

# Analysis und Use of Data (2): OLAP

*Knut Hinkelmann*

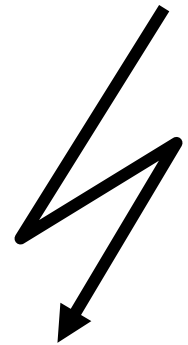
# Objectives of Data Modeling

## ■ operative databases

- ◆ **Transaction processing** by high granularity
- ◆ **Progress friendliness** by normalisation
- ◆ **Storage efficiency** by normalisation

## ■ Data Warehouses

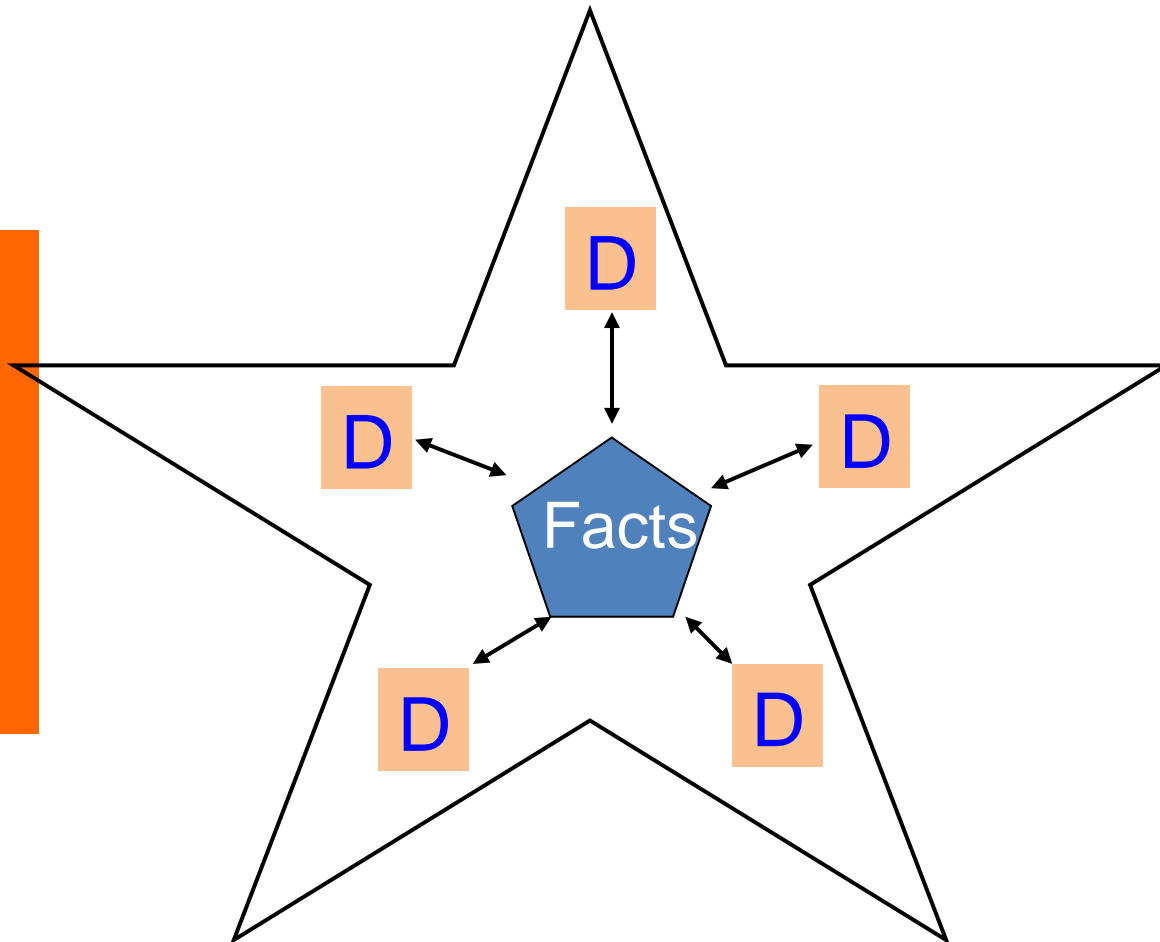
- ◆ **user friendliness** by low granularity
- ◆ **query efficiency** by denormalisation
- ◆ **close to business** by indicators and dimensions



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

cf. [Lusti, 1999]

