ConnectMe: Uma abordagem de modelagem ER/EER com geração sintética de dados para testes

Daniel de Oliveira Trindade¹

¹Instituto de Computação – Universidade Federal do Amazonas (UFAM) 69067-005 – Amazonas – AM – Brazil

daniel.trindade@icomp.ufam.edu.br

Resumo. Este artigo apresenta o desenvolvimento de uma rede social rudimentar denominada ConnectMe, com foco na modelagem conceitual utilizando diagrama ER/EER. O trabalho detalha a evolução do modelo conceitual até a sua implementação física, incluindo a aplicação do processo de normalização de modelos até a 3FN. Em conjunto apresenta uma abordagem recente do uso de IA's generativas para a geração de scripts para gerar dados sintéticos que serão usados para testar a estrutura físca da rede social.

1. Introdução

Os sistemas de bancos de dados surgiram da necessidade de gerenciar eficientemente grandes volumes de dados nas organizações, evoluindo desde sistemas de arquivos tradicionais até sistemas mais sofisticados [Elmasri and Navathe 2011]. O modelo relacional, proposto por Codd[Codd 1970], trouxe uma nova abordagem matemática para a organização de dados, que posteriormente levou ao desenvolvimento dos Sistemas Gerenciadores de Bancos de Dados Relacionais (SGBDR). O modelo Entidade-Relacionamento (ER), introduzido por Chen[Chen 1976], revolucionou a modelagem conceitual, permitindo representar dados de forma mais em conformidade com a realidade.

A prática da modelagem de dados é uma habilidade fundamental para estudantes e profissionais de tecnologia, em especial em um contexto onde sistemas mais complexos como redes sociais que demandam uma estrutura robusta e escalável [Silberschatz et al. 2020]. E desenvolver projetos acadêmicos que simulam nas devidadas proporções a experiência prática de como normalizar, manter a integridade referencial e otimizar consultas é fundamental.

De maneira tradicional, o teste desses sistemas requeria dados que normalmente eram gerados manualmente ou scripts básico de população. Com o advento das IAs generativas [Brown et al. 2023], surge um novo modo de automatizar a criação de scripts que sejam mais sofisticados, e que gerem conjuntos melhores de dados para representar cenários reais. Este trabalho explora essa abordagem através do desenvolvimento do ConnectMe, uma rede social simplificada que serve como caso de estudo para aplicação de conceitos de modelagem ER/EER e utilização de IA na geração de dados de teste.

2. Modelagem Conceitual do ConnectMe

2.1. Análise de Requisitos

A ConnectMe é uma plataforma online que permite aos usuários compartilhar informações, se conectar com amigos, participar de grupos de interesse e interagir por

meio de postagens, comentários e mensagens privadas. O objetivo da ConnectMe é fornecer uma experiência social envolvente e personalizada para seus usuários, promovendo conexões significativas e facilitando a comunicação e a interação online. A ConnectMe possui as seguintes características principais:

- **Perfil do Usuário**: Cada usuário possui um perfil que inclui informações pessoais como nome, foto, biografia, localização, data de nascimento, interesses, etc.
- **Postagens**: Os usuários podem fazer postagens em seu próprio perfil ou em grupos aos quais pertencem. As postagens podem incluir texto, imagens, vídeos e outros tipos de mídia.
- Grupos: Os usuários podem criar e participar de grupos com base em interesses comuns, hobbies, profissão, localização geográfica, etc. Os grupos servem como fóruns de discussão onde os membros podem compartilhar conteúdo relevante e interagir.
- Conexões: Os usuários podem se conectar uns aos outros, formando uma rede de amigos e contatos. As conexões permitem que os usuários vejam as postagens e atualizações de seus amigos em seus feeds de notícias.
- Mensagens Privadas: Os usuários podem enviar mensagens privadas uns aos outros para iniciar conversas individuais ou em grupo.

