



# Bases de Dados

## Aula 1: Introdução aos BDs + Informações sobre a Disciplina

**Prof. Eduardo Corrêa Gonçalves**

Data 08/08/2023

# Tópicos

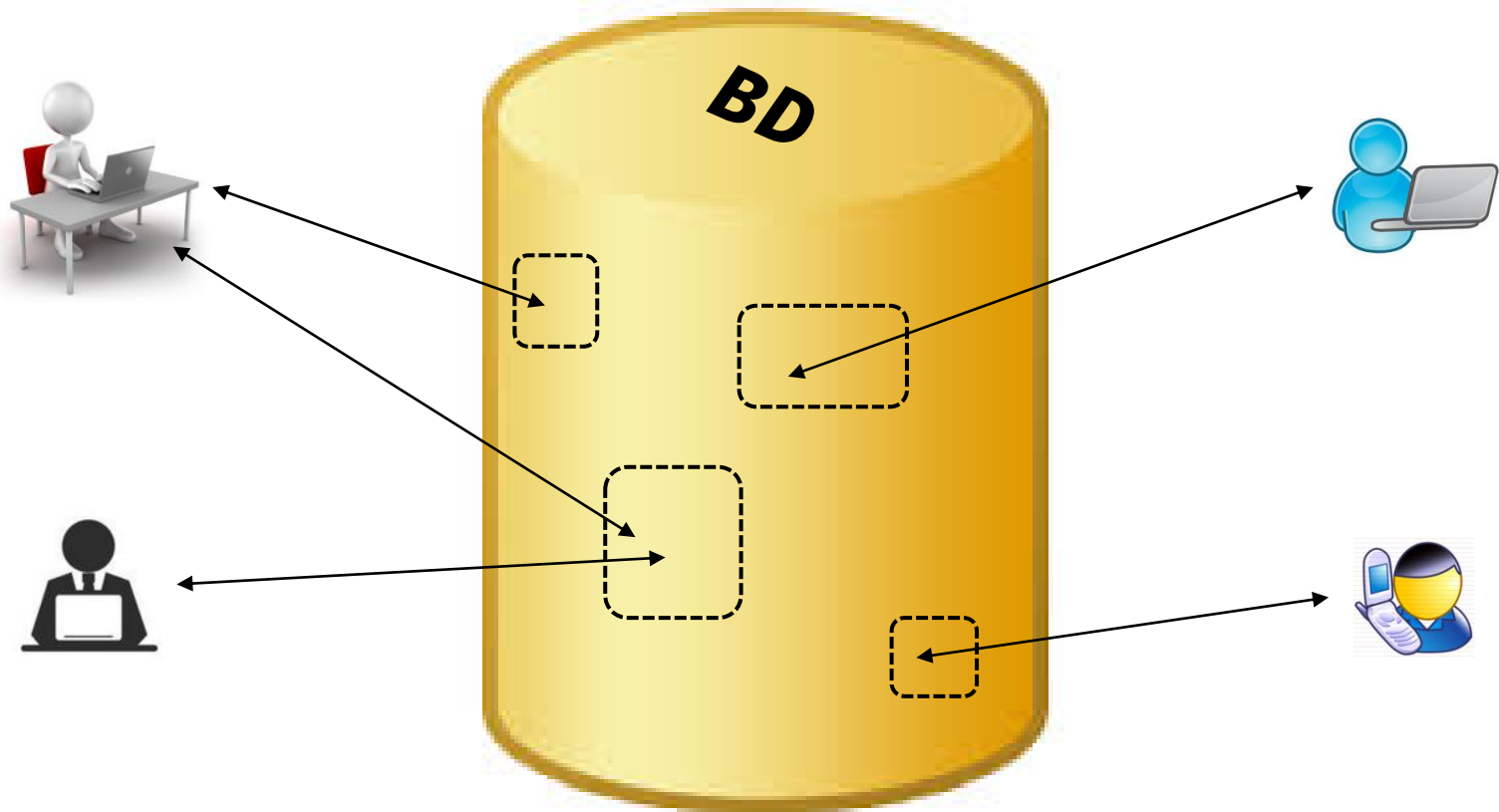
- **Introducao aos Bancos de Dados**
  - O que e Banco de Dados?
  - Sistemas Gerenciadores de Bancos de Dados: Visao Geral
- **Apresentacao do Curso**
  - Objetivos
  - Ementa
  - Bibliografia
  - Criterio de Avaliacao
  - Softwares Utilizados no Curso

# **Introducao aos Bancos de Dados**

# Introducao (1/2)

## O que e Banco de Dados?

- De maneira simplificada podemos definir **banco de dados (BD)** como um **repositorio central** de informacoes que podem ser consultadas e/ou atualizadas por **diversos usuarios** simultaneamente.



