INSY 437 - MANAGING DATA & DATABASES

DATABASE DESIGN FOR A SOCIAL NETWORKING PLATFORM



Kaibo Zhang: 261110409
Malo Thomassin: 260952200
George Gueorguiev: 261052000
Sofia Karim: 260945464
Aissatou Fanny: 261009120



TABLE OF CONTENTS

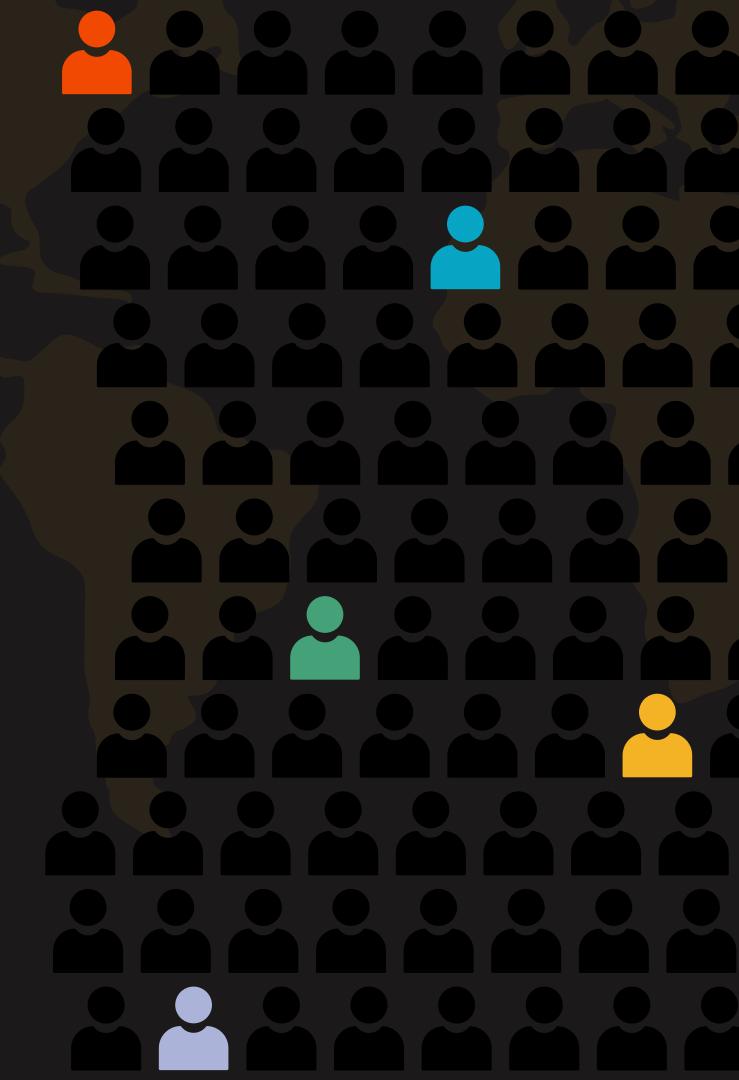
- 1. Overview of our Business Scenario
- 2. Mission Statement & Objective
- 3. ERD
- 4. Relational Schema Overview
- 5. Queries
- 6. Learning Experience & Takeaways

BUSINESS OVERVIEW

LockedIn is committed to revolutionizing professional networking, providing a clutter free platform designed to facilitate meaningful connections and career advancement.

Core Values:

- 1. **User-Centric Design:** Crafting an intuitive platform tailored to professionals' needs.
- 2. **Meaningful Connections**: Fostering genuine connections crucial for career growth.
- 3. Seamless Experience: Streamlined functionalities for a hassle-free networking journey.



MISSION STATEMENT

"The purpose of the envisioned database system for LockedIn is to maintain the data that is used to facilitate connections, foster professional growth, and provide valuable insights into various industries and job markets. LockedIn is geared toward professionals seeking networking made simple and vouches to provide a unified and undistracted user experience to its clients."