Em conjunto também é solicitado que seja possível fazer as seguintes consultas:

- Consulta de perfil do usuário: Recuperar informações do perfil de um usuário específico, incluindo nome, foto, biografia, etc.
- Consulta de conexões de um usuário: Obter a lista de amigos de um determinado usuário.
- Consulta de postagens de um usuário: Recuperar todas as postagens feitas por um usuário específico, ordenadas por data de publicação (as mais recentes primeiro).
- Consulta de postagens em um grupo: Listar 20 as postagens mais recentes feitas em um grupo específico.
- Consulta de mensagens privadas: Listar as 10 mensagens privadas mais recentes trocadas entre 2 usuários dados como entrada.
- Consulta de busca de usuários: Dado uma string como entrada, buscar os usuários cujos nomes contenham a string fornecida.
- Consulta de tendências: Listar o identificador dos 5 posts com mais interações nos últimos 7 dias.
- Consulta de análise de engajamento: Dado um post de um usuário, fazer a contagem de quantos usuários interagiram com ele nos últimos 7 dias.

2.2. Entidades e Relacionamentos

Depois de fazer a análise dos requisitos e ponderar sobre como seria a estrutura com base no que foi pedido, foi destacado as seguintes entidades:

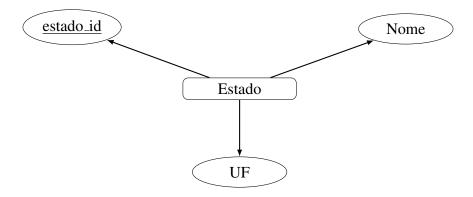


Figura 1. Diagrama da entidade Estado e seus atributos.

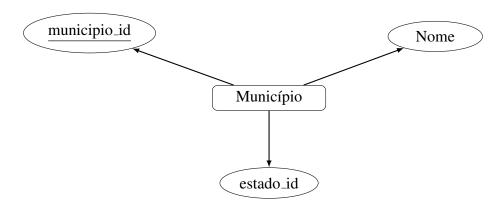


Figura 2. Diagrama da entidade Município e seus atributos.

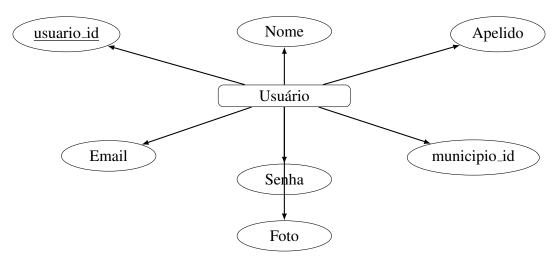


Figura 3. Diagrama da entidade Usuário e seus atributos.

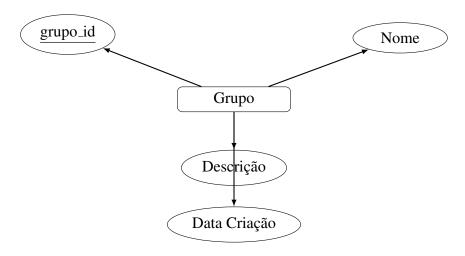


Figura 4. Diagrama da entidade Grupo e seus atributos.

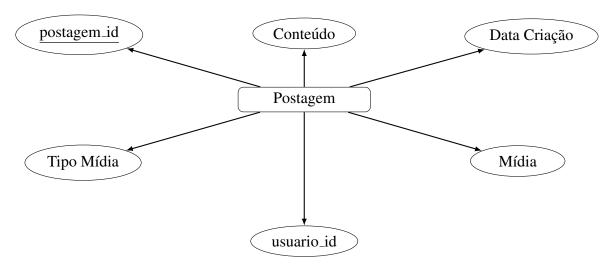


Figura 5. Diagrama da entidade Postagem e seus atributos.

