BI-Tools Frontend:
# 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
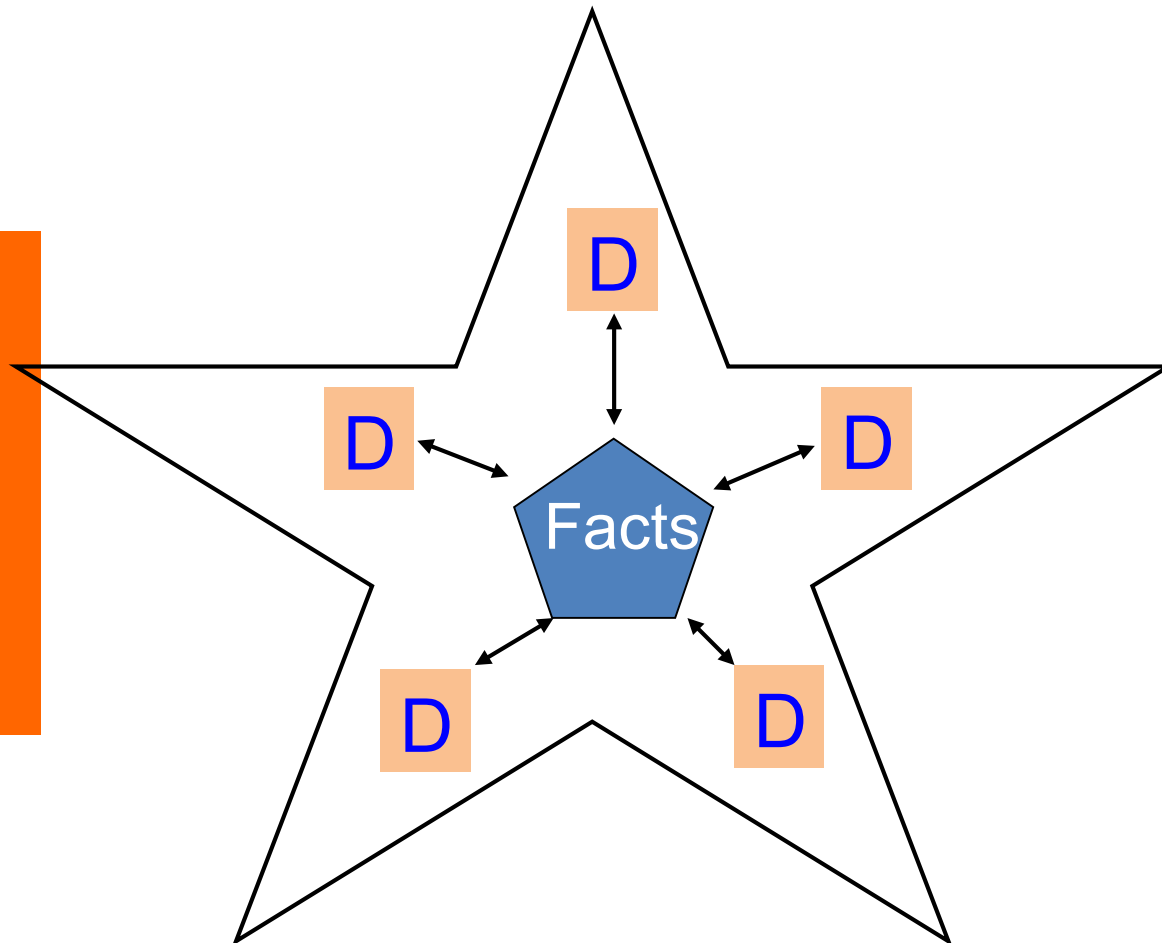
- **Data Warehouses**
  - ♦ **user friendliness** by low granularity
  - ♦ **query efficiency** by denormalisation
  - ♦ **close to business** by indicators and dimensions
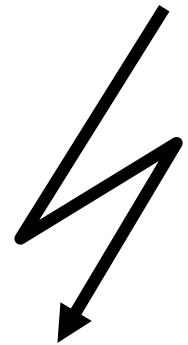
➢ Data model of operative databases cannot be reused for data warehouses

cf. [Lusti, 1999]

