



# SQL para Análise de Dados

Prof. Marcondes Alexandre

# Agenda

- **Visão Geral da Arquitetura SGBD Relacional**
- **Modelagem de Dados Relacional**

# Visão Geral da Arquitetura SGBD Relacional

---

# Sistema de Arquivos

Primeira arquitetura de sistemas para armazenamento e manipulação de dados e geração de informação.

- **Inconvenientes desta arquitetura:**
  - Definição das estruturas de arquivos inseridas no próprio código do aplicativo (alta dependência entre dados e aplicação) dificultando os trabalhos de manutenção;
    - compartilhamento de um arquivo por vários programas.
    - definição das estruturas de arquivos duplicadas nos programas.

# Sistema de Arquivos

- Arquivos e programas de um mesmo sistema são desenvolvidos de forma isolada por diferentes programadores e até mesmo em linguagens diferentes.
- **Inconsistência, redundância**, dificuldade de acesso, **isolamento de dados** e problemas com segurança.
- Falta de gerenciamento para **acessos concorrentes** aos dados e **recuperação de dados**.

# Sistema de Arquivos

- **Inconsistência e redundância de dados:**
  - Se a mesma informação está repetida em diversos lugares (arquivos diferentes) ela:
    - está redundante, aumentando os custos de armazenamento e;
    - pode passar para um estado inconsistente, com valores divergentes nas diferentes réplicas.

# Sistema de Arquivos

- **Dificuldade de acesso aos dados:**
  - A geração de informação pode surgir, durante o tempo em que o sistema está em produção, sob diferentes aspectos. Cada requisição de informação diferente no sistema de arquivos, vai gerar a necessidade da criação de um programa aplicativo. A recuperação de informação não é atendida de modo eficiente.

# Sistema de Arquivos

- **Isolamento de dados:**
  - Os dados estão armazenados em arquivos distintos, que não possuem qualquer tipo de relacionamento direto, e ainda podem conter diferentes formatos para o mesmo dado.
- **Problemas de integridade:**
  - É difícil manter “restrições de integridade” automaticamente, por exemplo:
    - O balanço de uma conta bancária não pode cair abaixo de um determinado valor.
    - Sempre que o saldo de uma conta for superior a um valor X, parte deste saldo deve ser automaticamente aplicado na poupança.



# Sistema de Arquivos

- **Problemas de atomicidade:**
  - Algumas operações em um sistema devem ser “atômicas” (indivisíveis). Ou essas operações são, em seu conjunto, executadas até o fim, ou nenhuma delas deve ser executada.
- **Problemas de segurança:**
  - Nem todos os usuários do sistema devem estar autorizados a ver/acessar todos os dados armazenados. Uma vez que os programas de aplicação são inseridos no sistema como um todo, é difícil implementar e garantir a efetividade de regras de segurança.

# Sistema de Arquivos

- **Anomalias no acesso concorrente:**
  - A melhora de desempenho em um sistema pode ocorrer pela execução simultânea de diversas operações. Geralmente, nos sistemas de arquivos, esta melhoria seria difícil de ser implementada sem levar a danos na consistência dos dados. Considere a seguinte situação num sistema bancário:

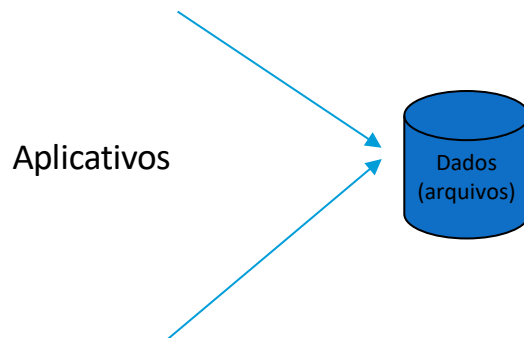
Suponha que o saldo de uma conta bancária A seja 500 reais. Se dois clientes retiram fundos desta conta A ao mesmo tempo (acesso concorrente à conta A), um estado inconsistente pode ocorrer se na execução das duas instâncias do programa de débito, ambos os clientes leiam o saldo antigo e retirem, cada um, seu valor correspondente, e seja então armazenado o valor restante.

