

Normalização

Banco de Dados: Teoria e Prática

André Santanchè e Luiz Celso Gomes Jr
Instituto de Computação - UNICAMP
Agosto de 2013

Qualidade de um Esquema?

Alocação de Tarefas para Membros de um Projeto

Planilha

Membro						
Id Membro	Nome	Papel	Tarefa			
			Id Tarefa	Descrição	Data de Inicio	Horas Alocadas
mel	Melissa	Gerente	1700	Planejamento e Orçamento	15/01/2012	80
			1701	Projeto do Sistema	15/02/2012	120
asd	Asdrúbal	Analista	1701	Projeto do Sistema	15/02/2012	180
			1705	Especificação da Arquitetura	01/03/2012	120
			1705	Detalhamento de Modelos	30/03/2012	200
dor	Doriana	Programador	1705	Detalhamento de Modelos	30/03/2012	120
			1850	Implementação de componentes	15/04/2012	1200
qui	Quincas	Programador	1850	Implementação de componentes	15/04/2012	2400

Alocação de Tarefas para Membros de um Projeto

Relação Universal

Membro			Tarefa			
IdMembro	Nome	Papel	Tarefa			
			IdTarefa	Descricao	DataInicio	HorasAlocadas
mel	Melissa	Gerente	1700	Planejamento e Orçamento	15/01/2012	80
mel	Melissa	Gerente	1701	Projeto do Sistema	15/02/2012	120
asd	Asdrúbal	Analista	1701	Projeto do Sistema	15/02/2012	180
asd	Asdrúbal	Analista	1705	Especificação da Arquitetura	01/03/2012	120
asd	Asdrúbal	Analista	1705	Detalhamento de Modelos	30/03/2012	200
dor	Doriana	Programador	1705	Detalhamento de Modelos	30/03/2012	120
dor	Doriana	Programador	1850	Implementação de componentes	15/04/2012	1200
qui	Quincas	Programador	1850	Implementação de componentes	15/04/2012	2400

Problemas?

Problemas? Redundância

Exercício 1

- Quais as vantagens e desvantagens de redundância de informação?

Problemas

- Redundância

- “raiz de vários males associados com esquemas relacionais”(Ramakrishnan, 2003)

- Medidas de Qualidade

- Semântica clara dos atributos
- Reduzir informações redundantes
- Reduzir nulos
- Não permitir a geração de tuplas espúrias

(Elmasri, 2011)

Exercício 2

- Considerando a tabela abaixo, dê exemplos de problemas causados pela redundância com:
 - inserção
 - exclusão
 - alteração

Táxi (PMDA)

<u>Placa</u>	Marca	Modelo	AnoFab
DKL4598	Wolksvagen	Gol	2001
DAE6534	Ford	Fiesta	1999
JDM8776	Wolksvagen	Santana	2002
DMZ1122	Wolksvagen	Gol	1995
DKL7878	Ford	Fiesta	2001
JJM3692	Chevrolet	Corsa	1999
DMN1012	Ford	Fiesta	2002

P = Placa
M = Marca
D = Modelo
A = AnoFab

Anomalias

- Inserção
 - Ex.: inserção de membro sem tarefas e vice-versa
- Exclusão
 - Ex.: exclusão de todas as tarefas de um membro
- Alteração
 - Ex.: modificação do nome de um membro em uma das tuplas

Dependência Funcional

- Permite a detecção dos problemas mencionados
- Propriedade definida a partir da semântica dos termos

Dependência Funcional

- “A dependência funcional $X \rightarrow Y$ vale sobre a relação R se, para cada instância possível r de R :
 - $t1 \in r, t2 \in r, \pi_x(t1) = \pi_x(t2)$ implica $\pi_y(t1) = \pi_y(t2)$
 - i.e., dada 2 tuplas em r , se os valores de X são iguais , então os de Y também devem ser.”

