



# PROGETTAZIONE E IMPLEMENTAZIONE DI UN DATAWAREHOUSE

Prof. Crescenzo Gallo

# Introduzione

La necessità di **conservare dati e informazioni** in modo permanente, perché potranno essere utili in momenti successivi, è un problema molto evidente nel mondo moderno e riguarda ormai un numero elevatissimo di persone ed agenti economici.

La conservazione e il successivo utilizzo di dati può costituire una fonte preziosa per **prendere decisioni** o per **aumentare l'attività economica** dell'impresa.

# Sistema informativo basato su Database relazionale ==> OLTP

## **Problemi:**

- Tempi di risposta
- Impossibilità di eseguire analisi più complesse sui dati
- Continua evoluzione dei dati
- Mancanza di dati storici adeguati
- Dati registrati in formati diversi

# Tempi di risposta

In un sistema transazionale i dati sono distribuiti su numerose tabelle per permettere l'efficace aggiornamento. In un'interrogazione complessa i dati devono essere ricercati su più tabelle, con un conseguente rallentamento della ricerca.

# Analisi complesse

Il sistema informativo \_\_\_\_\_ (OLTP - Online Transaction Processing) non è stato studiato per consentire analisi articolate.

Un sistema OLTP, infatti, non permette di:

- studiare una grande quantità di dati
- vedere i dati da prospettive diverse
- supportare i processi decisionali

Infatti i dati gestiti e conservati all'interno dei sistemi transazionali non si prestano facilmente al tipo di analisi di cui gli utenti hanno bisogno.

# Altri limiti

- **Continua evoluzione dei dati** = i dati in un sistema transazionale non sono affidabili se usati all'interno di un processo decisionale; essi cambiano in continuazione per via delle operazioni effettuate attraverso l'OLTP.
- **Mancanza di dati storici adeguati** = le applicazioni di un sistema transazionale sono progettate per gestire processi riguardanti il momento attuale dell'azienda, senza fare riferimento a dati nel passato.
- **Dati registrati in formati diversi** = applicazioni diverse possono usare tecnologie e piattaforme diverse, rendendone onerosa l'integrazione in un formato comune.

# DATA WAREHOUSE

Come risolvere queste problematiche?

- Progettazione e Implementazione di un DATA WAREHOUSE.

# DATA WAREHOUSE

Un **Data Warehouse** (dall'inglese *magazzino dati*), è un archivio informatico contenente grandi quantità di dati di un'organizzazione. Sono progettati per consentire di produrre facilmente relazioni ed analisi.

