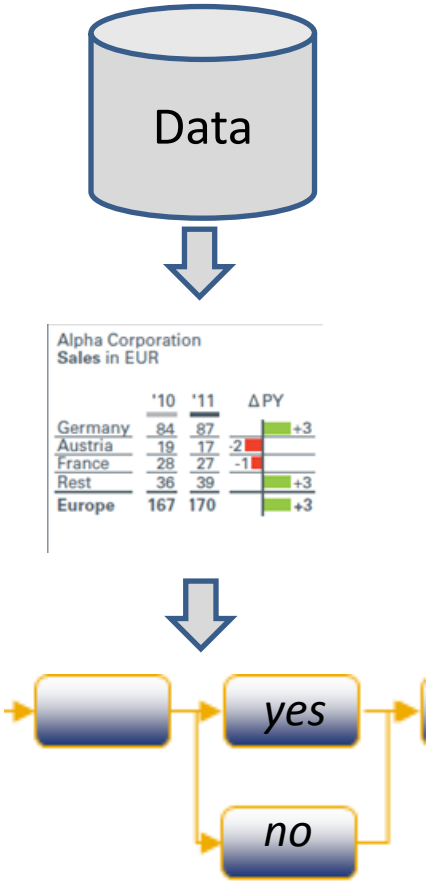
University of Applied Sciences Northwestern Switzerland
School of Business

Business Intelligence

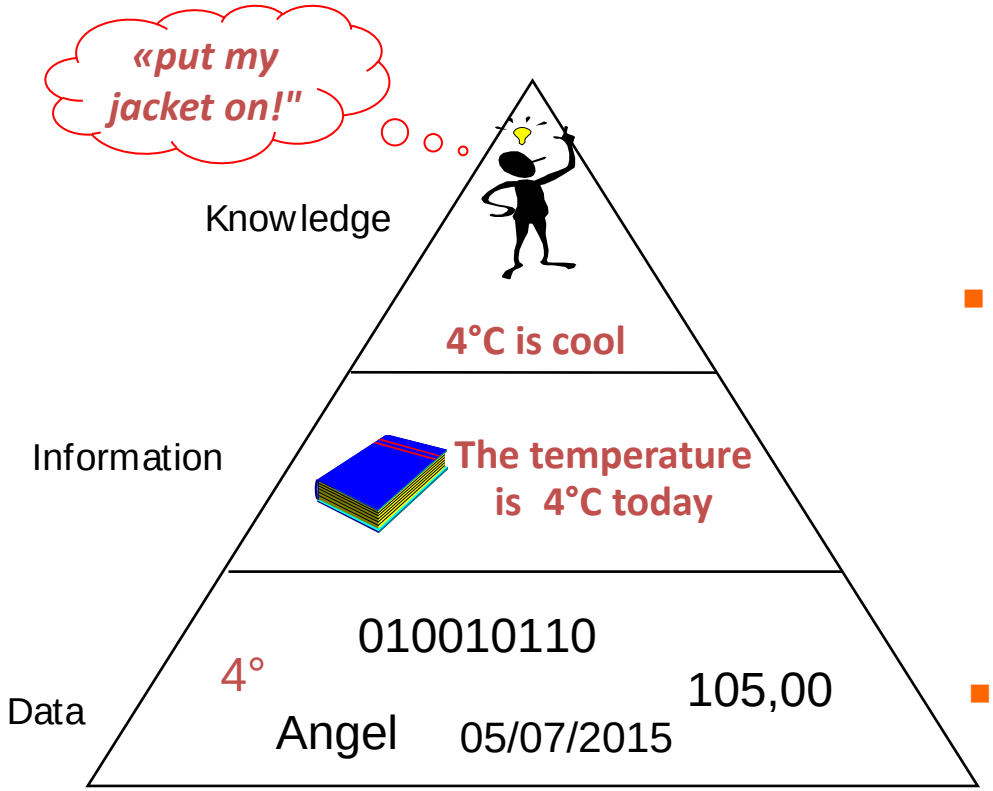
Knut Hinkelmann

Business Intelligence – Definition(s)

- *Hans-Peter Luhn (1958):* «the ability to apprehend the interrelationships of presented **facts** in such a way as to **guide action** towards a desired goal»
- *Sabherwal (2011):* «We define BI as providing decision makers with valuable information and knowledge by leveraging a variety of sources of data as well as structured and unstructured information. […] The key intellectual output of BI is **knowledge that enables decision making** with **information and data being the inputs**.»
- *Howson (2007):* Business Intelligence allows people at all levels of an organisation to **access, interact with and analyse data** to **manage the business, improve performance, discover opportunities, and operate efficiently.**



Data, information and knowledge

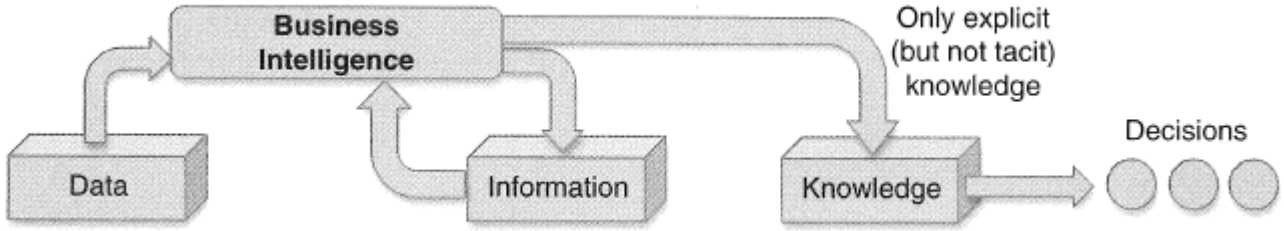


- **Knowledge** enables decisions and actios
  - originates from messages (information), experience, insight
  - is embedded into the beliefs and opinions of its owner
- **Information** is an interpretation of data, often assembled in messages
  - influences the judgment and behaviour of the recipient and
  - that has a significance (relevance, purpose)
- **Data** is a set of facts and/or signals
  - Do not have meaning by itself
  - To understand data you need an interpretation

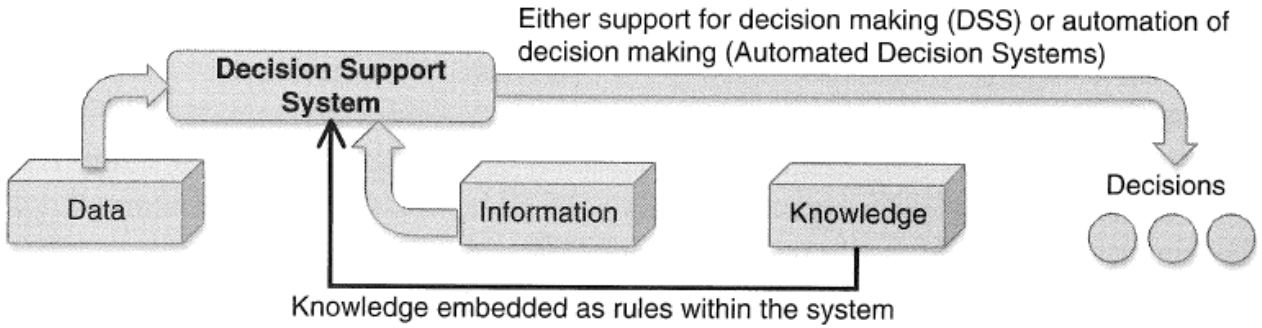
Analytic vs. transaction processing

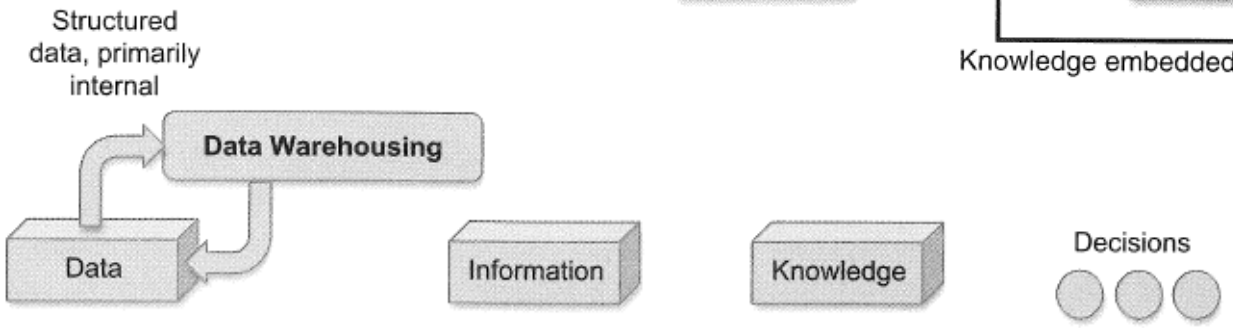
- BI focuses on **analytic** processing instead of **transaction** processing
  - transaction processing supports *execution of core business processes*
    - *use knowledge*
  - analytic processing supports *insight and decision-making*
    - *create knowledge*

Business Intelligence vs. Data warehousing, Decision Support Systems…



vs.



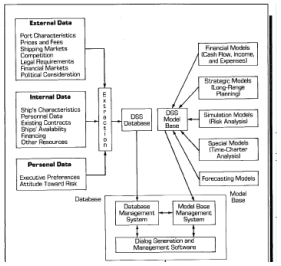


figures taken from Sabherwal 2011

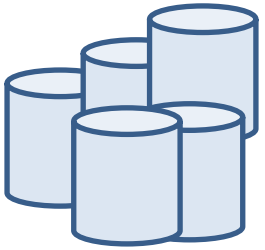
History of BI

1960 1970 1980 1990 2000 2010

- first model-driven decision support systems
- theory development, rising academic interest (DSS)
- Data warehouses
- OLAP
- data-driven approaches
- «Business Intelligence»
- Web-based BI systems
- Web mining
- e-commerce, targeted marketing, CRM
- «Business Analytics»
- Web Analytics
- «predictive analytics»
- «Big data»



Models



Data

Perspectives on BI – pain points

MARKETING

For targeted campaigns, we would urgently need a data basis that is harmonised with sales […] ideally on an integrated platform where we can communicate with sales.

MANAGEMENT

*I told my people that I wanted to retrieve some numbers myself from my laptop. I then got **access to various (!) systems** […] I finally gave up and now have an employee who does nothing but **create reports** for me […]*

SALES

*In most review meetings, we spend half the time discussing whose figures are the right ones because everyone **brings their own reporting**. I have the impression that for any key figure it is **possible to produce any value from the raw data**.*

SUPERVISORY BOARD

Why weren't you able to preview that trend? All our competitors seem to have reacted long before we did!

Why introduce BI? – primary motivations

- **Drive company strategy**
  - being able to connect strategising/planning to measuring of impact (do not manage «blindly»)
- **Growth and competitiveness:**
  - anticipate market trends and adapt R&D accordingly
  - better customer relationships through better-targeted offers
  - better leverage of customer potential (cross-/up-selling)
  - optimise business processes

business drivers

- **Single point of truth**
  - no by-pass reporting, consistent data
- **Cost reduction**
  - faster access to information
  - automation of reports, self-service BI
  - no interference of analytics with operational systems

technical drivers

Decision making

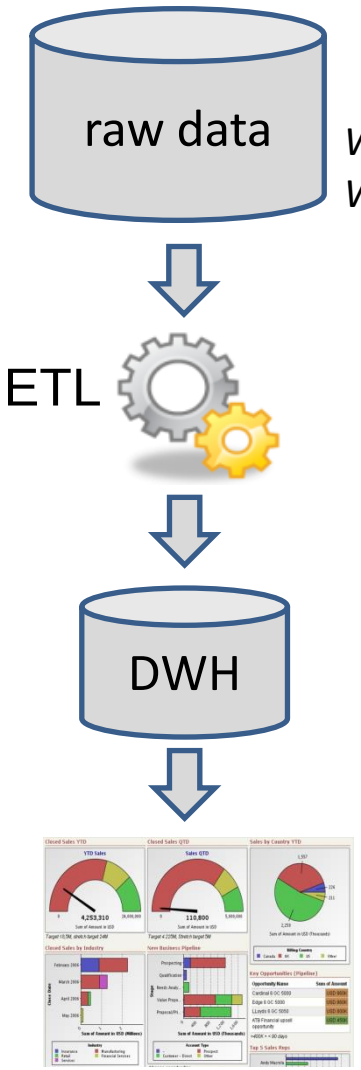
- **Decision making =** *The action of selecting among alternatives to achieve a goal*
  - ♦ each alternative leads to a different future
  - ♦ what is needed is the ability to predict the futures
- **Options:**
  1. predict based on gut feeling
     - cheap in the first place
     - risk of low-quality decisions
  2. Experiment with real system (try out)
     - too risky
     - too time-consuming (only one set of conditions at a time)
  3. Predict based on the past:
     - Data collection is time-consuming
     - difficult to determine when to stop and make a decision
     - Too little or too much information

Drivers for BI

BI and fact-based decision making

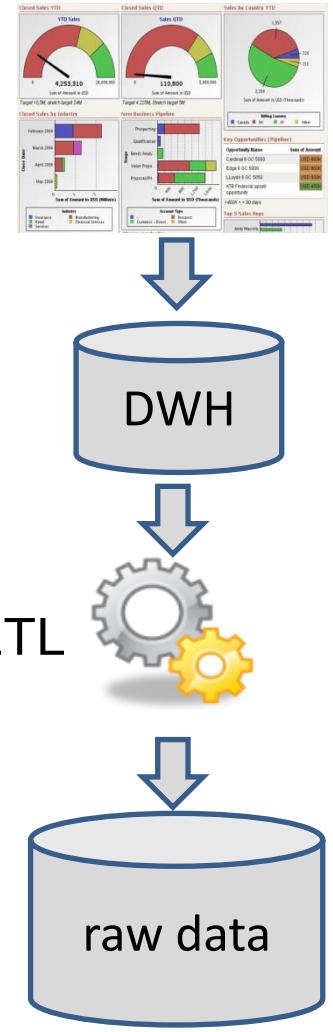
- Fact-based decisions are based on information
- BI supports decision making by providing that information, usually in the following way:
  - the human decision maker (HDM) formulates the decision problem
  - the HDM manually identifies the questions that need to be answered in order to take an informed decision
  - the HDM consults a BI tool to get the answers, usually by querying or browsing (e.g. OLAP)
  - the HDM uses the answers to take an informed decision

