

Einführung in Datenbanken

Kapitel 12: Relationale Algebra in SQL

Prof. Dr. Stefan Brass

Martin-Luther-Universität Halle-Wittenberg

Wintersemester 2019/20

<http://www.informatik.uni-halle.de/~brass/db19/>

Inhalt

- ① Einleitung
- ② UNION
- ③ Outer Join

Beispiel-Datenbank

STUDENTEN

<u>SID</u>	VORNAME	NACHNAME	EMAIL
101	Lisa	Weiss	...
102	Michael	Grau	NULL
103	Daniel	Sommer	...
104	Iris	Winter	...

AUFGABEN

<u>ATYP</u>	<u>ANR</u>	THEMA	MAXPT
H	1	ER	10
H	2	SQL	10
Z	1	SQL	14

BEWERTUNGEN

<u>SID</u>	<u>ATYP</u>	<u>ANR</u>	PUNKTE
101	H	1	10
101	H	2	8
101	Z	1	12
102	H	1	9
102	H	2	9
102	Z	1	10
103	H	1	5
103	Z	1	7

Von Relationaler Algebra zu SQL (1)

- Ein Ausdruck in relationaler Algebra

$$\pi_{A_1, \dots, A_n} \left(\sigma_F (R_1 \times \dots \times R_m) \right)$$

wird in SQL geschrieben als

```
SELECT A1, . . . , An
FROM   R1, . . . , Rm
WHERE  F
```

- Das Schlüsselwort **SELECT** entspricht also der Projektion.
- Die Relationen unter **FROM** werden mit \times verknüpft.
- Unter **WHERE** wird die Selektionsbedingung geschrieben.

Umgekehrt ist aber nicht jede WHERE-Bedingung auch als Selektionsbedingung möglich (keine Unteranfragen).

Von Relationaler Algebra zu SQL (2)

- Haben verschiedene R_j Attribute mit gleichem Namen A , schreibt man $R_j.A$, um den Bezug eindeutig zu machen.
- Man kann aber auch Umbenennungen verwenden:

$$\pi_{A_1, \dots, A_n} \left(\sigma_F \left(\rho_{X_1}(R_1) \times \dots \times \rho_{X_m}(R_m) \right) \right)$$

wird in SQL geschrieben als

```
SELECT  A1, ..., An
FROM    R1 X1, ..., Rm Xm
WHERE   F
```

Man kann die X_i auch als Kopien der Relationen R_i verstehen.

- SQL arbeitet mit Multimengen von Tupeln, Duplikate sind also möglich.
- Bei Bedarf verwende man **SELECT DISTINCT** statt **SELECT**.

Von Relationaler Algebra zu SQL (3)

- SQL hat einen **UNION**-Operator, um Ergebnisse mehrerer **SELECT**-Ausdrücke zu kombinieren.

Dieser wird unten eingeführt und ausführlich erklärt.

- Es gibt entsprechend auch andere Mengenoperationen in SQL, diese werden aber selten verwendet:
 - **EXCEPT** für Mengendifferenz,
 - **INTERSECT** für den Schnitt.

Meist verwendet man eher Unteranfragen (z.B. **NOT EXISTS**).

- Unteranfragen unter **FROM** erlauben den typischen Algebra-Stil, dass man Relationen schrittweise verknüpft.

Inhalt

1 Einleitung

2 UNION

3 Outer Join

UNION (1)

- In SQL kann man die Ergebnisse von zwei Anfragen mit **UNION** verknüpfen.

$R \cup S$ ist die Menge aller Tupel, die in R , in S , oder in beiden sind.

- UNION wird benötigt, da es sonst keine Möglichkeit gibt, **eine Ergebnisspalte mit Werten aus mehreren Tabellenspalten** (Eingabespalten) zu konstruieren.

Dies wird z.B. benötigt, wenn Subklassen durch verschiedene Tabellen repräsentiert werden. Z.B. könnte es eine Tabelle STUDENTEN und eine andere Tabelle GASTHÖRER geben.

- UNION wird auch für Fallunterscheidungen verwendet (um **if ... then ... else ...** darzustellen).

