



WBA0178\_v2.0

# Banco de Dados Relacional e Big Data





# Banco de Dados e Big Data: União Necessária

Bloco 1

Cassio Rodolfo Aveiro da Silva



## Dados: o novo petróleo

- ▶ Os dados são a nova **fonte de riqueza e vantagem competitiva** para a maioria das empresas.
- ▶ Por meio destes dados é que são processadas todas as **informações** necessárias e de **alto poder empresarial**.
- ▶ Definições importantes:
  - ✓ **Dados:** características brutas, quantitativas ou qualitativas sobre algo.
  - ✓ **Informação:** dados tratados, organizados e com significância.

## Dados: o novo petróleo

- ▶ Os dados, tendo essa enorme importância, necessitam de um **lugar seguro para serem armazenados**.
- ▶ Necessitam, também, de um **ambiente interativo**, que permitem uma relação dos dados entre si.
- ▶ Para essa relação, os dados devem ser **organizados e identificáveis** para o acesso.
- ▶ Os **bancos comuns** de armazenamento de dados são **insuficientes**.
- ▶ É preciso o uso de **Bancos de Dados Relacionais**.

# Banco de Dados Relacional

- ▶ Baseados no **modelo relacional**.
- ▶ Armazenamento de dados por meio de **tabelas**/matrizes.
- ▶ Linhas com endereçamento próprio (**ID ou chaves**).
- ▶ Colunas funcionais ou de **atributos associados** a cada dado.
- ▶ Dados com os mesmos atributos são relacionáveis com tabelas **compartilhando** da mesma coluna.
- ▶ Processamento e acesso rápido de dados e informações, respectivamente.



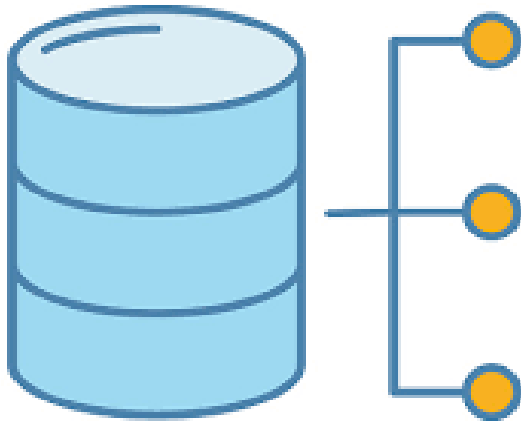
# Modelo Relacional

- ▶ No começo, o **acesso** era **restrito**.
- ▶ Dificuldade na **autonomia**.
- ▶ Necessidade de **compartilhamento de dados**.
- ▶ Solução:
  - ✓ **Modelo Relacional:** mecanismo de armazenamento mais rápido e dinâmico, o qual permite uma interação e associatividade entre dados.

# Modelo Relacional

- Linguagem de Consulta Estruturada (SQL - *Structured Query Language*): álgebra relacional.

Figura 1 - Estruturação de dados



Fonte: bsd55/iStock.com.

# Aplicações do Banco de Dados Relacional

- ▶ Empresas de **pequeno a grande porte**:
  - ✓ Armazenamento seguro.
  - ✓ Acesso rápido.
  - ✓ Confiabilidade.
- ▶ Primeiros modelos datam de **1970**.
- ▶ Exemplos de clientes: cartórios, delegacias, hospitais e órgãos públicos (como delegacias e prefeituras).



# Acessos ao Banco de Dados Relacional

- ▶ A **praticidade** e **agilidade** são os **requisitos** principais para o acesso ao banco de dados.
- ▶ O Banco de Dados Relacional dispõe de mecanismo que facilitam o acesso:
  - ✓ **Códigos de acesso:** utilizados para identificar um dado ou conjunto de dados específicos, amplamente requisitados repetitivamente.
  - ✓ **Acesso simultâneo:** vários usuários podem acessar os mesmos dados de diferentes banco de dados relacionados entre si.

