

# **Sistemi direzionali e modello multidimensionale**

Prof. Piercarlo Giolito

## Data warehousing e tecnologia OLAP

### *Argomenti trattati.*

- **Evoluzione dei Sistemi Informativi Decisionali**
- **Il modello dei dati multidimensionale**
- **Scenari decisionali**
- **Architettura di un sistema direzionale**

## Evoluzione dei Sistemi Informativi Direzionali o Decisionali

1950

### Transaction Processin Systems (TPS)

Reportistica contabile, sono sistemi che hanno riscontrato un discreto successo.

1960

### Management Information Systems (MIS)

- Vero primo tentativo di funzioni a supporto delle decisioni con report extracontabili;
- Difetto: poco maneggevoli. Erano basati su **rapporti predefiniti**;  
Informazioni mancanti e di difficile reperimento nella mole di dati.

1970

### Dai MIS ai Decision Support Systems (DSS) [Scott Marton, Gerrity,...]

- **Accesso tempestivo e ad hoc alla base dati** aziendale;
- Funzioni di sintesi dei dati (analisi statistiche);
- Expert Systems costruiti da esperti del dominio;

1980

- Knowledge Based Systems per sintesi di opinioni, valutazioni e consigli tramite il ragionamento simbolico e sistemi esperti;

### Executive Information Systems (EIS)

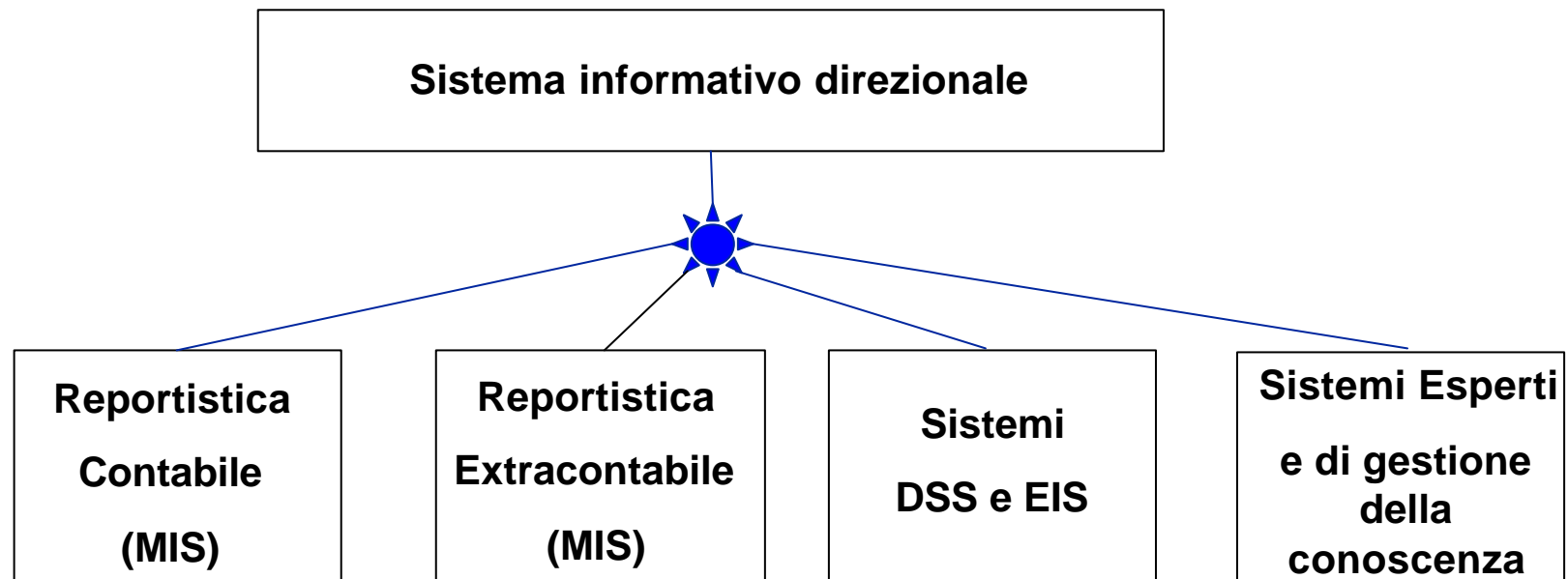
*Cruscotti* di immediata lettura per analisi di dati: facili, ma poco potenti.

## L'offerta attuale

**Business Intelligence:** metodi e strumenti per convertire dati in informazioni, informazioni in conoscenza e conoscenza in piani di sviluppo (D. Loshin 2003).

Proposta originale di Howard Dresner del Gartner Group (1989)

Tecnologia **data warehouse** e **OLAP** (Online Analytical Processing)



## Componenti di un sistema **Business Intelligence**

### Data warehouse:

Insieme di dati e strumenti per effettuare vari tipi di interrogazioni a carattere generalmente statistico e analitico.

I dati, in genere provenienti da fonti diverse, sono attinti dal **sistema operativo** dell'azienda e organizzati in modo opportuno.

### OLAP (Online Analytical Processing)

Manipolazione dei dati modellati con **strutture multidimensionali**.

### EIS (Executive Information System)

Interfacce ad oggetti prestabilite per analisi dei dati.

### Data Mining

Sistemi per sintetizzare relazioni sconosciute all'utente dai dati disponibili attraverso tecniche avanzate di esplorazione e di modellistica:

- Reti neurali
- Alberi di classificazione
- Algoritmi genetici
- Clustering

Un sistema di **data warehouse** tratta insiemi di dati estratti da vari sottosistemi aziendali e memorizzati in un proprio data base.

### **Caratteristiche di un data warehouse:**

- accento posto sui fatti e sulle informazioni da essi generati (dati interpretati secondo uno schema predefinito).
- integra dati provenienti da fonti eterogenee, a volte in conflitto.
- storicizza dati che spesso nei sistemi operazionali sono sovrascritti.
- l'arco temporale considerato (4 o più anni) è spesso superiore al lasso temporale gestito dai sistemi operazionali (1 o 2 anni).
- struttura lo schema interpretativo dei dati in **data mart** che corrispondono a punti di vista specifici spesso dettati da precise divisioni aziendali.

