



Data Science e Tecnologie per le Basi di Dati

Esame del 16 Giugno 2023

Domanda 1

Risposta corretta

Punteggio ottenuto 1,00 su 1,00

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

Un istituto di analisi di mercato vuole analizzare le informazioni raccolte sulle vendite di prodotti dolciari prodotti da una multinazionale negli ultimi anni.

I prodotti sono realizzati in stabilimenti dislocati in diverse aree geografiche. La multinazionale produce prodotti dolciari di diverso tipo (ad esempio merendine, torte, biscotti), e in modo da venire incontro alle esigenze di diversi regimi alimentari (ad esempio vegetariano o vegano).

La multinazionale effettua la vendita dei prodotti attraverso diversi punti vendita al dettaglio distribuiti in varie aree geografiche.

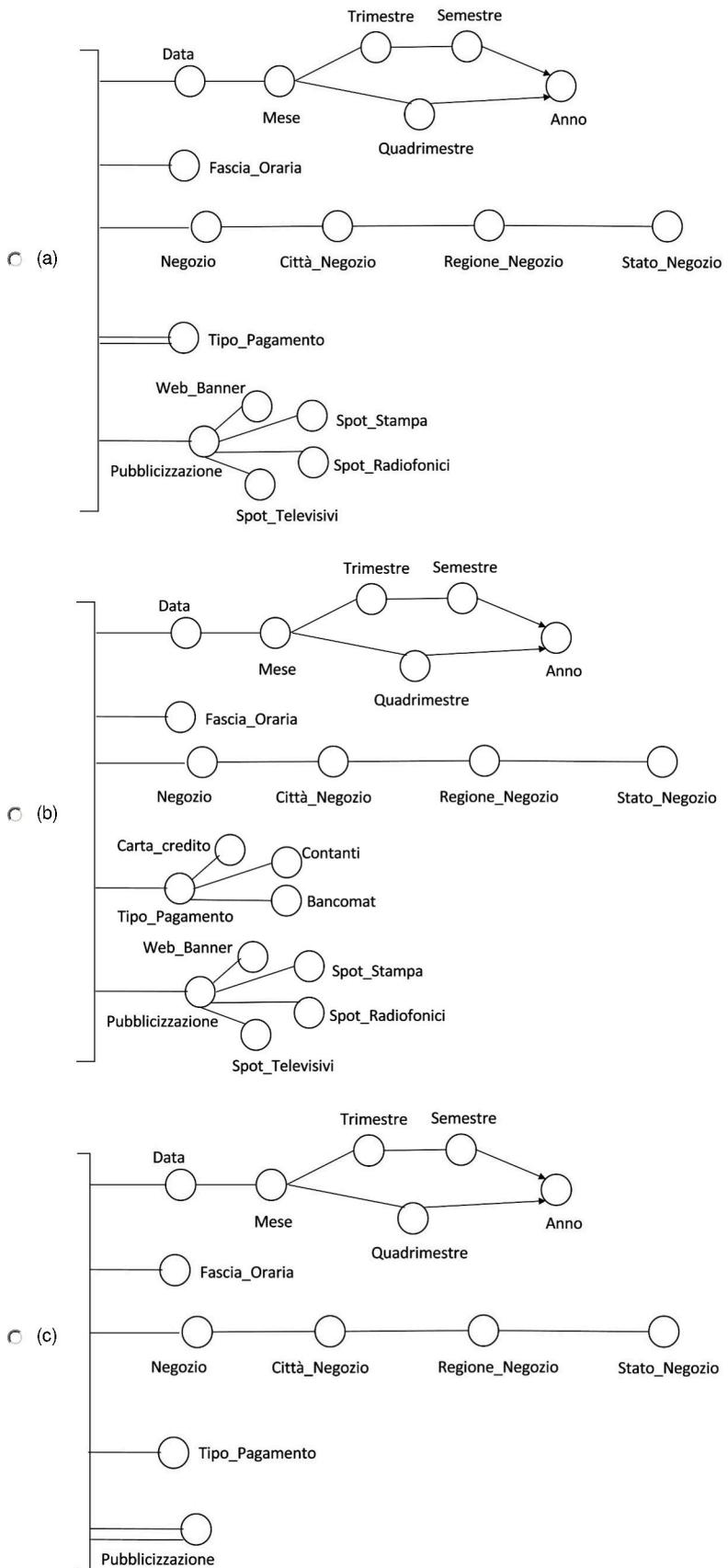
Per analizzare l'andamento delle vendite dei diversi prodotti rispetto a diversi segmenti di mercato, la multinazionale tiene traccia, attraverso le carte fedeltà, di alcune informazioni sui clienti che hanno acquistato i prodotti, quali il genere, la fascia di età e il tipo di impiego.

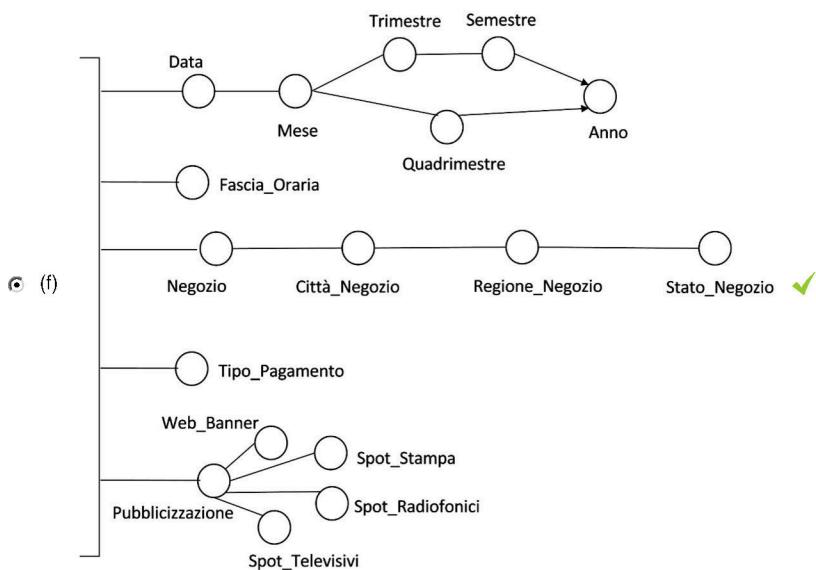
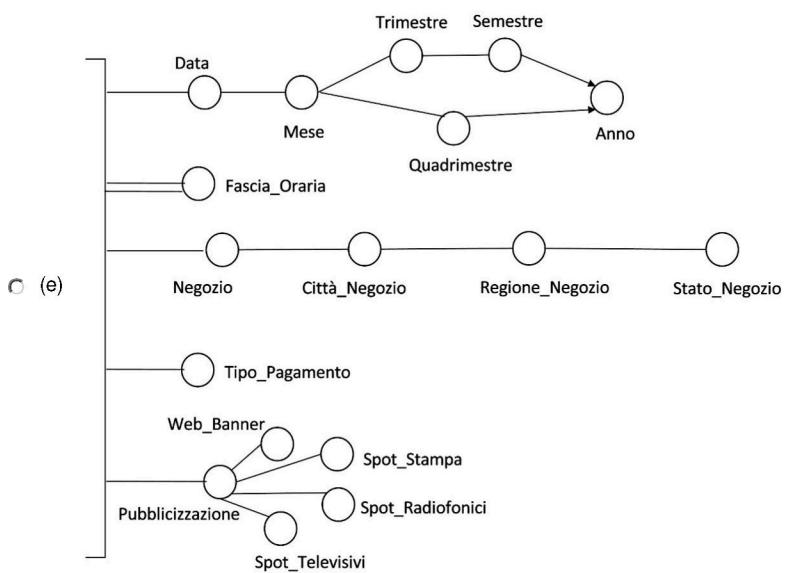
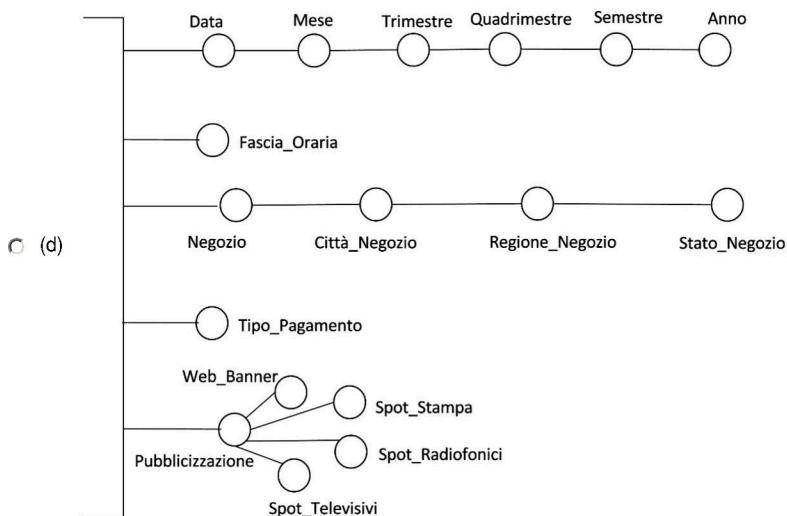
La multinazionale si avvale di diversi canali di comunicazione per pubblicizzare i suoi prodotti mediante spot pubblicitari radiofonici, televisivi o su banner su siti web.

