



Trabalho Prático – Análise de Dados Abertos

Objetivo: O objetivo deste trabalho é projetar e implementar um banco de dados relacional para uma aplicação de análise de dados abertos governamentais. O projeto do banco de dados deverá seguir a abordagem *top down* vista em aula. Na fase de projeto conceitual, será criado um diagrama entidade-relacionamento (ER) e as restrições de integridade necessárias que descrevam de forma precisa o banco de dados projetado. Na fase de projeto lógico, esse diagrama será mapeado para um esquema relacional que descreva a estrutura de cada uma das relações (tabelas) que comporão o banco de dados e as restrições de integridade necessárias para garantir que esse esquema relacional seja uma representação do esquema conceitual inicialmente criado. Definido o esquema relacional em nível lógico, no projeto físico, o banco de dados correspondente deverá ser criado (usando SQL). A seguir, o banco de dados deverá ser instanciado com dados reais. Consultas em SQL deverão ser formuladas e testadas sobre o banco de dados projetado através de uma aplicação Web. Essas consultas deverão ser: especificadas de acordo com as características definidas a seguir e relevantes para a aplicação analítica considerada.

Características básicas do banco de dados (40%): Para definição do tema, cada grupo deverá escolher um conjunto de dados do Portal Brasileiro de Dados Abertos (<http://dados.gov.br>). Para o conjunto de dados escolhido, deverão ser apresentados o diagrama ER + restrições de integridade e o esquema relacional correspondentes, atendendo aos seguintes requisitos:

- Pelo menos 4 tipos de entidade, cada um tipo deverá possuir, além de do seu atributo identificador (chave), pelo menos 2 outros atributos;
- Pelo menos 3 tipos de relacionamento, sendo pelo menos 1 deles com cardinalidade N:N e incluindo pelo menos um atributo;
- Finalmente, pelo menos um tipo de entidade deverá incluir um atributo multivalorado e um atributo composto.

Características básicas das consultas (40%): Deverão ser especificadas e executadas um total de 10 consultas em SQL, sendo:

- 1 consulta envolvendo apenas as operações de seleção e projeção;
- 2 consultas envolvendo a junção de apenas duas relações, sendo que uma deve conter junção externa;
- 2 consultas envolvendo a junção de três ou mais relações;
- 1 consulta envolvendo uma das operações sobre conjuntos (união, diferença ou intersecção);
- 2 consultas envolvendo funções de agregação;
- 2 consultas envolvendo subconsultas aninhadas.

Características avançadas (20%): O banco deve armazenar também dados de tipo não-convencional, tais como texto, imagens, som, vídeo, etc. Além das 10 consultas obrigatórias, especificar ao menos 2 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). Além disso, deve-se programar uma aplicação Web que acesse o banco de dados produzido para realizar as consultas elaboradas pelo grupo e apresente os resultados retornados.

Requisitos:

- O trabalho deverá ser feito em grupos de **4 alunos**;
- Recomenda-se o uso da ferramenta **BrModelo** para a modelagem dos dados;
- O **SGBD MySQL** deverá ser utilizado para implementação do banco de dados e execução das consultas;
- A linguagem de programação para o desenvolvimento da aplicação Web poderá ser escolhida pelo grupo (por ex.: Java, PHP, Python, etc.);
- A avaliação compreenderá o relatório final entregue e a apresentação em aula. Além da funcionalidade básica prevista, a avaliação irá considerar também a criatividade do grupo em desenvolver funcionalidades extras.

Calendário:

28/06/16 (terça-feira): Data limite para enviar ao grupo de e-mails da disciplina (bd-2016-1@googlegroups.com) os seguintes dados:

- Componentes do grupo de trabalho (nome e DRE dos alunos);
- Tema a ser trabalhado pelo grupo e conjunto de dados escolhido do Portal Brasileiro de Dados Abertos (conjuntos de dados disponíveis em: <http://dados.gov.br/dataset/>)
 - Para facilitar a identificação, indicar o nome e a URL do conjunto de dados selecionado

Cada grupo deverá selecionar um conjunto de dados diferente, no caso de mais de um grupo indicar o mesmo, o grupo que enviar o e-mail primeiro terá preferência, devendo o(s) outro(s) grupo(s) efetuar(em) uma nova escolha.

24/07/16 (domingo): Entrega do relatório final para o e-mail da professora (giseli.ufrj@gmail.com). O relatório deve descrever o trabalho desenvolvido em formato de artigo científico. O artigo deve conter no máximo 10 páginas e seguir o *template* para artigos da Sociedade Brasileira de Computação - SBC (<http://www.sbc.org.br/documentos-da-sbc/summary/169-templates-para-artigos-ecapitulos-de-livros/878-modelosparapublicaodeartigos>).

O relatório final deve apresentar e descrever a modelagem do banco de dados criado e as consultas realizadas. Essas consultas devem ser definidas através de uma explicação textual e do respectivo comando SQL. O relatório também deve incluir descrições e explicações sobre a aplicação Web desenvolvida. Além disso, é solicitado que sejam indicadas as atividades realizadas por cada um dos membros do grupo.

26 e 28/07/16 (terça e quinta-feira): Apresentação dos trabalhos desenvolvidos. 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, explicações e demonstração da aplicação Web desenvolvida.