

Cognome e nome: _____ Matricola: _____ Turno: _____

Riportare sui fogli i seguenti dati: cognome, nome, matricola e turno di laboratorio.

Esame di SQL

Punteggi massimi:

- Domande 1 e 2 svolte perfettamente: 24;
- Domande 1 e 3 svolte perfettamente: 24;
- Domande 2 e 3 svolte perfettamente: 28;
- Domande 1, 2 e 3 svolte perfettamente: 33.

Lo svolgimento corretto di una sola domanda non permette il raggiungimento della sufficienza.

Le seguenti relazioni definiscono una base di dati “**Medaglie**” per gestire i piazzamenti nelle competizioni sportive individuali.

ATLETA(Codice, Cognome, Nome, Nazione, AnnoNascita, Genere)

SPECIALITÀ(Codice, Sport, Disciplina, Nome, Genere)

PIAZZAMENTO(Competizione, Anno, Specialità, Atleta, Posizione)

PIAZZAMENTO(Atleta) referencia ATLETA(Codice),

PIAZZAMENTO(Specialità) referencia SPECIALITÀ(Codice),

“Genere” in ATLETA, può assumere i valori “M” ed “F”.

“Genere” in SPECIALITÀ può assumere i valori “M” (maschile), “F” (femminile) ed “MF” (“Misto”).

“Sport” indica la categoria sportiva (per esempio: “Ciclismo”, “Atletica”, “Sport acquatici”).

“Disciplina” indica la disciplina sportiva (per esempio: “Ciclismo su pista”, “Corsa”, “Nuoto”).

“Nome” in SPECIALITÀ indica una descrizione, per esempio: “100 metri rana maschile”.

Competizione in PIAZZAMENTO indica la competizione sportiva in cui è stato ottenuto il piazzamento (per esempio: “Mondiali di Nuoto”, “Olimpiadi Invernali”, ecc.).

Anno indica l’edizione di una determinata competizione.

Posizione è un intero che indica la posizione ottenuta in gara. Le prime tre posizioni corrispondono alle classiche tre medaglie (oro, argento, bronzo).

Tutti gli altri attributi sono autoesplicativi.

Domanda 1 (bassa complessità).

Contare le specialità miste in cui, tra le prime tre posizioni di una stessa gara si sono piazzat* almeno un’atleta di genere femminile e un atleta di genere maschile, nat* o entramb* dopo il 2000, o entramb* prima del 1990.

Soluzione 1.

```
SELECT COUNT(DISTINCT s.Codice)
FROM specialità s JOIN piazzamento p1 ON (s.Codice=p1.Specialità) JOIN
    piazzamento p2 ON (s.Codice=p2.Specialità AND p1.Anno=p2.Anno AND
    p1.Competizione=p2.Competizione) JOIN atleta a1
    ON (a1.Codice=p1.Atleta) JOIN atleta a2 ON (a2.Codice=p2.Atleta)
WHERE s.Genere='MF' AND p1.Posizione<=3 AND p2.Posizione<=3 AND
    ((a1.AnnoNascita>2000 AND a2.AnnoNascita>2000) OR (a1.AnnoNascita<1990 AND
    a2.AnnoNascita<1990)) AND a1.Genere='F' AND a2.Genere='M';
```

Domanda 2 (media/alta complessità).

Per ogni disciplina esclusivamente femminile, indicare il cognome e nome delle atlete italiane che hanno vinto più di tre medaglie d’oro, non necessariamente nella stessa competizione.

Soluzione 2.

```
SELECT s.Disciplina, a.Nome, a.Cognome
FROM specialità s JOIN piazzamento p ON (s.Codice=p.Specialità) JOIN
    atleta a JOIN (a.Codice=p.Atleta)
WHERE s.Genere='F' AND p.Posizione=1 AND a.Genere='F' AND a.Nazione='Italia'
    AND NOT EXISTS (
    SELECT *
    FROM specialità s2
    WHERE s2.Sport=s.Sport AND s.Disciplina=s2.Disciplina AND
    s2.Genere<>'F')
GROUP BY s.Disciplina, a.Codice
HAVING COUNT(*)>3;
```

Domanda 3 (media/alta complessità).

