UNIVERSIDADE FEDERAL DE MINAS GERAIS

INSTITUTO DE CIÊNCIAS EXATAS

DEPARTAMENTO DE CIÊNCIA DA COMPUTAÇÃO

DISCIPLINA: INTRODUÇÃO A BANCO DE DADOS - DCC011

TRABALHO PRÁTICO 2

Integrantes:

Bernardo Figueiredo Cipriano

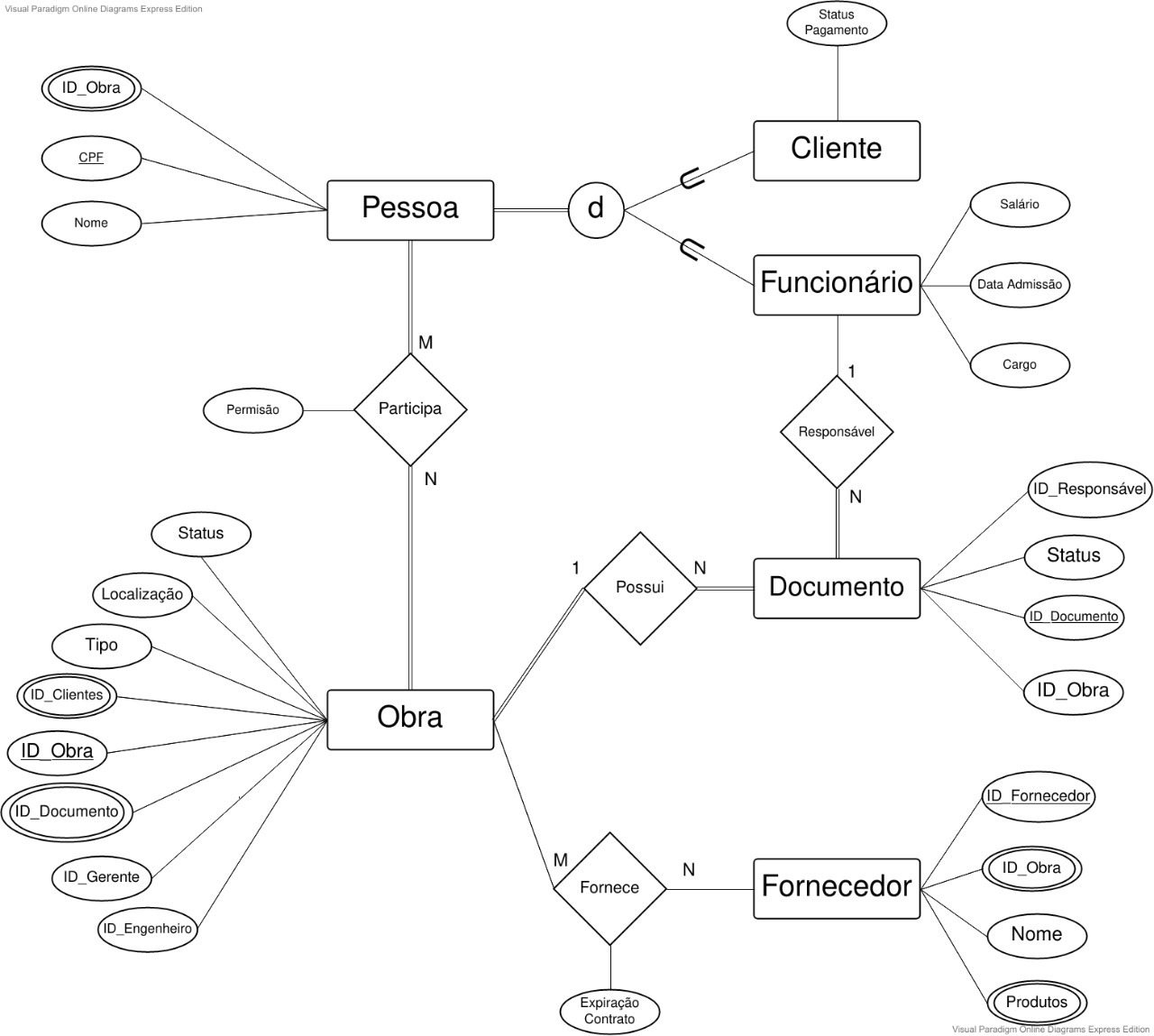
Gabriel Saraiva Espeschit

Leandro Luiz Duarte Teixeira

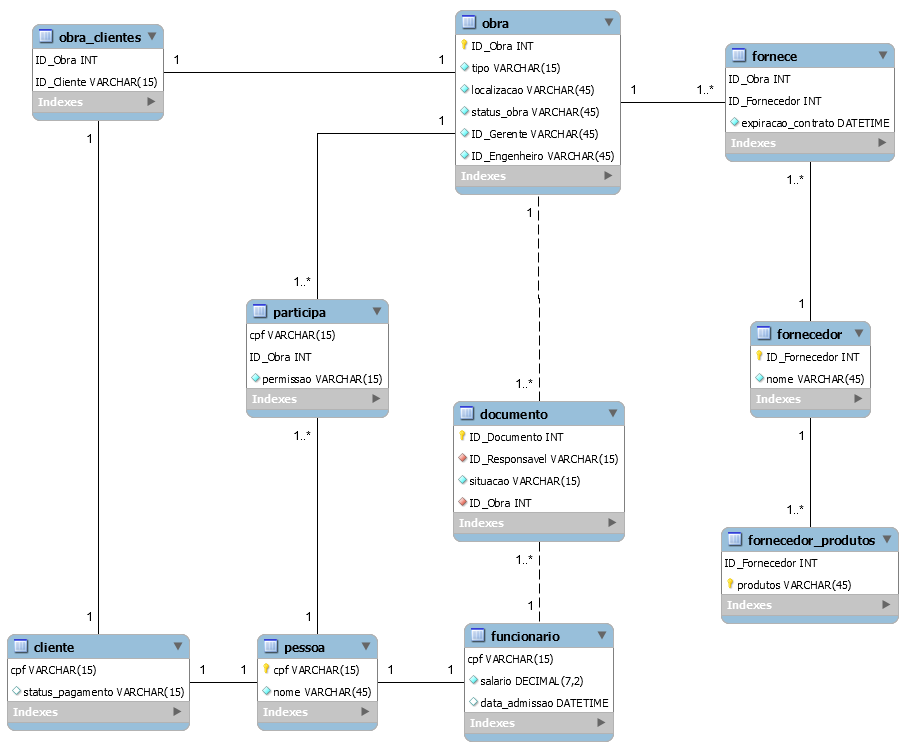
Thiago Lages Rocha

Victor Portugal de Vasconcellos

**Projeto Conceitual**



**Projeto Lógico**



Este é o projeto lógico que fizemos a partir do Diagrama Entidade-Relacionamento apresentado acima. A modelagem do fornecedor foi a mais objetiva; ele possui um ID e um nome na própria tabela. Para implementar os produtos, foi usada uma tabela fornecedor-produtos. Foi cogitado também criar uma tabela só para os produtos e fazer disso uma relação M x N. Essa ideia, porém, foi descartada; fazer isso exigiria uma padronização da forma de apresentação dos produtos.

Da forma como implementamos, um fornecedor está livre para fornecer “cimento” enquanto um outro pode optar por fornecer apenas “cimento da marca Mauá”, sem exigir que o primeiro modifique os dados para apresentar todas as marcas e variedades que ele oferta ou que o segundo generalize o produto mesmo quando ele só vende uma marca específica que pode não ser a desejada para aquela obra específica.