# Star Schema for Relational Data Warehouses /Marts to support OLAP

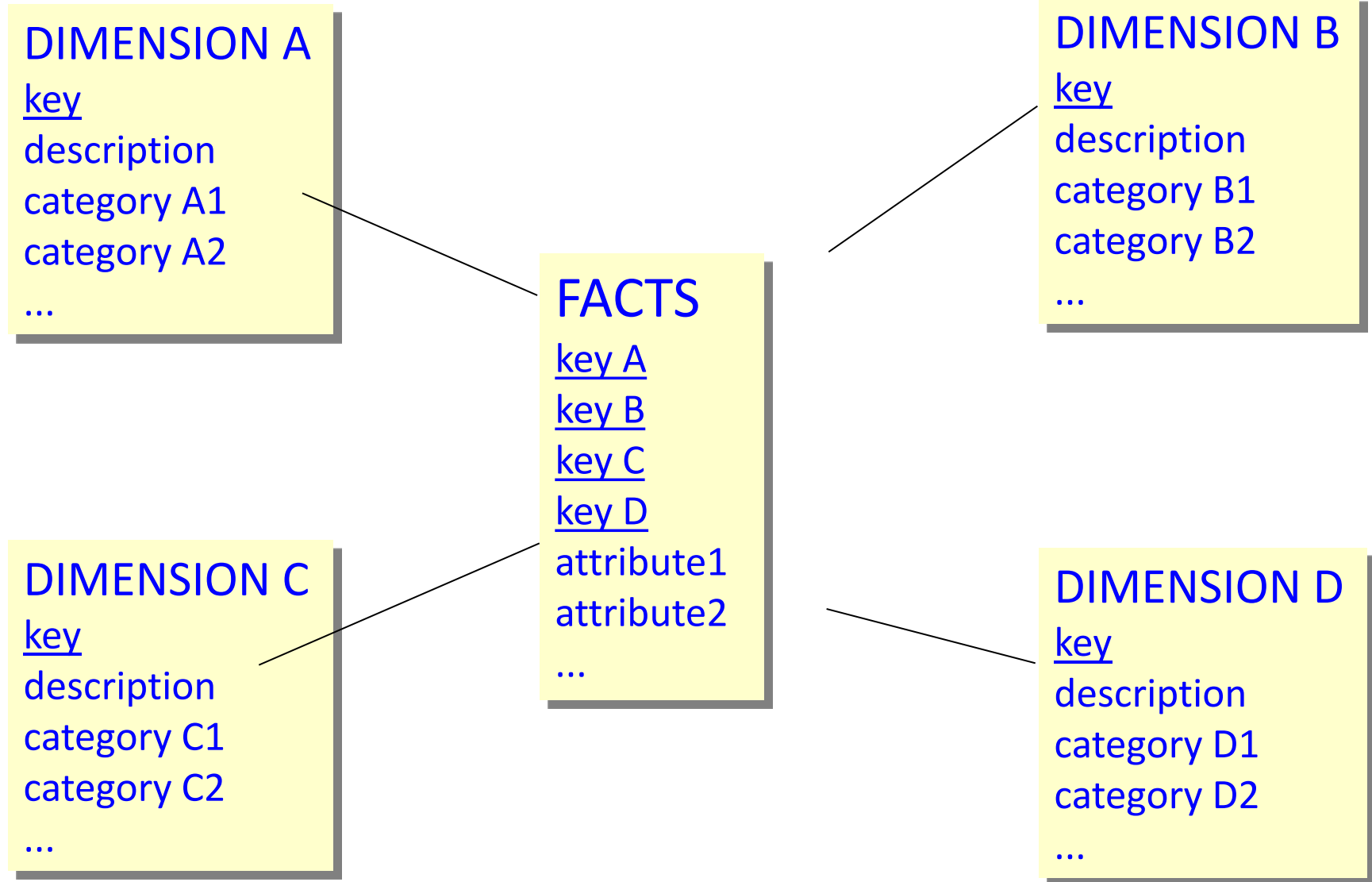


Star Schema:

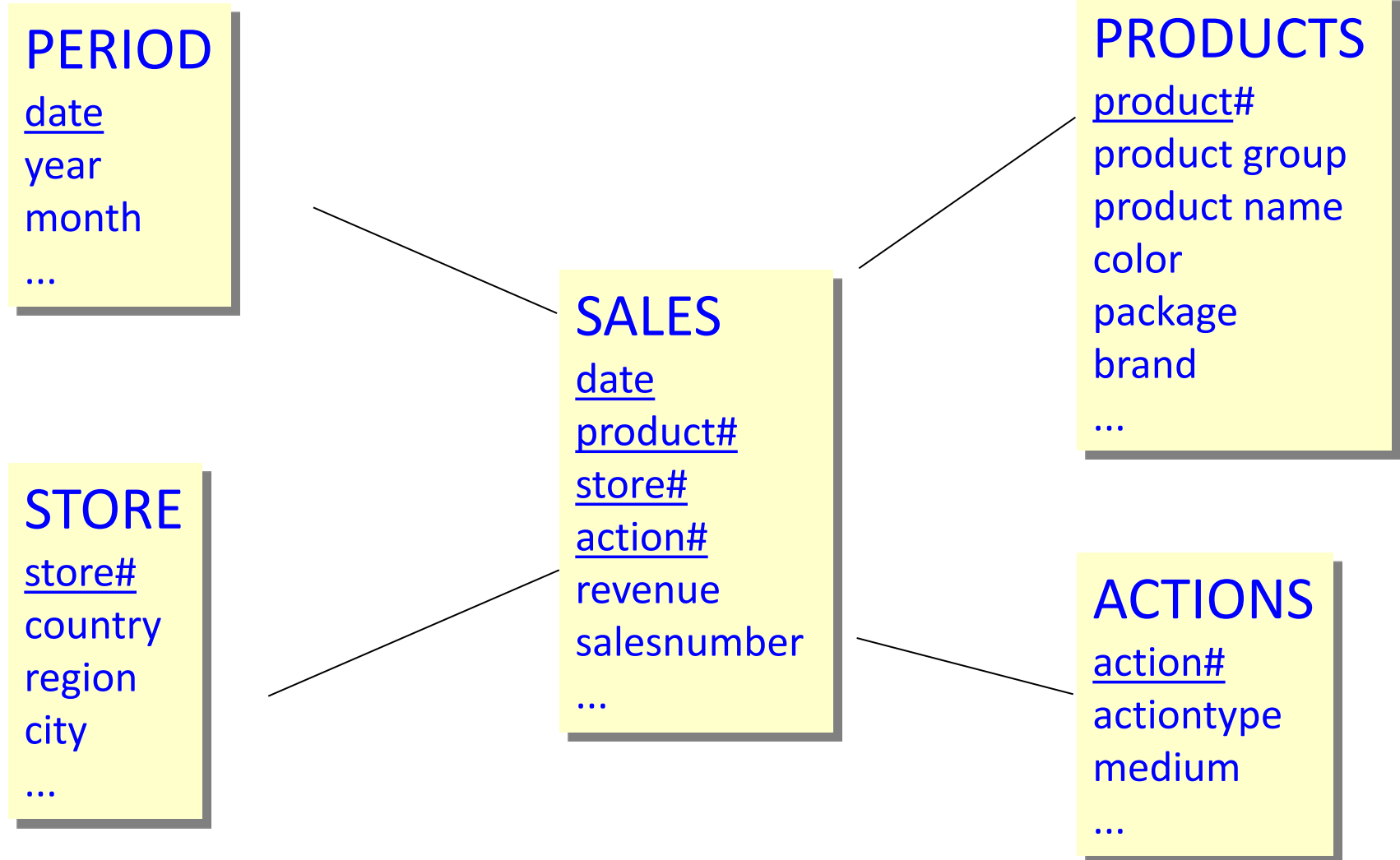
logical database schema, which places dimension tables of a relational database around a fact table for easy querying

Mapping of multidimensional data to two-dimensional tables.

