

Bases de Dados



2012/13

<http://ssdi.di.fct.unl.pt/bd>

João Leite (jleite@fct.unl.pt)

Funcionamento

■ Aulas:

- ✦ Aulas Teóricas (2h, até fim de Abril)
- ✦ Aulas Práticas (começam no dia 6 de Março)
 - ❖ Exercícios
 - ❖ Trabalho Prático

■ Avaliação:

- ✦ 50% - 2 Testes ou Exame de Recurso (mínimo 10 valores)
- ✦ 50% - Trabalho prático (várias etapas com prazos bem definidos)

■ Página da cadeira:

<http://ssdi.di.fct.unl.pt/bd>

■ Mailing list:

<https://lists.fct.unl.pt/mailman/listinfo/di-bd>

Algumas Regras (o que espero de vós)

- Pontualidade.
 - ✦ No caso excepcional de chegarem atrasados, devem entrar pela porta de trás, tentando não perturbar a aula.
- Não usar portáteis durante as aulas teóricas.
 - ✦ Se estiverem habituados a usar os portáteis para tirar notas, então devem vir para a fila da frente.
- Manter silêncio durante as aulas.
 - ✦ A não ser, claro está, para participar na aula.
- Fazer um esforço para compreender a matéria antes de colocar dúvidas, por e-mail ou no horário de dúvidas.

O que podem esperar de mim...

- Assiduidade e pontualidade.
 - ✦ A menos de algum imprevisto...
- Disponibilidade para esclarecimento de dúvidas, de forma presencial ou por e-mail.
- Colocação dos slides pouco tempo após cada aula teórica.
- Respeito pelos prazos estabelecidos.
- Preparar e dar as aulas da melhor forma que sei...

Capítulo 1: Introdução

- Função dos Sistemas de Bases de Dados
- Visão dos dados
- Modelos de dados
- Linguagem de Definição de Dados
- Linguagem de Manipulação de Dados
- Gestão de Transacções
- Gestão de Armazenamento
- Administrador da Base de Dados
- Utilizadores da Base de Dados
- Estrutura Global do Sistema

Sistemas de Gestão de Bases de Dados (SGBDs)

- Colecção de dados inter-relacionados (Base de Dados)
- Conjunto de programas para aceder aos dados
- Deve fornecer um ambiente de utilização *conveniente* e *eficiente*.
- Exemplos de aplicações de Bases de Dados:
 - Banca: todas as transacções e movimentos
 - Companhias aéreas: reservas, horários
 - Universidades: inscrições, notas
 - Vendas: clientes, produtos, compras
 - Indústria: produção, inventário, pedidos, cadeia de fornecimento
 - Recursos humanos: registos dos empregados, salários, impostos
 - ...

Funções de um SGBD

- Antigamente, os SGBDs eram construídos directamente sobre os sistemas de ficheiros.
- Inconvenientes dos sistemas de ficheiros no armazenamento de dados:
 - Múltiplos ficheiros e formatos
 - Isolamento de dados
 - Dificuldade de relacionar dados
 - Duplicação de informação em ficheiros diferentes
 - Redundância e inconsistência de dados
 - Dificuldades no acesso aos dados
 - Necessidade de escrever um novo programa para efectuar uma nova tarefa
 - Problemas de integridade
 - Restrições de integridade (e.g. saldo da conta > 0) estão incluídas no código dos programas
 - Difícil alterar ou adicionar novas restrições

Funções de um SGBD (Cont.)

