

## Analisi dei dati per il supporto alle decisioni

Elena Baralis  
Politecnico di Torino

## Supporto alle decisioni aziendali

- La maggior parte delle aziende dispone di enormi basi di dati contenenti dati di tipo operativo
  - queste basi di dati costituiscono una potenziale miniera di informazioni utili
- Sistemi per il supporto alle decisioni permettono di
  - analizzare lo stato dell'azienda
  - prendere decisioni *rapide e migliori*

## Supporto alle decisioni aziendali

- Analisi e previsione dell'evoluzione della domanda
- Individuazione di aree critiche
- Chiarezza dei conti e trasparenza finanziaria
  - reporting, pratiche antifrode e anticiclaggio
- Definizione e realizzazione di strategie vincenti
  - contenimento di costi e aumento di profitti

## Business Intelligence

- Intelligence: da *intus legere*
- Disciplina di supporto alla decisione strategica aziendale
- Obiettivo: trasformazione dei dati aziendali in informazioni fruibili
  - a livelli diversi di dettaglio
  - per applicazioni di analisi
- Tipologia di utenza eterogenea
- Necessaria un'adeguata infrastruttura hardware e software di supporto

## Ambiti applicativi

- Industrie manifatturiere: gestione ordini e spedizioni, supporto clienti
- Distribuzione: profilo utenti, gestione magazzino
- Servizi finanziari: analisi acquisti (carta di credito)
- Assicurazioni: analisi richieste indennizzo, riconoscimento frodi
- Telecomunicazioni: analisi delle chiamate, riconoscimento frodi
- Servizi pubblici: analisi dell'utilizzo
- Sanità: analisi dei risultati

## Data warehouse

- Base di dati per il supporto alle decisioni, che è mantenuta *separatamente* dalle basi di dati operative dell'azienda
  - Dati
    - orientati ai soggetti di interesse
    - integrati e consistenti
    - dipendenti dal tempo, non volatiliutilizzati per il supporto alle decisioni aziendali
- W. H. Inmon, Building the data warehouse, 1992*

## Perché dati separati?

- Prestazioni
  - ricerche complesse riducono le prestazioni delle transazioni operative
  - metodi di accesso diversi a livello fisico
- Gestione dei dati
  - informazioni mancanti (storico)
  - consolidamento dei dati
  - qualità dei dati (problema di inconsistenze)

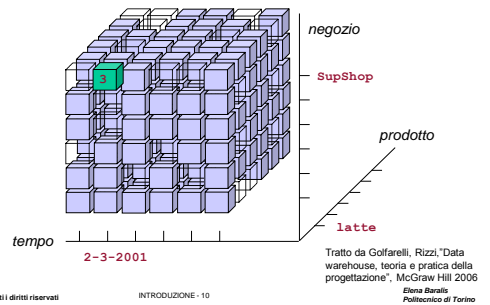
## Struttura e analisi dei dati

Elena Baralis  
Politecnico di Torino

## Rappresentazione multidimensionale

- Dati rappresentati come un (iper)cubo con tre o più dimensioni
- Misure su cui si esegue l'analisi: elementi individuati all'intersezione delle dimensioni
- Data warehouse per l'analisi delle vendite di una catena di supermercati
  - assi dimensionali: prodotto, negozio, tempo
  - misure: quantità venduta, importo della vendita, ...

## Rappresentazione multidimensionale

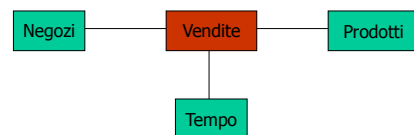


## Rappresentazione relazionale: modello a stella

- Misure numeriche memorizzate nella *tabella dei fatti*
  - gli attributi contengono valori numerici
- Le *dimensioni* descrivono il contesto di ogni misura nella tabella dei fatti
  - contengono molti attributi descrittivi

## Esempio

Data warehouse per l'analisi delle vendite di una catena di supermercati



## Dimensione del data warehouse

- dimensione tempo: 2 anni x 365 giorni
- dimensione negozio: 300 negozi
- dimensione prodotto: 30.000 prodotti, di cui 3.000 venduti ogni giorno in ogni negozio
- numero di righe della tabella dei fatti:
 
$$730 \times 300 \times 3000 = 657 \text{ milioni}$$

⇒ spazio occupato dalla tabella dei fatti ≈ 21GB

## Strumenti di analisi dei dati

- Analisi OLAP: calcolo di funzioni aggregate complesse
  - necessità di fornire supporto a diversi tipi di funzione aggregata (esempi: media mobile, top ten)
- Analisi dei dati mediante tecniche di data mining
  - varie tipologie di analisi
  - pesante componente algoritmica

## Strumenti di analisi dei dati

- Presentazione
  - attività distinta dalla ricerca: i dati ottenuti da una ricerca possono essere rappresentati mediante diversi tipi di strumenti di rappresentazione
- Ricerca di motivazioni
  - esplorazione dei dati mediante approfondimenti (esempio: drill down)

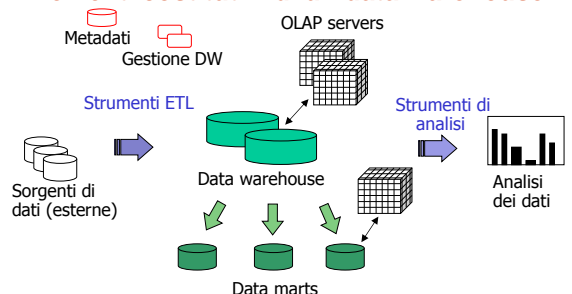
## Architetture per data warehouse

Elena Baralis  
Politecnico di Torino

## Architetture per data warehouse

- Separazione tra elaborazione transazionale e analisi dei dati
  - evitare le architetture a un livello
- Architetture a due o più livelli
  - separano in misura diversa i dati in ingresso nel data warehouse dai dati oggetto dell'analisi
  - maggiormente scalabili

