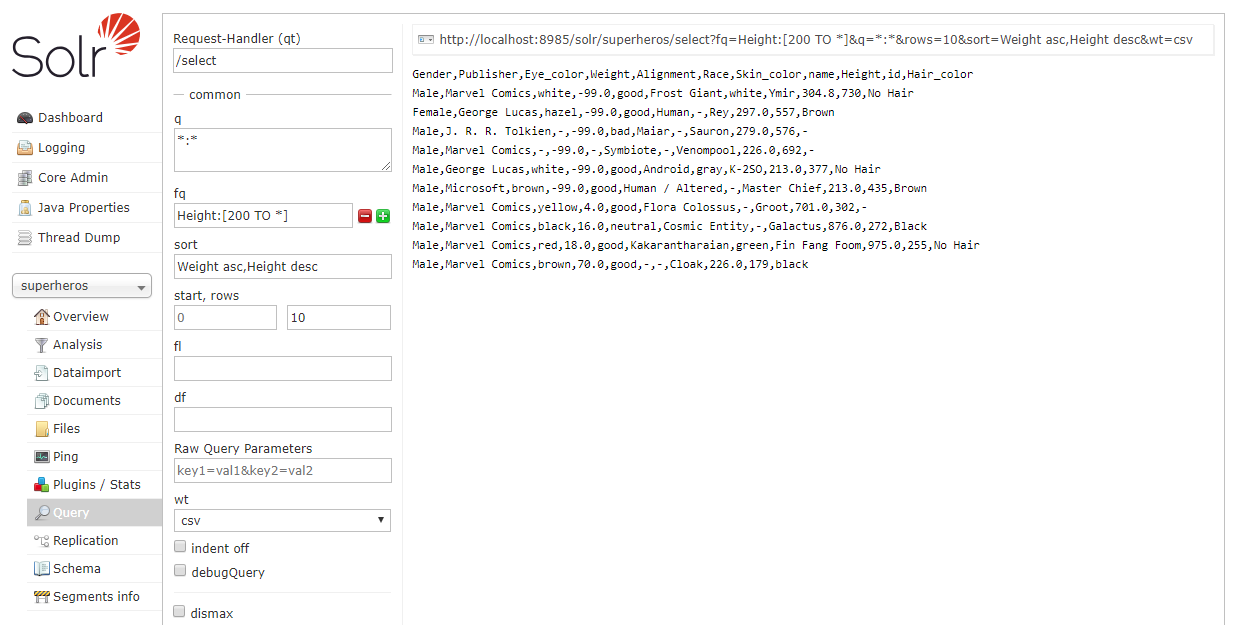
**Query Time: 61msec**

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

****

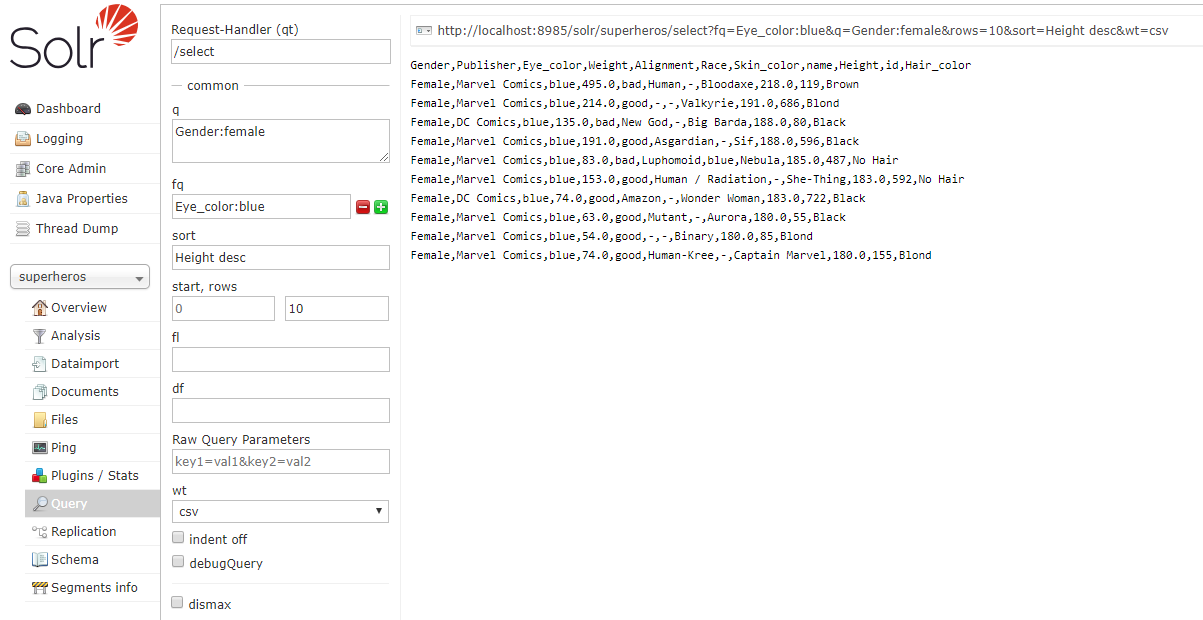
**Query Time: 5msec**

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

