

Dependências Funcionais em BDs Relacionais



Melissa Lemos

melissa@inf.puc-rio.br

1

Introdução

- Agrupamento de atributos
 - Até agora assumimos que os atributos são agrupados para formar um esquema de uma relação usando o
 - bom senso do projetista do BD ou
 - pelo mapeamento do esquema especificado em MER para o esquema relacional.
- Agora vamos ver uma maneira formal de quantificar o quanto o agrupamento de atributos em um esquema de uma relação é melhor que outro.

2

Medidas informais de qualidade do projeto do esquema relacional

1. Semântica dos atributos
2. Redução de tuplas com valores redundantes
3. Redução de tuplas com valores nulos
4. Desabilitação da possibilidade de geração de tuplas espúrias

3

Semântica dos atributos

- Ao agrupar os atributos em uma relação, assumimos que um certo significado está associado a estes atributos.
- Ou seja, devemos saber como os valores de atributos de uma tupla se relacionam.

NORMA 1

Projete uma relação tal que seja fácil explicar o seu significado.

Não combine atributos de várias entidades e relacionamentos em uma mesma relação.

4

Semântica dos atributos

- EMPLOYEE (ENAME, SSN, BDATE, ADDRESS, DNUMBER*)
- DEPARTMENT (DNAME, DNUMBER, DMGRSSN*) Chave Estrangeira
- DEPT_LOCATIONS (DNUMBER*, DLOCATION) Atributo Multivalorado
- PROJECT (PNAME, PNUMBER, PLOCATION, DNUM)
- WORKS_ON (SSN*, PNUMBER*, HOURS) Relacionamento M:N

5

Semântica dos atributos

- EMP_DEPT (ENAME, SSN, BDATE, ADDRESS, DNUMBER, DNAME, DMGRSSN) Exemplo de um projeto ruim
 - Mistura atributos de entidades distintas no mundo real (empregado e departamento)
- EMP_PROJ (SSN, PNUMBER, HOURS, ENAME, PNAME, PLOCATION)
 - Mistura atributos de entidades distintas no mundo real (empregado e projeto)

6

Redução de tuplas com valores redundantes

- Minimizar o espaço de armazenamento que as relações ocupam.
- Exemplo:
 - DEPARTMENT
(DNAME, DNUMBER, DMGRSSN*)
 - Os atributos DNAME e DMGRSSN aparecem apenas uma vez
 - X
 - EMP_DEPT
(ENAME, SSN, BDATE, ADDRESS, DNUMBER, DNAME, DMGRSSN)
 - Repetição dos atributos que pertencem ao departamento (DNUMBER, DNAME, DMGRSSN)

7

Redução de tuplas com valores redundantes

- Anomalia de Inserção
 - EMP_DEPT
(ENAME, SSN, BDATE, ADDRESS, DNUMBER, DNAME, DMGRSSN)
 - Ao inserir uma tupla nova em EMP_DEPT, é necessário ter cuidado para caso o departamento já exista em outra(s) tupla(s), seus valores têm que ser idênticos.
 - Isto não seria necessário no outro caso já que só existiria uma tupla com os valores de departamento na tabela DEPARTMENT

8

Redução de tuplas com valores redundantes

- Anomalia de Inserção
 - EMP_DEPT
(ENAME, SSN, BDATE, ADDRESS, DNUMBER, DNAME, DMGRSSN)
 - Ao inserir uma tupla nova em EMP_DEPT, é necessário ter cuidado para caso o departamento já exista em outra(s) tupla(s), seus valores têm que ser idênticos.
 - Isto não seria necessário no outro caso já que só existiria uma tupla com os valores de departamento na tabela DEPARTMENT

9

Redução de tuplas com valores redundantes

- Anomalia de Inserção
 - EMP_DEPT
(ENAME, SSN, BDATE, ADDRESS, DNUMBER, DNAME, DMGRSSN)
 - É difícil inserir um novo departamento que ainda não tem empregado em EMP_DEPT. Teria que colocar valores nulos em todos os atributos de empregado. Isto causa problemas porque SSN é chave primária, e ainda cada tupla deveria corresponder a um empregado e não a um departamento.

10

Redução de tuplas com valores redundantes

- Anomalia de Remoção
 - EMP_DEPT
(ENAME, SSN, BDATE, ADDRESS, DNUMBER, DNAME, DMGRSSN)
 - Ao remover o último empregado alocado a um departamento, a informação deste departamento é removida também do BD.

11

Redução de tuplas com valores redundantes

- Anomalia de Atualização
 - EMP_DEPT
(ENAME, SSN, BDATE, ADDRESS, DNUMBER, DNAME, DMGRSSN)
 - Ao atualizar uma informação do departamento em uma tupla, é necessário atualizar todas as tuplas, senão o BD ficará inconsistente.
 - Exemplo: Atualização do gerente de um departamento.

12

Redução de tuplas com valores redundantes

NORMA 2

Projete os esquemas de uma relação de tal forma que nenhuma anomalia de inserção, remoção e atualização estejam presentes.

