

# IAK WI

Herausgeber

**Thomas Barton**

**Burkhard Erdlenbruch**

**Frank Herrmann**

**Christian Müller**

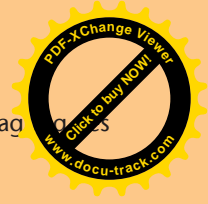
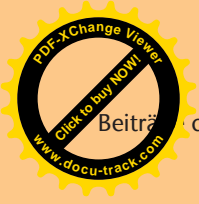
**Joachim Schuler**

Herausforderungen an die Wirtschaftsinformatik:

## Management und IT

unterstützt durch:

HOCHSCHULE PFORZHEIM 



Beitrag zur Fachtagung »**Management und IT**« im Rahmen der 25. Jahrestag  
Arbeitskreises Wirtschaftsinformatik an Fachhochschulen (AKWI)  
vom 16.09. bis 18.09.2012 an der Hochschule Pforzheim

Autoren:

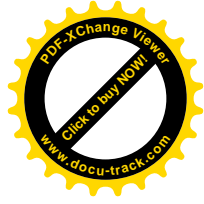
Wolfgang Alm, Christoph Arnold, Yves Augustin, Thomas Barton, Yvonne Bentz,  
Bettina C.K. Binder, Werner Burkard, Roberto Buscemi, Sven Denecken, Armin Elbert,  
Martin Gaubitz, Andreas Heberle, Frank Herrmann, Johannes Hoen, Georg Rainer Hofmann,  
Pezzotta Jeff, Norbert Ketterer, Oliver Kuchler, Elvira Kuhn, Martin Kütz, Christian Ludwig,  
Frank Morelli, Nils Mosbach, Christian Müller, Gerald Münzl, Gervais Jocelyn Ndongmo,  
Rainer Neumann, Djieukeng Ngoune Neylor, Kurt Porkert, Ute Riemann, Harald Ritz,  
Peter Scholz, Mathias Schröder, Bert O. Schulze, Meike Schumacher, Christian Seel,  
Carlo Simon, Bruno Sticht, Michael Tax, Stephan Thesmann, Heiko Thimm,  
Philipp Walter, Wendelin Wiedeking, Patrick M. Wolff,

Verlag News & Media, Berlin

**ISBN 978-3-936527-30-8**



Arbeitskreis Wirtschaftsinformatik an Fachhochschulen (AKWI)



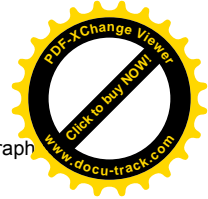
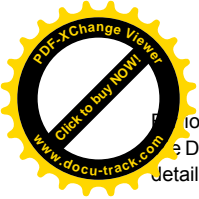
# **Management und IT**

Tagungsband zur AKWI-Fachtagung  
vom 16. bis 18.09.2012 an der Hochschule Pforzheim

herausgegeben von  
Thomas Barton, Burkhard Erdlenbruch, Frank Herrmann,  
Christian Müller, Joachim Schuler

Unterstützt durch den  
Studiengang Wirtschaftsinformatik an der Hochschule Pforzheim

Verlag News & Media, Berlin



Bibliographische Information der Deutschen Bibliothek:  
⇒ Deutsche Bibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliographie  
Detaillierte bibliographische Daten sind im Internet über <http://dnb.ddb.de> abrufbar.

## Management und IT

Tagungsband zur AKWI-Fachtagung vom 16. bis 18.09.2012  
an der Hochschule Pforzheim

Herausgeber:

Prof. Dr. Thomas Barton, *Fachhochschule Worms*,  
[barton@fh-worms.de](mailto:barton@fh-worms.de)

Prof. Dr. Burkhard Erdlenbruch, *Hochschule Augsburg*,  
[Burkhard.Erdlenbruch@hs-augsburg.de](mailto:Burkhard.Erdlenbruch@hs-augsburg.de)

Prof. Dr.-Ing. Frank Herrmann, *Hochschule Regensburg*,  
[Frank.Herrmann@informatik.fh-regensburg.de](mailto:Frank.Herrmann@informatik.fh-regensburg.de)

Prof. Dr. Christian Müller, *Technische Hochschule Wildau [FH]*,  
[christian.mueller@th-wildau.de](mailto:christian.mueller@th-wildau.de)

Prof. Dr. Joachim Schuler, *Hochschule Pforzheim*,  
[joachim.schuler@hs-pforzheim.de](mailto:joachim.schuler@hs-pforzheim.de)

Redaktion:

Teamarbeit der Herausgeber

Redaktionsschluss: 01.08.2012

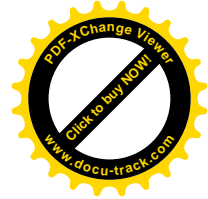
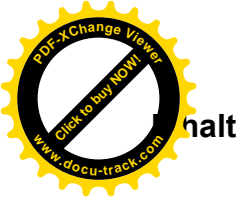
Erscheinungstermin: 16.09.2012

# HOCHSCHULE PFORZHEIM

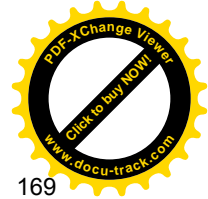
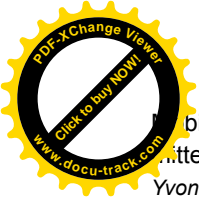
Die Herstellung dieses Tagungsbandes erfolgte mit freundlicher Unterstützung durch den Studiengang Wirtschaftsinformatik an der Hochschule Pforzheim

Verlag News & Media, von Amsberg, Berlin

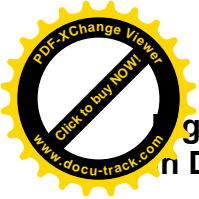
ISBN 978-3-936527-30-8



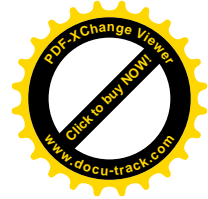
Vorwort des Sprechers des Arbeitskreises Wirtschaftsinformatik an Fachhochschulen (AKWI) .....	5
Vorwort der Herausgeber .....	7
Cloud-Lösungen – vom Hype zum Trend Herausforderungen und Chancen aus der Sicht des IT-Management <i>Bert O. Schulze, Sven Denecken</i> .....	9
Eignung unterschiedlicher Faktenmodellierungen in Data Warehouse-Systemen <i>Christoph Arnold, Harald Ritz</i> .....	23
Data-Mining Routinen in SAP BI 7.0 <i>Yves Augustin, Norbert Ketterer</i> .....	41
Generierung von Simulationsmodellen aus ereignisgesteuerten Prozessketten <i>Christian Müller</i> .....	55
Unterstützung einer selbststeuerenden Planung durch eine geeignete Erweiterung des SAP Systems <i>Frank Herrmann</i> .....	65
Cloud Computing als neue Herausforderung für Management und IT <i>Thomas Barton, Gerald Münzl</i> .....	77
Der Weg zur Enterprise-Cloud <i>Andreas Heberle, Rainer Neumann</i> .....	105
Zentrales Sicherheitsmanagement für Cloud-Infrastrukturen <i>Christian Seel, Philipp Walter, Peter Scholz, Johannes Hoen</i> .....	119
Entscheidungskonzept für die Transformation einer IT-Landschaft durch Cloud-Service-Einsatzszenarien <i>Kurt Porkert</i> .....	133
Was machen neue Informationstechnologien wie Internet und iDevices aus unseren Unternehmen? Auswirkungen neuer Technologien auf die Unternehmensprozesse und Arbeitsplatzgestaltung. <i>Elvira Kuhn, Pezzotta Jeff, Djieukeng Ngoune Neylor, Gervais Jocelyn Ndongmo</i>	151



Mobile Management- und Vertriebsunterstützung in einem mittelständischen Unternehmen auf Basis von SAP BW <i>Yvonne Bentz, Bruno Sticht, Armin Elbert, Thomas Barton</i> .....	169
Nachhaltige Unternehmens-Fitness durch Messung des IT-Unterstützungspotenzials in Geschäftsprozess-Reifegradmodellen <i>Frank Morelli, Roberto Buscemi, Wendelin Wiedeking</i> .....	179
Optimierung bestehender Geschäftsprozesse durch den Einsatz mobiler Applikationen am Beispiel eines mittelständischen Unternehmens mit diskreter Fertigung <i>Frank Morelli, Mathias Schröder</i> .....	191
Berechnung des IT-Wertbeitrags <i>Martin Kütz</i> .....	205
Dokumentenmanagement auf Basis von JSF-Komponenten <i>Werner Burkard, Nils Mosbach</i> .....	219
mainproject – ITIL-Wissenstransfer im Rahmen des ESF-Programms <i>Meike Schumacher, Michael Tax, Georg Rainer Hofmann, Wolfgang Alm</i> .....	243
Steuerung von IT-Services anhand von Prozesskennzahlen <i>Oliver Kuchler, Carlo Simon</i> .....	259
Strategische Planung der Anwendungslandschaft unter Einsatz von Gruppenentscheidungsunterstützungssystemen Eine Nutzenbetrachtung und ein Ausblick auf das Forschungsprojekt GRUPO-MOD <i>Heiko Thimm</i> .....	271
Onlinemarketing: Effektivität und Effizienz von SEO- und SEM-Maßnahmen <i>Stephan Thesmann, Patrick M. Wolff, Martin Gaubitz, Christian Ludwig</i> .....	295
Bewertung nichtfinanzieller Nutzeffekte in IT-Projekten <i>Martin Kütz</i> .....	315
Einführung einer SAP Prozessdefinition und -implementierung in einem Projekt bei Giesecke & Devrient <i>Bettina C.K. Binder, Ute Riemann</i> .....	329
Autoren .....	353



# Angewandte Analyse unterschiedlicher Faktenmodellierungen in Data Warehouse-Systemen



*Christoph Arnold, Harald Ritz*

## 1 Einleitung

Data Warehousing und Business Intelligence sind in der heutigen IT-Welt fest installierte Begriffe. Auch wenn die Begriffsverständnisse sehr unterschiedlich sein können, so verfolgen deren Anwender und Entwickler stets das Ziel der besseren Informationsversorgung. Diese Informationen sind eine notwendige Grundlage für die Generierung von Wissen, das für weitere Entscheidungen benötigt wird.