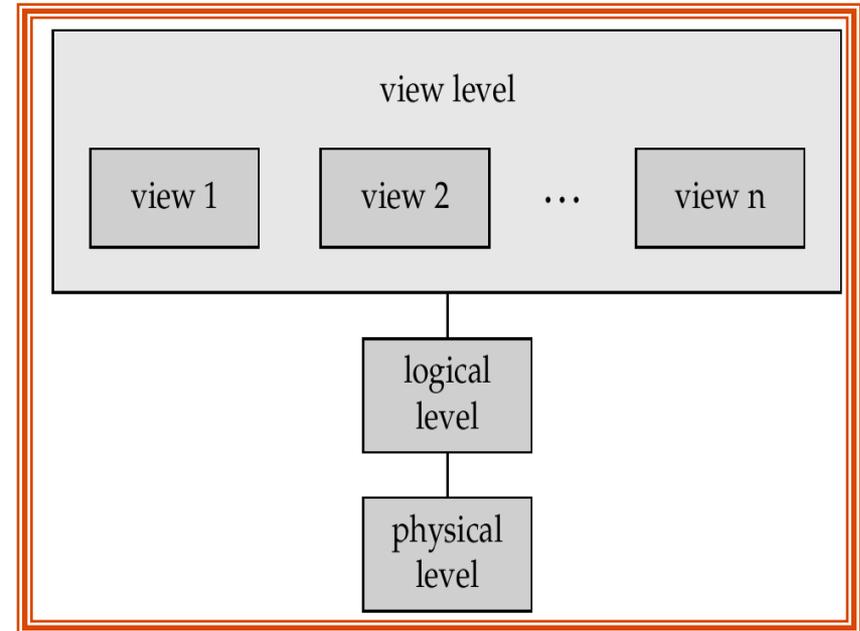
- Inconvenientes dos sistemas de ficheiros (cont.)
 - Atomicidade das alterações
 - Falhas podem colocar a base de dados num estado inconsistente com alterações parciais já efectuadas.
 - E.g. transferência de dinheiro de uma conta para outra ou deve ser totalmente realizada ou nenhuma alteração deve ser efectuada
 - Acessos concorrentes por diversos utilizadores
 - Acessos concorrentes são necessários por motivos de eficiência
 - Os acessos concorrentes não controlados podem originar inconsistências
 - E.g. duas pessoas a consultarem um saldo e a alterá-lo ao mesmo tempo
 - Problemas de segurança
- Os Sistemas de Gestão de Bases de Dados pretendem oferecer soluções para todos estes problemas.

Recapitulando...

- Um SGBD consiste:
 - ✦ Colecção de dados inter-relacionados (Base de Dados)
 - ✦ Conjunto de programas para aceder aos dados
 - ✦ Deve fornecer um ambiente de utilização *conveniente* e *eficiente*.
- Por um lado, a eficiência é importante (acesso rápido aos dados).
- Por outro lado, queremos esconder certos detalhes de implementação...

Níveis de Abstracção

- **Nível físico:** descreve como um registo (e.g., cliente) é armazenado.
- **Nível lógico:** descreve os dados armazenados na base de dados, assim como as relações entre os dados.
- **Nível das vistas:** as aplicações ocultam os detalhes dos tipos de dados. As vistas também podem esconder informação (e.g., salário) por motivos de segurança.



- Semelhante a tipos de dados nas linguagens de programação
 - **Nível físico:** blocos de dados no disco.
 - **Nível lógico:** tipos e relações entre tipos, usados ao nível da programação.
 - **Nível das vistas:** aplicações que escondem os tipos do utilizador.

Instâncias e Esquemas

- As Bases de Dados Variam ao longo do tempo...
- **Instância** – o conteúdo de uma base de dados num instante de tempo
 - Análogo ao valor de uma variável
- **Esquema**– a estrutura lógica da base de dados
 - e.g., a base de dados é constituída por informação sobre clientes, contas e as relações entre si.
 - Análogo ao tipo de uma variável num programa
 - **Esquema físico**: desenho da base de dados ao nível físico
 - **Esquema lógico**: desenho da base de dados ao nível lógico
- **Independência física dos dados** – a 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 estar bem definidas de modo a que alterações numa parte não influenciem grandemente outras partes.

Modelos de Dados

- Um conjunto de ferramentas para descrever
 - dados
 - relações entre dados
 - semântica dos dados
 - restrições sobre os dados
- Exemplos:
 - ✦ Modelo Entidade-Relações (para desenho de Bases de Dados)
 - ✦ Modelo Relacional
 - ✦ Modelos de dados baseados em objectos
 - ✦ Modelos de dados semi-estruturados (XML)
 - ✦ Outros modelos:
 - Modelo hierárquico
 - Modelo em rede

Linguagem de Definição de Dados (DDL)

