

„Konsistente Datenanalyse durch Data Warehouse Systeme“

Hochschule für Technik und Wirtschaft Dresden

Fakultät Informatik/Mathematik

Prof. Dr.-Ing. Axel Toll

Dresden, den 27. April 2016

Weltweites Datenvolumen

...

2009: 0,8 Zettabyte

2010: 1,2 Zettabyte

2011: 1,8 Zettabyte

2012: 2,8 Zettabyte

2015: 8,6 Zettabyte

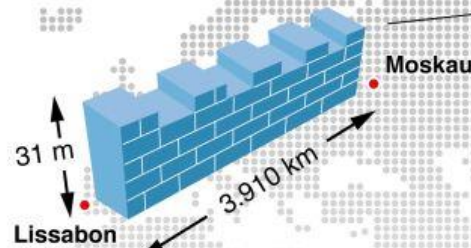
...

*Verdopplung
alle 2 Jahre*

1,8 ZettaByte
= 200 Milliarden HD-Filme
mit jeweils 2 h Länge
= 47 Millionen Jahre Filme

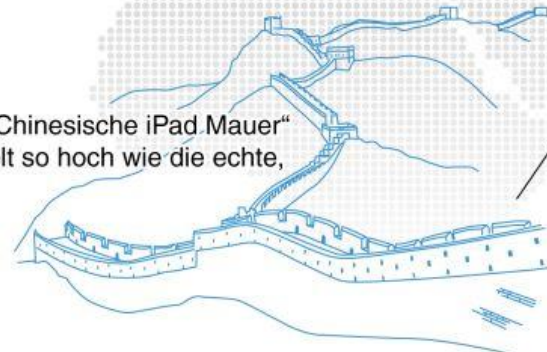
Wie groß sind 1,8 Zettabyte?

Um diese Datenmenge zu speichern, benötigt man 57,5 Mrd. Apple iPads.
Damit ließen sich imposante Bauwerke errichten:



oder

eine „Chinesische iPad Mauer“
doppelt so hoch wie die echte,



oder eine Fläche aus aneinandergelegten iPads so groß wie die Stadtflächen von München und Berlin zusammen.



Quelle: IDC Digital Universe Studie, EMC

1 Zettabyte = 10^{21} Byte = 1.000.000.000.000.000.000.000 Byte

[IDC-Studie "Extracting value from Chaos"]

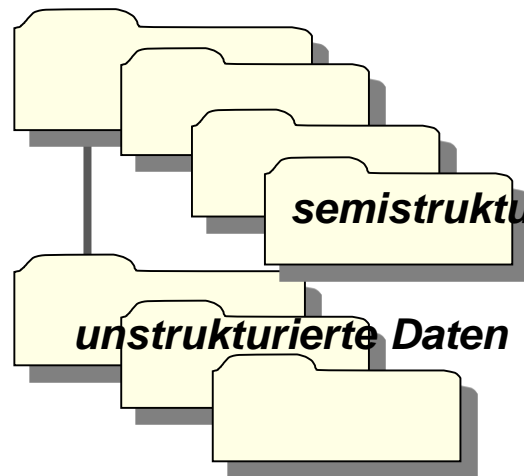
[<http://de.statista.com/statistik/daten/studie/267974/umfrage/prognose-zum-weltweit-generierten-datenvolumen/>: (21.04.2016)]

Datenkategorien und Organisationsformen der Speicherung

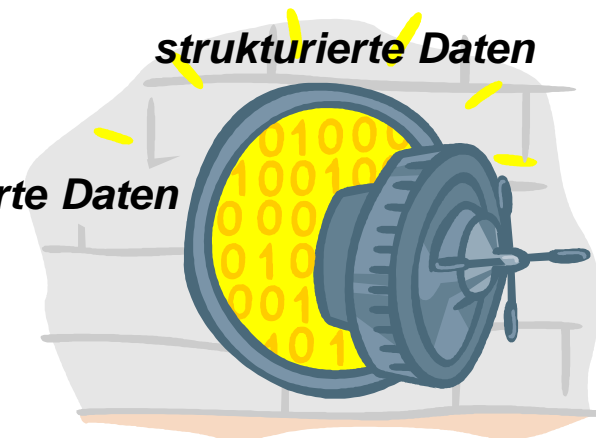
Datenkategorien:

- strukturierte Daten
(Tabellen mit Feldern, XML, ...)
- unstrukturierte Daten
(Texte, Videos, Bilder, ...)
- semistrukturierte Daten
(Dokumente, XML,...)

Dateisystem

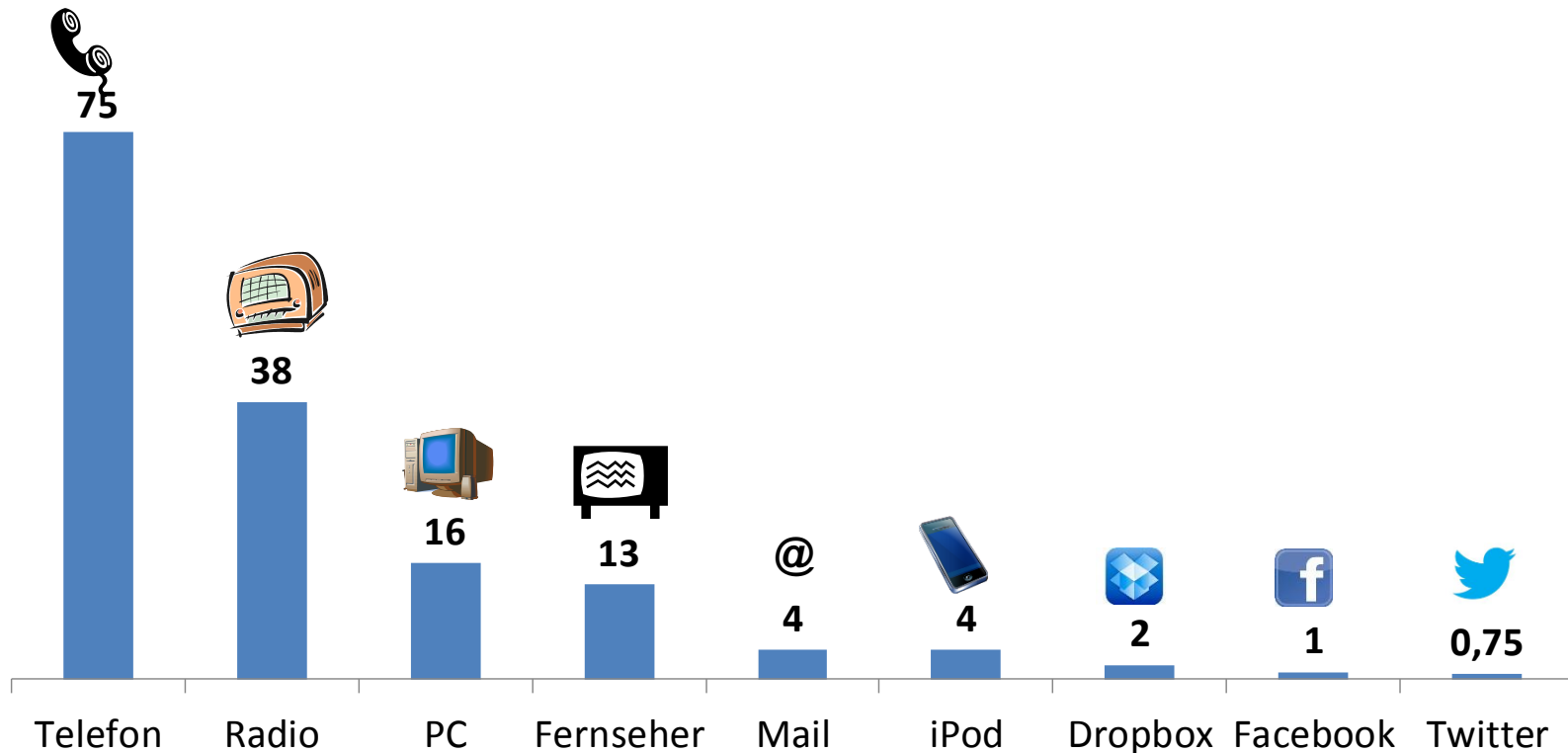


Datenbanksystem



Beschleunigtes Wachstum der Nutzerzahlen

Zeitdauer in Jahren bis zum Erreichen von 50 Millionen Nutzern:



Und die Anfragen aller Nutzer müssen vom System stets
und in akzeptabler Antwortzeit bearbeitet werden!

