



Politecnico
di Torino

Data Science e Tecnologie per le Basi di Dati

[Dashboard](#) ► [CORSI](#) ► [01SQMOV_0](#) ► [Introduzione](#) ► [Simulazione esame \(09 Febbraio 2022\)](#) ►
[Anteprima](#)

Iniziato	lunedì, 9 gennaio 2023, 17:38
Stato	Completato
Terminato	lunedì, 9 gennaio 2023, 17:38
Tempo impiegato	28 secondi
Valutazione	0,00 su un massimo di 31,00 (0%)

Domanda 1

Risposta non data

Punteggio max.:
1,50**Cardinalità (1.5 punti, penalità -15% per ogni risposta sbagliata)**

Sono date le seguenti tabelle:

APPARTAMENTO(CodA, NomeAppartamento, CodH, Indirizzo, Citta', Regione, Tipo)
 HOST(CodH, Nome, Cognome, Citta', Regione, email, numeroTelefono)
 UTENTE(CodU, Nome, Cognome, Citta', Regione, dataiscrizione)
 AFFITTO(CodA, DataInizio, CodU, DataFine, Costo, MetodoPagamento)

Sono date le seguenti cardinalità:

- $\text{card}(\text{APPARTAMENTO}) = 5 \cdot 10^4$ tuple,
 - Valori distinti di Tipo = 5
- $\text{card}(\text{HOST}) = 10^4$ tuple,
 - Valori distinti di Citta' = 100
 - Valori distinti di Regione = 20
- $\text{card}(\text{UTENTE}) = 10^7$ tuple,
 - Valori distinti di Citta' = 10^4
 - Valori distinti di Regione = 20
- $\text{card}(\text{AFFITTO}) = 5 \cdot 10^7$ tuple,
 - $\text{MIN}(\text{DataInizio}) = 1/1/2002$, $\text{MAX}(\text{DataInizio}) = 31/12/2021$
 - Fattore di riduzione di costo $>50 = 1/5$

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

having COUNT(*)>2 = 1/5

having COUNT(DISTINCT MetodoPagamento)>1 = 1/2

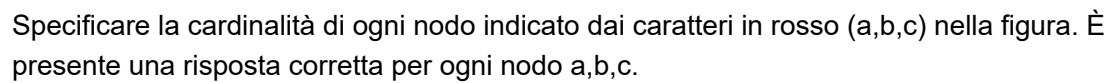
Si consideri la seguente query:

```

select AFF1.CodA, COUNT(DISTINCT MetodoPagamento)
from Appartamento A, Utente U, Affitto AFF1
where AFF1.CodU=U.CodU and A.CodA=AFF1.CodA and
      AFF1.DataInizio≥1/1/2021 and A.Tipo='Loft' and
      U.Regione<>'Piemonte'
      and AFF1.CodA IN ( SELECT AFF2.CodA
                        FROM Affitto AFF2, Appartamento A2, Host H
                        WHERE A2.CodH=H.CodH and AFF2.CodA=A2.CodA
                        and H.Regione='Liguria' and AFF2.Costo>50
                        GROUP BY AFF2.CodA
                        having COUNT(*)>2)
group by AFF1.CodA
having COUNT( DISTINCT MetodoPagamento)>1

```

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



☐ (a) $c: < 5 \cdot 10^2$

☐ (b) $b: 10^6$

☐ (c) $b: 10^4$

☐ (d) $b: 25 \cdot 10^5$

☐ (e) $c: \approx 10^4$

☐ (f) $c: \approx 5 \cdot 10^4$

☐ (g) $b: 5 \cdot 10^3$

☐ (h) $a: 10^6$

☐ (i) $a: 10^5$

☐ (j) $a: 10^4$

☐ (k) $a: 5 \cdot 10^5$

☐ (l) $c: < 50$

La risposta corretta è: a: $5 \cdot 10^5$, b: $5 \cdot 10^3$, c: $< 5 \cdot 10^2$

Domanda 2

Risposta non data

Punteggio max.:

1,00

Teoria - 1 punto (penalità 15% per risposta sbagliata)

Viene data una collection "invoices", contenente le fatture di una compagnia telefonica verso i suoi clienti. Per ciascuna fattura, sono presenti le informazioni del cliente e dei consumi, oltre all'importo totale dovuto. Il seguente è un esempio di documento estratto dalla collection:

```
{
  "_id" : ObjectId("61fa5b8f6f631bb5339dc4b7"),
  "customer" : {
    "first" : "DOROTHY",
    "last" : "DAVIS",
    "age" : 29,
    "gender" : "F"
  },
  "month" : 6,
  "year" : 2021,
  "total_due_EUR" : 6.6,
  "consumption" : {
    "minutes" : 552,
    "internet_GB" : 2.96,
    "texts" : 15
  }
}
```

Per un dato anno, si può calcolare il *fatturato annuo* come la somma di tutte le voci *total_due_EUR* all'interno delle fatture per quell'anno. Si vuole estrarre, da questa collection, il fatturato annuo più basso. Quale delle seguenti query soddisfa la richiesta?

☐ (a)

```
db.invoices.aggregate([
  {
    $group: {
      _id: "$year",
      "min_revenue": {
        "$min": "$total_due_EUR"
      }
    }
  },
  {
    $group: {
      _id: null,
      "revenue": {
        "$sum": "$min_revenue"
      }
    }
  }
])
```

☐ (b)

```
db.invoices.aggregate([
  {
    $group: {
      _id: "$year",
      "min_revenue": {
        "$sum": {
          "$min": "$total_due_EUR"
        }
      }
    }
  }
])
```

☐ (c)

```
db.invoices.aggregate([
  {
    $group: {
      _id: null,
      "revenue": {
        "$min": {
          "$sum": "$total_due_EUR"
        }
      }
    }
  }
])
```

☐ (d)

```
db.invoices.aggregate([
  {
    $group: {
      _id: null,
      "revenue": {
        "$sum": "$total_due_EUR"
      }
    }
  },
  {
    $group: {
      _id: "$year",
      "min_revenue": {
        "$min": "$revenue"
      }
    }
  }
])
```

☐ (e)

```
db.invoices.aggregate([
  {
    $group: {
      _id: "$year",
      "revenue": {
        "$sum": "$total_due_EUR"
      }
    }
  },
  {
    $group: {
      _id: null,
      "min_revenue": {
        "$min": "$revenue"
      }
    }
  }
])
```

☐ (f)

```
db.invoices.aggregate([
  {
    $group: {
      _id: "$year",
      "min_revenue": {
        "$min": {
          "$sum": "$total_due_EUR"
        }
      }
    }
  }
])
```

☐ (g)

```
db.invoices.aggregate([
  {
    $group: {
      _id: null,
      "revenue": {
        "$min": "$total_due_EUR"
      }
    }
  },
  {
    $group: {
      _id: "$year",
      "min_revenue": {
        "$sum": "$revenue"
      }
    }
  }
])
```

☐ (h)

```
db.invoices.aggregate([
  {
    $group: {
      _id: null,
      "min_revenue": {
        "$sum": {
          "$min": "$total_due_EUR"
        }
      }
    }
  }
])
```

Risposta errata.

La risposta corretta è:

```
db.invoices.aggregate([
  {
    $group: {
      _id: "$year",
      "revenue": {
        "$sum": "$total_due_EUR"
      }
    }
  },
  {
    $group: {
      _id: null,
      "min_revenue": {
        "$min": "$revenue"
      }
    }
  }
])
```

Domanda 3

Risposta non data

Punteggio max.:
1,50**Indici (1.5 punti, penalità -15% per ogni risposta sbagliata)**

Sono date le seguenti tabelle:

APPARTAMENTO(CodA, NomeAppartamento, CodH, Indirizzo, Citta', Regione, Tipo)HOST(CodH, Nome, Cognome, Citta', Regione, email, numeroTelefono)UTENTE(CodU, Nome, Cognome, Citta', Regione, dataiscrizione)AFFITTO(CodA, DataInizio, CodU, DataFine, Costo, MetodoPagamento)

Sono date le seguenti cardinalità:

- $\text{card}(\text{APPARTAMENTO}) = 5 \cdot 10^4$ tuple,
 - Valori distinti di Tipo = 5
- $\text{card}(\text{HOST}) = 10^4$ tuple,
 - Valori distinti di Citta' = 100
 - Valori distinti di Regione = 20
- $\text{card}(\text{UTENTE}) = 10^7$ tuple,
 - Valori distinti di Citta' = 10^4
 - Valori distinti di Regione = 20
- $\text{card}(\text{AFFITTO}) = 5 \cdot 10^7$ tuple,
 - $\text{MIN}(\text{DataInizio}) = 1/1/2002$, $\text{MAX}(\text{DataInizio}) = 31/12/2021$
 - Fattore di riduzione di costo $> 50 = 1/5$

