Bases de Dados

Universidade do Porto
Faculdade de Engenharia
FEUP

- INTRODUÇÃO
- MODELOS CONCEPTUAIS
  - Diagrama de Classes UML
  - Modelo Entidade-Associação (E-A)
- MODELO RELACIONAL
- LINGUAGEM DE DEFINIÇÃO DE DADOS
- INTERROGAÇÃO DE DADOS
  - **ÁLGEBRA RELACIONAL**
  - Linguagem de Manipulação de Dados (LMD)

Observação: baseado em slides desenvolvidos pelo Prof. Gabriel David da FEUP

Índice

- Linguagens relacionais
- Operações da ágebra relacional
- Leis algébricas
- Linguagem de interrogação
- Extensões à álgebra relacional

Linguagens relacionais

- Notações para expressar perguntas:
  - Algébrica – aplicação de operadores a relações
  - Lógica – fórmula que os tuplos da resposta devem satisfazer
- Limitações da álgebra relacional:
  - independência física dos dados
    - linguagem só com construções relativas ao modelo de dados (operações sobre relações) que não dependam da implementação
  - optimização das perguntas
    - restrição no poder expressivo: sem recursividade (incapaz de computar o fecho transitivo)
  - relações finitas
    - Noção de complementar é proibida

Operações da álgebra relacional

Reunião

1 - Reunião $R \cup S$

- é o conjunto dos tuplos que estão em R, em S, ou em ambas
  - R e S da mesma aridade (mesma aridade = mesmo nº de atributos)
  - Domínios dos atributos de R e S devem ser compatíveis:
    - Domínios do 1º atributo de R e S devem ser compatíveis
    - Domínios do 2º atributo de R e S devem ser compatíveis
    - etc.

R

A	B	C
a	b	c
d	a	f
c	b	d

S

D	E	F
b	g	a
d	a	f

$R \cup S$

A	B	C
a	b	c
d	a	f
c	b	d
b	g	a

Operações da álgebra relacional

Intersecção

2 - Intersecção $R \cap S$

- contém os tuplos que pertencem a R e a S simultaneamente
  - R e S da mesma aridade
  - R e S com domínios compatíveis
- $R \cap S = R - (R-S)$

R

A	B	C
a	b	c
d	a	f
c	b	d

S

D	E	F
b	g	a
d	a	f

$R \cap S$

A	B	C
d	a	f

Operações da álgebra relacional

Diferença

3 - Diferença R – S

- tuplos de R que não estão em S
  - R e S da mesma aridade
  - R e S com domínios compatíveis

R

A	B	C
a	b	c
d	a	f
c	b	d

S

D	E	F
b	g	a
d	a	f

R - S

A	B	C
a	b	c
c	b	d

Operações da álgebra relacional

Produto cartesiano

4 - Produto cartesiano R×S

- aridades de R e S são k_1 e k_2 ⇒ aridade de R×S é $k_1 + k_2$
  - contém todos os tuplos tais que os primeiros k_1 componentes formam um tuplo de R e os restantes k_2 componentes formam um tuplo em S

R

A	B	C
a	b	c
d	a	f
c	b	d

S

D	E	F
b	g	a
d	a	f

R×S

A	B	C	D	E	F
a	b	c	b	g	a
a	b	c	d	a	f
d	a	f	b	g	a
d	a	f	d	a	f
c	b	d	b	g	a
c	b	d	d	a	f

Operações da álgebra relacional

Projecção

5 - Projecção $\Pi_{i_1, i_2, \ldots, i_m}(R)$

- para cada tuplo em R existe um tuplo na projecção com os componentes (e pela ordem) indicados pelos i_j
  - se aridade de R for k então os $i_j \in 1, \ldots, k$ são distintos; aridade da projecção é m
  - números de componentes podem ser substituídos por atributos, se existirem: $\Pi_{1,3}(R) = \Pi_{A,C}(R)$

R

A	B	C
a	b	c
d	a	f
c	b	d

$\Pi_{1,3}$ (R)

A	C
a	c
d	f
c	d

Projecção escolhe colunas da tabela

Operações da álgebra relacional

Selecção

6 - Selecção σ_F (R)

