

@author: aar.velasquez@gmail.com

@date: 2020/10/18

Capítulo 8: El álgebra relacional y Cálculo relacional

El conjunto básico de operaciones para el modelo relacional formal es el álgebra relacional.

Estas operaciones permiten a un usuario para especificar solicitudes de recuperación básicas como expresiones de álgebra relacional. El resultado de una consulta de recuperación es una nueva relación. Las operaciones de álgebra producen así nuevas relaciones, que se pueden manipular aún más utilizando operaciones del mismo álgebra.

La secuencia de operaciones de álgebra relacional forma una expresión de álgebra relacional, cuyo resultado también será una relación que representa el resultado de una consulta de base de datos (o solicitud de recuperación).

Importancia del álgebra relacional:

- proporciona una base formal para las operaciones del modelo relacional.
- se utiliza como base para implementar y optimizar consultas en el módulo de optimización y procesamiento de consultas que son parte integral de las relaciones del lenguaje de consulta para RDBMS.

Mientras que el álgebra define un conjunto de operaciones para el modelo relacional, el cálculo relacional proporciona un lenguaje declarativo de alto nivel para especificar relaciones de consultas.

El cálculo relacional es importante porque tiene una base firme en lógica matemática y porque el lenguaje de consulta estándar (SQL) para RDBMS tiene algunos de sus fundamentos en una variación del cálculo relacional conocida como tupla cálculo relacional.

El álgebra relacional a menudo se considera una parte integral de los datos relacionales. modelo. Sus operaciones se pueden dividir en dos grupos.

- conjunto de operaciones de la teoría matemática de conjuntos.
- operaciones desarrolladas específicamente para bases de datos relacionales, que incluyen SELECT, PROJECT, y JOIN, entre otros.

8.1 Operaciones relacionales unarias:

Seleccionar y proyectar

SELECT:

se usa para elegir un subconjunto de las tuplas de una relación que satisface una condición de selección

En general, la operación SELECT se denota por $\sigma < \text{condición de selección} > (R)$

PROJECT:

Si pensamos en una relación como una tabla, la operación SELECT elige algunas de las filas de la tabla mientras descarta otras filas. La operación PROYECTO, por otro lado, selecciona ciertas columnas de la tabla y descarta las otras columnas. Si nosotros estamos interesados solo en ciertos atributos de una relación, usamos la operación PROYECTO para proyectar la relación sobre estos atributos solamente.

$\pi \text{ Lname, Fname, Salario } (\text{EMPLEADO})$

La operación PROYECTO elimina cualquier tupla duplicada, por lo que El resultado de la operación PROYECTO es un conjunto de tuplas distintas y, por lo tanto, una relación.

Secuencias de operaciones y operación RENAME

RENAME:

El general

La operación RENAME cuando se aplica a una relación R de grado n se denota por cualquiera de los siguientes tres formas:

$\rho S (B_1, B_2, \dots, B_n) (R)$ o $\rho S (R)$ o $\rho (B_1, B_2, \dots, B_n) (R)$

8.2 Operaciones de álgebra relacional de la teoría de conjuntos

Las operaciones UNION, INTERSECTION, y MINUS

UNION:

Por ejemplo, para recuperar los números de seguro social de todos.

empleados que trabajan en el departamento 5 o supervisan directamente a un empleado que trabaja en el departamento 5, podemos utilizar la operación UNION

```
DEP5_EMPS ← σ Dno = 5 ( EMPLOYEE )
RESULT1 ← π Ssn ( DEP5_EMPS )
RESULT2 ( Ssn ) ← π Super_ssn ( DEP5_EMPS )
RESULT ← RESULT1 U RESULT2
```

Se utilizan varias operaciones de la teoría de conjuntos para fusionar los elementos de dos conjuntos en formas, incluyendo UNION , INTERSECTION y SET DIFFERENCE (también llamado MINUS o EXEPT)

Podemos definir las tres operaciones UNION , INTERSECTION y SET DIFFERENCE en dos relaciones R y S compatibles con la unión como sigue:

UNION:

El resultado de esta operación, denotado por $R \cup S$, es una relación que incluye todas las tuplas que están en R o en S o en R y S. Duplicar se eliminan las tuplas.

INTERSECTION:

El resultado de esta operación, denotado por $R \cap S$, es una relación que incluye todas las tuplas que están tanto en R como en S.

