

Esercizio 1 : Qual è la funzione principale delle chiavi surrogate (es. 'cliente_sk') in uno schema a stella?

Esercizio 2: Quale strategia di estrazione dati (Extract) è più adatta per aggiornamenti frequenti su grandi volumi di dati, minimizzando il carico sul sistema sorgente?

Esercizio 3: Quali sono le differenze pratiche tra SCD Type 1 e Type 2?

Esercizio 4: Feature Engineering: Che cos'è l'analisi **RFM** e quali sono le tre metriche che la compongono?

Esercizio 5: Come si gestiscono i dati mancanti prima del Machine Learning?

Esercizio 6: Come si collegano i database OLTP e DW?

Esercizio7: Qual è uno dei principali svantaggi dell'utilizzo di Microsoft Excel per attività di Business Intelligence,

Esercizio 8: Scenario: Il Data Warehouse "TechStore"

Stai lavorando come Data Analyst per TechStore. Il database è strutturato secondo uno **Star Schema** con le seguenti tabelle chiave:

1. **fact_vendite**: Contiene le metriche di vendita. Colonne chiave: prodotto_sk (FK), data_sk (FK), ricavi, quantita.
 2. **dim_prodotto**: Contiene i dettagli dei prodotti. Colonne chiave: prodotto_sk (PK), categoria, nome_prodotto.
 3. **dim_data**: Contiene i dettagli temporali. Colonne chiave: data_sk (PK), anno, mese.
-

Esercizio: Analisi delle Performance per CATEGORIA

Obiettivo di Business: Il direttore vendite vuole capire quali categorie di prodotti hanno performato meglio nel **2024**.

La tua sfida: Scrivi una query SQL che restituiscia:

1. L'anno (dalla tabella dim_data).
2. La categoria del prodotto (dalla tabella dim_prodotto).
3. I ricavi totali (somma dei ricavi).

4. La quantità totale venduta.

Requisiti:

- Filtra i dati solo per l'anno **2024**.
- Raggruppa i risultati per anno e categoria.
- Ordina i risultati in base ai **ricavi totali** in ordine decrescente (dal più alto al più basso).

Esercizio 9: "Il Viaggio del Dato: Da TechStore alla Previsione"

Scenario: Sei il Lead Data Engineer di **TechStore**. Il management vuole implementare un sistema per prevedere quali clienti abbandoneranno il servizio (Churn Prediction). Hai davanti a te una lista di **7 azioni tecniche** necessarie per completare il progetto, ma sono state mescolate.

Il tuo compito: Metti in ordine cronologico corretto (dalla 1 alla 7) le seguenti azioni per costruire la pipeline completa.

Lista delle Azioni (Disordinate):

- [A] Dividere il dataset processato in **Training Set (80%)** e **Test Set (20%)**.
- [B] Eseguire il **Lookup** tra i dati estratti e le tabelle dimensionali per sostituire le chiavi naturali (es. cliente_id) con le **Chiavi Surrogate** (es. cliente_sk).
- [C] Definire il **Grain** della Fact Table (es. "Una riga per ogni prodotto venduto in un ordine").
- [D] Applicare fit_transform() sulle feature numeriche del Training Set per gestire i valori mancanti (Imputazione) e scalare i dati.
- [E] Creare metriche di **Feature Engineering** come RFM (Recency, Frequency, Monetary) aggregando i dati storici.
- [F] Utilizzare l'estensione **dblink** per estrarre i dati grezzi dal database OLTP (techstore) verso l'area di staging del DWH.
- [G] Progettare lo **Star Schema** creando le tabelle vuote per Dimensioni (dim_cliente, dim_prodotto) e Fatti (fact_vendite).

Esercizio 10: Esercizio: "L'Algoritmo SCD Type 2: Aggiornare Mario Rossi"

Scenario: Nel Data Warehouse di TechStore, il cliente **Mario Rossi** (cliente_id: 1001) ha appena superato la soglia di spesa e deve passare dal segmento "**Standard**" al segmento "**Premium**". Il sistema ETL deve gestire questo cambiamento usando la logica **SCD Type 2** per mantenere lo storico.

Il tuo compito: Metti in ordine cronologico corretto (dalla 1 alla 5) le seguenti azioni eseguite dall'algoritmo ETL.

Lista delle Azioni (Disordinate):

- [A] Eseguire una INSERT per creare una nuova riga con una **nuova Surrogate Key**, il segmento "Premium", data_inizio odierna e is_current = TRUE.
- [B] Eseguire un UPDATE sul record attivo esistente impostando data_fine a oggi e il flag is_current a FALSE (chiusura del record).
- [C] Caricare le nuove transazioni di vendita nella **Fact Table** collegandole alla **nuova Surrogate Key** generata per Mario Rossi.
- [D] Confrontare i valori in arrivo dal sistema sorgente (Segmento: "Premium") con i valori attuali nel Data Warehouse (Segmento: "Standard") per rilevare il cambiamento.
- [E] Identificare il record corrente di Mario Rossi nella tabella dimensionale usando la chiave naturale (cliente_id = 1001) e il filtro is_current = TRUE.

Esercizio 11 Qual è l'obiettivo fondamentale della Business Intelligence (BI)?

- A. Automatizzare la scrittura di codice SQL per gli analisti di dati.
 - B. Sostituire completamente i database operazionali (OLTP) con un Data Warehouse.
 - C. Trasformare dati grezzi in insight azionabili per supportare decisioni strategiche.
 - D. Creare esclusivamente dashboard visive per il top management aziendale.
-

Esercizio 12 Nella fase 'Transform' del processo ETL, a cosa serve principalmente l'aggregazione dei dati?

- A. A riassumere grandi volumi di dati per migliorare le performance delle query e creare viste analitiche.
 - B. A estrarre i dati da fonti eterogenee come file CSV e API.
 - C. A correggere errori e standardizzare i formati dei dati grezzi.
 - D. A caricare i dati trasformati nel Data Warehouse in modo incrementale.
-

Esercizio 13 Qual è la funzione principale delle chiavi surrogate (es. 'cliente_sk') in uno schema a stella?

- A. Garantire che ogni transazione nella tabella dei fatti sia unica.
- B. Contenere le metriche quantitative misurabili come i ricavi e la quantità.
- C. Fornire il contesto descrittivo per l'analisi, come il nome del cliente o la categoria del prodotto.
- D. Collegare la tabella dei fatti (*fact_vendite*) alle tabelle dimensionali (es. *dim_cliente*).