# Star Schema



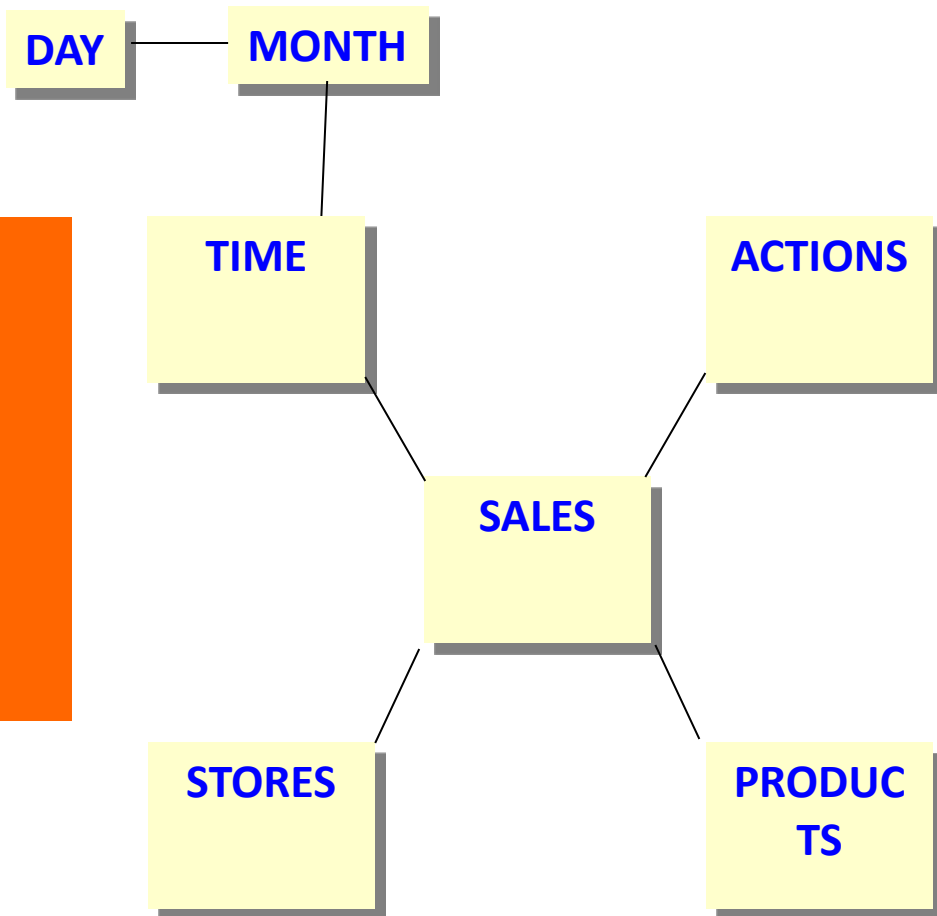
# Example of a Start Schema for Retail



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]



A solid orange vertical bar is positioned on the left side of the slide.

# ***MULTIDIMENSIONAL ANALYSIS: OLAP***

## OLAP – definition

- *OLAP = An approach to swiftly answer multi-dimensional analytical (MDA) queries (Wikipedia)*
  - ◆ Main functionality: **aggregate and de-aggregate fact measures**
    - e.g. sales/bookings by product/customer/sales rep/time
    - e.g. receipts/failures/stock by part/supplier

### ■ Distinction from (static) reports: FASMI criteria

- **FASMI:** Fast Analysis of Shared Multidimensional Information (synthesised from Codd's "12 laws" of OLAP)
  - ◆ fast: response time in the order of seconds
  - ◆ analysis: intuitive, user-friendly, with flexible queries
  - ◆ shared: multi-user access
  - ◆ multidimensional: offer conceptual multidimensional view, independently of physical storage
  - ◆ information: scale to large data volumes



# Information needs requiring OLAP analyses (1)

What are the profitabilities by customer segment?

What is the number of visits to the branch office compared to ATM and e-banking?

How many receipts, rejects and parts per million value do we have per supplier

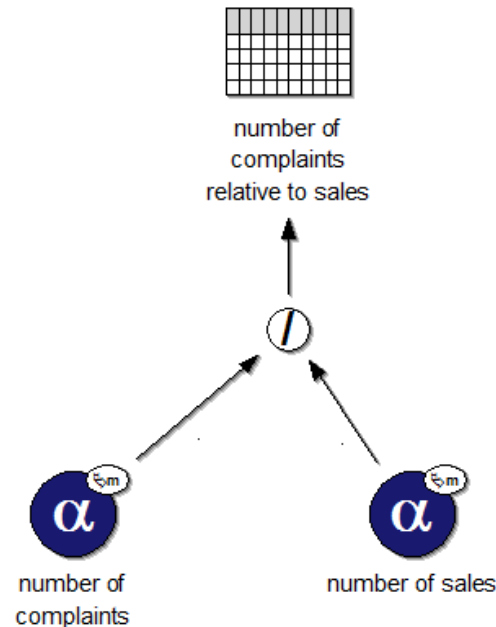
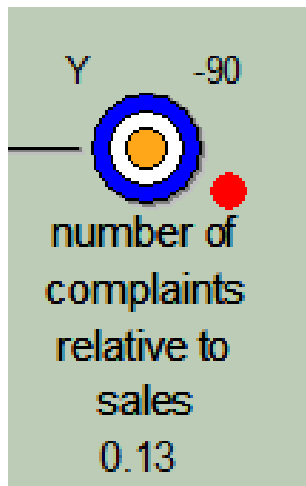
What sellers' strategies work best to yield high and profitable sales?

Where are campaigns more successful and where less?

Which sales are attributable to sales rep XYZ?

## Information needs requiring OLAP analyses (2)

- **Monitoring** of strategic goals: analyse deviations from target values, find root causes:



- Who is complaining?
  - ◆ Number of complaints per customer (segment)
- What are they complaining about?
  - ◆ Number of complaints per product/service
- When did they complain?
  - ◆ Number of complaints per week / month



## Information needs requiring OLAP analyses (3)

- **Exploring / analysing** to support decisions: find patterns.



*who  
(targets)?*

*when?*

*which  
channel(s)?*

Inquiries about the product or similar ones:

- Who is buying?
  - ◆ sales per customer (segment)
- Which channels do they use?
  - ◆ sales per channel
- When do they buy?
  - ◆ sales per quarter / month
- Where do they buy?
  - ◆ sales per region

***Example: how to market a new product (a «pinot gris» from Alsace)?***

# Requirements for BI tools – front-end

## 1. indicator-driven:

- ◆ 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

*BSC tools  
dashboards,  
reports*

*OLAP*

## 2. explorative:

- ◆ support queries for particular facts, needed to make decisions in core business processes
- ◆ support ad-hoc aggregation of numbers e.g. to make decisions about a particular product, customer, supplier or sales rep
- ◆ making prediction, e.g. to optimise sales and marketing strategies

*OLAP*

*data mining  
tools*



A solid orange vertical bar is positioned on the left side of the slide.

