

Data Science e Tecnologie per le Basi di Dati

Iniziato sabato, 9 settembre 2023, 18:26

Stato Completato

Terminato sabato, 9 settembre 2023, 20:05

Tempo impiegato 1 ora 38 min.

In ritardo 3 min. 52 secondi

Valutazione 0,00 su un massimo di 31,00 (0%)

Domanda 1

Risposta non data

Punteggio max.:

1,00

1 punto (penalità 15% per ogni risposta sbagliata)

Si vogliono analizzare le informazioni relative ad una piattaforma online che offre esercizi di inglese. La piattaforma permette agli utenti svolgere esercizi su diversi argomenti della grammatica inglese e a diversi livelli di difficoltà, e di verificare in modo automatico se la soluzione proposta sia corretta. La piattaforma assegna un punteggio agli esercizi svolti.

La piattaforma offre un ampio numero di esercizi in modalità open access, insieme ad alcuni esercizi a pagamento. Per guidare l'utente nella selezione degli esercizi di interesse, ciascun esercizio è descritto con l'indicazione degli argomenti di grammatica affrontati nell'esercizio e il livello di difficoltà dell'esercizio (ad esempio se è un esercizio di livello 'elementary' oppure 'advanced'). Per supportare le analisi di interesse sono anche memorizzati l'età dell'utente che sta svolgendo l'esercizio, l'elenco dei certificati di inglese conseguiti e la provenienza geografica dell'utente.

Si vuole analizzare il tempo medio per lo svolgimento di ciascun esercizio e il punteggio medio ottenuto sugli esercizi svolti in base a:

- data, giorno della settimana, mese, bimestre, trimestre, semestre, anno
- fascia oraria (un valore tra 'mattina', 'pomeriggio', 'sera' e 'notte')
- età dell'utente (un valore tra <15, tra 16 e 20, tra 21 e 30, maggiore di 31) e il titolo di studio più alto conseguito (ad esempio Laurea triennale, oppure laurea magistrale o diploma di scuola superiore)
- città, regione e nazione dell'utente che ha svolto esercizio
- livello di complessità dell'esercizio (un valore tra 'elementary', intermediate, 'advanced' e 'proficiency') e l'indicazione se l'esercizio sia disponibile in modalità open-access oppure a pagamento
- tematica dell'esercizio (ad esempio 'verbi', oppure 'frasi interrogative' o 'pronomi') e per ogni tematica l'elenco degli argomenti di grammatica oggetto dell'esercizio (ad esempio 'present tense' e 'passive voice' oppure 'interrogative form' e 'passive voice')
- la lista dei certificati di inglese posseduti dall'utente che sta svolgendo il test (uno o più valori tra 'PET', 'FIRST', 'ADVANCED', 'PROFICIENCY')

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) Tempo totale per lo svolgimento degli esercizi
- (b) Punteggio totale conseguito sugli esercizi svolti
- (c) Numero totale di esercizi svolti
- (d) Punteggio medio conseguito sugli esercizi svolti
- (e) Tempo totale per svolgimento degli esercizi open-access
- (f) Numero totale utenti
- (g) Durata media per lo svolgimento degli esercizi
- (h) Numero totale di esercizi svolti mensilmente

Risposta errata.

La risposta corretta è: Numero totale di esercizi svolti, Tempo totale per lo svolgimento degli esercizi, Punteggio totale conseguito sugli esercizi svolti

Domanda 2

Risposta non data

Punteggio max.: 10

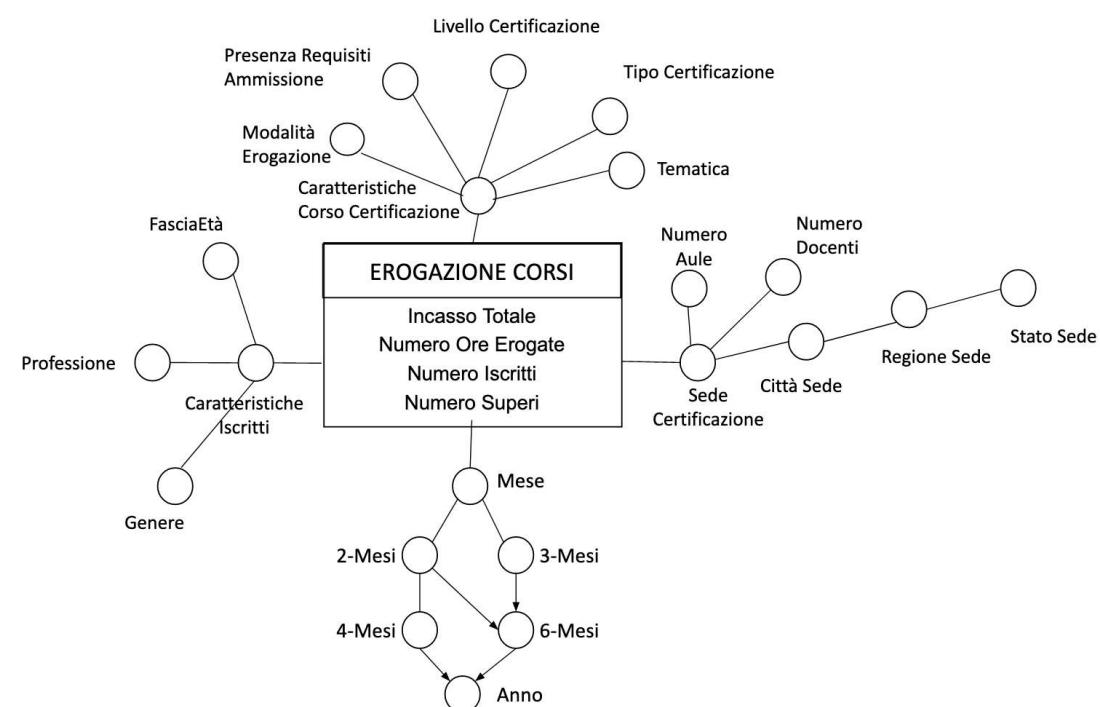
4,00

4 punti (nessuna penalità per risposta sbagliata)

Il seguente data warehouse descrive l'andamento dei corsi di certificazione erogati da una società internazionale di certificazione. I corsi di certificazione sono erogati in sedi di certificazione, che sono sedi della società distribuite in tutto il mondo. Le sedi sono anche caratterizzate dal numero di aule e di docenti disponibili. Le metriche da analizzare sono l'incasso totale dei corsi, il numero di ore erogate, il numero di persone iscritte e il numero di persone che hanno superato l'esame di certificazione alla fine del corso (NumeroSuperi).

Il data warehouse memorizza le caratteristiche principali dei corsi e relative certificazioni in termini di tematica, modalità di erogazione (presenza, online, mista), tipo di certificazione e relativo livello (base, intermedio o avanzato), e presenza o meno di prerequisiti (PresenzaRequisitiAmmissione, attributo booleano). Gli iscritti sono caratterizzati da fascia di età (≤ 20 , > 21 and ≤ 30 , > 31 and ≤ 45 , > 45), professione e genere (M, F).

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



CARATTERISTICHE-CORSO-CERTIFICAZIONE (IDCarattCorsoCert, Tematica, TipoCertificazione, LivelloCertificazione, PresenzaRequisitiAmmissione, ModalitàErogazione)

CARATTERISTICHE-ISCRITTI (IDCarattIscritti, FasciaEtà, Professione, Genere)

SEDE-CERTIFICAZIONE (IDSedeCertificazione, CittàSede, RegioneSede, StatoSede, NumeroAule, NumeroDocenti)

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

EROGAZIONE-CORSI ([IDCarattCorsoCert](#), [IDCarattIscritti](#), [IDSedeCertificazione](#), [IDTempo](#),
o_IncassoTotale, NumeroOreErogate, NumeroIscritti, NumeroSuperi)

Considerando le sedi con un numero di docenti maggiore di 50, separatamente per ogni fascia di età degli iscritti, anno e città della sede, visualizzare

- Il numero totale di ore erogate e il relativo incasso
 - Il numero medio di ore erogate per bimestre (attributo 2-Mesi)
 - La percentuale dell'incasso totale rispetto all'incasso complessivo della regione della sede, anno e fascia di età degli iscritti

- la posizione in una graduatoria (rank) in ordine decrescente rispetto al numero totale di ore erogate, separatamente per stato della sede.

```
SELECT FasciaEtà, Anno, CittàSede,  
SUM(NumerOreErogate), SUM(IncassoTotale),  
SUM(NumerOreErogate)/ COUNT (DISTINCT 2-Mesi)  
100* SUM(IncassoTotale)/SUM(SUM(IncassoTotale)) OVER (PARTITION BY Anno,  
RegioneSede, FasciaEtà),  
RANK() OVER (PARTITION BY StatoSede ORDER BY SUM(NumerOreErogate)  
DESC)  
FROM EROGAZIONE-CORSI EC, TEMPO T, SEDE-CERTIFICAZIONE CO,  
CARATTERISTICHE-ISCRITTI E  
WHERE EC.IDTempo=T.IDTempo AND  
EC.IDSedeCertificazione=CO.IDSedeCertificazione AND  
E.IDCarattCorsoCert=EC.CarattCorsoCert  
AND NUmeroDocenti > 50  
GROUP BY FasciaEtà, Anno, CittàSede, RegioneSede, StatoSede
```

Domanda 3

Risposta non data

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

Sono date le seguenti tabelle:

NAVE(CodN, TipoNave, Capienza, AnnoCostruzione, CodC)
COMPAGNIA(CodC, Nome, Indirizzo, Città, Paese, DataFondazione)
UTENTE(CodU, Nome, Cognome, DataNascita, Indirizzo, Città, Paese)
PRENOTAZIONE(CodN, DataInizio, CodU, DataFine)

Sono date le seguenti cardinalità:⁴

- card(NAVE) = $2 \cdot 10^4$ tuple,
valori distinti di TipoNave = 4
 $\text{MIN}(\text{AnnoCostruzione})= 1990$, $\text{MAX}(\text{AnnoCostruzione})= 2009$
- card(COMPAGNIA) = $5 \cdot 10^2$ tuple,
valori distinti di Paese = 5
- card(UTENTE) = $8 \cdot 10^5$ tuple,
valori distinti di Paese = 20
- card(PRENOTAZIONE)= $12 \cdot 10^6$ tuple,
 $\text{MIN}(\text{DataInizio}) = 1/1/2007$, $\text{MAX}(\text{DataInizio}) = 31/12/2021$

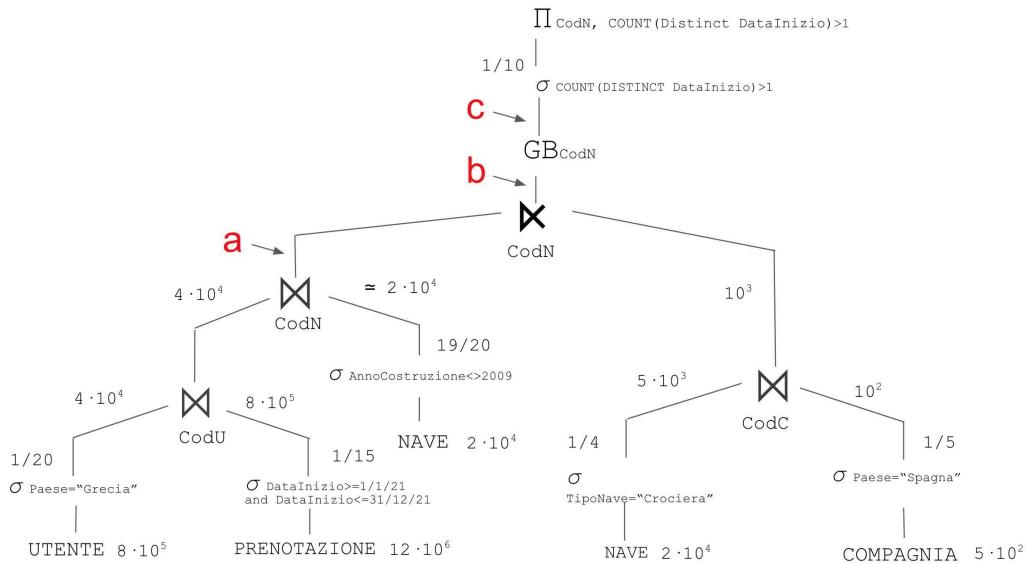
Inoltre, sono dati i seguenti fattori di riduzione per le clausole having:

- Having COUNT(Distinct DataInizio)>1 $\approx 1/10$

Si consideri la seguente query:

```
SELECT N1.CodN, count(Distinct DataInizio)
FROM NAVE N1, PRENOTAZIONE P, UTENTE U
WHERE N1.CodN=P.CodN AND P.CodU=U.CodU
AND U.Paese="Grecia"
AND N1.AnnoCostruzione<>2009
AND P.DataInizio>=1/1/21 and P.DataInizio<=31/12/21
AND N1.CodN IN (
    SELECT N2.CodN
        FROM NAVE N2, COMPAGNIA C
        WHERE N2.CodN=C.CodC
        AND N2.TipoNave="Crociera"
        AND C.Paese="Spagna" )
group by N1.CodN
Having COUNT(Distinct DataInizio)>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) c: $\approx 10^2$
- (b) a: $\approx 4 \cdot 10^3$
- (c) c: $\approx 10^4$
- (d) c: $\approx 10^3$
- (e) b: $\approx 2 \cdot 10^2$
- (f) c: $< 10^5$
- (g) b: $< 2 \cdot 10^5$
- (h) b: $\approx 2 \cdot 10^4$
- (i) a: $\approx 2 \cdot 10^5$
- (j) a: $\approx 4 \cdot 10^5$
- (k) a: $\approx 4 \cdot 10^4$
- (l) b: $\approx 2 \cdot 10^3$

Risposta errata.

La risposta corretta è: a: $\approx 4 \cdot 10^4$, b: $\approx 2 \cdot 10^3$, c: $\approx 10^3$

Domanda 4

Risposta non data

Punteggio max.: 1,50

1.5 punti (penalità 15% per risposta sbagliata)

Data una feature f caratterizzata da n campioni (f_1, f_2, \dots, f_n) , la normalizzazione z-score di f_i è data da $\frac{f_i - \mu}{\sigma}$, dove μ è la media di f e σ la sua deviazione standard.

Viene data una feature f con media μ e deviazione standard σ .

Quale delle seguenti affermazioni sulla normalizzazione z-score è corretta?

- (a) Normalizza tutti i punti nell'intervallo $[0, +\sigma]$
- (b) Normalizza tutti i punti nell'intervallo $[-\sigma, +\sigma]$
- (c) Normalizza tutti i punti nell'intervallo $[-1, +1]$
- (d) Normalizza tutti i punti nell'intervallo $[-\sigma^2, +\sigma^2]$
- (e) Nessuna delle altre risposte è corretta
- (f) Normalizza tutti i punti nell'intervallo $[0, +1]$

Risposta errata.

La risposta corretta è: Nessuna delle altre risposte è corretta

Domanda 5

Risposta non data

Punteggio max.: 1,50

1.5 punti (penalità 15% per risposta sbagliata)

Viene data una collection “employees”, contenente le informazioni dei dipendenti di un’azienda. Per ciascun dipendente, informazioni su età, dipartimento e salario sono note. Il seguente è un esempio di documento estratto dalla collection:

```
{  
    "_id" : ObjectId("62b97ce2850f3cffcbaab699"),  
    "first" : "SUSAN",  
    "last" : "DAVIS",  
    "age" : 26,  
    "compensation" : 65000,  
    "department" : "HR"  
}
```

Si vuole estrarre da questa collection il compenso medio dei dipendenti con età minore di 25 anni, separatamente per ogni dipartimento. Quale delle seguenti query soddisfa la richiesta?

(a)

```
db.employees.aggregate([  
    {  
        $match: {  
            "age": {  
                $gt: 25  
            },  
        }  
    },  
    {  
        $group: {  
            _id: null,  
            avg: {  
                $avg: "$compensation"  
            }  
        }  
    }  
])
```

(b)

```
db.employees.aggregate([
  {
    $group: {
      _id: "$department",
      avg: {
        $avg: "$compensation"
      }
    }
  },
  {
    $match: {
      "age": {
        $gt: 25
      }
    }
  }
])
```

(c)

```
db.employees.aggregate([
  {
    $match: {
      "age": {
        $lt: 25
      }
    }
  },
  {
    $group: {
      _id: null,
      avg: {
        $avg: "$compensation"
      }
    }
  }
])
```

(d)

```
db.employees.aggregate([
  {
    $group: {
      _id: "$department",
      avg: {
        $avg: "$compensation"
      }
    }
  },
  {
    $match: {
      "age": {
        $lt: 25
      },
    }
  }
])
```

(e)

```
db.employees.aggregate([
  {
    $match: {
      "age": {
        $gt: 25
      },
    }
  },
  {
    $group: {
      _id: "$department",
      avg: {
        $avg: "$compensation"
      }
    }
  }
])
```

(f)

```
db.employees.aggregate([
  {
    $group: {
      _id: null,
      avg: {
        $avg: "$compensation"
      }
    },
    {
      $match: {
        "age": {
          $lt: 25
        },
      }
    }
  }
])
```

(g)

```
db.employees.aggregate([
  {
    $group: {
      _id: null,
      avg: {
        $avg: "$compensation"
      }
    },
    {
      $match: {
        "age": {
          $gt: 25
        },
      }
    }
  }
])
```

(h)

```
db.employees.aggregate([
  {
    $match: {
      "age": {
        $lt: 25
      }
    }
  },
  {
    $group: {
      _id: "$department",
      avg: {
        $avg: "$compensation"
      }
    }
  }
])
```

Risposta errata.

La risposta corretta è:

```
db.employees.aggregate([
  {
    $match: {
      "age": {
        $lt: 25
      }
    }
  },
  {
    $group: {
      _id: "$department",
      avg: {
        $avg: "$compensation"
      }
    }
  }
])
```

Domanda 6

Risposta non data

Punteggio max.:

1,50

1 punto (penalità 15% per risposta sbagliata)

- Precisione(C) è la frazione di predizione corrette rispetto a tutte le predizioni fatte per la classe C
- Recall(C) è la frazione di predizioni corrette rispetto a tutti i punti che appartengono alla classe C

Vengono dati due vettori, y_{pred} , y_{true} , che contengono le predizioni effettuate da un classificatore e la ground truth, rispettivamente.

$y_{\text{true}}: [\text{B B A A C A B B C A}]$

$y_{\text{pred}}: [\text{A B B A B A B B C C}]$

Quali sono precisione e richiamo per la classe A?

- (a) Precisione: 0.5, Richiamo: 0.6667
- (b) Precisione: 0.6, Richiamo: 0.75
- (c) Precisione: 0.6667, Richiamo: 0.5
- (d) Precisione: 0.75, Richiamo: 0.5
- (e) Nessuna delle altre risposte è corretta
- (f) Precisione: 0.6, Richiamo: 0.6667
- (g) Precisione: 0.75, Richiamo: 0.6
- (h) Precisione: 0.5, Richiamo: 0.5
- (i) Precisione: 0.6, Richiamo: 0.5

Risposta errata.

La risposta corretta è: Precisione: 0.6667, Richiamo: 0.5

Domanda 7

Risposta non data

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

Sono date le seguenti tabelle:

NAVE(CodN, TipoNave, Capienza, AnnoCostruzione, CodC)
COMPAGNIA(CodC, Nome, Indirizzo, Città, Paese, DataFondazione)
UTENTE(CodU, Nome, Cognome, DataNascita, Indirizzo, Città, Paese)
PRENOTAZIONE(CodN, DataInizio, CodU, DataFine)

Sono date le seguenti cardinalità:

- $\text{card}(\text{NAVE}) = 2 \cdot 10^4$ tuple,
valori distinti di TipoNave = 4
 $\text{MIN}(\text{AnnoCostruzione}) = 1990$, $\text{MAX}(\text{AnnoCostruzione}) = 2009$
- $\text{card}(\text{COMPAGNIA}) = 5 \cdot 10^2$ tuple,
valori distinti di Paese = 5
- $\text{card}(\text{UTENTE}) = 8 \cdot 10^5$ tuple,
valori distinti di Paese = 20
- $\text{card}(\text{PRENOTAZIONE}) = 12 \cdot 10^6$ tuple,
 $\text{MIN}(\text{DataInizio}) = 1/1/2007$, $\text{MAX}(\text{DataInizio}) = 31/12/2021$

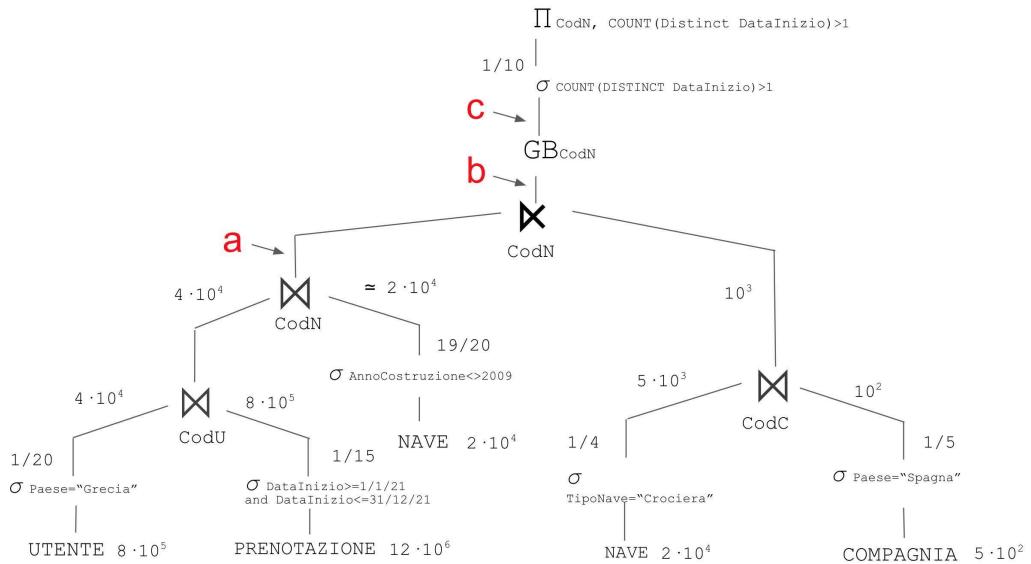
Inoltre, sono dati i seguenti fattori di riduzione per le clausole having:

- Having COUNT(Distinct DataInizio) > 1 $\approx 1/10$

Si consideri la seguente query:

```
SELECT N1.CodN, count(Distinct DataInizio)
FROM NAVE N1, PRENOTAZIONE P, UTENTE U
WHERE N1.CodN=P.CodN AND P.CodU=U.CodU
AND U.Paese="Grecia"
AND N1.AnnoCostruzione<>2009
AND P.DataInizio>=1/1/21 and P.DataInizio<=31/12/21
AND N1.CodN IN (
    SELECT N2.CodN
    FROM NAVE N2, COMPAGNIA C
    WHERE N2.CodN=C.CodC
    AND N2.TipoNave="Crociera"
    AND C.Paese="Spagna" )
group by N1.CodN
Having COUNT(Distinct DataInizio)>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 IndexH ON PRENOTAZIONE(DataInizio) - B+-Tree
- (b) CREATE INDEX IndexF ON Nave(AnnoCostruzione) - B+-Tree
- (c) CREATE INDEX IndexB ON NAVE(TipoNave) - B+-Tree
- (d) CREATE INDEX IndexI ON UTENTE(Paese) - HASH
- (e) CREATE INDEX IndexD ON COMPAGNIA(Paese) - B+-Tree
- (f) CREATE INDEX IndexC ON COMPAGNIA(Paese) - HASH
- (g) Nessuna - le strutture fisiche accessorie non migliorerebbero le prestazioni dell'interrogazione.
- (h) CREATE INDEX IndexA ON NAVE(TipoNave) - HASH
- (i) CREATE INDEX IndexJ ON UTENTE(Paese) - B+-Tree
- (j) CREATE INDEX IndexE ON Nave(AnnoCostruzione) - HASH
- (k) CREATE INDEX IndexG ON PRENOTAZIONE(DataInizio) - HASH

Risposta errata.