[Angeli, Kundler: Der Online Shop Handbuch für Existenzgründer]

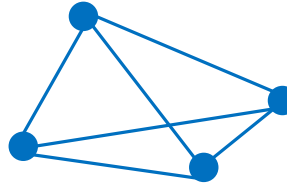
[<http://de.statista.com/themen/101/medien/infografik/501/adaption-von-technologien/>] (21.04.2016)

[<http://www.forbes.com/sites/victoriabarret/2011/10/18/dropbox-the-inside-story-of-techs-hottest-startup/>] (21.04.2016)

Verteilte Datenhaltungssysteme - ACID versus BASE

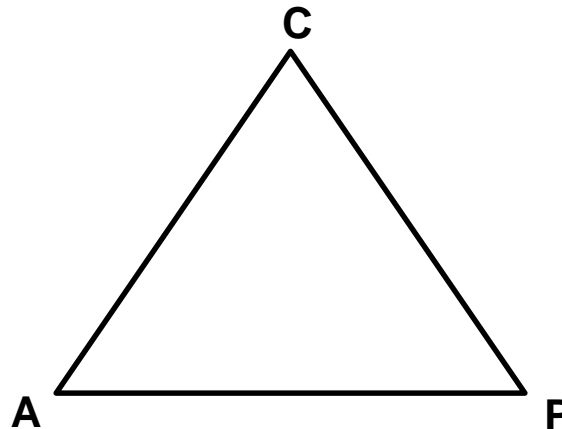
Verteilte Datenhaltungssysteme

Datenhaltung verteilt auf mehreren Knoten



CAP-Theorem

[Eric Brewer 2000]



C ... Consistency (Konsistenz)

Alle Knoten enthalten zur gleichen Zeit die gleichen Daten.

A ... Availability (Performance)

Alle Anfragen werden mit akzeptabler Antwortzeit geliefert.

P ... Partition Tolerance (Verfügbarkeit)

Das System arbeitet auch beim Ausfall einzelner Knoten weiter.

Klassisches Datenbanksystem:

ACID-Prinzip (**A**tomicity, **C**onsistency, **I**solation, **D**urability)

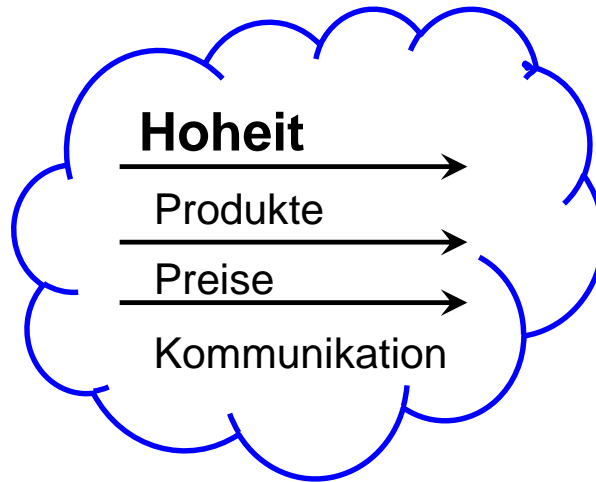
NoSQL-Datenbanken:

BASE-Prinzip (**B**asically **A**vailable, **S**oft State, **E**ventual Consistency)

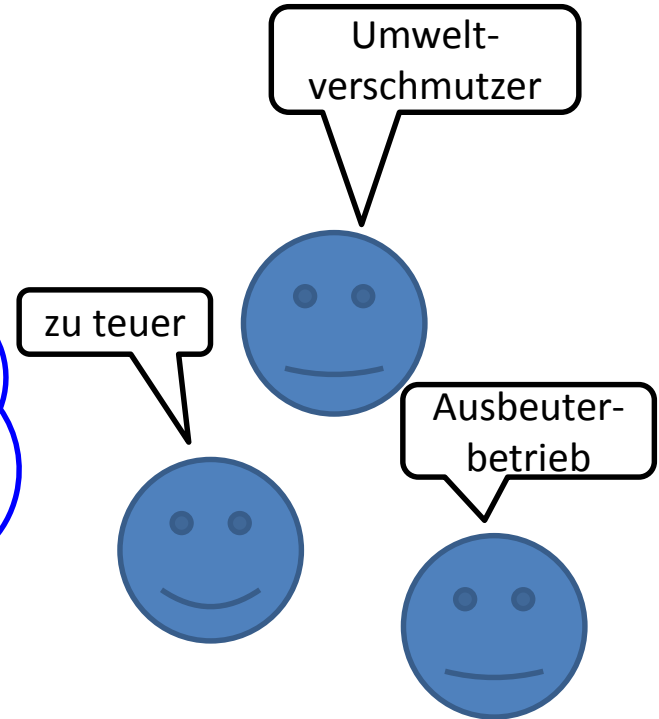
Eventual Consistency ... "verzögerte" Konsistenz

Verlagerung der Hoheit

Unternehmen

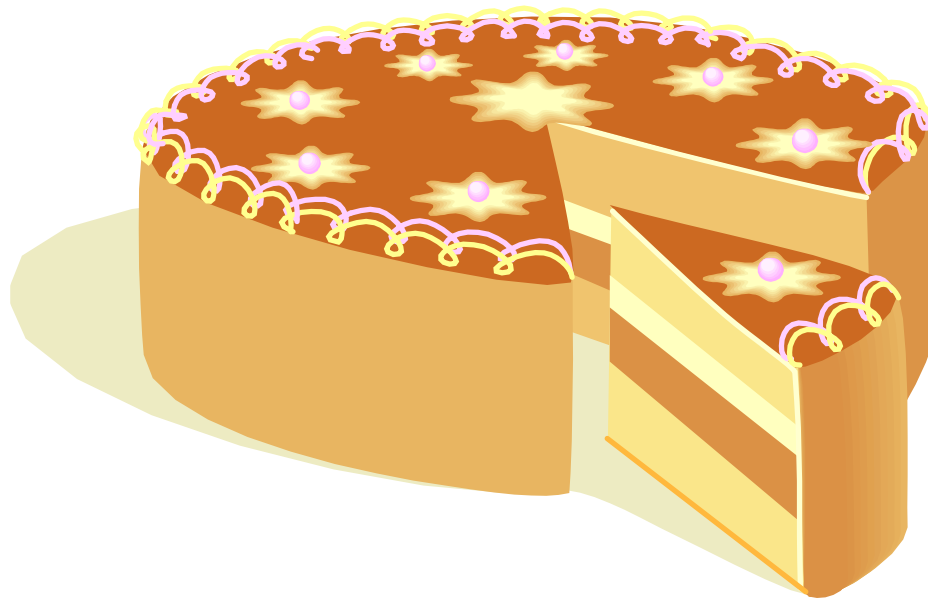


Kunden

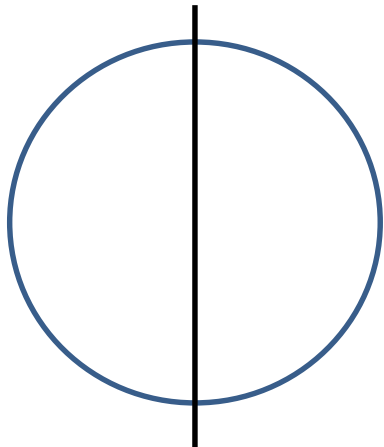


in Anlehnung an [Schönefeld: Praxisleitfaden Enterprise 2.0]