## Introducao (2/2)

- Nos dias atuais, interagimos diretamente ou indiretamente com diferentes BDs todos os dias.
  - Conversas ao celular, transacoes bancarias, compras em supermercados sao exemplos de operacoes que produzem registros em BDs.
  - Sempre que voce visita um Web site importante (Google, Globo.com, Amazon.com etc.) existe um BD “**por tras da cena**” para fornecer a informacao que voce requisita. Na verdade, ate mesmo os Web sites mais simples costumam fazer uso de BDs.
- As empresas utilizam BDs para manter todas as informacoes importantes sobre seus **clientes, fornecedores, produtos e servicos**.
- Alem disso, os bancos de dados representam o componente central da maioria das **pesquisas cient ficas** atuais.

# Propriedades-chave dos BDs (1/3)

- Agora vamos detalhar as características de um BD!
- Um BD possui **quatro propriedades**:
  - É projetado, construído e populado com um propósito específico. Os dados do BD representam algum aspecto do mundo real, denominado **minimundo**. Qualquer mudança no minimundo deve ser refletida no banco de dados.
  - Armazena conteúdo de interesse para **muitos usuários** e não apenas um.
  - Seus dados são enxergados e manipulados de maneira unificada, podendo ser acessados a partir de um **repositorio central**.
  - Ao invés de estarem estruturados em planilhas ou arquivos CSV, os dados do BD são estruturados em um software especial e muito poderoso denominado **Sistema Gerenciador de Banco de Dados (SGBD)**.

# Propriedades-chave dos BDs (2/3)

- **Exemplo – IMDb** (Internet Movie Database):
- **Minimundo:** BD que oferece informacoes completas sobre filmes (título, resumo, diretor, elenco, ano, generos, etc.) e artistas. Usuarios do mundo todo podem realizar consultas por diversos filtros, alem de cadastrar producoes, avaliar os filmes, cadastrar comentarios e realizar muitas outras acoes.
- **Usuarios:** Fãs e profissionais de cinema.
- **Repositorio Central:**
  - [www.imdb.com](http://www.imdb.com)
- **SGBD:** Não divulgado.



# Propriedades-chave dos BDs (3/3)

- Os bancos de dados podem ser de **qualquer tamanho** ou **complexidade**.
  - Nos dias atuais, os bancos de dados das empresas rotineiramente armazenam **terabytes** ( $10^{12}$  bytes) ou **petabytes** ( $10^{15}$  bytes) de dados, que são servidos aos seus usuários.
- **Exemplo – IMDb:**
  - O tamanho deste banco de dados não é divulgado, algumas estatísticas interessantes podem ser obtidas\* no endereço <http://www.imdb.com/stats>:
  - 15.463.008 produções cadastradas:
    - 2.909.681 Drama
    - 2.160.251 Comédia
    - 1.099.627 Romance
    - ...
  - 531.906.818 de itens de dados (fotos, dados do elenco, prêmios, etc.).



# SGBD (1/2)

## O que é SGBD?

- Abreviação de **Sistema Gerenciador de Banco de Dados** (*Database Management System - DBMS*)
  - Software especial e muito sofisticado que permite a **definição**, **construção**, **manipulação** e **compartilhamento** de bancos de dados entre vários usuários e aplicações.
  - Ele é também utilizado para garantir a **segurança** dos dados, protegendo-os contra falhas (ex: falhas de hardware) ou tentativas de acesso não autorizado.
  - Alguns exemplos de SGBD's famosos:

**ORACLE**  
DATABASE

  
Microsoft  
SQL Server

 **mongoDB**

 **SQLite**

  
PostgreSQL

 **MySQL**

## SGBD (2/2)

- Os SGBD's representam uma das categorias de software mais desenvolvidas e confiáveis dentre todas em que podemos classificar os sistemas de computação.
  - Existem poucas unanimidades nessa área, mas uma delas sem dúvida é: “use um SGBD para gerenciar seus dados!”.
- Há 2 tipos de SGBD que podem ser utilizados:
  - **SGBD's relacionais:** mais tradicionais e populares, empregam tabelas para representar os dados e adotam SQL como linguagem de consulta.
  - **SGBD's NoSQL:** mais modernos, nem sempre representam os dados em tabelas. Ao longo do curso forneceremos mais detalhes.
- Atualmente, é raro encontrar uma empresa que não utilize um SGBD, seja relacional ou NoSQL.

# Categorias de Usuarios de um SGBD (1/2)

- As 3 categorias tradicionais sao:
- **DBA** (*database administrator*): e o responsavel pela instalacao, autorizacao do acesso, monitoramento do uso, enfim da administracao geral do SGBD.
  - Em uma empresa, um mesmo SGBD podera gerenciar (“tomar conta de”) diversas bases de dados, cada qual referente a um dos diferentes sistemas da empresa.
    - **Ex.:** BD do sistema de RH, BD do sistema de vendas, BD do sistema de fornecedores, etc.
    - O DBA e o responsavel por zelar pelo “bom estado de saude” do SGBD como um todo.
- **Projetista de Banco de Dados:** e um usuario responsavel pelo projeto e criacao do banco de dados de um determinado sistema da empresa. Para tal, ele precisa receber autorizacao do DBA.
- **Usuario Final:** usuarios que consultam ou atualizam o banco de dados, atraves de algum sistema.

