

Udiddit, a social news aggregator

Introduction

Udiddit, a social news aggregation, web content rating, and discussion website, is currently using a risky and unreliable Postgres database schema to store the forum posts, discussions, and votes made by their users about different topics.

The schema allows posts to be created by registered users on certain topics, and can include a URL or a text content. It also allows registered users to cast an upvote (like) or downvote (dislike) for any forum post that has been created. In addition to this, the schema also allows registered users to add comments on posts.

Here is the DDL used to create the schema:

```
CREATE TABLE bad_posts (  
    id SERIAL PRIMARY KEY,  
    topic VARCHAR(50),  
    username VARCHAR(50),  
    title VARCHAR(150),  
    url VARCHAR(4000) DEFAULT NULL,  
    text_content TEXT DEFAULT NULL,  
    upvotes TEXT,  
    downvotes TEXT  
);  
  
CREATE TABLE bad_comments (  
    id SERIAL PRIMARY KEY,  
    username VARCHAR(50),  
    post_id BIGINT,  
    text_content TEXT  
);
```

Part I: Investigate the existing schema

As a first step, investigate this schema and some of the sample data in the project's SQL workspace. Then, in your own words, outline three (3) specific things that could be improved about this schema. Don't hesitate to outline more if you want to stand out!

1. In bad_post table
 - Upvote should be number
 - Downvote should be number
2. In bad_comment table
 - Username should be not null
 - Post_id should be not null
3. Schema missing some table for specifically queries: users , topics ...
4. Need some index for quickly query

Part II: Create the DDL for your new schema

Having done this initial investigation and assessment, your next goal is to dive deep into the heart of the problem and create a new schema for Udiddit. Your new schema should at least reflect fixes to the shortcomings you pointed to in the previous exercise. To help you create the new schema, a few guidelines are provided to you:

1. Guideline #1: here is a list of features and specifications that Udiddit needs in order to support its website and administrative interface:
 - a. Allow new users to register:
 - i. Each username has to be unique
 - ii. Usernames can be composed of at most 25 characters
 - iii. Usernames can't be empty
 - iv. We won't worry about user passwords for this project
 - b. Allow registered users to create new topics:
 - i. Topic names have to be unique.
 - ii. The topic's name is at most 30 characters
 - iii. The topic's name can't be empty
 - iv. Topics can have an optional description of at most 500 characters.
 - c. Allow registered users to create new posts on existing topics:
 - i. Posts have a required title of at most 100 characters
 - ii. The title of a post can't be empty.
 - iii. Posts should contain either a URL or a text content, **but not both**.
 - iv. If a topic gets deleted, all the posts associated with it should be automatically deleted too.
 - v. If the user who created the post gets deleted, then the post will remain, but it will become dissociated from that user.
 - d. Allow registered users to comment on existing posts:
 - i. A comment's text content can't be empty.
 - ii. Contrary to the current linear comments, the new structure should allow comment threads at arbitrary levels.
 - iii. If a post gets deleted, all comments associated with it should be automatically deleted too.
 - iv. If the user who created the comment gets deleted, then the comment will remain, but it will become dissociated from that user.
 - v. If a comment gets deleted, then all its descendants in the thread structure should be automatically deleted too.

- e. Make sure that a given user can only vote once on a given post:
 - i. Hint: you can store the (up/down) value of the vote as the values 1 and -1 respectively.
 - ii. If the user who cast a vote gets deleted, then all their votes will remain, but will become dissociated from the user.
 - iii. If a post gets deleted, then all the votes for that post should be automatically deleted too.
- 2. Guideline #2: here is a list of queries that Udiddit needs in order to support its website and administrative interface. Note that you don't need to produce the DQL for those queries: they are only provided to guide the design of your new database schema.
 - a. List all users who haven't logged in in the last year.
 - b. List all users who haven't created any post.
 - c. Find a user by their username.
 - d. List all topics that don't have any posts.
 - e. Find a topic by its name.
 - f. List the latest 20 posts for a given topic.
 - g. List the latest 20 posts made by a given user.
 - h. Find all posts that link to a specific URL, for moderation purposes.
 - i. List all the top-level comments (those that don't have a parent comment) for a given post.
 - j. List all the direct children of a parent comment.
 - k. List the latest 20 comments made by a given user.
 - l. Compute the score of a post, defined as the difference between the number of upvotes and the number of downvotes
- 3. Guideline #3: you'll need to use normalization, various constraints, as well as indexes in your new database schema. You should use named constraints and indexes to make your schema cleaner.
- 4. Guideline #4: your new database schema will be composed of five (5) tables that should have an auto-incrementing id as their primary key.

