Datenbanken

Normalformentheorie

<u>Anomalien</u>

- **Einfüge-Anomalie**
  Es soll ein neuer Kurs eingetragen werden : KNR = 24, Kurs = DBV, Kurspreis = 5.000
  => Dies geht jedoch erst, wenn er auch einen Teilnehmer hat.
- **Änderungs-Anomalie**
  Der Teilnehmer Müller (TNR = 2100) zieht nach Bochum um.
  => Alle Datensätze, die die TNR 2100 enthalten, müssen geändert werden.
- **Lösch-Anomalie**
  Der C-Kurs (KNR = 17) fällt aus.
  => Wenn Krüger an keinem weiteren Kurs teilnimmt, verschwinden seine Teilnehmerinformationen.

Kurse

TNR	Name	Adresse	KNR	Kurs	Kurspreis	Preis
2100	Müller	Essen	17	C	3.000	2.700
2100	Müller	Essen	18	C++	3.500	3.150
2101	Müller	München	23	DB	5.000	4.500
2103	Hansen	Hamburg	18	C++	3.500	2.800
2103	Hansen	Hamburg	23	DB	5.000	4.000
2544	Schmidt	Frankfurt	23	DB	5.000	5.000
2544	Schmidt	Frankfurt	24	Java	3.500	3.500
2378	Krüger	München	17	C	3.000	3.000
...	...	...	...	...	...	...

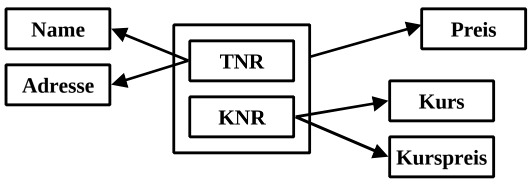
<u>Funktionale Abhängigkeiten</u>

Gegeben sei eine Relation R(A1, A2, ..., An) und die Menge der Attribute A:= {A1, A2, ..., An}.

Funktionale Abhängigkeit (FD : Functional Dependency)
Eine Attributkombination C (C $\subseteq$ A) heißt <u>funktional abhängig</u> von einer Attributkombination B (B $\subseteq$ A), wenn in jedem möglichen Tupel von R die Ausprägung von C durch die Ausprägung von B eindeutig bestimmt ist : <u>$B \rightarrow C$</u>.

Man sagt auch : B bestimmt C funktional (=eindeutig). oder: B ist **Determinante** von C.
<u>$B \not\rightarrow C$</u> : C ist nicht funktional abhängig von B

Beispiel : TNR $\rightarrow$ Name / TNR $\rightarrow$ Adresse / TNR $\rightarrow$ Name,Adresse / Name,Adresse $\rightarrow$ TNR
TNR,KNR $\rightarrow$ Preis / TNR,KNR $\rightarrow$ Name,Adresse,Kurs,Preis / TNR $\not\rightarrow$ Kurs



Superschlüssel (Superkey, Identifikator)
Eine Attributkombination D (D $\subseteq$ A) ist ein Superschlüssel von R, wenn die Menge aller Attribute A funktional abhängig von D ist : <u>$D \rightarrow A$</u>.
Beispiel : TNR,KNR $\rightarrow$ Name,Adresse,Kurs,Kurspreis,Preis
TNR,KNR,Name $\rightarrow$ Adresse,Kurs,Kurspreis,Preis

Volle funktionale Abhängigkeit
Eine Attributkombination C (C $\subseteq$ A) heißt <u>voll funktional abhängig</u> von einer Attributkombination B (B $\subseteq$ A), wenn C funktional abhängig ist von B und es keine Teilmenge von B gibt, von der C funktional abhängig ist : <u>$B \overset{\bullet}{\rightarrow} C$</u>.

Beispiel : TNR,KNR $\rightarrow$ Name,Adresse => TNR $\overset{\bullet}{\rightarrow}$ Name,Adresse
TNR,KNR $\rightarrow$ Kurs,Kurspreis => KNR $\overset{\bullet}{\rightarrow}$ Kurs,Kurspreis
TNR,KNR $\rightarrow$ Preis => TNR,KNR $\overset{\bullet}{\rightarrow}$ Preis

Schlüsselkandidat
Ein Schlüsselkandidat ist ein Superschlüssel D (D $\subseteq$ A) von R, von dem die Menge aller Attribute A voll funktional abhängig ist : $\underline{D \overset{\bullet}{\rightarrow} A}$.
Für die Relation R wird ein Schlüsselkandidat zum **Primärschlüssel** gewählt.

