1. Disk math-
   a) Transfer time
   b) Capacity of Cylinder
   c) Average number of head move if its on 4000th track. There are 8001 tracks numbered 0 to 8000.

2. B+ Tree and B tree math
   Block size = 4096-byte, key size = 8-byte, pointer size = 12 byte
   a) Maximum number if key hold on block?
   b) Maximum number of pointers to records of level 3 B+ Tree?
   c) Maximum number of pointers to records of level 3 B Tree?

3. Variable length math- 27000 records, 1 primary block hold 12 record and full, overflow block hold 4 records…
   a) I/O? b) dense index? C) sparse index?

4. Query math…. Cluster relation… find I/O?

5. Extensible/Linear hashing 65 % capacity…

6. Hash join… I/O

DBMS 2020/2021

1. (6 points)

Consider a disk with the following characteristics:
The disk rotates at 3000 rpm.
The disk has 7000 tracks per surface. The tracks are numbered from 1 to 7000.
The disk has 8 surfaces.
Each track is divided into 16 sectors, and a sector holds 512 bytes.
Blocks consist of 4 sectors. 20% of the circumference of each track is occupied by gaps.
The time it takes the head to move n tracks is 1+0.003n ms.

a) What is the transfer time for one block? 4.75
b) Assume that the head is currently located at track 7000. We want to fulfil request
Track different from 7000. How many tracks does the head have to move

a)

2.

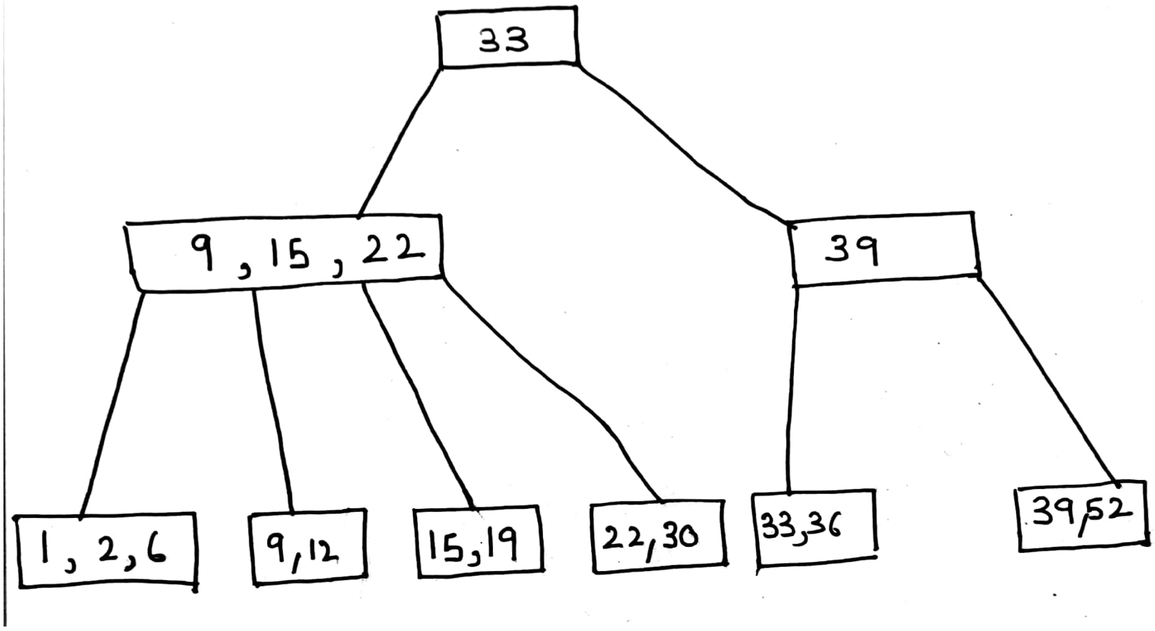
3. (9 points)

Consider B+ tree of order 3

a) Delete from the tree below the key 36.

b) Insert into the tree below the key 5. (Do not use the result of a)!)

c) What is the maximum number of indexed keys we can have in a B+ tree of order 3 with four levels?



4. (13 points)

Suppose blocks consist of 4096 bytes. We want to store the relation R (A, B, C). The size of A is 16 bytes, the size of B is 27 bytes, and the size of C is 35 bytes. Each record consists of 3 data fields and a record header. The record header consists of four fields with 4 bytes for each of them. A block has a block header that consists of eight fields with 6 bytes for each of them. We have T(R)= 100000. For part a), c) and d) fields can start at any byte.

a) what is the minimum number of blocks that we need to store the relation R (A, B, C) if we use spanned storage of records? Assume that the management of spanned records does not require additional space.

b) Repeat a) when we have the additional requirement that fields must start at a byte that is a multiple of 8.

c) We want to create a sparse index for the attribute C on the file created in a). A pointer has a size of 8 bytes. The blocks are described above. The index entries themselves do not have a header. The index entries are stored unspanned. What is the minimum number of blocks that we need for the index file?

d) Consider the query $\pi_{C}(\sigma_{A=17}(R))$. Assume that a projection is not eliminating duplicates. We have V (R, A) =25 and DOM (R, A) = 200. We assume that 17 is an arbitrary domain value. Calculate the minimum number of blocks to store the estimated result relation. The records of the result relation do not have a header. The blocks are as described above. Use unspanned storage of records.

5. (6 Points)

Below are some statistics for four relations W, X, Y and Z.

W (a, b)	X (b, c)	Y (c, d)	Z (c, d)
T(W) = 20	T(X) = 900	T(Y) = 800	T(Z) = 50
V (W, a) = 10	V (X, b) = 10	V (Y, c) = 4	V (Z, c) = 2
V (W, b) = 5	V (X, c) = 30	V (Y, d) = 5	V (Z, d) = 1

We want to perform the natural join for all four relations. What is the join-order selected by the greedy algorithm? What is its cost (measured by the sum of the sizes of the intermediate relations)? Provide the results of all intermediate steps.

6. (6 Points)

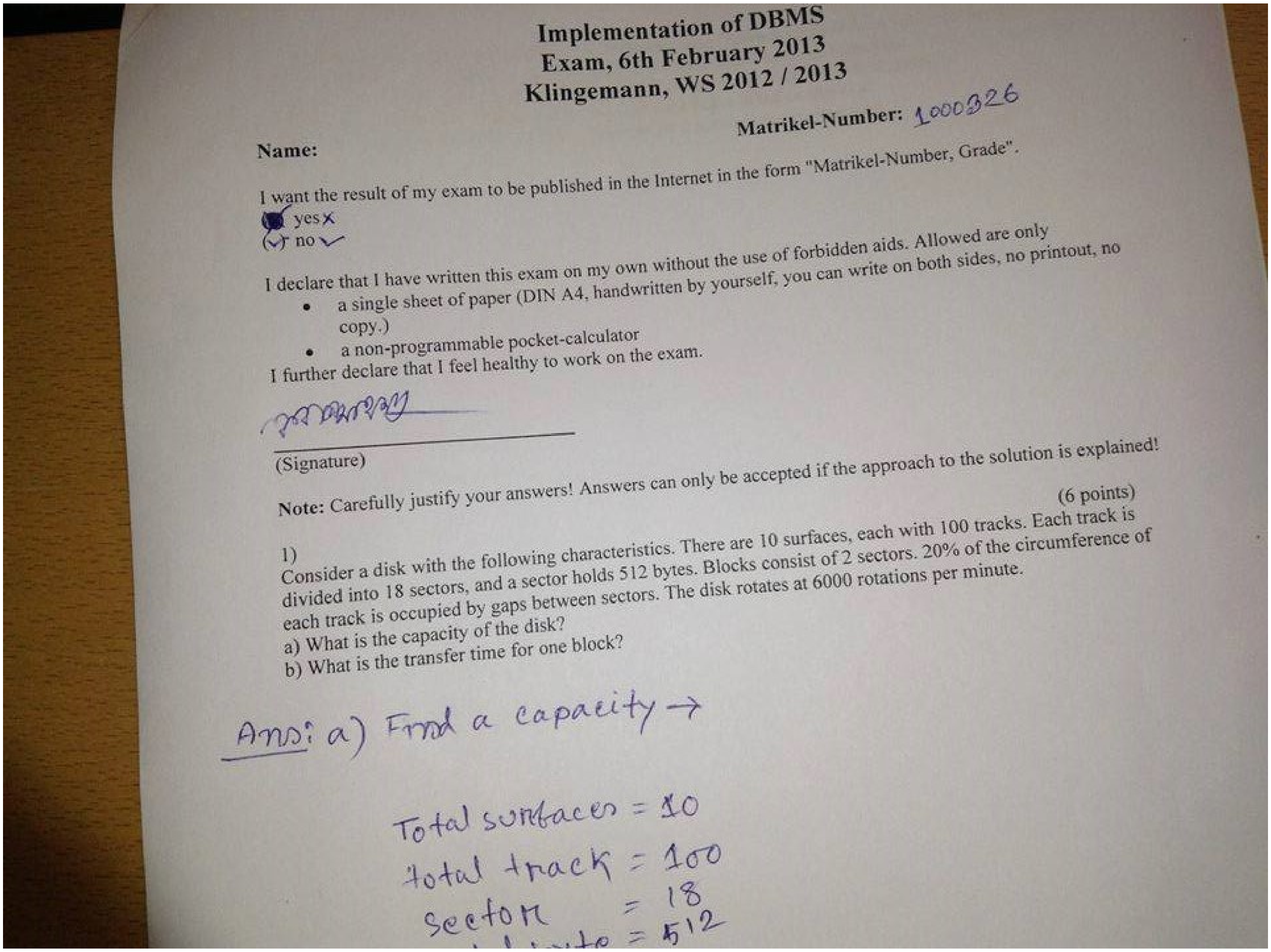
We have two relations R (A, B) and S (B, C) with the following statistics:

- T (R) = 120000, B(R) = 6000, V(R, A) =60000, V(R, B) = 10000
- T (S) = 90000, B(S) = 30000, V (S, B) = 90000, V(S, C) = 10000

Assume that the relations are clustered, i.e., are stored contiguously. The attribute B is a key in S and a foreign key in R. For the relation S we have an index on the attribute B. The index has three levels. Ignore the final output I/O cost.

a) We want to perform a natural join of R and S by using an index join algorithm. We assume that the index is completely in main memory. What is the required number of I/O's?

b) We want to perform a natural join of R and S by using an index join algorithm. We assume that the root of index is in the main memory but the blocks from the other levels are not. What is the required number of I/O's?



Implementation of DBMS
Exam, 6th February 2013
Klingemann, WS 2012 / 2013

Matrikel-Number: 1000326

Name:

I want the result of my exam to be published in the Internet in the form "Matrikel-Number, Grade".

● yes ✗
(✓) no ✓

I declare that I have written this exam on my own without the use of forbidden aids. Allowed are only

- a single sheet of paper (DIN A4, handwritten by yourself, you can write on both sides, no printout, no copy.)
- a non-programmable pocket-calculator

I further declare that I feel healthy to work on the exam.

(Signature)

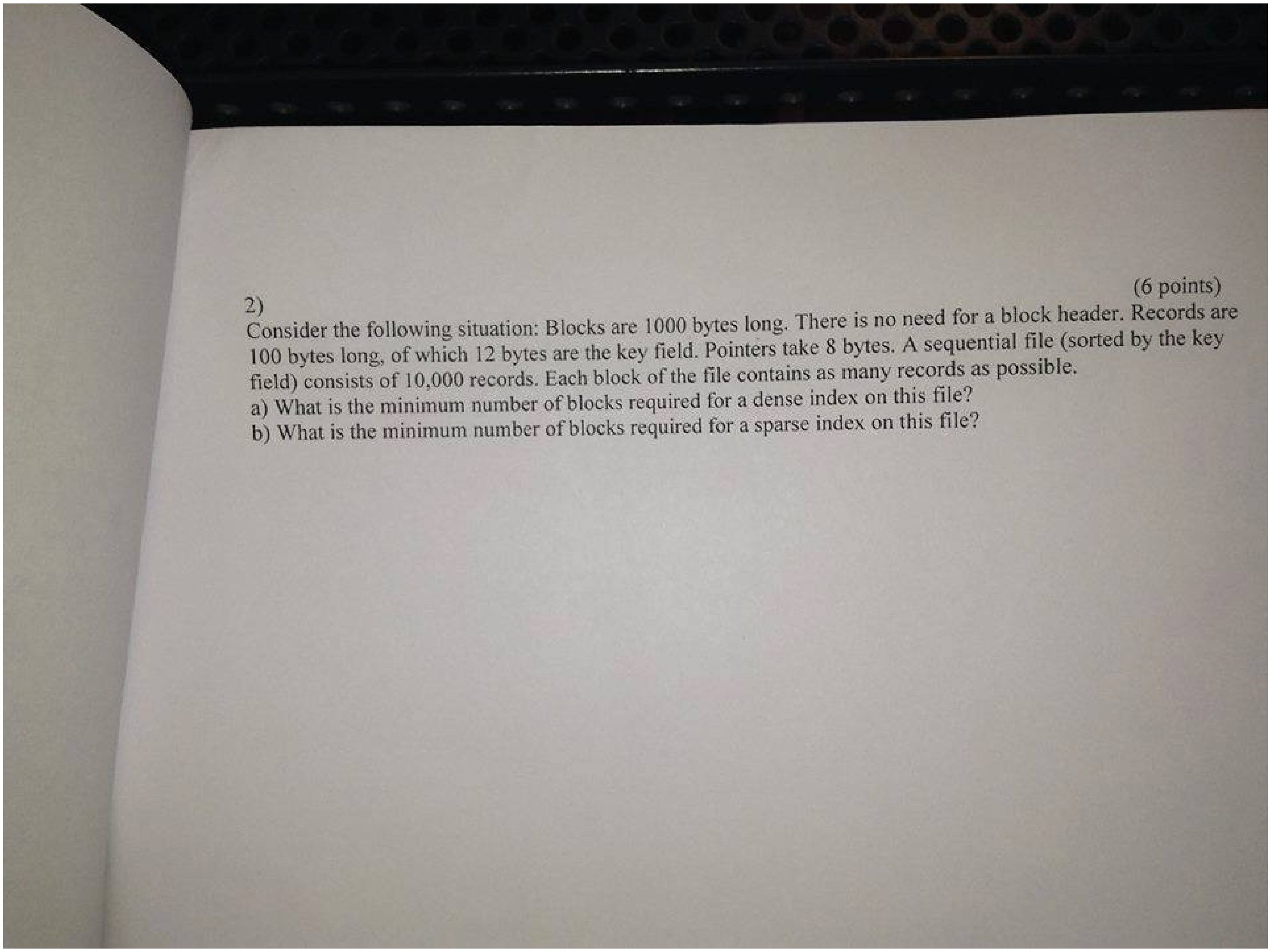
Note: Carefully justify your answers! Answers can only be accepted if the approach to the solution is explained!

(6 points)

1)
Consider a disk with the following characteristics. There are 10 surfaces, each with 100 tracks. Each track is divided into 18 sectors, and a sector holds 512 bytes. Blocks consist of 2 sectors. 20% of the circumference of each track is occupied by gaps between sectors. The disk rotates at 6000 rotations per minute.
a) What is the capacity of the disk?
b) What is the transfer time for one block?

