

Banco de Dados

Normalização

- Definição de Normalização.
- Primeira Forma Normal (1FN).
- Segunda Forma Normal (2FN).
- Terceira Forma Normal (3FN).

A terminologia utilizada na normalização, envolve basicamente a teoria de conjuntos (que foi a base para concepção da teoria de normalização), o conceito de banco de dados relacional e a aplicação convencional em ambiente de processamento de dados.

Conjunto	Banco de Dados	Processamento de Dados
Relação	Tabela	Arquivo
Domínio	Possíveis valores para coluna	Possíveis valores para o campo
Tupla	Linha	Registro

- Processo para avaliar e corrigir estruturas e tabelas de modo a minimizar as redundâncias de dados, reduzindo, assim, a probabilidade de anomalias (em inserção, exclusão ou atualização de dados).
- Atua por meio de uma série de estágios chamados formas normais onde os mais comuns são:
 - Primeira forma normal (1NF);
 - Segunda forma normal (2NF);
 - Terceira forma normal (3NF).
- Consiste da decomposição de esquemas para evitar anomalias nos dados.
- “Bom design” evita redundância de dados e anomalias de inserção, exclusão e atualização.

- 3NF é melhor que 2NF, que, por sua vez, é melhor que 1NF.
- Para a maioria das finalidades dos projetos de bancos de dados comerciais, 3NF é o mais alto nível que se precisa chegar no processo de normalização.
- Não se deve assumir que o nível mais alto de normalização seja sempre o mais desejável.
- Mecanismo formal para analisar esquemas de relações baseado nas suas chaves e nas dependências funcionais entre seus atributos.
- Projeto conceitual bem estrutura resulta naturalmente em esquemas normalizados.

- Desnormalização – produz uma forma normal mais baixa.
- O preço a pagar pela melhora de desempenho decorrente da desnormalização é a maior redundância de dados.

No processo de normalização deve-se chegar nas características descritas a seguir:

- Cada tabela representa um único assunto.
- Nenhum item de dados será armazenado desnecessariamente em mais de uma tabela.
- Todos os atributos não primários (atributos que não fazem parte da chave primária) de uma tabela são dependentes da chave primária.
- Todas as tabelas estão livres de anomalias de inserção, atualização e exclusão.

- O objetivo da normalização é garantir que todas as tabelas estejam, pelo menos, na terceira forma normal (3NF).
- Formas normais de nível superior provavelmente não serão encontradas em um ambiente comercial.
- O processo de normalização trabalha em uma relação por vez.
- Ocorre progressiva separação da relação (tabela) em um conjunto de novas relações (tabelas) baseadas nas dependências identificadas

Primeira Forma Normal (1FN)

- Um entidade está na primeira forma normal quando nenhum de seus atributos (na estrutura) possuir repetições. Ou seja, se e somente se todos os seus atributos contiverem apenas valores atômicos (simples, indivisíveis).
- Isso quer dizer que os atributos são únicos (indivisíveis) ou que possuem apenas um valor por célula. Assim, é preciso verificar cada um dos atributos para identificar se este possui um único valor para cada ocorrência (registro) da entidade.
- O valor (conteúdo) do atributo (campo) pode, e normalmente é, ser diferente em cada ocorrência (registro) da entidade. Ou seja, é preciso evitar a ocorrência de mais de um atributo com a mesma característica em um mesmo registro.
- Faz parte da definição formal de uma relação (tabela).
- Estabelecida para não permitir atributos multivalorados, atributos compostos e suas combinações.

Conversão para a Primeira Forma Normal (1FN)

<u>CPF</u>	Nome	Sexo	Localizacao	Telefone
1234	Markus	M	Curitiba, PR, Brasil	555-222, 555-333
4321	João	M	Arapoti, PR, Brasil	111-222, 111-333
5678	Lidia	F	Florianópolis, SC, Brasil	444-555, 444-777
8765	Lucia	F	Uberlândia, MG, Brasil	222-333

Como pode-se perceber existem dados compostos (no campo Localizacao, por exemplo: Curitiba, PR, Brasil) e dados multivalorados (no campo Telefone, por exemplo: 555-222, 555-333). Neste caso, primeiro é preciso separar os dados compostos do campo (Localizacao) em 3 campos (Cidade, UF, Pais); segundo, por conta dos dados multivalorados do campo (Telefone), criar uma nova entidade (NumTelefone) e levar a chave primária (CPF) da entidade original para a nova entidade (NumTelefone) gerada. E para essa nova entidade criada, definir uma chave composta (concatenada), a saber: CPF e Telefone.

