

Databases Autumn 2025

Hand-In Exercise 5

December 4, 2025

Aiysha Frutiger

Jannick Seper

Luis Tritschler

Total Points	
---------------------	--

Task 1

Task 2

i Parameters given in the sheet:

Parameter	Symbol	Value
Attribute size	s_a	6 B
Tuple size	s_t	12 B
Cardinality of R	$\text{Card}(R)$	50 000
Page size	page	8192 B
Fill degree (data pages)	f_{data}	0.9
Fill degree (index pages)	f_{index}	0.7
Page header	head	48 B
Pointer size	s_p	6 B

ii **FF(A)**

Attribute A is uniformly distributed in [0..999], therefore **FF(A = a) = 1/1000**.

iii **FF(B)**

Attribute B is not uniformly distributed:

- 40 000 tuples are uniformly distributed over the values [0..99] (100 values)

The tuples are spread uniformly across 100 values so each value occurs $\frac{40\,000}{100} = 400$ times.

Therefore **FF(B = 10) = $\frac{400}{50\,000}$** .

- 10 000 tuples are uniformly distributed over the values [100..999] (900 values)

Here the tuples are spread uniformly across 900 different values so each value occurs $\frac{10\,000}{900} \approx 11.11$ times.

Therefore **FF(B = 500) = $\frac{10\,000/900}{50\,000}$** .

1. No index (Layout R)

Page-layout: $\frac{(\text{page-head}) \cdot f_{\text{data}}}{r_{\text{avg}} + p_{\text{slot}}} \Rightarrow x = 407$

(with $r_{\text{avg}} = s_t$ and $p_{\text{slot}} = s_p$)

Number of data pages: $\lceil \text{Card}(R)/x \rceil \Rightarrow \underline{N\text{Pages}(R) = 123}$

Since there is no index: $\Rightarrow \mathbf{C(A = 10) = C(A = 500) = C(B = 10) = C(B = 500) = 123}$

2. Indirect B+ tree on A (Layout RA)

Leaf capacity: $\left\lceil \frac{(\text{page-head} - 2 \cdot p_{\text{leaf}}) \cdot f_{\text{index}}}{k + r_k \cdot p_{\text{rec}}} \right\rceil \Rightarrow \underline{t_{\text{leaf}} = 18}$

(with $k = s_a$, $r_k = 50$, $p_{\text{leaf}} = p_{\text{rec}} = s_p$)

Number of leaf pages: $\lceil \frac{n_k}{t_{\text{leaf}}} \rceil \Rightarrow \underline{n_{\text{leaf}} = 56}$

(with $n_k = 1000$)

Inner node capacity: $\left\lceil \frac{((\text{page-head}) \cdot f_{\text{index}}) - p}{k + p} \right\rceil \Rightarrow \underline{e_i = 474}$

(with $p = s_p$)

Height: $\lceil \log_{e_i+1}(n_{\text{leaf}}) \rceil + 1 \Rightarrow \mathbf{h = 2}$

Index-only selection cost: $(h - 1) + \lceil FF(A = a) \cdot n_{\text{leaf}} \rceil \Rightarrow \mathbf{C(A = 10) = C(A = 500) = 2}$

Queries on B require table scan: $\Rightarrow \mathbf{C(B = 10) = C(B = 500) = 123}$

3. Indirect B+ tree on B (Layout RB)

The index structure parameters are the same as RA:

$$\Rightarrow t_{\text{leaf}} = 18, n_{\text{leaf}} = 56, e_i = 474,$$

$$\text{Height: } \lceil \log_{e_i+1}(n_{\text{leaf}}) \rceil + 1 \Rightarrow \mathbf{h} = 2$$

$$\text{Queries on A require table scan: } \Rightarrow \mathbf{C(A = 10) = C(A = 500) = 123}$$

$$\text{Index-only selection cost: } (h-1) + \lceil FF(B=b) \cdot n_{\text{leaf}} \rceil \Rightarrow \mathbf{C(B = 10) = C(B = 500) = 2}$$

4. Two indirect indexes on A and B (Layout RAB)

Both RA and RB exist.

Heights and Cost identical (to calculation of indexes):

$$\Rightarrow \mathbf{h} = 2$$

$$\Rightarrow \mathbf{C(A = 10) = C(A = 500) = 2}$$

$$\Rightarrow \mathbf{C(B = 10) = C(B = 500) = 2}$$

5. Clustered, direct index on A (Layout RA\$)

Means physically sorted by A \rightarrow Leaf pages = table pages.

The parameters are the same as before:

$$\Rightarrow x = 407, n_{\text{leaf}} \rightarrow NPages(R) = 123, e_i = 474,$$

$$\text{Height: } \lceil \log_{e_i+1}(n_{\text{leaf}}) \rceil + 1 \Rightarrow \mathbf{h} = 2$$

$$\text{Cost for A: } (h-1) + \lceil FF(A=a) \cdot n_{\text{leaf}} \rceil \Rightarrow \mathbf{C(A = 10) = C(A = 500) = 2}$$

$$\text{B has no usable index: } \Rightarrow \mathbf{C(B = 10) = C(B = 500) = 123}$$

6. Combined B+ tree on (A,B) (Layout RC)

Combined search key: $s_A + s_B \Rightarrow k = 12$.

$$\text{Number of distinct key pairs: } 1000 \cdot 1000 \Rightarrow n_k = 1\,000\,000.$$

$$\text{Average number of tuples per key pair: } \frac{Card(R)}{n_k} \Rightarrow r_k = 0.05.$$

$$\text{Leaf capacity (page-layout indirect index): } \left\lceil \frac{(\text{page-head} - 2 \cdot p_{\text{leaf}}) \cdot f_{\text{index}}}{k + r_k \cdot p_{\text{rec}}} \right\rceil \Rightarrow t_{\text{leaf}} = 462.$$

$$\text{Number of leaf pages: } \left\lceil \frac{n_k}{t_{\text{leaf}}} \right\rceil \Rightarrow n_{\text{leaf}} = 2165$$

$$\text{Inner node capacity: } \left\lceil \frac{((\text{page-head}) \cdot f_{\text{index}}) - p}{k + p} \right\rceil \Rightarrow e_i = 316.$$

$$\text{Height of the index: } \lceil \log_{e_i+1}(n_{\text{leaf}}) \rceil + 1 \Rightarrow \mathbf{h} = 3$$

$$\text{Cost for queries on A: } (h-1) + \lceil FF(A=a) \cdot n_{\text{leaf}} \rceil \Rightarrow \mathbf{C(A = 10) = C(A = 500) = 5}.$$

$$\text{B alone cannot use the combined index (no leading A): } \Rightarrow \mathbf{C(B = 10) = C(B = 500) = 123}.$$

Summary:

Layout	A=10	B=10	A=500	B=500	Height
R	123	123	123	123	no index
RA	2	123	2	123	2
RB	123	2	123	2	2
RAB	2	2	2	2	2
RA\$	2	123	2	123	2
RC	5	123	5	123	3

Task 3