Ans: a) Find a capacity →

Total surfaces = 10
total track = 100
sector = 18
= 512



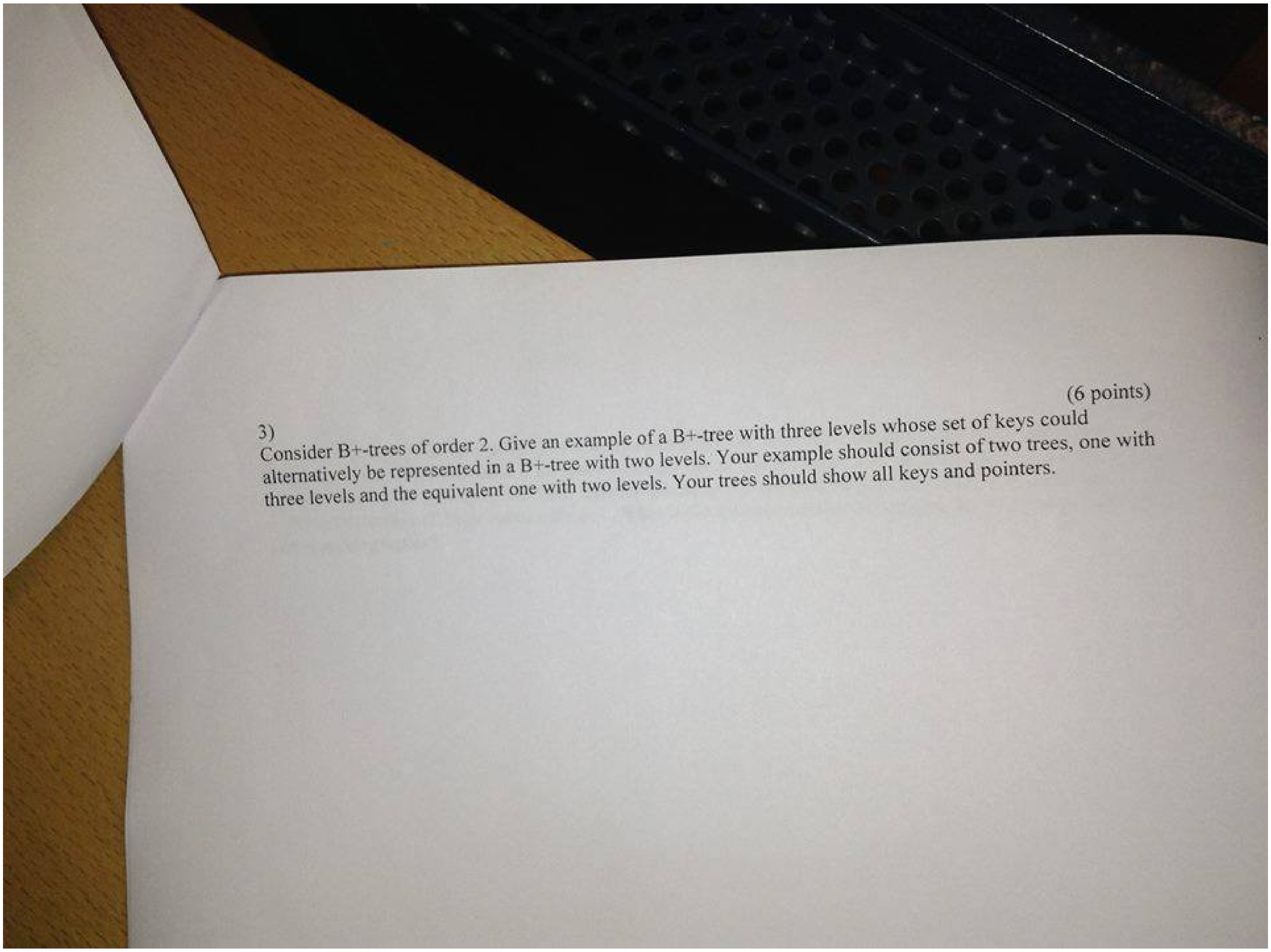
(6 points)

2)

Consider the following situation: Blocks are 1000 bytes long. There is no need for a block header. Records are 100 bytes long, of which 12 bytes are the key field. Pointers take 8 bytes. A sequential file (sorted by the key field) consists of 10,000 records. Each block of the file contains as many records as possible.

a) What is the minimum number of blocks required for a dense index on this file?

b) What is the minimum number of blocks required for a sparse index on this file?



(6 points)

3)
Consider B+-trees of order 2. Give an example of a B+-tree with three levels whose set of keys could alternatively be represented in a B+-tree with two levels. Your example should consist of two trees, one with three levels and the equivalent one with two levels. Your trees should show all keys and pointers.

(6 points)

4)
Suppose we store relation R(a, b, c) in a partitioned hash table with 1024 buckets. That is, the hash function produces 10 bits. Queries about R all specify exactly one of the attributes a, b, or c, and each of the attributes is equally likely to be specified. The hash function produces 5 bits based only on the value of a, 3 bits based only on the value on b and 2 bits based only on c. What is the average number of buckets that must be searched to find matching tuples?

(6 points)

5)
Let R(A, B), S(B, C) and T(A, B, C) be three relations. Consider the following relational algebra expressions:

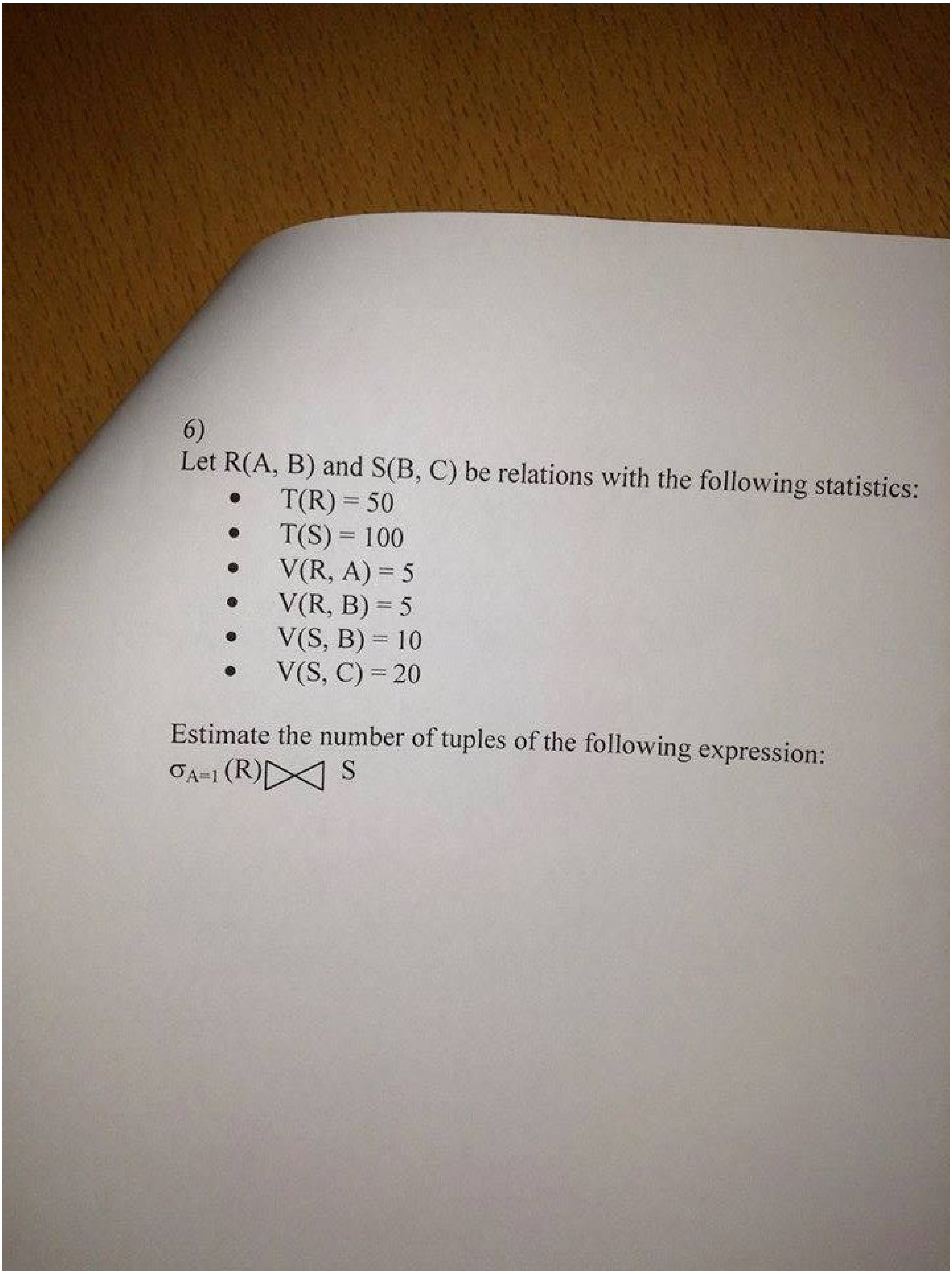
E1: $\Pi_{A,C}[\{\sigma_{B<10}(R) \bowtie S\} - T]$
E2: $\Pi_{A,C}[\{R \bowtie \sigma_{B<10}(S)\} - \sigma_{B<10}(T)]$
E3: $\Pi_{A,C}[\{\Pi_{A}(R) \bowtie \sigma_{B<10}(S)\} - T]$

