

**Benutzerzentrierte
Datenbankanfragen**

von Stefanie Breske

TOP-K-RETRIEVAL IV

- ◎ Motivation
 - Das Beispiel

- ◎ 2 Ansätze
 - Grundidee
 - Funktionsweise

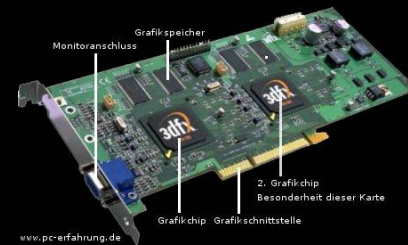
- ◎ Die Gegenüberstellung

Motivation

- ◎ Klaus und Bärbel beim Laptopkauf
 - Akkulaufzeit vs. Grafik
- ◎ zeitliche Optimierung durch Vorberechnung



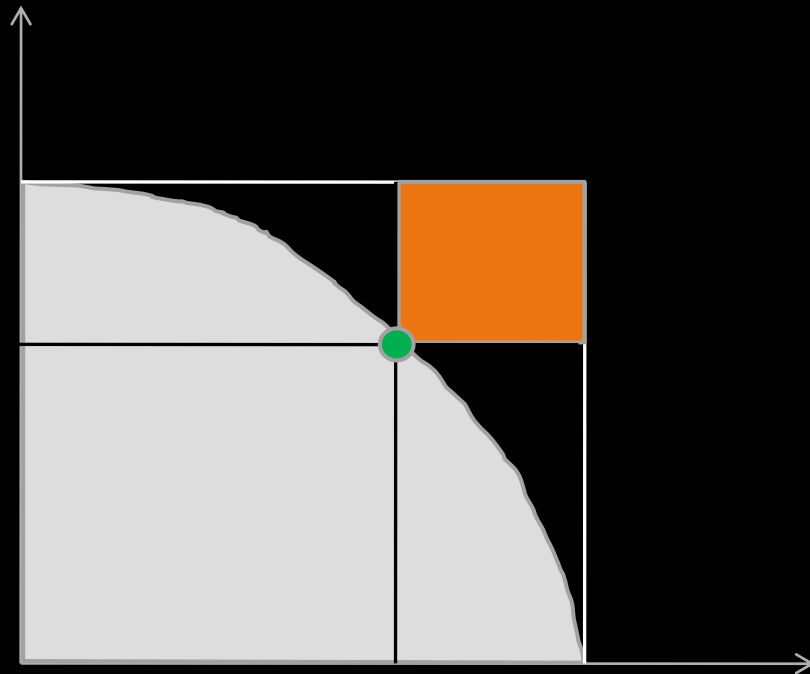
Datenbestand



ID	Hersteller	Grafik Benchmark	Akkulaufzeit in min
1	Dell	12.000	50
2	Samsung	7.500	100
3	Toshiba	6.500	250
4	Sony	8.000	120
5	Acer	12.500	115
6	Apple	14.000	145
7	Lenovo	10.000	170

Fagin - Korrelation

Grafik



Akkulaufzeit

Onion Technique



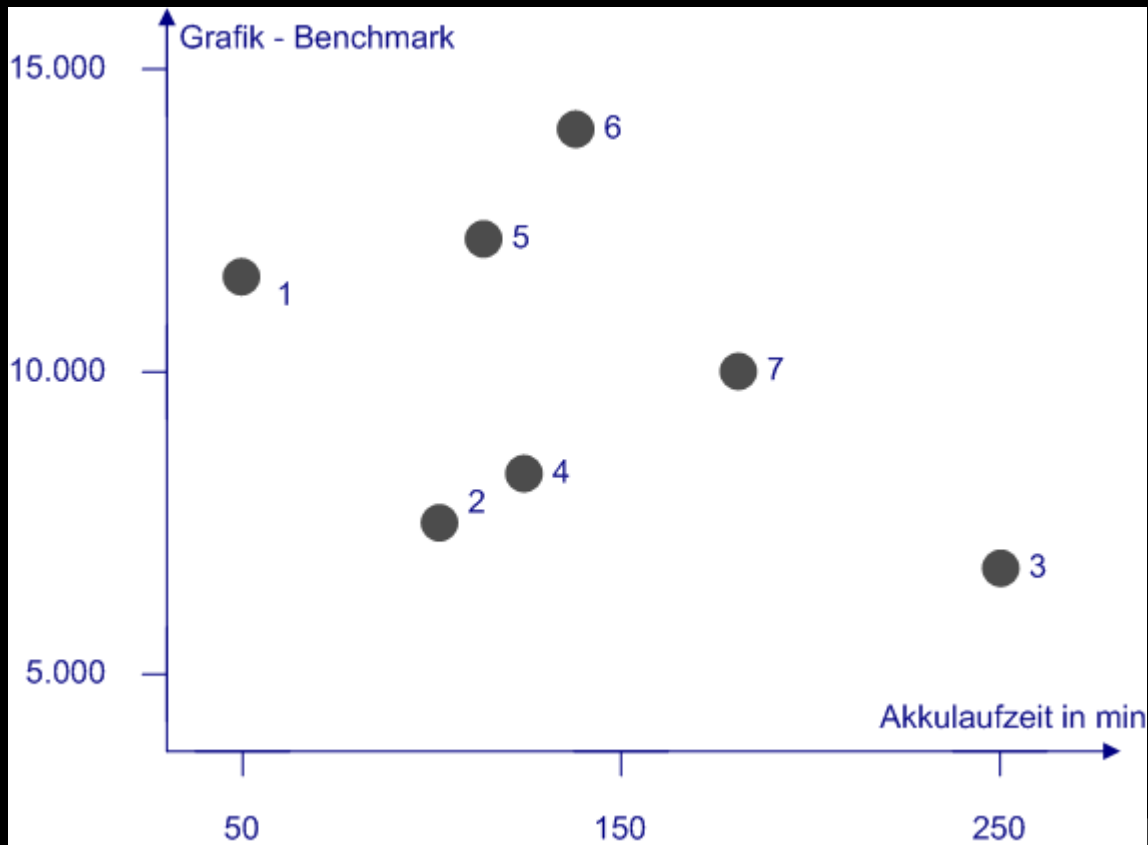
◎ Grundidee:

- Aufteilung der Daten in Schichten
- Geometrische Eigenschaft konvexer Hüllen
- Berücksichtigung von Korrelation

Onion - Funktionsweise

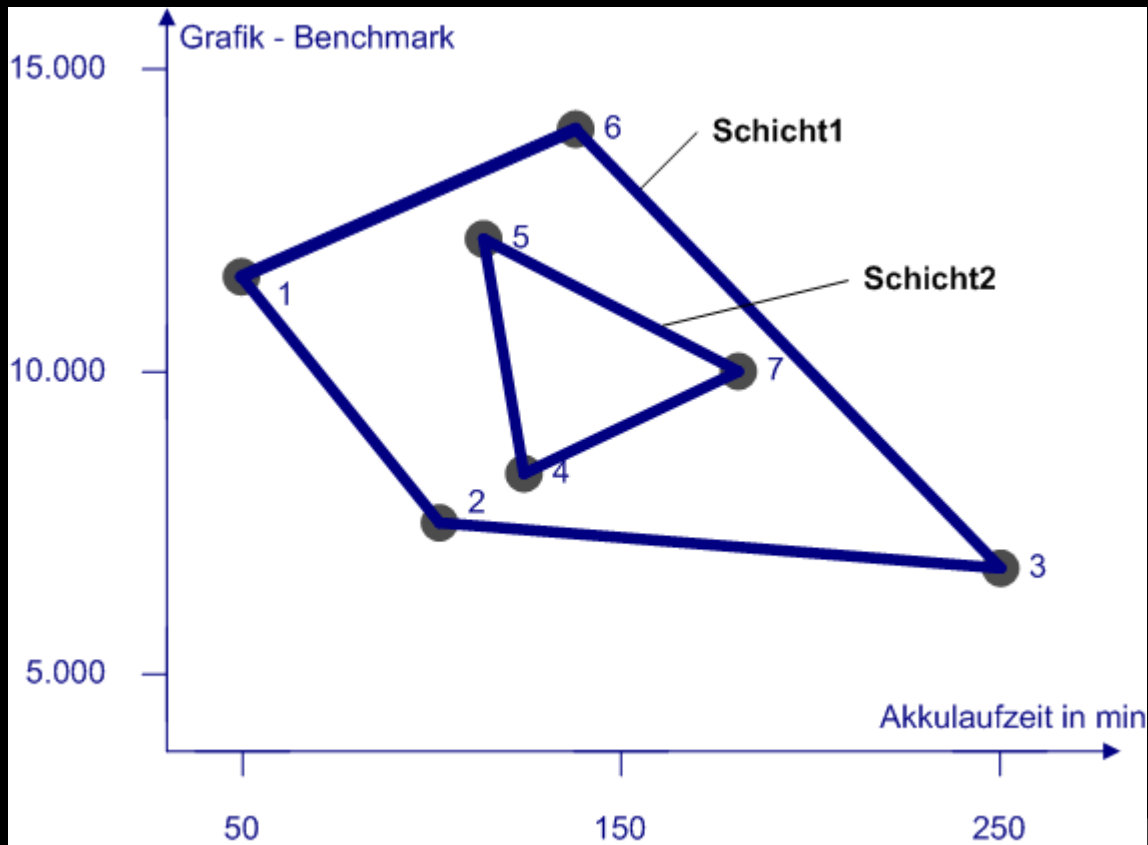


Vorverarbeitung: Indexerstellung



Onion - Funktionsweise

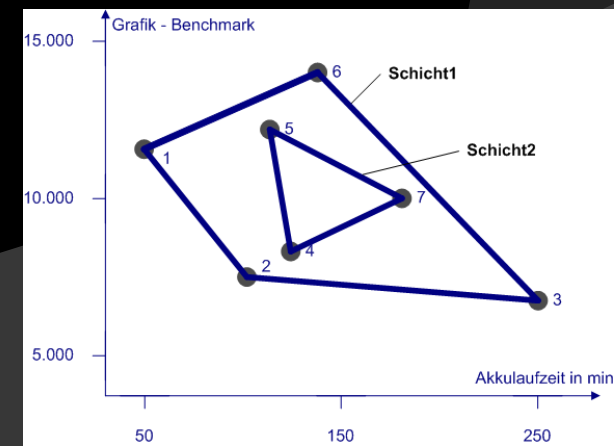
Vorverarbeitung: Indexerstellung



Onion - Funktionsweise

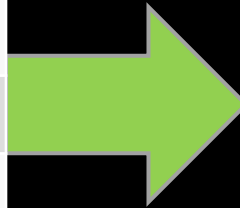
Schicht 1		
ID	Hersteller	Score
1	Dell	7,8
2	Samsung	9,0
3	Toshiba	17,6
6	Apple	14,3

Verteilung: 0,4 x Grafik + 0,6 x Akkulaufzeit



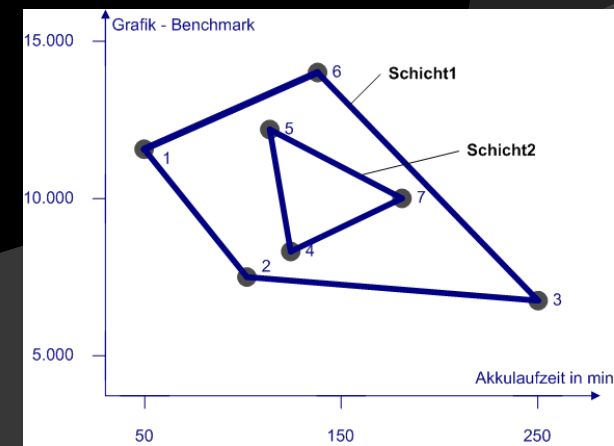
Onion - Funktionsweise

Schicht 1		
ID	Hersteller	Score
1	Dell	7,8
2	Samsung	9,0
3	Toshiba	17,6
