Basi di Dati

Algebra Relazionale

Raffaella Gentilini

October 11, 2020

Algebra Relazionale: Introduzione

Algebra Relazionale

Linguaggio per l'interrogazione di BD basate sul modello relazionale:

- E' costituita da un insieme di operatori unari e binari su istanze di relazioni:
  - ciascun operatore ha come argomento una o due istanze di relazioni e produce una nuova istanza di relazione.
- E' un linguaggio procedurale: viene specificata la sequenza di operazioni necessarie per ottenere il risultato atteso.
- L'algebra relazionale e' importante perche':
  - fondamento formale per le operazioni nel modello relazionale
  - base per implementare ed ottimizzare le interrogazioni nei RDBMS commerciali
  - alcuni suoi concetti sono incorporati nel linguaggio di interrogazione SQL, standard per i RDBMS

Operatori Algebra Relazionale: Overview

Gli operatori dell'algebra relazionale possono essere classificati in:

1. Operatori Insiemistici:
   - unione, intersezione, differenza, prodotto cartesiano
2. Operatori Propriamente Relazionali:
   - ridenominazione, selezione, proiezione, concatenazione (join), divisione

Non tutti i precedenti operatori sono primitivi: l'intersezione, la concatenazione e la divisione possono infatti essere espressi mediante i restanti operatori (lo vedremo).

La Ridenominazione

- Inizieremo ad illustrare gli operatori insiemistici dell'algebra relazionale.
- Poiche' semplice ed utile nel contesto delle operazioni insiemistiche, presenteremo prima l'operatore relazionale di ridenominazione.

La Ridenominazione $\rho_{\_}(r)$

Data un' istanza di relazione r sullo schema di relazione $R(A_1, \ldots, A_n)$, la ridenominazione:

$$\rho_{S(A_1 \rightarrow B_1, \ldots, A_n \rightarrow B_n)}(r)$$

del nome di relazione R con il nome di relazione S e degli attributi $A_1, \ldots, A_n$ con $B_1, \ldots, B_n$ e' definita dall' istanza di relazione s sullo schema di relazione $S(B_1, \ldots, B_n)$, dove:

$$s = \{\{(B_1, v_1), \ldots, (B_n, v_n)\} \mid \{(A_1, v_1), \ldots, (A_n, v_n)\} \in r\}$$

Notazioni

Nel seguito utilizzeremo le notazioni:

- $\rho_S(r)$ per indicare una ridenominazione del solo nome di relazione
- $\rho_{A_1 \to B_1, \ldots, A_j \to B_j}(r)$ per indicare una ridenominazione dei soli attributi $A_1, \ldots, A_j$ in $B_1, \ldots, B_j$
- $S(B_1, \ldots, B_n) \leftarrow R(A_1, \ldots, A_n)$ oppure $S \leftarrow R$ per indicare la ridenominazione in una sequenza di operazioni dell'algebra relazionale

Operatori insiemistici

La seguente definizione introduce la nozione di compatibilita all'unione per due schemi di relazione, utile a definire nel seguito gli operatori insiemistici dell'algebra relazionale di unione, intersezione e differenza.

Relazioni Compatibili all'Unione

Le relazioni $R(A_1, \ldots, A_n)$, $S(B_1, \ldots, B_m)$ sono dette compatibili all'unione sse:

- hanno lo stesso grado, ovvero $n = m$
- $\forall i = 1 \ldots n : Dom(A_i) = Dom(B_i)$

L'Unione

Unione di Relazioni

Siano r, s sue istanza di relazione i cui schemi ($R(X)$ ed $S(Y)$, rispettivamente) sono compatibili all'unione. L'unione applicata ad r, s, indicata con $r \cup s$ e' una relazione sull'insieme di attributi X contenente le tuple che appartengono ad r oppure ad s:

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

L'Intersezione

Intersezione di Relazioni

Siano r, s sue istanza di relazione i cui schemi ($R(X)$ ed $S(Y)$, rispettivamente) sono compatibili all'unione. L'intersezione applicata ad r, s, indicata con $r \cap s$ e' una relazione sull'insieme di attributi X contenente le tuple che appartengono sia ad r che ad s:

$$r \cap s = \{t \mid t \in r \wedge t \in s\}$$

La Differenza

Differenza di Relazioni

Siano r, s sue istanza di relazione i cui schemi ($R(X)$ ed $S(Y)$, rispettivamente) sono compatibili all'unione. La differenza $r \setminus s$ e' una relazione sull'insieme di attributi X contenente le tuple che appartengono ad r ma non appartengono ad s:

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

Operatori Insiemistici: Esempio

R

A	B	C
1	x	a
1	x	b
2	x	a
3	x	a

S

M	N	O
1	x	a
1	y	b

$R \cup S$

A	B	C
1	x	a
1	x	b
2	x	a
3	x	a
1	y	b

$R \cap S$

A	B	C
1	x	a

$S \setminus R$

M	N	O
1	y	b

Il Prodotto Cartesiano

L'ultimo operatore insiemistico dell'algebra relazionale e' il prodotto cartesiano.

Prodotto Cartesiano di Relazioni

Siano $R(X), S(Y)$ due schemi di relazione tali che $X = \{A_1, \ldots, A_n\}$, $Y = \{B_1, \ldots, B_m\}$ ed $X \cap Y = \emptyset$ e si considerino due istanze di relazione r, s sugli schemi $R(X), S(Y)$. L'operatore di prodotto cartesiano $r \times s$ produce un'istanza di relazione formata da tutte le tuple che e' possibile ottenere unendo le tuple di r ed s:

$$r \times s =$$

$$\{\{(A_1, v_1), \ldots, (A_n, v_n)\} \cup \{(B_1, v_1), \ldots, (B_n, v_m)\} \mid$$

$$\{(A_1, v_1), \ldots, (A_n, v_n)\} \in r \wedge \{(B_1, v_1), \ldots, (B_n, v_m)\} \in s\}$$

Prodotto Cartesiano: Esempio

LINGUE

LINGUA	% POP.
Cinese	11.9
Spagnolo	5.6
Inglese	4.9

STATI

STATO	MADRELINGUA
Messico	spagnolo
Colombia	"

LINGUA	% POP	STATO	MADRELINGUA
cinese	11.9	Messico	spagnolo
spagnolo	5.6	Messico	"
inglese	4.9	Messico	"
cinese	11.9	Colombia	"
spagnolo	5.6	"	"
inglese	4.9	"	"

LINGUE × STATI

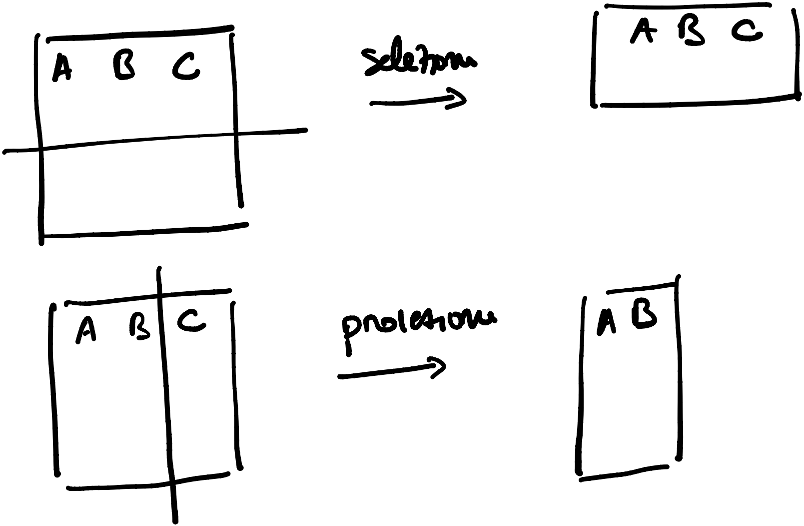
Selezione e Proiezione: Introduzione

Selezione e Proiezione

Gli operatori di selezione e proiezione sono operatori unari che svolgono funzioni complementari:

- La selezione produce come risultato un'istanza di relazione costituita da un sottoinsieme di tuple dell'istanza di relazione in input
- La proiezione produce come risultato un'istanza di relazione costituita da un sottoinsieme di colonne della tabella che illustra l'istanza di relazione in input.

