

Die Data Warehousing Kolumne

Diese Kolumne stellt regelmäßig Aspekte aus dem Bereich Data Warehousing zur Diskussion. Beiträge bitte direkt an die Betreuer dieser Kolumne: Manfred Jeusfeld (jeusfeld@kub.nl) oder Heinrich Jasper (jasper@tu-freiberg.de).

In den folgenden Beiträgen stellen sich Arbeitsgruppen vor, die das Thema Data Warehousing behandeln: Zuerst die Forschungsgruppe Wissensbasen (WIBAS) des FORWISS in München, anschließend die GI-Arbeitskreise "Konzepte des Data Warehousing" (AK der Fachgruppe 2.5.1) und "Modellierung und Nutzung von Data Warehouse-Systemen" (AK der Fachgruppe 5.10).

Forschungs- und Entwicklungsaktivitäten der Forschungsgruppe Wissensbasen (WIBAS) von FORWISS im Bereich Data Warehousing

Die WIBAS Forschungsgruppe von FORWISS (<http://www.wibas.forwiss.tu-muenchen.de/>) führt unter der Leitung von Prof. Rudolf Bayer, Ph.D. und Dr. Volker Markl Grundlagenforschung, Entwicklung und Technologietransfer in den Bereichen Data Warehousing, Online Datenbanken, digitale Bibliotheken, Metadatenmanagement und Zugriffsmethoden durch. Aktivitäten im Bereich Data Warehousing von WIBAS finden im Rahmen der Projekte MISTRAL, EDITH, System 42 und ESTEDI statt. Insgesamt sind 7 FORWISS Wissenschaftler im Warehousing Bereich tätig.

Im Projekt MISTRAL (<http://mistral.in.tum.de>) werden der UB-Baum, eine mehrdimensionale Zugriffsmethode, und dessen Anwendungen, insbesondere für Data Warehousing, untersucht. Dieses Grundlagenprojekt entwickelt und verfeinert Algorithmen für mehrdimensionale Datenbankzugriffe. Daneben wird eine API mit UB-Baum Algorithmen oberhalb verschiedener kommerzieller Datenbanksysteme entwickelt. Weiterführende Untersuchungen befassen sich mit Anfrageoptimierung und der Verwaltung ausgedehnter Objekte für geographische Data Warehousing Anwendungen.

Im EDITH Projekt (<http://edith.in.tum.de>) wird das relationale Datenbanksystem TransBase um spezielle Konstrukte erweitert, die eine hochperformante Abarbeitung von Anfragen auf typischen Warehouse-Schemata erlaubt. EDITH wird von der Europäischen Union im IST Programm gefördert und im internationalen Rahmen mit Partnern aus Italien, Griechenland und Polen durchgeführt. Pilot-Anwendungen im Rahmen des EDITH Projekts sind das Retail-Warehouse von Kotsovolos sowie ein GIS Datawarehouse mit Daten der Stadt Lodz. Das im Vorgängerprojekt MDA mit TransAction Software entwickelte System TransBase/Hypercube, das UB-Bäume als Zugriffsmethode in TransBase völlig transparent für den Benutzer bereitstellt, wurde mit dem IST Prize 2001 von der Europäischen Kommission und von Eurocase ausgezeichnet. Bei MDA wurde der UB-Baum in TransBase/HyperCube für das Konsumgüter-Warehouse der Gesellschaft für Konsumforschung (GfK) eingesetzt und erzielte erhebliche Antwortzeitverkürzungen für typische GfK Reports.

Im Rahmen des Projekts System 42 (www.forwiss.de/~system42) beschäftigt sich die FORWISS Forschungsgruppe Wissensbasen mit methodischer Unterstützung beim Betrieb und Aufbau von Data-Warehouse-Systemen. Die wichtigsten Themenschwerpunkte des Projekts sind:

- Eine Sichten-basierte Methodik zur graphischen Modellierung von Data Warehouses und deren Tool-Unterstützung. (BabelFish).
- Die Untersuchung von Schema-Evolution in Data-Warehouse-Umgebungen (FIESTA). Dies umfaßt die graphische Spezifikation von Schemaänderungen auf konzeptueller Ebene und die automatische Adaption der Implementierung (inkl. der Adaption der vorhanden Instanzen).
- Die Modellierung und Vorhersage von Benutzerverhalten in OLAP Umgebungen (PROMISE). Dabei werden Regelmäßigkeiten im Anfrageverhalten vom OLAP-Benutzern automatisch vom OLAP System gelernt und zur Antizipation künftiger Anfragen genutzt. Dies ermöglicht die Verbesserung von semantischen Cache-Verfahren z. B. durch spekulative Materialisierungstechniken.

Die Übertragung von Data-Management-Techniken aus dem Data-Warehouse-Bereich auf Supercomputing-Anwendungen ist Thema des EU-Projekts ESTEDI (www.estedi.org). Diese Anwendungen (z. B. kosmologische, chemische und strömungsdynamische Simulationen) erzeugen große Mengen an Daten (mehrere GB pro Tag), die zu verwaltet und zu analysieren sind. Bisher wird dazu allerdings kaum Datenbanktechnologie verwendet. Die Datenbestände weisen interessant Ähnlichkeiten mit "klassischen" Data-Warehouse-Anwendungsgebieten auf, z. B. das analyseorientierte Zugriffsverhalten und inhärent multidimensionale Eigenschaften (meist Raum- und Zeitbezug). Im Rahmen des ESTEDI Projekts wird bei FORWISS ein multidimensionales Datenbanksystem für dieses Anwendungsgebiet erweitert. Schwerpunkte bilden hierbei der Online Zugriff auf Tertiärspeichermedien unter Berücksichtigung der multidimensionalen Datencharakteristik und die Parallelisierung und Optimierung der Ausführung von multidimensionalen Operationen auf hierarchisch partitionierten multidimensionalen Datenräumen.

