

Data Warehousing

Anton Michlmayr
e9925744@student.tuwien.ac.at

Mai 2002

1 Die reale Praxis

In diesem Kapitel wird beschrieben wie Data Warehousing in der Praxis tatsächlich eingesetzt wird. Zuerst wird näher erklärt, welche Konzeptionen sich durchgesetzt haben. Danach werden typische Einsatzmöglichkeiten und -bereiche beschrieben und auch ein detailliertes Beispiel für ein Data Warehouse vorgestellt. Schlussendlich wird ausführlich auf die Klippen von Data Warehouse-Projekten eingegangen, Gründe für das Scheitern solcher Projekte genannt und erklärt wie man ein solches Scheitern verhindert kann.

1.1 Akzeptanz und Durchsetzungsbedingungen

Das ursprüngliche Konzept eines Data Warehouses war das unternehmensweite, zentrale Data Warehouse, in dem sich alle unternehmensrelevanten dispositiven Daten befinden sollten. Dieser Ansatz ist als gescheitert zu betrachten. Wesentliche Gründe dazu sind: (siehe [MuBe98])

- Ein zentrales Data Warehouse erfordert ein unternehmensweites Datenmodell. Aus den Zeiten von CASE (Computer Aided Software Engineering) weiß man aber, daß ein solches Modell im allgemeinen nicht realisierbar ist.
- Versucht man jedoch ein solches Datenmodell aufzubauen, so mündet dies in ein langwieriges Projekt, bei dem kurz- und mittelfristig keine Ergebnisse sichtbar werden. Das macht die Fachabteilungen unruhig, und man baut an der Informatik vorbei eigene, isolierte Punktlösungen in Form von Data Marts.
- Gelingt es dennoch, ein zentrales, umfangreiches Data Warehouse als Annäherung an das unternehmensweite Data Warehouse zu bauen, so findet eine solche Kompromisslösung in der Regel keine Akzeptanz bei den Fachabteilungen.

Als Alternative zu dem unternehmensweiten, zentralen Data Warehouse haben sich so die sogenannten Data Marts entwickelt. Unter Data Marts kann man sich eine Art "kleines Data Warehouse" vorstellen, wobei mehrere solche Data Marts dann zu einem globalen Data Warehouse zusammengefasst werden. In einem großen Unternehmen könnte so zum Beispiel jede Abteilung ein eigenes Data Mart betreiben, wobei unternehmensweit dann durch das globale Data Warehouse auf alle Data Marts zugegriffen wird.

Weitere Entwicklungen, die bereits vorgestellt wurden, sind die virtuellen und die verteilten Data Warehouses. Letztere haben in jüngster Vergangenheit vor allem durch das Internet immer mehr an Bedeutung gewonnen.

1.2 Einsatzbereiche und "typische" Nutzung

In diesem Kapitel wird näher auf die möglichen Einsatzbereiche von Data Warehousing eingegangen. Dabei wird jedoch, wie bereits im Kapitel über Konzepte und Technik von Data Warehousing, der technische Hintergrund nur kurz erklärt und das Hauptaugenmerk auf nicht-technische Aspekte gelegt. Zu klären ist also wo, wie und warum Data Warehouses eingesetzt werden.

Das folgende Fallbeispiel stellt eine mögliche typische Nutzung eines Data Warehouses dar. Aufgrund der Anschaulichkeit des Beispiels wurde es komplett übernommen. (siehe [Lu02] Seite 133)

1.2.1 Fallbeispiel eines Data Warehouses

Eine Einzelhandelskette besteht aus Dutzenden von Filialen, die alle Tausende von Produkten führen. Die meisten Produkte werden mit Bar Codes erfasst, die von POS-Systemen gelesen werden. Mehrere operative Systeme verwalten neben den Produktdaten Verkaufs-, Liefer-, Lager- und Marktforschungsinformationen.

