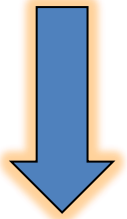
Compact Course in Data Mining

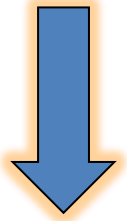
Data Mining and Business Intelligence

Professor Dr. Gholamreza Nakhaeizadeh

Why BI ?

„Business Intelligence“





About 50.000.000 results (0,37 seconds)

Worldwide BI, Analytics and Performance Management Revenue Estimates for 2010 (Millions of U.S. Dollars)

Source: Gartner (March 2011)

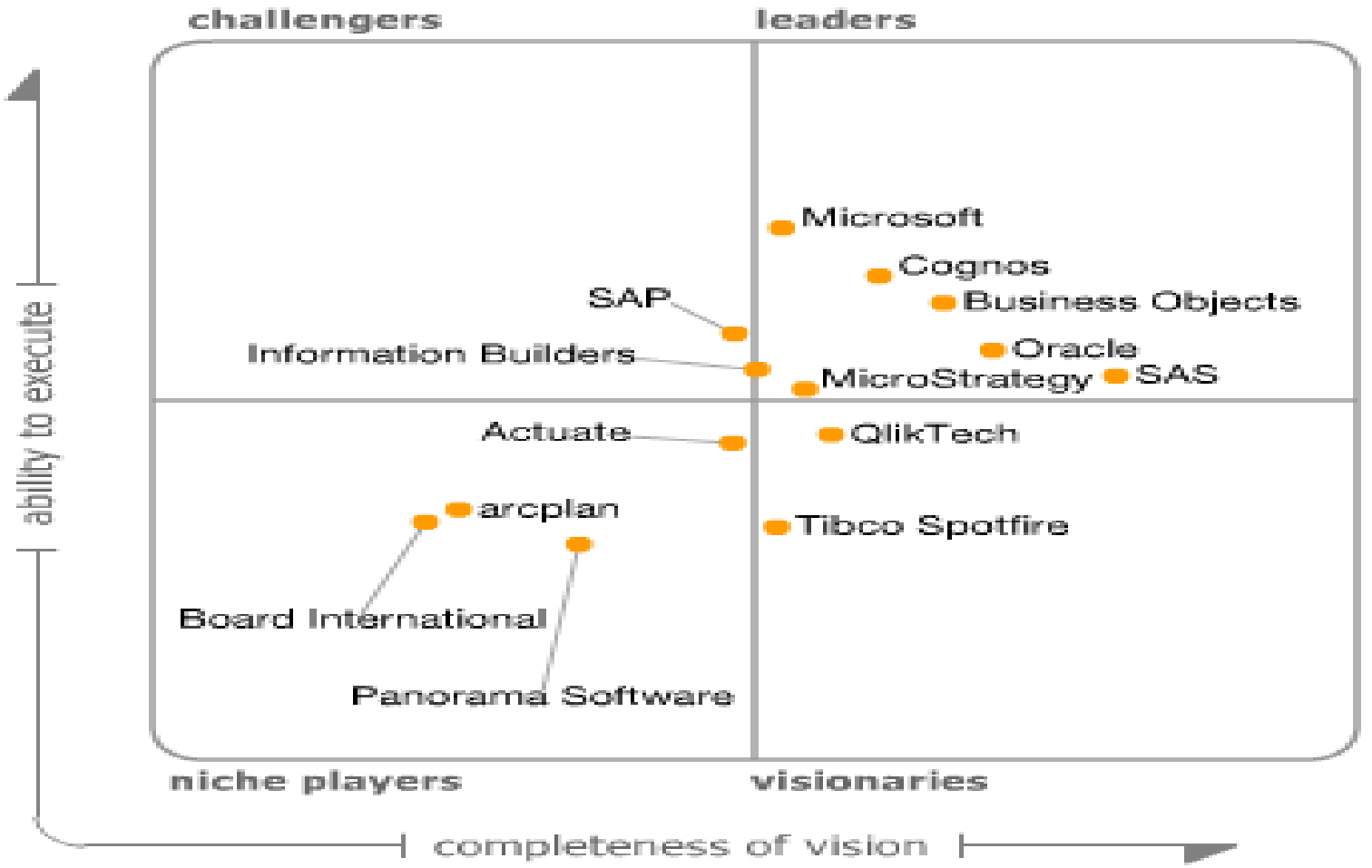
Company	2010 Revenue	2010 Market Share (%)	2009 Revenue	2009 Market Share (%)	2009-2010 Growth
SAP	2,413.1	22.9	2,066.2	22.3	16.8
Oracle	1,645.8	15.6	1,350.5	14.6	21.9
SAS Institute	1,386.5	13.2	1,324.6	14..3	4.7
IBM	1,222.0	11.6	1,135.6	12.2	7.6
Microsoft	913.7	8.7	739.5	8.0	23.6
Other Vendors	2,940.6	27.9	2,661.5	28.7	10.5
Total	**10,521.8**	**100.0**	**9,277.9**	**100.0**	**13.4**

Worldwide BI, Analytics and Performance Management Revenue Estimates for 2011 (Millions of U.S. Dollars)

Company	2011 Revenue	2011 Market Share (%)	2010 Revenue	2010 Market Share (%)	2010-2011 Growth (%)
SAP	2,883.5	23.6	2,413.1	23.0	19.5
Oracle	1,913.5	15.6	1,645.8	15.7	16.3
SAS Institute	1,542.8	12.6	1,386.5	13.2	11.3
IBM	1,477.6	12.1	1,222.0	11.6	20.9
Microsoft	1,059.9	8.7	913.7	8.7	16.0
Other Vendors	3,363.8	27.5	2,931.1	27.9	14.8
Total	**12,241.0**	**100.0**	**10,512.2**	**100.0**	**16.4**

Source: Gartner (March 2012)

