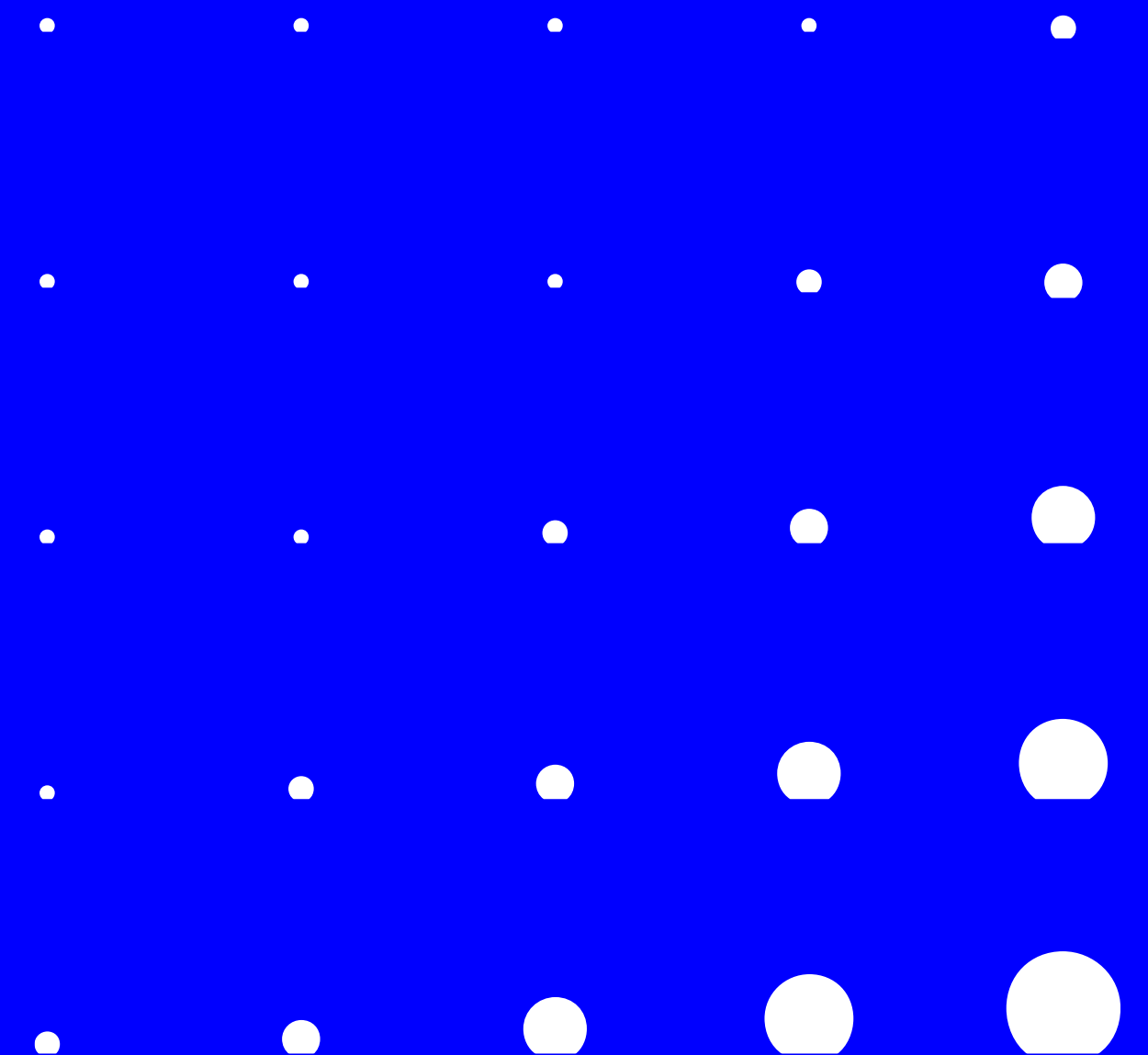


Relationenalgebra: Grundlagen, Selektion, Projektion



Was ist Relationenalgebra?

- In der Theorie der Datenbanken versteht man unter einer Relationenalgebra oder einer relationalen Algebra eine formale Sprache, mit der sich Abfragen über einem relationalen Schema formulieren lassen.
- Sie erlaubt es, Relationen
 - miteinander zu verknüpfen oder
 - zu reduzieren und
 - komplexere Informationen daraus herzuleiten.
- Die Relationenalgebra definiert Operationen, die sich auf einer Menge von Relationen anwenden lassen.
- Damit lassen sich beispielsweise Relationen
 - verknüpfen,
 - filtern oder
 - umbenennen.

Was ist Relationenalgebra?

