

Analysis and Use of Data

Knut Hinkelmann

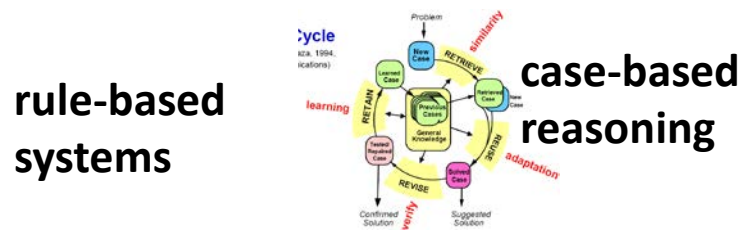
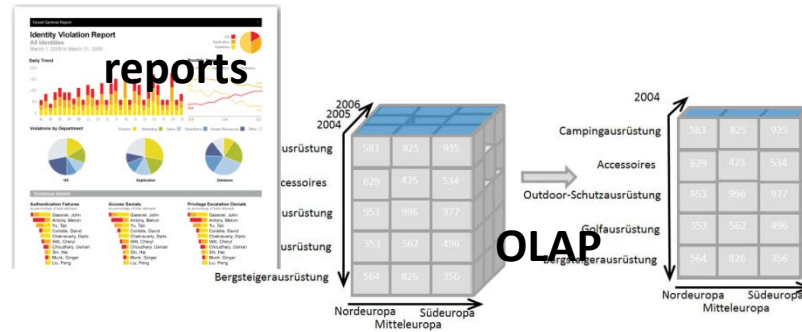
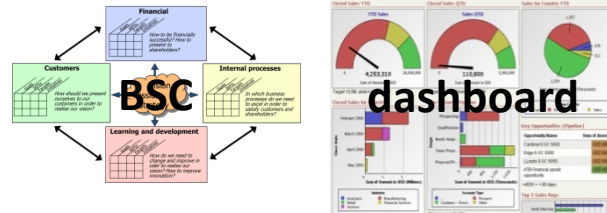
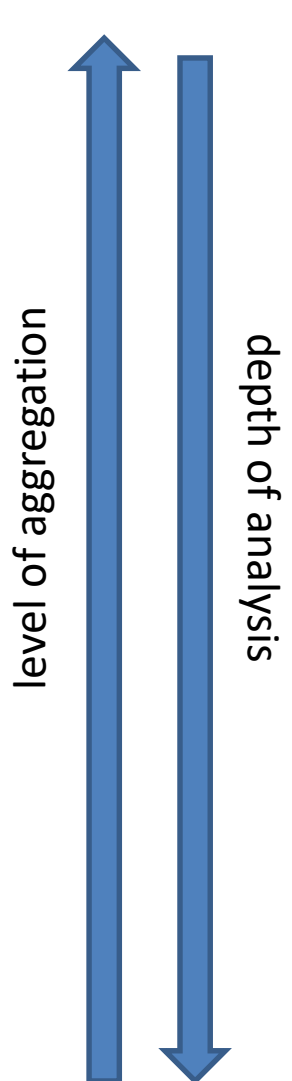
Specific requirements

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



BI front-ends and the questions they answer



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 ...

Make decision

8. apply rules
9. reuse cases



Analysis: BI tools – front-end

1. monitoring:

- ◆ 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 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

OLAP

query tools
CBR

data mining
tools



Requirements for BI tools – front-end

1. monitoring :

- ◆ 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*



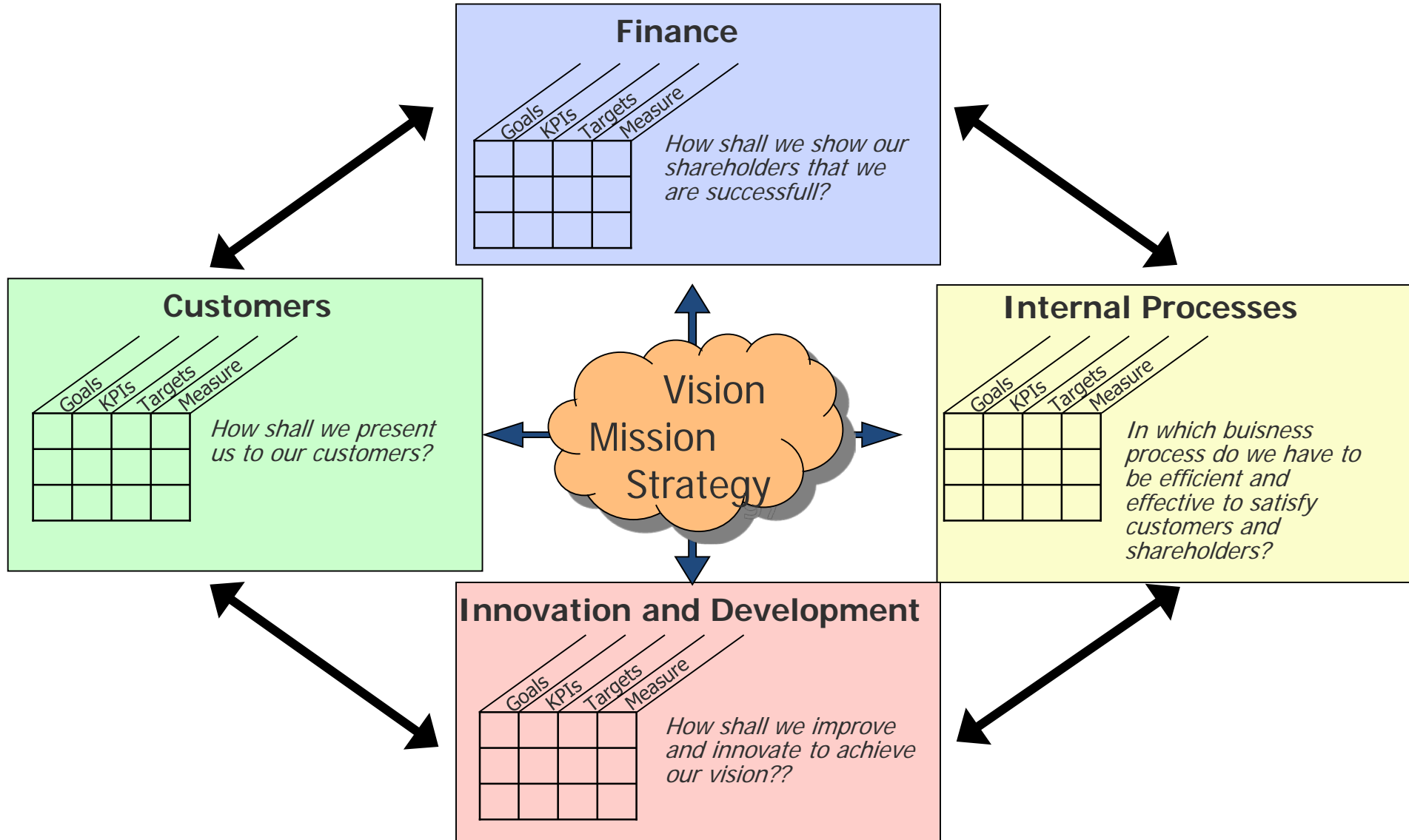
A solid orange vertical bar is positioned on the left side of the slide.