Beispiel : TNR,KNR $\overset{\bullet}{\rightarrow}$ Name,Adresse,Kurs,Kurspreis,Preis

Herleitungsregeln (Inferenzregeln)
- <u>Armstrong-Axiome</u>
  - <u>Reflexivität</u> : C $\subseteq$ B => B $\rightarrow$ C (triviale funktionale Abhängigkeit)
    Bsp. Name,Adresse $\rightarrow$ Name
  - <u>Verstärkung</u> : B $\rightarrow$ C => B,D $\rightarrow$ C,D
    Bsp. TNR $\rightarrow$ Name => TNR,Kurs $\rightarrow$ Name,Kurs
  - <u>Transitivität</u> : B $\rightarrow$ C $\wedge$ C $\rightarrow$ D => B $\rightarrow$ D
    Bsp. KNR$\rightarrow$ Kurs $\wedge$ Kurs$\rightarrow$ Kurspreis => KNR$\rightarrow$ Kurspreis
- weitere Herleitungsregeln
  - <u>Vereinigungsregel</u> : B $\rightarrow$ C $\wedge$ B $\rightarrow$ D => B $\rightarrow$ C,D
    Bsp. TNR$\rightarrow$Name $\wedge$ TNR$\rightarrow$Adresse=>TNR$\rightarrow$Name,Adresse
  - <u>Dekompositionsregel</u> : B $\rightarrow$ C,D => B $\rightarrow$ C $\wedge$ B $\rightarrow$ D
    Bsp. TNR$\rightarrow$Name,Adresse=>TNR$\rightarrow$Name $\wedge$ TNR$\rightarrow$Adresse
  - <u>Pseudotransitivitätsregel</u> : B $\rightarrow$ C $\wedge$ C,D $\rightarrow$ E => B,D $\rightarrow$ E
    Bsp. Kurs$\rightarrow$KNR $\wedge$ KNR,TNR$\rightarrow$Preis => Kurs,TNR$\rightarrow$Preis

<u>Normalisierung</u>

Normalisierung
Unter der Normalisierung versteht man die Zerlegung von Relationen in normalisierte Relationen, mit der Bedingung, dass alle enthaltenen Informationen (Verlustlosigkeit) alle funktionalen Abhängigkeiten erhalten bleiben (Abhängigkeitserhaltung).

Projektion
Eine Projektion P einer Relation R ist eine Relation, deren Attribute eine Teilmenge der Attribute von R darstellen. (Eine Projektion ist eine Relation.)

Beispiel: Kursteilnehmer (TNR,Name,Adresse), Kurspreise (TNR,KNR,Kurspreis,Preis)

Zerlegung
Eine Relation R ist in die Projektionen R1, R2, ..., Rn zerlegt, wenn die Menge der Attribute aller Projektionen der Menge der Attribute von R entspricht.

Beispiel : Teilnehmer (TNR,Name,Adresse) und Kurse (KNR,Kurs,Kurspreis,Preis)

Verlustlosigkeit (verlustfreie Zerlegung)
Eine Zerlegung einer Relation R in die Projektionen R1, R2,...,Rn ist verlustfrei, wenn R einem Verbund der Projektionen entspricht.

Beispiel : Kursteilnehmer (TNR,Name,Adresse)
Kurse (KNR,Kurs,Kurspreis)
Kurspreise (KNR,TNR,Preis)

```
select   kt.TNR, kt.Name, kt.Adresse, k.KNR, k.Kurs, k.Kurspreis,  kp.Preis
from     Kursteilnehmer kt, Kurse k, Kurspreis kp
where    kt.tnr = kp.tnr and kp.knr = k.knr ;
```

Heath-Theorem
Eine Relation R(A,B,C) (A, B, C repräsentieren Attributmengen von R) kann in die Projektionen R1(A,B) und R2(A,C) verlustfrei zerlegt werden, wenn $A \rightarrow B$ gilt.