- contém os tuplos de R que satisfazem F
  - a fórmula F pode envolver
    - ⇒ operandos constantes ou número de componente ($i)
    - ⇒ operadores aritméticos de comparação (<, =, >, ≤, ≠, ≥)
    - ⇒ operadores lógicos (∧, ∨, ¬) (e, ou, não)
  - Números de componentes podem ser substituídos por atributos, se existirem: $\sigma_{\$2=b}(R) = \sigma_{B=b}(R)$

R

A	B	C
a	b	c
d	a	f
c	b	d

$\sigma_{\$2=b}$(R)

A	B	C
a	b	c
c	b	d

Selecção escolhe linhas da tabela

Operações da álgebra relacional

Quociente

7 - Quociente R / S

- aridade de R é r e de S é s, r>s, S≠∅
  - contém os (r-s)-tuplos $(a_1, \ldots, a_{r-s})$ tais que, para **todos** os s-tuplos $(a_{r-s+1}, \ldots, a_r)$ em S, o tuplo $(a_1, \ldots, a_r)$ está em R

R

A	B	C	D
a	b	c	d
a	b	e	f
⇒ b	c	e	f
e	d	c	d
e	d	e	f
⇒ a	b	d	e

/

S

E	F
c	d
e	f

=

R/S

A	B
a	b
e	d

7	3
1	2

R/S

A	B
a	b
e	d

×

S

E	F
c	d
e	f

=

(R/S) × S

A	B	E	F
a	b	c	d
a	b	e	f
e	d	c	d
e	d	e	f

$2 \times 3 + 1 = 7$

Operações da álgebra relacional

Quociente: explicação alternativa

- Forma de proceder à divisão
  - reordenar as colunas de forma a que as últimas correspondam ao divisor
  - ordenar a tabela pelas primeiras colunas
  - cada subtuplo das primeiras colunas pertence ao resultado se o conjunto de subtuplos das últimas colunas que lhe corresponde contiver o divisor

R

A	B	C	D
a	b	c	d
		e	f
		d	e
b	c	e	f
e	d	c	d
		e	f

/

S

E	F
c	d
e	f

=

R/S

A	B
a	b
e	d

Operações da álgebra relacional

Quociente: expressão

R

A	B	C	D
a	b	c	d
a	b	e	f
b	c	e	f
e	d	c	d
e	d	e	f
a	b	d	e

S

E	F
c	d
e	f

T

A	B
a	b
b	c
e	d

T×S

A	B	E	F
a	b	c	d
a	b	e	f
b	c	c	d
b	c	e	f
e	d	c	d
e	d	e	f

- $T = \Pi_{1, \ldots, r-s}(R)$ = universo dos tuplos possíveis no resultado
- $W = (T \times S) - R$ = todas as linhas T combinadas com S mas que não estão em R, i.e., em que a condição falha
- $V = \Pi_{1, \ldots, r-s}(W)$ = tuplos que não interessam
- $R / S = T - V$ = tuplos que interessam
  - reunindo numa só expressão algébrica

$$R / S = \Pi_{1, \ldots, r-s}(R) - \Pi_{1, \ldots, r-s}[(\Pi_{1, \ldots, r-s}(R) \times S) - R]$$

W

A	B	E	F
b	c	c	d

V

A	B
b	c

R/S

A	B
a	b
e	d

Operações da álgebra relacional

Exercícios

Quais as frases verdadeiras?

a – $(R-S) \cup S = R$

b – $(R-S) \cup S \supseteq R$

c – $(R-S) \cup (R \cap S) = R$

d – $(R-S) \cup (S-R) = (R \cup S) - (R \cap S)$

e – $(R/S) \times S = R$

f – $(R/S) \times S \subseteq R$

- **Respostas**
  - **Erradas - a, e**
  - **Correctas - b, c, d, f**

Operações da álgebra relacional

Junção

8 - θ-Junção $R \underset{i\theta j}{\bowtie} S$

- se a aridade de R for r e a de S s, a aridade da θ-junção é r+s
- contém os tuplos do produto cartesiano de R por S tais que o componente i está na relação θ com o componente r+j (i.e., o correspondente ao j em S).

- Expressão da θ-junção
  - $R \underset{i\theta j}{\bowtie} S = \sigma_{\$i\ \theta\ \$(r+j)} (R \times S)$
    - se θ for =, a operação designa-se equijunção
    - (7, 8, 9) é um tuplo pendente de R pois não aparece na θ-junção

R

A	B	C
1	2	3
4	5	6
7	8	9

S

D	E
3	1
6	2

$R \underset{2<1}{\bowtie} S = R \underset{B<D}{\bowtie} S$

A	B	C	D	E
1	2	3	3	1
1	2	3	6	2
4	5	6	6	2

Operações da álgebra relacional

Junção natural

9 - Junção natural $R \bowtie S$

- só é aplicável se os componentes dos tuplos em R e S forem designados por atributos.
- a operação implícita na junção natural é a igualdade dos atributos com o mesmo nome.
- cada par de atributos iguais dá origem a um único atributo, com o mesmo nome, no resultado

- expressão:

$$R \bowtie S = \Pi_{i_1, \ldots, i_m} (\sigma_{R.A_1 = S.A_1 \wedge \ldots \wedge R.A_k = S.A_k} (R \times S))$$

  - k é o número de atributos comuns a R (aridade r) e S (aridade s) e m= r+s-k

R

A	B	C
a	b	c
d	b	c
b	b	f
c	a	d

S

B	C	D
b	c	d
b	c	e
a	d	b

R ⋈ S

A	B	C	D
a	b	c	d
a	b	c	e
d	b	c	d
d	b	c	e
c	a	d	b

Operações da álgebra relacional

Junção externa

- tuplos pendentes, isto é desemparelhados, quer em R quer em S, desaparecem na θ-junção e na junção natural
- a **junção externa (**θ- ou natural) inclui os tuplos pendentes de R ou S completados a nulos
  - (7, 8, 9) é um tuplo pendente de R pois não aparece na θ-junção
  - (b, b, f) idem, na junção natural

R

A	B	C
1	2	3
4	5	6
7	8	9

S

D	E
3	1
6	2

R $\overset{+}{\bowtie}$ S

$_{B<D}$

A	B	C	D	E
1	2	3	3	1
1	2	3	6	2
4	5	6	6	2
7	8	9	⊥	⊥

R

A	B	C
a	b	c
d	b	c
b	b	f
c	a	d

S

B	C	D
b	c	d
b	c	e
a	d	b

R $\overset{+}{\bowtie}$ S

A	B	C	D
a	b	c	d
a	b	c	e
d	b	c	d
d	b	c	e
c	a	d	b
b	b	f	⊥

Operações da álgebra relacional

Semi-junção

10 - Semi-junção $R \ltimes S$

- projecção nos atributos de R da junção natural de R e S
- $R \ltimes S = \Pi_R(R \bowtie S)$
  - R em Π_R representa os atributos de R (o seu esquema); em $R \bowtie S$ R representa a relação (a instância)
- outra expressão: $R \ltimes S = R \bowtie \Pi_{R \cap S}(S)$
- dá os tuplos de R que têm par em S

R

A	B	C
a	b	c
d	b	c
b	b	f
c	a	d

S

B	C	D
b	c	d
b	c	e
a	d	b

$R \ltimes S$

A	B	C
a	b	c
d	b	c
c	a	d

Operações da álgebra relacional

Relações com atributos

- na junção natural e na semi-junção os atributos são importantes; para os tornar explícitos escreve-se $R(A_1, ..., A_n)$
- é possível renomear colunas e fazer junções naturais como:

S = S(B,C,D)

B	C	D
b	c	d
b	c	e
a	d	b

S(E,F,G) ⋈ S(G,H,I)

E	F	G	H	I
a	d	b	c	d
a	d	b	c	e

- uma junção natural entre duas relações sem atributos comuns redunda num produto cartesiano porque, após este, não há nenhuma selecção a fazer (equivalente a fazer uma selecção com a condição *True*)
- **R(A,B,C) ⋈ S(G,H,I) = R × S**

Leis algébricas

- Reunião
  - associativa: $R \cup (S \cup T) = (R \cup S) \cup T$
  - comutativa: $R \cup S = S \cup R$
- Produto cartesiano
  - associativo: $R \times (S \times T) = (R \times S) \times T$
  - não comutativo: $R \times S \neq S \times R$
- Junção natural
  - associativa e comutativa (independência da ordem das colunas devida aos atributos): $R \bowtie S = S \bowtie R$
  - Por isso $\bowtie$ generaliza facilmente: $R = R_1 \bowtie \ldots \bowtie R_n$
    - R contém os tuplos μ tais que, para $1 \leq i \leq n$, μ restringido aos atributos de R_i é um tuplo de R_i
- θ - junção
  - não é comutativa mas é associativa (no caso de os índices serem válidos)

$$R \underset{i\,\theta_1\,j}{\bowtie} (S \underset{k\,\theta_2\,l}{\bowtie} T) = (R \underset{i\,\theta_1\,j}{\bowtie} S) \underset{(r+k)\,\theta_2\,l}{\bowtie} T$$