Per ogni edizione di ogni competizione sportiva, stampare il numero di medaglie d'oro, d'argento e di bronzo e il numero totale di medaglie ottenute da ogni nazione.

Soluzione 3.

```
WITH medaglieoro AS (
    SELECT p.Competizione, p.Anno, a.Nazione, COUNT(*) AS NumOri
    FROM piazzamento p JOIN atleta a ON (p.Atleta=a.Codice)
    WHERE p.Posizione=1
    GROUP BY p.Competizione, p.Anno, a.Nazione),
medaglieargento AS (
    SELECT p.Competizione, p.Anno, a.Nazione, COUNT(*) AS NumArgenti
    FROM piazzamento p JOIN atleta a ON (p.Atleta=a.Codice)
    WHERE p.Posizione=2
    GROUP BY p.Competizione, p.Anno, a.Nazione),
medagliebronzo AS (
    SELECT p.Competizione, p.Anno, a.Nazione, COUNT(*) AS NumBronzi
    FROM piazzamento p JOIN atleta a ON (p.Atleta=a.Codice)
    WHERE p.Posizione=3
    GROUP BY p.Competizione, p.Anno, a.Nazione)
SELECT mo.Competizione, mo.Anno, mo.Nazione, mo.NumOri, ma.NumArgenti,
    mb.NumBronzi, mo.NumOri+ma.NumArgenti+mb.NumBronzi AS NumMedaglie
FROM medaglieoro mo JOIN medaglieargento ma ON (mo.Competizione=ma.Competizione AND
    mo.Anno=ma.Anno AND mo.Nazione=ma.Nazione) JOIN medagliebronzo mb ON
    (mo.Competizione=mb.Competizione AND mo.Anno=mb.Anno AND mo.Nazione=mb.Nazione);
```

Esame di Teoria

Domanda 1 (9 punti).

Con riferimento alla base di dati “Medaglie”:

A. (4 punti) Esprimere in Algebra Relazionale l'interrogazione

Elencare gli atleti e le atlete francesi che hanno vinto medaglie in almeno due specialità diverse.

B. (5 punti) Esprimere, nel calcolo relazionale su tuple con dichiarazione di range, la seguente domanda:

Elencare le specialità in cui sono passati più di 20 anni tra la conquista di una qualsiasi medaglia e la successiva da parte di un* atleta italian*.

Soluzione 1.

A. Una possibile soluzione è la seguente:

$$(\rho_{P1 \leftarrow \text{PIAZZAMENTO}}(\sigma_{\text{Posizione} \leq 3}(\text{piazzamento}))) \bowtie_{\alpha} \pi_{P1.Atleta}(\sigma_{\text{Nazione} = 'Francia'}(\rho_{P2 \leftarrow \text{PIAZZAMENTO}}(\sigma_{\text{Posizione} \leq 3}(\text{piazzamento})))) \bowtie_{\beta} \text{atleta}))$$

dove

$$\alpha = P1.Atleta = P2.Atleta \wedge P1.Specialità \neq P2.Specialità$$

$$\beta = \text{atleta.Codice} = P1.Atleta$$

B. Una possibile soluzione è la seguente (sono possibili altre interpretazioni della domanda):

$$\{s.Codice, | s(SPECIALITÀ) | \exists p(\text{PIAZZAMENTO}) \exists a(\text{ATLETA}) (p.Specialità=s.Codice \wedge p.Posizione \leq 3 \wedge p.Atleta=a.Codice \wedge a.Nazione='Italia' \wedge \exists p'(\text{PIAZZAMENTO}) \exists a'(\text{ATLETA}) (p'.Specialità=s.Codice \wedge p'.Posizione \leq 3 \wedge p'.Atleta=a'.Codice \wedge a'.Nazione='Italia' \wedge (p'.Anno-p.Anno) > 20 \wedge \neg \exists p''(\text{PIAZZAMENTO}) (p''.Specialità=s.Codice \wedge p''.Posizione \leq 3 \wedge p''.Anno > p.Anno \wedge p''.Anno < p'.Anno \wedge \exists a''(\text{ATLETA}) (p''.Atleta=a''.Codice \wedge a''.Nazione='Italia')))) \}$$

Domanda 2 (8 punti).