Are some of the expressions equivalent to each other?

Ar

Now,

Capa



6)

Let R(A, B) and S(B, C) be relations with the following statistics:

- T(R) = 50
- T(S) = 100
- V(R, A) = 5
- V(R, B) = 5
- V(S, B) = 10
- V(S, C) = 20

Estimate the number of tuples of the following expression:

$\sigma_{A=1}(R) \bowtie S$

Q.7 We have three relations with the following statistics:

- B(Order)= 3000 blocks
- B(Cust)= 1000 blocks
- B(Book)=100 blocks

Assume that the relations are stored contiguously, i.e, they are clustered relations. You should further assume that each operation uses memory efficiently. You can ignore final output I/O cost.

a. We want to perform a selection of "Price<10" over block. We have 10 blocks of main memory. What is the required number of I/Os?
b. We want to perform a one pass join of Order and Cust. How many main-memory-blocks do you need?
c. We want to perform a one pass join of order and Cust and have sufficient main memory. What is the required number of I/Os?
d. We want to perform a hash join (without the "hybrid optimization") of order and Book. How main memory-blocks do you need?
e. We want to perform a hash join (without the "hybrid optimization") of order and Book and have sufficient main-memory. What is the required number of I/Os?

DBMS

Q.1 Consider a disk with the following characteristics:
There are 10 surfaces each with 100 tracks. Each track is divided into 18 sectors and a sector holds 512 bytes. Blocks consist of 2 sectors. 20% of the circumference of each track is occupied by gaps between sectors. The disk rotates at 6000 rotations per minute.

a. What is the capacity of the disk?
b. What is the transfer time for one block?

Q.2 Consider the following situation: Blocks are 1000 bytes long. There is no need for a block header. Records are 100 bytes long, of which 12 bytes are the key filed. Pointers take 8 bytes. A sequential file (sorted by the key filed) consists of 10000 records. Each block of the file contains as many records as possible.

a. What is the minimum number of block required for a dense index on this file?
b. What is the minimum number of blocks required for a sparse index on this file?

Q.3 Consider B+-tree of order 2. Give an example of a B+ tree with three levels whose set of keys could alternatively be represented in a B+ tree with two levels. Your example should consist of two trees, one with three levels and the equivalent one with two levels. Your trees should show all keys and pointers.

Q.4 Suppose we store relation R (a, b, c) in a partitioned hash table with 1024 buckets. That is, the hash function produces 10 bits. Queries about R all specify exactly one of the attributes a, b or c and each of the attributes is equally likely to be specified. The hash function produces 5 bits based only on the value of a, 3 bits based only on the value on b and 2 bits based only on c. What is the average number of buckets that must be searched to find matching tuples?

Q.5 Let R(A,B), S(B,C) and T(A,B,C) be three relations. Consider the following relational algebra expressions:

E1: $\pi_{A,C}[\{\sigma_{B<10}(R) \bowtie S\}-T]$

E2: $\pi_{A,C}[\{R \bowtie \sigma_{B<10}(S)- \sigma_{B<10}(T)]$

E3: $\pi_{A,C}[\{\pi_{A}(R) \bowtie =\sigma_{B<10}(S)\}-T]$

Are some of the expressions equivalents to each other?

Q.6 Let R(A,B), S(B,C) be relations with the following statistics:
T(R)=50
T(S)=100
V(R,A)=5
V(R,B)=5
V(S,B)=10
V(S,C)=20

Estimate the number of tuples of the following expression: $\sigma_{A}=1(R) \bowtie S$

Q.7 We have three relations with the following statistics:

- B(Order)= 3000 blocks
- B(Cust)= 1000 blocks
- B(Book)=100 blocks

Assume that the relations are stored contiguously, i.e, they are clustered relations. You should further assume that each operation uses memory efficiently. You can ignore final output I/O cost.

a. We want to perform a selection of “Price<10” over block. We have 10 blocks of main memory. What is the required number of I/Os?
b. We want to perform a one pass join of Order and Cust. How many main-memory-blocks do you need?
c. We want to perform a one pass join of order and Cust and have sufficient main memory. What is the required number of I/Os?
d. We want to perform a hash join (without the “hybrid optimization”) of order and Book. How main memory-blocks do you need?
e. We want to perform a hash join (without the “hybrid optimization”) of order and Book and have sufficient main-memory. What is the required number of I/Os?

https://classes.soe.ucsc.edu/cmps277/

1) Question was on disk to Calculate

(i) Capacity

(ii) Transfer time.

2) Disk 1 and Disk 2 are data disks.
Disk 3 is mirror of Disk 1 and Disk 4 is parity checker (Recovery disk) of disk 2 and disk 3.

Identify which of the pairs of disks can be recovered if crashed simultaneously.

a) Disk 1 and Disk 3

b) Disk 2 and Disk 3

c) Disk 2 and Disk 4

d) Disk 1 and Disk 4

e) Disk 1 and Disk 2

3) ~~# Records = 3000, 5 records per block, The Index used is B+ tree Sparse of order 10.~~ or a

3) A file has 3000 records. A block can store either 5 records or node of a B^+ tree of order 10. The index used is sparse. Calculat the blocks required to store ~~Relation~~ file and index.

4) R has no duplicates and S may have duplicates. Identify whether which of the following are equivalent.

(i) $\delta(R \times S)$

(ii) $\delta(\delta(R) \times S)$

(iii) $\delta(R \times \delta(S))$

(iv) $R \times \delta(S)$

(v) $\delta R \times S$.

5) Question on greedy algorithm

6) Question was framed in such a way u need to have good understanding of every term in Linear Hashing (like n, m, i, C)

7) $B(R) = 9000$ $B(S) = 4000$.

we are using optimised variant of Hashing and ~~but we are~~ regular Merge variant. Main memory of 20 blocks.

a) Minimum of passes required in Merge join.

b) No of I/O required to join.

c) No of passes in Hash join.

d) No of I/O's required to join.

