Data Warehouses

Letizia Tanca (*)
Politecnico di Milano

(*) with the kind support of Rosalba Rossato

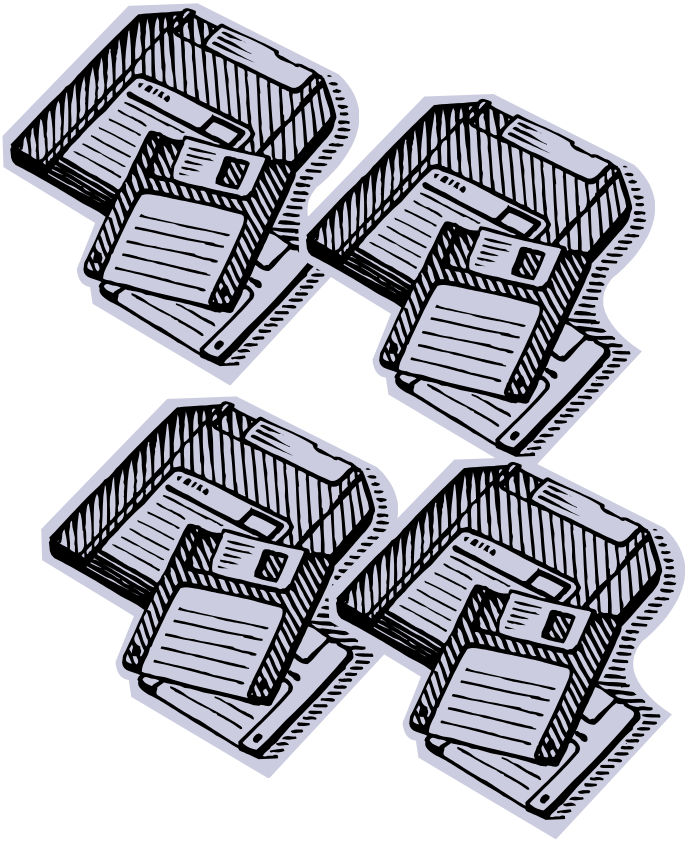
Outline

- What is a Data Warehouse
- Data Warehouse Architecture
- Data Warehouse Design

What is a Data Warehouse

- Data should be integrated across the enterprise(s)
- Summary data provide real value to the organization
- Historical data hold the key to understanding data over time
- What-if capabilities are required

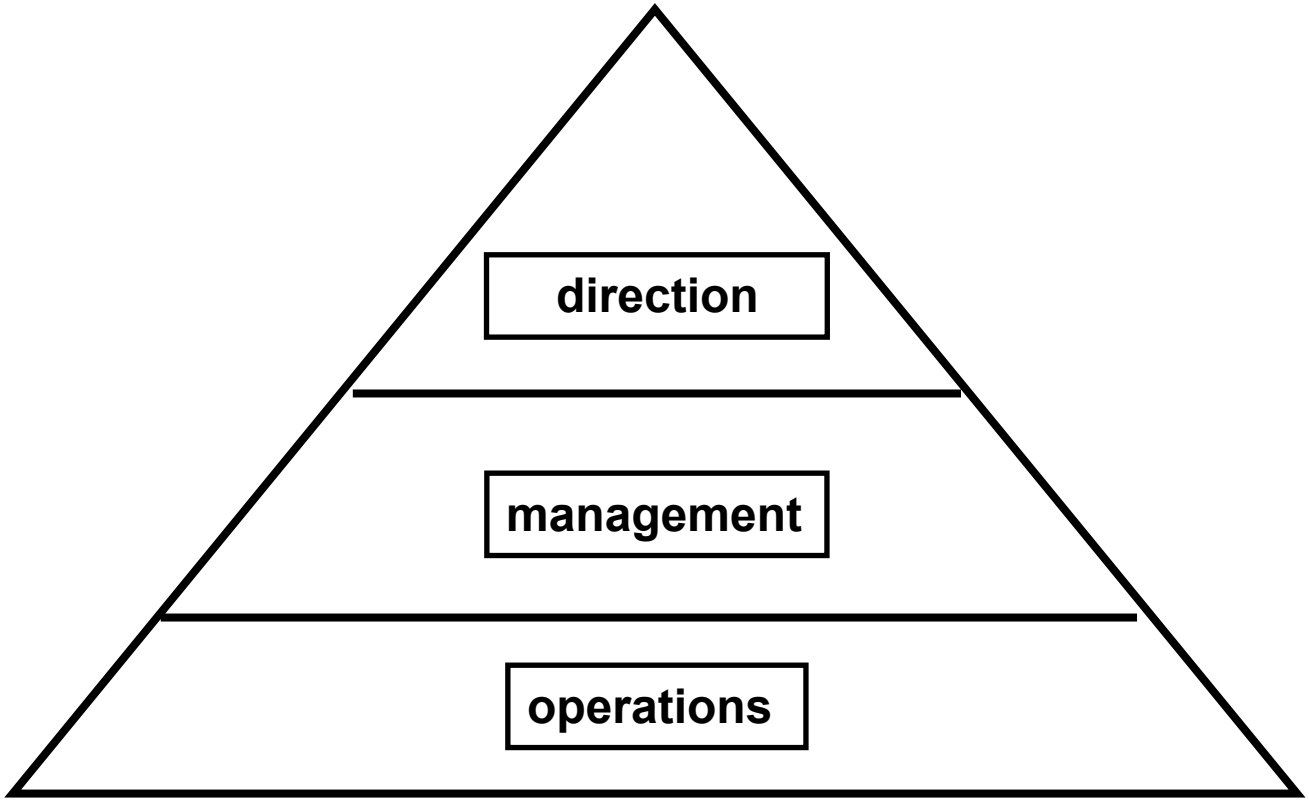
What is a Data Warehouse?



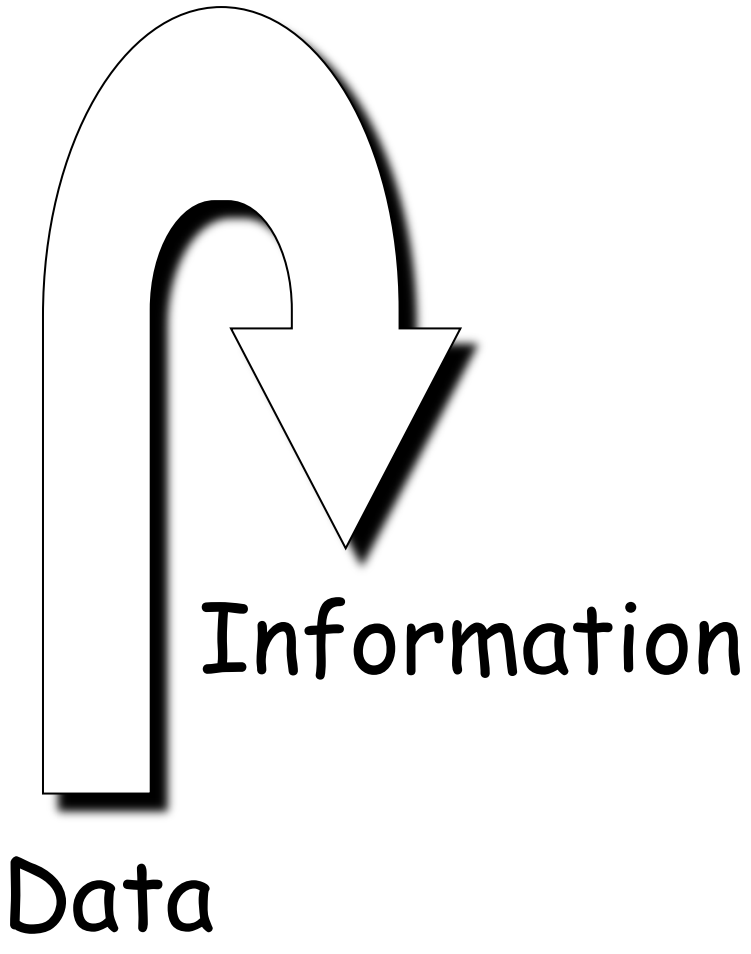
A single, complete and consistent store of data obtained from a variety of different sources made available to end users, *so that they can understand and use it in a business context.*

[Barry Devlin]

Business Processes' Pyramid



An alternative definition of Data Warehouse



A data warehouse is (the result of) a process for *transforming data into information* and for making it available to users *in a timely enough manner to make a difference.*

[Forrester Research, April' 96]

Data Warehouse

- **As a dataset:** decision support database maintained separately from the organization's operational database
- **As a process:** technique for assembling and managing data from various sources with the purpose of answering business questions. Thus making decisions that were not previously possible

Data Warehouse

- A Data Warehouse is a
  - **subject-oriented,**
  - **integrated,**
  - **time-varying,**
  - **non-volatile**

collection of data that is used primarily in **organizational decision making**.