6	Apple	14,3



Schicht 1		
ID	Hersteller	Score
3	Toshiba	17,6
6	Apple	14,3
2	Samsung	9,0
1	Dell	7,8

Verteilung: 0,4 x Grafik + 0,6 x Akkulaufzeit



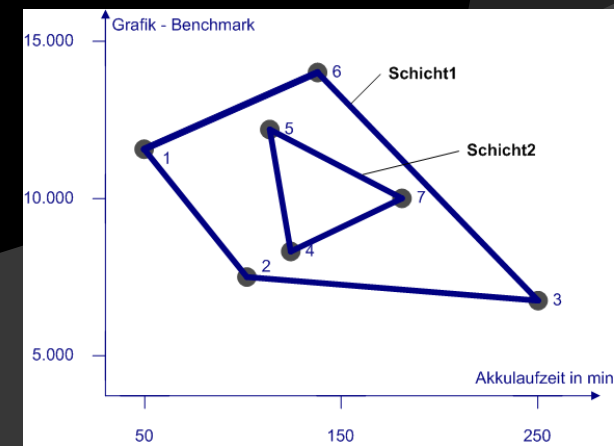
Onion - Funktionsweise

Schicht 1		
ID	Hersteller	Score
3	Toshiba	17,6
6	Apple	14,3
2	Samsung	9,0
1	Dell	7,8

K = 3

Ergebnis	
Hersteller	Score

Kandidaten	
Hersteller	Score



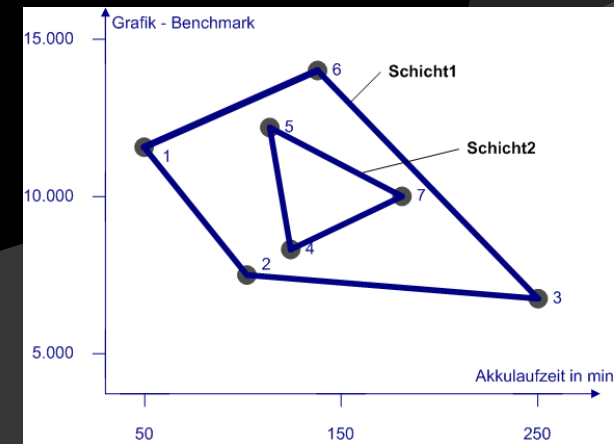
Onion - Funktionsweise

Schicht 1		
ID	Hersteller	Score
3	Toshiba	17,6
6	Apple	14,3
2	Samsung	9,0
1	Dell	7,8

Ergebnis	
Hersteller	Score
Toshiba	17,6

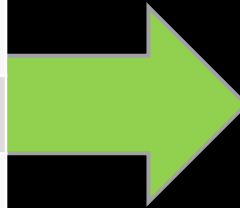
Kandidaten	
Hersteller	Score
Apple	14,3
Samsung	9,0

