

Dr. Hans-Georg Kemper
Dipl.-Kfm. Ralf Finger
Dipl.-Kfm. Klaus van Marwyk

Köln, im Februar 1997

Konzeption und Realisierung eines Data Warehouse mit dem SAS System

-

Ein Projektseminar
des Lehrstuhls für Wirtschaftsinformatik,
insbesondere Informationsmanagement der Universität zu Köln

1. Zielsetzung

Im WS 1996/97 wurde im Rahmen eines Projektseminars Studenten des Hauptstudiums der Wirtschaftsinformatik bzw. der Betriebswirtschaftslehre die Möglichkeit geboten, in kleinen Projektgruppen (max. 3 Studenten) mit Hilfe des SAS Systems ein Data Warehouse aufzubauen und es anschließend für verschiedene dispositive Anwendungen heranzuziehen.

Aufgabe der Studentengruppen war es,

- ein betriebswirtschaftliche Fachkonzept auf der Basis einer Fallstudie zu entwickeln,
- ein geeignetes Datenmodell für das Data Warehouse zu entwerfen,
- die Daten aus verschiedenen operativen Quellen in das Data Warehouse zu integrieren,
- ein Executive Information System zur Unterstützung des Top-Managements zu entwickeln,
- ein Berichtswesen für Controllingzwecke zu realisieren sowie
- Data Mining Analysen für Fragestellungen des Marketing zu erstellen.

2. Aufbau des Projektseminars

Das Projektseminars startete mit einer 2-tägigen Blockveranstaltung. In komprimierter Form wurden den Studenten die Grundlagen moderner Data Warehouse-Konzepte vermittelt, wobei speziell dem Bereich der Datenmodellierung mit seinen spezifischen Ausprägungen des Star- und Snow Flake-Schema besondere Beachtung geschenkt wurde. Weiterhin erhielten die Projektteilnehmer an diesen beiden Tagen eine Einführung in die Architektur, die Entwicklungsumgebung und das Datenmanagement des SAS Systems und konnten erste praktische Erfahrungen am Rechner mit SAS sammeln. Am Ende des Einführungsblocks wurde den Studentengruppen das Fallstudienpapier ausgehändigt, das aus einer Dokumentation der Aus-

gangssituation und einer globalen Beschreibung des Zielzustands bestand. Im Detail erhielten die Studentengruppen die textliche Situationsbeschreibung, konkrete Angaben über den Aufbau der unternehmensexternen Rohdaten, das Entity Relationship-Diagramm der normalisierten operativen internen Daten, ein Organigramm des Unternehmens und Rechnungsformulare sowie Artikellisten.

Von diesem Zeitpunkt an wurde die Veranstaltung - wie es sich für Projekte gehört - in Entwicklungsabschnitte unterteilt, die jeweils mit einer studentenseitig zu moderierenden Milestone-Sitzung abgeschlossen wurden. Abbildung 1 zeigt den Ablauf der Veranstaltungsreihe.