# END USER VIEW – PREPARING A CUBE (RELATIONAL)

# Preparing a cube in a relational OLAP tool (1)

- Prerequisite: data is stored in a star schema in the DWH
- Procedure for importing the star:

The screenshot shows a 'Data Source Wizard' window with the following sections:

- Select Source Type:** SB Sales
- Source Type:** Database Table(s)
- Database Table(s):** `SBSales`.`date`, `SBSales`.`product`, **Selected:** `SBSales`.`sales`, `SBSales`.`store`
- Fact Table:** `SBSales`.`sales`
- Join(s):**
  - `SBSales`.`sales`.date\_key - INNER JOIN - `SBSales`.`date`.date\_key
  - `SBSales`.`sales`.product\_key - INNER JOIN - `SBSales`.`product`.product\_key
  - `SBSales`.`sales`.store\_key - INNER JOIN - `SBSales`.`store`.store\_key

1. select the DB connection

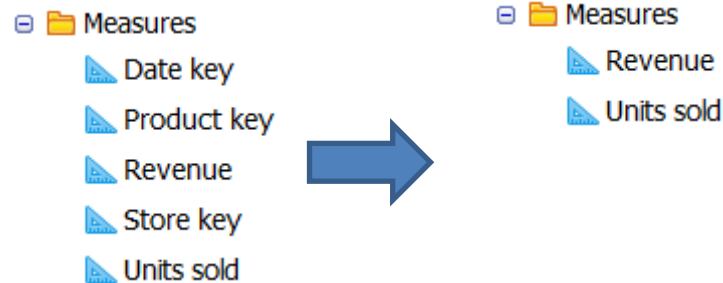
2. specify fact and dimension tables

3. specify joins

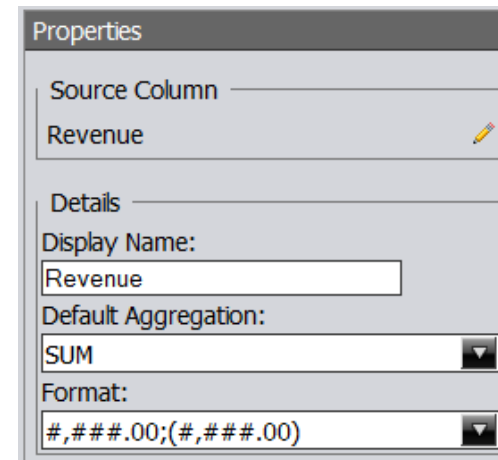


# Preparing a cube in a relational OLAP tool (2)

- Procedure for configuring facts: «customize model»



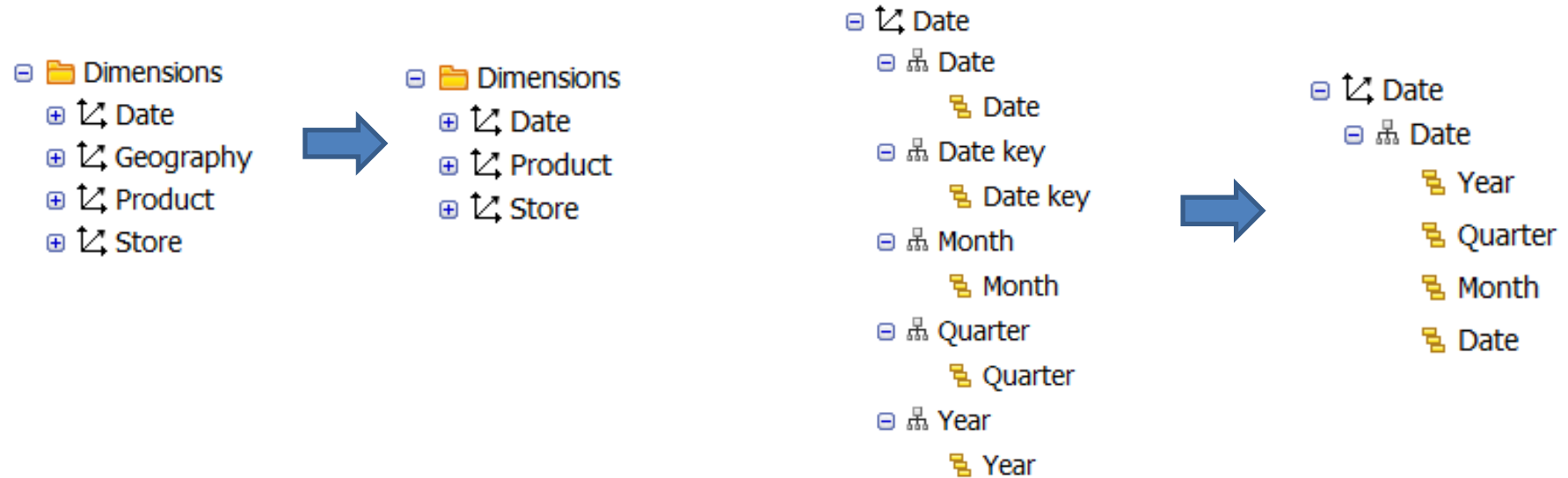
1. *remove measures that are actually dimension keys*



2. *specify aggregations and number formats*

# Preparing a cube in a relational OLAP tool (3)

## ■ Procedure for configuring dimensions:



*1. select the desired dimensions, possibly rename*

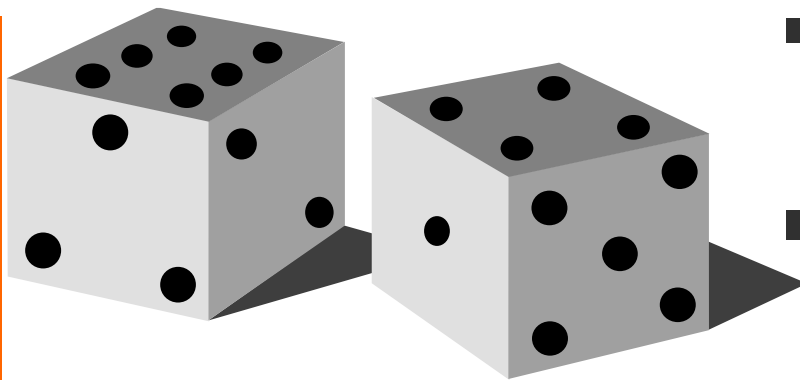
*2. specify dimension hierarchies*



A solid orange vertical bar is positioned on the left side of the slide.