DEFINITION OF KPIS

Definition of KPIs starts with the Goals

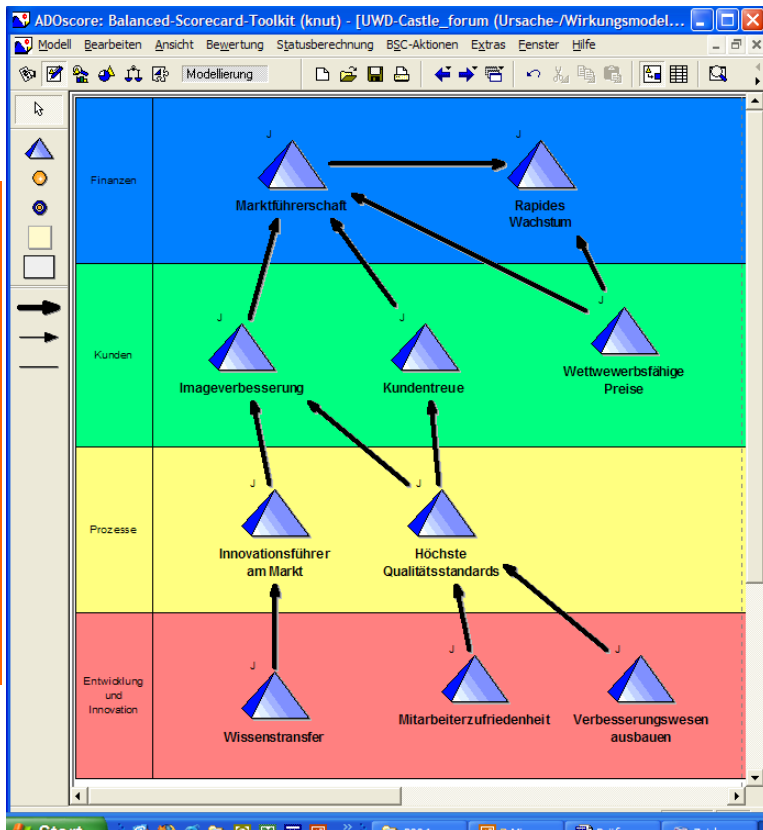
- KPIs (Key Performance Indicators) help to measure the achievement of goals.
- This means, KPIs only make sense, if the goals are known
- Steps for KPI definition:
 1. Define goals
 2. Determine KPIs for the goals
 3. Determine target values for the KPIs
 4. Measure the KPIs periodically
 5. Show KPIs and their values in a dashboard or report

Example: Balanced Scorecard



1. Defining Goals

Example: ADOscore ¹⁾

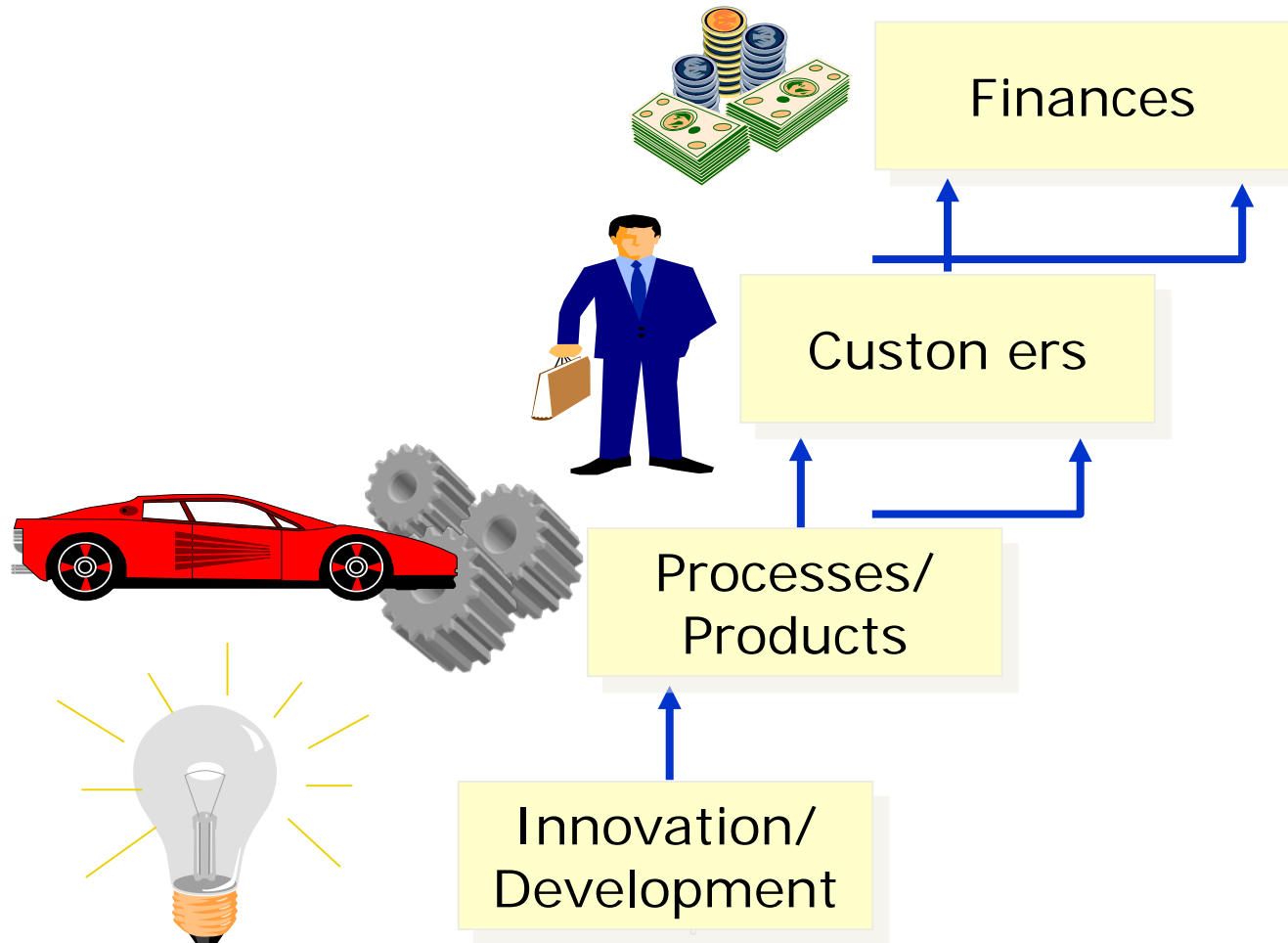


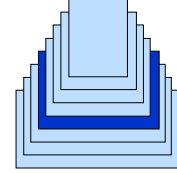
- Determine goals based on strategy and success factors
- Assign goals to perspectives
- Cause-effect relationship can help to determine means of a goal is not achieved:
 - ◆ Which other goals have an influence on goal X?

