

Bases de Dados

Módulo 14A1: Normalização

Prof. André Bruno de Oliveira

14/05/24 08:21

Tópicos

- **Introdução**
- **Conceitos Preliminares Atomicidade de Valores**
 - Superchave e chave
 - Dependência Funcional
 - Anomalias

Normalização

- Para este estudo vamos usar o “minimundo” do *IMDb ENCE descrito da seguinte forma:
- BD que oferece as seguintes informações sobre filmes produzidos em diferentes épocas:

título;

ano;

resumo;

elenco;

gêneros;

país;

duração;

avaliação;



Respondendo ao IBGE 26/2/2019



Institucional IBGE 3/7/2018



Me chamo IBGE 1/6/2016



IBGE 80 anos 3/12/2015



O que é Bioma e quais são os 6 Biomas brasileiros? • IBGE Explica
7/2/2023



Como saber os dados mais atuais da economia brasileira? • IBGE Explica
18/4/2022

- * Site sobre Internet Movie Database (IMDb) - é, como o nome sugere, uma base de dados sobre todo tipo de produção audiovisual.

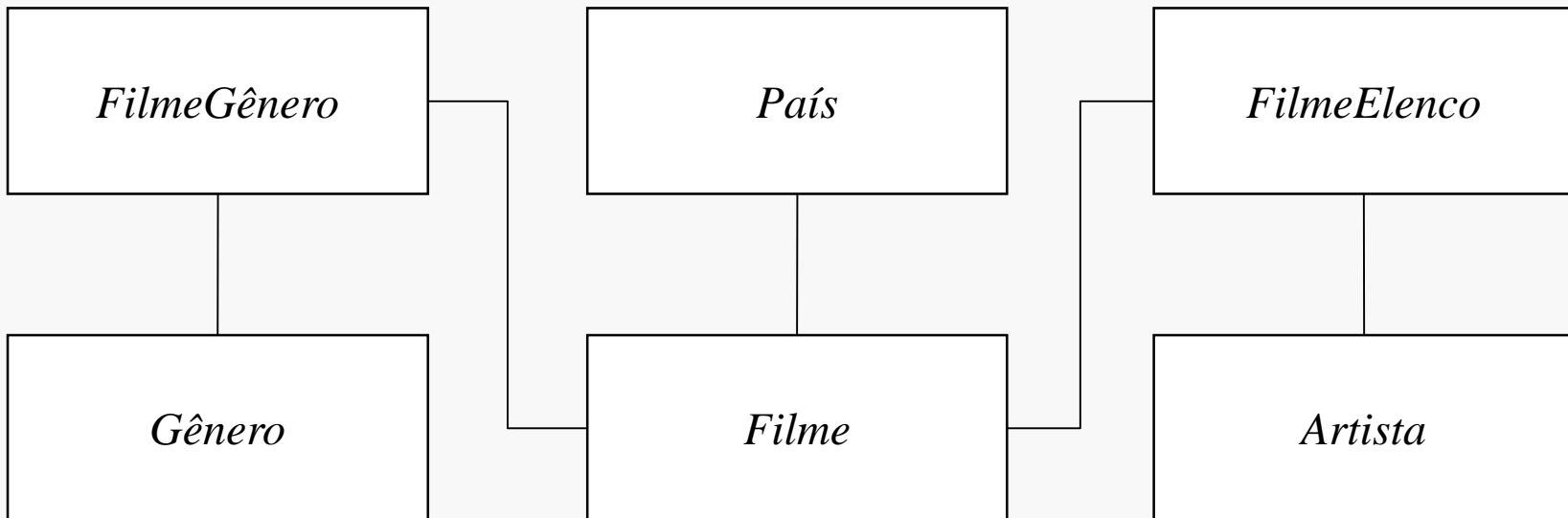
Introdução

- O BD poderia ter sido definido com apenas uma relação:
- $F(\text{título}, \text{ano}, \text{resumo}, \text{elenco}(\text{nome}, \text{sexo}), \text{gêneros}, \text{país}(\text{nome}, \text{sigla}), \text{duração}, \text{avaliação})$

título	ano	resumo	elenco	gênero	país	duração	avaliação
O discreto charme da burguesia	1972	A surreal story centered around six middle-class people and their consistently interrupted attempts to have a meal together	Fernando Rey (M), Delphine Seyring (F)	Comédia	França(FK)	102	8,0
Ensina-me a Viver	1971	Young, rich, and obsessed with death, Harold finds himself changed forever when he meets lively septuagenarian Maude.	Ruth Gordon (F), Bud Cort (M)	Comédia, Drama, Romance	Estados Unidos US)	95	8,0
Estado de Sítio	1972	In Uruguay in the early 1970s, an official of the US is kidnapped by a group of urban guerillas	Yves Montand (M)	Drama, Thriller	França(FK)	102	7,9
...