Conversão para a Primeira Forma Normal (1FN)

Pessoa

<u>CPF</u>	Nome	Sexo	Localizacao	Telefone
1234	Markus	M	Curitiba, PR, Brasil	555-222, 555-333
4321	João	M	Arapoti, PR, Brasil	111-222, 111-333
5678	Lidia	F	Florianópolis, SC, Brasil	444-555, 444-777
8765	Lucia	F	Uberlândia, MG, Brasil	222-333

Pessoa

<u>CPF</u>	Nome	Sexo	Cidade	UF	Pais
1234	Markus	M	Curitiba	PR	Brasil
4321	João	M	Arapoti	PR	Brasil
5678	Lidia	F	Florianópolis	SC	Brasil
8765	Lucia	F	Uberlândia	MG	Brasil

NumTelefone

<u>CPF</u>	<u>Telefone</u>
1234	555-222
1234	555-333
4321	111-222
4321	111-333
5678	444-555
5678	444-777
8765	222-333

Conversão para a Primeira Forma Normal (1FN)

<u>CodCD</u>	NomeCD	Grav	Preco	TotalTemp	Faixa	Musica	Autor	Tempo	
001	CD_A	RCA	20,00		01	MusicaA	X	3:00	
					02	MusicaB	X	3:10	
002	CD_B	EMI	18.00		01	MusicaC	Y	3:05	
					02	MusicaD	Y	3:00	

Como pode-se perceber existem diversos atributos que se repetem nesta tabela, a saber: Faixa, Autor e Tempo. Neste caso, é preciso separar a informação que se repete em um nova entidade e levar a chave primária da entidade original para a nova entidade gerada (senão, não será possível relacionar as informações das duas entidades).

Em seguida, é preciso criar uma nova chave para a nova entidade, para tanto, identifica-se um campo (atributo), que unido à chave da entidade original, formará uma chave composta (ou chave concatenada) na nova entidade. Caso não seja possível essa identificação, pode-se criar uma nova chave.



Conversão para a Primeira Forma Normal (1FN)

CD

<u>CodCD</u>	NomeCD	Grav	Preco	TotalTemp	Faixa	Musica	Autor	Tempo	
001	CD_A	RCA	20,00		01	MusicaA	X	3:00	
					02	MusicaB	X	3:10	
002	CD_B	EMI	18.00		01	MusicaC	Y	3:05	
					02	MusicaD	Y	3:00	

CD

<u>CodCD</u>	NomeCD	Grav	Preco	TotalTemp
001	CD_A	RCA	20,00	
002	CD_B	EMI	18.00	

Item_CD

<u>CodCD</u>	<u>Faixa</u>	Musica	Autor	Tempo
001	01	Musica1	X	3:00
001	02	Musica2	X	3:10
002	01	Musica10	Y	3:05
002	02	Musica20	Y	3:00

Em uma tabela relacional, diz-se que uma coluna C2 depende funcionalmente de uma coluna C1 (ou que uma coluna C1 determina a coluna C2) quando, em todas as linhas da tabela, para cada valor de C1 que aparece na tabela, aparece o mesmo valor de C2.

A Dependência Funcional deve ser explicitamente definida por alguém que conheça a semântica dos atributos de uma relação.

É utilizada para identificar as chaves primárias de uma tabela.

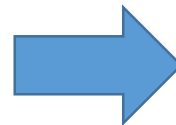
Pessoa

CPF	Nome
1234	Markus
4321	João
5678	Lidia
8765	Lucia

Ao analisar o caso abaixo, identifica-se que para este domínio não se pode duplicar o CPF porque ele deve ser único.

