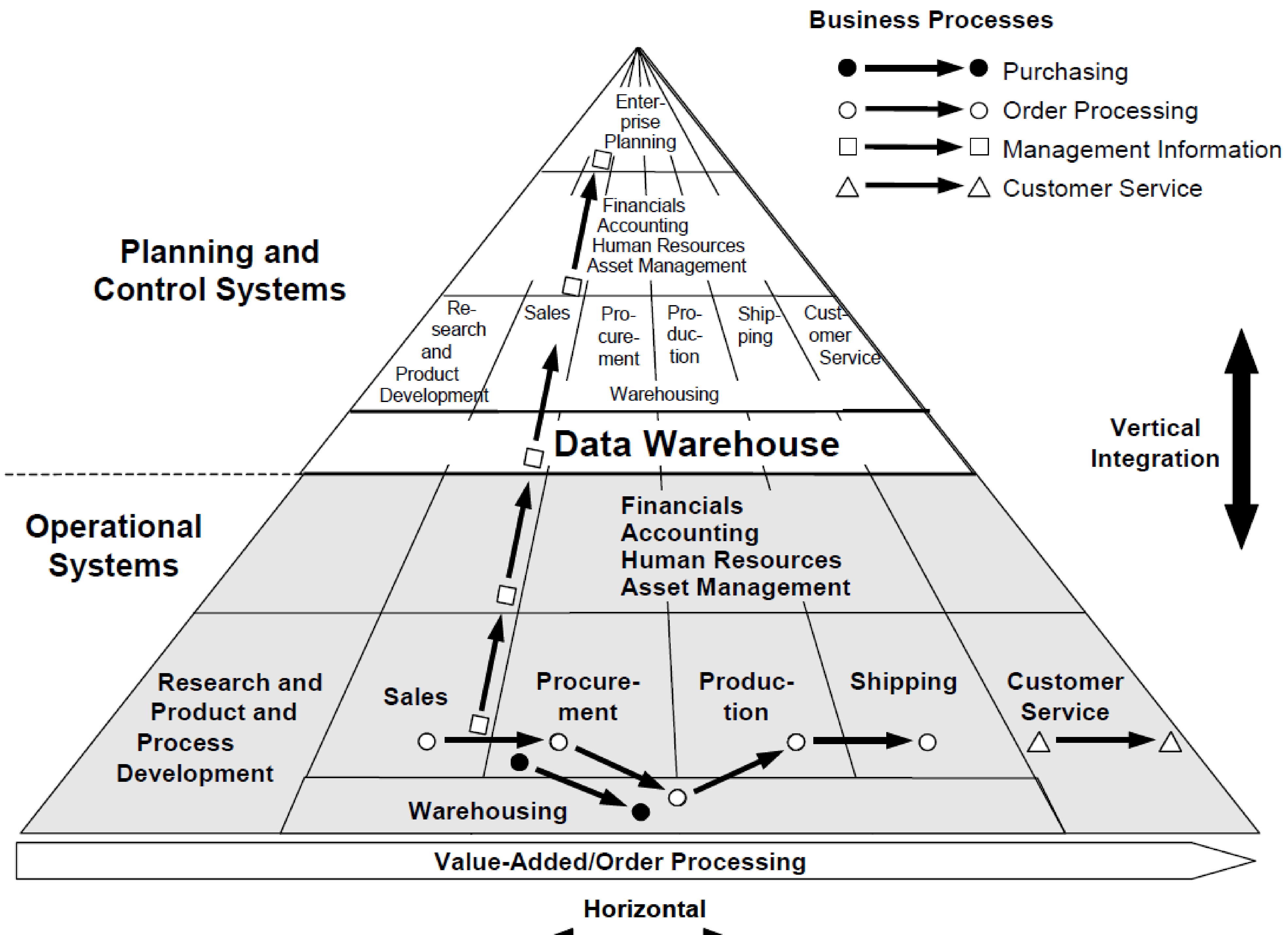
Navigation

2 Descriptive Analytics

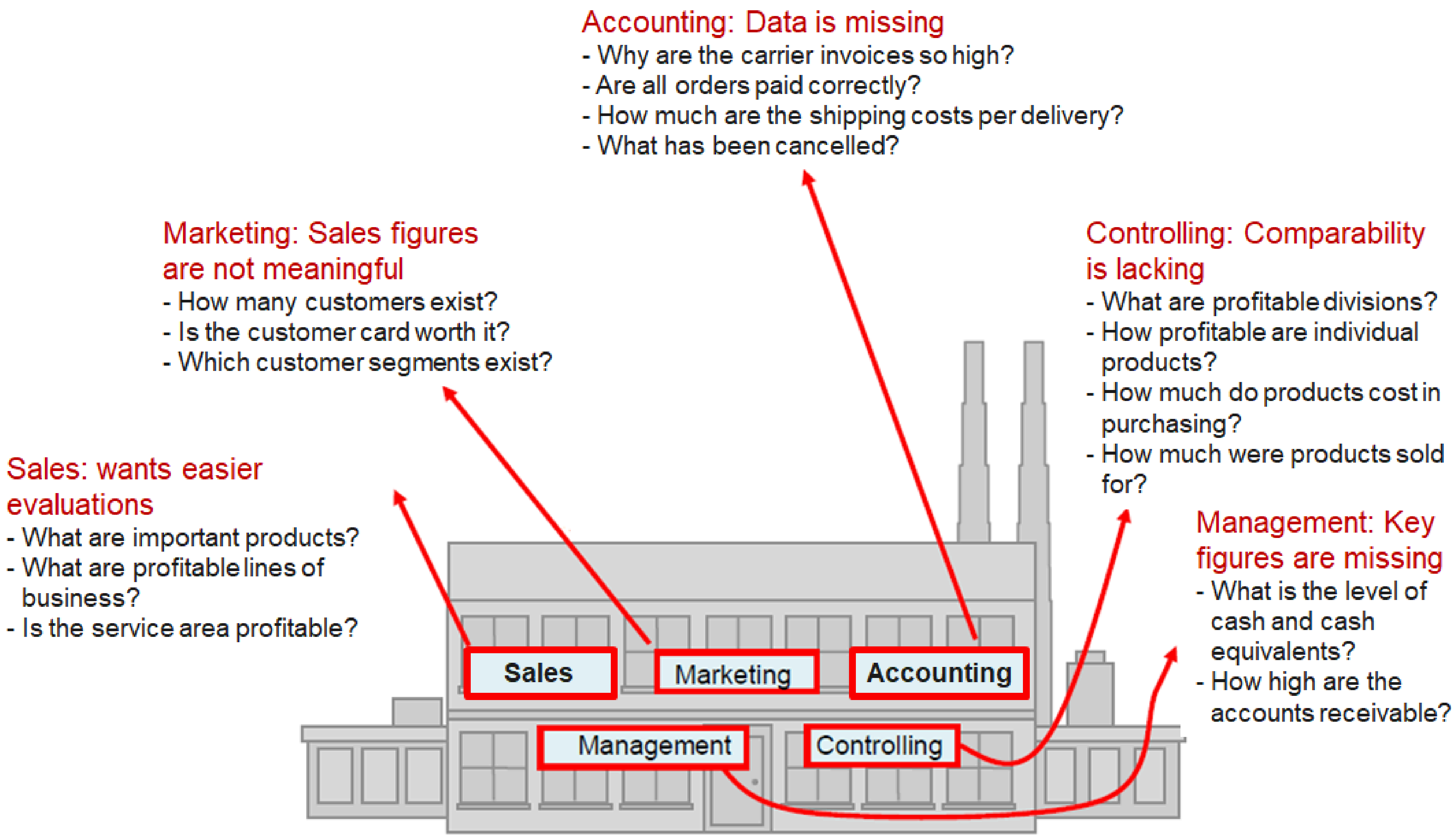
2.1 Data Warehouse Systems

2.2 Online Analytical Processing (OLAP)

Application Systems Pyramid



Typical Descriptive Questions



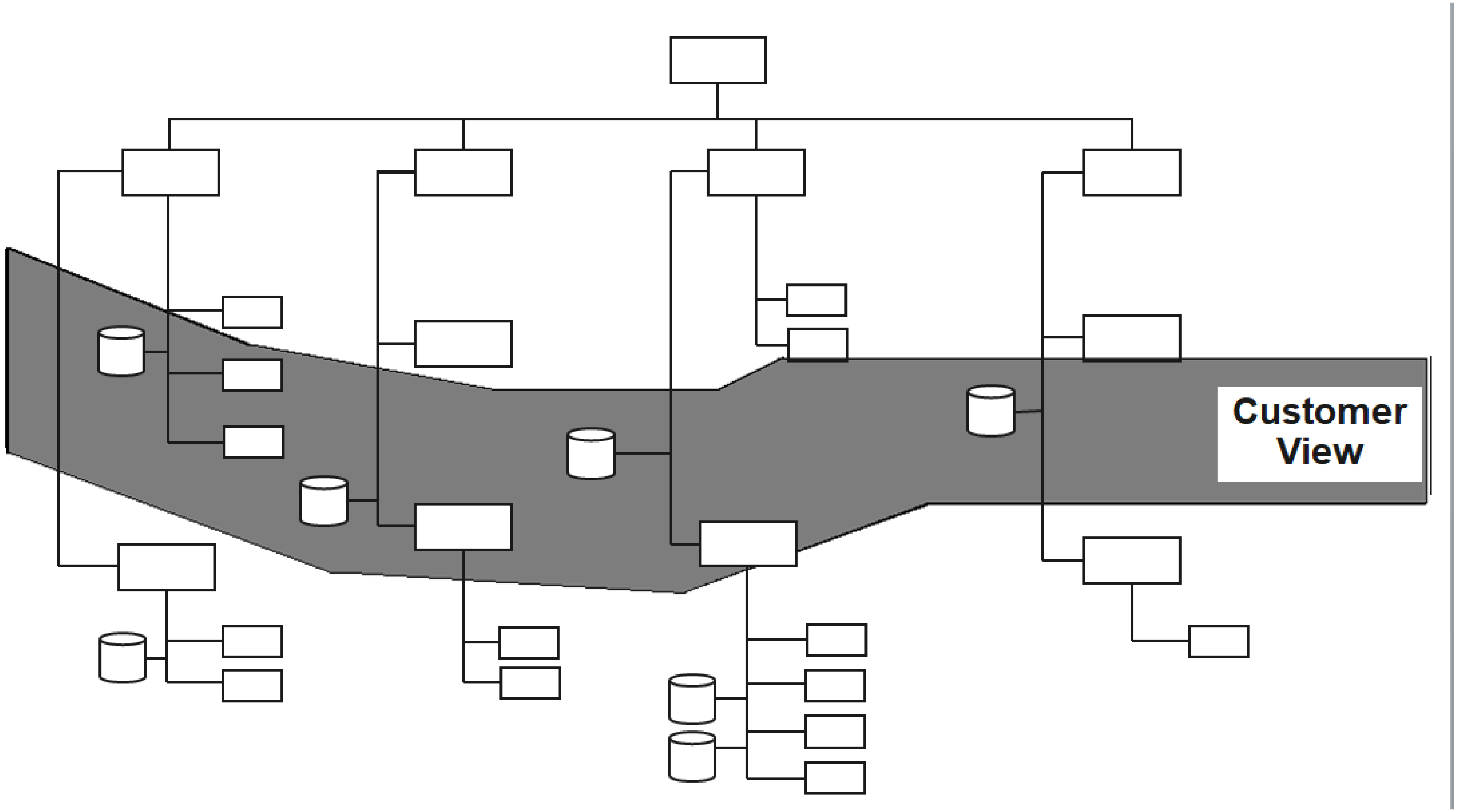
Source: Alfred Schlaucher, Oracle, p. 15

