

BASE DE DADOS

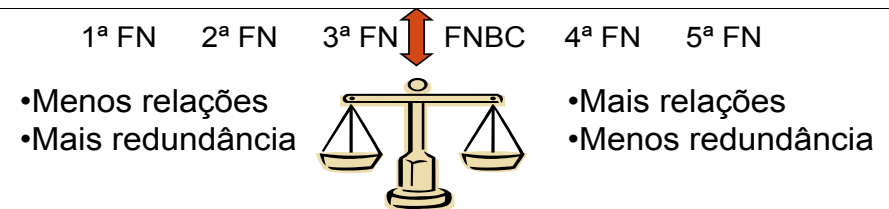


Teórico-Práticas

NORMALIZAÇÃO

Normalização

- Com base nas dependências funcionais, multivalor e de junção define-se o processo de normalização de dados aplicado ao modelo relacional. (já abordado nas aulas teóricas)
- 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.



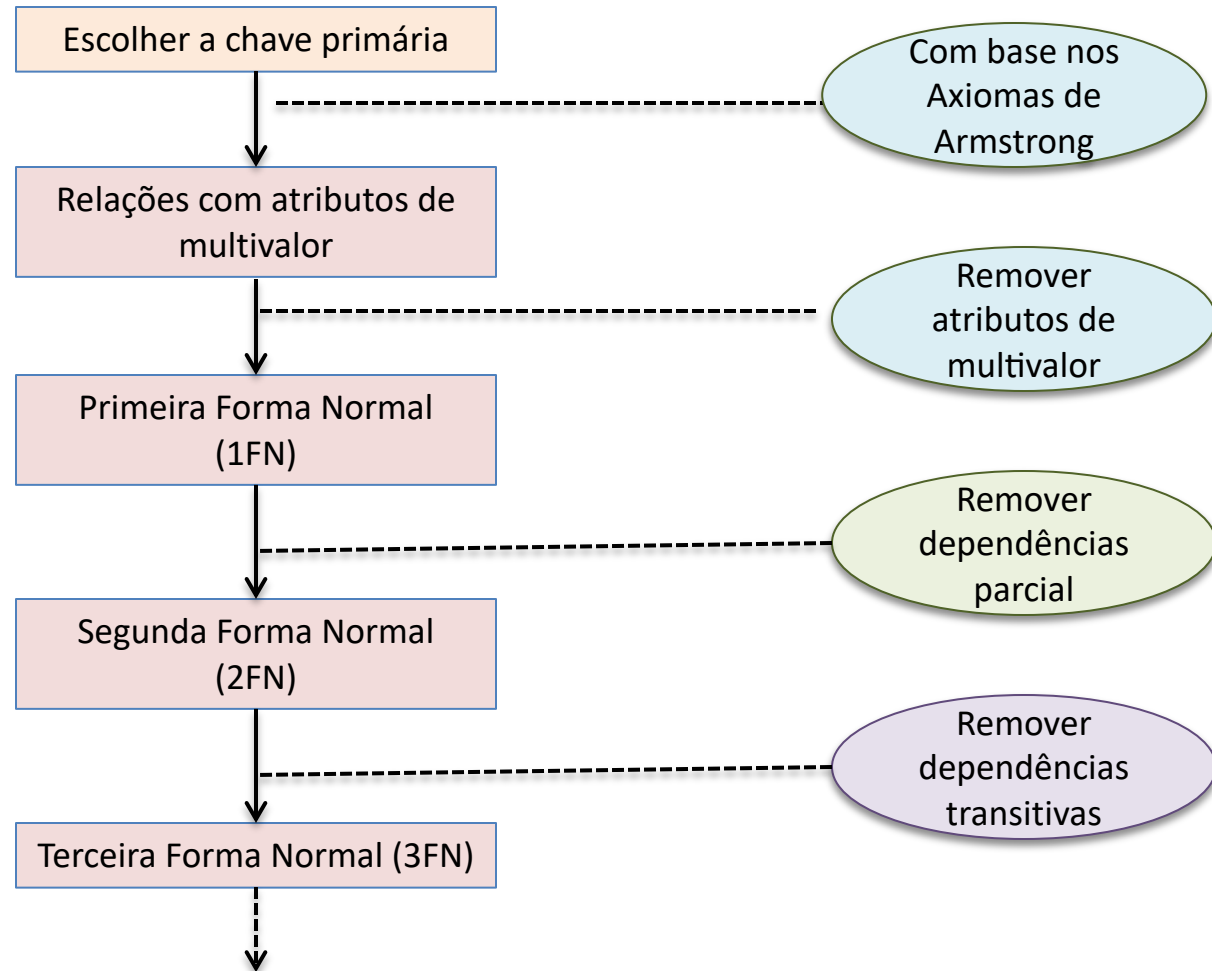
Normalização

Formas Normais:

| | |
|-------------------|-----------------------------------|
| 1FN 2FN 3FN | |
| FNBC (Boyce-Codd) | => <i>Dependências Funcionais</i> |
| 4FN | => <i>Dependência Multivalor</i> |
| 5FN | => <i>Dependência de Junção</i> |

Normalização

Processo



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

Normalização - Escolha da Chave Primária

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 $AB \rightarrow CE \Rightarrow AB \rightarrow C$ e $AB \rightarrow E$
2. Transitividade $AB \rightarrow C$ e $C \rightarrow D \Rightarrow AB \rightarrow CD$
3. Transitividade $E \rightarrow AB$ e $AB \rightarrow CD \Rightarrow E \rightarrow ABCD$

Normalização - Primeira Forma Normal

- Uma relação está na 1FN se:
 - Os atributos chave estão definidos
 - Não existem grupos repetitivos
 - 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

Normalização - Primeira Forma Normal

Suponhamos

Aluno (idAluno, nome, morada, (idDisciplina, nomeDisciplina))