# Categorias de Usuarios de um SGBD (2/2)

- ... Porem, uma nova categoria surgiu nos ultimos anos:
  - **Cientista de Dados**: usuario interessado em consultar, coletar, explorar e analisar os dados armazenados em um SGBD, normalmente utilizando uma linguagem de consulta como a SQL.
- Nessa disciplina, os alunos atuarao em dois papeis:
  - Cientista de Dados
  - Projetista de Banco de Dados

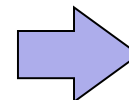
# Um SGBD deve oferecer as seguintes funcionalidades:

- 1. Permitir com que usuarios **definam** e **construam BDs**, especificando suas estruturas de armazenamento (esquemas) com o uso de uma **linguagem de definicao de dados**.



• Criar uma **estrutura** para armazenar filmes que devera conter **campos** para registrar as seguintes informacoes:

- **Titulo do filme** (alfanumerico)
- **Ano** (numero inteiro)
- **Resumo** (alfanumerico)
- **Sigla do pa s** onde filme foi produzido (alfanumerico)
- **Duracao** em minutos (numero inteiro)
- **Avaliacao** do publico (numero real).



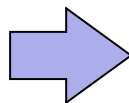
```
CREATE TABLE
Filme (
    Titulo TEXT,
    Ano INT,
    Resumo TEXT,
    Pais TEXT,
    Duracao INT,
    Avaliacao NUM
);
```

## Um SGBD deve oferecer as seguintes funcionalidades:

2. Dar aos usuarios a habilidade de consultar (*querying*) os dados e modifica-los, com o uso de uma **linguagem de manipulacao de dados**.



Inserir o  
filme:  
"Turma da  
Mônica:  
Laços"

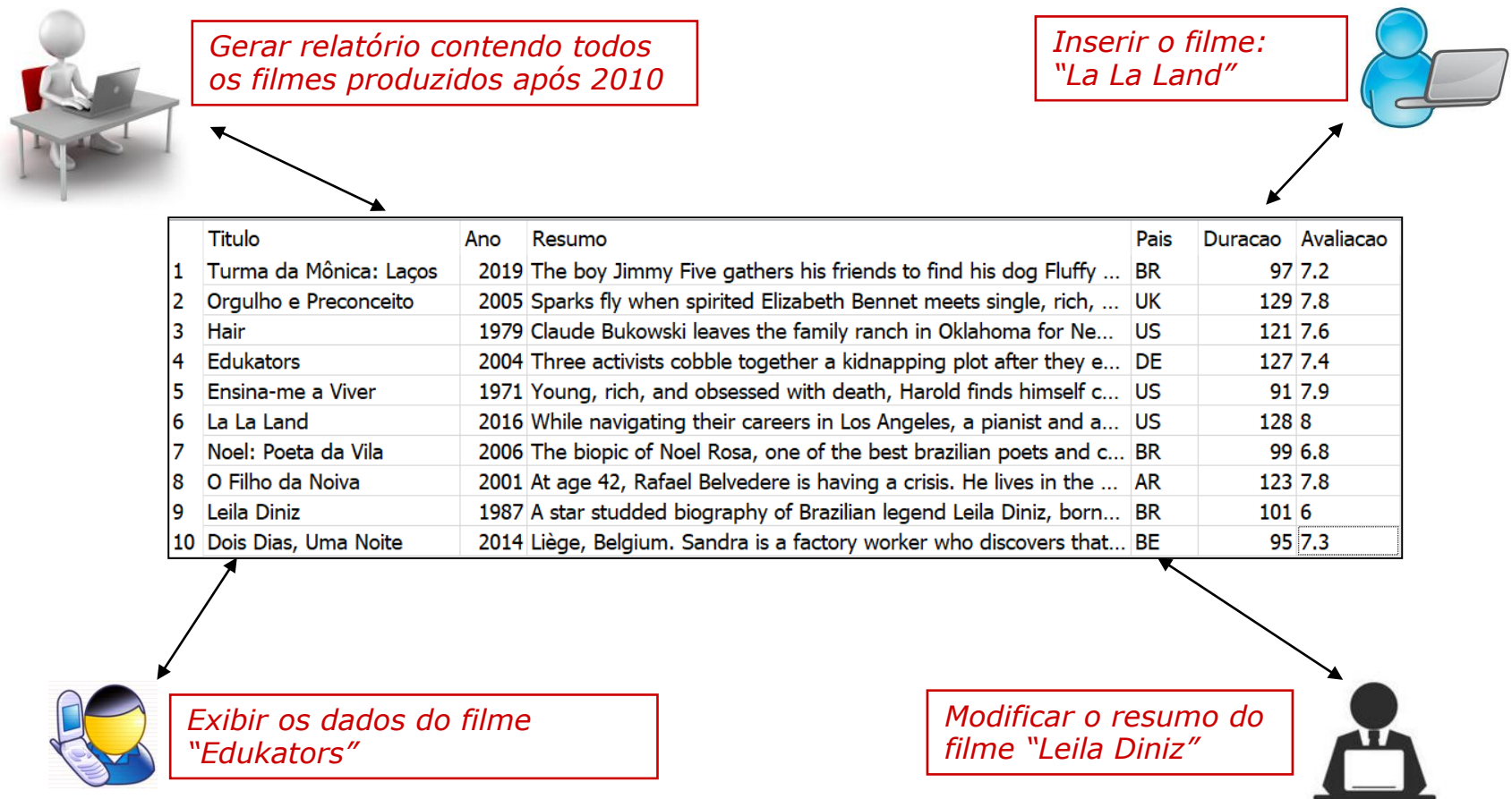


```
INSERT INTO Filme (  
Titulo,  
Ano,  
Resumo,  
Pais,  
Duracao,  
Avaliacao)  
VALUES (  
'Turma da Mônica: Laços',  
2019,  
'The boy Jimmy Five gathers his  
friends to find his dog Fluffy that  
has been kidnapped.',  
'BR',  
97,  
7.2);
```