Characteristics of Operational and Analytical Databases

Frankfurt School of Finance & Management

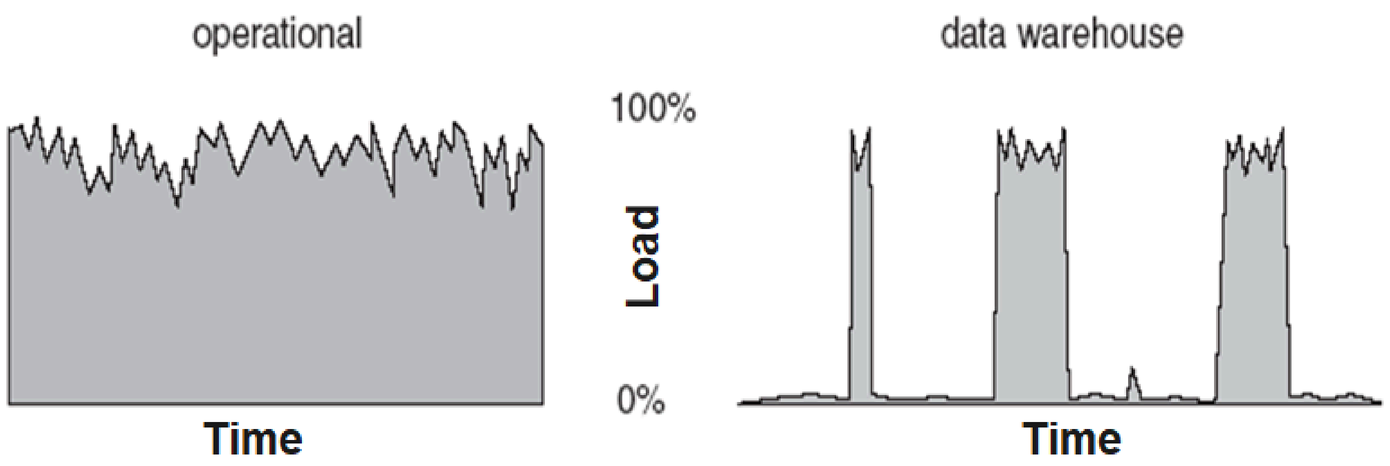
Characteristics	Operational Data	Analytical Data
Target group	Operational employees	Management, Analysts, Divisional manager
Access frequency	high	slow to medium
Data volume per access	low to medium volume	high volume
Required response time	very short	short to medium
Level of data	detailed	aggregated, processed
Queries	predictable, periodic	unpredictable, ad hoc
Given period	current	past to future
Time horizon	1–3 months	several years up to decades

Contradiction between Operational Data Architecture and Analytical Information Needs



Source: Jung/Winter (2000): Data Warehousing Strategie: Erfahrungen, Methoden, Visionen, Berlin, p. 7

The hardware load differs in operational application systems and data warehouse systems over time



=> A data warehouse relieves the

Definition: Data Warehouse

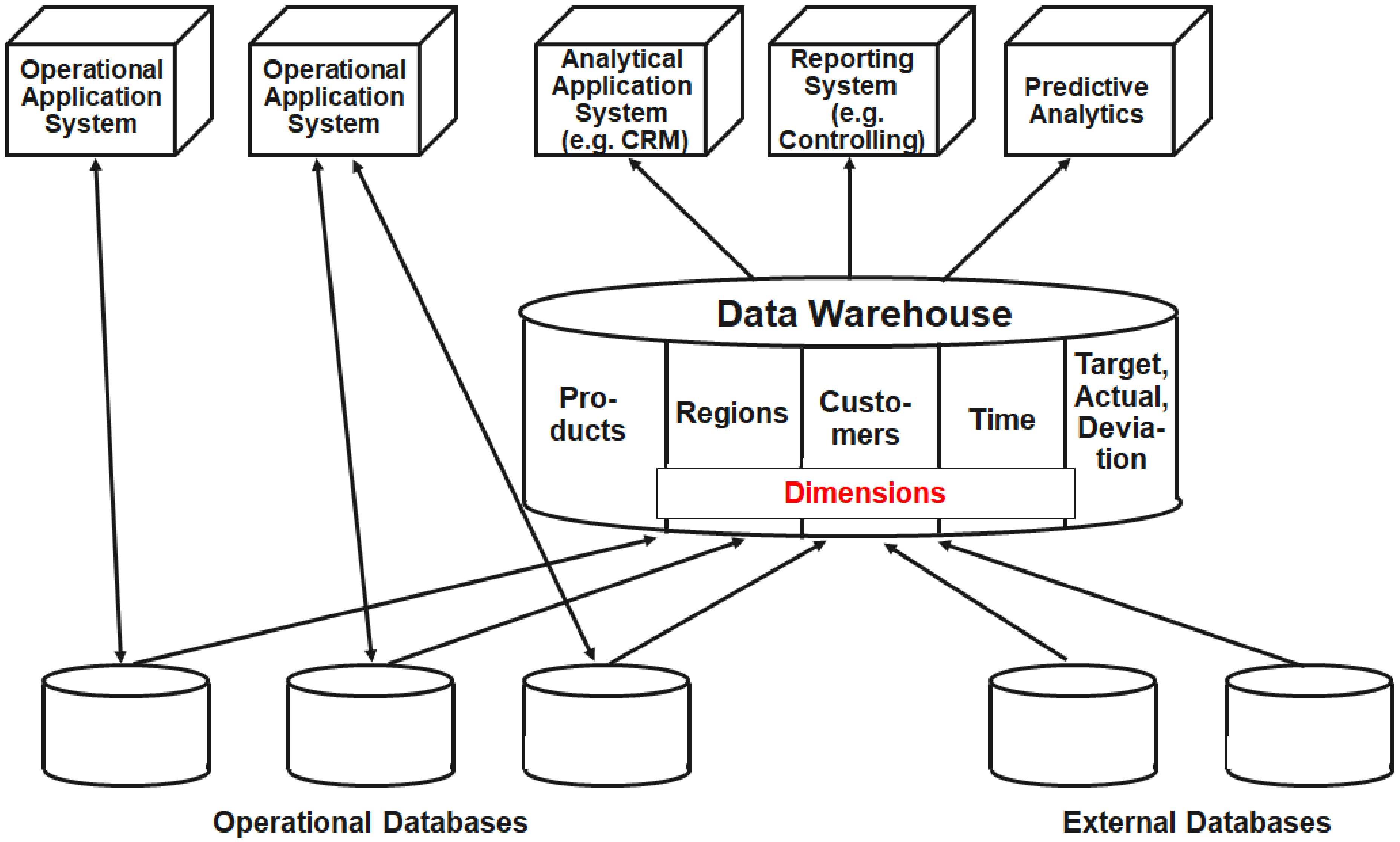
The term Data Warehouse was coined by Bill Inmon in 1990, which he defined in the following way:

"A data warehouse is a subject-oriented, integrated, time-variant and non-volatile collection of data in support of management's decision making process."

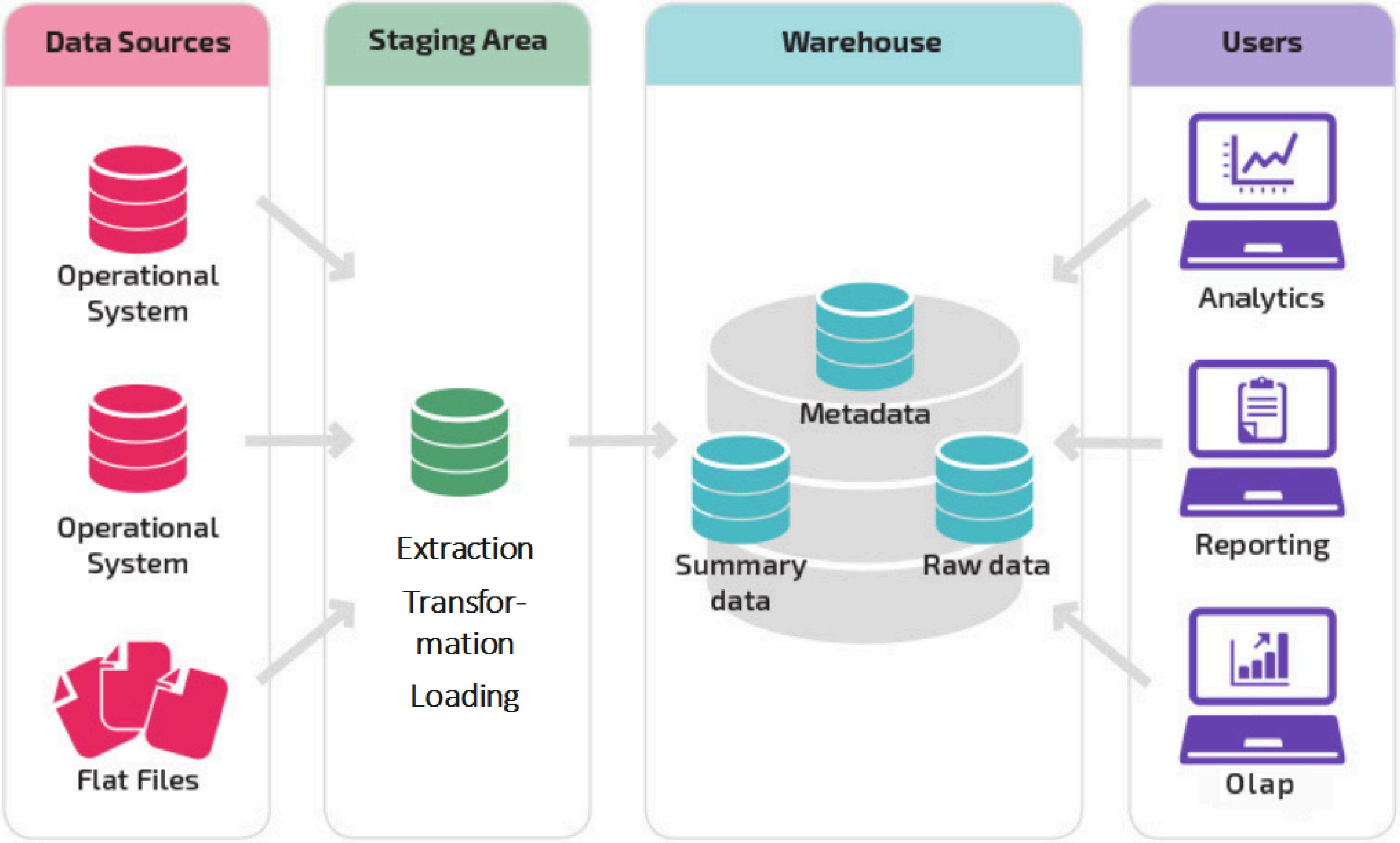
He defined the terms in the sentence as follows:

Subject Oriented: Data that gives information about

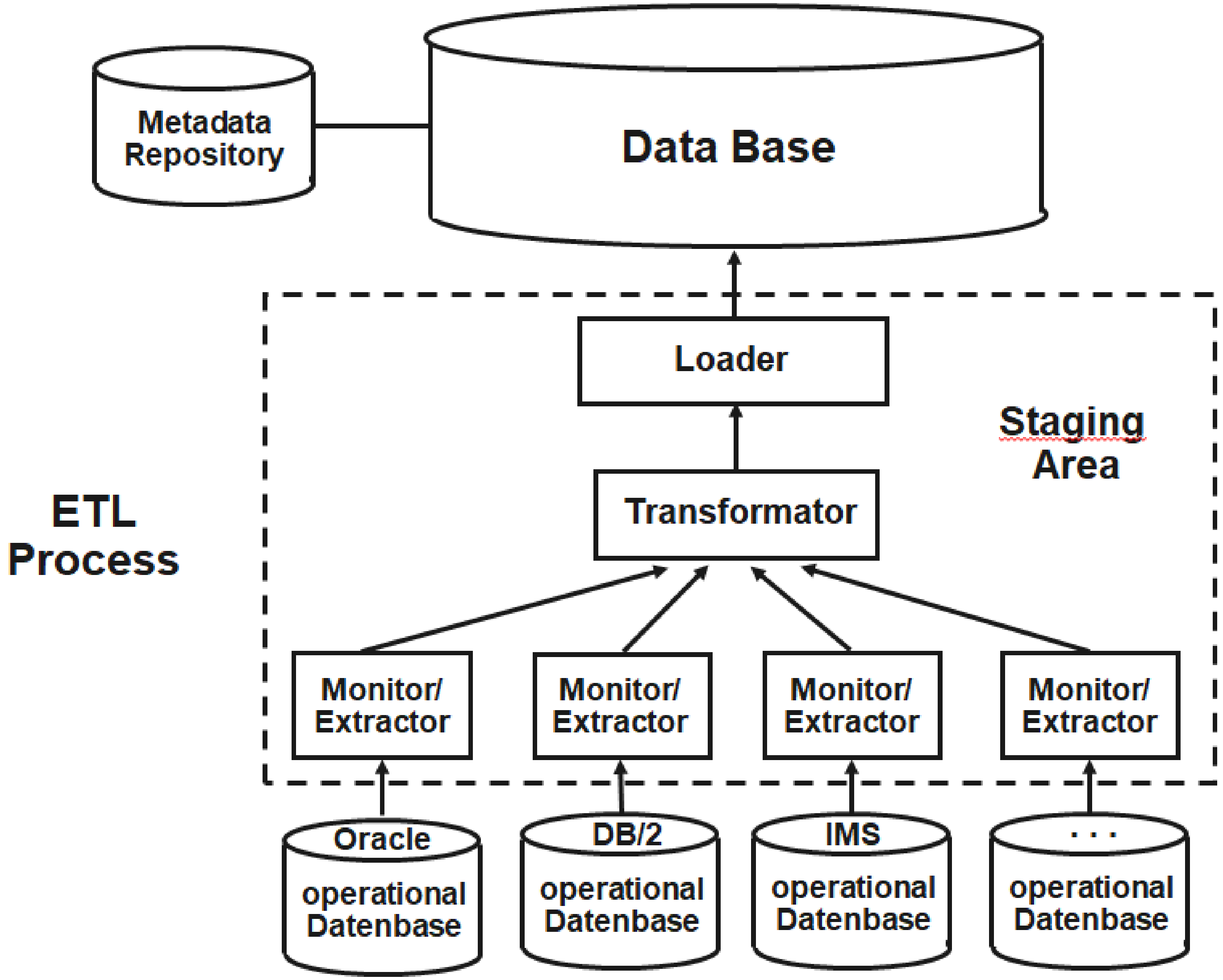
The Data Warehouse Concept



Layers of a Data Warehouse System



Components of a Data Warehouse

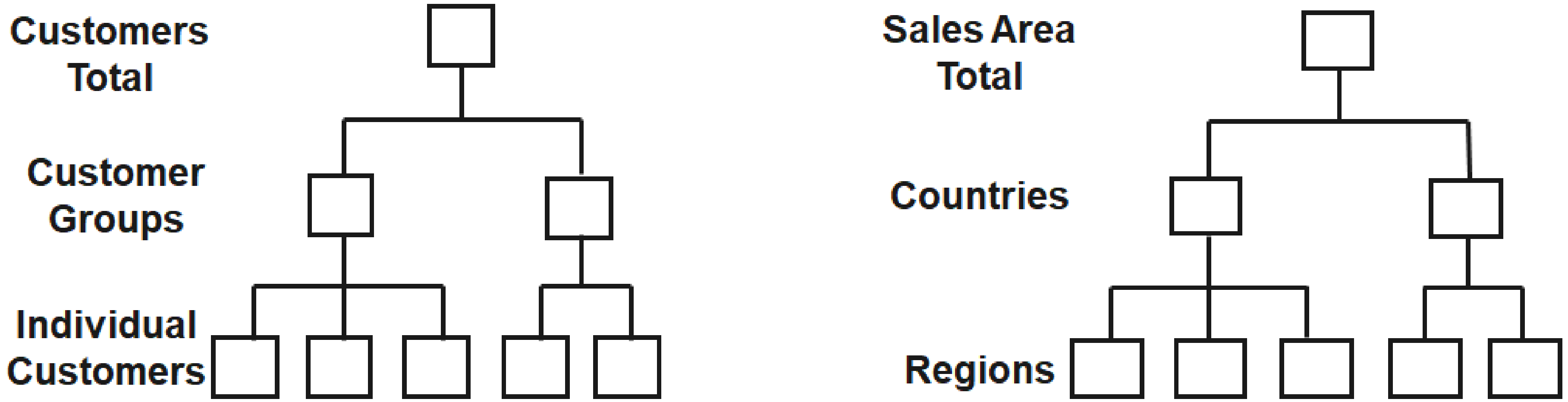


Structure of the Database

The database represents the core of the data warehouse. It contains current and historical data from all relevant areas of the company. In contrast to the operational databases, the data here is not structured according to the business processes and functions of the company, but is aligned with the facts that determine the company. They are organized in such a way that they can be viewed from different dimensions.

Frequently used dimensions are:

Examples of Aggregation Levels in a Data Warehouse



ETL-Tools

The data in the data warehouse can normally only be accessed in a read-only manner. The only exception to this are ETL tools, which are responsible for transferring data from internal and external information sources. They automate the process of data acquisition from these sources.

The timeliness with which the transfer to the database is performed can be defined flexibly, e.g., within defined time intervals, immediately after each change,

Staging Area

The staging area is a temporary cache that performs the ETL process. It includes the extraction, transformation and loading of the data into the data warehouse.

Neither the operational systems nor the data warehouse are affected by the process. The data can be transferred incrementally from the operational systems without having to

The Metadata Repository

Frankfurt School of Finance & Management

Metadata supports data management and serves as a descriptor for an object that holds some data or information. In data warehouses, it is collectively organized in a catalog called the metadata repository.

Metadata is classified into three types:

- **Technical metadata:** Provides information about the structure of data, where it resides, and other

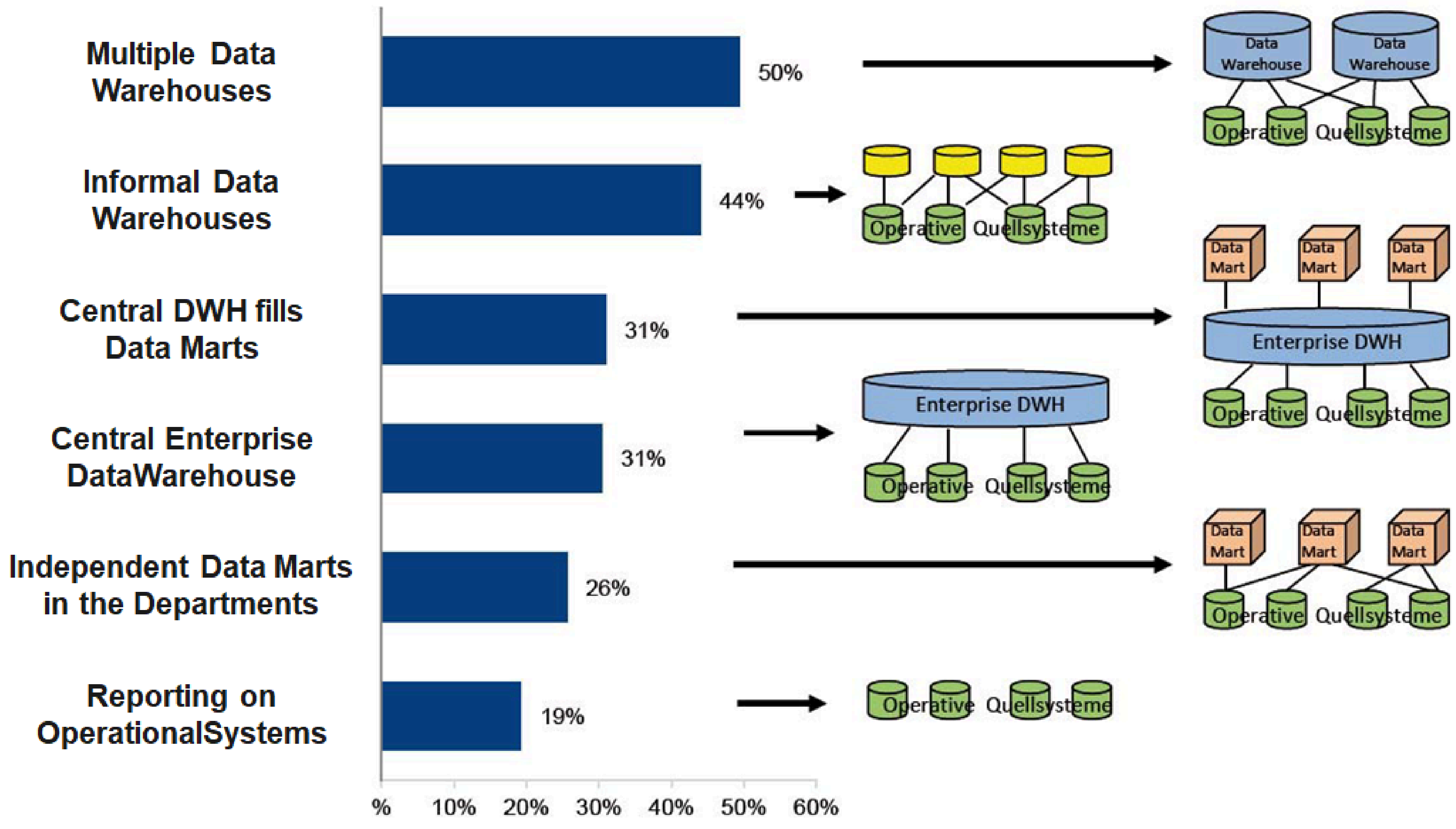
Data Warehouse and Data Mart

A single-subject data warehouse is typically referred to as a Data Mart, while Data Warehouses are generally enterprise in scope.

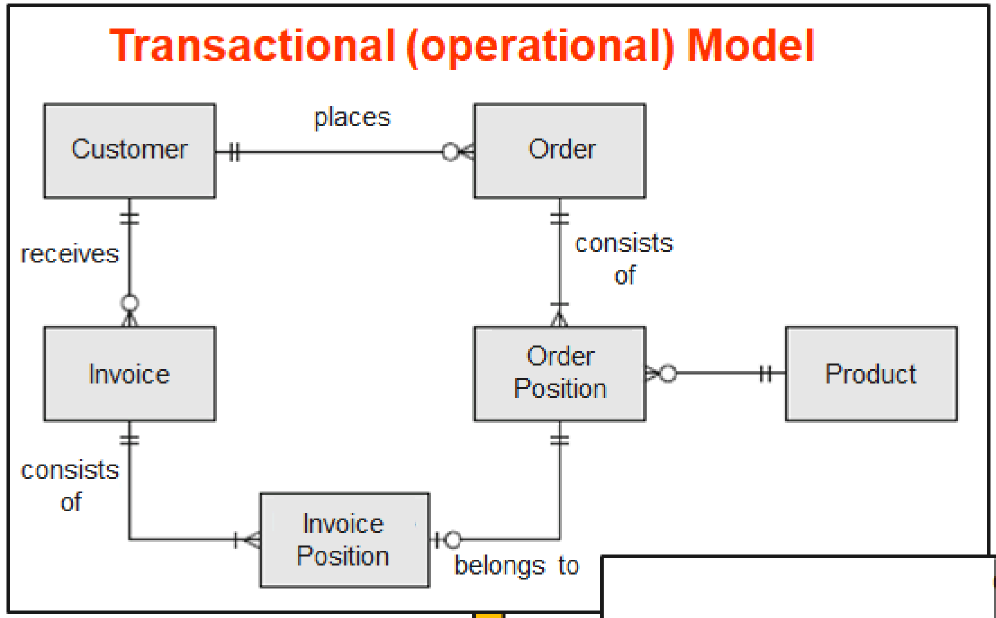
Comparison: Data Warehouse and Data Mart

	Data Mart	**Data Warehouse**
Application	Division/Department	Company
Data design	Optimized	Generalistic
Data volume	From gigabyte range	From terabyte range
Origin	Task oriented	Data model oriented
Requirements	Specific	Versatile

