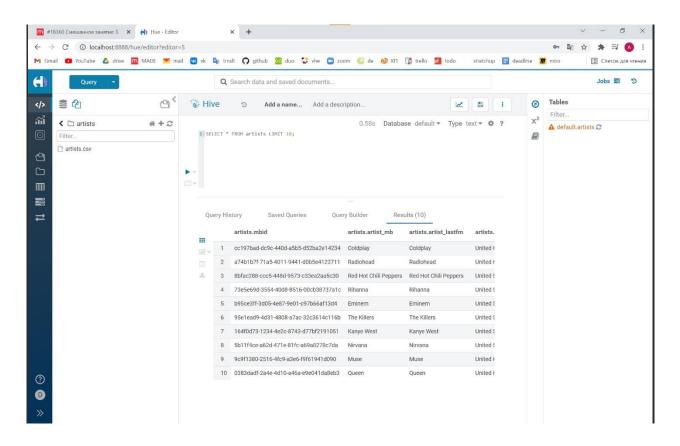
1. Сделать таблицу artists в Hive и вставить туда значения, используя датасет

```
create table artists
(mbid string,
artist_mb string,
artist_lastfm string,
country_mb string,
country_lastfm string,
tags_mb array<string>,
tags_lastfm array<string>,
listeners_lastfm double,
scrobbles_lastfm double,
ambiguous_artist boolean)
row format
delimited fields terminated by ','
collection items terminated by '; '
tblproperties ("skip.header.line.count"="1");
load data inpath '/artists.csv' overwrite into table artists;
```



- 2. HiveQL
- а) Найти исполнителя с максимальным числом скробблов

```
SELECT artists.artist_lastfm
FROM artists
WHERE artists.scrobbles_lastfm in
          (SELECT Max(artists.scrobbles_lastfm) from artists);
```



b) Найти самый популярный тэг на ластфм

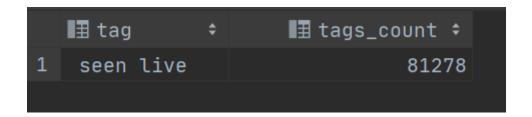
```
SELECT tag, count(*) tags_count
FROM artists

LATERAL VIEW explode(tags_lastfm) item_val AS tag

GROUP BY tag

ORDER BY tags_count DESC

LIMIT 1;
```



с) Найти самые популярные исполнители 10 самых популярных тегов ластфм

1<	< 10 rows ∨ > > S + − ^{DB} ■ DDL D	DML 🚓 🖈
	■ artist_mb ÷	■ listeners_lastfm ÷
1	Coldplay	5381567
2	Radiohead	4732528
3	Red Hot Chili Peppers	4620835
4	Rihanna	4558193
5	Eminem	4517997
6	The Killers	4428868
7	Kanye West	4390502
8	Nirvana	4272894
9	Muse	4089612
10	Queen	4023379

d) Найти самые популярные теги ластфм в самой музыкальной стране (где больше всего исполнителей)

```
SELECT tag, country_lastfm, count(*) as count

FROM artists

LATERAL VIEW explode(tags_lastfm) artists AS tag

WHERE artists.country_lastfm in

(SELECT countries.country_lastfm

FROM (SELECT country_lastfm, count(*) as count

FROM artists

WHERE length(country_lastfm) > 0

GROUP BY country_lastfm

ORDER BY count DESC

LIMIT 1) countries)

GROUP BY tag, country_lastfm

ORDER BY count DESC

LIMIT 10;
```

K	< 10 rows > >	S *	
	■ tag \$	■ country_lastfm ÷	■ count ÷
1	american	United States	17926
2	USA	United States	12347
3	seen live	United States	9603
4	rock	United States	8177
5	All	United States	7199
6	alternative	United States	5830
7	indie	United States	5219
8	pop	United States	4508
9	experimental	United States	4382
10	electronic	United States	4123