**Introdução - Fundamentos de Banco de Dados**

Esta disciplina se propõe ao estudo dos principais fundamentos de bancos de dados, para isto apresentando a modelagem de dados, operações relacionas aos bancos de dados e as técnicas de normalização de um modelo propiciando ao aluno o conhecimento para projetar um Banco de Dados desde o início.

Acompanhe a apresentação da disciplina, elaborada pelo Professor Marino Hilário Catarino - [vídeo](https://www.youtubeeducation.com/watch?v=G5w3lxeJZOE).

**Planejamento de estudos**

Esta disciplina é de 80 horas

Está previsto que você use cerca de 4 horas/semana

**Objetivos gerais**

Apresentar uma base sólida em fundamentos de banco de dados; introduzir como representar soluções de problemas utilizando modelos e diagramas de banco de dados, a fim de que seja possível organizar dados e informação em estruturas de banco de dados.

**Parte 1 -** [**Introdução aos sistemas de bancos de dados**](https://sites.google.com/faculdadeimpacta.com.br/fbd-p1?pli=1&authuser=2) **-** [**Vídeo**](https://www.youtubeeducation.com/watch?v=LiT1NoTy3ds)

**Apresentação**

Um banco de dados (BD) é uma coleção ou conjunto de dados inter-relacionados com uma estrutura regular que é armazenado de forma organizada para serem utilizados em situações específicas.

O Dado é toda a informação que possui um significado conforme um determinado contexto e que pode ser armazenado. Podemos exemplificar como um catálogo de filmes ou um sistema de controle de estoque em uma empresa.

**Objetivos**

A grande maioria dos bancos de dados são armazenados e acessados através de um software denominado Sistema Gerenciador de Banco de Dados (SGBD). O SGBD possui recursos que permitem uma interação do usuário e a manipulação das informações do banco de dados. Temos como exemplos de SGBD o SQL Server, Oracle, MySQL e o PostgreSQL. Em um SGBD, o modelo de dados mais utilizado para representar e armazenar os dados é o modelo relacional, no qual as estruturas são representadas por tabelas compostas por linhas e colunas formando as tuplas.

**Competências da disciplina**

Uma vantagem principal do SGBD é a persistência dos dados, porém ele possui outras propriedades que são muito importantes: abstrações de dados, escalabilidade, durabilidade, consistência e controle de concorrência, entre outras características.

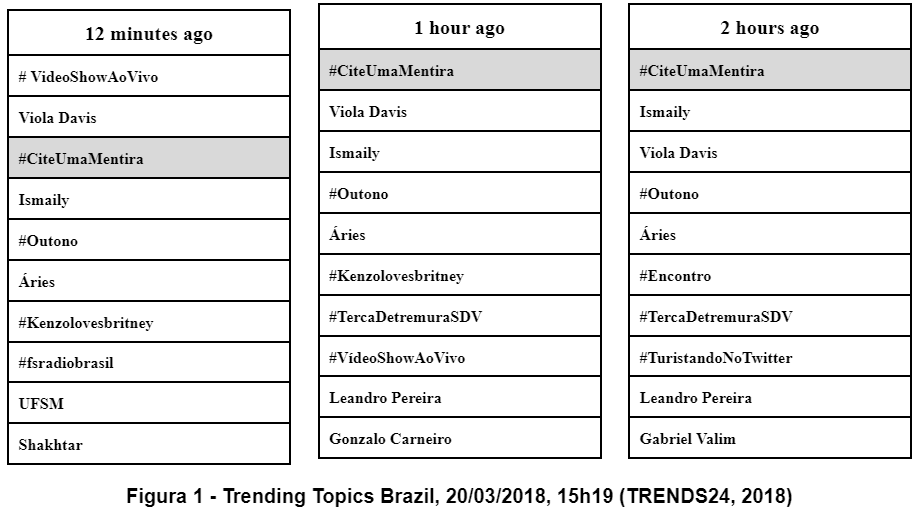
**Introdução**

Existem muitos tipos de banco de dados e eles estão presentes em nosso dia-a-dia: quando você vai ao supermercado, quando envia uma mensagem para alguém pelo celular, quando lê seus e-mails. Todas estas informações são armazenadas em bancos de dados e é o que estudaremos neste material.

Segundo Rangel et al (2014), os bancos de dados são amplamente utilizados e constituem a parte essencial de quase todas as empresas, independentemente do seu ramo de atividade.

Todas estas informações são dependentes de bancos de dados pois demandam grandes volumes de dados para serem armazenados e recuperados, sem contar mecanismos de recuperação em caso de falhas.