Selezione e Proiezione: Intuizione



La selezione

L'operatore di selezione e' denotato con il simbolo σ_F, dove F rappresenta la condizione di selezione.

Condizione di Selezione

Data $R(X)$, una condizione di selezione su X e' una formula proposizionale F, ovvero una formula ottenuta combinando con i connettivi $\wedge, \vee$ e $\neg$ condizioni atomiche (clausole) del tipo $A\theta B$ oppure $A\theta c$, dove:

- θ e' un operatore di confronto $\theta \in \{=, \neq, >, \geq, \leq, <\}$
- A e B sono attributi in X sui cui valori il confronto θ abbia senso
- c e' una costante compatibile con il dominio di A

Data una tupla t in un'istanza di relazione r sullo schema $R(X)$:

- $A\theta B$ e' vera su t sse $t[A]$ e' in relazione θ con $t[B]$
- $A\theta c$ e' vera su t sse $t[A]$ e' in relazione θ con c
- $F_1 \vee F_2, F_1 \wedge F_2, \neg F_1$ hanno l'usuale significato.

L'Operatore di Selezione

La Selezione σ_F

Data un'istanza di relazione r sullo schema $R(X)$ ed una condizione di selezione F (su X), l'operazione di selezione $\sigma_F(r)$ produce una relazione su X che contiene le sole tuple di r su cui F e' vera:

$$\sigma_F(r) = \{t \mid t \in r \wedge t \models F\}$$

Selezione: Esempio

LINGUE

LINGUA	% POP
Cinese	11.9
Spagnolo	5.9
Inglese	4.9
Hindu	4.4
Arabo	4.2

$\sigma_{\%\,POP > 10}$ (LINGUE)

LINGUA	% POP
Cinese	11.9

L'Operatore di Proiezione

La Proiezione π_Y

Data un'istanza di relazione r sullo schema $R(X)$ ed un sottoinsieme di attributi $Y \subseteq X$, l'operazione di proiezione $\pi_Y(r)$ e' definita da:

$$\pi_Y(r) = \{t[Y] \mid t \in r\}$$

ovvero $\pi_Y(r)$ contiene le tuple su Y ottenute dalle tuple di r considerando solo i valori su Y.

Proiezione: Esempio

$\pi_{\text{CAPITALE}}(\text{STATI})$

STATI

STATO	CAPITALE	LINGUA	SUPERFICIE
Austria	Vienna	Tedesco	83871
Cipro	Nicosia	Greco Turco	9251
Germania	Berlino	Tedesco	357022
Italia	Roma	Italiano	301340
Russia	Mosca	Russo	3492550

CAPITALE
Vienna
Nicosia
Berlino
Roma
Mosca

La Concatenazione: Introduzione al Join

L'operatore derivato dell'algebra relazionale di join (concatenazione):

- E' indicato con il simbolo $\bowtie_F$, dove F indica la condizione di join.
- E' usato per unire tuple logicamente collegate, provenienti da due relazioni, in tuple singole.
- Puo' essere definito mediante gli operatori primitivi (ovvero non derivati) di selezione e prodotto cartesiano.

Il θ-Join

La Concatenazione (θ-join) $\bowtie_F$

Siano $R(X), S(Y)$ due schemi di relazione tali che $X \cap Y = \emptyset$ e siano r, s due istanze di $R(X), S(Y)$. Sia F una formula proposizinale su condizioni atomiche del tipo $A\theta B$, dove $A \in X$, $B \in Y$ e θ e' un operatore di confronto.

