Normalisierung in Datenbanken

Ziel: **Redundanzen** (Wiederholungen von Daten-Einträgen) vermeiden, denn dies führt zu Daten-**Inkonsistenzen** (in sich nicht stimmige Einträge).

Vorgehen: Man teilt die Tabellenstruktur in mehrere Tabellen auf, gemäß bestimmter Normalisierungsregeln. Im Allgemeinen sind drei Normalisierungsregeln gebräuchlich (I, II, III). Sie stellen das „Ausmaß“ dar, nach dem eine Datenbank gegen Anomalien ((schlechte) „Besonderheiten“) und Redundanzen geschützt ist.

Wichtig: Beim Anwenden der Normalisierungsregeln dürfen **keine** Daten verloren gehen!

Eine Beispieltabelle zu den Normalisierungsregeln (aber noch unbearbeitet!) sei nun die folgende:

**Tabelle Prüfungen**

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Sch.ID | Name | Geb.Dat. | Adresse | FB-ID | FB-Name | FB-Lehrer | Prfg.Nr. | Note |
| 0001 | Max Müller | 24.02.1994 | Seestr. 2, 83052 Bruckmühl | 07 | Physik | Obermaier | 07/21 | 2,7 |

Anm.: Sch.ID = Schüler-ID; FB-ID = Fachbereichs-ID etc.; Prfg.Nr. = Prüfungs-Nummer

Diese Tabelle befindet sich noch in **keiner** Normalform. Hier muss unter Anderem der Name des Schülers für jedes Prüfungsfach wiederholt eingetragen werden! Abgesehen von dem unnötigen Aufwand können hier auch Anomalien z.B. durch Tippfehler entstehen!

I. Normalform (I.NF)

Definition: Eine Relation (Tabelle oder Verbund von Tabellen) ist in **I. Normalform**, wenn alle Attribute (Spalten) **elementare** (auch genannt: **atomare**) Daten beinhalten. Sprich: Die Daten sind nicht weiter „zerstückelbar“.

Übergang von obiger Tabelle „Prüfungen“ in die I.NF:

**Tabelle Prüfungen**

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Sch.ID | Name | Vorname | Geb. Dat. | Str./  Nr. | PLZ | Ort | FB-ID | FB-Name | FB-Lehrer | Prf. Nr. | Note |
| 0001 | Müller | Max | 24.02.1994 | See-str 2 | 83052 | Bruck-mühl | 07 | Physik | Ober-maier | 07/21 | 2,7 |

Geänderte Spalten sind orange hervorgehoben

Allgemeines Vorgehen zur Umwandlung in die I.NF ist also:

* Identifiziere die Attribute (Spalten), wo nicht-elementare/nicht-atomare Dateneinträge vorhanden sind
* Teile diese in weitere Spalten auf, so dass nur noch elementare Daten in jeder Spalte übrig bleiben

Noch vorhandene Nachteile in dieser Tabelle (in der I.NF):

* Redundanz: z.B. wird die Adresse des selben Schülers **pro Prüfung** gespeichert
* Änderungsanomalie: Falls sich z.B. die Adresse eines Schülers ändert, muss diese Änderung für jede Prüfung extra vorgenommen werden
* Einfügeanomalie: Ein Schüler wird nur erfasst, wenn er bereits eine Prüfung absolviert hat
* Löschanomalie: z.B. geht Information (auch) über den Fachbereich (FB) verloren, falls dort (noch) keine Prüfung vorliegt

II. Normalform

Definition: Eine Relation ist in zweiter Normalform (II.NF), wenn sie **schon in I.NF ist** und jedes Nichtschlüsselelement von jedem Schlüssel **voll** funktional abhängig ist (sprich: Man kann vom Schlüssel eindeutig auf die entsprechenden Nichtschlüssel-Spalten schließen).  
Im Beispiel sind die Spalten **Name, Vorname, Geb.Dat., Str./Nr., PLZ, Ort** und, wenn derjenige nur in einem Fach geprüft wurde, auch **FB-ID, FB-Name, FB-Lehrer**, eindeutig durch die Spalte **Sch.ID** bestimmt. Jedoch gibt es Probleme mit der Noten-Spalte: Dort braucht man einen zusammengesetzten Schlüssel aus **Sch.ID** und **Prfg.Nr.**!

