Data Science e Tecnologie per le Basi di Dati

Iniziato	sabato, 9 settembre 2023, 18:24
Stato	Completato
Terminato	sabato, 9 settembre 2023, 19:44
Tempo impiegato	1 ora 20 min.
Valutazione	**0,00** su un massimo di 31,00 (**0**%)

Domanda 1

Risposta non data

Punteggio max.: 1,00

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

Notazione (r/w)N(V): read/write dell'oggetto V da parte della transazione N

Definizioni:

- Reads-from: ri(x) reads-from wj(x) se wj(x) precede ri(x), i $\neq$ j, non ci sono altre wk(x) tra le due operazioni
- Final write: wi(x) è una final write se è l'ultima operazione di scrittura di x all'interno dello schedule

Due schedule sono view-equivalenti se hanno gli stessi set reads-from e final-write

Viene dato il seguente schedule di 4 transazioni

S = w1(x) r4(x) r4(z) w3(z) w1(y) r3(y) r2(z) w3(y)

Quale delle seguenti affermazioni è vera?

- (a) S è view serializable perchè è view equivalent allo schedule seriale w3(z) r3(y) w3(y) r4(x) r4(z) r2(z) w1(x) w1(y)
- (b) S è view serializable perchè è view equivalent allo schedule seriale r4(x) r4(z) w3(z) r3(y) w3(y) r2(z) w1(x) w1(y)
- (c) S è view serializable perchè è view equivalent allo schedule seriale w1(x) w1(y) r4(x) r4(z) r2(z) w3(z) r3(y) w3(y)
- (d) S è view serializable perchè è view equivalent allo schedule seriale w1(x) w1(y) r4(x) r4(z) w3(z) r3(y) w3(y) r2(z)
- (e) S è view serializable perchè è view equivalent allo schedule seriale w1(x) w1(y) r2(z) w3(z) r3(y) w3(y) r4(x) r4(z)
- (f) S non è view serializable perchè non è view equivalent ad alcuno schedule seriale

Risposta errata.

La risposta corretta è: S è view serializable perchè è view equivalent allo schedule seriale w1(x) w1(y) r4(x) r4(z) w3(z) r3(y) w3(y) r2(z)

Domanda 2

Risposta non data

Punteggio max.: 1,00

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

In base al teorema CAP (selezionare l'opzione corretta):

- (a) Nessuna delle altre opzioni è corretta
- (b) Un sistema può garantire solo uno fra Consistency, Availability e Partition tolerance
- (c) In caso di partizione, viene garantita la Consistency
- (d) Non è possibile avere un sistema che garantisca Consistency, Availability e Partition tolerance, tutti simultaneamente
- (e) In caso di partizione, è sempre possibile garantire sia Consistency, sia Availability
- (f) In caso di partizione, viene garantita la Availability

Risposta errata.

La risposta corretta è: Non è possibile avere un sistema che garantisca Consistency, Availability e Partition tolerance, tutti simultaneamente

Domanda 3

Risposta non data

Punteggio max.: 1,00

Teoria (MongoDB) (1 punto, penalità -15% per risposta sbagliata)
Dalla documentazione di MongoDB sugli operatori $in, $nin:

$nin

Sintassi: `{ field: { $nin: [ valore1, valore2, ... valoreN ] } }`

L'operatore $nin seleziona i documenti in cui:

- il valore *field* non e' nell'array specificato, oppure
- il valore *field* non esiste

$in

Sintassi: `{ field: { $in: [ valore1, valore2, ... valoreN ] } }`

L'operatore $in seleziona i documenti in cui il valore *field* e' uguale a uno dei valori nell'array specificato

Il seguente è un documento preso da una collection in mongoDB denominata "*purchases*", che mantiene informazioni sugli acquisti effettuati in un supermercato.

```
{ date: 2021-01-10,
  total_cost: 52.56,
  payment_method: "credit_card",
  fidelity_card_number: null,
  items: [ "uova", "birra", "mele" ]
}
```

Sulla base della struttura inferibile da questo documento, quale delle seguenti query MongoDB estrae l'elenco di acquisti che contengono uova o birra (or inclusivo), effettuati da utenti con carta fedeltà?

(a)
```
db.purchases.find({
  $or: [
    { fidelity_card_number: { $ne: null } },
    {
      items: {
        $nin: [ "uova", "birra" ]
      }
    }
  ]
})
```

(b)
```
db.purchases.find({
  $or: [
    { fidelity_card_number: { $ne: null } },
    {
      items: {
        $in: [ "uova", "birra" ]
      }
    }
  ]
})
```

(c)
```
db.purchases.find({
  $or: [
    { fidelity_card_number: null },
    {
```

```
        items: {
          $nin: [ "uova", "birra" ]
        }
      }
    ]
  })
```

(d)
```
db.purchases.find({
  fidelity_card_number: null,
  items: {
    $in: [ "uova", "birra" ]
  }
})
```

(e)
```
db.purchases.find({
  $or: [
    { fidelity_card_number: null },
    {
      items: {
        $in: [ "uova", "birra" ]
      }
    }
  ]
})
```

(f)
```
db.purchases.find({
  fidelity_card_number: { $ne: null },
  items: {
    $in: [ "uova", "birra" ]
  }
})
```

(g)
```
db.purchases.find({
  fidelity_card_number: null,
  items: {
    $nin: [ "uova", "birra" ]
  }
})
```

(h)
```
db.purchases.find({
  fidelity_card_number: { $ne: null },
  items: {
    $nin: [ "uova", "birra" ]
  }
})
```

Risposta errata.

La risposta corretta è:
```
db.purchases.find({
  fidelity_card_number: { $ne: null },
  items: {
    $in: [ "uova", "birra" ]
```

```
    }
})
```

Domanda 4

Risposta non data

Punteggio max.: 1,00

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

Per due punti n-dimensionali p1, p2, la distanza di Manhattan (cityblock) è definita come:

$$d(p1, p2) = \sum_{k=1}^{n} |p1_k - p2_k|$$

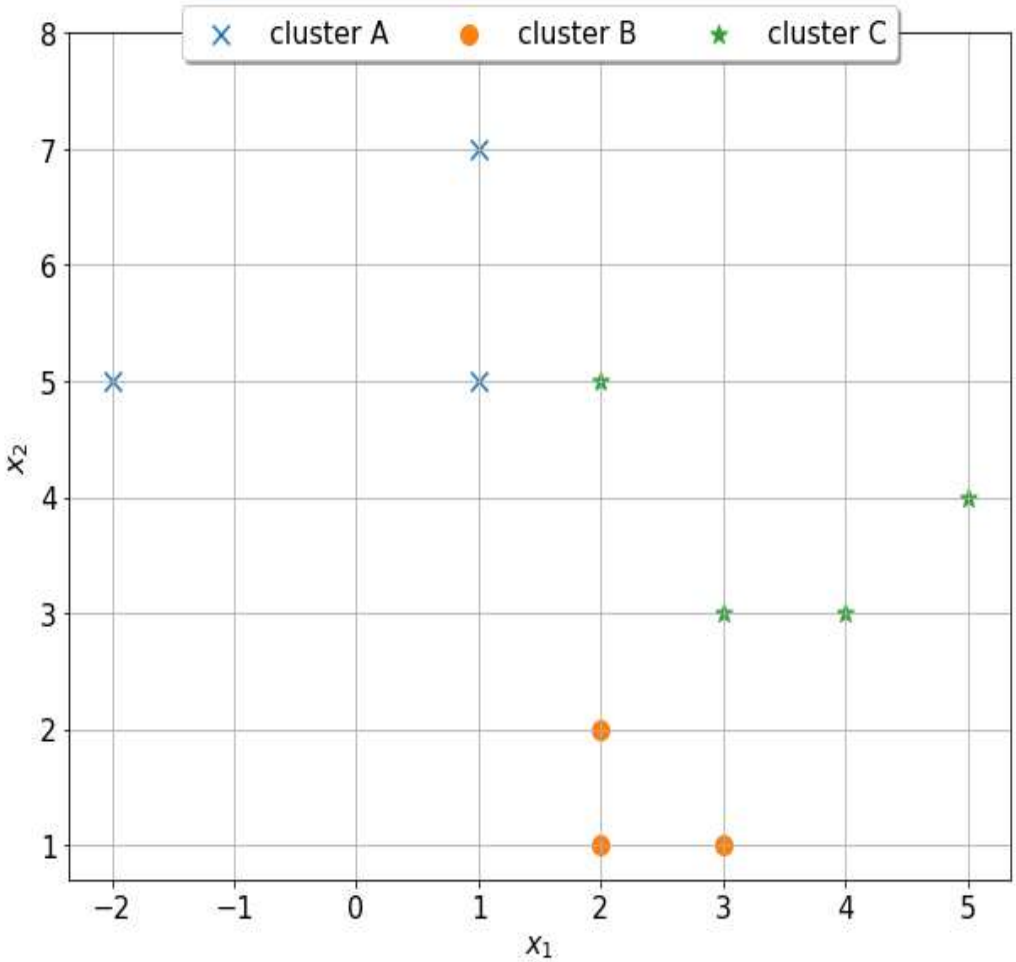
(k rappresenta la k-esima dimensione)

Per ciascun punto i, possiamo definire:

- a(i) come la dissimilarità media di i da tutti i punti dello stesso cluster
- b(i) come la piu' bassa dissimilarita' media di i dai cluster a cui i non appartiene
- $silhouette(i) = \frac{b(i)-a(i)}{max(a(i),b(i))}$

È dato un dataset bidimensionale costituito da 10 punti.

Su questo dataset viene effettuato un clustering. I tre cluster (A, B, C) sono identificati dai simboli X, O, * (vedi figura e legenda).



Si vuole valutare la qualità del clustering ottenuto tramite silhouette score. In particolare, si vuole calcolare la silhouette del punto (1, 5) appartenente al cluster A, utilizzando la distanza di Manhattan.

Quale dei seguenti è il valore corretto per la silhouette del punto specificato?

- (a) 1/6
- (b) 1/3

- (c) 4/5
- (d) 5/4
- (e) 2/5
- (f) 3/5

Risposta errata.

La risposta corretta è: 1/3

Domanda 5

Risposta non data

Punteggio max.: 2,00

Teoria (FP-Growth) (2 punti, penalità -15% per risposta sbagliata)

È dato il seguente dataset transazionale.

B D
A C D E
B D
A E
B C D E
A B C D
A B
A B C E
B C D
A B C E

Si vogliono estrarre gli itemset frequenti con minsup = 2 (un itemset e' frequente se ha almeno supporto 2) **applicando l'algoritmo FP-Growth**.

Il Conditional Pattern Base dell'itemset ED (ED-CPB) contiene i prefix-paths dell'item D all'interno dell'E-conditional FP-tree generato al passo di ricorsione precedente, insieme al loro support count.

Durante l'applicazione dell'algoritmo, quali sono i contenuti del Conditional Pattern Base di ED (ED-CPB)?

(a)

{A}	1
{C}	1

(b)

{B}	1
{C}	1

(c)

{A, C}	2
{A}	1

(d)

{A, B}	1
{B}	1

(e)

{ B, C }	1
{ B }	1

(f)

{A, C }	1
{ B, C }	1

(g)

{ A, B }	1
{ A }	1

(h)

{ B, C }	2
{ A, C }	1

Risposta errata.

La risposta corretta è:

{A, C }	1
{ B, C }	1

Domanda 6

Risposta non data

Punteggio max.: 1,50

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

Sono date le seguenti tabelle:

```
MUSEO(CodM, NomeMuseo, Indirizzo, Città, Regione,
TipoMuseo)
SERVIZIO(CodS, NomeServizio, TipoServizio)
MUSEO-OFFRE-SERVIZIO(CodM, CodS)
ABBONATO(CodAbb, Nome, Cognome, DataNascita,
TipoAbbonamento)
VISITA(CodAbb, Data, CodM, durataVisita)
```

Sono date le seguenti cardinalità:

- card(MUSEO) = 10^4 tuple,

  Valori distinti di Regione = 20
- card(SERVIZIO) = 100 tuple,

  Valori distinti di TipoServizio = 10
- card(MUSEO-OFFRE-SERVIZIO) =$2 \cdot 10^5$ tuple,
- card(ABBONATO) = 10^6 tuple,

  MIN(DataNascita) = 1/1/1920, MAX(DataNascita) = 31/12/2000

  Valori distinti di TipoAbbonamento = 5
- card(VISITA)= $3 \cdot 10^7$ tuple,

  MIN(Data) = 1/1/2005, MAX(Data) = 31/12/2019

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

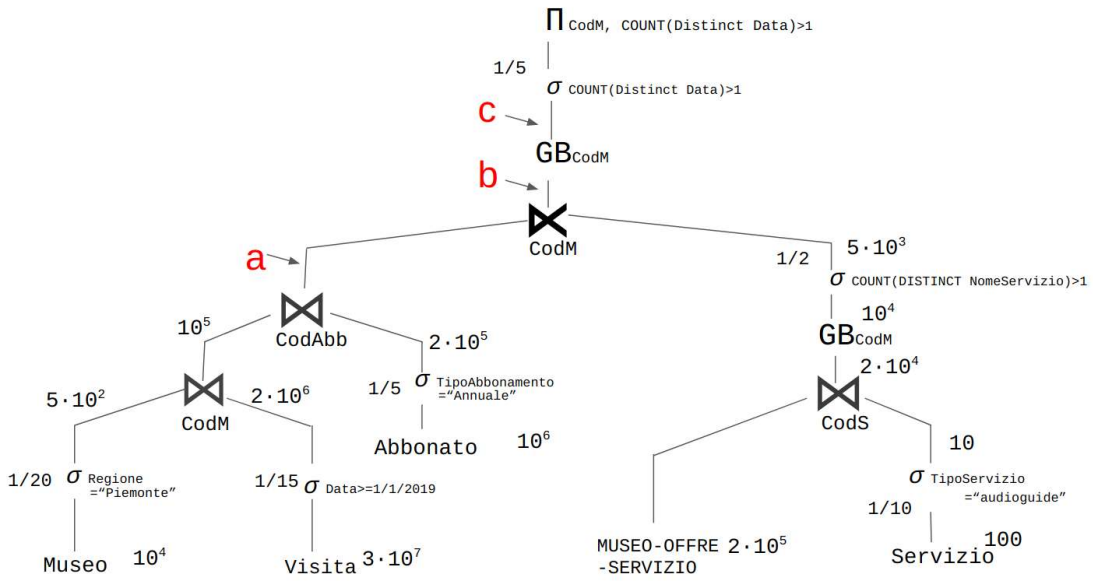
Having COUNT(DISTINCT NomeServizio)>1 = 1/2

