

# ***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 antiriciclaggio
- 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)

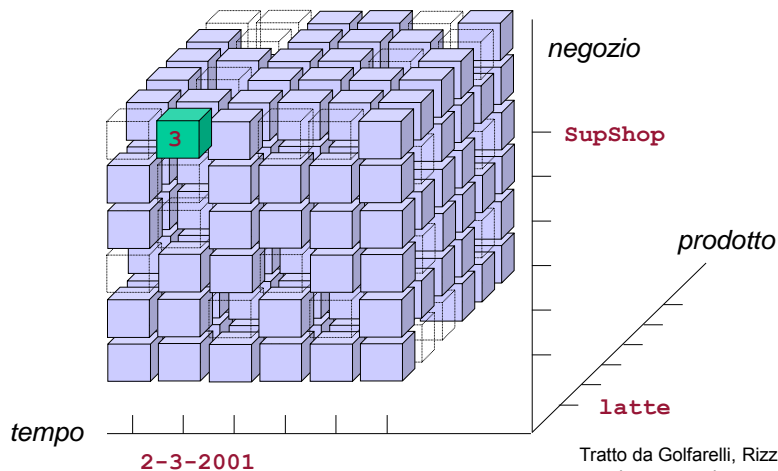
## *Struttura e analisi dei dati*

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



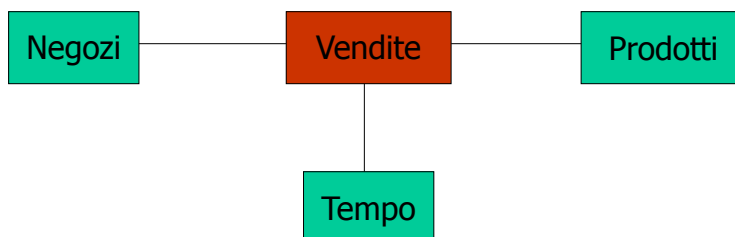
Tratto da Golfarelli, Rizzi, "Data warehouse, teoria e pratica della progettazione", McGraw Hill 2006

## 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:  
 $730 \times 300 \times 3000 = 657$  milioni

⇒ spazio occupato dalla tabella dei fatti  $\approx$  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*

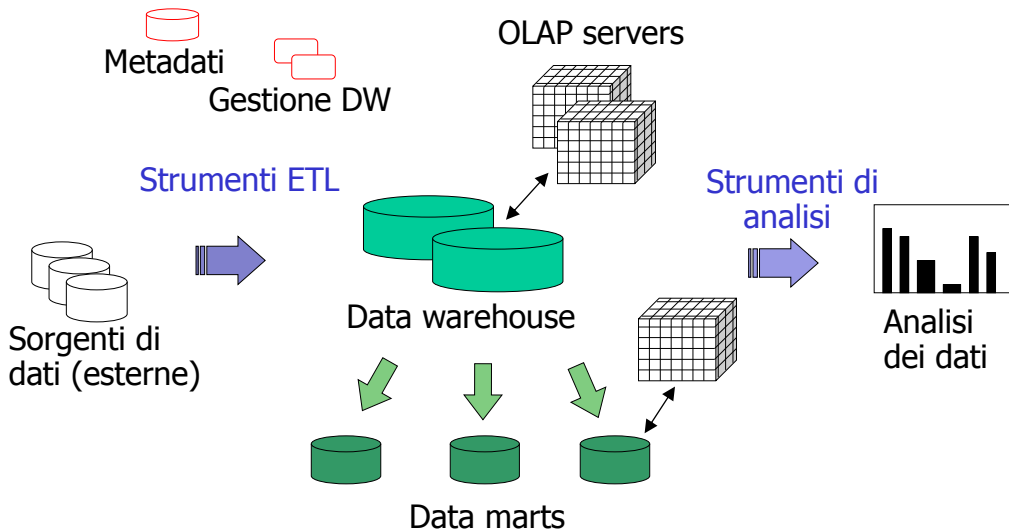
Elena Baralis  
Politecnico di Torino



## 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)

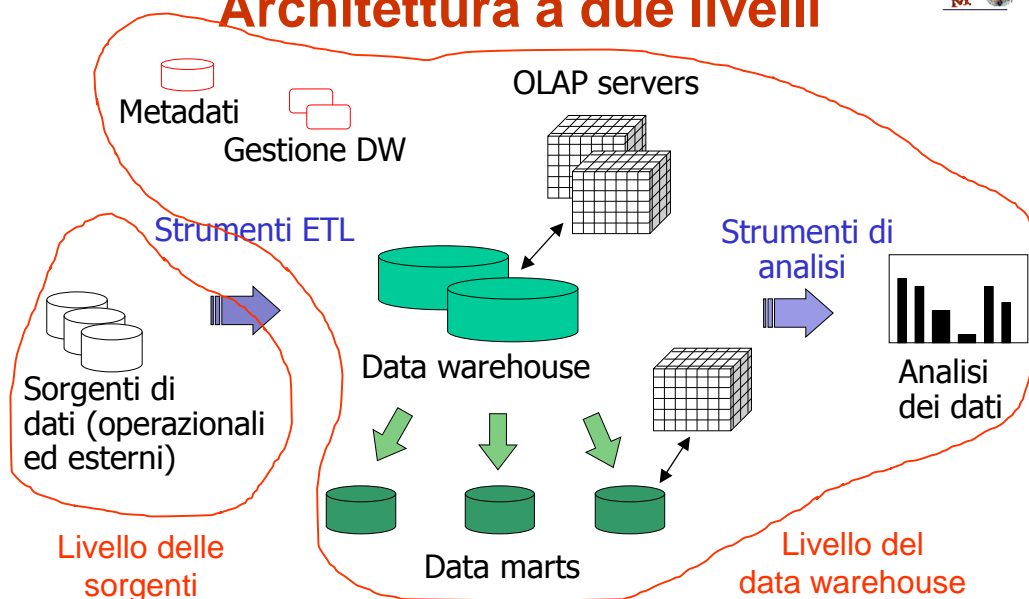
*Caricamento:* propagazione degli aggiornamenti al data warehouse