Entscheidungsprozesse im Unternehmen werden durch Globalisierung, sich angleichende Konkurrenz und stetig ändernde Rahmenbedingungen immer komplexer und folgenschwerer. Dadurch wächst der Rohstoff Daten zu einem wettbewerbsentscheidenden Faktor an. Durch seine potenzielle Wandlung über Informationen in Wissen entwickelt er sich zunehmend zum vierten Produktionsfaktor neben Arbeit, Boden und Kapital (vgl. [WiMW09], S. 118 f.). Die Optimierung der Informationsversorgung und die darauf folgende Wissensgenerierung werden auch in der Zukunft eine bedeutende Rolle spielen. Durch neue Technologien und bessere Ausbildung der Mitarbeiter werden die Anforderungen an eine solche Informationsversorgung auch zukünftig wachsen.

Durch diese wachsenden Anforderungen und die Komplexität bei der Arbeit mit diesen Systemen werden Anwender häufig überfordert, was unmittelbar zu einer Minderung der Qualität der Informationsversorgung führt. Aber auch Entwickler dieser Systeme können sich bei der Menge der Anforderungen und Möglichkeiten schnell verlaufen und so bereits eine mangelbehaftete Grundlage schaffen. Daher ist es notwendig, diese wachsende Komplexität zu analysieren, zu strukturieren und zu bewerten, um auch in Zukunft eine hohe Informationsqualität gewährleisten zu können. Wer diese Aufgabe bewerkstelligt, wird konkurrenzfähig bleiben und kann sogar eigene Wettbewerbsvorteile erzielen. Allerdings ist diese Aufgabe mit einem einmaligen Konzept nicht abgeschlossen. Vielmehr ist die kontinuierliche Wissensgenerierung und -strukturierung ein anhaltender Prozess, der auch zukünftig fortgeführt werden muss. Zentrale Aspekte sind dabei die zur Analyse verfügbaren betrieblichen Kennzahlen (Fakten).

Dieser Artikel liefert einen Überblick über die gängigen Methoden bei der Faktenmodellierung in Data Warehouse-Systemen und analysiert dabei individuell,



Die Einsatzgebiete durch die jeweilige Methode abgedeckt werden können. Dabei wird darauf geachtet, einen möglichst breiten Querschnitt der Anwendungsfälle zu behandeln, um für unterschiedlichste Situationen eine passende Modellierungsmethode zu bieten.

Dazu wurden etablierte Modelle renommierter Autoren – allen voran Ralph Kimball – analysiert und durch langjährige Erfahrungen von Entwicklern ergänzt. Diese Ergebnisse bilden die Grundlage für diese Arbeit.

Im folgenden zweiten Kapitel werden grundlegend die unterschiedlichen Modellierungsmethoden vorgestellt. Das dritte Kapitel analysiert und evaluiert die Methoden und stellt deren Stärken und Schwächen heraus. Es schließt mit einem Ausblick auf die vielfältigen Möglichkeiten mittels unterschiedlicher Kombinationen von Modellen. Im abschließenden vierten Kapitel werden aktuelle Anforderungen und Technologien behandelt, um so einen Ausblick auf die weiteren Entwicklungen dieses Themengebiets aufzuzeigen.

## 2 Faktenmodellierungen in Data Warehouse-Umgebungen

### 2.1. Datenmodellierung

Die Grundlage für nachfolgende Arbeiten schafft die Modellierung der Datenlandschaft. Ziel ist es, ein Datenmodell zu konzipieren, welches den fachlichen Anforderungen bestens genügt. Ein nicht geeigneter Entwurf wird sich auch durch alle folgenden Bereiche ziehen. Daher ist die Modellierung ein wesentlicher Faktor, der über den Erfolg eines Projekts entscheidet, da er je nach eingesetztem System auch direkt die Anwenderzufriedenheit beeinflusst.

Zur logischen Datenmodellierung stehen verschiedene Methoden zur Verfügung. Diese unterscheiden sich in ihren zugrunde liegenden Konzepten und ergeben dabei unterschiedliche Datenmodelle.

Die beiden gebräuchlichsten Methoden in der multidimensionalen Datenmodellierung sind das Star-Schema und das Snowflake-Schema. Dazu gibt es weitere Abwandlungen, welche ein Schema um besondere Eigenschaften ergänzen und bei Bedarf eingesetzt werden können (vgl. [BaGü09], S. 213-225; [Fark11], S. 27-32 und [KeBM10], S. 66-71).

Das Star-Schema zeichnet sich vor allem durch seinen einfachen strukturellen Aufbau aus. Zentral ist die Faktentabelle positioniert. Sie wird durch die einzelnen Dimensionen um Stammdateninformationen erweitert, enthält aber als einzige Tabelle die individuellen Daten über ein Faktum (z. B. eine Transaktion). Durch die Anordnungen der Dimensionstabellen entsteht der Eindruck eines Sternumrisses, woher der Name Star-Schema stammt (vgl. [KiRo02], S. 21f.). Nachfolgende Abbildung verdeutlicht diesen Aufbau an einem Beispiel zur Umsatzanalyse (entnommen aus [KeBM10], S. 68).



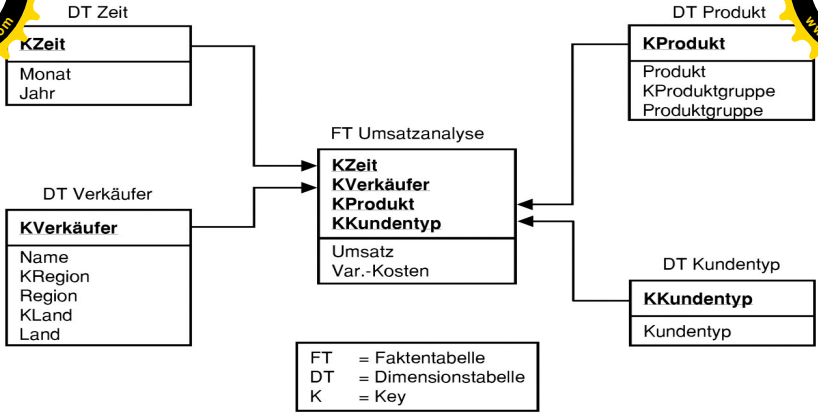
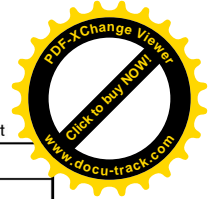
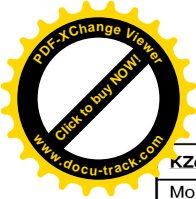


Abb. 1: Beispiel eines Star-Schemas ([KeBM10], S. 68)

Im Gegensatz zum Star-Schema versucht das Snowflake-Schema die Dimensionstabellen zu normalisieren. Dadurch lässt sich Speicherplatz einsparen und die Datenkonsistenz kann einfacher sichergestellt werden (vgl. [KiRo02], S. 21).

Welche Methode der Entwickler am Ende wählt, muss die jeweilige Situation entscheiden. Meistens ist das Star-Schema die beste Wahl, da der Speicherplatz der Dimensionstabellen vernachlässigbar ist im Vergleich zur Faktentabelle. Auch Änderungsanomalien treten durch Historisierungsmethoden meistens nicht auf. Sollte es der Fall sein, dass sich Dimensionen häufig ändern, und sollen die Änderungen auch auf historische Fakten übertragen werden, so bietet sich das Snowflake-Schema an (vgl. [KiRo02], S. 21).

## 2.2. Faktentabellen

Faktentabellen enthalten alle Aktivitäten, die bei der Untersuchung von Geschäftsprozessen relevant sind (vgl. [KiCa04], S. 45f.). So enthalten sie beispielsweise Daten über verkaufte Produkte an Kaufhauskassen, Retourendaten in Industrieunternehmen oder Kurse an der Börse.