Esta estrutura não se encontra na 1FN, uma vez que as colunas **idDisciplina** e **nomeDisciplina** admitem um conjunto de valores

| <u>idAluno</u> | nome | morada | idDisciplina | nomeDisciplina |
|----------------|--------|--------|--------------|-------------------------------|
| A1 | João | Rua A | D1, D2, D3 | Matemática, Economia, Direito |
| A2 | Ana | Rua B | D1, D4 | Matemática, Física |
| A3 | Pedro | Rua C | D1, D2 | Matemática, Economia |
| A4 | Filipa | Rua D | D1 | Matemática |

Normalização - Primeira Forma Normal

| <u>Id_aluno</u> | nome | morada |
|-----------------|--------|--------|
| A1 | João | Rua A |
| A2 | Ana | Rua B |
| A3 | Pedro | Rua C |
| A4 | Filipa | Rua D |

Aluno (idAluno, nome, morada)

| <u>Id_aluno</u> | <u>idDisciplina</u> | nomeDisciplina |
|-----------------|---------------------|----------------|
| A1 | D1 | Matematica |
| A1 | D2 | Economia |
| A1 | D3 | Direito |
| A2 | D1 | Matematica |
| A2 | D4 | Fisica |
| A3 | D1 | Matematica |
| A3 | D2 | Economia |
| A4 | D1 | Matematica |

AlunoInscrito (idAluno, idDisciplina, nomeDisciplina)

Normalização - Segunda Forma Normal

- 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

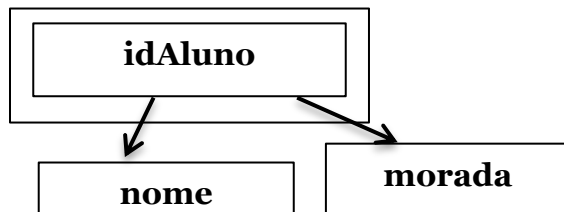
Normalização - Segunda Forma Normal

Conversão da estrutura para a 2 FN

1. 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
2. **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