## 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
    - CWMI (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

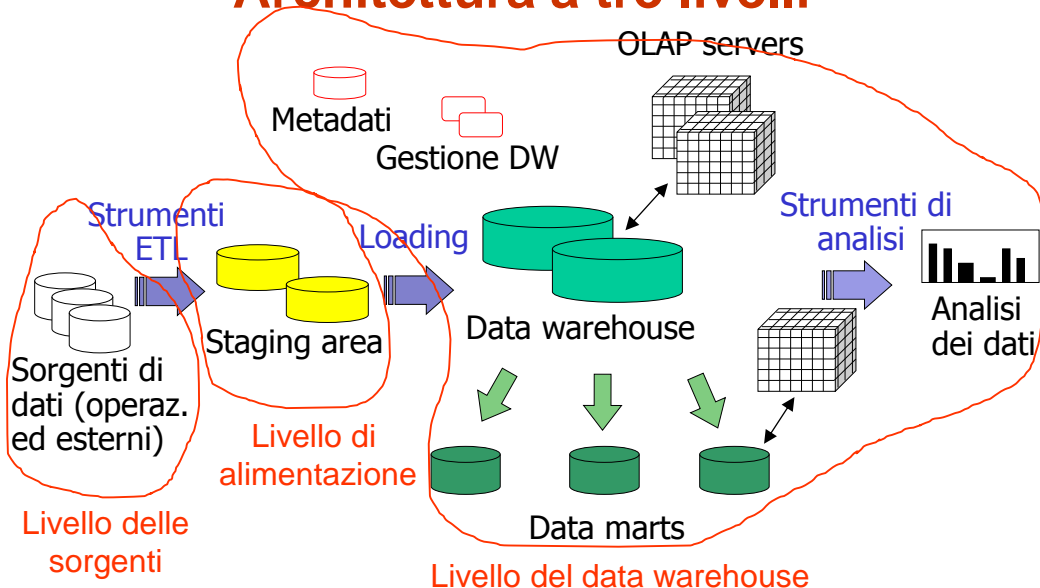
## Architettura a due livelli



## 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)

## Architettura a tre livelli



## 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

## Data warehouse Progettazione

Elena Baralis  
Politecnico di Torino

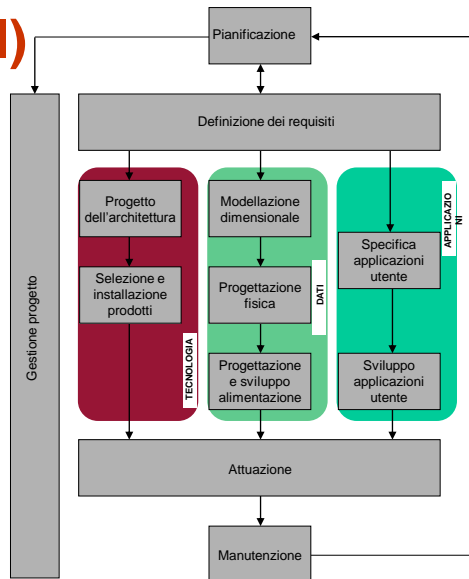
## Fattori di rischio

- 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

## 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

# Business Dimensional Lifecycle (Kimball)



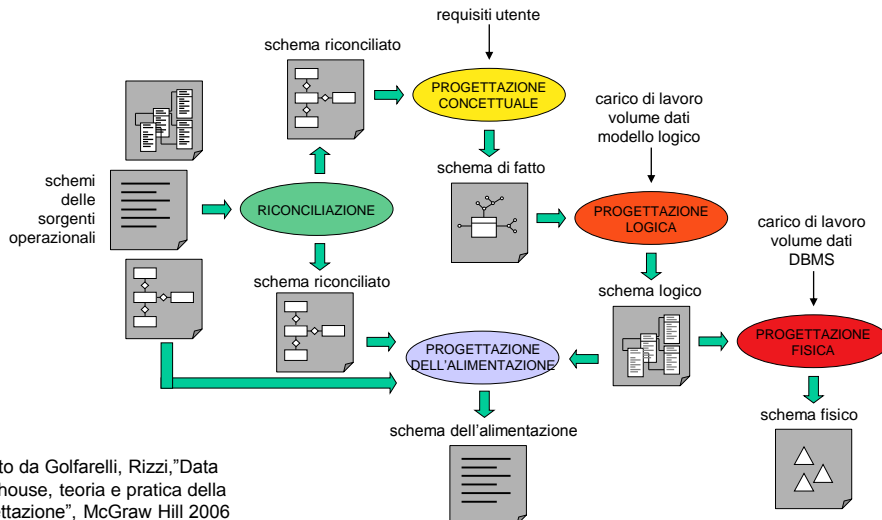
Tratto da Golfarelli, Rizzi, "Data warehouse, teoria e pratica della progettazione", McGraw Hill 2006

Elena Baralis  
Politecnico di Torino

Copyright – Tutti i diritti riservati

INTRODUZIONE - 31

# Progettazione di data mart



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

Elena Baralis  
Politecnico di Torino

Copyright – Tutti i diritti riservati

INTRODUZIONE - 32



## *Analisi dei requisiti*

Elena Baralis  
Politecnico di Torino

## *Analisi dei requisiti*

- 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

## 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

## 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

## **Progettazione concettuale**

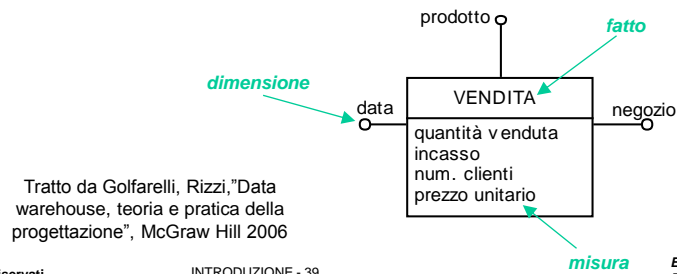
Elena Baralis  
Politecnico di Torino

## **Progettazione concettuale**

- Non esiste un formalismo di modellazione comunemente accettato
  - il modello ER non è adatto
- Dimensional Fact Model (Golfarelli, 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)



