

Übung 4

Prof. Dr. Alexandros Nanopoulos, Christoph Freudenthaler
Wirtschaftsinformatik und Maschinelles Lernen (ISMLL)
Universität Hildesheim

Abgabe bis: 06.06.2009

Aufgabe 1: Dimensional Modelling I (20 Punkte)

Ein renommiertes Buchhandlungsunternehmen aus dem Großraum Hannover möchte zur Unterstützung seines neu errichteten Online-Vertriebes ein DWH erstellen, um die Entwicklung dieses viel versprechenden, neuen Absatzweges zu beobachten, zu analysieren und um gegebenenfalls rechtzeitig gegensteuern zu können. Dabei wurdest Du als neuer Mitarbeiter des beratenden IT-Unternehmens *UniConsult* nun beauftragt, ein Datenmodell für ihren Kern-Geschäftsprozess *Buchverkauf* zu erstellen.

Dieser Geschäftsprozess soll auf der Granularitätsebene des verkauften Buches aufgezeichnet werden und gleichzeitig die wichtigsten zur Analyse notwendigen Informationen beinhalten, die durch die zur Verfügung stehenden (transaktionsbasierten) Informationssysteme bereit gestellt werden können. Der CIO ist sich jedoch noch nicht sicher, welche Informationen er im DWH über den Geschäftsprozess aufzeichnen möchte. Ihm ist nach einem Gespräch mit den Fachabteilungen aber klar, dass die Abbildung des Geschäftsprozesses folgende Punkte auf jeden Fall zu enthalten hat:

- das verkaufte Buch und dessen charakteristischen Attribute (Titel, ISBN, Einkaufspreis, Verkaufspreis, Verkaufsdatum,...)
- den Verlag des Buches
- den Kunden, dessen Daten durch die online-Plattform aufgezeichnet werden (demografische Daten, Emailadresse, Kreditkartennummer,...)
- der oder die Autoren des Buches

Die Einspeisung der Daten in das DWH soll täglich erfolgen. Im Sinne einer späteren Anpassung an neue Anforderungen soll das Datenmodell möglichst flexibel sein. Die Bücher werden in verschiedene Kategorien wie Wissenschaft, Krimi, Kinderbücher, Kochbücher,... eindeutig eingeteilt. Den CIO interessieren besonders weiters Informationen zu den Zu- und Auslieferern der Bücher, da das Unternehmen momentan eine Vielzahl unterschiedlicher Lieferanten beauftragt und in Zukunft eine Evaluation dieser Geschäftspartner möglich sein soll.

Erstelle durch *dimensionales Modellieren* ein Abbild des Geschäftsprozesses *Buchverkauf* im Star-Schema, das den genannten Anforderungen entspricht. Welche Dimensionen und Attribute hast Du gewählt? Warum?

Aufgabe 2: Dimensional Modelling II (5 Punkte)

Nachdem Du das Star-Schema für den Buchverkauf entwickelt hast, fällt Dir auf, dass durch die tägliche Beladung des DWH oftmals nicht alle Daten vorhanden sein können. So zum Beispiel kann im Zeitpunkt des Verkaufes noch nicht klar sein, welches Unternehmen das Buch an den Kunden liefert. Wie gehst Du bei der Modellierung damit um?

Aufgabe 3: Dimensional Modelling III (5 Punkte)

Die Aufzeichnungen der Buchverkäufe findet sekundengenau statt, die Speicherung der Verkäufe bislang nur auf Tagesebene. Es soll nun das Datenmodell so erweitert werden, dass eine sekundengenaue Analyse der Verkäufe möglich wird. Ein erster Ansatz dazu wäre, dass man die Zeitdimension von Tages- auf Sekundengranularität verfeinert. Angenommen, das DWH läuft nun 10 Jahre. Wieviele Datensätze (Zeilen) würde die Zeitdimension in etwa umfassen? Wie kann man das effizienter (platzsparender) lösen? Wieviele Datensätze würde Dein Vorschlag in den 10 Jahren akkumulieren?