

Cognome e nome: _____

Matricola: _____

Esame di SQL (rispondere su fogli separati rispetto alla parte di teoria)

Le seguenti relazioni definiscono una base di dati “NerdFlix” per gestire una piattaforma di streaming online di serie TV. Gli attributi sottolineati sono le chiavi primarie delle relazioni.

SERIETV(NomeSerie, Produttore, Nazione, Genere)

EPISODIO(NomeSerie, NumStagione, NumEpisodio, TitoloEpisodio, Anno)

AUDIO(NomeSerie, NumStagione, NumEpisodio, Lingua, AdattatoreDialoghi*, Originale)

SOTTOTITOLI(NomeSerie, NumStagione, NumEpisodio, Lingua)

Vincoli di integrità referenziale:

EPISODIO(NomeSerie) referencia SERIETV(NomeSerie),

AUDIO(NomeSerie,NumStagione,NumEpisodio) referencia

EPISODIO(NomeSerie,NumStagione,NumEpisodio),

SOTTOTITOLI(NomeSerie,NumNumStagione,NumEpisodio) referencia

EPISODIO(NomeSerie,NumStagione,NumEpisodio).

“Genere” può assumere i valori “Animazione”, “Dramma”, “Fantascienza” e “Commedia”. “Originale” può assumere i valori True (se l’audio è quello della lingua originale) e False (se si tratta di un doppiaggio). I rimanenti attributi sono autoesplicativi. AdattatoreDialoghi può essere NULL.

Con riferimento alla base di dati "NerdFlix" esprimere in SQL le seguenti interrogazioni.

Domanda 1 (bassa complessità).

Trovare le serie TV giapponesi di genere “Animazione” con dialoghi adattati da “Gualtiero Cannarsi”.

Soluzione 1.

```
SELECT DISTINCT s.NomeSerie
FROM serietv s JOIN episodio e ON (s.NomeSerie=e.NomeSerie)
      JOIN audio a ON (a.NomeSerie=e.NomeSerie AND a.NumStagione=e.NumStagione
                        AND a.NumEpisodio=e.NumEpisodio)
WHERE s.Genere='Animazione' AND s.Nazione='Giappone'
      AND a.AdattatoreDialoghi='Gualtiero Cannarsi';
```

Domanda 2 (media complessità).

Mostrare le serie TV di cui sono state prodotte almeno tre stagioni con sottotitoli in romeno nel periodo 2015-2019.

Soluzione 2.

```
SELECT e.NomeSerie
FROM episodio e JOIN sottotitoli s ON (e.NomeSerie=s.NomeSerie
      AND e.NumStagione=s.NumStagione AND e.NumEpisodio=s.NumEpisodio)
WHERE s.Lingua='Romeno' AND e.Anno>=2015 AND e.Anno<=2019
GROUP BY e.NomeSerie
HAVING COUNT(DISTINCT e.NumStagione)>=3;
```


Esame di Teoria (rispondere su fogli separati rispetto alla parte di SQL)

Domanda 1 (8 punti).

Con riferimento alla base di dati “NerdFlix”:

A. Esprimere in Algebra Relazionale l'interrogazione

Elencare le serie TV doppiate^(*) in almeno due lingue^() (oltre all'originale).**

^(*) è sufficiente che esista il doppiaggio di un solo episodio della serie

^(**) il doppiaggio può riguardare anche episodi diversi della stessa serie

B. Esprimere, nel calcolo dei predicati su tuple con dichiarazione di range, la seguente domanda:

Elencare le serie TV drammatiche (nome e nazione) per cui, di ogni episodio, esiste solo l'audio in lingua originale.

Soluzione 1.

A. Una possibile soluzione (in algebra relazionale) è la seguente:

$$\pi_{A1.NomeSerie} \left(\sigma_{A1.Originale=False \wedge A2.Originale=False} \left(\rho_{A1 \leftarrow AUDIO}(audio) \bowtie_{\theta_1} \rho_{A2 \leftarrow AUDIO}(audio) \right) \right)$$

dove $\theta_1 := A1.NomeSerie = A2.NomeSerie$

B. Una possibile soluzione è la seguente:

$$\{s.NomeSerie, s.Nazione \mid s(SERIETV) \mid s.Genere='Dramma' \wedge \\ \neg \exists a(AUDIO)(a.NomeSerie=s.NomeSerie \wedge a.Originale=False)\}$$

Domanda 2 (9 punti).

A. Riportare la definizione di chiusura di un insieme di attributi.

B. Dire se i seguenti due insiemi di dipendenze funzionali sono equivalenti, dimostrandolo:

$$F1 = \{A \rightarrow CE, C \rightarrow AB, AE \rightarrow F\}, \quad F2 = \{AB \rightarrow CE, A \rightarrow B, C \rightarrow B, C \rightarrow AF\}$$

C. Data una relazione R(ABCDEF), dire se R è in 3NF rispetto all'insieme F2 della domanda B motivando la risposta. Se non lo è, scomporla in 3NF esplicitando tutti i passi.

Soluzione 2.

A. Vedere libro di testo o slide.

B. Applicando una qualsiasi delle definizioni di equivalenza si trova facilmente che F1 ed F2 sono equivalenti.

Il modo più semplice di procedere è di calcolare la copertura minimale di F1 ed F2 e calcolare la chiusura solo per le d.f. non sintatticamente incluse in entrambi gli insiemi.