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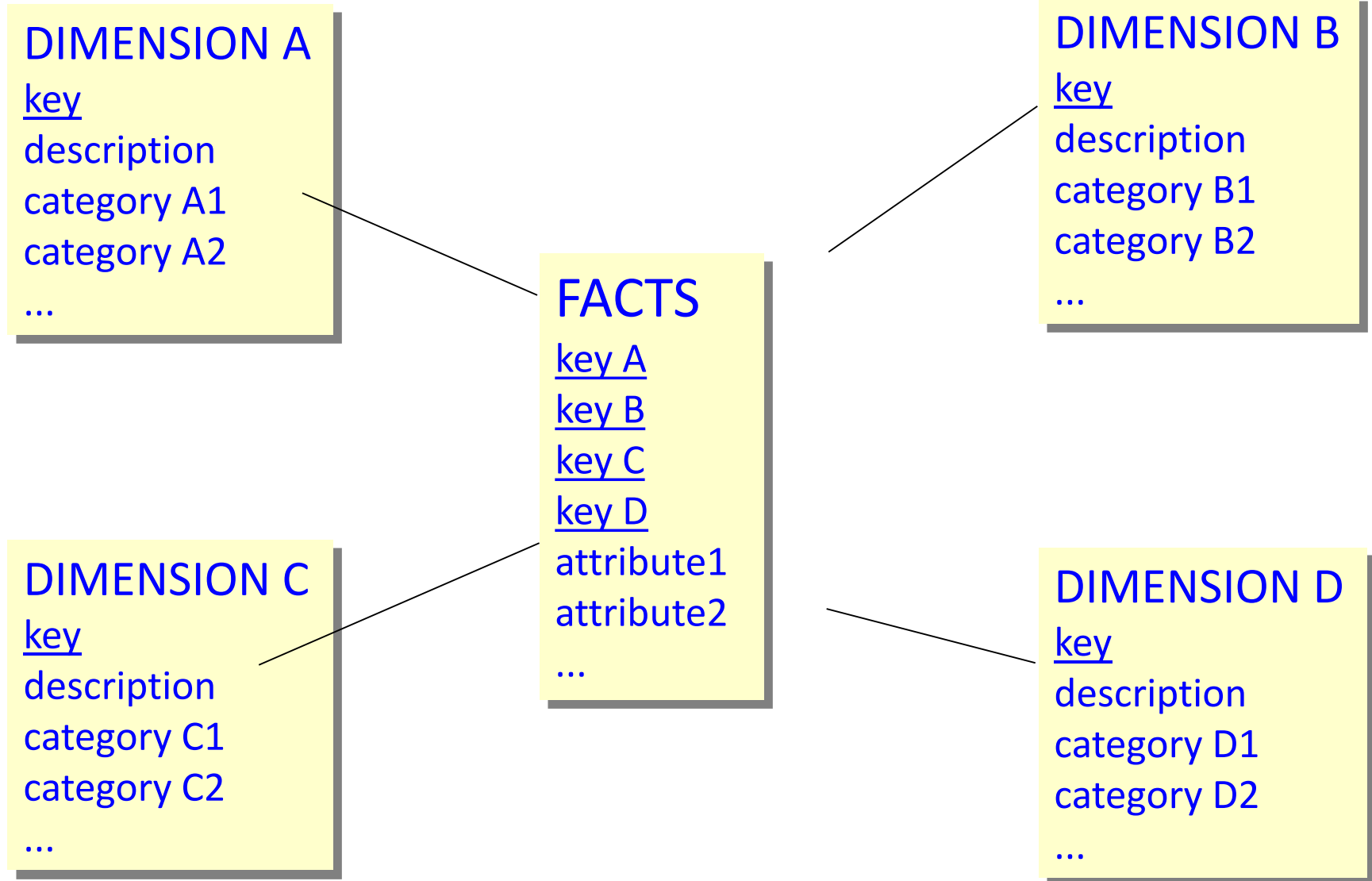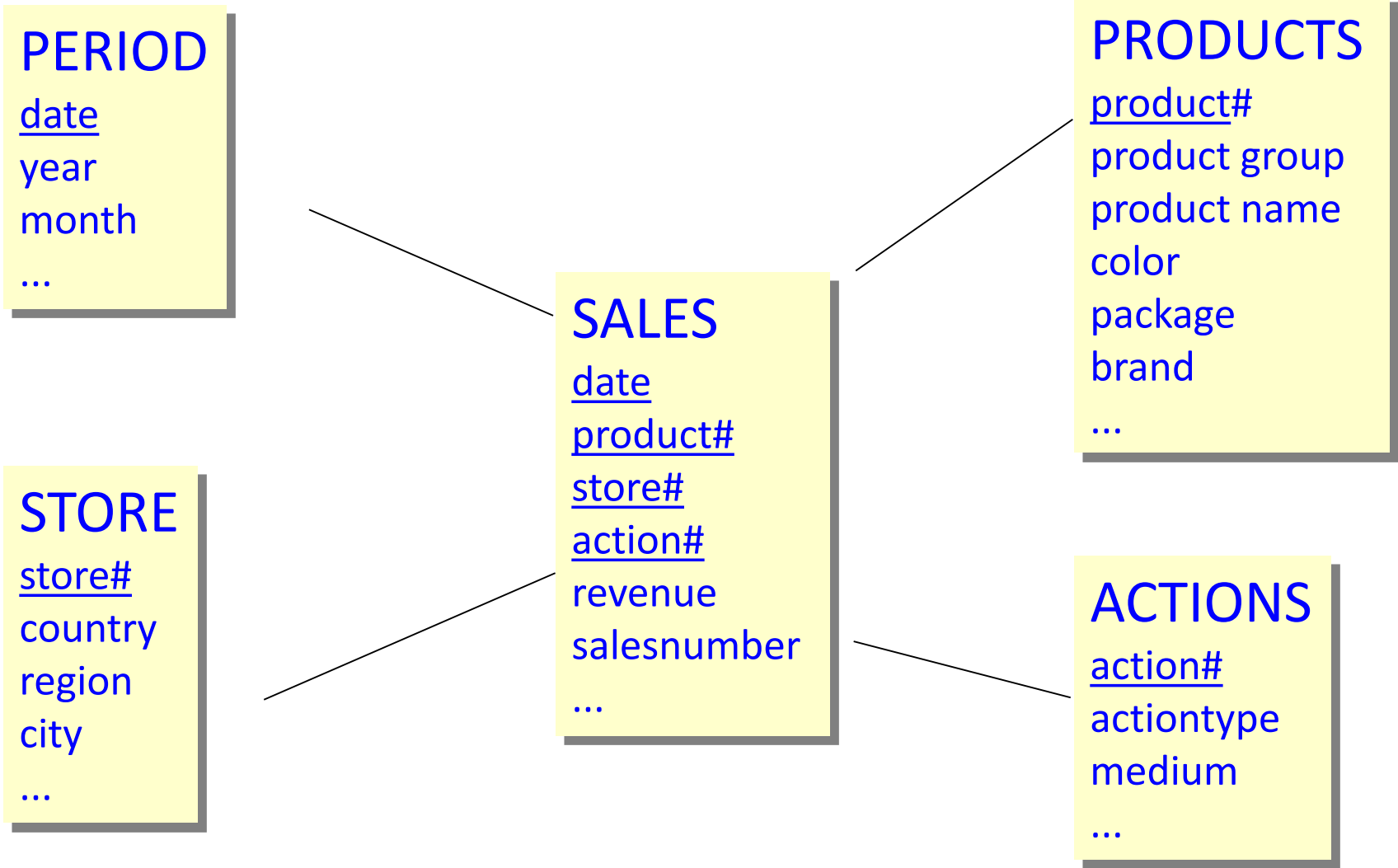
D

