

#### **Campus Igarassu**

#### Banco de Dados Normalização

Milton Secundino de Souza Júnior – Msc milton.junior@igarassu.ifpe.edu.br

- Um dos objetivos principais do gerenciamento de Bancos de Dados é manter a integridade dos dados nele armazenados, e para esse fim, algumas regras precisam ser consideradas;
- Algumas dessas regras são garantidas pelo próprio gerenciador
  - tais como unicidade da chave, ligação entre relações através da chave estrangeira, etc.
- Dutras regras são definidas nos programas de aplicação que ficam responsáveis por mantê-las.

- È necessário que as relações sejam bem fundamentadas, no sentido de evitar redundâncias que possam gerar, entre outros problemas, inconsistência de dados;
- Para procurar garantir esse aspecto, foi desenvolvida uma técnica chamada Normalização.

- Redundância é a causa de vários problemas com esquemas relacionais:
  - armazenamento redundante, anomalias de inserção, de exclusão e de atualização;
- Restrições de integridade podem ser usadas para identificar esquemas com esses problemas e para sugerir refinamentos.

- Principal técnica de refinamento: a decomposição de um esquema em subesquemas;
- A decomposição deve ser usada cuidadosamente:
  - Há motivos para se decompor uma relação?
  - A decomposição pode causar problemas?

Considere o esquema:

```
Pacientes (Id, Nome, Endereço, Telefone, Sexo, Data_nascimento, Sigla_convênio, Nome_convênio, Endereço_convênio, Telefone_convênio)
```

- Esse é um exemplo de mau projeto!
  - Os dados de pacientes e os de convênios não deveriam estar na mesma tabela.

Pacientes (Id, Nome, Endereço, Telefone, Sexo, Data\_nascimento, Sigla\_convênio, Nome\_convênio, Endereço\_convênio, Telefone\_convênio)

#### Por que?

- Os dados de um convênio (nome, endereço e telefone do convênio) são repetidos para cada paciente associado a esse convênio;
- Por exemplo, os dados da UNIMED serão repetidos para cada um de seus associados.

#### Anomalia de Inserção:

- Quando se inserir um paciente é preciso inserir também os dados do convênio, mesmo que já estejam cadastrados;
- Não é possível inserir um convênio sem inserir também um paciente;

#### Anomalia de Exclusão:

Ao se excluir um paciente, se este for o único associado de um convênio então os dados do convênio serão perdidos;

#### Anomalia de Modificação:

Para se modificar os dados de um convênio, é preciso atualizar os mesmos dados em todas as tuplas de pacientes que estejam associados àquele convênio.

- Dependências funcionais (DFs) são restrições de integridade mais gerais que as restrições de chave;
- Exemplo de dependência funcional:

```
{Sigla_convênio} → {Nome_convênio, Endereço_convênio, Telefone_convênio}
```

Leia-se: **Sigla\_convênio** determina funcionalmente Nome\_convênio, Endereço convênio e Telefone convênio.

Significado: "Se duas linhas da tabela Pacientes tiverem o mesmo valor de Sigla\_convênio, então elas tem de ter o mesmo valor de Nome\_convênio, de Endereço\_convênio e de Telefone\_convênio."

▶ Identifique as dependências funcionais:

Cod_Emp	Nome	Cargo	Salario
001	João da Silva	Analista Júnior	1.000,00
003	Maria João	Programador Pleno	1.200,00
006	Joice Casa	Analista Júnior	1.000,00
009	Caio Carvalho	Programador Pleno	1.200,00
002	Carlos Villa	Analista Sênior	2.200,00

Identifique as dependências funcionais:

Num_NF	Cod_Produto	Descricao	Preco	Data	Quantidade
001	PC	Computador	1.500,00	15/04/1998	02
001	IMP	Impressora	500,00	15/04/1998	01
002	PC	Computador	1.500,00	16/04/1998	01
003	PC	Computador	1.500,00	14/04/1998	03