- O exemplo mostra como inserir informacoes basicas de um novo filme na estrutura *Filme* com o uso da instrucao INSERT da SQL.

# Um SGBD deve oferecer as seguintes funcionalidades:

3. Suportar o armazenamento de **grandes quantidades de dados**, permitindo o acesso **eficiente** as informacoes por multiplos usuarios, de forma concorrente.



## Um SGBD deve oferecer as seguintes funcionalidades:

4. Oferecer um sistema de **backup** que permita a **recuperacao de dados** que sejam corrompidos (por motivo de falha de hardware ou software) ou perdidos de forma acidental.

- **Ex:** um usuario pode acidentalmente excluir os dados de um filme, o que exigira com que o DBA recupere esses dados.

5. Oferecer mecanismos para **controle de acesso** nao autorizado.

- Quando multiplos usuarios compartilham o acesso a um banco de dados, e comum que muitos deles nao estejam autorizados a acessar todas as informacoes contidas no banco.
  - **Ex:** (i) uma parte da base de dados do IMDb e restrita para usuarios assinantes; (ii) alguns usuarios do IMDb podem alterar dados de filmes, enquanto outros podem apenas consulta-los.
- Atraves do sistema de seguranca e controle de acesso do SGBD, o DBA pode especificar as restricoes de acesso associadas a cada perfil/conta de usuario.



## Um SGBD deve oferecer as seguintes funcionalidades:

6. Garantir que as **restricoes de integridade** associadas aos dados (definidas pelo Projetista do Banco) sejam sempre respeitadas, para que o banco mantenha-se sempre em um estado consistente.

- **Exemplos:**

- Não permitir com que um mesmo filme (título + ano) seja cadastrado duas vezes (restricao de chave).
- Não permitir que um texto seja gravado no campo referente ao ano do filme (restricao de tipo).
- Não permitir que a nota relativa a avaliacao do filme seja inferior a 0 ou superior a 10 (violacao de valor de dominio).
- Garantir que o elenco de um filme contenha apenas artistas que tenham sido previamente cadastrados na estrutura "Artista" (integridade referencial).

# A Era pre-SGBD

- Antes dos SGBD's se tornarem populares, os programas usados pelas empresas costumavam manter seus bancos de dados em arquivos proprietários (arquivos que "pertenciam" ao próprio programa). Isto gerava uma série de inconvenientes:
- Cada arquivo possuía uma estrutura distinta (escolhida pelo programador). Desta forma, tem-se por exemplo que os dados dos sistemas de vendas e do sistema de fornecedores poderiam ter formatos inteiramente distintos e incompatíveis.
- Os dados da empresa acabavam espalhados em diversos arquivos, ao invés de armazenados em um repositório central.
- Alterar o formato de um banco de dados (por exemplo, incluir uma nova variável) era um processo complicado, pois era preciso mexer no código do programa que gerenciava o banco.
- Era mais difícil impedir os acessos não autorizados, já que os dados ficavam espalhados em diferentes arquivos e não centralizados no SGBD, que só pode ser acessado por quem possui conta e senha.
- Os programadores precisavam implementar em seus programas as complicadas técnicas para disponibilizar arquivos para atualizações concorrentes. Também precisavam programar as técnicas de indexação responsáveis por garantir a recuperação eficiente de informações.

