

FIAP GRADUAÇÃO

DATA SCIENCE: BIG DATA, BI & DATA ENGINEERING

BUILDING RELATIONAL DATABASE

PROF. TADEU KANASHIRO proftadeu.kanashiro@fiap.com.br
PROFa. RITA DE CÁSSIA rita@fiap.com.br

PROFESSOR: PERÍODO NOTURNO



- **EXPERIÊNCIA PROFISSIONAL:** Mais de 15 anos atuando na área de tecnologia, focado em análise e estruturação de dados;
- **ÁREAS DE ATUAÇÃO:** Big Data, Analytics Engineering, Business Intelligence e Database Marketing;
- **INDÚSTRIAS:** Saúde, setor imobiliário farmacêutica, fintech, financeiro, Internet, telecomunicações, educação e filantropia.
- **MBA:** Big Data (Data Science);
- **GRADUAÇÃO:** Sistemas de Informações e Gestão Financeira.

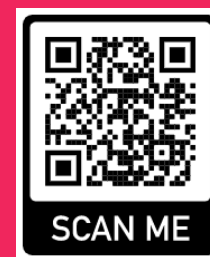
Professor: Tadeu Kanashiro



proftadeu.kanashiro@fiap.com.br



<https://www.linkedin.com/in/tadeukanashiro/>



AGENDA

- Exercício da Aula Anterior;
- Normalização (em Banco de Dados Relacionais);
- Introdução as Formas Normais;
- 1ª Forma Normal;
- 2ª Forma Normal;
- 3ª Forma Normal;
- Processo de Normalização.

EXERCÍCIO DA AULA ANTERIOR

PRODUTO	Tabela cadastral de produtos			
COLUNA	TIPO	CONSTRAINT	SENSÍVEL	ROTINA LIMPEZA
COD_PRODUTO	NUMÉRICO	PK	N	N/A
COD_CATEGORIA	NUMÉRICO	FK	N	N/A
PRODUTO	TEXTO		N	N/A

CATEGORIA	Tabela domínio que armazena todas as categoriza disponíveis para os produtos			
COLUNA	TIPO	CONSTRAINT	SENSÍVEL	ROTINA LIMPEZA
COD_CATEGORIA	NUMÉRICO	PK	N	N/A
CATEGORIA	TEXTO		N	N/A

ITEM_PRODUTO	Tabela associativa oriunda do PRODUTO e da MATERIA_PRIMA			
COLUNA	TIPO	CONSTRAINT	SENSÍVEL	ROTINA LIMPEZA
COD_PRODUTO	NUMÉRICO	PK, FK	N	N/A
COD_MATERIA_PRIMA	NUMÉRICO	PK, FK	N	N/A
QUANTIDADE	NUMÉRICO		N	N/A

MATERIA_PRIMA	Tabela cadastral de matérias primas			
COLUNA	TIPO	CONSTRAINT	SENSÍVEL	ROTINA LIMPEZA
COD_MATERIA_PRIMA	NUMÉRICO	PK	N	N/A
MATERIA_PRIMA	TEXT		N	N/A

NORMALIZAÇÃO (EM BANCO DE DADOS RELACIONAIS)

I INTRODUÇÃO: NORMALIZAÇÃO

- A **normalização** é um **procedimento que examina** os **atributos** de uma entidade **com** o **objetivo de evitar anomalias** que possam ocorrer na inclusão, na exclusão ou na alteração de uma ocorrência específica em uma entidade;
- Esse procedimento procura simplificar os atributos, nas respectivas ocorrências, eliminando grupos repetitivos, dependência parcial da chave primária, dependências transitivas, dados redundantes e dependência de atributos multivalorados, contribuindo para a estabilização do modelo de dados e para a redução de manutenções;
- A normalização é baseada em relações matemáticas (teoria dos conjuntos) e tem como objetivo estruturar relações entre as entidades, garantindo que não existam anomalias ou redundâncias no armazenamento de dados. As **anomalias ocorrem durante** o processo de **manipulação dos dados** e se tornam severas durante as **operações** de **inclusão**, **alteração e exclusão** de dados.

ANOMALIA: NA INCLUSÃO DE DADOS

- Ocorre quando **o registro de uma entidade possui dados incompletos**. Nesse caso, os dados podem ser parte de outras entidades ainda não definidas, ocasionando, por exemplo, a existência de nulos.

