

Fundamentos de Banco de Dados

Álgebra Relacional



Prof. Manoel Ribeiro

manoel@opencare.com.br



OPENCARE[®]
Inovação e Tecnologia

Álgebra relacional

- Um modelo de banco de dados precisa incluir um conjunto de operações para manipular o banco de dados.
- O conjunto básico de operações para o modelo relacional é a **álgebra relacional**.
- Essas operações permitem que um usuário especifique as solicitações de recuperação básicas como expressões da álgebra relacional.
- O resultado de uma recuperação é uma nova relação, que pode ter sido formada de uma ou mais relações.
- As operações da álgebra, assim, produzem novas relações, que podem ser manipuladas usando operações da mesma álgebra.

Álgebra relacional

- Uma sequência de operações da álgebra relacional forma uma **expressão da álgebra relacional**, cujo resultado também será uma relação que representa o resultado de uma consulta de banco de dados.
- A álgebra relacional fornece um alicerce formal para as operações do modelo relacional.
- Ela também é usada como base para a implementação e otimização de consultas nos módulos de otimização e processamento de consulta dos SGBDRs.
- As operações e funções essenciais nos módulos internos da maioria dos sistemas relacionais são baseadas nas operações da álgebra relacional.
- Alguns dos conceitos da álgebra relacional são incorporados na linguagem SQL.

Álgebra relacional x SQL

- Uma consulta em álgebra relacional é escrita como uma sequência de operações que, quando executadas, produz o resultado exigido.
- Logo, é necessário especificar em que ordem executar as operações de consulta.
- Já a linguagem SQL oferece uma interface de linguagem de consulta declarativa de nível mais alto, de modo que somente é necessário especificar qual deve ser o resultado, deixando as decisões de como executar a consulta para o SGBD.

Operações da álgebra relacional

- As operações da álgebra relacional podem ser divididas em dois grupos:
 - Operações da teoria de conjunto da matemática.
 - UNIÃO, INTERSECÇÃO, DIFERENÇA DE CONJUNTO e PRODUTO CARTESIANO.
 - Operações desenvolvidas especificamente para bancos de dados relacionais.
 - SELEÇÃO, PROJEÇÃO e JUNÇÃO, entre outras.

Operação SELEÇÃO

- Selecciona todas as tuplas que satisfazem a condição de seleção de uma relação R.
- Símbolo: σ (sigma)
- Notação: $\sigma_{\langle \text{condição de seleção} \rangle}(R)$
- Seleccionar a tupla FUNCIONARIO cujo departamento é 4:
 - $\sigma_{Dnr=4}(\text{FUNCIONARIO})$
- Seleccionar as tuplas cujo salário é maior do que R\$ 30.000,00:
 - $\sigma_{\text{Salario} > 30000}(\text{FUNCIONARIO})$

Exemplo

- Selecionar as tuplas para todos os funcionários que ou trabalham no departamento 4 e ganham mais de R\$ 25.000,00 por ano, ou trabalham no departamento 5 e ganham mais de R\$ 30.000,00:
- $\sigma_{(Dnr=4 \text{ AND } Salario > 25000) \text{ OR } (Dnr=5 \text{ AND } Salario > 30000)}$ (FUNCIONARIO)

SELEÇÃO e SQL

- Em SQL, a condição SELEÇÃO é especificada na cláusula WHERE:

$\sigma_{Dnr=4 \text{ AND Salario}>25.000}$ (FUNCIONARIO)

```
SELECT .  
FROM   FUNCIONARIO  
WHERE  Dnr=4 AND Salario>25.000;
```


Operação PROJEÇÃO

- Produz uma nova relação com apenas alguns dos atributos de R e remove tuplas duplicadas.
- Símbolo: π (pi)
- Notação: $\pi_{\langle \text{lista de atributos} \rangle}(R)$
- Exemplo:
 - $\pi_{\text{Sexo, Salario}}(\text{FUNCIONARIO})$
- Consulta SQL correspondente:

```
SELECT  DISTINCT Sexo, Salario
FROM    FUNCIONARIO
```

Sexo	Salario
M	30.000
M	40.000
F	25.000
F	43.000
M	38.000
M	25.000
M	55.000

SELEÇÃO e PROJEÇÃO

Operação	Descrição	Notação
SELECT	Seleciona todas as tuplas que satisfazem a condição de seleção de uma relação R.	$\sigma_{\langle \text{condição de seleção} \rangle}(R)$
PROJECT	Produz uma nova relação com apenas alguns dos atributos de R e remove tuplas duplicadas.	$\pi_{\langle \text{lista de atributos} \rangle}(R)$

