



Sistemi Informativi Aziendali

Umberto Nanni

Introduzione al Data Warehousing per
Sistemi Informativi Aziendali

a. Modello Multidimensionale & OLAP

Cos'è il Data Warehousing

Collezione di metodi, tecnologie e strumenti di ausilio al “lavoratore della conoscenza” (manager, gestore, analista) per condurre analisi dei dati finalizzate all’attuazione di processi decisionali e/o a migliorare la gestione del patrimonio informativo.

Cos'è un Data Warehouse

Un Datawarehouse è una collezione di dati

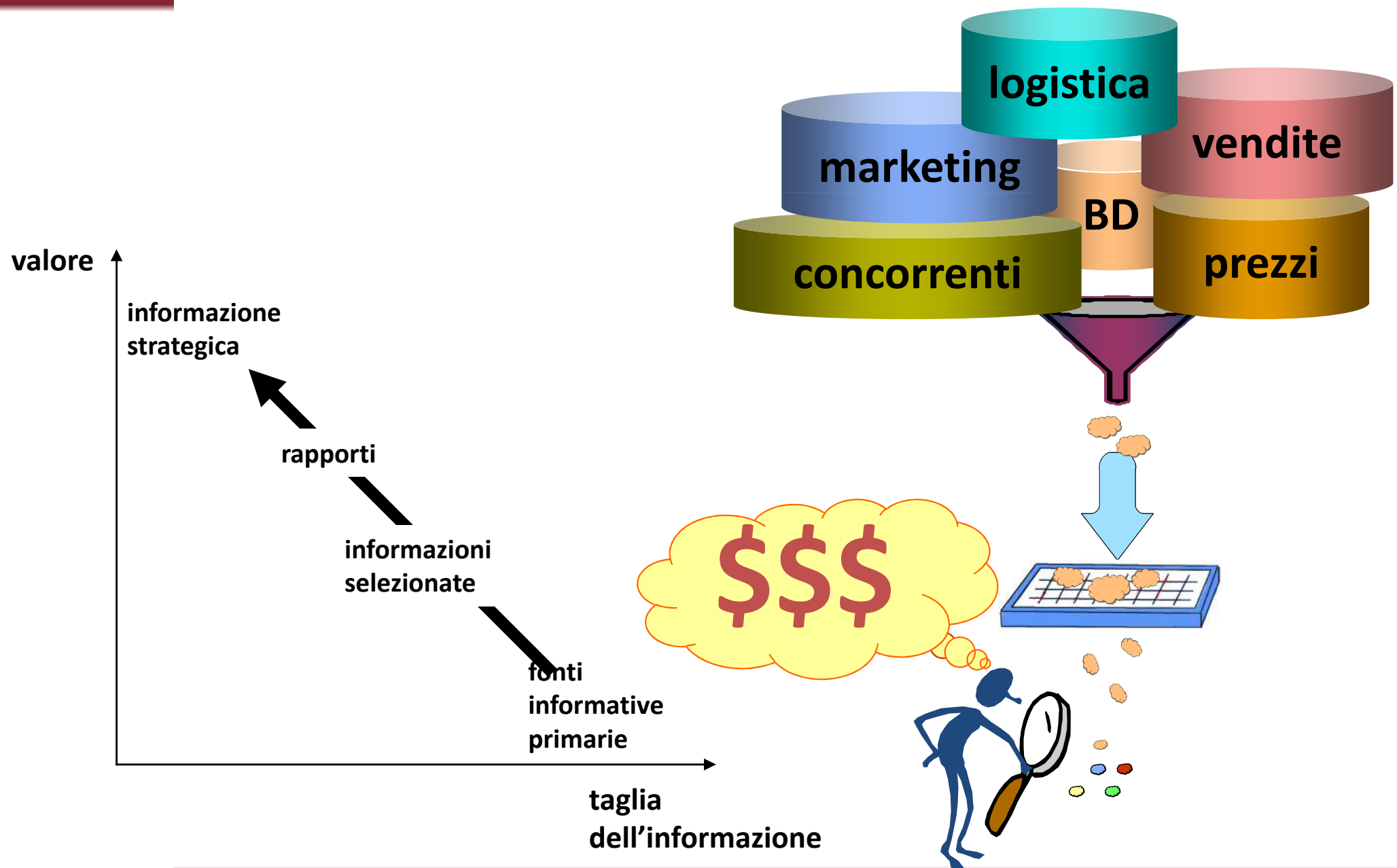
- integrati (anche oltre l'orizzonte dell'organizzazione)
- consistenti (nonostante l'origine eterogenea)
- focalizzati su un'area di interesse
- articolati su un orizzonte temporale significativo
- permanenti (non volatili, storicizzati)

Finalità del Data Warehouse

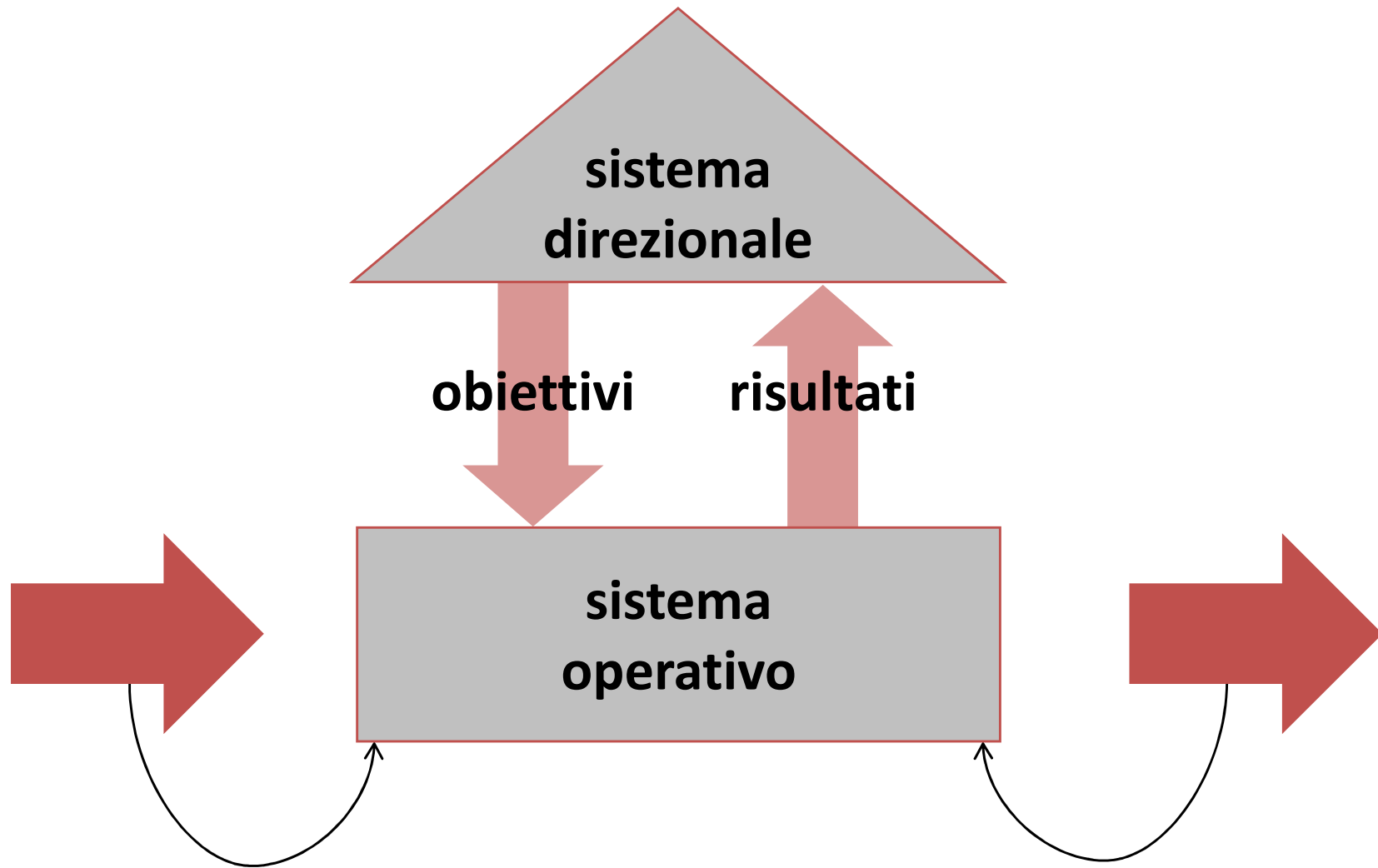
L'analisi dei dati alimentata da un Data Warehouse offre supporto per:

- prendere decisioni
- individuare ed interpretare fenomeni
- fare previsioni sul futuro
- controllare un sistema complesso

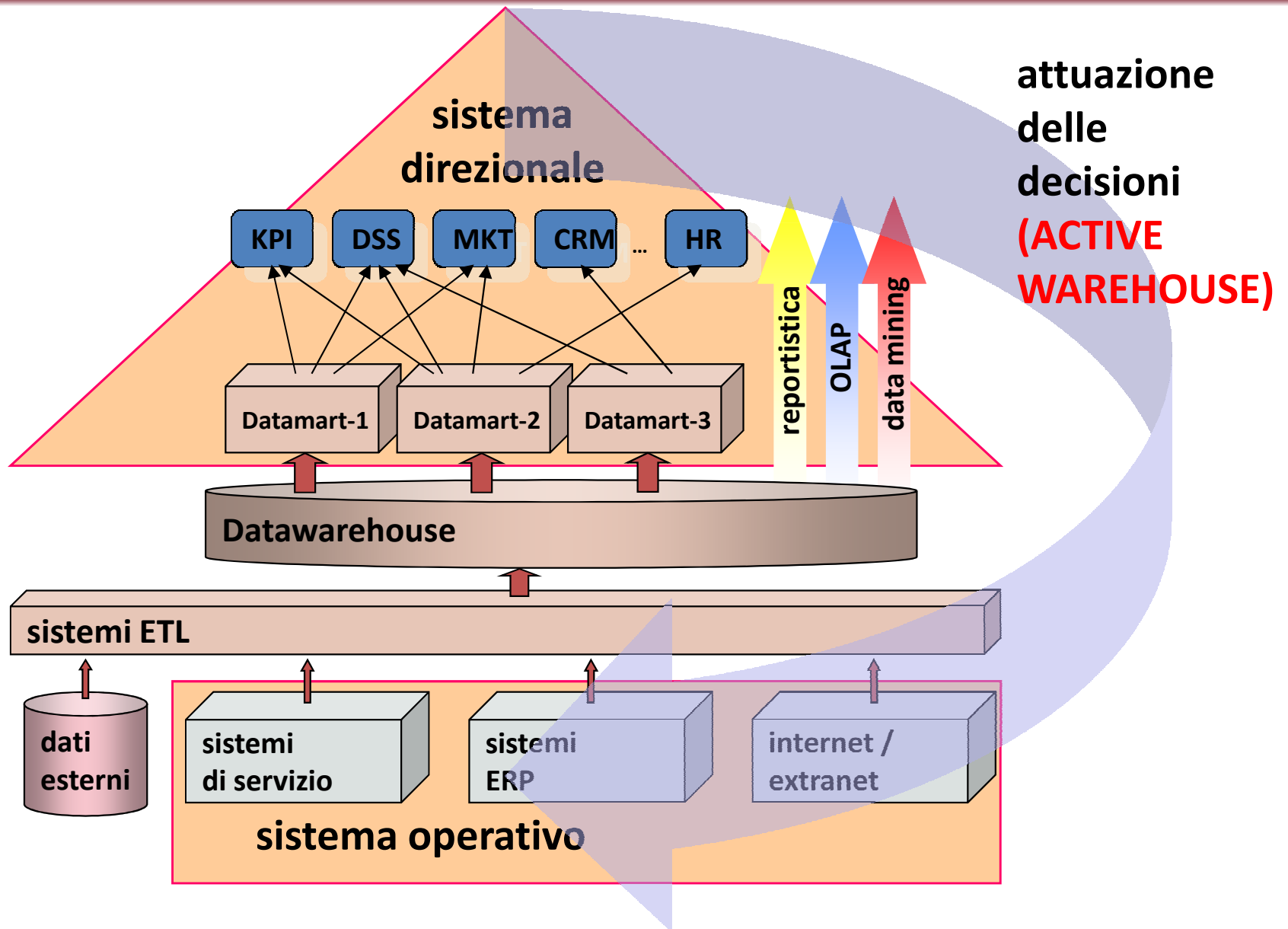
Valore e quantità di informazione



Macrostruttura di un'organizzazione



Ruolo del DW nell'organizzazione



OLTP & OLAP

OLTP - On-Line Transaction Processing

- transazioni, recovery, consistenza (scrittura e/o lettura)
- brevi, frequenti, elevato livello di concorrenza
- intervento su piccole quantità di dati
- dati accurati aggiornati all'istante

OLAP - On-Line Analytical Processing

- operazioni di sola lettura
- poche operazioni, basso livello di concorrenza
- intervento su grandissime quantità di dati
- dati articolati nel tempo ma essenzialmente statici

Separazione tra (OLTP) e (OLAP)

- carico computazionale completamente differente
- accorgimenti differenti:
 - DB: delicata sincronizzazione delle attività
 - DW: dati statici (aggiornamenti occasionali)
- integrazione con attività aziendale:
 - DB operativo: integrazione
 - DW: descrizione, storicizzazione
- dati di interesse
 - DB: minimale
 - DW: massimale

Alcuni ambienti per il Data Warehouse

- Oracle
- IBM DB2-UDB
- Microsoft SQL-Server
- Sybase IQ
- Microstrategy
- Teradata (NCR)
- Netezza – Cognos
- Business Objects (acquisito da SAP, Ottobre 2007)

Architettura per Datawarehouse: aspetti in gioco

- separazione OLTP, OLAP
- scalabilità
- estendibilità
- sicurezza
- amministrabilità

Scelta architettura per il Datawarehouse

- determinata da scelte progettuali
- condizionata da / condiziona scelta del prodotto
- condiziona costo / rende possibile alternative ed integrazioni future (quantitative e/o qualitative)
- condiziona costo del trattamento dati
- condiziona effettiva PERMANENZA dati

Data Mart

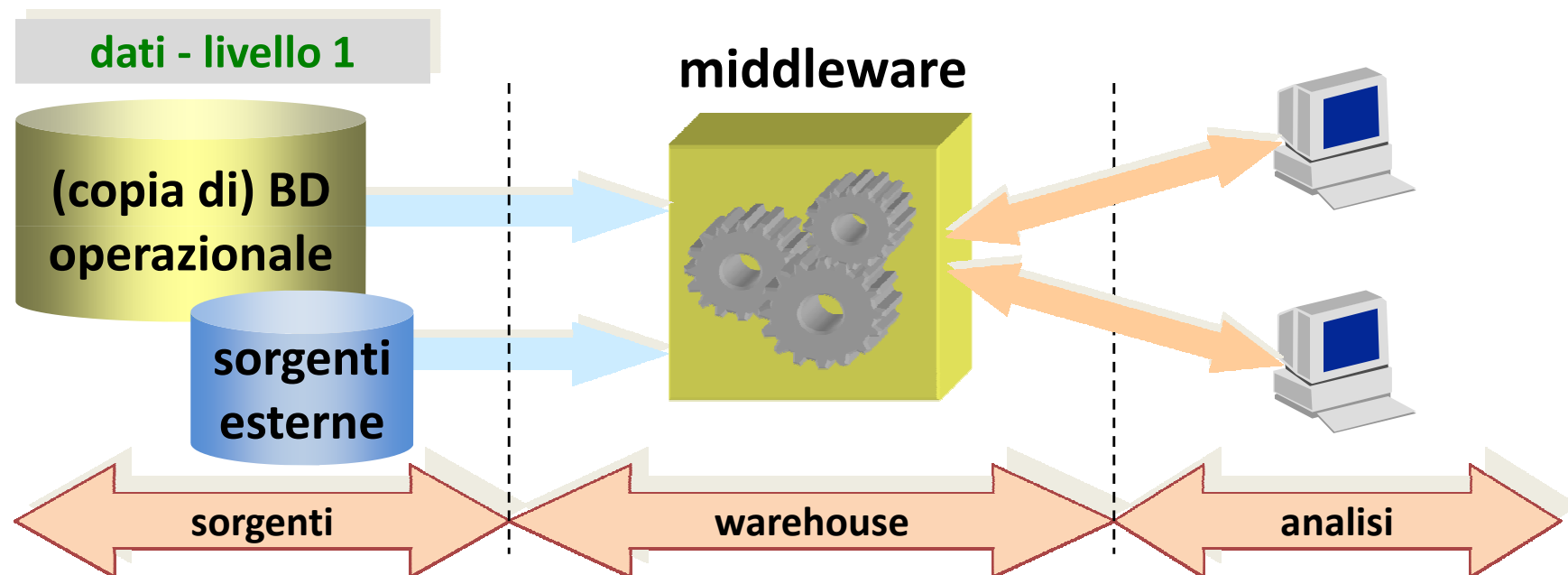
Collezione di dati focalizzata su particolare profilo di utente o su particolare obiettivo di analisi

