



Hochschule für Angewandte Wissenschaften Hamburg  
*Hamburg University of Applied Sciences*

# **Ausarbeitung Anwendungen 1 WiSe 2010/11**

Jason Hung Vuong  
Business Intelligence

## Inhaltsverzeichnis

<b>1</b>	<b>Einleitung</b>	<b>3</b>
1.1	Motivation . . . . .	3
1.2	Gliederung . . . . .	4
<b>2</b>	<b>Business Intelligence</b>	<b>4</b>
2.1	Business Intelligence Definition und Aufbau . . . . .	4
2.2	Data Warehouse . . . . .	6
2.2.1	Data Mart . . . . .	7
2.2.2	Architekturen . . . . .	7
2.3	Analyseverfahren . . . . .	9
2.3.1	Freie Datenrecherche . . . . .	10
2.3.2	Data Mining . . . . .	10
2.3.3	Online Analytical Processing (OLAP) . . . . .	10
<b>3</b>	<b>Risiken</b>	<b>11</b>
3.1	Risikofaktoren und Gegenmaßnahmen . . . . .	11
<b>4</b>	<b>Zusammenfassung und Ausblick</b>	<b>12</b>
4.1	Zusammenfassung . . . . .	12
4.2	Ausblick . . . . .	12
	<b>Literatur</b>	<b>14</b>

# 1 Einleitung

In diesem Kapitel wird die Motivation, die hinter dieser Arbeit steckt erläutert und eine kurze Übersicht zum Aufbau der Arbeit gegeben.

## 1.1 Motivation

In jedem Unternehmen sind heute zu Tage laufend Entscheidungen zu treffen. Einige dieser Beschlüsse haben große Auswirkungen auf die Zukunft des Unternehmens. Es ist aus diesem Grund wichtig, dass die Entscheidungen auf fundierten Informationen basieren. Zu diesem Zweck werden in Firmen viele Daten gespeichert, welche daraufhin spezifisch für das Ziel aufbereitet und analysiert werden.

Es lässt sich beispielweise das Kundenverhalten analysieren um das Produktangebot zu optimieren oder aber auch Ressourcenauslastungen bestimmen. Eine grafische Veranschaulichung ist dabei möglich.

Die gesammelten Daten können dabei auch Informationen enthalten, welche nicht offensichtlich sind. Es kann somit neues Wissen aufgedeckt werden, was daraufhin zur Strategieverbesserungen eingesetzt werden kann.

Das Problem bei großen Unternehmen besteht darin, dass die Daten auf verschiedene Standorte verteilt sein können und evtl. auch nicht alle Daten digital erfasst sind. Es werden heute zur Analyse und Aufbereitung der Daten daher sogenannte Business Intelligence Systeme bzw. BI-Suites eingesetzt. Diese extrahieren Daten aus Produktivsystemen, bereinigen und speichern sie in einer Form ab, welche für ein spezifisches Analyseziel besonders geeignet ist. Die Analyse läuft dann unabhängig von den Produktivsystemen und beinträchtigen diese demnach nicht.

Business Intelligence Systems können in Ihrem Umfang variieren. Die Architektur beeinflusst dabei maßgeblich die Performanz des Systems. Aus diesem Grund ist es wichtig sich mit dem Entwurf des Systems bzw. dem Kern des Systems zu befassen. Der Schwerpunkt dieser Ausarbeitung liegt demnach darin Data Warehouse Architekturen zu betrachten, um damit die Basis zu schaffen ein performantes System in Betrieb zu nehmen.

Ziel ist es im späteren Projektverlauf, im Rahmen eines Kooperationsvertrags mit dem Energieversorger "Hamburg Energie", ein solches Business Intelligence System mit einer geeigneten Architektur zu entwerfen, aufzusetzen und daraufhin die verschiedenen Anbieter von Business Intelligence Lösungen zu vergleichen und deren Stärken und Schwächen zu analysieren.