Sequências de operações

- Expressão em linha:

$\pi_{Pnome, Unome, Salario}(\sigma_{Dnr=5}(FUNCIONARIO))$

- Sequência de operações, dando nome a cada relação intermediária:

$$FUNCS_DEPT5 \leftarrow \sigma_{Dnr=5}(FUNCIONARIO)$$
$$RESULTADO \leftarrow \pi_{Pnome, Unome, Salario}(FUNCS_DEP5)$$

Renomeação de atributos

- É possível renomear os atributos nas relações intermediárias e de resultado.

$TEMP \leftarrow \sigma_{Dnr=5}(FUNCIONARIO)$

$R(\text{Primeiro_nome}, \text{Ultimo_nome}, \text{Salario}) \leftarrow$
 $\pi_{Pnome, Unome, Salario}(TEMP)$

Operação RENOMEAR

- Notação $\rho_{S(B_1, B_2, \dots, B_n)}(R)$ ou $\rho_S(R)$ ou $\rho_{(B_1, B_2, \dots, B_n)}(R)$
- Símbolo: ρ (rho)
- S é o nome da nova relação.
- B_1, \dots, B_n são os novos nomes de atributos.
- Em SQL é obtida por apelidos usando AS:

```
SELECT  F.Pnome AS Primeiro_nome, F.Unome  
        AS Ultimo_nome, F.Salario AS Salario  
FROM    FUNCIONARIO AS F  
WHERE   F.Dnr=5,
```

Operações UNIÃO, INTERSECÇÃO, SUBTRAÇÃO

Operação	Descrição	Notação
UNION	Produz uma relação que inclui todas as tuplas em R_1 ou R_2 ou em ambos (R_1 e R_2). R_1 e R_2 precisam ser compatíveis na união.	$R_1 \cup R_2$
INTERSECTION	Produz uma relação que inclui todas as tuplas em ambos (R_1 e R_2). R_1 e R_2 precisam ser compatíveis na união.	$R_1 \cap R_2$
DIFFERENCE	Produz uma relação que inclui todas as tuplas em R_1 que não estão em R_2 . R_1 e R_2 precisam ser compatíveis na união.	$R_1 - R_2$

Operações UNIÃO, INTERSECÇÃO, SUBTRAÇÃO

- Recuperar o CPF de todos os funcionários que ou trabalham no departamento 5 ou supervisionam diretamente um funcionário que trabalha no departamento 5:

$\text{Resultado} \leftarrow \pi_{\text{Cpf}}(\sigma_{\text{Dnr}=5}(\text{FUNCIONARIO})) \cup \pi_{\text{Cpf_supervisor}}(\sigma_{\text{Dnr}=5}(\text{FUNCIONARIO}))$

$\text{FUNCS_DEP5} \leftarrow \sigma_{\text{Dnr}=5}(\text{FUNCIONARIO})$

$\text{RESULTADO1} \leftarrow \pi_{\text{Cpf}}(\text{FUNCS_DEP5})$

$\text{RESULTADO2(Cpf)} \leftarrow \pi_{\text{Cpf_supervisor}}(\text{FUNCS_DEP5})$

$\text{RESULTADO} \leftarrow \text{RESULTADO1} \cup \text{RESULTADO2}$

Operação PRODUTO CARTESIANO ou PRODUTO CRUZADO

- Produz uma relação que tem os atributos de R_1 e R_2 e inclui como tuplas todas as possíveis combinações de tuplas de R_1 e R_2 .
- Notação: $R_1 \times R_2$
- Normalmente não faz sentido usar essa operação de forma isolada.
- Ela é mais útil quando seguida por uma seleção que combina valores de atributos vindos de relações componentes.
- O correspondente em SQL seria a operação CROSS JOIN ou a ausência de condição de junção na cláusula WHERE.