Lo schema di un data warehouse segue il **modello dei dati multidimensionale**

Il **modello dei dati multidimensionale** è adeguato per un'analisi dei dati da diverse prospettive, tecnicamente chiamate **dimensioni**.  
 I dati, cioè le **misure** di interesse, sono raccolti nella **fact table (tabella di fatto: fatto vendita)**.

Esempio bidimensionale:

**Tabella di fatto: fatto vendita**

punto vendita	tipo prodotto	valore del venduto
Torino	quaderni	1000
Milano	matite	1500
Torino	matite	700
Torino	zainetti	5000
Milano	quaderni	1400
Milano	zainetti	8000
Asti	zainetti	400

Eventi

Tabella delle Dimensioni su

valore del venduto		punto vendita		
		Asti	Milano	Torino
tipo prodotto	matite	0	1500	700
	quaderni	0	1400	1000
	zainetti	400	8000	5000

La rappresentazione a matrice con le due dimensioni (**coordinate**) esibisce gli stessi indicatori (misure) della tabella di fatto in forma semplice e comoda sia in lettura sia per totali di varia natura.

Tabella delle Dimensioni su

valore del venduto		punto vendita			totali
		Asti	Milano	Torino	
prodotto tipo	matite	0	1500	700	2200
	quaderni	0	1400	1000	2400
	zainetti	400	8000	5000	13400

In sintesi i concetti base del data warehouse sono:

**Fatto**, Concetto d'interesse per il processo decisionale (strategico, innovazione dei processi e dei prodotti, analisi catena distributiva,...)

**Misura**, proprietà numerica di un fatto, descrive un aspetto quantitativo di un indicatore d'interesse.



**Dimensione**, proprietà con dominio finito di un fatto, descrive una coordinata d'analisi.

**Evento**, istanza di fatto

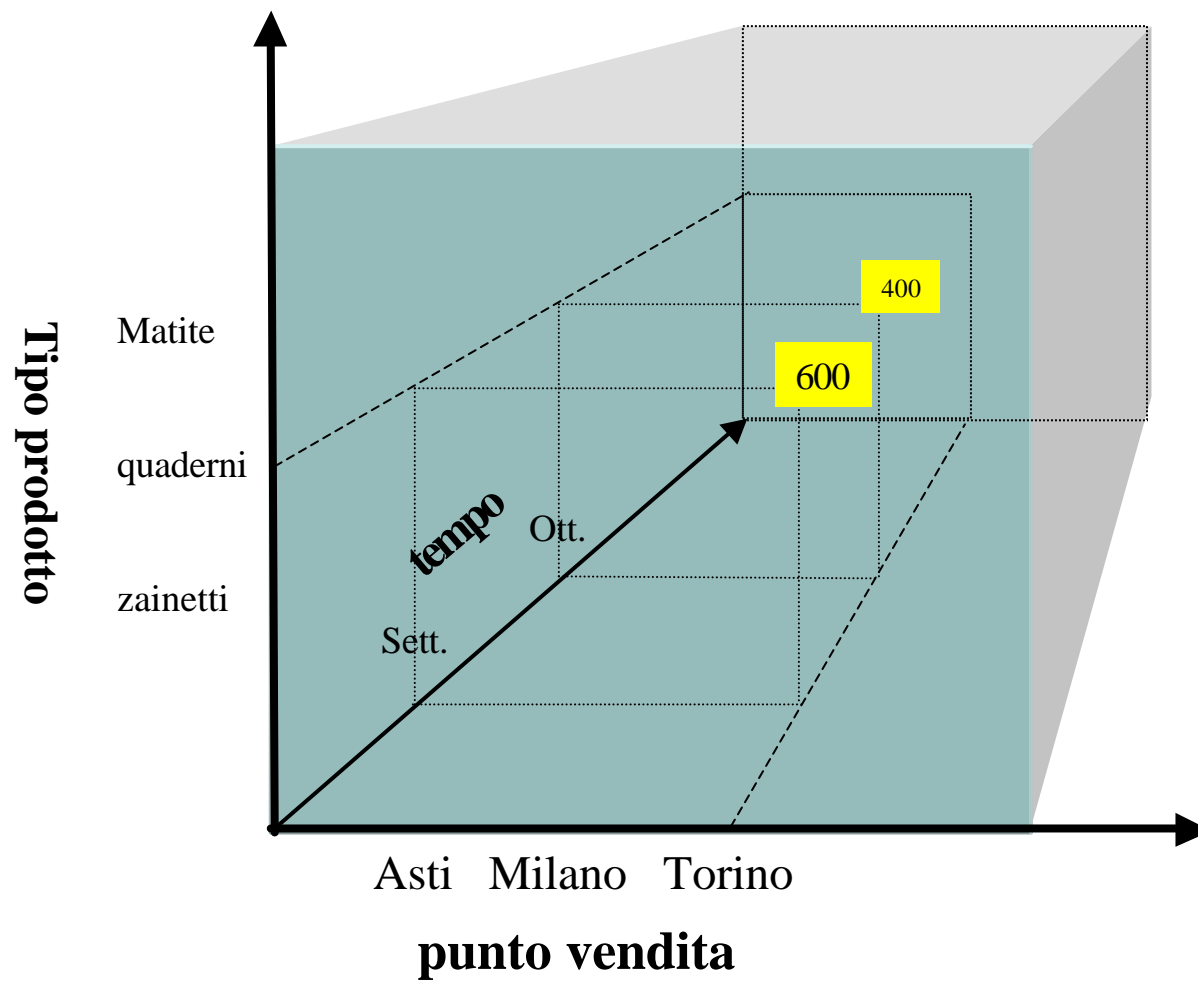
Tabella di fatto con aggiunta di una dimensione: la **dimensione tempo**.  
 Supponiamo che i dati si riferiscano ad uno specifico anno, poniamo 2000.

punto vendita	tipo prodotto	tempo	valore venduto
Torino	quaderni	Settembre	600
Torino	quaderni	Ottobre	400
Milano	matite	Settembre	800
Milano	matite	Ottobre	700
Torino	matite	Settembre	200
Torino	matite	Ottobre	500
Torino	zainetti	Settembre	2500
Torino	zainetti	Ottobre	2500
Milano	quaderni	Ottobre	1400
Milano	zainetti	Settembre	3000
Milano	zainetti	Ottobre	5000
Asti	zainetti	Ottobre	400

Schema relazionale  
 della tabella di fatto

<b>VENDITE</b>
tipo-prodotto
punto-vendita
tempo
valore-venduto

## Generalizzazione multidimensionale



Il modello multidimensionale può essere esplorato da vari punti di vista (operazione di **slice**):

Quaderni punto vendita/mese		punto vendita			totali
		Asti	Milano	Torino	
tempo	Sett.	0	0	600	600
	Ott.	0	1400	400	1800

Torino tipo prodotto/mese		tipo prodotto			totali
		matite	quaderni	zainetti	
tempo	Sett.	200	600	2500	3300
	Ott.	500	400	2500	3400

Il numero di dimensioni può, senza difficoltà concettuali, superare la fatidica soglia delle 3 dimensioni intuitive. L'ipercubo risultante è sempre esplorato da vari punti di vista (Esempio precedente di operatore **slice**).

