

Aula 1 – Conceitos fundamentais

Objetivos

Compreender os conceitos básicos sobre banco de dados.

Identificar as principais características de um SGBD.

1.1 Introdução

Todos nós, em nosso dia a dia, temos a necessidade de armazenar e recuperar dados. Cadernos de endereços, listas de telefones, dados financeiros, receitas, enfim, estamos sempre lançando mão de memórias auxiliares. O ideal seria que estas memórias pudessem ser seguras, confiáveis e estivessem disponíveis quando precisássemos dela. Um dado guardado, mas que não sabemos como recuperar é praticamente tão inútil quanto se não existisse. Assim, temos uma grande necessidade de armazenar informações que não se encontram isoladas, como por exemplo, as fichas de matrícula de um aluno que contém informações diversas sobre o mesmo, conforme a Figura 1.1.

INSTITUTO FEDERAL DE EDUCAÇÃO, CIÊNCIA E TECNOLOGIA DE SÃO PAULO	
FICHA INDIVIDUAL DO ALUNO	
NOME:	
ENDEREÇO:	
RG:	
PRONTUÁRIO:	
CURSO:	
ANO INGRESSO:	

Figura 1.1: Ficha de um aluno

Fonte: Elaborada pelo autor

Além de uma forma adequada para definir o armazenamento destas informações, os usuários desejam realizar operações sobre esta coleção de dados, tais como: adicionar novos dados, consultar um determinado subconjunto de dados, atualizar ou modificar a estrutura dos dados e eliminar informações desnecessárias ou que já perderam a validade por um motivo ou outro, de desatualização ou desuso.

Uma solução para este problema foi apresentada com o advento da tecnologia em Bancos de Dados (BD, *Database* em Inglês). Para que você comece a dominar o tema em estudo apresentamos várias definições, assim um banco

de dados é uma coleção de dados relacionados. Os dados são fatos que podem ser gravados e possuem um significado implícito. Assim, considere o exemplo anterior da ficha com dados de um aluno ou uma lista telefônica. Esses dados podem ser armazenados em uma ficha, uma agenda ou em um computador por meio de aplicativos como a Microsoft® Excel ou Access. Essas informações têm um significado implícito, formando um banco de dados. Algumas definições mais precisas de Banco de Dados podem ser conhecidas em (ELMASRI & NAVATHE, 2005):

- a) Um Banco de Dados é uma coleção logicamente coerente de dados com um determinado significado inerente. Isto significa que um conjunto aleatório de dados não pode ser considerada um Banco de Dados.
- b) Um Banco de Dados é projetado, construído e composto por um conjunto de dados para um propósito específico (como no nosso exemplo acima para armazenar um conjunto de informações de alunos). Existe um grupo de usuários ou algumas aplicações pré-concebidas onde estes dados serão utilizados.
- c) Um Banco de Dados representa aspectos de uma parte restrita do mundo real, denominado de minimundo. Alterações que ocorra no minimundo são refletidas em todo o Banco de Dados.

Um banco de dados computadorizado é mantido por um grupo de aplicativos chamado de Sistema de Gerenciamento de Bancos de Dados (SGBD), que é uma coleção de programas que permite ao usuário definir, construir e manipular um Banco de Dados para as mais diversas aplicações. O objetivo principal de um sistema de banco de dados é possibilitar um ambiente que seja adequado e eficiente para uso na recuperação e armazenamento de informações.

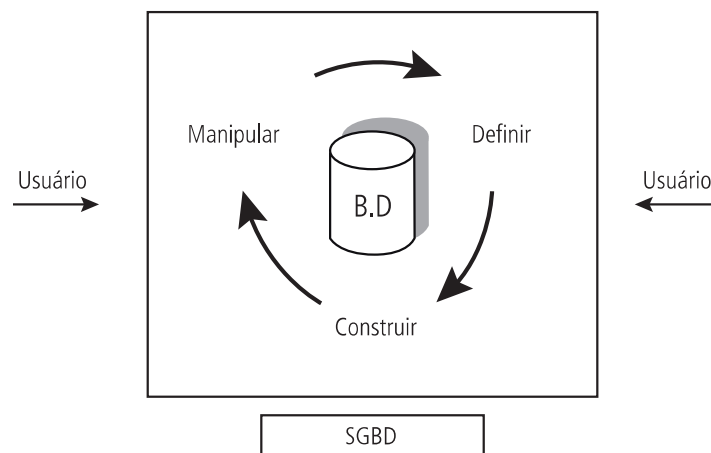


Figura 1.2: Sistema de Gerenciamento de Banco de Dados

Fonte: Elaborada pelo autor

- a) Banco de Dados e seu software são juntos denominados de Sistema de Bancos de Dados (SBD).

