

6. Übungsblatt

Besprechung: 12.1.2016 (Gruppe A), 12.1.2016 (Gruppe B)

Aufgabe 1

a) Gegeben sei eine Folge von 128 Bits:

$$B = 1 \times 0; 20 \times 1; 3 \times 0; 79 \times 1; 25 \times 0 \quad (\text{bzw. } 0 \underbrace{1 \dots 1}_{20 \times} 000 \underbrace{1 \dots 1}_{79 \times} \underbrace{0 \dots 0}_{25 \times}).$$

Geben Sie die WAH-Kodierung von B an. Nehmen Sie dabei eine **Word-Länge von 32 Bit** an.

b) Gegeben seien die WAH-kodierten Darstellungen von zwei Bitfolgen B_X und B_Y ¹

$$B_{X,WAH} = |010 \underbrace{1 \dots 1}_{28 \times} 0| 000 \underbrace{1 \dots 1}_{29 \times} |$$

$$B_{Y,WAH} = |11 \underbrace{0 \dots 0}_{28 \times} 010| .$$

Zeigen Sie, wie die Konjunktion

$$B_X \wedge B_Y$$

auf Basis von $B_{X,WAH}$ und $B_{Y,WAH}$ berechnet werden kann, möglichst **ohne** $B_{X,WAH}$ **und** $B_{Y,WAH}$ **zu dekodieren**. Benutzen Sie dazu das Verfahren, dass in Übungsprojekt 6 vorgestellt wurde und auf dem Artikel [1] basiert.

Aufgabe 2

In Übungsblatt 3 haben Sie einen einfachen ETL Prozess kennengelernt, den Sie hier um das inkrementelle Laden der Faktentabelle *VerkaufEinzelhandel* erweitern sollen. Dazu betrachten Sie nur einen Ausschnitt des Star Schemas von Übungsblatt 3 und der zugehörigen Datenquellen aus der AWC OLTP Datenbank, der nur die für dieses Laden relevanten Tabellen enthält wie in Abbildung 1 gezeigt.

¹Zur besseren Lesbarkeit wurden Word-Grenzen durch ‘|’ markiert.

OLTP DB → Data Warehouse

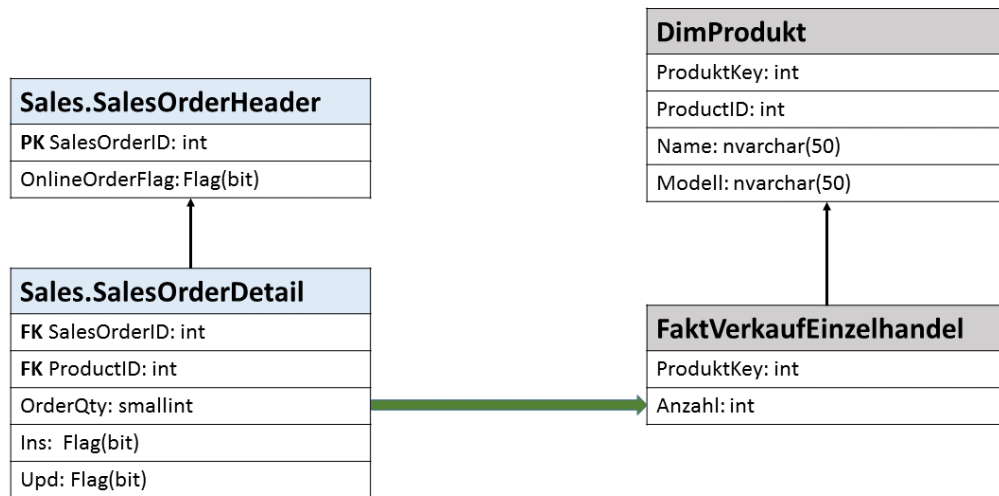


Abbildung 1: Quell- und Zieltabellen für das inkrementelle Laden des Faktus *VerkaufEinzelhandel*

Entwerfen Sie dazu den ETL Prozess mit den Komponenten in Tabelle 1, deren Funktionalität Sie wie in der Tabelle vorgegeben beschreiben sollten. Das Laden der Dimensionstabelle brauchen Sie nicht mehr zu beschreiben, sondern finden diesen Teil in der Lösung zu Übungsblatt 3 in Moodle. Dort finden Sie auch ein Beispiel für den Entwurf eines ETL Prozesses mit den Komponenten aus Tabelle 1.

