

Data Warehousing

Konzept einer rechnergestützten Übung mit SAS

Andreas Hilbert
TU Dresden
Münchner Platz 3
01187 Dresden
andreas.hilbert@tu-dresden.de

Alexander E. Müller
TU Dresden
Münchner Platz 3
01187 Dresden

alexander.mueller@tu-dresden.de

Stefan Sommer
Tu Dresden
Münchner Platz 3
01187 Dresden
stefan.sommer@tu-dresden.de

Zusammenfassung

Das Interesse an einer Ausbildung an einem Business Intelligence System eines Marktführer wird gerade in Zeiten einer angespannten wirtschaftlichen Situation für immer mehr Studierende zu einem interessanten Punkt in der proaktiven Gestaltung des eigenen Lebenslaufs.

Der Lehrstuhl für Wirtschaftsinformatik – Business Intelligence Research an der Technischen Universität Dresden trägt dieser Entwicklung unter anderem durch eine stetige Evolution seines Lehrkonzepts Rechnung. Diese Weiterentwicklung des Lehrkonzepts wurde auch nicht zuletzt dadurch notwendig, dass immer mehr Studierende diese zusätzliche Qualifikation mit in ihren stetig voller werdenden Stundenplan aufnehmen möchten. So musste sich das Konzept von einer, die ganze Dauer des Semesters einnehmenden Lehrveranstaltung, hin zu einem komprimierten, nur wenige Wochen dauernden Intensivkurs entwickeln, der dennoch ganzheitlich den Prozess des Data Warehousing abdeckt. Die aktuelle Struktur erlaubt es, durch geschickte Kombination von Frontalunterricht und Selbstlerneinheiten, die Studierenden mittels gerade einmal sechs Terminen an die Business Intelligence Plattform von SAS heranzuführen und das notwendige Handwerkszeug zu vermitteln, die Module Management Console, Enterprise Guide und OLAP Cube Studio soweit bedienen zu können, dass die eigenständige Bearbeitung einer praxisnahen Fallstudie problemlos möglich ist.

Auf der Seite der Lehrenden ist es durch den Aufbau der Lehrveranstaltung möglich, dass ein einzelner Mitarbeiter zwischen 25 und 50 studentische Teilnehmer simultan betreuen kann. Ermöglicht wird dies u.a. durch den Einsatz neuester Erkenntnisse sowie die nahtlose Integration innovativer Lehrmittel wie Pod- und Screencasts.

Schlüsselwörter: Data Warehousing, E-Learning, Wiki, Screencast, ETL, EPG

1 Lehrinhalte an der Professur

Die Professur für Wirtschaftsinformatik – Business Intelligence Research versteht sich als Forschungs-, Lehr- und Transferzentrum der Wirtschaftsinformatik und angrenzender Gebiete der Betriebswirtschaftslehre.

Den zentralen Gegenstand der Forschungsaktivitäten bilden zum einen die Business Intelligence mit ihren Facetten Data Warehousing, Online Analytical Processing, Data, Text und Web Mining sowie der Einsatz von Business Intelligence in unterschiedlichen Anwendungsdomänen, z. B. im e-Marketing, im CRM, im HR-Management oder im Hochschulbereich.

Im Rahmen der Lehre erwerben die Studierenden wissenschaftlich fundierte Kenntnisse und Fähigkeiten zur Analyse, Beschreibung, Gestaltung und Bewertung von Management-Support-Systemen (BI-Systeme).



Abbildung 1: Lehrkonzept der Professur

Damit werden sie in die Lage versetzt, die zum erfolgreichen Führen von Unternehmen benötigten Daten und Informationen zu identifizieren, eine analyseorientierte Speicherung dieser Informationen in Data Warehouses zu konzipieren und zu implementieren sowie unbekannte Zusammenhänge in den Daten mithilfe des Data Mining aufzudecken und somit zur Wertschöpfung dieser Unternehmen beizutragen. In Abbildung 1 ist die Aufteilung dieser Inhalte in die verschiedenen Lehrveranstaltungen für Bachelor bzw. Diplom/Master zu erkennen. Das Data Warehousing ist dem Schwerpunkt Business zugeordnet. Der Praxisteil dieser Lehrveranstaltung entspricht der Rechnerübung, deren Konzept im Folgenden dargestellt wird.

