

Technische Komponenten von Business Intelligence

Tamara Seitz
Matrikel Nr.: h0353839

LV 0402 Seminar aus Informationswirtschaft SS 07

LV-Leiter:

PD Dr. Michael Hahsler und o. Univ. Prof. Dr. Wolfgang Janko

Abteilung für Informationswirtschaft
Institut für Informationsverarbeitung und Informationswirtschaft

Wirtschaftsuniversität Wien

Augasse 2-6

A-1090 Wien, AUSTRIA

Telefon: ++43-31336-5231

Fax: ++43-31336-739

Technische Komponenten von Business Intelligence

Zusammenfassung

In dieser Arbeit wird auf die technischen Komponenten im Business Intelligence Prozess eingegangen. Im logischen Ablauf des Prozesses werden zuerst die Daten aus dem Operativsystem und diversen externen Quellen selektiert und gesammelt. Nach der Bereinigung und Aufbereitung werden sie im Data Warehouse gesammelt. Das Data Warehouse kann entweder als multidimensionales Modell (nach R. Kimball), oder als relationales Modell (nach W.H. Inmon) realisiert werden. Zur Darstellung der Daten werden in weiterer Folge entweder ein einfaches Reportsystem oder, bereits zur genaueren Analyse ein OLAP System verwendet. Wenn diese Darstellungen zur Informationsfindung nicht ausreichen, werden die Daten mit Hilfe von Data Mining-Verfahren weiter analysiert. Die gängigsten Aufgaben und Verfahren von Data Mining werden abschließend im Überblick dargestellt.

Stichworte: Business Intelligence, Data Warehouse, multidimensionale Datenmodell, relationale Datenmodell, Berichtswesen, OLTP, OLAP, Kennzahlensystem, Data Mining

Abstract

In this paper the technical components and their characteristics within the Business Intelligence Prozess are explained in detail. Within the logistical workflow of this process in a first step data from the operational System and diverse other sources are collected. With several selection and filter mechanism the data is cleaned. Prepared like this the data is stored in the data warehouse. The data warehouse can be realized either in a Multidimensional Modeling (developed by R. Kimball) or as a Relational Modeling (developed by W.H.Immon). For evaluation purposes simple reporting systems or for higher analysis more sophisticated systems for example OLAP reporting system are used to summarize the stored data from the data warehouse. If reporting systems are not sophisticated enough to find the relevant information for decision making, the stored data will be analysed furthermore with Data Mining procedures. The most popular procedures and functions of data mining are finally listed in an overview.

Keywords: Business Intelligence, Data Warehouse, multidimensional Modeling, relational Modeling, Reportingsystem, OLTP, OLAP, Performance Measurement System, Data Mining

Inhaltsverzeichnis

Inhaltsverzeichnis.....	3
1 Einführung.....	4
2 Allgemeines zu Business Intelligence.....	6
3 Business Intelligence Komponenten.....	9
3.1 ÜBERBLICK ÜBER DIE PHASEN DER INFORMATIONSVERRARBEITUNG.....	9
3.2 BEREITSTELLUNG VON DATEN.....	9
3.2.1 Das operative System.....	10
3.2.2 Das Data Warehouse.....	12
3.3 AUFBEREITUNG UND DARSTELLUNG VON INFORMATION.....	17
3.3.1 Kommunikation im Unternehmen.....	17
3.3.2 Reporting.....	18
3.3.3 OLAP.....	19
3.3.4 Kennzahlensysteme.....	23
3.4 ANALYSE VON INFORMATIONEN.....	26
3.4.1 Data Mining.....	26
4 Zusammenfassung.....	32
5 Literaturverzeichnis.....	33
6 Abbildungsverzeichnis.....	35

1 Einführung

Diese Arbeit beschäftigt sich mit dem Thema Business Intelligence im Allgemeinen. Es wird ein Überblick über die einzelnen Phasen und Komponenten von BI Systemen gegeben und versucht die Anwendung von bestimmten Methoden kurz darzustellen.

Die Arbeit beginnt mit einer allgemeinen Begriffsdefinition und einem Überblick über die Geschichte von Business Intelligence.

Es wird auf die Phasen von Business Intelligence im technischen Sinne eingegangen. Die drei Phasen Datenbereitstellung, Darstellung der Daten und Analyse der Daten werden im Einzelnen betrachtet.

Bei der Informationsbereitstellung werden, wie schon erwähnt, die Daten aus den verschiedenen Bereichen der Unternehmen gesammelt. Nicht selten ist auch das Bereitstellen der Daten, und das bereitet oftmals Probleme, unternehmensübergreifend notwendig. Das Sammeln der Daten geschieht meist in einem Data Warehouse. Die Aufbereitung der Daten aus im operativen System für das Data Warehouse und die damit verbundenen Grundgedanken, was zum Beispiel im Bereich der Performance beachtet werden muss, werden anschließend diskutiert. In diesem Zusammenhang werden weiters Data Warehouse Architekturen als Datenpool für Auswertungen betrachtet. Es gibt prinzipiell zwei verschiedene Möglichkeiten, wie ein Data Warehouse aufgebaut werden kann. Einerseits das multidimensionale Modell von Ralph Kimball, andererseits das relationale Modell von Bill Inmon, welche beide kurz vorgestellt werden vorgestellt. Beim multidimensionalen Modell werden die Unternehmensbereiche in verschieden Data Marts unterteilt und das Data Warehouse als so genanntes Star Schema modelliert. Im Gegensatz dazu sieht Bill Inmon das Data Warehouse als zentrales Objekt im Unternehmen, welches in seinem Modell als Corporate Information Factory ausgedrückt wird.

Der nächste Schritt im Aufbau eines Business Intelligence Systems ist die Aufbereitung und Darstellung von Information. In diesem Kapitel wird zuerst auf die Kommunikation im Unternehmen durch die Veränderungen durch Globalisierung und technischen Fortschritt eingegangen.

Es gibt einige Möglichkeiten der Informationen des BI Systems darzustellen. Einerseits existiert der Begriff Berichtswesen, welcher in weiterer Folge noch nach den Regeln zur Auslösung der Berichterstellung in aktive und passive, beziehungsweise in Ad Hoc Berichte unterteilt wird. Eine andere Möglichkeit neben dem Berichtswesen ist der so genannte OLAP Prozess (Online Analytical Process). Durch diese Theorie kann der Benutzer sehr interaktiv und ohne besondere (Programmier-)Vorkenntnisse die Daten im Data Warehouse ansehen und die relevanten Informationen auf sehr unterschiedliche Arten betrachten. Er kann sich beispielsweise entweder einen Gesamtüberblick der Daten verschaffen, oder zum Beispiel bis in den elementarsten Datensatz „zoomen“. OLAP ist eine intuitive und interaktive Methode, welche auch dem Benutzer, welcher über keine ausgeprägte IT-Erfahrung verfügt, viele Möglichkeiten lässt.

In der Arbeit wird nun OLAP im Allgemeinen kurz beschrieben, und darauf folgend mit seinen wichtigsten Regeln definiert. Die Grundoperationen Slicing, Dicing, Pivoting, Drill-Down and Roll-Up und Drill-Across and Drill-Through werden vorgestellt.

Eine weitere Darstellung von Daten im Unternehmen ist die Einteilung der Daten in Kennzahlen. Um diese für den Benutzer aufzubereiten werden gerne Methoden wie zum

Beispiel die Balance Score Card verwendet. Hierbei wird das Unternehmen aus vier verschiedenen Perspektiven betrachtet. Es gibt die Finanzperspektive, die Kundenperspektive, die Prozessperspektive und die Entwicklungsperspektive. Für diese Perspektiven werden verschiedene Kennzahlen angewandt, welche die Erreichung sowie den Fortschritt und Erfüllungsgrad der Ziele kontrolliert.

Für den Fall, dass Information mit Hilfe der bereits genannten Werkzeugen noch nicht ausreichend erkannt und dargestellt werden kann, werden diese in einem weiteren Schritt mit Hilfe analytischer Methoden ausgewertet. Die Analyse der Informationen umfasst die Ausarbeitung der gesammelten Daten in Bezug auf Menge, Inhalt und Verteilung. Hier werden mit Hilfe von Data Mining viele statistische Verfahren angewandt. Der Begriff Data Mining wird weiter unterteilt in Data Mining, Text Mining und Web Mining. Data- und Text Mining werden häufig unter dem Oberbegriff „Knowledge Discovery in Databases“, auch unter der Abkürzung KDD zu finden, geführt. Es handelt sich um prozessorientierte Komponenten zum Auffinden und Aufbereiten von Rohdaten, zum Entdecken von Wissen, sowie logischen Zusammenhängen dieser Daten. Während Data Mining sich jedoch vermehrt mit strukturierten Daten beschäftigt werden in Text Mining sämtliche Methoden vereinigt, um neues und relevantes Wissen zu finden, welches sich in großen unstrukturierten Textdokumenten befindet. Durch die schnelle Entwicklung und zunehmende Verbreitung des World Wide Web gewinnt auch das Web Mining immer mehr an Bedeutung. Nach diesem kurzen allgemeinen Überblick von Data Mining werden nun dessen verschiedenen Aufgaben und Methoden untersucht. Als Aufgaben unterscheiden sich: das Klassifikationsmodell, das Regressionsmodell, das Assoziationsmodell und das Clustering. Die dazugehörigen, gängigsten Verfahren sind: das Entscheidungsbaumverfahren, Bildung von künstlichen neuronalen Netzen, das Clusterverfahren und das Verfahren zur Assoziationsanalyse. Wobei alle diese Begriffe kurz vorgestellt werden. Abschließend werden in einem Fazit kurz die gewonnenen Erkenntnisse zusammengefasst.

Im 2. Kapitel wird der Begriff Business Intelligence kurz erläutert und sein geschichtlicher Verlauf dargestellt. Nach dieser allgemeinen Einführung in das Thema werden im 3. Kapitel die Phasen der Datenverarbeitung einerseits im Überblick dargestellt, und in weiterer Folge detailliert mit den zugehörigen technischen Werkzeugen beschrieben. Das Kapitel 3.1 beginnt mit einer allgemeinen Beschreibung der Phasen des Business intelligence Prozesses. Das Kapitel 3.2 beschäftigt sich mit der Bereitstellung von Daten, das Kapitel 3.3 führt mit der Darstellung der Information fort und im Kapitel 3.4 wird auf die Möglichkeit von weiteren Analysen eingegangen. Zum Schluss wird im Kapitel 4 die Ergebnisse dieser Arbeit zusammengefasst.

2 Allgemeines zu Business Intelligence

Die Veränderungen im Unternehmensumfeld und die Verschärfung der Wettbewerbsbedingungen auf Grund der weltweiten Marktöffnung bergen Chancen und Risiken für große und mittelständische Unternehmen [KeMU06]. In Zeiten, in denen Informationsbeschaffung und –speicherung sehr einfach geworden ist, wird versucht durch möglichst viel Information einen Wettbewerbsvorsprung zu erzielen.

Durch rapide wachsende Speicher- und Verarbeitungskapazitäten der EDV Systeme einerseits und der zunehmenden Umweltkomplexität andererseits kommt es zu einem Überfluss an Daten. Das Problem besteht darin, dass ohne deren Analyse das Unternehmen aus den Daten keinen Nutzen ziehen kann. Daraus ergibt sich paradoxerweise statt der erhofften Informationsflut vielmehr ein Mangel an Information. Denn je mehr Daten in einem Unternehmen anfallen, desto schwieriger wird es, auf die relevanten Daten zu fokussieren und desto größer ist die Gefahr, solche zu übersehen.

Der Einsatz betriebswirtschaftlicher Anwendungssysteme zur Unterstützung und Durchführung operativer Geschäftsprozesse führt oft dazu, dass Mitarbeiter verschiedener Abteilungen separate Datenbanken pflegen. Dazu kommt noch, dass, zwischen den Abteilungen oft ein ungenügender Informationsaustausch stattfindet, so dass einerseits möglicherweise ein Teil der Daten an mehreren Stellen redundant gespeichert wird – und dort jeweils gepflegt werden muss – und andererseits die Daten, die sich gegenseitig ergänzen könnten, nicht zueinander in Beziehung gesetzt werden, weil effiziente Zugriffsmöglichkeiten fehlen. Neben den unternehmensinternen Daten stehen in diesem Prozess auch oft noch externe Daten zur Verfügung.