Alternative:

1. Data Mart dipendente: il DM è sottoinsieme e/o aggregazione di dati presenti nel DW primario
→ DM estratto da DW
2. Data Mart indipendente: il DM è sottoinsieme e/o aggregazione di dati presenti nel DB operativo
→ $DW = \bigcup_i (DM_i)$, ossia il DM è parte del DW
3. soluzione ibrida, combinando le precedenti

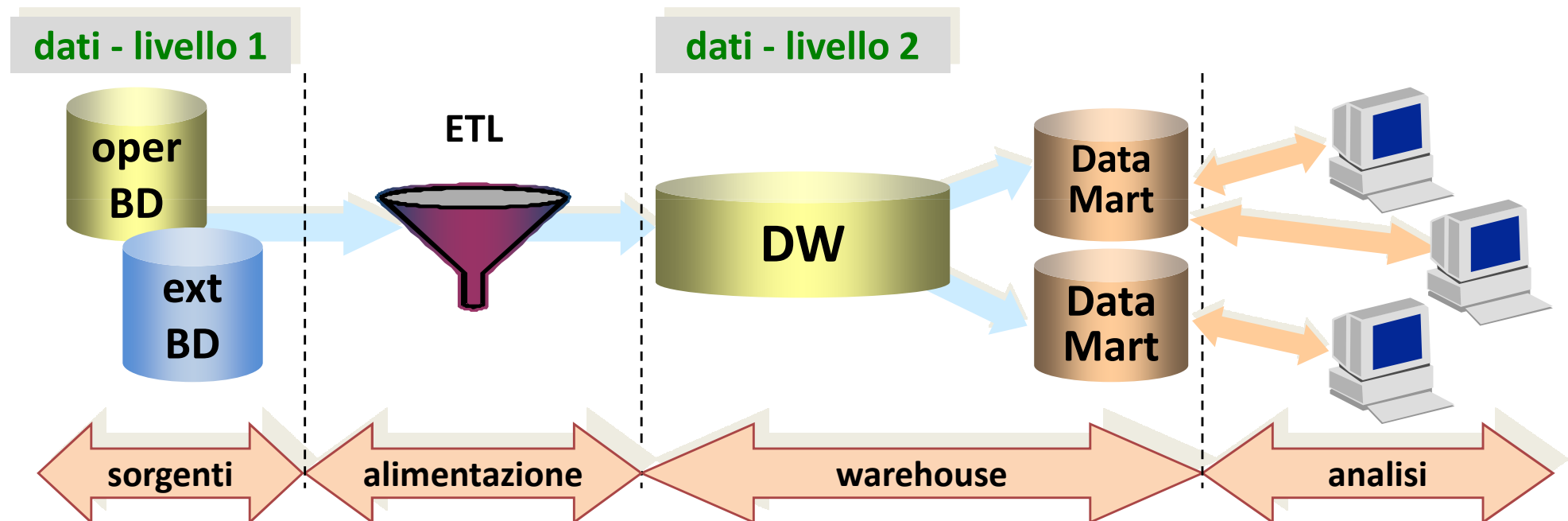
Architettura DW: 1 Livello

- esiste solo il DB operativo
- DW virtuale (non rispetta separazione OLTP-OLAP)
- dati coincidenti con DB operativo
- difficile integrazione con altre sorgenti



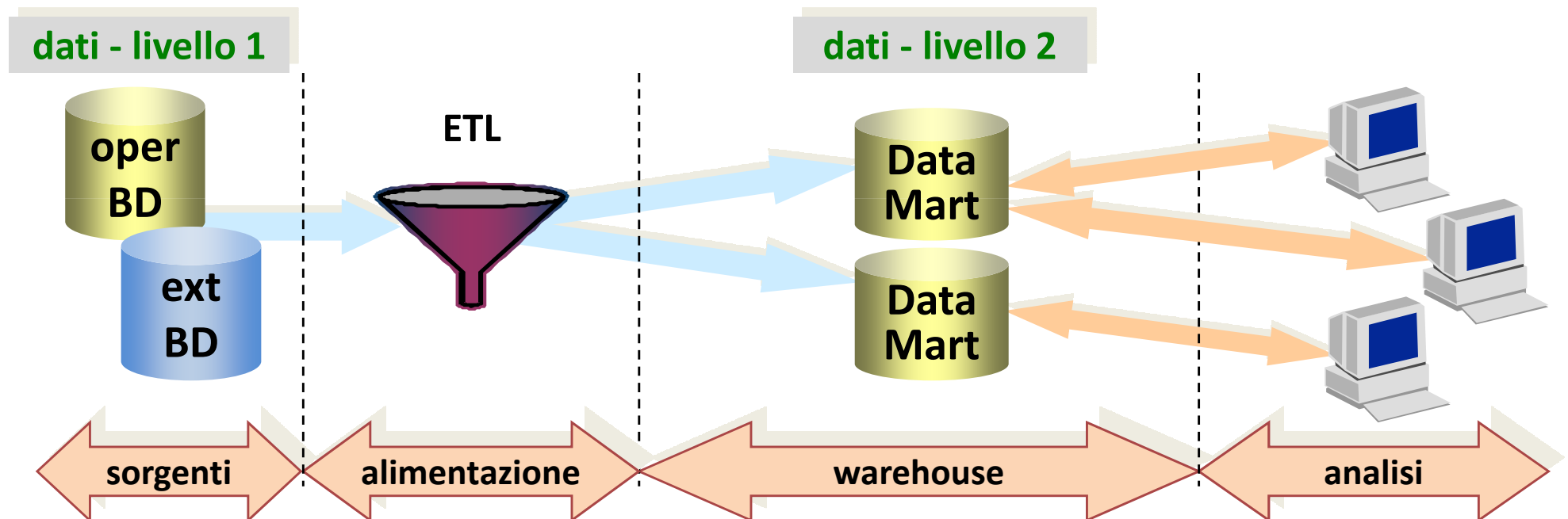
Architettura DW: 2 Livelli – DM dipendenti

- fonti dati integrate con sorgenti esterne
- esecuzione su piattaforma software dedicata
- ETL: Extraction, Transformation, Loading
- materializzazione del DW
- realizzazione e materializzazione di Data Mart



