

NORMALIZAÇÃO

MOTIVAÇÃO

◦ Tipos de chave

- **Chave candidata:**

- Todas as opções de identificação única de um registro;

- **Chave primária:**

- Chave escolhida como identificação única;

- **Chave estrangeira:**

- Chave que representa uma relação com outra tabela, ou seja, representa a chave primária da tabela relacionada.



MOTIVAÇÃO

- Após a construção do **modelo conceitual** dos dados é feita a transformação para um **modelo lógico** (esquema relacional);
- O conjunto de tabelas obtido representa a estrutura da informação de um modo natural e completo;
- Mas, é necessário avaliar o **grau de redundâncias** existente;
- A normalização tem como objetivo modificar o conjunto de tabelas obtido, pela transformação do modelo conceitual em um outro conjunto de tabelas equivalente, menos redundante e mais estável.



MOTIVAÇÃO

- O que são **relações (tabelas)** bem estruturadas?
 - Contem um mínimo de redundância;
 - Permitem aos usuários inserir, modificar e apagar linhas (registros) em tabelas, sem erros ou inconsistências.



RELAÇÕES BEM ESTRUTURADAS

- Observando as duas tabelas a seguir, qual representa uma relação bem estruturada?

| Empregado 1 | | | |
|-------------|-----------|------------------|---------|
| Emp_ID | Nome | Depart | Salário |
| 100 | Margarida | Marketing | 42000 |
| 140 | Hélio | Financeira | 39000 |
| 110 | Cristóvão | Sist. Informação | 41500 |
| 190 | Lourenço | Contabilidade | 38000 |
| 150 | Susana | Marketing | 38500 |

| Empregado 2 | | | | | |
|-------------|-----------|------------------|---------|------------|------------------|
| Emp_ID | Nome | Depart | Salário | Curso | Dta de conclusão |
| 100 | Margarida | Marketing | 42000 | SPSS | 19-06-2002 |
| 100 | Margarida | Marketing | 42000 | Pesquisa | 07-10-2002 |
| 140 | Hélio | Financeira | 39000 | IVA | 08-12-2002 |
| 110 | Cristóvão | Sist. Informação | 41500 | SPSS | 01-12-2002 |
| 110 | Cristóvão | Sist. Informação | 41500 | C++ | 22-04-2002 |
| 190 | Lourenço | Contabilidade | 38000 | Investimen | 07-05-2002 |
| 150 | Susana | Marketing | 38500 | SPSS | 19-06-2002 |
| 150 | Susana | Marketing | 38500 | TQM | 12-08-2003 |

- R: É a 1ª, uma vez que só apresenta informações sobre um empregado e qualquer modificação, como p.ex. o salário, será limitada a uma única linha da tabela;
- A 2ª contém informações sobre os empregados e os cursos que eles frequentaram e a chave primária para esta tabela terá de ser uma combinação entre *EmpID* e *Curso*.



RELAÇÕES BEM ESTRUTURADAS

- Redundâncias em uma tabela podem resultar em erros e inconsistências, quando o usuário tentar atualizar os dados na tabela.
- São três os tipos de erros:
 - **Erros de inserção** – caso se pretenda inserir um novo empregado na tabela 2, como a chave é *Emp_ID* e *Curso*, só é possível efetuar a inserção se ambas existirem;
 - **Erros de remoção** – se, p.ex., os dados do empregado 140 for apagado da tabela, também se perde a informação sobre o curso que esse empregado frequentou. *Será que é isso que se pretende?*
 - **Erros de modificação** – imagine que o empregado 100 tem um aumento de salário. Então esse aumento tem de ser registrado em cada uma das linhas em que a existem ocorrências desse empregado. *Não é muito prático, certo?*



RELAÇÕES BEM ESTRUTURADAS

- Então, a tabela ***Empregado_2*** não é uma relação bem estruturada e pode ser dividida em duas relações:

- Empregado*
- Curso*

| Empregado 2 | | | | | |
|-------------|-----------|------------------|---------|------------|------------------|
| Emp_ID | Nome | Depart | Salário | Curso | Dta de conclusão |
| 100 | Margarida | Marketing | 42000 | SPSS | 19-06-2002 |
| 100 | Margarida | Marketing | 42000 | Pesquisa | 07-10-2002 |
| 140 | Hélio | Financeira | 39000 | IVA | 08-12-2002 |
| 110 | Cristóvão | Sist. Informação | 41500 | SPSS | 01-12-2002 |
| 110 | Cristóvão | Sist. Informação | 41500 | C++ | 22-04-2002 |
| 190 | Lourenço | Contabilidade | 38000 | Investimen | 07-05-2002 |
| 150 | Susana | Marketing | 38500 | SPSS | 19-06-2002 |
| 150 | Susana | Marketing | 38500 | TQM | 12-08-2003 |

- O conceito de **Normalização** tem por base esta explicação e a capacidade de gerar relações bem estruturadas.