Introdução

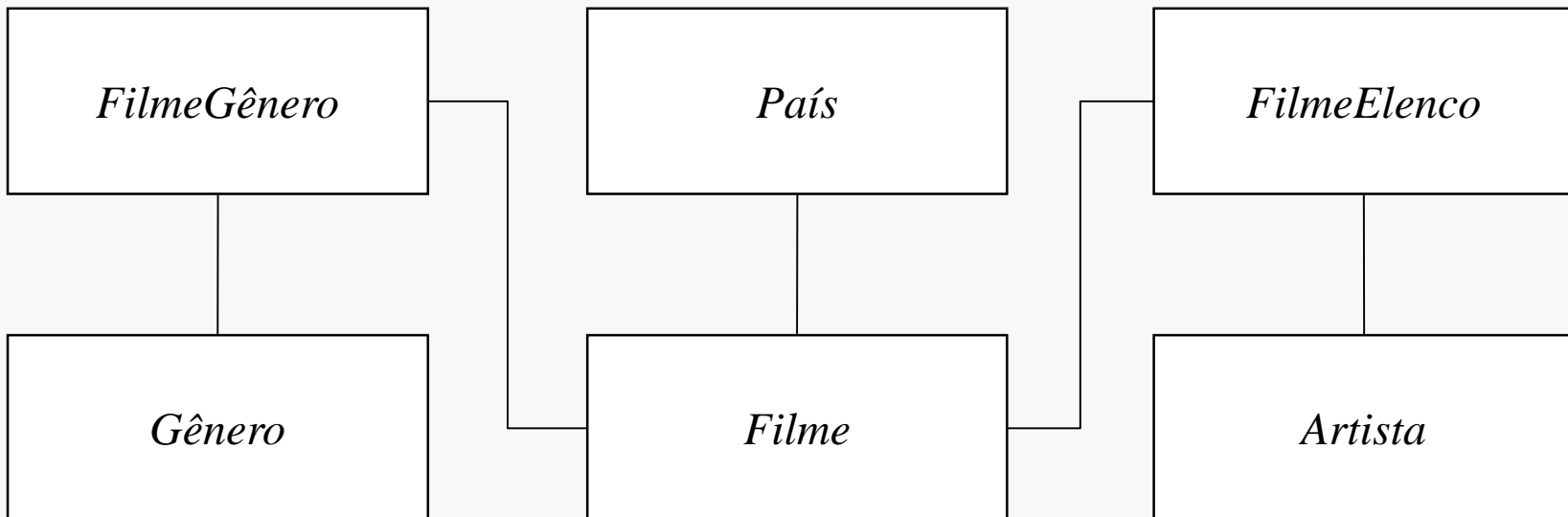
- Desde de o início do curso estamos trabalhando com banco de dados e preocupando somente com as relações entre as tabelas e seus atributos através de um conhecimento tácito que permite representar o minimundo numa base de dados. O BD de *Filme* pode ser usado como exemplo.



- Este BD possui 6 relações com seus atributos como país, gênero da tabela *Filme* que permitem aplicar junções com as relações *País* e *FilmeGênero* para produzir, por exemplo, resultados (relatórios). Por que há 6 relações e não uma relação única com todas as informações sem a necessidade de ter que ficar fazendo junções?
- A **normalização** nos ajuda a responder este tipo de pergunta.

Introdução

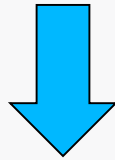
- Este BD possui 6 relações com seus atributos como **país**, **gênero** da tabela *Filme* que permitem aplicar junções com as relações *País* e *FilmeGênero* para produzir, por exemplo, resultados (relatórios).
- Por que há 6 relações e não uma relação única com todas as informações sem a necessidade de ter que ficar fazendo junções?
- A **normalização** nos ajuda a responder este tipo de pergunta.



Introdução

- Normalização
 - **Processo** de “decomposição” de relações:
 - Uma relação é substituída por diversas outras que, em conjunto, incluem todos os atributos da relação original.

$F(\underline{\text{titulo}}, \underline{\text{ano}}, \text{resumo}, \text{elenco}(\text{nome}, \text{sexo}), \text{gêneros}, \text{país}(\text{nome}, \text{sigla}), \text{duração}, \text{avaliação})$



$Filme(\underline{\text{titulo}}, \underline{\text{ano}}, \text{resumo}, \text{pais}, \text{duração}, \text{avaliacao})$