## Elementi costitutivi di un data warehouse



## Data warehouse e data mart

**Warehouse aziendale:** contiene informazioni sul funzionamento di "tutta" l'azienda

- processo di modellazione funzionale esteso
- progettazione e realizzazione richiedono molto tempo

**Data mart:** sottoinsieme dipartimentale focalizzato su un settore prefissato

- due possibilità
  - alimentato dal data warehouse primario
  - alimentato direttamente dalle sorgenti
- realizzazione più rapida
- richiede progettazione attenta, in modo da evitare problemi di integrazione in seguito

## Server per data warehouse

- Server ROLAP (Relational OLAP)
  - DBMS relazionale esteso
    - rappresentazione compatta di dati sparsi
  - estensioni SQL per aggregati
  - metodi di accesso speciali che realizzano le operazioni di accesso in modo efficiente
- Server MOLAP (Multidimensional OLAP)
  - dati rappresentati in forma matriciale (multidimensionale) proprietaria
    - dati sparsi richiedono compressione
  - primitive OLAP speciali
- Server HOLAP (Hybrid OLAP)

## Strumenti ETL

- Processo di preparazione dei dati da introdurre nel data warehouse
  - Extraction
  - Transformation
  - Loading
- Processo eseguito durante
  - il primo popolamento del DW
  - l'aggiornamento periodico dei dati

## Processo ETL

**Estrazione:** acquisizione dei dati dalle sorgenti

**Pulitura:** operazioni volte al miglioramento della qualità dei dati (correttezza e consistenza)

**Trasformazione:** conversione dei dati dal formato operativo a quello del data warehouse (integrazione)

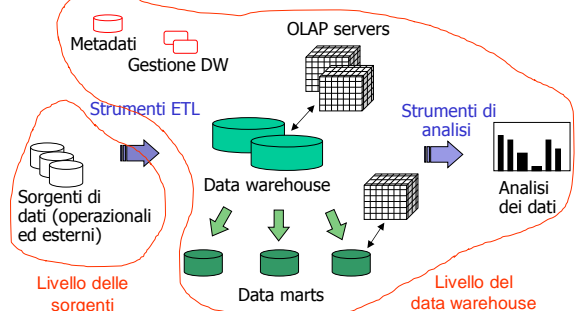
**Caricamento:** propagazione degli aggiornamenti al data warehouse

## Metadati

Metadati = dati sui dati

- Diversi tipi di metadati
  - per trasformazione e caricamento: descrivono i dati sorgenti e le trasformazioni necessarie
    - utile usare una notazione comune per dati sorgente e dati risultanti dalle trasformazioni
    - CWM (Common Warehouse Metadata Initiative): standard proposto da OMG per l'interscambio di dati tra strumenti DW e repository di metadati in ambienti eterogenei e distribuiti
  - per la gestione dei dati: descrivono la struttura dei dati presenti nel data warehouse
    - anche per dati derivati, quali le viste materializzate
  - per la gestione delle query: dati sulla struttura delle query e monitoraggio della loro esecuzione
    - codice SQL della query
    - piano di esecuzione
    - uso di memoria e CPU

## Architettura a due livelli



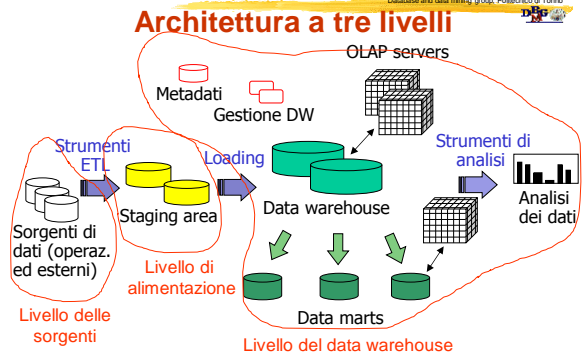
## Caratteristiche delle architetture a 2 livelli

- Disaccoppiamento dalle sorgenti
  - possibilità di gestire dati esterni al sistema OLTP
  - modellazione dei dati adatta all'analisi OLAP
  - progettazione fisica del data warehouse mirata al carico analitico
- Facilità di gestione delle differenti granularità temporali dei dati operazionali e analitici
- Separazione del carico transazionale da quello analitico
- Necessità di svolgere "al volo" la preparazione dei dati (ETL)

Copyright - Tutti i diritti riservati

INTRODUZIONE - 25

Elena Baralis  
Politecnico di Torino



Copyright - Tutti i diritti riservati

INTRODUZIONE - 26

Elena Baralis  
Politecnico di Torino

## Caratteristiche delle architetture a 3 livelli

- *Staging area*: area di transito che permette di separare l'elaborazione ET dal caricamento nel data warehouse
  - permette operazioni complesse di trasformazione e pulizia dei dati
  - offre un modello integrato dei dati aziendali, ancora vicino alla rappresentazione OLTP
  - talvolta denominata Operational Data Store (ODS)
- Introduce ulteriore ridondanza
  - aumenta lo spazio necessario per i dati

Copyright - Tutti i diritti riservati

INTRODUZIONE - 27

Elena Baralis  
Politecnico di Torino

## Data warehouse Progettazione

Elena Baralis  
Politecnico di Torino

Copyright - Tutti i diritti riservati

INTRODUZIONE - 28

Elena Baralis  
Politecnico di Torino

## Fattori di rischio

- Aspettative elevate degli utenti
  - il data warehouse come soluzione dei problemi aziendali
- Qualità dei dati e dei processi OLTP di partenza
  - dati incompleti o inaffidabili
  - processi aziendali non integrati e ottimizzati
- Gestione "politica" del progetto
  - collaborazione con i "detentori" delle informazioni
  - accettazione del sistema da parte degli utenti finali