2 Konzept der Rechnerübung

Die Rechnerübung Data Warehousing ermöglicht es pro Jahr etwa 40–50 Studierenden der Diplom- und Masterstudiengänge an der Technischen Universität Dresden die theoretisch erworbenen Kenntnisse an einem der führenden Business Intelligence Systeme umzusetzen. Die Veranstaltung unterliegt aufgrund ihres sehr computernahen Charak-

ters gewissen Rahmenbedingungen und Einschränkungen auf die – neben dem eigentlichen Konzept – in den folgenden Kapiteln eingegangen wird.

2.1 Technische Rahmenbedingungen

Im aktuellen Wintersemester 2009/10 kommt voraussichtlich zum letzten Mal die bereits langjährig eingesetzte SAS 9.1.3 SP4 Plattform zusammen mit dem Enterprise Guide 4.1 zum Einsatz. Als Betriebssystem dient Windows XP Professional SP3. Das Betriebssystem wird der Professur extern vorgegeben. Da die Kapazitäten nicht vorhanden sind, 50 Studierenden simultan Zugriff auf einen eigenen Computerarbeitsplatz zu gewähren, wird auf die öffentlichen Computerräume der Wirtschaftswissenschaftlichen Fakultät zurückgegriffen. Dort kommt ein spezielles Nutzerkonto zum Einsatz, dessen Anmeldedaten nur den Übungsteilnehmer im Gegenzug für das Unterzeichnen einer Verschwiegenheitserklärung mitgeteilt werden. In diesen öffentlichen Computerräumen kommt aufgrund etablierter Installationsroutinen bis auf Weiteres Windows XP zum Einsatz. Diese Installationsroutinen sind es auch, die eine Installation der SAS-Plattform über ein sogenanntes MSI-Paket erfordern. Dies bedeutet, an einem Mustercomputer eine Installation von Hand durchzuführen, welche dann in eine selbstinstallierende Datei umgewandelt wird. Diese Datei wird anschließend zusammen mit den anderen Softwarekomponenten auf allen Rechnern automatisiert installiert. Für eine SAS-Installation die mit lokalen Servern arbeitet bedeutet dies, dass nach dieser zentralisierten Grundinstallation das SAS-System auf jedem Computer dezentral – durch einen Mitarbeiter – konfiguriert werden muss. Diese manuelle Einrichtung nimmt die meiste Zeit bei der Installation in Anspruch, weshalb eine Lösung geschaffen wurde, die benötigte Zeit zu minimieren.

The image shows two overlapping Windows command prompt windows. The top window, titled '_runme_4_new_config.cmd - Editor', contains the following text:

```
@echo off
echo SCHRITT 1 #####
call "O:\SAS_Konfiguration\SAS-Installation-Pool\b_batches\1_Konfiguration von SAS.cmd"
echo SCHRITT 2 DANACH Management Console #####
call "O:\SAS_Konfiguration\SAS-Installation-Pool\b_batches\2_Metadatenserver_starten.cmd"
```

The bottom window, titled '1_Konfiguration von SAS.cmd - Editor', contains the following text:

```
@echo off
echo SAS wird konfiguriert...
O:\SAS_Konfiguration\cfgwizard\windows\setup.exe -DOVERRIDE_PROPERTIES_FILE="C:\configuration.properties" -i silent
pause
```

Abbildung 2: Batchdateien für die rechnerindividuelle Konfiguration

Wie in Abbildung 2 ersichtlich bilden mehrere ineinandergreifende Batchdateien, die auf eine rechnerindividuelle configuration.properties-Datei zugreifen das Herzstück der Lösung.

