

# Basi di dati attive

Una base di dati è ATTIVA quando consente la definizione e la gestione di regole di produzione (*regole attive o trigger*).

Tali regole vengono attivate in modo automatico al verificarsi di specifici *eventi* sulla base di dati e, verificate alcune *condizioni*, eseguono *azioni* sulla base di dati stessa.

# Basi di dati attive

STRUTTURA generale dei trigger

**when** <event>

**if** <condition>

**then** <action>

<event>: sono eventi che corrispondono a modifiche dei dati (insert, delete, update).

<condition>: è una condizione booleana specificata in SQL.

<action>: è una sequenza di comandi SQL che producono un'azione sulla base di dati.

# Basi di dati attive

Sintassi di un trigger in SQL-3

CREATE TRIGGER <nome>

BEFORE | AFTER | INSTEAD OF <evento>

ON <tabella>

[REFERENCING <variabili>]

WHEN ( <condizione SQL> )

<sequenza di comandi SQL>

[FOR EACH {ROW | STATEMENT}]

# Basi di dati attive

<evento>: INSERT | DELETE | UPDATE

<variabili>:

OLD AS <nome tuple vecchie>

NEW AS <nome tuple nuove>

OLD\_TABLE AS <nome tabella vecchia>

NEW\_TABLE AS <nome tabella nuova>

# Basi di dati attive

Esempio di trigger in SQL-3:

Studente(matricola, nome, anno)

Esame(matricola, insegnamento, data, voto)

```
CREATE TRIGGER T1 AFTER UPDATE
ON Studente REFERENCING NEW AS n
WHEN (n.anno > 2 AND 7 > count(SELECT *
FROM Esame WHERE matricola = n.matricola )
ROLLBACK
FOR EACH ROW
```

# Data Warehousing (DW)

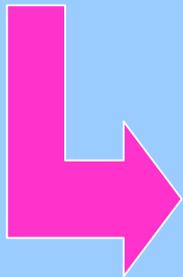
Data warehousing come processo per memorizzare dati storici ed integrati, eventualmente memorizzati in luoghi diversi, da usare come supporto al sistema di decisione aziendale



# Esigenze che portano al DW

Uso di dati a fini decisionali:

- grandi quantità di dati
- più basi di dati, tra loro eterogenee
- i dati devono essere raggruppati, classificati, riassunti



Tecnologia ad hoc  
(DBMS non sono sufficienti)

