

# INTEGRAÇÃO DE DADOS

---

Disciplina: SCC0542 - Tópicos Especiais em Banco de Dados

Estagiária PAE: Dayse Silveira de Almeida

Professora: Dr<sup>a</sup>. Cristina D. A. Ciferri



# Roteiro

- Fundamentação teórica;
  - Integração de esquemas
  - Integração de instâncias
- Exemplos:
  - Supermercado
  - Publicações
  - Viagens

# Fundamentação Teórica

## Integração de Dados

- Problema enfrentado por aplicações que precisam acessar várias fontes de dados
  - Autônomas e heterogêneas
  - Criadas independentemente
- **Compartilhamento de dados**
- Combinação de Bancos de Dados

# Fundamentação Teórica

## Esquemas e Instâncias

- Descrição do Banco de Dados x o próprio Banco de dados
  - Esquema – descrição especificada durante o projeto
    - Descrição dos dados, relacionamentos, semântica, restrições
  - Instância – conjunto de dados no Banco de Dados em determinado momento
    - Cada entidade tem seu próprio conjunto de instâncias atuais

# Fundamentação Teórica

## Integração de Esquemas

- Surge devido a representações não idênticas do mesmo conceito
  - A existência de conflitos entre representações é devido ao fato de que diferentes usuários modelam um mesmo mini mundo de maneiras distintas, de acordo com as suas percepções
    - **Relativismo semântico**

# Fundamentação Teórica

## Integração de Esquemas

- Integração depende da identificação de similaridades e diferenças entre os elementos dos diferentes esquemas
- Os conflitos podem ser divididos em grupos:
  - **Conflitos de nome** - nomes utilizados para representar os diferentes elementos existentes nos esquemas a serem integrados
    - Problema dos sinônimos - diferentes nomes podem ser aplicados ao mesmo elemento
      - Ex.: Nome **cliente** utilizado em um esquema e **comprador** em outro, para representar todos os clientes atendidos por uma loja
    - Problema dos homônimos - mesmo nome pode ser aplicado a diferentes elementos
      - Ex.: Nome **nome** utilizado para representar o nome de um empregado em um esquema **Empregado**, e também aplicado ao esquema **Departamento**

# Fundamentação Teórica

## Integração de Esquemas

- Os conflitos podem ser divididos em grupos:
  - **Conflitos semânticos** - mesmo elemento é modelado nos diferentes esquemas, porém representando conjuntos que se sobrepõem
    - Ou, o conjunto de instâncias do elemento de um esquema é mais abrangente do que o conjunto de instâncias do elemento do outro esquema
    - Ou, em um esquema, uma entidade pode ser tratada de forma mais genérica e em outra, mais específica
      - Ex. 1:
        - Uma fonte utiliza uma classe **estudante** para representar todos os alunos de uma universidade
        - Outra fonte usa uma classe **estudante\_grad** representando apenas os alunos de graduação da universidade
      - Ex. 2: Uma fonte pode tratar entidades **pessoa** de forma genérica, enquanto que outra fonte pode considerar diferentes tipos de pessoa, como **estudante** e **docente**

# Fundamentação Teórica

## Integração de Esquemas

- Os conflitos podem ser divididos em grupos:
  - **Conflitos estruturais** – diferentes estruturas são utilizadas para modelar o mesmo conceito representado em diferentes fontes
    - Ex. 1: conflito de tipos de dados
      - Data pode ser armazenada em uma fonte no formato **mês/ano**, e em outra **ano**
    - Ex. 2:
      - Sexo pode ser armazenado em uma fonte no formato *string* **feminino** ou **masculino**
      - Em uma segunda fonte como caractere **F** ou **M**
      - Em uma terceira fonte, como valor booleano **0** ou **1**

# Fundamentação Teórica

## Integração de Instâncias

- Ambiguidade – quais entidades de fontes distintas são similares
  - Entidades do mundo real são identificadas por valores de seus atributos, e não por identificadores únicos
    - Ex.1:
      - $e1 = (\{\text{"Distributed query processing in a relational database system"}\}, \{\text{"169-180"}\}, \{\text{"Robert S. Epstein"}; \text{"Michael Stonebraker"}; \text{"Eugene Wong"}\})$ , e
      - $e2 = (\{\text{"Distrited query processing in a relational database system"}\}, \{\text{"169-179"}\}, \{\text{"Epstein, R.S."}; \text{"Stonebraker, M."}; \text{"Wong, E."}\})$
    - Ex. 2:
      - $e1 = (\{\text{"Feminino"}\})$  e;
      - $e2 = (\{\text{"F"}\})$  e;
      - $e3 = (\{1\})$ .

# Fundamentação Teórica

## Integração de Instâncias

- Conflito de valores de atributos – diferentes fontes podem possuir valores conflitantes sobre a mesma entidade ou atributo
  - Ex. 1: o valor do atributo {"**Distributed** query processing in a relational database system"} de e1 e {"**Distrited** query processing in a relational database system"} de e2
  - Ex. 2: o valor do atributo {"Feminino"} de e1, {"F"} de e2 e, {"1"} de e3

## Exemplos

# Supermercado

1. Cada vendedor/distribuidor leva um relatório dos produtos entregues e quantidade de itens. Objetivo: fazer um relatório geral para controle de estoque.

