

Projeto Bottom-Up de banco de Dados

Prof. Alexandre Tagliari Lazzaretti
lazzaretti10@gmail.com

Introdução

- Projeto de banco de dados
 - Parte integrante do desenvolvimento de um sistema de informação;
 - Preocupa-se com a representação adequada dos dados operacionais;
 - Atividades:
 - Definição de esquemas de dados em diferentes níveis de abstração;
 - Nível conceitual, lógico e físico;
 - Tipos de projeto (**modelo relacional**):
 - Top-down
 - Bottom-Up

■ Top-Down – Objetivos

- Projeto Conceitual
 - preocupação: correta abstração do mundo real
(captura correta da semântica da aplicação)
- Projeto Lógico + Físico
 - preocupação: escolhas corretas no mapeamento para o modelo do SGBD (relacional), levando em conta uma boa performance de acesso
(distribuição adequada dos dados em tabelas)

■ Bottom-Up - Objetivos

- Ênfase nas descrições de dados já existentes na organização
 - arquivos eletrônicos, fichários, documentos formatados (pedido, NF), relatórios, ...
- Processo também chamado de engenharia reversa de BD
 - aplicado em casos onde existem fontes de dados ou sistemas informatizados (legados) sem BD
- Cinco etapas
 1. coleta de fontes de dados
 2. representação em uma tabela não-normalizada
 3. normalização
 4. integração de esquemas relacionais das fontes
 5. engenharia reversa do esquema relacional

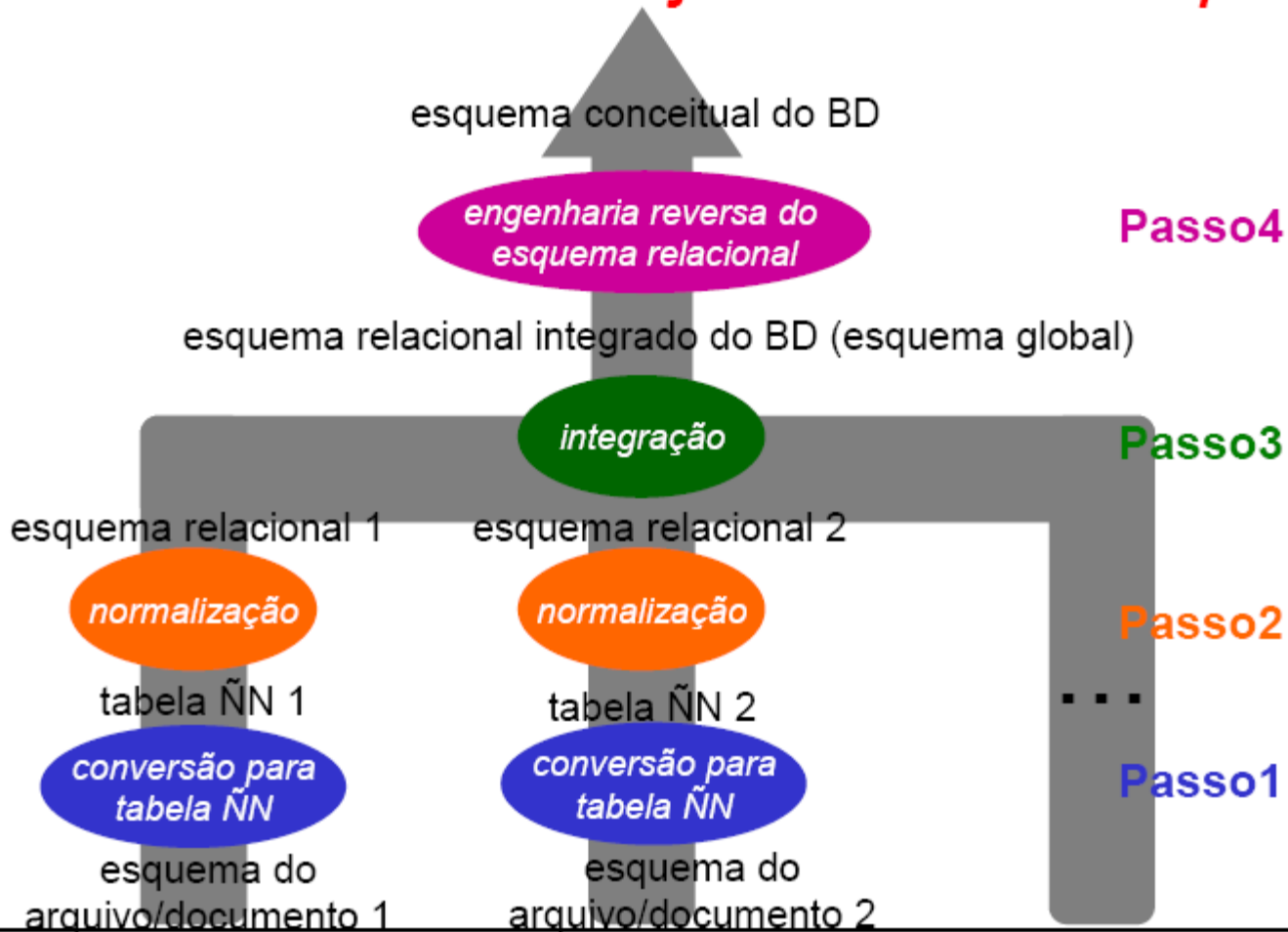
Projeto *Bottom-Up* de BD

- Uma organização que deseja definir um BD pode já conter descrições de dados
 - sistemas legados
 - gerenciamento de arquivos de dados
 - ambientes não-informatizados
 - dados manuscritos
 - fichários, relatórios, documentos, ...
 - ...

