



Data Science e Tecnologie per le Basi di Dati

Esame del 21 Febbraio 2023

Valutazione 25,78 su un massimo di 32,00 (81%)

Domanda 1

Risposta non data

Punteggio max.: 1,50

1.5 punti (-15% penalità per risposta sbagliata)

La policy di MAX (complete) linkage prevede che la distanza fra due cluster X e Y sia calcolata come:

$$\text{dist}(X, Y) = \max_{x \in X, y \in Y} \text{dist}(x, y)$$

dove $\text{dist}(x, y)$ è una distanza che può essere definita fra coppie di punti.

Per un dataset di 5 punti viene calcolata la seguente matrice di distanze.

	a	b	c	d	e
a	0	10	6	5	13
b	10	0	21	12	25
c	6	21	0	4	11
d	5	12	4	0	2
e	13	25	11	2	0

Si applica il clustering gerarchico agglomerativo per estrarre 3 cluster. Viene utilizzata la policy di "MAX linkage" (complete linkage).

Quali sono i 3 cluster ottenuti?

- ☐ (a) {a}, {b}, {c, d, e}
- ☐ (b) Non è possibile rispondere alla domanda con le informazioni a disposizione
- ☐ (c) {c, e}, {a, d}, {b}
- ☐ (d) {a, d}, {b, e}, {c}
- ☐ (e) {d, e}, {a}, {b, c}
- ☐ (f) Nessuna delle altre risposte è corretta
- ☐ (g) {d, e}, {a, b}, {c}
- ☐ (h) {b}, {d, e}, {a, c}

Risposta errata.

La risposta corretta è: {b}, {d, e}, {a, c}

Domanda 2

Risposta corretta

Punteggio ottenuto 5,00 su 5,00

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

Sono date le seguenti tabelle:

Brano(CodB, Titolo, Genere, CodA)
Artista(CodA, Nome, Cognome, Nazionalità, DataNascita)
User(CodU, Nome, Cognome, Nazionalità, DataNascita, Email)
LikeUser(CodU, CodB, Data, Piattaforma)

Sono date le seguenti cardinalità:

- $\text{card}(\text{BRANO}) = 10^8$ tuple
 - Valori distinti di Genere = 10
- $\text{card}(\text{ARTISTA}) = 5 \cdot 10^6$ tuple
 - Valori distinti di Nazionalità = 100
 - $\text{MIN}(\text{DataNascita}) = 1/1/1900$, $\text{MAX}(\text{DataNascita}) = 31/12/1999$
- $\text{card}(\text{USER}) = 2 \cdot 10^7$ tuple
 - $\text{MIN}(\text{DataNascita}) = 1/1/1930$, $\text{MAX}(\text{DataNascita}) = 31/12/2004$
 - Valori distinti di Nazionalità = 100
- $\text{card}(\text{LIKEUSER}) = 10^{10}$ tuple
 - Valori distinti di Piattaforma = 5
 - $\text{MIN}(\text{Data}) = 1/1/2003$, $\text{MAX}(\text{Data}) = 31/12/2022$

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

- Having $\text{COUNT}(\ast) \geq 150 = 1/5$

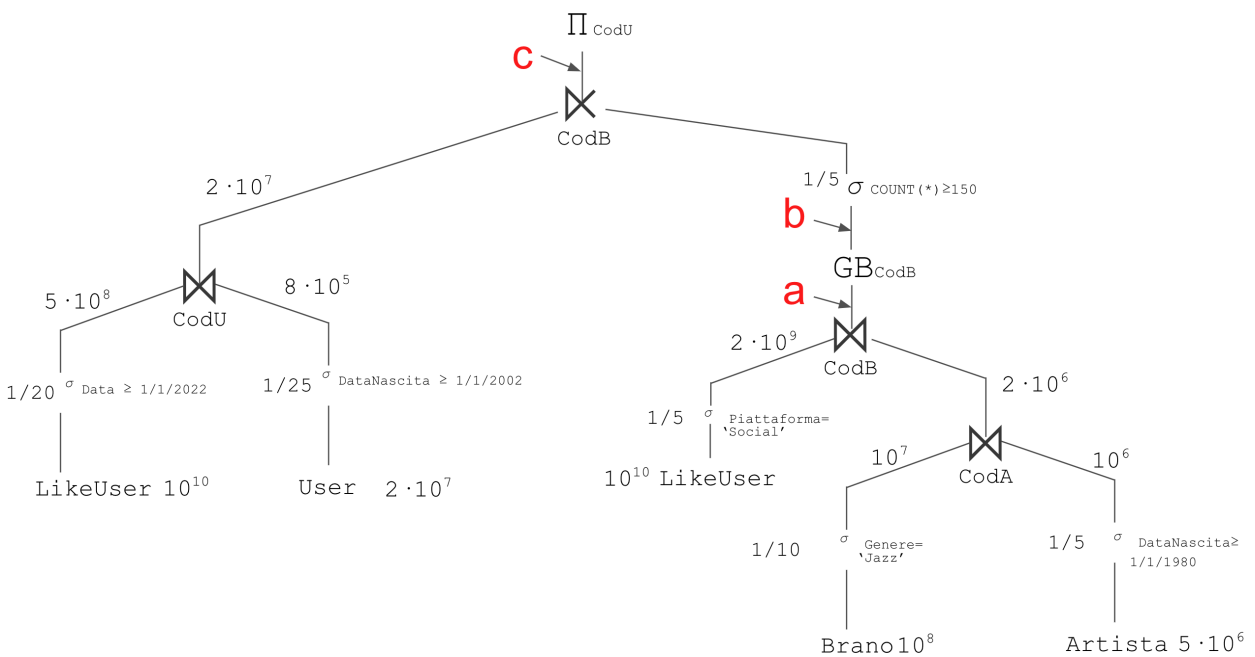
Si consideri la seguente query:

```
select LU1.CodU
from User U, LikeUser LU1
where U.CodU=LU1.CodU
and LU1.Data ≥ 1/1/2022
and U.DataNascita ≥ 1/1/2002
and LU1.CodB IN (SELECT LU2.CodB
FROM Brano B, Artista A, LikeUser LU2
WHERE B.CodA=A.CodA and LU2.CodB=B.CodB
and LU2.Piattaforma='Social'
and A.DataNascita ≥ 1/1/1980
and B.Genere='Jazz'
GROUP BY LU2.CodB
HAVING COUNT(*) ≥ 150
```

Cardinalità

(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 cardinalità di **(a)**:

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

Punteggio ottenuto 5,00 su 5,00

La risposta corretta è: $4 \cdot 10^7$

Seleziona la risposta corretta per la cardinalità di **(b)**:

- ☐ 10^8 ☐ 10^7 ☒ $2 \cdot 10^6$ ✓ ☐ 10^4

Punteggio ottenuto 5,00 su 5,00

La risposta corretta è: $2 \cdot 10^6$

Seleziona la risposta corretta per la cardinalità di **(c)**:

- ☐ $8 \cdot 10^6$ ☒ $8 \cdot 10^4$ ✓ ☐ $2 \cdot 10^5$ ☐ $2 \cdot 10^7$

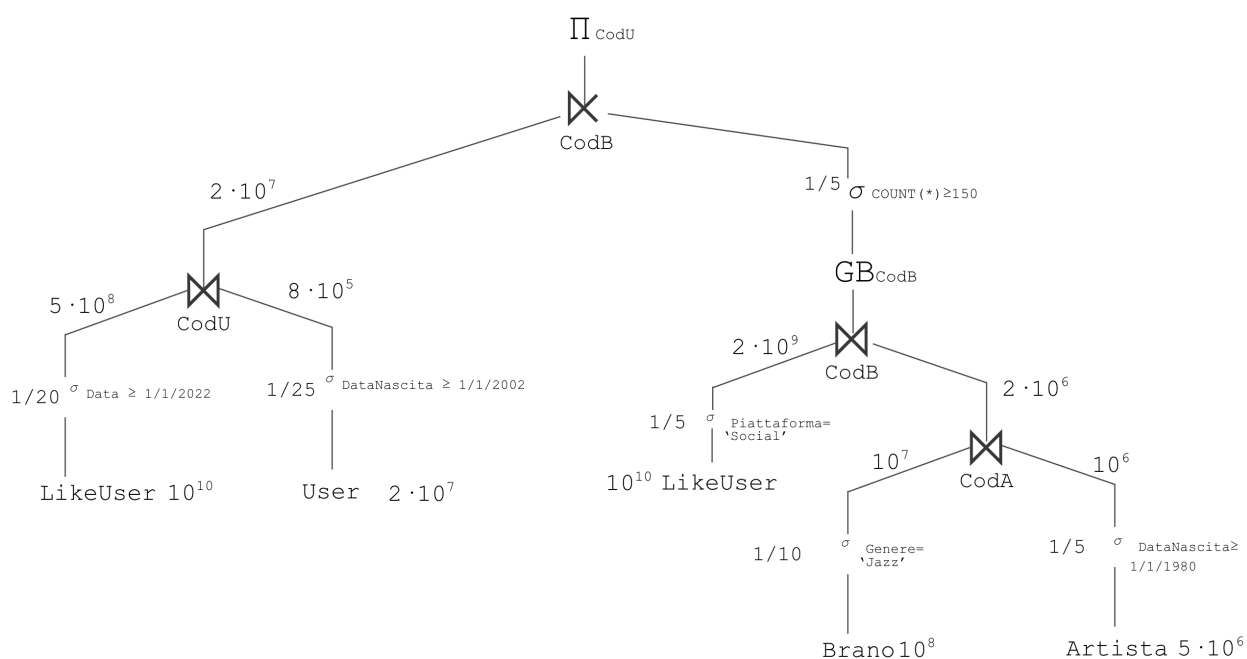
Punteggio ottenuto 5,00 su 5,00

La risposta corretta è: $8 \cdot 10^4$

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 LIKEUSER

- ☒ CREATE INDEX IndexB ON LIKEUSER(Data) - B+- Tree ✓
- ☐ Nessuna struttura fisica accessoria su questa tabella migliorerebbe le prestazioni dell'interrogazione
- ☐ CREATE INDEX IndexA ON LIKEUSER(Data) - HASH

Punteggio ottenuto 3,00 su 3,00

La risposta corretta è: CREATE INDEX IndexB ON LIKEUSER(Data) - B+- Tree

Tabella LIKEUSER

- ☐ CREATE INDEX IndexD ON LIKEUSER(Piattaforma) - B+- Tree
- ☐ CREATE INDEX IndexC ON LIKEUSER(Piattaforma) - HASH
- ☒ Nessuna struttura fisica accessoria su questa tabella migliorerebbe le prestazioni dell'interrogazione ✓

Punteggio ottenuto 3,00 su 3,00

La risposta corretta è: Nessuna struttura fisica accessoria su questa tabella migliorerebbe le prestazioni dell'interrogazione

Tabella USER

- ☐ CREATE INDEX IndexE ON USER(DataNascita) - HASH
- ☐ Nessuna struttura fisica accessoria su questa tabella migliorerebbe le prestazioni dell'interrogazione
- ☒ CREATE INDEX IndexF ON USER(DataNascita) - B+- Tree ✓

Punteggio ottenuto 3,00 su 3,00

La risposta corretta è: CREATE INDEX IndexF ON USER(DataNascita) - B+- Tree

Tabella BRANO

- ☒ CREATE INDEX IndexG ON BRANO(Genere) - HASH ✓
- ☐ Nessuna struttura fisica accessoria su questa tabella migliorerebbe le prestazioni dell'interrogazione
- ☐ CREATE INDEX IndexH ON BRANO(Genere) - B+- Tree

Punteggio ottenuto 3,00 su 3,00

La risposta corretta è: CREATE INDEX IndexG ON BRANO(Genere) - HASH

Tabella ARTISTA

- ☐ CREATE INDEX IndexI ON ARTISTA(Nazionalità) - HASH
- ☒ Nessuna struttura fisica accessoria su questa tabella migliorerebbe le prestazioni dell'interrogazione ✓
- ☐ CREATE INDEX IndexJ ON ARTISTA(Nazionalità) - B+- Tree

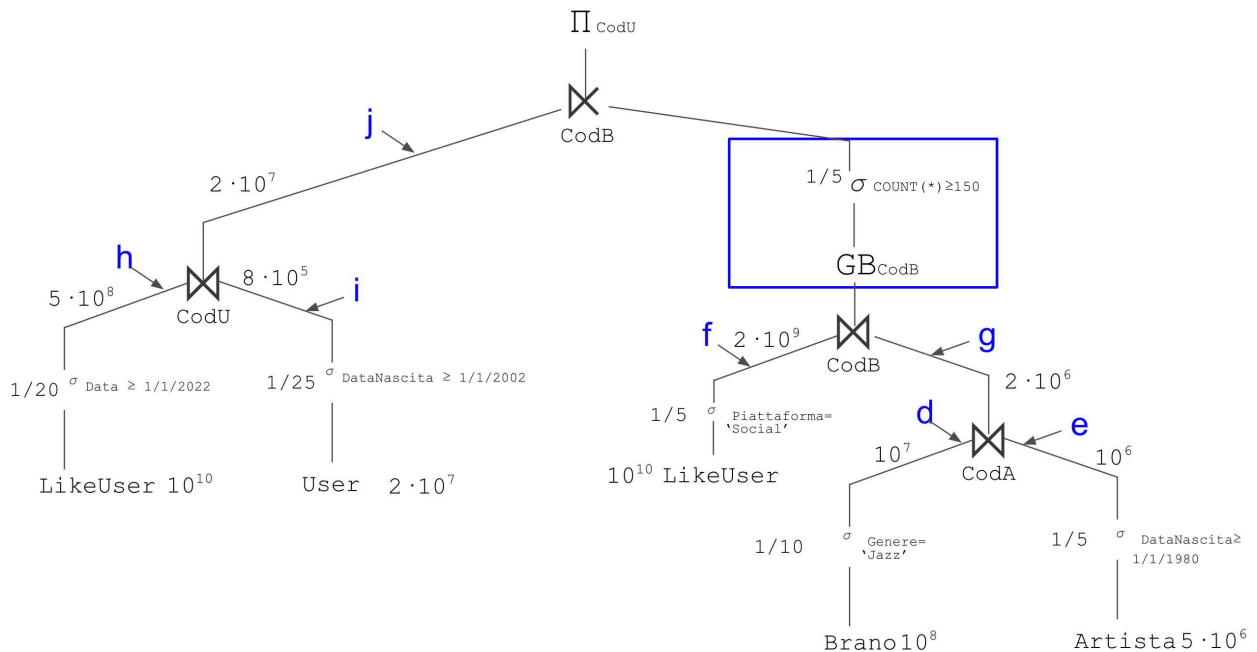
Punteggio ottenuto 3,00 su 3,00

La risposta corretta è: Nessuna struttura fisica accessoria su questa tabella migliorerebbe le prestazioni dell'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'anticipo della GROUP BY **GROUP BY LU2.CodB HAVING COUNT(*)≥150** rappresentata nel riquadro. Selezionare la soluzione che consente la massima efficienza nell'esecuzione della query (se esiste).