**Tabelle Prüfungen**

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Sch.ID | Name | Vorname | Geb. Dat. | Str./  Nr. | PLZ | Ort | FB-ID | FB-Name | FB-Lehrer | Prf. Nr. | Note |
| 0001 | Müller | Max | 24.02.1994 | See-str 2 | 83052 | Bruck-mühl | 07 | Physik | Ober-maier | 07/21 | 2,7 |

Erklärung: Orangene Spalten hängen von der roten Primärschlüssel-Spalte alleine ab; die gelbe Spalte hängt vom zusammengesetzten Schlüssel aus roter und grüner Spalte ab

Übergang zur II.NF:

**Tabelle Schülerdaten**

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Sch.ID | Name | Vorn. | Geb.Dat. | Str./Nr. | PLZ | Ort | FB-ID | FB-Name | FB-Lehrer |
| 0001 | Müller | Max | 24.02.1994 | Seestr. 2 | 83052 | Bruck-mühl | 07 | Physik | Ober-maier |

Rote Spalte: Primärschlüssel-Spalte

**Tabelle Prüfungen**

|  |  |  |
| --- | --- | --- |
| Sch.ID | Prfg.Nr. | Note |
| 0001 | 07/21 | 2,7 |

Grüne Spalten: Zusammengesetzter Schlüssel für die gelbe Spalte

Bemerkung: Eine Relation (Tabelle oder Verbund von Tabellen!), in der jeder Schlüssel nur aus einem Attribut (Spalte!) besteht, ist schon in der II. NF!

Allgemeines Verfahren zur Normalisierung in II.NF:

* Überprüfe, ob es ein Nichtschlüsselattribut (Spalte!) gibt, das **nicht** voll funktional abhängig vom Primärschlüssel ist – also: Überprüfe, ob eine oder mehrere Spalten von einem eventuellen zusammengesetzten Schlüssel abhängen.
* Falls es ein solches Nichtschlüsselattribut gibt, welches wiederum durch einen erweiterten, zusammengesetzten Schlüssel identifizierbar ist, entferne dieses Nichtschlüsselattribut aus der Tabelle und konstruiere eine neue Tabelle mit dem zusammengesetzten Schlüssel als Schlüssel und dem Nichtschlüsselattribut.

Noch vorhandene Nachteile:

* Redundanz: FB-Name und FB-Lehrer wird (immer noch!) pro Schüler gespeichert
* Update-Anomalie: Falls sich der FB-Lehrer ändert, muss dies pro Schüler geändert werden
* Insertion-Anomalie: Ein neuer Fachbereich wird (immer noch!) nur dann erfaßt, wenn es einen Schüler dieses Fachbereiches gibt.

III. Normalform

Definition: Eine Relation ist in III.NF, wenn sie **schon in der II.NF ist** und jedes Nichtschlüsselattribut **nur** vom Schlüssel voll funktional abhängig ist. Das heißt, es darf keine sogenannten „transitiven“ Abhängigkeiten („untereinander-Abhängigkeiten“) geben.  
Im Beispiel (in der II.NF) wäre die Spalte „FB-ID“ ein (weiterer) Schlüssel für die Spalten „FB-Name“ und „FB-Lehrer“. Das heißt, die eigentliche Nichtschlüssel-Spalte „FB-ID“, die ja wiederum von „Sch.ID“ abhängt, ist selbst ein neuer Schlüssel (Stichwort: „untereinander-Schlüssel“; transitiver Schlüssel)

Übergang zur III. NF:

**Tabelle Schüler**

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Sch.ID | Name | Vorn. | Geb.Dat. | Str./Nr. | PLZ | Ort | FB-ID |
| 0001 | Müller | Max | 24.02.1994 | Seestr. 2 | 83052 | Bruckmühl | 07 |

**Tabelle Fachbereich**

|  |  |  |
| --- | --- | --- |
| FB-ID | FB-Name | FB-Lehrer |
| 07 | Physik | Obermaier |

**Tabelle Prüfungen**

|  |  |  |
| --- | --- | --- |
| Sch.ID | Prfg.Nr. | Note |
| 0001 | 07/21 | 2,7 |

Rote Spalten: Primärschlüssel; Grüne Spalten: Zusammengesetzter Schlüssel für die gelbe Spalte

Bemerkung: Relationen, in denen solche transitiven Abhängigkeiten nicht auftauchen (es also neben dem Primärschlüssel keine Untereinander-Abhängigkeiten mehr gibt), sind schon in III.NF.

