



Übung Datenbanksysteme II
Anfrageoptimierung II

Tobias Bleifuß

Teilweise basierend auf Folien von Leon Bornemann, Maximilian Jenders und Thorsten Papenbrock

Agenda

1. Nachbesprechung Hausaufgabe 4
2. Kardinalitätsabschätzung für Joins
 - Annahmen
 - Abschätzung
 - Histogramme
 - Mehrere Join-Attribute
 - Mehrere Join-Relationen
3. Join-Reihenfolge
 - Dynamische Programmierung
4. Aufgabenblatt 5

**Übung DBS II –
Anfrage-
optimierung II**

07.01.19 / 09.01.19
Tobias Bleifuß

Nachbesprechung Hausaufgabe 4

- Aufgabe 1.a

a) Betrachten Sie folgende Anfrage:

```
EXPLAIN SELECT title FROM movies;
```

Warum wird hier ein sequential Scan ausgeführt obwohl ein Index zur Verfügung steht?
Erweitern Sie die Anfrage, sodass ein Index-Scan verwendet wird. **2 P**

**Übung DBS II –
Anfrage-
optimierung II**

07.01.19 / 09.01.19
Tobias Bleifuß

Nachbesprechung Hausaufgabe 4

- Aufgabe 1.b

Betrachten Sie folgende Anfrage:

```
EXPLAIN SELECT * FROM movies,directors
      WHERE movies.title = directors.title
      AND directors.pseudonym IS NOT null;
```

Nennen Sie zwei verschiedene Möglichkeiten die Anfrage zu verändern, sodass ein anderer Join-Algorithmus verwendet wird (lediglich die beiden Tabellen, sowie die JOIN-Bedingung muss gleich bleiben, die Semantik der neuen Anfrage darf eine völlig andere sein).

**Übung DBS II –
Anfrage-
optimierung II**

07.01.19 / 09.01.19
Tobias Bleifuß

Nachbesprechung Hausaufgabe 4

- Aufgabe 1.c

- c) Schreiben Sie eine sinnvolle Anfrage, bei welcher der Optimierer einen Merge-Join verwendet. **2 P**

**Übung DBS II –
Anfrage-
optimierung II**

07.01.19 / 09.01.19
Tobias Bleifuß

Nachbesprechung Hausaufgabe 4

- Aufgabe 2

```
EXPLAIN ANALYZE SELECT m.releaseyear,d.name,p.author
FROM movies m JOIN directors d JOIN editors e JOIN plots p
ON (e.title = p.title) ON (d.title = e.title) ON (m.title = e.title)
WHERE d.name = e.name AND p.author = 'Anonymous';
```

a) An welcher Stelle hat der Optimierer sich deutlich verschätzt?

**Übung DBS II –
Anfrage-
optimierung II**

07.01.19 / 09.01.19
Tobias Bleifuß

Nachbesprechung Hausaufgabe 4

■ Aufgabe 2

`enable_bitmapscan` (boolean)

Enables or disables the query planner's use of bitmap-scan plan types. The default is on.

`enable_hashagg` (boolean)

Enables or disables the query planner's use of hashed aggregation plan types. The default is on.

`enable_hashjoin` (boolean)

Enables or disables the query planner's use of hash-join plan types. The default is on.

`enable_indexscan` (boolean)

Enables or disables the query planner's use of index-scan plan types. The default is on.

`enable_indexonlyscan` (boolean)

Enables or disables the query planner's use of index-only-scan plan types. The default is on.

`enable_material` (boolean)

<https://www.postgresql.org/docs/9.5/runtime-config-query.html>

- c) Entwerfen Sie nun analog zu b) einen QEP, der die längstmögliche Zeit braucht. Erklären Sie kurz warum dieser Plan ungünstig ist. Sollte ihr QEP länger als die 100-fache Zeit des O-QEPs brauchen dürfen Sie die Zeitmessung abbrechen. **4 P**

**Übung DBS II –
Anfrage-
optimierung II**

07.01.19 / 09.01.19
Tobias Bleifuß

Kardinalitätsschätzung: Annahmen

- Natural Join $R(X,Y) \bowtie S(Y,Z)$
 - Equijoin analog
- **Annahme 1:**
Containment of Value Sets
 - Beispiel: Liste für Y: [A,B,C,D,E]
- **Annahme 2:**
Preservation of Value Sets
 - $V(R \bowtie S, Z) = V(S, Z)$
 - Z ist nicht Joinattribut