- Especificação da notação para definição do esquema da base de dados
 - E.g.

```
create table account (  
    account-number char(10),  
    balance integer)
```
- O compilador da DDL gera um conjunto de tabelas armazenadas num *dicionário de dados*.
- O dicionário de dados contém metadados (dados sobre dados)
 - Esquema de bases de dados
 - Restrições de integridade
 - Restrições de domínio; Integridade de referência; Asserções
 - Autorização
- Linguagem de *definição* de dados e *armazenamento*
 - Linguagem onde se especificam as estruturas de armazenamento e métodos de acesso utilizados pela base de dados
 - Normalmente uma extensão da linguagem da definição de dados

Ling. de Manipulação de Dados (DML)

- Linguagem para aceder e manipular os dados organizados de acordo com o respectivo modelo de dados
 - A DML também é conhecida por linguagem de consulta
- Duas classes de linguagens
 - **Procedimental** – o utilizador especifica quais os dados que se pretendem, bem como a forma de os obter
 - **Declarativa (Não-procedimental)** – o utilizador especifica quais os dados pretendidos sem especificar a forma de os obter
- A SQL é a linguagem de consulta mais utilizada

Bases de Dados Relacionais

- Uma base de dados relacional é baseada no modelo de dados relacional
- Os dados e as relações entre os dados são representadas por um conjunto de relações (semelhante a tabelas)
- Inclui uma DML e uma DDL
- A maioria dos sistemas de bases de dados relacionais comerciais usam a linguagem SQL (que inclui DML e DDL).

Modelo Relacional

- Exemplo de conjunto de dados no modelo relacional

Atributos



<i>Customer-id</i>	<i>customer-name</i>	<i>customer-street</i>	<i>customer-city</i>	<i>account-number</i>
192-83-7465	Johnson	Alma	Palo Alto	A-101
019-28-3746	Smith	North	Rye	A-215
192-83-7465	Johnson	Alma	Palo Alto	A-201
321-12-3123	Jones	Main	Harrison	A-217
019-28-3746	Smith	North	Rye	A-201

Exemplo de Base de Dados Relacional

<i>customer-id</i>	<i>customer-name</i>	<i>customer-street</i>	<i>customer-city</i>
192-83-7465	Johnson	12 Alma St.	Palo Alto
019-28-3746	Smith	4 North St.	Rye
677-89-9011	Hayes	3 Main St.	Harrison
182-73-6091	Turner	123 Putnam Ave.	Stamford
321-12-3123	Jones	100 Main St.	Harrison
336-66-9999	Lindsay	175 Park Ave.	Pittsfield
019-28-3746	Smith	72 North St.	Rye

(a) The *customer* table

<i>account-number</i>	<i>balance</i>
A-101	500
A-215	700
A-102	400
A-305	350
A-201	900
A-217	750
A-222	700

(b) The *account* table

<i>customer-id</i>	<i>account-number</i>
192-83-7465	A-101
192-83-7465	A-201
019-28-3746	A-215
677-89-9011	A-102
182-73-6091	A-305
321-12-3123	A-217
336-66-9999	A-222
019-28-3746	A-201

(c) The *depositor* table

SQL

- SQL: linguagem não-procedimental de uso generalizado
 - E.g. encontrar o nome de cliente com customer-id 192-83-7465

```
select customer.customer-name
from customer
where customer.customer-id = '192-83-7465'
```
 - E.g. procurar os saldos de todas as contas detidas pelo cliente com customer-id 192-83-7465

```
select account.balance
from depositor, account
where depositor.customer-id = '192-83-7465' and
depositor.account-number = account.account-number
```
- As aplicações geralmente acedem a bases de dados por intermédio de:
 - Extensões às linguagens permitindo SQL embutido:
 - Interface de aplicações (e.g. ODBC/JDBC) permitindo o envio de consultas SQL para a base de dados

Desenho de Bases de Dados

O processo de desenhar a estrutura geral de uma base de dados.