#### **Dependências funcionais**

Num\_NF → Data
Cod\_Produto → Descricao,Preco
Num\_NF,Cod\_Produto → Quantidade

- Uma restrição de chave é um caso especial de DF: a chave determina funcionalmente todos os outros atributos da tabela.
- Como Id é chave da tabela Pacientes, temos que:

{Id} → {Nome, Endereço, Telefone, Sexo, Data\_nascimento, Sigla\_convênio, Nome\_convênio, Endereço\_convênio, Telefone\_convênio}

- Certas DFs causam redundância!
- Por exemplo → Para cada associado de um convênio, os dados do convênio são repetidos na tabela Pacientes;
- A causa desse problema é a DF.

```
{Sigla_convênio} → {Nome_convênio, Endereço_convênio, Telefone_convênio}
```

#### Então...

#### Tudo Claro??



### Projeto de Banco de Dados

- O objetivo do projeto de um BD relacional
  - Gerar um conjunto de esquemas de relações que permitam armazenar informações sem redundância desnecessária;
  - Recuperar informações facilmente.

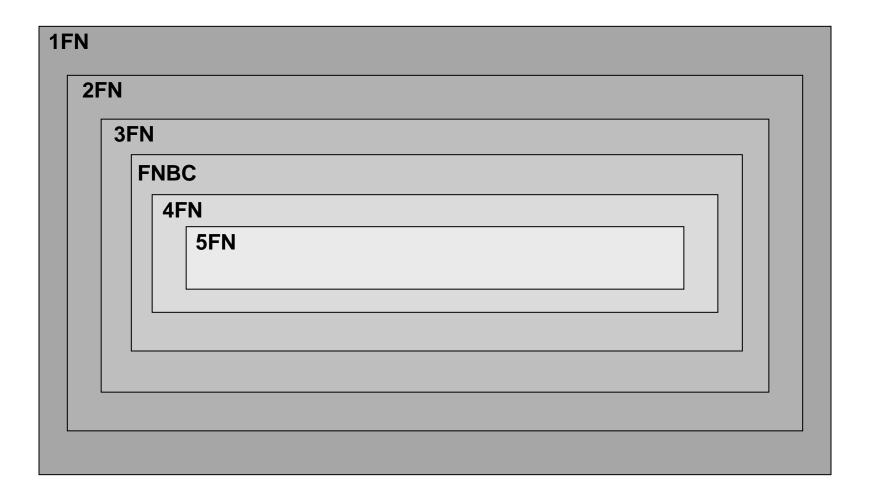
## Projeto de Banco de Dados

- Conjunto de regras que ajudam na definição de bancos de dados que não contenham redundância desnecessária e que permitam o fácil acesso às informações;
- Método permitindo identificar a existência de problemas (anomalias) no projeto de um BD relacional.

- Conjunto de regras que ajudam na definição de bancos de dados que não contenham redundância desnecessária e que permitam o fácil acesso às informações;
- Método permitindo identificar a existência de problemas (anomalias) no projeto de um BD relacional.

- Converte progressivamente uma tabela em tabelas de grau e cardinalidade menores até que pouca ou nenhuma redundância de dados exista:
  - Há diferentes níveis de normalização, de acordo com as condições atendidas;
  - A hierarquia entre as formas normais indica que uma tabela só pode estar numa forma mais avançada se, além de atender as condições necessárias, já estiver na forma normal imediatamente anterior.

# Níveis de Normalização



- Se a normalização é bem sucedida:
  - O espaço de armazenamento dos dados diminui;
  - A tabela pode ser atualizada com maior eficiência;
  - A descrição do BD será imediata.

- A finalidade das regras de normalização é evitar anomalias de atualização no banco de dados:
  - Anomalias de inserção
    - ▶ Evitar a repetição desnecessária de dados (redundância);
  - Anomalias de alteração
    - Evitar inconsistências e reduzir o esforço para a atualização dos dados;
  - Anomalias de exclusão
    - Evitar a perda de informações associadas a um dado registro.

