

Banco de Dados - INE 5323


Projeto de Banco de Dados Relacionais

Prof. Mario Dantas

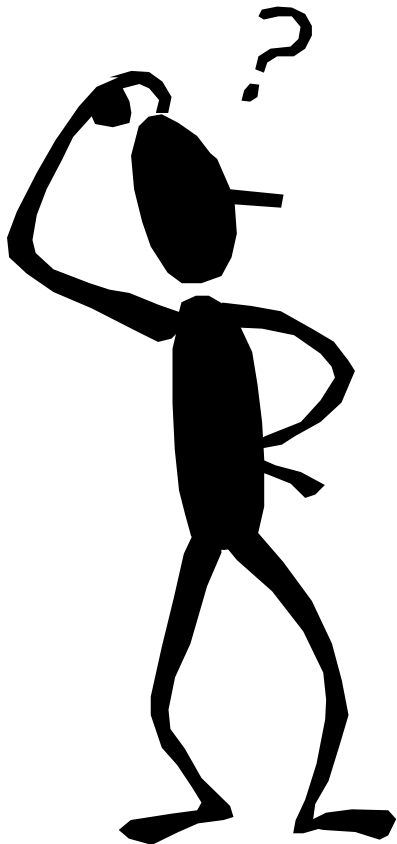
Projeto de Banco de Dados Relacionais

- Armadilhas
- Decomposição
- Dependências Funcionais e Normalização
- Dependências Multivaloradas e Normalização
- Abordagens Alternativas

Projeto de Banco de Dados Relacionais

- Armadilhas 
- Decomposição
- Dependências Funcionais e Normalização
- Dependências Multivaloradas e Normalização
- Abordagens Alternativas

Armadilhas no Projeto de Banco de Dados Relacional



O que seriam armadilhas no projeto de banco de dados relacional ?

Armadilhas no Projeto de Banco de Dados Relacional

O projeto de banco de dados relacional requer uma boa coleção de sistemas relacionais. Um mau projeto pode conduzir :

(1) Repetição de informações

(2) Inabilidade para representar certas informações

Armadilhas no Projeto de Banco de Dados Relacional

- Projete metas:
 - Evite dados redundantes
 - Garanta que haja relacionamentos entre atributos
 - Facilite a verificação de atualizações por violação da restrição de integridade do banco de dados

Observações Gerais

- Imagine a dependência funcional entre :

nome_agência \blacktriangleleft cidade_agência

Qual o seu *sentimento* sobre a dependência apresentada ?

Esta dependência é importante e deve se realizar

Observações Gerais

- Imagine a dependência funcional entre :

nome_agência \blacktriangle número_empréstimo

Qual o seu *sentimento* sobre a dependência apresentada ?

Esta dependência não é importante e não deve se realizar

Exemplo

Imagine que tenhamos dois esquemas (*estáveis*) e que por uma razão qualquer alguém proponha a união destes esquemas em um único esquema.

Qual sua opinião sobre a união de esquemas (*estáveis*) ?

Exemplo

Qual sua opinião sobre a união de esquemas
(estáveis) ?

*Esta união dever ser analisada cuidadosamente, pois na maioria da vezes não devemos realizar união de esquemas **estáveis**.*

Exemplo

- Esquema_agência = (nome_agência, fundos, cidade_agência)

nome_agência	fundos	cidade_agência
Downtown	9.000.000	Brooklyn
Redwood	2.100.000	Palo alto
Perryridge	1.700.000	Horseneck

- Esquema_empréstimo = (nome_agência, número_empréstimo, total)

Nome_agência	Número_empréstimo	Nome_cliente	total
Downtown	17	Jones	1000
Redwood	23	Smith	2000
Perryridge	15	Jackson	1500

Exemplo

Resultado :

- Esquema_empréstimo = (nome_agência, fundos, cidade_agência, número_empréstimo, nome_cliente, total)

nome_agência	fundos	cidade-agência	número-empréstimo	nome_cliente	total
Downtown	9.000.000	Brooklyn	17	Jones	1000
Redwood	2.100.000	Palo alto	23	Smith	2000
Perryridge	1.700.000	Horseneck	15	Jackson	1500

Observações Gerais

Problemas que podemos inferir dessa união :

- Não podemos representar diretamente a informação relativa a uma agência (somente se houver um empréstimo na agência).

