

Bases de Dados

Engenharia Informática (2ºano)

Tecnologias Digitais e Gestão (2ºano)

Capítulo 1. Introdução aos ambientes de Base de Dados

João Paulo Pereira | jprp@ipb.pt
Marisa Ortega | marisa.ortega@ipb.pt
David Dias | davide.dias@ipb.pt
Tiago Santos | tiago.santos@ipb.pt

2025



Índice

● Conteúdo

- ➔ 1. Introdução aos ambientes de Base de Dados
 - ⇒ - Noção de Sistema de Informação
 - ⇒ - A Informação nas Organizações
 - ⇒ - Tecnologias de Informação
 - ⇒ - Gestão de Informação
- ➔ 2. Sistemas de Gestão de Bases de Dados
 - ⇒ - Abordagem e Vantagens
 - ⇒ - Arquitetura de um SGBD
 - ⇒ - Tipos de Utilizadores num SGDB
- ➔ 3. Modelação e Normalização de Dados
 - ⇒ - Manutenção da Integridade
 - ⇒ - Redundância e Chaves
 - ⇒ - Diagramas E-R
 - ⇒ - Modelo Relacional
- ➔ 4. Álgebra Relacional e SQL - (MySQL)
 - ⇒ - Conceitos e aplicação de Álgebra Relacional
 - ⇒ - Ferramentas de Administração MySQL
 - ⇒ - Comandos DDL
 - ⇒ - Comandos DML
- ➔ 5. Introdução às Bases de Dados NoSQL
 - ⇒ Bases de Dados NoSQL,
 - ⇒ Tipos de Bases de Dados NoSQL,
 - ⇒ Bases de dados orientadas a documentos
- ➔ 6. MongoDB
 - ⇒ Estruturas JSON e BSON,
 - ⇒ Modelação de dados,
 - ⇒ Criação de coleções e documentos,
 - ⇒ Operações CRUD e agregação,
 - ⇒ Indexação e transações



Avaliação

- Alternativa 1 - (Ordinário, Trabalhador) (Final, Recurso, Especial)
 - ➔ Exame Final Escrito - 50% (Nota Mínima Obrigatória de 7 Valores)
 - ➔ Trabalhos Práticos - 50% (Avaliações Práticas, resolvidas em Sala de Aula)
- Alternativa 2 - (Trabalhador) (Especial)
 - ➔ Exame Final Escrito - 100% (No PC)
- Importante:
 - ➔ Presenças obrigatórias: 80%
 - ➔ Datas tarefas práticas (Avaliações Práticas, resolvidas em Sala de Aula)
 - ⇒ 1) Modelo relacional – Semana 4 (10%)
 - ⇒ 2) SQL – Semana 12 (25%)
 - ⇒ 3) MongoDB – Semana 15 (15%)



Necessidade de Informação

- Antes de mais, deparamos com o conceito de informação que têm uma relação estreita com o conceito de dados.
 - ➔ Dados
 - ⇒ São um conjunto de factos diretamente resultantes de observações considerados isoladamente.
 - ➔ Informação
 - ⇒ Será todo o conhecimento que se pode derivar dos dados quer por simples interação destes, quer por complexos processos de raciocínio.
- Qualquer organização para sobreviver necessita de informação, quer para poder interagir com o seu meio ambiente, quer para permitir a interação entre as diferentes componentes da organização.
- Este fluxo de informação permite que a organização possa alcançar os objetivos que se propõem atingir.



Tipos de Informação

- Numa organização pode-se identificar dois tipos de informação:
 - ➔ INFORMAÇÃO OPERACIONAL
 - ⇒ É usada diariamente e permite que a organização leve a cabo eficientemente as suas atividades de rotina.
 - ➔ INFORMAÇÃO DE GESTÃO
 - ⇒ A que suporta o processo de tomada de decisão dos níveis de gestão;
 - ⇒ é uma informação menos rotineira que a operacional.
- A informação para as organizações é hoje, se não o mais importante, pelo menos um dos recursos cuja gestão e aproveitamento mais influencia o seu sucesso.
- É também considerada e utilizada em muitas organizações como:
 - ➔ um fator estruturante e um instrumento de gestão;
 - ➔ uma arma estratégica indispensável para a obtenção de vantagens competitivas, podendo ser o agente crítico que determina o seu sucesso ou decadência.



Características da Informação

- Raio de ação da informação
 - ➔ Informação gestão: Amplo
 - ➔ Informação operacional: Reduzido
- Nível de detalhe da informação
 - ➔ Informação gestão: Geral
 - ➔ Informação operacional: Muito detalhada
- Síntese da informação
 - ➔ Informação gestão: Geralmente sumariada
 - ➔ Informação operacional: Geralmente em bruto
- Idade da informação
 - ➔ Informação gestão: Pode ser antiga
 - ➔ Informação operacional: Tem de ser atual
- Exatidão da informação
 - ➔ Informação gestão: Estimativas podem ser suficientes
 - ➔ Informação operacional: Precisa de ser exata
- Tipo de informação
 - ➔ Informação gestão: Mais qualitativa
 - ➔ Informação operacional: Mais quantitativa
- Fontes da informação
 - ➔ Informação gestão: Principalmente externa
 - ➔ Informação operacional: Principalmente interna



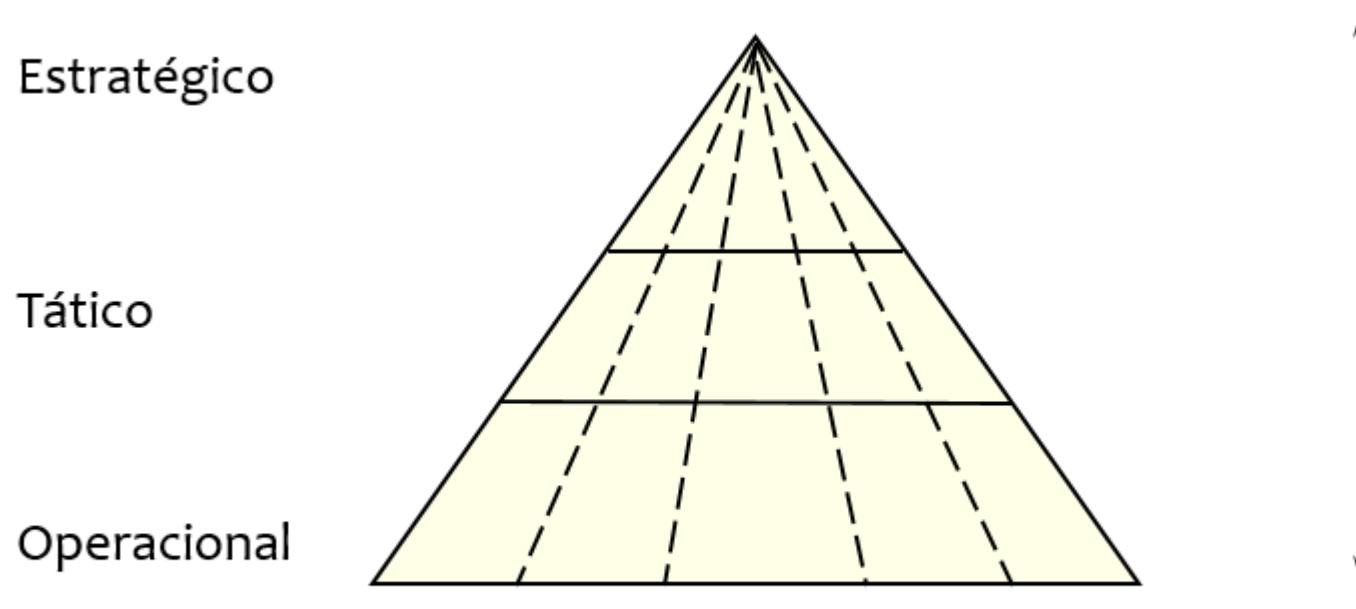
Valor da Informação

- Atualmente a informação é aceite como um recurso que qualquer tipo de organização tem de gerir com grande determinação para “sobreviver” ...
 - ➔ ... funciona como uma arma estratégica indispensável à obtenção e manutenção de vantagens competitivas.
- A informação, como recurso que é, tem de ser gerida com a mesma determinação dos restantes recursos das organizações.
- Qualquer tipo de informação deve estar disponível na organização de uma forma:
 - ➔ facilmente utilizável e compreensível,
 - ➔ estar disponível a tempo e
 - ➔ ser obtida a custo razoável.



Fluxo de Informação nas organizações

- Em qualquer organização observa-se:

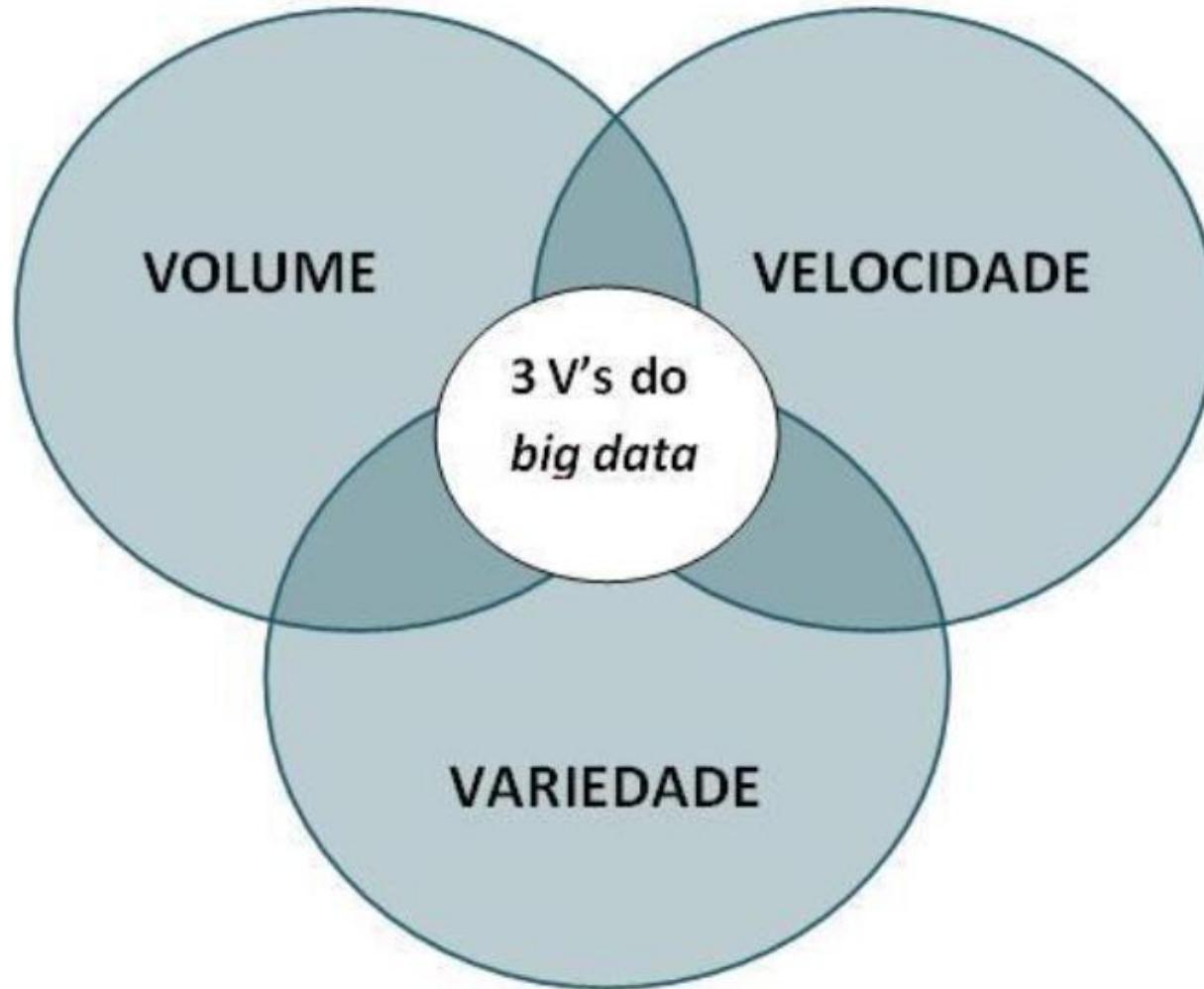


Big Data

- Big Data é o termo que descreve o gigantesco volume de dados que refletem os negócios no dia a dia.
- Mas o importante não é a quantidade de dados, mas sim o que as empresas fazem com os dados que realmente importam.
- O conceito ganhou força no início dos anos 2000. Definição de big data como os três Vs:
 - ➔ Volume
 - ⇒ As organizações armazenam dados de uma grande variedade de fontes, incluindo transações comerciais, redes sociais e informações de sensores ou dados transmitidos de máquina a máquina.
 - ⇒ No passado, armazenar tamanha quantidade de dados teria sido um problema – mas as novas tecnologias têm aliviado a carga.
 - ➔ Velocidade
 - ⇒ Os dados fluem a uma velocidade sem precedentes e devem ser tratados em tempo hábil.
 - ⇒ Dados provenientes de RFID, sensores, celulares e contadores inteligentes estão a impulsionar a necessidade de lidar com enormes quantidades de dados em tempo real, ou quase real.
 - ➔ Variedade
 - ⇒ Os dados são gerados em todos os tipos de formatos - de dados estruturados, dados numéricos em bases de dados tradicionais, até documentos de texto não estruturados, e-mail, vídeo, áudio, dados de cotações da bolsa e transações financeiras.



Big Data



Big Data

- Assim:
 - ➔ Nas empresas, os dados são armazenados a todo o momento, formando grandes volumes de dados.
 - ➔ Os dados armazenados contêm informações ocultas de grande relevância para o negócio.
 - ➔ Devido ao grande volume de dados, a extração destas informações não é uma tarefa trivial.
 - ➔ Existe uma grande necessidade de teorias e ferramentas para o auxílio dos analistas para a extração e análise de informações úteis.
- Com o constante crescimento e amadurecimento dos sistemas de informação, torna-se cada vez mais necessária a utilização de técnicas de Data Mining, capazes de extrair padrões de comportamento que forneçam, entre outros, informação intrínseca nos dados, com sentido e viável no domínio do negócio em questão.



Necessidade de Sistemas de Informação (SI) nas organizações

- As organizações precisam de ter um tipo especial de sistema - SISTEMA DE INFORMAÇÃO - o qual serve todos os outros subsistemas da organização ligando-os através da informação
- Os SI têm adquirido uma importância crescente, proporcional à relevância que estes sistemas têm vindo a assumir na sociedade.
- Pode-se afirmar que deparamos com um SI em quase todos os ramos da atividade a servir de apoio a organizações (Comerciais, industriais, científicas, etc.), tornando-lhes possível a manipulação eficiente de grandes quantidades de informação
- O controlo do grande caudal de informação tão característico das sociedades contemporâneas, é um fator decisivo para o sucesso de uma organização.
- É necessário ter em atenção, como o recurso a um SI adequado pode influenciar o aumento da produtividade nas diversas áreas de atividade.

“ ...a possibilidade de atingir estes objetivos depende, naturalmente, da qualidade do sistema de informação”



Sistema de Informação (SI) - definição

- “Sistema de informação é o meio que providencia os meios de armazenamento, geração e distribuição de informação com o objetivo de suportar as funções de operação e gestão de uma organização.” Layzell & Loucopolus, 1987
- Assim, partindo da definição comum de sistema, que sugere um grupo organizado de componentes agregados para interagir e cumprir um objetivo bem definido, podemos facilmente concluir que um SI é um sistema com a finalidade de produzir a já referida informação.
- É necessário referir que um SI não é forçosamente um Sistema Informático.
 - ➔ Podemos facilmente conceber um sistema cujo suporte físico seja um conjunto bem organizado de arquivos e ficheiros de papel.
- Os avanços tecnológicos que a microinformática sofreu nos últimos anos dotaram os computadores de cada vez maiores capacidades de processamento e armazenamento de informação a preços cada vez mais reduzidos, e fizeram destes uma ferramenta indispensável na composição de um SI.



Sistema de Informação (SI)

- Assim, podemos identificar algumas funções:
 - ➔ Recolher dados, tais como factos, indicadores, rumores, etc.
 - ➔ Armazenar dados num dado suporte (computador, armário, memória das pessoas, etc.).
 - ➔ Manipular/Tratar dados, agregando, interpretando, etc.
 - ➔ Disponibilizar/Distribuir informação útil, apresentando-a aos utilizadores no formato e no suporte mais adequados.
- Os sistemas de informação podem ser manuais ou mais vulgarmente computadorizados ou mistos.
 - ➔ Sistema de informação automatizado: É a parte do sistema de informação cujo tratamento e execução é realizado total ou parcialmente em computador, recorrendo às Tecnologias de Informação e das Comunicações (TIC).



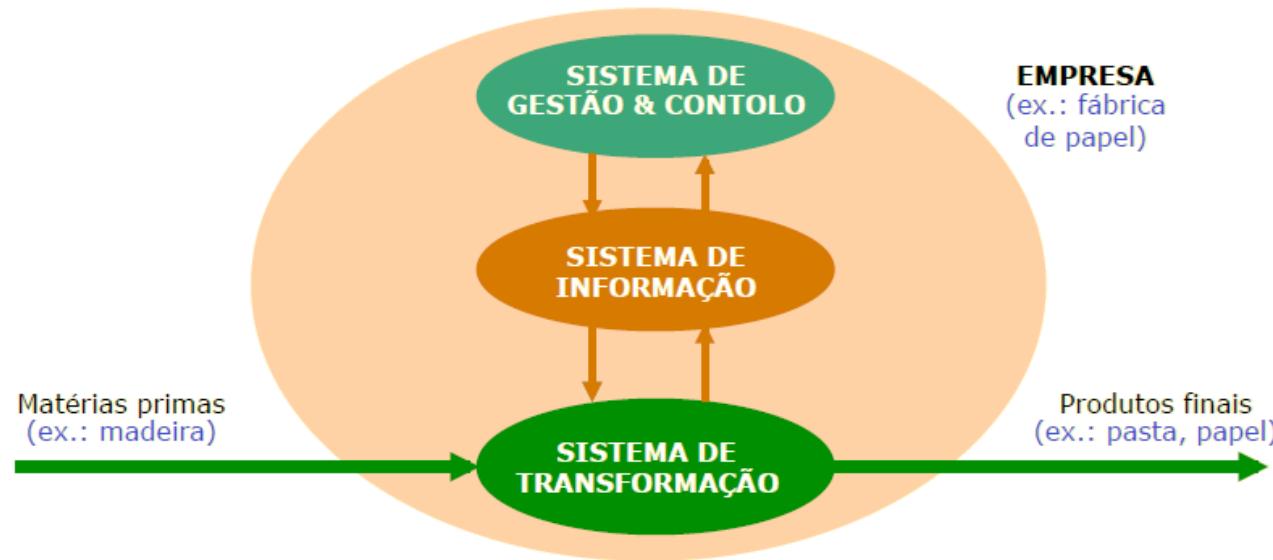
Sistema de Informação (SI)

- Os SI vistos como um sistema:

- ➔ Objetivo - orientar a tomada de decisão.
- ➔ Componentes - dados, sistema de processamento de dados, canal de comunicação.
- ➔ Estrutura - maneira como os diferentes processamento de dados estão ligados entre si.
- ➔ Comportamento - conjunto de procedimentos que se seguem para obter os dados, os processar e os enviar.
- ➔ Ciclo de Vida - se organização muda -» sist. informação muda; se organização não precisa mais do sist. informação -» sist. informação morre.



O SI como um sub-sistema de uma organização



“A finalidade dos sistema de informação é fornecer as informações certas, para as pessoas certas, no momento certo, na quantidade certa e no formato certo”



A Importância do SI

- O desenvolvimento de sistemas de informação tornou-se uma atividade de muita importância pois põe em causa o sucesso das organizações.
- O valor adicionado pelos SI:
 - ➔ Ajudar organizações ou indivíduos a melhorar os produtos ou os processos da organização;
 - ➔ Melhorar a qualidade;
 - ➔ Auxiliar na gestão, especialmente na tomada de decisões,
 - ➔ Criar e solidificar vantagens comparativas.
- Os SI podem apoiar diversas atividades, tais como:
 - ➔ Aquisição de conhecimentos;
 - ➔ Projeto de planos de ação;
 - ➔ Decisão da melhor alternativa;
 - ➔ Implantação e monitorização da escolha.



Tecnologias de Informação (TI)

- Ainda que conceptualmente seja aceitável a existência de SI sem a participação de Computadores...
 - ➔ ... a observação da realidade permite concluir que são muito raras as organizações que não integram computadores no seu SI.
- Assim, pode-se redefinir os SI com uma perspetiva mais Organizacional :
 - ➔ Sistema de Informação é uma combinação de procedimentos, informação, pessoas e TI, organizadas para o alcance de objetivos de uma organização.
- Apesar de insistentemente os autores “fugirem” de definir TI, pode-se dizer que, numa perspetiva estritamente tecnológica...
 - ➔ São o conjunto de equipamentos e suportes lógicos (hardware e software), que permitem executar tarefas como aquisição, transmissão, armazenamento, recuperação e exposição de dados.



Tecnologias de Informação (TI)

- Alguns autores agrupam as TI em:

- ➔ Hardware

- ⇒ Sistemas de Computação, Computadores Pessoais, Estações de Trabalho, Impressoras, discos, etc.

- ➔ Software de Sistema

- ⇒ Sistemas Operativos, Sistemas Gestão de Bases de Dados, Compiladores e interpretadores de Linguagens de Programação, etc.

- ➔ Comunicações

- ⇒ Hardware, software e serviços de comunicações.

- ➔ Ferramentas de Desenvolvimento

- ⇒ Ferramentas CASE, Linguagens de 4^a Geração, Ferramentas de Prototipagem.

- ➔ Software de Aplicação

- ⇒ Sistemas Periciais, Processamento de Texto, Correio Eletrónico, CAD-CAM, Sistemas de Informação de Gestão, Sistemas de Informação Executivos, Sistemas de Apoio à Decisão, Aplicações genéricas (Folhas de Cálculo, etc.), Aplicações específicas (Salários, Contabilidade, etc.).