NORMALIZAÇÃO

- **Redundância** – armazenamento repetido de dados, que resultam em problemas de:
 - **Manutenção** – alteração ou remoção de registros implica acessos a várias partes da base, tornando-se difícil manter a coerência;
 - **Custo** de espaço de armazenamento;
 - **Desempenho** – as operações de acesso aos discos são as que mais afetam o desempenho dos sistemas.



PROCESSO DE NORMALIZAÇÃO

A Normalização é uma técnica de modelagem de dados, ligada ao modelo relacional, que pretende definir um conjunto de tabelas prontas para serem implementadas numa base de dados relacional.

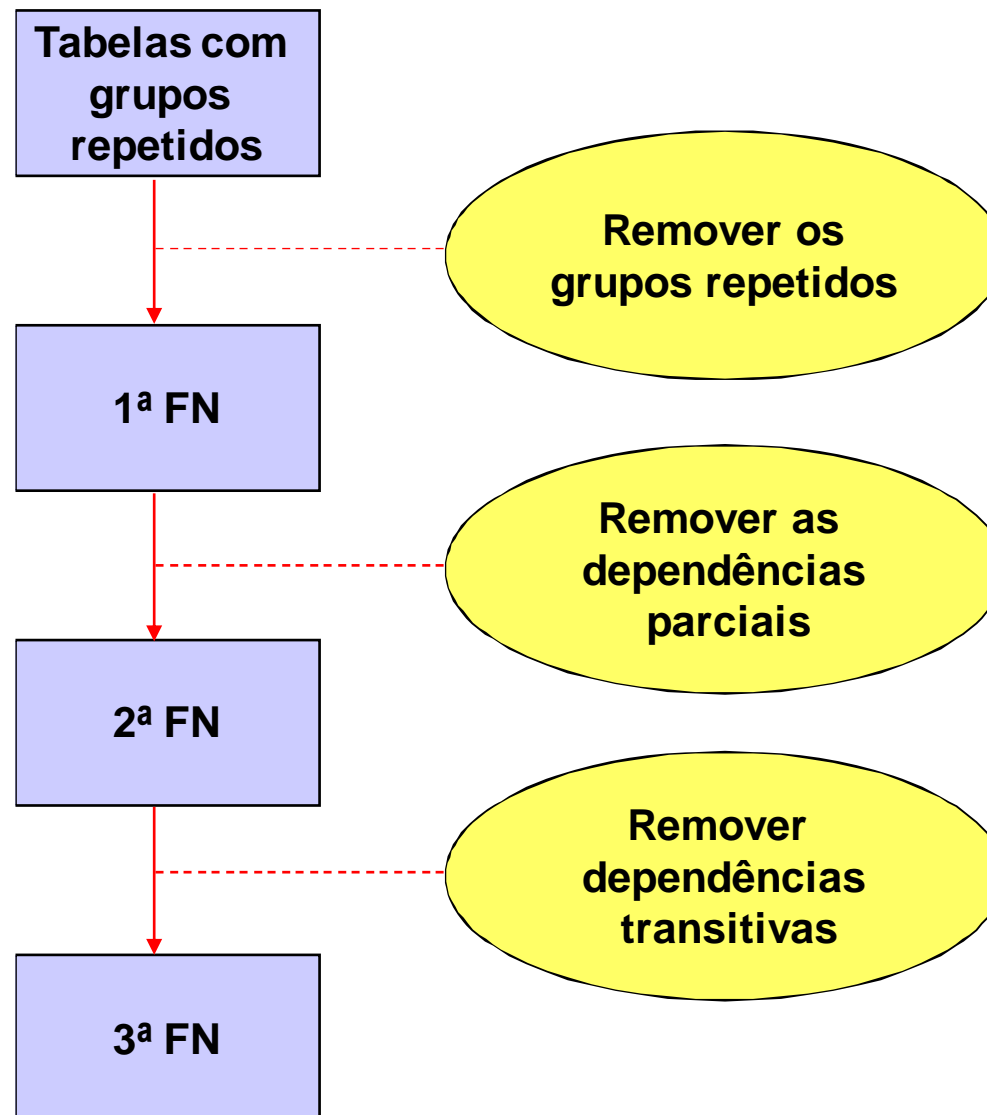
Pode ser aplicada a um conjunto de dados brutos
(1 tabela → relação universal)
ou a partir de estruturas já existentes.



