

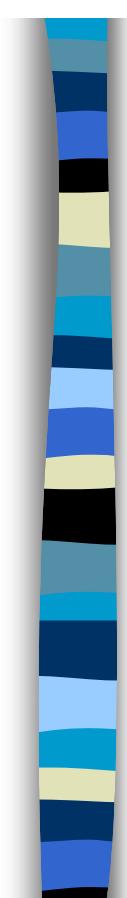


Il Data Warehousing

Prof. Stefano Rizzi
Alma Mater Studiorum - Università di Bologna

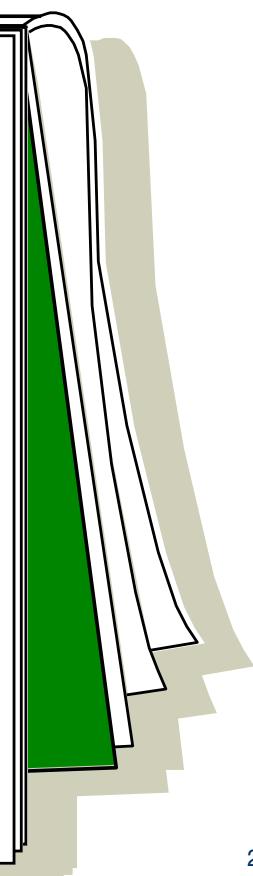


1



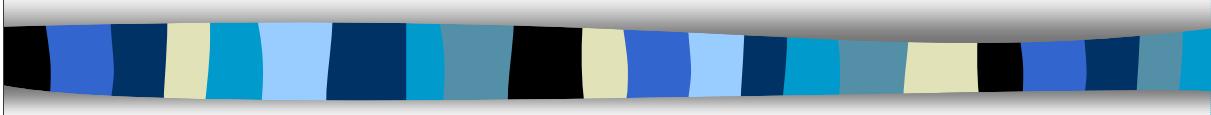
Sommario

- Il ruolo della business intelligence e del sistema informativo
 - ✓ Il ruolo dell'informatica in azienda
 - ✓ La "risorsa informazione"
- Introduzione al data warehousing
 - ✓ Caratteristiche del processo e architetture
 - ✓ Il modello multidimensionale
 - ✓ Accesso ai dati
- Il ciclo di sviluppo
 - ✓ Analisi e riconciliazione delle sorgenti
 - ✓ Analisi dei requisiti
 - ✓ Progettazione concettuale
 - ✓ Il carico di lavoro e il volume dati
 - ✓ Progettazione logica
 - ✓ Progettazione dell'alimentazione
 - ✓ Progettazione fisica



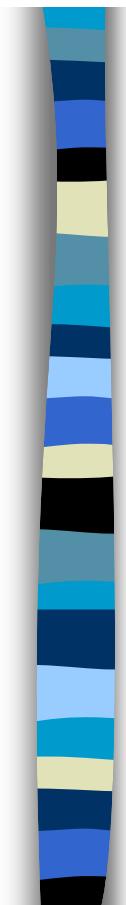
2

Il ruolo della BI e del sistema informativo



Prof. Stefano Rizzi

L'informatica in azienda

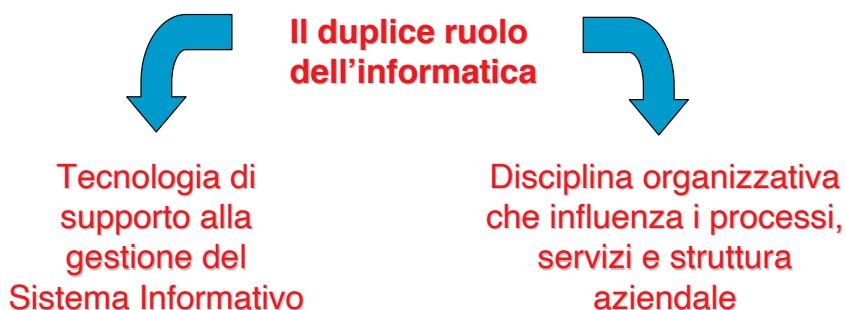


- La funzione svolta dalle basi di dati in ambito aziendale è stata fino a qualche anno fa solo quella di memorizzare dati operazionali, ossia dati generati da operazioni svolte all'interno dei processi gestionali
- L'informatica è vista come una scienza di supporto che permette di rendere più rapide ed economiche le operazioni di gestione delle informazioni ma che non crea di per sé ricchezza



L'evoluzione dei sistemi informativi

- Il ruolo dei Sistemi Informatici è radicalmente cambiato dai primi anni '70 a oggi. I sistemi informatici si sono trasformati da semplici **strumenti per migliorare l'efficienza dei processi** a **elementi centrali dell'organizzazione aziendale** in grado di rivoluzionare la struttura dei processi aziendali



5

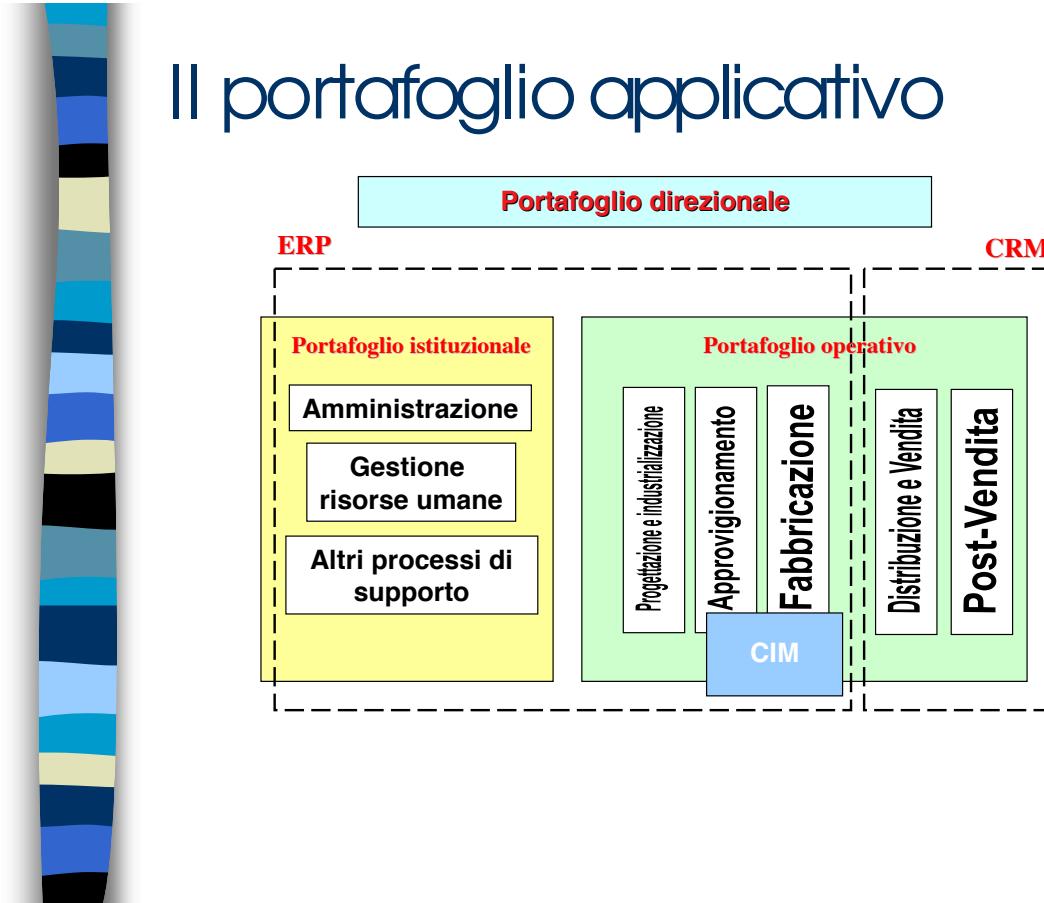


Motivazioni per migliorare il Sistema Informativo

- Trasformazione dell'economia:** l'economia moderna è basata sulle conoscenze e sull'informazione ed è caratterizzata da una breve vita dei prodotti che richiede decisioni tempestive
- Trasformazione dell'impresa:** per operare con profitto in un sistema economico altamente competitivo le aziende hanno bisogno di dotarsi di una struttura flessibile e snella in grado di reagire rapidamente alle mutate situazioni esterne/interne
- Globalizzazione:** con l'allargamento dei mercati a livello mondiale nasce l'esigenza del controllo di mercati a larga scala

6

Il portafoglio applicativo



7

Il portafoglio direzionale

- E' l'insieme delle applicazioni utilizzate dai manager aziendali per:
 - ✓ Analizzare lo stato dell'azienda
 - ✓ Prendere decisioni rapide
 - ✓ Prendere le decisioni migliori
- Si parla anche di *piattaforma per la Business Intelligence*, ossia...

Disciplina che consente a chi deve decidere in azienda di capire, attraverso soluzioni software, i fattori chiave del business e conseguentemente di prendere le migliori decisioni in quel momento

8



Business intelligence

- Si parla di piattaforma poiché per consentire ai manager analisi potenti e flessibili è necessario definire un'apposita infrastruttura hardware e software di supporto composta da:
 - ✓ Hardware dedicato
 - ✓ Infrastrutture di rete
 - ✓ DBMS
 - ✓ Software di back-end
 - ✓ Software di front-end
- Il ruolo chiave di una piattaforma di business intelligence è la *trasformazione dei dati aziendali in informazioni* fruibili a diversi livelli di dettaglio

9

Dai dati alle informazioni

- L'informazione è un bene a valore crescente, necessario per pianificare e controllare le attività aziendali con efficacia
- Essa costituisce la materia prima che viene trasformata dai sistemi informativi, come i semilavorati vengono trasformati dai sistemi di produzione

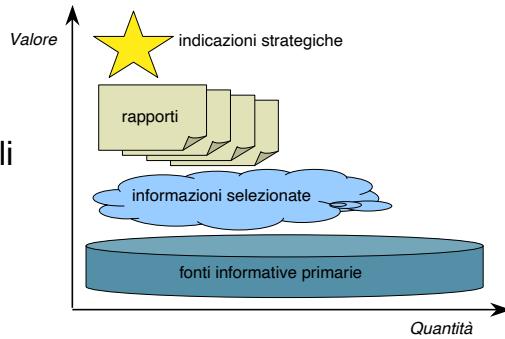
~~dati = informazione~~

- Spesso la disponibilità di troppi dati rende arduo, se non impossibile, estrarre le informazioni veramente importanti

10

Dai dati alle informazioni

- Per ogni azienda è fondamentale poter disporre in maniera rapida e completa delle informazioni necessarie al processo decisionale: le indicazioni strategiche sono estrapolate principalmente dalla mole dei dati operazionali contenuti nei database aziendali, attraverso un procedimento di selezione e sintesi progressiva



11

Il nuovo ruolo dell'informatica

- L'aumento esponenziale del volume dei dati operazionali ha reso il calcolatore l'unico supporto adatto al processo decisionale
- L'utilizzo massiccio di tecniche di analisi dei dati aziendali ha reso il sistema informativo un elemento strategico per la realizzazione del business
- Il ruolo dell'informatica è passato da passivo strumento per la registrazione delle operazioni a fattore decisivo per la individuazione di elementi critici dell'organizzazione e di potenziali aree di business

12

I sistemi di supporto alle decisioni

- Negli anni '80 nascono i sistemi di supporto alle decisioni (*decision support system*):

l'insieme delle tecniche e degli strumenti informatici atti a estrapolare informazioni da un insieme di dati memorizzati su supporti elettronici

Ruolo del DSS

Nel Passato	Nel Futuro
Descrivere il passato	Anticipare il futuro
Ridurre i costi	Aumentare i profitti
Descrivere i problemi	Suggerire i cambiamenti da apportare

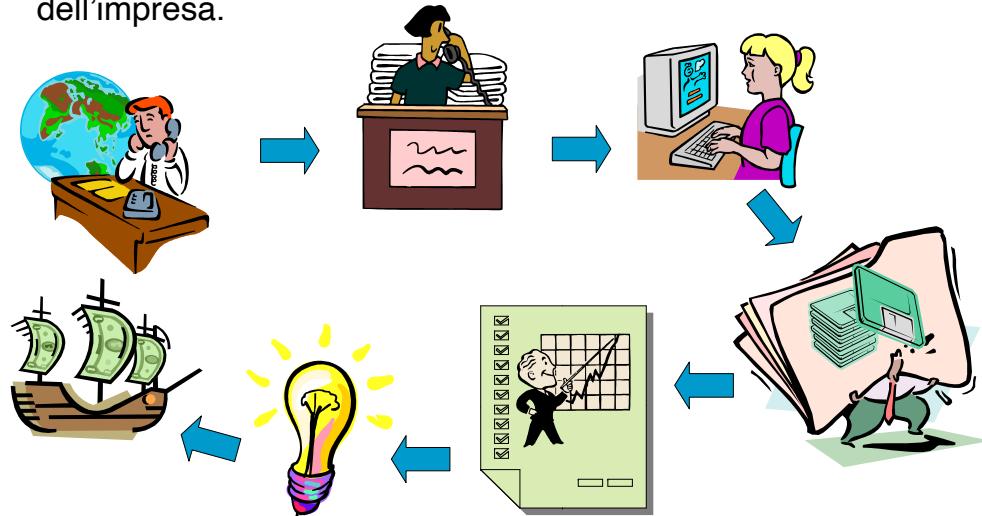
13

Introduzione al Data Warehousing

Prof. Stefano Rizzi

Uno scenario tipico...

