



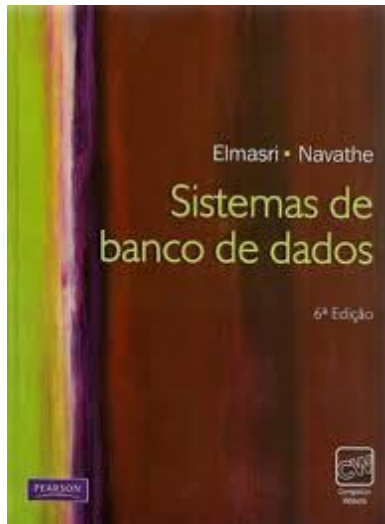
Unidade 23– Conceitos de Mineração de Dados – Parte 1



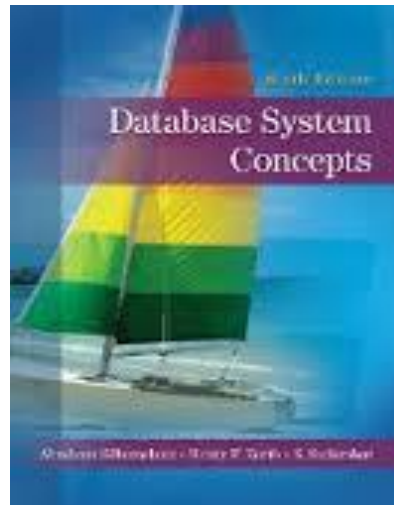
Prof. Aparecido V. de Freitas
Doutor em Engenharia
da Computação pela EPUSP



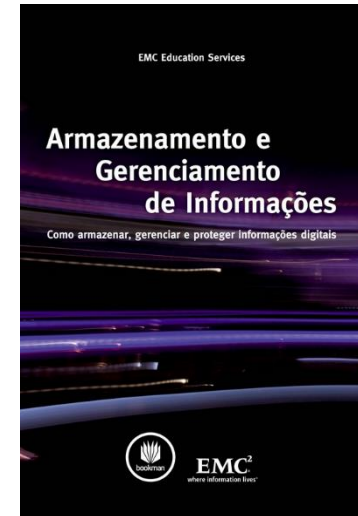
Bibliografia



Sistemas de Banco de Dados
Elmasri / Navathe 6ª edição



Sistema de Banco de Dados
Korth, Silberschatz – Sixth Edition



<http://education.EMC.com/ismbook>



Introdução

- ✓ Nas últimas décadas, muitas organizações têm gerado uma grande quantidade de **dados** na forma de **arquivos** e **banco de dados**;





Organizações dependem dos dados ...

- ✓ Passagens aéreas
- ✓ Sistemas de Telefonia
- ✓ Comércio Eletrônico
- ✓ Sistemas Bancários
- ✓ Montadoras
- ✓ Cartões de Crédito
- ✓ Redes Sociais
- ✓ ...





Qual a dificuldade em processar esses dados com SQL?

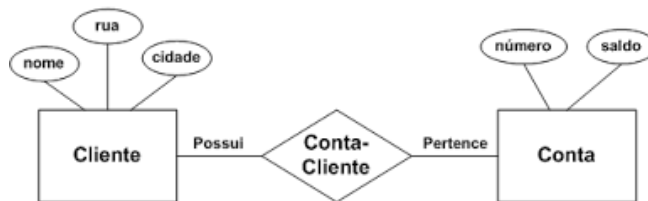




SQL



- ✓ **SQL** é uma linguagem estruturada de consulta, o qual assume que o usuário conhece o esquema do banco de dados;
- ✓ **SQL** dá suporte a operações de Álgebra Relacional que permite que o usuário selecione linhas e colunas de dados das tabelas, assumindo que os dados tenham uma determinada estrutura.





