

Einsatz des Enterprise Guide in der Lehre - Data Warehousing zum Anfassen und Explorieren -

Hans-Günter Lindner
FH Frankfurt am Main
Nibelungenplatz 1
60318 Frankfurt am Main
lindner@fb2.fh-frankfurt.de

Zusammenfassung

Im Rahmen der Veranstaltung "Data Warehouse" für Studierende der Wirtschaftsinformatik an der FH Frankfurt am Main wurden vier Methoden zur Vermittlung der Funktionsweise von Data Warehouses genutzt. Im ersten Ansatz wurde ein technisch komplett realisiertes Warehouse der "on demand Lösung" www.crmondemand.de von Siebel zur Verfügung gestellt. Der zweite und dritte "explorative" Ansatz vermittelte Data Warehousing mit Hilfe von MS Excel und humanITs InfoZoom, einem Knowledge Browser zur Erstellung von Lean Warehouses. Im dritten Ansatz wurden Data Warehouse-Funktionalitäten mit Enterprise Guide 2.1 von SAS und dem MDDB Modul erstellt. Der Artikel beschreibt die Erfahrungen der drei Ansätze. Alle drei Methoden weisen Vor- und Nachteile auf. Der Nutzen des SAS Enterprise Guide für die Umsetzung, das Verständnis und die Machbarkeit eines Warehouses werden beschrieben und Regeln für den erfolgreichen Einsatz abgeleitet.

Schlüsselworte: Data Warehouse, Enterprise Guide, MDDB, CRM

1 Ziel der Vorlesung „Data Warehouse“

Zielgruppen der Vorlesung „Data Warehouse“ an der Fachhochschule Frankfurt am Main sind Studenten der Wirtschaftsinformatik ab dem 4. Semester und Studenten im Hauptstudium Informatik. Diese Studenten besitzen Wissen über die Grundlagen der Vorlesung Datenbanken. Die Vorlesung und Übung Data Warehouse an der FH Frankfurt am Main vermittelt Fähigkeiten und Fertigkeiten für den Aufbau von Data Warehouses und Business Intelligence Lösungen, d.h. die Studenten sollen Data Warehouse Konzepte verstehen, sich in praktische Problemstellungen schnell einarbeiten können, komplexe Sachverhalte einfach und verständlich vermitteln können, sich Lösungen schnell erarbeiten können, den Aufbau eines Data Warehouses praktisch durchführen können, in der Praxis sofort kompetent mitarbeiten können und für den Arbeitsmarkt sehr gut positioniert sein.

Die Veranstaltung besteht aus zwei Vorlesungsstunden und zwei Übungsstunden. Das Skript beinhaltet Grundlagen aus [1], [3], [4], [5]. Darüber hinaus kommen Folien von SAS und Inhalte zu EAI, u.a. von SeeBeyond zum Einsatz. Das Skript ist unter <http://www.fh-frankfurt.de/~lindner/> zu finden. Im folgenden wird auf die Konzeption der Übung eingegangen.

2 Didaktischer Aufbau der Übung

In der Übung sollen die Studenten lernen,

1. abteilungsspezifische Kennzahlen zu bestimmen,
2. konzeptuell zu modellieren,
3. Datenbankschemata zu analysieren,
4. Wertebereiche der Attribute zu analysieren,
5. Kennzahlen zu bilden,
6. Kennzahlenmodelle miteinander kombinieren zu können,
7. Fakten und Dimensionen zu extrahieren,
8. konzeptuelle Modell in unterschiedliche logische Modelle zu übertragen,
9. unterschiedliche Softwaresysteme für das Data Warehousing bedienen zu können und
10. die Analyseergebnisse praxisgerecht darstellen zu können.

Dazu wird in einem ersten Schritt das Szenario „Beratung eines KMU“ vorgegeben und die Studenten in Übungsgruppen eingeteilt. Diese Übungsgruppen entsprechen Abteilungen wie z.B. Geschäftsführung, Vertrieb, Vertriebsunterstützung und Kundendienst. Größere Datentabellen werden den Studenten für die Übung werden den Studenten schrittweise zur Verfügung gestellt.