Bereits in den 60er Jahren wurden erste Schritte in Richtung unterstützender Anwendungen für betriebliche Entscheidungsträger unternommen. Damalige Projekt hatten das Ziel, Management Information Systeme (MIS) zu erstellen, welche die nunmehr vorhandenen Daten der aufkommenden Dialog- und Transaktionssysteme in die Planungs- und Kontrollprozesse direkt zu integrieren. Diese Systeme führten aber auf Grund der damaligen technisch noch nicht so fortgeschrittenen Entwicklung bestenfalls zu einer Automatisierung des Standardberichtssystems[ChGl06]

Die darauf folgenden Decision Support Systeme (DSS) versuchten darum nicht allein Daten in Entscheidungs- und Planungsprozesse einfließen zu lassen, sondern zusätzlich Szenarien darzustellen und mögliche Lösungen zu untersuchen. Es wurde jedoch keine eigene, zentrale Datenhaltung angestrebt, sondern Modelle und Lösungen wurden von jeder der betroffenen Abteilung gesondert betrachtet und gespeichert. Das führte zu redundanter Datenhaltung und Aussagen und Ergebnissen, welche sich im Vergleich zwischen den Abteilungen widersprachen[ChGl06]

In den 80er Jahren wurden, nun auf Basis verbesserter Infrastruktur im Unternehmen und zunehmender Vernetzung, Executive Information Systems (EIS) entwickelt, welche auf die Unterstützung der obersten Managementebenen abzielten. Mit Hilfe neuartiger Anwendungen zur Kommunikationsunterstützung und Möglichkeiten zu ad hoc Abfragen, auch für ungeübte bzw. EDV-unerfahrene Benutzer wollten diese Systeme vor allem die oberste Managementebene unterstützen[ChGl06].

Das Data Warehouse bietet eine gemeinsame Datenbasis, nachdem die Daten aus den operativen Systemen gesammelt, bereinigt und vervollständigt wurden. Davon ausgehend

bietet es interaktive Abfragen und analytische Betrachtung der Daten mit Hilfe von Online Analytical Processes (OLAP) [ChGI06].

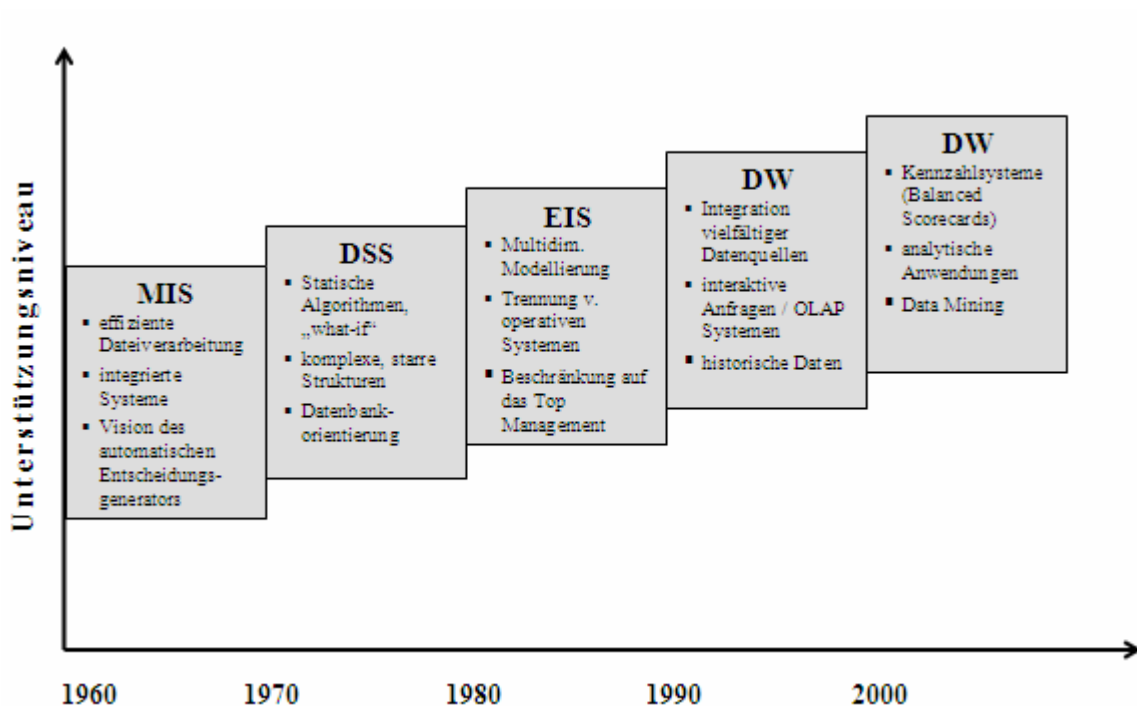


Abbildung 1: Entscheidungsunterstützungssysteme im Zeitverlauf nach [HuWi05]

In Abbildung 1 kann man die schematische Darstellung der Entwicklung des Business Intelligence Prozesses im Verlauf der Jahre sehen. Zusammenfassend kann man sagen, dass im Laufe der Zeit die Techniken und Möglichkeiten der entscheidungsunterstützenden Werkzeuge permanent verbessert wurden. Die Systeme lernten von ihren Vorgängern und wurden zunehmend erfolgreicher [HuWi05].

Der Begriff Business Intelligence wurde in den 90er Jahren von Howard Dresner, dem Urheber der Bezeichnung folgendermaßen definiert:

„all the technologies that help business make decision based on fact“ [Nylu99]

Während die Gartner Group, der Dresner in weiterer Folge im Jahr 1992 beitrug, nachstehende Definition hervorbrachte:

„Business Intelligence is the process of transforming data into information and, through discovery into knowledge.“ [TUMU05]

Es wird hier bewusst gemacht, dass mit dem Begriff Business Intelligence einerseits die technischen Aspekte andererseits aber auch die betriebswirtschaftlichen Prozesse in den Vordergrund gestellt werden können. Eine weitere Möglichkeit ist, Business Intelligence als

Prozess welcher die interaktiven Handlungen und Datenaustausch zwischen den einzelnen Bereichen bzw. Abteilungen des Unternehmens, mit dem vorrangigen Ziel optimale Information für den Wettbewerb zu gewinnen, bezeichnet:

„Business Intelligence (BI) bezeichnet den analytischen Prozess, der – fragmentierte - Unternehmens- und Wettbewerbsdaten in handlungsgerichtetes Wissen über Fähigkeiten, Positionen, Handlungen und Ziele der betrachteten internen oder externen Handlungsfelder (Akteure und Prozesse) transformiert.“[GRGe00]

An diesen verschiedenen Definitionen ist deutlich erkennbar, dass Business Intelligence ein breites Spektrum an Anwendungen und Technologien umfassen kann. Die Bezeichnung Business Intelligence kann vielseitig sein, und die Definition ob Business Intelligence im engeren Sinne, sprich auf rein technologischen Ebene, oder als Teil der Unternehmenspolitik, d.h. im weiteren Sinne gemeint ist, muss explizit abgegrenzt werden.

In dieser Arbeit wird vorrangig auf die technischen Aspekte von Business Intelligence eingegangen und anhand der Phasen der Informationsaufbereitung erläutert.

3 Business Intelligence Komponenten

In dieser Arbeit werden die einzelnen Komponenten von Business Intelligence im Zuge der betreffenden Phasen des Informationsverarbeitungsprozesses beschrieben.

3.1 Überblick über die Phasen der Informationsverarbeitung

Der Prozess der Informationsverarbeitung zur Erstellung einer Business Intelligence Umgebung im Unternehmen wird in der Literatur oft als Abfolge von mehreren, hintereinander auszuführenden Arbeitsschritten beschrieben, wie zum Beispiel in [GrBe99]. In dieser Arbeit wird von einem Prozess mit 3 Phasen ausgegangen:

1. Sammlung und Bereitstellung der Basisdaten und Integration der Daten
2. Verbreitung und Anwendung der Daten als Informationen
3. Weitere Auswertung der Daten, etwa durch OLAP oder Data Mining

Im ersten Schritt werden Daten, auf Grund ihrer Relevanz ausgewählt, aus den verschiedenen operativen Systemen des Unternehmens gesammelt. Anschließend werden die zur gemeinsamen Weiterverarbeitung vorbereitet und vereinheitlicht, auf Fehler und Vollständigkeit geprüft und schließlich in einem eigenem System, in der Regel einem Data Warehouse, gespeichert und eventuell verdichtet.

Diese, nun aus technischer Sicht homogenen, vollständigen Daten werden im zweiten Schritt zur Darstellung aufbereitet. Hier werden vorzugsweise graphische Oberflächen und Reportsysteme verwendet. Man kann auch OLAP zur Darstellung der Daten verwenden, dieses verfügt jedoch auch schon über weiterführende Möglichkeiten zur Analyse der Daten. Hiermit wird nun in der dritten Phase fortgesetzt, wobei die Daten nach Auffälligkeiten und interessanten Mustern untersucht werden. Diese Untersuchungen werden unter dem Begriff Data Mining zusammengefasst. Der Begriff Data Mining beinhaltet vielfältige Verfahren zur Analyse von Daten.

3.2 Bereitstellung von Daten

In der Praxis kommen unternehmensrelevante Daten meistens nicht aus einem zentralen System, sondern sind auf verschiedene Subsysteme, oftmals redundant, verteilt. Die Abfrage nach bestimmten Informationen ist dadurch schwierig realisierbar und fehlerbehaftet, da verschiedene Datenbasen unterschiedliche Ergebnisse liefern können.

Es gibt viele Wege, auf denen Daten ins Unternehmen fließen können. Der größte Teil an Daten, die im Unternehmen anfallen, entsteht intern, das heißt in den operativen Systemen. Diese Daten werden beim Durchlaufen der unterschiedlichen Prozesse, welche sich in erster Linie mit dem Tagesgeschäft beschäftigen, generiert. Es wird jedoch auch eine Vielzahl von Daten aus externen Quellen gesammelt, welche für das Unternehmen als Zusatzinformationen einen weiteren Informationsvorsprung bringen sollen. Diese gesammelten Daten sind durch die verschiedenen Quellsysteme und die unterschiedliche Erfassung und Speicherung nicht in einer einheitlichen Form. Einerseits unterscheiden sie sich in ihrer Qualität andererseits im Speicherformat. Außerdem kann es im Unternehmen durch teilweise redundante Datenhaltung zu Inkonsistenzen kommen. Weiters können die Daten in den operativen

Systemen auf Grund technischer oder menschlicher Fehler fehlerhafte Elemente enthalten. Diese Fehler sind für das operative System nicht von Bedeutung, haben aber im weiteren Business Intelligence Prozess negative Auswirkungen auf die Qualität der Analysen.

Aus diesem Grund wird in der Phase der Vorverarbeitung die Datenqualität des selektierten Datenpools untersucht und bearbeitet, bevor sie in einem einheitlichen Format in einem gemeinsamen System gespeichert werden [Ahle01].

3.2.1 Das operative System

Im operativen Bereich eines Unternehmens werden heutzutage zur Verwaltung der umfangreichen und ständig wachsenden Datenbeständen zumeist Datenbanksysteme eingesetzt. Sie sorgen für eine dauerhafte und fehlerrobuste Speicherung sowie einen effizienten Zugriff. Dieser Prozess lässt sich mit dem Begriff des Online Transactional Processing (OLTP) umschreiben. Bei der Informationsverarbeitung in operativen Systemen wird in der Regel aktuelle Detailinformation, die gemäß den Anforderungen der darauf zugreifenden operativen Anwendungssysteme abgespeichert sind, satzweise abgefragt. Die Verarbeitung von Transaktionen, das heißt Lese- und Schreiboperationen auf kurzfristig veränderlichen Datenbeständen, steht hierbei im Vordergrund. Es ist üblich, dass mehrere Benutzer sich derselben Systeme und Datenbestände bedienen, wie beispielsweise Auskunft-, Buchungs- oder Bestellsystemen [KeMU06].

OLTP-Anwendungen zeichnen sich dadurch aus, dass sie nur relativ kleine Datenmengen pro Transaktion zu verarbeiten haben und dass sie nur auf dem jüngsten, aktuell gültigen Zustand des Datenbestandes operieren. Die typische Datenstruktur besteht meist aus einem relationalen Datenbanksystem mit normalisierten, nicht hierarchischen Tabellen [KiRo02].

Wesentlich zur Gewährleistung der Qualität der Daten und Prozesse bei OLTP ist einerseits Transaktionssicherheit bei parallelen Anfragen, Minimierung der Antwortzeit von Anfragen sowie einem möglichst hohen Datendurchsatz. Durch die Beachtung von folgenden Transaktionskriterien wird sichergestellt, dass die Konsistenz der Datenbank erhalten bleibt und Daten nie unvollständig oder inkonsistent gespeichert werden [HäRa01]:

- *Atomarität*: Transaktionen werden entweder vollständig ausgeführt oder mit Hilfe der Aktion Rollback rückgängig gemacht.
- *Konsistenz*: Der Datenbestand nach einer Transaktion ist konsistent. Alle Integritätsbedingungen sind erfüllt.
- *Isolation*: Parallel ausgeführte Transaktionen können sich nicht gegenseitig beeinflussen. Jeder Datensatz der bearbeitet wird, ist zu dieser Zeit für alle anderen Operationen gesperrt.
- *Dauerhaftigkeit*: Auch bei etwaigen Systemabstürzen bleibt das Ergebnis einer abgeschlossenen Transaktion mit Hilfe von so genannten Pufferpools erhalten.