Copyright - Tutti i diritti riservati

INTRODUZIONE - 29

Elena Baralis  
Politecnico di Torino

## Progettazione di data warehouse

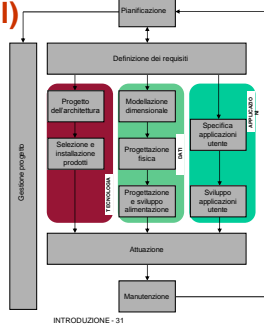
- Approccio top-down
  - realizzazione di un data warehouse che fornisca una visione globale e completa dei dati aziendali
  - costo significativo e tempo di realizzazione lungo
  - analisi e progettazione complesse
- Approccio bottom-up
  - realizzazione incrementale del data warehouse, aggiungendo data mart definiti su settori aziendali specifici
  - costo e tempo di consegna contenuti
  - focalizzato separatamente su settori aziendali specifici

Copyright - Tutti i diritti riservati

INTRODUZIONE - 30

Elena Baralis  
Politecnico di Torino

## Business Dimensional Lifecycle (Kimball)



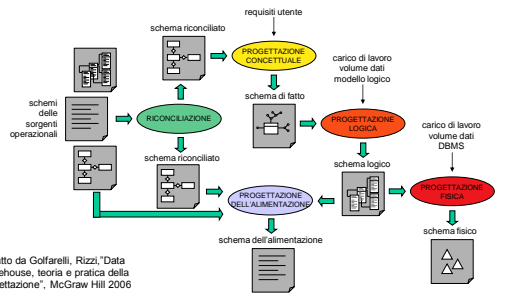
Tratto da Goffarelli, Rizzi, "Data warehouse, teoria e pratica della progettazione", McGraw Hill 2006

Elena Baralis  
Politecnico di Torino

Copyright - Tutti i diritti riservati

INTRODUZIONE - 31

## Progettazione di data mart



Tratto da Goffarelli, Rizzi, "Data warehouse, teoria e pratica della progettazione", McGraw Hill 2006

Copyright - Tutti i diritti riservati

INTRODUZIONE - 32

Elena Baralis  
Politecnico di Torino

## Analisi dei requisiti

Elena Baralis  
Politecnico di Torino

- Raccoglie
  - le esigenze di analisi dei dati che dovranno essere soddisfatte dal data mart
  - i vincoli realizzativi dovuti ai sistemi informativi esistenti
- Fonti
  - business users
  - amministratori del sistema informativo
- Il data mart prescelto è
  - strategico per l'azienda
  - alimentato da (poche) sorgenti affidabili

Copyright - Tutti i diritti riservati

INTRODUZIONE - 33

Elena Baralis  
Politecnico di Torino

Copyright - Tutti i diritti riservati

INTRODUZIONE - 34

Elena Baralis  
Politecnico di Torino

## Requisiti applicativi

- Descrizione degli eventi di interesse (fatti)
  - ogni fatto rappresenta una categoria di eventi di interesse per l'azienda
    - esempi: (per il CRM) reclami, servizi
  - caratterizzati da dimensioni descrittive (granularità), intervallo di storicizzazione, misure di interesse
  - informazioni raccolte in un glossario
- Descrizione del carico di lavoro
  - esame della reportistica aziendale
  - interrogazioni espresse in linguaggio naturale
    - esempio: numero di reclami per ciascun prodotto nell'ultimo mese

Copyright - Tutti i diritti riservati

INTRODUZIONE - 35

Elena Baralis  
Politecnico di Torino

## Requisiti strutturali

- Periodicità dell'alimentazione
- Spazio disponibile
  - per i dati
  - per le strutture accessorie (indici, viste materializzate)
- Tipo di architettura del sistema
  - numero di livelli
  - data mart dipendenti o indipendenti
- Pianificazione del deployment
  - avviamento
  - formazione

Copyright - Tutti i diritti riservati

INTRODUZIONE - 36

Elena Baralis  
Politecnico di Torino

## Progettazione concettuale

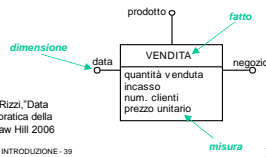
Elena Baralis  
Politecnico di Torino

## Progettazione concettuale

- Non esiste un formalismo di modellazione comunemente accettato
  - il modello ER non è adatto
- Dimensional Fact Model (Goffarelli, Rizzi)
  - per uno specifico fatto, definisce schemi di fatto che modellano
    - dimensioni
    - gerarchie
    - misure
  - modello grafico a supporto della progettazione concettuale
  - offre una documentazione di progetto utile sia per la revisione dei requisiti con gli utenti, sia a posteriori

## Dimensional Fact Model

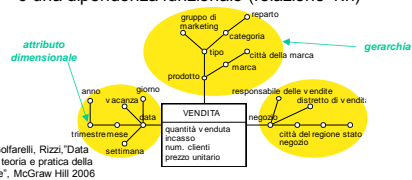
- Fatto
  - modella un insieme di eventi di interesse (vendite, spedizioni, reclami)
  - evolve nel tempo
- Dimensione
  - descrive le coordinate di analisi di un fatto (ogni vendita è descritta dalla data di effettuazione, dal negozio e dal prodotto venduto)
  - è caratterizzata da numerosi attributi, tipicamente di tipo categorico
- Misura
  - descrive una proprietà numerica di un fatto, spesso oggetto di operazioni di aggregazione (ad ogni vendita è associato un incasso)



Tratto da Goffarelli, Rizzi, "Data warehouse, teoria e pratica della progettazione", McGraw Hill 2006