Exemplo 1

- A tabela **Aluno** já está na 1ª FN e como a chave primária contém apenas um atributo ela também está na 2ª FN




| <u>Id_aluno</u> | nome | morada |
|-----------------|--------|--------|
| A1 | João | Rua A |
| A2 | Ana | Rua B |
| A3 | Pedro | Rua C |
| A4 | Filipa | Rua D |

Normalização - Segunda Forma Normal

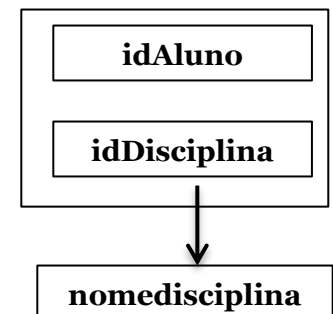
Conversão da estrutura para a 2 FN

- A tabela **AlunoInscrito** encontra-se na 1ª FN mas a sua chave primária é composta
- Necessário decompor a tabela **AlunoInscrito** pois existe uma dependência funcional entre o atributo não-chave **nomeDisciplina** e apenas parte da chave primária, com o atributo **idDisciplina**



| <u>Id_aluno</u> | <u>idDisciplina</u> | nomeDisciplina |
|-----------------|---------------------|----------------|
| A1 | D1 | Matemática |
| A1 | D2 | Economia |
| | | |

idDisciplina → nomeDisciplina



Normalização - Segunda Forma Normal

Conversão da estrutura para a 2 FN

| <u>Id_aluno</u> | nome | morada |
|-----------------|--------|--------|
| A1 | João | Rua A |
| A2 | Ana | Rua B |
| A3 | Pedro | Rua C |
| A4 | Filipa | Rua D |

Aluno(idAluno, nome, morada)

| <u>idDisciplina</u> | nomeDisciplina |
|---------------------|----------------|
| D1 | Matemática |
| D2 | Economia |
| D3 | Direito |
| D4 | Física |

Disciplina(idDisciplina, nomeDisciplina)

| <u>Id_aluno</u> | <u>idDisciplina</u> |
|-----------------|---------------------|
| A1 | D1 |
| A1 | D2 |
| A1 | D3 |
| A2 | D1 |
| A2 | D4 |
| A3 | D1 |
| A3 | D2 |
| A4 | D1 |

AlunoInscrito(idAluno(FK), idDisciplina(FK))

Normalização - Terceira Forma Normal

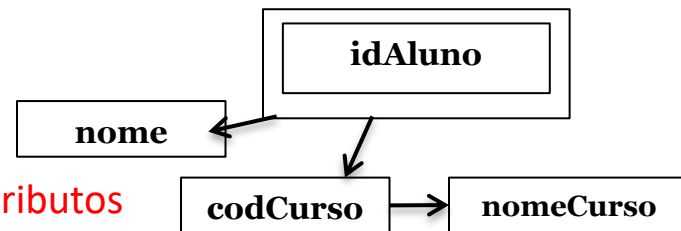
➤ Uma relação está na 3FN se:

➤ 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



Esta tabela não se encontra na 3FN porque o atributo não-chave **nomeCurso** depende funcionalmente do atributo **codCurso**

| <u>Id_aluno</u> | nome | codCurso | nomeCurso |
|-----------------|--------|----------|-------------|
| A1 | João | 01 | Informática |
| A2 | Ana | 02 | Civil |
| A3 | Pedro | 01 | Informática |
| A4 | Filipa | 03 | Química |

Aluno(idAluno, nome, codCurso, nomecurso)

Normalização - Terceira Forma Normal

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

Normalização - Terceira Forma Normal

Conversão da estrutura para a 3FN

- A tabela está na 2FN mas não está na 3FN
- Necessário decompor a tabela **Aluno** pois existe uma dependência funcional (transitiva) entre o atributo não-chave **codCurso** e o atributo **nomeCurso**

| <u>Id_aluno</u> | nome | codCurso |
|-----------------|--------|----------|
| A1 | João | 01 |
| A2 | Ana | 02 |
| A3 | Pedro | 01 |
| A4 | Filipa | 03 |

