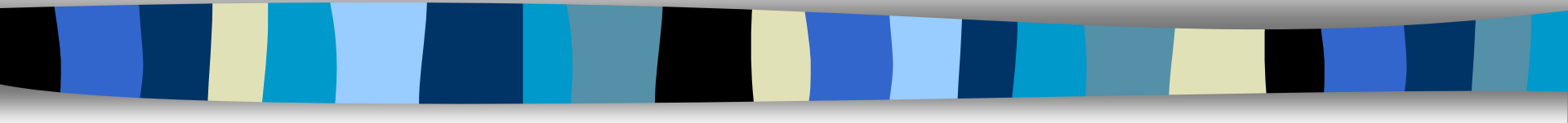


Introduction to Business Intelligence



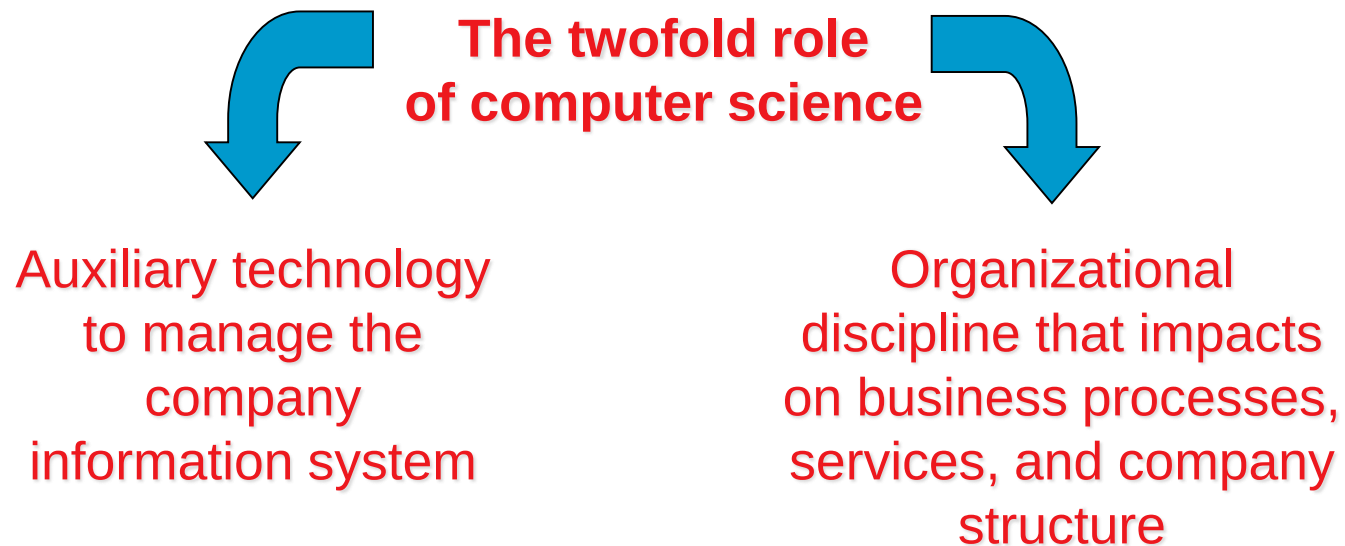
ICT in companies

- Up to some years ago, the main goal of databases in companies has been that of storing **operational data**, i.e., data generated by operations carried out within business processes
- Computer science was seen as a **subsidiary discipline** that makes information management faster and cheaper, but does not create profits in itself



The evolution of information systems

- The role of computer science in companies has radically changed since the early 70's. ICT systems turned from simple **tools to improve process efficiency** into **key factors of company organizations** capable of deeply impacting on the structure of business processes



The new role of computer science in decision making

- An **exponential increase** in operational data has made computers the only tools suitable for providing data for decision-making performed by business managers
- The massive use of techniques for analyzing enterprise data made information systems a **key factor to achieve business goals**

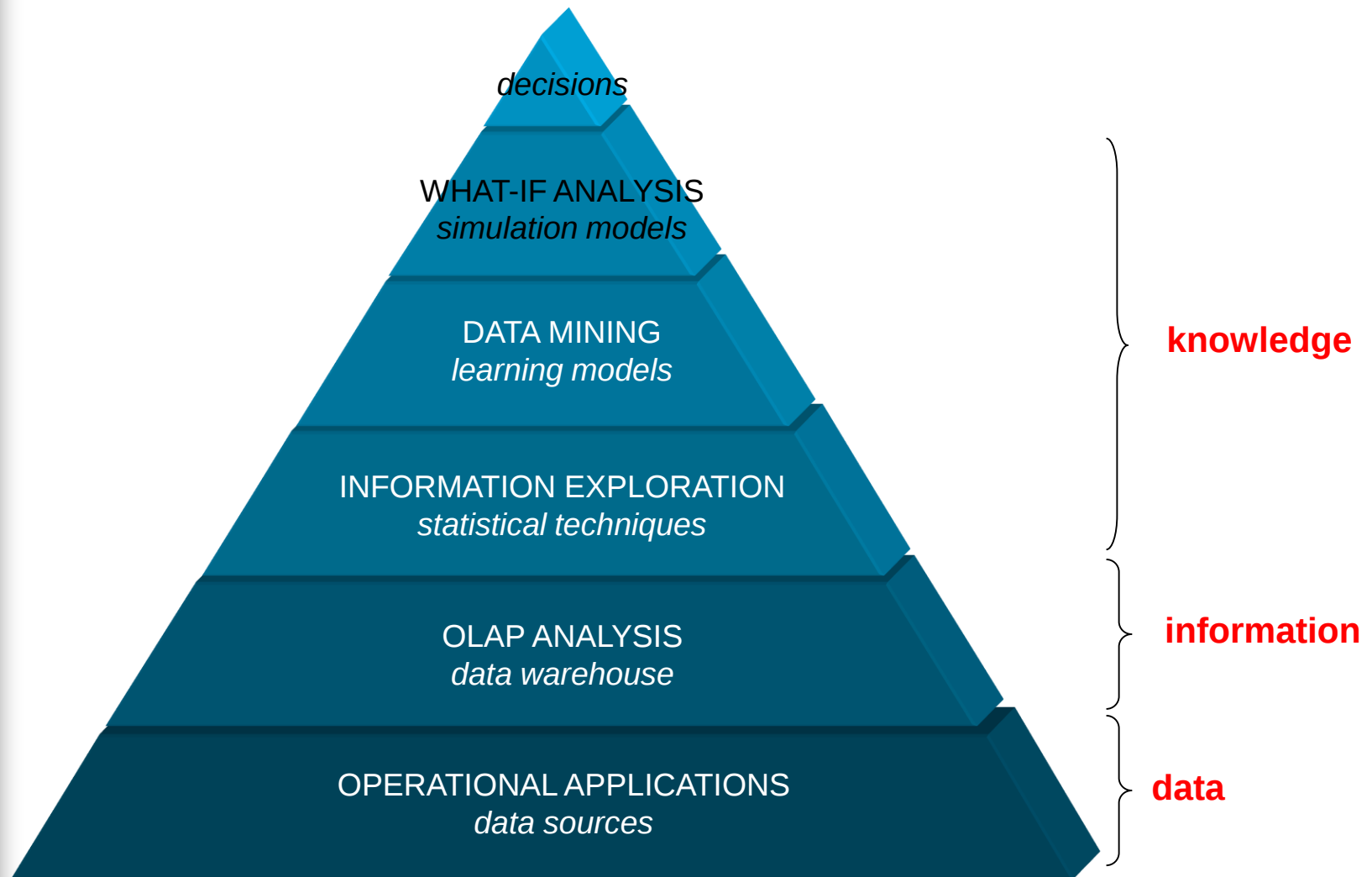


Business intelligence

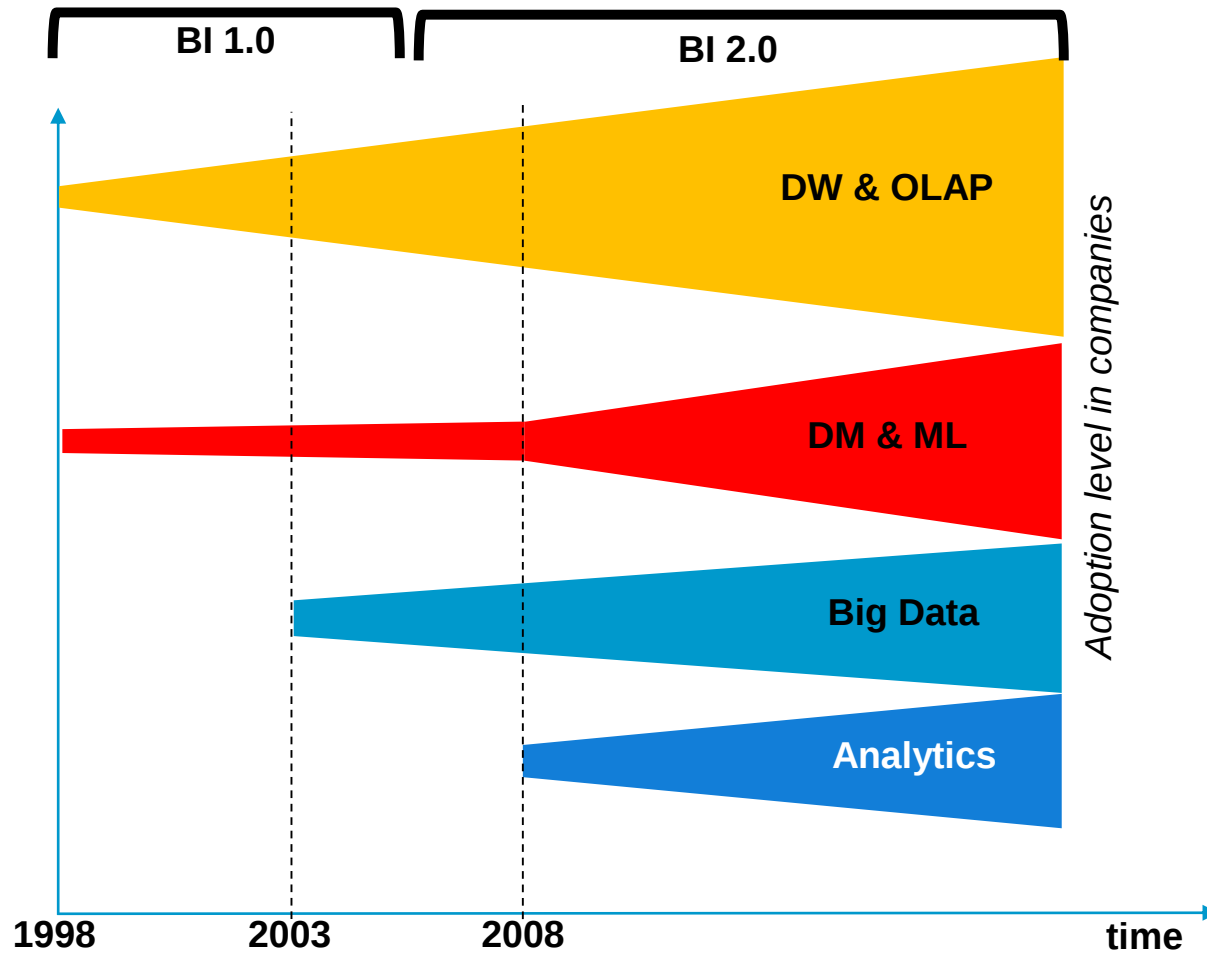


- A set of tools and techniques that enable a company to transform its business data into timely and accurate **information for the decisional process**
 - ✓ Business intelligence systems are used by decision makers to **get a comprehensive knowledge of the business** and of the factors that affect it, as well as to **define and support their business strategies**
 - ✓ The goal is to enable data-based decisions aimed at gaining competitive advantage, improving operative performance, responding more quickly to changes, increasing profitability and, in general, **creating added value for the company**

The BI pyramid



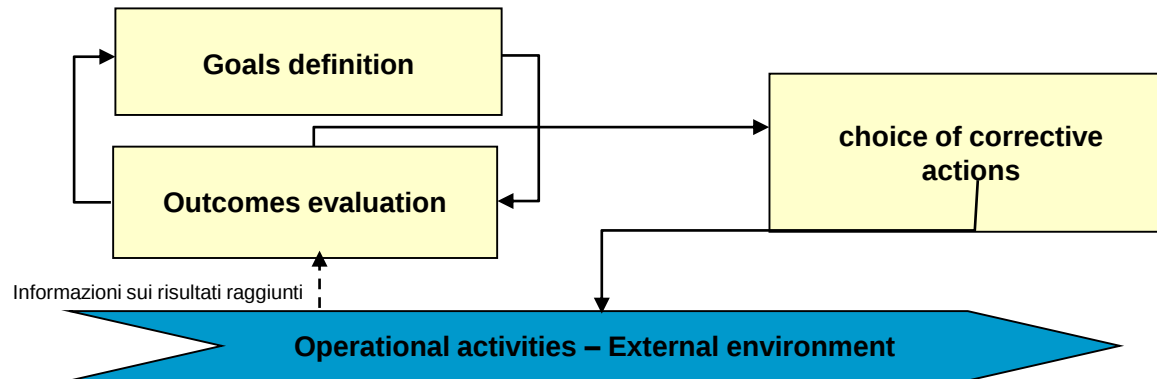
Spread of BI over time



Qualitative example

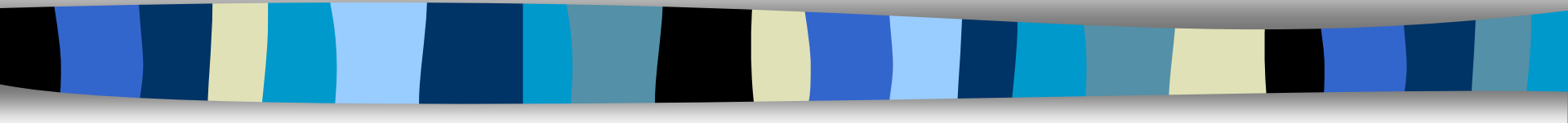
Decisional cycle in BI

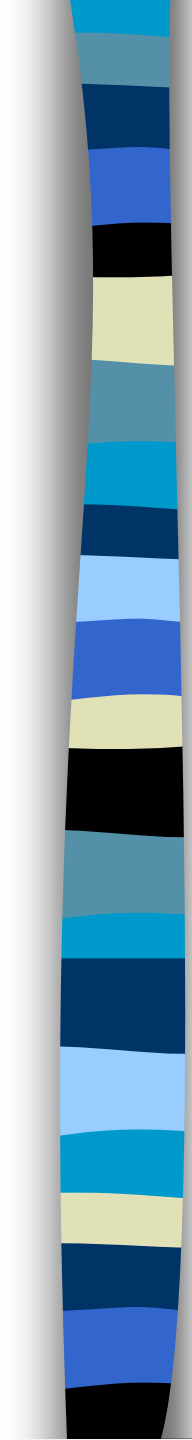
- They support the decisional process providing information to manager and knowledge worker
- Their reference model is the control loop:



- The managerial processes differ from the operational ones since:
 - Are based on indicators, that synthetic and aggregated
 - Processing is periodical rather than continuous
 - They rely on operational IS since they extract data from them

Data Warehousing





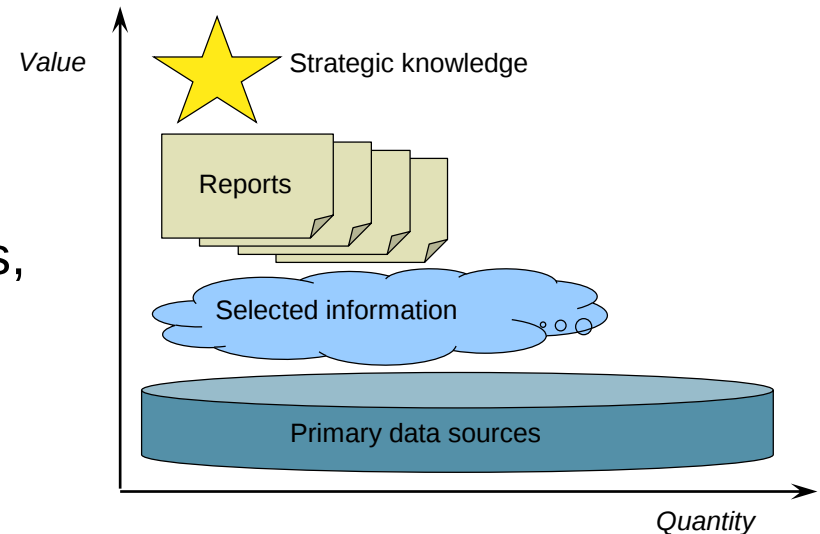
From data to information

- Information assets are immensely valuable to any enterprise, and because of this, these assets must be properly stored and readily accessible when they are needed
- However, the availability of too much data makes the extraction of the most important information difficult, if not impossible

~~data = information~~

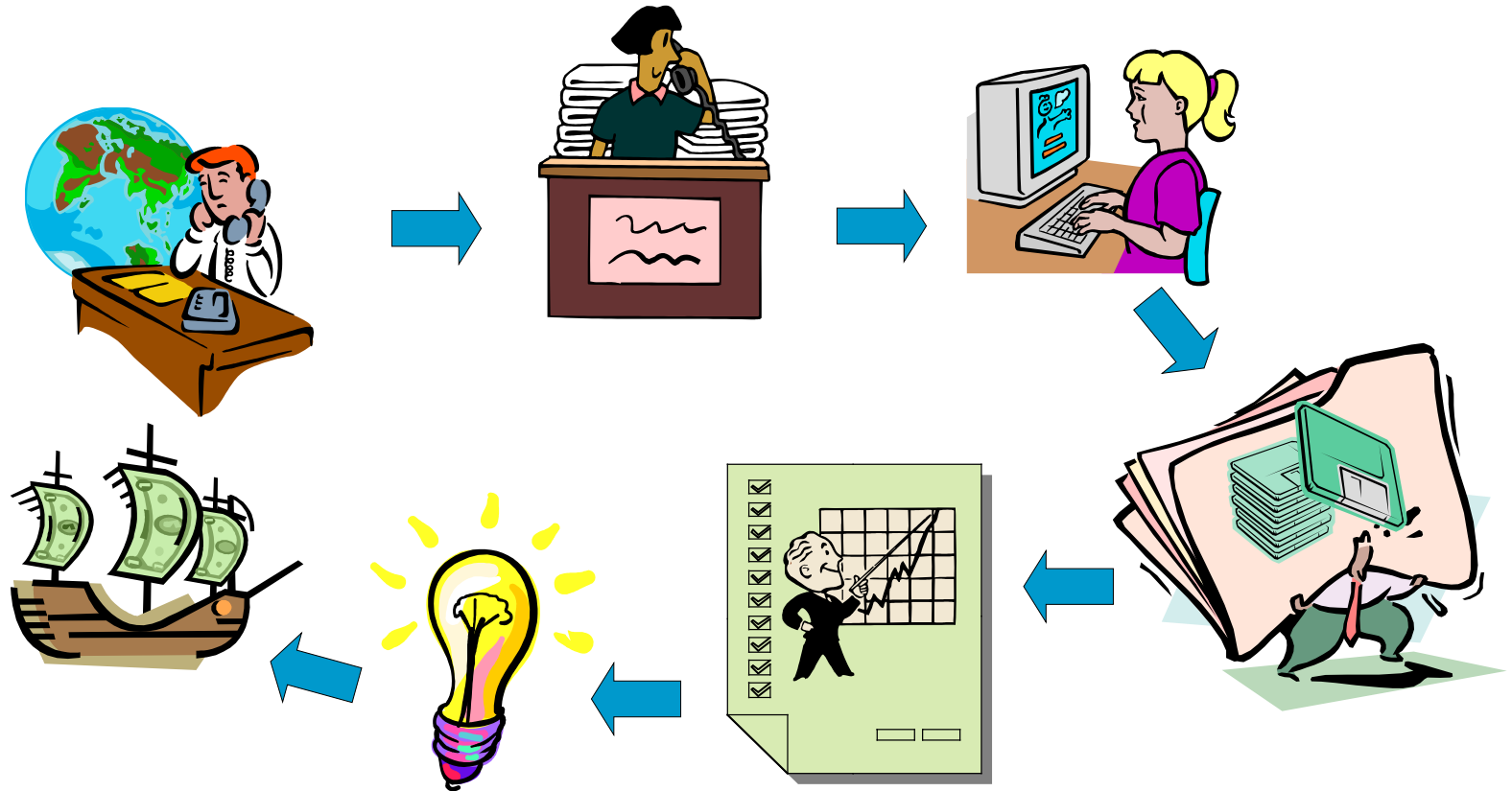
From data to information

- A company must have quick access to all the information needed for decision making
- Strategic knowledge is extracted from the huge amount of operational data stored in enterprise databases, through a selection and aggregation process



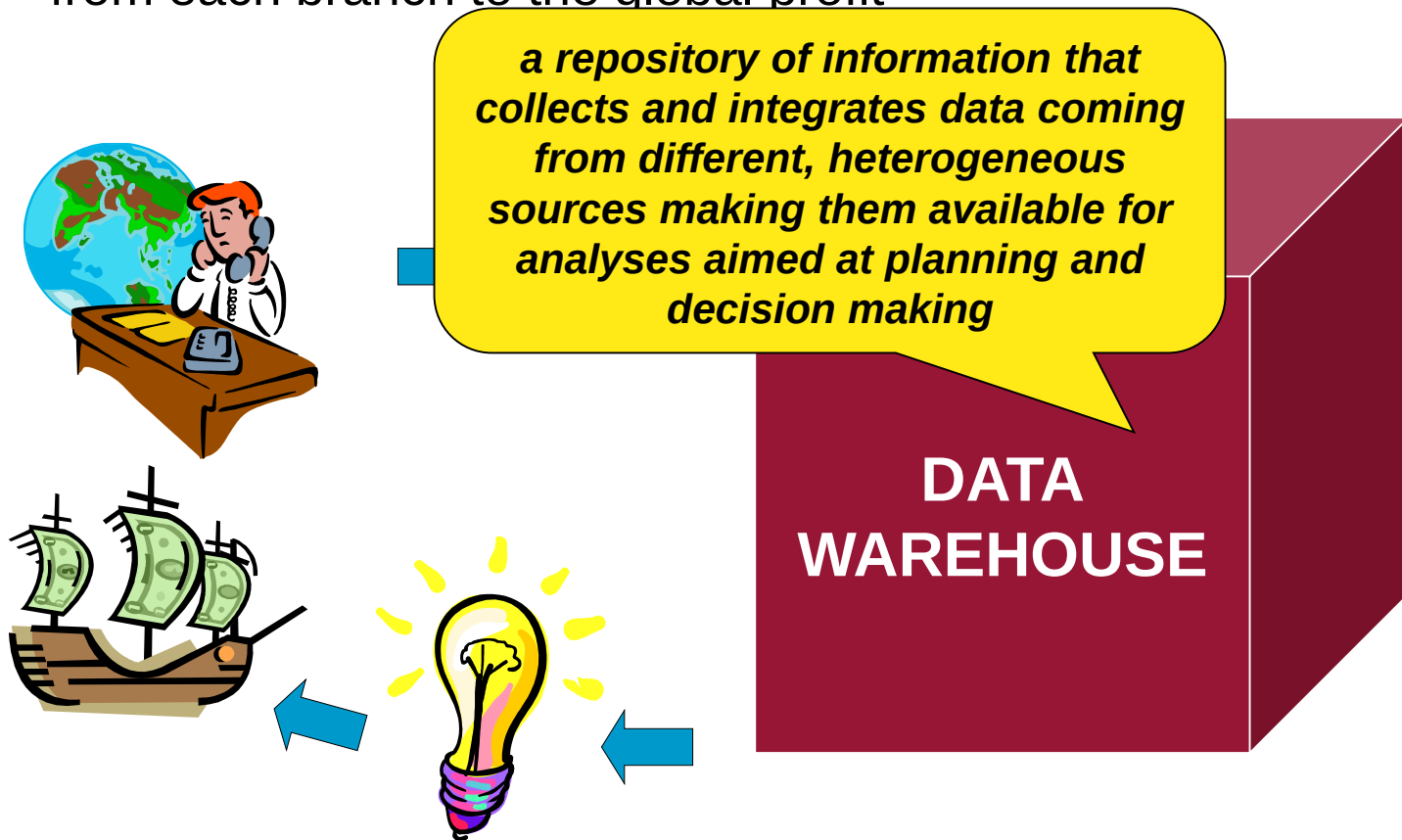
A typical scenario...

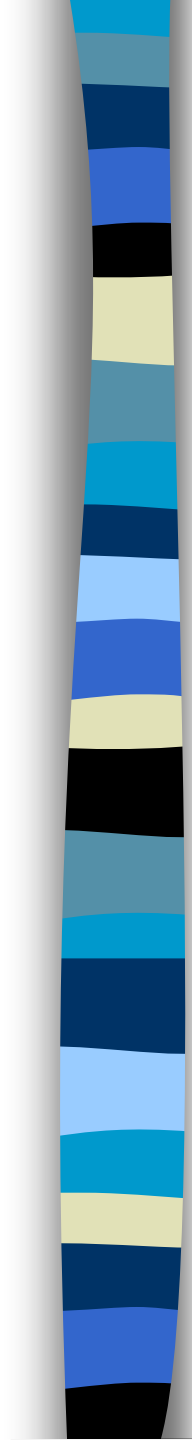
- ... is that of a large company, with several branches, whose managers wish to quantify and evaluate the contribution given from each branch to the global profit



A typical scenario...

- ... is that of a large company, with several branches, whose managers wish to quantify and evaluate the contribution given from each branch to the global profit





Queries

■ OLTP:

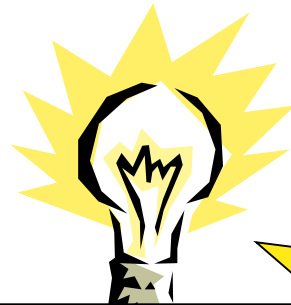
- ✓ They execute transactions that generally read/write a small number of tuples from/to many tables connected by simple relations
- ✓ The essential workload core is “frozen” in application programs, and ad hoc data queries are occasionally run for data maintenance

■ OLAP:

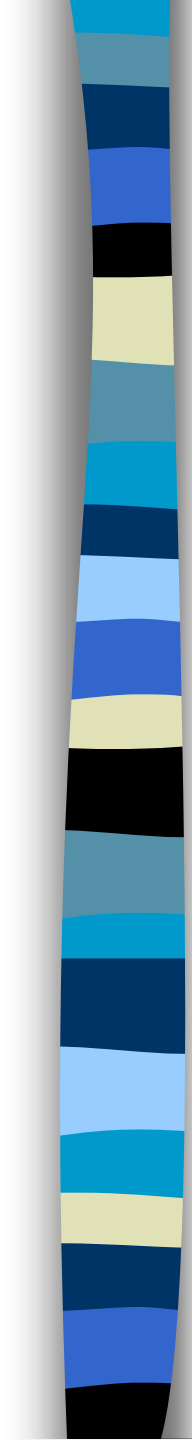
- ✓ Dynamic, multidimensional analyses that need to scan a huge amount of records to process a set of numeric data summing up the performance of an enterprise
- ✓ Interactivity is an essential property for analysis sessions, so the actual workload constantly changes as time goes by

OLTP and OLAP

- A mix of analytical queries with transactional routine queries inevitably slows down the system, and this does not meet the needs of users of both types of queries

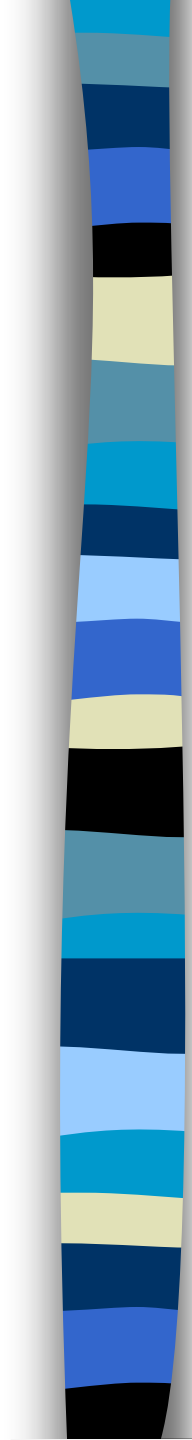


separate *online analytical processing (OLAP)* from *online transactional processing (OLTP)* by creating a new repository that integrates data from various sources and then makes data available for analysis and evaluation aimed at decision-making processes



Some fields of application

- **Trade** (sales and claims analyses, shipment and inventory control, customer care and public relations)
- **Craftsmanship** (production cost control, supplier and order support)
- **Financial services** (risk analysis and credit cards, fraud detection)
- **Transport industry** (vehicle management)
- **Telecommunication services** (call flow analysis and customer profile analysis)
- **Health care service** (patient admission and discharge analysis and bookkeeping in accounts departments)
-



Data Warehousing:

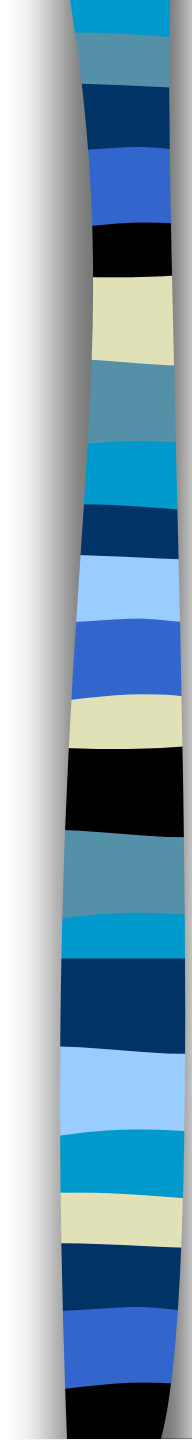
- A collection of methods, techniques, and tools used to support *knowledge workers* (senior managers, directors, managers, and analysts) to conduct data analyses that help with performing decision-making processes and improving information resources

User claims

- ✎ *We have heaps of data, but we cannot access them!*
- ✎ *How can people playing the same role achieve substantially different results?*
- ✎ *We want to select, group, and manipulate data in every possible way!*
- ✎ *Show me just what matters!*
- ✎ *Everyone knows that some data is wrong!*

R. Kimball, The Data Warehouse Toolkit





Requirements for the warehousing process

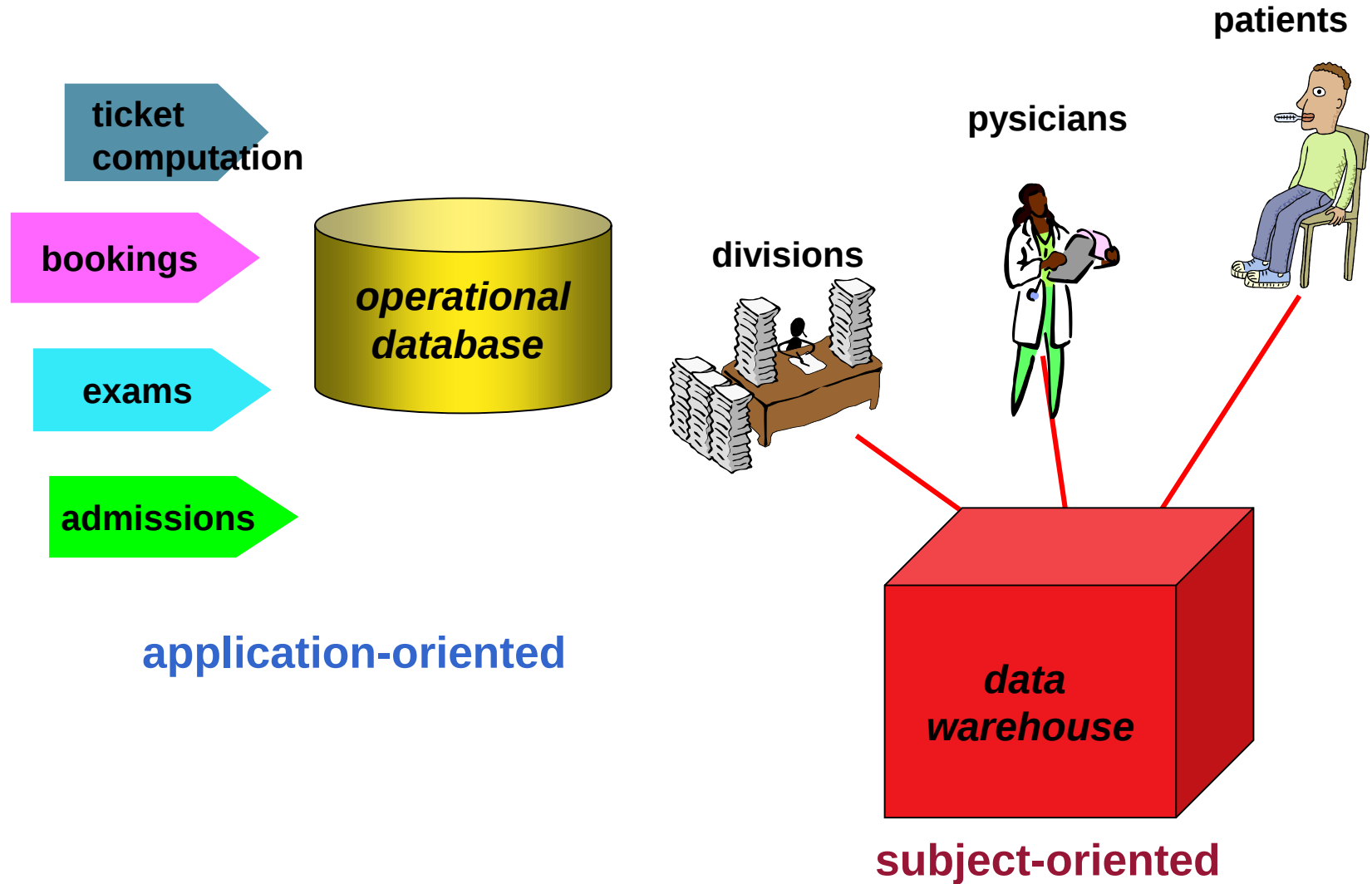
- **accessibility** to users not very familiar with IT and data structures
- **integration** of data on the basis of a standard enterprise model
- **query flexibility** to maximize the advantages obtained from the existing information
- **information conciseness** allowing for target-oriented and effective analyses
- **multidimensional representation** giving users an intuitive and manageable view of information
- **correctness and completeness** of integrated data

The Data Warehouse

- In the middle of this process, a data warehouse is a data repository that fulfills the requirements

- *A **data warehouse** is a collection of data that supports decision-making processes. It provides the following features:*
 - ✓ *It is subject-oriented;*
 - ✓ *It is integrated and consistent;*
 - ✓ *It shows its evolution over time and it is not volatile*

...subject-oriented



...integrated and consistent

Data warehouses take advantage of multiple data sources, such as data extracted from production and then stored to enterprise databases, or even data from a third party's information systems. A data warehouse should provide a unified view of all the data.



...shows its temporal evolution

Operational DB



**Limited historical content,
time is often not part of the
keys, data are updated**

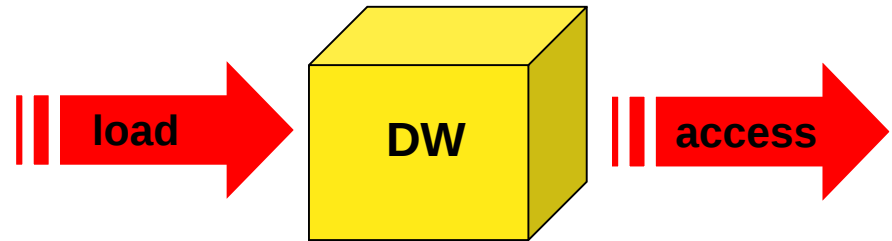
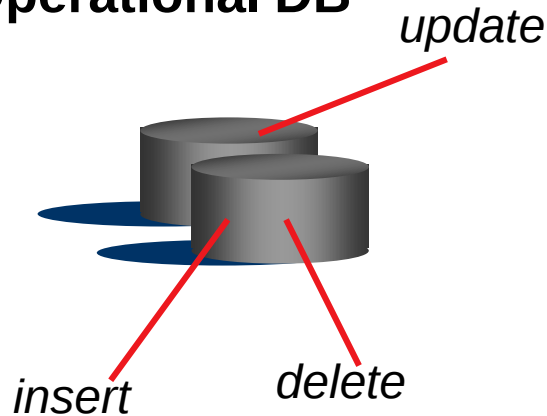
DW



**Rich historical content,
time is part of the keys,
a snapshot of data taken at a
given time cannot be updated**

...non volatile

Operational DB

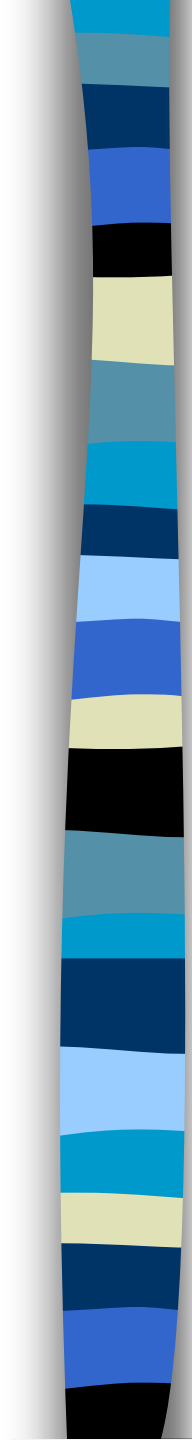


Huge data volumes:
from 50 GBs to some TBs
in a few years

- ✓ no need for advanced transaction management techniques required by operational applications
- ✓ key problems are query-throughput and resilience

Summarizing

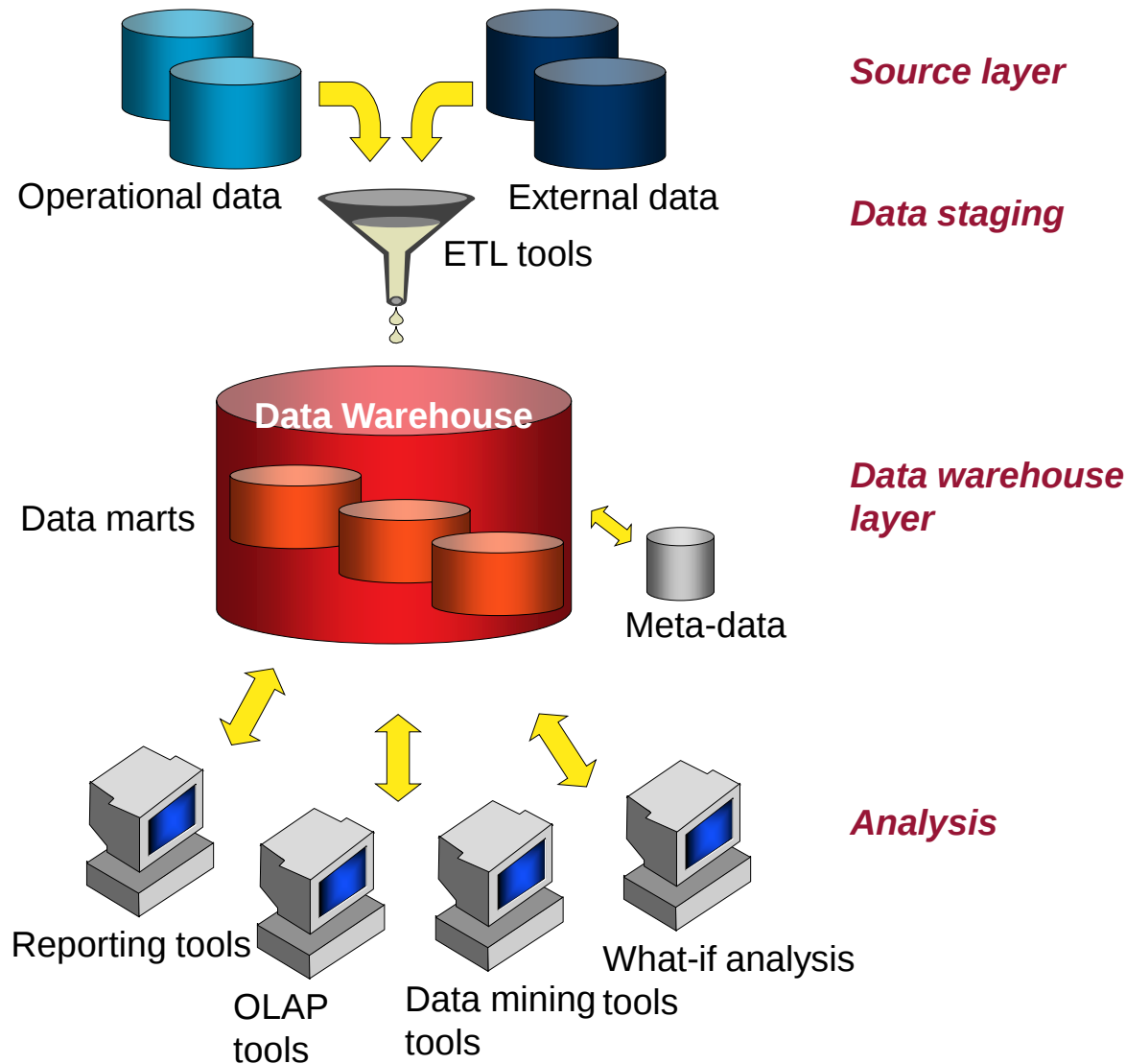
	<i>Operational DBs</i>	<i>Data warehouses</i>
users	thousands	hundreds
workload	predefined transactions	<i>ad hoc</i> analysis queries
access	to hundreds of records, read and write	to millions of records, mostly read-only
goal	application-dependent	decision support
data	elementary, numeric and alphanumeric	aggregated, mostly numeric
data integration	application-based	subject-based
quality	in terms of integrity	in terms of consistency
temporal span	current data	current and historical data
update	continuous	periodic
model	normalized	multidimensional
optimization	for OLTP accesses on a fraction of database	for OLAP accesses on a large part of database



Architectural requirements

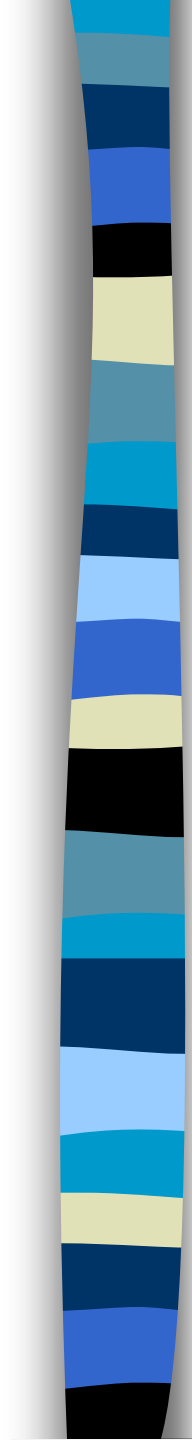
- ✓ **Separation** Analytical and transactional processing should be kept apart as much as possible
- ✓ **Scalability** Hardware and software architectures should be easy to upgrade as the data volume, which has to be managed and processed, and the number of users' requirements, which have to be met, progressively increase
- ✓ **Extendibility** The architecture should be able to host new applications and technologies without redesigning the whole system
- ✓ **Security** Monitoring accesses is essential because of the strategic data stored in data warehouses
- ✓ **Administerability** Data warehouse management should not be overly difficult

Two-layer architectures



DATA MART:

A subset or an aggregation of the data stored to a primary data warehouse. It includes a set of information pieces relevant to a specific business area, corporate department, or category of users.