Projeto *Bottom-Up* de BD

- Objetivo
 - construção dos esquemas conceitual e lógico do BD a partir de fontes de dados da organização
- Projeto *Bottom-Up* é também chamado de processo de Engenharia Reversa de BD
- Foco
 - engenharia reversa de BD relacional

Processo de Projeto *Bottom-Up*



Exemplo de Fonte de Dados: Relatório

RELATÓRIO DE ALOCAÇÃO A PROJETO

CÓDIGO DO PROJETO: LSC001

TIPO: Novo Desenv.

DESCRIÇÃO: Sistema de Estoque

CÓDIGO DO EMPREGADO	NOME	CATEGORIA FUNCIONAL	SALÁRIO	DATA DE INÍCIO NO PROJETO	TEMPO ALOCADO AO PROJETO
---------------------	------	---------------------	---------	---------------------------	--------------------------

2146	João	A1	4	1/11/91	24
3145	Silvio	A2	4	2/10/91	24
6126	José	B1	9	3/10/92	18
1214	Carlos	A2	4	4/10/92	18
8191	Mário	A1	4	1/11/92	12

CÓDIGO DO PROJETO: PAG02

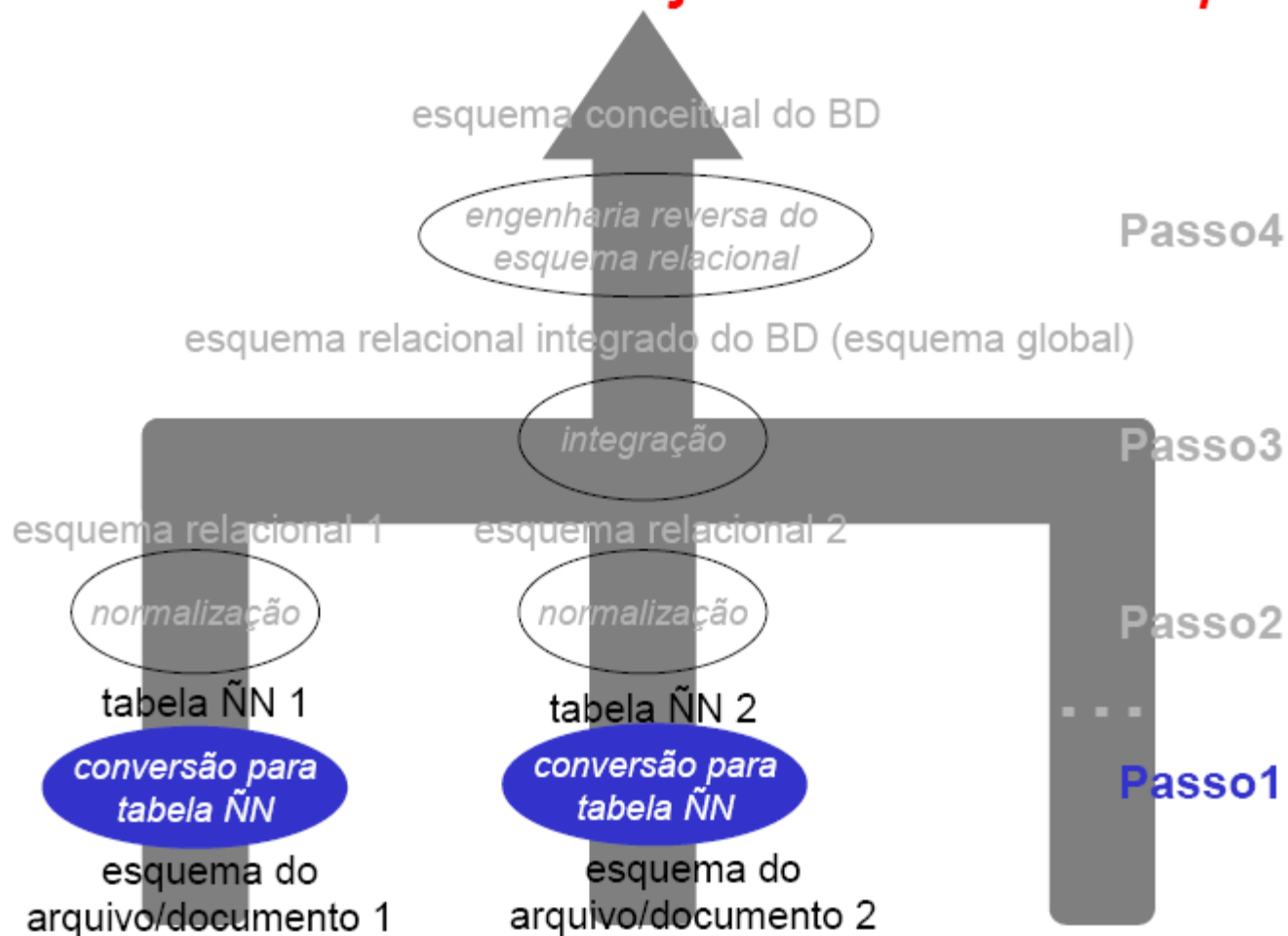
TIPO: Manutenção

DESCRIÇÃO: Sistema de RH

CÓDIGO DO EMPREGADO	NOME	CATEGORIA FUNCIONAL	SALÁRIO	DATA DE INÍCIO NO PROJETO	TEMPO ALOCADO AO PROJETO
---------------------	------	---------------------	---------	---------------------------	--------------------------

8191	Mário	A1	4	1/05/93	12
4112	João	A2	4	4/01/91	24
6126	José	B1	9	1/11/92	12

Processo de Projeto *Bottom-Up*



Passo1 – Conversão para Tabela ÑÑ

- Objetivo
 - obtenção de uma descrição da fonte de dados independente do seu formato
 - relatório, arquivo, ...
 - padronização da representação das fontes para facilitar o processo de projeto
- Forma de representação: Tabela ÑÑ
 - pode ter uma ou mais tabelas aninhadas
 - tabela aninhada
 - possui atributos multivalorados
 - atributo que ao invés de conter valores atômicos, contém múltiplos valores ou uma tabela que pode, por sua vez, ser aninhada

Exemplo de Tabela ÑN

CódProj	Tipo	Descr	Emp					
			CodEmp	Nome	Cat	Sal	DataIni	TempAl
LSC001	Novo Desenv.	Sistema de Estoque	2146	João	A1	4	1/11/91	24
			3145	Silvio	A2	4	2/10/91	24
			6126	José	B1	9	3/10/92	18
			1214	Carlos	A2	4	4/10/92	18
			8191	Mário	A1	4	1/11/92	12
PAG02	Manutenção	Sistema RH	8191	Mário	A1	4	1/05/93	12
			4112	João	A2	4	4/01/91	24
			6126	José	B1	9	1/11/92	12