# Acessos ao Banco de Dados Relacional

- ▶ O acesso simultâneo pode gerar **congestionamentos** na rede e **conflitos**, sobretudo se muitos usuários estiverem editando o mesmo dado.
- ▶ Surgem as técnicas de **bloqueio de acesso**:
  - ✓ Bloqueio parcial (tabela).
  - ✓ Bloqueio total (registro).
  - ✓ Gestor de acesso.
  - ✓ *Edudrive* - governo brasileiro.

## Banco de Dados Relacional: tendências

- ▶ As mudanças contínuas da **Tecnologia de Informação** e de seus respectivos **Sistemas de Informação** geram a necessidade de empresas possuírem bancos de dados:
  - ✓ Personalizáveis.
  - ✓ Com prazo de validade.
  - ✓ Compatível com as tecnologias contemporâneas.
  - ✓ Compatibilidade com os sistemas.

## Banco de Dados Relacional: tendências

- ▶ Atualmente, há uma **enorme complexidade e volume de dados** a serem processados.
- ▶ A **gestão manual** desses dados é praticamente **impossível**.
- ▶ Necessidade de um Banco de Dados **Autônomo**:
  - ✓ Evolução dos Banco de Dados Relacionais.
  - ✓ Independência e automação.
  - ✓ Inteligência Artificial (IA) - auto evolução.



# Banco de Dados e Big Data: União Necessária

Bloco 2

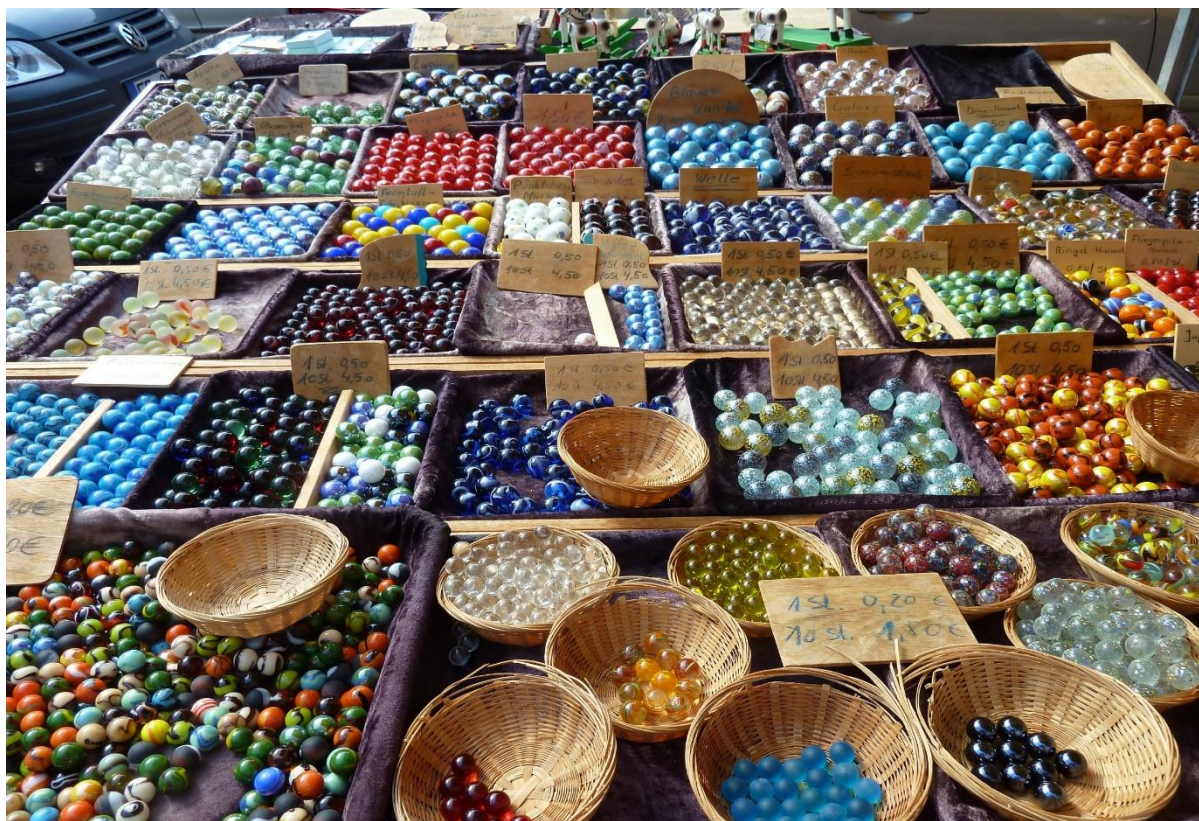
Cassio Rodolfo Aveiro da Silva





# Compreendendo o que são banco de dados

Figura 2 - Representação de dados e banco de dados



Fonte: <https://pxhere.com/pt/photo/939420>. Acesso em: 6 nov. 2020.