Beispiel :	Kursteilnehmer	(TNR,Name,Adresse)
	Kurse	(TNR,KNR,Kurs,Kurspreis,Preis)
	select	kt. TNR,kt.Name,kt.Adresse,k.KNR,Kurs,k.Kurspreis,k.Preis
	from	Kursteilnehmer kt, Kurs k
	where	kt.TNR = k.TNR ;

Abhängigkeitserhaltung
Eine Zerlegung einer Relation R in die Projektionen R1, R2,...,Rn ist abhängigkeitserhaltend, wenn alle funktionalen Abhängigkeiten in R in den Projektionen überprüft werden können.
(D.h., dass eine Projektion unabhängig von den anderen Projektionen verändert werden kann)

Beispiel : Bsp. zu Zerlegung ist nicht abhängigkeitserhaltend : TNR,KNR $\rightarrow$ Preis

Rissanen-Thoerem
2 Projektionen R1 und R2 von R sind unabhängig dann und nur dann, wenn gilt :
- Jede funktionale Abhängigkeit in R kann aus R1 und R2 abgeleitet werden.
- Die gemeinsamen Attribute von R1 und R2 bilden einen Schlüssel für mindestens eine der beiden Relationen.

Beispiel : Name und Adresse können nicht aufgeteilt werden : Name,Adresse $\rightarrow$ TNR

Übung **Lieferung**

SNR	Status	Ort	PNR	Menge
S1	20	München	P1	300
S1	20	München	P2	200
S1	20	München	P3	400
S1	20	München	P4	200
S1	20	München	P5	100
S1	20	München	P6	100
S2	10	Hamburg	P1	300
S2	10	Hamburg	P2	400
S3	10	Hamburg	P2	200
S4	20	Frankfurt	P2	200
S4	20	Frankfurt	P4	300
S4	20	Frankfurt	P5	400

Normalformen

1.Normalform (1NF)
Eine Relation R ist in 1NF, wenn alle ihre Domänen nur einfache (elementare, atomare) Werte enthalten.

Kurse

TNR	Name	Adresse	Preis
2100	Müller	Essen	{2.700,3.1500}
2101	Müller	München	{4.500}
2103	Hansen	Hamburg	{2.800,4.000}
2544	Schmidt	Frankfurt	{5.000,3.500}
2378	Krüger	München	{3.000}

NF^2 : non-first normal form
In NF^2-Modellen gibt es mengen- und relationenwertige Attribute. Ein NF^2-Modell wird auch geschachteltes relationales Modell (nested relational model) genannt.

2.Normalform (2NF)

Eine Relation R ist in 2NF genau dann, wenn sie in 1NF ist und jedes nicht-Schlüssel-Attribut voll funktional abhängig von einem Schlüsselkandidaten ist.

Kursteilnehmer

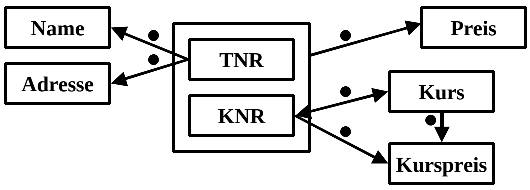
TNR	Name	Adresse
2100	Müller	Essen
2101	Müller	Hamburg
2103	Hansen	München
2544	Schmidt	Frankfurt
2378	Krüger	München

Kurse

KNR	Kurs	Kurspreis
17	C	3.000
18	C++	3.500
23	DB	5.000
24	Java	3.500

Kurspreise

TNR	KNR	Preis
2100	17	2.700
2100	18	3.150
2100	23	4.500
2103	18	2.800
2103	23	4.000
2544	23	5.000
2544	24	3.500
2378	17	3.000



3.Normalform (3NF)

Eine Relation ist dann und nur dann in 3NF, wenn sie in 2NF ist und jedes nicht-Schlüssel-Attribut nicht transitiv abhängig ist von einem Schlüsselkandidaten. Sie ist auch dann in 3NF, wenn sich eine transitive Abhängigkeit nur über Schlüsselkandidaten herleitet.