Mas, existe diferença entre informação e dados? É de fundamental importância conhecer a diferença entre estes dois conceitos. De acordo com Rangel et al (2014), “Os dados são considerados fatos brutos, o que indica que os fatos ainda não foram processados para revelar seu significado”. Por exemplo, todas as mensagens enviadas pelo Twitter em 1 minuto que, segundo Domo (2018), são da ordem de 456mil por minuto! Certamente, há muita informação nestas mensagens (twits) mas, primeiro é necessário que se faça um tratamento delas. Da mesma forma, pense em quantas compras são registradas por dia em um rede de lojas como o Carrefour ou o Walmart. Todos são dados brutos precisando ser lapidados. Estes dados “lapidados” são a informação, como, por exemplo: “Quanto se vendeu de sabão em pó hoje na rede de lojas?” ou, no caso do twitter, os “trending topics”. Um exemplo dos Trending Topics pode ser visto na figura 1.



Para se conseguir extrair tais informações de um número tão grande de dados, é preciso conhecer seu contexto e sua finalidade. As informações são essenciais para as empresas na tomada de decisão, independente da área que estejam inseridas: governamental, privado ou filantrópicas.

Antes do uso dos SGBDs (Sistemas Gerenciadores de Banco de Dados), as empresas armazenavam seus dados em sistemas de arquivos, que não tinham a capacidade nem de compartilhamento nem de proteção a estes dados.

Mas, e o que são os SGBDs? São os softwares que gerenciam e guardam os dados, dentre os quais podemos destacar o Oracle, SQL Server, Postgree, MySQL, Firebird, etc. Os softwares citados são apenas os relacionais, já que existem diversos tipos de SGBD, sendo estes os mais utilizados atualmente e fruto de avaliação neste material.

Para criar um banco de dados em um SGBDR (Sistema Gerenciador de Banco de Dados Relacional), precisamos criar um projeto que possui três níveis de abstração: Conceitual, Lógico e Físico.

O nível conceitual é composto pelo Projeto Conceitual que é o Modelo Entidade-Relacionamento utilizando-se o Diagrama Entidade-Relacionamento, onde estará o nível mais alto de abstração do projeto e, nenhum SGBD deve ser levado em consideração. Posteriormente, baseado neste projeto conceitual, será elaborado o projeto lógico que está situado no modelo Relacional, portanto, fazendo uso do Esquema de Relações do Banco de Dados, as quais devem ser aplicadas as regras de normalização para eliminação das redundâncias que possam haver. Neste nível, um SGBDR já pode ser considerado pois há pequenas diferenças entre eles que precisam ser levadas em conta.

“A função da normalização é atuar como um filtro sobre as entidades e os relacionamentos, eliminando alguns elementos sem causar perda de informação nas entidades e nas relações.” (RANGEL et al 2014). Estas regras de normalização são chamadas de formas normais e são em um total de 6, sendo:

1FN – Primeira Forma Normal

2FN – Segunda Forma Normal

3FN – Terceira Forma Normal

FNBC – Forma Normal de Boyce Codd

4FN – Quarta Forma Normal

5FN – Quinta Forma Normal

Findas estas fases, inicia-se o projeto físico que é elaboração dos scripts de construção das tabelas no SGBDR selecionado. Baseando-se então no Projeto Lógico que contém o esquema de relações do banco de dados, cria-se o script em Linguagem SQL, no dialeto do SGBDR escolhido para, em seguida aplicá-lo e criar efetivamente o conjunto de tabelas.

A Linguagem SQL (Structure Query Language ou Linguagem Estruturada de Consulta) é a que se utiliza para criação e manutenção dos bancos de dados relacionais. Ela está subdividida em três, a saber:

DDL – Data Definition Language ou Linguagem de Definição de Dados

DML – Data Manipulation Languagem ou Linguagem de Manipulação dos Dados

DCL – Data Control Language ou Linguagem de Controle de Dados

De acordo com Rangel et al (2014), a DDL (Data Definition Language) fornece recursos para definir objetos e controlar dados. São estes comandos que serão responsáveis pela estruturação do banco de dados, como, por exemplo, a criação das tabelas e índices.

Já o subconjunto DML (Data Manipulation Language), conforme afirmam Rangel et al (2014), objetivam mecanismos para manipular e gerenciar o banco de dados, permitindo assim que sejam inseridos, modificados, apagados e pesquisados quaisquer dados armazenados neles.