- ☒ E' possibile anticiparla nel ramo f ✓
- ☐ E' possibile anticiparla nel ramo d
- ☐ Non è possibile anticipare la Group By GROUP BY LU2.CodB HAVING COUNT(*)≥150
- ☐ E' possibile anticiparla nel ramo e
- ☐ E' possibile anticiparla nel ramo g
- ☐ E' possibile anticiparla nel ramo i
- ☐ E' possibile anticiparla nel ramo j
- ☐ E' possibile anticiparla nel ramo h

Punteggio ottenuto 20,00 su 20,00

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

- 1) La risposta corretta è : $4 \cdot 10^7$
- 2) La risposta corretta è : $2 \cdot 10^6$
- 3) La risposta corretta è : $8 \cdot 10^4$
- 4) La risposta corretta è : CREATE INDEX IndexB ON LIKEUSER(Data) - B+- Tree
- 5) La risposta corretta è : Nessuna struttura fisica accessoria su questa tabella migliorerebbe le

prestazioni dell'interrogazione

6) La risposta corretta è : CREATE INDEX IndexF ON USER(DataNascita) - B+- Tree

7) La risposta corretta è : CREATE INDEX IndexG ON BRANO(Genere) - HASH

8) La risposta corretta è : Nessuna struttura fisica accessoria su questa tabella migliorerebbe le prestazioni dell'interrogazione

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

Domanda 3

Risposta corretta

Punteggio ottenuto 2,00 su 2,00

2 punti (-15% penalità per risposta sbagliata)

Viene data la seguente base dati transazionale.

Transactions	
0	B C
1	A D E
2	A D E
3	A B C
4	C E
5	B C
6	B D
7	A D E
8	A B C E
9	A E

Si applichi l'algoritmo Apriori per estrarre gli itemset frequenti. Si utilizzi minsup = 2 (un itemset e' frequente se compare in almeno 2 transazioni).

Quali sono gli itemset di lunghezza 3 che vengono generati da Apriori **dopo i passi di join e prune** (con principio Apriori), **prima del conteggio del supporto** nella base dati?

- ☐ (a) Nessuna delle altre risposte e' corretta
- ☒ (b) ABC, ACE, ADE ✓
- ☐ (c) ABC, ABE
- ☐ (d) ABC, ADE

- ☐ (e) ABD, ABE, ACD, ECD
- ☐ (f) ABC, ABD, ABE, ACD, ACE, ADE, ECD
- ☐ (g) ABC, ABD, ABE, ACD, ACE, ADE, BCD, BCE
- ☐ (h) ABC, ABD, ACE, ADE
- ☐ (i) Non e' possibile rispondere alla domanda con le informazioni a disposizione

Risposta corretta.

La risposta corretta è: ABC, ACE, ADE

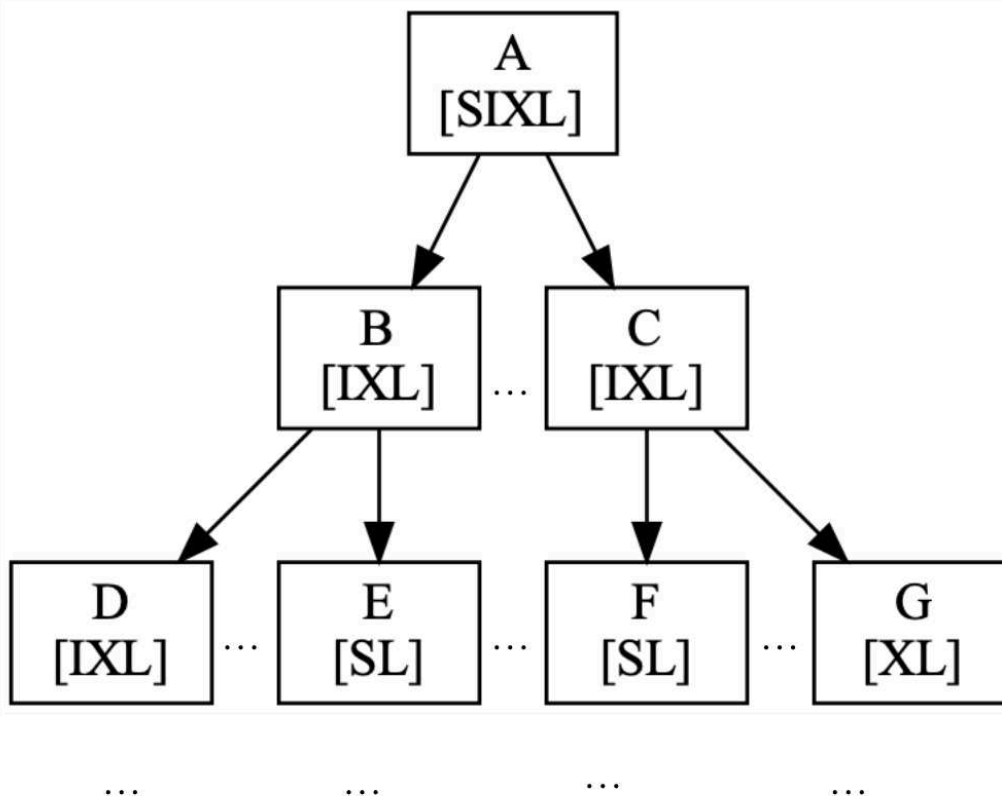
Domanda 4

Risposta corretta

Punteggio ottenuto 1,00 su 1,00

1 punto (penalità 15% per risposta sbagliata)

Il seguente albero mostra un sottoinsieme di una base dati con i diversi livelli di granularità (e.g. tabelle/frammenti/record). Per ogni nodo dell'albero, sono riportati i lock già acquisiti da altre transazioni tra [parentesi quadre].



Si tenga presente che, nell'albero, ciascun nodo eredita sempre un lock di livello almeno pari a quello del nodo padre, anche se non espressamente indicato nell'immagine.

Quale delle seguenti sequenze di lock può essere acquisita da una nuova transazione?

- ☐ (a) ISL su A, ISL su B, IXL su E
- ☒ (b) ISL su A, ISL su B, SL su E ✓
- ☐ (c) ISL su A, ISL su B, SL su D
- ☐ (d) IXL su A, IXL su B, IXL su E
- ☐ (e) ISL su A, SL su B, SL su E
- ☐ (f) ISL su A, ISL su B, XL su E
- ☐ (g) SIXL su A, ISL su B, ISL su E

Risposta corretta.

La risposta corretta è: ISL su A, ISL su B, SL su E

Domanda 5

Risposta corretta

1 punto (penalità 15% per risposta sbagliata)

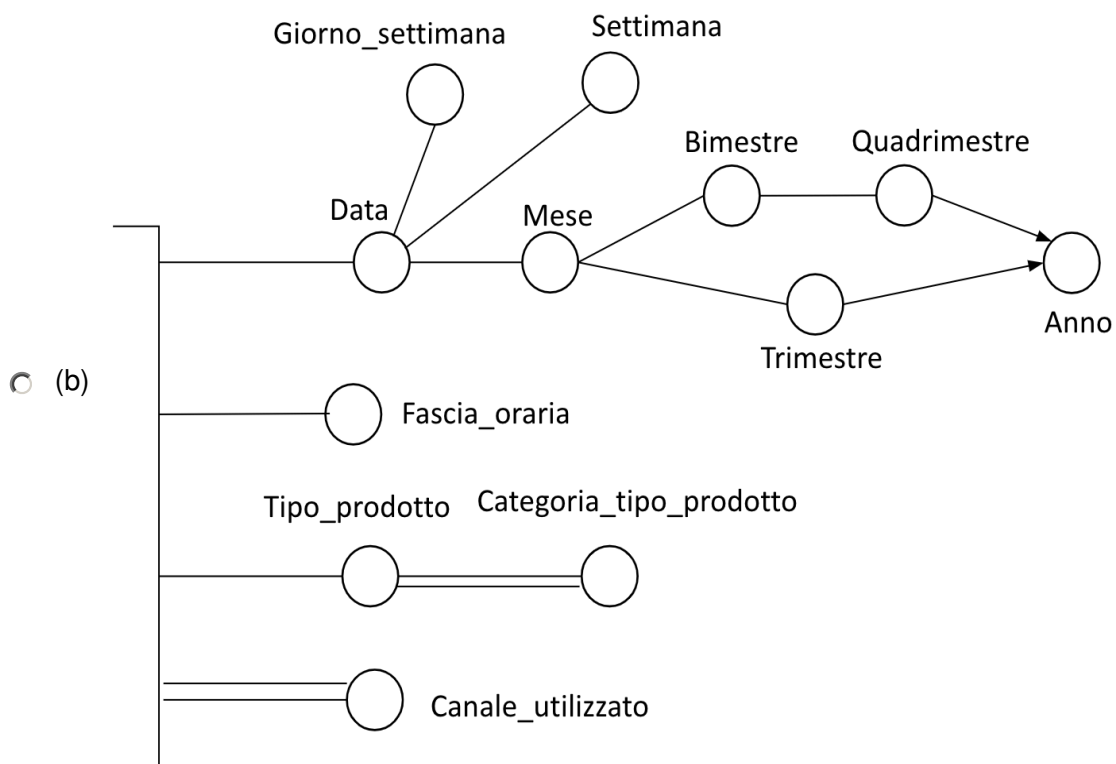
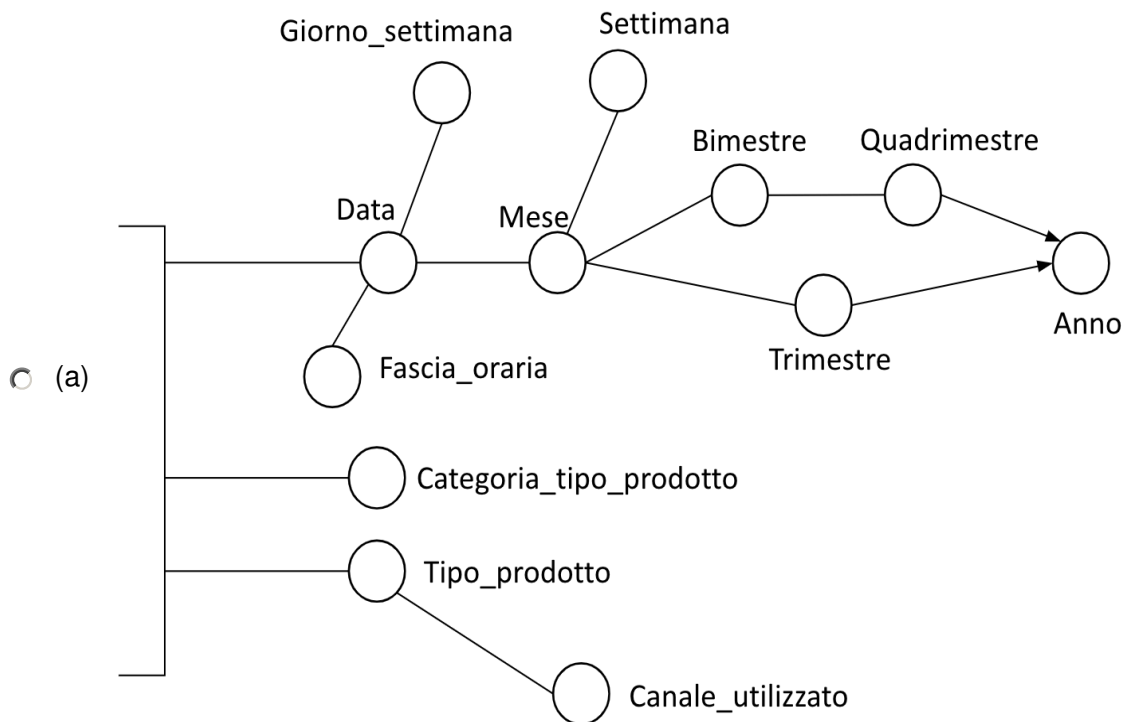
Si vogliono analizzare le informazioni relative ad una catena di call center che svolge attività per campagne di vendita su diversi tipi di prodotti forniti da una società multinazionale.

