

Algebra e calcolo relazionale

Ripasso

Le 7 Virtù del DBMS

persistenza
affidabilità
volume
condivisione
riservatezza
efficienza
efficacia

I 4 Livelli di astrazione

Livello fisico (o interno)
Livello logico (aggiunta del significato al livello fisico, senza dettagli implementativi)
Livello esterno (altri modi per vedere un unico livello logico)
Livello concettuale (per modellare il mondo reale)

Le Tabelle

Tabella è una struttura bidimensionale dove ogni colonna è caratterizzata da un nome univoco, ed ogni riga rappresenta una n-upla
Tutte le relazioni matematiche possono essere rappresentate come tabelle
Non tutte le tabelle sono rappresentazioni di relazioni matematiche
Una tabella rappresenta una relazione se

- le righe sono diverse fra loro
- le intestazioni delle colonne sono diverse tra loro
- i valori di ogni colonna sono fra loro omogenei

Recupero

L'informazione incompleta – NULL

Cosa significa non avere il valore di un attributo

- non esiste quel valore
- il valore esiste ma non è noto
- il valore non è rappresentabile con quella struttura dati (pi-greco)
- etc..

Perché è necessario avere un valore NULL

Non conviene usare valori del dominio come “Jolly” non utilizzati perché:

- i valori Jolly potrebbero diventare non più non utilizzati
- è sempre necessario interpretare il valore secondo una codifica non esplicita

Non sempre ad un attributo A di una relazione R può essere concesso di assumere il valore NULL: è necessario definire dei “vincoli” sul modello

I Vincoli

Definizione di vincolo:

restrizione imposta alle possibili ennuple di uno schema di basi di dati affinché tali tuple modellino in modo corretto una realtà specifica

Cioè i vincoli rappresentano i confini imposti dalla realtà alle infinite combinazioni dei dati insiemistici e matematici.

Oppure anche un vincolo è una proprietà che deve essere soddisfatta dalle istanze che rappresentano informazioni corrette per l'applicazione

E' sempre quindi relativo ad un'applicazione, non ad un dominio del possibile matematico.

I vincoli sono di diversi tipi :

- intrarelazionali (relativi ai valori di una singola ennupla o del singolo record)
 - di dominio (che riguardano i valori ammissibili per un certo attributo)
 - di ennupla (che legano tra loro i valo i dei vari domini di una ennupla)
- relazionali (che coinvolgono una relazione nella sua interezza)

- interrelazionali (cioè tra ennuple di diverse relazioni)

I vincoli intrarelazionali hanno la caratteristica di poter essere valutati considerando soltanto la singola ennupla, mentre quelli interrelazionali richiedono la valutazione di più relazioni contemporaneamente.

I vincoli relazionali hanno invece la caratteristica di riguardare una singola relazione, ma coinvolgono tutte le ennuple di tale relazione, possono essere visti come un caso degenerare di interrelazionale.

Ma pur essendo un caso “degenerare” hanno un'importanza particolare proprio nel modello relazionale per la definizione delle CHIAVI delle relazioni

Chiavi ed identificazione delle ennuple

Definizione intuitiva di chiave di una relazione:

insieme di attributi che identifica in modo univoco le ennuple di una relazione

Definizione matematica di chiave:

- sia data una relazione r
- un insieme K di attributi è detto superchiave per r se r non contiene due ennuple distinte t_1 e t_2 per le quali $t_1[K] = t_2[K]$
- K è detto chiave per r se non esiste un sottoinsieme K_i di K che sia a sua volta superchiave di r (cioè K è una superchiave minimale di r)

Spesso si rischia analizzando solo i dati di esempio di una relazione di identificare come chiave un set di attributi che in realtà (cioè nel mondo reale che quello schema di relazioni vuole modellare) non ha tale caratteristica.

ATTENZIONE: nei DBMS relazionali non è necessario che per ogni tabella (relazione) sia definita una chiave, ma in tal caso non sarà sempre possibile operare sulla singola ennupla in quanto potrà capitare di avere in una tabella una serie di ennuple tra loro indistinguibili.

Questo non è di solito un bene, ma bisogna vedere cosa in realtà quello schema modella.

Esempi di rappresentazione di informazioni nel modello relazionale

Codifica delle informazioni (tabelle di trascodifica)

Strutture nidificate (o master-detail)

Strutture ricorsive (gerarchia, albero, grafo comunque connesso)

Algebra e calcolo relazionale

Introduzione ai linguaggi

Il DBMS è una scatola nera, ma dobbiamo interagire con esso.