- Die Ergebnisse aller Operationen sind ebenfalls Relationen.
- Aus diesem Grund bezeichnet man die Relationenalgebra als abgeschlossen.
- Ihre Bedeutung hat die Relationenalgebra als theoretische Grundlage für Abfragesprachen in relationalen Datenbanken.
- Hier werden die Operationen der relationalen Algebra in so genannten Datenbankoperatoren implementiert.
- Wenn jede Operation der relationalen Algebra in der Abfragesprache durch (mindestens) einen Ausdruck umgesetzt werden kann, heißt sie relational vollständig; der Ausdruck kann hierbei mehrere Datenbankoperatoren verknüpfen.
- SQL ist relational vollständig.

Was ist Relationenalgebra?

- Ein übliches minimales System aus Operationen besteht aus den sechs Operationen:
 - Projektion,
 - Selektion,
 - Kreuzprodukt,
 - Vereinigung,
 - Differenz und
 - Umbenennung.
- Alle weiteren Operationen lassen sich aus diesen minimalen Operationen zusammensetzen.

Noch ein paar Begriffe...

- Eine Relation ist eine Datenbanktabelle.
Sie besteht aus Attributen und Tupeln.
- Ein Attribut beschreibt den Typ eines möglichen Attributwertes und bezeichnet ihn mit einem Attributnamen.
- Ein Tupel stellt eine konkrete Kombination von Attributwerten dar und wird im Datenbankbereich auch als Datensatz bzw. Tabellenzeile bezeichnet.

Grundoperator: Die Selektion σ (Sigma)

Was macht man bei einer Selektion?

- Bei der Selektion kann man mit einem Ausdruck festlegen, welche Tupel in die Ergebnismenge aufgenommen werden sollen.
- Dieser Ausdruck wird auch Prädikat oder Selektionsbedingung genannt, es ist immer ein Bool'scher Ausdruck.
- Es werden also einzelne Tupel, also Tabellenzeilen, ausgeblendet.
- Bei der Selektion selektiert man nur einzelne Datensätze, also einzelne Tabellenzeilen aus der Menge aller Tabellenzeilen.
- Sei R eine Relation, dann schreibt man: $\sigma_{\text{Ausdruck}}(R)$

$$\sigma_{\text{Ausdruck}}(R) := \{t | t \in R \wedge t \text{ erfüllt Ausdruck}\}$$

R:			R[A=1]:			R[C>6]:		
A	B	C	A	B	C	A	B	C
1	2	4	1	2	4	4	6	7
4	6	7	1	6	7	1	6	7
1	6	7						
8	6	1						

- Beispiel in SQL:
SELECT * FROM student WHERE matr > 700000;
 - R := student
 - Ausdruck := matr > 700000

Selektion: Beispiel

Mitarbeiter

PNr	Name	Vorname	Abteilung
001	Huber	Erwin	01
002	Mayer	Hugo	01
003	Müller	Anton	02

Alle Mitarbeiter von Abteilung 01:

$\sigma_{\text{Abteilung}=01}(\text{Mitarbeiter})$

PNr	Name	Vorname	Abteilung
001	Huber	Erwin	01
002	Mayer	Hugo	01

Selektion, Kreuzprodukt -> Join

Mitarbeiter \times Abteilungen

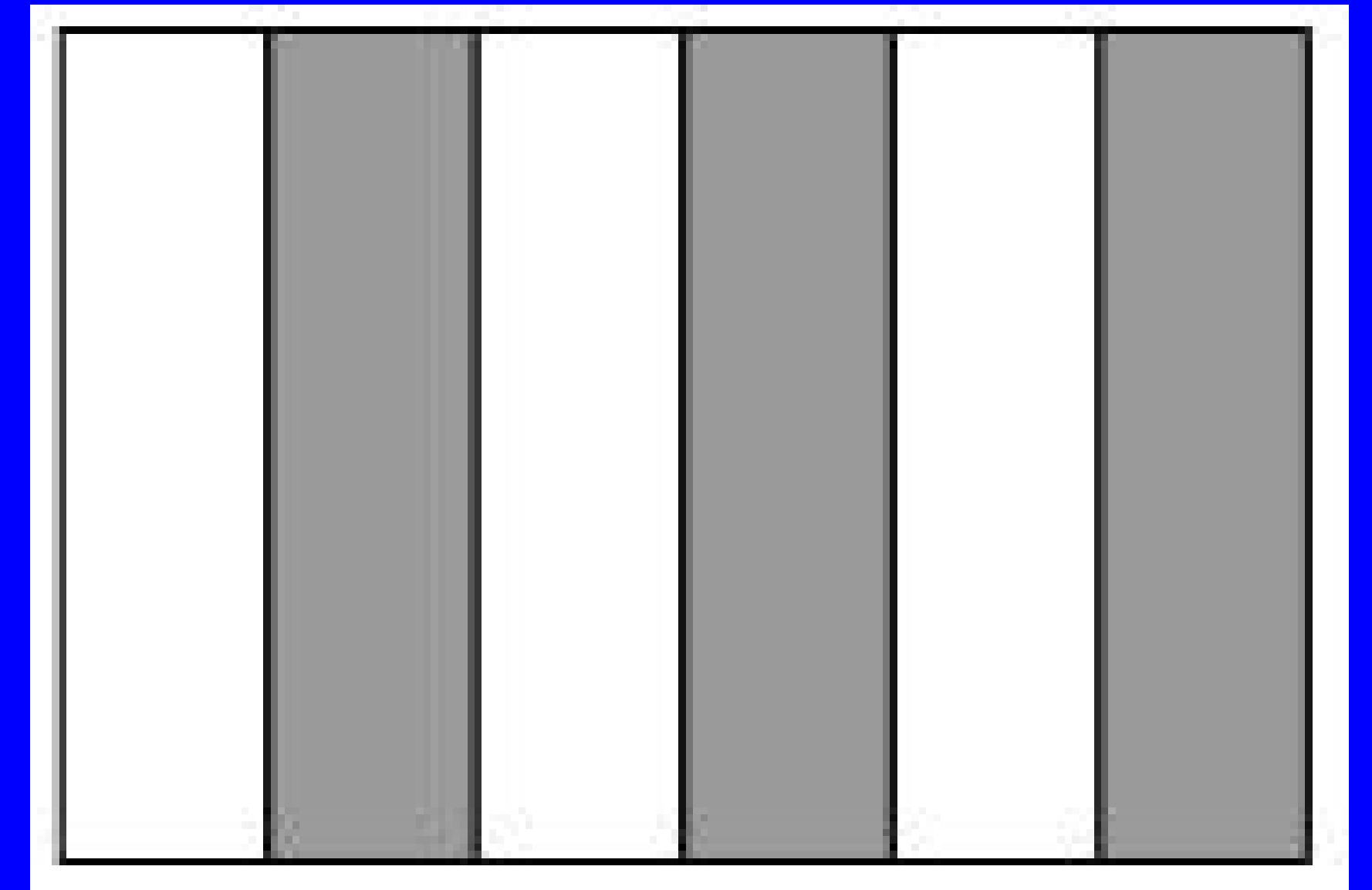
PNr	Name	Vorname	Abteilung	ANr	Abteilungsname
001	Huber	Erwin	01	01	Buchhaltung
001	Huber	Erwin	01	02	Produktion
002	Mayer	Hugo	01	01	Buchhaltung
002	Mayer	Hugo	01	02	Produktion
003	Müller	Anton	02	01	Buchhaltung
003	Müller	Anton	02	02	Produktion

$\sigma_{\text{Abteilung}=\text{ANr}}$ (Mitarbeiter \times Abteilungen)

PNr	Name	Vorname	Abteilung	ANr	Abteilungsname
001	Huber	Erwin	01	01	Buchhaltung
002	Mayer	Hugo	01	01	Buchhaltung
003	Müller	Anton	02	02	Produktion

Die Kombination aus Selektion und Kreuzprodukt heißt **Join**

Grundoperator: Die Projektion Π (P_i)



Was macht man bei einer Projektion?

- Bei der Selektion kann man mit einem Ausdruck festlegen, welche Tupel in die Ergebnismenge aufgenommen werden sollen.
- Dieser Ausdruck wird auch Prädikat oder Selektionsbedingung genannt.
 - Es werden also einzelne Tupel, also Tabellenzeilen, ausgeblendet.
- Bei der Projektion wählt man nur einzelne Spalten aus der Menge aller Tabellenspalten aus.
- β (Beta) nennt man dabei auch Projektionsliste oder Attributliste.
 - Das ist die Menge der Tabellenspalten, die man anzeigen lassen will.
 - Duplikate in der Ergebnisrelation werden dabei eliminiert.
- Die angegebenen Spalten müssen natürlich auch in R enthalten sein.
- Sei R eine Relation, dann schreibt man:
 $\Pi_{\beta}(R)$

Projektion

$$\pi_{\beta}(R) := \{t_{\beta} | t \in R\}$$

R:			R[A,B]:		R[A]:
A	B	C	A	B	A
1	2	3	1	2	1
4	5	6	4	5	4
1	3	8	1	3	1

- Beispiel in SQL:
SELECT nachname, matrikelnummer FROM student;
 - $R := \text{student}$
 - $\beta := \text{nachname, matrikelnummer}$

Projektion: Beispiel

Mitarbeiter

PNr	Name	Vorname	Abteilung
001	Huber	Erwin	01
002	Mayer	Josef	01
003	Müller	Anton	02
004	Mayer	Maria	01

$\pi_{\text{Name, Abteilung}}(\text{Mitarbeiter}) = \dots$

Zwischenergebnis:

Name	Abteilung
Huber	01
Mayer	01
Müller	02
Mayer	01

↪ **Duplikate**

Elimination →

Name	Abteilung
Huber	01
Mayer	01
Müller	02