Dadurch dass OLTP-Systeme auf die Prozesse des operativen Tagesgeschäfts ausgerichtet sind, sind sie aus mehreren Gründen für die Analyse im Zuge des Business Intelligence Prozesses ungeeignet [SGBS04]:

- Das Konzept von OLTP-Systemen ist auf den operativen und nicht auf den analytischen Gebrauch orientiert.

- Diese Systeme sind durch das historische Wachstum oftmals unsystematisch gewachsen und als diesem Grund für automatisierte Analysen nicht geeignet
- Mit der Bearbeitung der operativen Tätigkeiten ist die maximale Auslastung der Systeme erreicht. Weitere Analyseprozesse würden die Systeme überfordern.
- Die Datenhaltung in den Systemen konzentriert sich auf die jüngsten, aktuellen Daten, historische Daten sind oft nicht mehr vorhanden.
- Die Speicherung von Analysestrukturen und -lösungen erfordert ein weiteres Speichersystem.

In der Abbildung 2 sieht man abschließend die Unterschiede zwischen der Datenhaltung für den operativen Bereich und der Datenspeicherung für den Business Intelligence Bereich im Überblick zusammengefasst.

	Charakteristika operativer Daten	Charakteristika dispositiver Daten
Ziel	Abwicklung der Geschäftsprozesse	Informationen für das Management: Entscheidungsunterstützung
Ausrichtung	Detaillierte, granulare Geschäftsvorfalldaten	Verdichtete, transformierte Daten; umfassendes Metadatenangebot
Zeitbezug	Aktuell; zeitpunktsbezogen; auf die Transaktion ausgerichtet	Unterschiedliche, aufgabenabhängige Aktualität; Historienbetrachtung
Modellierung	Altbestände oft nicht modelliert (funktionsorientiert)	Sachgebiets- o. themenbezogen, standardisiert u. endbenutzerorientiert
Zustand	Häufig redundant; inkonsistent	Konsistent modelliert; kontrollierte Redundanz
Update	Laufend und konkurrierend	Ergänzend; Fortschreibung abgeleiteter, aggregierter Daten
Queries	Strukturiert, meist statischer Programmcode	Ad-hoc für komplexe, ständig wechselnde Fragestellungen und vorgefertigte Standardauswertungen

Abbildung 2: Gegenüberstellung von operativen und dispositiven Daten nach [KeMU06]

3.2.2 Das Data Warehouse

Die Ideen und Ausführungen zum Thema Data Warehouse wurden aus einer früheren Arbeit der Autorin [Seit06] übernommen.

Das Data Warehouse ist eine zentrale Sammlung aller wichtigen Kennzahlen des Unternehmens. Dieses Systems beschäftigt sich nicht mit den operativen Vorgängen des Unternehmens sondern um die logische dispositive Organisation von unternehmensrelevanten Daten, damit Informationen flexibel und rasch abrufbar sind. Das System muss außerdem gewährleisten, dass alle Mitarbeiter die gleichen Informationen in derselben Aktualität erhalten.

Daten von verschiedenen Quellsystemen werden in das System integriert und in regelmäßigen Zeitabständen permanent in das System gespeichert. Dabei werden die Daten auf Vollständigkeit und Fehler geprüft und eventuelle korrigiert. Alle Änderungen werden im System dokumentiert werden.

Einer der Kernpunkte im Design des Data Warehouse ist die Granularität. Diese ist für die Flexibilität und Wiederverwendbarkeit der Daten im System mitverantwortlich. Das ist der Schlüssel zu Benutzerfreundlichkeit. Weitere Aufmerksamkeit bei der Modellierung sollte der Flexibilität der Datenstrukturen geboten werden, um in weiterer Folge auf noch nicht definierte Auswertungen und Fragen der Zukunft einzugehen.

Aus technischer Sicht hat man die Möglichkeit die logisch zusammengehörenden, zentral verwalteten Daten physisch entweder gemeinsam an einem Speicherplatz abzulegen, oder in verteilten Datenhaltungssystemen zu speichern.

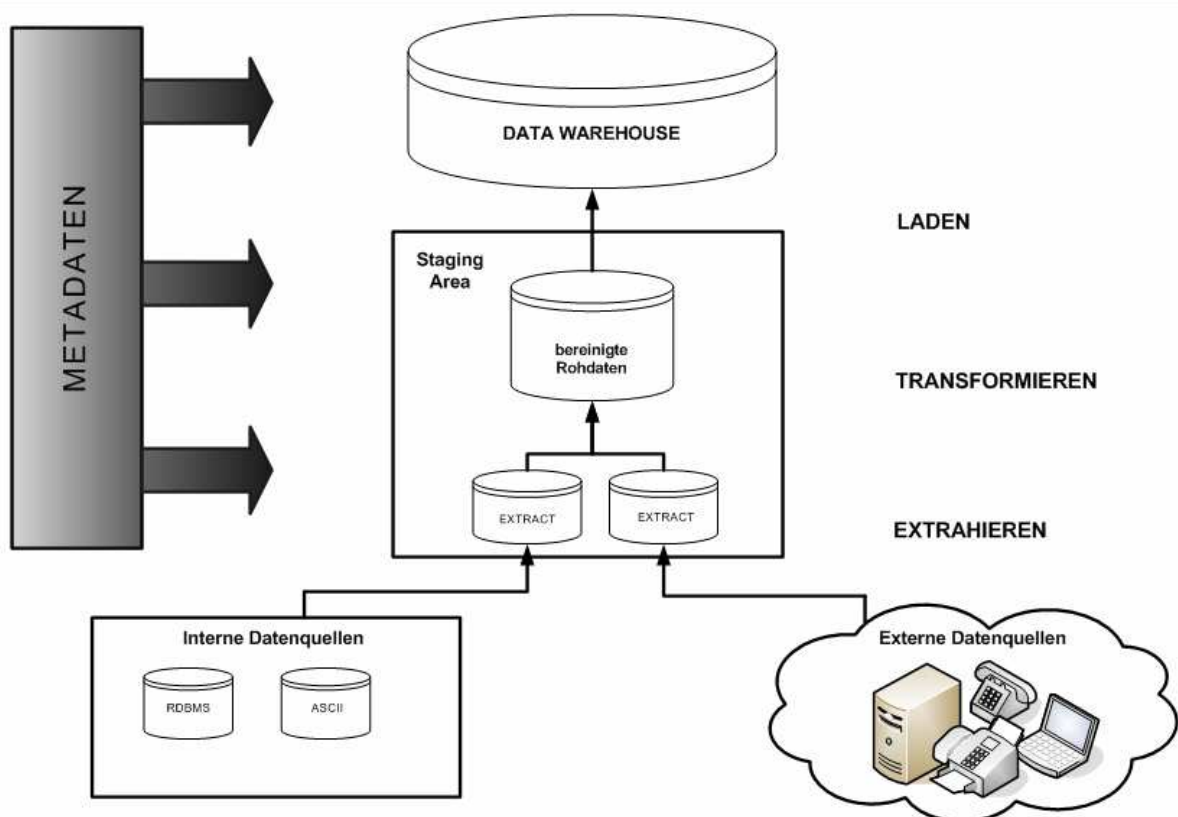


Abbildung 3: ETL-Prozess nach [ORDI03]

Wie in Abbildung 3 ersichtlich entstehen die Daten in verschiedenen operativen Systemen kommen aber auch aus externen Quellen. Der Transformationsprozess, auch ETL-Prozess genannt (Extract, Transformation, Loading), hat die Aufgabe die Daten von diesen Systemen in das Data Warehouse zu überführen. Dieser Prozess ist die Schnittstelle zwischen dem operativen und dem dispositiven Systemen und setzt sich aus folgenden Teilprozessen zusammen[KeMU06]:

Filterung: Die Daten werden aus den verschiedenen Quellen ausgewählt, zwischengespeichert und bereinigt.

Harmonisierung: Die gefilterten und bereinigten Daten werden zusammengeführt und auf dieselbe, möglichst detaillierte, Granularität gebracht.

Aggregation: In dieser Stufe werden die Daten um eventuelle Hierarchien erweitert.

Anreicherung: In der letzten Prozessstufe wird bereits mit den Daten eventuelle Kennzahlen berechnet und diese mitgespeichert, um für Standardberechnungen Performance-Vorteile bei späteren Abfragen zu erzielen.

3.2.2.1 Das relationale DWH nach Inmon

Im folgenden Kapitel wird das Datenmodell von W.H. Inmon [Inmo02] beschrieben, welches nach den relationalen Gesichtspunkten modelliert wird. Dieses wird in der Literatur auch Corporate Information Factory genannt.

Die Corporate Information Factory nach Inmon richtet sich nach so genannten Subjects. Es ist im Gegensatz zu den operativen Systemen weder applikations- noch prozessorientiert. Diese Subjects sind inhaltliche Kernbereiche, wie zum Beispiel Kunden oder Produkte. Jedes einzelne ist als Folge von zusammengehörenden Tabellen realisiert. Eine Subject Area kann 10, 100 oder mehr Tabellen beinhalten, welche über einen gemeinsamen Schlüssel miteinander in Beziehung stehen.

Erster Schritt zur Bildung eines solchen Modells ist auch hier die Integration aller Daten. Dazu wird festgelegt welche Daten das DWH enthalten soll, in welchen Organisationseinheiten diese vorkommen und wie sie benannt wurden. Anschließend folgt eine Kategorisierung und Strukturierung dieser Daten.

Das „overall corporate Model“ ist die Bezeichnung für ein Modell, das eine Überordnung für das operative Modell und das Modell des DWH bildet. Eine schematische Darstellung eines solchen Modells ist in Abbildung 4 zu sehen. Nachdem die Daten aus dem operativen Modell für das DWH gesammelt, bewertet und abgelegt wurden, können relevante Daten zum System hinzugefügt werden, d.h. Sekundärdaten, wie zum Beispiel gebräuchliche Kalkulationen über Sachverhalte, werden berechnet und beigelegt. Es werden im Weiteren die Schlüssel der Datensätze zur Historisierung um Zeitelemente erweitert. Dazu werden allen Einträgen neben dem Schlüssel noch ein Datum hinzugefügt, über welches ihre aktiven Zeiträume bestimmbar sind. Die Struktur erlaubt eine Unterscheidung in Daten der letzten Zeiträume, welche in sehr hoher Genauigkeit abgespeichert werden und Daten die schon länger im DWH sind, welche immer wieder in Strukturen mit geringerer Genauigkeit gespeichert werden. Je älter die Daten sind, desto höher ist ihre Granularität.

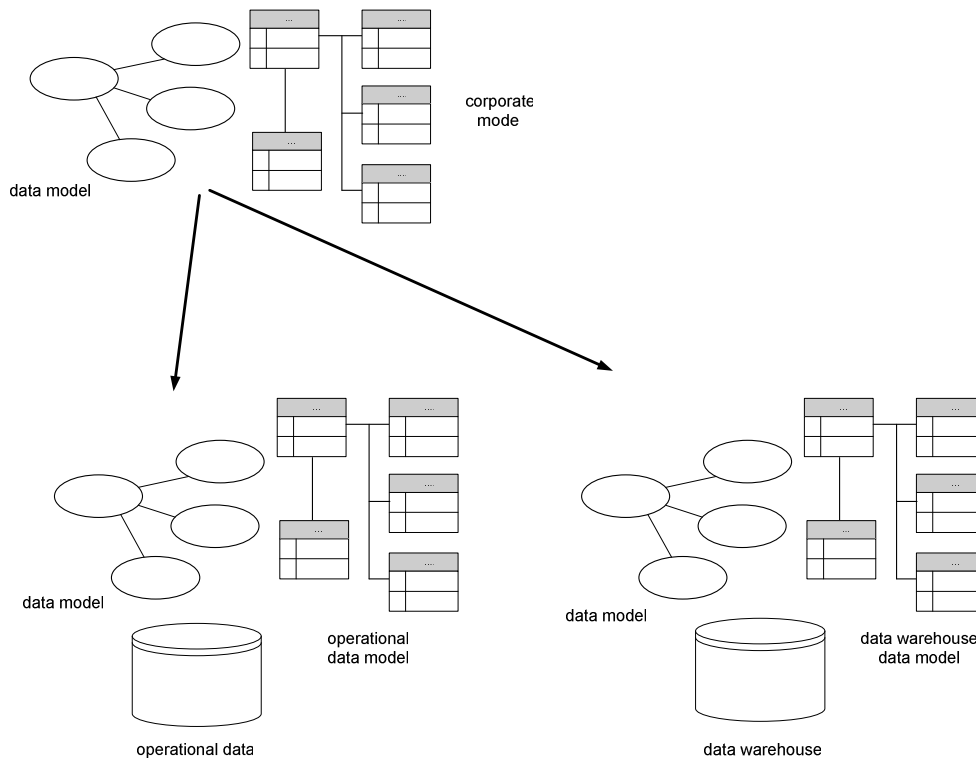


Abbildung 4: unterschiedliche Modellebenen nach Inmon [Inmo02]

Nachdem alle notwendigen Informationen für die Datenspeicherung und –verwaltung ermittelt wurden, wird das Modell in 3 Ebenen gebildet: das *high-level* modeling, *mid-level* modeling und das *low-level* modeling.