(Ramakrishnan, 2003)

Normalização e Formas Normais

- Normalização
 - Se baseia nas formas normais
- Formas normais progressivas
 - 1FN, 2FN, 3FN e 4FN

Primeira Forma Normal (1FN)

- Não contém tabelas aninhadas e atributos multivalorados

Alocação de Tarefas para Membros de um Projeto

Planilha

Membro						
Id Membro	Nome	Papel	Tarefa			
			Id Tarefa	Descrição	Data de Inicio	Horas Alocadas
mel	Melissa	Gerente	1700	Planejamento e Orçamento	15/01/2012	80
			1701	Projeto do Sistema	15/02/2012	120
asd	Asdrúbal	Analista	1701	Projeto do Sistema	15/02/2012	180
			1705	Especificação da Arquitetura	01/03/2012	120
			1705	Detalhamento de Modelos	30/03/2012	200
dor	Doriana	Programador	1705	Detalhamento de Modelos	30/03/2012	120
			1850	Implementação de componentes	15/04/2012	1200
qui	Quincas	Programador	1850	Implementação de componentes	15/04/2012	2400

Membro(IdMembro, Nome, Papel,
Tarefa(IdTarefa, Descricao, DataInicio, HorasAlocadas))

Segunda Forma Normal (2FN)

- Está na 1FN
 - Não contém dependências parciais
 - Coluna que depende de parte da chave primária
- (Heuser, 2004)

Exemplo

IdMembro	IdTarefa	Nome	Papel	Descricao	DataInicio	HorasAlocadas
mel	1700	Melissa	Gerente	Planejamento e Orçamento	15/01/2012	80
mel	1701	Melissa	Gerente	Projeto do Sistema	15/02/2012	120
asd	1701	Asdrúbal	Analista	Projeto do Sistema	15/02/2012	180
asd	1705	Asdrúbal	Analista	Especificação da Arquitetura	01/03/2012	120
asd	1705	Asdrúbal	Analista	Detalhamento de Modelos	30/03/2012	200
dor	1705	Doriana	Programador	Detalhamento de Modelos	30/03/2012	120
dor	1850	Doriana	Programador	Implementação de componentes	15/04/2012	1200
qui	1850	Quincas	Programador	Implementação de componentes	15/04/2012	2400

Terceira Forma Normal (3FN)

- Está na 2FN
- Não contém dependências transitivas
 - $X \rightarrow Y$ e $Y \rightarrow Z$

Exemplo

Táxi (PMDA)

<u>Placa</u>	Marca	Modelo	AnoFab
DKL4598	Wolkswagen	Gol	2001
DAE6534	Ford	Fiesta	1999
JDM8776	Wolkswagen	Santana	2002
DMZ1122	Wolkswagen	Gol	1995
DKL7878	Ford	Fiesta	2001
JJM3692	Chevrolet	Corsa	1999
DMN1012	Ford	Fiesta	2002

P = Placa
M = Marca
D = Modelo
A = AnoFab

Se estiver garantido que não existem dois carros de mesmo Modelo em Marcas diferentes:

Dependência Funcional (DF): $D \rightarrow M$

Decomposição

Decomposição em: PDA e DM

<u>Placa</u>	Modelo	AnoFab
DKL4598	Gol	2001
DAE6534	Fiesta	1999
JDM8776	Santana	2002
DMZ1122	Gol	1995
DKL7878	Fiesta	2001
JJM3692	Corsa	1999
DMN1012	Fiesta	2002

Modelo	Marca
Gol	Wolkswagen
Fiesta	Ford
Santana	Wolkswagen
Corsa	Chevrolet

P = Placa
M = Marca
D = Modelo
A = AnoFab

Exercício 3

- Normalize a tabela abaixo. Faça o modelo ER para as versões normalizada e desnormalizada da tabela. Qual é o problema em um projeto que modela a versão desnormalizada do ER?

<u>Placa</u>	Marca	Modelo	AnoFab
DKL4598	Wolksvagen	Gol	2001
DAE6534	Ford	Fiesta	1999
JDM8776	Wolksvagen	Santana	2002
DMZ1122	Wolksvagen	Gol	1995
DKL7878	Ford	Fiesta	2001
JJM3692	Chevrolet	Corsa	1999
DMN1012	Ford	Fiesta	2002

Quarta Forma Normal (4FN)

- Está na 3FN
- Não contém dependências multivaloradas

Exemplo

- Livro(ISBN, autores, assuntos)
(Guimarães, 2003)
- Relação entre três entidades:
 - Utilizacao(Projeto, CodEmp, CodEquip)
(Heuser, 2004)

Outras Formas Normais

- Boyce/Codd Normal Form (BCNF)
- Quinta Forma Normal (5FN)

Decomposição

- Relação $P \rightarrow P\text{MDA}$
 - Se
 - $P \rightarrow \text{DA}$
 - $D \rightarrow M$
 - Pode ser decomposta em
 - PMDA
 - DM

