

Boyce-Codd-Normalform

Unser Entwurf der BERATER-Relation wirft Probleme bei Aktualisierungen auf.

Definition: Ein *Kandidatenschlüssel* ist ein Attribut oder eine Menge von Attributen, der als Primärschlüssel verwendet werden kann. Der Kandidatenschlüssel kennzeichnet die Zeile eindeutig und ist bezüglich dieser Eigenschaft minimal.

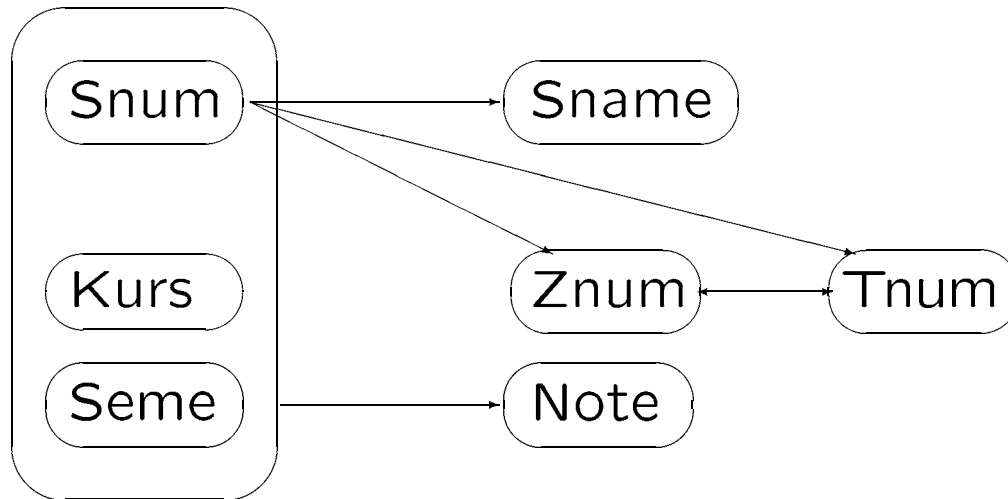
Definition: Gegeben sei eine funktionale Abhängigkeit (FA)

$$A_1, \dots, A_n \longrightarrow B$$

mit A_1, \dots, A_n minimal.

Dann heißt A_1, \dots, A_n eine *Determinante*.

Beim BERATER



gibt es nur einen Kandidatenschlüssel $\langle \text{Snum}, \text{Kurs}, \text{Seme} \rangle$. (Welche Mindestmengen an Attributen bestimmen alle anderen?)

Die Determinanten sind: $\langle \text{Snum}, \text{Kurs}, \text{Seme} \rangle$, $\langle \text{Snum} \rangle$, $\langle \text{Znum} \rangle$, $\langle \text{Tnum} \rangle$. Beachte: bei gegenseitigen funktionalen Abhängigkeiten gibt es 2 Determinanten.

Definition: Eine Relation befindet sich in *Boyce-Codd-Normalform* (*BCNF*), wenn jede Determinante ein Kandidatenschlüssel ist.

Bemerkung: Es gibt mehrere Normalformen. In dieser Vorlesung befassen wir uns mit 1NF und BCNF.

Bei der allgemeinen Relation BERATER gibt es

Kandidatenschlüssel	Determinanten
$\langle \text{Snum}, \text{Kurs}, \text{Seme} \rangle$	$\langle \text{Snum}, \text{Kurs}, \text{Seme} \rangle$
	$\langle \text{Snum} \rangle$
	$\langle \text{Tnum} \rangle$
	$\langle \text{Znum} \rangle$

Die allgemeine Relation BERATER befindet sich **nicht** in BCNF.

Die allgemeine Vorgehensweise bei der Zerlegung einer Relation

1. Entwicklung der allgemeinen Relation für die Datenbank.
2. Feststellung aller FAs.
3. Feststellung, ob die Relation in BCNF ist. Wenn ja, ist man fertig; wenn nein, zerlegen.
4. Die 2 vorigen Schritte für jede neue Relation wiederholen.

Vorgehensweise bei einer Zerlegung:

Ist $R(A, B, C, D, E, \dots)$ nicht in BCNF, sucht man eine FA $C \rightarrow D$, aufgrund derer R nicht in BCNF ist, d. h. wo C eine Determinante aber kein Kandidatenschlüssel ist.

Man bildet die zwei neuen Relationen

$$R1(A, B, C, \hat{D}, E, \dots) \text{ und } R2(C, D)$$

wobei der abhängige Teil der FA $C \rightarrow D$ aus R herausgenommen wird. Dieses ergibt $R1$. $R2$ ergibt sich aus den Attributen der FA $C \rightarrow D$.

Definition: $R2(C, D)$ heißt die *Projektion* aus R .

In BERATER gibt es die Determinanten Snum, Tnum, und Znum, die keine Kandidatenschlüssel sind.