## 1.2 Gliederung

Im 2. Kapitel wird zunächst der Begriff Business Intelligence genauer erläutert. Daraufhin wird in Kapitel 2.2 auf die Kernkomponente eines Business Intelligence Systems eingegangen, dem Data Warehouse. Dabei wird der Aufbau erläutert und die verschiedenen Architekturvarianten mit ihren Vor- und Nachteilen beschrieben. Im Anschluss werden die verschiedenen Analyseverfahren im Kapitel 2.3 unterschieden, die auf den Daten des Data Warehouses operieren. Das bereits erwähnte Data Mining aber auch OLAP sind hierbei die Schlüsselwörter. Im Anschluss wird im Kapitel 3 auf mögliche Risikofaktoren eingegangen und eventuelle Gegenmaßnahmen erläutert. Abschließend folgt im Kapitel 4 eine kurze Zusammenfassung und ein Ausblick mit Themenbereichen, die interessant sein könnten für das anschließende Projekt.

## 2 Business Intelligence

In diesem Kapitel wird zunächst der Begriff Business Intelligence näher erläutert und im darauf Folgenden auf das Data Warehouse eingegangen und dessen verschiedene Architekturen. Abschließend wird in diesem Kapitel auf Analyseverfahren eingegangen.

### 2.1 Business Intelligence Definition und Aufbau

Der Begriff Business Intelligence wird im Allgemeinen als Sammelbegriff für Techniken zur Konsolidierung, Analyse und Bereitstellung von Daten zur Entscheidungsunterstützung verstanden [uDS10].

Die analytischen Techniken werden dabei wie in der Abbildung 1 sichtbar ist auf verschiedene Datenbanken angewendet, um Muster herauszufinden und Managern und Angestellten des Unternehmens somit Hilfestellung bei der Entscheidungsfindung zu bieten und dadurch die Rentabilität zu steigern.

Aus der Abbildung 1 wird deutlich wie eine Business Intelligence System aufgebaut ist. Dem Business Intelligence Ordnungsrahmen sind die externen und operativen Quellsysteme vorgelagert auf die nicht genauer eingegangen wird. Das System besteht aus verschiedenen Komponenten, welche in Schichten untergliedert sind.

Die unterste Ebene, die Datenbereitstellung, hat die Aufgabe aus den Quellsystemen konsistente, stimmige Daten zu generieren und adäquat abzulegen. Man unterscheidet dabei zwischen strukturierten und unstrukturierten Daten. Strukturierte Daten werden in der Regel mit Hilfe von Data Warehouse Konzepten bereitgestellt worauf im folgenden Kapitel etwas näher