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

having COUNT(*) > 2 = 1/5

having COUNT(DISTINCT MetodoPagamento) > 1 = 1/2

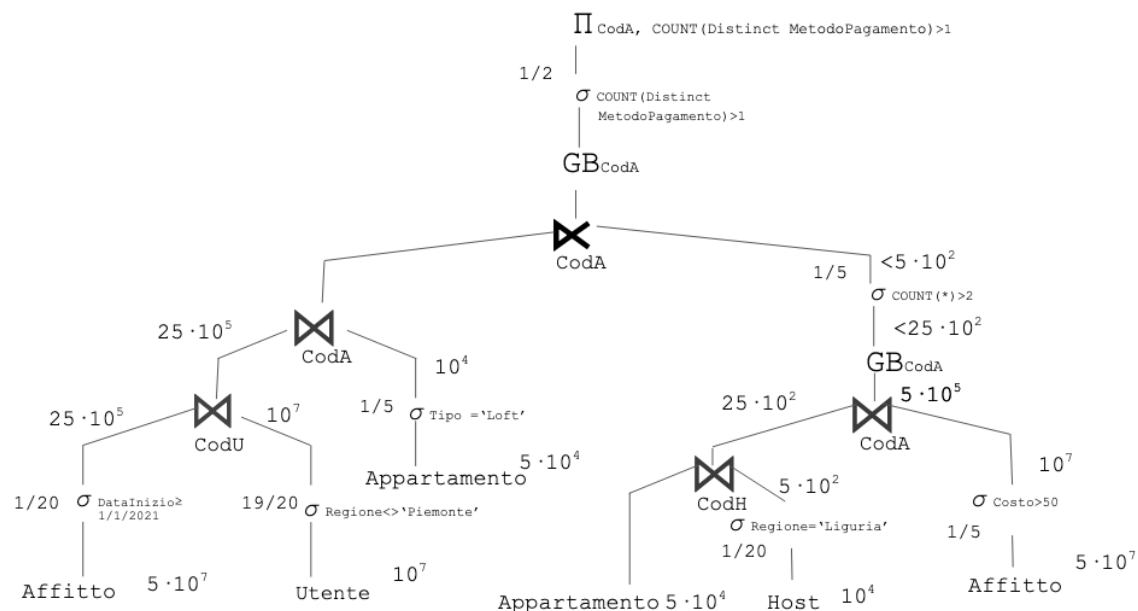
Si consideri la seguente query:

```

select AFF1.CodA, COUNT(DISTINCT MetodoPagamento)
from Appartamento A, Utente U, Affitto AFF1
where AFF1.CodU=U.CodU and A.CodA=AFF1.CodA and
      AFF1.DataInizio ≥ 1/1/2021 and A.Tipo='Loft' and
      U.Regione <> 'Piemonte'
      and AFF1.CodA IN ( SELECT AFF2.CodA
                        FROM Affitto AFF2, Appartamento A2, Host H
                        WHERE A2.CodH=H.CodH and AFF2.CodA=A2.CodA
                        and H.Regione='Liguria' and AFF2.Costo > 50
                        GROUP BY AFF2.CodA
                        having COUNT(*) > 2)
group by AFF1.CodA
having COUNT( DISTINCT MetodoPagamento) > 1

```

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



Si selezionino le strutture fisiche accessorie per migliorare le prestazioni dell'interrogazione (se possibile). È possibile selezionare più risposte corrette.

Scegli una o più alternative:

- ☐ (a) CREATE INDEX IndexE ON APPARTAMENTO(Tipo) - HASH
- ☐ (b) CREATE INDEX IndexD ON UTENTE(Regione) - B+-Tree
- ☐ (c) CREATE INDEX IndexH ON HOST(Regione) - B+-Tree
- ☐ (d) CREATE INDEX IndexJ ON AFFITTO(Costo) - B+-Tree
- ☐ (e) CREATE INDEX IndexA ON AFFITTO(DataInizio) - HASH
- ☐ (f) CREATE INDEX IndexC ON UTENTE(Regione) - HASH
- ☐ (g) CREATE INDEX IndexB ON AFFITTO(DataInizio) - B+-Tree
- ☐ (h) CREATE INDEX IndexF ON APPARTAMENTO(Tipo) - B+-Tree
- ☐ (i) CREATE INDEX IndexI ON AFFITTO(Costo) - HASH
- ☐ (j) Nessuna - le strutture fisiche accessorie non migliorerebbero le prestazioni dell'interrogazione
- ☐ (k) CREATE INDEX IndexG ON HOST(Regione) - HASH

Risposta errata.

La risposta corretta è: CREATE INDEX IndexB ON AFFITTO(DataInizio) - B+-Tree,
CREATE INDEX IndexG ON HOST(Regione) - HASH

Domanda 4

Risposta non data

Punteggio max.:
1,00**Progettazione DW - 1 punto (penalità 15% per risposta sbagliata)**

Si vogliono analizzare le informazioni relative agli accessi che sono stati effettuati dai pazienti presso il reparto di Pronto Soccorso degli ospedali italiani. Il reparto di Pronto Soccorso garantisce gli interventi diagnostico-terapeutici di emergenza/urgenza in risposta al bisogno manifestato dall'utenza.

I pazienti possono raggiungere il reparto di Pronto Soccorso in ambulanza o autonomamente con un mezzo proprio. Ai pazienti che accedono al pronto soccorso viene assegnato un codice colore in base ai sintomi denunciati dal paziente e ai segnali vitali riscontrati. Il codice colore può assumere un solo valore tra rosso, giallo, verde e bianco in base al livello di gravità della condizione del paziente e quindi il grado di urgenza nel procedere con la visita del paziente.

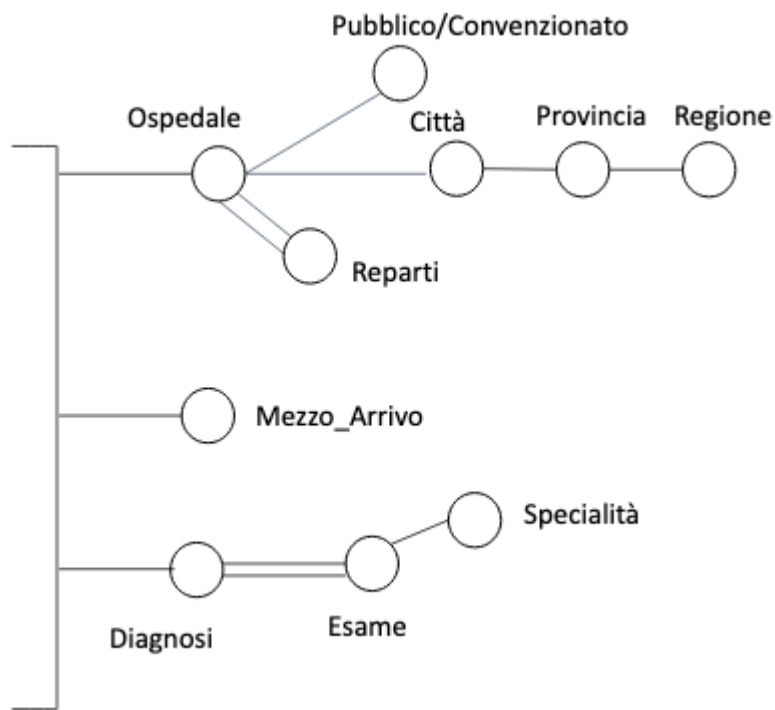
Una volta ricevuto il codice colore, il paziente viene visitato dai medici in servizio presso il reparto. Qualora ritenuto necessario, il paziente è anche sottoposto ad uno o più esami medici che permettano di approfondire e chiarire la sua condizione. Al termine, viene rilasciata una diagnosi. Nei casi più critici viene effettuato il ricovero del paziente presso l'ospedale; altrimenti il paziente viene dimesso.

Si vuole analizzare (1) il numero medio di ore di permanenza in pronto soccorso per paziente e (2) il numero medio di visite per medico, in base a:

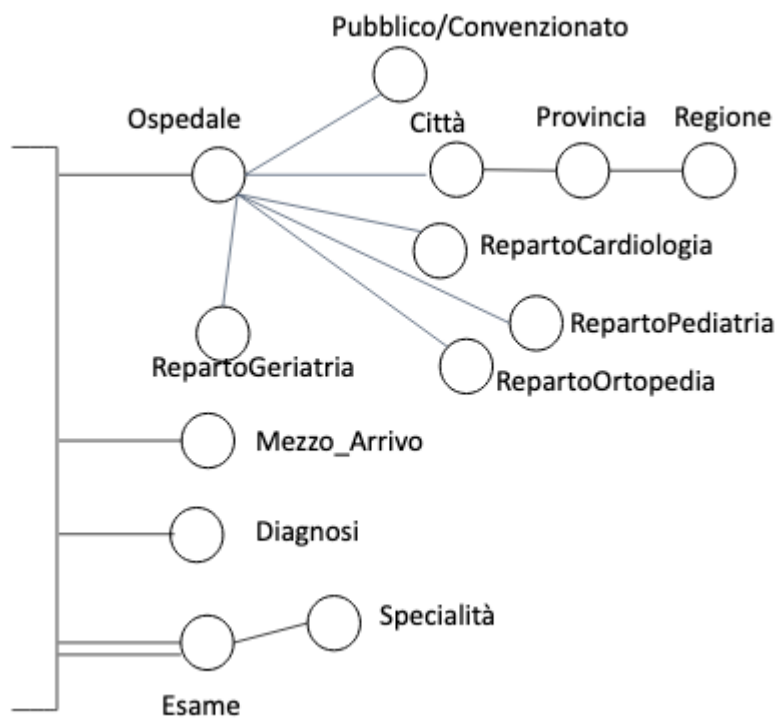
- modalità di arrivo al Pronto Soccorso (ambulanza oppure mezzo proprio)
- codice colore (bianco, verde, giallo e rosso)
- ospedale, caratterizzato dalla sua collocazione geografica (espressa in termini di città, provincia, regione), e dall'indicazione se siano presenti nell'ospedale uno o più tra i seguenti reparti: cardiologia, pediatria, ortopedia e geriatria. Per ogni ospedale è anche noto se sia una struttura pubblica o accreditata
- data, giorno del mese, giorno festivo, mese, mese dell'anno, quadrimestre, semestre, anno
- sesso e fascia di età del paziente (<13, 13-18, 19-40, 41-70, >70)
- diagnosi rilasciata, caratterizzata dall'elenco degli esami medici eseguiti. Ogni esame è caratterizzato dalla specialità medica a cui si riferisce. Inoltre, ogni esame può essere richiesto in più diagnosi diverse.