# Perché la tecnologia corrente non è sufficiente?

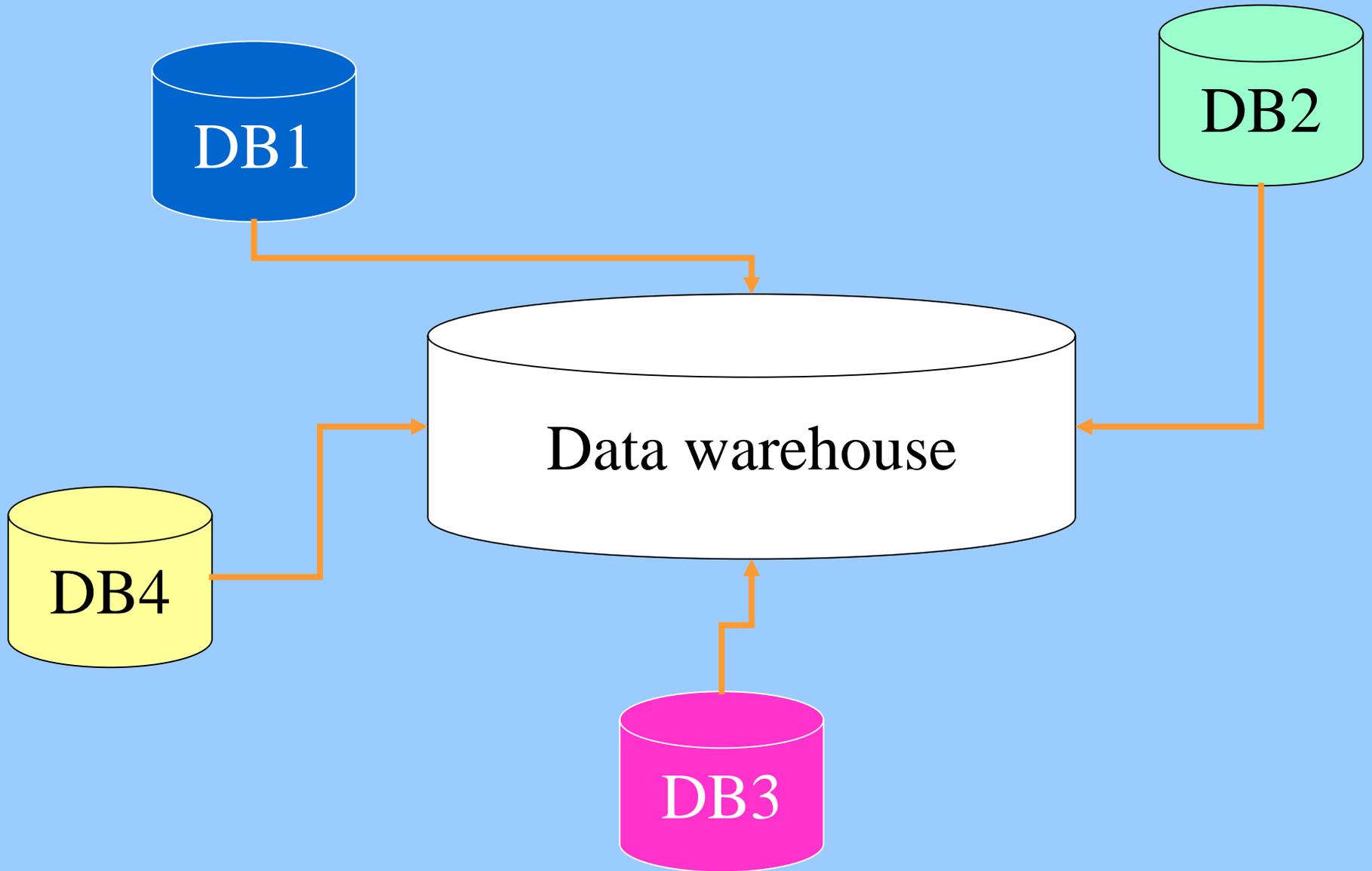
- sistemi eterogenei
- basse prestazioni
- DBMS non adeguati al supporto decisionale
- problemi di sicurezza

# Soluzione

Nuova base di dati  data warehouse

Nuovo sistema di gestione  
dati  sistema di data  
warehousing

# Il data warehouse



# Il data warehouse

## Collezione di dati

orientata ai soggetti

integrata

correlata alla variabile tempo

non-volatile

usata principalmente per il supporto alle decisioni

# Il data warehouse

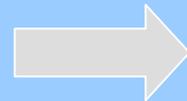
*Orientata ai soggetti*

Considera soggetto di interesse all'organizzazione e non in funzione dei processi organizzativi

# Il data warehouse

## *Integrata*

I dati arrivano da fonti diverse



diversi formati

integrazione



dati memorizzati in  
formati consistenti

Esempio: sesso dell'individuo può essere  
rappresentato come “M” e “F” o come “0” e “1”

