```
drop table if exists raw_tweets;
drop table if exists tweets;
drop table if exists positive_hashtags_per_day;
drop table if exists count_positive_hashtags_per_day;
drop table if exists top5_positive_hashtags_per_day;

create table raw_tweets (json string);
load data local inpath 'sample.json' into table raw_tweets;

create table tweets as
    select get_json_object(json, "$.text") as text,
        unix_timestamp(get_json_object(json, "$.created_at"),
            "EEE MMM d HH:mm:ss Z yyyy") as ts_created
    from raw_tweets;

create table positive_hashtags_per_day as
    select from_unixtime(ts_created, 'yyyy-MM-dd') as dt,
        lower(hashtag) as hashtag from tweets
            lateral view explode(split(text, ' ')) b as hashtag
        where ts_created is not null
            and hashtag rlike "^#[a-zA-Z0-9]+$"
            and text rlike "^.*[\;:]-?\\).*$";

create table count_positive_hashtags_per_day as
    select dt, hashtag, count(*) as cnt from positive_hashtags_per_day
        group by dt, hashtag;

add file topN.py;
create table top5_positive_hashtags_per_day as
    reduce dt, hashtag, cnt
    using 'topN.py 5' as dt, hashtag, cnt
    from
    (select dt, hashtag, cnt from count_positive_hashtags_per_day
        distribute by dt sort by dt, cnt desc) cnts;

select * from top5_positive_hashtags_per_day;
```