Um so wenig wie möglich Interaktion mit dem Installierenden zu verursachen, wird die –silent Option des Configuration Wizard verwendet. Die aktive Beteiligung des mit der Installation betrauten Mitarbeiters reduziert sich so auf wenige Klicks, um die einzelnen

Schritte zu bestätigen. Der faktische Aufwand pro Rechner lies sich so auf weniger als eine Minute reduzieren.

2.2 Unterrichtseinheiten

Der Aufbau der Rechnerübung unterlag in den letzten Semestern einer erheblichen Evolution. Bedingt wurde dies maßgeblich durch die Einführung der Bachelor- und Masterstudiengänge. Eine Entzerrung der Lehr- und Prüfungsbelastung wurde gefordert und führte dazu, dass die initial über ein ganzes Semester abgehaltene Rechnerübung in eine Blockveranstaltung umgewandelt wurde, die binnen sechs Wochen mit jeweils einem Übungs- und einem Tutoriumstermin pro Woche, die Inhalte vermittelt.

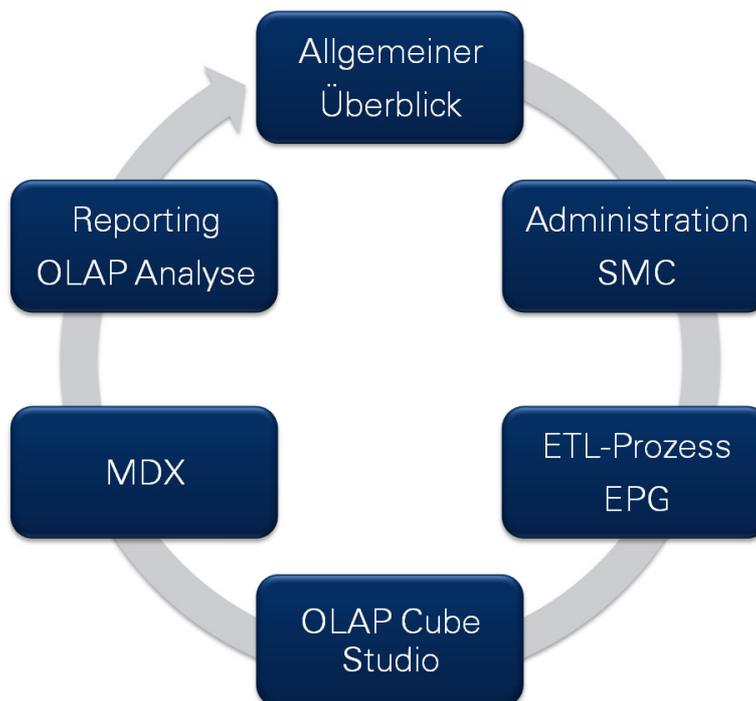


Abbildung 3: Inhalte der Unterrichtseinheiten

Aus diesen beiden 90-minütigen Terminen ergibt sich pro Woche ein betreutes Angebot von 180 Minuten, wobei der Besuch des Tutoriums optional ist. Wie Abbildung 3 zeigt, existieren momentan fünf thematisch fokussierte Veranstaltungen sowie zu Beginn ein Termin, der einen allgemeinen Überblick über die gesamte eingesetzte SAS-Plattform vermittelt.

2.2.1 Allgemeiner Überblick

Im allgemeinen Überblick wird den Studierenden vermittelt, dass die SAS Management Console (SMC) eine javabasierte grafische Verwaltungsoberfläche zur Administration des gesamten SAS Systems ist und einen Single Point of Control entlang der Intelligence Value Chain bietet. Ebenso wird erläutert, dass die SMC sowohl lokale als auch entfernte, über das Netzwerk erreichbare, SAS-Server administrieren kann.

