

Coding Challenge: Social Media Data Analysis

Scenario

You are working as a **Database Analyst** for a new social media platform called **ConnectHub**. The company stores user data, posts, likes, friendships, and comments in a any database. Your task is to **analyze and query the database** to derive insights about user activity and relationships.

The challenge will test your ability to use **subqueries, joins, and stored procedures** in SQL.

Database Schema

1. Users

Column	Type	Description
user_id	INT (PK)	Unique ID of the user
username	VARCHAR	User's profile name
email	VARCHAR	User's email address
join_date	DATE	Date user joined

2. Posts

Column	Type	Description
post_id	INT (PK)	Unique ID of the post
user_id	INT (FK)	Author of the post
content	TEXT	Post text/content
post_date	DATETIME	Date and time of posting

3. Likes

Column	Type	Description
like_id	INT (PK)	Unique like ID

Column	Type	Description
user_id	INT (FK)	User who liked
post_id	INT (FK)	Post that was liked
like_date	DATETIME	Date and time of like

4. Comments

Column	Type	Description
comment_id	INT (PK)	Unique comment ID
post_id	INT (FK)	Post commented on
user_id	INT (FK)	Comment author
comment_text	TEXT	Comment content
comment_date	DATETIME	Date and time of comment

5. Friendships

Column	Type	Description
friendship_id	INT (PK)	Unique ID
user_id1	INT (FK)	User 1
user_id2	INT (FK)	User 2
since_date	DATE	Date friendship started

Tasks

Task 1:

1. Retrieve all posts along with the username of the author.
2. Find all comments on each post along with the commenter's username.

Task 2:

3. Find the top 3 users with the most posts.
4. Retrieve posts that have more likes than the **average number of likes per post**.
5. Find users who have never posted anything but have liked posts.

Task 3:

6. Get a list of all friends of a specific user (say user_id = 3).
7. Retrieve posts that were liked by **friends of a given user** (nested join scenario).

Task 4:

8. Create a stored procedure GetUserActivity(userId) that returns:
 - Total posts by the user
 - Total likes given by the user
 - Total likes received on the user's posts
 - Total comments made by the user

Task 5: Challenge Query

9. Find the **most influential user** (the user whose posts have the highest total likes + comments).

Expected Outcomes

- Demonstrate **understanding of JOINS** (INNER, LEFT, SELF JOIN), **subqueries (scalar, correlated, nested)**, and **stored procedures**.
- They should be able to **interpret real-world problems** into SQL solutions.
- Can optimize queries using **indexes and group by/having clauses**.