Imagine um armário de aço, com várias gavetas, em cada gaveta contém alguma informação (como a ficha do aluno) que estão agrupadas de acordo com seu tipo. O armário no caso é forma de gerenciamento dos dados ali contidos, lá podemos: inserir, excluir, selecionar ou alterar algum documento que ali contenha. Neste primeiro momento podemos pensar que um banco de dados computacional consiste em “levar” os dados deste armário de aço para o computador, porém seguiremos algumas regras para que o armazenamento seja mais eficiente.

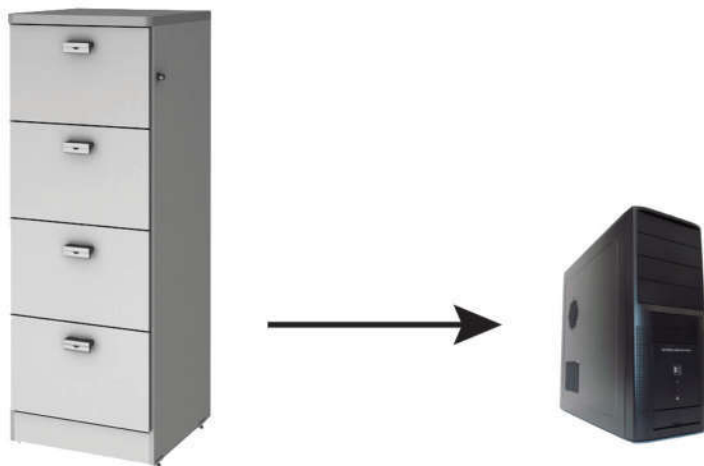


Figura 1.3: Armazenamento de dados

Fonte: Elaborada pelo autor

1.2 História

Os primeiros SBDs foram lançados no mercado no final da década de 60 e eram conceitualmente muito simples, de acordo com as necessidades das aplicações da época. No início as empresas que impulsionaram este segmento foram a IBM, ORACLE e SYBASE. Assim, como outras tecnologias na computação, os fundamentos de bancos de dados relacionais nasceram na empresa IBM, nas décadas de 1960 e 1970, por meio de pesquisas de funções de automação de escritório. Foi durante um período da história nas quais as empresas verificaram que era necessário empregar muitas pessoas, além do alto custo para fazer trabalhos, tais como: armazenar e organizar seus arquivos. Por este motivo, eram importantes os esforços e investimentos em pesquisa para obter-se um meio mais barato e eficiente de armazenamento de dados.

Em 1970 a IBM publicou o primeiro trabalho sobre bancos de dados relacionais. Este trabalho tratava sobre o uso de cálculo e álgebra relacional para que usuários não técnicos pudessem manipular grande quantidade de informações.

Devido à natureza técnica e a relativa complicação matemática, o significado e proposição do trabalho não foram prontamente realizados. Assim, a IBM montou um grupo de pesquisa conhecido como System R.

O projeto do Sistema R tinha como proposta a criação de um sistema de banco de dados relacional o qual eventualmente se tornaria um produto. Os primeiros protótipos foram utilizados por várias organizações, como MIT Sloan School of Management. Novas versões foram testadas com empresas de aviação.

O Sistema R evoluiu para SQL/DS tornando-se o DB2 (Database 2, em inglês). O grupo do Sistema R criou a Structured Query Language (SQL) - Linguagem de Consulta Estruturada. Esta linguagem tornou-se um padrão na indústria para bancos de dados relacionais e hoje em dia é um padrão ISO (International Organization for Standardization).

