

# Mineração de Dados

Banco de Dados

# Mineração de Dados - Motivação

---

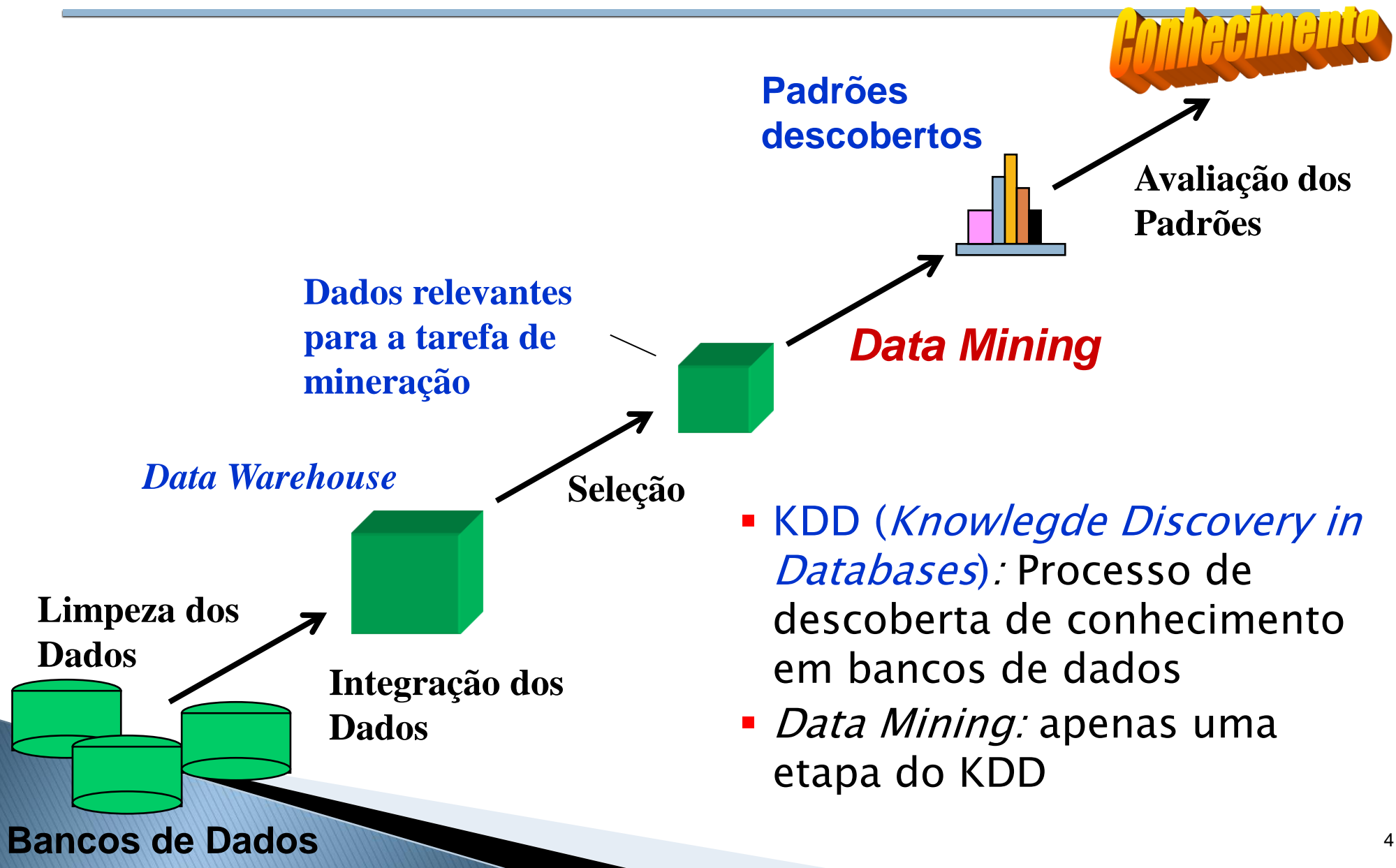
- ▶ Crescimento rápido do tamanho e do número de bancos de dados disponíveis em:
  - Aplicações comerciais;
  - Indústria;
  - Internet, etc.
- ▶ Muitos dados x pouca informação
- ▶ A mineração de dados surge como uma ferramenta poderosa para suprir a **necessidade de descobrir informações úteis**, tais como **padrões escondidos** em grandes bancos de dados.

# Mineração de Dados

---

- ▶ Corresponde em aplicar a uma **base de dados adequadamente preparada, algoritmos e técnicas para descoberta de padrões** interessantes nesses dados.
- ▶ É apenas uma etapa de um processo maior, conhecido como **processo de descoberta de conhecimentos** em bancos de dados.
  - KDD – Knowledge Discovery in Databases

# Data Mining: uma etapa do processo KDD



# Mineração de Dados

---

- ▶ Tarefas de Mineração de Dados:
  - Mineração de Regras de Associação
  - Mineração de Padrões Sequenciais
  - Classificação
  - Agrupamento
  - Detecção de Anomalias (*Outliers*)

# Mineração de Regras de Associação

- ▶ Problema de minerar regras de associação:
  - Dado um conjunto de registros, cada um contendo um número qualquer de itens,
  - Encontrar regras de dependência que podem ser usadas para prever a ocorrência de um item baseada na ocorrência de outros.

