



UAlg FCT

UNIVERSIDADE DO ALGARVE
FACULDADE DE CIÊNCIAS E TECNOLOGIA

Bases de Dados

INTRODUÇÃO

Agenda

- Exemplos de aplicações de Bases de Dados
- Objetivos dos Sistemas de Gestão de Bases de Dados
- Modelos de Dados
- Linguagens
 - *Data definition language (DDL)*
 - *Data manipulation language (DML)*
- *Estrutura de um sistema de Gestão de Bases de Dados*

Exemplos de aplicações de Bases de Dados

Informação Organizacional

- Vendas: produtos, clientes, ...
- Contabilidade: pagamentos, recebimentos, ativos, ...
- Recursos humanos: empregados, salários, ...

Banca

- Informação de clientes, contas bancárias, cartões, transações, ...

Universidades

- Alunos, professores, matrículas, avaliações, ...

Transportes

- Bilhetes, horários, ...

Telecomunicações

- Chamadas, mensagens, tarifários, ...

Hospitais / Clínicas

- Trabalho prático ;)

Objetivos dos Sistemas de Gestão de Bases de Dados

Antigamente, as aplicações interagiam diretamente com os sistemas de gestão de ficheiros, levando a problemas como:

Redundância e inconsistência de dados: dados armazenados em vários formatos de ficheiro, resultando, normalmente na duplicação de informação em ficheiros diferentes

Dificuldade em aceder aos dados

- Necessário escrever código específico para cada nova tarefa

Isolamento de dados

- Vários ficheiros e vários formatos

Problemas de integridade

- Restrições de integridade (por exemplo, $\text{nota} \geq 0$ & $\text{nota} \leq 20$) ficam “enterradas” no código da aplicação em vez de serem declaradas explicitamente
- Difícil adicionar novas restrições ou alterar as existentes

Objetivos dos Sistemas de Gestão de Bases de Dados (cont.)

Atomicidade das atualizações

- Podem deixar a BD num estado inconsistente
- Exemplo: transferência entre contas bancárias

Acesso simultâneo (concorrência) por vários utilizadores

- Concorrência necessária por questões de desempenho
- Concorrência descontrolada pode levar a inconsistências (exemplo: dois levantamentos de 100€ numa conta com saldo de 120€)

Problemas de segurança

- Limitar o acesso a diferentes perfis de utilizadores

Os sistemas de gestão bases de dados oferecem soluções para todos os problemas mencionados anteriormente

Modelos de dados

Conjunto de “ferramentas” para descrever:

- Dados
- Relacionamento de dados
- Semântica de dados
- Restrições de dados

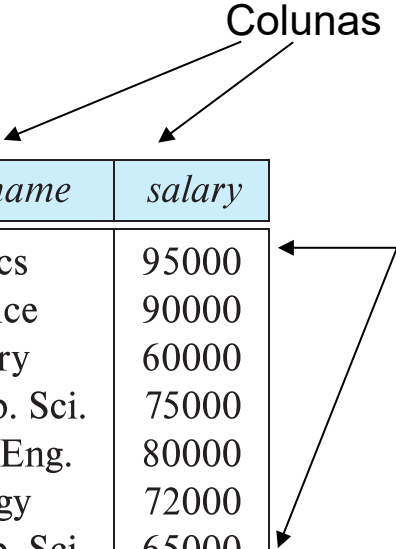
Exemplos de modelos:

- Modelo Relacional
- Modelo de Entidade Associação (ER do inglês Entity-Relationship)
- Modelos de dados baseados em objetos
- Modelos de dados semiestruturados (XML)

Modelo Relacional

Dados armazenados em várias tabelas

Exemplo de dados tabulares no modelo relacional



<i>ID</i>	<i>name</i>	<i>dept_name</i>	<i>salary</i>
22222	Einstein	Physics	95000
12121	Wu	Finance	90000
32343	El Said	History	60000
45565	Katz	Comp. Sci.	75000
98345	Kim	Elec. Eng.	80000
76766	Crick	Biology	72000
10101	Srinivasan	Comp. Sci.	65000
58583	Califieri	History	62000
83821	Brandt	Comp. Sci.	92000
15151	Mozart	Music	40000
33456	Gold	Physics	87000
76543	Singh	Finance	80000

Instâncias e esquemas

Semelhante a *tipos* e *variáveis* em linguagens de programação

Esquema lógico – a estrutura lógica geral da base de dados

- Exemplo: A base de dados consiste num conjunto de clientes e contas e no relacionamento entre eles
- Análogo ao tipo de informação de uma variável num programa

Esquema físico – a estrutura física geral da base de dados

Exemplo – o conteúdo real da base de dados num determinado momento

- Análogo ao valor de uma variável

Independência do nível físico

Independência do nível físico – capacidade de modificar o esquema físico sem alterar o esquema lógico

- As aplicações dependem do esquema lógico
- Em geral, as interfaces entre os vários níveis e componentes devem ser bem definidas para que mudanças em algumas partes não influenciem seriamente as outras.