UNION (2)

- Die Unteranfragen, die durch UNION verknüpft werden, müssen Tabellen mit der gleichen Anzahl von Spalten liefern. Die Datentypen der korrespondierenden Spalten müssen kompatibel sein.

Die Attributnamen müssen nicht übereinstimmen. Oracle und SQL Server verwenden im Ergebnis die Attributnamen der ersten Unteranfrage.

DB2 verwendet ggf. künstliche Spaltennamen (1, 2, ...).

- SQL unterscheidet zwischen

- **UNION**: \cup mit Duplikatelimination und

Es werden alle Duplikate entfernt, nicht nur solche, die durch die Vereinigung entstanden sind.

- **UNION ALL**: Konkatenation (erhält Duplikate).

Duplikatelimination ist ziemlich teuer.

UNION (3)

- Geben Sie für jeden Studenten die Gesamtpunktzahl für Hausaufgaben aus (auch für Studierende, die keine Aufgaben abgegeben haben).

```
SELECT  VORNAME, NACHNAME, SUM(PUNKTE) AS SUMME
FROM    STUDENTEN S, BEWERTUNGEN B
WHERE   S.SID = B.SID AND ATYP = 'H'
GROUP BY S.SID, VORNAME, NACHNAME
UNION ALL
SELECT  VORNAME, NACHNAME, 0 AS SUMME
FROM    STUDENTEN S
WHERE   S.SID NOT IN (SELECT SID
                      FROM  BEWERTUNGEN
                      WHERE  ATYP = 'H')
```

UNION (4)

- Erstellen Sie Noten für die Studenten basierend auf Hausaufgabe 1:

```
SELECT S.SID, S.VORNAME, S.NACHNAME, 1 NOTE  
FROM STUDENTEN S, BEWERTUNGEN B  
WHERE S.SID=B.SID AND B.ATYP='H' AND B.ANR=1  
AND B.PUNKTE >= 9
```

UNION ALL

```
SELECT S.SID, S.VORNAME, S.NACHNAME, 2 NOTE  
FROM STUDENTEN S, BEWERTUNGEN B  
WHERE S.SID=B.SID AND B.ATYP='H' AND B.ANR=1  
AND B.PUNKTE >= 7 AND B.PUNKTE < 9
```

UNION ALL

...

Vereinigung vs. Verbund

Aufgabe:

- Zwei Alternativen zur Repräsentation der Bewertungen für Hausaufgaben und Klausuren sind:

STUDENT	H	Z	E
Jim Ford	95	60	75
Ann Lloyd	80	90	95

STUDENT	ATYP	PZT
Jim Ford	H	95
Jim Ford	Z	60
Jim Ford	E	75
Ann Lloyd	H	80
Ann Lloyd	Z	90
Ann Lloyd	E	95

- Übersetzen Sie in beiden Richtungen mittels SQL.

Bedingte Ausdrücke (1)

- Während UNION eine klassische Lösung für Fallunterscheidungen ist, haben neuere DBMS-Versionen auch bedingte Ausdrücke (im SQL-Standard seit SQL-92).
- Z.B.: Ausgabe der Ergebnisse von Lisa Weiss, dabei Aufgabentyp ausgeschrieben:

```
SELECT CASE WHEN ATYP='H' THEN 'Hausaufgabe'
           WHEN ATYP='Z' THEN 'Zwischenklausur'
           WHEN ATYP='E' THEN 'Endklausur'
           ELSE 'Unbekannte Kat.' END,
        ANR, PUNKTE
FROM   STUDENTEN S, BEWERTUNGEN B
WHERE  S.SID = B.SID
AND    VORNAME = 'Lisa' AND NACHNAME = 'Weiss'
```


Beispiel-Datenbank

STUDENTEN

<u>SID</u>	<u>VORNAME</u>	<u>NACHNAME</u>	<u>EMAIL</u>
101	Lisa	Weiss	...
102	Michael	Grau	NULL
103	Daniel	Sommer	...
104	Iris	Winter	...

