



Data Science e Tecnologie per le Basi di Dati

Esame del 23 Gennaio 2023

Valutazione 31,75 su un massimo di 32,00 (99%)

Domanda 1

Risposta non data

Non valutata

Questa non è una domanda dell'esame.

Puoi utilizzare l'area di testo sottostante per appunti o bozze (per esempio, per scrivere i passi intermedi di un esercizio).

Eventuali commenti/feedback per il docente possono essere scritti qui.

Il testo inserito in questo esercizio non verrà considerato in fase di correzione dell'esame.

Domanda 2

Risposta corretta

Punteggio ottenuto 1,00 su 1,00

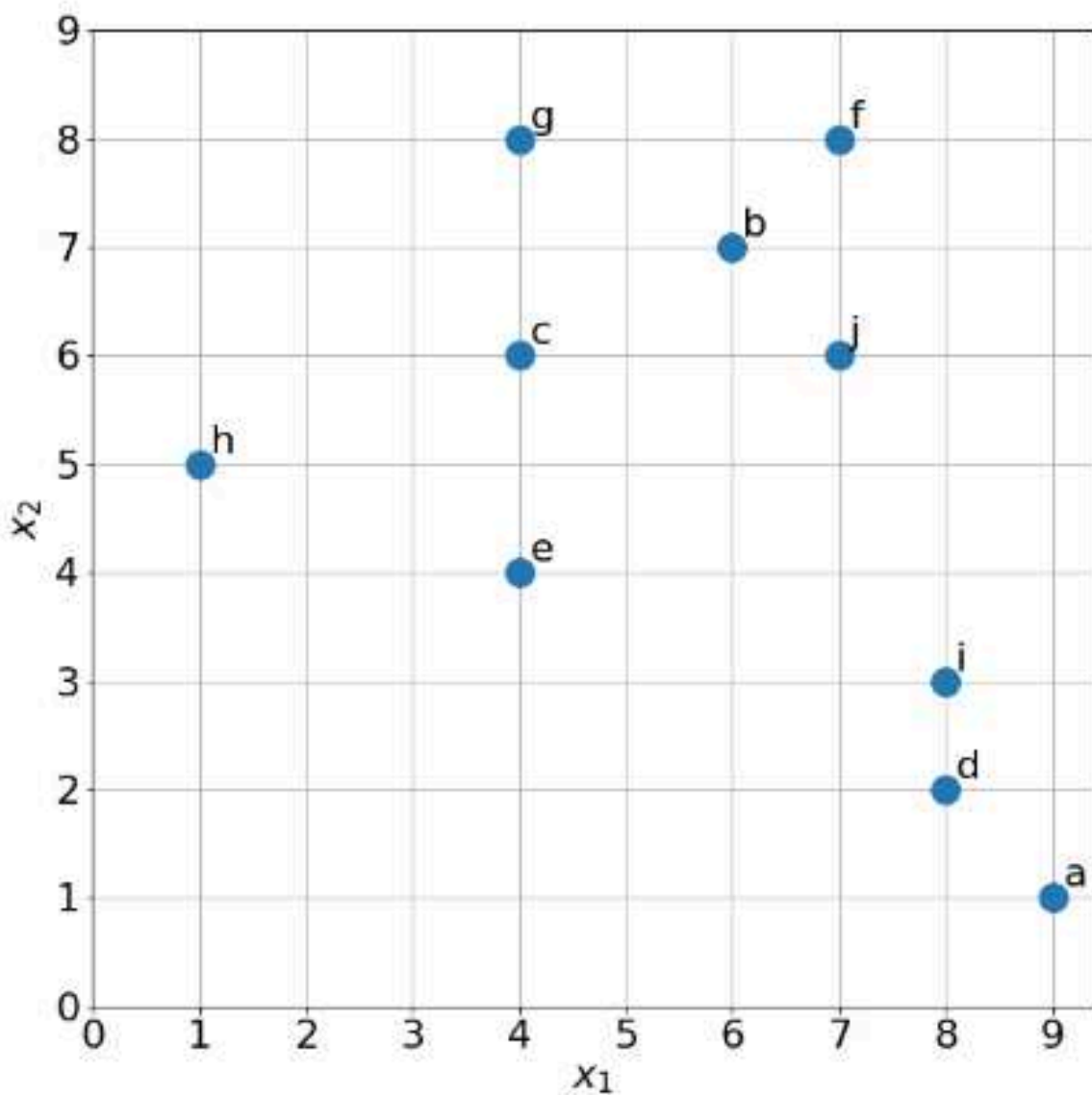
1 punto (penalità 15% per risposta sbagliata)

La distanza di Manhattan (o cityblock) tra due punti a e $b \in \mathbb{R}^n$ viene definita come:

$$d(a, b) = \sum_{k=1}^n |a_k - b_k|$$

(k rappresenta la k-esima dimensione)

DBSCAN viene applicato al seguente dataset bidimensionale. Gli iperparametri utilizzati sono $\epsilon = 2.5$, minpoints = 2 (un punto viene considerato "core" se ha almeno minpoints punti entro un vicinato di raggio ϵ). Viene utilizzata la distanza di Manhattan per calcolare la distanza fra coppie di punti.



A ciascun punto viene assegnata una categoria fra noise, border e core.
Quali categorie vengono assegnate ai punti h, b, i?

- ☐ (a) h noise, b core, i core
- ☒ (b) h noise, b core, i border ✓
- ☐ (c) h core, b core, i border
- ☐ (d) h border, b core, i core

- ☐ (e) Nessuna delle altre risposte e' corretta
- ☐ (f) h core, b border, i border
- ☐ (g) h core, b core, i core
- ☐ (h) h border, b core, i border
- ☐ (i) h border, b border, i border
- ☐ (j) h noise, b border, i border

Risposta corretta.

La risposta corretta è: h noise, b core, i border

Domanda 3

Risposta corretta

Punteggio ottenuto 1,00 su 1,00

1 punto (penalita' 15% per risposta sbagliata)

Notazione:

- $rN(V)$: read dell'oggetto V da parte della transazione N
- $wN(V)$: write dell'oggetto V da parte della transazione N

Viene dato il seguente schedule di 3 transazioni:

$S = w_0(z) w_0(y) r_2(y) r_0(y) w_2(z) r_1(z) w_2(x) r_2(x) w_1(x) w_1(z)$

S è conflict serializable perchè è conflict equivalent allo schedule seriale:

-
- ☐ (a) S non è conflict serializable perchè non è conflict equivalent a nessuno schedule seriale
 - ☒ (b) $w_0(z) w_0(y) r_0(y) r_2(y) w_2(z) w_2(x) r_2(x) r_1(z) w_1(x) w_1(z)$ ✓
 - ☐ (c) $r_1(z) w_1(x) w_1(z) r_2(y) w_2(z) w_2(x) r_2(x) w_0(z) w_0(y) r_0(y)$
 - ☐ (d) $r_2(y) w_2(z) w_2(x) r_2(x) w_0(z) w_0(y) r_0(y) r_1(z) w_1(x) w_1(z)$
 - ☐ (e) $r_2(y) w_2(z) w_2(x) r_2(x) r_1(z) w_1(x) w_1(z) w_0(z) w_0(y) r_0(y)$
 - ☐ (f) $r_1(z) w_1(x) w_1(z) w_0(z) w_0(y) r_0(y) r_2(y) w_2(z) w_2(x) r_2(x)$
 - ☐ (g) $w_0(z) w_0(y) r_0(y) r_1(z) w_1(x) w_1(z) r_2(y) w_2(z) w_2(x) r_2(x)$
 - ☐ (h) $w_0(z) w_0(y) w_2(z) r_1(z) w_2(x) r_2(x) w_1(x) w_1(z) r_2(y) r_0(y)$

Risposta corretta.

La risposta corretta è: $w_0(z) w_0(y) r_0(y) r_2(y) w_2(z) w_2(x) r_2(x) r_1(z) w_1(x) w_1(z)$

Domanda 4

Completo

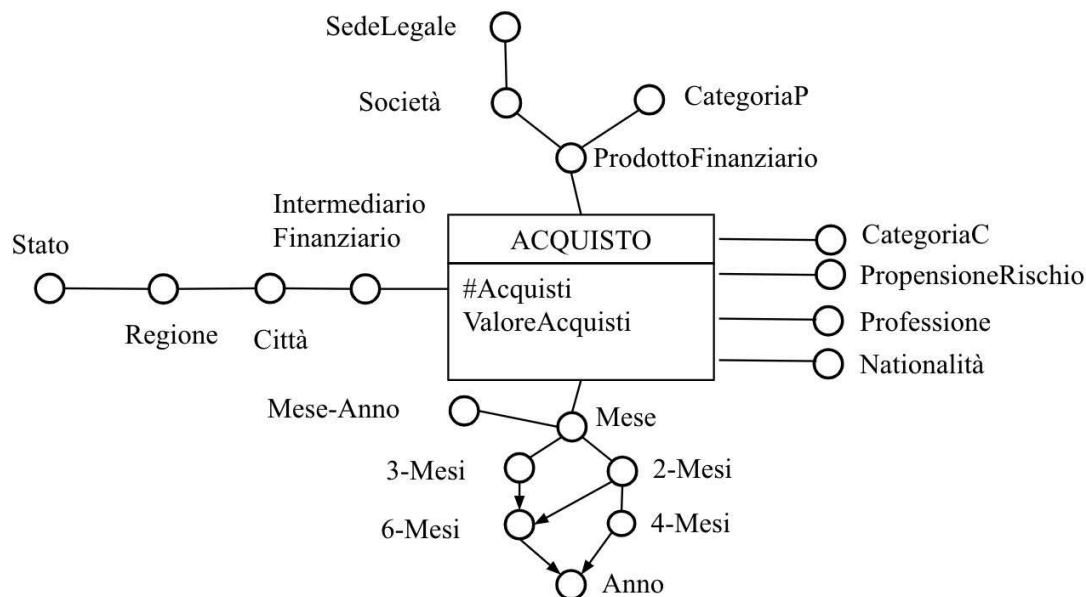
Punteggio ottenuto 10,75 su 11,00

11 punti totali (nessuna penalità per risposta sbagliata)

Il seguente data warehouse descrive l'andamento degli acquisti dei prodotti finanziari (azioni, titoli, fondi di investimento, etc) offerti da diverse società tramite intermediari finanziari (banca, broker, agenzie, etc). Ogni prodotto finanziario è emesso da una specifica società ed afferisce ad una sola

categoria (titoli di massa, contratti derivati, etc). Per ogni intermediario finanziario sono noti la città, la regione e lo stato in cui è presente. Il data warehouse memorizza la categoria del cliente (senior, medium, junior), la propensione al rischio (alta, media e bassa), la professione e la nazionalità dei clienti che hanno acquistato nel tempo i prodotti finanziari tramite gli intermediari finanziari.