Linguagem de Definição de Dados (Data Definition Language - DDL)

Notação para definir o esquema de uma base de dados

Exemplo:

```
create table instructor (  
    ID          char(5),  
    name        varchar(20),  
    dept_name   varchar(20),  
    salary      numeric(8,2))
```

O compilador DDL gera dados acerca da tabela e armazena-os num dicionário de dados (***data dictionary***)

Dicionário de dados contém metadados (i.e., dados acerca de dados)

- Esquema da base de dados
- Restrições de Integridade
 - Exemplo: chave primária
- Autorização
 - Quem pode aceder ao quê

Linguagem de Manipulação de Dados (Data Manipulation Language - DML)

Linguagem para aceder e atualizar os dados organizados pelo modelo de dados

- DML também conhecido como *query language*

Existem dois tipos de linguagem de manipulação de dados

- **DML procedimental** — exige que o utilizador especifique os dados que são necessários **e como obtê-los**.
- **DML declarativo** — exige que o utilizador especifique os dados que são necessários **sem especificar como obtê-los**.

DMLs declarativas são, normalmente, mais fáceis de aprender e usar do que DMLs procedimentais.

DMLs declarativas também são chamadas de DMLs não procedimentais

A parte de uma DML que envolve obtenção de informação é chamada de *query language*.

SQL Query Language

A linguagem SQL é declarativa (não-procedimental). Uma consulta recebe como entrada várias tabelas (possivelmente apenas uma) e retorna uma única tabela.

Exemplo para encontrar todos os professores no departamento de informática

```
select name  
from instructor  
where dept_name = 'informática'
```

Iremos estudar SQL com detalhe no âmbito da Unidade Curricular de Bases de Dados

Acesso à base de dados a partir de uma aplicação

Linguagens como SQL não permitem desenvolver uma aplicação completa. Por exemplo, SQL não permite suportar ações como receber o input de utilizadores e apresentar os resultados numa interface customizada como por exemplo num ambiente web.

Esse tipo de funcionamento deve ser desenvolvido em linguagens de programação genéricas, como C/C++/C#, Java ou Python, ou plataformas *low-code* (Outsystems, Salesforce, ServiceNow) que incorporam queries SQL para interagir com a Base de Dados.

Desenho de bases de dados

Processo de desenho da estrutura da base de dados:

Desenho lógico – Decisões sobre o esquema da base de dados. Requer que se encontre o conjunto “adequado” de relações (tabelas).

- Decisão de negócio – Que atributos devem ser guardados na base de dados?
- Decisão de bases de dados – Que relações (tabelas) devemos ter e como é que os atributos devem ser “distribuídos” nessas relações (tabelas)?

Desenho físico – Decisões sobre a organização física da base de dados

Sistema de Gestão de Bases de Dados

Um sistema de gestão de bases de dados (SGBD) é decomposto em módulos que lidam com diferentes responsabilidades. Os componentes podem ser divididos em:

- Gestor de armazenamento
- Processador de queries
- Gestor de transações

Gestor de armazenamento

Módulo que fornece a interface entre o armazenamento físico dos dados da base de dados e as aplicações e queries submetidas para o SGBD. É responsável por:

- Interação com a gestão de ficheiros do Sistema Operativo
- Armazenamento, acesso e atualização eficiente de dados

Inclui:

- Gestor de autorização e de integridade
- Gestor de transações
- Gestor de ficheiros
- Gestor de *buffer*

Gestor de armazenamento (cont.)

O gestor de armazenamento implementa diversas estruturas de dados como parte da implementação do sistema físico:

- Ficheiros de dados — armazenam a própria base de dados;
- Dicionário de dados — armazena metadados sobre a estrutura da base de dados, em particular o seu esquema;
- Índices - fornecem acesso rápido a dados. Um índice cria ponteiros para dados que contêm um valor específico.