Motive für die Einrichtung eines Data Warehouse sind einerseits die Integration verschiedenster Datenquellen in einer analytischen Datenbank und andererseits die Schaffung besserer Grundlagen für die bereits eingesetzten Data Mining-Werkzeuge in der Marktforschung. Die Unternehmensleitung verspricht sich insbesondere fundierte Antworten auf Fragen wie "Welche Werbeaktion hat sich prozentual am stärksten auf den Umsatz ausgewirkt?" oder "Wo wird eine bestimmte Produktkategorie am besten angeboten?".

1.2.2 Detailliertes Beispiel eines Data Warehouses

Die französische Eisenbahngesellschaft SCNF (Societe Nationale des Chemins de Fer Francais) ist die größte und bedeutendste Eisenbahngesellschaft in Frankreich. Sie befördert jährlich 800 Millionen Passagiere und 135 Millionen Tonnen an Gütern durch Frankreich, England und Belgien.

Um die eingesetzten Ressourcen optimal zu nutzen und die Preise zu bestimmen bzw. das Verkehrsaufkommen im Vorhinein abschätzen zu können, entschloss man sich zur Implementation eines Data Warehouses. Ein weiterer Grund dafür war, dass das vorher eingesetzte Reservierungssystem nicht in der Lage war eine derartig hohe Anzahl an Zügen effizient zu verwalten und darauf aufbauend Schlüsse auf das zukünftige Verkehrsaufkommen zu ziehen. Das Management forderte ein Passagiermanagementsystem, das in der Lage ist, die Züge optimal zu füllen indem auf populäre Strecken die Preise erhöht wurden, sowie versucht wurde Kunden auf weniger populären Strecken mit Sonderangeboten anzulocken.

Mit dem aufgebauten Data Warehouse, dessen Gesamtumfang nur vage mit dem Terabyte-Bereich angegeben wird, ist das Management der SCNF nun in der Lage sehr detaillierte Informationen über die in Anspruch genommenen Leistungen festzuhalten:

- Wie viele Tickets wurden für welche Strecke gekauft?
- Wo wurden diese Tickets erworben?
- Wie teuer waren diese Tickets?
- Wurden Mahlzeiten angefordert?
- Wurden Mietwagen angefordert?
- u.s.w.

Mit Hilfe dieser Daten kann das Reservierungssystem nun, ähnlich wie bei Fluggesellschaften, vorhersagen wie viele Leute voraussichtlich einen Zug reservieren und wie viele davon nicht erscheinen werden. Dementsprechend können dann die Plätze überbelegt werden.

Die detaillierten technischen Hintergründe dieses Data Warehouses werden dem Leser erspart und wurden überdies von der SCNF auch nicht preisgegeben. Bekannt ist jedoch, dass die Daten auf einem massiv parallelen System verarbeitet werden. Laut SNCF spart sich das Unternehmen jährlich rund 11 Millionen Dollar durch den Einsatz ihres Data Warehouses.

1.2.3 Weitere Einsatzbereiche

Data Warehousing wird im Großen und Ganzen eigentlich überall eingesetzt wo mit größeren Mengen an Daten hantiert wird. Ein sehr häufiges Einsatzgebiet ist das Management von Kundendaten, worauf später im Kapitel über Datenschutz noch eingegangen wird.

- Das Kaufverhalten bzw. die Vorhersage, wie Kunden auf Produkte reagieren ist ein enorm wichtiger Faktor in der Wirtschaft. Daher ist auch eine **Analyse des Kaufverhaltens** besonders wichtig, um

darauf aufbauend versuchen zu können genau dieses Kaufverhalten der Kunden zu manipulieren. Ein Beispiel für eine solche Manipulation sind die vielen Online-Shops im Internet. Hierbei wird jeder Einkauf jedes Kunden gespeichert um dann einerseits auf das Kaufverhalten des Kunden Rückschlüsse ziehen zu können. Das kann dann zum Beispiel dazu führen, dass diesem Kunden später ähnliche oder ergänzende Produkte zu diesem bereits getätigten Kauf angeboten werden. Andererseits kann aus der Summe aller Einkäufe wertvolle Information erhalten werden, welche Produkte oft zusammen erstanden werden. Diese werden dann entweder gruppiert angeboten oder es wird beim Einkauf eines Produktes auf die Existenz des anderen Produktes hingewiesen. Als gutes Beispiel hierfür dient der Online-Shop von www.amazon.com. Wenn man dort eine bestimmte CD kaufen will, so wird einem sofort auf andere Alben dieser Band oder auf Bands mit ähnlicher Musikrichtung aufmerksam gemacht.

