	Lehrveranstaltung	Datenbanken und Informationssysteme
	Aufgabe	Data-Warehousing
	Bearbeitungsbeginn	KW25
	Bearbeitungsende	KW27

Aufgabe 7: Data-Warehousing

In dieser Aufgabe erstellen Sie ein Data-Warehouse zur Analyse von Verkäufen verschiedener Artikel der Kaufhauskette „Superstore“. Ein Manager der Kaufhauskette soll durch ein Tool in die Lage versetzt werden, die erfolgten Verkäufe von Artikeln in diversen Kaufhäusern zu analysieren. Interessant sind für einen Manager vor allem, wie sich die Verkaufszahlen von Artikeln oder Produktfamilien pro Shop, Region bzw. Land in verschiedenen Zeiträumen wie Tag, Monat, Quartal oder Jahr entwickelt haben.

7.1 ETL-Prozess

Um das Data-Warehouse mit Daten zu füllen, wird im ersten Schritt vor der Datenanalyse ein ETL-Prozess implementiert. Die für die Analyse benötigten Daten liegen dabei in verschiedenen Systemen vor:

- Die Daten zu einzelnen Kaufhäusern sowie deren Artikelsortiment befindet sich innerhalb einer Unternehmensdatenbank. Mit der Unternehmensdatenbank können Sie sich via JDBC über `jdbc:db2://vsis4.informatik.uni-hamburg.de:50001/VSISP` verbinden. Die Tabellen für die Daten liegen alle innerhalb des Schemas DB2INST1.

Leider existiert für die Tabellen in der Unternehmensdatenbank keine richtige Dokumentation mehr. Es lassen sich aber leicht folgende Hierarchien über die Fremdschlüsselbeziehungen in den Tabellen erkennen:




sowie



- Die Anzahl der verkauften Waren und deren Umsatz pro Kaufhaus, Tag und Artikel liefert ein Kassensystem jeweils als CSV-Datei aus. Die Daten für die Verkäufe der letzten 5 Monate finden Sie unter <http://vsis-www.informatik.uni-hamburg.de/teaching/ss-10/dis/materialien/sales.zip>. Die erste Zeile der CSV-Datei enthält eine kurze Beschreibung der jeweiligen Spalten, der Umsatz ist jeweils in Euro ausgegeben.

Implementieren sie für den ETL-Prozess eine Java-Anwendung, welche Daten aus beiden Datenquellen extrahiert und diese in ein von Ihnen zu definierendes Zielschema des Data-Warehouses überführt; das Ziel-Schema im Data-Warehouse soll dabei als Stern-Schema realisiert werden. Achten Sie bei der Transformation auf die Konvertierung von Daten (Zahlenformate, Datumsformate) und Schemata (Hierarchien in flache Dimensionstabellen) für das Ziel-Schema des Data-Warehouses. Versehen Sie dabei Ihre Klassen mit geeigneten Ausgaben z.B. auf der Konsole, so dass sich die Abläufe im System nachvollziehen lassen.

Als weitere Nebenbedingung soll der ETL-Prozess erkennen können, ob die zu importierenden Daten schon im Data-Warehouse in den Dimensionstabellen oder Faktentabellen vorhanden sind oder nicht.

	Lehrveranstaltung	Datenbanken und Informationssysteme
	Aufgabe	Data-Warehousing
	Bearbeitungsbeginn	KW25
	Bearbeitungsende	KW27

7.2 Auswertung der Daten

Das Data-Warehouse kann nun mit Hilfe des selbst implementierten ETL-Tools gefüllt und für die Datenanalyse benutzt werden. Dafür benötigt der Manager allerdings eine weitere Anwendung, mit deren Hilfe er im Datenwürfel navigieren kann. Implementieren Sie deshalb für die Datenanalyse eine Java-Anwendung, welche Daten der folgenden Kreuztabelle ausgeben kann:

	Verkäufe	Artikel 1	Artikel 2	...	Gesamt
Hamburg	Quartal 1, 2009	12	48
	Quartal 2, 2009	31	12
	Quartal 3, 2009	50	1
	Quartal 4, 2009	2	0
	Gesamt	95	61
Bayern	Quartal 1, 2009	11	88
	Quartal 2, 2009	12	99
	Quartal 3, 2009	15	75
	Quartal 4, 2009	9	12
	Gesamt	47	274
...
Gesamt		142	335

Die Java-Anwendung soll weiterhin in der Lage sein, entlang der Dimensionen (Drill-Across) und der Klassifikationshierarchien zu navigieren (Drill-Down, Roll-Up). Benutzen sie dafür den vorgestellten Star-Query bzw. die SQL-Erweiterungen der DB2 wie GROUPING SETS, CUBE oder ROLLUP.

Hinweise

- Die Verbindung zur Datenbank im ETL-Tool sowie im Analysetool sollte via JDBC erfolgen. Als Alternative kann auch Hibernate genutzt werden.
- Die Ausgabe der Kreuztabelle in Aufgabe 7.2 kann via Konsole erfolgen.
- Eine Abgabe der Ergebnisse ist nicht erforderlich; der Erfolg der Aufgabenbearbeitung wird in den Präsenzübungen in der KW27 überprüft. Stellen Sie sicher, dass Ihre Bearbeitung der Aufgabe in KW27 auf jeden Fall abgeschlossen und flüssig präsentierbar ist.