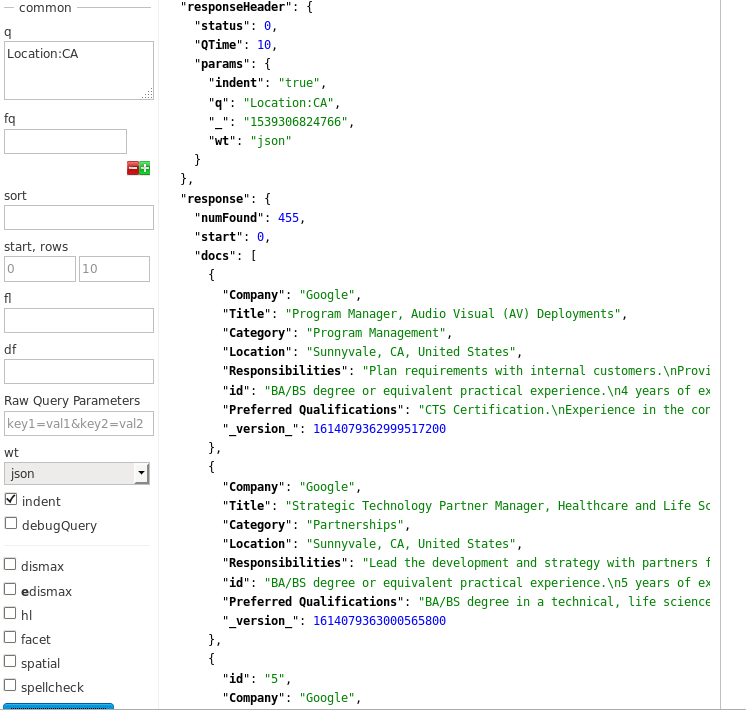
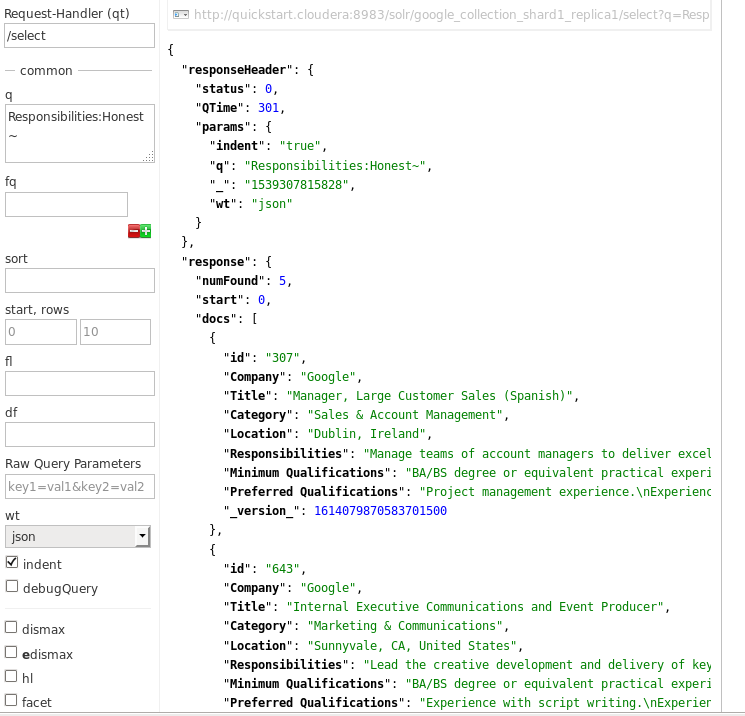
Part 1: Hive

Create external table superhero (ID int, name string, Gender string, Eye\_color string, Race string, Hair\_color string, Height double, Publisher string, Skin\_color string, Alignment string, Weight double)  
row format delimited fields terminated by ','  
tblproperties("skip.header.line.count"="1");  
  
load data local inpath '/home/cloudera/Desktop/data/heroes\_information.csv' overwrite into table superhero;  
  
--10 Intuitive Queries in HQL  
  
--1 Average weight by gender  
select Gender, avg(Weight) from superhero group by Gender;  
  