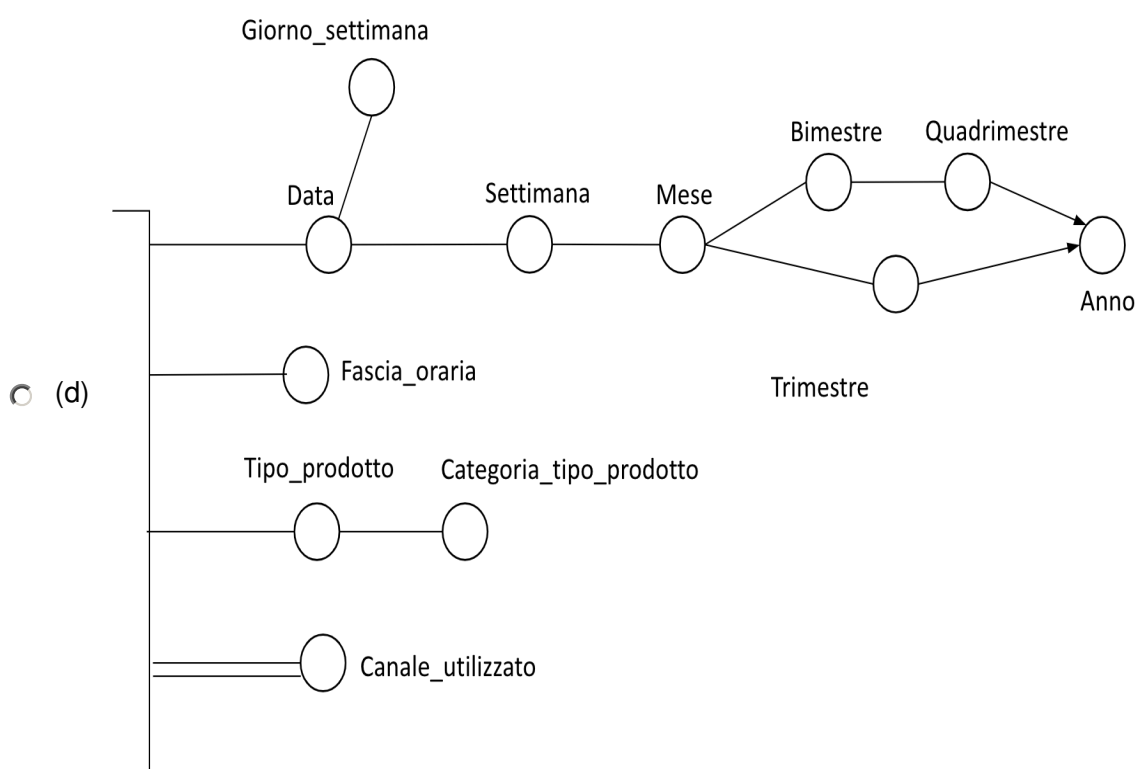
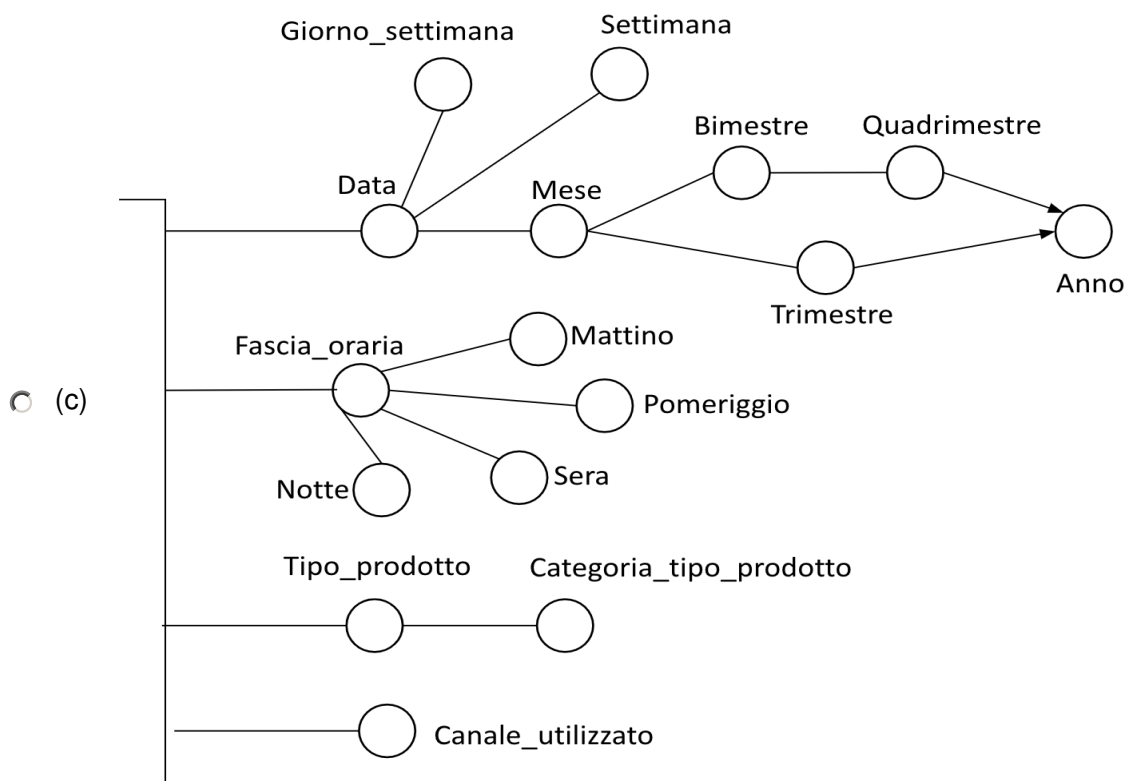
Ciascun call center fornisce un servizio multilingua e multicanale per contattare i potenziali clienti e proporre i prodotti. I diversi canali di comunicazione utilizzati da ciascun call center includono, oltre alle chiamate telefoniche, ad esempio anche le email e i messaggi sms. Ciascun call center è inoltre in grado di operare supportando più lingue straniere (ad esempio inglese, francese e tedesco). Pertanto ciascun call center può contattare clienti presenti in diversi stati nel mondo.

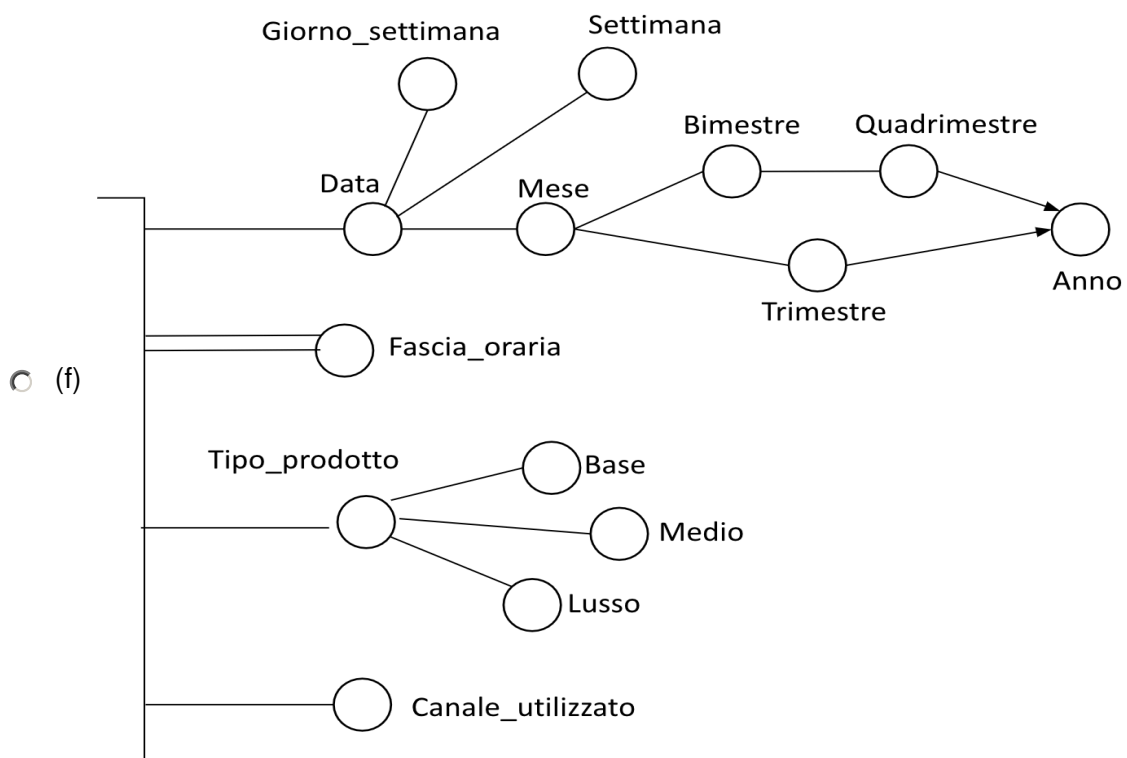
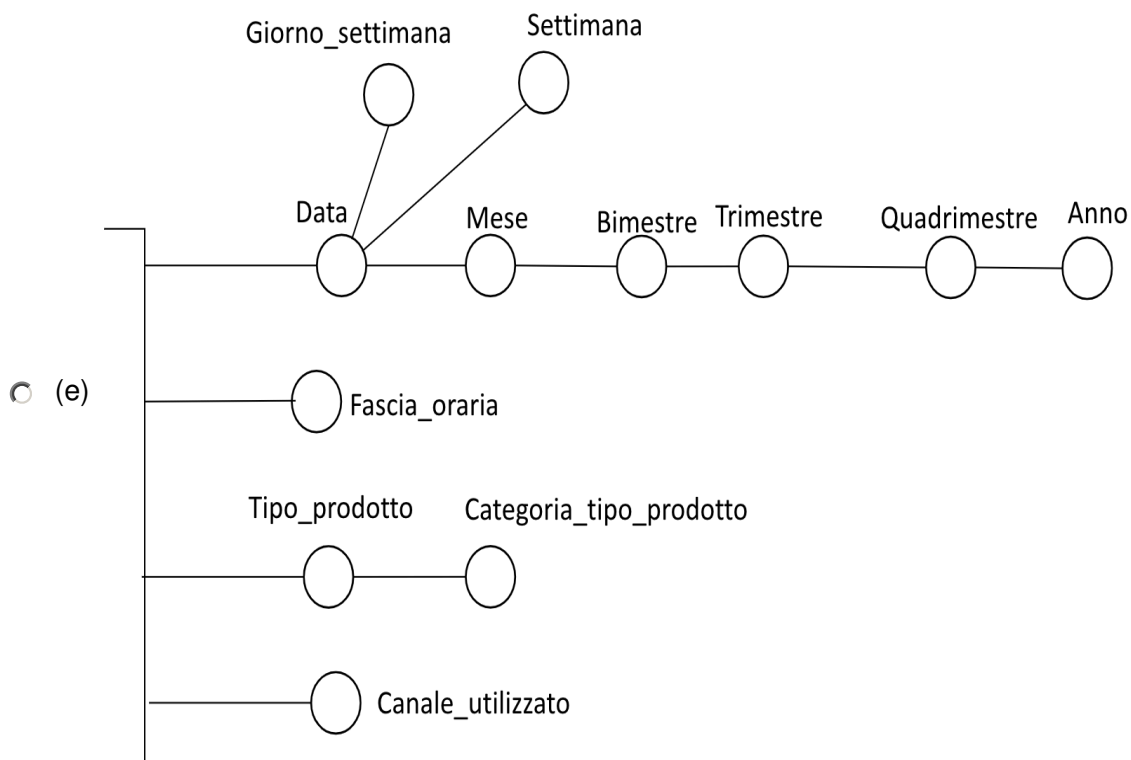
Si vuole analizzare (1) il numero medio di clienti contattati per prodotto e (2) il numero medio di risposte positive ricevute per prodotto, in base a:

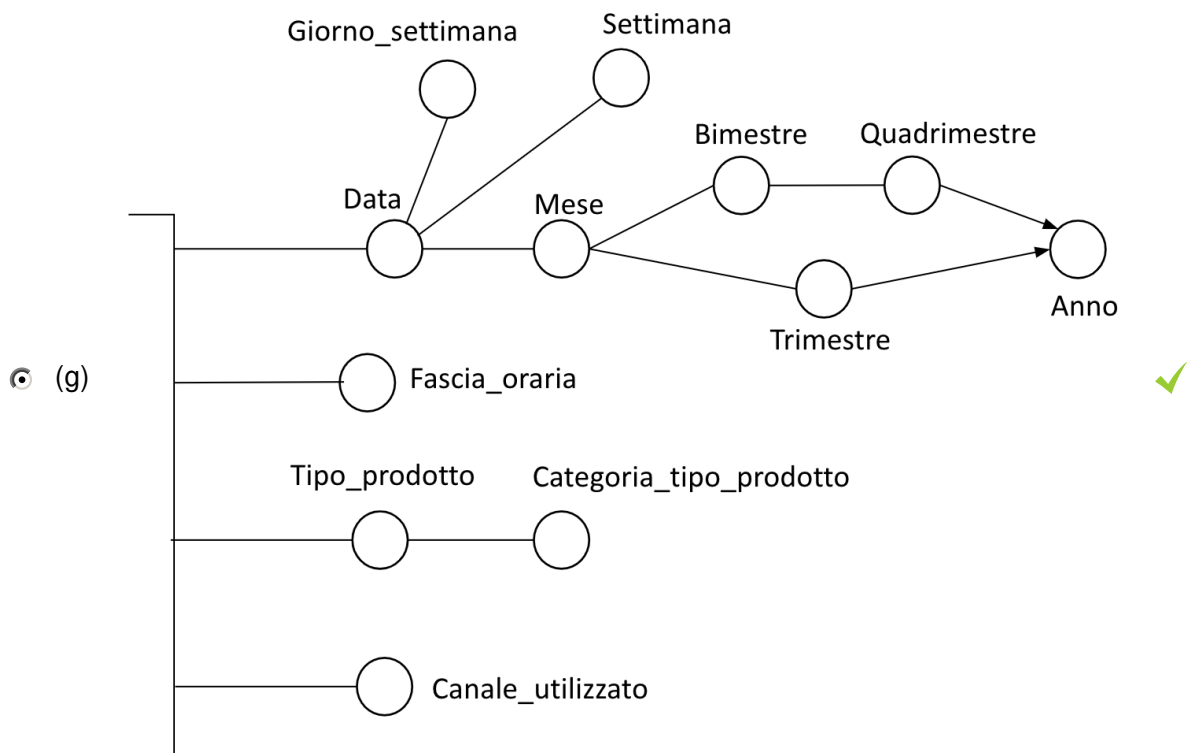
- call center, caratterizzato da un nome, dalla sua collocazione geografica (espressa in termini di città, stato e continente in cui si trova il call center), dalla dimensione del call center (espressa come piccola, media, o grande, in base al numero di dipendenti) e dall'indicazione delle lingue supportate dal call center (uno o più lingue tra inglese, francese, italiano, tedesco e spagnolo)
- informazione temporale su quando il call center ha contattato il cliente, espressa in termini di data, giorno della settimana, settimana, mese, bimestre, trimestre, quadrimestre, anno e fascia oraria (un valore tra mattino, pomeriggio, sera, notte)
- tipo di prodotto proposto al cliente contattato
- categorizzazione del tipo di prodotto in base, medio e lusso
- tipo di canale utilizzato per contattare il cliente (ossia chiamata telefonica oppure email o messaggio sms)
- stato e continente del cliente contattato
- caratteristiche del cliente contattato, in termini di genere, fascia di età del cliente (un valore tra meno di 20 anni, tra 20 e 40 anni, tra 40 e 60 anni, e più di 60 anni) e tipo di occupazione del cliente (ad esempio impiegato oppure libero professionista o pensionato)