D   D

Facts

D   D

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

PERIOD
date
year
month
...

PRODUCTS
product#
product group
product name
color
package
brand
...

SALES
date
product#
store#
action#
revenue
salesnumber
...

STORE
store#
country
region
city
...

ACTIONS
action#
actiontype
medium
...

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

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

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

BSC tools dashboards, reports

OLAP

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:
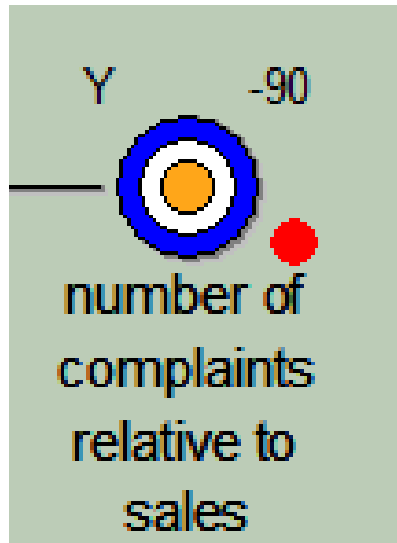
- ■ Who is complaining?
  - ♦ Number of complaints per customer (segment)

- ■ 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?
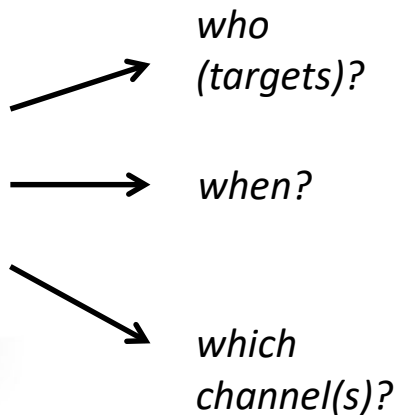
which channel(s)?

Example: how to market a new product (a «pinot gris» from Alsace)?
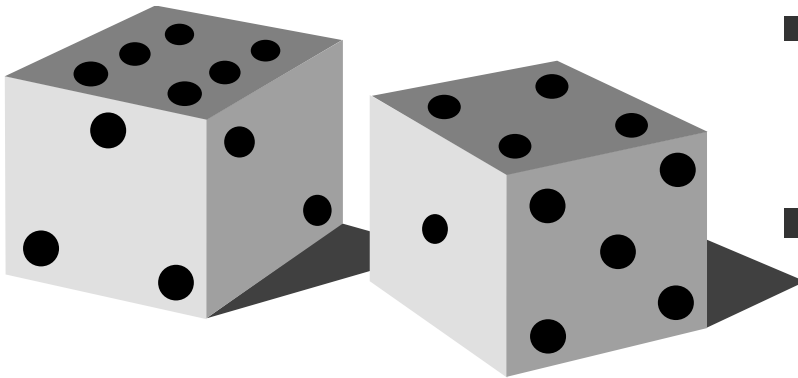
Inquiries about the product or similar ones:

■ 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

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 interfact (e.g. as a table)

■ Slicing
  ♦ Contraining 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-dimentional cube

