

Normalização de Esquemas de Banco de Dados

Prof. Carlos Bazilio
carlosbazilio@id.uff.br

Motivação



Aluno (matr, nome, cr, cod)

Curso (cod, nome, carga_horária)

X

Aluno (matr, nome, cr, cod, nome_curso, carga_horária)

Por que um mapeamento é melhor que o outro?

Informalmente temos ...

- Usualmente utilizamos o bom-senso para definir nossos esquemas de banco de dados
- Exemplos:
 - Agrupamos atributos em relações semanticamente pertinentes
 - Reduzimos informação redundante de tuplas
 - Reduzimos a possibilidade de valores NULL em tuplas

Normalização

- Processo de análise formal de relações de forma a se identificar quão bem projetadas estas estão
- Existem várias formas normais (1FN, 2FN, 3FN, ...) e o processo de análise verifica em qual forma normal uma relação se enquadra
- Para adequarmos uma relação a alguma forma normal, usualmente subdividimos (decompomos) a relação
- Para entender este processo, algumas definições precisam ser vistas

Definições

- Dependência Funcional (DF): $X \rightarrow Y$ / X e Y contém conjunto de atributos de R , Y é determinado/definido por X
- DF é uma propriedade da semântica dos atributos
- Exemplos:
 - $\text{matr} \rightarrow \{ \text{nome}, \text{cpf}, \text{cr} \}$
 - $\{ \text{matr}, \text{nome} \} \rightarrow \{ \text{nome}, \text{cpf}, \text{cr} \}$
 - $\text{cpf} \rightarrow \{ \text{matr}, \text{nome}, \text{cr} \}$
 - $\{ \text{matr}, \text{cod_disciplina} \} \rightarrow \text{nota}$

DFs Inferidas

- DFs que são inferidas à partir de outras dependências
- Exemplo: “matr \rightarrow cr” pode ser inferida de “matr \rightarrow { nome, cr }”
- Regras de inferência:

$X \supseteq Y, X \rightarrow Y$	$\{X \rightarrow YZ\} \models X \rightarrow Y \wedge X \rightarrow Z$
$\{X \rightarrow Y\} \models XZ \rightarrow YZ$	$\{X \rightarrow Y, X \rightarrow Z\} \models X \rightarrow YZ$
$\{X \rightarrow Y, Y \rightarrow Z\} \models X \rightarrow Z$	$\{X \rightarrow Y, WY \rightarrow Z\} \models WX \rightarrow Z$

Definições

- O conjunto de todas as dependências DF é chamado de clausura ou fechamento de F, denotado por F^+
- F cobre E (E é coberto por F) se $DF(E) \subseteq F^+$
- $DF(E) \equiv DF(F)$ se $E^+ = F^+$, ou seja, E cobre F e F cobre E
- $A_1A_2\dots A_n \rightarrow B$ é uma DF trivial se B é um dos A's

Definições

- Chave: define funcionalmente todos os outros atributos numa relação
- Superchave: chave que não precisa ser mínima
- Chaves candidatas: possíveis chaves de uma relação
- Chave primária: uma chave candidata escolhida para indexar uma relação
- Chaves secundárias: as chaves que não foram escolhidas
- Atributo primário: membro de alguma chave candidata

Decomposição sem Perda de Informação

Aluno (matr, nome, cr, cod) =>

A1 (matr, nome) e A2 (matr, cr, cod)

X

Aluno (matr, nome, cr, cod) =>

A3 (matr, cr) e A4 (cr, nome, cod)

- A junção natural das decomposições deve recuperar a relação original!*

Formas Normais

- Processo de análise formal de determinados esquemas de relações com base em suas DFs e chaves primárias para alcançar propriedades desejáveis:
 - (1) Minimização de redundância
 - (2) Minimização de anomalias de inserção, exclusão e atualização
- Esquemas que não satisfazem as formas normais são decompostos em relações menores para que sejam reavaliados

Formas Normais

- Esta reestruturação deve garantir 2 propriedades:
 - Junção sem perda ou não aditiva, para geração apenas de tuplas legítimas
 - Preservação da dependência

1ª. Forma Normal (1FN)