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



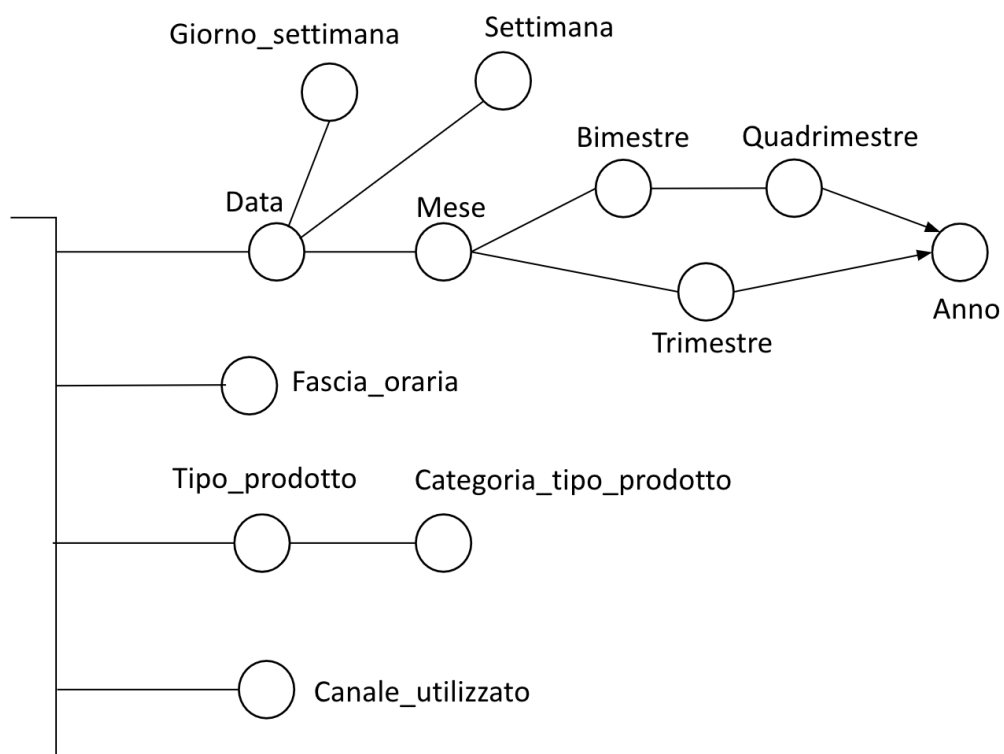






Risposta corretta.

La risposta corretta è:



Domanda 6

Risposta corretta

Punteggio ottenuto 1,00 su 1,00

1 punto (penalità 15% per risposta sbagliata)

Si vogliono analizzare le informazioni relative ad una catena di call center che svolge attività per campagne di vendita su diversi tipi di prodotti forniti da una società multinazionale.

Ciascun call center fornisce un servizio multilingua e multicanale per contattare i potenziali clienti e proporre i prodotti. I diversi canali di comunicazione utilizzati da ciascun call center includono, oltre alle chiamate telefoniche, ad esempio anche le email e i messaggi sms. Ciascun call center è inoltre in grado di operare supportando più lingue straniere (ad esempio inglese, francese e tedesco). Pertanto ciascun call center può contattare clienti presenti in diversi stati nel mondo.

Si vuole analizzare (1) il numero medio di clienti contattati per prodotto e (2) il numero medio di risposte positive ricevute per prodotto, in base a:

- call center, caratterizzato da un nome, dalla sua collocazione geografica (espressa in termini di città, stato e continente in cui si trova il call center), dalla dimensione del call center (espressa come piccola, media, o grande, in base al numero di dipendenti) e dall'indicazione delle lingue supportate dal call center (uno o più lingue tra inglese, francese, italiano, tedesco e spagnolo)
- informazione temporale su quando il call center ha contattato il cliente, espressa in termini di data, giorno della settimana, settimana, mese, bimestre, trimestre, quadrimestre, anno e fascia oraria (un valore tra mattino, pomeriggio, sera, notte)
- tipo di prodotto proposto al cliente contattato
- categorizzazione del tipo di prodotto in base, medio o lusso
- tipo di canale utilizzato per contattare il cliente (ossia chiamata telefonica oppure email o messaggio sms)
- stato e continente del cliente contattato
- caratteristiche del cliente contattato, in termini di genere, fascia di età del cliente (un valore tra meno di 20 anni, tra 20 e 40 anni, tra 40 e 60 anni, e più di 60 anni) e tipo di occupazione del cliente (ad esempio impiegato oppure libero professionista o pensionato)

Selezionare dall'elenco tutti e soli gli attributi necessari per modellare correttamente le misure nella tabella dei fatti richieste dalle specifiche (sono possibili più risposte corrette).

Scegli una o più alternative:

- ☐ (a) Numero totale call center
- ☐ (b) Numero totale di chiamate
- ☒ (c) Numero totale di prodotti proposti ✓
- ☐ (d) Numero medio di chiamate per call center

- ☐ (e) Numero totale operatori
- ☒ (f) Numero totale di risposte positive ✓
- ☒ (g) Numero totale clienti contattati ✓
- ☐ (h) Numero medio giornaliero di risposte positive

Risposta corretta.

La risposta corretta è: Numero totale clienti contattati, Numero totale di risposte positive, Numero totale di prodotti proposti

Domanda 7

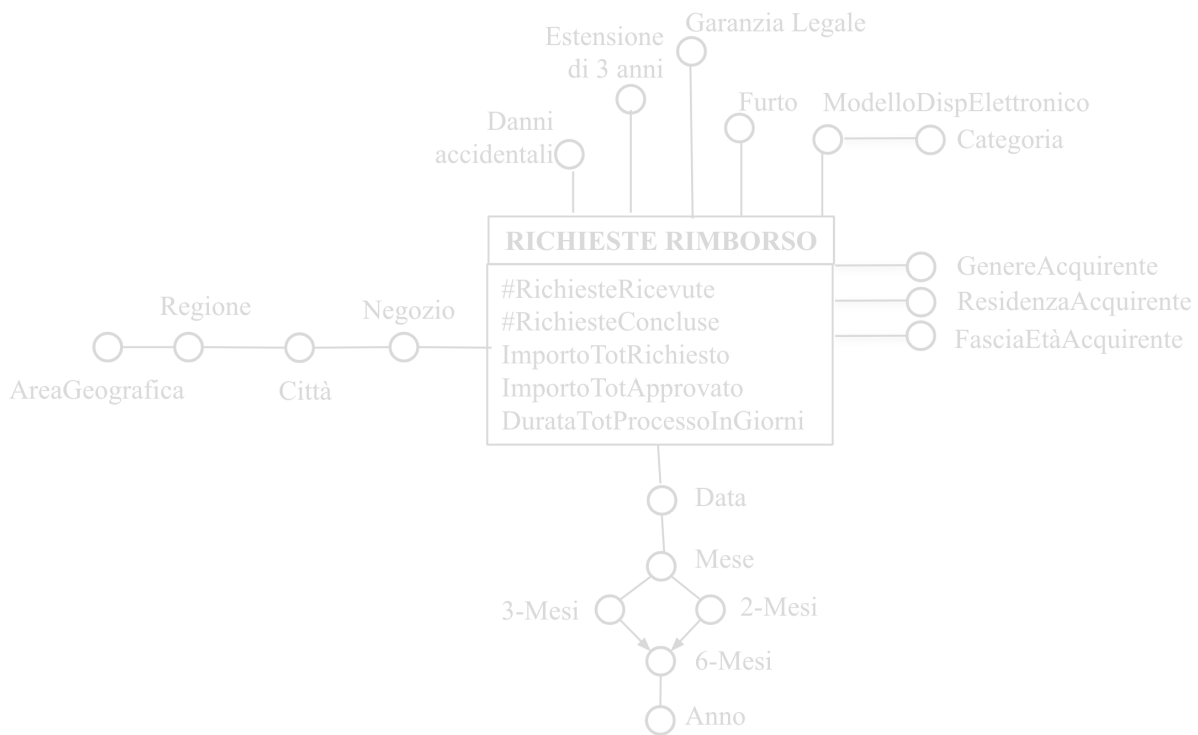
Completo

Punteggio ottenuto 2,50 su 3,00

3 punti (nessuna penalità per risposta sbagliata)

Il seguente data warehouse descrive l'andamento nel tempo delle RICHIESTE DI RIMBORSO ricevute dai diversi negozi a valle delle sottoscrizioni di coperture assicurative per modelli di dispositivi elettronici. Nello specifico, l'analisi deve essere effettuata in funzione del modello del dispositivo elettronico e relativa categoria, elenco delle coperture assicurative sottoscritte (GaranziaLegale, Estensione3Anni, DanniAccidentali, Furto). E' possibile sottoscrivere più coperture assicurative per lo stesso modello di dispositivo (questo elenco rappresenta una configurazione e ogni attributo assume valore Y/N). Il data warehouse memorizza il genere, la residenza e la fascia di età dell'acquirente. Per ogni negozio che gestisce le richieste di rimborso sono noti la città, la regione e la corrispondente area geografica. Le metriche da analizzare sono il numero di richieste ricevute, il numero di richieste concluse, l'importo totale richiesto, l'importo totale approvato, la durata del tempo di processamento delle richieste espresso in giorni (DurataTotProcesso). Le metriche devono essere analizzate per ogni data, mese, 2-mese, 3-mese, 6-mese, anno.

Il data warehouse è caratterizzato dal seguente schema concettuale e dal corrispondente schema logico.



NEGOZIO(IDNegozio, Negozio, Città, Regione, AreaGeografica)
 MODELLO-DISP-ELETTRONICO (IDModDispEle, ModelloDispElettronico, Categoria)
 JUNK-CARAT-COPERTURE-ASSICURATIVE (IDJCCA, GaranziaLegale, Estensione3Anni, DanniAccidentali, Furto)
 JUNK-CARAT-ACQUIRENTE (IDJCA, Genere, Residenza, FasciaEtà)
 TEMPO(IDTempo, Data, Mese, 2-Mesi, 3-Mesi, 6-Mesi, Anno,)
 RICHIESTE-RIMBORSO (IDNegozio, IDModDispEle, IDJCCA, IDJCA, IDTempo, #RichiesteRicevute, #RichiesteConcluse, ImportoTotRichiesto, ImportoTotApprovato, DurataTotProcesso)

Considerando le coperture assicurative che includono solo la garanzia legale (attributo GaranziaLegale), separatamente per trimestre e città, visualizzare:

- La percentuale di richieste concluse rispetto a quelle ricevute
- La durata media del tempo di processingo per richiesta ricevuta

Associare ad ogni record visualizzato la posizione in un ranking:

- in funzione dell'importo complessivo approvato separatamente per area geografica (1 per il record con il più basso valore complessivo approvato)
- in funzione della differenza tra importo richiesto e importo approvato (1 per il record con il più alto valore della differenza tra importo richiesto e importo approvato complessivi) separatamente per anno

```

SELECT 3M, CITTA, 100*SUM(#RICHIESTECONCLUSE)/SUM(#RICHIESTERICEVUTE),
SUM(DURATATOTPROCESSO)/SUM(#RICHIESTERICEVUTE),
RANK() OVER (PARTITION BY AREAGEOGRAFICA ORDER BY

```

```
SUM(IMPORTOTOTAPPROVATO)),  
    RANK() OVER (PARTITION BY ANNO ORDER BY SUM(IMPORTOTOTRICHIESTO)-  
SUM(IMPORTOTOTAPPROVATO) DESC)
```

```
FROM TEMPO T, NEGOZIO N, RICHIESTE-RIMBORSO RR, JUNK-CARAT-COPERTURE-  
ASSICURATIVE JC
```

```
WHERE T.IDTEMPO=RR.IDTEMPO AND N.IDNEGOZIO=RR.IDNEGOZIO AND  
JC.IDJCCA=NN.IDJCCA AND GARANZIALEGALE=1
```

```
GROUP BY 3M, CITTA, AREAGEOGRAFICA, ANNO
```

```
Select 3-Mesi, Città, 100*SUM(#RichiesteConcluse)/SUM(#RichiesteRicevute),  
SUM(DurataTotProcesso)/SUM(#RichiesteRicevute)  
RANK() OVER (PARTITION BY AreaGeografica ORDER BY SUM(ImportoTotApprovato) )  
RANK() OVER (PARTITION BY Anno ORDER BY SUM(ImportoTotRichiesto) -  
SUM(ImportoTotApprovato) DESC)  
FROM NEGOZIO, TEMPO, JUNK-CARAT-COPERTURE-ASSICURATIVE, RICHIESTE-  
RIMBORSO  
WHERE join AND GaranziaLegale = 'Y' AND Estensione3Anni = 'N' AND DanniAccidentali = 'N'  
AND Furto= 'N'  
GROUP BY 3-Mesi, Città, AreaGeografica, Anno
```

Commento:
Condizione specificata nella WHERE incompleta SOLO)

Domanda 8

Risposta non data

Punteggio max.: 1,00

1 punto (-15% penalità per risposta sbagliata)

Un albero decisionale viene addestrato su un dataset contenente N punti, M attributi e P classi.

