

VU Datenbanksysteme, Kapitel 7

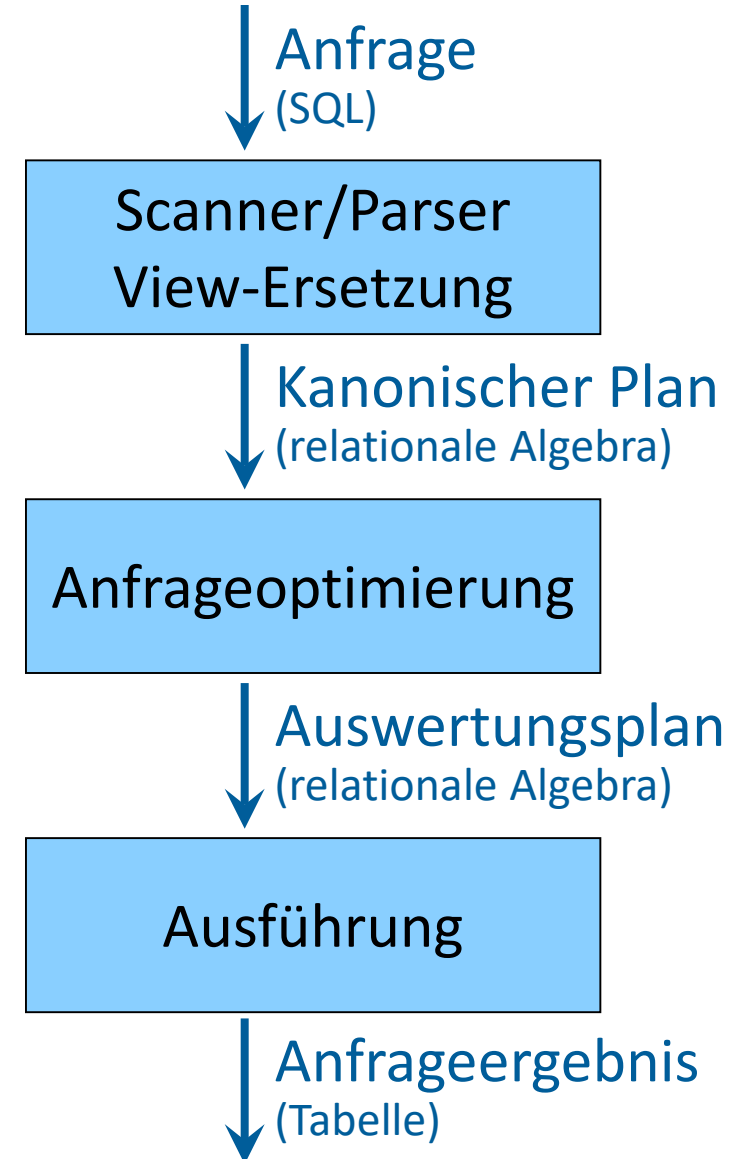
Relationale Anfrage-Optimierung

Christian Böhm

WS 2024/25: Kapitel 6 wird ab Freitag nachgeholt!

Aufgabe der Optimierung

- SQL ist deklarativ:
Anfrage spezifiziert, **WAS** das Ergebnis sein soll (Eigenschaften).
 - Relationale Algebra ist prozedural:
Ausdruck legt fest, **WIE** das Ergebnis berechnet wird.
- ➔ Übersetzung der deklarativen Anfrage in effizienten, prozeduralen Auswertungsplan



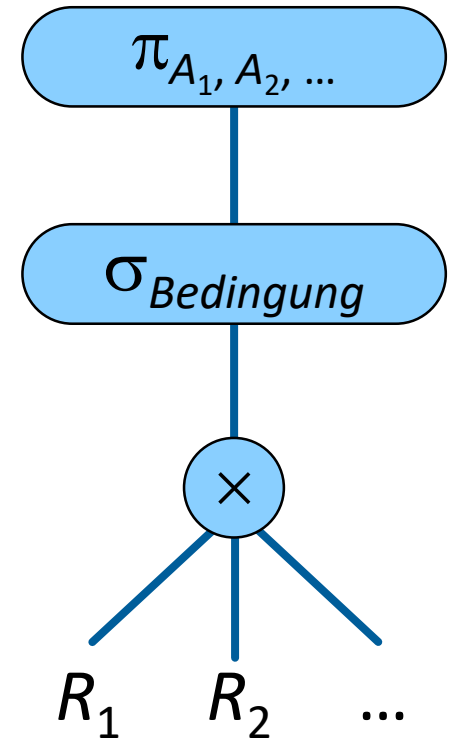
Kanonischer Auswertungsplan

1. Bilde Kreuzprodukt aller beteiligten Relationen,
2. Führe darauf die Selektionsbedingung aus,
3. Bilde die Projektion.

