Lenguajes de Consulta

Fuentes http://www-db.stanford.edu/~ullman/fcdb.html
Fundamentals of Database Systems, Elmasri y Navathe
Database System Concepts,Silberschatz, Korth, Sudarshan

Lenguajes

- Permiten recuperación de datos de la bd
- Varios lenguajes para MR
  - Lenguajes formales
  - Basados en lógica y algebra
  - Facilita optimización
- Lenguajes
  - Acceso eficiente a bd
  - No cálculos complejos
  - Declarativos vs. imperativos

Lenguajes de Consulta Relacional

- Base implementación SQL
- Algebra Relacional
  - Imperativa
  - Útil representación planes de consulta
- Cálculo Relacional
  - Declarativa
  - Enfasis QUÉ y no en CÓMO

Lenguajes de Consulta Relacional

- Consultas aplicables a instancias de relación
- Producen instancias relación
- Esquema relaciones entrada fijo
- Esquemas basados en posición vs. basados en nombres de atributos

ALGEBRA RELACIONAL

Algebra Relacional

- Algebra
  - Operandos: relaciones o variables de relación
  - Operadores: aplicados sobre relaciones
  - Operaciones se pueden componer

Algebra usada como lenguaje de consulta para las relaciones

Algebra Relacional: *Core*

- *Selección* (σ) Selecciona un subconjunto de *filas o tuplas* de una relación (horizontal)
- *Proyección* (π) Conserva algunas *columnas* de la relación (vertical)
- *Producto-cartesiano* ($\times$) Permite combinar dos relaciones
- *Diferencia-conjuntos* (—) tuplas en r1 no en r2.
- *Union* ($\cup$) Tuplas en r1 o en r2
- *Renombramiento* (ρ)

Algebra Relacional

Operadores Unarios

- *Selección* (σ)
- *Proyección* (π)
- *Renombramiento* (ρ)

Operaciones de conjunto

- *Producto-cartesiano* ($\times$)
- *Diferencia-conjuntos* (—)
- *Union* ($\cup$)
- *Intersección* ($\cap$)

Algebra Relacional

Operadores Relacionales Binarios

- *Join* ($\bowtie$)
- *División*

Otras Operaciones

- *OUTER JOIN, OUTER UNION*
- *Funciones Agregadas (SUM, COUNT, AVG, MIN, MAX)*

Selección σ

$R1 := \sigma_C (R2)$

- *C* es una condición(predicado) aplicable a attributes de R2
- R1 contiene todas las tuplas de R2 que satisfacen *C*
- Esquemas R1 y R2 iguales

Selección σ

$\sigma_c(r) = \{t \mid t \in r \text{ y } c(t)\}$

c es una fórmula del cálculo proposicional formada por términos y operadores $\wedge$ (and), $\vee$ (or), $\neg$ (not)

Cada término puede ser:

<atributo> *op* <atributo> o <constante>

donde *op* puede ser: $=, \neq, >, \geq, <, \leq$

Selección

Relación r

A	B	C	D
α	α	1	7
α	β	5	7
β	β	12	3
β	β	23	10

$\sigma_{A=B \wedge D > 5}(r)$

A	B	C	D
α	α	1	7
β	β	23	10

Ejemplo

Relación Matrícula

estudiante	curso	calif
Luis	BD	3.5
José	MDI	4.7
Maria	BD	3.5
Luis	MDI	4.0

Tabulado-Luis= $\sigma_{\text{estudiante=“Luis”}}(\text{Matricula})$

estudiante	curso	calif
Luis	BD	3.5
Luis	MDI	4.0

Selección σ

Por conmutatividad una secuencia (cascada) de operaciones SELECT se puede aplicar en cualquier orden

$$\sigma_{<cond1>}(\sigma_{<cond2>}(\sigma_{<cond3>}(R)) = \sigma_{<cond2>}(\sigma_{<cond3>}(\sigma_{<cond1>}(R)))$$

Selección σ

Una secuencia de operaciones SELECT se puede reemplazar por la conjunción de condiciones

$$\sigma_{<cond1>}(\sigma_{<cond2>}(\sigma_{<cond3>}(R)) = \sigma_{<cond1>AND< cond2>AND < cond3>}(R)))$$

Selección σ

El número de tuplas en el resultado de un SELECT es menor que (o igual a) el número de tuplas de la relación de entrada R

Proyección π

$R1 := \pi_L (R2)$

- L lista de atributos $(A_1, A_2, \ldots, A_n)$ del esquema R2
- R1 está formada por tuplas de R2, con los atributos en el orden de la lista L
- tuplas duplicadas se eliminan si hay alguna

(conjuntos no tienen elementos duplicados)

Proyección

Relación *r*

A	B	C
α	10	1
α	20	1
β	30	1
β	40	2

$$\Pi_{A,C}(r)$$

A	C
α	1
α	1
β	1
β	2

=

A	C
α	1
β	1
β	2

Proyección: Propiedades

El total de tuplas resultado de $\pi_{<list>}(R)$ es siempre menor o igual al total de tuplas de R

- Si la lista de atributos incluye una llave de R, el número de tuplas en el resultado de la proyección es igual al total de tuplas en R

- PROYECCIÓN no es commutativa

  $\pi_{<list1>}(\pi_{<list2>}(R)) = \pi_{<list1>}(R)$ sobre <list2> conteniendo los atributos en <list1>

Ejemplo

Relación Matricula

estudiante	curso	calif
Luis	BD	3.5
José	MDI	4.7
Maria	BD	3.5
Luis	MDI	4.0

$$\text{Estudiantes} = \pi_{\text{estudiante}} (\text{Matricula})$$

estudiante
Luis
José
María

Proyección Extendida

La lista *L*, en π_L, puede contener expresiones aplicables a los atributos:

- Aritméticas como $A{+}B \rightarrow C$
- Ocurrencias duplicadas sobre el mismo atributo

Ejemplo

R =

A	B
1	2
3	4

$\Pi_{A+B \rightarrow C, A}$ (R) =

C	A
3	1
7	3

Producto Cartesiano

R3 := R1 ✕ R2

- Cada tupla t1 de R1 se combina con cada tupla t2 de R2
- La concatenación t1t2 es una tupla de R3
- El esquema de R3 se forma con los atributos de R1 y de R2 en el orden en que aparecen
- Si el mismo atributo A aparece en ambas relaciones R1 y R2, usar R1.*A* y R2.*A*

Ejemplo R3 = R1 X R2

R1

A	B
1	2
3	4

R2

B	C
5	6
7	8
9	10

R3

A	R1.B	R2.B	C
1	2	5	6
1	2	7	8
1	2	9	10
3	4	5	6
3	4	7	8
3	4	9	10

Producto Cartesiano

Relaciones

A	B
α	1
β	2

r

C	D	E
α	10	a
β	10	a
β	20	b
γ	10	b

s

r x *s*

A	B	C	D	E
α	1	α	10	a
α	1	β	10	a
α	1	β	20	b
α	1	γ	10	b
β	2	α	10	a
β	2	β	10	a
β	2	β	20	b
β	2	γ	10	b

Producto Cartesiano

$$r \times s = \{t\,q \mid t \in r \wedge q \in s\}$$

- Se asume que los atributos en r(R) y s(S) son disjuntos ($R \cap S = \emptyset$)
- Si ($R \cap S \neq \emptyset$) necesidad de usar renombramiento

Renombramiento ρ

- Operador ρ produce un nuevo esquema de relación
- $R1 := \rho_{R1(A1,\ldots,An)}(R2)$ construye relación R1 con atributos $A_1,\ldots, A_n$ y las mismas tuplas de R2
- Notación simplificada

  $R1(A1,\ldots,An) := R2$

Renombramiento

- Permite nombrar atributos o relaciones
- Formas de renombramiento
  - $\rho_{S\ (B1,\ B2,\ \ldots,\ Bn\ )}(R)$ cambia
    - Por S el nombre de la relación
    - por B1, B1, …..Bn los nombres de los aributos o columnas

Renombramiento

- Formas de renombramiento
  - $\rho_S(R)$ cambia
    - Cambia únicamente por S el nombre de la relación
  - $\rho_{(B1,\ B2,\ \ldots,\ Bn\ )}(R)$
    - Cambia unicamente los nombres de las columnas o atributos por B1, B1, …..Bn

Unión $\cup$

- Operador binario $r \cup s$

$$r \cup s = \{t \mid t \in r \vee t \in s\}$$

  - *r, s* deben tener *la misma* **aridad**
  - Devuelve tuplas que están en r, s ó ambas
  - Tuplas duplicadas se eliminan
  - Dominios de atributos **compatibles** (mismo tipo de valores columnas correspondientes)
  - Relaciones UNION compatibles

Relaciones Unión Compatibles

Dos relaciones r y r’con esquemas $R(A_i:D_i)$, $R'(A_i:D_i)$ y cardinalidades m y m’, respectivamente, son compatibles cuando ambas están definidas sobre el mismo dominio.

$$\forall A_i \; \exists \; A'_j \mid dom(A_i)=dom(A'_j)$$

$$\forall \; A'_j \; \exists \; A_j \mid dom(A'_j)=dom(A_j)$$

Unión $\cup$

Relaciones

A	B
α	1
α	2
β	1

A	B
α	2
β	3

s

$r \cup s$

A	B
α	1
α	2
β	1
β	3

Observación

Cuando los esquemas R y R’no se corresponden (el nombre de los atributos son diferentes en las relaciones, o están en distinto orden) se sugiere renombrar los atributos en la relación resultante

Ejemplo

Calcular la unión de las siguientes relaciones

M

Cedula-M	Nombre-M
3232330	María Ruíz
3455775	Rosa Toro
3216789	Julia Muñoz

H

Nombre-H	Cedula-H
6756989	Luis Lara
6453446	Jose Cruz

Diferencia –

Relaciones

A	B
α	1
α	2
β	1

r

A	B
α	2
β	3

s

$r - s$

A	B
α	1
β	1

Diferencia

$$r - s = \{t \mid t \in r \wedge t \notin s\}$$

- Relaciones UNION compatibles
  - *r* y *s* deben tener la misma aridad
  - Dominios de atributos de *r* y *s* deben ser compatibles

Composición de Operaciones

- Expresiones con múltiples operadores
- $\sigma_{A=C}(r \times s)$

r x s

A	B	C	D	E
α	1	α	10	a
α	1	β	10	a
α	1	β	20	b
α	1	γ	10	b
β	2	α	10	a
β	2	β	10	a
β	2	β	20	b
β	2	γ	10	b

$\sigma_{A=C}(r \times s)$

A	B	C	D	E
α	1	α	10	a
β	2	β	10	a
β	2	β	20	b

Definición Formal

Sean E_1 y E_2 dos expresiones relacionales

Las siguientes son expresiones del álgebra

- $E_1 \cup E_2$
- $E_1 - E_2$
- $E_1 \times E_2$
- $\sigma_p (E_1)$, p es un predicado sobre los atributos en E_1
- $\prod_S(E_1)$, S lista de atributos en E_1
- $\rho_x (E_1)$, x es un nombre nuevo para el resultado de E_1