Existe a necessidade na nova relação de existir valores em uma tupla sobre *número_empréstimo*, *total* e *nome_cliente* ;

Observações Gerais


Problemas que podemos inferir dessa união :

- Devemos inserir *valores nulos* nos campos para atualização (por exemplo).

Observações Gerais

- As atualizações serão mais *custosas*, pois deveremos executar *mais operações* no esquema da relação (ou seja no projeto lógico).

Projeto de Banco de Dados Relacionais

- Armadilhas
- Decomposição 
- Dependências Funcionais e Normalização
- Dependências Multivaloradas e Normalização
- Abordagens Alternativas

Decomposição

- Decompor a relação Esquema_empréstimo:

Esquema_agência_cliente = (nome_agência, cidade_agência, fundos, nome_cliente)

Esquema_cliente_empréstimo = (nome_cliente, número_empréstimo, total)

Decomposição

- Todos atributos do esquema original (R) devem aparecer na decomposição (R_1, R_2):

$$R = R_1 \cup R_2$$

- Decomposição sem perda na junção

Para todas as possíveis relações r no esquema R

$$r = \Pi_{R_1}(r) \bowtie \Pi_{R_2}(r)$$

Exemplo de uma Decomposição sem perda na junção

- Decomposição de $R = (A, B)$

A	B
α	1
α	2
β	1

r

$R_1 = (A) \quad R_2 = (B)$

A

α
β

$\Pi_A(r)$

B

1
2

$\Pi_B(r)$

- $\Pi_A(r) \bowtie \Pi_B(r)$

A	B
α	1
α	2
β	1
β	2

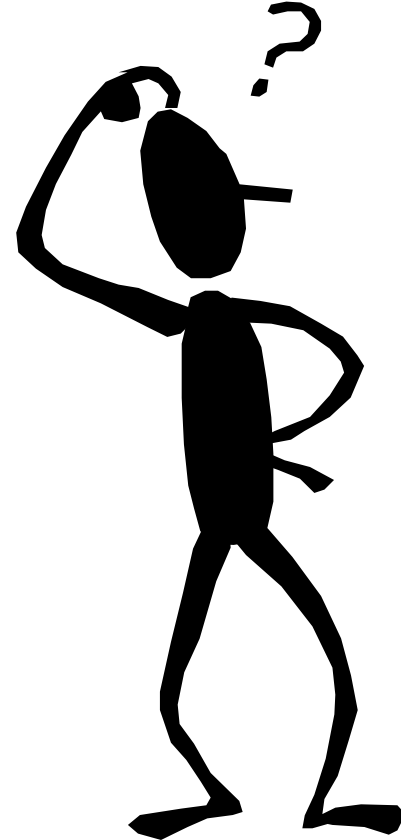
Decomposição

Exemplo - Considere a relação *Cinema* apresentada a seguir

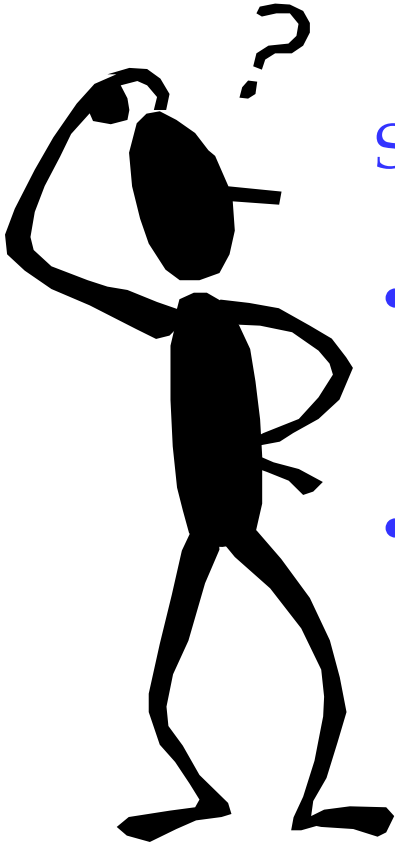
Título	Ano	Tempo	Tipo	Studio	Estrela
Star Wars	1977	124	Cor	Fox	Carrie Fisher
Star Wars	1977	124	Cor	Fox	Mark Hmill
Star Wars	1977	124	Cor	Fox	Harrison Ford
Mighty Ducks	1991	104	Cor	Disney	Emilio Estevez
Wayne's W.	1992	95	Cor	Param.	Dana Carvey
Wayne's W.	1992	95	Cor	Param.	Mike Meyers

Decomposição

Qual a *melhor* decomposição da relação *Cinema* no seu ponto de vista ?