Tipos de Dados

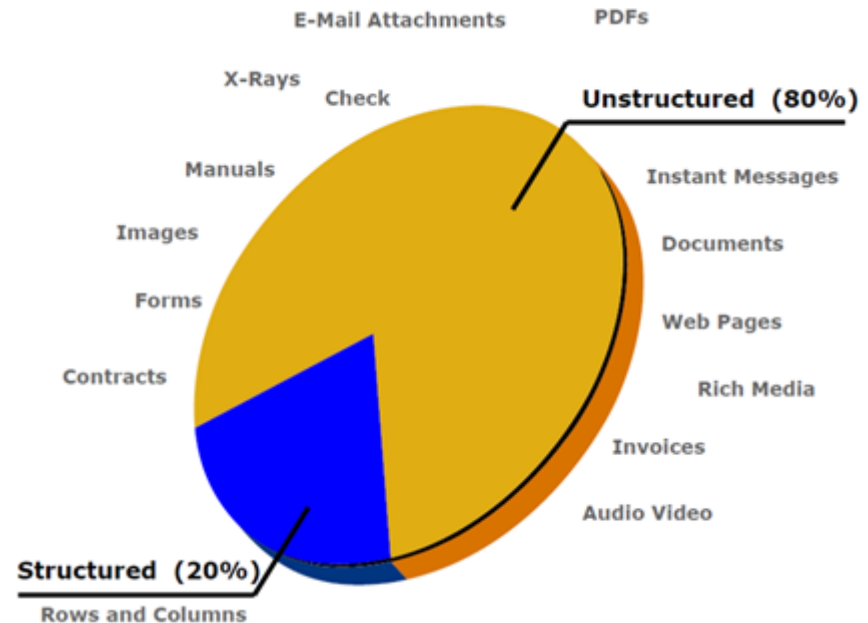
- ✓ **Estruturados:** organizados em linhas e colunas em um formato definido de forma rígida, de modo que aplicativos possam recuperá-los e processá-los com eficiência. (**SGBD**)
- ✓ **Não-estruturados:** Seus elementos não estão organizados na forma de linhas e colunas, sendo, portanto, difíceis de serem consultados e recuperados por aplicativos empresariais. Exemplos: mensagens de email, arquivos .pdf, .doc, etc.





