

Banco de dados

Felipe Marx Benghi

fbenghi@gmail.com

Objetivos

- [] Apresentação do plano de aulas da disciplina (Ementa)
- [] Combinados
- [] Ferramentas para disciplina
- [] Conscientização sobre ética, segurança, autenticação, integridade e responsabilidade dos dados (LGPD).

Sugestões de sites e leituras

- Bibliografia

[Português]ELMASRI, R.; NAVATHE, S. B. Sistemas de Banco de Dados. 6ª ed. São Paulo: Pearson, 2011.

[Inglês][Livro - Database Design \(full\)](#).

[Inglês][Geeks for Geeks](#)

- Dados abertos

[Google Data Search](#)

[Kaggle](#)

Ferramentas

MySQL Workbench

<https://dev.mysql.com/downloads/workbench/>

Visual Code (opcional)

<https://code.visualstudio.com/>

Extensions

[Marp](#) Cria uma apresentação com Markdown

Lei Geral de Proteção dos Dados (LGPD)

- Objetivo: proteger dados pessoais
- Obrigatória desde 01/08/2021

Dados pessoais

“ Dados pessoais são informações sobre uma pessoa física identificada ou identificável ”

Exemplos:

- Dados de funcionários
- Dados de clientes
- Dados do fornecedor
- Dados agregados (para análise)
- Dados médicos

Identificador Direto ou Indireto

Um dado pessoal pode ser:

- Identificador Direto
 - Aplicável a uma pessoa
 - Ex. CPF, **Pelé**
 - Identificador Indireto
 - Aplicável possivelmente a mais de uma pessoa
 - Data de nascimento, Endereço de IP, nacionalidade
- Importante: vários identificadores indiretos podem permitir a identificação do indivíduo**

Dados Confidenciais

Dados (identificadores indiretos) que apresentam maior risco aos **direitos** dos indivíduos.

- Exemplo: vida sexual ou orientação sexual, dados genéticos, opiniões políticas, dados biométricos, crenças religiosas, origem racial ou étnica

Normalmente é necessário ter consentimento para armazenamento de tais dados

Agentes de processamento

Precisam estar em conformidade com a LGPD (independentemente da localização), agentes que:

1. Processam dados pessoais no Brasil
2. Coletam dados pessoais no Brasil ou de indivíduos no Brasil
3. Oferecem bens ou serviços a indivíduos no Brasil

Agentes de processamento devem: manter registros de quais dados armazenam, do que pode ser removido e do que pode ser destruído.

Direitos dos titulares (Gratuitos, adequados e imediato)

Titulares têm o direito a:

- Avisos de privacidade - ser informados sobre quais dados foram coletados e com quem são compartilhados.
- Portabilidade - mediante solicitação, os dados podem ser transferidos e devem estar em formato legível.
- Exclusão - dados podem ser mantidos por motivos legais ou regulatórios.
- Correção dados incompletos, imprecisos ou desatualizados.

E eu com isso...

Um administrador do Banco de Dados precisa garantir que o sistema está de acordo com a LGPD!

Administradores de Banco de dados têm acesso a mais dados do que usuários normais!



WARNING

**VIEWER DISCRETION
IS ADVISED**

Conceitos Básicos

Dados

Fatos que podem ser gravados e tem um significado implícito

Exemplos: Nomes, telefones, email, dia do nascimento

Banco de dados (Database)

- Conjunto de Dados que se relacionam
- Possui alguma fonte em comum de onde os dados são derivados, algum grau de interação com eventos do mundo real, e uma audiência interessada nos seu conteúdo

Persistência

- A persistência de dados é a garantia de que um dado foi salvo e que poderá ser recuperado quando necessário no futuro
- São registros permanentes e que não são perdidos quando há o encerramento da sessão

Exemplo de sistemas para persistência de dados: Livros, Arquivos (físicos e digitais), Banco de dados

Persistência de dados em Arquivos

Forma clássica de armazenamento de dados. Funciona, mas:

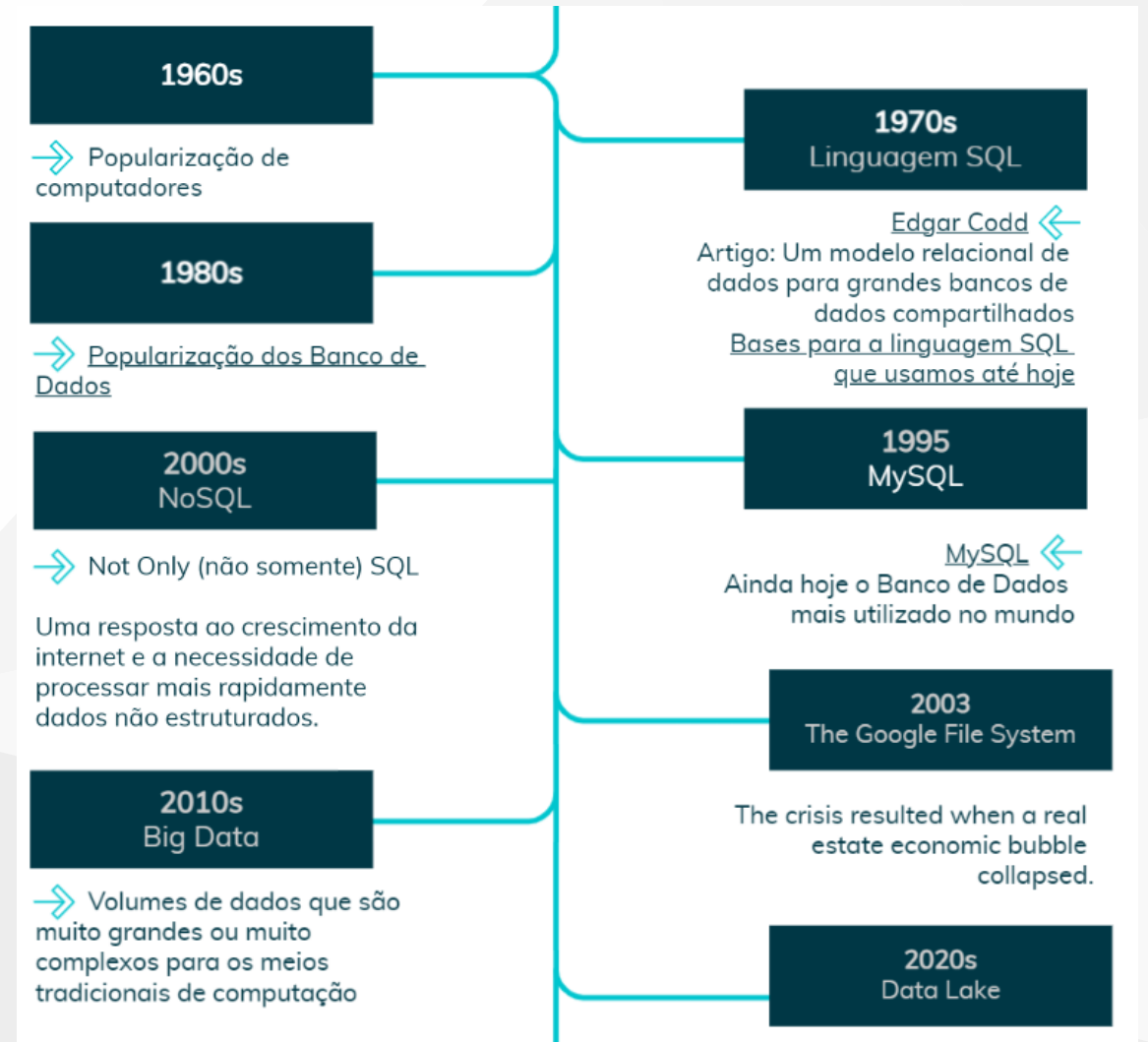
- Permite redundância / mais de uma instância de um atributo
 - Mesma informação armazenada em lugares diferentes
 - Mesma informação diferente
- Não garante integridade / permite dados (in)corretos e (in)consistentes

- Sem isolamento / uma modificação é imediatamente visível. Exemplo: Alguém querendo ler o dado, enquanto a escrita ainda não foi terminada
- Com problemas de segurança: Sem proteção para quem pode ler / escrever ou controle de acesso
- Sem acesso concorrente: Só uma pessoa pode escrever num arquivo ao mesmo tempo

Solução:



Evolução dos bancos de dados



MySQL

- Sistema de gerenciamento de banco de dados com código aberto
- Sistema de banco de dados mais popular do mundo - em pesquisa do Stack Overflow de 2020, ainda 55,6% dos participantes o utilizavam.
- MySQL ainda é usado por alguns dos websites mais famosos do mundo como Airbnb, Uber, LinkedIn, Facebook, Twitter, and YouTube (<https://www.oracle.com/database/what-is-database/#link4>)

NoSQL

- Not only SQL (não somente SQL)
- Este formato de banco de dados veio como uma resposta ao crescimento da internet e da necessidade de processar mais rapidamente dados não estruturados.

Google

- Artigos do Google que mostraram para o mundo como processar grandes quantidades de dados usando computação distribuída
 - 2003 - The Google File System
 - 2004 - MapReduce - Simplified Data Processing on Large Clusters
 - 2006 - Big table - A Distributed Storage System for Structured Data

Big Data

- Descreve um volume de dados que são ou muito grandes ou muito complexos para serem processados pelos métodos tradicionais de computação.
- 3Vs para definir Big Data
 - Dados que contém uma grande VARIEDADE
 - Dados que chegam em VOLUME crescente
 - Dados que chegam com VELOCIDADE crescente
- Exemplo: Posts, tweets, images, and video clips

Data Lake

- É um repositório centralizado para armazenar, processar dados estruturados, semiestruturados ou não estruturados.
- Exemplo de empresa: Snow Flake - empresa com a maior estreia na bolsa de valores americana em 2020