SELECT A_1, A_2, \dots
FROM R_1, R_2, \dots
WHERE *Bedingung*



$\pi_{A_1, A_2, \dots}(\sigma_{\text{Bedingung}}(R_1 \times R_2 \times \dots))$



Wir unterscheiden...

- **Logische Anfrageoptimierung:**

Optimierungstechniken, die den Auswertungsplan umbauen (d.h. die Reihenfolge der Operatoren im Baum verändern)

- **Physische Anfrageoptimierung:**

- z.B. Auswahl einer geeigneten Auswertungsstrategie für die Join-Operation
- Entscheidung, ob für eine Selektion ein Index verwendet wird oder nicht
- und wenn ja, welcher (bei unterschiedlichen Alternativen).

Es handelt sich also um die Auswahl eines geeigneten Algorithmus für jede Operation im Auswertungsplan.

Weiters unterscheiden wir...

- **Regelbasierte Optimierung:**

Es gibt zahlreiche Regeln zur Umformung in der relationalen Algebra, die in den meisten Fällen zur Verbesserung der Performanz führen:

- „Push Selection“: Selektionen möglichst früh (vor Joins) ausführen,
- Elimination leerer Teilbäume, Erkennen gleicher Teilbäume...

- **Kostenbasierte Optimierung:**

Schätzung der „Kosten“ (z.B. Laufzeit), die bei Ausführung entstehen

- Histogramme
- Parametrische Verteilungen
- Stichproben

Regelbasierte Anfrageoptimierung

Restrukturierungsalgorithmus:

1. Aufbrechen der Selektion und des Kreuzprodukts
2. Verschieben der Selektion(en) so weit wie möglich nach unten im Auswertungsplan („**Push Selection**“)
3. Zusammenfassen von Selektion und Kreuzprodukt zu Join
4. Einfügen und Verschieben von Projektionen so weit wie möglich nach unten („**Push Projection**“)
5. Zusammenfassen hintereinander folgender Selektionen.

Beispiel

Kunde	KNr	Name	Adresse	Region	Saldo
	201	Klein	Horn	Waldviertel	200 000
	337	Horn	Ebenthal	Weinviertel	100 000
	444	Berger	Illmitz	Seewinkel	300 000
	108	Weiss	Kuchl	Tennengau	500 000

Bestellt	BNr	Datum	KNr	PNr
	221	10.05.2021	201	12
	312	11.05.2021	201	4
	401	20.05.2021	337	330
	456	13.05.2021	444	330
	458	14.05.2021	444	98

