

# Business Intelligence

📅 Year 2024-2025

## ▼ Guidelines

1. Identificare i FATTI di interesse
  - a. Dimensioni
  - b. Metriche
2. Dipendenze funzionali delle dimensioni
  - a. Gerarchie
  - b. Eventuali dimensioni DEGENERI
3. Schema logico
  - a. Presenza di dimensioni condivise
  - b. Dimensioni degeneri eventualmente inseribili nella Fact table
4. **Risoluzione queries**
  - a. Quali tabelle interrogare (JOINS)
  - b. Predicati da selezionare
  - c. Granularità (Group By, Partition By)
  - d. Calcolo delle metriche (finestre di calcolo, OVER)
5. **Viste Materializzate**
  - a. Carico di Lavoro (tabelle di interesse, predicati, granularità)
  - b. Vista a fattor comune (predicati presenti in tutte le query)
  - c. Fattore di riduzione comune
  - d. Granularità minima per ogni dimensione (Ordine di grandezza di "risparmio")

## Schema logico

Tempo(TimeID, Data, M, 3M, 4M, 6M, Anno)  
Magazzini(MagazzinoID, Magazzino, Città, Provincia, Regione)  
UsoMagazzini(TimeID, MID, m2Liberi, m2Totali)  
Polizze(TimeID, MagazzinoID, ModelID, NumProdotti, ValoreProdotti)  
Modello(ModelID, Modello, Categoria)

## ▼ Queries

Sono di seguito riportate **alcune** delle interrogazioni frequenti di interesse per la dirigenza della ditta:

- a. Relativamente al primo trimestre dell'anno 2013, considerando solo i magazzini della città di Torino, trovare per ogni coppia (magazzino,data) il valore complessivo di prodotti presenti in tale data nel magazzino e il valore complessivo medio giornaliero di prodotti presenti nel magazzino nel corso della settimana precedente la data in esame (data in esame inclusa).

```
SELECT Magazzino, Data, SUM(ValoreProdotti)
      AVG(SUM(ValoreProdotti)) OVER (Partition By Magazzino
                                     Order By Data
                                     RANGE BETWEEN INTERVAL '6' Days preceding and curren
                                     -- oppure [Row 6 Preceding and current row]
      )

FROM Magazzini M, Polizze P, Tempo T
WHERE M.MagazzinoID = P.MagazzinoID and T.TimeID = P.TimeID and
      T.Trimestre = "1-2013" and Città = "Torino"

GROUP BY Magazzino, Data
```

- b. Relativamente all'anno 2004, trovare per ogni coppia (città,data) la percentuale di superficie libera giornaliera nella città. Associare ad ogni coppia un attributo di rank legato alla percentuale di superficie libera giornaliera nella città (1 per la coppia con la più bassa percentuale di superficie libera giornaliera).

```
SELECT città, data, 100 * SUM(M2Liberi) / SUM(M2Totali) AS PercLibera,
      RANK() OVER (ORDER BY SUM(M2Liberi) / SUM(M2Totali) ASC)

FROM Magazzini M, Tempo T, UsoMagazzini U

WHERE M.MagazzinoID = U.MagazzinoID and T.TimeID = U.TimeID and anno = 2004

GROUP BY città, data
```

- c. Relativamente ai primi sei mesi dell'anno 2014, trovare per ogni coppia (magazzino,data) la percentuale di superficie libera giornaliera.

```
SELECT Magazzino, Data, 100 * SUM(M2Liberi) / SUM(M2Totali) AS PercLibera

FROM Magazzini M, Tempo T, UsoMagazzini U

WHERE M.MagazzinoID = U.MagazzinoID and T.TimeID = U.TimeID and 6M = '1-2014'

GROUP BY Magazzino, Data
```

- d. Relativamente all'anno 2013, trovare per ogni coppia (magazzino,mese) il valore complessivo medio giornaliero di prodotti presenti.

```

SELECT Magazzino, Mese, AVG(ValoreProdotti),
      SUM(ValoreProdotti) / COUNT(DISTINCT Data)

FROM Magazzini M, Polizze P, Tempo T

WHERE M.MagazzinoID = P.MagazzinoID and T.TimeID = P.TimeID and anno = 2013

GROUP BY Magazzino, Mese