R		S	
X	Y	Y	Z
-	A	A	-
-	B	B	-
-	C	C	-
-	D	D	-
-		E	-
		A	-
		B	-
		C	-
		D	-
		E	-

Übung DBS II – Anfrage- optimierung II

07.01.19 / 09.01.19
Tobias Bleifuß

Kardinalitätsschätzung: Schätzungen für Natural Join

Kostenschätzung für $R(X,Y) \bowtie S(Y,Z)$:

■
$$T(R \bowtie S) = \frac{T(R) \cdot T(S)}{\max [V(R,Y), V(S,Y)]}$$
 Warum Maximum?

Was ist mit $V(R \bowtie S, Y)$?

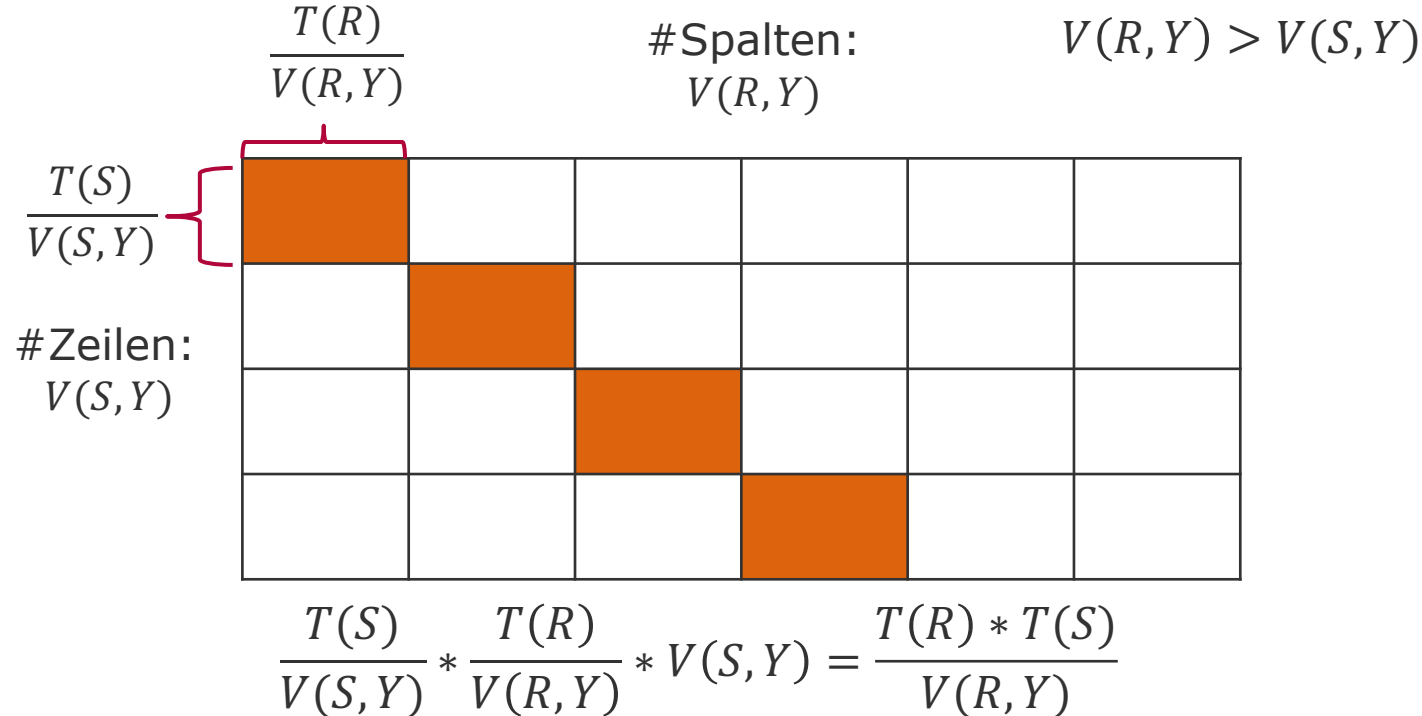
■
$$V(R \bowtie S, Y) = \min [V(R,Y), V(S,Y)]$$
 Wieso Minimum?

