

Banco de Dados I

Projeto Acadêmico

Gestão da Tecnologia da Informação - 2025.01

Objetivo geral

Desenvolver um banco de dados completo, a partir de dados abertos, passando pelas etapas de modelagem lógica, modelagem física, inserção de dados e execução de consultas, promovendo o raciocínio lógico, o domínio técnico de SQL e o uso de um sistema gerenciador de banco de dados.

Equipes

Os projetos poderão ser feitos por equipe de até 3 membros. Ou seja, é possível realizar o projeto individualmente, em duplas ou em trios. Independentemente do total de participantes, o escopo do projeto e os critérios de correção serão os mesmos.

Etapas do Projeto

1. Formação da Equipe e Escolha do Conjunto de Dados

- Discentes deverão selecionar se irão desenvolver o projeto de forma individual, em dupla ou em trios.
- A equipe deverá selecionar um ou mais conjuntos de dados do portal de dados abertos do Governo Federal ou algum conjunto de dados abertos. Exemplos:
 - <https://dados.gov.br>
 - <https://dados.gov.br/dados/organizacoes/visualizar/instituto-brasileiro-c>
- Equipes distintas não poderão selecionar o mesmo conjunto de dados.
- Após formar a equipe e selecionar a base de dados, preencher o formulário: <https://forms.gle/zPvuxey2tdECNv9cA> até 02/06/25

2. Modelo lógico

- Criar o modelo lógico no MySQL Workbench, com as entidades, atributos, relacionamentos e cardinalidades, com clareza e correção conceitual do conjunto de dados abertos selecionados.

3. Dicionário de Dados

- Elaborar um dicionário de dados contendo o nome de cada tabela, a descrição de cada campo, seu tipo, unidade de medida (se aplicável), e eventuais observações relevantes.

4. Modelo físico

- Criar o modelo físico, com a estrutura do banco em SQL (CREATE TABLE), utilizando tipos de dados adequados expostos no dicionário e implementando restrições de integridade (chaves primárias, estrangeiras, not null, etc.).

5. População das tabelas

- Preencher as tabelas com os dados reais extraídos e tratados a partir dos arquivos abertos selecionados.
- É importante que sejam inseridos ao menos 10.000 registros no banco de dados a partir do conjunto de dados selecionado.

6. Consultas SQL

- Elaborar ao menos as seguintes consultas SELECT, sendo que:
 - 3 consultas com filtros (WHERE) e uso de funções agregadas (AVG, COUNT, MAX, etc.);
 - 2 consultas com o uso do operador LIKE;
 - 1 consultas com uso de colunas do tipo DATA e/ou TEMPO;
 - 1 consulta com subconsulta;
 - 2 consultas livres, obrigatoriamente utilizando GROUP BY e/ou ORDER BY;
- Elaborar ao menos 2 consultas de atualização, sendo que:
 - 1 consulta deve atualizar apenas os dados de uma tabela
 - 1 consulta deve atualizar os dados em 3 tabelas usando as chaves estrangeiras
- Elaborar ao menos 2 consultas de remoção, sendo que:
 - 1 consulta deve remover os dados de uma tabela
 - 1 consulta deve remover os dados de ao menos 2 tabelas usando as chaves estrangeiras

7. Apresentação do projetos

- A equipe deverá realizar a apresentação do projeto, expondo a motivação do projeto, o conjunto de dados escolhida, o modelo lógico, o modelo físico, o dicionário de dados e as consultas SQL produzidas.

Requisitos

- O banco de dados deverá possuir ao menos 10 tabelas e no máximo 20 tabelas.
- O banco de dados deverá estar normalizado.

- Todas as tabelas deverão possuir algum registro armazenado.
- Todos os dados inseridos nas tabelas deverão ser oriundos do conjunto de dados selecionados. Ou seja, não deve-se gerar dados sintéticos ou dados artificiais por meio de inteligências artificiais generativas.
 - Apenas poderão ser usados dados sintéticos nas consultas de atualização.
- O banco de dados deverá, obrigatoriamente, possuir dados no formato textual, numérico e data.
- Todos os artefatos produzidos devem ser focados para utilização no MySQL Workbench.

Diretrizes

O banco de dados desenvolvido deve conter **no mínimo 10.000 registros**, distribuídos de forma equilibrada entre as diferentes tabelas. Essa distribuição deve representar um modelo de dados coerente e realista, permitindo a realização de análises significativas.

Sobre a distribuição dos dados

- Evite a concentração de registros em uma única tabela (exemplo: 9.990 em uma e apenas 1 nas demais).
- Os dados devem refletir **relacionamentos consistentes entre as tabelas**, possibilitando consultas com subconsultas, filtros, agregações e agrupamentos.
- A qualidade e o equilíbrio da base de dados serão critérios de avaliação, principalmente no que se refere à capacidade analítica do banco e à fidelidade do modelo relacional.

Consultas SQL

As equipes devem desenvolver **consultas que respondam a perguntas relevantes** relacionadas aos dados utilizados, com foco em gerar **informações úteis para a sociedade**.

- **Consultas sem propósito analítico não serão aceitas.**
Exemplo de consulta inválida: Somar todos os IDs da tabela de cidades.
Essa operação não traz nenhuma informação útil, pois se baseia em um campo técnico.
- **Consultas com valor informativo serão valorizadas.**
Exemplo de consulta válida: Listar o total de cidades que perderam população entre 2010 e 2022.
Esta análise gera um dado relevante para reflexão e tomada de decisão.
- O objetivo é **extrair conhecimento, identificar padrões e levantar hipóteses** com base nos dados, sempre com foco em um propósito informativo e socialmente relevante.

Uso das Tabelas

- **Todas as tabelas** do banco devem ser utilizadas em pelo menos uma consulta.
- **Consultas redundantes ou com pequenas variações** (como apenas trocar colunas exibidas) não serão consideradas válidas.
- As consultas devem explorar **o modelo relacional de forma completa**, demonstrando domínio sobre os dados e suas inter-relações. Ou seja, não basta concentrar todas as consultas em apenas uma tabela.

Caso tenha dúvidas, procure o professor. Boa Sorte!