- .. è quello di una grande azienda, con numerose filiali, i cui dirigenti desiderano quantificare e valutare il contributo dato da ciascuna di esse al rendimento commerciale globale dell'impresa.



15

Uno scenario tipico...

- .. è quello di una grande azienda, con numerose filiali, i cui dirigenti desiderano quantificare e valutare il contributo dato da ciascuna di esse al rendimento commerciale globale dell'impresa.



16

OLTP e OLAP

- Mescolare interrogazioni “analitiche” e “transazionali” di routine porta a inevitabili rallentamenti che rendono insoddisfatti gli utenti di entrambe le categorie.



separare l'elaborazione di tipo analitico (**OLAP**, On-Line Analytical Processing) da quella legata alle transazioni (**OLTP**, On-Line Transactional Processing), costruendo un nuovo raccoglitrice di informazioni che integri i dati provenienti da sorgenti di varia natura, li organizzi in una forma appropriata e li renda disponibili per scopi di analisi e valutazione finalizzate alla pianificazione e al processo decisionale

17

Alcune aree di utilità

- **Commercio** (analisi delle vendite e dei reclami, controllo di spedizioni e inventari, cura del rapporto con i clienti)
- **Manifattura** (controllo dei costi di produzione, supporto fornitori e ordini)
- **Servizi finanziari** (analisi del rischio e delle carte di credito, rivelazione di frodi)
- **Trasporti** (gestione parco mezzi)
- **Telecomunicazioni** (analisi del flusso delle chiamate e del profilo dei clienti)
- **Sanità** (analisi di ricoveri e dimissioni, contabilità per centri di costo)
-

18

Data Warehousing:

- Una collezione di metodi, tecnologie e strumenti di ausilio al *knowledge worker* (dirigente, amministratore, gestore, analista) per condurre analisi dei dati finalizzate all'attuazione di processi decisionali e al miglioramento del patrimonio informativo.

19

Le lamentele

- ⇒ *abbiamo montagne di dati ma non possiamo accedervi!*
- ⇒ *come è possibile che persone che svolgono lo stesso ruolo presentino risultati sostanzialmente diversi?*
- ⇒ *vogliamo selezionare, raggruppare e manipolare i dati in ogni modo possibile!*
- ⇒ *mostratemi solo ciò che è importante!*
- ⇒ *tutti sanno che alcuni dati non sono corretti!*

R. Kimball, The Data Warehouse Toolkit



20



Caratteristiche del processo di warehousing

- **accessibilità** a utenti con conoscenze limitate di informatica e strutture dati;
- **integrazione dei dati** sulla base di un modello standard dell'impresa;
- **flessibilità di interrogazione** per trarre il massimo vantaggio dal patrimonio informativo esistente;
- **sintesi** per permettere analisi mirate ed efficaci;
- **rappresentazione multidimensionale** per offrire all'utente una visione intuitiva ed efficacemente manipolabile delle informazioni;
- **correttezza e completezza** dei dati integrati.

21

Il Data Warehouse

- Al centro del processo, il data warehouse è un contenitore di dati che si fa garante dei requisiti esposti.

➤ *Un Data Warehouse è una collezione di dati di supporto per il processo decisionale che presenta le seguenti caratteristiche:*

- ✓ *è orientata ai soggetti di interesse;*
- ✓ *è integrata e consistente;*
- ✓ *è rappresentativa dell'evoluzione temporale;*
- ✓ *non volatile.*

22

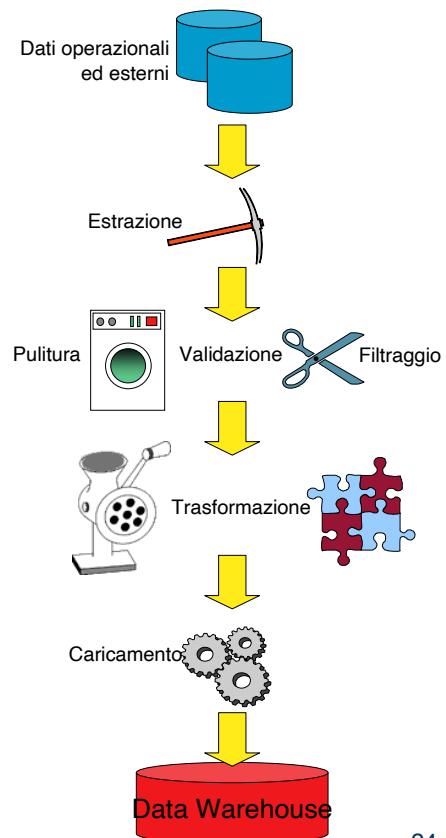
...orientato ai soggetti



23

...integrato e consistente

Il DW si appoggia a più fonti di dati eterogenee: dati estratti dall'ambiente di produzione, e quindi originariamente archiviati in basi di dati aziendali, o addirittura provenienti da sistemi informativi esterni all'azienda. Di tutti questi dati il DW restituisce una visione unificata.



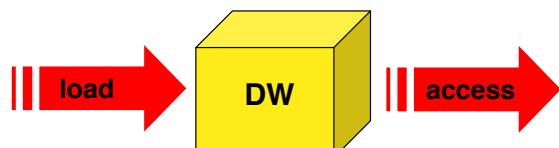
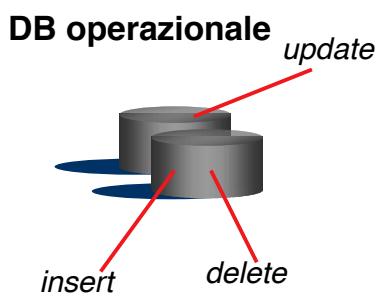
24

...rappresentativo dell'evoluzione temporale



25

...non volatile



Grandi volumi di dati:
da 20 GB a qualche TB
in pochi anni

- ✓ in un DW non sono necessarie tecniche sofisticate di gestione delle transazioni come invece richiesto da un sistema operazionale
- ✓ problemi chiave diventano il query-throughput e la resilienza

26



Le interrogazioni

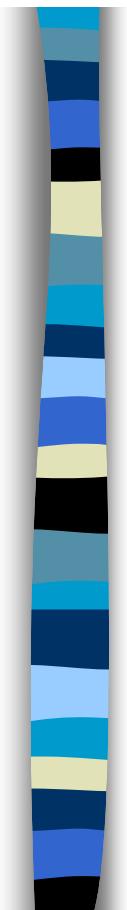
■ OLTP:

- ✓ Le interrogazioni eseguono transazioni che leggono e scrivono un ridotto numero di record da diverse tabelle legate da semplici relazioni.
- ✓ Il nucleo sostanziale del carico di lavoro è “congelato” all’interno dei programmi applicativi.

■ OLAP:

- ✓ Le interrogazioni effettuano un’analisi dinamica e multidimensionale che richiede la scansione di un’enorme quantità di record per calcolare un insieme di dati numerici di sintesi che quantifichino le prestazioni dell’azienda.
- ✓ L’interattività è una caratteristica irrinunciabile delle sessioni di analisi e fa sì che il carico di lavoro effettivo vari continuamente nel tempo.

27



Riassumendo:

	<i>Database operazionali</i>	<i>Data warehouse</i>
utenti	migliaia	centinaia
carico di lavoro	transazioni predefinite	interrogazioni di analisi <i>ad hoc</i>
accesso	a centinaia di record, in lettura e scrittura	a milioni di record, per lo più in lettura
scopo	dipende dall’applicazione	supporto alle decisioni
dati	elementari, sia numerici sia alfanumerici	di sintesi, prevalentemente numerici
integrazione dei dati	per applicazione	per soggetto
qualità	in termini di integrità	in termini di consistenza
copertura temporale	solo dati correnti	dati correnti e storici
aggiornamenti	continui	periodici
modello	normalizzato	denormalizzato, multidimensionale
ottimizzazione	per accessi OLTP su una frazione del database	per accessi OLAP su gran parte del database
sviluppo	a cascata	iterativo

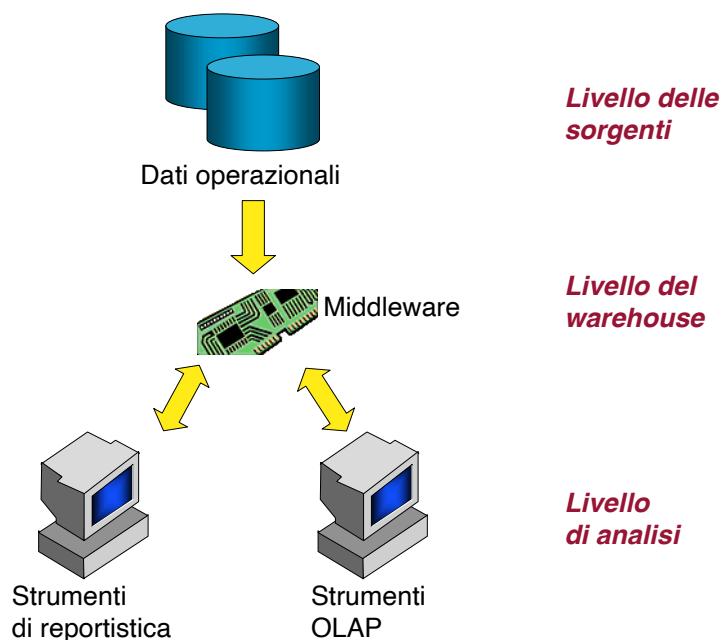
28

Architetture: requisiti

- ✓ **Separazione:** l'elaborazione analitica e quella transazionale devono essere mantenute il più possibile separate.
- ✓ **Scalabilità:** l'architettura hardware e software deve poter essere facilmente ridimensionata a fronte della crescita nel tempo dei volumi di dati da gestire ed elaborare e del numero di utenti da soddisfare.
- ✓ **Estendibilità:** deve essere possibile accogliere nuove applicazioni e tecnologie senza riprogettare integralmente il sistema.
- ✓ **Sicurezza:** il controllo sugli accessi è essenziale a causa della natura strategica dei dati memorizzati.
- ✓ **Amministrabilità:** la complessità dell'attività di amministrazione non deve risultare eccessiva.