Operação PRODUTO CARTESIANO ou PRODUTO CRUZADO

- Exemplo: Listar os nomes dos dependentes de cada funcionária, exibindo o nome da funcionária e o nome do dependente.

```
FUNC_MULHERES  $\leftarrow \sigma_{\text{Sexo}='F'}(\text{FUNCIONARIO})$ 
```

```
FUNCNOMES  $\leftarrow \pi_{\text{Pnome}, \text{Unome}, \text{Cpf}}(\text{FUNC_MULHERES})$ 
```

```
FUNC_DEPENDENTES  $\leftarrow \text{FUNCNOMES} \times \text{DEPENDENTE}$ 
```

```
DEPENDENTE_PARTIC  $\leftarrow \sigma_{\text{Cpf}=\text{Fcpf}}(\text{FUNC_DEPENDENTES})$ 
```

```
RESULTADO  $\leftarrow \pi_{\text{Pnome}, \text{Unome}, \text{Nome\_dependente}}(\text{DEPENDENTE\_PARTIC})$ 
```

Operação JUNÇÃO

- Produz todas as combinações de tuplas de R_1 e R_2 que satisfazem a condição de junção.
- Notação: $R_1 \bowtie_{\langle \text{condição de junção} \rangle} R_2$
- Exemplo: Obter o nome do gerente de cada departamento.

```
DEP_GER ← DEPARTAMENTO  $\bowtie_{\text{Cpf\_gerente=Cpf}}$  FUNCIONARIO  
RESULTADO ←  $\pi_{\text{Dnome, Unome, Pnome}}$ (DEP_GER)
```

Operação EQUIJUNÇÃO

- Produz todas as combinações de tuplas de R_1 e R_2 que satisfazem uma condição de junção apenas com comparações de igualdade.
- Os exemplos anteriores de junção são exemplos de equijunção.

Operação JUNÇÃO NATURAL

- O mesmo que EQUIJOIN, exceto que atributos de junção de R_2 não são incluídos na relação resultante. Se os atributos de junção tiverem os mesmos nomes, eles nem sequer precisam ser especificados.
- Notação:
 - $R_1 *_{\langle \text{condição de junção} \rangle} R_2$ OU
 - $R_1 *_{(\langle \text{atributos de junção1} \rangle), (\langle \text{atributos de junção2} \rangle)} R_2$ OU
 - $R_1 * R_2$

Operação JUNÇÃO NATURAL

$\text{PROJETO_DEP} \leftarrow \text{PROJETO} \star$
 $\rho_{(\text{Dnome}, \text{Dnum}, \text{Cpf_gerente}, \text{Data_inicio_gerente})}(\text{DEPARTAMENTO})$

$\text{DEP} \leftarrow \rho_{(\text{Dnome}, \text{Dnum}, \text{Cpf_gerente}, \text{Data_inicio_gerente})}$
 (DEPARTAMENTO)
 $\text{PROJETO_DEP} \leftarrow \text{PROJETO} \star \text{DEP}$

$\text{LOCAL_DEP} \leftarrow \text{DEPARTAMENTO} \star \text{LOCALIZA-}$
 CAO_DEP

Operação DIVISÃO

- Produz uma relação $R(X)$ que inclui todas as tuplas $t[X]$ em $R_1(Z)$ que aparecem em R_1 em combinação com toda tupla de $R_2(Y)$, onde $Z = X \cup Y$.
- Notação: $R_1(Z) \div R_2(Z)$

R	
A	B
a1	b1
a2	b1
a3	b1
a4	b1
a1	b2
a3	b2
a2	b3
a3	b3
a4	b3
a1	b4
a2	b4
a3	b4

S
A
a1
a2
a3

T
B
b1
b4

Operação DIVISÃO

- $T \leftarrow R \div S$

R	
A	B
a1	b1
a2	b1
a3	b1
a4	b1
a1	b2
a3	b2
a2	b3
a3	b3
a4	b3
a1	b4
a2	b4
a3	b4

S
A
a1
a2
a3

T
B
b1
b4

Operação DIVISÃO

- Exemplo: recuperar os nomes dos funcionários que trabalham em todos os projetos em que “João Silva” trabalha.

$$\text{SILVA} \leftarrow \sigma_{\text{Pnome}='João' \text{ AND } \text{Unome}='Silva'}(\text{FUNCIONARIO})$$
$$\text{SILVA_PNRS} \leftarrow \pi_{\text{Pnr}}(\text{TRABALHA_EM} \bowtie_{\text{Fcpf}=\text{Cpf}} \text{SILVA})$$
$$\text{CPF_PNRS} \leftarrow \pi_{\text{Fcpf}, \text{Pnr}}(\text{TRABALHA_EM})$$
$$\text{CPFS}(\text{Cpf}) \leftarrow \text{CPF_PNRS} \div \text{SILVA_PNRS}$$
$$\text{RESULTADO} \leftarrow \pi_{\text{Pnome}, \text{Unome}}(\text{CPFS} \star \text{FUNCIONARIO})$$