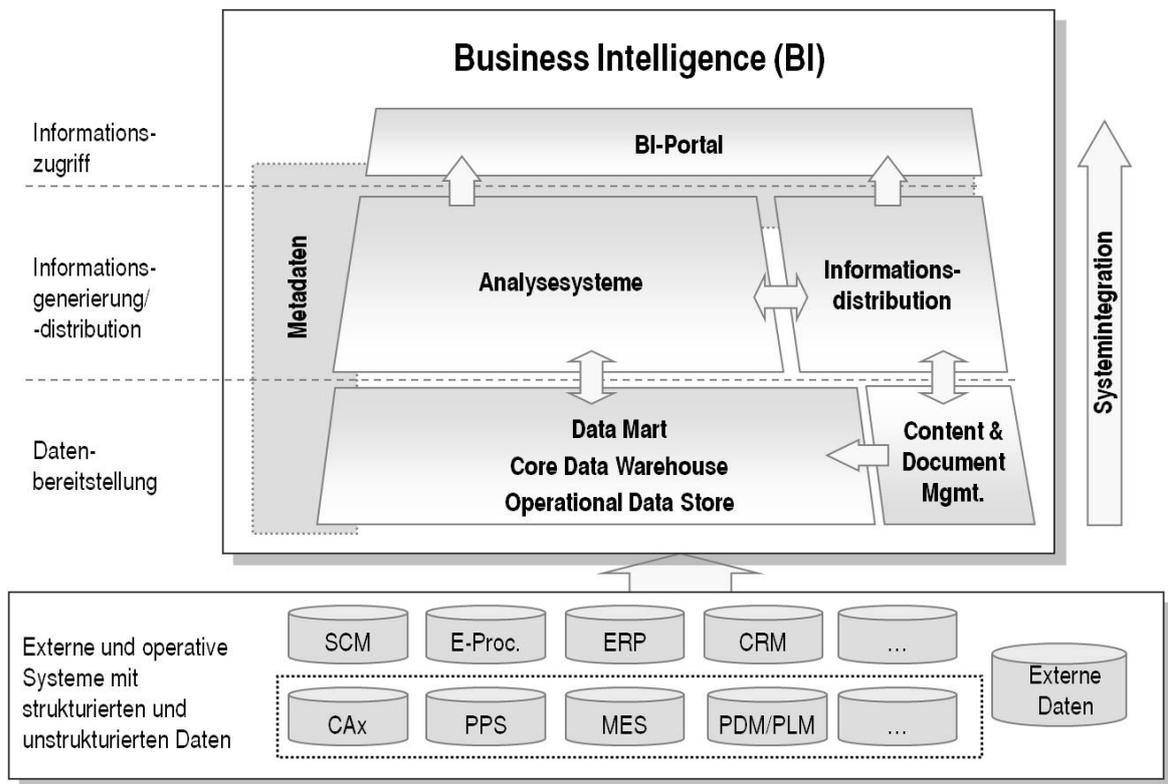


Abbildung 1: Business Intelligence Ordnungsrahmen  
[uDS10]

eingegangen wird. In neueren Ansätzen ist auch ein sogenannter Operational Data Store (ODS) integriert, welche eine Vorstufe eines analytischen Data Warehouses darstellt. Unstrukturierte Daten werden im Gegensatz zu strukturierten Daten dagegen in sog. Content- und Document Management Systemen vorgehalten. Content Management Systeme (CMS) binden dabei Informationsobjekte in beliebiger elektronischer Darstellungsform, wie Daten, Texte, Graphiken, Bilder oder Audio-/Videsequenzen. Hierbei erfolgt eine strikte Trennung von Inhalt Layout und Struktur. Document Management Systeme (DMS) ermöglichen es (eingescannte) digitalisierte Dokumente bereitzustellen. Dabei werden die Dokumente archiviert, verschlagwortet und versioniert.

Die Mittlere Schicht beinhaltet die Informationsgenerierung und -distribution. Im Kontext der Informationsgenerierung lassen sich unterschiedliche Systeme abgrenzen, die sich im Grad ihrer Anwendungsausrichtung, der Nutzungsfrequenz, der erforderlichen IT-Kompetenz, der Benutzer und der Form der Nutzungsinitiierung unterscheiden. Die Schnittstelle zwischen genannten Analysesystemen und den unstrukturierten Daten ist dabei die Komponente der Informationsdistribution. Sie stellt sicher, dass die im Business Intelligence-Kontext erzeug-

ten Daten archiviert werden und ermöglicht es Analysesysteme mit den unstrukturierten Daten anzureichern.

Auf oberster Ebene findet der Informationszugriff statt. Hier sind komfortable Benutzerschnittstellen erforderlich, um die vielfältigen steuerungsrelevanten Informationen abrufen zu können. Der Zugriff erfolgt hierbei in der Regel über sogenannte Portale, die dem Benutzer über das Firmen-Intranet einen zentralen Einstiegspunkt für verschiedene Analysesysteme bietet.

## 2.2 Data Warehouse

Man sah am Aufbau des Business Intelligence Systems in Abbildung 1 des vorhergehenden Kapitels, dass die Datenaufbereitung und Speicherung die Grundvoraussetzung für leistungsfähige Analysesysteme sind, da sie auf diesen Daten operieren. Zuvor wurde auch zwischen unstrukturierten und strukturierten Daten unterschieden. Dabei wurde erwähnt, dass strukturierte Daten nach dem Data-Warehouse-Konzept abgelegt werden. Es stellt sich jedoch zunächst die Frage was eine Data-Warehouse ist.

Ein Data Warehouse ist nichts anderes, als eine Datenbank zur dauerhaften Verwaltung analyserelevanter Daten [uDS10]. Das bedeutet, dass die enthaltenen Daten nicht gelöscht werden, wodurch eine Historie entsteht. Die enthaltenen Daten sind dabei speziell für Analysezwecke ausgelegt, wodurch das Datenbank-Schema auch von den speziellen Analyseprozessen abhängt und nach denen definiert wird. Die Schwierigkeit liegt jedoch auch genau an diesem Punkt, da es komplex ist ein Datenbankschema zu entwerfen, welches für alle Analysezwecke geeignet ist.