Beim Entwurf können Sie sich an folgenden Punkten orientieren

1. Gehen Sie davon aus, dass Insert Flags *Ins* und das Update Flags *Upd* in der Tabelle *SalesOrderDetail* der Quelldatenbank automatisch gesetzt werden, wobei ein Update Flag für ein Tupel *nicht* mehr gesetzt wird, wenn schon ein Insert Flag gesetzt ist.
2. Überlegen Sie, wie Sie entsprechend den drei verschiedenen Gruppen von Tupeln aus *SalesOrderDetail*, d.h. mit Insert Flag, mit Update Flag oder ohne Flag, das Aggregat "Anzahl" in der Faktentabelle berechnen und ob insbesondere eine inkrementelle Berechnung möglich ist.
3. Achten Sie dabei darauf dass der ETL Prozess die Flags richtig verwaltet.

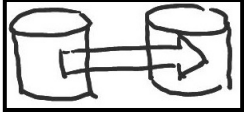

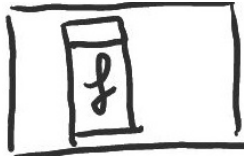
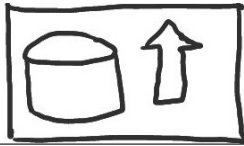


Kontrollfluss	
Datenflusstask 	Benennen Sie den Datenflusstask. Verbinden Sie mehrere Tasks zu einem Kontrollfluss. Geben Sie bei paralleler Ausführung die logische Bedingung zur Zusammenführung der Kontrollflüsse an.
Datenfluss Geben Sie den Namen des Datenflusstasks aus dem Kontrollfluss an	
Datenquelle 	Geben Sie an, welche Daten aus welcher Quelle extrahiert werden, z.B. durch einen vereinfachten SQL oder Algebra Ausdruck. Machen Sie dabei insbesondere den Join Typ deutlich. Geben Sie das Schema der extrahierten Daten in der Form Relationenname(Name 1. Spalte, . . . , Name m. Spalte) an.
Spaltenberechnung 	Geben Sie an, wie der Wert der Spalte aus den Daten im Datenfluss berechnet werden soll, z.B. durch Pseudocode. Dabei können Sie einzelne Werte von Tupeln im Datenfluss durch das Schema bezeichnen, das durch Datenquellen und Transformationen definiert wird. Benennen Sie die neu berechnete Spalte. Dabei können Sie eine bereits bestehende Spalte im Datenfluss überschreiben oder eine neue Spalte erzeugen.
Lookup 	Geben Sie das Kriterium für die Suche in einer weiteren Datenquelle an. Beschreiben Sie, wie gefundene Daten in den Datenfluss geschrieben werden sollen.
SQL Skript 	Geben Sie den vereinfachten SQL Befehl an und das Datenziel, in dem er ausgeführt werden soll.
SQL Operatoren	Daten im Datenfluss können auch durch alle bekannten SQL Operatoren verarbeitet werden, die eventuell weitere Angaben benötigen, z.B. Join Bedingung und Join Typ. Diese Operatoren können nur auf dem im Datenfluss über Datenquellen und Transformationen definierten Schema arbeiten und keine weiteren Daten verwenden.
Datenziel 	Geben Sie die Relation im Datenziel an und die Zuordnung von Spalten im Datenfluss zu Spalten in dieser Relation.

Tabelle 1: Komponenten für den Entwurf des ETL Prozesses

Literatur

- [1] Kesheng Wu, Ekow J. Otoo und Arie Shoshani: Optimizing bitmap indices with efficient compression. *ACM Trans. Database Syst*, Band 31, Nummer 1. 2006, Seiten 1-38.