# Compreendendo o que são banco de dados

- ▶ **Conjunto de dados** registrados é comumente chamado de **arquivo**.
- ▶ Os banco de dados devem ser estruturados e organizados para gerarem um **acesso ágil aos arquivos**:
  - ✓ Tempo de aquisição.
  - ✓ Confiabilidade.



# Compreendendo o que são banco de dados

- ▶ Pontos críticos em um Banco de Dados:
  - ✓ **Redundância:** ter os mesmos dados disponíveis em diferentes áreas de uma mesma organização.
  - ✓ **Inconsistência:** falta de constantes atualizações nos dados.
  - ✓ **Integração:** compartilhamento de dados, comunicação efetiva e relação com os sistemas de informação.

# Compreendendo o que são banco de dados

- ▶ Surgem conceitos sobre o **modelo de banco de dados**.

Modelo de um Banco de dados é **descrição formal de sua estruturação** apresentada.

Há três modelos principais:

- I. **Conceitual.**
- II. **Lógico.**
- III. **Organizacional.**

# Compreendendo o que são banco de dados

- ▶ Como existem inúmeros dados e bancos de dados disponíveis às empresas e aos desenvolvedores, surgem os **Sistemas de Gerência de Banco de Dados (SGBD)**:
  - ✓ Incorporação de funções.
  - ✓ Recuperação de dados.
  - ✓ Alteração de dados.
  - ✓ Manutenção de grandes repositórios de dados.



# Banco de Dados e Big Data: União Necessária

Bloco 3

Cassio Rodolfo Aveiro da Silva



# As necessidades por trás do tratamento de dados

- ▶ Sobretudo em empresas de grande porte, o **volume de dados** a serem processados é **gigantesco**.
- ▶ Logo, são necessários **sistemas de alto poder de análise de dados e modelagem de dados**.
- ▶ Há três principais concepções para esse tratamento de dados:
  - ✓ Tecnologias de grandes dados.
  - ✓ Ferramentas de inteligência de negócios.
  - ✓ Ferramentas para análises.

# As necessidades por trás do tratamento de dados

- ▶ Essas concepções são conhecidas como **pilares do tratamento de dados**.
- ▶ Todo projeto empresarial necessita de tratamento adequado de dados que englobe:
  - ✓ **Seleção da ferramenta adequada.**
  - ✓ **Análise dos dados propriamente dita.**
  - ✓ **Transformação em informação significativa.**
- ▶ Um projeto bem estruturado é sinônimo de um projeto eficiente e de sucesso.



# As necessidades por trás do tratamento de dados

Figura 3 - Centrais de informação



Fonte: <https://pxhere.com/pt/photo/1058721>. Acesso em: 6 nov. 2020.



# Desenvolvimento de projetos baseados na análise de dados

- ▶ Um projeto se inicia com **perguntas de pesquisa**, as quais darão o direcionamento de todo o seu desenvolvimento.
- ▶ Com os objetivos definidos e claros, alguns fatores exigem atenção:
  - ✓ **Escalabilidade.**
  - ✓ **Alta disponibilidade.**
  - ✓ **Flexibilidade.**

# Desenvolvimento de projetos baseados na análise de dados

## ► Escalabilidade:

- ✓ Quantitativo de usuários simultâneos.
- ✓ Medida preventivas contra picos de acesso.

## ► Alta disponibilidade.

- ✓ Garantia de acesso.
- ✓ Em tempos real.
- ✓ Ex.: redes sociais.

# Desenvolvimento de projetos baseados na análise de dados

## ► Flexibilidade:

- ✓ Modo de armazenamento.
- ✓ Forma de acesso por diferentes meios.