# Il data warehouse

*Correlata alla variabile tempo*

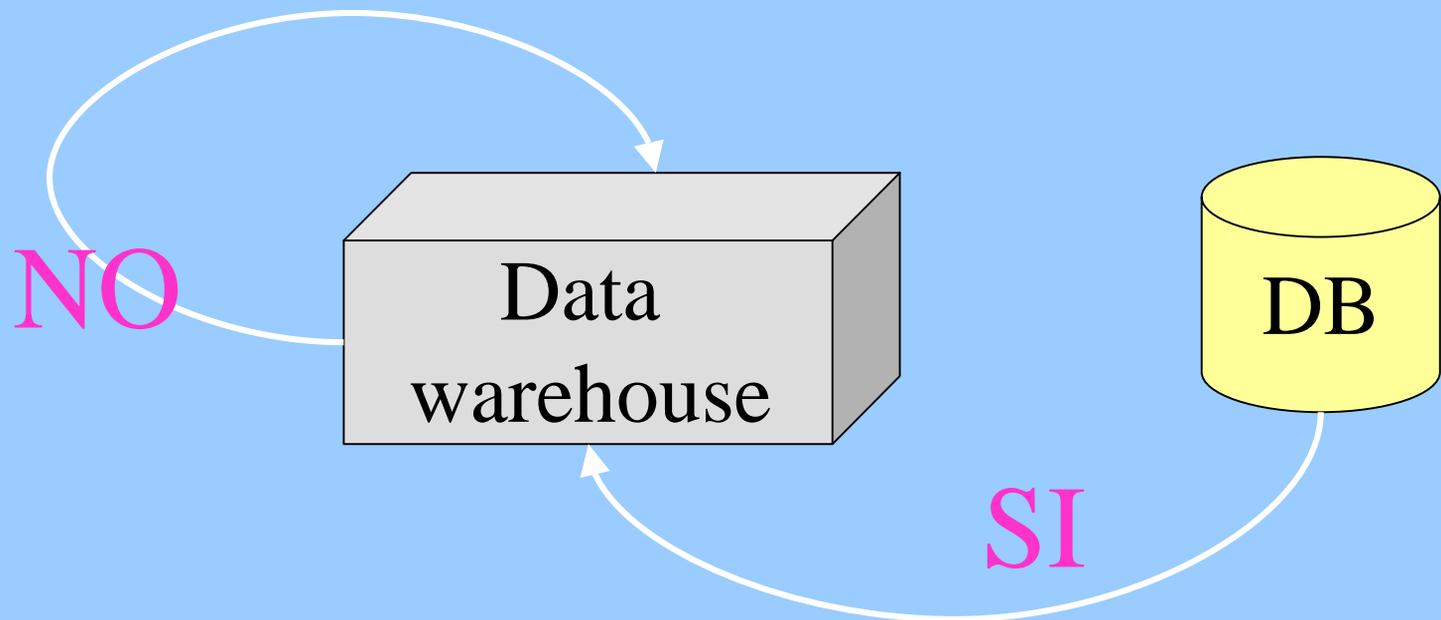
Presenza di dati storici per eseguire confronti, previsioni e per individuare tendenze

Esempio sistema bancario: dati relativi ai vecchi clienti per stabilire quanti clienti chiuderanno il conto nell'anno seguente

# Il data warehouse

*Non volatile*

Una volta che le informazioni sono inserite nel data warehouse, non possono essere modificate ma solo ricaricate