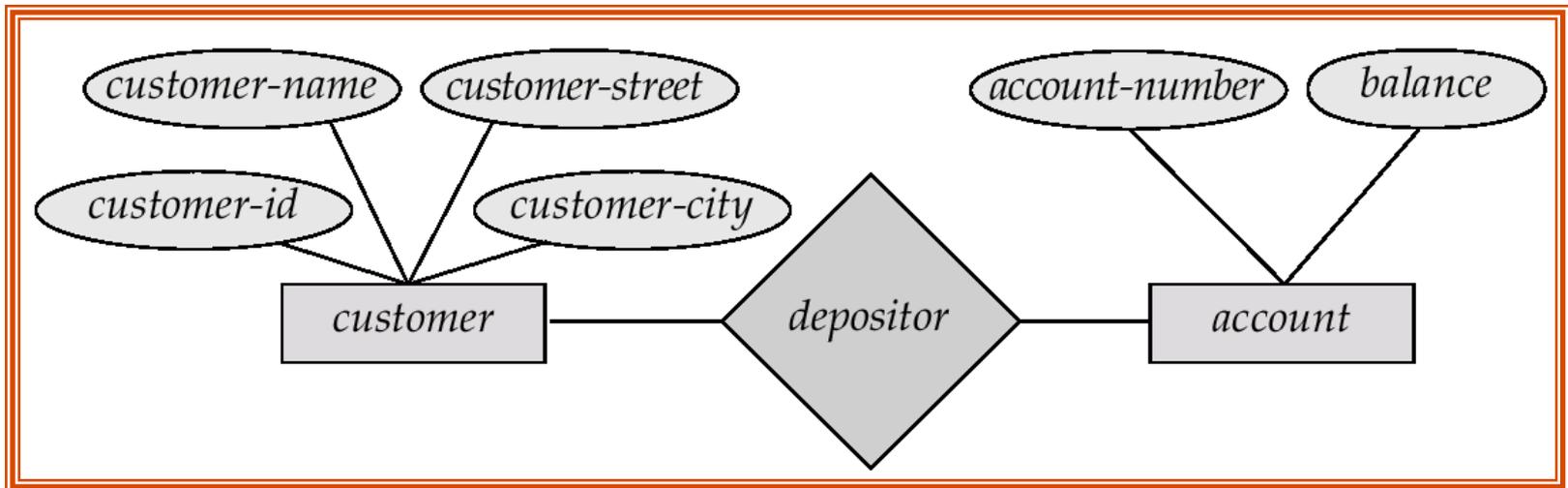
- **Desenho Lógico:** decidir o esquema da Base de Dados.
Encontrar uma boa colecção de Esquemas de Relação.
 - **Decisão empresarial:** que atributos devem ser mantidos na Base de Dados?
 - **Decisão Informática:** Que esquemas de relação devem ser usados e como dividir os vários atributos pelos vários esquemas de relação?
- **Desenho Físico:** Decidir a forma (layout) físico da Base de Dados

Modelo Entidades Relações

- Modelização ER do mundo
 - Entidades (objectos)
 - E.g. customers, accounts, bank branch
 - Relações entre entidades
 - E.g., a conta A-101 pertence ao cliente Johnson
 - O conjunto de associações *depositante* associa clientes a contas
- Amplamente utilizado no desenho de bases de dados
 - O desenho de base de dados no modelo ER é normalmente convertido para o desenho no modelo relacional (a seguir) que é utilizado para processamento e armazenamento

Modelo ER (cont.)

Exemplo de um esquema no modelo ER



Outros Modelos de Dados

- **Objecto-relacional**: estende o modelo relacional com a noção de objectos
- **XML**: permite a modelização de dados semi-estruturados

Gestão do Armazenamento

- O gestor do armazenamento é um módulo de programa que fornece uma interface entre os dados de baixo nível armazenados na base de dados e as aplicações e consultas submetidas ao sistema.
- O gestor de armazenamento é responsável pelas seguintes tarefas:
 - interacção com o gestor de ficheiros
 - armazenamento, recuperação e alteração eficientes dos dados