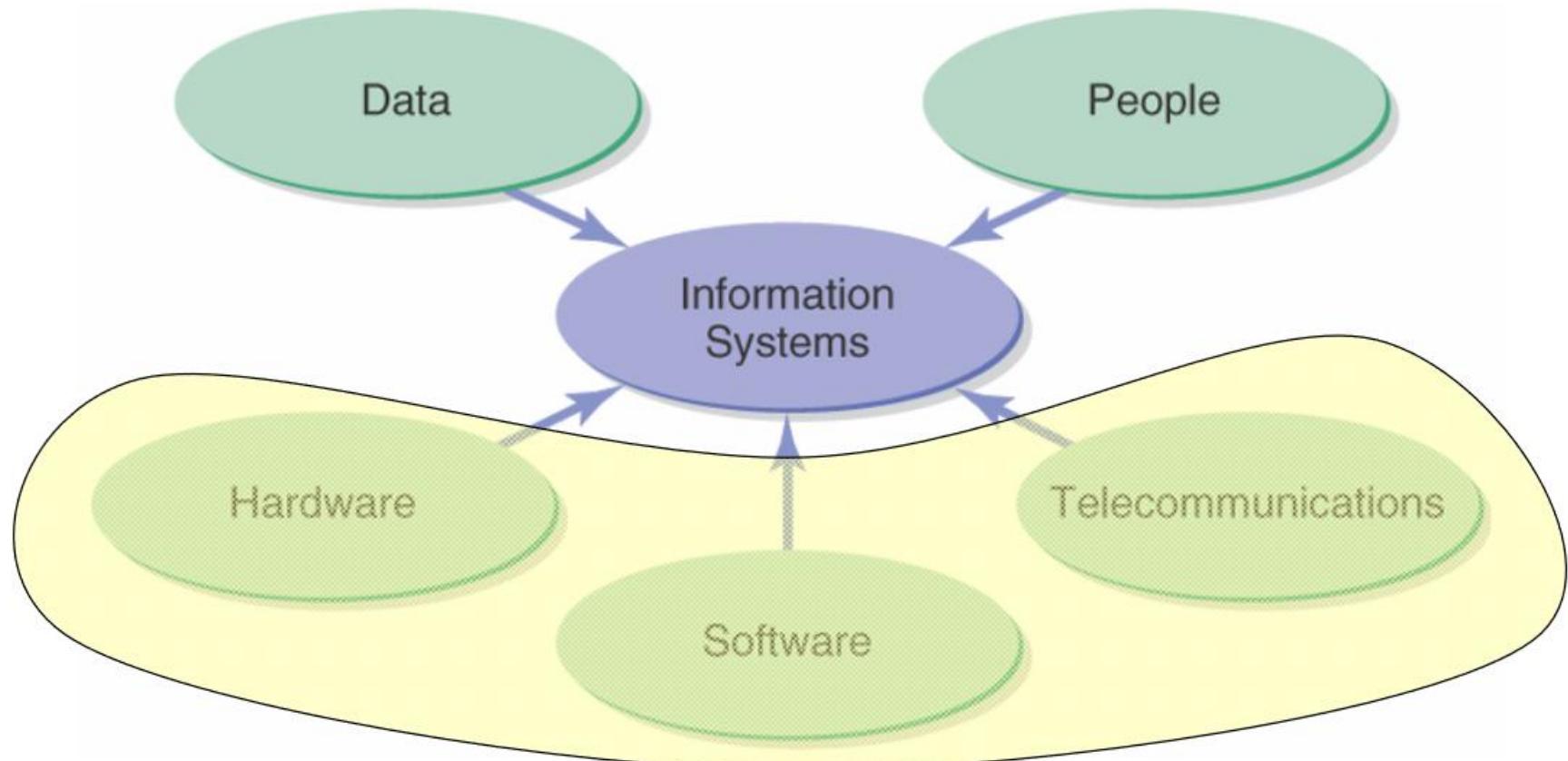


Tecnologias de Informação (TI)

- Em síntese, poderá dizer-se que as TI tornaram-se numa arma competitiva, tendo os gestores que entender como a sua utilização afeta o ambiente competitivo e a estratégia das organizações.
- Sistema de Informação versus Sistema Informático:
 - ➔ Todas as empresas têm um Sistema de Informação;
 - ➔ O Sistema de Informação inclui dados, informação, hardware, software, comunicações,...
 - ➔ Os Sistemas de Informação não são sinónimos de Informática;
 - ➔ O sistema informático é apenas a parte automatizada dos sistemas de informação, recorrendo às tecnologias da informação e das comunicações;
 - ➔ A informação deve ser entendida como um recurso, que tem de ser devidamente gerido.
- As TI constituem a infraestrutura tecnológica da informação, fornecendo a plataforma sobre a qual a empresa pode construir o seu sistema de informação.

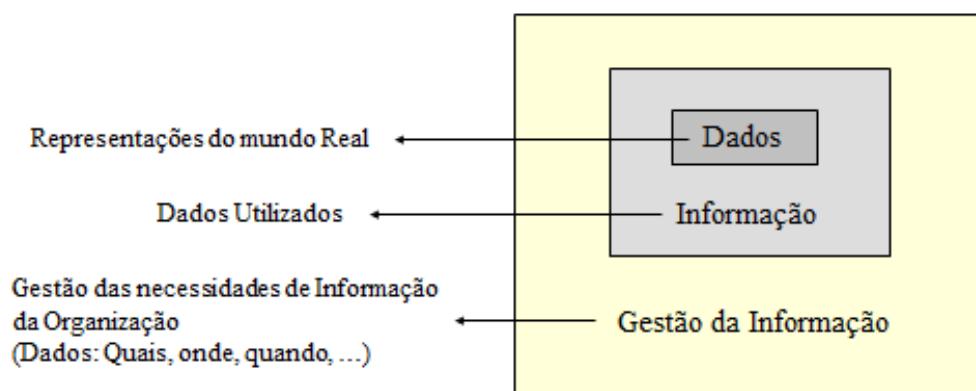


Sistema de Informação (SI)



Gestão da Informação

- A Informação como qualquer outro dos recursos vitais deve ser gerida, pelo que deve constituir o cerne de uma área funcional da gestão da organização a que se chama Gestão da Informação.
- A principal função desta função é o de manter uma visão global dos dados da organização, de modo a satisfazer as suas necessidades de informação possibilitando o cumprimento da missão que justifica a sua existência.
- Para satisfazer as necessidades das organizações é essencial a determinação de quais, onde e quando devem os dados estar presentes na vida da Organização.



Gestão da Informação

- A gestão da informação relaciona-se com questões do tipo:
 - ➔ “que dados são necessários?”
 - ➔ “onde são necessários?”
 - ➔ “quando são necessários?”
 - ➔ “que quantidade/qualidade?”
- Arrastadas pela importância que reconhecem à informação, muitas organizações não se apercebem (ou não reagem), a alguns excessos na procura e manutenção da informação.
- Em função do papel que a informação pode desempenhar nas atividades da organização, pode ser classificada como:
 - ➔ Critica (Essencial à sobrevivência da organização)
 - ➔ Útil (Essencial para uma boa gestão da organização)
 - ➔ Interessante (Essencial para a obtenção de vantagens competitivas pela utilização do SI)
 - ➔ Sem Interesse (Essencial para nada ...)



Gestão da Informação

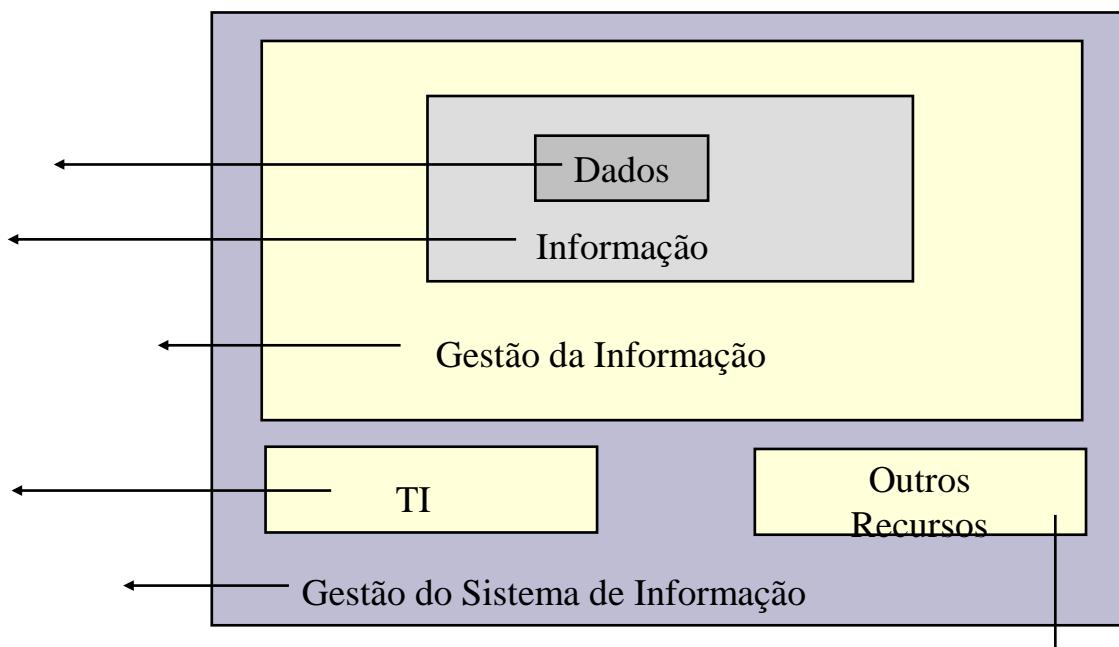
- A gestão da informação relaciona-se com questões do tipo:
 - ➔ “que dados são necessários?”
 - ➔ “onde são necessários?”
 - ➔ “quando são necessários?”
 - ➔ “que quantidade/qualidade?”
- Arrastadas pela importância que reconhecem à informação, muitas organizações não se apercebem (ou não reagem), a alguns excessos na procura e manutenção da informação.
- Em função do papel que a informação pode desempenhar nas atividades da organização, pode ser classificada como:
 - ➔ Critica (Essencial à sobrevivência da organização)
 - ➔ Útil (Essencial para uma boa gestão da organização)
 - ➔ Interessante (Essencial para a obtenção de vantagens competitivas pela utilização do SI)
 - ➔ Sem Interesse (Essencial para nada ...)



Gestão de SI

- Gestão de Sistemas de Informação é a Gestão do Recurso Informação e de todos os recursos envolvidos no planeamento, desenvolvimento, exploração e manutenção do SI.

Representações do mundo Real
 Dados Utilizados
 Gestão das necessidades de Informação da Organização
 (Dados: Quais, onde, quando, ...)
 Equipamentos, Suportes lógicos e Suportes metodológicos
 Gestão do recurso informação e de todos os recursos envolvidos no planeamento, desenvolvimento, exploração e manutenção do SI



Recursos humanos,
financeiros, etc.

Da Gestão da Informação à Gestão de SI



Bases de Dados

Engenharia Informática (2ºano)

Tecnologias Digitais e Gestão (2ºano)

Capítulo 2. Sistemas de Gestão de Bases de Dados (SGBD)

João Paulo Pereira | jprp@ipb.pt
Marisa Ortega | marisa.ortega@ipb.pt
David Dias | davide.dias@ipb.pt
Tiago Santos | tiago.santos@ipb.pt

2025



Índice

● Conteúdo

- ➔ 1. Introdução aos ambientes de Base de Dados
 - ⇒ - Noção de Sistema de Informação
 - ⇒ - A Informação nas Organizações
 - ⇒ - Tecnologias de Informação
 - ⇒ - Gestão de Informação
- ➔ 2. Sistemas de Gestão de Bases de Dados
 - ⇒ - Abordagem e Vantagens
 - ⇒ - Arquitetura de um SGBD
 - ⇒ - Tipos de Utilizadores num SGDB
- ➔ 3. Modelação e Normalização de Dados
 - ⇒ - Manutenção da Integridade
 - ⇒ - Redundância e Chaves
 - ⇒ - Diagramas E-R
 - ⇒ - Modelo Relacional
- ➔ 4. Álgebra Relacional e SQL - (MySQL)
 - ⇒ - Conceitos e aplicação de Álgebra Relacional
 - ⇒ - Ferramentas de Administração MySQL
 - ⇒ - Comandos DDL
 - ⇒ - Comandos DML
- ➔ 5. Introdução às Bases de Dados NoSQL
 - ⇒ Bases de Dados NoSQL,
 - ⇒ Tipos de Bases de Dados NoSQL,
 - ⇒ Bases de dados orientadas a documentos
- ➔ 6. MongoDB
 - ⇒ Estruturas JSON e BSON,
 - ⇒ Modelação de dados,
 - ⇒ Criação de coleções e documentos,
 - ⇒ Operações CRUD e agregação,
 - ⇒ Indexação e transações



Conceitos de Bases de Dados

- Base de Dados:
 - ➔ Conjunto estruturado de informação, organizada em Campos, Registos e Tabelas, que se inter-relacionam por intermédio de chaves e que permitem o acesso a diferentes perfis de utilizadores.
- Dados:
 - ➔ São elementos de Informação que permitem descrever as Entidades ou Eventos.
- Evento:
 - ➔ Acontecimento ou ocorrência relativa a uma entidade do qual se pretende guardar Dados e Informação.
- Informação:
 - ➔ Conjunto de dados que descrevem entidades ou eventos, organizados de uma forma significativa e relevante para um determinado contexto.
- Campo:
 - ➔ Componente elementar de um registo onde se armazenam os Dados que caracterizam as Entidades.
- Tabela:
 - ➔ Conjunto organizado de registos que armazenam a informação relacionada com as Entidades do sistema, segundo uma perspetiva matricial.
- Registo:
 - ➔ Corresponde a uma entrada individual na base de dados onde se armazena informação relacionada com as Entidades ou Eventos.