AUFGABEN

<u>ATYP</u>	<u>ANR</u>	<u>THEMA</u>	<u>MAXPT</u>
H	1	ER	10
H	2	SQL	10
Z	1	SQL	14

BEWERTUNGEN

<u>SID</u>	<u>ATYP</u>	<u>ANR</u>	<u>PUNKTE</u>
101	H	1	10
101	H	2	8
101	Z	1	12
102	H	1	9
102	H	2	9
102	Z	1	10
103	H	1	5
103	Z	1	7

Verbunde in SQL-92 (3)

- Aktuelle Systeme unterstützen den Standard nur teilweise:
 - SQL-92 Verbunde können nicht in Oracle 8i verwendet werden, in Oracle 9i dagegen fast alle.
 - Einige Verbundtypen werden in DB2, SQL Server und Access unterstützt, aber der “natürliche Verbund” nicht. Man muss in diesen Systemen die Bedingung explizit aufschreiben:

```
SELECT SID, B.ANR, (PUNKTE/MAXPT)*100
FROM   BEWERTUNGEN B INNER JOIN AUFGABEN A
      ON B.ATYP = A.ATYP AND B.ANR = A.ANR
WHERE  B.ATYP = 'H'
```

Verbunde in SQL-92 (4)

- Mit der expliziten Verbundbedingung ist die Anfrage nicht kürzer als die äquivalente Anfrage in klassischer Syntax (Verbund-Bedingung unter `WHERE`).
- Die Ausdruckskraft von SQL wird durch die neuen Verbund-Konstrukte nicht vergrößert.
 - Jede Anfrage mit den neuen Verbund-Konstrukten kann in eine äquivalente Anfrage ohne diese Konstrukte überführt werden.
- Der Grund, warum Verbunde zu SQL hinzugefügt wurden, ist wahrscheinlich der “**äußere Verbund**”: Hierfür ist die äquivalente Formulierung in SQL-86 deutlich länger.

Äußerer Verbund: Wdh.

- Der normale Verbund eliminiert Tupel ohne Verbundpartner:

$$\begin{array}{|c|c|} \hline A & B \\ \hline a_1 & b_1 \\ \hline a_2 & b_2 \\ \hline \end{array} \bowtie \begin{array}{|c|c|} \hline B & C \\ \hline b_2 & c_2 \\ \hline b_3 & c_3 \\ \hline \end{array} = \begin{array}{|c|c|c|} \hline A & B & C \\ \hline a_2 & b_2 & c_2 \\ \hline \end{array}$$

- Der linke äußere Verbund stellt sicher, dass Tupel der linken Tabelle auch im Ergebnis vorhanden sind:

$$\begin{array}{|c|c|} \hline A & B \\ \hline a_1 & b_1 \\ \hline a_2 & b_2 \\ \hline \end{array} \ltimes \begin{array}{|c|c|} \hline B & C \\ \hline b_2 & c_2 \\ \hline b_3 & c_3 \\ \hline \end{array} = \begin{array}{|c|c|c|} \hline A & B & C \\ \hline a_1 & b_1 & \\ \hline a_2 & b_2 & c_2 \\ \hline \end{array}$$

Zeilen der linken Seite werden, falls notwendig, mit "Null" aufgefüllt.

Es gibt auch einen rechten und einen vollen äußeren Verbund.

Äußerer Verbund in SQL (1)

- Z.B. Anzahl der Abgaben pro Hausaufgabe. Falls es keine Abgabe gibt, soll 0 ausgegeben werden:

```
SELECT  A.ANR, COUNT(SID)
FROM    AUFGABEN A LEFT OUTER JOIN BEWERTUNGEN B
        ON A.ATYP = B.ATYP AND A.ANR = B.ANR
WHERE   A.ATYP = 'H'
GROUP BY A.ANR
```

- Im Ergebnis des linken äußeren Verbunds treten alle Übungen auf. In Übungen ohne Abgaben werden die Attribute SID und PUNKTE mit Nullwerten aufgefüllt.
- COUNT(SID) zählt nur Zeilen, wo SID nicht Null ist.