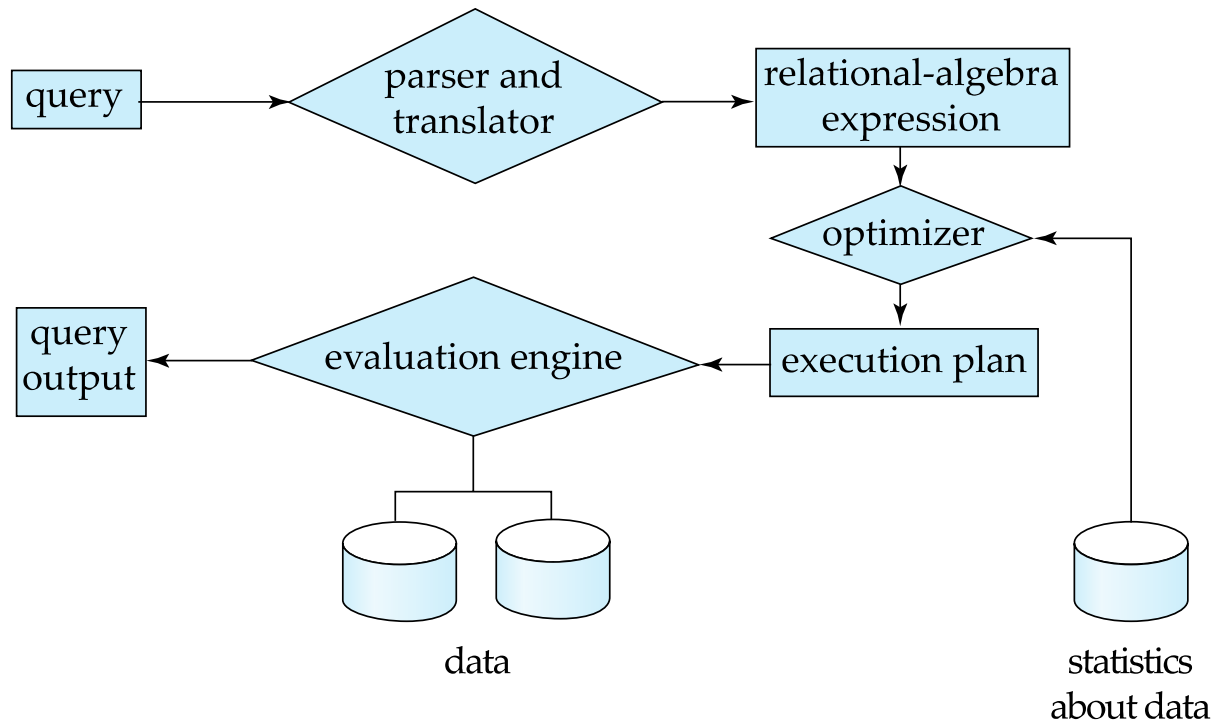
Processador de queries

Os componentes do processador de queries incluem:

- Interpretador de DDL — interpreta instruções DDL e regista as definições no dicionário de dados.
- Compilador DML — traduz instruções DML num plano de avaliação (instruções de baixo nível que o mecanismo de execução de queries processa).
 - O compilador DML realiza a otimização da query; ou seja, escolhe o plano de execução de menor custo entre as várias alternativas.
- Mecanismo de avaliação (execução) da query — executa instruções de baixo nível geradas pelo compilador DML.

Processador de queries

1. Análise e tradução
2. Otimização
3. Avaliação (execução)



Gestão de transações

Uma **transação** é um conjunto de operações que executa uma única função lógica numa base de dados

O componente de gestão de transações garante que a base de dados permanece num estado consistente (correto), apesar de falhas do sistema (por exemplo, falhas de energia e bloqueios no sistema operativo) e falhas de transação.

O gestor do controlo de concorrência controla a interação entre as transações simultâneas (concorrentes) para garantir a consistência da base de dados.

Breve História dos Sistemas de Gestão de Bases de Dados

Década de 1950 e início de 1960:

- Processamento de dados usando fitas magnéticas para armazenamento
 - As fitas forneciam apenas acesso sequencial
- Cartões perfurados para entrada de dados

Final da década de 1960 e 1970:

- Os discos rígidos permitiam acesso direto aos dados
- Modelos de dados de rede e hierárquicos em uso generalizado
- Ted Codd define o modelo de dados relacional
 - Oracle lança primeiro sistema de gestão de bases de dados relacional comercial
- Processamento de transações de alto desempenho (para a época)

História dos Sistemas de Banco de Dados (Cont.)

Década de 1980:

- SQL torna-se o padrão industrial
- Sistemas de bases de dados paralelos e distribuídos
 - IBM, Teradata
- Sistemas de bases de dados orientados a objetos

Década de 1990:

- Grandes aplicações de suporte à decisão e *data mining*
- Grandes armazéns de dados (*datawarehouse*)

Anos 2000

- Sistemas de armazenamento de big data
 - Google BigTable , Yahoo Pnuts
 - Sistemas “ NoSQL ”
- Análise de big data: além do SQL
 - Map-Reduce

Década de 2010

- *SQL reloaded*
 - Front-end SQL para sistemas Map-Reduce
 - Sistemas de bases de dados massivamente paralelos
 - Bases de dados *in-memory multi-core*

Leitura recomendada

Capítulo 1 do livro:

- Database system concepts. Abraham Silberschatz, Henry Korth and S. Sudarshan, McGraw-Hill, 7th edition, 2020 (ISBN: 9780078022159)