Hasso Plattner Institut

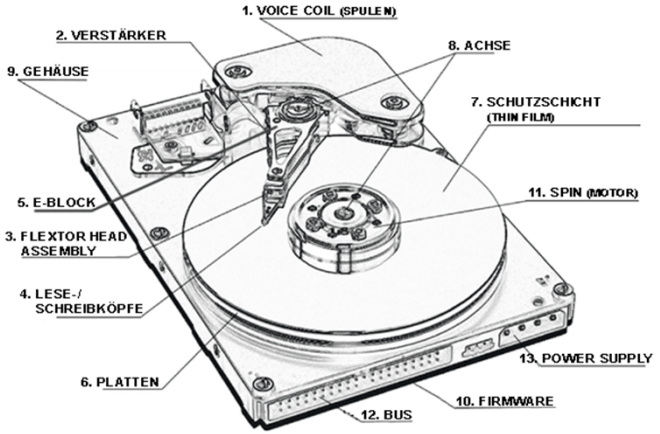
IT Systems Engineering | Universität Potsdam

Übung Datenbanksysteme II

Physische Speicherstrukturen

Leon Bornemann

Folien basierend auf
Thorsten Papenbrock,
Maximilian Jenders



Übung Datenbanksysteme II

- Übung
  - Leon Bornemann (Leon.Bornemann@hpi.de)

- Tutoren
  - Lukas Budach
  - Jonas Zimmermann

- Mailingliste
  - <Mailadresse wird ergänzt>
  - Hilfe zur Selbsthilfe: Fragen gegenseitig beantworten
  - Tutoren und Mitarbeiter lesen mit
  - Keine Email erhalten? -> Melden

Übung Datenbanksysteme II

- Abgabesystem
  - Abgabe *immer* in Zweiergruppen
  - Abgabe und Korrektur nur noch digital
  - Abgaben anderer Kurse bitte ignorieren
  - URL: http://www.dcl.hpi.uni-potsdam.de/submit/
  - Alle dort einmal für den Kurs DBS II eintragen
  - Anschließend Lösungen dort hochladen
- Besorgt euch Schlüssel :)
- Übungsthemen ≠ Hausaufgabenthemen

Übung Datenbanksysteme II

- Allgemeine Hinweise
  - Lösungen immer begründen!
    - Seid Präzise (Niemand liest gerne lange Texte :))
    - Im Zweifel 1-2 Sätze mehr
    - Wenn ihr Annahmen trefft, schreibt diese klar hin und argumentiert knapp warum diese sinnvoll sind
  - Bei Rechenaufgaben
    - In der Klausur habt ihr keine Rechenhilfe → Kopfrechnen
      - (dementsprechend werden halbwegs vernünftige Zahlen herauskommen)
    - In den Übungsaufgaben ist das nicht immer der Fall
  - Rechentipps:
    - Nutzt Potenzen und kürzt Brüche: 512 / 64 lässt sich in Exponentialform viel einfacher ausrechen
    - „Komische Zahlen" ruhig als Variablen benutzen

Übung Datenbanksysteme II

- Konkretes zu Übungen
  - Postgresql Installation wird benötigt
  - Es wird eine c/c++ Programmieraufgabe geben
    - Gerne jetzt schon Kenntnisse auffrischen

- Abwesenheit im Januar
  - Geburt meines Sohnes :)
  - Vertretung: Tobias Bleifuß (tobias.bleifuss@hpi.de)

Übung Datenbanksysteme II

- Fragen?
  - Zum Organisatorischen?
  - Zum Übungszettel?
- Gebt kontinuierliches Feedback

Erinnerung:

Dezimal- vs. Binär-Einheiten

Dezimalpräfixe			Binärpräfixe	
Name (Symbol)	Bedeutung[G 1]	Unterschied (gerundet)	IEC-Name (IEC-Symbol)	Bedeutung
Kilobyte (kB)[G 2]	10^3 Byte = 1.000 Byte	2,40 %	Kibibyte (KiB)[G 3]	2^{10} Byte = 1.024 Byte
Megabyte (MB)	10^6 Byte = 1.000.000 Byte	4,86 %	Mebibyte (MiB)	2^{20} Byte = 1.048.576 Byte
Gigabyte (GB)	10^9 Byte = 1.000.000.000 Byte	7,37 %	Gibibyte (GiB)	2^{30} Byte = 1.073.741.824 Byte
Terabyte (TB)	10^{12} Byte = 1.000.000.000.000 Byte	9,95 %	Tebibyte (TiB)	2^{40} Byte = 1.099.511.627.776 Byte
Petabyte (PB)	10^{15} Byte = 1.000.000.000.000.000 Byte	12,6 %	Pebibyte (PiB)	2^{50} Byte = 1.125.899.906.842.624 Byte
Exabyte (EB)	10^{18} Byte = 1.000.000.000.000.000.000 Byte	15,3 %	Exbibyte (EiB)	2^{60} Byte = 1.152.921.504.606.846.976 Byte
Zettabyte (ZB)	10^{21} Byte = 1.000.000.000.000.000.000.000 Byte	18,1 %	Zebibyte (ZiB)	2^{70} Byte = 1.180.591.620.717.411.303.424 Byte
Yottabyte (YB)	10^{24} Byte = 1.000.000.000.000.000.000.000.000 Byte	20,9 %	Yobibyte (YiB)	2^{80} Byte = 1.208.925.819.614.629.174.706.176 Byte

1. ↑ SI-Präfixe sind nur für SI-Einheiten standardisiert; Byte ist keine SI-Einheit
2. ↑ wird gelegentlich mit „KB“ abgekürzt
3. ↑ wird gelegentlich mit „KB“ abgekürzt, um den Unterschied zu „kB“ zu kennzeichnen (nicht standardisiert)

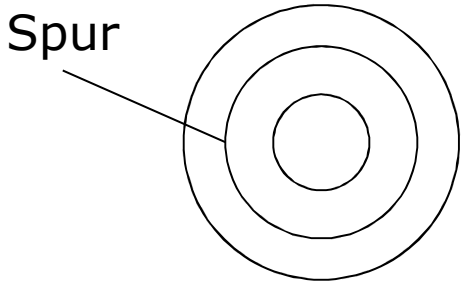
Für uns alles das gleiche!

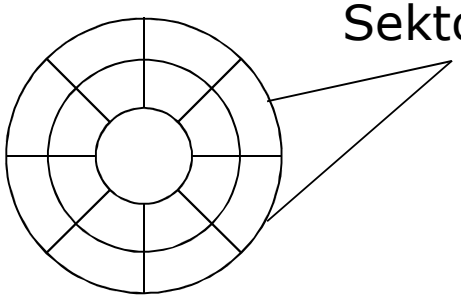
Nur Zweierpotenzen!

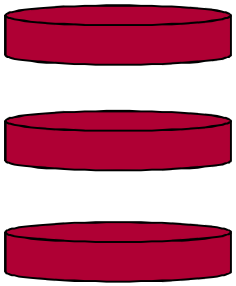
Begriffe

- Sektor: Abschnitt einer Spur (fester Länge)
- Spur: Ein Kreis auf einer Oberfläche
- Platte und Oberfläche: selbsterklärend
- Zylinder: alle Spuren eines bestimmten Radius
  - Über die gesamte Höhe
- *Block – besteht aus 1 oder mehr Sektoren (use-case abhängig)*

Auswendig Lernen! :)





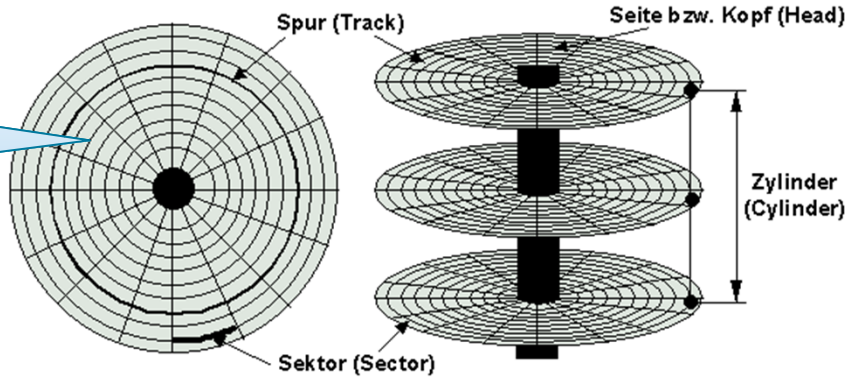


Physische Speicherstrukturen

- Sektorgröße: 512 Byte
- Sektoren pro Spur: Ø 64
- Spuren pro Oberfläche: 2048
- Oberflächen pro Platte: 2
- Anzahl Platten: 5
- Rotationsgeschwindigkeit: 5000 U/min
- Lesekopfbewegung über n Spuren: (1 + 0,002 · n) ms
- Blockgröße: 1024 Byte
- Lücken pro Spur: 10 %