Having COUNT(Distinct Data)>1 = 1/5

Si consideri la seguente query:

```
select V.CodM, count(Distinct V.Data)
from MUSEO M, ABBONATO ABB, VISITA V
where M.CodM=V.CodM and ABB.CodAbb=V.CodAbb and V.Data
>=1/1/2019
and ABB.TipoAbbonamento='Annuale'
and M.Regione='Piemonte'
and V.CodM IN (    SELECT MOS.CodM
                FROM SERVIZIO S, MUSEO-OFFRE-SERVIZIO M
OS
                WHERE S.CodS=MOS.CodS
                      and TipoServizio='audioguide'
                      GROUP BY MOS.CodM
                      HAVING COUNT(DISTINCT NomeServizi
                      o)>1 )
group by V.CodM
having COUNT(Distinct V.Data)>1
```

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



Specificare la cardinalità di ogni nodo indicato dai caratteri in rosso (a,b,c) nella figura. È presente una risposta corretta per ogni nodo a,b,c.

Scegli una o più alternative:

- [ ] (a) a: $2 \cdot 10^4$
- [ ] (b) a: 10^5
- [ ] (c) c: $\simeq 5 \cdot 10^2$
- [ ] (d) c: $\simeq 5 \cdot 10^3$
- [ ] (e) b: 10^4
- [ ] (f) b: 10^2
- [ ] (g) b: 10^5
- [ ] (h) a: 10^2
- [ ] (i) c: $\simeq 10^5$
- [ ] (j) b: $5 \cdot 10^3$
- [ ] (k) c: $\simeq 10^4$
- [ ] (l) a: 10^4

Risposta errata.

La risposta corretta è: a: $2 \cdot 10^4$, b: 10^4, c: $\simeq 5 \cdot 10^2$

Domanda 7

Risposta non data

Punteggio max.: 2,00

Anticipo Group By (2 punti, penalità -15% per risposta sbagliata)

Sono date le seguenti tabelle:

```
MUSEO(CodM, NomeMuseo, Indirizzo, Città, Regione,
TipoMuseo)
SERVIZIO(CodS, NomeServizio, TipoServizio)
MUSEO-OFFRE-SERVIZIO(CodM, CodS)
ABBONATO(CodAbb, Nome, Cognome, DataNascita,
TipoAbbonamento)
VISITA(CodAbb, Data, CodM, durataVisita)
```

Sono date le seguenti cardinalità:

- card(MUSEO) = 10^4 tuple,

  Valori distinti di Regione = 20
- card(SERVIZIO) = 100 tuple,

  Valori distinti di TipoServizio = 10
- card(MUSEO-OFFRE-SERVIZIO) =$2 \cdot 10^5$ tuple,
- card(ABBONATO) = 10^6 tuple,

  MIN(DataNascita) = 1/1/1920, MAX(DataNascita) = 31/12/2000

  Valori distinti di TipoAbbonamento = 5
- card(VISITA)= $3 \cdot 10^7$ tuple,

  MIN(Data) = 1/1/2005, MAX(Data) = 31/12/2019

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

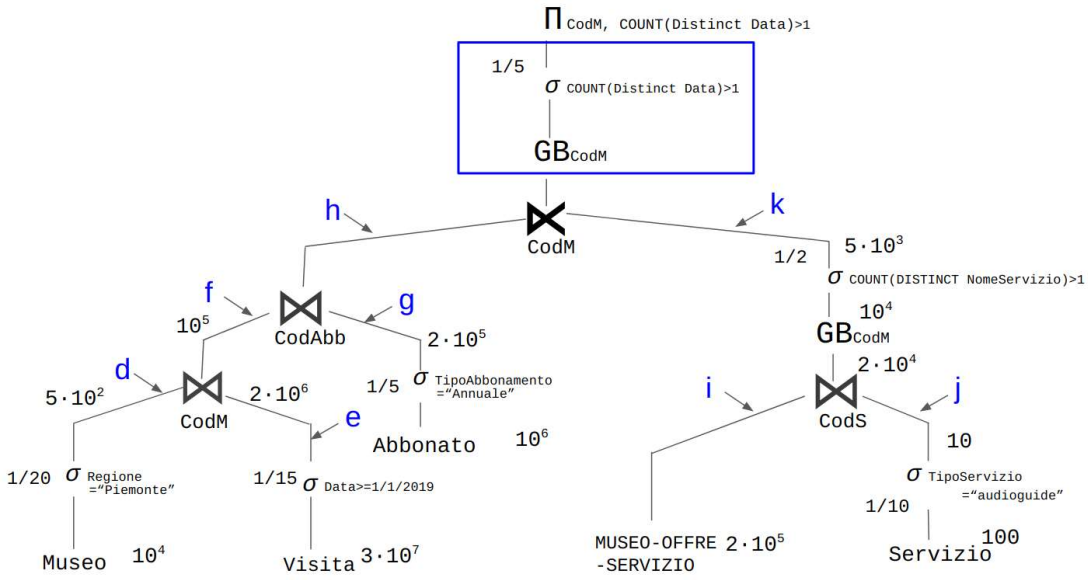
Having COUNT(DISTINCT NomeServizio)>1 = 1/2

Having COUNT(Distinct Data)>1 = 1/5

Si consideri la seguente query:

```
select V.CodM, count(Distinct V.Data)
from MUSEO M, ABBONATO ABB, VISITA V
where M.CodM=V.CodM and ABB.CodAbb=V.CodAbb and V.Data
>=1/1/2019
and ABB.TipoAbbonamento='Annuale'
and M.Regione='Piemonte'
and V.CodM IN (    SELECT MOS.CodM
                FROM SERVIZIO S, MUSEO-OFFRE-SERVIZIO M
OS
                WHERE S.CodS=MOS.CodS
                      and TipoServizio='audioguide'
                      GROUP BY MOS.CodM
                      HAVING COUNT(DISTINCT NomeServizi
                      o)>1 )
group by V.CodM
having COUNT(Distinct V.Data)>1
```

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



Analizzare l'anticipazione della GROUP BY **`group by CodM Having COUNT(DISTINCT Data)>1`**, rappresentata nel riquadro. Selezionare la soluzione che consente la massima riduzione di cardinalità (se esiste).