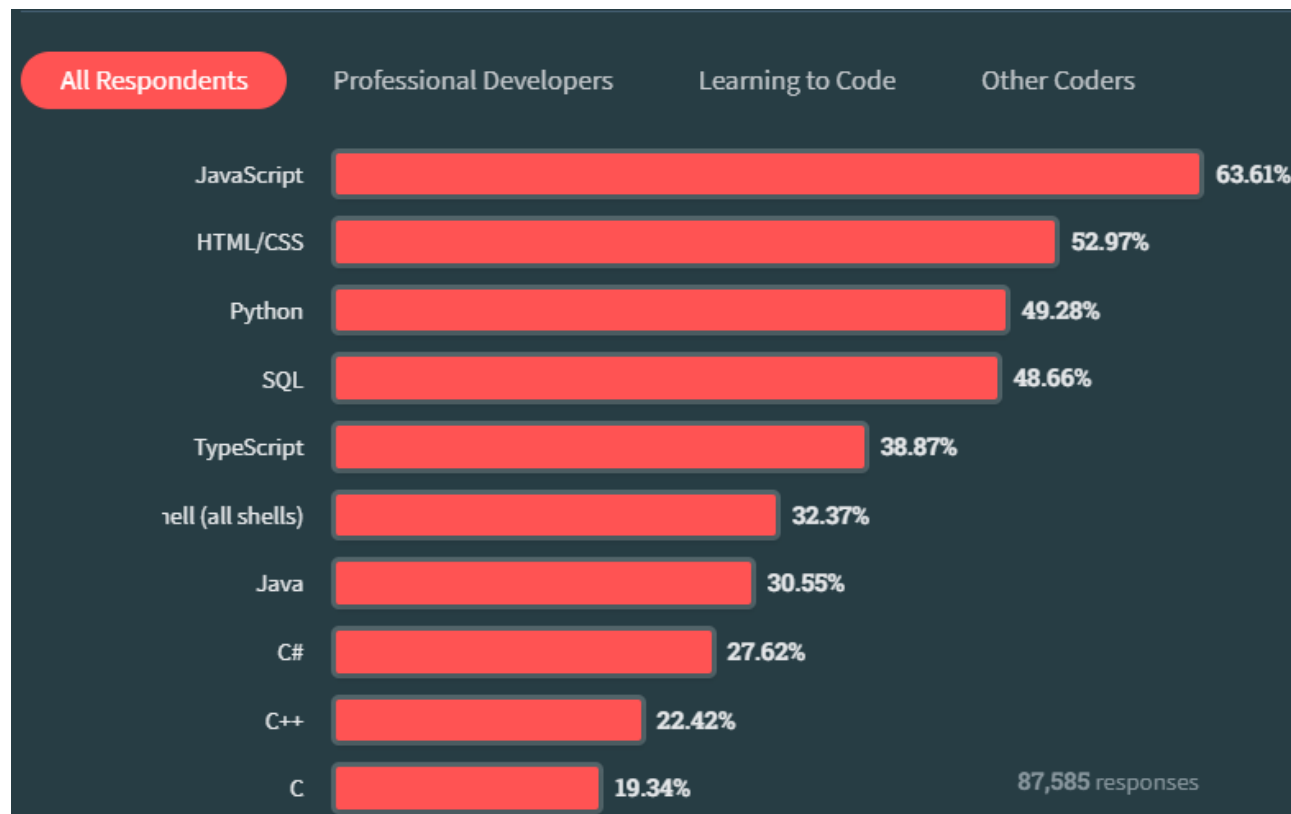
# **Informacoes Sobre a Disciplina**

# Objetivos da Disciplina

- Os objetivos da disciplina **Bases de Dados** são:
  1. Habilitar o aluno a **projetar, construir, atualizar e consultar** bancos de dados.
    - **Enfase** especial é dada aos **SGBDs relacionais**, tecnologia predominante no mercado.
    - Mas também abordaremos os SGBDs NoSQL.
  2. Tornar o aluno proficiente em **SQL** (*Structured Query Language*)
    - Trata-se da **linguagem padrão** para manipular SGBDs relacionais.
    - Como vocês são alunos de Estatística, o curso prioriza a utilização da SQL para **explorar** e **combinar** bases de dados, além de outras tarefas relacionadas a **Ciência de Dados**.

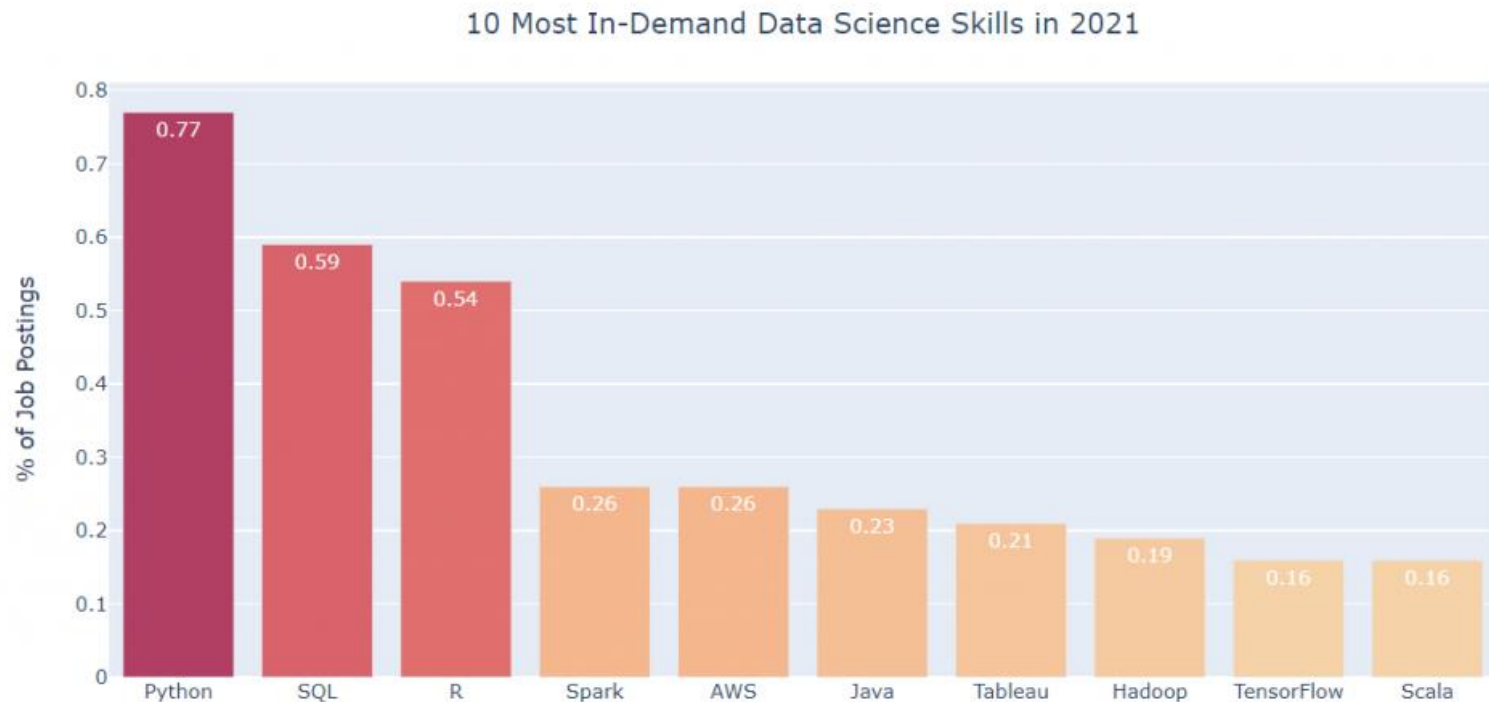
# Importancia da SQL (1/2)

