

Matricola: _____

Cognome: _____

Nome: _____

Basi di dati

II Prova in itinere del 27 febbraio 2018 (soluzioni)

Avvertenze: è severamente vietato consultare libri e appunti.

Durata 2h

Domande di teoria

- a) (2) Si illustri l'operatore di proiezione dell'algebra relazionale (sintassi, semantica, esempio d'uso)

- b) (2) Date le due seguenti relazioni: $R1(\underline{A}, B, C, D)$ e $R2(\underline{B}, E, F^*, G)$ si scriva in algebra relazionale:

b.1) un'espressione che restituisca una relazione con un solo attributo Z contenente l'insieme dei valori distinti contenuti nell'attributo B di R1 che non sono contenuti nell'attributo E di R2;

b.2) un'espressione ottimizzata che contenga **solo** un join naturale, una selezione su R2 e una proiezione su R1 e produca come risultato le tuple t di R2 tali che $t[G] < 6$ e $t[E] = 0$ e per le quali esista una tupla t' di R1 dove $t[B] = t'[B]$.

ALGEBRA RELAZIONALE (è obbligatorio rispondere ai quesiti 1.a, 1.b, 1.c e 4.a)

Dato il seguente schema relazionale contenente le informazioni che descrivono una serie di tornei di tennis svolti o programmati nel 2018:

TORNEO(Nome, Città, Stato);

GIORNATA(Torneo, NumeroGiornata, Data)

TENNISTA(Codice, Nome, Cognome, Età, Nazionalità)

PARTITA(Torneo, NumeroGiornata, OraInizio, Giocatore1, Giocatore2, Durata*, Vincitore*)

Vincoli d'integrità referenziale:

GIORNATA.Torneo → TORNEO

PARTITA.(Torneo, NumeroGiornata) → GIORNATA, PARTITA.Giocatore1 → TENNISTA,

PARTITA.Giocatore2 → TENNISTA, PARTITA.Vincitore → TENNISTA

Ulteriori vincoli d'integrità: NumeroGiornata ∈ {1,2,3,4,5,6,7,8,9,10,11,12,13,14,15}

1. Formulare in algebra relazionale **ottimizzata** le seguenti interrogazioni:

1.a (3) *Trovare il nome e la città dei tornei francesi dove nella prima e nella seconda giornata non sono previste partite dopo le 20.30.*

1.b (3) *Trovare il nome e il cognome dei tennisti italiani che hanno giocato almeno due partite nel torneo "T1".*

1.c (3) *Trovare le partite che si svolgono oggi o domani dopo le 12.30 nei tornei che non si tengono in Italia riportando il nome del torneo, la città dove si svolge, la data e l'ora d'inizio della partita e il cognome e la nazionalità dei due giocatori.*

2. Formulare in algebra relazionale le seguenti interrogazioni:

2.a (3) *Trovare la partita di durata massima che si è svolta in Francia riportando il nome del torneo, la città dove si è svolto, la data della partita e il cognome del vincitore.*

2.b (3) *Trovare il nome e il cognome dei tennisti che hanno partecipato (sono coinvolti in almeno una partita) a tutti i tornei svoltisi in Italia.*

3. Supponendo che le relazioni abbiano le seguenti cardinalità:

- **TORNEO:** 60 (tornei francesi = 5)
- **GIORNATA:** 240
- **TENNISTA:** 80 (tennist italiani = 10)
- **PARTITA:** 1400 (partite vinte da tennisti italiani in tornei francesi = 50)

3.a (3) calcolare la dimensione dei risultati intermedi (in termini di numero di valori) in tutti i nodi dell'albero che rappresenta la seguente interrogazione:

$$\Pi_{\{\text{Cognome, Nome, Città}\}}(\sigma_{\text{Nazionalità}='Italia' \wedge \text{Stato}='Francia'}((\text{PARTITA} \bowtie_{\text{Torneo} = \text{Nome}} \text{TORNEO}) \bowtie_{\text{Vincitore} = \text{Codice}} \Pi_{\{\text{Codice, Cognome, Nazionalità}\}}(\text{TENNISTA}))$$

3.b (3) produrre la versione ottimizzata della precedente interrogazione

3.c (3) calcolare la dimensione dei risultati intermedi (in termini di numero di valori) in tutti i nodi dell'albero che rappresenta la versione ottimizzata prodotta al punto precedente.