Le metriche da analizzare sono il numero di prodotti acquistati (#Acquisti) e il corrispondente valore (ValoreAcquisti). Le metriche devono essere analizzate per ogni mese, bimestre, trimestre, quadrimestre, semestre, anno.



INTERMEDIARIO-FINANZIARIO(IDIntermediarioFinanziario, IntermediarioFinanziario, Città, Regione, Stato)
 PRODOTTO-FINANZIARIO(IDProdottoFinanziario, ProdottoFinanziario, Società, SedeLegale, CategoriaP)
 JUNK-CARATTERISTICHE-CLIENTE(IDJCC, CategoriaC, PropensioneRischio, Professione, Nazionalità)
 TEMPO(IDTempo, Mese, Mese-Anno, 2-Mesi, 3-Mesi, 4-Mesi, 6-Mesi, Anno)
 ACQUISTO(IDIntermediarioFinanziario, IDProdottoFinanziario, IDJCC, IDTempo, #Acquisti, ValoreAcquisti)

Dato lo schema logico precedente, scrivere nel box sottostante in SQL esteso le seguenti interrogazioni, separate da uno spazio:

• (3 punti) Query 1

Visualizzare per ogni intermediario finanziario e quadrimestre il numero totale di acquisti e il corrispondente valore. Associare ad ogni record un rango:

- che identifica la posizione del record in funzione del valore complessivo degli acquisti (1 per il valore complessivo più basso)

- che identifica la posizione del record in ordine decrescente del numero di acquisti, separatamente per stato dell'intermediario finanziario

- **(4 punti) Query 2**

Per i prodotti finanziari appartenenti alla categoria contratti derivati (CategoriaP='Contratti Derivati'), visualizzare separatamente per ogni intermediario finanziario e quadrimestre

- il valore complessivo degli acquisti,
- il valore cumulativo degli acquisti dall'inizio dell'anno al trascorrere dei quadrimestri, separatamente per intermediario finanziario,
- il valore complessivo dei prodotti indipendente dall'intermediario finanziario

- **(4 punti) Query 3**

Visualizzare per ogni prodotto finanziario, professione del cliente e trimestre

- il numero degli acquisti,
- il valore medio per acquisto,
- il valore medio mensile degli acquisti,
- la percentuale del valore degli acquisti rispetto al valore complessivo dei prodotti finanziari appartenenti alla stessa Categoria, separatamente per professione del cliente e trimestre

```
SELECT IntermediarioFinanziario, Stato, 4-Mesi,
       SUM(#Acquisti), SUM(ValoreAcquisti),
       RANK() OVER (ORDER BY SUM(ValoreAcquisti)),
       RANK() OVER (PARTITION BY Stato ORDER BY SUM(#Acquisti) DESC)
FROM INTERMEDIARIO-FINANZIARIO IF, TEMPO T, ACQUISTO A
WHERE A.IDIntermediarioFinanziario = IF.IDIntermediarioFinanziario AND A.IDTempo =
T.IDTempo
GROUP BY IntermediarioFinanziario, Stato, 4-Mesi;
```

-- Interpretazione personale: La seconda rank fa una classifica separata per ogni stato.
 -- Se si volessero separare anche i quadrimestri bisognerebbe fare "PARTITION BY Stato, 4-Mesi"

```
SELECT IntermediarioFinanziario, 4-Mesi, Anno,
       SUM(ValoreAcquisti),
       SUM(SUM(ValoreAcquisti)) OVER (PARTITION BY Anno, IntermediarioFinanziario ORDER
BY 4-Mesi),
       SUM(SUM(ValoreAcquisti)) OVER (PARTITION BY 4-Mesi)
FROM PRODOTTO-FINANZIARIO PF, INTERMEDIARIO-FINANZIARIO IF, TEMPO T,
ACQUISTO A
WHERE A.IDProdottoFinanziario = PF.IDProdottoFinanziario AND A.IDIntermediarioFinanziario =
IF.IDIntermediarioFinanziario AND
       A.IDTempo = T.IDTempo AND CategoriaP = 'Contratti Derivati'
GROUP BY IntermediarioFinanziario, 4-Mesi, Anno;
```

```

SELECT ProdottoFinanziario, CategoriaP, Professione, 3-Mesi,
    SUM(#Acquisti),
    SUM(ValoreAcquisti) / SUM(#Acquisti),
    SUM(ValoreAcquisti) / COUNT(DISTINCT Mese),
    100 * SUM(ValoreAcquisti) / SUM(SUM(ValoreAcquisti)) OVER (PARTITION BY CategoriaP,
Professione, 3-Mesi)
FROM PRODOTTO-FINANZIARIO PF, JUNK-CARATTERISTICHE-CLIENTE JCC, TEMPO T,
ACQUISTO A
WHERE A.IDProdottoFinanziario = PF.IDProdottoFinanziario AND A.IDJCC = JCC.IDJCC AND
A.IDTempo = T.IDTempo
GROUP BY ProdottoFinanziario, CategoriaP, Professione, 3-Mesi;

```

Commento:

Query 2 Prima finestra di calcolo manca la finestra di aggregazione (ROWS UNBOUNDED PRECEDING) -0,25)

Domanda 5

Risposta corretta

Punteggio ottenuto 3,00 su 3,00

3 punti totali (penalità 15% per risposta sbagliata)

Si vogliono analizzare le informazioni relative alle iscrizioni alle scuole presenti sul territorio italiano. Ciascuna scuola offre alcuni laboratori (ad esempio per attività teatrale o musicale) e alcuni servizi a supporto degli studenti (ad esempio servizio mensa o supporto pomeridiano allo studio).

Gli alunni possono presentare una certificazione per un bisogno educativo speciale (ad esempio DSA o BSE). Il numero di iscritti nelle scuole potrebbe variare durante l'anno a causa di eventuali cambi di istituto o arrivi da altri istituti.

Si vuole analizzare (1) la percentuale di alunni presenti rispetto al numero di alunni iscritti e (2) il rapporto tra il numero di alunni presenti e il numero di docenti che svolgono lezione in base a:

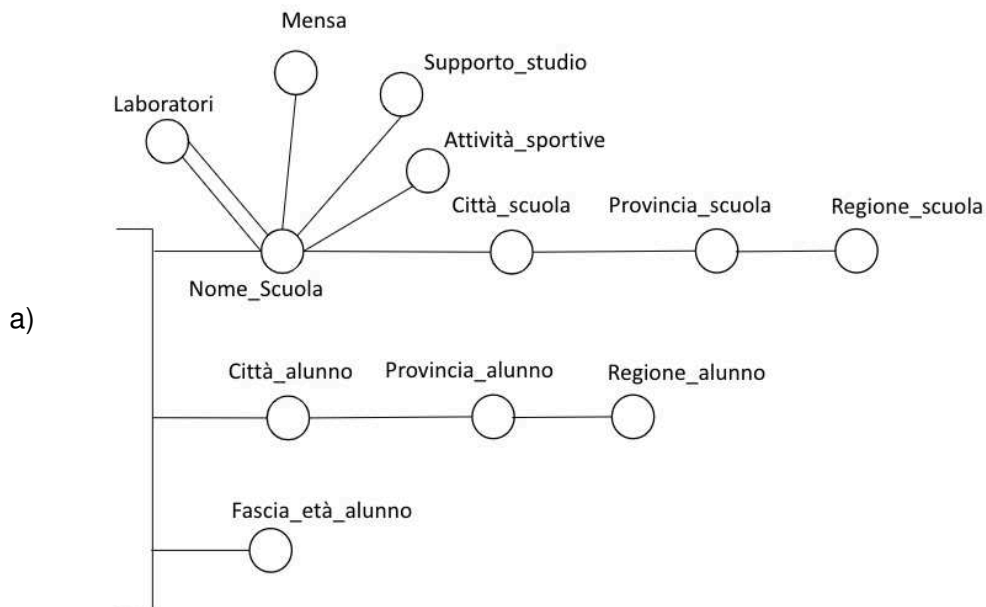
- genere dell'allievo
- data, settimana, mese, mese dell'anno, quadrimestre, semestre, anno in cui l'alunno era presente a scuola e fascia oraria di presenza a scuola (un valore tra i seguenti: 8:00-9:59, 10:00-12:59, 13:00-17:00)
- nome scuola, caratterizzata dalla sua collocazione geografica (espressa in termini di città, provincia, regione) e da tutti i servizi offerti dalla scuola (ad esempio mensa, supporto allo studio, corso pomeridiano per attività sportive, etc.). Ogni scuola è anche caratterizzata dai laboratori offerti dalla scuola (uno o più valori tra i seguenti: laboratorio STEM, laboratorio di