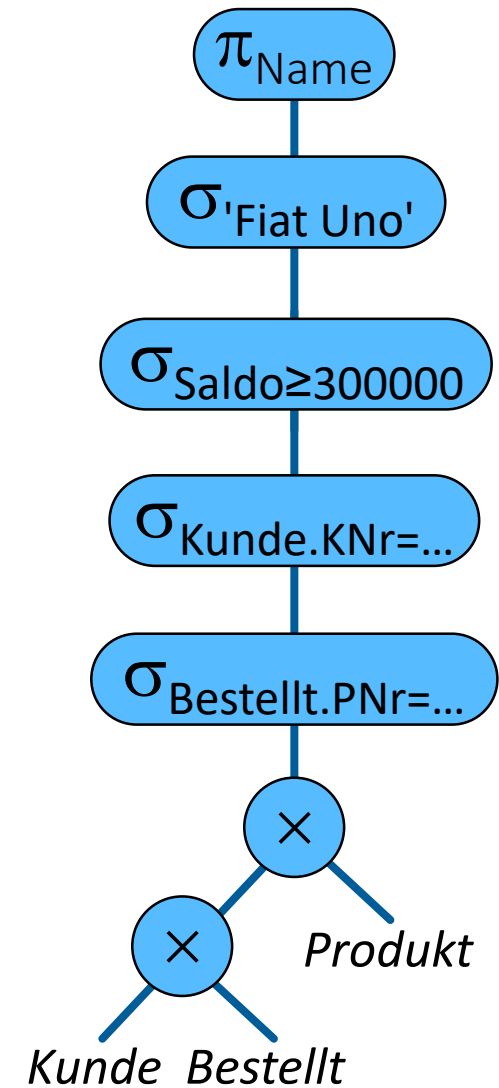
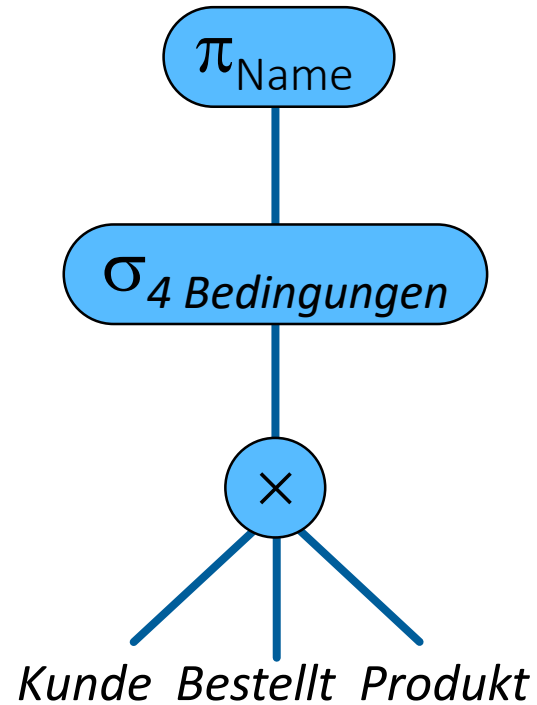
Produkt	PNr	Bezeichnung	Anzahl	Preis
	12	BMW 318i	10	40 000
	4	Golf 5	40	25 000
	330	Fiat Uno	5	18 000
	98	Ferrari 380	1	180 000
	14	Opel Corsa	14	17 000

- **Aufgabe:**

Welche *guten Kunden* ($\geq 300.000\text{€}$)
haben einen *Fiat Uno* bestellt?