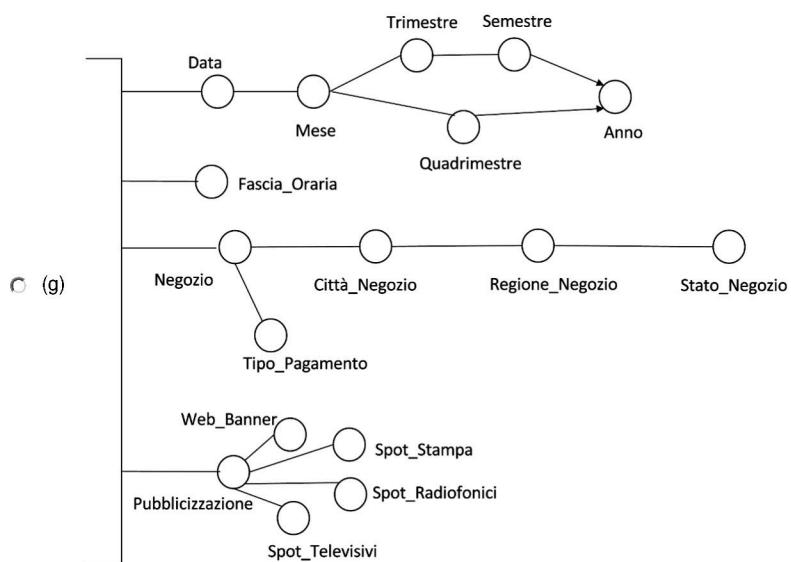
L'istituto di analisi di mercato vuole analizzare il costo medio delle azioni di pubblicizzazione e il profitto di vendita complessivo in base a:

- stabilimento di produzione, caratterizzato dall'informazione sulla città, regione, stato in cui è collocato, e dal responsabile dello stabilimento
- prodotto, caratterizzato dal nome, dal tipo di prodotto (ad esempio torta, merendina, barretta, etc), l'indicazione dei regimi alimentari compatibili (uno o più valori tra vegetariano, vegano, celiachia), la lista degli ingredienti e la lista degli allergenici.
- genere, fascia di età, tipo di impiego del cliente che ha effettuato l'acquisto. La fascia d'età è un valore tra < 20 anni, tra 20 e 29 anni, tra 30 e 49 anni, e maggiore di 50 anni. Il tipo di impiego è un valore tra libero professionista, dipendente, altro.
- negozio in cui il prodotto è stato acquistato, caratterizzato dalla città regione e stato in cui il negozio è collocato.
- il tipo di pagamento effettuato (un valore tra contanti, carta di credito e bancomat)
- giorno, mese, trimestre, quadrimestre semestre, anno in cui si è verificato l'acquisto e fascia oraria (un valore tra mattino, pomeriggio, sera)
- canali di comunicazione usati per la pubblicizzazione del prodotto (uno o più valori tra i seguenti spot pubblicitari radiofonici, spot pubblicitari a mezzo stampa, spot pubblicitari televisivi, banner pubblicitari web)

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

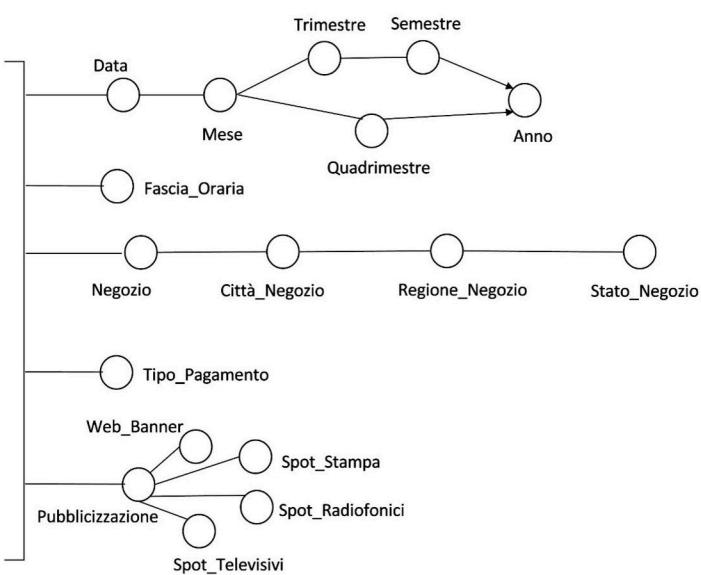






Risposta corretta.

La risposta corretta è:



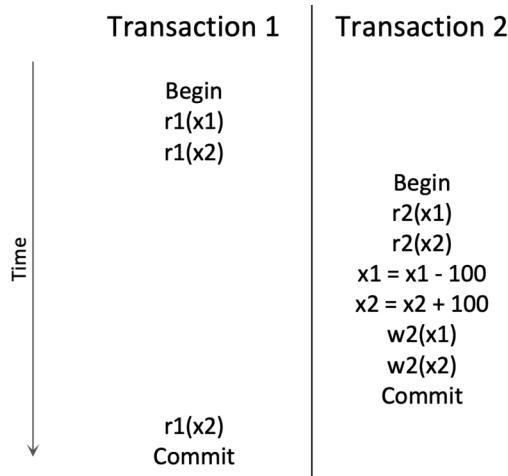
Domanda 2

Risposta non data

Punteggio max.: 1,00

1 punto (- 15% penalità per risposta errata)

Il seguente diagramma mostra l'esecuzione di due transazioni in parallelo in assenza di uno scheduler. Quale tipo di anomalia si verifica?

**Notazione:**

- $rX(o)$: Operazione di lettura da parte della transazione X dell'oggetto o
- $wX(o)$: Operazione di scrittura da parte della transazione X dell'oggetto o

- (a) Lost update
 (b) Ghost update (b)
 (c) Dirty read
 (d) Non si verifica alcuna anomalia
 (e) Ghost update (a)
 (f) Inconsistent read

Risposta errata.

La risposta corretta è: Inconsistent read

Domanda 3

Completo

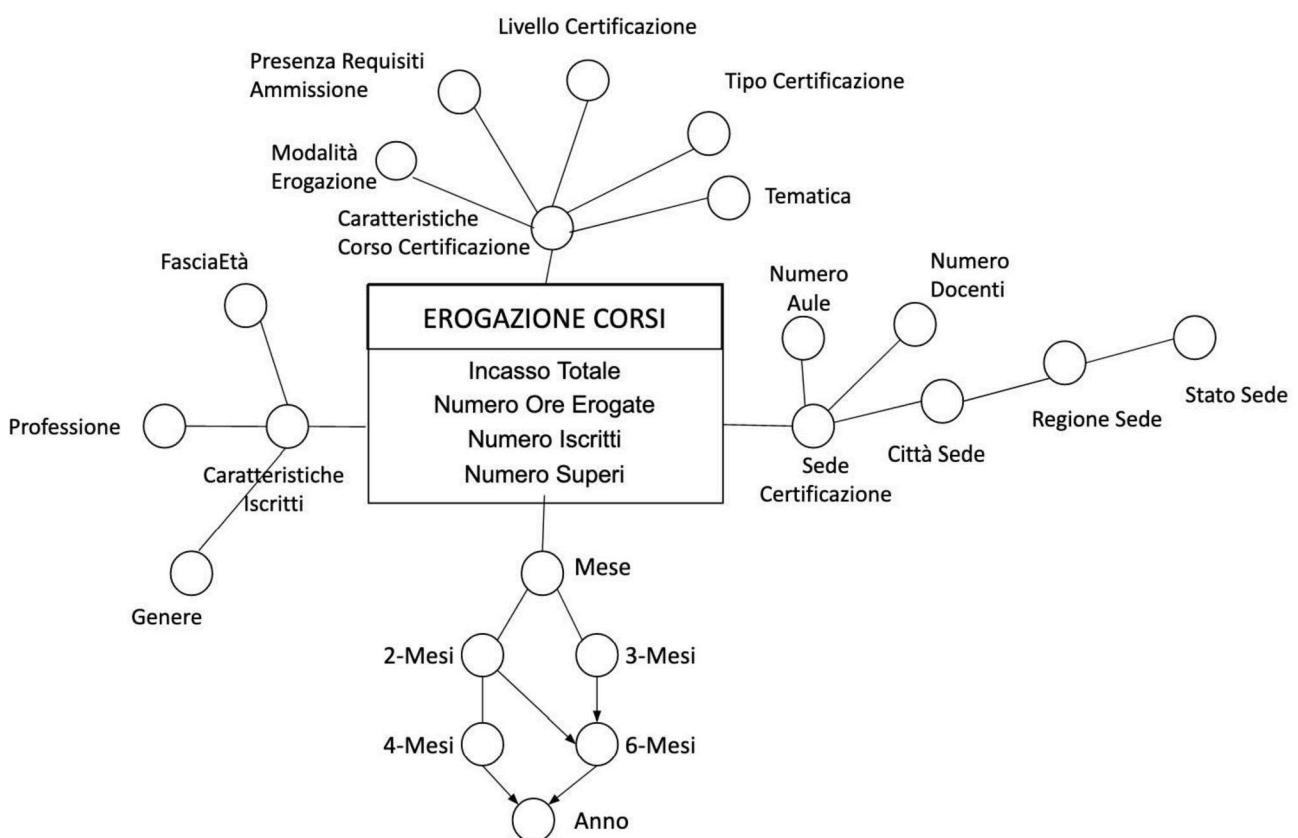
Punteggio ottenuto 3,00 su 4,00

Query SQL esteso 1 (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)

Separatamente per tipo certificazione e bimestre (attributo 2-Mesi), visualizzare:

- l'incasso,
- il rapporto tra il numero di superi e il numero di iscritti,
- la percentuale del numero di superi rispetto al totale semestrale di superi per tipo di certificazione
- il totale cumulativo dell'incasso al trascorrere dei bimestri, separatamente per semestre e tipo di certificazione.

Si effettui l'analisi separatamente per genere.

```

SELECT TIPOCERTIFICAZIONE,2-MESI,GENERE,
SUM(INCASSOTOTALE),
SUM(NUMEROSUPERI)/SUM(NUMEROISCRITTI),
100*SUM(NUMEROSUPERI)/SUM(SUM(NUMEROSUPERI)) OVER(PARTITION BY 6-MESI, TIPOCERTIFICAZIONE),
SUM(SUM(INCASSOTOTALE))OVER(PARTITION BY 6-MESI, TIPOCERTIFICAZIONE ORDER BY 2-MESI ROWS UNBOUNDED PRECEDINGS)
FROM EROGAZIONE-CORSI EC,TEMPO T,CARATTERISTICHE-CORSO-CERTIFICAZIONE CCC,CARATTERISTICHE-ISCRITTI CI
WHERE EC.IDTEMPO=T.IDTEMPO AND EC.IDCARATTCORSOCERT=CCC.IDCARATTCORSOCERT AND
EC.IDCARATTISCRITTI=CI.CARATTISCRITTI
GROUP BY GENERE, TIPOCERTIFICAZIONE,2-MESI
  
```

```

SELECT TipoCertificazione, 2-Mesi, Genere
SUM(Incasso Totale), SUM(Numeroiscritti)/SUM(Numerosuperi),
  
```

```

SELECT TipoCertificazione, 2-Mesi, Genere
SUM(Incasso Totale), SUM(Numeroliscritti)/SUM(NumeroSUPERI),
100*SUM(NumeroSUPERI)/SUM(SUM(NumberoSUPERI)) OVER (PARTITION BY 6-Mesi, TipoCertificazione, Genere),
SUM(SUM(IncassoTotale)) OVER (PARTITION BY 6-Mesi, TipoCertificazione, Genere ORDER BY 2-Mesi ROWS UNBOUNDED PRECEDING)
FROM EROGAZIONE-CORSI EC, TEMPO T, CARATTERISTICHE-CORSO-CERTIFICAZIONE CC, CARATTERISTICHE-ISCRITTI CI
WHERE EC.IDTempo=T.IDTempo AND CC.IDCarattCorsoCert=EC.IDCarattCorsoCert AND CI.IDCarattIscritti=EC.IDCarattIscritti
GROUP BY TipoCertificazinoe, 2-Mesi, 6-Mesi, Genere

```

Commento:

Manca 6-mesi nella GB

Manca Genere nelle due partition

Domanda 4

Risposta non data

Punteggio max.: 1,00

1 punto (-15% per risposta sbagliata)

Il seguente e' un esempio di documento preso dalla collection *guests* di un ristorante, dove sono inserite le informazioni circa gli ospiti di un ristorante.

```
{
  "firstName": "John",
  "lastName": "Smith",
  "yearOfBirth": 1990,
  "dietaryRequirements": "vegetarian",
  "table": 12
}
```

Si vuole estrarre, per ogni tavolo (*table*) il numero di ospiti con richieste di tipo vegetariano o vegano (*dietaryRequirements* = "vegetarian" o "vegan").