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

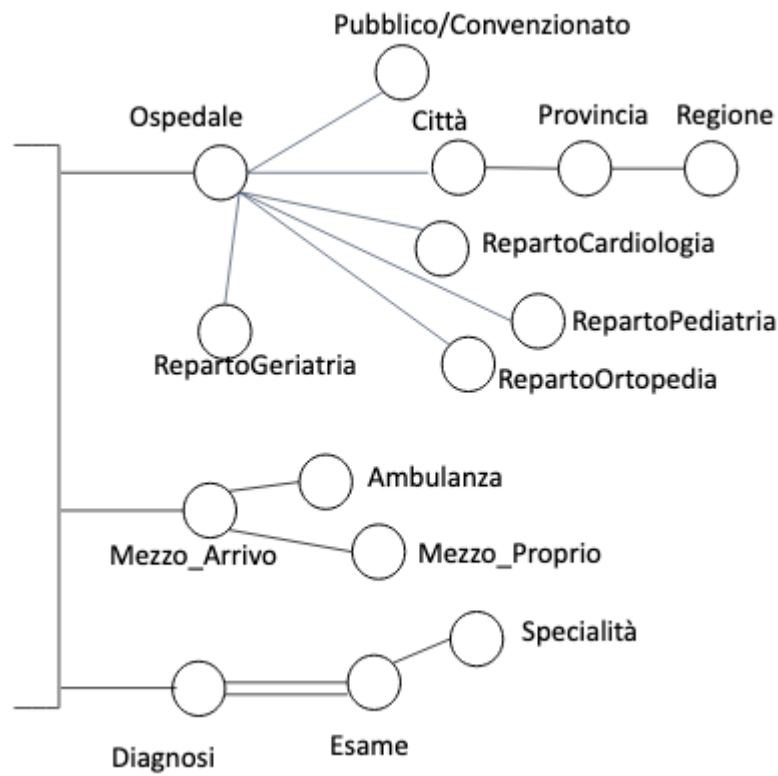
☐ (a)



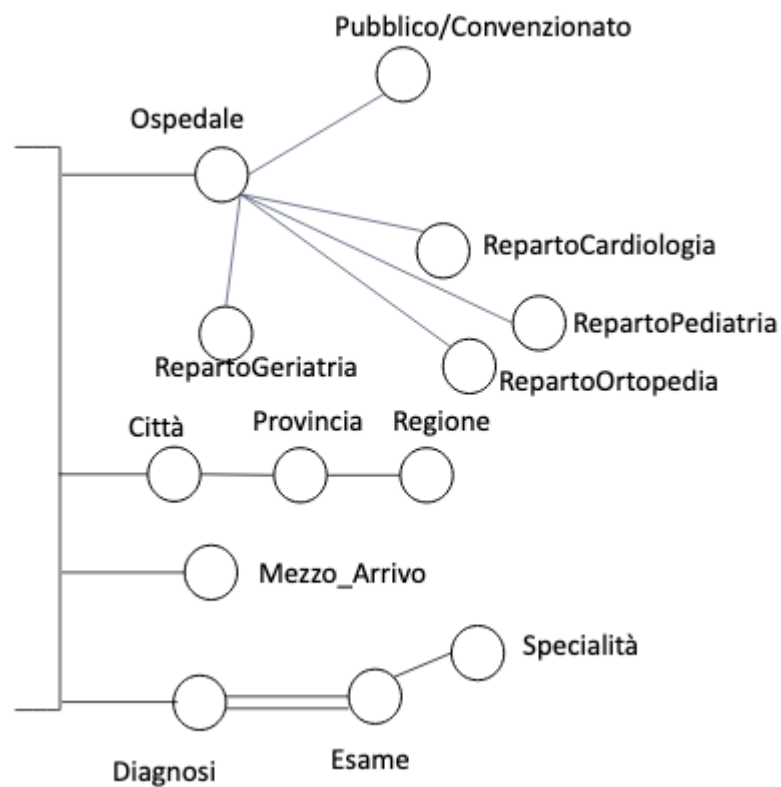
☐ (b)



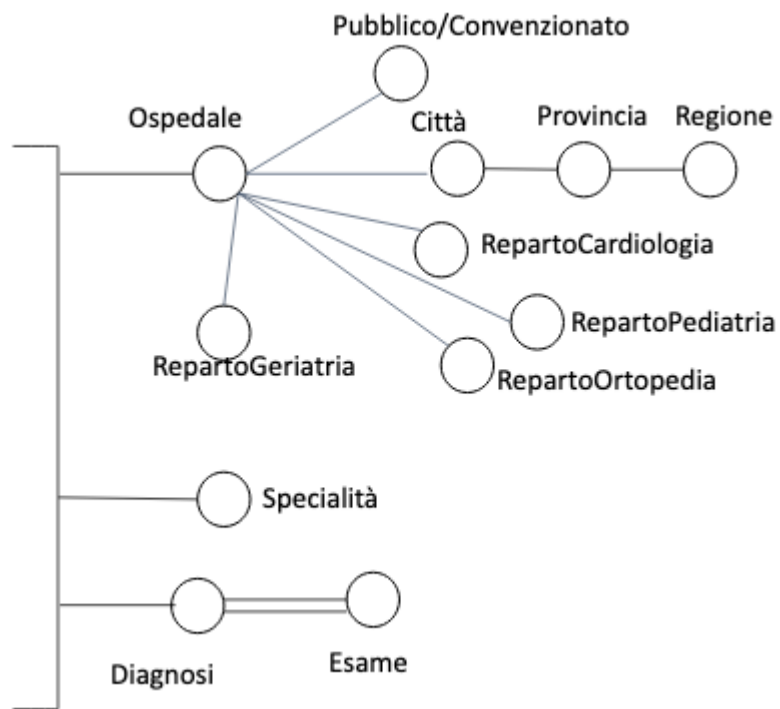
☐ (c)



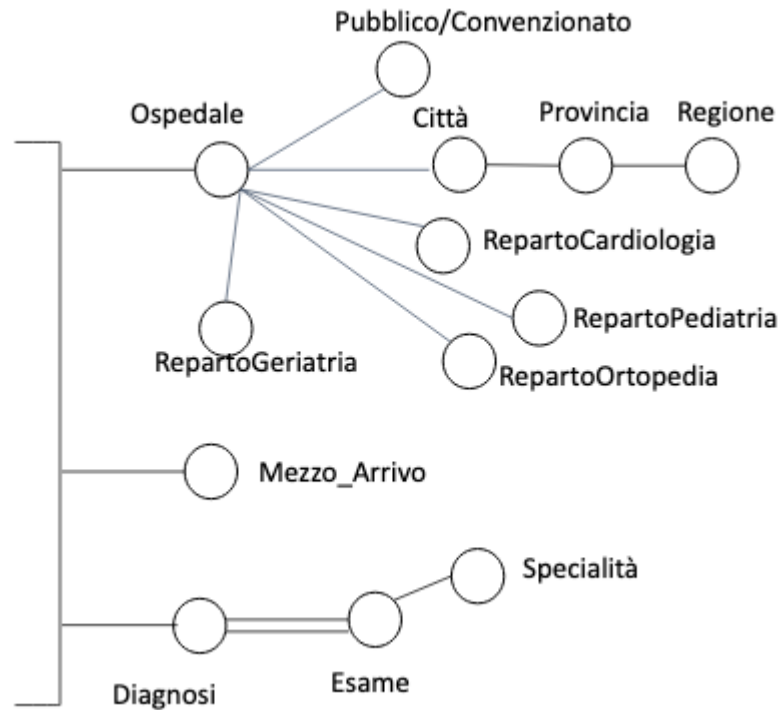
☐ (d)



☐ (e)

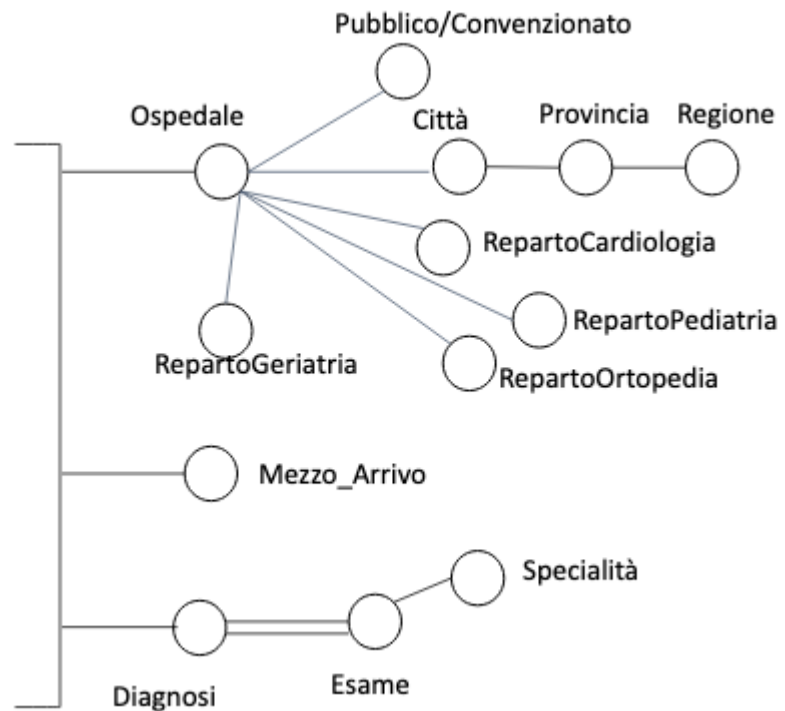


☐ (f)



Risposta errata.

La risposta corretta è:



Domanda 5

Risposta non data

Punteggio max.:

1,50

Teoria - 1.5 punti (penalità 15% per risposta sbagliata)

Quale delle seguenti affermazioni relative al clustering **non** è corretta?

- ☐ (a) il clustering gerarchico richiede di poter calcolare distanze tra punti, tra punti e cluster e tra cluster
- ☐ (b) aumentando il valore di k nell'algoritmo k-means, normalmente si riduce il valore di SSE (sum of squared errors)
- ☐ (c) l'algoritmo DBSCAN non permette di specificare il numero di cluster desiderati
- ☐ (d) il risultato di un clustering gerarchico può essere visualizzato tramite un dendrogramma
- ☐ (e) nell'algoritmo DBSCAN, i clusters sono regioni ad alta densità di punti, separati da regioni a bassa densità
- ☐ (f) utilizzando l'algoritmo k-means, il centroide di un cluster corrisponde sempre ad un punto appartenente a quel cluster

Risposta errata.

La risposta corretta è: utilizzando l'algoritmo k-means, il centroide di un cluster corrisponde sempre ad un punto appartenente a quel cluster

Domanda 6

Risposta non data

Punteggio max.:

1,50

Teoria - 1.5 punti (penalità 15% per risposta sbagliata)

Viene dato il seguente schedule di quattro transazioni:

$S = W4(Z) R2(Z) R1(X) W3(Y) R4(X) W2(X) R1(Y) W1(Y) R4(Y)$

Viene usata la seguente sintassi:

- $Ra(b)$ lettura da parte della transazione a dell'oggetto b
- $Wa(b)$ scrittura da parte della transazione a dell'oggetto b

Quale delle seguenti affermazioni è vera?