- **”Database Marketing”**: Aufgrund des immer mehr wachsenden Wettbewerbs- und Konkurrenzdruck machen sich Firmen seit vielen Jahren Gedanken um ihre Prozesse effizienter und kundenfreundlicher zu gestalten. Ziel all dieser Gedanken soll sein die Kunden individuell und bedarfsgerecht bedienen zu können. Dies bindet einerseits ”alte” Kunden an das Unternehmen, da sie mit den angebotenen Leistungen zufrieden sind, andererseits bewirkt es, dass neue Kunden gewonnen werden.

Vor einigen Jahren genügte es noch eine einfache Kundendatenbank mit Name und Adresse zu führen um die Angebote an die potentielle Kunden zu bringen. Durch den immer größer werdenden Wettbewerb ist es aber nun notwendig viele Daten über Kunden zu besitzen und ihnen individuelle Angebote machen zu können. Dieses Ziel kann durch den Einsatz eines Data Warehouses erreicht werden. Für weiterführende Informationen zum Thema ”Database Marketing” verweise ich auf [MuBe98] Seite 477 - 492.

- **”Controlling”**: Ein sehr wichtiger Einsatzbereich für das Data Warehousing ist das ”Controlling”. Das Ziel dabei ist, aus allen vorhandenen Daten eines Unternehmen, Zusammenhänge zu erkennen oder Trends und Entwicklungen vorauszusagen. Konkret soll den Entscheidungsträger - also dem oberen Management - eines Unternehmen bei ihren Entscheidungen geholfen werden.

In diesem Bereich ist jedoch besondere Vorsicht geboten, da eine einzige falsch getroffene Entscheidung das gesamte System in Verruf bringen kann und somit die Akzeptanz bzw. das Vertrauen der Benutzer sofort sinkt. Weiterführende Informationen mit Einsatzbeispielen im Bereich des ”Controlling” sind in [MuBe98] Seite 555ff zu finden.

1.3 Scheitern und Rettungsversuche

Im folgenden Kapitel wird näher auf die Klippen von Data Warehouse-Projekten eingegangen. Dabei werden typische Gründe für das Scheitern solcher Projekte beschrieben und kurz darauf eingegangen, wie so etwas vermieden werden kann.

1.3.1 Performance

Performance ist ein wichtiges Kriterium für Data Warehouse-Projekte, da die meisten Data Warehouses aus riesigen Datenmengen (Terabyte-Bereich) bestehen. Die Erstellung von Berichten oder das Ergebnis einer Abfrage sollte in möglichst kurzer Zeit erfolgen, andererseits wird das Data Warehouse vom Anwender nicht akzeptiert werden. Die Optimierung eines Data Warehouses stellt sich aber als schwieriges Problem dar, weil viele analytische Datenbanken viel umfangreicher als ihr operativen Quellen sind:

- Data Warehouses enthalten nicht nur laufenden, sondern auch historische Daten.
- Es werden zusätzlich auch Ableitungen und Zusammenfassungen gespeichert.
- Sie stellen nicht nur Daten für bekannte Anwendungen bereit, sondern auch Information zur Lösung unbekannter Probleme. (z.B. explorative Daten für Data Mining-Verfahren)
- Sie enthalten neben interne Daten auch Information aus externen Quellen.

Dennoch ist eine Optimierung in allen Fällen mit mehr oder weniger großem Erfolg möglich. Im folgenden werden Beispiele erläutert, wie Data Warehouses effizient gestaltet werden können, um möglichst optimale Performance zu erzielen.

