

Databases Autumn 2025

Hand-In Exercise 5

December 4, 2025

Aiysha Frutiger
Jannick Seper
Luis Tritschler

Total Points	

Task	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	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 $\mathbf{FF(A = a) = 1/1000}$.

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 $\mathbf{FF(B = 10) = \frac{400}{50\ 000} = 0.008}$.

- 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 $\mathbf{FF(B = 500) = \frac{10\ 000/900}{50\ 000} \approx 0.000222}$.

1. No index (Layout R)

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

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

Number of data pages: $\lceil Card(R)/x \rceil \implies NPages(R) = 123$

Since there is no index: $\implies C(A = 10) = C(A = 500) = C(B = 10) = C(B = 500) = 123$

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

Leaf capacity: $t_{leaf} = \left\lfloor \frac{(page-head-2s_p)\cdot f_{index}}{k+r_k\cdot s_p} \right\rfloor \implies t_{leaf} = 18$

(with $k = s_a$, $n_k = 1000$, $r_k = 50$)

Number of leaf pages: $\lceil n_k/t_{leaf} \rceil \implies n_{leaf} = 56$

Inner node capacity: $\left\lfloor \frac{(page-head)\cdot f_{index}-s_p}{k+s_p} \right\rfloor \implies e_i = 474$

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

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

Queries on B require table scan: $\implies C(B = 10) = C(B = 500) = 123$

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

The index structure parameters are the same as RA:

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

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

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

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

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

Both RA and RB exist.

Heights and Cost identical (to calculation of indexes):

$$\Rightarrow h = 2$$

$$\Rightarrow C(A = 10) = C(A = 500) = 2$$

$$\Rightarrow C(B = 10) = C(B = 500) = 2$$

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

Means physically sorted by A → Leaf pages = table pages.

The parameters are the same as before:

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

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

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

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

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

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	2	123	2	123	2

Task 3