Mesmo a IBM sendo a empresa que inventou o conceito original e o padrão SQL, ela não produziu o primeiro sistema comercial de banco de dados. Isto foi realizado pela Honeywell Information Systems Inc., tendo seu sistema lançado em junho de 1976. O sistema foi desenvolvido com base em muitos dos princípios que a IBM concebeu, mas foi modelado e implementado fora da IBM.



Um banco de dados relacional organiza seus dados em **relações**. Cada relação pode ser vista como uma **tabela** onde cada coluna corresponde a **atributos** da relação e as linhas correspondem às **tuplas** ou elementos da relação. Em uma nomenclatura mais próxima àquela de sistemas de arquivos, muitas vezes as **tuplas** são denominadas **registros** e os **atributos**, **campos**.

1.3 Usuários (atores de um banco de dados)

Um Banco de Dados pode apresentar diversos usuários cada qual com uma necessidade em especial, e com um envolvimento diferente com os dados. Os usuários podem ser classificados em categorias:

- **Administradores de bancos de dados (DBA)**: em qualquer organização onde muitas pessoas compartilham muitos recursos, existe a necessidade de um administrador chefe para supervisionar e gerenciar estes re-

curso. Em um ambiente de base de dados, o recurso primário é a própria base de dados e os recursos secundários são o próprio SGBD e *software* relacionados. A administração desses recursos é de responsabilidade do DBA ("Database Administrator"). O DBA é responsável por autorizar acesso à base de dados e coordenar e monitorar seu uso, como também é responsável por problemas, tais como: quebra de segurança ou baixo desempenho. Em grandes organizações, existe uma equipe de DBAs.

- **Analistas de bancos de dados (projetistas)**: possuem a responsabilidade de identificar os dados a serem armazenados no BD e pela escolha da estrutura apropriada utilizada para armazená-los. Eles devem se comunicar com os possíveis usuários do BD, obter a visão dos dados que cada um possui, integrando-as de forma a se obter uma representação adequada de todos os dados. Estas visões são então analisadas e, posteriormente, integradas para que, ao final, o projeto da base de dados possa ser capaz de dar suporte aos requisitos de todos os grupos de usuários.
- **Usuários finais**: existem profissionais que precisam ter acesso à base de dados para consultar, modificar e gerar relatórios. A base de dados existe para estes usuários. Podem ser usuários comuns do sistema, analistas de negócios ou usuários avançados que necessitam de uma maior interação com o BD.
- **Analistas de sistemas e programadores de aplicações**: os analistas determinam os requisitos dos usuários finais e desenvolvem especificações para transações que atendam estes requisitos, e os programadores implementam estas especificações, como: programas, testando, depurando, documentando e dando manutenção no mesmo. É importante que, tanto analistas quanto programadores, estejam a par dos recursos oferecidos pelo SGBD.

1.4 Níveis de abstração

A arquitetura ANSI/SPARC prevê múltiplas visões de dados, um esquema conceitual (lógico) e um esquema interno (físico). Um SBD se divide em geral em três níveis:

- **Nível externo**: possui as diversas descrições do BD de acordo com os grupos de usuários.

- **Nível conceitual:** descreve a estrutura de todo o BD para uma determinada comunidade de usuários, ocultando detalhes sobre a organização física dos dados e apresentando a descrição lógica dos dados e das ligações existentes entre eles.
- **Nível interno:** descreve a estrutura de armazenamento físico dos dados do BD, descreve o modelo físico dos dados que inclui detalhes sobre os caminhos de acesso aos dados internamente.

