

Orientações para elaboração do 1º trabalho de Tecnologias Big Data

2º semestre 2021 – Prof. Bianca

Tema: **Análise de dados estruturados usando SQL**

Implementação de um banco de dados relacional, na nuvem e utilizar a linguagem SQL para fazer análise dos dados. Exibir os resultados mais relevantes em tabelas e gráficos.

1. Organização do conteúdo do trabalho

Os trabalhos devem ter os seguintes tópicos:

- a) Introdução, que deve ser uma breve descrição sobre o banco de dados.
- b) O projeto conceitual do banco de dados (Diagram Entidade-relacionamento)
- c) Os scripts de 10 consultas SQL para análise de dados. Esses scripts devem incluir comandos de consulta úteis em análise de dados, tais como:
 1. Junções e suas variações (join, left join, right join)
 2. Agregações (Group by, having, max, min, avg, sum)
 3. Subconsultas e Funções (not in, when, date_format, concat, matemáticas)
 4. Ordenações (order by, limit)
 5. Funções analíticas (partition, rank, etc)

Obs: Apresentar as consultas SQL **mais relevantes** do trabalho com destaques para a sintaxe dos comandos utilizados e efeitos do comando no BD. Não repetir comandos, por exemplo, não pode usar as mesmas funções e operadores e só mudar o nome das tabelas. Consultas simples usadas apenas para apresentar uma amostra dos dados não são consideradas na contagem.
- d) Visualizações das consultas mais relevantes (máximo 3).

2. Tema

O tema do banco de dados é de livre escolha dos participantes. Entretanto, o banco de dados deve ter no mínimo 4 tabelas relacionadas.

3. Forma de entrega e apresentação

O trabalho deve ser publicado em um notebook ou em uma página da web e o link para este notebook/página deverá ser enviado via **MOODLE até 28 de setembro**, data em que o grupo fará uma breve apresentação para a turma.

Quanto à apresentação do trabalho

- Todos os membros do grupo devem estar presentes no momento da apresentação.
- Cada apresentação terá duração de 10 minutos

Critérios de Avaliação

Conteúdo	Pontuação
Introdução	1
DER	1
Consultas	5
Visualização dos resultados	1
Apresentação	1
Dados (volume, qualidade, fonte)	1

Exemplos:

1. Um exemplo do Kaggle que apresenta uma boa documentação e fundamentação do SQL
<https://www.kaggle.com/biancapedrosa/data-analysis-using-sql>

2. Um exemplo do Github com bons exemplos de consultas, incluindo funções analíticas do SQL. <https://github.com/guibaptista/analise-imdb>

Referências bibliográficas

SQL: <https://dev.mysql.com/doc/refman/8.0/en/window-functions-usage.html>

Fontes de dados: <https://www.kaggle.com/>, <https://brasil.io/home/>