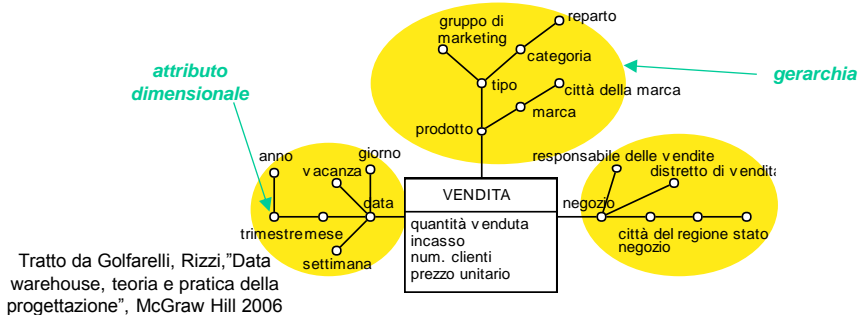
Copyright – Tutti i diritti riservati

INTRODUZIONE - 39

Elena Baralis  
Politecnico di Torino

## 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)

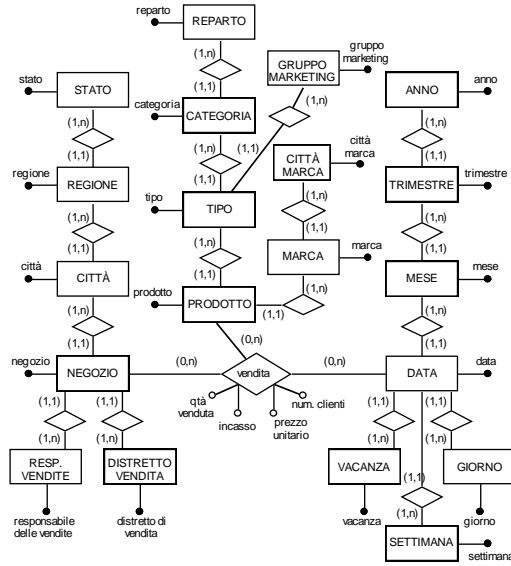


Copyright – Tutti i diritti riservati

INTRODUZIONE - 40

Elena Baralis  
Politecnico di Torino

# Corrispondenza con l'ER



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

Copyright – Tutti i diritti riservati

INTRODUZIONE - 41

Elena Baralis  
Politecnico di Torino

# Aggregazioni

Elena Baralis  
Politecnico di Torino

Copyright – Tutti i diritti riservati

INTRODUZIONE - 42

Elena Baralis  
Politecnico di Torino

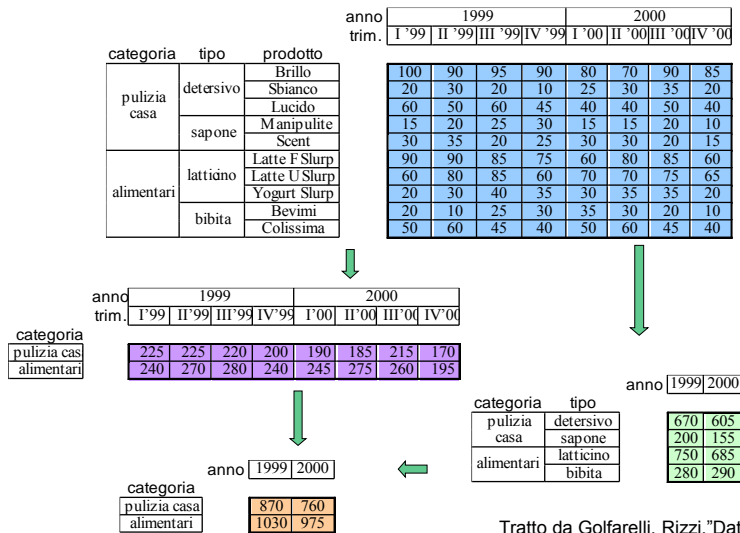
## Aggregazione

- 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



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

Elena Baralis  
Politecnico di Torino

Copyright – Tutti i diritti riservati

INTRODUZIONE - 45

## 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

Copyright – Tutti i diritti riservati

INTRODUZIONE - 46

Elena Baralis  
Politecnico di Torino

## 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

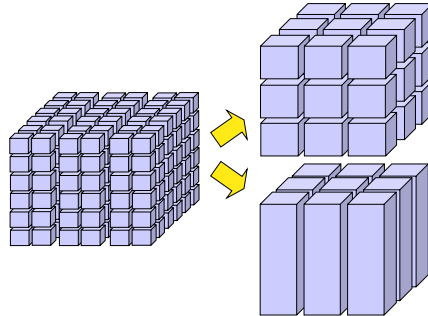
## Analisi OLAP

- Operazioni di ricerca disponibili
  - roll up, drill down
  - slice and dice
  - pivot di tabelle
  - ordinamento
- Le operazioni possono essere
  - combinate 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 Golfarelli, Rizzi, "Data warehouse, teoria e pratica della progettazione", McGraw Hill 2006