TABELA: COLABORADOR_DEPARTAMENTO					
CODIGO	NOME	SEXO	DT_NASC	COD_DEPTO	NOME_DEPTO
0001	JOÃO	M	20/02/1950	100	TECNOLOGIA
0002	MARIA	F	19/08/1967	101	FINANCEIRO
0003	JOSÉ	M	23/12/1970	101	FINANCEIRO
0004	JOANA	F	26/03/1968	102	RECURSOS HUMANOS
NULO				103	MARKETING
NULO				104	CONTABILIDADE

ANOMALIA: NA EXCLUSÃO DE DADOS

- Ocorre quando **a exclusão de um dado pode causar a eventual perda de outro.**

TABELA: COLABORADOR_DEPARTAMENTO					
CODIGO	NOME	SEXO	DT_NASC	COD_DEPTO	NOME_DEPTO
0001	JOÃO	M	20/02/1950	100	TECNOLOGIA
0002	MARIA	F	19/08/1967	101	FINANCEIRO
0003	JOSÉ	M	23/12/1970	101	FINANCEIRO
0004	JOANA	F	26/03/1968	102	RECURSOS HUMANOS
NULO				103	MARKETING
NULO				104	CONTABILIDADE

ANOMALIA: NA ALTERAÇÃO DE DADOS

- Ocorre quando **há** a **redundância** no armazenamento de dados, havendo a **necessidade de atualização em diferentes entidades**.

TABELA: COLABORADOR_DEPARTAMENTO					
CODIGO	NOME	SEXO	DT_NASC	COD_DEPTO	NOME_DEPTO
0001	JOÃO	M	20/02/1950	100	TECNOLOGIA
0002	MARIA	F	19/08/1967	101	FINANCEIRO
0003	JOSÉ	M	23/12/1970	101	FINANCEIRO
0004	JOANA	F	26/03/1968	102	RECURSOS HUMANOS
NULO				103	MARKETING
NULO				104	CONTABILIDADE

I DEFINIÇÃO E OBJETIVOS DA NORMALIZAÇÃO

- A normalização é uma técnica da análise de dados que visa determinar a melhor formação para uma estrutura de dados;
- Principais objetivos:
 - Eliminar detalhes que dificultam as operações sobre os dados;
 - Minimizar as redundâncias e os consequentes riscos de inconsistências;
 - Reduzir e facilitar as manutenções.

INTRODUÇÃO AS FORMAS NORMAIS

AS FORMAS NORMAIS

- Com o **objetivo de** estabelecer critérios para realizar a **normalização** das entidades, **Edgar Codd**, em 1.972, **definiu** as **Formas Normais** (FN), as quais foram apresentadas:
 - 1ª Forma Normal (**1FN**);
 - 2ª Forma Normal (**2FN**);
 - 3ª Forma Normal (**3FN**).
- Posteriormente, foram acrescentadas:
 - Forma Normal de Boyce-Codd (FNBC) ou Boyce-Codd Normal Form (BCNF);
 - 4ª Forma Normal (4FN);
 - 5ª Forma Normal (5FN).

DESCRITIVOS DAS FORMAS NORMAIS

- 1ª Forma Normal (1FN): forma normal básica, que verifica a aderência (compatibilidade) da tabela ao Modelo Relacional Normalizado (MRN);
- 2ª Forma Normal (2FN): verifica dependência parcial da chave primária;
- 3ª Forma Normal (3FN): verifica dependência transitiva;
- Forma Normal de Boyce-Codd): pode ser considerada com um aperfeiçoamento da 3FN;
- 4ª Forma Normal (4FN): verifica dependência multivalorada;
- 5ª Forma Normal (5FN): verifica dependência de junção;
- **Geralmente um Banco de Dados já é considerado normalizado se estiver na 3FN.** Dificilmente, em um projeto de Modelagem de Banco de Dados, há a implementação das outras formas (FNBC, 4FN e 5FN).



1ª FORMA NORMAL

I DEFINIÇÃO CONCEITUAL

“Uma estrutura de dados encontra-se na 1ª Forma Normal se todas as suas colunas não apresentarem atributos multivalorados, ou seja, atributos que podem conter mais de um valor para uma mesma instância da entidade.”