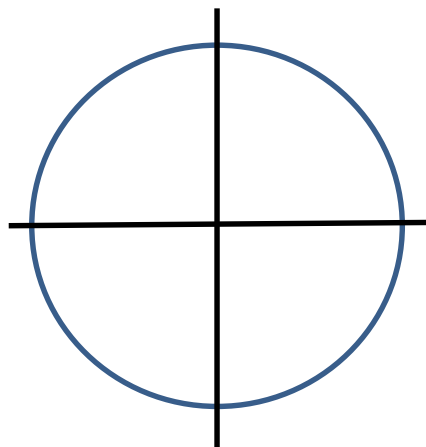
Teilungseffekt



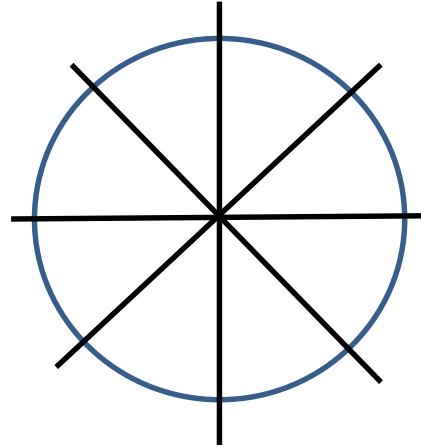
2 Nutzer



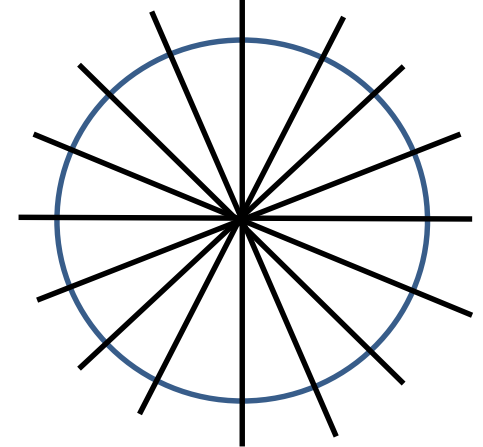
4 Nutzer



8 Nutzer

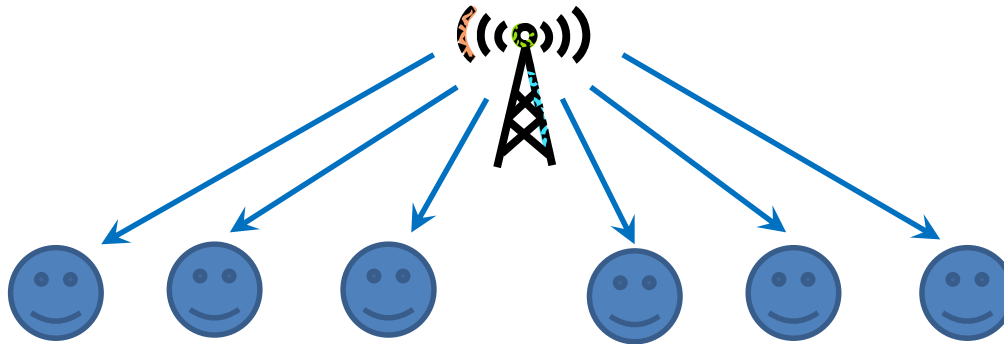


16 Nutzer



Netzwerkeffekte (1)

Fernsehen / Rundfunk



Wert V (Value) in Abhängigkeit der Teilnehmeranzahl N

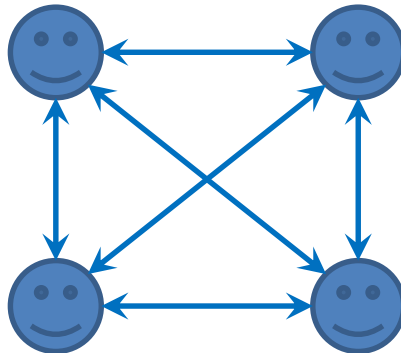
$$V \sim N \text{ [Sarnoff]}$$

$$V = N$$

$$V(1.000) = 1.000$$

$$V(1.001) = 1.001$$

Telefonnetz



obere Abschätzung: $V \sim N^2$ [Metcalfe]

$$V(N) = N(N - 1)/2$$

$$V(1.000) = 1.000 \cdot 999 / 2 = 499.500$$

$$V(1.001) = 1.001 \cdot 1.000 / 2 = 500.500$$

wirtschaftliche Abschätzung:

$$V \sim N \cdot \lg(N) \text{ [Odlyzko]}$$

$$V(N) = N \cdot \lg(N)$$

$$V(1.000) = 1.000 \cdot \lg(1.000) = 1.003$$

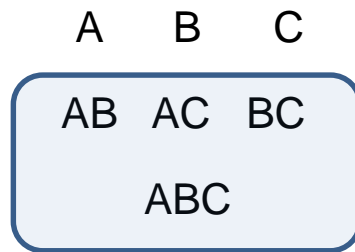
$$V(1.001) = 1.000 \cdot \lg(1.001) = 1.003,43$$

Netzwerkeffekte (2)

Netz mit Teilmengen

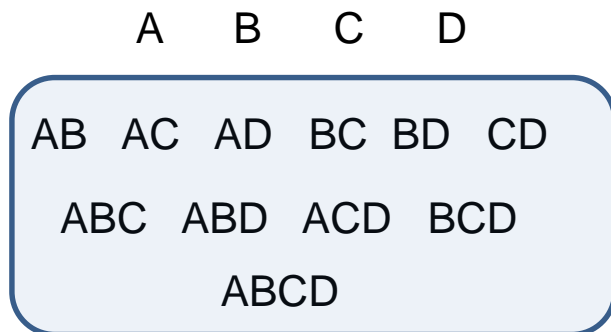
3 Teilnehmer

=> 4 potentielle Teilmengen



4 Teilnehmer

=> 11 potentielle Teilmengen



obere Abschätzung: $V \sim 2^N$ [Reed]