Architectural Variants of Data Warehouses in Practice

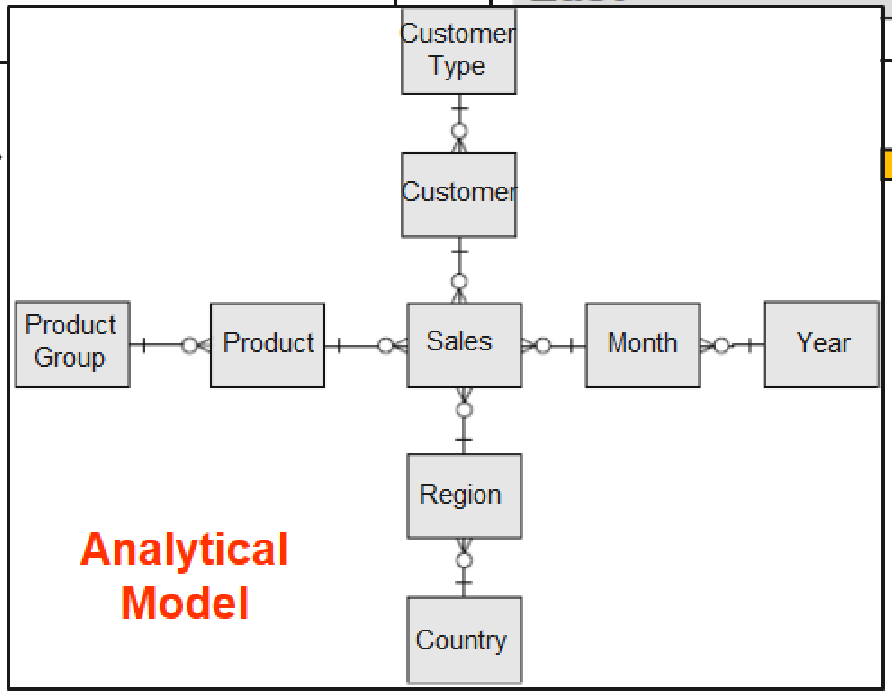


From Transactions to Reports



Report

Sales per Region and Month				
year:	2004			
Region/Month	Jan	Feb	Mar	...
North	460	526	556	...
South	234	334	364	...
Middle	56	100	130	...
West	120	170	200	...
East	89	130	160	...



Source: Müller/Lenz (2013), p. 16f

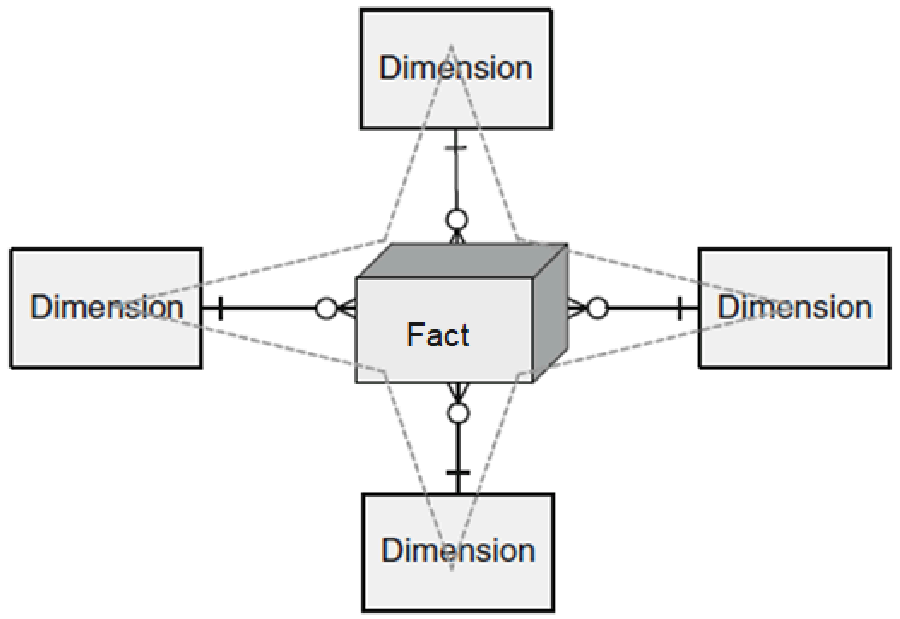
Example of a Dashboard

Source: https://www.bluemargin.com/power-bi-full-blog/topic/private-equity-data-intelligence

Star Schema (I)

Frankfurt School of Finance & Management

The Star schema has a central fact table and exactly one dimension table for each dimension.

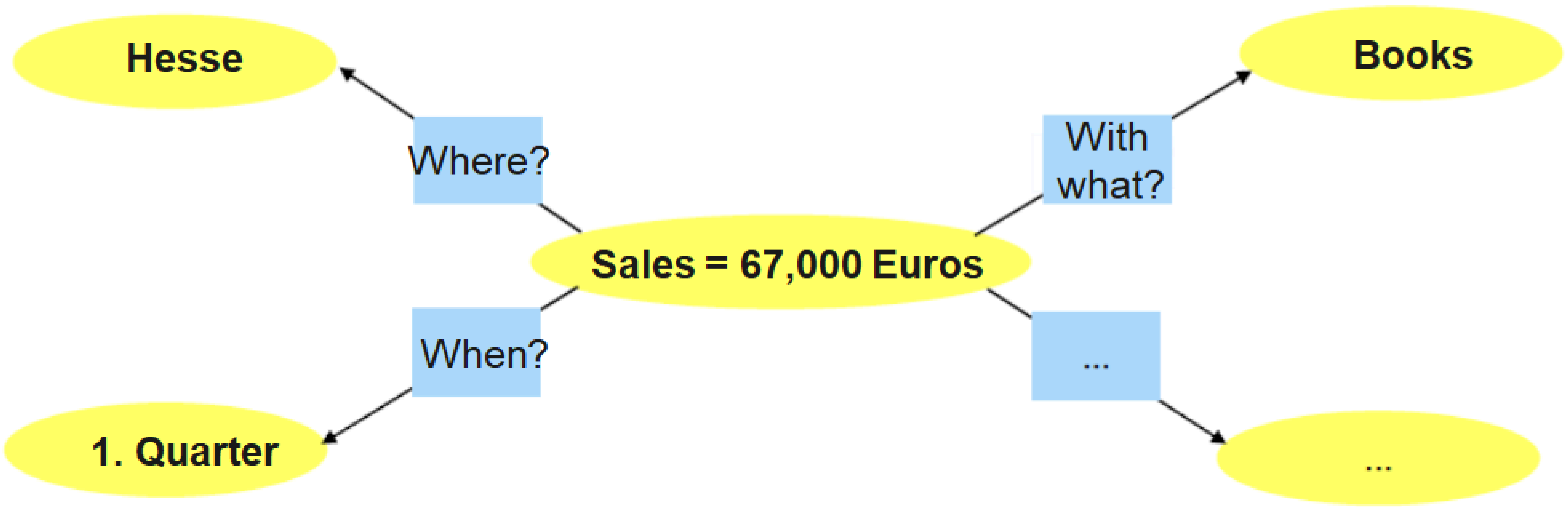


Source: Müller Lenz (2013), p. 60f

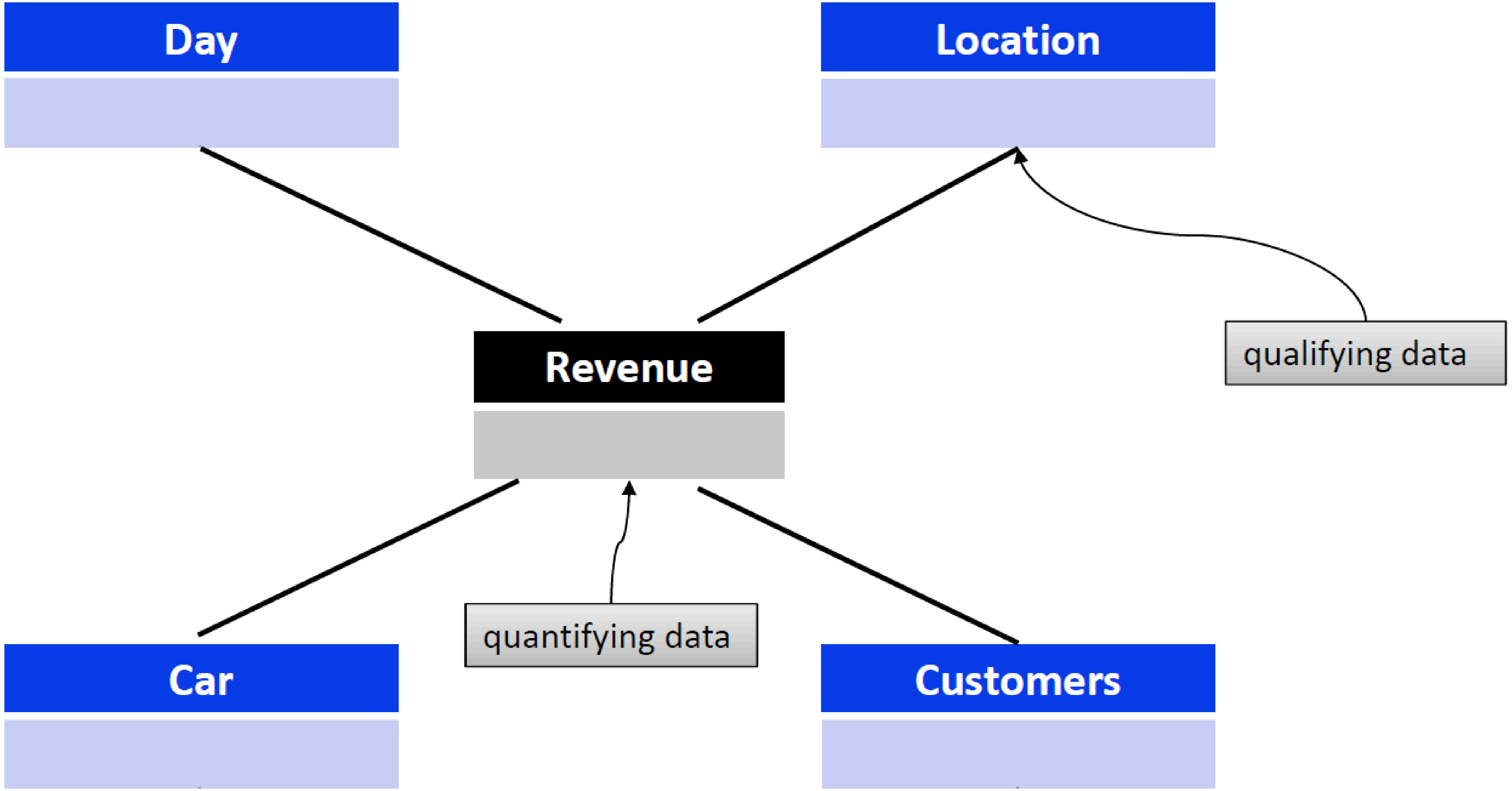
The cardinality between the fact table and a dimension table is n:1, i.e. a fact (e.g. sales: 67,000 euros) is described by one

Star Schema (II)