2.3. Modelo ER/EER

Verificado as entidades com seus devidos atributos e feito a abstração do problema, o resultado foi o seguinte modelo ER/EER:

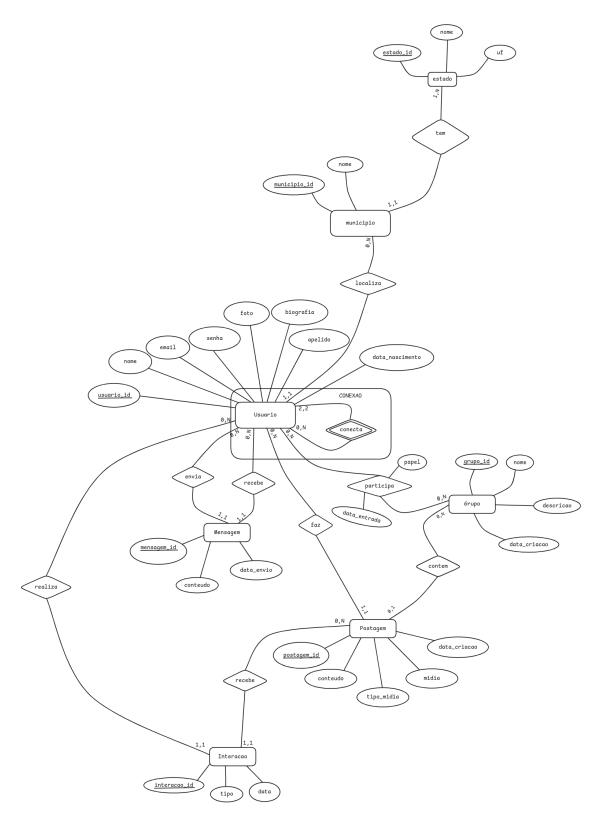


Figura 6. Modelo Entidade-Relacionamento (ER) do sistema ConnectMe.

No modelo acima (Figura 7), estão representadas as principais entidades, relacionamentos e atributos do sistema.

2.4. Normalizando o sistema

Depois de montado o modelo, foi realizado o processo de normalizar o banco até 3FN, que se mostrou necessário e o resultado se encontra abaixo:

- **ESTADO**(estado_id, nome, uf)
- MUNICIPIO(municipio_id, nome, #estado_id)
 MUNICIPIO[estado_id] = ESTADO[estado_id]
- USUARIO(<u>usuario_id</u>, nome, apelido, email, senha, foto, biografia, data_nascimento, #<u>municipio_id</u>)
 USUARIO[<u>municipio_id</u>] = MUNICIPIO[<u>municipio_id</u>]
- **GRUPO**(grupo_id, nome, descricao, data_criacao)
- POSTAGEM(postagem_id, conteudo, data_criacao, tipo_midia, midia, #usuario_id, #grupo_id)

 POSTAGEM[usuario_id] = USUARIO[usuario_id]

 POSTAGEM[grupo_id] = GRUPO[grupo_id]
- **MENSAGEM**(<u>mensagem_id</u>, conteudo, data_envio, #<u>remetente_id</u>, #<u>destinatario_id</u>)

MENSAGEM[remetente_id] = USUARIO[usuario_id]
MENSAGEM[destinatario_id] = USUARIO[usuario_id]

- CONEXAO(conexao_id, data_conexao, #usuario_id1, #usuario_id2)

 CONEXAO[usuario_id1] = USUARIO[usuario_id]

 CONEXAO[usuario_id2] = USUARIO[usuario_id]
- INTERACAO(interacao_id, tipo, data, #usuario_id, #postagem_id)
 INTERACAO[usuario_id] = USUARIO[usuario_id]
 INTERACAO[postagem_id] = POSTAGEM[postagem_id]
- **MEMBRO_GRUPO**(membro_id, papel, data_entrada, #usuario_id, #grupo_id)

 MEMBRO_GRUPO[usuario_id] = USUARIO[usuario_id]

 MEMBRO_GRUPO[grupo_id] = GRUPO[grupo_id]