### Produto

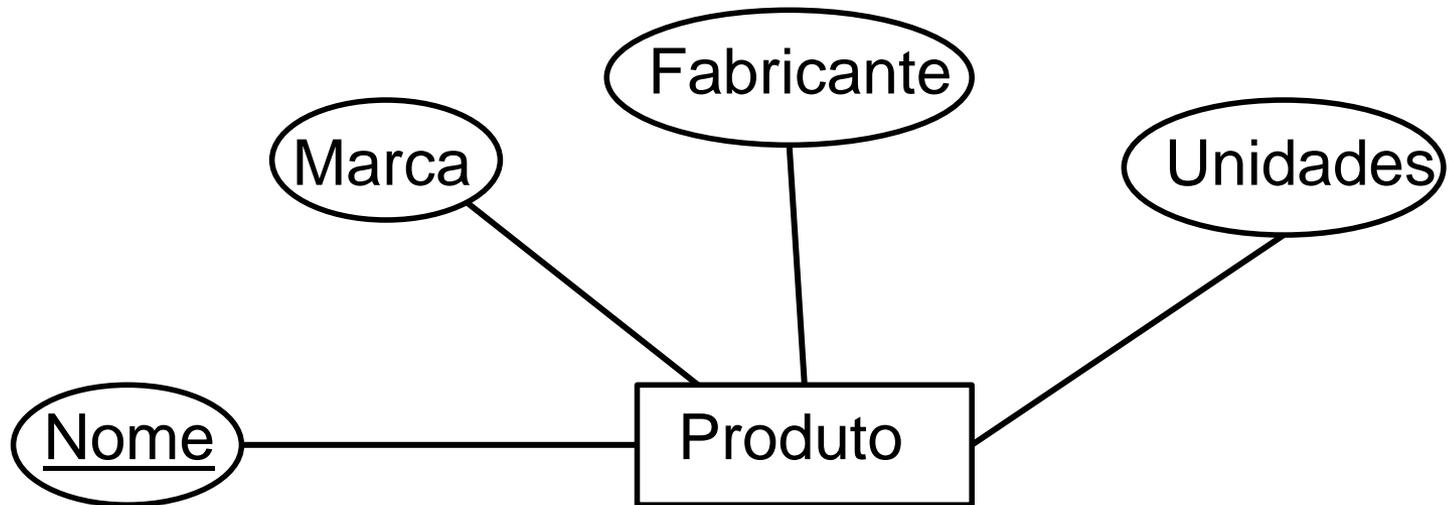
Nome	Marca	Fabricante	Unidades
Arroz	Cristal	Cristal Alimentos LTDA.	1200
Massa de tomate	Elefante	Knorr	250
Fralda	Pampers	P&G	700

- Integração de esquemas:
  - **NomeProduto** x **Nome**, **Unidades** x **Quantidade**
    - Conflito de nomes - sinônimos
  - **Fabricante** x **Null** já que é representante daquele fabricante
    - Conflito estrutural – nome de atributo x entidade
- Integração de instâncias:
  - e1 = {"massa de tomate"}, e2 = {"molho de tomate"}, e3 = {"extrato de tomate"}.

# Exemplos

## Supermercado

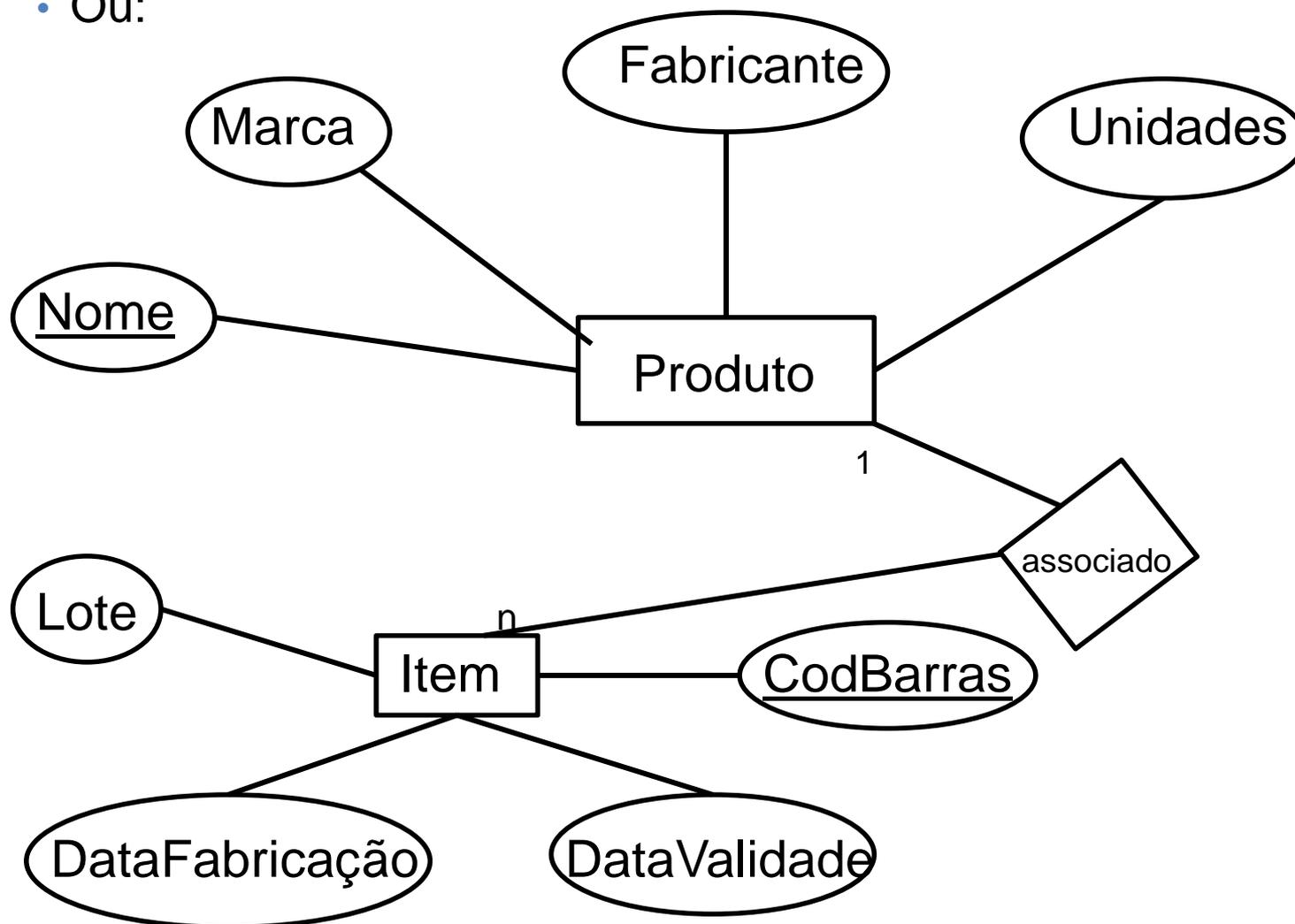
- Modelo E-R:



# Exemplos

## Supermercado

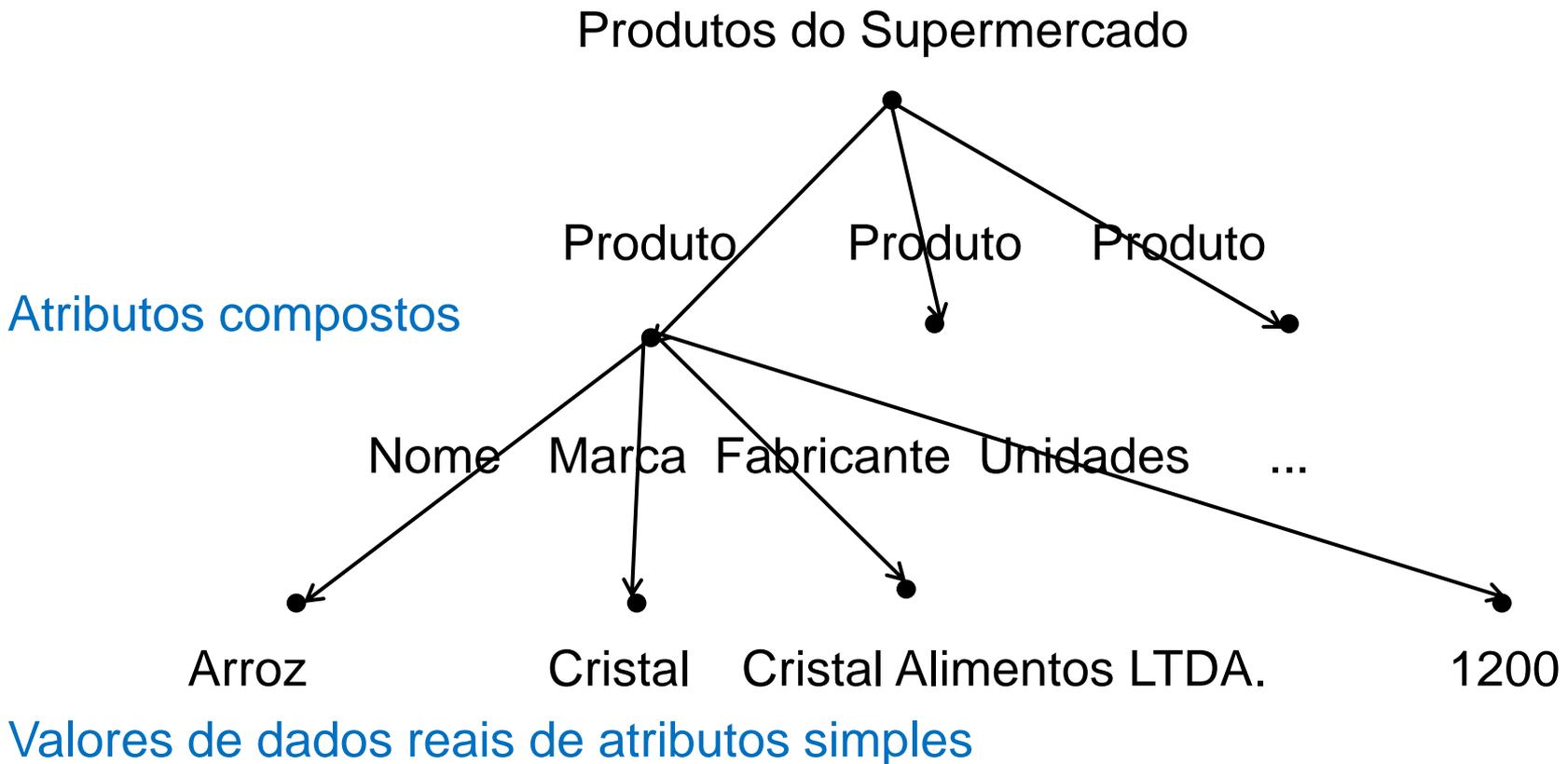
- Ou:



# Exemplos

## Supermercado

- Representando dados semiestruturados como um grafo:



# Exemplos

## Supermercado

- Modelos de dados hierárquico em XML:

```
<?xml version = "1.0" standalone = "yes">
```

```
<Produtos>
```

```
<Produto>
```

```
<Nome>Arroz</Nome>
```

```
<Marca>Cristal</Marca>
```

```
<Fabricante>Cristal Alimentos LTDA.</Fabricante>
```

```
<Unidades>1200</Unidades>
```

```
</Produto>
```

```
<Produto>
```

```
<Nome>Massa de tomate</Nome>
```

```
<Marca>Elefante</Marca>
```

```
<Fabricante>Knorr</Fabricante>
```

```
<Unidades>250</Unidades>
```

```
</Produto>
```

```
<Produto>
```

```
<Nome>Fralda</Nome>
```

```
<Marca>Pampers</Marca>
```

```
<Fabricante>P&G</Fabricante>
```

```
<Unidades>700</Unidades>
```

```
</Produto>
```

```
</Produtos>
```

Sem esquema

# Exemplos

## Publicações de Livros

- Publicação

Natureza	Título	Edição	Editora	DataPublicacao	Autores
Literatura	Iracema	1ª	L&PM	2002	José de Alencar
Sistemas Distribuídos	Sistemas Distribuídos: Princípios e Paradigmas	2ª	Prentice Hall	2007	Andrew Tanenbaum
Redes de Computadores	Redes de Computadores e a Internet: uma abordagem top-down	3ª	Pearson	2006	James F. Kurose, Keith W. Ross
Ficção	O Guia do Mochileiro das Galáxias	1ª	Sextant e	2010	Douglas Adams

# Exemplos

## Publicações de Livros

- Integração de esquemas:
  - *DataPublicação* – **ano** x **mês/ano**
    - Conflito estrutural - conflito de tipos de dados
  - ***Autores*** x ***Autor***
    - Conflito de nomes – sinônimos
  - ***Autor*** como nome de atributo de um tipo de entidade (**Publicação**) x ***autor*** como entidade, contendo **nome** e **contato**
    - Conflito estrutural

