



Instituto de Tecnologia - ITec
Ciência da Computação

Banco de Dados

Normalização de Dados



- ❑ Como avaliar a qualidade do Projeto de um Banco de Dados?
 - ❑ Análise Informal (princípios de BD); Semântica de atributos; Prevenção de anomalias; Redução de valores nulos; Análise Formal/Normalização.
- ❑ **Normalização** é um conjunto de regras (formas normais).
 - ❑ procedimento que examina os atributos de uma entidade com o objetivo de evitar anomalias que possam ocorrer na inclusão, exclusão ou na alteração de uma ocorrência específica.
 - ❑ Decomposição para esquemas menores que satisfaçam as propriedades desejáveis.
- ❑ **Objetivos:**
 - ❑ Reduzir as redundâncias e dependência de dados
 - ❑ Garantir integridade, evitando que informações sem sentido sejam inseridas
 - ❑ Reduzir a necessidade de reestruturar as relações quando novos tipos de dados são introduzidos



- ❑ A normalização é necessária (embora não suficiente) a um bom projeto de banco de dados relacional
- ❑ Na prática normalização é uma ferramenta usada para verificar se o projeto satisfaz algumas regras básicas do Modelo Relacional
- ❑ São **CINCO formas normais** (+ FN Boyce/Codd), porém na prática considera-se normalizada uma entidade/tabela que atenda às **três primeiras** formas normais.
 - ❑ Uma forma normal engloba todas as anteriores, ou seja, para que uma tabela esteja na 2^a FN deve estar na 1^a FN e assim por diante



Dependência Funcional (DF)

- ❑ O atributo **B** é funcionalmente dependente do Atributo **A** se, em qualquer instante, um valor de **A** determina, de modo único, o valor correspondente em **B**, na mesma relação.
 - ❑ Exemplo: Empregado(Numero#, Nome, Salario)
 - ❑ Nome é funcionalmente dependente do Numero, pelo fato de cada Numero estar associado sempre ao mesmo Nome.
- ❑ Para representar dependência funcional usamos:
 - ❑ **Numero -> Nome** (Nome depende funcionalmente de Numero) ou Numero funcionalmente determina Nome





Dependência Funcional Total (DFt)

- Os atributos de uma tabela devem depender da chave primária e somente dela
- Um Atributo **C** é dependente funcional total da chave primária composta pelos atributos **A** e **B**, quando for funcionalmente dependente de **A** e **B** e não dependente de qualquer parte da chave primária (ou seja A ou B)
 - Exemplo: Alocação(NumEmp@#, CodProj@#, QdtHorasTrab)
 - A quantidade de horas trabalhadas é determinada, de modo único, pela composição da matrícula do empregado e do código do projeto, porque significa a quantidade de horas trabalhadas pelo empregado em determinado projeto



Dependência Funcional Parcial(DFp)

- O atributo **C** é parcialmente dependente da chave primária, composta pelos atributos **A** e **B** quando for funcionalmente dependente de **A** **ou** **B**, e não em ambos (A e B).
 - Exemplo: Material (CodMat#, CodForn@#, NomeForn, PrecoMat)
 - O atributo NomeForn é funcionalmente dependente somente do atributo CodForn. Independentemente dos materiais fornecidos.
- Isto consiste em uma anomalia no banco de dados que deve ser evitada. Solução:
 - Material (codMat#, PrecoMat, codForn@)
 - Fornecedor (codForn#, NomeForn)





Dependência Funcional Transitiva(DFT)

- O atributo **C** é dependente funcional transitivo de **A** se **C** é funcionalmente dependente de **B** e **B** funcionalmente dependente de **A**, na mesma relação.
 - Exemplo: Empregado (Num#,Nome,DataAdm,CodProj,DataTermProj)
 - O atributo DataTermProj é dependente do atributo CodProj, que por sua vez, é funcionalmente dependente do atributo Num. Então DataTermProj é dependente transitivos de Num.
- A dependência funcional transitiva é uma anomalia que deve ser evitada.
 - Solução:** cria uma nova relação que será composta pelo atributo(s) que são dependentes funcionais transitivos tendo como chave primária o atributo que determina a transitividade CodProj
 - Emprego(Num#,Nome,DataAdm,CodProj@)
 - Projeto(CodProj#, DataTermProj)



Primeira Forma Normal (1^a FN)

- Um esquema está na 1^a FN se todos os seus atributos forem atômicos (não possuem valores que formam atributos compostos) e monovalorados
 - quando todas as repetições de informação foram removidas e enviadas para uma nova tabela
- Exemplo:
 - ALUNO = {MATRÍCULA# + NOME + ENDEREÇO}
 - ENDEREÇO é um atributo composto, ENDEREÇO = {RUA/NUMERO + BAIRRO + CIDADE + UF}.
- Para colocar na 1FN ?
 - ALUNO = {MATRÍCULA#+NOME+RUA/NUMERO+BAIRRO+CIDADE+UF}





Qualidade dos esquemas relacionais

Tabela: Empregado

Matr#	Nome	Cargo	NomeCargo	Projeto#	DataFim	Horas
120	João	1	Programador	1	17/03/20	32
120	João	1	Programador	8	12/02/20	12
121	Hélio	1	Programador	1	17/03/20	44
121	Hélio	1	Programador	8	12/04/20	21
121	Hélio	1	Programador	12	21/04/20	28
270	Gabriel	2	Analista	8	12/04/20	10

- A chave primária para a tabela Empregados é (**Matr, CodProj**)
- Nesse exemplo, a tabela está na 1ª FN, mas não ocorreu redução da redundância de dados
 - Precisamos realizar outros passos de normalização para termos um bom projeto.