-
- ☐ (a) S è view serializable perchè è view equivalent allo schedule $R1(X) W3(Y) R1(Y) R4(X) W2(X) W1(Y) R4(Y) W4(Z) R2(Z)$
 - ☐ (b) S è view serializable perchè è view equivalent allo schedule $W3(Y) R1(Y) R1(X) W1(Y) W2(X) R2(Z) W4(Z) R4(X) R4(Y)$
 - ☐ (c) S è view serializable perchè è view equivalent allo schedule $W3(Y) R1(Y) R1(X) W1(Y) W4(Z) R4(X) R4(Y) W2(X) R2(Z)$
 - ☐ (d) S è view serializable perchè è view equivalent allo schedule $W1(Y) R1(Y) R1(X) W3(Y) W4(Z) R4(X) R4(Y) W2(X) R2(Z)$
 - ☐ (e) S non è view serializable perchè non è view equivalent a nessuno schedule seriale
 - ☐ (f) S è view serializable perchè è view equivalent allo schedule $W3(Y) W4(Z) R4(X) R4(Y) R1(Y) R1(X) W1(Y) W2(X) R2(Z)$

Risposta errata.

La risposta corretta è: S è view serializable perchè è view equivalent allo schedule $W3(Y) R1(Y) R1(X) W1(Y) W4(Z) R4(X) R4(Y) W2(X) R2(Z)$

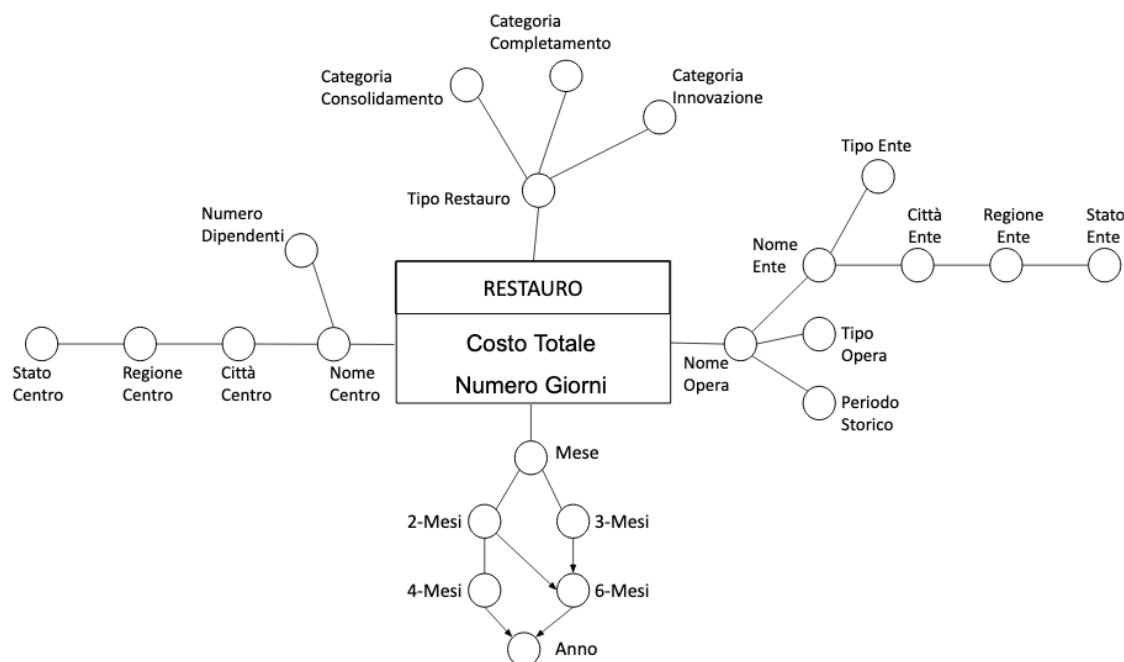
Domanda 7

Risposta non data

Punteggio max.:
4,00**SQL esteso - 4 punti (nessuna penalità per risposta sbagliata)**

Il seguente data warehouse descrive le attività di restauro svolte dai restauratori affiliati a centri di restauro, che effettuano restauri di opere (i tipi di opera possono essere dipinti, affreschi, sculture, ecc.) disponibili presso diversi enti (musei, fondazioni, chiese, ecc.). I tipi di restauro (conservativo, ecc.) possono ricadere in una o più delle seguenti tre categorie: consolidamento, completamento, innovazione (questi attributi sono booleani e rappresentano una configurazione delle categorie di restauro).

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



CENTRO_RESTAURO (IDCentro, NomeCentro, CittàCentro, RegioneCentro, StatoCentro, NumeroDipendenti)

OPERA (IDOpera, NomeOpera, TipoOpera, PeriodoStorico, NomeEnte, TipoEnte, CittàEnte, RegioneEnte, StatoEnte)

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

TIPO_RESTAURO (IDTipoRestauro, TipoRestauro, CategoriaConsolidamento, CategoriaCompletamento, CategoriaInnovazione)

RESTAURO (IDOpera, IDCentro, IDTempo, IDTipoRestauro, CostoTotale, NumeroGiorni)

Considerando solo i centri di restauro con più di 20 dipendenti, separatamente per ogni tipo di opera, anno e nome del centro, visualizzare

- Il numero totale di giorni di restauro
- Il numero medio di giorni di restauro per trimestre (attributo 3-Mesi)
- La percentuale del costo totale rispetto al costo totale complessivo della regione in cui si trova il centro
- la posizione in una graduatoria (rank) in ordine crescente rispetto al numero totale di giorni di restauro, separatamente per ogni anno e tipo di opera

```
SELECT NomeCentro, TipoOpera, Anno,
SUM(NumeroGiorni),
SUM(numeroGiorni)/COUNT(DISTINCT 3-Mesi),
100 * SUM(CostoTotale)/SUM(SUM(CostoTotale)) OVER (Partition by TipoOpera,
RegioneCentro, Anno),
RANK() OVER (PARTITION BY Anno, TipoOpera ORDER BY SUM(NumeroGiorni))
FROM RESTAURO R, TEMPO T, CENTRO_RESTAURO C, OPERA O
WHERE R.IDTempo = T.IDTempo AND R.IDCentro = C.IDCentro AND
O.IDOpera=R.IDOpera
AND NumeroDipendenti > 20
GROUP BY NomeCentro, TipoOpera, Anno, RegioneCentro
```

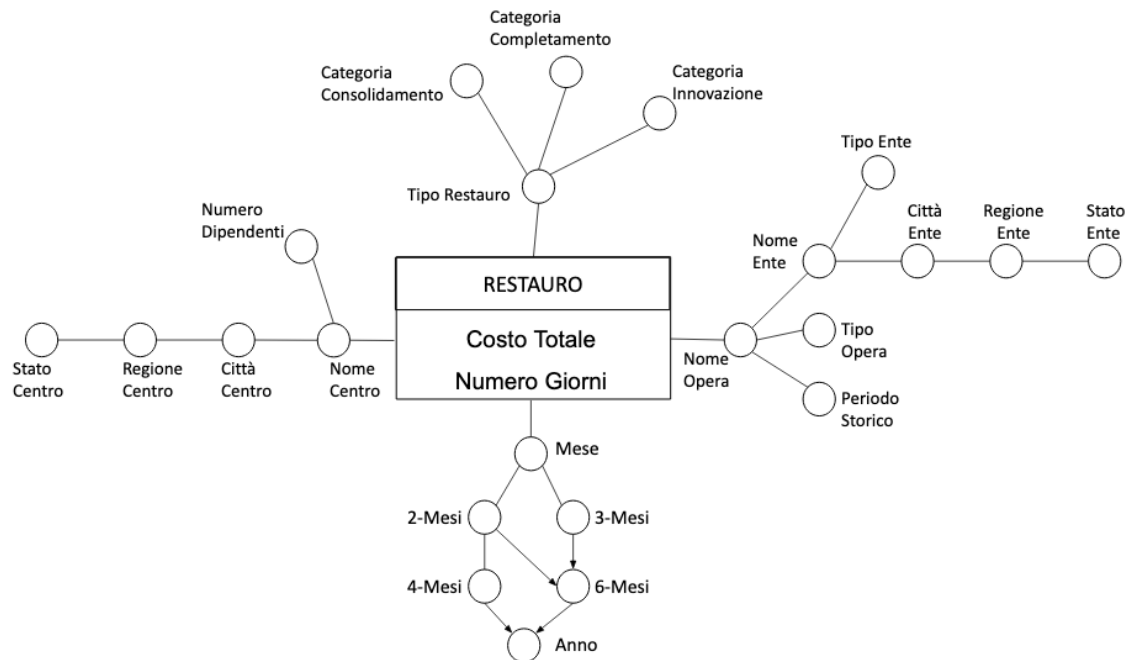
Domanda 8

Risposta non data

Punteggio max.:
4,00**Vista materializzata - 4 punti (nessuna penalità per risposta sbagliata)**

Il seguente data warehouse descrive le attività di restauro svolte dai restauratori affiliati a centri di restauro, che effettuano restauri di opere (i tipi di opera possono essere dipinti, affreschi, sculture, ecc.) disponibili presso diversi enti (musei, fondazioni, chiese, ecc.). I tipi di restauro (conservativo, ecc.) possono ricadere in una o più delle seguenti tre categorie: consolidamento, completamento, innovazione (questi attributi sono booleani e rappresentano una configurazione delle categorie di restauro).

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



CENTRO_RESTAURO (IDCentro, NomeCentro, CittàCentro, RegioneCentro, StatoCentro, NumeroDipendenti)

OPERA (IDOpera, NomeOpera, TipoOpera, PeriodoStorico, NomeEnte, TipoEnte, CittàEnte, RegioneEnte, StatoEnte)

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

TIPO_RESTAURO (IDTipoRestauro, TipoRestauro, CategoriaConsolidamento, CategoriaCompletamento, CategoriaInnovazione)

RESTAURO (IDOpera, IDCentro, IDTempo, IDTipoRestauro, CostoTotale, NumeroGiorni)

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

- Considerando solo i centri restauri siti in Germania, visualizzare il numero complessivo di giorni di restauro e il relativo costo separatamente per tipo restauro e anno.
- Visualizzare il costo medio dei restauri per giorno di lavoro separatamente per semestre e regione del centro restauro.
- Considerando i tipi di restauro appartenenti alla categoria innovazione, visualizzare il costo totale dei restauri separatamente per stato in cui è sito un centro restauro.

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.

Si ricorda che le istruzioni SQL per le query (a), (b) e (c) non sono richieste.