Conceitos de Bases de Dados

- Exemplo

Tabela cliente

Num_Cliente	Localidade	E_mail
1000	Bragança	joao@mail.pt
1100	Lisboa	joana@mail.pt
1200	Bragança	maria@mail.pt

Tabela encomenda

Num_Cliente	Num_Encomenda	Data_Encomenda
1000	123	01/03/2022
1200	150	01/03/2022
1200	245	12/03/2022

- Características das Bases de Dados

- Nas Bases de Dados, a Informação pode ser visualizada sob a forma de Tabelas, com uma linha para cada Registo ou Entrada e uma série de colunas para cada Campo ou Propriedade, o que origina que a inserção e manipulação dos dados seja feita de forma:
- ⇒ Organizada...
 - ⇒ Consistente...
 - ⇒ Integra...
 - ⇒ Não Redundante...
 - ⇒ Segura...

- Redundância:

- Duplicação desnecessária dos dados com inconvenientes:
- ⇒ Maior custo com armazenamento dos dados;
 - ⇒ Operações de atualização desnecessárias;
 - ⇒ Aumento custos com Pessoal;
 - ⇒ Maior possibilidade de Inconsistência dos dados.

- Inconsistência:

- A existência de redundância, pode levar a que determinados processos de atualização de dados, não atualizem os campos redundantes, penalizando-se assim a Integridade da Base de dados.

- Integridade:

- Existência de dados com conteúdos únicos.

- Segurança:

- A atribuição de determinadas permissões de acordo com os perfis de utilização, limita os acessos e processamentos a efetuar à base de dados, garantindo-se assim um maior nível de segurança e proteção dos dados.



Conceitos de Bases de Dados

● Redundância

- ➔ A redundância ocorre sempre que o mesmo facto é armazenado mais do que uma vez na mesma base de dados, ou quando factos derivados são armazenados numa base de dados.
- ➔ Os problemas que a redundância pode trazer a um sistema de bases de dados podem ser resumidos nos pontos seguintes:
 - ⇒ Problemas de manutenção:
 - Redundância significa o armazenamento repetido dos mesmos dados.
 - Como resultado, uma simples alteração ou remoção pode implicar o acesso a várias partes da base de dados, tornando-se difícil manter a coerência dos dados armazenados.
 - ⇒ Custos de espaço de armazenamento:
 - Apesar de, atualmente, não ser um problema significativo (o preço das memórias secundárias desce cada vez mais), não deixa de ser importante evitar ocupar espaço de armazenamento com algo que não acrescenta nada ao que já existe armazenado.
 - ⇒ Problemas de desempenho:
 - Apesar de não ser tão evidente, a redundância pode também trazer problemas de desempenho ao sistema.
 - A razão para isso é que a razão mais importante para um mau desempenho de um SGBD são os acessos aos discos onde as bases de dados se encontram armazenadas.
 - Se a redundância for significativa, isso implicará mais acessos aos discos para trazer os mesmos dados.
- ➔ Como exemplo de redundância, considere-se a relação Empregados e Projetos, a qual mostra os relacionamentos existentes entre projetos e os empregados que nele participam.

Código de Empregado	Número do Projeto	Orçamento	Tempo Total
E1	PROJ1	20	20
E3	PROJ1	20	16
E2	PROJ2	17	35
E2	PROJ3	84	42
E3	PROJ2	17	17
E2	PROJ1	20	83
E4	PROJ3	84	41
	PROJ4	90	

A coluna “Orçamento” do projeto é armazenado várias vezes, tantas quanto o número de pessoas (coluna Código de Empregado) participantes no projeto.

As desvantagens deste tipo de situações são várias:

- 1) Se o orçamento de um projeto muda, então teremos de alterar mais do que uma linha da tabela;
- 2) De cada vez que um pessoa é incluída num projeto, é necessário incluir também o orçamento do projeto;
- 3) Um projeto sem pessoas atribuídas, como PROJ4, só terá a coluna Orçamento preenchida; as restantes ficarão vazios. Isto significa que uma simples operação de introdução de uma pessoa num projeto será realizada de modos diferentes, de acordo com o estado atual da base de dados. Por exemplo, para inserir uma pessoa no projeto PROJ2 resume-se a inserir um nova linha na tabela Empregados e Projetos; para acrescentar uma pessoa ao projeto PROJ4, deve-se alterar o valor das colunas que estão vazias na linha correspondente a esse projeto.

Conceitos de Bases de Dados

● Manutenção da Integridade

→ A manutenção da integridade de uma base de dados relacional é fundamental para o seu bom funcionamento, devendo ser assegurada pelo SGBD (Sistema de Gestão de Bases de Dados) ou pelas aplicações que interagem com os dados.

→ Assim, uma base de dados relacional deve assegurar três tipos de integridade:

⇒ Integridade de Domínio

→ Esta é a forma mais elementar de manutenção de integridade. Consiste em que o SGBD faça respeitar as restrições que determinados atributos impõem sobre os valores que lhes podem ser atribuídos.

→ Exemplo:

- Numa relação que guarde informação sobre pessoas, o atributo Sexo apenas poderá aceitar um de dois valores possíveis: um valor que represente o sexo masculino, e outro valor que represente o sexo feminino.
- Nesta mesma relação, o atributo Data de Nascimento apenas poderá aceitar valores que representem datas válidas. Por exemplo, a data 30 de Fevereiro de 2005 é inválida e, portanto, não poderá ser aceite pelo atributo Data de Nascimento.

⇒ Integridade de Entidade

→ O problema da integridade de entidade coloca-se em relação à possível existência de valores nulos como valores de um atributo de uma relação.

→ Na terminologia das bases de dados, um valor nulo não significa um valor igual a zero; antes, significa que o valor não existe, ou que é desconhecido, ou que simplesmente não se aplica nessa ocorrência da relação.

→ O princípio da integridade de entidade estipula que nenhum atributo pertencente à chave primária de uma relação pode conter valores nulos.

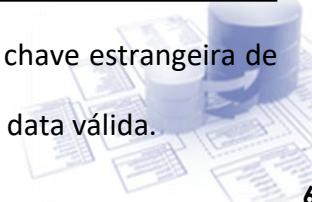
- Uma vez que a chave primária constitui o identificador dos tuplos da relação, a violação deste princípio equivaleria a admitir que a base de dados poderia registar informações de algo não identificável. Esta possibilidade é rejeitada pelo modelo relacional.

⇒ Integridade Referencial

→ O princípio da integridade referencial estabelece que um valor de uma chave estrangeira tem obrigatoriamente de existir como valor de um atributo da chave primária da tabela relacionada com aquela chave estrangeira.

- Ou seja, um valor que não exista como chave primária na sua tabela não pode ser usado como chave estrangeira de uma outra tabela.

→ Em rigor, nenhum atributo que represente datas poderá aceitar qualquer valor que não represente uma data válida.



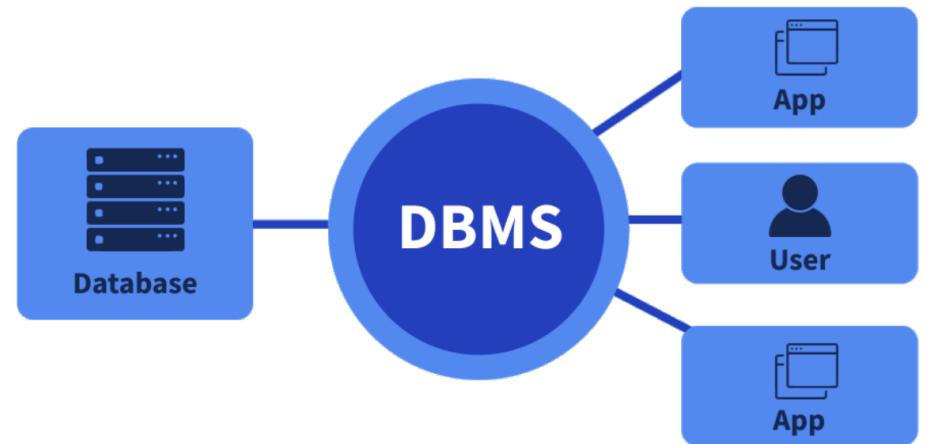
Conceitos de Bases de Dados

- Objetivos das Bases de Dados:
 - ➔ Facilitar o tratamento de grandes volumes de dados;
 - ➔ Processar informação de forma centralizada;
 - ➔ Assegurar a independência dos dados face aos processamentos;
 - ➔ Evitar a redundância no registo dos dados;
 - ➔ Evitar inconsistências ao nível dos processamentos de dados;
 - ➔ Garantir integridade dos dados;
 - ➔ Facilitar a partilha dos dados;
 - ➔ Garantir a confidencialidade e a segurança dos dados.



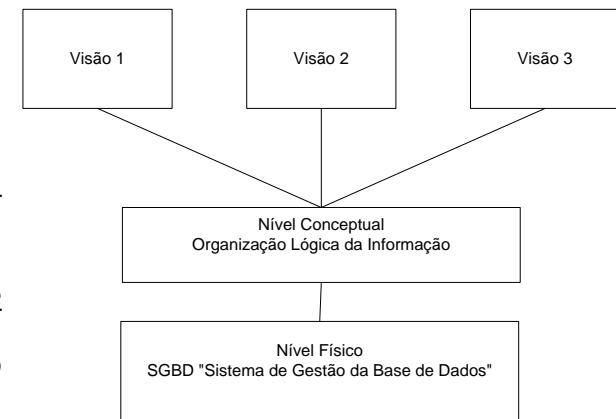
SGBD (Sistema de Gestão de Bases de Dados)

- O que é um SGBD (Sistema de Gestão de Bases de Dados)
 - ➔ O SGBD é uma porção de software que faz a ligação entre os utilizadores, ou programas de aplicação, e a base de dados propriamente dita.
 - ➔ É tarefa do SGBD conciliar os pedidos de acesso à base de dados originados pelos vários utilizadores.
 - ➔ Esse software deve proporcionar aos utilizadores um ambiente flexível, eficiente e de fácil utilização para a consulta, introdução, alteração e remoção de informação na base de dados.
- Os SGBDs são usados geralmente para:
 - ➔ criar ou gerir utilizadores
 - ➔ consultar dados
 - ➔ alterar dados e tabelas
 - ➔ excluir dados
 - ➔ relacionar tabelas
 - ➔ importar e exportar dados



SGBD (Sistema de Gestão de Bases de Dados)

- O modo como a informação está fisicamente armazenada não é relevante para o utilizador da base de dados...
 - ➔ pelo contrário, esse conhecimento apenas aumentaria a complexidade da operação.
- Nestas condições, uma função importante do SGBD é **apresentar ao utilizador uma visão abstrata da informação, escondendo os detalhes sobre como a informação está fisicamente armazenada** e, assim, simplificar a ação do utilizador.
- Normalmente, distinguem-se três níveis de abstração da informação, definidos como:
 - ➔ Nível Físico:
 - ⇒ É o mais baixo nível de abstração e descreve o modo como a informação está fisicamente armazenada nos dispositivos físicos apropriados (exemplo: HD, SSD,etc).
 - ➔ Nível Conceptual (ou lógico):
 - ⇒ Este nível é intermédio e descreve qual a informação que está armazenada na base de dados, e as várias relações internas a essa informação.
 - ⇒ Este nível é ainda relativamente complexo devido à inclusão de toda a informação sobre a organização dos dados existentes na base de dados.
 - ⇒ No entanto, o nível conceptual já não inclui a forma como a informação está fisicamente armazenada.
 - ➔ Nível de Visualização (ou aplicação):
 - ⇒ Também designado por nível de interação ou nível externo, é o nível com maior abstração da informação e contempla apenas uma parte da informação contida na base de dados.
 - ⇒ A informação disponível a este nível é ajustada às necessidades e competências de cada utilizador específico.
 - Em virtude da diversidade de utilizadores, existem normalmente várias visualizações para a mesma base de dados.