Anche le misure possono essere più di una.

In tal caso ogni cella dell'ipercubo dovrà contenere più valori.

Le conseguenti azioni di aggregazione dovranno selezionare opportunamente le misure da cui estrarre le informazioni volute.

Estensione dell'esempio precedente:

<b>Dimensioni</b>	<b>misure</b>
tipo prodotto	valore venduto
punto vendita	quantità venduta
periodo	giacenza
rivenditore	.....
cliente (dettaglio, grossista)	
.....	

## Categorizzazione schematica delle dimensioni:

<b>Prodotto</b>	Es.: tipo prodotto,...
<b>Responsabilità</b>	Es.: punto vendita, rivenditore,...
<b>Tempo</b>	Es.: mese, bimestre, anno,...
<b>Cliente</b>	Es.: dettaglio, grossista, diretto, indiretto,...
<b>Processo</b>	Es.: processi gestionali per segmentare costi,...
<b>Altro</b>	

## Proprietà delle misure:

**Tipologia** del valore, quale valore corrente, budget iniziale, ...

**Trasformazione e calcolo**, cioè valore in input dall'operazionale o calcolato

**Fonte** della misura

**Riferimento temporale**: date, time stamp,...

Lo schema multidimensionale favorisce la multidimensionalità del ragionamento umano: **Che cosa? Chi? Quando? Come? Risultato?....**

## Gerarchia delle dimensioni

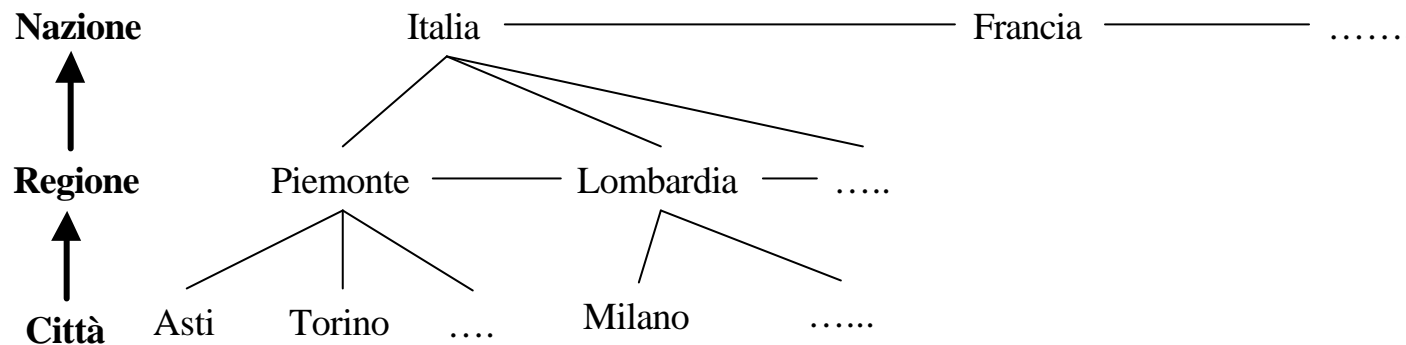
Ogni dimensione può essere strutturata in una **gerarchia** di variabili che **rappresentano diversi livelli di aggregazione**.

La pianificazione delle gerarchie, come quella delle dimensioni, è un passo cruciale del progetto di un sistema decisionale.

Esempio: Dimensione **punto vendita** (Asti, Milano, Torino)

Variabile della  
dimensione

Valori



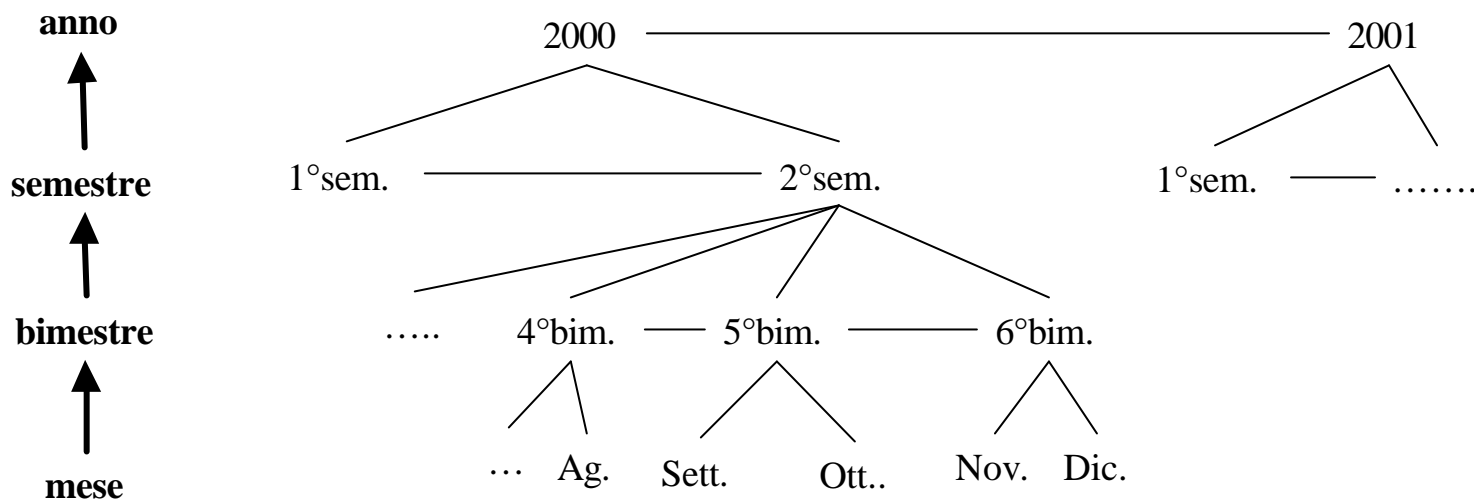
Molto articolata, di norma, la gerarchia sulla dimensione tempo, parametro ricorrente di analisi dei dati a fini decisionali.