[Bill Inmon, Building the Data Warehouse, 1996]

Dimensions of a Data Warehouse

- Data warehouses are very large databases
  - Terabytes (10^{12} bytes):
  - Petabytes (10^{15} bytes): e.g. Geographic Information Systems
  - Exabytes (10^{18} bytes): e.g. National Medical Records
  - Zettabytes (10^{21} bytes): e.g. Weather reports, including images
  - Zottabytes (10^{24} bytes): e.g. Intelligence Agency Videos

DW is a specialized DB

Standard DB (OLTP)

- Mostly updates
- Many small transactions
- Mb - Gb of data
- Current snapshot
- Index/hash on p.k.
- Raw data
- Thousands of users (e.g., clerical users)

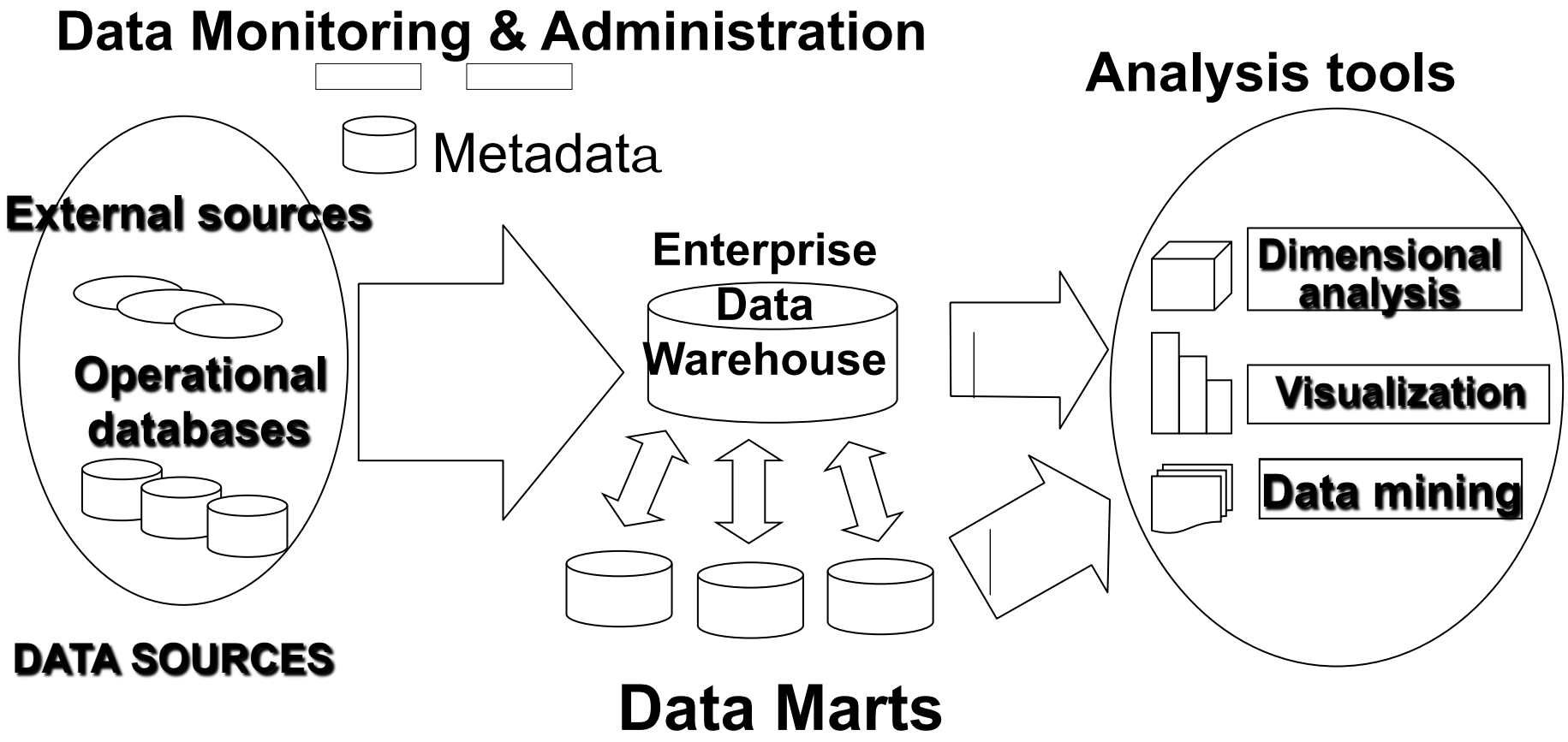
Warehouse (OLAP)

- Mostly reads
- Queries are long and complex
- Gb - Tb of data
- History
- Lots of scans
- Summarized, reconciled data
- Hundreds of users

Where is a DW useful

- Commerce: sales and complaints analysis, client fidelization, shipping and stock control
- Manufacturing plants: production cost control, provision and order support
- Financial services: risk and credit card analysis, fraud detection
- Telecommunications: call flow analysis, subscribers' profiles
- Healthcare structures: patients' ingoing and outgoing flows, cost analysis

Architecture for a data warehouse



OLAP-oriented data models

- must support sophisticated analyses and computations over different dimensions and hierarchies
- Most appropriate data model: data cube
- Cube dimensions are the search keys
- Each dimension may be hierarchical
  - DATE {DAY-MONTH-TRIMESTER-YEAR}
  - PRODUCT {BRAND - TYPE - CATEGORY}

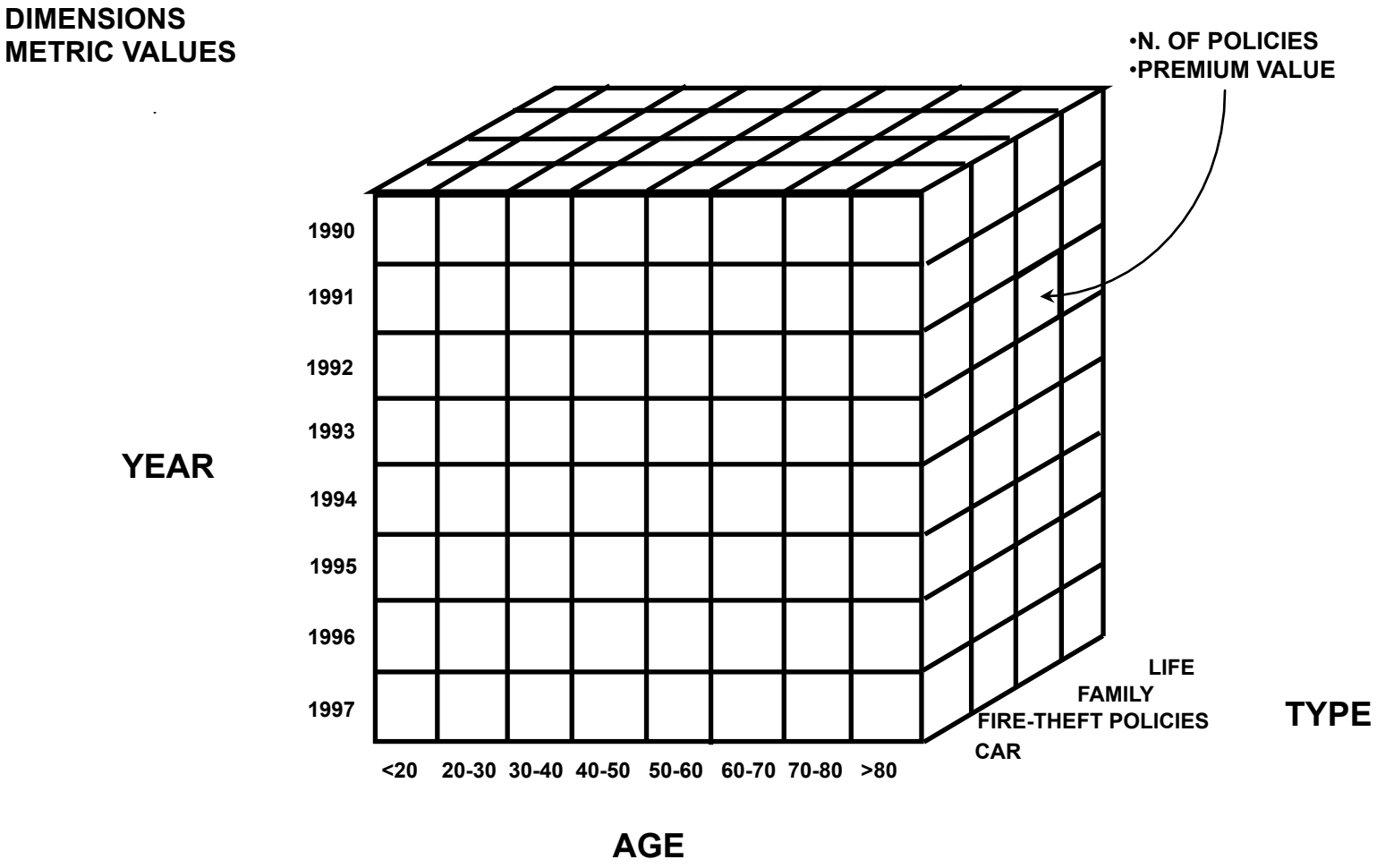
    (e.g. LAND ROVER - CARS - VEHICLES)
- Cube cells contain metric values

Examples of data warehouse queries

- Show total sales across all products at increasing aggregation levels for a geography dimension, from state to country to region, for 1999 and 2000.
- Create a cross-tabular analysis of our operations showing expenses by territory in South America for 1999 and 2000. Include all possible subtotals.
- List the top 10 sales representatives in Asia according to sales revenue for automotive products in year 2000, and rank their commissions.

