

Instruções para execução do projeto e escrita do relatório

O relatório deve seguir as **regras de escrita recomendadas** no documento de mesmo nome (postado no *Moodle*). O projeto deve conter os seguintes itens:

1. Capa

As seguintes informações devem ser fornecidas na capa do projeto: o nome da instituição, o nome do curso, o nome da disciplina, o nome do professor responsável, o nome do projeto e o **número e nome do tema, o nome dos participantes e a fase/data de entrega do projeto**.

2. Índice ou sumário

O índice deve enumerar os títulos das seções que compõem o projeto e as suas respectivas páginas de início. As seções devem ser numeradas.

3. Breve descrição do minimundo

Descreva brevemente sobre o problema a ser resolvido. **Defina um mínimo de 3 tabelas para o esquema de BD do tema atribuído. Foque no conjunto de relações/tabelas que vai, de fato, ser utilizado neste projeto e confira se está cumprindo com os requisitos na Seção 3.1.** Apresente o esquema de BD final (modelo relacional).

3.1 Especificação de Consultas em SQL

Cada grupo terá como diferencial um tema: conjunto de duas consultas. **As consultas deverão satisfazer os critérios abaixo apresentados.**

Nessa seção, devem ser especificados os enunciados das duas consultas do projeto. **Para cada uma das consultas do projeto, precisam ser definidos: os campos de busca, os operadores de comparação e os campos de visualização dos resultados.**

A SQL deve ser apresentada para as consultas. Se necessário, os valores parametrizados podem ser especificados entre `< >`, indicando que os valores podem variar (são variáveis), como no exemplo abaixo:

Exemplo: recupere todos os dados dos clientes dada uma faixa etária.

```
SELECT * FROM cliente  
  
WHERE idade > <idade_ini> AND idade <= <idade_fim>
```

Os grupos que apresentarem exatamente as mesmas consultas e/ou interfaces serão penalizados: **não serão aceitas consultas idênticas de diferentes grupos.**

As consultas têm por objetivo permitir a busca a partir de qualquer uma das palavras/valores do conjunto de campos da condição (**WHERE**). A **busca absoluta** utiliza **expressões exatas**, que são as palavras especificadas pelo usuário. A consulta de uma **busca relativa** é construída através de **expressões regulares**, pois obtém resultados que contêm (parcial ou totalmente) as palavras especificadas pelo usuário.

As consultas devem satisfazer:

1. cinco (5) ou mais campos de visualização do resultado (cláusula **SELECT**).

2. Envolver duas ou mais tabelas com condições de seleção (uso da cláusula `where` não apenas para as condições de junção) e/ou aninhamento de consultas (subconsultas) **E**
3. ser parametrizadas (condições no **WHERE** com parâmetros) para se tornarem *stored procedures* **E**
4. 1 de busca absoluta e 1 de busca relativa **E**
5. uma das consultas deve incluir algum resultado numérico ou ordenação, a partir do qual possa ser construído um *ranking*.

No relatório é obrigatório especificar, para cada consulta, os campos da busca e os da visualização dos resultados. Para os campos de busca, deve ser especificado se a busca será absoluta ou relativa. Também, pode se pensar em acrescentar operadores de comparação (`>` ou `>=` ou `==` ou `<` ou `<=` ou `<>`) nas condições de busca das consultas. As decisões sobre as consultas têm de ser refletidas na interface Web.

Exemplo: *recupere nome, telefone e endereço dos clientes dada uma faixa etária e um trecho do nome.*

Supondo o esquema da tabela: `clientes(cpf, nome, telefone, endereço, datanasc, idade)`

Campos de visualização do resultado: `nome, endereço, telefone, datanasc, idade`

Campos de busca (ou das condições): `idade (absoluta), nome (relativa)`

Operadores das condições: `idade (>=, <=), nome (ILIKE)`

4. Populando o BD

Nesta seção especifique quantos registros têm cada tabela do BD e o seu tamanho (em *bytes*). Para isto, crie uma tabela com três colunas: nome da tabela, número de registros, tamanho (*bytes*). Destaque, também, se os dados foram gerados com *scripts* de autoria própria ou com alguma ferramenta de geração de dados. Caso seja usada ferramenta(s) de geração automática de dados, mencione qual(is) no relatório.

Não há restrições para popular o BD. Poderá ser populado utilizando *script* ou ferramenta de geração de dados sintéticos para todas as tabelas que envolvem os dados das duas consultas.