A titolo di esempio possiamo supporre la tabella di fatto con dati per ogni mese.

La dimensione tempo può essere strutturata per bimestri, semestri, anni, ecc...

Variabile della  
dimensione

Valori



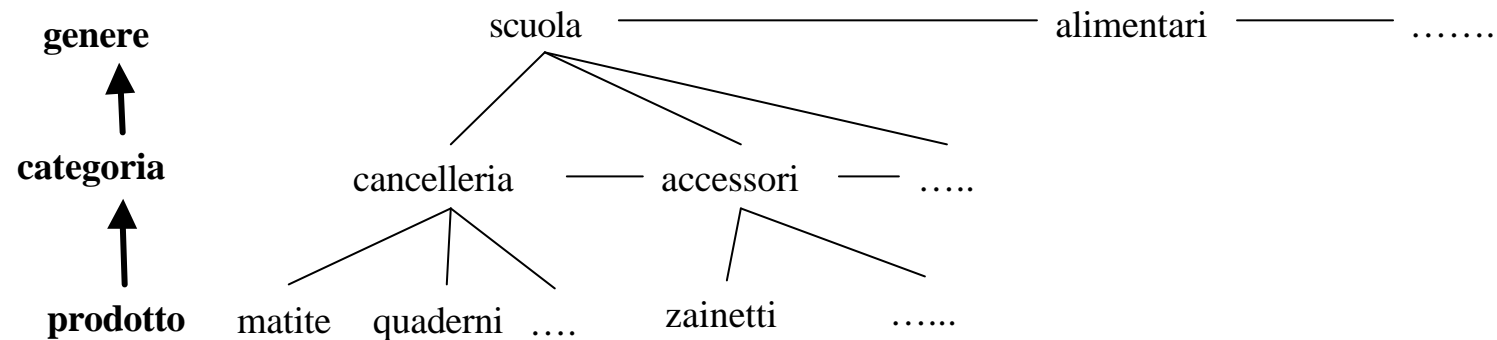
Anche i prodotti possono essere strutturati in gerarchie.  
A titolo di puro esempio si abbia:

Dimensione

Tipo prodotto (matite, quaderni, zainetti)

Variabile della  
dimensione

Valori



Il modello multidimensionale può essere esplorato con varie forme di aggregazione avvalendosi delle gerarchie sulle dimensioni.

Prodotti per la scuola:

Totale scuola punto vendita/mese		punto vendita				
		Asti	Milano	Torino		totali
tempo	Sett.	0	3800	3300		7100
	Ott.	400	7100	3400		10900

Altra aggregazione: il venduto in Italia:

Totale Italia tipo prodotto/mese		tipo prodotto				
		matite	quaderni	zainetti		totali
tempo	Sett.	1000	600	5500		7100
	Ott.	1200	1800	7900		10900

Vendite regionali dei prodotti per la scuola:

Analisi scuola Regione/anno		punto vendita		
		Piemonte	Lombardia	totali
tempo	Sett.	3300	3800	7100
	Ott.	3800	7100	10900

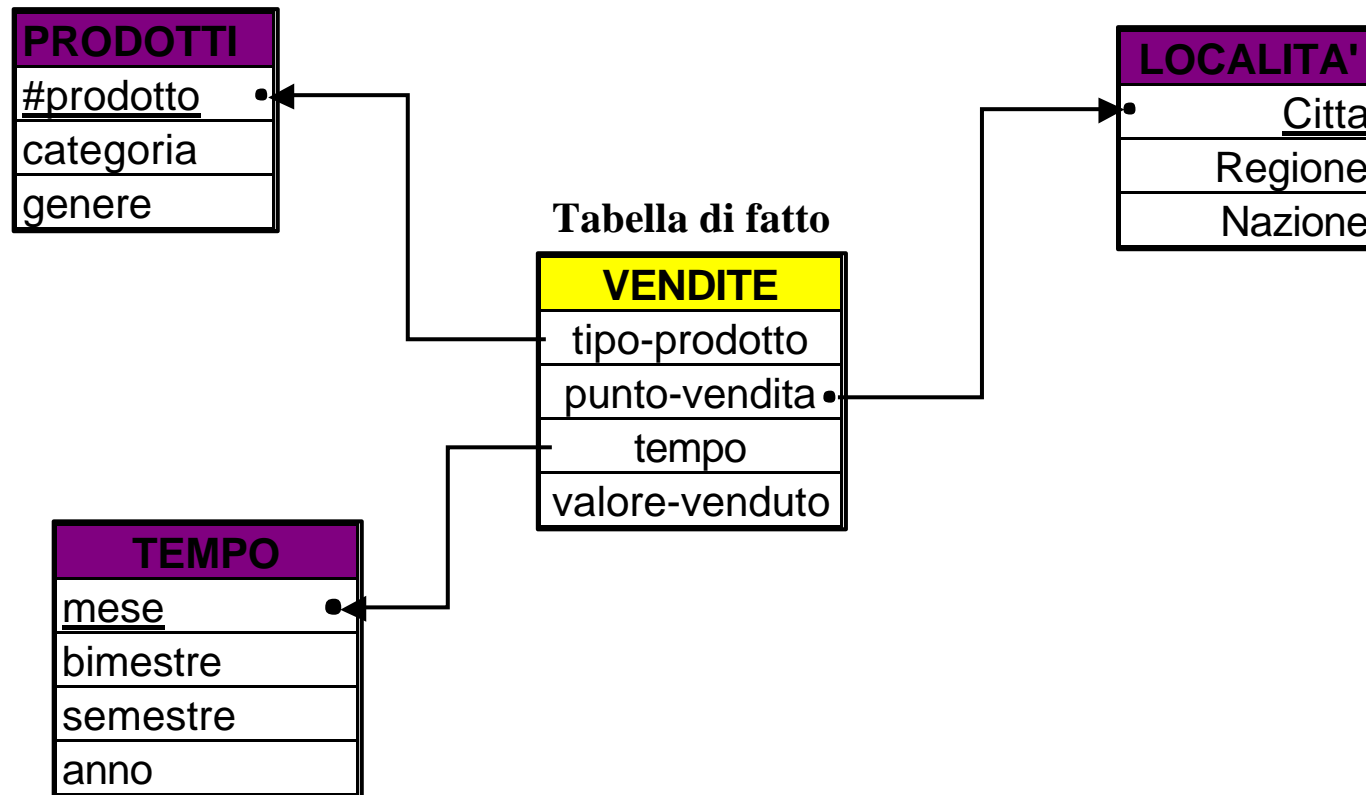
Vendite bimestrali in Italia:

Italia		tipo prodotto			
tipo prodotto/bimestre		matite	quaderni	zainetti	totali
tempo	5° bim.	2200	2400	13400	18000

Analisi 5°bim		punto vendita		
		Asti	Milano	Torino
tipo prodotto	matite	0	1500	700
	quaderni	0	1400	1000
	zainetti	400	8000	5000

Schema relazionale “a stella”:

Tabella dei fatti e tabelle con la descrizione delle dimensioni gerarchizzate.



Si osservino le tabelle denormalizzate

L'analisi multidimensionale dei dati è chiamata  
**OLAP**: On Line Analytical Processing.

### Operatori OLAP

Gli operatori OLAP sono alla base delle applicazioni DSS e EIS

**Roll-up**: diminuisce il dettaglio aggregando lungo la gerarchia della dimensione.

Es.: indicatori aggregati su Regione, Nazione, bimestre, ecc...

**Drill-down**: aumenta il dettaglio della dimensione (zoom).

Es.: dopo aver analizzato i dati aggregati per Regione si richiede un dettaglio a livello di città.

**Push-down**: elimina una dimensione riclassificandola misura in aggiunta alle altre misure.

Esempio di operatore **push-down**: **Tipo-prodotto** riclassificato misura, mentre rimangono le dimensioni **punto-vendita e tempo**.

Contare i tipi diversi dei prodotti venduti per città e mese:

#tipi prodotti punto vendita/mese		punto vendita		
		Asti	Milano	Torino
tempo	Sett.	0	2	3
	Ott.	1	3	3

Altri operatori OLAP:

**Pull:** inverso dell'operatore Push, classifica dimensione una misura.

**Slice:** già incontrato, distrugge un descrittore "sezionando" l'ipercubo in corrispondenza di un particolare valore del descrittore cancellato.

**Dice:** limita una dimensione considerando solo un range dei suoi valori.

**Select:** seleziona, con predicati, i valori d'interesse dei domini.

**Pivoting:** esplorazione dell'ipercubo da punti di vista diversi.

## Quando il modello multidimensionale è inadeguato

Coppie di dimensioni legate da relazioni funzionali sono inadeguate.

Es.: Città e relativo cap.

Ogni riga della vista bidimensionale ha al più un valore.

### Coppie di dimensioni inadeguate

punto vendita	cap	valore del venduto
Torino	10100	1000
Milano	20100	1500
Asti	14100	700
Susa	10059	5000
Roma	10000	1400
Genova	16100	8000
Novara	28100	400

		cap						
		10000	10059	10100	14100	16100	20100	28100
punto vendita	Asti				700			
	Genova					8000		
	Milano						1500	
	Novara							400
	Roma	1400						
	Susa		5000					
	Torino			1000				

## Scenari decisionali

Gli scenari decisionali sono legati ai modelli di analisi (Codd) adottati:

- **Exegetical**, uso statistico dei dati estratti.
- **Categorical**, classificazioni drill-down o roll-up.
- **Contemplative**, o analisi “what if” di variabili dipendenti in funzione di variabili indipendenti istanziate opportunamente.
- **Formulaic**, fissate le variabili dipendenti si chiede quali valori devono assumere le variabili indipendenti per poter ottenere i risultati predeterminati.

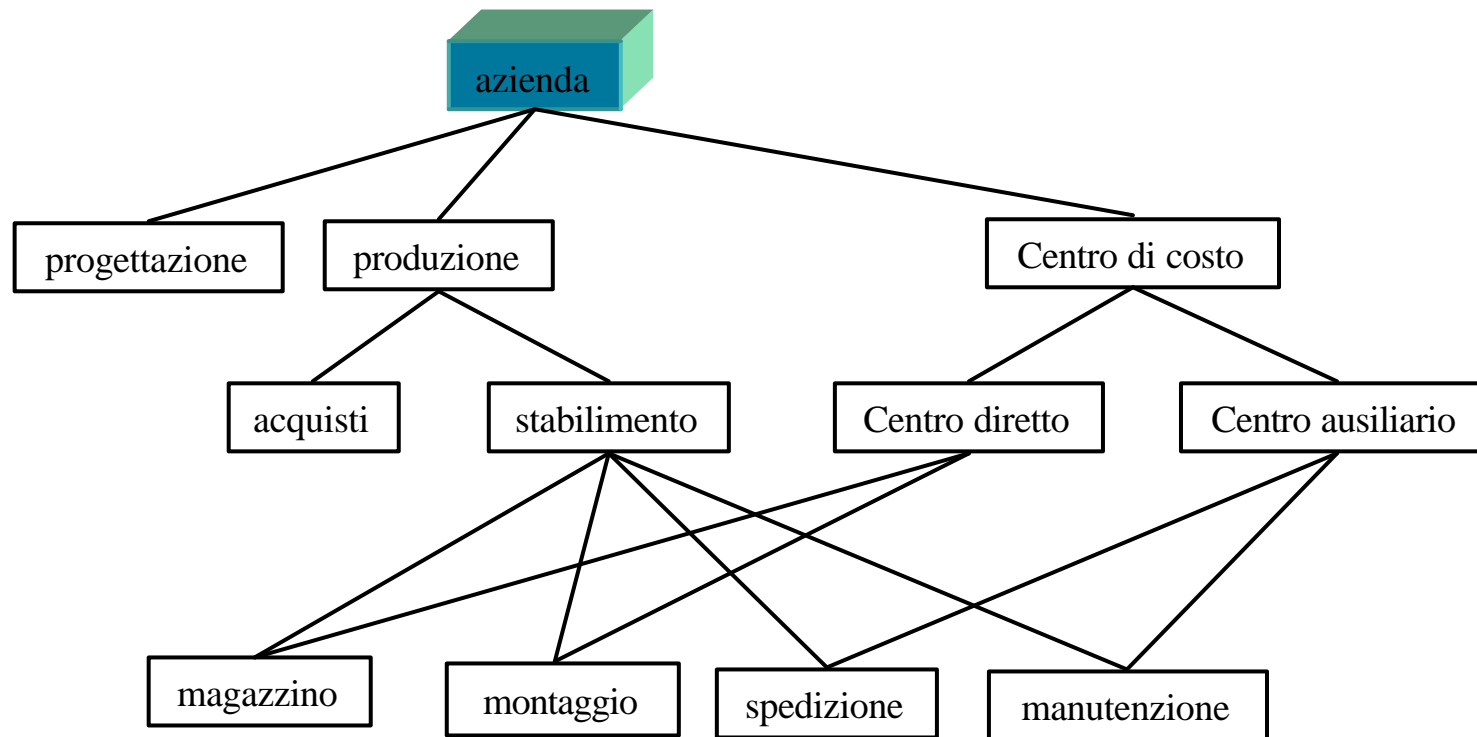
In queste lezioni ci limitiamo ad approfondire le tipiche dimensioni di un'azienda