Qual e' il numero massimo di nodi foglia che puo' avere l'albero addestrato?

- ☐ (a) M nodi
- ☐ (b) $P + M$ nodi
- ☐ (c) P nodi
- ☐ (d) N nodi
- ☐ (e) $N + P$ nodi
- ☐ (f) Nessuna delle altre risposte e' corretta
- ☐ (g) $N + M$ nodi
- ☐ (h) Non e' possibile rispondere con le informazioni a disposizione

Risposta errata.

La risposta corretta è: N nodi

Domanda 9

Risposta corretta

Punteggio ottenuto 1,00 su 1,00

1 punto (-15% per risposta sbagliata)

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

B(T0) CK(T0) CK(T0) B(T1) I1(o4) Commit(T1) Commit(T0) B(T2) I2(o0) B(T3) B(T4) U2(o3) U4(o4) I2(o3) I2(o2) FAILURE

Notazione:

- Tn: Id della transazione n
- B(Tn): Begin della transazione Tn
- CK(Ta,Tb,...): checkpoint delle transazioni Ta, Tb, ...
- 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 operazioni vengono eseguite per un warm restart?

- ☐ (a) UNDO T0,T1 REDO T2,T3,T4
- ☐ (b) UNDO T3,T4 REDO T0,T1,T2
- ☐ (c) UNDO T0,T1,T2 REDO T3,T4
- ☐ (d) UNDO T1,T2,T3,T4 REDO T0
- ☐ (e) UNDO T0,T3,T4 REDO T1,T2
- ☐ (f) Nessuna delle altre risposte e' corretta
- ☒ (g) UNDO T2,T3,T4 REDO T0,T1 ✓

Risposta corretta.

La risposta corretta è: UNDO T2,T3,T4 REDO T0,T1

Domanda 10

Completo

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.

Nell'esercizio 5 ho inteso settimana come settimana dell'anno.

Ad esempio 07/01/2023 corrisponde alla settimana 2, ,ma questa settimana include più date (7)

nella domanda 1 sicuramente {bc} è il cluster ottenuto al primo passo, per proseguire il calcolo si dovrebbe ricalcolare la matrice delle distanze

Domanda 11

Completo

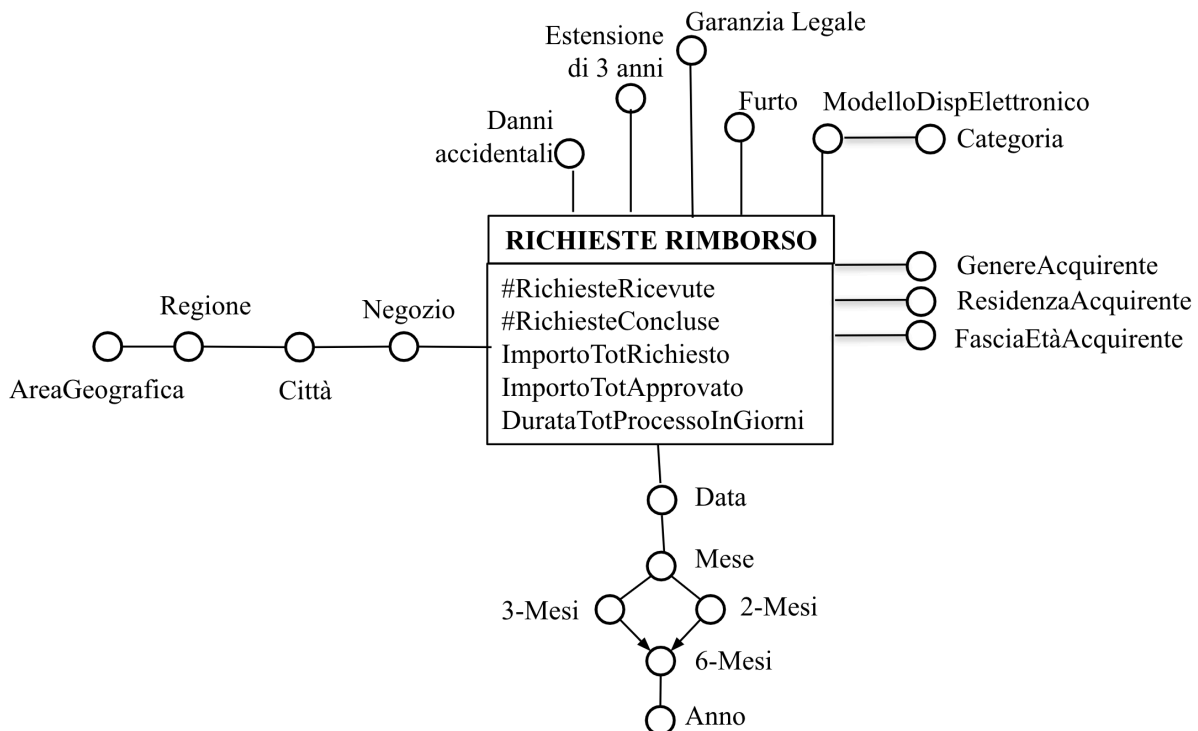
Punteggio ottenuto 3,50 su 4,00

4 punti (nessuna penalità per risposta sbagliata)

Il seguente data warehouse descrive l'andamento nel tempo delle RICHIESTE DI RIMBORSO ricevute dai diversi negozi a valle delle sottoscrizioni di coperture assicurative per modelli di dispositivi elettronici. Nello specifico, l'analisi deve essere effettuata in funzione del modello del dispositivo elettronico e relativa categoria, elenco delle coperture assicurative sottoscritte (GaranziaLegale, Estensione3Anni, DanniAccidentali, Furto). E' possibile sottoscrivere più coperture assicurative per lo stesso modello di dispositivo (questo elenco rappresenta una configurazione e ogni attributo assume valore Y/N). Il data warehouse memorizza il genere, la residenza e la fascia di età dell'acquirente. Per ogni negozio che gestisce le richieste di rimborso sono noti la città, la regione e la corrispondente area geografica. Le metriche da analizzare sono il numero di richieste ricevute, il numero di richieste concluse, l'importo totale richiesto, l'importo totale approvato, la durata del tempo di processamento delle richieste espresso in giorni

(DurataTotProcesso). Le metriche devono essere analizzate per ogni data, mese, 2-mese, 3-mese, 6-mese, anno.

Il data warehouse è caratterizzato dal seguente schema concettuale e dal corrispondente schema logico.



NEGOZIO(IDNegozio, Negozio, Città, Regione, AreaGeografica)
 MODELLO-DISP-ELETTRONICO (IDModDispEle, ModelloDispElettronico, Categoria)
 JUNK-CARAT-COPERTURE-ASSICURATIVE (IDJCCA, GaranziaLegale, Estensione3Anni, DanniAccidentali, Furto)
 JUNK-CARAT-ACQUIRENTE (IDJCA, Genere, Residenza, FasciaEtà)
 TEMPO(IDTempo, Data, Mese, 2-Mesi, 3-Mesi, 6-Mesi, Anno,)
 RICHIESTE-RIMBORSO (IDNegozio, IDModDispEle, IDJCCA, IDJCA, IDTempo, #RichiesteRicevute, #RichiesteConcluse, ImportoTotRichiesto, ImportoTotApprovato, DurataTotProcesso)

Considerando gli anni precedenti al 2020, separatamente per modello del dispositivo elettronico, semestre (attributo 6-Mesi) e genere dell'acquirente, visualizzare:

- L'importo approvato medio per richiesta conclusa
- L'importo cumulativo approvato al trascorrere dei semestri, separatamente per modello del dispositivo elettronico
- L'importo complessivo richiesto indipendentemente dal modello del dispositivo, semestre e Genere

```
SELECT MODELLODISPELETTRONICO, 6M, GENERE,
SUM(IMPORTOTOTAPPROVATO)/SUM(#RICHIESTECONCLUSE),
```



```

SUM(SUM(IMPORTOTOTAPPROVATO)) OVER (PARTITION BY
MODELLODISPELETTRONICO, 6M
ORDER BY MESE ROWS UNBOUNDED PRECEDING)
SUM(SUM(IMPORTOTOTRICHIESTO)) OVER ()

FROM MODELLO-DISP-ELETTRONICO M, TEMPO T, JUNK-CARAT-ACQUIRENTE J,
RICHIESTE-RIMBORSO R

WHERE M.IDMODDISPELE=R.IDMODDISPELE AND T.IDTEMPO=R.IDTEMPO AND
J.IDJCA=N.JCA AD ANNO<2020

GROUP BY MODELLODISPELETTRONICO, 6M, GENERE

```

```

Select ModelloDispElettronico, 6-Mesi, Genere, SUM(ImportoTotApprovato)/SUM(
#RichiesteConcluse),
SUM(SUM(ImportoTotApprovato)) OVER (PARTITION BY ModelloDispositivo ORDER BY 6-Mesi
ROWS UNBOUNDED PRECEDING)
SUM(SUM(ImportoTotRichiesto)) OVER ()
FROM JUNK-CARAT-ACQUIRENTE, TEMPO, RICHIESTE-RIMBORSO, MODELLO-DISP-
ELETTRONICO
WHERE join AND Anno < 2020
GROUP BY ModelloDispElettronico, 6-Mesi, Genere

```

Commento:
Seconda metrica: attributo in più nella partition

Domanda 12

Completo

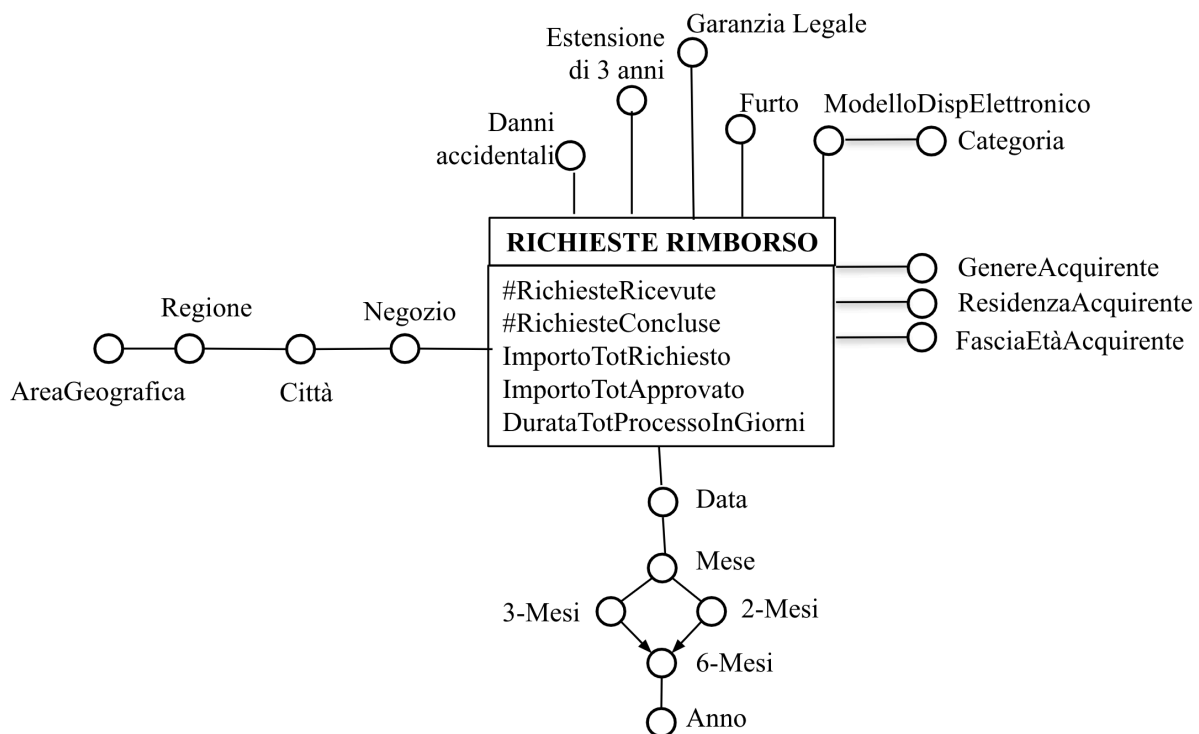
Punteggio ottenuto 4,00 su 5,00

5 punti (nessuna penalità per risposta sbagliata)

Il seguente data warehouse descrive l'andamento nel tempo delle RICHIESTE DI RIMBORSO ricevute dai diversi negozi a valle delle sottoscrizioni di coperture assicurative per modelli di dispositivi elettronici. Nello specifico, l'analisi deve essere effettuata in funzione del modello del dispositivo elettronico e relativa categoria, elenco delle coperture assicurative sottoscritte (GaranziaLegale, Estensione3Anni, DanniAccidentali, Furto). E' possibile sottoscrivere più coperture assicurative per lo stesso modello di dispositivo (questo elenco rappresenta una configurazione e ogni attributo assume valore Y/N). Il data warehouse memorizza il genere, la residenza e la fascia di età dell'acquirente. Per ogni negozio che gestisce le richieste di rimborso

sono noti la città, la regione e la corrispondente area geografica. Le metriche da analizzare sono il numero di richieste ricevute, il numero di richieste concluse, l'importo totale richiesto, l'importo totale approvato, la durata del tempo di processamento delle richieste espresso in giorni (DurataTotProcesso). Le metriche devono essere analizzate per ogni data, mese, 2-mese, 3-mese, 6-mese, anno.

Il data warehouse è caratterizzato dal seguente schema concettuale e dal corrispondente schema logico.