1) ADOscore is a tool from BOC to graphically support the development of scorecards



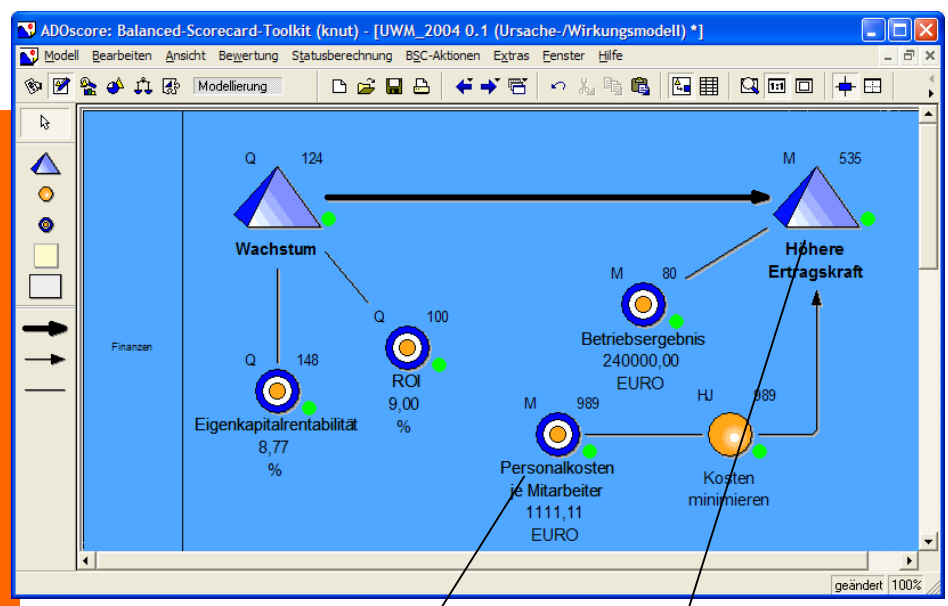
Principle: Cause-Effects are directed towards Success of the Enterprise





2. Determine KPIs

Example: ADOscore

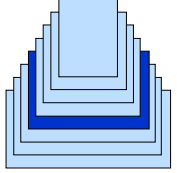


KPI

goal

- For each goal appropriate indicators are determined
- KPIs are indicators to measure, whether a goal is achieved
- Specify indicators such that
 - ... the achievement of the goal can be recognized from the indicator
 - ... the behaviour of involved peoples in directed toward the goal achievement



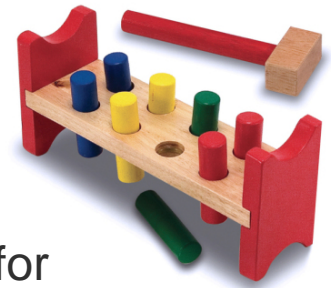


3. Specify Target Values for KPIs



- Specify target values for the indicators
 - ◆ Types of targets:
 - minimal value (if a high value is expected)
 - e.g. ROI, revenue, number of customer
 - maximal value (if low value is beneficial)
 - e.g. costs, time
 - interval (if value should be in a specific range)
 - e.g. number of employees

Choosing the right KPIs – example 1



The company WoodToy Inc. designs and produces funny wooden toys for children and adults. An important goal in their strategy is to increase innovation, i.e. to have more innovative products. Recently, they have introduced an idea management system where employees from the R/D department can place ideas for new innovative products. The following two KPIs have been proposed for measuring the achievement of their strategic goal “*increase innovation*”:

- ◆ KPI 1: Number of ideas in the idea management system
- ◆ KPI 2: Percentage of sales revenue attributable to new products that came out of the idea management system

Which KPI do you consider more suitable in order to measure achievement of the strategic goal?



Choosing the right KPIs – example 2



The company YourERP sells licences for ERP software that they develop. In all licences, technical support is included at no additional cost. YourERP has learned that the satisfaction of their customers with their products depends to a large degree on how fast support issues are solved. They have defined the goal “*solve support issues faster*”. The YourERP management has come up with the following alternative measures for that goal:

- ◆ KPI 1: average time (number of days) to first response
- ◆ KPI 2: average time (number of days) to final ticket resolution
- ◆ KPI 3: first contact resolution (= number of tickets resolved with the first response)

Which KPI do you consider more suitable in order to measure achievement of the strategic goal?



Choosing the right KPIs – example 3



A telecommunications company sells many contracts to new and existing customers through telesales campaigns. The goal is to *maximise the profit that results from the campaigns*. The company has the ability to create predictive models of customer behaviour. Two indicators have been discussed for measuring the achievement of the goal:

- ◆ KPI 1: revenue from all contracts sold through campaigns
- ◆ KPI 2: the net value (over the next 5 years) of all contracts sold through campaigns

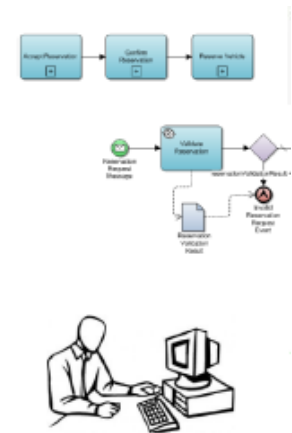
***Which behaviour will each of these KPIs trigger?
Which one is better suited for measuring achievement
of the goal?***



Choosing the right KPIs – example 4

A public administration serves citizens with various services. The underlying business processes are complex and knowledge-intensive, i.e. many exceptional situations can arise. In order to be able to measure more accurately the time needed for case completion, the management has introduced a workflow management system through which all cases should be handled. They are planning to measure the following KPI

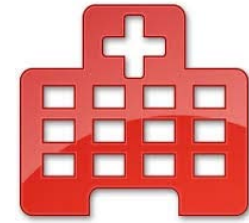
- ◆ KPI: number of cases completed within a month



***Do you foresee any problems with this KPI?
Which? What do you propose to avoid them?***



Choosing the right KPIs – example 5



In Switzerland, physicians' activities are billed using so-called «tax points» that are proportional to the (assumed) difficulty of the activity and that are defined by the law for each activity. However, in some areas – e.g. psychosomatic medicine – tax points are systematically lower for almost all activities. A hospital wishes to assess the achievement of the strategic goal «increase efficiency of physicians» via the KPI

- ◆ KPI: number of hours per tax point.



Do you foresee any problems with this KPI? Which? What do you propose to avoid them?