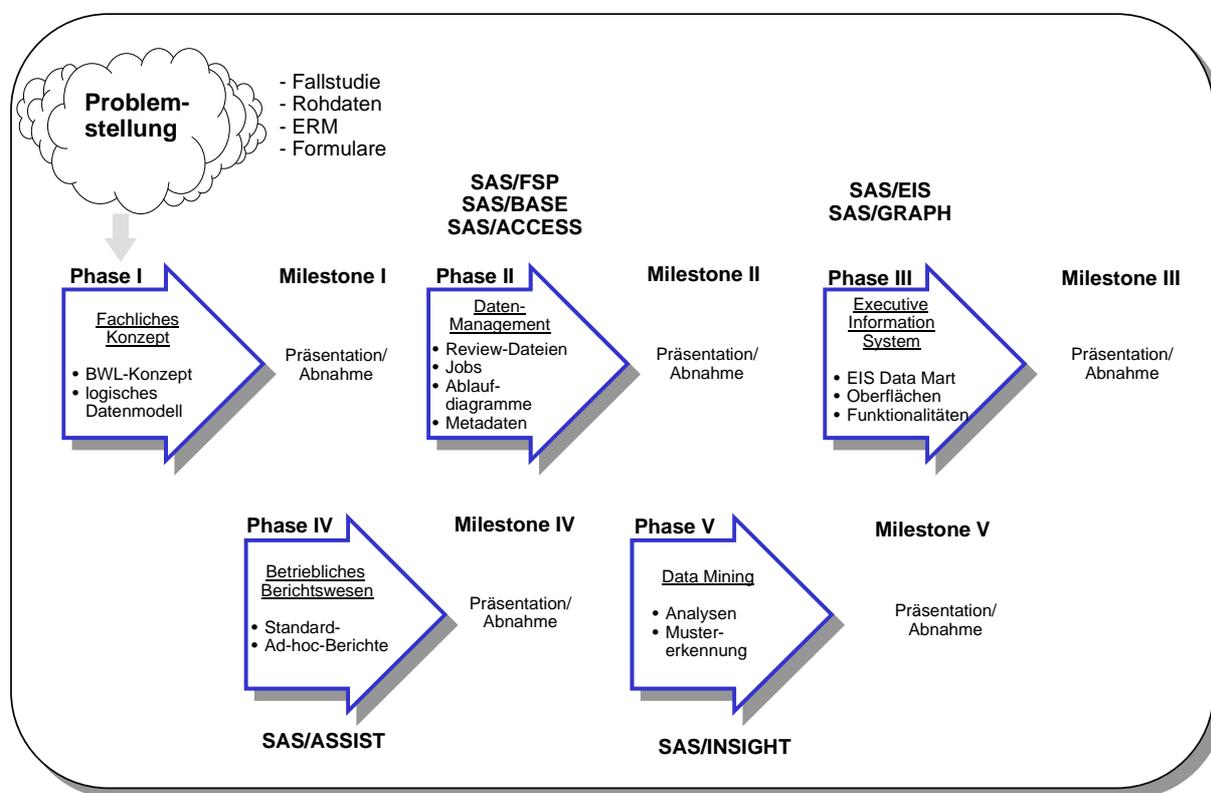


Abb. 1: Aufbau des Projektseminars

Phase 1: Fachliches Konzept

In einer Fallstudie können naturgemäß nicht sämtliche Details antizipiert und festgelegt werden. Auch die vorliegende Beschreibung des Anwendungsfalles enthielt demnach noch große Freiräume. Im Rahmen des zu erarbeitenden Fachkonzeptes sollte daher das „Projekt-Szenario“ detailliert und konkretisiert werden. Zu diesem Zwecke mußten zunächst die im Text der

Fallstudie beschriebene inhaltliche Aufgabenstellung unter betriebswirtschaftlichen Gesichtspunkten analysiert werden und insbesondere die Anspruchsgruppen, die Berichtsdimensionen und die betriebswirtschaftlichen Kennzahlen festgelegt werden. Aufbauend auf diesen Vorüberlegungen waren die logischen Datenstrukturen zu erarbeiten, wobei eine für die Applikationsausrichtung entsprechende Form der Modelldarstellung in Form des Star- oder Snow Flake-Schemas von den Gruppen zu wählen war.

Phase 2: Datenmanagement

In der Phase Datenmanagement stand die Übernahme der relevanten Daten aus den in der Fallstudie angebenen Quellen im Mittelpunkt. Insbesondere sollten mit Hilfe der SAS Module FSP, BASE und ACCESS die hierfür erforderlichen Jobs und Review-Dateien entwickelt und mit Hilfe von Kommentaren und zusätzlich zu erstellenden Ablaufdiagrammen dokumentiert werden. Da in den Übernahmedateien bewußt „Verunreinigungen“ - wie fehlende Werte, falsche Formatierungen - eingebracht worden waren, sollten von den Projektteilnehmern zusätzlich entsprechende Dokumentationen und Korrekturen vorgenommen werden.