Operação DIVISÃO

$CPFS(Cpf) \leftarrow CPF_PNRS \div SILVA_PNRS$

$RESULTADO \leftarrow \pi_{Pnome, Unome}(CPFS * FUNCIONARIO)$

CPF_PNRS

Fcpf	Pnr
12345678966	1
12345678966	2
66688444476	3
45345345376	1
45345345376	2
33344555587	2
33344555587	3
33344555587	10
33344555587	20
99988777767	30
99988777767	10
98798798733	10
98798798733	30
98765432168	30
98765432168	20
88866555576	20

SILVA_PNRS

Pnr
1
2

CPFS

Cpf
12345678966
45345345376

Operações da álgebra relacional

Operação	Descrição	Notação
SELECT	Seleciona todas as tuplas que satisfazem a condição de seleção de uma relação R.	$\sigma_{\langle \text{condição de seleção} \rangle}(R)$
PROJECT	Produce uma nova relação com apenas alguns dos atributos de R e remove tuplas duplicadas.	$\pi_{\langle \text{lista de atributos} \rangle}(R)$
THETA JOIN	Produce todas as combinações de tuplas de R_1 e R_2 que satisfazem a condição de junção.	$R_1 \bowtie_{\langle \text{condição de junção} \rangle} R_2$
EQUIJOIN	Produce todas as combinações de tuplas de R_1 e R_2 que satisfazem uma condição de junção apenas com comparações de igualdade.	$R_1 \bowtie_{\langle \text{condição de junção} \rangle} R_2$ OU $R_1 \bowtie_{(\langle \text{atributos de junção1} \rangle), (\langle \text{atributos de junção2} \rangle)} R_2$
NATURAL JOIN	O mesmo que EQUIJOIN, exceto que atributos de junção de R_2 não são incluídos na relação resultante. Se os atributos de junção tiverem os mesmos nomes, eles nem sequer precisam ser especificados.	$R_1 \bowtie_{\langle \text{condição de junção} \rangle} R_2$ OU $R_1 \bowtie_{(\langle \text{atributos de junção1} \rangle), (\langle \text{atributos de junção2} \rangle)} R_2$ OU $R_1 \bowtie R_2$

Operações da álgebra relacional

Operação	Descrição	Notação
UNION	Produz uma relação que inclui todas as tuplas em R_1 ou R_2 ou em ambos (R_1 e R_2). R_1 e R_2 precisam ser compatíveis na união.	$R_1 \cup R_2$
INTERSECTION	Produz uma relação que inclui todas as tuplas em ambos (R_1 e R_2). R_1 e R_2 precisam ser compatíveis na união.	$R_1 \cap R_2$
DIFFERENCE	Produz uma relação que inclui todas as tuplas em R_1 que não estão em R_2 . R_1 e R_2 precisam ser compatíveis na união.	$R_1 - R_2$
CARTESIAN PRODUCT	Produz uma relação que tem os atributos de R_1 e R_2 e inclui como tuplas todas as possíveis combinações de tuplas de R_1 e R_2 .	$R_1 \times R_2$
DIVISION	Produz uma relação $R(X)$ que inclui todas as tuplas $t[X]$ em $R_1(Z)$ que aparecem em R_1 em combinação com toda tupla de $R_2(Y)$, onde $Z = X \cup Y$.	$R_1(Z) \div R_2(Z)$

Operações da álgebra relacional

- As operações SELEÇÃO e PROJEÇÃO são **operações unárias** que ocorrem sobre relações isoladas.
- A operação de JUNÇÃO e outras **operações binárias** operam sobre duas relações.
- A JUNÇÃO combina tuplas relacionadas baseadas em condições de junção.
- Duas relações são **compatíveis na união** se tiverem o mesmo número de atributos e cada par correspondente de atributos tem o mesmo domínio.