```

```

SELECT Magazzino, Mese, AVG(TotProdotti)

FROM (SELECT Magazzino, Data, Mese, SUM(Valore Prodotti) as TotProdotti
FROM Magazzini M, Polizze P, Tempo T
WHERE M.MagazzinoID = P.MagazzinoID and T.TimeID = P.TimeID and anno = 2013
GROUP BY Magazzino, Data, Mese)

GROUP BY Magazzino, MESE

```

```

SELECT DISTINCT Magazzino, Mese
      AVG(SUM(Valore Prodotti)) OVER (Partition by Mese, Magazzino)

FROM Magazzini M, Polizze P, Tempo T

WHERE M.MagazzinoID = P.MagazzinoID and T.TimeID = P.TimeID and anno = 2013

GROUP BY Magazzino, Mese, Data

```

m1		Modello 1	M1	150	100
m1		Modello 1	M1	50	100
m1		Modello 1	M1	100	100

$(100 + 50 + 20 + 50) / 4 - ((100 + 50) + 50 + 20) / 3$

1. Relativamente all'anno 2015, trovare per ogni regione il valore complessivo medio giornaliero di prodotti presenti nella regione.
2. Relativamente all'anno 2014, trovare per ogni coppia(mese, regione) la percentuale di superficie libera giornaliera nella regione.

### ▼ Vista materializzata

Definire la vista materializzata che potrebbe essere utilizzata per rispondere efficientemente a tutte le seguenti interrogazioni:

- Relativamente al primo semestre degli anni 2015 e 2016, considerando solo i magazzini nella regione Piemonte , trovare per ogni coppia (provincia del magazzino, mese) il valore complessivo di prodotti presenti e il relativo numero.

- Relativamente all'anno 2013, trovare per ogni provincia del magazzino e semestre il valore mensile medio di prodotti.
- Relativamente all'anno 2014, trovare per ogni regione del magazzino e trimestre il valore complessivo di prodotti presenti nella regione.

**Tabella:** Magazzini, Tempo, Polizze

**Filtri:** 6M in {"1-2015", "1-2016"}, regione = "Piemonte"

**Group by:** Provincia, Mese,

**Misure:** SUM(ValoreProdotti), SUM(NumProdotti)

**Tabella:** Tempo, Magazzini, Polizze

**Filtri:** anno = 2013

**Group by:** Provincia, 6M

**Misure:** SUM(ValoreProdotti) / SUM(NumProdotti) / COUNT(DISTINCT Mese)

- *Remark 1:* dividere per SUM(NumProdotti) non è strettamente necessario in quanto il testo potrebbe anche essere interpretato come "media giornaliera del totale", in ogni caso la vista materializzata non cambierebbe.

- *Remark 2:* COUNT(DISTINCT Mese) non compare nella vista perché agevolmente ricalcolabile e inoltre perché in quel livello di aggregazione sarebbe sempre 1.

**Tabella:** Tempo, Magazzini, Polizze

**Filtri:** anno=2014

**Group by:** Regione, 3M

**Misure:** SUM(ValoreProdotti)

```
SELECT Provincia, Mese, 3M, 6M, Regione, SUM(ValoreProdotti), SUM(NumProdotti)
FROM Magazzini, Tempo, Polizze
WHERE <join conditions>
GROUP BY Provincia, Mese, 3M, 6M, Regione
```

## Cardinalità

Dal testo:

*Il data warehouse realizzato deve contenere le informazioni relative agli ultimi 10 anni. Al fine di una corretta realizzazione del data warehouse sono state fornite le seguenti informazioni:*

- o Numero di modelli diversi di prodotti: ~100*
- o Numero di categorie diverse di prodotti: ~10*
- o Numero di magazzini: ~500*
- o Numero di magazzini di Torino: ~5*
- o Numero di città: ~200*
- o Numero di regioni: ~20*
- o Numero di province: ~90*

- TEMPO: 10 anni \* 365 → 10 \* 12 **riduzione 1/10**
- Magazzini: 500 → 90 → **riduzione 1/10**
- Prodotti: 100 → 1 → **riduzione 1/100**

**Riduzione totale  $10^{-4}$**