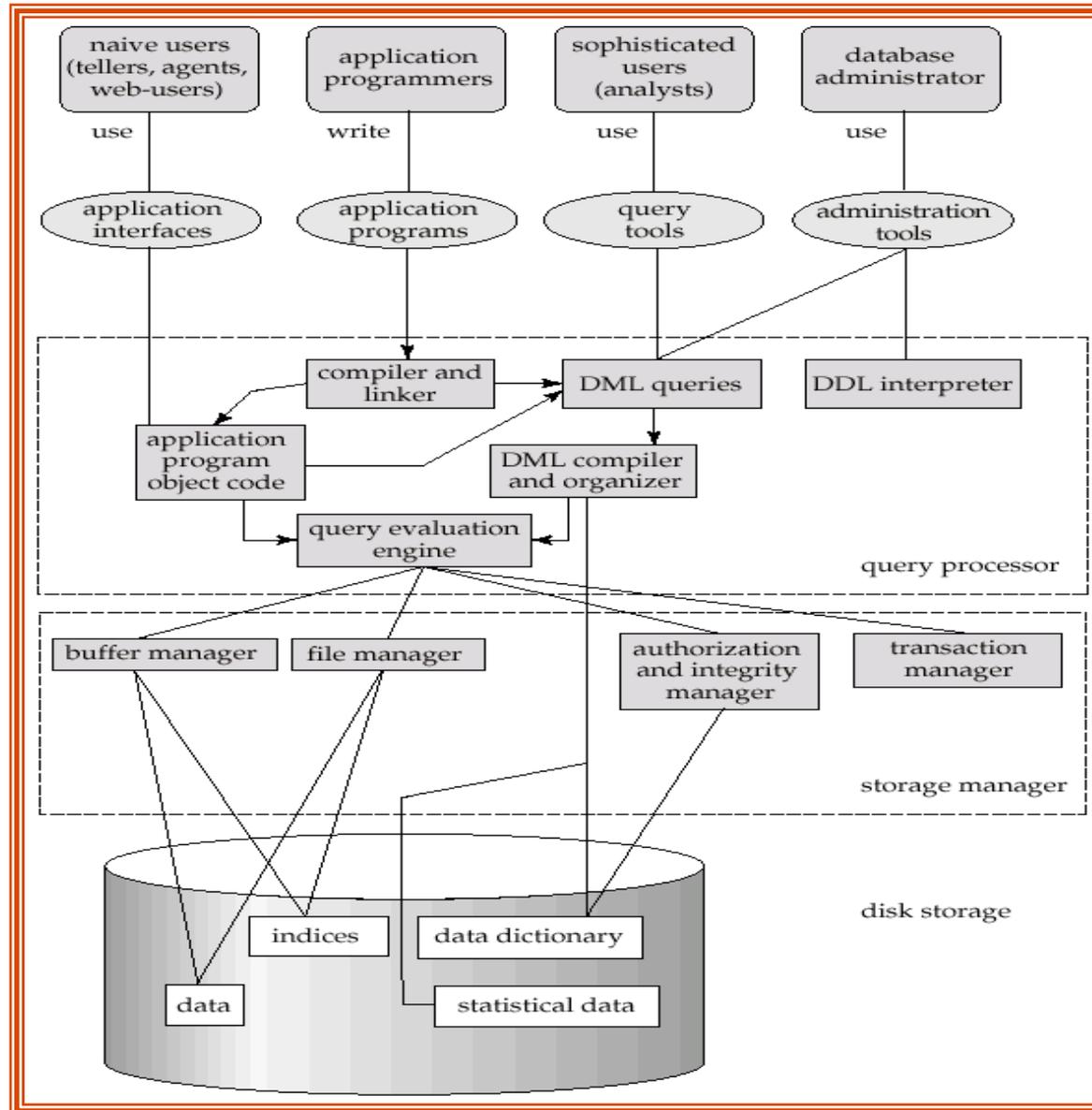
Processamento de consultas

- Módulos:
 - Interpretação e tradução
 - Optimização
 - Avaliação
- Formas alternativas de avaliação de uma dada consulta
 - Expressões equivalentes
 - A diferença de custo entre uma boa e uma má forma de avaliação de uma consulta pode ser enorme
- É necessário estimar o custo das operações
 - Depende de dados estatísticos mantidos pela Base de Dados sobre as relações que armazena
 - Necessita estimar estatísticas para resultados intermédios para calcular o custo de uma expressão complexa