- (a) E' possibile anticiparla nel ramo **h**
- (b) Non e' possibile anticipare la Group By **`group by CodM Having COUNT(DISTINCT Data)>1`**
- (c) E' possibile anticiparla nel ramo **d**
- (d) E' possibile anticiparla nel ramo **e**
- (e) E' possibile anticiparla nel ramo **g**
- (f) E' possibile anticiparla nel ramo **k**
- (g) E' possibile anticiparla nel ramo **f**
- (h) E' possibile anticiparla nel ramo **i**
- (i) E' possibile anticiparla nel ramo **j**

Risposta errata.

La risposta corretta è: E' possibile anticiparla nel ramo **h**

Domanda 8

Risposta non data

Punteggio max.: 1,50

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

Sono date le seguenti tabelle:

```
MUSEO(CodM, NomeMuseo, Indirizzo, Città, Regione,
TipoMuseo)
SERVIZIO(CodS, NomeServizio, TipoServizio)
MUSEO-OFFRE-SERVIZIO(CodM, CodS)
ABBONATO(CodAbb, Nome, Cognome, DataNascita,
TipoAbbonamento)
VISITA(CodAbb, Data, CodM, durataVisita)
```

Sono date le seguenti cardinalità:

- card(MUSEO) = 10^4 tuple,

  Valori distinti di Regione = 20
- card(SERVIZIO) = 100 tuple,

  Valori distinti di TipoServizio = 10
- card(MUSEO-OFFRE-SERVIZIO) =$2 \cdot 10^5$ tuple,
- card(ABBONATO) = 10^6 tuple,

  MIN(DataNascita) = 1/1/1920, MAX(DataNascita) = 31/12/2000

  Valori distinti di TipoAbbonamento = 5
- card(VISITA)= $3 \cdot 10^7$ tuple,

  MIN(Data) = 1/1/2005, MAX(Data) = 31/12/2019

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

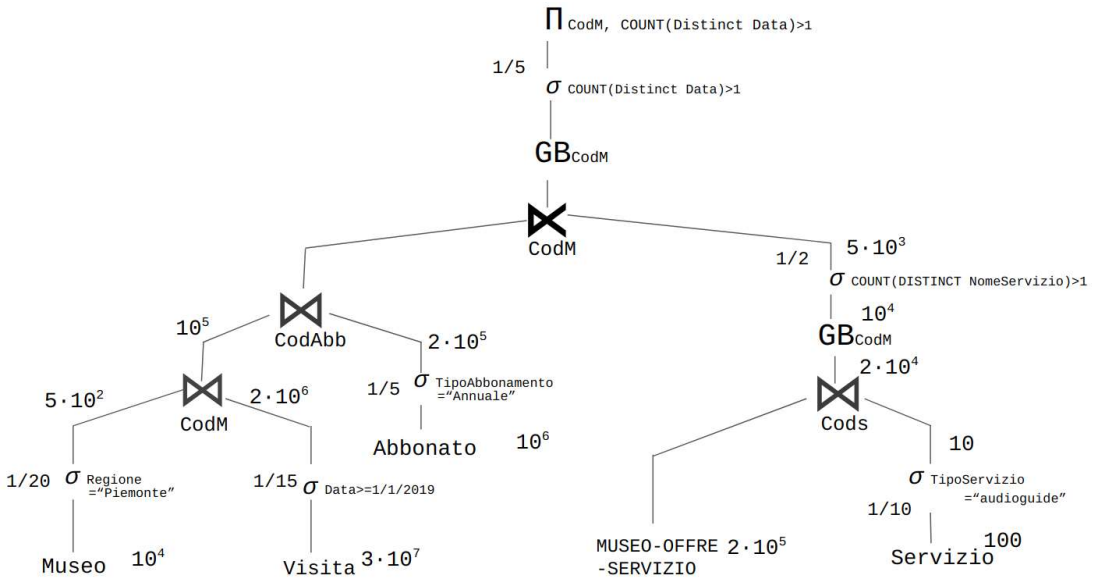
Having COUNT(DISTINCT NomeServizio)>1 = 1/2