Die allgemeine Relation:

BERATER(Snum, Kurs, Seme, Sname, Znum, Tnum, Note)

$Snum \rightarrow Znum$; $Snum \rightarrow Tnum$; $Znum \rightarrow Tnum$; $Tnum \rightarrow Znum$

Je nachdem mit welcher FA man anfängt, entstehen verschiedene Entwürfe. Man muss dann alle Entwürfe prüfen, um festzustellen, welcher für die Applikation am geeignetsten ist.

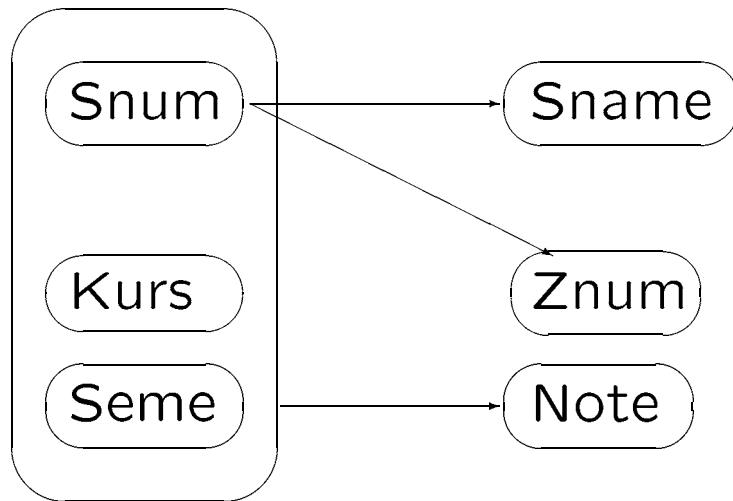
Faustregel: Suchen Sie eine Kette der Form

$$A \longrightarrow B \longrightarrow C$$

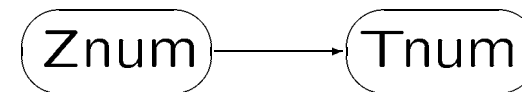
und projizieren Sie das rechte Ende.

Beim BERATER: $\text{Snum} \longrightarrow \text{Znum} \longrightarrow \text{Tnum}$

R_1



R_2



$R_2(\underline{Znum}, Tnum)$ (ist in BCNF)

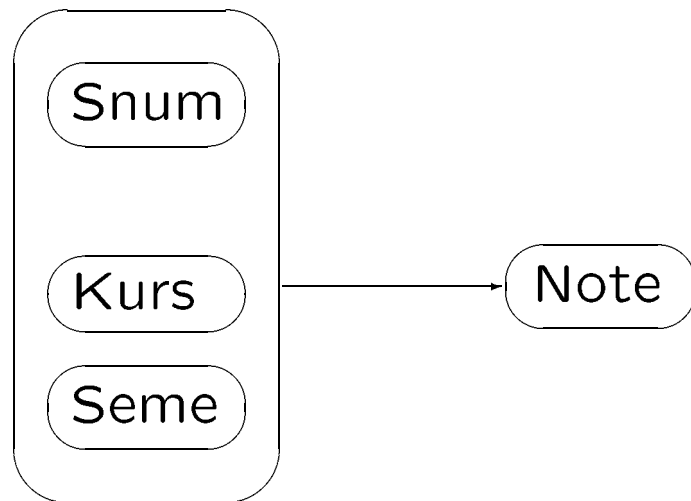
$R_1(\underline{Snum}, Kurs, Seme, Sname, Znum, Note)$

Jetzt haben wir die FAs: $Snum \rightarrow Sname$ und $Snum \rightarrow Znum$.
Dieses können wir als zusammengesetztes Attribut betrachten

$Snum \rightarrow Sname, Znum$

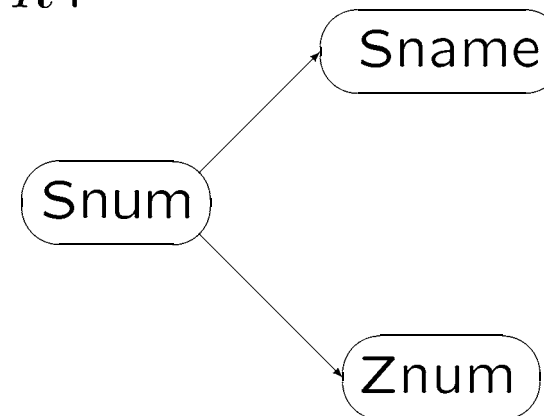
Projizieren ergibt

$R3$



$R3(\underline{Snum}, Kurs, Seme, Note)$

$R4$



$R4(\underline{Snum}, Sname, Znum)$

$R2(\underline{Znum}, Tnum)$

$R3(\underline{Snum}, \underline{Kurs}, \underline{Seme}, Note)$

$R4(\underline{Snum}, Sname, Znum)$

$R3$

Snum	Kurs	Seme	Note
3215	MAT122	W88	1,4
3215	PHY120	S88	2,5
3215	WUR330	W88	3,1
3215	MAT122	S87	1,4
3462	MAT122	W87	2,5
3462	MAT130	S87	1,5
3462	PSY220	W88	3,6
3567	PHY230	S87	3,0
3567	WER220	S88	2,5
3567	PHY141	W88	1,2
4756	MUS389	W88	4,0