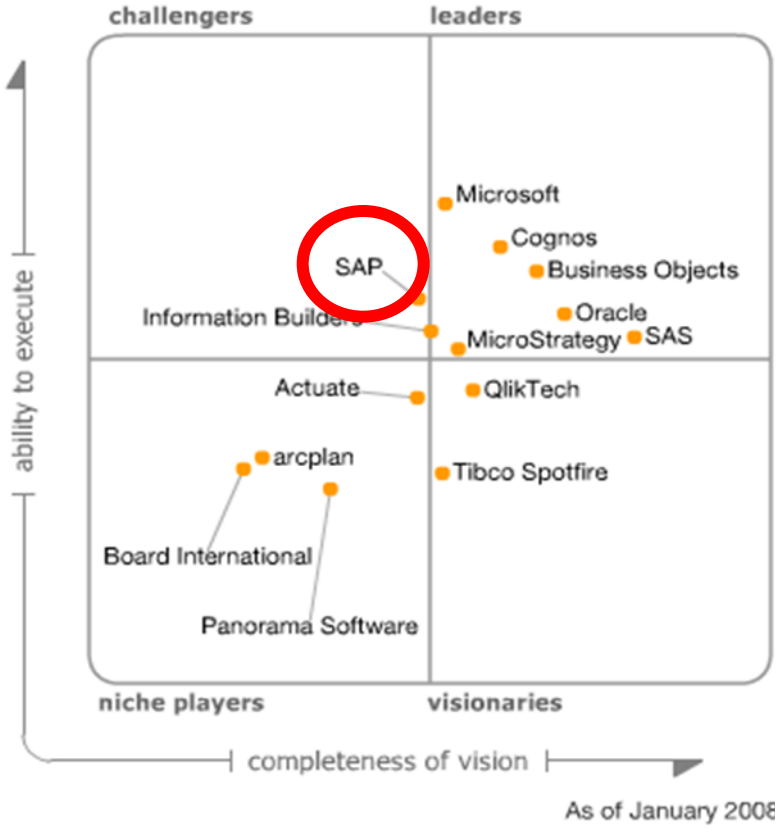
Gartner Magic Quadrant for Business Intelligence Platforms



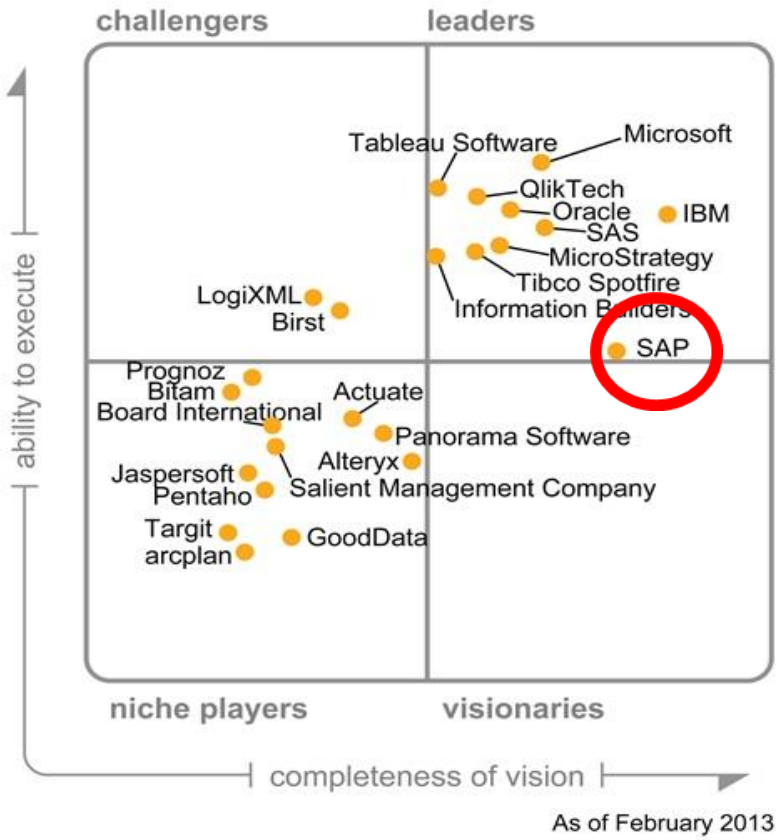
Source: Gartner (January 2008)

Gartner Magic Quadrant for Business Intelligence Platforms

2008



2013



Source: Gartner (February 2013)

Gartner Magic Quadrant for Business Intelligence Platforms 2015

What is BI ?

Some definitions (theoretical point of view)

- BI is a broad category of Management Information Systems, applications and technologies for gathering, storing, analyzing, and providing access to data to help enterprise users make better business. decisions.

- The process firms go through to gather, store and ana

- A critical activity that helps companie
decisions, as well as increas loyalty,
streamline operations, i nt and even enable
processes.

- A tegy, vision and architecture that continuously se ization's operations and direction with its strategic B. goals.

"One version of the truth"

The theory that the more you know about your customers and the B. problem you're trying to solve, the better you're able to solve it.

- An effective BI system provides corporations with "one version of the truth".

Source: http://www.managementlogs.com/2004/09/what-is-bi-really-multiple-versions-of.html

.....

BI-History

Business intelligence was defined in an October 1958 IBM Journal article by Hans Peter Luhn.