3 Genutzte Softwaresysteme und Lerneffekte

Die Studenten sollen einen Überblick über mehrere Systeme kommerzielle Softwaresysteme für Data Warehousing erhalten und die diese praxisgerecht nutzen können. Um die Probleme der Praxis zu kennen, ist die Vermittlung von Ansätzen des Change Managements wichtig. Deshalb kommen sowohl vorkonfigurierte Systeme mit integriertem Data Warehouse, klassische Werkzeuge als auch Möglichkeiten für die „Erste-Hilfe“ zum Einsatz:

- Siebel crmondemand.de ist ein vorkonfiguriertes CRM-System mit voller Funktionalität und dem integrierten Data Warehouse Siebel Analytics, das von IBM „on demand“ gehostet wird,
- Excel als Erste-Hilfe, das in der Praxis als erster Ansatz für Analysen genutzt wird,
- humanIT InfoZoom als Knowledge Browser für den Überblick und das visuelle Data Mining sowie
- SAS 9 mit dem EnterpriseGuide für den Aufbau von Data Warehouses und die Analyse den Demodaten, das einen Schwerpunkt in den Übungen bildet.

3.1 Siebel/IBM crmondemand.de

crmondemand ist ein Kundenbindungssystem, das über Web direkt als Lösung verfügbar ist. Es basiert auf dem Siebel System 7 sowie dem Siebel Analytics Server. Es vereint typische CRM-Funktionen, die durch ein vorkonfiguriertes Data Warehouse unterstützt werden. Die gesamte Applikation und das Data Warehouse sind individuell veränderbar. Einen Überblick gibt Abbildung 1.

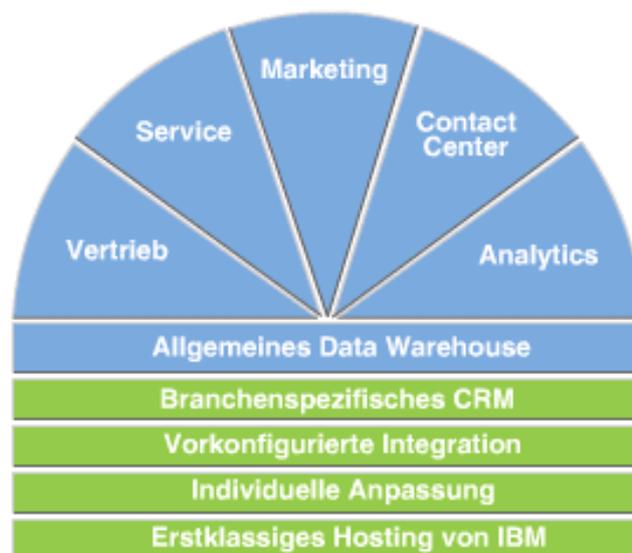


Abbildung 1: Funktionen von crmondemand

Funktionen, die insbesondere im Zusammenhang mit Data Warehouse von Bedeutung sind, sollen intensiv genutzt werden. Diese sind Importwerkzeuge für Kundendaten und Konfiguration neuer Berichte und Dashboards. Abbildung 2 zeigt die Liste

vorkonfigurierter Berichte (im Hintergrund) und das Beispiel des Dashboards, das als interaktiver Bericht kombinierte Listen und Grafiken darstellt. Die Studenten arbeiten jedoch nicht mit einem bereits gefüllten System, sondern müssen die Daten erst schrittweise integrieren.