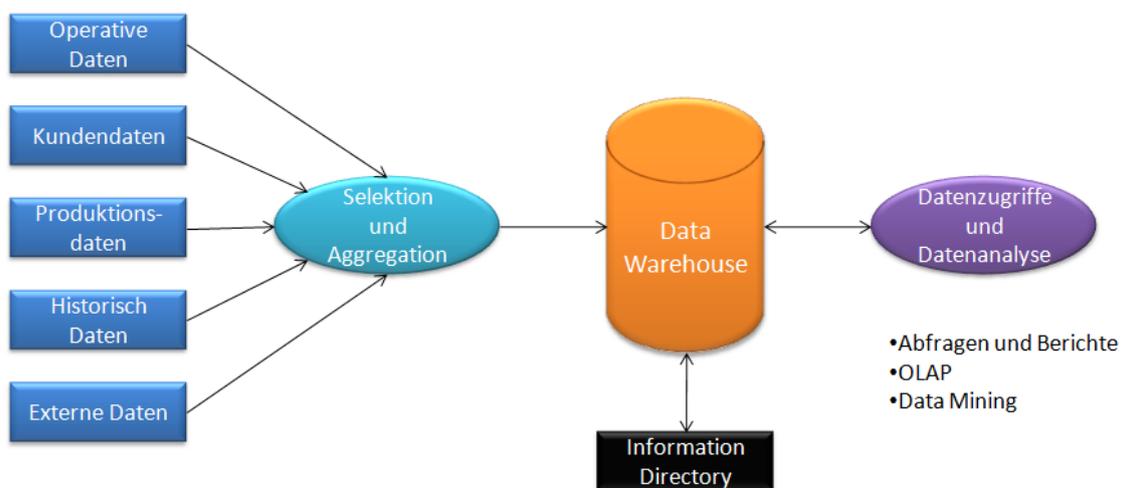


Abbildung 2: Komponenten eines Data Warehouse

Die Daten stammen wie bereits erwähnt auch verschiedenen Quellen eines Unternehmens. Ein Data Warehouse extrahiert dabei sowohl aus historischen und aktuellen Beständen die Daten und kombiniert diese auch mit Externen Daten. Nach der Neustrukturierung der Daten werden sie im Data Warehouse zentral zusammengefasst und stehen daraufhin für Analysen und die Entscheidungsfindung zur Verfügung. Das Information Directory gibt dabei Managern, Auskunft über die im Data Warehouse gespeicherten Daten. Das Buch nach Laudon [uDS10] bietet zu diesem Themenbereich eine kompakte Beschreibung.

### 2.2.1 Data Mart

Ein Begriff der dem Data Warehouse sehr nahe steht lautet Data Mart. Es handelt sich dabei um ein kleines Data Warehouse [uDS10], die nur einen Teil der Unternehmensdaten für eine spezielle Funktion oder Gruppe von Benutzern, beispielweise einer Abteilung bereitstellen.

Die Verwendung Data Marts ist in bestimmten Situationen einfach erforderlich, wie beispielsweise bei Daten, die Datenschutzrichtlinien unterliegen. Gehaltspersonaldateien sollten beispielsweise nicht für alle Mitarbeiter zugänglich sein. Organisatorische Gründe, wie z.B. um die Unabhängigkeit zwischen Abteilungen zu gewährleisten wäre ein weiterer Grund der für die Nutzung von Data Marts spricht. Data Marts können zudem, da sie nur Teildaten des Data Warehouses enthalten, die Fülle der Daten auf ein analyserelevantes Minimum reduzieren, was sie zugleich performanter macht. Da die Analyseprozesse beim Data Mart Einsatz auch nicht alle auf eine zentrale Datenbank zugreifen, verteilen sie außerdem die Last. Ein weiterer Vorteil, ist die Eigenständigkeit dieser, somit müssen sie nicht den Aktualisierungszyklen des zentralen Data Warehouses folgen.