La risposta corretta è: CREATE INDEX IndexH ON PRENOTAZIONE(DataInizio) - B+-Tree, CREATE INDEX IndexI ON UTENTE(Paese) - HASH

Domanda 8

Risposta non data

Punteggio max.: 1,00

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

B(T1) B(T2) I2(o1) Commit (T2) CK(T1) B(T3) I3(o2) I1(o1) D3(o2) Commit(T3)

U1(o1) **FAILURE**

Notazione:

- T_n = Id della transazione n
- $B(T_n)$ = Begin di T_n
- CK = checkpoint
- $U_n(o_x)$ = update eseguito da T_n sull'oggetto o_x ; stessa notazione per I (insert) e D (delete)

Quali operazioni vengono eseguite per un **warm restart**?

- (a) UNDO = {}, REDO = {T3}
- (b) UNDO = {T1}, REDO = {T3}
- (c) UNDO = {T3}, REDO = {T1}
- (d) UNDO = {T1, T2}, REDO = {T3}
- (e) UNDO = {T2}, REDO = {T3}
- (f) Nessuna delle altre risposte è corretta
- (g) UNDO = {T1, T2}, REDO = {}

Risposta errata.

La risposta corretta è: UNDO = {T1}, REDO = {T3}

Domanda 9

Risposta non data

Punteggio max.: 10

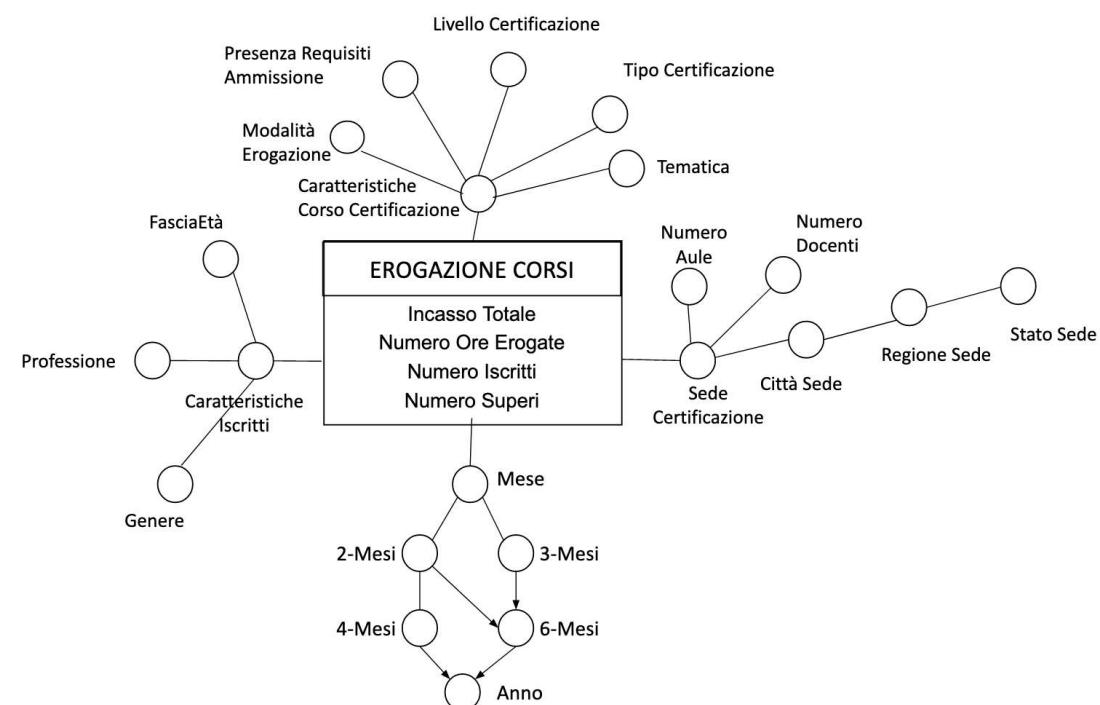
3,00

3 punti (nessuna penalità per risposta sbagliata)

Il seguente data warehouse descrive l'andamento dei corsi di certificazione erogati da una società internazionale di certificazione. I corsi di certificazione sono erogati in sedi di certificazione, che sono sedi della società distribuite in tutto il mondo. Le sedi sono anche caratterizzate dal numero di aule e di docenti disponibili. Le metriche da analizzare sono l'incasso totale dei corsi, il numero di ore erogate, il numero di persone iscritte e il numero di persone che hanno superato l'esame di certificazione alla fine del corso (NumeroSuperi).

Il data warehouse memorizza le caratteristiche principali dei corsi e relative certificazioni in termini di tematica, modalità di erogazione (presenza, online, mista), tipo di certificazione e relativo livello (base, intermedio o avanzato), e presenza o meno di prerequisiti (PresenzaRequisitiAmmissione, attributo booleano). Gli iscritti sono caratterizzati da fascia di età (≤ 20 , > 21 and ≤ 30 , > 31 and ≤ 45 , > 45), professione e genere (M, F).

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



CARATTERISTICHE-CORSO-CERTIFICAZIONE (IDCarattCorsoCert, Tematica, TipoCertificazione, LivelloCertificazione, PresenzaRequisitiAmmissione, ModalitàFrogazione)

CARATTERISTICHE-ISCRITTI (IDCarattIscritti, FasciaEtà, Professione, Genere)

SEDE-CERTIFICAZIONE (IDSedeCertificazione, CittàSede, RegioneSede, StatoSede, NumeroAule, NumeroDocenti)

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

EROGAZIONE-CORSI ([IDCarattCorsoCert](#), [IDCarattIscritti](#), [IDSedeCertificazione](#), [IDTempo](#),
o IncassoTotale NumeroOreFrogate NumeroIscritti NumeroSuperi)

Separatamente per regione della sede di certificazione e trimestre (attributo 3-Mesi), considerando i corsi erogati con modalità mista e tematica “Machine Learning”, visualizzare:

- l'incasso, il numero di iscritti, il numero di superi,
 - la percentuale del numero di iscritti rispetto rispetto al totale annuale per regione della sede di certificazione

- il totale cumulativo dell'incasso al trascorrere dei trimestri, separatamente per anno e regione della sede di certificazione

```
SELECT RegioneSede, 3-Mesi,
SUM(Incasso Totale), SUM(Numerolscritti), SUM(NumeroSperi),
100* Sum(Numberolscritti)/SUM(Sum(Numberolscritti)) OVER (PARTITION BY Anno,
RegioneSede),
SUM(SUM(IncassoTotale)) OVER (PARTITION BY Anno, RegioneSede ORDER BY 3-
Mese ROWS UNBOUNDED PRECEDING)
FROM EROGAZIONE-CORSI EC, TEMPO T, SEDE-CERTIFICAZIONE SC,
CARATTERISTICHE-CORSO-CERTIFICAZIONE CC
WHERE EC.IDTempo=T.IDTempo AND
EC.SedeCertificazione=SC.IDSedeCertificazione AND
CC.IDCarattCorsoCert=EC.IDCarattCorsoCert
AND ModalitàErogazione= 'Mista' AND Tematica = 'Machine Learning'
GROUP BY RegioneSede, 3-Mesi, Anno
```

Domanda 10

Risposta non data

Punteggio max.:

1,00

1 punto (penalità 15% per risposta sbagliata)

Si vogliono analizzare le informazioni relative ad una piattaforma online che offre esercizi di inglese. La piattaforma permette agli utenti svolgere esercizi su diversi argomenti della grammatica inglese e a diversi livelli di difficoltà, e di verificare in modo automatico se la soluzione proposta sia corretta. La piattaforma assegna un punteggio agli esercizi svolti.

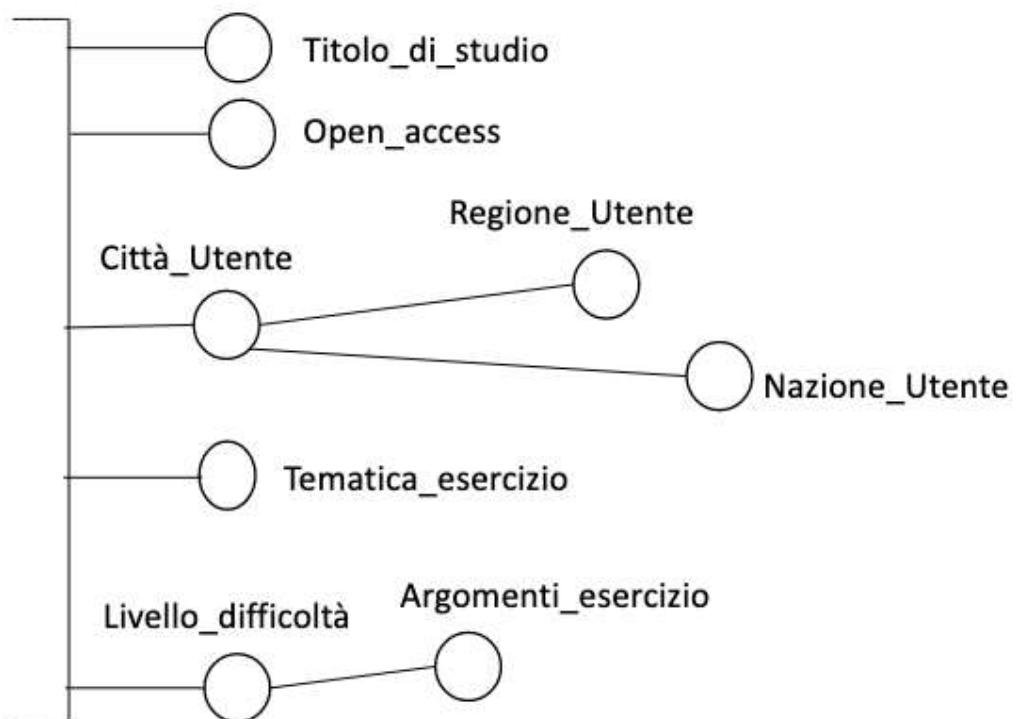
La piattaforma offre un ampio numero di esercizi in modalità open access, insieme ad alcuni esercizi a pagamento. Per guidare l'utente nella selezione degli esercizi di interesse, ciascun esercizio è descritto con l'indicazione degli argomenti di grammatica affrontati nell'esercizio e il livello di difficoltà dell'esercizio (ad esempio se è un esercizio di livello 'elementary' oppure 'advanced'). Per supportare le analisi di interesse sono anche memorizzati l'età dell'utente che sta svolgendo l'esercizio, l'elenco dei certificati di inglese conseguiti e la provenienza geografica dell'utente.