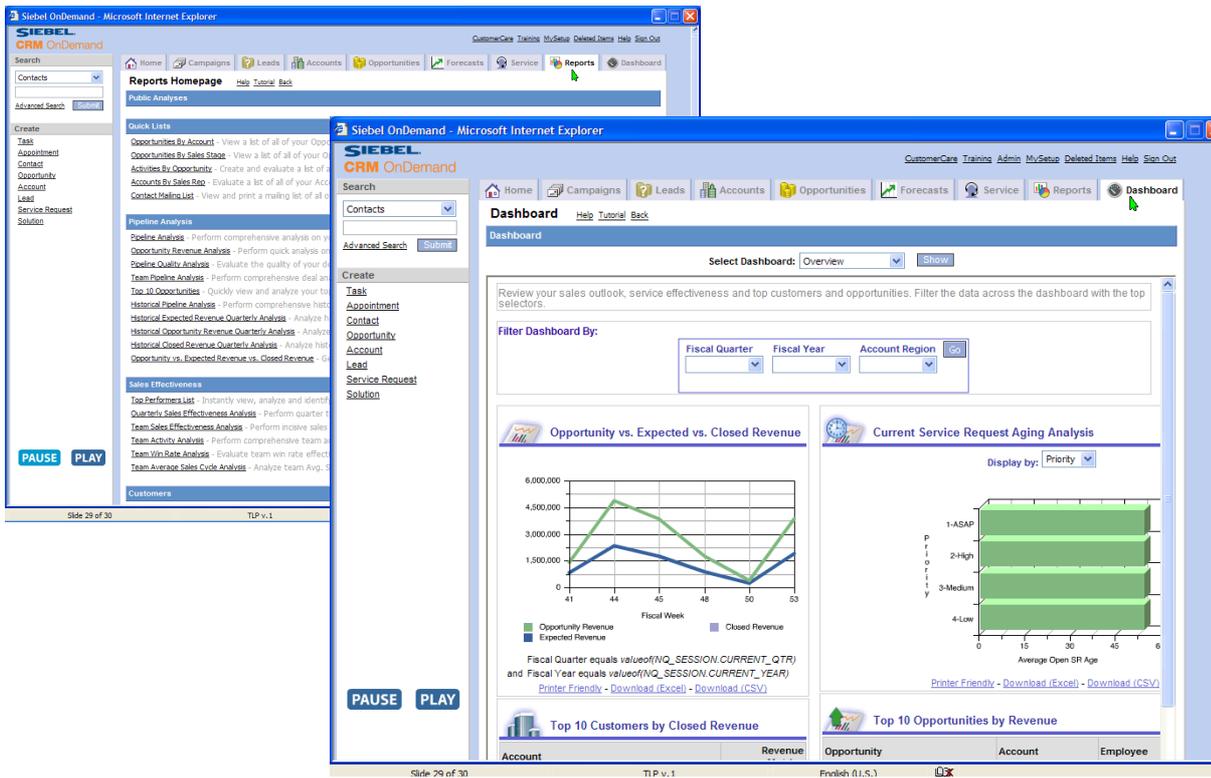


Abbildung 2: Berichtsliste und Dashboard in crmondemand

Probleme des Change Managements werden behandelt, wenn Daten manuell aber koordiniert in das System eingegeben und gemeinsame Kennzahlensysteme mit den zugehörigen Berichten entwickelt werden sollen.

Folgende Lerneffekte mit diesem System ergeben sich:

- Der Aufbau eines vorkonfigurierten Data Warehouses ist sofort erkennbar,
- ein Warehouse ohne Kenntnis der Datenqualität ist nicht sinnvoll,
- Probleme der Datenmenge und Werteverteilung werden erkannt,
- Change-Management-Probleme bei der Umsetzung werden spürbar.

Insbesondere die Nutzung eines nicht mit Daten gefüllten Systems bringt einen Lerneffekt, da zuerst mit Freude die Funktionen erkundet werden, aber die Eingabe und

Strukturierung der Daten sich nicht als einfach herausstellt und ein konzeptionelles Vorgehen erfordert. Die Notwendigkeit des strukturierten Vorgehens wird für die Studenten zum Muss. Das schrittweise Herangehen an größere Datenmengen wird in einem nächsten Schritt mit Hilfe von MS Excel vollzogen.

3.2 MS Excel

MS Excel ist die Standard-Tabellenkalkulation, die in jedem Unternehmen zu finden ist. Analysen werden meist zuerst mit diesem Werkzeug durchgeführt. Die Einschränkungen zeigen sich dann mit zunehmenden Datenmengen und bei der notwendigen Analyse der Wertebereiche der einzelnen Attribute und deren Zusammenhänge. Abbildung 3 zeigt als Hintergrundfenster eine Tabelle mit 29 Attributen und 10.165 Datensätzen (Hintergrund) und eine Pivot-Tabelle mit einer ABC-Kundenbewertung.