P = Placa
M = Marca
D = Modelo
A = AnoFab

Decomposição Problemas

- “Algumas queries se tornam muito caras
- Dadas as instâncias das relações decompostas, nós podemos não conseguir reconstruir a instância correspondente da relação original
- A verificação de algumas dependências podem exigir a junção de instâncias das relações decompostas”

(Ramakrishnan, 2003)

Exemplo de Decomposição

Táxi (PMDA)

<u>Placa</u>	Marca	Modelo	AnoFab
DKL4598	Wolkswagen	Gol	2001
DAE6534	Ford	Fiesta	1999
JDM8776	Wolkswagen	Santana	2002
DMZ1122	Wolkswagen	Gol	1995
DKL7878	Ford	Fiesta	2001
JJM3692	Chevrolet	Corsa	1999
DMN1012	Ford	Fiesta	2002

P = Placa
M = Marca
D = Modelo
A = AnoFab

Se estiver garantido que não existem dois carros de mesmo Modelo em Marcas diferentes:

Dependência Funcional (DF): $D \rightarrow M$

Exemplo de Decomposição (cont.)

Decomposição em: PDA e DM

<u>Placa</u>	Modelo	AnoFab
DKL4598	Gol	2001
DAE6534	Fiesta	1999
JDM8776	Santana	2002
DMZ1122	Gol	1995
DKL7878	Fiesta	2001
JJM3692	Corsa	1999
DMN1012	Fiesta	2002

Modelo	Marca
Gol	Wolkswagen
Fiesta	Ford
Santana	Wolkswagen
Corsa	Chevrolet

P = Placa
M = Marca
D = Modelo
A = AnoFab

Decomposição 'sem perda na junção'

$\pi_{\text{Placa, Modelo, AnoFab}}$ (PMDA)

<u>Placa</u>	Marca	Modelo	AnoFab
DKL4598	Wolkswagen	Gol	2001
DAE6534	Ford	Fiesta	1999
JDM8776	Wolkswagen	Santana	2002
DMZ1122	Wolkswagen	Gol	1995
DKL7878	Ford	Fiesta	2001
JJM3692	Chevrolet	Corsa	1999
DMN1012	Ford	Fiesta	2002

Decomposição 'sem perda na junção'

$$\pi_{\text{Placa, Modelo, AnoFab}}(\text{PMDA}) = \text{PDA}$$

<u>Placa</u>	Modelo	AnoFab
DKL4598	Gol	2001
DAE6534	Fiesta	1999
JDM8776	Santana	2002
DMZ1122	Gol	1995
DKL7878	Fiesta	2001
JJM3692	Corsa	1999
DMN1012	Fiesta	2002

Decomposição 'sem perda na junção'

$\pi_{\text{Modelo,Marca}}(\text{PMDA})$

<u>Placa</u>	Marca	Modelo	AnoFab
DKL4598	Wolkswagen	Gol	2001
DAE6534	Ford	Fiesta	1999
JDM8776	Wolkswagen	Santana	2002
DMZ1122	Wolkswagen	Gol	1995
DKL7878	Ford	Fiesta	2001
JJM3692	Chevrolet	Corsa	1999
DMN1012	Ford	Fiesta	2002

Decomposição 'sem perda na junção'

$$\pi_{\text{Modelo,Marca}}(\text{PMDA}) = \text{DM}$$

Modelo	Marca
Gol	Wolksvagen
Fiesta	Ford
Santana	Wolksvagen
Corsa	Chevrolet

Decomposição ‘sem perda na junção’

$$\pi_{\text{Placa, Modelo, AnoFab}}^{\text{(PMDA)}} \bowtie \pi_{\text{Modelo, Marca}}^{\text{(PMDA)}} = \text{PMDA}$$

<u>Placa</u>	Modelo	Marca	AnoFab
DKL4598	Gol	Wolksvagen	2001
DAE6534	Fiesta	Ford	1999
JDM8776	Santana	Wolksvagen	2002
DMZ1122	Gol	Wolksvagen	1995
DKL7878	Fiesta	Ford	2001
JJM3692	Corsa	Chevrolet	1999
DMN1012	Fiesta	Ford	2002

Mais sobre Junções Sem Perda

Táxi (PMDA)

<u>Placa</u>	Marca	Modelo	AnoFab
DKL4598	Wolkswagen	Perua	2001
DAE6534	Ford	Fiesta	1999
JDM8776	Wolkswagen	Santana	2002
DMZ1122	Wolkswagen	Perua	1995
DKL7878	Ford	Fiesta	2001
JJM3692	Chevrolet	Perua	1999
DMN1012	Ford	Fiesta	2002

Neste caso, dois carros de mesmo Modelo pertencem a Marcas diferentes.