Si vuole analizzare il tempo medio per lo svolgimento di ciascun esercizio e il punteggio medio ottenuto sugli esercizi svolti in base a:

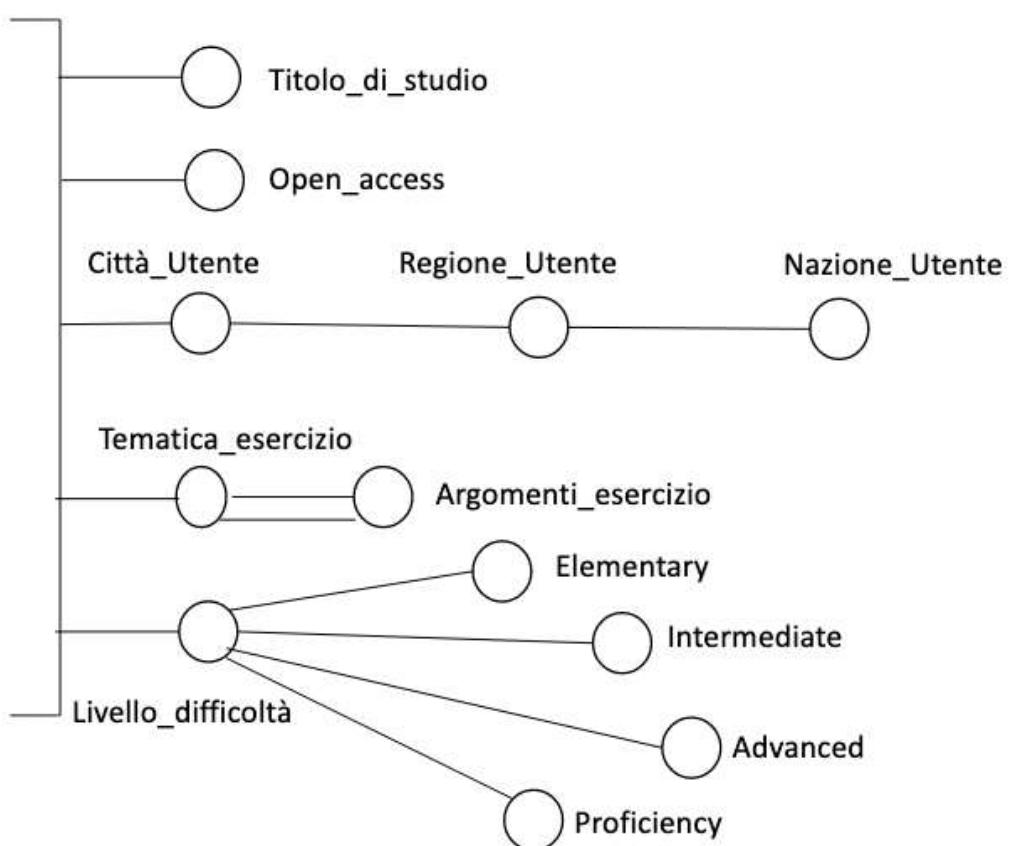
- data, giorno della settimana, mese, bimestre, trimestre, semestre, anno
- fascia oraria (un valore tra 'mattina', 'pomeriggio', 'sera' e 'notte')
- età dell'utente (un valore tra <15, tra 16 e 20, tra 21 e 30, maggiore di 31) e il titolo di studio più alto conseguito (ad esempio Laurea triennale, oppure laurea magistrale o diploma di scuola superiore)
- città, regione e nazione dell'utente che ha svolto esercizio
- livello di complessità dell'esercizio (un valore tra 'elementary', intermediate, 'advanced' e 'proficiency') e l'indicazione se l'esercizio sia disponibile in modalità open-access oppure a pagamento
- tematica dell'esercizio (ad esempio 'verbi', oppure 'frasi interrogative' o 'pronomi') e per ogni tematica l'elenco degli argomenti di grammatica oggetto dell'esercizio (ad esempio 'present tense' e 'passive voice' oppure 'interrogative form' e 'passive voice')
- la lista dei certificati di inglese posseduti dall'utente che sta svolgendo il test (uno o più valori tra 'PET', 'FIRST', 'ADVANCED', 'PROFICIENCY')

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

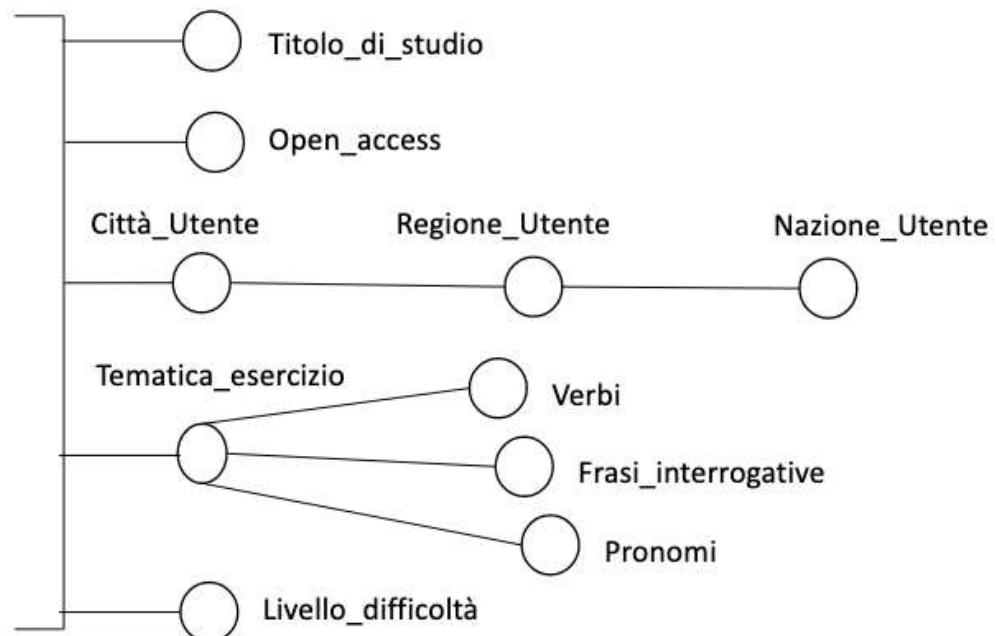
○ (a)



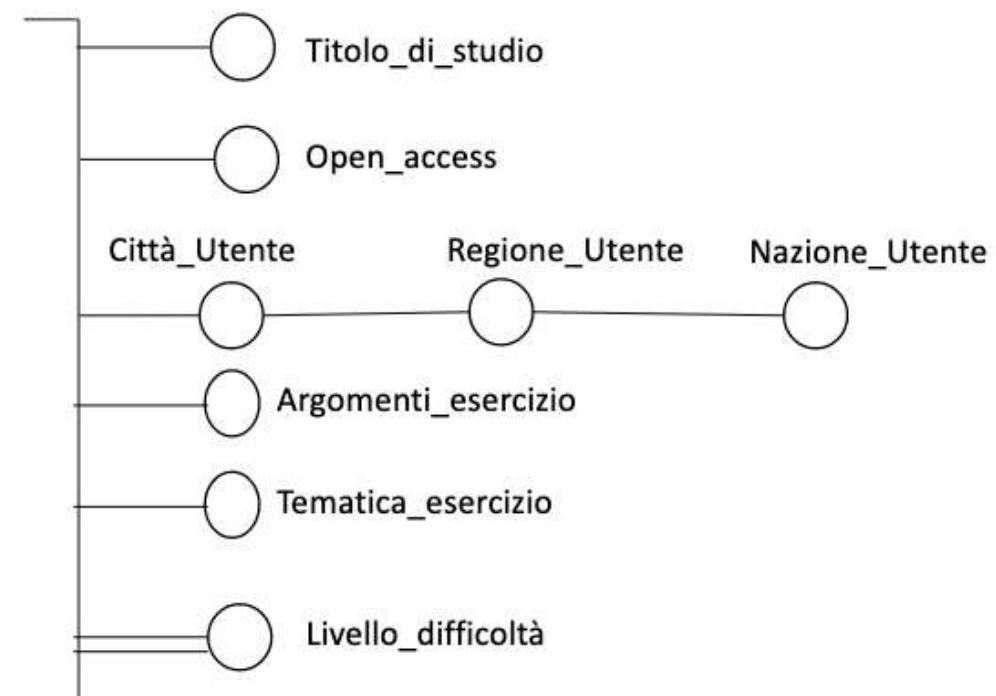
○ (b)



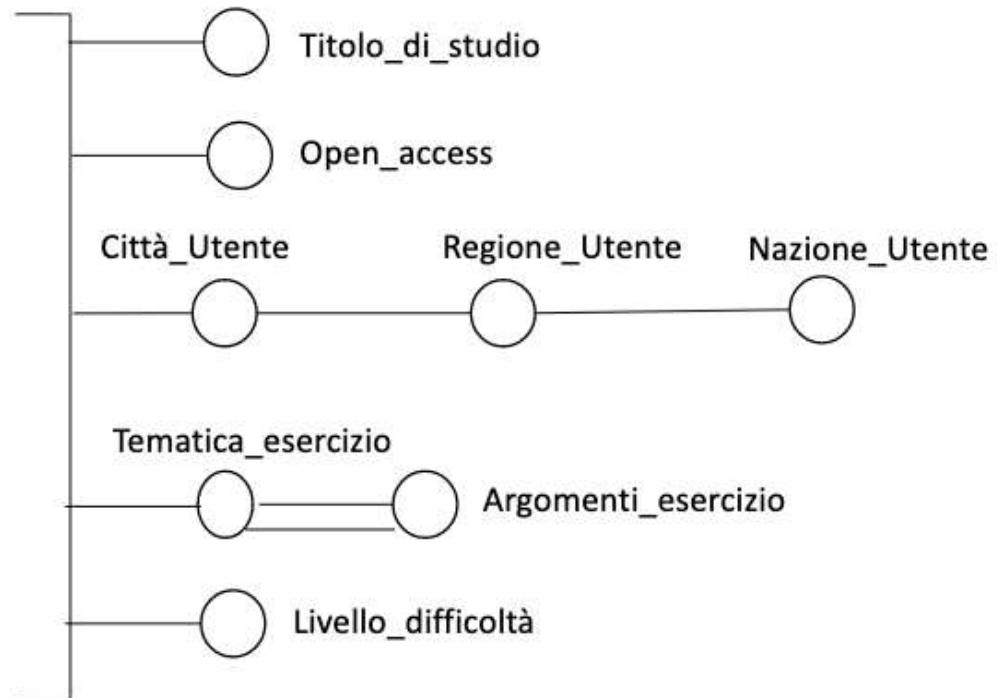
○ (c)