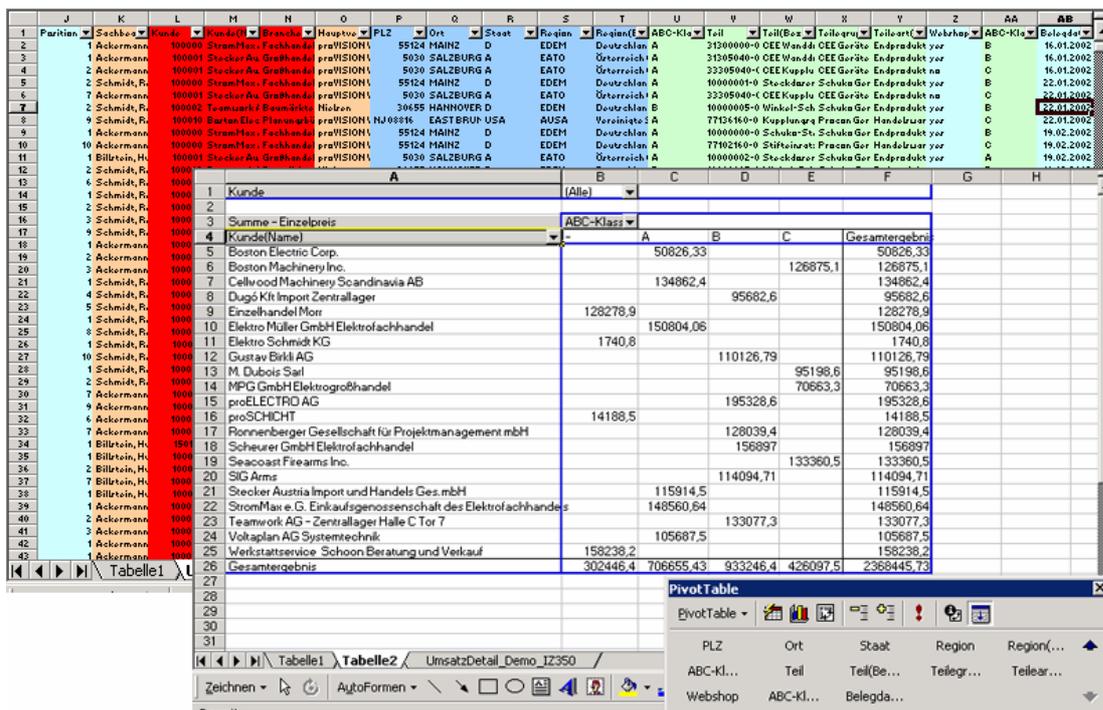


Abbildung 3: Farblich strukturierte Datenbasis und Pivottabelle

Das vorgegebene Szenario und die Gruppeneinteilung bleiben gleich. Aufgabe ist es, die möglichen Dimensionen farblich zu markieren. Eine derartige Markierung findet sich in Abb. 3. Im Anschluss daran soll ein Star-Schema konzipiert und Berichte in

Excel erstellt werden. Es werden alle Schritte mit Ausnahme von Schritt 8 (logische Datenmodelle) durchgeführt

Als Lerneffekte ergeben sich

- Gewöhnung an größere Datenmengen,
- erste Dimensionierungen zu erlernen,
- Nutzung einfacher Werkzeuge wie z.B. Pivot-Tabelle,
- Grenzen herkömmlicher Werkzeugen erkennen und
- MS Office als „Erste-Hilfe-Paket“ zur Datenanalyse kennen lernen.

Es zeigt sich, dass häufig zuerst Formeln und „Auto-Filter“ eingesetzt werden; eine Analyse der Wertebereiche findet bei Anfängern kaum statt, obwohl der Schritt prinzipiell bekannt ist. Kennzahlen und diverse Auswertungen werden vorgenommen, jedoch fehlt häufig die Abstimmung im Team und das gemeinsame Verständnis der Basisbegriffe für die Kennzahlen wie z.B. was ein „A-Kunde“ ist.

Die Aufgabe, ein Star-Schema zu erstellen und die Faktentabelle zu spezifizieren erzwingt das gewünschte schematische Vorgehen. „Kennzahlen“ wurden in der Übung zum wichtigsten Diskussionsthema innerhalb der Lerngruppe.

3.3 humanIT InfoZoom

InfoZoom ist ein Knowledge Browser zur Analyse großer Datentabellen mit mehreren Millionen Datensätzen auf dem PC [2], [6]. Mit Hilfe visueller Komprimierung werden die Merkmale der Attribute einer Tabelle als horizontale Staffeln dargestellt. Importwerkzeuge für Tabellen und Datenwürfel ermöglichen den InfoZoom-Nutzer einen direkten Zugriff auf Datenbanktabellen oder Würfel in Data Warehouses.

Die visuelle Komprimierung ergibt sich wie folgt: Die Breite eines Segments des Staffeln zeigt die Häufigkeit der Datensätze oder den Anteil am Umsatz etc. In Abbildung 4 sehen wir, dass ca. 50% aller Verkäufe in den Regionen Nord und Süd erfolgen und das Produkt MP3onDemand den größten Teil der Verkäufe ausmacht. Die Farbgebung ist für den Umsatz aktiviert (Ampel-Icon in Attribut „Revenue“) und zeigt, dass mit dem meistverkauften Produkt ein geringerer Umsatz im Vergleich zu den anderen erzielt wurde. Ebenso ist sofort ersichtlich, dass Geschäftsführer und Vorstände (CxO) als Zielgruppen lukrativ erscheinen. Insofern sind größere Datenmengen mit InfoZoom leicht zu überblicken und Auffälligkeiten schnell ersichtlich (Visual Data Mining). Ein Klick auf ein Segment eines Staffeln oder auf gewählte Bereiche liefert einen Drill-Down.