New Father

Howard Dresner

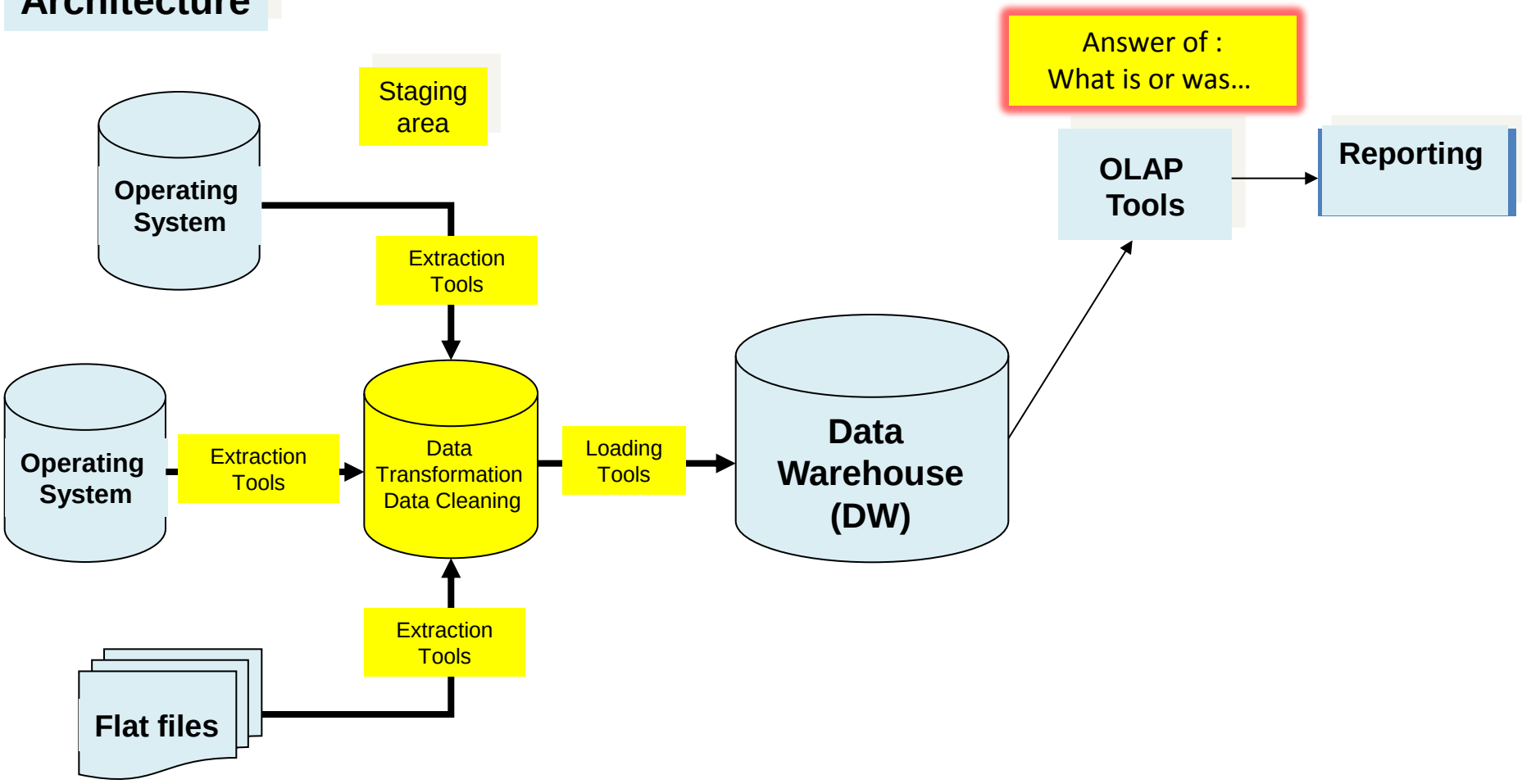
The term "business intelligence" was refounded in 1989 **(31 years after Luhn)** by Howard Dresner an analyst of Gartner Group.

Now: Dresner Advisory Services, LLC

http://en.wikipedia.org/wiki/Business_intelligence

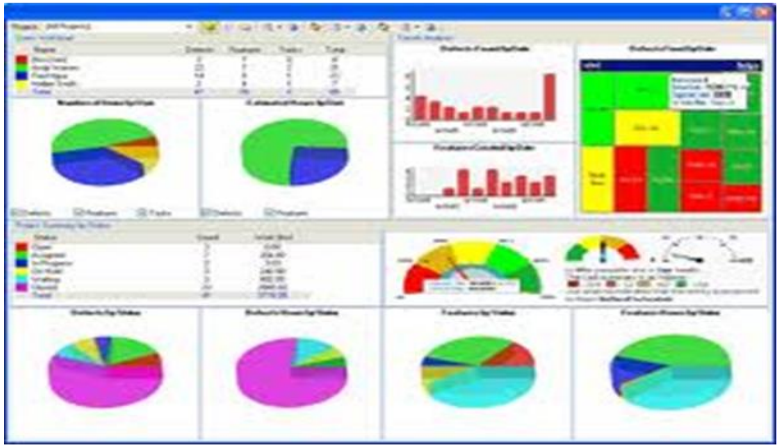
Main Components of a Business Intelligent (BI) System

Architecture



ETL: Extraction, Transformation, Loading

Reporting tools



Where is the intelligence ?

What is „Business“ in BI ?

In this paper, *business is a collection of activities carried on for whatever purpose*, be it science, technology, commerce, industry, law, government, defense, et cetera.

Business is not just:

- **Commerce**
- **CRM**
- **Finance**
- **….**

What about „Intelligence“ in BI ?

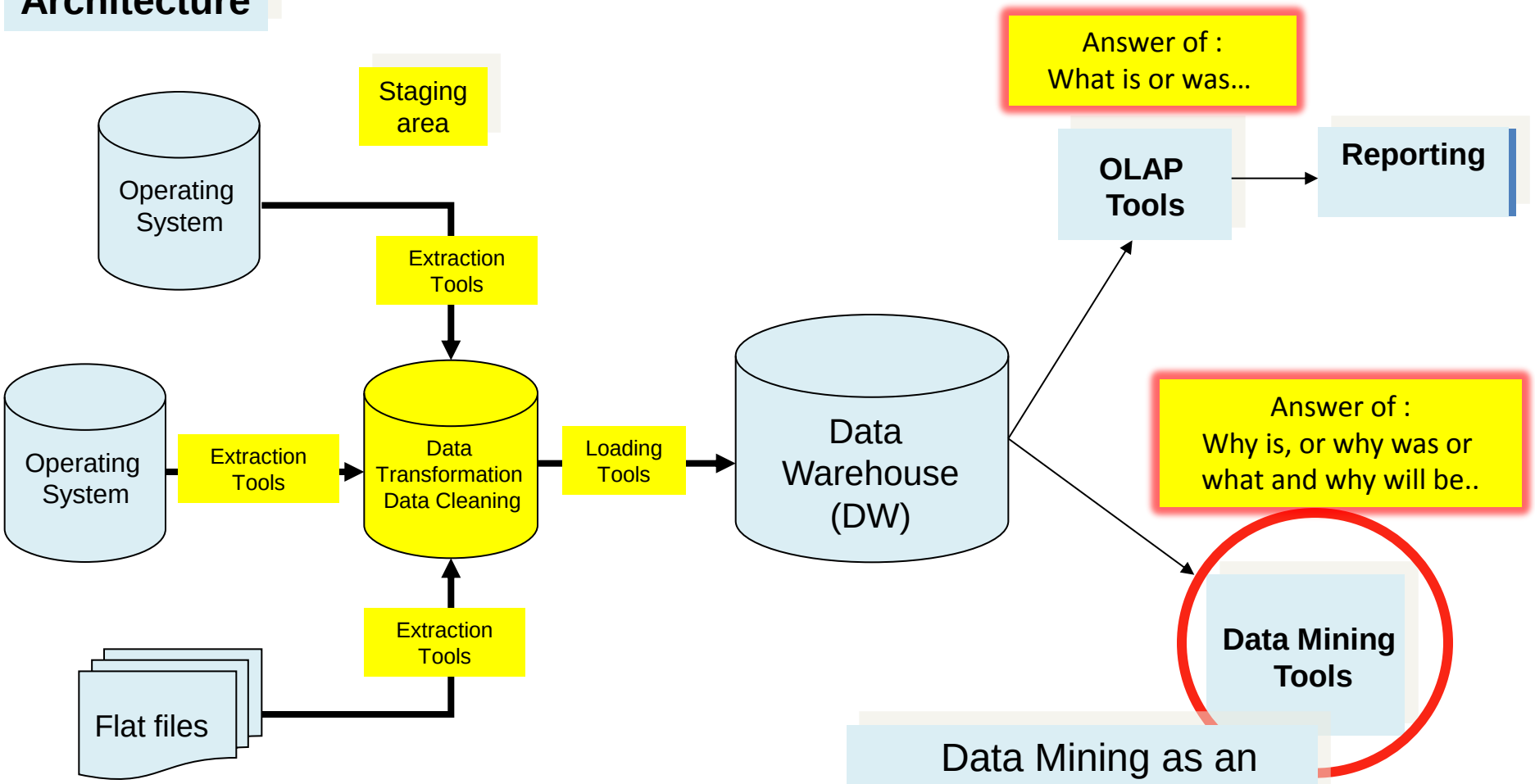
The notion of intelligence is also defined here, in a more general sense, as *"the ability to apprehend the interrelationships of presented facts in such a way as to guide action towards a desired goal."*