Frankfurt School of Finance & Management

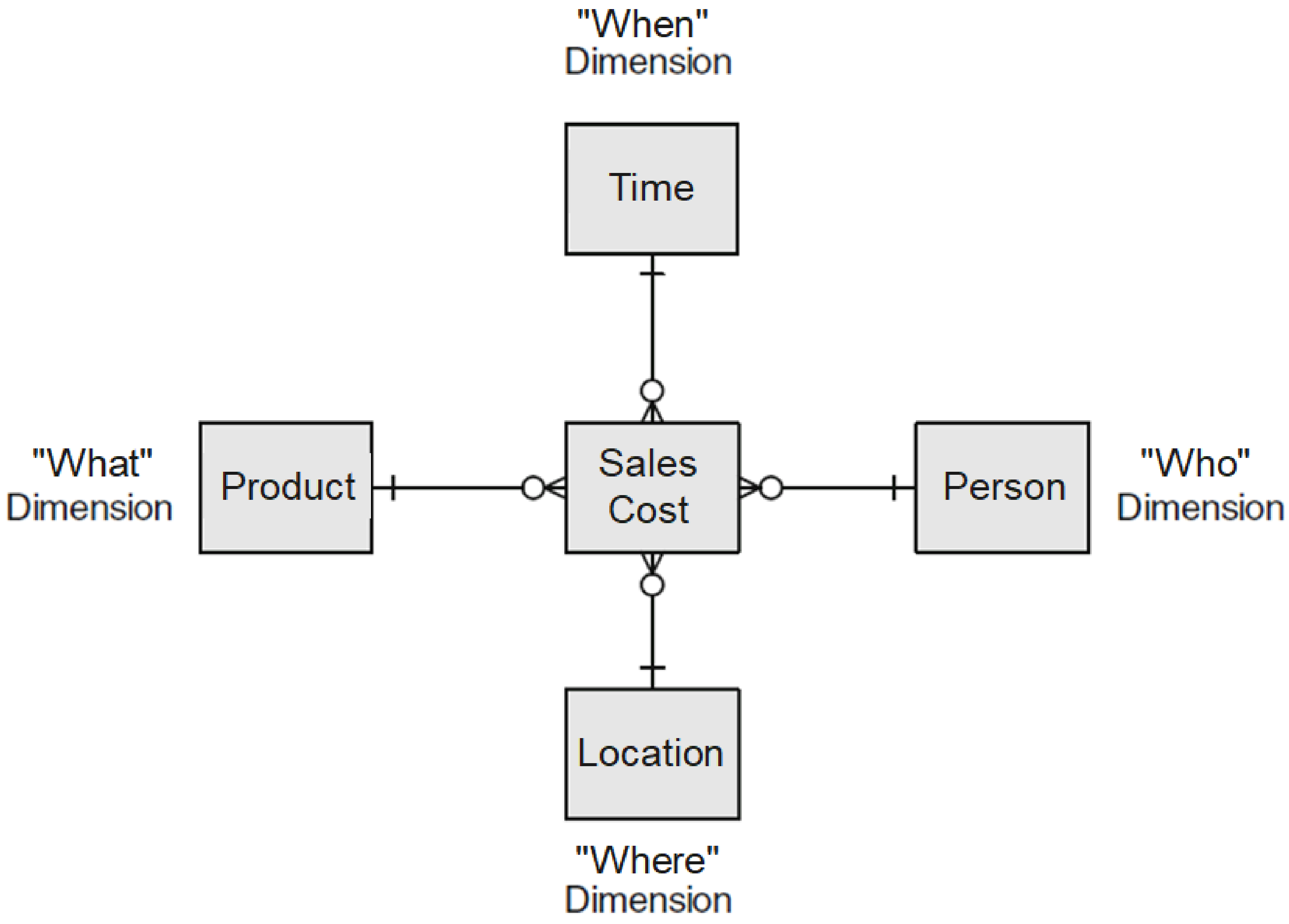


Star Schema (III)

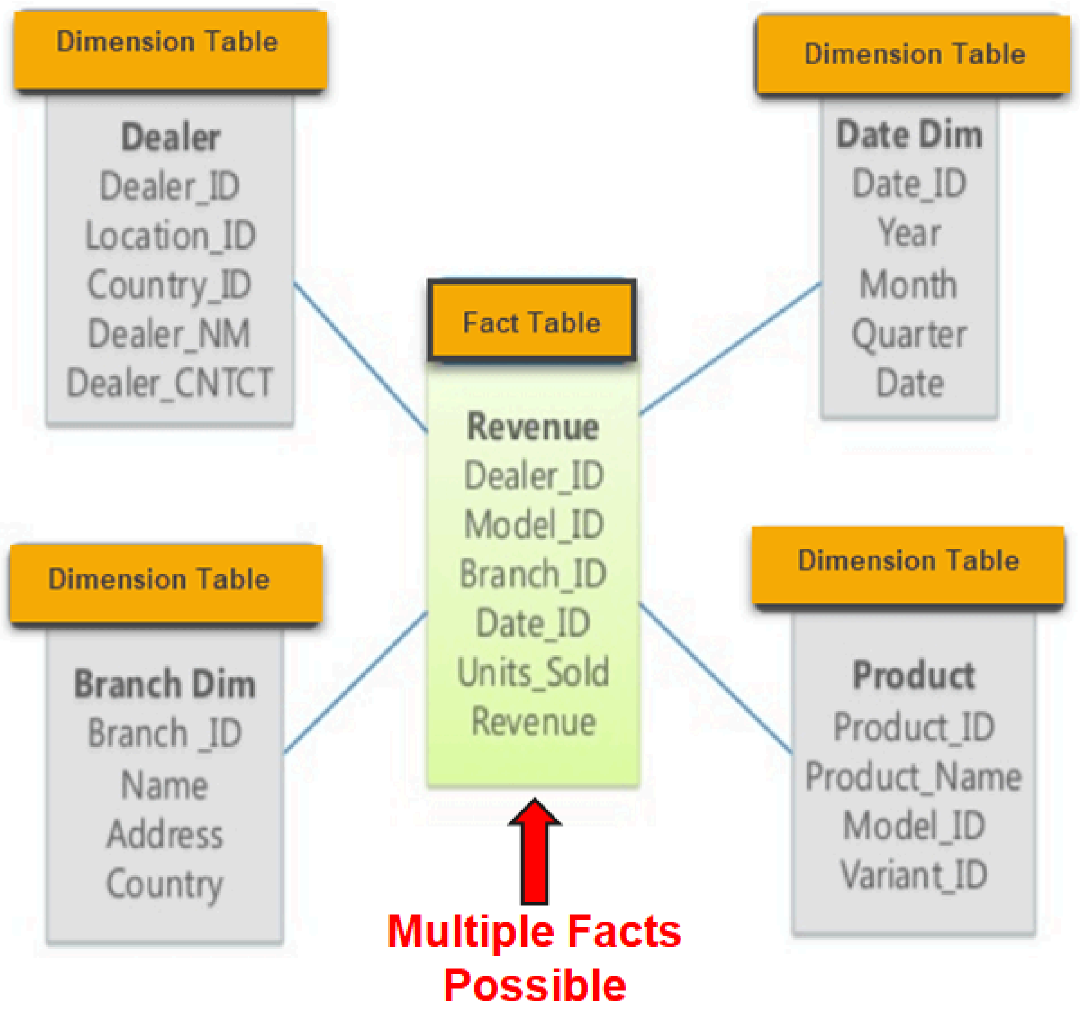


 # rentals

Star Schema Table Design (I)



Star Schema Table Design (II)



Source: https://www.guru99.com/star-snowflake-data-warehousing.html

Star Schema Example

Frankfurt School of Finance & Management

F (Sales)

Agency_ID	CustomerID	CarID	Revenue
HE_001	45678	BMW0017	330
HE_002	42973	BMW0008	260
HE_004	81115	VW1704	160
BW_001	81115	FORD2231	890

D_1 (Agency)

Agency_ID	Ag_City	State	Country
HE_001	Frankfurt	Hesse	BRD
HE_002	Fulda	Hesse	BRD
HE_004	Bensheim	Hesse	BRD
BW_001	Mannheim	Baden-W.	BRD

D_2 (Car)

CarID	C-Company	CarClass
BMW0017	BMW	Luxury
BMW0008	BMW	Family
VW1704	VW	Convertible
FORD2231	Ford	Eco-friendly

D_2 (Customer)

CustomerID	CustomerType	CustomerGroup
45678	Premium	Prefered
42973	Silver	Prefered
81115	Regular	Normal

Exercise 1

An online wine retailer plans to design a data warehouse to collect key figures regarding its wine sales. The relevant part of the operational database consists of the following tables:

- **CUSTOMER** (ID, name, address, telephone, birthday, gender)
- **WINE** (ID, name, type, year, bottle_price, class)
- **CLASS** (ID, name, region)

Queries on the Star Schema

Frankfurt School of Finance & Management

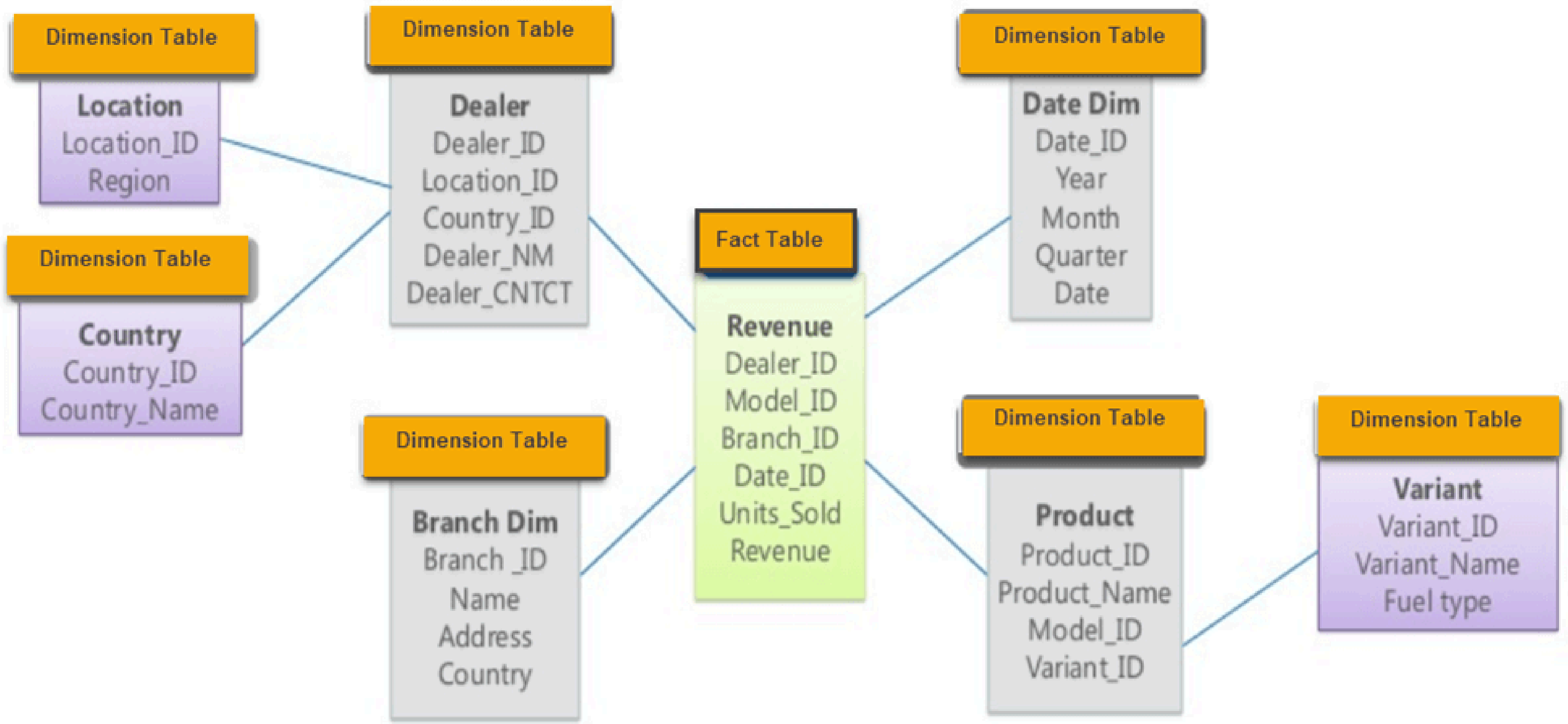
Example:

In which years were the most cars purchased by female customers in Hesse in the 1st quarter?

select z.Year as Year, sum(v.Amount) as Total_Amount
from Branch f, Product p, Time z, Customer k, Sales v
where z.Quarter = 1 and k.gender = ‘f’ and p.Product_type = ‘Car’ and f.state = ‘Hesse’ and v.Date = z.Date and v.ProductID = p.ProductID and v.Branch = f.Branchname and v.CustomerID = k.CustomerID

Snowflake Schema

In the Snowflake schema, the redundancies in the Star schema are resolved in the dimension tables and the hierarchy levels are each modeled with their own tables:



Star Schema vs. Snowflake Schema

– Speed of query processing is in favor of Star scheme

– Data volume tends to be lower with Snowflake schema

– Snowflake schema is easier to change (maintenance)