Allgemeines Verfahren zur Normalisierung in III.NF:

* Falls ein Nichtschlüsselattribut von **anderen** Nichtschlüsselattributen voll funktional abhängig ist, entferne dieses
* Erstelle eine neue Tabelle mit dem neuen Schlüssel und den davon abhängigen Attributen/Spalten

Beispiel: CD-Datenbank (auch auf wikipedia.de zu finden)

CD-Datenbank

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| CD\_ID | Album | Grd.Jahr | Ersch.Jahr | Titelliste |
| 4711 | Anastacia – Not That Kind | 1999 | 2000 | {1. Not That Kind, 2. I‘m Outta Love, 3. Cowboys & Kisses} |
| 4712 | Pink Floys – Wish You Were Here | 1965 | 1975 | {1. Shine On You Crazy Diamond} |
| 4713 | Anastacia – Freak Of Nature | 1999 | 2001 | {1. Paid My Dues} |

(Grd.Jahr = Gründungsjahr der Band/des Interpreten; Ersch.Jahr = Erscheinungsjahr des Albums)

* Zur I. NF:

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| * CD-Datenbank | | | | | | |
| CD\_ID | Albumtitel | Interpret | Grd.Jahr | Ersch.Jahr | Track | Titel |
| 4711 | Not That Kind | Anastacia | 1999 | 2000 | 1 | Not That Kind |
| 4711 | Not That Kind | Anastacia | 1999 | 2000 | 2 | I‘m Outta Love |
| 4711 | Not That Kind | Anastacia | 1999 | 2000 | 3 | Cowboys & Kisses |
| 4712 | Wish You Were Here | Pink Floyd | 1965 | 1975 | 1 | Shine On You Crazy Diamond |
| 4713 | Freak Of Nature | Anastacia | 1999 | 2001 | 1 | Paid My Dues |

Farbgebung: Überlegungen in Richtung II.NF: Rot = Primärschlüssel; Hellrot = davon voll funktional abhängige Spalten; Grün = 2. Komponente eines zusammengesetzten Schlüssels; Gelb: Spalte, die abhängt vom zusammengesetzten Schlüssel „CD\_ID“ und „Track“.

* Zur II. NF:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Albumdaten | | | | |
| CD\_ID | Albumtitel | Interpret | Grd.Jahr | Ersch.Jahr |
| 4711 | Not That Kind | Anastacia | 1999 | 2000 |
| 4712 | Wish You Were Here | Pink Floyd | 1965 | 1975 |
| 4713 | Freak Of Nature | Anastacia | 1999 | 2001 |

|  |  |  |
| --- | --- | --- |
| Trackliste | | |
| CD\_ID | Track | Titel |
| 4711 | 1 | Not That Kind |
| 4711 | 2 | I‘m Outta Love |
| 4711 | 3 | Cowboys & Kisses |
| 4712 | 1 | Shine On You Crazy Diamond |
| 4713 | 1 | Paid My Dues |

Farbgebung: Überlegungen in Richtung III.NF: Rot: Primärschlüssel; Grün: Weiterer, transitiver, „untereinander-“Schüssel für die goldene Spalte (Interpretenname hängt immer direkt mit dem Gründungsjahr zusammen!)

--→ zur III.NF:

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Albumdaten | | | | | | | |
| CD\_ID | | Albumtitel | | | Ersch.Jahr | | Interpret.-ID |
| 4711 | | Not That Kind | | | 2000 | | 311 |
| 4712 | | Wish You Were Here | | | 1975 | | 312 |
| 4713 | | Freak Of Nature | | | 2001 | | 311 |
| Interpretendaten | | | | | | | |
| Interpret.-ID | | | | Int.Name | | Grd.Jahr | |
| 311 | | | | Anastacia | | 1999 | |
| 312 | | | | Pink Floyd | | 1965 | |
| Trackliste | | | | | | | |
| CD\_ID | Track | | Titel | | | | |
| 4711 | 1 | | Not That Kind | | | | |
| 4711 | 2 | | I‘m Outta Love | | | | |
| 4711 | 3 | | Cowboys & Kisses | | | | |
| 4712 | 1 | | Shine On You Crazy Diamond | | | | |
| 4713 | 1 | | Paid My Dues | | | | |