### 2.2.2 Architekturen

Die Architektur eines Informationssystems beschreibt die einzelnen Systembausteine hinsichtlich ihrer Art, ihrer funktionalen Eigenschaften und ihres Zusammenwirkens [uWM10]. Daher ist es nicht überraschend, dass auch Data Warehouse-/Data-Mart-Architekturvarianten existieren, bei denen die Vorteile von den Data Marts ausgespielt werden können. Es werden in diesem Unterkapitel nun ausgewählte Architekturen vorgestellt und die Vor- und Nachteile diskutiert. Das Buch nach Bauer [Grü04] beschäftigt sich ausschließlich mit Data Warehouse Systemen, jedoch bietet auch das Buch nach Kemper [uWM10] eine gute Beschreibung dieser.

#### Unabhängige Data Marts

Die dezentrale Architektur mit unabhängigen Data Marts ist dann sinnig, wenn keine übergreifenden Auswertungen nötig sind, was jedoch selten der Fall ist. Die Daten sind dabei

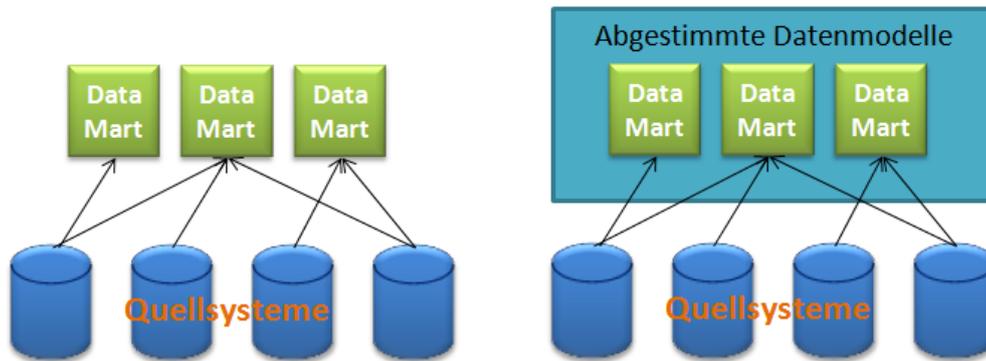


Abbildung 3: Unabhängige Data Marts (links) und unabhängige abgestimmt Data Marts (rechts)

für jeden Anwendungsbereich isoliert, was eine übergreifende Auswertung verhindert. Wenn solche Architekturen vorzufinden sind, handelt es sich meist um eine historisch gewachsene Struktur. Der Nachteil daran besteht darin, dass die Data-Marts sich überlappen und dadurch eine Mehrfachaufbereitung stattfindet. Eine Verbesserung der Unabhängigen Data Marts, liegt dann vor wenn die Datenmodelle der Data Marts abgestimmt sind. Die Daten werden zwar auch hier mehrfach aufbereitet, doch wird dadurch die Konsistenz und Integrität der Daten sichergestellt. Data Marts sind applikationsnah und stoßen daher schnell an ihre Grenzen.

#### Zentrales Core Data Warehouse

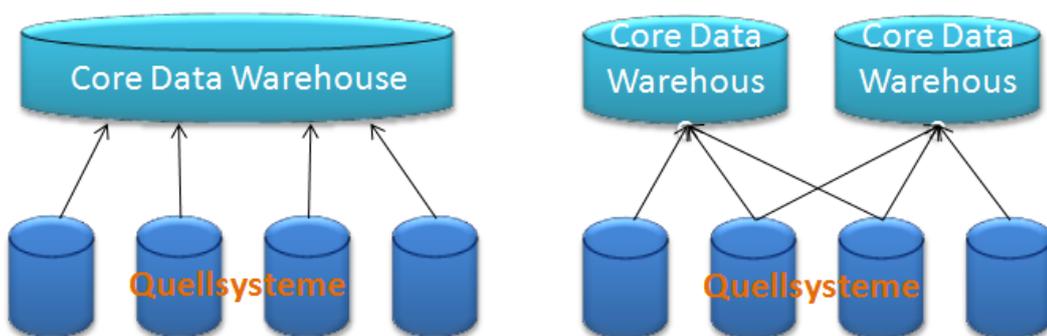


Abbildung 4: Zentrales Core Data Warehouse (links) und der Einsatz mehrerer Core Data Warehouses (rechts)

Der Verzicht auf Data Marts kann eine alternative für kleinere Lösung sein. Die Verwendung eines Core-Data Warehouses (C-DWH) ist nur für eine überschaubare Anzahl von Endbenutzern geeignet und nur für ein geringes Datenvolumen, da die Performance sonst leidet

und die Administration erschwert ist. Gibt es im geschäftlichen Unternehmensumfeld mehrere unterschiedliche Produkt- oder Marktstrukturen, können auch mehrere C-DWH eingesetzt werden. Dies kommt häufig bei spartenorientierten Unternehmen zum Einsatz.

