**Questions**

1. Cosa si intende per database?

Per database si intende una collezione di dati, un oggetto che ha memoria e capacità di organizzare i dati secondo un criterio logico. I database si suddividono in **relazionali (SQL),** ovvero database strutturati in tabelle collegate tra loro mediante relazioni e **non relazionali (NoSQL)**.

1. Cos’è un DBMS?

**DBMS** acronimo di **database management system** è un software progettato per consentire operazioni di creazione, manipolazione e interrogazione di un database in modo efficiente e sicuro. Nel caso di modelli relazionali il sistema di gestione delle basi dati è detto RDBMS (Relational Database Management System).

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

Le principali clausole di uno statement SELECT sono in ordine di esecuzione logica:

1. **FROM** : che consente di indicare la tabella o l’unione delle tabella dalle quali originano i dati che si vuole interrogare.
2. **WHERE :**  consente di filtrare i record restituiti secondo uno o più condizioni logiche
3. **GROUP BY :** consente di raggruppare i record restituiti per ogni combinazione univoca dei campi indicati nella group by list, al fine di applicare un funzione di aggregazione
4. **HAVING** : consente di filtrare il risultato del raggruppamento secondo uno o più criteri di ricerca
5. **SELECT** : consente di visualizzare i campi richiesti nella select list
6. **ORDER BY** : consente di applicare un ordinamento ai record restituiti dalla select
7. 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).

|  |  |  |
| --- | --- | --- |
| **Nome** | **Sesso** | **Skipass Acquistati** |
| Luca | M | 5 |
| Antonio | M | 3 |
| Laura | F | 10 |
| Francesca | F | 2 |
| Federico | M | 11 |

SELECT Sesso, SUM(Skipass Acquistati)

FROM [nome\_tabella]

GROUP BY Sesso;

|  |  |
| --- | --- |
| **Sesso** | **Tot Skipass Acquistati** |
| M | 19 |

Nello scenario proposto abbiamo diversi sciatori che comprano degli skipass durante una stagione sciistica. Voglio estrapolare il numero totale di skipass acquistati da un solo sesso. In questo caso GROUP BY è utilizzato per aggregare i dati secondo il campo sesso per poi applicarvi una funzione di aggregazione quale la somma.

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

**OLTP** è l’acronimo di Online Transactional Processing ed indica la gestione dei dati transazionali generati da operazioni CRUD. Si utilizza in casi di intensa attività di scrittura, quindi transazioni CRUD brevi ma frequenti. Diversamente **OLAP** (Online Analytical Processing) è utilizzato per analisi complesse su grandi volumi di dati a supporto della business intelligence, e costituisce il cuore della data warehouse.

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

Una join permette di unire più tabelle e quindi di selezionare e visulizzare campi appartenenti a tabelle differenti. Una subquery consiste invece in una query innestata all’interno di un'altra ed è lo scopo di estrapolare record da un result set proveniente da un’altra query.

1. Cosa si intende per DML e DDL?

DDL sta per data definition language e comprende tutte le istruzioni utili a definire oggetti, quindi creare, modificare o eliminare tabelle.

DML sta invece per data manipulation language e comprende tutte le istruzioni utili ad interrogare, modificare i dati all’interno delle tabelle.

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

SELECT YEAR([campo\_data]) FROM [nome\_tabella]

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

Un’espressione composta da due espressioni unite da un operatore AND risulta vera se entrambe le espressioni sono vere. Mentre un’espressione costituita da due espressioni unite da un operatore OR risulta vera se almeno una delle due espressioni è vera.

1. È possibile innestare una query nella clausola SELECT?

Si è possibile

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

OR serve ad individuare se almeno una delle due condizioni è verificata, IN serve ad individuare se un elemento fa parte di una lista di elementi.

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

Si

**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*.

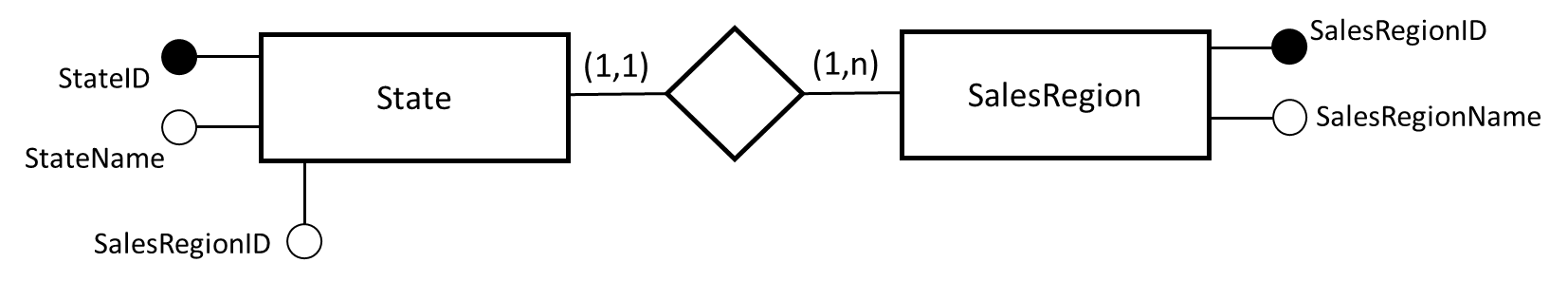
È necessario progettare e implementare fisicamente un database che modelli lo scenario garantendo l’**integrità referenziale** e la **minimizzazione della ridondanza dei dati.**

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