## ► Esses fatores influenciam na **estrutura do banco de dados** a ser utilizado pelos projetos:

- ✓ Custos.
- ✓ Gerenciamento.
- ✓ Segurança de informações.

# Desenvolvimento de projetos baseados na análise de dados

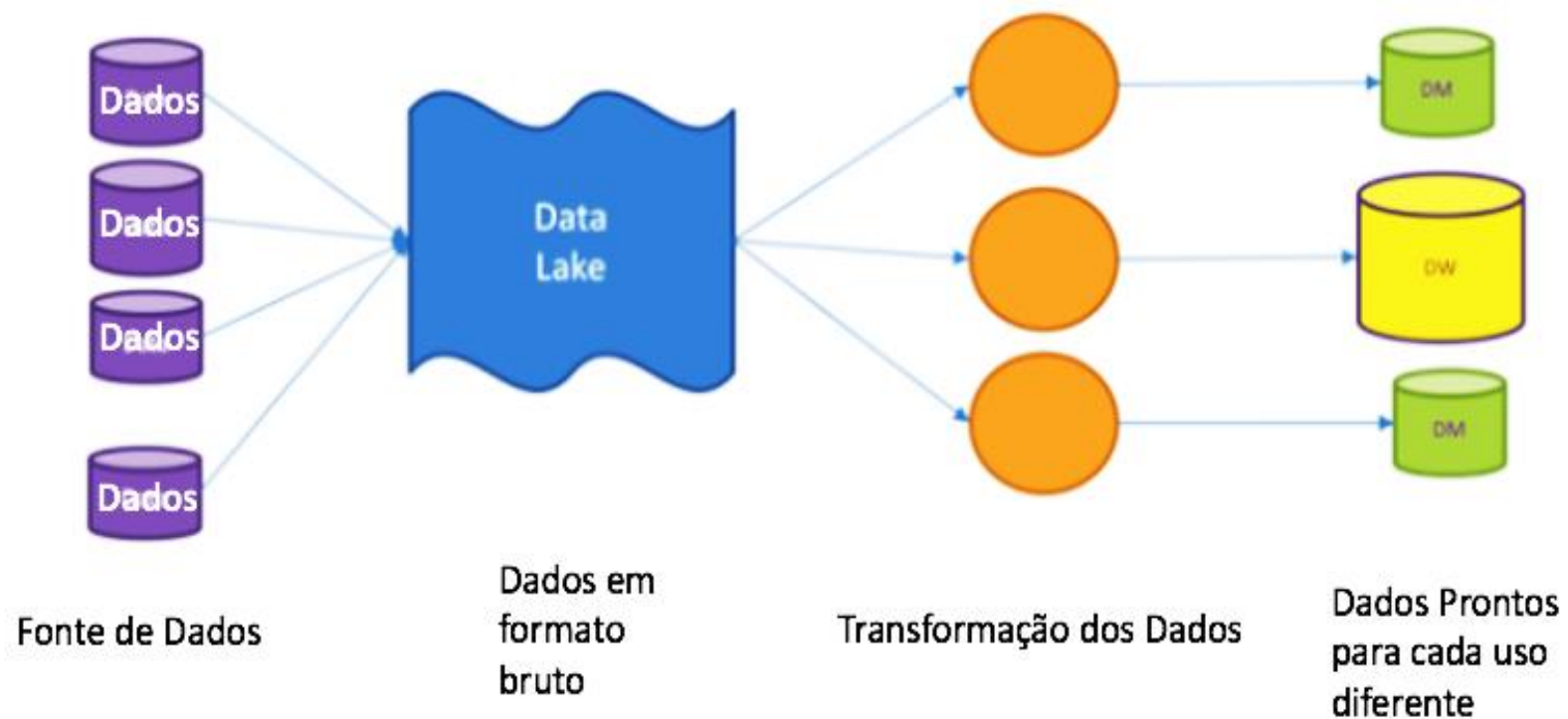
- ▶ É importante que você compreenda o que são os chamados **repositórios**.
- ▶ Os repositórios nada mais são do que **locais** em que os dados serão armazenados, compreendendo também a **forma** como são alocados.
- ▶ Dentre os principais e mais utilizados pela Big Data está o **Data Lake**.

