

1. Was wird unter dem Differential Snapshot Problem und Record Linkage verstanden? Wo liegt der Unterschied

Die triviale Herangehensweise besteht in der vollständigen Evaluierung sämtlicher Tupel (Jeder-gegen-Jeden).

- (a) Wie kann der Aufwand für die Aufgabe der Eliminierung von Duplikaten gelöst werden? Welche Vor- und Nachteile haben die einzelnen Techniken?
- (b) Wie kann die Datenqualität im Datawarehouse weiter verbessert werden?

2. Vergleichen Sie die unterschiedlichen Anforderungen an den Speicherplatz für die Umsetzung des Datenwürfels durch MOLAP- und ROLAP-Techniken. Dafür seien Folgende Daten gegeben:

- (a) 1 Fakt; 3 Dimensionen mit je 1000 Werten; Füllgrad des Würfels 20%; 1 Attribut = 8byte
- (b) 1 Fakt; 5 Dimensionen mit je 1000 Werten; Füllgrad 20%; 1 Attribut = 8byte
- (c) 1 Fakt; 3 Dimensionen mit je 1000 Werten; Füllgrad 50%; 1 Attribut = 8byte
- (d) 1 Fakt; 5 Dimensionen mit je 1000 Werten; Füllgrad 50%; 1 Attribut = 8byte

3. MOLAP ist die direkte Speicherung eines multidimensionalen Datenwürfels in einer meist proprietären Datenstruktur. Da die Menge der Daten und ihre Persistenz eine Speicherung auf der Festplatte erfordern, müssen die Datenstrukturen *linearisiert* werden.

Erstellen Sie ein *Java*- oder *C++*-Programm, welches 2-dimensional eingeordnete Fakten aufnimmt und wieder auslesen kann. Das "Mini-DWH" soll folgende Operationen unterstützen:

- Ausgabe einzelner Werte (exact match),
- Ausgabe einer Zeile (partial match),
- Ausgabe einer Spalte

Die Datenhaltung soll dabei – ähnlich der Festplattenspeicherung – über ein *eindimensionales* Array implementiert werden. An dieser Stelle ist keine Schachtelung von Array-Objekten erlaubt.

Eine kleine Testanwendung soll die Funktionalität illustrieren. Welche Schlussfolgerungen lassen sich bezüglich MOLAP ableiten?

Diese Aufgabe ist Voraussetzung für den Schein.

4. Erweitern Sie das Programm aus Aufgabe 3 derart, dass mehr als 2 Dimensionen unterstützt werden. Ergeben sich daraus Folgen für die Anfragebearbeitung?