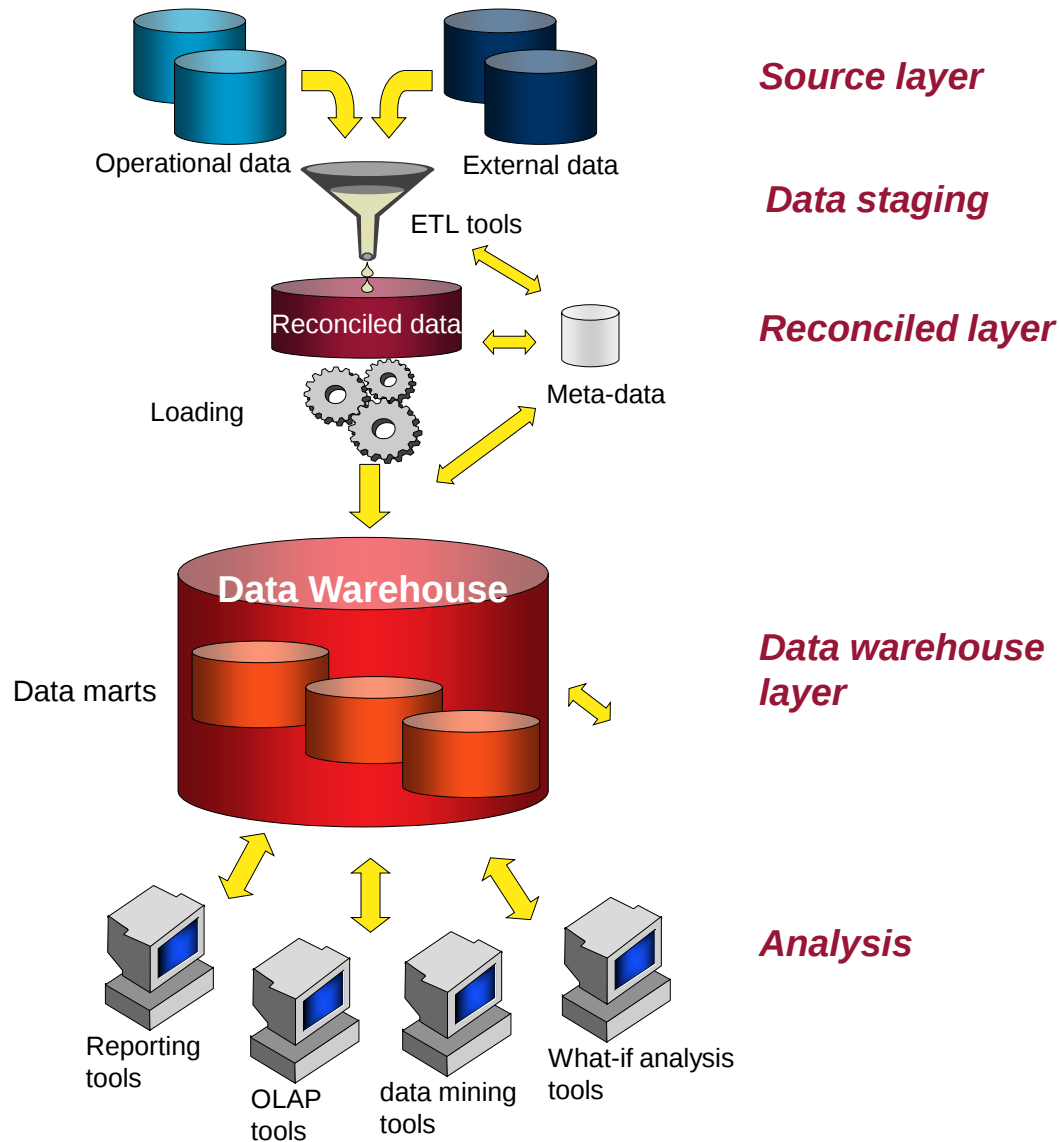


Two-layer architectures

■ Pros:

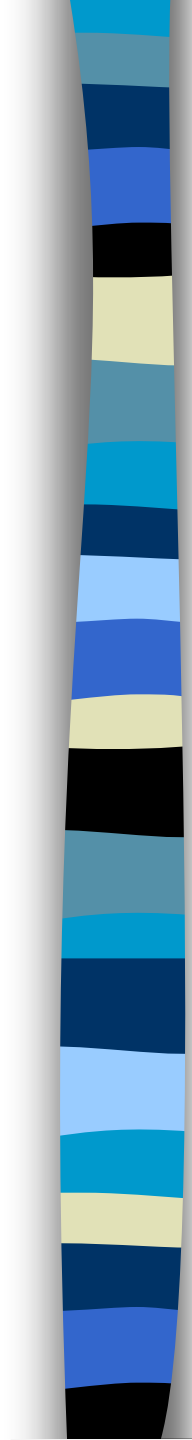
- ✓ At the level of the warehouse a quality information is continuously available even when, for technical or organizational issues, access to the sources is temporarily denied
- ✓ The analytical queries performed on the DW does not interfere with the handling of transactions at the operational level, whose reliability is essential to business operations
- ✓ The logical organization of the DW is based on the multidimensional model, while the sources typically offer relational models
- ✓ There is temporal and granularity discrepancy between OLTP systems, handling current at the maximum level of detail, and OLAP systems that operate on historical and summarized data
- ✓ At the DW level you can use specific techniques to optimize applications for performance analysis and reporting

Three-layer architectures



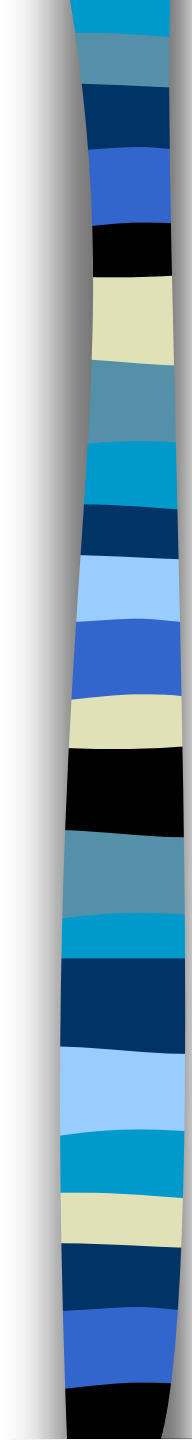
RECONCILED DATA:

operational data obtained after integrating and cleansing source data. As a result, those data are integrated, consistent, appropriate, current, and detailed



Three-layer architectures

- The main advantage of the reconciled data level is that it creates a common data model and reference for the whole company, while introducing a clear distinction between issues related to the extraction and integration of data from sources and those inherent the DW feeding
- On the other hand, reconciled data introduce additional redundancy compared with source operational data

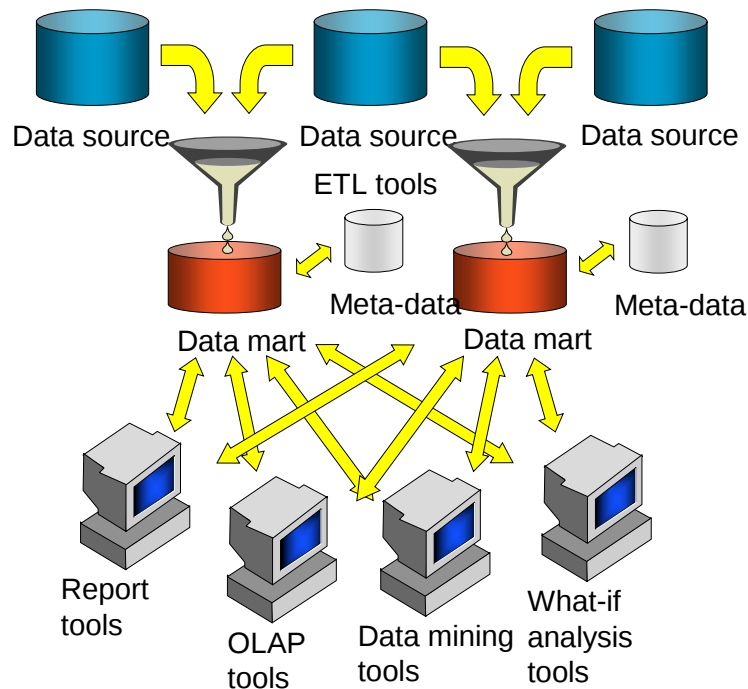


Architectures: another classification

- Data mart independent
- Data mart bus
- Hub-and-spoke
- Centralized data warehouse
- Federation

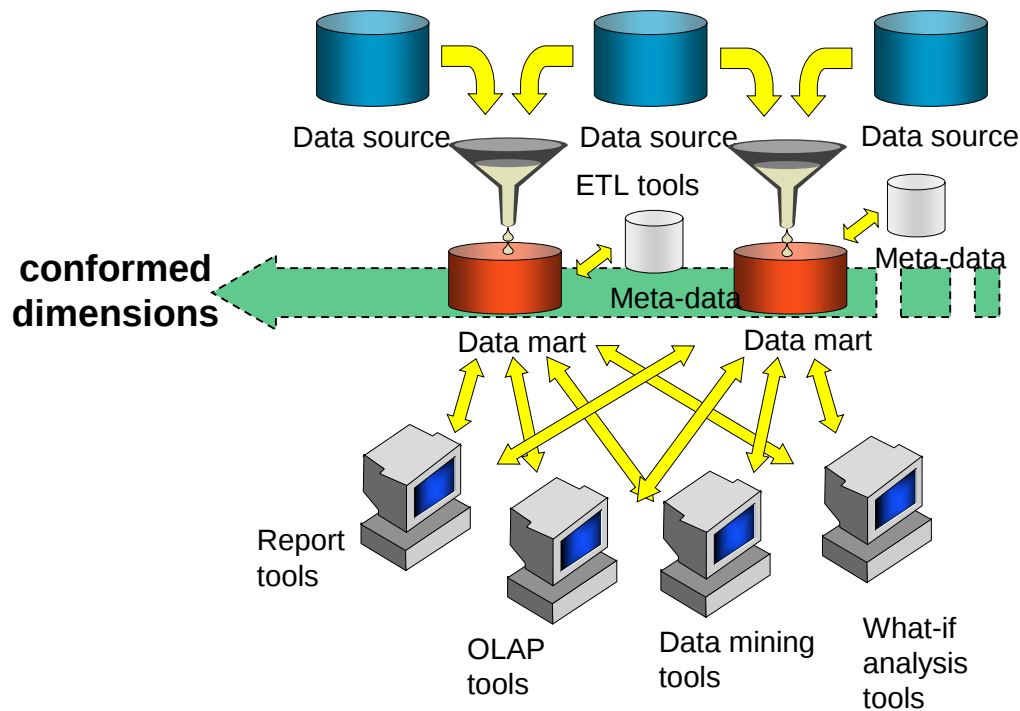
Data mart independent

- First approach to data warehousing
- Inconsistency issues



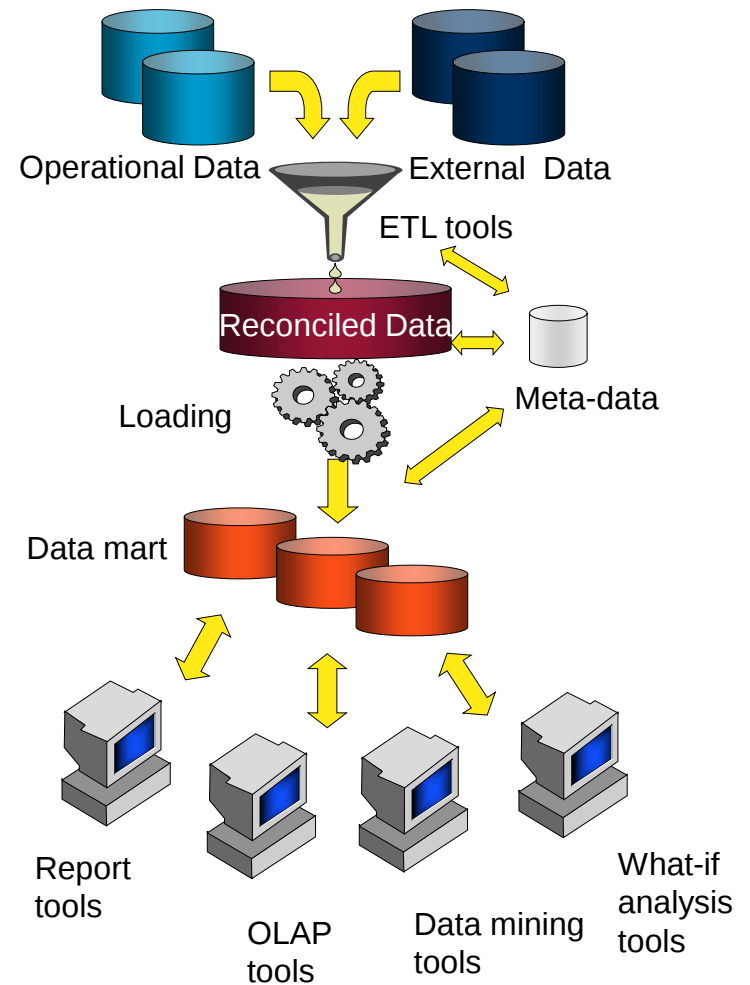
Data mart bus

- Approach suggested by [Kimball](#)
- Logical level integration
- “Enterprise view”