One interpretation of

Knowledge Discovery in data
(Data Mining)

Main Components of a Business Intelligent (BI) System

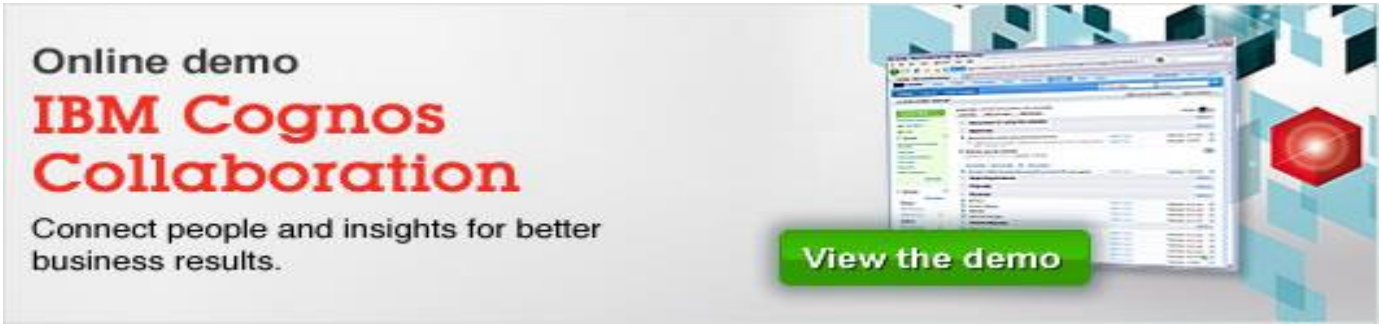
Architecture



ETL: Extraction, Transformation, Loading

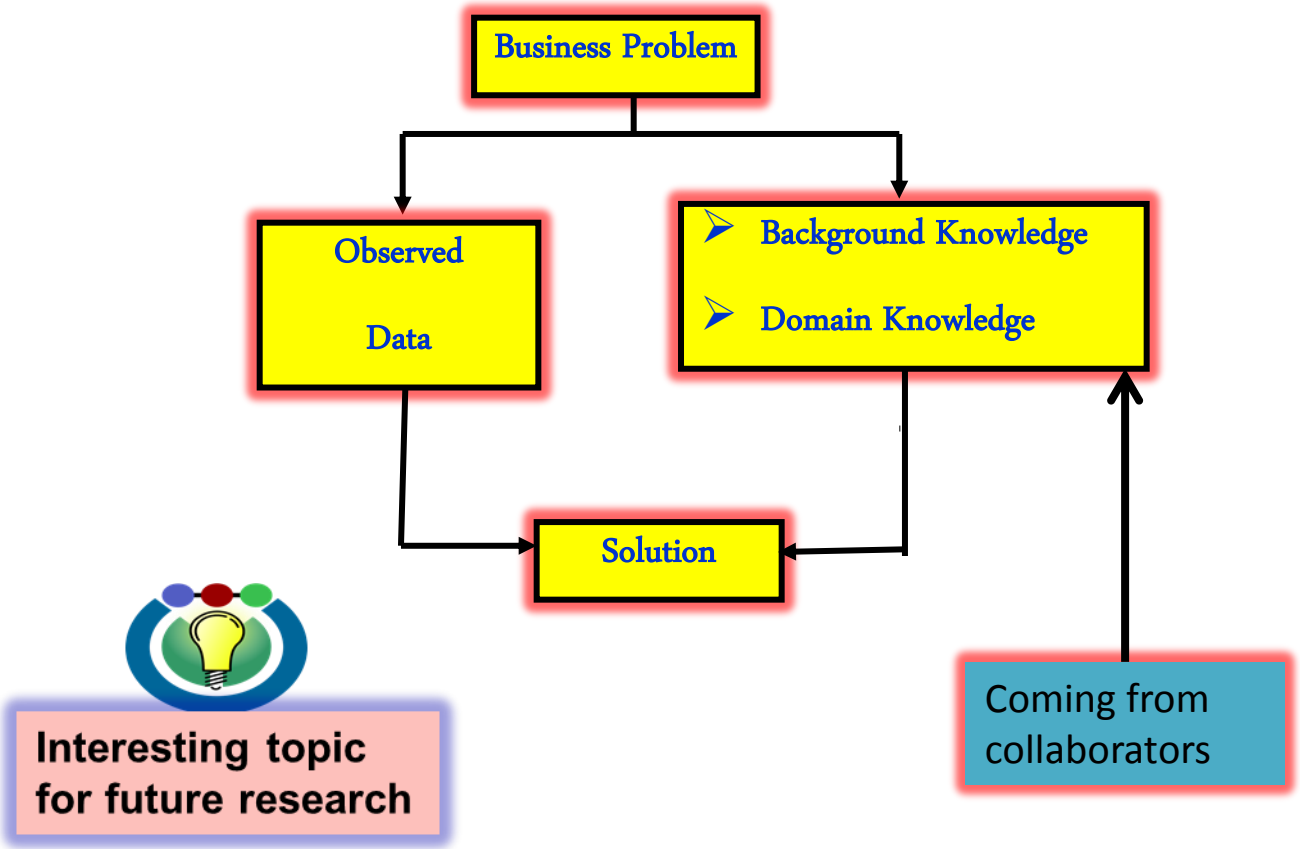
Future trend in BI and the roll of Intelligence

- **Collaborative BI**



Collaborative BI

Data and Knowledge Driven Problem Solving in Healthcare



Future trend in BI and the roll of Intelligence

- Mobile BI

The current top three features for Mobile BI:

- Viewing
- Alerting
- KPI Monitoring

Mobile BI

In future needed technology for performing data mining:

- Intelligent Sampling
- High speed learning algorithms
- Learning from small training datasets
- Sophisticated content extraction algorithms

Launching of a Business Intelligence Competency Centre (BICC)

Launching of a Business Intelligence Competency Centre (BICC)

What is BICC ?