# END USER VIEW – OLAP OPERATIONS

# Dicing and Slicing



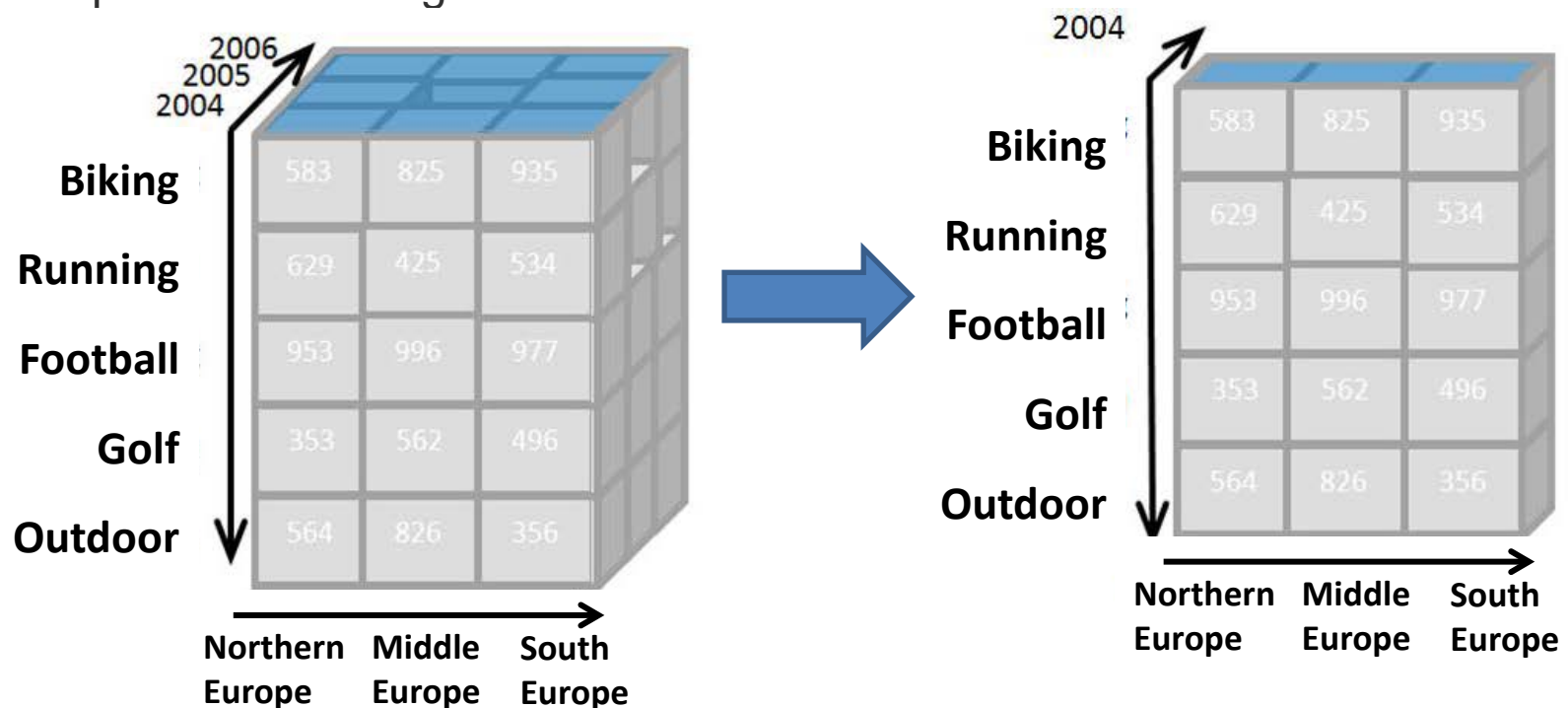
- An OLAP cube can be regarded as a multidimensional cube
- From a cube only two dimensions are visible on a two-dimensional interface (e.g. as a table)
- Slicing
  - ◆ Constraining one dimension
- Dicing
  - ◆ Constraining several dimensions
- Pivoting
  - ◆ "turning" the cube to show other dimension
- Roll-up/Drill-down – Split/Merge
  - Aggregate or detailing views

# OLAP Operation - Slicing

Reduction of the dimensions in a multi-dimensional cube

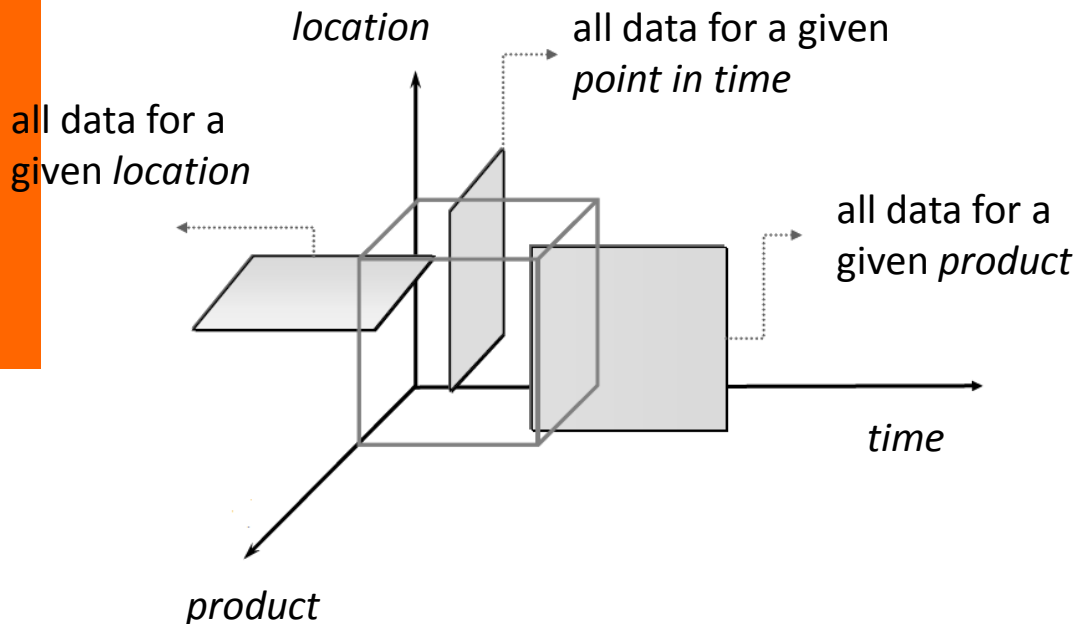
- Constraining one dimension to a particular value

Example: Constraining time



# OLAP Operation – Slicing

- **Slicing** is selecting a group of cells from the entire cube by constraining one dimension to a *particular value*.