3. Implementação do ConnectMe

Depois do passo anterior, com base no diagrama e no modelo relacional. Foi implementado os scripts de criação das tabelas e relacionamentos no banco de dados.

```
CREATE DATABASE connectme;
  USE connectme;
 CREATE TABLE ESTADO (
   estado_id INT AUTO_INCREMENT PRIMARY KEY,
   nome VARCHAR (100) NOT NULL,
   uf CHAR(2) NOT NULL
8);
10 CREATE TABLE MUNICIPIO (
   municipio_id INT AUTO_INCREMENT PRIMARY KEY,
11
  nome VARCHAR (100) NOT NULL,
  estado_id INT NOT NULL,
    FOREIGN KEY (estado_id) REFERENCES ESTADO(estado_id)
14
15);
17 CREATE TABLE USUARIO (
```

```
usuario_id INT AUTO_INCREMENT PRIMARY KEY,
    nome VARCHAR (100) NOT NULL,
    apelido VARCHAR (100) NOT NULL,
20
    email VARCHAR(100) UNIQUE NOT NULL,
    senha VARCHAR (255) NOT NULL,
23
    foto VARCHAR (255),
    biografia TEXT,
24
    data_nascimento DATE NOT NULL,
25
    municipio_id INT NOT NULL,
    FOREIGN KEY (municipio_id) REFERENCES MUNICIPIO(municipio_id)
27
28);
29
 CREATE TABLE GRUPO (
    grupo_id INT AUTO_INCREMENT PRIMARY KEY,
31
    nome VARCHAR (100) NOT NULL,
32
33
    descricao TEXT,
   data_criacao TIMESTAMP DEFAULT CURRENT_TIMESTAMP
35 );
36
37 CREATE TABLE POSTAGEM (
    postagem_id INT AUTO_INCREMENT PRIMARY KEY,
    conteudo TEXT NOT NULL,
39
    data_criacao TIMESTAMP DEFAULT CURRENT_TIMESTAMP.
40
    tipo_midia VARCHAR(50),
41
    midia VARCHAR (255),
42
43
    usuario_id INT NOT NULL,
    grupo_id INT,
44
    FOREIGN KEY (usuario_id) REFERENCES USUARIO(usuario_id),
    FOREIGN KEY (grupo_id) REFERENCES GRUPO(grupo_id)
46
47 );
48
49 CREATE TABLE MENSAGEM (
    mensagem_id INT AUTO_INCREMENT PRIMARY KEY,
50
    conteudo TEXT NOT NULL,
51
    data_envio TIMESTAMP DEFAULT CURRENT_TIMESTAMP,
52
    remetente_id INT NOT NULL,
53
    destinatario_id INT NOT NULL,
54
    FOREIGN KEY (remetente_id) REFERENCES USUARIO(usuario_id),
55
    FOREIGN KEY (destinatario_id) REFERENCES USUARIO(usuario_id)
56
57 );
58
59 CREATE TABLE CONEXAO (
   conexao_id INT AUTO_INCREMENT PRIMARY KEY,
60
    data_conexao TIMESTAMP DEFAULT CURRENT_TIMESTAMP,
61
    usuario_id1 INT NOT NULL,
62
   usuario_id2 INT NOT NULL,
63
   FOREIGN KEY (usuario_id1) REFERENCES USUARIO(usuario_id),
    FOREIGN KEY (usuario_id2) REFERENCES USUARIO(usuario_id)
65
66 );
67
68 CREATE TABLE INTERACAO (
    interacao_id INT AUTO_INCREMENT PRIMARY KEY,
69
    tipo VARCHAR(50) NOT NULL,
70
    data TIMESTAMP DEFAULT CURRENT_TIMESTAMP,
71
   usuario_id INT NOT NULL,
postagem_id INT NOT NULL,
```

```
FOREIGN KEY (usuario_id) REFERENCES USUARIO(usuario_id),
    FOREIGN KEY (postagem_id) REFERENCES POSTAGEM(postagem_id)
76);
77
78 CREATE TABLE MEMBRO_GRUPO (
79
    membro_id INT AUTO_INCREMENT PRIMARY KEY,
    papel VARCHAR(50) NOT NULL,
80
    data_entrada TIMESTAMP DEFAULT CURRENT_TIMESTAMP,
81
   usuario_id INT NOT NULL,
    grupo_id INT NOT NULL,
83
    FOREIGN KEY (usuario_id) REFERENCES USUARIO(usuario_id),
84
    FOREIGN KEY (grupo_id) REFERENCES GRUPO(grupo_id)
85
 );
86
87
88 CREATE INDEX idx_usuario_nome ON USUARIO(nome);
89 CREATE INDEX idx_postagem_data ON POSTAGEM(data_criacao);
90 CREATE INDEX idx_mensagem_data ON MENSAGEM(data_envio);
gi CREATE INDEX idx_conexao_usuarios ON CONEXAO(usuario_id1, usuario_id2);
```