# Gestione del data warehouse

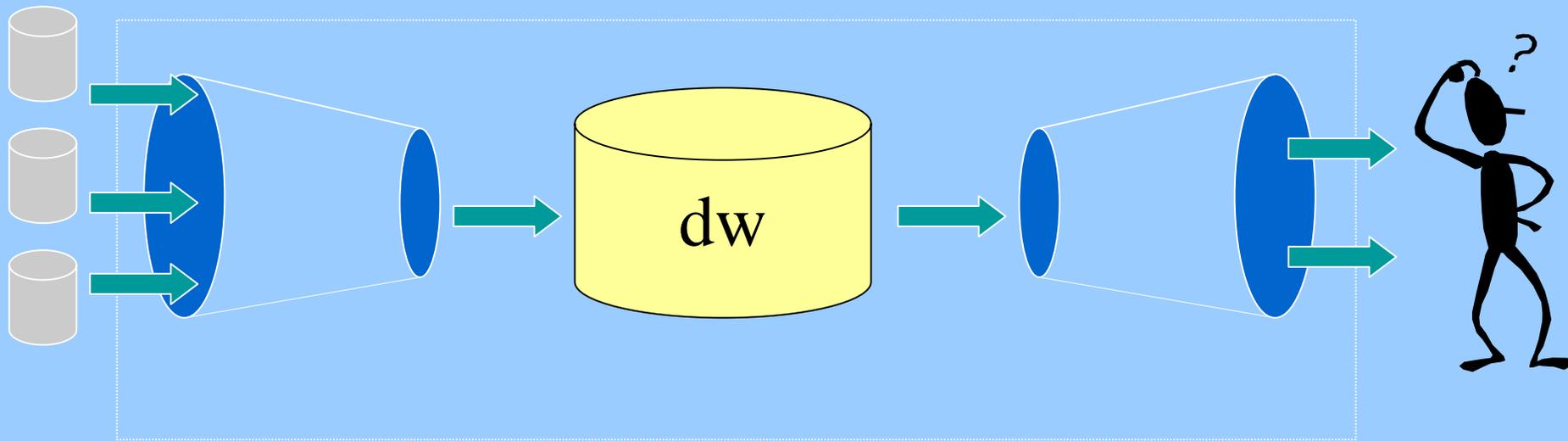


# Una prima architettura funzionale

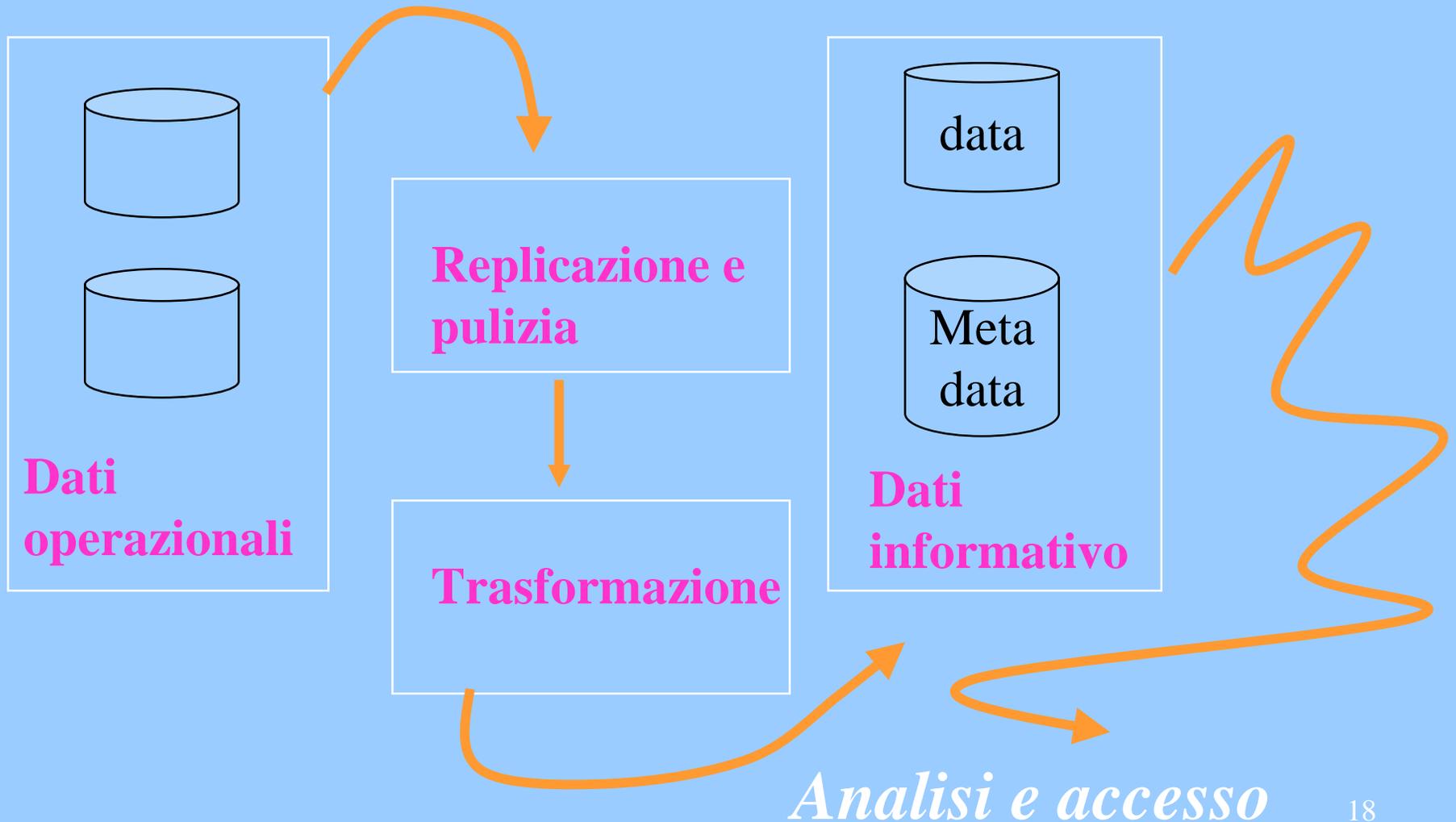
acquisizione

memorizzazione

accesso