Se alguma anomalia estiver presente, deixe-a clara e tenha certeza que os programas que atualizam o BD operam corretamente.

13

Redução de tuplas com valores nulos

- Se muitos atributos não são aplicados a várias tuplas, ficamos com muitos valores nulos.
 - Desperdício de espaço
 - Problemas para o entendimento do significado do valor nulo
 - Este atributo não se aplica a tupla?
 - O valor deste atributo para esta tupla é desconhecido?
 - O valor deste atributo existe mas ainda não foi gravado?

14

Redução de tuplas com valores nulos

NORMA 3

Evite o quanto possível colocar atributos que frequentemente terão valores nulos. Se for inevitável, certifique-se se o valor nulo é a minoria dos casos para as tuplas da relação.

- Exemplo: Se 10% dos empregados tem escritórios, não há muita justificativa para acrescentar o atributo OFFICE_NUMBER na tabela EMPLOYEE. Ao invés disso, crie uma tabela EMP_OFFICES(ESSN, OFFICE_NUMBER) que possui somente as tuplas dos empregados que possuem escritório.

15

Desabilitação da possibilidade de geração de tuplas espúrias

- Exemplo:
 - Utilização das relações:
 - EMP_LOCS (ENAME, PLOCATION)
 - EMP_PROJ1 (SSN, PNUMBER, HOURS, PNAME, PLOCATION)
 - ao invés da
 - EMP_PROJ (SSN, PNUMBER, HOURS, ENAME, PNAME, PLOCATION)
 - Se fizermos a junção natural de EMP_LOCS e EMP_PROJ1, obteremos tuplas com informações erradas porque o atributo de junção PLOCATION não está coerente. PLOCATION não é chave estrangeira em EMP_LOCS e não é chave primária em EMP_PROJ1.

16

Desabilitação da possibilidade de geração de tuplas espúrias

NORMA 4

Projete esquemas de relações que a JUNCTÃO pela condição de igualdade possa ser aplicada em atributos que sejam chaves primárias e estrangeiras de forma que não apareçam resultados espúrios.

Não tenha relações que contenham atributos que se casem, mas que não sejam combinações de chaves estrangeira e primária. Se isto for inevitável, não faça junção entre estes atributos.

17

Dependências Funcionais Definição

- Uma dependência funcional (DF) é uma restrição entre dois conjuntos de atributos do BD.
- Suponha que exista um esquema de BD relacional com n atributos A_1, A_2, \dots, A_N e que o BD seja uma única relação $R = \{A_1, A_2, \dots, A_N\}$.

18

Dependências Funcionais

Definição

- $R = \{A_1, A_2, \dots, A_N\}$.
- Seja X e Y dois conjuntos de atributos que são sub-conjuntos de R
- Uma dependência funcional $X \rightarrow Y$ especifica uma restrição nas possíveis tuplas que formam um estado da relação r de R .
- Os valores do componente Y de uma tupla em r dependem de, ou são determinados por, valores do componente X .
- Ou ainda, os valores do componente X determinam funcionalmente os valores do componente Y .
- Existe uma dependência funcional entre X e Y ou Y é dependente funcionalmente de X .

19

Dependências Funcionais

Exemplo

- EMP_DEPT
(ENAME, SSN, BDATE, ADDRESS, DNUMBER, DNAME, DMGRSSN)
- EMP_PROJ
(SSN, PNUMBER, HOURS, ENAME, PNAME, PLOCATION)
- $SSN \rightarrow ENAME$
 - O valor do número SSN determina unicamente o nome do empregado.
- $PNUMBER \rightarrow \{PNAME, PLOCATION\}$
 - O valor do número do projeto determina unicamente o nome do projeto e sua localização.
- $\{SSN, PNUMBER\} \rightarrow HOURS$
 - A combinação dos valores de SSN e PNUMBER determinam o número de horas que um empregado trabalha em um projeto.

20

Dependências Funcionais

Regras de Inferência

- F é o conjunto de dependências funcionais que são especificados para um esquema de uma relação R .
 - Geralmente o projetista do esquema define as dependências funcionais que são óbvias.
 - Entretanto, existem várias outras dependências funcionais que podem ser deduzidas.
- O conjunto de todas as dependências funcionais é chamado de *fechamento* de F , denotado por F^+ .

21

Dependências Funcionais

Regras de Inferência - Exemplo

- EMP_DEPT
(ENAME, SSN, BDATE, ADDRESS, DNUMBER, DNAME, DMGRSSN)
- $F = \{SSN \rightarrow \{ENAME, BDATE, ADDRESS, DNUMBER\}, DNUMBER \rightarrow \{DNAME, DMGRSSN\}\}$
- Mas é possível inferir que
 - $SSN \rightarrow \{DNAME, DMGRSSN\}$
 - $SSN \rightarrow SSN$
 - $DNUMBER \rightarrow DNAME$

22

Dependências Funcionais

Regras de Inferência - Exemplo

- Para determinar uma forma sistemática de inferir as dependências funcionais é necessário ter um conjunto de regras de inferência que são usadas para inferir novas dependências funcionais a partir das já existentes.
- $F \vdash X \rightarrow Y$ diz que a dependência funcional $X \rightarrow Y$ foi inferida de F .