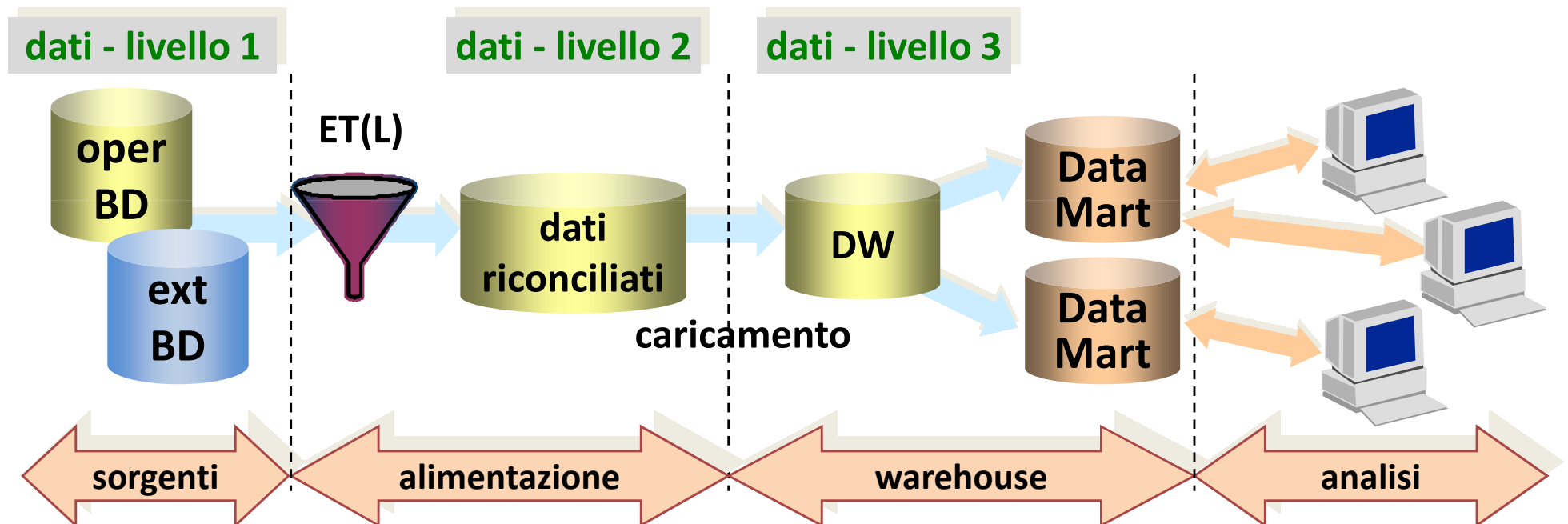
Architettura DW: 2 Livelli – DM indipendenti

- vengono materializzati i soli Data Mart
- non esiste un DW primario materializzato
- DW = unione dei DM



Architettura DW: 3 Livelli

- viene introdotto il livello dei dati riconciliati (operational data store)
- separazione in due fasi delle attività ETL:
 1. estrazione/trasformazione
 2. caricamento



Metadati

- metadati interni
- riguardano l'amministratore del DW (es.: sorgenti, trasformazioni, schemi, utenti, etc.)
- metadati esterni
- di interesse degli utenti (es.: unità di misura, le possibili aggregazioni)
- STANDARD
- CWM - Common Warehouse Model (OMG), definito mediante:
 - UML (Unified Modeling Language)
 - XML (eXtensible Markup Language)
 - XMI (XML Metadata Interchange)

ETL: Extraction, Transformation, Loading

Dati Operazionali, Dati Esterni

- estrazione
- pulitura - validazione - filtraggio
- trasformazione

Dati Riconciliati

- caricamento

Data Warehouse

Estrazione

- iniziale:
 - finalizzata alla creazione del DW
- successiva:
 - statica (integrale rispetto alle sorgenti)
 - incrementale
 - log
 - timestamp

Pulitura

- intervento a livello di VALORI:
- duplicati
- inconsistenze
 - violazione dominio
 - violazione dipendenze funzionali
- valori nulli
- uso improprio campi
- ortografia
- abbreviazioni (non omogenee)

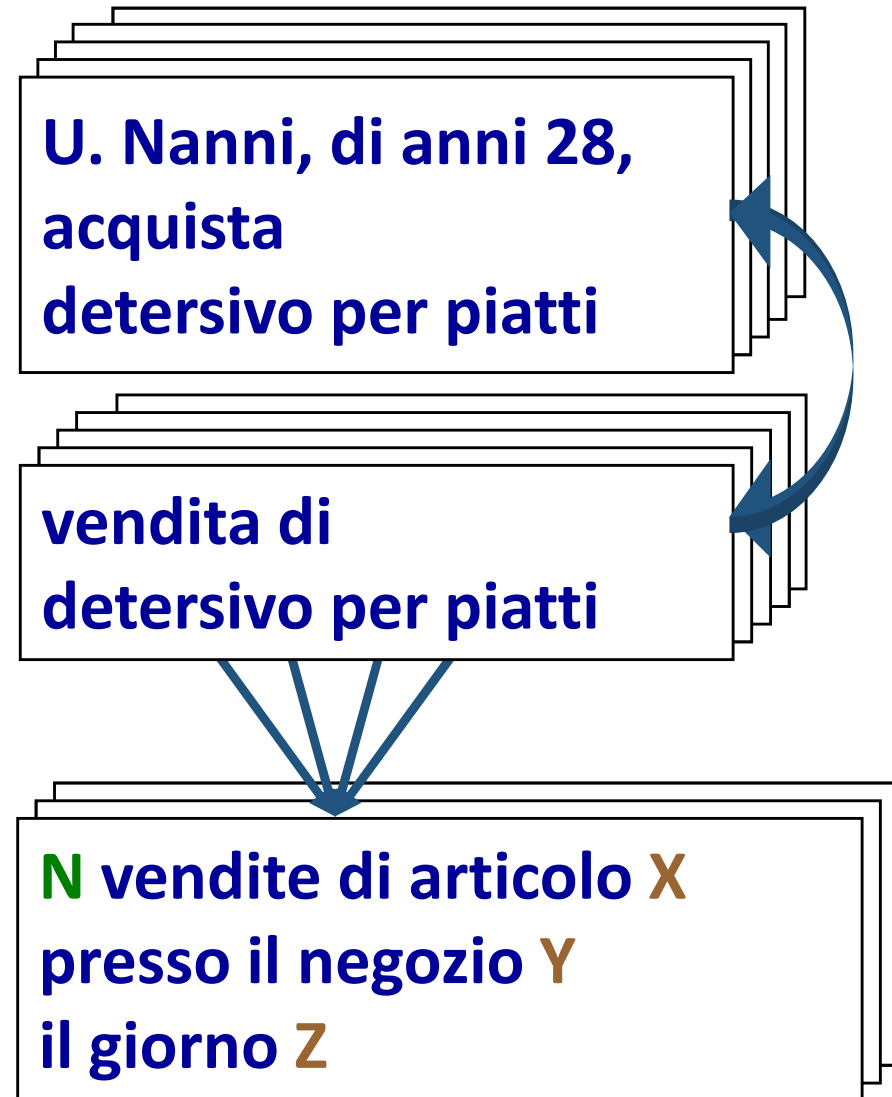
Trasformazione

- intervento a livello di FORMATI:
- disallineamento formati
- overloading campi
- codifiche non omogenee

Caricamento (Loading)

- Refresh:
caricamento ex-novo dell'intero DW
- Update:
modifiche intervenute

Modello multidimensionale



Eventi e Fatti

- **EVENTO** (nella realtà) e **FATTO** (di interesse) sono termini il cui significato è determinato dalla granularità (livello di dettaglio) della rappresentazione multidimensionale del Data Warehouse

Esempio di FATTO:

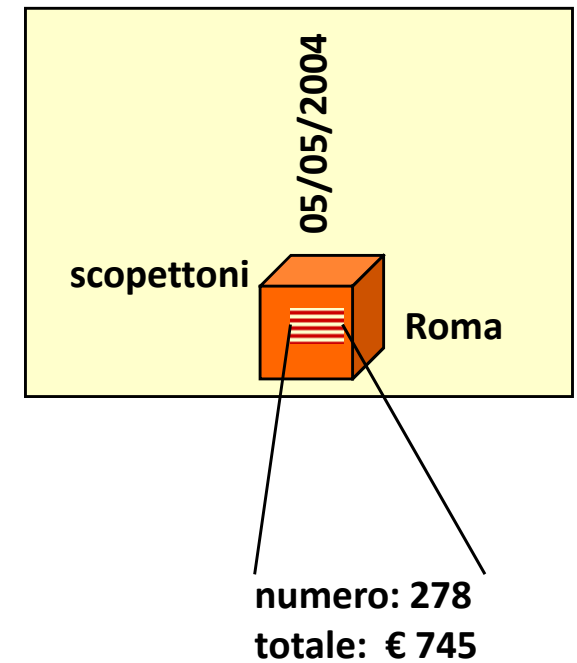
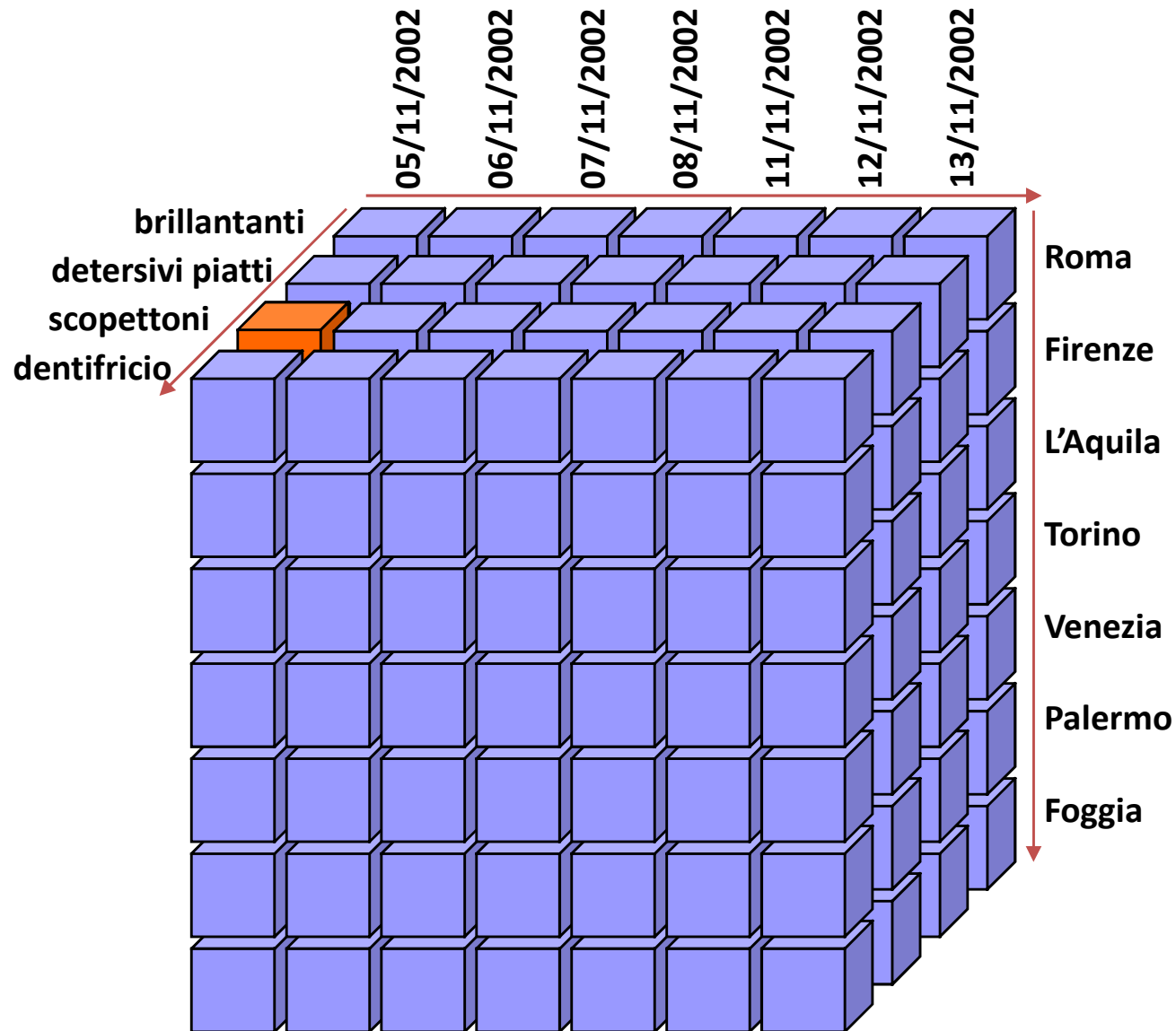
il giorno 02/05/2004

a Roma

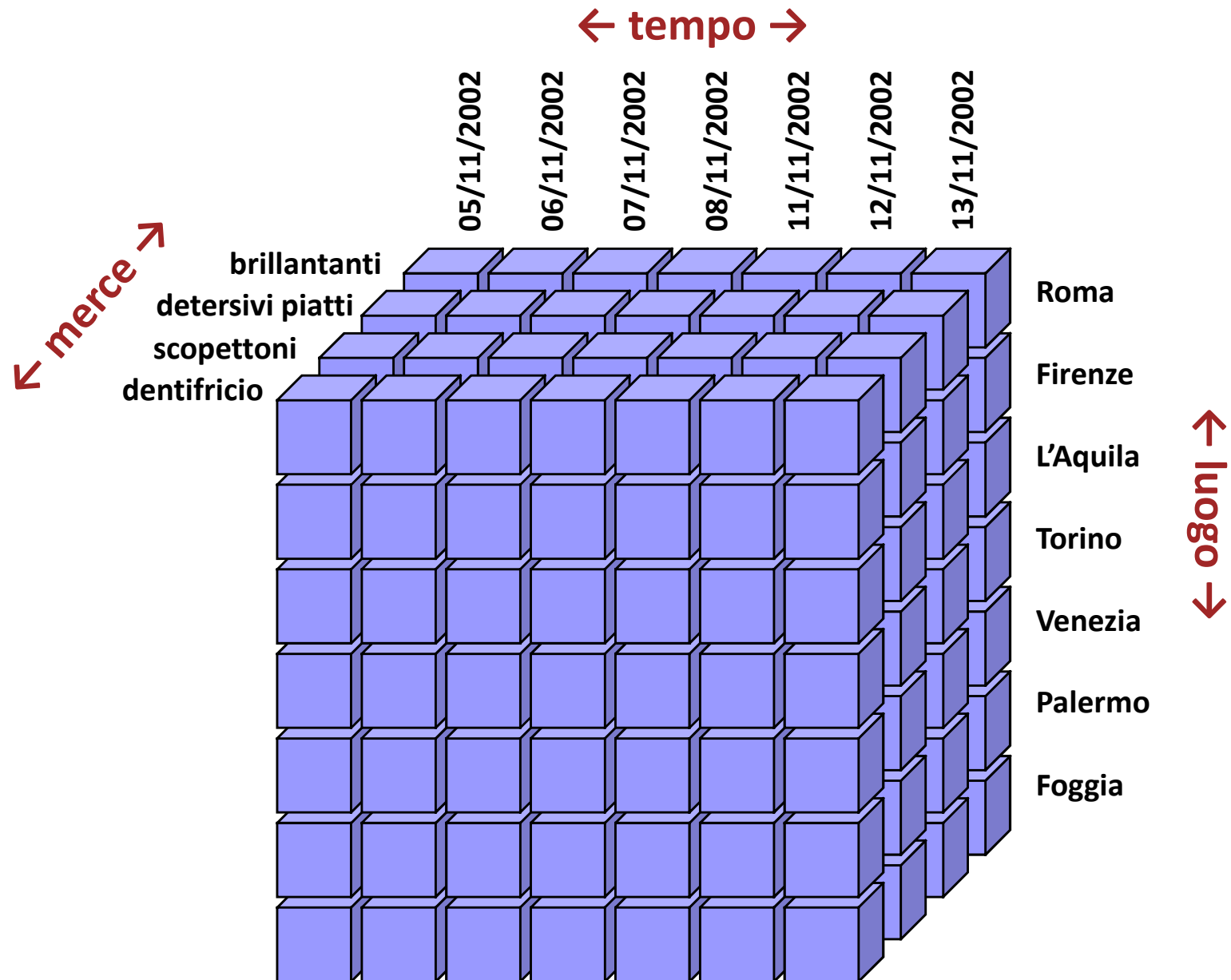
sono stati venduti 278 scopettoni

con un incasso di 745 Euro

Cubo multidimensionale



Dimensioni



Schema relazionale per il cubo multidimensionale

TABELLA(dimens-1, ..., dimens-k, misura-1, ..., misura-h)

Esempio:

VENDITE(prodoto, città, data, numVendite, incasso)

dimensioni

(o attributi di classificazione)

misure

(o variabili, o metriche, o indicatori, ...)