Mais sobre Junções Sem Perda

Decomposição em: PDA e DM

<u>Placa</u>	Modelo	AnoFab
DKL4598	Perua	2001
DAE6534	Fiesta	1999
JDM8776	Santana	2002
DMZ1122	Perua	1995
DKL7878	Fiesta	2001
JJM3692	Perua	1999
DMN1012	Fiesta	2002

Modelo	Marca
Perua	Wolksvagen
Fiesta	Ford
Santana	Wolksvagen
Perua	Chevrolet

Mais sobre Junções Sem Perda

PDA  DM não corresponde a PMDA

<u>Placa</u>	Modelo	Marca	AnoFab
DKL4598	Perua	Wolkswagen	2001
DKL4598	Perua	Chevrolet	2001
DAE6534	Fiesta	Ford	1999
JDM8776	Santana	Wolkswagen	2002
DMZ1122	Perua	Wolkswagen	1995
DMZ1122	Perua	Chevrolet	1995
DKL7878	Fiesta	Ford	2001
JJM3692	Perua	Chevrolet	1999
JJM3692	Perua	Wolkswagen	1999
DMN1012	Fiesta	Ford	2002

Decomposição Preservando Dependência

Taxi (PMDNAI)

<u>Placa</u>	Marca	Modelo	Nr	AnoFab	Identificação
DKL4598	Wolkswagen	Gol	001	2001	WG001
DAE6534	Ford	Fiesta	002	1999	FF002
JDM8776	Wolkswagen	Santana	003	2002	WS003
DMZ1122	Wolkswagen	Gol	004	1995	WG004
DKL7878	Ford	Fiesta	005	2001	FF005
JJM3692	Chevrolet	Corsa	006	1999	CC006
DMN1012	Ford	Fiesta	007	2002	FF007

P = Placa; M = Marca; D = Modelo; N = Nr; A = AnoFab; I = Identificação

A identificação é composta a partir da Marca e Modelo mais um número sequencial (Nr). Há duas DFs:

$D \rightarrow M$ e $MDN \rightarrow I$

Decomposição Preservando Dependência

Decomposição em: PDNAI e DM

<u>Placa</u>	Modelo	Nr	AnoFab	Identificação
DKL4598	Gol	001	2001	WG001
DAE6534	Fiesta	002	1999	FF002
JDM8776	Santana	003	2002	WS003
DMZ1122	Gol	004	1995	WG004
DKL7878	Fiesta	005	2001	FF005
JJM3692	Corsa	006	1999	CC006
DMN1012	Fiesta	007	2002	FF007

Modelo	Marca
Gol	Wolkswagen
Fiesta	Ford
Santana	Wolkswagen
Corsa	Chevrolet

Para se verificar a DF: MDN → I é necessário realizar uma junção das relações, portanto a decomposição não Preserva a Dependência.

Exercício 4

- Num banco de dados, em quais casos redundância deve ser evitada? Em quais casos ela é necessária? Dê exemplos de aplicações em cada caso.

Referências

- Codd, Edgar Frank (1970) **A relational model of data for large shared data banks**. Communications ACM 13(6), 377-387.
- Elmasri, Ramez; Navathe, Shamkant B. (2005) **Sistemas de Bancos de Dados**. Addison-Wesley, 4ª edição em português.
- Elmasri, Ramez; Navathe, Shamkant B. (2010) **Sistemas de Banco de Dados**. Pearson, 6ª edição em português.
- Guimarães, Célio (2003) **Fundamentos de Bancos de Dados: Modelagem, Projeto e Linguagem SQL**. Editora UNICAMP, 1ª edição.

Referências

- Heuser, Carlos Alberto (2004) **Projeto de Banco de Dados**. Editora Sagra Luzzato, 5^a edição.
- Ramakrishnan, Raghu; Gehrke, Johannes (2003) **Database Management Systems**. McGraw-Hill, 3rd edition.

André Santanchè

<http://www.ic.unicamp.br/~santanche>

License

- These slides are shared under a Creative Commons License. Under the following conditions: Attribution, Noncommercial and Share Alike.
- See further details about this Creative Commons license at: <http://creativecommons.org/licenses/by-nc-sa/3.0/>