- Constraining one dimension to a particular value

Example: Constraining time
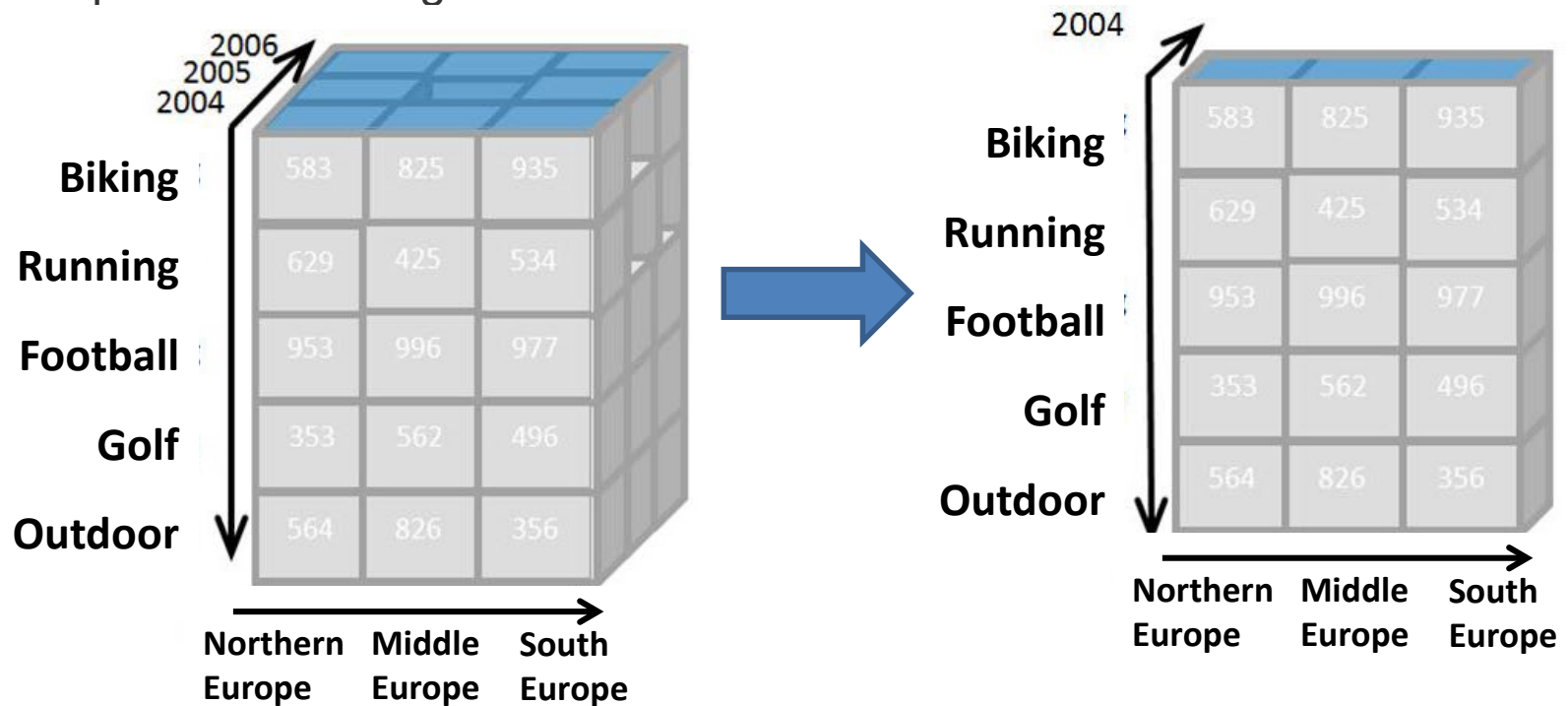
OLAP Operation – Slicing