**Übung DBS II –
Anfrage-
optimierung II**

07.01.19 / 09.01.19
Tobias Bleifuß

Kardinalitätsschätzung: Visualisierung

Natural Join $R(X,Y) \bowtie S(Y,Z)$



**Übung DBS II –
Anfrage-
optimierung II**

07.01.19 / 09.01.19
Tobias Bleifuß

Kardinalitätsschätzung: Histogramme

- Histogram-Typen:
 - **Equi-width**
 - Equi-height
 - Most frequent values

```
SELECT Januar.Tag, Juli.Tag
FROM Januar, Juli
WHERE Januar.temp = Juli.temp
```

Wertebereich Temperatur °F	Januar	Juli
0-9	40	0
10-19	60	0
20-29	80	0
30-39	50	0
40-49	10	5
50-59	5	20
60-69	0	50
70-79	0	100
80-89	0	60
90-99	0	10

Übung DBS II – Anfrage- optimierung II

07.01.19 / 09.01.19
Tobias Bleifuß

Chart **11**

Kardinalitätsschätzung: Histogramme

- Histogramm-Typen:
 - Equi-width
 - **Equi-height**
 - Most frequent values
- Beispiel equi-height:
 - R: Buckets [0-24],[25-74],[75-99]
jeweils 100 Tupel pro Bucket
 - S: Buckets [0-19],[20-99]
jeweils 200 Tupel pro Bucket

	[0-19]	[20-24]	[25-74]	[75-99]	Σ
R	80	20	100	100	300
S	200	12,5	125	62,5	400
Kard.	$80 \cdot 200 / 20$	$20 \cdot 12,5 / 5$	$100 \cdot 125 / 50$	$100 \cdot 62,5 / 25$	1350

Übung DBS II – Anfrage- optimierung II

07.01.19 / 09.01.19
Tobias Bleifuß

Chart **12**

Kardinalitätsschätzung: Mehrere Joinattribute

Verallgemeinerung für mehrere Joinattribute:

- Teile durch das Produkt aller Maxima von $V(R, y)$ und $V(S, y)$ für jedes Joinattribut y

R(a,b,c)	S(b,c,d)
T(R)=1000	T(S)=2000
V(R,b)=30	V(S,b)=50
V(R,c)=100	V(S,c)=40

Geschätzte Kardinalität für
 $R(a,b,c) \bowtie S(b,c,d)$:

$$\frac{1000 * 2000}{50 * 100} = 400$$

**Übung DBS II –
Anfrage-
optimierung II**

07.01.19 / 09.01.19
Tobias Bleifuß

Kardinalitätsschätzung: Mehrfach-Join über ein Attribut

X(a,b)	Y(a,b)	Z(a,b)
T(X)=100	T(Y)=200	T(Z)=300
V(X,a)=10	V(Y,a)=20	V(Z,a)=30
V(X,b)=3	V(Y,b)=2	V(Z,b)=1

$$T(X \bowtie Y) = 100 \cdot 200 / (20 \cdot 3) = 333,33$$

$$T((X \bowtie Y) \bowtie Z) = 333,33 \cdot 300 / (30 \cdot 2) = 1666,66$$

- *Preservation of Value Sets* gilt nicht, wenn das Attribut ein Join-Attribut war!
- Wegen *Containment of Value Sets* wissen wir aber, dass nach dem Join $V(X \bowtie Y, a) = 10$ und $V(X \bowtie Y, b) = 2$ sein müssen.
 - *Containment of Value Sets*: Werte eines Attributs, das in mehreren Relationen auftaucht, werden vom Beginn einer festen Liste gewählt.
 - Falls $V(R, Y) \leq V(S, Y) \Rightarrow$ Jeder Y-Wert in R taucht in S auf
 - Realistisch? Wann?
 - *Preservation of Value Sets*: Anzahl der distinct-Werte eines nicht-Joinattributs bleiben erhalten.
 - $V(R \bowtie S, X) = V(R, X)$

Übung DBS II – Anfrage- optimierung II

07.01.19 / 09.01.19
Tobias Bleifuß

Kardinalitätsschätzung: Mehrere Relationen

Verallgemeinerung für mehrere Relationen:

- Multipliziere die Anzahl der Tupel in jeder Relation; teile für jedes Attribut A, was mindestens zweimal auftaucht, durch alle außer dem kleinsten $V(R,A)$

