

Questions

1) Cosa si intende per database?

Collezione di dati organizzata in modo logico e coerente per la gestione delle operazioni CRUD (Create, Read, Update, Delete) si tratta di operazioni di creazione/inserimento, lettura, modifica o aggiornamento, eliminazione di dati.

2) Cos'è un DBMS?

Acronimo che sta per Database Management System. È un sistema software che consente di gestire database tramite CRUD, DDL (data definition language), DML (data manipulation language), DCL (data control language).

3) Indica le principali clausole di uno statement SELECT in ordine di esecuzione logica. Descrivi per ciascuna delle clausole indicate la logica di funzionamento.

FROM consente di specificare la tabella da interrogare. Al suo interno possono essere presenti join, che combinano record appartenenti a tabelle diverse, query innestata, union e view.

WHERE → consente di filtrare le righe. Può essere definito come una (o più) condizione che filtra la ricerca

GROUP BY → consente di eseguire dei raggruppamenti o aggregazioni per ogni combinazione univoca dei campi indicati nella group by list.

HAVING → filtro sui risultati del group by. Possono essere presenti una o più condizioni di ricerca

SELECT → consente di indicare quali colonne devono essere restituite. Può contenere costanti, query innestate, operazioni su colonne.

ORDER → consente di fare ordinamenti.

4) Descrivi, immaginando uno scenario a te familiare, il concetto di group by. Utilizza l'approccio che ritieni più efficiente per trasmettere il concetto (suggerimento: disegna anche una sola tabella in Excel o in word con poche colonne e pochi record e descrivi, basandosi sulla tabella stessa, un esempio di group by).

Il GROUP BY consente di eseguire dei raggruppamenti o aggregazioni per ogni combinazione univoca dei campi indicati nella GROUP BY LIST. L' GROUP BY raggruppa le righe che hanno gli stessi valori in righe.

Si propone un esempio esplicativo.

Si propone un database in cui sono registrate le vendite in Sicilia di prodotti per la cura del corpo. Il database riporta la tabella SALES_Sicily che consta di tre colonne: Product_ID, TotalAmount e la Location, come mostrato in figura.

SALES_Sicily		
product_ID	TotalAmount	Location
1	5,99	Palermo
2	6,25	Trapani
3	3,56	Catania
4	5,36	Palermo
5	7,5	Palermo
6	9,6	Catania
7	3,5	Trapani

Si vuole conoscere il totalAmount per Città siciliana al fine di monitorare l'andamento delle vendite. Si applica la seguente query su SQL

```

SELECT
    Location,
    SUM(TotalAmount)
FROM
    Sales_Sicily
GROUP BY
    Location;

```

si ottiene il seguente risultato

Tabella restituita dalla query	
Location	TotalAmount_City
Palermo	18,85
Trapani	9,75
Catani	13,16

5) Descrivi la differenza tra uno schema OLTP e uno schema OLAP.

OLTP: online transaction processing → il cui scopo è garantire la consistenza e la sicurezza delle transazioni stesse. È possibile effettuare manipolazione, elaborazione del dato, gestione di operazioni/transazioni (CRUD). Ha una struttura normalizzata che funge da garante per l'integrità e l'accuratezza del dato riducendo al minimo la ridondanza. OLTP è normalizzato in modo da minimizzare la ridondanza di dati e garantire l'integrità referenziale dei dati.

OLAP: analisi dei dati a supporto di report e visualizzazione a valle dell'ETL (estrazione, trasformazione e caricamento)

6) Dato un medesimo scenario di analisi, qual è la differenza in termini di risultato ottenibile tra una join e una subquery?

La differenza è che la join è tendenzialmente più pesante. Nella join si possono richiedere nella SELECT le colonne sia di un'ipotetica tabella A che B, mentre nel caso di una sub query non possiamo selezionare colonne della tabella B ma può essere usata come filtro.

7) Cosa si intende per DML e DDL?

Si tratta di istruzioni utilizzate in istruzioni SQL. Le istruzioni possono essere classificate in

- DDL (Data Definition Language) che comprende istruzioni utili a creare, modificare ed eliminare oggetti.
- DML (Data Manipulation Definition) comprende istruzioni per interrogare e modificare i dati all'interno di tabelle.

