1 Stunde, insgesamt 50 Punkte. Freitextaufgaben gaben ca 4,5 Punkte, die meisten anderen 2 Punkte. Aufgaben hatten zwischen TeilnehmerInnen teils andere Zahlen. Grundsätzlich gab es eher Zeitdruck.

Disks & Storage

1) Definition of temporal locality True or False?
2) Capacity calculation

Consider a hard disk with the following characteristics:

- 4 platters with 8 surfaces total
- 32768 $= 2^{15}$ tracks per surface
- 256 $= 2^{8}$ sectors per track (this is a simplified model where all tracks have the same number of sectors)
- 2048 $= 2^{11}$ bytes per sector
- 8192 $= 2^{13}$ bytes per block
- 20% of the tracks are used for gaps
- 5400 rotations per minute
- 1 ms to start/stop the write head (1 for both operations together)
- 1 ms per 6000 cylinders that have to be passed by the read/write head

What is the overall of capacity of the disk in GiB (1 GiB = 1024^3 Bytes)?

(Please answer just with a number without units. Round the final result to the nearest integer.)

Zahlen aus Klausur:

- 4 platters with 8 surfaces
- 16384 = 2^14 tracks per surface
- 256 = 2^8 sectors per track
- 1024 = 2^10 bytes per sector

Lösung:

32 GB

3) Elevator algorithm (**nur Endergebnis abgefragt** [completion of last request])

- Average transfer time for a block: 0,33
- Average rotational latency is 4,17
- It takes 0,5ms to start and 0,5ms to stop the read/write head
- The read/write head can pass 4000 cylinders in 1ms

Initially, the read/write head is at cylinder 32000

Request time	Cylinder
0	16000
1	8000
2	12000
20	64000
30	48000
40	56000

(hier hatten wir unterschiedliche Cylinder Zahlen und Ergebnisse)

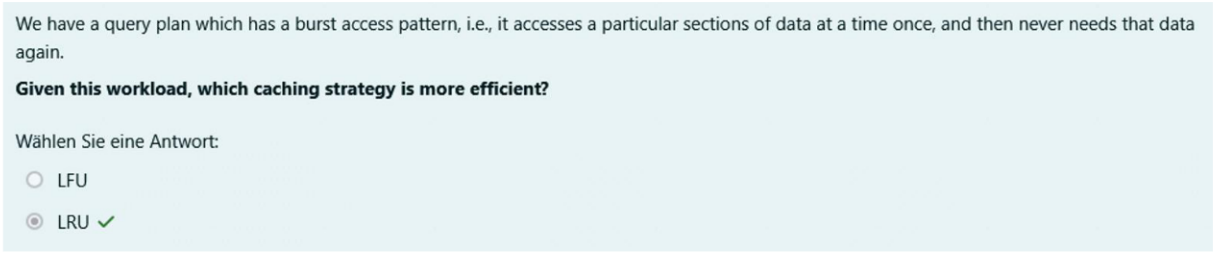
4) Gegeben: Relation R(employee_id INT, payment_id INT, amount INT, date, DATE)
- Int = 4 bytes, date = 4 bytes
- Block size = 32 bytes
- When reading 5 blocks, how many records can be retrieved by the query SELECT ID, a WHERE date = today

Anzahl bei n-Ary:

Anzahl bei Decomposition:

Lösung: n-ary: 10, Decomp: 13

Disks & Storage

5) 