Quale delle seguenti query indirizza la richiesta?

(a)

```
db.guests.aggregate([
  {
    $group: {
      _id: "$table",
      count: { $sum: {} }
    }
  },
  {
    $match: {
      $or: [
        { dietaryRequirements: "vegetarian" },
        { dietaryRequirements: "vegan" }
      ]
    }
  }
])
```

(b)

```
db.guests.aggregate([
  {
    $group: {
      _id: "$table",
      count: { $sum: 1 }
    }
  },
  {
    $match: {
      $and: [
        { dietaryRequirements: "vegetarian" },
        { dietaryRequirements: "vegan" }
      ]
    }
  }
])
```

(c)

```
db.guests.aggregate([
  {
    $group: {
      _id: "$table",
      count: { $sum: {} }
    }
  },
  {
    $match: {
      $and: [
        { dietaryRequirements: "vegetarian" },
        { dietaryRequirements: "vegan" }
      ]
    }
  }
])
```

(d)

```
db.guests.aggregate([
  {
    $match: {
      $or: [
        { dietaryRequirements: "vegetarian" },
        { dietaryRequirements: "vegan" }
      ]
    }
  },
  {
    $group: {
      _id: "$table",
      count: { $sum: 1 }
    }
  }
])
```

(e)

```
db.guests.aggregate([
  {
    $group: {
      _id: "$table",
      count: { $sum: 1 }
    }
  },
  {
    $match: {
      $or: [
        { dietaryRequirements: "vegetarian" },
        { dietaryRequirements: "vegan" }
      ]
    }
  }
])
```

(f)

```
db.guests.aggregate([
  {
    $match: {
      $and: [
        { dietaryRequirements: "vegetarian" },
        { dietaryRequirements: "vegan" }
      ]
    }
  },
  {
    $group: {
      _id: "$table",
      count: { $sum: {} }
    }
  }
])
```

(g)

```
db.guests.aggregate([
  {
    $match: {
      $or: [
        { dietaryRequirements: "vegetarian" },
        { dietaryRequirements: "vegan" }
      ]
    }
  },
  {
    $group: {
      _id: "$table",
      count: { $sum: {} }
    }
  }
])
```

(h)

```

db.guests.aggregate([
{
  $match: {
    $and: [
      { dietaryRequirements: "vegetarian" },
      { dietaryRequirements: "vegan" }
    ]
  }
},
{
  $group: {
    _id: "$table",
    count: { $sum: 1 }
  }
}
])

```

Risposta errata.

La risposta corretta è:

```

db.guests.aggregate([
{
  $match: {
    $or: [
      { dietaryRequirements: "vegetarian" },
      { dietaryRequirements: "vegan" }
    ]
  }
},
{
  $group: {
    _id: "$table",
    count: { $sum: 1 }
  }
}
])

```

Domanda 5

Risposta corretta

Punteggio ottenuto 1,50 su 1,50

Indici (1,5 punti, penalità -15% per ogni risposta sbagliata)

Sono date le seguenti tabelle:

PERSONALE(PId, Nome, Cognome, Email, Telefono, DataNascita, AnniEsperienza, SId)
SPECIALIZZAZIONE(SIId, Nome, Categoria)
CONDOMINIO(CId, Nome, Indirizzo, Regione, Categoria)
INTERVENTO(Data, CId, PId, Importo)

Sono date le seguenti cardinalità:

- card(PERSONALE) = 10^5 tuple
 - MIN(AnniEsperienza) = 1, MAX(AnniEsperienza) = 30
 - MIN(DataNascita) = 1/1/1930, MAX(DataNascita) = 31/12/1999
- card(SPECIALIZZAZIONE) = 10^2 tuple
 - Valori distinti di Categoria = 10

- $\text{card}(\text{CONDOMINIO}) = 10^9$ tuple
 - Valori distinti di CATEGORIA = 10
- $\text{card}(\text{INTERVENTO}) = 2 \cdot 10^9$ tuple
 - MIN(Importo) = 100, MAX(Importo) = 100000,
 - MIN(Data) = 1/1/2003, MAX(Data) = 31/12/2022

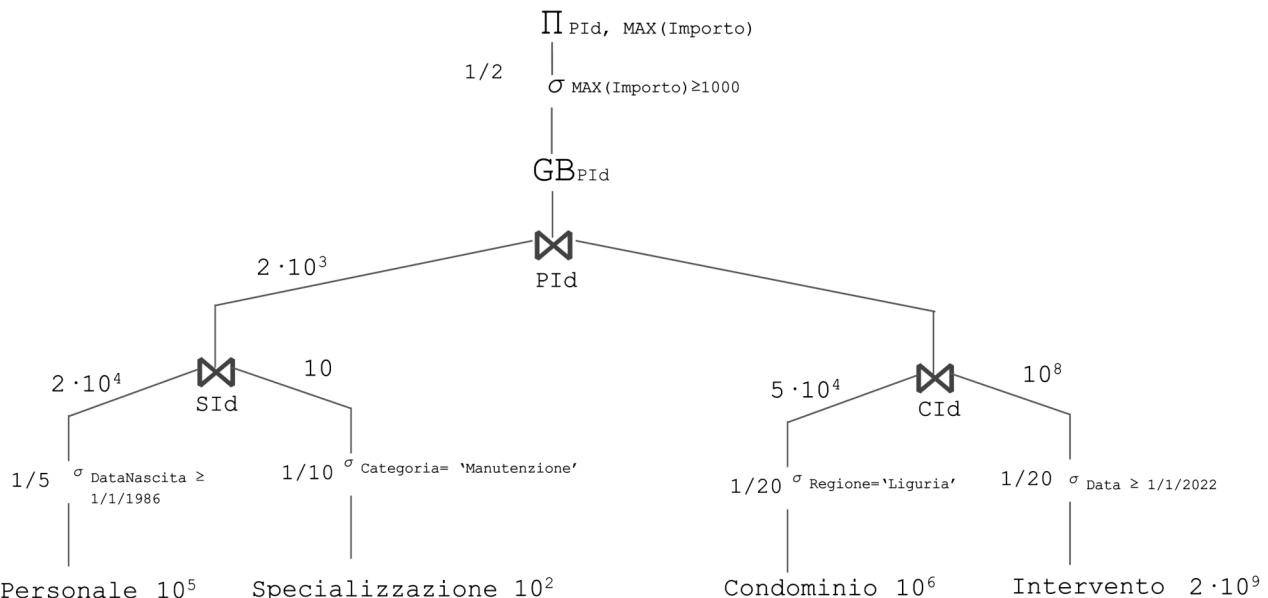
Inoltre, sono dati i seguenti fattori di riduzione per le clausole having:

Having $\text{MAX}(\text{Importo}) \geq 1000 = 1/2$

Si consideri la seguente query:

```
SELECT I.PId, MAX(Importo)
FROM INTERVENTO I, PERSONALE P, SPECIALIZZAZIONE S, CONDOMINIO C
WHERE I.PId=P.PId and P.SId=S.SId and I.CId=C.CId
      and S.CATEGORIA='Manutenzione' and C.Regione='Liguria' and I.Data >= 1/1/2022
      and P.DataNascita >= 1/1/1986)
GROUP BY I.PId
HAVING MAX(Importo) >= 1000
```

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 IndexD ON SPECIALIZZAZIONE(CATEGORIA) - B+-Tree
- (b) CREATE INDEX IndexH ON INTERVENTO(Data) - B+-Tree ✓
- (c) CREATE INDEX IndexA ON PERSONALE(DataNascita) - HASH
- (d) CREATE INDEX IndexG ON INTERVENTO(Data) - HASH
- (e) CREATE INDEX IndexE ON CONDOMINIO(Regione) - HASH ✓
- (f) CREATE INDEX IndexF ON CONDOMINIO(Regione) - B+-Tree
- (g) CREATE INDEX IndexC ON SPECIALIZZAZIONE(CATEGORIA) - HASH
- (h) CREATE INDEX IndexB ON PERSONALE(DataNascita) - B+-Tree
- (i) Nessuna - le strutture fisiche accessorie non migliorerebbero le prestazioni dell'interrogazione

Risposta corretta.

La risposta corretta è: CREATE INDEX IndexE ON CONDOMINIO(Regione) - HASH, CREATE INDEX IndexH ON INTERVENTO(Data) - B+-Tree

Domanda 6

Risposta corretta

Punteggio ottenuto 1,00 su 1,00

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

- La precisione di una classe e' definita come la frazione di punti correttamente predetti per quella classe, rispetto a tutti i punti predetti per la classe stessa
- Il richiamo di una classe e' definito come la frazione di punti correttamente predetti per quella classe, rispetto a tutti i punti appartenenti alla classe stessa

Vengono dati y_{pred} e y_{true} , vettori rispettivamente di predizioni e di ground truth per un test set di 10 punti.

y_{pred} : [B C A A A A B C C C]
 y_{true} : [C B A A B C B C A C]

Quali sono la precisione e il richiamo per la classe A?

Tutte le risposte sono fornite nel formato (precisione, richiamo).

- (a) (1/3, 1/2)
- (b) (2/3, 1/3)
- (c) (1/2, 1/2)
- (d) Nessuna delle altre risposte e' corretta
- (e) (1/2, 1/3)
- (f) (1/2, 2/3) ✓
- (g) (2/3, 1/2)
- (h) (1/3, 2/3)

Risposta corretta.