A. (4 punti) Enunciare la definizione di equivalenza di due insiemi F e G di dipendenze funzionali utilizzando la nozione di chiusura di un insieme di attributi.

B. (4 punti) Dati: R(L, M, N, O, P, Q, R) e F = { O → M,R; N → L,M,Q; M,N → O,P; R → N,O } dire se R è in 3FN e se non lo è decomporla in relazioni in 3FN esplicitando tutti i passaggi.

Soluzione 2.

A. Dati due insiemi di dipendenze funzionali F e G, F≡G se e solo se $\forall X \rightarrow Y \in F, Y \subseteq X^+_G$ e $\forall X \rightarrow Y \in G, Y \subseteq X^+_F$

B. Per prima cosa è necessario identificare le chiavi candidate. In questo caso abbiamo tre chiavi K1={O}, K2={N}, K3={R} (il calcolo della chiusura lo dimostra). Le dipendenze funzionali sono tutte di tipo superchiave, quindi la relazione è già in 3NF.

Domanda 3 (8 punti).

Con riferimento alla base di dati “Medaglie” e i seguenti dati quantitativi:

$CARD(ATLETA) = 100\ 000 = 10^5$
 $CARD(SPECIALITÀ) = 1\ 000 = 10^3$
 $CARD(PIAZZAMENTO) = 1\ 000\ 000 = 10^6$
 $VAL(Nazione, ATLETA) = 200$
 $VAL(Anno, PIAZZAMENTO) = 100$
 $VAL(Disciplina, SPECIALITÀ) = 500$
 $VAL(Atleta, PIAZZAMENTO) = CARD(ATLETA)$
 $VAL(Specialità, PIAZZAMENTO) = CARD(SPECIALITÀ)$

disegnare gli alberi sintattici prima e dopo l'ottimizzazione logica (**utilizzando tutti i passi dell'algoritmo di ottimizzazione**) e calcolare il numero di tuple “mosse” prima e dopo l'ottimizzazione logica della seguente query:

$\sigma_{Nazione = 'Italia' \wedge Anno=2022 \wedge Disciplina='Nuoto'}((atleta \bowtie_{atleta.Codice=piazzamento.Atleta} \text{piazzamento}) \bowtie_{piazzamento.Specialità=specialità.Codice} specialità)$

Soluzione 3.

Prima dell'ottimizzazione:

- Costo $r_1 = (atleta \bowtie_{atleta.Codice=piazzamento.Atleta} \text{piazzamento})$: $CARD(ATLETA) \times CARD(PIAZZAMENTO) = 10^5 \times 10^6 = 10^{11}$
- Cardinalità $|r_1| = CARD(PIAZZAMENTO) = 10^6$ (equijoin attraverso la chiave esterna)
- Costo $r_2 = (r_1 \bowtie_{piazzamento.Specialità=specialità.Codice} specialità)$: $CARD(r_1) \times CARD(SPECIALITÀ) = 10^6 \times 10^3 = 10^9$
- Cardinalità $|r_2| = CARD(PIAZZAMENTO) = 10^6$ (equijoin attraverso la chiave esterna)
- Costo della selezione: $|r_2| = 10^6$
- Costo totale = $10^{11} + 10^9 + 10^6 \approx 10^{11}$.

Applichiamo ora l'algoritmo di ottimizzazione (primi 6 passi):

Per prima cosa, si portano le selezioni verso le foglie:

$\sigma_{Nazione = 'Italia'}(atleta) \bowtie_{atleta.Codice=piazzamento.Atleta} \sigma_{Anno='2022'}(\text{piazzamento}) \bowtie_{piazzamento.Specialità=specialità.Codice} \sigma_{Disciplina='Nuoto'}(specialità)$

Calcoliamo la variante della sequenza di join di costo minimo. Per questo occorre prima calcolare la cardinalità delle selezioni (ci servirà anche dopo):