NEGOZIO(IDNegozio, Negozio, Città, Regione, AreaGeografica)

MODELLO-DISP-ELETTRONICO (IDModDispEle, ModelloDispElettronico, Categoria)

JUNK-CARAT-COPERTURE-ASSICURATIVE (IDJCCA, GaranziaLegale, Estensione3Anni, DanniAccidentali, Furto)

JUNK-CARAT-ACQUIRENTE (IDJCA, Genere, Residenza, FasciaEtà)

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

RICHIESTE-RIMBORSO (IDNegozio, IDModDispEle, IDJCCA, IDJCA, IDTempo, #RichiesteRicevute, #RichiesteConcluse, ImportoTotRichiesto, ImportoTotApprovato, DurataTotProcesso)

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

- Considerando l'area geografica nord, separatamente per regione e anno, visualizzare il numero complessivo di richieste concluse, l'importo complessivo approvato e la durata media mensile di processamento delle richieste.
- Considerando le coperture assicurative che includono i danni accidentali (attributo

DanniAccidentali), ma non il furto (attributo Furto), separatamente per mese e area geografica, visualizzare il numero complessivo di richieste concluse e la differenza tra l'importo complessivo richiesto e quello approvato.

c. Considerando gli anni 2021 e 2022, separatamente per semestre (attributo 6-Mesi) e regione, visualizzare il numero complessivo di richieste concluse e il corrispondente importo medio approvato.

Dato lo schema logico precedente, si svolgano le seguenti attività

1. 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 ViewRimborsi  
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 ViewRimborsi.

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 ViewRimborsi in caso di inserimento di un nuovo record nella tabella dei fatti RICHIESTE-RIMBORSO.

--pseudo query--

a)

```
SELECT REGIONE, ANNO, SUM(#RICHIESTECONCLUSE),  
SUM(IMPORTOTOTAPPROVATO), SUM(DURATATOTPROCESSAMENTO),  
SUM(#RICHIESTERICEVUTE)  
FROM RICHIESTE-RIMBORSO R, TEMPO T, NEGOZIO N  
WHERE JOIN AND AREAGEOGRAFICA=NORD  
GB REGIONE, ANNO
```

b)

```
SELECT SUM(#RICHIESTECONCLUSE), SUM(IMPORTOTOTRICHIESTO),  
SUM(IMPORTOTOTAPPROVATO)  
FROM JUNK-CARAT-COPERTURE-ASSICURATIVE JCC, NEGOZIO N, TEMPO T, RICHIESTE-  
RIMBORSO R  
WHERE JOIN AND DANNIACCIDENTALI=1 AND FURTO=0  
GB MESE, AREAGEOGRAFICA
```

c)

```
SELECT SUM(#RICHIESTECONCLUSE), SUM(IMPORTOTOTAPPROVATO)  
FROM TEMPO T, NEGOZIO N, RICHIESTE-RIMBORSO R  
WHERE ANNO>=2020 AND ANNO<=2021  
GB 6M, REGIONE
```

Blocco A

```
SELECT MESE, REGIONE, 6M, ANNO, AREAGEOGRAFICA, DANNIACCIDENTALI, FURTO,  
    SUM(#RICHIESTECONCLUDE) AS CONCLUDE,  
    SUM(#IMPORTOTOTAPPROVATP) AS IMPAPP,  
    SUM(DURATATOTPROCESSAMENTO) AS DURATA,  
    SUM(#RICHIESTERICEVUTE) AS RICEVUTE,  
    SUM(IMPORTOTOTRICHIESTO) AS IMPRIC  
FROM TEMPO T, NEGOZIO N, JUNK-CARAT-COPERTURE-ASSICURATIVE J, RICHIESTE-  
RIMBORSO R  
WHERE T.IDTEMPO=R.IDTEMPO AND N.IDNEGOZIO=R.IDNEGOZIO AND  
J.IDJCCA=R.IDJCCA  
GROUP BY MESE, REGIONE, 6M, ANNO, AREAGEOGRAFICA, DANNIACCIDENTALI, FURTO
```

--insieme minimale: MESE, REGIONE

```
CREATE OR REPLACE TRIGGER AGG_RIMBORSO  
AFTER INSERT ON RICHIESTE-RIMBORSO  
FOR EACH ROW  
DECLARE  
MYMESE, MY6M, MYANNO DATE;  
MYREG, MYAREA VARCHAR(20);  
MYDANNI, MYFURTO VARCHAR(5);  
N NUMBER;  
BEGIN  
  
SELECT MESE, 6M, ANNO INTO MYMESE, MY6M, MYANNO  
FROM TEMPO  
WHERE IDTEMPO=:NEW.IDTEMPO;  
  
SELECT REGIONE, AREAGEOGRAFICA INTO MYREG, MYAREA  
FROM NEGOZIO  
WHERE IDNEGOZIO=:NEW.IDNEGOZIO;  
  
SELECT DANNIACCIDENTALI, FURTO INTO MYDANNI, MYFURTO  
FROM JUNK-CARAT-COPERTURE-ASSICURATIVE  
WHERE IDJCCA=:NEW.IDJCCA;  
  
SELECT COUNT(*) INTO N  
FROM VIEWRIMBORSI  
WHERE MESE=MYMESE AND REGIONE=MYREG;  
  
IF (N>0) THEN  
    UPDATE VIEWRIMBORSI  
    SET CONCLUDE=CONCLUDE+:NEW.#RICHIESTECONCLUDE AND  
    IMPAPP=IMPAPP+:NEW.IMPORTOTOAPPROVATO AND  
    DURATA=DURATA+:NEW.DURATATOTPROCESSAMENTO AND  
    RICEVUTE=RICEVUTE+:NEW.#RICHIESTERICEVUTE AND
```

```

    IMPRIC=IMPRIC+:NEW.IMPORTOTOTRICHIESTO
    WHERE MESE=MYMESE AND REGIONE=MYREGIONE;
ELSE
    INSERT VIEWRIMBORSI(...)
    VALUES(MYMESE, MYREG,MY6M, MYANNO, MYAREA, MYDANNI, MYFURTO,
:NEW.#RICHIESTERICEVUTE, :NEW.IMPORTOTOTAPPROVATO,
        :NEW.DURATATOTPROCESSAMENTO, :NEW.#RICHIESTERICEVUTE,
:NEW.IMPORTOTOTRICHIESTO);
END IF;
END;

```

Query di interesse:

- (a) Select Regione, Anno, SUM(#RichiesteConcluse), SUM(ImportoTotApprovato),
SUM(DurataTotProcesso)/COUNT(DISTINCT Mese)
FROM NEGOZIO, TEMPO, RICHIESTE-RIMBORSO
WHERE join AND AreaGeografica = 'Nord'
GROUP BY Regione, Anno
- (b) Select Mese, AreaGeografica, SUM(#RichiesteConcluse), SUM(ImportoTotRichiesto) -
SUM(ImportoTotApprovato)
FROM NEGOZIO, TEMPO, JUNK-CARAT-COPERTURE-ASSICURATIVE, RICHIESTE-RIMBORSO
WHERE join AND DanniAccidentali = 'Y' AND Furto = 'N'
GROUP BY Mese, AreaGeografica
- (c) Select 6-Mesi, Regione, [Anno,] SUM(#RichiesteConcluse),
SUM(ImportoTotApprovato)/SUM(#RichiesteConcluse)
FROM JUNK-CARATTERISTICHE-CLIENTE, TEMPO, SOTTOSCRIZIONI
WHERE join AND (Anno = 2021 OR Anno = 2022)
GROUP BY 6-Mesi, Regione [, Anno]

1. Blocco A – Query per vista materializzata

```

SELECT Mese, 6-Mesi, Anno, Regione, Area Geografica, DanniAccidentali, Furto,
SUM(#RichiesteConcluse) AS RichiesteTot, SUM(ImportoTotRichiesto) AS ImportoRichiestoTot,
SUM(ImportoTotApprovato) AS ImportoApprovatoTot, SUM(DurataTotProcesso) AS DurataTot
FROM JUNK-CARAT-COPERTURE-ASSICURATIVE C, NEGOZIO N, TEMPO T, RICHIESTE-RIMBORSO R
WHERE C.IDJCCA = R.IDJCCA AND N.IDNegozio = R.IDNegozio AND T.IDTempo = R.IDTempo
GROUP BY Mese, 6-Mesi, Anno, Regione, Area Geografica, DanniAccidentali, Furto

```

2. Identificatore

Mese, Regione, DanniAccidentali, Furto

3. Trigger

```

CREATE OR REPLACE TRIGGER MaintenanceViewRimborsi
AFTER INSERT ON RICHIESTE-RIMBORSO
FOR EACH ROW

DECLARE
VarAnno, VarSemestre VarMese DATE;
VarDanniAccidentali, VarFurto BOOLEAN;
VarRegione, VarAreaGeo VARCHAR(10);
N INTEGER;

BEGIN
    SELECT Mese, 3-Mesi, Anno INTO VarMese, VarTrimestre, VarAnno
    FROM TEMPO
    WHERE IDTempo = :NEW.IDTempo;

    SELECT DanniAccidentali, Furto INTO VarDanniAccidentali, VarFurto
    FROM JUNK-CARAT-COPERTURE-ASSICURATIVE
    WHERE IDJCCA = :NEW.IDJCCA;

    SELECT Regione, AreaGeografica INTO VarRegione, VarAreaGeo
    FROM NEGOZIO
    WHERE IDNegozio = :NEW.IDNegozio;

    SELECT COUNT(*) INTO N
    FROM ViewRimborsi
    WHERE Mese = VarMese AND DanniAccidentali = VarDanniAccidentali AND Furto =
VarFurto AND Regione = VarRegione;

    IF N>0 THEN
        UPDATE ViewRimborsi
        SET RichiesteTot = RichiesteTot + :NEW.#RichiesteConclude, ImportoRichiestoTot =
ImportoRichiestoTot + :NEW.ImportoTotRichiesto,
        ImportoApprovatoTot = ImportoApprovatoTot + :NEW.ImportoTotApprovato, DurataTot =
DurataTot + :NEW.DurataTotProcesso
        WHERE Mese = VarMese AND DanniAccidentali = VarDanniAccidentali AND Furto =
VarFurto AND Regione = VarRegione;
    ELSE
        INSERT INTO ViewRimborsi(...) VALUES (VarMese, VarTrimestre, varAnno,
VarRegione, VarAreaGeo, VArDanniAccidentali, VarFurto, :NEW.#RichiesteConclude,
:NEW.ImportoTotRichiesto, :NEW.ImportoTotApprovato, :NEW.DurataTotProcesso);
    END IF;

END;

```

Commento:

Errori

- Aggregato SUM(#RICHIESTERICEVUTE) non necessario in def. vista
- Mancano attributi Furto, DanniAccidentali in insieme minimale

Domanda 13

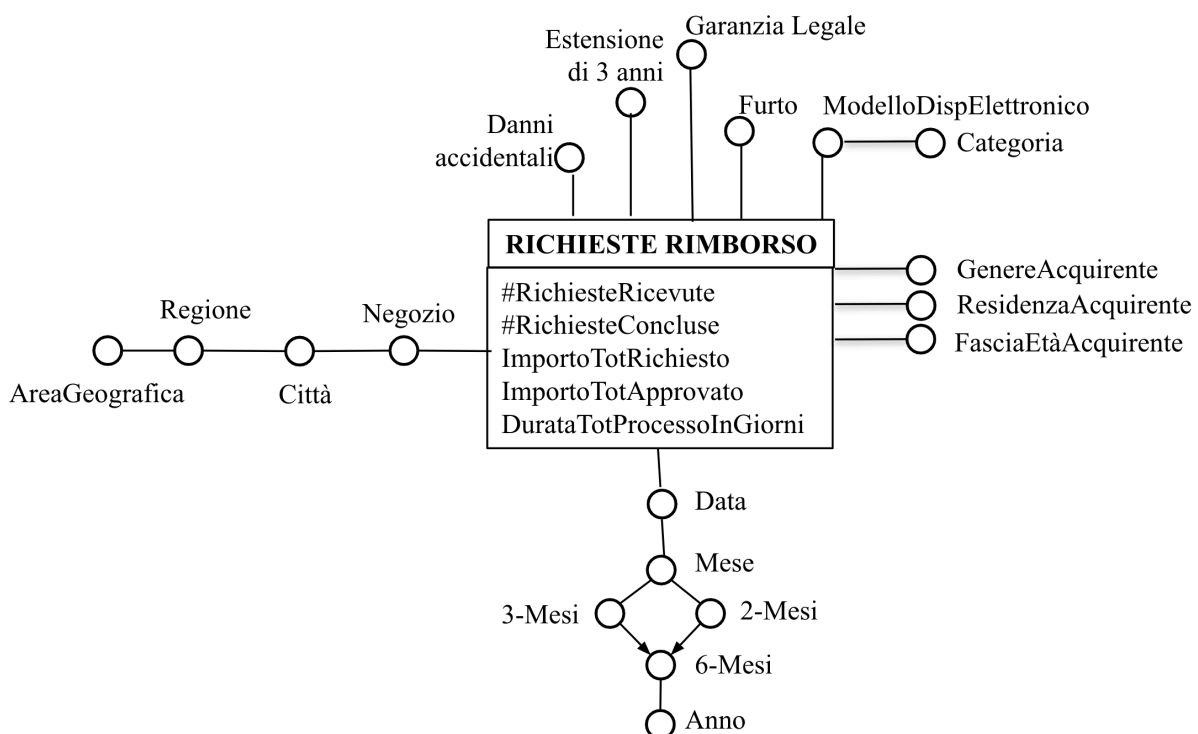
Completo

Punteggio ottenuto 4,00 su 4,00

4 punti (nessuna penalità per risposta sbagliata)

Il seguente data warehouse descrive l'andamento nel tempo delle RICHIESTE DI RIMBORSO ricevute dai diversi negozi a valle delle sottoscrizioni di coperture assicurative per modelli di dispositivi elettronici. Nello specifico, l'analisi deve essere effettuata in funzione del modello del dispositivo elettronico e relativa categoria, elenco delle coperture assicurative sottoscritte (GaranziaLegale, Estensione3Anni, DanniAccidentali, Furto). E' possibile sottoscrivere più coperture assicurative per lo stesso modello di dispositivo (questo elenco rappresenta una configurazione e ogni attributo assume valore Y/N). Il data warehouse memorizza il genere, la residenza e la fascia di età dell'acquirente. Per ogni negozio che gestisce le richieste di rimborso sono noti la città, la regione e la corrispondente area geografica. Le metriche da analizzare sono il numero di richieste ricevute, il numero di richieste concluse, l'importo totale richiesto, l'importo totale approvato, la durata del tempo di processamento delle richieste espresso in giorni (DurataTotProcesso). Le metriche devono essere analizzate per ogni data, mese, 2-mese, 3-mese, 6-mese, anno.