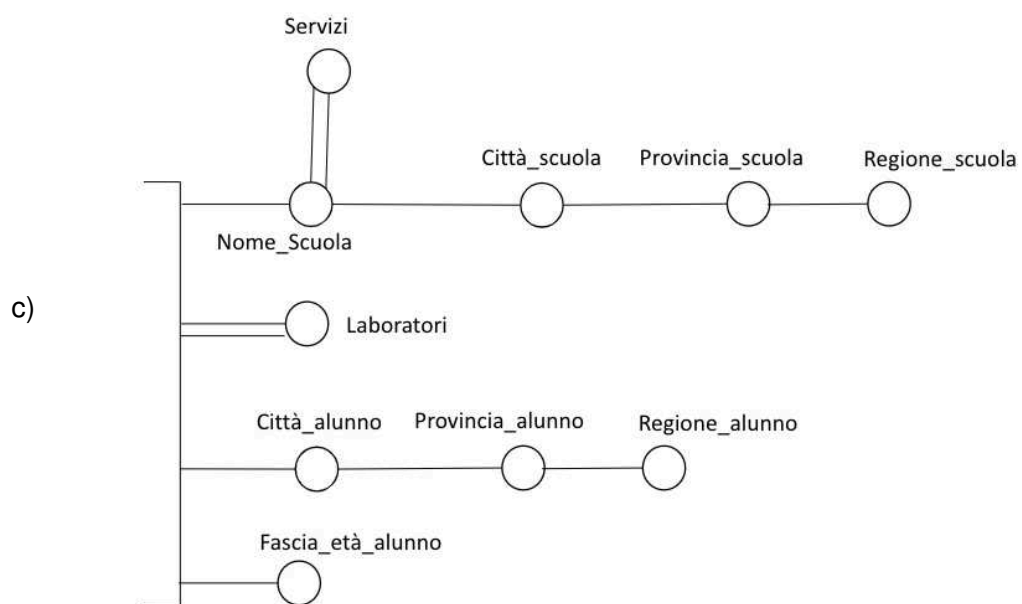
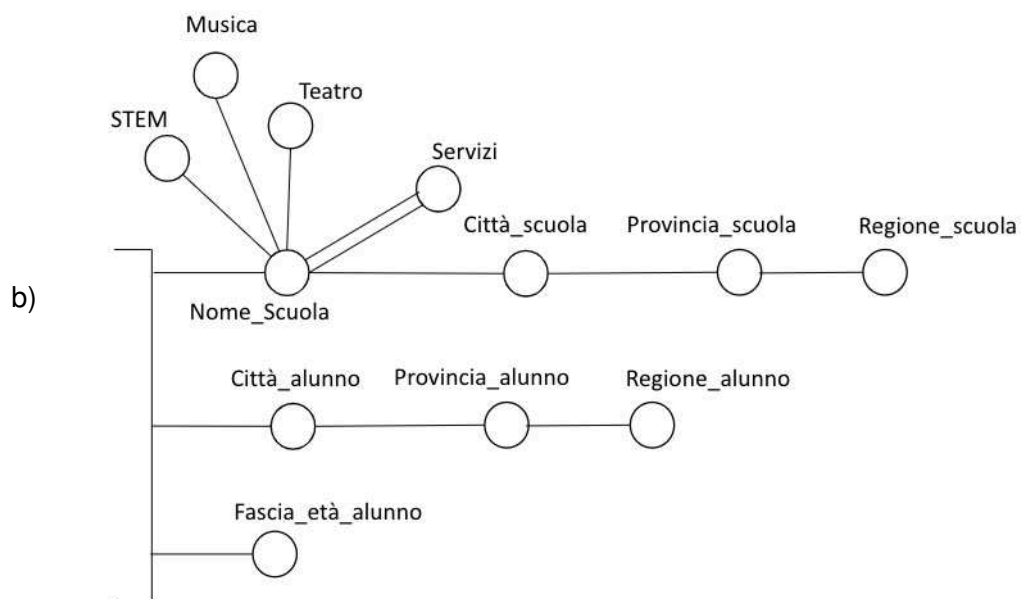
- teatro, laboratorio di musica)
- città, provincia e regione dell'alunno
- certificazione del bisogno educativo eventualmente presentata dall'allievo (un valore tra i seguenti: DSA, BSE, nessuno)
- fascia d'età dello studente (un valore tra i seguenti: tra 6 e 9 anni, tra 10 e 14 anni, più di 14 anni)

SCHEMA CONCETTUALE 1

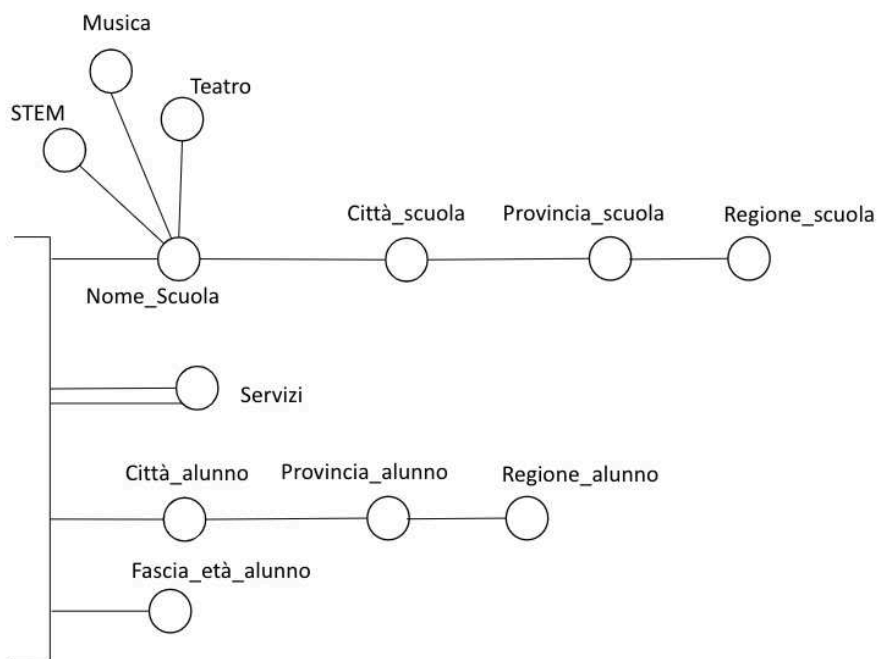
1 punto (penalità 15% per risposta sbagliata)

Selezionare, tra le dimensioni proposte di seguito, quelle che soddisfano le richieste descritte nelle specifiche del problema.

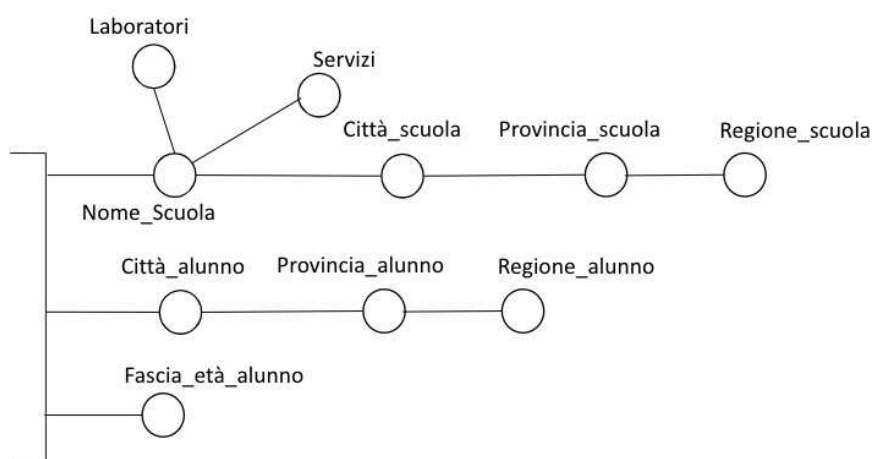




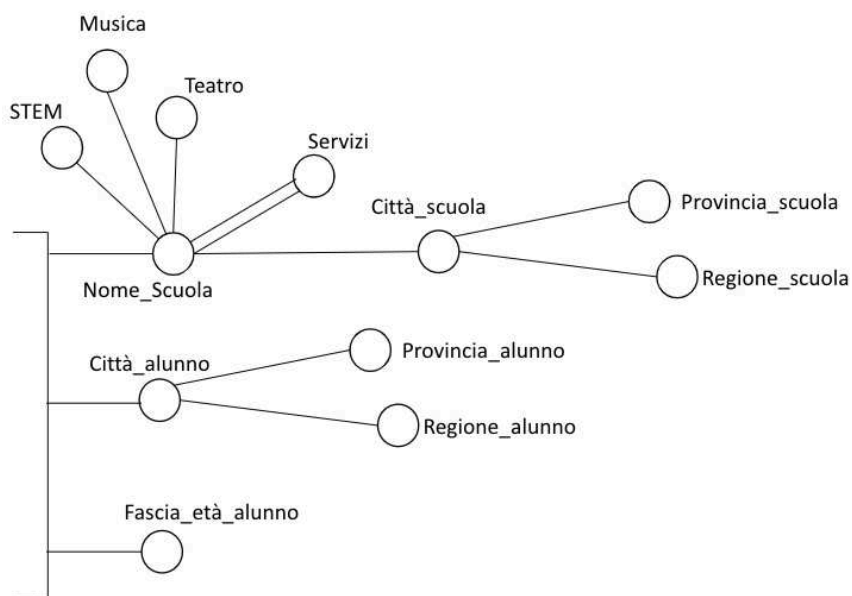
d)