O volume de armazenamento mínimo é 500.000 tuplas por tabela. **Recomendação:** cada grupo deve conversar com a professora para discutir quais as tabelas com esta restrição.

Ferramentas recomendadas para geração de dados sintéticos:

- BigQuery - bigquery.cloud.google.com
- Parallel Data Generation Framework (PDGF) - <http://www.paralldatageneration.org/>
- Generate data - <http://www.generatedata.com/>
- SQL Data Generator (*trial*) - <http://www.red-gate.com/products/sql-development/sql-data-generator/>
- Database Test Data Generator - <http://www.databasetestdata.com/>
- Datanamic Data Generator MultiDB (*trial*) - <http://www.datanamic.com/datagenerator/index.html>

- Mockaroo - <https://www.mockaroo.com/>
- Fake Name Generator - <https://www.fakenamegenerator.com/>

O **script final do esquema de BD deve ser incluído na entrega** e postado no *Moodle*. Caso o *backup* ficar com mais de 50Mb, deverá ser compartilhada a área onde vai ficar disponível o *backup* do BD e entregue o *link* pelo *Moodle*. Caso use *scripts* de autoria própria, insira os mesmos junto com o *backup* do BD, da forma especificada na respectiva tarefa do *Moodle*.

5. Técnicas de acesso eficiente ao Banco de Dados

Para a otimização das consultas, precisam ser definidos as estruturas de indexação e outros recursos necessários para a recuperação eficiente dos resultados das consultas.

Essa seção deve apresentar as técnicas de indexação e otimização empregadas, descrevendo as estruturas de indexação criadas e outros recursos empregados para a recuperação eficiente dos resultados das consultas com suas respectivas **justificativas**. Para utilizar outros recursos, leia o material de sintonização do projeto físico de BD.

- Especifique a versão do PostgreSQL (`select version()`) que está sendo usada. As novas versões podem implementar melhoras no desempenho do otimizador, acarretando assim em diferentes planos para as consultas.
- As consultas devem ser executadas várias vezes para observar o tempo de processamento (recomendado: 5 vezes). O relatório precisa especificar quantas amostragens foram executadas e a média de tempos de execução deve ser apresentada. Lembre-se de limpar o *buffer* de memória (memória cachê) antes da execução de uma consulta, pois a consulta, quando rodada de novo já está carregada na memória e vai ser mais rápida, e se não estiver?

Limpando o cachê do PostgreSQL no Linux

```
#!/usr/bin/sudo bash

service postgresql stop
sync
echo 3 > /proc/sys/vm/drop_caches
service postgresql start
```

Limpando o cachê no Windows

```
programa RAMMap v1.52
https://docs.microsoft.com/en-us/sysinternals/downloads/rammap
```

Como parte da justificativa, para cada consulta incluir: uma **tabela comparativa** contendo as versões da consulta e a média do tempo de processamento para cada uma, e a *% de diferença entre as diferentes versões da mesma consulta*. A Tabela 1, abaixo, é só um exemplo de como podem ser expostos os resultados.

	Consulta Inicial	Consulta Otimizada	Diferença (%)
Tempo de Execução	973.005ms	44.616ms	95,41

Tabela 1. Exemplo de apresentação de resultados de tempo de execução

Fórmula para cálculo do % = $((V_{ot} - V_{ini})/V_{ini} \times 100)$

- O **desafio** nesta fase é o tempo de execução das consultas. Isto é, conseguir que o tempo de execução seja abaixo de 3 segundos ou que a diferença de execução na Tabela 1 seja de mais de 60%.
- Os planos das consultas precisam ser incluídos, com uma discussão/explicação sobre o mesmo. Para cada consulta, apresentar o plano da versão inicial e da versão com melhor desempenho.
- Devem ser incluídos comentários sobre quais técnicas foram usadas para melhorar o desempenho. Exemplos: variações algébricas, uso de operadores, uso de junção em contraposição de aninhamento de consultas, uso de UNION vs. operador lógico OR, etc.
- Outros comentários: [Leia com atenção a Seção 5.1.](#)

5.1 Considerações para otimizar as consultas

Apenas quando aplicável: um breve relato das transformações realizadas para cada uma das relações iniciais do banco de dados, até o esquema final de cada tabela do BD (**se necessário**). Para tanto, devem ser discutidos os seguintes fatores:

- Os tipos de dados dos campos PK e FK de uma tabela influenciam no tempo de execução ou plano de execução da consulta?
- Justificar ações inusuais ou de desnormalização que tenham sido implementadas.