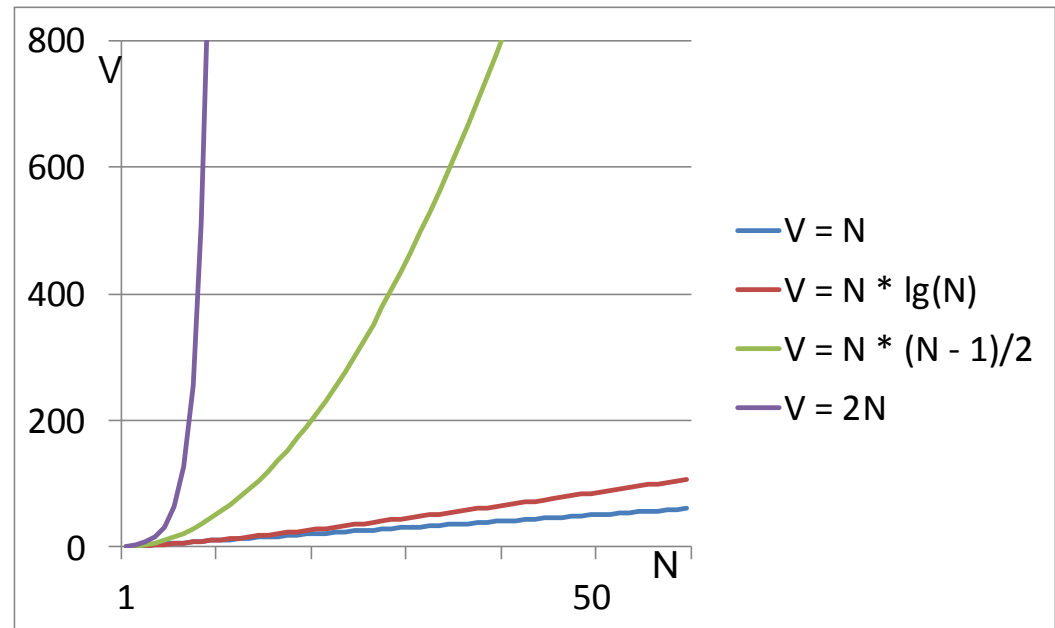
$$V(N) = 2^N - N - 1$$

$$V(3) = 8 - 3 - 1 = 4$$

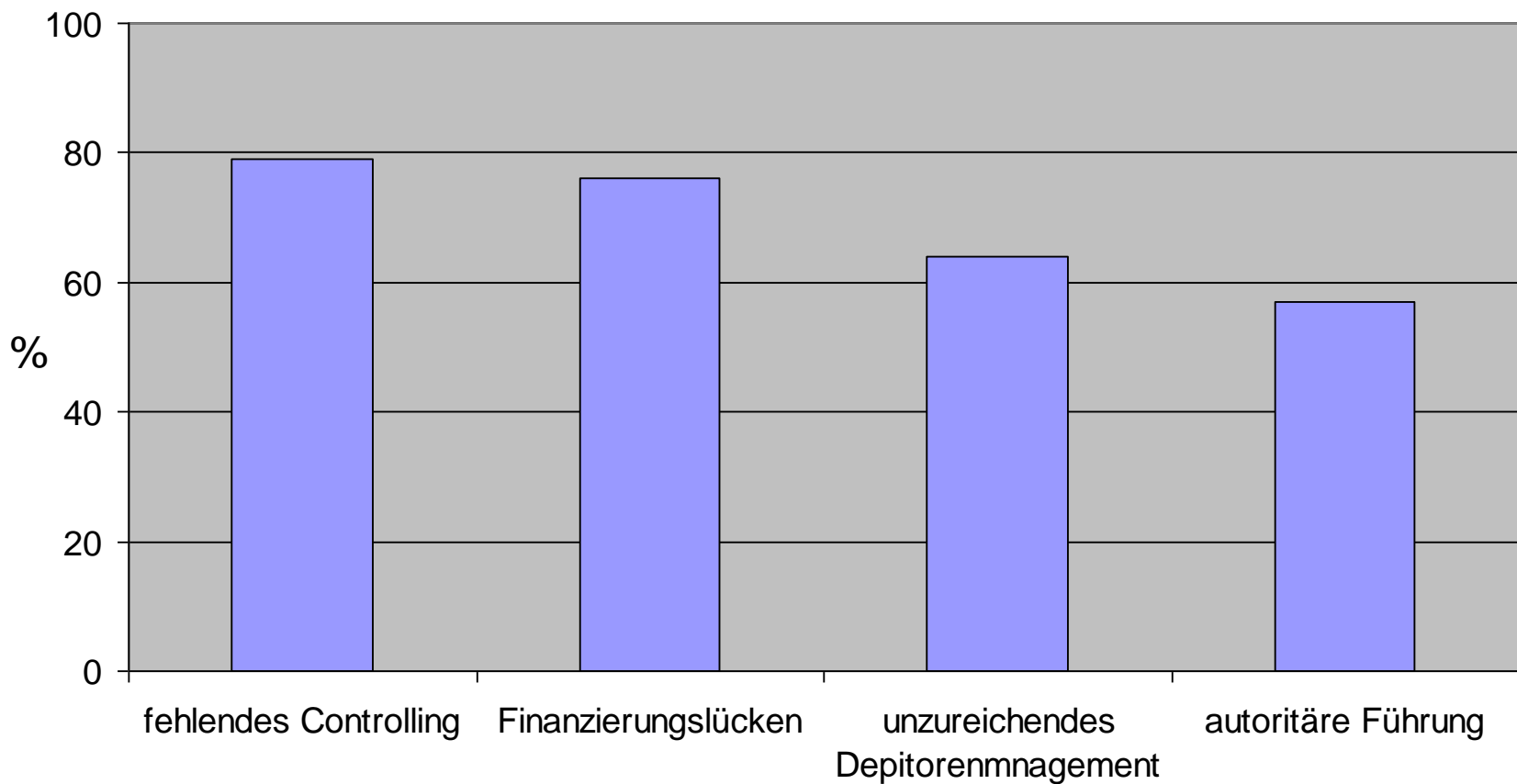
$$V(4) = 16 - 4 - 1 = 11$$

$$V(100) = 2^{100} - 100 - 1 = 1,07 \cdot 10^{30}$$

$$V(1000) = 2^{1000} - 1000 - 1 = 1,07 \cdot 10^{301}$$



Hauptgründe für das Scheitern von Unternehmen



[nach: Insolvenzen, Neugründungen, Löschungen Jahr 2007; Creditreform Wirtschaftsforschung]



u.a. ist BI die Basis für ein erfolgreiches Controlling

Ausgangslage - Problem und Fragen

Situation:

- technologischer Fortschritt
- internationale Verflechtung von Unternehmen
- Liberalisierung der Märkte



neuartige Problemstellungen, gekennzeichnet durch

- große Datenmengen
- hohe Vernetztheit
- steigende Dynamik
- Intransparenz

Problem:

- Die Beherrschung derartig komplexer Situationen stellt hohe Anforderungen an die Datenanalyse.

Fragen:

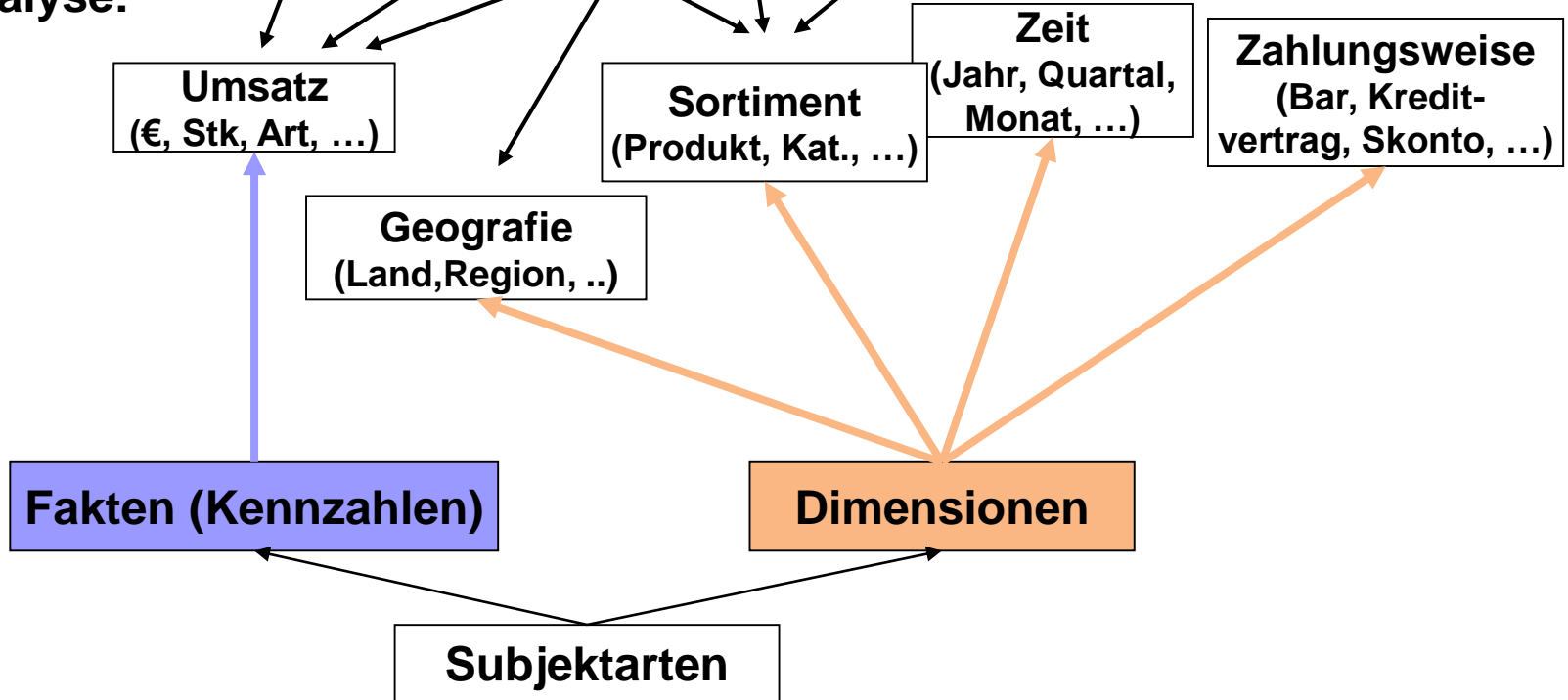
- Auf welcher Grundlage können Entscheidungen getroffen werden?
- Kann Informatik (im weitesten Sinne) hilfreich sein?
- Existieren systematische Ansätze, die eine „automatisierte“ Bereitstellung von Basisdaten für die Analyse ermöglichen?
- Was kann man vom Computer erwarten?
- ...

Subjektorientierung der Analyse

Unternehmens-
funktionen:



Subjekte
der Analyse:



Beispiel der Kennzahlendefinition über Balanced Scorecard

Finanzwirtschaftliche Perspektive

Ziele	Leistungsmaßstäbe
Erfolg	Umsatzrentabilität, ROI
Ertragswachstum	Umsatzwachstum/Zielsegment
Kostenreduzierung	% indirekte Kosten am Gesamt
Liquidität	Cash to Cash - Cycle

Kundenperspektive

Ziele	Leistungsmaßstäbe
Kundenrentabilität	Nettoerfolg je Kunde
Marktanteil	% Umsatz im Zielsegment
Kundenaquisition	% Umsatz mit Neukunden
Kundentreue	Umsatzsteigerung Altkunden
Kundenzufriedenheit	Zufriedenheitsindex aus Umfrage
Qualität, Zeit	Fehlerquote, Reparaturzeit