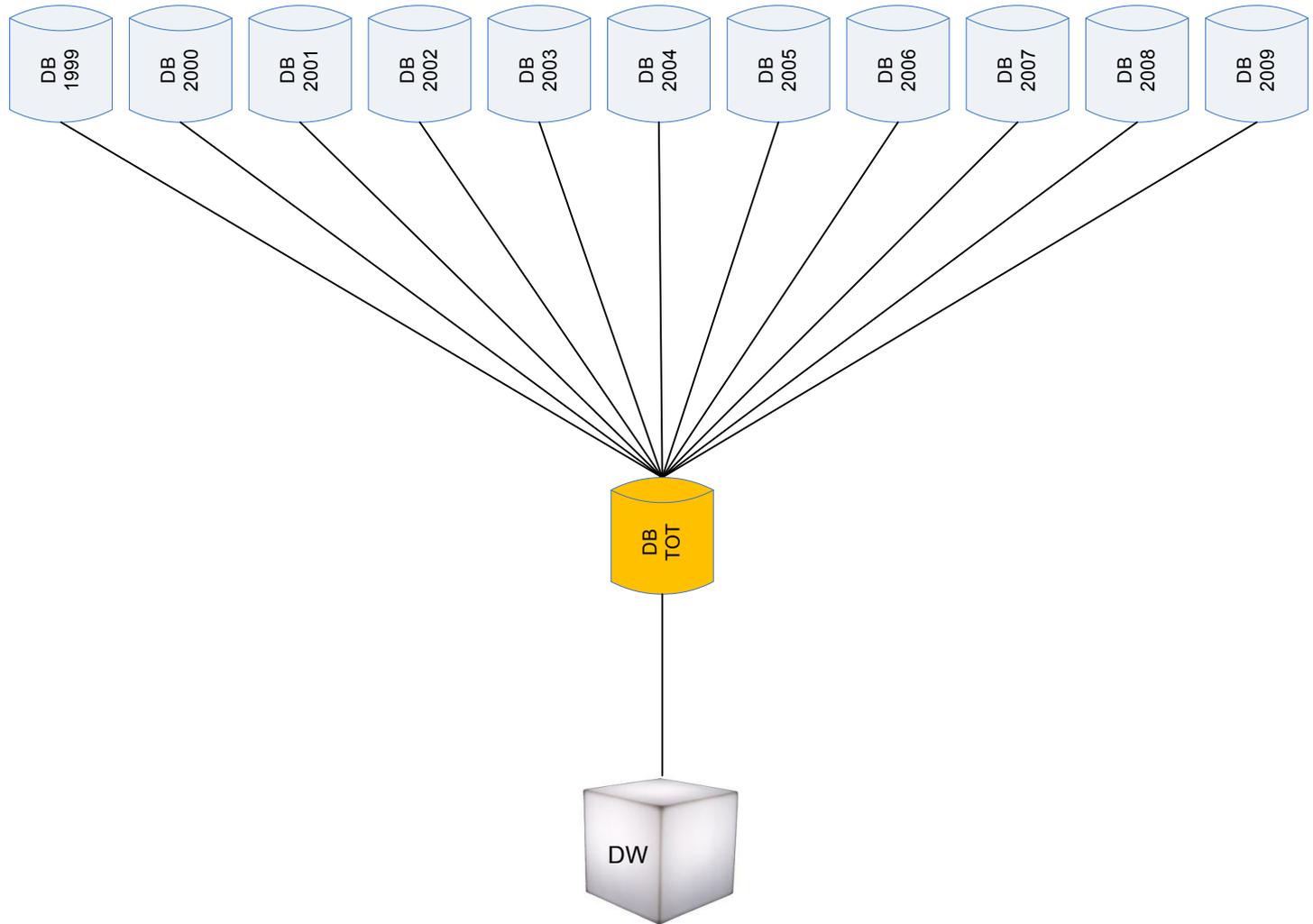
# Definizione DW (Inmon)

Un DW è una raccolta di dati:

- **integrata,**
- **subject oriented,**
- **time variant**
- **non-volatile**

di supporto ai processi decisionali

# Progettazione DW



# Progettazione DW

Basi di dati del DW → **Multidimensionali** (M-OLAP)