Por fim, a sub linguagem DCL (Data Control Language) é responsável pelo controle de acesso aos dados do banco de dados, com o gerenciamento de usuários e a criação de regras para realização de pesquisas, inserções, modificações e exclusões, por exemplo.

**Glossário de Conceitos**

1) Atributo: abstração de uma propriedade de uma entidade ou de um relacionamento.

2) Banco de Dados: sistema de armazenamento de dados cujo objetivo é registrar e guardar

informações importantes que poderão ser acessadas quando necessário.

3) BI (Business Intelligence ou Inteligência de Negócios): utiliza conceitos em que as informações são coletadas, armazenadas e analisadas, tendo como base fatos reais e/ou hipóteses. Esses sistemas auxiliam na gestão organizacional e no processo de tomada de decisões.

4) Dado atômico: tipo de dado considerado básico, ou seja, indivisível.

5) Dado não atômico: tipo de dado considerado complexo, divisíveis (fragmentados).

6) Entidade: abstração de um fato do mundo real para o qual se deseja manter seus dados no banco de dados.

7) ERP’s: são sistemas de gestão empresarial que possibilitam a integração de todos os dados e processos de uma empresa, melhorando o fluxo de informações.

8) Gerenciamento de Banco de Dados: utiliza Sistemas Gerenciadores de Banco de Dados (SGBDs).

9) Modelagem Conceitual: nível mais alto de abstração cujo objetivo é representar os requisitos de dados do domínio da aplicação (independente do modelo de banco de dados).

10) Modelagem Lógica: representação da modelagem conceitual em um modelo de banco de dados.

11) Modelagem Física: constitui um esquema SQL para a modelagem lógica (depende exclusivamente do SGBD).

12) Relacionamento: abstração de uma associação entre (ocorrências de) entidades.

13) Sistema Gerenciador de Banco de Dados (SGBD): coleção de programas responsáveis pela criação e manutenção de banco de dados. (RANGEL et al 2014)

**Perspectiva Histórica**

Com o passar do tempo, as empresas descobriram que a quantidade de dados gerados pelos sistemas de informação aumentava imensamente e que tornava-se cada vez mais dispendioso o processo de armazenagem destes dados. Desta forma, percebeu-se que era valido o esforço para descobrir uma forma mais eficiente de armazenar todos estes dados. (SANCHES, 2005)

Da mesma forma que muitas tecnologias da computação, os fundamentos de banco de dados relacionais surgiram na empresa IBM, em meados dos anos 1960 e 1970, como fruto de pesquisas por ela desenvolvidas. Diversas outras pesquisas desta época culminaram em outros modelos de bancos de dados como os hierárquicos e de rede. (RANGEL et al, 2014)

Em 1970, Ted Codd, então pesquisador da IBM publicou o primeiro artigo sobre bancos de dados relacionais, o qual tratava sobre o uso de cálculo e álgebra relacional. Ele procurava o desenvolvimento de um sistema que fosse capaz de acessar as informações com comandos em inglês. Como fruto deste artigo, a IBM idealizou o projeto System R. (SANCHES, 2005)

Este projeto visava criar um sistema de banco de dados relacional que deveria tornar-se um produto, o que posteriormente ocorreu, sendo primeiro o SQL/DS e depois o DB2, que ela produz até os dias atuais. A linguagem que foi criada pelo projeto System/R foi a linguagem SQL (Structure Query Language), que tornou-se o padrão para bancos de dados relacionais e atualmente é um padrão ISO (International Organization for Standardization). (RANGEL et al, 2014)

Os primeiros protótipos foram utilizados por muitas organizações, como o MIT Sloan School of Management. A IBM, no entanto, manteve o System/R em segundo plano por vários e decisivos anos. (SANCHES, 2005)

No final dos anos 1970, surge uma empresa, a Oracle, fruto do vislumbre de seu criador, Larry Elison que percebeu uma oportunidade onde outras empresas não haviam visto. Ele encontrou uma descrição de um protótipo funcional de um banco de dados relacional e descobriu que nenhuma empresa tinha se empenhado em comercializar essa tecnologia. (WIKIPEDIA, 2010)

Ellison e os cofundadores da Oracle Corporation, Bob Miner e Ed Oates, perceberam que havia um tremendo potencial de negócios ali, tornando-a assim uma das maiores empresas de software empresarial do mundo, cuja primeira versão comercial foi lançada em 1979 sob o nome de RSI. Atualmente chama-se Oracle Database e encontra-se na versão 12g. (FARIAS, 2018)

Vídeo - UNIVESP. Banco de dados - aula 01 - [visão geral sobre banco de dados e motivação](https://youtu.be/pmAxIs5U1KI?si=ay5wrMjtskeBYiOd).