Äußerer Verbund in SQL (2)

- Äquivalente Anfrage in SQL-86 (12 vs. 5 Zeilen):

```

SELECT  A.ANR, COUNT(*)
FROM    AUFGABEN A, BEWERTUNGEN B
WHERE   A.ATYP = 'H' AND B.ATYP = 'H'
AND     A.ANR = B.ANR
GROUP BY A.ANR
UNION ALL
SELECT  A.ANR, 0
FROM    AUFGABEN A
WHERE   A.ATYP = 'H'
AND     A.ANR NOT IN (SELECT B.ANR
                       FROM    BEWERTUNGEN B
                       WHERE   B.ATYP = 'H')

```

Äußerer Verbund in SQL (3)

- Geben Sie für jeden Studenten die Anzahl der abgegebenen Hausaufgaben aus (einschließlich 0).
- Die folgende Anfrage funktioniert nicht: Studenten ohne Hausaufgaben werden nicht aufgelistet.

```
SELECT VORNAME, NACHNAME, COUNT(ANR) Falsch!
FROM   STUDENTEN S LEFT OUTER JOIN BEWERTUNGEN B
      ON S.SID = B.SID
WHERE  B.ATYP = 'H'
GROUP BY S.SID, VORNAME, NACHNAME
```

- Der äußere Verbund wird konstruiert, bevor die WHERE-Bedingung ausgewertet wird.

Äußerer Verbund in SQL (4)

- Mögliche Verbundpartner dürfen nicht nach Konstruktion des äußeren Verbundes eliminiert werden.
- Man muss die Hausaufgabenergebnisse selektieren bevor der äußere Verbund gemacht wird:

```
SELECT VORNAME, NACHNAME, COUNT(B.ANR)
FROM   STUDENTEN S LEFT OUTER JOIN
      (SELECT SID, ANR
       FROM   BEWERTUNGEN
       WHERE  ATYP = 'H') B
ON S.SID = B.SID
GROUP BY S.SID, VORNAME, NACHNAME
```

Äußerer Verbund in SQL (5)

- Man kann auch die Bedingung für die rechte Tabelle in die Verbundbedingung integrieren:

```
SELECT VORNAME, NACHNAME, COUNT(B.ANR)
FROM   STUDENTEN S LEFT OUTER JOIN BEWERTUNGEN B
      ON S.SID = B.SID AND B.ATYP = 'H'
GROUP BY S.SID, VORNAME, NACHNAME
```

- SQL-92 lässt jede WHERE-Bedingung, die sich nur auf die rechte oder linke Tabelle bezieht, zu.
(Das sollte aber nicht missbraucht werden.)

Es scheint, dass DB2 und Access keine Unteranfragen in der ON-Klausel zulassen. In Access müssen komplexere Bedingungen in Klammern eingeschlossen werden.

Äußerer Verbund in SQL (6)

- Bedingungen für die linke Tabelle machen in einem linken äußeren Verbund wenig Sinn.
- Z.B. betrachte man diese Anfrage:

```
SELECT A.ATYP, A.ANR, B.SID, B.PUNKTE
FROM   AUFGABEN A LEFT OUTER JOIN BEWERTUNGEN B
      ON A.ATYP = 'H' AND B.ATYP = 'H'
      AND A.ANR = B.ANR
```

- Aufgabe:
Wird A.ATYP = 'Z' in der Ausgabe auftauchen?

ja nein

Äußerer Verbund in SQL (7)

- Zum Teil kann NOT EXISTS durch einen äußeren Verbund ersetzt werden (z.B. für MySQL vor Ver. 4.1).
- Z.B. Studenten ohne eine gelöste Hausaufgabe:

```
SELECT S.SID, S.VORNAME, S.NACHNAME  
FROM   STUDENTEN S LEFT OUTER JOIN BEWERTUNGEN B  
      ON S.SID = B.SID AND B.ATYP = 'H'  
WHERE  B.ATYP IS NULL
```

- Natürlich kann man statt B.ATYP jedes Attribut von BEWERTUNGEN auf Null testen.

Der Test auf den Nullwert prüft, ob das aktuelle STUDENTEN-Tupel einen Verbundpartner gefunden hat.