PROCESSO DE NORMALIZAÇÃO



ETAPAS DA NORMALIZAÇÃO



NORMALIZAÇÃO

- A normalização é baseada na análise das **dependências funcionais**:
 - Uma dependência funcional é uma relação particular entre 2 atributos;
 - $A \rightarrow B$, ou seja, **Emp_ID \rightarrow Nome**;
 - B é funcionalmente dependente de A
 - A determina B
 - B depende de A



NORMALIZAÇÃO

- Um atributo pode estar dependente de 2 ou mais atributos, em vez de um único.
- São exemplos:
 - ISBN → Título do livro
 - N_Funcionario → Departamento
 - Se se conhece o *N_Funcionario* (atributo único) é possível determinar o *Departamento* (um funcionário só pode pertencer a um departamento)
 - **Atenção:** A análise das instâncias não é suficiente para avaliar a existência de dependências funcionais. Só o conhecimento do domínio do problema, obtido através da análise de requisitos é um método válido para essa conclusão.



NORMALIZAÇÃO

| N_funcionario | Nome_Proprio | Sobrenome | Departamento |
|---------------|--------------|-----------|--------------|
| 1021 | Sofia | Reis | 900 |
| 1022 | Afonso | Reis | 700 |
| 1023 | Antonio | Cardoso | 900 |

- *Departamento* → *N_funcionario* ?
 - Não, pois *Departamento* 900 \Rightarrow {1021,1023}
- *N_funcionario* → *Departamento* ?
 - Sim, pois se é conhecido o *N_funcionario* (atributo único) é possível determinar o *Departamento* (um funcionário só pode pertencer a um departamento)



NORMALIZAÇÃO

| N_funcionario | Nome_Proprio | Sobrenome | Departamento |
|---------------|--------------|-----------|--------------|
| 1021 | Sofia | Reis | 900 |
| 1022 | Afonso | Reis | 700 |
| 1023 | Antonio | Cardoso | 900 |

- **Nome_proprio \rightarrow N_funcionario ?**
 - Não, pois podem existir funcionários com o mesmo nome \Rightarrow podem haver múltiplos valores de *N_funcionario* para o mesmo *Nome_próprio*
- ***N_funcionario* \rightarrow *Sobrenome*?**
 - Apesar de dois funcionários terem o mesmo sobrenome, se é conhecido o *N_funcionario* determina-se um só *Sobrenome*
- ***N_funcionario* \rightarrow todos os restantes atributos**



FORMAS NORMAIS

Primeira Forma Normal (1FN) **Eliminar redundâncias**

Uma tabela se encontra na 1FN se todas as suas colunas são elementares (**atributos atômicos**).



1ª FORMA NORMAL (ELIMINAR REDUNDÂNCIAS)

Encomendas (#encomenda, n_cliente, cliente, endereço_cliente, dt_encomenda, **cod_produto, produto, qtd_encomendada)**



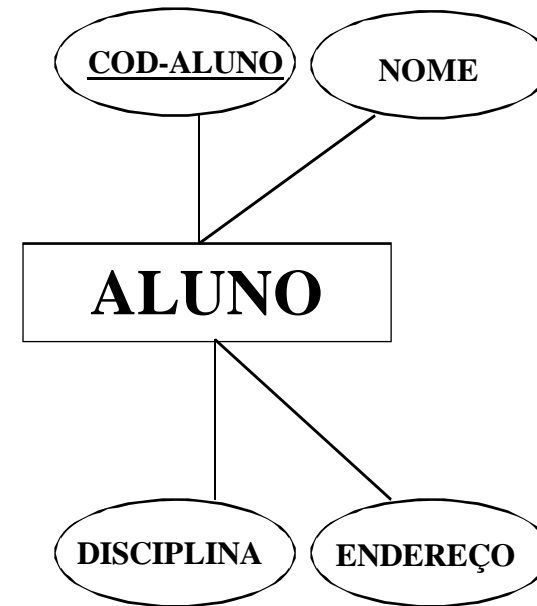
Encomendas (#encomenda, n_cliente, cliente, endereço_cliente, dt_encomenda)

Itens_Encomenda (#encomenda, #item_encomenda, **cod_produto, produto, qtd_encomendada)**