La risposta corretta è: (1/2, 2/3)

Domanda 7

Completo

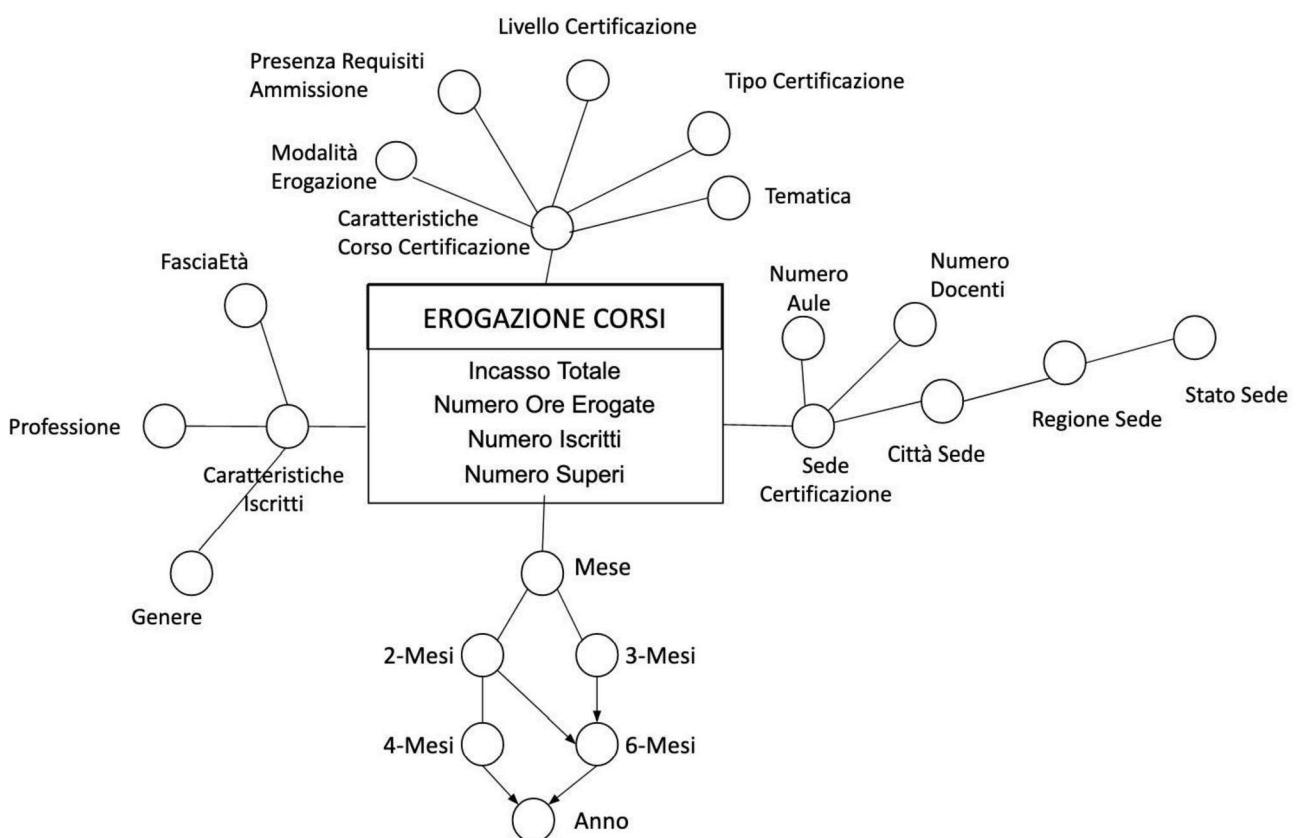
Punteggio ottenuto 4,25 su 5,00

Vista materializzata (5 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)

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

- Considerando solo gli iscritti di genere maschile nella fascia d'età > 45, separatamente per regione sede di certificazione e professione, visualizzare numero di iscritti totale, numero di superi totale e incasso totale.
- Separatamente per stato sede di certificazione e mese, visualizzare l'incasso totale mensile e l'incasso cumulativo annuale al trascorrere dei mesi.
- Separatamente per professione, considerando solo la regione sede di certificazione Piemonte, visualizzare il numero medio mensile di ore erogate e l'incasso medio mensile.

Dato lo schema logico precedente, si svolgono le seguenti attività

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

```
CREATE MATERIALIZED VIEW ViewCorsi
BUILD IMMEDIATE
REFRESH FAST ON COMMIT
AS
Blocco A
```

- Definire l'**insieme minima** di attributi che permette di identificare le tuple appartenenti alla vista materializzata ViewCorsi.
- Si ipotizzi che la gestione della vista materializzata (tabella derivata) sia svolta mediante trigger. Scrivere il trigger per propagare le modifiche alla vista materializzata ViewCorsi in caso di inserimento di un nuovo record nella tabella dei fatti EROGAZIONE-CORSI.

```
SELECT REGIONESEDE, PROFESSIONE,
SUM(NUMEROISCRITTI), SUM(NUMEROSUPERI), SUM(INCASSOTOTALE)
```

```

FROM CARATTERISTICHE-ISCRITTI CI,SEDE-CERTIFICAZIONE SC, EROGAZIONE-CORSI EC
WHERE EC.IDSEDECERTIFICAZIONE=SC.IDSEDECERTIFICAZIONE AND
EC.IDCARATTISCRITTI = CI.IDCARATTISCRITTI AND GENERE='M' AND FASCIAETA='>45'
GROUP BY REGIONESEDE,PROFESSIONE

SELECT STATOSEDE,MESE,
SUM(INCASSOTOTALE),
SUM(SUM(INCASSOTOTALE))OVER(PARTITION BY STATOSEDE ORDER BY MESE),
FROM EROGAZIONE-CORSI EC,TEMPO T,SEDE-CERTIFICAZIONE SC
WHERE EC.IDTEMPO = T.IDTEMPO AND SC.IDSEDECERTIFICAZIONE = EC.IDSEDECERTIFICAZIONE
GROUP BY STATOSEDE,MESE

```

```

SELECT PROFESSIONE,
SUM(INCASSOTOTALE)/COUNT(DISTINCT MESE),
SUM(NUMEROOREEROGATE)/COUNT(DISTINCT MESE)
FROM CARATTERISTICHE-ISCRITTI CI, EROGAZIONE-CORSI EC, TEMPO T,SEDE-CERTIFICAZIONE SC
WHERE CI.IDCARATTISCRITTI = EC.CARATTISCRITTI AND EC.IDTEMPO = T.IDTEMPO AND
EC.IDSEDECERTIFICAZIONE = SC.IDSEDECERTIFICAZIONE AND REGIONESEDE = 'PIEMONTE'
GROUP BY PROFESSIONE

```

```

SELECT REGIONESEDE,PROFESSIONE,STATOSEDE,MESE,GENERE,FASCIAETA,SUM(INCASSOTOTALE) AS TOTINCASSO,
SUM(NUMEROISCRITTI) AS TOTISCRITTI, SUM(NUMEROSUPERI) AS TOTSUPERI, SUM(NUMEROOREEROGATE) AS TOTORE
FROM CARATTERISTICHE-ISCRITTI CI,SEDE-CERTIFICAZIONE SC, TEMPO T, EROGAZIONE-CORSI EC
WHERE CI.IDCARATTISCRITTI=EC.IDCARATTISCRITTI AND T.IDTEMPO = EC.IDTEMPO AND SC.IDSEDECERTIFICAZIONE =
EC.IDSEDECERTIFICAZIONE
GROUP BY REGIONESEDE,PROFESSIONE,STATOSEDE,MESE,REGIONESEDE,GENERE,FASCIAETA

```

INSIEME MINIMALE: FASCIAETA,PROFESSIONE,GENERE,REGIONESEDE,MESE

```

CREATE OR UPDATE TRIGGER MAINTAINVIEWCORSI
AFTER INSERT ON EROGAZIONE-CORSI
FOR EACH ROW
DECLARE
VARMSE DATE;
VARREGIONESEDE,VARPROFESSIONE,VARSTATOSEDE,VARGENERE,VARFASCIAETA VARCHAR(20);
N INTEGER;
BEGIN

SELECT MESE INTO VARMSE
FROM TEMPO
WHERE IDTEMPO = :NEW.IDTEMPO

SELECT REGIONESEDE,STATOSEDE INTO VARREGIONESEDE,VARSTATOSEDE
FROM SEDE-CERTIFICAZIONE
WHERE IDSEDECERTIFICAZIONE = :NEW.IDSEDECERTIFICAZIONE

SELECT PROFESSIONE,GENERE,FASCIAETA INTO VARPROFESSIONE,VARGENERE,VARFASCIAETA
FROM CARATTERISTICHE-ISCRITTI
WHERE ID.CARATTISCRITTI = :NEW.IDCARATTISCRITTI

SELECT COUNT(*) INTO N
FROM VIEWCORSI
WHERE MESE=VARMSE AND REGIONESEDE=VARREGIONESEDE AND STATOSEDE=VARSTATOSEDE AND
PROFESSIONE=VARPROFESSIONE AND GENERE=VARGENERE AND FASCIAETA=VARFASCIAETA

IF N>0 THEN
UPDATE VIEWCORSI
SET TOTINCASSO = TOTINCASSO + :NEW.TOTALEINCASSO,TOTISCRITTI = TOTISCRITTI + :NEW.NUMEROISCRITTI,
TOTSUPERI = TOTSUPERI + :NEW.NUMEROSUPERI, TOTORE = TOTORE + :NEW.NUMEROOREEROGATE
WHERE MESE=VARMSE AND REGIONESEDE=VARREGIONESEDE AND STATOSEDE=VARSTATOSEDE AND GENERE=VARGENERE AND
FASCIAETA=VARFASCIAETA AND PROFESSIONE=VARPROFESSIONE
ELSE
INSERT INTO
VIEWCORSI(...)VALUES(VARMSE,VARREGIONESEDE,VARSTATOSEDE,VARGENERE,VARFASCIAETA,VARPROFESSIONE,:NEW.NUMEROSUPERI,
:NEW.NUMEROISCRITTI,:NEW.NUMEROOREEROGATE,:NEW.TOTALEINCASSO);

```

END IF;

1. Blocco A - Query per vista materializzata

```
SELECT Genere, Professione, FasciaEtà, RegioneSede, StatoSede, Mese, Anno,
SUM(IncassoTotale) AS TotIncasso, SUM(NumerOreErogate) AS OreTotali, SUM(NumerOlscritti) AS IscrittiTotali, SUM(NumeroSperi) AS SperiTotali
FROM CARATTERISTICHE-ISCRITTI CI, SEDE-CERTIFICAZIONE SE, TEMPO T, EROGAZIONE-CORSI EA
WHERE CI.IDCarattIscritti = EA.IDCarattIscritti
AND SE.IDSedeCertificazione = EA.IDSedeCertificazione
AND T.IDTempo = EA.IDTempo
GROUP BY Genere, Professione, FasciaEtà, RegioneSede, StatoSede, Mese, Anno
```