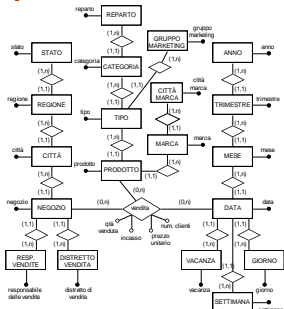
## Dimensional Fact Model

- Gerarchia
  - rappresenta una relazione di generalizzazione tra un sottoinsieme di attributi di una dimensione (gerarchia geografica per la dimensione negozio)
  - è una dipendenza funzionale (relazione 1:n)



Tratto da Goffarelli, Rizzi, "Data warehouse, teoria e pratica della progettazione", McGraw Hill 2006

## Corrispondenza con l'ER



Tratto da Goffarelli, Rizzi, "Data warehouse, teoria e pratica della progettazione", McGraw Hill 2006

## Aggregazioni

Elena Baralis  
Politecnico di Torino

## Aggregazione

- Processo di calcolo del valore di misure a granularità meno fine di quella presente nello schema di fatto originale
  - la riduzione del livello di dettaglio è ottenuta risalendo lungo una gerarchia
  - operatori di aggregazione standard: SUM, MIN, MAX, AVG, COUNT
- Caratteristiche delle misure
  - additive
  - non additive: non aggregabili lungo una gerarchia mediante l'operatore di somma
  - non aggregabili

## Classificazione delle misure

- Misure di flusso
  - possono essere valutate cumulativamente alla fine di un periodo di tempo
  - sono aggregabili mediante tutti gli operatori standard
  - esempi: quantità di prodotti venduti, importo incassato
- Misure di livello
  - sono valutate in specifici istanti di tempo (snapshot)
  - non sono additive lungo la dimensione tempo
  - esempi: livello di inventario, saldo del conto corrente
- Misure unitarie
  - sono valutate in specifici istanti di tempo ed espresse in termini relativi
  - non sono additive lungo nessuna dimensione
  - esempio: prezzo unitario di un prodotto

## Operatori di aggregazione

categoria	tipo	prodotto	1999				2000			
			I '99	II '99	III '99	IV '99	I 00	II 00	III 00	IV 00
pulizia casa	detersivo	Briio	100	90	95	90	80	70	90	85
		Stano	20	30	20	19	25	30	15	20
		Lucida	60	50	60	45	30	40	50	40
sapone	Marmorata	Marmorata	15	20	25	30	15	15	20	10
		Scat	30	35	20	25	30	30	20	15
		Latte Sharp	80	80	85	75	60	80	85	60
alimenti	latticini	Latte Sharp	60	80	85	60	80	70	75	65
		Yogurt Sharp	20	30	40	35	30	35	35	20
		Beviri	20	10	25	30	35	30	20	10
		Colosima	50	60	45	40	50	60	45	40

categoria	tipo	1999				2000			
		I '99	II '99	III '99	IV '99	I 00	II 00	III 00	IV 00
pulizia casa	detersivo	225	225	220	200	190	185	215	170
		250	230	280	240	245	275	260	195
alimenti	latticini	200	200	200	200	190	185	215	170
		200	200	200	200	190	185	215	170

Tratto da Gofarelli, Rizzi, "Data warehouse, teoria e pratica della progettazione", McGraw Hill 2006

## Operazioni di analisi dei dati

- Calcolo di funzioni aggregate lungo una o più dimensioni
  - necessità di fornire supporto a diversi tipi di funzione aggregata (esempi: media mobile, top ten)
- Operazioni di confronto, essenziali per confrontare l'andamento degli affari (esempio: confronto dei dati delle vendite in mesi diversi)
  - è difficile eseguire confronti utilizzando solo il linguaggio SQL
- Analisi dei dati mediante tecniche di data mining

## Strumenti di interfaccia

- L'utente può interrogare il data warehouse mediante strumenti di vario tipo:
- ambiente controllato di query
  - strumenti specifici di query e generazione rapporti
    - Es: [Google Data Analytics](#)
  - strumenti di data mining

## Ambiente controllato di query

- Sono definite
  - ricerche complesse con struttura prefissata (normalmente parametrica)
  - procedure specifiche di analisi
  - rapporti con struttura prefissata
- È possibile introdurre elementi specifici del settore economico considerato
- È necessario lo sviluppo di codice ad hoc
  - si utilizzano stored procedures, applicazioni contenute in packages, join e aggregazioni predefinite
  - sono disponibili strumenti flessibili per la gestione della reportistica, che permettono di definire layout, periodicità di pubblicazione, liste di distribuzione



## Ambiente di query ad hoc

- È possibile definire interrogazioni OLAP di tipo arbitrario, progettate al momento dall'utente
  - formulazione delle interrogazioni mediante tecniche point and click, che generano automaticamente istruzioni SQL
  - si possono definire interrogazioni (tipicamente) complesse
  - interfaccia basata sul paradigma dello spreadsheet
- Una sessione di lavoro OLAP permette raffinamenti successivi della stessa interrogazione
- Utile quando i rapporti predefiniti non sono adeguati

## OLAP

Elena Baralis  
Politecnico di Torino

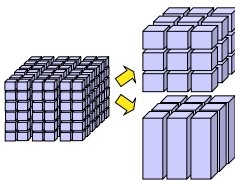
## Analisi OLAP

- Operazioni di ricerca disponibili
  - roll up, drill down
  - slice and dice
  - pivot di tabelle
  - ordinamento
- Le operazioni possono essere
  - combinare tra loro nella stessa query
  - eseguite in una sequenza di raffinamenti successivi della stessa query che forma la sessione di lavoro OLAP