A Business Intelligence Competency Center is a cross-functional organizational team that has defined tasks, roles, responsibilities for supporting and promoting the effective use of Business Intelligence across an organization.

Source: http://en.wikipedia.org/wiki/Business_Intelligence_Competency_Center

A Business Intelligence Competency Center is a:

- permanent and formal organizational structure
- staffed internally (i.e. by the organization) with individuals from the business and IT
- with defined tasks, roles, responsibilities and processes
- supporting and promoting the effective use of Business Intelligence to drive the business strategy.

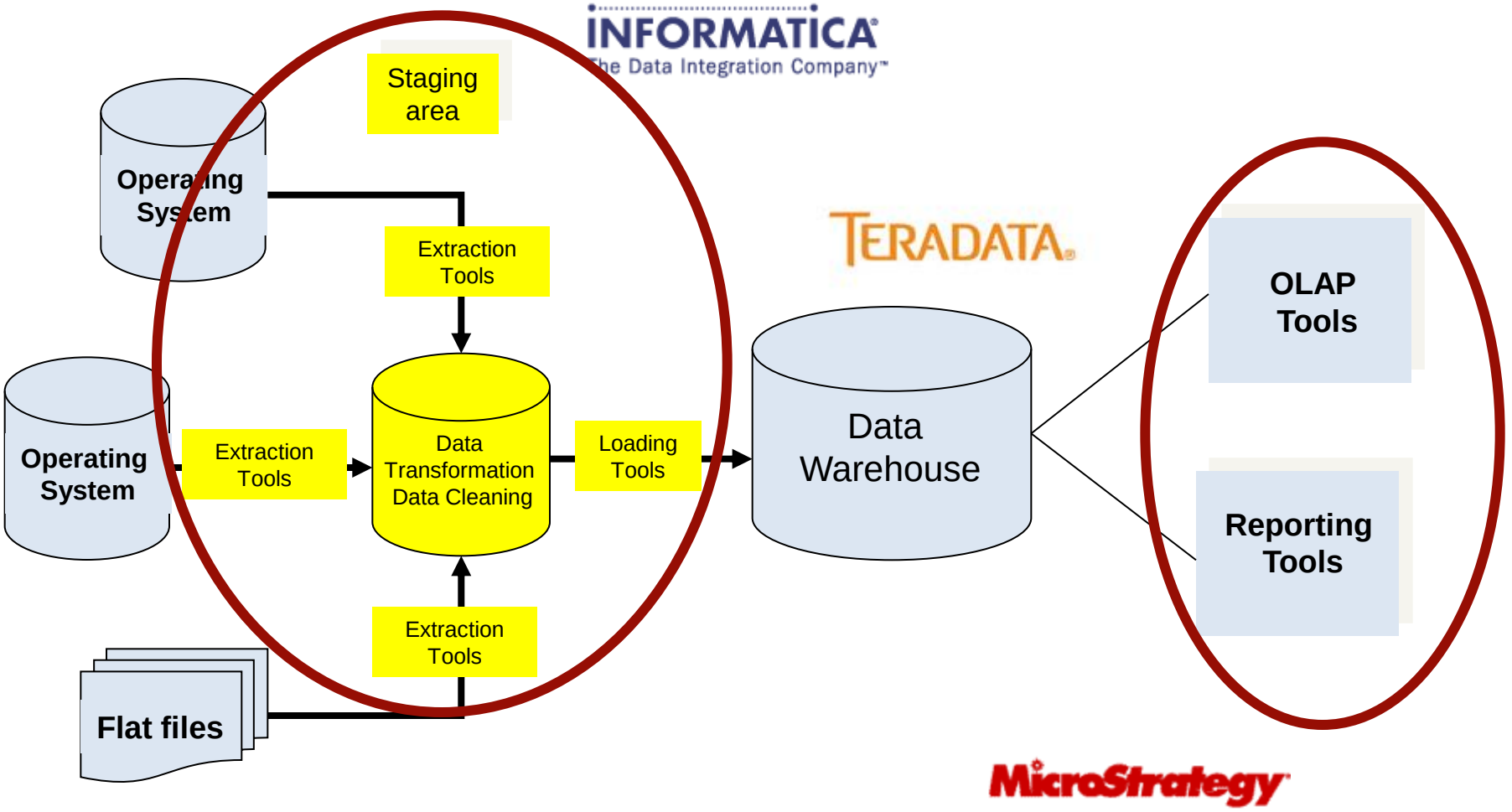
Source: http://www.computerworld.com/pdfs/sas_free_wp.pdf

Example : Project AQUA at Daimler AG

- „AQUA (Advanced Quality Analysis) is the standardized platform for
  - *Product and diagnosis reliability,*
  - *Fixed First Visit controlling*
  - *as well as market and dealer controlling at Daimler*
- The system identifies quality problems in
  - *Development,*
  - *Production*
  - *After-Sales*
- *“There were many different systems for data analysis in use,* the idea was, therefore, to transfer *all the data into a consistent system"*