Transação	Itens Comprados
1	Leite, pão, coca
2	Cerveja, pão
3	Coca, leite, fralda, cerveja
4	Pão, Leite, fralda, cerveja
5	Leite, pão, coca, fralda, cerveja
6	Leite, suco, iogurte, cerveja

**Algumas regras que poderiam ser Descobertas pelo software de mineração**

**{Pao} → {Leite}**

**{Fralda, Leite} → {Cerveja}**

# Mineração de Regras de Associação

- ▶ Uma regra é considerada interessante quando a mesma possui **suporte** e **confiança** superiores aos valores mínimos fornecidos pelo usuário.
  - **Suporte de um item** (ou conjunto de itens): número de transações em que o item aparece.
  - **Suporte de uma regra  $A \rightarrow B$** : número de transações em que aparecem os itens  $A \cup B$ .
  - **Confiança de uma regra  $A \rightarrow B$** :  $\text{sup}(A \cup B) / \text{sup}(A)$
- ▶ **Exemplo**: se o banco de dados abaixo fosse passado para o software de mineração e o usuário indicasse um suporte mínimo = 3 e confiança = 70%, pelo menos a regra  $\{\text{pão}\} \rightarrow \{\text{leite}\}$  seria descoberta pelo software de mineração.

Transação	Itens Comprados
1	Leite, pão, coca
2	Cerveja, pão
3	Coca, leite, fralda, cerveja
4	Pão, Leite, fralda, cerveja
5	Leite, pão, coca, fralda, cerveja
6	Leite, suco, iogurte, cerveja

**Sup (Pão) = 4**  
**Sup(Leite) = 5**  
**Sup ( $\{\text{Pão}\} \rightarrow \{\text{Leite}\}$ ) = 3**

**Conf ( $\{\text{Pão}\} \rightarrow \{\text{Leite}\}$ ) =  $3/4 = 75\%$**   
**Conf( $\{\text{Leite}\} \rightarrow \{\text{Pão}\}$ ) =  $3/5 = 60\%$**

# Mineração de Regras de Associação – Aplicação

---

## ► Marketing e Promoção de Vendas

- Considere que a seguinte regra tenha sido descoberta:

*{Refrigerante, ...} → {Batata Frita}*

- **Batata Frita como conseqüente**: pode ajudar a determinar o que pode ser feito para aumentar as vendas de batata frita.
- **Refrigerante como antecedente**: pode ser usado para identificar quais produtos seriam afetados caso a loja deixasse de vender refrigerante.



# Mineração de Regras de Associação

---

- ▶ Como aprimorar o negócio a partir da regra descoberta  $\{p\tilde{a}o\} \rightarrow \{leite\}$  ?
- ▶ **Estratégia 1: colocar leite e pão próximos para encorajar a compra conjunta** desses produtos quando o cliente visitar a loja.



# Mineração de Regras de Associação

---

- ▶ Como aprimorar o negócio a partir da regra descoberta  $\{p\tilde{a}o\} \rightarrow \{leite\}$  ?
- ▶ **Estratégia 2:** colocar **leite e pão em lados opostos** da loja pode levar o consumidor de tais produtos a **comprar também outros itens** no caminho.



# Mineração de Regras de Associação

---

- ▶ Como aprimorar o negócio a partir da regra descoberta  $\{\text{pão}\} \rightarrow \{\text{leite}\}$  ?
- ▶ **Estratégia 3**: fazer uma **promoção dos itens** para venda conjunta.



Leve duas unidades  
do Leite X

+



E ganhe 50% de desconto  
no pão francês

# Mineração de Padrões Seqüenciais

---

- ▶ Tem como objetivo identificar **sequências de eventos** (ou objetos) que ocorrem **frequentemente em bancos de dados temporais**.
  - Descobrir a **evolução de sintomas** apresentados em **pacientes**.
  - Descobrir a **evolução das compras** realizadas por **clientes**.
  - Descobrir **caminhos (sequências de páginas acessadas)** na Web frequentemente **percorridos pelos usuários**.

# Mineração de Padrões Seqüenciais

## Banco de Dados de Seq. de Transações

Cliente	Seqüências de transações
001	< {TV}, {DVD} >
002	< {computador, câmera, CD} >
003	< {TV, VCR}, {CD}, {DVD} >
004	< {livro, relógio, computador} >

*Clientes que compram  
TV, posteriormente  
compram DVD-player.*



## Padrão Seqüencial

< {TV} , {DVD} >

- ❑ O Padrão seqüencial acima seria descoberto se o usuário informasse o banco de dados à esquerda e por exemplo um **suporte mínimo de 50%**.
- ❑ Assim, o mecanismo de mineração deve **encontrar todos os padrões seqüenciais**, de todos os tamanhos possíveis, que apareçam **em mais de 50% das seqüências de transações**.

# Padrões Seqüenciais - Aplicações

---

## ▶ Marketing Direcionado

- Descobrindo um padrão seqüencial, como

$\langle \{\text{TV}\}, \{\text{DVD}\} \rangle$

a empresa pode lançar uma **propaganda do produto DVD direcionada** àqueles **clientes que já compraram o produto TV**, sabendo que haverá **maior probabilidade de efetividade**.

- O padrão  $\langle \{\text{TV}\}, \{\text{DVD}\} \rangle$  é um exemplo comum, mas o mecanismo de mineração pode encontrar muitos padrões importantes, não óbvios como este citado.

# Projeto Final

---

- ▶ Aplique Regras de Mineração de Banco de Dados no Projeto Final
  - Busque dados reais para o banco de dados. Caso não seja possível, procure uma massa de dados mais próxima do real possível.
  - Mostre os resultados da mineração e mostre como o grupo deseja investir nestes resultados.