Copyright – Tutti i diritti riservati

INTRODUZIONE - 53

Elena Baralis  
Politecnico di Torino

## Roll up

Metrics Customer Region	Dollar Sales North-East	Mid-Atlantic	South-East	Central	South	North-West	South-West	England	France	Germany	Canada
Month											
Jan 97	\$ 620	\$ 753	\$ 30	\$ 660	\$ 2.405	\$ 1.312	\$ 440	\$ 1.002	\$ 1.002	\$ 383	\$ 210
Feb 97	\$ 258	\$ 252	\$ 800	\$ 975	\$ 160	\$ 582	\$ 744	\$ 310	\$ 799	\$ 118	\$ 337
Mar 97	\$ 648	\$ 244	\$ 148	\$ 250	\$ 1.085	\$ 2.961	\$ 650	\$ 1.240	\$ 119	\$ 142	\$ 96
Apr 97	\$ 787	\$ 588	\$ 447	\$ 486	\$ 226	\$ 506	\$ 601	\$ 119	\$ 550	\$ 85	
May 97	\$ 1.350	\$ 245	\$ 936	\$ 159	\$ 664	\$ 626	\$ 107	\$ 135	\$ 200	\$ 177	\$ 230
Jun 97	\$ 842	\$ 582	\$ 1.281	\$ 937	\$ 240	\$ 774	\$ 176	\$ 1.139	\$ 552	\$ 254	\$ 745
Jul 97	\$ 652	\$ 690	\$ 486	\$ 1.293	\$ 605	\$ 303	\$ 818	\$ 103	\$ 124	\$ 173	\$ 66
Aug 97	\$ 1.783	\$ 304	\$ 1.032	\$ 170	\$ 398	\$ 356	\$ 432	\$ 190	\$ 241	\$ 407	\$ 259
Sep 97	\$ 581	\$ 778	\$ 3.558	\$ 587	\$ 440	\$ 1.652	\$ 1.071	\$ 315	\$ 210	\$ 202	
Oct 97	\$ 2.293	\$ 1.840	\$ 600	\$ 656	\$ 1.900	\$ 718	\$ 1.210	\$ 427	\$ 320	\$ 520	\$ 65
Nov 97	\$ 39	\$ 1.602	\$ 1.082	\$ 1.187	\$ 842	\$ 759	\$ 745	\$ 232	\$ 101	\$ 1.037	\$ 37
Dec 97	\$ 381	\$ 1.588	\$ 343	\$ 118	\$ 1.459	\$ 635	\$ 2.021	\$ 259	\$ 210	\$ 119	\$ 189
Jan 98	\$ 311	\$ 1.174	\$ 2.634	\$ 3.130	\$ 954	\$ 2.083	\$ 1.351	\$ 747	\$ 426	\$ 447	\$ 1.141
Feb 98	\$ 2.918	\$ 702	\$ 1.123	\$ 1.336	\$ 1.227	\$ 3.887	\$ 546	\$ 268	\$ 277	\$ 282	
Mar 98	\$ 2.459	\$ 1.523	\$ 1.178	\$ 4.708	\$ 1.420	\$ 3.514	\$ 1.948	\$ 1.705	\$ 276	\$ 1.168	\$ 63
Apr 98	\$ 407	\$ 841	\$ 524	\$ 712	\$ 133	\$ 2.486	\$ 49	\$ 390	\$ 1.298	\$ 221	\$ 46
May 98	\$ 657	\$ 1.721	\$ 440	\$ 148	\$ 80	\$ 1.310	\$ 303	\$ 104	\$ 657	\$ 65	
Jun 98	\$ 699	\$ 1.096	\$ 898	\$ 353	\$ 902	\$ 839	\$ 230	\$ 155	\$ 105	\$ 75	
Jul 98	\$ 586	\$ 1.897	\$ 412	\$ 226	\$ 406	\$ 361	\$ 1.628	\$ 267	\$ 1.011	\$ 41	\$ 184
Aug 98	\$ 894	\$ 326	\$ 792	\$ 1.832	\$ 1.199	\$ 295	\$ 1.816	\$ 277	\$ 102	\$ 118	\$ 115
Sep 98	\$ 338	\$ 3.179	\$ 505	\$ 427	\$ 99	\$ 2.976	\$ 885	\$ 135	\$ 85	\$ 1.110	\$ 510
Oct 98	\$ 544	\$ 413	\$ 1.467	\$ 209	\$ 679	\$ 706	\$ 556	\$ 480	\$ 485	\$ 99	\$ 160
Nov 98	\$ 671	\$ 459	\$ 1.471	\$ 2.066	\$ 701	\$ 716	\$ 986	\$ 1.127	\$ 154	\$ 440	\$ 361
Dec 98	\$ 836	\$ 2.096	\$ 1.726	\$ 3.642	\$ 995	\$ 1.740	\$ 1.943	\$ 1.143	\$ 366	\$ 307	\$ 118



Metrics Customer Region	Dollar Sales North-East	Mid-Atlantic	South-East	Central	South	North-West	South-West	England	France	Germany	Canada
Quarter											
Q1 1997	\$ 1.526	\$ 1.249	\$ 978	\$ 1.885	\$ 3.650	\$ 4.855	\$ 1.834	\$ 2.552	\$ 1.920	\$ 643	\$ 663
Q2 1997	\$ 2.979	\$ 1.415	\$ 2.664	\$ 1.582	\$ 1.130	\$ 1.906	\$ 884	\$ 1.393	\$ 1.402	\$ 516	\$ 975
Q3 1997	\$ 3.016	\$ 1.772	\$ 5.076	\$ 2.050	\$ 1.443	\$ 2.311	\$ 2.321	\$ 608	\$ 575	\$ 782	\$ 325
Q4 1997	\$ 2.711	\$ 5.030	\$ 2.025	\$ 1.961	\$ 3.601	\$ 2.112	\$ 3.976	\$ 918	\$ 531	\$ 1.675	\$ 291
Q1 1998	\$ 5.288	\$ 3.399	\$ 4.935	\$ 9.174	\$ 3.601	\$ 9.484	\$ 3.044	\$ 2.720	\$ 979	\$ 1.897	\$ 1.204
Q2 1998	\$ 1.773	\$ 3.658	\$ 1.862	\$ 1.213	\$ 1.115	\$ 4.635	\$ 352	\$ 724	\$ 2.110	\$ 391	\$ 121
Q3 1998	\$ 1.818	\$ 5.402	\$ 1.709	\$ 2.485	\$ 1.704	\$ 3.632	\$ 4.329	\$ 679	\$ 1.198	\$ 1.269	\$ 809
Q4 1998	\$ 2.051	\$ 2.968	\$ 4.664	\$ 5.917	\$ 1.775	\$ 3.162	\$ 3.485	\$ 2.750	\$ 1.005	\$ 846	\$ 639