The platform offers a **focused** networking experience, devoid of *unnecessary*, *distracting* features found on other social networking platforms. Our scope delineates what LockedIn has to offer...



PROFESSIONAL PROFILES

CONTENT CREATION AND SHARING

THE SCOPE



With LockedIn, distractions are removed.

Stay Locked In, Stay Ahead.

ACTIVE BUSINESS

PROFILES

NETWORKING TOOLS

COMMENTS / REACTIONS

B2B SALES TOOLS

JOB SEARCH TOOLS

RECRUITMENT TOOLS

DIRECT MESSAGING



MISSION OBJECTIVES P.1

MAINTAIN

To maintain (enter, update, and delete) data on users, user backgrounds, user skills, and user activities and interactions. To maintain (enter, update, and delete) data on companies and company backgrounds.

To maintain (enter, update, and delete) data on job postings and applications.

SEARCH

To perform searches on user profiles, backgrounds, and activity.

To perform searches on companies and schools.

To perform searches on job postings and applicants.





MISSION OBJECTIVES P.2

TRACK

To track the status of job openings and applicants.

To track the status of company and school involvement.

To track LockedIn's competitive business performance and ability to create professional connections (i.e. analyzing job postings, platform offerings of opportunities, and user popularity).

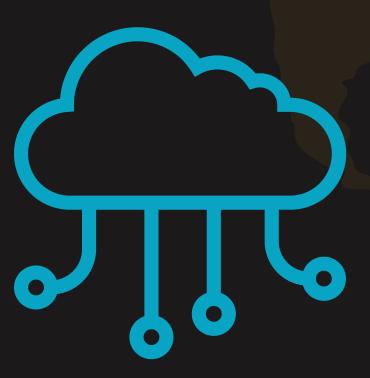
REPORT

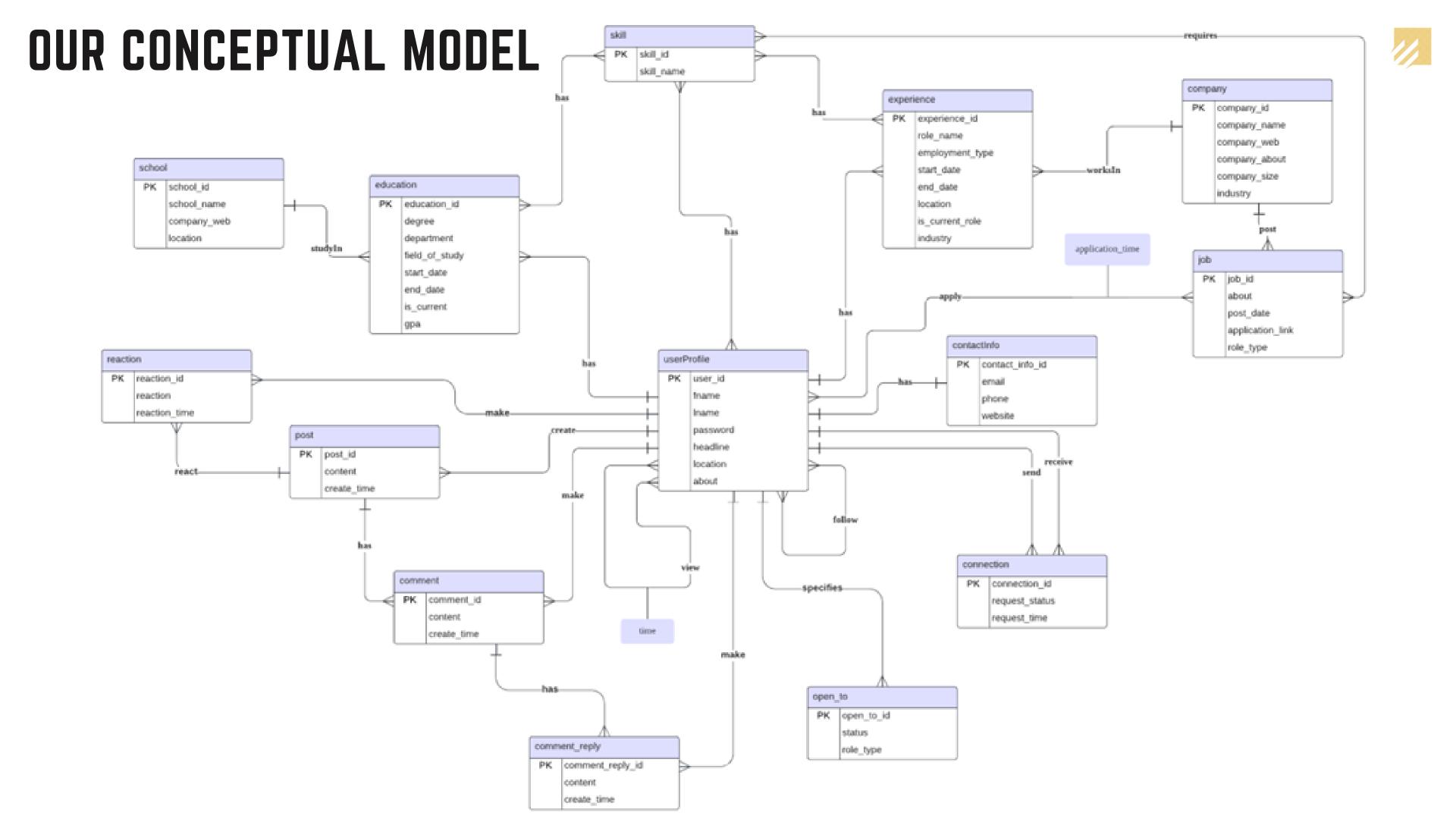
To report on users' experiences, skills, backgrounds, interactions, and activities.