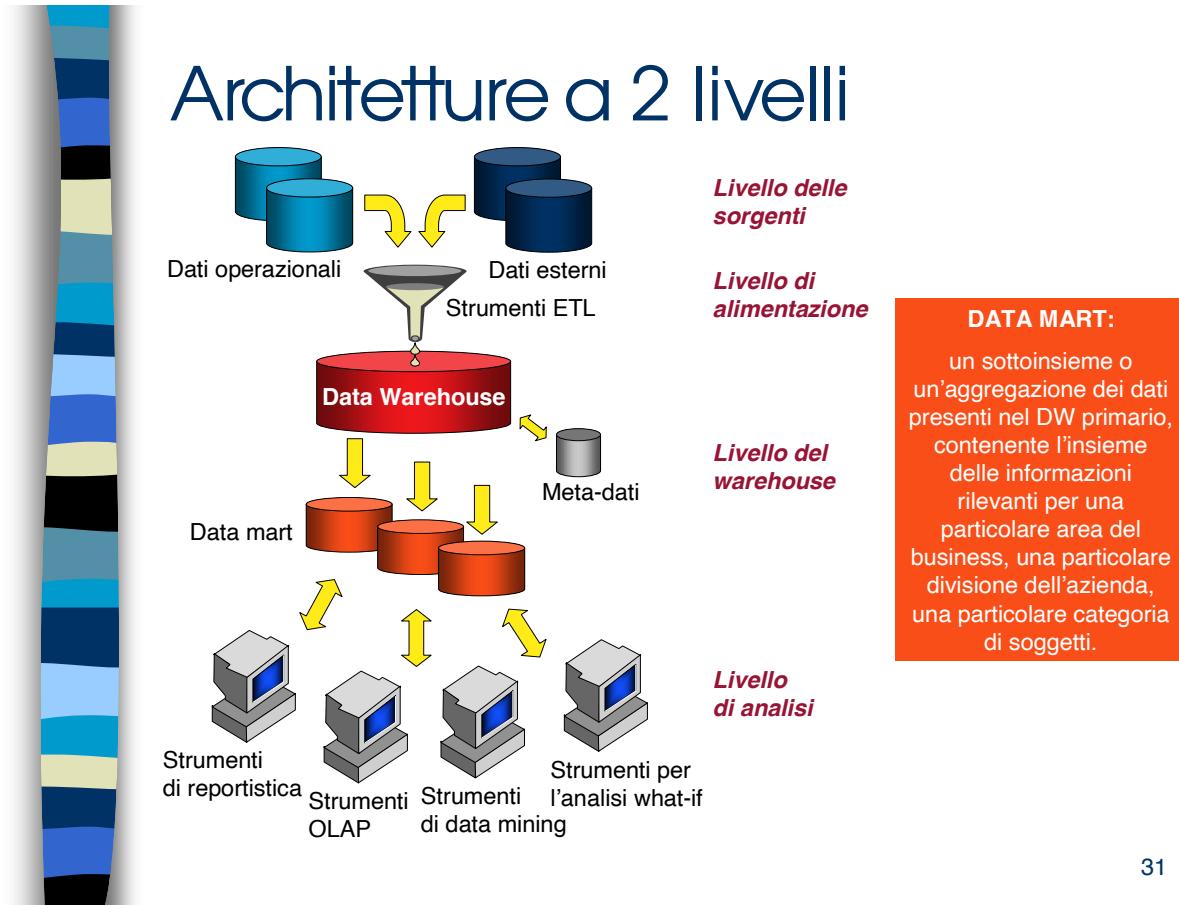
29

Architetture a 1 livello



30

Architetture a 2 livelli



31

Architetture a 2 livelli

- I data mart alimentati dal DW primario sono detti **dipendenti**. Per i sistemi collocati all'interno di realtà aziendali medio-grandi essi sono utili:
 - ✓ come blocchi costruttivi durante la realizzazione incrementale del DW;
 - ✓ in quanto delineano i contorni delle informazioni necessarie a un particolare tipo di utenti per le loro interrogazioni;
 - ✓ poiché, essendo di dimensioni inferiori al DW primario, permettono di raggiungere prestazioni migliori
- In alcuni contesti si preferisce adottare data mart alimentati direttamente dalle sorgenti, detti **indipendenti**
 - ✓ L'assenza di un DW primario snellisce le fasi progettuali, ma determina uno schema complesso di accessi ai dati e ingenera il rischio di inconsistenze tra i data mart

32

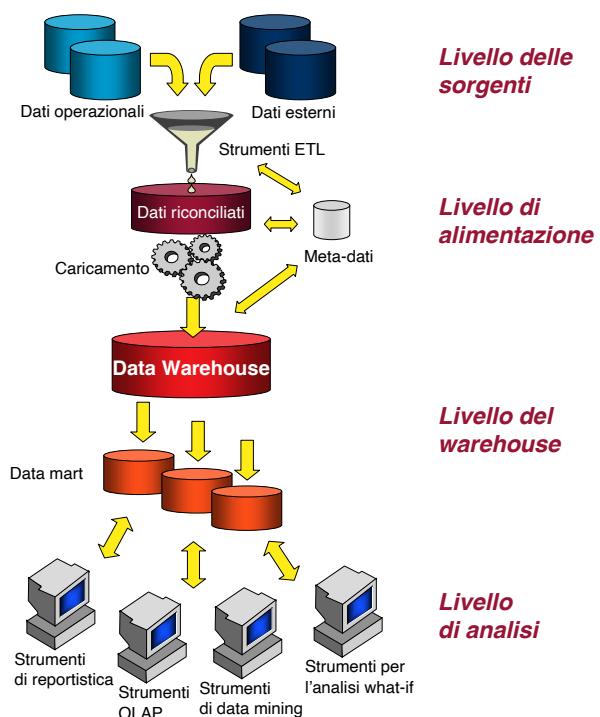
Architetture a 2 livelli

■ Vantaggi:

- ✓ A livello del warehouse è continuamente disponibile informazione di buona qualità anche quando, per motivi tecnici oppure organizzativi, è temporaneamente precluso l'accesso alle sorgenti
- ✓ L'interrogazione analitica effettuata sul DW non interferisce con la gestione delle transazioni a livello operazionale, la cui affidabilità è essenziale per il funzionamento dell'azienda
- ✓ L'organizzazione logica del DW è basata sul modello multidimensionale, mentre le sorgenti offrono in genere modelli relazionali o semi-strutturati
- ✓ C'è una discordanza temporale e di granularità tra sistemi OLTP, che trattano dati correnti e al massimo livello di dettaglio, e sistemi OLAP che operano su dati storici e di sintesi
- ✓ A livello del warehouse è possibile impiegare tecniche specifiche per ottimizzare le prestazioni per applicazioni di analisi e reportistica

33

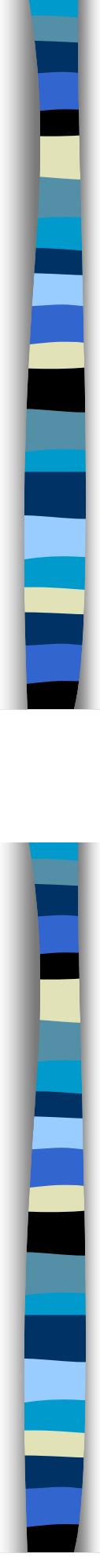
Architetture a 3 livelli



DATI RICONCILIATI:

dati operazionali
ottenuti a valle del
processo di
integrazione e ripulitura
dei dati sorgente: quindi
dati integrati,
consistenti, corretti,
volatili, correnti e
dettagliati

34



Architetture a 3 livelli

- Il vantaggio principale del livello dei dati riconciliati è che esso crea un modello di dati comune e di riferimento per l'intera azienda, introducendo al contempo una separazione netta tra le problematiche legate all'estrazione e integrazione dei dati dalle sorgenti e quelle inerenti l'alimentazione del DW
- D'altro canto, i dati riconciliati introducono un'ulteriore ridondanza rispetto ai dati operazionali sorgente

35

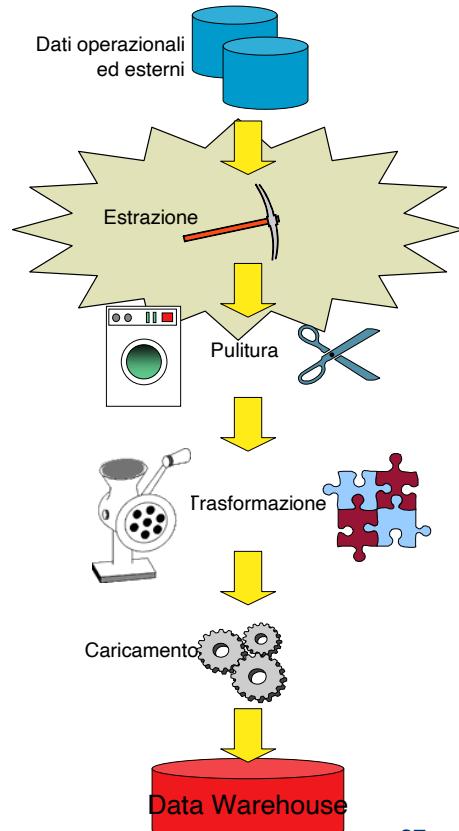
ETL

- Il ruolo degli strumenti di *Extraction, Transformation and Loading* è quello di alimentare una sorgente dati singola, dettagliata, esauriente e di alta qualità che possa a sua volta alimentare il DW (*riconciliazione*)
- Durante il processo di alimentazione del DW, la riconciliazione avviene in due occasioni: quando il DW viene popolato per la prima volta, e periodicamente quando il DW viene aggiornato.
 - ✓ estrazione
 - ✓ pulitura
 - ✓ trasformazione
 - ✓ caricamento

36

Estrazione

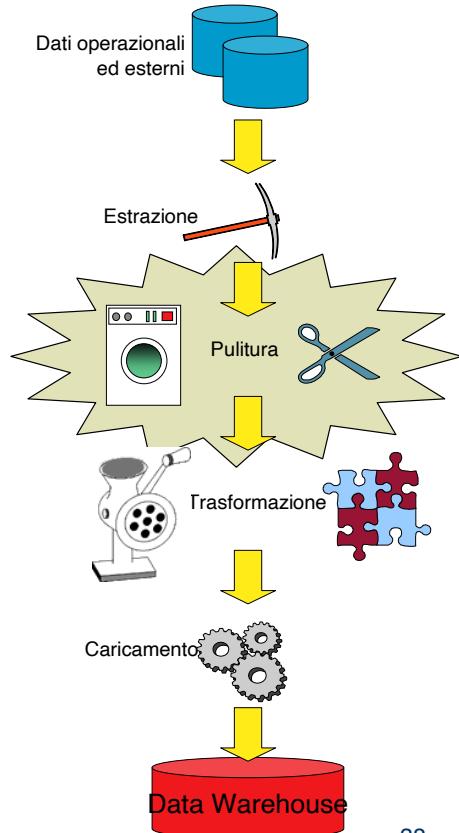
- I dati rilevanti vengono estratti dalle sorgenti.
 - ✓ L'estrazione **statica** viene effettuata quando il DW deve essere popolato per la prima volta e consiste concettualmente in una fotografia dei dati operazionali.
 - ✓ L'estrazione **incrementale** viene usata per l'aggiornamento periodico del DW, e cattura solamente i cambiamenti avvenuti nelle sorgenti dall'ultima estrazione
 - basata sul log mantenuto dal DBMS operazionale
 - basata su time-stamp
 - guidata dalle sorgenti
- La scelta dei dati da estrarre avviene principalmente in base alla loro qualità.



37

Pulitura

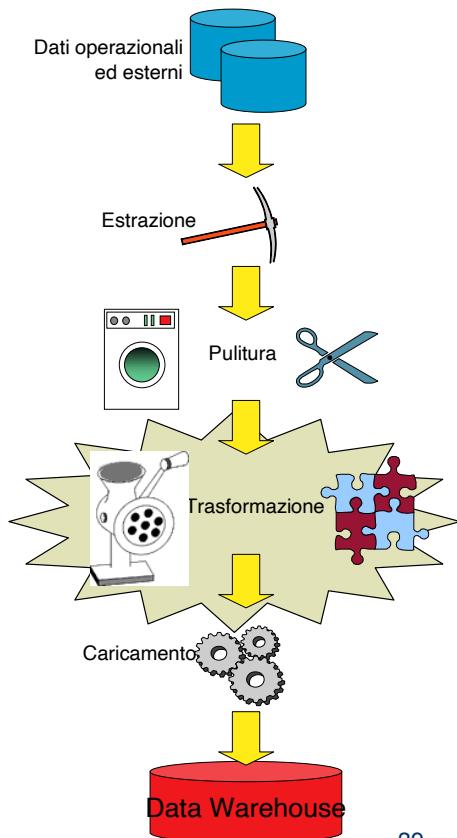
- Si incarica di migliorare la qualità dei dati delle sorgenti
 - ✓ dati duplicati
 - ✓ inconsistenza tra valori logicamente associati
 - ✓ dati mancanti
 - ✓ uso non previsto di un campo
 - ✓ valori impossibili o errati
 - ✓ valori inconsistenti per la stessa entità dovuti a differenti convenzioni
 - ✓ valori inconsistenti per la stessa entità dovuti a errori di battitura



