Prof. Dr. Gunter Saake
Lehrstuhl Datenbank- und Informationssysteme

Datenbanken II (SS 2019)

Exercise 8

Task 1 Given the following table ist folgende Tabelle:

Name	MatrNr	Sex	Start of studies	Study Program
Tom	15874	M	2002	INF
Claudia	18091	W	2006	CV
Klaus	14342	M	2003	INF
Susanne	11012	W	2001	INF

(a) Compress the Sex using Bit-Vector Encoding!

(b) Compress the Start of Studies using Delta Encoding!

(c) Compress the Study Program using Run-Length Encoding!

(d) Compress the Name using Dictionary Encoding!

Task 2 Given the following data set:

8, 12, 9, 5, 1, 3, 10, 14, 4, 6, 7, 11, 2, 13, 15, 16

Sort the data set using external sorting (Partition Size: 2). Why is external sorting suitable for sorting large data sets in databases?

Task 3 Scans

(a) What are the different kinds of a scan?

(b) Sketch the principles of the scan strategies?

(c) When do you use which scan strategy?

(d) What is a scan semantic and what is it used for?

Task 4 Explain the following join algorithms:

(a) Nested-Loop Join

(b) Block Nested-Loop Join

(c) Sort-Merge Join

(d) Hash Join

Use these strategies to join the Tables 1 and 2 on Person.PID=Residence.PID. For hashing, use the hash function h(x) = x mod 3 and a block size of 2 for the block nested loop join. Which of these algorithms could adjusted also be used for grouping?

PID	First Name	Last Name
1	Mareicke	Müller
2	Tom	Meier
3	Frank	Schmitt
5	Stefan	Schulz

Tabelle 1: Person

PID	Town	Street	Number
2	Hannover	Lindenstr	7
5	Magdeburg	Schillerstr	5
3	Hannover	Breiter Weg	3
1	Hamburg	Hafenstr	14
3	Berlin	Gartenstr	1

Tabelle 2: Wohnort

Task 5 Which problems exist for classical hash-joins in multi-core systems? How can the Radix join solve these problems?

Good Luck!