

Stack Exchange Data Analysis

About this Exercise

Stack Exchange Is a network of sites, allowing different communities learn and share their knowledge.

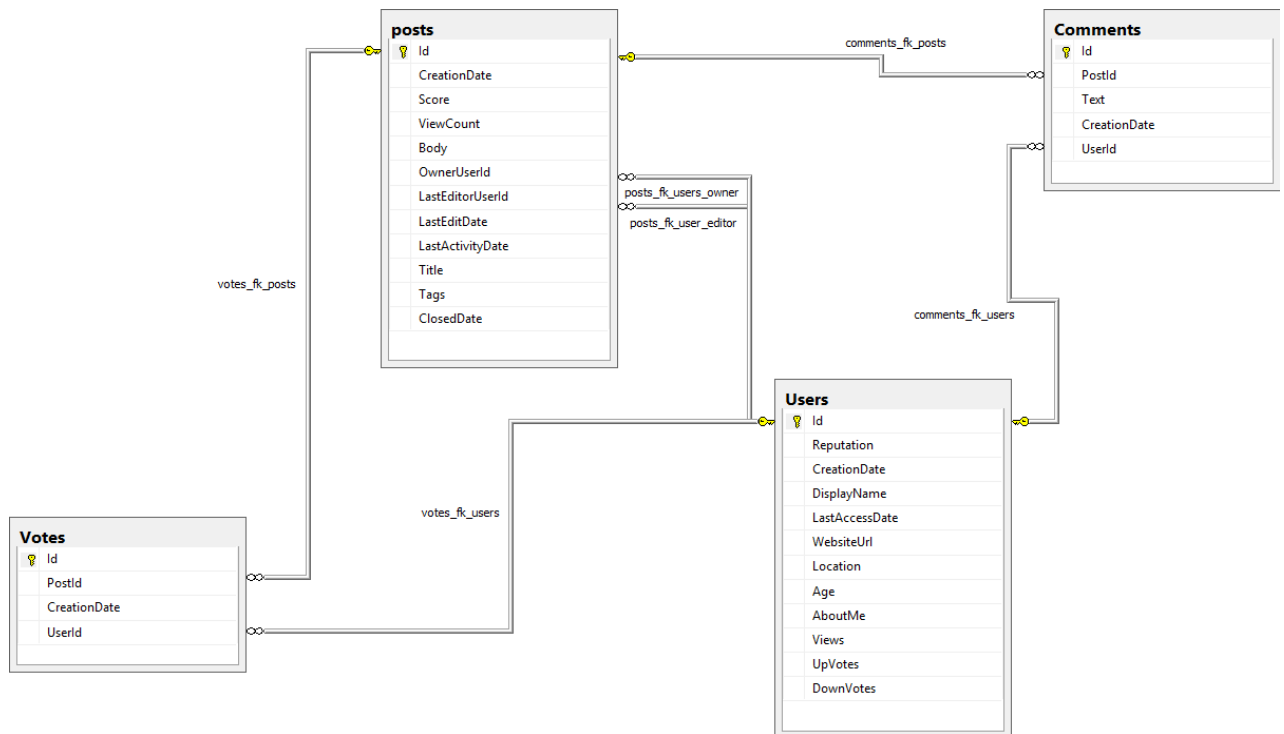
Movies & TV Stack Exchange – is one of these sites, focused entirely on questions and answers regarding movies and series.

The following database is built according to its structure, and simulates some of the data it has.

Diagram

In [15]:

Out[15]:



ERD General Description

- **Posts** – Each post can have different comments, each post can be voted by different users (in order to improve its ranking)
- **Votes** – Each vote consists of a voting ID, user number (FK), post number (FK), and a date of creation
- **Comments** – The details regarding the various posts comments

Exercises

The following exercises are constructed in ascending difficulty order. First ones are pretty easy, made for warm-up.

The advanced exercises, on the other hand, can be quite challenging.

Basic Analysis

1. How many post were made each year ?
2. How many votes were made in each day of the week (Sunday, Monday, Tuesday, etc.) ?
3. List all comments created on September 19th, 2012
4. List all users under the age of 33, living in London

Advanced Analysis

1. Display the number of votes for each post title
2. Display posts with comments created by users living in the same location as the post creator
3. How many users have never voted ?
4. Display all posts having the highest amount of comments
5. For each post, how many votes are coming from users living in Canada ? What's their percentage of the total number of votes
6. How many hours in average, it takes to the first comment to be posted after a creation of a new post
7. Whats the most common post tag ?
8. Create a pivot table displaying how many posts were created for each year (Y axis) and each month (X axis)