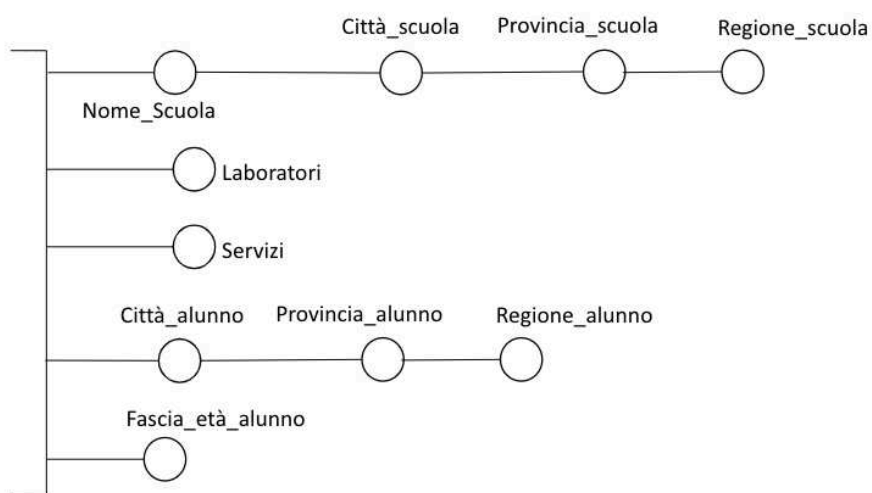
e)



f)



g)



- ☐ c
 ☐ a
 ☐ e
 ☐ g
 ☒ b ✓
 ☐ d
 ☐ f

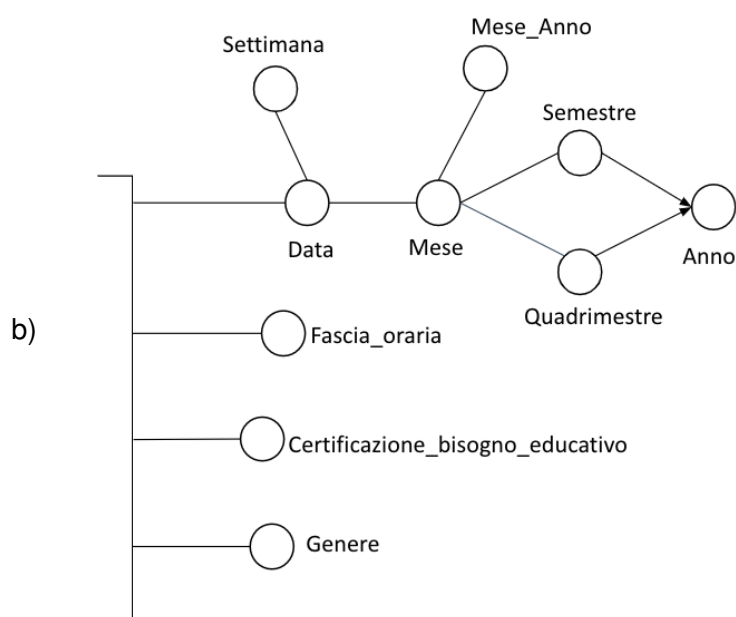
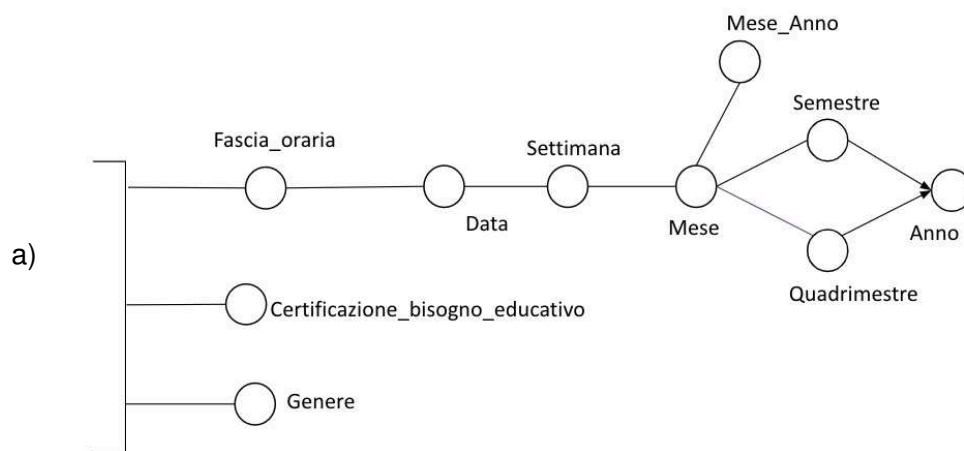
Punteggio ottenuto 1,00 su 1,00

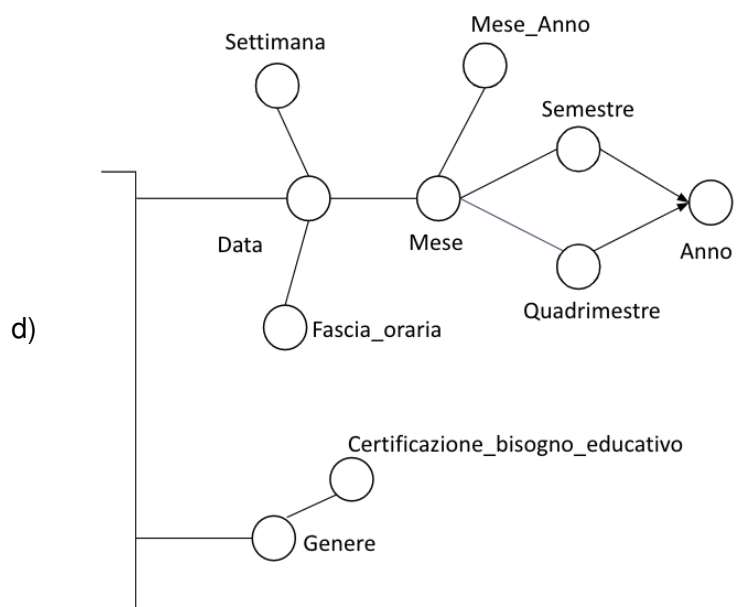
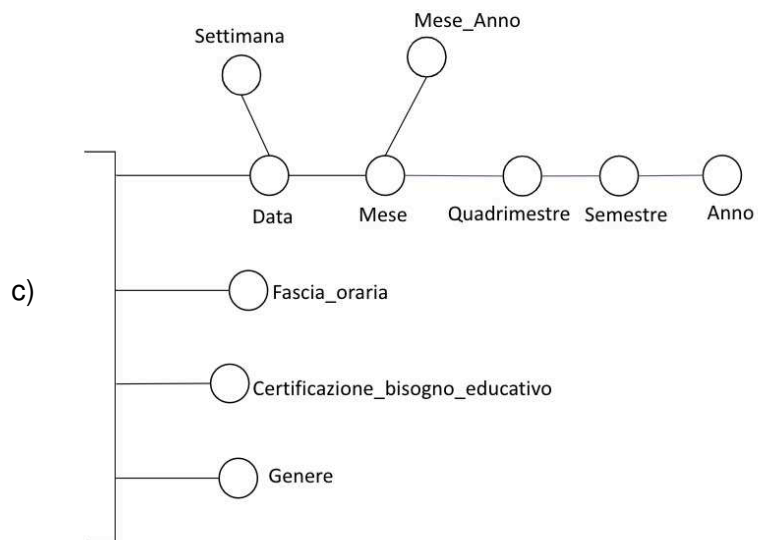
La risposta corretta è: b

Schema Concettuale 2

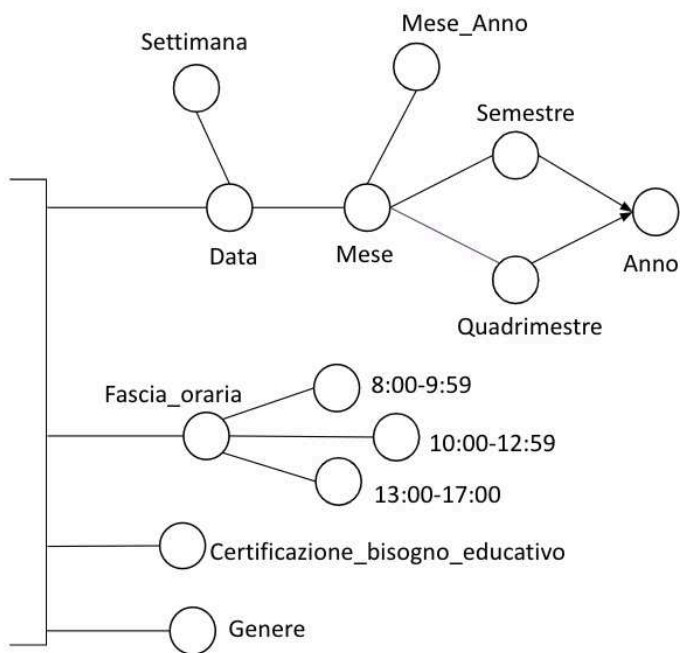
1 punto (penalità 15% per risposta sbagliata)

Selezionare, tra le dimensioni proposte di seguito, quelle che soddisfano le richieste descritte nelle specifiche del problema.