Considere uma única tabela <u>Vendas</u> para representar as informações sobre os negócios de uma loja de CDs:

NOME_CLIENTE	COD_CD	MUSICA	CANTOR	PRECO	DATA_COMPRA
Alice Nóbrega	215621	Bem que se quis	Marisa Monte	R\$ 20,00	21/03/2003
	:	••		•••	•••
Juliano Moreira	878650	Corcovado	Tom Jobim	R\$ 25,00	10/06/2003

- Caso fosse preciso registrar a compra de 5 CDs iguais para um mesmo cliente, as seguintes anomalias seriam observadas:
  - Anomalia de inserção
    - Redundância em todas as colunas (5 linhas iguais na tabela);
  - Anomalia de alteração
    - A mudança no preço do CD deveria ser feita em todas as linhas correspondentes da tabela;
  - Anomalia de exclusão
    - Só haveria registro dos CDs que fossem comprados; se a única venda de um CD fosse apagada, não haveria mais informações sobre aquele CD.

- Conceito: Uma variável de relação (tabela) está em IFN se, e somente se, em todo valor válido dessa variável de relação, cada tupla contém exatamente um valor para cada atributo
- Os atributos devem ser atômicos (indivisíveis)
- Atributos compostos ou multivalorados devem ser representados por novas linhas ou novas tabelas

- Exemplo: Tabela Controle de Faltas numa Escola
- A tabela abaixo não está na I FN

COD TURMA	<u>ALUNO</u>	PROFESSOR	SALA	CAPACIDADE	QTE_FALTAS
BD1032	Alice Luna Juliano Camargo Márcio Andrade	Bruno Pereira	101	50	02 00 04
		•••			•

 Os atributos Aluno e Qte\_Faltas não são atômicos (há mais de um valor para cada registro)

- Passos para obtenção da IFN em uma tabela
  - Identificar a chave primária da tabela;
  - Identificar os atributos compostos ou multivalorados;
  - Incluir uma coluna/linha para cada atributo composto/multivalorado.

A tabela abaixo está na IFN (atributos atômicos)

COD TURMA	<u>ALUNO</u>	PROFESSOR	SALA	CAPACIDADE	QTE_FALTAS
BD1032	Alice Luna	Bruno Pereira	101	50	02
BD1032	Juliano Camargo	Bruno Pereira	101	50	00
BD1032	Márcio Andrade	Bruno Pereira	101	50	04

De próximo passo é observar se ela está também na 2FN

#### Então...

#### Tudo Claro??



- Conceito I: uma variável de relação está em 2FN se, e somente se, ela está em IFN e todo atributo não-chave é irredutivelmente dependente da chave primária;
- Conceito 2: uma variável de relação está em 2FN se, e somente se, ela está em 1FN e, para tabelas com chave primária composta, cada coluna não-chave depende de toda a chave, e não de apenas uma parte dela;
- Dica: tabelas em IFN e com Chave Primária simples estão automaticamente em 2FN.

A tabela abaixo está na IFN mas não está na 2FN:

COD TURMA	<u>ALUNO</u>	PROFESSOR	SALA	CAPACIDADE	QTE_FALTAS
BD1032	Alice Luna	Bruno Pereira	101	50	02
BD1032	Juliano Camargo	Bruno Pereira	101	50	00
BD1032	Márcio Andrade	Bruno Pereira	101	50	04

 Os atributos Professor, Sala e Capacidade dependem apenas de Cod\_Turma (repetição para todos os alunos da turma)

#### Passos para obtenção da 2FN em uma tabela

- Deixá-la em IFN;
- Identificar os atributos que não fazem parte da chave primária da tabela;
- Para cada um desses atributos, analisar se seu valor é determinado por parte ou pela totalidade da chave;
- Criar novas tabelas para os atributos parcialmente dependentes, incluindo a parte da chave correspondente, e retirá-los da tabela original.