Interne Prozessperspektive

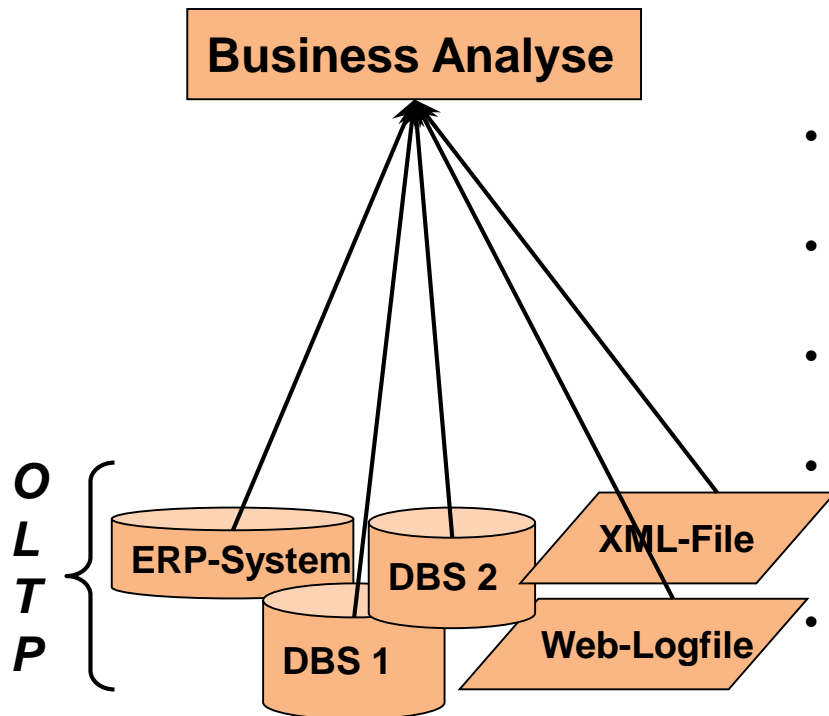
Ziele	Leistungsmaßstäbe
Innovationsprozess - Effektivität, - Zykluszeiten	Umsatz neuer Produkte, Time to Market, BreakEvenTime
Betriebszyklus: - Zykluszeiten, - Effizienz, - Qualität	Durchlaufzeiten, Kosten der Prozesse, Fehlerquote, first pass yields

Lern – und Entwicklungsperspektive

Ziele	Leistungsmaßstäbe
Mitarbeiterproduktivität	Wertschöpfung je Mitarbeiter
Mitarbeitertreue	Fluktuationsquote
Weiterbildung	Aufgabendeckungsziffer
Informationssystem	Informationsdeckungsziffer

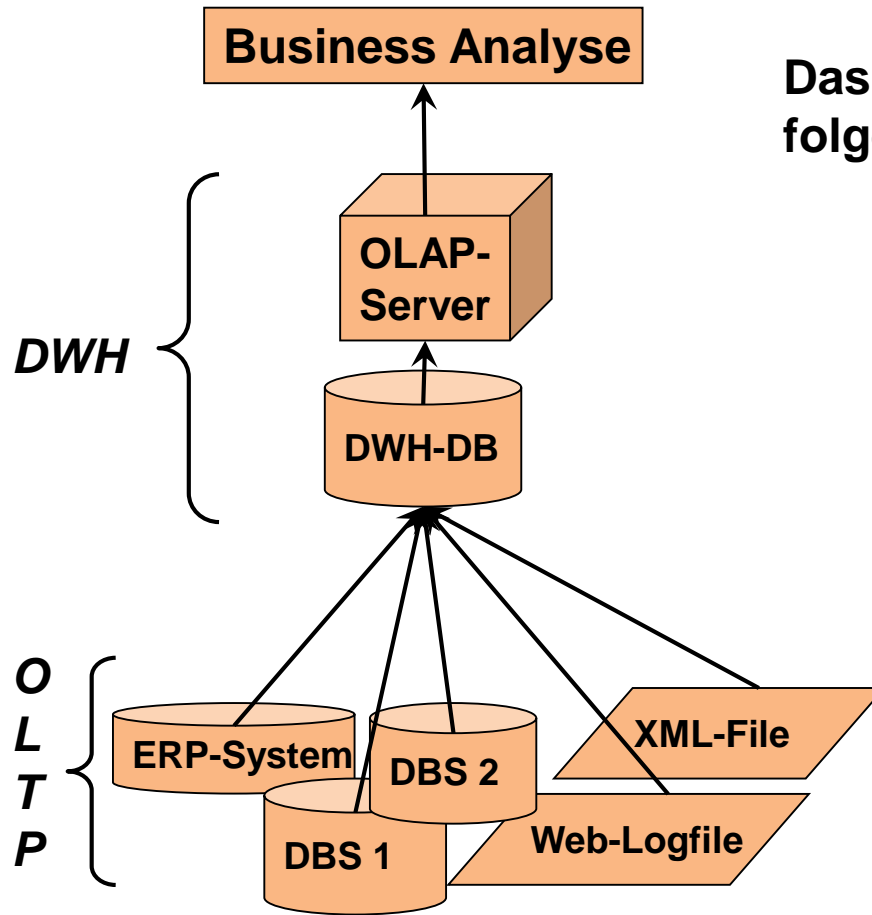
IT-Basiertes Controlling ohne Data Warehouse (DWH)

Hindernisse bei der Auswertung von Datenbestände ohne Einsatz eines Data Warehouses:



- Behinderung von Zugriff auf und Zusammenführung von Daten durch **heterogene IS-Strukturen**
- unbefriedigende und oft schwer überprüfbare **Qualität und Aktualität der Daten**
- Behinderung der Berichtserstellung durch **mangelnde Kapazität in der IT.**
- **Auslagerung historischer Daten** auf langsame, nicht online verfügbare Datenspeicher
- Erhebliche **Mehrbelastung der vorhandenen IS-Infrastruktur** (Rechner, Netzwerke, etc...) durch die notwendigen Analysen
- **Fehlende Nachvollziehbarkeit von Analysen** wegen permanenter Datenänderungen
- **unbefriedigende Performance** beim Umgang mit großen Datenbeständen

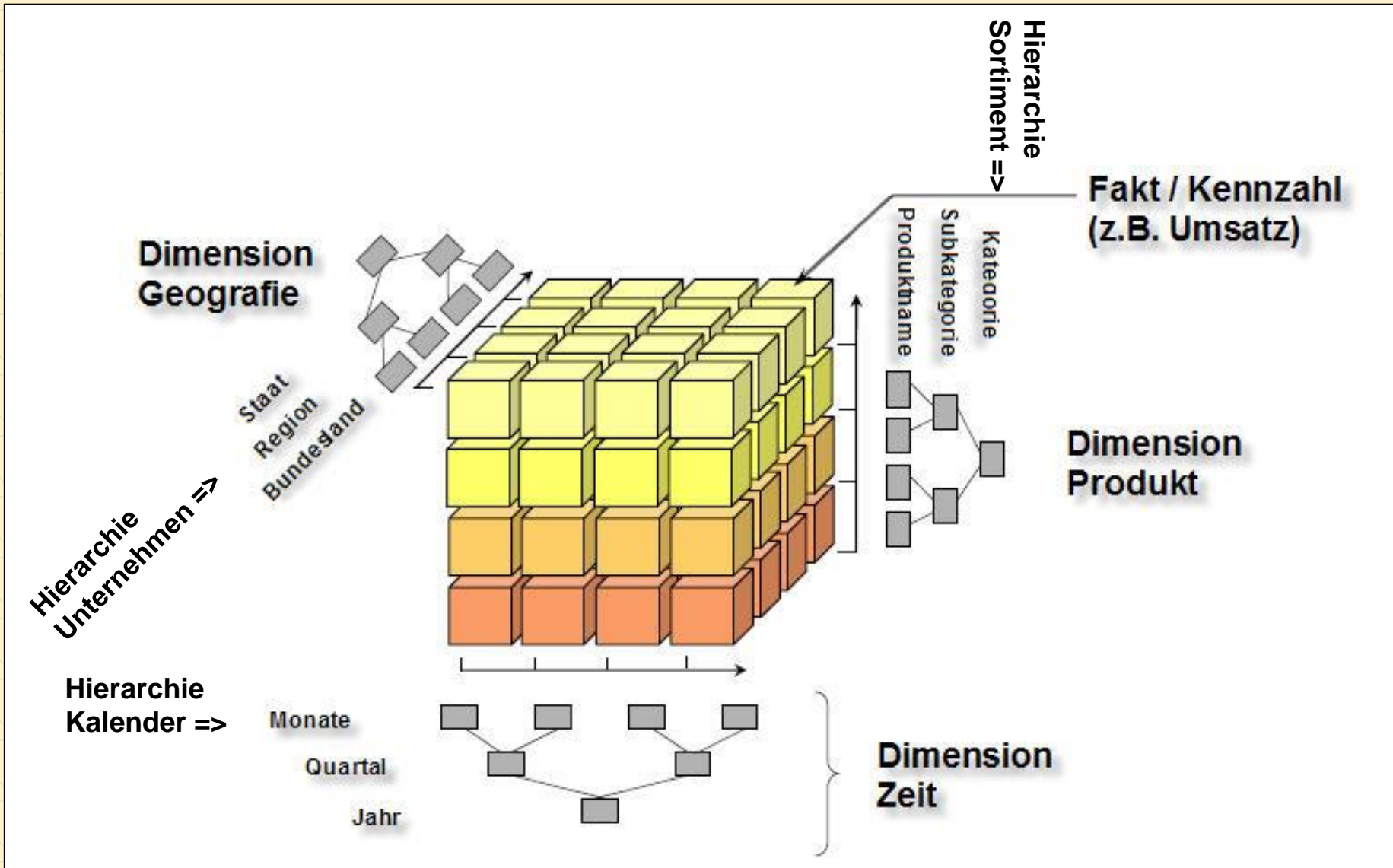
IT-Basiertes Controlling mit Data Warehouse (DWH)