Qualidade dos esquemas relacionais

Matr #	Nome	CodCargo	NomeCargo	Projeto #	DataFim	Horas
120	João	1	Programador	1	17/03/20	32
120	João	1	Programador	8	12/02/20	12
121	Hélio	1	Programador	1	17/03/20	44
121	Hélio	1	Programador	8	12/04/20	21
121	Hélio	1	Programador	12	21/04/20	28
270	Gabriel	2	Analista	8	12/04/20	10

Anomalias da 1FN:

- Inserção: não podemos inserir um empregado sem que este esteja alocado num projeto, nem um projeto sem que haja um empregado trabalhando nele.
- Remoção: se precisarmos remover um projeto, as informações de empregados que estiverem lotados apenas naquele projeto serão perdidas.
- Atualização: se um empregado for promovido de cargo teremos que atualizar os atributos CodCargo e NomeCargo em todas as tuplas(linhas) nas quais aquele empregado está presente.



Segunda Forma Normal (2^a FN)

- Um esquema está na 2^a FN quando, além de encontrar-se na primeira forma normal, cada coluna não chave depende da chave primária completa.
- Para que uma relação esteja na 2FN havendo uma chave primária composta, todos os dados que não são chaves devem depender de toda chave primária (não apenas de parte da chave).
- Exemplo: relação que represente o estoque de um estabelecimento
 - ESTOQUE = {PRODUTO# + ALMOX # + END_ALMOX + UNID_EST + QTD + PRECO}
 - Não está na 2FN porque alguns dados não dependem somente de parte da chave, como END_ALMOX depende só de ALMOX, e UNID_EST depende só de PRODUTO. Com a normalização ficaria?
 - Estoque = {PRODUTO# + UNID_EST}
 - Almoxarifado = {ALMOX# + END_ALMOX}
 - Produto = {PRODUTO# + ALMOX@ + QTD + PRECO}



- 2FN: Exemplo 2 => TABELA Estudante-Disciplina

E #	Enome	Sexo	Idade	D #	Dnome	Opinião
E1	João	M	25	D1	Mat	Boa
E1	João	M	25	D2	Quim	Má
E1	João	M	25	D3	Fis	Boa
E2	Maria	F	22	D2	Quim	Satisf
E2	Maria	F	22	D3	Fis	Satisf
E2	Maria	F	22	D4	Est	Má
E3	João	M	27	D2	Quim	Boa
E3	João	M	27	D3	Fis	Boa

- Conclusão: Enome, sexo, Idade e Dnome devem ser retirados de ESTUDANTE-DISCIPLINA





2FN: Exemplo 2 => Estudante-Disciplina

ESTUDANTE			
E #	Enome	Sexo	Idade
E1	João	M	25
E2	Maria	F	22
E3	João	M	27

DISCIPLINA	
D #	Dnome
D1	Mat
D2	Quim
D3	Fis
D4	Est

AVALIAÇÃO		
E #@	D #@	Opinião
E1	D1	Boa
E1	D2	Pobre
E1	D3	Boa
E2	D2	Satisfatória
E2	D3	Satisfatória
E2	D4	Pobre
E3	D2	Boa
E3	D3	Boa



Terceira Forma Normal (3^a FN)

Um esquema está na 3FN quando, além de estar na 2FN, toda coluna não chave depende diretamente de chave primária, isto é, quando não há dependências funcionais *transitivas* ou *indiretas*.

Exemplo: relação - dados referentes às músicas:

MÚSICA = {CODIGO# + TITULO + GENERO + PAIS_ORIGEM}

Supondo que o PAÍS_ORIGEM refere-se ao GÊNERO musical e não a música, apesar de estar na 2FN, a relação não está na 3FN, pois existe dependência entre GÊNERO e PAÍS_ORIGEM.

Com a normalização ficaria:

MÚS_1 = {CODIGO# + TITULO + GENERO@}

MÚS_2 = {GENEROS# + PAIS_ORIGEM}



3FN – Exemplo 2:

Empregado

Matrícula #	Nome	CodCargo	NomeCargo
120	João	1	Programador
121	Hélio	1	Programador
270	Gabriel	2	Analista

Assim além da tabelas **Projeto** e **Alocação**, teremos a tabela **Cargo** com alteração da tabela Empregado

EMPREGADO		
Matrícula #	Nome	CodCargo @
120	João	1
121	Hélio	1
270	Gabriel	2

CARGO	
CodCargo #	Nome
1	Programador
2	Analista

RESUMO

- Uma relação está em 3FN se todas as colunas da tabela são funcionalmente dependentes da chave inteira e nada além da chave.
- Existem outros casos especiais que requerem mais níveis de normalização: **Boyce-Codd**, **4FN** e **5FN**.
 - Livro: Projeto de Banco de Dados (Carlos A. Heuser)
- Normalmente após a aplicação das regras de normalização, algumas tabelas acabam sendo divididas em duas ou mais.
- Este processo colabora significativamente para a estabilidade do modelo de dados e reduz consideravelmente as necessidades de manutenção e melhor o desempenho do SGBD na recuperação de informações