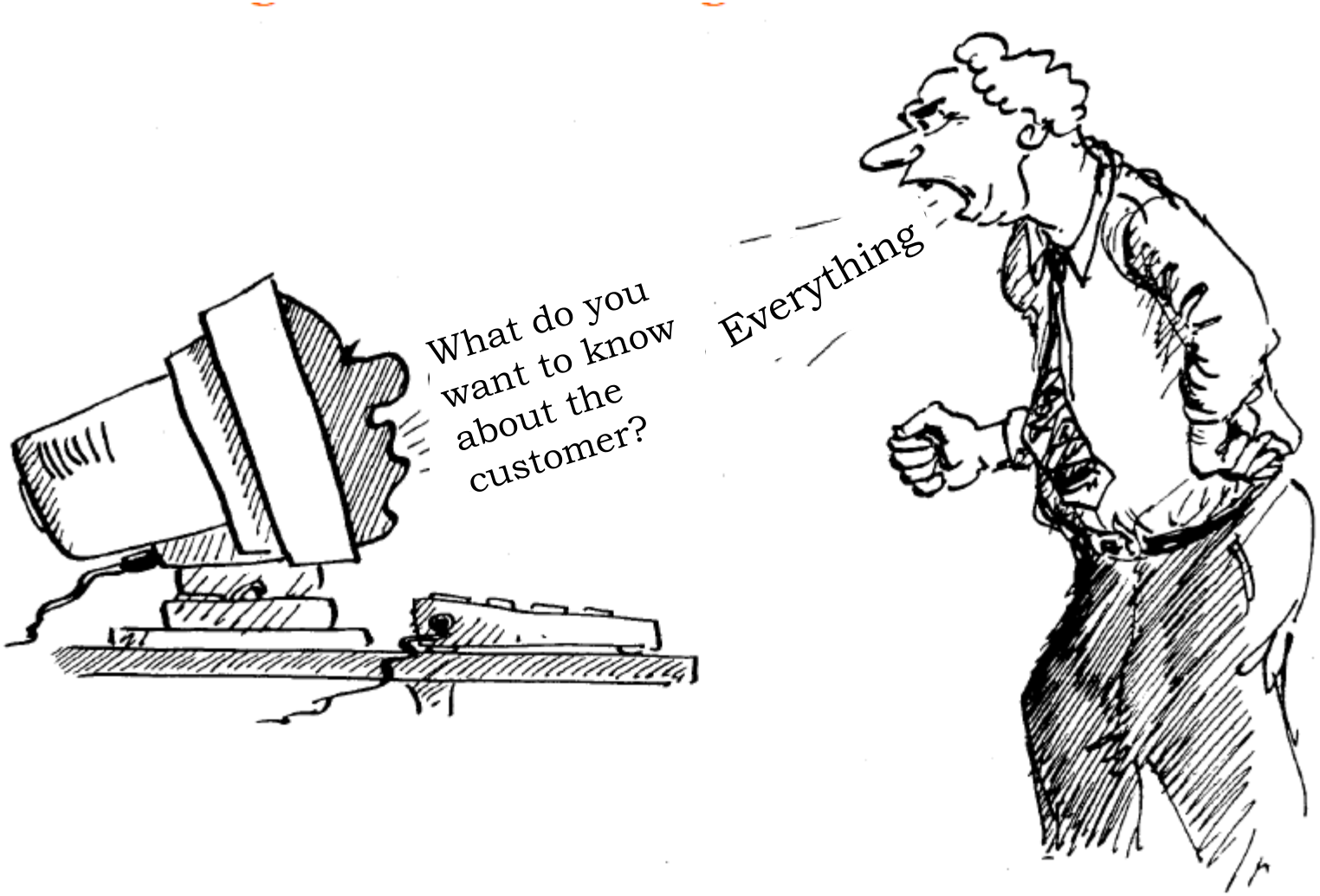
Data-driven vs. business-driven BI



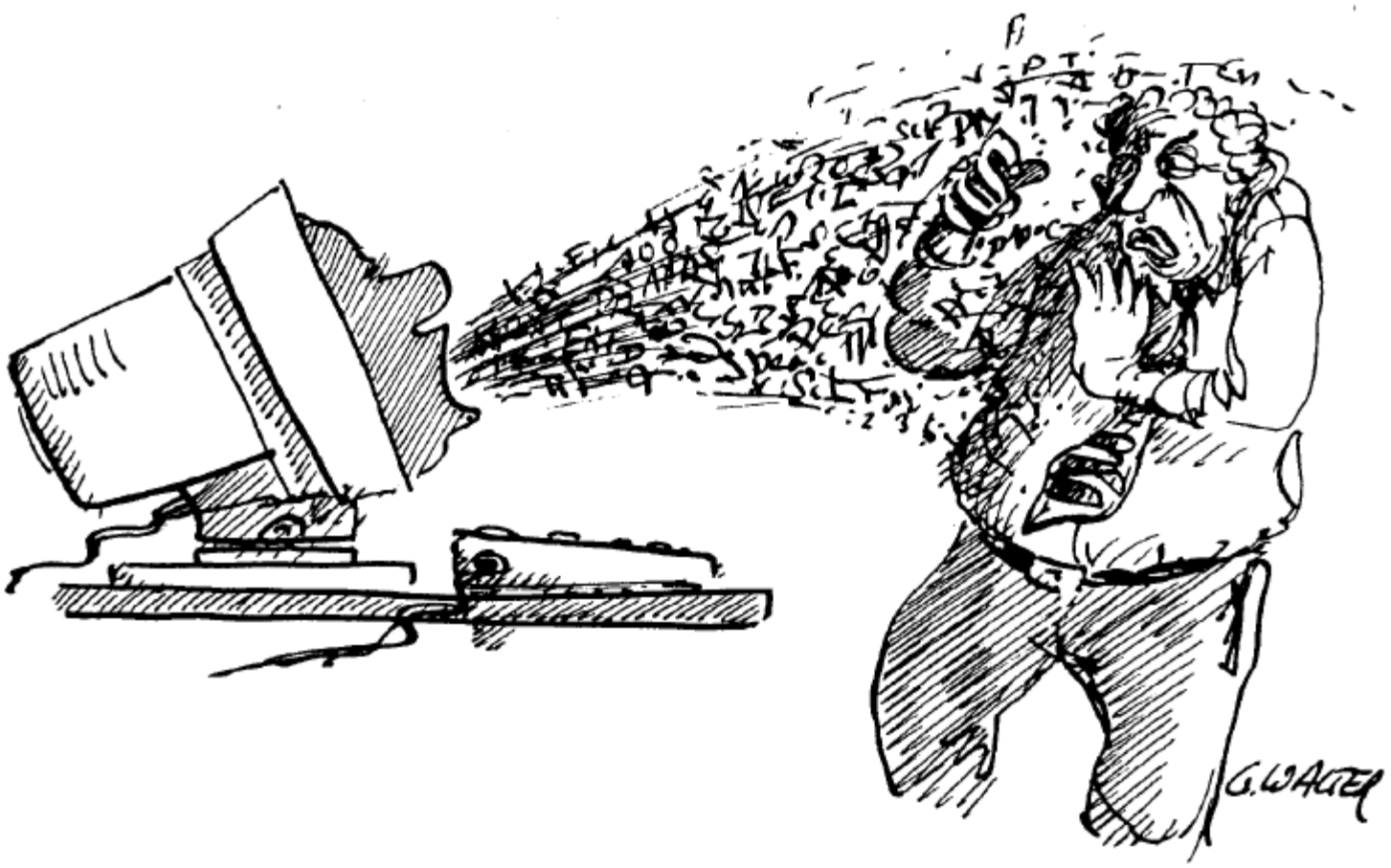
We have data.
What can we do with it?

We have questions.
How do we get answers?





adapted from slides by Dani Schnider



adapted from slides by Dani Schnider

Questions addressed by BI – task

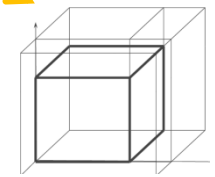
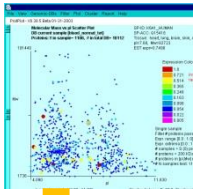
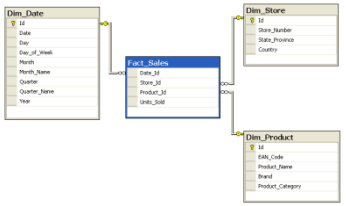
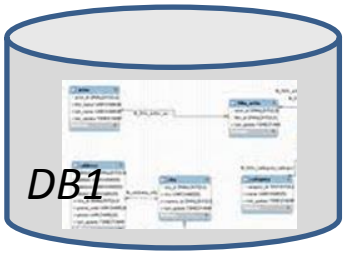
- «We have questions» may only be the second step
- **For the health insurance company**
  - **Where do the questions come from?**
    - identify decisions that (will) need to be taken
    - identify possible actions that can be taken as a result of decisions
  - **compile a list of questions that a BI system could answer**
    - Make sure they are really formulated as questions! (see the examples!)
    - Concentrate on specific (rather than very general) questions! Questions should be ones that can be answered based on data alone, not involving any judgment or argumentation. They should not start with «how»!
    - categorise these questions by assigning them to the different parts of the value chain
    - Which actions will possibly be taken as a result of knowing the answer to these questions

Where questions come from

- Generally speaking, companies need information to
  - monitor and improve performance
  - recognize and mitigate risks
  - recognize and seize opportunities

- All this can happen both on a strategic and an operative level

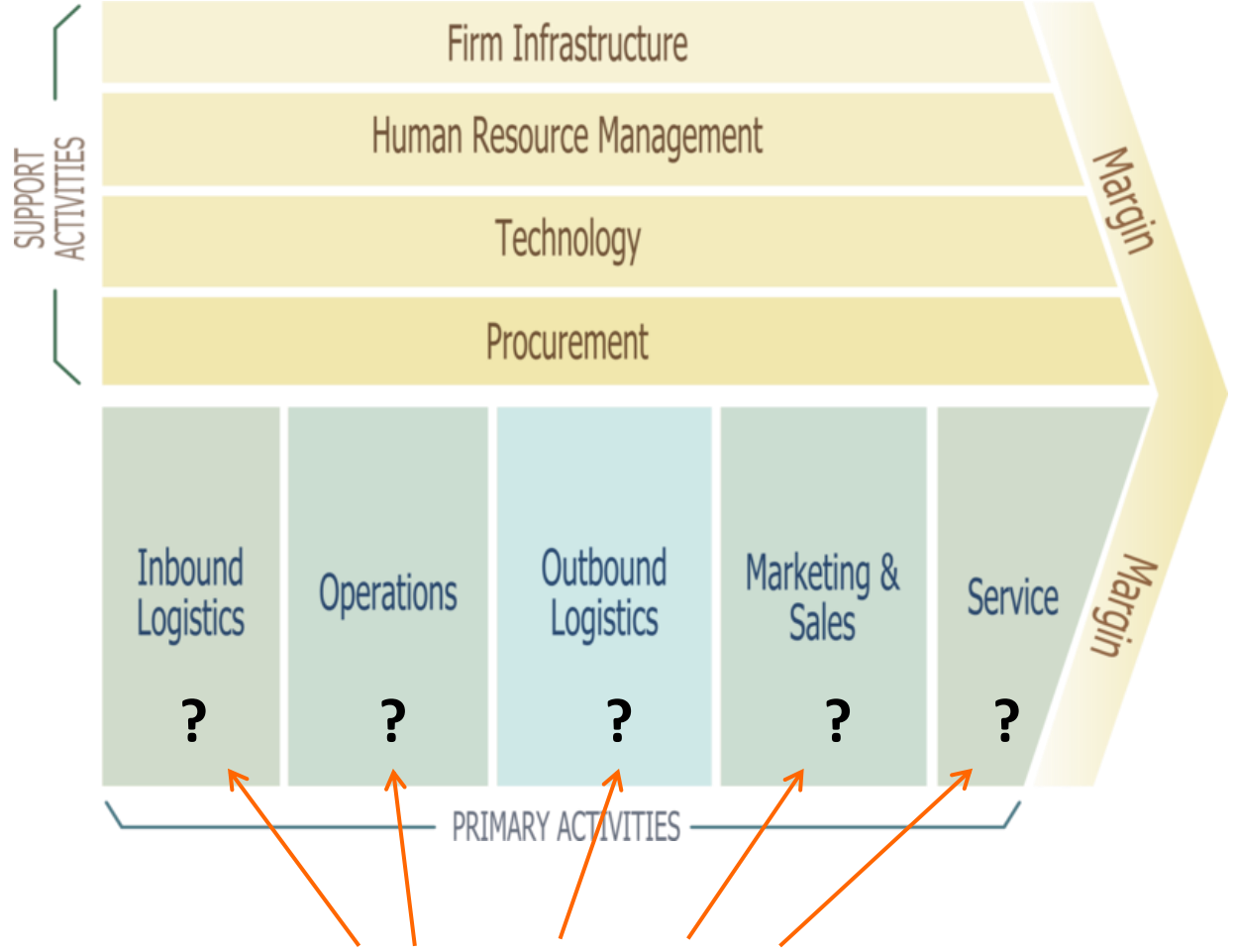
BI overview

	strategic	***operative***
Questions	• What are our goals? • Are we reaching our goals? • If not, where is the problem?	• Which customers are interested in the new product? • Which department has how many efforts that cannot be booked?
Analyses	 *measure, aggregate, visualise*	 *Ad hoc queries, OLAP*  *find patterns (data mining)*
	dimensional modelling	
	 **DWH**	
	ETL **ETL** **IE**	
raw data	 *DB1* *DB2*  	

Strategic decisions…

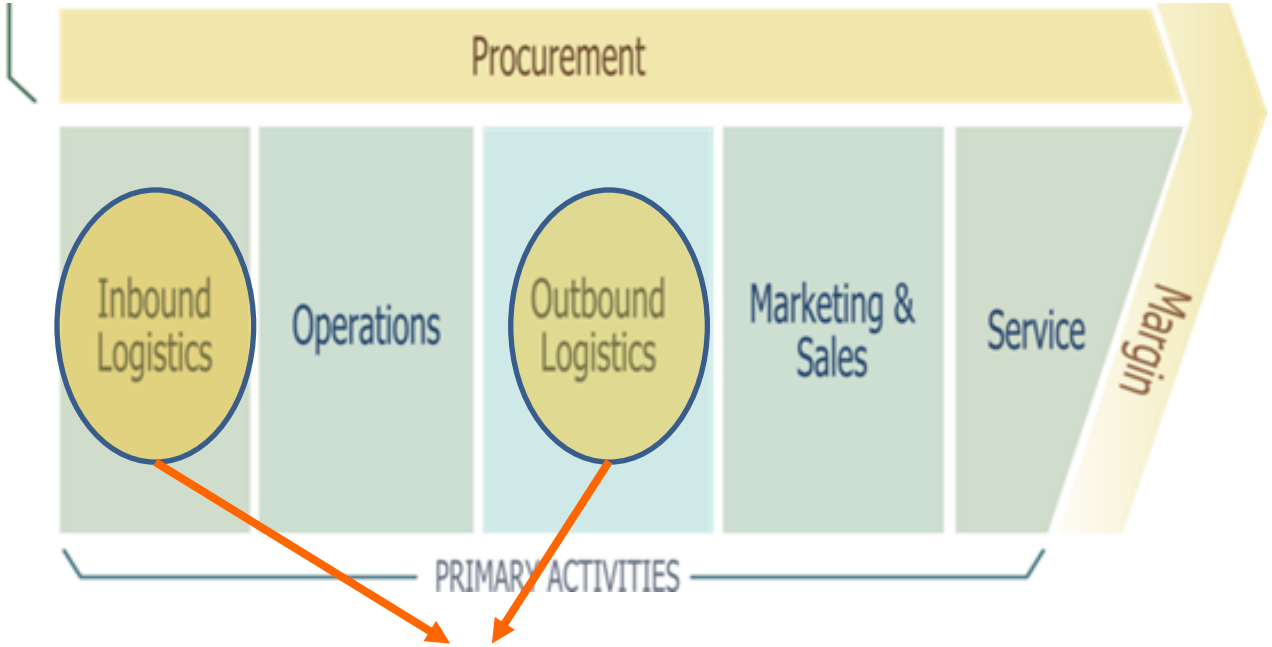
- Business Performance Management:
  - "how to perform better as a company?"
  - BI helps to achieve that by enabling measurement of achievement of strategic goals via KPIs
    1. Define **strategy**
    2. Define **goals**
       - e.g., identify key business processes to be improved, derive (concrete) strategic goals
       - for each goal, define KPIs and target values
    3. **Measure**
       - current values of KPIs (dashboard/cockpit)
       - analyse / compare current to targeted values
    4. **Decide**…
       - understand the (possible) deviation of KPI values from target!

Operative decisions: where BI creates value…



decisions to be taken in corresponding business processes?

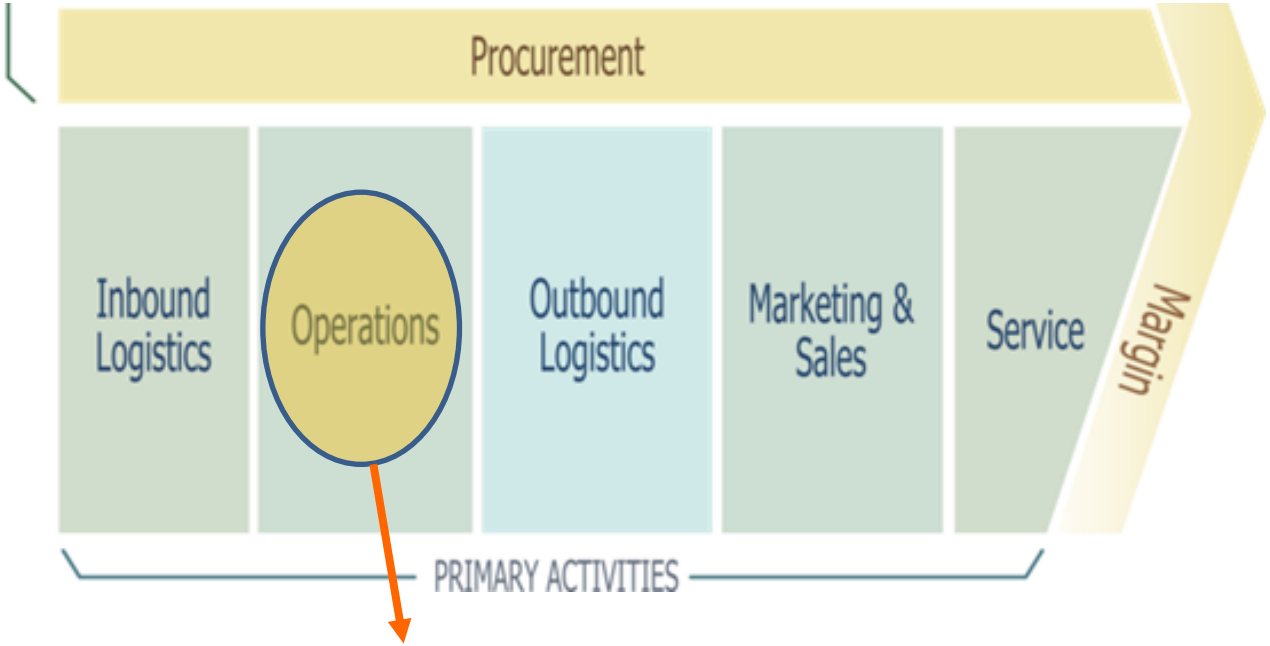
Operative decisions: where BI creates value…



Logistics:
the process of planning, implementing and controlling the efficient, effective flow and storage of goods, services and related information from the point of origin to the point of consumption for the purpose of conforming to customer requirements

- ***how to best use resources (inbound)?***
  - which parts to order, in which quantity, at what time, from which supplier?
- ***how to optimise processes (outbound)?***
  - which route/channel to use, how to schedule deliveries?

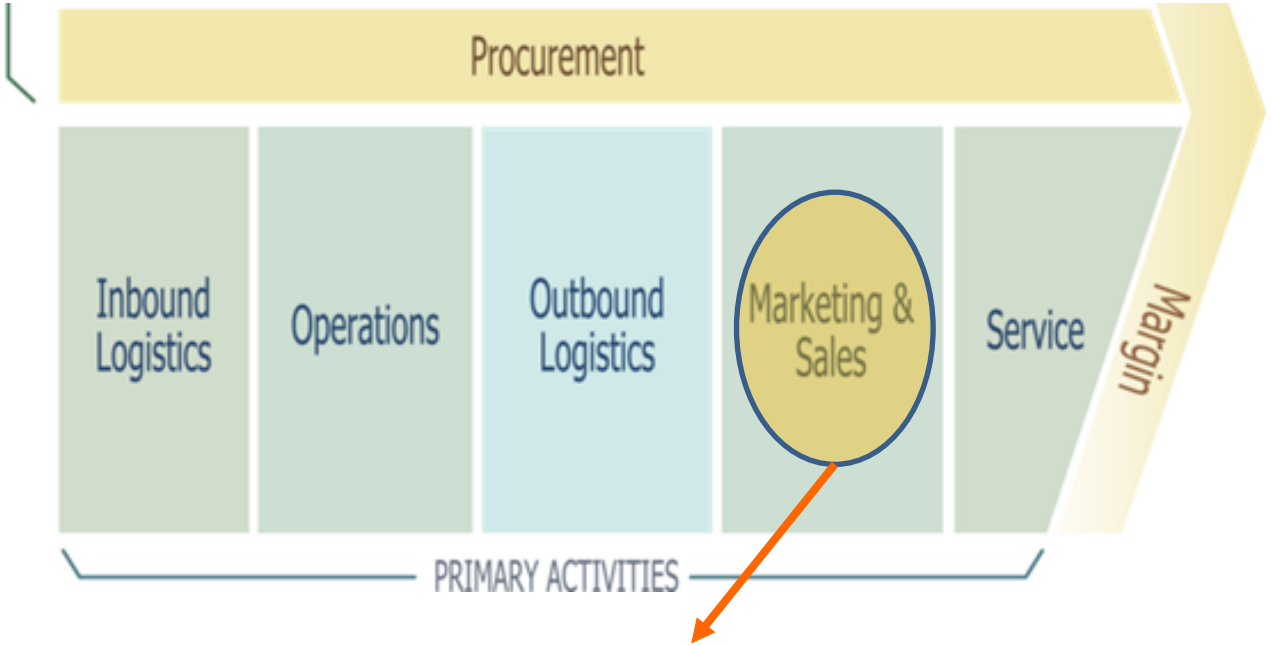
Operative decisions: where BI creates value…



Operations:
activities associated with the functions of transforming inputs into the final product form, such as machining, packaging, assembly, equipment maintenance, testing, printing, and facility operations.