4. Populando o banco

Com a criação das tabelas, foi utilizado a IA generativa chamada claude da Anthropic para gerar scripts que irão popular o nosso banco com dados para testar a modelagem feita.

Script 1. Script para popular usuários

```
ı # Funcaoo para popular a tabela de usuarios
 def populate_usuarios(cursor, municipio_ids, num_usuarios=3000):
   Funcaoo para popular a tabela de usuarios no banco de dados.
    :param cursor: O cursor do banco de dados para execucaoo das queries.
    :param municipio_ids: Lista de IDs de municipios para atribuicao aos
     usuarios.
   :param num_usuarios: Numero de usuarios a serem criados (default:
    :return: Lista de IDs dos usuarios criados.
    usuarios = [] # Lista para armazenar os IDs dos usuarios criados
11
    used_emails = set() # Conjunto para garantir que os e-mails sejam
12
     unicos
13
14
   # Loop para criar os usuarios
    for _ in tqdm(range(num_usuarios), desc="Populando usuarios", unit="
15
     usuario"):
      while True:
16
        # Gerando um e-mail unico para o usuario
17
        email = fake.email()
18
        if email not in used emails:
19
          used_emails.add(email)
20
          break
21
22
      # Gerando os outros dados do usuario
23
      data_nascimento = fake.date_of_birth(minimum_age=13, maximum_age
24
     =80)
```

```
usuario = (
25
        fake.name(), # Nome do usuario
26
        fake.first_name(), # Apelido
27
        email, # E-mail unico
28
        fake.password(), # Senha
29
        fake.image_url(), # URL de imagem (foto)
30
        fake.text(max_nb_chars=200), # Biografia
31
        random.choice(municipio_ids), # ID do municipio
        data_nascimento # Data de nascimento
33
34
35
      try:
36
        # Inser o do usuario na tabela USUARIO
37
        cursor.execute("""
38
          INSERT INTO USUARIO (nome, apelido, email, senha, foto,
39
     biografia,
                      municipio_id, data_nascimento)
40
          VALUES (%s, %s, %s, %s, %s, %s, %s)
41
        """, usuario)
42
43
        # Adiciona o ID do usu rio recem-criado a lista
45
        usuarios.append(cursor.lastrowid)
46
      except mysql.connector.IntegrityError:
47
        # Caso ocorra erro de integridade (por exemplo, e-mail duplicado)
      , tenta criar o pr ximo usu rio
        continue
49
50
    return usuarios # Retorna a lista de IDs dos usuarios criados
51
```

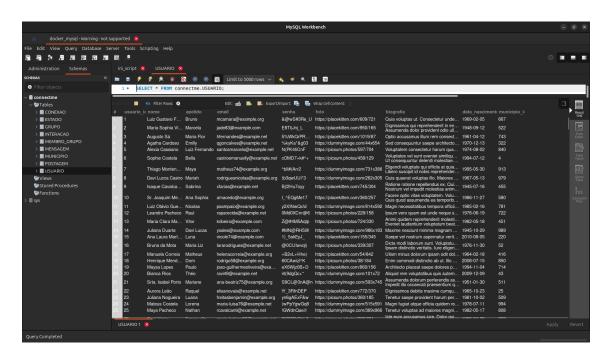


Figura 7. Exemplo da tabela usuário populada pelo script.

5. Implementação das consultas

Para verificar se o modelo está de acordo com o solicitado, iremos implementar 8 queries e que foram solicitadas.