L'operatore di θ-join $r \bowtie_F s$ produce una relazione formata da tutte le combinazioni di tuple che soddisfano la condizione di join F:

$$r \bowtie_F s = \sigma_F(r \times s)$$

θ-Join: Esempio

CONTINENTI

NOME	SUPERFICIE	STATI
Asia	44579000	51
America	42549000	35
Africa	30221532	54
Europa	10300734	48

STATI

STATO	CAPITALE	CONTINENTE
Marocco	Rabat	Africa
Mali	Bamako	Africa
Italia	Roma	Europa
Francia	Parigi	Europa
Ruanda	Kigali	Africa
Brasile	Brasilia	America
Giordania	Amman	Asia

- reperire le capitali degli stati che appartengono a continenti con una superficie di almeno 40.000.000 km^2

$RIS(CAP) \leftarrow \pi_{capitale}(\sigma_{superficie > 40.000.000}(STATI \bowtie_{continente = nome} CONTINENTI))$

Equijoin

Equijoin

Un equijoin e' un θ-join dove la condizione di join F e' una congiunzione di condizioni atomiche di uguaglianze.

- CONTINENTE $\bowtie_{Nome = Continente}$ STATI è un esempio di equijoin.

Il Join Naturale: Premesse

- Il risultato di un'operazione di equijoin e' una relazione dove ogni tupla assume gli stessi valori sugli attributi di equijoin.
- tali attributi appaiono dunque ripetuti nel risultato.
- L'operatore (derivato) di join naturale permette di eliminare le suddette duplicazioni di un equijoin.
- In particolare, date due istanze di relazioni r, s il join naturale produce una relazione le cui tuple sono ottenute concatenando le tuple di r ed s che hanno gli stessi valori sugli attributi comuni.

Il Join Naturale

Join Naturale $\bowtie$

Siano $R(X), S(Y)$ due schemi di relazione, sia $Z = X \cap Y$ e si considerino due istanze r, s di $R(X), S(Y)$.
La concatenazione naturale (join naturale) $r \bowtie s$ e' definita come:

$$r \bowtie s = \{t \cup t' \mid t \in r \wedge t' \in s \wedge t[Z] = t'[Z]\}$$

Join Naturale: Esempio

Si considerino le relazioni della slide 23 :

CONTINENTE (NOME , SUPERFICIE , STATI)
STATI (STATO , CAPITALE , CONTINENTE)

La sequenza di operazioni :

C (CON , SUP , N_S) ← CONTINENTI (NOME , SUPERFICIE , STATI)
S (ST , CAP , CON) ← STATI (STATO , CAPITALE , CONTINENTE)
C ⋈ S

produce la relazione :

CON	SUP	N_S	ST	CAP
Africa	30221532	54	Marocco	Rabat
"	"	"	Mali	Bamako
Europa	10300734	48	Italia	Roma
"	"	"	Francia	Parigi
Africa	30221532	55	Ruanda	Kigali
America	42549000	35	Brasile	Brasilia
Asia	44579000	51	Singapore	Ankara

dove la colonna in comune (di S e C) CON appare solo una volta !

Join Naturale: Esempio

- Recuperare le capitali degli stati memorizzati nella BD della slide 26 che appartengono a continenti con una superficie di almeno 40.000.000 di Km^2.

$C(CON, SUP, N\_S) \leftarrow CONTINENTI(NOME, SUPERFICIE, STATI)$

$S(ST, CAP, CON) \leftarrow STATI(STATO, CAPITALE, CONTINENTE)$

$RIS \leftarrow \pi_{CAP}(\sigma_{SUP > 40.000.000}(C \bowtie S))$

Il risultato sarà:

RIS

CAP
Brasilia
Ammon