Im High Level Datenmodell (Entity Relationship Level im Folgenden kurz ERD genannt) wird der Umfang der Integration definiert. Dieser Schritt dient der Abgrenzung von irrelevanten Daten, und verhilft dem Modell sich als abgeschlossene Einheit zu repräsentieren. Anschließend werden alle Sichten der verschieden Benutzer zu einer ERD zusammengefasst.

In der *mid-level* Modellierung, auch Data Item Set (kurz DIS) bezeichnet, werden die ERDs zur logischen Organisation der Daten, in der die detaillierte Sicht der Attribute und Beziehungen dargestellt wird. In der *low-level* Modellierung, dem technischen Modell der Daten, werden die Beziehungen in der ERD-Ebene konkretisiert.

Bei der Umsetzung des technischen Modells wird das DIS-Modell mit den physikalischen Eigenschaften kreiert, welche über das DIS-Modell spezifiziert wurden. Es wird nun eine Serie von relational orientierten Tabellen generiert. Hier fließen Überlegung zur Sicherung der Performance und Realisierung bzw. in weiterer Folge die Reduktion der Schreib-/Lesezugriffe auf die Datenbank ein.

Die Speicherung der Tabellen in normalisierte und non-redundante Tabellen ist einer der zentralen Punkte. Normalisierte Daten bringen Flexibilität und sind an die verschiedenen Granularitäten sehr gut anpassbar. Die Basisstruktur der Daten beinhalten jedenfalls den Key und ein dazugehöriges Zeitattribut. Somit ist die Speicherung im DWH darauf ausgerichtet, dass jeder Datensatz der ins DWH gelangt wird als neuer Datensatz angefügt wird. Weiters enthält ein Datensatz direkte Daten und Fremdschlüssel zu anderen Tabellen, so genannte

Sekundärdaten. Auch hier muss auf die Zielsetzungen von guter Performanz Rücksicht genommen werden.

3.2.2.2 Das multidimensionale DWH nach Kimball

Die Multidimensionale Modellierung nach Ralph Kimball wird im Weiteren nach den Grundsätzen von [KiRo03] erklärt.

Als ersten Schritt zur Erstellung des Modells ist die Ermittlung und Festlegung der modellrelevanten Geschäftsfelder des Unternehmens. Danach werden die Dimensionen definiert. Abschließend werden Verbindungen der Dimensionstabellen zueinander mit Hilfe der Definition und Erstellung der Faktentabelle festgelegt.

Das Modell von Kimball ist nach der Idee eines „Standard bus interface“ realisiert, in dem alle relevanten Unternehmensbereiche eingegliedert sind. Anhand dieser werden Data Marts abgrenzt und im Zuge dessen Datenüberschneidungen untersucht. Diese Überschneidungen, sind zum Beispiel, dass in mehreren Unternehmensbereiche Informationen zu den verschiedenen Produkten bearbeitet werden. Damit diese Daten mehrfach verwendet werden können, gibt es im Modell so genannte Conformed Dimensions (siehe Abbildung 5).

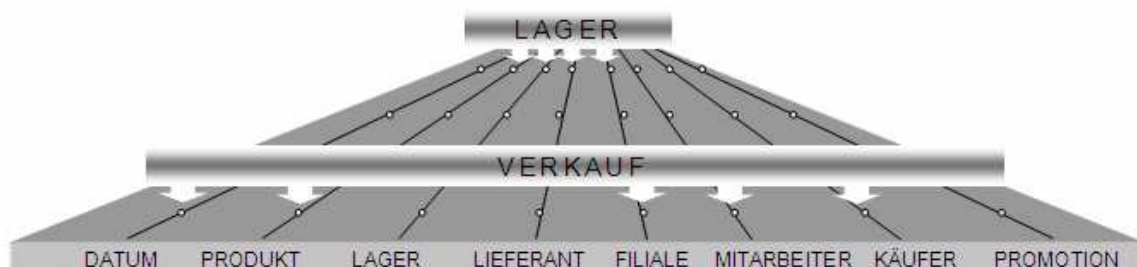


Abbildung 5: Conformed Dimensions nach [KiRo03]

Diese Conformed Dimensions sind Dimensionen, die so organisiert sind, dass sie von den unterschiedlichen Data Marts gemeinsam genutzt werden können

In diesem Schritt wird auch die Granularität der Daten festgelegt. Hierzu sollte überlegt werden, wie detailliert die Daten der Fakten-Tabellen vorhanden sein sollen. Dabei ist zu beachten, dass Daten zwar immer verdichtet angezeigt und danach weiterbearbeitet werden können, aber Daten nur in diesem Detaillierungsgrad vorhanden sind, in dem sie gespeichert wurden.

In den Dimensionstabellen sind die Beschreibungen zu den Daten aus den Faktentabellen enthalten. Nach diesen Attributen werden Reports erstellt. Daher sollten diese Daten gut lesbar und ausreichend beschreibend sein.

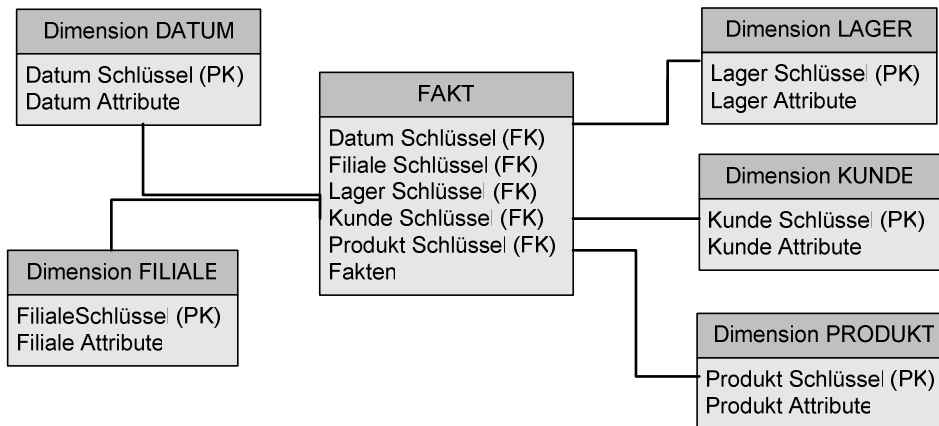


Abbildung 6: Star Schemas nach [KiRo03]

Das multidimensionale Modell von Kimball ist im Idealfall durch so genannte Starschemas realisiert. Das sind Datenmodelle, welche keine Normalisierungen aufweisen, wie man in Abbildung 6 ersehen kann. Die Hierarchie der Daten werden, wo immer es Sinn macht nicht normalisiert, sondern redundant gespeichert. Snowflake-Strukturen (siehe Abbildung 7), welche normalisierte Tabellen enthalten, werden nur in Ausnahmefällen verwendet. Kritiker, welche auf den unnötig beanspruchten Speicherplatz verweisen, wird die relativ geringe Größe von Dimensionen entgegengehalten. Dem grundlegenden Konzept des Data Warehouse folgend, ist hierbei nicht der Speicherplatz, sondern die Möglichkeit zur Erstellung von einfachen und performanten Abfragen entscheidend.

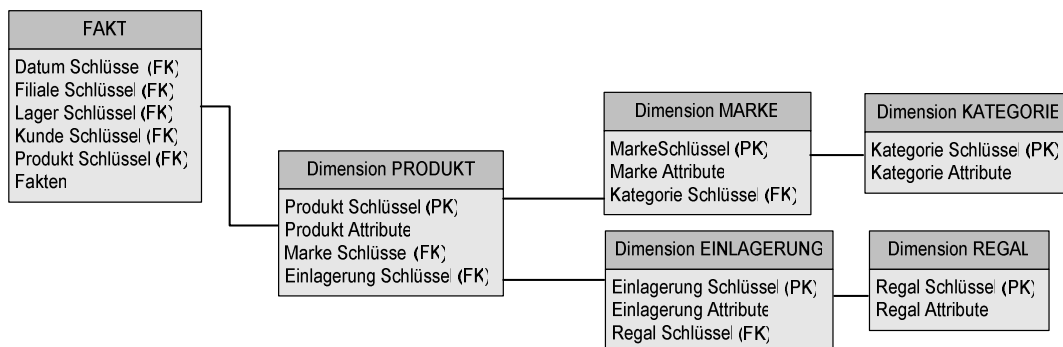


Abbildung 7: Snowflake Schema nach [KiRo03]

Die optimale Anzahl nach Kimballs Ausführungen liegen zwischen 5 bis 15 Dimensionen. Nachdem das Konzept der Dimensionen feststeht, wird die dazugehörige Faktentabelle definiert. Durch die Faktentabelle werden die Dimensionen mit Hilfe ihrer Schlüssel zueinander in Beziehung gestellt. Die Faktentabelle enthält neben dieser Fremdschlüssel zu den Dimensionen vor allem numerische und additive Felder. Diese Verbindung der Dimensionen und des Faktens ergibt schlussendlich den so genannten Star des Data Mart.

Die Erstellung des multidimensionalen Modells geht nach iterativen Vorgängen voran. Nach jedem abgeschlossenen Prozess wird kontrolliert, ob das definierte Ergebnis erreicht wurde. Falls das nicht der Falls ist, wird in die Vorherige Phase zurück verzweigt. Damit wird gewährleistet, dass man in der Modellierung zu einem optimalen Ergebnis gelangt

3.3 Aufbereitung und Darstellung von Information

Nachdem die Daten im Data Warehouse gespeichert wurden, gibt es eine Vielzahl an Möglichkeiten, um diese als Information im Unternehmen weiterzugeben. Die einfachsten Varianten bieten Reportsysteme zur Generierung von verschiedenen Berichten an, die Auswertung und Darstellung über OLAP oder Kennzahlensysteme, wie die Balanced Scorecard, geht bereits einen Schritt weiter von der reinen Darstellung der Information in Richtung Analysemethoden. Bevor diese aufgezählten Methoden im Weiteren vorgestellt werden, erfolgt ein kurzer Überblick über die Relevanz und Möglichkeiten von Kommunikation im Unternehmen.

3.3.1 Kommunikation im Unternehmen

Aufgrund der Veränderung der Wettbewerbssituation, der Informations- und Kommunikationstechnik und sowie der Veränderung der gesellschaftlicher Werte hat sich auch die Art der Kommunikation im Unternehmen gewandelt.

Durch die weltweite Globalisierung aufgrund von technischen Entwicklungen spielen regionale oder nationale Grenzen bei der Definition und Koordination wirtschaftlicher Aktivitäten eine immer geringere Rolle. Außerdem werden durch den erheblich vereinfachten, weltweiten Zugriff Wissensträger und Wissensbestände vereinfacht zugänglich gemacht. Durch die neuartigen Vernetzungsmöglichkeiten von Prozessen und Personen werden Spezialisierung und Qualifizierung von Menschen konzentriert auf Organisationen irrelevant. Die veränderten Wettbewerbsbedingungen verlangen von den Unternehmen Flexibilität und Innovationsfähigkeit. Einerseits wird eine Fülle von Daten geboten, andererseits muss daraus die relevante Information gefiltert werden. In Abhängigkeit von der Aufgabenstellung und den daraus resultierenden Kommunikationserfordernissen lässt sich der Einsatz von geeigneten Kommunikationsmitteln mit Hilfe von Kommunikationsmanagement planen, organisieren und kontrollieren. Bei der Wahl der Kommunikationsmittel ist immer deren Aufgabenbezug zu berücksichtigen. Die gezielte Auseinandersetzung mit dem Produktionsfaktor Information ist als wichtige Aufgabe der Unternehmensführung zu planen, zu organisieren und zu kontrollieren.