# Una seconda architettura funzionale



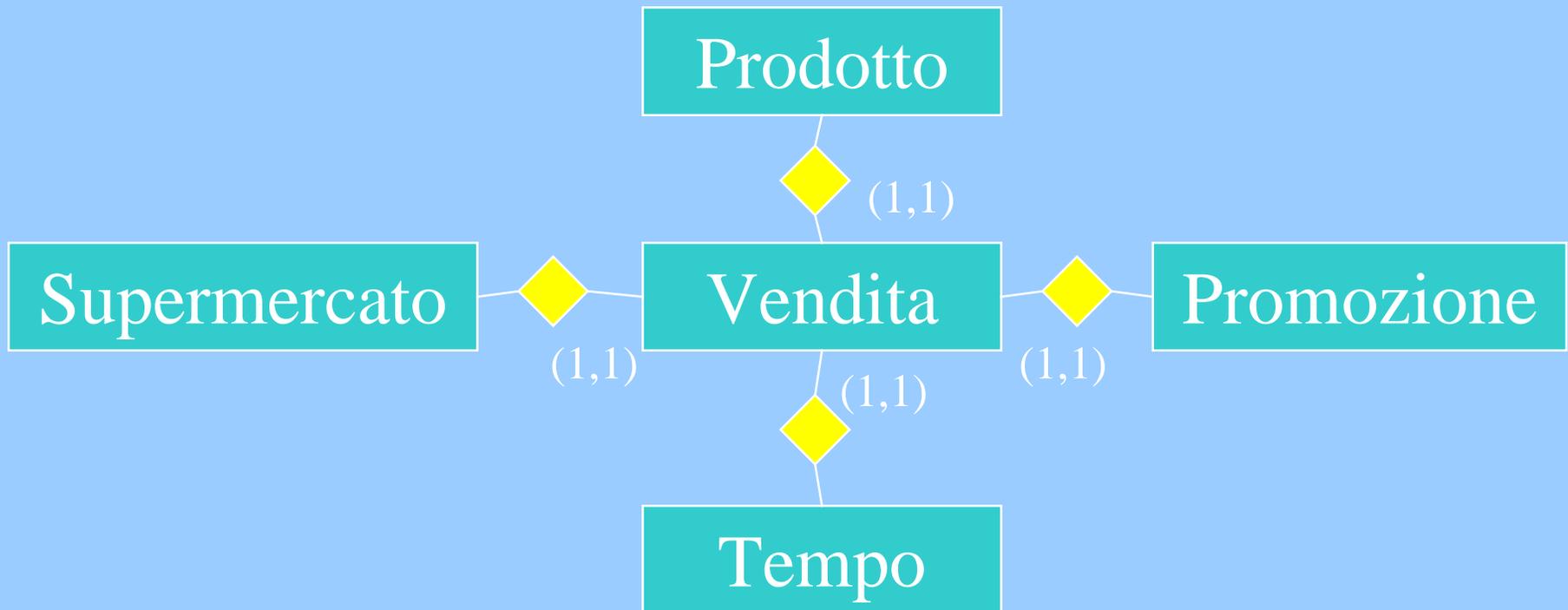
# Schema di un DW

E' strutturato in sottoparti separate dette DATA MART per le quali risulta chiaro l'obiettivo dell'analisi.