Galaxy Schema

Frankfurt School of Finance & Management

The Galaxy schema contains more than one fact table. The fact tables share some but not all dimensions. The Galaxy schema thus represents more than one data cube (multi-cubes)

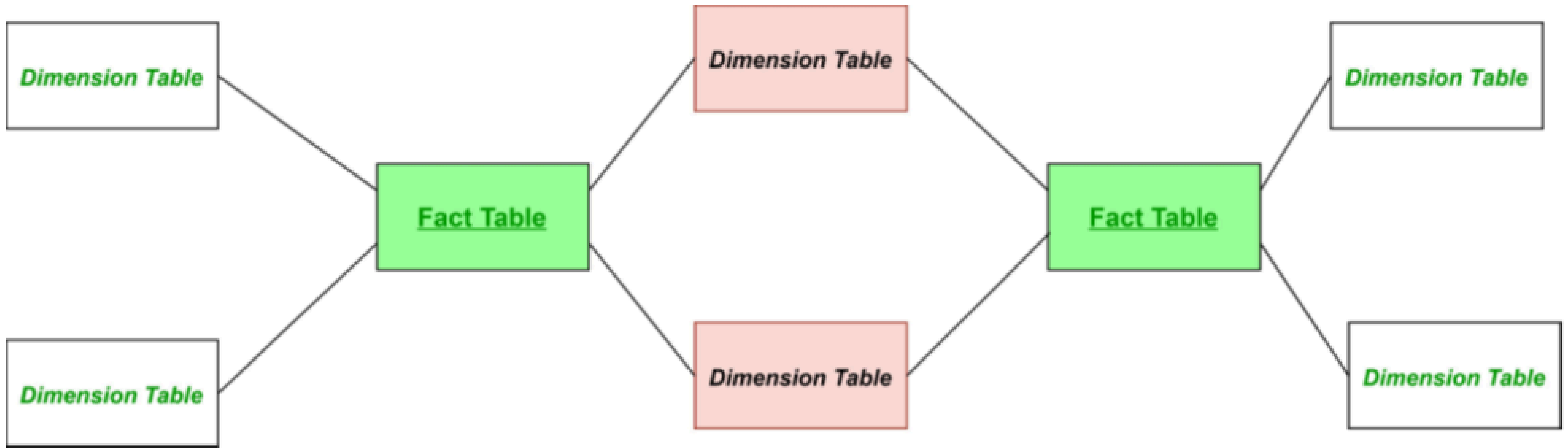
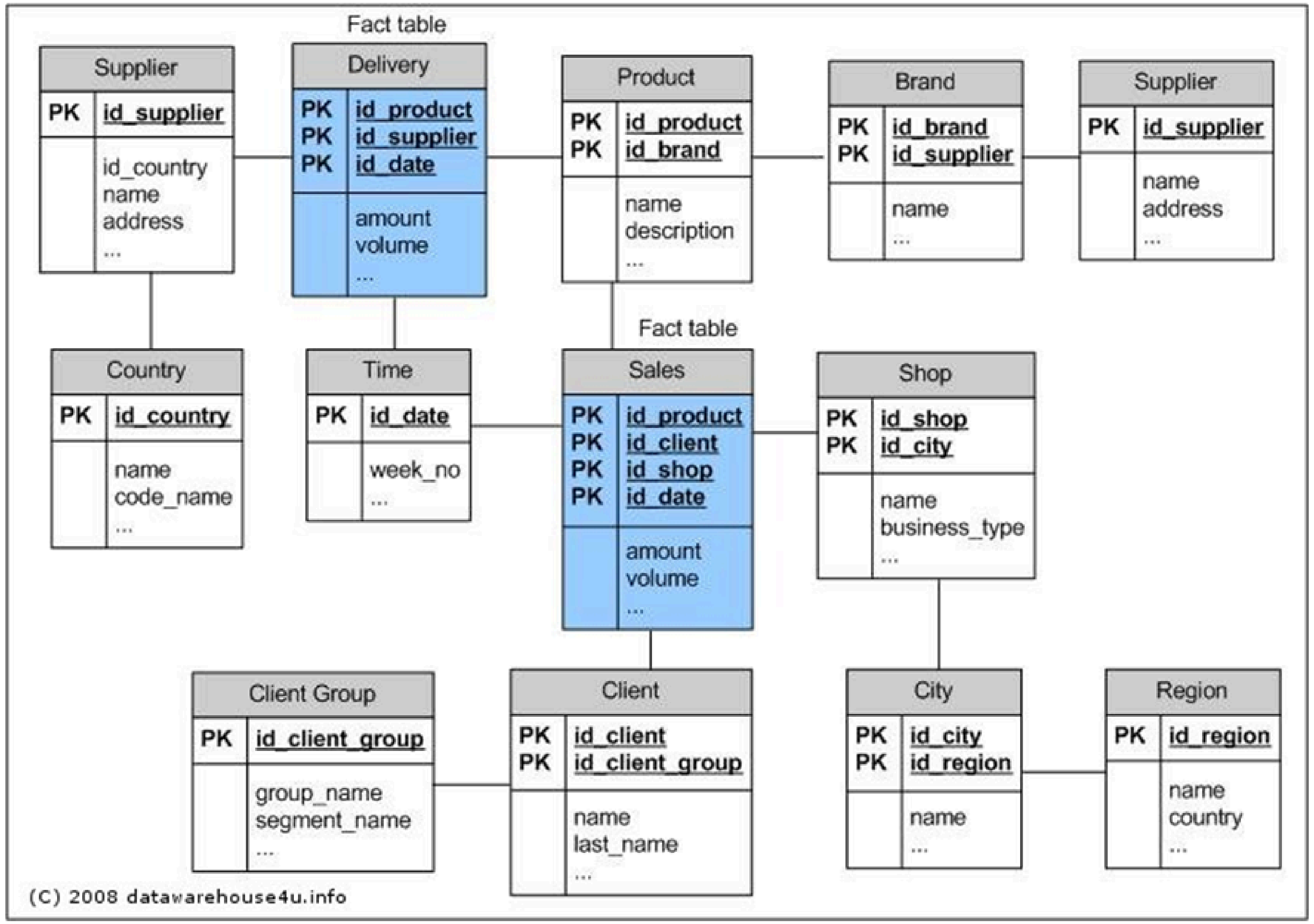


Table Structure of the Galaxy Schema



DW Planning – User Questions (I)

- What is the ratio of private to corporate customers?
- Does this quantity ratio correspond to the turnover ratio?
- Do customer characteristics such as occupational group, living situation, education, etc. play a role in turnover?
- What are the customers with the highest sales?
- What is the development of sales over time? Are there any shifts in relation to customer characteristics?
- Can additional customer segments be formed, e.g. if customer characteristics are combined?

DW Planning – User Questions (II)

Frankfurt School of Finance & Management

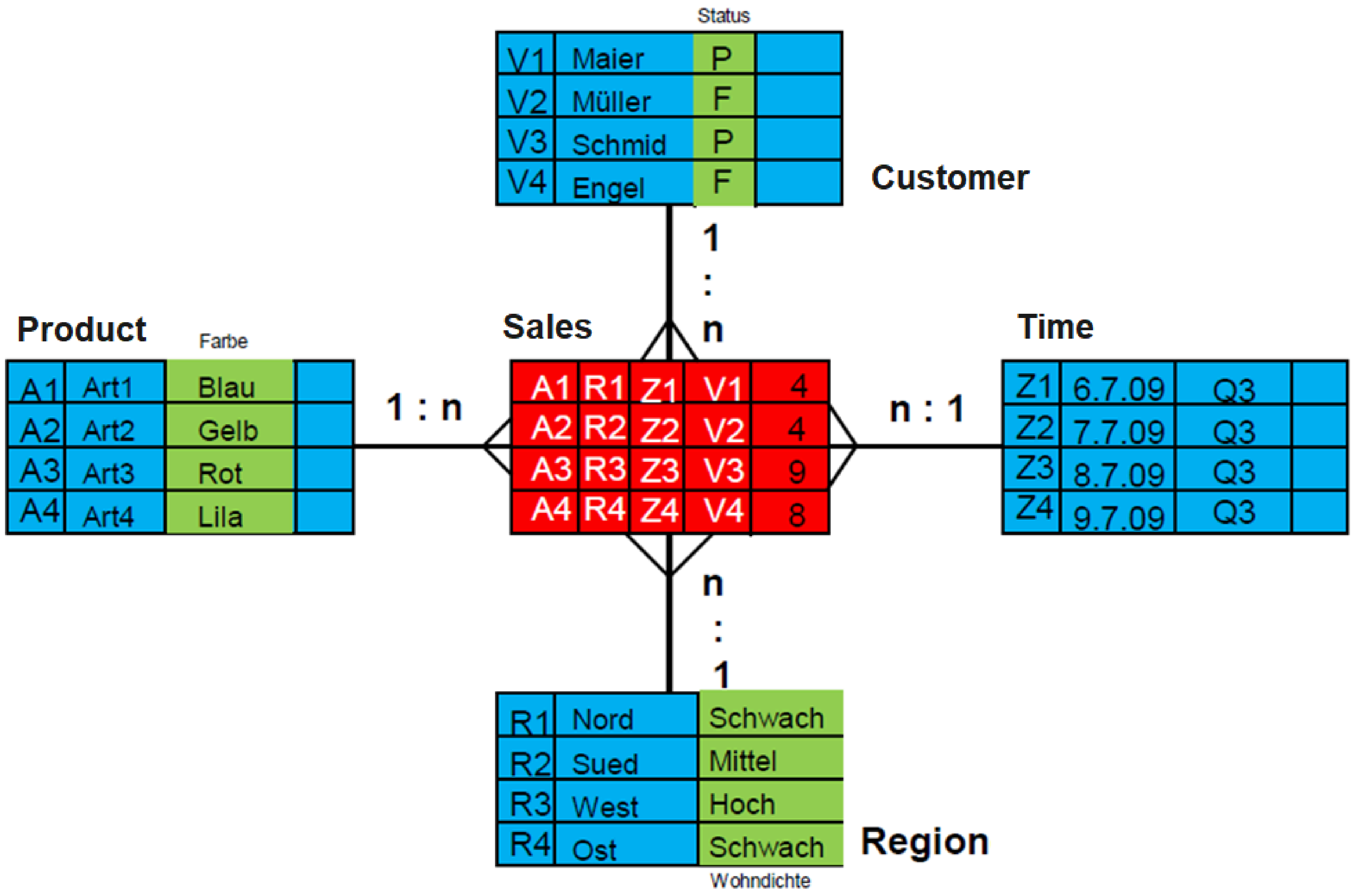
- What is the ratio of private to corporate customers?
- Does this quantity ratio correspond to the turnover ratio?
- Do customer characteristics such as occupational group, living situation, education, etc. play a role in turnover? What are the customers with the highest sales?
- What is the development of sales over time? Are there any shifts in relation to customer characteristics?
- Can additional customer segments be formed, e.g. if customer characteristics are combined?
- Are there regional focal points in relation to customer characteristics?
- Are there regional focuses related to products / product groups?
- Are there correlations between customer characteristics and products?
- What are the top-selling products? Top sellers/bottom sellers lists?
- Are there seasonal influences on the sales development depending on product or customer characteristics?

Transaction Data
Master Data
Properties Master Data

Master Data ➡ Candidates for Dimensions
Transaction Data ➡ Candidates for Facts

Source: Alfred Schlaucher, Oracle, p. 28

Result: Star Schema



Source: Alfred Schlaucher, Oracle, p. 40

Exercise 2

A restaurant chain wants to build a management information system. The application system architecture is to be aligned with the data warehouse concept. An OLAP system is chosen for report generation.

The company maintains various restaurants, which can be differentiated by region and by renovation status. The Leonardo-Campus restaurant belongs to the North region and is New from the renovation status. The restaurant Nordblick is also located in the North region, but its renovation status is Old. The Spätzleburg restaurant, on the other hand, is in the

Exercise 3

As an employee of the Controlling department, you will be given the task of setting up a data warehouse for sales and headcount analyses.