As tabelas abaixo estão em 2FN

COD TURMA	ALUNO	QTE_FALTAS
BD1032	Alice Luna	02
BD1032	Juliano Camargo	00
BD1032	Márcio Andrade	04
	•••	

COD TURMA	PROFESSOR	SALA	CAPACIDADE
BD1032	Bruno Pereira	101	50
LG1512	Marina Lucena	101	50
JV8796	Ana Barbosa	101	50
		•••	

#### Então...

#### Tudo Claro??



## Terceira Forma Normal (3FN)

- Conceito I: uma variável de relação está em 3FN se, e somente se, ela está em 2FN e todo atributo não-chave é dependente de forma não transitiva da chave primária;
- Conceito 2: uma variável de relação está em 3FN se, e somente se, ela está em 2FN e todo atributo não-chave depende apenas da chave, e não de outros atributos não-chave;
- Dica: tabelas em 2FN e com nenhum ou um atributo além da chave estão automaticamente em 3FN.

### Terceira Forma Normal (3FN)

A tabela abaixo está em 2FN, mas não está em 3FN

COD TURMA	PROFESSOR	SALA	CAPACIDADE
BD1032	Bruno Pereira	101	50
LG1512	Marina Lucena	101	50
JV8796	Ana Barbosa	101	50
•••			

▶ O atributo Capacidade depende do atributo Sala, e não da chave Cod Turma.

# Terceira Forma Normal (3FN)

#### Passos para obtenção da 3FN em uma tabela

- Deixá-la em 2FN;
- Identificar os atributos que não participam da chave primária da tabela;
- Para cada um desses atributos, analisar se seu valor é determinado por algum outro atributo não pertencente à chave primária;
- Criar novas tabelas para os atributos que não dependem exclusivamente da chave, incluindo o atributo determinante correspondente, e retirá-los da tabela original.

# Terceira Forma Normal (3FN)

As tabelas abaixo estão em 3FN

COD TURMA	<u>ALUNO</u>	QTE_FALTAS
BD1032	Alice Luna	02
BD1032	Juliano Camargo	00
BD1032	Márcio Andrade	04

COD TURMA	PROFESSOR	<u>SALA</u>
BD1032	Bruno Pereira	101
LG1512	Marina Lucena	101
JV8796	Ana Barbosa	101

<u>SALA</u>	CAPACIDADE
101	50
201	40
301	50

## Regras Gerais - Normalização

- IFN: Eliminar atributos multivalorados ou compostos;
- 2FN: Eliminar atributos que dependem apenas de parte da chave primária composta;
- > 3FN: Eliminar atributos que dependem de atributos não-chave.

### Normalização

- Aumentar o nível de normalização contribui para melhorar a qualidade do projeto do banco de dados;
- Há ainda outras três formas normais (FNBC, 4FN e 5FN), cada uma com suas restrições;
- Essas outras formas normais não serão vistas nesse curso.

### Então...

#### Tudo Claro??



#### Exercícios...

- ▶ A tabela a seguir representa as vendas numa loja de CDs;
- Considerando as formas normais vistas (IFN, 2FN e 3FN), indicar quais são atendidas pelo projeto;
- Caso alguma delas não seja atendida, identifique o problema e proponha as mudanças necessárias.

#### Exercício

#### **TABELA VENDAS**

#### **Chave composta**

<u>CLIENTE</u>	COD CD	CANTOR	MUSICA	DURACAO	PRECO	DATA_COMPRA
			Beija Eu	2:20		
	215621	Marisa Monte	Chocolate	3:05	R\$ 20,00	
Alice Nóbrega						21/03/2007
			Corcovado	2:50		
	878650	Tom Jobim	Sabiá	2:10	R\$ 25,00	
			•••			

- A tabela Vendas não está na IFN, pois há vários atributos não atômicos;
- Para deixá-la em IFN, é preciso dividir esses atributos em linhas.