$R2$

Znum	Tnum
120DH	2136
238VH	2344
345VH	3321

$R4$

Snum	Sname	Znum
3215	G. Jonas	120DH
3462	A. Schmid	238VH
3567	J. Neider	120DH
4756	K. Alex	345VH

Rückblick auf die Anomalien

Einfügen Neuer Student ohne Kurs. Dieses führt zu einer neuen Zeile in $R4(\underline{Snum}, Sname, Znum)$.

Aktualisieren Ändern einer Telefonnummer. Bei der allgemeinen Relation kam dieselbe Telefonnummer mehrmals vor.

Löschen Bei der allgemeinen Relation verschwanden Daten beim Löschen des letzten Studenten aus einem Kurs. Bei $R3$ bestehen diese Daten weiter.

Leider gibt es jetzt drei Relationen statt nur eine. Dieses führt dazu, dass die SQL-Befehle komplizierter werden, um auf die Daten zuzugreifen.

```
CREATE TABLE SNOTE (SNUM CHAR(4),  
                    KURS CHAR(6),  
                    SEME CHAR(3),  
                    NOTE NUMBER(1,1),  
                    PRIMARY KEY (SNUM, KURS, SEME));
```

```
CREATE TABLE ZT (ZNUM CHAR(5) PRIMARY KEY,  
                 TNUM CHAR(4));
```

```
CREATE TABLE SNAMADR (SNUM CHAR(4) PRIMARY KEY,  
                      SNAME CHAR(25),  
                      ZNUM CHAR(5));
```

Alle Noten für den Studenten 3462?

```
SELECT NOTE FROM SNOTE WHERE SNUM=3462;
```

Telefonnummer des Studenten 3567?

```
SELECT TNUM FROM ZT, SNAMADR WHERE SNUM=3567  
AND SNAMADR.ZNUM=ZT.ZNUM;
```

Eine andere Zerlegung von BERATER

In dem oben aufgeführten Datenmodell projizierten wir mit $Znum \rightarrow Tnum$ als letztes Glied in der Kette $Snum \rightarrow Znum \rightarrow Tnum$.

Möglich wäre auch die Kette $Snum \rightarrow Tnum \rightarrow Znum$. Dieses ergibt eine Zerlegung:

$R2(\underline{Tnum}, Znum)$

$R3(\underline{Snum}, \underline{Kurs}, \underline{Seme}, \underline{Note})$

$R4(\underline{Snum}, \underline{Sname}, Tnum)$

Dieses Datenmodell ist genau so gültig wie das erste. Welches für die Applikation besser eignet ist, hängt davon ab wie die Datenbank benutzt wird bzw. welches Modell der Auftraggeber bevorzugt.

Allgemeine Bemerkungen zu der Vorgehensweise beim Zerlegen

1. Benutzt man statt dem rechten Ende einer Kette $A \longrightarrow B \longrightarrow C$ das linke Ende, würde man aus der allgemeinen Relation $R(\underline{A}, B, C)$ die Relationen $R1(\underline{A}, C)$ und $R2(\underline{A}, B)$ erhalten. In diesem Datenmodell ist die FA $B \longrightarrow C$ verlorengegangen. Beim Eingeben von Daten können sich *falsche Beziehungen* zwischen Werten für B und C einschleichen.

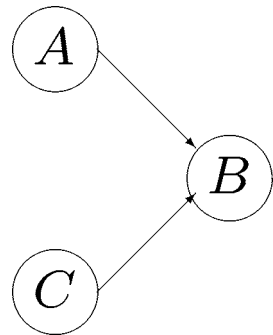
$R1$

A	C
9	4
8	3

$R2$

A	B
9	2
8	2

2. Spezialfall: Ein Attribut ist von zwei verschiedenen Determinanten abhängig.



$R(\underline{A}, B, \underline{C})$

Diese Relation steht nicht in BCNF: $\langle A, C \rangle$ ist der einzige Kandidatenschlüssel, und es gibt Determinanten $\langle A \rangle$ und $\langle C \rangle$.

Würde man projizieren, so würde eine FA verloren gehen, z. B. $R1(\underline{A}, \underline{C})$ und $R2(\underline{A}, B)$.

In diesem Fall sollte man $R(\underline{A}, B, \underline{C})$ in

$$R1(\underline{A}, B) \quad \text{und} \quad R2(\underline{C}, B)$$

zerlegen. Diese Vorgehensweise heißt die *Synthesemethode* und weicht von der Standardzerlegung ab.

Die Synthesemethode geht davon aus, dass alle FAs mit gleichen Determinanten in Gruppen eingeteilt werden. Diese werden zu neuen Relationen.