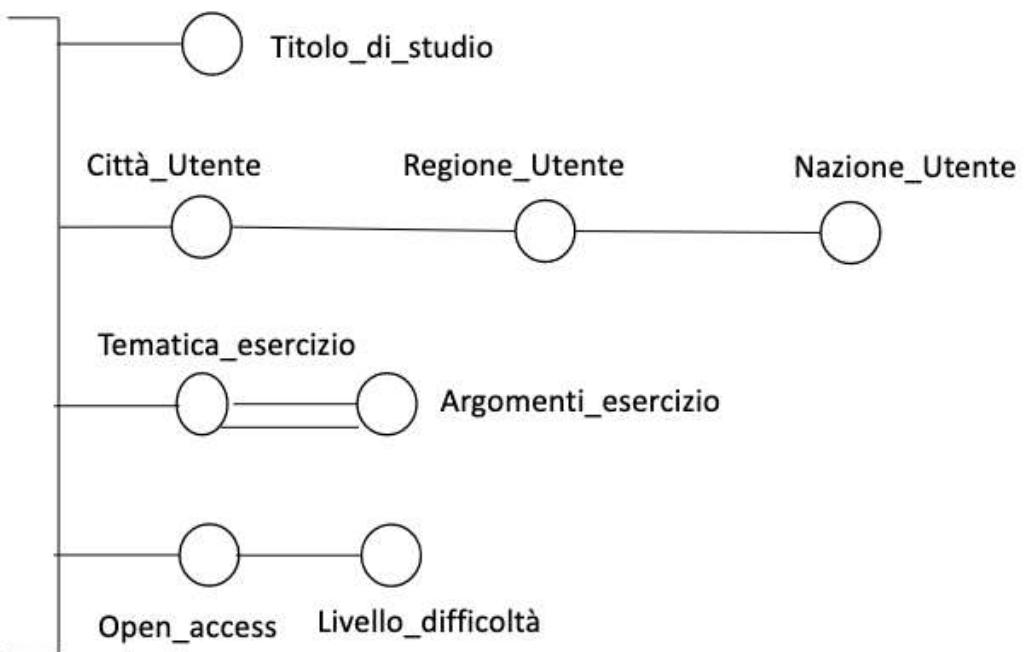
○ (d)



(e)

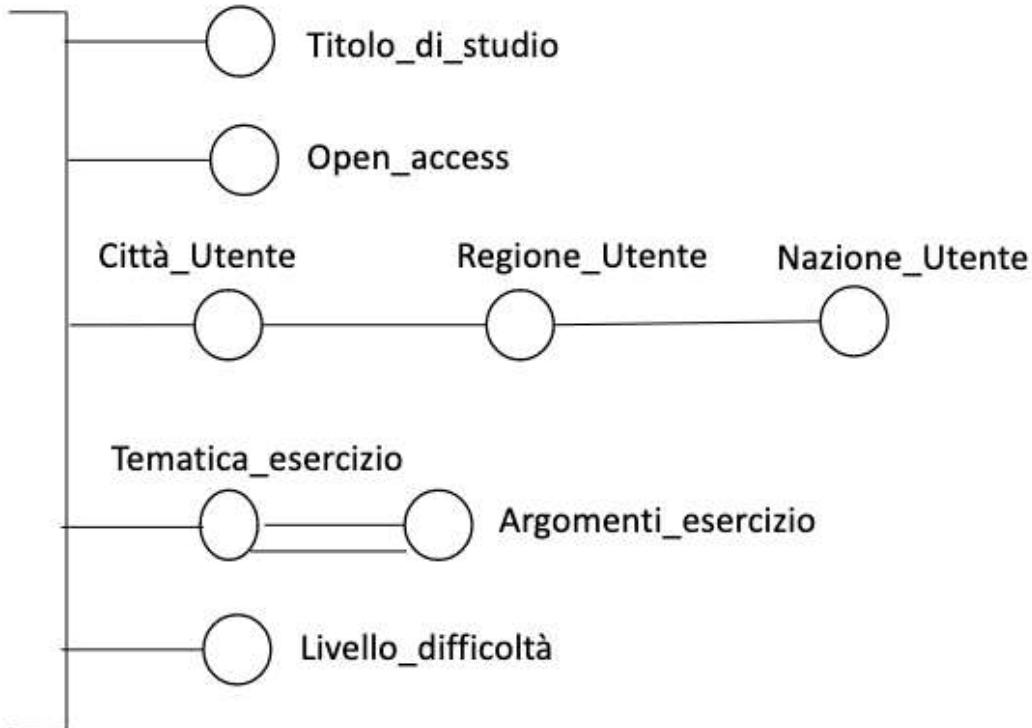


(f)



Risposta errata.

La risposta corretta è:



Domanda 11

Risposta non data

Punteggio max.:
1,50

1.5 punti (penalità 15% per risposta sbagliata)

Viene dato il seguente schedule di cinque transazioni:

S = W2(x) R4(x) R1(x) W2(Y) R5(Y) R3(x) W3(z) R4(z) W2(z) R4(Y) W3(Y)

Quale delle seguenti affermazioni è vera?

-
- (a) S è view serializable perchè è view equivalent allo schedule seriale W3(Y) W3(z) R3(x) W2(x) W2(Y) W2(z) R1(x) R5(Y) R4(x) R4(z) R4(Y)
 - (b) S è view serializable perchè è view equivalent allo schedule seriale W3(Y) W3(z) R3(x) R1(x) R5(Y) R4(x) R4(z) R4(Y) W2(x) W2(Y) W2(z)
 - (c) S è view serializable perchè è view equivalent allo schedule seriale W2(x) W2(Y) R1(x) W3(z) R3(x) R4(x) R4(Y) R4(z) R5(Y) W3(Y) W2(z)
 - (d) S è view serializable perchè è view equivalent allo schedule seriale R1(x) R5(Y) R4(x) R4(z) R4(Y) W3(Y) W3(z) R3(x) W2(x) W2(Y) W2(z)
 - (e) S è view serializable perchè è view equivalent allo schedule seriale W2(x) W2(Y) W2(z) R4(x) R4(z) R4(Y) R1(x) R5(Y) W3(Y) W3(z) R3(x)
 - (f) Nessuna delle altre risposte è corretta

Risposta errata.

La risposta corretta è: Nessuna delle altre risposte è corretta

Domanda 12

Risposta non data

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

L'algoritmo di clustering KMeans viene applicato ad un dataset con 15 attributi e 1000 record. Il numero di cluster che si vuole ottenere è pari a 4 e l'algoritmo esegue 300 iterazioni prima di convergere. Quale delle seguenti affermazioni è corretta:

- (a) l'algoritmo calcola $15 \times 4 \times 1000 \times 300$ distanze tra punti durante la sua intera esecuzione
- (b) l'algoritmo calcola 1000×4 distanze tra punti durante la sua intera esecuzione
- (c) l'algoritmo calcola $1000 \times 4 \times 300$ distanze tra punti durante la sua intera esecuzione
- (d) l'algoritmo calcola $15 \times 1000 \times 300$ distanze tra punti durante la sua intera esecuzione
- (e) durante l'esecuzione dell'algoritmo, vengono calcolate 15×1000 distanze tra punti ad ogni iterazione
- (f) nessuna delle asserzioni è corretta

Risposta errata.

La risposta corretta è: l'algoritmo calcola $1000 \times 4 \times 300$ distanze tra punti durante la sua intera esecuzione

Domanda 13

Risposta non data

Punteggio max.:

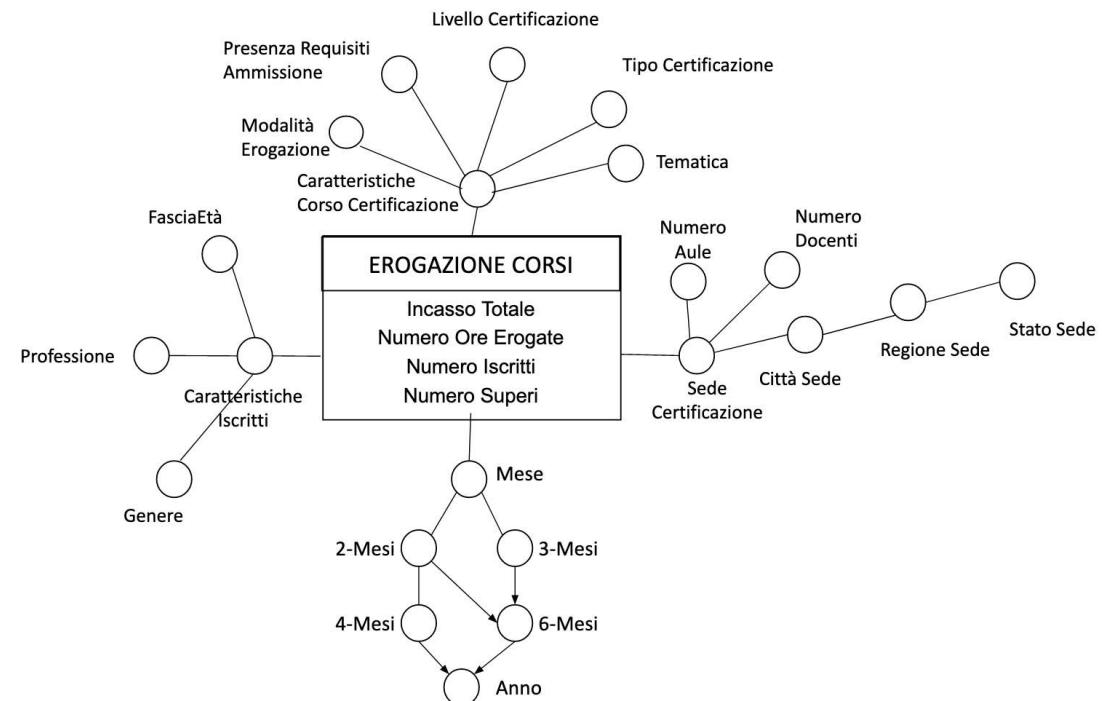
4,00

4 punti (nessuna penalità per risposta sbagliata)

Il seguente data warehouse descrive l'andamento dei corsi di certificazione erogati da una società internazionale di certificazione. I corsi di certificazione sono erogati in sedi di certificazione, che sono sedi della società distribuite in tutto il mondo. Le sedi sono anche caratterizzate dal numero di aule e di docenti disponibili. Le metriche da analizzare sono l'incasso totale dei corsi, il numero di ore erogate, il numero di persone iscritte e il numero di persone che hanno superato l'esame di certificazione alla fine del corso (NumeroSuperi).

Il data warehouse memorizza le caratteristiche principali dei corsi e relative certificazioni in termini di tematica, modalità di erogazione (presenza, online, mista), tipo di certificazione e relativo livello (base, intermedio o avanzato), e presenza o meno di prerequisiti (PresenzaRequisitiAmmissione, attributo booleano). Gli iscritti sono caratterizzati da fascia di età (<=20, >21 and <=30, >31 and <=45, >45), professione e genere (M, F).

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



CARATTERISTICHE-CORSO-CERTIFICAZIONE (IDCarattCorsoCert, Tematica, TipoCertificazione, LivelloCertificazione, PresenzaRequisitiAmmissione, ModalitàErogazione)

CARATTERISTICHE-ISCRITTI (IDCarattIscritti, FasciaEtà, Professione, Genere)

SEDE-CERTIFICAZIONE (IDSedeCertificazione, CittàSede, RegioneSede, StatoSede, NumeroAule, NumeroDocenti)

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

EROGAZIONE-CORSI (IDCarattCorsoCert, IDCarattIscritti, IDSedeCertificazione, IDTempo, IncassoTotale, NumeroOreErogate, NumeroIscritti, NumeroSuperi)

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

- Visualizzare incasso totale e numero totale di ore erogate per ogni tipo di certificazione e anno.
- Per le certificazioni di livello 'base', visualizzare l'incasso medio per iscritto, separatamente per ogni semestre (attributo 6-Mesi) e professione dell'iscritto.

- c. Considerando solo gli scritti di genere femminile e il livello di certificazione ‘avanzato’, visualizzare la differenza tra il numero totale di iscritti e il numero totale di superi (attributo NumeroSuperi), separatamente per trimestre (attributo 3-Mesi) e fascia d’età.

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 ViewCorsi  
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 ViewCorsi in caso di inserimento di un nuovo record nella tabella dei fatti EROGAZIONE-CORSI.