Flowers GmbH has two branches in Hesse, one in Rhineland-Palatinate and three in France. Product categories include cut flowers, garden flowers and bridal jewelry. Within garden flowers, the rose family includes the genera roses and wild roses. Products in the rose genus include thornless roses and roses with thorns. The company distinguishes between corporate and private

Subject of OLAP

Online Analytical Processing (OLAP) is software technology focused on the analysis of dynamic, multidimensional data.

The goal is to provide executives with the information they need quickly, flexibly and interactively.

OLAP functionality can be characterized as dynamic, multidimensional analysis of consolidated enterprise data, which includes the following components:

- Calculations of key figures over different dimensions and aggregation levels

Facts and Dimensions

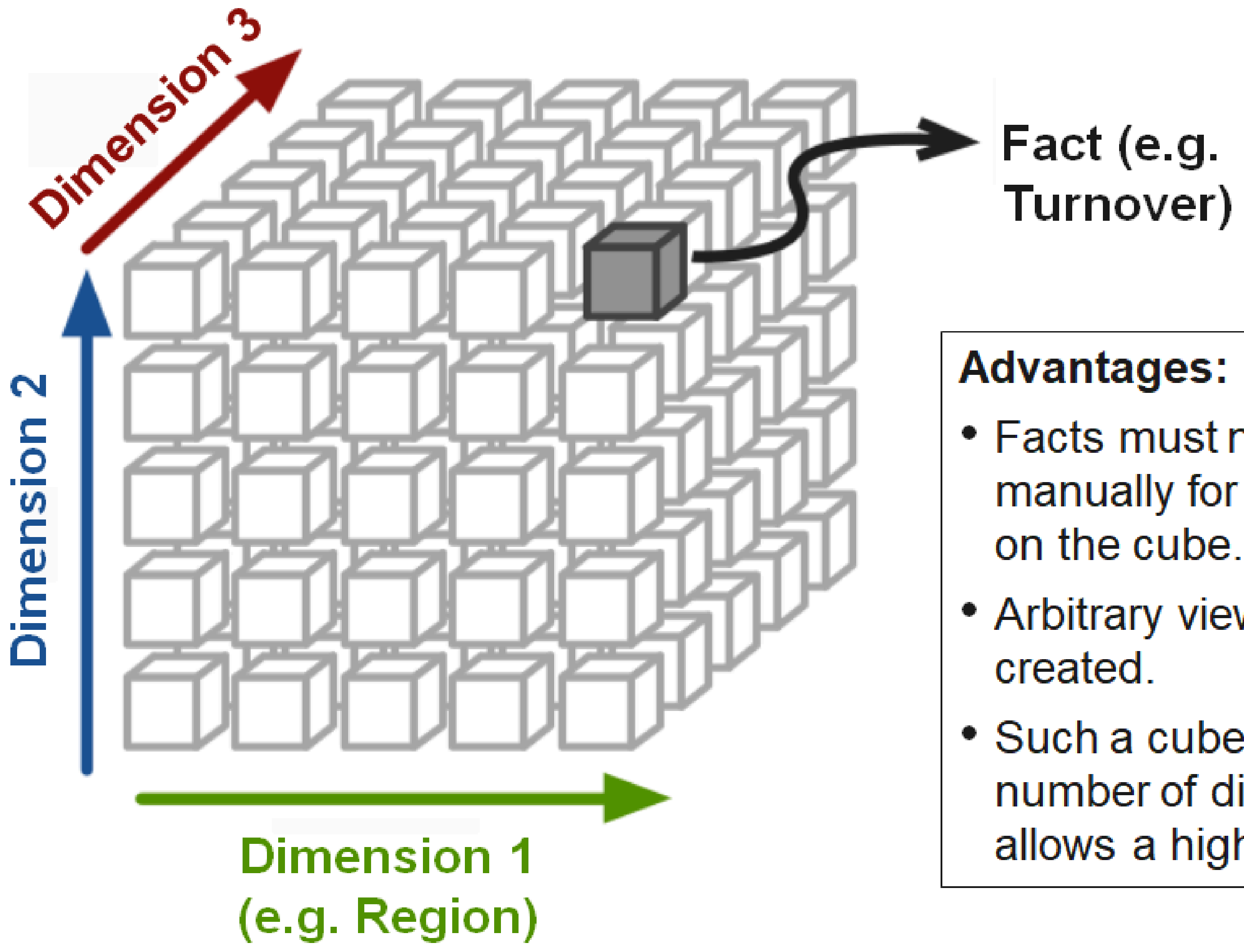
Turnover		Product Group									
		Accessories		Desktops		Laptops		Monitors		Printers	
Region	Year	Actual	Target	Actual	Target	Actual	Target	Actual	Target	Actual	Target
Austria	2019	1435	1200	3444	3500	3728	4100	1808	1780	2001	2000
	2020	3109	3145	3453	3920	7558	9000	1481	1610	1285	1320
Brazil	2019	2812	3000	7294	7350	7529	8200	3636	3550	3895	3880
	2020	6079	5991	6972	6720	14780	15500	2952	3150	2591	2640
France	2019	2910	3150	7153	6650	7425	8200	3549	3560	4013	4000
	2020	6373	6840	6594	6720	15320	16300	2823	2870	2541	2580
Germany	2019	4349	4500	10392	10500	11355	11890	5663	5630	6061	6090
	2020	9370	9064	10130	10080	23103	24600	4290	4270	3877	3780
Great Britain	2019	3619	3600	9034	9100	8900	8610	4440	4440	4895	5020
	2020	8218	7964	8432	8680	19480	18900	3536	3500	3227	3120
Italy	2019	2927	2850	6839	6650	7534	7790	3553	3550	3967	4000
	2020	6215	5991	6711	6720	15325	16300	2852	2800	2653	2700
Switzerland	2019	2977	3150	7061	6650	7551	8200	3683	3710	4066	4130
	2020	6163	5991	6589	6720	15162	14700	2837	2660	2556	2520
USA	2019	14651	14700	35613	35700	37595	37310	18391	18380	20154	20050
	2020	30900	30433	32913	32760	74830	73900	14137	14070	12991	13020

Qualifying Information (Dimensions)

Quantifying Information (Facts)

Organizing Data as Hypercube

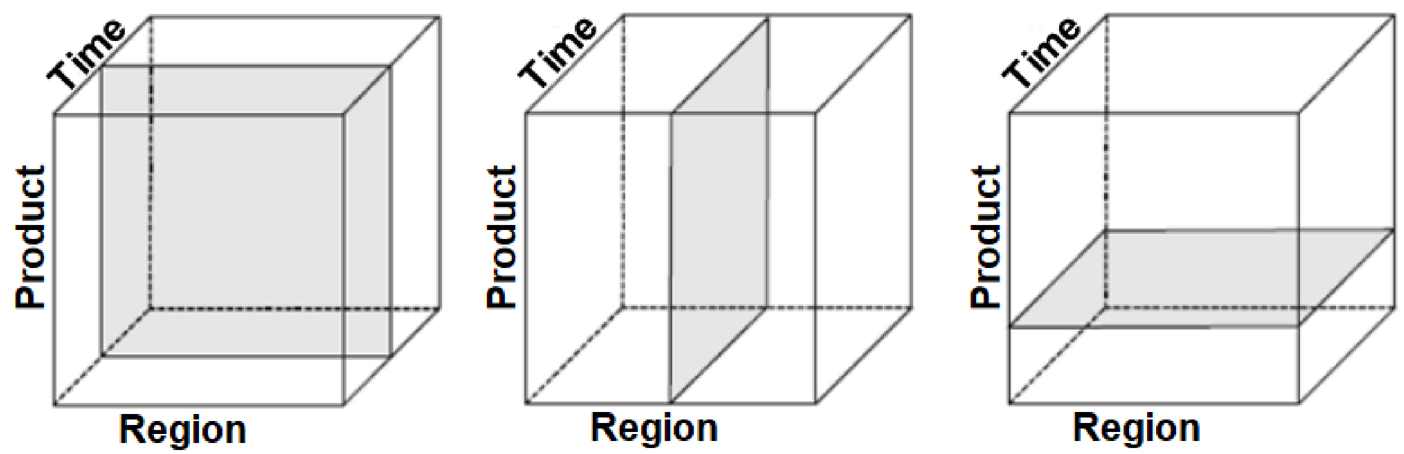
Frankfurt School of Finance & Management



Advantages:

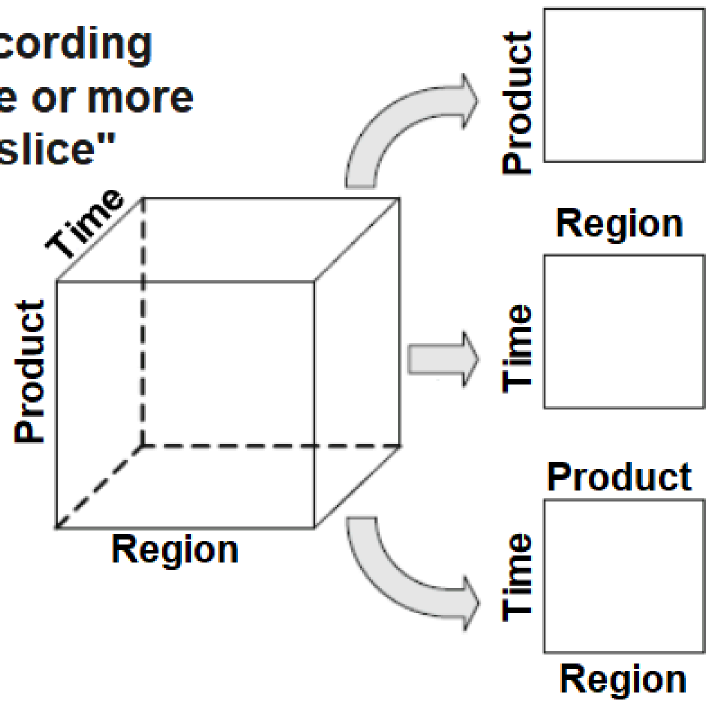
- Facts must not be calculated manually for any chosen view on the cube.
- Arbitrary views can be easily created.
- Such a cube can have any number of dimensions which allows a higher complexity.

OLAP Operations: Slicing and Dicing



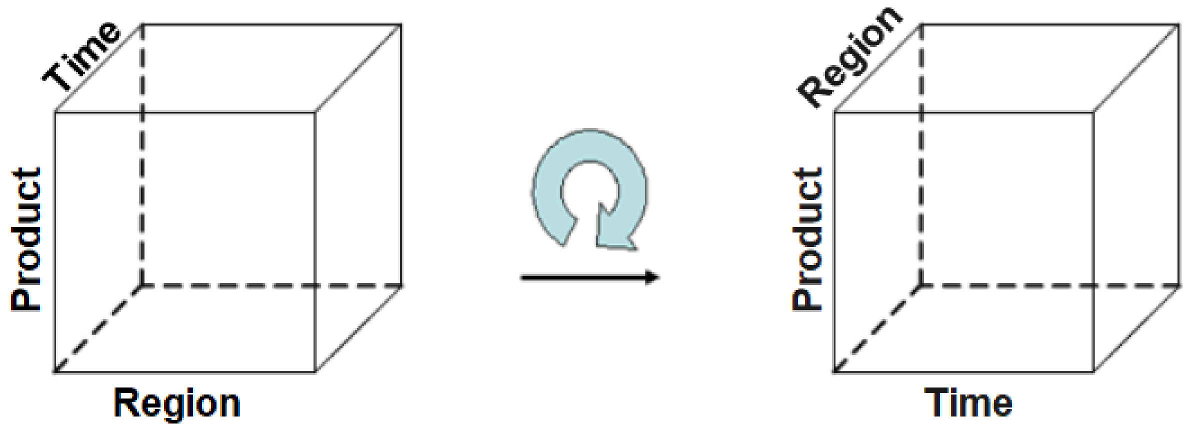
The **slicing** operator selects tuples from the cube according to the selection condition. The operator specifies one or more values for at least one dimension and thus slices a "slice" from the data cube.