K = 3



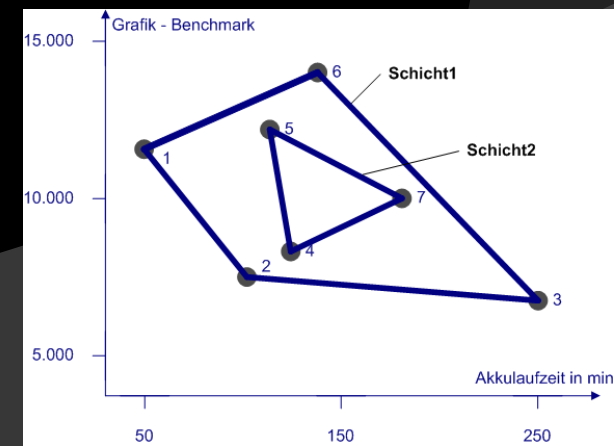
Onion - Funktionsweise

Schicht 2		
ID	Hersteller	Score
4	Sony	10,4
5	Acer	11,9
7	Lenovo	14,2



Schicht 2		
ID	Hersteller	Score
7	Lenovo	14,2
5	Acer	11,9
4	Sony	10,4

Verteilung: 0,4 x Grafik + 0,6 x Akkulaufzeit



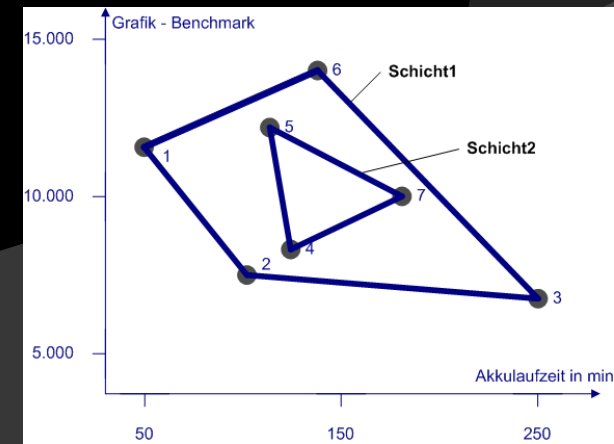
Onion - Funktionsweise

Schicht 2		
ID	Hersteller	Score
7	Lenovo	14,2
5	Acer	11,9
4	Sony	10,4

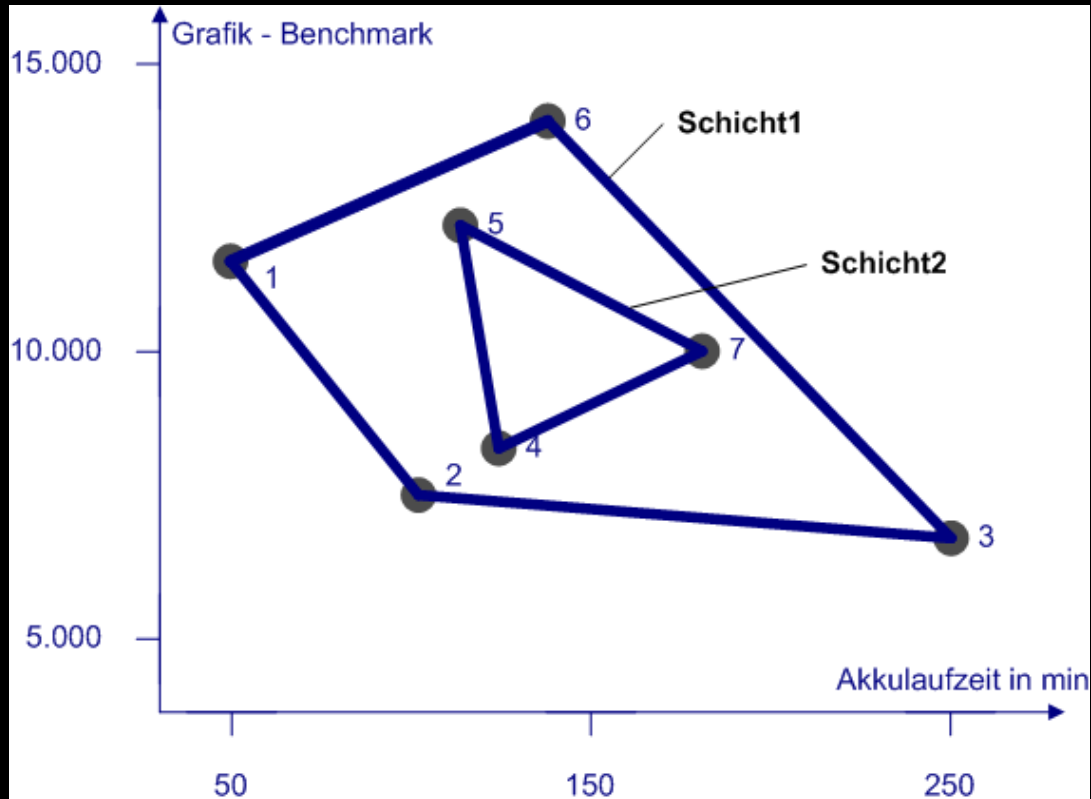
Ergebnis	
Hersteller	Score
Toshiba	17,6

Kandidaten	
Hersteller	Score
Apple	14,3
Samsung	9,0

K = 3



Onion



Ergebnis		
ID	Hersteller	Score
3	Toshiba	17,6
6	Apple	14,3
7	Lenovo	14,2

PREFER

◎ Grundidee:

- Gruppierung von Menschen
- Vorberechnung von Views

Nutzer-Profil: In einer Minute zum richtigen Notebook:

- ➔ [Das ideale Notebook für Einsteiger](#)
- ➔ [Allround-Nutzer: Ein Notebook für alle Gelegenheiten](#)
- ➔ [Multimedia- und Spiele-Notebook: Keine Kompromisse!](#)
- ➔ [Ein Notebook fürs Büro: Sparsam und pragmatisch](#)
- ➔ [Notebooks für spezielle Situationen und Berufsgruppen](#)
- ➔ [Klein, aber fein: Günstige Netbooks für unterwegs](#)

PREFER: Funktionsweise

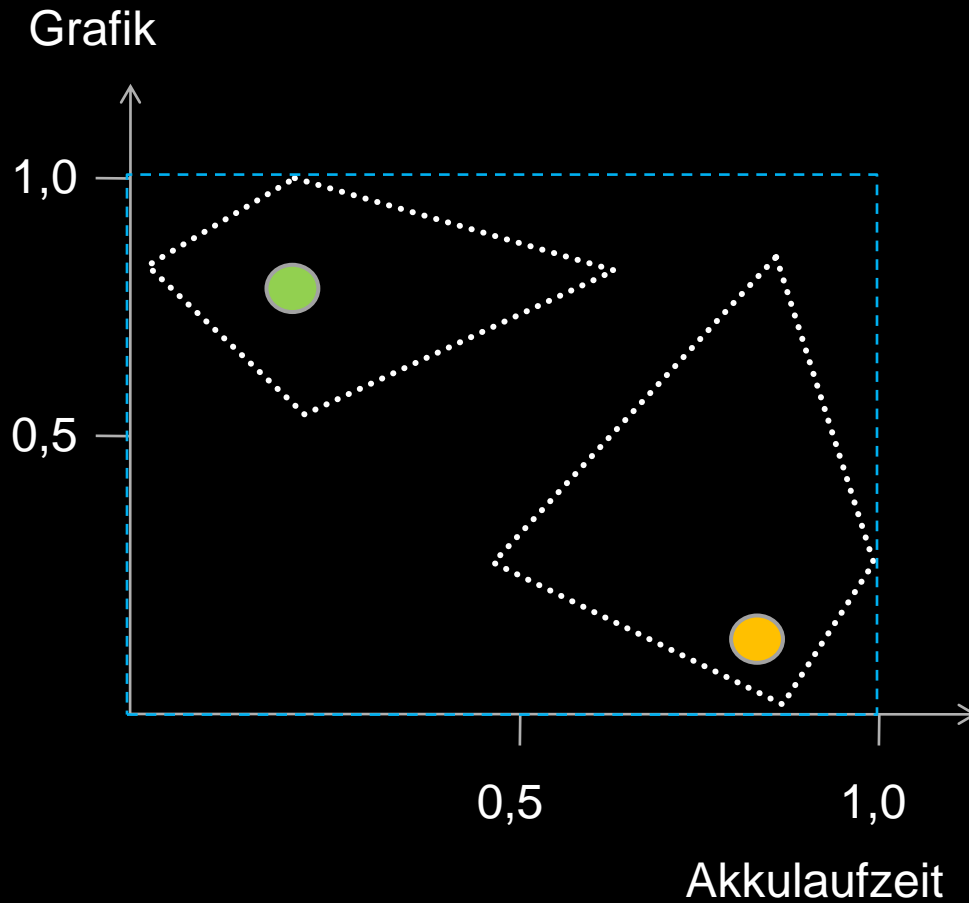
Vorverarbeitung

- Relation
- Präferenzfunktionen
- Platzbeschränkung

**View-
Erzeugung**



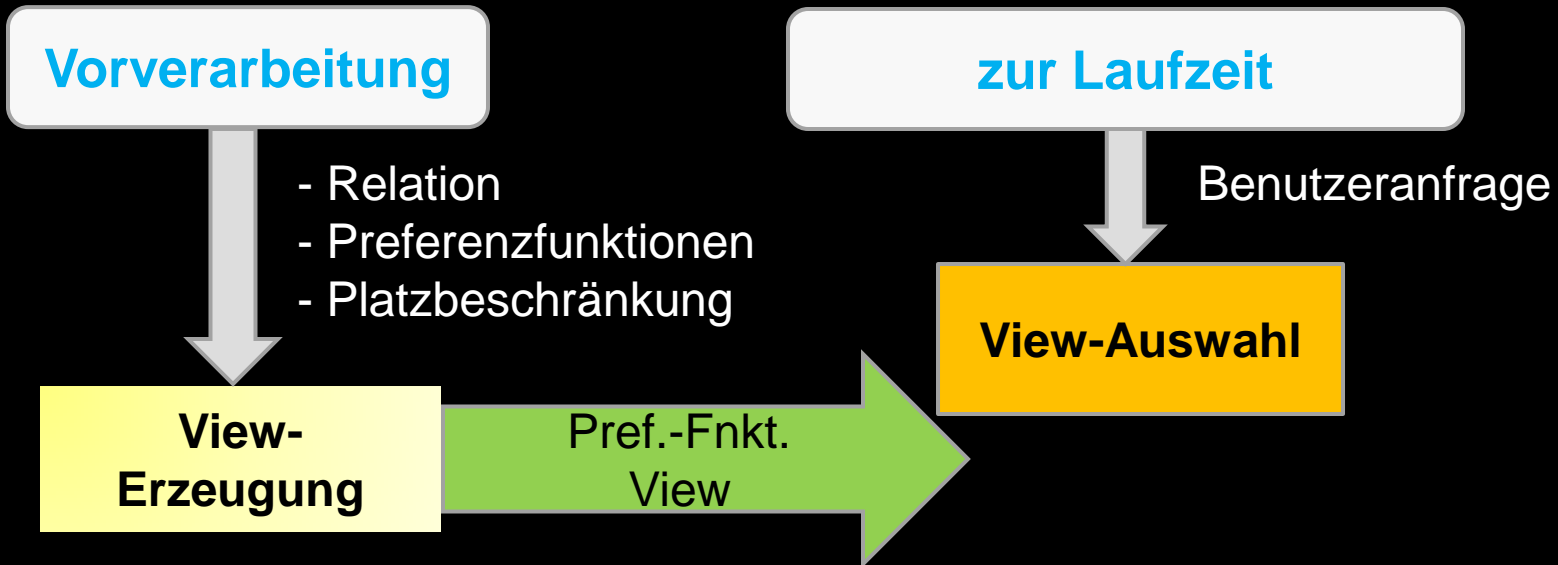
PREFER: Views



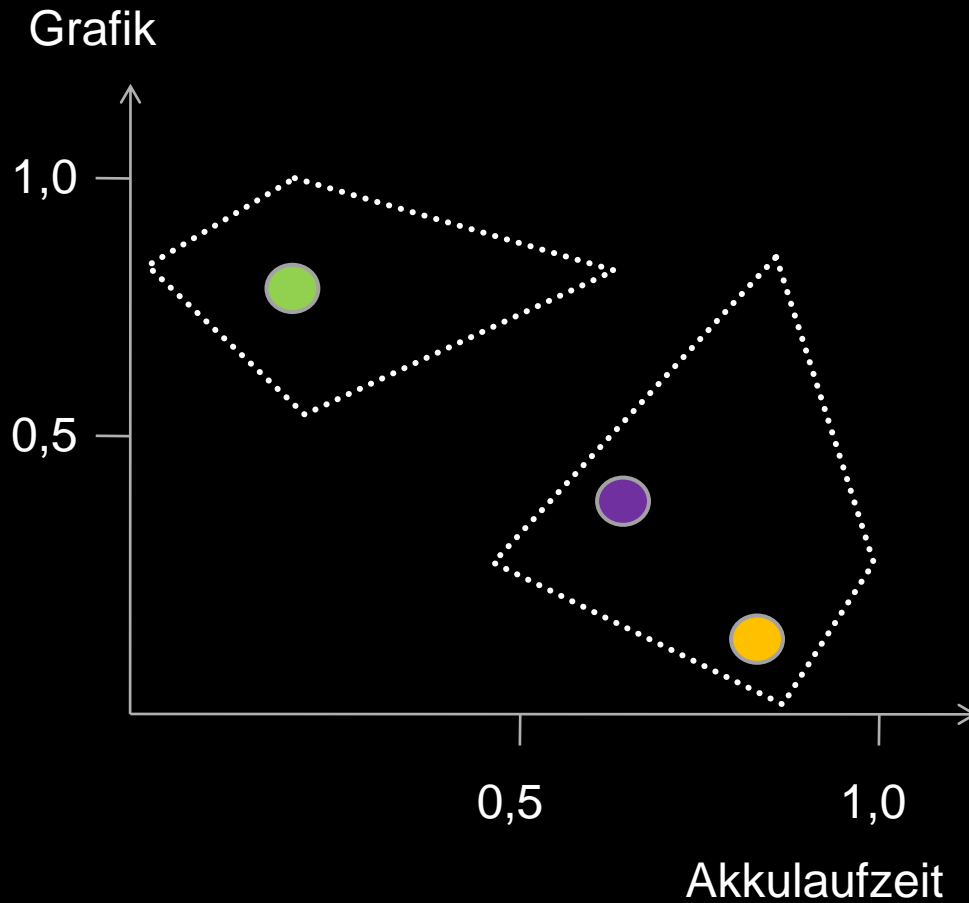
Ranked View 1
0,9 x Grafik +
0,1 x Akku

Ranked View 2
0,2 x Grafik +
0,8 x Akku

PREFER: Funktionsweise