Il data warehouse è caratterizzato dal seguente schema concettuale e dal corrispondente schema logico.



NEGOZIO(**IDNegozio**, Negozio, Città, Regione, AreaGeografica)
 MODELLO-DISP-ELETTRONICO (**IDModDispEle**, ModelloDispElettronico, Categoria)
 JUNK-CARAT-COPERTURE-ASSICURATIVE (**IDJCCA**, GaranziaLegale, Estensione3Anni, DanniAc
 cidentali, Furto)
 JUNK-CARAT-ACQUIRENTE (**IDJCA**, Genere, Residenza, FasciaEtà)
 TEMPO(**IDTempo**, Data, Mese, 2-Mesi, 3-Mesi, 6-Mesi, Anno,)
 RICHIESTE-RIMBORSO (**IDNegozio**, **IDModDispEle**, **IDJCCA**, **IDJCA**, **IDTempo**, #RichiesteRicevu
 te, #RichiesteConcluse, ImportoTotRichiesto, ImportoTotApprovato, DurataTotProcesso)

Visualizzare separatamente per negozio, categoria del dispositivo elettronico e bimestre (attributo 2-mesi):

- La percentuale dell'importo richiesto che è stato approvato
- La differenza tra il numero di richieste ricevute e quelle concluse
- L'importo complessivo approvato indipendentemente dal negozio
- Il rapporto tra l'importo approvato e l'importo complessivo di tutti i negozi siti nella stessa città, separatamente per categoria e bimestre

```
SELECT NEGOZIO, CATEGORIA, 2M,
100*SUM(IMPORTOTOTAPPROVATO)/SUM(IMPORTOTOTRICHIESTO),
SUM(#RICHIESTERICEVUTE)-SUM(#RICHIESTECONCLUSE),
SUM(SUM(IMPORTOTOTAPPROVATO)) OVER (PARTITION BY CATEGORIA, 2M)
SUM(IMPORTOAPPROVATO)/SUM(IMPORTOTOTRICHIESTO) OVER (PARTITION BY
CATEGORIA, CITTA, 2M)

FROM TEMPO T, NEGOZIO N, MODELLO-DISP-ELETTRONICO M, RICHIESTE-RIMBORSO R

WHERE T.IDTEMPO=R.IDTEMPO AND N.IDNEGOZIO = R.IDNEGOZIO AND
M.IDMODDISPELE=R.IDMODDISPELE

GROUP BY NEGOZIO, CATEGORIA, 2M, CITTA
```

```
SELECT Negozio, Categoria, 2-Mesi,
100*SUM(ImportoTotApprovato)/SUM(ImportoTotRichiesto),
SUM(#RichiesteRicevute) - SUM( #RichiesteConcluse),
SUM(SUM(ImportoPremi)) OVER (PARTITION BY Categoria, 2-Mesi)
SUM(ImportoTotApprovato)/SUM(SUM(ImportoTotApprovato)) OVER (PARTITION BY Città,
Categoria, 2-Mesi)
FROM, RICHIESTE-RIMBORSO, TEMPO, MODELLO-DISP-ELETTRONICO, NEGOZIO(
WHERE join
GROUP BY Negozio, Categoria, 2-Mesi, Città
```

Commento:

Domanda 14

Risposta corretta

Punteggio ottenuto 1,00 su 1,00

1 punto (penalità 15% per risposta sbagliata)

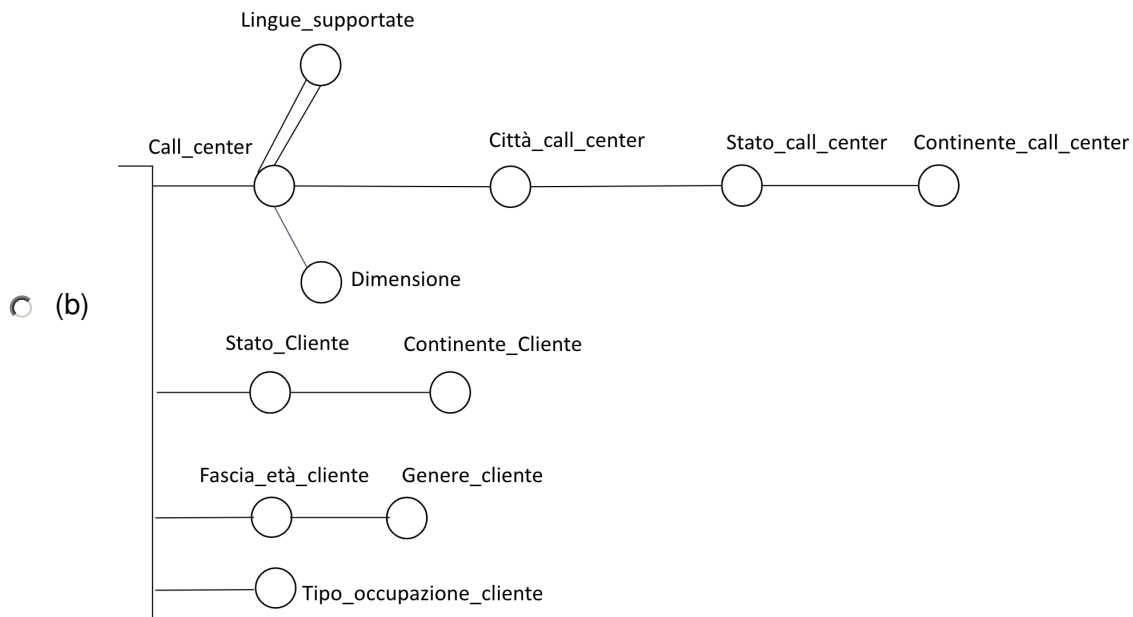
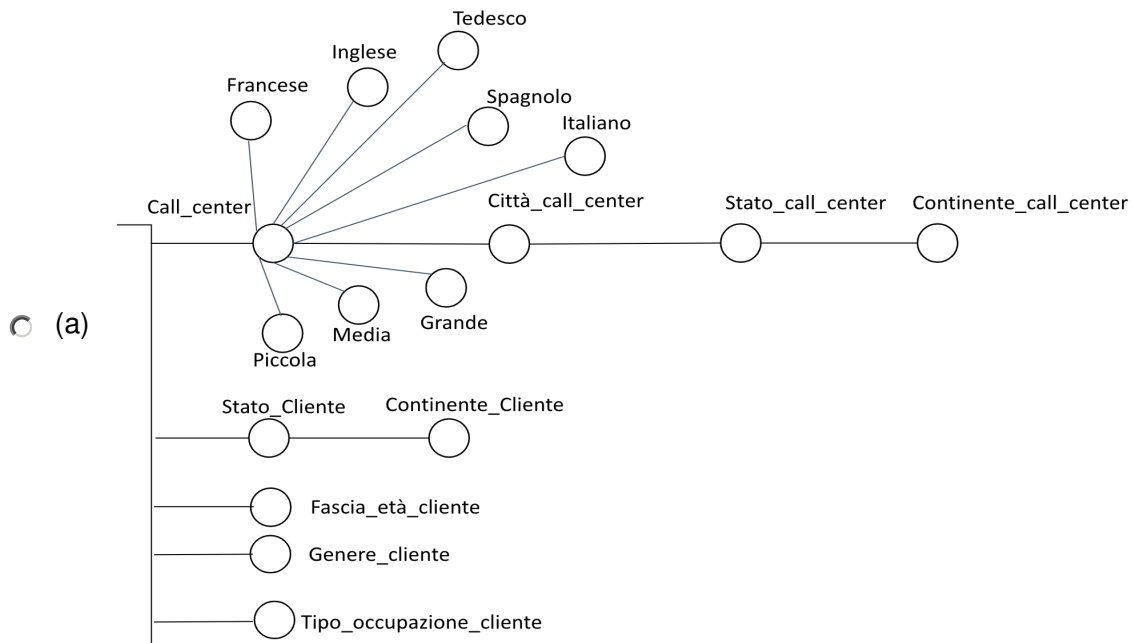
Si vogliono analizzare le informazioni relative ad una catena di call center che svolge attività per campagne di vendita su diversi tipi di prodotti forniti da una società multinazionale.

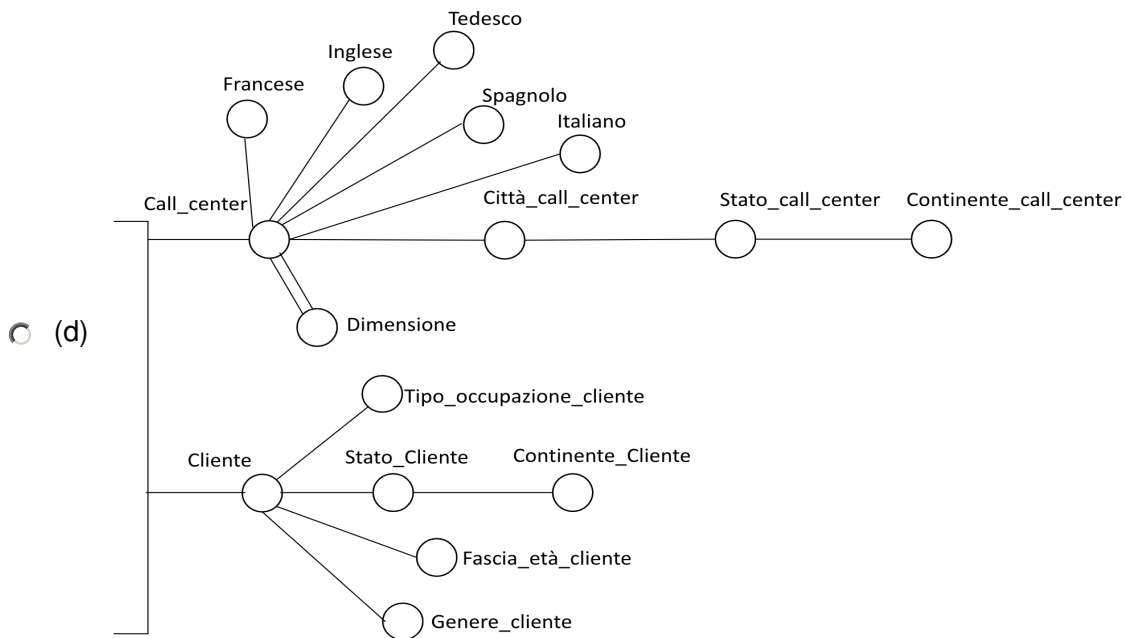
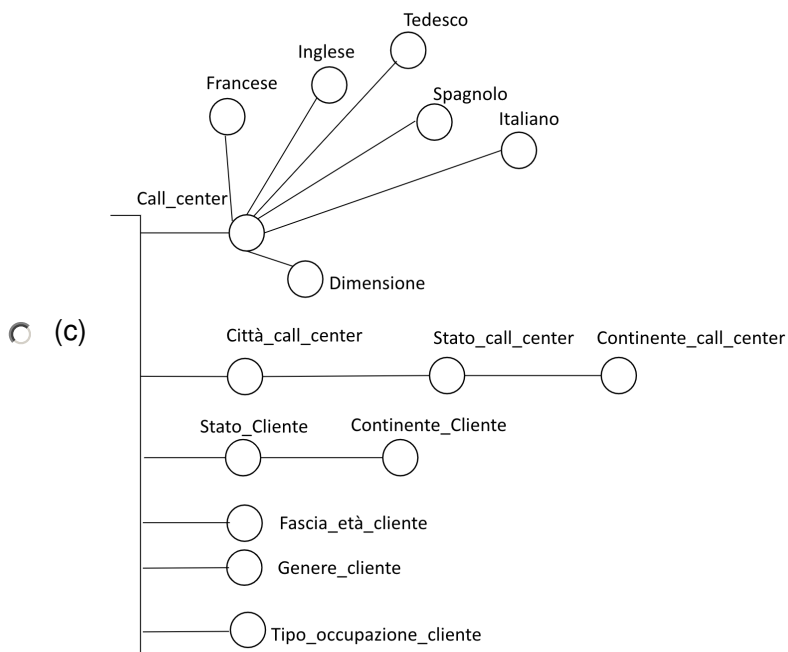
Ciascun call center fornisce un servizio multilingua e multicanale per contattare i potenziali clienti e proporre i prodotti. I diversi canali di comunicazione utilizzati da ciascun call center includono, oltre alle chiamate telefoniche, ad esempio anche le email e i messaggi sms. Ciascun call center è inoltre in grado di operare supportando più lingue straniere (ad esempio inglese, francese e tedesco). Pertanto ciascun call center può contattare clienti presenti in diversi stati nel mondo.