The **dicing** operator selects a subset of the dimensions of the cube and uses it to cut out a smaller "sub-cube".



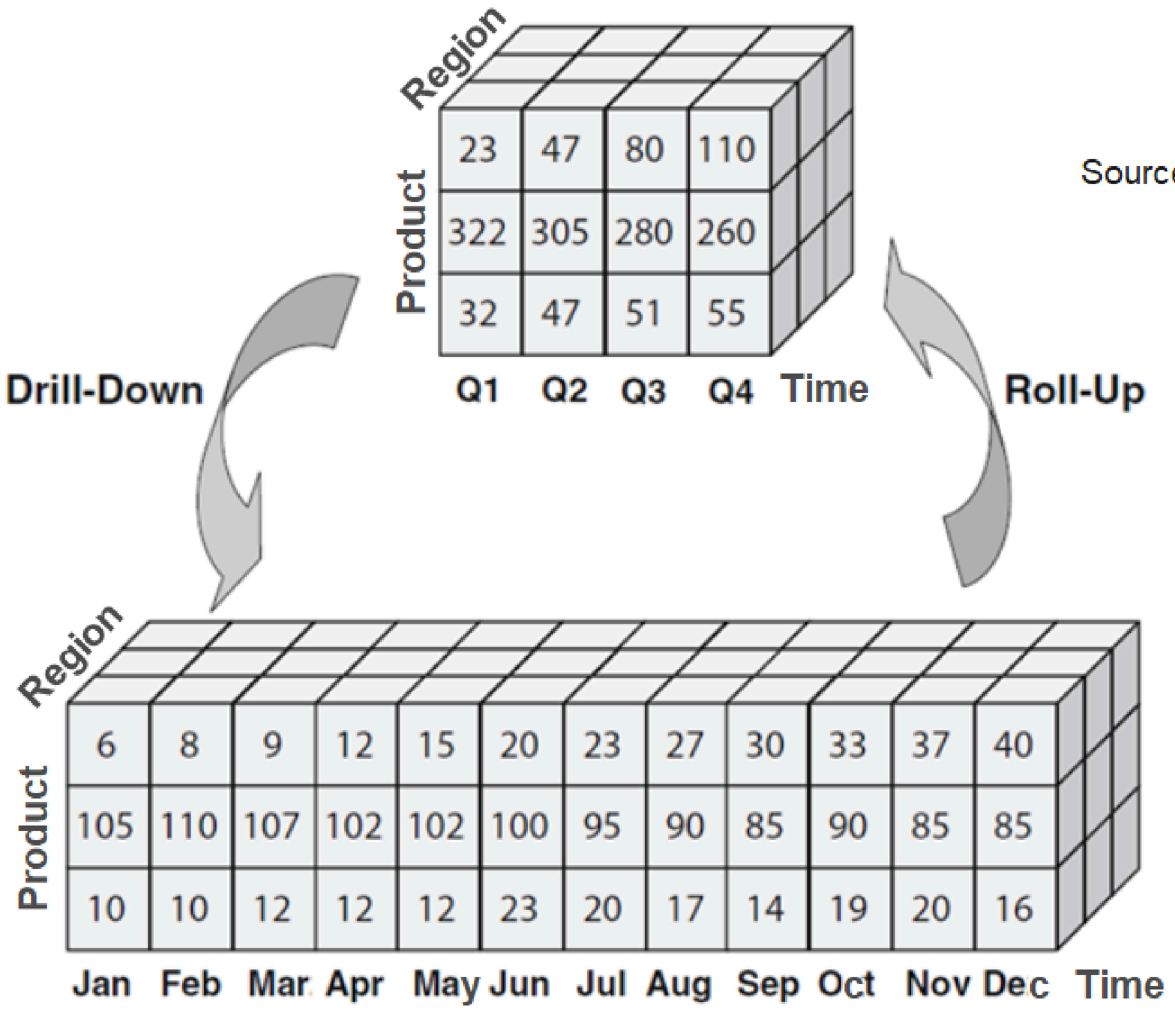
OLAP Operations: Rotate

The **rotate** operator rotates the hypercube around one of its axes:



OLAP Operations: Drill Down and Roll Up

The **roll-up** operator corresponds to aggregation and goes from the specific up to the general in the dimensional hierarchy; conversely, **drill-down** goes as disaggregation from the general down to the specific in the hierarchy.



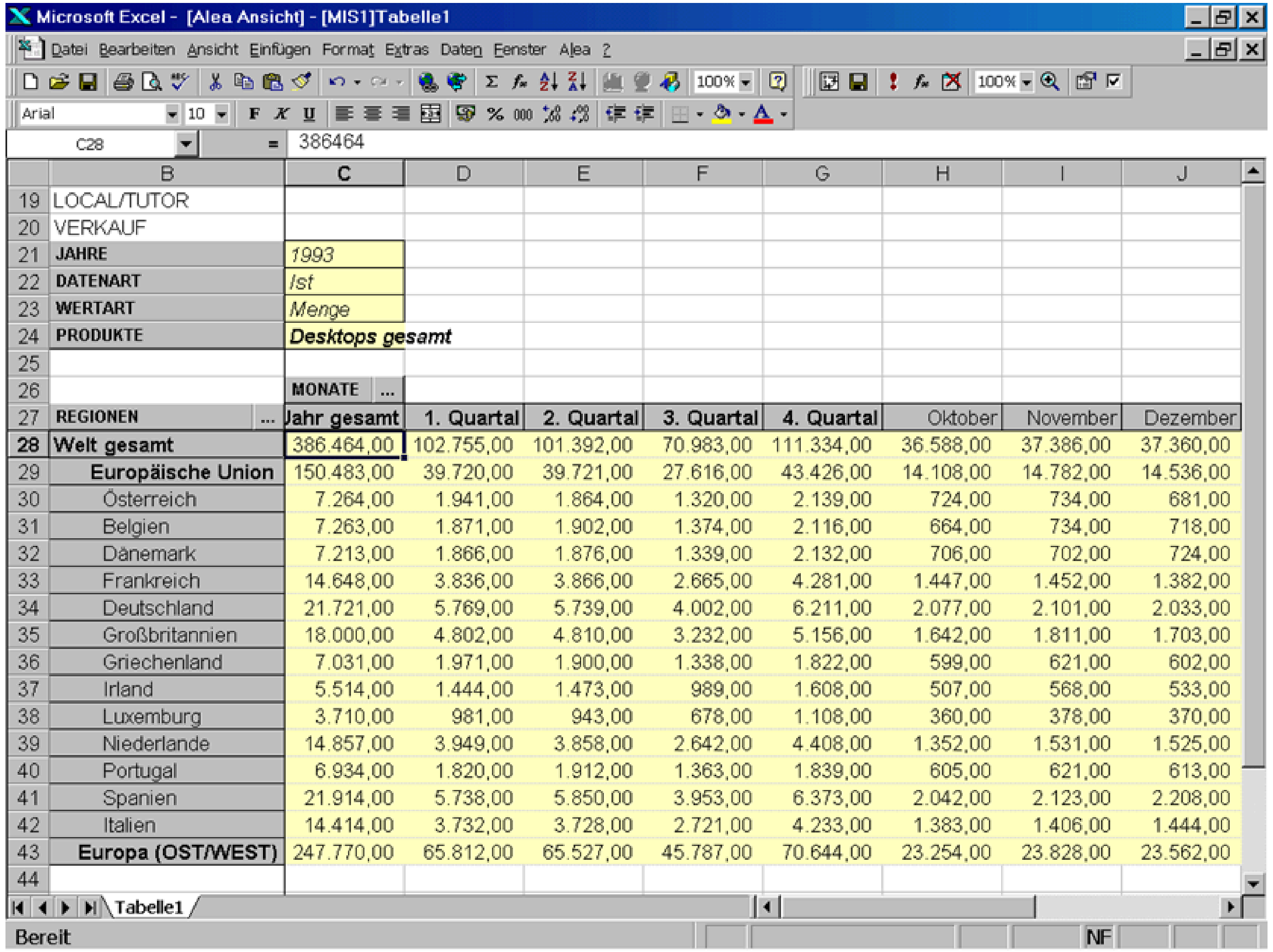
Source: Müller/Lenz (2013), p. 55

OLAP Operations: Nest

The result of a nest operation physically represents a two-dimensional matrix. It is extended by displaying different hierarchy levels of one or more dimensions on one axis (column or row) in nested form. In the following example, three dimensions are displayed:

Turnover		Product Group					
Region	Year	Accessories	Desktops	Laptops	Monitors	Printers	Total
Austria	2019	1435	3444	3728	1808	2001	12416
	2020	3109	3453	7558	1481	1285	16886
Brazil	2019	2812	7294	7529	3636	3895	25166
	2020	6079	6972	14780	2952	2591	33374
France	2019	2910	7153	7425	3549	4013	25050
	2020	6373	6594	15320	2823	2541	33651
Germany	2019	4349	10392	11355	5663	6061	37820
	2020	9370	10130	23103	4290	3877	50770
Great Britain	2019	3619	9034	8900	4440	4895	30888
	2020	8218	8432	19480	3536	3227	42893
Italy	2019	2927	6839	7534	3553	3967	24820
	2020	6215	6711	15325	2852	2653	33756

OLAP as Excel plug-in



Microsoft Excel - [Alea Ansicht] - [MIS1]Tabelle1

Datei Bearbeiten Ansicht Einfügen Format Extras Daten Fenster Alea ?

C28 = 386464

	B	C	D	E	F	G	H	I	J
19	LOCAL/TUTOR								
20	VERKAUF								
21	JAHRE	1993							
22	DATENART	Ist							
23	WERTART	Menge							
24	PRODUKTE	Desktops gesamt							
25									
26		MONATE ...							
27	REGIONEN ...	Jahr gesamt	1. Quartal	2. Quartal	3. Quartal	4. Quartal	Oktober	November	Dezember
28	Welt gesamt	386.464,00	102.755,00	101.392,00	70.983,00	111.334,00	36.588,00	37.386,00	37.360,00
29	Europäische Union	150.483,00	39.720,00	39.721,00	27.616,00	43.426,00	14.108,00	14.782,00	14.536,00
30	Österreich	7.264,00	1.941,00	1.864,00	1.320,00	2.139,00	724,00	734,00	681,00
31	Belgien	7.263,00	1.871,00	1.902,00	1.374,00	2.116,00	664,00	734,00	718,00
32	Dänemark	7.213,00	1.866,00	1.876,00	1.339,00	2.132,00	706,00	702,00	724,00
33	Frankreich	14.648,00	3.836,00	3.866,00	2.665,00	4.281,00	1.447,00	1.452,00	1.382,00
34	Deutschland	21.721,00	5.769,00	5.739,00	4.002,00	6.211,00	2.077,00	2.101,00	2.033,00
35	Großbritannien	18.000,00	4.802,00	4.810,00	3.232,00	5.156,00	1.642,00	1.811,00	1.703,00
36	Griechenland	7.031,00	1.971,00	1.900,00	1.338,00	1.822,00	599,00	621,00	602,00
37	Irland	5.514,00	1.444,00	1.473,00	989,00	1.608,00	507,00	568,00	533,00
38	Luxemburg	3.710,00	981,00	943,00	678,00	1.108,00	360,00	378,00	370,00
39	Niederlande	14.857,00	3.949,00	3.858,00	2.642,00	4.408,00	1.352,00	1.531,00	1.525,00
40	Portugal	6.934,00	1.820,00	1.912,00	1.363,00	1.839,00	605,00	621,00	613,00
41	Spanien	21.914,00	5.738,00	5.850,00	3.953,00	6.373,00	2.042,00	2.123,00	2.208,00
42	Italien	14.414,00	3.732,00	3.728,00	2.721,00	4.233,00	1.383,00	1.406,00	1.444,00
43	Europa (OST/WEST)	247.770,00	65.812,00	65.527,00	45.787,00	70.644,00	23.254,00	23.828,00	23.562,00
44									

Tabelle1

Bereit NF

Architecture of an OLAP System