Esquema relacional de Cursos

1 Curso(codcurso, designacur)

2 Disciplina(coddis, sigla, designadis)

3 Turma(codcurso➔Curso, ano, letra)

4 Professor(bip, nome, morada, telefone, habilitação, grupo)

5 Aluno(bia, nome, morada, telefone, data_nasc, [codcurso, ano, letra]➔Turma)

6 Plano(codcurso➔Curso, coddis➔Disciplina)

7 Inscrito(coddis➔Disciplina, bia➔Aluno, resultado)

8 Lecciona(bip ➔Professor, coddis➔Disciplina, [codcurso, ano, letra]➔Turma)

Só as tabelas 4 – 7 são necessárias para os exercícios seguintes.

Linguagem de Interrogação

- Álgebra Relacional pode ser usada como linguagem de interrogação à BD
- *P1 - Relativamente à BD "Cursos" (ver atrás), quais os nomes dos professores do 12º grupo?*

$$\Pi_{nome}\ [\ \sigma_{grupo\ =\ '12'}\ (\ Professor\ )\ ]$$

- *P2 - Quais os nomes e datas de nascimento dos alunos do curso 'CG1' nascidos antes de 1983?*

$$\Pi_{nome,\ data_nasc}\ [\ \sigma_{codcurso\ =\ 'CG1'\ \wedge\ data_nasc\ <\ 1983\text{-}01\text{-}01}\ (\ Aluno\ )\ ]$$

Linguagem de interrogação

Perguntas com junção

- *P3 - Nomes dos alunos inscritos à disciplina 327?*
  - nenhuma relação contém nomes de alunos e códigos de disciplina
  - mas a junção Aluno ⋈ Inscrito = R contém:
  - R (bia, nome, morada, telefone, data_nasc, codcurso, ano, letra, coddis, resultado)

$$(\text{ i }) - \Pi_{nome} \; [\, \sigma_{coddis = 327} \, (\text{Aluno} \bowtie \text{Inscrito}) \,]$$

$$(\text{ ii }) - \Pi_{nome} \; [\, \text{Aluno} \bowtie \sigma_{coddis = 327} \, (\text{Inscrito}) \,]$$

  - esta maneira de ligar informações no modelo de dados dá muita liberdade para exprimir perguntas arbitrárias mas exige uma fase de optimização para executar (ii) mesmo que a pergunta seja (i)

Núcleo da álgebra relacional : Π, σ, ⋈

Extensões à álgebra relacional

Eliminação de duplicados

- R
- R' = δ(R)
  - elimina os tuplos repetidos de R

- *P4 – Quais os alunos inscritos a alguma disciplina?*
  - Π_{nome} [Aluno $\bowtie$ δ [Π_{bia} (Inscrito)]]
  - A relação Inscrito contém os bi dos alunos tantas vezes quantas as cadeiras a que o aluno está inscrito
  - Assim, é necessário eliminar os bi repetidos para que os nomes dos alunos não apareçam repetidos

Extensões à álgebra relacional

Renomeações de atributos

- R (A, B, C)
- R'(X,Y, Z) = $\Pi_{X = A, Y = B, Z = C}$ [R(A, B, C)]
  - onde não houver ambiguidades, a simples menção dos atributos, em conjunto com o nome da relação, faz a renomeação

    OU
    - R' (X, Y, Z) = R (A, B, C)

    OU
    - R' = $\Pi_{X = A, Y = B, Z = C}$ (R)
  - expressões aritméticas (+, -, *, /)
- este mecanismo serve para dar nomes a expressões
- S = $\Pi_{W = A * B - C, U = C/B,A}$ (R)

  OU
- S (W, U, A)= $\Pi_{A * B - C, C/B, A}$ (R)

Linguagem de interrogação

Expressões artméticas

- P5 – *Obtenha a relação das inscrições com as classificações inflaccionadas de 20%.*
  - $\Pi_{\text{coddis, bia, resultado, novo = resultado*1.2}}$ (Inscrito)

