


Banco de Dados I

Engenharia Reversa
e Normalização



O que é Engenharia Reversa





Conceito

- A **Engenharia Reversa** é o processo de descobrir os princípios tecnológicos de um dispositivo/objeto ou de um sistema com a análise de suas estrutura, função e operação.
- Diretamente falando a Engenharia Reversa geralmente é a arte de **desmontar** para saber como funciona.



Aplicação

- A **Engenharia reversa** consiste em usar a criatividade para, a partir de uma solução pronta, retirar todos os possíveis conceitos novos ali empregados.
- É o processo de análise de um artefato (um aparelho, um componente elétrico, um programa de computador, etc.) e dos detalhes de seu funcionamento, geralmente com a intenção de construir um novo aparelho ou programa que faça a mesma coisa, sem realmente copiar alguma coisa do original.
- Objetivamente a engenharia reversa consiste em, por exemplo, desmontar uma máquina para descobrir como ela funciona.

Exemplo

Jerry Can (Latas Jerry)



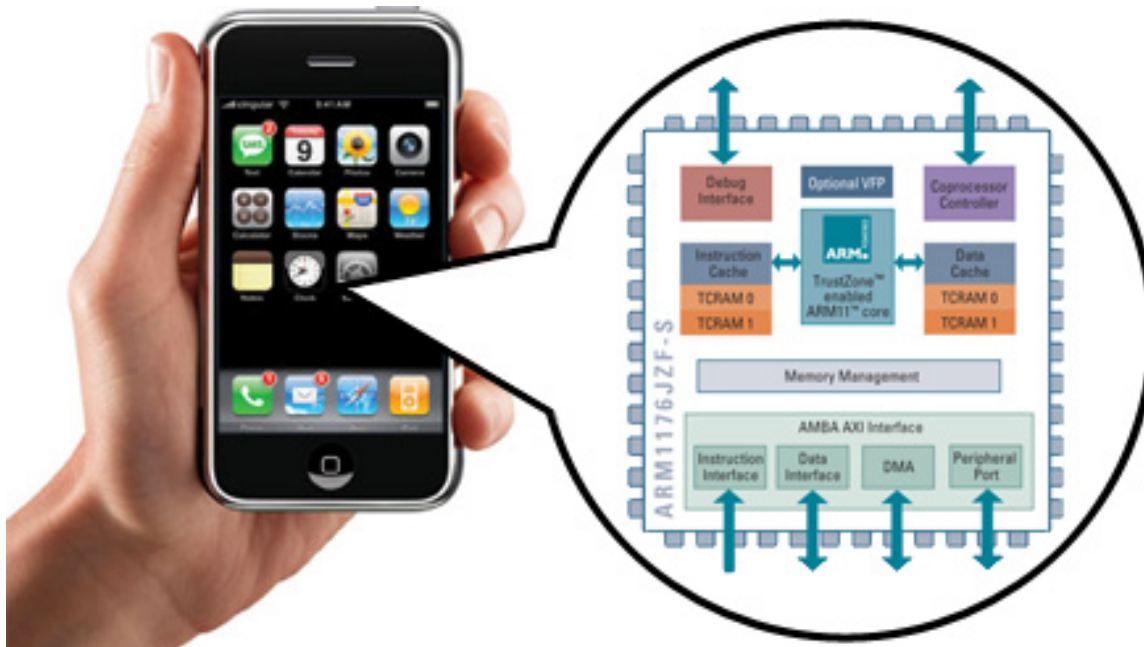
- Criado pelos Alemães na Segunda Guerra Mundial
- Copiada pelas forças britânicas e americanas



Exemplo

- Enquanto a Apple não divulga oficialmente as especificações de hardware do iPhone, entusiastas do novo telefone examinam o hardware através de técnicas de **engenharia reversa**, para descobrir suas especificações e características.