Das DWH ist eine Datenbank, mit folgenden Merkmalen:

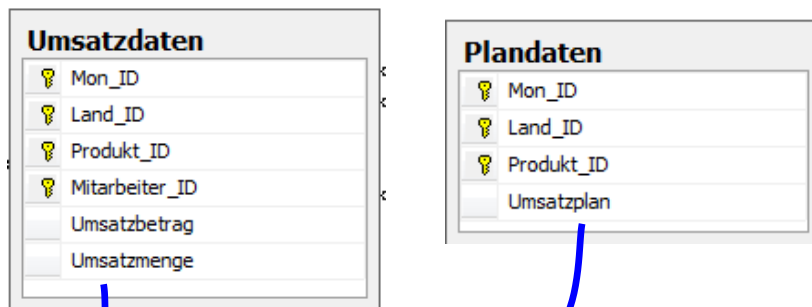
- Speicherung historische Daten
- Speicherung aggregierter Daten verschiedener Granulate
- hoch performant bei Abfragen großer Datenmengen
- optimiertes Design für die Business Analyse (spezielle Datenmodelle)
- Automatisierung der Datenbereitstellung

Veranschaulichung der Multidimensionalität

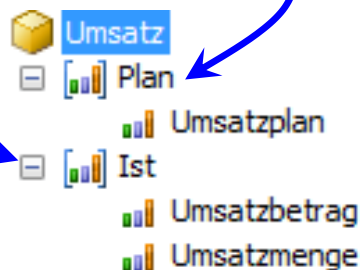


Fakttabellen – Kennzahlengruppen - Dimensionsverwendung

Fakttabellen:



Kennzahlen-
gruppen:



Dimensions-
verwendung:

Measuregruppen		
Dimensionen	Plan	Ist
Produkt	Produktname	Produktname
Zeit	Monat	Monat
Mitarbeiter	[-]	Mitarbeiter
Geografie	Bundeland	Bundeland

Auswertung der Kennzahl *Umsatzplan* nach Dimension *Mitarbeiter* nicht sinnvoll !

Zentrale Definition abgeleiteter Kennzahlen

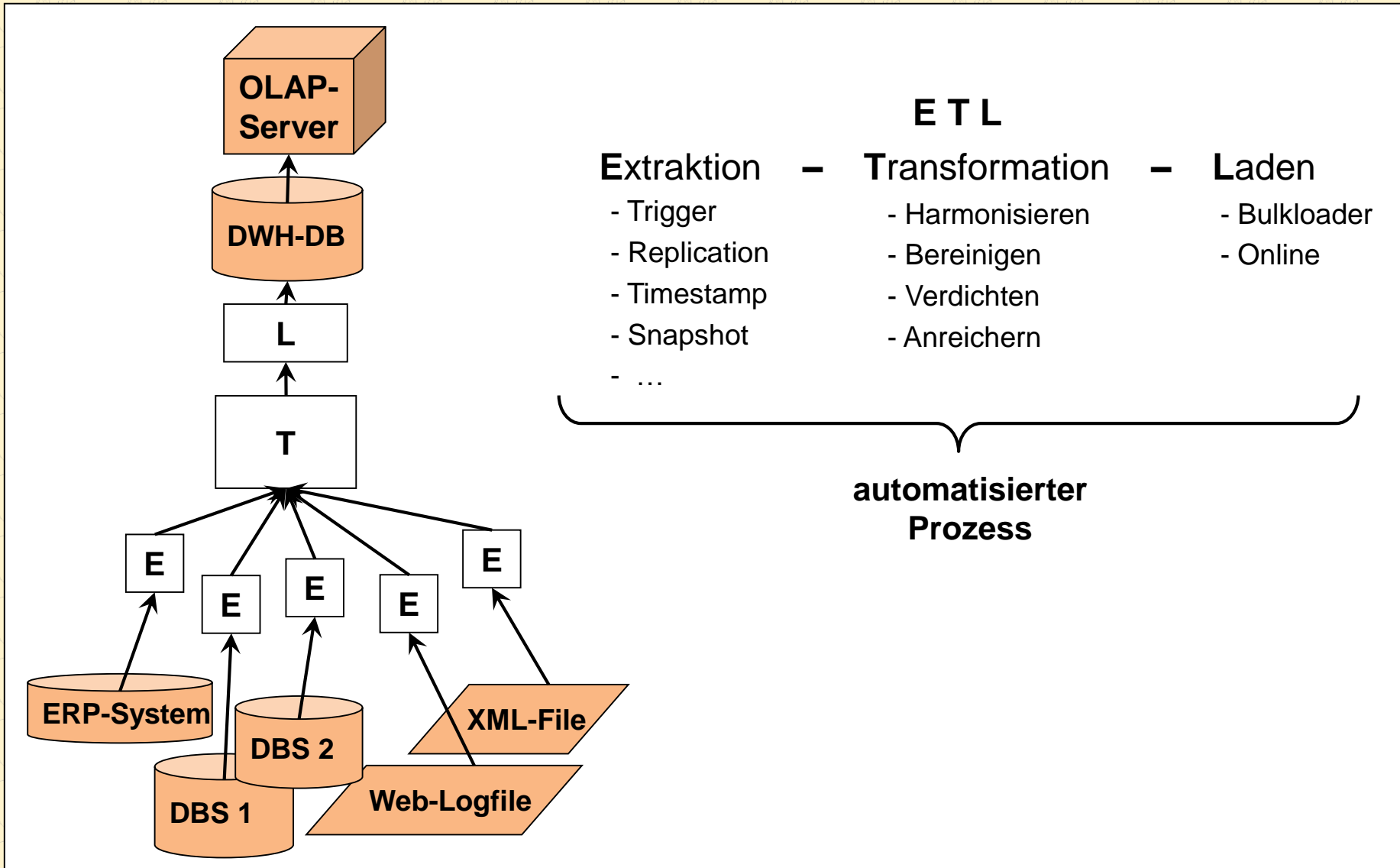
Bsp.: Wie groß ist der Anteil der Umsätze der Subkategorie an dem Umsatz der jeweiligen übergeordneten Kategorie und am Gesamtumsatz?

```
WITH
  MEMBER [Measures].[Anteil Subkategorie am Gesamtumsatz]
    AS '([Measures].[Umsatzbetrag] * 100) / ([Measures].[Umsatzbetrag],
[Produkt].[Kategorie].[All])'
    , FORMAT_STRING='###.00\%'
  MEMBER [Measures].[Anteil der Subkategorie an deren Kategorie]
    AS '([Measures].[Umsatzbetrag] * 100) / ([Measures].[Umsatzbetrag],
[Produkt].[Kategorie])'
    , FORMAT_STRING='###.00\%'
SELECT
  { [Measures].[Umsatzbetrag], [Measures].[Anteil Subkategorie am
Gesamtumsatz],
    [Measures].[Anteil der Subkategorie an deren Kategorie] } on
columns,
  { NONEMPTY([Produkt].[Kategorie].Children *
[Produkt].[Subkategorie].Children) } on rows
FROM [Umsatz]
```

		Umsatzbetrag	Anteil Subkategorie am Gesamtumsatz	Anteil der Subkategorie an deren Kategorie
Backwaren	Bagels	1.807.831,59 €	19,81%	71,72%
Backwaren	Muffins	265.937,09 €	2,91%	10,55%
Backwaren	Brot	446.957,79 €	4,90%	17,73%
Milchprodukte	Käse	1.064.149,59 €	11,66%	29,54%

■ ■ ■

Automatisierte Bereitstellung der Daten in das DWH



Flexible Analyse konsistenter DWH-Datenbestände

Beispiel der Business Analyse via Drag and Drop mittels Excel PivotTable als Front End

- Webapplication
- Smartphone
- Office
- Management Dashboard
- ...

