

Normalização

14 de outubro de 2020 11:08

Normalização: é um processo que consiste em estruturar a informação em tabelas na forma mais adequada a fim de evitar:

- Redundâncias desnecessárias;
- Problemas associados à inserção, eliminação e atualização de dados.

Problemas de Redundância de Dados:

- Armazenamento redundante
 - Mesmos dados gravados em vários locais
 - Menos espaço disponível para gravar outros dados
- Anomalias - Incoerências que podem existir aquando da escrita de dados
 - Inserção
 - Pode não ser possível inserir dados, sem serem fornecidos outros, não relacionados
 - Uma alternativa seria usar NULL nos outros dados, mas nem sempre é possível
 - Atualização
 - Pode existir uma incoerência nos dados se apenas uma das cópias for atualizada
 - Eliminação
 - Pode não ser possível apagar dados, sem apagar outros, não relacionados

Processos de Normalização:

- Baseada nas Dependências funcionais (DFs);
- Garante consistência na construção do sistema: redução de anomalias.
- redução de redundância;
- Existem algumas regras para a normalização da base de dados.
 - Cada regra é chamada de " **FORMA NORMAL (FN)**".
 - Condição usando chaves e DFs de uma relação para certificar se um esquema de relação está numa forma normal específica

Do processo de normalização emergem **três tipos de dependências entre os dados: funcionais, multivalor e de junção.**

As dependências funcionais **referem-se à semântica dos dados e não ao seu conteúdo.**

Dependências Funcionais:

- Numa relação R, diz-se que o atributo y é funcionalmente dependente de x ($x, y \in R$), se e apenas se, em qualquer instante, cada valor de x em R tem associado apenas um valor de y em R

Uma dependência funcional para R é uma expressão da forma R:

$X \rightarrow Y$, onde X e Y são conjuntos de atributos de R



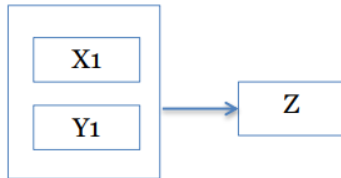
Onde y depende funcionalmente de x.

Nota:

A chave primária de uma relação determina sempre os restantes atributos, isto é, todos eles são dependentes funcionalmente da chave.

Dependência funcional Total:

- Numa relação R, o atributo y é funcionalmente dependente total de x ($x, y \in R$), no caso de x ser um atributo composto, se e apenas se, é funcionalmente dependente de x e não é funcionalmente dependente de qualquer subconjunto dos atributos de x



Dependência Funcional Transitiva:

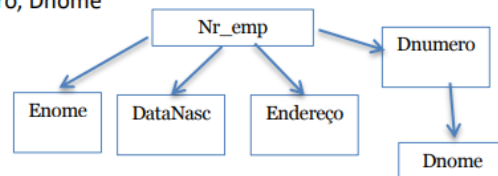
- Uma dependência funcional $R: x \rightarrow y$ é transitiva, se existe um atributo z que não é um subconjunto de x, tal que $x \rightarrow z$ e $z \rightarrow y$

Exemplo: Considere o seguinte esquema com suas dependências funcionais:

Empregado= Nr_emp, Enome, DataNasc, Endereço, Dnumero, Dnome

DF's: Nr_emp \rightarrow Enome, DataNasc, Endereço, Dnumero

Dnumero \rightarrow Dnome



A dependência funcional $\text{Nr_emp} \rightarrow \text{Dnome}$ é transitiva para Dnumero, pois ambas as dependências $\text{Nr_emp} \rightarrow \text{Dnumero}$ e $\text{Dnumero} \rightarrow \text{Dnome}$ são asseguradas e **Dnumero não é nem chave primária nem um subconjunto da chave da relação.**

Dependências Multivalor:

- A dependência multivalor apenas se verifica em casos em que a relação tem pelo menos 3 atributos.
 - Numa relação R, o atributo y tem uma dependência funcional multivalor relativamente a x ($x, y \in R$), se para cada par de tuplos de R contendo os mesmos valores de x, também existe um par de tuplos de R correspondentes à troca dos valores de y no par original