Faktentabellen bestehen nur aus Fremdschlüsseln auf die Dimensionen (die zusammen den Primärschlüssel ergeben) und Kennzahlen. Diese enthalten Messwerte über das Faktum und liegen in der Regel in numerischer Form vor. Sämtliche Beschreibungsmerkmale, die nicht in direkter Abhängigkeit des Faktums stehen, sollten hingegen in die Dimensionstabellen ausgelagert werden (vgl. [KiRo02], S. 17-19).

Wichtig ist an dieser Stelle, diese Daten möglichst auf atomarer Ebene zu erfassen. Das bedeutet, dass nicht die gesamte Transaktion aggregiert erfasst wird, sondern jeder einzelne Artikel. So werden die Analysemöglichkeiten



nt eingeschränkt und der Anwender hat später viele Freiheiten, durch se  
Daten zu navigieren. Ralph Kimball drückt diesen Sachverhalt wie folgt aus.



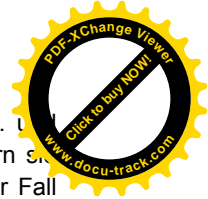
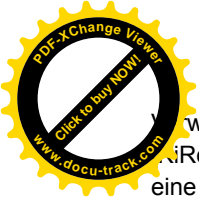
*„Atomic data that has not been aggregated is the most expressive data;  
this atomic data should be the foundation for every fact table design  
in order to withstand business users' ad hoc attacks  
where they pose unexpected queries.“ ([KiRo02], S. 22f.)*

Wie Kimball bereits erwähnt, ist diese atomare Erfassung der Daten nur die Grundlage für jedes Faktentabellendesign. So ist es auch möglich, weitere Faktentabellen zu erstellen, in denen Daten in aggregierter Form vorliegen. Die Vorteile dieser Methode liegen in der deutlich besseren Abfrageperformance für aggregierte Daten und der Möglichkeit, Daten dennoch auf atomarer Ebene auszuwerten. Diese Tabellen sind die Grundlage für spätere Analysen. Sie sind auch die mit Abstand größten Tabellen in dem System und enthalten häufig mehrere Milliarden Datensätze. Ein Richtwert der Größenverhältnisse ist 90/10. Mindestens 90 % des Speicherplatzes wird dabei für die Faktentabelle verwendet und nur maximal zehn Prozent benötigen die Dimensionstabellen (vgl. [KiRo02], S. 18 u. S. 21).

Faktentabellen sind ein entscheidender Faktor für den späteren Erfolg der kompletten Business-Intelligence-Umgebung. Ihr Einfluss auf die Performance und die Nutzerfreundlichkeit im Reporting ist sehr groß und eine mangelhafte Modellierung führt auch in den Folgearbeiten zu Problemen. Dadurch ergibt sich eine hohe Relevanz der Faktenmodellierung, die den Fokus dieses Artikels begründet. Nur durch ein gutes Design lassen sich optimale Ergebnisse der Gesamtarchitektur erreichen.

Eine wichtige Entscheidung zu Beginn der Faktenmodellierung ist die Anzahl der Faktentabellen. Vielen Anforderungen genügt eine einzige Faktentabelle, die atomare Fakten enthält. Problematisch wird es, wenn unterschiedliche Prozesse betrachtet werden, wie zum Beispiel Einkauf und Verkauf. Diese Transaktionen sind nahezu identisch aufgebaut. Daher sollten sich Faktentabellen immer auf einen Prozess beziehen, um keine Kollisionen zu verursachen. Die Ergebnisse wären unbrauchbar, wenn Einkäufe versehentlich mit den Verkäufen vermischt würden.

Ebenso wichtig wie die Prozessorientierung ist die Granularität, d. h. die Feinkörnigkeit der Fakten. Wie zuvor erwähnt, kann eine unterschiedliche Granularität bewusst durch Aggregation hervorgerufen werden. Monatlich summierte Transaktionen lassen sich nicht mit atomaren Transaktionen kombinieren. Für jeden notwendigen Detailgrad sollte eine eigene Faktentabelle verwendet werden, um Folgekomplikationen zu vermeiden (vgl. [KiRo02], S. 121f.).



Wandlung mit der Granularität ist die Additivität (vgl. [Fark11], S. 86f. und [iRo], S. 17). Die Werte einer Spalte sollten stets additiv sein, sofern sie eine Kennzahl enthalten, um sie aggregieren zu können. Es kann der Fall eintreten, dass diese Werte nur über bestimmte Dimensionen additiv sind. Bestandsgrößen sind z. B. nicht über die Dimension Zeit additiv aggregierbar, jedoch über die Dimension Ort. Dann spricht man von Semi-Additivität. Beschreibungsmerkmale oder nicht numerische Werte sind grundsätzlich nicht-additiv. Dies schließt jedoch nicht aus, dass auch numerische Werte nicht-additiv sind. So können Temperaturen z. B. nicht über mehrere Monate, aber auch nicht über mehrere Orte addiert werden. Eine Ausnahmeaggregation wäre in diesem Fall die Durchschnittsberechnung.

### **2.3. Faktenmodellierung**

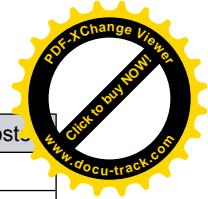
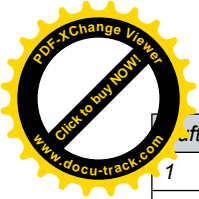
Es existieren zahlreiche unterschiedliche Ansätze, wie Faktentabellen modelliert werden können. Jede Methode eignet sich für ganz bestimmte Anforderungen und besitzt eigene Vor- und Nachteile. In den folgenden Abschnitten werden einige der am häufigsten benötigten Modellierungen vorgestellt. Dabei wird stets das Ziel verfolgt, die Anforderungen darzustellen, auf die diese Modellierung am besten passt.

Dieser Punkt ist gleichzeitig auch der schwierigste, da auf den ersten Blick ähnliche Umgebungen häufig sehr unterschiedliche Anforderungen haben. Zur erfolgreichen Auswahl einer geeigneten Methode muss der Entwickler über gewisse Fähigkeiten zum Wissenstransfer verfügen. Dieser Artikel soll dabei helfen, verschiedene Anforderungen möglichst einfach zu erkennen und einzuschätzen, welche Stärken der unterschiedlichen Methoden bei jeweils vorliegenden Situationen zum Tragen kommen könnten. Eine einfache Schritt-für-Schritt-Anleitung kann es allerdings nicht geben. Der letzte Schritt der Auswahl kann nur unterstützt, aber nicht abgenommen werden.

Die Bezeichnungen für die folgenden Modellierungen wurden entweder der gängigen Literatur entnommen oder sie basieren auf den Grundideen der jeweiligen Methode und stammen von den Autoren.

### **2.4. Einfache Starmodellierung**

Eine relativ einfache Modellierungsmethode ist die Modellierung in die Breite (im Folgenden genannt: Einfache Starmodellierung). Dabei wird für jede Kennzahl eine eigene Spalte in der Tabelle angeboten. Auf diese Weise stehen alle Kennzahlen gemeinsam in einem Datensatz. So entspricht diese Methode der einfachsten Modellierung mit dem Star-Schema, bei dem alle Merkmale in einem Datensatz erfasst werden. Im vorliegenden Beispiel verweist DatumId auf eine eigene Dimension, welche weitere Informationen zu diesem Faktum bietet, nach denen die Daten aggregiert und gruppiert werden können.



Auftragld	Datumld	Fahrtkosten	Arbeitszeit	Materialkosten
1	2	140	5	100
2	45	20	30	0

Tab. 1: Einfache Starmodellierung

Ein sehr großer Vorteil dieser Methode ist die gute Performance bei bestimmten Abfragen. Da alle auftragsbezogenen Messwerte in einer Zeile stehen, können diese schnell verarbeitet werden, wenn gesamte Aufträge ausgewertet werden sollen. Man muss dazu nur alle Datensätze durchlaufen und die Messwerte aggregieren.

Im Gegensatz dazu sind Analysen über bestimmte Kennzahlen sehr ressourcenaufwendig. Möchte man nur alle berechneten Materialkosten auswerten, so müssen trotzdem sämtliche weiteren Datensätze mitsamt ihren Spalten geladen werden. Bei der Betrachtung von einer Kennzahl ist das Laden von vielen weiteren Spalten notwendig, jedoch sind diese Daten unbrauchbar und unnötig geladen. Dieses Problem wird unter anderem durch spaltenbasierte Datenbanksysteme kompensiert. Zur Vertiefung dieser Thematik vgl. u. a. [AbMH08], [Bößw10], [PIZe11] und [StCe05].

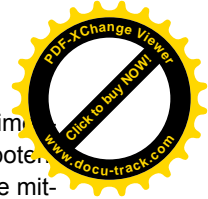
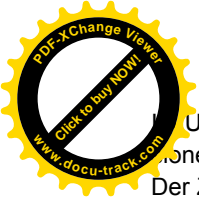
## 2.5. Degenerierte Dimensionen

In der Regel verweisen Schlüssel in Faktentabellen auf Datensätze in Dimensionstabellen. Im Gegensatz dazu existieren bei Degenerierten Dimensionen keine Fremdschlüsselverweise auf Dimensionstabellen. In der Faktentabelle sind Spalten vorhanden, die auf den ersten Blick nur wie Fremdschlüssel auf Dimensionstabellen wirken (vgl. [KiRo02], S. 50). Auch das folgende Beispiel enthält mit der Spalte Auftragld eine Degenerierte Dimension (kursiv hervorgehoben).