- ***how to improve efficiency and effectiveness of processes?***
  - which resources to allocate, in which quantity, …

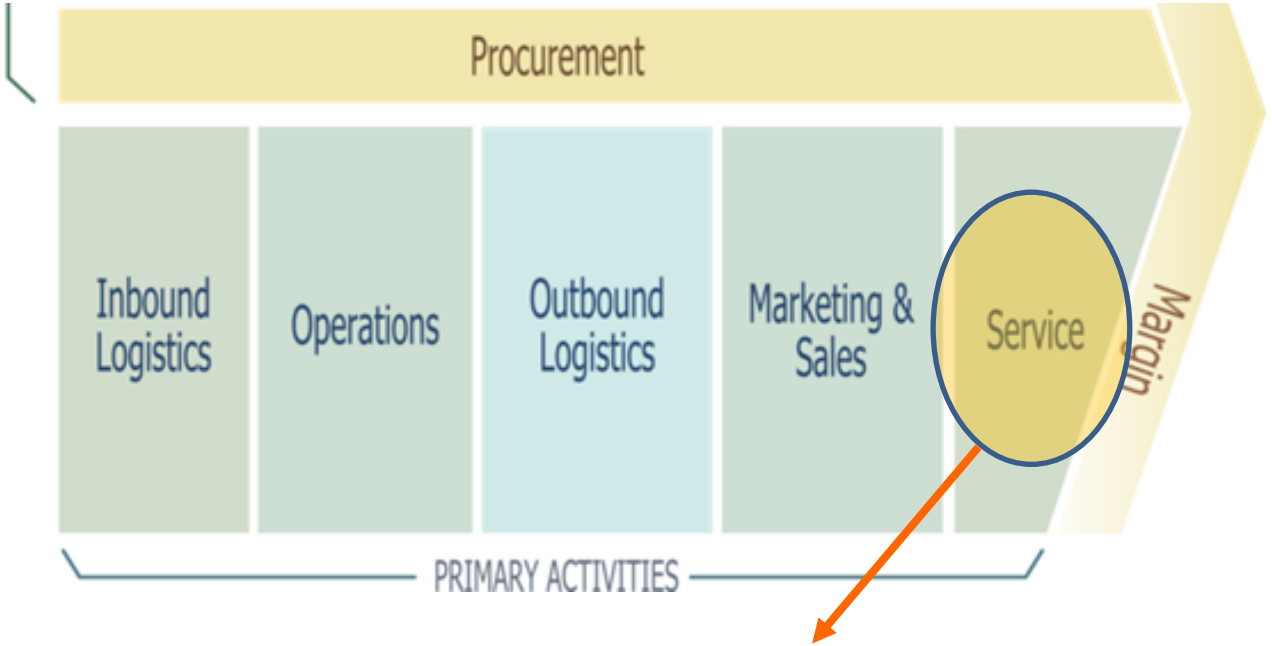
Operative decisions: where BI creates value…



Marketing/Sales:
activities associated with the functions of providing the means by which buyers can purchase the product and inducing them to do so, such as advertising, promotion, quoting, pricing, channel and sales force management.

- ***how to understand and best address the market?***
  - cross-selling: which offers to make?
  - where to place products in stores?
  - which customers to approach with a campaign?
  - client profitability: which customers to treat with special care?
  - pricing decisions

Operative decisions: where BI creates value…



Service:
activities associated with the functions of providing service to enhance or maintain the value of the product, such as installation, repair, training, parts supply, and product adjustment.

- ***how to meet customer requirements and anticipate problems?***
  - which distribution channels to use for service delivery?
  - which quality problems to address first?
  - Attrition prediction: which customers to retain with special offers?

Question types – summary

- Types of questions identified:
  - **query** for particular numbers or facts
    - *e.g. list of all policies that have been lost, list of all complaints, list of treatments that have been billed twice, list of high-value customers…*
  - **compute a measure or KPI by aggregating numbers**
    - *e.g. cost, margin, turnover, profitability*
  - **analyse KPIs / facts in different ways**
    - *e.g. sales/bookings by product/customer/sales rep/time*
    - *e.g. receipts/failures/stock by part/supplier*
    - *e.g. number of clicks/purchases by buyer/seller/page*
  - **predict**
    - *e.g. predict fraudulent transactions/claims*
    - *e.g. predict if a customer will buy a product*
    - *e.g. detect types of customers or types of complaints*

Monitor and improve performance

- Requirement on **strategic level**: be able to measure if strategic goals are achieved
  - e.g. be able to measure the satisfaction of our customers over the last year
    → so that we can decide to change our customer service model

- Requirement on **operative level**: monitor performance within certain business processes, in small time intervals
  - e.g. find out that/why (individual) customers are not satisfied today
    → so that we can decide to call them and find a solution

Recognise and mitigate risks

- Requirement on **strategic level**: be able to recognise general threats to our business
  - e.g. become aware that sales in certain product category are dropping dramatically (which is threatening our whole business)
    → so that we can revise our product portfolio

- Requirement on **operative level**: be able to recognise risks related to individual processes, customers, suppliers, employees, …
  - e.g. in telecommunications, be able to predict if a customer is going to cancel (or not renew) her contract
    → so that we can decide to make a special offer to that customer

Recognise and seize opportunities

- Requirement on **strategic level**: be able to recognise general opportunities for our business
  - e.g. become aware that (potential) customers are asking for a certain kind of product or product feature in social media
    → so that we can decide to develop such a product

- Requirement on **operative level**: be able to recognise opportunities related to individual process instances, customers, suppliers, employees…
  - e.g. recognise that we can cross-sell a certain product to an existing customer
    → so that we can decide to make the customer aware of that product

Specific requirements

When analysing requirements for a certain company's future BI solution, usually at least the following need to be fixed:

- Which are the strategic goals → which are our **KPIs**?

  e.g. revenue, delivery time, profitability, …

- By which criteria should KPI values be **grouped and/or filtered**?

  e.g. by customer, by sales rep, by region, by date, …

- Which **drill-down paths** should be possible?

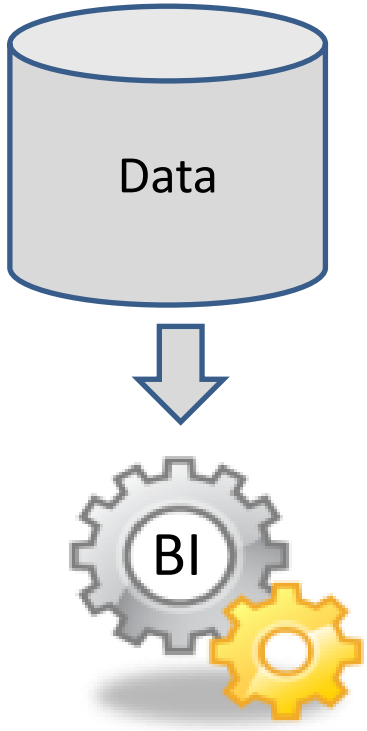
  e.g. date: year → quarter → month → …

  e.g. region: country → region → city

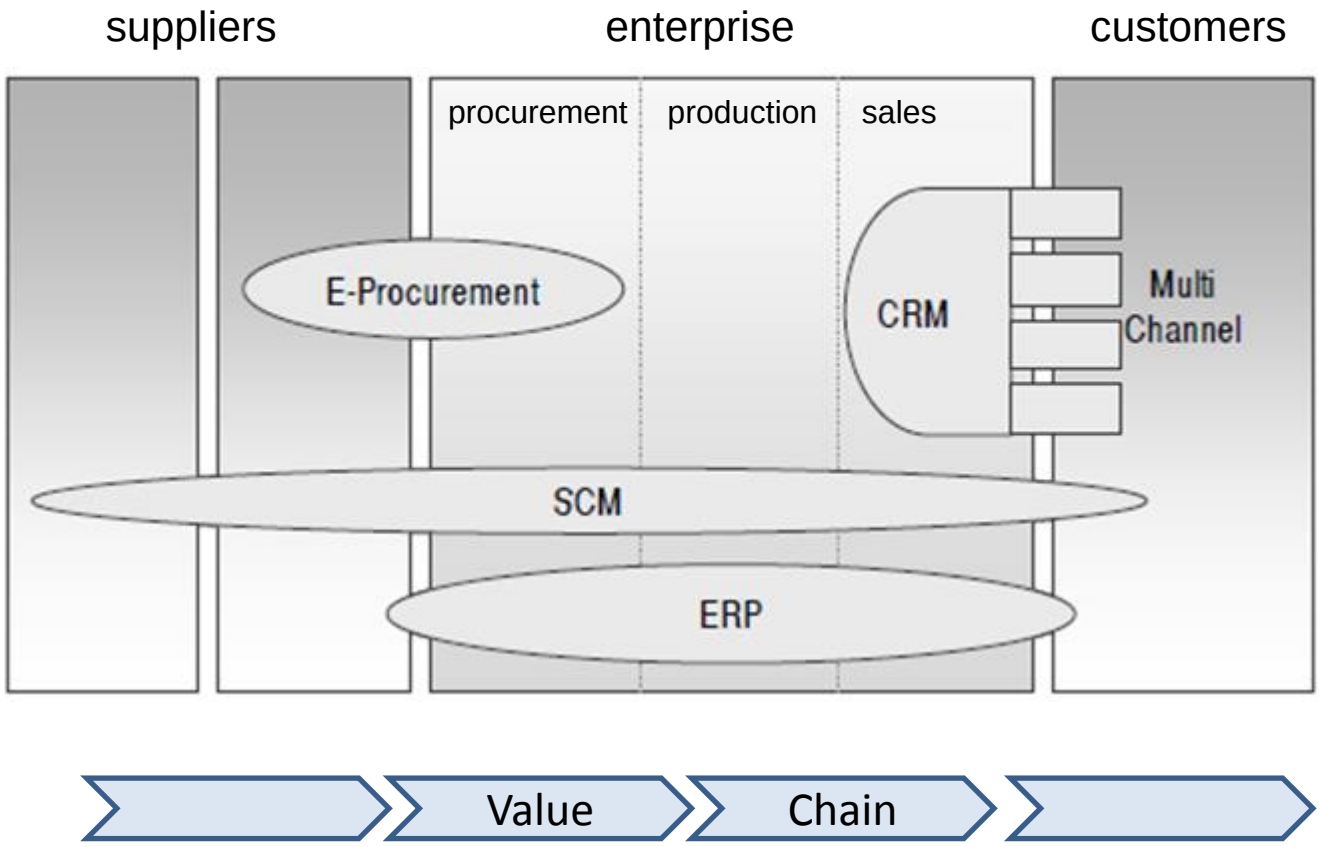
Data

Remember…

- … transform **raw data** into meaningful and useful **information**…
- **Raw data** is the starting point!



Where the data come from… (1)



adapted from Kemper et al. 2004

Where the data come from (2)

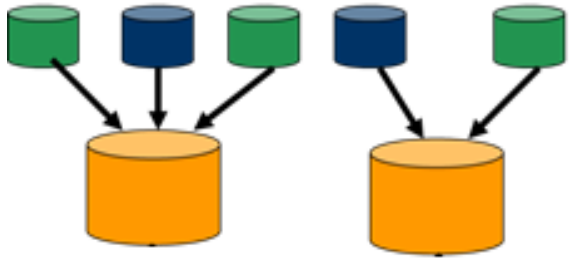
- Internal data sources:
  - (Transactional) standard business applications: SCM, ERP, CRM, …
  - Legacy databases
  - Web data: clickstreams from server logs, application logs
  - textual documents (from DMS, CMS, intranet, email,…)

structured-ness

- External data sources:
  - Web and web 2.0

BI tools – backend

- Observations:
  - ♦ many questions involve multiple (types of) data
  - ♦ sometimes the data can be expected to originate from more than one source system
  - ♦ for answering the questions, data from various sources needs to be connected
    - example: «Which is the best way to distribute product XYZ to customers?" → involves information about customers (e.g. profitability, behaviour) as well as about channels (e.g. cost of each channel)



Planning Data vs. Operative Data (1)

- **operative data:** generated by and used in processing operational transactions (on-line transaction processing, OLTP)
  - many concurrent users access and modify the same data
  - focus on transactions
  - example: booking/reservation systems
- **planning data:** used for decision support
  - read-only data

following Kemper et al. ch 2.1

Planning Data vs. Operative Data(2)

	Operative data	**Planning data**
users	clerk, IT professional	knowledge worker, decision maker
Function/goal	Support day to day operations (value adding business processes)	decision support
DB design	application-oriented	subject-oriented
data	current, up-to-date detailed information on business events, flat relational	historical, summarized, multidimensional integrated, consolidated
usage	Continuous, repetitive, concurrent	ad-hoc
access	read/write index/hash on primary key	lots of scans
Queries	Static, transactions embedded in application code	Ad-hoc, for changing information needs
# records accessed	tens	millions
#users	thousands	hundreds
DB size	100MB-GB	100GB-TB
metric	transaction throughput	query throughput, response

adapted from http://www.slideshare.net/idnats/data-warehousing-and-data-mining-presentation-725476

Data Warehouse – BI Backend

Data warehouse

- *A data warehouse is*
  - *"a **copy** of transaction data specifically structured **for querying and reporting" (Kimball et al. 2008)***

  or

  - *"an environment […] comprising a data store and […] tools for data extraction, loading, storage, access, query and reporting […] to **support decision-oriented management queries**" (Bashein/Markus, 2000)*

- 4 essential characteristics (Inmon 2005):
  - **Subject oriented:** data are organized around sales, products, etc.
  - **Integrated**: data are integrated to provide a comprehensive view
  - **Time variant:** historical data are maintained
  - **Nonvolatile:** data are not updated by users

Subject orientation

- Example: insurance

operational systems are organised around functional applications

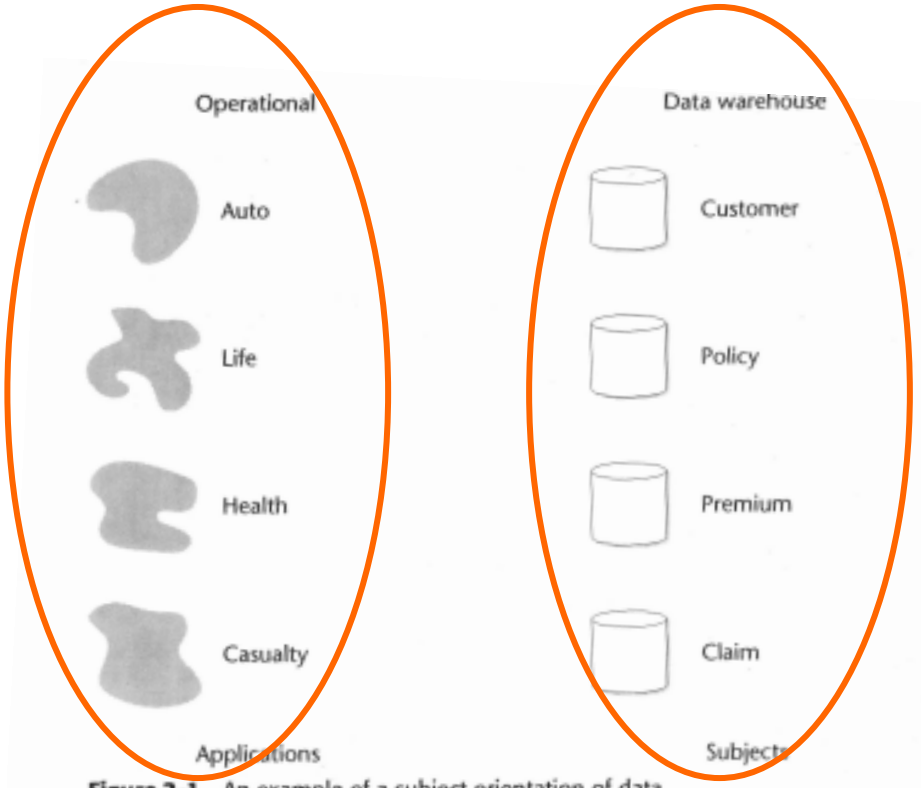


Figure 2-1 An example of a subject orientation of data.

analytical systems should be organised around the major subject areas of an insurance!

figure taken from B. Inmon: Building the data warehouse.