A variável multivalorada “ID\_Obra” presente em Fornecedor pode ser obtida implicitamente através da tabela Fornece, que representa a relação entre o fornecedor e a obra. Ao inserir as restrições de modificação na tabela Fornece, foi decidido inserir bloqueio a mudança e cascata a remoção em ambas as chaves. A lógica é, uma vez firmado um contrato entre as duas partes e inserida a data de expiração, nenhuma outra relação pode vir a herdar essa data de expiração sem o firmamento de um novo contrato. Por fins de segurança, se for necessário trocar o ID de um fornecedor ou de uma obra, os contratos previamente definidos deverão ser cancelados e refirmados. Por outro lado, se um fornecedor ou obra deixa de existir, todos os contratos a eles relacionados podem ser removidos.

Ao modelar a obra, foi levado muito em consideração as potenciais exigências do mundo real também. Além dos dados monovalorados que são mais objetivos, toda obra deve ter ao menos um documento. Porém, inserir uma restrição de integridade forçaria os usuários do banco de dados a só inserirem uma obra no sistema uma vez que seu documento tenha sido registrado. Nós achamos melhor permitir que a obra apareça no sistema no momento em que ela tem um engenheiro e um gerente já registrados. Por outro lado, os documentos só podem ser inseridos no sistema se possuírem uma obra correspondida. E, seguindo a mesma linha do fornecedor-obra, não é possível modificar a chave de um responsável ou obra enquanto estes tiverem documentos atribuídos a eles. É possível remover uma obra com documentos, isso gerará uma reação em cascata, porém não é possível remover um funcionário responsável por documento enquanto ele ainda possuir documentos atribuídos

Situação parecida acontece com a relação obra-cliente. Uma forma alternativa de implementar seria inserir o ID da obra como chave estrangeira na tabela cliente, visto que todo cliente deve possuir uma obra relacionada (caso contrário, não deve ser considerado cliente ainda). Todavia, optamos por criar uma tabela obra-clientes justamente porque ela permite aos usuários do sistema inserir um cliente que ainda está escolhendo a obra que deseja comprar. Outra coisa importante para se notar sobre a relação obra-cliente é que há um bloqueio a remoção inserido apenas no ID\_Obra. Um cliente pode decidir desistir de uma obra a qualquer momento sem dar uma satisfação direta às respectivas obras. Porém, antes de deletar uma obra, é importante garantir que todos os seus clientes tenham sido removidos ou atribuídos a outras obras.

**Consultas**

Para carregamento do banco de dados com um volume significativo de dados, desenvolveu-se um programa em *Python*, responsável por gerar um script SQL contendo aproximadamente 50 mil comandos *insert*. O código consiste em geração aleatória de dados a partir de algumas opções fornecidas, atentando-se à unicidade dos atributos que são chaves primárias (CPF e IDs).

Abaixo estão descritas as consultas realizadas sobre esses dados. A descrição textual é acompanhada de duas opções de comando SQL, testadas a fim de determinar-se a influência da forma como a consulta é especificada sob seu desempenho (comando JOIN, consultadas aninhadas etc.).

O grupo salienta que, em função da enorme quantidade de dados e do espaço limitado do trabalho, um máximo de 10 linhas para cada consulta (de 1 a 10) está sendo representada no trabalho, porém na maioria dos casos, este número é bem superior a isso.

Na hora de formular as consultas, buscamos pensar em um conjunto que conseguisse cobrir da melhor forma possível as relações, tentando capturar as informações singulares que podem ser obtidas através de diversas tabelas.

De um ponto de vista do SQL, também tentamos selecionar ao menos uma query para testar cada uma das seguintes hipóteses:

• Se é mais barato (computacionalmente falando) fazer um join de forma explícita entre duas tabelas ou pesquisar, através da chave estrangeira, qual elemento da segunda tabela atende aos requisitos e através disso encontrar o dado correspondente na primeira tabela

• Se, para eliminar dados repetidos, é mais eficiente usar group by ou distinct.

• Se, para descobrir o maior valor em uma tabela, é melhor usar max ou colocar os dados em ordem decrescente e limitar a pesquisa para exibir apenas um dado.