SGBD (Sistema de Gestão de Bases de Dados)

- Níveis de abstração da informação

➔ É frequente que a estrutura de uma base de dados seja alterada ao longo do tempo para melhor responder aos requisitos que lhe são endereçados, e assim aumentar a sua capacidade e eficiência.

⇒ A possibilidade de alterar a definição de um nível de abstração de informação sem afetar o nível seguinte designa-se por independência da informação.

➔ De acordo com os níveis de abstração considerados, existem dois níveis de independência da informação:

⇒ Independência da informação física:

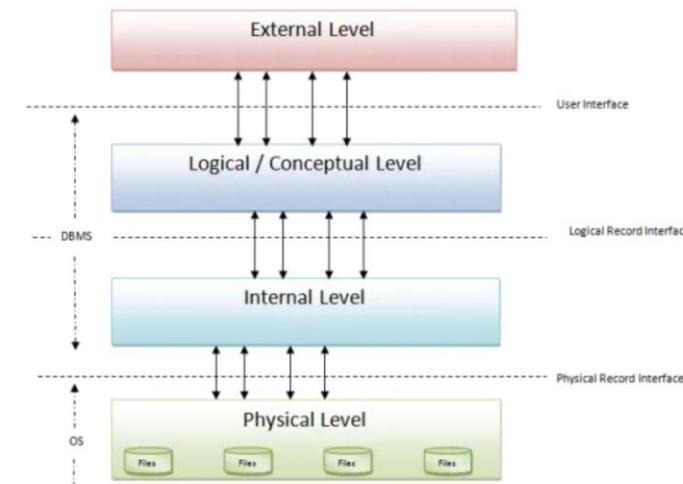
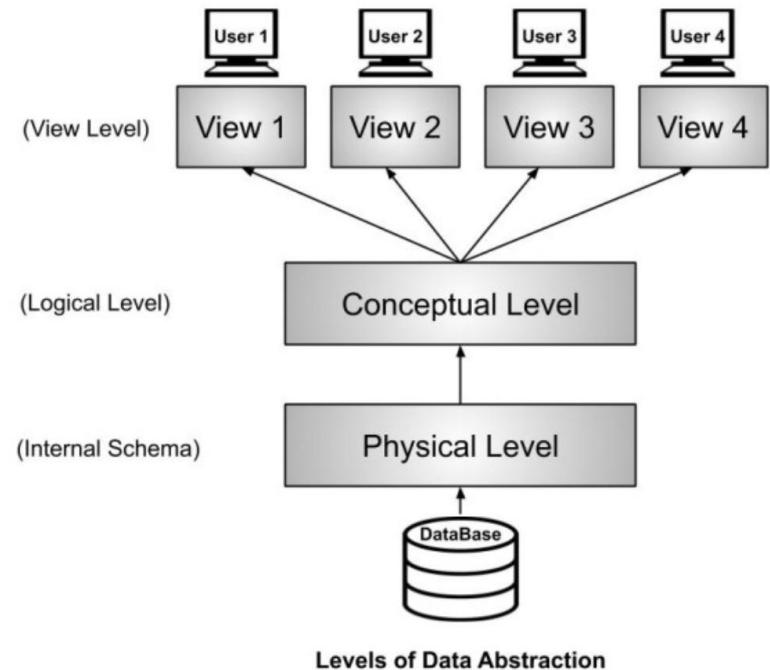
→ É a capacidade de alterar a estrutura de armazenamento físico da informação sem alterar a estrutura ao nível conceptual, e por conseguinte, sem alterar os programas de aplicação.

→ Alterações ao nível físico são esporadicamente efetuadas para aumentar a eficiência do armazenamento da informação e, consequentemente, o desempenho dos programas de aplicação que acedem a uma determinada base de dados.

⇒ Independência da informação lógica:

→ É a possibilidade de alterar a base de dados ao nível conceptual sem alterar os programas de aplicação.

→ Esta propriedade é difícil de conseguir, dado que os programas de aplicação baseiam-se na descrição da informação ao nível conceptual.



SGBD (Sistema de Gestão de Bases de Dados)

- Níveis de abstração da informação
 - ➔ A interação das aplicações e dos utilizadores com o SGBD faz-se à custa de linguagens apropriadas.
 - ➔ Ao nível conceptual, a estrutura de uma base de dados é especificada por um conjunto de instruções numa Linguagem de Definição de Dados (DDL) a qual deve permitir que o esquema de uma base de dados possa ser definido.
 - ⇒ O resultado dessa definição é um conjunto de tabelas armazenadas naquilo a que se chama dicionário de dados, o qual pode ser visto como um depósito central que descreve e define o significado de toda a informação usada no esquema da base de dados.
 - ➔ Os SGBDs modernos também disponibilizam aos utilizadores uma Linguagem de Manipulação de Dados (DML), que permite manipular (isto é, consultar, inserir, apagar ou alterar) os dados armazenados numa base de dados.
 - ⇒ A linguagem padrão que permite definir e manipular os dados armazenados numa base de dados chama-se SQL ("Structured Query Language", isto é, "Linguagem de Perguntas Estruturadas").



SGBD (Sistema de Gestão de Bases de Dados)

- Os dois principais modelos de SGBD são os relacionais e os não relacionais, e a diferença básica entre eles é a estrutura utilizada para organizar os dados.
- Modelos de bases de dados:
 - ➔ 1- Relacional (SQL)
 - ⇒ São o tipo mais comum de SGBD. Expressam os dados em tabelas e relacionamentos através de chaves, e utilizam a linguagem SQL para relacionar os dados.
 - ⇒ Aplicações: Oracle SQL, Microsoft SQL Server, PostgreSQL, MySQL, MariaDB.
 - ➔ 2- Não relacional (NoSQL)
 - ⇒ São bases de dados que não usam somente SQL ou o modelo relacional. São ideais para ambientes onde o modelo de dados é muito volátil (por exemplo: eventos em um website).
 - ⇒ Aplicações: Cassandra, Redis, MongoDB, CouchDB, Apache Hbase, Riak.
 - ➔ 3- Graph
 - ⇒ É um tipo de NoSQL que representa os dados em um grafo (graph) de vértices e arestas. Ideal para dados que estão altamente relacionados entre si (por exemplo, modelo de recomendação).
 - ⇒ Aplicações: Neo4j, Azure Cosmos DB, SAP HANA, Sparksee, Oracle Spatial and Graph, OrientDB



Utilizadores de Bases de Dados

- Relativamente ao tipo de utilizadores que interagem com os SGBDs, é possível distinguir duas classes de utilizadores com interesses e posturas diferentes perante estes sistemas:
 - ➔ O administrador da base de dados e os utilizadores propriamente ditos:
 - ⇒ Administrador
 - ⇒ Utilizadores



Utilizadores de Bases de Dados

● Administrador

➔ O administrador de uma base de dados é o responsável máximo pelo bom funcionamento de todo o SGBD. As suas funções mais importantes são:

- ⇒ Definição da estrutura: A estrutura da base de dados é definida por um conjunto de instruções que são transformadas pelo compilador da linguagem de definição de dados do SGBD num conjunto de definições que são armazenadas no dicionário de dados.
- ⇒ Alteração da estrutura: Às vezes, torna-se necessário alterar a estrutura de uma base de dados já existente. Esta ação é também definida por um conjunto de instruções utilizando a linguagem de definição de dados disponibilizada pelo SGBD.
- ⇒ Autorização de acesso aos dados: É tarefa do administrador especificar quais os dados que podem ser consultados ou alterados por cada tipo de utilizador que necessita de aceder à base de dados.
- ⇒ Condições de integridade: Cabe ao administrador definir quais as condições que devem ser satisfeitas pela informação existente na base de dados. Por outras palavras, o administrador da base de dados deve definir todas as restrições que garantem a integridade dos dados armazenados.



Utilizadores de Bases de Dados

● Utilizadores

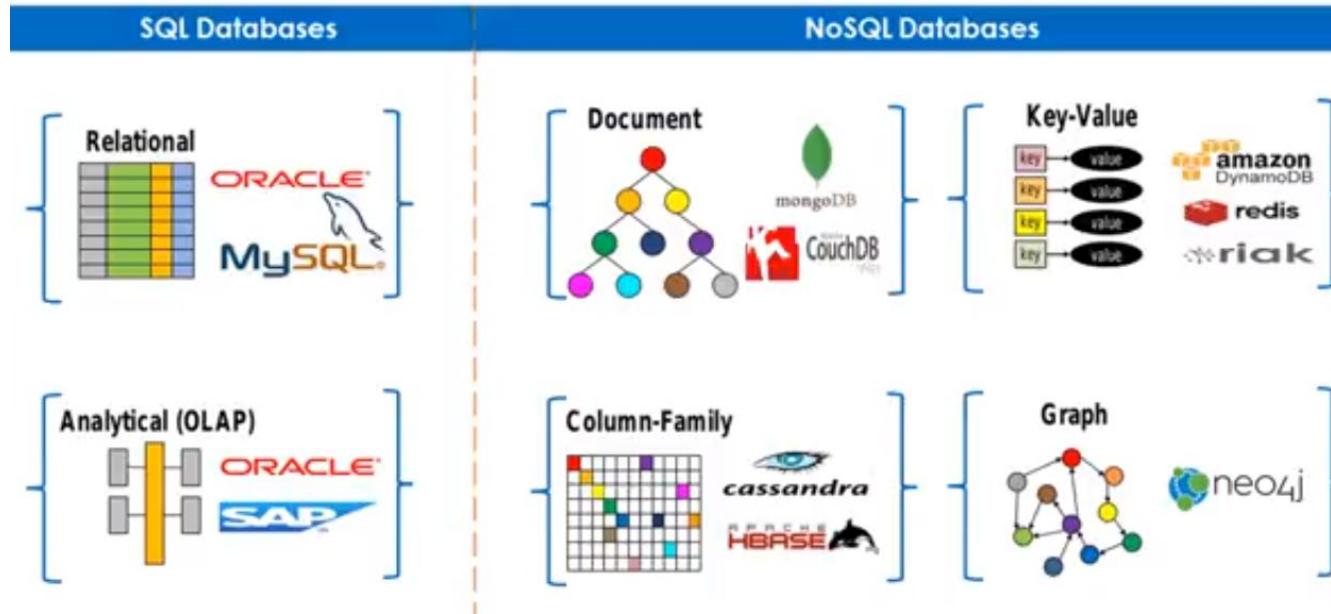
- ➔ O perfil de um utilizador de um SGBD é definido consoante o tipo de interação que este necessita de ter com o sistema.
- ➔ Basicamente, existem dois tipos de utilizadores:
 - ⇒ Programadores de aplicações:
 - São profissionais de informática que, através de linguagens de programação de alto nível (por exemplo, Pascal, C, Visual Basic, etc.) e através do esquema conceptual da base de dados, desenvolvem as aplicações informáticas que os utilizadores finais irão depois utilizar para aceder à base de dados.
 - ⇒ Utilizadores finais:
 - São a razão de existir da base de dados. Acedem ao sistema para consultar, alterar, adicionar ou remover dados, sob as condições de acesso definidas e impostas pelo administrador da base de dados. Estes utilizadores não poderão nunca alterar a estrutura (isto é, o esquema conceptual) da base de dados.