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

Copyright – Tutti i diritti riservati

INTRODUZIONE - 54

Elena Baralis  
Politecnico di Torino

## Roll up

Category	Year	Metrics Customer Region	Dollar Sales									
			North-East	Mid-Atlantic	South-East	Central	South	North-West	South-West	England	France	Germa
Electronics	1997		\$ 138	\$ 1.774	\$ 384	\$ 138	\$ 2.346	\$ 2.554	\$ 2.184	\$ 566	\$ 199	\$
	1998		\$ 1.184	\$ 4.529	\$ 1.892	\$ 7.222	\$ 651	\$ 9.488	\$ 476	\$ 2.683	\$ 462	\$ 7
Food	1997		\$ 759	\$ 682	\$ 729	\$ 262	\$ 989	\$ 469	\$ 807	\$ 156	\$ 615	\$ 1
	1998		\$ 538	\$ 925	\$ 959	\$ 677	\$ 213	\$ 1.503	\$ 261	\$ 165	\$ 175	\$ 1
Gifts	1997		\$ 2.532	\$ 1.355	\$ 1.854	\$ 1.413	\$ 2.535	\$ 2.132	\$ 1.904	\$ 908	\$ 375	\$ 1.0
	1998		\$ 1.955	\$ 2.785	\$ 2.800	\$ 2.695	\$ 1.813	\$ 2.844	\$ 1.778	\$ 1.158	\$ 717	\$ 6
Health & Beauty	1997		\$ 624	\$ 640	\$ 1.217	\$ 647	\$ 588	\$ 754	\$ 654	\$ 143	\$ 292	\$ 3
	1998		\$ 611	\$ 887	\$ 566	\$ 382	\$ 499	\$ 1.162	\$ 1.044	\$ 273	\$ 72	
Household	1997		\$ 5.354	\$ 4.112	\$ 5.410	\$ 4.446	\$ 3.058	\$ 3.974	\$ 2.654	\$ 3.545	\$ 2.875	\$ 1.9
	1998		\$ 5.787	\$ 5.320	\$ 5.416	\$ 6.812	\$ 4.334	\$ 5.008	\$ 7.588	\$ 2.139	\$ 3.649	\$ 2.7
Kid's Korner	1997		\$ 201	\$ 398	\$ 485	\$ 186	\$ 409	\$ 323	\$ 396	\$ 105	\$ 34	\$
	1998		\$ 247	\$ 422	\$ 441	\$ 380	\$ 221	\$ 592	\$ 290	\$ 198	\$ 19	\$
Travel	1997		\$ 624	\$ 505	\$ 564	\$ 386	\$ 300	\$ 978	\$ 416	\$ 48	\$ 38	\$
	1998		\$ 608	\$ 359	\$ 1.096	\$ 611	\$ 464	\$ 316	\$ 575	\$ 257	\$ 198	\$



Category	Year	Metrics Dollar Sales	
		Year	Dollar Sales
Electronics	1997		\$ 10.616
	1998		\$ 29.299
Food	1997		\$ 5.300
	1998		\$ 5.638
Gifts	1997		\$ 15.215
	1998		\$ 20.047
Health & Beauty	1997		\$ 6.042
	1998		\$ 5.685
Household	1997		\$ 38.283
	1998		\$ 50.391
Kid's Korner	1997		\$ 2.559
	1998		\$ 2.943
Travel	1997		\$ 4.497
	1998		\$ 4.792

Tratto da Golfarelli, 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

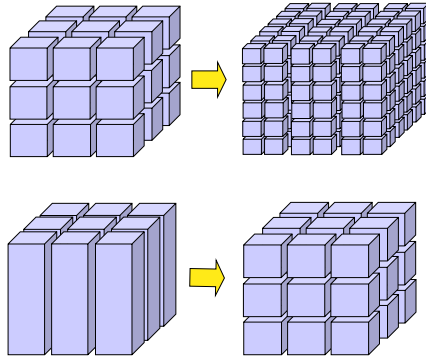
- Aumento di dettaglio dei dati mediante
  - l'aumento del livello di dettaglio di una delle dimensioni presenti, con la riduzione di livello in una gerarchia
    - esempio: da raggruppamento per città e mese a raggruppamento per negozio e mese
  - l'aggiunta di una nuova dimensione
    - esempio: da 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 Golfarelli, 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

Metrics Customer Region	Dollar Sales										
	North-East	Mid-Atlantic	South-East	Central	South	North-West	South-West	England	France	Germany	Canada
Quarter Q1 1997	\$ 1,526	\$ 1,249	\$ 978	\$ 1,885	\$ 3,650	\$ 4,855	\$ 1,834	\$ 2,552	\$ 1,920	\$ 643	\$ 663
Q2 1997	\$ 2,979	\$ 1,415	\$ 2,664	\$ 1,582	\$ 1,130	\$ 1,906	\$ 884	\$ 1,393	\$ 1,402	\$ 516	\$ 975
Q3 1997	\$ 3,016	\$ 1,772	\$ 5,076	\$ 2,050	\$ 1,443	\$ 2,311	\$ 2,321	\$ 608	\$ 575	\$ 782	\$ 325
Q4 1997	\$ 2,711	\$ 5,030	\$ 2,028	\$ 1,961	\$ 3,601	\$ 2,112	\$ 3,976	\$ 918	\$ 531	\$ 1,676	\$ 291
Q1 1998	\$ 5,288	\$ 3,399	\$ 4,935	\$ 9,174	\$ 3,601	\$ 9,484	\$ 3,844	\$ 2,720	\$ 979	\$ 1,897	\$ 1,204
Q2 1998	\$ 1,773	\$ 3,658	\$ 1,862	\$ 1,213	\$ 1,115	\$ 4,635	\$ 352	\$ 724	\$ 2,110	\$ 391	\$ 121
Q3 1998	\$ 1,818	\$ 5,402	\$ 1,709	\$ 2,485	\$ 1,704	\$ 3,632	\$ 4,329	\$ 679	\$ 1,198	\$ 1,269	\$ 809
Q4 1998	\$ 2,051	\$ 2,968	\$ 4,664	\$ 5,917	\$ 1,775	\$ 3,162	\$ 3,485	\$ 2,750	\$ 1,005	\$ 646	\$ 639