38

Trasformazione

- Converte i dati dal formato operazionale sorgente a quello del DW. La corrispondenza con il livello sorgente è complicata dalla presenza di fonti distinte eterogenee, che richiede una complessa fase di integrazione
 - ✓ presenza di testi liberi che nascondono informazioni importanti
 - ✓ utilizzo di formati differenti per lo stesso dato
- Per l'alimentazione dei dati riconciliati:
 - ✓ conversione e normalizzazione (operano a livello di formato di memorizzazione e di unità di misura per uniformare i dati)
 - ✓ matching (stabilisce corrispondenze tra campi equivalenti in sorgenti diverse)
 - ✓ selezione (riduce il numero di campi e di record rispetto alle sorgenti)
- Per l'alimentazione del DW:
 - ✓ la normalizzazione è sostituita dalla denormalizzazione
 - ✓ si introduce l'aggregazione, che realizza le opportune sintesi dei dati



39

Pulitura e trasformazione

Carlo Bianchi
P.zza Grande 12
50126 Bologna (I)

Normalizzazione

nome:	
cognome:	
indirizzo:	
CAP:	
città:	
nazione:	

Carlo
Bianchi
P.zza Grande 12
50126
Bologna
I

nome:	Carlo
cognome:	Bianchi
indirizzo:	Piazza Grande 12
CAP:	50126
città:	Bologna
nazione:	Italia

Standardizzazione

Correzione

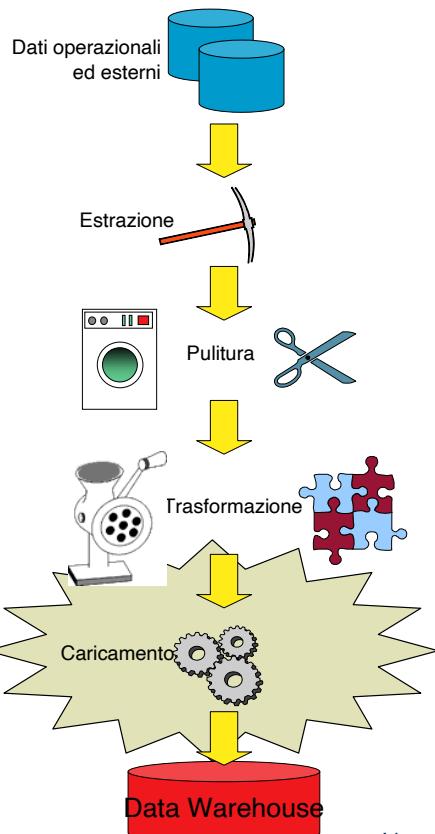
nome:	Carlo
cognome:	Bianchi
indirizzo:	Piazza Grande 12
CAP:	40126
città:	Bologna
nazione:	Italia

40

Caricamento

■ Il caricamento dei dati nel DW

- ✓ **Refresh:** i dati del DW vengono riscritti integralmente, sostituendo quelli precedenti (tecnica normalmente utilizzata solo per popolare inizialmente il DW)
- ✓ **Update:** i soli cambiamenti occorsi nei dati sorgente vengono aggiunti nel DW (tecnica normalmente utilizzata per l'aggiornamento periodico del DW)



41

Verso il modello multidimensionale

“Che incassi sono stati registrati l’anno passato per ciascuna regione e ciascuna categoria di prodotto?”

“Che correlazione esiste tra l’andamento dei titoli azionari dei produttori di PC e i profitti trimestrali lungo gli ultimi 5 anni?”

“Quali sono gli ordini che massimizzano gli incassi?”

“Quale di due nuove terapie risulterà in una diminuzione della durata media di un ricovero?”

“Che rapporto c’è tra i profitti realizzati con spedizioni di meno di 10 elementi e quelli realizzati con spedizioni di più di 10 elementi?”

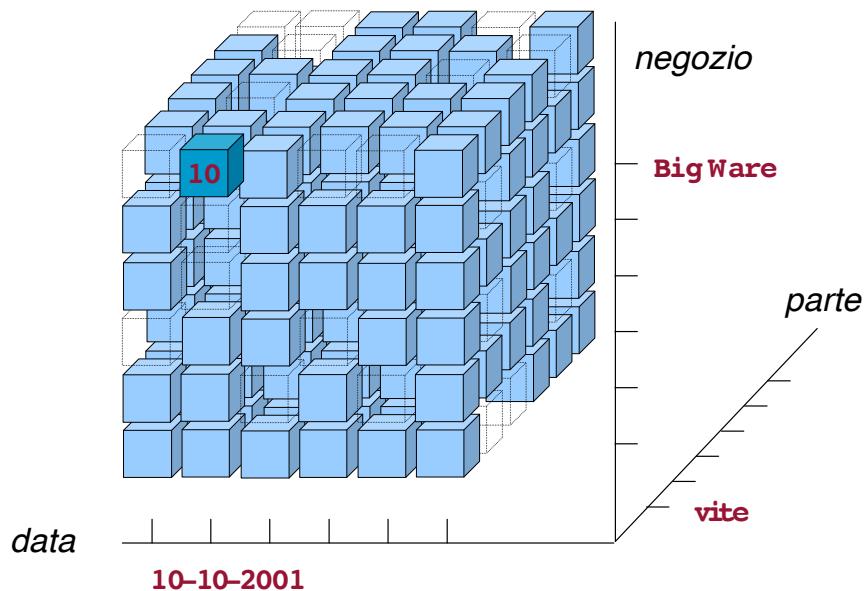
42

Il modello multidimensionale

- È il fondamento per la rappresentazione e l'interrogazione dei dati nei data warehouse.
- I *fatti* di interesse sono rappresentati in *cubi* in cui:
 - ✓ ogni cella contiene *misure* numeriche che quantificano il fatto da diversi punti di vista;
 - ✓ ogni asse rappresenta una *dimensione* di interesse per l'analisi;
 - ✓ ogni dimensione può essere la radice di una *gerarchia* di attributi usati per aggregare i dati memorizzati nei cubi base.

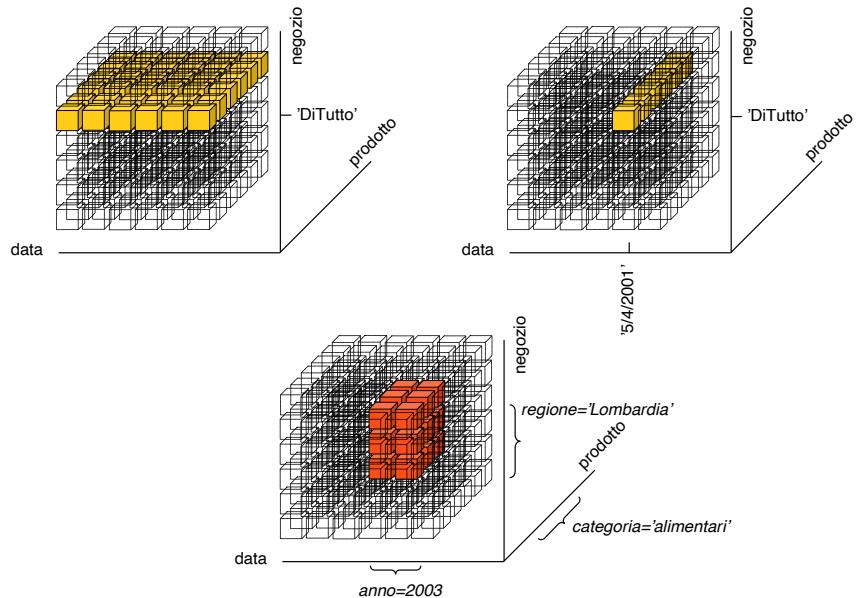
43

Il cubo delle vendite



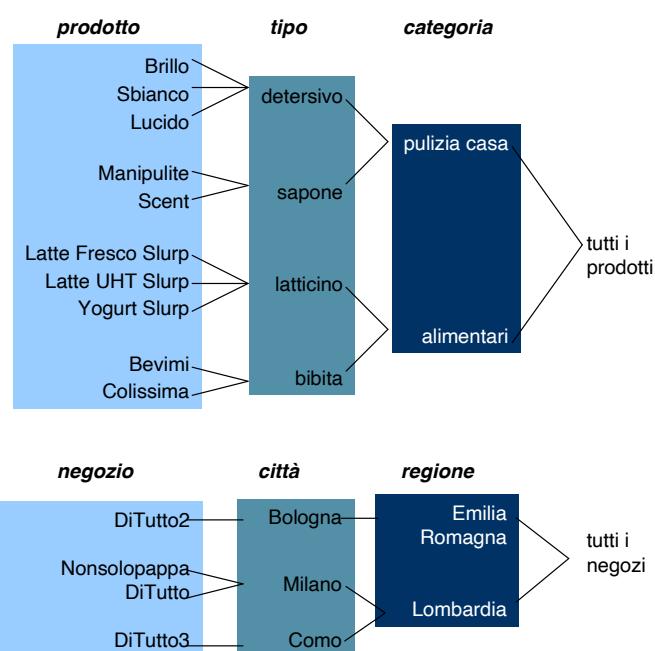
44

Slicing and dicing



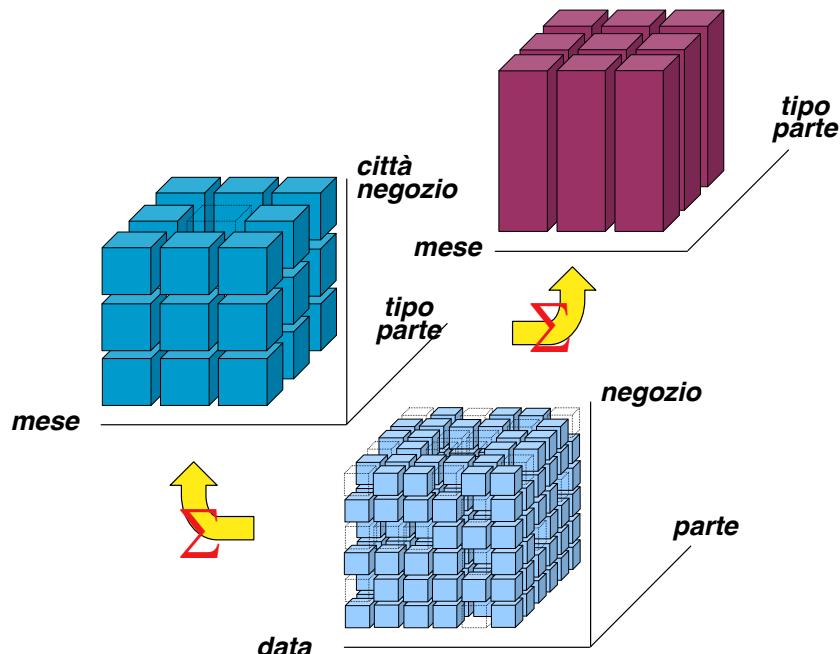
45

Le gerarchie

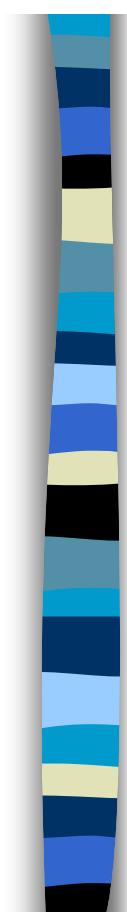


46

Aggregazione



47

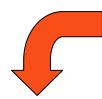


Aggregazione

	DiTutto	DiTutto2	Nonsolopappa
1/1/2000	—	—	—
2/1/2000	10	15	5
3/1/2000	20	—	5
.....
1/1/2001	—	—	—
2/1/2001	15	10	20
3/1/2001	20	20	25
.....
1/1/2002	—	—	—
2/1/2002	20	8	25
3/1/2002	20	12	20
.....



	DiTutto	DiTutto2	Nonsolopappa
Gennaio 2000	200	180	150
Febbraio 2000	180	150	120
Marzo 2000	220	180	160
.....
Gennaio 2001	350	220	200
Febbraio 2001	300	200	250
Marzo 2001	310	180	300
.....
Gennaio 2002	380	200	220
Febbraio 2002	310	200	250
Marzo 2002	300	160	280
.....



