

## Übungsblatt 10: Speicher- und Indexstrukturen im DWH

### 1. Speicherung

Gegeben sei ein Datenwürfel mit fünf Dimensionen, jeweils 1000 Dimensionselemente und zwei Kennzahlen. Überschlagen Sie den Speicherbedarf bei MOLAP- und ROLAP-Speicherung bei einem Füllgrad von 50%, 20% und 5%.

### 2. Decomposition Storage Model (DSM)

Geben Sie für die Relation Verkauf eine Zerlegung im DSM an.

Anzahl	Ort	Kunde	Produkt
5	Magdeburg	Meier	Rotwein
3	Ilmenau	Müller	Rotwein
4	Magdeburg	Meier	Weißwein
11	Magdeburg	Kunze	Weißwein
10	Ilmenau	Kunze	Weißwein
7	Erfurt	Müller	Weißwein

### 3. Dwarf

Geben Sie einen Dwarf für die Tabelle aus Aufgabe 2 an.

### 4. Bitmap-Index

Legen Sie einen Standard-Bitmap-Index für die folgende Kundentabelle an.

KNr	Nachname	Vorname	Stadt	Geburtsjahr
555666	Maier	Thomas	Berlin	1983
555667	Mustermann	Max	Ulm	1984
555668	Schulz	Maik	Magdeburg	1990
555669	Spörl	Meier	Ulm	1990
555670	Marzer	Klaus	Magdeburg	1984
555671	Nandau	Tilo	Magdeburg	1983
555672	Schaber	Mike	Ulm	1991

Überlegen Sie sich verschiedene Strategien für die einzelnen Spalten. Was sind mögliche Einsatzgebiet für Bitmap-Indexe? Diskutieren Sie Vor- und Nachteile. Wie werden Exact-Match-, Partial-Match und Range-Queries unterstützt?

### 5. Bereichskodierter Bitmap-Index

Erstellen Sie einen bereichskodierten Bitmap-Index über das Geburtsjahr der Tabelle Kunde aus Aufgabe 4. Welche Vektoren müssen genutzt werden, um die Kunden zu finden, die zwischen 1985 und 1990 geboren wurden?

### 6. Intervallkodierter Bitmap-Index

Überlegen Sie sich passende Intervalle eines intervallkodierten Bitmap-Index für das Geburtsjahr der Kundentabelle. Welchen Einfluss hat die Größe des Intervalls auf den Speicherbedarf und den Berechnungsaufwand? Welche Vektoren müssen genutzt werden, um die Kunden zu finden, die zwischen 1985 und 1990 geboren wurden?

### 7. Bitmap-Indexe im Vergleich

Die 16 Bundesländer seien als Zahlen von 1..16 in lexikographischer Reihenfolge kodiert. Legen Sie jeweils einen Standard-Bitmap-Index, einen Mehrkomponenten-Bitmap-Index (4+4), einen bereichskodierten Bitmap-Index und einen intervallkodierten Index (Intervallbreite 5) an. Geben Sie die Bit-Kodierungen jeweils in einer Tabelle für alle Bundesländer an.

Wie müssten folgende Anfragen in den einzelnen Varianten beantwortet werden?

- a) Gesucht wird das Land Sachsen-Anhalt
- b) Alle Bundesländer lexikographisch nach Hessen.
- c) Die Bundesländer von Hessen bis Thüringen.

### 8. CUBE-Operator

Geben Sie für die Berechnung aller Gruppierungskombinationen der Menge (Produkt, Filiale, Verkaufstag) drei Paare von Kombinationen an, die direkt ohne Sortierung auseinander abgeleitet werden können, sowie drei Paare von Kombinationen, die eine Sortierung erfordern.