--    14.821428571428571  
--Female    27.265  
--Male    52.051587301587304  
  
  
--2 Number of superheroes from each publisher  
select Publisher, count(name) as count group by Publisher;  
  
--ABC Studios    4  
--DC Comics    215  
--Dark Horse Comics    18  
--George Lucas    14  
--Hanna-Barbera    1  
  
--3 Average height by hair color  
select Hair\_color, avg(Height) from superhero group by Hair\_color;  
  
--No Hair    148.67066666666665  
--Orange    249.0  
--Orange / White    175.0  
--Pink    165.0  
  
--4 Number of heroes by gender and skin color  
select Gender, Skin\_color, count(name) as count from superhero group by Gender, Skin\_color;  
  
--Male    -    449  
--Male    black    1  
--Male    blue    4  
--Male    blue-white    1  
  
--5 What are the races and how many heroes in each  
select Race, count(name) as count from superhero group by Race;  
  
--Mutant    63  
--Mutant / Clone    1  
--New God    3  
--Neyaphem    1  
  
--6 Number of good, bad, or neutral superheroes  
select Alignemnt, count(name) from superhero group by Alignment;  
  
--    7  
--bad    207  
--good    496  
--neutral    24  
  
--7 Average weight of good vs bad guys  
select Alignment, avg(Weight) from superhero group by Alignment;  
  
--    18.714285714285715  
--bad    65.61650485436893  
--good    31.876767676767678  
--neutral    111.45833333333333  
  
--8 Are there more female good or bad guys  
select Alignment, count(name) from superhero where Gender = 'Female' group by Alignment;  
  
--bad    35  
--good    161  
--neutral    4  
  
--9 Are there more male good or bad guys  
select Alignment, count(name) from superhero where Gender = 'Male' group by Alignment;  
  
-    6  
bad    165  
good    316  
neutral    18  
  
--10 The weight of all the superheroes combined  
select sum(Weight) from superhero;  
  