Business Analyse

OLAP-Server

	A	B	C	D	E	F
1	Unternehmen	(Mehrere Elemente)				
2						
3	Umsatzbetrag	Spaltenbeschriftung				
4		Backwaren		Backwaren Ergebnis		Milchprodukte
5	Zeilenbeschriftung	Bagels	Muffins	Brot		Käse
6	2010	542498,24	83037,76	138247,79	763783,79	270416,82
7	1. Quartal	111933,24	19868,82	37679,53	169481,59	58862,34
8	Januar	32702,59	6695,48	13673,27	53071,34	19352,57
9	Februar	34263,72	6657,62	12076,82	52998,16	19416,2
10	März	44966,93	6515,72	11929,44	63412,09	20093,57
11	2. Quartal	170960,88	18763,16	36785,61	226509,65	55553,94
12	April	47626,33	5944,7	12294,05	65865,08	18757,33
13	Mai	78934,26	6203,17	12285,64	97423,07	19219,12
14	Juni	44400,29	6615,29	12205,92	63221,5	17577,49
15	3. Quartal	131721,56	19097,94	28727,92	179547,42	64055,38
16	Juli	45369,35	6854,96	11171,72	63396,03	21010,91
17	August	46651,38	6788,08	5446,57	58886,03	18450,23
18	September	39700,83	5454,9	12109,63	57265,36	24594,24
19	4. Quartal	127882,56	25307,84	35054,73	188245,13	91945,16
20	Oktober	39794,72	6388,91	11593,63	57777,26	30638,76
21	November	40456,59	6100,05	10941,99	57498,63	31827,21
22	Dezember	47631,25	12818,88	12519,11	72969,24	29479,19
23	2011	588345,35	83009,47	139646,14	811000,96	394230,72
24	1. Quartal	130273,53	19580,2	30885,52	180739,25	92523,73
25	Januar	40093,77	6433,36	12914,19	59441,32	30907,38
26	Februar	41848,59	6373,87	11560,19	59782,65	31082,41
27	März	48331,17	6772,97	6411,14	61515,28	30533,94
28	2. Quartal	193697,79	18881,97	42433,23	255012,99	96469,15

PivotTable-Feldliste

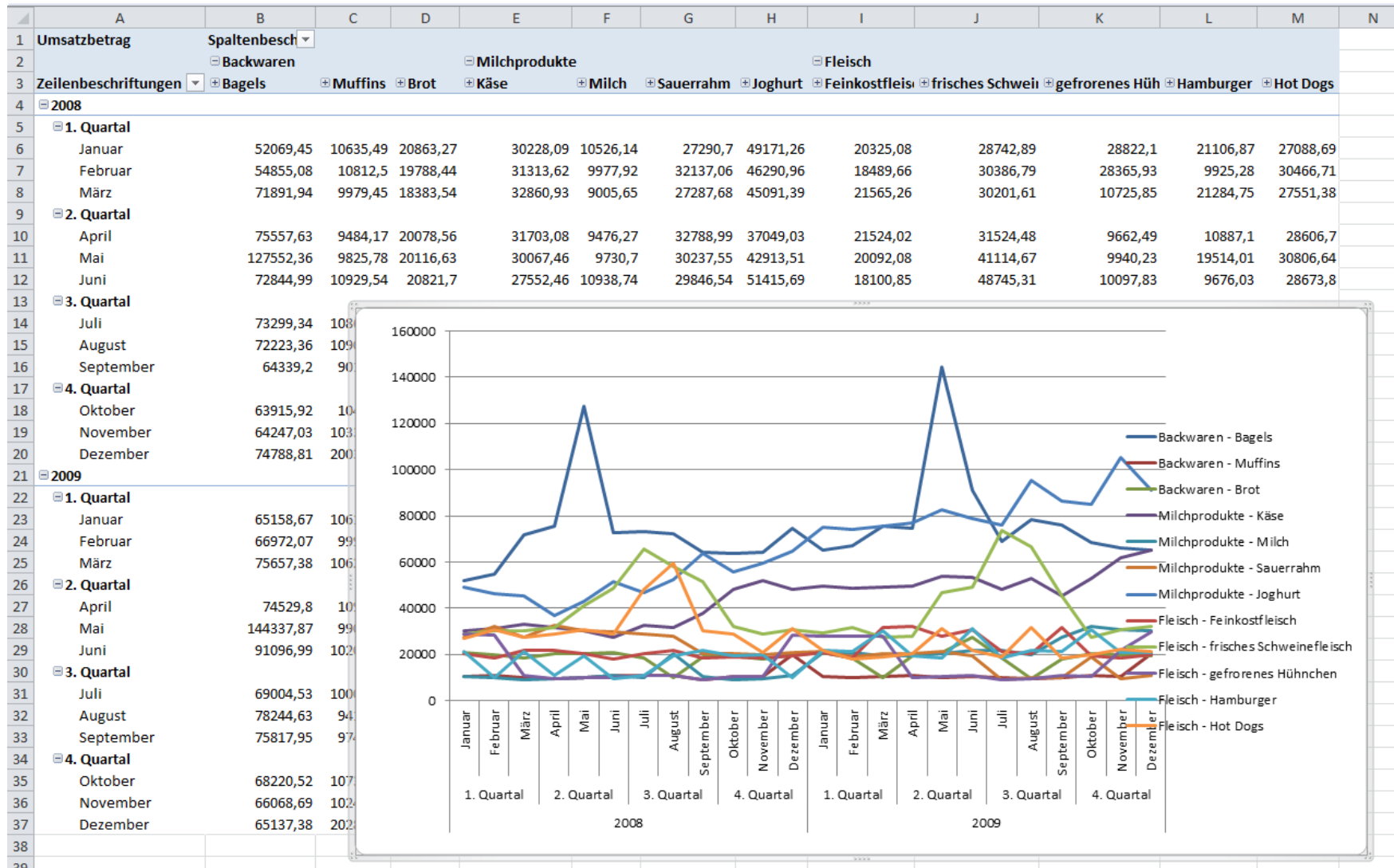
Wählen Sie die Felder aus, die Sie dem Bericht hinzufügen möchten:

- Umsatzdaten**
 - Umsatzbetrag
 - Umsatzdaten Anzahl
 - Umsatzmenge
 - Umsatzsteuer
- Geografie**
 - Unternehmen
 - Weitere Felder
- Mitarbeiter**
 - Management
- Produkt**
 - Sortiment
 - Weitere Felder
- Zeit**
 - Kalender

Felder zwischen den Bereichen unten ziehen:

Layoutaktualisierung zurückstellen

Auswertung des DWH mittels Excel PivotTables (1)