- **Archivierung:** Nicht mehr benötigte, sogenannte "schlafende" Daten (engl. dormant data) werden archiviert. Überwachungsprogramme (Monitore) protokollieren alle Abfragen mit und halten fest mit welcher Häufigkeit Daten noch verwendet werden. Bei der Archivierung handelt es sich um die ergiebigste Optimierungsmaßnahme.
- **Modularisierung:** Die Unterteilung eines Data Warehouses in mehrere Data Marts ist ebenfalls der Performance dienlich. Es kann zum Beispiel versucht werden Routine- und explorative Benutzer nicht auf den selben Data Marts arbeiten zu lassen, da sie sich gegenseitig in ihren Abfragen behindern.

- **Denormalisierung:** Bei dieser Optimierungsmaßnahme werden Tabellen dritter oder höherer Normalform in die 2. Normalform transformiert. Das verschlechtert zwar die Speichereffizienz und Fortschreibungsfreundlichkeit, verbessert aber die Antwortzeit merklich. Beispiele für Denormalisierung sind Vorberechnung, Spaltenreplikation und der Vorverbund.
- **Partitionierung:** Dabei wird das physische Datenmodell so modifiziert, dass sich die Laufzeit- oder Speichereffizienz bei Ein- und Ausgabeoperationen verbessert. Die Zeilenpartitionierung bildet neue Tabellen, die aus den gleichen Spalten, jedoch aus verschiedenen Zeilen besteht während die Spaltenpartitionierung dafür sorgt, dass gleiche Attributwerte zusammen gespeichert werden.
- **Indexierung:** Die Adresse eines Datensatzes wird entweder durch ein Inhaltsverzeichnis oder durch eine Hashfunktion ermittelt, womit dann direkt statt sequentiell auf den gesuchten Satz zugegriffen werden kann. Die Indexisierung wirkt sich positiv auf die Abfrageeffizienz aus, der Speicheraufwand steigt jedoch mit zunehmender Zahl der indizierten Attribute.
- **Mehrprozessorsysteme:** Die Last wird auf mehrere Prozessoren verteilt, die Daten werden also parallel verarbeitet. Während bei den symmetrischen Mehrprozessorsystemen (SMP) wenig Prozessoren eng miteinander verbunden werden, wird bei den massiv parallelen Mehrprozessorsystemen (MMP) eine nur lose Koppelung von vielen Prozessoren samt deren Peripherie vorgenommen.
- **Cluster:** SMP-Systeme werden oft über eine Hochgeschwindigkeitsverbindung zu sogenannten Clustern verknüpft, wobei die einzelnen Knoten auf ihre lokalen Festplatten und einen gemeinsamen Externspeicher zugreifen. Cluster sind besser skalierbar und fehlertoleranter als ein einzelnes SMP.

1.3.2 Datenschutz und Datensicherheit

Prinzipiell sind Datenschutz und Datensicherheit immer wichtige Punkte sobald es um das Speichern von Daten geht. Vor allem bei sogenannten "sensiblen" Daten (unter die zum Beispiel Personendaten fallen) ist auf diese Aspekte besonderer Wert zu legen.

Man stelle sich vor, ein grosses Multi-Unternehmen besteht aus Tochter-Unternehmen unterschiedlichster Branchen (Sportartikel, Bücher- und Zeitschriften, Lebensmittel, Banken, Versicherungen,...) und speichert alle Kundendaten in einem riesigen Data Warehouse. Werden diese Daten dann miteinander verknüpft und analysiert kann auf das Kaufverhalten jeden einzelnen geschlossen werden. Es ist sicher nicht im Interesse des Kunden, wenn

dieses Unternehmen dann genau protokolliert, wann er oder sie Geld abhebt, wo diese Transaktionen durchgeführt werden oder welche Produkte erstanden werden. Noch schlimmer wäre es, wenn diese Daten bzw. daraus resultierende Rückschlüsse auf das Kaufverhalten an dritte weitergegeben wird. Für weiterführende Informationen verweise ich auf [MuBe98] Seite 101 - 124.

