**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?
13. Un database è una raccolta di dati organizzati in modo logico e coerente per facilitarne la manipolazione (interrogazione, inserimento, modifica, eliminazione).
14. Un DBMS , Database Management System, è un sistema software progettato per consentire la creazione, la manipolazione e lʼinterrogazione di database.

FROM: specifica la tabella da cui vengono selezionate le colonne

WHERE: determina una condizione per il quali I dati vengono filtrati

GROUP BY: Raggruppa le righe con valori identici in colonne specifiche e consente di applicare funzioni di aggregazione

HAVING: filtra i gruppi di dati in base a una condizione specifica dopo aver utilizzato GROUP BY

SELECT: restituisce le colonne selezionate

ORDER BY: raggruppa i dati in base a uno o più colonne

1. Il comando GROUP BY viene utilizzato per raggruppare i risultati di una query in base a una o più colonne. Ad esempio, immagina di avere una tabella vendite con le colonne prodotto, anno, e quantità\_venduta. Utilizzando GROUP BY, possiamo raggruppare le vendite per prodotto e calcolare la quantità totale venduta per ciascuno.

SELECT Product\_ID , Order\_Date , SUM(Sales) AS Total\_Sales FROM Sales GROUP BY Product\_ID , Order\_Date

|  |  |  |
| --- | --- | --- |
| Product\_ID | Order\_Date | Sales |
| 1 | 2024-01-01 | 100 |
| 1 | 2024-01-01 | 100 |
| 2 | 2024-01-01 | 100 |

|  |  |  |
| --- | --- | --- |
| Product\_ID | Order\_Date | Total\_Sales |
| 1 | 2024-01-01 | 200 |
| 2 | 2024-01-01 | 100 |

1. La gestione dei dati transazionali tramite sistemi informatici è chiamata OLTP (Online Transaction Processing). Si usa per gestire le transazione quotidiane dette CRUD, mantenedone l’integrità e la coerenza . Il Sistema OLAP serve per l’analisi e l’aggregazione di grandi volume di dati a support della BI. Un esempio di OLTP serve per fare una transazione finanziaria , mentre OLAP per fare un analisi dei dati di vendita.
2. Nello stesso scenario se usassi una JOIN combinerei due o più tabelle in base ad una condizione comune, quindi il risultato sarà un’unica tabella con valori collegati, se usassi una subquery, che è una query all'interno di un'altra query, il cui valore restituito viene utilizzato nella query principale, otterrei un'interrogazione con un filtro complesso
3. DML, Data Manipulation Language, comprende le istruzioni per interrogare e modificare i dati allʼinterno di tabelle ( esempio INSERT, UPDATE, SELECT, DELETE). DDL , Data Definition Language, comprende le istruzioni utili a definire (creare, modificare, eliminare) oggetti (tabelle per esempio con CREATE,ALTER,DROP).

8) SELECT YEAR(data) as Year FROM nome\_tabella;

SELECT YEAR(CURRDATE()) as CURR\_YEAR;

9) AND: Richiede che entrambe le condizioni siano vere.

OR: richiede che almeno una delle condizione sia vera.

10) si , è possibile

11) Con l’operatore OR la query restituisce tutti i record per cui è vera o lʼuna o lʼaltra condizione di ricerca, con lʼoperatore IN si ricerca la corrispondenza in una lista di valori specificati.

Esempio Select Nomi\_Colonne

From Nome\_Tabella

Where Nome\_Tabella IN (Corrispondenza1,Corrispondenza2,ect..)

Select Nomi\_Colonne

From Nome\_Tabella

Where Nome\_Tabella =Condizione1, OR =Condizione2, OR =Condizione 3 ect…

12) Sì, l'operatore logico BETWEEN include anche gli estremi del range specificato.

**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

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*.

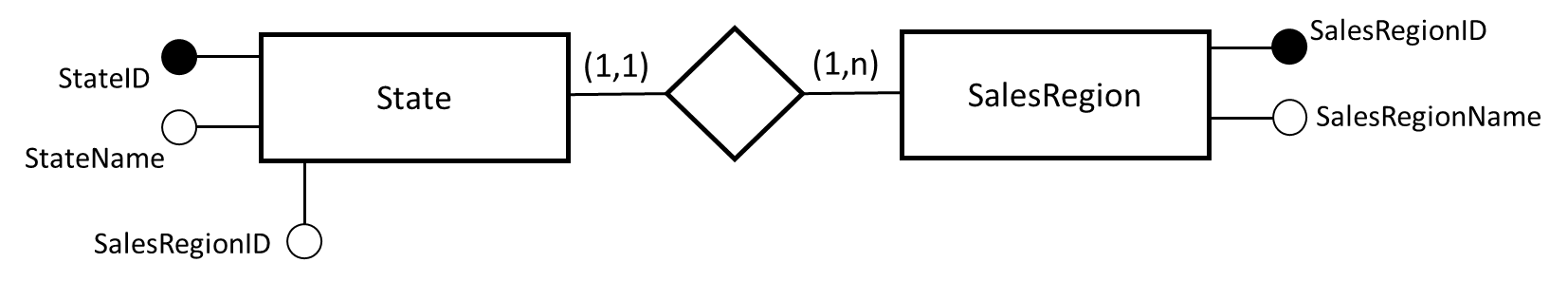
È necessario progettare e implementare fisicamente un database che modelli lo scenario garantendo l’**integrità referenziale** e la **minimizzazione della ridondanza dei dati.**

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

Task 2:

la tabella category è collegata alla product (1:N);

la tabella product è collegata alla category (N:1) e alla sales(1:N) ;

la tabella region è collegata alla costumer (1:N) e alla sales (1:N);

la tabella sales è collegata alla sales (N:1), alla region (N:1) e alla costumer (1:N);

la tabella costumer è collegata alla region (N:1) e alla sales (N:1).