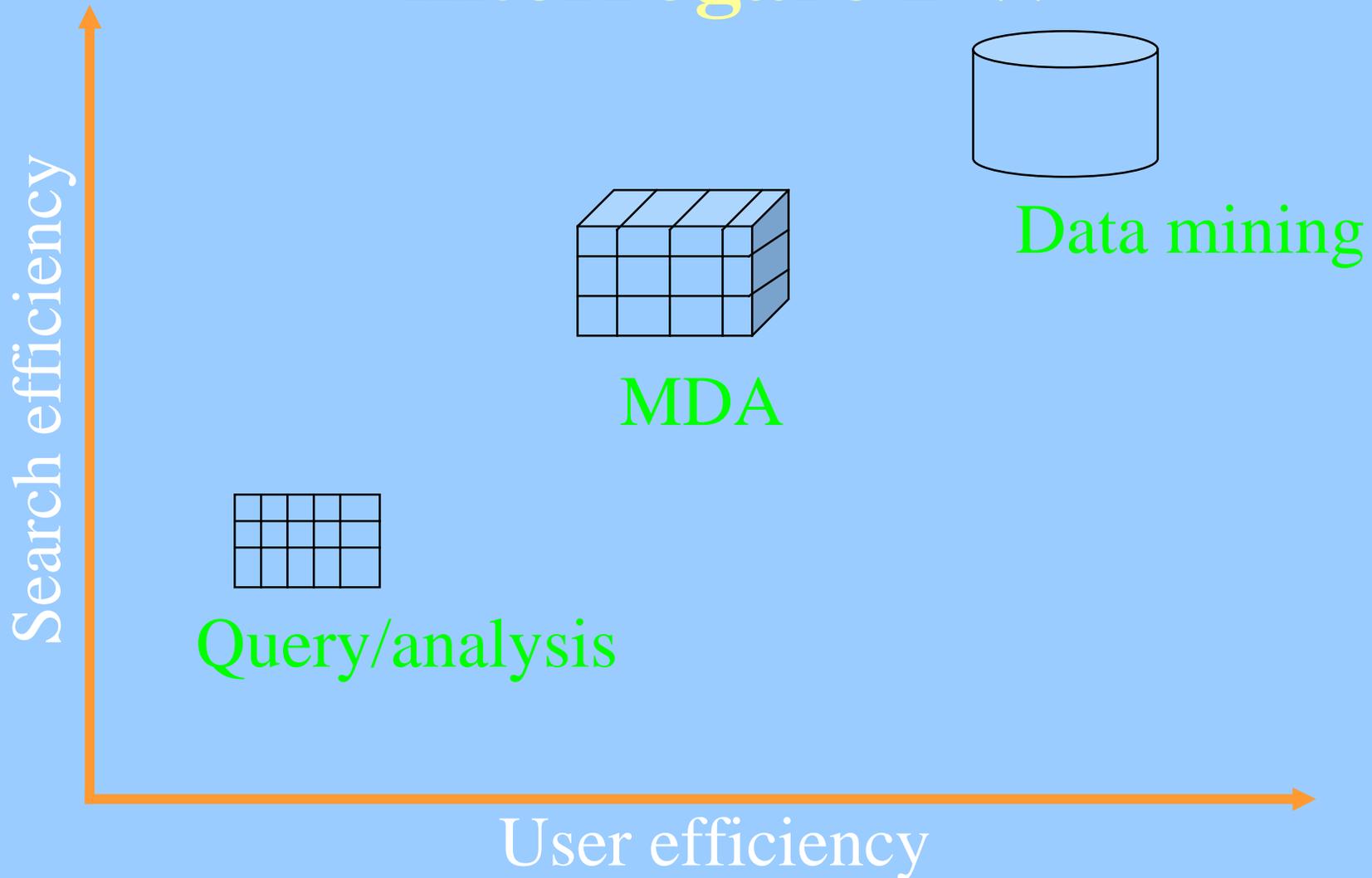
Ogni DATA MART ha uno schema concettuale che spesso si presenta con una struttura a STELLA.

# Schema di un DATA MART

Schema a *stella*: al centro l'entità che rappresenta i **fatti**, alle estremità le entità che rappresentano le **dimensioni** dell'analisi.