1.3.3 Weitere Gründe

Gründe für das Scheitern von Data Warehouse-Projekten: (siehe [Ku99] Seite 306ff und [Dorsay])

- Das Projektmanagement schätzt den Gesamtaufwand falsch ein oder die zu erreichenden Projektziele werden überschätzt bzw. die Erwartungshaltungen der Anwender werden zu unrealistisch. Hierbei ist zu beachten, daß ein Data Warehouse nicht das Allheilmittel für alle betrieblichen wie organisatorischen Probleme in einem Unternehmen ist, sondern nur als Entscheidungsunterstützung dienen soll.
- Die Qualität der extrahierten, operativen, internen und externen Daten ist mangelhaft. Die darauf aufbauenden Auswertungen liefern ungenaue Ergebnisse und führen zu falschen Rückschlüssen. Dieser Effekt kann unter dem Stichwort "Garbage In Garbage Out" zusammengefaßt werden. Die Datenhygiene ist ein wichtiger Baustein, um ein Data Warehouse auch erfolgreich umsetzen zu können.
- Einsatz von proprietären analytischen Werkzeugen. Nur die Konzepte von offenen, generischen und plattformübergreifenden Systemen sind im Umfeld des Data Warehousings erfolgversprechend.
- Das erste Data Warehouse eines Unternehmen wird sofort unternehmensweit geplant und durchgeführt. Besser ist zuerst ein kleines Pilot-Projekt für einen Teilbereich des Unternehmens zu starten um erst dann ein größeres Projekt zu initiieren.
- Bei der Auswahl des Front-End Tools wird hauptsächlich auf dessen Preis und nicht auf Benutzerfreundlichkeit Wert gelegt. Data Warehouse-Projekte sind zumeist so umfangreich und komplex, dass die Kosten des Front-End Tools (das über Erfolg und Misserfolg entscheiden kann) keine Rolle spielen.
- Nach Abschluß der Start- bzw. Projektevaluierungsphase herrscht noch immer kein Konsens über die Projektziele, Vorteile bzw. Nutzen des Projektes, Gewichtung des funktionalen Umfangs, oder es wurde kein offizieller Projektauftrag erstellt.
- Die Projektziele werden nach Ende der Startphase in Frage gestellt.

- Die Evaluierung der am Markt erhältlichen Werkzeuge wird entweder aus Zeitgründen mangelhaft oder aber zu umfangreich durchgeführt.
- Erweiterung des funktionellen Umfangs. Neue Anforderungen werden noch zusätzlich (zumeist von den Anwendern gefordert) in das Projekt mit aufgenommen, ohne dabei den Projektplan entsprechend zu adaptieren.
- Der Machtpromotor des Projektes geht verloren
- Interne Widerstände gegen das Projekt durch andere Fachabteilungen
- Unter den Projektmitarbeitern macht sich eine negative Stimmung breit, eine allgemeine Resignation verbreitet sich. Es wird an der Machbarkeit bzw. der Erreichung der gesteckten Ziele gezweifelt.

Es sollte noch angemerkt werden, daß viele dieser Punkte auch für "normale" IT-Projekte Gültigkeit haben und nicht nur Konfliktpunkte im Bereich des Data Warehousing sind.

Der Auslöser für das Scheitern von Data Warehouse-Projekten ist auch oft, dass solche Projekte verglichen mit herkömmlichen IT-Projekten andere Bewertungs- und Planungsmethoden bedürfen. Auf diese Methoden wird hier aber nicht näher eingegangen, dem interessierten Leser wird dafür [JoDW01] empfohlen.