- **SQL-Anfrage:**

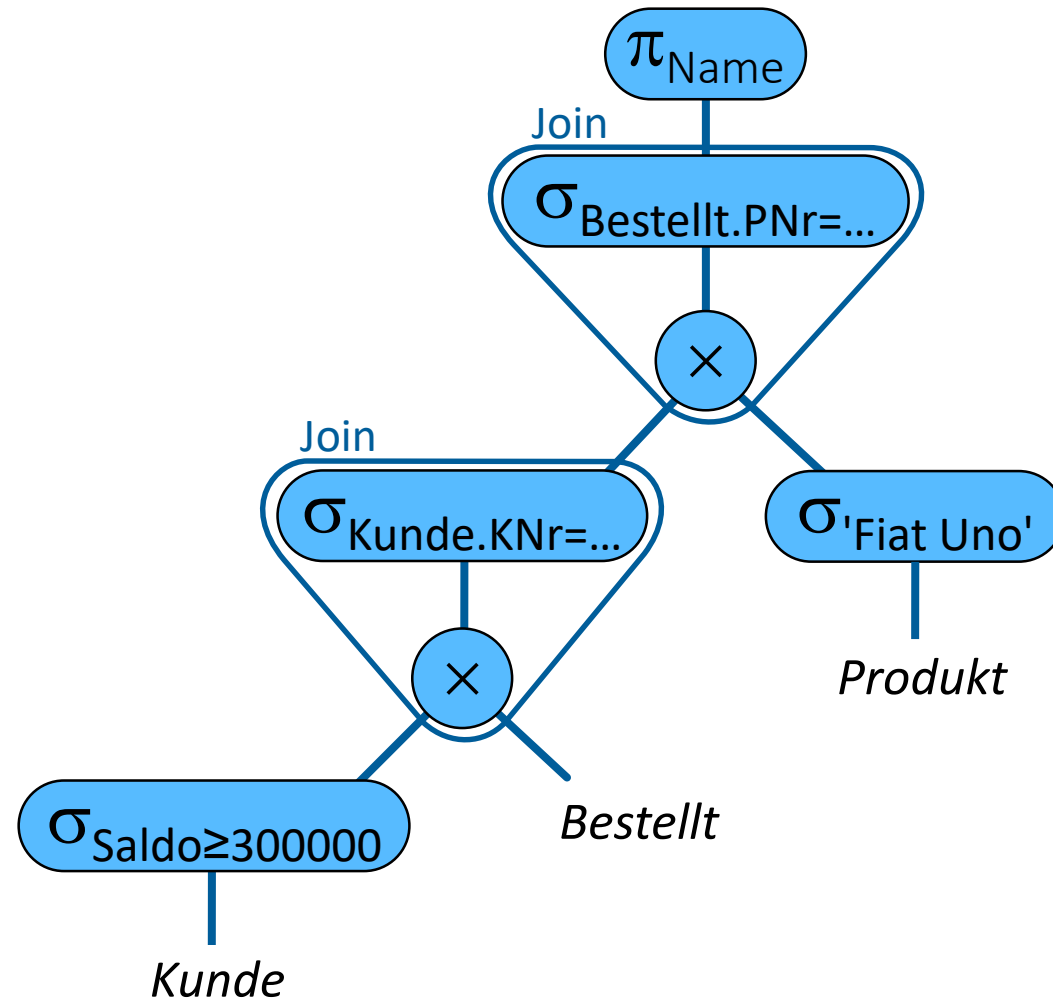
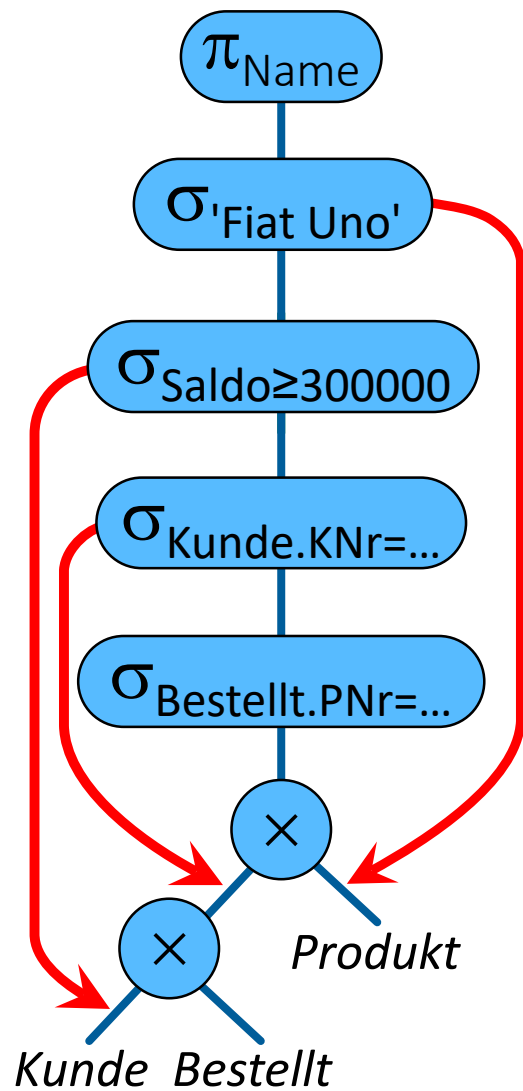
```
SELECT Name
FROM Kunde, Bestellt, Produkt
WHERE Bezeichnung = 'Fiat Uno'
AND Saldo >= 300000
AND Kunde.KNr = Bestellt.KNr
AND Bestellt.PNr = Produkt.PNr
```



Kunde	KNr	Name	Adresse	Region	Saldo
	201	Klein	Horn	Waldviertel	200 000
	337	Horn	Ebenthal	Weinviertel	100 000
	444	Berger	Illmitz	Seewinkel	300 000
	108	Weiss	Kuchl	Tennengau	500 000

Bestellt	BNr	Datum	KNr	PNr
	221	10.05.2021	201	12
	312	11.05.2021	201	4
	401	20.05.2021	337	330
	456	13.05.2021	444	330
	458	14.05.2021	444	98