	Date	
	+ Q3	+ Q4
	Kennzahlen	Kennzahlen
Product	● Revenue	● Revenue
<input type="checkbox"/> All Products	8.925,00	34.925,00
<input type="checkbox"/> accessories	1.810,00	3.150,00
<input type="checkbox"/> mountain bikes	7.115,00	31.775,00



slice: Country = Austria

	Date	
	+ Q3	+ Q4
	Kennzahlen	Kennzahlen
Product	● Revenue	● Revenue
<input type="checkbox"/> All Products	215,00	7.230,00
<input type="checkbox"/> accessories	215,00	170,00
<input type="checkbox"/> mountain bikes		7.060,00

Slicer: [Country=Austria]

# Layers as Additional Dimensions

PowerPlay Special Edition - [PPlay1 of NASDAQ (Explorer)]

File Edit View Explore Format Tools Window Help

Nasdaq Index Fiscal Year Country of HQ Nasdaq Composite

1995 Layer 1 of 4

	North America	Europe	Asia Pacific	Country of HQ
Bank	23'011	na	na	22'915
Biotechnology	51'968	13'261	na	55'798
Computer	153'948	64'796	661'789	148'835
Industrial	148'311	158'937	71'397	147'450
Insurance	270'406	na	na	260'487
Other Finance	94'996	19'167	na	94'777
Telecommunications	303'266	80'198	na	293'631
Transportation	390'476	704'485	na	386'088
Nasdaq Index	149'300	123'553	169'795	147'877

23'011 (Bank.North America)

## ■ Layers:

- ◆ Showing 3 (instead of only 2) dimensions
- ◆ For each value of the third dimension an additional layer (Drag and Drop Fiscal Year to the layer symbol)

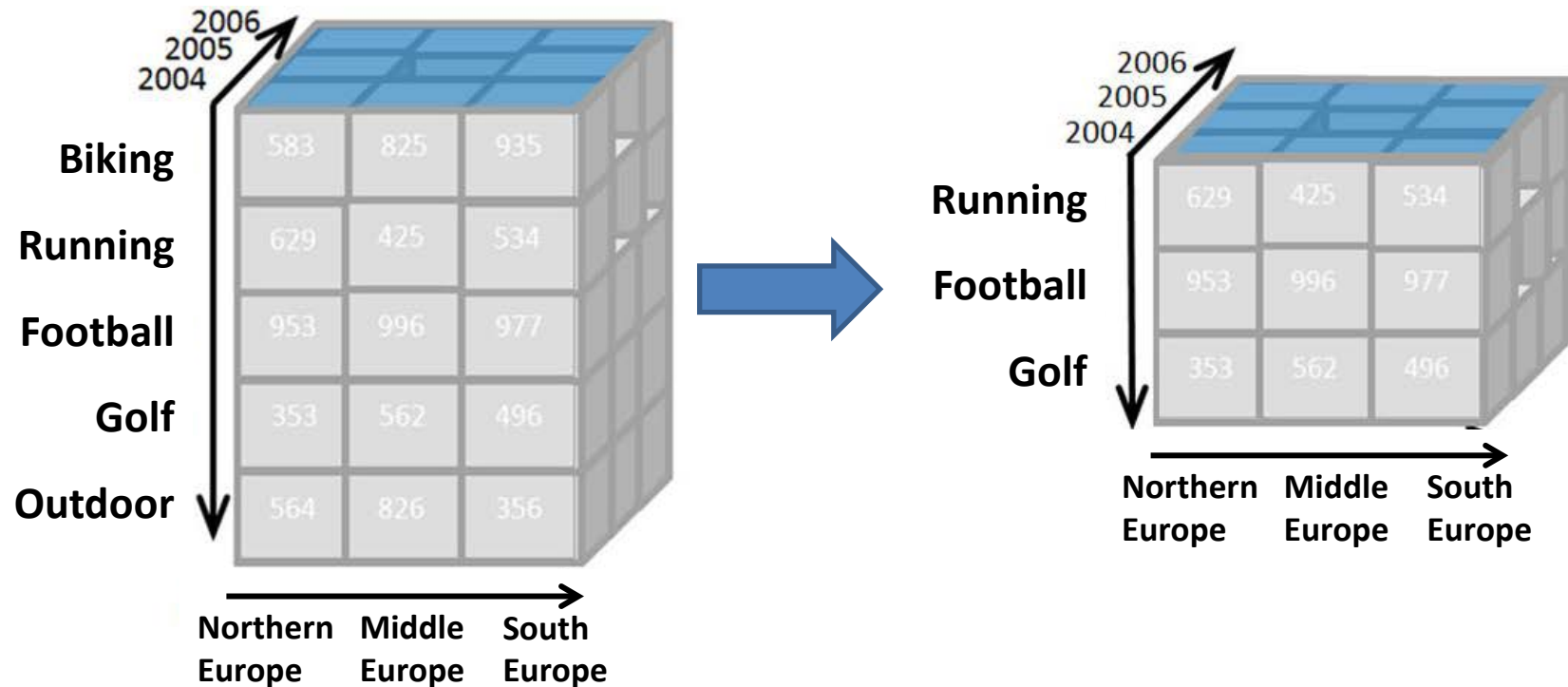
## ■ Example

- ◆ Showing Branch and Country of HQ
- ◆ Slicing: Change to a different Fiscal Year

# OLAP Operation - Dicing

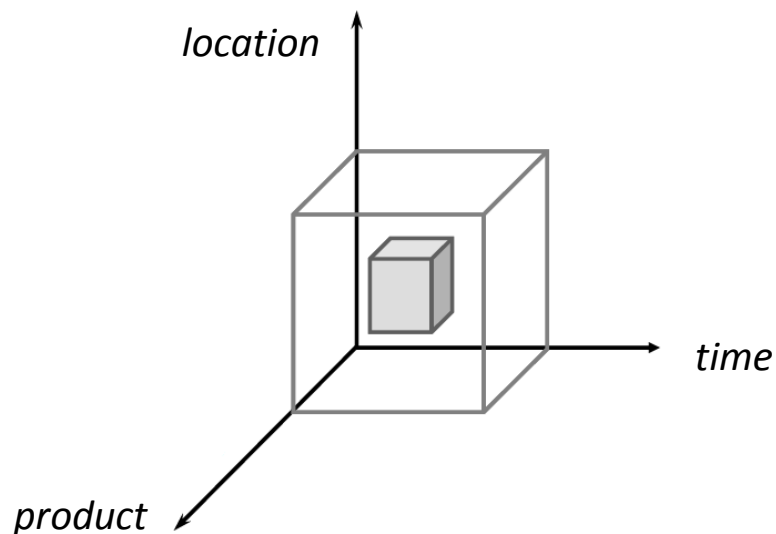
Slicing in several dimensions

- ◆ Creating a smaller cube , showing only part of the cube



# OLAP Operations – Dicing

- **Dicing** involves selecting a subset of cells by specifying a *range of attribute values for more than one dimension*.