Verbundsyntax: Übersicht (1)

- SQL-92 hat folgende Verbundtypen:
 - **[INNER] JOIN**: Gewöhnlicher Verbund.
 - **LEFT [OUTER] JOIN**: Erhält Tupel der linken Tabelle.
 - **RIGHT [OUTER] JOIN**: Erhält alle Tupel von rechts.
 - **FULL [OUTER] JOIN**: Erhält alle Eingabetupel.
 - **CROSS JOIN**: Kartesisches Produkt \times .
 - **UNION JOIN**: Diese Vereinigung füllt die Spalten der anderen Tabelle mit Nullwerten auf.

UNION JOIN wurde in SQL-99 gelöscht (ist aber interessante Idee).

- Schlüsselworte in **[...]** sind optional.

Verbundsyntax: Übersicht (2)

- Mögliche Spezifikationen der Verbundbedingung:
 - Schlüsselwort **NATURAL** vor dem Verbundnamen.
 - “**ON** *<Bedingung>*” folgt dem Verbund.
 - “**USING** (A_1, \dots, A_n) ” folgt dem Verbund.

USING listet alle Verbundattribute (z.B. um den natürlichen Verbund zu spezifizieren). Attribute mit den Namen A_1, \dots, A_n müssen in beiden Tabellen auftauchen. Die Verbundbedingung ist dann $R.A_1 = S.A_1 \wedge \dots \wedge R.A_n = S.A_n$. NATURAL ist äquivalent zu USING mit allen gleichen Attributnamen.

- Nur eines der Konstrukte kann verwendet werden.
- CROSS JOIN und UNION JOIN haben keine Verbundbedingung.

Verbandsyntax: Übersicht (3)

- Nach dem Standard liefern der NATURAL Join und der Join mit USING eine Tabelle mit nur einer Kopie der gemeinsamen Attribute.
- Diese Attribute sind die ersten Spalten im Ergebnis. Man kann sie nicht mit Tupelvariablen ansprechen.

```
SELECT *  
FROM BEWERTUNGEN B NATURAL JOIN AUFGABEN A
```

- Die Ergebnisspalten sind ATYP, ANR, B.SID, B.PUNKTE, A.THEMA, A.MAXPT (in dieser Reihenfolge).

Es ist unzulässig, sich auf B.ATYP oder A.ATYP zu beziehen: Es kann nur ATYP verwendet werden (das gleiche gilt für ANR).

Verbundsyntax: Übersicht (4)

- Man kann in der FROM-Klausel auch sowohl Verbunde verwenden, als auch weitere Tupelvariablen deklarieren (getrennt durch “,”).
- Das Ergebnis eines Verbundes zweier Tabellen kann mit einer dritten Tabelle verbunden werden (usw.):

```
SELECT ...
FROM   R LEFT JOIN S ON R.A=S.B
       LEFT JOIN T ON S.C=T.D
```

- Um Join-Ausdrücke kann man Klammern setzen, nicht um normale Tupelvariablen-Deklarationen.

Für den (...) -Ausdruck kann man keine Tupelvariable einführen.

Verbundsyntax: Details (1)

- Oracle 9i unterstützt die SQL-92 Verbunde.

Einschließlich des Natural Join, aber ohne **UNION JOIN** (der in SQL:1999 entfernt wurde). Oracle 8i unterstützte keinen der SQL-92 Verbunde.

- Innerer und äußerer Verbund mit **ON** funktionieren auch in DB2, SQL Server, Access und MySQL.

In Access und MySQL ist das Schlüsselwort **INNER** nicht optional.

- **USING** und **NATURAL** funktionieren nur in Oracle 9i.

NATURAL existiert auch in MySQL, aber MySQL vereinigt die gleichen Attribute nicht. Das verletzt den SQL-92 Standard.

Verbundsyntax: Details (2)

- **CROSS JOIN** wird nur in Oracle 9i, SQL Server und MySQL, aber nicht in Access und DB2 unterstützt.

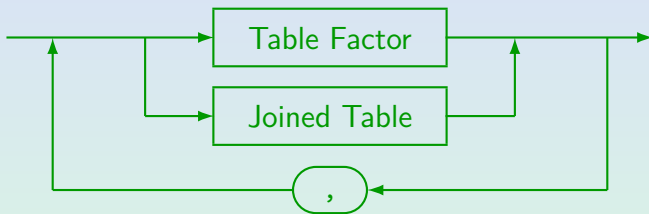
Da man ein Komma für den **CROSS JOIN** schreiben kann, ist dies auch nicht sehr sinnvoll.