# Sistema de Arquivos

- Instanciando o problema ...
  - Ambos leêm o valor 500;
  - Um tira 50 reais (resultando 450 reais) e o outro 100 reais (resultado 400 reais);
  - Dependendo de qual execução do programa de débito registre o saldo restante primeiro, o valor do saldo da conta será 450 ou 400 reais, quando deveria ser 350 reais.

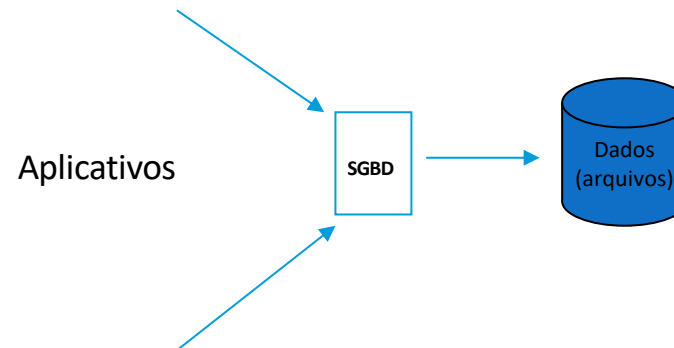
# Sistema de Arquivos

## Sistema de arquivos



O acesso/gerenciamento aos/dos dados é feito diretamente pelos programas aplicativos.

## Sistema de Banco de Dados



O acesso/gerenciamento aos/dos dados é feito pelo SGBD. O SGBD funciona como uma interface entre o BD e os programas aplicativos.

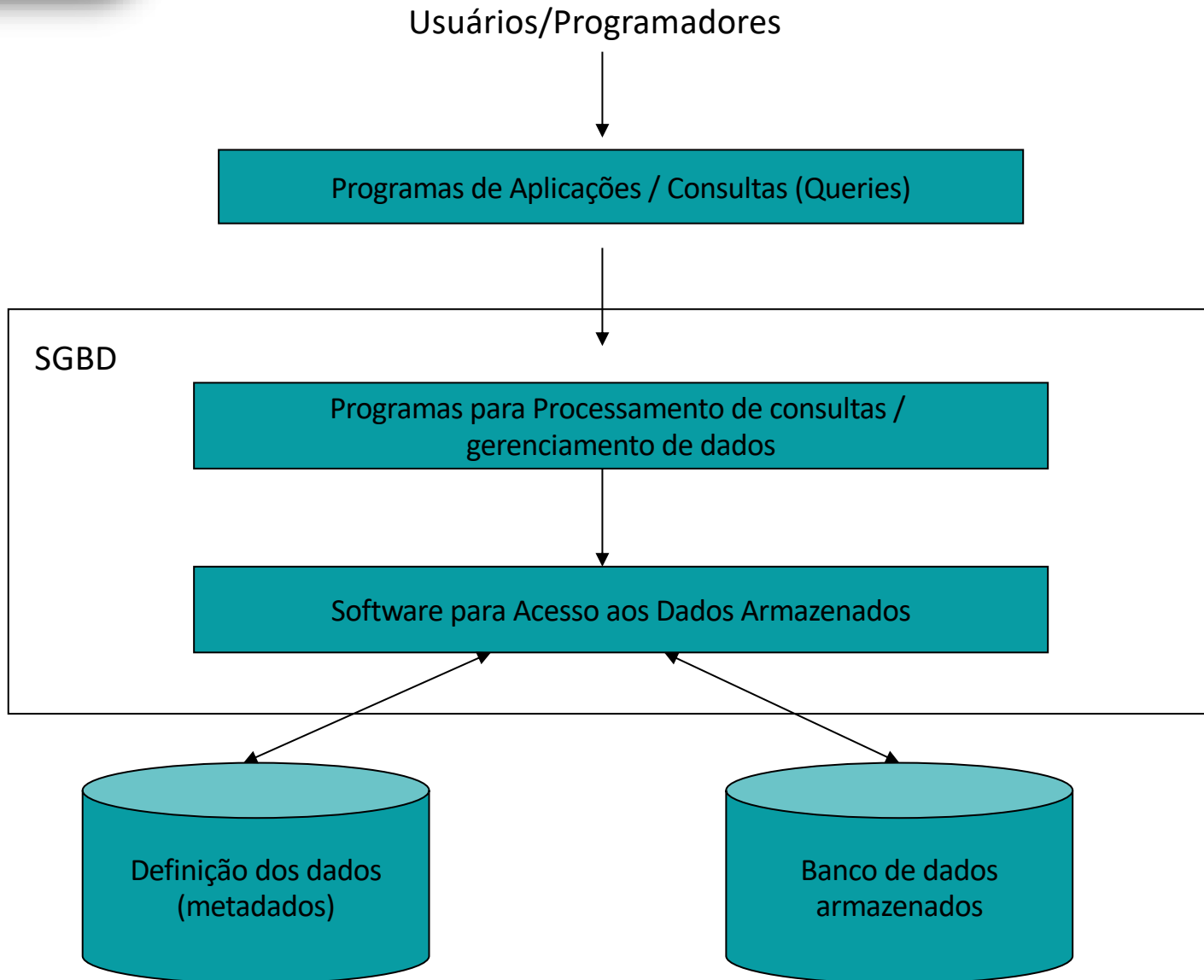
# Definições ...