# Exemplos

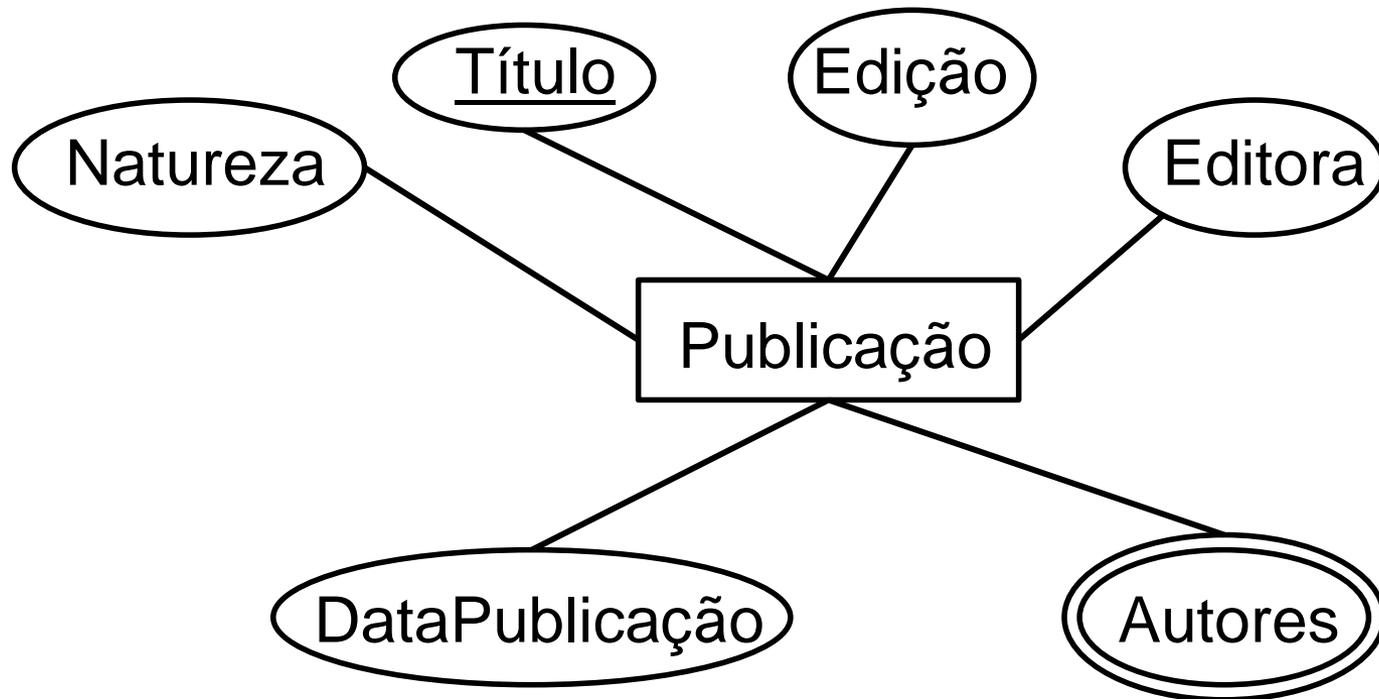
## Publicações de Livros

- Integração de instâncias:
  - Ambiguidade –
    - $e_1 = (\{\text{"Redes de Computadores e a Internet: uma abordagem top-down"}\}, \{\text{"Kurose, James F."}; \text{"Ross, Keith W."}\}, \{2006\})$ , e
    - $e_2 = (\{\text{"Redes de Computadores e a Internet: uma abordagem top-down"}\}, \{\text{"Kurose, J. F."}; \text{"Ross, K. W."}\}, \{2005\})$
  - Conflito de valores de atributos –
    - $e_1 = \{\text{"James F. Kurose"}; \text{"Keith W. Ross"}\}$ ,  $e_2 = \{\text{"Kurose, J. F."}; \text{"Ross", "K. W."}\}$ ,  $e_3 = \{\text{"Kurose, James F."}; \text{"Ross, Keith W."}\}$
    - $e_1 = \{2006\}$  e  $e_2 = \{2005\}$

# Exemplos

## Publicações de Livros

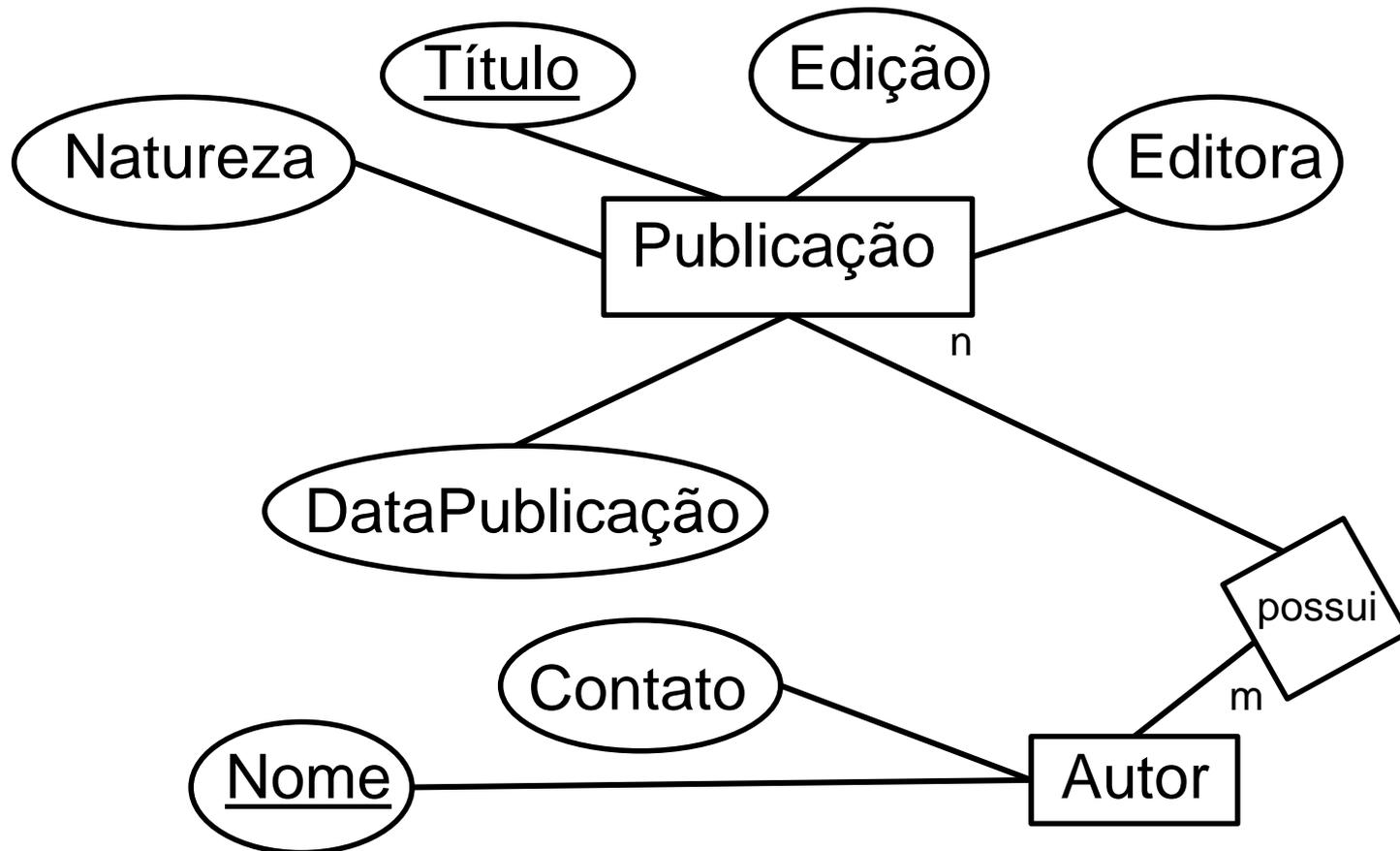
- Modelo E-R:



# Exemplos

## Publicações de Livros

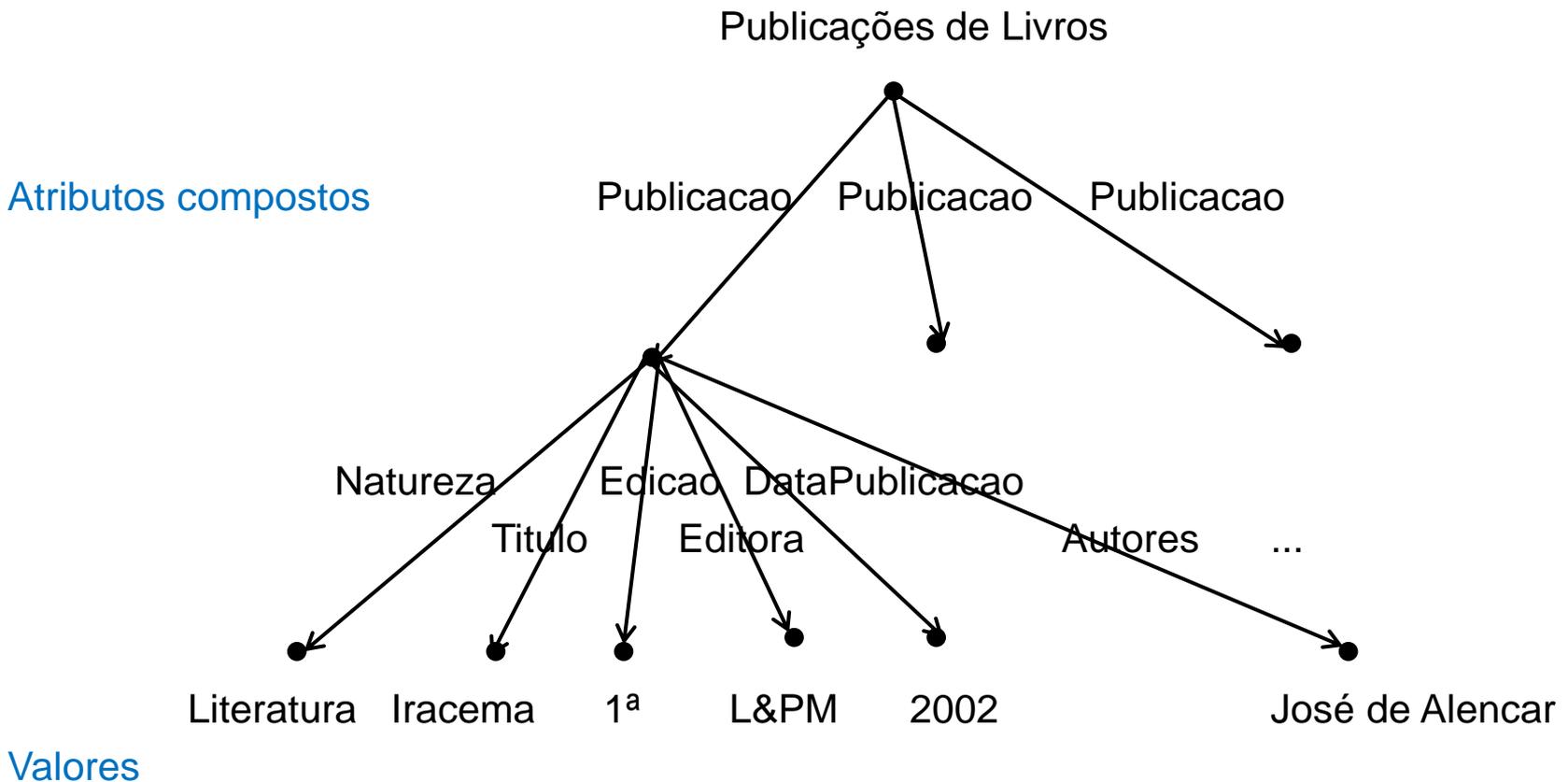
- Ou:



## Exemplos

# Publicações de Livros

- Representando dados semiestruturados como um grafo:



# Exemplos

## Publicações de Livros

- Modelos de dados hierárquico em XML:

```
<?xml version = "1.0" standalone = "yes">
```

Sem esquema

```
<Publicações>
```

```
<Publicacao>
```

```
<Natureza>Literatura</Natureza>
```

```
<Titulo>Iracema</Titulo>
```

```
<Edicao>1ª</Edicao>
```

```
<Editora>L&PM</Editora>
```

```
<DataPublicacao>2002</DataPublicacao>
```

```
<Autores>José de Alencar</Autores>
```

```
</Publicacao>
```

```
<Publicacao>
```

```
<Natureza>Sistemas Distribuídos</Natureza>
```

```
<Titulo>Sistemas Distribuídos: Princípios e Paradigmas</Titulo>
```

```
<Edicao>2ª</Edicao>
```

```
<Editora>Prentice Hall</Editora>
```

```
<DataPublicacao>2007</DataPublicacao>
```

```
<Autores>Andrew Tanenbaum</Autores>
```

```
</Publicacao>
```

```
...
```

```
</Publicacoes>
```

# Exemplos

## Relatório de Vôos

### Passageiro

Nome	RG	CPF	Endereço	Telefone
------	----	-----	----------	----------

### Voo

Numero	Data	HoraSaida	HoraChegada	Origem	Destino
--------	------	-----------	-------------	--------	---------

### Aviao

Modelo	Fabricante	Capacidade
--------	------------	------------

### Piloto

Patente	HorasVoo
---------	----------

### Empregado

Nome	DataNasc	CPF	Salario	Endereco	Telefone
------	----------	-----	---------	----------	----------