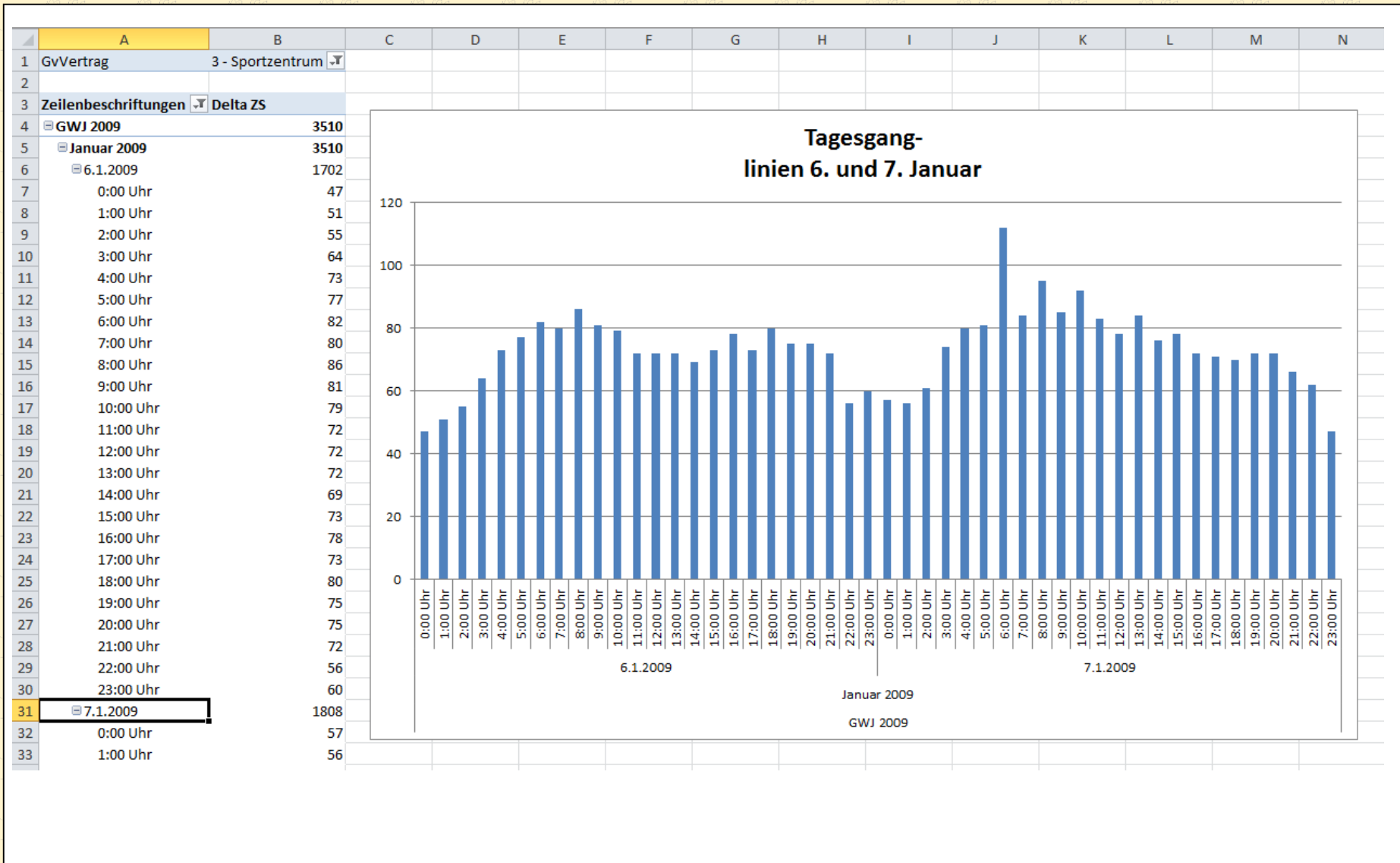


Auswertung des DWH mittels Excel PivotTables (2)

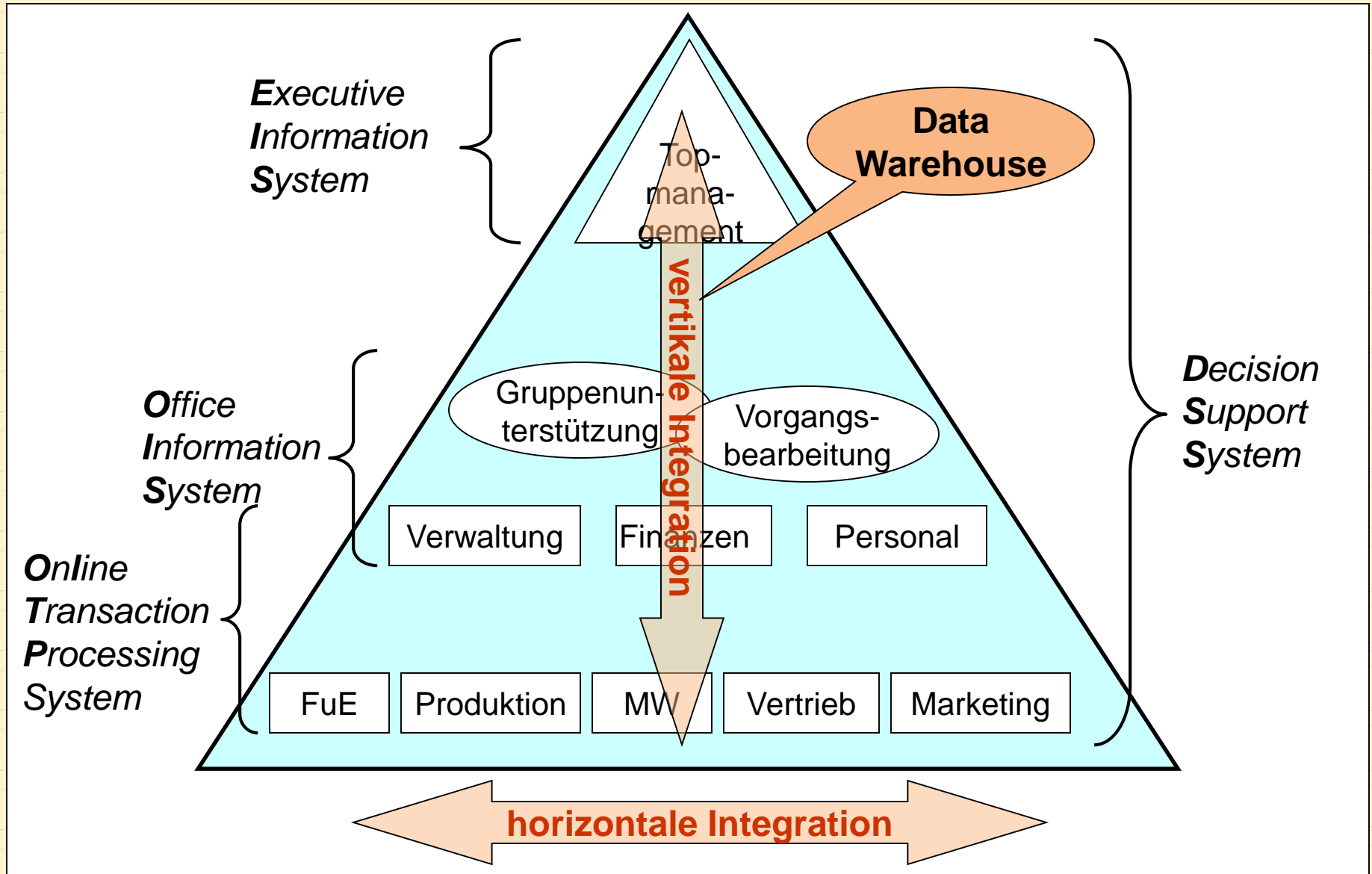
	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	AA	
1		Sp																										
2		GWJ 2009																										
3		Oktober	Novemb	Dezemb	Januar 2	Februar	März 20	April 20	Mai 200	Juni 200	Juli 2009	August 2	September 2009															
4	Zeilenbeschriftungen	Lmax	Lmax	Lmax	Lmax	Lmax	Lmax	Lmax	Lmax	Lmax	Lmax	Lmax	Lmax	Lmax	Lmax	Lmax	Lmax	Lmax	Lmax	Lmax	Lmax	Lmax	Lmax	Lmax	Lmax	Lmax	Lmax	Status
5	1 - Mustervertrag 1	0,51	0,62	0,65	0,59										0,18	0,17	0,19	0,60	0,65									
6	000102 - Abnahmestellen Preisgruppe 001	0,51	0,62	0,65	0,59										0,18	0,17	0,19	0,60	0,65									
7	15 - Wohnheim Block B	0,51	0,62	0,65	0,59										0,18	0,17	0,19	0,60	0,65									
8	2 - Verbundvertrag	0,46	0,56		1,06	0,93	1,08	0,64	0,43	0,50	0,41	0,42	0,40	1,08														
9	000201 - Verträge ohne Grundpreis		0,11		1,06	0,93	1,08	0,64	0,43	0,50	0,41	0,42	0,40	1,08														
10	3 - Sportzentrum				1,06	0,93	1,08	0,64	0,43	0,50	0,41	0,42	0,40	1,08														
11	7 - Schulheizwerk				0,70																							
12	8 - Heizwerk WG 1																											
13	184 - Ontario KG																											
14	189 - Schweizer Orangen AG		0,11																									
15	000205 - Vertrag Spezialheizwerk	0,46	0,56		0,58																							
16	304 - Merkur KG	0,46	0,56		0,58																							
17																												
18																												
19																												
20																												
21																												
22																												
23																												
24																												
25																												
26																												
27																												
28																												
29																												
30																												
31																												

Drill-Down

Auswertung des DWH mittels Excel PivotTables (3)



Integration von IT-Systemen



Vielen Dank für Ihre Aufmerksamkeit

Papa, wie seid ihr eigentlich früher
ins Internet gekommen, als es
noch keine Computer gab?



[<http://geschaut.com>] (11.04.2013)