Metrics Customer City	Dollar Sales												
	Arlin	San Pedro	Springfield	Chappel Hill	Scranburg	Pebble Beach	Martinsville	Maddon	Peoria	Pecos	Lake Barkley	Alcameda	Fingers Lake
Quarter Q1 1997	\$ 675										\$ 39		
Q2 1997				\$ 203						\$ 53			\$ 135
Q3 1997				\$ 276								\$ 252	\$ 63
Q4 1997	\$ 215	\$ 124			\$ 113	\$ 45	\$ 192	\$ 348				\$ 79	\$ 98
Q1 1998			\$ 140	\$ 174			\$ 85	\$ 12	\$ 17		\$ 237	\$ 30	\$ 119
Q2 1998													
Q3 1998	\$ 734					\$ 25	\$ 1,535						
Q4 1998						\$ 219	\$ 119	\$ 142		\$ 85	\$ 1,533		

Tratto da Golfarelli, 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

Category	Metrics Dollar Sales		
	Year	1997	1998
Electronics		\$ 10.616	\$ 29.299
Food		\$ 5.300	\$ 5.638
Gifts		\$ 16.315	\$ 20.047
Health & Beauty		\$ 6.042	\$ 5.665
Household		\$ 38.383	\$ 50.391
Kid's Korner		\$ 2.559	\$ 2.943
Travel		\$ 4.497	\$ 4.792



Category	Metrics Customer Region Year	Dollar Sales											
		North-East		Mid-Atlantic		South-East		Central		South		North-West	
		1997	1998	1997	1998	1997	1998	1997	1998	1997	1998	1997	1998
Electronics		\$ 138	\$ 1.184	\$ 1.774	\$ 4.529	\$ 384	\$ 1.892	\$ 138	\$ 7.232	\$ 2.346	\$ 651	\$ 2.554	\$ 9.488
Food		\$ 759	\$ 538	\$ 682	\$ 928	\$ 729	\$ 959	\$ 262	\$ 973	\$ 588	\$ 213	\$ 469	\$ 1.503
Gifts		\$ 2.532	\$ 1.955	\$ 1.355	\$ 2.785	\$ 1.854	\$ 2.800	\$ 1.413	\$ 2.695	\$ 2.535	\$ 1.813	\$ 2.132	\$ 2.844
Health & Beauty		\$ 624	\$ 611	\$ 640	\$ 887	\$ 1.317	\$ 566	\$ 647	\$ 382	\$ 588	\$ 499	\$ 754	\$ 1.162
Household		\$ 5.354	\$ 5.787	\$ 4.112	\$ 5.320	\$ 5.410	\$ 5.416	\$ 4.446	\$ 6.812	\$ 3.058	\$ 4.334	\$ 3.974	\$ 5.008
Kid's Korner		\$ 201	\$ 247	\$ 398	\$ 422	\$ 485	\$ 441	\$ 186	\$ 380	\$ 408	\$ 221	\$ 323	\$ 592
Travel		\$ 624	\$ 608	\$ 505	\$ 559	\$ 564	\$ 1.096	\$ 386	\$ 611	\$ 300	\$ 464	\$ 978	\$ 316

Tratto da Golfarelli, 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

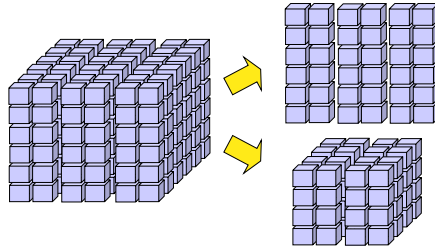
- Riduzione del volume dei dati da analizzare
  - selezione di un sottoinsieme mediante predicati
    - slice: predicato di uguaglianza che seleziona una "fetta"
      - esempio: Anno=2005
    - dice: combinazione di predicati che seleziona un "cubetto"
      - esempio: Categoria='Alimentari' and Città='Torino'

