

DCC011: INTRODUÇÃO A BANCO DE DADOS - TRABALHO PRÁTICO 2

INTRODUÇÃO:

O objetivo deste trabalho é projetar e implementar um banco de dados relacional para análise de dados abertos governamentais. O projeto do banco de dados deverá seguir um processo *bottom-up*, iniciado a partir da análise de um conjunto de dados existente, passando pela normalização de seu esquema relacional, e mapeamento do esquema relacional normalizado para o esquema ER correspondente. O conjunto de dados reais escolhido deverá ser então inserido no banco de dados normalizado. Por fim, consultas analíticas deverão ser formuladas e executadas sobre esse banco de dados.

Características básicas (até 40%): Para definição do tema, cada grupo deverá escolher um conjunto de dados atualizado em 2025 no portal dados.gov.br/dados/conjuntos-dados.¹ Para o conjunto de dados escolhido, deverão ser apresentados o esquema relacional normalizado e o diagrama ER correspondentes, atendendo aos seguintes requisitos:

- Pelo menos 4 tipos de entidade, cada tipo com ao menos 2 atributos (além de um atributo identificador);
- Pelo menos 3 tipos de relacionamento, ao menos 1 com cardinalidade M:N;
- Pode ser necessário alterar os dados originais para atender a esses requisitos.

Consultas (até 40%): Deverão ser especificadas e executadas um total de **10 consultas em SQL**, sendo:

- 2 consultas envolvendo as operações de seleção e projeção;
- 3 consultas envolvendo a junção de duas relações;
- 3 consultas envolvendo a junção de três ou mais relações;
- 2 consultas envolvendo funções de agregação sobre o resultado da junção de pelo menos duas relações.

Análise dos resultados (até 20%): Além dos requisitos listados acima, os relatórios produzidos serão avaliados quanto à criatividade e profundidade das análises desenvolvidas a partir dos resultados retornados pelas várias consultas.

Aplicação Streamlit (até 20% EXTRA): Serão concedidos pontos extras aos grupos que programarem uma aplicação web via Streamlit² para visualização interativa dos dados modelados por meio das consultas produzidas.

OBSERVAÇÕES:

1. O trabalho deverá ser feito em **grupo de 3-4 alunos**.
2. O SGBD **SQLite** embarcado em um **Jupyter notebook** deverá ser utilizado para implementação do banco de dados e execução das consultas. **DICA:** Para acelerar a inserção de dados, desative temporariamente a validação de chaves estrangeiras (**PRAGMA foreign_keys = 0**).
3. A avaliação compreenderá o relatório (notebook) final entregue e a apresentação.

CALENDÁRIO:

04/11 (ter): Entrega da proposta (.pdf, máx. 1 página, via Moodle): definição do grupo e do tema

A proposta deverá descrever o conjunto de dados escolhido do portal dados.gov.br/dados/conjuntos-dados. Para evitar temas duplicados, informe seu grupo e dataset escolhido na seguinte planilha: <http://bit.ly/4hyDcnl>. Na proposta escrita (1 página), deverão ser listadas as entidades e relacionamentos presentes, o número de instâncias de cada entidade e relacionamento, e possíveis investigações (em alto nível) a serem formuladas sobre esses dados. A proposta deve ser submetida via Moodle.

25/11 (ter): Entrega do relatório final (.ipynb + .pdf, verifique o template no Moodle)

O relatório final deve ser feito no formato notebook (.ipynb), incluindo as seguintes seções (um template está disponível na atividade criada no Moodle): (1) Título; (2) Membros (nome e matrícula); (3) Descrição dos dados (como foram processados?); (4) Diagrama ER; (5) Diagrama relacional; (6) Consultas; (7) Autoavaliação dos membros

Idealmente, o processamento dos dados deverá ser feito no próprio notebook e implementado na Seção 3.

Os diagramas ER e relacional produzidos deverão ser anexados nas Seções 4 e 5, respectivamente.

Na Seção 6, cada consulta deverá ser definida através de uma explicação textual e do respectivo comando SQL, acompanhado do resultado da consulta. Gráficos também podem ser incluídos para melhor visualização dos resultados.

Na Seção 7, a autoavaliação deverá descrever as atividades realizadas individualmente por cada membro.

25/11 (ter), 27/11 (qui): Apresentações em sala

Cada grupo terá 10 min (+5 min para perguntas) para apresentar o tema do trabalho, a modelagem desenvolvida (esquemas ER e relacional), e uma seleção das várias consultas formuladas (incluindo sua explicação em alto nível e seus resultados retornados). Para apresentação em sala, o grupo deverá preparar slides incluindo os diagramas, consultas selecionadas e seus respectivos resultados (possivelmente com gráficos). Poderão apresentar também a aplicação Streamlit desenvolvida.

¹ Conjuntos de dados originalmente disponibilizados em formato SQL **não** poderão ser utilizados.

² <https://www.alura.com.br/artigos/streamlit-compartilhando-sua-aplicacao-de-dados-sem-dor-de-cabeca>