Integration, Time variance

- **Integration:** provide a «single version of the truth»
  - remove redundancy, inconsistency, semantic contradictions
  - The nasty bit… See ETL processes later
- **Time variance:** every record in the DWH has a specified moment or period when it was valid; data is collected over a long time period

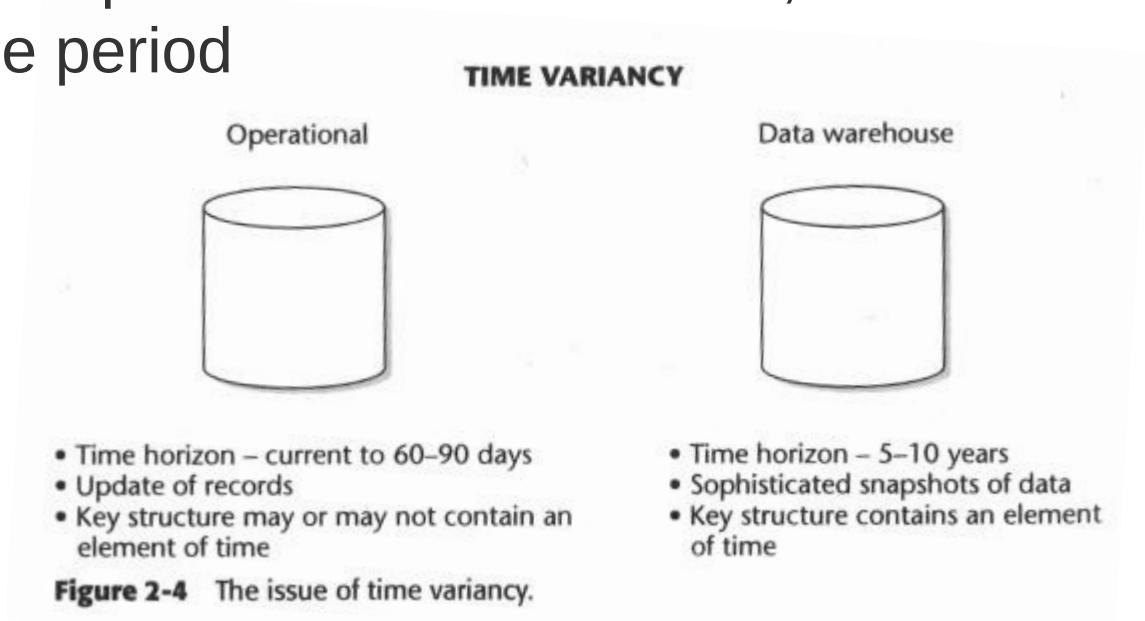


Figure 2-4 The issue of time variancy.

figure taken from B. Inmon: Building the data warehouse.

Non-volatility

- **Non-volatility:** data is not updated by end users on a regular basis
  - bulk loading, «read-only» access

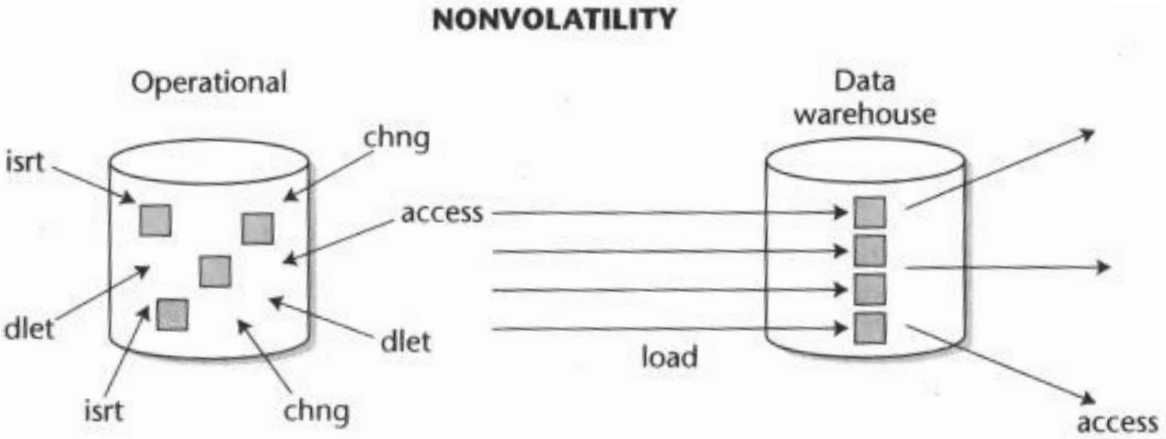


Figure 2-3 The issue of nonvolatility.

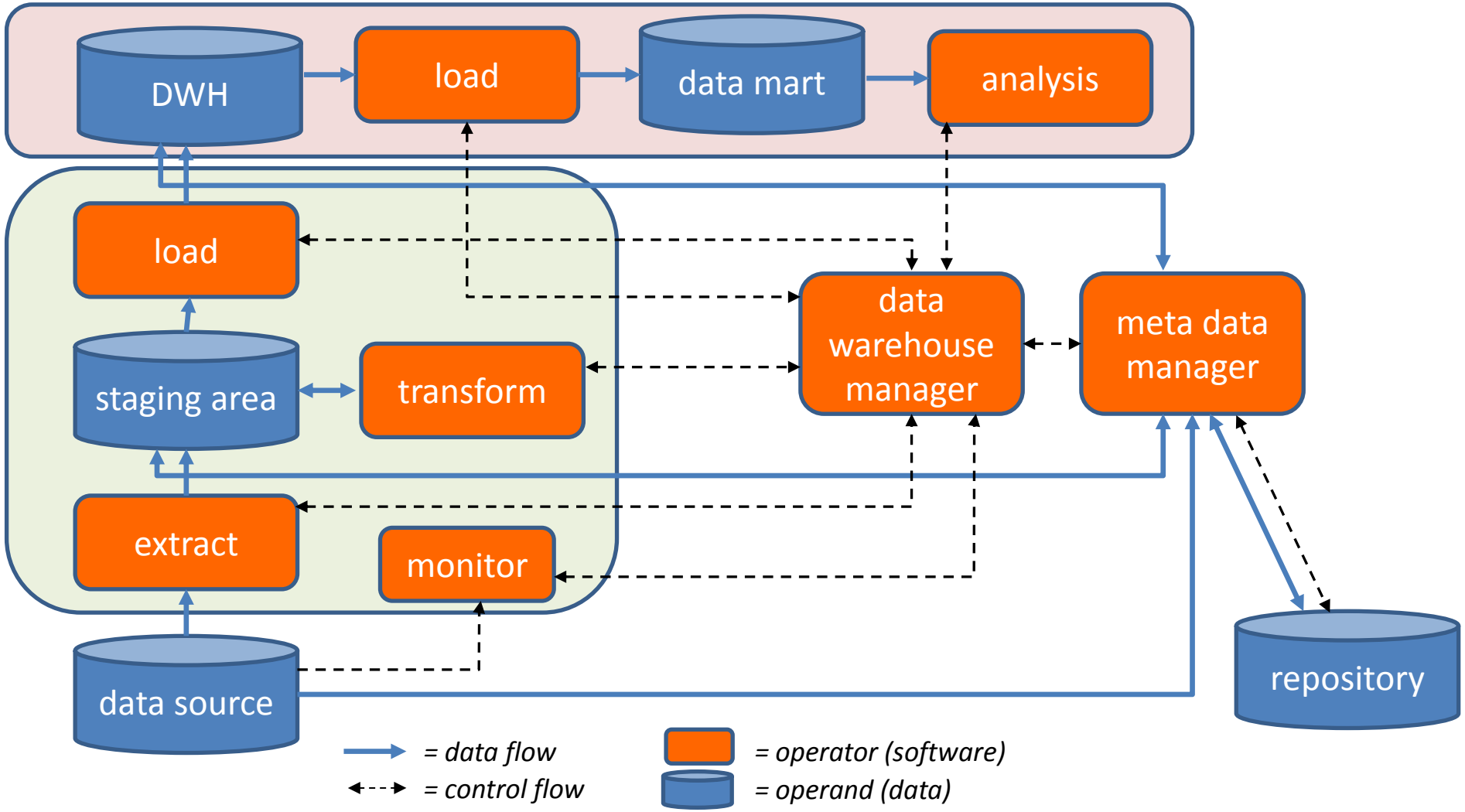
figure taken from B. Inmon: Building the data warehouse.

Data warehousing

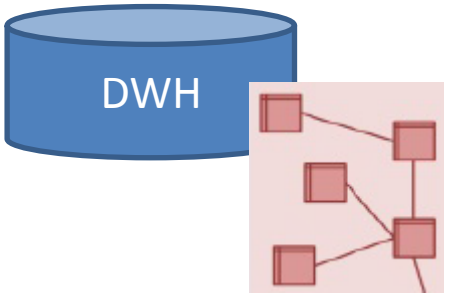
- *Data warehousing is the entire process of data **extraction, transformation, and loading** of data to the warehouse and the access of the data by end users and applications.*

following: http://www.terry.uga.edu/~hwatson/dw_tutorial.ppt

Reference architecture – overview

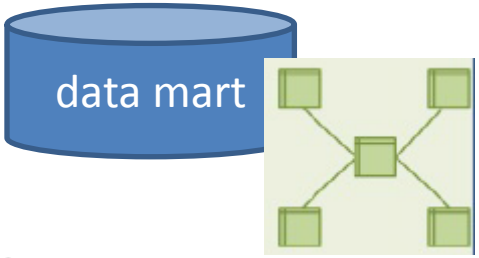


Reference architecture – DWH



- DWH = integrated data basis for all analyses
  - integration means both schema and data integration of various sources => «single point of truth» (no by-passes allowed!)
  - purpose: flexibility for re-using the data in multiple anlyses, no focus on (and hence no pre-aggregations for) particular types of questions
  - provides a collection of data that can be used to build data marts for specific analyses
  - not always present because of high cost: building data marts for specific analysis purposes directly from source data is often cheaper…

Reference architecture – data mart

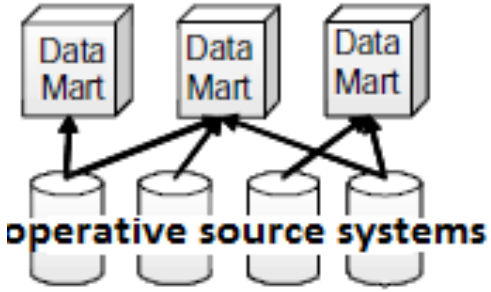


- each data mart is designed for specific analyses
  - data is extracted from data warehouse in a fast mode («bulk loading»), represents only a relevant subset of the DWH data,
  - transformed into target schema most suitable for the required analyses (very often a multidimensional model) and
  - sometimes aggregated
  - offers access to data as well as processing functionality (e.g. cumputation of sums, mean, variance,…) via query languages such as SQL or MDX

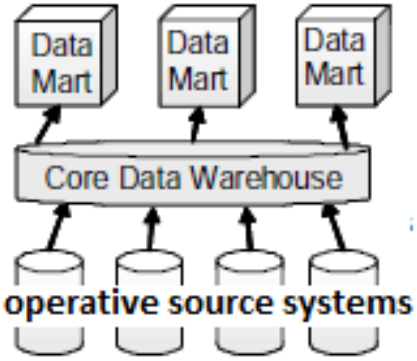
Data Marts

- A **data mart** stores data for a *limited* number of subject areas, such as marketing and sales data. It is used to support *specific* applications.
  - An **independent** data mart is created directly from source systems.
  - A **dependent** data mart is populated from a data warehouse.

independent

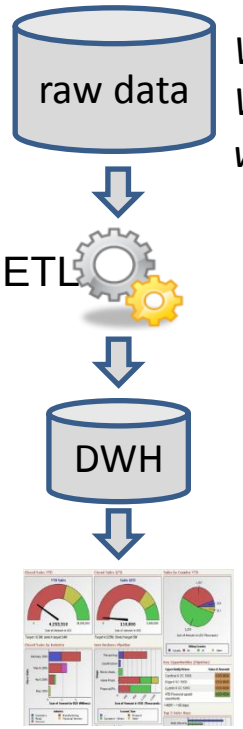


dependent



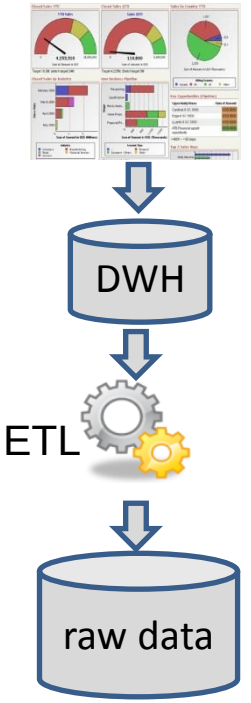
following Kemper et al. fig. 2.4

DWH architecture philosophies



We have data. What can we do with it?

We have questions. How do we get answers?

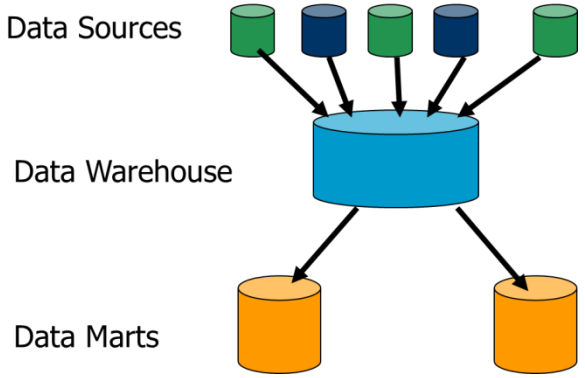


The requirements for the data warehouse cannot be known until it is partially populated and in use. […] Therefore, data warehouses cannot be designed the same way as the classical requirements-driven system. On the other hand anticipating requirements is still important. Reality lies somewhere in between.
(from W.H. Inmon: «Building the Data Warehouse»)

DWH architectures

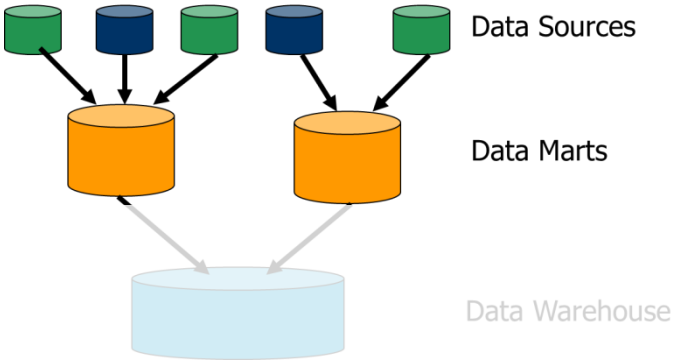
Enterprise Datawarehouse

- Advocate: Bill Inmon
- also called «hub-and-spoke»
- **strategy:** aggregate all enterprise data into one core DWH, derive dependent data marts as subsets as needed

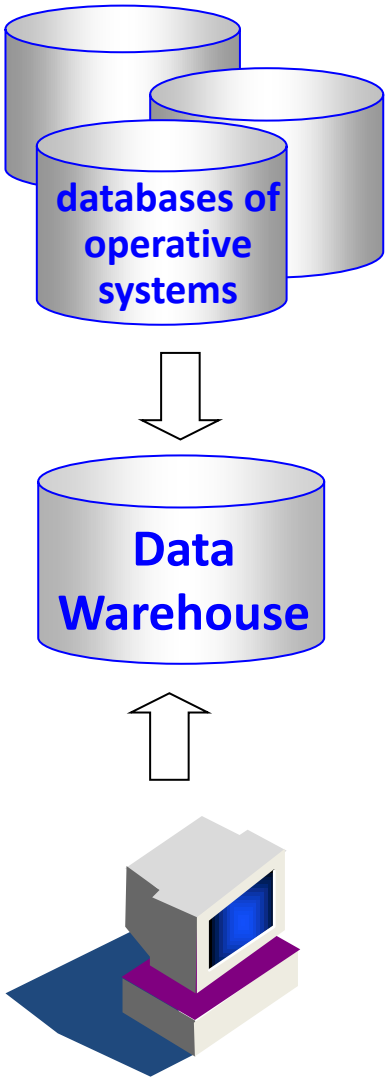


Independent data marts

- Advocate: Ralph Kimball
- often done in reality («historic reasons»)
- **strategy:** build data marts from source systems, re-use dimension tables where possible. Possibly join marts into a centralised DWH later



Data Warehouse



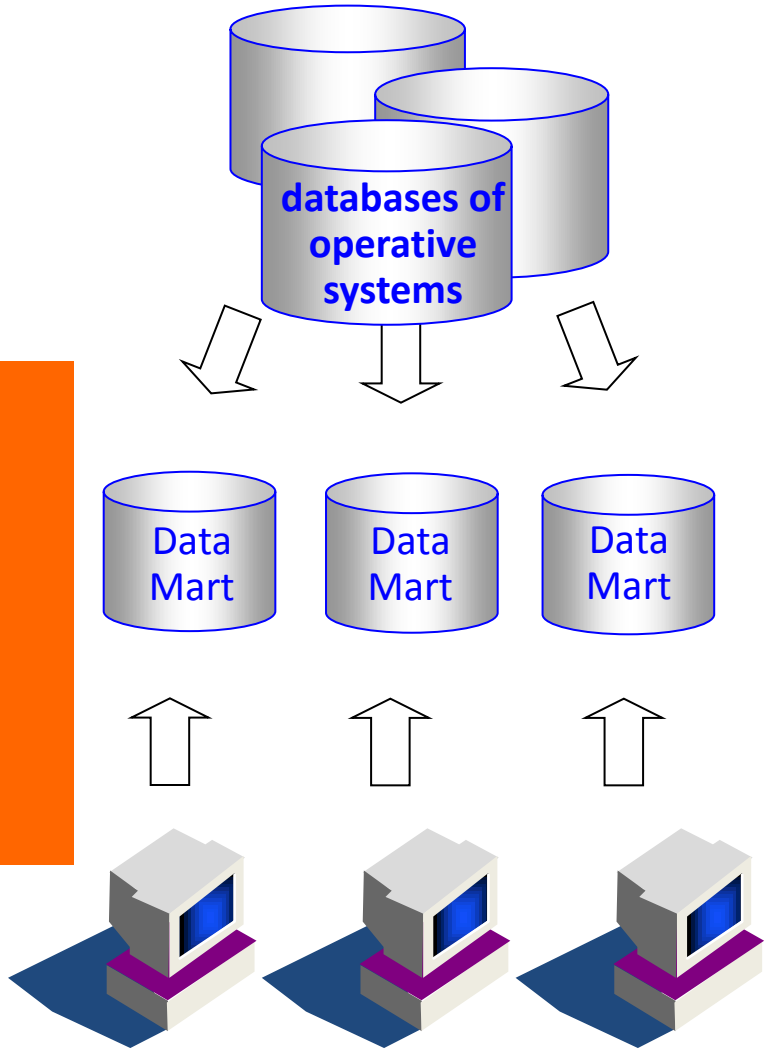
A Data Warehouse is a database which supports strategic decisions by providing ...