Copyright – Tutti i diritti riservati

INTRODUZIONE - 60

Elena Baralis  
Politecnico di Torino

## Slice and dice



Tratto da Golfarelli, 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	Metrics Customer Region									
		North-East	Mid-Atlantic	South-East	Central	South	North-West	South-West	England	France	Germa
Electronics	1997	\$ 138	\$ 1,774	\$ 384	\$ 138	\$ 2,346	\$ 2,554	\$ 2,184	\$ 566	\$ 199	\$
	1998	\$ 1,184	\$ 4,529	\$ 1,892	\$ 7,232	\$ 651	\$ 9,488	\$ 476	\$ 2,683	\$ 462	\$ 7
Food	1997	\$ 759	\$ 682	\$ 729	\$ 262	\$ 588	\$ 469	\$ 807	\$ 156	\$ 615	\$ 1
	1998	\$ 538	\$ 925	\$ 959	\$ 677	\$ 213	\$ 1,503	\$ 261	\$ 165	\$ 175	\$ 1
Gifts	1997	\$ 2,532	\$ 1,355	\$ 1,854	\$ 1,413	\$ 2,535	\$ 2,132	\$ 1,904	\$ 908	\$ 375	\$ 1.0
	1998	\$ 1,955	\$ 2,785	\$ 2,800	\$ 2,695	\$ 1,813	\$ 2,844	\$ 1,778	\$ 1,158	\$ 717	\$ 6
Health & Beauty	1997	\$ 624	\$ 640	\$ 1,317	\$ 647	\$ 588	\$ 754	\$ 654	\$ 143	\$ 292	\$ 3
	1998	\$ 611	\$ 887	\$ 566	\$ 382	\$ 499	\$ 1,162	\$ 1,044	\$ 273	\$ 72	\$
Household	1997	\$ 5,354	\$ 4,112	\$ 5,410	\$ 4,446	\$ 3,058	\$ 3,974	\$ 2,654	\$ 3,545	\$ 2,875	\$ 1.9
	1998	\$ 5,787	\$ 5,320	\$ 5,416	\$ 6,812	\$ 4,334	\$ 5,008	\$ 7,588	\$ 2,139	\$ 3,649	\$ 2.7
Kid's Korner	1997	\$ 201	\$ 398	\$ 485	\$ 186	\$ 409	\$ 323	\$ 396	\$ 105	\$ 34	\$
	1998	\$ 247	\$ 422	\$ 441	\$ 380	\$ 221	\$ 592	\$ 290	\$ 198	\$ 19	\$
Travel	1997	\$ 624	\$ 505	\$ 564	\$ 386	\$ 300	\$ 978	\$ 416	\$ 48	\$ 38	\$
	1998	\$ 608	\$ 559	\$ 1,096	\$ 611	\$ 464	\$ 316	\$ 573	\$ 257	\$ 198	\$



File Detail:											
Year = 1998											
Category	Metrics Customer Region	Dollar Sales									
		North-East	Mid-Atlantic	South-East	Central	South	North-West	South-West	England	France	Germany
Electronics		\$ 1,184	\$ 4,529	\$ 1,892	\$ 7,232	\$ 651	\$ 9,488	\$ 476	\$ 2,683	\$ 462	\$ 702
Food		\$ 538	\$ 925	\$ 959	\$ 677	\$ 213	\$ 1,503	\$ 261	\$ 165	\$ 175	\$ 100
Gifts		\$ 1,955	\$ 2,785	\$ 2,800	\$ 2,695	\$ 1,813	\$ 2,844	\$ 1,778	\$ 1,158	\$ 717	\$ 686
Health & Beauty		\$ 611	\$ 887	\$ 566	\$ 382	\$ 499	\$ 1,162	\$ 1,044	\$ 273	\$ 72	\$
Household		\$ 5,787	\$ 5,320	\$ 5,416	\$ 6,812	\$ 4,334	\$ 5,008	\$ 7,588	\$ 2,139	\$ 3,649	\$ 2,791
Kid's Korner		\$ 247	\$ 422	\$ 441	\$ 380	\$ 221	\$ 592	\$ 290	\$ 198	\$ 19	\$ 69
Travel		\$ 608	\$ 559	\$ 1,096	\$ 611	\$ 464	\$ 316	\$ 573	\$ 257	\$ 198	\$ 55

Tratto da Golfarelli, 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

Subcategory	Metrics Customer City	Dollar Sales											
		Alton	Akron	Albon	Alcameda	Alka	Allagash	Alta	Altoola	Amestra	Amsterdam	Andersonville	Annap
Audio						\$ 85							
Automotive							\$ 30						
Chocolate		\$ 42	\$ 42		\$ 50		\$ 20	\$ 22	\$ 44				\$
Christmas		\$ 30				\$ 25	\$ 30	\$ 15					
Classic Toys						\$ 7	\$ 26					\$ 38	
Coffee				\$ 9									
Comfort					\$ 59	\$ 59							
Furniture							\$ 485						
Gadgets							\$ 199	\$ 79	\$ 79				
Games & Puzzles							\$ 17	\$ 45			\$ 45		
Gift Baskets				\$ 55	\$ 43								\$
Golf		\$ 25						\$ 25	\$ 14			\$ 25	
Hearth									\$ 15				
Jewelry		\$ 75			\$ 189	\$ 24	\$ 77	\$ 189	\$ 24				
Kitchen						\$ 55	\$ 21		\$ 75				\$ :
Lawn & Garden		\$ 75	\$ 100		\$ 15	\$ 63	\$ 100	\$ 180	\$ 67	\$ 40		\$ :	
Learning		\$ 16						\$ 37					
Meat & Cheese		\$ 40			\$ 20	\$ 200	\$ 139	\$ 20				\$ 25	
Miscellaneous		\$ 200	\$ 1,320						\$ 993				
Natural Remedies		\$ 13							\$ 13				
Pets		\$ 215	\$ 26	\$ 26			\$ 30	\$ 68	\$ 115	\$ 25		\$ 34	\$ :
Plants & Flowers		\$ 65	\$ 65	\$ 65				\$ 50	\$ 60				\$ :
Safety & Security								\$ 30	\$ 22		\$ 22		
Skin Care													
Sleeping				\$ 18									
Toys & Accessories							\$ 29	\$ 185	\$ 744				\$ :



Filter Details:  
 Category = Electronics  
 AND  
 Dollar Sales > 80  
 AND  
 Customer Region = North-West  
 AND  
 Year = 1997

Subcategory	Metrics Customer City	Dollar Sales						
		Alta	Armstrong	Avery Heights	Lane	Mt. Everest	San Fransisco	
Audio			\$ 98		\$ 123	\$ 85		
Comfort				\$ 118		\$ 1,495		
Gadgets		\$ 199					\$ 199	