0,5 Start
0,002 · n Bewegung
0,5 Stop

Aufteilung der Sektoren
auf den Spuren muss
nicht tortenförmig sein!
→ evtl. außen mehr Sektoren



Aufgabe 1:

Speicherkapazität berechnen

- Sektorgröße: 512 Byte
- Sektoren pro Spur: Ø 64
- Spuren pro Oberfläche: 2048
- Oberflächen pro Platte: 2
- Anzahl Platten: 5

a) Gesucht: Kapazität einer Spur

K_{Spur} = Sektorgröße · Sektoren pro Spur = $2^9 B \cdot 2^6 = 2^{15} B$

b) Gesucht: Kapazität einer Oberfläche

$K_{Oberfläche} = K_{Spur} \cdot$ Spuren pro Oberfläche = $2^{15} B \cdot 2^{11} = 2^{26}$

c) Gesucht: Kapazität der Festplatte

$K_{Platte} = K_{Oberfläche} \cdot$ Anzahl Oberflächen = $2^{26} \cdot 2 \cdot 5 = 10 \cdot 2^{26}$

Aufgabe 1:

Speicherkapazität berechnen

- Sektorgröße: 512 Byte
- Sektoren pro Spur: Ø 64
- Spuren pro Oberfläche: 2048
- Oberflächen pro Platte: 2
- Anzahl Platten: 5

d) Gesucht: Anzahl Zylinder auf dieser Festplatte

2048 (= Anzahl Spuren)

e) Gesucht: Sind die folgenden Blockgrößen zulässig?

- 256 Byte — Nein: Blöcke können nicht kleiner als 1 Sektor sein
- 1.024 Byte — Ja: Block umfasst genau 2 Sektoren
- 51.200 Byte — Nein: Pro Spur gibt es nur $2^{15}B$ = 32.768 Byte

Aufgabe 2:

Latenzzeiten Begriffe

- Kommunikationszeit:
  - Zeit, die der Lesekopf braucht um die Informationen über den Bus zu schicken (vernachlässigbar klein)
- Suchzeit (Seek-Time)
  - Zeit, die der Lesekopf braucht um die richtige Spur zu erreichen
- Rotationszeit (rotational latency)
  - Zeit, die der Lesekopf braucht bis der richtige Teil der Spur unter ihn rotiert ist.
- Transferzeit
  - Zeit, die der Lesekopf braucht um alle Daten zu lesen
- Latenzzeit (latency)
  - Summe aus allen

Auswendig Lernen! :)

Aufgabe 2:

Latenzzeiten berechnen

- Sektorgröße: 512 Byte
- Sektoren pro Spur: Ø 64
- Spuren pro Oberfläche: 2048
- Rotationsgeschwindigkeit: 5000 U/min
- Lesekopfbewegung über n Spuren: $(1 + 0{,}002 \cdot n)$ ms
- Blockgröße: 1024 Byte
- Lücken pro Spur: 10 %

a) Gesucht: Minimale, maximale und durchschnittliche Latenzzeit zum Lesen eines Blocks

Latenzzeit = Kommunikation + Seek + Rotation + Transfer

vernachlässigbar

für alle drei Fälle gleich

Aufgabe 2:

Latenzzeiten berechnen

- Sektorgröße: 512 Byte
- Sektoren pro Spur: Ø 64
- Spuren pro Oberfläche: 2048
- Rotationsgeschwindigkeit: 5000 U/min
- Lesekopfbewegung über n Spuren: (1 + 0,002 · n) ms
- Blockgröße: 1024 Byte
- Lücken pro Spur: 10 %

a) Gesucht: Minimale, maximale und durchschnittliche Latenzzeit zum Lesen eines Blocks

Latenzzeit = Kommunikation + Seek + Rotation + Transfer

Rotationszeit = 1 U / Rotationsgeschw. = 1 / 5000 min = 12 ms

T_{Sektor} = 12ms * 0.9 / 64 = 0,168ms
$T_{Lücke}$ = 12ms * 0.1 / 64 = 0,018ms

Blocklänge = 2 Sektoren + 1 Lücke (2. Lücke ist ja egal)