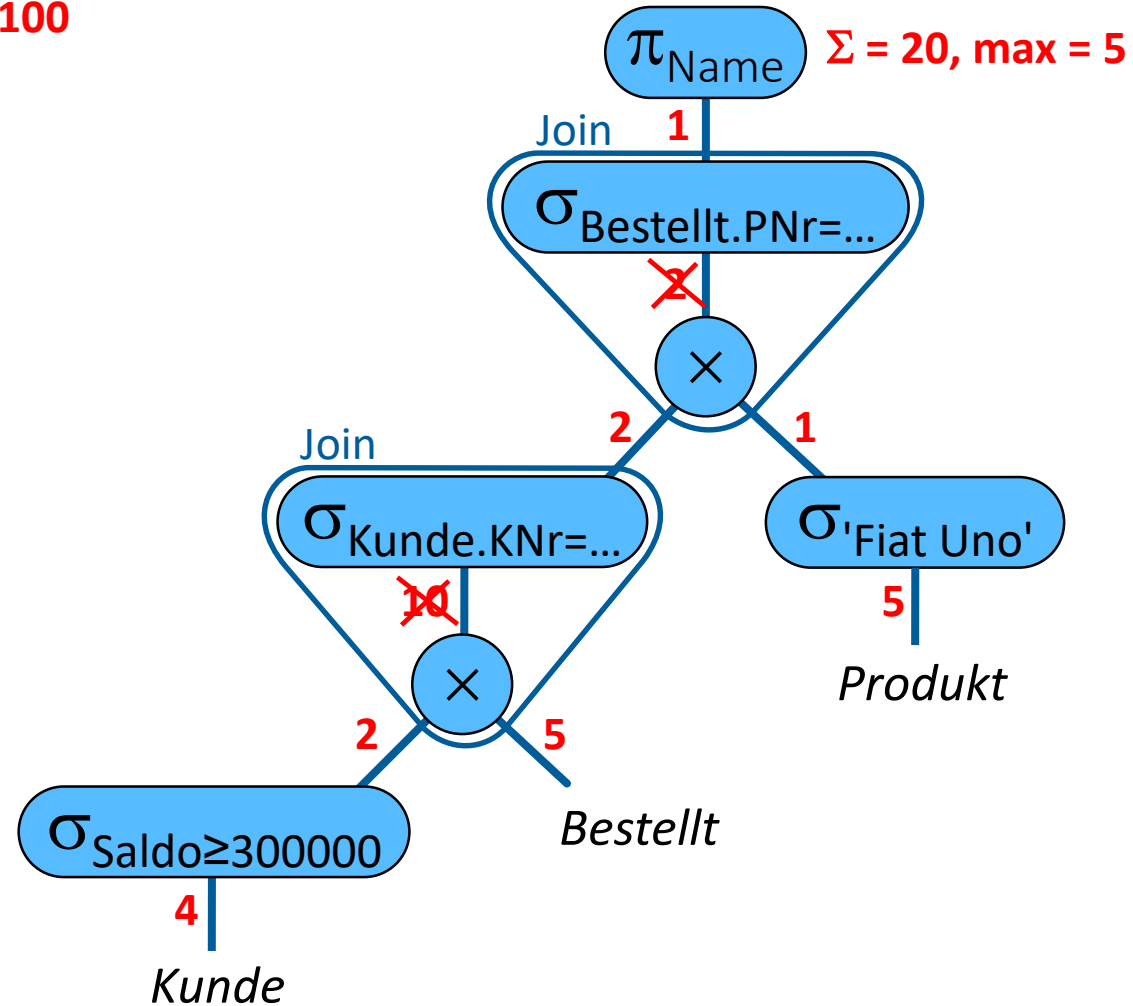
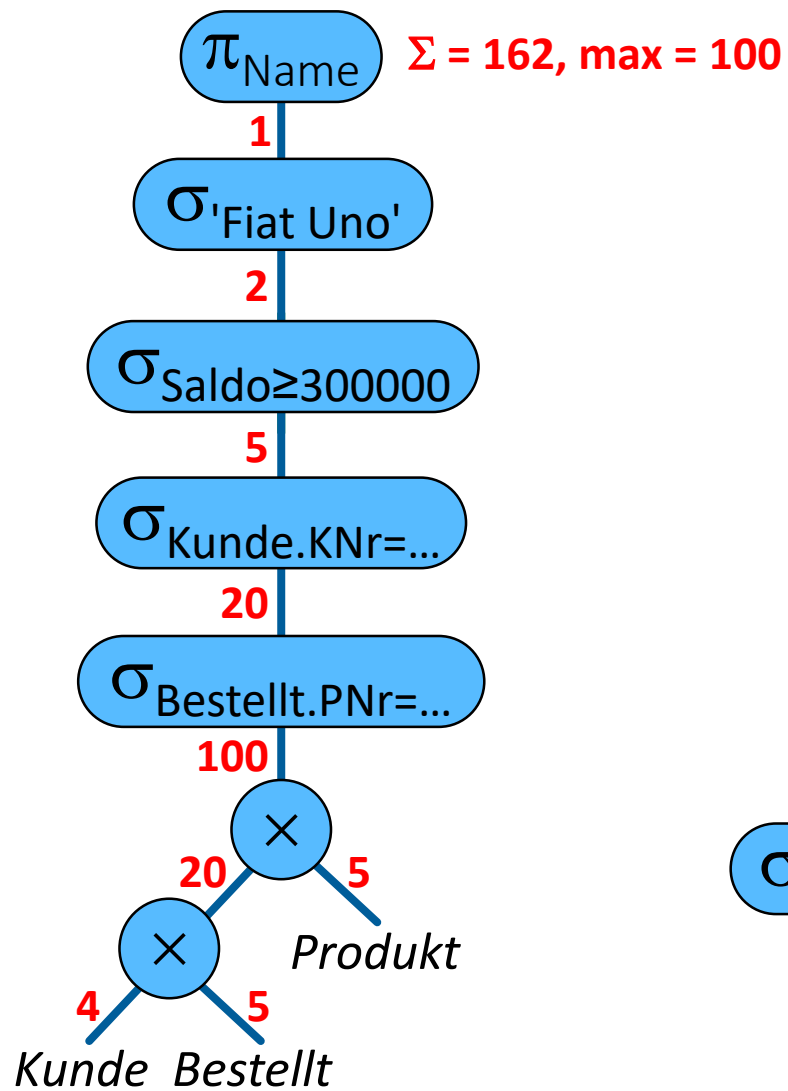
Produkt	PNr	Bezeichnung	Anzahl	Preis
	12	BMW 318i	10	40 000
	4	Golf 5	40	25 000
	330	Fiat Uno	5	18 000
	98	Ferrari 380	1	180 000
	14	Opel Corsa	14	17 000