TRADUZINDO O CONCEITO ATRIBUTO MULTIVALORADO

TABELA: CLIENTE				
CODIGO	NOME	SEXO	D_NASC	TELEFONE
0001	JOÃO	M	20/02/1950	(11) 5050-4040
0001	JOÃO	M	20/02/1950	(11) 9-7878-9090
0002	MARIA	F	19/08/1967	(11) 3535-1111
0003	JOSÉ	M	23/12/1970	(11) 5055-6060
0003	JOSÉ	M	23/12/1970	(11) 9-2222-3535
0003	JOSÉ	M	23/12/1970	(11) 9-7676-3434
0004	JOANA	F	26/03/1968	(11) 5040-7766

1ª FORMA NORMAL

- Uma entidade está na 1FN se:
 - Não possuir atributos multivalorados ou grupos repetitivos;
 - Todos os atributos estiverem no formato atômico, isto é, não forem compostos por múltiplas partes;
 - Existe uma chave primária que identifica apenas uma ocorrência da entidade;
 - As ocorrências da entidade forem todas distintas entre si.

TABELA: TELEFONE		
CODIGO	COD_CLIENTE	TELEFONE
1	0001	(11) 5050-4040
2	0001	(11) 9-7878-9090
3	0002	(11) 3535-1111
4	0003	(11) 5055-6060
5	0003	(11) 9-2222-3535
6	0003	(11) 9-7676-3434
7	0004	(11) 5040-7766

TABELA: CLIENTE			
CODIGO	NOME	SEXO	D_NASC
0001	JOÃO	M	20/02/1950
0002	MARIA	F	19/08/1967
0003	JOSÉ	M	23/12/1970
0004	JOANA	F	26/03/1968

TRADUZINDO O CONCEITO

ATRIBUTO COMPOSTO

TABELA: CLIENTE				
CODIGO	NOME	SEXO	D_NASC	ENDERECO
0001	JOÃO	M	20/02/1950	AV. LINS DE VASCONCELOS, 1222 - ACLIMAÇÃO
0002	MARIA	F	19/08/1967	AV. PAULISTA, 1106 - CERQUEIRA CÉSAR
0003	JOSÉ	M	23/12/1970	RUA OLIMPÍADAS, 186 - VILA OLÍMPIA
0004	JOANA	F	26/03/1968	ALAMEDAARAGUAIA, 1142 - ALPHAVILLE

1ª FORMA NORMAL

- Uma entidade está na 1FN se:
 - Não possuir atributos multivalorados ou grupos repetitivos;
 - Todos os atributos estiverem no formato atômico, isto é, não forem compostos por múltiplas partes;
 - Existe uma chave primária que identifica apenas uma ocorrência da entidade;
 - As ocorrências da entidade forem todas distintas entre si.

TABELA: CLIENTE					
CODIGO	NOME	SEXO	D_NASC	LOGRADOURO	BAIRRO
0001	JOÃO	M	20/02/1950	AV. LINS DE VASCONCELOS, 1222	ACLIÇÃO
0002	MARIA	F	19/08/1967	AV. PAULISTA, 1106	CERQUEIRA CÉSAR
0003	JOSÉ	M	23/12/1970	RUA OLIMPIADAS, 186	VILA OLÍMPIA
0004	JOANA	F	26/03/1968	ALAMEDAARAGUAIA, 1142	ALPHAVILLE

2ª FORMA NORMAL

I DEFINIÇÃO CONCEITUAL

“Uma estrutura de dados está na segunda 2ª Forma Normal se ela já estiver na 1ª Forma Normal e não há colunas não chaves dependentes de uma chave primária parcial dessa tabela.”

TRADUZINDO O CONCEITO DEPENDÊNCIA PARCIAL

TABELA: PROJETO					
COD_PROF *	COD_PROJ *	HORAS_PROJ	PROFISSIONAL	PROJETO	CLIENTE_PROJ
0001	1010	120	JOÃO	BIG DATA	BANCO +DINHEIRO
0001	1020	300	JOÃO	BI	CORRETORA TOP
0002	1010	60	MARIA	BIG DATA	BANCO +DINHEIRO
0003	1030	400	JOSÉ	DATA SCIENCE	SAUDE IDOSOS
0003	1040	120	JOSÉ	SCRUM	STARTUP NET
0003	1050	80	JOSÉ	ANALYTICS	VAREJO BRASIL
0004	1010	100	JOANA	BIG DATA	BANCO +DINHEIRO

(*) Chave Primária Composta: Cod_Prof e Cod_Proj

2ª FORMA NORMAL

- Uma entidade está na 2FN se:
 - Estiver na 1FN;
 - Não possuir atributos com dependência parcial da chave primária.

