W8D4

**QUESTIONS**

1. **Cosa si intende per database?**

Un DB è un contenitore digitale dove vengono immagazzinati dei dati, spesso in Tabelle. In ogni Tab abbiamo le colonne dette “Campi” e di conseguenza le righe dette anche “Record”.

1. **Cos’è un DBMS?**

Il Database Management System è un software che ci permette di gestire un DB. Un esempio è MySQL da noi usato in questo mese.

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

(1) “FROM”, stiamo dicendo da quale tabella o quali (JOIN) prendiamo questi dati;

(2) “WHERE”, stiamo dicendo che c’è una condizione per la quale scegliamo dei dati da restituire; (3) “GROUP BY”, stiamo dicendo di raggruppare le righe che hanno almeno un valore uguale in una Colonna o più;

(4) “HAVING”, ha la stessa funzione di Where, ma filtra solo I gruppi creati da “Group BY”; (5) “SELECT”, stiamo dicendo quali colonne usciranno in Output; (6) “DISTINCT”, stiamo dicendo di non mostrarci dei dati duplici dalle colonne nella “Select”; (7) “ORDER BY”, stiamo dicendo di ordinare una tab, in base ad una Colonna dichiarata; (8) “LIMIT/ OFFSET”, viene utilizzato per restituire un Tot di righe che abbiamo dichiarato, oppure per saltare un numero di righe che sempre abbiamo dichiarato.

1. **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).**

Abbiamo una tab con (Modello Auto) e (Casa Automobilistica). Noi vogliamo Vedere quante auto ha la nostra Concessionaria per Casa Automobilistica. Il Group By insieme all’operazione Count ci restituirà quante Auto ci sono per Casa Automobilistica.

|  |  |
| --- | --- |
| Modello Auto | Casa Automobilistica |
| Mustang | Ford |
| Focus St Line | Ford |
| Model X | Tesla |
| Stelvio | Alfa Romeo |
| Model S | Tesla |

ES. Ford – 2; Tesla – 2; Alfa Romeo – 1.

1. **Descrivi la differenza tra uno schema OLTP e uno schema OLAP.**

L’uso dello Schema OLTP e OLAP è diverso, a seconda di come vogliamo organizzare e utilizzare I dati. OLTP lo usiamo per gestire transazioni quotidiane, ad alta velocità di lettura e scrittura, strutture semplice per evitare dati ridondanti. OLAP invece lo usiamo più per strategie marketing, Business Intelligence, dove ci sono operazioni complesse come aggregazioni e confronti e ci sono Grandi volume di dati

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

Cambiano alcuni fattori di come vengono visualizzati i dati in Output.

Ad esempio. Se con una JOIN combiniamo delle tabelle e visualizzeremo dunque una sola tab combinata, con SUBQUERY i valori potrebbero essere utilizzati come filtro.

Rivedi (4) Con Join vedremo unite le tab di Modelli auto e Casa Automobilistica ( MUSTANG – FORD), con SUBQUERY potremmo chiedere in Output le Case Automobilistiche con almeno 2 auto presenti in Negozio.

1. **Cosa si intende per DML e DDL?**

Sono dei comandi usati per gestire un DB. Il DDL sono i comandi per creare, modificare e definire una struttura di un DB. Es. Create Database, Create Table …

Il DML invece comprende i comandi che gestiscono i dati presenti nelle tab. Es. Select, Insert into…

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

Possiamo usare la funzione: Select Year(colonna) From (Tab)

1. **Qual è la differenza tra gli operatori logici AND e OR?**

And restituisce “Vero” solo se TUTTE le condizioni sono vere.

OR restituisce “Vero” se almeno UNA delle condizioni è vera.

1. **È possibile innestare una query nella clausola SELECT?**

Si, questo tipo di Query viene chiamata SUBQUERY Scalare

1. **Qual è la differenza tra l’operatore logico OR e l’operatore logico IN?**

Principalmente hanno entrambe la stessa funzione, ma con la differenza della sintassi, in quanto si scrivono in maniera diversa. Una differenza che si presenta è che con Or possiamo confrontare Colonne diverse, mentre con IN noi indichiamo una sola colonna.

1. **L’operatore logico BETWEEN include anche gli estremi del range specificato?**

Si, include anche gli estremi espressi

**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*

Immagine che contiene nero, oscurità

Il contenuto generato dall'IA potrebbe non essere corretto.

*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