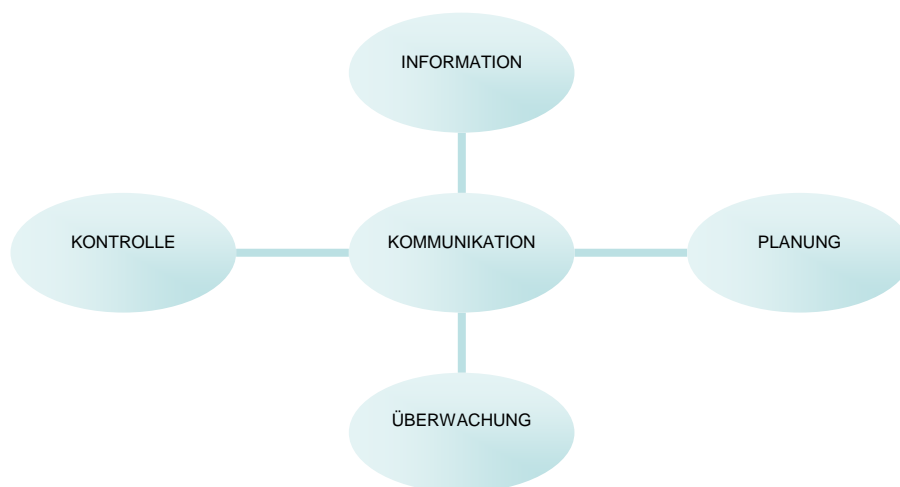


Abbildung 8: Kommunikation im Unternehmen

In Abbildung 8 sind die einzelnen Aufgaben von Kommunikation dargestellt. Die Zielsetzung eines Informationsmanagements ist die Sicherstellung eines effektiven und wirtschaftlichen Einsatzes von Information. Mit diesem Fokus sind technische, organisatorische und personelle Bedingungen zu gestalten, wobei besonders der Informationsbedarf und seine Deckung für alle wesentlichen Verwendungszwecke Vorrang hat. Unter Informationsbedarf versteht man Art, Menge und Qualität der Informationen, die zur Erfüllung einer Aufgabe in einer bestimmten Zeit benötigt werden. [PiRW98]

3.3.2 Reporting

Die Ideen und Ausführungen zum Thema Data Reporting wurden aus einer früheren Arbeit der Autorin [Seit06] übernommen.

Berichtssysteme helfen bei der Lösung von betriebswirtschaftlichen Aufgaben indem sie unternehmensrelevante Informationen auswerten und bereitstellen.

„Abfragen und Berichtssysteme erlauben die einfache Auswertung von Dateien und Datenbanken (Datenextraktion und –aggregation) und die ansprechende Präsentation der Ergebnisse in fester oder variabler Form.“ [siehe HaNe05 S.778]

Nach [KeMU06] beinhaltet das betriebliche Berichtswesen die Prozesse Berichtsgestaltung und -erstellung, Berichtsverteilung, Berichtsverwaltung sowie die Berichtsaufnahme und –diskussion. Auf die Abläufe des Berichtswesens wird im Folgenden nicht näher eingegangen, da in dieser Arbeit ausschließlich die technischen Aspekte beschrieben werden.

Eine Einteilung der Berichtssysteme nach [KeMU06] erfolgt in erster Linie auf Grund der Benutzerinteraktion:

- Aktive Berichtssysteme
 - Periodische Berichtssysteme
 - Aperiodische Berichtssysteme
- Passive Berichtssysteme
 - Ad hoc Berichtssysteme

Aktive Berichtssysteme bedürfen keiner Interaktion des Benutzers, sie werden automatisiert generiert. Hierbei kann man unterscheiden, ob diese Generierung in regelmäßigen Abständen passiert, dann gehören sie zu der Gruppe der periodischen Berichtssystemen, oder, wie bei den Aperiodischen Berichtssystemen, nur bei unvorhergesehenen Ereignissen und Überschreitungen von Grenzwerten durch so genannte Frühwarnsysteme. Passive Berichtssysteme sind nicht automatisiert, hier muss der Benutzer den Bericht manuell explizit anfragen. In dieser Gruppe sind Ad-hoc Berichte sehr beliebt, welche der Benutzer an seine eigenen Bedürfnisse und Fragestellungen anpassen kann.

Es sind viele verschiedene Software Lösungen zur Erstellung und Gestaltung der Berichte vorhanden. Die Werkzeuge welche mit graphischer Oberfläche und Drag-and Drop Techniken dem Benutzer Hilfe bei der Berichtserstellung bieten, sind besonders für EDV-unerfahrenen Mitarbeitern gedacht, welche ihre Anfragen nicht an IT Spezialisten weiterleiten müssen, sondern selbst auswerten können. Hierbei werden die Daten oft mit Hilfe von OLAP (Online Analytical Processing) Werkzeuge bearbeitet. OLAP Operationen, auf welche im nächsten Kapitel noch näher eingegangen wird, ermöglichen auch technisch unerfahrenen Benutzern eine einfache Datenauswahl und Durchführung von Analysen im System [KeMU06].

Es ist auch möglich Ad-hoc Berichte mit Hilfe von freier Datenrecherche zu erstellen. Freie Datenrecherche bezeichnet die Bearbeitung und Illustration von Daten über eine Datenmanipulationssprache (DML Sprache) wie z.B. SQL (Structured Query Language) oder MDX (Multidimension Expression). Für die Erstellung dieser Berichte sind jedoch Programmierkenntnisse das heißt im Weiteren doch eher umfangreiche IT Kenntnisse des Erstellers notwendig.

Jedoch kann man mit Hilfe dieser Werkzeuge alle genannte Arten der Berichtswesen erstellen, das impliziert sowohl die Erzeugung von periodischen und aperiodischen als auch von interaktiven und ad hoc Berichten.

3.3.3 OLAP

OLAP ist ein Instrument zur Datenanalyse und Berichtserstellung, welche über ein standardisiertes Berichtswesen hinausgeht. Dem Benutzer ist es möglich intuitiv Auswertungen und Analysen zu gestalten, somit können entscheidungsrelevante Daten zum richtigen Zeitpunkt angeboten werden.

Aus logischer Sicht wird bei diesem Prozess ein mehrdimensionaler Datenraum aufgespannt, in dem sich der Benutzer mittels Ad-hoc Fragen bewegt [KeMU06].

Die Definition der Anforderungen an OLAP Systeme wurde mit Hilfe der folgenden zwölf Codd'schen Regeln von E.F. Codd standardisiert [Knob05, OLAP01]:

1) *Multidimensionale konzeptionelle Sichtweise*

Die Sicht auf die Daten soll aus verschiedenen Dimensionen erfolgen können und der Benutzer soll darin flexibel navigieren können.

2) *Transparenz*

Der transparente Zugriff auf Daten verschiedener Quellen ist möglich.

3) *Zugriffsmöglichkeit*

Es sind Zugriffe auf Daten von internen und externen Quellen möglich.

4) *Gleichbleibende Antwortzeit bei der Berichterstellung*

Die Antwortzeit ist unabhängig vom Datenvolumen und der Anzahl der Dimensionen.

5) *Client-Server-Architektur*

Die Prozesse Speicherung, Verarbeitung und Darstellung werden getrennt realisiert.

6) *Generische Dimensionalität*

Alle Dimensionen werden gleich behandelt.

7) *Dynamische Behandlung dünn besetzter Matrizen*

Das physische Schema wird an die Datenverteilung und Dimensionen angepasst.

8) *Mehrbenutzer-Unterstützung*

Konkurrierende Zugriffe werden durch Sicherheits- und Integritätsmechanismen sowie Zugangsrechte gesteuert.

9) *Uneingeschränkte kreuzdimensionale Operationen*

Die automatische Ableitung der Berechnungen, die sich aus den Hierarchiebeziehungen der Dimensionen ergeben ist gegeben.

10) Intuitive Datenbearbeitung

Es gibt eine ergonomische, intuitive Benutzeroberfläche, welche eine flexible Navigation über Datenmöglich macht.

11) Flexible Berichterstellung

Es ist möglich, Berichten mit beliebiger Datenanordnung zu erstellen.

12) Unbegrenzte Anzahl von Dimensionen und Klassifikationsebenen

Es existieren keine Einschränkungen hinsichtlich der Anzahl unterstützter Dimensionen und deren Aggregationsebenen.

Ein paar Jahre später, 1995, wurden von N. Pendse die Definition für „Fast Analysis of Shared Multidimensional Information“ (FASMI) publiziert. Hierin werden die fünf grundlegenden Eigenschaften von OLAP zusammengefasst [OLAP01]:

1) Geschwindigkeit (Fast):

Die Antwortzeit der meisten Anwenderanfragen ist in der Regel 5 Sekunden, bei komplexeren Anfragen vergehen max. 20 Sekunden.

2) Analysemöglichkeit (Analysis):

Es ist die Möglichkeit zur anwenderfreundlichen, intuitiven Analyse gegeben.

3) Sicherheit (Shared):

Die Systeme sind mehrbenutzerfähig und es existieren Sicherheitsstandards wie Zugriffsrechte auf Zellenebene sowie Sperrverfahren.

4) Multidimensionalität (Multidimensional):

Es ist eine multidimensionale, konzeptuelle Sicht unabhängig vom zugrunde liegenden Datenbanksystem gewährleistet

5) Kapazität (Information):

Es existiert keine Begrenzung des Datenvolumens oder der Dimensionalität.

Mitarbeiter, welche mit OLAP im Unternehmen ad-hoc Auswertungen erstellen wollen, sind meistens keine Datenbankspezialisten, sondern hauptsächlich aus Vertriebs- und Controllingabteilungen und verfügen über eher geringen IT-Kenntnissen. Diese Gegebenheit fordert sowohl eine intuitive Bedienbarkeit, als auch Vielseitigkeit und Flexibilität des Systems.

Wie bereits in den Regeln von Codd und auch im „FASMI“ definiert, ist es bei OLAP-Systemen also festgelegt, dass aus der meist relationalen Struktur der Datenbank eine Darstellung in mehrdimensionalen Sichten generiert wird.

Um diese Multidimensionalität zu erklären, wird das System oftmals als so genannter Data Cube oder Datenwürfel vorgestellt.

Jede Dimension des Modells bildet eine Achse des Würfels. Am besten kann man dieses Modell mit Hilfe von 3 Dimensionen darstellen, wobei die Achsen zum Beispiel das Produkt, den Standort und die Zeit darstellen (siehe Abbildung 9) [ChGI06].

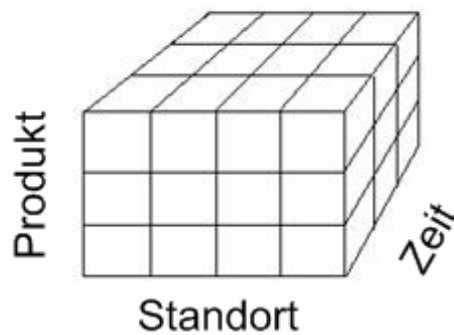


Abbildung 9: Data Cube mit 3 Dimensionen

In diesem Datenwürfel kann nun nach Bedarf navigiert werden. Die wichtigsten Navigationsmöglichkeiten werden nun überblicksmäßig dargestellt [KeMU06]:

Pivotierung

Unter Pivotierung versteht man das Rotieren des Würfels um die eigene Achse. Wenn nur zwei Dimensionen dargestellt werden, kann man diese in der Ansicht austauschen, indem man sozusagen den Würfel um eine Dimension dreht. Diese Darstellung wird mit Hilfe von Pivot-Tabellen realisiert.

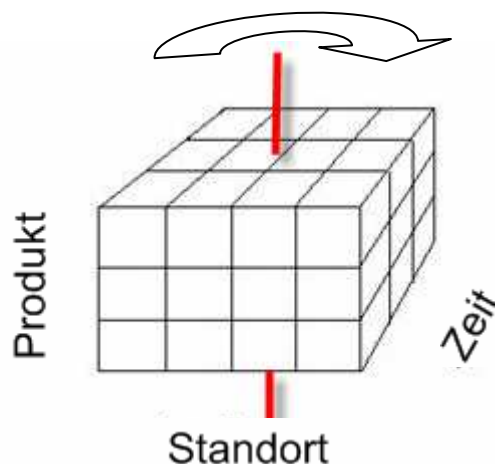


Abbildung 10: Pivotierung des Data Cube

In Abbildung 10 würden statt der Attribute Produkt und Standort nun Produkt und Zeit die Tabelle bilden.

Roll-up & Drill-Down

Viele der Dimension sind in sich noch in Hierarchien untergliedert. Die Dimension Produkt hat zum Beispiel noch Attribute wie Produktgruppe oder Brand (Marke), welche hierarchische Strukturierung zulassen. Um in diesen Hierarchien zu navigieren, kann man Operationen wie Roll-Up oder Drill-Down anwenden.

Beim Roll-Up werden alle Einzelwerte zu einem weiter oben liegenden Hierarchieattribut verdichtet. Es ist zum Beispiel möglich, sich die Daten einerseits auf Produktebene anzusehen, oder mit Hilfe einer Roll-Up Operation diese Werte zu aggregieren und nur mehr die Darstellung auf Ebene der Produktgruppe zu betrachten. Wenn nun diese Produkte und Ihre Jahresumsätze dargestellt wurden, werden in der aggregierten Darstellung die Jahresumsätze auf alle Produktgruppen gezeigt. Ein Beispiel dazu ist in Abbildung 11 dargestellt.

Die Funktion Drill-Down hat genau die entgegengesetzte Funktion. Diese Operation bewirkt eine detaillierte Darstellung der Daten. Es muss jedoch bereits bei der Modellierung des Data Warehouse überlegt werden, bis zu welchem Detaillierungsgrad die Daten vorhanden sein sollen.