### Core Data Warehouse mit abhängigen Data Marts und Data Warehouse Architekturmix

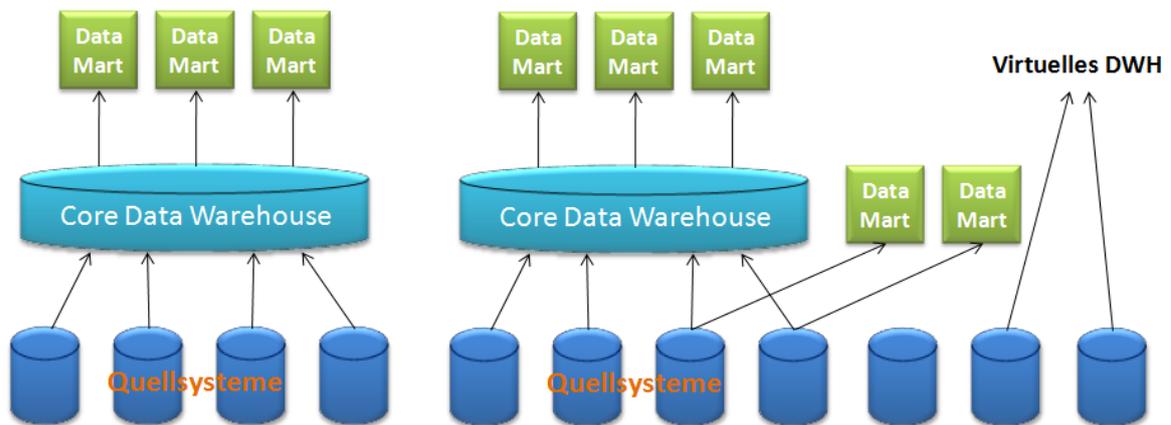


Abbildung 5: Core Data Warehouse mit abhängigen Data Marts (links) und Data Warehouse Architekturmix (rechts)

Ein zentrales Data Warehouse, führte zu Performanceproblemen und Administrationschwierigkeiten. Die Erweiterung um abhängige Data Marts behebt dies. Die Data Marts sind mit dem C-DWH Daten konsistent und die Daten werden nur einmalig zentral aufbereitet. Ein weiter Vorteil liegt darin dass durch die Lastverteilung mehrere Benutzer auf die Daten zugreifen können ohne Performanceeinbußen zu haben. Hier kommen die im vorherigen Kapitel erwähnten Data Mart Vorteile zur Geltung. In der Praxis ist am häufigsten ein Mix der Architekturen zu finden, die aus C-DWH, abhängigen und unabhängigen Data Marts besteht. Es sind auch sogenannte virtuelle Data Warehouses vorhanden, von welchen man spricht wenn man direkte Datenzugriffe durchführt.

## 2.3 Analyseverfahren

In den vergangenen Kapiteln wurde beschrieben was Business Intelligence ist und die Schichten eines Business Intelligence Systems erläutert. Daraufhin wurde erklärt wie auf der untersten Schicht, der Datenbereitstellungsschicht, die strukturierten Daten in das Data Warehouse extrahiert werden und insbesondere wurde auf die Architektur des Data Warehouses eingegangen. Es wurde auch erwähnt, dass das Data Warehouse die Basis für die Analyseverfahren ist. In diesem Kapitel wird nun auf die Analyseverfahren näher eingegangen, welche auf den Daten des Data Warehouses operieren.

### 2.3.1 Freie Datenrecherche

Ein Analyseverfahren ist die freie Datenrecherche. Hierbei wird die Datenmanipulationssprache (DML) verwendet um Teildaten zu recherchieren. Es lassen sich Informationen bestimmen, wie zum Beispiel die Anzahl der verkauften Produkte des Typs X in einer bestimmten Kalenderwoche. Der Vorteil darin besteht, dass direkte Anfragen generell performanter sind. Durch inkompetente Recherchen kann die Performanz jedoch auch stark sinken. Ein weiterer Nachteil ist, dass hierfür IT-Kompetenzen vorausgesetzt werden.