SQL (sono ammesse solo le parole chiave: SELECT, FROM, WHERE, EXISTS, IN, AND, OR, NOT, DISTINCT)

4. Dato lo schema relazionale sopra riportato, formulare in SQL le seguenti interrogazioni:

4.a (4) *Trovare per ogni partita del torneo di nome "T2" la partita immediatamente successiva nella stessa giornata, riportando nel risultato la giornata, l'ora di inizio e il cognome dei giocatori della prima partita e l'ora di inizio e il cognome dei giocatori della seconda partita.*

SOLUZIONI

Dato il seguente schema relazionale contenente le informazioni che descrivono una serie di tornei di tennis svolti o programmati nel 2018:

TORNEO(Nome, Città, Stato);

GIORNATA(Torneo, NumeroGiornata, Data)

TENNISTA(Codice, Nome, Cognome, Età, Nazionalità)

PARTITA(Torneo, NumeroGiornata, Orainizio, Giocatore1, Giocatore2, Durata*, Vincitore*)

Esercizio 1

1.a (3) *Trovare il nome e la città dei tornei francesi dove nella prima e nella seconda giornata non sono previste partite dopo le 20.30.*

$$\Pi_{\{Nome, Città\}}(\sigma_{Stato='Francia'}(TORNEO))$$

-

$$\Pi_{\{Nome, Città\}}(\sigma_{Stato='Francia'}(TORNEO)) \bowtie_{Nome = Torneo} \Pi_{\{Torneo\}}($$

$$\sigma_{(NumeroGiornata=1 \vee NumeroGiornata=2) \wedge Orainizio > 20.30} (PARTITA)))$$

1.b (3) *Trovare il nome e il cognome dei tennisti italiani che hanno giocato almeno due partite nel torneo "T1".*

$$r = \Pi_{\{NumeroGiornata, Orainizio, Codice\}}($$

$$\Pi_{\{NumeroGiornata, Orainizio, Giocatore1, Giocatore2\}}(\sigma_{Torneo='T1'} (PARTITA)) \bowtie$$

$$Giocatore1 = Codice \vee Giocatore2 = Codice$$

$$\Pi_{\{Codice\}}(\sigma_{Nazionalità='Italiana'} (TENNISTA)))$$

$$\Pi_{\{Nome, Cognome\}}(\Pi_{\{Codice\}}(r \bowtie_{\rho_{x \rightarrow x}(r)})) \bowtie TENNISTA)$$

P: (NumeroGiornata \neq NumeroGiornata \vee Orainizio \neq Orainizio) \wedge Codice = Codice

1.c (4) *Trovare le partite che si svolgono oggi o domani dopo le 12.30 nei tornei che non si tengono in Italia riportando il nome del torneo, la città dove si svolge, la data e l'ora d'inizio della partita e il cognome e la nazionalità dei due giocatori.*

$$\Pi_{\{Torneo, Città, Orainizio, Cognome, Nazionalità, Cognome2, Nazionalità2\}}($$

$$((((\Pi_{\{Torneo, NumeroGiornata, Orainizio, Giocatore1, Giocatore2\}}(\sigma_{Orainizio > 12.30} (PARTITA))$$

$$\bowtie \sigma_{Data=27/2/18 \vee Data=28/2/18} (GIORNATA))$$

$$\bowtie \Pi_{\{Torneo, Città\}}(\sigma_{Stato \neq 'Italia'} (TORNEO)))$$

$$\bowtie \Pi_{\{Codice, Cognome, Nazionalità\}} (TENNISTA))$$

Giocatore1 = Codice

$$\bowtie \Pi_{\{Codice2, Cognome2, Nazionalità2\}} (\rho_{x \rightarrow x2} (TENNISTA))))$$

Giocatore2 = Codice2

2.a (3) *Trovare la partita di durata massima che si è svolta in Francia riportando il nome del torneo, la città dove si è svolto, la data della partita e il cognome del vincitore.*

$$r = \Pi_{\{\text{Torneo}, \text{NumeroGiornata}, \text{Città}, \text{Vincitore}, \text{Durata}\}} \left(\Pi_{\{\text{Torneo}, \text{NumeroGiornata}, \text{Vincitore}, \text{Durata}\}} (\text{PARTITA}) \bowtie_{\text{Torneo} = \text{Nome}} \Pi_{\{\text{Nome}, \text{Città}\}} (\sigma_{\text{Stato} = \text{'Francia'}} (\text{TORNEO})) \right)$$

$$\Pi_{\{\text{Torneo}, \text{NumeroGiornata}, \text{Città}, \text{Cognome}, \text{Data}\}} \left(((r - \Pi_{\{\text{Torneo}, \text{NumeroGiornata}, \text{Città}, \text{Vincitore}, \text{Durata}\}} (r \bowtie_{\text{durata} < \underline{\text{durata}}} \rho_{x \rightarrow x} (r))) \bowtie_{\text{Vincitore} = \text{Codice}} \text{GIORNATA}) \bowtie \text{TENNISTA} \right)$$