6. Programação com Banco de Dados

Essa seção deve apresentar as descrições das funções, com a descrição dos parâmetros de entrada/saída. Deve ser acrescentado o *script* documentado de cada um.

Como parte da programação em BD, é **obrigatório** a definição e construção de:

- duas *stored procedures* (que devem corresponder às consultas definidas). Em qualquer caso, explicitar no relatório qual a função de cada uma.

7. Controle de acesso de usuários

Essa seção deve descrever as diferentes visões de usuários do BD e sua relação com os usuários da sua aplicação. Para isto, criem o **grafo/diagrama de concessões de usuários** e uma tabela como no exemplo da Tabela 2, abaixo.

	gerente	padeiro	vendedor	admin
produto	S, I, U	S, I, U	S, I, U	Sem acesso
encomenda	S, I, U, D	S, I, U, D	S, I, U, D	Sem acesso
cliente	S, I, U	S, I, U	S, I, U	Sem acesso
funcionário	S, I, U	Sem acesso	Sem acesso	S(cpf_funcio, senha, cargo, nome_funcio)
pedido	S, I, U, D	S, I, U, D	S, I, U, D	Sem acesso
possui	S, I, U, D	S, I, U, D	S, I, U, D	Sem acesso
produz	S, I, U	S, I, U	S	Sem acesso

Tabela 2. Exemplo de perfis de usuários, seu papel e privilégios
(SELECT (S), INSERT (I), DELETE (D) e UPDATE (U))

Deve ser acrescentado o *script* documentado dos usuários criados, acompanhado do *script* das autorizações concedidas a cada um.

Descrever, também, como deveria ser feita a conexão de uma aplicação com o BD e qual usuário de BD, dos propostos, seria utilizado para estabelecer a conexão com cada tela (inicial, de CRUD, de consulta, etc.). Criem o gráfico da Figura 1 para ilustrar a descrição. **Não é para fazer uma aplicação! e sim para esboçar como o controle de acesso foi planejado para uma aplicação.**

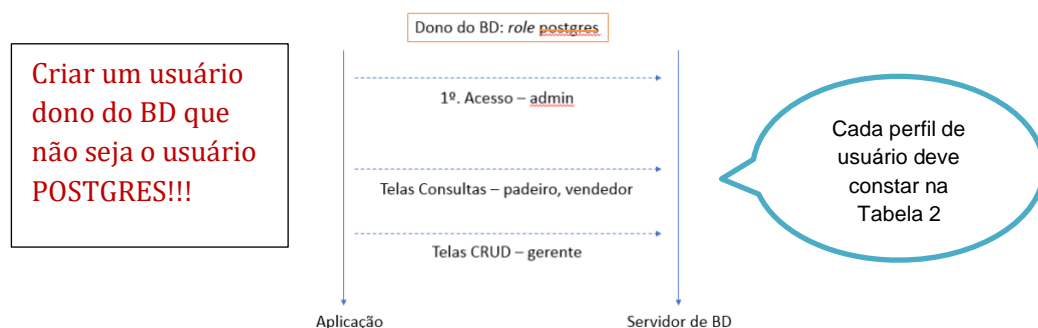


Figura 1. Conexões usando perfis de usuários

8. Considerações finais: comentários próprios sobre o desenvolvimento do projeto, experiências, dificuldades encontradas e observações que permitam estabelecer uma comparação dos resultados entre as versões inicial e final do projeto.

Referências Bibliográficas OU Bibliografia

Não é obrigatório!

Use **Referências** Bibliográficas se o material listado for citado no texto do relatório. Use **Bibliografia** se foi utilizado para desenvolver o trabalho, mas não foi citado no texto do relatório.

Entrega em cada fase do projeto

O cronograma de entrega do relatório acompanha o cronograma do projeto.

- **Fase intermediária:** entrega até a Seção **5**.
- **Fase Final:** entrega até a Seção **8**.

ATENÇÃO: NA ENTREGA DA SEGUNDA PARTE DO PROJETO, DEVEM SER ENTREGUES AS CORREÇÕES PROPOSTAS PELA PROFESSORA REFERENTE À PRIMEIRA PARTE DO PROJETO. DE FORMA SIMILAR, NA ENTREGA DAS FASES RESTANTES (SE FOR O CASO).