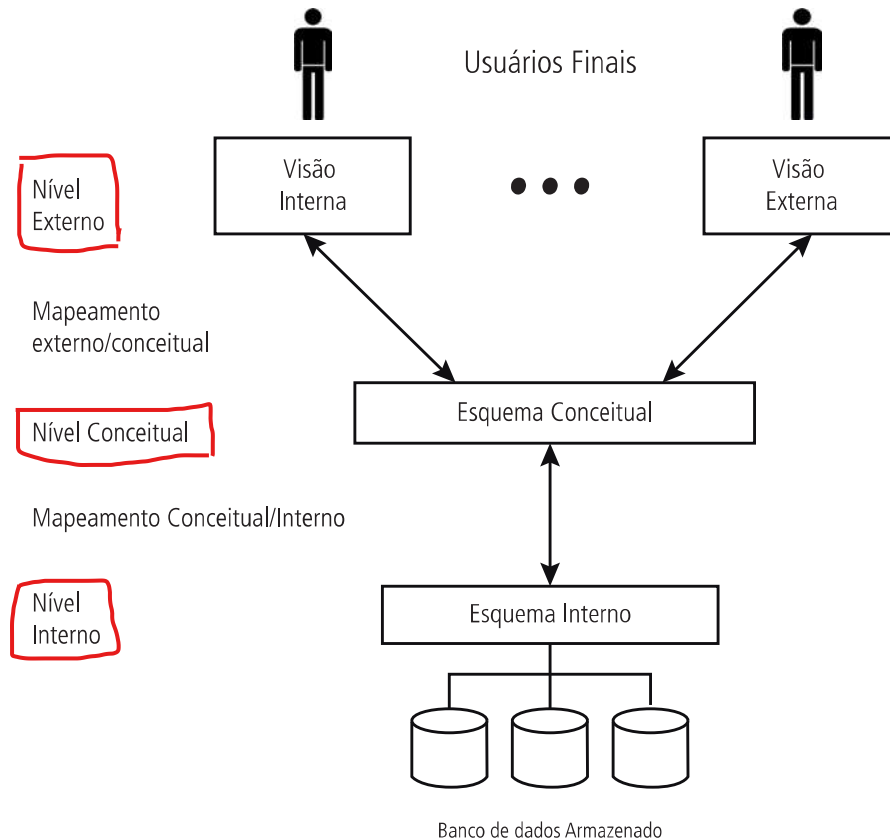


Figura 1.4: Visões de um banco de dados

Fonte: ELMASRI & NAVATHE, 2005



Sobre os níveis de abstração de um banco de dados podemos pensar assim: Nível externo é o que o usuário pensa e quer e deseja visualizar, nível conceitual é como o projetista irá implementar o banco de dados, e nível interno é como estes dados serão armazenados, formas de acesso físico por exemplo.

Observe que os três níveis apresentam apenas descrições dos dados. Como os três níveis apresentam descrições diferentes para os mesmos dados, torna-se necessário converter uma representação em outra, ou seja, definir mapeamentos de dados entre os níveis.

1.5 SGBD

Um Sistema de Gerenciamento de Bancos de Dados tem como principais propriedades:

- **Controle de redundância:** em um sistema tradicional de controle de arquivos cada usuário normalmente apresenta seus próprios arquivos armazenando o conjunto de dados que é de seu interesse, e nestes casos é comum ocorrer redundância de dados. Esta redundância consiste no armazenamento de uma mesma informação em locais diferentes, o que pode provocar sérios problemas. Alguns destes problemas consistem inicialmente no aumento de esforço computacional para realizar a atualização destes dados; aumento do espaço necessário para o armazenamento dos dados. O problema mais sério é que a representação dos dados desta forma pode tornar-se inconsistente, pois duas informações podem aparecer em locais distintos, mas apresentando valores diferentes. Em um sistema de BD as informações só se encontram armazenadas em um único local ou estão existindo duplicação controlada dos dados.
- **Compartilhamento dos dados:** um SGBD deve incluir um *software* para o controle de concorrência ao acesso dos dados em um ambiente multiusuário, de forma que possibilite o compartilhamento dos dados, garantindo que se vários usuários tentem realizar operações de atualização sobre um mesmo conjunto de dados, o resultado destas operações possa ser correto e consistente.
- **Controle de acesso:** quando vários usuários compartilham os dados, é comum que alguns não apresentem autorização para acesso a todo o BD. Por exemplo, os dados do faturamento de uma empresa podem ser considerados confidenciais e, desse modo, somente pessoas autorizadas devem ter acesso. Além disso, pode ser permitido, a alguns usuários, apenas a recuperação dos dados. Já, para outros, são permitidas a recuperação e a modificação. Assim, o tipo de operação de acesso - recuperação ou modificação - pode também ser controlado. Tipicamente, usuários e grupos de usuários recebem uma conta protegida por senhas, que é usada para se obter acesso à base de dados, o que significa dizer que contas diferentes possuem restrições de acesso diferentes. Um SGBD deve fornecer um subsistema de autorização e segurança, que é usado pelo DBA para criar contas e especificar restrições nas contas. O SGBD deve então obrigar estas restrições automaticamente.