- Stack Overflow Developer Survey 2023
- Relacao das 10 linguagens mais utilizadas em aplicativos e sistemas



## Importancia da SQL (2/2)

- Abaixo o resultado da análise dos requisitos exigidos em 15.000 ofertas de emprego no exterior em 2021 \*















- Fonte <https://www.kdnuggets.com/2021/04/most-demand-skills-data-scientists.html>

# SGBDs Relacionais

• Abaixo, o ranking com os 10 SGBD's mais populares de acordo com o site **db-engines**\*

- 7 deles são SGBD's relacionais.

Rank			DBMS	Database Model	Score		
Aug 2023	Jul 2023	Aug 2022			Aug 2023	Jul 2023	Aug 2022
1.	1.	1.	Oracle +	Relational, Multi-model 	1242.10	-13.91	-18.70
2.	2.	2.	MySQL +	Relational, Multi-model 	1130.45	-19.89	-72.40
3.	3.	3.	Microsoft SQL Server +	Relational, Multi-model 	920.81	-0.78	-24.14
4.	4.	4.	PostgreSQL +	Relational, Multi-model 	620.38	+2.55	+2.38
5.	5.	5.	MongoDB +	Document, Multi-model 	434.49	-1.00	-43.17
6.	6.	6.	Redis +	Key-value, Multi-model 	162.97	-0.80	-13.43
7.	 8.	 8.	Elasticsearch	Search engine, Multi-model 	139.92	+0.33	-15.16
8.	 7.	 7.	IBM Db2	Relational, Multi-model 	139.24	-0.58	-17.99
9.	9.	9.	Microsoft Access	Relational	130.34	-0.38	-16.16
10.	10.	10.	SQLite +	Relational	129.92	-0.27	-8.95

- Fonte: <https://db-engines.com/en/ranking>

# Ementa

## TEORIA

- **Modelo Relacional**
  - Visao Geral e Terminologia
  - Restricoes de Integridade
  - Algebra Relacional
  - Projeto de BD
  - Outros Objetos de BD ( ndices, visoes, ...)
  - Controle de Transacoes
- **NoSQL:** Document Databases, Graph Databases
- **BDs Anal ticos:** DW, Data Lakes, ...

## PRATICA

- **SQL / NoSQL**
  - Criacao de Esquemas.
  - Modificacao de Dados.
  - Restricoes de Integridade.
  - Consultas Basicas
  - Juncao de Tabelas
  - Indices, Visoes
  - Controle de Transacoes
  - SQL para Ciencia de Dados
  - Consultas em SGBDs NoSQL



# Material de Estudo

- Notas de Aula\* e Listas de Exercícios

- Serão sempre disponibilizadas na minha pasta no site <https://aulasence.ibge.gov.br/>

\***OBS:** as notas de aula foram criadas com base no conteúdo dos livros abaixo:

***Fundamentals of Database Systems***, 7<sup>th</sup> Edition

Ramez Elmasri & Shamkant B. Navathe

Pearson, 2016.

***Database Systems: The Complete Book***, 2<sup>nd</sup> Edition

Hector Garcia-Molina, Jeffrey D. Ullman & Jennifer Widom

Pearson, 2009.

***NoSQL Distilled***

Sadalage, P. J. & Fowler, M.

Addison-Wesley, 2012.