### 2.3.2 Data Mining

Analyseverfahren die dem Begriff Data Mining zuzuordnen sind, haben als Ziel aus bestehenden Daten durch verschiedene Techniken verborgene Muster und Beziehungen aufzudecken, woraus Regeln abgeleitet werden können für ein zukünftiges Verhalten [uDS10].

Bei der Methode des Clustering, werden vorhanden Informationen in mögliche Gruppen, in sogenannte Cluster zusammengefasst und bei der Klassifikation Informationen einer vorhandenen Klasse zugeordnet. Assoziationsanalysen untersuchen welche Beziehungen zwischen einzelnen Merkmalsausprägungen bestehen und die Abweichungsanalyse hat das Ziel stark abweichende Merkmalsausprägungen aufzuspüren. Ereigniszusammenhänge und deren Auswirkungen werden durch Regressionsanalyse ermittelt.

### 2.3.3 Online Analytical Processing (OLAP)

OLAP ist eine Technik um Daten nach mehreren Dimensionen bzw. aus mehreren Perspektiven zu analysieren [uDS10]. Der Benutzer legt hierbei zu Beginn alle zu betrachtenden Kriterien fest und kann dann die Daten, welche man sich in Form eines Dimensionswürfels vorstellen kann aus verschiedenen Sichten betrachten.

Mit Pivotierung ist es möglich die Reihenfolge von Spalten und Zeilen zu ändern und mit Rotationen kann der Blickwinkel auf den aktuellen Ausschnitt des Würfels geändert werden. Um die Granularität zu verändern, werden die Operationen Roll-Up und Drill-Down verwendet, wobei Roll-Down das Ergebnis fein-granularer macht und Roll-Up grob-granularer. Bezogen auf ein Filialnetze, würde es bedeuten dass man zwischen Bezirkssicht und Ländersicht wechseln kann.

Die Operation Drill-Across führt zum Wechsel des betrachteten Würfels. Es werden dann beispielsweise Umsätze statt Verkäufe betrachtet. Das Drill-Through führt dagegen zum Austausch der physikalischen Datenbasis. Einschränkungen von mehreren Dimensionen/Ebenen

werden als Dice bezeichnet, wohingegen die Einschränkung einer Dimension/Ebene Slice genannt wird.

Die Ausarbeitung meines Kommilitonen Andrew Tarasenko befasst sich detailliert mit diesem Thema.

## 3 Risiken

Es soll wie bereits in der Motivation erwähnt ein Business Intelligence System eingeführt werden. In diesem Kapitel werden mögliche Risiken und deren Gegenmaßnahmen genauer erläutert.

### 3.1 Risikofaktoren und Gegenmaßnahmen

Man hat gesehen, dass über den Entwurf des Data Warehouses, der Extraktion der Daten und die Analyse eine gewisse Zeit nötig ist, bis man Ergebnisse vorweisen kann. Dies stellt ein Risiko dar, da man sich früh festlegen muss welche Daten relevant sind. Man muss bei Fehlentscheidungen die Schemata des Data Warehouses und Data Marts evtl. überarbeiten, was zusätzliche Zeit in Anspruch nimmt.

Was außerdem Unsicherheiten bietet ist die schlechte Informationslage bezüglich der "Hamburg Energie". Es ist noch nicht ganz klar, was konkret für das Unternehmen umgesetzt werden soll und in welchem Umfang. Um dem entgegen zu wirken ist eine frühzeitige Absprache mit den Verantwortlichen notwendig.

Desweiteren sind echte Kunden-Testdaten für das zu aufzusetzende Business Intelligence System nur bedingt für Testzwecke verfügbar, da diese speziellen Datenschutzrichtlinien unterliegen. Um dennoch ein mit dem Originalsystem funktionierendes System entwerfen zu können, muss genau in Erfahrung gebracht werden, wie die Datensätze aussehen und dafür eigene Testdaten generiert werden.

## 4 Zusammenfassung und Ausblick

In diesem Kapitel wird kurz zusammengefasst, was in den vorherigen Kapiteln beschrieben wurde und zudem ein Ausblick auf das geben womit man sich im Anschließenden Projekt befassen könnte.