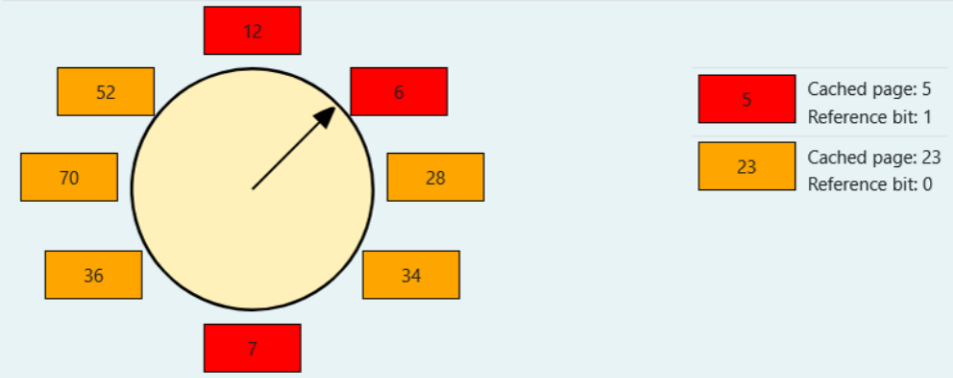
6) Cache hit / misses

Gegeben war ein gefüllter Cache und Anfragen, angegeben werden sollte für jede Anfrage, ob diese ein Hit oder Miss ist

7) CLOCK

Genau so, nur andere sequence:

Consider the CLOCK (a.k.a. Second Chance) algorithm. The system cache has space for 8 pages, and the current content of the cache is shown in the figure below.



Assume that when a page is accessed and it is already in the cache, its reference bit is set to 1 (if it is 0, it is changed to 1; if it is 1, then it is left as 1). The clock hand does not move in such a case (i.e., it stays where it was before this access). The clock hand only moves when the accessed page is not in the cache and the cache is looking for a page to evict by moving clockwise. When a new page is brought into the cache its reference bit is set to 1.

Also assume that in the first figure; red box indicates that the reference bit is set to 1, and orange box indicates that reference bit is set to 0.

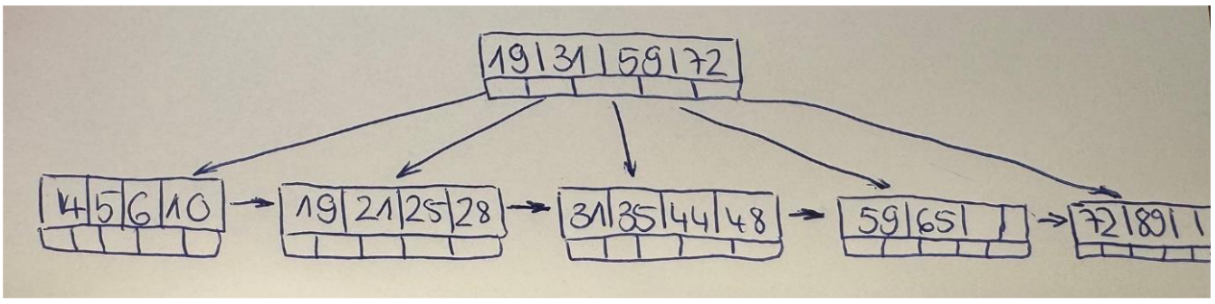
Consider the following sequence of page requests to the cache:

> 81, 15, 36, 6, 42, 3, 69, 55

Fill in the pages of the clock with regards to the page numbers after the last request has been completed (after all the pages in the series are accessed).

Multidimensional Indexing

8) Multiple Choice, ein B+ tree ist gegeben, drei Zahlen sollen eingefügt werden, welcher ist der resultierende B+ tree? Es wurde eine neue Root-Node gebildet und zwei Leaf nodes gesplittet (Achtung: Inner nodes mussten auch richtige Zahlen enthalten)



Einzufügen: 24, 58, 71

9) Second-Level Sparse Index

Data file:

- The data file contains 15 million tuples.
- Each tuple requires 4096 bytes on disk

Disk properties:

- The size of a block disk is 8192 bytes

Index properties:

- The size of a search key is 12 bytes
- The size of a pointer to a block is 12 bytes
- Index entries do not span blocks on disk

How many blocks are required for storing the second-level sparse index on the data file, if the first-level index is also sparse?

Lösung:

65 blocks for the second level index

10) Freitext: Indexing technique

Gegeben wurden Definition von zwei Tabellen, die erste enthielt Longtitude, Latitude, foreign_key Timezone o.ä. Zweite Tabelle enthielt timezone und offset o.ä.