#### **TABELA VENDAS**

<u>CLIENTE</u>	COD CD	CANTOR	MUSICA	DURACAO	PRECO	DATA_COMPRA
Alice Nóbrega	215621	Marisa Monte	Beija Eu	2:20	R\$ 20,00	21/03/2007
Alice Nóbrega	215621	Marisa Monte	Chocolate	3:05	R\$ 20,00	21/03/2007
Alice Nóbrega	878650	Tom Jobim	Corcovado	2:50	R\$ 25,00	21/03/2007
Alice Nóbrega	878650	Tom Jobim	Sabiá	2:10	R\$ 25,00	21/03/2007
	111				•••	•••

- A tabela Vendas não está na 2FN, pois há atributos que dependem apenas de parte da chave primária composta;
- Para deixá-la em 2FN, é preciso criar uma nova tabela

#### **TABELA VENDAS**

CLIENTE	COD CD	DATA_COMPRA
Alice Nóbrega	215621	21/03/2007
Alice Nóbrega	878650	21/03/2007

#### **TABELA CDs**

COD CD	CANTOR	MUSICA	DURACAO	PRECO
215621	Marisa Monte	Beija Eu	2:20	R\$ 20,00
215621	Marisa Monte	Chocolate	3:05	R\$ 20,00
878650	Tom Jobim	Corcovado	2:50	R\$ 25,00
878650	Tom Jobim	Sabiá	2:10	R\$ 25,00

- A tabela CDs não está na 3FN, pois há atributos que dependem de atributos não-chave;
- Para deixá-la em 3FN, é preciso criar uma nova tabela.

#### **TABELA CDs**

COD_CD	CANTOR	MUSICA	PRECO
215621	Marisa Monte	Beija Eu	R\$ 20,00
215621	Marisa Monte	Chocolate	R\$ 20,00
878650	Tom Jobim	Corcovado	R\$ 25,00
878650	Tom Jobim	Sabiá	R\$ 25,00

#### **TABELA MUSICAS**

MUSICA	DURACAO
Beija Eu	2:20
Chocolate	3:05
Corcovado	2:50
Sabiá	2:10

CLIENTE	COD_CD	CANTOR	MUSICA	DURACAO	PRECO	DATA_COMPRA
Alice Nóbrega	215621	Marisa Monte	Beija Eu	2:20	R\$ 20,00	21/03/2007
Alice Nóbrega	215621	Marisa Monte	Chocolate	3:05	R\$ 20,00	21/03/2007
Alice Nóbrega	878650	Tom Jobim	Corcovado	2:50	R\$ 25,00	21/03/2007
Alice Nóbrega	878650	Tom Jobim	Sabiá	2:10	R\$ 25,00	21/03/2007

#### **TABELA VENDAS**

#### **TABELA MUSICAS**

MUSICA	DURACAO
Beija Eu	2:20
Chocolate	3:05
Corcovado	2:50
Sabiá	2:10

# CLIENTE COD CD DATA\_COMPRA Alice Nóbrega 215621 21/03/2007 Alice Nóbrega 878650 21/03/2007 ... ... ...

#### TABELA CDs

COD_CD	CANTOR	MUSICA	PRECO
215621	Marisa Monte	Beija Eu	R\$ 20,00
215621	Marisa Monte	Chocolate	R\$ 20,00
878650	Tom Jobim	Corcovado	R\$ 25,00
878650	Tom Jobim	Sabiá	R\$ 25,00

## Observação

- Veja que as tabelas Vendas, CDs e Musicas já estão em 3FN, mas ainda apresentam algumas redundâncias desnecessárias
  - Cliente e Data\_Compra na tabela Vendas;
  - Cod\_CD, Cantor e Preco na tabela CDs;
- Para resolver esses problemas, seria necessário aplicar outras formas normais.

### Então...

#### Tudo Claro??



#### Exercício

- A tabela a seguir representa os pedidos de produtos de software para uma loja e não obedece nenhuma das formas normais vistas (IFN, 2FN e 3FN);
- Indique os passos para deixá-la em cada uma dessas formas normais.