Tipos de SGBD mais usados

- SQL e NoSQL

SQL and NoSQL



- SGBD relacionais mais populares



- SGBD orientados a documento mais populares



Bases de Dados

Engenharia Informática (2ºano)

Tecnologias Digitais e Gestão (2ºano)

Capítulo 3. Modelação e Normalização de Dados

João Paulo Pereira | jprp@ipb.pt
Marisa Ortega | marisa.ortega@ipb.pt
David Dias | davide.dias@ipb.pt
Tiago Santos | tiago.santos@ipb.pt

2025



Índice

● Conteúdo

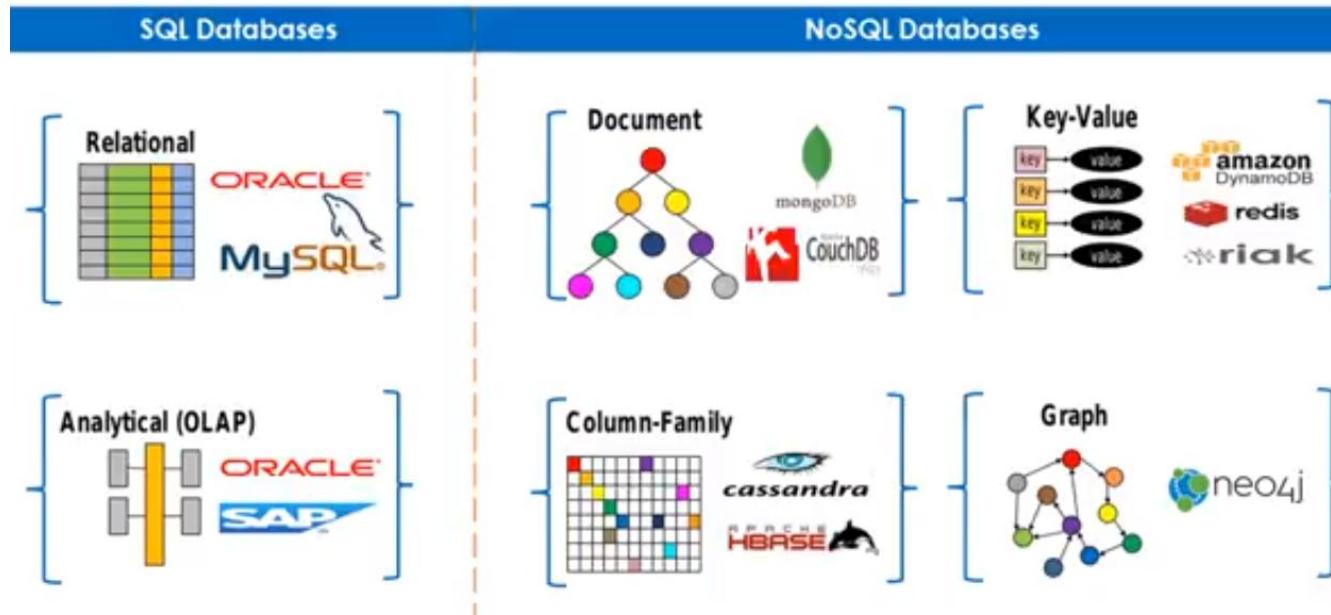
- ➔ 1. Introdução aos ambientes de Base de Dados
 - ⇒ - Noção de Sistema de Informação
 - ⇒ - A Informação nas Organizações
 - ⇒ - Tecnologias de Informação
 - ⇒ - Gestão de Informação
- ➔ 2. Sistemas de Gestão de Bases de Dados
 - ⇒ - Abordagem e Vantagens
 - ⇒ - Arquitetura de um SGBD
 - ⇒ - Tipos de Utilizadores num SGDB
- ➔ 3. Modelação e Normalização de Dados
 - ⇒ - Manutenção da Integridade
 - ⇒ - Redundância e Chaves
 - ⇒ - Diagramas E-R
 - ⇒ - Modelo Relacional
- ➔ 4. Álgebra Relacional e SQL - (MySQL)
 - ⇒ - Conceitos e aplicação de Álgebra Relacional
 - ⇒ - Ferramentas de Administração MySQL
 - ⇒ - Comandos DDL
 - ⇒ - Comandos DML
- ➔ 5. Introdução às Bases de Dados NoSQL
 - ⇒ Bases de Dados NoSQL,
 - ⇒ Tipos de Bases de Dados NoSQL,
 - ⇒ Bases de dados orientadas a documentos
- ➔ 6. MongoDB
 - ⇒ Estruturas JSON e BSON,
 - ⇒ Modelação de dados,
 - ⇒ Criação de coleções e documentos,
 - ⇒ Operações CRUD e agregação,
 - ⇒ Indexação e transações



Tipos de SGBD mais usados

- SQL e NoSQL

SQL and NoSQL



- SGBD relacionais mais populares



- SGBD orientados a documento mais populares



Abordagem relacional

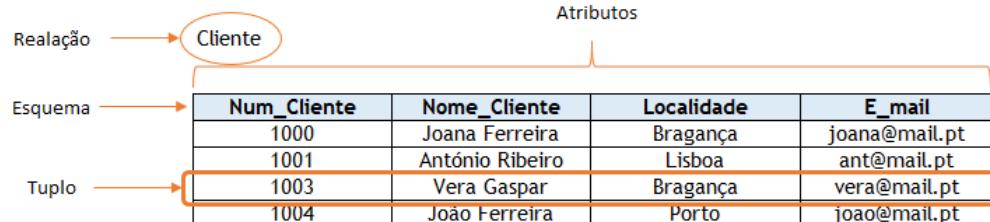
- Álgebra Relacional
 - ➔ A Álgebra Relacional é uma linguagem processual de questões.
 - ➔ Consiste num conjunto de operações que têm uma ou duas relações como entradas e produzem uma nova relação como resultado (ou saída).
 - ➔ Pode-se dizer que a Álgebra Relacional é uma linguagem formal de interrogação a bases de dados relacionais.
- As operações básicas da Álgebra Relacional são as operações equivalentes da teoria dos conjuntos (reunião, intersecção, diferença, produto cartesiano), e algumas operações específicas que são úteis ao modelo relacional:
 - ➔ Seleção de tuplos (linhas) numa relação;
 - ➔ Seleção de atributos de uma relação (projeção);
 - ➔ Junção natural de duas relações;
 - ➔ Semi-junção natural de duas relações;
 - ➔ Divisão de duas relações;
 - ➔ Renomeação de relações ou de atributos de uma relação.



Abordagem relacional

- Modelo relacional

- O modelo relacional tornou-se o modelo conceptual de bases de dados mais difundido e utilizado pelos SGBDs atualmente existentes no mercado.
- Devido à sua relativa simplicidade, o modelo relacional contribuiu decisivamente para a massificação da utilização da tecnologia de bases de dados nas organizações.
- A grande maioria dos SGBDs disponíveis no mercado atualmente são baseados no modelo relacional de dados: Oracle, SQL Server, MySQL, PostgreSQL, etc.
- O modelo relacional é baseado na teoria matemática dos conjuntos, no qual a informação e as suas associações internas são descritas:
 - ⇒ por várias **relações** (também designadas como **tabelas**), por serem constituídas por linhas e colunas.
 - ⇒ A relação é a estrutura fundamental do modelo relacional. Uma relação é uma estrutura bidimensional com um determinado **esquema**, e zero ou mais **instâncias**.
 - ⇒ O esquema de uma relação é constituído por um ou mais **atributos** (também chamados de **campos**, que correspondem a cada uma das colunas da tabela), os quais traduzem o **tipo de dados** que serão armazenados na relação. Os atributos são as características da entidade real que a relação pretende representar.



- No modelo relacional, as relações possuem algumas características que devem ser sempre tidas em conta:
 - ⇒ O valor de cada atributo num tuplo é atómico, isto é, no cruzamento de uma linha (tuplo) com uma coluna (atributo), só é possível encontrar um valor.
 - ⇒ Os atributos de uma relação devem ter identificadores distintos. Apesar de num mesmo esquema de base de dados relacional poderem surgir atributos com identificadores idênticos, estes atributos terão que pertencer a relações distintas.
 - ⇒ Os tuplos de uma relação devem ser distintos, isto é, numa relação não podem existir dois tuplos com os mesmos valores para todos os atributos. Por outras palavras, não pode haver linhas duplicadas numa relação.
 - ⇒ A ordem dos tuplos numa relação não tem qualquer significado.
 - ⇒ A ordem dos atributos no esquema de uma relação não tem qualquer significado.
 - ⇒ Todos os valores de um atributo provêm do seu domínio. Porém, o valor de alguns atributos num tuplo poderá ser desconhecido, ou não existir. Para estes casos, diz-se que o valor do atributo é nulo.

Abordagem relacional

- Modelo relacional

- ➔ A arquitetura de uma base de dados relacional é vista pelo utilizador apenas como um conjunto de tabelas relacionadas, independentemente das técnicas de organização interna da informação, ao nível do suporte físico.
- ➔ O SGBD é responsável pela implementação do sistema, ao nível do software, “escondendo” do utilizador as complexidades da gestão interna dos dados e possibilitando uma visão lógica da estrutura da informação, compatível com o modelo conceptual.
- ➔ Deve-se ter em conta que o modelo de bases de dados relacionais é uma teoria:
 - ⇒ Os SGBDs permitem transformar essa teoria em prática, ou seja, permitem implementar o modelo nos computadores.
- ➔ Exemplo:
 - ⇒ As tabelas representam uma implementação das relações Clientes e Encomendas, respetivamente, usando o modelo relacional.
 - ⇒ O objetivo destas tabelas é guardar os dados referentes aos clientes e às encomendas, respetivamente, na base de dados.
 - ⇒ Estas tabelas relacionam os clientes com as encomendas através do atributo **Num_Cliente**, que representa o número de um cliente (por exemplo: o cliente “1000 / Joana” tem 2 encomendas).

