

Modelagem de dados

Agenda:

- Resumindo o modelo relacional;
- Visão geral da álgebra relacional;
- Normalização;



Bancos de dados relacionais

- Um dos principais tipos de bancos de dados são os bancos de dados relacionais.
- Pode-se dizer que se trata de um alicerce formal, ou uma teoria, que é a base para a construção de SGBDs de alto desempenho.
- O modelo relacional representa o banco de dados como uma coleção de relações.
- Cada linha de uma tabela representa uma coleção de valores de dados relacionados, as RelVars. Uma linha representa um fato que normalmente corresponde a uma entidade ou relacionamento do mundo real. Os nomes da tabela e de coluna são usados para ajudar a interpretar o significado dos valores em cada linha.

Linha = registro = tupla = entidade e conjunto entidade = tabela = relação

Bancos de dados relacionais

- Constantemente descrito como tendo os três aspectos:
 - Aspecto estrutural: percepção ao usuário como tabelas;
 - Aspecto de integridade: tabelas satisfazem restrições de integridade;
 - Aspecto manipulador: possuem operadores de restrição, projeção e junção.
-
- O aspecto manipulador garante operações básicas da álgebra relacional.

Bancos de dados relacionais

- Banco de dados EMPRESA
- Temos de fato a percepção dos dados em tabelas.

DEPTO#	NOMEDEPTO	ORCAMENTO
D1	Marketing	10M
D2	Desenvolvimento	12M
D3	Pesquisa	5M

EMP#	NOMEEMP	DEPTO#	SALÁRIO
E1	Lopez	D1	40K
E2	Cheng	D1	42K
E3	Finzi	D2	30K
E4	Saito	D2	35K

Bancos de dados relacionais

Banco de dados EMPRESA

Restrição: operação que extrai linhas específicas de uma tabela. Também chamada de seleção.

Projeção: operação que extrai colunas específicas de uma tabela.

Junção: operação que une duas ou mais tabelas com base em valores comuns em uma coluna comum.

DEPTO#	NOMEDEPTO	ORCAMENTO
D1	Marketing	10M
D2	Desenvolvimento	12M
D3	Pesquisa	5M

EMP#	NOMEEMP	DEPTO#	SALÁRIO
E1	Lopez	D1	40K
E2	Cheng	D1	42K
E3	Finzi	D2	30K
E4	Saito	D2	35K

Bancos de dados relacionais

Banco de dados EMPRESA

Restrição:

DEPTOs onde ORCAMENTO > 8M

DEPTO#	NOMEDEPTO	ORCAMENTO
D1	Marketing	10M
D2	Desenvolvimento	12M

DEPTOs onde ORCAMENTO = 5M

DEPTO#	NOMEDEPTO	ORCAMENTO
D3	Pesquisa	5M

Bancos de dados relacionais

Banco de dados EMPRESA

Projeção:

DEPTOs sobre DEPTO#, ORCAMENTO

DEPTO#	ORCAMENTO
D1	10M
D2	12M
D3	5M

Bancos de dados relacionais

Banco de dados EMPRESA

Junção:

DEPTOs e EMPs sobre DEPTO#

DEPTO#	NOMEDEPTO	ORCAMENTO	EMP#	NOMEEMP	SALARIO
D1	Marketing	10M	E1	Lopez	40K
D1	Marketing	10M	E2	Cheng	42K
D2	Desenvolvime nto	12M	E3	Finzi	30K
D2	Desenvolvime nto	12M	E4	Saito	35K

Bancos de dados relacionais

- Importante!!!!
 -
 - O resultado de cada operação é uma nova tabela!
 -
 - Isto leva à uma importante propriedade dos bancos de dados relacionais, o **fechamento** (closure): como a saída de uma operação é uma tabela com os mesmos tipos de dados da tabela de entrada, a saída de uma operação pode ser a entrada de outra.
 -
- Projeção de junção;
Junção de restrições;
Restrição de uma projeção.

Bancos de dados relacionais

- Importante, parte 2, a revanche!!!!
-
- Observe que os sistemas relacionais só exigem que o banco de dados seja percebido pelo usuário como tabelas. Elas são a estrutura lógica, e não a física;
-
- Bancos de dados relacionais satisfazem a um princípio chamado O PRINCÍPIO DA INFORMAÇÃO, onde todo o conteúdo de informação do BD é representado de um único modo, i.e., como valores explícitos em posições de colunas em linhas de tabelas. Em outras palavras, não há ponteiros conectando dados de uma tabela à outra.

Bancos de dados relacionais

Chaves:

- É preciso ter uma maneira de especificar como as tuplas dentro de uma determinada relação são distinguidas. Isto deve ser expresso em termos de seus atributos, i.e., os valores de atributo de uma tupla precisam ser tais que possam identificar unicamente a tupla.
- Uma **superchave** é um único atributo ou conjunto de atributos que nos permite identificar unicamente uma tupla da relação.
- **Chave primária** é, então, uma restrição importante em uma tabela.
 - Deve ser um atributo determinante.