$W(A, B)$	$X(B, C)$	$Y(C, D)$	$Z(D, A)$
$T(W) = 100$	$T(X) = 200$	$T(Y) = 300$	$T(Z) = 400$
$V(W, A) = 20$			$V(Z, A) = 100$
$V(W, B) = 60$	$V(X, B) = 50$		
	$V(X, C) = 100$	$V(Y, C) = 50$	
		$V(Y, D) = 50$	$V(Z, D) = 40$

- Geschätzte Kardinalität für $W \bowtie_B X \bowtie_C Y \bowtie_{D,A} Z$:

$$\frac{100 * 200 * 300 * 400}{100 * 60 * 100 * 50} = 80$$

**Übung DBS II –
Anfrage-
optimierung II**

07.01.19 / 09.01.19
Tobias Bleifuß

Join-Reihenfolge

Gegeben: Folgende Relationen und deren Statistiken

$W(A, B)$	$X(B, C)$	$Y(C, D)$	$Z(D, A)$
$T(W) = 100$	$T(X) = 200$	$T(Y) = 300$	$T(Z) = 400$
$V(W, A) = 20$			$V(Z, A) = 100$
$V(W, B) = 60$	$V(X, B) = 50$		
$V(X, C) = 100$		$V(Y, C) = 50$	
		$V(Y, D) = 50$	$V(Z, D) = 40$

Gesucht: Optimale Join-Reihenfolge für $W \bowtie X \bowtie Y \bowtie Z$

Bestimme die Join-Reihenfolge als Left Deep Tree. Nutze dazu Dynamische Programmierung und gib die Tabellen aller Zwischenschritte an. Verwende als Kostenmaß die Summe der Zwischenergebnisgrößen.

**Übung DBS II –
Anfrage-
optimierung II**

07.01.19 / 09.01.19
Tobias Bleifuß

Join-Reihenfolge

$W(A, B)$	$X(B, C)$	$Y(C, D)$	$Z(D, A)$
$T(W) = 100$	$T(X) = 200$	$T(Y) = 300$	$T(Z) = 400$
$V(W, A) = 20$			$V(Z, A) = 100$
$V(W, B) = 60$	$V(X, B) = 50$		
	$V(X, C) = 100$	$V(Y, C) = 50$	
		$V(Y, D) = 50$	$V(Z, D) = 40$

	{W}	{X}	{Y}	{Z}
Kardinalität	100	200	300	400
Kosten	0	0	0	0
Opt. Plan	W	X	Y	Z

Kleinste Relation links

	{W,X}	{W,Y}	{W,Z}	{X,Y}	{X,Z}	{Y,Z}
Kard.	$100 \cdot 200 / 60 = 333,33$	$100 \cdot 300 = 30000$	$100 \cdot 400 / 100 = 400$	$200 \cdot 300 / 100 = 600$	$200 \cdot 400 = 80000$	$300 \cdot 400 / 50 = 2400$
Kosten	0	0	0	0	0	0
Opt. Plan	$W \bowtie X$	$W \bowtie Y$	$W \bowtie Z$	$X \bowtie Y$	$X \bowtie Z$	$Y \bowtie Z$

Übung DBS II – Anfrage- optimierung II

07.01.19 / 09.01.19
Tobias Bleifuß

Join-Reihenfolge

$W(A,B)$	$X(B,C)$	$Y(C,D)$	$Z(D,A)$
$T(W) = 100$	$T(X) = 200$	$T(Y) = 300$	$T(Z) = 400$
$V(W,A) = 20$			$V(Z,A) = 100$
$V(W,B) = 60$	$V(X,B) = 50$	$V(Y,C) = 50$	
	$V(X,C) = 100$	$V(Y,D) = 50$	$V(Z,D) = 40$

	$\{W,X\}$	$\{W,Y\}$	$\{W,Z\}$	$\{X,Y\}$	$\{X,Z\}$	$\{Y,Z\}$
Kard.	$100 \cdot 200 / 60 = 333,33$	$100 \cdot 300 / 100 = 300$	$100 \cdot 400 / 100 = 400$	$200 \cdot 300 / 100 = 600$	$200 \cdot 400 / 50 = 800$	$300 \cdot 400 / 50 = 2400$
Kosten	0	0	0	0	0	0
Opt. Plan	$W \bowtie X$	$W \bowtie Y$	$W \bowtie Z$	$X \bowtie Y$	$X \bowtie Z$	$Y \bowtie Z$

Kardinalität ist für alle möglichen Joinreihenfolgen gleich!