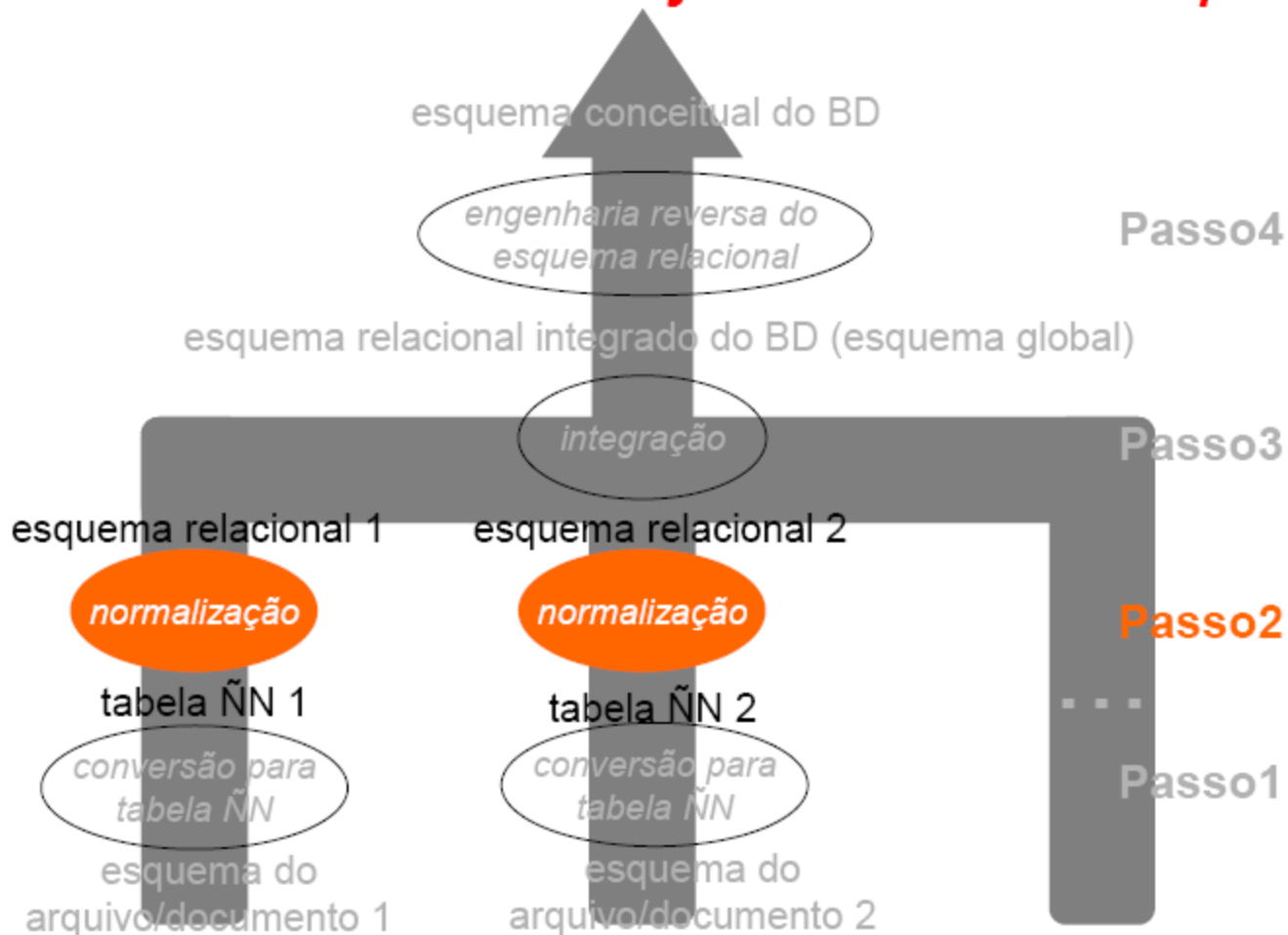
Tabela aninhada em uma linha de projeto

Representação na Forma de Tabela ÑÑ

Projetos (codProj, tipo, descr,
(codEmp, nome, categ, sal, dataIni, tempoAloc))

indicam-se as chaves primárias (CPs) de cada tabela

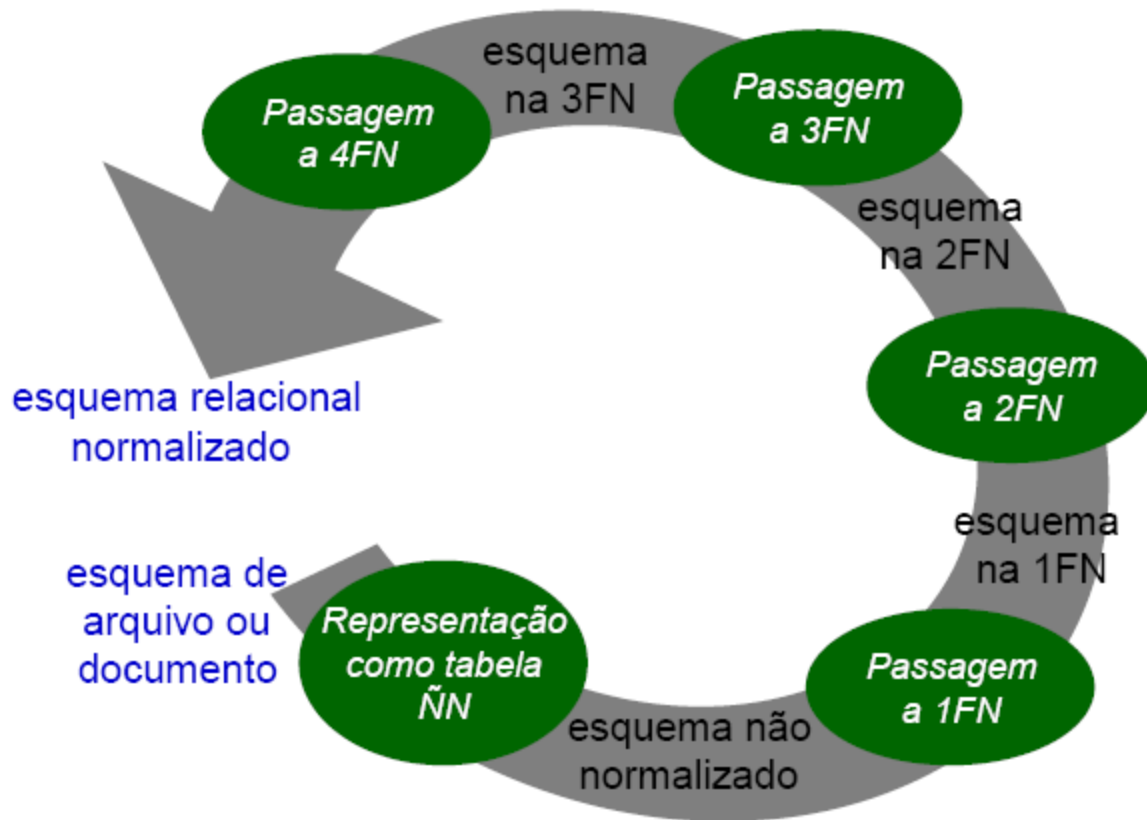
Processo de Projeto *Bottom-Up*



Processo de Normalização

- Objetivos
 - eliminar redundâncias em uma tabela
 - geração de novas tabelas não-redundantes
 - agrupar dados relevantes em uma mesma tabela de modo a constituir uma futura entidade ou relacionamento do esquema conceitual
 - análise de *dependências funcionais*