Decomposição



Seria interessante termos os seguintes esquemas :

- Um esquema semelhante ao anterior sem o nome das estrelas;
- Um segundo esquema onde pudéssemos ter o título do filme, o ano e o nome da estrela

Decomposição

Exemplo - Do nosso primeiro esquema podemos afirmar que a relação abaixo está correta ?

Título	Ano	Tempo	Tipo	Studio
Star Wars	1977	124	Cor	Fox
Star Wars	1977	124	Cor	Fox
Star Wars	1977	124	Cor	Fox
Mighty Ducks	1991	104	Cor	Disney
Wayne's W.	1992	95	Cor	Param.
Wayne's W.	1992	95	Cor	Param.

Decomposição

Exemplo - Do nosso primeiro esquema temos a seguinte forma correta para a relação *Filme* :

Título	Ano	Tempo	Tipo	Studio
Star Wars	1977	124	Cor	Fox
Mighty Ducks	1991	104	Cor	Disney
Wayne's W.	1992	95	Cor	Param.

Observações Gerais

Note que a relação *Filme* elimina *anomalias* (redundâncias, anomalias de atualização e remoção).

Exemplos são :

- o tempo do filme que aparece somente uma vez;

Observações Gerais

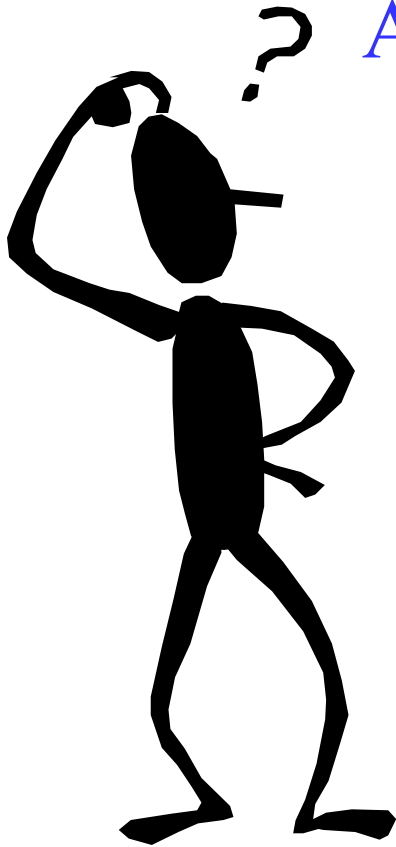
- atualizações, como por exemplo, somente vão afetar uma tupla, não existe o caso de inconsistência de mesmo filme com tempos diferentes ;
- não existe mais o risco de eliminar uma estrela e o filme ser eliminado.

Decomposição

Exemplo - E com relação a nossa relação *estrela*, estaria correta a proposta a seguir ?

Título	Ano	Estrela
Star Wars	1977	Carrie Fisher
Star Wars	1977	Mark Hmill
Star Wars	1977	Harrison Ford
Mighty Ducks	1991	Emilio Estevez
Wayne's W.	1992	Dana Carvey
Wayne's W.	1992	Mike Meyers

Decomposição



A relação *estrela* está em uma boa forma.

O que vocês podem dizer sobre os atributos título e ano na relação *estrela* ?

Embora pareçam redundantes estes atributos forma a chave para um filme.

Meta — Imagine uma seguinte teoria para:

- Decidir quando uma relação particular R é uma *boa forma*.

Meta — Imagine uma seguinte teoria para:

- No caso em que a relação R não é uma *boa forma*, decompor em um conjunto de relações $\{R_1, R_2, \dots, R_n\}$ de modo que :
 - cada relação é uma *boa forma*
 - a decomposição é uma decomposição sem perda na junção

Meta — Imagine uma seguinte teoria para:

- Nossa teoria é baseada em :
 - dependências funcionais
 - dependências multivaloradas

Projeto de Banco de Dados Relacionais

- Armadilhas
- Decomposição
- Dependências Funcionais e Normalização ←
- Dependências Multivaloradas e Normalização
- Abordagens Alternativas



Dependências Funcionais e Normalização

Melhoria de Desempenho em BD Relacionais

Objetivo : Evitar redundâncias através da divisão de relações.

Desafio : Manter a semântica da informação. É necessário que seja possível reconstituir relações originais.