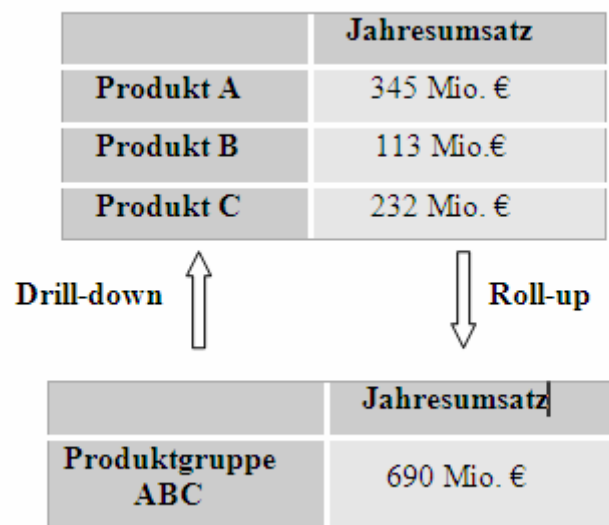


Abbildung 11: Roll-up & Drill Down

Slice & Dice

Die Operationen Slice und Dice bieten die Möglichkeit in einer großen Menge an Daten, die relevanten Informationen herauszufiltern. Dies wird bei Slicing, was soviel bedeutet wie „in Scheiben schneiden“ durch die Reduktion einer Dimension zu einem Wert realisiert. Man kann sich zum Beispiel alle Daten zu einem bestimmten Produkt ansehen (siehe Abbildung 12).

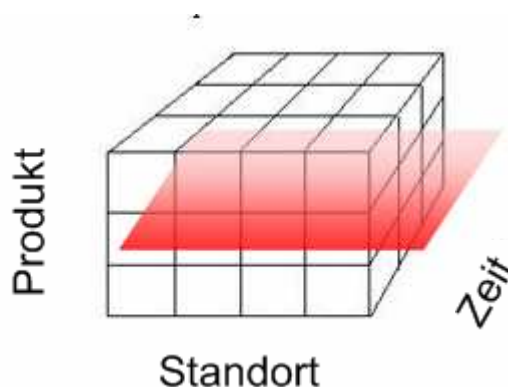


Abbildung 12: Selektion aus den Daten mit Hilfe der Slice-Operation

Die Dice-Operation ermöglicht einen kleinen Ausschnitt aus den Gesamtdaten in Richtung mehrerer Dimensionen, welcher jedoch mehrdimensional bleibt. In Abbildung 13 ist diese Selektion schematisch dargestellt:

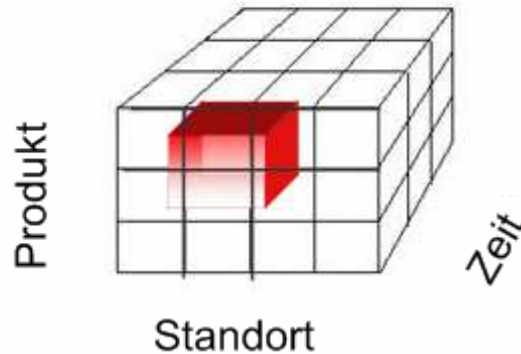


Abbildung 13: Selektion aus den Daten mit Hilfe der Dice-Operation

Mit Hilfe der oben beschriebenen Operationen ist es für den Benutzer nun möglich interaktiv auf einer graphischen Benutzeroberfläche durch Drag & Drop der Dimensionsschaltflächen auf verschiedene Positionen beliebige Auf- und Einsichten in den Datenwürfel zu erreichen.

3.3.4 Kennzahlensysteme

Kennzahlensysteme beschreiben eine Kombination von relevanten Kennzahlen des Unternehmens zur Darstellung und Beurteilung von Sachverhalten. Diese Kennzahlen werden nach einer definierten Systematik ausgewählt und zu einander in Beziehung gestellt, um verschiedene Interpretationsspielräume und Auslegungssichten zu minimieren und den Sachverhalt einheitlich darzustellen.

Es gibt verschiedene Arten von Kennzahlensystemen. Es wird in eindimensionale, mehrdimensionale und Partialansätze unterschieden.

In eindimensionalen Kennzahlensystemen wie zum Beispiel das DuPont System, welches den Return on Investment darstellt, wird eine rein monetäre Spitzenkennzahl dargestellt. Dieses Kennzahlensystem ist das Grundsystem, worauf viele andere Systeme aufbauen.

In den mehrdimensionalen Systemen, wie dem Tableau de Bord und Balanced Scorecard, werden neben den monetäre auch nicht-monetäre Spitzenkennzahlen dargestellt. Bei Kennzahlensystemen mit Partialansätzen, wie beim Target Costing oder dem Benchmarking wird die Leistung einer Unternehmung im eingeschränkten Sinne beurteilt [COPO07].

Beim Business Intelligence Prozess werden Daten häufig mit Hilfe des Balanced Scorecard Modell dargestellt. Deshalb wird dieses nachfolgend genauer erklärt:

Die Balanced Scorecard wurde im Jahr 1992 erstmals durch Robert S. Kaplan und David P. Norton veröffentlicht. Das Konzept beruht auf der Darstellung der vier verschiedenen Blickwinkel, welche berücksichtigt werden, um eine ausgeglichene Steuerung zu gewährleisten [Wiki07].

In Abbildung 14 sind die 4 Perspektiven welche in die Strategieüberlegungen mit einbezogen werden, dargestellt [COPO07a]:

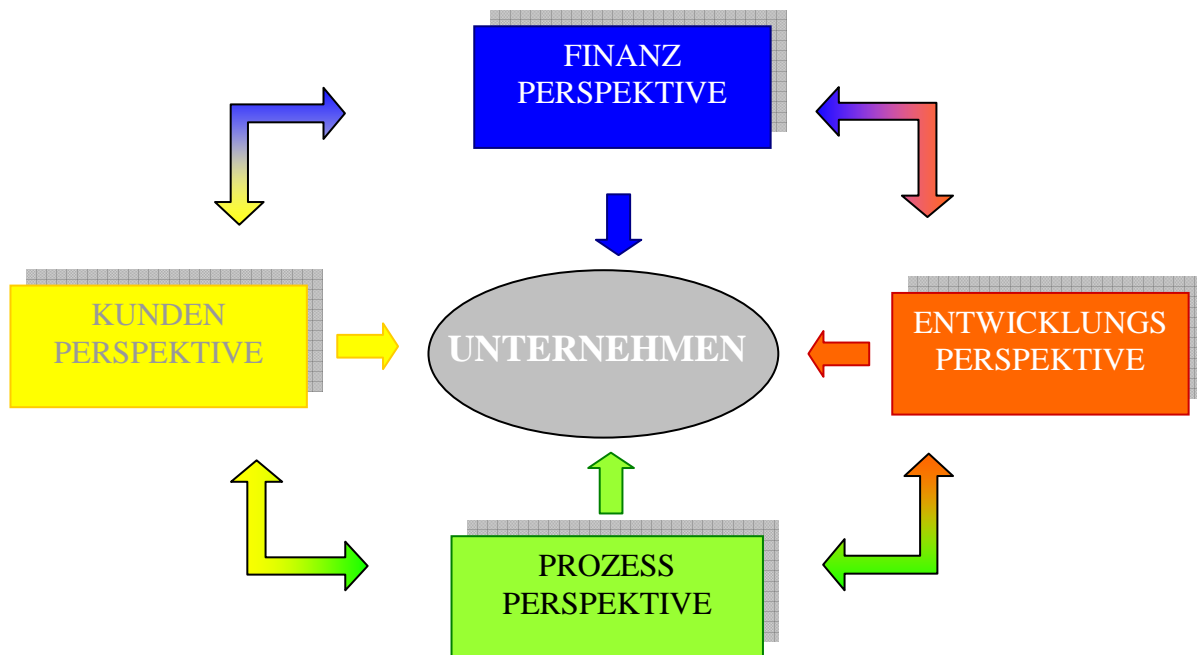


Abbildung 14: Die 4 Perspektiven der Balanced Scorecard

a) Finanzperspektive

Finanzielle Kennzahlen sollen erkennen lassen, ob eine Strategie das Gesamtergebnis des Unternehmens verbessert. Klassische Kennzahlen sind in diesem Bereich: Rentabilität, Ergebnis, Finanzkraft, Wachstum und weitere Ziele wie z.B. Kostensenkung.

Die Finanzperspektive steht des Weiteren noch mit allen anderen Perspektiven in einer Beziehung, da alle Kennzahlen, welche im Endeffekt auch über finanzielle Ergebnisse definiert werden.

b) Kundenperspektive:

Die Kundenperspektive definiert die strategischen Ziele hinsichtlich der Kunden- und Marktsegmente. Hier fallen unter anderem folgende Fragen an: „Wie schätzt der Kunde das Unternehmen ein?“ Es werden somit die Kunden- und Marktsegmente des Unternehmens untersucht.

b) Prozessperspektive:

Die Prozessperspektive untersucht die wichtigsten, internen Kernprozesse. „Was muss intern getan werden, um die definierten Ziele im Bereich Kunden und Finanz zu erreichen“.

c) Entwicklungsperspektive:

Die Kennzahlen der Entwicklungsperspektive beschreiben die Ziele der Lern- und Entwicklungsprozesse des Unternehmens, wobei hier das Augenmerk wieder auf den Prozessen zu den vorhergegangenen Perspektiven liegt. Es werden Fragen nach dem “Wie kann sich das Unternehmen verbessern und Innovationen einführen“ gestellt. Hierbei wird die Einbindung der Mitarbeiter in Unternehmensprozesse, und somit das tatsächliche Ergebnis

der Organisationsprozesse im Unternehmen, überprüft. Relevante Zielgrößen dafür sind zum Beispiel Mitarbeiterzufriedenheit, Mitarbeiterproduktivität und Weiterbildung.

Nachdem die Perspektiven, welche eventuell noch durch weitere, für das Unternehmen wichtige Sichten ergänzt werden können, als Konstrukt vorhanden sind, werden für jede dieser Bereiche mit Hilfe von Kennzahlen Ziele definiert. Diese können dann kontrolliert und gesteuert werden.

3.4 Analyse von Informationen

Wenn durch Darstellung der Daten mit Hilfe von OLAP-Werkzeugen oder Balanced Scorecard keine bzw. unzureichende Informationen gewonnen werden, gibt es weiterführende Möglichkeiten die Daten mit Hilfe von statistischen Methoden zu analysieren.

In der Analyse von Daten gibt es prinzipiell zwei Verfahren, nach denen vorgegangen wird, um Daten zu untersuchen. Einerseits gibt es Fragestellungen, bei denen eine Behauptung aufgestellt wird, welche mit Hilfe der vorhandenen Daten untersucht wird. Dabei spricht man von hypothesengetriebenen Fragestellungen, auch „Top-Down“ Probleme genannt. Wenn andererseits aber Datenbestände untersucht werden, welchen keine konkreten Annahmen zugrunde liegen spricht man von hypothesefreien Fragestellungen, auch „Bottom-Up“ Probleme genannt [Knob00]. Die Abbildung 15 verdeutlicht das Schema:

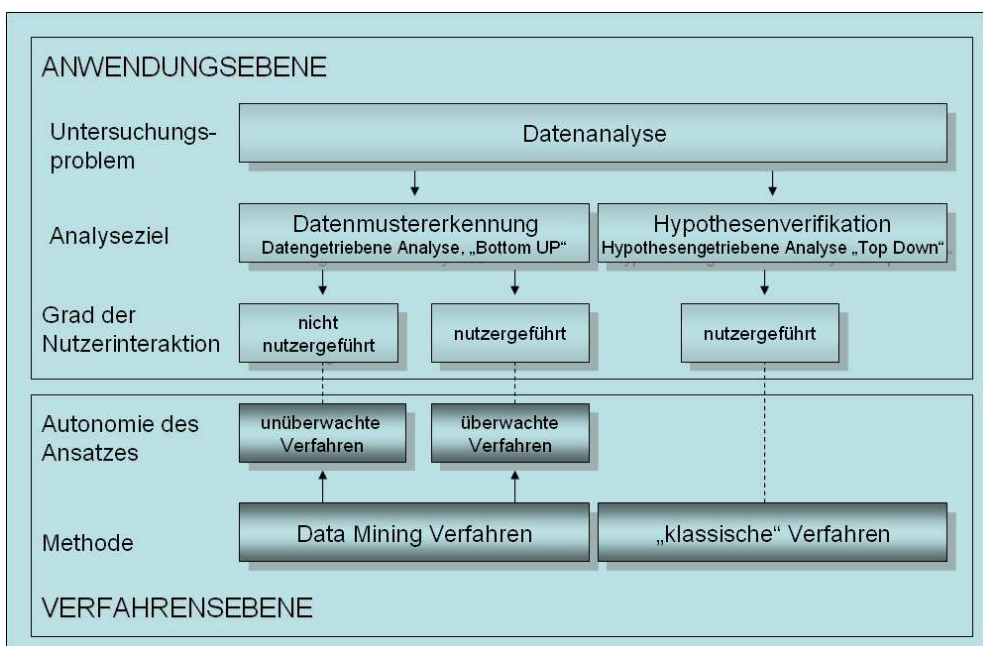


Abbildung 15: Einordnung von Data Mining nach [Knob00]

3.4.1 Data Mining

Data Mining umfasst Möglichkeiten um Daten mit Hilfe verschiedener statistischer Methoden zu untersuchen. Es wird gern im Zuge von Business Intelligence Prozessen eingesetzt um entweder Hypothesen zu finden oder zu prüfen. Ziel ist hier, in großen Datenmengen interessante Zusammenhänge herauszufinden, welche mit freiem Auge nicht aufgedeckt würden. Diese Verfahren der Suche in Datenbeständen nach Mustern und Zusammenhängen ist sehr komplex und während einfache Reports bei Bedarf von vielen Anwendern generiert werden, setzen vor allem Spezialisten Data Mining Methoden ein.