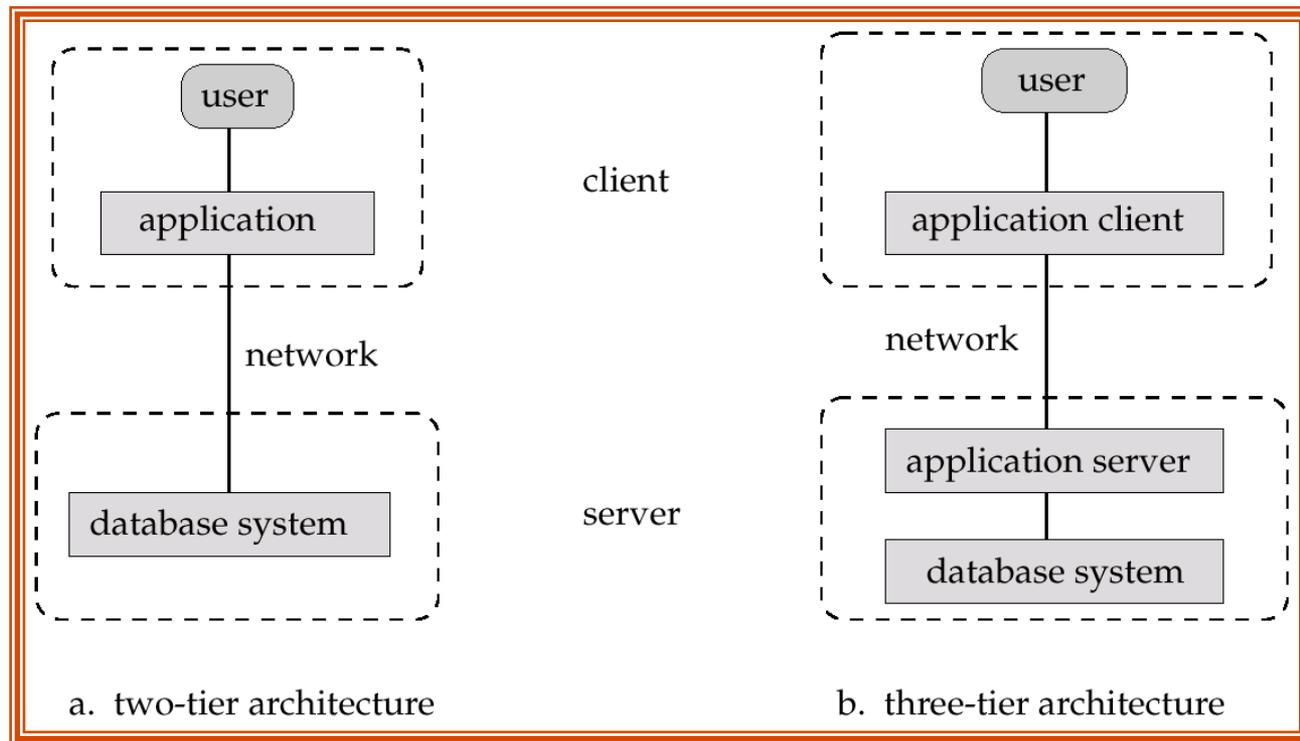
Gestão de Transacções

- Uma *transacção* é um conjunto de operações que efectuam uma função lógica na aplicação de base de dados
- A componente de gestão de transacções garante que a base de dados se mantém num estado consistente (correcto) apesar de falhas no sistema (e.g., falta de energia eléctrica e paragens abruptas do sistema operativo) e de transacções falhadas.
- O gestor de controlo de concorrência coordena a interacção entre transacções concorrentes para garantir a consistência da base de dados.

Estrutura Genérica do Sistema



Arquiteturas de Aplicação



- **Arquitetura de duas camadas:** E.g. programas clientes recorrendo a ODBC/JDBC para comunicar com a base de dados
- **Arquitetura de três camadas:** E.g. aplicações web e aplicações construídas recorrendo a "software intermediário", e.g. Forms)

Utilizadores da Base de Dados

Os utilizadores diferenciam-se pela forma esperada de interacção com o sistema

- **Programadores de aplicações** – interagem com o sistema através de chamadas DML
- **Utilizadores sofisticados** – constróem pedidos numa linguagem de consulta a bases de dados
- **Utilizadores especializados** – escrevem aplicações de bases de dados especializadas que não se enquadram no espírito do processamento de dados tradicional
- **Utilizadores** – chamam uma das aplicações que foi construída previamente
 - E.g. pessoas a acederem a uma base de dados através da Web, caixas, pessoal de secretariado

Administrador da Base de Dados

- Coordena todas as actividades do sistema de base de dados; o administrador da base de dados compreende bem dos recursos e necessidades de informação da empresa.
- As funções do administrador de bases de dados incluem:
 - Definição do esquema
 - Definição dos métodos de acesso e estrutura de armazenamento
 - Modificação do esquema e da organização física
 - Dar aos utilizadores autorizações de acesso à base de dados
 - Especificar restrições de integridade
 - Servir de ligação entre os utilizadores
 - Monitorar a performance e responder a alterações nos requisitos.