1ª FORMA NORMAL (ELIMINAR REDUNDÂNCIAS)

○ Valores atômicos:

- cidade do cliente (Curitiba)
- Nome do cliente (João Carlos)
- Idade do aluno (30 anos)

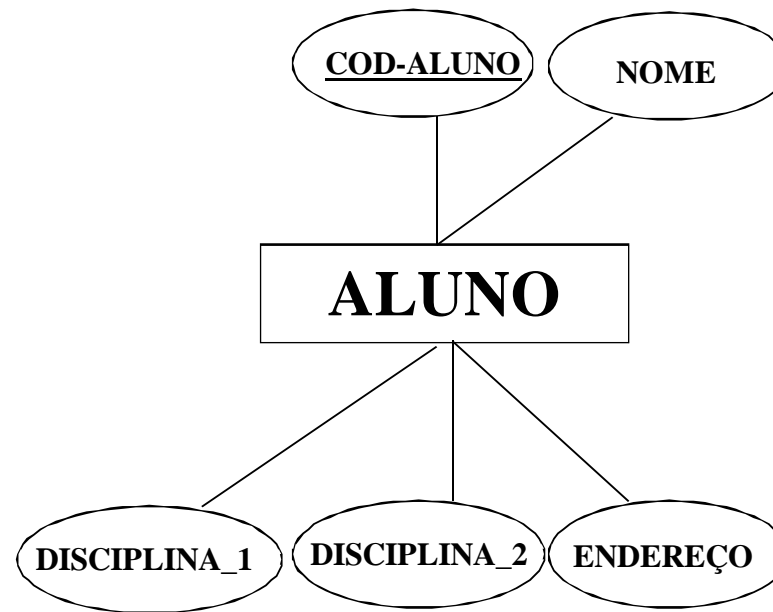


○ Valores não atômicos:

- Disciplina do aluno (matemática, física) é multivalorado, ou seja, o aluno pode estar matriculado em mais de uma disciplina.



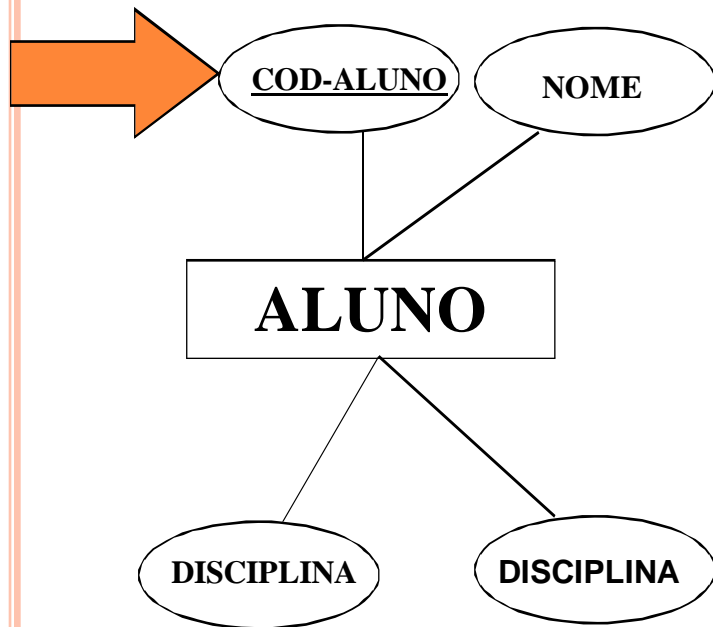
1ª FORMA NORMAL (ELIMINAR REDUNDÂNCIAS)



- Não está na 1FN quando possuir atributos repetidos.



1ª FORMA NORMAL (ELIMINAR REDUNDÂNCIAS)



| <u>CodAluno</u> | Nome | Morada | Disciplina |
|-----------------|-----------|--------|------------|
| 1214 | Rui Costa | Rua A | Português |
| 1214 | Rui Costa | Rua A | Matemática |
| 1214 | Rui Costa | Rua A | Física |
| 1250 | Ana Maria | Rua B | Latim |
| 1250 | Ana Maria | Rua B | Português |
| 1250 | Ana Maria | Rua B | Inglês |

- Não é PK (chave primária), apenas FK (chave estrangeira) ou chave candidata.

1ª FN

- Normalize (1FN) a seguinte relação:

Num_estudante, Nome_estudante, curso,
num_disciplinas, nome disciplinas, cod_professor,
nome_professor, grau_professor, nota

- a) Identificar a chave primária da entidade;
- b) Identificar o grupo repetitivo e removê-lo da entidade;
- c) Criar uma nova entidade com a chave primária da entidade anterior e o grupo repetitivo.