Cliente

Num_Cliente	Nome_Cliente	Localidade	E_mail
1000	Joana Ferreira	Bragança	joana@mail.pt
1001	António Ribeiro	Lisboa	ant@mail.pt
1003	Vera Gaspar	Bragança	vera@mail.pt
1004	João Ferreira	Porto	joao@mail.pt

Encomenda

Num_Cliente	Num_Encomenda	Data_Encomenda
1000	123	01/03/2022
1000	150	01/03/2022
1004	245	12/03/2022

Abordagem relacional

● Modelo Entidade – Relacionamento (DER)

- ➔ O desenvolvimento de um sistema de informação baseado em bases de dados implica o desenvolvimento de um **modelo conceptual de dados** onde se descreve a estrutura lógica dos dados, que depois será traduzido para um modelo particular de base de dados.
 - ⇒ Um SGBD permite que os dados sejam definidos em termos de um modelo conceptual de dados, o qual é, por sua vez, uma descrição de alto nível dos dados.
 - ⇒ Essa descrição permite ignorar muitos detalhes de armazenamento de baixo nível, em especial as interações com o sistema operativo para o acesso aos ficheiros físicos nos quais os dados serão efetivamente guardados.
- ➔ O modelo de dados Entidade-Relacionamento (mais conhecida pela sigla E-R) foi desenvolvido para auxiliar o projeto de bases de dados, através da especificação de um esquema lógico que define a organização ou a estrutura da base de dados.
- ➔ A modelação ER é a técnica de modelação de dados mais popular, dado ser simples, de fácil compreensão e leitura e um bom meio de discussão e análise em processos de definição para esquemas de bases de dados.
- ➔ O modelo E-R é muito útil para modelar, com um diagrama simples, um sistema de informação completo.
 - ⇒ O diagrama criado é depois melhor detalhado com a ajuda de outra técnica de modelação denominada por normalização.
- ➔ Informação do SI é organizada recorrendo ao conceito de:
 - ⇒ entidade,
 - ⇒ relação e
 - ⇒ atributo.



Abordagem relacional

- Modelo Entidade – Relacionamento (DER)

- ➔ Entidade

- ⇒ Qualquer coisa, real ou abstrata, sobre a qual guardamos dados para produzir informação requerida pelo o SI em estudo.
 - ⇒ Os dados que guardamos sobre as entidades são os atributos da entidade.
 - ⇒ Uma entidade tem de ter pelo menos dois atributos.
 - ⇒ Não há entidades “padrão” - a importância de uma “coisa” varia com a natureza e objetivo do SI em estudo.
 - ⇒ Os nomes das entidades devem ser um substantivo no singular:
 - Exemplo: Aluno, Requisição, Cliente, Fornecedor, etc.
 - ⇒ A entidade representa o conjunto de todas as ocorrências possíveis:
 - Exemplo: Entidade - Aluno; duas das suas ocorrências podem ser Ana Cardoso, Joaquim Santos



Abordagem relacional

- Modelo Entidade – Relacionamento (DER)

➔ Atributos

- ⇒ Para cada entidade é necessário conhecer as suas propriedades relevantes para o sistema.
- ⇒ Atributo é qualquer propriedade de uma entidade.
→ Exemplo: N° Cliente e Nome de Cliente
- ⇒ São sempre elementos atómicos (indivisíveis) de informação.
- ⇒ Um atributo poderá assumir diferentes valores de um domínio.

⇒ Tipos de Atributos

- Identificadores (Chaves candidatas)
- Descritores

Identificador

	Num_Cliente	Nome_Cliente	Localidade
	1000	Joana Ferreira	Bragança
	1001	António Ribeiro	Lisboa
	1003	Vera Gaspar	Bragança
	1004	João Ferreira	Porto

⇒ De todos os atributos de uma entidade algum (ou alguns) identificam uma ocorrência específica dessa entidade distinguindo-a das restantes.

→ A estes atributos chamamos **identificadores ou chaves candidatas**

- Para que um atributo seja identificador é preciso que não existam duas ocorrências distintas dessa entidade em que esse atributo assuma o mesmo valor.
 - Exemplo : O N° Cliente é o único identificador para cliente no caso presente

→ Aos restantes atributos, que só descrevem ou caracterizam as ocorrências da entidade, chamamos descritores.



Abordagem relacional

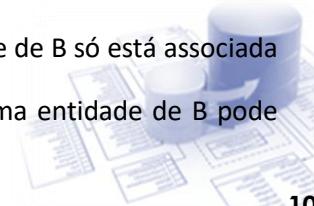
- Modelo Entidade – Relacionamento (DER)

- ➔ Relação (Associação)

- ➔ Um **relacionamento** é uma associação entre duas ou mais entidades.
 - Os relacionamentos são introduzidos no modelo E-R depois de as entidades terem sido identificadas, porque torna-se necessário perceber e definir o modo como as entidades do sistema se associam ou se relacionam entre si.
 - A todos os relacionamentos entre conjuntos de entidades dá-se o nome de **conjunto de relacionamentos**.
- ➔ As entidades não existem isoladas, estão associadas entre si no SI.
 - O nome da relação deve ser expresso na voz ativa, ser significativo e não muito longo.
- ➔ Exemplo:
 - O Cliente **faz** Encomenda
- ➔ Grau de Associação (Forma como duas entidades podem estar relacionadas):
 - Para cada ocorrência da entidade tipo A há uma e uma só ocorrência da entidade tipo B.
 - Para cada ocorrência da entidade tipo A há uma ou nenhuma ocorrência da entidade tipo B.
 - Para cada ocorrência da entidade tipo A há uma ou mais ocorrências da entidade tipo B.
 - Para cada ocorrência da entidade tipo A há uma, várias ou nenhuma ocorrência da entidade tipo B.
- ➔ Assim, uma associação pode relacionar :
 - Duas entidades entre si (binária)
 - Várias Entidades entre si (Complexa)
 - Uma entidade consigo própria (Unária)
- ➔ Um relacionamento entre entidades é caracterizado pela sua cardinalidade, ou seja, a quantificação de entidades que podem se relacionar através de um determinado relacionamento.
- ➔ Considerando dois conjuntos de entidades A e B, a cardinalidade dos relacionamentos entre entidades de A e de B pode ser de 3 tipos distintos:
 - **Um para um (1:1)**: uma entidade de A está associada apenas a uma entidade de B, e uma entidade de B está associada apenas a uma entidade de A.
 - **Um para vários (1:N)**: uma entidade de A está associada a qualquer número de entidades de B, mas uma entidade de B só está associada a uma única entidade de A.
 - **Vários para vários (N:N)**: uma entidade de A pode estar associada a qualquer número de entidades de B, e uma entidade de B pode estar associada a qualquer número de entidades de A

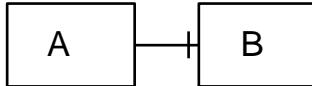
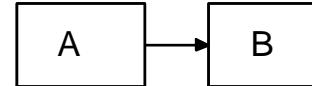
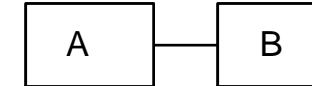
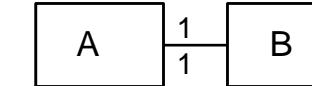
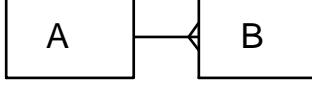
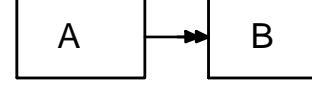
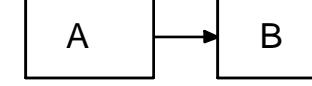
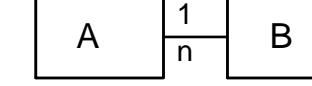
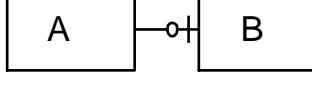
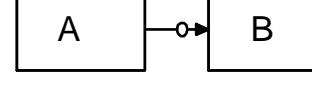
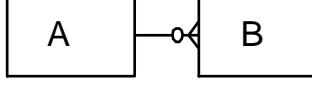
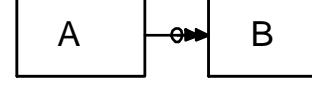
Ao analisar um sistema, a primeira análise para distinguir entidades e associações pode socorrer-se da notação utilizada pelas organizações:

- Substantivos: Para fazer referência a entidades
- Verbos: Para fazer referência a associações



Abordagem relacional

- Grau de Associação (Cardinalidade)

Notação	Pé de Galinha	Ross	Bachman	Chen
Um A está associado com um B				
Um A está associado com mais B's				
Um A está associado com zero ou um B				
Um A está associado com zero, um ou mais B's				



Abordagem relacional

- Modelo Entidade – Relacionamento (DER)

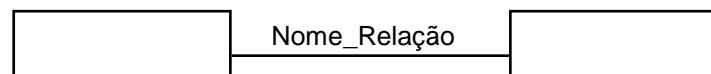
➔ Notação do DER

⇒ Entidade é sempre representada por um retângulo.

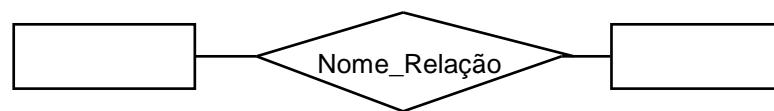


⇒ Tipo de Relação

→ sobre a linha



→ dentro de um losango

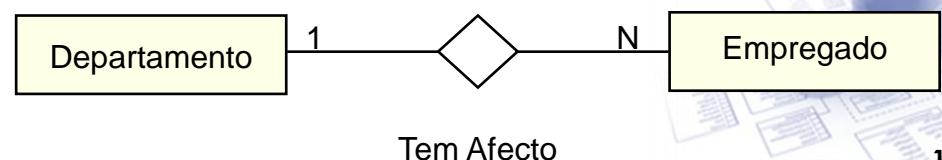


➔ Exemplo:

Uma associação que relate funcional com departamento pode limitar-se a indicar-nos:

⇒ Os funcionários afetos a um departamento;

⇒ A que departamento está afeto um funcionário.



Abordagem relacional

- Como construir um Diagrama ER?

→ Identificar as entidades

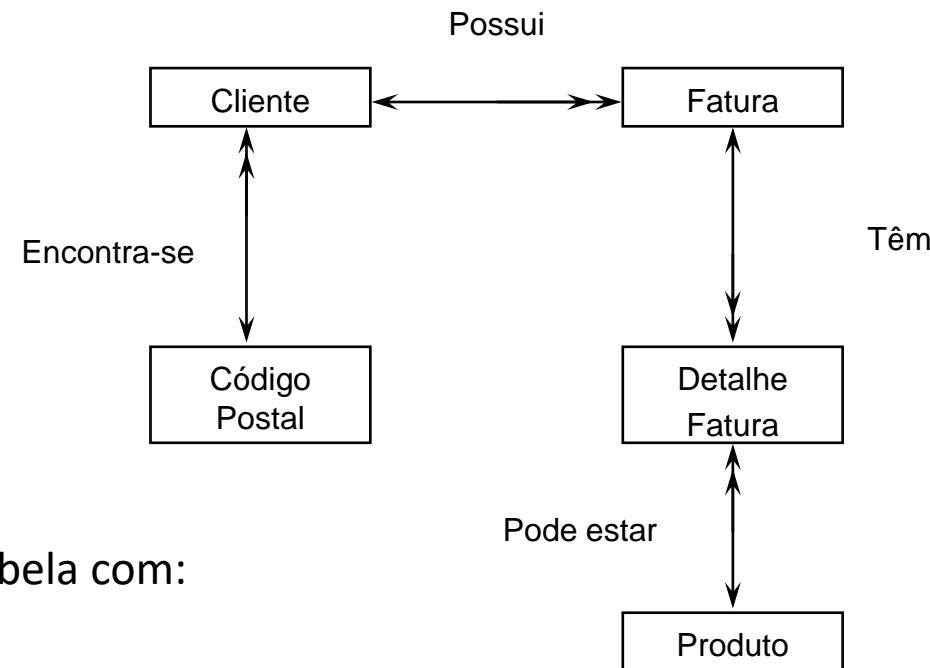
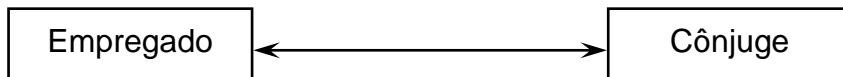
⇒ Sobre o que temos que guardar dados

→ Olhar para cada par de entidades e verificar se existe alguma relação entre elas...