- **UNION JOIN** unterstützt keins der fünf Systeme.

Man kann aber in SQL-92 (und z.B. Oracle, DB2, SQL Server, nicht Access) eine Unteranfrage unter **FROM** schreiben, die **UNION** oder **UNION ALL** enthält. Mit etwas mehr Aufwand kann also der Union Join simuliert werden. Nebenbei bemerkt, ist es etwas seltsam, dass z.B. "**FROM A NATURAL JOIN B**" in SQL-92 zulässig ist, aber "**FROM A UNION B**" nicht. Auch lässt SQL-92 die Schreibweise "**FROM (SELECT * FROM A UNION SELECT * FROM B) X**" zu, aber das gleiche mit "**NATURAL JOIN**" statt "**UNION**" liefert einen Syntaxfehler [Date/Darwen, 1997, S. 148].

Verbundsyntax: Formal (1)

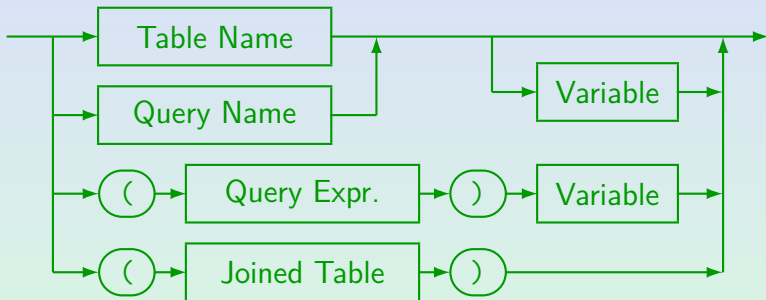
Quell-Liste (nach FROM):



- Die Syntaxgraphen sind eine vereinfachte Version der Grammatik aus dem SQL-2008 Standard. Deswegen sind die Namen der Graphen (syntaktische Kategorien, Nichtterminalsymbole) hier in Englisch. Es wurden allerdings viele kompliziertere Konstrukte weggelassen.
- Der Standard hat eine syntaktische Kategorie "Table Reference", die (im wesentlichen) der Alternative zwischen "Table Factor" und "Joined Table" entspricht.

Verbundsyntax: Formal (2)

Table Factor:



- “Query Name” ist mit der WITH-Klausel definiert.
- “Query Expression” ist eine Unteranfrage (inkl. ggf. WITH, UNION, ORDER BY).
- Der Standard und viele Systeme erlauben “AS” vor der Variable.
- Der Standard+einige DBMS erlauben Spalten-Umbenennung: $R\ X(A, B)$.

Verbundsyntax: Formal (3)

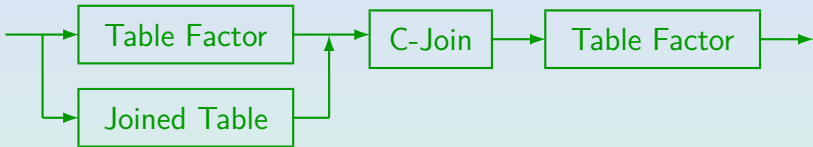
Joined Table:



- Noch zum “Table Factor”: Man beachte, dass Tupelvariablen bei Tabellennamen optional sind, bei Unteranfragen nötig, und bei geklammerten Join-Ausdrücken nicht erlaubt sind (bei einigen Systemen geht es doch, aber es ist nicht portabel). Tupelvariablen können natürlich für die einzelnen am Join beteiligten Tabellen eingeführt werden.
- “Qualified Join” ist mit ON oder USING, s.u.