Processo de Normalização



1ª Forma Normal (1FN)

- *“Uma tabela está na 1FN se ela não possui tabelas aninhadas”*
- **Duas alternativas**
 1. manter todos os dados em uma única tabela
 2. gerar uma tabela para cada aninhamento

1FN – Alternativa 1

ÑN: Projetos (codProj, tipo, descr,
(codEmp, nome, categ, sal, dataIni, tempoAloc))



1FN: Projetos (codProj, tipo, descr, codEmp, nome, categ,
sal, dataIni, tempoAloc)

Desvantagem: dados de diferentes entidades estão misturados!

- mais difícil identificar posteriormente os relacionamentos entre estas entidades

1FN – Alternativa 2

ÑN: Projetos (codProj, tipo, descr,
(codEmp, nome, categ, sal, dataIni, tempoAloc))



1FN: Projetos (codProj, tipo, descr)
Alocações (codProj, codEmp, nome, categ, sal,
dataIni, tempoAloc)

- CP da tabela externa migra para a tabela aninhada
- Qual a CP da tabela aninhada?

Exemplos

Boleto de Entrega

ÑN: Pedidos (codCli, nomeCli, nroPed, nroRampa)



1FN: Pedidos (codCli, nomeCli, nroPed, nroRampa)

Relatório de Pedidos

ÑN: Pedidos (nroPed, dataPed, codCli, nomeCli,
nroTel), (codPeça, descrPeça, qtdePedida))



1FN: Pedidos (nroPed, dataPed, codCli, nomeCli)
Telefones (nroPed, nroTel)
Itens (nroPed, codPeça, nomePeça, qtde))

Dependência Funcional (DF)

- Conceito necessário para o entendimento da segunda e terceira formas normais
- Definição
 - um atributo A_2 *depende funcionalmente* de um atributo A_1 (ou um atributo A_1 *determina* um atributo A_2) quando, em todas linhas da tabela, para cada valor de A_1 que aparece na tabela, aparece o mesmo valor de A_2

Dependência Funcional - Exemplo

...	Código	...	Salário
	E1		500
	E3		450
	E2		500
	E1		500
	E3		450
	E2		500

Código ? Salário

DF Total e DF Parcial

- DF Total

- se um atributo A_x depende funcionalmente de **todos** os atributos que compõem a CP de uma tabela T, diz-se que A_x possui **DF total** da CP de T

- DF Parcial

- se um atributo A_x depende funcionalmente apenas de **alguns** atributos (não todos!) que compõem a CP de uma tabela T, diz-se que A_x possui **DF parcial** da CP de T

2ª Forma Normal (2FN)

- *“Uma tabela está na 2FN se ela estiver na 1FN e não possuir DFs parciais”*
 - tabelas com DFs parciais devem ser desmembradas em tabelas que possuam DFs totais
- Tabelas cuja CP possui apenas um atributo estão automaticamente na 2FN

2FN - Exemplo

1FN: Projetos (CodProj, Tipo, Descr)

Alocações (CodProj, CodEmp, Nome, Cat, Sal,
DataIni, TempoAloc)

DFs: CodEmp ? Nome, Cat, Sal (DF Parcial!)
(CodProj, CodEmp) ? DataIni, TempoAloc



2FN: Projetos (CodProj, Tipo, Descr)

Alocações (CodProj, CodEmp, DataIni, TempoAloc)

Empregados (CodEmp, Nome, Cat, Sal)

DF Transitiva ou Indireta

- Se um atributo **não-chave** A_x possui **DF total** da CP de uma tabela T e também possui DF total de um ou mais atributos **não-chave** de T, então diz-se que A_x possui **DF transitiva ou indireta** da CP de T

3ª Forma Normal (3FN)

- *“Uma tabela está na 3FN se ela estiver na 2FN e não possuir DFs transitivas”*
 - tabelas com DFs transitivas devem ser desmembradas em tabelas que não possuam tais DFs
- Tabelas que possuem zero ou apenas um atributo que não faz parte da CP estão automaticamente na 3FN

3ª Forma Normal (3FN)

2FN: Projetos (CodProj, Tipo, Descr)
Alocações (CodProj, CodEmp, DataIni, TempoAloc)
Empregados (CodEmp, Nome, Cat, Sal)