To report on companies.

To report on schools and new graduates entering the job market.

To report on job openings, applicants, and skills.







RELATIONAL SCHEMA OVERVIEW

userProfile: One-to-One with contact

information

follow: Many-to-Many between users

view: Many-to-Many between users

connection: One-to-Many between users

reaction: One-to-Many between users and

post reactions

comment: One-to-Many between users and

post comments

comment_reply: One-to-Many between users,

comments, and comment replies

skill: Many-to-Many between users,

experience, job and skills

experience: One-to-Many between users and

work experiences

job: One-to-Many between companies and job

postings

user_apply: Many-to-Many between users

and job applications

school: One-to-Many relationship with

education

education: One-to-Many between users and

education

open_to: One-to-Many between users and

openness preferences

QUERIES

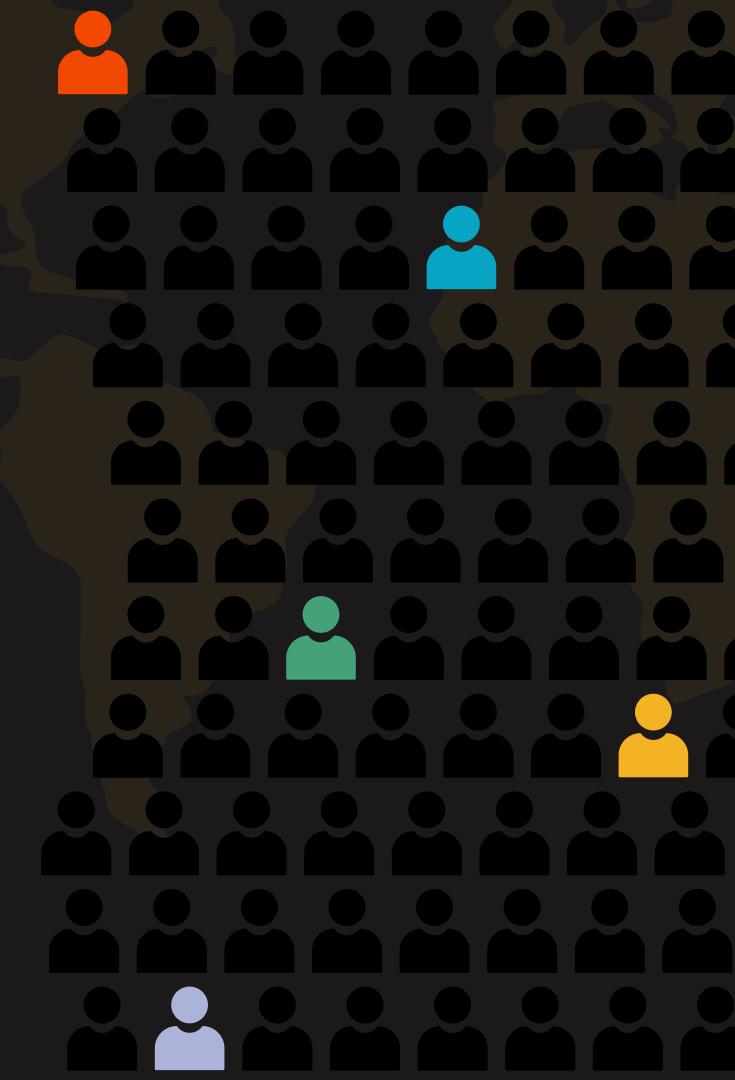
AT LOCKEDIN'S MANAGEMENT TEAM, WE CRAFTED 20 UNIQUE QUERIES CATEGORIZED INTO:



LET'S TAKE A LOOK AT THE TWO QUERIES WE'VE HIGHLIGHTED...

SCENARIO 1: FIND THE MOST POPULAR USERS

We're interested in influencers on LockedIn for potential brand deals. To determine the most popular users, we will base ourselves on users with the most number of connections. We would like to see if a user's popularity is linked to the number of posts they have also created. Rank users in terms of their number of connections (in descending order) while also linking their corresponding number of posts. Our query creates a new column with the ratio of connections to the number of posts that a user has, to see if these two fields are linked.



SCENARIO 1 QUERY BREAKDOWN:

- 1. Calculate Connection-Post Ratio:
 For each user, we calculate the ratio of their total connections to their total posts.
- 2. Gather User Data: Collect the first and last name, total posts, total connections, and connection-post ratio for each user.
- 3. Combine Data: Aggregate the data obtained from the two seperate subqueries using a UNION operation.
- 4. Output: User first and last name, total posts, total connections, and connection-post ratio for each user.

```
SELECT
    u.fname AS first_name,
    u.lname AS last name,
    combined data.total posts,
    combined_data.total_connections,
    IF(combined data.total posts > 0, combined data.total connections / combined data.total posts, 0)
AS connection post ratio
FROM (
    SELECT
        p.user_id,
        IFNULL(p.total posts, 0) AS total posts,
        IFNULL(c.total_connections, 0) AS total_connections
    FROM (
        SELECT user id, COUNT(*) AS total posts
        FROM post
        GROUP BY user_id
    ) AS p
    LEFT JOIN (
        SELECT sender AS user_id, COUNT(*) AS total_connections
        FROM connection
        GROUP BY sender
    ) AS c ON p.user_id = c.user_id
    SELECT
       c.user id,
        IFNULL(p.total_posts, 0) AS total_posts,
        IFNULL(c.total_connections, 0) AS total_connections
    FROM (
        SELECT sender AS user_id, COUNT(*) AS total_connections
        FROM connection
        GROUP BY sender
    ) AS c
    LEFT JOIN (
        SELECT user_id, COUNT(*) AS total_posts
        FROM post
        GROUP BY user id
    ) AS p ON p.user_id = c.user_id
) AS combined data
JOIN userprofile u ON combined data.user id = u.user id
ORDER BY total_connections DESC;
```