# Criterio de Avaliacao

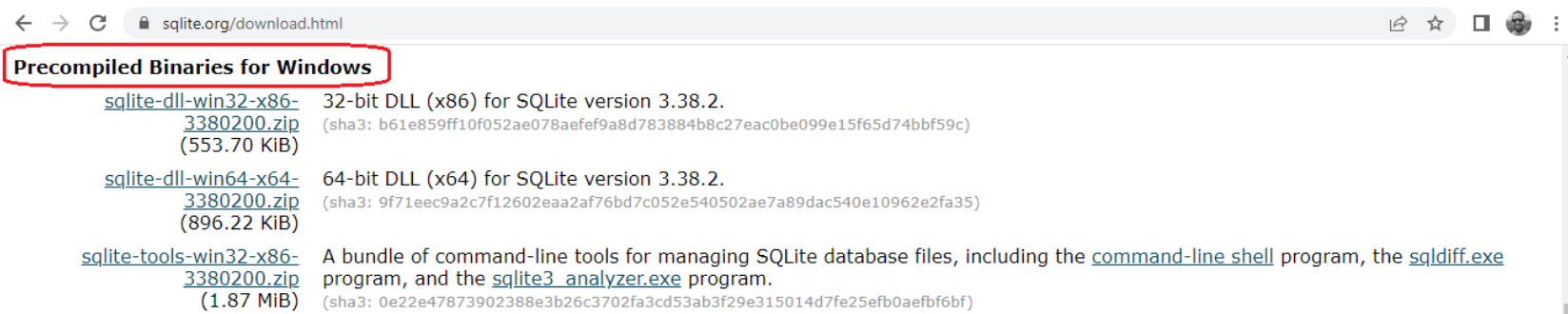
- O criterio abaixo sera adotado tanto para a VAE1 como para a VAE2
  - Nota obtida na **prova**
  - Simulados e listas de exerc cios poderao acrescentar ate 0,5 na nota final.
- Para tirar duvidas com o professor, envie mensagem preferencialmente pelo **Microsoft Teams**
  - Basta ir no chat e procurar meu nome (Eduardo Correa Goncalves)
  - Se nao conseguir pelo teams, tente por e-mail: [eduardo.ence@gmail.com](mailto:eduardo.ence@gmail.com)

# Softwares Utilizados no Curso

- Os seguintes softwares serao utilizados no curso:
  - **SQLite**: utilizaremos este software como um “emulador” de SGBD.
  - **SQLite Studio**: permite manipular o SQLite atraves de uma interface grafica.
- Adicionalmente, em **aulas opcionais** cobriremos os softwares abaixo:
  - **sqldf** (pacote R): permite com que data frames possam ser consultados atraves da linguagem SQL no ambiente R.
  - **RSQLite** (pacote R): permite a conexao do R com o SQLite.
  - **Python**: veremos como integrar o Python com os SGBDs SQLite e MySQL.

# Download do SQLite

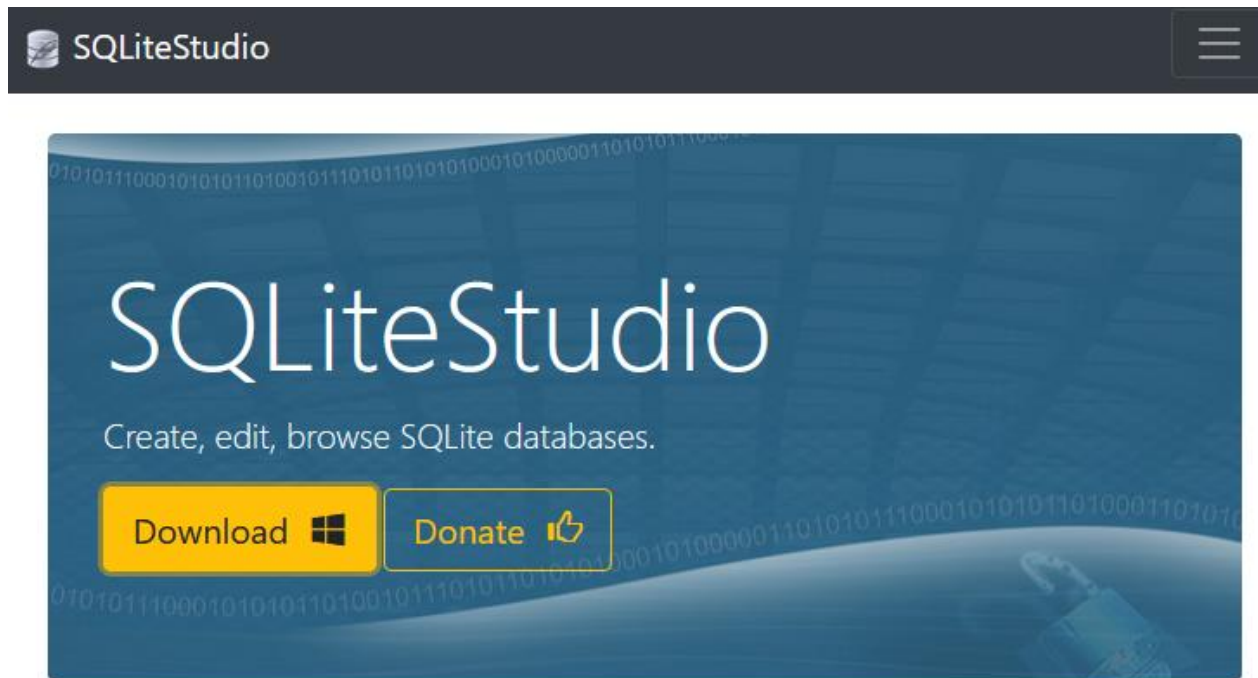
- Acesse o endereço: <http://www.sqlite.org/download.html>