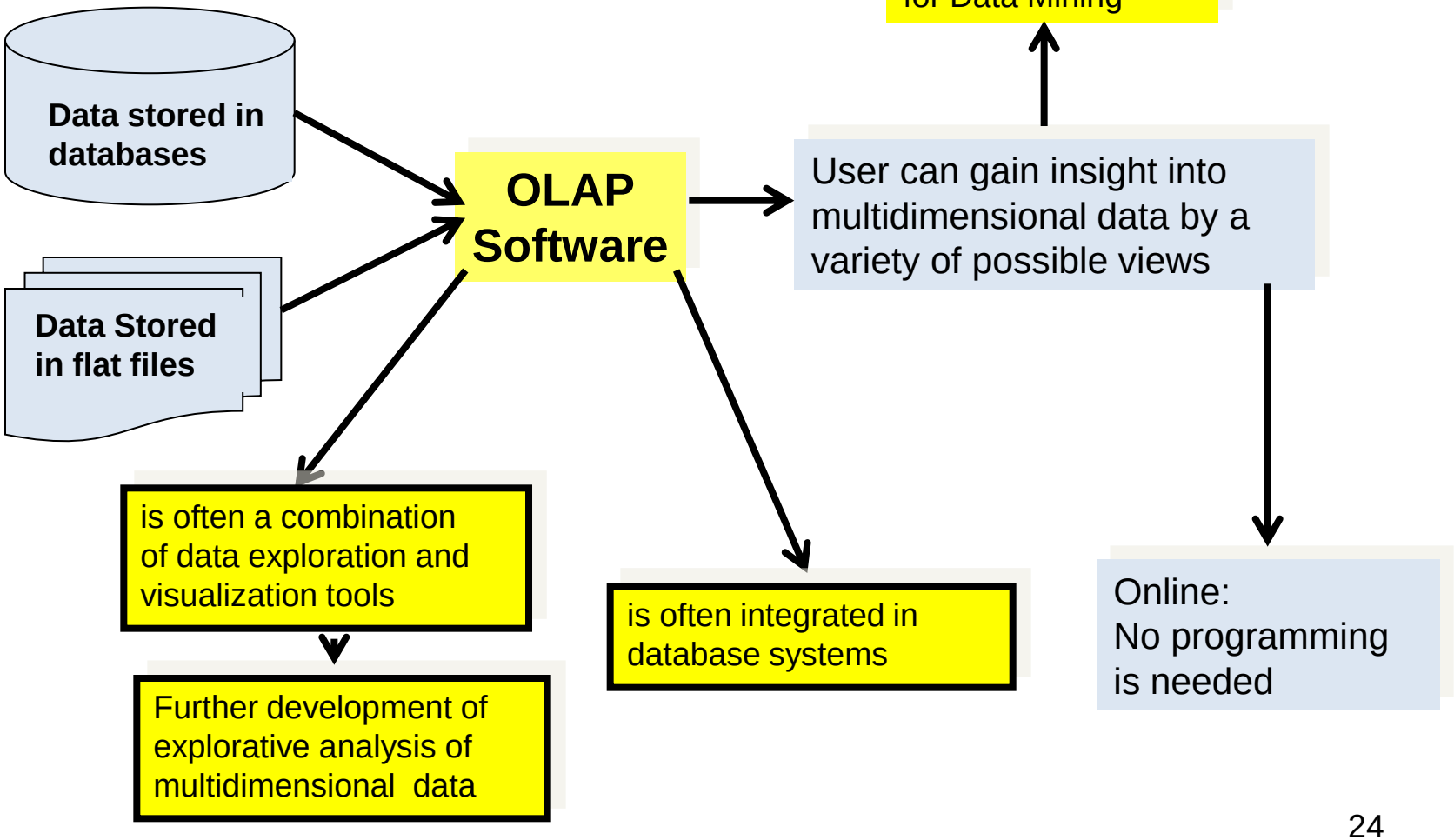
Source: http://www.teradataemea.com/Lisbon/Speakers/Parallel-Sessions-Speakers/page.aspx/975?xf_itemId=37 and http://www.monitor.co.at/index.cfm?storyid=9952&mark=daimler

The Providers, Example



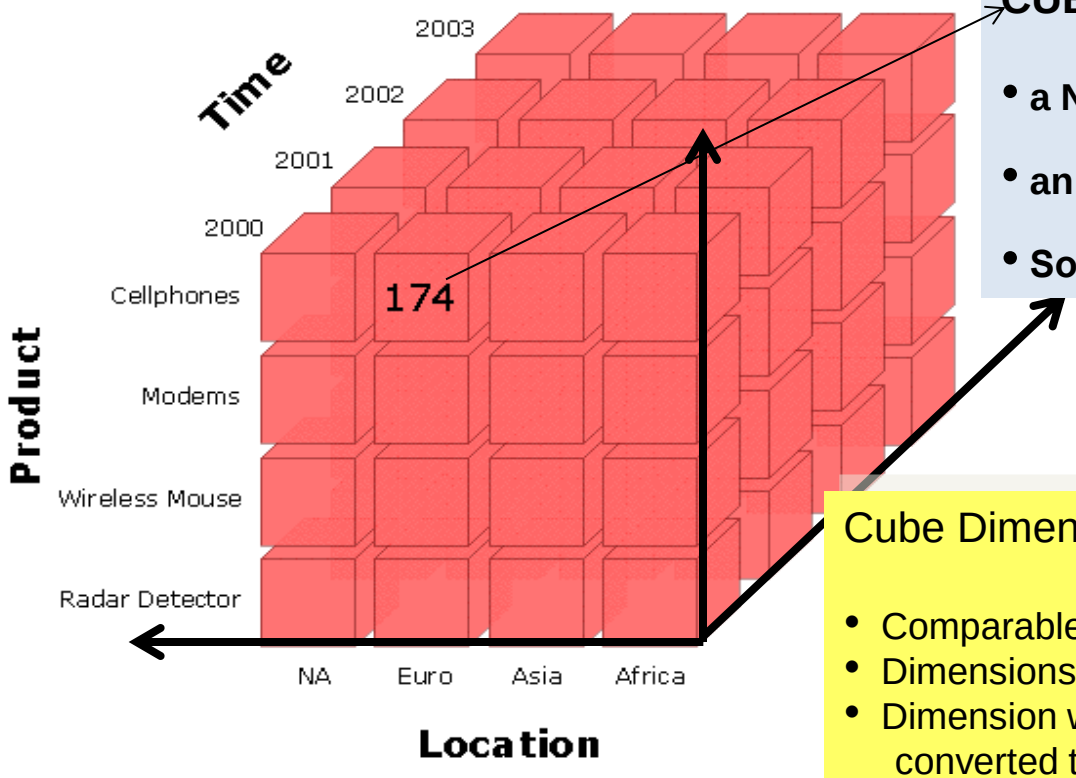
OLAP

OLAP: Online Analytical Processing



OLAP

OLAP-CUBE:
Analysis in OLAP is done by using OLAP-CUBES



CUBE Measure: content of a cell can be

- **a Number (number of cell phones produced in Europe in 2000)**
- **an amount (total sales in $ of cell phones produced in Europe in 2000)**
- **Sometimes called "target quantity"**