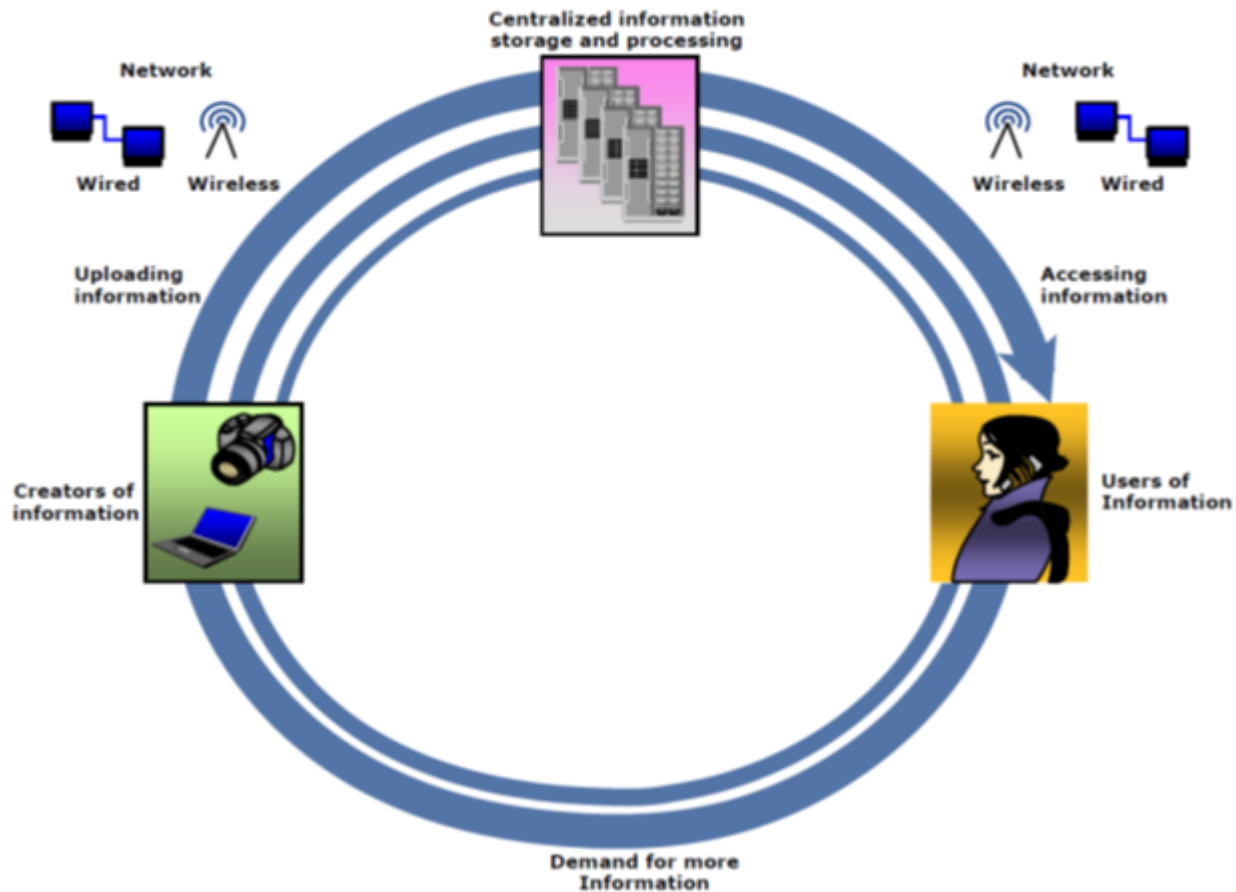
Dados não estruturados

- ✓ Alvo de preocupação das empresas;
- ✓ A maioria dos dados corporativos não são estruturados;
- ✓ Requerem mais espaço e gerenciamento.





Ciclo Virtuoso das Informações





Como então lidar com dados não estruturados?





Mineração de Dados (Data Mining)

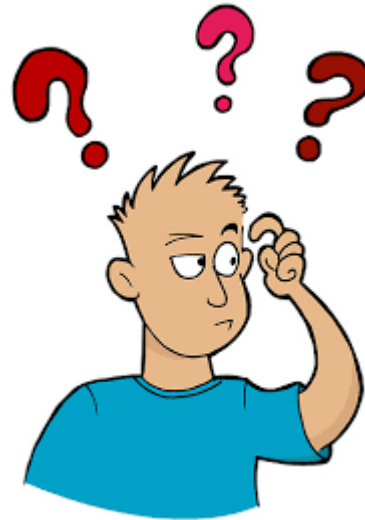


- ✓ Área cujo interesse é atuar na descoberta de informações em termos de **padrões** ou **regras** com base em grandes quantidades de dados;
- ✓ Para ser útil na prática, precisa ser executada de modo eficiente em grandes arquivos e banco de dados;
- ✓ Utiliza técnicas de áreas como **aprendizado de máquina, estatística, redes neurais e algoritmos genéticos**.





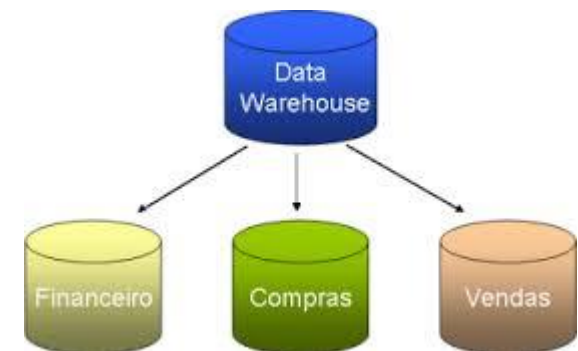
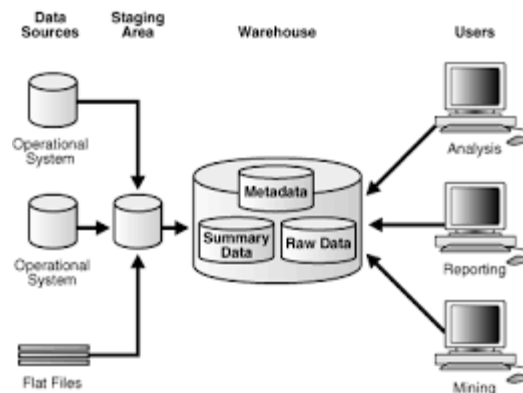
Qual a relação entre Data Warehouse e Data Mining?





Data Warehouse

- ✓ O objetivo de um Data Warehouse (Armazém de Dados) é dar suporte à tomada de decisão com dados;
- ✓ Corresponde a uma base de dados histórica projetada para dar suporte à tomada de decisão;
- ✓ A mineração de dados pode ser usada em conjunto com Data Warehouses para auxiliar nessa tomada de decisão.
- ✓ O uso bem sucedido das aplicações de mineração de dados dependerá, primeiro, da construção de um Data Warehouse.





Descoberta de Conhecimento em Banco de dados (KDD)

⌚ Mineração de dados, na verdade, é um passo de um processo maior conhecido por **KDD** (**knowledge-discovery in databases**);

⌚ O processo de descoberta de conhecimento compreende seis fases:

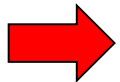
- ✓ Seleção de dados;
- ✓ Limpeza de dados;
- ✓ Enriquecimento;
- ✓ Transformação ou Codificação de dados;
- ✓ Mineração de dados;
- ✓ Relatório e exibição da informação descoberta.