	$\{W,X,Y\}$	$\{W,X,Z\}$	$\{W,Y,Z\}$	$\{X,Y,Z\}$
Kardinalität	$333,33 \cdot 300 / 100 = 1000$	$333,33 \cdot 400 / 100 = 1333,33$	$400 \cdot 300 / 50 = 2400$	$600 \cdot 400 / 50 = 4800$
Kosten	333,33	333,33	400	600
Opt. Plan	$(W \bowtie X) \bowtie Y$	$(W \bowtie X) \bowtie Z$	$(W \bowtie Z) \bowtie Y$	$(X \bowtie Y) \bowtie Z$

Kosten: (Kardinalität + Kosten) des Zwischenergebnisses

Join-Reihenfolge

$W(A, B)$	$X(B, C)$	$Y(C, D)$	$Z(D, A)$
$T(W) = 100$	$T(X) = 200$	$T(Y) = 300$	$T(Z) = 400$
$V(W, A) = 20$			$V(Z, A) = 100$
$V(W, B) = 60$	$V(X, B) = 50$		
	$V(X, C) = 100$	$V(Y, C) = 50$	
		$V(Y, D) = 50$	$V(Z, D) = 40$

	$\{W, X\}$	$\{W, Y\}$	$\{W, Z\}$	$\{X, Y\}$	$\{X, Z\}$	$\{Y, Z\}$
Kard.	$100 \cdot 200 / 60 = 333,33$	$100 \cdot 300 = 30000$	$100 \cdot 400 / 100 = 400$	$200 \cdot 300 / 100 = 600$	$200 \cdot 400 = 80000$	$300 \cdot 400 / 50 = 2400$
Kosten	0	0	0	0	0	0
Opt. Plan	$W \bowtie X$	$W \bowtie Y$	$W \bowtie Z$	$X \bowtie Y$	$X \bowtie Z$	$Y \bowtie Z$

	$\{W, X, Y\}$	$\{W, X, Z\}$	$\{W, Y, Z\}$	$\{X, Y, Z\}$
Kardinalität	$333,33 \cdot 300 / 100 = 1000$	$333,33 \cdot 400 / 100 = 1333,33$	$400 \cdot 300 / 50 = 2400$	$600 \cdot 400 / 50 = 4800$
Kosten	333,33	333,33	400	600
Opt. Plan	$(W \bowtie X) \bowtie Y$	$(W \bowtie X) \bowtie Z$	$(W \bowtie Z) \bowtie Y$	$(X \bowtie Y) \bowtie Z$

Übung DBS II – Anfrage- optimierung II

07.01.19 / 09.01.19
Tobias Bleifuß

Join-Reihenfolge

$W(A,B)$	$X(B,C)$	$Y(C,D)$	$Z(D,A)$
$T(W) = 100$	$T(X) = 200$	$T(Y) = 300$	$T(Z) = 400$
$V(W,A) = 20$			$V(Z,A) = 100$
$V(W,B) = 60$	$V(X,B) = 50$		
	$V(X,C) = 100$	$V(Y,C) = 50$	
		$V(Y,D) = 50$	$V(Z,D) = 40$

	$\{W,X,Y\}$	$\{W,X,Z\}$	$\{W,Y,Z\}$	$\{X,Y,Z\}$
Kardinalität	$333,33 \cdot 300 / 100 = 1000$	$333,33 \cdot 400 / 100 = 1333,33$	$400 \cdot 300 / 50 = 2400$	$600 \cdot 400 / 50 = 4800$
Kosten	333,33	333,33	400	600
Opt. Plan	$(W \bowtie X) \bowtie Y$	$(W \bowtie X) \bowtie Z$	$(W \bowtie Z) \bowtie Y$	$(X \bowtie Y) \bowtie Z$

Plan	$((W \bowtie X) \bowtie Y) \bowtie Z$	$((W \bowtie X) \bowtie Z) \bowtie Y$	$((W \bowtie Z) \bowtie Y) \bowtie X$	$((X \bowtie Y) \bowtie Z) \bowtie W$
Kard.	$1000 \cdot 400 / (100 \cdot 50) = 80$	$1333,33 \cdot 300 / (100 \cdot 50) = 80$	$2400 \cdot 200 / (100 \cdot 60) = 80$	$4800 \cdot 100 / (100 \cdot 60) = 80$
Kosten	1333,33	1666,66	2800	5400