LOGICAL MODELS FOR OLAP

AN INSURANCE COMPANY DATA CUBE



Dimensional Fact Model

- Allows one to describe a set of fact schemata
- The components of a fact schema are:
  - Facts
  - Measures
  - Dimensions
  - Dimension Hierarchy

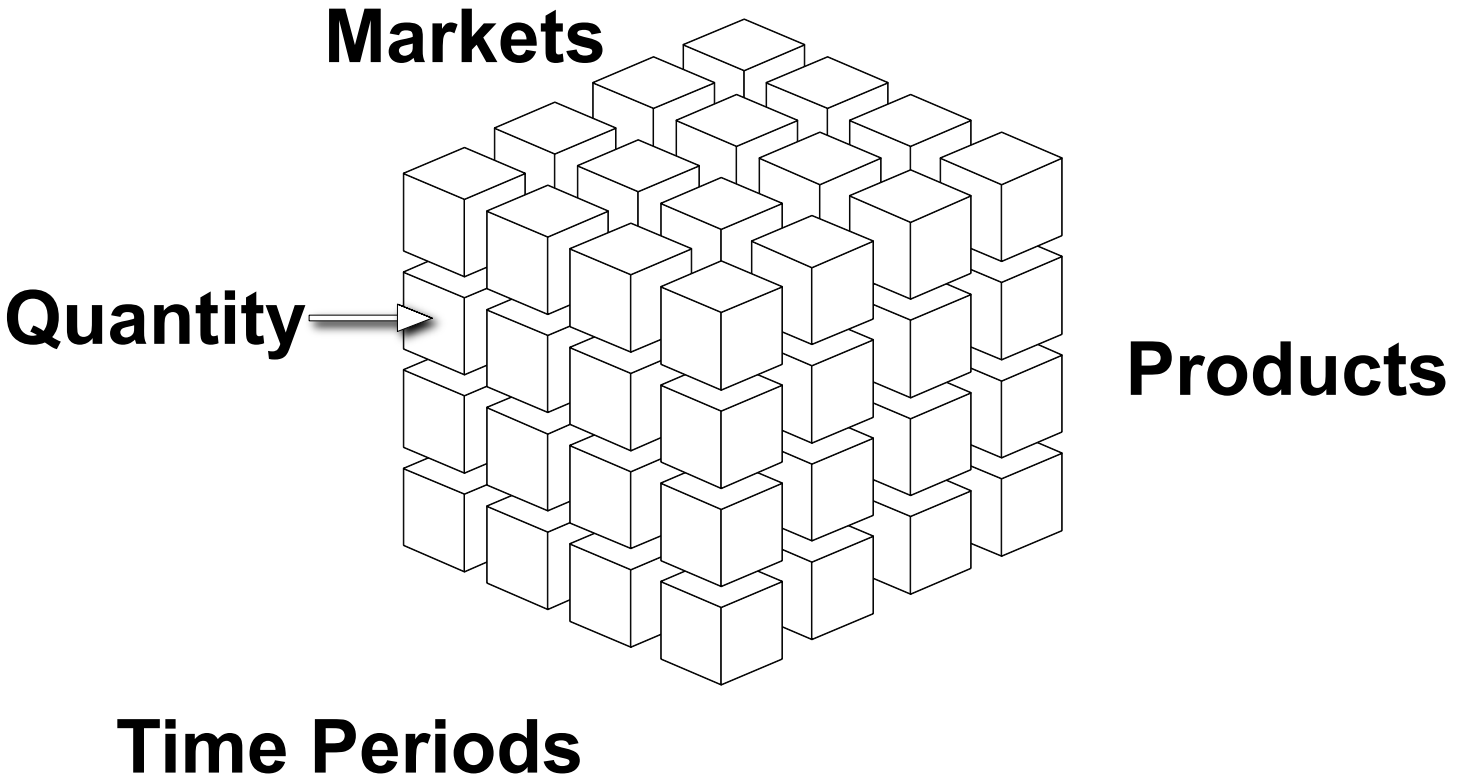
DFM: elements

- A fact is a concept that is relevant for the decisional process; typically it models a set of events of the organization
- A measure is a numerical property of a fact
- A dimension is a fact property defined w.r.t. a finite domain; it describes an analysis coordinate for the fact

Examples

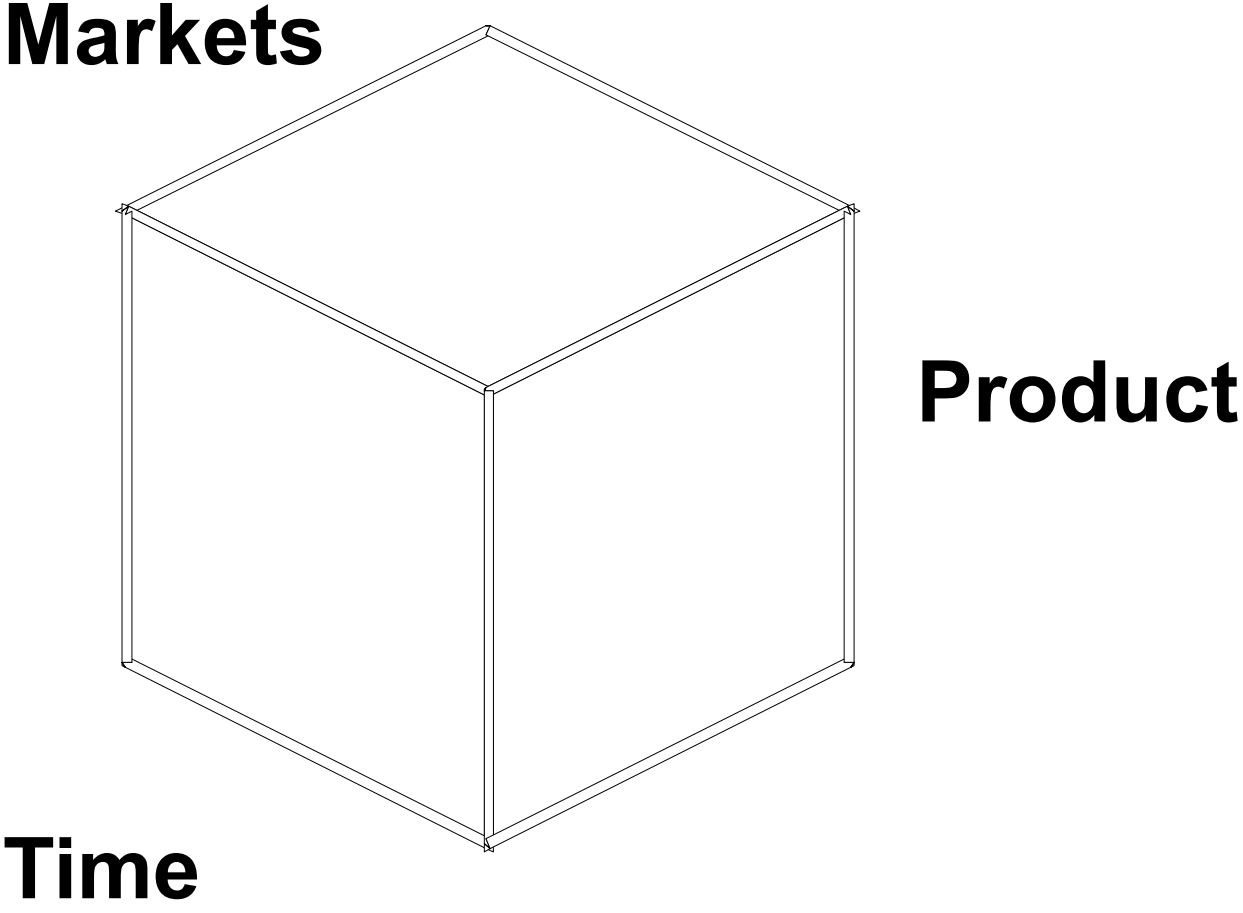
- **Store chain**
  - Fact: **sales**
  - Measures: **sold quantity, gross income**
  - Dimensions: **product, time, zone**
- **Telecom Operator**
  - Fact: **phone call**
  - Measures : **cost, duration**
  - Dimensions: **caller subscriber,called subscriber, time**