• Se, para descobrir dados que estão em uma tabela e não estão na outra, é melhor usar not in ou fazer um left join na primeira e selecionar apenas os dados em que a chave estrangeira tem valor nulo.

• Os impactos potenciais de colocar uma subquery após a palavra chave from.

• Se é mais eficiente procurar um dado através da seleção de dados que atendem à condição desejada ou através da procura dos dados que não estão entre os que não atendem a essa condição.

Levando esses pontos em consideração, as queries escolhidas foram, agrupadas pelo principal critério que satisfazem, as seguintes:

Apenas seleção e projeção (2)

1 – Selecionar os CPFs dos funcionários que ganham entre 2000 e 3000 reais.

1.1 - *SELECT CPF from Funcionario WHERE salario > 2000 AND salario < 3000*

1.2 - *SELECT CPF from Funcionario WHERE CPF in (SELECT CPF from Funcionario WHERE salario > 2000) AND CPF in (SELECT CPF from Funcionario WHERE salario < 3000)*

2 – Selecionar os status dos sobrados que serão construídos em Belo Horizonte.

2.1 - *SELECT status\_obra FROM Obra WHERE tipo = 'Sobrado' and localizacao = 'Belo Horizonte'*

2.2 - *SELECT status\_obra FROM Obra WHERE ID\_Obra IN (SELECT ID\_Obra FROM Obra WHERE tipo = 'Sobrado') AND ID\_Obra IN (SELECT ID\_Obra from Obra WHERE localizacao = 'Belo Horizonte')*

*1- CPF*

|  |
| --- |
| *CPF* |
| *1001507072660* |
| *1001340174804* |
| *100100206832* |
| *1000976424525* |
| *1000899768784* |
| *100073542407* |
| *1000712211403* |
| *1000125264950* |
| *100002098563* |
| *1000013692887* |

*2- status\_obra*

|  |
| --- |
| *CPF* |
| *Iniciada* |
| *Aguardando pagamento* |
| *Entregue* |
| *Entregue* |
| *Aguardando autorizacao* |
| *Aguardando autorizacao* |
| *Aguardando pagamento* |
| *Iniciada* |
| *Aguardando autorizacao* |
| *Aguardando pagamento* |

Junção de duas relações (3)

3 – Selecionar o nome de todas as empresas que fornecem telhas.

3.1 - *SELECT nome FROM Fornecedor a JOIN Fornecedor\_Produtos b ON a.ID\_Fornecedor = b.ID\_Fornecedor WHERE produtos = 'Telha'*

3.2 - *SELECT nome FROM Fornecedor WHERE ID\_Fornecedor IN (SELECT ID\_Fornecedor FROM Fornecedor\_Produtos WHERE produtos = 'Telha')*

4 – Selecionar os nomes dos funcionários que estão há ao menos três anos na empresa.

4.1 - *SELECT nome FROM Funcionario a JOIN Pessoa b ON a.CPF = b.CPF WHERE CURRENT\_TIMESTAMP - data\_admissao >= 3*

4.2 - *SELECT nome FROM Pessoa a JOIN Funcionario b ON a.cpf = b.cpf WHERE a.cpf NOT IN (SELECT cpf FROM Funcionario WHERE CURRENT\_TIMESTAMP - data\_admissao < 3)*

*3 - Nome do Fornecedor)*

|  |
| --- |
| *nome* |
| *Dicico* |
| *Duas Irmas* |
| *Telhanorte* |

*4 - nome (Funcionario)*

|  |
| --- |
| *nome* |
| *Valentina Goncalves Oliveira* |
| *Gustavo Oliveira Menezes* |
| *Bernardo Barreto Novaes* |
| *Lucas Moraes Mello* |
| *Rebeca Souza Vasconcelos* |
| *Bruna Garcia Mello* |
| *Antonio Novaes Ribeiro* |
| *Miguel Moraes Marques* |
| *Julia Silva Mello* |
| *Vitoria Martinez Araujo* |

5 – Selecionar o nome e os produtos do fornecedor que fornece a maior quantidade de produtos.