Melhoria de Desempenho em BD Relacionais



É importante a redução (ou evitar) das redundâncias observando as chamadas *formas normais*.

Melhoria de Desempenho em BD Relacionais

Relação com presença de redundância

ISBN	Título	Autor	Edição	Editora
0130319953	BD	Garcia	1a.	PrenHall
0130319953	BD	Ullman	1a.	PrenHall
0130319953	BD	Widom	1a.	PrenHall
0130319953	BD	Garcia	2a.	PrenHall
0130319953	BD	Ullman	2a.	PrenHall
0130319953	BD	Widom	2a.	PrenHall

Dependência Funcional

Um conjunto de atributos Y é dependente de um conjunto de atributos X , expresso por $X \rightarrow Y$, se para cada tupla na relação, os valores de X determinam os valores de Y .

Melhoria de Desempenho em BD Relacionais

Dependência Funcional

EXEMPLOS

(1) O ISBN determina o título e editora

ISBN \rightarrow Título, Editora

(2) O ISBN não determina o autor.

ISBN \rightarrow Autor

Melhoria de Desempenho em BD Relacionais

Chaves como Dependências Funcionais

Nota : Uma chave é um conjunto mínimo de atributos cujos valores determinam de maneira única uma tupla em uma relação.

Melhoria de Desempenho em BD Relacionais

Chaves como Dependências Funcionais

Fornecida uma relação R : Se existir uma dependência funcional $X \rightarrow R$, onde X se apresenta como mínimo, então X é uma chave.

Melhoria de Desempenho em BD Relacionais

Chaves como Dependências Funcionais

Metas de Projeto de um Banco de Dados: Fazer com que todas as dependências funcionais apresentem dependências de chaves, sem perder a informação semântica

Melhoria de Desempenho em BD Relacionais

Chaves como Dependências Funcionais

A parte esquerda de qualquer dependência funcional $X \rightarrow R$ é uma *superchave*.

Melhoria de Desempenho em BD Relacionais

Chaves como Dependências Funcionais

Ident	Nome	Sobrenome	CEP	Estado	Endereço
9988	Maria	Silva	88040	SC	Rua Souza ...
9900	Fátima	Alvarez	22440	RJ	Rua Carlos ...
8890	Lúcia	Baez	70900	DF	Quadra Sul
8880	Monica	Qiz	22040	RJ	Rua Nasce
8800	Marcia	Tronco	88030	SC	Rua Poti

Melhoria de Desempenho em BD Relacionais

Chaves como Dependências Funcionais

Ident	Telefone
9980	048 - 2242222
9900	021- 2398543
8890	061 - 4483625
8880	021 - 5222222
8800	048 - 331 6666

Derivando Dependências Funcionais : Transitividade

Algumas vezes nem todas as *dependências funcionais* estão explicitamente definidas. Como então podemos achar estas relações ?

- Através de técnicas como o *data mining* (*olhando nos dados*)
- *Observando conhecidas* dependências funcionais.

Melhoria de Desempenho em BD Relacionais

Derivando Dependências Funcionais : Transitividade

Teorema : Se $A \rightarrow B$ e $B \rightarrow C$, então $A \rightarrow C$ (transitividade)

Melhoria de Desempenho em BD Relacionais

Normalização

É um processo no qual esquemas de relação inadequados são decompostos através da *quebra* de atributos em esquemas de relações menores e mais apropriados.

Melhoria de Desempenho em BD Relacionais

1ª Forma Normal

Esta forma normal é caracterizada pela existência de atributos com apenas valores atômicos (indivisíveis).

Melhoria de Desempenho em BD Relacionais

1ª Forma Normal - Exemplos

Nº	Departamento	Gerente	Localização
1	Produção	Ricardo	Florianópolis
1	Produção	Jaqueline	Joinville
2	Perfuração	Cesar	Joinville
3	Exploração	Silvio	Florianópolis

Você acredita que a relação acima está na 1ª Forma Normal ?

Melhoria de Desempenho em BD Relacionais

1ª Forma Normal

Como podemos colocar a relação anterior na primeira forma normal ?