FORMAS NORMAIS

Segunda Forma Normal (2FN)

Dependência total da Chave

O conceito de **dependência funcional** é apresentado considerando uma tabela **R** com a estrutura **R(x,y,z,w)**.

Diz-se que a **coluna z** é funcionalmente dependente da **coluna x** se, para um dado valor da **coluna x**, o valor da **coluna z** é sempre o mesmo.

Pode dizer-se que **x determina z**, ou que **z depende funcionalmente de x**, $x \rightarrow z$.



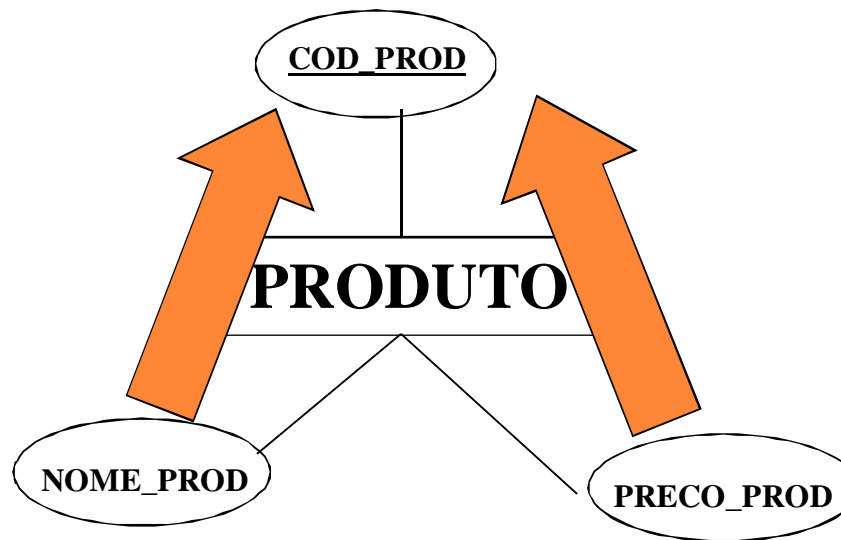
FORMAS NORMAIS

Segunda Forma Normal (2FN) Dependência total da Chave

Uma tabela está na 2FN quando estiver na 1FN e as suas colunas que não são chave primária dependerem inteiramente da chave primária.