Im Anschluss werden der Enterprise Guide (EPG) und sein grafischer Aufbau näher betrachtet. Es wird ein Einblick in die Nutzung des Projektdesigners, der Logfileanzeige und der Serverliste gegeben. Gefolgt wird diese EPG-Kurzeinführung von der Definition relevanter und häufig vorkommender Begriffe wie SAS Data Library, Library Engine oder LIBREF. Der allgemeine Überblick schließt mit einer Betrachtung des OLAP Cube Studios, insbesondere mit der Vorstellung des Cube Designers um den Studierenden vor Augen zu führen, warum sie in den späteren Schritten gewisse Vorgänge durchführen müssen.

2.2.2 SAS Management Console

Ab der Veranstaltung zur Management Console sind die Foliensätze für die fünf thematisch fokussierten Rechnerübungen ähnlich aufgebaut. Es ist, wie in Abbildung 4 und 5 erkennbar, immer eine detailliert beschriftete Sammlung von Screenshots der

Einrichten eines Metadatenprofils Ablauf



TU Dresden Professur für Wirtschaftsinformatik – Business Intelligence Research

- 1. Eingabemaske
 - Einfach den **Button „Weiter“** drücken.
- 2. Eingabemaske
 - Den **Namen sasadm** für das Metadatenprofil vergeben und mit dem **Button „Weiter“** bestätigen.

- 3. Eingabemaske
 - Eingabe des Rechnernamens, der sich vorne oben auf dem Gehäuse des Rechners befindet, z.B. **P14K**, und mit dem Zusatz **.iln.wiwi.tu-dresden.de** versehen.
→ Bsp: **P14K.iln.wiwi.tu-dresden.de**

Data Warehousing WS 2009/10 – Seite 6

Abbildung 4: Folienaufbau 1

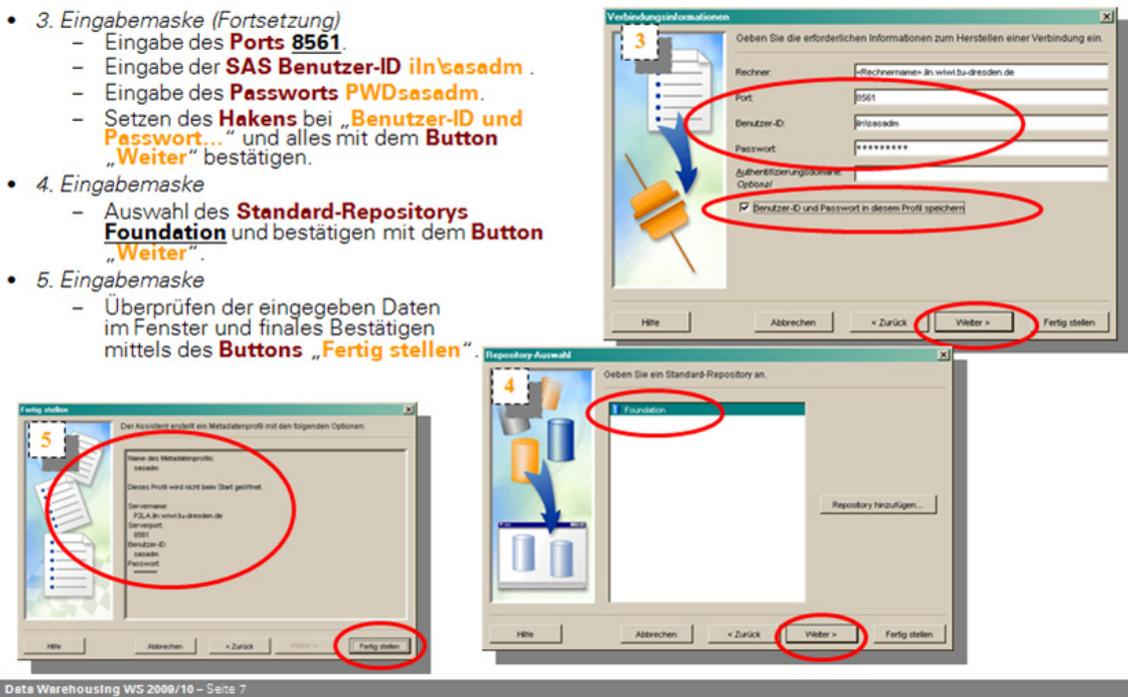
einzelnen Schritte, die der Nutzer zum Erreichen eines vorab gestellten Ziels durchführen muss. Die vorzunehmenden Einstellungen sind jeweils einzeln auf einem Bild erfasst und relevante Stellen sind durch ein rotes Oval oder Rechteck hervorgehoben. Die durch die Screenshots dokumentierten Lösungswege erstrecken sich, wie ebenfalls in Abbildung 4 und 5 ersichtlich, oftmals über mehrere Seiten des Foliensatzes und jede Folie kann wiederum mehrere Einzelschritte mit den zugehörigen Erläuterungen enthalten.