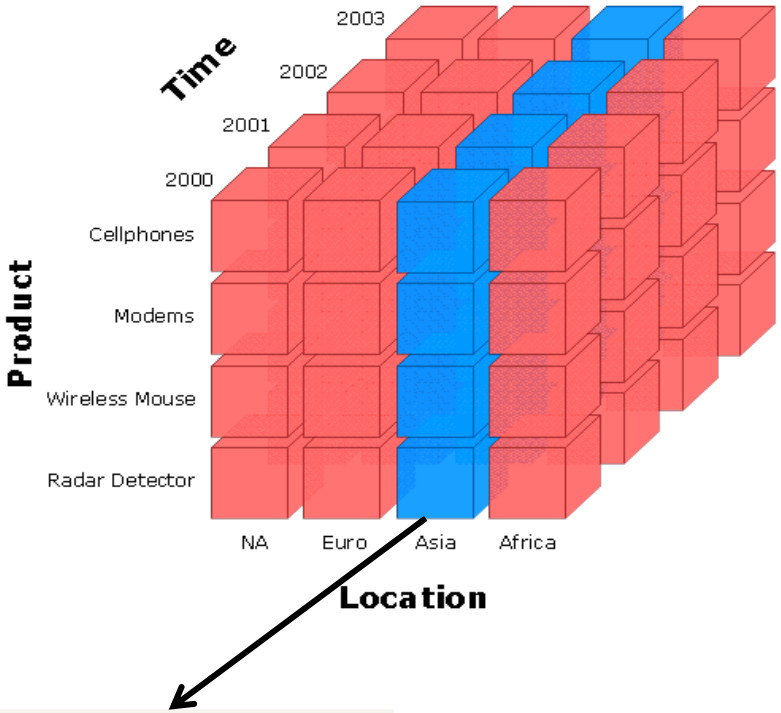
Cube Dimensions:

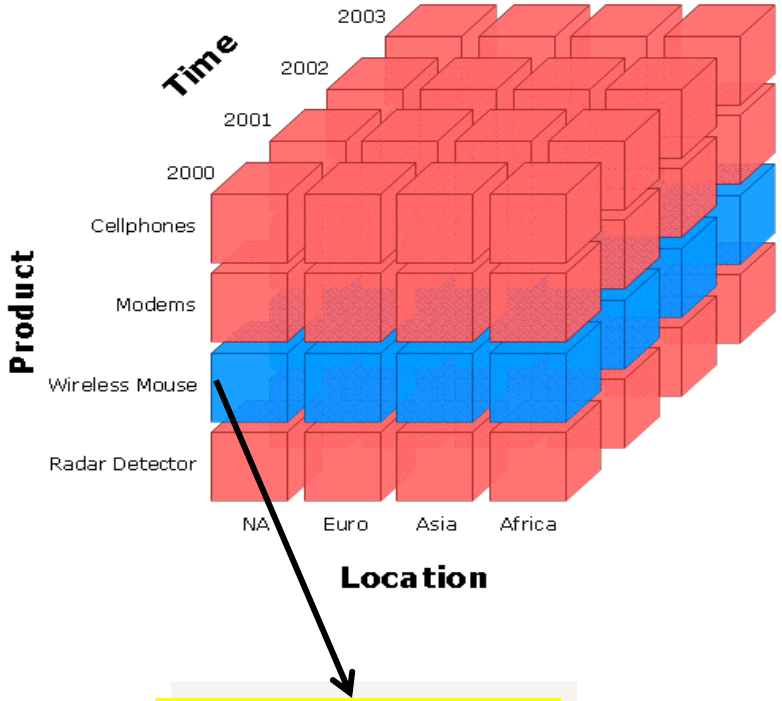
- Comparable with attributes in Data Mining
- Dimensions have nominal values (called categories)
- Dimension with continuous categories have to be converted to nominal categories
- In the reality, the number of Dimensions is often more than 3 (Hypercube)

Source of the cube fig. in this and the following pages: http://training.inet.com/OLAP/Cubes.htm

OLAP

Slicing: Selecting a value of a dimensional and consider all the cells belong to other dimensions





