

17.02.2025 Week 1 exercises: Storage

Exercise 1:

Given relations R and S, and the following query:

```
select Ra from R, S
where Rc = Sb and 5<Ra<20 and 40<Rb<50 and 49<Sa<65
```

Give the execution steps of the query in a row store, and in a column store. Specifically, provide the output of each operator given a row store or a column store.

R			S	
Ra	Rb	Rc	Sa	Sb
3	12	12	17	11
16	34	34	49	35
56	75	53	58	62
9	45	23	99	44
11	49	78	64	29
27	58	65	37	78
8	97	33	53	19
41	75	21	61	81
19	42	29	32	26
35	55	0	50	23

Exercise 2:

Assume a single-relation schema with a relation $R(A_0, \dots, A_{127})$, with 128 4-byte integer columns and $||R||$ number of records. Consider a columnar layout. The page size is 8kB. Assume page metadata consumes no space. Consider select-project queries (no joins, no aggregates) which use a total of k columns (by using them in the select predicate theta or choosing them for output). Compute the cost measured in number of pages read; disregard seeks and in-core computation time.

Exercise 3:

You have a 100 Gb table of the following format:

```
<int col1, int col2, int col3, int col4, int col5>.
```

The following queries are very frequent on this table:

Q1: SELECT col1 FROM tbl

Q2: SELECT col3,col5 FROM tbl

Q3: SELECT col2 FROM tbl

Q4: SELECT col4 FROM tbl

Design a good storage layout for this table that would help on the performance of these four queries. Explain your answer briefly.

Exercise 4:

Assume the tables are stored in column layout. Which compression techniques presented in the lecture would you choose for the following tables? You are allowed to combine compression techniques. Assume that you can bit-pack the keys of dictionary keys.

(Note: do not think about possible future data for this table, just try to compress the given data as well as possible.)

visitors		
ID	Downloads	IP_address
13	217	138.92.122.175
81	0	138.92.122.195
42	6	138.92.122.182
25	4	138.92.122.181
21	52	138.92.122.177
56	4	138.92.122.188
78	2	138.92.122.191
30	1	138.92.122.185
23	0	138.92.122.179
80	2	138.92.122.193
27	3	138.92.122.183
82	0	138.92.122.197

issues		
ID	Status	Subject
1	In Progress	Migrate Moodle site to Turnkey VM
2	In Progress	Come up with backup strategy
5	New	Test recovery of Moodle site
6	Resolved	Drink fifth coffee
7	Resolved	Buy next coffee
8	Resolved	Test group selection for students
9	Resolved	Set up Moodle site