Per interagire con il DBMS vengono definiti dei linguaggi per le Basi di Dati

Che tipo di operazioni possono essere fatte sul DBMS ?

- Interrogazione sui dati (Query)
- Manipolazione dei dati
- Definizione del modello dei dati

Funzioni dei linguaggi:

Interrogazione e manipolazione dei dati :

DML: data manipulation language

Definizione del modello dei dati :

DDL: data definition language

Tipologie dei linguaggi:

Dichiarativi - specificano le proprietà del risultato ("che cosa")

Procedurale - specificano le modalità di generazione del risultato ("come")

Esempi di linguaggi di interrogazione per DB:

Algebra relazionale – matematico di tipo procedurale basato sull'insiemistica

Calcolo relazionale – teorico matematico di tipo dichiarativo

SQL (Structured Query Language) – reale e solo parzialmente dichiarativo

QBE (Query by Example) – reale di tipo dichiarativo

Datalog – reale di programmazione logica basato su regole ed inferenza

L'Algebra relazionale

L'Algebra relazionale è una formulazione rigorosa e matematica che permette di operare sulle relazioni, dove le relazioni vengono viste come insiemi di n-uple manipolate attraverso operatori appunto insiemistici.

Si basa su:

- una serie di operatori insiemistici
- che operano sulle relazioni
- che possono essere tra loro combinati
- e che producono come risultato delle altre relazioni

Unione, intersezione, differenza

Sono le tre operazioni base dell'insiemistica ed hanno una corrispondenza diretta con gli operatori della logica booleana.

Sono operazioni binarie, cioè si applicano ad una coppia di relazioni (o insiemi)

Si possono applicare solo a relazioni che hanno lo stesso insieme di attributi o domini

Questa limitazione è di tipo matematico ma anche logico: non si sommano mele con pere

Il risultato di queste operazioni è sempre una relazione il cui schema è a sua volta uguale a quello delle due (o più) relazioni di partenza.

Unione : tutti gli elementi di un insieme o dell'altro – **OR - U**

Intersezione : tutti gli elementi di un insieme e dell'altro – **AND - \cap**

Differenza : tutti gli elementi di un insieme che non sono anche dell'altro - **NOT - -**

Queste operazioni possono essere tra loro liberamente combinate ed esistono diversi modi per ottenere lo stesso risultato, seguendo le varie proprietà associative, commutative, etc. ad esempio:

$$\mathbf{r \cap (s \cup t) = (r \cap s) \cup (r \cap t)}$$

In particolare l'intersezione può essere definita a partire dalla differenza :

$$\mathbf{r \cap s = r - (r - s)}$$

Quindi l'insieme minimo di operandi è dato da Unione e Differenza.

Selezione

La selezione è un operatore “monadico”, cioè che si applica ad una sola relazione

Produce una relazione il cui schema è identico a quello della relazione di partenza (detta “operando”)

La relazione ottenuta tramite una **selezione** contiene un sottoinsieme delle n-uple dell'**operando**, in particolare tutte e sole quelle che soddisfano una **condizione** definita dalla selezione stessa.

La selezione ha quindi tre elementi:

1. operando (relazione di partenza)
2. risultato (relazione risultante)
3. condizione (regola o funzione booleana che debbono soddisfare le n-uple dell'operando per appartenere alla relazione risultante)

Sintassi della selezione: **SEL** *Condizione* (*Operando*)

Una selezione conterrà al massimo lo stesso numero di n-uple dell'operando, ma tendenzialmente meno in funzione del livello di “ristrettezza” delle condizioni applicate.

Proiezione

La selezione è anche lei un operatore “monadico”.

Produce una relazione il cui schema è un sottoinsieme di quello della relazione di partenza (detta “operando”)

La relazione ottenuta tramite una **proiezione** contiene al massimo un numero di n-uple pari a quello delle n-uple dell'operando.

La proiezione ha tre elementi:

4. operando (relazione di partenza)
5. risultato (relazione risultante)
6. lista degli attributi dell'operando che devono appartenere alla relazione risultante

Sintassi della proiezione: $\text{PROJ}_{\text{ListaAttributi}}(\text{Operando})$

E' importante fare delle considerazioni sulla cardinalità del risultato di una proiezione.

ATTENZIONE:

parliamo di relazioni MATEMATICHE, quindi che NON contengono n-uple uguali.