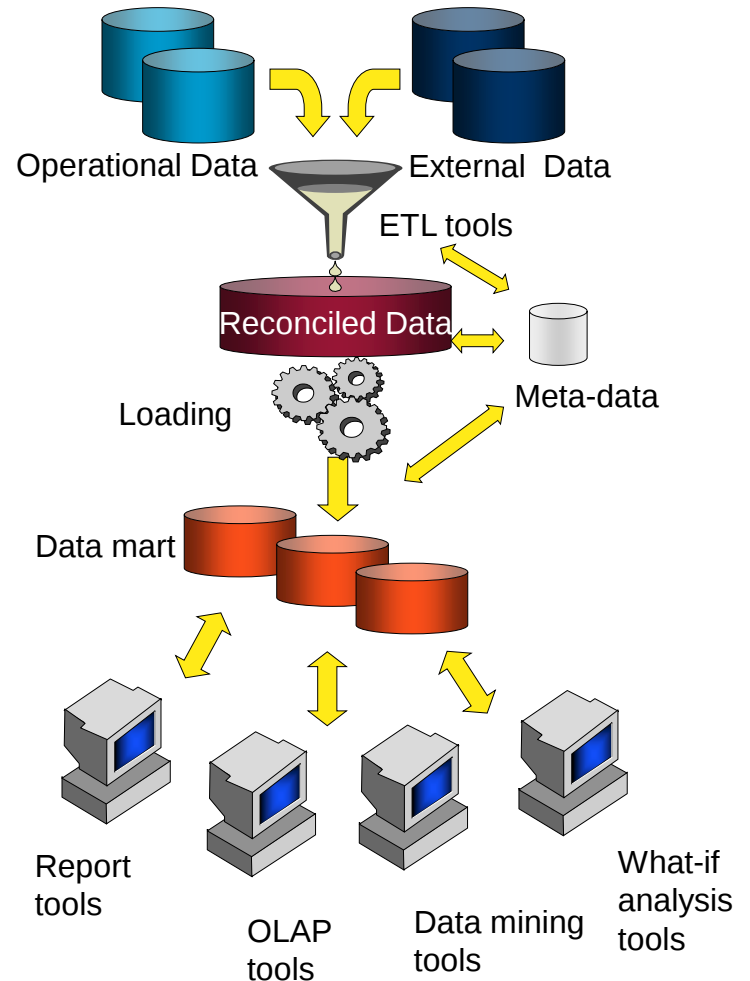
Hub-and-spoke

- One of the most used architectures in medium to large environments



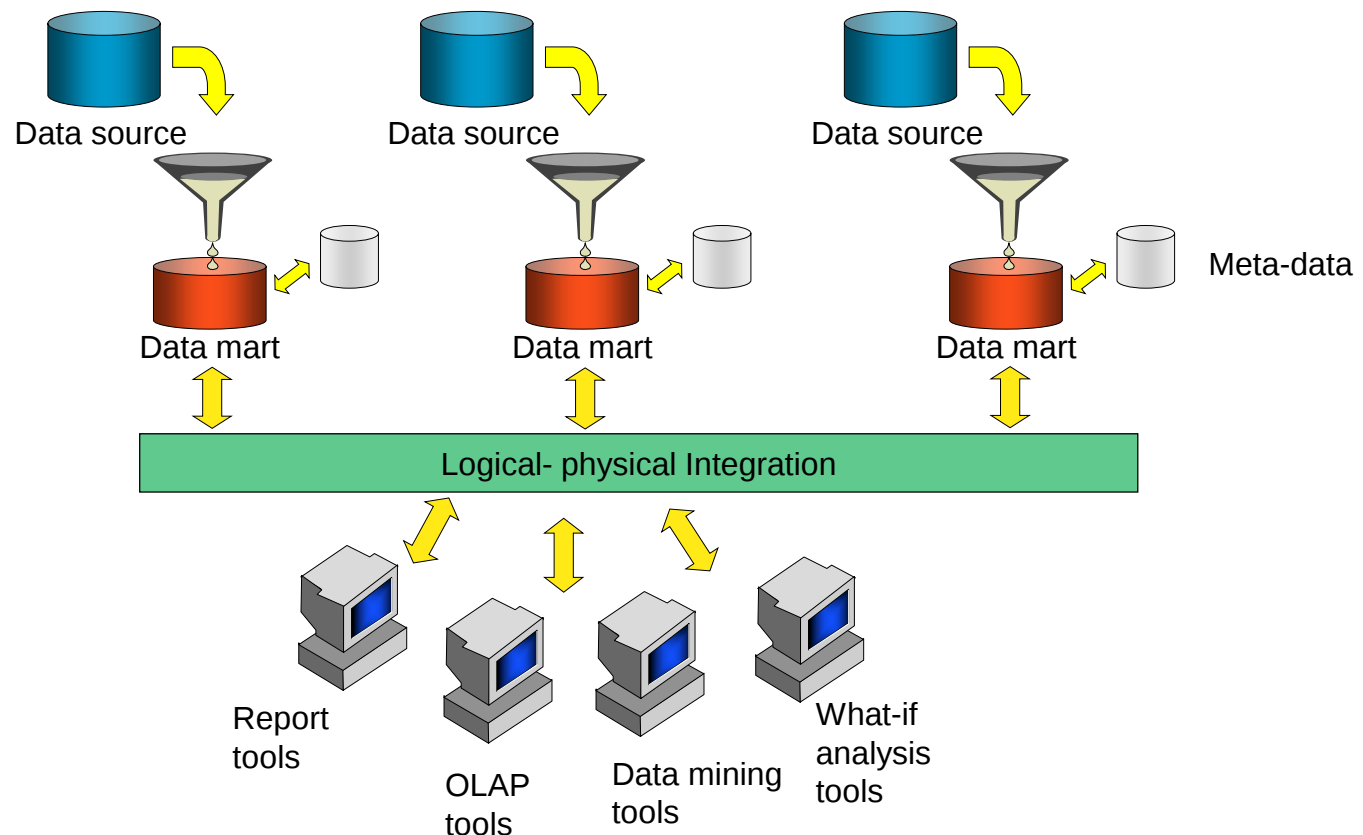
Centralized Data warehouse

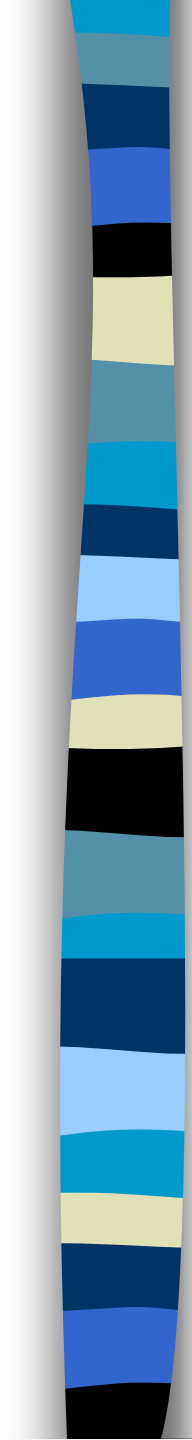
- Approach suggested by [Inmon](#)
- Logical, more than physical, implementation of a hub-and-spoke architecture



Federation

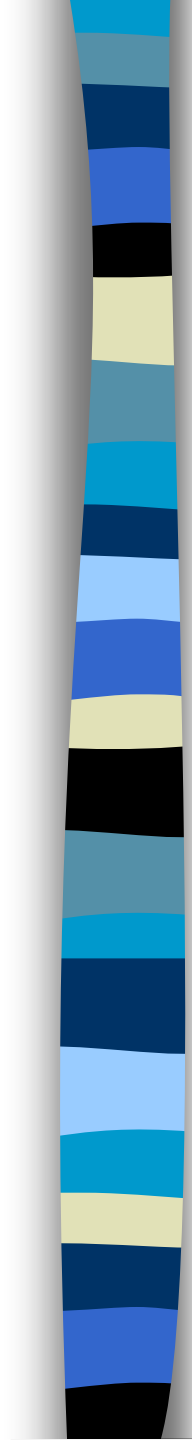
- Good for highly dynamic environments (mergers and acquisitions)
- Problem of effective and efficient integration





Choosing between architectures

- Information interdependence among organizational units in company
 - ✓ encourages the adoption of enterprise-wide architectures
- Urgency of the data warehousing project
 - ✓ encourages the adoption of “fast” architectures
- Constraints on economic and human resources
- Role of the project within the business strategy
 - ✓ independent data marts vs. hub-and-spoke
- Compatibility with existing platforms
- Skills of the IT staff
- Organizational position of the sponsor of the project
 - ✓ enterprise architectures vs. departmental architectures



ETL

- ETL (*Extraction, Transformation and Loading*) processes extract, integrate and clean data from operational sources to feed the data warehouse layer
- In a three-layer architecture, ETL processes actually feed the reconciled data layer—a single, detailed, comprehensive, top-quality data source—that in its turn feeds the data warehouse
- ETL takes place once when a data warehouse is populated for the first time, then it occurs every time that data warehouse is regularly updated
 - ✓ extraction
 - ✓ cleaning
 - ✓ transformation
 - ✓ loading

Extraction

- Relevant data are extracted from operational sources
 - ✓ *static extraction is used when a data warehouse is populated for the first time*
 - ✓ *Incremental extraction is used to update data warehouses regularly, and it seizes the changes applied to source since their latest extraction*



Cleaning

- The cleaning phase is supposed to improve data quality—normally quite poor in sources

- ✓ duplicate data
- ✓ inconsistent values that are logically associated
- ✓ missing data
- ✓ unexpected use of fields
- ✓ impossible or wrong data
- ✓ inconsistent values for a single entity because different practices were used
- ✓ inconsistent values for a single entity because of typing mistakes



Transformation

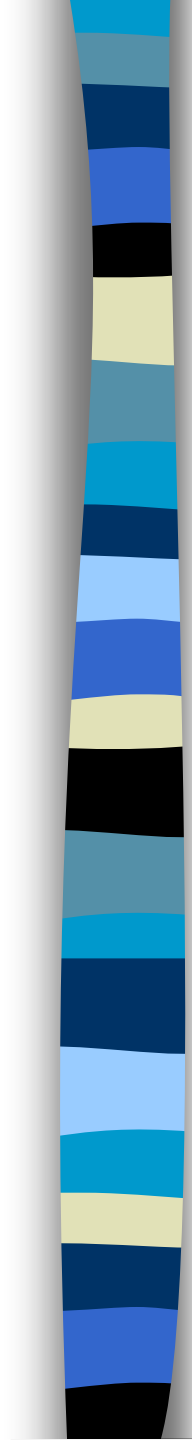
- It converts data from its operational source format into a specific data warehouse format
- Establishing a mapping between the source data layer and the data warehouse layer is generally made difficult by the presence of many different, heterogeneous sources



Loading

- It loads data into the DW
 - ✓ **Refresh:** data warehouse data is completely rewritten. This means that older data is replaced
 - ✓ **Update:** only those changes applied to source data are added to the data warehouse





Towards the multidimensional model

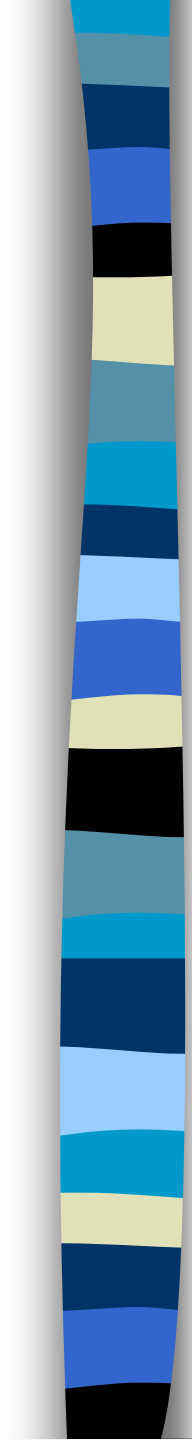
"What business were registered last year for each region and each product category?"

"What is the relationship between performance of shares of PC manufacturers and quarterly profits along the last 5 years?"

"What are the types of orders that will maximize revenues?"

"Which of two new therapies is more effective for decreasing the average length of hospitalizations?"

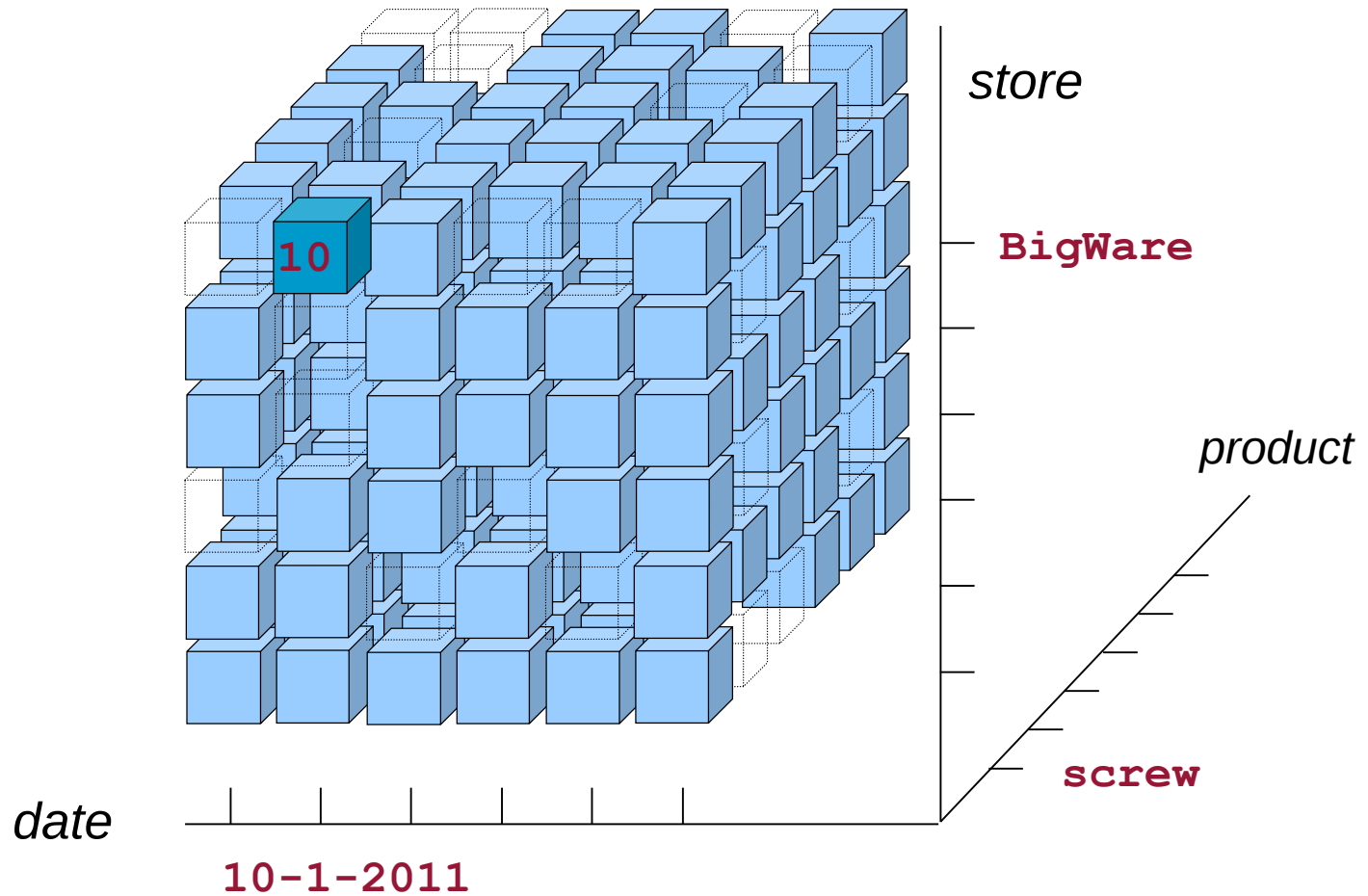
"What is the relationship between the profits made with shipments of less than 10 items and those made with shipments of more than 10 items?"