Multidimensional Representation



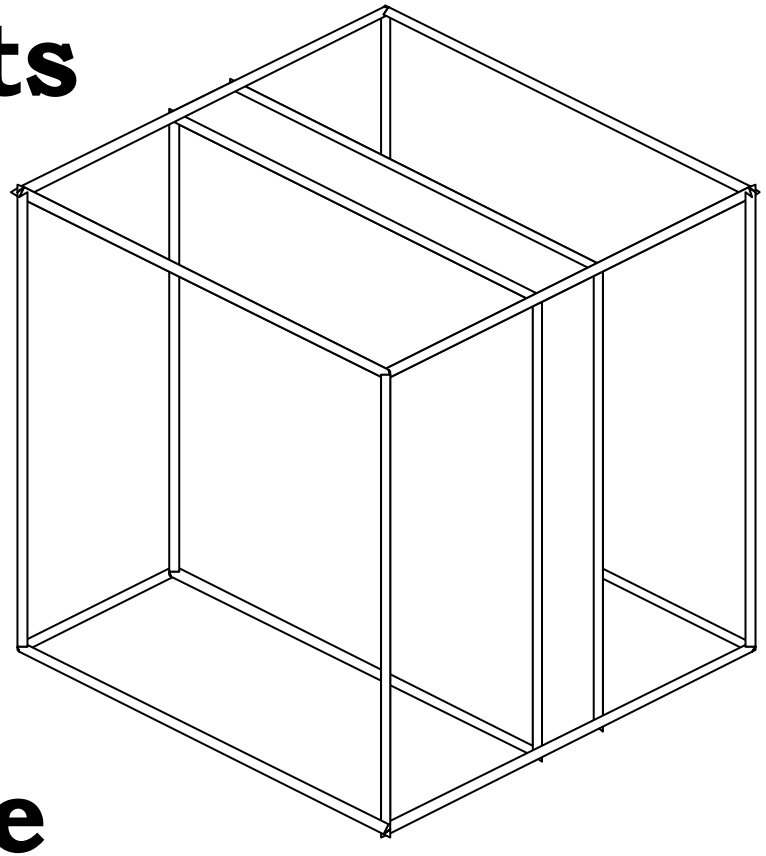
Sales Cube

Multidimensional data views



The area manager examines product sales
of his/her own markets

Markets

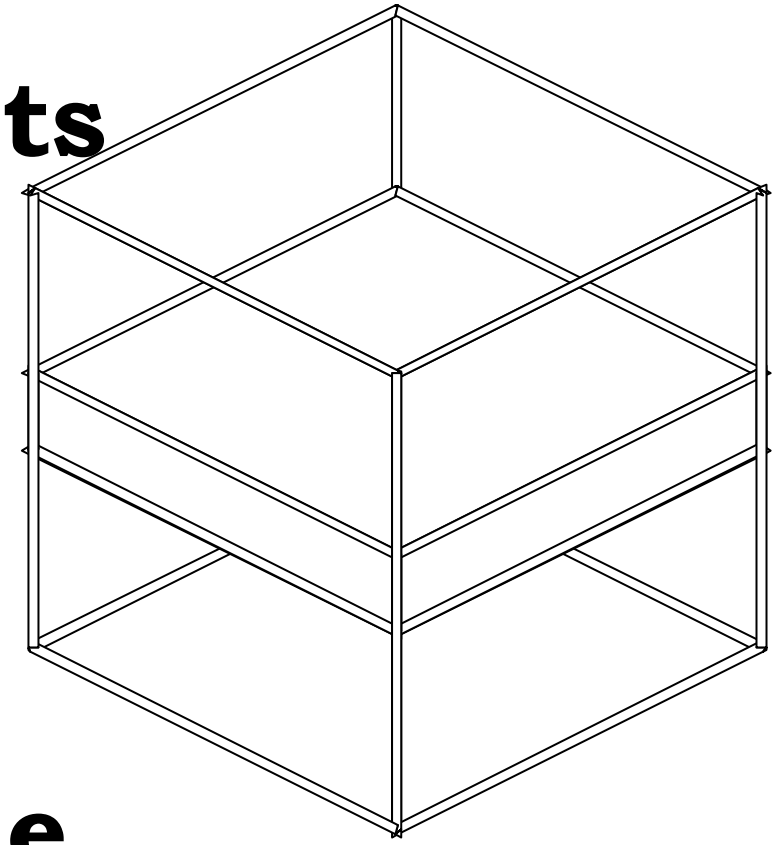


Products

Time

Product manager examines the sales of a specific product in all periods and in all markets

Markets

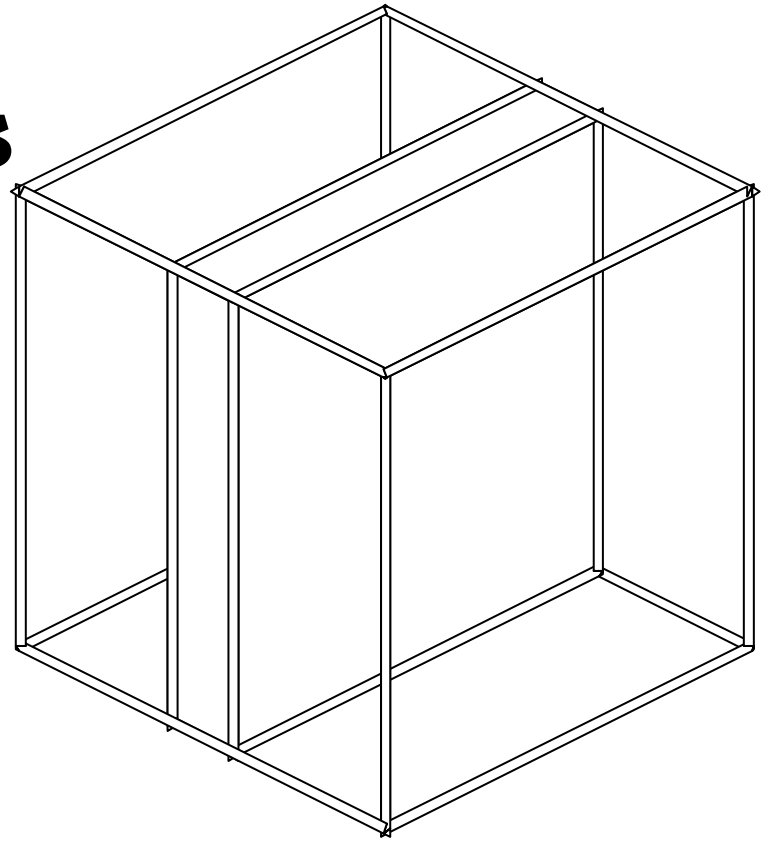


Products

Time

Financial manager examines product sales in all markets, for the current period and the previous one

Markets

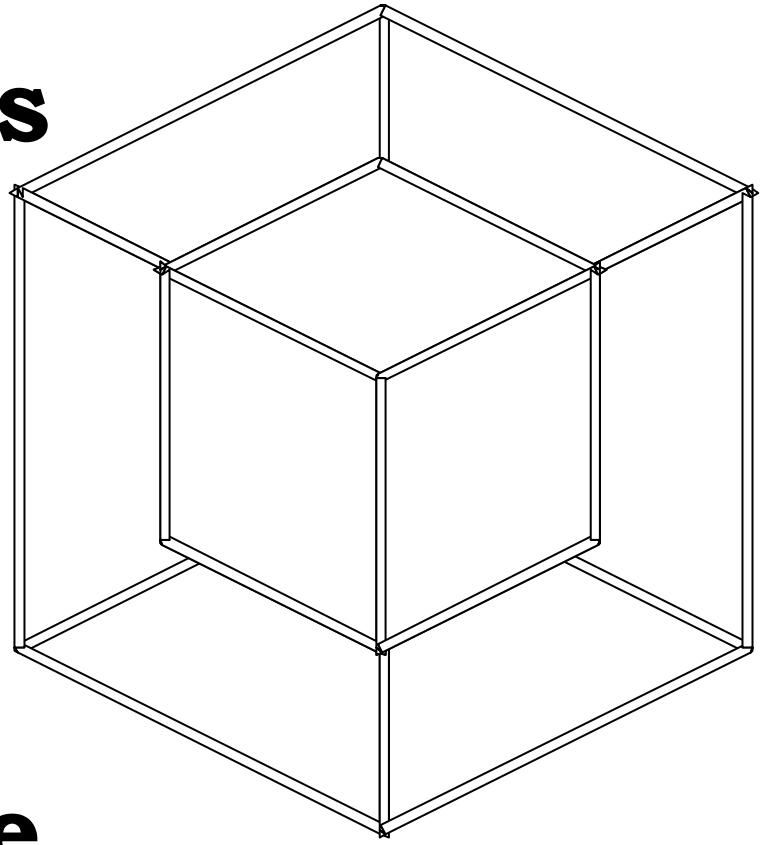


Products

Time

The strategic manager concentrates on a category of products, a specific region and a medium time span