Kursteilnehmer

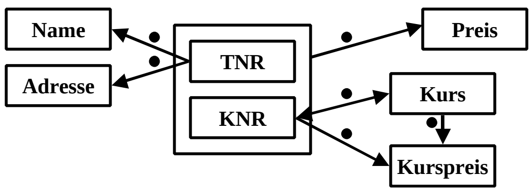
TNR	Name	Adresse
2100	Müller	Essen
2101	Müller	Hamburg
2103	Hansen	München
2544	Schmidt	Frankfurt
2378	Krüger	München

Kurse

KNR	Kurs	Kurspreis
17	C	3.000
18	C++	3.500
23	DB	5.000
24	Java	3.500

Kurspreise

TNR	KNR	Preis
2100	17	2.700
2100	18	3.150
2100	23	4.500
2103	18	2.800
2103	23	4.000
2544	23	5.000
2544	24	3.500
2378	17	3.000



KNR ® Kurs ® Kurspreis

Kurspreis ist transitiv abhängig von KNR
aber :
Kurs ist Schlüsselkandidat

Kursteilnehmer

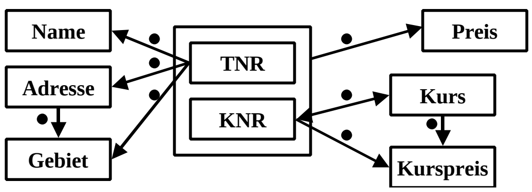
TNR	Name	Adresse	Gebiet
2100	Müller	Essen	NRW
2101	Müller	Hamburg	HH
2103	Hansen	München	BAY
2544	Schmidt	Frankfurt	HES
2378	Krüger	München	BAY

Kurse

KNR	Kurs	Kurspreis
17	C	3.000
18	C++	3.500
23	DB	5.000
24	Java	3.500

Kurspreise

TNR	KNR	Preis
2100	17	2.700
2100	18	3.150
2100	23	4.500
2103	18	2.800
2103	23	4.000
2544	23	5.000
2544	24	3.500
2378	17	3.000



TNR ® Adresse ® Gebiet

Gebiet ist transitiv abhängig von TNR

=>

Kursteilnehmer

TNR	Name	Adresse
2100	Müller	Essen
2101	Müller	Hamburg
2103	Hansen	München
2544	Schmidt	Frankfurt
2378	Krüger	München

Gebiet

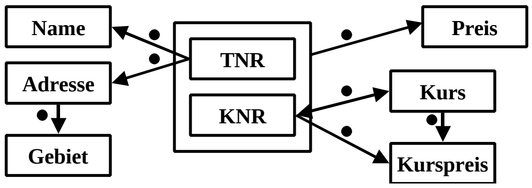
Adresse	Gebiet
Essen	NRW
Hamburg	HH
München	BAY
Frankfurt	HES

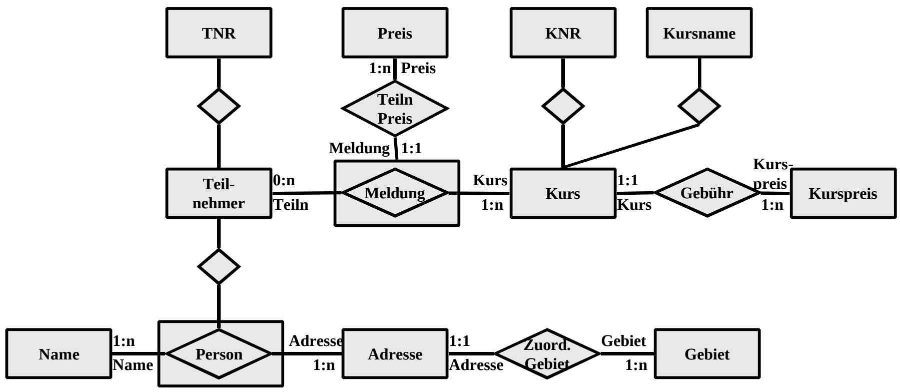
Kurse

KNR	Kurs	Kurspreis
17	C	3.000
18	C++	3.500
23	DB	5.000
24	Java	3.500

Kurspreise

TNR	KNR	Preis
2100	17	2.700
2100	18	3.150
2100	23	4.500
2103	18	2.800
2103	23	4.000
2544	23	5.000
2544	24	3.500
2378	17	3.000





Boyce-Codd-Normalform (BCNF)

Eine Relation R ist in BNCF, wenn jede Determinante von R auch Schlüsselkandidat von R ist.