<i>Auftragld</i>	Datumld	Produktld	Betrag	Menge
1	12	51	23,99	1
1	12	138	4,95	2
1	12	14	0,50	4

Tab. 2: Degenerierte Dimensionen

Der Zweck dieser Dimension ist es, bei Analysen die Möglichkeit zu bekommen, zusammenhängende Datensätze ebenso zusammenhängend auszuwerten. Dies eignet sich in der Regel besonders gut für Eltern-Kind-Beziehungen zwischen Fakten. Das oben gezeigte Beispiel macht dies anhand eines Supermarkts deutlich, in dem Transaktionen an der Kasse stattfinden und diese aus unterschiedlichen Produkten bestehen.



Unterschied zu einer realen Dimension werden bei Degenerierten Dimensionen keine weiteren Informationen für diese Pseudo-Dimension angeboten. Der Zweck, der mit einer solchen Spalte verfolgt wird, ist es, Datensätze miteinander zu gruppieren. Wächst eine Dimensionstabelle (nahezu) mit jedem Eintrag in der Faktentabelle, kann eine Degenerierte Dimension vorliegen, die nicht als solche modelliert wurde (vgl. [Beck03]).

## 2.6. Fact Constellation-Schema

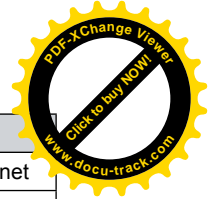
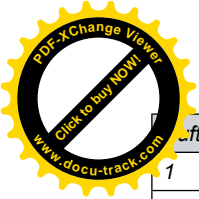
Eine besondere Form des Galaxy-Schemas (Star-Schema mit mehreren Faktentabellen) stellen vorberechnete, aggregierte Faktentabellen dar. Auch hier werden mehrere Faktentabellen eingesetzt. Diese behandeln allerdings keine unterschiedlichen Geschäftsvorfälle, sondern nur einen einzigen. Dabei werden unterschiedliche Granularitäten und Aggregationsstufen in verschiedenen Faktentabellen gespeichert (vgl. [BaGü09], S.221).

Ein Fact Constellation-Schema tritt in der Regel dann auf, wenn vorberechnete Aggregate gebildet werden. Da diese nicht in die normale Faktentabelle geschrieben werden sollten, um in dieser die gleiche Granularität zu gewährleisten, besteht die Möglichkeit, diese in weitere Faktentabellen auszulagern. Konkret könnte man im vorherigen Beispiel zwei weitere Faktentabellen erstellen, welche die Daten auf Monats- bzw. auf Jahresbasis halten. Diese zusätzlichen Faktentabellen werden ebenfalls an die anderen Dimensionen angebunden, um sie möglichst transparent in der Auswertung verwenden zu können.

Diese Methode ist sehr gut dafür geeignet, eine saubere Struktur innerhalb der Faktentabellen zu erhalten. Liegt der Fokus auf der Performance, ist es häufig interessant, während des ETL-Prozesses bereits Berechnungen auszuführen. Unter diesen Umständen bietet es sich an, auf das Fact Constellation-Schema zurückzugreifen.

## 2.7. Typisierte Fakten

Wie bei der einfachen Star-Modellierung kennengelernt, können Kennzahlen in sehr vielen unterschiedlichen Ausprägungen auftreten und dadurch die Faktentabelle sehr breit werden lassen. Für viele unterschiedliche Ausprägungen (Typen) derselben Kennzahl bietet sich daher die Modellierung von Typisierten Fakten an. Diese Methode ging aus Ideen von Mitarbeitern der Infomotion GmbH, Frankfurt a. M., und Ausführungen Kimballs (vgl. [KiRo02], S. 109f. u. S. 178f.) hervor. Sie wurde durch die Autoren weiter ergänzt und ausgearbeitet, indem eine konkrete Spalte ergänzt wurde, die den Typ explizit angibt.



AuftragId	DatumId	Menge	Stunden	Kennzahl	Typ
1	2	0	5	Fahrtzeit	Berechnet
1	2	12	0	Material	Kulanz
2	12	0	2	Arbeitszeit	Berechnet
2	12	2	0	Material	Berechnet

Tab. 3: Typisierte Fakten

Bereits auf den ersten Blick wird bei dieser Modellierungsmethode ersichtlich, dass sich die Breite der Faktentabelle enorm reduzieren lässt. Sie hat nun eine feste Breite und sofern keine Messgrößen bei der Modellierung vergessen wurden, muss die Struktur der Tabelle im Nachhinein nicht mehr geändert werden. Für jede Messgröße (Menge, Stunden, Kosten etc.) ist dabei eine eigene Spalte vorgesehen, um die Additivität zu gewährleisten.

### 2.8. Generische Fakten

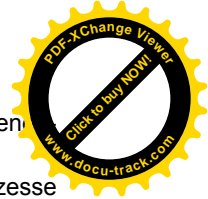
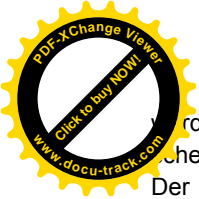
Sehr nah mit den Typisierten Fakten sind die Generischen Fakten verwandt. Ist es bei Ersteren noch so, dass für jede Messgröße eine eigene Spalte vorhanden sein muss, um die Additivität und die gleiche Granularität zu gewährleisten, so wird bei dieser Methode gegen diesen Ansatz gezielt verstoßen, um das Modell nochmals zu vereinfachen.

Diese Methode wurde aus der vorherigen abgeleitet. Sie als Weiterentwicklung zu bezeichnen wäre nicht ganz richtig, da sie einen anderen Ansatz verfolgt wie die Typisierten Fakten. Sie ist eher eine Zweigentwicklung, die durch einen neuen Aspekt geprägt wird. Im Gegensatz zu den Typisierten Fakten gibt es bei Generischen Fakten nur noch eine Spalte, die den Kennzahlenwert enthält. Dieser Wert wird durch drei Spalten (Kennzahl, Messgröße und Typ) spezifiziert.

AuftragId	DatumId	Wert	Messgröße	Kennzahl	Typ
1	2	5	Stunden	Fahrtzeit	Berechnet
1	2	12	Menge	Material	Kulanz
2	12	2	Stunden	Arbeitszeit	Berechnet
2	12	2	Menge	Material	Berechnet

Tab. 4: Generische Fakten

Der größte Vorteil an diesem Modell ist die Flexibilität in Bezug auf unterschiedliche Kennzahlausprägungen (berechnet, Kulanz, Gewährleistung etc.) und Messgrößen (Stunden, Menge, Kosten etc.). Diese lassen sich zu jeder Zeit erweitern und verändern, ohne dass die Struktur der Tabelle angepasst



werden muss. Im Gegensatz zu den Typisierten Fakten besitzen die Genannten Fakten also eine absolute, feste Breite.

Der Nachteil ist jedoch, dass die Komplexität der verbundenen Prozesse (ETL, Reporting) sehr stark wächst, da einige Fallunterscheidungen notwendig werden. Im Gegensatz zu der Einfachen Starmodellierung aus Kapitel 2.4 werden mit dieser Modellierung mehr Datensätze erzeugt, da jeder Messwert in einem eigenen Datensatz gespeichert wird.

## 2.9. Faktendimensionen

Der letzte Ansatz zur Faktentabellenmodellierung, der an dieser Stelle vorgestellt wird, sind die sogenannten Faktendimensionen (vgl. [KiRo02], S. 269-271). Mit diesen lassen sich Fakten aus der Faktentabelle über eine spezielle Dimension weiter definieren. Dadurch hat man die Möglichkeit, Faktenausprägungen als eigene Dimension darzustellen und somit deren gesamte Funktionalität zu erhalten.

AuftragId	DatumId	FaktDimId	Wert
1	11	2	65
2	14	4	95
3	15	6	5

Tab. 5: Faktendimension – Faktentabelle

FaktDimId	Kennzahl	Typ	Messgröße
1	Fahrtzeit	Berechnet	Stunden
2	Fahrtzeit	Gewährleistung	Stunden
3	Fahrtzeit	Kulanz	Stunden
4	Arbeitszeit	Berechnet	Stunden
5	Arbeitszeit	Gewährleistung	Stunden
6	Arbeitszeit	Kulanz	Stunden
7	Material	Berechnet	Menge
8	Material	Gewährleistung	Menge
9	Material	Kulanz	Menge

Tab. 6: Faktendimension – Faktendimensionstabelle

Deutlich zu erkennen ist bereits, dass die Faktentabelle nur noch sehr schmal ist und in ihrer Struktur auch bei Änderungen der Kennzahlen bestehen bleibt. Mit feststehenden und sich nicht oder selten ändernden Kennzahlen ist dies eine sehr vorteilhafte Modellierung, da sie sehr wenig Platz in der Faktentabelle verbraucht. Sie eignet sich außerdem dafür, Kennzahlen, die in unterschiedlichen Ausprägungen gleichzeitig vorkommen, darzustellen.



dem gezeigten Beispiel wurde bereits der Ansatz der Generischen Faktendimensionen aufgegriffen. Dies führt zu einer sehr flexiblen Nutzung der Faktendimensionen.