Aufgabenblatt 5

Aufgabe 1: Kardinalitätsschätzung

Gegeben seien die folgenden Relationen und deren Statistiken:

$W(a, b)$	$X(b, c)$	$Y(c, d)$	$Z(d, a)$
$T(W) = 300$	$T(X) = 600$	$T(Y) = 900$	$T(Z) = 1200$
$V(W, a) = 30$	$V(W, b) = 60$	$V(X, b) = 50$	$V(Z, a) = 300$
	$V(X, c) = 100$	$V(Y, c) = 50$	
		$V(Y, d) = 60$	$V(Z, d) = 40$

Schätze die Kardinalität der Ergebnisrelationen der folgenden Ausdrücke:

9 P

- a) $W \bowtie X \bowtie Y \bowtie Z$
- b) $\sigma_{a=10}(W)$
- c) $\sigma_{c=20}(Y)$
- d) $\sigma_{c=20}(Y) \bowtie Z$
- e) $W \times Y$
- f) $\sigma_{d>10}(Z)$
- g) $\sigma_{a=1 \wedge d=2}(Z)$
- h) $\sigma_{c>1 \wedge d=2}(Y)$
- i) $X \bowtie_{X.b=Z.d} Z$

**Übung DBS II –
Anfrage-
optimierung II**

07.01.19 / 09.01.19
Tobias Bleifuß

Aufgabenblatt 5

Aufgabe 2: Join-Kardinalität

Gegeben sind zwei Relationen $R(A, B)$ und $S(B, C)$. Beide Relationen enthalten 20 unterschiedliche Werte im Attribut B , wobei die Werte in R den Werten in S entsprechen. Es gilt also $V(R, B) = V(S, B) = 20$. Die Werteverteilungen in $R(B)$ und $S(B)$ sind durch folgendes Histogramm beschrieben, welches die Häufigkeit der 4 häufigsten Werte angibt:

	0	1	2	3	4	andere Werte
$R.B$	5	6	4	5	*	32
$S.B$	10	8	5	*	7	48

Die mit * gekennzeichneten Werte gehören nicht zu den vier jeweils häufigsten Werten, sondern zu den "anderen Werten". Schätze nun unter Verwendung des Histogramms die Kardinalität des Joins über $R(A, B) \bowtie S(B, C)$ ab. Schätze anschließend die Kardinalität ohne das Histogramm zu verwenden (wie üblich Annahme der Gleichverteilung aller 20 Attributwerte). Vergleiche die beiden Ergebnisse. 5 P

Übung DBS II – Anfrage- optimierung II

07.01.19 / 09.01.19
Tobias Bleifuß

Aufgabe 3: Join-Reihenfolge

Gegeben seien die folgenden Relationen und deren Statistiken:

$E(a, b, c)$	$F(a, b, d)$	$G(a, c, d)$	$H(b, c, d)$
$T(E) = 1000$	$T(F) = 2000$	$T(G) = 3000$	$T(H) = 4000$
$V(E, a) = 1000$	$V(F, a) = 50$	$V(G, a) = 50$	
$V(E, b) = 50$	$V(F, b) = 100$		$V(H, b) = 40$
$V(E, c) = 20$		$V(G, c) = 300$	$V(H, c) = 100$
	$V(F, d) = 200$	$V(G, d) = 500$	$V(H, d) = 400$

Bestimme die Join-Reihenfolge als Left Deep Tree. Gib dazu alle Tabellen des Algorithmus der Dynamischen Programmierung an. Was ist die optimale Join-Reihenfolge, wie hoch sind deren Kosten und welche Kardinalität hat der Join am Ende?

Hinweis: Bei dieser Aufgabe können bei der Berechnung für Kardinalität und Kosten Bruchzahlen herauskommen. Auch wenn man in der Realität keine Kardinalität von z.B. 1/10 Tupeln erwartet könnt ihr in dieser Aufgabe damit ruhig weiterrechnen.

10 P

Übung DBS II –
Anfrage-
optimierung II

07.01.19 / 09.01.19
Tobias Bleifuß