2. Identificatore

Genere, Professione, FasciaEtà, RegioneSede, Mese

3. Trigger

```
CREATE OR REPLACE TRIGGER TriggerViewCorsi
AFTER INSERT ON EROGAZIONE-CORSI
FOR EACH ROW
DECLARE

VarAnno DATE, VarMese DATE;
VarRegione, VarStato varchar(10);
VarGenere, VarProfessione, VarFasciaEtà varchar(10);
N INTEGER;

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

SELECT Genere, Professione, FasciaEtà INTO VarGenere, VarProfessione, VarFasciaEtà
FROM CARATTERISTICHE-ISCRITTI
WHERE IDCarattIscritti = :NEW.IDCarattIscritti;

SELECT RegioneSede, StatoSede INTO VarRegione, VarStato
FROM SEDE-CERTIFICAZIONE
WHERE IDSedeCertificazione = :NEW.IDSedeCertificazione;

SELECT COUNT(*) INTO N
FROM ViewCorsi
WHERE Mese = VarMese AND Genere = VarGenere
AND Professione= VarProfessione AND FasciaEtà = VarFasciaEtà
AND RegioneSede = varRegione;

IF N>0 THEN
    UPDATE ViewCorsi
    SET IncassoTotale = IncassoTotale + :NEW.IncassoTotale,
        OreTotali = OreTotali + :NEW.NumerOreErogate,
        IscrittiTotali = IscrittiTotali + :NEW.NumerOlscritti,
        SperiTotali = SperiTotali + :NEW.NumeroSperi
    WHERE Mese = VarMese AND Genere= VarGenere
    AND Professione= VarProfessione AND FasciaEtà= VarFasciaEtà
    AND RegioneSede = VarRegione;
ELSE
    INSERT INTO ViewCorsi (...) VALUES (VarMese, VarAnno, VarGenere, VarProfessione, VarFasciaEtà, VarRegione, VarStato, :NEW.IncassoTotale,
    :NEW.NumerOreErogate, :NEW.NumerOlscritti, :NEW.NumeroSperi);
END IF;
END
```

Commento:
Erori

- Manca attributo Anno in select e GB della definizione vista
- Attributo StatoSede non necessario in predicato selezione di select count(*) e di update
- Manca END trigger (non valutato)

Domanda 8

Risposta corretta

Punteggio ottenuto 1,00 su 1,00

1 punto (15% di penalità per una risposta sbagliata)

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:

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

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

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

Risposta corretta.

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

Domanda 9

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.

Domanda 10

Completo

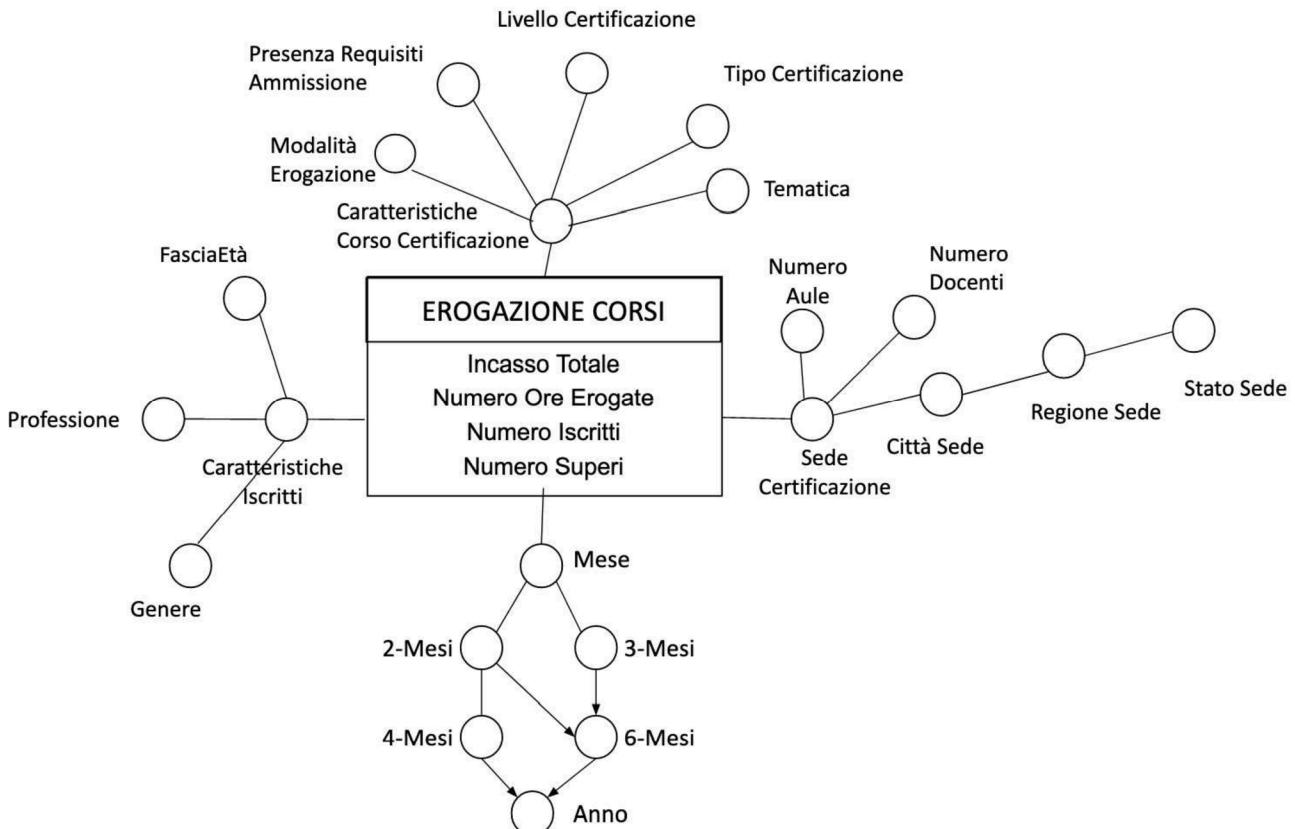
Punteggio ottenuto 3,50 su 4,00

Query SQL esteso 2 (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)

Considerando i corsi con modalità di erogazione online, separatamente per professione degli iscritti, trimestre (3-Mesi) e stato della sede di certificazione visualizzare

- Il numero di iscritti e il numero di superi
- l'incasso medio per numero di iscritti
- l'incasso medio mensile
- l'incasso complessivo separatamente per professione degli iscritti, stato della sede di certificazione e anno,
- la posizione in una graduatoria (rank) in ordine decrescente rispetto al numero di superi, separatamente per anno.

SELECT PROFESSIONE,3-MESI,STATOSEDE
SUM(NUMEROSUPERI),SUM(NUMEROISCRITTI),

```
SUM(INCASSOTOTALE)/COUNT(DISTINCT MESE),
SUM(SUM(INCASSOTOTALE))OVER(PARTITION BY PROFESSIONE,STATOSEDE,ANNO)
RANK()OVER(PARTITION BY ANNO ORDER BY SUM(NUMEROSUPERI) DESC)
FROM TEMPO T,EROGAZIONE-CORSI EC,CARATTERISTICHE-ISCRITTI EC,SEDE-CERTIFICAZIONE SC
WHERE T.IDTEMPO=EC.IDTEMPO AND EC.IDCARATTISCRITTI=CI.IDCARATTISCRITTI AND
SC.IDSEDECERTIFICAZIONE=EC.IDSEDECERTIFICAZIONE AND MODALITAEROGAZIONE='ONLINE'
GROUP BY PROFESSIONE,3-MESI,STATOSEDE
```

```
SELECT Professione, 3-Mesi, StatoSede
SUM(Numerolscritti), SUM(Numerosuperi), SUM(IncassoTotale)/SUM(Numerolscritti)
SUM(IncassoTotale)/ COUNT (DISTINCT Mese)
SUM(SUM(IncassoTotale)) OVER (PARTITION BY A Professione, Anno, StatoSede),
RANK() OVER (PARTITION BY Anno ORDER BY SUM(Numerosuperi) DESC)
FROM EROGAZIONE-CORSI EC, TEMPO T, SEDE-CERTIFICAZIONE CO, CARATTERISTICHE-ISCRITTI E, CARATTERISTICHE-CORSO-CERTIFICAZIONE CCC
WHERE EC.IDTempo=T.IDTempo AND EC.IDSedeCertificazione=CO.IDSedeCertificazione AND E.IDCarattCorsoCert=EC.CarattCorsoCert
AND CCC.IDCarattCorsoCert= EC.IDCarattCorsoCert
AND ModalitàErogazione, ="online"
GROUP BY Professione, 3-Mesi, StatoSede, Anno
```

Commento:
Manca Anno nella GB

Domanda 11

Risposta errata

Punteggio ottenuto -0,30 su 2,00

2 punti (penalità 15% per risposta errata)

Viene data la base dati transazionale mostrata qui sotto. Si applichi l'algoritmo Apriori per l'estrazione di itemset frequenti.

Transactions

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

Il valore di minsup è 2 (un itemset è frequente se appare in almeno 2 transazioni).

Quali sono gli itemset candidati di lunghezza 3 che vengono generati nel join step e poi prunati nel prune step?

Si utilizzi l'ordinamento alfabetico per stabilire un ordine fra gli items.

- (a) { ABE, BCE, CDE }
- (b) { ABC, ABE, ACE, BCE, CDE } X
- (c) { ABD, ACD, ADE, BCD, BDE }
- (d) { ABC, ABD, ABE, ACD, ACE, ADE, BCD, BCE, BDE, CDE }
- (e) { ABC, ACE }
- (f) {}

Risposta errata.

La risposta corretta è: {}

Domanda 12

Risposta corretta

Punteggio ottenuto 1,00 su 1,00

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

Un istituto di analisi di mercato vuole analizzare le informazioni raccolte sulle vendite di prodotti dolciari prodotti da una multinazionale negli ultimi anni.

I prodotti sono realizzati in stabilimenti dislocati in diverse aree geografiche. La multinazionale produce prodotti dolciari di diverso tipo (ad esempio merendine, torte, biscotti), e in modo da venire incontro alle esigenze di diversi regimi alimentari (ad esempio vegetariano o vegano).

La multinazionale effettua la vendita dei prodotti attraverso diversi punti vendita al dettaglio distribuiti in varie aree geografiche.