PREFER: View-Auswahl

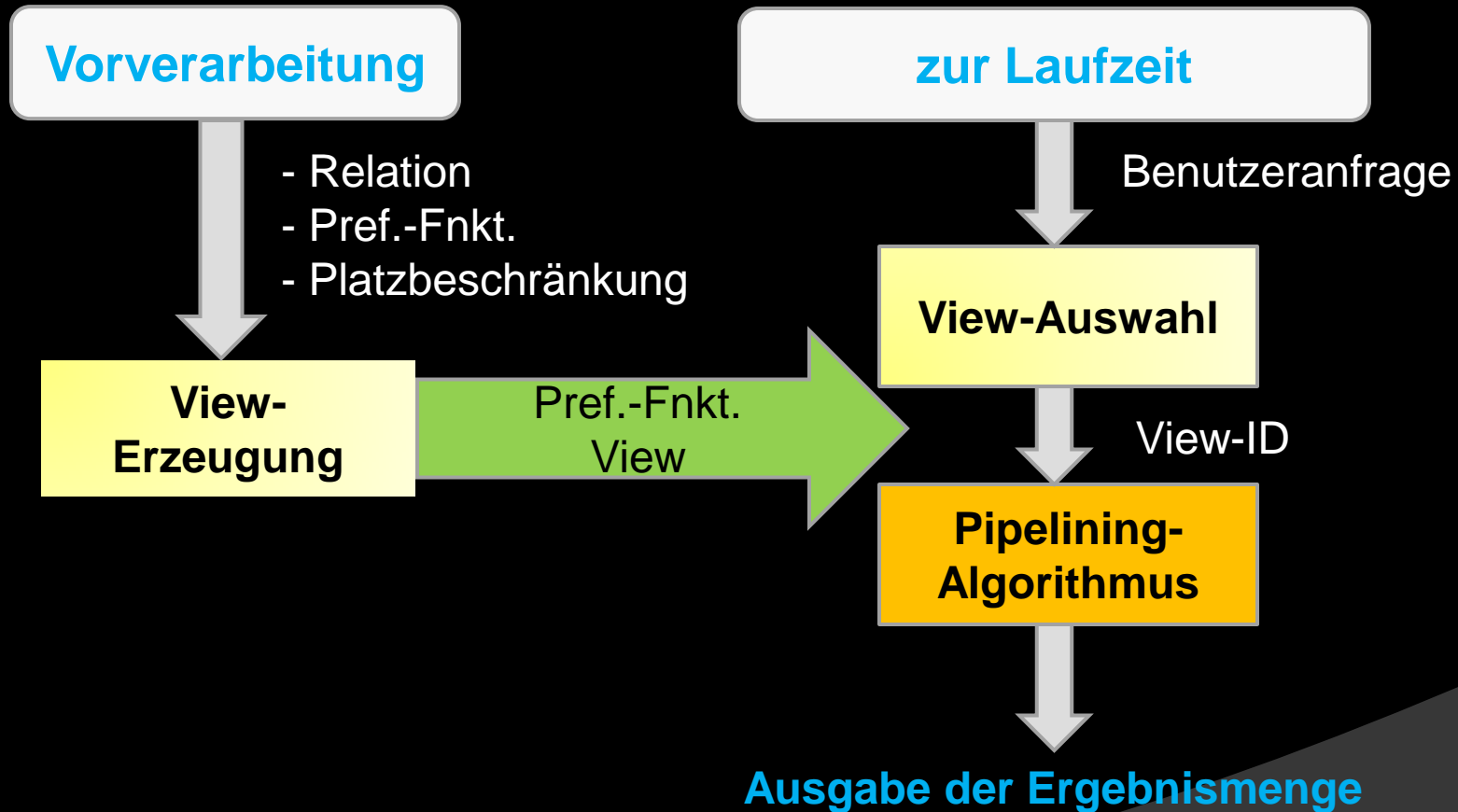


Ranked View 1
0,9 x Grafik +
0,1 x Akku

Ranked View 2
0,2 x Grafik +
0,8 x Akku

Benutzeranfrage
0,4 x Grafik +
0,6 x Akku

PREFER: Funktionsweise



PREFER: Pipelining-Algorithmus

Ranked View 2

ID	Hersteller	VS	BS
3	Toshiba	21,3	17,6
7	Lenovo	15,6	14,2
6	Apple	14,4	14,3
5	Acer	11,7	11,9
4	Sony	11,2	10,4
2	Samsung	9,5	9,0
1	Dell	6,4	7,8