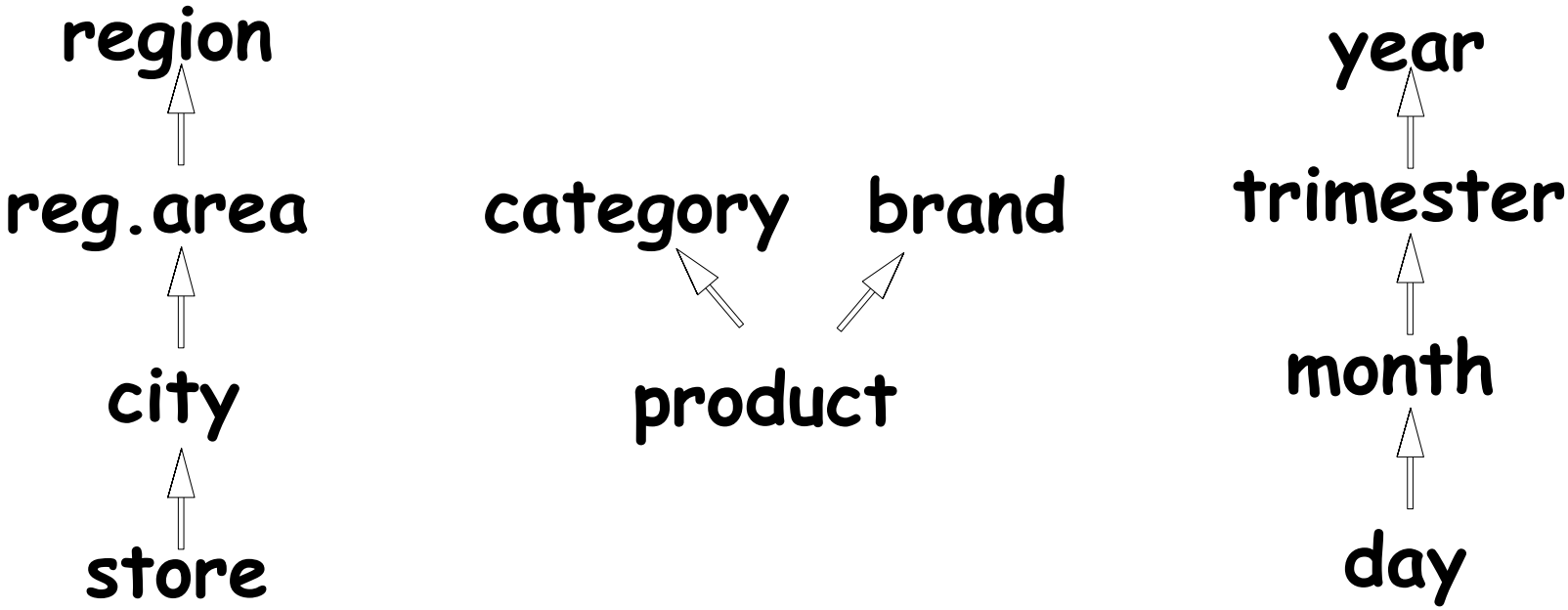
Markets



Products

Time

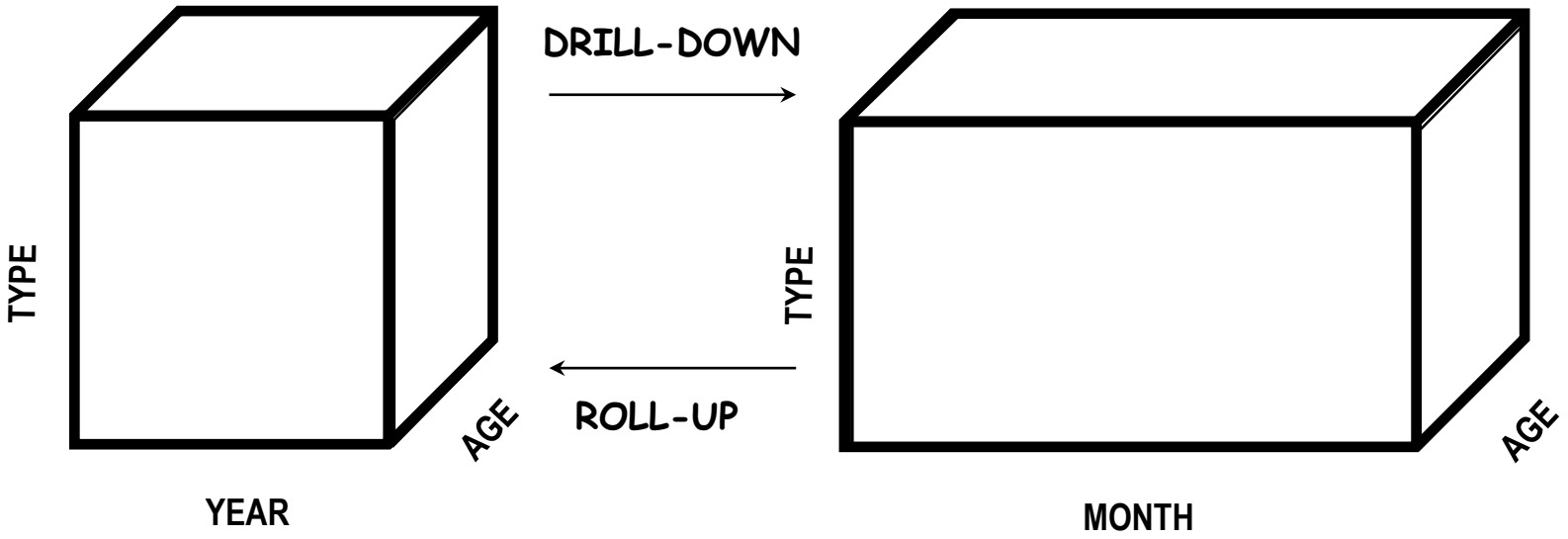
Dimensions and hierarchies



OLAP OPERATIONS

- **roll-up**
  - Aggregates data at a higher level – **e.g. last year's sales volume per product category and per region**
- **drill-down**
  - De-aggregates data at the lower level – **e.g. for a given product category and a given region, show daily sales**
- **slice-and-dice**
  - Applies selections and projections, which reduce data dimensionality
- **pivoting**
  - Selects two dimensions to re-aggregate data (cube re-orientation)
- **ranking**
  - Sorts data according to predefined criteria
- **traditional operations** (select, project, join, derived attributes, etc.)