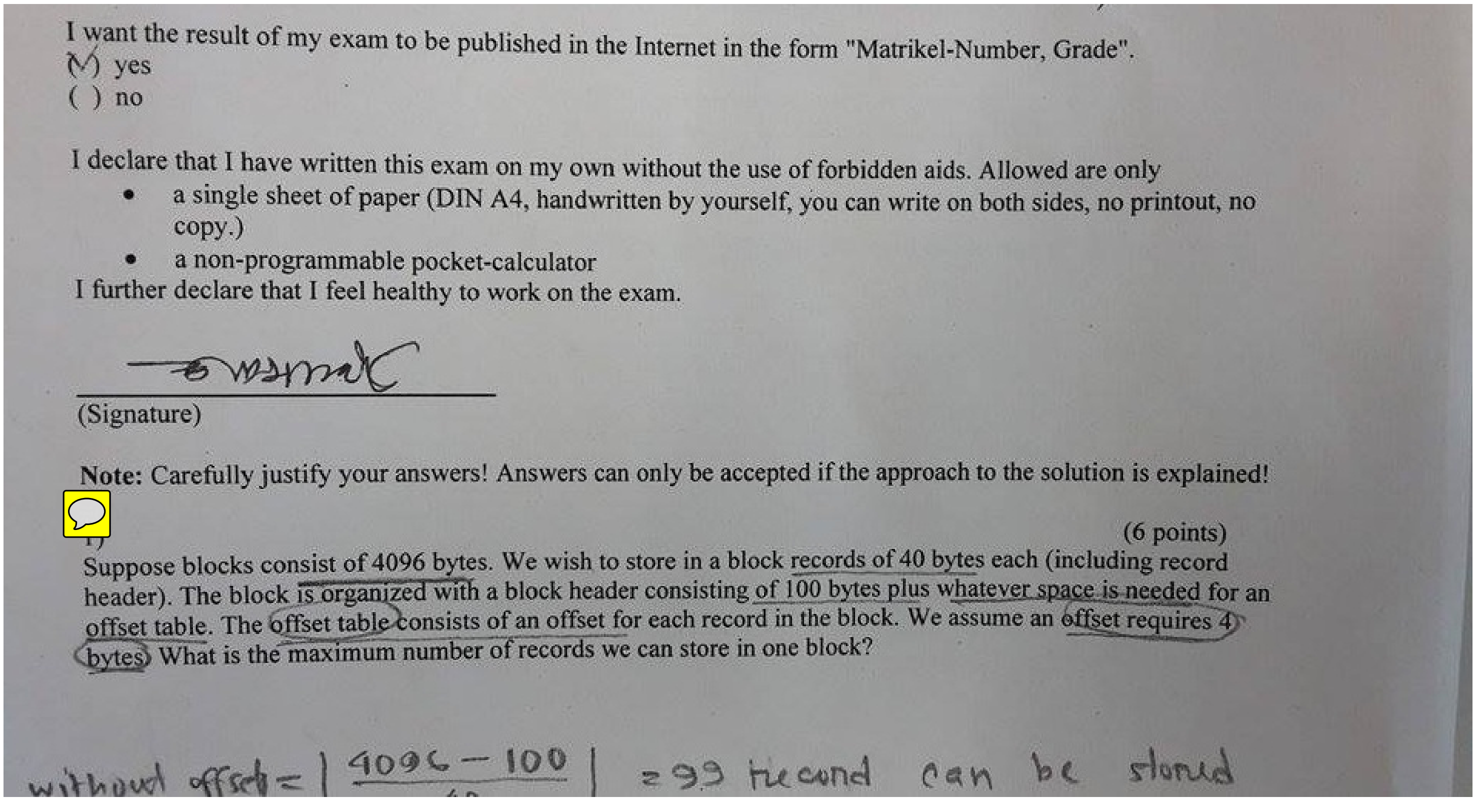
e) S is clustered index on B and non-clustered on C. Identity the optimal query plan and estimate the value. V(S, b) = 5000 V(S, c) = 1,000,000.

$\sigma_{B=10 \wedge C=2}(S)$.

T(S) = 1,000,000

B(S) = 400,000

{S is clustered and design}



I want the result of my exam to be published in the Internet in the form "Matrikel-Number, Grade".

(✓) yes

( ) no

I declare that I have written this exam on my own without the use of forbidden aids. Allowed are only

- a single sheet of paper (DIN A4, handwritten by yourself, you can write on both sides, no printout, no copy.)
- a non-programmable pocket-calculator

I further declare that I feel healthy to work on the exam.

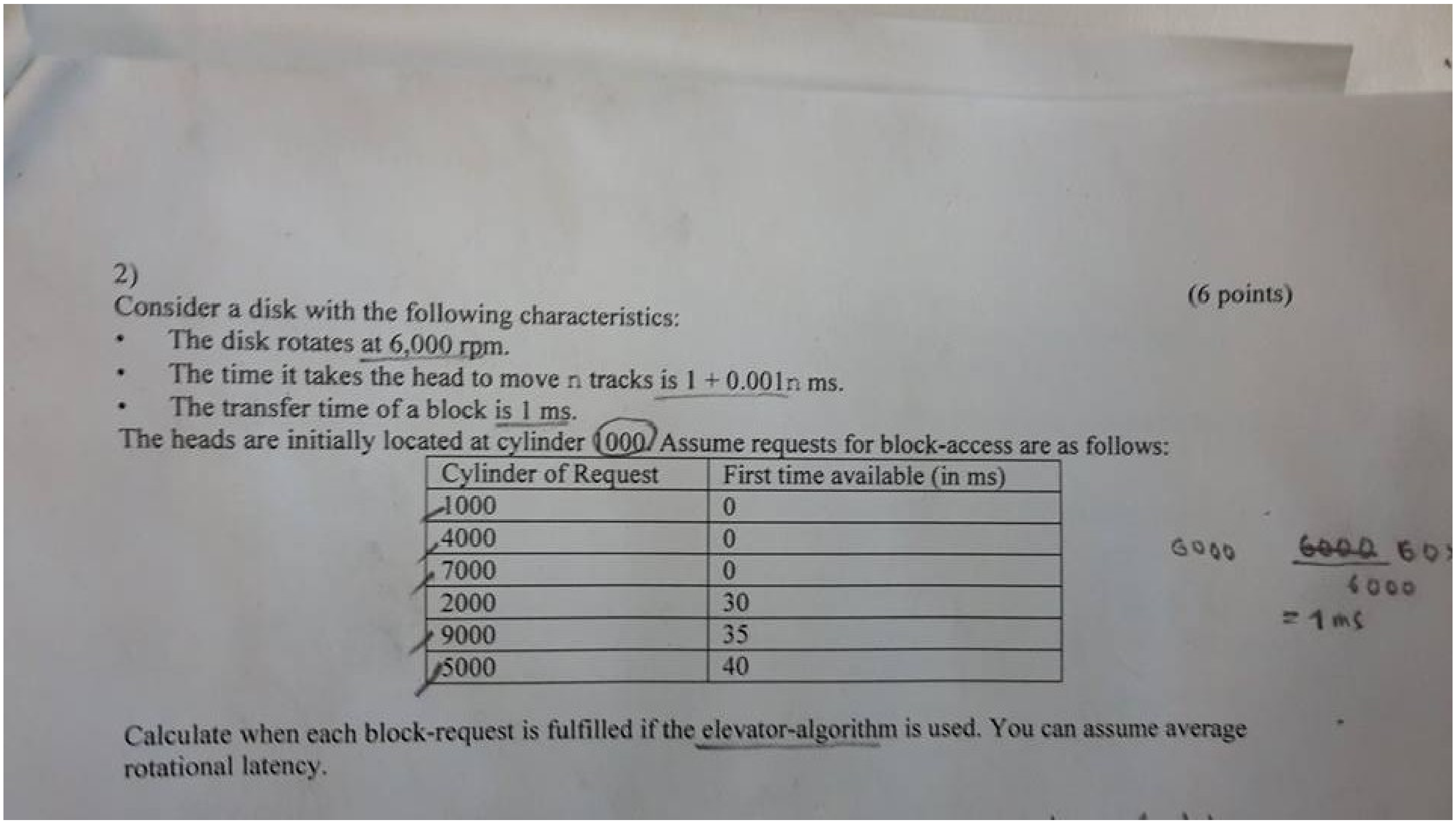
(Signature)

Note: Carefully justify your answers! Answers can only be accepted if the approach to the solution is explained!