1.3.4 Erfolgsrezept für Data Warehouse-Projekte

Demgegenüber stehen folgende Faktoren, die für den Erfolg eines Data Warehouse-Projektes stehen (können). Diese Auflistung ist natürlich weder vollständig noch als immer zutreffend anzusehen da jedes Projekt seine Eigendynamik hat und es nicht möglich ist alle Projekte in eine Schublade zu werfen. Garant für erfolgreiche Data Warehousing-Projekte sind nach [Ku99] (Seite 309ff) folgende Punkte:

- Eine klar definierte Rolle des Data Warehouse-Projektes innerhalb der IT-Architektur des Unternehmens.
- Die Implementierung eines Data Warehouse-Konzepts verpflichtet das Projektmanagement eine Reihe von Einflußfaktoren sicherzustellen. Vor allem muß eine verbindliche (schriftliche) Verpflichtung der obersten Top-Management-Ebene (Top-Level Management Commitment) gegenüber dem Data Warehouse-Projekt eingefordert werden. Weiterhin ist speziell auf die finanzielle Absicherung des Projektes, effiziente Ausnützung aller Personalressourcen. Ganz wesentlich ist in dieser Hinsicht auch eine realistische Zeitplanung des Projektmanagements.

- Eine entsprechende Evaluierung und Auswahl des externen Beraters ist von gleich hoher Bedeutung wie die Evaluierung der Hardware-Architektur und den verwendeten Decision Support-Lösungen.
- Eine gesamtheitlich prozeßorientierte Sicht auf das Data Warehouse. Ein Data Warehouse-Projekt ist weit mehr als nur die geschickte Kombination von leistungsfähiger Hardware und moderner analytischer Software.
- Standards und Open Source-Ansätze sind proprietären Ansätzen vorzuziehen. Der Austausch zwischen den einzelnen Systemkomponenten unterschiedlicher Hersteller muß gegeben sein.

Weitere "Critical Success Factors" für das Bauen eines Data Warehouses sind nach [Dorsay]:

- Der Projektleiter sollte Erfahrung im Data Warehousing haben, da Data Warehouse-Projekte und "normale" IT-Projekte völlig unterschiedlicher Natur sind.
- Die Benutzeranforderungen müssen genau gesammelt und analysiert werden. Es ist wichtig zu wissen, was die User mit ihren Berichten vorhaben.
- Es sollte unbedingt ein Pilotprojekt erfolgreich beendet werden. Erst dann sollte mit dem Start des eigentlichen Projektes begonnen werden.
- Hauptaugenmerk eines solchen Projektes muss die benutzerfreundliche Bedienung sein. Das System muss einfach zu benutzen sein und sinnvolle Resultate liefern um wirklich akzeptiert zu werden. Data Warehouse-Projekte sind viel heikler als andere Projekte. Schon eine einzige falsche (Management-)Entscheidung, die auf dem System basiert kann dazu führen, dass das Vertrauen in das System verloren geht.

2 Bewertung

Am Beginn dieses Kapitels möchte ich darauf hinweisen, daß eine Bewertung von Data Warehouse-Projekten aus mehreren Gründen problematisch ist. Die meiste (vor allem im Internet zu findende) Literatur über beendete Data Warehouse-Projekte besteht aus sogenannten "Success Stories" von diversen Unternehmen, die oft keine wirklich objektive Sichtweise auf ihre eigenen Data-Warehouse-Projekte besitzen. Weiters ist anzumerken, daß es zwangsläufig zu einer Selektion kommt, welche die sogenannten "Failure Stories" - also jene Projekte, die gescheitert sind bzw. ihr Ziel verfehlt haben - verständlicherweise nicht publiziert werden. [SemDW96] Dieser Effekt tritt

jedoch nicht nur bei Data-Warehouse-Projekten auf, sondern ist generell bei der Bewertung von Projekten - welcher Art diese auch immer sein mögen - zu beachten.

Daher kann eine Bewertung von Data Warehousing nicht vollzogen werden, ohne den Themenbereich von mehreren Seiten her zu betrachten und erst dann aus allen daraus erhaltenen Erkenntnissen Schlüsse zu ziehen.

2.1 Data Warehousing - bloßer Hype?

In [Dorsay] ist zu lesen, dass Data Warehousing der eindeutig am meisten überbewertete Hype in der Geschichte der Computerindustrie sei. Der Autor vergleicht die seiner Meinung nach leeren Versprechungen des Data Warehousing mit jenen der CASE-Tools in den achtziger Jahren. Es seien zu viele verschiedene Produkte und vor allem so viele Experten ohne wirkliche Sachkenntnis.