Pessoa

CPF	Nome
1234	Markus
4321	João
5678	Lidia
8765	Lucia



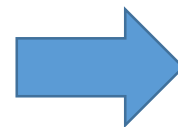
Pessoa

CPF	Nome
1234	Markus
4321	João
5678	Lidia
8765	Lucia
8765	Ana

Ao analisar o caso abaixo, identifica-se que para este domínio não se pode duplicar o CPF porque ele deve ser único. E nesta simulação temos a duplicidade do CPF: 8765.

Pessoa

CPF	Nome
1234	Markus
4321	João
5678	Lidia
8765	Lucia



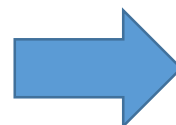
Pessoa

CPF	Nome
1234	Markus
4321	João
5678	Lidia
8765	Lucia
8765	Ana

Ao analisar o caso abaixo, identifica-se que para este domínio não se pode duplicar o CPF porque ele deve ser único. E nesta simulação temos a duplicidade do CPF: 8765, ou seja, isso está errado.

Pessoa

CPF	Nome
1234	Markus
4321	João
5678	Lidia
8765	Lucia



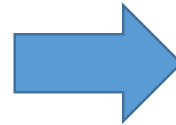
Pessoa

CPF	Nome
1234	Markus
4321	João
5678	Lidia
8765	Lucia
8765	Ana

Ao analisar o novo caso, identificamos que existem duas pessoas com o mesmo nome (homônimas), ou seja, são pessoas distintas com CPFs diferentes, porém com o mesmo nome, isso é permitido.

Pessoa

CPF	Nome
1234	Markus
4321	João
5678	Lidia
8765	Lucia



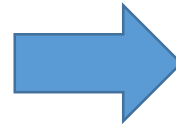
Pessoa

CPF	Nome
1234	Markus
4321	João
5678	Lidia
8765	Lucia
2244	Lidia

Ao analisar o novo caso, identificamos que existem duas pessoas com o mesmo nome (homônimas), ou seja, são pessoas distintas porém com o mesmo nome, isso é permitido. Então, podemos perceber que **CPF determina** nome (isso é dependência funcional), entretanto, **Nome** não determina **CPF** porque pode haver nomes iguais.

Pessoa

CPF	Nome
1234	Markus
4321	João
5678	Lidia
8765	Lucia



Pessoa

<u>CPF</u>	Nome
1234	Markus
4321	João
5678	Lidia
8765	Lucia
2244	Lidia

CPF → Nome

- Uma entidade (relação/tabela) está na 2FN quando todos os seus atributos não chave dependem unicamente da chave.
- Uma entidade (relação/tabela) encontra-se na 2FN se e somente se estiver na 1FN e não contiver dependências parciais, ou seja, para estar na 2FN não deve possuir atributo não-chave funcionalmente determinado por parte da chave primária.
 - Dependência Parcial: ocorre quando uma coluna depende apenas de uma parte da chave primária composta.
- Reforçando, uma entidade (relação/tabela) está em 2NF quando estiver na 1NF; não tiver dependências parciais; e nenhum atributo for dependente apenas de uma parte da chave primária (composta).

Segunda Forma Normal (2FN)

Este é um domínio definido para armazenar algumas informações de projeto e alocação de pessoas (empregados) em projetos.

EmpregadoProjeto

<u>CPF_Emp</u>	<u>Cod_Pj</u>	Hrs_Trab	Nome_Emp	Nome_Pj	Local_Pj
1234	AMS-200	200	Markus	SistSem	MG
4321	UCM-300	300	João	SistAmz	RO
5678	ATO-400	220	Lidia	SistSoj	MG
1234	UCM-300	280	Markus	SistAmz	PR

Segunda Forma Normal (2FN)

Podemos identificar os campos (CPF_Emp e Cod_Pj) como chave. Ainda nesta tabela, percebemos que o campo (Hrs_Trab) depende da chave (CPF_Emp e Cod_Pj) completa.