	DiTutto	DiTutto2	Nonsolopappa
2000	2400	2000	1600
2001	3200	2300	3000
2002	3400	2200	3200

<i>Totali:</i>	DiTutto	DiTutto2	Nonsolopappa
	9000	6500	7800

48

Tecniche di analisi dei dati

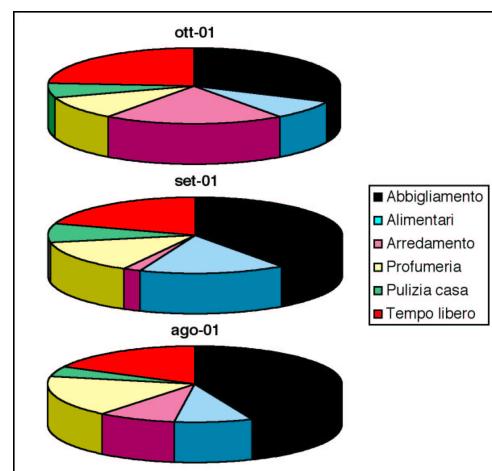
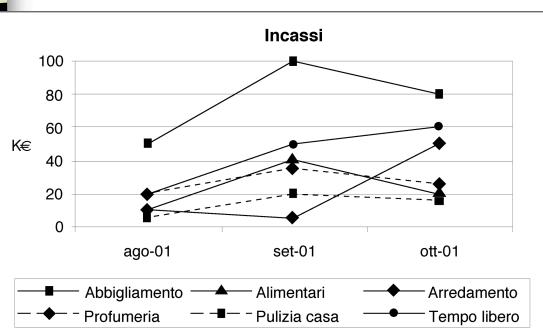
- Una volta che i dati sono stati ripuliti, integrati e trasformati, occorre capire come trarne il massimo vantaggio informativo
- Esistono in sostanza tre approcci differenti, supportati da altrettante categorie di strumenti, all'interrogazione di un DW da parte degli utenti finali:
 - ✓ *reportistica*: non richiede conoscenze informatiche
 - ✓ *OLAP*: richiede all'utente di ragionare in modo multidimensionale e di conoscere l'interfaccia dello strumento grafico utilizzato
 - ✓ *data mining*: richiede all'utente la conoscenza dei principi che stanno alla base degli strumenti utilizzati

49

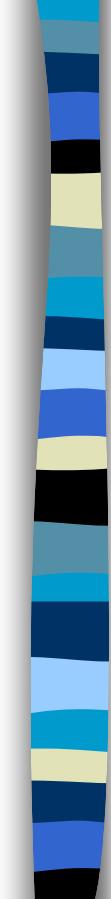
Reportistica

orientato agli utenti che hanno necessità di accedere, a intervalli di tempo predefiniti, a informazioni strutturate in modo pressoché invariabile

incassi (K€)	Ottobre 2001	Settembre 2001	Agosto 2001
Abbigliamento	80	100	50
Alimentari	20	40	10
Arredamento	50	5	10
Profumeria	25	35	20
Pulizia casa	15	20	5
Tempo libero	60	50	20



50



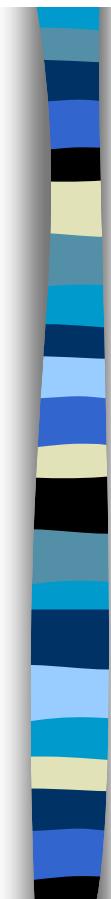
OLAP

- È la principale modalità di fruizione delle informazioni contenute in un DW
- Consente, a utenti le cui necessità di analisi non siano facilmente identificabili a priori, di analizzare ed esplorare interattivamente i dati sulla base del modello multidimensionale
- Mentre gli utenti degli strumenti di reportistica svolgono un ruolo essenzialmente passivo, gli utenti OLAP sono in grado di costruire attivamente una sessione di analisi complessa in cui ciascun passo effettuato è conseguenza dei risultati ottenuti al passo precedente
 - ✓ estemporaneità delle sessioni di lavoro
 - ✓ richiesta approfondita conoscenza dei dati
 - ✓ complessità delle interrogazioni formulabili
 - ✓ orientamento verso utenti non esperti di informatica



interfaccia flessibile, facile
da usare ed efficace

51

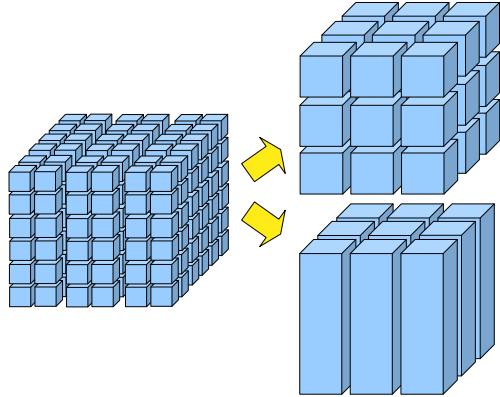
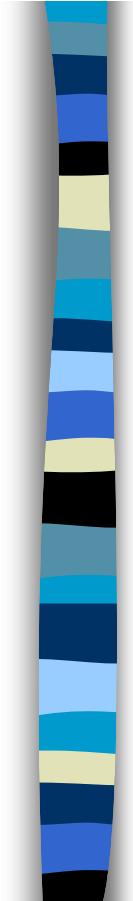


OLAP: sessione

- Una sessione OLAP consiste in un *percorso di navigazione* che riflette il procedimento di analisi di uno o più fatti di interesse sotto diversi aspetti e a diversi livelli di dettaglio. Questo percorso si concretizza in una sequenza di interrogazioni spesso formulate non direttamente, ma per differenza rispetto all'interrogazione precedente
- Ogni passo della sessione di analisi è scandito dall'applicazione di un **operatore OLAP** che trasforma l'ultima interrogazione formulata in una nuova interrogazione
- Il risultato delle interrogazioni è di tipo multidimensionale; gli strumenti OLAP rappresentano tipicamente i dati in modo tabellare evidenziando le diverse dimensioni mediante intestazioni multiple, colori ecc.

52

OLAP: operatori

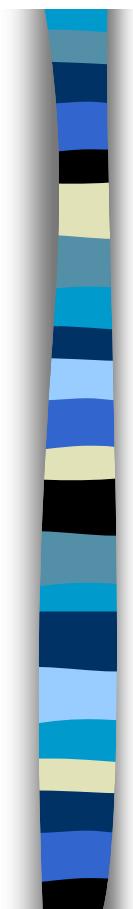


roll-up

Aumenta
l'aggregazione dei dati
eliminando un livello di
dettaglio da una
gerarchia

53

OLAP: operatori



Incassi Mensili Anni 97 e 98
per Regione del Cliente

Metrics	Dollar Sales	North-East	Mid-Atlantic	South-East	Central	South	North-West	South-West	England	France	Germany	Canada
Customer Region												
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	\$ 357	
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	\$ 652	\$ 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	\$ 2.291	\$ 1.840	\$ 600	\$ 656	\$ 1.300	\$ 718	\$ 1.210	\$ 427	\$ 220	\$ 520	\$ 65	
Nov 97	\$ 39	\$ 1.602	\$ 1.082	\$ 1.187	\$ 842	\$ 759	\$ 745	\$ 232	\$ 101	\$ 1.037	\$ 37	
Dec 97	\$ 381	\$ 1.598	\$ 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	\$ 2.518	\$ 702	\$ 1.123	\$ 1.336	\$ 1.227	\$ 3.887	\$ 545	\$ 268	\$ 277	\$ 282		
Mar 98	\$ 2.459	\$ 1.523	\$ 1.178	\$ 4.708	\$ 1.420	\$ 3.514	\$ 1.948	\$ 1.705	\$ 276	\$ 1.168	\$ 6	
Apr 98	\$ 407	\$ 841	\$ 524	\$ 712	\$ 133	\$ 2.486	\$ 49	\$ 390	\$ 1.298	\$ 221	\$ 46	
May 98	\$ 667	\$ 1.721	\$ 440	\$ 148	\$ 80	\$ 1.310	\$ 303	\$ 104	\$ 657	\$ 65		
Jun 98	\$ 699	\$ 1.096	\$ 898	\$ 353	\$ 902	\$ 639	\$ 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	\$ 395	\$ 1.740	\$ 1.943	\$ 1.143	\$ 366	\$ 307	\$ 118	

roll-up

Incassi Trimestrali per
Regione del Cliente

Metrics	Dollar Sales	North-East	Mid-Atlantic	South-East	Central	South	North-West	South-West	England	France	Germany	Canada
Customer Region												
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	\$ 1.676	\$ 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	\$ 846	\$ 639	



54

OLAP: operatori

Incassi Annuali Anni 97 e
98 per Categoria di prodotto e per
ciascuna Regione

Category	Year	Metrics										
		Dollar Sales	Customer Region	North-East	Mid-Atlantic	South-East	Central	South	North-West	South-West	England	France
Electronics	1997	\$ 138		\$ 1.774	\$ 384	\$ 138	\$ 2.346	\$ 2.554	\$ 2.184	\$ 566	\$ 199	\$ 1
	1998	\$ 1.184		\$ 4.529	\$ 1.892	\$ 7.232	\$ 651	\$ 9.488	\$ 476	\$ 2.693	\$ 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	\$ 10
	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	\$ 5.009	\$ 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	\$

roll-up

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

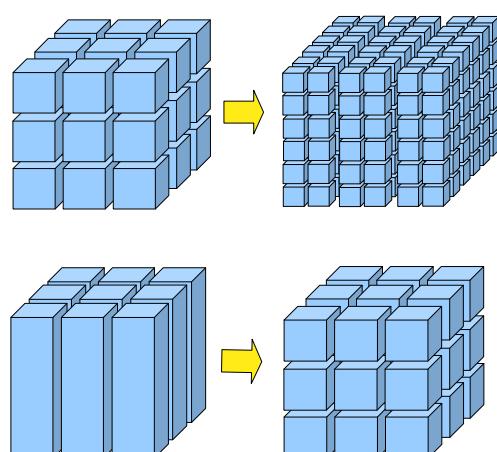
Incassi Annuali per
Categoria di prodotto

55

OLAP: operatori

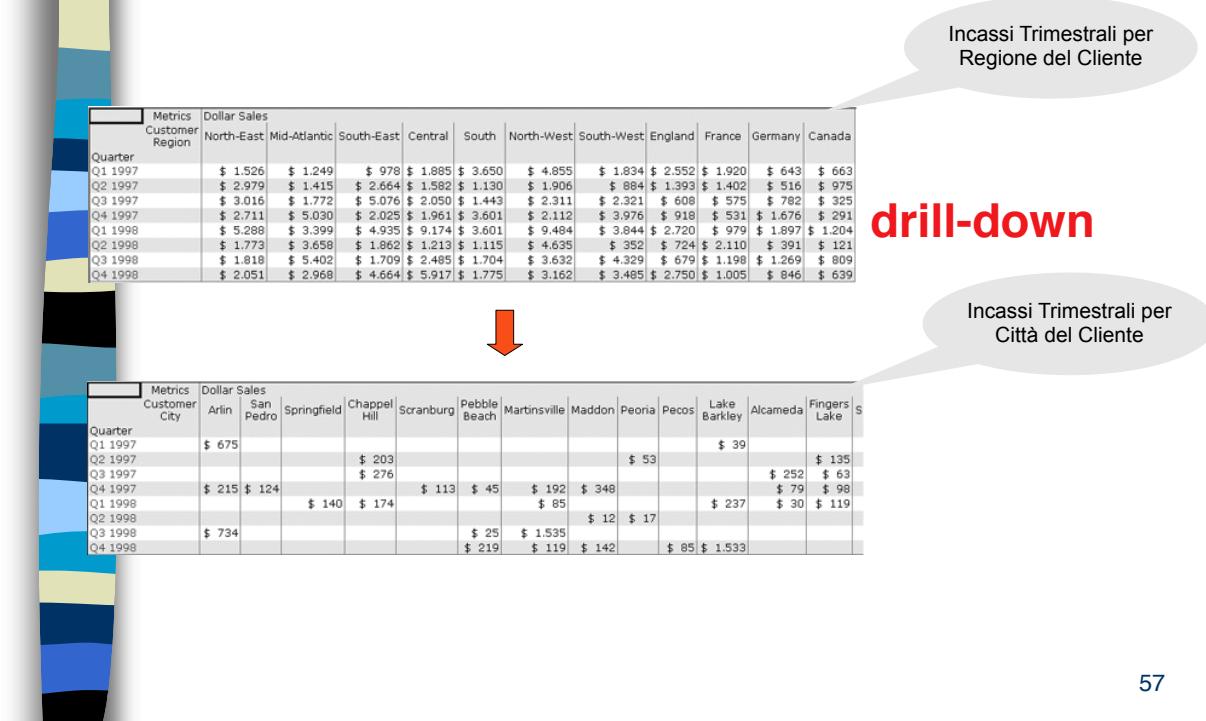
drill-down

Diminuisce
l'aggregazione dei dati
introducendo un ulteriore
livello di dettaglio in una
gerarchia