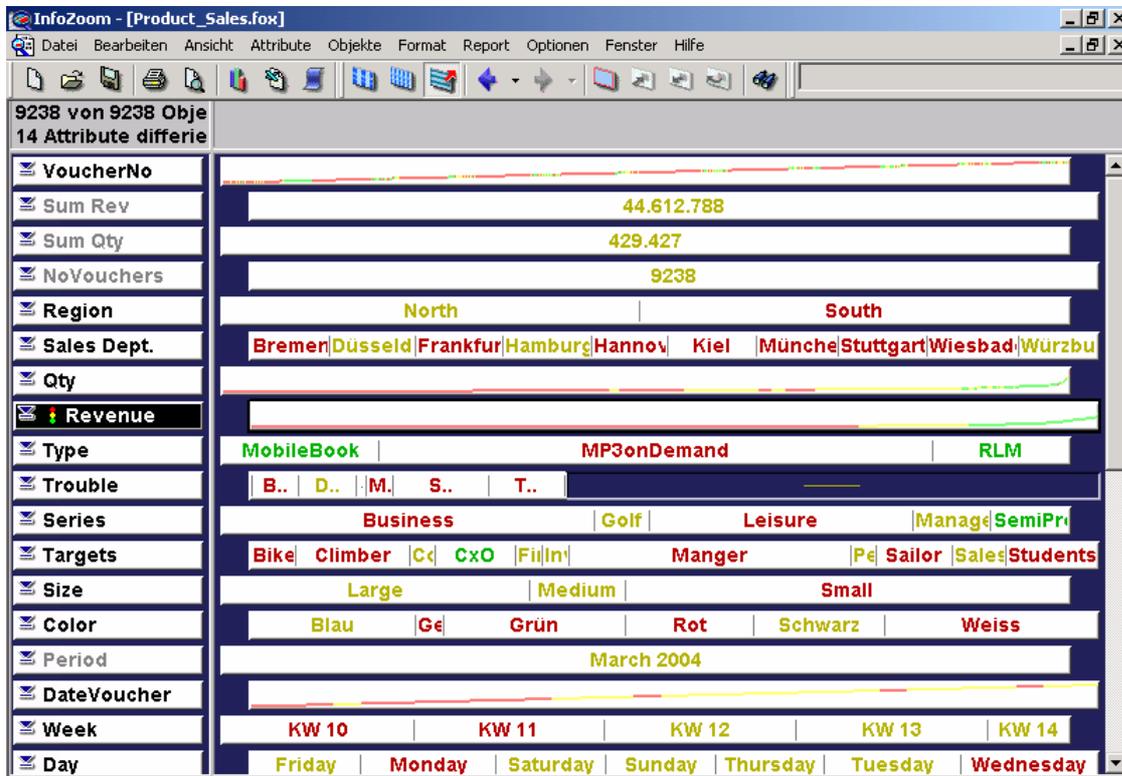


Abbildung 4: Übersicht in InfoZoom

Die komprimierte Ansicht in InfoZoom ermöglicht, die Werteverteilungen von Attributen zu inspizieren. Eine geschachtelte Sortierung (Klick auf einen der Pfeile in der Linken Spalte) nach „Revenue“, „Sales Dept.“ und dann „Region“ zeigt die Abbildung 5. Deutlich erkennbar ist, dass die größten Umsätze in Hamburg zu finden sind. Zusätzlich wurde mit dem Mauszeiger der Bereich der höchsten Mengen in Bremen gewählt und es erscheinen die gleichen Absatzmengen in anderen Verkaufsbüros als vertikaler Balken dunkel markiert. Die Werteverteilung lässt sich somit leicht erschließen.