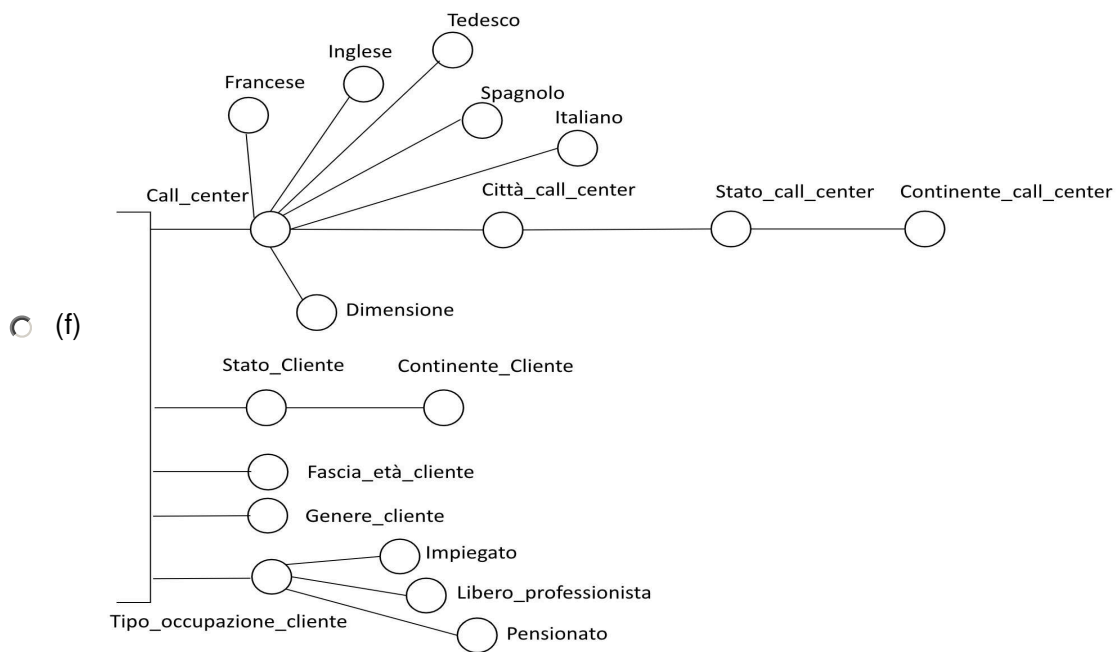
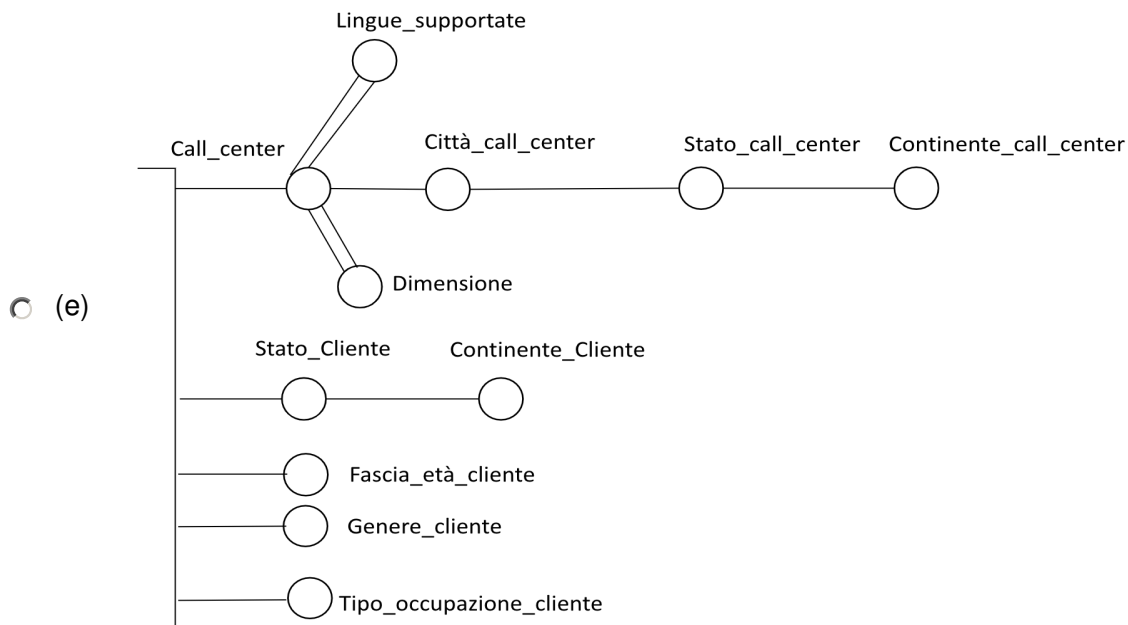
Si vuole analizzare (1) il numero medio di clienti contattati per prodotto e (2) il numero medio di risposte positive ricevute per prodotto, in base a:

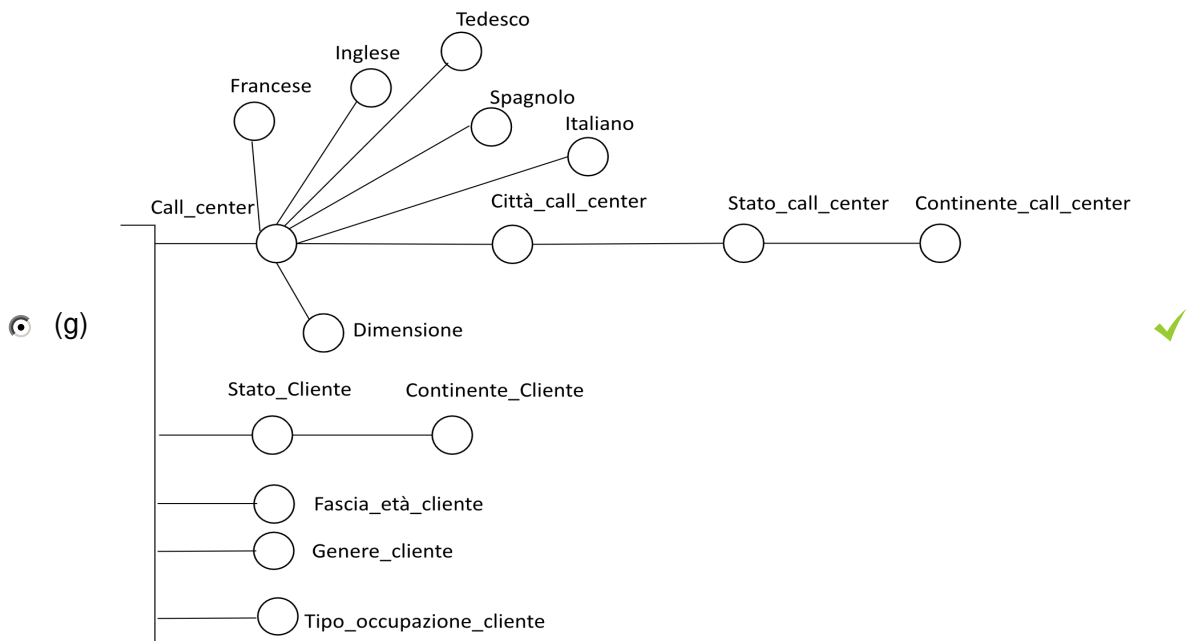
- call center, caratterizzato da un nome, dalla sua collocazione geografica (espressa in termini di città, stato e continente in cui si trova il call center), dalla dimensione del call center (espressa come piccola, media, o grande, in base al numero di dipendenti) e dall'indicazione delle lingue supportate dal call center (uno o più lingue tra inglese, francese, italiano, tedesco e spagnolo)
- informazione temporale su quando il call center ha contattato il cliente, espressa in termini di data, giorno della settimana, settimana, mese, bimestre, trimestre, quadrimestre, anno e fascia oraria (un valore tra mattino, pomeriggio, sera, notte)
- tipo di prodotto proposto al cliente contattato
- categorizzazione del tipo di prodotto in base, medio o lusso
- tipo di canale utilizzato per contattare il cliente (ossia chiamata telefonica oppure email o messaggio sms)
- stato e continente del cliente contattato
- caratteristiche del cliente contattato, in termini di genere, fascia di età del cliente (un valore tra meno di 20 anni, tra 20 e 40 anni, tra 40 e 60 anni, e più di 60 anni) e tipo di occupazione del cliente (ad esempio impiegato oppure libero professionista o pensionato)

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



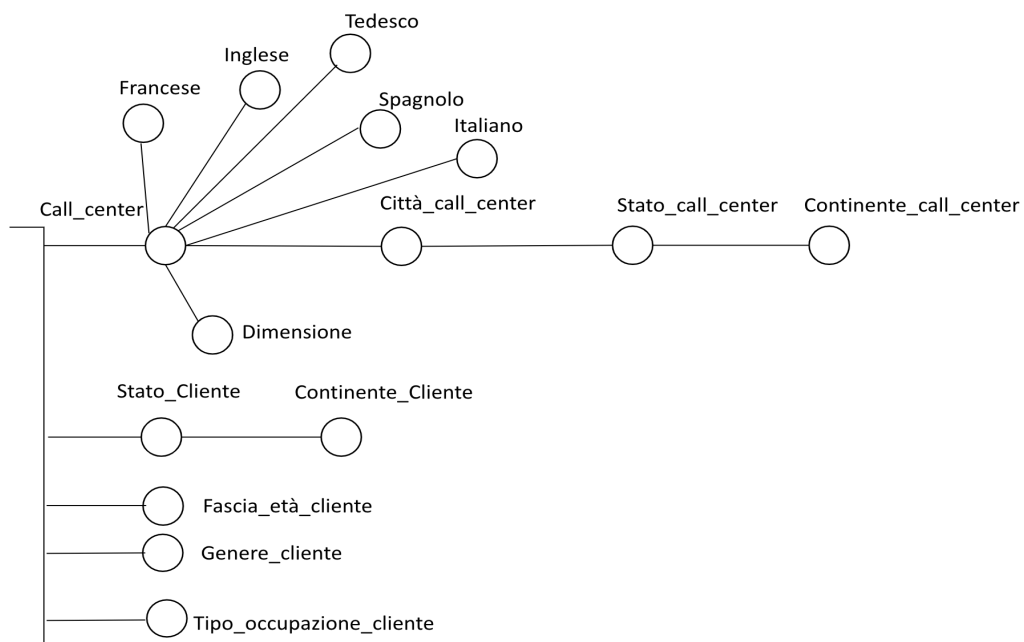






Risposta corretta.

La risposta corretta è:



Domanda 15

Risposta errata

Punteggio ottenuto -0,23 su 1,50

1.5 punti (-15% penalità per risposta sbagliata)

Nota: in MongoDB e' possibile effettuare raggruppamenti (\$group) su attributi multipli.

Per fare cio', e' sufficiente specificare come chiave (_id) un oggetto nella forma: { "chiave1": attributo1, "chiave2": attributo2, ... }.

In questo caso, verra' effettuato un raggruppamento sugli attributi attributo1, attributo2, Ciascun documento restituito dall'aggregazione avra' valore _id nel formato:

```
"_id": {  
  "chiave1": "attributo1_valore1",  
  "chiave2": "attributo2_valore1",  
  ...  
}
```

Viene data una collection MongoDB utilizzata per registrare le presenze per una piattaforma di apprendimento online. Ciascuna presenza viene registrata come un documento all'interno di una collection "presenze". Il seguente e' un esempio di documento estratto dalla collection.

```
{  
  "student": {  
    "first": "MARGARET",  
    "last": "MOORE"  
  },  
  "class_name": "Computer science",  
  "lesson_number": 9,  
  "duration": 120  
}
```

Si vuole estrarre, per ciascun corso, il **numero massimo di partecipanti a una singola lezione**. Quale delle seguenti query soddisfa la precedente richiesta?

Si consideri la coppia (nome del corso, numero della lezione) sufficiente a identificare univocamente qualunque lezione tenuta.

☐ (a)

```

db.collection.aggregate([
  {
    "$group": {
      "_id": {
        "class": "$class_name",
        "lesson": "$lesson_number"
      },
      "max_participants": {
        "$sum": {
          "$max": 1
        }
      }
    }
  }
])

```

☐ (b)

```

db.collection.aggregate([
  {
    "$group": {
      "_id": "$_id.class",
      "max_participants": {
        "$max": "$participants"
      }
    },
  },
  {
    "$group": {
      "_id": {
        "class": "$class_name",
        "lesson": "$lesson_number"
      },
      "participants": {
        "$sum": 1
      },
    }
  }
])

```

☐ (c) Nessuna delle altre risposte permette di indirizzare la query di interesse.

☒ (d)

```

db.collection.aggregate([
  {
    "$group": {
      "_id": {
        "class": "$class_name",
        "lesson": "$lesson_number"
      },
      "participants": {
        "$sum": 1
      },
    }
  },
  {
    "$group": {
      "_id": "$_id",
      "max_participants": {
        "$max": "$participants"
      }
    }
  },
])

```



○ (e)

```

db.collection.aggregate([
  {
    "$group": {
      "_id": {
        "class": "$class_name",
        "lesson": "$lesson_number"
      },
      "participants": {
        "$sum": 1
      },
    }
  },
  {
    "$group": {
      "_id": "$_id.class",
      "max_participants": {
        "$max": "$participants"
      }
    }
  },
])

```


○ (f)

```
db.collection.aggregate([
  {
    "$group": {
      "_id": {
        "class": "$class_name",
      },
      "max_participants": {
        "$max": {
          "$sum": 1
        }
      }
    }
  }
])
```

○ (g)

```
db.collection.aggregate([
  {
    "$group": {
      "_id": {
        "class": "$class_name",
        "lesson": "$lesson_number"
      },
      "max_participants": {
        "$max": {
          "$sum": 1
        }
      }
    }
  }
])
```

○ (h)

```

db.collection.aggregate([
  {
    "$group": {
      "_id": "$_id",
      "max_participants": {
        "$max": "$participants"
      }
    },
  },
  {
    "$group": {
      "_id": {
        "class": "$class_name",
        "lesson": "$lesson_number"
      },
      "participants": {
        "$sum": 1
      },
    },
  }
])

```

○ (i)

```

db.collection.aggregate([
  {
    "$group": {
      "_id": {
        "class": "$class_name",
      },
      "max_participants": {
        "$sum": {
          "$max": 1
        }
      }
    },
  }
])

```

Risposta errata.

La risposta corretta è:

```
db.collection.aggregate([
  {
    "$group": {
      "_id": {
        "class": "$class_name",
        "lesson": "$lesson_number"
      },
      "participants": {
        "$sum": 1
      },
    }
  },
  {
    "$group": {
      "_id": "$_id.class",
      "max_participants": {
        "$max": "$participants"
      }
    }
  },
])
```