- **Dados:** são fatos que podem ser gravados e que possuem um significado implícito.
- **Banco de Dados (BD):** é uma coleção de dados relacionados:
  - Representa aspectos do mundo real (minimundo ou universo de discurso) e mudanças no mundo real devem ser refletidas no BD.
  - É uma coleção lógica e coerente de dados com algum significado inerente. Uma organização randômica de dados não pode ser considerada um BD.
  - Um BD é construído em atendimento a uma proposta específica.

# Definições ...

- **Um Sistema Gerenciador de Banco de Dados (SGBD)** é uma coleção de programas que permite aos usuários criar e manter um banco de dados.
  - É um sistema de software de propósito geral que facilita os processos de definição, construção, manipulação e compartilhamento de bancos de dados entre vários usuários e aplicações.
- **Sistema de Banco de Dados (SBD):** o banco de dados mais ou software SGBD

# Sistema de Banco de Dados



# Exemplo 1

ALUNO	Nome	Número	Turma	Curso
	Smith	17	1	CC
	Brown	8	2	CC

CURSO	Nome_curso	Número_curso	Créditos	Departamento
	Introdução à Ciência da Computação	CC1310	4	CC
	Estrutura de Dados	CC3320	4	CC
	Matemática Discreta	MAT2410	3	MATH
	Banco de Dados	CC3380	3	CC



DISCIPLINA	Id_disciplina	Número_curso	Semestre	Ano	Instrutor
	85	MAT2410	Segundo	98	King
	92	CC1310	Segundo	98	Anderson
	102	CC3320	Primeiro	99	Knuth
	112	MAT2410	Segundo	99	Chang
	119	CC1310	Segundo	99	Anderson
	135	CC3380	Segundo	99	Stone

HISTÓRICO	Número_Aluno	Id_disciplina	Nota
	17	112	B
	17	119	C
	8	85	A
	8	92	A
	8	102	B
	8	135	A

PRÉ_REQUISITO	Número_Curso	Número_Pré_Requisito
	CC3380	CC3320
	CC3380	MAT2410
	CC3320	CC1310

# Características do uso de um SBD

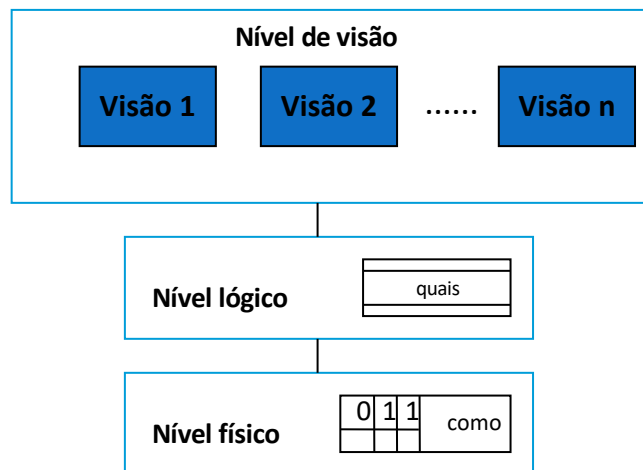
- Natureza autodescritiva do sistema de banco de dados;
  - Catálogo: metadados
- Isolamento entre programas e os dados, e a abstração de dados;
- Suporte para as múltiplas visões dos dados;
- Compatilhamento de dados e processamento de **transações** de multiusuários.
  - Concorrência