### 3 Evaluation

Die zuvor vorgestellten Faktenmodellierungsmethoden eignen sich alle für bestimmte Anwendungsfälle. Es gibt weder *die* Modellierung, nach der man sich immer richten kann, noch muss bei jedem Projekt das Rad neu erfunden werden. Daher werden die Methoden nachfolgend anhand eines Anforderungskatalogs untersucht und beurteilt. Dies soll dabei helfen, die Stärken und Schwächen jeweils besser einzuschätzen, um bei der täglichen Arbeit in der Entwicklung begründete Entscheidungen für oder wider bestimmten Ansätzen treffen zu können.

#### 3.1. Anforderungskatalog

Der Anforderungskatalog definiert alle Anforderungen, die bei der Evaluation berücksichtigt werden. Dabei sind folgende fünf Kriterien eine Auswahl aus einer Vielzahl von Möglichkeiten, besitzen allerdings nach Meinung der Autoren eine hohe Relevanz für die Entwicklung einer Business Intelligence-Umgebung:

- Datenverarbeitungsleistung
- Komplexität
- Anpassbarkeit
- Datenvolumen
- Reportingfähigkeit

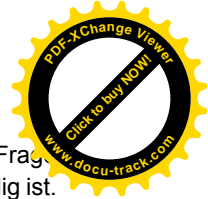
##### *Datenverarbeitungsleistung*

Die Datenverarbeitungsleistung betrifft die Antwortzeit. Da in der Regel enorme Datenmengen verarbeitet werden, muss ein Ziel sein, die Last möglichst gering zu halten, um ein schnelles Antwortverhalten beim Anfordern eines Reports zu erzielen. Eine Richtlinie zur Beurteilung der Leistung ist die DIN-Norm 66273 zur „Messung und Bewertung der Leistung von DV-Systemen“ (vgl. [DIfN91]).

##### *Komplexität*

Bei der Komplexität liegt der Fokus auf den einzelnen Komponenten, die mit der Lösung verbunden sind. Sie werden in der Konzeption geplant, später implementiert und ggf. in der Wartung angepasst. Eine hohe Komplexität ist mit einem größeren Zeitaufwand verbunden, da das Zusammenwirken der verschiedenen Komponenten geplant und nachvollzogen werden muss.





### **Anpassbarkeit**

Ein weiterer Punkt ist die Anpassbarkeit der Lösung. Dahinter steckt die Frage nach dem Aufwand, der für die Reaktion auf neue Bedingungen notwendig ist. Da Business Intelligence-Umgebungen keine statischen Anwendungen sind, entwickeln sich diese ständig weiter, sobald die Anwender neue Anforderungen an das System stellen.

### **Datenvolumen**

Trotz sinkender Speicherpreise können Datenvolumen anfallen, die sich mit heutigen Mitteln nicht anforderungsgerecht verarbeiten lassen. Vor allem Umgebungen, die eine enorm große Datenmenge in kurzer Zeit generieren, bereiten Probleme bei der zeitnahen Auswertung, da sie rechenintensiv aggregiert und gefiltert werden müssen.

### **Reportingfähigkeit**

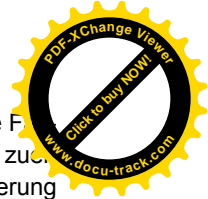
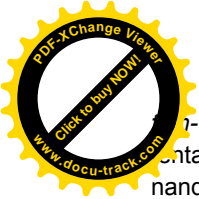
Schließlich grenzt die Wahl einer bestimmten Modellierung bereits die spätere Reportingfähigkeit ein. Hier muss darauf geachtet werden, dass die Anforderungen sich bestmöglich mit den Möglichkeiten decken. Ein sonst optimal abgestimmtes Datenmodell ist nutzlos, wenn es nicht in der Lage ist, Daten in der Form anzubieten, die der Endanwender benötigt.

## **3.2. Vorgehen bei der Beurteilung**

Die vorgestellten Methoden eignen sich jeweils für ganz bestimmte Anwendungsfälle. Zu deren Beurteilung wird im Folgenden ausgehend vom Anforderungskatalog dargestellt, welche Methoden sich bei der Fokussierung auf diesen Punkt anbieten. Dabei ist es wichtig, sich über die Ziele im Klaren zu sein. Natürlich ist es stets erstrebenswert, möglichst alle Disziplinen optimal zu erfüllen. Allerdings führen wirtschaftliche und technische Rahmenbedingungen in der Regel dazu, dass der Fokus auf einen bestimmten Aspekt gelegt werden muss. So möchte man zwar eine hohe Datenverarbeitungsleistung und gute Anpassbarkeit erreichen, allerdings können diese Bestrebungen in Konflikt zueinander stehen. Ist man sich darüber bewusst, welche Priorität die einzelnen Kriterien haben, lässt sich eine bessere Gewichtung der Vor- und Nachteile erzielen. Die folgende Beurteilung sollte unter diesen Kriterien interpretiert werden.

## **3.3. Datenverarbeitungsleistung**

Möchte man eine hohe Datenverarbeitungsleistung erzielen, so ist es empfehlenswert, bereits während des ETL-Prozesses aggregierte Tabellen zu erstellen, auf die im Reporting sehr schnell zugegriffen werden kann, ohne dass Massen an atomaren Daten verarbeitet werden müssen. Das *Fact Constella-*



*n*-Schema bietet hierfür einen sehr guten Ansatz. Es werden mehrere Faktentabellen unterschiedlicher Granularität verwendet, die in Beziehung zueinander stehen. Allerdings muss an dieser Stelle bereits bei der Modellierung eine Entscheidung für die späteren Reportingszenarien getroffen werden. Es ist möglich, für alle Aggregationsstufen Faktentabellen anzubieten. Weiß man z. B., dass im Berichtswesen häufig kumulierte Monatsdaten benötigt werden, so kann dies berücksichtigt werden. Allerdings bringt dies keine Vorteile bei der Analyse auf Wochenbasis (hingehen auf Quartalsbasis schon).

Die *Einfache Starmodellierung* kann ihre Schlichtheit mithilfe einer spaltenorientierten Datenbank ebenfalls um eine sehr hohe Datenverarbeitungsleistung ergänzen. Eine zeilenorientierte Datenbank kann dennoch eine höhere Geschwindigkeit erzielen, wie sie in gefilterten oder verknüpften Modellen erreicht werden kann. Allerdings hängt diese Aussage sehr stark von der Umgebung und den Systemen ab. Grundsätzlich ist die Verbundoperation (Join) kein Performanceproblem mehr in aktuellen Datenbanken (vgl. [KiRo02], S. 40).

### 3.4. Komplexität

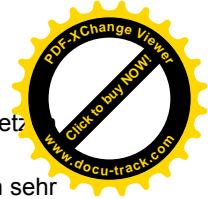
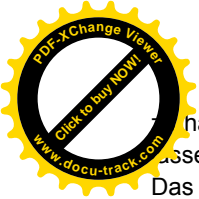
Komplexität erhöht den Wartungsaufwand, führt zu längeren Entwicklungszeiten und kostet somit Geld. Daher sollte bei jedem Projekt darauf geachtet werden, diese Komplexität möglichst gering zu halten. Hier gilt die Regel „So komplex wie nötig, so einfach wie möglich“.

Die geringste Komplexität kann die *Einfache Starmodellierung* aufweisen. Sie entwickelt sich nur in die Breite und wird durch die normalen Dimensionen ergänzt. Dieser Standardfall ist für jeden Entwickler sofort nachvollziehbar und kann sehr einfach ergänzt werden. Allerdings kann bei einer sehr breiten Tabelle mit der Zeit die Übersichtlichkeit schwinden. Bei mehreren Hundert Spalten kann es schnell passieren, dass neue Anforderungen zu einer neuen Spalte führen, obwohl deren Werte sich bereits in einer bestehenden befinden oder sich von ihr ableiten lassen.

Eine Methode, die mit sehr geringem Zusatzaufwand neue Funktionalität in ein Datenmodell bringen kann, sind *Degenerierte Dimensionen*. Durch eine einzelne Spalte lassen sich so beispielsweise Hierarchien und Gruppen abbilden. Durch eine strukturierte Nomenklatur lassen sich diese Spalten auch von den üblichen Dimensionen einfach abgrenzen, um so die Übersicht nicht zu gefährden.

### 3.5. Anpassbarkeit

Mit der Zeit wachsen und ändern sich auch die Anforderungen an ein System. Um nicht regelmäßig dazu gezwungen zu sein das Datenmodell anzupassen, sind flexible Datenmodelle wichtig. Sie erlauben es, die Daten so generisch



halten, dass sich auch neue Anforderungen mit der Umgebung umsetzen lassen.