- ✓ high-volume and
- ✓ regular excerpts from
- ✓ operative databases
- ✓ by periods and
- ✓ often aggregated[1]
- ✓ also for ad hoc[2] analysis

Source: [Lusti, 1999]

1) combined, consolidated (e.g. als sum, average, indicators)
2) without preparation, in contrast to standardized analysis

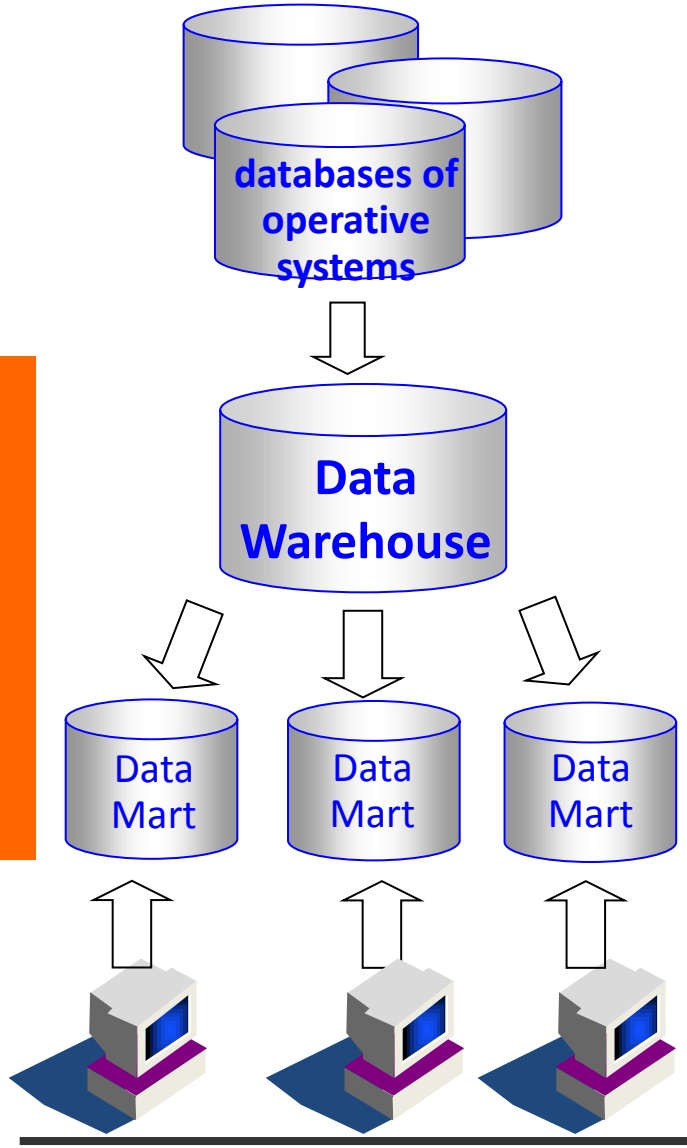
Modularisation: Enterprise Data Mart



- Distribute data to several Data Marts
- Advantages:
  - data model easier to understand
  - efficient access
- Problem: overall analyses
- Coordination:
  - Loading cycles
  - Data model:

    Attributes with equal meaning meaning should have the same identifier , key, datatype in all data marts

Hierarchical Architecture



- Coordination of local Data Marts by an Enterprise Data Warehouse (EDW)
- Objective of the EDW:
  - Extraction, integration and distribution of the data
  - integrated data model
- Objective of the Data Marts:
  - Queries and analysis on parts of data
  - Adaptation to needs of an organsation unit or process
- Data distribution
  - time-based
  - event-driven
  - on demand

Data Marts: Departmental vs. Enterprise?

- Question: should data marts be enterprise-wide or departmental?
  - Answer 1: Data marts should be organised around business processes (orders, invoices,…), not department boundaries!
  - Answer 2: … but they don't necessarily have to be enterprise-wide (depends on the business process)!

Reference architecture – ETL

- Extract
  - read source data into staging area as indicated by monitor
    - e.g. from file created by triggers or tables created by replication services
  - control the selection of data that should be copied
  - Extractions can happen…
    - …periodically
    - …on human request
    - …event-based (e.g. when a certain number of changes has occurred)
    - …upon each change
- Transform: data cleaning and schema integration (more on this later)
- Load: copy transformed data into DWH

ETL – Extract,Transform, Load

ETL process

- The process of
  - **extracting** relevant data from source systems (e.g. transaction-based ones)
  - **transforming** the data into the target format defined for the DWH or data mart
  - **loading** the data into the DWH

- The nasty, time-consuming and hence costly bit of data warehouse design
  - => do not underestimate the possible dirtiness of data!!!

ETL process – transformation tasks

- **Transformation =** adapting data, data quality and schemas to the requirements of users
  - **Filtering:** remove syntactic and semantic defects of data
  - **Harmonisation:** map source schemas to the target schema of the DWH
    - syntactic harmonisation: schema integration + data integration
    - business harmonisation
  - **Aggregation:** aggregate data along dimension hierarchies (e.g. «customer», «customer segment», «all»)
  - **Enrichment:** pre-compute values of frequent interest and store as new attributes
    - on the basis of harmonised/aggregated data

ETL - Harmonization

- These are parts of tables that should be integrated in a DWH. What harmonisation asks/problems do you see?

CustomerID	Name	City
11	Peter	Rom
15	Paul	Camerino
18	Mary	Olten
25	Joe	Bern

PurchaseID	CustomerID	Date	ProductID
1002	11	5 May 2015	SE4256
1003	18	5 May 2015	EA4516
1004	11	6 May 2015	EA4516
1005	25	6 May 2015	RG3452

ComplaintID	Complaint	Person
36536	Return	George
44363	Failure	Paul
46344	Failure	John

ETL – Harmonisation: Schema integration

Problem	characteristics	Example: data source 1	Example: data source 2	Solution
Synonyms	Attributes with different names have identical meaning	Attribute «employee» contains employee name	Attribute «staff» contains employee name	Choose an attribute name
Homonyms	Same attribute name refers to attributes with different meaning	Attribute «partner» refers to name of customer	Attribute «partner» refers to name of supplier	Choose different attribute names

ETL – Harmonisation: data integration (1)

Problem	characteristics	Example: data source 1	Example: data source 2	Solution
Deviating primary keys (synonyms)	Same entity has different id in different operational DBs	Customer «Smith» has id 376_ACC in accounting application	Customer «Smith» has id 7843_CC in call center application	Record linkage: identify identical entities via overlapping attribute values; use mapping table

- How to detect entity identity?

ETL – Harmonisation: data integration (2)

- Mapping tables: allow to map updates in sources to DWH records

AD_SYS	...	customer	LOADTIME
AD-FX8257		Müller	31DEC2009:23:03:08
AD-FH2454		Meier	31DEC2009:23:03:08
AD-FX7059		Schulz	31DEC2009:23:03:08
AD-FT2567		Schmitz	31DEC2009:23:03:08
...	...	...	...

AC_SYS	customer	customerStatus
3857_ACC	Müller	A
3525_ACC	Meier	A
3635_ACC	Schulz	A
3566_ACC	Schmitz	B
...	...	...

CC_SYS	cust_grp	customer	LOADTIME
59235395	retail	Müller	31DEC2009:23:03:08
08485356	industry	Meier	31DEC2009:23:03:08
08555698	industry	Schulz	31DEC2009:23:03:08
85385386	retail	Schmitz	31DEC2009:23:03:08
...	...	...	...

AD=customer service
CC = call center
AC = accounting

Kunde_ID	cust_id	...	AD_SYS	CC_SYS	AC_SYS	...	LOADTIME
0001	Müller		AD-FX8257	59235395	3857_ACC		31DEC2009:23:03:08
0002	Meier		AD-FH2454	08485356	3525_ACC		31DEC2009:23:03:08
0003	Schulz		AD-FX7059	08555698	3635_ACC		31DEC2009:23:03:08
0004	Schmitz		AD-FT2567	85385386	3566_ACC		31DEC2009:23:03:08
...	...	...	...	...	...	...	...

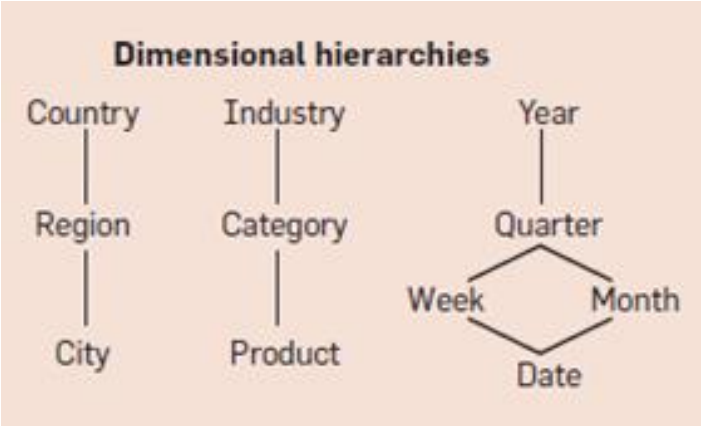
adapted from Kemper et al.

ETL – business harmonisation

- **adjust figures/values**
  - consolidate figures from various databases based on their (business) meaning, e.g. apply rules to map location- or department-specific value deviations
  - convert currencies and units (e.g. inch → cm)
- **adjust granularity**
  - decide for a level of granularity (e.g. monthly or quarterly)
  - harmonise according to period (source systems may have differing granularity, e.g. quarters vs. years)
  - aggregate all values on that level (e.g. sum all records/receipts of one day together)

ETL - Aggregation

- Aggregate data based on dimensional hierarchy
  - ♦ usually, aggregates are pre-computed for performance reasons
  - ♦ introduces «controlled redundancy»
  - ♦ aggregates become invalid when hierarchies and/or source data change…

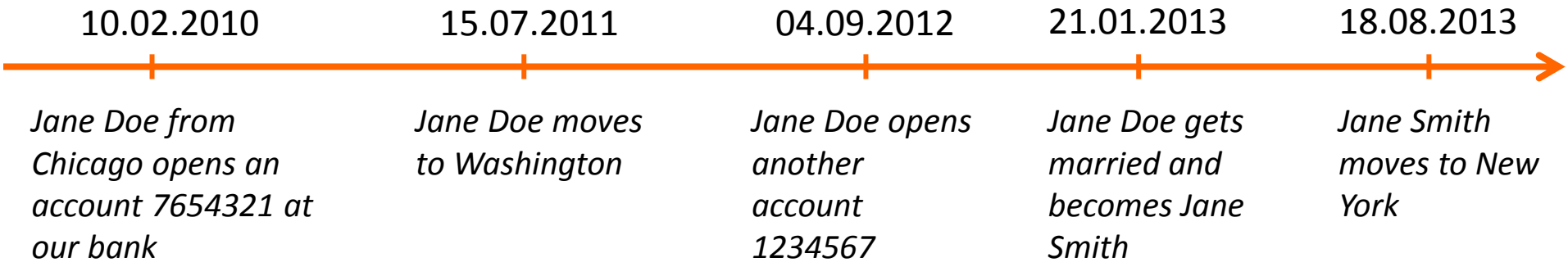


ETL - Enrichment

- add new attributes that are functions of existing data; compute these functions and store the result
  - sums, averages or more complicated computations (e.g. profitability)
  - based on harmonised and/or aggregated data
  - same motivation as aggregation: performance
  - introduces another «controlled redundancy»

Slowly Changing Dimensions

Example: customer dimension change



- who's the owner of the bank account 1234567?
  - as of today: Jane Smith from New York
  - as of 31.12.2012: Jane Doe from Washington
  - as of 31.12.2011: there is no such bank account

Type I: no history

Cust_id	Cust_name	Cust_city	…
1	John Allan	Chicago	…
2	Chris Lee	Boston	…
3	Jane Doe	Chicago	…

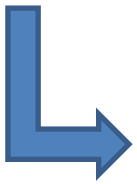
Cust_id	Cust_name	Cust_city	…
1	John Allan	Chicago	…
2	Chris Lee	Boston	…
3	Jane Doe	Washington	…

15.07.2011

old value is simple overwritten with new value

Type II: full history

Cust_id	Cust_name	Cust_city	Valid from	Valid to
1	John Allan	Chicago	10.03.2008	
2	Chris Lee	Boston	02.06.2010	
3	Jane Doe	Chicago	10.02.2010	



Cust_id	Cust_name	Cust_city	Valid from	Valid to
1	John Allan	Chicago	…	
2	Chris Lee	Boston	…	
3	Jane Doe	Chicago	10.02.2010	14.07.2011
3	Jane Doe	Washington	15.07.2011	20.01.2013
3	Jane Smith	Washington	21.01.2013	17.08.2013
3	Jane Smith	New York	18.08.2013	

18.08.2013

- every intermediate state is documented, validity range of values is signalled via «valid from», «valid to» attributes
- «valid from» becomes part of primary key

Type III: limited history

Cust_id	Previous Cust_name	Current Cust_name	Effective date cust_name	Previous Cust_city	Current Cust_city	Effective date cust_city
1		John Allan			Chicago	10.03.2008
2		Chris Lee			Boston	02.06.2010
3		Jane Doe			Chicago	10.02.2010

18.08.2013

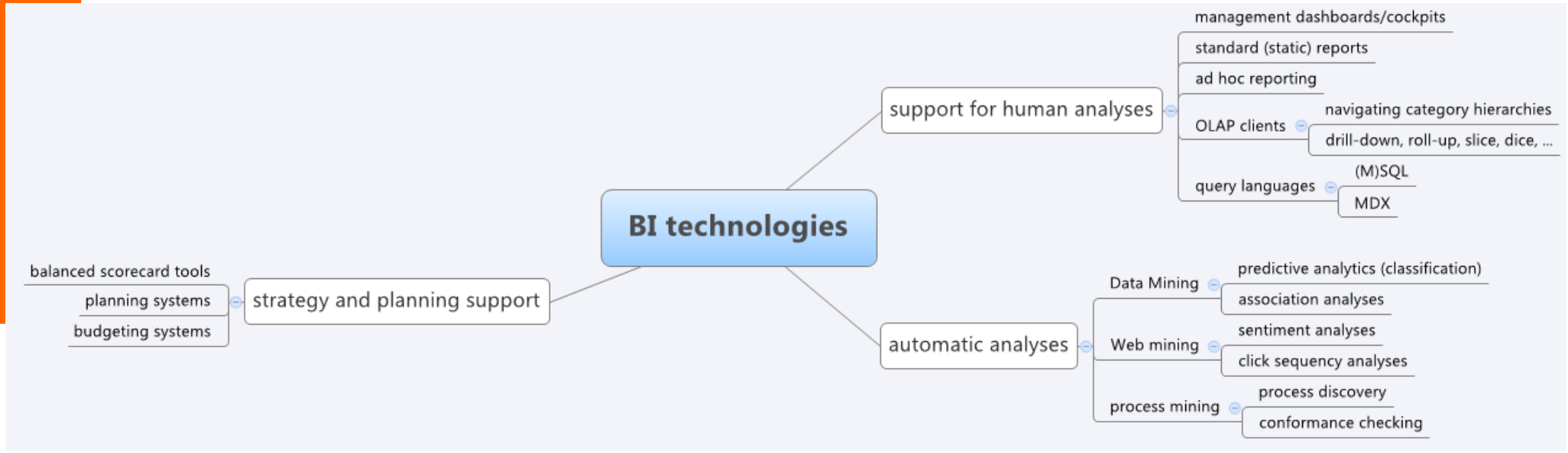
Cust_id	Previous Cust_name	Current Cust_name	Effective date cust_name	Previous Cust_city	Current Cust_city	Effective date cust_city
1		John Allan			Chicago	10.03.2008
2		Chris Lee			Boston	02.06.2010
3	Jane Doe	Jane Smith	21.01.2013	Washington	New York	18.08.2013

- keeps the n previous values, each in a separate new column (in the example: n=1)
- effective date column(s) show(s) when the change occured

Analysis und Use of Data – BI Frontend

BI front-ends

- contribution of systems ranges from…
  - ♦ support of strategic planning, over…
  - ♦ simple aggregation of data (e.g. in dashboards/reports) to…
  - ♦ complex statistical analysis, data mining



Analysis: BI tools – front-end

1. strategic:
   - ♦ support definition of KPIs and their connection to data
   - ♦ offer possibilities to monitor status of indicators
   - ♦ offer a possibility for drill-down, e.g. when indicators are off-target, to understand cause for deviation

2. operative:
   - ♦ support ad-hoc aggregation of numbers e.g. to make decisions about a particular product, customer, supplier or sales rep
   - ♦ support queries for particular facts, needed to make decisions in core business processes
   - ♦ predict (mainly) customer behaviour to optimise sales and marketing strategies, find patterns

BSC tools

dashboards, reports

OLAP

OLAP

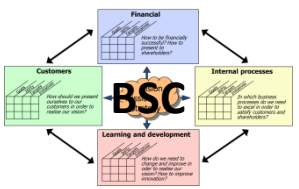
query tools
CBR

data mining tools

BI front-ends and the questions they answer

level of aggregation

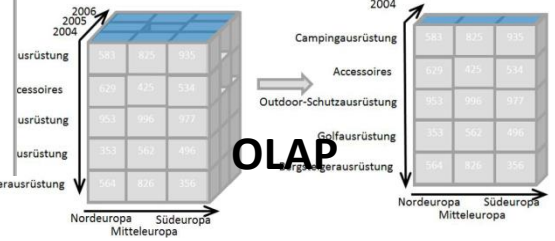
depth of analysis



Set goals and measure…

1. definition of KPIs

2. KPI measurement

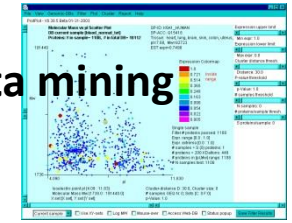


Understand why…

3. group by …

4. filter by …

5. drill-down

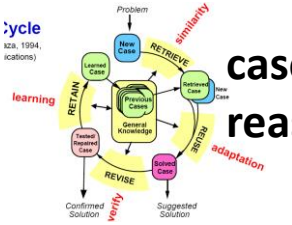


Find new patterns

6. predict …

7. find patterns …

rule-based systems



case-based reasoning

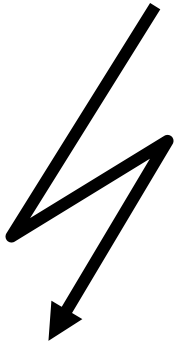
Make decision

8. apply rules

9. reuse cases

Objectives of Data Modeling

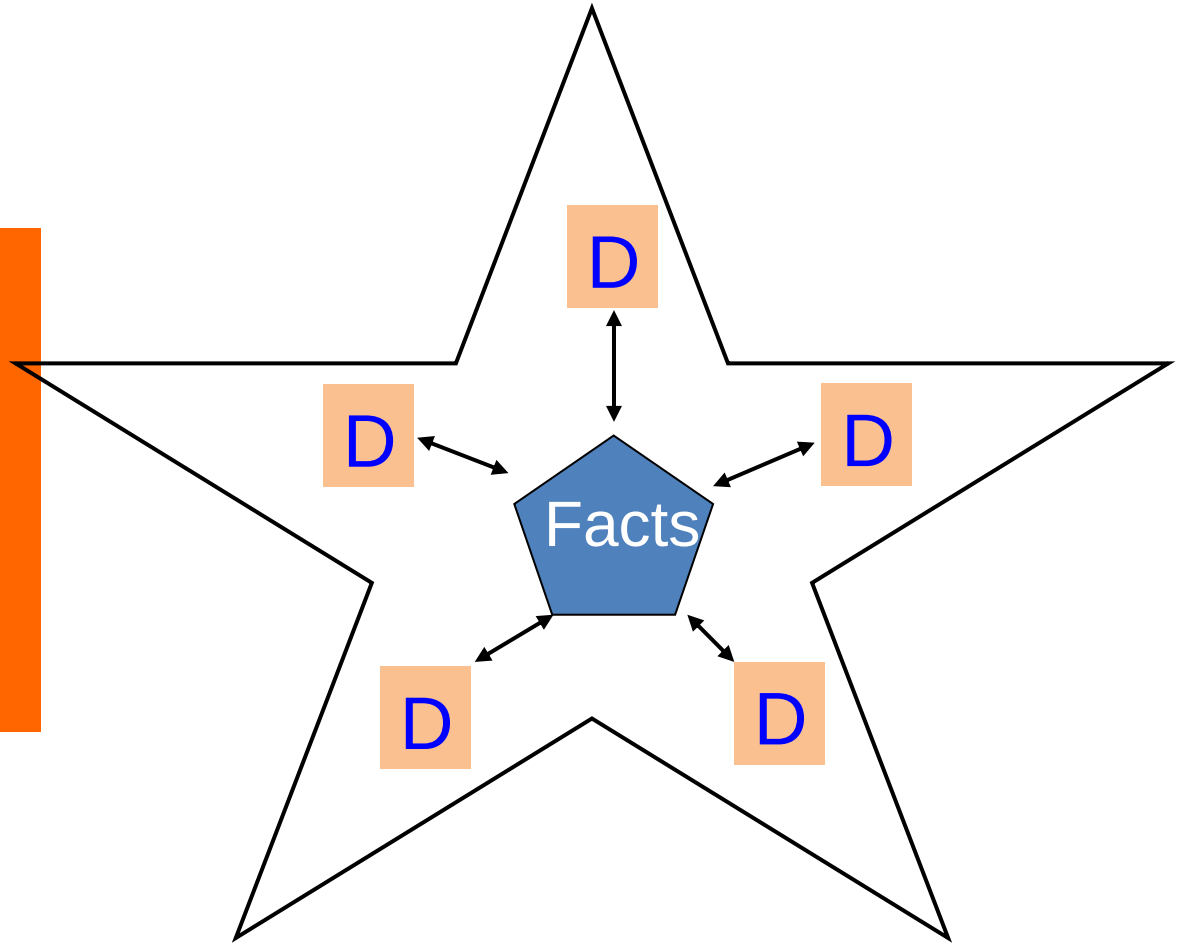
- **operative databases**
  - **Transaction processing** by high granularity
  - **Progress friendliness** by normalisation
  - **Storage efficiency** by normalisation
- **Data Warehouses**
  - **user friendliness** by low granularity
  - **query efficiency** by denormalisation
  - **close to business** by indicators and dimensions



➢ Data model of operative databases cannot be reused for data warehouses

cf. [Lusti, 1999]

Star Schema for Relational Data Warehouses /Marts to support OLAP

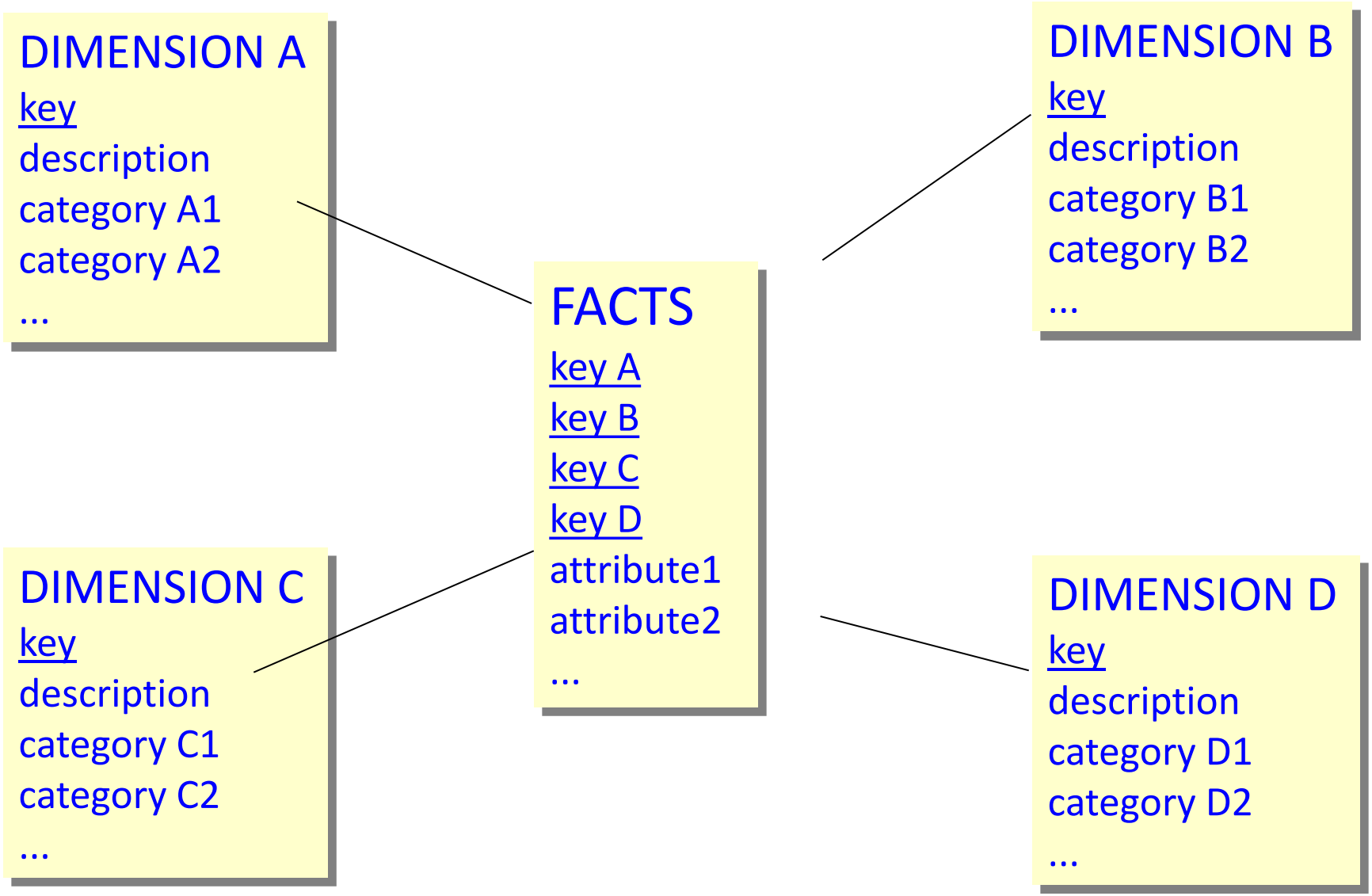


Star Schema:

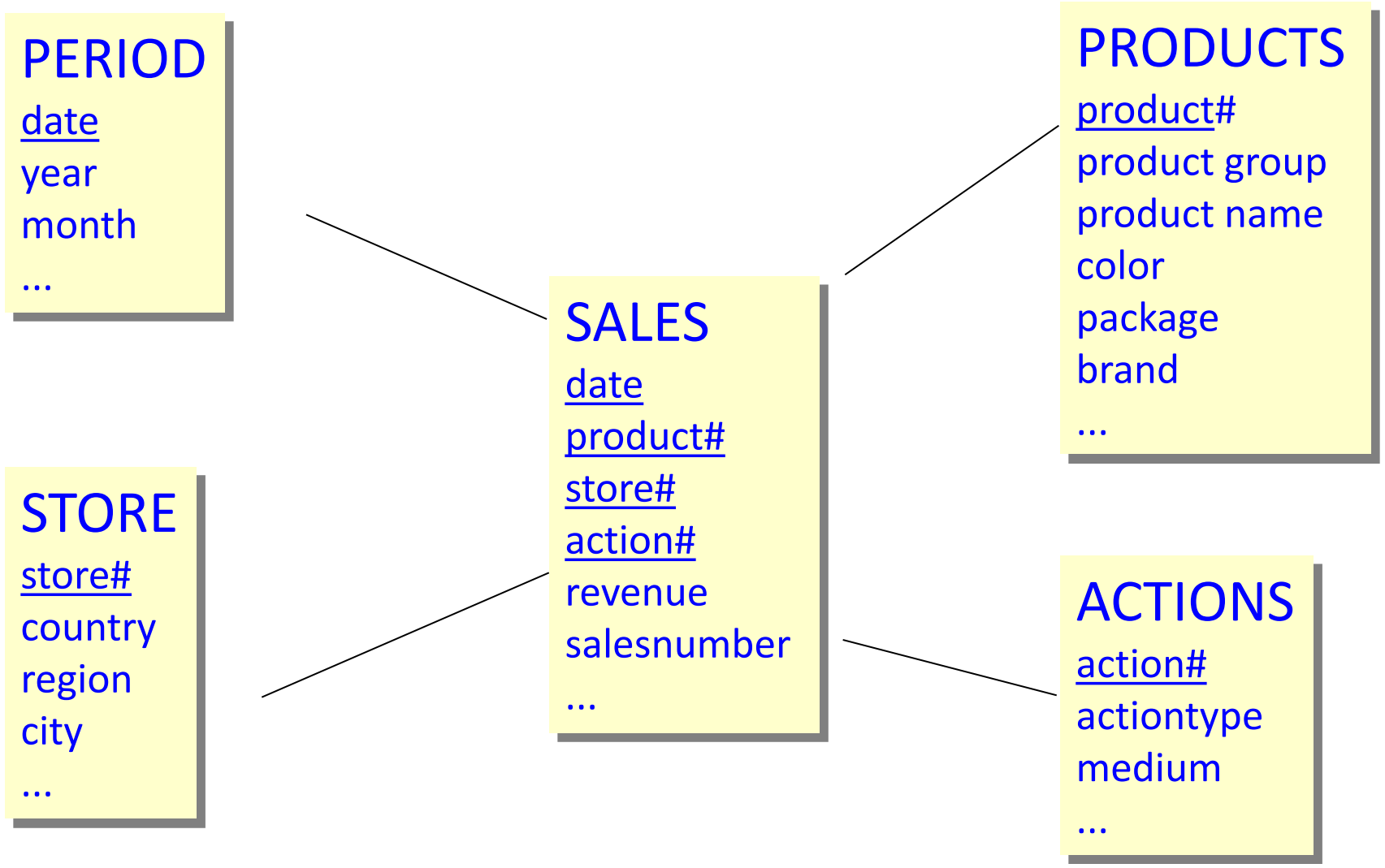
logical database schema, which places dimension tables of a relational database aroung a fact table for easy querying

Maping of multidimensional data to two-dimensional tables.

Star Schema


DIMENSION A
key
description
category A1
category A2
...
DIMENSION B
key
description
category B1
category B2
...
FACTS
key A
key B
key C
key D
attribute1
attribute2
...
DIMENSION C
key
description
category C1
category C2
...
DIMENSION D
key
description
category D1
category D2
...

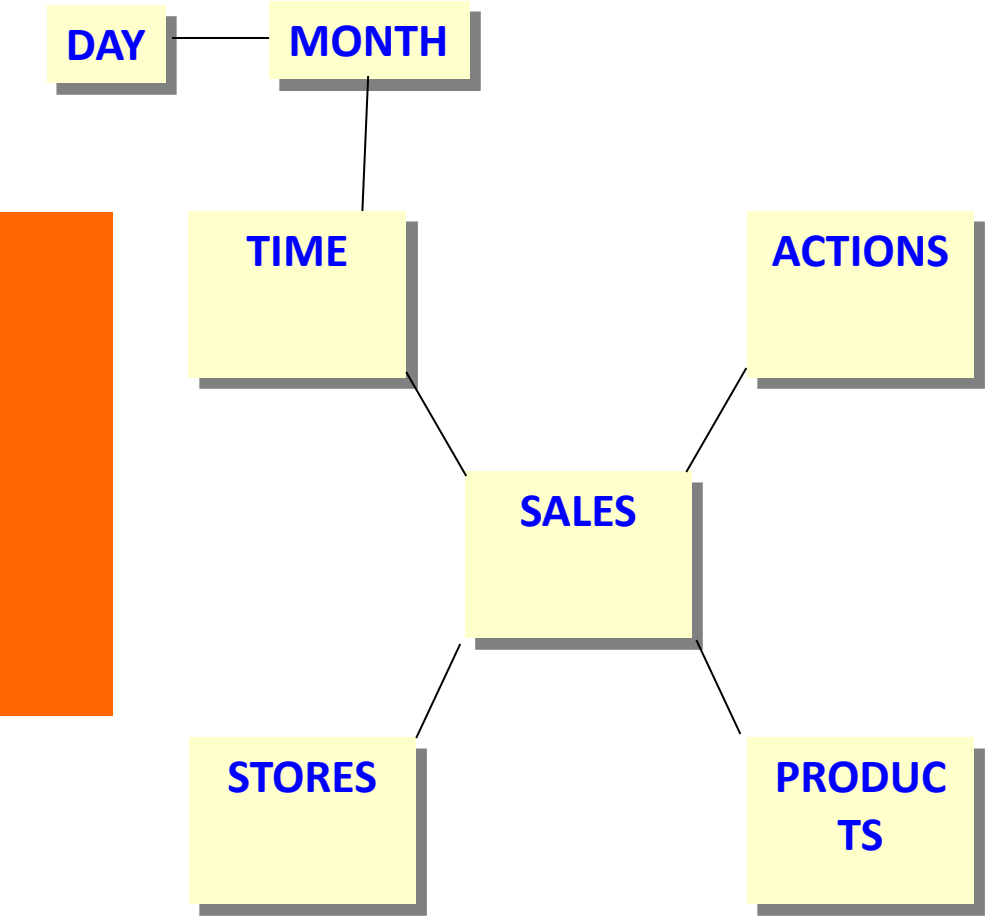
Example of a Start Schema for Retail



SALES: Key from the four foreign keys give 1:n relation to dimension tables

Snowflake Schema

Minimizing Redundancy by Normalisation of the Dimension Time



- The denormalised dimension TIME has the attributes

  TIME(Year, Month, Day, MonthName, DayName, Weekday?)

- Normalised TIME

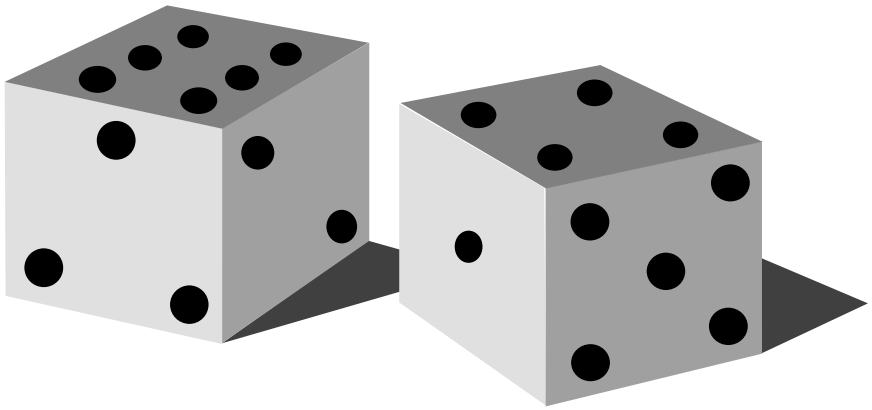
  TIME(Year, Month, Day, Weekday?)