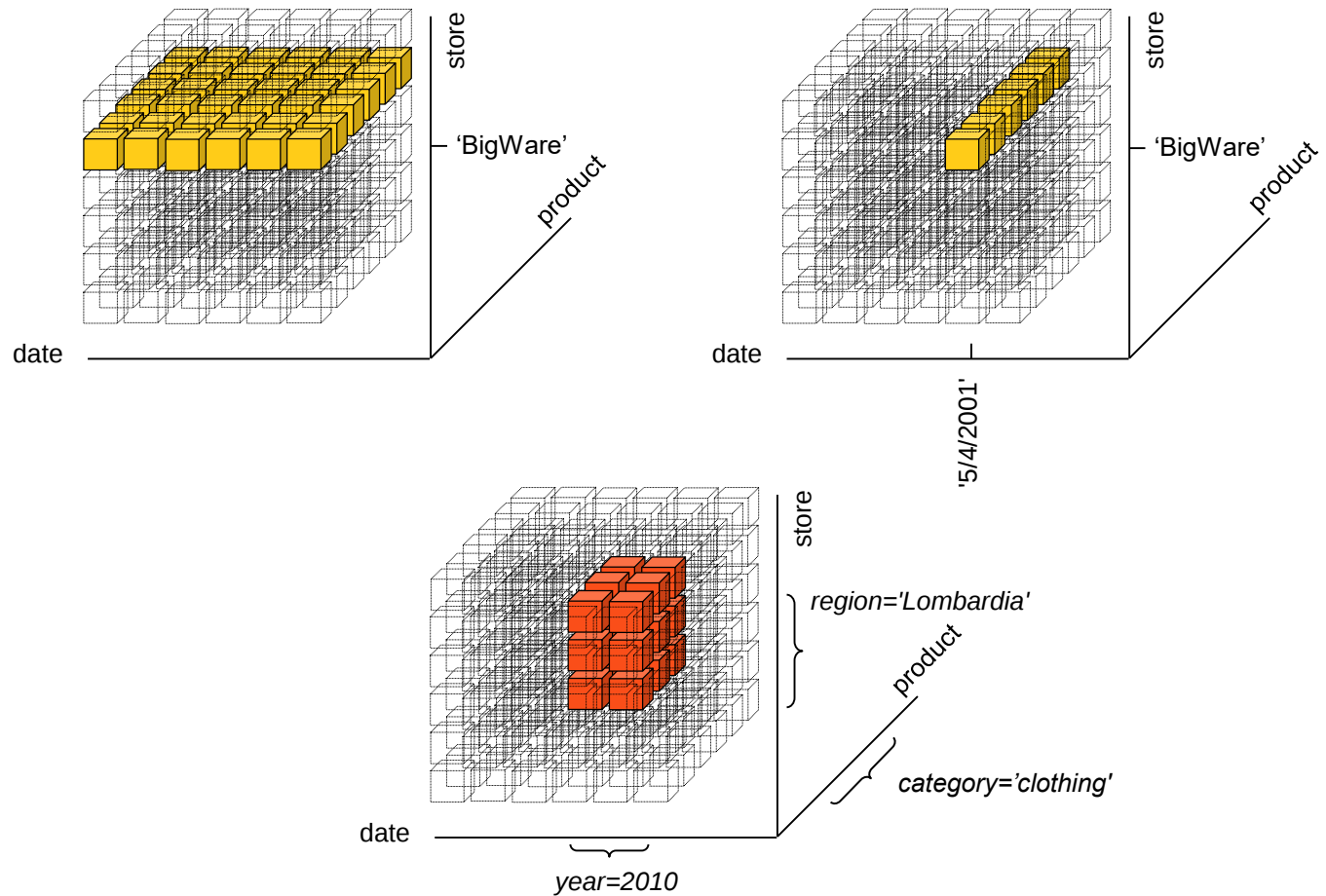
The multidimensional model

- It is the key for representing and querying information in a DW
- *Facts* of interest are represented in *cubes* where:
 - ✓ each cell stores numerical *measures* that quantify the fact from different points of view;
 - ✓ each axis is a *dimension* for analyzing measure values;
 - ✓ each dimension can be the root of a *hierarchy* of attributes used to aggregated measure values

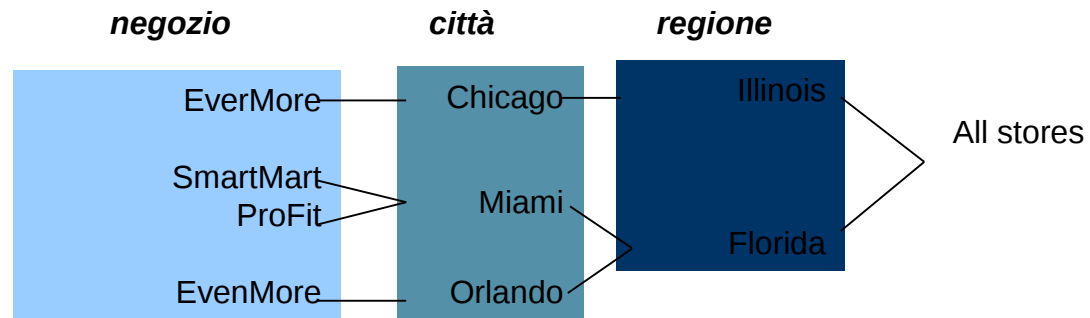
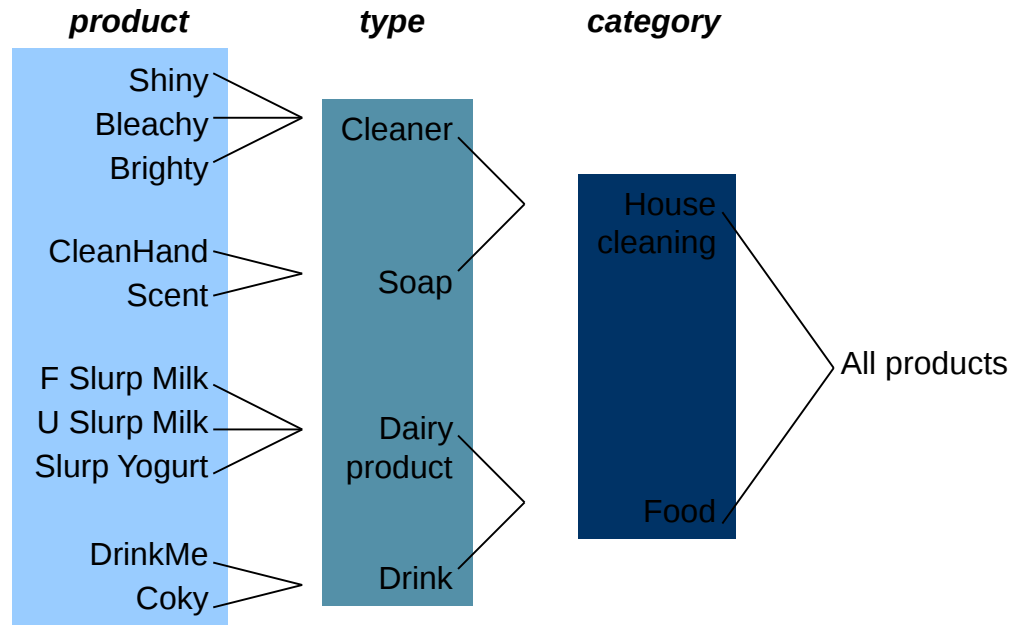
The Sales cube



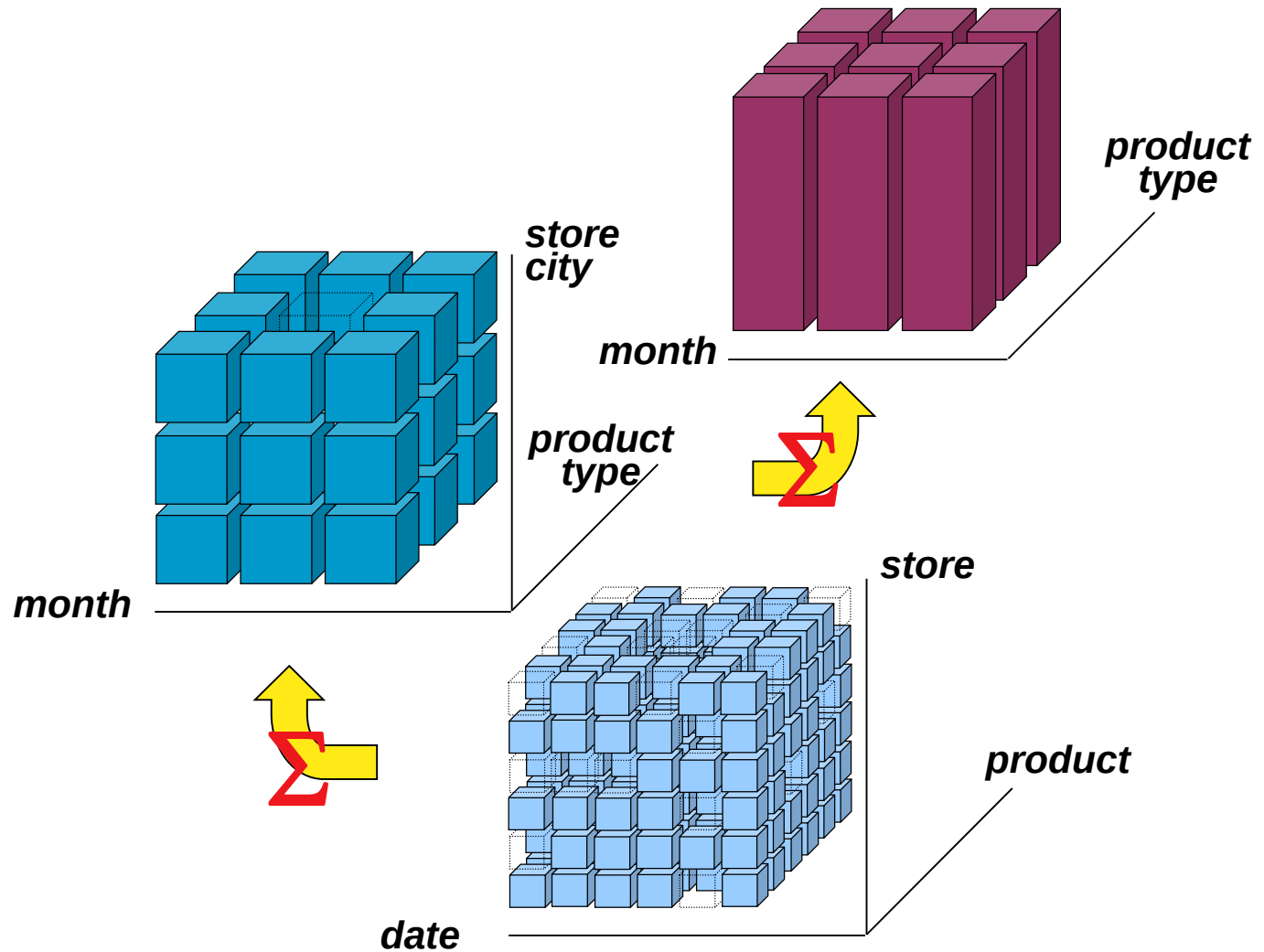
Slicing and dicing



Hierarchies

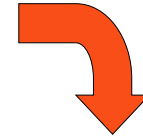


Aggregation

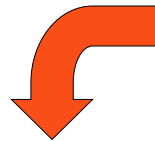


Aggregation

	BigWare1	BigWare2	NotOnlyJelly
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
.....			



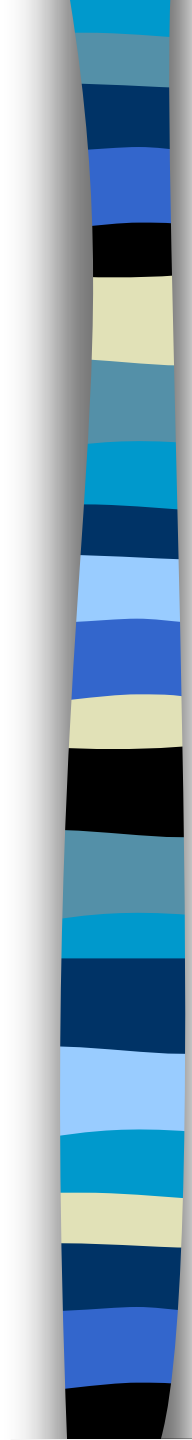
	BigWare1	BigWare2	NotOnlyJelly
January 2000	200	180	150
February 2000	180	150	120
March 2000	220	180	160
.....			
January 2001	350	220	200
February 2001	300	200	250
March 2001	310	180	300
.....			
January 2002	380	200	220
February 2002	310	200	250
March 2002	300	160	280
.....			