Das Modell der *Generischen Fakten* lässt dem Entwickler diesbezüglich sehr viele Freiheiten. Es gibt eine Struktur vor, die unterschiedliche Werte halten und bereitstellen kann. So lässt sich diese sehr einfach um neue Kennzahlen erweitern, ohne dass das Schema verändert werden muss.

Eine noch flexiblere Modellierung bieten die *Faktendimensionen*. Änderungen müssen in der Faktendimension an einem Datensatz vorgenommen werden und sind fortan für alle verknüpften Daten gültig. Dieser Vorteil ist auch aus der Normalisierung transaktionaler Systeme bekannt.

Die Auswirkungen auf bereits bestehende Reports sind ebenfalls nur minimal. Da mittels eines künstlichen Schlüssels referenziert wird, ändern sich diese bei Anpassungen nicht und bleiben bestehen. Änderungen an den textlichen Beschreibungen können nur Probleme verursachen, wenn auf diese explizit gefiltert wurde.

### **3.6. Datenvolumen**

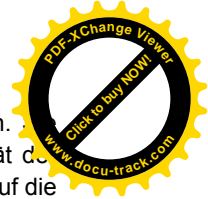
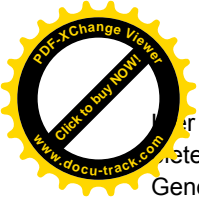
Eng mit der Datenverarbeitungsleistung verwandt ist das Datenvolumen. Ein möglichst geringes Datenvolumen führt zu einer schnellen Verarbeitung der Daten (allerdings nicht zwingend zu einem größeren Durchsatz).

Die effizienteste Methode ist es, wie bereits bei der Datenverarbeitungsleistung erwähnt, mehrere Faktentabellen mittels *Fact Constellation-Schema* zu verwenden. Hier fällt nicht nur die Aggregation weg (was zu einer hohen Datenverarbeitungsleistung führt), sondern es entfallen auch enorm viele Datenmengen. Anstatt großer Mengen atomarer Daten zu übertragen, müssen nur noch aggregierte Einzelwerte übermittelt werden.

Das Datenvolumen in *Faktendimensionen* ist gemischt. Ist die Faktendimension selbst im Vergleich zur eigentlichen Faktentabelle klein, so lassen sich Einsparungen erzielen. Die redundanten und häufig mit Beschreibungsmerkmalen ausgestatteten Faktendimensionen befinden sich getrennt in einer eigenen Tabelle. Die große Faktentabelle enthält hingegen nur einen einzigen Verweis auf diese Dimension. Würde man die Informationen der Faktendimension gemeinsam mit dem Faktum abspeichern, müssten sehr viel mehr Daten übertragen werden.

### **3.7. Reportingfähigkeit**

Wie bereits an anderen Stellen erwähnt, schränkt jede Entscheidung für eine Modellierung die späteren Reportingmöglichkeiten bereits ein. Ein Modell, das eine möglichst große Flexibilität beibehält, sollte in sehr variablen Anforderungsszenarien bevorzugt werden. Um die Daten auch für Anwender einfach auswertbar bereitzustellen, sollte die Komplexität des Reportings ebenfalls gering gehalten werden.



Man kann vor allem die *Typisierten Fakten* ihren Vorteil ausspielen. Sie bieten flexible Auswertungsmöglichkeiten, ohne jedoch die Komplexität der Generischen Fakten mit sich zu bringen. Bei ihnen muss nicht explizit auf die Additivität geachtet werden, da sich die entsprechenden Werte im Gegensatz zu den Generischen Fakten in unterschiedlichen Spalten befinden. Diese Regeln können zwar je nach Reportingumgebung bereits durch den Entwickler implementiert werden, allerdings erhöhen sie in der Regel die Komplexität des Berichtswesens an sich.

*Faktendimensionen* verlangen einen etwas anderen Ansatz in der Analyse. Man arbeitet nicht mehr mit Objekten wie „Fahrzeit“, „Materialmengen“ oder „Arbeitskosten“, sondern es existiert ein generisches „Wert“-Objekt, welches über Dimensionen gesteuert werden kann. So kann man das Objekt „Wert“ auswerten, indem man es zum Beispiel auf die Dimension „Fahrzeit“ und die Ausprägung „Kulanz“ einschränkt. Auf diese Weise lässt sich ein sehr intuitives und flexibles Reporting erzielen.

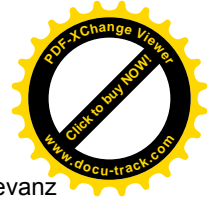
### **3.8. Kombinationsmöglichkeiten**

Mit einer Modellierungsmethode lassen sich in der Regel nicht alle Anforderungen gerecht umsetzen. Von daher ist es notwendig, im Datenmodell unterschiedliche Methoden sinnvoll zu kombinieren, um eine möglichst hohe Flexibilität und Erfüllung der Anforderungen zu erreichen.

Durch das Ergänzen einfacher Methoden wie Degenerierten Dimensionen oder des Fact Constellation-Schemas lassen sich in bestehende Modelle ohne großen Aufwand neue Funktionen einbinden.

Eine Möglichkeit ist es, die Einfache Starmodellierung mit dem Fact Constellation-Schema zu kombinieren. Auf diese Weise erhält man einfach strukturierte Faktentabellen, die Daten unterschiedlicher Granularität enthalten. Diese Kombination ermöglicht einen äußerst lauffzeiteffizienten Zugriff auf die Fakten und sie lässt sich sehr gut skalieren, indem neue Faktentabellen ergänzt werden.

Eine andere Modellierung ist auf möglichst hohe Anpassbarkeit an zukünftige Anforderungen ausgelegt. Mit Hilfe von Generischen Fakten und Faktendimensionen, die die generischen Kennzahlen spezifizieren, lassen sich sehr dynamische Strukturen schaffen. Dieses Modell kann sogar für unterschiedliche Anwendungsfälle verwendet werden, da der Typ einer Kennzahl allein durch die Faktendimension bestimmt wird.



## Zusammenfassung und Ausblick

Das Ziel dieses Artikels ist es, dem Leser ein Verständnis für die Relevanz der Modellierung von Fakten in Data Warehouse-Systemen zu vermitteln und ihm Anhaltspunkte zu geben, anhand derer er in seinen eigenen Data Warehouse-Projekten Entscheidungen für oder wider bestimmten Ansätzen treffen kann. Dadurch ist er in der Lage, selbstständig auf Basis seiner jeweils individuellen Umgebung Kombinationen zu bilden, die seine Anforderungen möglichst genau erfüllen.

Man wird häufig auf Konflikte treffen, die eine gewisse Menge an Kompromissen erfordert. Es kann kein einfacher Leitfaden vorgegeben werden, nach dem sich ein Entwickler richten kann. Allerdings hilft es bei der Orientierung, sich über Möglichkeiten, Vor- und Nachteile bewusst zu sein. Auf dieser Basis lassen sich begründete und nachhaltige Entscheidungen treffen.

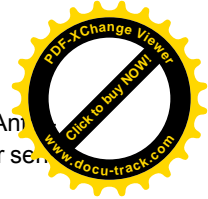
In Kapitel 3 wurden die zentralen Evaluierungsergebnisse textlich erläutert. Die nachfolgende Tabelle stellt nochmals die vollständigen Untersuchungsergebnisse vergleichend gegenüber.

	<i>Einf. Starmod.</i>	<i>Typ. Fakten</i>	<i>Gen. Fakten</i>	<i>Faktendimensionen</i>	<i>Deg. Dimensionen</i>	<i>F. C. Schema</i>
DV-Leistung	◐	◐	◐	◐	⊘	●
Komplexität	●	◐	○	◐	●	○
Anpassbarkeit	○	◐	◐	●	⊘	○
Datenvolumen	○	○	◐	◐	⊘	●
Reportingfähigkeit	○	●	◐	●	◐	◐

Tab. 7: Vergleichende Evaluierungsergebnisse

### Legende:

Grafik	Beschreibung
⊘	Kriterium trifft nicht zu
○	Merkmale vorhanden, aber keine Eignung
◐	Merkmale vorhanden und mäßige Eignung
◑	Merkmale vorhanden und mittlere Eignung
●	Merkmale vorhanden und gute Eignung



In diesem Artikel vorgestellten Methoden sind aus Erfahrungen und Anforderungen der letzten Jahrzehnte entstanden. Mit ihnen lassen sich daher sehr gut klassische Anforderungen an das Berichtswesen modellieren.

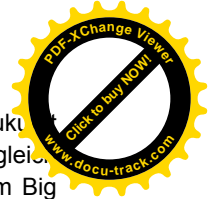
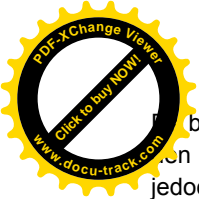
Aktuell wird unter dem Schlagwort *Big Data* (vgl. [BaGr12]) zunehmend die Entwicklung diskutiert, dass immer mehr potenziell auswertbare Daten in einem Unternehmen anfallen und dies häufig mit einer enormen Granularität verbunden ist. Mit RFID, Maschinensensoren oder Beiträgen in sogenannten sozialen Netzwerken lassen sich nahezu alle Zustandsänderungen einer Umwelt festhalten und in aggregierter Form auswerten.