Melhoria de Desempenho em BD Relacionais

1ª Forma Normal - Exemplos

Nº	Departamento	Gerente
1	Produção	Ricardo
2	Perfuração	César
3	Exploração	Silvio

Melhoria de Desempenho em BD Relacionais

1ª Forma Normal - Exemplos

Nº	Localização
1	Florianópolis
2	Joinville
3	Florianópolis

Melhoria de Desempenho em BD Relacionais

2ª Forma Normal

Esta forma normal é caracterizada pela existência da dependência funcional de todos os atributos não chave com relação a chave primária da relação.

3ª Forma Normal

Uma relação R está na *3ª Forma Normal* se seus atributos não forem dependentes funcionalmente de outros atributos não chave, ou ainda, se seus atributos não chave não forem transitivamente dependentes das chaves primárias.

Relembrando Transitividade

Uma dependência funcional $X \rightarrow Y$ é tida como *transitiva*, se existir um conjunto de atributos não chave Z , que observe as seguintes dependências : $X \rightarrow Z$ e $Z \rightarrow Y$

Melhoria de Desempenho em BD Relacionais

Forma Normal Boyce-Codd

Uma relação R está na forma normal Boyce-Codd se existir uma dependência não trivial na relação que :

$$(A_1, A_2, \dots, A_n) \rightarrow B$$

e

(A_1, A_2, \dots, A_n) é uma superchave para a relação R

Projeto de Banco de Dados Relacionais

- Armadilhas
- Decomposição
- Dependências Funcionais e Normalização
- Dependências Multivaloradas e Normalização ←
- Abordagens Alternativas

Melhoria de Desempenho em BD Relacionais

Dependências Multivaloradas

Imagine que em um BD de uma biblioteca temos uma relação com os ISBNs dos livros, nomes dos autores e números de cópias.

ISBN	Autor	Cópias
85-7323-169-6	Dantas	1,2,3
0-13031-995-3	Molina Ulman Widom	1,2,3

Melhoria de Desempenho em BD Relacionais

Dependências Multivaloradas

Como podemos normalizar esta relação ?

ISBN	Autor	Cópias
85-7323-169-6	Dantas	1,2
0-13031-995-3	Molina Ulman Widom	1,2

Melhoria de Desempenho em BD Relacionais

Relação Normalizada para a Forma Normal Boyce-Codd

ISBN	Autor	Cópias
85-7323-169-6	Dantas	1
85-7323-169-6	Dantas	2
0-13031-995-3	Molina	1
0-13031-995-3	Molina	2
0-13031-995-3	Ullman	1
0-13031-995-3	Ullman	2
0-13031-995-3	Widom	1
0-13031-995-3	Widom	2

Melhoria de Desempenho em BD Relacionais

A relação anterior não tem dependências funcionais, ou seja está na *Forma Normal Boyce-Codd*.

Mas ainda temos *redundâncias* por quê ?

Temos *dependências multivaloradas*

Melhoria de Desempenho em BD Relacionais

4^a Forma Normal (4FN)

Uma relação **R** está na *4FN* se a parte esquerda da relação de dependência de todas as MVDs não triviais são superchaves.

Em outras palavras, podemos dizer que a *4FN* é a forma Boyce-Codd com a ausência de atributos multivalorados independentes.

Melhoria de Desempenho em BD Relacionais

Dependência Multivalorada (MVD)

Uma MVD é estabelecida quando, para um dado conjunto de valores de um determinado atributo X

Melhoria de Desempenho em BD Relacionais

Como Normalizar a relação abaixo ?

ISBN	Autor	Cópias
85-7323-169-6	Dantas	1
85-7323-169-6	Dantas	2
0-13031-995-3	Molina	1
0-13031-995-3	Molina	2
0-13031-995-3	Ullman	1
0-13031-995-3	Ullman	2
0-13031-995-3	Widom	1
0-13031-995-3	Widom	2

Melhoria de Desempenho em BD Relacionais

Teremos a primeira relação como :

ISBN	Autor
85-7323-169-6	Dantas
85-7323-169-6	Dantas
0-13031-995-3	Molina
0-13031-995-3	Molina
0-13031-995-3	Ullman
0-13031-995-3	Ullman
0-13031-995-3	Widom
0-13031-995-3	Widom

Melhoria de Desempenho em BD Relacionais

Teremos a segunda relação como :

ISBN	Cópias
85-7323-169-6	1
85-7323-169-6	2
0-13031-995-3	1
0-13031-995-3	2
0-13031-995-3	1
0-13031-995-3	2
0-13031-995-3	1
0-13031-995-3	2

Melhoria de Desempenho em BD Relacionais

Formas Normais