8) Quali istruzioni possono utilizzare per estrarre l'anno da un campo data? Proponi degli esempi.

L'istruzione per estrarre l'anno da un campo data è YEAR.

Viene usato solo per memorizzare l'anno, occupa 1 byte e il formato restituito è del tipo 'YYYY'.

Si propone di seguito l'esempio di una tabella definita "Matricola" che contiene tutte le informazioni di studenti universitari. Nella tabella è presente la colonna "data_immatricolazione". Se volessi analizzare l'anno di immatricolazione potrei utilizzare la funzione YEAR.

```

SELECT
  Id_matricola,
    YEAR (data_immatricolazione) Anno_immatricolazione,
FROM
  Matricola

```

tabella MATRICOLA che inserisco nella FROM

id_matricola	Data_immatricolazione
0785	12-10-2020
0953	5-9-2018
...	...

tabella restituita.

id_matricola	Anno_immatricolazione
0785	2020
0953	2018
...	...

9) Qual è la differenza tra gli operatori logici AND e OR?

Sia AND che OR sono due operatori logici che vengono utilizzati all'intero di query nella WHERE o nell'HAVING. L'operatore logico AND restituisce TRUE se entrambe le espressioni booleane sono soddisfatte. L'operatore OR restituisce TRUE se almeno una delle espressioni booleane identificate viene soddisfatta.

10) È possibile innestare una query nella clausola SELECT?

Si è possibile inserire una query nella SELECT. Di seguito si propone un esempio su una tabella che abbiamo informazioni su studenti immatricolati universitari. Nella SELECT vi è una subquery in cui viene calcolato gli anni che gli studenti impiegano per finire il corso, calcolato come valore medio.

```

SELECT
  id_studente,
  nome,
  anni_studio,
  (SELECT AVG(anni_studio) FROM matricola) AS media_anni_studio
FROM matricola;

```

la media verrà effettuata su tutta la colonna "anni_studio".

Tabella originaria MATRICOLA

id_studente	Nome	Anni_studio
0785	Mario Rossi	3
0784	Enrico Verdi	5
0584	Luigi Bianchi	4
...

tabella restituita

id_studente	Nome	Anni_studio	media_anni_studio
0785	Mario Rossi	3	4
0784	Enrico Verdi	5	4
0584	Luigi Bianchi	4	4
...

11) Qual è la differenza tra l'operatore logico OR e l'operatore logico IN?

Si tratta di due operatori logici.

OR restituisce TRUE se una delle espressioni booleane è verificata. Confronta due elementi

IN vuole in ingresso una lista, essa restituisce TRUE se l'operatore è uguale a una delle espressioni elencate.

Confronta campi di un elenco.

L'IN equivale ad eseguire tanti OR quanti solo i valori inseriti nella lista.

Si propone un esempio esplicativo in cui l'utilizzo di IN e OR restituisce lo stesso risultato.

<pre>SELECT Id_matricola YEAR(anno_immatricolazione) Count(*) AS Immatricolati_nel_covid FROM Matricola WHERE YEAR(anno_immatricolazione) ='2020' OR '2021' OR '2022'</pre>	<pre>SELECT Id_matricola YEAR (anno immatricolazione) Count (*) AS Immatricolati_nel_covid FROM Matricola WHERE YEAR (anno immatricolazione) IN ('2020', '2021' , '2022')</pre>
---	---

12) L'operatore logico BETWEEN include anche gli estremi del range specificato?

Si. Un esempio di applicazione

```
SELECT *
FROM matricola
WHERE anni_studio BETWEEN 3 AND 5;
```

Se volessi visualizzare gli anni di studio tra 3 e 5 anni escludendo gli estremi di questo intervallo allora la query andrebbe scritta in questo modo

```
SELECT *
FROM matricola
WHERE anni_studio > 3 AND anni_studio < 5
```

CASE STUDY

ToysGroup è un'azienda che distribuisce articoli (giocattoli) in diverse aree geografiche del mondo. I prodotti sono classificati in categorie e i mercati di riferimento dell'azienda sono classificati in regioni di vendita. In particolare:

1) Le entità individuabili in questo scenario sono le seguenti:

- Product
- Region
- Sales

2) Le relazioni tra le entità possono essere descritte nel modo seguente:

- Product e Sales → Un prodotto può essere venduto tante volte (o nessuna) per cui è contenuto in una o più transazioni di vendita. Ciascuna transazione di vendita è riferita ad uno solo prodotto
- Region e Sales → Possono esserci molte o nessuna transazione per ciascuna regione. Ciascuna transazione di vendita è riferita ad una sola regione

3) Le entità Product e Region presentano delle gerarchie:

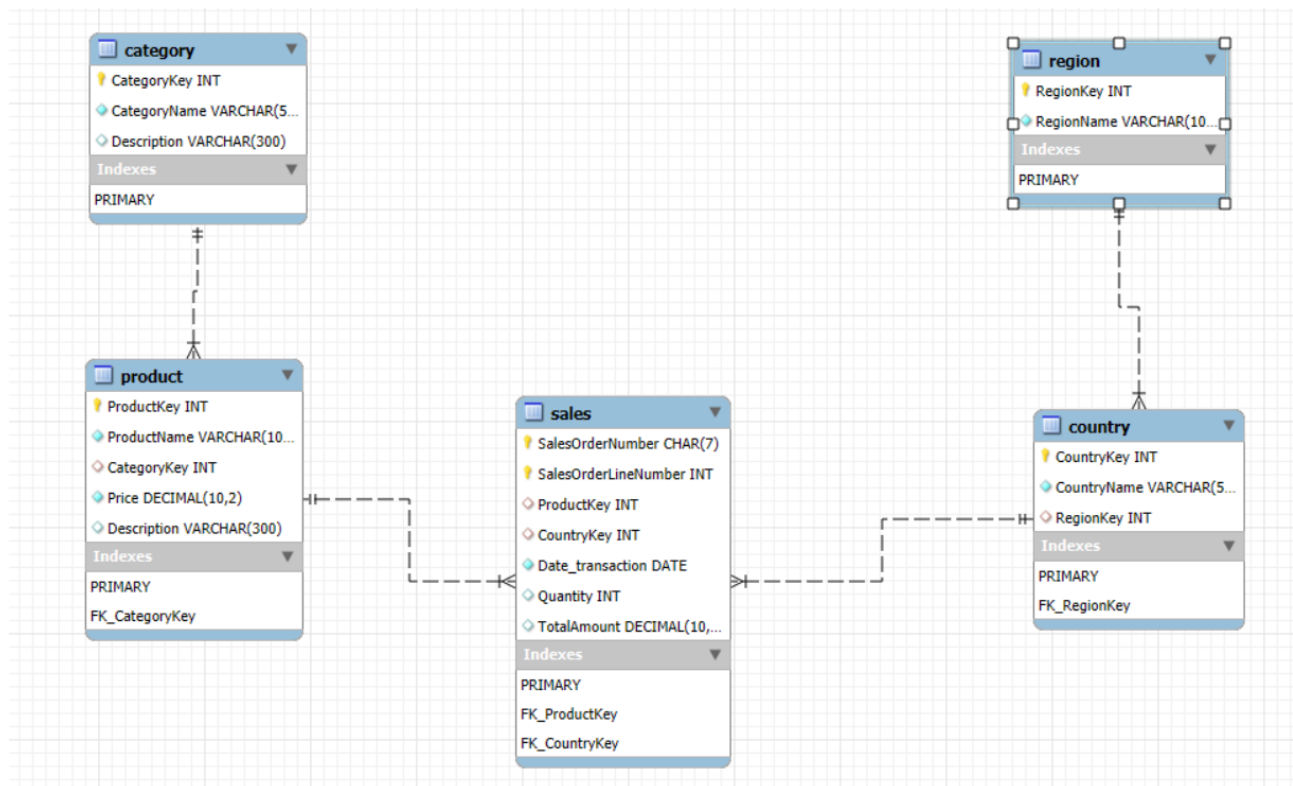
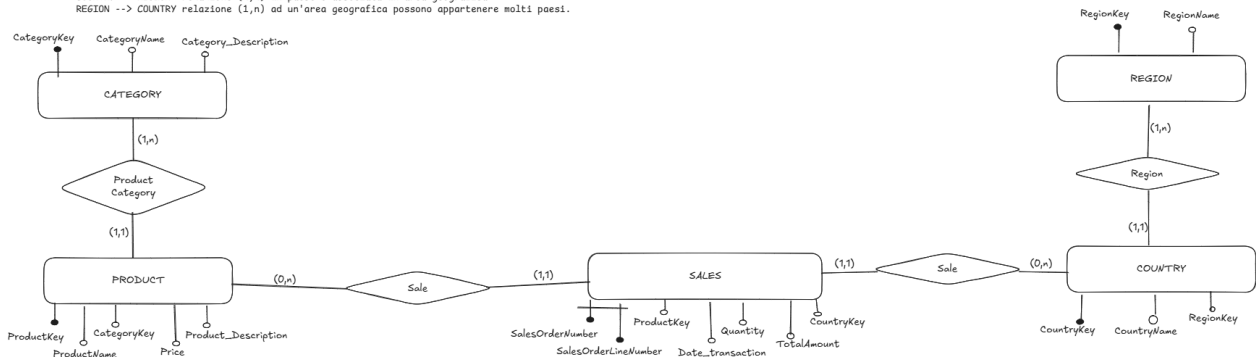
- L'entità **prodotto** contiene, oltre alle informazioni del singolo prodotto, anche la descrizione della categoria di appartenenza. L'entità prodotto contiene quindi una gerarchia: un prodotto può appartenere ad una sola categoria mentre la stessa categoria può essere associata a molti prodotti diversi. Esempio: gli articoli 'Bikes-100' e 'Bikes-200' appartengono alla categoria Bikes; gli articoli 'Bike Glove M' e 'Bike Gloves L' sono classificati come Clothing.
- L'entità **regione** contiene una gerarchia: più stati sono classificati in una stessa regione di vendita e una stessa regione di vendita include molti stati. Esempio: gli stati 'France' e 'Germany' sono classificati nella region WestEurope; gli stati 'Italy' e 'Greece' sono classificati nel mercato SouthEurope. È necessario progettare e implementare fisicamente un database che modelli lo scenario garantendo l'integrità referenziale e la minimizzazione della ridondanza dei dati. In altre parole, progetta opportunamente un numero di tabelle e di relazioni tra queste sufficiente a garantire la consistenza del dato.