	BigWare1	BigWare2	NotOnlyJelly
2000	2400	2000	1600
2001	3200	2300	3000
2002	3400	2200	3200



	BigWare1	BigWare2	NotOnlyJelly
<i>Totale:</i>	9000	6500	7800



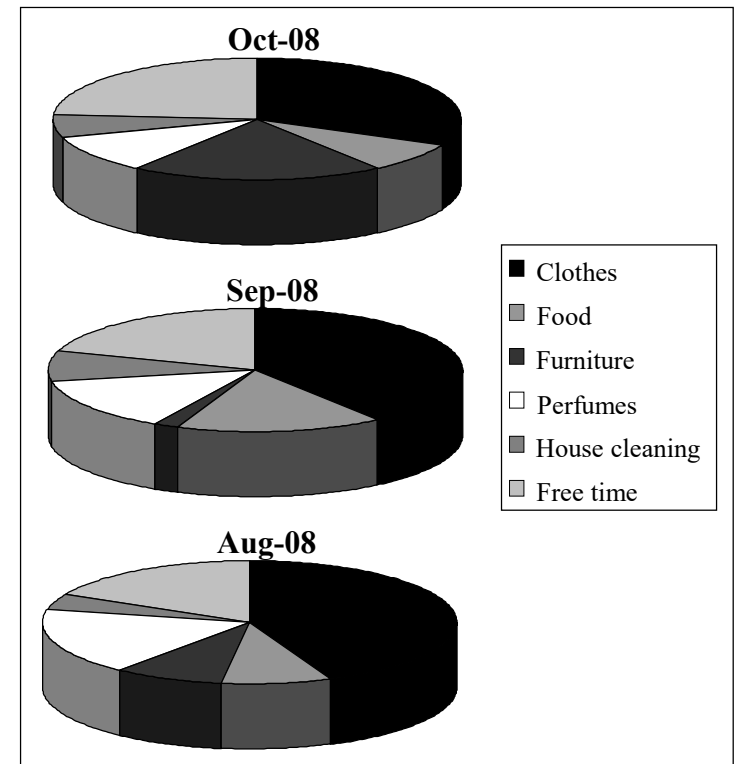
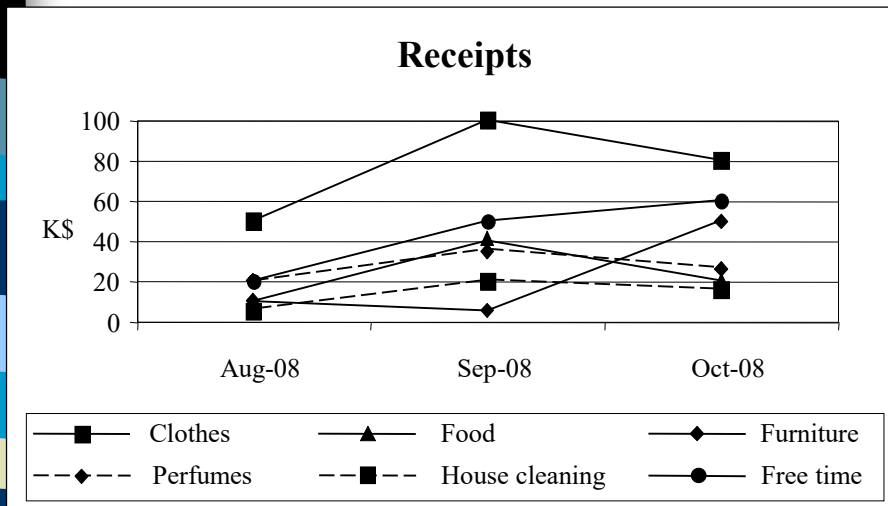
Data analysis techniques

- Once data have been cleaned, integrated, and transformed, users must be enabled to take maximum advantage from the resulting information assets
- There are two different approaches for querying data warehouses, supported by different types of tools:
 - ✓ *reporting*: no ICT skills required
 - ✓ *OLAP*: users must be able to reason according to the multidimensional paradigm, and they must be acquainted with the visual interface of the tool

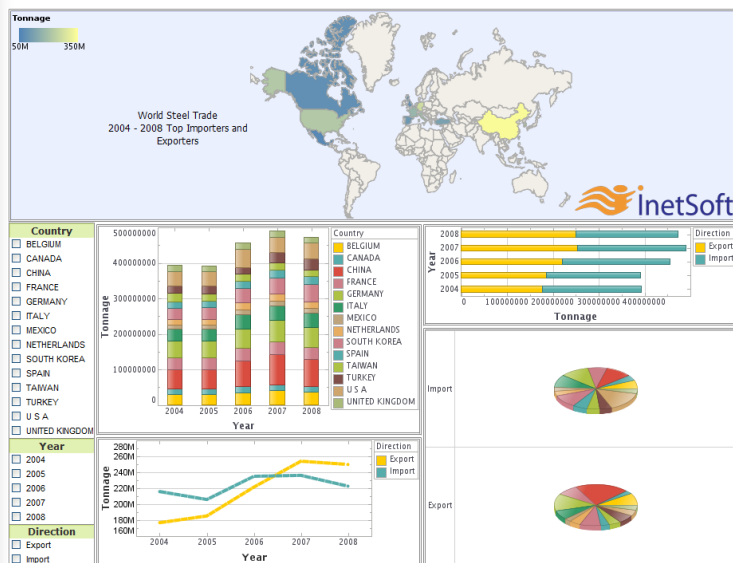
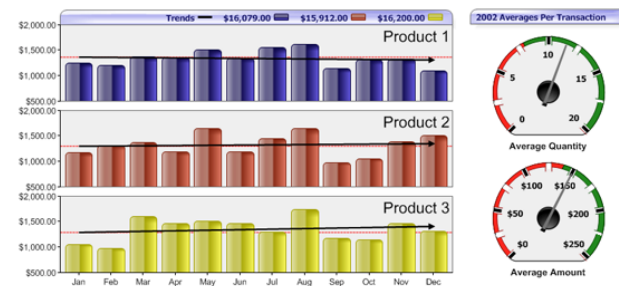
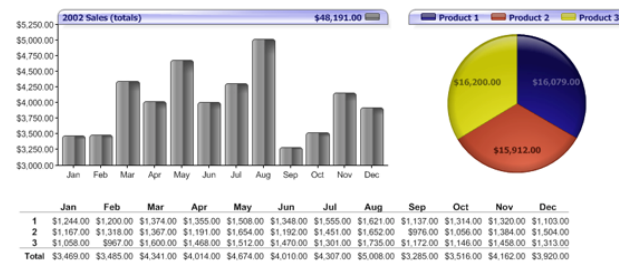
Reporting

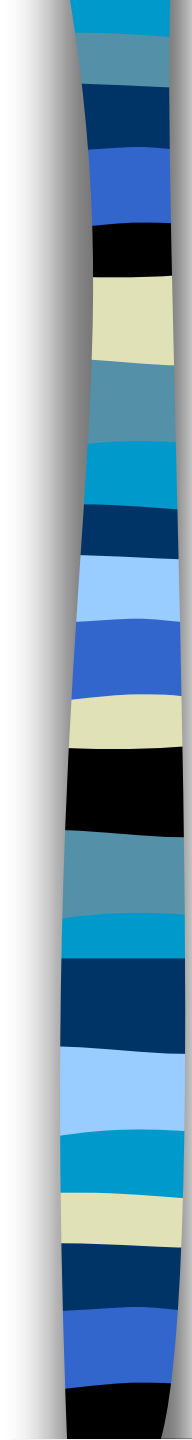
geared towards users
who need to regularly
access information
with fixed structure

Receipts (K\$)	Oct. 2008	Sep. 2008	Aug. 2008
Clothes	80	100	50
Food	20	40	10
Furniture	50	5	10
Perfumes	25	35	20
House cleaning	15	20	5
Free time	60	50	20



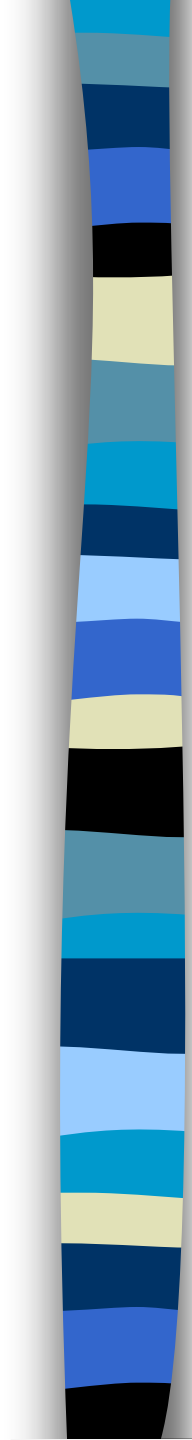
Reporting





OLAP

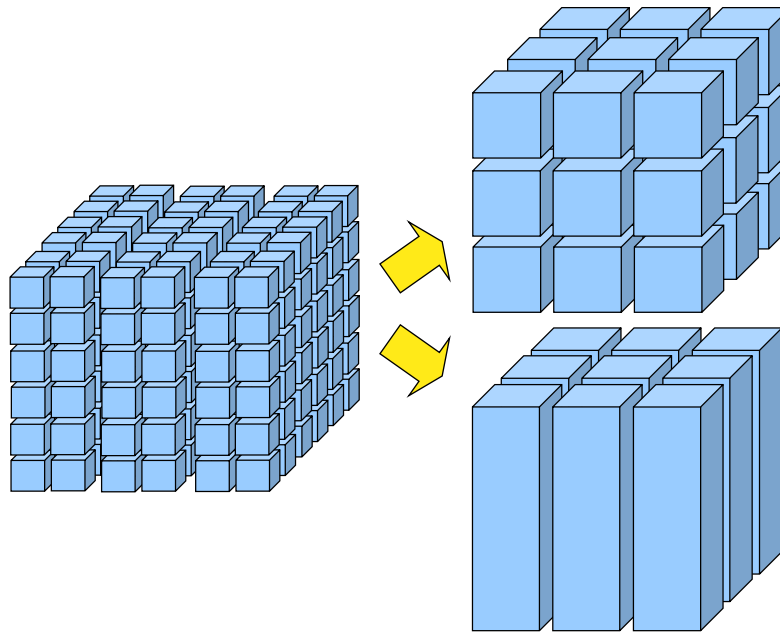
- OLAP is the **main way** to exploit information in a data warehouse
- It gives end-users, whose analysis needs are not easy to define beforehand, the opportunity to **analyze and explore data interactively** on the basis of the multidimensional model
- While users of reporting tools play a passive role, OLAP users are able to **actively** carry out a complex analysis session, where each step is the result of the outcome of preceding steps



OLAP session

- An OLAP session consists of a *navigation path* that corresponds to an analysis process for facts according to different viewpoints and at different detail levels. This path is turned into a sequence of queries, which are not issued directly, but differentially expressed with reference to the previous query
- Every step of an analysis session is characterized by an *OLAP operator* that turns the latest query into a new one
- The results of queries are multidimensional; OLAP tools typically use tables to display data, with multiple headers, colors, and other features to highlight data dimensions

OLAP operators



roll-up

OLAP operators

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	\$ 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	\$ 1.588	\$ 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	\$ 2.518	\$ 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	\$ 63
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	\$ 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	\$ 395	\$ 1.740	\$ 1.943	\$ 1.143	\$ 366	\$ 307	\$ 118

roll-up

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	\$ 1.676	\$ 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

OLAP operators

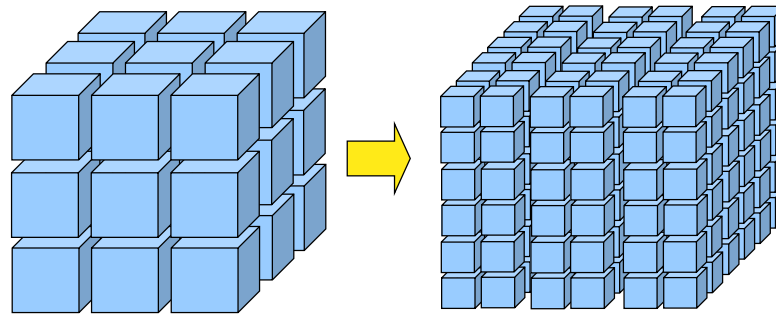
Category	Year	Metrics	Dollar Sales									
		Customer Region	North-East	Mid-Atlantic	South-East	Central	South	North-West	South-West	England	France	Germany
Electronics	1997		\$ 138	\$ 1,774	\$ 384	\$ 138	\$ 2,346	\$ 2,554	\$ 2,184	\$ 566	\$ 199	\$
	1998		\$ 1,184	\$ 4,529	\$ 1,892	\$ 7,232	\$ 651	\$ 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	\$

roll-up

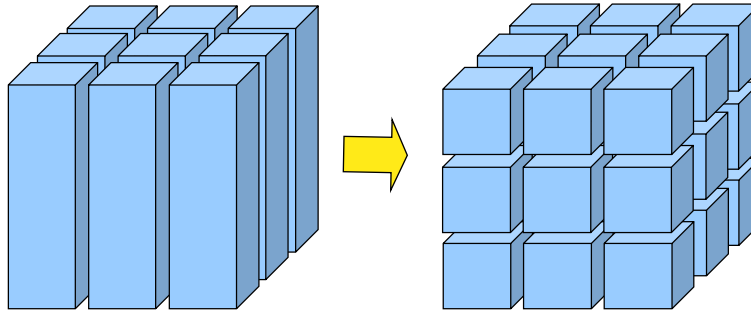


Category	Metrics		Dollar Sales
	Year		
Electronics	1997		\$ 10,616
	1998		\$ 29,299
Food	1997		\$ 5,300
	1998		\$ 5,638
Gifts	1997		\$ 16,315
	1998		\$ 20,047
Health & Beauty	1997		\$ 6,042
	1998		\$ 5,665
Household	1997		\$ 38,383
	1998		\$ 50,391
Kid's Korner	1997		\$ 2,559
	1998		\$ 2,943
Travel	1997		\$ 4,497
	1998		\$ 4,792

OLAP operators



drill-down



OLAP operators

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	\$ 1.676	\$ 291
Q1 1998	\$ 5.288	\$ 3.399	\$ 4.935	\$ 9.174	\$ 3.601	\$ 9.484	\$ 3.844	\$ 2.720	\$ 979	\$ 1.897	\$ 1.201
Q2 1998	\$ 1.773	\$ 3.658	\$ 1.862	\$ 1.213	\$ 1.115	\$ 4.635	\$ 352	\$ 724	\$ 2.110	\$ 391	\$ 12
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

drill-down



Metrics Customer City	Dollar Sales													
	Arlin	San Pedro	Springfield	Chappel Hill	Scrانبург	Pebble Beach	Martinsville	Maddon	Peoria	Pecos	Lake Barkley	Alcameda	Fingers Lake	S
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				\$ 237	\$ 30	\$ 119	
Q2 1998								\$ 12	\$ 17					
Q3 1998	\$ 734					\$ 25	\$ 1.535							
Q4 1998						\$ 219	\$ 119	\$ 142		\$ 85	\$ 1.533			