--32102.0

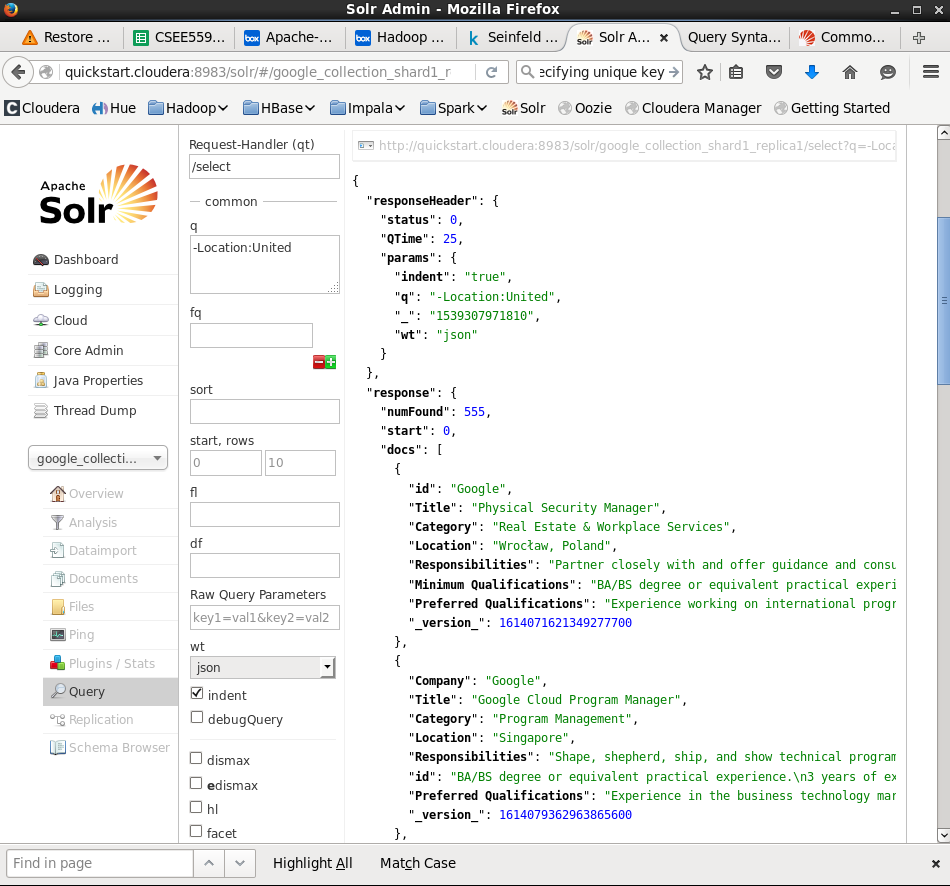
Part 2: Solr



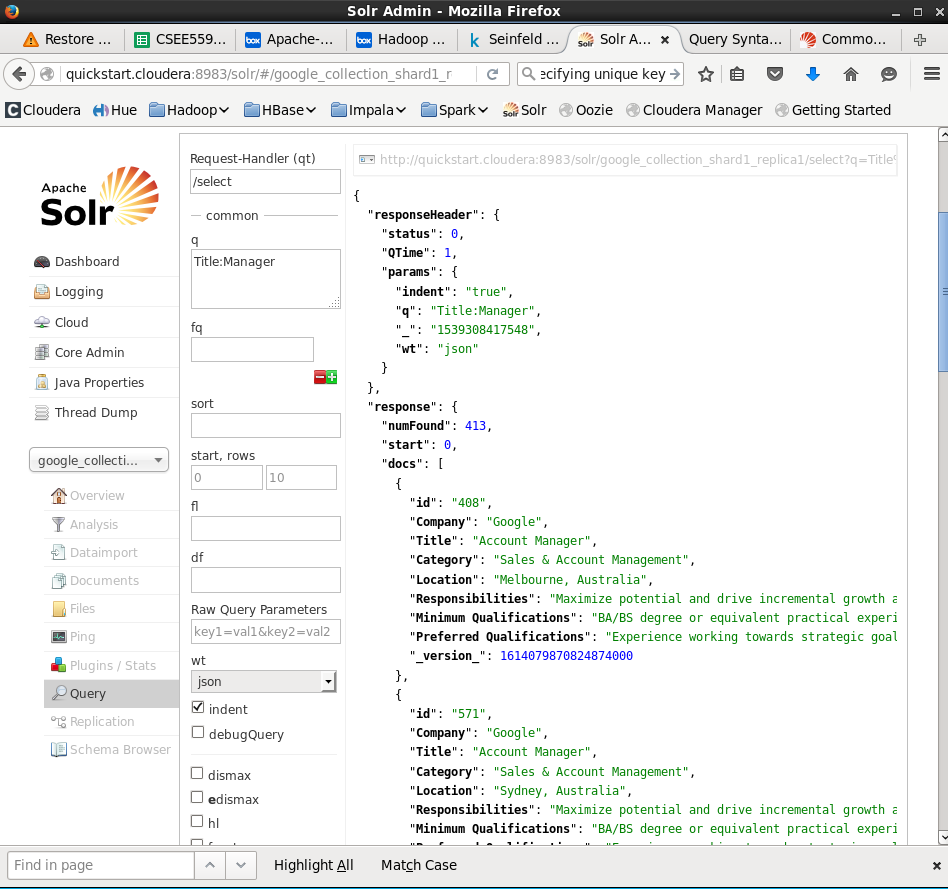
Query 1: Jobs that are in California



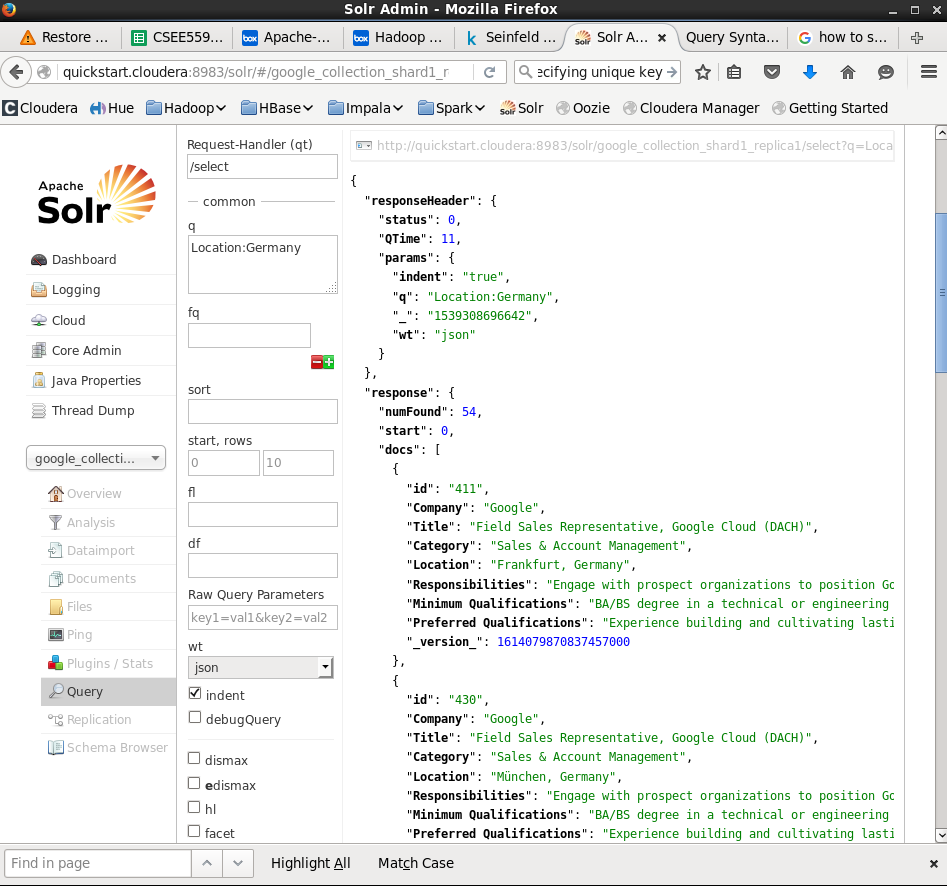
Query 2: Jobs where the Responsibilities include being honest.



Query 3: Jobs that are located in the United States.



Query 4: Jobs that are for Manager positions.



Query 5: Jobs that are located in Germany

Part 3: Cassandra

# Cassandra Use Case: Spotify

Spotify would need to have a dataset that keeps track of the number of listens for a given song, as well as some other information. Spotify has a ton of songs, so the scalability of this solution would need to be very large, which is where Cassandra comes in.