OLAP OPERATIONS



Roll-up

Metrics	Dollar Sales									
Customer Region / Month	North-East	Mid-Atlantic	South-East	Central	South	North-West	South-West	England	France	Germany
Jan 97	$ 620	$ 753	$ 30	$ 660	$ 2.405	$ 1.312	$ 440	$ 1.002	$ 1.002	$ 383
Feb 97	$ 258	$ 252	$ 800	$ 975	$ 160	$ 582	$ 744	$ 310	$ 799	$ 118
Mar 97	$ 648	$ 244	$ 148	$ 250	$ 1.085	$ 2.961	$ 650	$ 1.240	$ 119	$ 142
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
Jun 97	$ 842	$ 582	$ 1.281	$ 937	$ 240	$ 774	$ 176	$ 1.139	$ 652	$ 254
Jul 97	$ 652	$ 690	$ 486	$ 1.293	$ 605	$ 303	$ 818	$ 103	$ 124	$ 173
Aug 97	$ 1.783	$ 304	$ 1.032	$ 170	$ 398	$ 356	$ 432	$ 190	$ 241	$ 407
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
Nov 97	$ 39	$ 1.602	$ 1.082	$ 1.187	$ 842	$ 759	$ 745	$ 232	$ 101	$ 1.037
Dec 97	$ 381	$ 1.588	$ 343	$ 118	$ 1.459	$ 635	$ 2.021	$ 259	$ 210	$ 119
Jan 98	$ 311	$ 1.174	$ 2.634	$ 3.130	$ 954	$ 2.083	$ 1.351	$ 747	$ 426	$ 447
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
Apr 98	$ 407	$ 841	$ 524	$ 712	$ 133	$ 2.486	$ 49	$ 390	$ 1.298	$ 221
May 98	$ 667	$ 1.721	$ 440	$ 148	$ 80	$ 1.310	$ 303	$ 104	$ 657	$ 65
Jun 98	$ 699	$ 1.096	$ 898	$ 353	$ 902	$ 839		$ 230	$ 155	$ 105
Jul 98	$ 586	$ 1.897	$ 412	$ 226	$ 406	$ 361	$ 1.628	$ 267	$ 1.011	$ 41
Aug 98	$ 894	$ 326	$ 792	$ 1.832	$ 1.199	$ 295	$ 1.816	$ 277	$ 102	$ 118
Sep 98	$ 338	$ 3.179	$ 505	$ 427	$ 99	$ 2.976	$ 885	$ 135	$ 85	$ 1.110
Oct 98	$ 544	$ 413	$ 1.467	$ 209	$ 679	$ 706	$ 556	$ 480	$ 485	$ 99
Nov 98	$ 671	$ 459	$ 1.471	$ 2.066	$ 701	$ 716	$ 986	$ 1.127	$ 154	$ 440
Dec 98	$ 836	$ 2.096	$ 1.726	$ 3.642	$ 395	$ 1.740	$ 1.943	$ 1.143	$ 366	$ 307

Metrics	Dollar Sales									
Customer Region / Quarter	North-East	Mid-Atlantic	South-East	Central	South	North-West	South-West	England	France	Germany
Q1 1997	$ 1.526	$ 1.249	$ 978	$ 1.885	$ 3.650	$ 4.855	$ 1.834	$ 2.552	$ 1.920	$ 643
Q2 1997	$ 2.979	$ 1.415	$ 2.664	$ 1.582	$ 1.130	$ 1.906	$ 884	$ 1.393	$ 1.402	$ 516
Q3 1997	$ 3.016	$ 1.772	$ 5.076	$ 2.050	$ 1.443	$ 2.311	$ 2.321	$ 608	$ 575	$ 782
Q4 1997	$ 2.711	$ 5.030	$ 2.025	$ 1.961	$ 3.601	$ 2.112	$ 3.976	$ 918	$ 531	$ 1.676
Q1 1998	$ 5.288	$ 3.399	$ 4.935	$ 9.174	$ 3.601	$ 9.484	$ 3.844	$ 2.720	$ 979	$ 1.897
Q2 1998	$ 1.773	$ 3.658	$ 1.862	$ 1.213	$ 1.115	$ 4.635	$ 352	$ 724	$ 2.110	$ 391
Q3 1998	$ 1.818	$ 5.402	$ 1.709	$ 2.485	$ 1.704	$ 3.632	$ 4.329	$ 679	$ 1.198	$ 1.269
Q4 1998	$ 2.051	$ 2.968	$ 4.664	$ 5.917	$ 1.775	$ 3.162	$ 3.485	$ 2.750	$ 1.005	$ 846

Roll-up

Category	Year	Metrics / Customer Region	Dollar Sales: 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
	1998		$ 1.184	$ 4.529	$ 1.892	$ 7.232	$ 651	$ 9.488	$ 476	$ 2.683	$ 462
Food	1997		$ 759	$ 682	$ 729	$ 262	$ 588	$ 469	$ 807	$ 156	$ 615
	1998		$ 538	$ 925	$ 959	$ 677	$ 213	$ 1.503	$ 261	$ 165	$ 175
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	$ 292
	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
	1998		$ 5.787	$ 5.320	$ 5.416	$ 6.812	$ 4.334	$ 5.008	$ 7.588	$ 2.139	$ 3.649
Kid's Korner	1997		$ 201	$ 396	$ 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

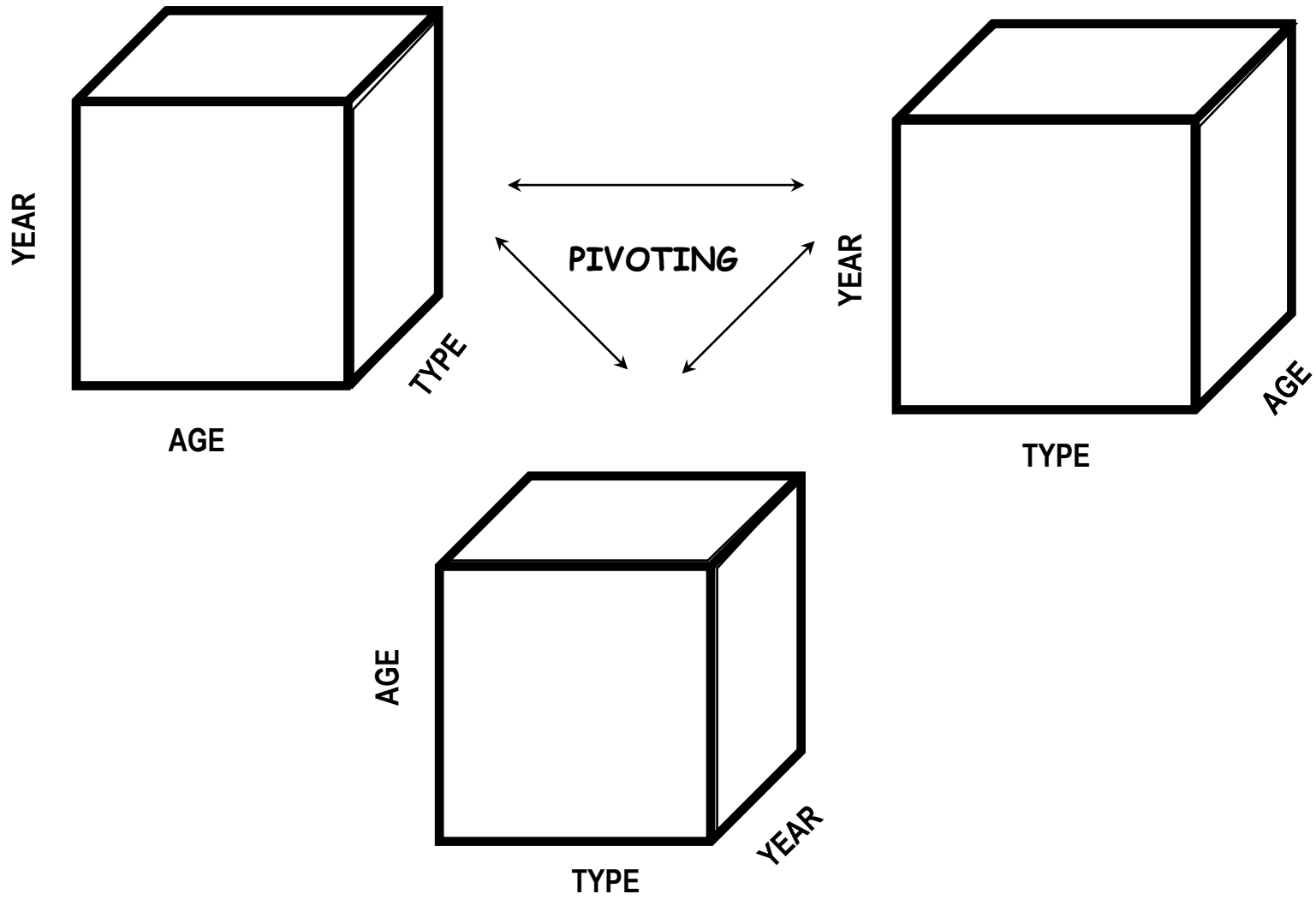
Category	Year	Metrics / Dollar Sales
Electronics	1997	$ 10.616
	1998	$ 29.299
Food	1997	$ 5.300
	1998	$ 5.638
Gifts	1997	$ 16.315
	1998	$ 20.047
Health & Beauty	1997	$ 6.042
	1998	$ 5.665
Household	1997	$ 38.383
	1998	$ 50.391
Kid's Korner	1997	$ 2.550
	1998	$ 2.943
Travel	1997	$ 4.497
	1998	$ 4.792

Drill-down

	Metrics	Dollar Sales	
	Year	1997	1998
Category			
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