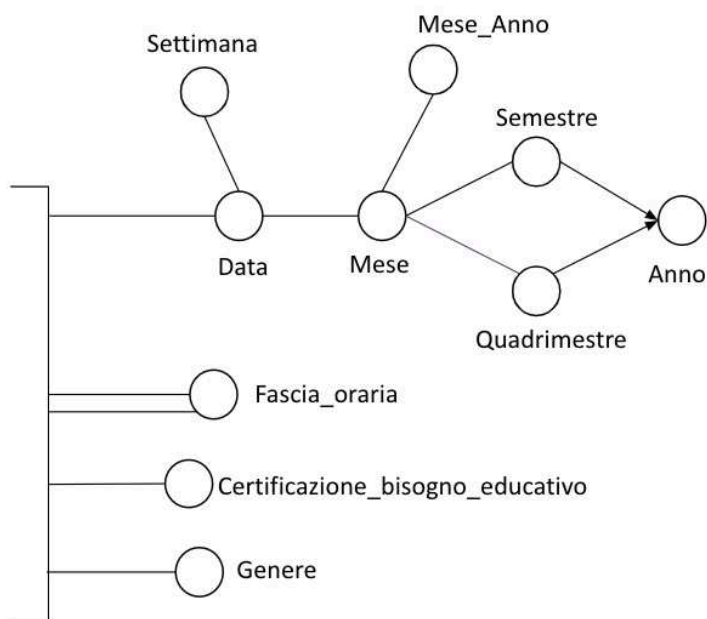


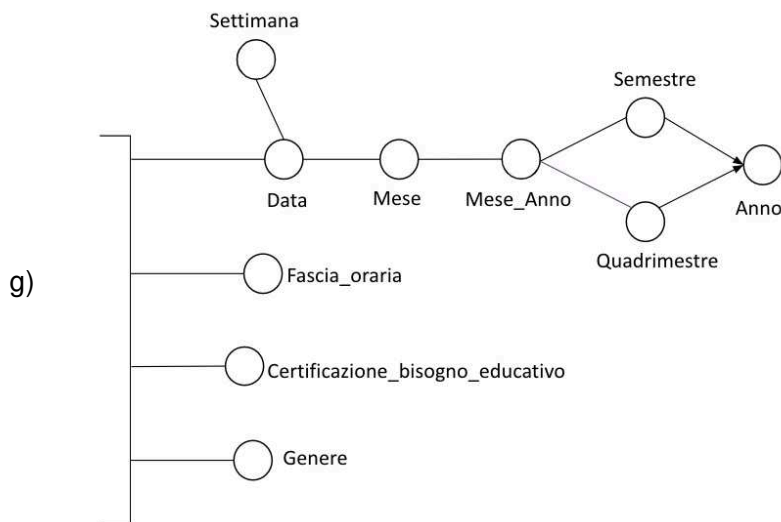


e)



f)





- ☒ b ✓
 ☐ a
 ☐ g
 ☐ f
 ☐ d
 ☐ c
 ☐ e

Punteggio ottenuto 1,00 su 1,00

La risposta corretta è: b

MISURE

1 punto (penalità 15% per risposta sbagliata)

Selezionare l'insieme di attributi necessari per modellare correttamente le misure nella tabella dei fatti richieste dalle specifiche.

- ☐ Numero totale di alunni presenti, Numero totale di ore di lezione, Numero totale scuole
- ☐ Numero totale di certificazioni, Numero totale alunni iscritti, Numero totale di docenti che svolgono lezione
- ☒ Numero totale di alunni presenti, Numero totale alunni iscritti, Numero totale di docenti che svolgono lezione ✓
- ☐ Numero totale di alunni presenti, Numero totale di docenti che svolgono lezione, Numero totale di classi
- ☐ Numero totale scuole, Numero totale di classi, Numero totale di ore di lezione
- ☐ Numero totale di alunni presenti, Numero totale alunni iscritti, Numero medio di alunni presenti
- ☐ Numero totale alunni iscritti, Numero totale scuole, Numero medio di alunni presenti
- ☐ Numero totale di ore di lezione, Numero totale di certificazioni, Numero totale di docenti che svolgono lezione

Punteggio ottenuto 1,00 su 1,00

La risposta corretta è: Numero totale di alunni presenti, Numero totale alunni iscritti, Numero totale di docenti che svolgono lezione

- 1) La risposta corretta è : b
- 2) La risposta corretta è : b
- 3) La risposta corretta è : Numero totale di alunni presenti, Numero totale alunni iscritti, Numero totale di docenti che svolgono lezione

Domanda 6

Risposta corretta

Punteggio ottenuto 5,00 su 5,00

5 punti totali (penalità -15% per ogni risposta sbagliata)

Sono date le seguenti tabelle:

Film(CodF, Titolo, CodR, CodC, Genere, Anno, RatingMedio)
Regista(CodR, Nome, Cognome, Nazionalità, DataNascita)
CasaDistribuzione(CodC, NomeCasa, DataFondazione, AreaSedeCentrale)
Utente(CodU, Nome, Cognome, Sottoscrizione)
Visualizzazione(CodF, CodU, DataUltimaVisualizzazione, Rating)

Sono date le seguenti cardinalità:

- $\text{card}(\text{FILM}) = 3 \cdot 10^6$ tuple,
 - Valori distinti di Genere = 20
 - Fattore di riduzione di $\text{RatingMedio} > 4.5$ 1/20
- $\text{card}(\text{REGISTA}) = 15 \cdot 10^4$ tuple,
 - Valori distinti di Nazionalità = 100
 - $\text{MIN}(\text{DataNascita}) = 1/1/1910$, $\text{MAX}(\text{DataNascita}) = 31/12/1999$
- $\text{card}(\text{CASA-DISTRIBUZIONE}) = 10^2$ tuple,
 - Valori distinti di AreaSedeCentrale = 10
- $\text{card}(\text{UTENTE}) = 3 \cdot 10^7$ tuple,
 - $\text{MIN}(\text{DataNascita}) = 1/1/1940$, $\text{MAX}(\text{DataNascita}) = 31/12/2009$

- Valori distinti di Sottoscrizione = 4
- $\text{card}(\text{VISUALIZZAZIONE}) = 12 \cdot 10^7$ tuple,

Inoltre, sono dati i seguenti fattori di riduzione per le clausole having:

Having COUNT(*) > 2 = 1/5

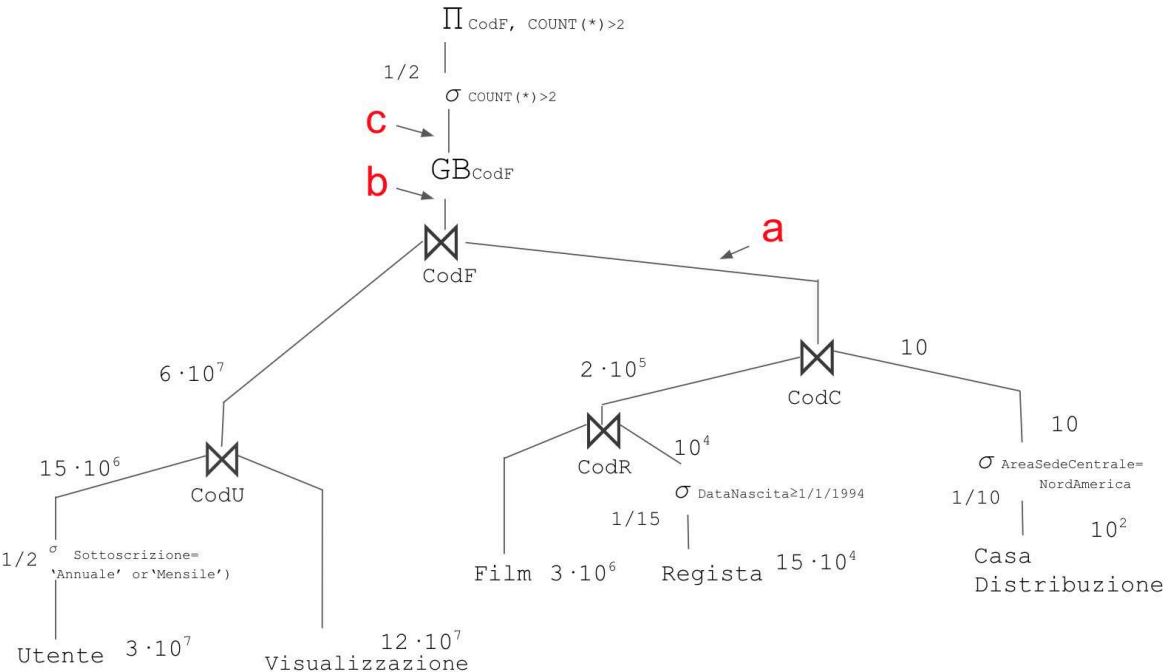
Si consideri la seguente query:

```
select F.CodF, COUNT(*)
from Utente U, Visualizzazione V, Film F, Regista R, CasaDistribuzione C
where V.CodU=U.CodU and F.CodR=R.CodR
and F.CodC=C.CodC and V.CodF=F.CodF
and (U.Sottoscrizione='Annuale' or U.Sottoscrizione='Mensile')
and C.AreaSedeCentrale='NordAmerica'
and R.DataNascita>=1/1/1994
group by F.CodF
having COUNT(*)>2)
```

Cardinalita'

(1.5 punti, penalità -15% per ogni risposta sbagliata)

La figura sottostante rappresenta il query tree per la query precedente.



