

Projekt zur Lehrveranstaltung Data Warehousing

Ausgangspunkt bildet die Testdatenbank *iw_shop*, die inhaltlich der Datenbank eines Online-Shops entspricht. Die Testdatenbank enthält 7 Tabellen mit insgesamt ca. 2.2 Mio Datensätzen.

- Tabelle *iw_customer* enthält alle Stammdaten der Kunden (anonymisiert)
- Tabelle *iw_sales* enthält alle Bestelldaten
- Tabelle *iw_article* enthält alle Artikelinformationen
- Tabelle *iw_payment* enthält alle Zahlungsinformationen
- Tabelle *iw_return_header* enthält alle Kopfdaten der Retouren, mit Beziehung zur ursprünglichen Bestellung
- Tabelle *iw_return_line* enthält alle Retourendaten auf Artikelbasis
- Tabelle *iw_code_reason* enthält einige wenige Datensätze mit den Übersetzungen diverser Codes

Die Datenbank soll so aufbereitet werden, dass sie für OLAP-Anwendungen genutzt werden kann. Hierbei ist der komplette Ablauf eines kleinen OLAP-Projekts zu vollziehen

Phase 1: Analyse

1. Beschreibung möglicher Anwendungen aus Business-Sicht
Zur Unternehmenssteuerung sind entsprechende Kennzahlen zu definieren, die langfristig den Erfolg des Unternehmens beeinflussen können. Die Datenanalyse sollte in unterschiedlichen Perspektiven erfolgen. In jeder Perspektive sind Ziele und Fragestellungen zu formulieren. Für jede Perspektive müssen Messwerte bzw. Kennzahlen definiert werden, die für die Steuerung genutzt werden können. Als methodische Orientierung kann die Balanced-Scorecard-Methode genutzt werden [1],[2].
2. Konzeptuelle Modellierung
Zur konzeptuellen Modellierung der OLAP-Datenbank sind geeignete Dimensionen, Fakten bzw. Kennzahlen eines multidimensionalen Modells zu spezifizieren. Hierfür ist die ADAPT-Technik zu nutzen.
3. Datenverarbeitungsanforderungen
Es sind mögliche Auswertungen und deren Anforderungen (z.B. Häufigkeit, Aktualität, Genauigkeit) zu beschreiben.

Das Ergebnis der Analyse ist schriftlich zu dokumentieren.

Phase 2: Entwurf des Data Warehouse

1. Relationale Umsetzung eines MDM-Schemas
Es ist ein Data-Warehouse-Schema auf der Basis des multidimensionalen Datenmodells zu entwerfen, das auf relationalen Strukturen umzusetzen ist (ROLAP). Dieses bildet die Grundlage für die Erzeugung von Data Cubes, die aus der Basis-Datenbank zu laden sind.
2. Optimierung der Data Cubes
Zur Verbesserung der Performance sollten mögliche Optimierungen des DWH Schemas betrachtet werden. Hierzu zählen z.B. Denormalisierung einzelner Tabellen, Dimensionsreduktion, Verkleinerung von Cubes, Partitionierung, Indexierung, Materialisierte Sichten.

Das Ergebnis des Entwurfs ist in Form einer kommentierten SQL-Datei, versehen mit notwendigen Erläuterungen zu bestimmten Entwurfsentscheidungen, zu dokumentieren.

Phase 3: Auswertungen und Visualisierung

1. Berichte und Diagramme für Datenbank-Auswertungen.
Es sind mindestens vier ausgewählte Auswertungen aus den Bereichen Bestellungen (z.B. Warenkorbanalyse, Retourenanalyse) bzw. Kunden (z.B. Kundenaktivität, Kundenherkunft) zu implementieren. Die Auswertung sollte graphisch erfolgen unter Nutzung eines Reporting Tools (z.B. PowerPivot).
2. Evaluation der Ergebnisse
Ausgehend von der Informationsbedarfsanalyse sind die Ergebnisse der Auswertungen hinsichtlich ihrer Aussagekraft und Steuerungsfähigkeit für das Unternehmen zu bewerten.

Das Ergebnis der Phase 3 wird in einem Abschlusskolloquium präsentiert.

Termine

Phase 1 und 2 (schriftlich bzw. elektronisch): 01.07.2015

Phase 3 (mündlich): 31.08.2015

Als Datenbank-Plattformen können genutzt werden:

- Oracle 11g: <http://www.imn.htwk-leipzig.de/oracle11>
- Microsoft SQL Server 2008 R2 (2010): <https://www.microsoft.com/de-de/server-cloud/products/sql-server/default.aspx>

Für die konzeptuelle Modellierung des MDM-Schemas wird das Visio Stencil für ADAPT empfohlen [3].

Für die Datenauswertung und Visualisierung steht das Excel Add-In PowerPivot bereit, das sowohl mit Oracle als auch mit MS SQL Server zusammenarbeiten kann [4].

Das Projekt ist in Zweiergruppen durchzuführen.

Referenzen:

- [1] Balanced Scorecard: http://de.wikipedia.org/wiki/Balanced_Scorecard
- [2] Freies BSC-Tool:
<http://www.smartdraw.com/balanced-scorecard/balanced-scorecard /software.htm>
- [3] ADAPT™: http://www.symcorp.com/tech_expertise_design.html
- [4] A. Ferrari, M. Russo: Microsoft Excel 2013: Building Data Models with PowerPivot. Microsoft Press, 2013.