Il data warehouse aziendale può comprendere più tabelle di fatto con relative gerarchie anche condivise. Sono detti **Data mart** insiemi di informazioni (fatti, dimensioni, aggregati tratti da tabelle di fatto,...) rilevanti per una particolare divisione o area di business aziendale, o anche per una particolare categoria di oggetti.

La scelta delle dimensioni e di una loro efficace composizione è vitale nella caratterizzazione dell'informazione aziendale.

Le gerarchie possono aggregare dati in vario modo.  
Uno stesso dato può partecipare a molteplici gerarchie.

Esempio: Dimensione Centri di costo





La **dimensione responsabilità** risponde all'esigenza di articolare gli indicatori secondo le strutture organizzative aziendali.

Aggregazione tipica: suddivisione regionale.

La dimensione responsabilità costituisce un elemento fondamentale della contabilità direzionale. La gerarchia organizzativa è accompagnata da una gerarchia di conti economici.

<b>mese</b>	<b>Divisione 1</b>			<b>Divisione ...</b>			<b>Azienda</b>			
<b>Voci di conto economico</b>	budget	eff.	diff.	budget	eff.	diff.	budget	eff.	diff.	anno prec.
vendite										
acquisti										
costo personale diretto										
ammortamenti										
altri costi										
voci varie										
utile										
<b>Indicatori operativi</b>										
N.ro auto fatturate										
N.ro auto prodotte										



La **dimensione cliente** è critica per comprendere la competitività dell'azienda. La segmentazione in classi è spesso intersecata con la dimensione prodotto (*Balanced Score Card*).

I totali degli indicatori qualitativi si ottengono con medie pesate.

	Cliente 1									.....
mese	Prodotto 1			Prodotto ...			Azienda			
Voci di conto economico	budget	eff.	diff.	budget	eff.	diff.	budget	eff.	diff.	anno prec.
vendite										
costo prodotto venduto										
ammortamenti										
altri costi										
voci varie										
utile										
<b>Indicatori operativi</b>										
N.ro auto fatturate										
N.ro ordini ricevuti										
Tempo consegna										
Indice soddisfazione										

**Dimensione attività e processi** gestionali. Ha lo scopo di disaggregare costi, ricavi e prestazioni dell'azienda lungo le attività della catena del valore.

Importanti gli incroci con la dimensione responsabilità e prodotto.

Sono significativi indicatori di efficienza. Es.: costo per unità di output.

La segmentazione per processo permette di controllare le prestazioni aziendali da una prospettiva complementare a quella dei centri di responsabilità.

Utile nella metodologie *Activity Based Costing* (ABC) che lega i costi aziendali alle attività, anziché ai centri di responsabilità.

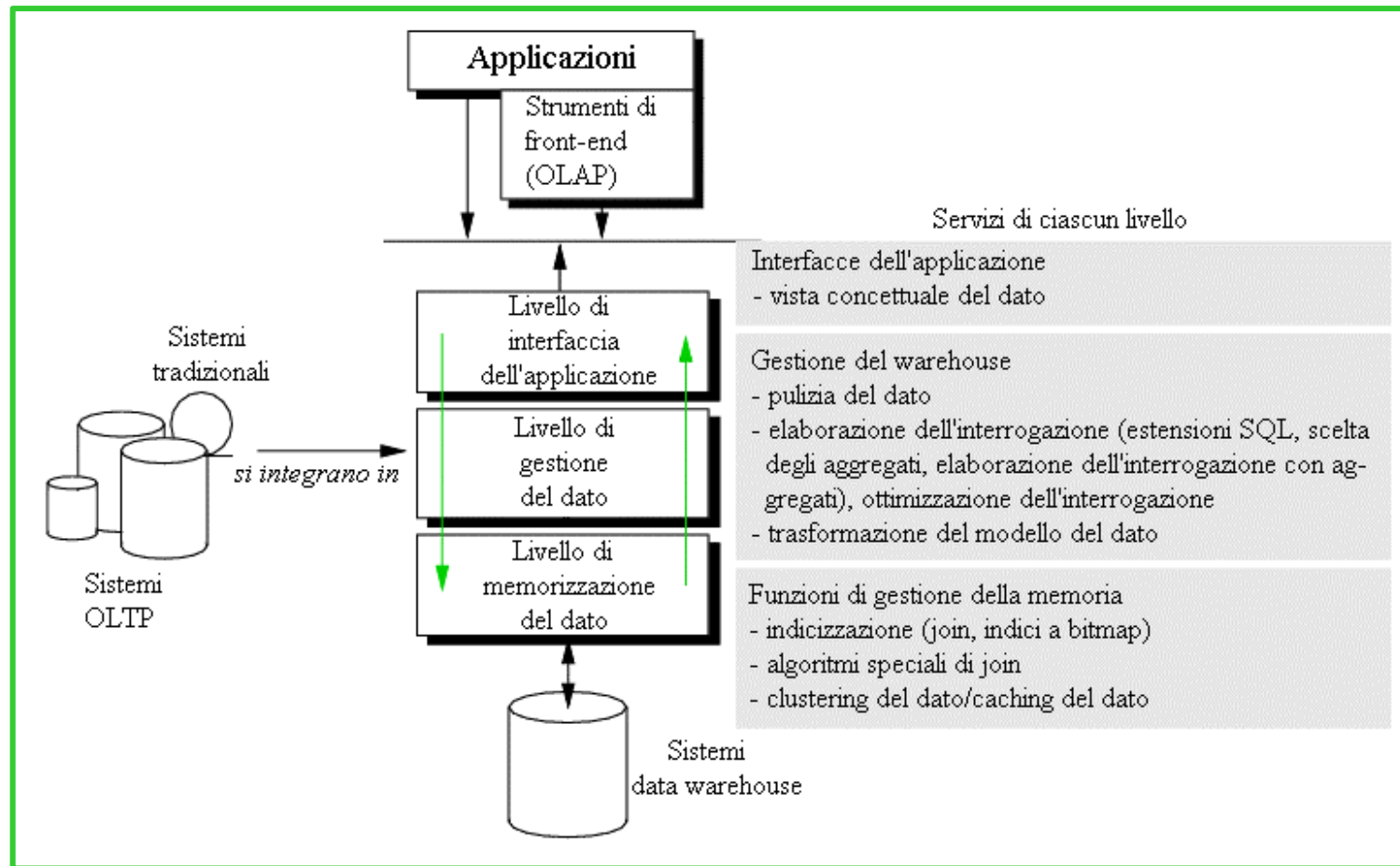
Nel seguente esempio le attività sono raggruppate in “**segmenti**” aziendali.