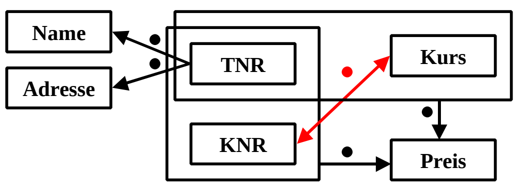
- BCNF fordert mehr als 3NF
- BNCF basiert nicht auf 1NF und 2NF
- Ist eine Relation in der 3NF und ist der Schlüsselkandidat die einzige Determinante, so ist die Relation in BCNF.

Teilnehmer

TNR	Name	Adresse
2100	Müller	Essen
2101	Müller	München
2103	Hansen	Hamburg
2544	Schmidt	Frankfurt
2378	Krüger	München

Kurse

TNR	KNR	Kurs	Preis
2100	17	C	2.700
2100	18	C++	3.150
2101	23	DB	4.500
2103	18	C++	2.800
2103	23	DB	4.000
2544	23	DB	5.000
2544	24	Java	3.500
2378	17	C	3.000



KNR ® Kurs
Kurs ® KNR

Weder KNR noch Kurs ist Schlüsselkandidat.

Teilnehmer

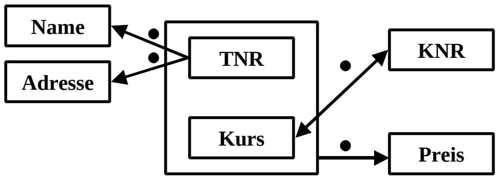
TNR	Name	Adresse
2100	Müller	Essen
2101	Müller	München
2103	Hansen	Hamburg
2544	Schmidt	Frankfurt
2378	Krüger	München

Kurse

KNR	Kurs
17	C
18	C++
23	DB
24	Java

Kurspreise

TNR	Kurs	Preis
2100	C	2.700
2100	C++	3.150
2101	DB	4.500
2103	C++	2.800
2103	DB	4.000
2544	DB	5.000
2544	Java	3.500
2378	C	3.000



Übung : Lieferung

Lieferung

SNR	Lieferant	Ort	PNR	Menge
S1	Schmitz	München	P1	300
S1	Schmitz	München	P2	200
S1	Schmitz	München	P3	400
S1	Schmitz	München	P4	200
S1	Schmitz	München	P5	100
S1	Schmitz	München	P6	100
S2	Koller	Hamburg	P1	300
S2	Koller	Hamburg	P2	400
S3	Huber	Hamburg	P2	200
S4	Krumm	Frankfurt	P2	200
S4	Krumm	Frankfurt	P4	300
S4	Krumm	Frankfurt	P5	400

Mehrwertige Abhängigkeit (MVD : Multivalued Dependency)

Seien B,C,D disjunkte Teilmengen von A deren Vereinigung A ist : A:= B ∪ C ∪ D. C heißt <u>mehrwertig abhängig</u> von B

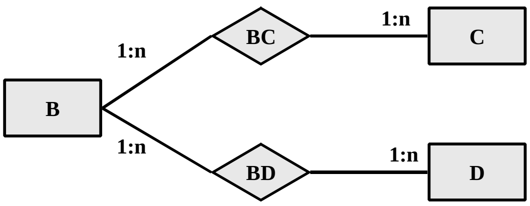
$B \rightarrow\rightarrow C$

wenn die Menge der möglichen Werte für Attribute aus C nur von B und nicht von D abhängig ist.

Eine mehrwertige Abhängigkeit heißt <u>trivial</u>, wenn C := ∅ oder C := A\B.

=>

Die Menge der möglichen Werte für Attr. aus D sind nur von B und nicht von C abh. : $B \rightarrow\rightarrow D$



B ® ® C

Die Menge der möglichen Werte für Attribute aus C ist nur von B und nicht von D abhängig ist.

B ® ® D

Die Menge der möglichen Werte für Attribute aus D ist nur von B und nicht von C abhängig ist.

BC (B,C)	BD (B,D)	R (B,C,D)	
b1 – c1	b1 – d1	b1 – c1 – d1	b2 – c1 – d1
b1 – c2	b1 – d3	b1 – c1 – d3	b2 – c1 – d3
b2 – c1	b2 – d1	b1 – c2 – d1	b2 – c3 – d1
b2 – c3	b2 – d2	b1 – c2 – d3	b2 – c3 – d3

R (B,C,D) und B ® ® C

Aus

(b,c1,d1) und (b,c2,d2)

folgt

(b,c1,d2) und (b,c2,d1)

Fagin-Theorem

Gegeben sei die Relation R (B,C,D) mit den Attributkombinationen B, C und D. Dann kann R verlustlos in die Projektionen R1 (B,C) und R2 (B,D) zerlegt werden genau dann wenn B $\rightarrow\rightarrow$ C.