# Exemplos

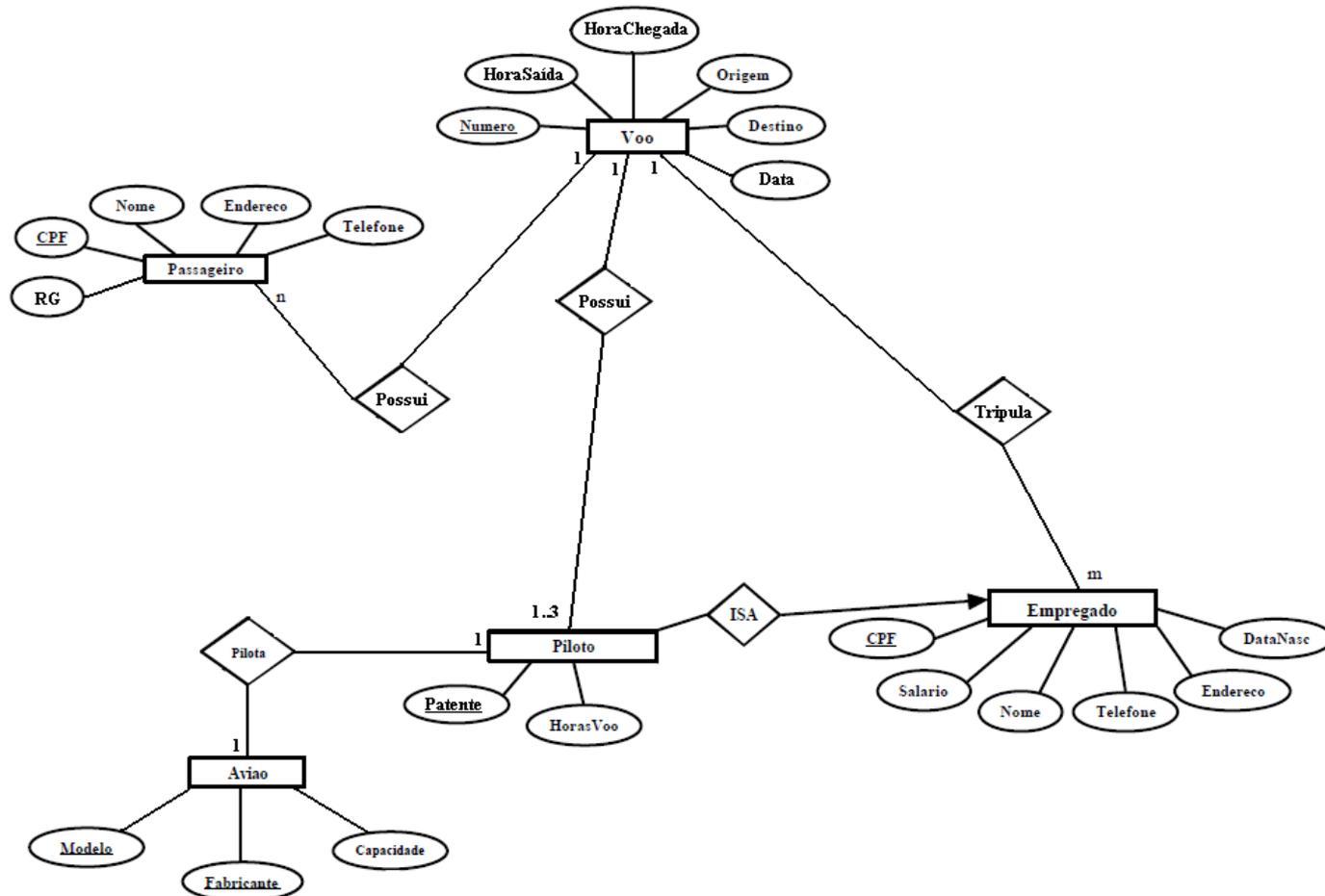
## Relatório de Vôos

- Integração de esquemas:
  - **HoraPart** x **HoraDePartida** x **HoraSaida**
    - Conflito de nomes - sinônimos
  - **Hora** como hora de saída e **Hora** como hora de chegada
    - Conflito de nomes – homônimos
  - **Nome** (entidade Passageiro x entidade Empregado)
  - **CPF** (entidade Passageiro x Empregado)
  - **Endereço** (Passageiro, x Empregado)
  - **Telefone** (Passageiro, x Empregado)
    - Conflito de nomes – homônimos
  - **Data** – **ano** x **mês/ano** x **dia/mês/ano**
    - Conflito estrutural - conflito de tipos de dados
  - **Passageiro** x **Passageiro1** para passageiros de 1ª classe, **PassageiroE** para passageiros de classe econômica
    - Conflito semântico