Er bemängelt weiter, dass in der Wirtschaft alle Firmen angeben an mehreren erfolgreichen Data Warehouse-Projekten beteiligt zu sein. Tatsächlich schlagen aber überdurchschnittlich viele solche Projekte fehl. Gründe für dieses Fehlschlagen wurden bereits eingehend erläutert.

Weiters ist zu lesen, dass in den meisten Publikationen - seien es Bücher, Papers oder Präsentationen - sehr optimistisch über Data Warehousing berichtet wird, womit wir ihm nach ausführlicher Recherche auch Recht geben müssen. Die Realität des Data Warehousing zeigt seiner Meinung nach jedoch ein ganz anderes Bild, wobei er von unbefriedigten Benutzererwartungen, verschwendeter Zeit und Geld sowie ruinierten Karrieren spricht.

2.2 Data Warehousing - echte Neuheit?

Prinzipiell ist anzumerken, dass die Notwendigkeit grosse Datenmengen zu verwalten sicher vorhanden ist. Glaubt man den Meinungen von Experten wird dieser Trend auch in Zukunft weiter zunehmen und die Menge der gespeicherten Daten wird explodieren. Dennoch kann Data Warehousing nicht als echte Neuheit vorgestellt werden, da die Idee eines Data Warehouses bzw. eines großen Daten-Pools schon länger existiert. Vielmehr wurden erst in den letzten Jahren die technischen Möglichkeiten geschaffen derartig grosse Datenmengen halbwegs erschwinglich speichern zu können (60 bis 80 GB fassende Festplatten sind für Otto Normalverbraucher heute schon fast Standard).

2.3 Abschließende Bemerkungen

Nach den Angaben von Survey.com liegt die durchschnittliche Menge an Daten in einem Data Warehouse im Jahr 2002 schon bei rund 1,2 TeraByte. Weiters sollen im Jahr 2004 30 mal mehr Daten verarbeitet werden als noch

1999. Unter Anbetracht dieser Zahlen kann angenommen werden, dass Data Warehousing auch in Zukunft eine nicht unbedeutende Rolle spielen wird.

Es wurde bereits darauf hingewiesen, dass Data Warehouse-Projekte anders als herkömmliche IT-Projekte geplant und durchgeführt werden sollten. Es muss unbedingt darauf geachtet werden, dass das System schlussendlich für die User akzeptabel ist und nicht an ihren Wünschen vorbei geht. Auch deren Erwartung vom Data Warehouse sollte von vorne herein realistisch bleiben, indem man ihnen das Data Warehouse nicht als das Allheilmittel alle betrieblichen Probleme verkaufen will.

Sind all diese Faktoren vorhanden und wird ein Data Warehouse-Projekt erfolgreich abgeschlossen, so ist dies sicher ein sehr nützliches Instrument für die Entscheidungsfindung eines Unternehmen.

Literatur

[MuBe98] Harry Mucksch, Wolfgang Behme, Das Data Warehouse-Konzept (3.Auflage), Gabler Verlag, 1998

[Lu02] Markus Lusti, Data Warehousing und Data Mining (2.Auflage), Verlag Springer, 2002

[Ku99] Andreas Kurz, Data Warehousing - Enabling Technology, MITP Verlag, 1999

[SemDW96] Christoph Breitner, Uwe Herzog, Jutta Muelle, Data Warehousing - Seminar im Sommersemester 1996 an der Uni Karlsruhe, <http://ubka.uni-karlsruhe.de/cgi-bin/psview?document=/ira/1996/29>

[Dorsay] Dr. Paul Dorsay, Data Warehouses, Ad Hoc Query Tools and other ways to destroy your company, http://www.dulcian.com/papers_by_topic.html#DataWarehousing

[JoDW01] Reinhard Jung, Robert Winter, Justification of Data Warehousing Projects, IRMA 2001, http://www.iwi.unisg.ch/iwiwebdaten/Publications/RJU_RWI_IRMA2001.pdf