4. Dashboard: Show Goal Achievement

Comparing values of KPIs with their target values

Example: ADOscore



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

DASHBOARDS

Definitions: Dashboards and Reports

A dashboard is a visual display of the most important information needed to achieve one or more objectives; consolidated and arranged on a single screen so the information can be monitored at a glance.²

from: C. Howson. Successful Business Intelligence

- **Dashboards** usually contain actual vs. targeted values of KPIs and thus...
 - ... are often closely related to strategic goals defined e.g. in a Balanced Scorecard
 - ... are used by management to visualise goal achievement
 - ... contain highly aggregated data
- A **report** is a summary of business data, typically in the form of tables and/or diagrams

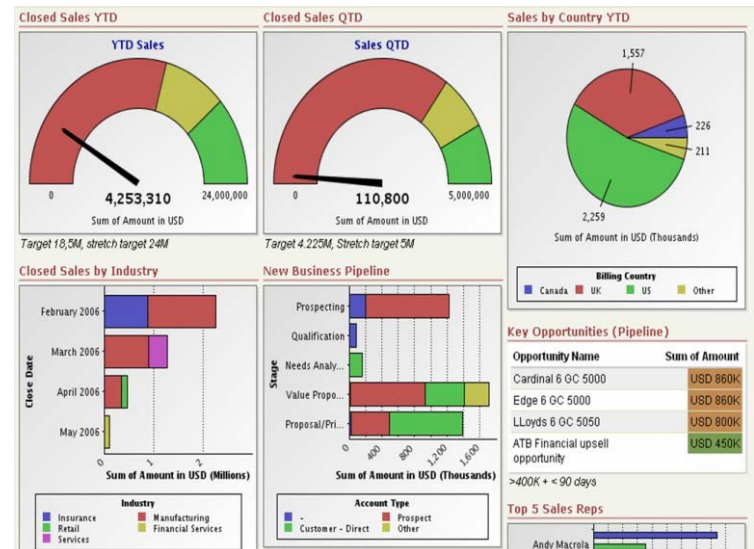
Dashboards – purpose

- Monitor performance of an organisation

When you want to know how your car is doing, the dashboard tells you, at a glance, whether your car is healthy and progressing at the right speed toward your destination.

A well-designed business dashboard helps you understand how healthy your organization is, and whether it's progressing at a fast enough pace toward its destination.

cited from «Dashboards at Eden, Inc.» case



Common dashboard elements (1)

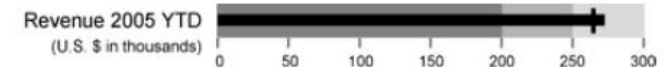
- **KPI visualisation:** (usually) needs to highlight
 - ◆ the target value
 - ◆ the actual value
 - ◆ the ranges of «red (poor), yellow (satisfactory), green (good)», if defined



thermometer



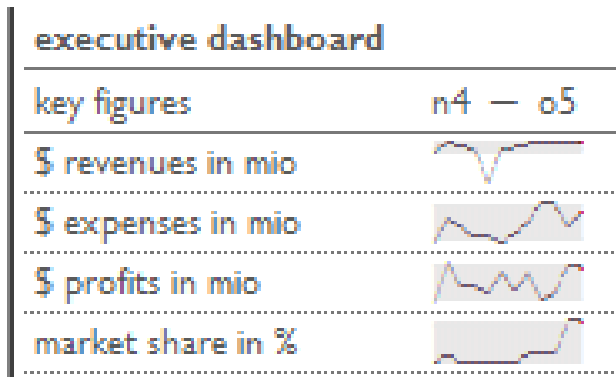
dial chart



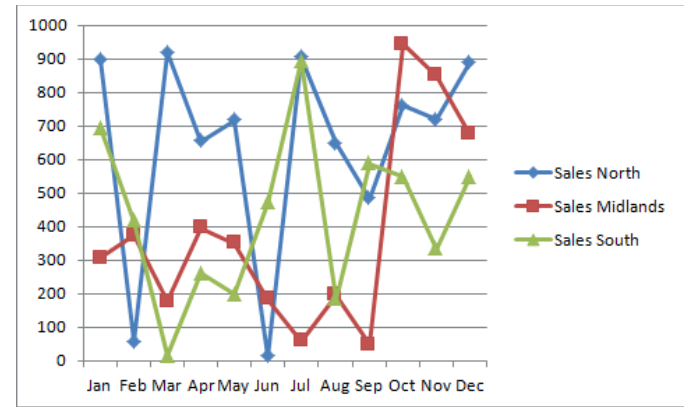
*bullet graph
(for color-blind!)*

Common dashboard elements (2)

- Date dimension: visualisation of **values over time**
 - ◆ usually displays time-line horizontally and values of a KPI vertically



spark lines
no coordinates
for general trend only



line charts
with coordinates, details
use different lines for different categories



Common dashboard elements (3)

- Place dimension: visualisation of **values by region**



regional heat map

encodes sales values with colors on a map

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

REPORTING

Reporting

- **Reporting** = gathering, presenting and formatting certain data in a meaningful way.



- *who needs the information?*
- *what decisions do they need to take?*
- *select relevant KPIs*
- *define aggregation and drill-down levels*
- *define reporting frequency and up-to-dateness*
- *how to turn data into information?*
- *define general way of presentation*
- *select visualisation paradigms*
- *(define interaction)*
- *define distribution channel(s)*
- *implement service*

Reporting – kinds and purpose

- **fixed reports:** help in monitoring important key figures and relationships on a regular basis
- **ad hoc reports:** help in making strategic and operative decisions by answering questions
- production reports: reports drawn from operational (transactional) systems or operational data stores (e.g. a list of open orders), used for operative decision-making only

