



Aula VI Banco de Dados e Gerenciamento de Informações

Eduardo Kinder Almentero

ekalmentero@gmail.com

Sumário da Aula

1. Armazenamento de dados em sistemas de arquivos
 - Hierarquia de dados
2. Abordagem de banco de dados para armazenamento de dados
 - Sistemas Gerenciadores de Banco de Dados
 - SGDBs relacionais
 - Consultas em um SGBD relacional
3. Utilizando banco de dados para melhorar o desempenho do negócio e a tomada de decisões
 - Big data
 - Armazém de dados (data warehouse)
 - Mineração de dados
 - Mineração da Web

Introdução

- Um **sistema de informação eficaz** deve fornecer aos seus usuários **informações precisas, livres de erros e no momento em que são necessárias**.
 - A **informação é relevante** quando é **útil e apropriada** para os tipos de **trabalho e decisões** que a exigem.
- O **gerenciamento de informações é essencial** para garantir as **qualidades mencionadas**.
 - Muitas organizações **não dispõem de informações úteis**, pois possuem **dados desorganizados** decorrentes de **sistemas de informação mal projetados**.

Hierarquia de Dados

- Um sistema de computador organiza os **dados** através de uma **hierarquia**, que começa com **bits** e **bytes**, e passa por **campos**, **registros**, **arquivos** e **banco de dados**.
- Um **bit** representa a menor unidade de dados que o computador pode manipular;
- Um **byte** é um conjunto de **8 bits** e, normalmente, representa um **dado significativo**, como um **caractere**, **símbolo** ou **algarismo**, de acordo com uma convenção estabelecida.
- O **agrupamento de caracteres** em uma **palavra**, um **conjunto de palavras** ou um **conjunto de algarismos** que forma um **número completo** pode ser considerado um **campo**.
- Um **conjunto de campos relacionados**, como os dados pessoais de um estudante, forma um **registro**.
- Um **conjunto de registros do mesmo tipo** é chamado de **arquivo**.

Hierarquia de Dados

- Um registro descreve uma entidade
 - Pode ser uma pessoa, lugar, objeto, etc.
- Cada **característica** ou **qualidade** descrevendo uma **entidade particular** é chamada de atributo.

Banco de dados

Arquivo

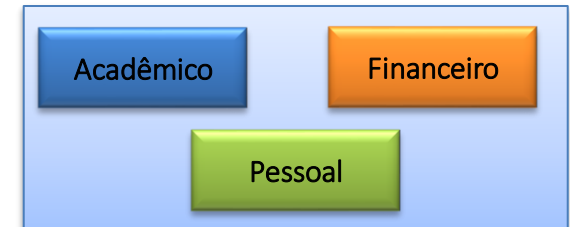
Registro

Campo

Byte

Bit

Banco de dados do estudante



TURMA			
Matrícula	Curso	Avaliação	Nota
015478	SI 039	P1	8,0
068743	SI 039	P1	7,0
698742	SI 039	P1	9,0

Matrícula	Curso	Avaliação	Nota
015478	SI 039	P1	8,5

SI 039

0101 0011 (S em ASCII)

0 ou 1

Figura adaptada de LAUDON, Kenneth C. et al. **Management information systems: Managing the digital firm**. Pearson Education India, 2007.

Problemas com o sistema de arquivos tradicional

- No ambiente organizacional havia a **tendência** de que cada **sistema evoluísse de forma independente**, sem uma **estratégia para toda a empresa**.
 - Cada sistema necessita do **seus próprios arquivos e software** para operar.
 - Por exemplo, os recursos humanos pode ter um arquivo de folha de pagamento, outro de plano de saúde, lista de e-mails, e etc.
- Esse processo levava a **criação de distintos sistemas de arquivo**, que eram **mantidos e operados por diferentes departamentos** (ou grupos).
- Ao longo do tempo, a **organização ficava sobrecarregada com centenas de programas** que são muito **difíceis de manter e gerenciar**.

