



1 Descrição do Projeto

Durante a evolução da pandemia da COVID-19 vários institutos uniram as forças para disponibilizar os dados a respeito de sua evolução. Um destes sites é o *Our World in Data*¹ que atualiza os dados diariamente por meio de arquivos CSV, XLSX e JSON. Por simplicidade organizaram os dados todos em um único arquivo. Há uma grande quantidade de dados repetidos indicando que não estão normalizados.

O projeto de vocês será fazer uma engenharia reversa neste arquivo de dados por meio das seguintes tarefas:

1. Normalização dos dados produzindo o esquema relacional equivalente
2. Criação do diagrama do esquema relacional
3. Implementação do esquema relacional no Banco de Dados *MySQL*
4. Carga dos dados no bancos de dados
5. Implementação de consultas

Será disponibilizado um arquivo de carga inicial de dados que você poderá manipular para preencher as tabelas do seu banco de dados.

2 “O que” deve ser entregue

1. **Modelo Lógico** na forma do diagrama representando o esquema Relacional que resultou da normalização. Utilize a notação *crow's foot* para os relacionamentos e a notação **IDEF1X** para as entidades (Retângulo com 3 compartimentos) (sugestão de software: *MySQL Workbench*)
2. **Modelo Físico** na forma do *script* de criação do Banco de Dados em SQL. O *script* deve ter em sua primeira parte a criação de todas as tabelas, respectivos atributos e restrições, entretanto as definições das restrições de integridade referencial (chaves estrangeiras) devem ser definidas após o código de criação de **todas** as tabelas.
3. **Script de Inserção** dos dados para as tabelas. O *script* de inserção será executado logo após o *script* de criação do banco de dados.
4. **Consultas em SQL** requisitadas

2.1 Consultas SQL

As seguintes consultas devem ser implementadas:

1. Quantos países estão na base de dados?
2. Qual o total de casos para o mundo no dia 01/02/2022?
3. Quais foram os 10 países com mais casos confirmados no mês de janeiro/2022 (ordem decrescente)?
4. Liste os 10 países com maior e os 10 com menor expectativa de vida.
5. Liste os continentes contendo o total de casos de cada um em 2021
6. Liste os países da Europa e inclua as informações do total de pessoas totalmente vacinadas em 2021, em valores absolutos e percentuais. Ordene o resultado em ordem decrescente pelo percentual de vacinados.

¹<https://ourworldindata.org/>

7. Liste os países informando o grau de restrições² que foram aplicadas à população (há um atributo com este índice (*stringency*) que engloba várias medidas como fechamento de escolas, proibição de viagens, ...) e o total de novos casos por milhão de habitantes confirmados para o mês de janeiro/2022.
8. Quais países não possuem informação de pacientes na UTI para o mês de janeiro/2022
9. Qual foi o dia com a maior quantidade de novos casos registrados de COVID-19 no Brasil?
10. Qual foi o dia com a maior quantidade de mortes confirmadas por COVID-19 no mundo?

3 “Como” deve ser feita a entrega

- Cada grupo deverá ter um representante que será o responsável pelo envio do projeto pelo Moodle.
- A entrega deve ser feita por meio de um arquivo compactado (.zip). O nome do arquivo deve ser composto pela matrícula dos integrantes do grupo, em ordem crescente, separada por um (_) (ex. aaaa_bbbb_cccc.zip).

O arquivo compactado deve conter:

- Um arquivo em formato pdf com nome “esquema_relacional.pdf” contendo o diagrama do esquema relacional resultante do processo de normalização.
 - Um arquivo contendo o *script* de criação do banco de dados em formato de texto puro com o nome “criacao.sql”
 - Um arquivo contendo o *script* de carga (inserção) do banco de dados em formato de texto puro com o nome “carga.sql”
 - Um arquivo contendo o *script* de consultas ao banco de dados em formato de texto puro com o nome “consultas.sql”
- Formate seu trabalho para que fique legível. Coloque os diagramas na orientação do papel (retrato ou paisagem) que privilegie uma boa visualização. Tenha cuidado com o layout do seu diagrama.

4 Critérios de Avaliação

A pontuação será assim distribuída:

- **20% - Análise e Normalização dos dados.** O resultado do processo deve evitar redundância dos dados, valores nulos desnecessários e armazenamento de atributos que podem ser derivados.
- **20% - Modelo Lógico-diagrama do Esquema Relacional).** Atenção ao layout e correta representação das tabelas e relacionamentos.
- **20% - Modelo Físico.** O código presente no arquivo deve criar toda a estrutura de seu banco de dados.
- **20% - Script de Inserção dos dados.** O código presente em seu arquivo deve fazer a carga em todas as tabelas de seu banco de dados.
- **20% - Consultas SQL.** As consultas devem aparecer na ordem correta (1, 2, ...) e devem produzir o resultado esperado usando seu banco de dados.

²Os países sem este valor informado devem ser descartados do resultado. Utilize a média do valor para o mês.

Haverá uma entrevista com cada grupo a respeito do projeto. A entrevista tem importância fundamental na avaliação.

O grupo receberá uma nota total e o representante será responsável por informar ao professor a distribuição desta nota entre os membros do grupo. Por exemplo, se o grupo possui 3 integrantes e a nota do grupo foi 15, a nota total do grupo será $3 \times 15 = 45$ e o representante informará de que forma os 45 pontos serão distribuídos entre os 3 integrantes.

5 Cronograma para o Projeto

- 23/03/2022 - Entrega do trabalho pelo Moodle.
- 25, 30/03 e 01/04/2022 - Entrevistas com os grupos