$L_{Transfer}$ = 0.168ms * 2 + 0,018ms = 0.35475ms

Aufgabe 2:
Latenzzeiten berechnen

- Sektorgröße: 512 Byte
- Sektoren pro Spur: Ø 64
- Spuren pro Oberfläche: 2048
- Rotationsgeschwindigkeit: 5000 U/min
- Lesekopfbewegung über n Spuren: $(1 + 0{,}002 \cdot n)$ ms
- Blockgröße: 1024 Byte
- Lücken pro Spur: 10 %

a) Gesucht: Minimale, maximale und durchschnittliche Latenzzeit zum Lesen eines Blocks

Latenzzeit = Kommunikation + Seek + Rotation + Transfer

$L_{Transfer\_min}$ = 0,35625 ms (= $L_{Transfer}$)
L_{Seek_min} = 0 ms (Lesekopf auf passender Spur)
$L_{Rotation\_min}$ = 0 ms (Lesekopf genau vor dem Block)

L_{min} = 0,35625 ms + 0 ms + 0 ms = 0,35625 ms

Aufgabe 2:
Latenzzeiten berechnen

- Sektorgröße: 512 Byte
- Sektoren pro Spur: Ø 64
- Spuren pro Oberfläche: 2048
- Rotationsgeschwindigkeit: 5000 U/min
- Lesekopfbewegung über n Spuren: $(1 + 0{,}002 \cdot n)$ ms
- Blockgröße: 1024 Byte
- Lücken pro Spur: 10 %

a) Gesucht: Minimale, maximale und durchschnittliche Latenzzeit zum Lesen eines Blocks

Latenzzeit = Kommunikation + Seek + Rotation + Transfer

$L_{Transfer\_max} = 0{,}35625$ ms (= $L_{Transfer}$)
$L_{Seek\_max} = (1 + 0{,}002 \cdot 2048)$ ms $= 5{,}096$ ms (über alle Spuren)
$L_{Rotation\_max} = 12$ ms (volle Rotation)

$L_{max} = 0{,}35625$ ms $+ 5{,}096$ ms $+ 12$ ms $= 17{,}45225$ ms

Aufgabe 2:
Latenzzeiten berechnen

- Sektorgröße: 512 Byte
- Sektoren pro Spur: Ø 64
- Spuren pro Oberfläche: 2048
- Rotationsgeschwindigkeit: 5000 U/min
- Lesekopfbewegung über n Spuren: (1 + 0,002 · n) ms
- Blockgröße: 1024 Byte
- Lücken pro Spur: 10 %

"1/3 der Spuren" ist eine gute An-näherung

≠ L_{Seek_max} / 2

a) Gesucht: Minimale, maximale und durchschnittliche Latenzzeit zum [illegible]n eines Blocks

Latenzzeit = Kommunikation + Se[illegible]k + Rotatio[illegible] + Transfer

$L_{Transfer\_avg}$ = 0,35625 ms (= $L_{Transfer}$)
L_{Seek_avg} = (1 + 0,002 · 2048 / 3) ms ≈ 2,355 ms (avg. Distanz)
$L_{Rotation\_avg}$ = 12 ms / 2 = 6 ms (halbe Rotation)

L_{avg} = 0,35625 ms + 2,355 ms + 6 ms = 8,71125 ms

Aufgabe 3:

Daten speichern

- Sektorgröße: 512 Byte
- Sektoren pro Spur: Ø 64
- Spuren pro Oberfläche: 2048
- Oberflächen pro Platte: 2
- Anzahl Platten: 5
- Blockgröße: 1024 Byte

- Datei mit 100.000 Tupeln der Größe 100 Byte
- Es gibt keine Tupel, die auf mehrere Blöcke aufgeteilt sind

a) Gesucht: Anzahl Tupel pro Block

$$\begin{aligned}\text{Tupel pro Block} &= \lfloor(\text{Blockgröße} / \text{Tupelgröße})\rfloor \\ &= \lfloor(1024\ \text{Byte} / 100\ \text{Byte})\rfloor \\ &= 10\end{aligned}$$

Aufgabe 3:

Daten speichern

- Sektorgröße: 512 Byte
- Sektoren pro Spur: Ø 64
- Spuren pro Oberfläche: 2048
- Oberflächen pro Platte: 2
- Anzahl Platten: 5
- Blockgröße: 1024 Byte