Kunde	KNr	Name	Adresse	Region	Saldo
	201	Klein	Horn	Waldviertel	200 000
	337	Horn	Ebenthal	Weinviertel	100 000
	444	Berger	Illmitz	Seewinkel	300 000
	108	Weiss	Kuchl	Tennengau	500 000

Bestellt	BNr	Datum	KNr	PNr
	221	10.05.2021	201	12
	312	11.05.2021	201	4
	401	20.05.2021	337	330
	456	13.05.2021	444	330
	458	14.05.2021	444	98

Produkt	PNr	Bezeichnung	Anzahl	Preis
	12	BMW 318i	10	40 000
	4	Golf 5	40	25 000
	330	Fiat Uno	5	18 000
	98	Ferrari 380	1	180 000
	14	Opel Corsa	14	17 000



1, 2, ...: Anzahl Tupel („Kosten“)

Kunde	KNr	Name	Adresse	Region	Saldo
	201	Klein	Horn	Waldviertel	200 000
	337	Horn	Ebenthal	Weinviertel	100 000
	444	Berger	Illmitz	Seewinkel	300 000
	108	Weiss	Kuchl	Tennengau	500 000

Bestellt	BNr	Datum	KNr	PNr
	221	10.05.2021	201	12
	312	11.05.2021	201	4
	401	20.05.2021	337	330
	456	13.05.2021	444	330
	458	14.05.2021	444	98

Produkt	PNr	Bezeichnung	Anzahl	Preis
	12	BMW 318i	10	40 000
	4	Golf 5	40	25 000
	330	Fiat Uno	5	18 000
	98	Ferrari 380	1	180 000
	14	Opel Corsa	14	17 000

Fazit

- Die Gesamt-Summe Σ der Tupel ist ein Maß für die Verarbeitungszeit, das Maximum **max** ein Maß für den Speicherbedarf der Anfrage.
- Nach Anwendung des (regelbasierten) Restrukturierungsalgorithmus ist der Auswertungsplan erheblich effizienter (**20** vs. **162** im Beispiel).
- Bei komplexeren Beispielen finden **kostenbasierte** Optimierer nochmals deutlich effizientere Auswertungspläne.
- Für jeden Operator im Auswertungsplan (z.B. Join) muss nun noch ein geeignetes Verfahren gewählt werden (physische Optimierung).

Das Verbesserungspotential beträgt viele Größenordnungen

(im Gggs. zu imperativen Programmen, dort oft nur einige Prozent)