## Roll up

- Riduzione di dettaglio dei dati mediante
  - la riduzione del livello di dettaglio di una delle dimensioni presenti, con l'aumento di livello in una gerarchia
    - esempio  
group by negozio, mese → group by città, mese
  - l'eliminazione di una delle dimensioni presenti
    - esempio  
group by prodotto, città → group by prodotto

## Roll up



Tratto da Goffarelli, Rizzi, "Data warehouse, teoria e pratica della progettazione", McGraw Hill 2006

## Roll up

Mesino	Cittadina										
Contabile	North-East	Mid-Atlantic	South-East	Central	South	North-West	South-West	England	France	Germany	Canada
Jan 97	\$ 450	\$ 750	\$ 20	\$ 200	\$ 200	\$ 120	\$ 400	\$ 100	\$ 200	\$ 200	\$ 200
Feb 97	\$ 500	\$ 750	\$ 200	\$ 200	\$ 200	\$ 200	\$ 200	\$ 200	\$ 200	\$ 200	\$ 200
Mar 97	\$ 500	\$ 750	\$ 200	\$ 200	\$ 200	\$ 200	\$ 200	\$ 200	\$ 200	\$ 200	\$ 200
Apr 97	\$ 500	\$ 750	\$ 200	\$ 200	\$ 200	\$ 200	\$ 200	\$ 200	\$ 200	\$ 200	\$ 200
May 97	\$ 500	\$ 750	\$ 200	\$ 200	\$ 200	\$ 200	\$ 200	\$ 200	\$ 200	\$ 200	\$ 200
Jun 97	\$ 500	\$ 750	\$ 200	\$ 200	\$ 200	\$ 200	\$ 200	\$ 200	\$ 200	\$ 200	\$ 200
Jul 97	\$ 500	\$ 750	\$ 200	\$ 200	\$ 200	\$ 200	\$ 200	\$ 200	\$ 200	\$ 200	\$ 200
Aug 97	\$ 500	\$ 750	\$ 200	\$ 200	\$ 200	\$ 200	\$ 200	\$ 200	\$ 200	\$ 200	\$ 200
Sep 97	\$ 500	\$ 750	\$ 200	\$ 200	\$ 200	\$ 200	\$ 200	\$ 200	\$ 200	\$ 200	\$ 200
Oct 97	\$ 500	\$ 750	\$ 200	\$ 200	\$ 200	\$ 200	\$ 200	\$ 200	\$ 200	\$ 200	\$ 200
Nov 97	\$ 500	\$ 750	\$ 200	\$ 200	\$ 200	\$ 200	\$ 200	\$ 200	\$ 200	\$ 200	\$ 200
Dec 97	\$ 500	\$ 750	\$ 200	\$ 200	\$ 200	\$ 200	\$ 200	\$ 200	\$ 200	\$ 200	\$ 200
Jan 98	\$ 500	\$ 750	\$ 200	\$ 200	\$ 200	\$ 200	\$ 200	\$ 200	\$ 200	\$ 200	\$ 200
Feb 98	\$ 500	\$ 750	\$ 200	\$ 200	\$ 200	\$ 200	\$ 200	\$ 200	\$ 200	\$ 200	\$ 200
Mar 98	\$ 500	\$ 750	\$ 200	\$ 200	\$ 200	\$ 200	\$ 200	\$ 200	\$ 200	\$ 200	\$ 200
Apr 98	\$ 500	\$ 750	\$ 200	\$ 200	\$ 200	\$ 200	\$ 200	\$ 200	\$ 200	\$ 200	\$ 200
May 98	\$ 500	\$ 750	\$ 200	\$ 200	\$ 200	\$ 200	\$ 200	\$ 200	\$ 200	\$ 200	\$ 200
Jun 98	\$ 500	\$ 750	\$ 200	\$ 200	\$ 200	\$ 200	\$ 200	\$ 200	\$ 200	\$ 200	\$ 200
Jul 98	\$ 500	\$ 750	\$ 200	\$ 200	\$ 200	\$ 200	\$ 200	\$ 200	\$ 200	\$ 200	\$ 200
Aug 98	\$ 500	\$ 750	\$ 200	\$ 200	\$ 200	\$ 200	\$ 200	\$ 200	\$ 200	\$ 200	\$ 200
Sep 98	\$ 500	\$ 750	\$ 200	\$ 200	\$ 200	\$ 200	\$ 200	\$ 200	\$ 200	\$ 200	\$ 200
Oct 98	\$ 500	\$ 750	\$ 200	\$ 200	\$ 200	\$ 200	\$ 200	\$ 200	\$ 200	\$ 200	\$ 200
Nov 98	\$ 500	\$ 750	\$ 200	\$ 200	\$ 200	\$ 200	\$ 200	\$ 200	\$ 200	\$ 200	\$ 200
Dec 98	\$ 500	\$ 750	\$ 200	\$ 200	\$ 200	\$ 200	\$ 200	\$ 200	\$ 200	\$ 200	\$ 200

Tratto da Goffarelli, Rizzi, "Data warehouse, teoria e pratica della progettazione", McGraw Hill 2006

## Roll up

Region	North-East	South-East	Central	South	North-west	South-west	England	France	Germany
Category	1997	\$ 1.08	\$ 1.774	\$ 884	\$ 1.01	\$ 2.344	\$ 1.034	\$ 1.664	\$ 1.00
Electronics	1998	\$ 1.084	\$ 1.826	\$ 920	\$ 1.001	\$ 2.408	\$ 1.078	\$ 1.682	\$ 1.001
Food	1997	\$ 790	\$ 682	\$ 720	\$ 280	\$ 460	\$ 800	\$ 136	\$ 11
Books	1998	\$ 1.000	\$ 1.000	\$ 1.000	\$ 1.000	\$ 1.000	\$ 1.000	\$ 1.000	\$ 1.000