## 4.1 Zusammenfassung

Es wurde deutlich, dass Geschäftsentscheidungen eine schwierige Angelegenheit sind und Business Intelligence die Entscheidungsfällung durch hochwertige, aufbereitete Daten erheblich unterstützt und dadurch die Qualität der Entscheidungen erhöht, was dem Unternehmenserfolg zu Gute kommt.

Der Kern der Analyseverfahren stellen die Daten des Data Warehouses dar, wobei die Architektur eine wichtige Rolle für die Performanz des Systems spielt. Es ist dabei wichtig zu wissen für welche Analysezwecke das System entworfen werden soll.

Mit den aufbereiteten Daten können diverse Analysen durchgeführt werden, die neue Erkenntnisse bringen.

## 4.2 Ausblick

Es existieren verschiedene Business Intelligence Suites wie beispielsweise

- SAP BusinessObjects [[SAP](#)]
- Cognos 10 - Intelligence Unleashed [[IBM](#)]
- Oracle Business Intelligence 11g [[Ora](#)]
- Pentaho [[Pen](#)]

Einer der Lösungen soll für den Energiekonzern Hamburg Energie eingesetzt werden. Womit man sich beschäftigen könnte wäre zum Einen, zu untersuchen worin sich die Business Intelligence Lösungen unterscheiden und ob und wie Erweiterungen sinnvoll möglich sind. Daraufhin könnte man Gedanken darüber verlieren, welche Lösung für einen Energiekonzern wie Hamburg Energie am geeignetsten ist und welche Data Warehouse Architektur am effizientesten wäre.

Interessant wäre auch sich damit zu befassen wie generell ein solches Informationssystem in eine bestehende Informationssystemlandschaft ohne großen Aufwand zu integrieren ist. Desweiteren kann es auch darum gehen, herauszufinden in wie weit solche Lösungen nicht nur den Unternehmenserfolg steigern, sondern auch die Umwelt entlasten durch geschickteren Einsatz von Ressourcen.

## Literatur

- [Baa08] Hans-Georg Kemper; Henning Baars. *Business Intelligence - Arbeits- und Übungsbuch*. Vieweg & Sohn Verlag, 1. edition, 2008. ISBN 978-3-8348-0434-1.
- [Ene] Hamburg Energie. Hamburg energie. <http://www.hamburgenergie.de/>. [Online Stand 26. Februar 2011].
- [Grü04] Andreas Bauer; Holger Grünzel. *Data-Warehouse-Systeme*. dpunkt Verlag, 2. edition, 2004. ISBN 3-89864-251-8.
- [IBM] IBM. Cognos 10 - intelligence unleashed. <http://www-01.ibm.com/software/de/data/cognos/>. [Online Stand 26. Februar 2011].
- [Ora] Oracle. Oracle business intelligence 11g. <http://www.oracle.com/oms/businessintelligence11g/>. [Online Stand 26. Februar 2011].
- [Pen] Pentaho. Pentaho. <http://www.pentaho.com/>. [Online Stand 26. Februar 2011].
- [Roh10] Sebastian Rohjans. Eine standardkonforme und semantikbasierte kommunikationsarchitektur für die automationsebene in smart grids. [http://www.offis.de/f\\_e\\_bereiche/energie/gruppen/](http://www.offis.de/f_e_bereiche/energie/gruppen/), 2010. [Online Stand 26. Februar 2011].
- [SAP] SAP. Sap businessobjects. <http://www.sap.com/germany/solutions/sapbusinessobjects/>. [Online Stand 26. Februar 2011].
- [uDS10] Kenneth C. Laudon; Jane P. Laudon und Detlef Schoder. *Wirtschaftsinformatik*. Pearson Studium, 2. edition, 2010. ISBN 978-3-8273-7348.
- [uDT09] Markus Grünwald und Dirk Taubner. Informatik spektrum 32 5 2009 - business intelligence. <http://www.springerlink.com/content/j110174l25161p7r/>, 2009. [Online Stand 26. Februar 2011].
- [uWM10] Hans-Georg Kemper; Henning Baars und Walid Mehanna. *Business Intelligence Grundlagen und praktische Anwendungen*. Vieweg + Springer Verlag, 3. edition, 2010. ISBN 978-3-8348-0719-9.