**Parte 2 -** [**SGBD e modelos de bancos de dados**](https://sites.google.com/faculdadeimpacta.com.br/fbd-p2?pli=1&authuser=2) **-** [**Vídeo**](https://www.youtubeeducation.com/watch?v=_zJu06l9vbY)

**Fundamentos de Banco de Dados**

O que é um banco de dados? Por que utilizamos banco de dados? De onde surgiram os bancos de dados? O que são os Sistemas Gerenciadores de Bancos de Dados e para que eles servem? Estas são algumas das questões que se pretende responder neste capítulo. Segundo Silberschatz, Korth e Sudarshan (2008), um banco de dados “é uma coleção de dados inter-relacionados, representando informações sobre um domínio específico”, ou seja, sempre que for possível agrupar informações que se relacionam e tratam de um mesmo assunto, posso dizer que tenho um banco de dados.

**Bancos de Dados**

Segundo Date (2000), em essência, um sistema de bancos de dados é apenas um sistema computadorizado de armazenamento de registros. O banco de dados pode, ele próprio, ser visto como o equivalente eletrônico de um armário de arquivamento.

O crescimento do uso de computadores está intimamente relacionado a utilização de bancos de dados pelos sistemas computacionais. O termo banco de dados é tão utilizado que precisa, primeiramente ser definido: “Um banco de dados é uma coleção de dados relacionados”.

Já por dados, podemos tomar quaisquer fatos conhecidos que possam ser armazenados e que possuam um significado implícito, como, por exemplo, nomes, números de documentos, telefones e e-mails.

Segundo Rangel et al (2014), “É importante que você entenda a diferença entre dados e informações. Os dados são considerados fatos brutos, o que indica que os fatos ainda não foram processados para revelar seu significado.”. Desta forma, ainda de acordo com o autor, “Para revelar seu significado, os dados brutos são processados de maneira apropriada, gerando, assim, as informações.”.

Desta forma, os bancos de dados começaram a crescer e o volume de dados tornava trabalhoso e dispendioso sua manutenção. A IBM iniciou um grupo de pesquisa, denominado System/R para pesquisar sobre uma maneira mais eficiente de armazenar-se administrar-se estes dados, de forma que posteriormente pudesse se tornar um produto. Um de seus pesquisadores (Ted Codd) publicou um artigo descrevendo detalhadamente como deveriam ser o uso de cálculos e álgebra relacional. Um dos leitores deste artigo era Larry Elison que posteriormente seria um dos fundadores da Oracle, o primeiro SGBDR (Sistema Gerenciador de Banco de Dados Relacional) comercial e atualmente, um dos principais do mercado. (FARIAS, 2018)

**Sistema Gerenciador de Bancos de Dados**

Um SGBD (Sistema Gerenciador de Banco de Dados) é uma coleção de programas que permite aos usuários criar e manter um banco de dados, permitindo-se a definição, manipulação e compartilhamento de dados entre aplicações diversas e seus usuários. (EMASRI; NAVATHE, 2011)

São várias as características que distinguem o funcionamento de um SGBD em relação ao modus operandi anterior a eles com uso de arquivos. Quando se trabalha com arquivos, cada programa deve definir e implementar todos os arquivos necessários para a execução de uma aplicação. (EMASRI; NAVATHE, 2011)

Um bom exemplo destes tipos de sistemas foram os programas em COBOL (Common Business Oriented Language ou Linguagem Comum Orientada para os Negócios) que foi uma linguagem de programação orientada ao processamento de bancos de dados comerciais. O COBOL foi criado no segundo semestre de 1959! A linguagem possui várias versões de padronização, sendo a última em 2002, que substituiu a versão do COBOL-85 (1985). (WIKIPEDIA, 2004)

Os programas em COBOL, como a versão Microbase (brasileira) trabalhavam com arquivos próprios, basicamente, texto ASCII, que podia ser acessado por qualquer editor de texto e ter seus dados modificados. Não havia controle de transações e a segurança era a que a aplicação fosse capaz de fornecer.