Verbindsyntax: Formal (4)

Cross Join:

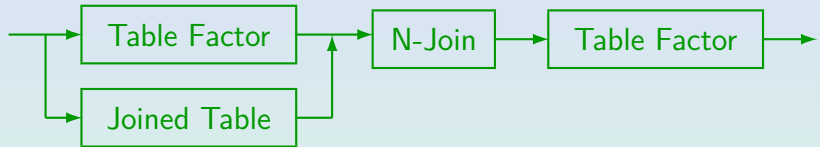


C-Join:



Verbindsyntax: Formal (5)

Natural Join:

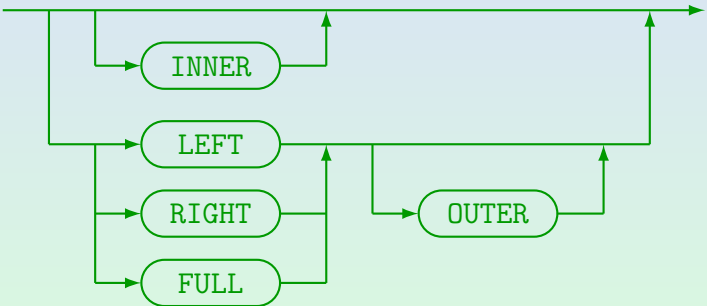


N-Join:



Verbandsyntax: Formal (6)

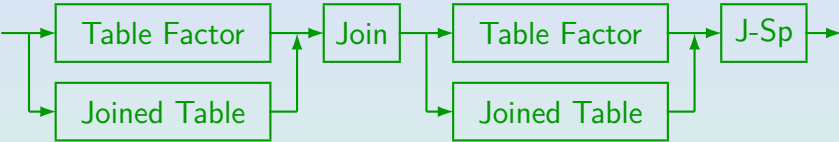
Join Type:



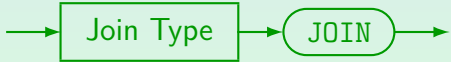
- Der Verbund-Typ ist optional. Wenn kein Typ angegeben ist, wird ein normaler ("innerer") Verbund genommen.

Verbindsyntax: Formal (7)

Qualified Join:



Join:

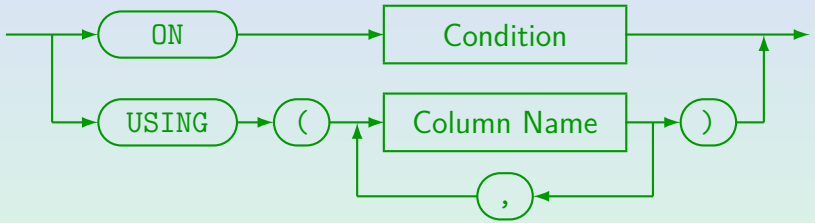


J-Sp ("Join Specification"):

- ON- oder USING-Klausel, siehe nächste Folie.

Verbandsyntax: Formal (8)

J-Sp ("Join Specification"):



Äußerer Verbund in Oracle (1)

- In Oracle wird der äußere Verbund traditionell unter WHERE spezifiziert (nicht länger nötig ab Version 9i).
- Statt der Bedingung $R.A = S.B$ schreibt man
 - $R.A = S.B(+)$ für den linken äußeren Verbund
 - $R.A(+) = S.B$ für den rechten äußeren Verbund

D.h. das Zeichen “(+)” wird an die Attribute angehängt, die durch Null ersetzt werden können.

D.h. dies erhält die Tupel der anderen Tabelle (die nicht mit “(+)” markiert sind). Viele syntaktische Restriktionen sichern, dass dies wirklich ein äußerer Verbund ist. Wird der Verbund über mehrere Attribute durchgeführt, müssen alle markiert werden. Man kann auch $S.B(+) = c$ mit einer Konstante c schreiben, oder z.B. $R.A = S.B(+) + 1$.

Äußerer Verbund in Oracle (2)

- Z.B. Anzahl der Abgaben pro HA (kann 0 sein):

```
SELECT  A.ATYP, A.ANR, COUNT(SID)
FROM    AUFGABEN A, BEWERTUNGEN B
WHERE   A.ATYP = B.ATYP(+) AND A.ANR = B.ANR(+)
GROUP BY A.ATYP, A.ANR
```

- Wie im äußeren Verbund von SQL-92, wird der äußere Verbund durchgeführt, bevor irgendeine andere Bedingung der WHERE-Klausel angewandt wird.

Egal in welcher Reihenfolge die Bedingungen stehen. Aber wie oben gezeigt, kann man eine Unteranfrage unter FROM machen, um vor dem äußeren Verbund zu selektieren.