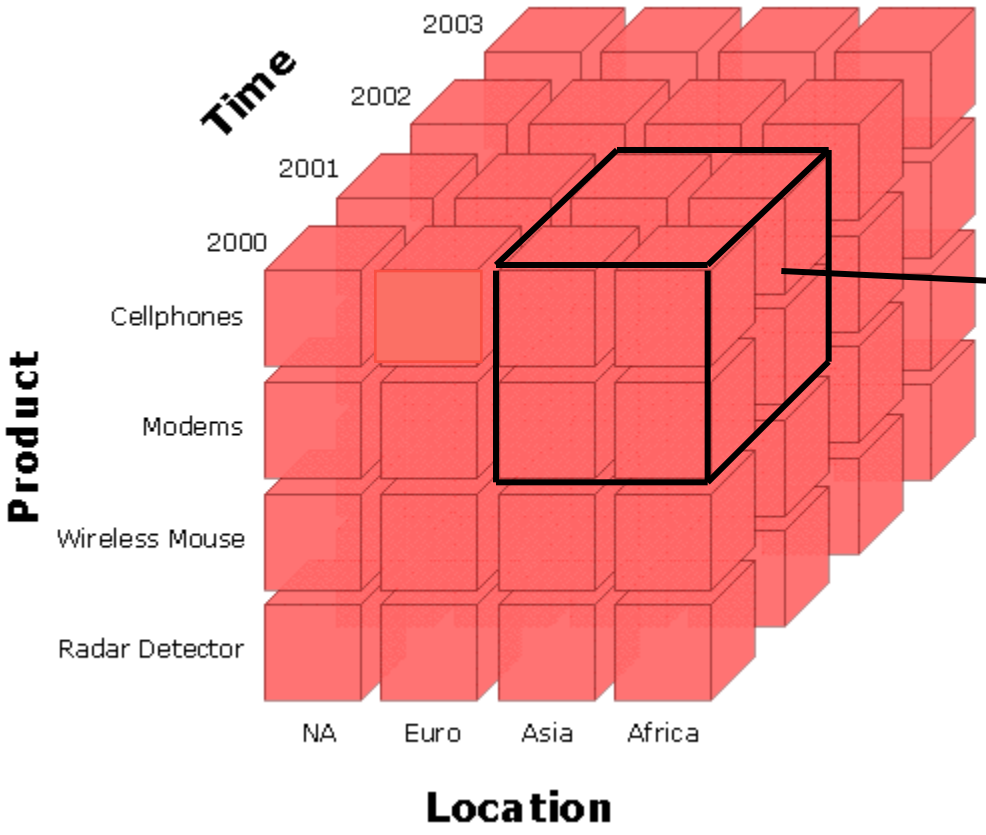
Slice Asia

Slice Wireless Mouse

Consist of 16 cells and 16 measures

OLAP

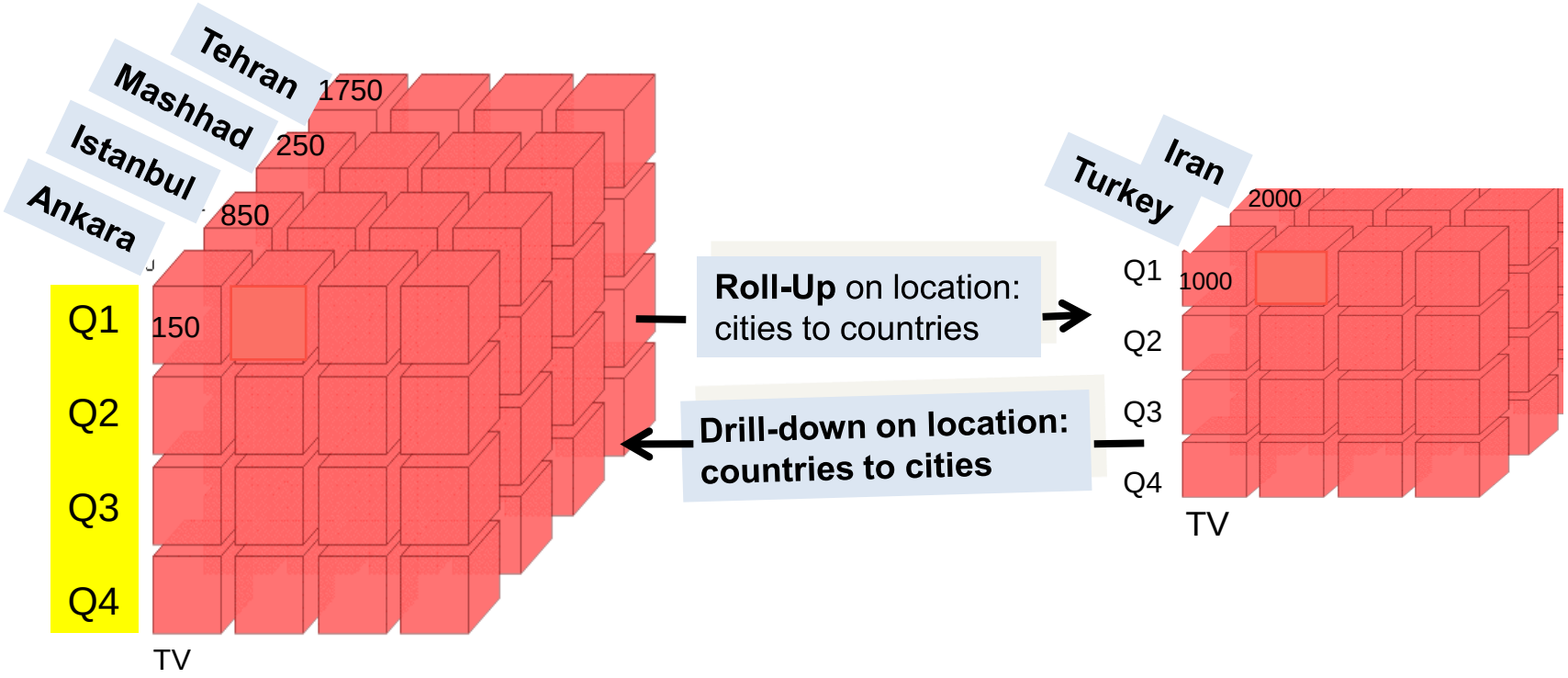
Dicing: selecting a subset of a cube on two or more dimensions



Dice operation involving 3 Dimensions:
(Location: Asia, Africa),
(Product: Modems, Cell phones)
and (Time: 2000, 2001)

OLAP

Roll-up : Aggregation by climbing up a category hierarchy



Drill-down : Going to more detailed data by stepping down a category hierarchy

OLAP

Other capabilities and functionalities

- Calculation Engine for
  - Ratios
  - Mean
  - Variance
  - …..
- Supporting functional modeling for:
  - Forecasting
  - Trend analysis
  - Other statistical computations and tests

OLAP

Other systems

- **ROLAP: Relational OLAP**
  - **OLAP software based on relational data bases**
  - **They have greater scalability than MOLAP but less efficiency**
- **MOLAP: Multidimensional OLAP**
  - **OLAP software based on multidimensional data models**
  - **Mapping multidimensional views directly to data cube array structures**
- HOLAP: Hybrid OLAP
  - Such systems combine ROLAP and MOLAP technologies
  - They benefit from the high scalability of ROLAP systems and faster computation of MOLAP systems
- OLAM: Online Analytical Mining
  - Integration of OLAP with Data Mining
    - Related to the concept "in-database Mining"

Data Warehouse (DWH)

Introduction

Development of DWH started in the beginning of 80s
DWH is an enterprise-wide *database* that serves as a databse for all kind of management support systems

Definition:

Several definition can be found for DW in the literature.
One often used is due to W. H. Inmon:

„A Data Warehouse is a subject-oriented, integrated, time-variant and non-volatile collection of Data in support of managements Decision support process.”

Technical potential benefits

- Integrated database systems for management support
- Discharge operational data processing systems
- Quick queries and reports due to the integrated data

Data Warehouse

Definition (continuous)

✓ Subject-Oriented:
Oriented to main subjects like Customer, Company, product, supplier,..
instead to concentrate on company's ongoing operations.

✓ Integrated:
Integrate data from different heterogeneous data sources
Relational databases flat files….
by application of data cleaning and data integration methods consistency in naming, encoding structure and attributes measures is fulfilled

✓ Time-variant : Analysis on temporal changes and developments requires the long-term storage of data in DW; therefore “time” is a main dimension of DW

✓ Nonvolatile: The data once stored in a DW should not change ; otherwise it is not possible to perform a realistic data analysis