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

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

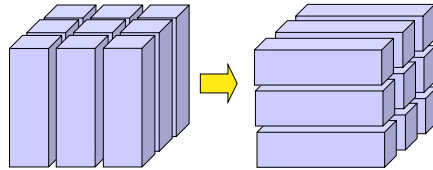
Copyright – Tutti i diritti riservati

INTRODUZIONE - 64

Elena Baralis  
Politecnico di Torino



## Pivot



Tratto da Golfarelli, 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	Metrics	Dollar Sales
	Year	
Electronics	1997	\$ 10.616
Electronics	1998	\$ 29.299
Food	1997	\$ 5.300
Food	1998	\$ 5.638
Gifts	1997	\$ 16.315
Gifts	1998	\$ 20.047
Health & Beauty	1997	\$ 6.042
Health & Beauty	1998	\$ 5.665
Household	1997	\$ 38.383
Household	1998	\$ 50.391
Kid's Korner	1997	\$ 2.559
Kid's Korner	1998	\$ 2.943
Travel	1997	\$ 4.497
Travel	1998	\$ 4.792



Category	Metrics	Dollar Sales	
	Year	1997	1998
Electronics		\$ 10.616	\$ 29.299
Food		\$ 5.300	\$ 5.638
Gifts		\$ 16.315	\$ 20.047
Health & Beauty		\$ 6.042	\$ 5.665
Household		\$ 38.383	\$ 50.391
Kid's Korner		\$ 2.559	\$ 2.943
Travel		\$ 4.497	\$ 4.792

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

Copyright – Tutti i diritti riservati

INTRODUZIONE - 66

Elena Baralis  
Politecnico di Torino

## Pivot

Category	Year	Metrics Customer Region									
		North-East	Mid-Atlantic	South-East	Central	South	North-West	South-West	England	France	Germa
Electronics	1997	\$ 138	\$ 1,774	\$ 384	\$ 138	\$ 2,346	\$ 2,554	\$ 2,184	\$ 566	\$ 199	\$
Electronics	1998	\$ 1,184	\$ 4,529	\$ 1,892	\$ 7,232	\$ 651	\$ 9,488	\$ 476	\$ 2,689	\$ 462	\$ 7
Food	1997	\$ 759	\$ 682	\$ 729	\$ 262	\$ 598	\$ 469	\$ 807	\$ 156	\$ 615	\$ 1
Food	1998	\$ 538	\$ 925	\$ 959	\$ 677	\$ 213	\$ 1,503	\$ 261	\$ 165	\$ 175	\$ 1
Gifts	1997	\$ 2,532	\$ 1,355	\$ 1,854	\$ 1,413	\$ 2,535	\$ 2,132	\$ 1,904	\$ 908	\$ 375	\$ 1.0
Gifts	1998	\$ 1,955	\$ 2,785	\$ 2,800	\$ 2,695	\$ 1,813	\$ 2,844	\$ 1,778	\$ 1,158	\$ 717	\$ 6
Health & Beauty	1997	\$ 624	\$ 640	\$ 1,317	\$ 647	\$ 588	\$ 754	\$ 654	\$ 143	\$ 292	\$ 3
Health & Beauty	1998	\$ 611	\$ 887	\$ 566	\$ 382	\$ 499	\$ 1,162	\$ 1,044	\$ 273	\$ 72	\$
Household	1997	\$ 5,354	\$ 4,112	\$ 5,410	\$ 4,446	\$ 3,058	\$ 3,974	\$ 2,654	\$ 3,545	\$ 2,875	\$ 1.9
Household	1998	\$ 5,787	\$ 5,320	\$ 5,416	\$ 6,812	\$ 4,334	\$ 5,008	\$ 7,588	\$ 2,139	\$ 3,649	\$ 2.7
Kid's Korner	1997	\$ 201	\$ 398	\$ 485	\$ 186	\$ 409	\$ 323	\$ 396	\$ 105	\$ 34	\$
Kid's Korner	1998	\$ 247	\$ 422	\$ 441	\$ 380	\$ 221	\$ 592	\$ 290	\$ 198	\$ 19	\$
Travel	1997	\$ 624	\$ 505	\$ 564	\$ 386	\$ 300	\$ 978	\$ 416	\$ 48	\$ 38	\$
Travel	1998	\$ 608	\$ 559	\$ 1,096	\$ 611	\$ 464	\$ 316	\$ 573	\$ 257	\$ 198	\$



Category	Year	North-East		Mid-Atlantic		South-East		Central		South		North-West	
		1997	1998	1997	1998	1997	1998	1997	1998	1997	1998	1997	1998
Electronics		\$ 138	\$ 1,184	\$ 1,774	\$ 4,529	\$ 384	\$ 1,892	\$ 138	\$ 7,232	\$ 2,346	\$ 651	\$ 2,554	\$ 9,488
Food		\$ 759	\$ 538	\$ 682	\$ 925	\$ 729	\$ 959	\$ 262	\$ 677	\$ 598	\$ 469	\$ 807	\$ 1,503
Gifts		\$ 2,532	\$ 1,955	\$ 1,355	\$ 2,785	\$ 1,854	\$ 2,800	\$ 1,413	\$ 2,695	\$ 2,535	\$ 1,813	\$ 2,132	\$ 2,844
Health & Beauty		\$ 624	\$ 611	\$ 640	\$ 887	\$ 1,317	\$ 566	\$ 647	\$ 382	\$ 588	\$ 499	\$ 754	\$ 1,162
Household		\$ 5,354	\$ 5,787	\$ 4,112	\$ 5,320	\$ 5,410	\$ 5,416	\$ 4,446	\$ 6,812	\$ 3,058	\$ 4,334	\$ 3,974	\$ 5,008
Kid's Korner		\$ 201	\$ 247	\$ 398	\$ 422	\$ 485	\$ 441	\$ 186	\$ 380	\$ 409	\$ 221	\$ 323	\$ 592
Travel		\$ 624	\$ 608	\$ 505	\$ 559	\$ 564	\$ 1,096	\$ 386	\$ 611	\$ 300	\$ 464	\$ 978	\$ 316

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

Copyright – Tutti i diritti riservati

INTRODUZIONE - 67

Elena Baralis  
Politecnico di Torino