1) (6 points)

Suppose blocks consist of 4096 bytes. We wish to store in a block records of 40 bytes each (including record header). The block is organized with a block header consisting of 100 bytes plus whatever space is needed for an offset table. The offset table consists of an offset for each record in the block. We assume an offset requires 4 bytes. What is the maximum number of records we can store in one block?

without offset = $\left| \frac{4096 - 100}{40} \right|$ = 99 record can be stored



2)

Consider a disk with the following characteristics: (6 points)

- The disk rotates at 6,000 rpm.
- The time it takes the head to move n tracks is 1 + 0.001n ms.
- The transfer time of a block is 1 ms.

The heads are initially located at cylinder 1000. Assume requests for block-access are as follows:

Cylinder of Request	First time available (in ms)
1000	0
4000	0
7000	0
2000	30
9000	35
5000	40

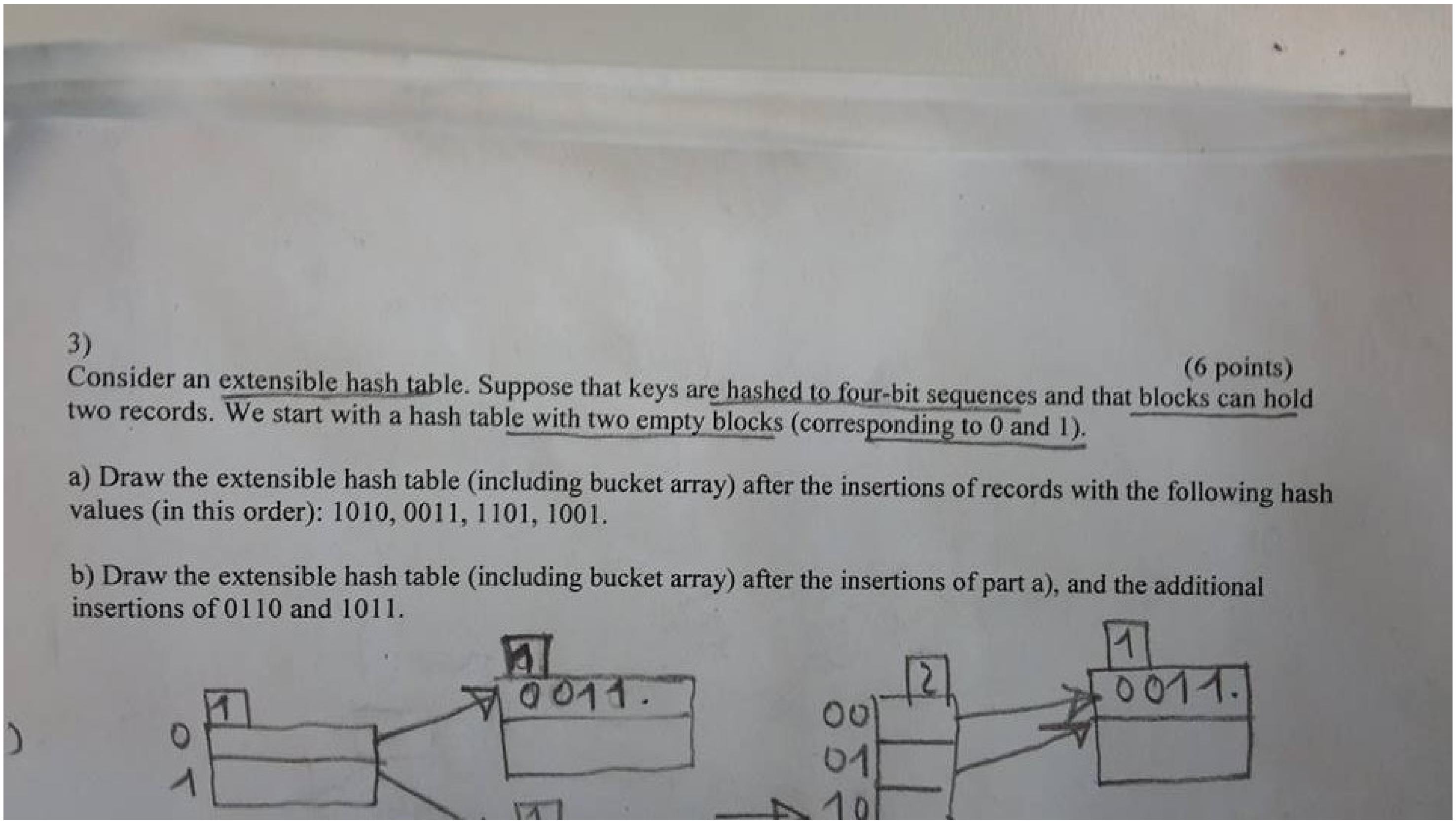
6000

~~6000~~ 60×

6000

= 1 ms

Calculate when each block-request is fulfilled if the elevator-algorithm is used. You can assume average rotational latency.



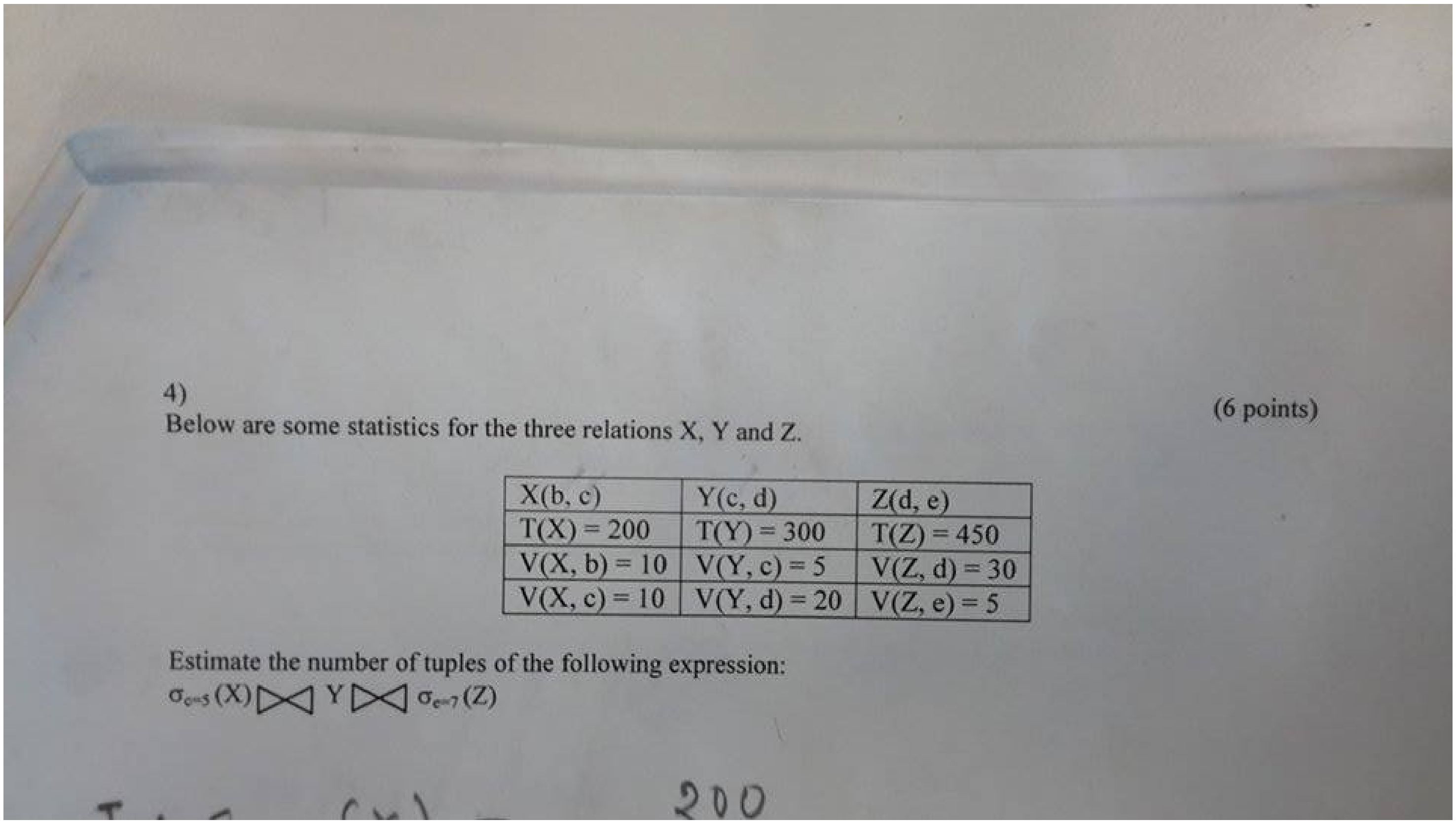
3)