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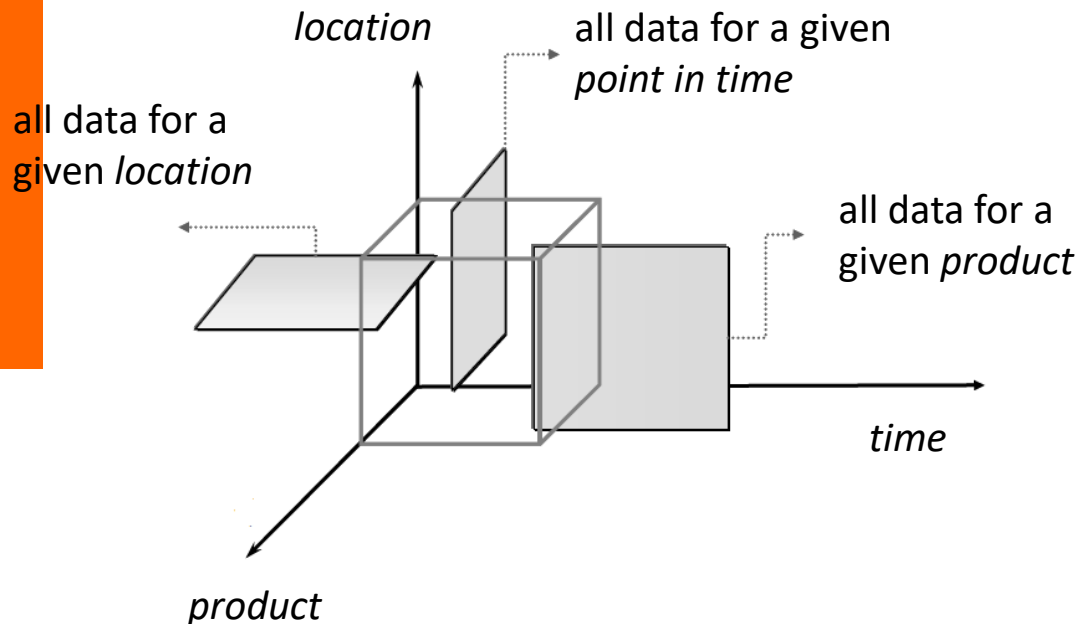
all data for a given *location*

location

all data for a given *point in time*

all data for a given *product*

time

product

	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

slice: Country = Austria

	Date	
	+ Q3	+ Q4
Product	● Revenue	● Revenue
− All Products	215,00	7.230,00
+ accessories	215,00	170,00
+ mountain bikes		7.060,00

Slicer: [Country=Austria]