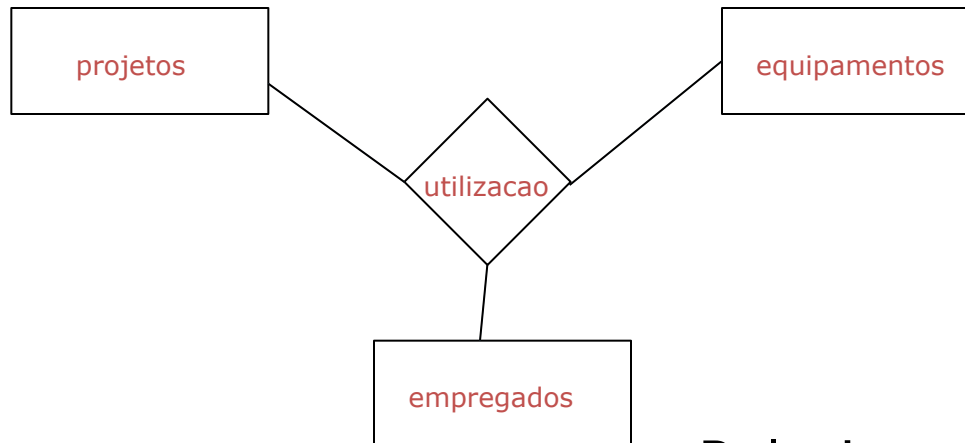
DFs: CodEmp ? Sal
CodEmp ? Cat ? Sal (DF Transitiva!)



3FN: Projetos (CodProj, Tipo, Descr)
Alocações (CodProj, CodEmp, DataIni, TempoAloc)
Empregados (CodEmp, Nome, Cat)
CategoriasFuncionais(Cat, Sal)

DF Multivalorada

“Uma coluna ou um conjunto de colunas depende multivaloradamente de uma coluna (determinante) da mesma tabela quando o valor do atributo determinante identifica repetidas vezes um conjunto de valores na coluna dependente.” (Heuser, 2009)



Relacionamento N:N:N

DF Multivalorada

utilizacoes

#@Codproj	#@Codemp	#@Codequip
1	1	1
1	2	1
1	3	1
1	1	2
1	2	2
1	3	2

- Codemp (possui um conjunto de valores 1,2,3) depende multivaloradamente de Codproj (coluna determinante valor 1)

4ª Forma Normal (4FN)

- *“Uma tabela está na 4FN sse ela estiver na 3FN e não possuir DFs multivaloradas”*
 - tabelas com DFs multivaloradas devem ser desmembradas em tabelas que não possuam tais DFs
- Tabelas que possuem CP composta por um ou dois atributos, ou que possuem atributos não-chave estão automaticamente na 4FN

4^a Forma Normal (4FN)

CodProj	CodEmp	CodEquip
1	1	1
1	2	1
1	3	1
1	1	2
1	2	2
1	3	2

CodProj ? ? CodEmp

CodProj ? ? CodEquip

3FN: ProjEmpEquip (CodProj, CodEmp, CodEquip)



4FN: ProjEmp (CodProj, CodEmp)
ProjEquip (CodProj, CodEquip)

Normalização – Dicas

- Inexistência de CPs
 - tabelas podem não ter atributos que realmente garantam identificação única de suas tuplas
 - dica: definir uma CP

ÑN: Projetos (CodProj, Tipo, Descr,
(Nome, Cat, Sal, DataIni, TempoAloc))



ÑN: Projetos (CodProj, Tipo, Descr,
(CodEmp, Nome, Cat, Sal, DataIni, TempoAloc))

Normalização – Dicas

- Dados irrelevantes, redundantes ou derivados
 - tabelas podem ter atributos que não necessariamente precisam ser mantidos no BD
 - dica: eliminar estes atributos

ÑN: Projetos (CodProj, Tipo, Descr, NroEmps, DataRel,
(CodEmp, Nome, Cat, Sal, DataIni, TempoAloc))



ÑN: Projetos (CodProj, Tipo, Descr,
(CodEmp, Nome, Cat, Sal, DataIni, TempoAloc))