	Metrics	Dollar Sales											
	Customer Region	North-East		Mid-Atlantic		South-East		Central		South		North-West	
	Year	1997	1998	1997	1998	1997	1998	1997	1998	1997	1998	1997	1998
Category													
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	$ 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	$ 409	$ 221	$ 323	$ 592
Travel		$ 624	$ 608	$ 505	$ 559	$ 564	$ 1.096	$ 386	$ 611	$ 300	$ 464	$ 978	$ 316

OLAP OPERATIONS

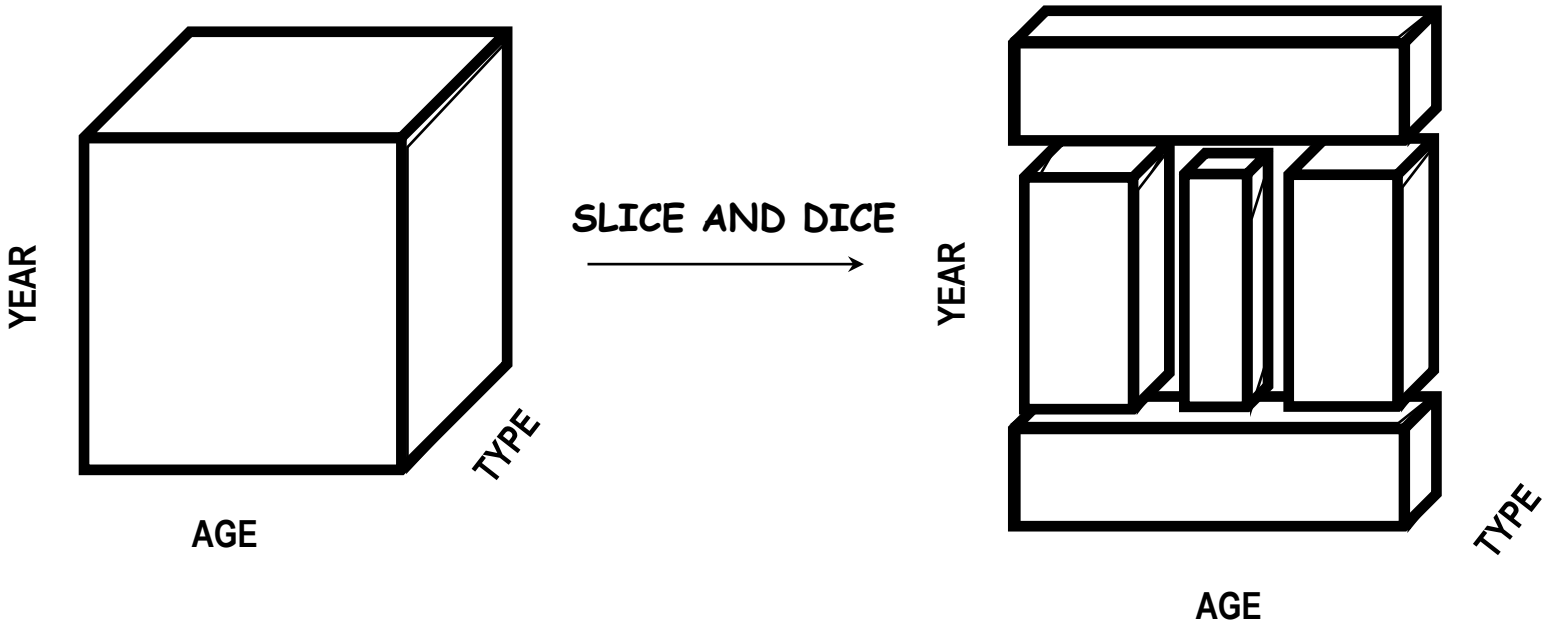


Pivoting

Category	Year	Metrics	Dollar Sales
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

	Metrics	Dollar Sales	
	Year	1997	1998
Category			
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

OLAP OPERATIONS



Slice and Dice

Metrics / Customer City / Subcategory	Dollar Sales										
	Afton	Akron	Albon	Alcameda	Alka	Allagash	Alta	Altoola	Amestra	Amsterdam	Andersonville
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		$ 76		
Lawn & Garden	$ 75		$ 100		$ 15	$ 63	$ 100		$ 180	$ 67	$ 40
Learning	$ 16							$ 37			
Meat & Cheese		$ 40		$ 20			$ 20				$ 25
Miscellaneous		$ 200	$ 1.320		$ 200	$ 139			$ 993		
Natural Remedies	$ 13								$ 13		
Pets	$ 215		$ 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

Metrics / Customer City / Subcategory	Dollar Sales					
	Alta	Armstrong	Avery Heights	Lane	Mt. Everest	San Fransisco
Audio		$ 98		$ 123	$ 85	
Comfort			$ 118		$ 1.495	
Gadgets	$ 199					$ 199

Slice and Dice

Category	Year	Metrics / Customer Region	Dollar Sales: 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
	1998		$ 1.184	$ 4.529	$ 1.892	$ 7.232	$ 651	$ 9.488	$ 476	$ 2.683
Food	1997		$ 759	$ 682	$ 729	$ 262	$ 588	$ 469	$ 807	$ 156
	1998		$ 538	$ 925	$ 959	$ 677	$ 213	$ 1.503	$ 261	$ 165
Gifts	1997		$ 2.532	$ 1.355	$ 1.854	$ 1.413	$ 2.535	$ 2.132	$ 1.904	$ 908
	1998		$ 1.955	$ 2.785	$ 2.800	$ 2.695	$ 1.813	$ 2.844	$ 1.778	$ 1.158
Health & Beauty	1997		$ 624	$ 640	$ 1.317	$ 647	$ 588	$ 754	$ 654	$ 143
	1998		$ 611	$ 887	$ 566	$ 382	$ 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
Travel	1997		$ 624	$ 505	$ 564	$ 386	$ 300	$ 978	$ 416	$ 48
	1998		$ 608	$ 559	$ 1.096	$ 611	$ 464	$ 316	$ 573	$ 257

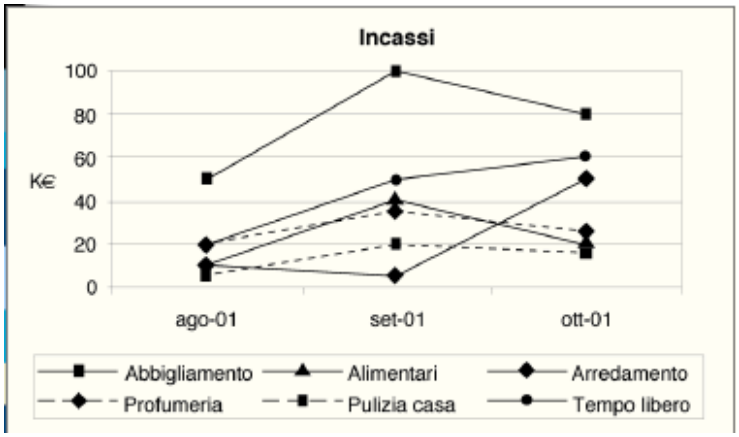
Filter Details:
Year = 1998

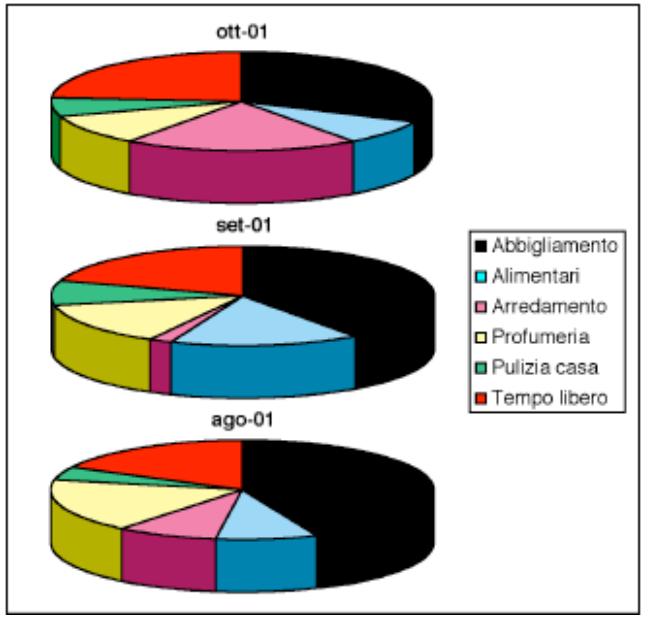
Category	Metrics / Customer Region	Dollar Sales: North-East	Mid-Atlantic	South-East	Central	South	North-West	South-West	England
Electronics		$ 1.184	$ 4.529	$ 1.892	$ 7.232	$ 651	$ 9.488	$ 476	$ 2.683
Food		$ 538	$ 925	$ 959	$ 677	$ 213	$ 1.503	$ 261	$ 165
Gifts		$ 1.955	$ 2.785	$ 2.800	$ 2.695	$ 1.813	$ 2.844	$ 1.778	$ 1.158
Health & Beauty		$ 611	$ 887	$ 566	$ 382	$ 499	$ 1.162	$ 1.044	$ 273
Household		$ 5.787	$ 5.320	$ 5.416	$ 6.812	$ 4.334	$ 5.008	$ 7.588	$ 2.139
Kid's Korner		$ 247	$ 422	$ 441	$ 380	$ 221	$ 592	$ 290	$ 198
Travel		$ 608	$ 559	$ 1.096	$ 611	$ 464	$ 316	$ 573	$ 257

VISUALIZATION and REPORTS

Data may be visualized graphically, in an Excel-like format: tables, hystograms, graphics, 3D surfaces, etc.