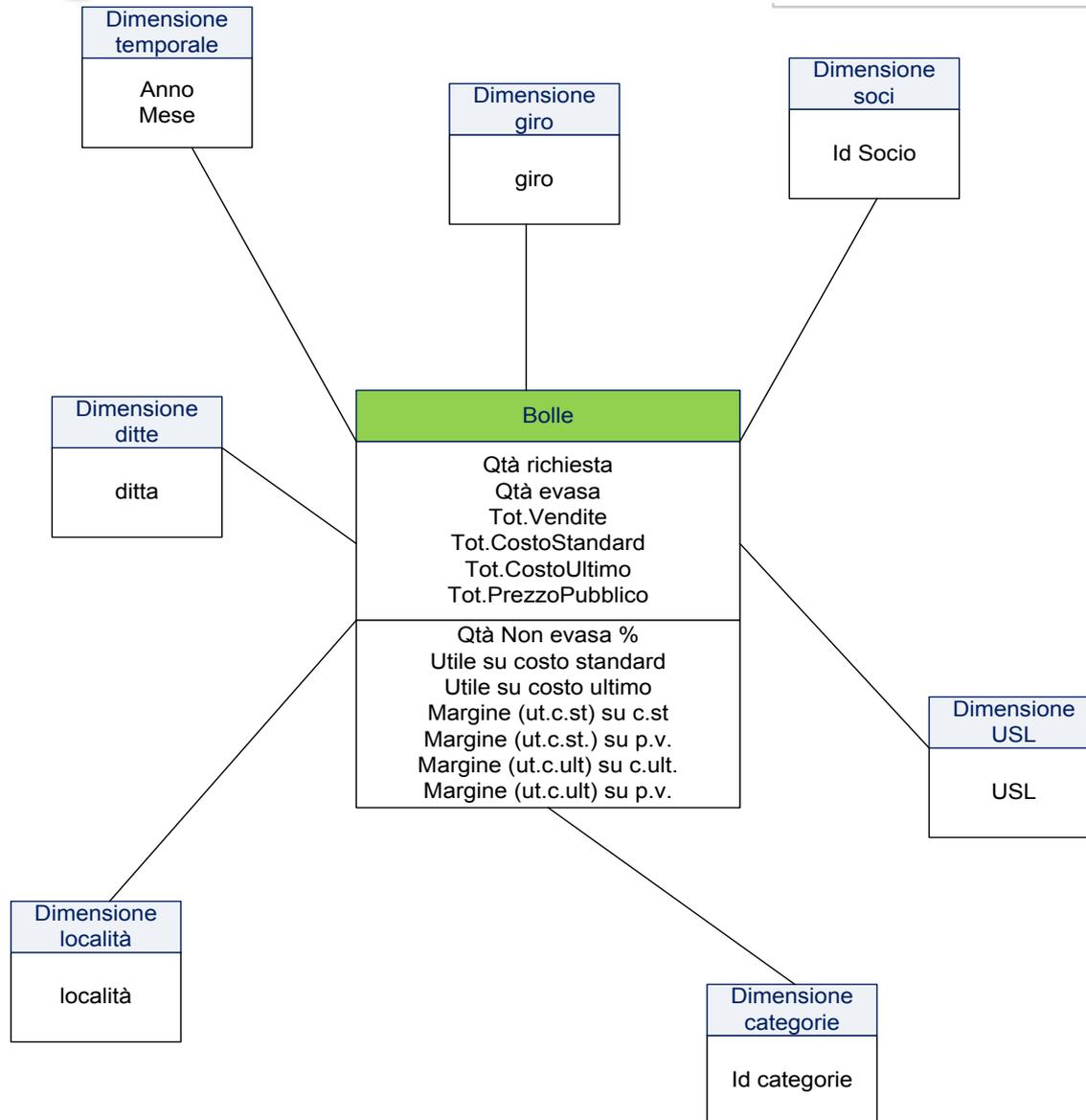
- Dati come **dimensioni** e non tabelle
- Supporta la registrazione dei dati in vari **livelli di aggregazione** (possibilità di effettuare *drill-down* e *roll-up*)

# Progettazione DW

Modello dati DW → **STAR schema**

- La **tabella dei fatti** contiene solo attributi che misurano il business oltre agli identificativi (le chiavi esterne)
- Le **tabelle dimensioni** contengono attributi che descrivono la dimensione, oltre agli identificativi (chiavi) che indicizzano e organizzano i dati della tabella dei fatti.

# Progettazione DW



# UN ESEMPIO DI ANALISI SUL CUBO

	+ Anno	
Categoria	- Totale Tempo	+ 2009
- Totale Categoria	€	€
2	€	€
3	€	€
4	€	€
5	€	€
6	€	€
7	€	€
A	€	€
B	€	€

- *Drill – down*

- *Roll - up*

# Analisi cubo su Microsoft Excel

	A	B	C	D	E	F	G	H	I	J
1	Prezzo Vendita	Etichette di colonna								
2	Etichette di riga	2	3	4	5	6	7	A	B	Totale complessivo
3	2009									
4	Totale complessivo									
5										
6										
7										
8										
9										
10										
11										

**Drill – down** dimensione temporale

**Filtro dati** → possibilità di inserire nella tabella solo i dati relativi ad uno o più mesi/anni, oppure relativi ad una o più categorie.