Kurse

Fach	Dozent	Lehrinhalt
Physik	Müller	Mechanik
Physik	Müller	Optik
Physik	Hansen	Mechanik
Physik	Hansen	Optik
Physik	Kroll	Mechanik
Physik	Kroll	Optik
Mathe	Schmidt	Algebra
Mathe	Schmidt	Geometrie

Dozent

Fach	Dozent
Physik	Müller
Physik	Hansen
Physik	Kroll
Mathe	Schmidt

Lehrinhalt

Fach	Lehrinhalt
Physik	Mechanik
Physik	Optik
Mathe	Algebra
Mathe	Geometrie

Jedes Fach hat mehrere Dozenten.

Jedes Fach hat mehrere Lehrinhalte.

```
select   d.Fach, d.Dozent, l.Lehrinhalt
from     Dozent d, Lehrinhalt l
where    d.Fach = l.Fach ;
```

4.Normalform (4NF)

Seien B,C $\subset$ A und B $\subset$ C := $\varnothing$.
R ist dann und nur dann in 4NF, wenn gilt : Gibt es in R eine mehrwertige Abhängigkeit, z.B. B $\rightarrow\rightarrow$ C, dann handelt es sich um eine funktionale Abhängigkeit und alle Attribute von R sind von B funktional abhängig (d.h. B $\rightarrow$ X für alle Attribute X von R).

- Die einzige funktionale/mehrwertige Abhängigkeit in einer Relation hat einen Schlüsselkandidat als Determinante.
- 4NF ist eine strengere Forderung als BCNF : Jede 4NF-Relation ist auch BCNF-Relation.
- Jede Relation kann ohne Informationsverlust in einen äquivalenten Satz von 4NF-Relationen zerlegt werden.
- Das Rissanen-Theorem gilt auch für mehrwertige Abhängigkeit.

Lernnachweise

Studien-gruppe	IM	Name	Vorname	Fach	Praktikum	Klausur
I4	126793	Kramm	Irene	PM	x	-
I4	347899	Bonsen	Alois	DB	x	x
I4	347899	Bonsen	Alois	PM	x	-
WI4	216565	Meier	Hans	DB	x	x
WI4	216565	Meier	Hans	PM	x	-
WI4	216565	Meier	Hans	PA	-	-
WI4	126793	Kramm	Irene	DB	x	x
WI4	126793	Kramm	Irene	PM	x	-
WI4	789789	Heller	Andrea	PA	-	-
WI4	126793	Kramm	Irene	PA	-	-
WI4	789789	Heller	Andrea	DB	x	x
I4	126793	Kramm	Irene	DB	x	x
WI4	789789	Heller	Andrea	PM	x	-

Studiengruppe

Studien-gruppe	IM	Fach
I4	126793	DB
I4	126793	PM
I4	347899	DB
I4	347899	PM
WI4	216565	DB
WI4	216565	PM
WI4	216565	PA
WI4	126793	DB
WI4	126793	PM
WI4	126793	PA
WI4	789789	DB
WI4	789789	PM
WI4	789789	PA

Student

IM	Name	Vorname
126793	Kramm	Irene
347899	Bonsen	Alois
216565	Meier	Hans
789789	Heller	Andrea