# Exemplos

## Relatório de Vôos

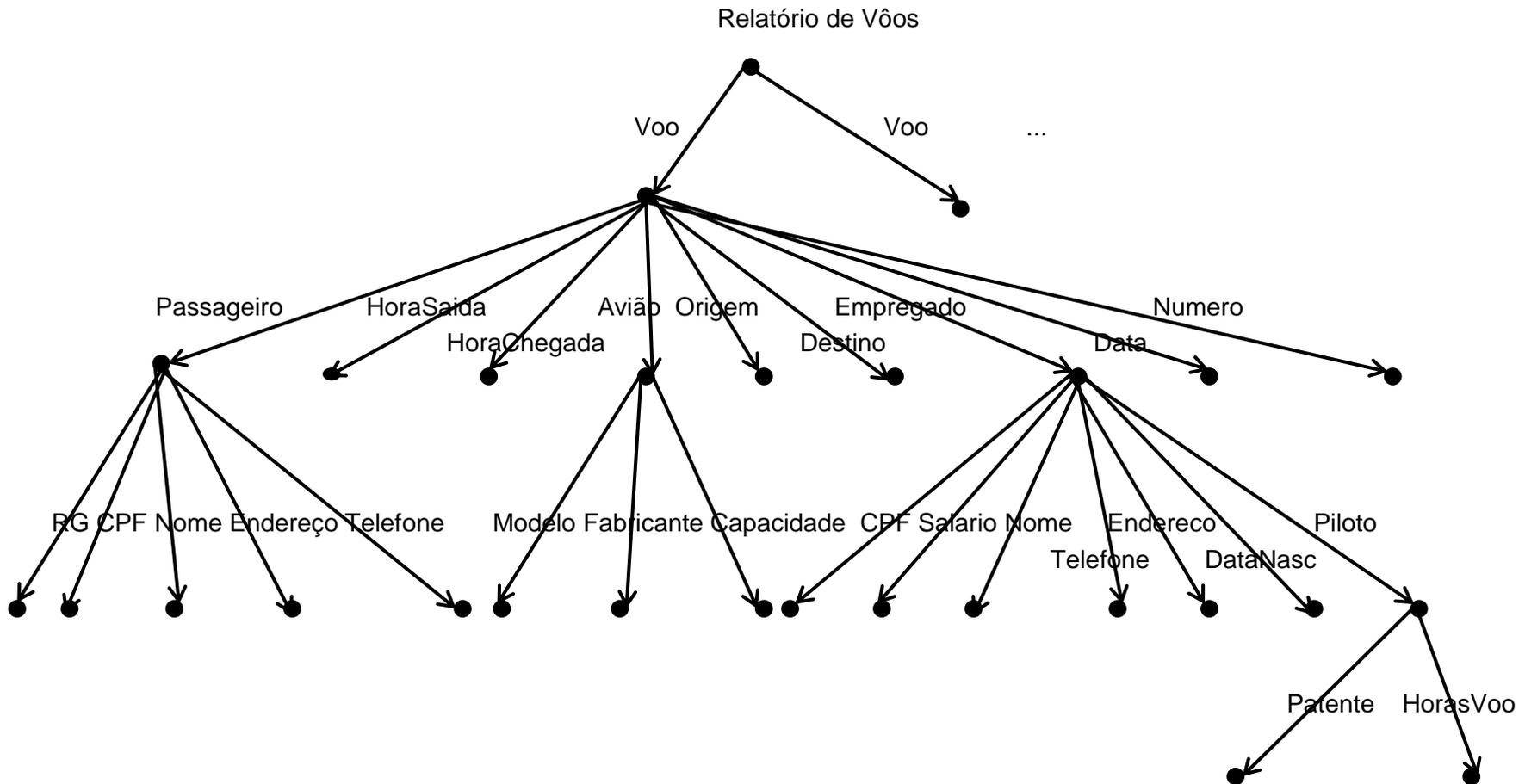
- Modelo E-R:



# Exemplos

## Relatório de Vôos

- Representando dados semiestruturados como um grafo:



# Exemplos

## Relatório de Vôos

- Modelos de dados hierárquico em XML:

```
<?xml version = "1.0" standalone = "yes">
```

```
<Vôos>
```

```
<Voo>
```

```
<Numero></Numero>
```

```
<HoraSaida></HoraSaida>
```

```
<HoraChegada></ HoraChegada >
```

```
<Origem></Origem>
```

```
<Destino></Destino>
```

```
<Data></Data>
```

```
<Passageiro>
```

```
<Nome></Nome>
```

```
<RG></RG>
```

```
<CPF></CPF>
```

```
<Endereco></Endereco>
```

```
<Telefone></Telefone>
```

```
</Passageiro>
```

```
<Passageiro>
  <Nome></Nome>
  <RG></RG>
  <CPF></CPF>
  <Endereco></Endereco>
  <Telefone></Telefone>
</Passageiro>
...
<Aviao>
  <Modelo></Modelo >
  <Fabricante></Fabricante >
  <Capacidade></Capacidade >
</Aviao>
<Empregado>
  <Nome></Nome>
  <Endereco></Endereco>
  <Telefone></Telefone>
  <DataNasc></DataNasc>
  <CPF></CPF>
  <Salario></Salario>
  <Piloto>
    <Patente></Patente>
    <HorasVoo></HorasVoo>
  </Piloto>
</Empregado>
```

```
<Empregado>
  <Nome></Nome>
  <Endereco></Endereco>
  <Telefone></Telefone>
  <DataNasc></DataNasc>
  <CPF></CPF>
  <Salario></Salario>
  <Piloto>
    <Patente></Patente>
    <HorasVoo></HorasVoo>
  </Piloto>
</Empregado>
<Empregado>
  <Nome></Nome>
  <Endereco></Endereco>
  <Telefone></Telefone>
  <DataNasc></DataNasc>
  <CPF></CPF>
  <Salario></Salario>
</Empregado>
...
</Voo>
<Voo>
....
</Voo>
...
</Vôos>
```