Запрос 1:

Выбрать те строки таблицы поста, которые являются вопросами (PostTypeId==1) и просмотры не меньше 50 тыс.

```
create table  liya.filter_post
        as select YEAR(CreationDate) as year,
          ViewCount/10000+Score/100+AnswerCount/20  as result ,
             Tags
         from default.posts
          where(PostTypeId==1 and ViewCount>50000);
```

MapReduce CPU Time Spent: 6 minutes 26 seconds 910 msec

Запрос 2:

Сплит по тегам

```
create table tag_split
          as select year, result, tag from filter_post
          LATERAL VIEW explode
                (split(substr(Tags,5,length(Tags)-8),"&gt;&lt;") ) my_Table as tag   ;
```

MapReduce CPU Time Spent: 15 seconds 290 msec

Запрос 3:

Суммировать result по годам и тегам

```
create table sum_by_year_and_tag
                as select year, tag , sum(result)  result f
                rom tag_split  group by year, tag;
```

Total MapReduce CPU Time Spent: 21 seconds 540 msec

Запрос 4:

Отсеиваем те теги, которые не вошли 2020

1)создаем список, которые вошли в 2020 год.

```
create table tag2020
                  as select tag from sum_by_year_and_tag
                  where year=2020;
```

MapReduce Total cumulative CPU time: 8 seconds 450 msec

2) Сверяемся с списком тегов за 2020 год.

```
create table filter_by_list
                as select year , s.tag, result
                from sum_by_year_and_tag as s
                inner join tag2020 as t on t.tag=s.tag;
```

Total MapReduce CPU Time Spent: 10 seconds 730 msec

Запрос 5:

По оставшимся данным пытаюсь построить функцию

$$f(year) = a * (year - 2008) + b\ ,$$

где ошибка $L = \sum_{(year,res)} (f(year) - res)^2 \to \min_{a,b}$

Получаем систему (5.1)

$$a \sum_{year} (year - 2008)^2 + b \sum_{year} (year - 2008) = \sum_{year} (year - 2008) * res$$

$$a \sum_{year} (year - 2008) + b \sum_{year} 1 = \sum_{year} res$$

Обозначим $A = \sum_{year} (year - 2008)^2$, $B = \sum_{year} (year - 2008)$,

$C = \sum_{year} res * (year - 2008)$, $D = \sum_{year} 1$, $E = \sum_{year} res.$

Тогда система (5.1) перепишется в (5.2)

$$a * A + b * B = C$$

$$a * B + b * D = E$$

Следующий запрос связан с вычислением коэффициентов A,B,C,D,E.

```
create table   coef_by_tag
       as select tag,
          sum ((year-2008)*(year-2008)) as A,
```

```
        sum (year-2008) as B,
        sum ((year-2008)*result) as C,
        count(*) as D,
       sum(result ) as E
    from  filter_by_list
     group by tag
```

MapReduce Total cumulative CPU time: 15 seconds 70 msec

Запрос 6:

Решение системы (5.2)

$$a = \frac{CD - BE}{AD - B^2} \quad \text{и } b = \frac{AE - BC}{AD - B^2}$$

Решение будет корректно при D>1.

Поэтому следующий запрос таким будет

```
create table tag_trand
        as select tag, (C*D-E*B)/(A*D-B*B) as a,
                        (E*A-B*C)/(D*A-B*B) as b
        from coef_by_tag where D>1  ;
```

Total MapReduce CPU Time Spent: 6 seconds 370 msec

Запрос 7:

Так как требуют определить топ 10 трендовых технологии, то я смотрю на результаты функции $f(2021)$ и упорядочиваю по убыванию

```
create table top10
          as select  tag,
          a*13+b as res
         from tag_trand
   order by res desc limit 10;
```

Total MapReduce CPU Time Spent: 13 seconds 540 msec

Результаты после запроса 7

angular	2145.5760821428494
typescript	996.9441111111091
reactjs	881.1714000000006
android-studio	868.4499939393937
android	785.672107692304
ios	721.4288076923094
python-3.x	699.7711615384621
xcode	380.3598423076928
jenkins	289.3589342857143
pip	263.34207121212137

Total MapReduce CPU Time Spent: 7 minutes 47 seconds 170 msec