# Evoluzione degli strumenti per interrogare DW



# Query e analisi

- Permettono di formulare interrogazioni senza scrivere un programma o conoscere SQL
- mettono a disposizione alcuni meccanismi di reporting: generazione di documenti (tabelle, grafici e altro) per presentare il risultato dell'analisi

# Formulazione delle query via interfaccia grafica

Esempio di interfaccia

dim1 (D1.a)	dim2 (D2.b)	fatti (F.c)	fatti (F.d)
valori presenti	valori presenti		
selezione	selezione		
group by	group by	OP1 <sub>Aggr</sub>	OP2 <sub>Aggr</sub>

# Query SQL equivalente all'interfaccia grafica

```
SELECT D1.a, D2.b, OP1Aggr(F.c), OP2Aggr(F.d)
FROM Fatti AS F, Dim1 AS D1, Dim2 AS D2
WHERE <predicati di join(F,D1)> AND
<predicati di join(F,D2)> AND <selezioni>
GROUP BY D1.a, D2.b
ORDER BY D1.a, D2.b
```

# Strumenti

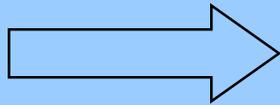
## Multi Dimensional Analysis

Permettono di creare viste multidimensionali su ordinarie tabelle bidimensionali (si possono anche usare particolari DBMS multidimensionali)

# MDA tools

## Esempio

vendite



Dati di dettaglio  
(fatti)

tempo

cliente

prezzo

magazzino

prodotto

Dimensioni  
(parametri  
analizzati)

# MDA tool

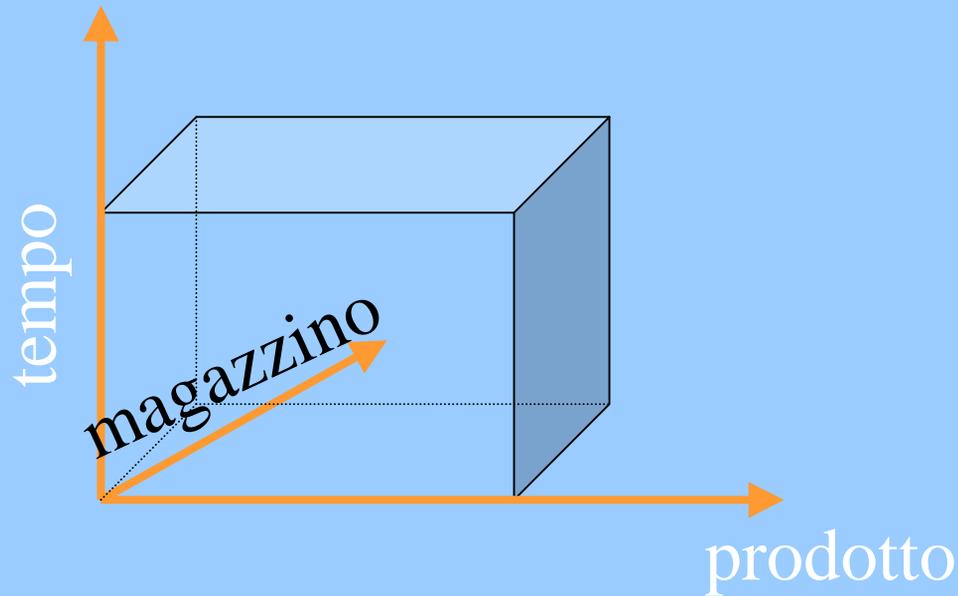
## Esempio di query

Quali sono i dati sulle vendite,  
classificati rispetto a:

- prodotto
- magazzino
- mese?