Por exemplo, em COBOL, a organização de arquivos indica como os registros são organizados em um arquivo, podendo ser sequencial ou indexado. Cada modo de organização podia ter diferentes formas de acesso. Isso significa que, a cada programa, estas configurações precisavam ser especificadas e, em outras linguagens seriam completamente diferentes. No caso de uma modificação na estrutura de um arquivo, todos os programas que fazem acesso aquele arquivo, independente da forma de acesso, se para escrita ou leitura, deveriam ser modificados. Este foi o grande problema do “Bug do Milênio”, onde, mesmo arquivos que não utilizassem as datas, se faziam acesso a um arquivo que contivesse datas, precisavam ser modificados.

Agora, multiplique este trabalho por centenas de programas. Insano! Se uma modificação semelhante precisar ser feita em um banco de dados, o trabalho será infinitamente menor.

| 01 | IDENTIFICATION DIVISION. |
| --- | --- |
| 02 | PROGRAM-ID. HELLO. |
| 03 |  |
| 04 | ENVIRONMENT DIVISION. |
| 05 | INPUT-OUTPUT SECTION. |

| 06 | FILE-CONTROL. |
| --- | --- |
| 07 | SELECT STUDENT ASSIGN TO OUT1 |
| 08 | ORGANIZATION IS SEQUENTIAL |
| 09 | ACCESS IS SEQUENTIAL |
| 10 | FILE STATUS IS FS. |
| 11 |  |
| 12 | DATA DIVISION. |
| 13 | FILE SECTION. |
| 14 | FS STUDENT |
| 15 | 01 STUDENT-FILE. |
| 16 | 05 STUDENT-ID PIC 9(5). |
| 17 | 05 NAME PIC A(25). |
| 18 | 05 CLASS PIC X(3). |
| 19 |  |
| 20 | WORKING-STORAGE SECTION. |
| 21 | 01 WS-STUDENT. |
| 22 | 05 WS-STUDENT-ID PIC 9(5). |
| 23 | 05 WS-NAME PIC A(25). |
| 24 | 05 WS-CLASS PIC X(3). |
| 25 |  |
| 26 | PROCEDURE DIVISION. |
| 27 | OPEN EXTEND STUDENT. |
| 28 | MOVE 1000 TO STUDENT-ID. |
| 29 | MOVE ‘Tim”TO NAME. |
| 30 | MOVE ‘10’TO CLASS. |
| 31 | WRITE STUDENT-FILE |

| 32 | END-WRITE. |
| --- | --- |
| 33 | CLOSE STUDENT. |
| 34 | STOP RUN. |

**Listagem 1 – Programa em COBOL. (TUTORIALSPOINT, 2018)**

Percebe-se então que o controle de redundância de dados, de segurança da informação e a eficiência das consultas ficam seriamente comprometidas. Nestas situações, o SGBD apresenta-se como solução mais robusta e indicada. Entretanto, não se deve usar SGBDs indiscriminadamente.

**Quando não utilizar um Sistema Gerenciador de Bancos de Dados**

Em algumas situações, como em aplicações muito simples e estáveis, provavelmente, a utilização do SGBD torne-a demasiadamente complicada, fazendo com que se perca sua simplicidade.

Segundo Elmasri e Navathe (2011, p. 17), “Apesar das vantagens de usar um SGBD, existem algumas situações em que esse sistema pode envolver custos adicionais desnecessários, que não aconteceriam no processamento de arquivos tradicional”.

Os custos adicionais quando da utilização de um SGBD devem-se aos seguintes fatores:

Alto investimento inicial em hardware, software e treinamento.

A generalidade que um SGBD oferece para a definição e o processamento de dados.

Esforço adicional para oferecer funções de segurança, controle de concorrência, recuperação e integridade. (ELMASRI; NAVATHE, 2011)

Desta forma, pode ser mais vantajoso utilizar sistemas de arquivos nas seguintes situações:

* Aplicações de banco de dados simples e bem definidas, para as quais não se espera muitas mudanças.
* Requisitos rigorosos, de tempo real, para alguns programas de aplicação, que podem ser atendidos devido as operações extras executadas pelo SGBD.
* Sistemas embarcados com capacidade de armazenamento limitada, onde um SGBD de uso geral não seria indicado.
* Nenhum acesso de múltiplos usuários aos dados. (ELMASRI; NAVATHE, 2011)

Vídeo - [SGBD - Conceitos Básicos](https://youtu.be/NnG7f60XPkQ?si=XB8GdUHBpMWx0mHa)

**Parte 3 -** [**SGBD e modelos de bancos de dados**](https://sites.google.com/faculdadeimpacta.com.br/fbd-p2?pli=1&authuser=2) **-** [**Vídeo**](https://www.youtubeeducation.com/watch?v=_zJu06l9vbY)