- Tuple prodotte da $\sigma_1 = \sigma_{Nazione = 'Italia'}(atleta) = 1/VAL(Nazione, ATLETA) \times CARD(ATLETA) = 1/200 \times 10^5 = 500$
- Tuple prodotte da $\sigma_2 = \sigma_{Anno=2022}(\text{piazzamento}) = 1/VAL(Anno, PIAZZAMENTO) \times CARD(PIAZZAMENTO) = 1/100 \times 10^6 = 10^4$
- Tuple prodotte da $\sigma_3 = \sigma_{Disciplina='Nuoto'}(specialità) = 1/VAL(Disciplina, SPECIALITÀ) \times CARD(SPECIALITÀ) = 1/500 \times 1000 = 2$

Data la configurazione delle condizioni dei due join, non è possibile mettere in join σ_1 con σ_3 , quindi la configurazione ottimale è:

$\sigma_{Nazione = 'Italia'}(atleta) \bowtie_{atleta.Codice=piazzamento.Atleta} (\sigma_{Anno='2022'}(\text{piazzamento}) \bowtie_{piazzamento.Specialità=specialità.Codice} \sigma_{Disciplina='Nuoto'}(specialità))$

poiché il costo del join tra σ_2 e σ_3 (2×10^4) è minore del costo del join tra σ_1 e σ_2 (5×10^6).

Costo dopo l'ottimizzazione:

- Costo $\sigma_1 = \sigma_{Nazione = 'Italia'}(atleta) = CARD(ATLETA) = 10^5$
- Costo $\sigma_2 = \sigma_{Anno='2022'}(\text{piazzamento}) = CARD(PIAZZAMENTO) = 10^6$
- Costo $\sigma_3 = \sigma_{Disciplina='Nuoto'}(specialità) = CARD(SPECIALITÀ) = 10^3$

- Costo join $r2 = \sigma_2 \bowtie_{\text{piazzamento.Specialità=specialità.Codice}} \sigma_3 = 2 \times 10^4$ (già calcolato)
- Tuple prodotte da $r2 = \sigma_2 \bowtie_{\text{piazzamento.Specialità=specialità.Codice}} \sigma_3 = \min\{1/\text{VAL}(\text{Specialità}, \sigma_2), 1/\text{VAL}(\text{Codice}, \sigma_3)\} \times \text{CARD}(\sigma_2) \times \text{CARD}(\sigma_3)$ dove
 $\text{VAL}(\text{Specialità}, \sigma_2) = \min\{\text{VAL}(\text{Specialità}, \text{PIAZZAMENTO}), \text{CARD}(\sigma_2)\} = \min\{10^3, 10^4\} = 10^3$ e
 $\text{VAL}(\text{Codice}, \sigma_3) = \min\{\text{VAL}(\text{Codice}, \text{SPECIALITÀ}), \text{CARD}(\sigma_3)\} = \min\{10^3, 2\} = 2$
 Quindi $\text{CARD}(r2) = \min\{1/10^3, 1/2\} \times 10^4 \times 2 = 1/10^3 \times 10^4 \times 2 = 20$
- Costo join $r1 = \sigma_1 \bowtie_{\text{atleta.Codice=piazzamento.Atleta}} r2 = 500 \times 20 = 10\,000 = 10^4$
- Costo totale $= 10^5 + 10^6 + 10^3 + 2 \times 10^4 + 10^4 \approx 10^6$.

Domanda 4 (8 punti).

- Illustrare la procedura per la ripresa a freddo (cold restart/dump restore).
- Dire, giustificando la risposta, se la seguente storia è compatibile con il protocollo 2PL (lock a due fasi):
 $S = r1(x), r2(x), w1(z), w2(y), w3(x), r3(y), w2(z)$

Soluzione 4.

- Si vedano gli appunti/slide.
- E' compatibile, e una possibile storia è la seguente:
 $LS1(x), r1(x), LS2(x), r2(x), LX1(z), w1(z), UN1(x,z), LX2(y), w2(y), LX2(z), UN2(x), LX3(x), w3(x), UN2(y), LS3(y), r3(y), UN3(y), w2(z), UN2(z).$