Problemas com o sistema de arquivos tradicional

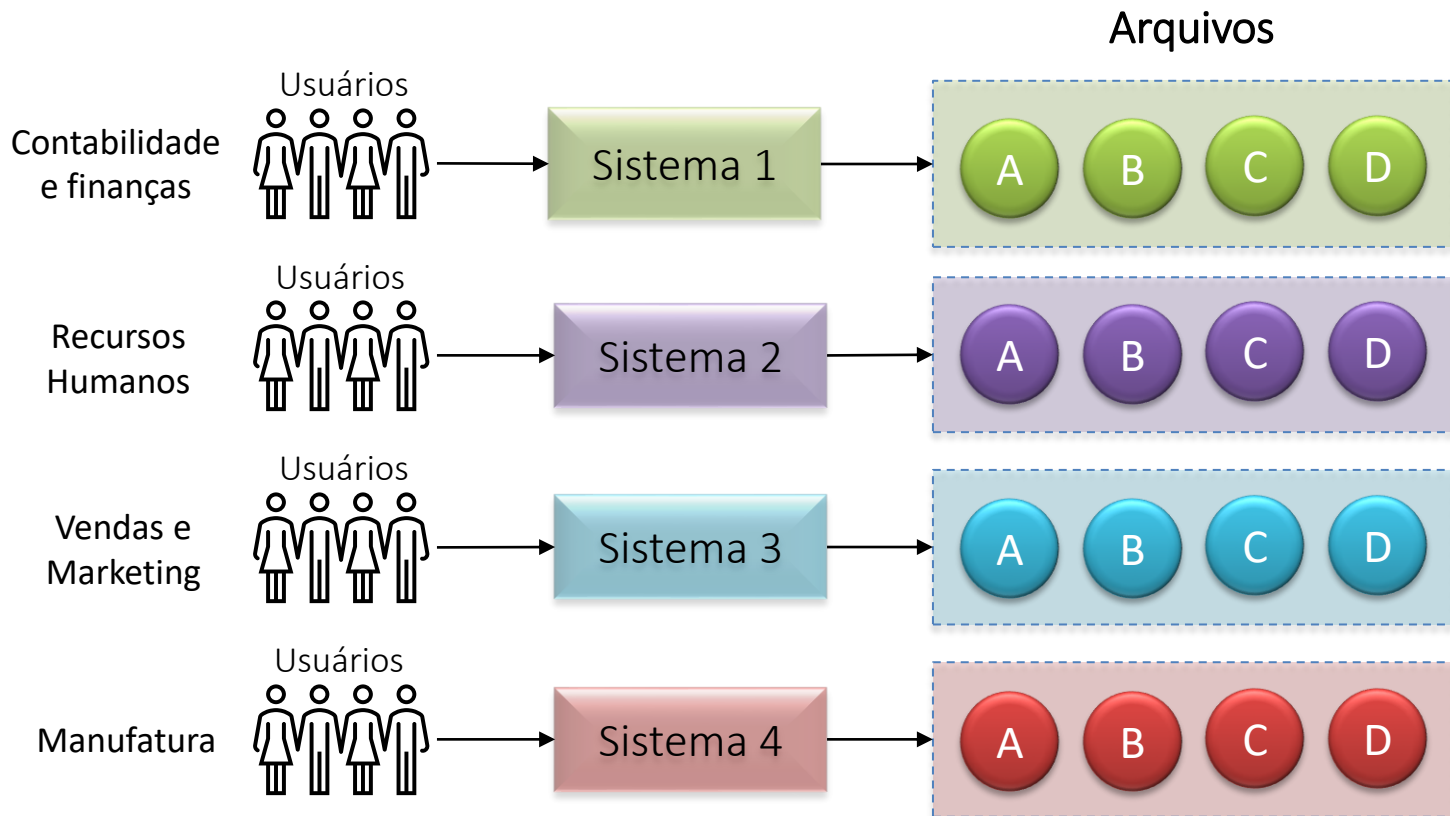


Figura adaptada de LAUDON, Kenneth C. et al. **Management information systems: Managing the digital firm**. Pearson Education India, 2007.