La relazione ottenuta tramite una **proiezione** contiene al massimo un numero di n-uple pari a quello delle n-uple dell'operando, ma può contenerne meno se:

la lista degli attributi della proiezione non è una superchiave dell'Operando

Se X è superchiave di R, allora numerosità ($\text{PROJ}_X(R)$) = numerosità (R)

Cioè

$$|\text{PROJ}_X(R)| = |R| \text{ se } X \text{ superchiave di } R$$

Considerazioni su Selezione e Proiezione

La selezione e la proiezione possono essere viste come due operazioni tra loro “ortogonali”.

Infatti :o

- la selezione “taglia” delle fette orizzontali di una tabella (appunto seleziona delle n-uple tra tutte quelle della relazione di partenza)
- la proiezione “taglia” delle fette verticali della stessa tabella (selezionando solo alcune delle colonne)

La selezione e la proiezione possono essere combinate per ottenere delle porzioni particolari di una tabella (o relazione)

Ridenominazione

La ridenominazione è anche un operatore “monadico”, che lascia inalterato il contenuto di una

relazione ma ne modifica lo schema cambiando il nome di alcuni attributi.

La funzione della ridenominazione è quella di permettere di trasformare due relazioni che hanno informazioni tra loro coerenti ma schemi i cui nomi differiscono, in modo da poter poi applicare su tali relazioni le altre operazioni dell'algebra relazionale.

È importante quando due o più relazioni esprimono particolari tipologie di un'unica entità che viene in esse dettagliata secondo maggiori dettagli trascurabili nell'interrogazione in atto.

La sintassi di una ridenominazione è: $REN\ y \leftarrow x\ (R)$

Join

Il join (combinazione, aggregazione) è uno degli operatori più interessanti dell'algebra relazionale poiché permette di correlare dati tra relazioni diverse.

$R_1 JOIN R_2$ è una relazione su $X_1 X_2$

$$\{ t \text{ su } X_1 X_2 \mid \text{esistono } t_1 \in R_1 \text{ e } t_2 \in R_2 \\ \text{con } t[X_1] = t_1 \text{ e } t[X_2] = t_2 \}$$

Cardinalità del JOIN

Il join di R_1 e R_2 contiene un numero di n-uple compreso fra zero e il prodotto di $|R_1|$ e $|R_2|$

NATURAL JOIN

Il NATURAL JOIN è un tipo di operazione che ci permette di correlare due o più tabelle sulla base di valori uguali in attributi contenenti lo stesso tipo di dati.

INNER JOIN - Interno

È un tipo di join in cui le righe delle tabelle vengono combinate solo se i campi collegati con join soddisfano una determinata condizione.

OUTER JOIN - Esterno

Con l'OUTER JOIN è possibile estrapolare anche quei dati, *appartenenti ad una delle tabelle*, che non verrebbero estrapolati nei tipi di join visti fino a questo momento. Infatti OUTER significa esterno; dati esterni al normale tipo di join.

Il join esterno completa, con valori nulli, le n-uple che verrebbero tagliate fuori da un join (interno).

Può essere di tre tipi:

- sinistro
- destro
- completo

Join e proiezioni

Prodotto cartesiano

theta-join

$R_1 \text{ JOIN}_{\text{Condizione}} R_2$

Equivalenza di espressioni

Push selections (se A è attributo di R_2)

$\text{SEL } A=10 (R_1 \text{ JOIN } R_2) = R_1 \text{ JOIN SEL } A=10 (R_2)$

Selezione con valori nulli

Viste (relazioni derivate)

Viste virtuali e materializzate

Viste e aggiornamenti

Conclusione Algebra relazionale

Operatori dell'algebra relazionale

unione, intersezione, differenza

selezione

proiezione

ridenominazione

join (join naturale, prodotto cartesiano, theta-join)

Calcolo relazionale

Calcolo e algebra

Calcolo e algebra sono "equivalenti"

per ogni espressione del calcolo relazionale che sia indipendente dal dominio esiste un'espressione dell'algebra relazionale equivalente a essa

per ogni espressione dell'algebra relazionale esiste un'espressione del calcolo relazionale equivalente a essa (e di conseguenza indipendente dal dominio)

Limiti dell'algebra e del calcolo relazionale

Chiusura transitiva

Supervisione(Impiegato, Capo)

Per ogni impiegato, trovare tutti i superiori (cioè il capo, il capo del capo, e così via)

Chiusura transitiva, come si fa?

Nell'esempio, basterebbe il join della relazione con se stessa, previa opportuna ridenominazione, ma servirebbero potenzialmente infinite join e ridenominazioni

Chiusura transitiva, impossibile!

Non esiste in algebra e calcolo relazionale la possibilità di esprimere l'interrogazione che, per ogni relazione binaria, ne calcoli la chiusura transitiva