$Pais(\underline{\text{sigla}}, \text{nome})$

$Artista(\underline{\text{nome}}, \text{sexo})$

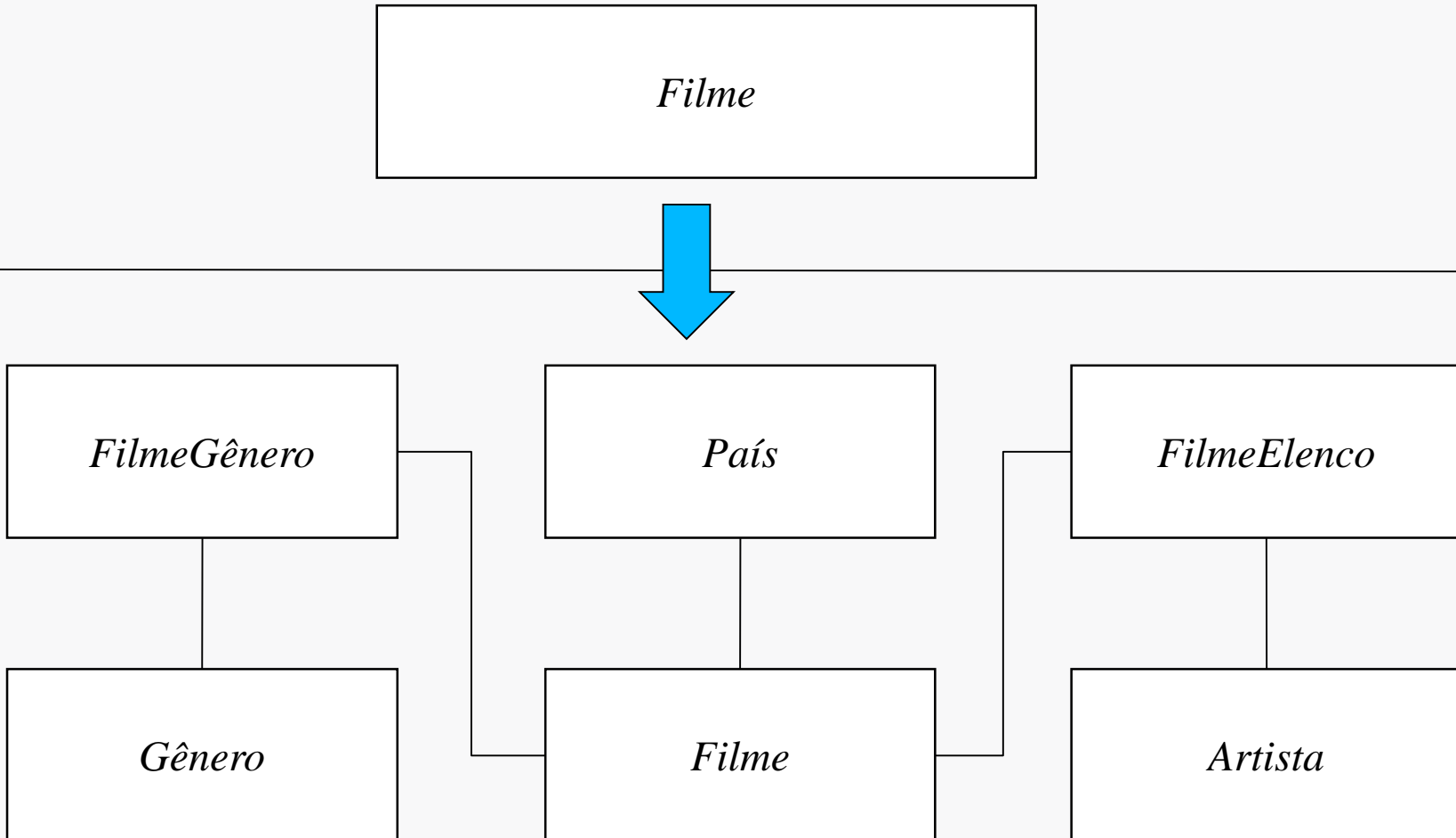
$Genero(\underline{\text{nome}})$

$FilmeGenero(\underline{\text{titulofilme}}, \underline{\text{ano}}, \underline{\text{nomegenero}})$

$FilmeElenco(\underline{\text{titulofilme}}, \underline{\text{ano}}, \underline{\text{nomeartista}})$

Introdução

- A **normalização** nos ajuda a organizar as relações para evitar problemas futuros.



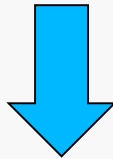
Normalização de Relações

- A normalização é aplicada com o intuito de que sejam atingidos dois objetivos em um projeto de BD:

1. Processo de “decomposição” de relações:

- Minimizar o armazenamento redundante da mesma informação.
- Desta forma, reduz-se a necessidade de múltiplas atualizações para manter a consistência dos dados entre múltiplas cópias da mesma informação.
 - Consequentemente, a possibilidade de ocorrência de **anomalias** no BD é **minimizada**.

F(titulo , ano , resumo, elenco(nome, sexo), gêneros, país(nome, sigla), duração, avaliação)



Filme (titulo, ano, resumo, país, duração, avaliacao)

País (sigla, nome)

Artista (nome, sexo)

Genero (nome)

FilmeGenero (titulofilme, ano, nomegenero)

FilmeElenco (titulofilme, ano, nomeartista)

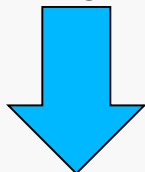
Normalização de Relações

- A normalização é aplicada com o intuito de que sejam atingidos dois objetivos em um projeto de BD:

2. Preservar a informação:

- Manter todos os conceitos relevantes do minimundo a ser modelado, incluindo entidades, seus relacionamentos e tipos de atributo.

F(titulo, ano, resumo, elenco(nome, sexo), gêneros, país(nome, sigla), duração, avaliação)



Filme (titulo, ano, resumo, país, duração, avaliacao)

País (sigla, nome)

Artista (nome, sexo)

Genero (nome)

FilmeGenero (titulofilme, ano, nomegenero)

FilmeElenco (titulofilme, ano, nomeartista)

Normalização de Relações

- O processo de normalização baseia-se nos seguintes conceitos (especialmente os dois últimos):
 - Atomicidade de Valores
 - Superchave e Chave
 - Dependência Funcional
- **IMPORTANTE:**
 - A normalização é um conceito essencial para a modelagem de SGBD relacionais.
 - Nos SGBD NoSQL a técnica costuma ser “relaxada”, o que significa que ela é aplicada de uma forma menos rigorosa.