Visão da álgebra relaciona

- As operações de Projeção e Restrição são operações da álgebra relacional que operam em uma única relação, e são chamadas de operações unárias.
-
- Já operações que operam em pares de relações são chamadas de operações binárias, como visto em JUNÇÃO.

Visão da álgebra relacional

- Operação Seleção(?):
- Assim como descrito em RESTRIÇÃO, esta operação seleciona tuplas que satisfaçam um determinado predicado.
- Usa-se a letra grega sigma σ para denotar a seleção. O predicado aparece como um subscrito de sigma (ϕ). A relação de argumento está entre parênteses após o sigma (R).

$$\sigma_{\phi}(R)$$

- No exemplo, para selecionar as tuplas da relação DEPTO onde ORCAMENTO for maior que 5M, escreve-se:

$$\sigma_{ORCAMENTO > 5M}(DEPTO)$$

Visão da álgebra relacional

- Operação de Projeção(π):
- Projeta uma nova tabela com as colunas especificadas.
- Usa-se a letra grega π para denotar a projeção. Os elementos a_i denotam as colunas da projeção, de forma sobrescrita à π , e R é a relação:

$$\pi_{(a_1, \dots, a_n)}(R)$$

$$\pi_{(nome, telefone)}(CLIENTE)$$

Visão da álgebra relacional

- Operação Renomear(ρ):

- Projeta uma nova tabela com as colunas especificadas.


- Usa-se a letra grega rho ρ para denotar a renomeação. Onde o resultado é idêntico ao original, exceto pelo fato de que o campo b de todas as tuplas da saída será renomeado para a :

$$\rho_{a/b}(R)$$

- Para mudar o campo preco para valorVenda, utiliza-se:


$$\rho_{valorVenda/preco}(PRODUTO)$$

Visão da álgebra relacional

- Operação Junção Natural  (bowtie – gravata borboleta):
- Escrita como $(R \bowtie S)$ onde R e S são relações, e o resultado é uma tabela com todas as combinações das tuplas em R e S, onde seus atributos sejam comuns.

<i>Empregado</i>		
Nome	IdEmp	DeptNome
Harry	3415	Finanças
Sally	2241	Vendas
George	3401	Finanças
Harriet	2202	Vendas

<i>Departamento</i>	
DeptNome	Gerente
Finanças	George
Vendas	Harriet
Produção	Charles

<i>Empregado</i>  <i>Departamento</i>			
Nome	IdEmp	DeptNome	Gerente
Harry	3415	Finanças	George
Sally	2241	Vendas	Harriet
George	3401	Finanças	George
Harriet	2202	Vendas	Harriet

Visão da álgebra relacional

- Operação semijunção (\bowtie) (\ltimes):
- Escrita como $(R \ltimes S)$ R e S são relações, e o resultado desta semijunção à esquerda é o conjunto de todas as tuplas em R para o qual exista uma tupla em S, com igual em seus atributos comuns:

<i>Empregado</i>			<i>Dept</i>		<i>Empregado \ltimes Dept</i>		
Nome	IdEmp	NomeDept	NomeDept	Gerente	Nome	IdEmp	NomeDept
Harry	3415	Finanças	Vendas	Bob	Sally	2241	Vendas
Sally	2241	Vendas	Vendas	Thomas	Harriet	2202	Produção
George	3401	Finanças	Produção	Katie			
Harriet	2202	Produção	Produção	Marco			

Visão da álgebra relacional

- Operação antijunção (\triangleright):
- Escrita como $(R \triangleright S)$, onde R e S são relações, e o resultado é parecido com a semijunção, mas o agora a saída será somente aquelas tuplas que não estão presentes nem em R e nem em S:

<i>Empregado</i>			<i>Departamento</i>		<i>Empregado \triangleright Departamento</i>		
Nome	IdEmp	DeptNome	DeptNome	Gerente	Nome	IdEmp	DeptNome
Harry	3415	Finanças	Vendas	Harriet	Harry	3415	Finanças
Sally	2241	Vendas	Produção	Charles	George	3401	Finanças
George	3401	Finanças					
Harriet	2202	Produção					

Normalização

- Dado um esquema da relação, precisamos decidir se ele é bom projeto ou se precisamos decompô-lo em relações menores.
- Tal decisão deve ser conduzida por um entendimento de quais problemas (se houver) surgem a partir do esquema corrente.

Para conduzir a este caminho, diversas formas normais foram propostas, onde se um esquema de relação está em uma dessas formas normais, sabemos que certos tipos de problemas não podem surgir.