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)
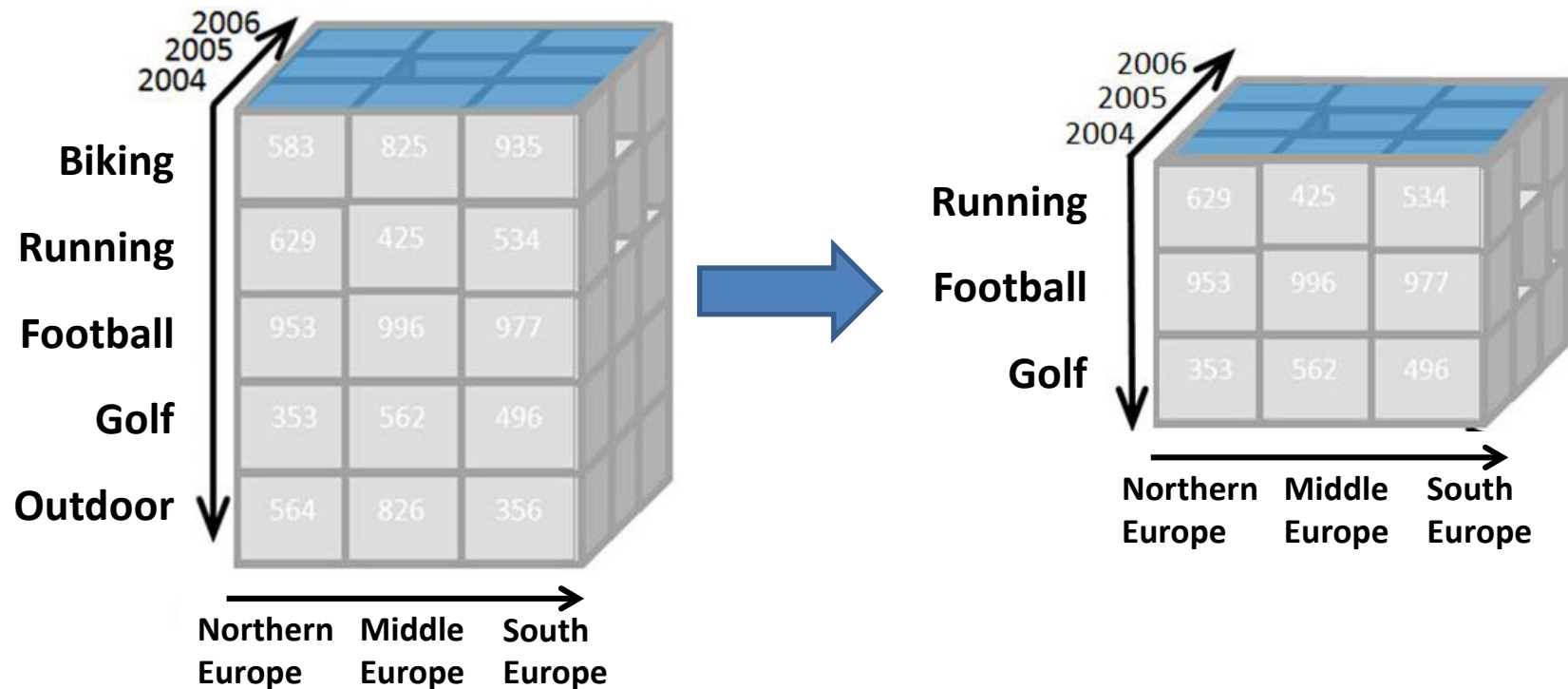
- Example
  - ♦ Showing Branch and Country of HQ
  - ♦ Slicing: Change to a different Fiscal Year
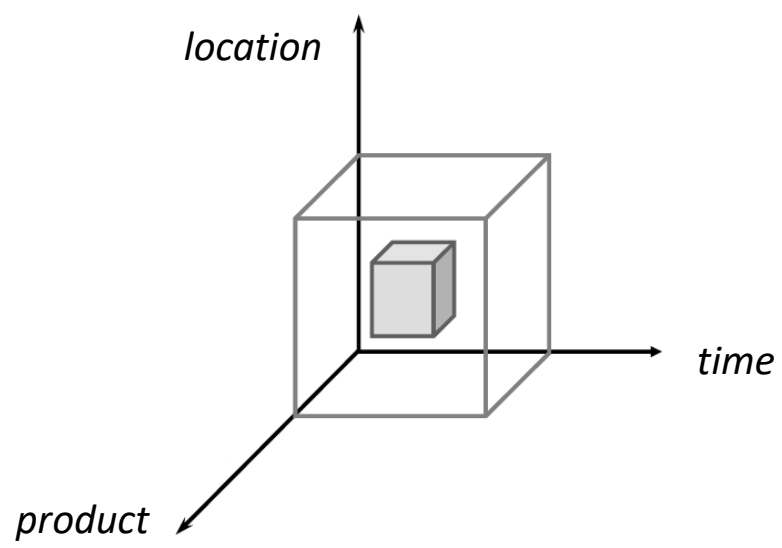
OLAP Operation - Dicing

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
⊟ All Products	8.925,00	34.925,00
⊞ accessories	1.810,00	3.150,00
⊞ 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

location

time

product

OLAP Operation - Pivoting

Rotating the cube to show other dimensions

OLAP Operation - 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

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

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

Product	Date	
	Q3	Q4
	● 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

Product	Date							
	Q3				Q4			
	Store name				Store name			
	All Store.Store names	Austria	Germany	Switzerland	All Store.Store names	Austria	Germany	Switzerland
	● 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

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
⊟ 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	● ⊞ Classic Cars	● ⊞ Motorcycles	● ⊞ Classic Cars	● ⊞ Trucks and Buses
⊟ 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	● − All Years	● + 2005	● − All Years	● + 2005	● − All Years	● − All Years
Danish Wholesale Imports	174	174				
Euro+ Shopping Channel			259	259		
Volvo Model Replicas, Co					46	171

Slicer: [Territory=EMEA]