Teoria Relacional/Atomicidade de Valores

- Domínio - D : Conjunto de valores atômicos
- Por atômico, queremos dizer que cada valor é de um tipo **simples**, **indivisível** ex : inteiro, real, alfanumérico, lógico).
- O valor **não** pode ser de um tipo “complexo” (ex: lista, vetor, objeto).
 - **Exemplos:**
 - **anos de produção:** possíveis anos de produção de filmes; cada valor deve ser um inteiro acima de 1874.
 - **notas de avaliação de filmes:** possíveis valores atribuídos para avaliações. Cada valor deve ser um real entre 0.0 e 10.0
 - **nomes de artistas:** o conjunto de valores alfanuméricos que representa nomes de pessoas.

Teoria Relacional/Atomicidade de Valores

- **Esquema Relação** – $R(A_1, A_2, \dots, A_n)$: nome de uma relação (R) e uma lista de n atributos (A_1, A_2, \dots, A_n).
- Cada atributo A_i é o nome do papel executado por algum domínio D no esquema de R .
- O domínio de A_i é denotado por $Dom(A_i)$.

Teoria Relacional/Atomicidade de Valores

- **Instância da Relação -r**: conjunto de **tuplas**

$$r = \{t_1, t_2, \dots, t_m\}$$

- Cada **tupla** t é uma lista de n valores

$$t = \langle v_1, v_2, \dots, v_n \rangle$$

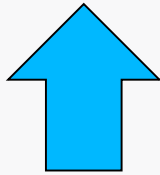
- Cada **valor** v_i , $1 \leq i \leq n$, é um elemento de $\text{Dom}(A_i)$ ou NULL.
- O **i-ésimo valor** de uma tupla t , corresponde ao atributo A_i e pode ser representado por $t.A_i$ ou $t[A_i]$.

Teoria Relacional/Atomicidade de Valores

- **Exemplo de tuplas de grau 5 e 1:**
 - $t = \langle \text{'Intocáveis'}, 2011, \text{'After he becomes bla bla ...'}, \text{'FR'}, 112, 8.6 \rangle$
 - $t[\text{titulo}] = \langle \text{'Intocáveis'} \rangle$ (grau 1)
 - $t[\text{ano}] = \langle 2011 \rangle$ (grau 1)

Teoria Relacional/Atomicidade de Valores

- É possível utilizar a notação $t[A_u, A_w, \dots, A_z]$ ou $t.(A_u, A_w, \dots, A_z)$ para representar os valores de um **conjunto específico** de atributos.
- $t[\text{titulo}, \text{ano}, \text{pais}] = \langle \text{'Intocáveis'}, 2011, \text{'FR'} \rangle$



Conjunto de 3 atributos



Tupla específica de 3 valores

Teoria Relacional/Superchave e chave

- Uma relação é um conjunto de tuplas.
 - Elementos de um conjunto são, por definição, distintos.
 - Ou seja: não devem existir duas tuplas em R com a mesma combinação de valores para todos os seus atributos.
 - Ex.: não podem existir dois filmes com o mesmo **titulo, ano, país, duração**.

Relação R

Titulo	Ano	Pais	duração
O discreto charme da burguesia	1972	FR	102
Ensina-me a viver	1971	US	95
Estado de Sítio	1972	FR	102
Caça-Fantasmas	1986	US	90
Caça-Fantasmas	2016	US	90

Teoria Relacional/Superchave e chave

- Na maioria dos casos, existem outros subconjuntos de atributos que jamais possuirão duas tuplas com a mesma combinação, para qualquer instância r (r é conjunto de tuplas) de R .
- Suponha que esse conjunto de atributos seja denotado por **SK**;
- Então, para quaisquer duas tuplas distintas t_1 e t_2 em uma instância r de R teremos a seguinte restrição : $t_1[\text{SK}] \neq t_2[\text{SK}]$
 - Ex.: não podem existir dois filmes com o mesmo {titulo, ano, país } ou {título, ano, duração } ou {titulo, ano}.

Relação R

Titulo	Ano	Pais	duração
O discreto charme da burguesia	1972	FR	102
Ensina-me a viver	1971	US	95
Estado de Sítio	1972	FR	102
Caça-Fantasmas	1986	US	90
Caça-Fantasmas	2016	US	90

Teoria Relacional/Superchave e chave

- Note que uma SK pode ter atributos redundantes:
 - {titulo, ano, **país, duração**}
- Portanto, um conceito mais útil é o de chave (*key*) que não tem redundância.
 - {titulo, ano}

Relação *R*

Título	Ano	Pais	duração
O discreto charme da burguesia	1972	FR	102
Ensina-me a viver	1971	US	95
Estado de Sítio	1972	FR	102
Caça-Fantasmas	1986	US	90
Caça-Fantasmas	2016	US	90

Teoria Relacional/Superchave e chave

- Definição Formal - Chave
 - Uma chave K de uma relação R é uma SK de R com a propriedade adicional de que a remoção de qualquer atributo A de K , resulta em um conjunto de atributos K' que não é uma SK de R .

A retirada de um atributo da chave de R retira a propriedade de identificador único.