(6 points)

Consider an extensible hash table. Suppose that keys are hashed to four-bit sequences and that blocks can hold two records. We start with a hash table with two empty blocks (corresponding to 0 and 1).

a) Draw the extensible hash table (including bucket array) after the insertions of records with the following hash values (in this order): 1010, 0011, 1101, 1001.

b) Draw the extensible hash table (including bucket array) after the insertions of part a), and the additional insertions of 0110 and 1011.



4) (6 points)

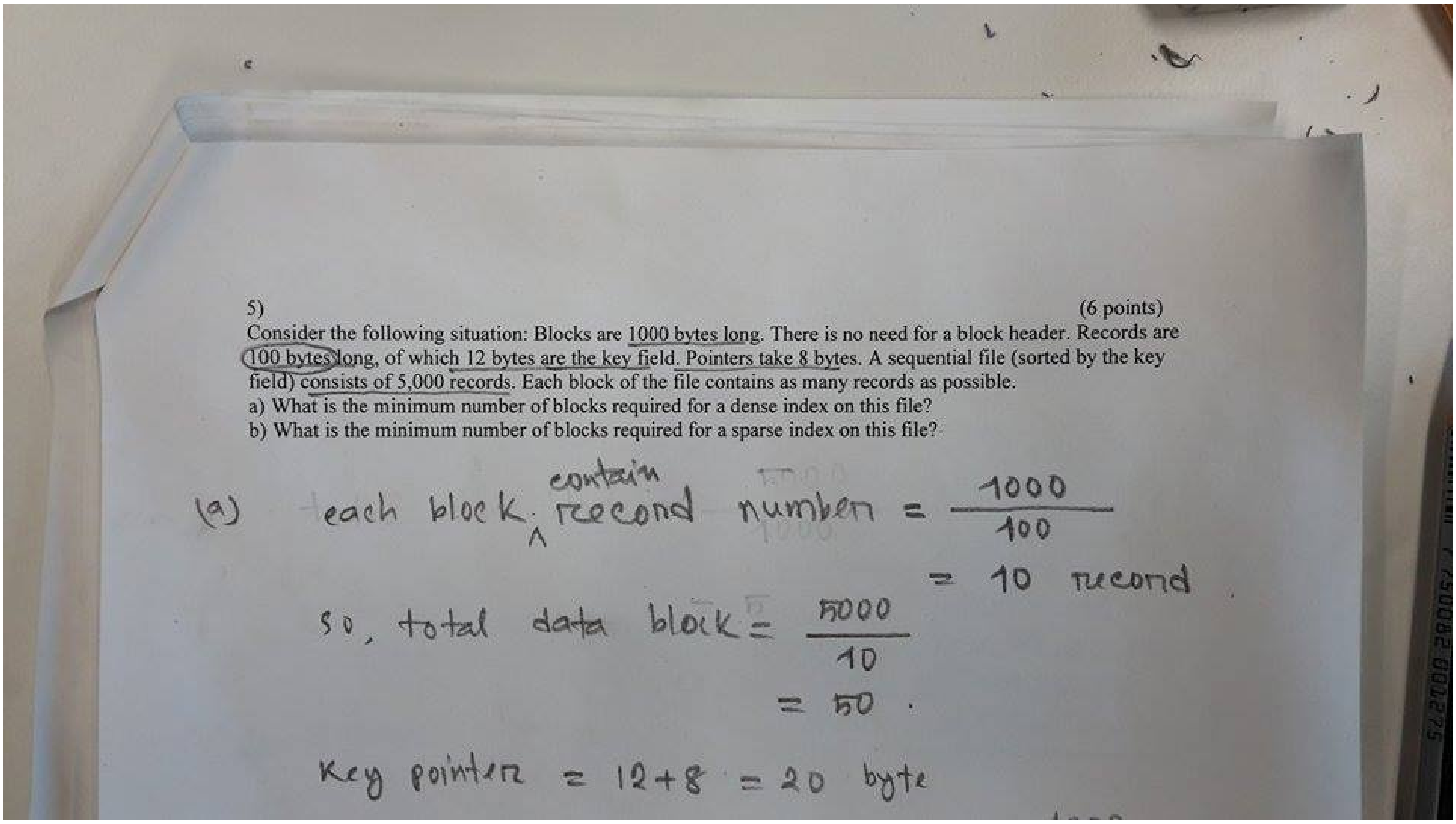
Below are some statistics for the three relations X, Y and Z.

X(b, c)	Y(c, d)	Z(d, e)
T(X) = 200	T(Y) = 300	T(Z) = 450
V(X, b) = 10	V(Y, c) = 5	V(Z, d) = 30
V(X, c) = 10	V(Y, d) = 20	V(Z, e) = 5

Estimate the number of tuples of the following expression:

$\sigma_{c=5}(X) \bowtie Y \bowtie \sigma_{e=7}(Z)$

200



5) (6 points)

Consider the following situation: Blocks are 1000 bytes long. There is no need for a block header. Records are 100 bytes long, of which 12 bytes are the key field. Pointers take 8 bytes. A sequential file (sorted by the key field) consists of 5,000 records. Each block of the file contains as many records as possible.

a) What is the minimum number of blocks required for a dense index on this file?

b) What is the minimum number of blocks required for a sparse index on this file?

(a) each block contain record number $= \frac{1000}{100}$

$= 10$ record.

so, total data block $= \frac{5000}{10}$

$= 50$.

Key pointer $= 12+8 = 20$ byte

DBMS 2020/2021

1. (6 points)

Consider a disk with the following characteristics:
The disk rotates at 3000 rpm.
The disk has 7000 tracks per surface. The tracks are numbered from 1 to 7000.
The disk has 8 surfaces.
Each track is divided into 16 sectors, and a sector holds 512 bytes.
Blocks consist of 4 sectors. 20% of the circumference of each track is occupied by gaps.
The time it takes the head to move n tracks is 1+0.003n ms.