Fach

Fach	Praktikum	Klausur
DB	x	x
PM	x	-
PA	-	-

Studienrichtung

Studien-gruppe	IM
I4	126793
I4	347899
WI4	216565
WI4	126793
WI4	789789

Studienfach

Studien-gruppe	Fach
I4	DB
I4	PM
WI4	DB
WI4	PM
WI4	PA

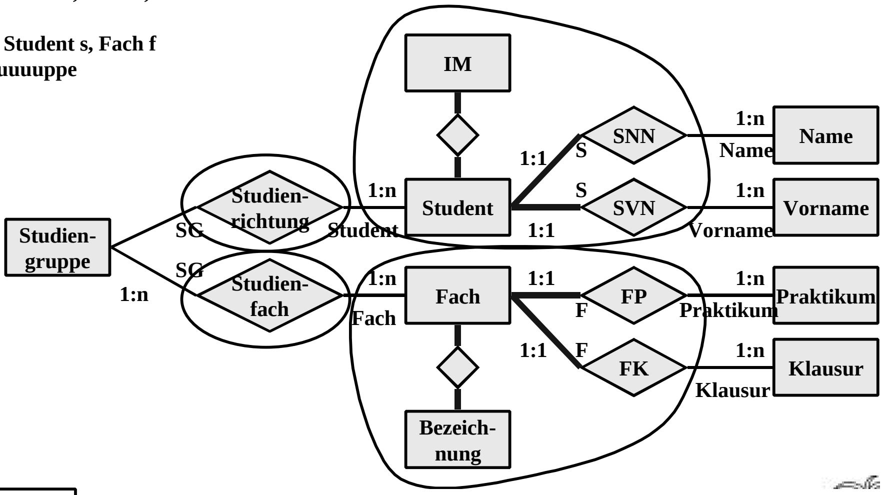
Student

IM	Name	Vorname
126793	Kramm	Irene
347899	Bonsen	Alois
216565	Meier	Hans
789789	Heller	Andrea

Fach

Fach	Praktikum	Klausur
DB	x	x
PM	x	-
PA	-	-

```
select   sr.Studiengruppe,sr.IM,s.Name,s.Vorname,sf.Fach,
         f.Praktikum,f.Klausur
from     Studienrichtung sr, Studienfach sf, Student s, Fach f
where    sr.Studiengruppe = sf.Studiengruuuuuuppe
         and
         sr.IM = s.IM
         and
         sf.Fach = f.Fach
;
```



Übung : Produktpalette

Produktpalette

Autotyp	Farbe	ReifenId	Jahreszeit	Geschw.
Audi	schwarz	M34-H	Winter	180
Toyota	weiss	A99-A	Allwetter	180
BMW	rot	M34-H	Winter	180
Audi	silber	C32-S	Sommer	250
BMW	rot	A99-9	Allwetter	180
BMW	schwarz	M34-H	Winter	180
Audi	schwarz	C32-S	Sommer	250
BMW	schwarz	C32-S	Sommer	250

Verbundabhängigkeit (JD : Join Dependency)

Seien B,C,D disjunkte Teilmengen von A deren Vereinigung A ist : A:= B $\cup$ C $\cup$ D. In R (B,C,D) existiert Verbundabhängigkeit, wenn R dem Verbund der Projektionen R1 (B,C), R2 (C,D) und R3 (D,B) entspricht.
Äquivalent : Wenn (bi,cj) $\in$ BC und (cj,dk) $\in$ CD und (dk,bi) $\in$ DB dann muss (bi,cj,dk) $\in$ BCD.

BC (B,C)	CD (C,D)	DB (D,B)	R (B,C,D)
b1 – c1	c1 – d1	d1 – b1	b1 – c1 – d1
b1 – c2	c1 – d3	d3 – b1	b1 – c1 – d3
b2 – c3	c2 – d1	d2 – b2	b1 – c2 – d1
	c3 – d2		b2 – c3 – d2

```
select   bc.b, cd.c, db.d
from     bc, cd, db
where    bc.c=cd.c and cd.d=db.d
         and  db.b=bc.b
;
```

5.Normalform (5NF)

Eine Relation ist in 5NF, wenn sie in 4NF ist und keine Verbundabhängigkeit existiert.

Vorsicht bei Beseitigung der Verbundabhängigkeiten : Es muss stets der vollständige Satz der Projektion betrachtet werden.

Normalisierungsprozess

1. Führe die unnormalisierten Relationen in normalisierte Relationen über.
2. Bilde Projektionen auf diese Relationen, so dass alle funktionalen Abhängigkeiten eliminiert werden, in denen die Determinante nicht Schlüssel ist. => BCNF
3. Bilde Projektionen dieser BCNF-Relationen, um alle mehrwertigen Abhängigkeiten zu eliminieren, die nicht gleichzeitig funktionale Abhängigkeiten sind. => 4NF
4. Bilde Projektionen dieser 4NF-Relationen, um alle Verbundabhängigkeiten zu eliminieren. => 5NF