Category	Region	1997	1998
Electronics	1997	\$ 1.08	\$ 1.084
Food	1997	\$ 790	\$ 1.000
Books	1997	\$ 1.000	\$ 1.000
North & South	1997	\$ 1.000	\$ 1.000
West	1997	\$ 1.000	\$ 1.000
East	1997	\$ 1.000	\$ 1.000
France	1997	\$ 1.000	\$ 1.000
Germany	1997	\$ 1.000	\$ 1.000

Tratto da Goffarelli, Rizzi, "Data warehouse, teoria e pratica della progettazione", McGraw Hill 2006

Copyright - Tutti i diritti riservati

INTRODUZIONE - 55

Elena Baralis Politecnico di Torino

## Drill down

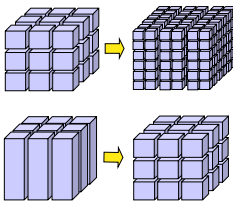
- Aumento di dettaglio dei dati mediante
  - l'aumento del livello di dettaglio di una delle dimensioni presenti, con la riduzione di livello in una gerarchia
    - esempio: da raggruppamento per città e mese a raggruppamento per negozio e mese
  - l'aggiunta di una nuova dimensione
    - esempio: da raggruppamento per città, raggruppamento per città e prodotto
- Spesso il drill down opera su un sottoinsieme dei dati di partenza

Copyright - Tutti i diritti riservati

INTRODUZIONE - 56

Elena Baralis Politecnico di Torino

## Drill down



Tratto da Goffarelli, Rizzi, "Data warehouse, teoria e pratica della progettazione", McGraw Hill 2006

Copyright - Tutti i diritti riservati

INTRODUZIONE - 57

Elena Baralis Politecnico di Torino

## Drill down

Region	North-East	South-East	Central	South	North-west	South-west	England	France	Germany	Canada
Quarter	1997	\$ 1.08	\$ 1.774	\$ 884	\$ 1.01	\$ 2.344	\$ 1.034	\$ 1.664	\$ 1.00	\$ 1.00
Q1 1997	\$ 293	\$ 445	\$ 221	\$ 266	\$ 608	\$ 266	\$ 266	\$ 419	\$ 266	\$ 266
Q2 1997	\$ 293	\$ 445	\$ 221	\$ 266	\$ 608	\$ 266	\$ 266	\$ 419	\$ 266	\$ 266
Q3 1997	\$ 293	\$ 445	\$ 221	\$ 266	\$ 608	\$ 266	\$ 266	\$ 419	\$ 266	\$ 266
Q4 1997	\$ 293	\$ 445	\$ 221	\$ 266	\$ 608	\$ 266	\$ 266	\$ 419	\$ 266	\$ 266



Customer	City	Product	Quantity	Price	Total
C1	Torino	Food	1	100	100
C2	Torino	Books	1	100	100
C3	Torino	Electronics	1	100	100
C4	Torino	Food	1	100	100
C5	Torino	Books	1	100	100
C6	Torino	Electronics	1	100	100
C7	Torino	Food	1	100	100
C8	Torino	Books	1	100	100
C9	Torino	Electronics	1	100	100
C10	Torino	Food	1	100	100

Tratto da Goffarelli, Rizzi, "Data warehouse, teoria e pratica della progettazione", McGraw Hill 2006

Copyright - Tutti i diritti riservati

INTRODUZIONE - 58

Elena Baralis Politecnico di Torino

## Drill down

Customer	Year	1997	1998
Electronics	1997	\$ 1.08	\$ 1.084
Food	1997	\$ 790	\$ 1.000
Books	1997	\$ 1.000	\$ 1.000
North & South	1997	\$ 1.000	\$ 1.000
West	1997	\$ 1.000	\$ 1.000
East	1997	\$ 1.000	\$ 1.000
France	1997	\$ 1.000	\$ 1.000
Germany	1997	\$ 1.000	\$ 1.000



Region	North-East	South-East	Central	South	North-west	South-west	England	France	Germany
Category	1997	\$ 1.08	\$ 1.774	\$ 884	\$ 1.01	\$ 2.344	\$ 1.034	\$ 1.664	\$ 1.00
Electronics	1998	\$ 1.084	\$ 1.826	\$ 920	\$ 1.001	\$ 2.408	\$ 1.078	\$ 1.682	\$ 1.001
Food	1997	\$ 790	\$ 682	\$ 720	\$ 280	\$ 460	\$ 800	\$ 136	\$ 11
Books	1998	\$ 1.000	\$ 1.000	\$ 1.000	\$ 1.000	\$ 1.000	\$ 1.000	\$ 1.000	\$ 1.000

Tratto da Goffarelli, Rizzi, "Data warehouse, teoria e pratica della progettazione", McGraw Hill 2006

Copyright - Tutti i diritti riservati

INTRODUZIONE - 59

Elena Baralis Politecnico di Torino

## Slice and dice

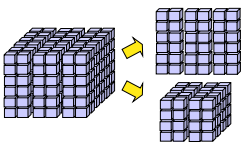
- Riduzione del volume dei dati da analizzare
  - selezione di un sottoinsieme mediante predicati
    - slice: predicato di uguaglianza che seleziona una "fetta"
      - esempio: Anno=2005
    - dice: combinazione di predicati che seleziona un "cubetto"
      - esempio: Categoria='Alimentari' and Città='Torino'

Copyright - Tutti i diritti riservati

INTRODUZIONE - 60

Elena Baralis Politecnico di Torino

## Slice and dice



Tratto da Goffarelli, Rizzi, "Data warehouse, teoria e pratica della progettazione", McGraw Hill 2006  
Copyright - Tutti i diritti riservati