- Datei mit 100.000 Tupeln der Größe 100 Byte
- Es gibt keine Tupel, die auf mehrere Sektoren aufgeteilt sind

b) Gesucht: Anzahl Blöcke für vollständige Datei

Blöcke für Datei = ⌈(Anzahl Tupel / Tupel pro Block)⌉
= ⌈(100.000 / 10)⌉
= 10.000

Aufgabe 3:

Daten speichern

- Sektorgröße: 512 Byte
- Sektoren pro Spur: Ø 64
- Spuren pro Oberfläche: 2048
- Oberflächen pro Platte: 2
- Anzahl Platten: 5
- Blockgröße: 1024 Byte

- Datei mit 100.000 Tupeln der Größe 100 Byte
- Es gibt keine Tupel, die auf mehrere Sektoren aufgeteilt sind

c) Gesucht: Maximale Anzahl Tupel auf der gesamten Festplatte

Aufgabe 3:

Daten speichern

- Sektorgröße: 512 Byte
- Sektoren pro Spur: Ø 64
- Spuren pro Oberfläche: 2048
- Oberflächen pro Platte: 2
- Anzahl Platten: 5
- Blockgröße: 1024 Byte

- Datei mit 100.000 Tupeln der Größe 100 Byte
- Es gibt keine Tupel, die auf mehrere Sektoren aufgeteilt sind

c) Gesucht: Maximale Anzahl Tupel auf der gesamten Festplatte

Blöcke pro Spur = Sektoren pro Spur / Sektoren pro Block

$$= 2^6B \;/\; (2^{10}B \;/\; 2^9B\;) = 2^5$$

Tupel pro Festplatte = Tupel pro Block · Blöcke pro Spur · Spuren pro Oberfläche · Oberflächen pro Platte · Anzahl Platten

$$= 10 \cdot 2^5 \cdot 2^{11} \cdot 2 \cdot 5 = 100 \cdot 2^{16}$$

Aufgabe 4:

Daten lesen

- Anzahl Zylinder: 8192
- Seekzeit für n Zylinder: (1 + 0,002 · n) ms
- Rotationszeit: Ø 6,5 ms
- Transferzeit: Ø 0,5 ms
- Initiale Lesekopfposition: 4000

- Anfragen:

Eintreffen der Anfrage	0 ms	3 ms	11 ms	19 ms
Angefragter Zylinder	6500	2000	8000	3500

a) Gesucht: Bearbeitung der Anfragen mit **First-Come, First-Served**

b) Gesucht: Bearbeitung der Anfragen mit dem **Elevator Algorithmus**

Daten Lesen
Wiederholung: Elevator Algorithm

- Lesekopf streicht über die Oberfläche
  - Hält an Zylinder an wenn dort IO-Anfrage benötigt wird
  - Dreht um falls in der momentanen Richtung keine IO-Anfrage mehr zu bearbeiten ist

Aufgabe 4:
Daten lesen

Eintreffen der Anfrage	0 ms	3 ms	11 ms	19 ms
Angefragter Zylinder	6500	2000	8000	3500

- Anzahl Zylinder: 8192
- Seekzeit für n Zylinder: (1 + 0,002 · n) ms
- Rotationszeit: Ø 6,5 ms
- Transferzeit: Ø 0,5 ms
- Initiale Lesekopfposition: 4000

Anfangsbeispiel

Start Time	Lesekopf-Position	Request Queue	Duration	Direction
0ms	4000	6500		

Aufgabe 4:
Daten lesen

Eintreffen der Anfrage	0 ms	3 ms	11 ms	19 ms
Angefragter Zylinder	6500	2000	8000	3500

- Anzahl Zylinder: 8192
- Seekzeit für n Zylinder: (1 + 0,002 · n) ms
- Rotationszeit: Ø 6,5 ms
- Transferzeit: Ø 0,5 ms
- Initiale Lesekopfposition: 4000

Anfangsbeispiel

Start Time	Lesekopf-Position	Request Queue	Duration	Direction
0ms	4000	6500	(1+0,002·2500)+6,5+0,5 =13ms	→
13ms				

Aufgabe 4:
Daten lesen

Eintreffen der Anfrage	0 ms	3 ms	11 ms	19 ms
Angefragter Zylinder	6500	2000	8000	3500