	Prodotto 1									
mese	Segmento 1			Segmento ...			Azienda			
Voci di conto economico	budget	eff.	diff.	budget	eff.	diff.	budget	eff.	diff.	anno prec.
vendite										
costi mat. prime										
costo trasformazione										
costo vendita										
costo assistenza										
costi comuni										
costo progettazione										
costo ricerca										
<b>Indicatori operativi</b>										
N.ro auto prodotte										
N.ro ordini ricevuti										
Indice qualità										
<b>Approvvigionamento</b>										
N.ro ordini emessi										
N.ro fornitori gestiti										
Organico										
<b>Trasformazione</b>										
.....										

## Tecnologie dei sistemi informativi direzionali

Molti autori preferiscono parlare di **data warehousing** piuttosto che di data warehouse perché non si tratta di una mera tecnologia (prodotti, interfacce, ecc...), ma piuttosto di un aggregato di processi operanti con strumenti ad hoc su dati esistenti che consentono l'estrazione, l'adattamento e la memorizzazione nelle tabelle di fatto dei dati di interesse gestite da specifici data base.

## Architettura logica di un data warehousing

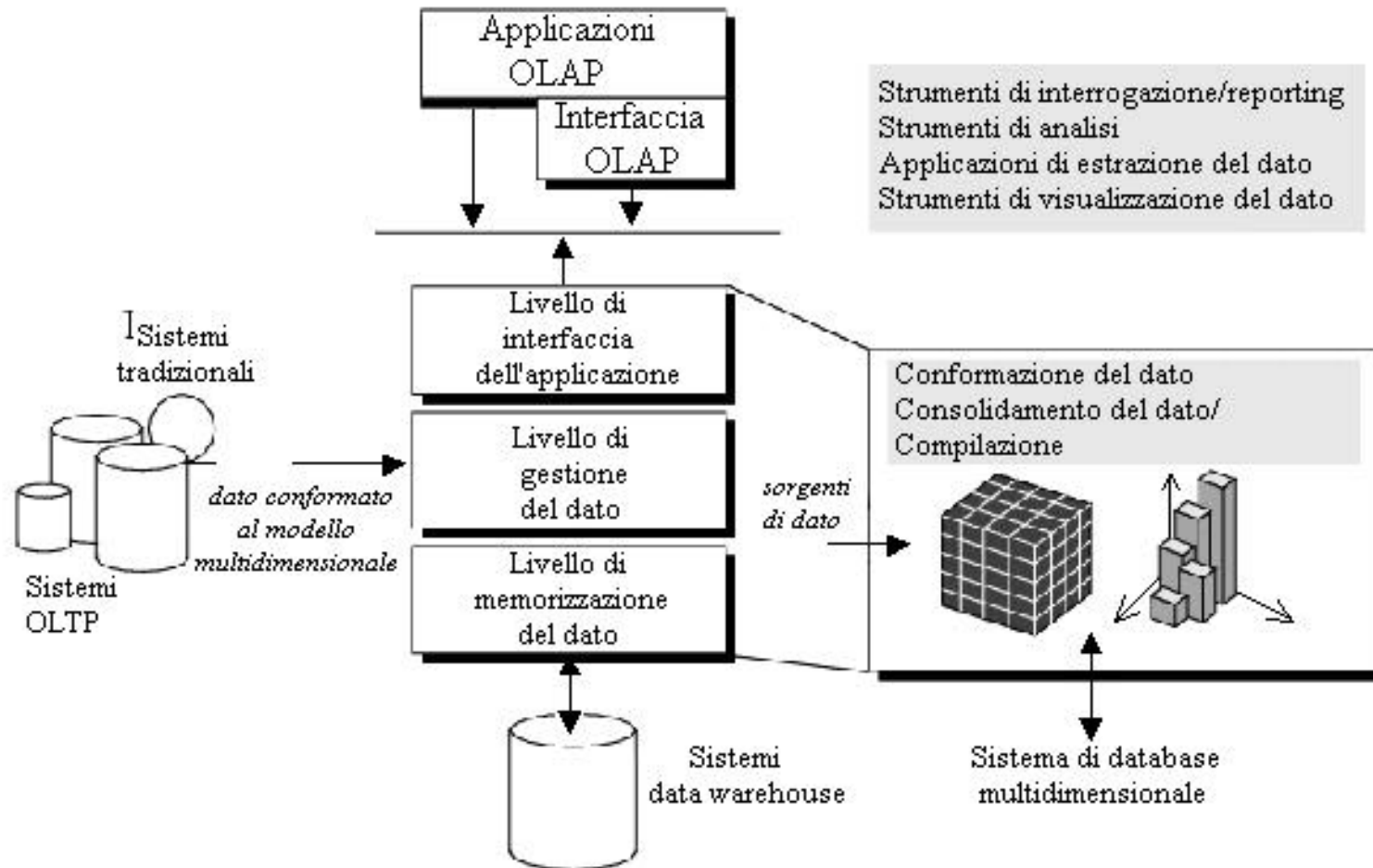


OLTP: On line Transactional Processing

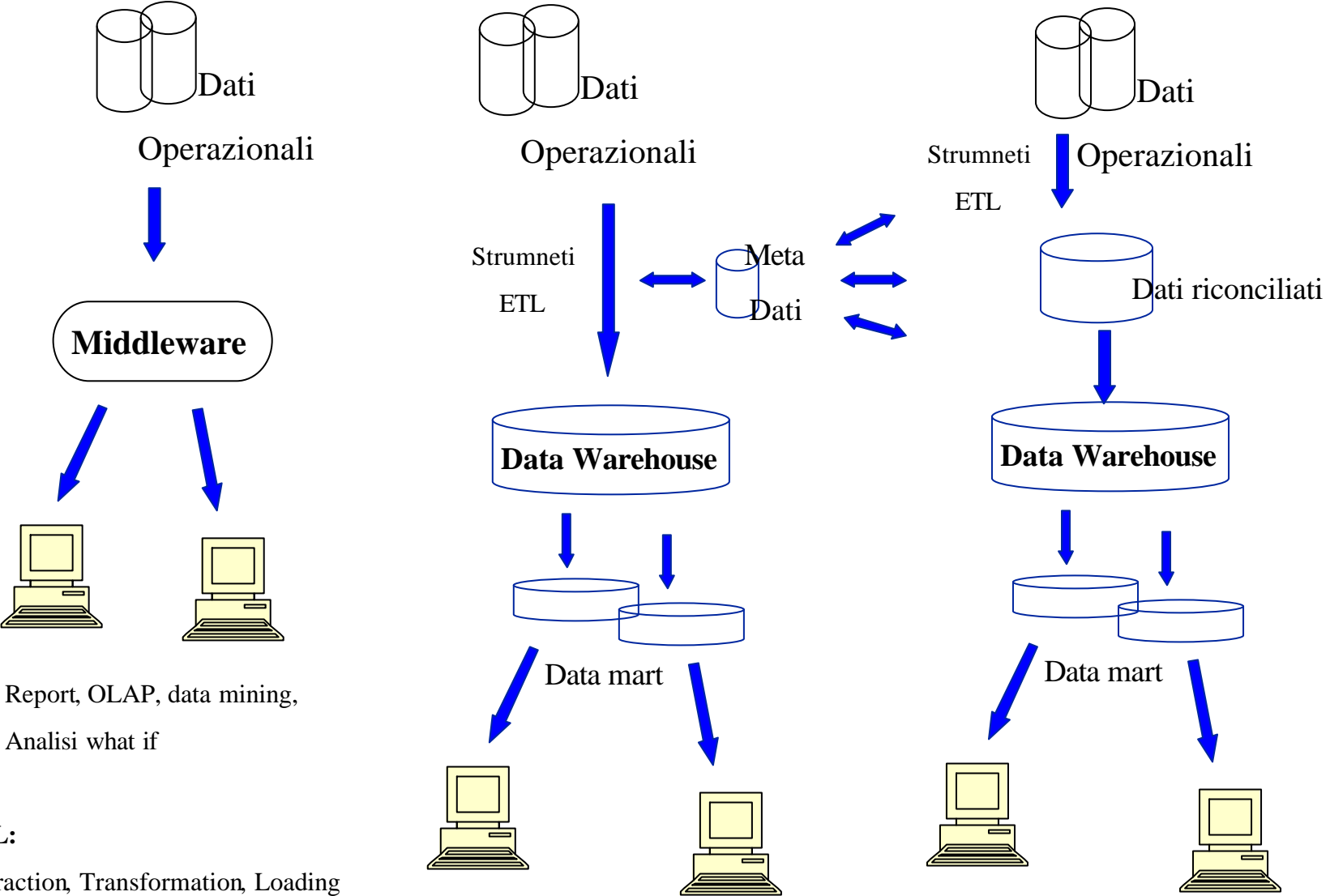
Architettura (2) del warehouse:

ROLAP, ossia implementazione su DBMS relazionale

MOLAP, ossia implementazione su DBMS specializzato multidimensionale