5.1 - *SELECT nome, produtos FROM Fornecedor a JOIN Fornecedor\_Produtos b on a.ID\_Fornecedor = b.ID\_Fornecedor WHERE a.ID\_Fornecedor IN (SELECT ID\_Fornecedor FROM (SELECT a.ID\_Fornecedor, COUNT(produtos) as qtd FROM Fornecedor\_Produtos a JOIN Fornecedor b ON a.ID\_Fornecedor = b.ID\_Fornecedor GROUP BY a.ID\_Fornecedor) as q1 WHERE qtd in (SELECT max(qtd) FROM (SELECT a.ID\_Fornecedor, COUNT(produtos) as qtd FROM Fornecedor\_Produtos a JOIN Fornecedor b ON a.ID\_Fornecedor = b.ID\_Fornecedor GROUP BY a.ID\_Fornecedor)as q1))*

5.2 - *SELECT nome, produtos FROM Fornecedor a JOIN Fornecedor\_Produtos b on a.ID\_Fornecedor = b.ID\_Fornecedor WHERE a.ID\_Fornecedor IN (SELECT ID\_Fornecedor FROM (SELECT a.ID\_Fornecedor, COUNT(produtos) as qtd FROM Fornecedor\_Produtos a JOIN Fornecedor b ON a.ID\_Fornecedor = b.ID\_Fornecedor GROUP BY a.ID\_Fornecedor ORDER BY qtd DESC LIMIT 1) as q1)*

*5 - Nome e Produtos*

|  |  |
| --- | --- |
| *nome* | *produtos* |
| *Patria Unida* | *Aditivos* |
| *Patria Unida* | *Areia* |
| *Patria Unida* | *Argamassa* |
| *Patria Unida* | *Cal* |
| *Patria Unida* | *Cimento* |
| *Patria Unida* | *Concreto* |
| *Patria Unida* | *Madeira* |
| *Patria Unida* | *Materiais Eletricos* |
| *Patria Unida* | *Materiais Hidraulicos* |
| *Patria Unida* | *Pedra Brita* |

Junções de três ou mais relações (3)

6 – Selecionar os status de pagamento de todas as obras localizadas em Belo Horizonte.

6.1 - *SELECT status\_pagamento FROM Pessoa a JOIN Participa b on a.CPF = b.CPF JOIN Obra c ON b.Id\_Obra = c.Id\_Obra JOIN Cliente d ON d.CPF = a.CPF WHERE c.Localizacao = 'Belo Horizonte'*

6.2 *- SELECT status\_pagamento FROM Cliente a JOIN Participa b on a.CPF = b.CPF JOIN Obra c ON b.Id\_Obra = c.Id\_Obra WHERE c.Localizacao = 'Belo Horizonte'*

*6 - Status de pagamento*

|  |
| --- |
| *status\_pagamento* |
| *Bloqueado* |
| *Nao efetuado* |
| *Autorizado* |
| *Bloqueado* |
| *Autorizado* |
| *Autorizado* |
| *Autorizado* |
| *Autorizado* |
| *Nao efetuado* |
| *Bloqueado* |

7 – Selecionar os nomes dos funcionários com salário maior ou igual a 3000 reais que não são responsáveis por nenhum documento. Apresentar os resultados em ordem alfabética.

7.1 - *SELECT nome FROM Funcionario a JOIN Pessoa b on a.cpf = b.cpf LEFT JOIN Documento c ON a.cpf = c.ID\_Responsavel WHERE salario >= 3000 AND c.ID\_Responsavel is null*

7.2 - *SELECT nome FROM Pessoa WHERE cpf IN (Select cpf FROM Funcionario WHERE cpf NOT IN (SELECT ID\_Responsavel FROM Documento) AND salario >= 3000) ORDER BY nome*

*7 - Nomes*

|  |
| --- |
| *nome* |
| *Amanda Menezes Boaventura* |
| *Amanda Menezes Oliveira* |
| *Amanda Novaes Araujo* |
| *Amanda Oliveira Barreto* |
| *Amanda Paiva Carvalho* |
| *Amanda Vieira Costa* |
| *Antonio Alves Ribeiro* |
| *Antonio Campos Paes* |
| *Antonio Maia Fernandes* |
| *Antonio Moraes Morais* |

8 – Selecionar os tipos, localizações e status das obras em que trabalha o funcionário com maior salário.