TABELA: PROFISSIONAL	
COD_PROF *	PROFISSIONAL
0001	JOÃO
0002	MARIA
0003	JOSÉ
0004	JOANA

TABELA: PROJETO		
COD_PROJ *	PROJETO	CLIENTE_PROJ
1010	BIG DATA	BANCO +DINHEIRO
1020	BI	CORRETORA TOP
1030	DATA SCIENCE	SAUDE IDOSOS
1040	SCRUM	STARTUP NET
1050	ANALYTICS	VAREJO BRASIL

TABELA: PROJETO_PROFISSIONAL		
COD_PROF *	COD_PROJ *	HORAS_PROJ
0001	1010	120
0001	1020	300
0002	1010	60
0003	1030	400
0003	1040	120
0003	1050	80
0004	1010	100

3ª FORMA NORMAL

I DEFINIÇÃO CONCEITUAL

“Uma estrutura de dados encontra-se na 3ª Forma Normal se já estiver na 2ª Forma Normal e se não existir a anomalia da dependência transitiva ou indireta entre um atributo e a chave primária, portanto uma estrutura de dados estará na 3ª Forma Normal se todos os seus atributos dependerem funcionalmente apenas da chave primária e de nenhum outro atributo.”

TRADUZINDO O CONCEITO DEPENDÊNCIA TRANSITIVA

TABELA: VENDA				
NOTA_FISCAL *	COD_VENDEDOR	NOME_VENDEDOR	COD_PRODUTO	QTDE_VENDIDA
1010	0001	JOÃO	100100	100
1011	0001	JOÃO	100200	320
1012	0002	MARIA	100100	150
1013	0003	JOSÉ	100100	240
1014	0003	JOSÉ	100300	300
1015	0003	JOSÉ	100400	400
1016	0004	JOANA	100100	250

(*) Chave Primária: NOTA_FISCAL

3ª FORMA NORMAL

- Uma entidade está na 3FN se:
 - Estiver na 2FN;
 - Não existir dependência transitiva entre os atributos, ou conjunto de atributos, não pertencentes à chave primária;
 - Todos os atributos que não pertencem à chave primária dependerem exclusivamente da chave primária.

TABELA: VENDA			
NOTA_FISCAL *	COD_VENDEDOR	COD_PRODUTO	QTDE_VENDIDA
1010	0001	100100	100
1011	0001	100200	320
1012	0002	100100	150
1013	0003	100100	240
1014	0003	100300	300
1015	0003	100400	400
1016	0004	100100	250

TABELA: VENDEDOR	
COD_VENDEDOR *	NOME_VENDEDOR
0001	JOÃO
0002	MARIA
0003	JOSÉ
0004	JOANA

PROCESSO DE NORMALIZAÇÃO

■ “PASSO A PASSO” PARA A NORMALIZAÇÃO

- Tabela Desnormalizada;
- 1FN: Remover os Atributos Multivalorados e Compostos;
- 2FN: Remover Dependências Parciais;
- 3FN: Remover Dependências Transitivas.

EXERCÍCIO

EXERCÍCIOS

1. Revisar os três modelos criados anteriormente e checar se estão até a 3ª forma normal.
2. Construir o modelo abaixo e garantir que a proposta esteja até a 3ª forma normal.
 - Um aplicativo de streaming de músicas lançou um concurso musical para a escolha das melhores performances (interpretações) musicais.
 - Cada música pode ter várias interpretações e cada interpretação pode ter artistas variados.
 - Os ouvintes votam na interpretação favorita (cada um possui apenas um único voto).
 - Obs.: criar os atributos que achar necessário.

Copyright © 2024 Prof. Tadeu Kanashiro e Prof. André Santos

Todos direitos reservados. Reprodução ou divulgação total ou parcial deste documento é expressamente proibido sem o consentimento formal, por escrito, do Professor (autor).