Der Begriff Data Mining umfasst einerseits die Untersuchung von strukturierten Daten, weiters gibt es aber auch die Möglichkeit mit Hilfe von Text Mining große Mengen an unstrukturierte Daten und Dokumente zu untersuchen. Hier werden die Inhalte der Dokumente erschlossen und eventuell interessante Verbindungen zwischen Dokumenten

gefunden. Data- und Text Mining werden häufig unter dem Oberbegriff „Knowledge Discovery in Databases“, auch unter der Abkürzung KDD zu finden, geführt. Es handelt sich um prozessorientierte Komponenten zum Auffinden und Aufbereiten von Rohdaten, zum Entdecken von Wissen, sowie logischen Zusammenhängen dieser Daten. Web Mining verbindet Data Mining mit Text Mining bezogen auf das Inter- bzw. Intranet. Einerseits wird das Benutzerverhalten von Usern mittels Web Usage Mining untersucht. Hierbei wird beobachtet, wie sich Benutzer auf bestimmten Internetseiten verhalten, wie oft diese besucht werden, was wozu der Benutzer Webseiten besucht (Informationen sammeln, Online-Käufe tätigen...)gemacht wird u.s.w. Andererseits wird beim Web Content Mining, ähnlich wie beim Text Mining der Inhalt der Webseiten analysiert [PREU02].

Es gibt verschiedene Problemstellungen bzw. Aufgabe, welche in Data Mining untersucht werden. Die gängigsten davon werden hier nun aufgeführt und erklärt. Danach werden die dazugehörigen Verfahren, welche üblicherweise zur Lösung des Problems verwendet werden, kurz dargestellt. Zum besseren Verständnis werden die Verfahren und Aufgaben zuvor überblicksmäßig in Abbildung 16 dargestellt.

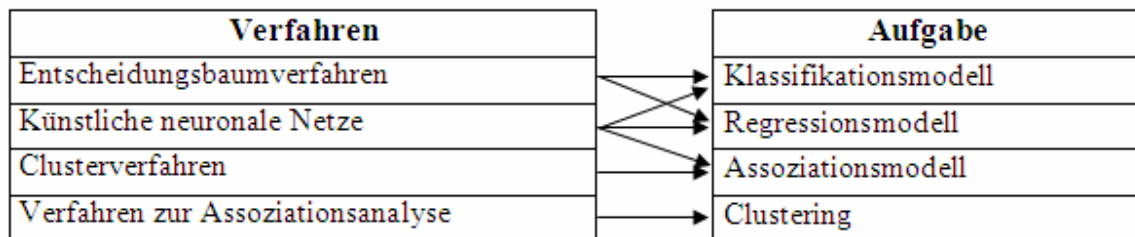


Abbildung 16: Data Mining Aufgaben und Verfahren im Überblick [ChC106],[Lust02]

3.4.1.1 Data Mining Aufgaben

Klassifikationsmodell

Bei einem Klassifikationsmodell werden mehrere, durch unterschiedliche Attributsausprägungen unterscheidbare Klassen bestimmt. Einzelne Elemente werden anschließend einer bestimmten Klasse zugeordnet, wenn die charakteristischen Eigenschaften bzw. das Verhalten zur Klasse passt. Diese Klassen mit den zugehörigen Elementen stellen die Trainingsdatenmenge dar. Auf Basis dieser Trainingsdatenmenge wird des Weiteren ein Klassifikationsmodell entwickelt. Dieses hilft später bei der Bestimmung von Klassenzugehörigkeit von neuen Elementen, sprich bei der Klassifikation. Das klassische Beispiel ist eine Kreditwürdigkeitsprüfung. Hier wird ein Kunde anhand seiner angegebenen Kundendaten als kreditwürdig oder –unwürdig eingestuft (siehe Abbildung 17) [ChC106].

Zur Erstellung dieser Klassifikationsmodelle werden z.B. neuronale Netzwerke oder Entscheidungsbäume eingesetzt. Beide Verfahren werden im Anschluss im Kapitel „Data Mining Verfahren“ besprochen.

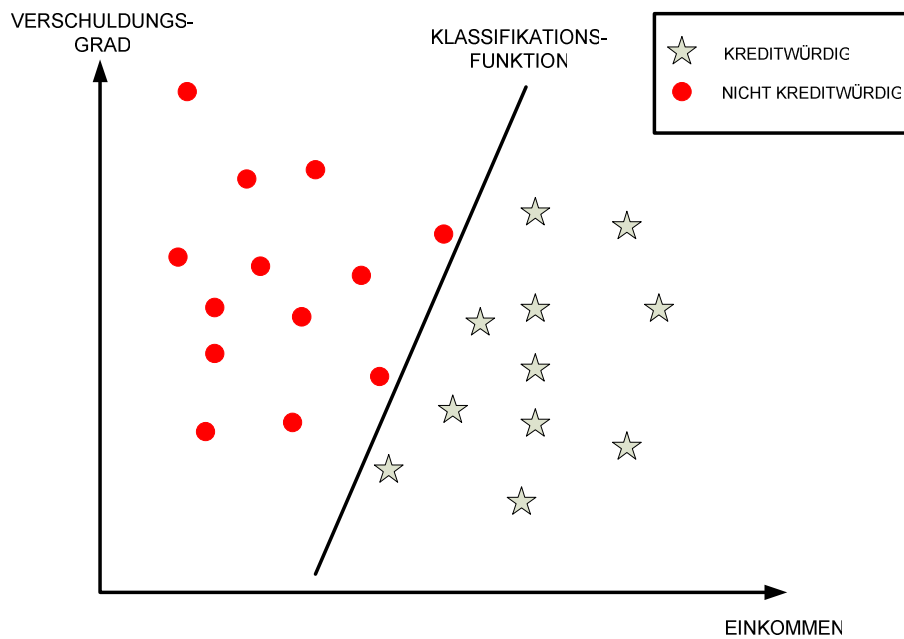


Abbildung 17: Klassifikation nach [GrBe99]

Regressionsmodell

Im Regressionsmodell wird eine abhängige Variable mit Hilfe von mehreren unabhängigen Variablen erklärt. Die Zielsetzung zur Klassifikation ist ähnlich, in diesem Fall werden jedoch keine Klassen gebildet. Regressionsmodelle werden gebildet um zum Beispiel Absatzprognosen oder Umsatzentwicklungen für das nächste Jahr zu berechnen [ChCl06].

Clustering

Beim Clustering werden Gruppen mit gleichen Merkmalen gebildet, wobei versucht wird, die Gruppenelemente zueinander möglichst ähnlich sind, sich zu den anderen Gruppen aber deutlich unterscheiden. Diese Ähnlichkeiten sind jedoch zuvor nicht bekannt, sondern sollen mit Hilfe des Clustering, welches durch einem Cluster-Verfahren aus der multivariaten Statistik zustande kommt, erkannt werden. Das Interpretieren der bestimmten Cluster erweist sich oft als schwierig. Clustering wird oft bei Betrachtung von Kunden auf einen bestimmten Produkt-Markt angewandt, um diesen zu segmentieren und neue Kundengruppen aufzufinden [ChCl06]. In Abbildung 18 wird dieses Verfahren veranschaulicht.

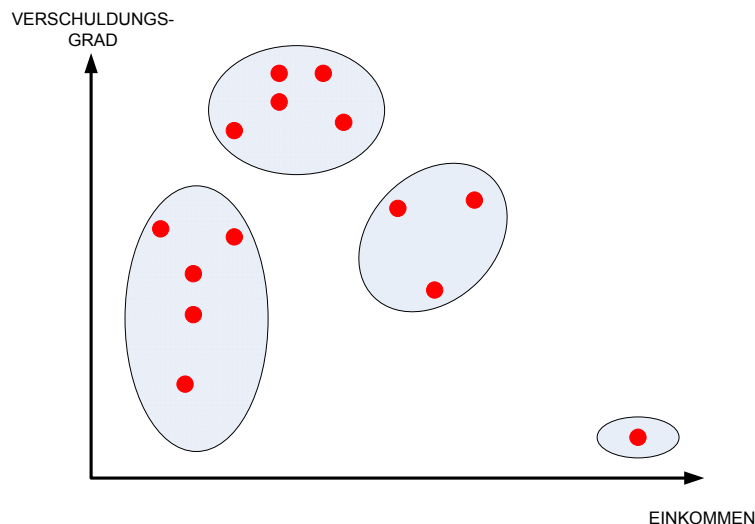


Abbildung 18: Clustering nach [GrBe99]

Assoziation

Unter Assoziation versteht man die Entdeckung von Abhängigkeiten. Hier werden Abhängigkeiten zwischen Merkmalen oder deren Ausprägungen aufgedeckt, welche mit einer „Wenn...dann“ Regel erklärt werden können. Es werden jedoch keine Abhängigkeiten als Hypothesen angenommen und untersucht, sondern es wird in den Datenbeständen nach solchen mit bestimmten Verfahren gesucht. Erst nachdem ein auffälliges Muster gefunden wurde, wird untersucht ob es sich wirklich um eine Abhängigkeit handelt und falls ja, Assoziationsregeln dazu aufgestellt. Ein Beispiel für die Assoziationsanalyse ist die Warenkorbanalyse. Dabei wird eine Menge von Kaufakten analysiert, um Aussagen über das Käuferverhalten zu entwickeln und sachliche Zusammenhänge aufzudecken. Eine Ergebnis einer Warenkorbanalyse könnte folgendes Muster sein: Wenn Kunden Bier kaufen, kaufen 60 % der Kunden auch Chips[ChCl06].

3.4.1.2 Data Mining Verfahren

Entscheidungsbaumverfahren

Entscheidungsbäume finden sowohl in der Regressionsanalyse als auch in der Klassifikationsanalyse Verwendung. Das Prinzip dieser Bäume ist die Aufspaltung der Trainingsdaten über ein in Klassen aufteilendes, signifikantes Merkmal. Es gibt zwei Verfahrensgruppen, nach denen die Datenmengen geteilt werden. Wenn die Aufteilung anhand nur eines Merkmales vorgenommen wird, spricht man von den klassischen univarianten Verfahren, werden Linearkombinationen der Attribute zur Aufteilung verwendet, spricht man von multivarianten Verfahren.

Jeder Baum besitzt Knoten und Kanten. Die Knoten stellen die Datenmengen dar. Die Kanten repräsentieren die Merkmale, aufgrund derer die Daten geteilt wurden. Während der Wurzelknoten die Gesamtmenge der Trainingsdaten enthält, werden die Blattknoten durch die schließlich Klassen repräsentierenden Daten gebildet.

Es existieren verschiedene Algorithmen, nach denen der Entscheidungsbaum gebildet werden kann. Diese lassen sich in drei verschiedene Verfahrensgruppen aufteilen [ChCI06]:

- concept learning systems (CLS) mit den Algorithmen:
 - ID3-Algorithmus
 - C4.5-Algorithmus
- Automatic-Interaction-Detection-Verfahren mit den Algorithmen:
 - AID-Algorithmus
 - THAID-Algorithmus
 - CHAID-Algorithmus
- Classification and Regression Tree-Verfahren (CART)

Künstliche neuronale Netze

Künstliche Neuronale Netze (KNN) versuchen durch Nachbildung der Arbeitsweise des menschlichen Gehirns Data Mining Aufgaben zu lösen. Künstliche Neuronale Netze sind als Netz unabhängiger Neuronen aufgebaut, welche jeweils in Schichten (Layers) angeordnet sind. Diese Schichten sind nur zu ihren Nachfolgern verbunden. Es gibt eine Eingabe- und eine Ausgabeschicht. Bei manchen Neuronalen Netzen dazwischen noch weitere Verbindungsschichten.