5.1. Consulta de perfil do usuário

```
SELECT

u.usuario_id,
u.nome,
u.email,
u.foto,
u.biografia,
u.data_nascimento,
m.nome AS municipio,
e.nome AS estado,
e.uf

FROM USUARIO u

JOIN MUNICIPIO m ON u.municipio_id = m.municipio_id

JOIN ESTADO e ON m.estado_id = e.estado_id

WHERE usuario_id = ?;--substituir o ? pelo usuario_id que quer procurar
```

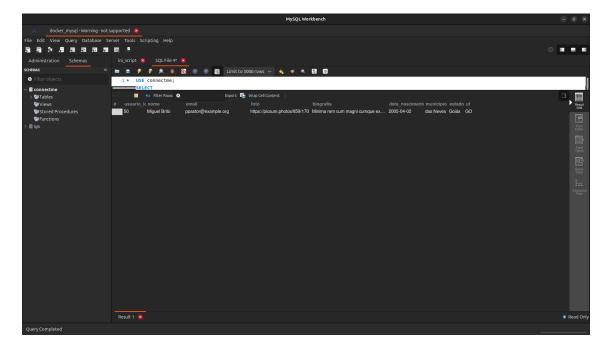


Figura 8. um exemplo de saída para a query 1

5.2. Consulta de conexões de um usuário

```
SELECT

u.usuario_id,

u.nome,

u.email

FROM

CONEXAO c

JOIN
```

```
USUARIO u ON (u.usuario_id = c.usuario_id1 AND c.usuario_id2 = ?) --
subtituir o ? por pelo usuario_id de que quer procurar
OR (u.usuario_id = c.usuario_id2 AND c.usuario_id1 = ?); --
subtituir o ? pelo mesmo usuario_id usado acima
```

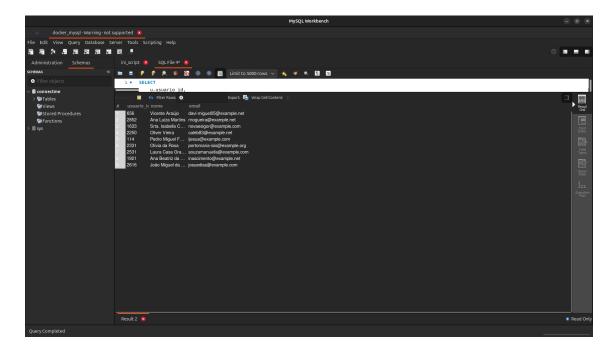


Figura 9. um exemplo de saída para a query 2

5.3. Consulta de postagens de um usuário

```
SELECT

*
FROM

POSTAGEM

WHERE

usuario_id = ? --subtituir o ? por pelo usuario_id de que quer
procurar

ORDER BY
data_criacao DESC;
```

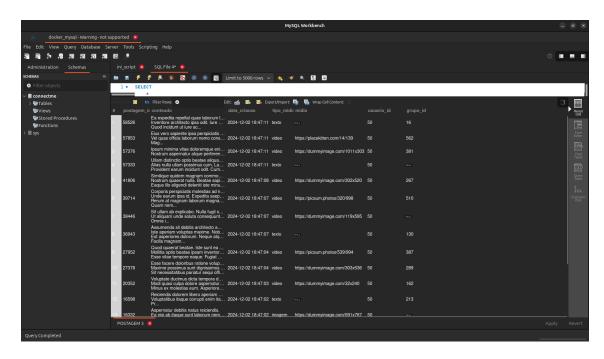


Figura 10. um exemplo de saída para a query 3

5.4. Consulta de postagens em um grupo

```
1 SELECT
   p.postagem_id,
   p.conteudo,
   p.data_criacao,
   p.tipo_midia,
   p.midia,
   p.usuario_id,
   u.nome AS nome_usuario
9 FROM
  POSTAGEM p
11 JOIN
   USUARIO u ON p.usuario_id = u.usuario_id
13 WHERE
  p.grupo_id = ? -- substituir o ? pelo id da postagem que quer
    procurar
15 ORDER BY
p.data_criacao DESC
17 LIMIT 20;
```