View = $0,2 \times \text{Grafik} + 0,8 \times \text{Akkulaufzeit}$

Benutzer = $0,4 \times \text{Grafik} + 0,6 \times \text{Akkulaufzeit}$

PREFER: Pipelining-Algorithmus

Ranked View 2

ID	Hersteller	VS	BS	VS-BS
3	Toshiba	21,3	17,6	3,7
7	Lenovo	15,6	14,2	1,4
6	Apple	14,4	14,3	0,1
5	Acer	11,7	11,9	-0,2
4	Sony	11,2	10,4	0,8
2	Samsung	9,5	9,0	0,5
1	Dell	6,4	7,8	-1,4

t1

Ergebnismenge

ID	BS

Watermark von t1: $17,6 - 1,4 = 16,2$

PREFER: Pipelining-Algorithmus

Ranked View 2

ID	Hersteller	VS	BS	VS - BS
3	Toshiba	21,3	17,6	3,7
7	Lenovo	15,6	14,2	1,4
6	Apple	14,4	14,3	0,1
5	Acer	11,7	11,9	-0,2
4	Sony	11,2	10,4	0,8
2	Samsung	9,5	9,0	0,5
1	Dell	6,4	7,8	-1,4

t1 →

Ergebnismenge

ID	BS

Watermark von t1: $17,6 - 1,4 = 16,2$

PREFER: Pipelining-Algorithmus

Ranked View 2

ID	Hersteller	VS	BS	VS - BS
3	Toshiba	21,3	17,6	3,7
7	Lenovo	15,6	14,2	1,4
6	Apple	14,4	14,3	0,1
5	Acer	11,7	11,9	-0,2
4	Sony	11,2	10,4	0,8
2	Samsung	9,5	9,0	0,5
1	Dell	6,4	7,8	-1,4

t1

Ergebnismenge

ID	BS
3	17,6

Watermark von t1: $14,2 - 1,4 = 12,8$

PREFER: Pipelining-Algorithmus

Ranked View 2

ID	Hersteller	VS	BS	VS - BS
3	Toshiba	21,3	17,6	3,7
6	Apple	14,4	14,3	0,1
7	Lenovo	15,6	14,2	1,4
5	Acer	11,7	11,9	-0,2
4	Sony	11,2	10,4	0,8
2	Samsung	9,5	9,0	0,5
1	Dell	6,4	7,8	-1,4

t1 →

Ergebnismenge

ID	BS
3	17,6
6	14,3
7	14,2

Watermark von t1: $14,2 - 1,4 = 12,8$

Gegenüberstellung

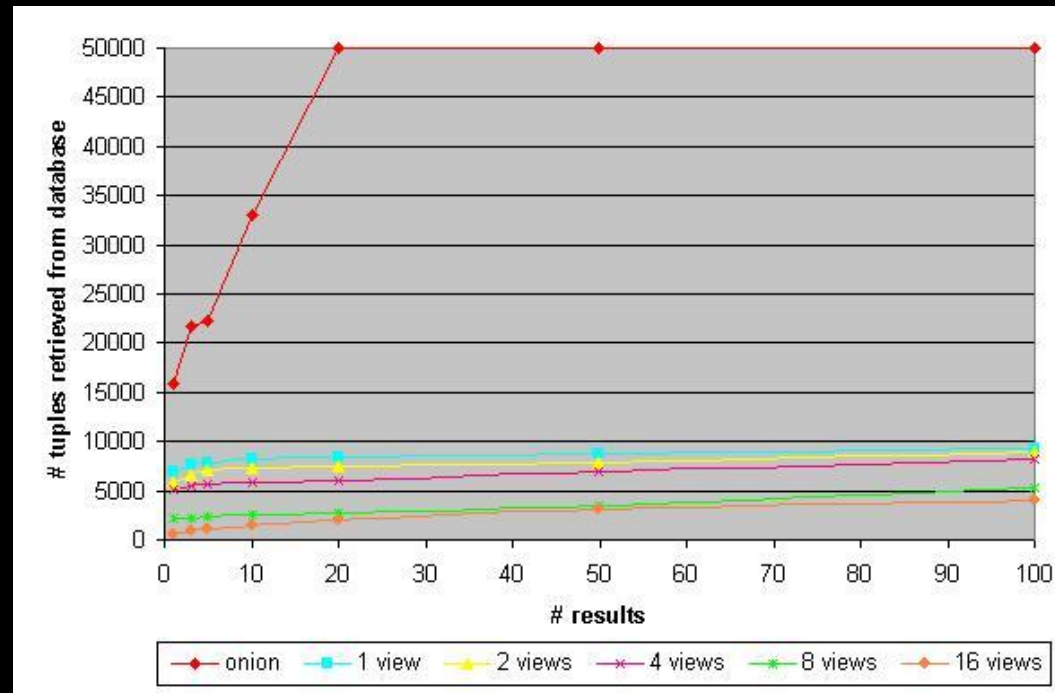
Ergebnis	
ID	Score
3	17,6
6	14,3
7	14,2

Ergebnis	
ID	BS
3	17,6
6	14,3
7	14,2

	Onion Technique	PREFER
Vorberechnung	Index	ranked Views
Verhalten bei steigender Attributanzahl	weniger Schichten → weniger effizient	kleinere Viewreichweite → mehr Views benötigt
worst case	k Schichten für Top-k	zu wenig Speicher

Direkter Vergleich

- ◎ 50.000 Datensätze
- ◎ 3 Attribute



- ◎ Indexerstellung: 2,5h
- ◎ Anzahl Schichten: 20

V I E L E N D A N K

E I N E N

S C H O E N E N

A B E N D

N O C H