EmpregadoProjeto

<u>CPF_Emp</u>	<u>Cod_Pj</u>	Hrs_Trab	Nome_Emp	Nome_Pj	Local_Pj
1234	AMS-200	200	Markus	SistSem	MG
4321	UCM-300	300	João	SistAmz	RO
5678	ATO-400	220	Lidia	SistSoj	MG
1234	UCM-300	280	Markus	SistAmz	PR

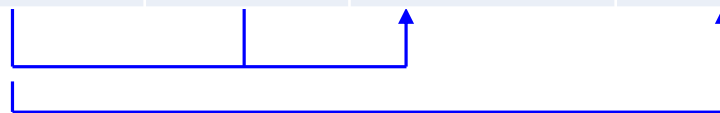


Segunda Forma Normal (2FN)

Aqui identificamos o campo (Nome_Emp) sendo dependente parcial da chave, ou seja, o campo (Nome_Emp) depende única e exclusivamente do campo (CPF_Emp). Neste caso, atributo não chave (Nome_Emp) sendo determinado por parte da chave primaria (composta).

EmpregadoProjeto

<u>CPF_Emp</u>	<u>Cod_Pj</u>	Hrs_Trab	Nome_Emp	Nome_Pj	Local_Pj
1234	AMS-200	200	Markus	SistSem	MG
4321	UCM-300	300	João	SistAmz	RO
5678	ATO-400	220	Lidia	SistSoj	MG
1234	UCM-300	280	Markus	SistAmz	PR



Segunda Forma Normal (2FN)

Nesta tabela identificamos mais duas dependências parciais (Nome_Pj e Local_Pj) que dependem somente de parte (Cod_Pj) da chave.

EmpregadoProjeto

<u>CPF_Emp</u>	<u>Cod_Pj</u>	Hrs_Trab	Nome_Emp	Nome_Pj	Local_Pj
1234	AMS-200	200	Markus	SistSem	MG
4321	UCM-300	300	João	SistAmz	RO
5678	ATO-400	220	Lidia	SistSoj	MG
1234	UCM-300	280	Markus	SistAmz	PR



Segunda Forma Normal (2FN)

Para estar na 2FN os campos não chave deveriam depender exclusivamente da chave primaria (composta) assim como o campo (Hrs_Trab). Isso implica em alguns problemas no momento de inserção, exclusão e atualização dos dados no banco de dados.

EmpregadoProjeto

<u>CPF_Emp</u>	<u>Cod_Pj</u>	Hrs_Trab	Nome_Emp	Nome_Pj	Local_Pj
1234	AMS-200	200	Markus	SistSem	MG
4321	UCM-300	300	João	SistAmz	RO
5678	ATO-400	220	Lidia	SistSoj	MG
1234	UCM-300	280	Markus	SistAmz	PR



Neste caso de inserção, podemos perceber que ao inserirmos um novo registro os campos (Nome_Pj) e (Local_Pj), para este caso, começam a ter dados duplicados, o que viola a 1FN. Ou seja, se incluirmos uma pessoa em um projeto que já existe, repetimos os dados do projeto, isso é uma redundância desnecessária.

EmpregadoProjeto

<u>CPF_Emp</u>	<u>Cod_Pj</u>	Hrs_Trab	Nome_Emp	Nome_Pj	Local_Pj
1234	AMS-200	90	Markus	SistSem	MG
4321	UCM-300	300	João	SistAmz	RO
5678	ATO-400	220	Lidia	SistSoj	MG
1234	UCM-300	90	Markus	SistAmz	PR
2244	AMS-200	100	Carlos	SistSem	MG

Neste caso de exclusão, se excluirmos a “Lidia”, os dados do projeto “SistSoj” também são excluídos. Isso é um problema de anomalia de exclusão dos dados ocasionado pela falta de normalização.

EmpregadoProjeto

<u>CPF_Emp</u>	<u>Cod_Pj</u>	Hrs_Trab	Nome_Emp	Nome_Pj	Local_Pj
1234	AMS-200	90	Markus	SistSem	MG
4321	UCM-300	300	João	SistAmz	RO
5678	ATO-400	220	Lidia	SistSoj	MG
1234	UCM-300	90	Markus	SistAmz	PR

Na atualização, ao atualizarmos os dados do “SistAmz”, por exemplo: o “Local_Pj”, precisamos percorrer e atualizar todos os campos como podemos perceber na tabela abaixo. Isso também é um problema de anomalia, agora de atualização dos dados, ocasionado pela falta de normalização.

