

FEBBRAIO 2020

Bologna

NOME
COGNOME
MATRICOLA

ESERCIZIO 1 (12 punti)

Dato il seguente schema relazionale, che modella i dati di un'applicazione che tiene traccia delle **catene montuose nel mondo**:

CATENAM(Nome, Lunghezza, Altitudine)

ESTENSIONE(NomeCatena, NomeStato)

STATO(Nome, Superficie, Continente)

Con i seguenti vincoli di integrità referenziale:

ESTENSIONE.NomeCatena → CATENAM.Nome

ESTENSIONE.NomeStato → STATO.Nome

a) (1 pt) Indicare quante righe e quante colonne produce in output (numero esatto o intervallo [min:max]) la query riportata sotto:

```
SELECT COUNT(*), NomeStato
FROM CATENAM, ESTENSIONE
WHERE (NomeCatena=Nome)
GROUPBY NomeStato
```

assumendo la seguente dimensione delle tabelle:

CATENAM → 40 righe ESTENSIONE → 50 righe

b) (1 pt) Riscrivere la query indicata sotto, evitando l'utilizzo del costrutto COUNT.

```
SELECT NomeStato
FROM CATENAM, ESTENSIONE
WHERE (NomeCatena=Nome)
GROUPBY NomeStato
HAVING (count(*) >=2)
```

c) (3 pt) Scrivere in **SQL** la query che calcola il numero totale delle catene montuose con altezza superiore ai 2000 metri che si estendono sia in Italia, sia in Germania, sia in Svizzera [VINCOLO: Non è possibile usare il join tra tabelle nella clausola **FROM**].

d) (3 pt) Scrivere in **SQL** la query che restituisce nome e superficie degli stati attraversati dalla catena della "Alpi Carniche" ma non dalla catena delle "Alpi Giulie".

e) (3 pt) Scrivere in **SQL** la query che restituisce i continenti su cui si estendono almeno 5 catene di lunghezza pari a 1000 km.

f) (1 pt) Assumendo che CATENAM sia una collezione in **MONGODB**, scrivere la query che restituisce i nomi delle catene (solo questo campo) con lunghezza inferiore ai 300km ed altitudine superiore ai 2000 metri.

ESERCIZIO 2 (12 punti)

Si vuole progettare una base di dati per un sistema di gestione delle edizioni del Giro d'Italia di ciclismo. Per ogni edizione, si vuole tenere traccia dell'anno di svolgimento (si svolge esattamente 1 edizione ogni anno), del nome/cognome dell'organizzatore, del numero complessivo dei km percorsi. Inoltre, si vogliono gestire le informazioni delle squadre partecipanti: ogni squadra dispone di un nome (univoco), un anno di fondazione, una sede, ed è composta da una lista di ciclisti. Ogni ciclista dispone di un codice (univoco), nome, cognome, anno di nascita e può avere una galleria di uno o più foto associate. Ogni edizione del Giro è composta da una serie di tappe, ciascuna caratterizzata da un numero progressivo (univoco all'interno di un'edizione ma non a livello globale), una località di partenza, una località di arrivo, un lunghezza in km. Per ogni località (di partenza o di arrivo) si vuole tenere traccia del nome, della regione, del numero di abitanti. Le tappe possono appartenere a due categorie (e solo a quelle): tappe in linea o cronosquadre. Delle tappe in linea si vogliono conoscere la descrizione e la difficoltà. Per le cronosquadre, si vuole tenere traccia della percentuale di km in pianura. Inoltre, si vuole tenere traccia del vincitore di ogni tappa: nel caso delle tappe in linea, il vincitore è un ciclista, nel caso delle cronosquadre, il vincitore è una squadra. Infine, si vogliono gestire le classifiche associate ad ogni edizione: ogni classifica dispone di un nome (es. maglia rosa/maglia verde/etc) ed un regolamento. Si vuole tenere traccia del posizionamento dei ciclisti nelle classifiche disponibili.

a) (6pt) Costruire il modello Entità-Relazione (E-R) della base di dati.

b) (4pt) Tradurre il modello E-R nel modello logico relazionale, preferendo (NELLA TRADUZIONE DI EVENTUALI GENERALIZZAZIONI, E SOLO IN QUESTE) la soluzione che minimizzi la presenza di valori NULL nelle tabelle generate. Indicare i vincoli di integrità referenziale tra gli attributi dello schema.

c) (2pt) Il campo "numero complessivo dei km" di un'edizione è una **ridondanza concettuale**, in quanto l'informazione può essere ottenuta anche sommando la lunghezza delle singole tappe. Valutare il costo, con o senza la ridondanza, delle seguenti operazioni:

- Aggiungere una tappa ad un'edizione esistente (Batch, 1 volte/mese).
- Visualizzare tutte le tappe di un'edizione esistente (Interattivo, 5 volte volte/mese).

Tabella dei volumi: 20 tappe per edizione. α (peso operazioni scrittura)=2, w_I (peso operazioni interattive)=1, w_B (peso operazioni batch)=0.5

ESERCIZIO 3 (4 punti)

Data la tabella:

FESTIVAL(NomeCantante, Durata, TitoloCanzone, NumeroPartecipazioni,
AnnoEdizione, NomePresentatore, Città, NomeTeatro, NumeroVittorie)

con le seguenti dipendenze funzionali:

AnnoEdizione → NomePresentatore Città
NomeCantante → NumeroPartecipazioni
TitoloCanzone → Durata
Città → NomeTeatro
NomeCantante → NumeroVittorie
AnnoEdizione NomeCantante → TitoloCanzone

a) (2 pt) Indicare una **superchiave** ed una **chiave** dello schema. **Giustificare la risposta.**

b) (2 pt) La relazione è in forma normale di Boyce e Codd (FNBC)? La relazione è in terza forma normale (3FN)? La relazione è in seconda forma normale (2FN)? **Giustificare le risposte.**

ESERCIZIO 4 (2 punti)

a) (1 pt) Indicare che tipo di problemi risolve l'algoritmo di Naive Bayes (NB), e descriverne il funzionamento (in sintesi).

b) b) (1 pt)) Dato il seguente stato dei log:

CK(T₀, T₁) B(T₂) B(T₃) D(T₂, O₆, B₆) C(T₀) CK(T₁, T₂, T₃) C(T₃) U(T₁, O₀, B₀, A₀) CK(T₁, T₂)
B(T₄) B(T₅) I(T₄, O₄, A₄) D(T₅, O₅, B₅) C(T₄) C(T₁) B(T₆) I(T₅, O₅, A₅) C(T₅) **guasto**

Applicando l'algoritmo di ripresa a caldo, indicare la composizione degli insiemi UNDO e REDO (ossia, quali sono le transazioni di cui occorre fare l'*undo* e di quali il *redo*).