Os tuplos que não têm valores repetidos, satisfazem por redução esta regra

Consideremos a relação $R = \{a, b, c\}$

➤ Existem 2 dependências multivalor

➤ $R: a \twoheadrightarrow b$

➤ $R: a \twoheadrightarrow c$

a	b	c
x1	y1	z1
x1	y1	z2
x1	y2	z1
x1	y2	z2
x3	y1	z1
x4	y3	z2

Dependência de junção:

- Uma dependência de junção numa relação só existe quando, dadas algumas projeções sobre a relação, apenas é possível reconstruir a relação inicial através de algumas junções bem específicas, mas não de todas

Consideremos a relação $R = \{a, b, c\}$ e três projecções:

- $P1 = \{a, b\}$, $P2 = \{a, c\}$, $P3 = \{b, c\}$
- Se não é possível reconstruir a relação com:
 - $P1$ e $P2$
 - $P2$ e $P3$
 - $P1$ e $P3$
- E o for, por exemplo, apenas com $P1$, $P2$ e $P3$...
- Diz-se que R possui uma dependência de junção

Regras de Inferência de DF's :

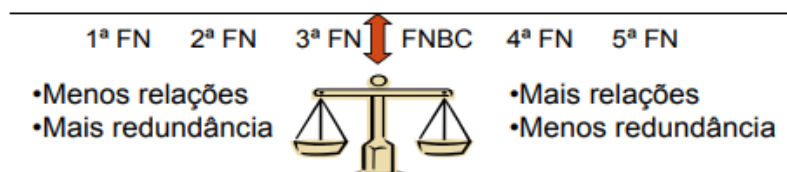
Dada uma relação R com um conjunto U de atributos e algumas dependências funcionais, é possível inferir outras dependências funcionais (triviais ou derivadas) usando os axiomas de Armstrong

Axiomas de Armstrong:

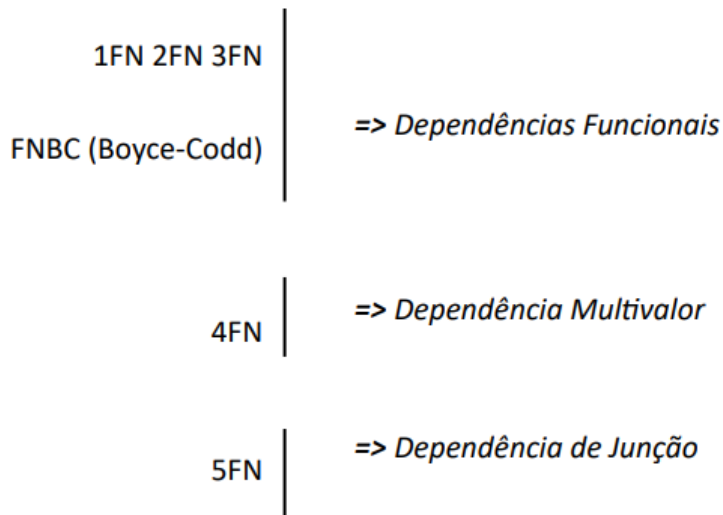
- | | |
|-------------------------|---|
| – União | – Se $X \rightarrow Y$ e $X \rightarrow Z$, então $X \rightarrow YZ$ |
| – Decomposição | – Se $X \rightarrow YZ$, então $X \rightarrow Y$ e $X \rightarrow Z$ |
| – Transitividade | – Se $X \rightarrow Y$ e $Y \rightarrow Z$, então $X \rightarrow Z$ |
| – Pseudo-transitividade | – Se $X \rightarrow Y$ e $WY \rightarrow Z$ então $XW \rightarrow Z$ |
| – Extensão (Aumento) | – Se $X \rightarrow YZ \subseteq U$, então $XZ \rightarrow YZ$ |
| – Reflexividade | – Se $X \supseteq Y$, então $X \rightarrow Y$ |

Aprofundando a Normalização:

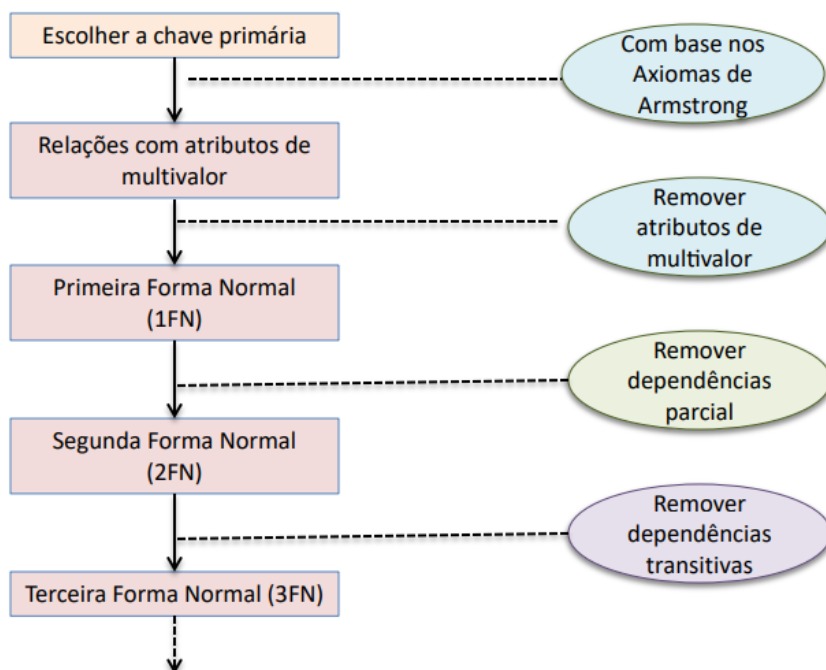
- Com base nas dependências funcionais, multivalor e de junção define-se o processo de normalização de dados aplicado ao modelo relacional.
- A hierarquia é composta por cinco formas normais (1a, 2a, 3a, 4a e 5a Forma Normal) e uma intermédia (Forma Normal de Boyce-Codd, entre a 3a e a 4a).
- Na prática, não deve ser levada às ultimas consequências, pois a proliferação de relações pode conduzir à deterioração do desempenho da Base de Dados.
- Na maioria dos casos opta-se por uma solução de compromisso entre a 3a Forma Normal e a Forma Normal de Boyce Codd.



Formas Normais:



Processo:



Como determinar a chave primária a partir de DF's :

- a) Seja a Relação $R(A,B,C,D)$ e as seguintes DF : $B \rightarrow D$ e $AB \rightarrow C$
 - A chave primaria da relação é **AB**.

Aplicou-se os axiomas de Armstrong.

1. Aumento à DF $B \rightarrow D \Rightarrow AB \rightarrow AD$
2. União $AB \rightarrow C$ e $AB \rightarrow AD \Rightarrow AB \rightarrow CD$

- b) Seja a Relação $R(A,B,C,D,E)$ e as seguintes DF : $AB \rightarrow CE$; $E \rightarrow AB$ e $C \rightarrow D$
 - As chaves candidatas da relação é **AB** e **E**.

Aplicou-se os axiomas de Armstrong.

1. Decomposição $AB \rightarrow CE \Rightarrow AB \rightarrow C$ e $AB \rightarrow E$
2. Transitividade $AB \rightarrow C$ e $C \rightarrow D \Rightarrow AB \rightarrow CD$
3. Transitividade $AB \rightarrow CD$ e $AB \rightarrow E \Rightarrow AB \rightarrow CDE$

ou

1. Decomposição $AB \rightarrow CE \Rightarrow AB \rightarrow C \text{ e } AB \rightarrow E$
2. Transitividade $AB \rightarrow C \text{ e } C \rightarrow D \Rightarrow AB \rightarrow CD$
3. Transitividade $E \rightarrow AB \text{ e } AB \rightarrow CD \Rightarrow E \rightarrow ABCD$

Uma relação está na 1FN se:

- Os atributos chave estão definidos
- Não existem grupos repetidos
- Todos os atributos estão definidos em domínios que contêm apenas valores atômicos, isto é, cada atributo só pode admitir valores elementares e não conjunto de valores
- Todos os atributos dependem funcionalmente da chave primária
- Visa eliminar a existência de grupos de valores repetidos --> A uma ocorrência da chave só pode corresponder uma ocorrência dos outros atributos não chave