  MONTH(Month, MonthName)

  DAY(Day, DayName)

Source: [Lusti, 1999]

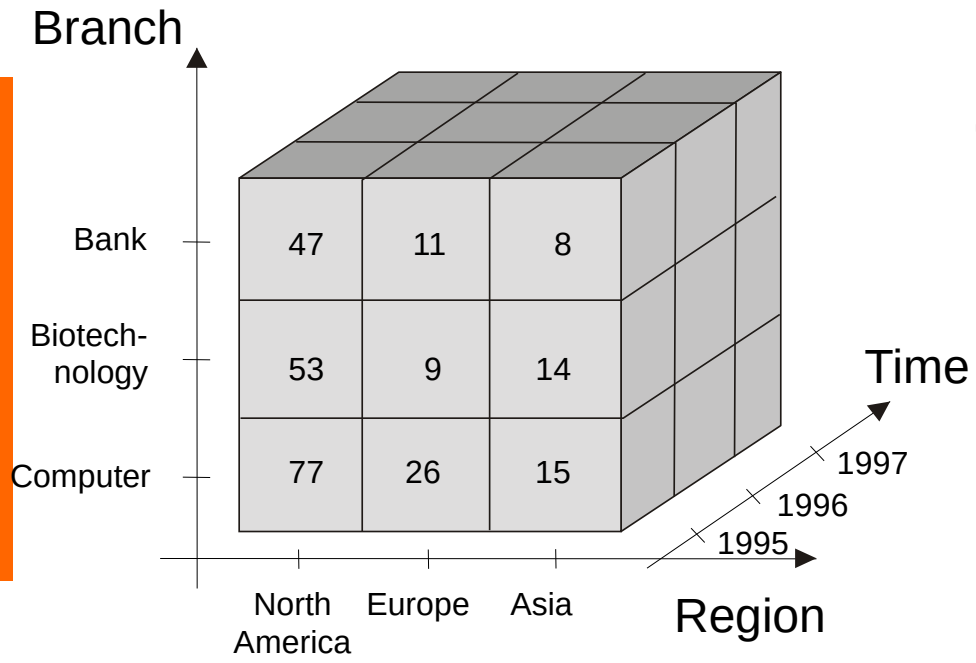
Dicing and Slicing



- An OLAP cube can be regarded as a multidimensional cube
- From a cube only two dimensions are visible on a two-dimensional interfact (e.g. as a table)
- Pivoting
  - "turning" the cube to show other dimension
- Slicing
  - changing the layer in the third dimension
- Dicing
  - Reducing dimensions

Slicing

Reduction of the dimensions in a multi-dimentional cube



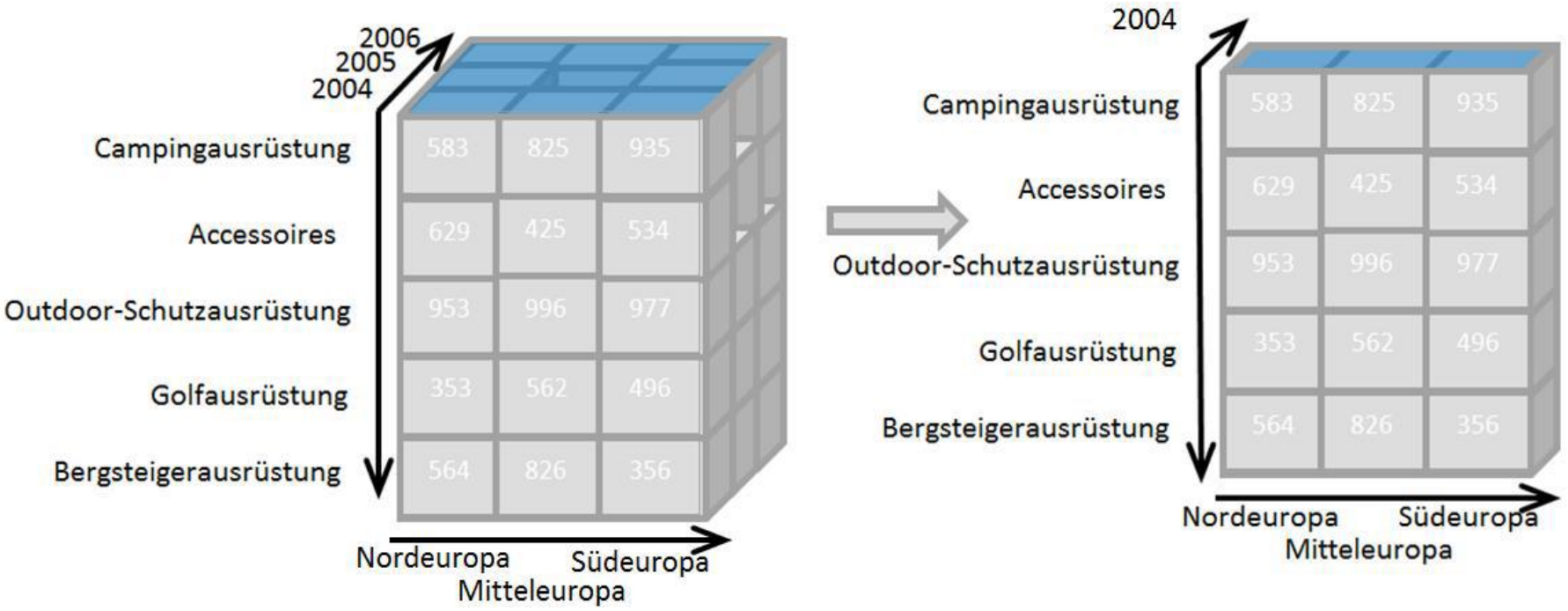
1995

	North America	Europe	Asia
Bank	47	11	8
Biotechn.	53	9	14
Computer	77	26	15

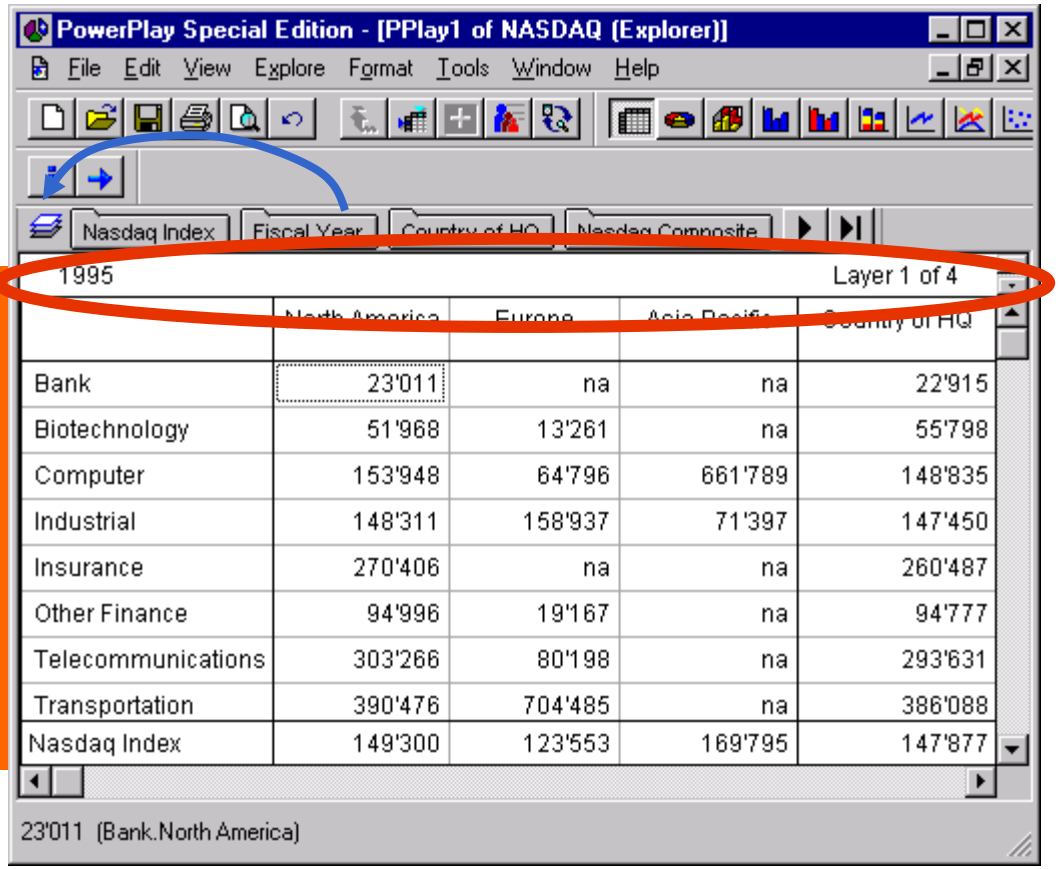
1996

	North America	Europe	Asia
Bank	43	15	8
Biotechn.	54	13	17
Computer	71	45	28

Slicing


2006
2005
2004
Campingausrüstung
Accessoires
Outdoor-Schutzausrüstung
Golfausrüstung
Bergsteigerausrüstung
583 825 935
629 425 534
953 996 977
353 562 496
564 826 356
Nordeuropa
Südeuropa
Mitteleuropa
2004
Campingausrüstung
Accessoires
Outdoor-Schutzausrüstung
Golfausrüstung
Bergsteigerausrüstung
583 825 935
629 425 534
953 996 977
353 562 496
564 826 356
Nordeuropa
Südeuropa
Mitteleuropa

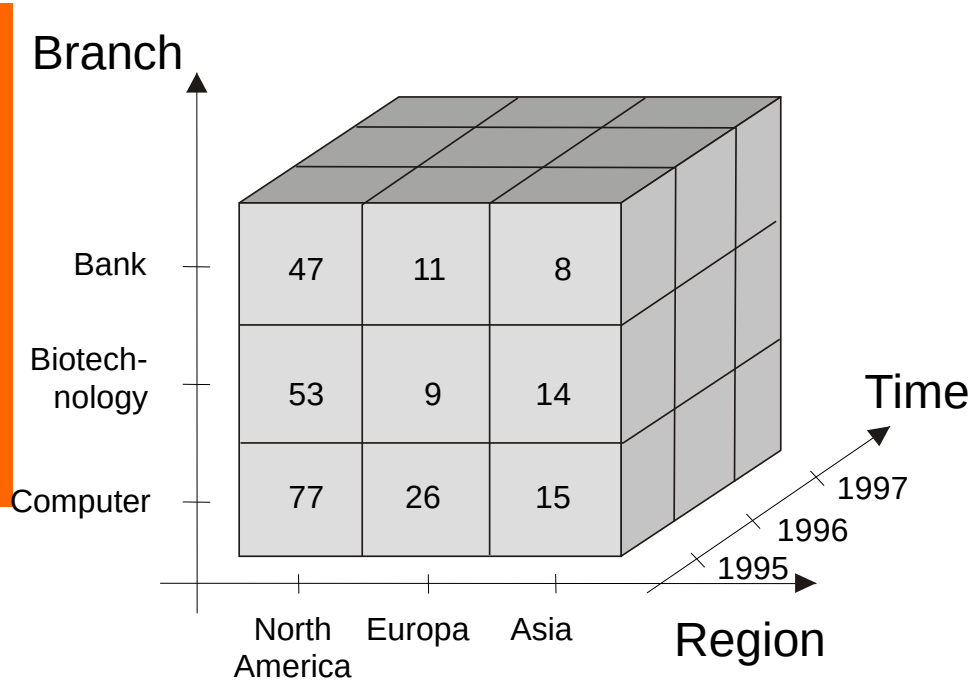
Layers as Additional Dimensions

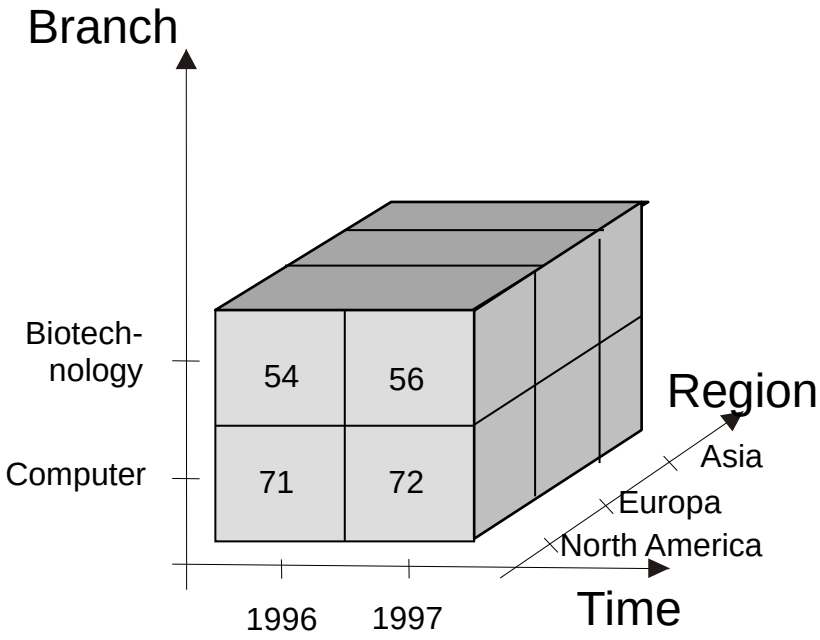


- **Layers**:
  - Showing 3 (instead of only 2) dimensions
  - For each value of the third dimions an additional layer (Drag and Drop Fiscal Year to the layer symbol)
- Example
  - Showing Branch and Country of HQ
  - Slicing: Change to a different Fiscal Year

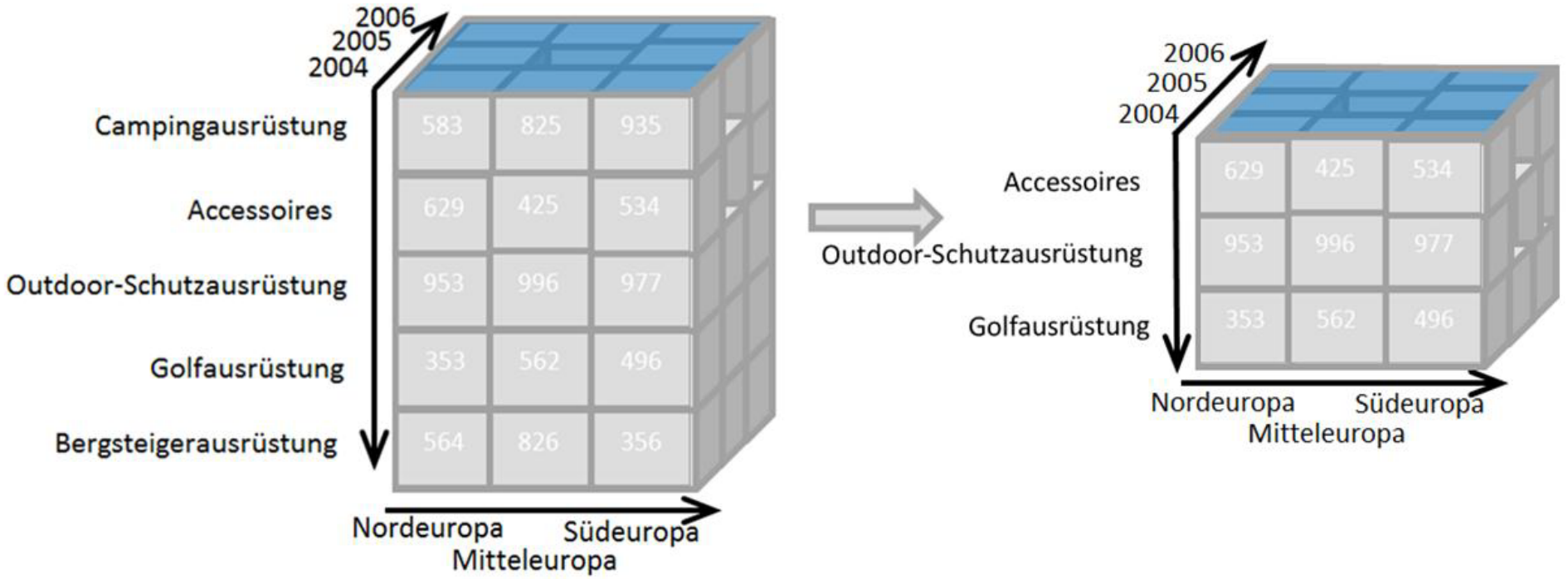
Dicing

- Slicing in several dimensions
  - Creating a smaller cube , showing only part of the cube



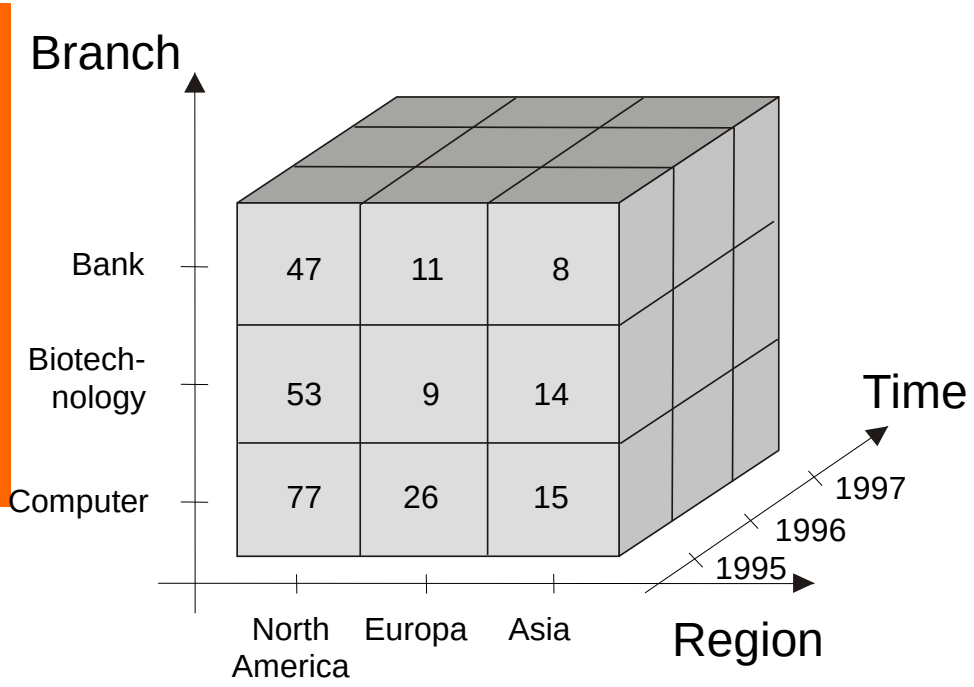


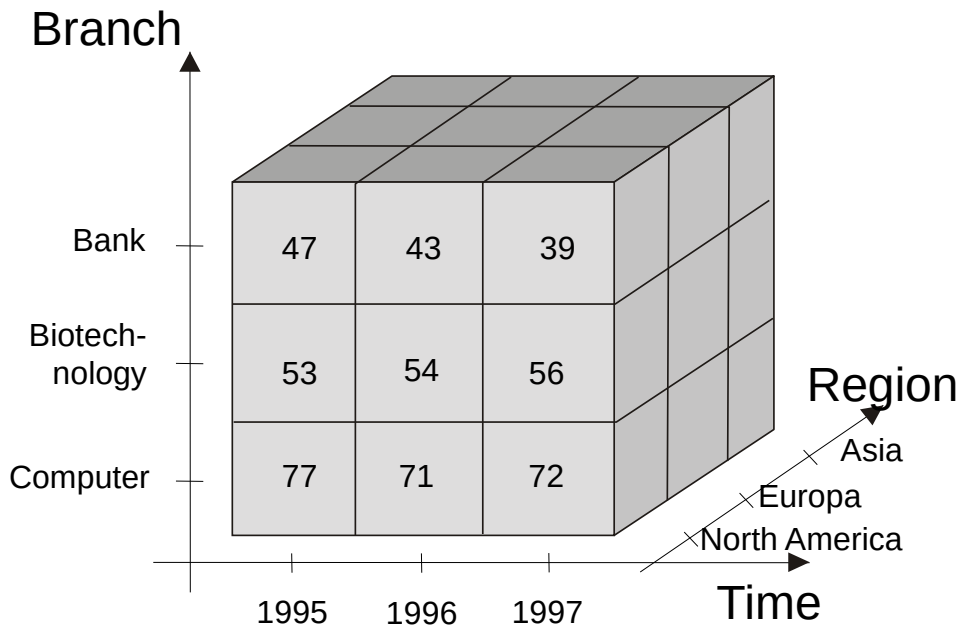
Dicing


2006
2005
2004
Campingausrüstung
Accessoires
Outdoor-Schutzausrüstung
Golfausrüstung
Bergsteigerausrüstung
583
825
935
629
425
534
953
996
977
353
562
496
564
826
356
Nordeuropa
Südeuropa
Mitteleuropa
Outdoor-Schutzausrüstung
2006
2005
2004
Accessoires
Golfausrüstung
629
425
534
953
996
977
353
562
496
Nordeuropa
Südeuropa
Mitteleuropa