EmpregadoProjeto

<u>CPF_Emp</u>	<u>Cod_Pj</u>	Hrs_Trab	Nome_Emp	Nome_Pj	Local_Pj
1234	AMS-200	90	Markus	SistSem	MG
4321	UCM-300	300	João	SistAmz	RO
5678	ATO-400	220	Lidia	SistSoj	MG
1234	UCM-300	90	Markus	SistAmz	RO

Segunda Forma Normal (2FN)

Para resolver estes problemas precisamos normalizar esta tabela, identificando o campos que são determinados pela chave primária e derivando novas tabelas, e também definindo a integridade referencial entre estas novas tabelas.

Segunda Forma Normal (2FN)

EmpregadoProjeto

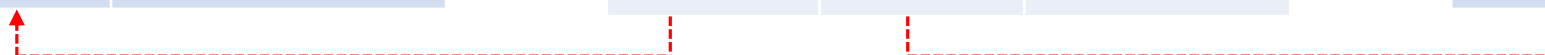
<u>CPF_Emp</u>	<u>Cod_Pj</u>	Hrs_Trab	Nome_Emp	Nome_Pj	Local_Pj
1234	AMS-200	200	Markus	SistSem	MG
4321	UCM-300	300	João	SistAmz	RO
5678	ATO-400	220	Lidia	SistSoj	MG
1234	UCM-300	280	Markus	SistAmz	PR



<u>CPF_Emp</u>	Nome_Emp
1234	Markus
4321	João
5678	Lidia

<u>CPF_Emp</u>	<u>Cod_Pj</u>	Hrs_Trab
1234	AMS-200	90
4321	UCM-300	300
5678	ATO-400	220
1234	UCM-300	90

<u>Cod_Pj</u>	Nome_Pj	Local_Pj
AMS-200	SistSem	MG
UCM-300	SistAmz	RO
ATO-400	SistSoj	MG



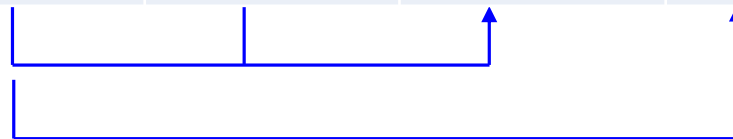
- Uma relação está em 3FN se e somente se estiver na 2FN e nenhum atributo não-primo (isto é, que não seja membro de uma chave) for **transitivamente dependente** da chave primária.
 - **Dependência Transitiva:** ocorre quando uma coluna, além de depender da chave primária de uma tabela, depende de outra coluna ou conjunto de colunas da tabela.
 - Uma relação para estar na 3FN não deve ter um atributo não-chave **funcionalmente determinado** por um outro atributo não-chave, ou seja, para uma tabela estar na 3FN deve estar na 2FN e não conter **dependências transitivas**.

Terceira Forma Normal (3FN)

Na tabela abaixo, podemos identificar que os campos que são determinados pela chave primária “CPF_Emp” são: Nome_Emp, Sexo_Emp e Cod_Dep.

EmpregadoDepartamento

<u>CPF_Emp</u>	Nome_Emp	Sexo_Emp	Cod_Dep	Nome_Dep	Ger_Dep
1234	Markus	M	1a	RH	Sheila
4321	João	M	2b	ADM	Tertuliano
5678	Lidia	F	3c	TI	Fabio
8765	Carlos	M	2b	ADM	Tertuliano



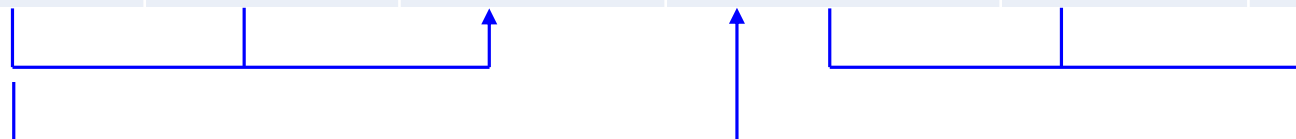
Terceira Forma Normal (3FN)

Neste caso, podemos identificar que os campos que são determinados pela chave primária “**CPF_Emp**” são: **Nome_Emp**, **Sexo_Emp** e **Cod_Dep**. E o campo **Nome_Dep** não depende do campo **CPF_Emp** mas sim, do campo **Cod_Dep** assim como o campo **Ger_Dep**. Ou seja, temos dois campos não chave (**Nome_Dep** e **Ger_Dep**) dependendo de outro campo (**Dep_Cod**) não chave.