Σ Valori

Categoria

Ditta

Giro

Località

Soci

Tempo

USL

Trascinare i campi nelle aree sottostanti:

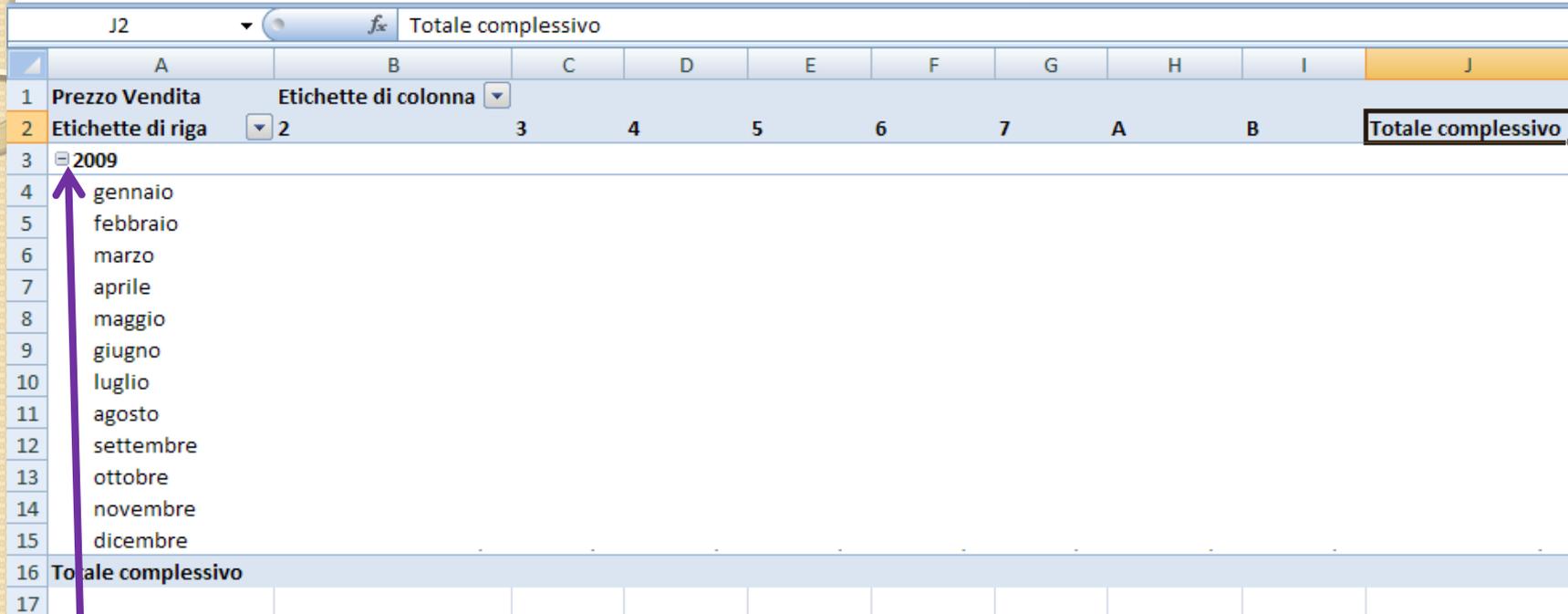
Filtro rapporto Etichette di col...

Categoria

Etichette di riga Σ Valori

Tempo Prezzo Vendita

# Analisi cubo su Microsoft Excel



The screenshot shows an Excel PivotTable with the following structure:

	A	B	C	D	E	F	G	H	I	J
1	Prezzo Vendita	Etichette di colonna								
2	Etichette di riga	2	3	4	5	6	7	A	B	Totale complessivo
3	2009									
4	gennaio									
5	febbraio									
6	marzo									
7	aprile									
8	maggio									
9	giugno									
10	luglio									
11	agosto									
12	settembre									
13	ottobre									
14	novembre									
15	dicembre									
16	Totale complessivo									
17										

**Roll-up** dimensione temporale