Künstliche Neuronale Netze sind lernfähig, d.h. es werden mit Hilfe von Trainingssets Klassifikationen von Daten erlernt. Dabei existiert eine Grundstruktur, welche die spezifischen Merkmale der Neuronen definiert, erstellt und mit Hilfe von Lernregeln an eine Datenmenge, dem so genannten Trainingsset angepasst. Hierbei werden jedoch meist viele Durchläufe das heißt im weiteren Sinne Trainingssets benötigt, und zu sinnvollen Ergebnissen zu gelangen. Hierbei unterscheidet man, je nach Art der Lernregeln zwischen Vorwärts gerichteten Künstlichen Neuronalen Netzen, welche zur Klassifikation oder Regression angewendet werden, und Kohonen Netzen welche zur Erkennung von Clustern eingesetzt werden [ChCI06].

Clusterverfahren

Clusterverfahren werden zur Bildung von Datengruppen bzw. Segmentierung eingesetzt.

Hierfür müssen als erstes geeignete Ähnlichkeitsmaße, einerseits zwischen den Datensätzen, als auch zwischen den einzelnen Clustern gefunden werden.

Cluster-Verfahren kann zwischen hierarchischen und partitionierende Verfahren unterteilt werden. Beim hierarchischen Verfahren entstehen die Cluster entweder nach dem agglomerativen Verfahren, bei dem in jedem Schritt ausgehend von einzelnen Elementen die ähnlichsten Daten zusammengefasst werden, oder nach dem diversiven Verfahren, bei dem ausgehend von einer Gesamtdatenmenge, in jedem Schritt die unterscheidbarsten Daten geteilt werden. Bei beiden Verfahren kann die Anzahl der Cluster je nach Beschaffenheit der Daten differenzieren. Die partitionierenden Verfahren gehen von einer definierten Clusteranzahl aus, und teilen die Daten in Gruppen größter Ähnlichkeit.

Weitere Ansätze zur Clusterbildung sind Fuzzy-Methoden oder Erwartungs-Maximierungsverfahren[ChCI06].

Verfahren zur Assoziationsanalyse

Mit Hilfe einer Assoziationsanalyse werden Beziehungen zwischen den Daten durch Regeln (Wenn...dann...) beschrieben. Der Supportwert ist die Maßzahl dafür, wie viele Datensätze im Verhältnis zu den Gesamtdaten diese Regel unterstützen. Der Konfidenzwert sagt aus, wie viele Datensätze diese Regel unterstützen, im Verhältnis zu den Datensätzen die nur die Prämisse der Regel unterstützen. Je größer diese Werte, desto bedeutender ist die Regel. Zur Erstellung einer Assoziationsanalyse werden nun für beide Werte ein Minimum definiert und dann mittels dieser Minima alle Regeln erstellt, deren Konfidenz- bzw. Supportwert darüber liegen. Hiermit lässt sich ein einfaches Assoziationsmodell erstellen. Weiterführend kann man dann nur mehr solche Regeln mit einbeziehen, welche besondere Restriktionen oder Taxoniomen berücksichtigen[ChCI06].

Die hier vorgestellten Verfahren werden nun dazu verwendet, um aus den Daten und in weiterer Folge den Informationen, welche im Business Intelligence Prozess zusammengetragen und gesammelt wurden, weitere Erkenntnisse zu gewinnen. Diese Einblicke wären auf Grund der riesigen Datenmengen ohne maschinelle bzw. funktionelle Hilfe nicht ersichtlich. Somit wurde der Weg zu einer noch genaueren Geschäftsprozessanalyse bereitet, welche dazu beitragen soll, dem Unternehmen zum Erfolg zu verhelfen.

4 Zusammenfassung

Der Prozess Business Intelligence bietet viele technische Möglichkeiten große Datenmengen zu sammeln, aufzubereiten, zu speichern und zu verarbeiten. Mit den Techniken des Berichtswesen, der OLAP-Werkzeuge und des Data Mining können Daten aufgrund vieler Gesichtspunkte analysiert und dargestellt werden. Trotzdem braucht jedes Unternehmen als Basis vollständige und sinnvolle Daten. Es hat keinen Wert Unmengen an Daten zu speichern und zu bearbeiten, welche zur Analyse und Prognose von Unternehmenszielen nicht geeignet sind. Ein weiterer relevanter Aspekt, welcher nicht von der Technik übernommen werden kann, ist die Kommunikation im Unternehmen. Wenn zuwenig oder in die falsche Richtung kommuniziert wird, können auch Analysetools nicht zum Erfolg helfen. Hier kann höchstens Unterstützung und Aufdeckung von Fehlverhalten durch die Analyse der Bereiche Mitarbeiter und Kunden gegeben werden. Jedoch sollte man nicht vergessen, dass sich nicht alle relevanten Aspekte für Erfolg und Misserfolg in Zahlen und Fakten ausdrücken lassen. Ein anderer Aspekt ist auch die Problematik mit dem Umgang der Daten. Durch verschiedensten Data Mining Methoden ist es möglich aus jedem Einkauf, jedem Telefonats oder jeden Bestellung Rückschlüsse auf das spezifische Verhalten des einzelnen, identifizierbaren Kunden zu ziehen. Es stellt sich die Frage, ob der gläserne Kunde sich mit den Grundsätzen des Datenschutzes im Einklang steht. Es stehen somit alle technischen Fortschritte im Business Intelligence Bereich gegenüber gestellt zu seiner vernünftigen, wohlüberlegten Nutzung.

5 Literaturverzeichnis

- [Ahle01] *Andrea Ahlemeyer-Stubbe*: Data Warehousing. IBAI-Report 7/2001
<http://www.ahlemeyer-stubbe.de/artikel/ibai0700.htm>
Letzter Abruf: 19.06.2007
- [Bode06] *Freimut Bodendorf*: Daten und Wissensmanagement. Zweite aktualisierte und erweiterte Auflage, Springer Verlag Berlin Heidelberg 2003,2006.
- [ChGl06] *Peter Chamoni, Peter Gluchowski*: Analytische Informationssysteme. Business Intelligence – Technologien und Anwendungen. 3., vollständig überarbeitete Auflage, Springer Verlag Berlin Heidelberg New York 2006.
- [Copo07] *Controlling Portal WWW*, Kennzahlensysteme:
<http://www.controllingportal.de/Fachinfo/Kennzahlen/Kennzahlen-Systeme.html>
Letzter Abruf: 19.06.2007
- [Copo07a] *Controlling Portal WWW*, Balanced Scorecard:
<http://www.controllingportal.de/Fachinfo/BSC/Balanced-Scorecard.html>
Letzter Abruf: 19.06.2007
- [GrBe99] *Heinz Lothar Grob, Frank Bensberg*: Das Data-Mining-Konzept. Institut für Wirtschaftsinformatik der westfälischen Wilhelms-Universität Münster 1999.
WWW: <http://miami.uni-muenster.de/servlets/DerivateServlet/Derivate-217/CGC8.pdf>
Letzter Abruf: 19.06.2007
- [GrGe00] *Martin Grothe, Peter Gentsch*: Business Intelligence – Aus Informationen Wettbewerbsvorteile gewinnen. Addison & Wesley, München 2000.
- [HaNe05] *Hans Robert Hansen; Gustaf Neumann*: Wirtschaftsinformatik 1. Grundlagen und Anwendungen. 9. Auflage, Lucius & Lucius Verlagsgesellschaft mbH, Stuttgart 2005.
- [HäRa05] *Theo Härder; Erhard Rahm*: Datenbanksysteme. Konzepte und Techniken der Implementierung. 2. Auflage, Springer Verlag Berlin Heidelberg 2001.
- [HuWi05] *Bernhard Humm Frank Wietek*: Architektur von Data Warehouses und Business Intelligence Systemen. Informatik Spektrum Nr.23Februar 2005 WWW:
http://www.sdm.de/web4archiv/objects/download/fachartikel/sdm_pub_is_humm_dwh.pdf. letzter Abruf am 10. Juni 2007
Letzter Abruf: 19.06.2007
- [Inmo02] *W.H. Inmon*: Building the Data Warehouse. Third Edition, Wiley Computer Publishing, New York 2002.
- [KeMU06] *Hans-Georg Kemper, Walid Mehanna, Carsten Unger*: Business Intelligence; Grundlagen und praktische Anwendungen. 2. Auflage, Viweg Verlag Wiesbaden 2004.
- [KiRo02] *Ralph Kimball; Margy Ross*: The Data Warehouse Toolkit. The Complete Guide to Dimensional Modeling. Second Edition, Wiley Computer Publishing, New York 2002.

- [Knob00] *Bernd Knobloch*: Der Data Mining-Ansatz zur Analyse betriebswirtschaftlicher Daten. Otto-Friedrichs-Universität Bamberg 2000.
- [Knob05] *Christian Knobloch*: Business Intelligence-Komponenten. Theorie, Realisierung, Entwicklungsprozess. W3L Verlag, Bochum 2005.
- [Nylu99] *Andrea L. Nylund*: Tracing the BI Family Tree, Knowledge Management, 1999
WWW: <http://www.visionassociates.com/FamilyTree.pdf>
Letzter Abruf: 19.06.2007
- [Oehl06] *Karsten Oehler*: Corporate Performance Management mit Business Intelligence Werkzeugen. Hanser Verlag 2006.
- [OLAP01] Vortrag Data Warehouse Technologieren: OLAP und Data Mining. 2001 : WWW: http://www.witi.cs.uni-magdeburg.de/iti_db/lehre/dw/dw0001/dw09.pdf
Letzter Abruf: 19.06.2007
- [ORDI03] *ORDIX AG*, WWW: http://www.ordix.de/ORDIXNews/2_2003/avs_abb3.gif
Letzter Abruf: 17.06.2007
- [PiRW98] *Arnold Picot, Ralf Reichwald, Rolf T. Wigand*: Die grenzenlose Unternehmung. Information, Organisation und Management. Lehrbuch zur Unternehmensführung im Informationszeitalter. Wiesbaden: Gabler 1998 WWW: <http://www.andranet.de/>
Letzter Abruf: 16.06.2007
- [PREU02] *Sarah Preuschhoff*: Business Intelligence – Gegenstand, Ansätze und Technologien. Arbeitspapiere Wissensmanagement Nr. 3/2002
WWW: <http://www.iuk.hdm-stuttgart.de/nohr/Km/KmAP/BusinessIntelligence.pdf>
Letzter Abruf: 17.06.2007
- [SaSi06] *Sai Sumathi, S.N. Sivanandam*: Introduction to data mining and its applications. Springer Verlag Berlin 2006.
- [Seit06] *Tamara Seitz* : Bakkalaureatsarbeit. Modellierung eines DWH für den Personalbereich der Wirtschaftsuniversität Wien. Wirtschaftsuniversität Wien 2006.
- [SGBS04] *Thomas Schildhauer, Martin Grothe, Matthias Braun, Matthias Schulze*: Business Intelligence Durch eBusiness Strategien und Prozesse verbessern. 1. Auflage BusinessVillage GmbH 2004.
- [TuMu05] *H. Krcmar*: Vortrag: Informationsmanagement
WWW: http://www.winfobase.de/lehre/lv_materialien.nsf/intern01/E395A94250203CB7C1256FE000310DA9/
Letzter Abruf: 19.06.2007
- [Wiki07] Wikipedia: Balanced Scorecard. WWW: http://en.wikipedia.org/wiki/Balanced_Scorecard, Letzter Abruf: 19.06.2007

6 Abbildungsverzeichnis

Abbildung 1: Entscheidungsunterstützungssysteme im Zeitverlauf nach [HuWi05]	7
Abbildung 2: Gegenüberstellung von operativen und dispositiven Daten nach [KeMU06] ...	11
Abbildung 3: ETL-Prozess nach [ORDI03]	12
Abbildung 4: unterschiedliche Modellebenen nach Inmon [Inmo02]	14
Abbildung 5: Conformed Dimensions nach [KiRo03]	15
Abbildung 6: Star Schemas nach [KiRo03]	16
Abbildung 7: Snowflake Schema nach [KiRo03]	16
Abbildung 8: Kommunikation im Unternehmen	17
Abbildung 9: Data Cube mit 3 Dimensionen	21
Abbildung 10: Pivotierung des Data Cube	21
Abbildung 11: Roll-up & Drill Down	22
Abbildung 12: Selektion aus den Daten mit Hilfe der Slice-Operation	22
Abbildung 13: Selektion aus den Daten mit Hilfe der Dice-Operation	23
Abbildung 14: Die 4 Perspektiven der Balanced Scorecard	24
Abbildung 15: Einordnung von Data Mining nach [Knobl00]	26
Abbildung 16: Data Mining Aufgaben und Verfahren im Überblick [ChCl06],[Lust02]	27
Abbildung 17: Klassifikation nach [GrBe99]	28
Abbildung 18: Clustering nach [GrBe99]	29