**Questions**

1. Cosa si intende per database?
2. Cos’è un DBMS?
3. Indica le principali clausole di uno statement SELECT in ordine di esecuzione logica. Descrivi per ciascuna delle clausole indicate la logica di funzionamento.
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).
5. Descrivi la differenza tra uno schema OLTP e uno schema OLAP.
6. Dato un medesimo scenario di analisi, qual è la differenza in termini di risultato ottenibile tra una join e una subquery?
7. Cosa si intende per DML e DDL?
8. Quali istruzioni possono utilizzare per estrarre l’anno da un campo data? Proponi degli esempi.
9. Qual è la differenza tra gli operatori logici AND e OR?
10. È possibile innestare una query nella clausola SELECT?
11. Qual è la differenza tra l’operatore logico OR e l’operatore logico IN?
12. L’operatore logico BETWEEN include anche gli estremi del range specificato?

1) Cosa si intende per database?

Un database è un insieme di dati strutturati, organizzati in tabelle, le quali contengono righe e colonne. Questa struttura permette che possano essere facilmente gestiti, aggiornati e consultati, permettendo un accesso e una manipolazione efficiente;

2) Cos’è un DBMS?

Un Database Management System (DBMS) è un software che consente di creare, gestire e manipolare database. Il DBMS fornisce strumenti per l'inserimento, la modifica, la cancellazione e la ricerca dei dati, assicurando integrità, sicurezza e consistenza delle informazioni. Il tipo di DBMS piu utilizzato è quello relazionale, dove I dati sono organizzati in tabelle;

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

Ordine di esecuzione logica:

FROM: Specifica la tabella o le tabelle (in caso di join) da cui vengono interrogati i dati;

WHERE: Filtra le righe restituite dalla clausula FROM in base alle condizioni di ricerca;

GROUP BY: Raggruppa le righe che hanno gli stessi valori in determinate colonne;

HAVING: Filtra i gruppi creati dal GROUP BY in base a condizioni specifiche;

SELECT: Specifica le colonne o le espressioni da restituire nel result set, ovvero nell’output della query;

ORDER BY: Consente di applicare ordinamenti. Ordina i risultati in base a una o piu colonne;

LIMIT: Limita il numero di righe nell'output della query a un numero specificato

4) Descrivi, immaginando uno scenario a te familiare, il concetto di group by. Immagina di avere una tabella Excel con le vendite di vari prodotti in diversi negozi. La tabella contiene colonne come "Prodotto", "Negozio" e "Vendite". Utilizzando GROUP BY, puoi raggruppare le vendite per prodotto, ottenendo così il totale delle vendite per ciascun prodotto. Esempio: Raggruppando per "Prodotto" ottieni:

Ponendo il caso di avere 4 prodotti, ripetuti perchè venduti piu volte da diversi negozi, raggruppando per prodotto otterrò un result set in cui i 4 prodotti sono indicati in modo distinto nella prima Colonna e affianco nella seconda Colonna avrò il totale delle vendite per ciascun prodotto

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

Online Transaction Processing (OLTP): È progettato per gestire operazioni transazionali rapide e ripetitive come inserimenti, aggiornamenti e cancellazioni. Le tabelle sono normalizzate per ridurre la ridondanza. Utilizzato per scenari con intense attività di scrittura, quindi per transazioni/operazioni CRUD brevi ma frequenti;

Online Analytical Processing (OLAP): È ottimizzato per le query analitiche e l'estrazione di dati. Le tabelle sono spesso denormalizzate per migliorare le prestazioni delle query complesse e dei report. Utilizzato per l’analisi complessa di grandi volume di dati;

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

La Join andrebbe a combinare dei record da due o piu tabelle, basandosi su una condizione comune (chiavi primarie e esterne). Il result set conterrebbe colonne di entrambe le tabelle in un’unica visualizzazione;

La Subquery è una query presente all’interno di un’altra query, con lo scopo di restituire uno o piu valori da utilizzare nella query principale. Usata per calcolare valori intermedi (conte, medie, minimi, massimi, somme o valori maggiori/minori);

7) Cosa si intende per DML e DDL?

Data Manipulation Language (DML) è utilizzato per gestire e manipolare I dati all’interno di un database; le istruzioni DML permettono la manipolazione dei dati, come l’inserimento (INSERT) , l’aggiornamento (UPDATE) , la cancellazione (DELETE ) e il recupero (SELECT ) dei dati;

Data Definition Language (DDL) invece è utilizzato per definire (CREATE) e gestire (ALTER e DROP); le istruzioni DDL consentono di creare, modificare e cancellare oggetti del database come tabelle, viste e schemi;

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

Per estrarre l’anno ad esempio da una colonna data (DataOrdine): SELECT YEAR(DataOrdine) AS Anno FROM TabellaOrdini); o ad esempio, raggruppare per anno calcolando le vendite fatte per ogni anno: SELECT YEAR(DataOrdine) as Anno, SUM(Vendite) AS TotaleVendite FROM TabellaOrdini GROUP BY Anno;

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

Con AND, qualora entrambe le condizioni booleane siano soddisfatte restituirebbe TRUE. Con OR, a differenza di AND, basterebbe che almeno una delle condizioni booleane fosse soddisfatta e restituirebbe TRUE;

10) È possibile innestare una query nella clausola SELECT?