Fonte: <http://www.pdaexpert.net/noticias/computacao-movel/especificacoes-de-hardware-do-iphone-descobertas/>





Engenharia Reversa

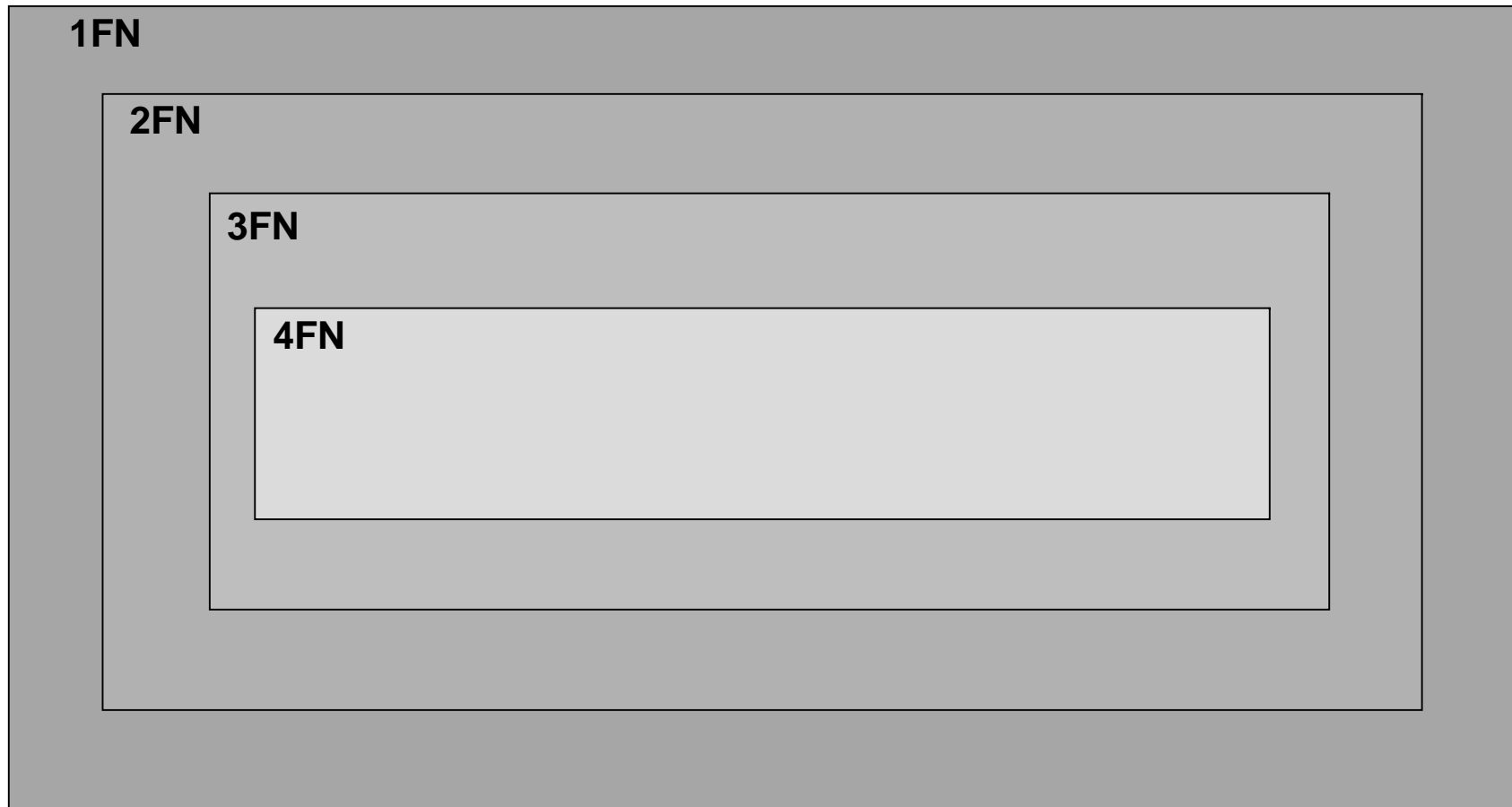
- A Engenharia Reversa utiliza o processo de normalização para os objetivos seguintes:
 - Reagrupar informações de forma a eliminar redundâncias de dados que possam existir nos arquivos;
 - Reagrupar informações de uma forma que permita a obtenção de um modelo ER.