Gegeben war, dass die meisten Queries Point Queries auf Latitute und Longitute sind, welches Indexing sollte gewählt werden? Warum?

Query Optimization

11) Cardinality of Selection

T(R) = 420000

V(R, a) = 30

V(R, b) = 20

V(R, c) = 10

V(R, d) = 15

!Nichts weiter gegeben! Keine Formeln!

Frage: What is the Cardinality of $\sigma_{b\ !=37 \vee d<80}$

12) Cardinality

Andere Zahlen, sonst Gleich dieser Aufgabe:

Consider the natural join $R(b,c,d) \bowtie S(b,c,e) \bowtie U(a,b,c)$, and the following important statistics:

R(b, c, d)	S(b, c, e)	U(a, b, c)
T(R) = 12000	T(S) = 42000	T(U) = 18000
		$V(U,a) = 100$
$V(R,b) = 300$	$V(S,b) = 150$	$V(U,b) = 40$
$V(R,c) = 50$	$V(S,c) = 100$	$V(U,c) = 150$
$V(R,d) = 100$		
	$V(S,e) = 70$	

What is the estimated cardinality of the natural join, i.e., the join on all common attributes?

13) Freitext DBMS technique

What is a DBMS technique to gain information on the data and cardinality of intermediate results (e.g. after join etc). What are problems?

Lösung (unsicher, da uns die Fragestellung auch nicht mehr ganz einfällt):

Heuristics provide good indicators, but real-world data is often scewed or follows other than normal distribution. -> Therefore not accurate results

Concurrency

14) ACID definition

In ACID, durability descibes that data from a query needs to be persistently written to disk before that query commits.

Lösung: Falsch, erst danach wird geschrieben

15) Precedence graph

Gegeben war eine Abfolge von Operationen von drei Transaktionen auf drei pages.

Es sollte angeben werden, welche Edges im resultierenden Precedence Graph enthalten sind.

16) Is the following schedule 2PL compliant? (drei Transaktionen, die concurrent Sachen lesen und schreiben)
17) Is the following schedule 2PL compliant? (nochmal)

Recovery

18) Definition of Analysis-Phase

A definition of the ANALYSIS Phase was given, asked if True or False

19) Undo / Redo

Ähnlich folgender Aufgabe, nur gab es nur einen Checkpoint (kein Begin / End), es mussten für die Einträge REDO und UNDO angegeben werden sowie die Einträge der active transaction table und dirty page table

Consider the following log records. The table contains the log sequence number (LSN), the transaction ID (TID), the previous log sequence number of the transaction (prevLSN), and the action of the log entry. The format of the UPDATE action is: (Page ID, Element ID, Old value, New value).

LSN	TID	prevLSN	ACTION	Redo Pass
1	1		BEGIN_TRANSACTION	
2	2		BEGIN_TRANSACTION	
3	3		BEGIN_TRANSACTION	
4	1	1	UPDATE (1, 1, "a", "A")	Ignore ✓
5	2	2	UPDATE (2, 2, "b", "B")	Ignore ✓
6	2	5	UPDATE (1, 3, "c", "C")	Ignore ✓
7			BEGIN_CHECKPOINT	
8	2	6	UPDATE (1, 4, "d", "D")	Redo ✓
9	2	8	UPDATE (3, 5, "e", "E")	Ignore ✗
10	2	9	COMMIT	
11	2	10	END_TRANSACTION	
12			END_CHECKPOINT	
13	3	3	UPDATE (2, 6, "f", "F")	
14	3	13	UPDATE (3, 7, "g", "G")	

The recovery process starts at LSN 7 with the following transaction table and dirty pages table.

Transaction table:

TID	lastLSN
1	4
2	6
3	3

Dirty page table:

PID	recLSN
1	8

Parallel and Distributed

20) Freitext: Gustafson vs. Amdahl

Why is Gustavson's Law more optimistic than Amdahl's law? + Write the assumptions and the influence on the perspective on scalability