Normalização – Dicas

- Dados relevantes, porém implícitos
 - dica: definir tais dados

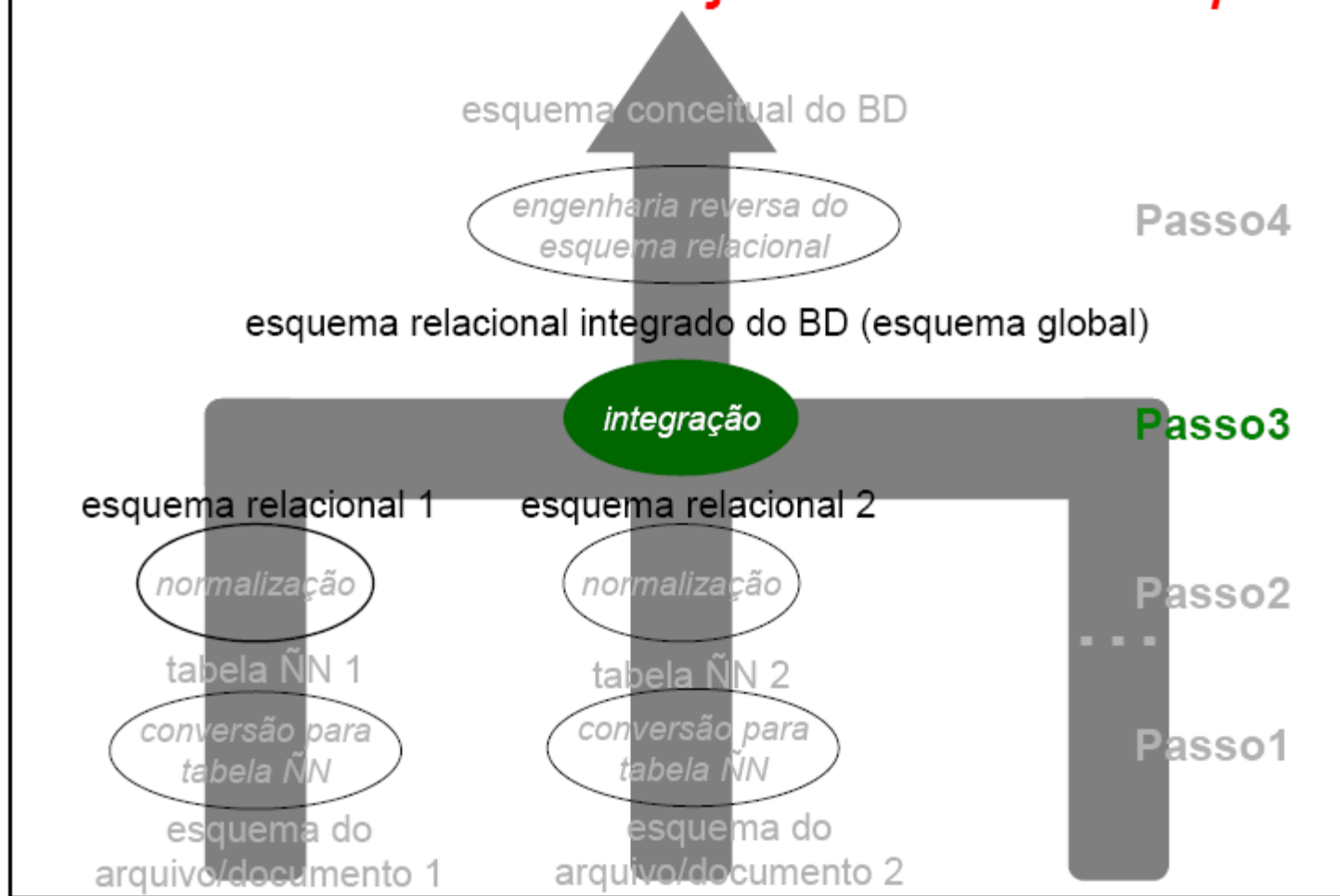
ÑN: Aprovação (CodCurso, Nome, CodCand, Nome, Endereço)

a ordem determina a classificação do candidato



ÑN: Aprovação (CodCurso, Nome, CodCand, Nome, Endereço, **OrdemClass**)

Processo de Projeto *Bottom-Up*



Integração de Esquemas Relacionais

- Integração de tabelas produzidas para fontes de dados diferentes
- Objetivos
 - integrar atributos de uma mesma entidade ou relacionamento que aparecem em tabelas diferentes
 - eliminar atributos redundantes presentes em tabelas diferentes ou mesmo tabelas redundantes

Integração de Esquemas Relacionais

- Etapas
 1. integração de tabelas com mesma CP
 2. integração de tabelas com chave contida
 3. verificação da 3FN
- Etapas 1 e 2 se aplicam a tabelas semanticamente equivalentes
 - entidades ou relacionamentos sinônimos

Integração de Tabelas com Mesma CP

- Tabelas que possuem CPs com os mesmos domínios e conteúdos

Tabela1: Projeto (CodProj, Tipo, Descrição)

Tabela2: Proj (Código, Descr, DataIni, CodDepto)



EsquemaGlobal:

Projeto (CódProj, Tipo, Descr, DataIni, CodDepto)

Integração de Tabelas com CP contida

- Todos os atributos de uma tabela x são CP e esta CP está contida na CP de uma tabela y
 - tabela x pode ser removida

Tabela1: **DiscCursadas** (CodA, CodD)

Tabela2: **Histórico** (Al, Disc, Sem, Nota)



EsquemaGlobal: **Histórico**(Al, Disc, Sem, Nota)

Verificação da 3FN

- A integração de dados em uma tabela pode provocar dependências entre atributos não-chave

Tabela1: Depto (CodD, Nome, CodGerente)

Tabela2: Departamento (Código, Local, NomeGer)



EsquemaGlobal:

Departamento(Código, Nome, Local, CodGerente,
NomeGerente)

volta à 2FN!

