**Questions**

1. Cosa si intende per database?

Per Database si intende una collezione di dati organizzata in modo logico e coerente per la gestione delle operazioni CRUD, ovvero le operazioni che si svolgono giornalmente all’interno di un database aziendale. Operazioni di creazione, lettura, modifica e eliminazione di record.

1. Cos’è un DBMS?

Per poter gestire un database transrelazionale ovvero un database relazionale dove vengono eseguite delle transazioni, bisogna utilizzare un software chiamato DBMS che ci consente la creazione e la manipolazione dei dati e interrogazione del database.

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

FROM – Indichi le tabelle all’interno del database selezionato da dove prendere le informazioni.

JOIN – Il Join è uno strumento che permette a 2 o più tabelle che hanno relazioni tra di esse(primary key e foreign key) di integrarsi. L’integrazione varia in base al tipo di Join utilizzato(Inner, Left, right, Outer) .

WHERE – Difatti è un filtro che puoi applicare prima di raggruppare i dati

GROUP BY - Raggruppa i dati secondo un criterio(attributo) che tu indichi

HAVING - Filtra i dati dopo il raggruppamento.

SELECT – Indichi gli attributi da restituire.

ORDER BY - Ordina i dati ottenuti in base al criterio da te indicato crescente o decrescente.

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).

Ipotizziamo di avere un result set tabella con 2 attributi differenti, IDProdotto, ImportoOrdine. Quindi abbiamo una tabella con diverse transazioni di vendita di più prodotti, se volessimo sapere ogni prodotto quanto fattura basterebbe raggruppare i dati (group by) per singolo prodotto (Order By IDProdotto) e vedere a quanto ammontano gli importi delle transazioni.

|  |  |
| --- | --- |
| ID prodotto | ImportoOrdine |
| 1 | 10 |
| 2 | 12 |
| 1 | 15 |

Se raggruppo per ID prodotto avrò

|  |  |
| --- | --- |
| ID prodotto | ImportoOrdine |
| 1 | 25 |
| 2 | 12 |

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

L’OLTP ovvero l’Online Transaction Processing come definisce la parola stessa “elaborazione online delle tranasazioni” ovvero si occupa di inserire, modificare ed eliminare tutte quelle transazioni (CRUD) generate e garantire lʼintegrità referenziale dei dati e la sicurezza delle transazioni stesse. Un sistema OLTP è ottimizzato per la gestione delle transazioni e non per l’analisi dei dati.

L’OLAP ovvero l’Online Analytical Processing si occupa di elaborazione analitica dei dati(in grandi volumi) a supporto della business intelligence. Per poter avere operazioni OLAP vi deve prima essere il data warehouse ovvero una riorganizzazione dei dati esistenti ottimizzati per l’analisi dei dati. Le query presentano tipicamente operazioni di aggregazione come SUM, AVG, COUNT e GROUP by.

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

Ritengo che se le query sono scritte entrambe correttamente allora le operazioni possano arrivare allo stesso risultato. La subquery sono molto utilii quando si richiedono la semplificazioni di alcune operazioni, le stesse che facendole tramite join risultano complesse. In generale quando è complesso fare operazioni where o l’having con la join si prova con la subquery per scomporre il problema in 2 parti e semplificarlo.

1. Cosa si intende per DML e DDL?

Sono 2 categorie di domande di Sql, gli acronimi stanno per Data Manipulation Language e Data Definition Language. Il DML come dice la stessa parola si occupa di manipolare i dati quindi si intendono operazione di insert, update, delete. Mentre il DDL come ci fa intuire la parola stessa si occupa di definire le strutture dei dati all’interno del Database tramite operazioni di Create, Alter, Drop. Quindi da un lato abbiamo operazioni di definizione delle strutture dati e dall’altro operazioni di manipolazione di dati.

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

SELECT YEAR(campodata)

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

L’operatore logico AND restituisce il valore “VERO” se entrambe le operazioni indicate sono vere viceversa restituisce “FALSO”.

L’operatore “OR” restituisce il valore “VERO” se una delle 2 operazioni indicate sono vere. Se entrambe sono vere o false restituisce “FALSO”.

1. È possibile innestare una query nella clausola SELECT?

Si, si chiamano subquery

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

L’operatore “OR” restituisce il valore “VERO” se una delle 2 operazioni indicate sono vere. Se entrambe sono vere o false restituisce “FALSO”.

Mentre l’operatore IN restituisce il valore “VERO” se il valore che si sta cercando è all’interno dei valori indicati nell’operatore IN.

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

Si, include anche gli estremi.

**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.