****

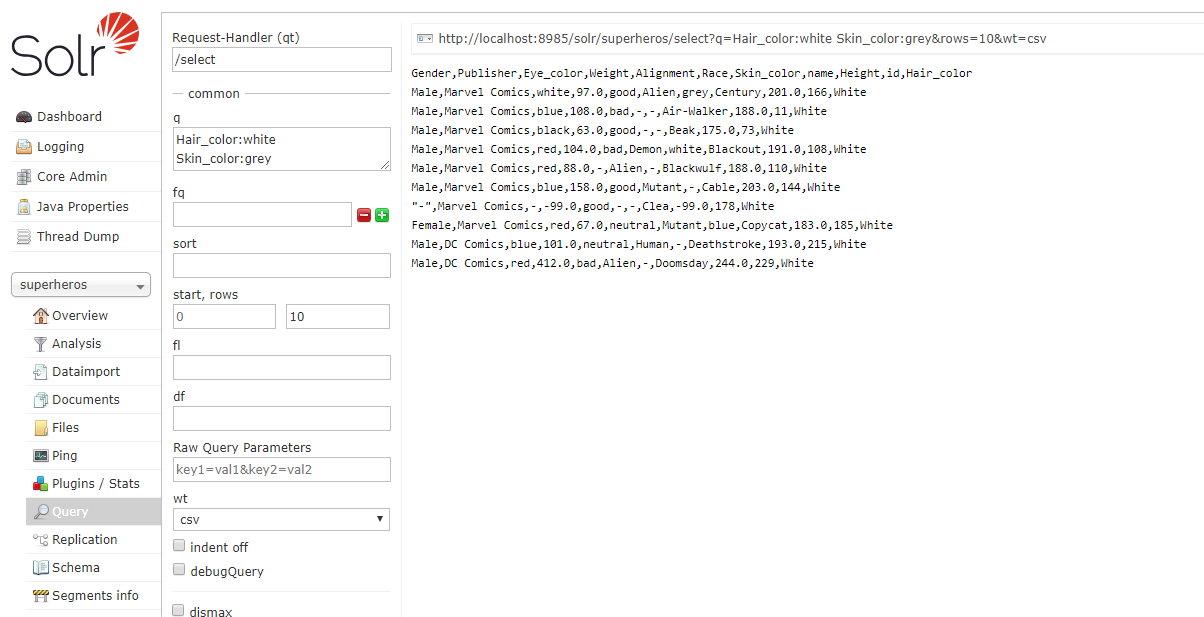
**Query Time: 1msec**

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

****

**Query Time: 3msec**

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

****

**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.