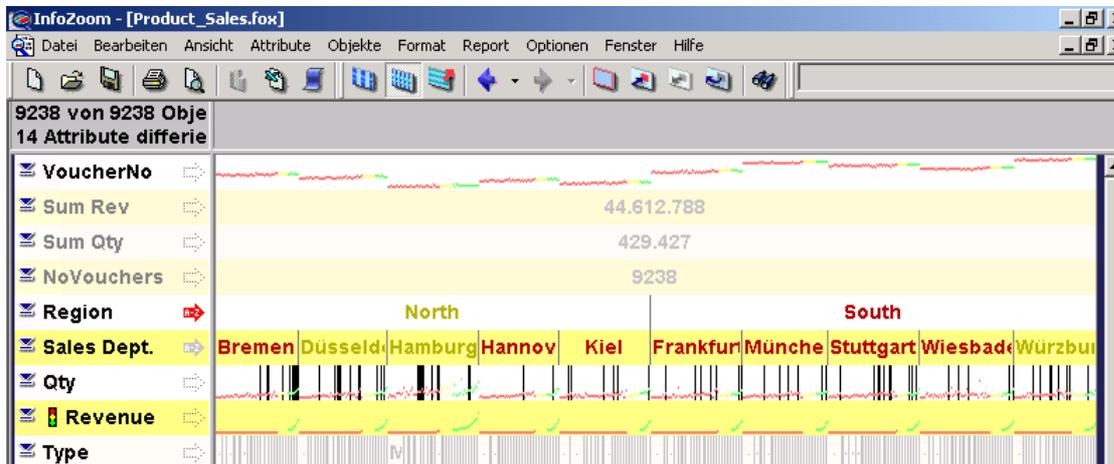


Abbildung 5: Visual Data Mining mit der InfoZoom-Komprimierung

In diesem Teil der Übung bleiben das Szenario und die Gruppen gleich. Die Datentabelle beinhaltet 18 Attribute und 9238 Datensätze. Wiederum werden alle Lernschritte aus Kapitel 2 mit Ausnahme der logischen Modellierung durchgeführt. Lerneffekte sind:

- Leichter Überblick über Daten durch geeignete Visualisierung und Exploration,
- Sichtbarmachen der Werteverteilungen steigert das Verständnis und eine intuitive Einschätzung der Datenqualität,
- Wichtigkeit der konzeptuellen Modellierung trotz leichtem Überblick wird erkannt,
- Angst vor großen Datenmengen wird genommen.

Mit InfoZoom können die Studenten sehr schnell produktiv arbeiten und viele detaillierte Ergebnisse in kurzer Zeit aufbereiten, da InfoZoom über Werkzeuge für das Reporting verfügt. Durch diese hohe Produktivität leidet im ersten Schritt das konzeptionelle Vorgehen. Durch gezielte Fragen wurde das Verständnis für Kennzahlensysteme und das konzeptuelle Modell gefördert.

3.4 SAS Enterprise Guide 2.1

Der Enterprise Guide (EG) ermöglicht einen leichten Einstieg in SAS 9. Die grafische Oberfläche liefert einen projektorientierten Zugang und fördert das mehrfach

geforderte konzeptionelle Denken. Die Studenten werden von Anfang an angehalten, klar strukturiert vorzugehen. Der EG zeigt links die Projektstruktur und die Aktionen auf den Daten. Somit können alle Schritte jederzeit nachvollzogen werden. Änderungen und Verbesserungen sind überschaubar und reproduzierbar. Die Abbildung 6 zeigt im rechten Hauptfenster die importierten Tabellen und eine erste Darstellung eines Zahlenwertes als Balkendiagramm. Als Beispieltabellen wurde die Demodatenbank „Nordwind“ von MS Access genutzt.