Fases – KDD (Seleção de dados)

- Ⓒ Consideremos um banco de dados transacional mantido por uma empresa de bens de consumo, com a seguinte estrutura: nome do cliente, CEP, telefone, data de compra, código do item, preço, quantidade e quantidade total;
- Ⓒ Na fase de **Seleção** de dados, **dados sobre um item específico podem ser selecionados**, ou clientes de uma determinada de uma determinada região.



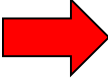
- ✓ **Seleção de dados;**
- ✓ **Limpeza de dados;**
- ✓ **Enriquecimento;**
- ✓ **Transformação ou Codificação de dados;**
- ✓ **Mineração de dados;**
- ✓ **Relatório e exibição da informação descoberta.**

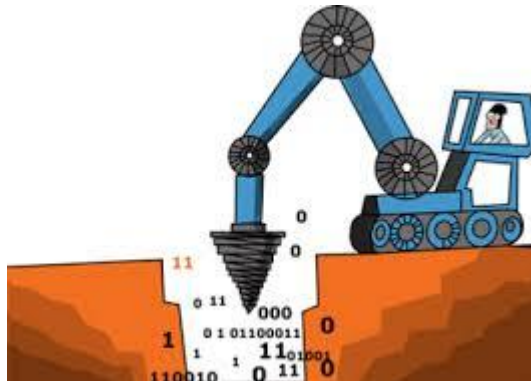




Fases – KDD (Limpeza de dados)

Na fase de **Limpeza** de dados, pode-se corrigir códigos postais inválidos ou eliminar-se registros com prefixos de telefone inválidos.

- 
- ✓ Seleção de dados;
 - ✓ **Limpeza de dados;**
 - ✓ Enriquecimento;
 - ✓ Transformação ou Codificação de dados;
 - ✓ Mineração de dados;
 - ✓ Relatório e exibição da informação descoberta.

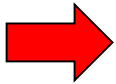




Fases – KDD (Enriquecimento de dados)

Na fase de **Enriquecimento** de dados, pode-se melhorar com fontes de informação adicionais. Por exemplo, dados sobre idade, renda e avaliação de crédito podem ser anexados à cada registro.

- ✓ Seleção de dados;
- ✓ Limpeza de dados;
- ✓ **Enriquecimento;**
- ✓ Transformação ou Codificação de dados;
- ✓ Mineração de dados;
- ✓ Relatório e exibição da informação descoberta.





Fases – KDD (Transformação ou Codificação de dados)

Na fase de **Transformação ou Codificação** de dados, pode-se manipular os dados para deixá-los de forma mais conveniente. Por exemplo, dados sobre renda podem ser divididos em faixas, códigos postais podem ser agregados em regiões, etc

- ✓ Seleção de dados;
- ✓ Limpeza de dados;
- ✓ Enriquecimento;
- ➔ ✓ **Transformação ou Codificação de dados;**
- ✓ Mineração de dados;
- ✓ Relatório e exibição da informação descoberta.





Pré-processamento dos dados

- Ⓢ Somente após o pré-processamento, as técnicas de mineração de dados são usadas para se extrair diferentes padrões e regras.

- ✓ **Seleção de dados;**
 - ✓ **Limpeza de dados;**
 - ✓ **Enriquecimento;**
 - ✓ **Transformação ou Codificação de dados;**
 - ✓ **Mineração de dados;**
 - ✓ **Relatório e exibição da informação descoberta.**
- } **Pré-processamento**





Qual o resultado da mineração de dados?





Resultados da Mineração de Dados

☉ A mineração de dados pode descobrir o seguinte tipo de informação:



- ✓ **Regras de Associação;**
- ✓ **Padrões sequenciais;**
- ✓ **Árvores de Classificação.**

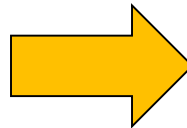




Regras de Associação



- @ Por exemplo, sempre que um cliente compra um item, ele ou ela pode também comprar alguma acessório associado ao item.





Padrões sequenciais



- ⌚ Suponha que um cliente compre uma câmera fotográfica e a financie em 6 meses. Após seis meses ele compra uma lente e também a financie.
- ⌚ Após mais seis meses, provavelmente o cliente comprará um outro acessório para o seu equipamento de fotografia.