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

Blocco A

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

```
Select TipoRestauro, 6-Mesi, RegioneCentro, Anno, StatoCentro, CategoricalInnovazione
SUM(NumeroGiorni) As NumGG, SUM(CostoTotale) As CostoTOT
FROM CENTRO_RESTAURO, RESTAURO, TEMPO, TIPO_RESTAURO
WHERE join
Group by TipoRestauro, 6-Mesi, RegioneCentro, Anno, StatoCentro, CategoricalInnovazione
```

```
CREATE OR REPLACE TRIGGER TriggerForViewCentroRestauro
AFTER INSERT ON RESTAURO
FOR EACH ROW
DECLARE
varAnno DATE, varSemestre DATE;
varTipoRestauro varchar(10); varStato varchar(10); varRegione varchar(10); N INTEGER;
varCatInnovazione Boolean;
BEGIN
    SELECT 6-Mesi, Anno INTO VarSemestre, varAnno
    FROM TEMPO
    WHERE IDTempo = :NEW.IDTempo;

    SELECT RegioneCentro, StatoCentro INTO varRegione, varStato
    FROM CENTRO_RESTAURO
    WHERE IDCentro = :NEW.Centro;

    SELECT TipoRestauro, CategoricalInnovazione INTO varTipoRestauro, varCatInnovazione
    FROM TIPO_RESTAURO
    WHERE IDTipoRestauro = :NEW.IDTipoRestauro;

    SELECT COUNT(*) INTO N
    FROM ViewCentroRestauro
    WHERE TipoRestauro= varTipoRestauro AND 6-Mesi=VarSemestre AND
    RegioneCentro=varRegione;

    IF N>0 THEN
        UPDATE ViewCentroRestauro
        SET NumGG = NumGG+ :NEW.NumeroGiorni, CostoTOT = CostoTOT +
        :NEW.CostoTotale
        WHERE TipoRestauro= varTipoRestauro AND 6-Mesi=VarSemestre AND
        RegioneCentro=varRegione;
    ELSE
        INSERT INTO ViewCentroRestauro(...) VALUES ( varTipoRestauro, VarSemestre,
        varRegione, varAnno, varStato, varCatInnovazione, :NEW.NumeroGiorni,
        :NEW.CostoTotale);
    END
```

Domanda 9

Risposta non data

Punteggio max.:

1,00

Teoria - 1 punto (penalità 15% per risposta sbagliata)

La correlazione di Pearson fra due variabili $x = \{x_1, x_2, \dots, x_n\}$ con media \bar{x} e $y = \{y_1, y_2, \dots, y_n\}$ con media \bar{y} viene definita come:

$$r_{xy} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}}$$

Un dataset è costituito da cinque features A, B, C, D, E. Si considerino i seguenti valori di correlazione tra le features, espressi attraverso il coefficiente di Pearson:

- $\text{corr}(A, B) = -1$
- $\text{corr}(B, C) = 0.4$
- $\text{corr}(C, D) = 1$
- $\text{corr}(D, E) = -0.3$

Quale delle seguenti affermazioni **non** è corretta:

- ☐ (a) quando la variabile C cresce, anche la variabile D cresce
- ☐ (b) $\text{corr}(C, E) = -0.3$
- ☐ (c) $\text{corr}(A, C) = 0.4$
- ☐ (d) le variabili A e B potrebbero contenere i seguenti valori: $A = [5, 6, 7, 8]$, $B = [-5, -6, -7, -8]$
- ☐ (e) le variabili C e D potrebbero contenere i seguenti valori: $C = [1, 2, 3, 4]$, $D = [1, 3, 5, 7]$
- ☐ (f) le variabili C e D potrebbero contenere valori identici

Risposta errata.

La risposta corretta è: $\text{corr}(A, C) = 0.4$

Domanda 10

Risposta non data

Punteggio max.:
2,00**Anticipo group by (2 punti, penalità -15% per ogni risposta sbagliata)**

Sono date le seguenti tabelle:

APPARTAMENTO(CodA, NomeAppartamento, CodH, Indirizzo, Citta', Regione, Tipo)
 HOST(CodH, Nome, Cognome, Citta', Regione, email, numeroTelefono)
 UTENTE(CodU, Nome, Cognome, Citta', Regione, dataiscrizione)
 AFFITTO(CodA, DataInizio, CodU, DataFine, Costo, MetodoPagamento)

Sono date le seguenti cardinalità:

- $\text{card}(\text{APPARTAMENTO}) = 5 \cdot 10^4$ tuple,
 - Valori distinti di Tipo = 5
- $\text{card}(\text{HOST}) = 10^4$ tuple,
 - Valori distinti di Citta' = 100
 - Valori distinti di Regione = 20
- $\text{card}(\text{UTENTE}) = 10^7$ tuple,
 - Valori distinti di Citta' = 10^4
 - Valori distinti di Regione = 20
- $\text{card}(\text{AFFITTO}) = 5 \cdot 10^7$ tuple,
 - $\text{MIN}(\text{DataInizio}) = 1/1/2002$, $\text{MAX}(\text{DataInizio}) = 31/12/2021$
 - Fattore di riduzione di costo $>50 = 1/5$

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

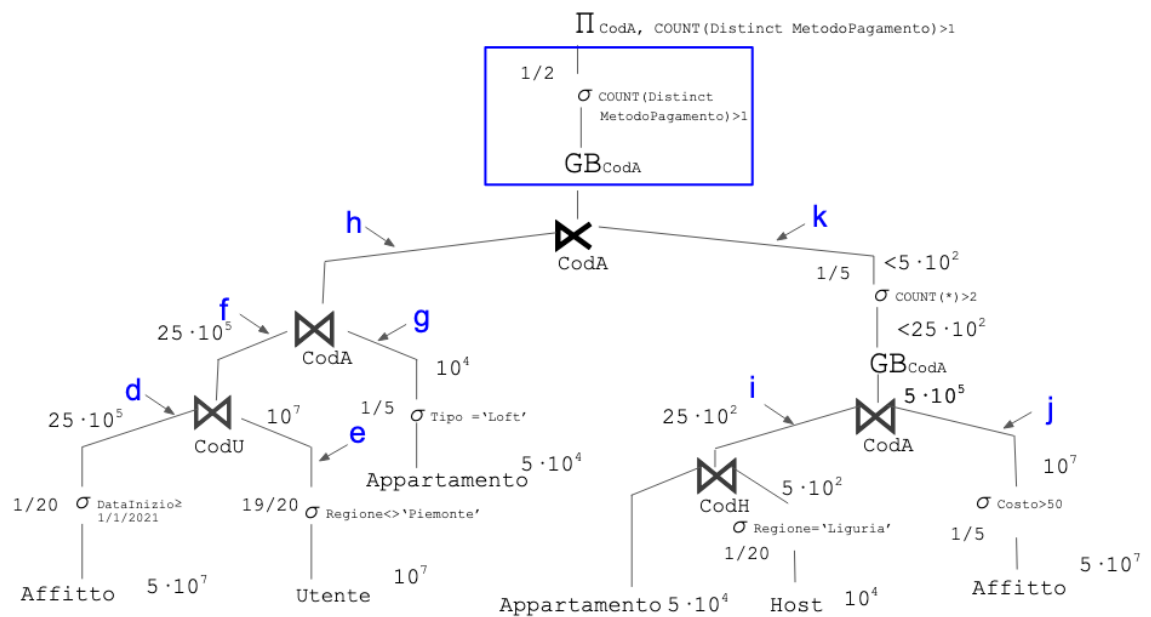
having COUNT(*)>2 = 1/5

having COUNT(DISTINCT MetodoPagamento)>1 = 1/2

Si consideri la seguente query:

```
select AFF1.CodA, COUNT(DISTINCT MetodoPagamento)
from Appartamento A, Utente U, Affitto AFF1
where AFF1.CodU=U.CodU and A.CodA=AFF1.CodA and
      AFF1.DataInizio≥1/1/2021 and A.Tipo='Loft' and
      U.Regione<>'Piemonte'
      and AFF1.CodA IN ( SELECT AFF2.CodA
                        FROM Affitto AFF2, Appartamento A2, Host H
                        WHERE A2.CodH=H.CodH and AFF2.CodA=A2.CodA
                        and H.Regione='Liguria' and AFF2.Costo>50
                        GROUP BY AFF2.CodA
                        having COUNT(*)>2)
group by AFF1.CodA
having COUNT( DISTINCT MetodoPagamento)>1
```

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



Analizzare l'anticipazione della GROUP BY **group by AFF1.CodA having COUNT(DISTINCT MetodoPagamento) > 1** rappresentata nel riquadro. Selezionare la soluzione che consente la massima efficienza nell'esecuzione della query (se esiste).

- ☐ (a) E' possibile anticiparla nel ramo d
- ☐ (b) E' possibile anticiparla nel ramo h
- ☐ (c) E' possibile anticiparla nel ramo k
- ☐ (d) Non e' possibile anticipare la Group BY **group by AFF1.CodA having COUNT(DISTINCT MetodoPagamento) > 1**
- ☐ (e) E' possibile anticiparla nel ramo f
- ☐ (f) E' possibile anticiparla nel ramo j
- ☐ (g) E' possibile anticiparla nel ramo i
- ☐ (h) E' possibile anticiparla nel ramo e
- ☐ (i) E' possibile anticiparla nel ramo g

Risposta errata.

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

Domanda 11

Risposta non data

Punteggio max.:
1,00**Progettazione DW - 1 punto (penalità 15% per ogni risposta sbagliata)**

Si vogliono analizzare le informazioni relative agli accessi che sono stati effettuati dai pazienti presso il reparto di Pronto Soccorso degli ospedali italiani. Il reparto di Pronto Soccorso garantisce gli interventi diagnostico-terapeutici di emergenza/urgenza in risposta al bisogno manifestato dall'utenza.

I pazienti possono raggiungere il reparto di Pronto Soccorso in ambulanza o autonomamente con un mezzo proprio. Ai pazienti che accedono al pronto soccorso viene assegnato un codice colore in base ai sintomi denunciati dal paziente e ai segnali vitali riscontrati. Il codice colore può assumere un solo valore tra rosso, giallo, verde e bianco in base al livello di gravità della condizione del paziente e quindi il grado di urgenza nel procedere con la visita del paziente.

Una volta ricevuto il codice colore, il paziente viene visitato dai medici in servizio presso il reparto. Qualora ritenuto necessario, il paziente è anche sottoposto ad uno o più esami medici che permettano di approfondire e chiarire la sua condizione. Al termine, viene rilasciata una diagnosi. Nei casi più critici viene effettuato il ricovero del paziente presso l'ospedale; altrimenti il paziente viene dimesso.