Aluno(idAluno, nome, codCurso(FK))

| <u>codCurso</u> | nomeCurso |
|-----------------|-------------|
| 01 | Informática |
| 02 | Civil |
| 01 | Informática |
| 03 | Química |

Curso(codCurso, nomecurso)

Normalização - 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 aplicativo !**

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

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

Normalização - Boyce Cood

- Exemplo:
 - Suponha a seguinte relação, que serve para registrar os alunos nas aulas laboratoriais
 - sabe-se que cada disciplina pode ser lecionada por vários docentes. No entanto cada docente só pode lecionar uma disciplina.

Laboratorios (id_aluno, coddisciplina, coddocente)

| <u>Id_aluno</u> | <u>coddisciplina</u> | coddocente |
|-----------------|----------------------|------------|
| 1180720 | Bddad | 15200 |
| 1180721 | Bddad | 15200 |
| 1180720 | Esinf | 15230 |
| 1180728 | Bddad | 15240 |

Normalização - Boyce Cood

- A relação satisfaz as três primeiras formas normais, no entanto não está na FNBC
- o atributo coddocente não é chave candidata, no entanto é um determinante
 - Uma vez que cada docente só pode lecionar uma disciplina, temos coddocente -> coddisciplina
- Para estar na FNBC exigirá a sua decomposição em duas relações

| <u>Id_aluno</u> | <u>coddocente</u> |
|-----------------|-------------------|
| 1180720 | 15200 |
| 1180721 | 15200 |
| 1180720 | 15230 |
| 1180728 | 15240 |

Docentes de cada aluno

| <u>coddocente</u> | <u>coddisciplina</u> |
|-------------------|----------------------|
| 15200 | Bddad |
| 15230 | Esinf |
| 15240 | Bddad |

Disciplina lecionada por cada docente

Normalização - 4ª e 5ª FN

- 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

Normalização - Conclusão

- 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

Normalização - Exercício 2

Normalize a estrutura apresentada

Considere a estrutura de dados seguinte, referente ao planeamento de produção de uma fábrica de artigos de plástico. A fábrica está estruturada em secções e cada secção é composta por diferentes centros de trabalho. Uma ordem de produção pode ser realizada em diversos centros de trabalho e utiliza diversas matérias-primas.



Planeamento de produção = {ordem_prod, produto, nome_produto, qtd_a_produzir, data_prev_inicio, data_prev_fim, data_real_inicio, data_real_fim, {secção, nome_secção, {centro_trabalho, desc_centro_trabalho}, localização}, {mat_prima, descrição_mp, qtd_mp}, percent_execução}

Normalização - Exercício 3

Desenhe o diagrama de dependências funcionais e normalize a estrutura apresentada



| Nr Factura | Data | codCliente | NomeCliente | CodProd | Descricao Produto | Valor | Quantidade | Desconto |
|------------|------------|------------|-------------|---------|-------------------|-------|------------|----------|
| 000257 | 01-07-2016 | 1234567 | João Gomes | 12 | Lápis Bic | 100 | 250 | 5% |
| 000257 | 01-07-2016 | 1234567 | João Gomes | 13 | Bloco de notas | 1000 | 200 | 5% |
| 000257 | 01-07-2016 | 1234567 | João Gomes | 15 | Caneta | 70 | 50 | 0% |
| 000258 | 01-07-2016 | 1234568 | Ana Marques | 12 | Lápis Bic | 100 | 400 | 6% |
| 000258 | 01-07-2016 | 1234568 | Ana Marques | 16 | Caderno | 500 | 350 | 6% |
| 000258 | 01-07-2016 | 1234568 | Ana Marques | 17 | Régua | 100 | 20 | 0% |