Leva a compreensão de que uma chave primária (PK) precisa ter o menor número de atributos possíveis.

Teoria Relacional/Superchave e chave

Uma chave satisfaz duas propriedades:

1. Duas tuplas distintas em qualquer instância da relação não podem possuir valores idênticos para todos os atributos da chave.
 2. Deve ser uma SK mínima, isto é, uma superchave onde não podemos remover nenhum atributo e ainda manter a restrição de unicidade garantida.
- Note que a primeira propriedade vale para superchaves e chaves. No entanto, a segunda apenas para chaves.

Teoria Relacional/Superchave e chave

- Toda chave é uma superchave . Entretanto, algumas superchaves não são chaves, pois a chave tem um conjunto mínimo de atributos que são únicos numa relação.
 - {titulo, ano, país, duração} é uma superchave.
 - {titulo, ano, país} é superchave.
 - {titulo, ano, duração} é uma superchave.
 - {titulo, ano} é uma **superchave** e uma **chave** (número mínimo de A_i)
 - {titulo} sozinho **não** é superchave e nem chave;
 - {ano} sozinho **não** é superchave e nem chave

Relação R

Título	Ano	Pais	duração
O discreto charme da burguesia	1972	FR	102
Ensina-me a viver	1971	US	95
Estado de Sítio	1972	FR	102
Caça-Fantasmas	1986	US	90
Caça-Fantasmas	2016	US	90

Teoria Relacional/Superchave e chave

- O valor da chave pode ser utilizado para identificar unicamente cada tupla na relação.
- **Ex.:** a matrícula ‘M001’ identifica unicamente o estudante ‘Milton’.

É importante observar que uma chave é determinada pelo significado dos atributos e deve valer para qualquer instância da relação.

Ex.: não podemos designar “nome” como chave, pois em algum momento pode ser inserido um estudante que tenha mesmo nome de outro já cadastrado em *Alunos*.

Relação *Alunos*

Matrícula	Nome	idade
M01	Milton	45
M033	Gilberto	29
M110	Clara	19
M076	Roberta	32

Teoria Relacional/Superchave e chave

- Chave Candidata e Chave Primária
 - Muitas vezes uma relação pode ter mais de uma chave.
 - Na relação *Imoveis* temos duas. **Elas são chamadas de chaves candidatas.**
 - **Inscrição**: número de inscrição do imóvel na Prefeitura.
 - **endereço + cep**: esses dois atributos em conjunto nunca possuirão valores iguais em duas tuplas.

Relação *Imoveis*

Inscricao	Endereco	Cep	Valor IPTU
123456	Rua alpha, 44	11111-000	120,00
999999	Rua alpha, 47	11111-000	120,0
235890	Rua André Cavalcante, 01	2222-000	278,0
692232	Rua André Cavalcante, 02	5555-999	300,0

Teoria Relacional/Superchave e chave

- Chave Candidata e Chave Primária
 - Neste caso, uma das chaves candidatas escolhida como **chave primária** PK (“inscrição”) e a outra continuará sendo chamada de **chave candidata** (“endereço”, “cep”).
 - Uma regra importante é escolher a chave mais simples (normalmente aquela com menor número de atributos) como PK.
 - As chaves candidatas não são sublinhadas: Imóveis(inscrição, endereço, cep, valor_ipitu).

Relação *Imoveis*

Inscricao	Endereco	Cep	Valor_ipitu
123456	Rua alpha, 44	11111-000	120,00
999999	Rua alpha, 47	11111-000	120,0
235890	Rua André Cavalcante, 01	2222-000	278,0
692232	Rua André Cavalcante, 02	5555-999	300,0

Teoria Relacional/Superchave e chave

- Chave Candidata e Chave Primária
 - Note que a definição de chave para esta relação está baseada no universo de informações visíveis, contudo é fundamental conhecer o minimundo para definir com maior precisão se um conjunto de atributos define uma chave. Assim, a escolha da chave deve ser baseada no levantamento de informações.

Relação *Imoveis*

Inscricao	Endereco	Cep	Valor_ipitu
123456	Rua alpha, 44	11111-000	120,00
999999	Rua alpha, 47	11111-000	120,0
235890	Rua André Cavalcante, 01	2222-000	278,0
692232	Rua André Cavalcante, 02	5555-999	300,0

Dependência Funcional

- Considere uma empresa que paga o **salário básico** de seus funcionários de uma maneira padronizada, que é definida pelo **cargo** e **categoria** de cada funcionário, semelhante a um plano de carreira.