DEPENDÊNCIA
FUNCIONAL

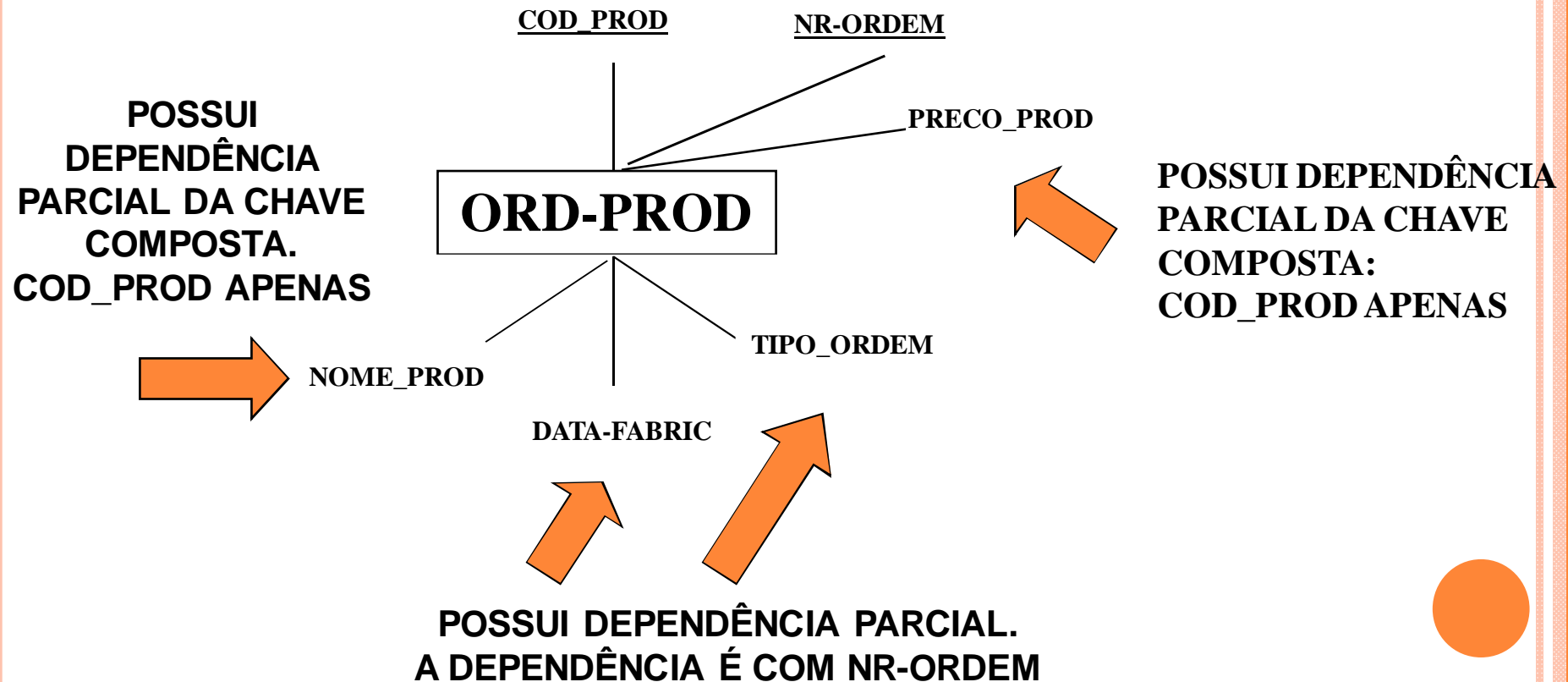
NA CHAVE
PRIMÁRIA



FORMAS NORMAIS

Segunda Forma Normal (2FN) Dependência total da Chave

EXEMPLOS DEPENDÊNCIA NÃO FUNCIONAL NA CHAVE COMPOSTA:



2ª FORMA NORMAL (DEPENDÊNCIA TOTAL DA CHAVE)

**Itens_Encomenda (#encomenda, #item_encomenda,
cod_produto, produto, qtd_encomendada)**



**Itens_Encomenda (#encomenda, #item_encomenda,
cod_produto, qtd_encomendada)**

Produtos (cod_produto, produto**)**



2ª FN

- Aplique a 2ª FN ao exemplo anterior (alunos, cursos etc.)

- a) Identificar os atributos que **não são** funcionalmente dependentes de toda a chave primária;
- b) Remover da entidade todos esses atributos identificados e criar uma nova entidade com eles;
- c) A chave primária da nova entidade será o atributo do qual os atributos removidos são funcionalmente dependentes.



FORMAS NORMAIS

Terceira Forma Normal (3FN) **(Eliminar dependências transitivas)**

Uma tabela está na 3FN quando:

- estiver na 2FN e
- se todos os atributos que não pertencem à chave primária, não dependerem de outro atributo que não a chave.



FORMAS NORMAIS

Terceira Forma Normal (3FN) **(Eliminar dependências transitivas)**

A 3FN impõe que cada coluna dependa apenas da chave primária e não de qualquer outra coluna ou conjunto de colunas.

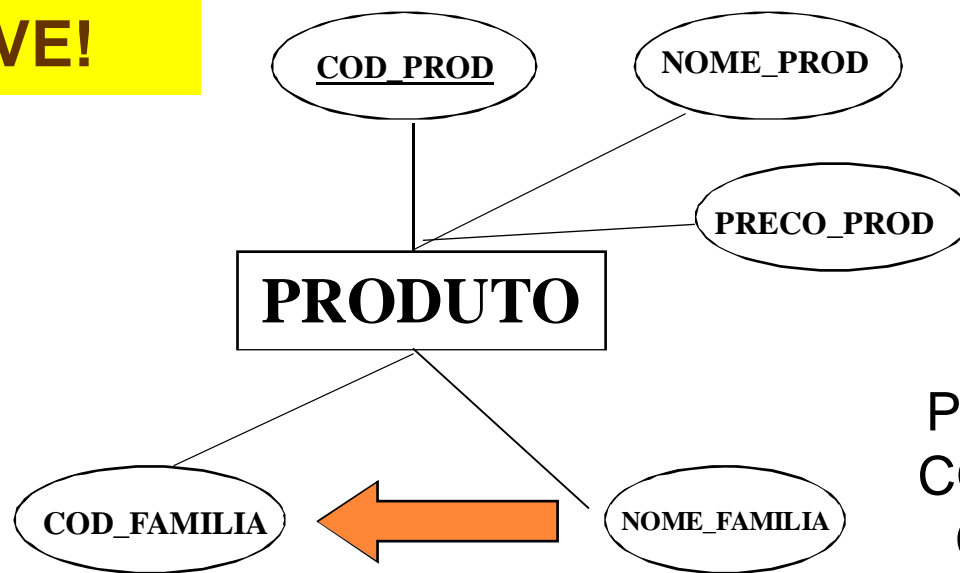
Assim todas as colunas que não pertencem à chave são mutuamente independentes.