Conjunto completo

- O conjunto de operações $\{ \sigma, \pi, \cup, \rho, -, \times \}$ é um conjunto completo, ou seja, qualquer uma das outras operações originais da álgebra relacional pode ser expressa como uma sequência de operações desse conjunto.

$$R \cap S \equiv (R \cup S) - ((R - S) \cup (S - R))$$

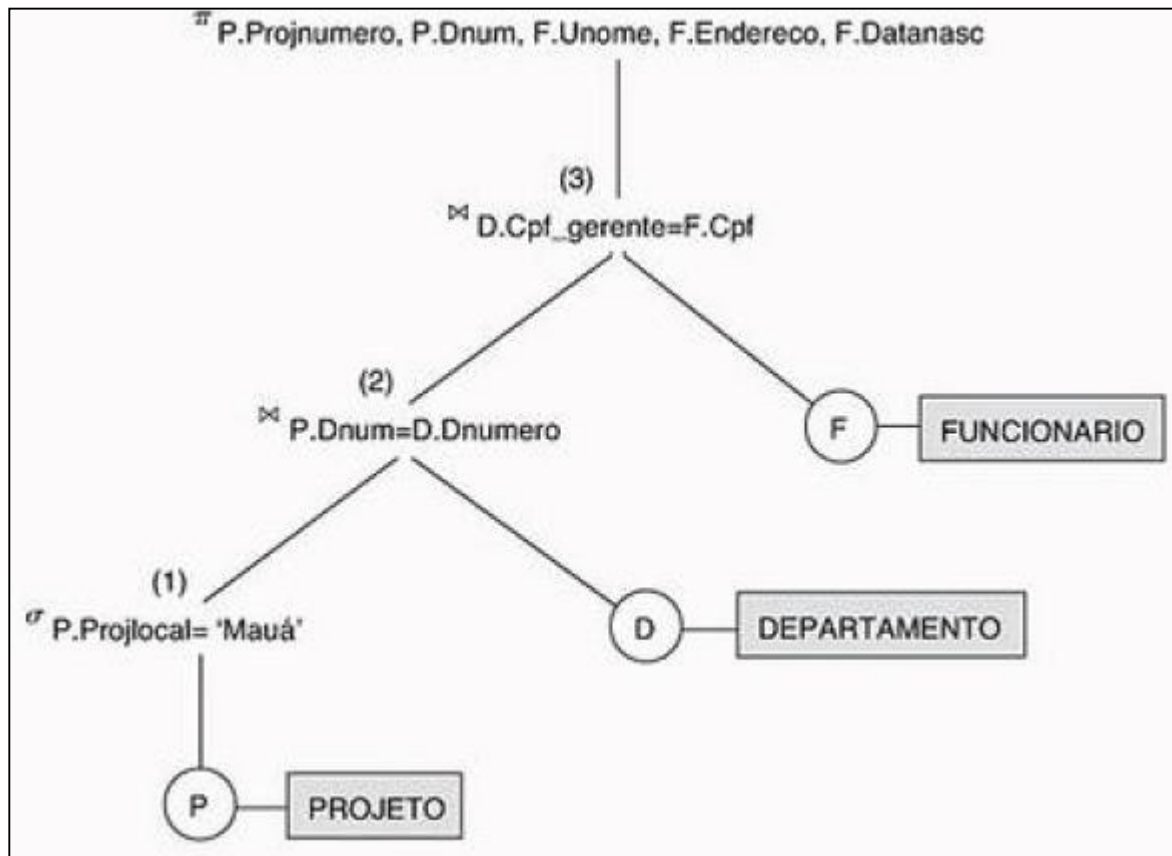
$$R \bowtie_{\langle \text{condição} \rangle} S \equiv \sigma_{\langle \text{condição} \rangle} (R \times S)$$

Árvore de consulta

- A notação chamada **árvore de consulta** ou **árvore de avaliação de consulta** ou **árvore de execução de consulta** costuma ser usada como uma estrutura de dados para representação interna da consulta em um SGBDR.
- Uma árvore de consulta é uma estrutura de dados em árvore que corresponde a uma expressão da álgebra relacional.
- Ela representa as relações de entrada da consulta como nós folha da árvore, e representa as operações da álgebra relacional com nós internos.
- A execução da árvore de consulta consiste em executar uma operação de nó interno sempre que seus operandos estiverem disponíveis, e depois substituir esse nó interno pela relação que resulta da execução da operação.
- A execução termina quando o nó raiz é executado e produz a relação de resultado para a consulta.

Árvore de consulta para expressão de álgebra relacional

$\pi_{\text{Projnumero, Dnum, Unome, Endereco, Datanasc}}(((\sigma_{\text{Projlocal}='Mauá'}(\text{PROJETO})) \bowtie_{\text{Dnum=Dnumero}} (\text{DEPARTAMENTO})) \bowtie_{\text{Cpf_gerente=Cpf}} (\text{FUNCIONARIO}))$



Operações adicionais

- Algumas solicitações comuns em bancos de dados não podem ser realizadas com as operações da álgebra relacional anteriormente descritas.
- Operações adicionais foram definidas para expressar essas solicitações.
- Elas melhoram a expressividade da álgebra relacional original.