- Anzahl Zylinder: 8192
- Seekzeit für n Zylinder: (1 + 0,002 · n) ms
- Rotationszeit: Ø 6,5 ms
- Transferzeit: Ø 0,5 ms
- Initiale Lesekopfposition: 4000

First-Come, First-Served

Start Time	Lesekopf-Position	Request Queue	Duration	Direction
0ms	4000	6500	(1+0,002·2500)+6,5+0,5 =13ms	→
13ms	6500	2000 8000	(1+0,002·4500)+6,5+0,5 =17ms	←
30ms	2000	8000 3500	(1+0,002·6000)+6,5+0,5 =20ms	→
50ms	8000	3500	(1+0,002·4500)+6,5+0,5 =17ms	←
67ms				

Aufgabe 4:

Daten lesen

Eintreffen der Anfrage	0 ms	3 ms	11 ms	19 ms
Angefragter Zylinder	6500	2000	8000	3500

- Anzahl Zylinder: 8192
- Seekzeit für n Zylinder: (1 + 0,002 · n) ms
- Rotationszeit: Ø 6,5 ms
- Transferzeit: Ø 0,5 ms
- Initiale Lesekopfposition: 4000

Elevator Algorithmus

Start Time	Lesekopf -Position	Request Queue	Duration	Direction
0ms	4000	6500	(1+0,002·2500)+6,5+0,5 =13ms	→
13ms	6500	8000 2000	(1+0,002·1500)+6,5+0,5 =11ms	→
24ms	8000	3500 2000	(1+0,002·4500)+6,5+0,5 =17ms	←
41ms	3500	2000	(1+0,002·1500)+6,5+0,5 =11ms	←
52ms				

Aufgabe 5:

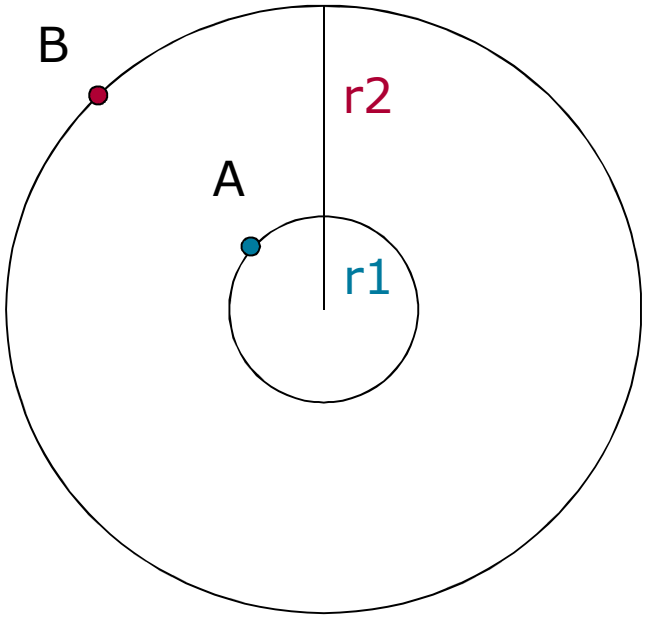
Daten auf verschiedenen Spuren

- Rotations- und Lesekopfgeschwindigkeit sind konstant, d.h.
  - Rotationslatenz unverändert
  - sequentielle Datentransferrate ist auf äußeren Spuren größer, falls dort mehr Sektoren pro Spur angelegt wurden
  - Was ist mit der Seektime?
    - Geschwindigkeit des Lesekopfes ist gleich…
    - … aber die durchschnittliche Anzahl an zu überquerenden Spuren ist in der Mitte am geringsten

Aufgabe 5:

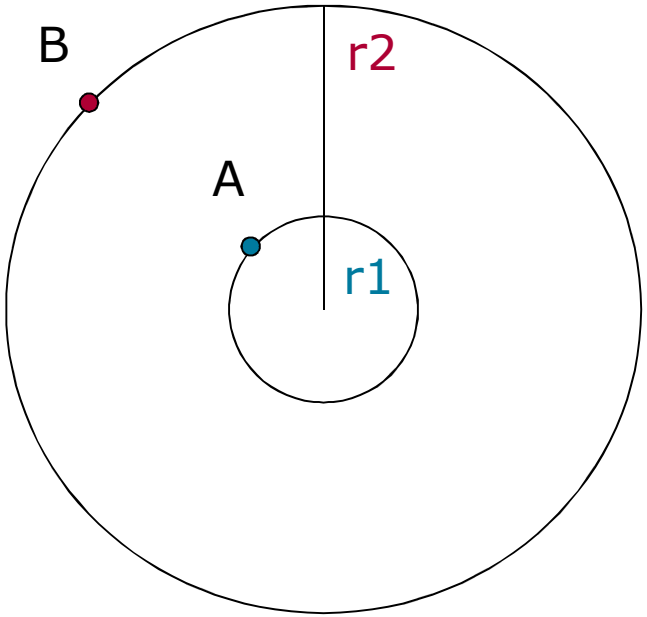
Daten auf verschiedenen Spuren

- Warum ist Lesen/Schreiben auf der äußeren Spur schneller?



Aufgabe 5:

Daten auf verschiedenen Spuren



- Beide Punkte legen eine Umdrehung in der gleichen Zeit zurück (Gleiche Winkelgeschwindigkeit)
- Zurückgelegte Strecke = Umfang Kreis = $2 \cdot \pi \cdot r$
- r1 < r2 → B bewegt sich mit größerer Geschwindigkeit als A (Bahnengeschwindigkeit)
- → Sektoren die auf Spur B liegen werden schneller gelesen

Aufgabe 5:

Daten auf verschiedenen Spuren

- Wo sollte man daher die folgenden Dateien für die genannten Zugriffe positionieren (innen, mitte, außen)?
  - seltene, sequentielle Scans …
    - einer großen Datei
    - einer kleinen Datei
  - häufiger, random Zugriff auf …
    - eine kleine Datei
    - eine große Datei per Index

Aufgabe 5:
Daten auf verschiedenen Spuren

- Wo sollte man daher die folgenden Dateien für die genannten Zugriffe positionieren (innen, mitte, außen)?
  - seltene, sequentielle Scans …
    - ◇ einer großen Datei: **außen**
      - Kosten dominiert durch sequentiellen Datentransfer
      - Sequentieller Datentransfer ist außen am schnellsten
    - ◇ einer kleinen Datei: **innen**
      - Kosten dominiert durch initialen Seek und Rotation (Lesen einer kleinen Datei ist effektiv Random I/O)
      - Innen wird nichts optimiert, aber das Lagern kleiner, selten zugegriffener Dateien tut hier am wenigsten weh

Aufgabe 5:

Daten auf verschiedenen Spuren

- Wo sollte man daher die folgenden Dateien für die genannten Zugriffe positionieren (innen, mitte, außen)?
  - häufiger, random Zugriff auf …
    - eine kleine Datei: **mitte**
      - Kosten dominiert durch Seek und Rotation
      - Seek wird in der Mitte minimiert (Wegen häufigem Zugriff ist die Optimierung der Seektime hier am wichtigsten)