# Banco de Dados X Abstração de Dados

- Um dos maiores benefícios dos sistemas de banco de dados é proporcionar aos usuários uma visão abstrata dos dados. O sistema é capaz de ocultar alguns detalhes sobre a forma de armazenamento e a manutenção dos dados.
- A eficiência da recuperação de informações está relacionada à forma como as estruturas de representação são projetadas e, dado a complexidade e importância destas representações, elas devem ser protegidas em níveis de abstrações.
- Estes níveis facilitam a manutenção do sistema e a interação dos usuários com os sistemas. São eles:

# Níveis de abstração

- **Nível físico:** um registro de dado pode ser descrito como um bloco consecutivo de memória (por exemplo, palavras ou bytes).
- **Nível lógico:** um registro de dado é descrito por um tipo definido (como um tipo em linguagem de programação) e as inter-relações entre dados são definidas.
- **Nível de visão:** subconjunto de dados que podem existir apenas durante a execução de uma operação (uma consulta ao banco de dados, por exemplo).



# Exemplo de Visão

HISTÓRICO- ESCOLAR	Nome_Aluno	Histórico Escolar do Aluno				
		Número_curso	Nota	Semestre	Ano	Id_Disciplina
	Smith	CC1310	C	Primeiro	99	119
		MAT2410	B	Primeiro	99	112
	Brown	MAT2410	A	Primeiro	98	85
		CC1310	A	Primeiro	98	92
		CC3320	B	Segundo	99	102
		CC3380	A	Primeiro	99	135

# Instâncias e Esquemas

**Esquema:** Projeto geral do banco de dados. Os esquemas são alterados com pouca frequência.

**Cliente** (nome\_cliente: *string*; seguro\_social: *string*;  
rua\_cliente: *string*; cidade\_cliente: *string*).

**Instância:** O conjunto de informações contidas em um determinado banco de dados, em um dado momento.

Cliente\_1: *Cliente*;

(João da Silva, 5.929.555.99, R. Vicente Machado, São Paulo)

Cliente\_2: *Cliente*;

(Marcos Pereira, 8.223.938.51, R. Carvalho Bueno, São Paulo)

# Independência entre Programas e Dados

- É a capacidade de modificar a definição dos esquemas em determinado nível, sem afetar o esquema de nível superior.
  - ***Independência física de dados:*** é a capacidade de modificar o esquema físico sem que, com isso, qualquer programa de aplicação precise ser reescrito. Modificações no nível físico são necessárias, ocasionalmente, para aprimorar desempenho. (mais fácil de ser alcançada nos SBDs)

# Independência entre Programas e Dados

- ***Independência lógica de dados:*** é a capacidade de modificar o esquema lógico sem que, com isso, qualquer programa de aplicação precise ser reescrito. Modificações no nível lógico são necessárias sempre que uma estrutura lógica do banco de dados é alterada (por exemplo, mudança do sistema monetário).