Problemas com o sistema de arquivos tradicional

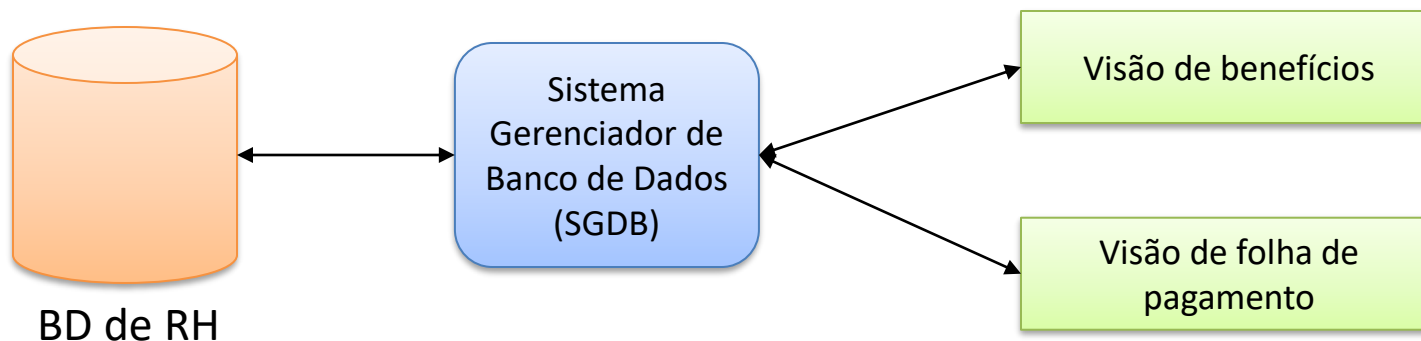
- Os problemas decorrentes desta abordagem são a **redundância e inconsistência**, falta de flexibilidade, dependência dados-programa, segurança de dados ruim e a incapacidade de compartilhar dados entre sistemas.
- A **redundância** é caracterizada pela presença de **dado duplicados em vários arquivos**, ou seja, o **mesmo dado** é **armazenado em mais de um lugar**.
 - Ocorre quando **grupos diferentes coletam o mesmo dado**, de forma independente, e o **armazenam em locais distintos**.
 - A **redundância** de dados **desperdiça recursos** de armazenamento, e também leva a **inconsistência de dados**.
- A **inconsistência de dados** ocorre quando o mesmo atributo possui valores diferentes (em locais distintos).
 - Ex.: o mesmo estudante pode ter notas distintas para mesma turma em arquivos distintos.

Problemas com o sistema de arquivos tradicional

- A dependência dados-programa
 - É comum uma mudança no programa ter impacto nos arquivos e causar uma inconsistência com os demais programas que os utilizam.
- Falta de flexibilidade
 - É difícil responder às necessidades que não foram previstas e permitir a geração de relatórios customizados, com informações relevantes para ações que não foram antecipadas.
- Segurança ruim
 - A ausência do gerenciamento de dados normalmente tem como consequência um **baixo controle sobre os acessos aos dados**.
- Incapacidade de compartilhar dados
 - Cada grupo possui seus próprios sistemas, e armazena os dados em seus próprios arquivos.

Abordagens para gerenciamento de dados

- O uso de banco de dados permite eliminar muitos dos problemas relatados.
 - Um banco de dados é uma **coleção de dados organizado** para servir a **muitos aplicativos**, de forma eficiente, **centralizando os dados** e **controlando dados redundantes**.
 - Em vez de armazenar os dados em arquivos separados para cada aplicativo, estes aparecem para os usuários como sendo **armazenados em um único lugar**.



Abordagens para gerenciamento de dados

- Sistemas Gerenciadores de Banco de Dados (SGDBs)
 - Ajuda a **controlar a redundância e inconsistência**;
 - **Minimiza o isolamento de arquivos**;
 - **Desacopla programas de dados**;
 - **Amplia o acesso e disponibilidade dos dados**;
 - Permite o **gerenciamento centralizado de dados**;
 - Incluem as **ferramentas para organizar, gerenciar e acessar os dados** armazenados no banco;
 - Permite **especificar a estrutura do conteúdo do banco de dados**, que é utilizada para **criação de tabelas** e definição das **características de cada campo**.

SGDBs Relacionais

- São os **mais utilizados** atualmente;
- Representam os dados como **tabelas bidimensionais**
 - Linhas (registros)
 - Colunas (campos)
- Cada tabela contém **dados sobre uma entidade**;
- Cada tabela possui **uma coluna** que é um **identificador único** para todas as informações de uma linha
 - É chamada de **chave primária**
- **Uma tabela pode possuir uma relação com outras tabelas**
 - Este relacionamento é possível através das **chaves primária e estrangeira**.

SGDBs Relacionais

Colunas (atributos, campos)

Fornecedor

ID	Nome	Rua	Cidade	Estado	CEP
6846	LTM Inc.	Rua Aristides Caire, 155	Rio de Janeiro	RJ	20775-090
9433	K. L. Peças	Rua Tripú, 328	São Paulo	SP	01235-000
6413	JP Corp.	Av. Doutor Arnaldo, 715	São Paulo	SP	01246-900
3642	OL Vendas	Av. Augusto de Lima 1120	Belo Horizonte	MG	30190-001

Linhas
(tuplas,
Registros)

Campo chave
primária

Parte

ID	Nome	Preco_Unidade	Fornecedor_ID
125	Espelho Lateral	98,25	6846
120	Compressor	741,21	3642
132	Trava da Porta	8,25	6846
181	Maçaneta	54,21	9433

Campo chave
estrangeira

Consultas em um SGBD relacional

- SQL (Structured Query Language) é a linguagem mais utilizada para manipulação de dados em um SGBQ relacional.
- Exemplo de consulta SQL:

```
SELECT Parte.ID, Parte.Nome, Parte.Preco_Unidade, Fornecedor.Nome,  
Fornecedor.Cidade  
FROM Parte, Fornecedor  
WHERE Parte.Fornecedor_ID = Fornecedor.ID AND Parte.ID = 125 OR  
Parte.ID = 181
```

Tabela resultante da consulta:

Parte.ID	Parte.Nome	Parte.Preco_Unidade	Fornecedor.Nome	Fornecedor.Cidade
125	Espelho Lateral	98,25	LTM Inc.	Rio de Janeiro
181	Maçaneta	54,21	K. L. Peças	São Paulo

Utilizando banco de dados para melhorar o desempenho do negócio e a tomada de decisões

- As empresas usam seus bancos de dados para **rastrear transações básicas**
 - Como pagamentos a fornecedores, processamento de pedidos, controle de clientes e pagamento de funcionários.
- Mas também precisam de bancos de dados para fornecer **informações** que **ajudarão** a empresa **administrar o negócio** com **mais eficiência** e **ajudar** os gerentes e funcionários a **tomar melhores decisões**.
- Se uma empresa deseja saber **qual produto é o mais popular** ou **quem é o cliente mais lucrativo**, a resposta está nos dados.

Big data

- Nos últimos anos houve um **aumento expressivo na geração de dados de dados**
 - tráfego da Web, mensagens de e-mail e conteúdo de mídia social (tweets, mensagens de status), bem como dados gerados por sensores ou sistemas de negociação eletrônicos.
- Esses dados podem ser **não-estruturados** ou **semiestruturados** e, portanto, **não são adequados SGBDs relacionais**, que organizam dados na forma de colunas e linhas.
- O termo **big data** é utilizado para **descrever esses conjuntos de dados com volumes tão grandes** que estão **além da capacidade de um DBMS típico** de capturar, armazenar e analisar.
 - não se refere a **nenhuma quantidade específica**, mas geralmente se refere a dados em o intervalo **petabyte** e **exabyte**
 - **Tweets** possuem tamanho limitado, mas geram aproximadamente **12 terabytes** por dia (dado de 2018).

Big data

- As **empresas** estão **interessadas** em big data porque eles podem **revelar mais padrões e anomalias relevantes** do que conjuntos de dados menores
 - Potencial para fornecer novos **insights sobre o comportamento do cliente, padrões climáticos**, atividades do mercado financeiro, ou outros fenômenos.
- No entanto, para **gerar valor comercial** desses **dados**, as organizações precisam de **novas tecnologias e ferramentas** capazes de **gerenciar e analisar dados não tradicionais** junto com seus dados corporativos tradicionais.