Per analizzare l'andamento delle vendite dei diversi prodotti rispetto a diversi segmenti di mercato, la multinazionale tiene traccia, attraverso le carte fedeltà, di alcune informazioni sui clienti che hanno acquistato i prodotti, quali il genere, la fascia di età e il tipo di impiego.

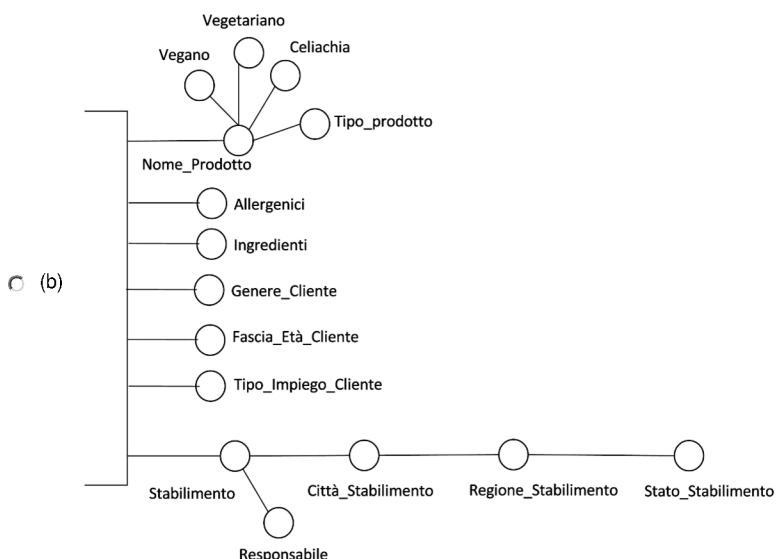
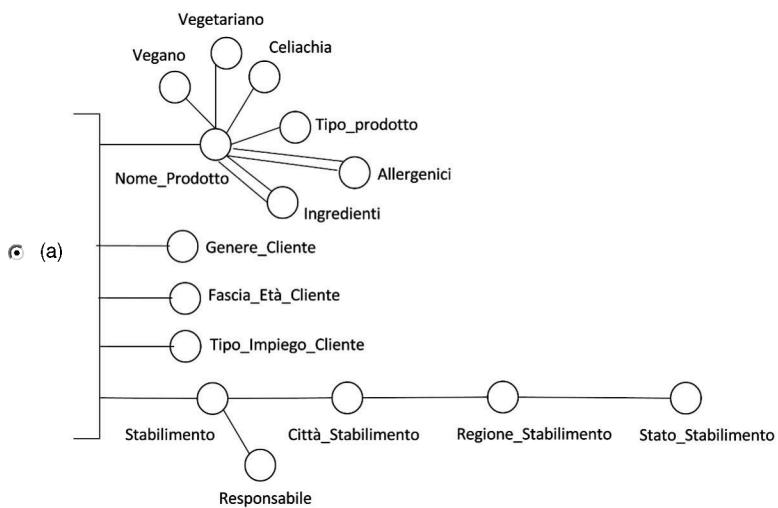
La multinazionale si avvale di diversi canali di comunicazione per pubblicizzare i suoi prodotti mediante spot pubblicitari radiofonici, televisivi o su banner su siti web.

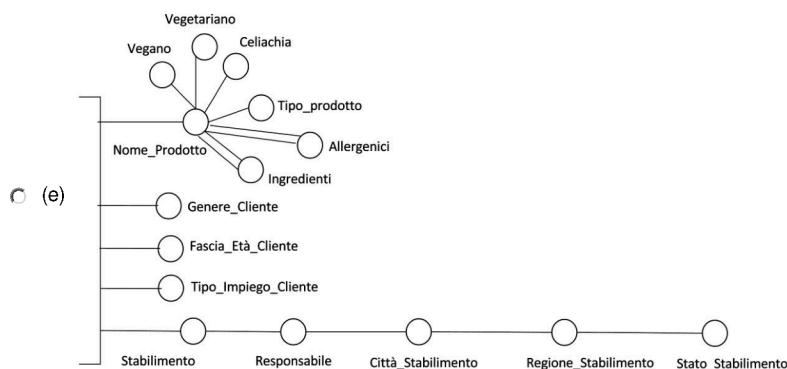
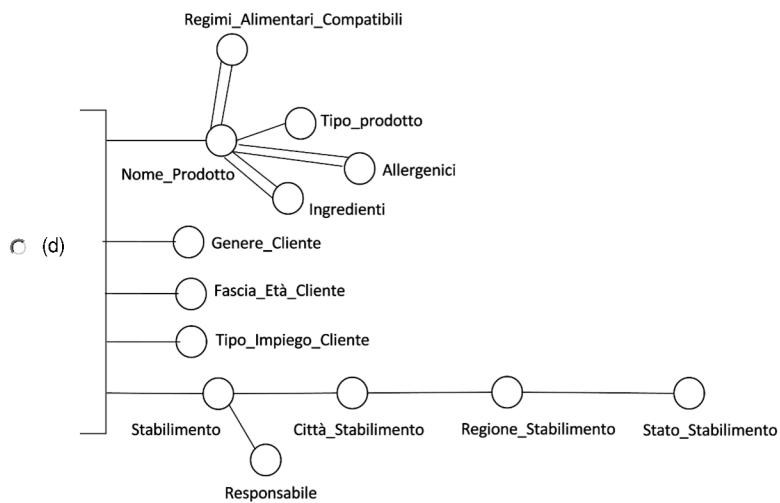
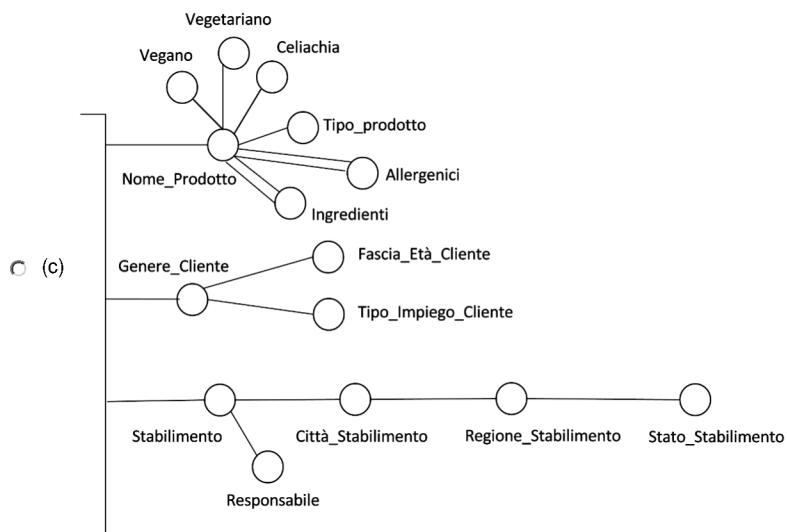
L'istituto di analisi di mercato vuole analizzare il costo medio delle azioni di pubblicità e il profitto di vendita complessivo in base a:

- stabilimento di produzione, caratterizzato dall'informazione sulla città, regione, stato in cui è collocato, e dal responsabile dello stabilimento

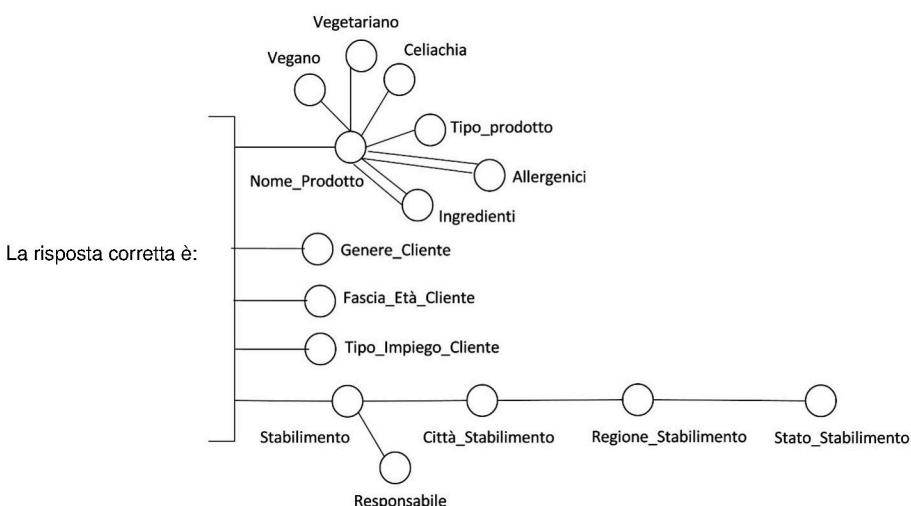
- prodotto, caratterizzato dal nome, dal tipo di prodotto (ad esempio torta, merendina, barretta, etc), l'indicazione dei regimi alimentari compatibili (uno o più valori tra vegetariano, vegano, celiachia), la lista degli ingredienti e la lista degli allergenici.
 - genere, fascia di età, tipo di impiego del cliente che ha effettuato l'acquisto. La fascia d'età è un valore tra < 20 anni, tra 20 e 29 anni, tra 30 e 49 anni, e maggiore di 50 anni. Il tipo di impiego è un valore tra libero professionista, dipendente, altro.
 - negozio in cui il prodotto è stato acquistato, caratterizzato dalla città regione e stato in cui il negozio è collocato.
 - il tipo di pagamento effettuato (un valore tra contanti, carta di credito e bancomat)
 - giorno, mese, trimestre, quadri mestre semestre, anno in cui si è verificato l'acquisto e fascia oraria (un valore tra mattino, pomeriggio, sera)
 - canali di comunicazione usati per la pubblicizzazione del prodotto (uno o più valori tra i seguenti spot pubblicitari radiofonici, spot pubblicitari a mezzo stampa, spot pubblicitari televisivi, banner pubblicitari web)

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





Risposta corretta.