- **Possibilidade de múltiplas interfaces:** diversos usuários com níveis diferenciados de conhecimento técnico representam necessidades diversas no que se refere aos tipos de interfaces fornecidas pelo SGBD. Interfaces para consultas de dados, programação, e interfaces baseadas em menus ou em linguagem natural, são exemplos de alguns tipos que podem estar disponíveis.



Um usuário com conhecimento técnico pode utilizar uma interface em que recupera dados mediante uma linguagem de consulta de dados, como SQL, já usuários com menor conhecimento técnico devem utilizar uma interface gráfica onde visualizam os dados e os selecionam para obter a consulta que necessitam ao invés de utilizarem uma linguagem para isto.

- **Representação de relacionamento complexo entre dados:** uma base de dados pode possuir uma variedade de dados que estão inter-relacionados de muitas formas. Um SGBD deve ter a capacidade de representar uma variedade de relacionamentos complexos entre dados, bem como recuperar e modificar dados relacionados de maneira fácil e eficiente.
- **Forçar restrições de integridade:** a maioria das aplicações de um banco de dados apresenta serviços que possibilitam garantir a integridade dos dados no BD. A restrição de integridade mais simples consiste na especificação do padrão de formato para os dados ou valores assumidos como um padrão.



Como exemplo de restrição de integridade, podemos pensar quando queremos armazenar o nome e a idade de uma pessoa. O nome deve ser uma cadeia de caracteres (*string*) menor que 50 caracteres alfabéticos, já a idade deve ser um dado numérico inteiro menor que 150. Estes são dois exemplos de restrições que podemos aplicar ao armazenar dados com intuito de garantir sua integridade.



Alguns exemplos de SGBDs modernos que apresentam as vantagens citadas: MySQL, Oracle, MS SQL Server PostgreSQL. Para exemplificar nossas aulas utilizaremos o SGBD MySQL, que recomendo a todos já realizarem o *download* no site: <http://www.mysql.com/downloads/mysql/>, e posteriormente realizarem a instalação.

- **Garantir *backup* e restauração de dados:** um SGBD deve prover recursos para realização de cópias de segurança e restauração caso ocorra falhas de *hardware* ou *software*. O subsistema de ***backup* e restauração** do SGBD é o responsável pela restauração. Por exemplo, se o sistema de computador falhar no meio da execução de um programa que esteja realizando uma alteração complexa na base de dados, o subsistema de restauração é responsável em assegurar que a base de dados seja restaurada no estado anterior ao início da execução do programa. Alternativamente, o subsistema de restauração poderia assegurar que o programa seja reexecutado a partir do ponto em que havia sido interrompido.

1.5.1 Vantagens de um SGBD

A escolha da tecnologia adequada de Banco de Dados e sua correta utilização trazem benefícios à maioria das organizações, tais como:

- **Potencial para garantir padrões:** a abordagem de base de dados permite que o Administrador do Banco (DBA) defina e force a padronização entre os usuários da base de dados em grandes organizações. Isso facilita a comunicação e a cooperação entre vários departamentos, projetos e usuários. Padrões podem ser definidos para: formatos de nomes, elementos de dados, telas, relatórios, terminologias, etc. O DBA pode obrigar a padronização em um ambiente de base de dados centralizado, muito mais facilmente que em um ambiente onde cada usuário ou grupo tem o controle de seus próprios arquivos e *software*.
- **Redução do tempo de desenvolvimento de aplicações:** um dos principais argumentos de venda para o uso da abordagem de um banco de dados é o tempo reduzido para o desenvolvimento de novas aplicações, tal como a recuperação de certos dados da base de dados para a impressão de novos relatórios. Projetar e implementar uma nova base de dados pode tomar mais tempo do que escrever uma simples aplicação de arquivos especializada. Porém, uma vez que a base de dados esteja em uso, geralmente o tempo para se criar novas aplicações, usando-se os recursos de um SGBD, é bastante reduzido. O tempo para se desenvolver uma nova aplicação em um SGBD é estimado em 1/4 a 1/6 do que o tempo de desenvolvimento, usando-se apenas o sistema de arquivos tradicional.
- **Independência de dados:** as aplicações de banco de dados não devem depender da forma como os dados estão representados e/ou armazenados.
- **Flexibilidade:** pode ser necessário alterar a estrutura de uma base de dados devido a mudanças nos requisitos. Por exemplo, um novo grupo de usuários pode surgir com necessidade de informações adicionais, não disponíveis atualmente na base de dados. Um SGBD moderno permite que tais mudanças na estrutura da base de dados sejam realizadas sem afetar a maioria dos programas de aplicações existentes.
- **Disponibilidade para atualizar as informações:** um SGBD disponibiliza o banco de dados para todos os usuários. Imediatamente após um usuário modificar uma base de dados, todos os outros usuários “sen-

tem” imediatamente esta modificação. Essa disponibilidade de informações atualizadas é essencial para muitas aplicações, tais como: sistemas de reservas de passagens aéreas ou bases de dados bancárias. Isso somente é possível devido ao subsistema de controle de concorrência e restauração do SGBD.

- **Economia de escala:** a abordagem de SGBDs permite a consolidação de dados e de aplicações, reduzindo-se, desse modo, o desperdício em atividades redundantes de processamento em diferentes projetos ou departamentos. Isto possibilita à organização como um todo investir em processadores mais poderosos e periféricos de armazenamento e de comunicação mais eficientes do que cada departamento adquirir seu próprio (menos potente) equipamento, o que reduz o custo total da operação e gerenciamento.

Apesar de todas as facilidades oferecidas por um banco de dados, um projeto de implantação pode gerar um alto custo inicial para a organização.

1.5.2 Linguagens

Uma vez que o projeto do BD tenha se completado e um determinado SGBD tenha sido escolhido para a sua implementação, o primeiro passo consiste em realizar uma especificação dos esquemas conceituais e internos, e os respectivos mapeamentos entre eles. Para estas etapas o SGBD oferece algumas linguagens apresentadas a seguir:

- **Linguagem de definição de dados (DDL - Data Definition Language):** utilizada pelos analistas e projetistas do BD para a definição dos esquemas do banco de dados. O SGBD também apresentará um interpretador para a DDL, o qual será responsável pelo processamento dos comandos da DDL, e realiza o armazenamento do esquema definido em estruturas internas do BD. Por exemplo, os comandos para criar, definir índice de uma tabela fazem parte da linguagem de definição de dados. Uma vez definido e preenchido o BD com os seus dados, estes normalmente sofrerão uma série de operações de acesso às informações nele armazenado.
- **Linguagem de manipulação de dados (DML):** o SGBD fornece esta linguagem para a especificação das operações de acesso ao banco. Os comandos da DML podem aparecer embutidos em outra linguagem (geralmente uma linguagem de programação de alto nível), e neste caso esta é denominada de Linguagem hospedeira, e DML é denominada de

Sublinguagem de dados. De outra forma, se DML for utilizada isoladamente de uma forma interativa, passa a ser denominada de Linguagem de consulta (ou "*query language*" como a SQL).

- **Linguagem de Controle de Dados (DCL)**: no controle de acesso e transações dos dados utiliza-se esta linguagem, que inclusive possibilita estabelecer os diversos níveis de segurança de cada usuário.

Resumo

Nesta aula você pode compreender a definição de um banco de dados como uma coleção de dados relacionados, na qual os dados significam fatos registrados. Um SGBD é um pacote de *software* para implementação e manutenção de dados computadorizados. Em seguida, discutimos as categorias principais de usuários de um banco de dados. Depois, apresentamos uma lista de capacidades que devem ser oferecidas por um *software* SGBD. Finalmente, vimos as linguagens envolvidas com banco de dados.