2.b (3) *Trovare il nome e il cognome dei tennisti che hanno partecipato (sono coinvolti in almeno una partita) a tutti i tornei svoltisi in Italia.*

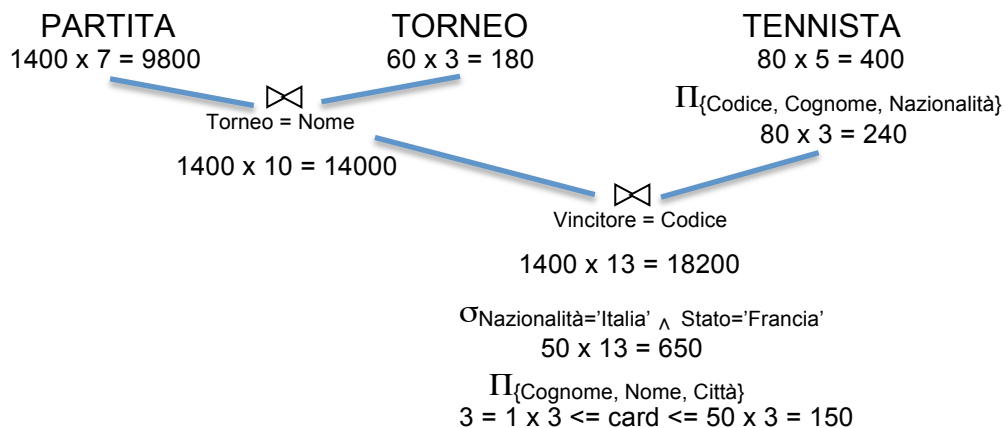
$$r = \Pi_{\{\text{Nome}\}} (\sigma_{\text{Stato} = \text{'Italia'}} (\text{TORNEO})) \bowtie \Pi_{\{\text{Codice}\}} (\text{TENNISTA})$$

$$\Pi_{\{\text{Nome}, \text{Cognome}\}} \left((\Pi_{\{\text{Codice}\}} (\text{TENNISTA}) - \Pi_{\{\text{Codice}\}} (r - \Pi_{\{\text{Nome}, \text{Codice}\}} ((\Pi_{\{\text{Codice}\}} (\text{TENNISTA}) \bowtie \text{PARTITA})_{\text{Giocatore1} = \text{Codice} \vee \text{Giocatore2} = \text{Codice}})) \bowtie_{\text{Torneo} = \text{Nome}} \sigma_{\text{Stato} = \text{'Italia'}} (\text{TORNEO}))) \bowtie \text{TENNISTA} \right)$$

3. Supponendo che le relazioni abbiano le seguenti cardinalità:

- **TORNEO**: 60 (tornei francesi = 5)
- **GIORNATA**: 240
- **TENNISTA**: 80 (tennististi italiani = 10)
- **PARTITA**: 1400 (partite vinte da tennisti italiani in tornei francesi = 50)

3.a (3) calcolare la dimensione dei risultati intermedi (in termini di numero di valori) in tutti i nodi dell'albero che rappresenta la seguente interrogazione:

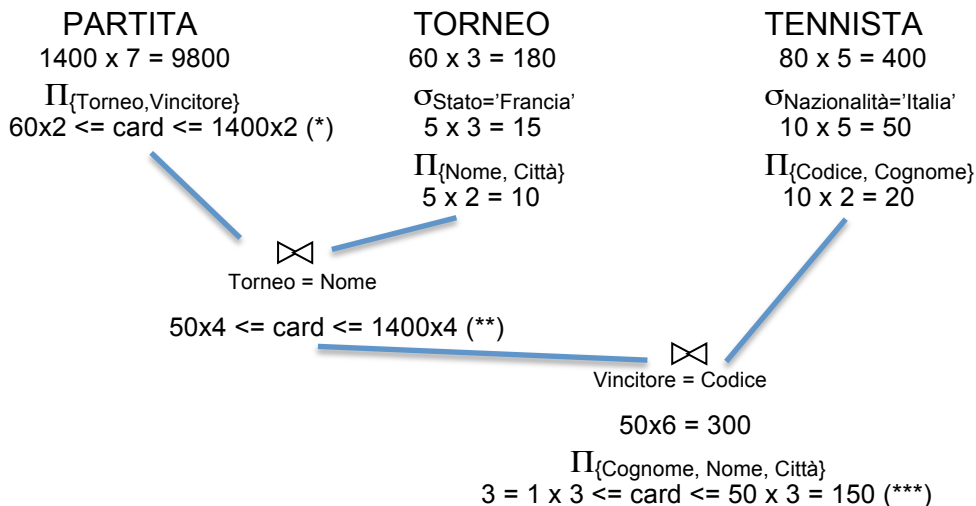
$$\Pi_{\{\text{Cognome, Nome, Città}\}}(\sigma_{\text{Nazionalità}='Italia' \wedge \text{Stato}='Francia'}((\text{PARTITA} \bowtie_{\text{Torneo} = \text{Nome}} \text{TORNEO}) \bowtie_{\text{Vincitore} = \text{Codice}} \Pi_{\{\text{Codice, Cognome, Nazionalità}\}}(\text{TENNISTA}))$$