Si vuole analizzare (1) il numero medio di ore di permanenza in pronto soccorso per paziente e (2) il numero medio di visite per medico, in base a:

- modalità di arrivo al Pronto Soccorso (ambulanza oppure mezzo proprio)
- codice colore (bianco, verde, giallo e rosso)
- ospedale, caratterizzato dalla sua collocazione geografica (espressa in termini di città, provincia, regione), e dall'indicazione se siano presenti nell'ospedale uno o più tra i seguenti reparti: cardiologia, pediatria, ortopedia e geriatria. Per ogni ospedale è anche noto se sia una struttura pubblica o accreditata
- data, giorno del mese, giorno festivo, mese, mese dell'anno, quadrimestre, semestre, anno
- sesso e fascia di età del paziente (<13, 13-18, 19-40, 41-70, >70)
- diagnosi rilasciata, caratterizzata dall'elenco degli esami medici eseguiti. Ogni esame è caratterizzato dalla specialità medica a cui si riferisce. Inoltre, ogni esame può essere richiesto in più diagnosi diverse.

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 ore di permanenza in pronto soccorso
- ☐ (b) Numero totale visite pediatriche
- ☐ (c) Numero totale medici
- ☐ (d) Numero medio mensile di ore di permanenza in pronto soccorso
- ☐ (e) Numero totale pazienti
- ☐ (f) Numero totale visite
- ☐ (g) Numero medio giornaliero di visite

☐ (h) Numero medio di visite per regione

Risposta errata.

La risposta corretta è: Numero totale ore di permanenza in pronto soccorso, Numero totale pazienti, Numero totale medici, Numero totale visite

Domanda 12

Risposta non data

Punteggio max.:
2,00**1 punto (penalità 15% per ogni risposta sbagliata)**

Si consideri la seguente base dati transazionale.

E
ABCE
AE
AE
ABCE
ACDE
AD
ABCD
BCDE
BCE

Si applichi l'algoritmo FP-growth per l'estrazione degli itemset frequenti (minsup = 2; un itemset è frequente se compare in almeno 2 transazioni).

Quali sono i pattern contenuti all'interno del B-CPB (B Conditional Pattern Base)?
Selezionare *tutti e soli* i pattern corretti, insieme ai loro supporti.

Scegli una o più alternative:

- ☐ (a) A: 4
- ☐ (b) E: 4
- ☐ (c) C: 1
- ☐ (d) E: 3
- ☐ (e) AC: 1
- ☐ (f) CE: 2
- ☐ (g) ACE: 2
- ☐ (h) AC: 3
- ☐ (i) CE: 4
- ☐ (j) C: 5
- ☐ (k) ACE: 3
- ☐ (l) A: 3

Risposta errata.

La risposta corretta è: CE: 2, ACE: 2, AC: 1

Domanda 13

Risposta non data

Punteggio max.:
1,00**Progettazione DW - 1 punto (penalità 15% per risposta sbagliata)**

Si vogliono analizzare le informazioni relative agli accessi che sono stati effettuati dai pazienti presso il reparto di Pronto Soccorso degli ospedali italiani. Il reparto di Pronto Soccorso garantisce gli interventi diagnostico-terapeutici di emergenza/urgenza in risposta al bisogno manifestato dall'utenza.

I pazienti possono raggiungere il reparto di Pronto Soccorso in ambulanza o autonomamente con un mezzo proprio. Ai pazienti che accedono al pronto soccorso viene assegnato un codice colore in base ai sintomi denunciati dal paziente e ai segnali vitali riscontrati. Il codice colore può assumere un solo valore tra rosso, giallo, verde e bianco in base al livello di gravità della condizione del paziente e quindi il grado di urgenza nel procedere con la visita del paziente.

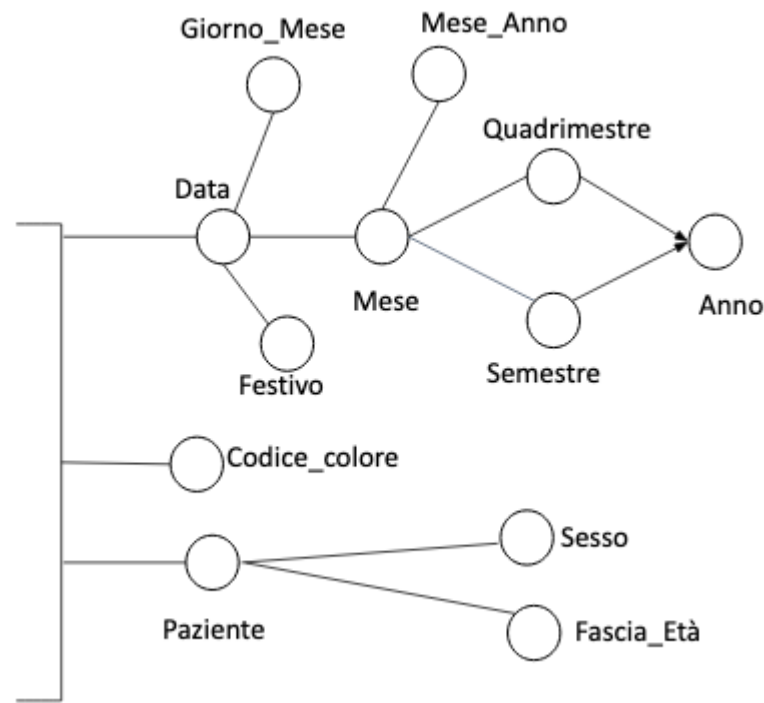
Una volta ricevuto il codice colore, il paziente viene visitato dai medici in servizio presso il reparto. Qualora ritenuto necessario, il paziente è anche sottoposto ad uno o più esami medici che permettano di approfondire e chiarire la sua condizione. Al termine, viene rilasciata una diagnosi. Nei casi più critici viene effettuato il ricovero del paziente presso l'ospedale; altrimenti il paziente viene dimesso.

Si vuole analizzare (1) il numero medio di ore di permanenza in pronto soccorso per paziente e (2) il numero medio di visite per medico, in base a:

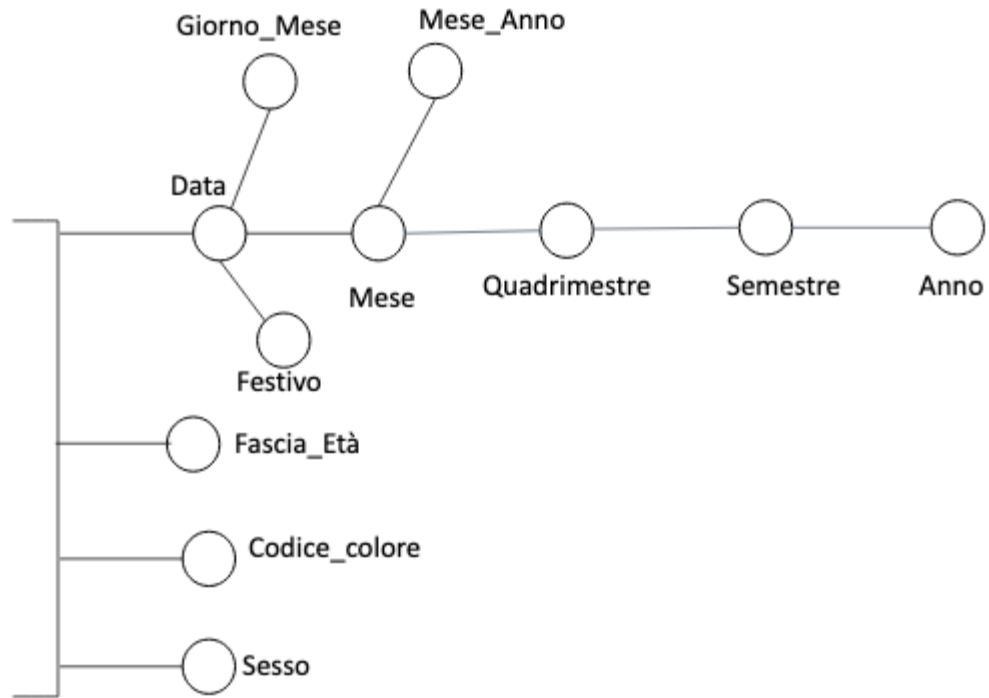
- modalità di arrivo al Pronto Soccorso (ambulanza oppure mezzo proprio)
- codice colore (bianco, verde, giallo e rosso)
- ospedale, caratterizzato dalla sua collocazione geografica (espressa in termini di città, provincia, regione), e dall'indicazione se siano presenti nell'ospedale uno o più tra i seguenti reparti: cardiologia, pediatria, ortopedia e geriatria. Per ogni ospedale è anche noto se sia una struttura pubblica o accreditata
- data, giorno del mese, giorno festivo, mese, mese dell'anno, quadrimestre, semestre, anno
- sesso e fascia di età del paziente (<13, 13-18, 19-40, 41-70, >70)
- diagnosi rilasciata, caratterizzata dall'elenco degli esami medici eseguiti. Ogni esame è caratterizzato dalla specialità medica a cui si riferisce. Inoltre, ogni esame può essere richiesto in più diagnosi diverse.