Então dizemos que **CPF_Emp** determina transitivamente **Dep_Nome** e **Dep_Ger**, porque **CPF_Emp** determina **Cod_Dep** e **Cod_Dep** determina **Nome_Dep**.

EmpregadoDepartamento

<u>CPF_Emp</u>	Nome_Emp	Sexo_Emp	Cod_Dep	Nome_Dep	Ger_Dep
1234	Markus	M	1a	RH	Sheila
4321	João	M	2b	ADM	Tertuliano
5678	Lidia	F	3c	TI	Fabio
8765	Carlos	M	2b	ADM	Tertuliano



Neste caso de inserção podemos perceber, que ao inserirmos um novo registro, os campos (Nome_Dep) e (Ger_Dep), para este caso, começam a ter dados repetidos. Ou seja, se incluirmos uma pessoa em um departamento que já existe, repetimos os dados do departamento, isso é uma redundância desnecessária.

EmpregadoDepartamento

<u>CPF_Emp</u>	Nome_Emp	Sexo_Emp	Cod_Dep	Nome_Dep	Ger_Dep
1234	Markus	M	1a	RH	Sheila
4321	João	M	2b	ADM	Tertuliano
5678	Lidia	F	3c	TI	Fabio
8765	Carlos	M	2b	ADM	Tertuliano
2244	Paulo	M	3c	TI	Fabio

Neste caso de exclusão, se excluirmos a “Lidia”, o departamento “TI” também será excluído. Isso é um problema de anomalia de exclusão.

EmpregadoDepartamento

<u>CPF_Emp</u>	Nome_Emp	Sexo_Emp	Cod_Dep	Nome_Dep	Ger_Dep
1234	Markus	M	1a	RH	Sheila
4321	João	M	2b	ADM	Tertuliano
5678	Lidia	F	3c	TI	Fabio
8765	Carlos	M	2b	ADM	Tertuliano

Na atualização, ao atualizarmos os dados do departamento “ADM”, por exemplo: o “Ger_Dep”, precisamos percorrer e atualizar todos os campos que tem o nome do departamento com “ADM”, como podemos perceber na tabela abaixo. Isso também é um problema de anomalia dos dados, ocasionado pela falta de normalização.

EmpregadoDepartamento

<u>CPF_Emp</u>	Nome_Emp	Sexo_Emp	Cod_Dep	Nome_Dep	Ger_Dep
1234	Markus	M	1a	RH	Sheila
4321	João	M	2b	ADM	Tertuliano
5678	Lidia	F	3c	TI	Fabio
8765	Carlos	M	2b	ADM	Tertuliano

Para resolver estes problemas, precisamos normalizar esta tabela identificando o campos que são determinados pela chave primária e os que têm dependências transitivas, removendo-os para novas tabelas derivadas do domínio analisado, e também definindo a integridade referencial entre estas novas tabelas.

Terceira Forma Normal (3FN)

EmpregadoDepartamento

<u>CPF_Emp</u>	Nome_Emp	Sexo_Emp	Cod_Dep	Nome_Dep	Ger_Dep
1234	Markus	M	1a	RH	Sheila
4321	João	M	2b	ADM	Tertuliano
5678	Lidia	F	3c	TI	Fabio
8765	Carlos	M	2b	ADM	Tertuliano



Empregado

<u>CPF_Emp</u>	Nome_Emp	Sexo_Emp	Cod_Dep
1234	Markus	M	1a
4321	João	M	2b
5678	Lidia	F	3c
8765	Carlos	M	2b

Departamento

<u>Cod_Dep</u>	Nome_Dep	Ger_Dep
1a	RH	Sheila
2b	ADM	Tertuliano
3c	TI	Fabio



Sistema
Fiep
nosso i é de indústria.

FIEP

SESI

SENAI

IEL