SET DIFFERENCE (o MINUS):

el resultado de esta operación, denotado por $R - S$, es una relación que incluye todas las tuplas que están en R pero no en S.

Observe que tanto UNION como INTERSECTION son operaciones conmutativas; es decir, $R \cup S = S \cup R$ y $R \cap S = S \cap R$. Tanto UNION como INTERSECTION pueden tratarse como operaciones n-arias aplicables a cualquier número de relaciones porque ambas son también operaciones asociativas; es decir, $R \cup (S \cup T) = (R \cup S) \cup T$ y $(R \cap S) \cap T = R \cap (S \cap T)$

La operación MINUS no es conmutativa; es decir, en general,

$$R - S \neq S - R$$

En SQL, hay tres operaciones: UNION , INTERSECT y EXCEPT , que corresponden a las operaciones establecidas aquí descritas. Además, existen operaciones multiset (UNION ALL , INTERSECT ALL y EXCEPT ALL) que no eliminan duplicados.

EL PRODUCTO CARTESIANO (PRODUCTO CRUZADO)

También conocida como CROSS PRODUCTO o UNIÓN CRUZADA, que se indica con \times .

La operación PRODUCTO CARTESIANO aplicada por sí misma generalmente no tiene sentido. Es sobre todo útil cuando va seguido de una selección que coincide con los valores de los atributos que provienen de las relaciones de los componentes. Por ejemplo, supongamos que queremos recuperar una lista de nombres de las dependencias de cada empleada abolladuras. Podemos hacer esto de la siguiente manera:

```
FEMALE_EMPS ← σ Sex = 'F' ( EMPLOYEE )
EMP_NAMES ← π Fname, Lname, Ssn ( FEMALE_EMPS )
EMP_DEPENDENTS ← EMP_NAMES × DEPENDENT
ACTUAL_DEPENDENTS ← σ Ssn = Essn ( EMP_DEPENDENTS )
RESULT ← π Fname, Lname, Dependent_name ( ACTUAL_DEPENDENTS )
```

8.3 Operaciones relacionales binarias: JOIN y DIVISION

La operación JOIN:

se usa para combinar tuplas relacionadas de dos relaciones en tuplas simples "más largas".

Variaciones de JOIN: EQUIJOIN y NATURAL JOIN

- El uso más común de JOIN implica condiciones de unión con comparaciones de igualdad solamente, tal JOIN, donde el único operador de comparación utilizado es =, se denomina EQUIJOIN.
- La definición estándar de NATURAL JOIN requiere que los dos atributos se unan (o cada par de atributos de combinación) tienen el mismo nombre en ambas relaciones. Si este no es el caso, primero se aplica una operación de cambio de nombre.

Un conjunto completo de operaciones de álgebra relacional

Se ha demostrado que el conjunto de operaciones de álgebra relacional $\{\sigma, \pi, \cup, \rho, -, \times\}$ es un juego completo; es decir, cualquiera de las otras operaciones originales de álgebra relacional puede ser expresado como una secuencia de operaciones de este conjunto. Por ejemplo, la INTERSECCIÓN. La operación se puede expresar usando UNION y MINUS de la siguiente manera:

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

La operación DIVISION

En general, la operación DIVISION se aplica a dos relaciones $R(Z) \div S(X)$, donde el

los atributos de S son un subconjunto de los atributos de R ; es decir, $X \subseteq Z$. Sea Y el conjunto de atributos de R que no son atributos de S ; es decir, $Y = Z - X$ (y por tanto $Z = X \cup Y$).

Notación para árboles de consultas

Un árbol de consulta es una estructura de datos de árbol que corresponde a una expresión de álgebra relacional. Representa las relaciones de entrada de la consulta como nodos hoja del árbol, y representa Resiente las operaciones del álgebra relacional como nodos internos.

8.4 Operaciones relacionales adicionales

Algunas solicitudes de bases de datos comunes, que se necesitan en aplicaciones comerciales para RDBMS: no se puede realizar con las operaciones originales de álgebra relacional.

En esta sección definimos operaciones adicionales ciones para expresar estas solicitudes. Estas operaciones realzan el poder expresivo de el álgebra relacional original.