Normalização

- Conjunto de regras que leva à construção de modelos mais robustos, com menos dependências entre seus elementos e menos redundância de informações.
- Normalização é uma atividade, ou conjunto de atividades, de verificação do modelo lógico visando o modelo ótimo.

Formas mais comuns (Date, 2004):

- 1ª Forma Normal (1FN)
- 2ª Forma Normal (2FN)
- 3ª Forma Normal (3FN)

21

Normalização

- Importante: ideia de que atributos não possuem qualquer subestrutura.
- O modelo E-R permite que conjuntos de entidades e conjuntos de relacionamentos tenham atributos com algum grau de subestrutura, mas ele também permite atributos de valores múltiplos(multivalorados) e atributos compostos.
- Quando criamos modelos com estes tipos de atributos, eliminamos as subestruturas.
- Para atributos compostos, espera-se que cada componente da composição do atributo seja um atributo por si só.
- Para atributos de valores múltiplos, cria-se uma tupla para cada item em um conjunto de valores múltiplos.

Normalização

1ª Forma Normal (1FN)

-
-
- Uma relação está em 1FN se todos atributos contém apenas valores atômicos, i.e., nenhuma lista nem conjuntos.
- Toda relação deve ter uma chave primária, e atributos compostos devem ser separados em novos atributos.

Normalização

1ª Forma Normal (1FN)

ex.:

nome	endereço	fone
Marcos da Silva	Rua 02, nº 10, centro, Formosa-go	66-6666-9999
Fabrício Carvalho	Avenida 14, nº 1, centro, Jataí	77-0000-00000

Normalização

1ª Forma Normal (1FN)

ex.:

id	nome	tipoLog	log	nro	bairro	cidade	uf	fone
1	Marcos da Silva	Rua	02	10	centro	Formosa	go	66-6666-9999
2	Fabício Carvalho	Avenida	14	1	centro	Jataí	go	77-0000-00000

Normalização

1ª Forma Normal (1FN)

ex.:

id	nome	tipoLog	log	nro	bairro	cidade	uf	fone
1	Marcos da Silva	Rua	02	10	centro	Formosa	go	66-6666-9999
2	Fabício Carvalho	Avenida	14	1	centro	Jataí	go	77-0000-00000

Normalização

1ª Forma Normal (1FN)

ex 2.:

id	telefone	cliente
01	6666-9999 2222-8787	cli1
02	8888-0000 2325-6767 2325-6868	cli2

Normalização

1ª Forma Normal (1FN)

ex 2.:

idCliente	telefone
01	6666-9999
01	2222-8787
02	8888-0000
02	2325-6767
02	2325-6868

id	cliente
01	cli1
02	cli2

Normalização

- 2ª Forma Normal (2FN)
-
- Uma tabela está na segunda forma normal se ela estiver na primeira forma normal e se cada coluna não-chave depender totalmente da coluna-chave.

×

Normalização

2ª Forma Normal (2FN)

Uma tabela está na segunda forma normal se ela estiver na primeira forma normal e se cada coluna não-chave depender totalmente da coluna-chave.

ex.:

cod	depósito	qtd	Telefone-depósito	nroContaineres
T2	RJ	467	222-3111	47
T2	SP	319	200-0011	32
T2	BSB	121	123-4567	13
H9	SP	578	200-0011	58
H9	RJ	227	222-3111	23

Normalização

2ª Forma Normal (2FN)

O telefone do depósito depende do campo depósito, porém não depende do campo CódigoPeca, e não está no 2FN.

ex.:

CódigoPeca	deposito	qtd	Telefone-depósito	nroContaineres
T2	RJ	467	222-3111	47
T2	SP	319	200-0011	32
T2	BSB	121	123-4567	13
H9	SP	578	200-0011	58
H9	RJ	227	222-3111	23

Normalização

2ª Forma Normal (2FN)

ex.:

CodigoPeca	deposito	qtd	nroContaineres
T2	RJ	467	47
T2	SP	319	32
T2	BSB	121	13
H9	SP	578	58
H9	RJ	227	23

Depósito	Telefone-depósito
RJ	222-3111
SP	200-0011
BSB	123-4567

Normalização

3ª Forma Normal (3FN)

Uma tabela está na terceira forma normal se ela estiver na segunda forma normal e se nenhuma coluna não-chave depender de outra coluna não-chave.

CodigoPeca	deposito	qtd	nroContaineres
T2	RJ	467	47
T2	SP	319	32
T2	BSB	121	13
H9	SP	578	58
H9	RJ	227	23

Normalização

3ª Forma Normal (3FN)

Como o número de containeres é a aproximação superior do número de peças dividido por 10, esta coluna não é necessária na 3ª forma normal.