Seleziona la risposta corretta per la cardinalita' di **(a)**:

- ☐ $2 \cdot 10^5$ ☐ 10^3 ☒ $2 \cdot 10^4$ ✓ ☐ $5 \cdot 10^3$

Punteggio ottenuto 5,00 su 5,00

La risposta corretta è: $2 \cdot 10^4$

Seleziona la risposta corretta per la cardinalita' di **(b)**:

- ☐ $2 \cdot 10^4$ ☐ 10^6 ☐ $6 \cdot 10^7$ ☒ $4 \cdot 10^5$ ✓

Punteggio ottenuto 5,00 su 5,00

La risposta corretta è: $4 \cdot 10^5$

Seleziona la risposta corretta per la cardinalita' di **(c)**:

- ☐ $2 \cdot 10^5$ ☒ $2 \cdot 10^4$ ✓ ☐ $5 \cdot 10^3$ ☐ 10^3

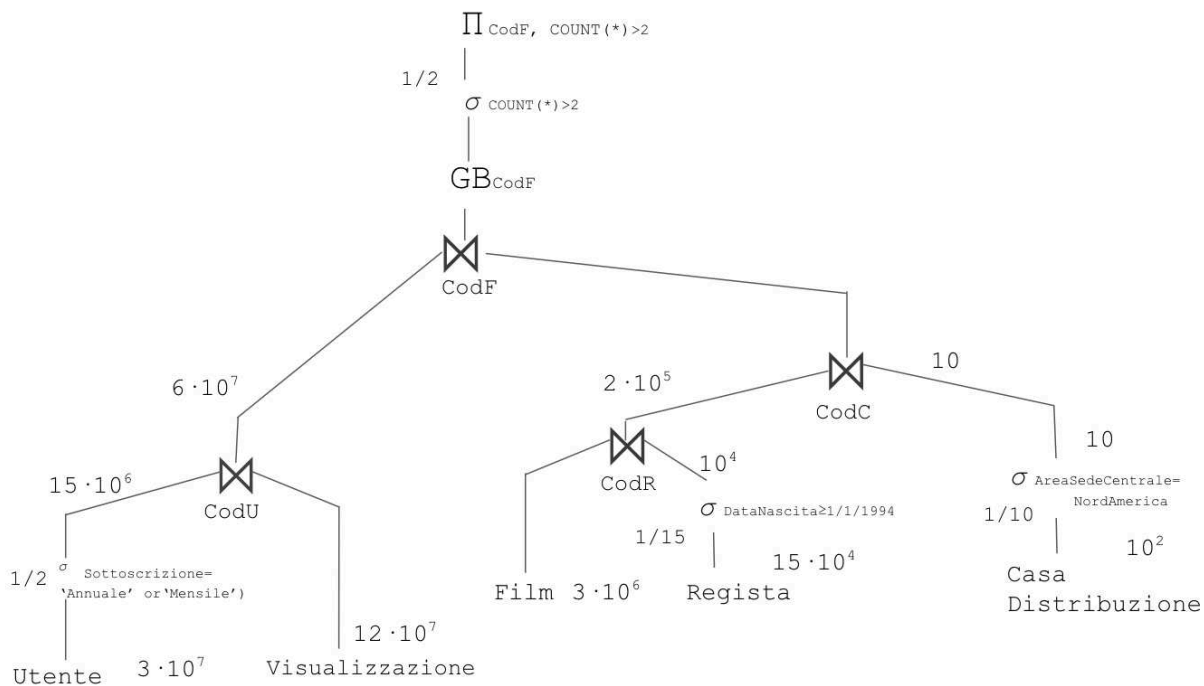
Punteggio ottenuto 5,00 su 5,00

La risposta corretta è: $2 \cdot 10^4$

Indici

(1.5 punti, penalità -15% per ogni risposta sbagliata)

La figura sottostante rappresenta il query tree per la query precedente.



Selezionare, per ogni tabella, le strutture fisiche accessorie per migliorare le prestazioni dell'interrogazione (se possibile) tra le opzioni di seguito.

Tabella VISUALIZZAZIONE

- ☐ CREATE INDEX IndexB ON VISUALIZZAZIONE(DataUltimaVisualizzazione) - B+-Tree
- ☒ Nessuna strutture fisica accessoria su questa tabella migliorerebbe le prestazioni della interrogazione ✓
- ☐ CREATE INDEX IndexA ON VISUALIZZAZIONE(DataUltimaVisualizzazione) - HASH

Punteggio ottenuto 3,00 su 3,00

La risposta corretta è: Nessuna strutture fisica accessoria su questa tabella migliorerebbe le prestazioni della interrogazione

Tabella UTENTE

- ☐ CREATE INDEX IndexC ON UTENTE(Sottoscrizione) - HASH
- ☐ CREATE INDEX IndexD ON UTENTE(Sottoscrizione) - B+-Tree
- ☒ Nessuna strutture fisica accessoria su questa tabella migliorerebbe le prestazioni della interrogazione ✓

Punteggio ottenuto 3,00 su 3,00

La risposta corretta è: Nessuna strutture fisica accessoria su questa tabella migliorerebbe le prestazioni della interrogazione

Tabella FILM

- ☐ CREATE INDEX IndexF ON FILM(Genere) - B+-Tree
- ☐ CREATE INDEX IndexE ON FILM(Genere) - HASH
- ☒ Nessuna strutture fisica accessoria su questa tabella migliorerebbe le prestazioni della interrogazione ✓

Punteggio ottenuto 3,00 su 3,00

La risposta corretta è: Nessuna strutture fisica accessoria su questa tabella migliorerebbe le prestazioni della interrogazione

Tabella REGISTA

- ☒ CREATE INDEX IndexH ON REGISTA(DataNascita) - B+-Tree ✓
- ☐ CREATE INDEX IndexG ON REGISTA(DataNascita) - HASH
- ☐ Nessuna strutture fisica accessoria su questa tabella migliorerebbe le prestazioni della interrogazione

Punteggio ottenuto 3,00 su 3,00

La risposta corretta è: CREATE INDEX IndexH ON REGISTA(DataNascita) - B+-Tree

Tabella DISTRIBUZIONE

- ☒ Nessuna strutture fisica accessoria su questa tabella migliorerebbe le prestazioni della interrogazione ✓
- ☐ CREATE INDEX IndexJ ON CASA-DISTRIBUZIONE(AreaSedeCentrale) - B+-Tree
- ☐ CREATE INDEX IndexI ON CASA-DISTRIBUZIONE(AreaSedeCentrale) - HASH

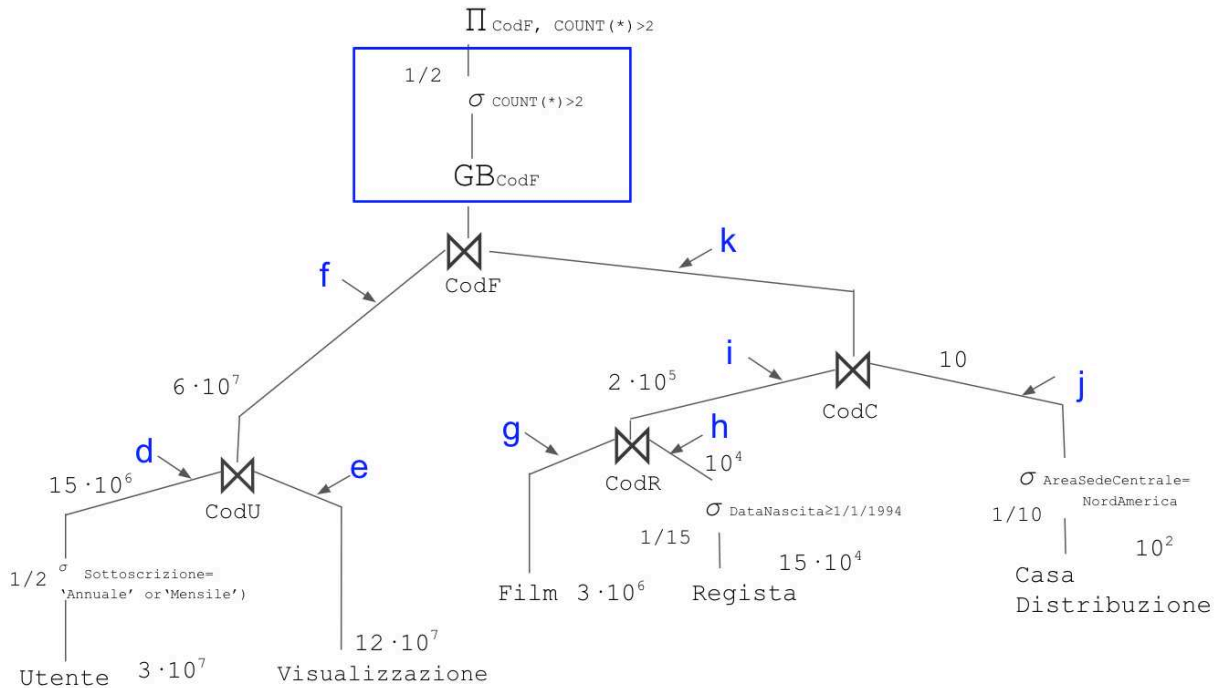
Punteggio ottenuto 3,00 su 3,00

La risposta corretta è: Nessuna strutture fisica accessoria su questa tabella migliorerebbe le prestazioni della interrogazione

Anticipo Group By

(2 punti, penalità -15% per ogni risposta sbagliata)

La figura sottostante rappresenta il query tree per la query precedente.



Analizzare l'antipico della GROUP BY **group by F.CodF having COUNT(*)>2** rappresentata nel riquadro. Selezionare la soluzione che consente la massima efficienza nell'esecuzione della query (se esiste).