Relação *Funcionários*

Matricula	Nome	Cargo	Categoria	Salário-básico	Vale_trasn porte
F002	Marta	Técnico	Pleno	1.200	300
F004	Zico	Engenheiro	Sênior	3.000	200
F005	Sócrates	Supervisor	Pleno	3.000	180
F007	Garrincha	Engenheiro	Sênior	3.000	150
F011	Leônidas	Engenheiro	Pleno	2.000	320
F013	Alisha	Supervisor	Pleno	3.000	240
F014	Quitéria	Engenheiro	Pleno	2.000	128
F022	Olga	Supervisor	Pleno	3.000	150
F025	Narges	Técnico	Pleno	1.200	320

Dependência Funcional

- Cargo='Técnico', Categoria='Pleno' → Salário = R\$1.200,00
- Cargo='Engenheiro', Categoria='Pleno' → Salário = R\$2.000,00
- Cargo='Engenheiro', Categoria='Sênior' → Salário = R\$3.000,00
- Cargo='Supervisor', Categoria='Pleno' → Salário = R\$3.000,00

Relação *Funcionários*

Matricula	Nome	Cargo	Categoria	Salário-básico	Vale_trasn porte
F002	Marta	Técnico	Pleno	1.200	300
F004	Zico	Engenheiro	Sênior	3.000	200
F005	Sócrates	Supervisor	Pleno	3.000	180
F007	Garrincha	Engenheiro	Sênior	3.000	150
F011	Leônidas	Engenheiro	Pleno	2.000	320
F013	Alisha	Supervisor	Pleno	3.000	240
F014	Quitéria	Engenheiro	Pleno	2.000	128
F022	Olga	Supervisor	Pleno	3.000	150
F025	Narges	Técnico	Pleno	1.200	320

Dependência Funcional

- Desta forma, o valor do “salário_básico” em uma tupla é determinado pelo valor conjunto dos atributos “cargo” e “categoria”.

Relação *Funcionários*

Matricula	Nome	Cargo	Categoria	Salário-básico	Vale_trasn porte
F002	Marta	Técnico	Pleno	1.200	300
F004	Zico	Engenheiro	Sênior	3.000	200
F005	Sócrates	Supervisor	Pleno	3.000	180
F007	Garrincha	Engenheiro	Sênior	3.000	150
F011	Leônidas	Engenheiro	Pleno	2.000	320
F013	Alisha	Supervisor	Pleno	3.000	240
F014	Quitéria	Engenheiro	Pleno	2.000	128
F022	Olga	Supervisor	Pleno	3.000	150
F025	Narges	Técnico	Pleno	1.200	320

Dependência Funcional

- “salário_básico” é funcionalmente dependente de “cargo” e “categoria”.
- Representação: $\{\text{cargo}, \text{categoria}\} \rightarrow \{\text{salário_básico}\}$

Relação *Funcionários*

Matricula	Nome	Cargo	Categoria	Salário-básico	Vale_trasn porte
F002	Marta	Técnico	Pleno	1.200	300
F004	Zico	Engenheiro	Sênior	3.000	200
F005	Sócrates	Supervisor	Pleno	3.000	180
F007	Garrincha	Engenheiro	Sênior	3.000	150
F011	Leônidas	Engenheiro	Pleno	2.000	320
F013	Alisha	Supervisor	Pleno	3.000	240
F014	Quitéria	Engenheiro	Pleno	2.000	128
F022	Olga	Supervisor	Pleno	3.000	150
F025	Narges	Técnico	Pleno	1.200	320

Dependência Funcional

- Definição Formal - Dependência Funcional
 - Uma dependência funcional (*functional dependency* - *FD*) em uma relação R , denotada por $X \rightarrow Y$ (X define Y), especifica a seguinte restrição sobre as tuplas que podem formar uma instância de R :
 - Se duas tuplas de R concordam em um conjunto de atributos X (*i.e.*, as tuplas possuem os mesmos valores para cada um dos atributos que compõem X), então elas devem também concordar em outra lista de atributos Y .

Dependência Funcional

- Dependência Funcional
 - **Em outras palavras:** quaisquer duas tuplas t_1 e t_2 que tenham $t_1[X] = t_2[X]$ devem ter também $t_1[Y] = t_2[Y]$.
 - Exemplo: O valor de um atributo A_i numa tupla de uma relação define o valor do atributo A_n (sendo $i \neq n$) e numa outra tupla qualquer da relação esta definição se mantém.

Dependência Funcional

- $X \rightarrow Y$
- Os valores do componente X de uma tupla determinam unicamente os valores do componente Y .
 - Existe dependência funcional de X para Y ;
 - Y é funcionalmente dependente de X .
 - X determina Y .

(são formas diferentes de descrever a mesma coisa)

Dependência Funcional

- Exemplo 1: dada uma determinada cidade (desconsiderando cidades homônimas) é possível saber a sua UF e com a UF a região.
- $\text{Cidade} \rightarrow \text{UF}$
- $\text{UF} \rightarrow \text{região}$
- UF é funcionalmente dependente de cidade e região é funcionalmente dependente de UF.
- Cidade determina UF e UF determina região.

Cidade	UF	região
Alegre	ES	Sudeste
Vila velha	ES	Sudeste
Curitiba	PR	Sul
Vitória	ES	Sudeste
Londrina	PR	Sul
Porto Alegre	RS	Sul

Dependência Funcional

- Exemplo 2: considere a relação *Filmes1* (titulo, ano, pais, duracao, artista)
- Observe que essa relação tenta “fazer muita coisa”. Ela mistura atributos das relações *Filme* e *FilmeElenco* de nosso exemplo do IMDb.
- Em breve vamos compreender que ela não representa um bom projeto! E que são as FD que nos indicam isso.