- Um atributo só deve conter valores atômicos
- Contra-exemplo:
Aluno (matr, nome, cr, cod, telefones)
- Para adequar esta relação numa 1FN, podemos:
 - Remover o atributo “telefones” da relação e criar uma relação específica para isso
 - Definir uma chave composta <matr, telefone>
 - Definir atributos para armazenar cada telefone

2ª. Forma Normal (2FN)

- Um esquema R está na 2FN se todo atributo não-primário A em R tem df total da chave primária de R
- Contra-exemplo:
 - Aluno (matr, nome, nota, cod, nome_curso, carga_horária)*
 - *matr* → *nome*, onde *nome* é atributo não-primário e *matr* é parte da chave primária que define funcionalmente *nome*
 - *nota* é o único atributo que depende totalmente da chave primária

2ª. Forma Normal (2FN)

Aluno (matr, nome, nota, cod, nome_curso, carga_horária)

- Para transformação da relação acima numa 2FN, podemos decompor a relação em 3:

Aluno (matr, nome)

Curso (cod, nome_curso, carga_horária)

Aluno_Curso (matr, cod, nota)

- Observe que na relação original o relacionamento entre aluno e curso é n:n
- Dica: para uma relação não estar na 2FN esta precisa possuir uma chave composta para que possa haver a dependência parcial da chave!

3^a. Forma Normal (3FN)

- Definição:
 - DF transitiva: $X \rightarrow Y$ é transitiva se $\exists Z$ e $X \rightarrow Z$ e $Z \rightarrow Y$
- Um esquema R está na 3FN se estiver na 2FN e se nenhum atributo não-primário de R não for transitivamente dependente da chave primária.

3ª. Forma Normal (3FN)

- Contra-exemplo:

Aluno (matr, nome, cr, cod, nome_curso, carga_horária)

- Observe que nesta relação um aluno só pode cursar 1 único curso
- Neste caso, temos as seguintes dependências transitivas:
 - matr \rightarrow cod
 - cod \rightarrow { nome_curso, carga_horária }

3ª. Forma Normal (3FN)

Aluno (matr, nome, cr, cod, nome_curso, carga_horária)

- Para adequar a relação acima para 3FN podemos decompô-la em 2 outras relações:

Aluno (matr, nome, cr, cod)

Curso (cod, nome_curso, carga_horária)

3ª. Forma Normal (3FN)

- Outra definição:
 - Uma relação está na 3FN se, para toda df não-trivial $X \rightarrow A$, vale:
 - X é uma superchave, ou;
 - A é um membro de alguma chave
- Para o exemplo dado, a segunda df não-trivial não é válida:
 - $matr \rightarrow cod$
 - $cod \rightarrow \{ nome_curso, carga_horária \}$

BCNF (Forma Normal Boyce-Codd)

- 3FN mais rígida:
 - Se R está na BCNF, então R está na 3FN
 - Se R está na 3FN, então não necessariamente R está na BCNF
- Uma relação R está na BCNF se, sempre que uma df não-trivial $X \rightarrow A$ existir em R, então X será uma superchave de R.
- Ou seja, somente a parte final da última definição da 3FN **não** está presente: “A é um membro de alguma chave”

Exemplo de 3FN que não é BCNF

- Suponha uma relação de Reservas de ingressos de cinemas com os seguintes atributos:
 - título: o nome de um filme
 - cinema: o nome do cinema onde o filme está sendo exibido
 - cidade: a cidade onde o filme será realizado
- Neste exemplo, um filme não é exibido em 2 cinemas numa mesma cidade
- Esta relação tem as seguintes dfs:
 - cinema \rightarrow cidade
 - {título, cidade} \rightarrow cinema

Exemplo de 3FN que não é BCNF

- Identificando chaves na relação
 - Os atributos isoladamente não podem ser chaves
 - {título, cidade} é uma chave
 - {cinema, título} também é uma chave
 - {cidade, cinema} não é chave pelo fato de podermos ter cinemas multiplex
- Com isso, a df não-trival “*cinema* → *cidade*” invalida a BCNF, pois *cinema* não é uma superchave na relação Reservas

Referências

- Fundamentals of Database Systems, Elmasri, Navathe, Pearson
- Database Systems, Hector Garcia-Molina, Jeffrey D. Ullman and Jennifer Widom, Pearson
- <https://holowczak.com/database-normalization/>