In diesen Szenarien stoßen klassische Data Warehouse-Umgebungen an ihre Grenzen. Sie sind trotz leistungsfähiger Hardware nicht dafür ausgelegt, derartige Datenmengen effizient zu verarbeiten. Auf diese Lücke wird durch neue Technologien wie *Apaches Hadoop* (vgl. [o.V.12a] und [Wart12]) und weitere Ansätze (wie z. B. *In-Memory Data Management*, vgl. [PIZe11]) zunehmend eingegangen. Diese bieten Lösungen für massiv-parallele und verteilte Abfragen. Eigens entwickelte Algorithmen (z. B. *MapReduce von Google* (vgl. [o.V.12b])) sorgen für eine effiziente Steuerung dieser Prozesse.

Die großen Datenbankanbieter wie Oracle und Microsoft bieten Lösungen an, wie sich ihre Datenbanken mitsamt den Datenmodellen mit *Hadoop* kombinieren lassen (vgl. [ShCD10]). Dadurch lassen sich viele Aspekte, die in diesem Artikel genannt wurden, auch auf solche Umgebungen übertragen. Hier spielt der Aspekt *Datenverarbeitungsleistung* zwar nur eine geringe Rolle, aber die Auswirkungen auf die Analysewerkzeuge müssen weiterhin beachtet werden. Gerade bei riesigen Datenmengen besitzen effiziente Datenstrukturen eine wichtige Rolle. Die Vorteile einer schnellen Datenabfrage im Hintergrund würden durch längere Analysezeiten schnell zunichte gemacht.

Abseits von den bisher genannten Datenquellen gibt es jedoch zusätzliche, die in der Vergangenheit kaum eine Rolle gespielt haben. Sogenannte „unstrukturierte Daten“ aus Telefongesprächen, Bildern und Nachrichten oder sogar Videos bieten großes Potenzial für die Zukunft. Das Erschließen dieser Datenquellen übersteigt jedoch die Möglichkeiten bisheriger ETL-Umgebungen.

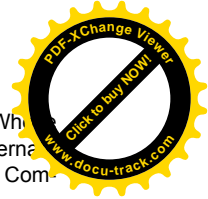
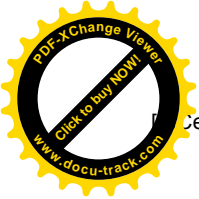
Diese zusätzlichen Datenquellen lösen aber keine bestehende Data Warehouse-Umgebung ab. In der Regel können sie ergänzend zu den bisherigen Datenbeständen genutzt werden. Für sie eignen sich andere Speicher- und Analysewerkzeuge besser. Eine Möglichkeit bieten hier *NoSQL*-Datenbanken (vgl. [Ed++11]), die in ihrer Datenhaltung wesentlich flexibler sind als relationale Datenbanken, die den Data Warehouse-Markt dominieren. Auf diese Szenarien können die hier wiedergegebenen Aspekte nur begrenzt angewendet werden.



bleibt festzuhalten, dass die klassischen Modelle auch in naher Zukunft einen Großteil der betrieblichen Analyseanforderungen abdecken, zugleich jedoch die sich abzeichnenden neuen Forschungsergebnisse aus dem Big Data-Umfeld sinnvoll und ergänzend zu den bestehenden Data Warehouse-Umgebungen eingesetzt werden sollten.

## Literatur

- [AbMH08] *Abadi, Daniel; Madden, Samuel; Hachem, Nabul*: Column-Stores vs. Row-Stores: How Different Are They Really? In: ACM SIGMOD'08, Vancouver, Juni 2008, S. 1-14.
- [BaGü09] *Bauer, Andreas; Günzel, Holger (Hrsg.)*: Data Warehouse Systeme: Architektur, Entwicklung, Anwendung. 3. überarbeitete und aktualisierte Aufl., Heidelberg: dpunkt-Verl., 2009.
- [BaGr12] *Bange, Carsten; Grosser, Timm*: Big Data – BI der nächsten Generation. In: Computerwoche, 9, 2012, S. 24-26.
- [Beck03] *Becker, Bob*: Kimball Design Tip #46: Another Look At Degenerate Dimensions. Kimball Group, 2003. Online: <http://www.kimballgroup.com/html/designtipsPDF/DesignTips2003/KimballDT46AnotherLook.pdf> (Abruf am 2012-05-21).
- [Bößw10] *Bößwetter, Daniel*: Spaltenorientierte Datenbanken. In: Informatik-Spektrum, 1, 2010, S. 61-65.
- [DifN91] *Deutsches Institut für Normung*: DIN 66273 – Messung und Bewertung der Leistung von DV-Systemen. Beuth Verlag, November 1991.
- [Ed++11] *Edlich, Stefan; Friedland, Achim; Hampe, Jens; Brauer, Benjamin; Brückner, Markus*: NoSQL – Einstieg in die Welt nichtrelationaler Web 2.0 Datenbanken. 2. Aufl., München: Carl Hanser Verl., 2011.
- [Fark11] *Farkisch, Kiumars*: Data-Warehouse-Systeme kompakt. Heidelberg: Springer Verl., 2011.
- [KeBM10] *Kemper, Hans-Georg; Baars, Henning; Mehanna, Walid*: Business Intelligence – Grundlagen und praktische Anwendungen. 3. überarbeitete und aktualisierte Aufl., Wiesbaden: Vieweg + Teubner Verl., 2010.
- [KiCa04] *Kimball, Ralph; Caserta, Joe*: The Data Warehouse ETL Toolkit: Practical Techniques for Extracting, Cleaning, Conforming, and Delivering Data. Indianapolis, Ind.: Wiley Publ., 2004.
- [KiRo02] *Kimball, Ralph; Ross, Margy*: The Data Warehouse Toolkit – The Complete Guide to Dimensional Modeling. 2. Ed., [Nachdr.]. New York [u.a.]: Wiley, 2002.
- [PIZe11] *Plattner, Hasso; Zeier, Alexander*: In-Memory Data Management, Heidelberg [u.a.]: Springer Verl. 2011
- [ShCD10] *Shankar, Shrikanth; Choi, Alan; Dijcks, Jean-Pierre*: Integrating Hadoop Data with Oracle Parallel Processing – An Oracle White Paper. Oracle Corporation, 2010. Online: <http://www.oracle.com/ocom/groups/public/@otn/documents/webcontent/130063.pdf> (Abruf am 2012-04-13).

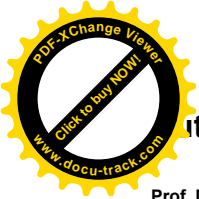


- [Ce05] *Stonebraker, Michael; Cetintemel, Ugur*: „One Size Fits All“: An Idea Whose Time Has Come and Gone. In: ICDE '05: Proceedings of the 21st International Conference on Data Engineering. Washington, DC, USA: IEEE Computer Society, 2005, S. 2–11.
- [o.V.12a] o. V.: Apache Hadoop-Homepage. The Apache Software Foundation, 2012. Online: <http://hadoop.apache.org/> (Abruf am 2012-03-19).
- [o.V.12b] o. V.: Introduction to Parallel Programming and MapReduce, Google Code University, 2012. Online: <http://code.google.com/intl/de-DE/edu/parallel/mapreduce-tutorial.html> (Abruf am 2012-05-21).
- [Wart12] *Wartala, Ramon*: Hadoop – zuverlässige, verteilte und skalierbare Big-Data-Anwendungen. München: Open Source Press, 2012.
- [WiMW09] *Wimmer, Rudolf; Meissner, Jens O.; Wolf, Patricia (Hrsg.)*: Praktische Organisationswissenschaft. Heidelberg: Carl-Auer-Systeme Verl., 2009.

## Kontakt

Prof. Dr. Harald Ritz  
Christoph Arnold, B. Sc. Wirtsch.-Inform.  
Technische Hochschule Mittelhessen (THM)  
Campus Gießen, Fachbereich MNI  
Wiesenstraße 14, 35390 Gießen  
T +49 641 309-2431  
[harald.ritz@mni.thm.de](mailto:harald.ritz@mni.thm.de)  
[kontakt@christoph-arnold.net](mailto:kontakt@christoph-arnold.net)



**Prof. Dr. Wolfgang Alm**

Hochschule Aschaffenburg  
Information Management Institut  
Labor für Informations- und Wissens-  
bewertungssysteme  
Würzburger Str. 45, 63743 Aschaffenburg  
T +49 6021 4206-700  
wolfgang.alm@h-ab.de

**Christoph Arnold, B. Sc. Wirtsch.-Inform.**

Technische Hochschule Mittelhessen (THM)  
Campus Gießen, Fachbereich MNI  
Wiesenstraße 14, 35390 Gießen  
kontakt@christoph-arnold.net

**Yves Augustin**

Hochschule Fulda  
Fachbereich Angewandte Informatik  
Marquardstraße 35, 36039 Fulda

**Prof. Dr. Thomas Barton**

Fachhochschule Worms  
Erenburgerstr. 19, 67543 Worms  
T +49 6241 509-253  
barton@fh-worms.de