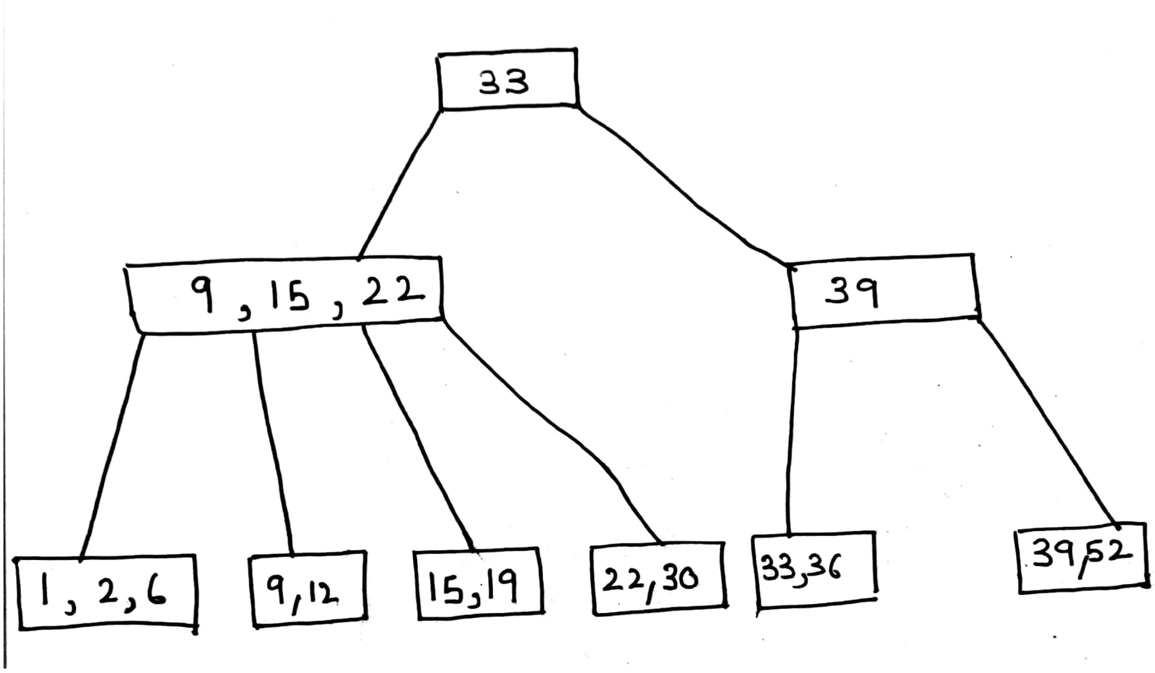
a) What is the transfer time for one block? 4.75
b) Assume that the head is currently located at track 7000. We want to fulfil request
   Track different from 7000. How many tracks does the head have to move

2.

3. (9 points)

Consider B+ tree of order 3

a) Delete from the tree below the key 36.
b) Insert into the tree below the key 5. (Do not use the result of a)!)
c) What is the maximum number of indexed keys we can have in a B+ tree of order 3 with four levels?



4. (13 points)

Suppose blocks consist of 4096 bytes. We want to store the relation R (A, B, C). The size of A is 16 bytes, the size of B is 27 bytes, and the size of C is 35 bytes. Each record consists of 3 data fields and a record header. The record header consists of four fields with 4 bytes for each of them. A block has a block header that consists of eight fields with 6 bytes for each of them. We have T(R)= 100000. For part a), c) and d) fields can start at any byte.

a) what is the minimum number of blocks that we need to store the relation R (A, B, C) if we use spanned storage of records? Assume that the management of spanned records does not require additional space.

b) Repeat a) when we have the additional requirement that fields must start at a byte that is a multiple of 8.

c) We want to create a sparse index for the attribute C on the file created in a). A pointer has a size of 8 bytes. The blocks are described above. The index entries themselves do not have a header. The index entries are stored unspanned. What is the minimum number of blocks that we need for the index file?

d) Consider the query $\pi_C(\sigma_{A=17}(R))$. Assume that a projection is not eliminating duplicates. We have V (R, A) =25 and DOM (R, A) = 200. We assume that 17 is an arbitrary domain value. Calculate the minimum number of blocks to store the estimated result relation. The records of the result relation do not have a header. The blocks are as described above. Use unspanned storage of records.

5. (6 Points)

Below are some statistics for four relations W, X, Y and Z.

W (a, b)	X (b, c)	Y (c, d)	Z (c, d)
T(W) = 20	T(X) = 900	T(Y) = 800	T(Z) = 50
V (W, a) = 10	V (X, b) = 10	V (Y, c) = 4	V (Z, c) = 2
V (W, b) = 5	V (X, c) = 30	V (Y, d) = 5	V (Z, d) = 1

We want to perform the natural join for all four relations. What is the join-order selected by the greedy algorithm? What is its cost (measured by the sum of the sizes of the intermediate relations)? Provide the results of all intermediate steps.

6. (6 Points)

We have two relations R (A, B) and S (B, C) with the following statistics:

- T (R) = 120000, B(R) = 6000, V(R, A) =60000, V(R, B) = 10000
- T (S) = 90000, B(S) = 30000, V (S, B) = 90000, V(S, C) = 10000

Assume that the relations are clustered, i.e., are stored contiguously. The attribute B is a key in S and a foreign key in R. For the relation S we have an index on the attribute B. The index has three levels. Ignore the final output I/O cost.

a) We want to perform a natural join of R and S by using an index join algorithm. We assume that the index is completely in main memory. What is the required number of I/O's?

b) We want to perform a natural join of R and S by using an index join algorithm. We assume that the root of index is in the main memory but the blocks from the other levels are not. What is the required number of I/O's?

Implement of DBMS 2023/2024 Question

Question 1. **Mark = 4**

Consider a disk with the following characteristics:
There are 50 surfaces with 20,000 tracks each
Tracks hold on average 1000 sectors of 512 bytes each
The disk rotates at 3,000 rpm
The time it takes the head to move n tracks is 1 + 0.001n ms
30% of each track is used for gaps
A block has 32 sectors
a) What is the capacity of the disk?
b) What is the transfer time of a block?

Question 2. Mark 4

Suppose we use a RAID level 4 scheme with four data disks and one redundant disk. Assume that the blocks are a single byte. Give the block of the redundant disk if the corresponding blocks of the data disks are:
01010110, 11000000, 00111011, and 11111011
Recover the block while the redundant disk holds 11001011.
What changes to the
corresponding blocks on the other disks must be made?

Question 3. **Mark 6**

Prof. given tree and we need to insert value and delete value:

Insert the keys 20, 40, 10, 30, 15, 35, 7, 26, 18, 22, 36, and 42 in this order into an initially empty B+-tree of order 2.

Delete from the B+-tree of order 2 you produced in task 1 the keys 7 in this order.

Question 4. **Mark 6**

Let R and S be relations, p a predicate with only R attributes, q a predicate with only S attributes and m a predicate with attributes from R and S. Show that the following rule holds. Use in your proof only the given rules in the box. Indicate in each step which rule you are using. **mpq(R A S) = m([p(R)] A [q(S)])**

Question 5. **Mark 6**

Below are some statistics for four relations W, X, Y and Z.
W(a, b) X(b, c) Y(c, d) Z(d, e)
T(W) = 100 T(X) = 200 T(Y) = 300 T(Z) = 400
V(W, a) = 20 V(X, b) = 50 V(Y, c) = 50 V(Z, d) = 40
V(W, b) = 60 V(X, c) = 100 V(Y, d) = 50 V(Z, e) = 100

a) Consider a query optimizer that uses statistical data. In particular, the following information is known about an attribute A of relation R. Attribute A is of type integer. Consider the query A=7(R). How many tuples are expected in the answer, assuming values are uniformly distributed over possible V(R, A) values?
b) Consider the query A=17(R). How many tuples are expected in the answer, assuming values are uniformly distributed over possible domain values?

Question 6.**Mark 8**

Suppose that blocks can hold either ten records or 99 keys and 100 pointers. Also assume that each B+-tree node is 60% full, i.e., it will have 59 keys and 60 pointers in case of interior nodes and 60 keys and 60 pointers to records in case of leaf nodes. We have a data file that is a sequential file, and the B+- tree is a sparse index, but each primary block of the data file has one overflow block. The primary blocks are full, and the overflow blocks are half full. However, records are in no particular order within primary block and its overflow block.

a) Calculate the total number of blocks needed for a 3,240,000-record file and the index.

b) Calculate the average number of disk I/O's needed to retrieve a record given its search key. You may assume that nothing is in memory initially, and that the search key is the primary key for the records.

Question 7.**Mark 7**

Consider a clustered relation R(A, B, C, D) that has a clustering index on A and a
non-clustering
index on each of the other attributes. The relevant parameters are: B(R) = 1000, T(R) =
5000, V(R, A)
= 20, V(R, B) = 1000, V(R, C) = 5000, and V(R, D) = 500. Give the best query plan for the
following
selection and the corresponding number of disk I/O's. You can ignore the cost for accessing the index