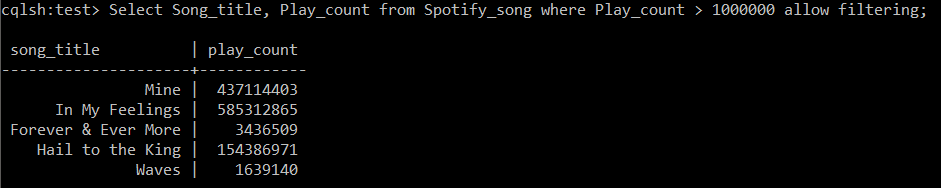
The fields would be (Song\_title text, Artist text, Album text, Play\_count integer)

As you can see, this dataset includes basic information about the song, as well as the number of listens the song has had.

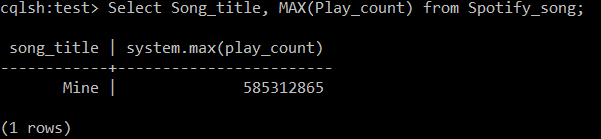
# Queries

1: Songs with over a million listens.

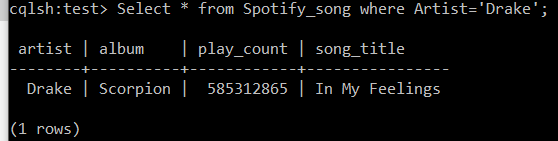
Select Song\_title, Play\_count from Spotify\_song where Play\_count > 1000000 allow filtering;



2: The most listened to song.

Select Song\_title, MAX(Play\_count) from Spotify\_song; 

3: Songs by Drake.

Select \* from Spotify\_song where Artist='Drake';