Esta condição permite eliminar dependências transitivas.



3ª FORMA NORMAL (ELIMINAR DEPENDÊNCIAS TRANSITIVAS)

**DEPENDÊNCIA
FUNCIONAL
APENAS COM
A CHAVE!**



Possui DF com
COD_FAMILIA e
COD_PROD



3ª FORMA NORMAL (ELIMINAR DEPENDÊNCIAS TRANSITIVAS)

Encomendas (#encomenda, n_cliente, cliente, endereço_cliente, dt_encomenda)



Encomendas (#encomenda, n_cliente, dt_encomenda)

Clientes (n_cliente, cliente, endereço_cliente)



3ª FORMA NORMAL (ELIMINAR DEPENDÊNCIAS TRANSITIVAS)

- Considere a seguinte tabela destinada a armazenar informação sobre jogos de futebol e árbitros:

JOGO (IdJogo, Estádio, Nome_arbitro, Categoria_arbitro).

- A tabela encontra-se na 3FN?
 - a) Identificar os atributos com DF além da chave;
 - b) Retirar estes atributos da entidade;
 - c) Gerar nova entidade com estes atributos e a chave é o atributo que possui a DF.



FORMAS NORMAIS

Cadastro de Funcionário

Funcionário: João da Silva

Endereço: Av. Maringá, S/N

Telefone: 3622-2010

Celular:

Nome dos filhos: Samuel, Maria, Joana

Data de Nascimento dos filhos: 1/01/92, 5/03/94, 10/12/97



Tela do
Sistema de
Controle de
venda

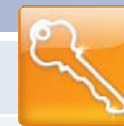


FORMAS NORMAIS

Tabela física do banco de dados


Funcionário

| Atributo | Tipo | Tamanho | Chave |
|------------------|---------|---------|-------|
| Cod_Funcionario | Integer | | PK |
| Nome_Funcionario | Varchar | 50 | |
| End_Funcionario | Varchar | 50 | |
| Telefone | Varchar | 12 | |
| Celular | Varchar | 12 | |
| Nome_filhos | Varchar | 50 | |
| Dt_Nasc_filhos | Varchar | 30 | |



FORMAS NORMAIS

Resultado

| Cod_funcionario | Nome_Funcionario | Nome_Filhos | Dt_Nasc_filhos |
|-----------------|---|----------------------|-----------------|
| 001 |  | João da Silva | Carlos da Silva |
| | | Pedro da Silva | 04/04/1980 |
| 002 | | Maria das Dores | 28/07/1985 |
| 003 | | Carla das Dores | 25/05/1980 |
| | | Claudio Soares | |
| 004 | Jose Sem Sobrenome | Carlos Sem Sobrenome | 10/03/1987 |
| | | Maria Sem Sobrenome | 23/09/1990 |

Esta tabela está mal projetada



FORMAS NORMAIS

Forma Normal de Boyce/Cood (BCFN)

A BCFN é um aperfeiçoamento da 3FN, destinada a lidar com situações em que se verifique a existência de mais do que uma chave candidata e que duas chaves candidatas possuam elementos comuns.

Uma relação está na BCFN quando todos os atributos estão dependentes da chave, de toda a chave e de nada mais do que a chave.



FORMA NORMAL DE BOYCE/COOD (BCFN)

(paciente, serviço) → médico

R(doente, serviço, médico)

(um doente, num determinado serviço hospitalar é sempre observado pelo mesmo médico e um médico só pertence a um e um só serviço)

**R1 (paciente,médico)
R2 (médico, serviço)**

Mas esta solução permite ter um paciente com 2 médicos do mesmo serviço

Tratamento a nível de aplicação, mantendo a relação R e R2 (criando redundância)



FORMA NORMAL DE BOYCE/COOD (BCFN)

- Imagine a relação:

$R(\underline{\text{CodAluno}}, \underline{\text{CodDisciplina}}, \text{Professor})$

na 3FN, onde os alunos frequentam várias disciplinas;

- Cada professor só leciona uma disciplina, mas uma disciplina pode ser lecionada por vários professores;
- Portanto, pode-se verificar a seguinte dependência funcional $(\text{CodAluno}, \text{CodDisciplina}) \rightarrow \text{Professor}$.



