



Banco de Dados (BD1) 2024.2  
Profs Veronica dos Santos (3WA),  
Sérgio Lifschitz (3WB) e Marcos Villas (3WC)  
E-mail: <vdsantos,sergio,villas>@inf.puc-rio.br  
URL: <http://home.ead.puc-rio.br/>  
Atendimento: Discord / Sala 414 RDC (SL)

## Trabalho Prático – Grau G1

**Prazo limite: até domingo 27/outubro/2024 12h**

Com o objetivo de praticar ao máximo os conceitos e fundamentos de sistemas de bancos de dados relacionais, os professores de INF1383 (BD1), Veronica, Villas e Sérgio, convidam os alunos da disciplina no segundo semestre de 2024 para projetar e construir uma base de dados relacional experimental que diz respeito ao mercado de EDUCAÇÃO SUPERIOR NO BRASIL.

Nas últimas duas décadas, o setor de Educação Superior no Brasil passou por um crescimento significativo, especialmente no número de **instituições privadas**. Esse aumento foi impulsionado pelo crescimento da demanda por qualificação profissional e pela ampliação das **políticas de inclusão educacional**. Além do ensino presencial, o **ensino a distância (EAD)** teve uma expansão acelerada, tornando-se uma opção importante para estudantes que buscam flexibilidade e menores custos. Porém a **taxa de evasão** em muitos cursos de graduação ainda é alta.

Anualmente, o INEP realiza o **Censo da Educação Superior no Brasil**. O objetivo da pesquisa é obter um panorama completo sobre a Educação Superior no Brasil, retratando as instituições de educação superior, públicas e privadas, que ofertam cursos de graduação e sequenciais de formação específica, bem como o perfil de seus alunos e docentes.

Os registros das instituições contemplam seus cursos e locais de oferta, infraestrutura das instituições de educação superior, vagas oferecidas, candidatos, matrículas, ingressantes, concluintes e docentes, nas diferentes formas de organização acadêmica e categoria administrativa. O perfil dos docentes e discentes revela questões sobre gênero, cor/raça, idade e nacionalidade e o curso de interesse. Para os discentes ainda é necessário registrar informações socio-econômicas, se são bolsistas (FIES e PROUNI) e PCD.

Com base nestes dados, gestores públicos e também privados podem extrair informações estatísticas para as atividades de acompanhamento e avaliação, realizar comparações entre cursos e entre instituições, elaborar programas de expansão, redução da evasão e de melhoria da qualidade desse nível de ensino, entre outros.

São muitas dimensões possíveis a explorar e não há propriamente um limite: cada grupo decidirá o alcance e detalhamento que pretende modelar e especificar. Visando permitir muita liberdade criativa aos alunos e seus grupos, propositalmente não há definição de quais dados e quais consultas e manuseios devem ser realizados.

Nesta **primeira parte do trabalho**, contando 20% do grau G1, cada grupo deverá propor um texto base que orientará a definição dos requisitos de dados e funcionais. Em seguida, o grupo deve modelar conceitualmente o que será o banco de dados a ser construído, e que será mapeado depois para um esquema relacional, visando dar suporte à persistência e às consultas e requisições diversas envolvendo dados neste domínio da Educação Superior no Brasil. Além de criar tabelas, será necessário incluir e atualizar alguns dados, validando as restrições de integridade estruturais. É necessário também sugerir consultas úteis e diversificadas, com expressões em linguagens relacionais conhecidas, como álgebra e SQL, exemplificando possíveis acessos úteis e os mais realistas possíveis, por potenciais usuários de um sistema envolvendo este tema particular.



Banco de Dados (BD1) 2024.2  
Profs Veronica dos Santos (3WA),  
Sérgio Lifschitz (3WB) e Marcos Villas (3WC),  
E-mail: <vdsantos,sergio,villas>@inf.puc-rio.br  
URL: <http://home.ead.puc-rio.br/>  
Atendimento: Discord / Sala 414 RDC (SL)

**Pede-se: ENTREGAR um relatório (como PDF!) em uma tarefa no EAD de INF1383 contemplando este trabalho, contendo as modelagens, definições e expressões utilizadas para as seguintes questões:**

**(Questão 1 – 0.5 ponto)** Propor um texto-enunciado que sirva de base para uma especificação formal de um banco de dados relacional sobre o tema EDUCAÇÃO SUPERIOR NO BRASIL. Usar algumas das observações aqui colocadas apenas como base para a criação do teu sistema.

Este texto deverá, necessariamente, ser utilizado como referência para todas as demais questões deste trabalho, mantendo a consistência de informações. Sugerimos que os alunos realizem buscas na WWW de notícias sobre o tema, consultando também o site do INEP e do Instituto SEMESP.