CHIAVE IDENTIFICATIVA

Dipendenza Funzionale: **dimensioni → misure**

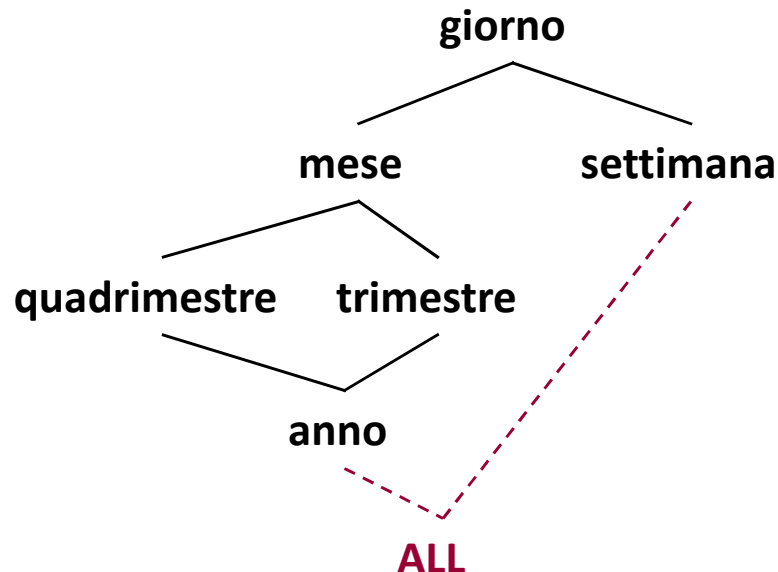
Nell'esempio: **prodoto, città, data → numVendite, incasso**

Gerarchia dimensionale

Ad ogni dimensione viene associata una gerarchia che ne raggruppa i valori a diverso livello di aggregazione (anche in modo ortogonale). Ciascun nodo della gerarchia è detto

ATTRIBUTO (DIMENSIONALE)

Esempio di gerarchia (dimensione temporale):



Valgono le DIPENDENZE FUNZIONALI:

giorno → settimana

giorno → mese

mese → trimestre

mese → quadrimestre

trimestre → anno

quadrimestre → anno

Accesso al Data Warehouse

Reportistica

- esigenze predefinite (ev., parametrizzate)
- estrazione automatica di informazioni

OLAP

- esigenze di analisi non identificabili a priori
- esplorazione interattiva dei dati, alla ricerca di informazioni di interesse

Data Mining

- esigenze di analisi non identificabili a priori
- esplorazione (semi) automatica dei dati

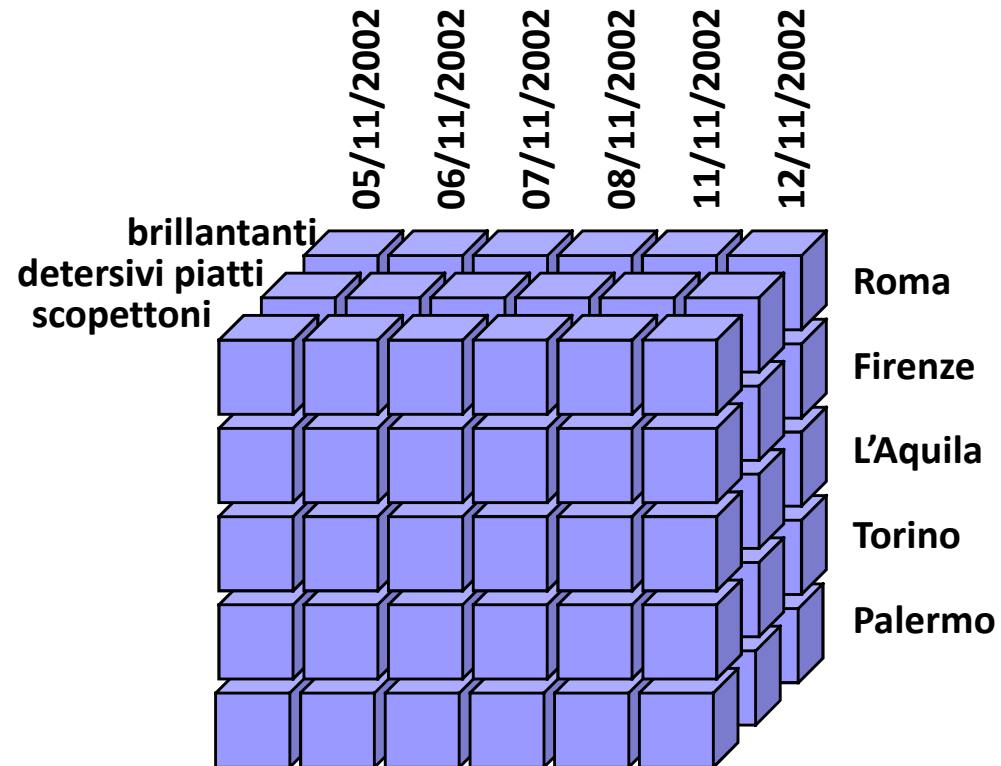
Reportistica

Rapporti/Report di struttura e forma predefinita

- interrogazione
 - tipicamente basata su restrizioni e/o aggregazioni
- presentazione
 - intercalata con testo
 - dati in forma tabellare e/o grafica
 - personalizzata con variabili d'ambiente (data, utente, ...)
- generazione
 - richiesta esplicita dell'utente
 - periodica
 - in base a condizioni predefinite
- distribuzione
 - preview/stampa, e-mail, web, ...

- modalità di attuazione:
 - costruzione di una sessione di analisi, articolata in una serie di passi, ciascuno dipendente dai risultati ottenuti in precedenza
- utenti tipici:
 - esperti del dominio, non necessariamente esperti informatici

OLAP: cubo di esempio



RESTRIZIONE (limitazione puramente quantitativa)

- selezione su specifici valori di attributi dimensionali
- nel caso si imponga l'uguaglianza con un singolo valore, l'operatore prende il nome di SLICING (scompare una dimensione)

AGGREGAZIONE (aumento del livello di astrazione)

- considerata 1 dimensione, raggruppare un attributo dimensionale ad un livello di maggiore granularità
- se, al limite, si aggrega al massimo livello (il più generico, costituito da un singolo valore), scompare una dimensione

