

## Esercizio 1

*Si considerino le seguenti specifiche relative alla realizzazione del sistema informativo per la gestione del catalogo dei corsi di studio all'estero proposti da un operatore del settore e si definisca il relativo **schema E/R** (nella metodologia proposta a lezione). Si evidenzino eventuali **vincoli inespressi e attributi derivati**.*

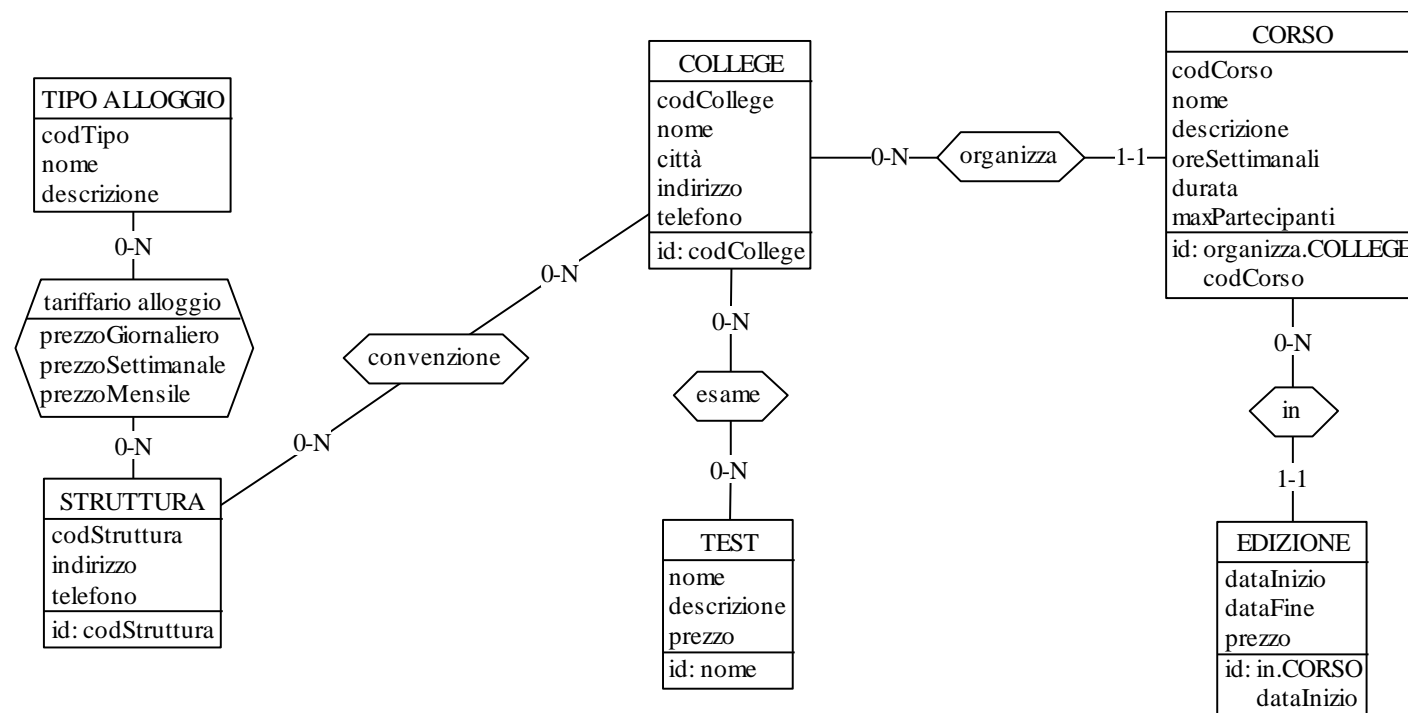
Si vuole realizzare un sistema informativo per la gestione del catalogo di corsi di studio della lingua inglese all'estero proposti da un operatore del settore.

I corsi vengono organizzati da college inglesi. Per ciascun college si memorizzano la città, il nome, l'indirizzo e il numero di telefono. Ogni college offre diversi corsi (es. standard, intensive, ecc.) identificati da un codice univoco all'interno del college; ciascun corso si ripete in diversi periodi dell'anno. Di ogni corso si registrano il nome, una descrizione, il numero di ore settimanali di lezione, la durata (in settimane) e il numero massimo di partecipanti. Ogni college ha un proprio tariffario per i corsi il cui costo dipende anche dal periodo di svolgimento.

I college organizzano anche un servizio di alloggio. Esistono diverse tipologie di alloggio (es. camera singola in ostello, camera doppia in ostello, sistemazione in famiglia, ecc.). Il sistema deve memorizzare l'elenco di strutture convenzionate con il college che offrono un determinato tipo di alloggio e il relativo prezzo giornaliero, settimanale e mensile. Il prezzo praticato da una struttura dipende solo dal tipo di alloggio ed è lo stesso per tutti i college convenzionati con tale struttura.

Alcuni college offrono anche la possibilità di sostenere un test finale in collaborazione con organizzazioni internazionali che gestiscono esami standardizzati (ad esempio Trinity, TOEFL, etc); il costo di un determinato test è lo stesso per tutti i college che lo offrono (ad esempio, il TOEFL costa 50 Euro indipendentemente dal college, il Trinity costa 40 Euro).

## Svolgimento



- Spiegare la differenza tra organizzazioni hash statiche e dinamiche e descrivere un'organizzazione dinamica a propria scelta.

