

## **Prüfungsschwerpunkte „Data Warehousing“**

Die Prüfungsklausur dauert 90 Minuten. Es dürfen keine Unterlagen verwendet werden. Der Schwerpunkt der Fragen wird zu Themengebieten aus Vorlesung und Seminar gehören, die durch das begleitende Projekt nicht oder nur peripher berührt wurden.

### **1. Einführung und Grundbegriffe**

OLAP vs. OLTP

DBMS-Techniken für Data Warehousing

Begriffe Data Warehouse, Data Warehousing, Data Mart, Business Intelligence

Charakteristika von Data Marts

5 V's in Big Data

### **2. Architektur von Data Warehouses**

Anforderungen an die Architektur

FASMI

Phasen des Data Warehousing

Arten von Datenquellen

DWS-Referenzarchitektur, Komponenten

### **3. Modellierung**

ROLAP - MOLAP - HOLAP Prinzip und Vergleich

Charakteristika Star Schema vs. Snowflake Schema

Slowly Changing Dimensions - unterschiedliche Berichtsanforderungen (z.B. as-is, as-of)

### **4. ETL-Prozess**

ETL-Prozess und Komponenten

Vergleich von Extraktionstechniken (Aktualität, Belastung)

Differential Snapshot Problem allgemein

Transformation

Aspekte von Heterogenität: Datenmodell, Schema, Datenrepräsentation

Datenfehler / Datenkonflikte auf Schema- und Instanzebene

Beispiele für Schema Mapping

Vermeidung von Datenfehlern, Data Profiling

Normalisierung und Standardisierung (Beispiele)

Verfahren zur Duplikaterkennung

Vergleichsfunktionen (distanzbasiert, q-Gramme, Soundex)

### **5. OLAP-Operationen**

Funktionsweise des CUBE- und ROLLUP-Operators in SQL (erläutern)

SQL:2003 Anwendung sequenzbasierter Operationen (Partitionierung, dynamische Fenster)

MDX vs. SQL (Unterschied allgemein)

### **6. Speicherung**

MDC-Tabellen

Bewertung von MOLAP / mögl. Nachteile multidimensionaler Speicherung

Vergleich Speicherverbrauch Array vs. Star Schema

Prinzip des Dwarf

Horizontale vs. vertikale Partitionierung in DWH (insb. Von Datenwürfeln)

Mini-Dimensionen zum Umgang mit sparse data

### **7. Column Stores**

Motivation für spaltenorientierte Speicherung

Row Store vs. Column Store

Speicherung von Column Stores: DSM, TID-Listen, PAX

Prinzip und zu lösende Aufgaben bei Hauptspeicher-DB

Run Length Encoding, Dictionary Encoding

Delta-Relationen

## **8. Indexstrukturen für DWH**

Klassifikation von Indexstrukturen (mehrdimensionale vs. zusammengesetzte Indexe)

Anwendung von Bitmap-Indexen

- \* Standard-Bitmap-Index
- \* Mehrkomponenten-Bitmap-Index
- \* Bereichskodierter Bitmap-Index
- \* Intervallkodierter Bitmap-Index

Auswahl von Indexen entsprechend Anfragetyp

Verbundindexe

## **9. Anfrageverarbeitung und materialisierte Sichten**

Ablauf einer Star Query anhand eines vorgegebenen Beispiels

Berechnung eines CUBE, Aggregationsgitter

Materialisierte Sichten:

- \* Grundgedanke
- \* Problemstellungen: Query Containment, Query Rewriting
- \* Aktualisierung materialisierter Sichten (Parameter)
- \* Konsistenzaspekte materialisierter Sichten

## **10. Ausblick Data Mining**

Zusammenhang DWH - Data Mining - Business Intelligence

Einordnung DWH in den Knowledge Discovery Prozess

Angrenzung von DW-Anwendungsfällen vs. Data-Mining-Verfahren