Summer Term 2019

Advanced Topics in Databases

Exercise 7

1. **[Group 17]** Classify BitWeaving, Column Imprints based on the following features:
   - storage structure vs access path
   - one- vs multi-dimensional

   Explain your categorization.

2. **[Group 1]** Given the following data set

$$C = [1, 0, 0, 1, 2, 3, 2, 3, 1, 0, 2, 3]$$

   (a) Build BitWeaving/H for the given data set. Assume a word size of 8-bit.What is the storage saving compared to storing each value as an 8-bit value?

   (b) Execute the query $C < 1$ and create the resulting bit mask.

3. **[Group 2]** Given the following data set:

$$C = [1, 0, 0, 1, 2, 3, 2, 3, 3, 3, 3, 3, 3, 3, 3, 3]$$

   (a) Build BitWeaving/V for the given data set. Assume a word size of 8-bit.What is the storage saving compared to storing each value as an 8-bit value?

   (b) Execute the query $C < 1$ and create the resulting bit mask.

   (c) Where is an early pruning possible?

4. **[Group 3]** Given the following dataset:

$$C = [12, 13, 18, 8, 2, 3, 2, 3, 1, 1, 1, 4, 6, 7, 8, 9]$$

   (a) Create a Column Imprint for the given data set. Assume 8-bit Imprints, 32-bit integer as column values and 16 byte cache lines. Sample the data range on every second value in the data space.

   (b) Execute the following queries on the created Imprint (you can express the result as bitmap or position list):

      i. $C < 5$,

      ii. $C = 8$,

      iii. $5 < C < 10$.

5. **[Group 4]** How well can Column Imprints and BitWeaving support the following use cases?

   (a) Equality predicates (e.g., $C = 4$)
   (b) Between predicates (e.g., $2 < C < 7$)
   (c) Multi-dimensional queries (e.g., $C_1 < 5$ AND $C_2 < 300$)
   (d) Analytical workloads with frequent insertions (e.g., daily sales).
   (e) Storing and querying strings (char or varchar)

6. **[Group 5]** Given the following dataset:

TID	C_1	C_2	C_3	C_4
T1	0	1	0	1
T2	1	1	1	0
T3	1	0	1	0
T4	0	0	0	2

   (a) Build a tree-based Elf with the following column Order: C_3,C_4,C_2,C_1.
   (b) Linearize the build elf.
   (c) Execute the following queries:
      i. $C_3 = 0$
      ii. $C_2 < 1$ AND $0 < C_3 < 2$
      iii. $C_1 = 1$

Good Luck!