SCENARIO 1 QUERY OUTPUT

first_name	last_name	total_posts	total_connections	connection_post_ratio
Christopher	Lopez	2	9	4.5000
Brittany	Turner	1	7	7.0000
Jennifer	Garcia	1	7	7.0000
Matthew	Rodriguez	0	5	0.0000
Alice	Johnson	0	5	0.0000
Michael	Brown	1	5	5.0000
Daniel	Wilson	0	5	0.0000
Kevin	Martinez	1	5	5.0000
Jane	Smith	1	5	5.0000
Amanda	Anderson	2	5	2.5000
Emily	Davis	0	4	0.0000
David	Young	1	3	3.0000
John	Doe	4	3	0.7500

SCENARIO 2: HIRE THE BEST FIT

SunPower Energy is hiring for a Cloud Computing Specialist. The company's HR department has registered on LockedIn and wants to check if there are any users on the platform who could be a good fit for this position. As the company values the extent to which a candidate's skill set matches with the job posting and their educational background, the following table is extracted. The fit-level is a percentage data value that represents the number of skills a person possesses that are also listed in the job posting.

```
SQL script:
SELECT a.user id, a.fit level, b.gpa, b.school name, b.location
FROM
        SELECT user id,
               COUNT(*) / (SELECT COUNT(*)
                           FROM job skill
                           WHERE job_id = 4) AS fit_level
        FROM user skill
        WHERE skill id IN
              (SELECT skill id
               FROM job skill
               WHERE job_id = 8)
        GROUP BY user id
        ORDER BY fit level DESC
   ) a
JOIN
        SELECT e.user id, e.gpa, s.school name, s.location
        FROM education e
        JOIN school s ON e.school id = s.school id
    ) b ON a.user_id = b.user_id;
```

SCENARIO 2 QUERY BREAKDOWN

- 1. Calculate Fit Level: For each user, determine how many skills they have that match those needed for the job. Then, compare this to the total skills required to get a fit level.
- 2. Gather Education Info: Collect GPA, school name, and location for each user from their education records.
- 3. Combine Data: Join the fit-level data with the education info based on the user ID.
- 4. Output: Display the user ID, fit level, GPA, school name, and location.

```
SQL script:
SELECT a.user id, a.fit level, b.gpa, b.school name, b.location
FROM
        SELECT user_id,
               COUNT(*) / (SELECT COUNT(*)
                           FROM job skill
                           WHERE job_id = 4) AS fit_level
        FROM user skill
        WHERE skill id IN
              (SELECT skill id
               FROM job_skill
               WHERE job_id = 8)
        GROUP BY user id
        ORDER BY fit level DESC
    ) a
JOIN
        SELECT e.user id, e.gpa, s.school name, s.location
        FROM education e
        JOIN school s ON e.school id = s.school id
    ) b ON a.user id = b.user id;
```

SCENARIO 2 QUERY OUTPUT

user_id	fit_level	gpa	school_name	location
user12	0.5000	3.80	Lakeside Arts Academy	Chicago
user15	0.5000	3.60	Sunrise Computer Science Institute	Austin
user13	0.5000	3.70	Green Valley Environmental Studies	Denver



CHALLENGES, LEARNINGS, AND OPPORTUNITIES.

CHALLENGES	Defining the scope of the project: determining which features provide value and need to be included on LockedIn	Populating each table with realistic dummy data was especially challenging due to the numerous interrelationships among values from different tables.	
LEARNINGS	This phase required a few iterations to finalize the ERD which takes into account all core features of the platform.	It was important to ensure that the data remained coherent when executing SQL queries, so another iterative process was necessary to refine and finalize the dummy data set.	

OPPORTUNITIES FOR LOCKEDIN

Going forward, more dummy data can be added to the database to get more realistic outputs when testing new queries. Moreover, a future objective is to integrate a new set of SQL queries to allow LockedIn to have its API which it can share with third-party developers and get integrated on other platforms. A second beta version of the database could be crafted to integrate additional prominent features, such as private messaging, which plays a pivotal role in fostering user interactions and enhancing their connectivity experience within our platform.

THANK YOU FOR READING!



PROFESSIONAL NETWORKING, NOTHING ELSE.