	Date	
	<input type="checkbox"/> Q3	<input type="checkbox"/> Q4
	Kennzahlen	Kennzahlen
Product	<input checked="" type="radio"/> Revenue	<input checked="" type="radio"/> Revenue
<input type="checkbox"/> All Products	8.925,00	34.925,00
<input type="checkbox"/> accessories	1.810,00	3.150,00
<input type="checkbox"/> mountain bikes	7.115,00	31.775,00

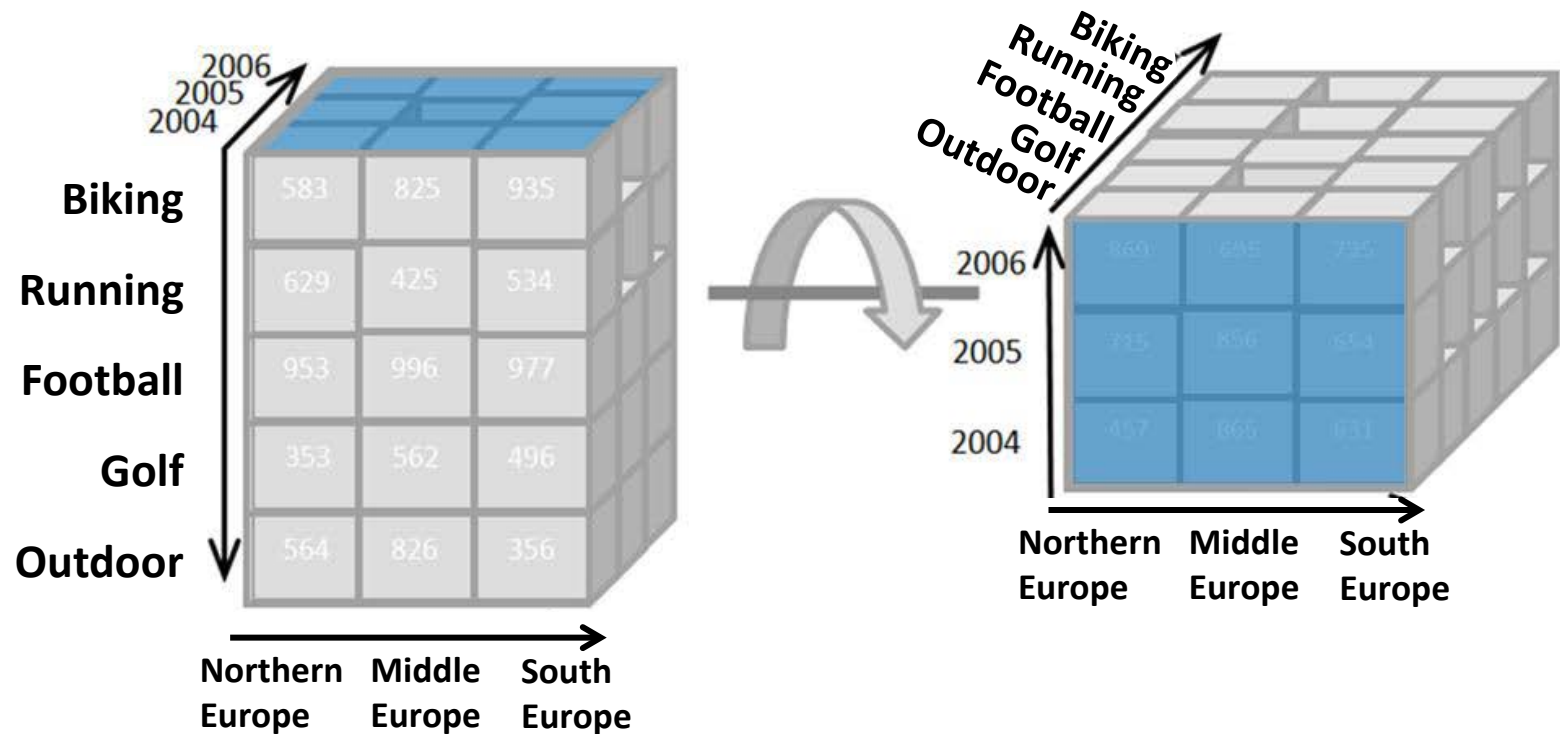


*dice: month = {September, October}*  
*products = {SB123, SB234}*

	Date	
	<input type="checkbox"/> September	<input type="checkbox"/> October
	Kennzahlen	Kennzahlen
Product	<input checked="" type="radio"/> Revenue	<input checked="" type="radio"/> Revenue
Mountain Bike SB123	2.300,00	13.800,00
Mountain Bike SB234	1.765,00	8.825,00

# OLAP Operation - Pivoting

Rotating the cube to show other dimensions





# OLAP Operation - Pivoting

The screenshot illustrates the pivoting operation in PowerPlay. The top window shows the 'Country of HQ' dimension selected in the dimension list. The bottom window shows the result of pivoting 'Country of HQ' into the column headers and 'Fiscal Year' into the row headers.