Normalização

- Conjunto de regras que ajudam na definição de bancos de dados que não contenham redundância desnecessária e que permitam o fácil acesso às informações
- 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
 - (Ex: só pode estar em 2FN se já estiver em 1FN)

Níveis de Normalização





Normalização

- 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

Antes

Com anomalias



Depois

Sem anomalias



Exemplo de Anomalias

- Considere uma única tabela Vendas 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 | 21/08/2008 |
| ... | ... | ... | ... | ... | ... |
| Juliano Moreira | 878650 | Corcovado | Tom Jobim | R\$ 25,00 | 10/06/2008 |



Exemplo de Anomalias

- 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

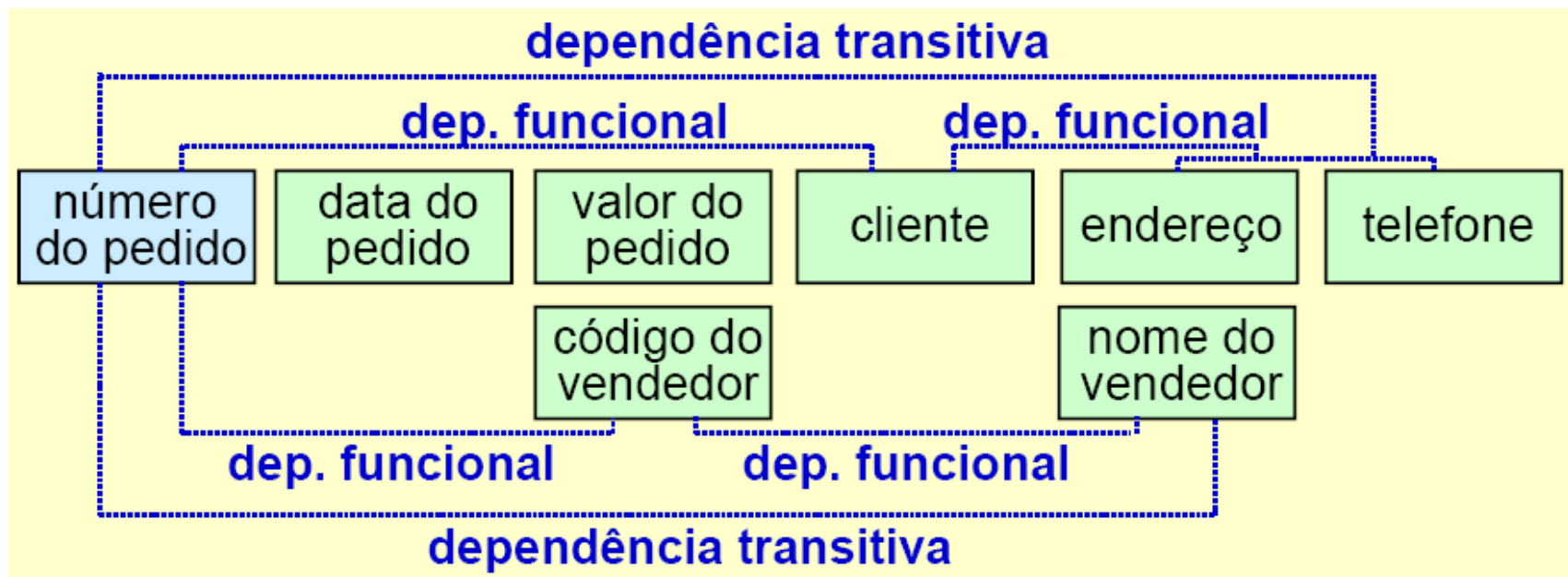
Dependência transitiva

- *Dependência Funcional Transitiva*

Um atributo A tem **dependência funcional transitiva** da chave primária, se ele possui **dependência funcional** em relação a um atributo B que possua **dependência funcional** dessa chave.

Ex: Entidade PEDIDO

• **endereço**, **telefone** e **nome do vendedor** são dependentes transitivos de **código do pedido**.





Primeira Forma Normal (1FN)

- **Conceito:** Uma variável de relação (tabela) está em 1FN 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), não podem conter tabelas aninhadas;
- Atributos compostos ou multivalorados devem ser representados por novas linhas ou novas tabelas.