Einrichten eines Metadatenprofils Ablauf



TU Dresden Professur für Wirtschaftsinformatik – Business Intelligence Research

- 3. Eingabemaske (Fortsetzung)
 - Eingabe des **Ports 8561**.
 - Eingabe der **SAS Benutzer-ID iiln\sasadm**.
 - Eingabe des **Passworts PWDsasadm**.
 - Setzen des **Hakens** bei „**Benutzer-ID und Passwort...**“ und alles mit dem **Button „Weiter“** bestätigen.
- 4. Eingabemaske
 - Auswahl des **Standard-Repositorys Foundation** und bestätigen mit dem **Button „Weiter“**.
- 5. Eingabemaske
 - Überprüfen der eingegebenen Daten im Fenster und finales Bestätigen mittels des **Buttons „Fertig stellen“**.



Data Warehousing WS 2009/10 – Seite 7

Abbildung 5: Folienaufbau 2

Inhaltlich wird den Studierenden in diesem Termin vermittelt, warum ein Metadatenprofil nötig ist und was die Einstellungen beim Anlegen eines solchen bewirken. Anschließend wird das lokale Anlegen von SAS- und Datenbankbibliotheken erläutert. Die Teilnehmer erfahren, wie sie sich ihre Datenquellen in Form von Bibliotheken verfügbar machen (meist über die ODBC-Schnittstelle) und wie Datenbanken sowie Staging-Areas definiert werden. Abschließend erfolgt eine Einführung in das Nutzermanagement und die Zugriffssteuerung mittels ACEs und ACTs.

2.2.3 ETL-Prozess im Enterprise Guide

Der Termin zum ETL-Prozess im Enterprise Guide (EPG) widmet sich nach einem kurzen Exkurs zum SAS-Code den einzelnen Schritten Extraktion, Transformation und Laden. Durch die eingeführten SAS Code Fragmente wie beispielsweise:

```
LIBNAME BIB_VFDB CLEAR;
LIBNAME BIB_VFDB META
METAOUT=DATA
LIBRARY="VFDB-Bibliothek"
METADATENREPOSITORY="Foundation";
```

werden Bibliotheken zugewiesen und aktualisiert. Die Beschäftigung mit den eigentlichen Schritten des ETL-Prozesses umfasst dann das verfügbar machen der in der SMC definierten Bibliotheken im Projekt Designer, das Gestalten von (komplexen) Filtern und Abfragen, um die Daten in das gewünschte Format zu bringen und letztlich das La-

den der fertigen Daten in eine Data Warehouse Struktur. Bei den Filtern und Abfragen wird ein besonderer Schwerpunkt sowohl auf die verschiedenen Möglichkeiten die Tabellen zu joinen gelegt als auch darauf, mit berechneten Spalten zu arbeiten.