  Outro exemplo:
  - $\Pi_{\text{coddis, bia, resultado, novo = resultado + (20-resultado)/10}}$ (Inscrito)

- nos parâmetros da projecção:
  - no membro direito só podem ser usados nomes de atributos do argumento de Π; no esquerdo só pode estar um atributo (novo…)

Extensões à álgebra relacional

Agregações

- Operadores de agregação
  - CNT (contagem), SUM (adição), AVG (média), MAX (máximo), MIN (mínimo)
- $S = \Pi_{V = CNT (B)} (R)$
  - S(V) tem um único valor, o número de tuplos de R com valor não nulo no atributo B (CNT (*) conta todas as linhas) → toda a relação agregada
- $T = \Pi_{A, M = MAX (B)} (R)$
  - T (A, M) tem tantos pares quantos os valores diferentes de A, sendo indicado para cada A o respectivo valor máximo de B (não nulo ...);
  - é feita uma partição segundo os atributos de projecção sem operadores de agregação e cada classe é agregada num só tuplo;
  - é possível misturar agregações e aritmética

Linguagem de interrogação

Perguntas com agregação

- P6 - *Qual o número de inscrições, nota média das inscrições, soma de todas as notas e número de resultados não nulos ?*
  - R(NI, M, T, NR) = $\Pi_{CNT(*),\ AVG(\ resultado),\ SUM(\ resultado\ ),\ CNT(\ resultado)}$ (Inscrito)
  - pode ser M ≠ T/NI se houver inscrições ainda sem resultado (valor nulo); tem que ser M = T/ NR

Inscrito

coddis	bia	resultado
PA	97	14
PA	38	12
ITI	97	17
ITI	25	14
H	97	10
H	25	

R

NI	M	T	NR
6	13.4	67	5

Linguagem de interrogação

Agregação com partição

- P7 - *Quais as notas mínima, média e máxima de cada disciplina (independentemente do aluno)?*
  - $R = \Pi_{coddis,\ MI = MIN(\ resultado\ ),\ ME = AVG(resultado\ ),\ MA = MAX(resultado)}$ (Inscrito)

Inscrito

coddis	bia	resultado
PA	97	14
PA	38	12
ITI	97	17
ITI	25	14
H	97	10
H	25	

R

coddis	MI	ME	MA
PA	12	13	14
ITI	14	15.5	17
H	10	10	10

Linguagem de interrogação

Quantificação existencial

- P8 - *Obtenha os códigos dos alunos com inscrição a* ***pelo menos uma*** *das disciplinas do curso LEEC.*
  - $\delta(\Pi_{BIA}(INSCRITO \bowtie \sigma_{CODCURSO= 'LEEC'}(PLANO)))$

Linguagem de interrogação

Quantificação universal

- P9 - *Obtenha o código dos alunos com inscrição a **todas** as disciplinas do curso LEEC.*
  - $A = \Pi_{BIA}$ (ALUNO)
    - conjunto dos alunos
  - $D = \Pi_{CODDIS}$ [$\sigma_{CODCURSO= \text{'LEEC'}}$ (PLANO)]
    - conjunto das disciplinas do curso LEEC
  - $A \times D$
    - conjunto de todos os pares (aluno, disciplina da LEEC)
  - $NI = A \times D - \Pi_{BIA,\ CODDIS}$ (INSCRITO)
    - pares (aluno, disciplina da LEEC) tais que o aluno não tem inscrição à disciplina
  - $R = A - \Pi_{BIA}$ (NI)
    - resultado (notar a dupla subtracção)
- $R = \Pi_{BIA,\ CODDIS}$(INSCRITO) / D

Query language

More aggregation

- P10 - Qual o nome e média actual do aluno com melhor média do curso 'LEEC'?
- Só a média actual dos alunos
  - $MA = \Pi_{bia,\ media\ =\ AVG(resultado\ )}\ (Inscrito \bowtie \sigma_{CODCURSO=\ 'LEEC'}\ (PLANO))$
- Só a média máxima
  - $MaxM = \Pi_{media\ =\ MAX(media)}\ (MA)$
- Aluno cuja média é igual à média máxima
  - $R= \Pi_{nome,\ MA.media}\ (Aluno \bowtie MA \bowtie MaxM)$