**Yvonne Bentz**

Fachhochschule Worms und  
Infocient Consulting GmbH  
Marie-Bernays-Platz 3, 68309 Mannheim  
yvonne.bentz@infocient.de

**Prof. Dr. Bettina C. K. Binder**

Controlling und strategische Unternehmensführung  
Hochschule Pforzheim  
Tiefenbronner Straße 65, 75175 Pforzheim  
T +49 7231 28-6682  
bettina.binder@hs-pforzheim.de

**Prof. Werner Burkard**

Hochschule Pforzheim  
Tiefenbronner Straße 65, 75175 Pforzheim  
T +49 7231 28-6693  
werner.burkard@hs-pforzheim.de

**Roberto Buscemi**

Hochschule Pforzheim  
Tiefenbronner Straße 65, 75175 Pforzheim

**Sven Denecken**

SAP AG  
Dietmar-Hopp-Allee 16, 69190 Walldorf  
sven.denecken@sap.com, Twitter: @SDenecken

**Dr. Armin Elbert**

Infocient Consulting GmbH  
Marie-Bernays-Platz 3, 68309 Mannheim  
armin.elbert@infocient.de

**Martin Gaubitz**

Hochschule Pforzheim  
Tiefenbronner Straße 65, 75175 Pforzheim

**Prof. Dr. Andreas Heberle**

Hochschule Karlsruhe  
Technik und Wirtschaft  
Fakultät für Informatik und Wirtschaftsinformatik  
Moltkestr. 30, 76133 Karlsruhe  
T +49 721 925-2969  
Andreas.Heberle@hs-karlsruhe.de

**Prof. Dr.-Ing. Frank Herrmann**

Hochschule Regensburg  
Innovationszentrum für Produktionslogistik  
und Fabrikplanung (IPF)  
Postfach 120327, 93025 Regensburg  
T +49 941 943-1307  
frank.herrmann@hs-regensburg.de

**Johannes Hoen**

INFOSERVE GmbH  
Am Felsbrunnen 15, 66119 Saarbrücken  
j.hoen@infoserve.de

**Prof. Dr. Georg Rainer Hofmann**

Hochschule Aschaffenburg  
Information Management Institut  
Labor für Informations- und Wissens-  
bewertungssysteme  
Würzburger Str. 45, 63743 Aschaffenburg  
T +49 6021 4206-700  
georg-rainer.hofmann@h-ab.de

**Pezzotta Jeff**

Fachhochschule Trier  
FB Wirtschaft, FR Wirtschaftsinformatik  
Organisation und Informationsmanagement  
Schneidershof, 54293 Trier

**Prof. Dr. Norbert Ketterer**

Hochschule Fulda  
Fachbereich Angewandte Informatik  
Marquardstraße 35, 36039 Fulda  
T +49 661 9640-323  
norbert.ketterer@informatik.hs-fulda.de

**Oliver Kuchler**

Provadis, School of International Management  
and Technology AG  
Fachbereich Wirtschaftsinformatik,  
Prozessmanagement  
Industriepark Höchst, Gebäude B845,  
65926 Frankfurt am Main

**Prof. Dr. Elvira Kuhn**

Fachhochschule Trier  
FB Wirtschaft, FR Wirtschaftsinformatik  
Organisation und Informationsmanagement  
Schneidershof, 54293 Trier  
T +49 651 8103-382 oder -299 (Sekt.)  
e.kuhn@fh-trier.de

**Prof. Dr. Martin Kütz**

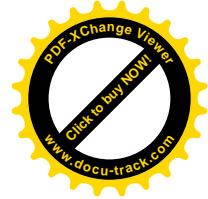
Hochschule Anhalt (FH)  
Fachbereich Informatik  
Postfach 1458, 06354 Köthen/Anhalt  
T +49 3496 67-3114  
martin.kuetz@inf.hs-anhalt.de

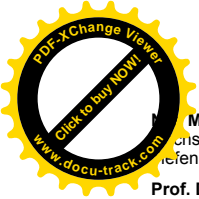
**Christian Ludwig**

Hochschule Pforzheim  
Tiefenbronner Straße 65, 75175 Pforzheim

**Prof. Dr. Frank Morelli**

Hochschule Pforzheim  
Tiefenbronner Straße 65, 75175 Pforzheim  
T +49 7231 28-6697  
frank.morelli@hs-pforzheim.de





### Mosbach

Hochschule Pforzheim  
Tiefenbronner Straße 65, 75175 Pforzheim

#### Prof. Dr. Christian Müller

Technische Hochschule Wildau  
Fachbereich Betriebswirtschaft/  
Wirtschaftsinformatik  
Bahnhofstraße, 15745 Wildau  
T +49 3375 508-956  
christian.mueller@th-wildau.de

#### Gerald Münzl

House of IT e.V. Darmstadt  
Mornwegstr. 30-32, 64293 Darmstadt  
muenzl@house-of-it.eu

#### Gervais Jocelyn Ndongmo

Fachhochschule Trier  
FB Wirtschaft, FR Wirtschaftsinformatik  
Organisation und Informationsmanagement  
Schneidershof, 54293 Trier

#### Prof. Dr. Rainer Neumann

Hochschule Karlsruhe  
Technik und Wirtschaft  
Fakultät für Informatik und Wirtschaftsinformatik  
Moltkestr. 30, 76133 Karlsruhe  
Rainer.Neumann@hs-karlsruhe.de

#### Djieukeng Ngoune Neylor

Fachhochschule Trier  
FB Wirtschaft, FR Wirtschaftsinformatik  
Organisation und Informationsmanagement  
Schneidershof, 54293 Trier

#### Prof. Dr. Kurt Porkert

Hochschule Pforzheim  
Tiefenbronner Straße 65, 75175 Pforzheim  
T +49 7231 28-6691  
kurt.porkert@hs-pforzheim.de

#### Dipl.-Informatikerin Ute Riemann, MBA

Business Enterprise Principal Consultant  
SAP Business Transformation Services

#### Prof. Dr. Harald Ritz

Technische Hochschule Mittelhessen (THM)  
Campus Gießen, Fachbereich MNI  
Wiesenstraße 14, 35390 Gießen  
T +49 641 309-2431  
harald.ritz@mni.thm.de

#### Prof. Dr. Peter Scholz

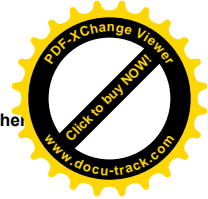
Hochschule Landshut  
Am Lurzenhof 1, 84036 Landshut  
peter.scholz@fh-landshut.de

#### Mathias Schröder

Hochschule Pforzheim  
Tiefenbronner Straße 65, 75175 Pforzheim

#### Dipl. Wirt.-Ing. Bert O. Schulze

SAP AG  
Dietmar-Hopp-Allee 16, 69190 Walldorf  
T +49 6227 7-66841  
bert.schulze@sap.com, Twitter: @BeSchulze



#### Dipl. Betriebswirtin (FH) Meike Schumacher

Hochschule Aschaffenburg  
Information Management Institut  
Labor für Informations- und Wissens-  
bewertungssysteme  
Würzburger Str. 45, 63743 Aschaffenburg  
T +49 6021 4206-700  
meike.schumacher@h-ab.de

#### Prof. Dr. Christian Seel

Hochschule Landshut  
Am Lurzenhof 1, 84036 Landshut  
christian.seel@fh-landshut.de

#### Prof. Dr. rer. nat. Carlo Simon

Provisdis, School of International Management  
and Technology AG  
Fachbereich Wirtschaftsinformatik,  
Prozessmanagement  
Industriepark Höchst, Gebäude B845,  
65926 Frankfurt am Main  
T +49 69 305-13278  
carlo.simon@provisdis-hochschule.de

#### Bruno Sticht

Constantia Flexibles  
Pirkmühle 14-16, 92712 Pirk  
bruno.sticht@constantia-hueck.com

#### Michael Tax, B.A.

Hochschule Aschaffenburg  
Information Management Institut  
Labor für Informations- und Wissens-  
bewertungssysteme  
Würzburger Str. 45, 63743 Aschaffenburg  
T +49 6021 4206-700  
michael.tax@h-ab.de

#### Prof. Dr. Stephan Thesmann

Hochschule Pforzheim  
Tiefenbronner Straße 65, 75175 Pforzheim  
T +49 7231-28-6095 (Sekretariat)  
stephan.thesmann@hs-pforzheim.de

#### Prof. Dr.-Ing. Heiko Thimm

Hochschule Pforzheim  
Professur für Quantitative Methoden und  
Informationstechnik  
Tiefenbronner Str. 65, 75175 Pforzheim  
T +49 7231 28-6451  
heiko.thimm@hs-pforzheim.de

#### Dr. Philipp Walter

INFOSERVE GmbH  
Am Felsbrunnen 15, 66119 Saarbrücken  
p.walter@infoserve.de

#### Wendelin Wiedeking

Hochschule Pforzheim  
Tiefenbronner Straße 65, 75175 Pforzheim

#### Patrick M. Wolff

Hochschule Pforzheim  
Tiefenbronner Straße 65, 75175 Pforzheim