# MDA tools

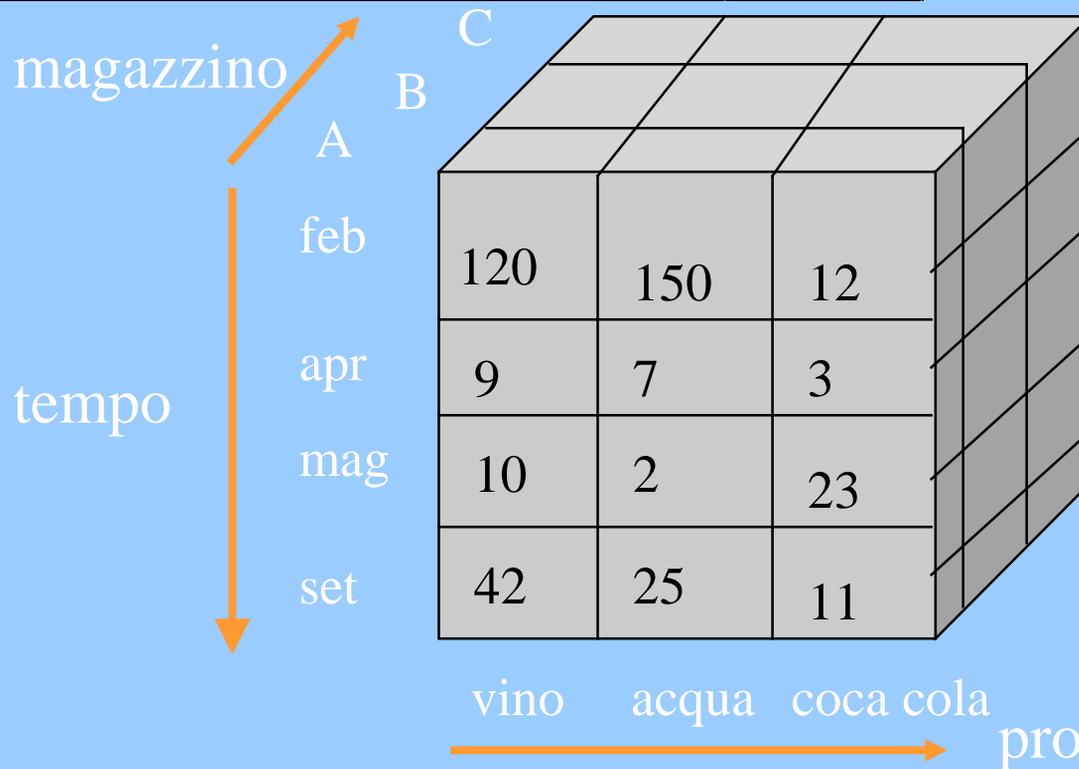
## Analisi rispetto a variabili multiple



- In genere: efficienti fino a 10 variabili

# MDA tools

vendite	prodotto	mese	magazzino	quantità	ricavi
1	vino	febbraio	A	120	150000
2	acqua	febbraio	B	34	34000
3	coca cola	aprile	A	245	380000
4	acqua	maggio	A	47	99000
5	acqua	settembre	C	55	
...	...	...	...	...	...



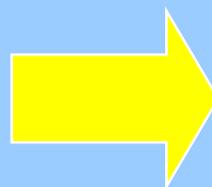
# MDA tools

Le strutture dati **OLAP (MDA)** si possono pensare come un **CUBO DI RUBIK** di dati, che gli utenti possono modificare in modi diversi per analizzare diversi scenari del tipo “what-if”

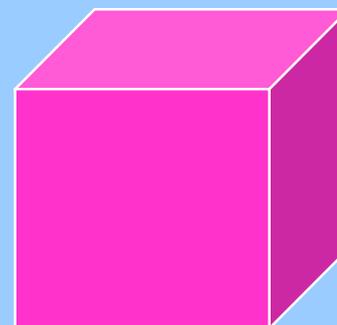
# Data Cube nei sistemi

Clausola: WITH CUBE su una interrogazione SQL con il GROUP BY.

Mese	Magazzino	Prodotto	Quantità
Gen	A	Vino	10
Feb	A	Vino	20
Mar	A	Acqua	30
Gen	B	Acqua	40
Feb	B	Vino	50
Mar	B	Vino	60
Gen	C	Vino	70
Feb	C	Acqua	80
Mar	C	Acqua	90



```
SELECT Mese,  
Magazzino, Prodotto,  
Sum(Quantità) FROM  
T GROUP BY Mese,  
Prodotto WITH CUBE
```



# Data Cube nei sistemi

Tabella risultato della query WITH CUBE del precedente lucido.

Mese	Prodotto	sum(Qta)
Gen	Vino	80
Feb	Vino	70
Mar	Vino	60
Gen	Acqua	40
Feb	Acqua	80
Mar	Acqua	120
Gen	ALL	120
Feb	ALL	150
Mar	ALL	180
ALL	Vino	210
ALL	Acqua	240
ALL	ALL	450

# MDA tools

## Due tipi di analisi:

1. *slice and dice*: si parte da un particolare aspetto e si naviga nella struttura multidimensionale

## 2. *drill*

*drill-down*: da informazioni più generali a informazioni meno generali

*drill-up*: da informazioni meno generali a informazioni più generali