Funcionamento da UC de Bases de Dados



Objectivo: Dotar os alunos das bases necessárias à concepção, construção e análise de bases de dados relacionais.

Objectivos – Saber

- Conhecer o modelo Entidade-Relação e o modelo relacional de dados, e os fundamentos do modelo objecto/relacional
- Álgebra relacional como linguagem formal de manipulação de bases de dados relacionais
- Os fundamentos da linguagem SQL (como linguagem de definição e manipulação de dados)
- Normalização de bases de dados com base em dependências funcionais e multivalor
- Breves noções de modelação de dados em XML, e de mecanismos para manipulação desses dados

Objectivos – Fazer

- Modelar, primeiro em ER e depois numa base de dados relacional, um problema de dimensão média (resultando em cerca de uma/duas dezenas de tabelas)
- Criar uma base de dados em SQL para problema médio implementando em SQL todos os mecanismos necessário à integridade dos dados
- Implementar uma interface simples para manipulação da base de dados
- Estar à vontade na formulação de perguntas complexas em SQL
- Saber usar os mecanismos base SQL do modelo objecto/relacional

Objetivos – Soft Skills

- Capacidade de trabalhar em equipa e de colaborar numa equipa
- Capacidade de modelação de problemas
- Saber elaborar relatório de análise, desenho e implementação de uma solução
- Gestão do tempo e cumprimento de prazos

Programa

- Introdução aos sistemas de bases de dados (1 aula – esta)
- Modelos de dados
 - * Modelo de Entidades e Relações (2 aulas)
 - * Modelo Relacional (1 aula)
- Linguagens de manipulação de bases de dados
 - * Álgebra relacional (2 aulas)
 - * SQL (3 aulas)
 - * Outras linguagens (1 aula)
- Integridade e Segurança de Bases de Dados
 - * Integridade de referência (1/2 aula)
 - * Asserções e triggers (1/2 aula)
 - * Segurança e autorizações (1/2 aula)
 - * Transacções (1/2 aula)
- Normalização de Bases de Dados
 - * Dependências funcionais e multi-valor (1 1/2 aula)
 - * Formas normais: 3^a, 4^a e de Boyce-Cood (1 1/2 aula)
- Discussão de outros modelos de bases de dados
 - * Bases de dados objectos/relacional (1 aula)
 - * XML (2 aulas)

Aulas Práticas

- Exercícios de modelo ER (1 aula)
- Exercícios de álgebra relacional (1 aula)
- Exercícios de SQL (3 aulas)
- Exercícios de Normalização de BDs (1 aula)
- Exercícios de XML (1/2 aula)
- Exercícios de Transacções (1/2 aula)
- Introdução ao Oracle® Application Express (2 aulas)
- Implementação do trabalho prático (1 aula)

Algumas Recomendações

■ Erros habituais:

- ✦ As aulas práticas são suficientes;
- ✦ Os testes/exames são uma questão de sorte pois são de escolha múltipla;
- ✦ Basta ir às aulas e dar uma vista de olhos pelos acetatos nas vésperas dos testes/exame;

■ A matéria não é complicada.

- ✦ No entanto, é exigido um conhecimento **profundo e rigoroso**.
- ✦ Este conhecimento **não se adquire na véspera** do teste/exame.
- ✦ Requer um certo amadurecimento que só se obtém com tempo.

Algumas Recomendações

- Aos alunos que só vêm as primeiras três/quatro aulas teóricas.
 - ✦ Não é boa ideia! ...pelo menos, estatisticamente, não tem sido.
 - ✦ Devem ler o livro com muita atenção para perceber bem os conceitos. **Ler os acetatos não chega!**
- Para os alunos que vêm às aulas:
 - ✦ O mais indicado é ler o livro antes de cada aula (ver planificação).