Query di interesse:

(1) Select TipoCertificazione, Anno, SUM(IncassoTotale), SUM(NumerOreErogate)
FROM CARATTERISTICHE-CORSO-CERTIFICAZIONE, EROGAZIONE-CORSI,
TEMPO
WHERE join
Group by TipoCertificazione, Anno

(2)
Select 6-Mesi, Professione, SUM(IncassoTotale)/SUM(Numerolscritti)
FROM CARATTERISTICHE-CORSO-CERTIFICAZIONE, CARATTERISTICHE-
ISCRITTI, EROGAZIONE-CORSI, TEMPO

WHERE join AND LivelloCertificazione = 'base'
Group by 6-Mesi, Professione

(3)

Select 3-Mesi, FasciaEtà, SUM(CostoTotale)
FROM CARATTERISTICHE-CORSO-CERTIFICAZIONE, CARATTERISTICHE-
ISCRITTI, EROGAZIONE-CORSI, TEMPO
WHERE join AND Genere = 'F' AND LivelloCertificazione = 'avanzato'
Group by 3-Mesi, FasciaEtà

====

Soluzione

Materialized view

```
SELECT TipoCertificazione, LivelloCertificazione, Genere, Professione, FasciaEtà, 3-  
Mesi, 6-Mesi, Anno, SUM(IncassoTotale) AS IncassoTot, SUM(NumerOreErogate) AS  
OreTot, SUM(Numeroliscritti) AS TotIscritti, SUM(NumeroSUPERI) AS TotSUPERI  
FROM CARATTERISTICHE-CORSO-CERTIFICAZIONE TC, CARATTERISTICHE-  
ISCRITTI CI, TEMPO T, EROGAZIONE CORSI EA  
WHERE TC.IDCarattCorsoCert = EA.IDCarattCorsoCert AND CI.IDCarattIscritti =  
EA.IDCarattIscritti AND T.IDTempo = EA.IDTempo  
GROUP BY TipoCertificazione, LivelloCertificazione, Genere, Professione, FasciaEtà,  
3-Mesi, 6-Mesi, Anno
```

====

Trigger

```
CREATE OR REPLACE TRIGGER CertificazioniViewMaintenance  
AFTER INSERT ON EROGAZIONE-CORSI  
FOR EACH ROW  
DECLARE  
  
VarAnno DATE, VarSemestre DATE, VarTrimestre DATE;  
VarTipoCert, VarLivelloCert varchar(10);  
VarGenere, VarProfessione, VarFasciaEtà varchar(10);  
N INTEGER;  
BEGIN  
    SELECT 3-Mesi, 6-Mesi, Anno INTO VarTrimestre, VarSemeste, varAnno  
    FROM TEMPO  
    WHERE IDTempo = :NEW.IDTempo;  
  
    SELECT Genere, Professione, FasciaEtà INTO VarGenere, VarProfessione,  
    VarFasciaEtà  
    FROM CARATTERISTICHE-ISCRITTI  
    WHERE IDCarattIscritti = :NEW.IDCarattIscritti;  
  
    SELECT TipoCertificazione, LivelloCertificazione INTO VarTipoCert,  
    VarLivelloCert  
    FROM CARATTERISTICHE-CORSO-CERTIFICAZIONE  
    WHERE IDCarattCorsoCert = :NEW.IDCarattCorsoCert;  
  
    SELECT COUNT(*) INTO N
```

```
FROM ViewCertificazioni
WHERE 3-Mesi = VarTrimestre AND Genere = VarGenere AND Professione =
VarProfessione AND FasciaEtà = VarFasciaEtà AND
    TipoCertificazione= varTipoCert AND LivelloCertificazione=VarLivelloCert;

IF N>0 THEN
    UPDATE ViewCertificazioni
    SET IncassoTot = IncassoTot + :NEW.IncassoTotale, OreTot = OreTot +
:NEW.NumeroOreErogate,
TotIscritti = TotIscritti + :NEW.NumerIscritti, TotSuperi = TotSuperi +
:NEW.NumeroSuperi
    WHERE 3-Mesi = VarTrimestre AND Genere = VarGenere AND Professione =
VarProfessione AND FasciaEtà = VarFasciaEtà AND
        TipoCertificazione= varTipoCert AND
LivelloCertificazione=VarLivelloCert;
ELSE
    INSERT INTO ViewCertificazioni(...) VALUES (VarTrimestre, VarSemeste, varAnno,
VarGenere, VarProfessione, VarFasciaEtà, VarTipoCert, VarLivelloCert,
:NEW.IncassoTotale, :NEW.NumeroOreErogate, :NEW.NumerIscritti,
:NEW.NumeroSuperi);
END
```

Domanda 14

Risposta non data

Punteggio max.:

2,00

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

Sono date le seguenti tabelle:

NAVE(CodN, TipoNave, Capienza, AnnoCostruzione, CodC)
COMPAGNIA(CodC, Nome, Indirizzo, Città, Paese, DataFondazione)
UTENTE(CodU, Nome, Cognome, DataNascita, Indirizzo, Città, Paese)
PRENOTAZIONE(CodN, DataInizio, CodU, DataFine)

Sono date le seguenti cardinalità:

- $\text{card}(\text{NAVE}) = 2 \cdot 10^4$ tuple,
valori distinti di TipoNave = 4
 $\text{MIN}(\text{AnnoCostruzione}) = 1990$, $\text{MAX}(\text{AnnoCostruzione}) = 2009$
- $\text{card}(\text{COMPAGNIA}) = 5 \cdot 10^2$ tuple,
valori distinti di Paese = 5
- $\text{card}(\text{UTENTE}) = 8 \cdot 10^5$ tuple,
valori distinti di Paese = 20
- $\text{card}(\text{PRENOTAZIONE}) = 12 \cdot 10^6$ tuple,
 $\text{MIN}(\text{DataInizio}) = 1/1/2007$, $\text{MAX}(\text{DataInizio}) = 31/12/2021$

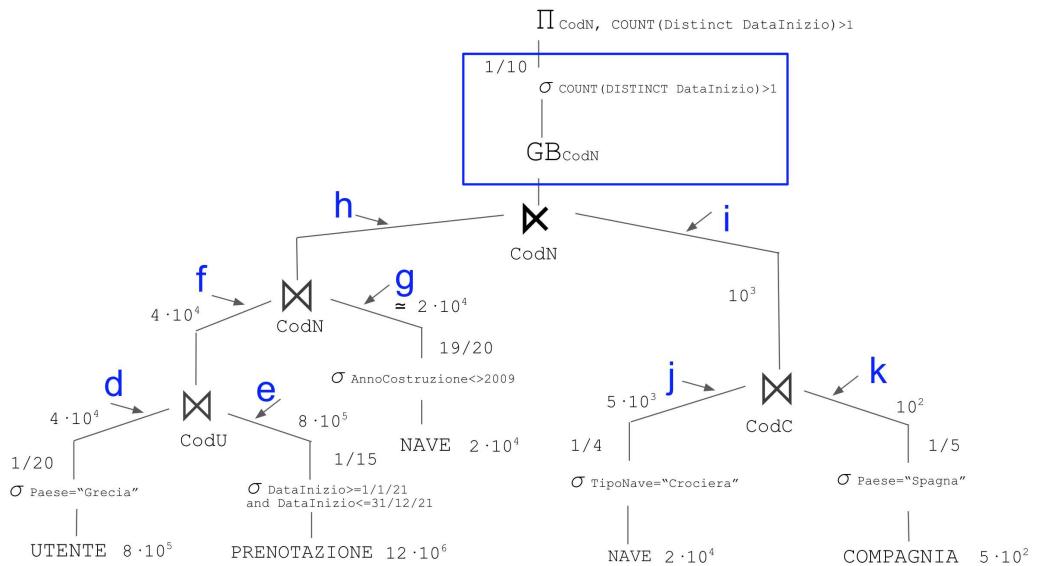
Inoltre, sono dati i seguenti fattori di riduzione per le clausole having:

- Having COUNT(Distinct DataInizio) > 1 $\approx 1/10$

Si consideri la seguente query:

```
SELECT N1.CodN, count(Distinct DataInizio)
FROM NAVE N1, PRENOTAZIONE P, UTENTE U
WHERE N1.CodN=P.CodN AND P.CodU=U.CodU
AND U.Paese="Grecia"
AND N1.AnnoCostruzione<>2009
AND P.DataInizio>=1/1/21 and P.DataInizio<=31/12/21
AND N1.CodN IN (
    SELECT N2.CodN
    FROM NAVE N2, COMPAGNIA C
    WHERE N2.CodN=C.CodC
    AND N2.TipoNave="Crociera"
    AND C.Paese="Spagna" )
group by N1.CodN
Having COUNT(Distinct DataInizio)>1
```

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



Analizzare l'anticipazione della GROUP BY group by N1.CodN having COUNT(DISTINCT DataInizio)>1 rappresentata nel riquadro. Selezionare la soluzione che consente la **massima riduzione di cardinalità** (se esiste).

- (a) Non è possibile anticipare la Group By group by N1.CodN having COUNT(DISTINCT DataInizio)>1
- (b) E' possibile anticiparla nel ramo e
- (c) E' possibile anticiparla nel ramo k
- (d) E' possibile anticiparla nel ramo g
- (e) E' possibile anticiparla nel ramo d
- (f) E' possibile anticiparla nel ramo h
- (g) E' possibile anticiparla nel ramo i
- (h) E' possibile anticiparla nel ramo f
- (i) E' possibile anticiparla nel ramo j

Risposta errata.

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

Domanda 15

Risposta non data

Punteggio max.:

1,00

1 punto (penalità 15% per risposta sbagliata)

Si vogliono analizzare le informazioni relative ad una piattaforma online che offre esercizi di inglese. La piattaforma permette agli utenti svolgere esercizi su diversi argomenti della grammatica inglese e a diversi livelli di difficoltà, e di verificare in modo automatico se la soluzione proposta sia corretta. La piattaforma assegna un punteggio agli esercizi svolti.

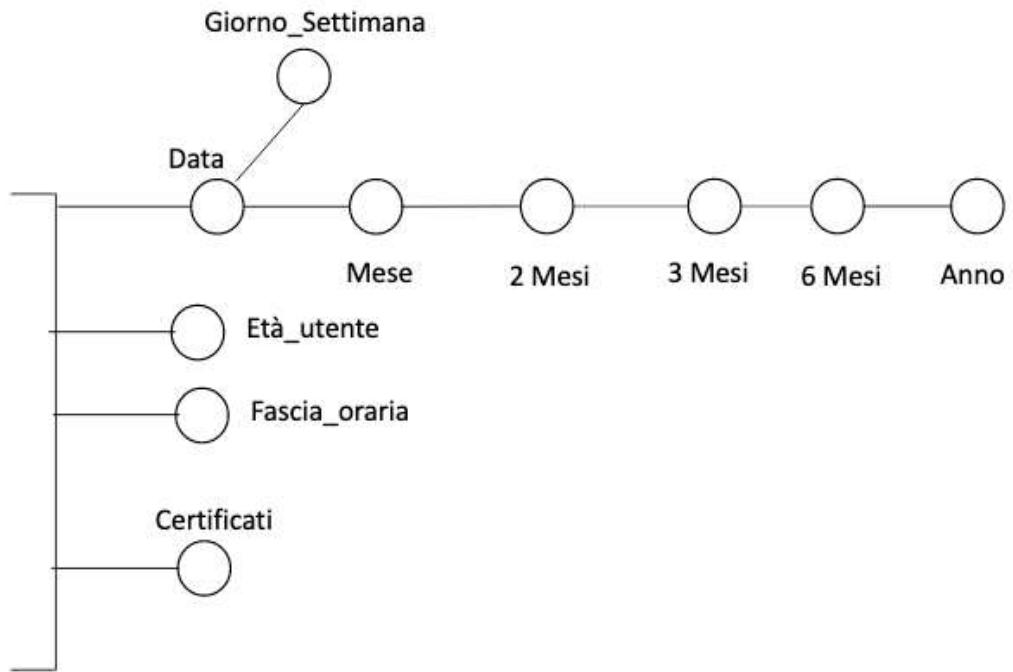
La piattaforma offre un ampio numero di esercizi in modalità open access, insieme ad alcuni esercizi a pagamento. Per guidare l'utente nella selezione degli esercizi di interesse, ciascun esercizio è descritto con l'indicazione degli argomenti di grammatica affrontati nell'esercizio e il livello di difficoltà dell'esercizio (ad esempio se è un esercizio di livello 'elementary' oppure 'advanced'). Per supportare le analisi di interesse sono anche memorizzati l'età dell'utente che sta svolgendo l'esercizio, l'elenco dei certificati di inglese conseguiti e la provenienza geografica dell'utente.