- Baixe os arquivos indicados em “**Precompiled Binaries for Windows**”. A versão mais recente na data em que esta aula foi elaborada é a 3.42.0.
  - “**sqlite-dll-win32**” ou “**sqlite-dll-win64**” (*dependendo do seu tipo de processador, 32 ou 64 bits – versão para Windows*).
  - “**sqlite-tools-win32**”
    - Preferencialmente, **descompacte** ambos para a pasta “**C:\bd**” do seu computador. Ou crie uma pasta “**bd**” em seu pen drive (o SQLite funciona bem mesmo quando executado a partir do pen drive).

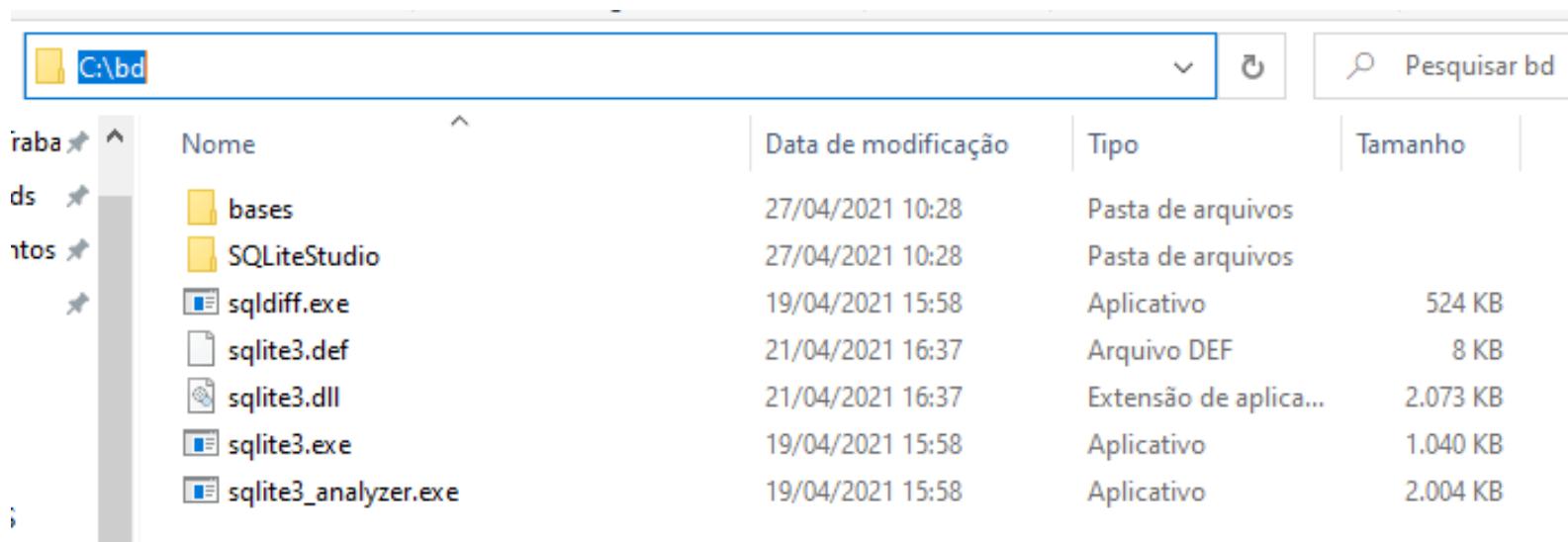
# Download do SQLite Studio

- Acesse o endereço: <https://sqlitestudio.pl/>
- Basta clicar no botao Download.
- Preferencialmente, descompacte a pasta SQLite Studio para "C:\bd" (ou para o seu pen drive).



# Estrutura de Pastas

- Como sugestao, utilize a seguinte estrutura de pastas para armazenar os softwares do curso e os bancos de dados que serao criados em cada aula:



Nome	Data de modificação	Tipo	Tamanho
bases	27/04/2021 10:28	Pasta de arquivos	
SQLiteStudio	27/04/2021 10:28	Pasta de arquivos	
sqldiff.exe	19/04/2021 15:58	Aplicativo	524 KB
sqlite3.def	21/04/2021 16:37	Arquivo DEF	8 KB
sqlite3.dll	21/04/2021 16:37	Extensão de aplica...	2.073 KB
sqlite3.exe	19/04/2021 15:58	Aplicativo	1.040 KB
sqlite3_analyzer.exe	19/04/2021 15:58	Aplicativo	2.004 KB

- Nesta estrutura, a pasta **bases** armazenara os bancos de dados criados em nossas aulas (bancos que serao criados com o uso do SQLite).