Korszerű adatbázisok - twitter lekérdezések

0. kellenek a tweet_id-ik

SELECT run_id, tweet_id, time FROM tweet_hour into mydb.korszeru0
WHERE time = '2012-12-24 00:00' OR time = '2012-12-25 00:00'

2004 283014124387065856 12/24/2012 12:00:00 AM
2004 283376514765029376 12/25/2012 12:00:00 AM

1. Hány tweet volt december 24-én?

select count(*) into mydb.korszeru1 from tweet where tweet_id>='283014124387065856' and tweet_id<'283376514765029376'

5111389

2. Hányban szerepet az XMAS szó ezek közül?

select count(*) into mydb.korszeru2 from tweet where run_id=2004 and tweet_id between 283014124387065856 and 283376514765029376 and text like '%XMAS%'

28281

3. Hányban szerepelt a http://www.youtube.com/watch?v=z8Vfp48laS8 ?

select count(*) into mydb.korszeru3 from tweet_url
where run_id=2004 and tweet_id between 283014124387065856 and 283376514765029376 and expanded_url='http://www.youtube.com/watch?v=z8Vfp48laS8'

3

4. Hány magyar nyelvű tweet volt?

select count(*) into mydb.korszeru4 from tweet where run_id=2004 and tweet_id between 283014124387065856 and 283376514765029376 and lang = 'hu'

938

5. Melyik tweetet retweetelték a legtöbbször aznap?

select top 1 retweeted_tweet_id from (
select retweeted_tweet_id,count(*) c from tweet_retweet where run_id=2004 and retweeted_tweet_id between 283014124387065856 and 283376514765029376 group by retweeted_tweet_id
) t into mydb.korszeru5 order by t.c

283094491844452352

6. Hányan retweeteltek aznap?

select count(distinct user_id) from tweet_retweet into mydb.korszeru6 where run_id=2004 and tweet_id between 283014124387065856 and 283376514765029376

808510

7. (Ki,kit) retweetelt gráfnak hány csúcsa, hány éle van?

Csúcs:

select count(distinct user_id)+count(distinct retweeted_user_id) from tweet_retweet
into mydb.korszeru7csucs
where run_id=2004 and ((tweet_id between 283014124387065856 and 283376514765029376) or (retweeted_tweet_id between 283014124387065856 and 283376514765029376))

1444845

Él:

select count(*) from (
select user_id, retweeted_user_id from tweet_retweet
where run_id=2004 and ((tweet_id between 283014124387065856 and 283376514765029376) or (retweeted_tweet_id between 283014124387065856 and 283376514765029376))
group by user_id, retweeted_user_id) t
into mydb.korszeru7el

1025096

Futási idők:
0. 0:0:0
1. 0:2:30
2. 0:0:57
3. 0:2:11
4. 0:0:2
5. 0:0:39
6. 0:0:5
7.
   a. 0:0:6
   b. 0:1:13