C. Per prima cosa, bisogna individuare le chiavi di R (a partire da F2). Si trovano facilmente le uniche due chiavi $K1=\{AD\}$ e $K2=\{CD\}$. Nessuna delle dipendenze funzionali è quindi di tipo superchiave, attributi primi o banale.

Calcolo l'insieme di copertura minimale.

Trasformo le d.f. mettendo un solo attributo a destra: $F2''=\{AB \rightarrow C, AB \rightarrow E, A \rightarrow B, C \rightarrow B, C \rightarrow A, C \rightarrow F\}$.

In $AB \rightarrow C$ e $AB \rightarrow E$, B è attributo estraneo (a causa di $A \rightarrow B$); il nuovo insieme è quindi:

$$F2''=\{A \rightarrow C, A \rightarrow E, A \rightarrow B, C \rightarrow B, C \rightarrow A, C \rightarrow F\}.$$

$A \rightarrow B$ è ridondante per transitività, perché $A \rightarrow C$ e $C \rightarrow B$. L'insieme di copertura minimale è quindi:

$$F2''=\{A \rightarrow C, A \rightarrow E, C \rightarrow B, C \rightarrow A, C \rightarrow F\}.$$

Raggruppo le d.f. con lo stesso antecedente e ottengo:

$$F2'' = \{A \rightarrow CE, C \rightarrow ABF\}.$$

Posso quindi scomporre R nelle relazioni R1(ACE) ($C \rightarrow A$ nella restrizione) e R2(CABF) (con $A \rightarrow C$ nella restrizione) cui devo aggiungere una relazione con la chiave AD (o CD), quindi R3(AD).

Domanda 3 (8 punti).

A. Spiegare la differenza tra ottimizzazione logica e ottimizzazione fisica di un'interrogazione.

B. Data la seguente query sulla base di dati “NerdFlix”:

$$\sigma_{audio.Lingua='Italiano' \wedge serietv.NomeSerie='Stranger Things'}((audio \bowtie episodio) \bowtie serietv)$$

dove i join sono tutti naturali, effettuare l'ottimizzazione logica sapendo che

$$CARD(SERIETV) = 1000$$

$$CARD(EPISODIO) = 50.000$$

$$CARD(AUDIO) = 100.000$$

$$VAL(Lingua, AUDIO) = 10$$

Soluzione 3.

A. Vedere libro di testo o slide.

B. Decomponendo l'AND della selezione e trasferendo le selezioni verso le foglie (passi 1 e 2 dell'algoritmo di ottimizzazione logica) ottengo:

$$(\sigma_{audio.Lingua='Italiano'}(audio) \bowtie episodio) \bowtie (\sigma_{serietv.NomeSerie='Stranger Things'}(serietv))$$

Esaminando le varianti dell'albero di parsificazione (passo 7 dell'algoritmo di ottimizzazione logica) scelgo

la variante di costo minimo: dato che $|\sigma_{audio.Lingua='Italiano'}(audio)| = \frac{1}{10} \cdot 100.000 = 10.000$,

$|episodio| = 50.000$ e, poiché NomeSerie è chiave primaria di serietv,

$|\sigma_{serietv.NomeSerie='Stranger Things'}(serietv)| = 1$, posticipiamo il join con la relazione episodio:

$$(\sigma_{audio.Lingua='Italiano'}(audio) \bowtie (\sigma_{serietv.NomeSerie='Stranger Things'}(serietv))) \bowtie episodio$$

Domanda 4 (8 punti).

A. Riportare la definizione di protocollo 2PL (two-phase lock) stretto.

Considerare la seguente storia interfogliata $S = r1(x), r1(y), r2(y), r3(x), w1(x), r2(x), w2(y)$

B. La storia S è conflict-serializzabile? Giustificare la risposta.

C. La storia S è view-serializzabile? Giustificare la risposta

D. (Caso 1) Nel caso la storia S sia view-serializzabile, dire se è anche compatibile con il 2PL stretto, giustificando la risposta.

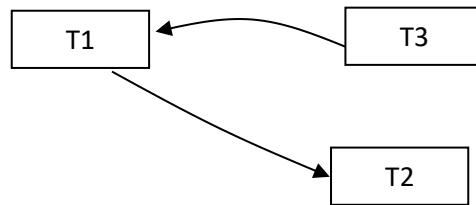
(Caso 2) Nel caso la storia S non sia view-serializzabile, è possibile togliere da S un'azione che la renda view-serializzabile? Se sì, quale?

N.B.: ovviamente, solo uno dei due casi dell'esercizio D è da svolgere.

Soluzione 4.

A. Vedere libro di testo o slide.

B e C. Il grafo dei conflitti è il seguente:



Poiché il grafo è aciclico, la storia è conflict-serializzabile e view-serializzabile.

D. (Caso 1) È compatibile con 2PL stretto perché, per esempio, è possibile aggiungere i lock ottenendo la storia seguente, compatibile con il 2PL stretto:

$S = LS1(x), r1(x), LS1(y), r1(y), LS2(y), r2(y), LS3(x), r3(x), UL3(x), LX1(x), w1(x), UL1(y), UL1(x), LS2(x), r2(x), LX2(y), w2(y), UL2(x), UL2(y)$

Si noti come i lock $LX1(x)$ e $LX2(y)$ sono semplici upgrade di lock shared precedentemente acquisiti dalle rispettive transazioni.