Kontakt:

Prof. Rudolf Bayer, Ph. D.

bayer@informatik.tu-muenchen.de

Dr. Volker Markl

volker.markl@forwiss.de

Arbeitskreis 5.10.4

"Modellierung und Nutzung von Data Warehouse-Systemen"

Prof. E.J. Sinz, M. Böhnlein, A. Ulbrich-vom Ende

Lehrstuhl für Wirtschaftsinformatik, Universität Bamberg

Feldkirchenstr. 21, D-96045 Bamberg

E-Mail: {elmar.sinz | michael.boehnlein | achim.ulbrich } @sowi.uni-bamberg.de

Die Gründung des Arbeitskreises 5.10.4. wurde am 15.10.1998 im Rahmen der Vollversammlung der Fachgruppe 5.10 "Informationssystemarchitekturen: Modellierung betrieblicher Informationssysteme (MobIS)" beschlossen. Die erste offizielle Veranstaltung des Arbeitskreises fand im Rahmen der Tagung MobIS im Oktober 1999 statt.

Ziel des Arbeitskreises ist es, Probleme im Bereich der Modellierung, des Betriebs und der Nutzung von Data Warehouse-Systemen gleichermaßen aus praktischer und wissenschaftlicher Sicht zu diskutieren und zu bewerten. Dabei ist der inhaltliche Schwerpunkt des Arbeitskreises durchgängig im Bereich der Wirtschaftsinformatik angesiedelt. Zu dem Themenschwerpunkt *Modellierung von Data Warehouse-Systemen* zählt neben der konzeptuellen und logischen Modellierung die Entwicklung von Vorgehensmodellen zur Erstellung von Data Warehouses, sowie Aspekte der Verwaltung von Metadaten. Darüber hinaus befasst sich der Arbeitskreis insbesondere mit Techniken zur Ermittlung des Informationsbedarfs von Entscheidungsträgern und Führungskräften. Im Rahmen *des Betriebs von Data Warehouse-Systemen* hingegen werden beispielsweise mögliche Sicherheitsanforderungen und Qualitätsaspekte von Data Warehouse-Systemen diskutiert. Dabei wird durch die Berücksichtigung kommerzieller Data Warehouse-Lösungen ein enger Bezug zur industriellen Praxis garantiert. Für die *Nutzung von Data Warehouse-Systemen* spielen vor allem die Einsatzgebiete in der Praxis und deren betriebswirtschaftliche Nutzenpotentiale eine entscheidende Rolle. Aus diesem Grund beschäftigt sich der Arbeitskreis sowohl mit Synergieeffekten zwischen E-Commerce-/E-Business- und Data Warehouse-Systemen als auch mit der umfassenden Integration von Data Mining-Umgebungen in Data Warehouse-Lösungen. Unter diesem Teilaspekt wird auch Mitarbeitern industrieller Projekte Raum für die Diskussion von praxisrelevanter Problemstellungen und Lösungen eingeräumt.

Trotz seines kurzen Bestehens erfreut sich der Arbeitskreis eines starken Zuwachses und zählt bereits 112 eingetragene Mitglieder. Hierbei ist besonderes die hohe Beteiligung von Vertretern aus der Praxis zu unterstreichen. Nach der ersten Veranstaltung auf der MobIS 1999 fanden bisher zwei weitere themenspezifische Workshops statt. Der erste Workshop (Freiburg, März 2000) wurde in Zusammenarbeit mit dem Kompetenzzentrum Data Warehousing Strategie der Universität St. Gallen (Prof. Dr. Winter, Dr. Jung) mit dem Themenschwerpunkt "Fachkonzeptentwurf und Metadaten beim Data Warehousing" organisiert. Ein weiterer Workshop des Arbeitskreises auf der MobIS 2000 richtete den Fokus auf "Möglichkeiten und Grenzen gängiger ETL-Werkzeuge".

Im April nächsten Jahres ist eine Veranstaltung in Zusammenarbeit mit dem Arbeitskreis "Konzepte des Data Warehousing" unter Koordination des Lehrstuhls von Prof. Chamoni (Universität Duisburg) geplant. Themenschwerpunkt dieses Workshops ist das "Knowledge Discovery in Databases". Für 2001 sind darüber hinaus eine themenspezifische Veranstaltung im Juni/Juli und ein Workshop auf der MobIS 2001 in Bamberg fest vorgesehen.

Weitere Informationen über den Arbeitskreis finden sich auf der Internetseite <http://ceus.uni-bamberg.de/ak5104>. Mitglieder steht neben Data Warehouse-spezifischen Inhalten auch eine Mailingliste und ein Diskussionsforum zur Verfügung.

Wir möchten uns hiermit noch einmal ganz herzlich bei unseren Mitglieder für Ihre rege Beteiligung und konstruktive Arbeit in den letzten beiden Jahren bedanken.