Proyección generalizada

extiende la operación de proyección al permitir funciones de los atributos que se incluirán en la lista de proyección. La forma generalizada se puede expresar como:

$$\pi_{F_1, F_2, \dots, F_n}(R)$$

donde F_1, F_2, \dots, F_n son funciones sobre los atributos en relación R y pueden involucrar operaciones aritméticas y valores constantes.

Funciones agregadas y agrupación

Ejemplos de tales funciones incluyen recuperar el salario promedio o total de todos los empleados o el número total de tuplas de empleados.

Las funciones comunes aplicadas a colecciones de valores numéricos incluyen SUM, AVERAGE, MAXIMUN y MINIMUN. La función COUNT se utiliza para contar tuplas o valores.

Operaciones de cierre recursivo

Un ejemplo de una operación recursiva es recuperar todos los supervisados de un empleado e en todos los niveles, es decir, todos los empleados e' supervisados directamente por e , todos los empleados e'' supervisados directamente por cada empleado e' , todos los

```
empleados e ' ' supervisan directamente  
visualizado por cada empleado e ' ', y así sucesivamente.
```

Operaciones OUTER JOIN

Las operaciones de unión que describimos anteriormente en La sección 8.3 son todas combinaciones internas. Esto equivale a la pérdida de información si el usuario quiere que el resultado de JOIN incluya todas las tuplas en una o más de las combinaciones relaciones con los ponentes.

Se desarrolló un conjunto de operaciones, llamadas combinaciones externas , para el caso en el que el usuario quiere mantener todas las tuplas en R, o todas las de S, o todas las de ambas relaciones en el resultado de JOIN , independientemente de si tienen o no tuplas coincidentes en la otra relación. Esto satisface la necesidad de consultas en las que tuplas de dos Las tablas deben combinarse haciendo coincidir las filas correspondientes, pero sin perder cualquier tupla por falta de valores coincidentes.

La operación UNIÓN EXTERIOR

La operación OUTER UNION fue desarrollada para tomar la unión de tuplas de dos relaciones que tienen algunos atributos comunes, pero no son compatibles con la unión (tipo).

UNIÓN EXTERIOR se puede aplicar a dos relaciones cuyos esquemas son ESTUDIANTE (Nombre , Ssn , Departamento , Asesor) e INSTRUCTOR (Nombre , Ssn , Departamento , Rango). Tuplas de las dos relaciones se emparejan basándose en tener la misma combinación de valores de los atributos compartidos: Nombre , Ssn , Departamento .

8.5 Ejemplos de consultas en álgebra relacional

8.6 El cálculo relacional de tuplas

Una expresión de cálculo especifica qué va a ser recuperado en lugar de cómo recuperarlo. Por tanto, el cálculo relacional es considerado un lenguaje no procesal .

Se considera un lenguaje de consulta relacional L relacionalmente completo si podemos expresar en L cualquier consulta que pueda expresarse en cálculo relacional.

El cálculo relacional es importante por dos razones:

- Tiene una base firme en lógica matemática.
- el lenguaje de consulta estándar (SQL) para RDBMS tiene su fundamento en el cálculo relacional de tuplas.

Variables de tupla y relaciones de rango

El cálculo relacional de tuplas se basa en especificar una serie de variables de tuplas .

La consulta de cálculo relacional de tupla tiene la forma:

$$\{t \mid \text{COND}(t)\}$$

donde t es una variable de tupla y $\text{COND}(t)$ es una expresión condicional (booleana) t que se evalúa como VERDADERO o FALSO para diferentes asignaciones de tuplas a la variable t .

Expresiones y fórmulas en cálculo relacional de tuplas

Una expresión general del cálculo relacional de tuplas tiene la forma:

$$\{t_1.A_j, t_2.A_k, \dots, t_n.A_m \mid \text{COND}(t_1, t_2, \dots, t_n, t_{n+1}, t_{n+2}, \dots, t_{n+m})\}$$

donde $t_1, t_2, \dots, t_n, t_{n+1}, \dots, t_{n+m}$ son variables de tupla, cada A_i es un atributo del relación en la que t_i varía, y COND es una condición o fórmula 13 de la relación tupla cálculo cional. Una fórmula se compone de átomos de cálculo de predicados, que pueden ser uno de los siguientes:

1. Un átomo de la forma $R(t_i)$, donde R es un nombre de relación y t_i es una variable tupla.
poder. Este átomo identifica el rango de la variable tupla t_i como la relación cuyo nombre es R . Se evalúa como VERDADERO si t_i es una tupla en la relación R , y se evalúa como FALSO en caso contrario.
2. Un átomo de la forma $t_i.A_{op} t_j.B$, donde op es una de las operaciones de comparación
tors en el conjunto $\{=, <, \leq, >, \geq, \neq\}$, t_i y t_j son variables de tupla, A es un atributo de la relación en la que t_i varía, y B es un atributo de la relación en la que t_j varía.
3. Un átomo de la forma $t_i.A_{op} c \text{ o } c_{op} t_j.B$, donde op es uno de los operadores en el conjunto $\{=, <, \leq, >, \geq, \neq\}$, t_i y t_j son variables de tupla, A es un atributo de la relación en la que t_i varía, B es un atributo de la relación en la que t_j varía, y c es un valor constante.

Los cuantificadores existenciales y universales

cuantificador universal (\forall) y el cuantificador existencial (\exists).

El cuantificador (\exists) se llama cuantificador existencial porque una fórmula $(\exists t)(F)$ es TRUE si existe alguna tupla que haga F TRUE. Para el cuantificador universal

$(\forall t) (F)$ es VERDADERO si todas las tuplas posibles que se pueden asignar a apariciones libres de t en F se sustituye por t , y F es VERDADERO para cada sustitución de este tipo.

Consultas de muestra en cálculo relacional de tuplas

Notación para gráficos de consulta

Estos tipos de consultas se conocen como consultas de selección-proyecto-uni3n porque solo involucran estas tres operaciones de 3lgebra relacional

La representaci3n de una consulta se denomina gr3fico de consulta .

Las relaciones en la consulta est3n representadas por nodos de relaci3n , que se muestran como c3rculos simples.

Los valores constantes, normalmente de las condiciones de selecci3n de la consulta, son representados por nodos constantes , que se muestran como c3rculos dobles u 3valos.

Las condiciones de uni3n y uni3n est3n representadas por los bordes del gr3fico (las l3neas que conectan los nodos)

Transformar los cuantificadores universal y existencial

Es posible transformar un cuantificador universal en un cuantificador existencial, y viceversa, para obtener una expresi3n equivalente.

Una transformaci3n general puede describirse informalmente como sigue:

formar un tipo de cuantificador en otro con negaci3n (precedido de NOT); Y y O reemplazarse entre s3; una f3rmula negada se vuelve innecesaria; y un la f3rmula negada se vuelve negada.

Uso del cuantificador universal en consultas Siempre que utilizamos un cuantificador universal, es bastante sensato seguir algunas reglas para asegurarse de que nuestra expresi3n tenga sentido.

Expresiones seguras

Una expresi3n segura en el c3lculo relacional es aquella que garantiza producir un n3mero finito de tuplas como resultado; de lo contrario, la expresi3n se denomina insegura . Por ejemplo, la expresi3n

$$\{t \mid \text{NO} (\text{EMPLEADO} (t)) \}$$

no es seguro porque produce todas las tuplas del universo que no son tuplas EMPLEADOS , que son infinitamente numerosos.

Se dice que una expresi3n es segura si todos los valores en su resultado son del dominio de la expresi3n.

8.7 El dominio del cálculo relacional

Existe otro tipo de cálculo relacional llamado cálculo relacional de dominio, o simplemente cálculo de dominio .

El cálculo de dominios difiere del cálculo de tuplas en el tipo de variables utilizadas en la fórmula. las: en lugar de tener un rango de variables en tuplas, las variables varían en un solo valores de dominios de atributos.

Para formar una relación de grado n para el resultado de una consulta, debemos tener n de estas variables de dominio, una para cada atributo. Una expresión del cálculo de dominio es de la forma

$$\{x_1, x_2, \dots, x_n \mid \text{COND} (x_1, x_2, \dots, x_n, x_{n+1}, x_{n+2}, \dots, x_{n+m})\}$$

donde $x_1, x_2, \dots, x_n, x_{n+1}, x_{n+2}, \dots, x_{n+m}$ son variables de dominio que van más allá de dominios (de atributos), y COND es una condición o fórmula del dominio cálculo relacional.