

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 o processo apresentado em aula. Em particular, deverão ser apresentados o diagrama entidade-relacionamento, bem como o esquema relacional referente ao banco de dados projetado. O banco de dados modelado deverá ser instanciado com dados reais. Consultas analíticas deverão ser formuladas e testadas sobre o banco de dados e discutidas quanto à sua eficiência.

Características básicas do banco de dados (até 30%): Para definição do tema, cada grupo deverá escolher um conjunto de dados do portal <http://dados.gov.br>. Para o conjunto de dados escolhido, deverão ser apresentados o diagrama ER e o esquema relacional 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, e ao menos 1 com um atributo multivalorado.
- Pode ser necessário alterar os dados originais para atender a esses requisitos.

Consultas (até 30%): 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.

Características avançadas (até 40%): Para avaliação de eficiência, cada uma das 10 consultas especificadas deverá ser formulada de pelo menos 2 formas diferentes considerando, por exemplo, opções como utilização ou não do operador JOIN, utilização ou não de consultas aninhadas, criação de índices para determinadas colunas, entre outras. Para deixar os dados e as consultas ainda mais interessantes, o banco pode armazenar também dados de tipo não-convencional, tais como texto, imagens, som, vídeo, etc. Além das 10 consultas obrigatórias, podem-se especificar consultas do tipo relatório (por exemplo, se for um banco de dados de indicadores sociais, um relatório seria composto por indicadores agrupados por região geográfica ou por grupos demográficos).

PONTOS EXTRA: *Serão concedidos pontos extra aos grupos que programarem uma interface gráfica (e.g., um website) para visualização interativa dos dados modelados por meio das consultas produzidas.*

OBSERVAÇÕES:

1. O trabalho deverá ser feito em **grupo de 2 ou 3 pessoas**;
2. O **SGBD MySQL** deverá ser utilizado para implementação do banco de dados e execução das consultas;
3. A avaliação compreenderá o relatório final entregue e a apresentação em aula. Além da funcionalidade básica prevista, a avaliação irá considerar a criatividade e a eficiência das consultas formuladas.

CALENDÁRIO:

29/05 (seg): Entrega da proposta: definição do grupo e do tema (**máx. 1 página; em aula**)

A proposta deverá descrever o conjunto de dados escolhido do portal <http://dados.gov.br>. Para evitar temas duplicados, informe seu grupo e dataset escolhido na seguinte planilha: <https://goo.gl/oy9ozZ>. Na proposta escrita (1 página), deverão ser descritas as entidades e relacionamentos presentes, o número de instâncias de cada entidade e relacionamento, e possíveis consultas a serem formuladas sobre esses dados.

14/06 (qua): Entrega do relatório final (**máx. 10 páginas; sala 5309, A/C Prof. Marcos até 17h**)

O relatório final deve descrever as características do banco de dados criado e as consultas realizadas. Essas consultas devem ser definidas através de uma explicação textual e do respectivo comando SQL, acompanhado do resultado da consulta e do tempo de execução em comparação a consultas alternativas. Deve conter também uma autoavaliação dos membros do grupo, incluindo a descrição das atividades realizadas individualmente. Evite desperdício de papel: utilize a frente e o verso de cada página, evite usar capa.

14/06 (qua), 19/06 (seg), e 21/06 (qua): Apresentação dos trabalhos

Cada grupo terá em torno de 15 min (a confirmar, dependendo do número de grupos) para apresentar o tema do trabalho, a modelagem desenvolvida (conceitual e relacional), uma seleção das várias consultas formuladas e de seus resultados, e uma descrição do trabalho desenvolvido por cada membro do grupo.