	North America	Europe	Asia	Country of HQ
Bank	56'023	na		55'026
Biotechnology	54'825	26'932		58'233
Computer	191'503	80'836		186'273
Industrial	172'717	149'610		171'596
Insurance	288'479	na		278'484
Other Finance	97'115	17'507		96'871
Telecommunications	298'746	91'536		289'971
Nasdaq Index	169'744	127'654		168'219

- 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



# OLAP Operations – roll-up/drill-down

- Attribute values often have a hierarchical structure.
  - ◆ e.g. products can be organised into product categories
  - ◆ For sales quantities, we can aggregate (**roll up**) the expenses across all the products in a group.
  - ◆ Conversely, we could split the total quantities (**drill down**) into ones for each product in the group

	Date	
	+ Q3	+ Q4
	Kennzahlen	Kennzahlen
Product	● Revenue	● Revenue
<input type="checkbox"/> All Products	8.925,00	34.925,00
<input type="checkbox"/> accessories	1.810,00	3.150,00
<input type="checkbox"/> mountain bikes	7.115,00	31.775,00



	Date	
	+ Q3	+ Q4
	Kennzahlen	Kennzahlen
Product	● Revenue	● Revenue
<input type="checkbox"/> All Products	8.925,00	34.925,00
<input type="checkbox"/> accessories	1.810,00	3.150,00
<input type="checkbox"/> mountain bikes	7.115,00	31.775,00
Mountain Bike SB123	2.300,00	13.800,00
Mountain Bike SB234	1.765,00	8.825,00
Mountain Bike SB345	3.050,00	9.150,00

screenshots taken from Pentaho CE)



# OLAP Operation – split/merge

- **Split** = show details for a value by *adding a dimension*
  - ◆ e.g. split sales by region
- **Merge** = remove a dimension such that data is aggregated for that dimension

	Date	
	<input type="checkbox"/> Q3	<input type="checkbox"/> Q4
	Kennzahlen	Kennzahlen
Product	● Revenue	● Revenue
<input type="checkbox"/> All Products	8.925,00	34.925,00
<input type="checkbox"/> accessories	1.810,00	3.150,00
<input type="checkbox"/> mountain bikes	7.115,00	31.775,00



	Date							
	<input type="checkbox"/> Q3				<input type="checkbox"/> Q4			
	Store name				Store name			
	<input type="checkbox"/> All Store.Store names	<input type="checkbox"/> Austria	<input type="checkbox"/> Germany	<input type="checkbox"/> Switzerland	<input type="checkbox"/> All Store.Store names	<input type="checkbox"/> Austria	<input type="checkbox"/> Germany	<input type="checkbox"/> Switzerland
	Kennzahlen	Kennzahlen	Kennzahlen	Kennzahlen	Kennzahlen	Kennzahlen	Kennzahlen	Kennzahlen
Product	● Revenue	● Revenue	● Revenue	● Revenue	● Revenue	● Revenue	● Revenue	● Revenue
<input type="checkbox"/> All Products	8.925,00	215,00	90,00	8.620,00	34.925,00	7.230,00	2.730,00	24.965,00
<input type="checkbox"/> accessories	1.810,00	215,00	90,00	1.505,00	3.150,00	170,00	430,00	2.550,00
<input type="checkbox"/> mountain bikes	7.115,00			7.115,00	31.775,00	7.060,00	2.300,00	22.415,00



# OLAP beyond the cube – drill-through

- **Drill-through** = enable viewing the original data (e.g. transactions) by changing to another data source or showing all relevant fact table rows

	Date	
	+ Q3	+ Q4
	Kennzahlen	Kennzahlen
Product	● Revenue	● Revenue
- All Products	↕8.925,00	↕34.925,00
+ accessories	↕1.810,00	↕3.150,00
+ mountain bikes	↕7.115,00	↕31.775,00



Drill Through Table for revenue										
^ year	● quarter	● month	● date	● product_category	● product_name	● country	● province	● store_name	● brand	● revenue
2013	Q3	September	28.09.2013	mountain bikes	Mountain Bike SB123	Switzerland	Bern	Store Bern	Swiss Bikes	2.300,00
2013	Q3	September	28.09.2013	mountain bikes	Mountain Bike SB234	Switzerland	Solothurn	Store Solothurn	Swiss Bikes	1.765,00
2013	Q3	September	29.09.2013	mountain bikes	Mountain Bike SB345	Switzerland	Bern	Store Bern	Swiss Bikes	3.050,00

Seite 1/1 Zeilen/Seite 10

# Example Application of OLAP (1)

## a) Protocol:

- ◆ **slice:** use region=EMEA as a slicer
- ◆ **merge:** remove product and time from the display
- ◆ *move order status to columns*
- ◆ **dice:** constrain order status to the values «disputed» and «on hold»
  
- ◆ *Answer: we have problems with*
  - *Danish Wholesale Imports*
  - *Euro+Shopping Channel*
  - *Volvo Model Replicas, Co*

	Kennzahlen	
	Quantity	
	Order Status	
Customers	● Disputed	● On Hold
<input type="checkbox"/> All Customers	433	217
Danish Wholesale Imports	174	
Euro+ Shopping Channel	259	
Volvo Model Replicas, Co		217

Slicer: [Territory=EMEA]

# Application of OLAP (2)

## b) Protocol:

- ◆ **split:** add the Product dimension to columns
- ◆ *Answer: the problematic customers have ordered «Classic Cars», «Motorcycles» and «Trucks and Buses»*

	Kennzahlen			
	Quantity			
	Order Status			
	Disputed		On Hold	
	Product		Product	
Customers	● <input type="checkbox"/> Classic Cars	● <input type="checkbox"/> Motorcycles	● <input type="checkbox"/> Classic Cars	● <input type="checkbox"/> Trucks and Buses
<input type="checkbox"/> All Customers	174	259	46	171
Danish Wholesale Imports	174			
Euro+ Shopping Channel		259		
Volvo Model Replicas, Co			46	171

Slicer: [Territory=EMEA]



# Application of OLAP (3)

## ■ Protocol:

- ◆ **split:** add the Time dimension to columns
- ◆ **dice:** constrain the Time dimension to «All years» and «2005»
- ◆ *Answer: in 2005, we've had disputes with Danish Wholesale Imports over 174 ordered classic cars and with Euro + Shopping Channel over 259 ordered motorcycles*

	Kennzahlen					
	Quantity					
	Order Status					
	Disputed			On Hold		
	Product			Product		
	+ Classic Cars		+ Motorcycles		+ Classic Cars	+ Trucks and Buses
	Time		Time		Time	Time
Customers	● <input type="checkbox"/> All Years	● <input checked="" type="checkbox"/> 2005	● <input type="checkbox"/> All Years	● <input checked="" type="checkbox"/> 2005	● <input type="checkbox"/> All Years	● <input type="checkbox"/> All Years
Danish Wholesale Imports	174	174				
Euro+ Shopping Channel			259	259		
Volvo Model Replicas, Co					46	171

Slicer: [Territory=EMEA]