8.1- *SELECT tipo, localizacao, status\_obra FROM Obra a JOIN Participa b ON a.ID\_Obra = b.ID\_Obra JOIN Funcionario c ON b.cpf = c.cpf WHERE c.salario in (Select Max(salario) from Funcionario)*

8.2 - *SELECT tipo, localizacao, status\_obra FROM Obra WHERE ID\_Obra IN (SELECT ID\_Obra FROM Participa WHERE CPF IN (SELECT CPF from Funcionario WHERE salario IN (Select Max(salario) from Funcionario)))*

*8 - tipo, localização e status da obra*

|  |  |  |
| --- | --- | --- |
| *tipo* | *localizacao* | *status\_obra* |
| *Sobrado* | *Fortaleza* | *Aguardando autorizacao* |
| *Casa* | *Recife* | *Em projeto* |

Funções de agregação (2)

9 – Selecionar o CPF, nome, o salário e a quantidade de obras do funcionário responsável pela maior quantidade de obras.

9.1 - *SELECT a.cpf, nome, salario, qtd FROM ( SELECT ID\_Responsavel as cpf, count(ID\_Documento) as qtd FROM Documento GROUP BY cpf) a JOIN Pessoa b on a.cpf = b.cpf JOIN Funcionario c ON b.cpf = c.cpf WHERE qtd IN (Select max(qtd) FROM (SELECT count(ID\_Documento) as qtd FROM Documento GROUP BY ID\_Responsavel) as sq1)*

9.2 - *SELECT b.cpf, nome, salario, count(Id\_Responsavel) as qtd FROM Documento a JOIN Pessoa b ON a.ID\_Responsavel = b.cpf JOIN Funcionario c ON b.cpf = c.cpf GROUP BY b.cpf ORDER BY qtd DESC LIMIT 1*

*9 - CPF, nome, salario, qtd*

|  |  |  |  |
| --- | --- | --- | --- |
| *CPF* | *nome* | *salario* | *qtd* |
| *80076235408* | *Laura Garcia Moraes* | *3410.00* | *8* |

10 – Selecionar, para cada tipo de obra, a quantidade de obras registradas no sistema com ao menos um contrato de fornecimento que vence até 2022.

10.1 - *SELECT tipo, count(\*) FROM Obra a JOIN (Select ID\_Obra FROM Fornece WHERE expiracao\_contrato < '2021-12-31' GROUP BY ID\_Obra) b ON a.ID\_Obra = b.ID\_Obra GROUP BY tipo*

10.2 - *SELECT tipo, count(\*) FROM Obra a JOIN (SELECT DISTINCT ID\_Obra FROM Fornece WHERE expiracao\_contrato < '2021-12-31') b ON a.ID\_Obra = b.ID\_Obra GROUP BY tipo*

*10 - tipo, count(\*)*

|  |  |
| --- | --- |
| *tipo* | *count(\*)* |
| *Quitinete* | *86* |
| *Apartamento* | *89* |
| *Casa* | *89* |
| *Flat* | *68* |
| *Sobrado* | *69* |
| *Duplex* | *72* |

**Resultados**

Abaixo são apresentados os resultados para os tempos obtidos em cada uma das consultas especificadas na seção anterior. Como dito anteriormente, um banco de dados foi criado e populado a partir de um script feito em Python que incluía diferentes valores para cada atributo, levando-se em conta a unicidade dos dados e também aqueles que não poderiam assumir valores nulos.