TASK 1

Task 1: Proponi una progettazione concettuale e logica della base dati La progettazione concettuale deve includere tutte le entità coinvolte e le relazioni tra queste. Per ciascuna entità indica l'attributo chiave e i principali attributi descrittivi (non è necessario indicare tutti gli attributi). Esempio di schema E/R

Di seguito sono riportati le due immagini relative alle due progettazioni sviluppate rispettivamente con Excalidraw e con il Reverse Engineer di MySQL.

SALES --> PRODUCT relazione (1,1) ad una transazione di vendita è associato un solo prodotto
 PRODUCT --> SALES relazione (0,n) potrebbero esserci dei prodotti mai venduti che quindi non compaiono in nessun record della tabella Sales. Al massimo un prodotto può essere venduto molte volte.
 PRODUCT --> CATEGORY relazione (1,1) ad un prodotto è associata una sola categoria.
 CATEGORY --> PRODUCT relazione (1,n) ad una categoria possono essere associati molti prodotti diversi.
 SALES --> COUNTRY relazione (1,1). Ad una transazione di vendita viene associato un paese in cui è venduto il prodotto
 COUNTRY --> SALES relazione (0,n) ci possono essere paesi che non hanno transazione di vendita, cioè non hanno mai comprato quei prodotti. Ci possono essere al più, un paese che ha acquistato tante volte, quindi che è presente in tanti record della tabella Sales
 COUNTRY --> REGION relazione (1,1) un paese è associato un'area geografica
 REGION --> COUNTRY relazione (1,n) ad un'area geografica possono appartenere molti paesi.



Task 2: Descrivi la struttura delle tabelle che reputi utili e sufficienti a modellare lo scenario proposto tramite la sintassi **DDL**. Implementa fisicamente le tabelle utilizzando il DBMS SQL Server (o altro). [Implementata in MySQL]

Task 3: Popola le tabelle utilizzando dati a tua discrezione (sono sufficienti pochi record per tabella; riporta le query utilizzate) [Implementata in MySQL]

Task 4: [Implementata in MySQL]