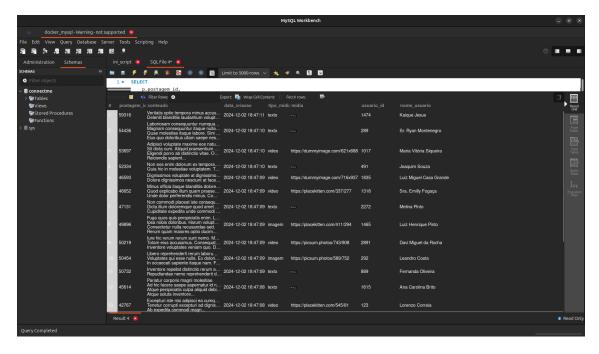


Figura 11. um exemplo de saída para a query 4

5.5. Consulta de mensagens privadas

```
SELECT

mensagem_id,
conteudo,
data_envio,
remetente_id,
destinatario_id

FROM

MENSAGEM

WHERE

(remetente_id = ? AND destinatario_id = ?) -- ex: subtituir o 1 ?
por 2 e o 2 por 3

OR (remetente_id = ? AND destinatario_id = ?) -- ex: subtituir o 1
? por 3 e o 2 por 2

ORDER BY

data_envio DESC

LIMIT 10;
```

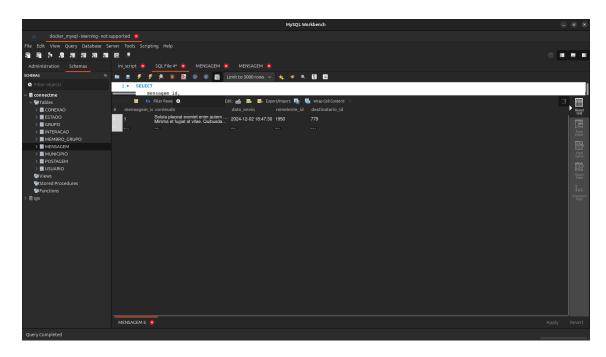


Figura 12. um exemplo de saída para a query 5

5.6. Consulta de busca de usuários

```
SELECT
usuario_id,
nome,
email,
municipio_id
FROM
USUARIO
WHERE
nome LIKE CONCAT('%', ?, '%'); -- substituir o ? pela string que quer
procurar, ex: 'Daniel'
```

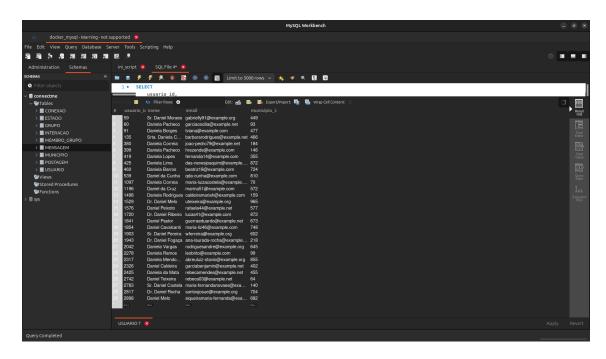


Figura 13. um exemplo de saída para a query 6

5.7. Consulta de tendências

```
1 SELECT
   p.postagem_id,
    p.conteudo AS conteudo_postagem,
    p.data_criacao AS data_postagem,
    p.tipo_midia,
    p.midia,
   u.nome AS nome_usuario,
    COUNT(i.interacao_id) AS total_interacoes,
10
      WHEN p.grupo_id IS NOT NULL THEN g.nome
      ELSE 'N o de grupo'
11
    END AS nome_grupo
13 FROM
    INTERACAO i
14
15 JOIN
   POSTAGEM p ON i.postagem_id = p.postagem_id
16
17 JOIN
   USUARIO u ON p.usuario_id = u.usuario_id
18
19 LEFT JOIN
  GRUPO g ON p.grupo_id = g.grupo_id
20
21 WHERE
  i.data >= CURDATE() - INTERVAL 7 DAY
22
23 GROUP BY
  p.postagem_id
24
25 ORDER BY
   total_interacoes DESC
26
27 LIMIT 5;
```