Árvores de Classificação



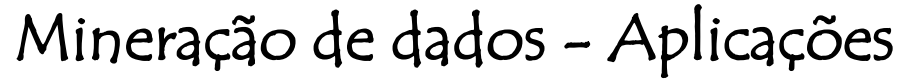
- ② Clientes podem ser classificados por frequência de visitas, tipos de financiamento utilizado, valor da compra, etc;
- ② Pode-se gerar estatísticas interessantes para essas classes de dados;





Que aplicações podem se beneficiar
da mineração de dados?





-





Mineração de dados – Objetivos

De um modo geral, os objetivos são:

- ✓ **Previsão**
- ✓ **Identificação**
- ✓ **Classificação**
- ✓ **Otimização;**





Mineração de dados – Objetivo Previsão (Predição)



@ A mineração de dados pode mostrar como certos atributos dos dados se comportarão no futuro;

@ Exemplos:

- ✓ Quanto de volume de vendas uma determinada loja gerará em um determinado período;
- ✓ Predizer o valor de uma ação três meses adiante;
- ✓ Predizer o percentual que será aumentado de tráfego na rede se a velocidade aumentar;
- ✓ Predizer o vencedor do campeonato baseando-se na comparação das estatísticas dos times.

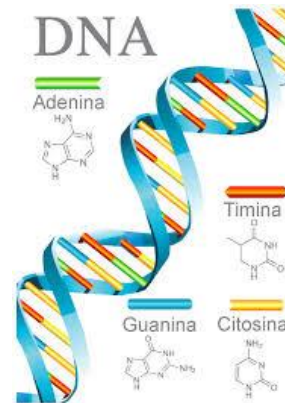




Mineração de dados – Objetivos Identificação



- @ Padrões de dados podem ser usados para se identificar a existência de um item, um evento ou uma atividade;
- @ Exemplos:
 - ✓ Intrusos tentando quebrar um sistema podem ser identificados pelos programas executados, arquivos acessados e tempo de CPU por sessão;
 - ✓ Em aplicações biológicas, a existência de um gene pode ser identificada por certas sequências de **DNA**;





Mineração de dados – Objetivos Classificação



- Ⓜ A mineração de dados pode **particionar** os dados de modo que diferentes **classes** ou **categorias** possam ser identificadas com base em combinações de parâmetros.
- Ⓜ Faz sentido analisar os relacionamentos dentre e entre categorias como problemas separados;
- Ⓜ Essa categorização pode servir para codificar os dados corretamente antes de submetê-los a mais mineração de dados;
- Ⓜ Exemplo:
 - ✓ Clientes em supermercados podem ser categorizados em compradores que buscam **desconto**, compradores **fiéis** a uma determina **marca** ou ainda compradores **eventuais**;

I AM NIKON

Panasonic

Canon

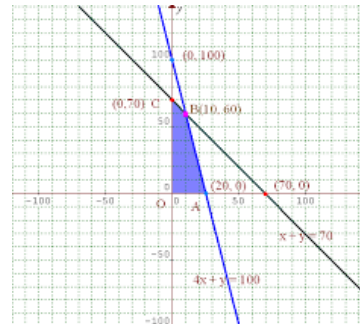




Mineração de dados – Objetivos Otimização



- Um objetivo relevante da Mineração de Dados pode ser **otimizar** o uso de recursos limitados, como tempo, espaço, dinheiro ou materiais e maximizar variáveis de saída como vendas ou lucros sob determinado conjunto de restrições;
- Como tal, esse objetivo da Mineração de Dados é semelhante aos problemas de **Pesquisa Operacional**;
- A **Pesquisa Operacional** é um ramo da Matemática Aplicada que faz uso de modelos matemáticos, estatísticos e algoritmos para ajuda à tomada de decisão. Trata da aplicação da ciência à solução de problemas gerenciais e administrativos.





Tipos de Conhecimentos descobertos pela Mineração de Dados

- ⌚ A mineração de dados enfoca o Conhecimento Indutivo, que descobre novas regras e padrões com base nos **dados fornecidos**;
- ⌚ É comum descrever-se o conhecimento descoberto durante a Mineração de Dados por:
 - ✓ **Regras de Associação;**
 - ✓ **Hierarquias de Classificação;**
 - ✓ **Padrões Sequenciais;**
 - ✓ **Padrões dentro de séries temporais;**
 - ✓ **Agrupamento.**