ONLINE ANALYTIC PROCESSING - OLAP

Knut Hinkelmann

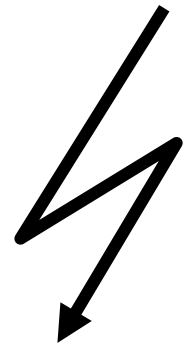
Objectives of Data Modeling

■ operative databases

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

■ Data Warehouses

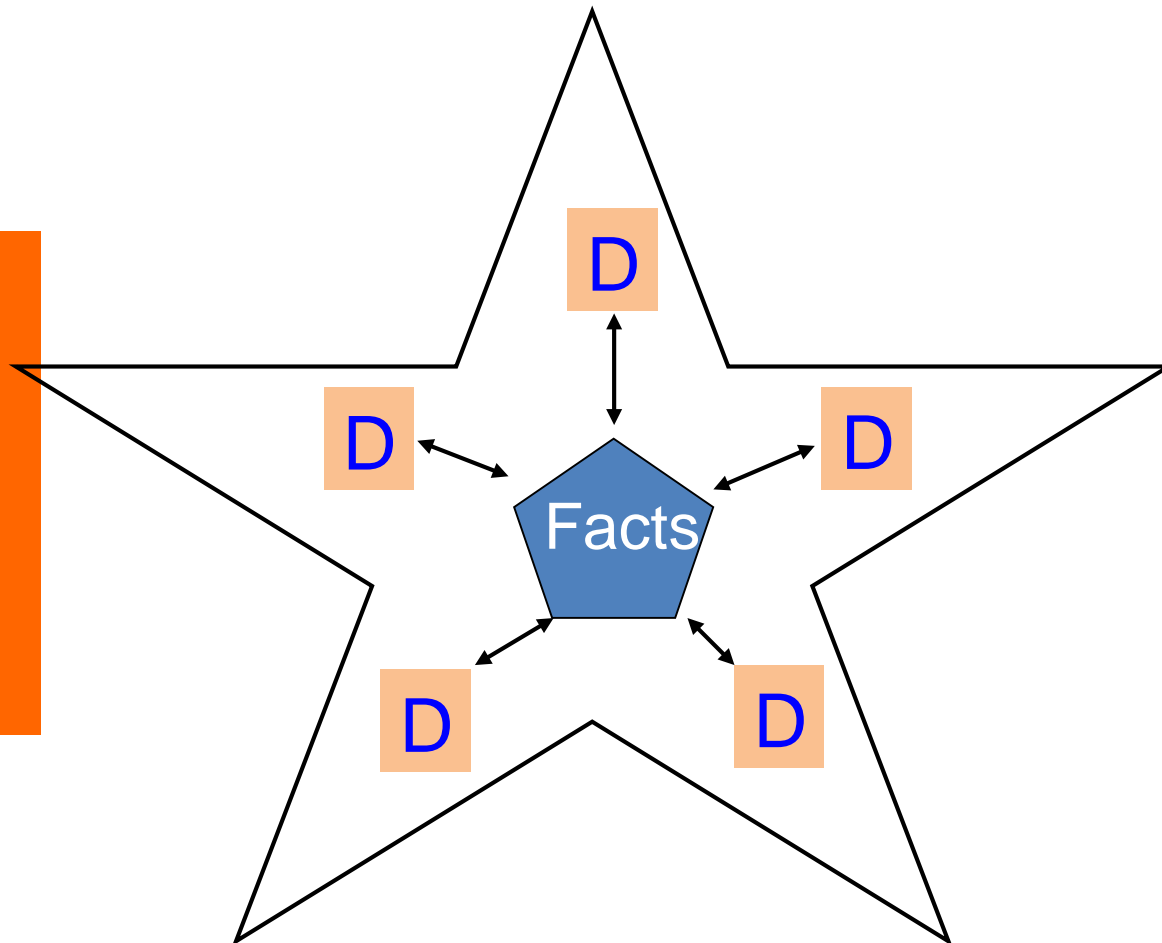
- ◆ **user friendliness** by low granularity
- ◆ **query efficiency** by denormalisation
- ◆ **close to business** by indicators and dimensions



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

cf. [Lusti, 1999]

Star Schema for Relational Data Warehouses /Marts to support OLAP

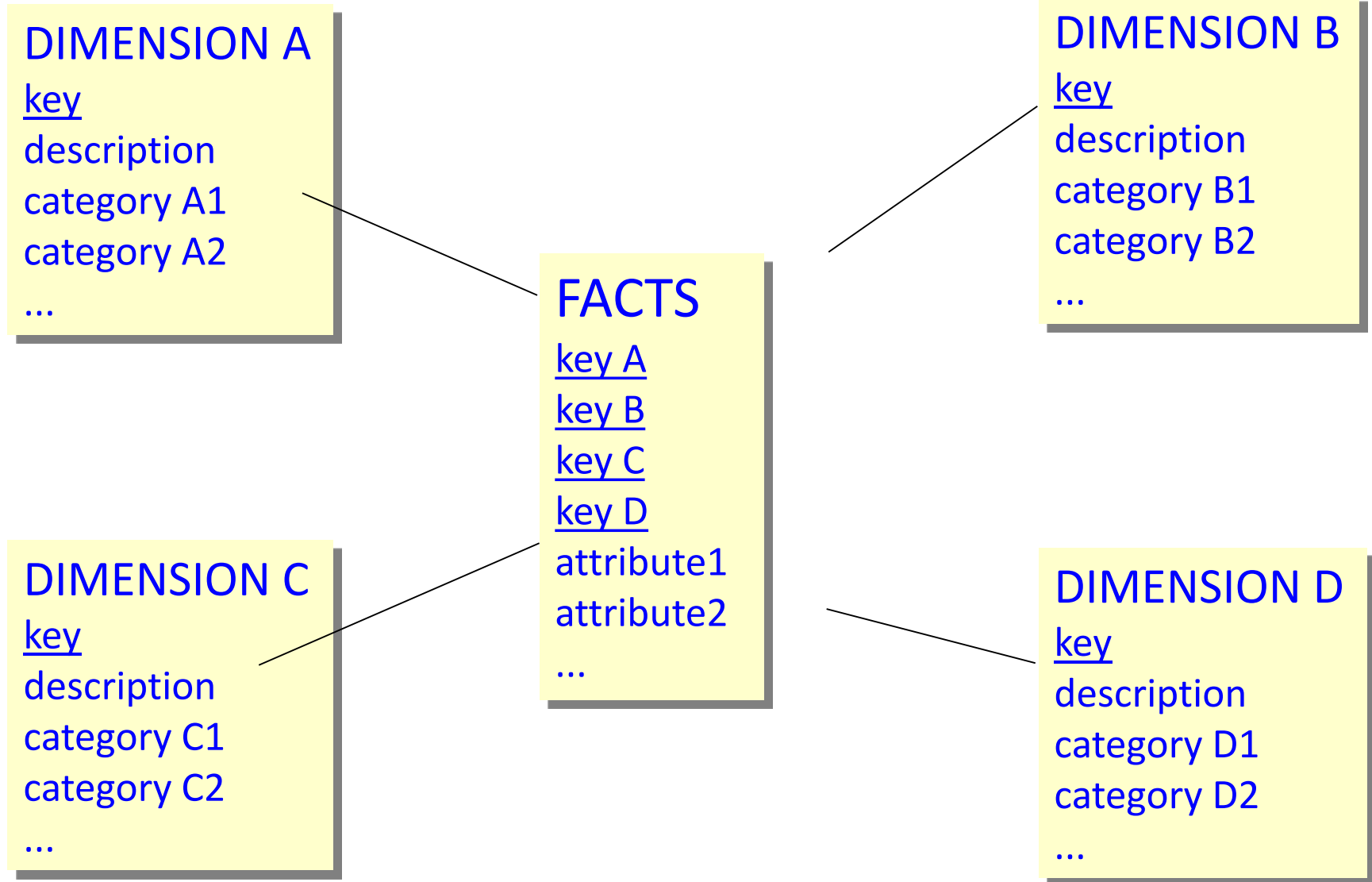


Star Schema:

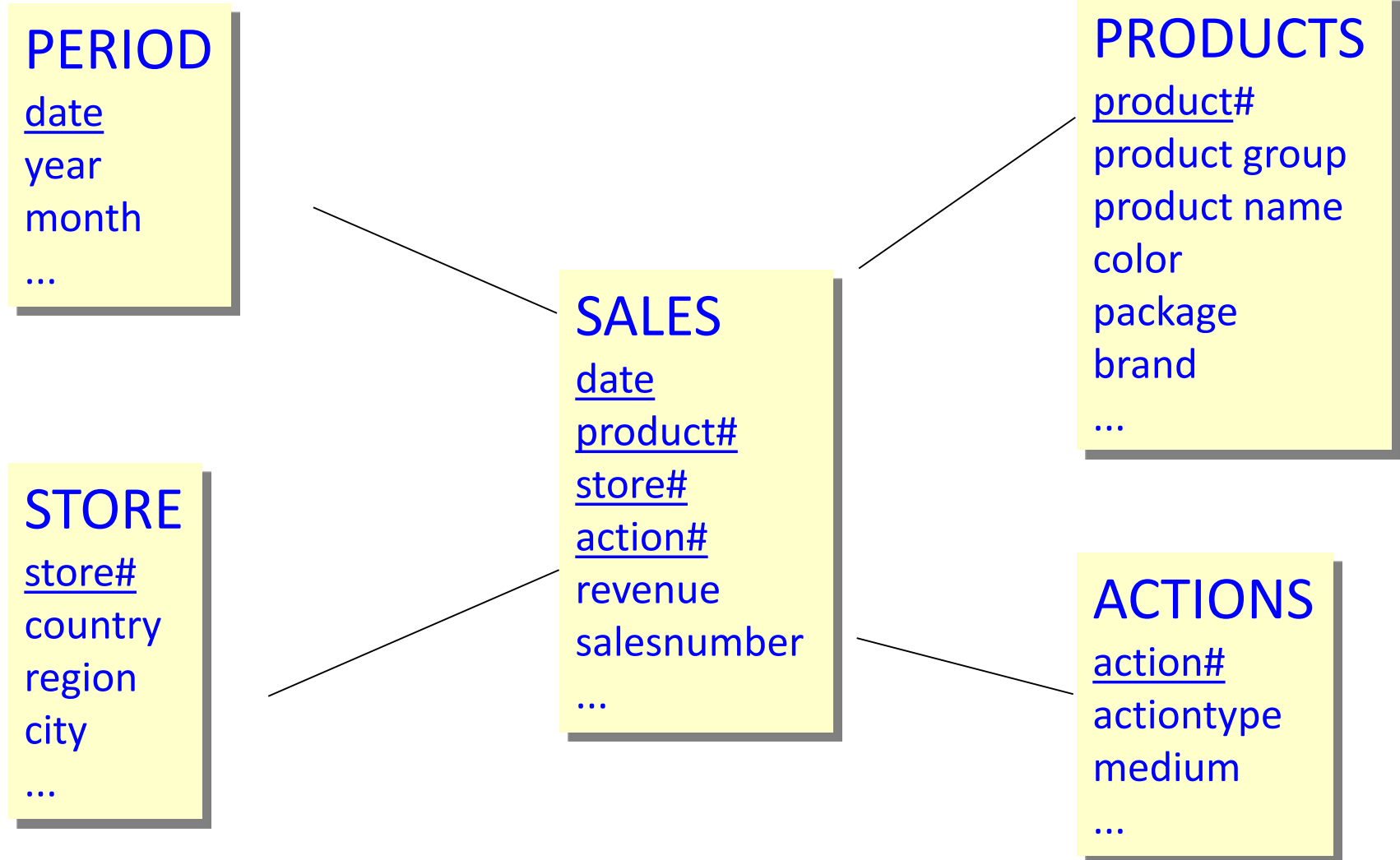
logical database schema, which places dimension tables of a relational database around a fact table for easy querying

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



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

MULTIDIMENSIONAL ANALYSIS: OLAP

Requirements for BI tools – front-end

1. monitoring:

- ◆ 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*



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

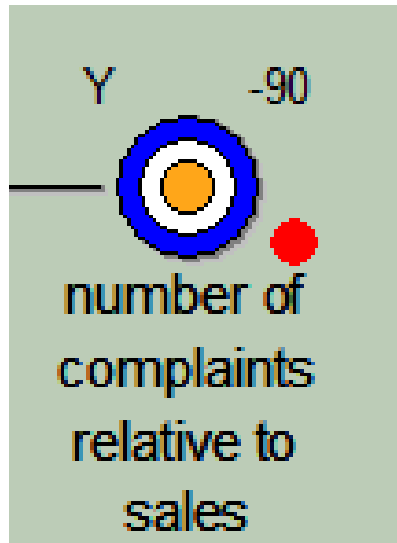
What sellers' strategies work best to yield high and profitable sales?

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:

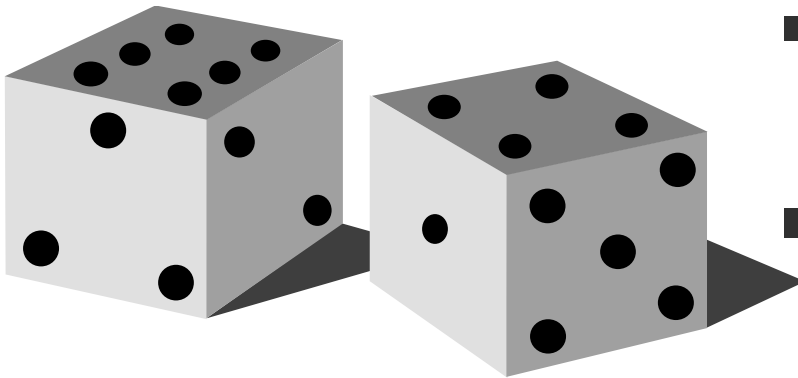
- Who is buying?
 - ◆ sales per customer (segment)
- Which channels do they use?
 - ◆ sales per channel
- When do they buy?
 - ◆ sales per quarter / month
- Where do they buy?
 - ◆ sales per region