Verificação da 3FN

EsquemaGlobal:

Departamento(Código, Nome, Local, CodGerente,
NomeGerente)



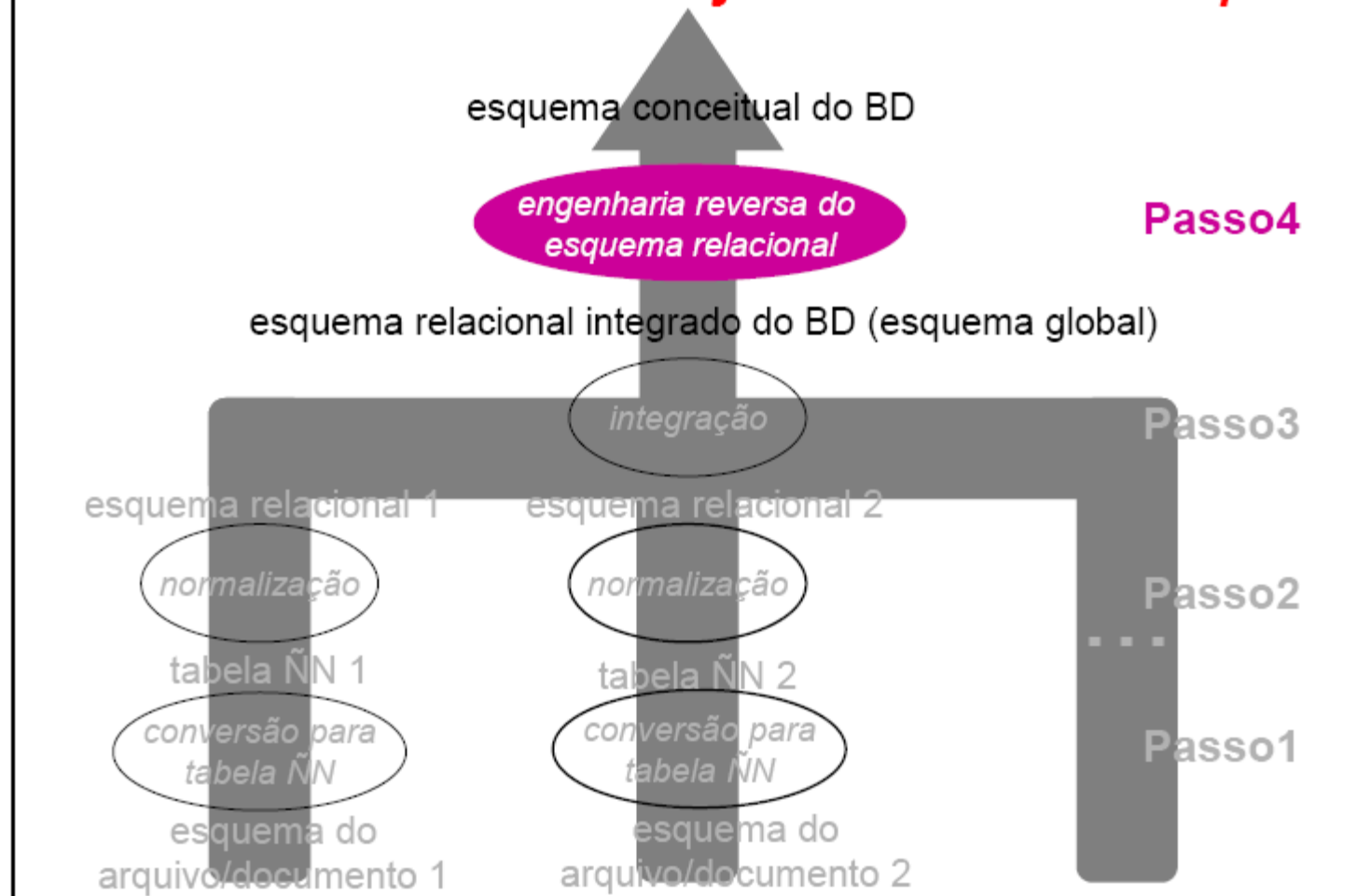
aplica-se novamente a 3FN!

EsquemaGlobal:

Departamento(Código, Nome, Local, CodGerente)
Gerente(Código, Nome)

Observação: a geração de uma nova tabela no esquema global exige que se verifiquem os demais casos de integração para ela!

Processo de Projeto *Bottom-Up*



Engenharia Reversa do Esquema Relacional Global

- Aplicação das regras de conversão de um esquema relacional em um esquema ER



Validação do Esquema Conceitual

- Relacionamentos redundantes
 - Exemplo: Sessão-Artigo; Sessão-Conferência;
Artigo-Conferência
- Relacionamentos em excesso
 - Exemplo: Artigo-(Autoria)-Pessoa
Artigo-(AutorPrincipal)-Pessoa
 - o fato do autor ser principal pode ser indicado no relacionamento *Autoria*
- Relacionamentos não detectados
 - Exemplo: Artigo(..., AssuntoPrincipal)
Tema(código, nome)
 - AssuntoPrincipal* pode se referir a um *Tema* e isto não ter sido percebido
- Relacionamentos errados
 - Exemplo: Telefones (nroPed, nroTel)
Clientes (codCli, nome)
 - telefones estão relacionados a clientes e não a pedidos

Projeto BU X Projeto TD

- Projeto BU **ajuda a validar** o projeto TD
 - complementar e/ou corrigir os esquemas conceitual e lógico obtidos através do projeto TD com base nos esquemas conceitual e lógico obtidos através do projeto BU