Phase 3: Executive Information System

In der nächsten Phase des Projektseminars wurde von den Gruppen das berichtsorientierte Informationssystem für das Top-Management realisiert. Mit Hilfe der Werkzeugkomponente SAS/EIS wurde hierbei von den Studenten ein entsprechender Data Mart mit geeigneten Metadaten erzeugt, so daß führungsrelevante Kennzahlen mittels einer „managergerechten“, intuitiven Benutzeroberflächen recherchiert werden konnten.

Phase 4: Betriebliches Berichtswesen

Für das Controlling unseres Fallstudienunternehmens sollte in dieser Projektphase eine Systemumgebung aufgebaut werden, mit dessen Hilfe sowohl die spontane Informationsabfrage in Form von sog. Ad-hoc-Berichten unterstützt als auch ein Set von vordefinierten Standardberichten erzeugt werden konnte. Die Standardberichte konnten hierbei vom späteren Benutzer über „Jobs“ (gespeicherte SAS/ASSIST-Befehlssequenzen) abgerufen werden. Da Ad-hoc-Berichte nicht detailliert vordefinierbar sind, sollte hier der spätere Benutzer mit Hilfe von SAS/ASSIST selbständig seine spontanen Recherchen durchführen können. Damit

jedoch eine angemessene Performance sichergestellt werden konnte, war es die Aufgabe der Projektteilnehmer, einen abfrageoptimierten Data Mart zu erzeugen, der bereits aufbereitete Daten für die Klasse der antizipierbaren Spontanabfragen enthielt.

Phase 5: Data Mining

Für die Marketing-Abteilung unseres Unternehmens wurde von den Studenten mit Hilfe der Werkzeugkomponente SAS/INSIGHT ein Bericht erstellt, der das Käuferverhalten und die regionalen Nachfrageschwankungen im Zeitverlauf dokumentiert. Ziel der Analyse war es, bislang nicht erkannte Zusammenhänge innerhalb des Datenbestandes „aufzuspüren“ und entsprechend zu visualisieren. Für diese Aufgabe, die häufig als Data Mining bezeichnet wird, mußte von den Gruppen eine angemessene Stichprobe aus den Kundenrechnungen extrahiert und mit Hilfe von statistischen Verfahren analysieren werden, wobei besonderer Wert auf eine angemessene Methodenauswahl und Form der Präsentation gelegt wurde.

3. Resümee

Natürlich war die von den Teilnehmern zu bewältigende Komplexität erheblich. Keiner der Teilnehmer verfügte vor dem Seminar über konzeptionelles Wissen zu Data Warehouse Architekturen. Keiner der Teilnehmer hatte vor dem Seminar mit dem SAS System gearbeitet. Und: Das Seminar hatte den Anspruch, eine möglichst breite Vielfalt der Nutzenpotentiale aufzuzeigen, die ein Data Warehouse auszeichnen können. Aus diesem Grunde mußten mehrere Module des SAS Systems herangezogen werden. Vor dem Hintergrund dieser Komplexität war es entscheidend, möglichst weitgehend die Praxisprobleme der Projektgruppen vor auszudenken und in wichtigen Momenten Lösungshilfen an die Hand zu geben.

Dennoch ist - wie häufig bei Projektarbeit festgestellt werden kann - gerade die Aufwandschätzung für die einzelnen Projektabschnitte eine kritische Größe. „Der Teufel steckt eben im Detail“, und das war auch bei unserem Seminar nicht selten der Fall. So wurde für die Studentengruppen die Projektarbeit zum Teil recht arbeitsintensiv. Da jedoch keiner der Studenten während der Seminarreihe „die weiße Fahne schwenkte“ und aufgab, sondern das Arbeitsklima vielmehr durch eine gelöste, konstruktive und engagierte Mitarbeit aller Beteiligten gezeichnet war, kann die Veranstaltungsreihe wohl trotz des hohen Arbeitsaufwands als gelungen bezeichnet werden.