Having COUNT(Distinct Data)>1 = 1/5

Si consideri la seguente query:

```
select V.CodM, count(Distinct V.Data)
from MUSEO M, ABBONATO ABB, VISITA V
where M.CodM=V.CodM and ABB.CodAbb=V.CodAbb and V.Data
>=1/1/2019
and ABB.TipoAbbonamento='Annuale'
and M.Regione='Piemonte'
and V.CodM IN (    SELECT MOS.CodM
                FROM SERVIZIO S, MUSEO-OFFRE-SERVIZIO M
OS
                WHERE S.CodS=MOS.CodS
                      and TipoServizio='audioguide'
                      GROUP BY MOS.CodM
                      HAVING COUNT(DISTINCT NomeServizi
                      o)>1 )
group by V.CodM
having COUNT(Distinct V.Data)>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 IndexJ ON Servizio(TipoServizio) - B+-Tree`
- (b) `CREATE INDEX IndexI ON Servizio(TipoServizio) - HASH`
- (c) `CREATE INDEX IndexD ON Visita(Data) - B+-Tree`
- (d) `CREATE INDEX IndexC ON Visita(Data) - HASH`
- (e) `CREATE INDEX IndexA ON Abbonato(TipoAbbonamento) - HASH`
- (f) `CREATE INDEX IndexF ON Museo(Regione) - HASH`
- (g) `CREATE INDEX IndexG ON Museo-offre-servizio(CodM) - B+-Tree`
- (h) `CREATE INDEX IndexH ON Museo-offre-servizio(CodM) - HASH`
- (i) `CREATE INDEX IndexB ON Abbonato(TipoAbbonamento) - B+-Tree`
- (j) `CREATE INDEX IndexE ON Museo(Regione) - B+-Tree`
- (k) `Nessuna - le strutture fisiche accessorie non migliorerebbero le prestazioni dell'interrogazione`

Risposta errata.

La risposta corretta è: `CREATE INDEX IndexD ON Visita(Data) - B+-Tree`, `CREATE INDEX IndexF ON Museo(Regione) - HASH`

Domanda 9

Risposta non data

Punteggio max.: 1,00

Schema concettuale (1 punto, penalità -15% per ogni risposta sbagliata)

Un istituto di ricerca sugli incendi vuole analizzare le informazioni raccolte da vari canali sugli incendi avvenuti in Europa negli ultimi anni.

Di ogni incendio sono noti il momento di inizio e le coordinate geografiche del luogo. Le coordinate sono associate al Distretto del Corpo Forestale che ne ha la competenza territoriale.

I vari Corpi Forestali, uno per ogni Stato dell'Unione Europea, hanno un Comando Nazionale che raccoglie le varie Sezioni locali dislocate sul territorio. Ogni Sezione raggruppa un sottoinsieme di Distretti. Un Distretto appartiene ad una sola Sezione.

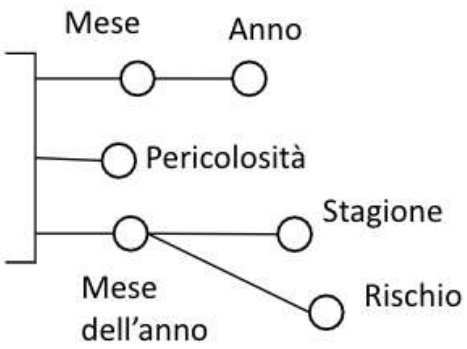
Inoltre, in base alle coordinate, il luogo dell'incendio è classificato in una o più di queste categorie geografiche: pianeggiante, collinare, montuoso, urbano, rurale, parco naturale. In base ai dati storici, il sistema associa ad ogni mese dell'anno un rischio potenziale di incendio su 4 livelli (0 basso, 1 medio, 2 alto, 3 molto alto). In base alle informazioni meteorologiche puntuali, il sistema associa ad ogni incendio una pericolosità su 3 livelli (0 basso, 1 medio, 2 alto).

Di ogni incendio è inoltre nota la durata in ore e l'estensione in km quadrati. Le informazioni di ogni incendio sono raccolte attraverso uno o più canali tra quelli elencati di seguito: satelliti, ricognizioni aeree, siti web, social network, ispezioni sul posto.

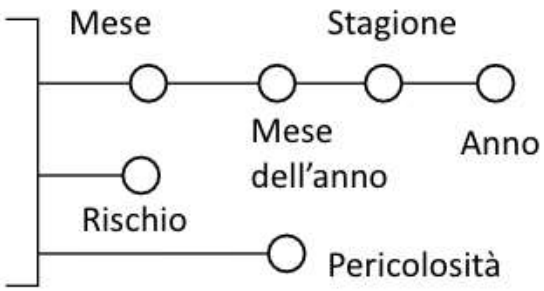
L'istituto di ricerca vuole analizzare la durata media espressa in ore e l'estensione media degli incendi espressa in km quadrati in base a:

- mese, mese dell'anno, anno, stagione (primavera dal 21 marzo al 21 giugno, estate 22/06-22/09, autunno 23/09-21/12, inverno 22/12-20/03)
- livelli di rischio, livelli di pericolosità
- categorie geografiche, Distretto, Sezione, Comando Nazionale
- canali di raccolta delle informazioni (satelliti, ricognizioni aeree, siti web, social network, ispezioni sul posto)

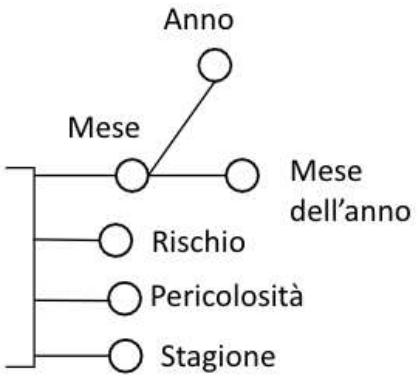
Selezionare, tra le dimensioni proposte di seguito, quella che soddisfa le richieste descritte nelle specifiche del problema.



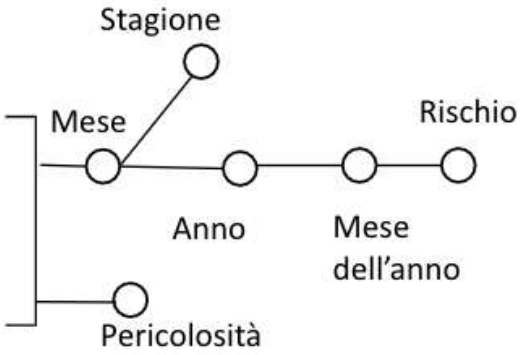
(a)



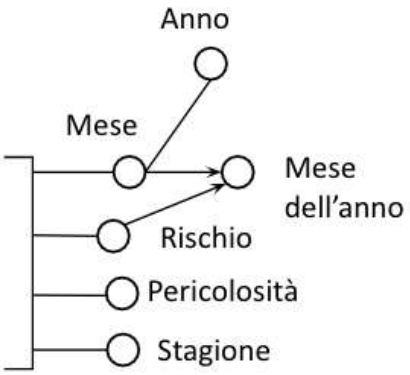
(b)


Anno
Mese
Mese
dell'anno
Rischio
Pericolosità
Stagione

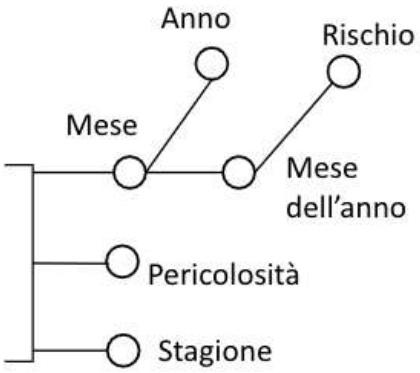
(c)


Stagione
Mese
Rischio
Anno
Mese
dell'anno
Pericolosità

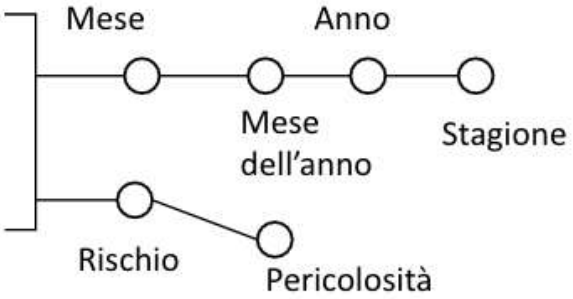
(d)


Anno
Mese
Mese
dell'anno
Rischio
Pericolosità
Stagione

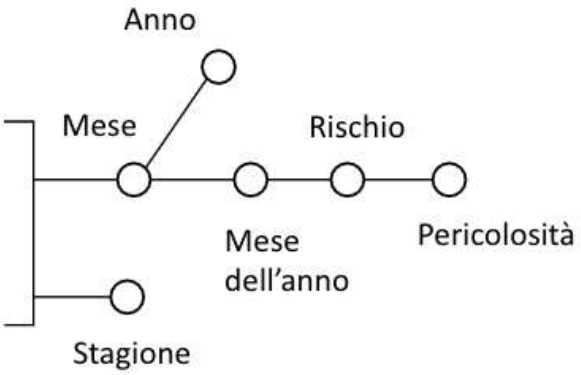
(e)


Anno
Rischio
Mese
Mese
dell'anno
Pericolosità
Stagione

(f)


Mese
Anno
Mese
dell'anno
Stagione
Rischio
Pericolosità

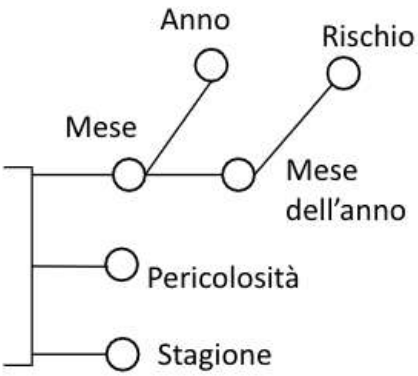
(g)



(h)

Risposta errata.

La risposta corretta è:



Domanda 10

Risposta non data

Punteggio max.: 1,00

Schema concettuale (1 punto, penalità -15% per ogni risposta sbagliata)

Un istituto di ricerca sugli incendi vuole analizzare le informazioni raccolte da vari canali sugli incendi avvenuti in Europa negli ultimi anni.

Di ogni incendio sono noti il momento di inizio e le coordinate geografiche del luogo. Le coordinate sono associate al Distretto del Corpo Forestale che ne ha la competenza territoriale.

I vari Corpi Forestali, uno per ogni Stato dell'Unione Europea, hanno un Comando Nazionale che raccoglie le varie Sezioni locali dislocate sul territorio. Ogni Sezione raggruppa un sottoinsieme di Distretti. Un Distretto appartiene ad una sola Sezione.

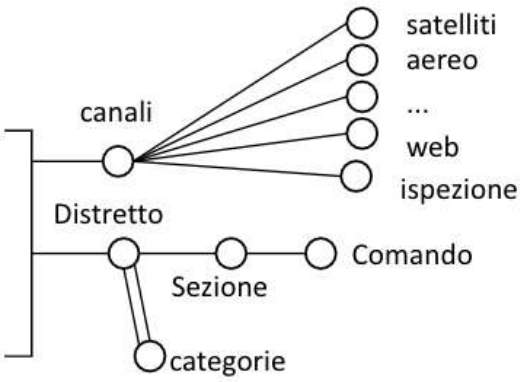
Inoltre, in base alle coordinate, il luogo dell'incendio è classificato in una o più di queste categorie geografiche: pianeggiante, collinare, montuoso, urbano, rurale, parco naturale. In base ai dati storici, il sistema associa ad ogni mese dell'anno un rischio potenziale di incendio su 4 livelli (0 basso, 1 medio, 2 alto, 3 molto alto). In base alle informazioni meteorologiche puntuali, il sistema associa ad ogni incendio una pericolosità su 3 livelli (0 basso, 1 medio, 2 alto).