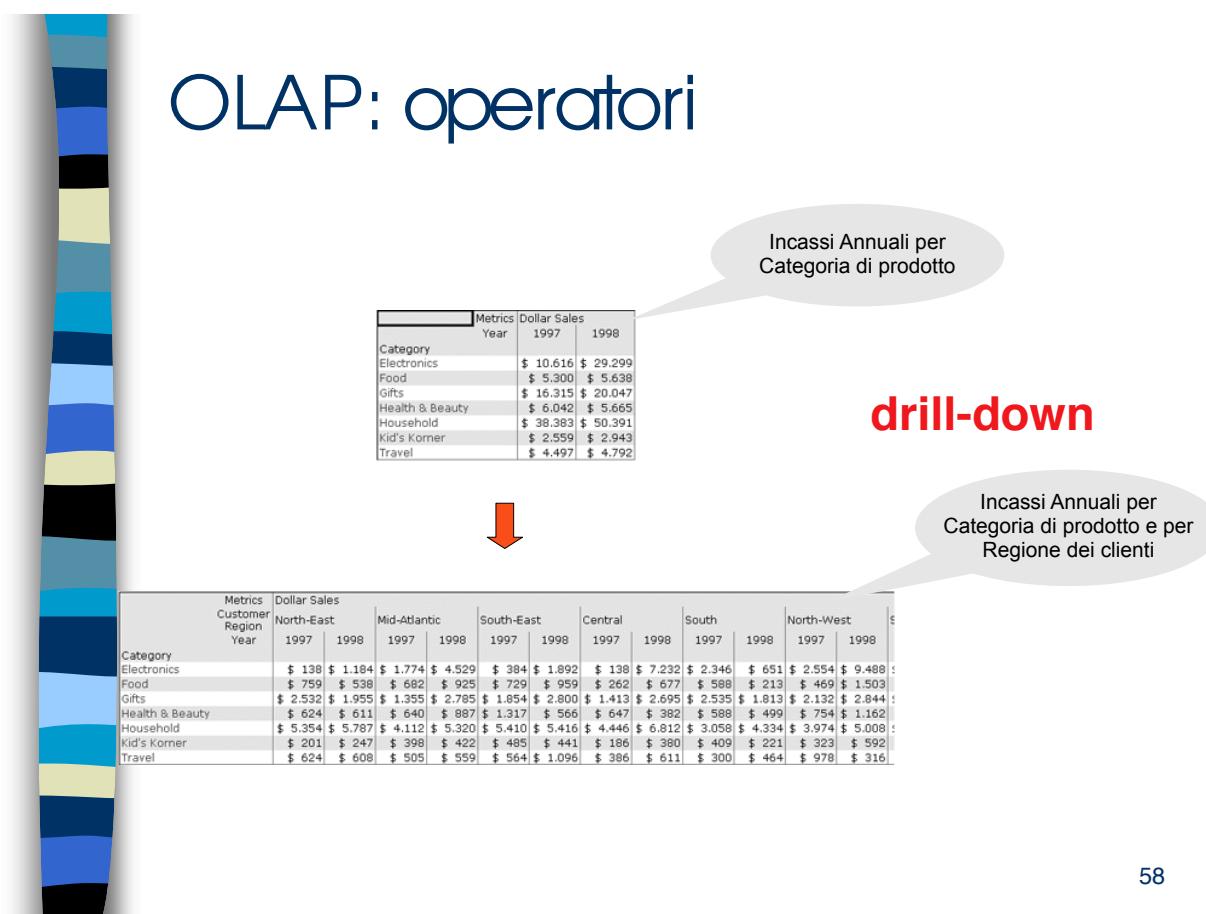
56

OLAP: operatori



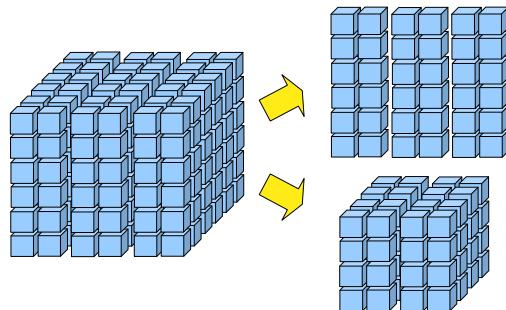
57

OLAP: operatori



58

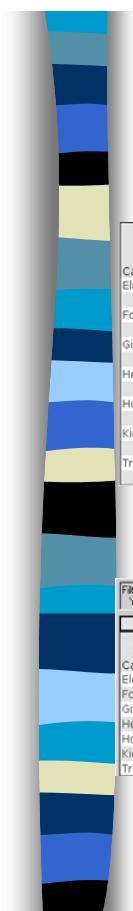
OLAP: operatori



slice-and-dice

Riduce la dimensionalità
del cubo

59



OLAP: operatori

Incassi Annuali Anni 97 e
98 per CATEGORIA di prodotto e per
ciascuna REGIONE

Category	Year	Metrics									
		Dollar Sales	Customer Region	North-East	Mid-Atlantic	South-East	Central	South	North-West	South-West	England
Electronics	1997	\$ 138	\$ 1.774	\$ 384	\$ 138	\$ 2.346	\$ 2.554	\$ 2.184	\$ 566	\$ 199	\$ 7
	1998	\$ 1.184	\$ 4.529	\$ 1.892	\$ 7.232	\$ 651	\$ 9.488	\$ 476	\$ 2.683	\$ 462	\$ 1
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	\$ 392	\$ 499	\$ 1.162	\$ 1.044	\$ 273		
Household	1997	\$ 5.354	\$ 4.112	\$ 5.410	\$ 4.446	\$ 3.058	\$ 3.974	\$ 2.654	\$ 3.545	\$ 1	
	1998	\$ 5.787	\$ 5.320	\$ 5.416	\$ 6.812	\$ 4.334	\$ 5.008	\$ 7.588	\$ 2.139	\$ 1	
Kid's Korner	1997	\$ 201	\$ 398	\$ 485	\$ 186	\$ 409	\$ 323	\$ 396	\$ 105		
	1998	\$ 247	\$ 422	\$ 441	\$ 380	\$ 221	\$ 592	\$ 290	\$ 198	\$ 19	\$ 38
Travel	1997	\$ 624	\$ 505	\$ 564	\$ 386	\$ 300	\$ 978	\$ 416	\$ 48		
	1998	\$ 608	\$ 559	\$ 1.096	\$ 611	\$ 464	\$ 316	\$ 573	\$ 257	\$ 198	\$ 55

slice-and-dice

Incassi Annuali Anno 98
per CATEGORIA di prodotto e per
ciascuna REGIONE

Category	Year	Metrics										
		Dollar Sales	Customer Region	North-East	Mid-Atlantic	South-East	Central	South	North-West	South-West	England	France
Electronics	1997	\$ 1.184	\$ 4.529	\$ 1.892	\$ 7.232	\$ 651	\$ 9.488	\$ 476	\$ 2.683	\$ 462	\$ 702	
	1998	\$ 538	\$ 925	\$ 959	\$ 677	\$ 213	\$ 1.503	\$ 261	\$ 165	\$ 175	\$ 100	\$ 686
Food	1997	\$ 759	\$ 682	\$ 729	\$ 262	\$ 588	\$ 469	\$ 807	\$ 156	\$ 615		
	1998	\$ 538	\$ 925	\$ 959	\$ 677	\$ 213	\$ 1.503	\$ 261	\$ 165	\$ 175	\$ 100	\$ 686
Gifts	1997	\$ 2.532	\$ 1.355	\$ 1.854	\$ 1.413	\$ 2.535	\$ 2.132	\$ 1.904	\$ 908	\$ 375		
	1998	\$ 1.955	\$ 2.785	\$ 2.800	\$ 2.695	\$ 1.813	\$ 2.844	\$ 1.778	\$ 1.158	\$ 717		
Health & Beauty	1997	\$ 624	\$ 640	\$ 1.317	\$ 647	\$ 588	\$ 754	\$ 654	\$ 143			
	1998	\$ 611	\$ 887	\$ 566	\$ 392	\$ 499	\$ 1.162	\$ 1.044	\$ 273			
Household	1997	\$ 5.354	\$ 4.112	\$ 5.410	\$ 4.446	\$ 3.058	\$ 3.974	\$ 2.654	\$ 3.545			
	1998	\$ 5.787	\$ 5.320	\$ 5.416	\$ 6.812	\$ 4.334	\$ 5.008	\$ 7.588	\$ 2.139			
Kid's Korner	1997	\$ 201	\$ 398	\$ 485	\$ 186	\$ 409	\$ 323	\$ 396	\$ 105			
	1998	\$ 247	\$ 422	\$ 441	\$ 380	\$ 221	\$ 592	\$ 290	\$ 198	\$ 19	\$ 69	
Travel	1997	\$ 624	\$ 505	\$ 564	\$ 386	\$ 300	\$ 978	\$ 416	\$ 48			
	1998	\$ 608	\$ 559	\$ 1.096	\$ 611	\$ 464	\$ 316	\$ 573	\$ 257	\$ 198	\$ 55	



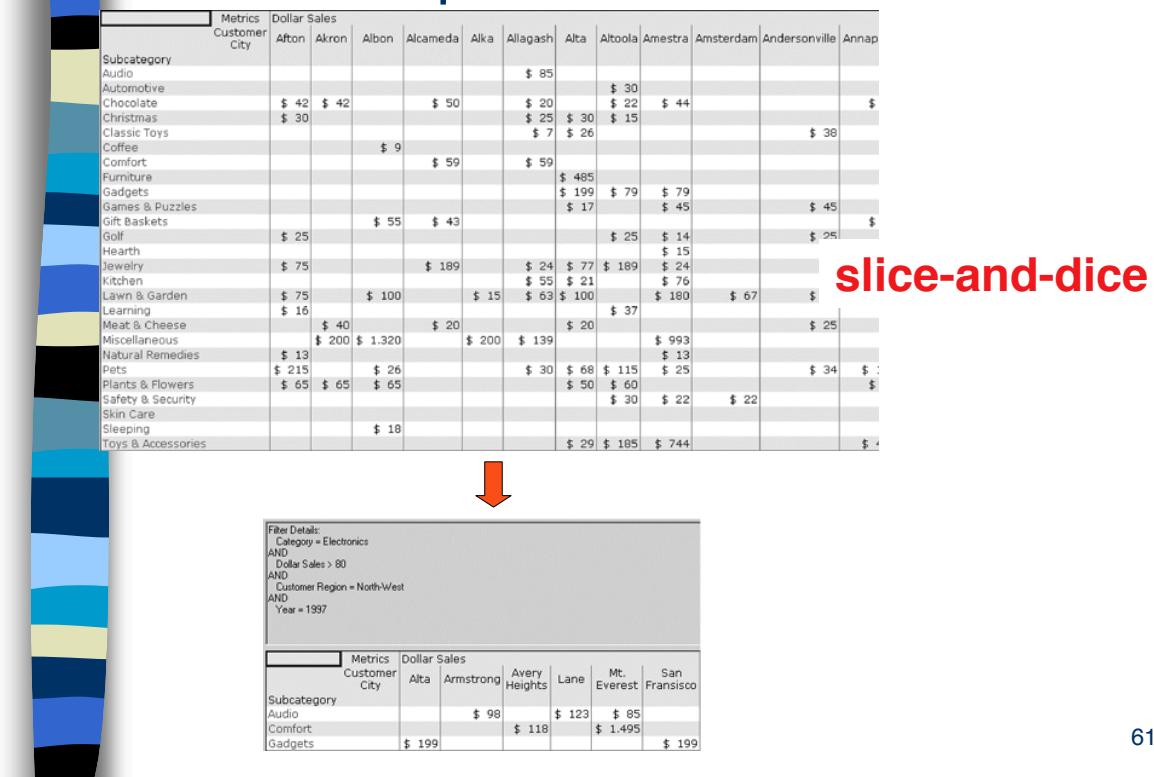
slice-and-dice

Incassi Annuali Anno 98
per CATEGORIA di prodotto e per
ciascuna REGIONE



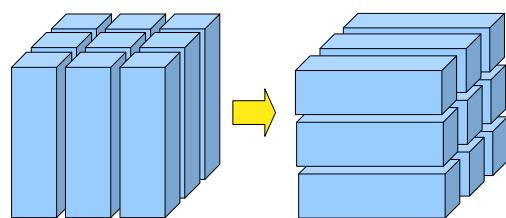
60

OLAP: operatori



61

OLAP: operatori



pivoting

Comporta un
cambiamento nella
modalità di
presentazione

62

OLAP: operatori



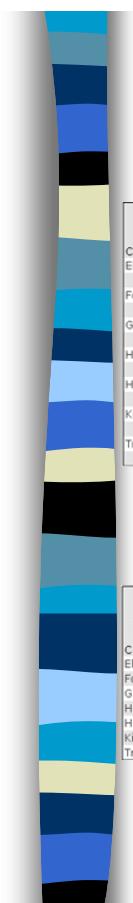
Category	Year	Metrics		Dollar Sales
		1997	1998	
Electronics	1997	\$ 10.616		
	1998	\$ 29.299		
Food	1997	\$ 5.300		
	1998	\$ 5.638		
Gifts	1997	\$ 16.315		
	1998	\$ 20.047		
Health & Beauty	1997	\$ 6.042		
	1998	\$ 5.665		
Household	1997	\$ 38.383		
	1998	\$ 50.391		
Kid's Korner	1997	\$ 2.559		
	1998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