Exemplo – 1FN

- Exemplo: Tabela Controle de Faltas numa Escola
- A tabela abaixo não está na 1 FN

| COD_TURMA | ALUNO | PROFESSOR | SALA | CAPACIDADE | QTE_FALTAS |
|-----------|-----------------|---------------|------|------------|------------|
| BD1032 | Alice Luna | Bruno Pereira | 101 | 50 | 02 |
| | Juliano Camargo | | | | 00 |
| | Márcio Andrade | | | | 04 |
| ... | ... | ... | ... | | ... |

- Os atributos Aluno e Qte_Faltas não são atômicos (há mais de um valor para cada registro)



Solução

- Construir uma única tabela com redundância de dados

Cria-se uma tabela na qual os dados das linhas externas à tabela aninhada são repetidos para cada linha da tabela aninhada.

- Construir uma tabela para cada tabela aninhada

Cria-se uma tabela referente a própria tabela que está sendo normalizada e uma tabela para cada tabela aninhada.



Exemplo – 1FN

- A tabela abaixo está na 1FN (atributos atômicos)

| COD_TURMA | ALUNO | 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 |
| ... | ... | ... | ... | ... | ... |



Obtenção da 1FN

- Os atributos não contém valores nulos

Do ponto de vista prático, não há restrições quanto à existência de valores nulos em coluna de uma tabela. Valores nulos desconhecidos são comuns, porém valores nulos inaplicáveis podem indicar irregularidades no modelo conceitual e talvez mereçam tabelas distintas, gerando uma especialização.

Obtenção da 1FN

- Passos para obtenção da 1FN em uma tabela
 1. Escolher a chave primária definir as chaves candidatas da tabela.
 2. Transformar os atributos compostos em atômicos.
 3. Eliminar os atributos multivalorados, gerando uma tabela para cada um dos conjuntos de itens repetitivos. A chave primária de cada tabela será a concatenação da chave da tabela original com um atributo da nova tabela que identifique cada linha de forma única.

Cada item repetitivo dará origem a uma linha nesta nova tabela.
- O próximo passo é observar se ela está também na 2FN

Segunda Forma Normal (2FN)

- **Conceito 1:** uma variável de relação está em 2FN se, e somente se, ela está em 1FN 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 1FN e com Chave Primária simples estão automaticamente em 2FN



Exemplo - 2FN

- A tabela abaixo está na 1FN mas não está na 2FN
 - Considere a chave composta (Cod_Turma, Nome)

| COD_TURMA | ALUNO | 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)

Exemplo - 2FN

- 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 |
| ... | ... | ... | ... |



Processo para obtenção da 2FN

- 1) Identificar as *colunas* que não participam da **chave primária** da tabela.
- 2) Para cada uma das colunas identificadas, analisar se existe **dependência parcial** da chave primária.

Para identificar a dependência parcial de uma coluna em relação à chave, deve-se indagar:

Para que o valor da coluna seja determinado, quais as partes da chave que devem ser conhecidas?

- 3) Para as colunas dependentes parcialmente da chave:
 - **Criar** novas tabelas que herdarão a chave parcial e todos os atributos que dependem dessa chave parcial. Essa chave parcial será a **chave primária** da tabela.
 - **Excluir** da tabela original todas as colunas com dependência parcial da chave.



Obtenção da 2FN

- Passos para obtenção da 2FN em uma tabela
 - Deixá-la em 1FN
 - 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



Terceira Forma Normal (3FN)

- **Conceito 1:** 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

Exemplo – 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



Processo para obtenção da 3FN

- 1) Identificar as *colunas* que não participam da *chave primária* da tabela.
- 2) Para cada uma das colunas identificadas, analisar se existe *dependência transitiva* da chave primária.

Para identificar a dependência transitiva de uma coluna deve-se indagar:

Qual outra coluna não pertencente à chave poderia determinar o valor da coluna em análise?

- 3) Para as colunas dependentes transitivamente da chave:
 - *Criar* novas tabelas que herdarão as colunas com dependência transitiva e também a coluna determinante da transitividade. Essa coluna será a *chave primária* da tabela criada.
 - *Excluir* da tabela original todas as colunas com dependência transitiva, *mantendo* a coluna determinante da transitividade.

Excluir também as colunas *derivadas* de outras.