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

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

END USER VIEW – OLAP OPERATIONS

Dicing and Slicing



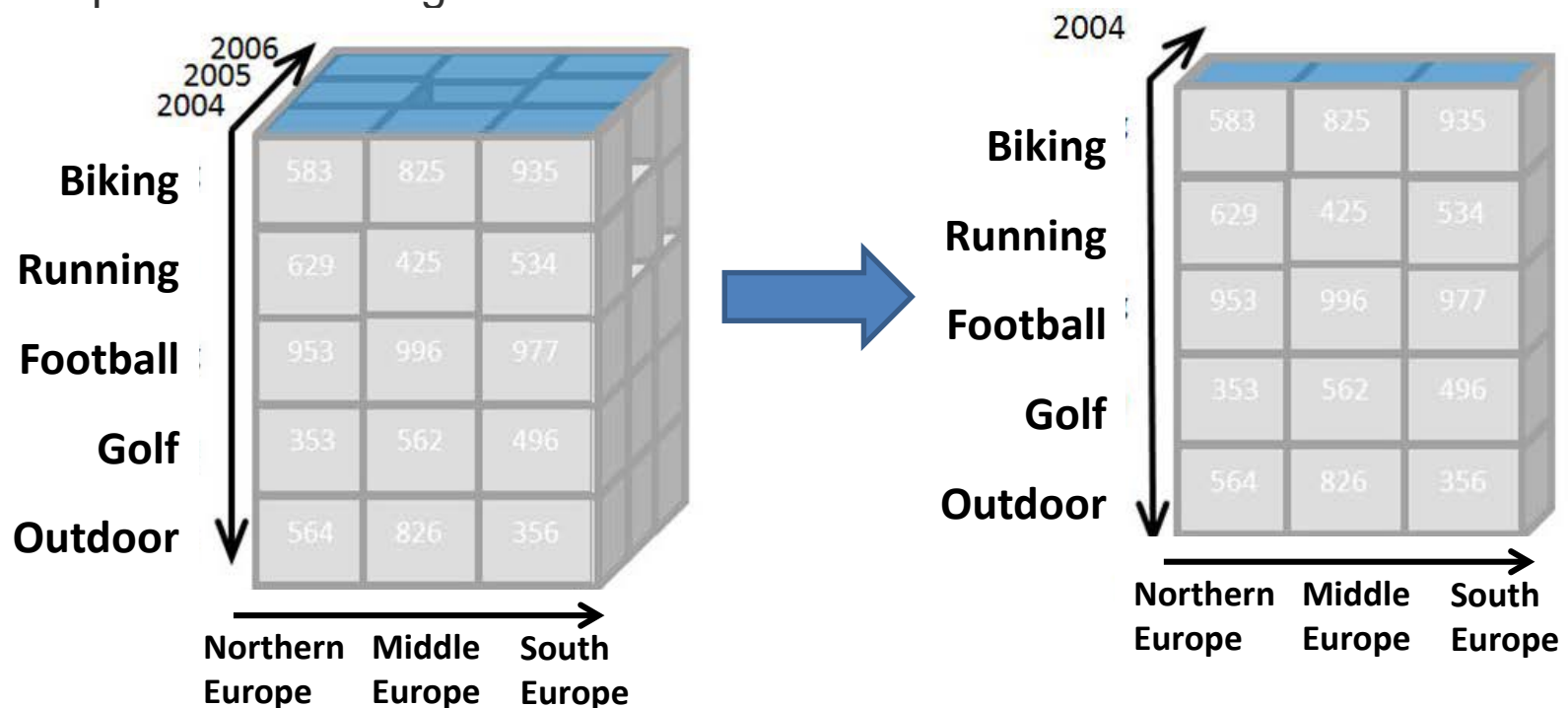
- An OLAP cube can be regarded as a multidimensional cube
- From a cube only two dimensions are visible on a two-dimensional interface (e.g. as a table)
- Slicing
 - ◆ Constraining one dimension
- Dicing
 - ◆ Constraining several dimensions
- Pivoting
 - ◆ "turning" the cube to show other dimension
-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

OLAP Operation - Slicing

Reduction of the dimensions in a multi-dimensional cube

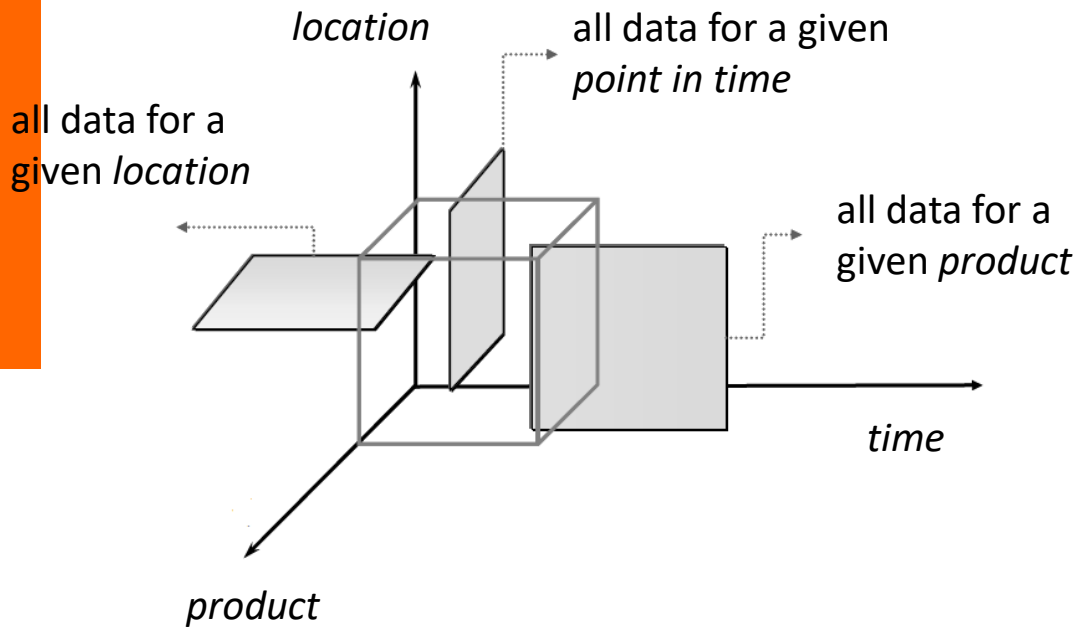
- Constraining one dimension to a particular value

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
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
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)