#### Exercício

#### **TABELA PEDIDOS**

NUM PEDIDO	DATA	FORNECEDOR	CNPJ	ENDERECO	COD PRODUTO	NOME	QUANT	PRECO
					033A	DOS	04	R\$ 130
003	20/01/03	CasaSoftware	8888	R. Lapa, 77	002M	Corel	01	R\$ 499
					145J	ABC	13	R\$ 256
004	27/01/03	BrasilSoftware	5555	Al. Itú, 49	002M	Corel	02	R\$ 450
					083P	ZAPT	10	R\$ 85
					145J	ABC	50	R\$ 110

- Para deixar a tabela em IFN, é preciso dividir os atributos não-atômicos em linhas;
- ▶ A chave da tabela é composta por Num\_Pedido e Cod Produto.

#### **TABELA PEDIDOS**

NUM_PEDIDO	DATA	FORNECEDOR	CNPJ	ENDERECO	COD_PRODUTO	NOME	QUANT	PRECO
003	20/01/03	CasaSoftware	8888	R. Lapa, 77	033A	DOS	04	R\$ 130
003	20/01/03	CasaSoftware	8888	R. Lapa, 77	002M	Corel	01	R\$ 499
003	20/01/03	CasaSoftware	8888	R. Lapa, 77	145J	ABC	13	R\$ 256
004	27/01/03	BrasilSoftware	5555	Al. Itú, 49	002M	Corel	02	R\$ 450
004	27/01/03	BrasilSoftware	5555	Al. Itú, 49	083P	ZAPT	10	R\$ 85
004	27/01/03	BrasilSoftware	5555	Al. Itú, 49	145J	ABC	50	R\$ 110

- Para deixar a tabela em 2FN, é preciso criar novas tabelas para os atributos que dependem apenas de parte da chave primária composta
  - Data, Fornecedor, CNPJ e Endereco dependem apenas de Num\_Pedido;
  - Nome depende apenas de Cod\_Produto;
  - Quantidade e Preco dependem da chave composta.

#### **TABELA PEDIDOS**

NUM_PEDIDO	DATA	FORNECEDOR	CNPJ	ENDERECO
003	20/01/03	CasaSoftware	8888	R. Lapa, 77
004	27/01/03	BrasilSoftware	5555	Al. Itú, 49

#### **TABELA PRODUTOS**

COD_PRODUTO	NOME
033A	DOS
002M	Corel
145J	ABC
083P	ZAPT

#### **TABELA ITENS DE PEDIDOS**

NUM PEDIDO	COD PRODUTO	QUANT	PRECO
003	033A	04	R\$ 130
003	002M	01	R\$ 499
003	145J	13	R\$ 256
004	002M	02	R\$ 450
004	083P	10	R\$ 85
004	145J	50	R\$ 110

- Para deixar as tabelas em 3FN, é preciso criar novas tabelas para os atributos dependentes de atributos nãochave
  - CNPJ e Endereco dependem de Fornecedor

#### **TABELA PEDIDOS**

NUM_PEDIDO	DATA	<u>CNPJ</u>	
003	20/01/03	8888	
004	27/01/03	5555	

#### **TABELA PRODUTOS**

COD_PRODUTO	NOME
033A	DOS
002M	Corel
145J	ABC
083P	ZAPT

#### **TABELA PEDIDOS**

NUM PEDIDO	COD PRODUTO	QUANT	PRECO
003	033A	04	R\$ 130
003	002M	01	R\$ 499
003	145J	13	R\$ 256
004	002M	02	R\$ 450
004	083P	10	R\$ 85
004	145J	50	R\$ 110

CNPJ	FORNECEDOR	ENDERECO
8888	CasaSoftware	R. Lapa, 77
5555	BrasilSoftware	Al. Itú, 49

### Então...

#### Tudo Claro??



#### Referências

Slides compostos a partir de materiais da professora Renata Viegas (UFPB)