



**Beatriz Saito Gobira**

## **Normalização**

Trabalho para obtenção de nota na matéria de banco de dados, onde haverá apenas criação de função (stored procedure).

Professor: Edson Castro

**Três Lagoas**

**2023**



O dataset selecionado consiste em um banco de dados que registra óbitos por doenças crônicas não transmissíveis, sendo composto por informações como datas de nascimento e óbito, tempo de vida, sexo, raça, escolaridade, municípios de ocorrência e residência, e detalhes da doença, tais como CID<sup>1</sup> numérico, capítulo, categoria e descrição. O acesso a esses dados está disponível no site <https://dados.gov.br/dados/conjuntos-dados/dados-doencas-cronicas-ses> para análises e entendimento adicionais

Para iniciar o processo, os dados foram transferidos de um arquivo CSV para uma tabela no banco de dados PostgreSQL. A primeira etapa de normalização (1FN) foi alcançada com a divisão dos dados em quatro tabelas distintas: Pessoa, Município\_Residencia, Município\_Ocorrência, Prontuário e Cids. Vamos agora aprofundar a compreensão de cada uma dessas tabelas.

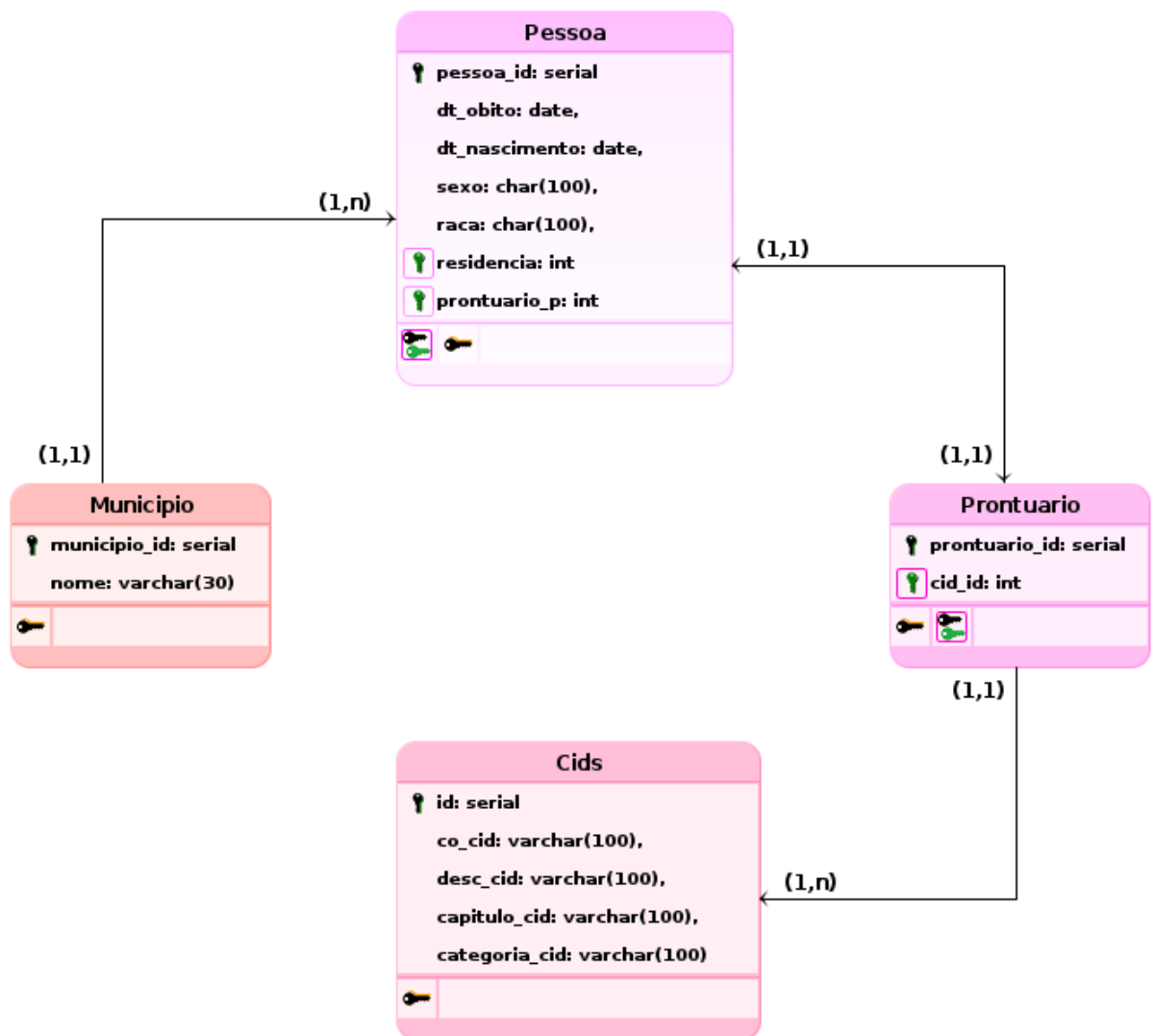
A tabela principal, denominada "tabela\_principal," armazena dados cruciais, como data de óbito, data de nascimento, idade, sexo, raça, escolaridade, municípios, CID da causa, descrição do CID, capítulo do CID e categoria do CID. Prosseguindo, foram criadas tabelas adicionais: "municipio," com as colunas município\_id e nome, "cids," com campos id, co\_cid, desc\_cid, capitolo\_cid, categoria\_cid, e "prontuario," com as colunas prontuario\_id (chave primária) e cid\_id (chave estrangeira), estabelecendo a relação entre prontuários médicos e dados de CID. A tabela "pessoa" foi desenvolvida para consolidar os dados cadastrais de cada paciente, marcando o início da segunda etapa da normalização.

Na segunda etapa, as tabelas "prontuario" e "municipio" foram ligadas à tabela "pessoa" para eliminar dependências parciais, garantindo a consistência e integridade dos dados.

Após a conclusão da normalização, se dá início ao processo de criptografia, para isso foi utilizado a biblioteca PGCRYPTO que é uma ferramenta robusta para implementar criptografia no PostgreSQL. Foi utilizada a função pgp\_sym\_encrypt que é focada em criptografia simétrica, onde é utilizada para cifrar informações sensíveis, como CID da doença, capítulo do CID, categoria do CID e detalhes da residência. Essa função aceita como parâmetros a coluna desejada a ser criptografada e a senha de criptografia, servindo como chave para o processo de cifragem.

No que tange às formas normais, a primeira forma normal (1FN) diz respeito à eliminação de grupos repetitivos de dados e à garantia de que cada coluna contenha apenas valores atômicos. A segunda forma normal (2FN) vai além, eliminando dependências parciais, assegurando que cada coluna dependa completamente da chave primária.

Para concluir, foi desenvolvida uma view que possibilita a denormalização dos dados, incluindo informações primordiais da tabela original, mas utilizando as tabelas geradas no processo de normalização. Essa view inclui a função de descriptografar os dados, exigindo a chave de segurança para a decifragem ou permitindo a passagem de parâmetros em branco, caso não haja senha. Este relatório é finalizado com os dados descriptografados e as colunas correspondentes da tabela original, proporcionando uma visão simplificada das tabelas normalizadas para melhor compreensão visual.



### Referencias:

<https://www.youtube.com/watch?v=eRaAMNjCFYw>

<https://www.youtube.com/watch?v=6ER9IWOk-cY>

<https://www.postgresql.org/docs/current/pgcrypto.html>