INTRODUZIONE - 61

Elena Baralis  
Politecnico di Torino

## Slice and dice

Category	Year	Metric (Dollar Sales)							
		Midwest	Southwest	South	Northwest	Southwest	England	France	Germany
Electronics	1997	\$ 1,128	\$ 1,174	\$ 1,044	\$ 1,128	\$ 1,044	\$ 1,144	\$ 1,144	\$ 1,144
Food	1999	\$ 1,184	\$ 1,200	\$ 1,184	\$ 1,184	\$ 1,184	\$ 1,184	\$ 1,184	\$ 1,184
Books	1997	\$ 1,184	\$ 1,184	\$ 1,184	\$ 1,184	\$ 1,184	\$ 1,184	\$ 1,184	\$ 1,184
Life	1999	\$ 1,184	\$ 1,184	\$ 1,184	\$ 1,184	\$ 1,184	\$ 1,184	\$ 1,184	\$ 1,184
Health & Beauty	1997	\$ 1,184	\$ 1,184	\$ 1,184	\$ 1,184	\$ 1,184	\$ 1,184	\$ 1,184	\$ 1,184
Household	1997	\$ 1,184	\$ 1,184	\$ 1,184	\$ 1,184	\$ 1,184	\$ 1,184	\$ 1,184	\$ 1,184
Prof & Expens	1997	\$ 1,184	\$ 1,184	\$ 1,184	\$ 1,184	\$ 1,184	\$ 1,184	\$ 1,184	\$ 1,184
Travel	1997	\$ 1,184	\$ 1,184	\$ 1,184	\$ 1,184	\$ 1,184	\$ 1,184	\$ 1,184	\$ 1,184
Travel	1999	\$ 1,184	\$ 1,184	\$ 1,184	\$ 1,184	\$ 1,184	\$ 1,184	\$ 1,184	\$ 1,184
Travel	1999	\$ 1,184	\$ 1,184	\$ 1,184	\$ 1,184	\$ 1,184	\$ 1,184	\$ 1,184	\$ 1,184

Tratto da Goffarelli, Rizzi, "Data warehouse, teoria e pratica della progettazione", McGraw Hill 2006  
Copyright - Tutti i diritti riservati

INTRODUZIONE - 62

Elena Baralis  
Politecnico di Torino

## Slice and dice

Category	Year	Metric (Dollar Sales)							
		Midwest	Southwest	South	Northwest	Southwest	England	France	Germany
Subcategory	Year	Midwest	Southwest	South	Northwest	Southwest	England	France	Germany
Subcategory	1997	\$ 1,184	\$ 1,184	\$ 1,184	\$ 1,184	\$ 1,184	\$ 1,184	\$ 1,184	\$ 1,184
Subcategory	1999	\$ 1,184	\$ 1,184	\$ 1,184	\$ 1,184	\$ 1,184	\$ 1,184	\$ 1,184	\$ 1,184
Subcategory	1997	\$ 1,184	\$ 1,184	\$ 1,184	\$ 1,184	\$ 1,184	\$ 1,184	\$ 1,184	\$ 1,184
Subcategory	1999	\$ 1,184	\$ 1,184	\$ 1,184	\$ 1,184	\$ 1,184	\$ 1,184	\$ 1,184	\$ 1,184
Subcategory	1997	\$ 1,184	\$ 1,184	\$ 1,184	\$ 1,184	\$ 1,184	\$ 1,184	\$ 1,184	\$ 1,184
Subcategory	1999	\$ 1,184	\$ 1,184	\$ 1,184	\$ 1,184	\$ 1,184	\$ 1,184	\$ 1,184	\$ 1,184
Subcategory	1997	\$ 1,184	\$ 1,184	\$ 1,184	\$ 1,184	\$ 1,184	\$ 1,184	\$ 1,184	\$ 1,184
Subcategory	1999	\$ 1,184	\$ 1,184	\$ 1,184	\$ 1,184	\$ 1,184	\$ 1,184	\$ 1,184	\$ 1,184
Subcategory	1997	\$ 1,184	\$ 1,184	\$ 1,184	\$ 1,184	\$ 1,184	\$ 1,184	\$ 1,184	\$ 1,184
Subcategory	1999	\$ 1,184	\$ 1,184	\$ 1,184	\$ 1,184	\$ 1,184	\$ 1,184	\$ 1,184	\$ 1,184
Subcategory	1997	\$ 1,184	\$ 1,184	\$ 1,184	\$ 1,184	\$ 1,184	\$ 1,184	\$ 1,184	\$ 1,184
Subcategory	1999	\$ 1,184	\$ 1,184	\$ 1,184	\$ 1,184	\$ 1,184	\$ 1,184	\$ 1,184	\$ 1,184
Subcategory	1997	\$ 1,184	\$ 1,184	\$ 1,184	\$ 1,184	\$ 1,184	\$ 1,184	\$ 1,184	\$ 1,184
Subcategory	1999	\$ 1,184	\$ 1,184	\$ 1,184	\$ 1,184	\$ 1,184	\$ 1,184	\$ 1,184	\$ 1,184

Copyright - Tutti i diritti riservati

INTRODUZIONE - 63

Elena Baralis  
Politecnico di Torino

## Pivot

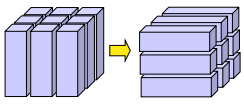
- Riorganizzazione dell'orientamento della struttura multidimensionale senza variare il livello di dettaglio
  - permette una visualizzazione più chiara delle stesse informazioni
  - la rappresentazione dei dati multidimensionali rimane sotto forma di "griglia"
    - due dimensioni sono gli assi principali della griglia
    - varia la posizione delle dimensioni nella griglia