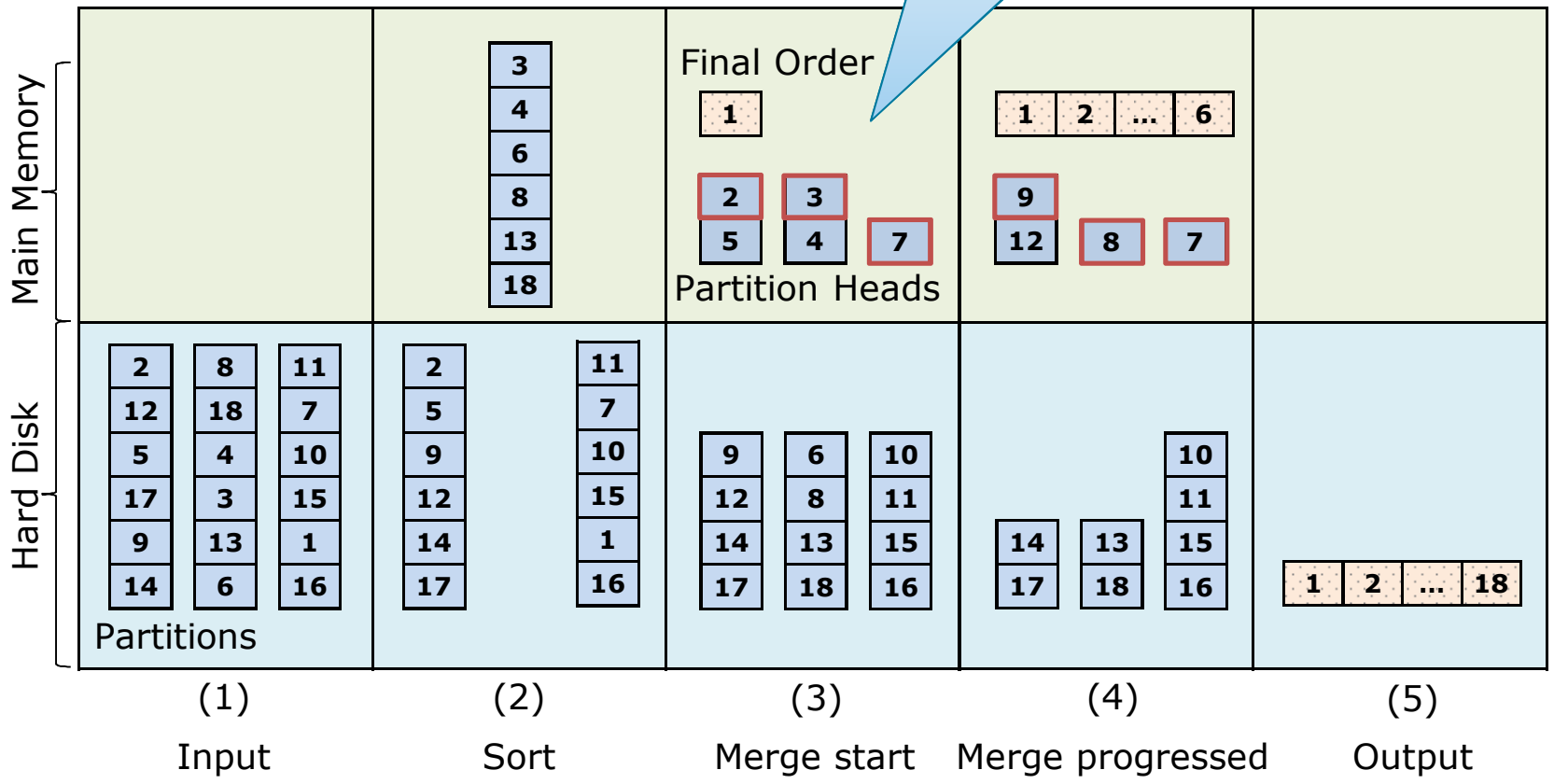
Annahme: Index Steht bei Datei

    - eine große Datei per Index: **innen**
      - Kosten dominiert durch häufigen Seek zwischen Datei und Index
      - Seek wird durch nahes Zusammenlegen minimiert
      - Beim Platzieren innen sparen wir den jeweils wertvollen äußeren und mittleren Platz

Aufgabe 6:

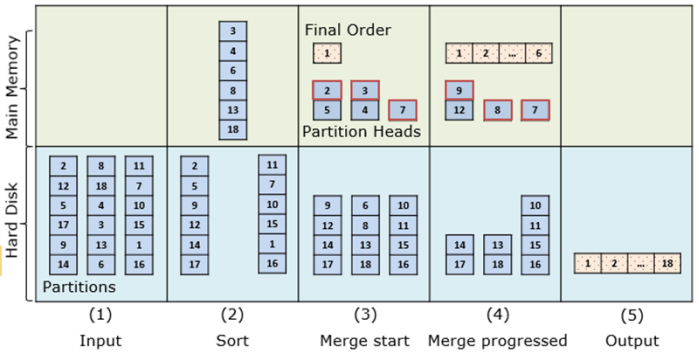
Daten sortieren: TPMMS

Algorithmus:



Aufgabe 6:

Daten sortieren: TPMMS



Gedankenspiel: Three-Phase-Multiway-Merge-Sort

1. Warum könnte er notwendig sein?

   Es passen nicht alle Köpfe vorsortierter Teillisten in den RAM.

2. Wie könnte ein Three-Phase-Multiway-Merge-Sort funktionieren?

   P1: Sort; P2: (Pre-)Merge jeweils n Teillisten; P3: Merge alle Listen

   Um wie viel steigen die Lese- und Schreibkosten?

   1 x alle Tupel lesen + 1 x alle Tupel schreiben

3. Die Partitionen müssen nicht unbedingt Hauptspeicher-groß sein.

   Welche Vorteile könnte eine Partitionsgröße $\frac{Hauptspeichergröße}{2}$ haben?

   Sortieren im RAM müsste nicht in-place sein und ist so O(n*log(n))