pivoting



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

63



OLAP: operatori

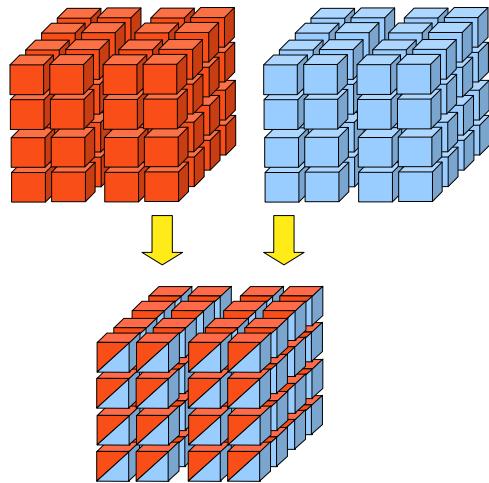
Category	Year	Customer Region	Metrics		Dollar Sales																					
					North-East		Mid-Atlantic		South-East		Central		South		North-West		South-West		England		France		Germany			
			1997	1998	1997	1998	1997	1998	1997	1998	1997	1998	1997	1998	1997	1998	1997	1998	1997	1998	1997	1998	1997	1998		
Electronics	1997		\$ 138	\$ 1.184	\$ 1.774	\$ 4.529	\$ 384	\$ 138	\$ 2.346	\$ 2.554	\$ 2.184	\$ 566	\$ 199	\$ 488	\$ 2.554	\$ 2.184	\$ 566	\$ 199	\$ 488	\$ 2.554	\$ 2.184	\$ 566	\$ 199	\$ 488	\$ 2.554	
	1998		\$ 1.184	\$ 4.529	\$ 4.529	\$ 1.774	\$ 729	\$ 729	\$ 262	\$ 588	\$ 469	\$ 469	\$ 469	\$ 469	\$ 807	\$ 807	\$ 156	\$ 615	\$ 1	\$ 1	\$ 1	\$ 1	\$ 1	\$ 1	\$ 1	\$ 1
Food	1997		\$ 759	\$ 538	\$ 682	\$ 925	\$ 729	\$ 729	\$ 262	\$ 588	\$ 469	\$ 469	\$ 469	\$ 469	\$ 807	\$ 807	\$ 156	\$ 615	\$ 1	\$ 1	\$ 1	\$ 1	\$ 1	\$ 1	\$ 1	\$ 1
	1998		\$ 538	\$ 1.955	\$ 925	\$ 2.785	\$ 959	\$ 959	\$ 677	\$ 213	\$ 1.503	\$ 1.503	\$ 261	\$ 261	\$ 165	\$ 165	\$ 175	\$ 175	\$ 1	\$ 1	\$ 1	\$ 1	\$ 1	\$ 1	\$ 1	\$ 1
Gifts	1997		\$ 2.532	\$ 2.532	\$ 1.355	\$ 1.355	\$ 1.854	\$ 1.854	\$ 1.413	\$ 1.413	\$ 2.535	\$ 2.535	\$ 2.132	\$ 2.132	\$ 1.904	\$ 1.904	\$ 908	\$ 908	\$ 375	\$ 375	\$ 1.0	\$ 1.0	\$ 1.0	\$ 1.0	\$ 1.0	\$ 1.0
	1998		\$ 2.532	\$ 2.532	\$ 1.955	\$ 1.955	\$ 2.785	\$ 2.785	\$ 2.800	\$ 2.800	\$ 2.695	\$ 2.695	\$ 1.813	\$ 1.813	\$ 2.844	\$ 2.844	\$ 1.778	\$ 1.778	\$ 1.158	\$ 1.158	\$ 717	\$ 717	\$ 6	\$ 6	\$ 6	\$ 6
Health & Beauty	1997		\$ 624	\$ 624	\$ 640	\$ 640	\$ 1.317	\$ 1.317	\$ 647	\$ 647	\$ 588	\$ 588	\$ 754	\$ 754	\$ 654	\$ 654	\$ 143	\$ 143	\$ 292	\$ 292	\$ 3	\$ 3	\$ 3	\$ 3	\$ 3	\$ 3
	1998		\$ 611	\$ 611	\$ 887	\$ 887	\$ 566	\$ 566	\$ 382	\$ 382	\$ 499	\$ 499	\$ 1.162	\$ 1.162	\$ 1.044	\$ 1.044	\$ 273	\$ 273	\$ 72	\$ 72	\$ 2	\$ 2	\$ 2	\$ 2	\$ 2	\$ 2
Household	1997		\$ 5.354	\$ 5.354	\$ 4.112	\$ 4.112	\$ 5.410	\$ 5.410	\$ 4.446	\$ 4.446	\$ 3.058	\$ 3.058	\$ 3.974	\$ 3.974	\$ 2.654	\$ 2.654	\$ 3.545	\$ 3.545	\$ 2.875	\$ 2.875	\$ 1.9	\$ 1.9	\$ 1.9	\$ 1.9	\$ 1.9	\$ 1.9
	1998		\$ 5.787	\$ 5.787	\$ 5.320	\$ 5.320	\$ 5.416	\$ 5.416	\$ 6.812	\$ 6.812	\$ 4.334	\$ 4.334	\$ 5.008	\$ 5.008	\$ 7.588	\$ 7.588	\$ 2.139	\$ 2.139	\$ 3.649	\$ 3.649	\$ 2.7	\$ 2.7	\$ 2.7	\$ 2.7	\$ 2.7	\$ 2.7
Kid's Korner	1997		\$ 201	\$ 201	\$ 398	\$ 398	\$ 485	\$ 485	\$ 186	\$ 186	\$ 409	\$ 409	\$ 323	\$ 323	\$ 396	\$ 396	\$ 105	\$ 105	\$ 34	\$ 34	\$ 1	\$ 1	\$ 1	\$ 1	\$ 1	\$ 1
	1998		\$ 247	\$ 247	\$ 422	\$ 422	\$ 441	\$ 441	\$ 380	\$ 380	\$ 221	\$ 221	\$ 592	\$ 592	\$ 290	\$ 290	\$ 198	\$ 198	\$ 19	\$ 19	\$ 1	\$ 1	\$ 1	\$ 1	\$ 1	\$ 1
Travel	1997		\$ 624	\$ 624	\$ 505	\$ 505	\$ 564	\$ 564	\$ 386	\$ 386	\$ 300	\$ 300	\$ 978	\$ 978	\$ 416	\$ 416	\$ 48	\$ 48	\$ 38	\$ 38	\$ 1	\$ 1	\$ 1	\$ 1	\$ 1	\$ 1
	1998		\$ 608	\$ 608	\$ 559	\$ 559	\$ 1.096	\$ 1.096	\$ 611	\$ 611	\$ 464	\$ 464	\$ 316	\$ 316	\$ 573	\$ 573	\$ 257	\$ 257	\$ 198	\$ 198	\$ 1	\$ 1	\$ 1	\$ 1	\$ 1	\$ 1

pivoting



Category	Year	Customer Region	Metrics		Dollar Sales																					
					North-East		Mid-Atlantic		South-East		Central		South		North-West		South-West		England		France		Germany			
			1997	1998	1997	1998	1997	1998	1997	1998	1997	1998	1997	1998	1997	1998	1997	1998	1997	1998	1997	1998	1997	1998		
Electronics	1997		\$ 138	\$ 1.184	\$ 1.774	\$ 4.529	\$ 384	\$ 138	\$ 2.346	\$ 2.554	\$ 2.184	\$ 566	\$ 199	\$ 488	\$ 2.554	\$ 2.184	\$ 566	\$ 199	\$ 488	\$ 2.554	\$ 2.184	\$ 566	\$ 199	\$ 488	\$ 2.554	
	1998		\$ 1.184	\$ 4.529	\$ 4.529	\$ 1.774	\$ 729	\$ 729	\$ 262	\$ 588	\$ 469	\$ 469	\$ 469	\$ 469	\$ 807	\$ 807	\$ 156	\$ 615	\$ 1	\$ 1	\$ 1	\$ 1	\$ 1	\$ 1	\$ 1	\$ 1
Food	1997		\$ 759	\$ 538	\$ 682	\$ 925	\$ 729	\$ 729	\$ 262	\$ 588	\$ 469	\$ 469	\$ 469	\$ 469	\$ 807	\$ 807	\$ 156	\$ 615	\$ 1	\$ 1	\$ 1	\$ 1	\$ 1	\$ 1	\$ 1	\$ 1
	1998		\$ 538	\$ 1.955	\$ 682	\$ 925	\$ 729	\$ 729	\$ 262	\$ 588	\$ 469	\$ 469	\$ 469	\$ 469	\$ 807	\$ 807	\$ 156	\$ 615	\$ 1	\$ 1	\$ 1	\$ 1	\$ 1	\$ 1	\$ 1	\$ 1
Gifts	1997		\$ 2.532	\$ 2.532	\$ 1.355	\$ 1.355	\$ 1.854	\$ 1.854	\$ 1.413	\$ 1.413	\$ 2.695	\$ 2.695	\$ 2.535	\$ 2.535	\$ 1.813	\$ 1.813	\$ 2.132	\$ 2.132	\$ 2.844	\$ 2.844	\$ 1.622	\$ 1.622	\$ 1.622	\$ 1.622	\$ 1.622	\$ 1.622
	1998		\$ 2.532	\$ 2.532	\$ 1.955	\$ 1.955	\$ 2.785	\$ 2.785	\$ 2.800	\$ 2.800	\$ 1.413	\$ 1.413	\$ 2.695	\$ 2.695	\$ 1.813	\$ 1.813	\$ 2.132	\$ 2.132	\$ 2.844	\$ 2.844	\$ 1.622	\$ 1.622	\$ 1.622	\$ 1.622	\$ 1.622	\$ 1.622
Health & Beauty	1997		\$ 624	\$ 624	\$ 640	\$ 640	\$ 887	\$ 887	\$ 566	\$ 566	\$ 647	\$ 647	\$ 382	\$ 382	\$ 588	\$ 588	\$ 499	\$ 499	\$ 754	\$ 754	\$ 1.162	\$ 1.162	\$ 1.162	\$ 1.162	\$ 1.162	\$ 1.162
	1998		\$ 624	\$ 608	\$ 505	\$ 559	\$ 564	\$ 564	\$ 1.096	\$ 1.096	\$ 611	\$ 611	\$ 386	\$ 386	\$ 611	\$ 611	\$ 300	\$ 300	\$ 464	\$ 464	\$ 978	\$ 978	\$ 316	\$ 316	\$ 316	\$ 316
Household	1997		\$ 5.354	\$ 5.787	\$ 4.112	\$ 5.320	\$ 5.410	\$ 5.416	\$ 4.446	\$ 4.446	\$ 6.812	\$ 6.812	\$ 3.058	\$ 3.058	\$ 4.334	\$ 4.334	\$ 3.974	\$ 3.974	\$ 5.008	\$ 5.008	\$ 2.844	\$ 2.844	\$ 2.844	\$ 2.844	\$ 2.844	\$ 2.844
	1998		\$ 5.787	\$ 5.787	\$ 5.320	\$ 5.320	\$ 5.416	\$ 5.416	\$ 4.446	\$ 4.446	\$ 6.812	\$ 6.812	\$ 3.058	\$ 3.058	\$ 4.334	\$ 4.334	\$ 3.974	\$ 3.974	\$ 5.008	\$ 5.008	\$ 2.844	\$ 2.844	\$ 2.844	\$ 2.844	\$ 2.844	\$ 2.844
Kid's Korner	1997		\$ 201	\$ 247	\$ 398	\$ 422	\$ 441	\$ 441	\$ 186	\$ 186	\$ 380	\$ 380</td														