# Transações

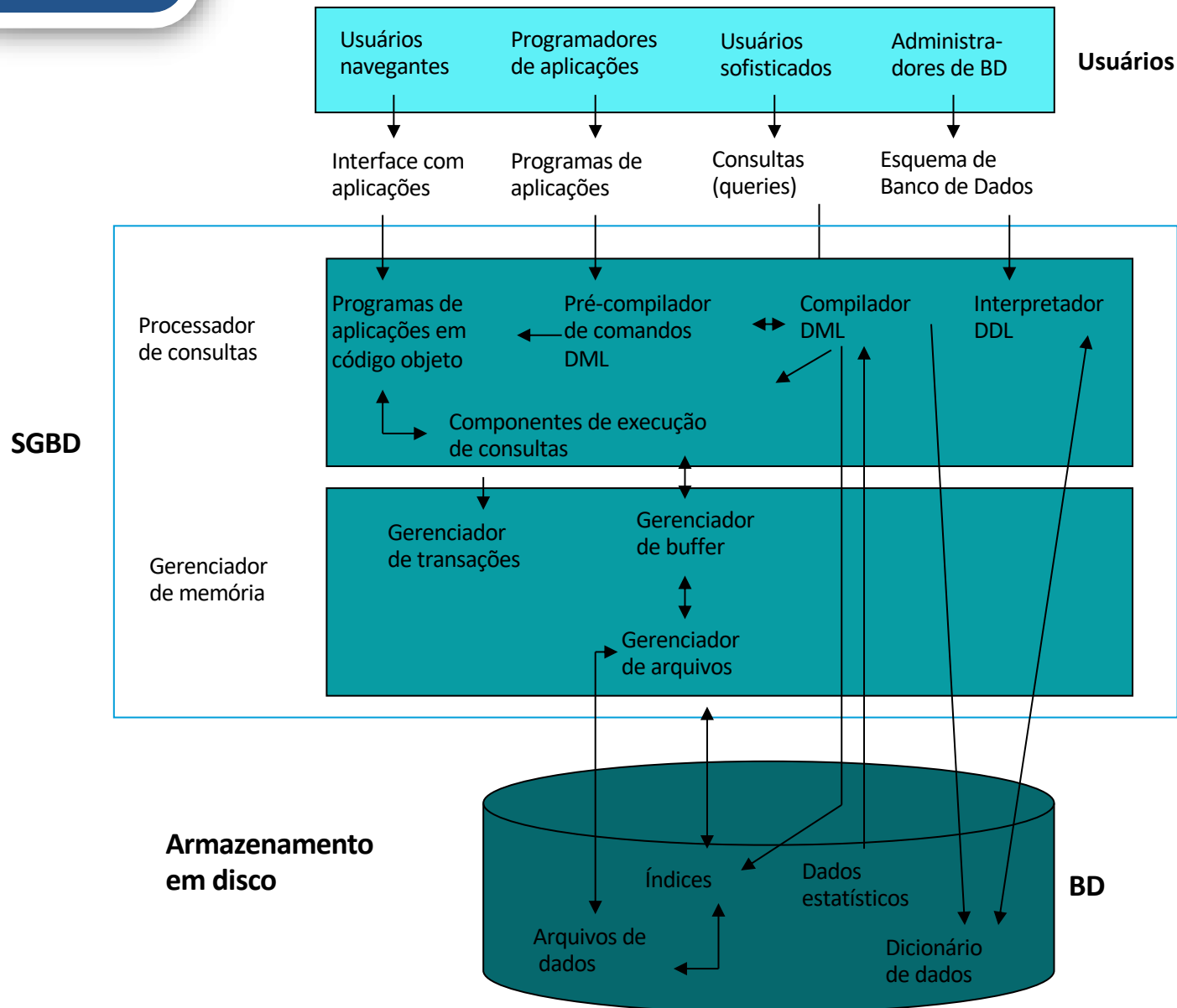
- **Definição:** coleção de operações que desempenha uma função lógica única dentro de uma aplicação do sistema de banco de dados.
- **Propriedades de uma transação (ACID):**
  - A** – atomicidade : ou todas as operações envolvidas na transação ocorrem, ou nenhuma delas deve ter efeito sobre o banco de dados. Tarefa do SGBD.
  - C** – consistência : ao final da execução da transação a consistência dos dados no banco devem ter sido mantidas. Tarefa do programador.

# Transações

**I – isolamento** : uma transação deve ter sua execução realizada de forma isolada em relação à execução de outras transações.

**D – durabilidade** : depois que uma transação é executada com sucesso, as modificações por ela realizadas devem ser mantidas no sistema, mesmo na ocorrência de falhas. Tarefa do SGBD.

# Visão Geral de um SGBD



# Linguagens de um SGBD

## 1. Linguagem de Definição de Dados

Um esquema de dados é especificado por uma conjunto de definições expressas por uma linguagem especial chamada *Linguagem de definição de dados* (do inglês *Data-Definition Language* – DDL).