- ☐ E' possibile anticiparla nel ramo h
- ☐ E' possibile anticiparla nel ramo g
- ☒ E' possibile anticiparla nel ramo f ✓
- ☐ E' possibile anticiparla nel ramo e
- ☐ Non e' possibile anticipare la Group BY group by F.CodF having COUNT(*)>2
- ☐ E' possibile anticiparla nel ramo k
- ☐ E' possibile anticiparla nel ramo j
- ☐ E' possibile anticiparla nel ramo d
- ☐ E' possibile anticiparla nel ramo i

Punteggio ottenuto 20,00 su 20,00

La risposta corretta è: E' possibile anticiparla nel ramo f

- 1) La risposta corretta è : $2 \cdot 10^4$
- 2) La risposta corretta è : $4 \cdot 10^5$
- 3) La risposta corretta è : $2 \cdot 10^4$
- 4) La risposta corretta è : Nessuna strutture fisica accessoria su questa tabella migliorerebbe le

prestazioni della interrogazione

5) La risposta corretta è : Nessuna strutture fisica accessoria su questa tabella migliorerebbe le prestazioni della interrogazione

6) La risposta corretta è : Nessuna strutture fisica accessoria su questa tabella migliorerebbe le prestazioni della interrogazione

7) La risposta corretta è : CREATE INDEX IndexH ON REGISTA(DataNascita) - B+-Tree

8) La risposta corretta è : Nessuna strutture fisica accessoria su questa tabella migliorerebbe le prestazioni della interrogazione

9) La risposta corretta è : E' possibile anticiparla nel ramo f

Domanda 7

Completo

Punteggio ottenuto 5,00 su 5,00

5 punti (nessuna penalità per risposta sbagliata)

Il seguente data warehouse descrive l'andamento degli acquisti dei prodotti finanziari (azioni, titoli, fondi di investimento, etc) offerti da diverse società tramite intermediari finanziari (banca, broker, agenzie, etc). Ogni prodotto finanziario è emesso da una specifica società ed afferisce ad una sola categoria (titoli di massa, contratti derivati, etc). Per ogni intermediario finanziario sono noti la città, la regione e lo stato in cui è presente. Il data warehouse memorizza la categoria del cliente (senior, medium, junior), la propensione al rischio (alta, media e bassa), la professione e la nazionalità dei clienti che hanno acquistato nel tempo i prodotti finanziari tramite gli intermediari finanziari.

Le metriche da analizzare sono il numero di prodotti acquistati (#Acquisti) e il corrispondente valore (ValoreAcquisti). Le metriche devono essere analizzate per ogni mese, bimestre, trimestre, quadrimestre, semestre, anno.



INTERMEDIARIO-FINANZIARIO(IDIntermediarioFinanziario, IntermediarioFinanziario, Città, Regione, Stato)

PRODOTTO-FINANZIARIO(IDProdottoFinanziario, ProdottoFinanziario, Società, SedeLegale, CategoriaP)

JUNK-CARATTERISTICHE-CLIENTE(IDJCC, CategoriaC, PropensioneRischio, Professione, Nationalità)

TEMPO(IDTempo, Mese, Mese-Anno, 2-Mesi, 3-Mesi, 4-Mesi, 6-Mesi, Anno)

ACQUISTO(IDIntermediarioFinanziario, IDProdottoFinanziario, IDJCC, IDTempo, #Acquisti, ValoreAcquisti)

Dato lo schema logico precedente, considerare le seguenti query di interesse:

- Considerando gli acquisti effettuati da clienti con nazionalità italiana e propensione al rischio 'alta' (attributo PropensioneRischio), visualizzare il numero complessivo di acquisti, il valore complessivo corrispondente e il valore medio per acquisto, separatamente per bimestre (attributo 2-Mesi) e regione dell'intermediario.
- Considerando gli anni 2021 e 2022, separatamente per anno e città dell'intermediario, visualizzare il valore complessivo e il valore medio bimestrale degli acquisti.
- Separatamente per nazionalità del cliente e semestre (attributo 6-Mesi), visualizzare il valore medio per acquisto.

Dato lo schema logico precedente, si svolgano le seguenti attività:

- Definire una vista materializzata con CREATE MATERIALIZED VIEW, in modo da ridurre il tempo di risposta delle query di interesse da (a) a (c) sopra riportate. In particolare si specifichi la query SQL associata al **Blocco A** nella seguente istruzione:

```
CREATE MATERIALIZED VIEW ViewAcquisti
BUILD IMMEDIATE
REFRESH FAST ON COMMIT
AS
```

Blocco A

2. Definire l'**insieme minimale** di attributi che permette di identificare le tuple appartenenti alla vista materializzata ViewAcquisti..

3. Si ipotizzi che la gestione della vista materializzata (tabella derivata) sia svolta mediante trigger. Scrivere il trigger per propagare le modifiche alla vista materializzata ViewAcquisti in caso di inserimento di un nuovo record nella tabella dei fatti ACQUISTI.

```
SELECT 2-Mesi, 6-Mesi, Anno, Città, Regione, Nazionalità, PropensioneRischio,
SUM(#Acquisti) AS TotAcquisti,
SUM(ValoreAcquisti) AS TotValore
FROM ACQUISTO A, TEMPO T, JUNK-CARATTERISTICHE-CLIENTE JCC, INTERMEDIARIO-
FINANZIARIO IF
WHERE A.IDTempo = T.IDTempo AND A.IDJCC = JCC.IDJCC AND A.IDIntermediarioFinanziario
= IF.IDIntermediarioFinanziario
GROUP BY 2-Mesi, 6-Mesi, Anno, Città, Regione, Nazionalità, PropensioneRischio;
```

Insieme minimale di attributi: (2-Mesi, Città, Nazionalità, PropensioneRischio)

```
CREATE OR REPLACE TRIGGER UpdateViewAcquisti
AFTER INSERT ON ACQUISTI
FOR EACH ROW
DECLARE
bim DATE,
sem DATE,
ann DATE,
cit VARCHAR(20),
reg VARCHAR(20),
naz VARCHAR(20),
pro VARCHAR(20),
N INTEGER;
BEGIN
SELECT 2-Mesi, 6-Mesi, Anno INTO bim, sem, ann
FROM TEMPO
WHERE IDTempo = :NEW.IDTempo;

SELECT Città, Regione INTO cit, reg
FROM INTERMEDIARIO-FINANZIARIO
```



```
WHERE IDIntermediarioFinanziario = :NEW.IDIntermediarioFinanziario;

SELECT Nazionalità, PropensioneRischio INTO naz, pro
FROM JUNK-CARATTERISTICHE-CLIENTE
WHERE IDJCC = :NEW.IDJCC;

SELECT COUNT(*) INTO N
FROM ViewAcquisti
WHERE 2-Mesi = bim AND Città = cit AND Nazionalità = naz AND PropensioneRischio = pro;

IF (N > 0) THEN
  UPDATE ViewAcquisti
  SET TotAcquisti += :NEW.#Acquisti, TotValore += :NEW.ValoreAcquisti
  WHERE 2-Mesi = bim AND Città = cit AND Nazionalità = naz AND PropensioneRischio = pro;
ELSE
  INSERT INTO ViewAcquisti(2-Mesi, 6-Mesi, Anno, Città, Regione, Nazionalità,
PropensioneRischio, TotAcquisti, TotValore)
  VALUES(bim, sem, ann, cit, reg, naz, pro, :NEW.#Acquisti, :NEW.ValoreAcquisti);
ENDIF
END
```

Commento:

Domanda 8

Risposta corretta

Punteggio ottenuto 2,50 su 2,50

2,5 punti (penalita' 15% per risposta sbagliata)

Viene data la seguente lista di transazioni.

abcd
acde
abc
abd
cde
acde
b
d
abde
acd

Viene applicato l'algoritmo FP-growth usando minsup = 2 (un itemset e' considerato frequente se compare in almeno due transazioni).

Qual e' il BC-CPB (BC-Conditional Pattern Base)?

-
- ☐ (a) ae: 1, a: 2
 - ☐ (b) abc: 1, a: 1
 - ☐ (c) ae: 1, a: 1
 - ☐ (d) ae: 1, a: 2
 - ☒ (e) ad: 1, a: 1 ✓
 - ☐ (f) ade: 1, a: 2
 - ☐ (g) ad: 3, a: 4
 - ☐ (h) ade: 2, a: 1

Risposta corretta.