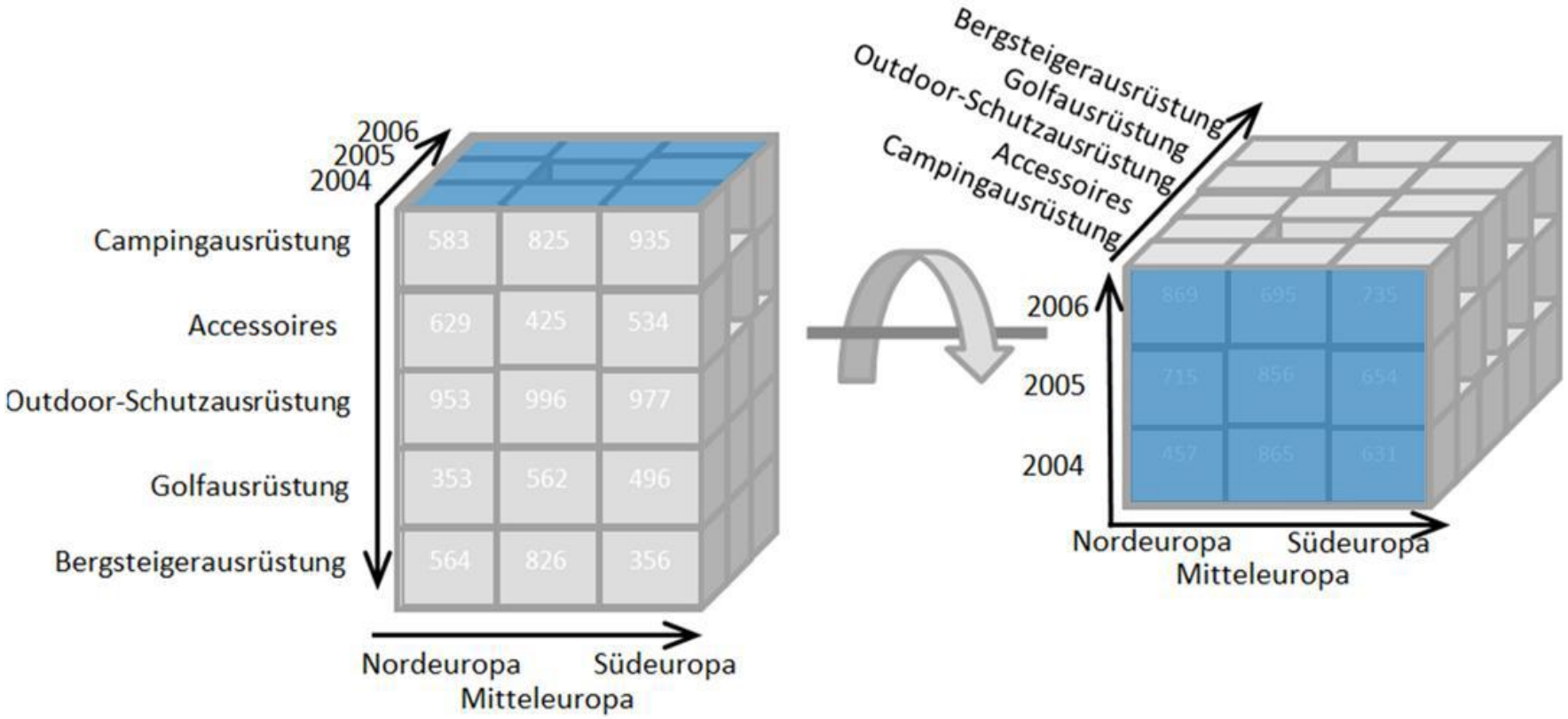
Pivoting

- Rotating the resultats to show different views, e.g.
  - Showing other dimensions

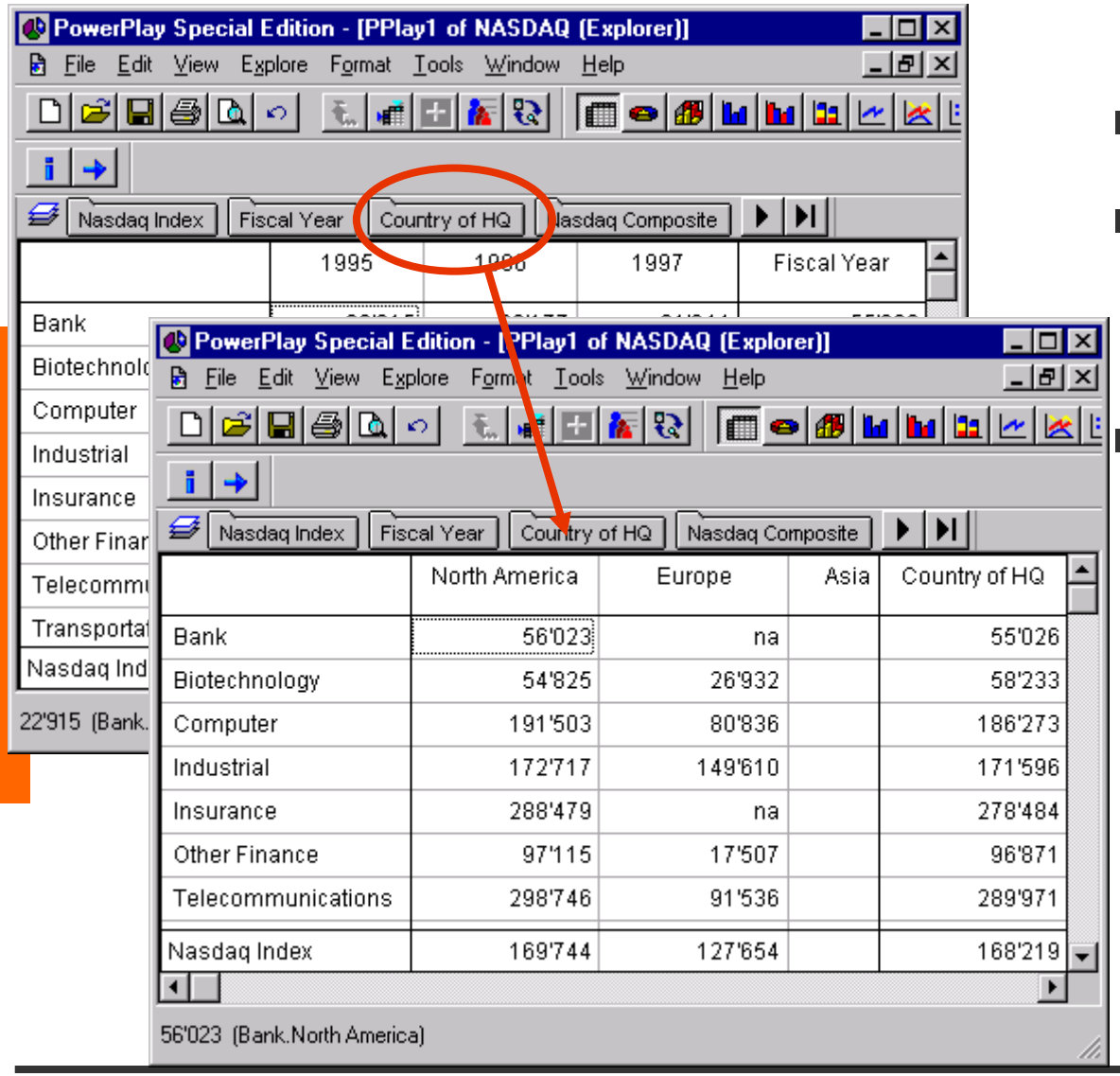




Pivoting


2006
2005
2004
Campingausrüstung
Accessoires
Outdoor-Schutzausrüstung
Golfausrüstung
Bergsteigerausrüstung
583
825
935
629
425
534
953
996
977
353
562
496
564
826
356
Nordeuropa
Südeuropa
Mitteleuropa
Bergsteigerausrüstung
Golfausrüstung
Outdoor-Schutzausrüstung
Accessoires
Campingausrüstung
2006
2005
2004
869
695
735
715
856
654
457
865
631
Nordeuropa
Südeuropa
Mitteleuropa

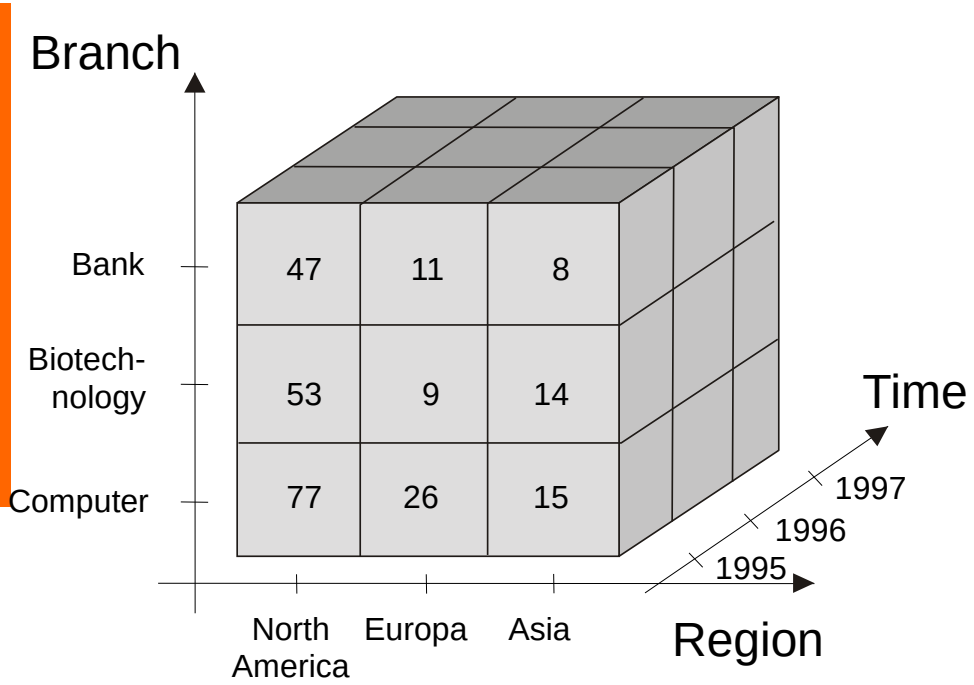
Pivoting

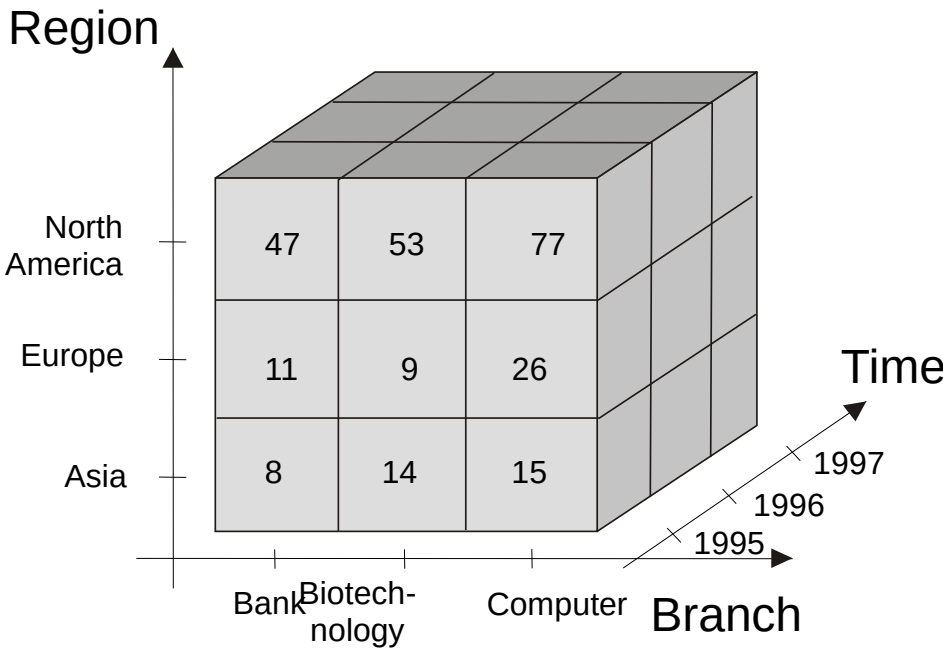


- Selection of dimensions
- Drag and Drop a dimension of the horizontal line to the window
- Example: Drag and Drop the dimension Country of HQ in the column headers and thus exchange Fiscal Year

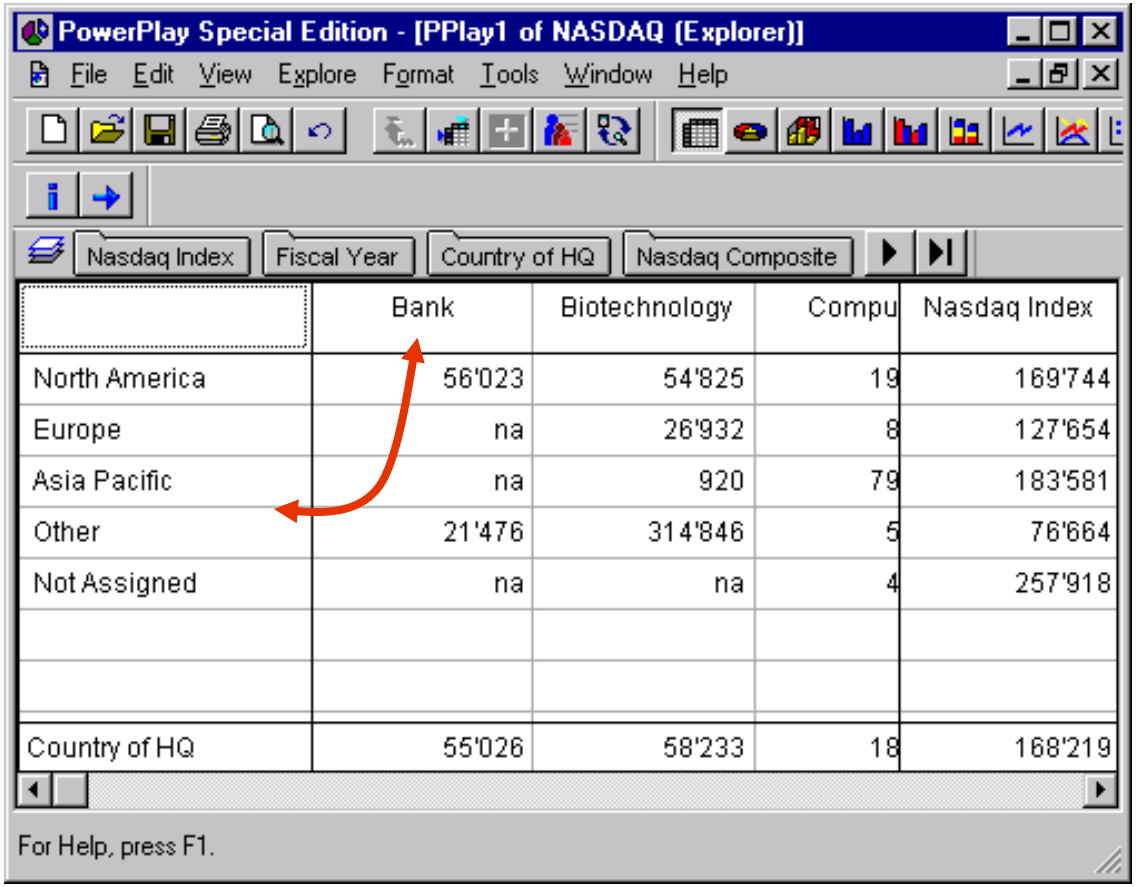
Pivoting

- Rotating the resultats to show different views, e.g.
  - Exchange of rows and columns





Pivoting: Exchange Rows and Columns



- **Pivoting:** Exchange rows and columns
- Example: Exchange dimensions Branch and Country of HQ