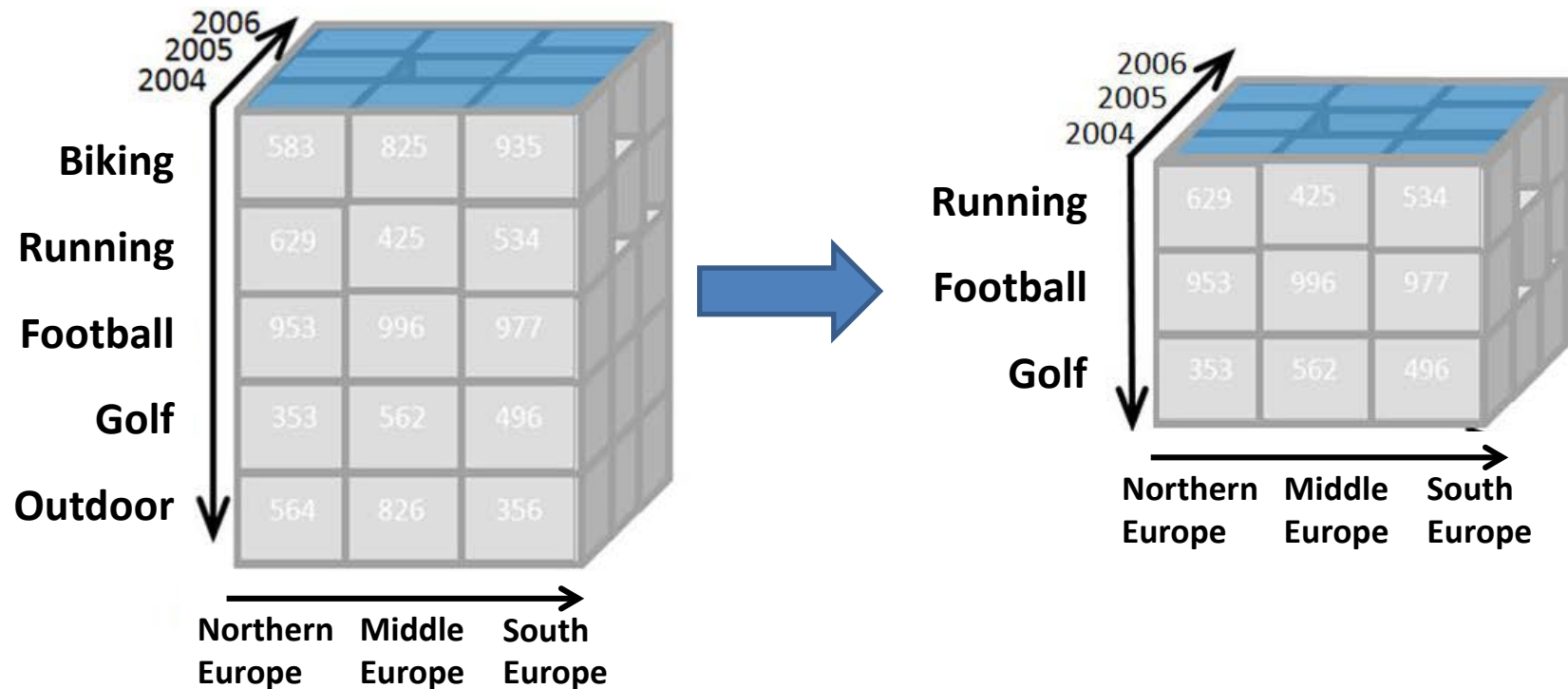
■ Example

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

OLAP Operation - Dicing

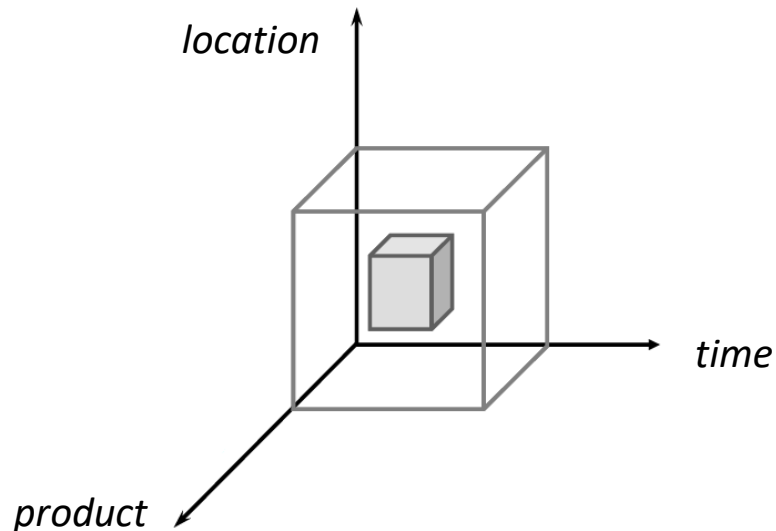
Slicing in several dimensions

- ◆ 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	
	+ Q3	+ Q4
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

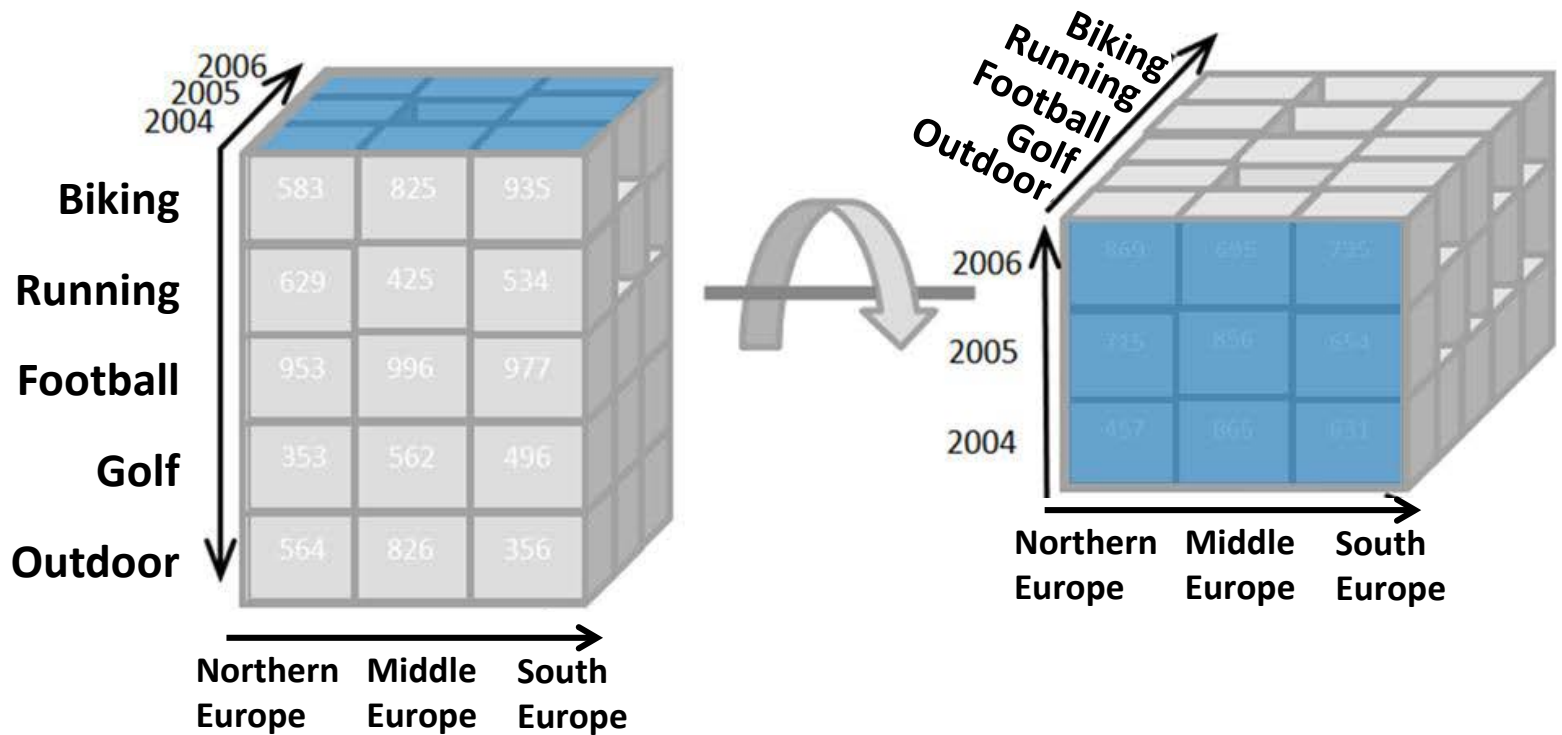


dice: month = {September, October}
products = {SB123, SB234}

	Date	
	+ September	+ October
Product	● Revenue	● 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	
	<input type="checkbox"/> Q3	<input type="checkbox"/> Q4
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
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	
	+ Q3	+ Q4
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



	Date							
	+ Q3				+ Q4			
	Store name				Store name			
	- All Store.Store names	+ Austria	+ Germany	+ Switzerland	- All Store.Store names	+ Austria	+ Germany	+ Switzerland
Product	● Revenue	● Revenue	● Revenue	● Revenue	● Revenue	● Revenue	● Revenue	● Revenue
- All Products	8.925,00	215,00	90,00	8.620,00	34.925,00	7.230,00	2.730,00	24.965,00
+ accessories	1.810,00	215,00	90,00	1.505,00	3.150,00	170,00	430,00	2.550,00
+ 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
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*

	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»*

	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*

	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]