Di ogni incendio è inoltre nota la durata in ore e l'estensione in km quadrati. Le informazioni di ogni incendio sono raccolte attraverso uno o più canali tra quelli elencati di seguito: satelliti, ricognizioni aeree, siti web, social network, ispezioni sul posto.

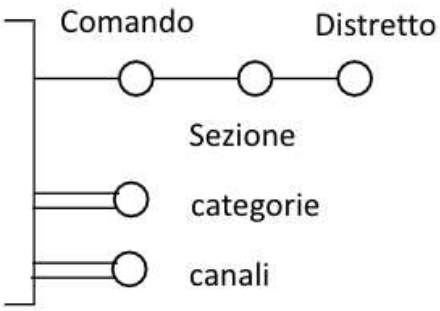
L'istituto di ricerca vuole analizzare la durata media espressa in ore e l'estensione media degli incendi espressa in km quadrati in base a:

- mese, mese dell'anno, anno, stagione (primavera dal 21 marzo al 21 giugno, estate 22/06-22/09, autunno 23/09-21/12, inverno 22/12-20/03)
- livelli di rischio, livelli di pericolosità
- categorie geografiche, Distretto, Sezione, Comando Nazionale
- canali di raccolta delle informazioni (satelliti, ricognizioni aeree, siti web, social network, ispezioni sul posto)

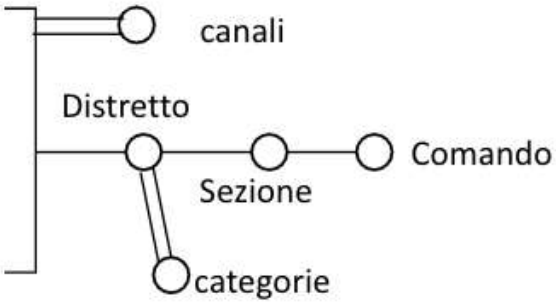
Selezionare, tra le dimensioni proposte di seguito, quella che soddisfa le richieste descritte nelle specifiche del problema.



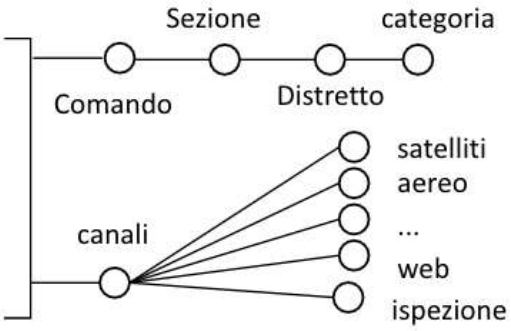
(a)



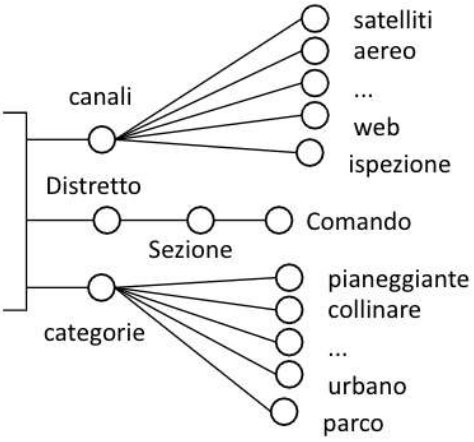
(b)


canali
Distretto
Comando
Sezione
categorie

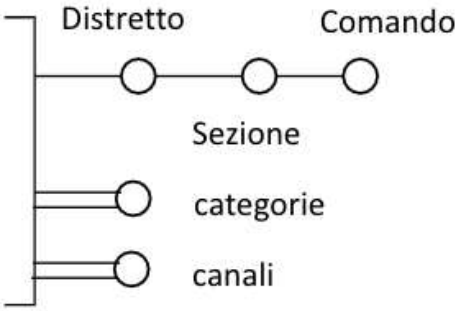
(c)


Sezione
categoria
Comando
Distretto
satelliti
aereo
...
canali
web
ispezione

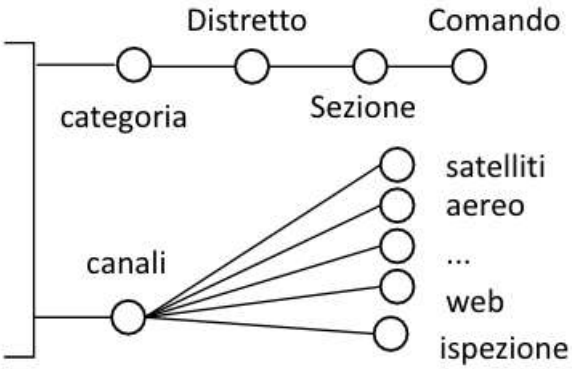
(d)


satelliti
aereo
canali
...
web
ispezione
Distretto
Comando
Sezione
pianeggiante
collinare
categorie
...
urbano
parco

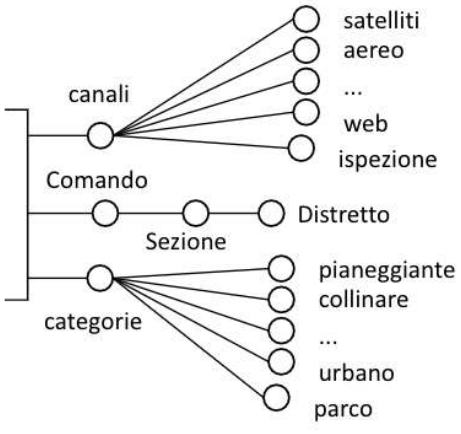
(e)


Distretto
Comando
Sezione
categorie
canali

(f)

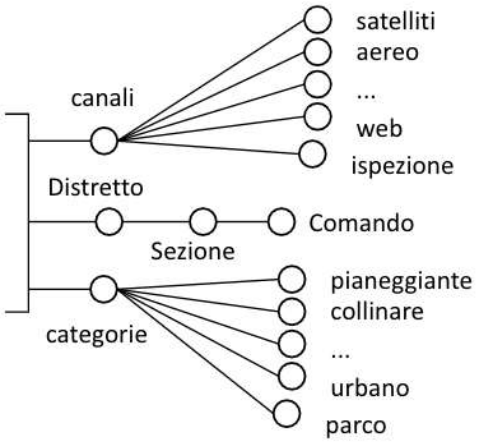

Distretto
Comando
categoria
Sezione
satelliti
aereo
canali
...
web
ispezione

(g)



(h)

Risposta errata.



La risposta corretta è:

Domanda 11

Risposta non data

Punteggio max.: 1,00

Misure (1 punto, penalità -15% per ogni risposta sbagliata)

Un istituto di ricerca sugli incendi vuole analizzare le informazioni raccolte da vari canali sugli incendi avvenuti in Europa negli ultimi anni.

Di ogni incendio sono noti il momento di inizio e le coordinate geografiche del luogo. Le coordinate sono associate al Distretto del Corpo Forestale che ne ha la competenza territoriale.

I vari Corpi Forestali, uno per ogni Stato dell'Unione Europea, hanno un Comando Nazionale che raccoglie le varie Sezioni locali dislocate sul territorio. Ogni Sezione raggruppa un sottoinsieme di Distretti. Un Distretto appartiene ad una sola Sezione.

Inoltre, in base alle coordinate, il luogo dell'incendio è classificato in una o più di queste categorie geografiche: pianeggiante, collinare, montuoso, urbano, rurale, parco naturale. In base ai dati storici, il sistema associa ad ogni mese dell'anno un rischio potenziale di incendio su 4 livelli (0 basso, 1 medio, 2 alto, 3 molto alto). In base alle informazioni meteorologiche puntuali, il sistema associa ad ogni incendio una pericolosità su 3 livelli (0 basso, 1 medio, 2 alto).

Di ogni incendio è inoltre nota la durata in ore e l'estensione in km quadrati. Le informazioni di ogni incendio sono raccolte attraverso uno o più canali tra quelli elencati di seguito: satelliti, ricognizioni aeree, siti web, social network, ispezioni sul posto.

L'istituto di ricerca vuole analizzare la durata media espressa in ore e l'estensione media degli incendi espressa in km quadrati in base a:

- mese, mese dell'anno, anno, stagione (primavera dal 21 marzo al 21 giugno, estate 22/06-22/09, autunno 23/09-21/12, inverno 22/12-20/03)
- livelli di rischio, livelli di pericolosità
- categorie geografiche, Distretto, Sezione, Comando Nazionale
- canali di raccolta delle informazioni (satelliti, ricognizioni aeree, siti web, social network, ispezioni sul posto)

Selezionare dall'elenco tutti e solo 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) Durata media degli incendi
- [ ] (b) Durata totale degli incendi
- [ ] (c) Numero medio giornaliero di incendi
- [ ] (d) Rapporto tra durata ed estensione degli incendi
- [ ] (e) Numero totale di incendi
- [ ] (f) Numero medio di incendi per kilometro quadrato
- [ ] (g) Estensione media degli incendi
- [ ] (h) Estensione totale degli incendi

Risposta errata.

La risposta corretta è: Numero totale di incendi, Durata totale degli incendi, Estensione totale degli incendi

Domanda 12

Risposta non data

Punteggio max.: 4,00

SQL Esteso 1 (4 punti)

Il seguente schema relazionale è relativo ad un datawarehouse per l'analisi del conferimento di rifiuti industriali sul territorio nazionale. Ogni tipo di rifiuto appartiene ad una sola categoria. Alcuni rifiuti sono etichettati come rifiuti speciali. La quantità indica il numero di oggetti smaltiti, il peso è espresso in kg, il valore in euro.

Smaltimento (CodT, CodP, CodD, CodR, quantità, peso, valore)
Tempo (CodT, Data, Mese, 2M, 3M, 4M, 6M, Anno, GiornoDellaSettimana, Festivo)
Prelievo (CodP, NomeAzienda, CodCittà)
Discarica(CodD, NomeDiscarica, CodCittà)
Città (CodCittà, Città, Provincia, Regione)
Rifiuti (CodR, Tipo, Categoria, Speciale)

Considerando solo i rifiuti speciali, separatamente per ogni mese e per ogni tipo di rifiuto calcolare:

- il numero di discariche diverse presso le quali sono stati conferiti i rifiuti
- il valore medio giornaliero dei rifiuti smaltiti
- il valore percentuale rispetto al totale mensile della categoria di rifiuti
- il valore mensile cumulativo dall'inizio del semestre (6M)