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

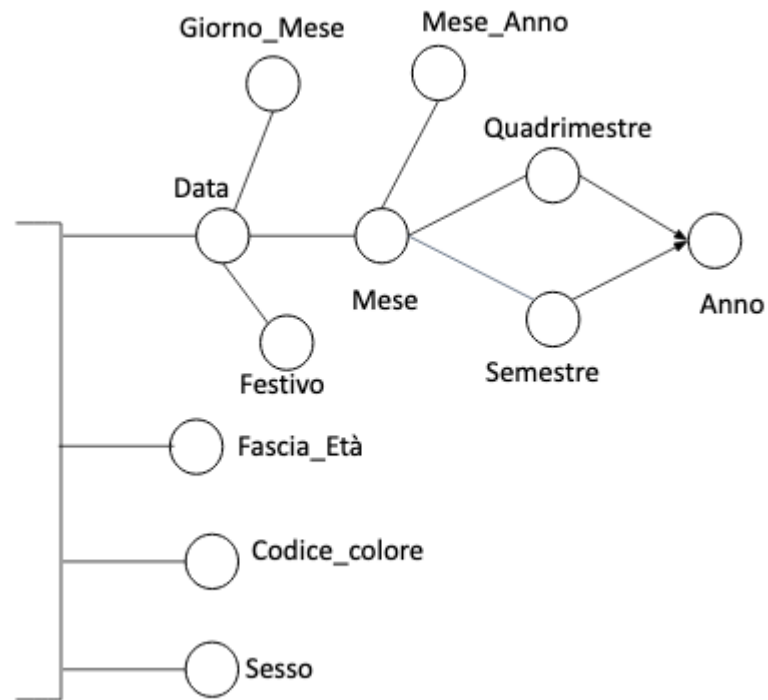
☐ (a)



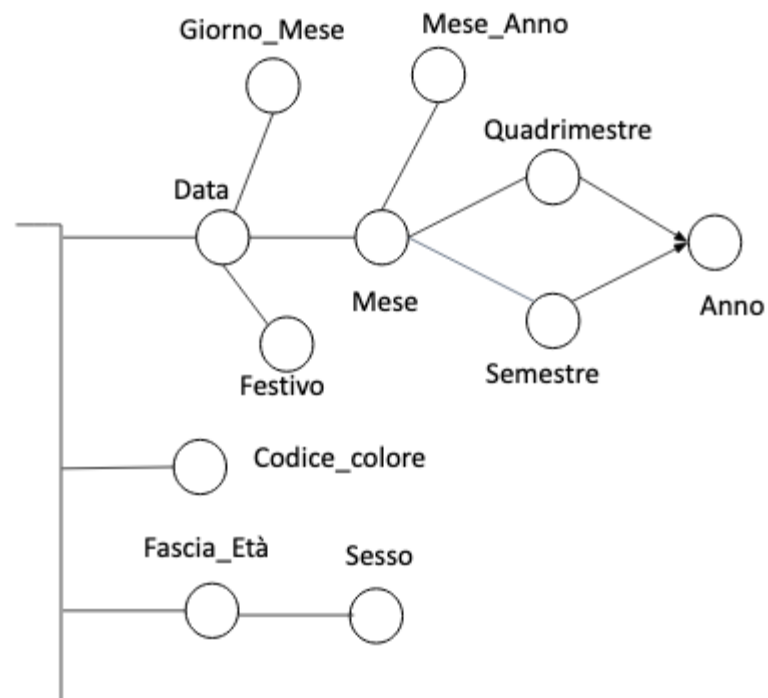
☐ (b)



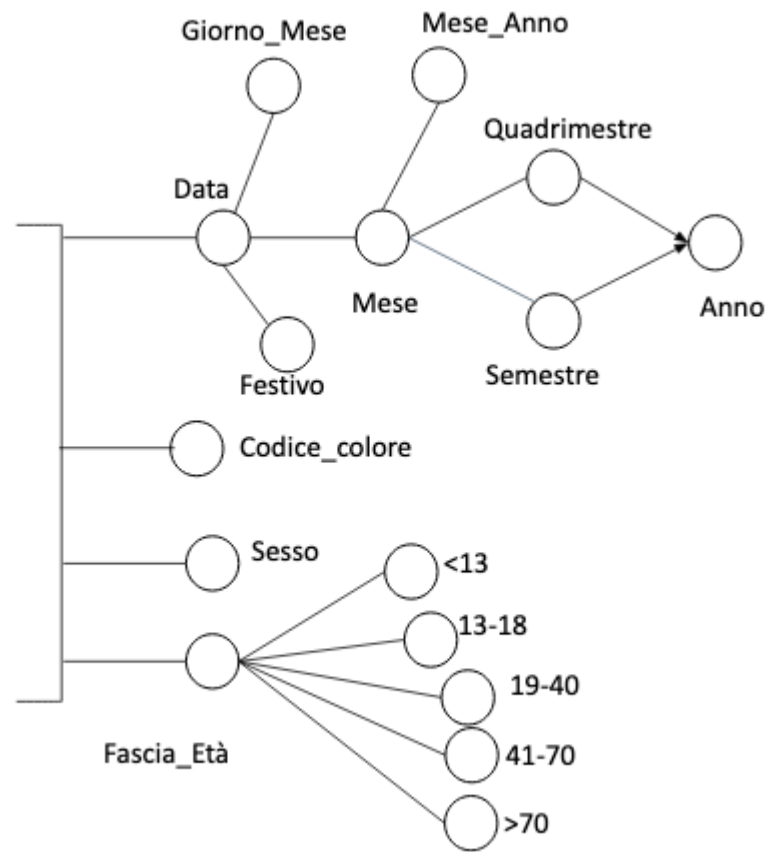
☐ (c)



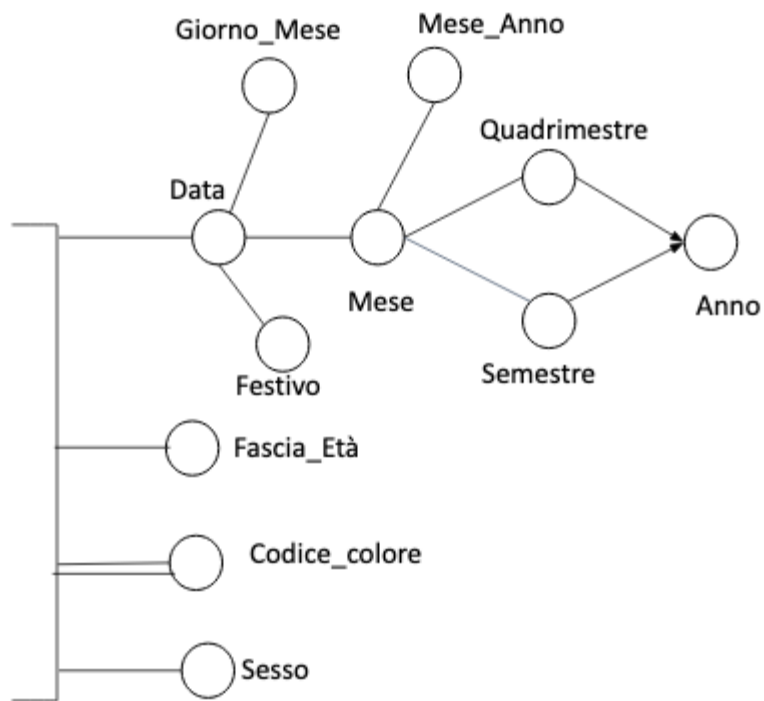
☐ (d)



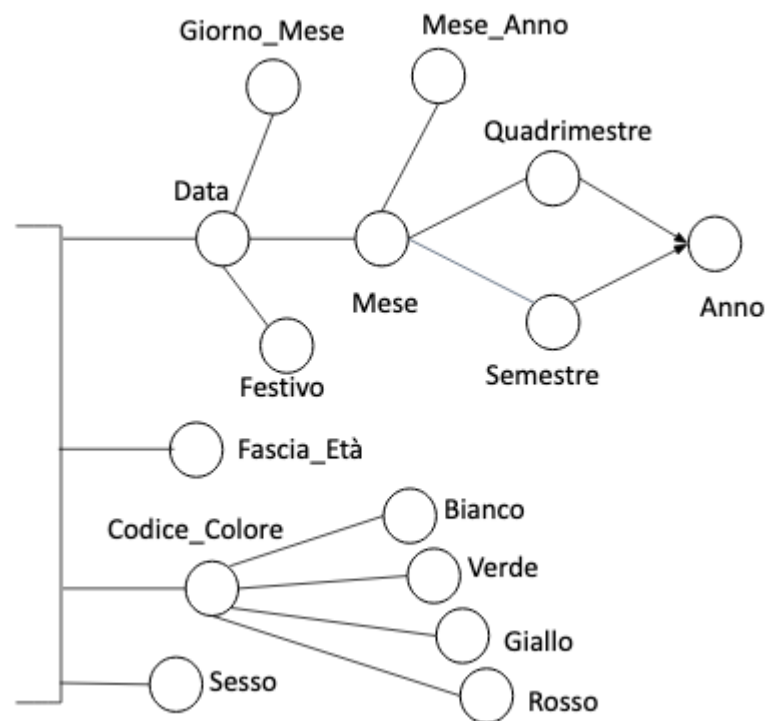
☐ (e)



☐ (f)

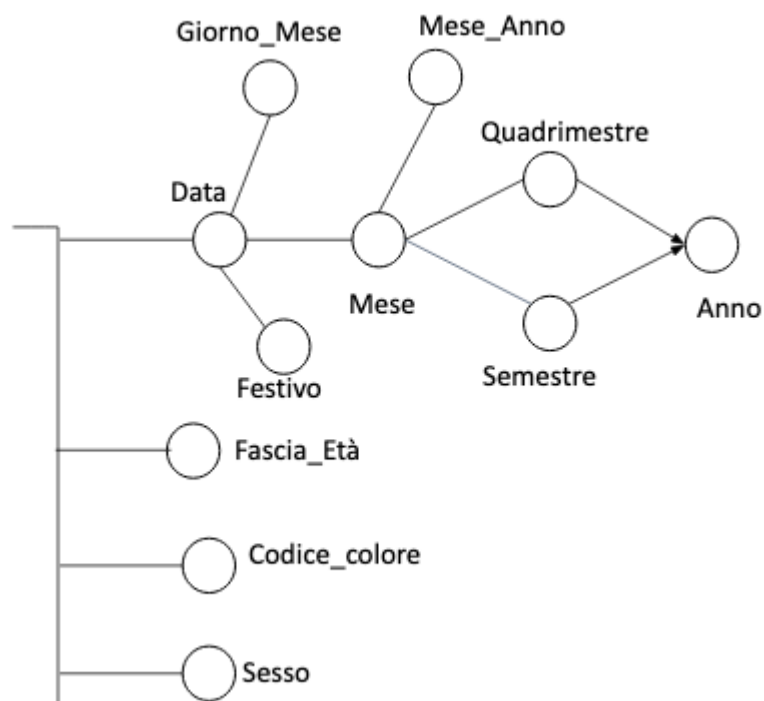


☐ (g)



Risposta errata.

La risposta corretta è:



Domanda 14

Risposta non data

Punteggio max.:
1,00**Teoria - 1 punto (penalità 15% per risposta sbagliata)**

Si consideri la seguente esecuzione del protocollo 2-phase commit per i database distribuiti. Il Transaction Manager (TM) ha inviato un messaggio di global commit a tutti i Resource Managers (RMs) che partecipano al protocollo. Tuttavia allo scadere del timeout il TM non ha ancora ricevuto tutti i messaggi di acknowledgement (ACK) dai Resource Managers (RMs).

Quale delle seguenti affermazioni è corretta?