Domanda 13

Risposta corretta

Punteggio ottenuto 1,50 su 1,50

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

Sono date le seguenti tabelle:

PERSONALE(PId, Nome, Cognome, Email, Telefono, DataNascita, AnniEsperienza, SId)
SPECIALIZZAZIONE(SId, Nome, Categoria)
CONDOMINIO(CId, Nome, Indirizzo, Regione, Categoria)
INTERVENTO(Data, CId, PId, Importo)

Sono date le seguenti cardinalità:

- $\text{card}(\text{PERSONALE}) = 10^5$ tuple
 - MIN(AnniEsperienza) = 1, MAX(AnniEsperienza) = 30
 - MIN(DataNascita) = 1/1/1930, MAX(DataNascita) = 31/12/1999
- $\text{card}(\text{SPECIALIZZAZIONE}) = 10^2$ tuple
 - Valori distinti di Categoria = 10
- $\text{card}(\text{CONDOMINIO}) = 10^6$ tuple
 - Valori distinti di Categoria = 10
- $\text{card}(\text{INTERVENTO}) = 2 \cdot 10^9$ tuple
 - MIN(Importo) = 100, MAX(Importo) = 100000,
 - MIN(Data) = 1/1/2003, MAX(Data) = 31/12/2022

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

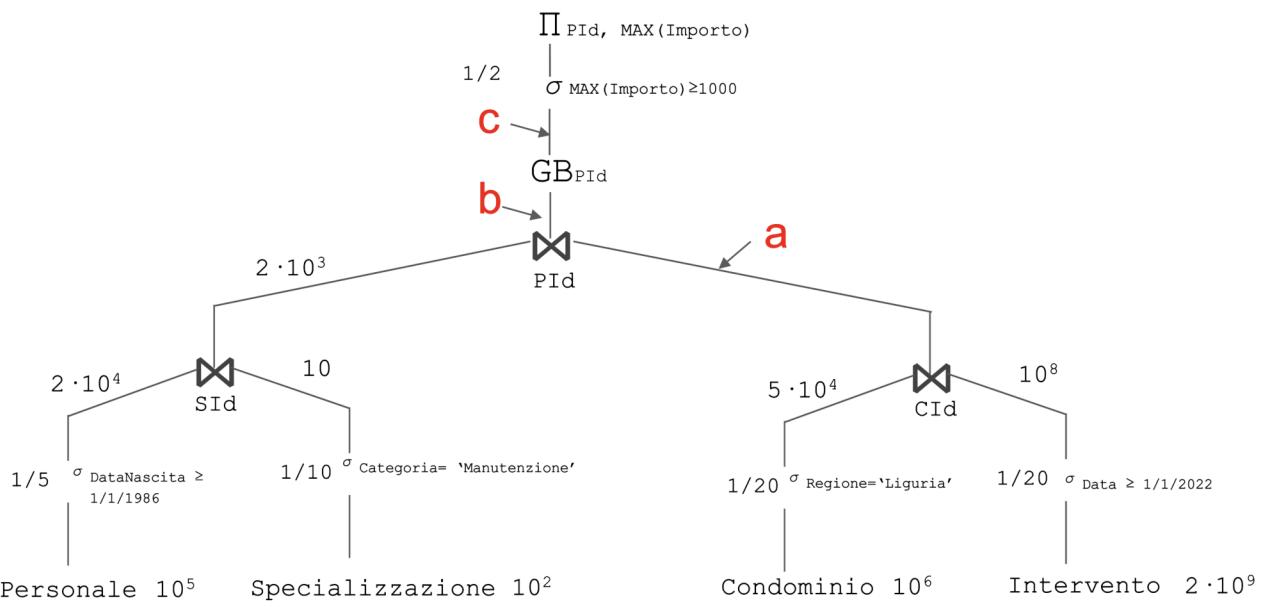
Having MAX(Importo) $\geq 1000 = 1/2$

Si consideri la seguente query:

```

SELECT I.PId, MAX(Importo)
FROM INTERVENTO I, PERSONALE P, SPECIALIZZAZIONE S, CONDOMINIO C
WHERE I.PId=P.PId and P.SId=S.SId and I.CId=C.CId
      and S.Categoria='Manutenzione' and C.Regione='Liguria' and I.Data >= 1/1/2022
      and P.DataNascita >= 1/1/1986)
GROUP BY I.PId
HAVING MAX(Importo)>=1000
  
```

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: $5 \cdot 10^4$
- (b) c: $2 \cdot 10^3$ ✓
- (c) b: $5 \cdot 10^3$
- (d) c: $2 \cdot 10^2$
- (e) b: 10^4
- (f) b: 10^5 ✓
- (g) a: $5 \cdot 10^6$ ✓
- (h) c: 10^5
- (i) a: $2 \cdot 10^7$
- (j) a: $5 \cdot 10^4$
- (k) b: 10^3
- (l) a: $2 \cdot 10^5$

Risposta corretta.

La risposta corretta è: a: $5 \cdot 10^6$, b: 10^5 , c: $2 \cdot 10^3$

Domanda 14

Parzialmente corretta

Punteggio ottenuto 0,52 su 1,00

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

Un istituto di analisi di mercato vuole analizzare le informazioni raccolte sulle vendite di prodotti dolciari prodotti da una multinazionale negli ultimi anni.

I prodotti sono realizzati in stabilimenti dislocati in diverse aree geografiche. La multinazionale produce prodotti dolciari di diverso tipo (ad esempio merendine, torte, biscotti), e in modo da venire incontro alle esigenze di diversi regimi alimentari (ad esempio vegetariano o vegano).

La multinazionale effettua la vendita dei prodotti attraverso diversi punti vendita al dettaglio distribuiti in varie aree geografiche.

Per analizzare l'andamento delle vendite dei diversi prodotti rispetto a diversi segmenti di mercato, la multinazionale tiene traccia, attraverso le carte fedeltà, di alcune informazioni sui clienti che hanno acquistato i prodotti, quali il genere, la fascia di età e il tipo di impiego.

La multinazionale si avvale di diversi canali di comunicazione per pubblicizzare i suoi prodotti mediante spot pubblicitari radiofonici, televisivi o su banner su siti web.

L'istituto di analisi di mercato vuole analizzare il costo medio delle azioni di pubblicizzazione e il profitto di vendita complessivo in base a:

- stabilimento di produzione, caratterizzato dall'informazione sulla città, regione, stato in cui è collocato, e dal responsabile dello stabilimento
- prodotto, caratterizzato dal nome, dal tipo di prodotto (ad esempio torta, merendina, barretta, etc), l'indicazione dei regimi alimentari compatibili (uno o più valori tra vegetariano, vegano, celiachia), la lista degli ingredienti e la lista degli allergenici.
- genere, fascia di età, tipo di impiego del cliente che ha effettuato l'acquisto. La fascia d'età è un valore tra < 20 anni, tra 20 e 29 anni, tra 30 e 49 anni, e maggiore di 50 anni. Il tipo di impiego è un valore tra libero professionista, dipendente, altro.
- negozio in cui il prodotto è stato acquistato, caratterizzato dalla città regione e stato in cui il negozio è collocato.
- il tipo di pagamento effettuato (un valore tra contanti, carta di credito e bancomat)
- giorno, mese, trimestre, quadrimestre semestre, anno in cui si è verificato l'acquisto e fascia oraria (un valore tra mattino, pomeriggio, sera)
- canali di comunicazione usati per la pubblicizzazione del prodotto (uno o più valori tra i seguenti spot pubblicitari radiofonici, spot pubblicitari a mezzo stampa, spot pubblicitari televisivi, banner pubblicitari web)

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 poiché possono essere necessarie più misure)

Scegli una o più alternative:

- (a) Numero prodotti venduti 
- (b) Profitto complessivo
- (c) Costo complessivo di pubblicizzazione 
- (d) Profitto medio
- (e) Costo massimo di pubblicizzazione
- (f) Costo medio di pubblicizzazione
- (g) Numero azioni di pubblicizzazione 

Risposta parzialmente esatta.

Hai selezionato correttamente 2.

La risposta corretta è: Numero azioni di pubblicizzazione, Costo complessivo di pubblicizzazione, Profitto complessivo

Domanda 15

Risposta non data

Punteggio max.: 2,00

2 punti (15% di penalità per una risposta sbagliata)

Un clustering puo' essere rappresentato come un insieme di n cluster $\{C_1, C_2, \dots, C_n\}$. Ciascun cluster C_i e' un insieme di punti che sono stati assegnati allo specifico cluster.

Dato un punto $x \in C_i$, la sua silhouette puo' essere calcolata come:

$$silh(x) = \frac{b(x) - a(x)}{\max(b(x), a(x))}$$

Dove:

- $a(x)$ e' la distanza media fra x e tutti gli altri punti in C_i , i.e.

$$a(x) = \frac{1}{|C_i|-1} \sum_{j \in C_i, j \neq x} dist(j, x)$$

- $b(x)$ e' la distanza minima tra x e i cluster a cui x non appartiene (la distanza fra un punto e un cluster e' calcolata come la distanza media fra il punto e tutti i punti nel cluster)

$$b(x) = \min_{C_k, k \neq i} \frac{1}{|C_k|} \sum_{j \in C_k} dist(j, x)$$

($dist(w, z)$ e' una qualche distanza fra i punti w e z)

La seguente e' una matrice di distanze di Manhattan fra 6 punti 2-dimensionalni.

	a	b	c	d	e	f
a	0	9	9	10	10	7
b	9	0	2	5	9	12
c	9	2	0	7	9	12
d	10	5	7	0	4	7
e	10	9	9	4	0	3
f	7	12	12	7	3	0

Un algoritmo di clustering e' stato utilizzato su questo dataset, ottenendo 3 clusters (0, 1, 2). Le etichette dei cluster assegnate ai 6 punti sono le seguenti:

a: 0

b: 0

c: 1

d: 1

e: 0

f: 2

Quali sono le silhouette dei punti **a** e **c**?

Le opzioni sono riportate come (silhouette a, silhouette c) utilizzando 4 cifre significative.

- (a) (+0.6111, +0.0952)
- (b) (+0.2632, +0.0476)
- (c) (+0.0476, +0.2632)
- (d) (-0.6111, -0.0952)
- (e) Nessuna delle opzioni e' corretta.
- (f) (-0.0476, -0.2632)
- (g) (-0.2632, -0.0476)
- (h) (-0.0952, -0.6111)
- (i) (+0.0952, +0.6111)

Risposta errata.