Si vuole analizzare il tempo medio per lo svolgimento di ciascun esercizio e il punteggio medio ottenuto sugli esercizi svolti in base a:

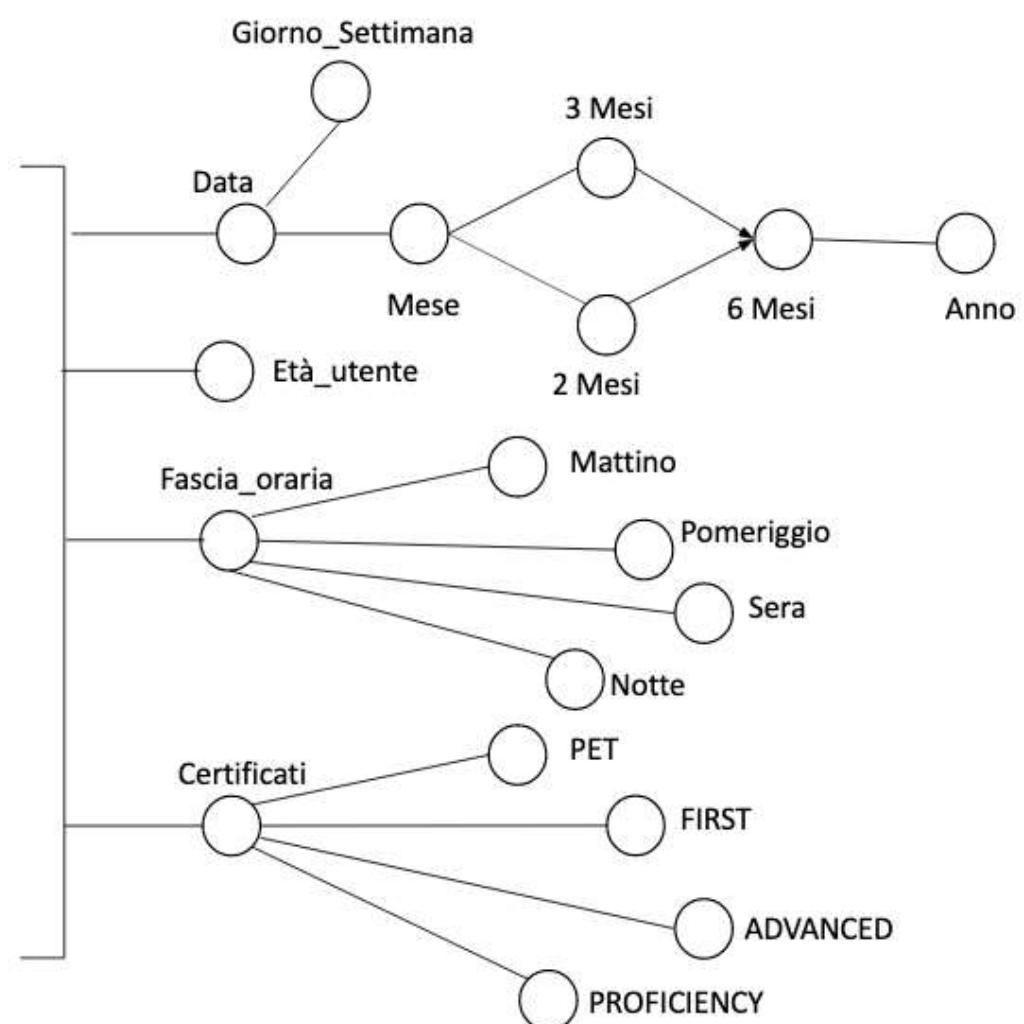
- data, giorno della settimana, mese, bimestre, trimestre, semestre, anno
- fascia oraria (un valore tra 'mattina', 'pomeriggio', 'sera' e 'notte')
- età dell'utente (un valore tra <15, tra 16 e 20, tra 21 e 30, maggiore di 31) e il titolo di studio più alto conseguito (ad esempio Laurea triennale, oppure laurea magistrale o diploma di scuola superiore)
- città, regione e nazione dell'utente che ha svolto esercizio
- livello di complessità dell'esercizio (un valore tra 'elementary', intermediate, 'advanced' e 'proficiency') e l'indicazione se l'esercizio sia disponibile in modalità open-access oppure a pagamento
- tematica dell'esercizio (ad esempio 'verbi', oppure 'frasi interrogative' o 'pronomi') e per ogni tematica l'elenco degli argomenti di grammatica oggetto dell'esercizio (ad esempio 'present tense' e 'passive voice' oppure 'interrogative form' e 'passive voice')
- la lista dei certificati di inglese posseduti dall'utente che sta svolgendo il test (uno o più valori tra 'PET', 'FIRST', 'ADVANCED', 'PROFICIENCY')

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

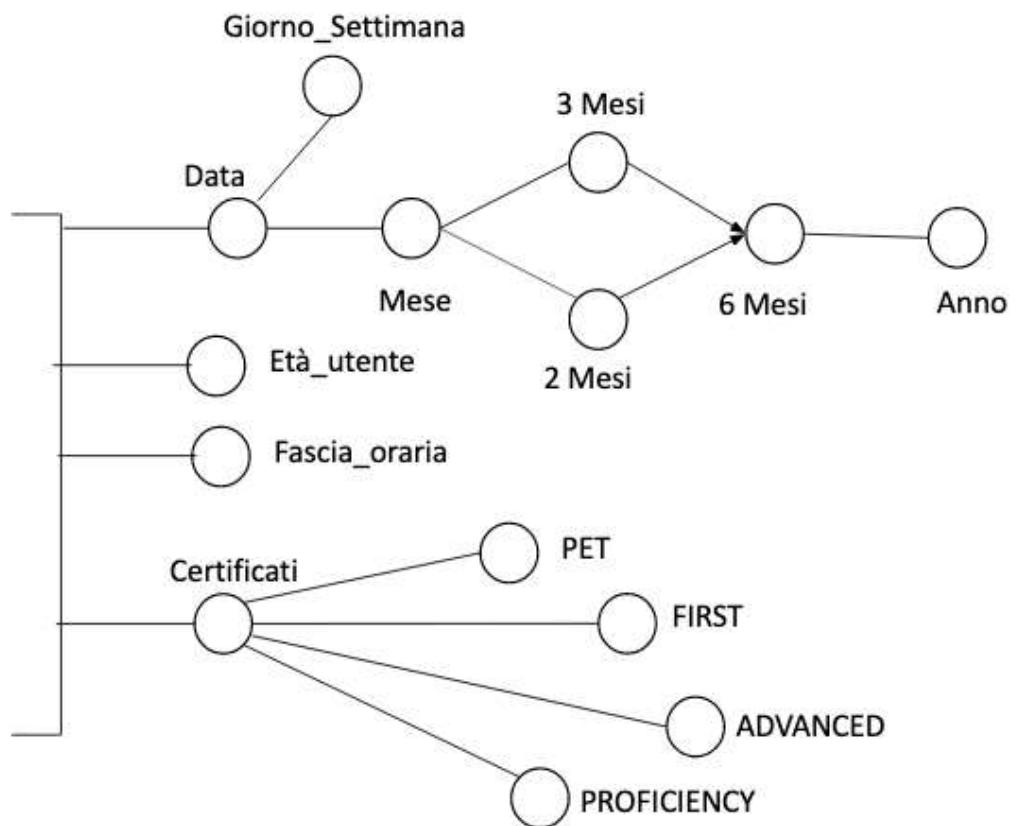
(a)



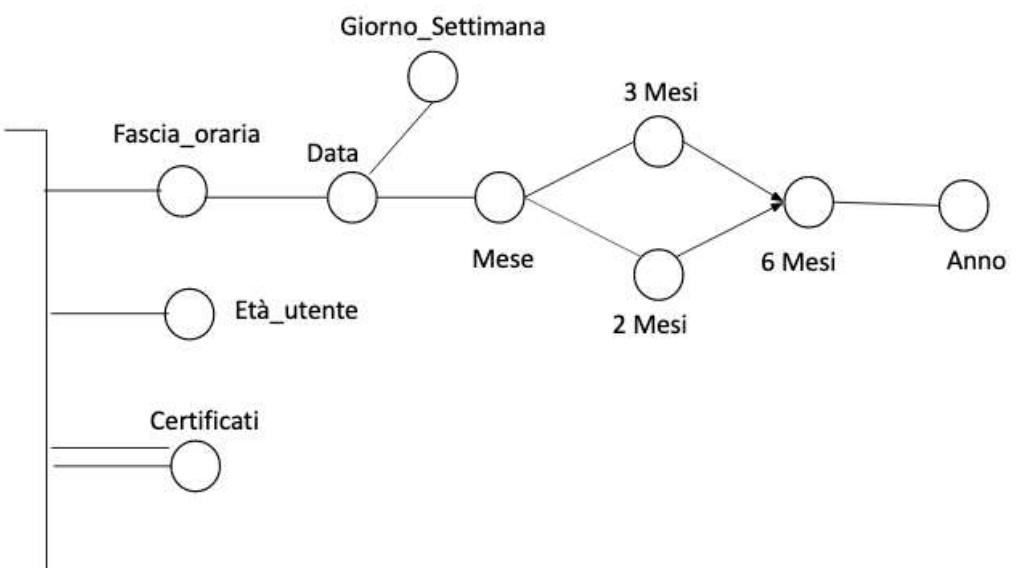
(b)



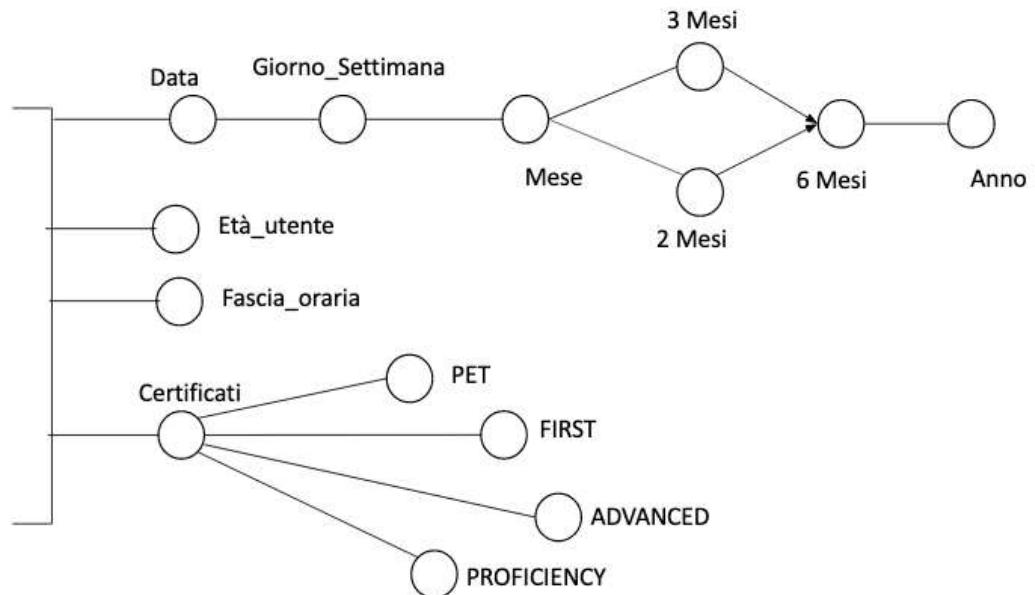
○ (c)