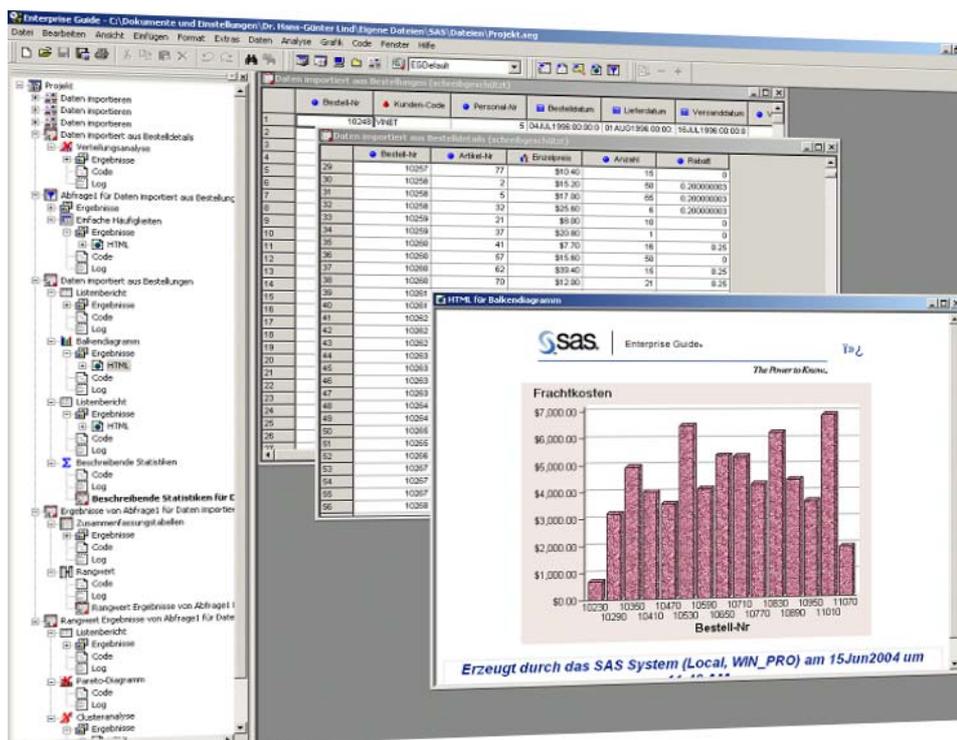


Abbildung 6: Importierte Tabellen im Enterprise Guide

Nach dem Import der Datenbanktabellen wurde die MDDDB-Funktion in SAS 9 [7] genutzt und damit mehrdimensionale Tabellen und interaktive Berichte erstellt (Abbildung 7). Die resultierende Benutzeroberfläche ermöglicht die Ansicht von Tabellen und Grafiken, die mit Drill-Down, Drill-Across etc. detailliert analysiert werden können. Ein tieferer und leicht verständlicher Einblick in die mehrdimensionale Datenmodellierung wird in kurzer Zeit möglich.

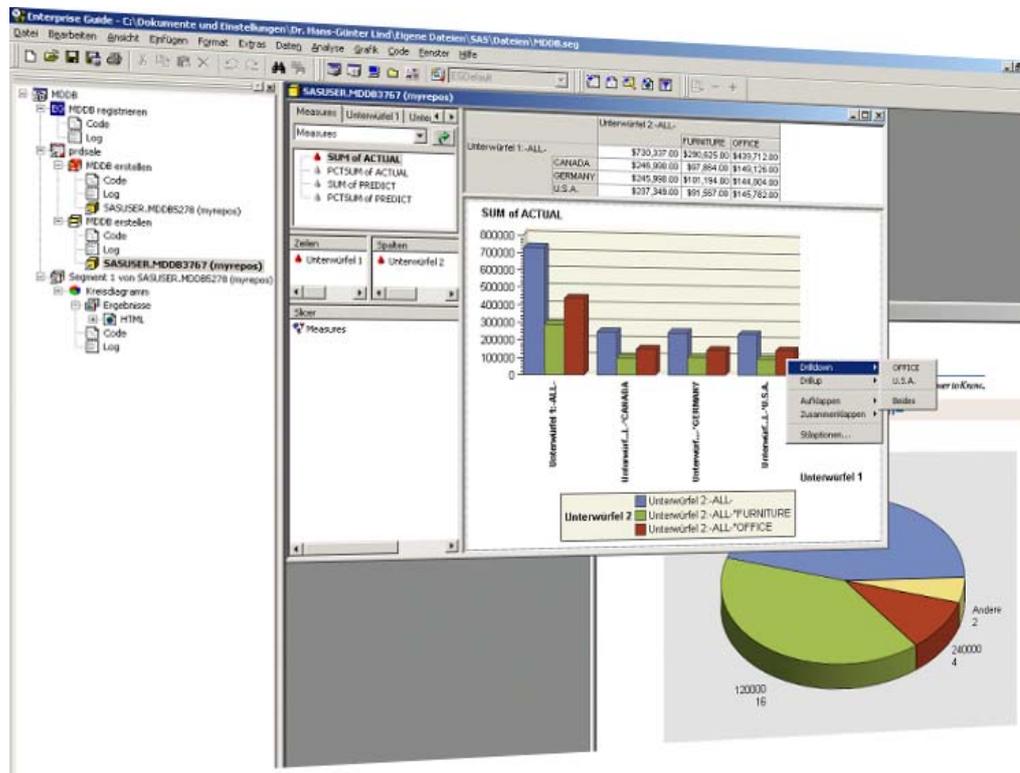


Abbildung 7: Mehrdimensionale Modellierung mit dem Enterprise Guide

Das Szenario „Beratung eines KMU“ wurde weiterhin beibehalten. Die teilweise zusammenhängenden Datentabellen wurden im EG mittels Joins kombiniert, um eine effektive ABC-Kundenanalyse zu ermöglichen. Alle Lernschritte aus Kapitel 2 wurden durchgeführt und als Projekt abgebildet.