Once you've taken the time to think about your new schema, write the DDL for it in the space provided here:

```

create table if not exists users
(
    user_id      serial              not null
        constraint users_pk primary key,
    username     varchar(25) unique          not null,
    last_login   timestamp default current_timestamp not null,
    constraint check_username
        check (length(trim(users.username)) > 0)
);

create table if not exists topics
(
    topic_id      serial              not null
        constraint topics_pk primary key,
    topic_name    varchar(30) unique          not null
        constraint check_topic_name
            check (length(trim(topics.topic_name)) > 0),
    topic_description varchar(500)
);

create table if not exists posts
(
    post_id      serial              not null
        constraint posts_pk
            primary key,
    user_id      integer
        constraint posts_users_user_id_fk
            references users
            on delete set null,
    topic_id     integer              not null
        constraint posts_topics_topic_id_fk
            references topics
            on delete cascade,
    title        varchar(100) not null,
    content      text,
    url          text,
    created_timestamp timestamp default current_timestamp not null,
    constraint check_content
        check ((url is not null and content is null) or (content is not null
and url is null)),
    constraint check_title
        check (length(trim(title)) > 0)
);

create index posts_url_index
    on posts (url);

create table if not exists comments
(
    comment_id    serial              not null
        constraint comments_pk
            primary key,

```

```

        content          text                      not null,
        parent_id        integer
            constraint comments_comments_comment_id_fk
                references comments
                on delete cascade,
        post_id          integer                      not null
            constraint comments_posts_post_id_fk
                references posts
                on delete cascade,
        user_id          integer
            constraint comments_users_user_id_fk
                references users
                on delete set null,
        created_timestamp timestamp default current_timestamp not null
    );

create table if not exists votes
(
    vote_id serial
        constraint votes_pk
            primary key,
    vote smallint not null
        constraint check_name
            check ((vote = 1) OR (vote = '-1'::integer)),
    user_id integer
        constraint votes_users_user_id_fk
            references users,
    post_id integer not null
        constraint votes_posts_post_id_fk
            references posts,
    constraint votes_unique
        unique (user_id, post_id)
);

create index votes_post_id_index
    on votes (post_id);
create index comments_post_id_index
    on comments (post_id);

-- i. List all the top-level comments (those that don't have a parent
comment) for a given post.
SELECT * FROM comments WHERE post_id = 26890 AND comments.parent_id IS NULL;

--Compute the score of a post, defined as the difference between the number
of upvotes and the number of downvotes
SELECT post_id, SUM(CASE WHEN vote = 1 THEN 1 ELSE -1 END) AS score
FROM votes WHERE post_id = <post_id> GROUP BY post_id;

```

Part III: Migrate the provided data

Now that your new schema is created, it's time to migrate the data from the provided schema in the project's SQL Workspace to your own schema. This will allow you to review some DML and DQL concepts, as you'll be using INSERT...SELECT queries to do so. Here are a few guidelines to help you in this process:

1. Topic descriptions can all be empty
2. Since the bad_comments table doesn't have the threading feature, you can migrate all comments as top-level comments, i.e. without a parent
3. You can use the Postgres string function **regexp_split_to_table** to unwind the comma-separated votes values into separate rows
4. Don't forget that some users only vote or comment, and haven't created any posts. You'll have to create those users too.
5. The order of your migrations matter! For example, since posts depend on users and topics, you'll have to migrate the latter first.
6. Tip: You can start by running only SELECTs to fine-tune your queries, and use a LIMIT to avoid large data sets. Once you know you have the correct query, you can then run your full INSERT...SELECT query.
7. **NOTE:** The data in your SQL Workspace contains thousands of posts and comments. The DML queries may take at least 10-15 seconds to run.

Write the DML to migrate the current data in bad_posts and bad_comments to your new database schema:

```
insert into users(username)
  select username
    from bad_comments
 union
  select username
  from bad_posts
 union
  select regexp_split_to_table(upvotes, ',') as username
  from bad_posts
 union
  select regexp_split_to_table(downvotes, ',') as username
  from bad_posts;

insert into topics(topic_name)
select distinct topic
from bad_posts;
```

```

insert into posts (title, url, content, topic_id, user_id)
select left(bp.title, 100), bp.url, bp.text_content, t.topic_id, u.user_id
from bad_posts as bp
    join topics as t on bp.topic = t.topic_name
    join users as u on bp.username = u.username;

insert into comments(content, post_id, user_id)
select bc.text_content, p.post_id, u.user_id
from bad_comments as bc
    join bad_posts as bp on bc.post_id = bp.id
    join posts as p on p.title = left(bp.title, 100)
    left join users as u on bc.username = u.username;

insert into votes (vote, user_id, post_id)
select vote, u.user_id, p.post_id
from (select distinct title as title, regexp_split_to_table(upvotes, ',') as
username, 1 as vote
    from bad_posts) as bp
    join posts as p on left(bp.title, 100) = p.title
    join users as u on u.username = bp.username;

insert into votes (vote, user_id, post_id)
select vote, u.user_id, p.post_id
from (select title as title, regexp_split_to_table(downvotes, ',') as
username, -1 as vote
    from bad_posts) as bp
    join posts as p on left(bp.title, 100) = p.title
    join users as u on u.username = bp.username;

```