La risposta corretta è: ad: 1, a: 1

Domanda 9

Risposta corretta

Punteggio ottenuto 1,50 su 1,50

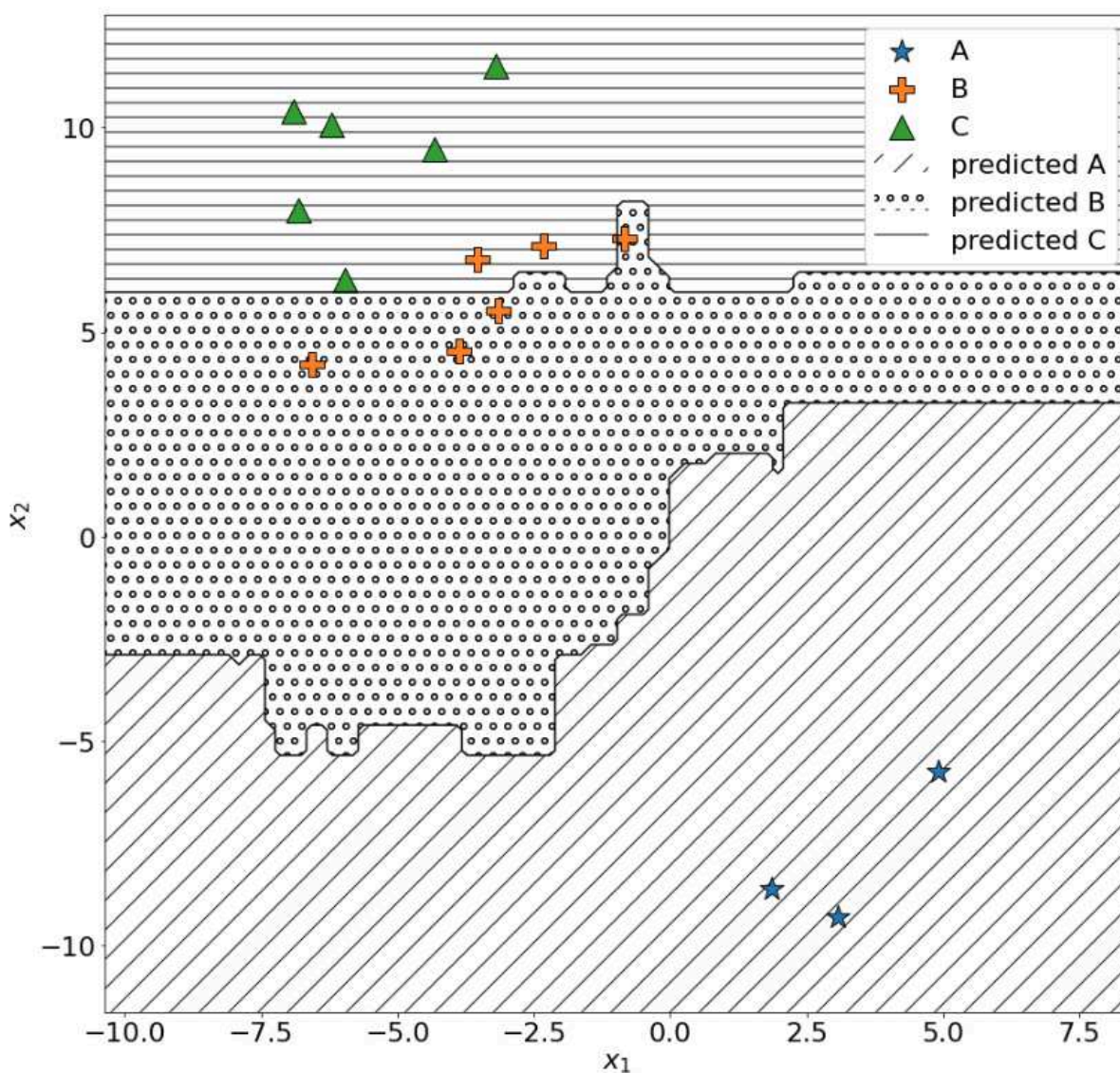
1 punto (penalità 15% per risposta sbagliata)

Precisione(X) è la frazione di predizioni corrette rispetto a tutte le predizioni fatte per la classe X

Richiamo(X) è la frazione di predizioni corrette rispetto a tutti i punti che appartengono alla classe X

Un classificatore basato su random forest e' stato addestrato con un dataset bidimensionale (features x_1 , x_2). Ciascun punto del dataset e' etichettato con un'etichetta di classe tra A, B, C (rispettivamente rappresentate come stella, croce, triangolo).

La seguente figura rappresenta un test set che e' stato utilizzato per validare il classificatore.



Le etichette assegnate del modello sono mostrate nella figura:

- Le aree con righe diagonali rappresentano zone in cui il modello predice classe A

- Le aree con dei pallini rappresentano zone in cui il modello predice classe B
- Le aree con righe orizzontali rappresentano zone in cui il modello predice classe C

Sulla base di queste informazioni, qual è la precisione per la classe C e il richiamo per la classe B?

- ☐ (a) precisione C: 3/4, richiamo B: 1
- ☐ (b) precisione C: 2/3 richiamo B: 1
- ☐ (c) precisione C: 2/3, richiamo B: 3/4
- ☒ (d) Nessuna delle altre risposte è corretta ✓
- ☐ (e) precisione C: 1, richiamo B: 1
- ☐ (f) precisione C: 1, richiamo B: 2/3
- ☐ (g) precisione C: 1, richiamo B: 3/4

Risposta corretta.

La risposta corretta è: Nessuna delle altre risposte è corretta

Domanda 10

Risposta corretta

Punteggio ottenuto 1,00 su 1,00

1 punto (penalità 15% per risposta sbagliata)

Il seguente è un documento preso da una collection MongoDB denominata "ingressi", che mantiene informazioni sugli ingressi dei clienti di una palestra.

```
{
  "first": "DOROTHY",
  "last": "MILLER",
  "age": 20,
  "gender": "F",
  "date": {
    "day": 16,
    "month": 8,
    "year": 2022
  },
}
```

Sulla base della struttura inferibile da questo documento, quali delle seguenti query su MongoDB estrae il numero massimo di ingressi effettuati nei mesi dell'anno in cui la durata media degli ingressi e' minore di 75 minuti?

☒ (a)

```
db.collection.aggregate([
  {
    $group: {
      _id: "$date.month",
      "avg_duration": {
        $avg: "$duration"
      },
      "count": {
        $sum: 1
      }
    }
  },
  {
    $match: {
      avg_duration: {
        $lte: 75
      }
    }
  },
  {
    $group: {
      _id: null,
      max_count: {
        $max: "$count"
      }
    }
  }
])
```



☐ (b)

```
db.collection.aggregate([
  {
    $match: {
      avg_duration: {
        $lte: 75
      }
    }
  },
  {
    $group: {
      _id: null,
      "avg_duration": {
        $avg: "$duration"
      },
      "count": {
        $sum: 1
      }
    }
  },
  {
    $group: {
      _id: "$date.month",
      max_count: {
        $max: "$count"
      }
    }
  }
])
```

○ (c)

```
db.collection.aggregate([
  {
    $match: {
      avg_duration: {
        $lte: 75
      }
    }
  },
  {
    $group: {
      _id: "$date.month",
      "avg_duration": {
        $avg: "$duration"
      }
    }
  },
  {
    $group: {
      _id: null,
      max_count: {
        $max: "$count"
      }
    }
  }
])
```

☐ (d)


```
db.collection.aggregate([
  {
    $group: {
      _id: "$date.month",
      "avg_duration": {
        $avg: "$duration"
      }
    }
  },
  {
    $match: {
      avg_duration: {
        $lte: 75
      }
    }
  },
  {
    $group: {
      _id: null,
      max_count: {
        $max: "$count"
      }
    }
  }
])
```

☐ (e)

```
db.collection.aggregate([
  {
    $group: {
      _id: null,
      "avg_duration": {
        $avg: "$duration"
      },
      "count": {
        $sum: 1
      }
    }
  },
  {
    $match: {
      avg_duration: {
        $lte: 75
      }
    }
  },
  {
    $group: {
      _id: "$date.month",
      max_count: {
        $max: "$count"
      }
    }
  }
])
```

☐ (f)

```
db.collection.aggregate([
  {
    $match: {
      avg_duration: {
        $lte: 75
      }
    }
  },
  {
    $group: {
      _id: "$date.month",
      "avg_duration": {
        $avg: "$duration"
      },
      "count": {
        $sum: 1
      }
    }
  },
  {
    $group: {
      _id: null,
      max_count: {
        $max: "$count"
      }
    }
  }
])
```

○ (g)

```
db.collection.aggregate([
  {
    $match: {
      avg_duration: {
        $lte: 75
      }
    }
  },
  {
    $group: {
      _id: null,
      "avg_duration": {
        $avg: "$duration"
      }
    }
  },
  {
    $group: {
      _id: "$date.month",
      max_count: {
        $max: "$count"
      }
    }
  }
])
```

○ (h)

```
db.collection.aggregate([
  {
    $group: {
      _id: null,
      "avg_duration": {
        $avg: "$duration"
      }
    }
  },
  {
    $match: {
      avg_duration: {
        $lte: 75
      }
    }
  },
  {
    $group: {
      _id: "$date.month",
      max_count: {
        $max: "$count"
      }
    }
  }
])
```

Risposta corretta.

La risposta corretta è:

```
db.collection.aggregate([
  {
    $group: {
      _id: "$date.month",
      "avg_duration": {
        $avg: "$duration"
      },
      "count": {
        $sum: 1
      }
    }
  },
  {
    $match: {
      avg_duration: {
        $lte: 75
      }
    }
  },
  {
    $group: {
      _id: null,
      max_count: {
        $max: "$count"
      }
    }
  }
])
```

Domanda 11

Risposta corretta

Punteggio ottenuto 1,00 su 1,00

1 punto (penalità 15% per risposta sbagliata)

È data la seguente sequenza di operazioni all'interno di un file di log:

B(T0) Abort(T0) CK() B(T1) B(T2) CK(T1,T2) I1(o1) D1(o1) Abort(T2) Commit(T1) B(T3) I3(o2) U3(o2) U3(o4) D3(o4) FAILURE

Notazione:

- Tn: Id della transazione n
- B(Tn): Begin della transazione Tn
- CK(Ta,Tb,...): checkpoint con transazioni Ta, Tb, ... non concluse
- Commit(Tn): commit della transazione Tn
- Abort(Tn): abort (rollback) della transazione Tn
- Un(ox): update eseguito dalla transazione Tn sull'oggetto ox
- In(ox): insert eseguito dalla transazione Tn sull'oggetto ox
- Dn(ox): delete eseguito dalla transazione Tn sull'oggetto ox

Quali sono gli insiemi di UNDO e REDO **finali** per il warm restart?

- ☐ (a) undo: [T2], redo[T3, T1]
- ☐ (b) undo: [T0, T2, T3], redo: [T1]
- ☒ (c) undo: [T2, T3], redo: [T1] ✓
- ☐ (d) undo: [T0, T1, T2, T3], redo: []
- ☐ (e) undo: [T1, T2, T3], redo: []
- ☐ (f) undo: [T0, T2], redo[T3, T1]
- ☐ (g) undo: [T0, T1,T3], redo: [T2]
- ☐ (h) undo: [T1,T3], redo: [T2]

Risposta corretta.

La risposta corretta è: undo: [T2, T3], redo: [T1]