⇒ Se existe:

→ Identificar o tipo de relação

→ Identificar o grau de associação



- Cada entidade dá normalmente origem a uma tabela com:

→ Identificador da entidade (Chave da Relação)

→ Restantes atributos da entidade (Descritores da Relação)

→ Identificadores de outras entidades que eventualmente lhes estejam associadas Chaves estrangeiras

Abordagem relacional

- Normalização

→ Técnica para analisar dados.

→ Problemas:

- ⇒ **redundância** da informação
- ⇒ mais **espaço em disco**,
- ⇒ mais **tempo de processamento**,
- ⇒ dificuldade de **manutenção da integridade** dos dados

→ O objetivo da normalização é eliminar a redundância de factos numa base de dados relacional.

⇒ Por outras palavras, através da análise das relações de uma base de dados relacional, são identificadas as **chaves primárias e estrangeiras dessas relações**, e as dependências funcionais existentes entre os atributos dessas relações, por forma a eliminar a redundância dos dados.

→ Características:

- ⇒ Permite derivar estruturas de dados flexíveis e sem anomalias.
- ⇒ Geralmente usado em conjunção com o diagrama ER.
- ⇒ A teoria da normalização foi construída à volta do conceito das Formas Normais
- ⇒ Diz-se que uma estrutura está numa determinada forma normal se satisfizer um conjunto de restrições.

→ Uma estrutura diz-se não normalizada quando nem sequer está na 1^a FN.

→ Restrições para uma estrutura se encontrar numa determinada forma normal:

- ⇒ Uma estrutura está na **1^a Forma Normal** se não houver grupos repetitivos de atributo(s), isto é cada atributo assume um só valor.
- ⇒ Uma estrutura está na **2^a Forma Normal** se já estiver na 1^a FN e cada atributo não chave depende inteiramente da chave.
- ⇒ Uma estrutura está na **3^a Forma Normal** se já estiver na 2^a FN e não há nenhuma dependência entre atributos não-chave.

Duas metodologias (podem ser usadas isoladamente ou combinadas):

- Normalização
- Entidades e relacionamentos

Dados não Normalizados

1. Decompor as estruturas com elementos de dados repetitivos em duas estruturas .

1^a Forma Normal

2. Para as estruturas que têm chaves com mais do que um elemento de dados, é necessário garantir que todos os outros elementos de dados dependem da chave como um todo. Dividir a estrutura, se necessário, para satisfazer esta condição.

2^a Forma Normal

3. Eliminar todas as dependências transitivas. Dividir a estrutura, se necessário, para atingir esta condição.

3^a Forma Normal



Abordagem relacional

● Normalização

- ➔ A normalização é um processo sistemático, conduzido por regras que traduzem as dependências funcionais entre os atributos, constituído por decomposições sucessivas de relações maiores em relações cada vez menores que, ao longo de vários passos, tenta identificar e remover a redundância própria das associações dos dados tal e qual existem no mundo real.
- ➔ Apesar de a normalização alterar a estrutura das relações, não pode alterar o conteúdo global da informação.
 - ⇒ Ou seja, durante o processo de normalização não pode haver perda de informação: deverá ser sempre possível regressar à estrutura original da base de dados.
- ➔ Para efetuar a normalização de uma base de dados, deve-se começar por analisar a informação da base de dados e estruturá-la com vista à construção das relações.
 - ⇒ Deve-se procurar incluir nessas relações todos os atributos considerados importantes.
- ➔ A normalização pode ser aplicada sobre um conjunto de dados “em bruto”, isto é, sobre a relação universal, ou então sobre as estruturas resultantes da modelação E-R, com o objetivo de as refinar.
- ➔ Redundância:
 - ⇒ A redundância ocorre sempre que o mesmo facto é armazenado mais do que uma vez na mesma base de dados, ou quando factos derivados são armazenados numa base de dados.
 - ⇒ Os problemas que a redundância pode trazer a um sistema de bases de dados podem ser resumidos nos pontos seguintes:
 - **Problemas de manutenção:** Redundância significa o armazenamento repetido dos mesmos dados. Como resultado, uma simples alteração ou remoção pode implicar o acesso a várias partes da base de dados, tornando-se difícil manter a coerência dos dados armazenados.
 - **Custos de espaço de armazenamento:** Apesar de, atualmente, não ser um problema significativo (o preço das memórias secundárias desce cada vez mais), não deixa de ser importante evitar ocupar espaço de armazenamento com algo que não acrescenta nada ao que já existe armazenado.
 - **Problemas de desempenho:** Apesar de não ser tão evidente, a redundância pode também trazer problemas de desempenho ao sistema. A razão para isso é que a razão mais importante para um mau desempenho de um SGBD são os acessos aos discos onde as bases de dados se encontram armazenadas. Se a redundância for significativa, isso implicará mais acessos aos discos para trazer os mesmos dados.

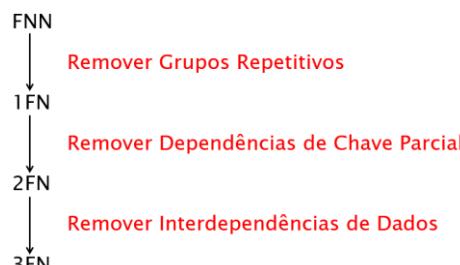
Abordagem relacional

● Normalização – Exemplo

- ➔ Informalmente, a normalização consiste em dividir uma tabela inicial em várias tabelas minimizando a redundância da informação.
- ➔ Iniciar pela identificação de uma chave
 - ⇒ A chave é um conjunto de campos, que identifica de forma única cada registo da tabela (Pode ser simples ou composta)

a) Estrutura não-normalizada

Numero_aluno
 Nome
 Morada
 Nome_mae
 Nome_pai
 Departamento
 Codigo_curso
 Nome_curso
Disciplinas*
 Código_disciplina
 Nome_disciplina
 Nota

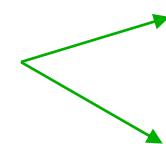


b) Estrutura na 1^ª Forma Normal

Numero_aluno
Nome
Morada
Nome_mae
Nome_pai
Departamento
Codigo_curso
Nome_curso

Numero_aluno
Código_disciplina
Nome_disciplina
Nota

(resultaram 2 estruturas na 1^ªFN)



c) Estrutura na 2^ª Forma Normal

Numero_aluno
Nome
Morada
Nome_mae
Nome_pai
Departamento
Código_curso
Nome_curso

Numero_aluno
Código_disciplina
Nota

(resultaram 3 estruturas na 2^ª FN)



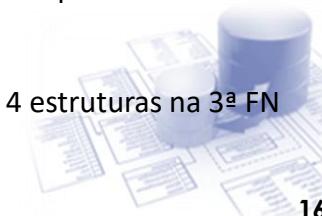
d) Estrutura na 3^ª Forma Normal

Numero_aluno
Nome
Morada
Nome_mae
Nome_pai
Código_curso
Nome_Curso

Numero_aluno
Código_disciplina
Nota

Código_disciplina
Nome_disciplina

(resultaram 4 estruturas na 3^ª FN)



Abordagem relacional

● Normalização

➔ O porquê da 1^aFN, 2^aFN e 3^aFN

⇒ Porquê a 1^aFN?

→ Base de dados relacionais não permitem registo de comprimento variável.

⇒ Porquê a 2^a FN?

→ Evitar redundância e consequentemente evitar inconsistência.

→ Diminuir o espaço necessário para guardar a informação.

⇒ Porquê a 3^a FN?

→ Evitar redundância e consequentemente evitar inconsistência.

→ Diminuir o espaço necessário para guardar a informação.

Exemplo de uma estrutura que não satisfaz a 2^aFN:

Numero_aluno

Codigo_disciplina

Nome_disciplina

Nota

- O Nome_disciplina é repetida para todas as ocorrências referentes a um aluno diferente (outro Número_aluno) mas que frequenta a mesma cadeira (Codigo_disciplina igual).
- Se temos 1000 alunos diferentes a frequentar a mesma disciplina repetimos 1000 vezes o mesmo nome da disciplina ---> redundância
- Se por qualquer razão muda o nome da disciplina temos que alterá-lo 1000 vezes ---> possibilidade de inconsistência

Exemplo de uma estrutura que não satisfaz a 3^aFN:

Numero_aluno

Nome

Morada

Nome da Mãe

Nome do Pai

Departamento

Código Curso

Nome Curso

- O Nome_curso é repetida para todas as ocorrências referentes a um aluno diferente (outro Número_aluno) que frequenta o mesmo curso (Código_curso igual).

- Se temos 1000 alunos diferentes a frequentar o mesmo curso repetimos 1000 vezes o mesmo nome da curso ---> redundância

- Se por qualquer razão muda o nome do curso temos que alterá-lo 1000 vezes ---> possibilidade de inconsistência

Abordagem relacional

- Normalização

→ Estruturas Finais - Combinar as diferentes estruturas resultantes da normalização numa estrutura final também normalizada.

→ **Todas as estruturas resultantes da normalização, (3^aFN) e que tenham o mesmo elemento de dados como chave, podem ser combinadas numa estrutura única garantindo que essa estrutura final também se encontra na 3^a FN.**

→ Exemplo:

⇒ Após normalização das diferentes estruturas definidas num determinado SI obtiveram as seguintes estruturas na 3^aFN.

A	B	C
---	---	---

A	D	E	F
---	---	---	---

W	F
---	---

W	P
---	---

A	D	P
---	---	---

W	F
---	---

W	P
---	---

⇒ Estruturas finais:

W	F	P
---	---	---

A	D	P	E	F
---	---	---	---	---

A	B	C
---	---	---

A	D	P
---	---	---

A	D	E	F
---	---	---	---

A	D	P	E	F
---	---	---	---	---

A	B	C
---	---	---



Abordagem relacional

- Normalização (Resumo)

Em geral, o processo de normalização é aplicado sobre uma estrutura não normalizada designada de relação universal, a qual deve conter todos os atributos dos quais é necessário guardar informação.

→ Passo 1: Nesta fase inicial do processo, o analista deve:

- Identificar quais os atributos que poderão ser repetidos.
- Identificar os atributos que poderão ser chave: um atributo que identifique os atributos não repetidos, e um outro atributo para identificar os atributos repetidos.

→ Passo 2: Aplicar as regras da 1.^a forma normal, isto é, separar os atributos repetidos dos atributos não repetidos. Para isso, o analista deve:

- Converter os atributos não atómicos em atributos atómicos, para que não seja possível incluir mais do que um valor em cada atributo de uma relação.
- Separar os atributos repetidos dos não repetidos, passando a considerá-los como elementos de uma nova relação. A chave primária da relação dos atributos não repetidos deve ser acrescentada, como chave estrangeira, à chave primária dos atributos repetidos.

→ Passo 3: Aplicar as regras da 2.^a forma normal; para isso, terá de verificar se os atributos não-chave de uma relação dependem da totalidade da chave primária dessa relação.

- Se a relação só tem um atributo como chave primária, e se essa relação já estiver na 1.^a forma normal, então a relação também se encontra na 2.^a forma normal.
- Se a chave primária é composta, e se algum atributo não-chave depende apenas de uma parte da chave primária, então a relação deverá ser decomposta, para que cada atributo dependa da totalidade da chave primária.

→ Passo 4: aplicação das regras da 3.^a forma normal, a qual obriga a analisar a dependência dos atributos não-chave de cada relação.

- O analista deverá procurar dependências funcionais entre os atributos não-chave de uma relação.
- Se uma relação que já esteja na 2.^a forma normal tiver, no máximo, um atributo não-chave, então essa relação também se encontra na 3.^a forma normal.
- Se existir algum conjunto de atributos não-chave de uma relação que tenha dependência funcional em relação a um outro conjunto de atributos não-chave daquela mesma relação, então a relação deverá ser decomposta, de modo que qualquer atributo não-chave dessa relação só dependa da chave primária da relação.