2.2.4 OLAP Cube Studio

Der vierte Termin widmet sich ganz dem OLAP Cube Studio. Nach einer genaueren Einführung in den Aufbau der grafischen Oberfläche wird der Fokus auf den Assistenten zum Erstellen eines OLAP-Würfels, den Cube Designer, gelegt. Im Cube Designer kommt als Basis für die Würfelerstellung das STAR-Schema zum Einsatz. Die Daten wurden im ETL-Prozess bereits in diesem, bis dahin nur aus der Theorie bekannten, Schema gespeichert. Die Möglichkeit eine Detailtabelle als Input zu verwenden wird im Rahmen dieser Rechnerübung nicht weiter thematisiert, aber interessierten Studierenden als weiterführende Aufgabe für ein Selbststudium empfohlen. In den weiteren Schritten zur Definition von Dimensionen, Ebenen und Hierarchien erkennen die Studierenden, wofür viele der bei der (theoretischen) Konzeption und praktischen Implementierung des STAR-Schemas im ETL-Prozess getätigten Maßnahmen wichtig waren. Ein kurzer Exkurs in das Lesen des im Hintergrund erstellten PROC OLAP-Codes schließt diese Lehreinheit.

2.2.5 Multidimensional Expressions

Da das Erlernen von Multidimensional Expressions (MDX) nicht in einer Unterrichtseinheit von 90 Minuten funktionieren kann, wird in dieser Stunde der Schwerpunkt bewusst mehr ganzheitlich gesehen. Es geht darum, den Studierenden die Idee hinter der Sprache nahe zu bringen und bereits fertige Statements interpretieren zu können. Neben dem aktiven Verstehen von Tupeln wie

```
(Source.[Eastern Hemisphere].Africa, Time.[2nd half].[4th
quarter],Route.Air, Measures.Packages)
```

oder auch anspruchsvolleren SELECT-Statements wie beispielsweise

```
SELECT
{[Measures].[Sales Amount], [Measures].[Tax Amount] } ON COLUMNS,
{[Date].[Fiscal].[Fiscal Year].&[2002], [Date].[Fiscal].[Fiscal
Year].&[2003] }
```

```
ON ROWS FROM [Adventure Works] WHERE
([Sales Territory].[Southwest])
```

ist es wichtig, dass die Studierenden am Ende der Veranstaltung einfache MDX Ausdrücke, abhängig von den Gegebenheiten (Dimensionen, Ebenen, Hierarchien) der vorliegenden Daten erstellen können. Als Beispiel für ein solches, selbst zu erstellendes Statement sei das folgende genannt:

```
(( Zeit.CurrentMember, [Measures].[UmsatzSUM]) / (
Zeit.CurrentMember.PrevMember, [Measures].[UmsatzSUM])) - 1
```

Dieser Ausdruck wird in einem vorliegenden Datenwürfel das Umsatzwachstum in Prozent zurückliefern. Eingegeben wird dieses Statement im EGP, was zum Ende dieser Veranstaltung den Übergang zum sechsten und letzten Termin herstellt.

2.2.6 Reporting und OLAP Analyse

Die letzte Lehrveranstaltung gliedert sich inhaltlich in zwei verwandte Teilbereich, das Erstellen von Reports und die OLAP-Analyse eines Datenwürfels. Als Einstieg in die Datenauswertung, erhalten die Teilnehmer eine Einführung in die verschiedenen Möglichkeiten, mit dem Enterprise Guide Reports auf Basis vorhandener Daten zu erstellen. Die einzelnen Reports werden abschließend in einem ansprechend gestalteten Bericht zusammengefasst.

Den Abschluss der Rechnerübung bildet das Durchführen von Analysen auf dem bis zu diesem Punkt Schritt für Schritt selbst erstellten OLAP Würfel. Die Studierenden entdecken die verschiedenen Navigationsarten, die der Enterprise Guide als grafische Schnittstelle zu den Würfeldaten bietet und erhalten Anleitungen, wie die Ergebnisse schneller wieder abrufbar (vgl. Abbildung 6) oder grafisch besser aufbereitet werden können.