La risposta corretta è: (-0.2632, -0.0476)

Domanda 16

Risposta corretta

Punteggio ottenuto 2,00 su 2,00

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

Sono date le seguenti tabelle:

```
PERSONALE(PId, Nome, Cognome, Email, Telefono, DataNascita, AnniEsperienza, SId)
SPECIALIZZAZIONE(SId, Nome, Categoria)
CONDOMINIO(CId, Nome, Indirizzo, Regione, Categoria)
INTERVENTO(Data, CId, PId, Importo)
```

Sono date le seguenti cardinalità:

- $\text{card}(\text{PERSONALE}) = 10^5$ tuple
 - MIN(AnniEsperienza) = 1, MAX(AnniEsperienza) = 30
 - MIN(DataNascita) = 1/1/1930, MAX(DataNascita) = 31/12/1999
- $\text{card}(\text{SPECIALIZZAZIONE}) = 10^2$ tuple
 - Valori distinti di Categoria = 10
- $\text{card}(\text{CONDOMINIO}) = 10^6$ tuple
 - Valori distinti di Categoria = 10
- $\text{card}(\text{INTERVENTO}) = 2 \cdot 10^9$ tuple
 - MIN(Importo) = 100, MAX(Importo) = 100000,
 - MIN(Data) = 1/1/2003, MAX(Data) = 31/12/2022

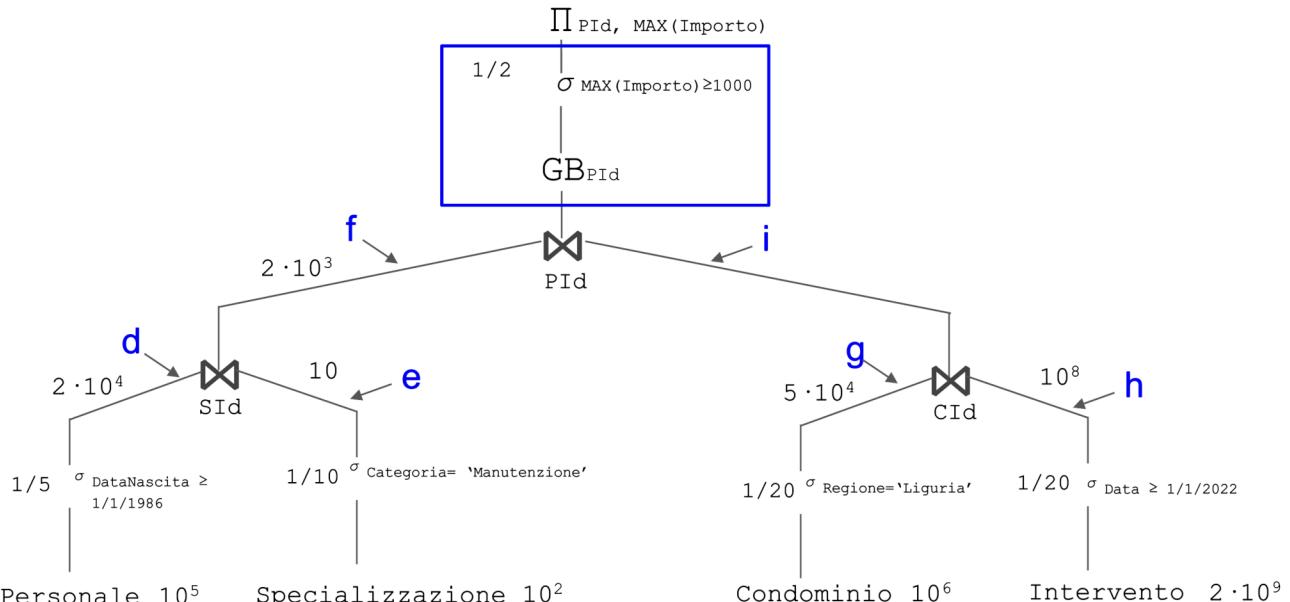
Inoltre, sono dati i seguenti fattori di riduzione per le clausole having:

Having MAX(Importo) $\geq 1000 = 1/2$

Si consideri la seguente query:

```
SELECT I.PId, MAX(Importo)
FROM INTERVENTO I, PERSONALE P, SPECIALIZZAZIONE S, CONDOMINIO C
WHERE I.PId=P.PId and P.SId=S.SId and I.CId=C.CId
      and S.Categoria='Manutenzione' and C.Regione='Liguria' and I.Data  $\geq 1/1/2022$ 
      and P.DataNascita  $\geq 1/1/1986$ )
GROUP BY I.PId
HAVING MAX(Importo)  $\geq 1000$ 
```

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



Analizzare l'anticipo della GROUP BY GROUP BY I.PId HAVING MAX(Importo) ≥ 1000 rappresentata nel riquadro. Selezionare la soluzione che consente la massima efficienza nell'esecuzione della query (se esiste).

- (a) E' possibile anticiparla nel ramo e
- (b) E' possibile anticiparla nel ramo f
- (c) Non e' possibile anticipare la Group BY Group BY **GROUP BY I.Pid HAVING MAX(Importo)≥1000**
- (d) E' possibile anticiparla nel ramo h
- (e) E' possibile anticiparla nel ramo d
- (f) E' possibile anticiparla nel ramo j
- (g) E' possibile anticiparla nel ramo g
- (h) E' possibile anticiparla nel ramo i ✓

Risposta corretta.

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

Domanda 17

Completo

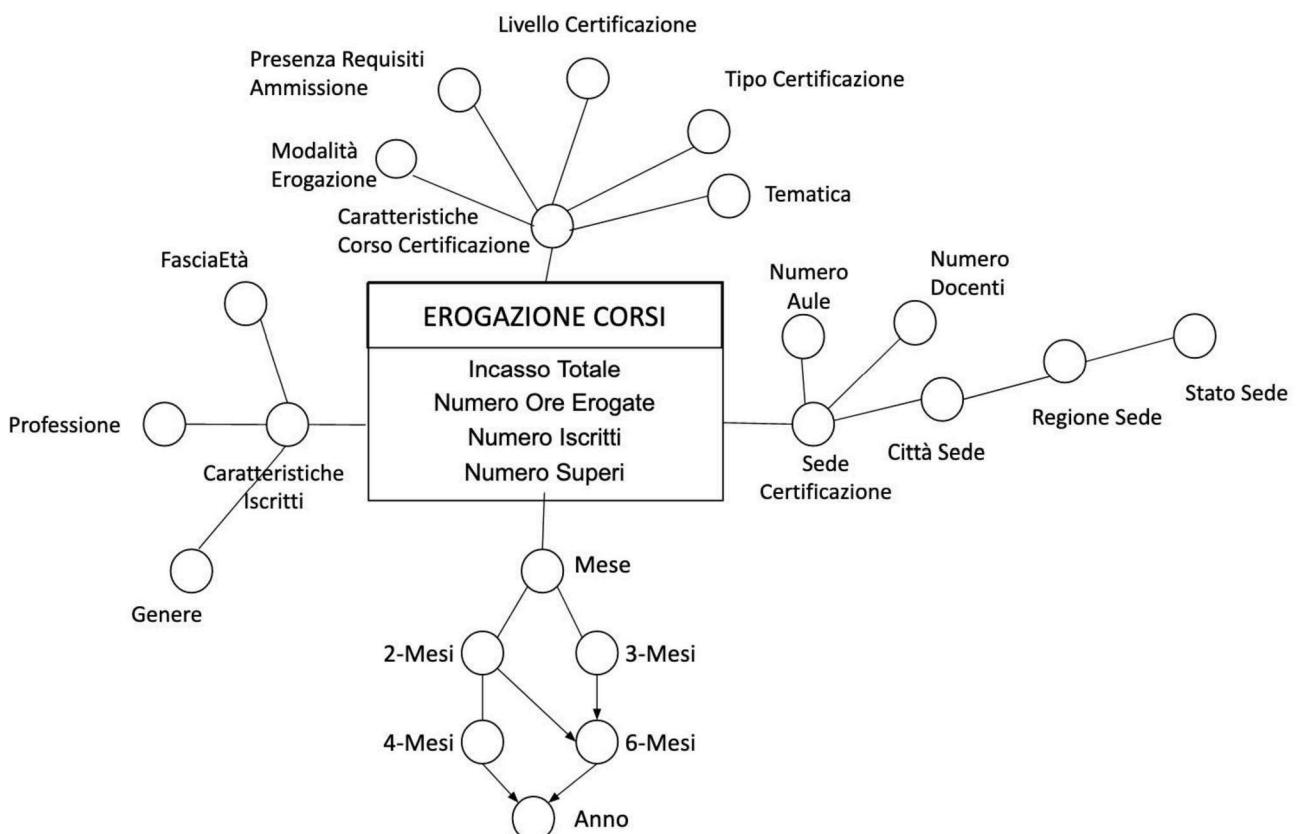
Punteggio ottenuto 1,00 su 3,00

Query SQL esteso 3 (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, 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 le sedi di certificazione site in Italia, separatamente per modalità di erogazione del corso e semestre (attributo 6-Mesi) visualizzare

- Il numero complessivo di ore erogate
- Il numero medio di superi per mese
- Il rapporto tra il numero di iscritti rispetto al numero complessivo di iscritti per anno e modalità di erogazione
- la posizione in una graduatoria (rank) in ordine decrescente rispetto al numero di ore complessive erogate.

Si effettui l'analisi separatamente per Tematica.

```

SELECT MODALITAEROGAZIONE,6-MESI,TEMATICA
SUM(NUMEROOREEROGATE),
SUM(NUMEROSUPERI)/COUNT(DISTINCT MESE),
SUM(NUMEROISCRITTI)/SUM(SUM(NUMEROISCRITTI))OVER(PARTITION BY ANNO,MODALITAEROGAZIONE),
RANK()OVER(ORDER BY SUM(NUMEROOREEROGATE) DESC)
FROM TEMPO T, EROGAZIONE-CORSI EC, CARATTERISTICHE-CORSO-CERTIFICAZIONE CCC, SEDE-CERTIFICAZIONE SC
WHERE T.IDTEMPO = EC.IDTEMPO AND EC.IDCARATTCORSOCERT=CCC.IDCARATTCORSOCERT AND
EC.IDSEDECERTIFICAZIONE = SC.IDSEDECERTIFICAZIONE AND STATOSEDE='ITALIA'
GROUP BY MODALITAEROGAZIONE,6-MESI,TEMATICA

```

```

SELECT TipoCertificazione, 2-Mesi, Genere
SUM(Incasso Totale), SUM(NumeroiScritti)/SUM(NumeroSUPERI),
100*SUM(NumeroSUPERI)/SUM(SUM(NumeroSUPERI)) OVER (PARTITION BY 6-Mesi, TipoCertificazione, Genere),
SUM(SUM(IncassoTotale)) OVER (PARTITION BY 6-Mesi, TipoCertificazione, Genere ORDER BY 2-Mesi ROWS UNBOUNDED PRECEDING)
FROM EROGAZIONE-CORSI EC, TEMPO T, CARATTERISTICHE-CORSO-CERTIFICAZIONE CC, CARATTERISTICHE-ISCRITTI CI
WHERE EC.IDTempo=T.IDTempo AND CC.IDCarattCorsoCert=EC.IDCarattCorsoCert AND CI.IDCarattIscritti=EC.IDCarattIscritti
GROUP BY TipoCertificazinoe, 2-Mesi, 6-Mesi, Genere

```

Commento:
Manca Anno nella GB

Manca Tematica nelle due partition

manca Tematica nella GB
Manca tematica nelle due finestre di calcolo