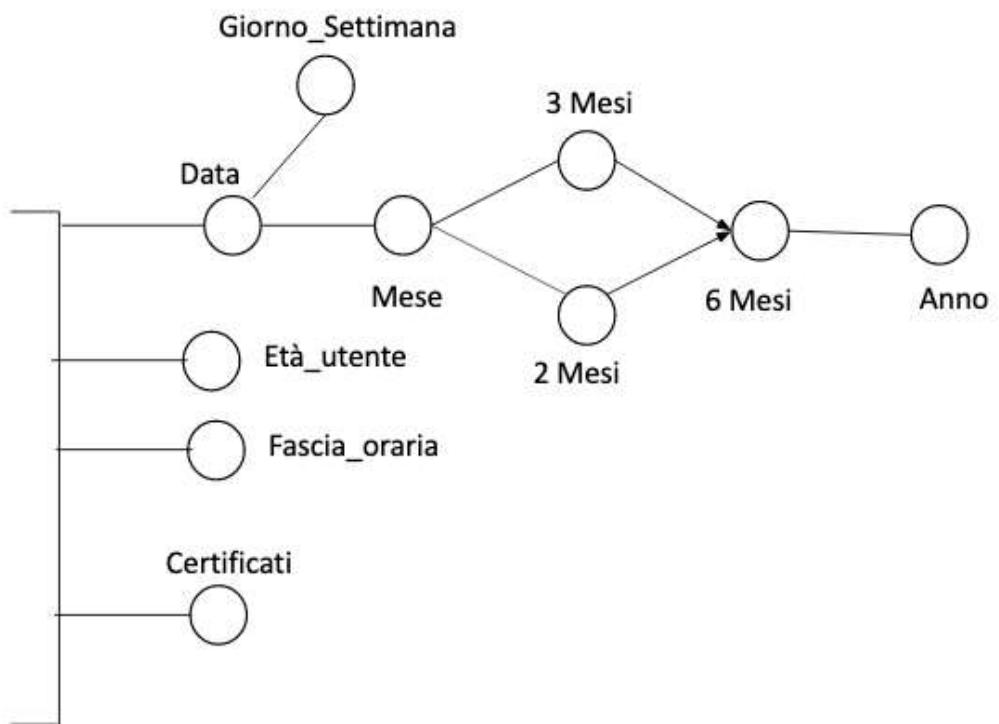
○ (d)



○ (e)

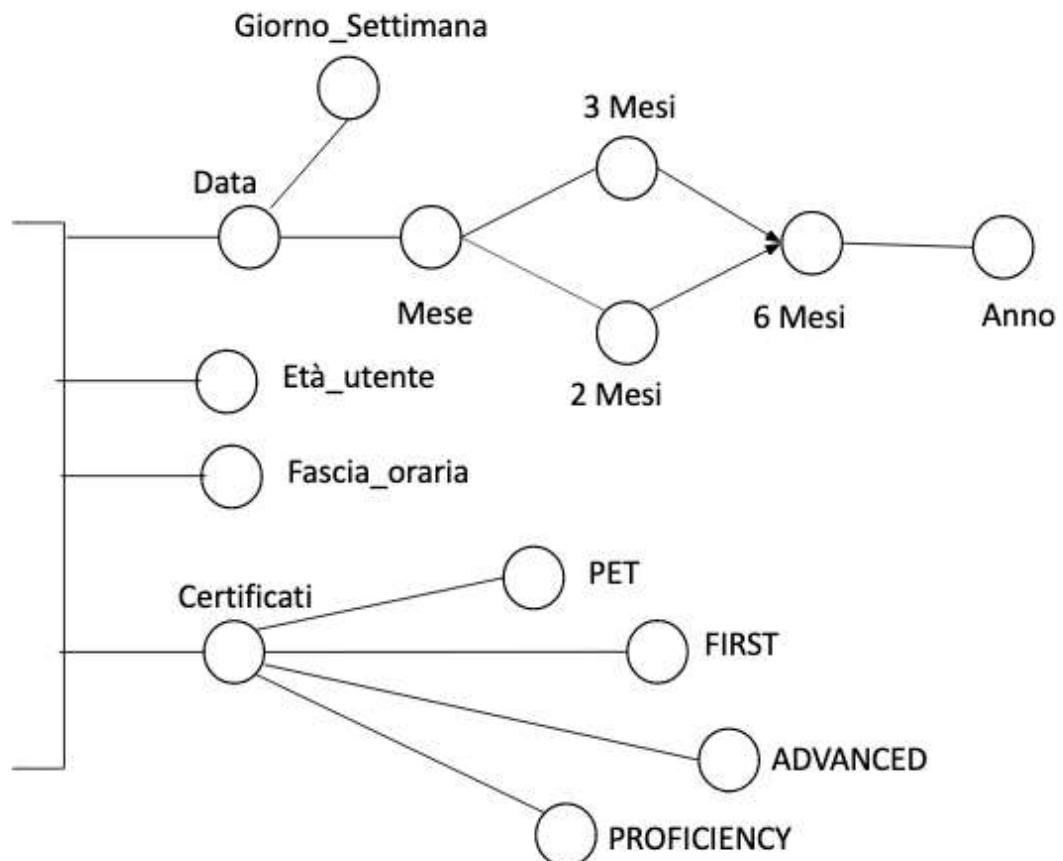


○ (f)



Risposta errata.

La risposta corretta è:



Domanda 16

Risposta non data

Punteggio max.:

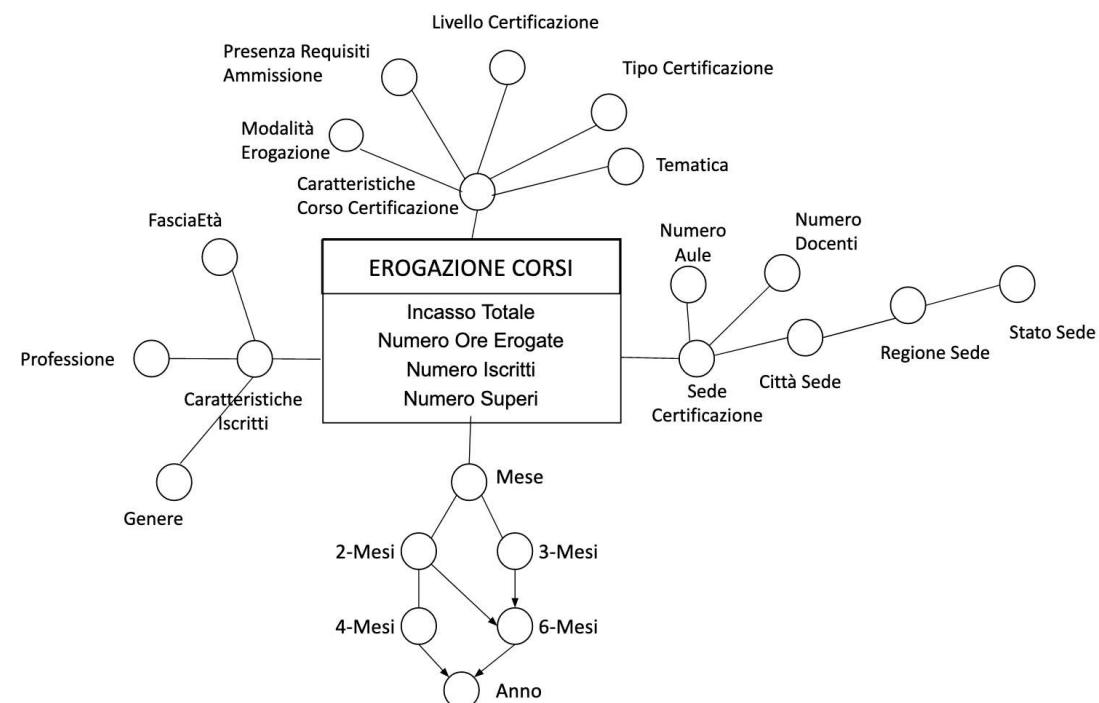
4,00

4 punti (nessuna penalità per risposta sbagliata)

Il seguente data warehouse descrive l'andamento dei corsi di certificazione erogati da una società internazionale di certificazione. I corsi di certificazione sono erogati in sedi di certificazione, che sono sedi della società distribuite in tutto il mondo. Le sedi sono anche caratterizzate dal numero di aule e di docenti disponibili. Le metriche da analizzare sono l'incasso totale dei corsi, il numero di ore erogate, il numero di persone iscritte e il numero di persone che hanno superato l'esame di certificazione alla fine del corso (NumeroSuperi).

Il data warehouse memorizza le caratteristiche principali dei corsi e relative certificazioni in termini di tematica, modalità di erogazione (presenza, online, mista), tipo di certificazione e relativo livello (base, intermedio o avanzato), e presenza o meno di prerequisiti (PresenzaRequisitiAmmisione, attributo booleano). Gli iscritti sono caratterizzati da fascia di età (≤ 20 , > 21 and ≤ 30 , > 31 and ≤ 45 , > 45), professione e genere (M, F).

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



CARATTERISTICHE-CORSO-CERTIFICAZIONE (IDCarattCorsoCert, Tematica, TipoCertificazione, LivelloCertificazione, PresenzaRequisitiAmmissione, ModalitàErogazione)

CARATTERISTICHE-ISCRITTI (IDCarattIscritti, FasciaEtà, Professione, Genere)

SEDE-CERTIFICAZIONE (IDSedeCertificazione, CittàSede, RegioneSede, StatoSede, NumeroAule, NumeroDocenti)

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

EROGAZIONE-CORSI (IDCarattCorsoCert, IDCarattIscritti, IDSedeCertificazione, IDTempo, IncassoTotale, NumeroOreErogate, NumeroIscritti, NumeroSuperi)

Considerando i corsi relativi alla tematica "Lingua Inglese" per il rilascio delle certificazioni di livello avanzato, separatamente per professione degli iscritti e quadri mestri (attributo 4-Mesi) visualizzare

- Il rapporto tra il numero di superi e il numero degli iscritti
- Il numero medio di iscritti per mese

- La percentuale del numero di superi rispetto al numero complessivo di superi per anno e professione degli iscritti
- la posizione in una graduatoria (rank) in ordine decrescente rispetto al numero di superi, separatamente per anno.

```
SELECT Professione, 4-Mesi  
SUM(NumeroSUPERI)/SUM(NUmeroIscritti),  
SUM(NUmeroSUPERI)/ COUNT (DISTINCT Mese)  
100* SUM(NUmeroSUPERI)/SUM(SUM(NmeroSUPERI)) OVER (PARTITION BY Anno,  
Professione),  
RANK() OVER (PARTITION BY Anno ORDER BY SUM(NumeroSUPERI) DESC)  
FROM EROGAZIONE-CORSI CD, TEMPO T, CARATTERISTICHE-ISCRITTI E,  
CARATTERISTICHE-CORSO-CERTIFICAZIONE CD  
WHERE CD.IDTempo=T.IDTempo AND E.IDCarattIscritti=CD.IDCarattIscritti AND  
F.IDCarattCorsoCert=CD.IDCarattCorsoCert  
AND Tematica = "Inglese" AND LivelloCertificazione= "Avanzato"  
GROUP BY Professione, 4-Mesi, Anno
```