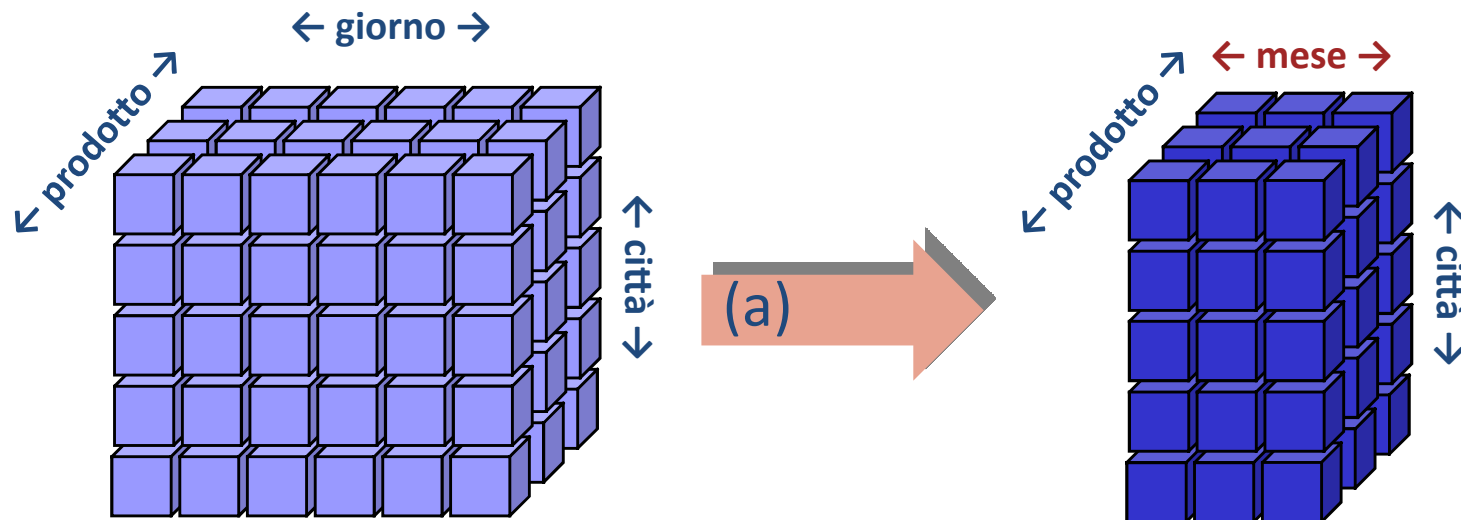
OLAP: Operatore Roll-Up

Aggregazione lungo una dimensione

→ uso di un operatore aggregativo su valori di misura

Esempi:

- a. vendite di ciascun prodotto per città e per mese
- b. vendite giornaliere per città (totale su tutti i prodotti)
- c. vendite giornaliere per prodotto (Nord/Centro/Sud)



OLAP: Operatori di aggregazione

DISTRIBUTIVI: calcolabili a partire da:

1. aggregati parziali
 - Esempi: SUM, MAX, MIN, ...

ALGEBRICI: calcolabili a partire da:

1. aggregati parziali
2. insieme finito di misure di supporto
 - Esempi: AVG, deviazione standard, ...

OLISTICI: calcolabili a partire da:

1. dati originari
 - Esempio: rango di una matrice

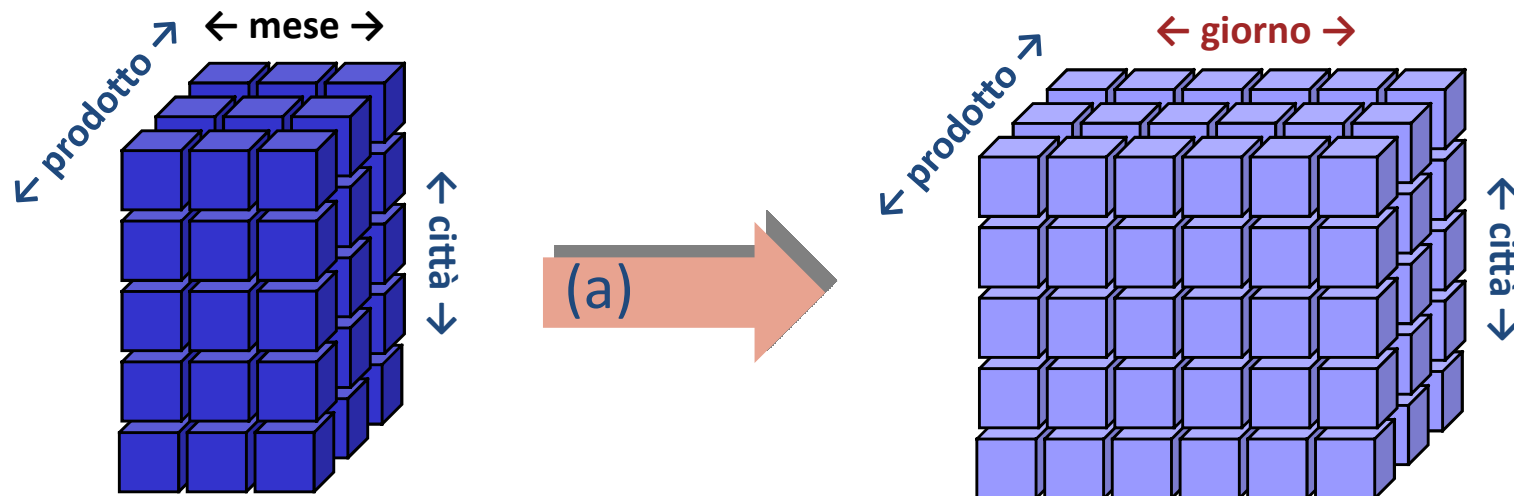
OLAP: Operatore Drill-Down

Disaggregazione lungo una dimensione

→ comporta la disponibilità dei dati disaggregati

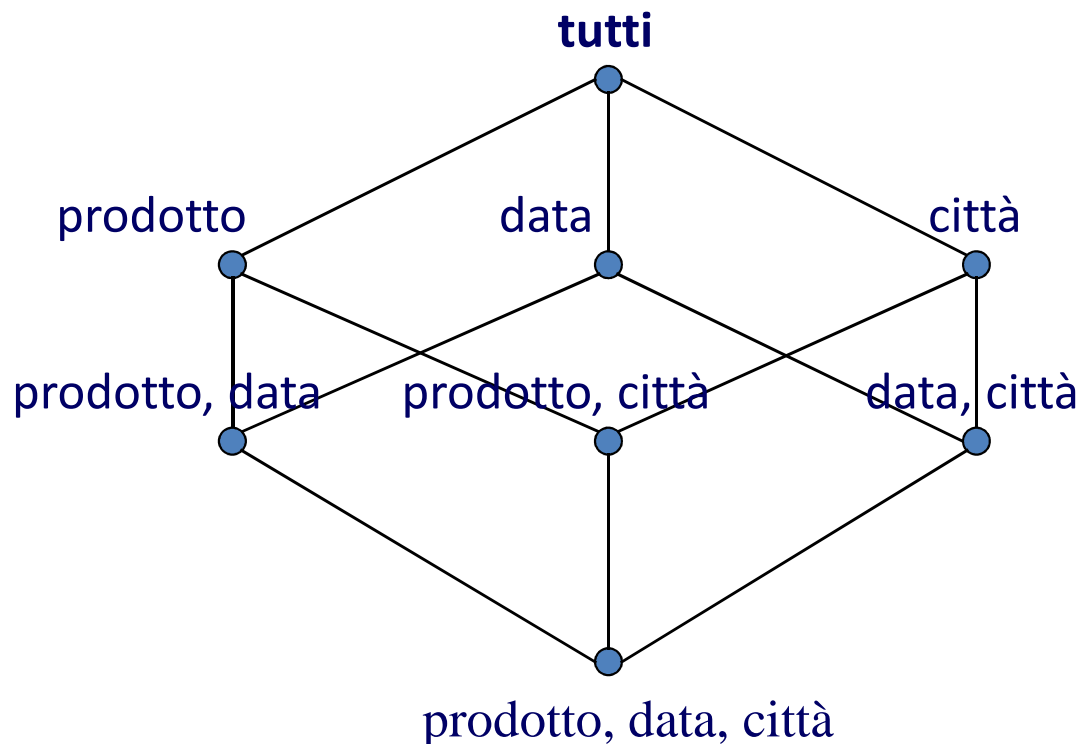
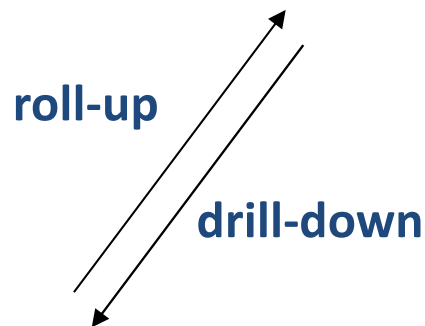
Esempi:

- passare dalle vendite di ciascun prodotto per città e per mese alle vendite giornaliere
- passare dalle vendite mensili per città (totale su tutti i prodotti) alle vendite giornaliere per città e per prodotto