A média entre 5 execuções foi feita e os melhores resultados estão em negrito. [INSERIR EXPLICAÇÕES DE MAIS LENTO/MAIS RÁPIDO]

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| --- | Execução 1 (ms) | Execução 2 (ms) | Execução 3 (ms) | Execução 4 (ms) | Execução 5 (ms) | Média  (ms) |
| 1.1 | 0.68288 | 0.6012 | 0.5774 | 0.6498 | 0.7778 | **0.657816** |
| 1.2 | 17.6227 | 1.8895 | 1.8434 | 1.7131 | 2.1922 | 5.05218 |
| 2.1 | 0.7796 | 0.132 | 0.6519 | 0.6158 | 0.7365 | **0.58316** |
| 2.2 | 1.7933 | 1.3656 | 0.877 | 0.904 | 0.949 | 1.17778 |
| 3.1 | 0.7962 | 0.3109 | 0.302 | 0.3214 | 0.3305 | 0.4122 |
| 3.2 | 0.4878 | 0.444 | 0.3963 | 0.3088 | 0.3256 | **0.3925** |
| 4.1 | 2.3778 | 1.7034 | 1.7168 | 1.6554 | 2.0087 | **1.89242** |
| 4.2 | 5.8893 | 6.0674 | 6.0329 | 6.2795 | 6.0817 | 6.07016 |
| 5.1 | 1.0104 | 0.7729 | 0.7576 | 0.6892 | 0.734 | 0.79282 |
| 5.2 | 0.302 | 0.6332 | 0.6213 | 0.65 | 0.345 | **0.5103** |
| 6.1 | 4.5673 | 5.1104 | 4.6522 | 4.867 | 5.5874 | 4.95686 |
| 6.2 | 4.3394 | 4.1033 | 4.0782 | 4.5382 | 4.3258 | **4.27698** |
| 7.1 | 7.7429 | 8.3491 | 8.0845 | 8.1533 | 8.7643 | **8.21882** |
| 7.2 | 17.3371 | 17.4532 | 17.0177 | 17.5433 | 17.3809 | 17.34644 |
| 8.1 | 54.5277 | 54.9069 | 58.5 | 56.9878 | 54.2288 | 55.83024 |
| 8.2 | 55.3473 | 55.0052 | 58.1211 | 54.9364 | 55.188 | **55.7196** |
| 9.1 | 2.6365 | 1.4487 | 1.7956 | 1.7331 | 1.6155 | **1.84588** |
| 9.2 | 6.9726 | 4.1398 | 3.3299 | 3.8192 | 4.2599 | 4.50428 |
| 10.1 | 1.7071 | 1.7625 | 1.7487 | 1.7125 | 1.9514 | **1.77644** |
| 10.2 | 1.9167 | 1.9344 | 2.1114 | 1.8866 | 1.8347 | 1.93676 |

**Conclusões**

O trabalho proposto foi de extrema relevância pois propôs ao grupo o desafio de fazer a modelagem de um sistema do mundo real e formalizá-lo em um banco de dados relacional. Para isso, foi necessário rever conceitos estudados durante toda a disciplina, aplicá-los de maneira correta e lidar com possíveis problemas ao longo do caminho.

A escolha do tema pelo grupo foi feita em conjunto, de maneira que várias ideias foram propostas e apenas uma foi selecionada. O modelo de uma construtora permitiu lidar não só com funcionários e clientes, mas também com os documentos e fornecedores que fazem parte da execução de uma obra de grande porte. Isto permitiu ao grupo modelar uma série de entidades e relacionamentos dos mais diversos a fim de exercitar os conceitos vistos em sala de aula.

A implementação se mostrou eficiente pois considerou-se uma empresa de médio porte, e portanto aproximadamente 50 mil entradas foram feitas no banco de dados, e ainda sim os tempos de consulta são muito pequenos, mesmo com consultas complexas. A média geral de 8ms se mostra irrisória, se comparado a tempos de resposta e de execução de sistemas e algoritmos mais complexos utilizando essa quantidade de dados.