



Aufgabe 1

Berechnen Sie die durch Dictionary-Kodierung¹ erzielte Kompressionsrate der Spalte BELNR von der Tabelle SAPHPB.ACDOCA unter folgenden Annahmen:

- a) Die unkomprimierte Spalte ist als Vektor mit Werten fester Breite umgesetzt. Die feste Breite entspricht der maximalen Zeichenanzahl der gespeicherten Werte in Bytes.
- b) Die komprimierte Spalte ist als Vektorpaar (Dictionary, Attributvektor) umgesetzt.

Dabei gelten für das Dictionary die Annahmen aus a).

Der Attributvektor speichert Wert-IDs fester Breite. Die feste Breite soll als möglichst kleine Bit-Anzahl gewählt werden (ohne Padding).

Die Größe des Attributvektors soll gegebenenfalls auf volle Bytes aufgerundet werden.

Aufgabe 2

Berechnen Sie die durch Run-Length-Encoding¹ erzielte Kompressionsrate der Spalte BELNR von der Tabelle SAPHPB.ACDOCA unter folgenden Annahmen:

- a) Die Spalte ist sortiert.
- b) Der mit Run-Length-Encoding komprimierte Attributvektor ist als Vektorpaar (Werte, Startpositionen) umgesetzt.

Der Werte-Vektor entspricht dem unkomprimierten Dictionary.

Der Startpositionen-Vektor speichert Offsets fester Breite. Die feste Breite soll als möglichst kleine Bit-Anzahl gewählt werden (ohne Padding).

Die Größe des Startpositionen-Vektors soll gegebenenfalls auf volle Bytes aufgerundet werden.

Aufgabe 3

Berechnen Sie die durch Delta-Kodierung¹ erzielte Kompressionsrate des in Aufgabe 1 spezifizierten Dictionary der Spalte BELNR von der Tabelle SAPHPB.ACDOCA unter folgenden Annahmen:

- a) Das Dictionary ist sortiert.
- b) Das mit Delta-Kodierung komprimierte Dictionary ist als Byte-Vektor umgesetzt. Alle Werte sind in einem Block komprimiert. Gemeinsame Präfixlängen, die Anzahl zusätzlicher Zeichen, sowie die einzelnen zusätzlichen Zeichen können und werden jeweils mit einem Byte repräsentiert.

Skizzieren Sie außerdem für die ersten fünf Werte des komprimierten Dictionary die gespeicherten Werte mit ihren Byte-Offsets.

Verwenden Sie die Daten der folgenden HANA-Datenbankinstanz als Basis für Ihre Lösungen (Der Zugriff auf die Datenbank ist nur aus dem HPI-Netz möglich. Ihr Laptop muss außerdem unter <https://byod.hpi.de/> für die Nutzung interner HPI-Ressourcen registriert sein.)

Host Name: vm-syene.eaalab.hpi.uni-potsdam.de (192.168.30.23)

Instance Nummer: 02

User Name: STUDENTS

Password: Students17



Abgabeanweisung

Die Aufgaben **müssen** in Zweierteams bearbeitet werden.

¹ Die Komprimierungsverfahren sind wie in der Vorlesung vorgestellt zu verwenden.

Ihre Lösungswege müssen nachvollziehbar sein (gegebenenfalls stichpunktartig begründet werden, inklusive verwendeter SQL-Statements).

Die Abgabe erfolgt über das HPI-Abgabesystem:

<https://www.dcl.hpi.uni-potsdam.de/submit/dashboard/>

Alle Studenten müssen sich dafür beim Abgabesystem anmelden.

Der Name des Partners muss bei der Abgabe der Ergebnisse angegeben werden.

Reichen Sie eine pdf-Datei mit ihren Lösungen ein.

Abgabefrist: 20. Juni 2019 Anywhere on Earth (AoE).

Die pünktliche Abgabe und das Bestehen der Übungsblätter sind Voraussetzung für die Zulassung zur Klausur.