Der Lerneffekt ergibt sich wie folgt:

- die konzeptionelle Vorgehensweise zum Aufbau eines Data Warehouse wird besser verankert als in den vorhergehenden Systeme (Ergebnisse reproduzierbar!),
- verifizierbare und reproduzierbare Bewertungen der Datenqualität werden begreifbar,
- die Kodierung lässt den Zusammenhang mit den Quelldatenbanken erkennen, Performanzkennzahlen des EG erlauben den Transfer der Übungen auf die Praxis,
- Aufbau einer mehrdimensionalen Datenstruktur (MDCB) ist leicht möglich,
- erstellen interaktiver Berichte erfolgt „quasi“ automatisch,

- klassische Navigation in Data Warehouses wird geübt,
- intuitive Navigation durch die selbst erstellten Datenstrukturen,
- Export der MDDB in Tabellenformat.

Abschließend wurden die mehrdimensionalen Würfel in Tabellen exportiert und mit MS Excel und InfoZoom analysiert um das ganzheitliche Verständnis zu fördern. Die resultierende Tabellendarstellung der MDDB macht Mehrdimensionalität begreifbar. Insgesamt lässt sich feststellen, dass die Studenten mit dem EG sehr schnell produktiv arbeiten können. Dabei fördert der Projektstrukturbaum in Kombination mit dem explorativen Vorgehen und der Nutzung der statistischen Funktionen die kognitive Modellierung zum Aufbau von und dem Umgang mit Data Warehouses erheblich.

4 Nutzen in der Lehre

Die Verwendung unterschiedlicher Werkzeuge nützt dem ganzheitlichen Verständnis beim Erlernen des Umgangs mit Data Warehouses und liefert einen praxisgerechten Überblick. Insbesondere der SAS Enterprise Guide verbindet strukturiertes inhaltliches Vorgehen, verlässliche Reproduzierbarkeit und Nachvollziehbarkeit sowie Exploration und technische Projektierung.

Der Lernerfolg mit dem SAS Enterprise Guide ist überzeugend:

- Die Nutzung ist einfach und sofort verständlich,
- Konzeptuelles Vorgehen wird gefördert,
- Datenqualität wird begreifbar (Statistikfunktionen „liefen zu Füßen“),
- gängige Probleme und Lösungsansätze können vermittelt werden und
- die Ergebnisse sind jederzeit nachvollziehbar.

Der gesamte Weg von den Quelldaten über das Data Warehouse bis zur Analyse kann mit Hilfe des SAS Enterprise Guide systematisch abgebildet werden. Deshalb wird der SAS Enterprise Guide im Sommersemester 2005 an der FH Frankfurt am Main verstärkt für die Vorlesung Data Warehouse eingesetzt werden.

Literatur

- [1] Bauer, A., Günzel, H. (Hrsg.) (2001): Data Warehouse Systeme: Architektur, Entwicklung, Anwendung, dpunkt.verlag, Heidelberg, 1. Auflage
- [2] Brüggenthies, A., Lindner, H.-G (1999).: InfoZoom – Implizites Wissen bewußt machen und vermitteln, Ein Werkzeug für das Knowledge Reporting. In: Gentsch, P. (Hrsg.): Methodenbasiertes Wissensmanagement. Gabler, Wiesbaden, 1999.
- [3] Inmon, W. H. (2002): Building the Data Warehouse, John Wiley & Sons, New York u. a., 3. Auflage
- [4] Kimball, R.; Ross, M. (2002): The Data Warehouse Toolkit – The Complete Guide to Dimensional Modeling, Wiley, New York, 2. Aufl.
- [5] Lane, P. (2001): Oracle9i Data Warehousing Guide, Release 1 (9.0.1), Redwood City, CA
- [6] Lindner, H.-G. (1999): Knowledge Reporting with InfoZoom - Simply know more: Visual Data Mining and Lean Data Warehousing for decision makers. In: SAP Design Guild 1999
http://www.sapdesignguild.org/index.asp?b=innovation/artikel/info_zoom.html&t=innovation/innovation_topnav.asp
- [7] SAS 2003: Creating and Exploiting MDDBs Using the SAS System, Course Notes.