### VEDERE LUCIDI SULL'ORGANIZZAZIONE DEI DATI

- E' data la query:

```
SELECT città, MIN(mq), MAX(mq), AVG(mq)
FROM APPARTAMENTO
WHERE numeroCamere = 2          (p1)
AND prezzo Giornaliero < 100    (p2)
GROUP BY città
```

sulla relazione:

APPARTAMENTO (codAppartamento, indirizzo, città, mq, numeroCamere, referente, telefono, prezzoGiornaliero)

sulla quale sono costruiti due indici: uno clustered su città e uno unclustered su numeroCamere.

Si calcoli la selettività dei predicati e si determini il miglior piano di accesso per la risoluzione della query tenendo conto dei seguenti dati:

NP = 4500, NT = 30000

NK<sub>città</sub> = 25, NK<sub>numeroCamere</sub> = 4

prezzoGiornaliero ∈ [50, 200], mq ∈ [35, 100]

Len(città) = 20 byte, Len(mq) = 4 byte, Len(numeroCamere) = 4 byte, Len(prezzoGiornaliero) = 4 byte,

Len(TID) = 4 byte

D = 1Kb, u=0.69

Si ipotizzi di utilizzare l'algoritmo Sort-Merge a Z=3 vie per l'eventuale ordinamento del risultato.

### Svolgimento

Calcoliamo innanzitutto il numero di foglie dei due indici:

$$NL_{città} = \left\lceil \frac{NK_{città} \times len(città) + NT \times len(TID)}{D \times u} \right\rceil = \left\lceil \frac{25 \times 20 + 30000 \times 4}{1024 \times 0.69} \right\rceil = 171$$

$$NL_{numeroCamere} = \left\lceil \frac{NK_{numeroCamere} \times len(numeroCamere) + NT \times len(TID)}{D \times u} \right\rceil = \left\lceil \frac{4 \times 4 + 30000 \times 4}{1024 \times 0.69} \right\rceil = 170$$

La selettività dei tre predicati è:

$$f(p1) = \frac{1}{NK_{numeroCamere}} = \frac{1}{4} \quad f(p2) = \frac{100 - \min(prezzoGiornaliero)}{\max(prezzoGiornaliero) - \min(prezzoGiornaliero)} = \frac{100 - 50}{200 - 50} = \frac{1}{3}$$

Il numero di tuple attese prima del GROUP BY risulta:

$$ET = NT \times f(p1) \times f(p2) = 30000 \times \frac{1}{4} \times \frac{1}{3} = 2500$$

Per quanto riguarda l'ordinamento del risultato, si suppone di utilizzare l'algoritmo di sort-merge a Z=3 vie. Il numero di pagine occupate dalla relazione da ordinare è:

$$NP_R = \left\lceil \frac{ET \times (\text{len}(\text{città}) + \text{len}(\text{mq}))}{D} \right\rceil = \left\lceil \frac{2500 \times 24}{1024} \right\rceil = 59$$

$$C_{\text{sort}} = 2 \times NP_R \times \lceil \log_Z NP_R \rceil = 472$$

Sono possibili tre piani d'accesso:

1. *Scansione sequenziale*

$$C_{\text{seq}} = NP = 4500$$

2. *Accesso tramite indice clustered su città*

Certamente non è conveniente perché il file è già ordinato su città e non ci sono predicati specifici su questo attributo; il costo d'accesso con indice sarà quindi superiore a quello con scansione sequenziale.

3. *Accesso tramite indice unclustered su numeroCamere*

$$C_{\text{numeroCamere}} = \lceil f(p1) \times NL_{\text{numeroCamere}} \rceil + EK_{\text{numeroCamere}} \times \Phi(ER, NP) + C_{\text{sort}}$$

$$= \left\lceil \frac{1}{4} \times 170 \right\rceil + \Phi\left(\frac{30000}{4}, 4500\right) + 472 = 43 + 3650 + 472 = 4165$$

Il piano d'accesso più conveniente è quello che fa uso dell'indice su numeroCamere.

Dato il seguente schema relazionale:

STUDENTE (Matricola, cognome, nome, luogoNascita)  
CORSO (codCorso, nomeCorso, numeroCrediti)  
ESAME (codCorso: CORSO, Matricola: STUDENTE, Anno, Voto)

1. Scrivere la query SQL che visualizza, per ogni studente, il numero degli esami superati nel 2005 e la media dei voti ottenuti, visualizzando anche gli studenti che non hanno superato esami nel 2005 (matricola, numeroEsami, votoMedio).
2. Scrivere la query SQL che visualizza gli studenti (Matricola, cognome, nome) per i quali il voto medio degli esami sostenuti nel 2007 è superiore al voto medio degli esami sostenuti nel 2006.
3. Scrivere un'espressione di algebra relazionale che visualizza la matricola degli studenti che hanno superato almeno un esame nel 2007 e nessuno nel 2006.

### Svolgimento

1. 

```
SELECT matricola, COUNT(codCorso), AVG(voto)
FROM STUDENTE S LEFT OUTER JOIN ESAME E ON (S.matricola=E.matricola AND anno=2005)
GROUP BY matricola
```
2. 

```
SELECT matricola, cognome, nome
FROM STUDENTE S, ESAME E
WHERE S.matricola = E.matricola
AND Anno=2007
GROUP BY Matricola, cognome, nome
HAVING AVG(voto) > (SELECT AVG(voto)
                    FROM ESAME E1
                    WHERE anno = 2006
                    AND E.matricola = E1.matricola)
```
3. 
$$\pi_{matricola}(STUDENTE \bowtie \sigma_{anno=2007}(ESAME)) - \pi_{matricola}(STUDENTE \bowtie \sigma_{anno=2006}(ESAME))$$