**(Questão 2 – 2.0 pontos)** Modelar conceitualmente esta base de dados de EDUCAÇÃO SUPERIOR NO BRASIL utilizando o Modelo de Entidades e Relacionamentos (MER) (não usar UML ou outros!), considerando necessariamente a sua representação diagramática que contempla as cardinalidades min-max para os relacionamentos entre entidades. Justifique suas escolhas quando houver alternativas de modelagem

Utilizar obrigatoriamente a ferramenta CASE (*Computer Aided Software Engineering*) chamada de **BRmodelo** (gratuita, versões Desktop-Java ou Web) para gerar o diagrama ER. Além de mostrar no diagrama, listar também, em separado, os atributos das entidades e/ou relacionamentos, descrevendo-os textualmente e conceitualmente, sem preocupações com detalhes de implementação. Devem ser especificados (1) qual domínio de dados para valores considerados válidos (e.g numérico, alfanumérico, alfabética, dia-mes-ano, apenas ano, numérico limitado entre 0 e 10, etc. e (2) a descrição de cada atributo, não apenas seu nome, pois nem sempre está clara a semântica associada.

**(Questão 3 – 1.0 ponto)** Enunciar pelo menos 3 (três) restrições de integridade semânticas (regras de negócio ou *business rules*), que você ou seu grupo entenda que deveriam ser controladas no sistema. Cuidado para não confundir com regras que podem ser controladas por restrições estruturais do modelo relacional! Têm de ser regras que independem do modelo de dados escolhido.

**(Questão 4 – 1.0 ponto)** Propor um esquema lógico-relacional que seja decorrente (mapeado!) do esquema conceitual ER definido na questão 1. Listar as tabelas, atributos e restrições de integridades estruturais do modelo relacional, como as chaves primárias e estrangeiras. O mapeamento deve ser feito usando as regras propostas pelos professores seguindo a proposta do livro Projeto de Banco de Dados. Quando houver mais de uma alternativa para o mapeamento é importante justificar a escolha.

**(Questão 5 – 1.5 ponto)** Listar os códigos SQL DDL utilizados para criação de todas as tabelas do esquema lógico da questão 2. Os códigos devem ser implementados por vocês no servidor PostgreSQL de BD1, com a conta de usuário do seu grupo. Apenas restrições estruturais (relacionais,



Banco de Dados (BD1) 2024.2  
Profs Veronica dos Santos (3WA),  
Sérgio Lifschitz (3WB) e Marcos Villas (3WC)  
E-mail: <vdsantos,sergio,villas>@inf.puc-rio.br  
URL: <http://home.ead.puc-rio.br/>  
Atendimento: Discord / Sala 414 RDC (SL)

diretamente definidas na criação das tabelas) devem ser consideradas aqui. Faz parte do exercício saber definir estas restrições.

Considerando que o esquema relacional conterá pelo menos 3 relações ou tabelas, para pelo menos uma destas tabelas o comando *ALTER TABLE* deve ser usado para adicionar a restrição de chave, com a definição da chave primária (PK). Escolher uma outra tabela, diferente daquela utilizada para a restrição de chave, para usar o comando *ALTER TABLE* na definição da(s) chave(s) estrangeira(s).

Escolher tipos de dados adequados para controle das restrições de integridade. Devem ser evitados tipos genéricos (char, number...) para domínios especializados, por exemplo, para datas.

**(Questão 6 – 1 ponto)** Listar os códigos SQL DML de inserção de dados (novas tuplas) em todas as tabelas. Por favor, considerar dados realistas e possíveis, por exemplo, obtidos de datasets públicos como por exemplo no site do **INEP**. Para cada tabela, exiba no relatório para esta questão, um exemplo de inserção de dados realizados. O script completo em SQL para inserção de todos os dados não deve fazer parte do relatório! Entretanto, é necessário referenciar em separado um link para o script completo que corresponda ao conjunto de todos os dados presentes no servidor de BD1.

*Dica: conjuntos de dados volumosos (centenas ou mesmo milhares de tuplas) podem ser bonificados pelos professores na avaliação do trabalho, pois serão úteis também na parte 2 do trabalho – grau G2.*

**(Questão 7 – 2.0 pontos)** Sugerir consultas em ALGEBRA RELACIONAL que sejam úteis para este sistema de bancos de dados. Propor um texto em português e a solução em linguagem álgebra correspondente.

Devem ser exibidas ao menos 5 (cinco) consultas de forma que todas as tabelas sejam contempladas em pelo menos uma das consultas sugeridas por você e seu grupo, e cada consulta contenha pelo menos um operador da álgebra relacional distinto das demais consultas. Dos operadores de conjuntos (UNION, INTERSECTION e EXCEPT) usar apenas um deles e um dos exemplos deve contemplar a divisão relacional.

**(Questão 8 – 1.0 ponto)** O mesmo da Questão 7, porém, agora, são pelo menos 3 (três) consultas para a linguagem SQL (parte 1). Propor um texto em português e a solução em SQL correspondente usando somente uma tabela na cláusula FROM.

Procurem evitar fazer “versões SQL das questões de álgebra”! Usem o máximo de criatividade possível para definir consultas distintas com linguagem relacional distinta! Assim como para álgebra, é necessário utilizar os diversos operadores da linguagem SQL apresentadas nas aulas, de forma a praticar o uso da linguagem e, conseqüentemente, aprender sobre como é o comportamento quando usamos SGBDs de mercado. Cada consulta pode ter alguns operadores repetidos mas pelo menos um deles diferente. Também é importante explorar o conceito de sub-consulta. No caso de SQL, as consultas devem ser implementadas e validadas no servidor PostgreSQL de BD1, em contas específicas para cada grupo, e devem retornar algumas tuplas.