Come in parte detto per la domanda numero 6, si è possibile, utilizzando una Subquery, ad esempio se volessi calcolare il Prezzo medio dei prodotti: SELECT NomeProdotto, (SELECT AVG(Prezzo) FROM Prezzi WHERE Prezzi.ProdottoID = Prodotti.ProdottoID) AS PrezzoMedio FROM Prodotti;

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

Con OR, basterebbe che almeno una delle condizioni booleane fosse soddisfatta e restituirebbe TRUE, ed è usato per specificare più condizioni singole che devono essere valutate separatamente. Con IN , restituirebbe TRUE se il valore di una colonna è uguale a uno qualsiasi dei valori presenti in un elenco specificato ed è usato per specificare un elenco di valori da confrontare con una colonna specifica;

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

Sì, l'operatore BETWEEN include sempre gli estremi del range. Ad esempio: BETWEEN 1 AND 10 include sia 1 che 10. In questo caso, se proprio volessi escludere 1 e 10 dal range, dovrei aggiungere al WHERE nomecolonna BETWEEN 1 AND 10; un: AND nomecolonna NOT IN (1,10);

**Case Study**

ToysGroup è un’azienda che distribuisce articoli (giocatoli) 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
* Categories (opzionale) – Jasmine a lezione lo ha fatto notare

1. Le relazioni tra le entità possono essere descritte nel modo seguente:

* Product e Sales
* Un prodotto puo’ 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

1. 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 puo’ appartenere ad una sola categoria mentre la stessa categoria puo’ 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*.

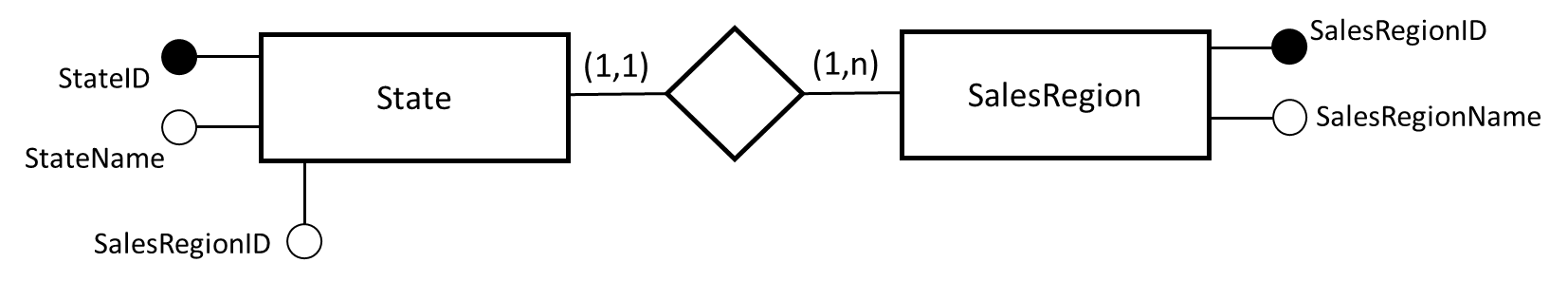
È necessario progettare e implementare fisicamente un database che modelli lo scenario garantendo l’**integrità referenziale** e la **minimizzazione della ridondanza dei dati.**

In altre parole, progetta opportunamente un numero di tabelle e di relazioni tra queste sufficiente a garantire la **consistenza del dato.**

**Task 1: 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*



*Lo schema proposto è puramente esemplificativo e non esaustivo o completo per la soluzione!*

La progettazione logica deve includere, per ciascuna tabella, tutte le colonne che poi verranno implementate fisicamente e deve esplicitare la cardinalità dei campi utilizzati per definire la relazione.

*Esempio di schema grafico delle tabelle e delle relazioni tra le stesse.*



n

1



*Il diagramma è puramente esemplificativo e non esaustivo o completo per la soluzione!*

**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).**

CREATE TABLE table\_name (

column1 datatype option,

column2 datatype option,

column3 datatype,

....

);

**Task 3: Popola le tabelle utilizzando dati a tua discrezione (sono sufficienti pochi record per tabella; riporta le query utilizzate)**

**Task 4: Dopo aver popolate le tabelle, scrivi delle query utili a:**

1. Verificare che i campi definiti come PK siano univoci. In altre parole, scrivi una query per determinare l’univocità dei valori di ciascuna PK (una query per tabella implementata).
2. Esporre l’elenco delle transazioni indicando nel result set il codice documento, la data, il nome del prodotto, la categoria del prodotto, il nome dello stato, il nome della regione di vendita e un campo booleano valorizzato in base alla condizione che siano passati più di 180 giorni dalla data vendita o meno (>180 -> True, <= 180 -> False)
3. Esporre l’elenco dei prodotti che hanno venduto, in totale, una quantità maggiore della media delle vendite realizzate nell’ultimo anno censito. (ogni valore della condizione deve risultare da una query e non deve essere inserito a mano). Nel result set devono comparire solo il codice prodotto e il totale venduto.
4. Esporre l’elenco dei soli prodotti venduti e per ognuno di questi il fatturato totale per anno.
5. Esporre il fatturato totale per stato per anno. Ordina il risultato per data e per fatturato decrescente.
6. Rispondere alla seguente domanda: qual è la categoria di articoli maggiormente richiesta dal mercato?
7. Rispondere alla seguente domanda: quali sono i prodotti invenduti? Proponi due approcci risolutivi differenti.
8. Creare una vista sui prodotti in modo tale da esporre una “versione denormalizzata” delle informazioni utili (codice prodotto, nome prodotto, nome categoria)
9. Creare una vista per le informazioni geografiche