Il valore minimo (1 tupla) si verifica quando le 50 partite sono state vinte nello stesso torneo dallo stesso giocatore, il valore massimo (50 tuple) quando tutte le combinazioni (torneo,vincitore) sono diverse.

3.b (3) produrre la versione ottimizzata della precedente interrogazione

$$\Pi_{\{\text{Cognome, Nome, Città}\}} \left(\left(\Pi_{\{\text{Torneo, Vincitore}\}}(\text{PARTITA}) \bowtie_{\text{Torneo} = \text{Nome}} \Pi_{\{\text{Nome, Città}\}}(\sigma_{\text{Stato}='Francia'}(\text{TORNEO})) \right) \bowtie_{\text{Vincitore} = \text{Codice}} \Pi_{\{\text{Codice, Cognome}\}}(\sigma_{\text{Nazionalità}='Italia'}(\text{TENNISTA})) \right)$$

3.c (3) calcolare la dimensione dei risultati intermedi (in termini di numero di valori) in tutti i nodi dell'albero che rappresenta la versione ottimizzata prodotta al punto precedente.



(*): Il valore minimo (60 tuple) si verifica quando nei 60 tornei tutte le partite con lo stesso vincitore (abbastanza improbabile), il valore massimo (1400 tuple) quando ho tutte le partite vinte da un giocatore diverso.

(**): il valore minimo (50 tuple) si verifica quando in totale le partite con vincitore in tornei francesi sono tutte e sole quelle vinte da giocatori italiani, il valore massimo (1400 tuple) quando tutte le partite si sono svolte in tornei francesi

(***): Il valore minimo (1 tupla) si verifica quando le 50 partite sono state vinte nello stesso torneo dallo stesso giocatore, il valore massimo (50 tuple) quando tutte le combinazioni (torneo, vincitore) sono diverse.

4. Dato lo schema relazionale sopra riportato, formulare in SQL le seguenti interrogazioni:

4.a (4) *Trovare per ogni partita del torneo di nome "T2" la partita immediatamente successiva nella stessa giornata, riportando nel risultato la giornata, l'ora di inizio e il cognome dei giocatori della prima partita e l'ora di inizio e il cognome dei giocatori della seconda partita.*

```
SELECT P1.NumeroGiornata, P1.Oralnizio as OraPartita, P2.Oralnizio as OraPartitaSuccessiva,
       T1.Cognome as Partita_Giocatore1, T2.Cognome as Partita_Giocatore2,
       T3.Cognome as PartitaSuccessiva_Giocatore1, T4.Cognome as PartitaSuccessiva_Giocatore2
FROM PARTITA as P1, PARTITA as P2, TENNISTA as T1, TENNISTA as T2,
     TENNISTA as T3, TENNISTA as T4
WHERE P1.NumeroGiornata = P2.NumeroGiornata AND P2.Oralnizio > P1.Oralnizio AND
      P1.Giocatore1 = T1.Codice AND P1.Giocatore2 = T2.Codice AND
      P2.Giocatore1 = T3.Codice AND P2.Giocatore2 = T4.Codice AND
      P1.Torneo = 'T2' AND P2.Torneo = 'T2' AND
      NOT EXISTS (SELECT 1 FROM PARTITA P3
                   WHERE P3.NumeroGiornata = P1.NumeroGiornata AND P3.Torneo = 'T2' AND
                        P3.Oralnizio > P1.Oralnizio AND P3.Oralnizio < P2.Oralnizio)
```