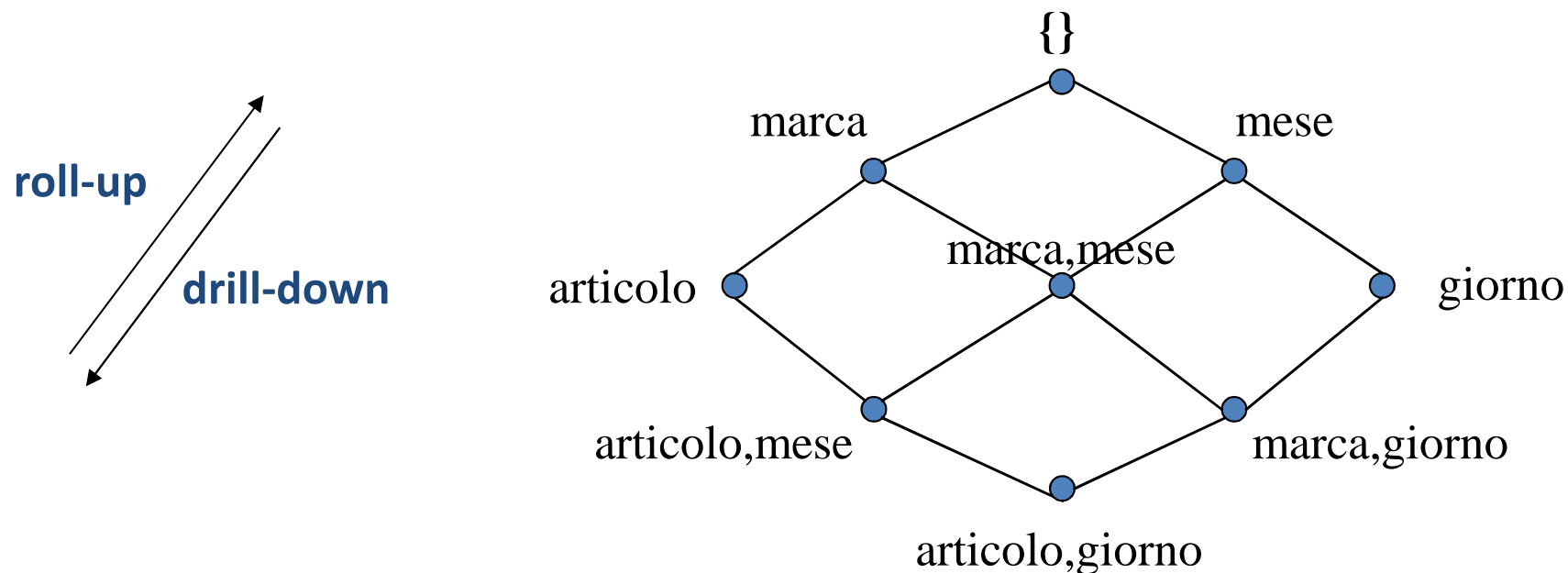
OLAP: Cuboidi (dimensioni)

- dato un cubo di dati, è possibile effettuare diverse operazioni di roll-up che conducono alla scomparsa di una dimensione
- dato un cubo con k dimensioni, esistono 2^k possibili cuboidi



OLAP: Cuboidi (livelli aggregazione)

- dato un cubo di dati, è possibile effettuare diverse operazioni di roll-up che aumentano il livello di aggregazione (con o senza la scomparsa di una dimensione)
- dato un cubo con k dimensioni di n_1, n_2, \dots, n_k livelli di aggregazione, esistono $n_1 \times n_2 \times \dots \times n_k$ possibili cuboidi



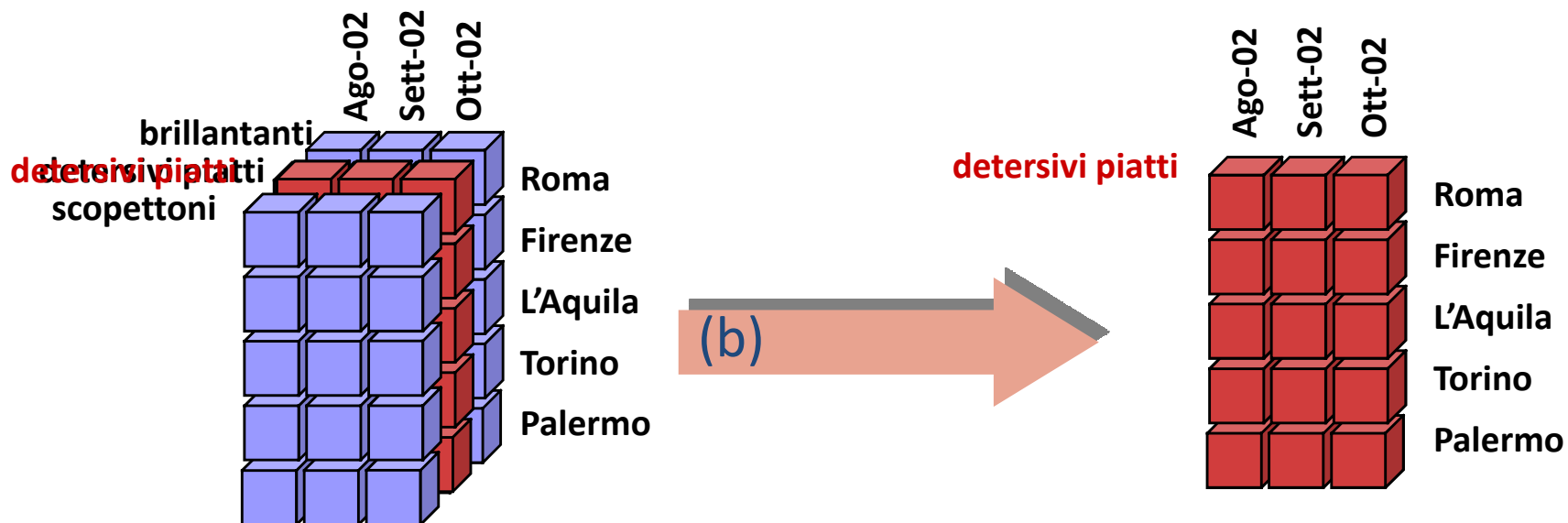
OLAP: Operatori Slice-and-Dice

Selezione (o, al limite, slicing)

Esempi:

passare dalle vendite mensili dei prodotti per città

- alle vendite mensili dei prodotti per le sole città del centro
- alle vendite mensili per città di detersivi per piatti



OLAP: Operatori di Pivoting

Rotazione delle dimensioni su una tabella bidimensionale

Esempio

passare da: vendite di (prodotti per mese) **per** (città)

a: vendite di (prodotti per città) **per** (mese)

		RM	FI	AQ	TO	PA
brillantanti	ago-02	34	23	12	56	65
	set-02	56	45	23	44	67
	ott-02	76	34	34	55	45
detersivi piatti	ago-02	57	46	35	79	88
	set-02	79	68	46	67	90
	ott-02	99	57	57	78	68
scopettoni	ago-02	46	35	24	68	77
	set-02	68	57	35	56	79
	ott-02	88	46	46	67	57

due
operazioni

		ago-02	set-02	ott-02
brillantanti	RM	34	56	76
	FI	23	45	34
	AQ	12	23	34
	TO	56	44	55
	PA	65	67	45
detersivi piatti	RM	57	79	99
	FI	46	68	57
	AQ	35	46	57
	TO	79	67	78
	PA	88	90	68
scopettoni	RM	46	68	88
	FI	35	57	46
	AQ	24	35	46
	TO	68	56	67
	PA	77	79	57

OLAP: Operatore Drill-Through

- accesso ai dati ad un livello di aggregazione più basso di quello offerto dal Data Warehouse
- accesso ai dati riconciliati (architettura a tre livelli) o al DB operativo
- costituisce un caso estremo di Drill-Down

OLAP: Operatori Drill-Across

- Correlare i dati provenienti da due o più cubi in base ai valori dimensionali, calcolando un'espressione in un nuovo cubo

Esempio:

A partire da

- vendite per prodotto, data, città
- dipendenti per città

calcolare:

- vendite medie per prodotto per dipendente