Operações adicionais

Operação	Descrição	Notação
GENERALIZED PROJECTION	Estende a operação de projeção, permitindo que as funções dos atributos sejam incluídas na lista de projeção. As funções sobre os atributos podem envolver operações aritméticas e valores constantes.	$\pi_{F_1, F_2, \dots, F_n}(R)$
AGGREGATE FUNCTION	Agrupar tuplas pelo valor de algum de seus atributos e aplicar função de agregação independentemente para cada grupo. <lista funções> é uma lista de pares (<função> <atributo>), onde <função> é uma das funções agregadas: SUM, AVERAGE, MAXIMUM, MINIMUM, COUNT.	$\langle \text{atributos de agrupamento} \rangle \mathcal{F} \langle \text{lista funções} \rangle (R)$
LEFT OUTER JOIN	Mantém cada tupla da relação R_1 (relação da esquerda). Se nenhuma tupla correspondente for encontrada em R_2 , então os atributos de R_2 no resultado da junção são preenchidos com valores NULL.	$R_1 \bowtie_{\langle \text{condição de junção} \rangle} R_2$

Operações adicionais

Operação	Descrição	Notação
RIGHT OUTER JOIN	Mantém cada tupla da relação R_2 (relação da direita). Se nenhuma tupla correspondente for encontrada em R_1 , então os atributos de R_2 no resultado da junção são preenchidos com valores NULL.	$R_1 \bowtie_{\langle \text{condição de junção} \rangle} R_2$
FULL OUTER JOIN	Mantém todas as tuplas nas relações da esquerda e da direita quando nenhuma tupla correspondente for encontrada, preenchendo-as com valores NULL, conforme a necessidade.	$R_1 \bowtie_{\langle \text{condição de junção} \rangle} R_2$
OUTER UNION	Faz a união de tuplas de duas relações que possuem alguns atributos comuns, mas não são compatíveis na união. Todas as tuplas de ambas as relações aparecem nos resultados. Tuplas com a mesma combinação de valores aparecerão somente uma vez no resultado. Os atributos não compatíveis na união também são mantidos no resultado.	

RA: A Relational Algebra Interpreter

<http://www.cs.duke.edu/~junyang/ra/>

```
java -jar ra.jar -h
```

```
java -jar ra.jar -l jdbc:postgresql:empresa -u postgres -P
```

```
java -jar ra.jar -l jdbc:postgresql:empresa -u postgres -p postgres
```

```
java -jar ra.jar empresa.properties
```

```
\help;
```

empresa.properties

```
\list;
```

```
url=jdbc:postgresql:empresa  
user=postgres  
password=postgres
```

```
\quit;
```

```
\project_{pname} funcionario;
```

RA: A Relational Algebra Interpreter

empresa.properties (segundo exemplo)

```
url=jdbc:postgresql://localhost/empresa  
user=postgres  
password=postgres
```

Exemplo de junção com renomeação de atributos com nomes iguais nas tabelas

- Veja esse exemplo, onde as relações 'funcionario' e 'dependente' possuem os atributos 'datanasc' e 'sexo' em comum que foram renomeados na relação 'dependente':
- **Consulta SQL:**

```
select pnome, nome_dependente  
from funcionario join dependente on cpf = fcpf;
```
- Álgebra relacional correspondente usando a notação do interpretador RA:

```
\project_{pname,nome_dependente} (funcionario \join_{cpf =  
fcpf} \rename_{fcpf, nome_dependente, dsexo, ddatanasc,  
parentesco} dependente);
```
- Perceba que o uso dos parênteses pode ser bastante importante. A expressão em álgebra relacional acima não funcionaria sem o uso dos parênteses em destaque.

Referências

- Elsmari, R., Navathe, Shamkant B. “Sistemas de Banco de Dados”. 6ª Edição, Pearson Brasil, 2011. → Capítulo 6
- Silberschatz, A., Korth, H., Sudarshan, S. “Sistema de Banco de Dados”. 5ª Edição, Editora Campus, 2006.
- RA: A Relational Algebra Interpreter
 - <http://www.cs.duke.edu/~junyang/ra/>
- Slides Prof. José Maria (UFC).

Obrigado!
Dúvidas, comentários, sugestões?

Prof. Manoel Ribeiro
Manoel@opencare.com.br



OPENCARE®
Inovação e Tecnologia