```
SELECT Mese, Tipo,
COUNT(DISTINCT CodD),
SUM(valore)/COUNT(DISTINCT Data),
100*SUM(valore) / SUM(SUM(valore)) OVER (PARTITION BY Mese, Categoria),
SUM(SUM((valore)) OVER (PARTITION BY Tipo, 6M
                        ORDER BY Mese
                        ROWS UNBOUNDED PRECEDING),
FROM Smaltimento S, Tempo T, Rifiuti R
WHERE S.CodT=T.CodT  AND S.CodR=R.CodR AND Speciali = TRUE
GROUP BY Mese, Tipo, Categoria, 6M;
```

Domanda 13

Risposta non data

Punteggio max.: 4,00

SQL Esteso 2 (4 punti)

Il seguente schema relazionale è relativo ad un datawarehouse per l'analisi del conferimento di rifiuti industriali sul territorio nazionale. Ogni tipo di rifiuto appartiene ad una sola categoria. Alcuni rifiuti sono etichettati come rifiuti speciali. La quantità indica il numero di oggetti smaltiti, il peso è espresso in kg, il valore in euro.

Smaltimento (CodT, CodP, CodD, CodR, quantità, peso, valore)
Tempo (CodT, Data, Mese, 2M, 3M, 4M, 6M, Anno, GiornoDellaSettimana, Festivo)
Prelievo (CodP, NomeAzienda, CodCittà)
Discarica(CodD, NomeDiscarica, CodCittà)
Città (CodCittà, Città, Provincia, Regione)
Rifiuti (CodR, Tipo, Categoria, Speciale)

Separatamente per ogni provincia della discarica e per ogni trimestre, calcolare:

- il peso medio degli oggetti smaltiti (i.e., quanto pesa in media un oggetto smaltito)
- la percentuale di oggetti smaltiti rispetto al totale trimestrale della regione della discarica
- la percentuale di oggetti smaltiti rispetto al totale annuale della provincia della discarica
- assegnare un rank trimestrale ad ogni provincia della discarica in base al peso totale degli oggetti smaltiti (la prima provincia è quella col totale più elevato)