O resultado da compilação dos parâmetros/comando DDL gera os **dicionário de dados** (arquivo de metadados).

# Linguagens de um SGBD

## 2. Linguagem de Manipulação de Dados

**Manipulação de dados** = recuperação de informações **do** banco de dados e inserção, remoção e alteração de dados **no** banco de dados.

A linguagem de manipulação de dados

(do inglês *data-manipulation language* – DML) é a linguagem que viabiliza o acesso e a manipulação do dados. Podem ser:

**Procedurais:** o usuário especifica procedimentos para recuperar os dados que necessita.

**Não-procedurais:** o usuário descreve os dados que necessita.

# Interfaces de um SGBD

- Interfaces Baseadas em Menus para os Clientes WEB ou Navegação;
- Interfaces Baseada em Formulários;
- Interfaces Gráficas para os Usuários;
- Interfaces em Linguagem Natural;
- Interface para Usuários Parametrizáveis

# Processamento de Consultas

- **Compilador DML:** traduz comandos DML em instruções de baixo nível, entendidos pelo componente de execução de consultas. Além disso, otimiza a solicitação do usuário.
- **Pre-compilador para comandos DML inseridos em programas de aplicação:** convertem comandos DML em chamadas de procedimentos normais da linguagem hospedeira. Interage com o compilador DML de modo a gerar o código apropriado.

# Processamento de Consultas

- **Interpretador DDL:** interpreta os comandos DDL e os registra no dicionário de dados.
- **Componentes para tratamento de consultas:** executa instruções de baixo nível gerada pelo compilador DML.



# Gerenciador de Memória

- Um dos principais objetivos de um Sistema de Banco de Dados é simplificar e otimizar o acesso aos dados.
- O desempenho de um SBD depende diretamente da eficiência das estruturas usadas na representação dos dados e do quanto este sistema está apto a operar essas estruturas de dados

# Gerenciador de Memória

- Um dos principais módulos de um SGBD é o *gerenciador de memória*, responsável por fazer a interface entre o armazenamento de dados em um nível mais baixo e as consultas e programas de aplicação submetidos ao sistema. Ele também realiza a interface do SBD com o Sistema de Arquivos do Sistema Operacional.
- É o *gerenciador de memória* quem traduz os diversos comandos DML em comandos de baixo nível de sistemas de arquivos.

# Gerenciador de Memória

- Composto por:
  - **Administração de buffer:** responsável pela intermediação de dados do disco para a memória principal e pela decisão de quais dados colocar em memória auxiliar.
  - **Administração de arquivos:** gerencia a alocação de espaço no armazenamento em disco e as estruturas de dados usadas para representar estas informações armazenadas em disco.

# Gerenciador de Memória

- Composto por:
  - **Gerenciamento de autorizações e integridade:** testam o cumprimento das regras de integridade e a permissão ao usuário no acesso ao dado.
  - **Gerenciamento de Transações:** cuida da execução das transações.

# Banco de Dados

- **Arquivo de dados:** armazena os dados (o banco de dados propriamente dito).
- **Dicionário de dados:** metadados.
- **Índices:** estrutura que otimizam o acesso aos itens de dados.
- **Estatística de dados:** armazena informações estatísticas relativas aos dados contidos no banco de dados. Essas informações são usadas pelo processador de consultas para seleção de meios eficientes para execução de consultas.

# Utilitários de um SGBD

- **Carregamento (*loading*)** : carrega arquivos e dados existentes dentro do banco de dados. Útil para transferência de dados entre SGBDs ou entre SGBDs e outros sistemas (são ferramentas de conversão).
- ***Backup***: cria uma cópia do banco de dados, geralmente descarregando (*dumping*) todo o banco de dados em uma fita (por exemplo). Também possibilita o *backup* incremental.

# Utilitários de um SGBD

- **Reorganização de arquivos:** reorganiza os arquivos do banco de dados em uma nova forma buscando melhorar seu desempenho.
- **Monitoramento de desempenho:** monitora o uso do BD e fornece estatísticas para o DBA, que pode tomar decisões para melhorar o desempenho.

# Laboratório 01

- Conectando a um SGBD Relacional para configuração e implantação de um banco de dados relacional.