Obtenção da 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

Exemplo - 3FN

- As tabelas abaixo estão em 3FN

| COD_TURMA | ALUNO | QTE_FALTAS |
|-----------|-----------------|------------|
| BD1032 | Alice Luna | 02 |
| BD1032 | Juliano Camargo | 00 |
| BD1032 | Márcio Andrade | 04 |
| ... | ... | ... |

| COD_TURMA | PROFESSOR | SALA |
|-----------|---------------|------|
| BD1032 | Bruno Pereira | 101 |
| LG1512 | Marina Lucena | 101 |
| JV8796 | Ana Barbosa | 101 |
| ... | ... | ... |

| SALA | CAPACIDADE |
|------|------------|
| 101 | 50 |
| 201 | 40 |
| 301 | 50 |
| ... | ... |



Regras Gerais - Normalização

- **1FN**: 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



Observação

- Aumentar o nível de normalização contribui para melhorar a qualidade do projeto do banco de dados
- Geralmente normalizamos até a 3ºFN.

Exercício

- A tabela abaixo representa as vendas numa loja de CDs. Considerando as formas normais vistas (1FN, 2FN e 3FN), indicar quais são atendidas pelo projeto. Caso alguma delas não seja atendida, identifique o problema e proponha as mudanças necessárias.

Chave composta

TABELA VENDAS

| CLIENTE | COD_CD | CANTOR | MUSICA | DURACAO | PRECO | DATA_COMPRA |
|---------------|--------|--------------|-----------------------|--------------|-----------|-------------|
| Alice Nóbrega | 215621 | Marisa Monte | Beija Eu Chocolate | 2:20 3:05 | R\$ 20,00 | 21/03/2003 |
| | 878650 | Tom Jobim | Corcovado Sabiá | 2:50 2:10 | R\$ 25,00 | |
| ... | ... | ... | ... | ... | ... | ... |

Solução – 1FN

- A tabela Vendas não está na 1FN, pois há vários atributos não atômicos
- Para deixá-la em 1FN, é preciso dividir esses atributos em linhas

TABELA VENDAS

| CLIENTE | COD_CD | CANTOR | MUSICA | DURACAO | PRECO | DATA_COMPRA |
|---------------|--------|--------------|-----------|---------|-----------|-------------|
| Alice Nóbrega | 215621 | Marisa Monte | Beija Eu | 2:20 | R\$ 20,00 | 21/03/2003 |
| Alice Nóbrega | 215621 | Marisa Monte | Chocolate | 3:05 | R\$ 20,00 | 21/03/2003 |
| Alice Nóbrega | 878650 | Tom Jobim | Corcovado | 2:50 | R\$ 25,00 | 21/03/2003 |
| Alice Nóbrega | 878650 | Tom Jobim | Sabiá | 2:10 | R\$ 25,00 | 21/03/2003 |
| ... | ... | ... | ... | ... | ... | ... |

Solução – 2FN

- 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 | 21/03/2003 |
| Alice Nóbrega | 878650 | 21/03/2003 |
| ... | ... | ... |

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 |

Solução – 3FN

- 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 |



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 Preço na tabela CDs
- Para resolver esses problemas, seria necessário aplicar outras formas normais

Exercício

- A tabela abaixo representa os pedidos de produtos de software para uma loja e não obedece nenhuma das formas normais vistas (1FN, 2FN e 3FN). Indique os passos para deixá-la em cada uma dessas formas normais.

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 |
| | | | | | 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 1FN, é preciso dividir os atributos não-atômicos em linhas
- A chave da tabela é composta por Num_Pedido e Cod_Produto

TABELA PEDIDOS

[illegible]



Solução - 2FN

- 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 Preço dependem da chave composta

Solução – 2FN

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 |

TABELA PRODUTOS

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

TABELA DADOS_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 |



Solução – 3FN

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

Solução – 3FN

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 |

TABELA PRODUTOS

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

TABELA DADOS_PEDIDOS

| NUM_PEDIDO | DATA | CNPJ |
|------------|----------|------|
| 003 | 20/01/03 | 8888 |
| 004 | 27/01/03 | 5555 |

TABELA FORNECEDORES

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