OLAP: operatori



drill-across

Permette di stabilire un collegamento tra due o più cubi correlati al fine di compararne i dati

65

OLAP: operatori

	Metrics	Dollar Sales	Quarter	Q1 1997	Q2 1997	Q3 1997	Q4 1997	Q1 1998	Q2 1998	Q3 1998	Q4 1998
Category											
Electronics		\$ 4.383	\$ 817	\$ 827	\$ 4.589	\$ 13.770	\$ 2.977	\$ 4.226	\$ 8.326		
Food		\$ 1.546	\$ 1.310	\$ 1.268	\$ 1.176	\$ 2.676	\$ 1.120	\$ 953	\$ 889		
Gifts		\$ 3.398	\$ 3.893	\$ 4.682	\$ 4.342	\$ 7.879	\$ 4.145	\$ 4.378	\$ 3.645		
Health & Beauty		\$ 1.826	\$ 878	\$ 1.904	\$ 1.434	\$ 2.156	\$ 898	\$ 1.207	\$ 1.404		
Household		\$ 9.314	\$ 8.124	\$ 9.331	\$ 11.614	\$ 17.453	\$ 7.604	\$ 12.698	\$ 12.436		
Kid's Korner		\$ 685	\$ 531	\$ 811	\$ 532	\$ 1.084	\$ 491	\$ 532	\$ 836		
Travel		\$ 603	\$ 1.293	\$ 1.456	\$ 1.145	\$ 1.507	\$ 719	\$ 840	\$ 1.726		

drill-across

	Quarter	Q1 1997	Q2 1997	Q3 1997	Q4 1997	Q1 1998	Q2 1998	Metrics	Discount	Dollar Sales	Discount	Dollar Sales	
Category													
Electronics		\$ 0	\$ 4.383	\$ 0	\$ 817	\$ 0	\$ 827	\$ 300	\$ 4.589	\$ 15	\$ 13.770	\$ 0	\$ 2.97
Food		\$ 25	\$ 1.546	\$ 0	\$ 1.310	\$ 0	\$ 1.268	\$ 38	\$ 1.176	\$ 0	\$ 2.676	\$ 0	\$ 1.12
Gifts		\$ 31	\$ 3.398	\$ 0	\$ 3.893	\$ 5	\$ 4.682	\$ 0	\$ 4.342	\$ 15	\$ 7.879	\$ 0	\$ 4.14
Health & Beauty		\$ 0	\$ 1.826	\$ 0	\$ 878	\$ 0	\$ 1.904	\$ 0	\$ 1.434	\$ 229	\$ 2.156	\$ 0	\$ 89
Household		\$ 0	\$ 9.314	\$ 228	\$ 8.124	\$ 175	\$ 9.331	\$ 35	\$ 11.614	\$ 5	\$ 17.453	\$ 211	\$ 7.60
Kid's Korner		\$ 0	\$ 685	\$ 0	\$ 531	\$ 32	\$ 811	\$ 40	\$ 532	\$ 0	\$ 1.084	\$ 0	\$ 49
Travel		\$ 0	\$ 603	\$ 0	\$ 1.293	\$ 200	\$ 1.456	\$ 0	\$ 1.145	\$ 0	\$ 1.507	\$ 0	\$ 71

66

Data mining

- Attività orientata a scoprire informazioni nascoste nei dati
 - ✓ In presenza di moli di dati molto elevate, l'utente non è sempre in grado di individuare tutti i pattern (modelli) significativi presenti
 - ✓ Il data mining raccoglie tecniche di intelligenza artificiale e pattern recognition per aiutare l'utente nella ricerca di pattern: è sufficiente indicare cosa e dove si vuole ricercare
 - Ricerche di mercato
 - Studio dell'efficacia del marketing
 - Segmentazione di mercato
 - Analisi delle abitudini di acquisto
 - Pianificazione aziendale
 - Modellazione degli investimenti
 - Rilevamento di attività fraudolente
 - Valutazione delle categorie di rischio
 - Riconoscimento di similarità tra sequenze di eventi
 - Valutazione di casi clinici e studio di modelli epidemiologici

67

Data mining: regole associative

- Consentono di determinare le regole di implicazione logica presenti nella base di dati, quindi di individuare i gruppi di affinità tra oggetti
- **Applicazioni:**
 - ✓ studio delle abitudini di acquisto per la pubblicità mirata e l'organizzazione della merce sugli scaffali (*market-basket analysis*)
 - ✓ studio della variabilità delle vendite in assenza di un certo prodotto

$\{\text{scarpe}\} \Rightarrow \{\text{calze}\}$

Chi compra
scarpe compra
anche calze

supporto=70 %

confidenza=85 %



La regola $X \Rightarrow Y$ ha supporto s se una frazione pari a s delle transazioni contengono tutti gli item in $X \cup Y$ es.: (il 70% delle transazioni include scarpe e calze)

La regola $X \Rightarrow Y$ ha confidenza c se una frazione pari a c delle transazioni in cui compare X contiene Y es.: (l'85% delle persone che comprano scarpe comprano anche calze)

La regola $A \Rightarrow D$ ha

- SUPPORTO = 60% perché $\{A, D\}$ compare in 6 transazioni su 10
- CONFIDENZA = 75% perché su 8 transazioni in cui compare A, in 6 compare anche D

La regola $D \Rightarrow A$ ha

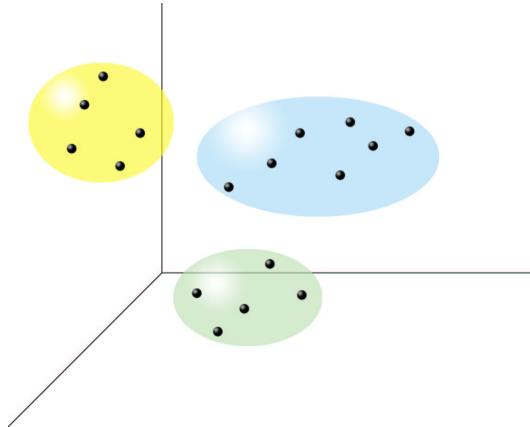
- SUPPORTO = 60%
- CONFIDENZA = 86% su 7 transazioni in cui compare D, in 6 compare anche A

TRANSAZIONE	ITEMS
1	A B C D
2	A B C
3	A D
4	B C D
5	A B D
6	B C
7	A C D
8	A B D
9	A B
10	A D

PROBLEMA: determinare tutte le regole associative che abbiano
 $SUPPORTO > minS$
e
 $CONFIDENZA > minC$

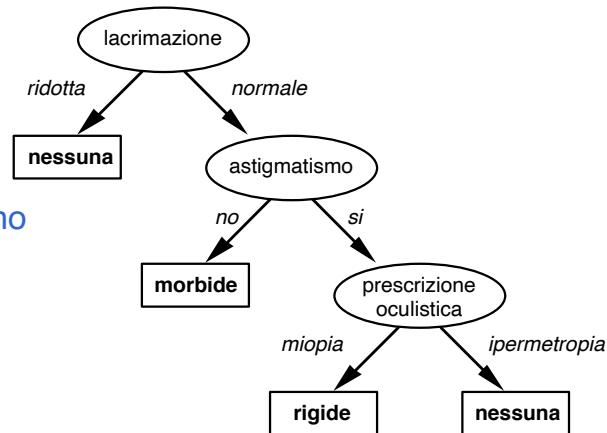
Data mining: clustering

- Data una popolazione di oggetti rappresentabili come punti in uno spazio multidimensionale in cui ciascuna dimensione corrisponde a una caratteristica di interesse, effettuare un clustering significa raggruppare gli oggetti in un ridotto numero di insiemi (**cluster**) che caratterizzino al meglio la popolazione
- **Applicazioni:**
 - ✓ segmentazione della clientela in categorie
 - ✓ valutazione di casi clinici sulla base della sintomatologia
 - ✓ analisi epidemiologica



Data mining: alberi decisionali

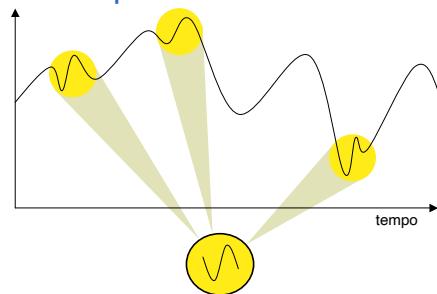
- Vengono usati per la comprensione di un particolare fenomeno poiché permettono di classificare, in ordine di importanza, le cause che portano al verificarsi di un evento
- **Applicazioni:**
 - ✓ valutazione delle categorie di rischio dei clienti per le società che concedono mutui e prestiti



70

Data mining: serie temporali

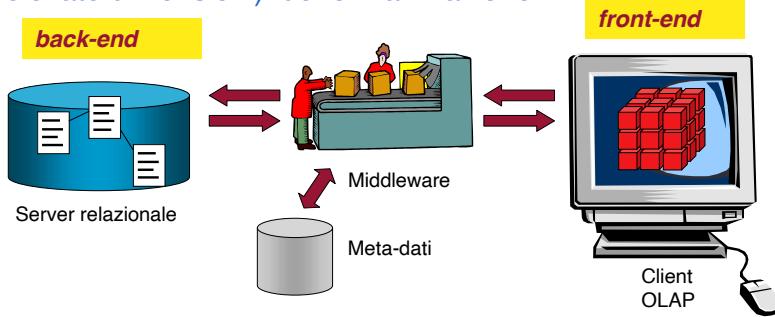
- Individuazione di pattern ricorrenti o atipici in sequenze di dati complesse
- **Applicazioni:**
 - ✓ identificazione di schemi associati alla crescita dei titoli di borsa
 - ✓ rilevazione di anomalie in un sistema di monitoraggio
 - ✓ studio delle correlazioni tra serie temporali distinte
 - ✓ identificazione di aziende con modelli di sviluppo simili
 - ✓ analisi dei percorsi di navigazione in siti web.



71

ROLAP

- Giustificato dall'enorme lavoro svolto in letteratura sul modello relazionale, dalla diffusa esperienza aziendale sull'utilizzo e l'amministrazione di basi di dati relazionali e dall'elevato livello di prestazioni e flessibilità raggiunto dai DBMS relazionali
 - ✓ Necessità di elaborare tipologie specifiche di schemi che permettano di traslare il modello multidimensionale sul modello relazionale: *schema a stella*.
 - ✓ Problema delle prestazioni (costose operazioni di join tra tabelle di elevate dimensioni): *denormalizzazione*.

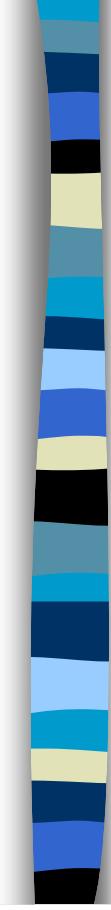


72

MOLAP

- Basato su un modello logico ad hoc sul quale i dati e le operazioni multidimensionali possono essere direttamente rappresentati.
- I dati vengono fisicamente memorizzati in vettori e l'accesso è di tipo posizionale.
 - ✓ Il grosso vantaggio dell'approccio MOLAP rispetto a quello ROLAP è che le operazioni multidimensionali sono realizzabili in modo semplice e naturale, senza necessità di ricorrere a join; le prestazioni risultano pertanto ottime.
 - ✓ Non esistendo ancora uno standard per il modello logico multidimensionale, le diverse implementazioni MOLAP hanno veramente poco in comune: in genere, solo l'utilizzo di tecnologie di ottimizzazione specifiche per trattare il problema della sparsità.

73



La qualità

La qualità di un processo misura la sua aderenza agli obiettivi degli utenti

- Fattori che caratterizzano la qualità dei dati in un DW:
 1. **Accuratezza:** la conformità tra il valore memorizzato e quello reale.
 2. **Attualità:** il dato memorizzato non è obsoleto.
 3. **Completezza:** non mancano informazioni.
 4. **Consistenza:** la rappresentazione dei dati è uniforme.
 5. **Disponibilità:** i dati sono facilmente disponibili all'utente.
 6. **Tracciabilità:** è possibile risalire alla fonte di ciascun dato.
 7. **Chiarezza:** i dati sono facilmente interpretabili