Copyright - Tutti i diritti riservati

INTRODUZIONE - 64

Elena Baralis  
Politecnico di Torino

## Pivot



Tratto da Goffarelli, Rizzi, "Data warehouse, teoria e pratica della progettazione", McGraw Hill 2006  
Copyright - Tutti i diritti riservati

INTRODUZIONE - 65

Elena Baralis  
Politecnico di Torino

## Pivot

Category	Year	Metric (Dollar Sales)	
		Midwest	Southwest
Electronics	1997	\$ 10,616	\$ 10,616
Food	1999	\$ 10,616	\$ 10,616
Books	1997	\$ 10,616	\$ 10,616
Life	1999	\$ 10,616	\$ 10,616
Health & Beauty	1997	\$ 10,616	\$ 10,616
Household	1999	\$ 10,616	\$ 10,616
Prof & Expens	1997	\$ 10,616	\$ 10,616
Travel	1997	\$ 10,616	\$ 10,616
Travel	1999	\$ 10,616	\$ 10,616

Copyright - Tutti i diritti riservati

INTRODUZIONE - 66

Elena Baralis  
Politecnico di Torino

Database and data mining group, Politecnico di Torino

## Pivot

Category	Year	Market (Order Dates)									
		North-east	North-east	South-east	Central	South	North-west	South-west	England	France	Germany
Economics	1997	\$ 1,124	\$ 1,274	\$ 384	\$ 1,108	\$ 1,344	\$ 1,014	\$ 1,134	\$ 564	\$ 1,194	\$
Economics	1998	\$ 1,244	\$ 1,494	\$ 494	\$ 1,324	\$ 1,554	\$ 1,224	\$ 1,344	\$ 674	\$ 1,414	\$ 1,464
Food	1997	\$ 704	\$ 854	\$ 254	\$ 734	\$ 884	\$ 604	\$ 674	\$ 334	\$ 744	\$ 1,154
Food	1998	\$ 784	\$ 934	\$ 284	\$ 814	\$ 964	\$ 684	\$ 754	\$ 394	\$ 864	\$ 1,274
Games	1997	\$ 1,014	\$ 1,164	\$ 364	\$ 1,044	\$ 1,194	\$ 814	\$ 924	\$ 464	\$ 1,034	\$ 1,084
Games	1998	\$ 1,094	\$ 1,244	\$ 394	\$ 1,124	\$ 1,274	\$ 894	\$ 1,004	\$ 504	\$ 1,114	\$ 1,164
Health & Beauty	1997	\$ 1,014	\$ 1,164	\$ 364	\$ 1,044	\$ 1,194	\$ 814	\$ 924	\$ 464	\$ 1,034	\$ 1,084
Health & Beauty	1998	\$ 1,094	\$ 1,244	\$ 394	\$ 1,124	\$ 1,274	\$ 894	\$ 1,004	\$ 504	\$ 1,114	\$ 1,164
Household	1997	\$ 1,014	\$ 1,164	\$ 364	\$ 1,044	\$ 1,194	\$ 814	\$ 924	\$ 464	\$ 1,034	\$ 1,084
Household	1998	\$ 1,094	\$ 1,244	\$ 394	\$ 1,124	\$ 1,274	\$ 894	\$ 1,004	\$ 504	\$ 1,114	\$ 1,164
Soft & Hardware	1997	\$ 1,014	\$ 1,164	\$ 364	\$ 1,044	\$ 1,194	\$ 814	\$ 924	\$ 464	\$ 1,034	\$ 1,084
Soft & Hardware	1998	\$ 1,094	\$ 1,244	\$ 394	\$ 1,124	\$ 1,274	\$ 894	\$ 1,004	\$ 504	\$ 1,114	\$ 1,164
Travel	1997	\$ 1,014	\$ 1,164	\$ 364	\$ 1,044	\$ 1,194	\$ 814	\$ 924	\$ 464	\$ 1,034	\$ 1,084
Travel	1998	\$ 1,094	\$ 1,244	\$ 394	\$ 1,124	\$ 1,274	\$ 894	\$ 1,004	\$ 504	\$ 1,114	\$ 1,164



Category	Year	Market (Order Dates)									
		1997	1998	1997	1998	1997	1998	1997	1998	1997	1998
Economics		\$ 1,124	\$ 1,274	\$ 384	\$ 1,108	\$ 1,344	\$ 1,014	\$ 1,134	\$ 564	\$ 1,194	\$ 1,464
Food		\$ 704	\$ 854	\$ 254	\$ 734	\$ 884	\$ 604	\$ 674	\$ 334	\$ 744	\$ 1,274
Games		\$ 1,014	\$ 1,164	\$ 364	\$ 1,044	\$ 1,194	\$ 814	\$ 924	\$ 464	\$ 1,034	\$ 1,164
Health & Beauty		\$ 1,014	\$ 1,164	\$ 364	\$ 1,044	\$ 1,194	\$ 814	\$ 924	\$ 464	\$ 1,034	\$ 1,164
Household		\$ 1,014	\$ 1,164	\$ 364	\$ 1,044	\$ 1,194	\$ 814	\$ 924	\$ 464	\$ 1,034	\$ 1,164
Soft & Hardware		\$ 1,014	\$ 1,164	\$ 364	\$ 1,044	\$ 1,194	\$ 814	\$ 924	\$ 464	\$ 1,034	\$ 1,164
Travel		\$ 1,014	\$ 1,164	\$ 364	\$ 1,044	\$ 1,194	\$ 814	\$ 924	\$ 464	\$ 1,034	\$ 1,164

Tratto da Goffarelli, Rizzi, "Data warehouse, teoria e pratica della progettazione", McGraw Hill 2006  
 Copyright - Tutti i diritti riservati

Elena Baralis  
 Politecnico di Torino

INTRODUZIONE - 67