FORMA NORMAL DE BOYCE/COOD (BCFN)

- No entanto, existem duas chaves candidatas compostas

(CodAluno, CodDisciplina) e (CodAluno, Professor)

e o atributo *Professor* não é chave candidata mas determina o *CodDisciplina*

(Professor \rightarrow CodDisciplina),

pois um professor só leciona uma disciplina.



FORMAS NORMAIS

Quarta Forma Normal (4FN)

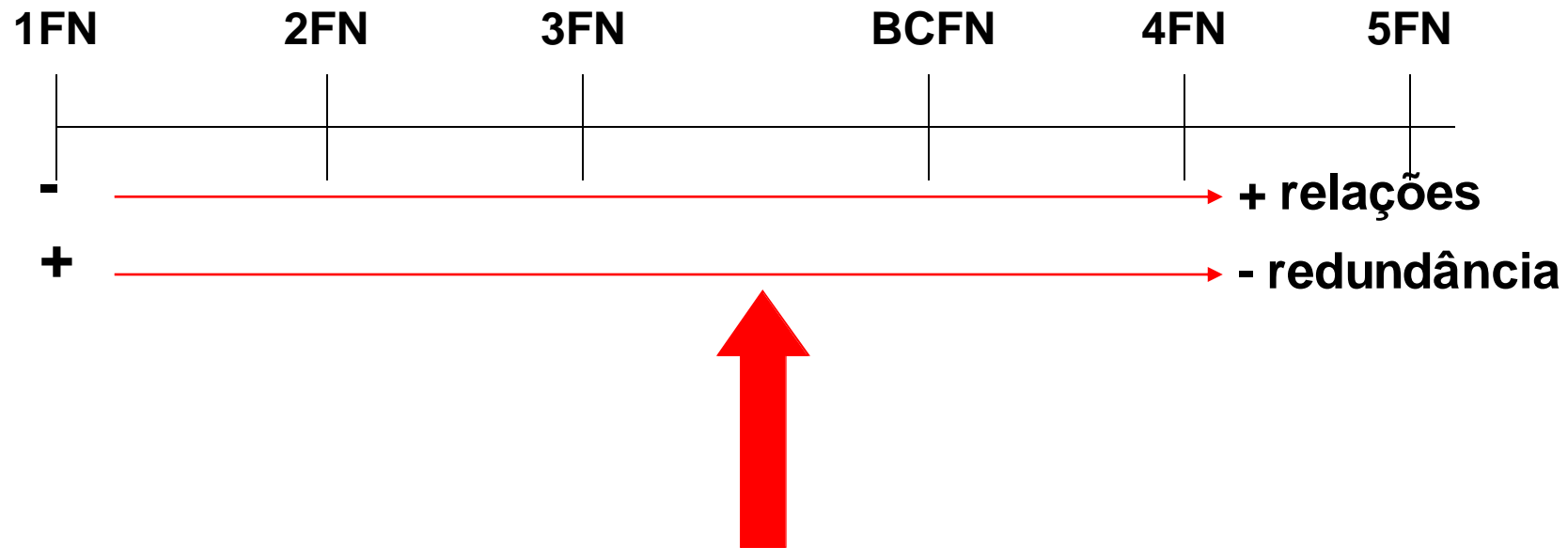
Quinta Forma Normal (5FN)

Em geral uma relação na BCNF já está na 4FN e 5FN, que surgem para resolver problemas muito raros:

- Uma relação encontra-se na 4FN, se está na BCNF e não existem dependências multivalor;
- Uma relação R está na 5FN se não puder ser mais decomposta sem perda de informação.



CONSEQUÊNCIAS DA NORMALIZAÇÃO



Compromisso, bom senso, avaliação do desempenho do sistema e, eventualmente, necessidade de **desnormalizar**



FORMAS NORMAIS

Desnormalização

O SGBD para além de implementar o modelo de dados de um universo, deve considerar alguns aspectos que permitem melhorias na utilização do sistema;

Trata-se de estabelecer um compromisso entre a flexibilidade do sistema e a viabilidade da sua utilização;

Pretende-se obter um esquema equilibrado, que não ponha em risco a integridade da BD, mas que, simultaneamente, tenha um desempenho razoável. Por essa razão, na maioria dos casos, o processo de normalização para na 3FN.

