

# Übung 3

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

Abgabe bis: 17.05.2009

## Aufgabe 1: (15 Punkte)

Kathryn Avery war in den letzten 6 Jahren Datenbankadministratorin für eine amerikanische Einzelhandelskette (Big Chain) . Vor kurzem wurde sie mit der Entwicklung des ersten Data Warehouses für Big Chain beauftragt. Das Projekt genießt die Unterstützung des Topmanagements und des CIO. Der Grund für die Entwicklung des Data Warehouses ist die Verbesserung der Berichtssysteme - vor allem des Verkaufs und des Marketings. Langfristig will man mit dem DWH auch das Customer Relationship Management verbessern. Deshalb war Kathryn auf einer Konferenz zu Data Warehousing und hat sich mit allerlei Fachbüchern eingedeckt. Trotzdem ist sie noch unsicher, wie sie die Entwicklung des Data Warehouses angehen soll. Soll es nach dem Ansatz von Inmon (Top-Down-Ansatz; abhängige Data Marts) gehen oder nach dem Ansatz von Kimball (Bottom-Up, unabhängige Data Marts).

Ursprünglich war Kathryn der Meinung, dass sie nun einen guten Überblick über beide Ansätze hat. Je länger sie aber darüber nachdenkt, welche der Konzepte sie übernehmen soll, fallen ihr nach und nach mehr Fragen ein. Bitte hilf ihr dabei.

1. Was sind die genauen Unterschiede zwischen dem Top-Down- und dem Bottom-Up-Ansatz?
2. Wie sehen die Faktoren aus, die bei der Auswahl zwischen den beiden Ansätzen helfen?
3. Was sollte sie nach einer baldigen Entscheidung für eine der beiden Ansätze tun?

## Aufgabe 2: (15 Punkte)

In der Vorlesung wurde besprochen, dass Enterprise Application Integration (EAI)-Lösungen mit einer hohen Wahrscheinlichkeit scheitern. Welche Gründe gibt es dafür? (Hinweise unter [http://www.ebizq.net/topics/int\\_sbp/features/3463.html](http://www.ebizq.net/topics/int_sbp/features/3463.html)) Wie ist Deine Meinung?

## Aufgabe 3: (5 Punkte)

Beantworte die folgenden Aussagen mit *wahr* oder *falsch*:

1. Der Top-Down-Ansatz von Inman für die Erstellung von DWH ist auch als *Data Mart*-Ansatz bekannt-
2. DBMS tendieren dazu, nicht redundant, normalisiert und relational zu sein, während DWH redundant, nicht normalisiert und nicht relational sind.
3. Datenbereinigung (data cleansing) ist ein kritischer und schwieriger Aspekt bei der DWH-Entwicklung und beinhaltet das Abstimmen sich unternehmensweit widersprechender Datendefinitionen (data definitions) und -formate.

**Beantworte die folgenden Multiple-Choice Aussagen:**

4. Welche der 4 angegebenen Möglichkeiten beschreibt eine Menge von (Software-)Tools zur Echtzeitdatenintegration verschiedener, unternehmensweiter Datenquellen wie relationalen Datenbanken, Webservices oder multidimensionalen Datenbanken (zB SAP)?
  - (a) Information integration
  - (b) Data management integration
  - (c) SQL data integration
  - (d) Enterprise Information Integration
5. Welcher der folgenden 4 Punkte beschreibt eine(n) Vorteil(e) von DWH?
  - (a) Konsolidierung der unternehmensweiten Daten
  - (b) Höhere Datenqualität und zeitliche Analysen
  - (c) Vereinfachung des Zugriffes auf die Daten
  - (d) Alle 3 Punkte zusammen
6. Ein klassisches DWH sammelt nicht nur \_\_\_\_\_ Daten, sondern auch quasi-Echtzeitdaten, die direkt über kabellose Verbindungen aus jeder operativen Datenquelle zum DWH gelangen.
  - (a) aktuelle
  - (b) vergangene
  - (c) zukünftige
  - (d) aktuelle, vergangene und zukünftige

**Fülle richtig aus:**

7. Das Design für die Entwicklung von Data Marts aber auch von DWH basiert auf dem Konzept des \_\_\_\_\_, welches im Gegensatz zum relationalen Modell für effizientere Analysen sorgt.
8. Die Hauptaspekte bzw. -fragen, die vom Begriff \_\_\_\_\_ beschrieben werden, sind die Datenmenge eines DWH, wie schnell das DWH wachsen wird, die Anzahl der gleichzeitig auf das System zugreifenden User und die Geschwindigkeit mit der Analysen abhängig von der Größe des DWH gemacht werden können.
9. \_\_\_\_\_ ist der Prozess des Ladens und Bereitstellens von Daten über das DWH, sobald sie durch ein operatives System erfasst wurden.
10. Der \_\_\_\_\_ ist die hauptverantwortliche Person bei der Administration und Verwaltung des DWH.

**Bonus (2 Punkte)**

Nenne 4 Vorteile eines gehosteten DWHs: