

1. Data Cubes

Für eine BI-Anwendung soll ein in-memory Datenbanksystem verwendet werden, das die Daten auch intern in Form eines Datenwürfels verwaltet. Für die Dimensionierung des DB-Servers ist es notwendig den Speicherbedarf für die Datenbank abzuschätzen.

Folgende Parameter spielen dabei eine Rolle:

- Anzahl der Dimensionen: d
- Anzahl der Ausprägungen (Werte) pro Dimension: n_1, \dots, n_d
- Speicherbedarf für eine Zelle: b Bytes

- a) Wie hoch ist der Speicherbedarf für diesen Würfel? (Größenordnung)
- b) Welche Metadaten müssen zur Verwaltung dieses Würfels zusätzlich gespeichert werden und wie hoch ist der Speicherbedarf für die Metadaten?
- c) Um welchen Faktor wird der Speicherbedarf höher wenn der Würfel um eine zusätzliche Dimension mit n_{d+1} Ausprägungen erweitert werden soll
- d) Um welchen Faktor wird der Speicherbedarf höher wenn in einer Dimension eine zusätzliche Ausprägung hinzukommt?
- e) Um welchen Faktor wird der Speicherbedarf höher wenn zusätzlich **alle** aggregierten Werte im Datenwürfel gespeichert werden sollen?

2. Pivot-Tabellen

Im Moodle finden Sie eine Excel-Datei (*sales.zip*) mit Verkaufsdaten eines fiktiven Onlineshops. Importieren Sie die Daten in ein Werkzeug das Pivot-Tabellen unterstützt (z.B.: MS Excel) und beantworten Sie folgenden Fragen:

- Wie viele Bestellungen wurden pro Monat ausgeliefert (Liniendiagramm)? ok
- Wie viele Bestellungen wurden pro Woche ausgeliefert (Liniendiagramm)? ok
- Wie hoch sind die Bestellsummen pro Monat (Liniendiagramm)? ok
- Wie hoch ist die durchschnittliche Bestellsumme pro Monat und pro Jahr?
- Wie sind die Aufträge über die Wochentage verteilt (Balkendiagramm)? Was stellen Sie dabei fest?
- Welche Kunden haben besonders häufig bestellt (top-10)?
- Welche Kunden haben besonders viel Umsatz generiert (top-10)?
- Eine Bestellung besteht aus mehreren Positionen. Wie hoch ist der relative Anteil der Bestellungen mit nur einer Position, mit zwei Positionen usw. (Balkendiagramm)?
- Für die Bestellungen mit nur einer Position, welche Artikel wurden am häufigsten bestellt (top-10)?
- Für den häufigsten Artikel und Bestellungen mit genau zwei Positionen, welche Artikel wurden am häufigsten gemeinsam mit dem ersten Artikel bestellt (top-10)?
- Welche Artikel wurden im Jänner 2011 besonders häufig bestellt?
- Welche Artikel wurden im Juni 2011 besonders häufig bestellt?
- Wie hoch schätzen Sie die Erträge für das angelaufene Monat (Jänner 2012)

Notiz: Gegebenenfalls müssen Sie die Datenbasis um zusätzliche Spalten erweitern.

3. Anwendungen

Beschreiben Sie einen Anwendungsfall aus ihrem näheren Umfeld für den ein Data Warehouse zweckmäßig erscheint.

- Welche Daten würden in dem Data Warehouse gespeichert?
- Welche Quellsysteme liefern die Daten?
- Wer soll das Data Warehouse benutzen und für welche Zwecke soll es genutzt werden?