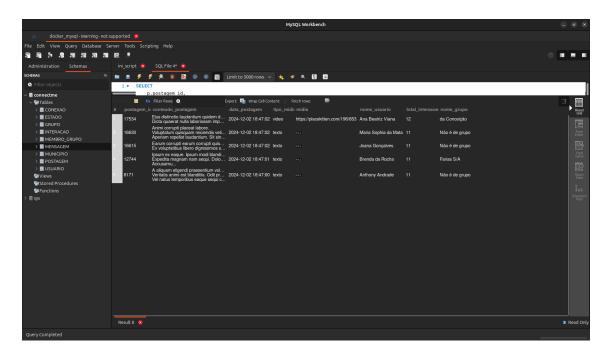


Figura 14. um exemplo de saída para a query 7

5.8. Consulta de tendências

```
SELECT
COUNT(DISTINCT i.usuario_id) AS total_usuarios_interagiram
FROM
INTERACAO i
JOIN
POSTAGEM p ON i.postagem_id = p.postagem_id
WHERE
p.postagem_id = ? -- Substitua '?' pelo postagem_id do post desejado
AND i.data >= CURDATE() - INTERVAL 7 DAY;
```

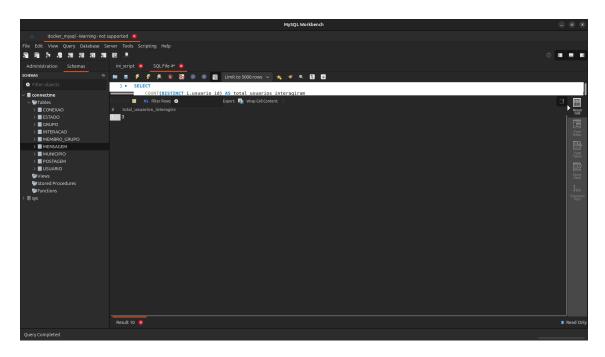


Figura 15. um exemplo de saída para a query 8

6. Conclusão

Neste trabalho, abordado o desenvolvimento de um sistema de banco de dados para o gerenciamento da rede social ConnectMe para o gerenciamento dos seus usuário, post, grupos e etc. A estrutura adotada foi possível devido ao uso da técnica de modelagem de dados ER, para garantir um sistema que atenda as necessidades de escalabilidade e robustez do sistema, essenciais para redes sociais de larga escala.

Além disso, a implementação de script usando IA generativa para auxiliar nos testes populando o banco de dados com dados falsos e realistas, mostrou-se uma estratégia eficaz no desenvolvimento de projetos reais. Não apenas no processo de teste, mas também contribuindo para a criação de um sistema mais alinhado com a realidade.

Este trabalho demonstrou a importância de um design bem construído e adoção de boa práticas garantem uma qualidade e eficiência no desenvolvimento de sistemas baseados em bancos de dados relacionais e como um projeto construído em passos bem definidos é importante para o sucesso. Por fim, a utilização de IA se consolidou como uma ferramenta de grande valia, para otimizar o ciclo de desenvolvimento de sistemas.

Referências

Brown, T. B. et al. (2023). Gpt-4 technical report. arXiv preprint arXiv:2303.08774.

Chen, P. P.-S. (1976). The entity-relationship model—toward a unified view of data. *ACM Transactions on Database Systems*, 1(1):9–36.

Codd, E. F. (1970). A relational model of data for large shared data banks. *Communications of the ACM*, 13(6):377–387.

Elmasri, R. and Navathe, S. B. (2011). *Fundamentals of Database Systems*. Addison-Wesley, 6 edition.

Silberschatz, A., Korth, H. F., and Sudarshan, S. (2020). *Database System Concepts*. McGraw-Hill, 7 edition.