Relação *Filmes1*

Titulo	Ano	Pais	duração	artista
Ensina-me a Viver	1971	US	95	Ruth
Ensina-me a Viver	1971	US	95	Bud Cort
Estado de Sítio	1972	FR	102	Yves Montand
A Vingança de Manon	1986	FR	113	Emmanuelle Béart
A Vingança de Manon	1986	FR	113	Yves Montando

Dependência Funcional

- Exemplo 2: considere a relação *Filmes1* (título, ano, país, duracao, artista).
- $\{\text{título, ano}\} \rightarrow \{\text{país, duracao}\}$ é uma FD.
 - Se duas tuplas possuem os mesmos valores para título e ano, elas terão o mesmo valor para país e o mesmo para duração.
- $\{\text{título, ano}\} \rightarrow \{\text{artista}\}$ não é uma FD.
 - Dado um filme, a instância abaixo mostra que é inteiramente possível que ele possua mais de um artista.

Relação *Filmes1*

Titulo	Ano	Pais	duração	artista
Ensina-me a Viver	1971	US	95	Ruth
Ensina-me a Viver	1971	US	95	Bud Cort
Estado de Sítio	1972	FR	102	Yves Montand
A Vingança de Manon	1986	FR	113	Emmanuelle Béart
A Vingança de Manon	1986	FR	113	Uves Montando

Dependência Funcional – Regras e Propriedades

- $X \rightarrow Y$ não implica em $Y \rightarrow X$
- $UF \rightarrow$ região é uma FD.
 - A mesma UF tem a mesma região em linhas diferentes.
- Porém região \rightarrow UF não é um FD.
 - A mesma região tem UF diferente em linhas diferentes.

Cidade	UF	região
Alegre	ES	Sudeste
Vila velha	ES	Sudeste
Curitiba	PR	Sul
Vitória	ES	Sudeste
Londrina	PR	Sul
Porto Alegre	RS	Sul

Dependência Funcional – Regras e Propriedades

- (ii) Se **X** é uma **chave** de *R*, isto implica que $X \rightarrow Y$ é verdadeiro para **qualquer subconjunto de atributos Y** de *R*.
- Exemplo:
 - Matrícula identifica unicamente cada tupla de Alunos .
 - Dado o valor da matrícula, sabemos o nome e a idade do aluno.
 - $\text{Matricula} \rightarrow \text{nome}$
 - $\text{Matricula} \rightarrow \text{idade}$

Relação Alunos

<u>matricula</u>	nome	idade
M001	Milton	45
M033	Gilberto	29
M110	Clara	19
M076	Roberta	32

Dependência Funcional – Regras e Propriedades

- (iii -**TRANSITIVIDADE**) Se em uma relação R , tem-se que $X \rightarrow Y$ e $Y \rightarrow Z$ representam FD válidas, então pode-se afirmar que $X \rightarrow Z$ também é uma FD válida.
- cidade \rightarrow UF e UF \rightarrow região são FD.
- Então cidade \rightarrow região também é uma FD.
- Diz-se que: “cidade \rightarrow região é transitiva em relação à UF”

Cidade	UF	região
Alegre	ES	Sudeste
Vila velha	ES	Sudeste
Curitiba	PR	Sul
Vitória	ES	Sudeste
Londrina	PR	Sul
Porto Alegre	RS	Sul

Dependência Funcional – Regras e Propriedades

- (iv-TRIVIALIDADE) $X \rightarrow Y$ é considerada **trivial** se $Y \subseteq X$.

(\subseteq - Y é subconjunto de X)

- **Exemplos de FD triviais:**

- $\{\text{titulo}, \text{ano}\} \rightarrow \{\text{titulo}\}$
- $\{\text{titulo}\} \rightarrow \{\text{titulo}\}$

- **Em contrapartida:**

- $\{\text{titulo}, \text{ano}\} \rightarrow \{\text{titulo}, \text{duração}\}$
- É **não-trivial**, pois $\{\text{titulo}, \text{duração}\} \not\subseteq \{\text{titulo}, \text{ano}\}$

Dependência Funcional – Regras e Propriedades

- (v - DECOMPOSIÇÃO) É possível decompor $A \rightarrow \{B, C, D\}$, em um conjunto de FD
 - $A \rightarrow B$
 - $A \rightarrow C$
 - $A \rightarrow D$
- A FD $\{\text{titulo}, \text{ano}\} \rightarrow \{\text{duração}, \text{pais}\}$ pode ser decomposta em:
 - $\{\text{titulo}, \text{ano}\} \rightarrow \{\text{duração}\}$
 - $\{\text{titulo}\} \rightarrow \{\text{pais}\}$

Relação *Filmes1*

Titulo	Ano	Pais	duração	artista
Ensina-me a Viver	1971	US	95	Ruth
Ensina-me a Viver	1971	US	95	Bud Cort
Estado de Sítio	1972	FR	102	Yves Montand
A Vingança de Manon	1986	FR	113	Emmanuelle Béart
A Vingança de Manon	1986	FR	113	Uves Montando

Dependência Funcional – Regras e Propriedades

- (v - DECOMPOSIÇÃO) É possível decompor $A \rightarrow \{B, C, D\}$, em um conjunto de FD
 - Porém, o lado esquerdo da FD não pode ser decomposto
 - $\{\text{título}\} \rightarrow \{\text{duração}\}$
 - $\{\text{ano}\} \rightarrow \{\text{duração}\}$
 - As FD acima não são válidas! O título sozinho não determina a duração (dois filmes com o mesmo título podem ter durações diferentes) e o ano sozinho também não define duração.