23

Dependências Funcionais

Regras de Inferência

- $F \vdash X \rightarrow Y$
 - a dependência funcional $X \rightarrow Y$ foi inferida de F .
- RI1 (reflexão):
 - Se $X \supseteq Y$, então $X \rightarrow Y$
 - Um conjunto de atributos sempre se determina ou determina seus sub-conjuntos.
- RI2 (incremental):
 - $\{X \rightarrow Y\} \vdash \{X, Z\} \rightarrow \{Y, Z\}$
 - Adicionar um conjunto de atributos a ambos os lados de uma dependência resulta em uma outra dependência válida.
- RI3 (transitiva):
 - $\{X \rightarrow Y, Y \rightarrow Z\} \vdash X \rightarrow Z$

24

Dependências Funcionais Regras de Inferência

- $F \vdash X \rightarrow Y$
 - a dependência funcional $X \rightarrow Y$ foi inferida de F .
- RI4 (decomposição):
 - $\{X \rightarrow \{Y, Z\}\} \vdash X \rightarrow Y$
- RI5 (união):
 - $\{X \rightarrow Y, X \rightarrow Z\} \vdash X \rightarrow \{Y, Z\}$
- RI6 (pseudo-transitiva):
 - $\{X \rightarrow Y, \{W, Y\} \rightarrow Z\} \vdash \{W, X\} \rightarrow Z$

25

Dependências Funcionais Regras de Inferência

- Armstrong em 1974 mostrou que o conjunto F^+ , que chamamos de fechamento de F , pode ser determinado a partir de F usando somente as regras de inferência RI1, RI2 e RI3.
- Por isso as regras de inferência RI1, RI2 e RI3 são conhecidas como **regras de inferência de Armstrong**.

26

Dependências Funcionais Regras de Inferência

- Seja
 - U um conjunto de atributos
 - F um conjunto de DFs sobre U
 - X um subconjunto de U
- $X^+ = \{A \mid F \vdash X \rightarrow A\}$
 - X^+ é o fecho de X com relação a F , ou
 - X^+ é o conjunto de atributos que são determinados funcionalmente por X baseado em F

27

Dependências Funcionais Regras de Inferência

- Algoritmo para determinação do fechamento X^+ de um conjunto de atributos X com relação a um conjunto de DFs F
 - Passo 0: $X^0 \leftarrow X$
 - Passo 1: Repetir
 - $X^{i+1} \leftarrow X^i \cup \{Z \mid V \rightarrow Z \in F, V \subseteq X^i\}$
 - até $X^{i+1} = X^i$
 - Passo 2: $X^+ = X^i$

28

Dependências Funcionais Exemplo

- EMP_PROJ
(SSN, PNUMBER, HOURS, ENAME, PNAME, PLOCATION)
- $F = \{SSN \rightarrow ENAME, PNUMBER \rightarrow \{PNAME, PLOCATION\}, \{SSN, PNUMBER\} \rightarrow HOURS\}$
- $(PNUMBER)^+ ??$
 - $X^0 \leftarrow \{PNUMBER\}$
 - $X^1 \leftarrow \{PNUMBER, PNAME, PLOCATION\}$
 - $(PNUMBER)^+ \leftarrow \{PNUMBER, PNAME, PLOCATION\}$

29

Dependências Funcionais Exemplo

- EMP_PROJ
(SSN, PNUMBER, HOURS, ENAME, PNAME, PLOCATION)
- $F = \{SSN \rightarrow ENAME, PNUMBER \rightarrow \{PNAME, PLOCATION\}, \{SSN, PNUMBER\} \rightarrow HOURS\}$
- $(SSN)^+ ??$
 - $X^0 \leftarrow \{SSN\}$
 - $X^1 \leftarrow \{SSN, ENAME\}$
 - $(SSN)^+ \leftarrow \{SSN, ENAME\}$

30

Dependências Funcionais

Exemplo

- EMP_PROJ
(SSN, PNUMBER, HOURS, ENAME, PNAME, PLOCATION)
- F = {SSN \rightarrow ENAME,
PNUMBER \rightarrow {PNAME, PLOCATION},
{SSN, PNUMBER} \rightarrow HOURS}
- (SSN, PNUMBER)⁺ ??
 - $X^0 \leftarrow \{\text{SSN, PNUMBER}\}$
 - $X^1 \leftarrow \{\text{SSN, PNUMBER, ENAME}\}$
 - $X^2 \leftarrow \{\text{SSN, PNUMBER, ENAME, PNAME, PLOCATION}\}$
 - $X^3 \leftarrow \{\text{SSN, PNUMBER, ENAME, PNAME, PLOCATION, HOURS}\}$
 - (SSN)⁺ $\leftarrow \{\text{SSN, PNUMBER, ENAME, PNAME, PLOCATION, HOURS}\}$