

## а) Исполнитель с максимальным числом скробблов:

```
1 SELECT artist_lastfm FROM
2 (SELECT artist_lastfm, scrobbles_lastfm FROM mytable ORDER BY scrobbles_lastfm DESC LIMIT 1) t;
3
```

WARNING: Hive-on-MR is deprecated in Hive 2 and may not be available in the future versions. Consider using a different execution engine (i.e. spark, tez) or using Hive 1.X releases.

Query History Saved Queries Query Builder Results (1)

artist\_lastfm



1 The Beatles

## б) Самый популярный тэг на ластфм

```
1 SELECT tag FROM
2 (SELECT tag, COUNT(tag) AS num_of_tag FROM mytable
3  LATERAL VIEW EXPLODE(SPLIT(LOWER(tags_lastfm), ' ')) tmpTable AS tag
4  WHERE tag != '' GROUP BY tag ORDER BY num_of_tag DESC LIMIT 1) t;
5
```

WARNING: Hive-on-MR is deprecated in Hive 2 and may not be available in the future versions. Consider using a different execution engine (i.e. spark, tez) or using Hive 1.X releases.

Query History Saved Queries Query Builder Results (1)

tag



1 seen live

## с) Самые популярные исполнители 10 самых популярных тегов ластфм

### Вариант 1 (brute force)

```
1 SELECT artist_lastfm FROM
2 (WITH tags_table AS
3   (SELECT tag, COUNT(tag) AS num_of_tag FROM mytable
4    LATERAL VIEW EXPLODE(SPLIT(LOWER(tags_lastfm), '; ')) tmpTable AS tag
5    WHERE tag != '' GROUP BY tag ORDER BY num_of_tag DESC LIMIT 10)
6 SELECT artist_lastfm, scrobbles_lastfm FROM
7   (SELECT artist_lastfm, scrobbles_lastfm, tags FROM mytable
8    LATERAL VIEW EXPLODE(SPLIT(LOWER(tags_lastfm), '; ')) tags_lastfm AS tags
9    WHERE tags != '') a
10 WHERE tags IN (SELECT tag FROM tags_table)
11 GROUP BY artist_lastfm, scrobbles_lastfm
12 ORDER BY scrobbles_lastfm DESC LIMIT 10) t;
```

```
2021-10-14 09:54:22 Dump the side table for tag. 1 with group count. 10 into file: file:/tmp/root/667a55d1-685e-435a-b5dd-447708a22316/hive_2021-10-14_09-54-12_492_8067343878755354541-1/-local-10007/HashTable-Stage-6/MapJoin-mapfile111--.hashtable
2021-10-14 09:54:23 Uploaded 1 File to: file:/tmp/root/667a55d1-685e-435a-b5dd-447708a22316/hive_2021-10-14_09-54-12_492_8067343878755354541-1/-local-10007/HashTable-Stage-6/MapJoin-mapfile111--.hashtable (536 bytes)
2021-10-14 09:54:23 End of local task; Time Taken: 0.117 sec.
```

Query History Saved Queries Query Builder Results (10)

artist_lastfm	
1	The Beatles
2	Radiohead
3	Coldplay
4	Muse
5	Arctic Monkeys
6	Pink Floyd
7	Linkin Park
8	Red Hot Chili Peppers
9	Lady Gaga
10	Metallica

## с) Самые популярные исполнители 10 самых популярных тегов ластфм

### Вариант 2 (более правильный)

```
1 WITH tags_table AS
2   (SELECT tag, COUNT(tag) AS num_of_tag FROM mytable
3    LATERAL VIEW EXPLODE(SPLIT(LOWER(tags_lastfm), ' ')) tmpTable AS tag
4    WHERE tag != '' GROUP BY tag ORDER BY num_of_tag DESC LIMIT 10),
5 result AS
6   (SELECT artist_lastfm, scrobbles_lastfm, tags FROM
7    (SELECT artist_lastfm, scrobbles_lastfm, tags FROM mytable
8     LATERAL VIEW EXPLODE(SPLIT(LOWER(tags_lastfm), ' ')) tags_lastfm AS tags
9     WHERE tags != '') a
10    WHERE tags IN (SELECT tag FROM tags_table))
11 SELECT artist_lastfm, max_scrobbles, tags
12 FROM (SELECT
13        artist_lastfm,
14        MAX(scrobbles_lastfm) OVER (PARTITION BY tags) max_scrobbles,
15        tags,
16        ROW_NUMBER() OVER (PARTITION BY tags ORDER BY scrobbles_lastfm DESC) row_num
17       FROM result) r
18 WHERE row_num = 1 ORDER BY max_scrobbles DESC;
```

WARNING: Hive-on-MR is deprecated in Hive 2 and may not be available in the future versions. Consider using a different execution engine (i.e. spark, tez) or using Hive 1.X releases.

SLF4J: Class path contains multiple SLF4J bindings.

SLF4J: Found binding in [jar:file:/opt/hive/lib/log4j-slf4j-impl-2.6.2.jar!/org/slf4j/impl/StaticLoggerBinder.class]

SLF4J: Found binding in [jar:file:/opt/hadoop-2.7.4/libexec/hadoop-common-2.7.4/lib/slf4j-log4j12.jar!/org/slf4j/impl/StaticLoggerBinder.class]

Query History

Saved Queries

Query Builder

Results (10)

	artist_lastfm	max_scrobbles	tags
1	The Beatles	517126254	rock
2	The Beatles	517126254	pop
3	Radiohead	499548797	seen live
4	Radiohead	499548797	indie
5	Radiohead	499548797	experimental
6	Radiohead	499548797	electronic
7	Radiohead	499548797	alternative
8	Lady Gaga	285469647	female vocalists
9	Ariana Grande	106673207	all
10	Joyce Manor	10993278	under 2000 listeners

d) Любой другой инсайт на ваше усмотрение:

### Исполнитель с максимальным числом скробблов (pandas-style)

```
1 SELECT artist_lastfm FROM mytable WHERE scrobbles_lastfm IN (SELECT MAX(scrobbles_lastfm) FROM mytable);
```

```
2021-10-15 14:14:12      Uploaded 1 File to: file:/tmp/root/667a55d1-685e-435a-b5dd-447708a22316/hive_2021-10-15_14-14-06_874_7162195166456177097-92/-local-10005/HashTable-Stage-3/MapJoin-mapfile631--.hashtable (282 bytes)
2021-10-15 14:14:12      End of local task; Time Taken: 0.124 sec.
```

Query History

Saved Queries

Query Builder

Results (1)

artist\_lastfm



1	The Beatles
---	-------------

### ТОП-5 исполнителей по числу скробблов из Ирландии:

```
1 SELECT artist_lastfm FROM
2 (SELECT artist_lastfm, scrobbles_lastfm FROM mytable WHERE country_lastfm = 'Ireland'
3 ORDER BY scrobbles_lastfm DESC LIMIT 5) t;
```

WARNING: Hive-on-MR is deprecated in Hive 2 and may not be available in the future versions. Consider using a different execution engine (i.e. spark, tez) or using Hive 1.X releases.

Query History

Saved Queries

Query Builder

Results (5)

artist\_lastfm



1	U2
2	The Cranberries
3	Enya
4	God Is An Astronaut
5	The Script