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





Examples of operators: initial table

month	product	Sales am. t	Quantity
February	pasta	130.000.000	45.000
March	pasta	140.000.000	50.000
April	pasta	135.000.000	51.000

Drill-down: adding a dimension

Drill-down on the zone

month	product	zone	quantity
February	pasta	north	15.000
February	pasta	east	17.000
February	pasta	center	13.000
March	pasta	north	18.000
March	pasta	east	18.000
March	pasta	center	14.000
April	pasta	north	18.000
April	pasta	east	17.000
April	pasta	center	16.000

Roll-up: dimension elimination

Roll-up on the month

product	zone	Quantity
pasta	north	51.000
pasta	east	52.000
pasta	center	43.000

Aggregate Queries

Examples:

- Total sales per product category, per supermarket, per day
- Total monthly sales for all the products, per supermarket
- Total monthly sales per category per supermarket
- Avg. monthly sales per category, for all supermarkets

Olap logical models

- **MOLAP** (*Multidimensional On-Line Analytical Processing*) stores data by using a multidimensional data structure (a "physical" data cube)

- **ROLAP** (*Relational On-Line Analytical Processing*) uses the relational data model to represent multidimensional data

The data cube in SQL

- It expresses all the possible tuple aggregations of a table
- It uses the new polymorphic value ALL

Data cube in SQL (ROLAP)

```
select Model, Year,
       Color, sum(Sales)
from Sales
where Model in {'Fiat','Ford'}
  and Color = 'Red'
  and Year between 1994 and 1995
group by (Model, Year, Color)
with cube
```

Relevant Facts

model	year	color	sales
fiat	1994	red	50
fiat	1995	red	85
ford	1994	red	80

All the data in the cube

model	year	color	sum (sales)
fiat	1994	red	50
fiat	1995	red	85
fiat	1994	ALL	50
fiat	1995	ALL	85
fiat	ALL	red	135
fiat	ALL	ALL	135
ford	1994	red	80
ford	1994	ALL	80
ford	ALL	red	80
ford	ALL	ALL	80
ALL	1994	red	130
ALL	1995	red	85
ALL	ALL	red	215
ALL	1994	ALL	130
ALL	1995	ALL	85
ALL	ALL	ALL	215

Roll up

ROLLUP enables a SELECT statement to calculate multiple levels of subtotals across a specified group of dimensions.

Using the ROLLUP operator instead of the CUBE operator will eliminate the results that contain a NULL (ALL) only in one column (except for the last).

In this case the aggregations only by model or only by year are not computed.

```
select Model, Year,
       Color, sum(Sales)
from Sales
where Model in {'Fiat','Ford'}
  and Color = 'Red'
  and Year between 1994 and 1995
group by (Model, Year, Color)
       with ROLLUP
```

The data after roll-up

model	year	color	sum(sales)
fiat	1994	red	50
fiat	1995	red	85
ford	1994	red	80
fiat	1994	ALL	50
fiat	1995	ALL	85
ford	1994	ALL	80
fiat	ALL	ALL	135
ford	ALL	ALL	80
ALL	ALL	ALL	215

A Simple Cross-Tabular Report With Subtotals

- A database containing: products, customers, sale channels,with the measures amount_sold and quantity_sold.
- Amounts in dollars, aggregated by country, sliced on France and US and the month of September, aggregated by channel
- Half of the values needed for this report would not be calculated just by means of a query that requested SUM(amount_sold) and did a GROUP BY(channel_desc, country_id).

CHANNEL	COUNTRY		
	France	US	Total
Internet	9,597	124,224	133,821
Direct Sales	61,202	638,201	699,403
Total	70,799	762,425	833,224

Cube creation

```
SELECT channels.channel_desc,
  countries.country_iso_code,
  TO_CHAR(SUM(amount_sold), '9,999,999,999')
  SALES$
FROM sales, customers, times, channels, countries
WHERE sales.time_id=times.time_id AND
  sales.cust_id=customers.cust_id AND
  sales.channel_id= channels.channel_id AND
  channels.channel_desc IN ('Direct Sales',
  'Internet') AND
  times.calendar_month_desc='2000-09' AND
  customers.country_id=countries.country_id AND
  countries.country_iso_code IN ('US','FR')
GROUP BY (channels.channel_desc,
  countries.country_iso_code)
with CUBE;
```

Resulting table

CHANNEL_DESC	COUNTRY	SALES ($)
ALL	ALL	833,224
ALL	FR	70,799
ALL	US	762,425
INTERNET	ALL	133,821
INTERNET	FR	9,597
INTERNET	US	124,224
DIRECT SALES	ALL	699,403
DIRECT SALES	FR	61,202
DIRECT SALES	US	638,201

Roll-up query

```
SELECT channels.channel_desc, calendar_month_desc,
  countries.country_iso_code,
  TO_CHAR(SUM(amount_sold), '9,999,999,999') SALES
  $
FROM sales, customers, times, channels, countries
WHERE sales.time_id=times.time_id AND
  sales.cust_id=customers.cust_id AND
  sales.channel_id= channels.channel_id AND
  channels.channel_desc IN ('Direct Sales',
  'Internet') AND times.calendar_month_desc IN
  ('2000-09', '2000-10') AND
  countries.country_iso_code IN ('GB', 'US')
GROUP BY (channels.channel_desc,
  calendar_month_desc, countries.country_iso_code)
with ROLLUP;
```

Roll-up result

CHANNEL_DESC	CALENDAR	COUNTRY	SALES$
INTERNET	2000-09	GB	228,241
INTERNET	2000-09	US	228,241
INTERNET	2000-09	ALL	456,482
INTERNET	2000-10	GB	239,236
INTERNET	2000-10	US	239,236
INTERNET	2000-10	ALL	478,473
INTERNET	ALL	ALL	934,955
DIRECT_SALES	2000-09	GB	1,217,808
DIRECT_SALES	2000-09	US	1,217,808
DIRECT_SALES	2000-09	ALL	2,435,616
DIRECT_SALES	2000-10	GB	1,225,584
DIRECT_SALES	2000-10	US	1,225,584
DIRECT_SALES	2000-10	ALL	2,451,169
DIRECT_SALES	ALL	ALL	4,886,784
ALL	ALL	ALL	5,821,739

A ROLL-UP QUERY PRODUCES…

… the following sets of rows:

- Regular aggregation rows that would be produced by GROUP BY without using ROLLUP.
- First-level subtotals aggregating across country_id for each combination of channel_desc and calendar_month.
- Second-level subtotals aggregating across calendar_month_desc and country_id for each channel_desc value.
- A grand total row.

Again, note that the roll-up operation, unlike the cube, does not produce the rows with the "all" value only in one column but the last, while it still retains the row containing all "ALL" values.

→ It is a kind of "progressive aggregation"

Typical DW dimensions

time: 730 days
stock houses: 300
products: 30.000
daily sales: 3.000
promotions: not more than one per product sold
sales: 730 x 300 x 3000 x 1 = 657 millions.
dimensions: 657 millions x 8 attributes
x 4 byte = 21gb.

References

- Stefano Rizzi : Data Warehouse Design: Modern Principles and Methodologies McGraw-Hill, 2009
- M. Golfarelli, S. Rizzi: Data Warehouse: teoria e pratica della progettazione McGraw-Hill, 2002.
- Matteo Golfarelli: Data Warehouse Life-Cycle and Design. Encyclopedia of Database Systems 2009: 658-664
- Stefano Rizzi: Business Intelligence. Encyclopedia of Database Systems 2009: 287-288
- Ralph Kimball: The Data Warehouse Toolkit: Practical Techniques for Building Dimensional Data Warehouses John Wiley 1996.
- On the Internet: Oracle® Database Data Warehousing Guide (see link on the course page)