# Desenvolvimento de projetos baseados na análise de dados

- ▶ O **Data Lake** é, basicamente, um banco de dados que armazena os dados advindos diretamente de um sistema operacional ou fornecedor.
- ▶ Os dados **são brutos**, sem processamento ou com **processamento leve**.
- ▶ Pode ser desenvolvido na **nuvem**.
- ▶ É fundamental:
  - ✓ **Presença da diretoria** e gestores sêniores.
  - ✓ Entendimento do **modo de organização e armazenamento**.

# Desenvolvimento de projetos baseados na análise de dados

Figura 4 - Utilização do Data Lake na área de Big Data



Fonte: adaptada de <https://medium.com/data-ops/the-data-lake-is-a-design-pattern-888323323c66>. Acesso em: 6 nov. 2020.

# Desenvolvimento de projetos baseados na análise de dados

- ▶ Mas o que é armazenamento em **nuvem**?
  - ✓ **Repositório** em servidores da **web**.
  - ✓ Podem ser **privados, públicos** ou de caráter **híbrido**.
  - ✓ As públicas se destinam às **pessoas** como um todo.
  - ✓ As privadas são utilizadas pelas **empresas**.
  - ✓ Computação nas nuvens (***Cloud Computing***).





# Teoria em Prática

Bloco 4

Cassio Rodolfo Aveiro da Silva



## Refleta sobre a seguinte situação

Você, diretor da empresa Google, a qual existem muitas áreas e vários programas que estão sob a mesma empresa, como e-mails, sites de busca, agendas virtuais, mapas e vários aplicativos. Trata-se de uma empresa multinacional, que atua mundialmente e possui um volume de dados gigantesco, na faixa de zetabytes. Com a maior parcela de seu lucro vindo da publicidade, ela é uma das maiores empresas do mundo, embora seja recente, pois, foi fundada em 4 de setembro de 1998. O crescimento da empresa foi muito rápido e, com isso, culminou em diversos produtos inovadores e tecnológicos no mercado. O Google é executado em mais de um milhão de servidores e processa mais de 5 bilhões de solicitações de pesquisa em vários países diferentes.

## Refleta sobre a seguinte situação

Já foi tido como o site mais visitado do mundo e considerado uma das melhores empresas para se trabalhar. Com todas essas informações sobre o Google, é possível perceber com o conteúdo estudado que estamos tratando com o maior volume de dados do mundo. Quais as características de profissionais que você contrataria para sua empresa? Quais os princípios aplicados aos dados você recomendaria ter para que sua empresa continuasse a ter sucesso?

## Norte para a resolução...

- O perfil do profissional a ser selecionado deve contemplar, primeiramente, os princípios e cultura organizacional do Google. Neste contexto, o perfil deve ser constituído por uma atitude inovadora, sempre em busca de novidades e almejando fazer a diferença não só na empresa, mas no mundo. Além disso, liderança, capacidade cognitiva e familiaridade com sistemas de informação e banco de dados é fundamental.
- Claramente, para manter o sucesso, o foco está em analisar os princípios de redundância, inconsistência e integração.





# Dica do Professor

Bloco 5

Cassio Rodolfo Aveiro da Silva



## Orientações para o bloco 5

**Filme:** *Banco de Dados Relacional / O que é?* Canal: Dicionário de Informática, disponível no Youtube.

*Banco de Dados Relacional e Não Relacional.* Canal: Data Lab, disponível no Youtube.

### Capítulo de livro:

ELMASRI, R.; NAVATHE, S. B. **Sistemas de banco de dados**. 6. ed. São Paulo: Pearson, 2011. Cap. 2, p. 19-36.

## Referências

AMARAL, F. **Introdução à ciência de dados, mineração de dados e big data**. Rio de Janeiro: Alta Books, 2016.

GRUS, J. **Data science do zero: primeiras regras com Python**. Rio de Janeiro: Alta Books, 2016.

HEUSER, C. A. **Projeto de bancos de dados**. Porto Alegre: Bookman, 2009.

NOGARE, D. **Do Banco de dados relacional à tomada de decisão**. São Paulo: Mvtech, 2014.





Bons estudos!