- ☐ (a) il TM definisce "global abort" come decisione globale per la transazione
- ☐ (b) il TM invia un messaggio di global abort a tutti i RMs da cui non ha ricevuto ACK
- ☐ (c) il TM conclude immediatamente il protocollo
- ☐ (d) nessuna delle altre risposte è corretta
- ☐ (e) è ripetuta la fase 1 del protocollo 2-phase commit
- ☐ (f) il TM imposta un nuovo timeout e invia nuovamente il messaggio di global commit a tutti i RMs da cui non ha ricevuto ACK

Risposta errata.

La risposta corretta è: il TM imposta un nuovo timeout e invia nuovamente il messaggio di global commit a tutti i RMs da cui non ha ricevuto ACK

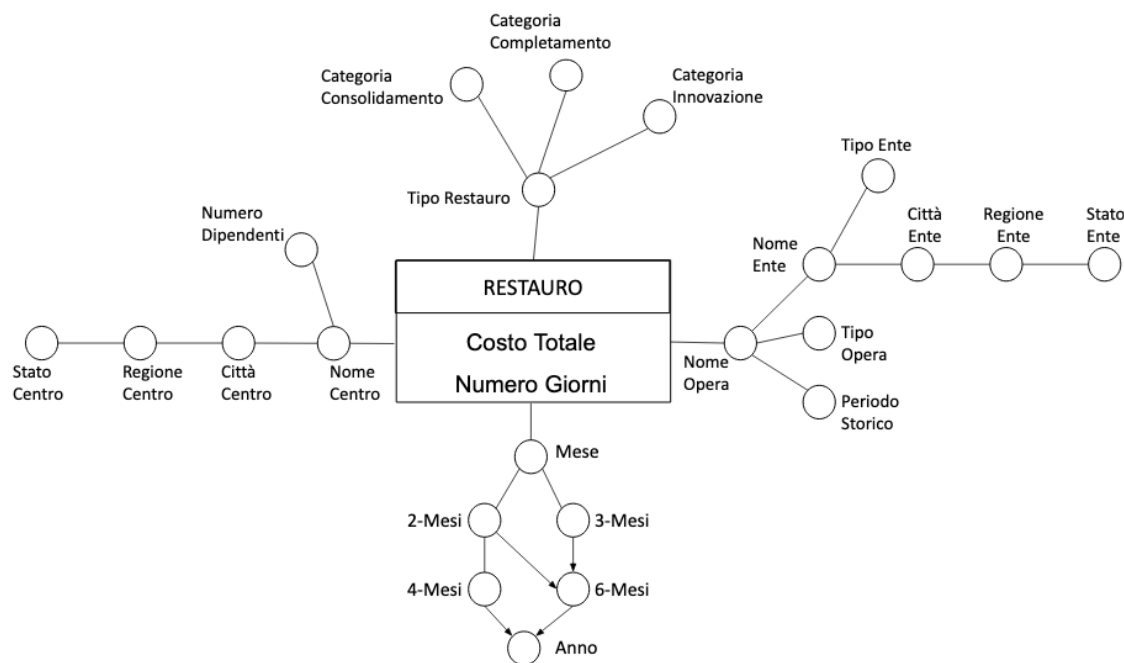
Domanda 15

Risposta non data

Punteggio max.:
3,00**SQL esteso - 3 punti (nessuna penalità per risposta sbagliata)**

Il seguente data warehouse descrive le attività di restauro svolte dai restauratori affiliati a centri di restauro, che effettuano restauri di opere (i tipi di opera possono essere dipinti, affreschi, sculture, ecc.) disponibili presso diversi enti (musei, fondazioni, chiese, ecc.). I tipi di restauro (conservativo, ecc.) possono ricadere in una o più delle seguenti tre categorie: consolidamento, completamento, innovazione (questi attributi sono booleani e rappresentano una configurazione delle categorie di restauro).

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



CENTRO_RESTAURO (IDCentro, NomeCentro, CittàCentro, RegioneCentro, StatoCentro, NumeroDipendenti)

OPERA (IDOpera, NomeOpera, TipoOpera, PeriodoStorico, NomeEnte, TipoEnte, CittàEnte, RegioneEnte, StatoEnte)

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

TIPO_RESTAURO (IDTipoRestauro, TipoRestauro, CategoriaConsolidamento, CategoriaCompletamento, CategoriaInnovazione)

RESTAURO (IDOpera, IDCentro, IDTempo, IDTipoRestauro, CostoTotale, NumeroGiorni)

Separatamente per città dell'ente e trimestre (attributo 3-Mesi), considerando solo le opere del periodo storico 'Rinascimento', visualizzare:

- il numero di tipi diversi di opera restaurati
- la percentuale del numero giorni di restauro rispetto al totale annuale per città dell'ente
- il totale cumulativo del costo al trascorrere dei trimestri, separatamente per anno

```
SELECT CittàEnte, 3-Mesi,
COUNT(DISTINCT TipoOpera),
100 * SUM(NumeroGiorni)/SUM(SUM(NumeroGiorni))
                        OVER (PARTITION BY CittàEnte, Anno) AS PercGiorni,
SUM(SUM(CostoTotale) OVER (Partition by CittàEnte, Anno
                        ORDER BY 3-Mesi
                        ROWS UNBOUNDED PRECEDING) AS CumCosto,
FROM RESTAURO R, TEMPO T, OPERA O
WHERE R.IDTempo = T.IDTempo AND R.IDOpera = O.IDOpera
AND PeriodoStorico = 'Rinascimento'
GROUP BY CittàEnte, 3-Mesi, Anno
```

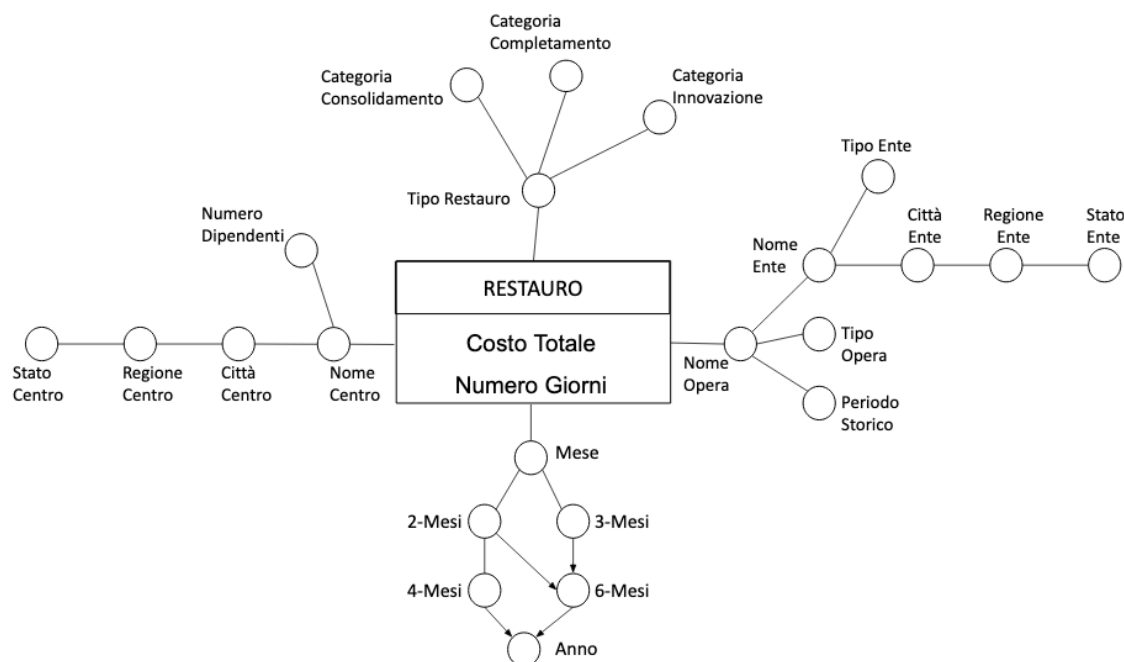
Domanda 16

Risposta non data

Punteggio max.:
4,00**SQL esteso - 4 punti (nessuna penalità per risposta sbagliata)**

Il seguente data warehouse descrive le attività di restauro svolte dai restauratori affiliati a centri di restauro, che effettuano restauri di opere (i tipi di opera possono essere dipinti, affreschi, sculture, ecc.) disponibili presso diversi enti (musei, fondazioni, chiese, ecc.). I tipi di restauro (conservativo, ecc.) possono ricadere in una o più delle seguenti tre categorie: consolidamento, completamento, innovazione (questi attributi sono booleani e rappresentano una configurazione delle categorie di restauro).

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



CENTRO_RESTAURO (IDCentro, NomeCentro, CittàCentro, RegioneCentro, StatoCentro, NumeroDipendenti)

OPERA (IDOpera, NomeOpera, TipoOpera, PeriodoStorico, NomeEnte, TipoEnte, CittàEnte, RegioneEnte, StatoEnte)

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

TIPO_RESTAURO (IDTipoRestauro, TipoRestauro, CategoriaConsolidamento, CategoriaCompletamento, CategoriaInnovazione)

RESTAURO (IDOpera, IDCentro, IDTempo, IDTipoRestauro, CostoTotale, NumeroGiorni)

Separatamente per anno e regione dell'ente, considerando solo i tipi di restauro che appartengono ad entrambe le categorie consolidamento e completamento, visualizzare

- il costo medio mensile
- la percentuale del costo totale rispetto al costo totale complessivo per anno e stato dell'ente
- la posizione in una graduatoria (rank), separatamente per anno, in ordine decrescente di numero medio di giorni di restauro mensili

```
SELECT Anno, RegioneEnte,
SUM(CostoTotale)/COUNT(DISTINCT Mese),
100 * SUM(CostoTotale)/SUM(SUM(CostoTotale)) OVER (PARTITION BY Anno, StatoEnte)
RANK() OVER (PARTITION BY Anno
              ORDER BY SUM(NumeroGiorni)/COUNT(DISTINCT Mese) DESC)
FROM RESTAURO R, TEMPO T, TIPO_RESTAURO TR, OPERA O
WHERE R.IDTempo = T.IDTempo AND R.IDTipoRestauro= TR. IDTipoRestauro AND
O.IDOpera=R.IDOpera
AND TR.CategoriaConsolidamento = 1 AND TR.CategoriaCompletamento = 1
GROUP BY Anno, RegioneEnte, StatoEnte
```