Uma relação está na 2FN se:

- Estiver na 1FN
- Cada atributo não chave depende funcionalmente da totalidade da chave
 - Não existem dependências parciais
 - Todos os atributos que não pertencem à chave dependem funcionalmente da chave no seu conjunto e
 - Não dependem de nenhum dos seus elementos ou subconjuntos tomados isoladamente

Conversão da estrutura para a 2 FN :

- Se a relação só tem um atributo como chave primária e se essa relação já estiver na 1FN, então a relação também se encontra na 2FN
- Se a chave primária é composta e se algum atributo não-chave depende apenas de uma parte da chave primária, então a relação deverá ser decomposta, para que cada atributo dependa da totalidade da chave primária

Uma relação está na 3FN se:

- Estiver na 2FN
- Nenhum dos seus atributos depende funcionalmente de atributos não chave
 - Nenhum dos atributos que não fazem parte da chave pode ser funcionalmente dependente de qualquer combinação dos restantes
 - Cada atributo depende apenas da chave e não de qualquer outro atributo ou conjunto de atributos

Conversão da estrutura para a 3FN:

1. Procurar dependências funcionais entre os atributos não-chave da relação
2. Se a relação que já está na 2FN e tiver apenas um atributo não-chave, então a relação também já se encontra na 3FN
3. Se existir algum conjunto de atributos não-chave na relação que tenha dependência funcional em relação a um outro conjunto de atributos não-chave da mesma relação, então a relação deve ser decomposta de modo a que qualquer atributo não-chave da relação só dependa da chave primária da relação

Boyce Codd

Uma relação está na forma normal de Boyce-Codd, se e apenas se, todos os seus atributos são funcionalmente dependentes da chave, de toda a chave e nada mais do que a chave

Consideremos a relação:

$$R = \{\underline{a}, b, c\}$$

E as dependências funcionais em R:

$$R: (a, b) \rightarrow c$$

$$R: c \rightarrow b$$

- R está na 3ª FN, mas tem uma dependência que invalida a forma normal de Boyce-Codd
Podia resolver-se criando duas relações:

R1 = {c, b} correspondente à dependência funcional R: $c \rightarrow b$

R2 = {a, c} correspondente à dependência funcional R: $(a, b) \rightarrow c$

... mas na verdade perdia-se a dependência funcional R: $(a, b) \rightarrow c$, que não se encontrando explicitamente incorporada no modelo relacional **teria de ser implementada no nível aplicacional !**

O ideal será então obter uma solução que, embora mais redundante, mantém todas as dependências funcionais, ou seja, não normalizar até Boyce-Codd...

$$R = \{\underline{a}, b, c\} \text{ e } R1 = \{\underline{c}, b\}$$

4ª Forma Normal (4FN)

- Uma relação está na 4ª forma normal, se está na Boyce-Codd, e se não existem dependências multivalor

5ª Forma Normal (5FN)

- Uma relação encontra-se na 5FN se não existem dependências de junção.
 - Verificam-se em situações muito raras e difíceis de detetar
 - Exige que se compreenda bem a semântica da relação

Conclusões:

- A 4ª e 5ª formas normais são raras e difíceis de detetar
- Frequentemente considera-se que uma relação na 3ª forma normal ou Boyce-Codd está num nível de normalização aceitável
- O nível de normalização deve ser pensado contra outros critérios
 - Por exemplo, um nível de normalização exagerado pode originar problemas de performance
- A redundância entre os dados não pode ser completamente eliminada
 - de facto, as chaves estrangeiras são também uma forma de redundância
- Problemas que a redundância pode trazer
 - Custo de espaço de armazenamento - a redundância implica ocupar espaço adicional com algo que não acrescenta nada ao que já existe armazenado
 - Manutenção - uma simples alteração ou remoção pode implicar o acesso a várias tabelas, tornando-se difícil manter a coerência dos dados armazenados
 - Desempenho - Se a redundância for significativa, isso implicará mais acessos a disco para trazer os mesmos dados