# Architettura (3) del warehouse: a uno, due e tre livelli



**ETL:**  
Extraction, Transformation, Loading

# I metadati

I metadati indicano sorgenti, uso, trasformazione dei dati nel DW

Schemi logici e fisici del DW, i vincoli

Unità di misura, le funzioni di aggregazione

Profili utente.

Strumenti di gestione dei metadati, possibilmente con interfaccia grafica:

Interrogazione

Esportazione ed importazione secondo lo standard CWM

CWM: Common Warehouse Metamodel sviluppato da OMG

che si appoggia su UML e XML

## Differenze tra database operazionali e data warehouse

<b>Aspetti</b>	<b>Database operazionali</b>	<b>Data Warehouse</b>
Utente	Progettista di sistema, amministratore di sistema, impiegato, addetto al data entry	Costruttore di decisioni, operatore della conoscenza, esecutivo
Funzione	Operazioni giornaliere, elaborazione della transazione on-line (OLTP)	Supporto alla decisione, elaborazione analitica on-line (OLAP)
Progetto di DB	Orientato all'applicazione	Orientato al soggetto
Dato	Corrente, aggiornato, atomico, relazionale (normalizzato), isolato	Storico, riassuntivo, denorm. multidimensionale, integrato
Uso	Ripetitivo, di routine	Ad hoc
Accesso	Transazione semplice di lettura/scrittura (solitamente coinvolge 1-3 tabelle)	Interrogazione complessa principalmente di lettura (coinvolge più tabelle)
Requisiti di sistema	Output prodotto dalla transazione, consistenza del dato	Output prodotto all'interrogazione, precisione del dato

***Fine DSS***