Dependência Funcional – Regras e Propriedades

- (vi) Uma FD é propriedade da semântica ou do significado dos atributos de uma relação R .
 - Os projetistas de BD utilizam o seu conhecimento a respeito da semântica dos atributos de R (como eles estão relacionados) para especificar as FD que são válidas para qualquer instância de R .
 - Isto significa que, dada uma relação populada, não é possível determinar quais as FD são válidas, a não ser que a relação entre os atributos seja conhecida. Porém, muitas vezes é possível determinar quais **FD não existem**.

Exemplo:

- **Possível FD:**
 - $\{\text{livro}\} \rightarrow \{\text{curso}\}$
- **Não são FD:**
 - $\{\text{professor}\} \rightarrow \{\text{curso}\}$
 - $\{\text{curso}\} \rightarrow \{\text{livro}\}$

Professor	Curso	Livro
Smith	Mineração	Han & Kamber
Smilth	Banco de dados	C. J. Date
Hall	Redes	Kurose & Ross
Brown	Mineração de dados	Zaki & Meira Jr.

Anomalias

- Uma relação mal projetada pode levar a redundância e a anomalias relacionadas.
 - Por exemplo, considere novamente a relação *Filmes1* (introduzida em slide anterior).
 - Observe que o “país” e “duração” para os filmes “Ensina me a Viver” e “A Vingança de Manon” são repetidos para cada artista do filme.
 - Esta repetição é redundante e introduz potencial para anomalias de atualização, exclusão e inclusão.

Relação *Filmes1*

titulo	ano	pais	duração	artista
Ensina-me a Viver	1971	US	95	Ruth
Ensina-me a Viver	1971	US	95	Bud Cort
Estado de Sítio	1972	FR	102	Yves Montand
A Vingança de Manon	1986	FR	113	Emmanuelle Béart
A Vingança de Manon	1986	FR	113	Yves Montando

Anomalias

- As anomalias podem trazer dificuldades para manter as informações devidamente atualizadas. Quando elas existem uma operação de atualização requer muito cuidado:
 - Exemplo, se for alterado o valor de duração em uma tupla e não alterar o mesmo valor numa outra tupla, isso pode trazer problemas, por exemplo, um relatório pode apresentar informações inconsistentes.

Relação *Filmes1*

titulo	ano	pais	duração	artista
Ensina-me a Viver	1971	US	95	Ruth
Ensina-me a Viver	1971	US	95	Bud Cort
Estado de Sítio	1972	FR	102	Yves Montand
A Vingança de Manon	1986	FR	113	Emmanuelle Béart
A Vingança de Manon	1986	FR	113	Yves Montando

Anomalias de Atualização

- Ocorre quando modifica-se uma informação em uma tupla, mas deixa-se de modificar a mesma informação em outra.
 - Exemplo:
 - Se descobrirmos que “**Ensina me a Viver**” na verdade tem 90 minutos, podemos modificar essa informação na primeira tupla, mas esquecer de modificar na segunda tupla.

Relação *FilmesI*

titulo	ano	pais	duração	artista
Ensina-me a Viver	1971	US	95 — 90	Ruth
Ensina-me a Viver	1971	US	95	Bud Cort
Estado de Sítio	1972	FR	102	Yves Montand
A Vingança de Manon	1986	FR	113	Emmanuelle Béart
A Vingança de Manon	1986	FR	113	Yves Montando

Anomalias de Exclusão

- Ocorre quando um dado conjunto de valores se torna vazio e, com isso, perdemos outra informação como efeito colateral.

Exemplo:

- Se excluirmos “Yves Montand” do conjunto de artistas do filme “Estado de Sítio”, então não haveriam mais artistas relacionados a esse filme.
- Consequentemente, a última tupla de “Estado de Sítio” iria desaparecer, em conjunto com as informações de que este é um filme francês com 102 minutos de duração.

Relação *FilmesI*

titulo	ano	pais	duração	artista
Ensina-me a Viver	1971	US	95	Ruth
Ensina-me a Viver	1971	US	95	Bud Cort
Estado de Sítio	1972	FR	102	Yves Montand
A Vingança de Manon	1986	FR	113	Emmanuelle Béart
A Vingança de Manon	1986	FR	113	Yves Montando

Anomalias de Inclusão

- Ocorre quando certos atributos não podem ser inseridos no BD sem a presença de outros atributos.
 - Exemplo:
 - Se desejarmos incluir a artista “**Juliette Binoche**” sem nenhum filme associado, a única forma seria colocar o valor NULL em todos os outros atributos (titulo, ano, pais, duracao).

Relação *Filmes1*

titulo	ano	pais	duração	artista
Ensina-me a Viver	1971	US	95	Ruth
Ensina-me a Viver	1971	US	95	Bud Cort
Estado de Sítio	1972	FR	102	Yves Montand
A Vingança de Manon	1986	FR	113	Emmanuelle Béart
A Vingança de Manon	1986	FR	113	Yves Montando
<i>NULL</i>	<i>NULL</i>	<i>NULL</i>	<i>NULL</i>	<i>Juliette Binoche</i>

Obrigado