CodigoPeca	deposito	qtd
T2	RJ	467
T2	SP	319
T2	BSB	121
H9	SP	578
H9	RJ	227

Normalização

Exercitando

Avalie a relação e normalize-a, caso necessário for.

nro	autor	assunto	titulo	editora	cidade_editora	ano_publicacao
1	Paulo Coelho	Exoterismo	O Alquimista	Objetiva	Rio de Janeiro	1995
2	Nicholas Sparks	Romance	A última música	Érica	São Paulo	2012
3	Paulo Coelho	Exoterismo	Onze Minutos	Objetiva	Rio de Janeiro	1998

1FN: Uma relação está em 1FN se todos atributos contém apenas valores atômicos, i.e., nenhuma lista nem conjuntos.

2FN: Uma tabela está na segunda forma normal se ela estiver na primeira forma normal e se cada coluna não-chave depender totalmente da coluna-chave.

3FN: Uma tabela está na terceira forma normal se ela estiver na segunda forma normal e se nenhuma coluna não-chave depender de outra coluna não-chave.

Normalização

Exercitando

1FN: Uma relação está em 1FN se todos atributos contém apenas valores atômicos, i.e., nenhuma lista nem conjuntos.

nro	autor	assunto	titulo	editora	cidade_editora	ano_publicacao
1	Paulo Coelho	Exoterismo	O Alquimista	Objetiva	Rio de Janeiro	1995
2	Nicholas Sparks	Romance	A última música	Érica	São Paulo	2012
3	Paulo Coelho	Exoterismo	Onze Minutos	Objetiva	Rio de Janeiro	1998

Análise: temos uma relação com chave primária e nenhuma coluna multivalorada.

Normalização

Exercitando

2FN: Uma tabela está na segunda forma normal se ela estiver na primeira forma normal e se cada coluna não-chave depender totalmente da coluna-chave.

nro	autor	assunto	titulo	editora	cidade_editora	ano_publicacao	idAssAut
1	Paulo Coelho	Exoterismo	O Alquimista	Objetiva	Rio de Janeiro	1995	1
2	Nicholas Sparks	Romance	A última música	Érica	São Paulo	2012	2
3	Paulo Coelho	Exoterismo	Onze Minutos	Objetiva	Rio de Janeiro	1998	1

idAssAut	autor	assunto
1	Paulo Coelho	Exoterismo
2	Nicholas Sparks	Romance

Normalização

Exercitando

3FN: Uma tabela está na terceira forma normal se ela estiver na segunda forma normal e se nenhuma coluna não-chave depender de outra coluna não-chave.

nro	autor	assunto	titulo	editora	cidade_editora	ano_publicacao	idEditora	idAssAut
1	Paulo Coelho	Exoterismo	O Alquimista	Objetiva	Rio de Janeiro	1995	1	1
2	Nicholas Sparks	Romance	A última música	Érica	São Paulo	2012	2	2
3	Paulo Coelho	Exoterismo	Onze Minutos	Objetiva	Rio de Janeiro	1998	1	1

idAssAut	autor	assunto
1	Paulo Coelho	Exoterismo
2	Nicholas Sparks	Romance

idEditora	editora	cidade_editora
1	Objetiva	Rio de Janeiro
2	Érica	São Paulo

Normalização

Exercitando

Resultado da normalização

nro	titulo	ano_publicacao	idEditora	idAssAut
1	O Alquimista	1995	1	1
2	A última música	2012	2	2
3	Onze Minutos	1998	1	1

idAssAut	autor	assunto
1	Paulo Coelho	Exoterismo
2	Nicholas Sparks	Romance

idEditora	editora	cidade_editora
1	Objetiva	Rio de Janeiro
2	Érica	São Paulo

Projeto de BD com E-R

Exercitando

1. Projete um banco de dados para controlar a campanha de times de um campeonato de futebol. O sistema deve ser capaz de armazenar os jogos realizados, os pontos em cada jogo, os jogadores em cada partida, bem como as estatísticas individuais de cada jogador em cada partida. O sistema deve servir para se controlar todos os times participantes do campeonato, independentemente da quantidade de times, e deve ser capaz de guardar dados de campeonatos de anos anteriores.
2. Construa um banco de dados para controlar os dados de uma seguradora de automóveis, onde cada cliente pode ter diversos veículos, desde que haja uma apólice para cada veículo. Deve ser possível se registrar os acidentes ocorridos.

Projeto de BD com E-R

Exercitando

3. Projete um banco de dados para uma pequena mercearia, considerando que é importante se guardar dados sobre os produtos e seus fornecedores, bem como considerar a necessidade de se guardar dados relativos às vendas realizadas. Não há a necessidade de se catalogar todos os clientes, nem relacionar todas as vendas aos clientes, a não ser que a venda seja para pagamento futuro, a famosa “notinha”.
4. Construa um banco de dados para uma faculdade, que deve possibilitar o registro dos alunos, dos cursos, professores e das turmas, além de possibilitar todo controle necessário em uma faculdade.