Armazém de Dados (Data Warehouses)

- Um armazém de dados é um banco de dados que armazena os **dados de potencial interesse** para os **tomadores de decisão** de forma **consolidada**.
- Os dados são **oriundos de muitos sistemas**, tais como sistemas para vendas, contas de clientes e manufatura e etc.
- Estes dados são **combinados com dados de fontes externas** e **transformados**, através da **correção de imprecisões e incompletudes** e **reestruturação** dos dados para **relatórios de gestão e análises**.
 - Isto **antes destes dados serem inseridos no armazém de dados**.
- O armazém de dados **controla o acesso** a estes dados, isto é, os disponibiliza para **qualquer pessoa** acessar **conforme necessário**, mas eles **não podem ser alterados**.

Mineração de Dados (Data Mining)

- A mineração de dados fornece insights sobre dados corporativos, encontrando padrões ocultos e relacionamentos em grandes bancos de dados e inferindo regras para prever o futuro comportamento.
- Os padrões e regras são usados para orientar a tomada de decisão e prever o efeito dessas decisões.
- Os tipos de informações que podem ser obtidas dos dados minerados incluem associações, sequências, classificações, clusters e previsão.

Mineração de Dados (Data Mining)

- Associações são ocorrências **relacionadas a um único evento**.
 - Ex.: os quando um determinado produto é comprado, outro relacionado é comprado 65% das vezes.
- Sequencias são eventos **ligados ao longo do tempo**.
 - Ex.: quando uma casa é comprada, em 65% das vezes uma nova geladeira é comprada no intervalo de duas semana.
- A **classificação** reconhece **padrões** que **descrevem o grupo a que um item pertence**.
 - Ex.: uma operadora de telefonia pode prever clientes que irão solicitar a portabilidade analisando o comportamentos dos que solicitaram anteriormente.
- O **agrupamento** funciona de maneira **similar a classificação** quando ainda **não foram definidos grupos**
 - Ex.: encontrar grupos de possíveis clientes para um determinado produto.
- A **previsão** usa uma **série de valores existentes** para **inferir outros valores**.
 - Ex.: encontrar padrões para determinar a quantidade de vendas no futuro.

Mineração de Texto (Text Mining)

- Acredita-se que **dados não estruturados**, a maioria na forma de **arquivos de texto**, respondem por mais de **80% das informações organizacionais úteis** e é uma das **principais fontes de big data** que as empresas desejam analisar.
- E-mails, memorandos, transcrições de call center, respostas a pesquisas e redes sociais, por exemplo, **são valiosos** para **encontrar padrões e tendências** que irão ajudar os gestores a **tomar melhores decisões** de negócios.
- **Ferramentas de mineração de texto** auxiliam as empresas a **analisar esses dados**.
- Por exemplo, um software de **análise de sentimento** é capaz de **minerar comentários de texto** em uma mensagem de **e-mail**, **rede sociais** ou **formulário de pesquisa** para detectar **opiniões favoráveis e desfavoráveis** sobre assuntos específicos.

Mineração da Web (Web Mining)

- A Web é outra fonte de dados não estruturados para revelar padrões, tendências e percepções sobre o comportamento do cliente.
- As empresas podem recorrer à mineração na Web para ajudá-las a compreender o comportamento do cliente, avaliar a eficácia de um determinado Web site, ou quantificar o sucesso de uma campanha de marketing.
 - Por exemplo, os profissionais de marketing usam o Google Trends e o Google Insights para serviços de pesquisa, que rastreiam a popularidade de várias palavras e frases usadas nas pesquisas do Google para saber o que as pessoas estão interessadas e o que são interessado em comprar.
- A mineração da Web procura padrões nos dados por meio da mineração de conteúdo e estrutura de uso.
 - Mineração de conteúdo da web é o processo de extração conhecimento do conteúdo das páginas da Web, que pode incluir texto, imagem, dados de áudio e vídeo.
 - A mineração da estrutura da Web examina os dados relacionados a estrutura de um determinado site.
 - Por exemplo, links apontando para um documento indicam a popularidade do documento, enquanto os links que saem de um documento indicam a riqueza ou talvez a variedade de tópicos cobertos no documento (análise comum em sites como Wikipédia).



FIM

Eduardo Kinder Almentero

ekalmentero@gmail.com