OLAP operators

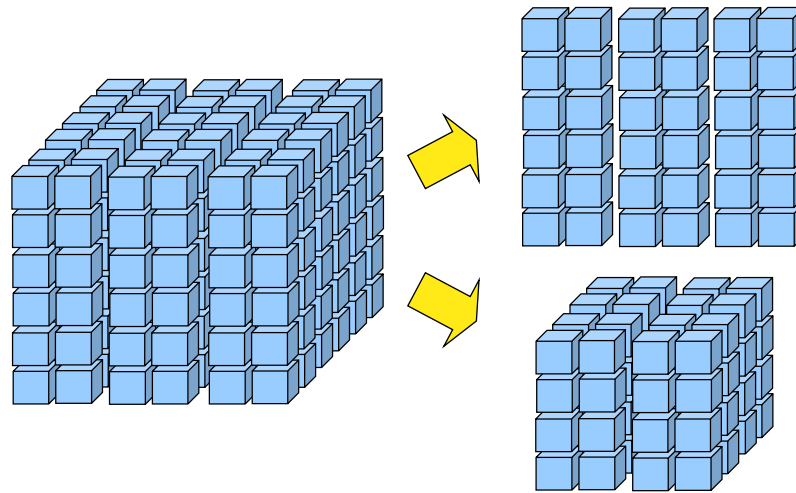
	Metrics	Dollar Sales	
	Year	1997	1998
Category			
Electronics		\$ 10.616	\$ 29.299
Food		\$ 5.300	\$ 5.638
Gifts		\$ 16.315	\$ 20.047
Health & Beauty		\$ 6.042	\$ 5.665
Household		\$ 38.383	\$ 50.391
Kid's Korner		\$ 2.559	\$ 2.943
Travel		\$ 4.497	\$ 4.792

drill-down



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	\$ 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 operators



slice-and-dice

OLAP operators

Category	Year	Metrics Customer Region	Dollar Sales									
			North-East	Mid-Atlantic	South-East	Central	South	North-West	South-West	England	France	Germany
Electronics	1997		\$ 138	\$ 1.774	\$ 384	\$ 138	\$ 2.346	\$ 2.554	\$ 2.184	\$ 566	\$ 199	\$ 702
	1998		\$ 1.184	\$ 4.529	\$ 1.892	\$ 7.232	\$ 651	\$ 9.488	\$ 476	\$ 2.683	\$ 462	\$ 1000
Food	1997		\$ 759	\$ 682	\$ 729	\$ 262	\$ 588	\$ 469	\$ 807	\$ 156	\$ 615	\$ 175
	1998		\$ 538	\$ 925	\$ 959	\$ 677	\$ 213	\$ 1.503	\$ 261	\$ 165	\$ 175	\$ 100
Gifts	1997		\$ 2.532	\$ 1.355	\$ 1.854	\$ 1.413	\$ 2.535	\$ 2.132	\$ 1.904	\$ 908	\$ 375	\$ 1.000
	1998		\$ 1.955	\$ 2.785	\$ 2.800	\$ 2.695	\$ 1.813	\$ 2.844	\$ 1.778	\$ 1.158	\$ 717	\$ 686
Health & Beauty	1997		\$ 624	\$ 640	\$ 1.317	\$ 647	\$ 588	\$ 754	\$ 654	\$ 143	\$ 292	\$ 344
	1998		\$ 611	\$ 887	\$ 566	\$ 382	\$ 499	\$ 1.162	\$ 1.044	\$ 273	\$ 72	\$ 198
Household	1997		\$ 5.354	\$ 4.112	\$ 5.410	\$ 4.446	\$ 3.058	\$ 3.974	\$ 2.654	\$ 3.545	\$ 2.875	\$ 1.900
	1998		\$ 5.787	\$ 5.320	\$ 5.416	\$ 6.812	\$ 4.334	\$ 5.008	\$ 7.588	\$ 2.139	\$ 3.649	\$ 2.791
Kid's Korner	1997		\$ 201	\$ 398	\$ 485	\$ 186	\$ 409	\$ 323	\$ 396	\$ 105	\$ 34	\$ 69
	1998		\$ 247	\$ 422	\$ 441	\$ 380	\$ 221	\$ 592	\$ 290	\$ 198	\$ 19	\$ 69
Travel	1997		\$ 624	\$ 505	\$ 564	\$ 386	\$ 300	\$ 978	\$ 416	\$ 48	\$ 38	\$ 55
	1998		\$ 608	\$ 559	\$ 1.096	\$ 611	\$ 464	\$ 316	\$ 573	\$ 257	\$ 198	\$ 55

slice-and-dice



Filter Details: Year = 1998												
Category	Metrics Customer Region	Dollar Sales										
		North-East	Mid-Atlantic	South-East	Central	South	North-West	South-West	England	France	Germany	Canada
Electronics		\$ 1.184	\$ 4.529	\$ 1.892	\$ 7.232	\$ 651	\$ 9.488	\$ 476	\$ 2.683	\$ 462	\$ 702	\$ 100
Food		\$ 538	\$ 925	\$ 959	\$ 677	\$ 213	\$ 1.503	\$ 261	\$ 165	\$ 175	\$ 100	\$ 100
Gifts		\$ 1.955	\$ 2.785	\$ 2.800	\$ 2.695	\$ 1.813	\$ 2.844	\$ 1.778	\$ 1.158	\$ 717	\$ 686	\$ 100
Health & Beauty		\$ 611	\$ 887	\$ 566	\$ 382	\$ 499	\$ 1.162	\$ 1.044	\$ 273	\$ 72	\$ 198	\$ 100
Household		\$ 5.787	\$ 5.320	\$ 5.416	\$ 6.812	\$ 4.334	\$ 5.008	\$ 7.588	\$ 2.139	\$ 3.649	\$ 2.791	\$ 100
Kid's Korner		\$ 247	\$ 422	\$ 441	\$ 380	\$ 221	\$ 592	\$ 290	\$ 198	\$ 19	\$ 69	\$ 100
Travel		\$ 608	\$ 559	\$ 1.096	\$ 611	\$ 464	\$ 316	\$ 573	\$ 257	\$ 198	\$ 55	\$ 100

OLAP operators

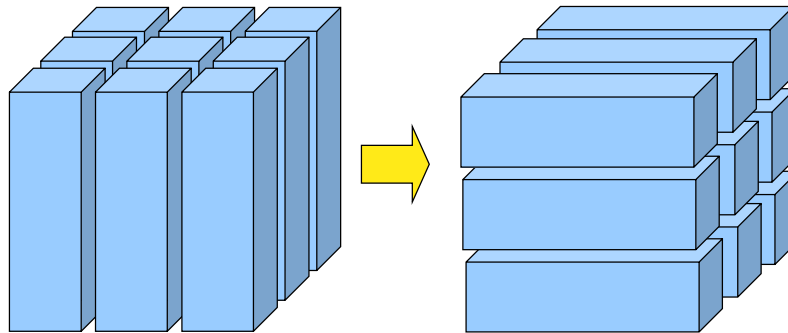
	Metrics	Dollar Sales											
Customer City		Afton	Akron	Albon	Alcameda	Alka	Allagash	Alta	Altoola	Amestra	Amsterdam	Andersonville	Annap
Subcategory													
Audio							\$ 85						
Automotive									\$ 30				
Chocolate		\$ 42	\$ 42		\$ 50		\$ 20		\$ 22	\$ 44			\$
Christmas		\$ 30					\$ 25	\$ 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			\$:

slice-and-dice



Filter Details:						
Category = Electronics						
AND						
Dollar Sales > 80						
AND						
Customer Region = North-West						
AND						
Year = 1997						
	Metrics	Dollar Sales				
Customer City		Alta	Armstrong	Avery Heights	Lane	San Francisco
Subcategory						
Audio			\$ 98		\$ 123	\$ 85
Comfort				\$ 118		\$ 1,495
Gadgets		\$ 199				\$ 199

OLAP operators



pivoting

OLAP operators

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



Category	Year	Metrics	Dollar Sales	
			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

pivoting

OLAP operators

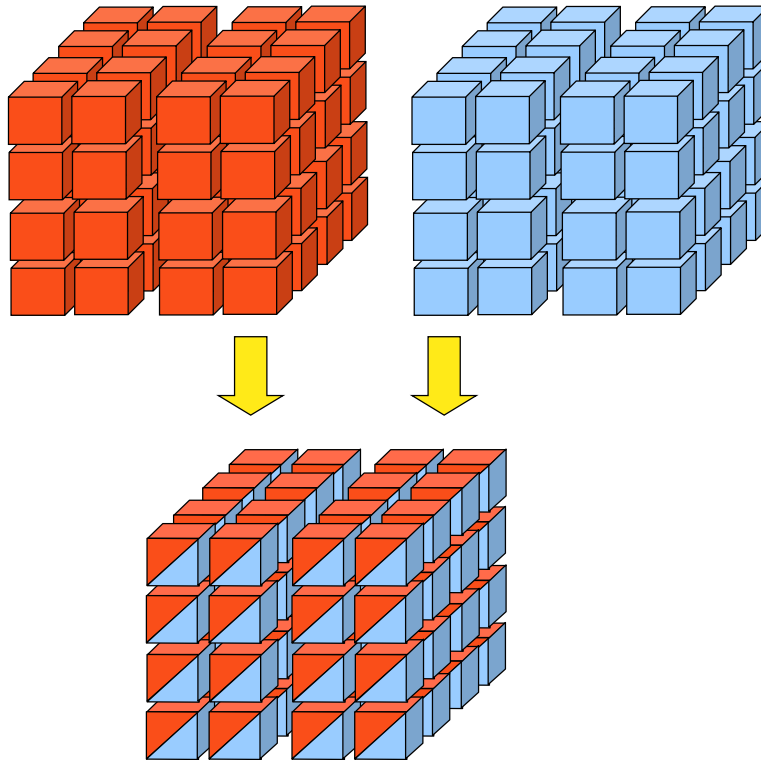
		Metrics Customer Region	Dollar Sales									
			North-East	Mid-Atlantic	South-East	Central	South	North-West	South-West	England	France	Germa
Category	Year											
Electronics	1997		\$ 138	\$ 1.774	\$ 384	\$ 138	\$ 2.346	\$ 2.554	\$ 2.184	\$ 566	\$ 199	\$
	1998		\$ 1.184	\$ 4.529	\$ 1.892	\$ 7.232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

pivoting



Category	Year	Metrics Customer Region	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	\$ 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 operators



drill-across

OLAP: operators

Category	Metrics	Dollar Sales							
	Quarter	Q1 1997	Q2 1997	Q3 1997	Q4 1997	Q1 1998	Q2 1998	Q3 1998	Q4 1998
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	\$ 12,898	\$ 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

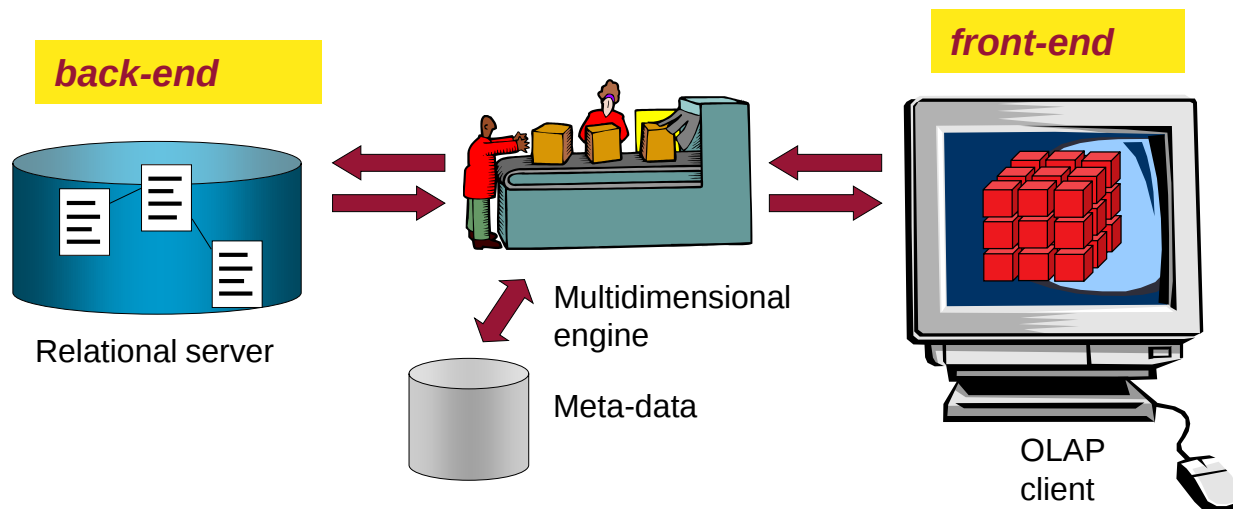


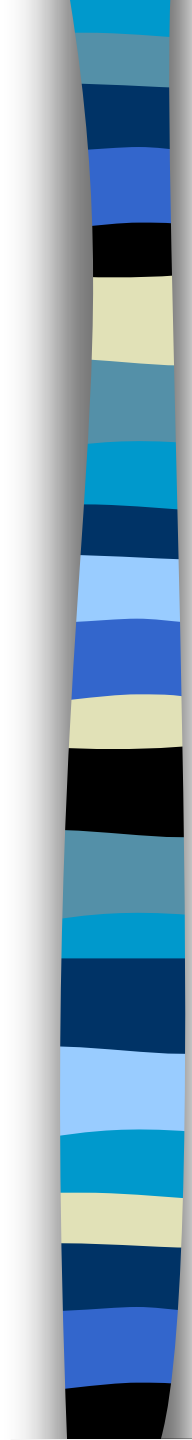
drill-across

Category	Quarter	Q1 1997		Q2 1997		Q3 1997		Q4 1997		Q1 1998		Q2 1998	
	Metrics	Discount	Dollar Sales	Discount	Dollar Sales	Discount	Dollar Sales	Discount	Dollar Sales	Discount	Dollar Sales	Discount	Dollar Sales
Electronics		\$ 0	\$ 4,383	\$ 0	\$ 817	\$ 0	\$ 827	\$ 300	\$ 4,589	\$ 15	\$ 13,770	\$ 0	\$ 2,977
Food		\$ 25	\$ 1,546	\$ 0	\$ 1,310	\$ 0	\$ 1,268	\$ 38	\$ 1,176	\$ 0	\$ 2,676	\$ 0	\$ 1,120
Gifts		\$ 31	\$ 3,398	\$ 0	\$ 3,893	\$ 5	\$ 4,682	\$ 0	\$ 4,342	\$ 15	\$ 7,879	\$ 0	\$ 4,145
Health & Beauty		\$ 0	\$ 1,826	\$ 0	\$ 878	\$ 0	\$ 1,904	\$ 0	\$ 1,434	\$ 229	\$ 2,156	\$ 0	\$ 898
Household		\$ 0	\$ 9,314	\$ 228	\$ 8,124	\$ 175	\$ 9,331	\$ 35	\$ 11,614	\$ 5	\$ 17,453	\$ 211	\$ 7,604
Kid's Korner		\$ 0	\$ 685	\$ 0	\$ 531	\$ 32	\$ 811	\$ 40	\$ 532	\$ 0	\$ 1,084	\$ 0	\$ 491
Travel		\$ 0	\$ 603	\$ 0	\$ 1,293	\$ 200	\$ 1,456	\$ 0	\$ 1,145	\$ 0	\$ 1,507	\$ 0	\$ 719

ROLAP (Relational OLAP)

- Why adopting the relational technology to store data to a data warehouse?
 1. huge amount of literature written about the relational model
 2. broadly available corporate experience with relational database usage and management
 3. top performance and flexibility standards of relational DBMSs
- Data are stored in a relational DBMS, in a detailed and pre-aggregated form
 - ✓ Need for specific types of schemata to map the multidimensional model onto the relational model: *star schema*
 - ✓ The performance problem requires *denormalization* to avoid expensive joins





MOLAP

(Multidimensional OLAP)

- Based on an ad hoc logical model that can be used to represent multidimensional data and operations directly
- Data are physically stored in arrays with positional access
 - ✓ The greatest advantage of MOLAP systems in comparison with ROLAP is that multidimensional operations can be performed in an easy, natural way without any need for complex join operations
 - ✓ However, MOLAP implementations have very little in common, because no multidimensional logical model standard has yet been set; generally, they simply share the usage of optimization techniques specifically designed for sparsity management