```
SELECT Provincia, 3M,
SUM(peso) / SUM(quantità),
100*SUM(quantità) / SUM(SUM(quantità)) OVER (PARTITION BY 3M, Regione),
100*SUM(quantità) / SUM(SUM(quantità)) OVER (PARTITION BY Provincia, Anno) ,
RANK() OVER (PARTITION BY 3M ORDER BY SUM(peso) DESC)
FROM Smaltimento S, Tempo T, Discarica D, Città C
WHERE S.CodT=T.CodT AND S.CodD=D.CodD AND D.CodCittà=C.CodCittà
GROUP BY Provincia, 3M, Regione, Anno;
```

Domanda 14

Risposta non data

Punteggio max.: 7,00

Trigger 1 (7 points)

Sono date le relazioni seguenti (le chiavi primarie sono sottolineate)

CONSEGNA_ COMPLETATA (CodRichiesta, CodRider, Data, Ora)
RIDER (CodRider, Nome, Cognome)
CONSEGNE_TOTALI (CodRider, Data, NumeroConsegneTotali, ImportoConsegneTotali)
RICHIESTA_CONSEGNA (CodRichiesta, Data, Ora, Importo, IndirizzoRistorante, IndirizzoConsegna, DistanzaKm)
CONSEGNA_ASSEGNATA (CodRichiesta, CodRider)
NOTIFICA (CodN, CodRider, NumConsegneDaEffettuare)

Si vogliono gestire in modo automatico alcune attività di una società che effettua consegne di cibo a domicilio. Si scriva il trigger per gestire la seguente attività.

Completamento di una consegna.
Un rider (fattorino) completa una consegna di cibo (inserimento di un nuovo record nella tabella CONSEGNA_COMPLETATA). Il trigger deve svolgere le seguenti attività.

(1) Deve aggiornare la tabella CONSEGNE_TOTALI modificando, per il rider che ha effettuato la consegna, il numero totale di consegne effettuate e l'importo totale delle consegne nella data considerata (attributo Data). È necessario gestire anche la situazione in cui la consegna considerata sia, per il rider, la prima della giornata e la corrispondente tupla non sia ancora presente nella tabella CONSEGNE_TOTALI. L'informazione sull'importo della consegna appena effettuata è presente nella tabella RICHIESTA_CONSEGNA. La consegna completata corrisponde alla richiesta identificata dall'attributo CodRichiesta nelle tabelle RICHIESTA_CONSEGNA e CONSEGNA_ASSEGNATA.

(2) Deve eliminare le informazioni relative alla consegna completata dalle tabelle RICHIESTA_CONSEGNA e CONSEGNA_ASSEGNATA.

(3) Deve verificare se vi sono consegne ancora da effettuare, per la data considerata, nella tabella RICHIESTA_CONSEGNA. Una consegna è ancora da effettuare se la richiesta di consegna (identificata da CodRichiesta) è presente nella tabella RICHIESTA_CONSEGNA, ma non nella tabella CONSEGNA_ASSEGNATA. Nel caso esista almeno una consegna da effettuare, deve notificare al rider il numero di consegne ancora da effettuare (inserimento nella tabella NOTIFICA, si gestisca anche il caso di tabella vuota). L'attributo CodN è un contatore che deve essere incrementato ogni volta che viene inserita una nuova notifica. Se non vi sono consegne da effettuare, il trigger termina senza svolgere altre attività.

```
 Create trigger ConsegnaCompletata
After insert on CONSEGNA_COMPLETATA
For each row
Declare
N number;
myImporto number;
X number;
Begin
---- Leggo l'importo della consegna
Select Importo into myImporto
From RICHIESTA_CONSEGNA
Where CodRichiesta = :new.CodRichiesta;

----- Verifico se è la prima consegna del rider
Select count(*) into N
From CONSEGNE_TOTALI
Where CodRider = :new.CodRider
And Data = :new.Data;

If (N<>0) then
---- Il rider ha già fatto consegne
```

```
---- aggiorno le consegne totali del rider
   Update CONSEGNE_TOTALI
   Set NumeroConsegneTotali = NumeroConsegneTotali + 1
         ImportoConsegneTotali = ImportoConsegneTotali + myImporto
   Where CodRider = :new.CodRider
   And Data = :new.Data;
Else
---- prima consegna della giornata per il rider
---- inserisco la nuova tupla
   Insert into CONSEGNE_TOTALI (CodRider, Data, NumeroConsegneTotali,
ImportoConsegneTotali)
   Values (:new.CodRider,:new.Data,1,myImporto);
endif;

---- Elimino le informazioni dalle tabelle
Delete from CONSEGNA_ASSEGNATA
Where CodRichiesta = :new.CodRichiesta;

delete from RICHIESTA_CONSEGNA
where CodRichiesta = :new.CodRichiesta;

---- verifico se esistono consegne da fare
Select count(*) into N
From RICHIESTA_CONSEGNA
Where Data = new.Data
And CodRichiesta not in
   (select CodRichiesta
    From CONSEGNA_ASSEGNATA);

If (N <>0) then
---- ci sono consegne da fare
---- notifico la disponibilità
   Select max(CodN) into X
   From NOTIFICA;
   If (X is null) then
      X:= 0;
   End if;
   Insert into NOTIFICA(CodN, CodRider, X)
   Values (X+1, :new.CodRider);
end if;
end;
```

Domanda 15

Risposta non data

Punteggio max.: 2,00

Trigger 2 (2 points)

Sono date le relazioni seguenti (le chiavi primarie sono sottolineate)

CONSEGNA_ COMPLETATA (CodRichiesta, CodRider, Data, Ora)
RIDER (CodRider, Nome, Cognome)
CONSEGNE_TOTALI (CodRider, Data, NumeroConsegneTotali, ImportoConsegneTotali)
RICHIESTA_CONSEGNA (CodRichiesta, Data, Ora, Importo, IndirizzoRistorante, IndirizzoConsegna, DistanzaKm)
CONSEGNA_ASSEGNATA (CodRichiesta, CodRider)
NOTIFICA (CodN, CodRider, NumConsegneDaEffettuare)

Si vogliono gestire in modo automatico alcune attività di una società che effettua consegne di cibo a domicilio. Si scriva il trigger per gestire la seguente attività.

Vincolo di integrità sull'importo minimo di una richiesta di consegna.
Quando la distanza in chilometri (attributo DistanzaKm della tabella RICHIESTA_CONSEGNA) tra ristorante e luogo di consegna è maggiore di 4Km, l'importo minimo della richiesta (attributo Importo) deve essere maggiore o uguale a 10. In caso di operazione di modifica della tabella RICHIESTA_CONSEGNA che causino la violazione del vincolo, si deve assegnare all'attributo Importo un valore pari a 10. Valutare attentamente *tutti* gli eventi innescanti sulla tabella RICHIESTA_CONSEGNA. Nel caso di UPDATE, si indichino esplicitamente *tutti e soli* gli attributi per cui l'azione di UPDATE deve innescare il trigger.

```
Create trigger check_min_importo
Before insert or update of Importo, DistanzaKm
on RICHIESTA_CONSEGNA
for each row
when new.Importo <10 and new.DistanzaKm > 4
Declare
Begin
   :new.Importo := 10;
End;
```

Domanda 16

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).

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