OLAP-Pivot-Tabelle: Navigationswege speichern



TU Dresden Professur für Wirtschaftsinformatik – Business Intelligence Research

- Durch Drill-up  und Drill-down  kann man sich durch den Würfel bewegen
- Bestimmte Navigationspfade werden regelmäßig aufgerufen -> Speicherung hilfreich!

Beispiel: Der Produktmanager der Warengruppe „Bikes“ möchte jede Woche die Auftragswerte für alle Artikel der Artikelgruppe „Mountainbikes“ überblicken.

Navigationspfade speichern und abrufen

Data Warehousing WS 2009/10 – Seite 29

Abbildung 6: OLAP Navigationswege speichern

2.3 Screencast und Wiki Integration

Um die Studenten zwischen den einzelnen Terminen zu unterstützen, wurde ein Wiki ins Leben gerufen, in dem die theoretischen Aspekte der Unterrichtseinheiten behandelt werden. Das Wiki wurde „von Studenten für Studenten“ in kleinen Projekten umgesetzt und wird in jedem Semester aktiv überarbeitet sowie erweitert. Es basiert auf der Software der Wikimedia, die auch die Grundlage für das im Jahr 2009 eingeführte SAS Wiki und auch die Wikipedia ist. Dieses von der Professur aktiv unterstützte Wiki umfasst mittlerweile mehrere hundert Beiträge zu den verschiedensten Aspekten der Business Intelligence und des Data Warehousing.

Speziell für kritische Einstellungen und Vorgänge, bei denen die Erfahrung der letzten Semester gezeigt hat, dass gehäuft Probleme auftreten, wurde – und wird weiterhin – eine Bibliothek mit Screencasts aufgebaut. Ein solcher Screencast ist eine kurze – meist kürzer als fünf Minuten – Aufzeichnung von Arbeitsschritten, die in der SMC, dem EPG oder dem Cube Studio durchgeführt werden. Diese Screencasts sind momentan bewusst nicht vertont, da die Studenten in den Computerräumen der Fakultät nur beschränkt Audio zur Verfügung haben.

2.4 Projektarbeit

Im Nachgang der Veranstaltung bildet eine inhaltlich an die Rechnerübung angelehnte Projektarbeit den Abschluss. Die Studenten müssen in Kleinstgruppen (2-3 Personen) einen (fiktiven) Fall bearbeiten. Für die Konzeption, die Implementierung in SAS und das Aufbereiten der Ergebnisse in Form einer 20-minütigen Präsentation stehen vier Wochen zur Verfügung. Bisherige Themen umfassten z. B. Zeitreihenanalysen anhand der Daten von finance.yahoo.com, Konzeption eines OLAP-Würfels für eine Restaurantkette, Identifikation der Unterschiede zwischen SAS 9.1.3 und SAS 9.2, bezogen auf die spezifischen Inhalte der Rechnerübung.

3 Erfolgsgeschichten

Die aktuelle Konzeption der Rechnerübung mit sechs Blockterminen und dem jeweils thematisch passenden Tutorium schneidet in der fakultätsinternen Lehrevaluation immer deutlich besser ab als der Fakultätsdurchschnitt. Dies ist ein klares Zeichen dafür, dass die Studenten dieses Lehrkonzept gut finden, wenn auch immer angemerkt wird, dass mehr eigenständiges Engagement vonseiten der Studierenden nötig ist als in anderen Rechnerübungen.

Studenten, die an der Professur den Schwerpunkt Business Intelligence/Data Warehouse erfolgreich gemeistert haben, berichten immer wieder, wie nützlich die erworbenen Kenntnisse bei der Arbeitssuche oder dem späteren Job sind. So ist es nicht verwunderlich, dass viele Alumni bei diversen wichtigen Firmen aus dem BI-Umfeld untergekommen sind. Exemplarisch können hier Altran, AT Kearney, Daimler, DETECON, PWC, SHS Viveon oder auch T-Systems genannt werden.

