

Esercizio sull'ottimizzazione e stima di costo

Si consideri il seguente schema relazionale contenente le ricette di una catena di ristoranti:

INGREDIENTE (Codice, Nome, Calorie);
COMPOSIZIONE (Ricetta, Ingrediente, Quantità)
RICETTA (CodiceRicetta, Nome, Regione, TempoPreparazione)

Nota: la quantità nella tabella COMPOSIZIONE è espressa in grammi

Vincoli di integrità referenziale:

COMPOSIZIONE.Ricetta → RICETTA,
COMPOSIZIONE.Ingrediente → INGREDIENTE

Data la seguente interrogazione SQL che consente di:

Trovare gli ingredienti usati in ricette della Regione Veneto, riportando, il codice della ricetta e il nome e le calorie dell'ingrediente.

```
SELECT  R.CodiceRicetta, I.Nome, I.Calorie
FROM    RICETTA R JOIN COMPOSIZIONE C ON R.CodiceRicetta = C.Ricetta
        JOIN INGREDIENTE I ON C.Ingrediente = I.Codice
WHERE   R.Regione = 'Veneto'
```

(1) Calcolare il costo dell'interrogazione in termini di numero di accessi a memoria secondaria sotto le seguenti ipotesi:

- la selezione su ricetta richiede una scansione sequenziale della tabella RICETTA
- l'ordine di esecuzione del join è (RICETTA ⋈ COMPOSIZIONE) ⋈ INGREDIENTE
- le operazioni di join vengono eseguite con la tecnica Nested Loop Join con una pagina di buffer disponibile per ogni tabella
- il risultato intermedio del primo join viene interamente memorizzato nel buffer
- $NP(INGREDIENTE) = 40$, $NP(COMPOSIZIONE) = 200$, $NP(RICETTA) = 12$
- $NR(INGREDIENTE) = 1200$, $NR(COMPOSIZIONE) = 13000$, $NR(RICETTA) = 260$
- $VAL(Regione, RICETTA) = 20$, $VAL(Ricetta, COMPOSIZIONE) = 260$

(2) Come cambia il costo se è disponibile un indice B+-tree sull'attributo Codice della tabella INGREDIENTE che ha profondità 2.

Si consideri il seguente schema relazionale contenente le ricette di una catena di ristoranti:

INGREDIENTE (Codice, Nome, Calorie);
 COMPOSIZIONE (Ricetta, Ingrediente, Quantità)
 RICETTA (CodiceRicetta, Nome, Regione, TempoPreparazione)

Applicando la formula per il costo della tecnica "Nested Loop Join" al caso specifico e secondo l'ordine di join indicato risulta:

A) Selezione ricette della regione Veneto e PRIMO JOIN

$\text{sel_veneto}(\text{RICETTA}) \bowtie \text{COMPOSIZIONE}$

$\text{COSTO A} = \text{NP}(\text{RICETTA}) + \text{NR}(\text{RICETTA con selezione Regione='Veneto'}) * \text{NP}(\text{COMPOSIZIONE})$

B) SECONDO JOIN

$\text{COSTO B} = (\text{nessun costo di lettura della tabella esterna: è già nel buffer}) + \text{NR}(\text{COMPOSIZIONE per le ricette con selezione Regione='Veneto'}) * \text{NP}(\text{INGREDIENTE})$

Date le statistiche seguenti:

TABELLA	NP	NR	VAL (Regione)	VAL (Ricetta)
RICETTA	12	260	20	
INGREDIENTE	40	1200		
COMPOSIZIONE	200	13000		260

$\text{COSTO A} = 12 + \text{NR}(\text{RICETTA}) / \text{VAL}(\text{Regione, RICETTA}) * 200 = 12 + 260 / 20 * 200 = 12 + 13 * 200 = 2612$

$\text{COSTO B} = 0 + (\text{stima delle tuple di COMPOSIZIONE che rappresentano ingredienti di ricette del Veneto}) * 40 = (\text{numero medio di tuple di COMPOSIZIONE per ricetta}) * (\text{numero di ricette del Veneto}) * 40 = (\text{NR}(\text{COMPOSIZIONE}) / \text{VAL}(\text{Ricetta, COMPOSIZIONE}) * (13) * 40 = (13000/260 * 13) * 40 = 650 * 40 = 26000$

$\text{COSTO TOTALE} = 28612$

Come cambia il costo se è disponibile un indice B+-tree sull'attributo Codice della tabella INGREDIENTE che ha profondità 2.

$\text{COSTO A} = \text{come prima}$

$\text{COSTO B} = (\text{nessun costo di lettura della tabella esterna: è già nel buffer}) + \text{NR}(\text{COMPOSIZIONE per le ricette con selezione Regione='Veneto'}) * (\text{Profondità Indice} + \text{selettività di Codice in INGREDIENTE})$

$\text{COSTO A} = 2612$

$\text{COSTO B} = (13000/260 * 13) * (2 + 1) = 650 * 3 = 1950$

$\text{COSTO TOTALE} = 4562$