Zweites Beispiel zur Übung:

Lernangeboteübersicht

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Sch.Nr | Name | Klasse | Lehrer | LernA.Nr | Beschr. | Zeit [h] |
| 1 | Ina Jürgens | 11a | Lempel | 2 | Tanz | 12 |
| 2 | Tom Schmidt | 12a | Breier | 3 | Chor | 22 |
| 3 | Franz Jäger | 11a | Lempel | 1,2,3 | Elektronik, Tanz, Chor | 15, 12, 2 |
| 4 | Ina Olsen | 11b | Sommer | 2 | Tanz | 5 |
| 5 | Paula Jürgens | 12a | Breier | 1 | Elektronik | 23 |

(Abkürzungen: Sch.Nr: Schülernummer; LernA.Nr:Lernangebotsnummer; Beschr.: Beschreibung)

* I. NF:

Lernangeboteübersicht

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Sch.Nr | Name | Vorname | Klasse | Lehrer | LernA.Nr | Beschr. | Zeit [h] |
| 1 | Jürgens | Ina | 11a | Lempel | 2 | Tanz | 12 |
| 2 | Schmidt | Tom | 12a | Breier | 3 | Chor | 22 |
| 3 | Jäger | Franz | 11a | Lempel | 1 | Elektronik | 15 |
| 3 | Jäger | Franz | 11a | Lempel | 2 | Tanz | 12 |
| 3 | Jäger | Franz | 11a | Lempel | 3 | Chor | 2 |
| 4 | Olsen | Ina | 11b | Sommer | 2 | Tanz | 5 |
| 5 | Jürgens | Paula | 12a | Breier | 1 | Elektronik | 23 |

* II. NF:

Schüler

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Sch.Nr | Name | Vorname | Klasse | Lehrer |
| 1 | Jürgens | Ina | 11a | Lempel |
| 2 | Schmidt | Tom | 12a | Breier |
| 3 | Jäger | Franz | 11a | Lempel |
| 4 | Olsen | Ina | 11b | Sommer |
| 5 | Jürgens | Paula | 12a | Breier |

Belegte Lernangebote

|  |  |  |  |
| --- | --- | --- | --- |
| Sch.Nr | LernA.Nr | Beschr. | Zeit [h] |
| 1 | 2 | Tanz | 12 |
| 2 | 3 | Chor | 22 |
| 3 | 1 | Elektronik | 15 |
| 3 | 2 | Tanz | 12 |
| 3 | 3 | Chor | 2 |
| 4 | 2 | Tanz | 5 |
| 5 | 1 | Elektronik | 23 |

* III.NF:

Schüler

|  |  |  |  |
| --- | --- | --- | --- |
| Sch.Nr | Name | Vorname | Klasse |
| 1 | Jürgens | Ina | 11a |
| 2 | Schmidt | Tom | 12a |
| 3 | Jäger | Franz | 11a |
| 4 | Olsen | Ina | 11b |
| 5 | Jürgens | Paula | 12a |

Klassen

|  |  |
| --- | --- |
| Klasse | Klassenlehrer |
| 11a | Lempel |
| 12a | Breier |
| 11b | Sommer |

Lernangebote

|  |  |
| --- | --- |
| Lernangebots-Nummer | Beschreibung |
| 1 | Elektronik |
| 2 | Tanz |
| 3 | Chor |

Belegte Lernangebote

|  |  |  |
| --- | --- | --- |
| Sch.Nr | LernA.Nr | Zeit [h] |
| 1 | 2 | 12 |
| 2 | 3 | 22 |
| 3 | 1 | 15 |
| 3 | 2 | 12 |
| 3 | 3 | 2 |
| 4 | 2 | 5 |
| 5 | 1 | 23 |

Weiteres Beispiel:

Film-Trilogien

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| ID | Regisseur/Trilogie | Geb.Jahr Reg. | Ersch.Jahr | Titel |
| 0001 | Peter Jackson – Der Herr der Ringe | 1961 | 2001 | {1. Die Gefährten, 2. Die zwei Türme, 3. Die Rückkehr des Königs} |
| 0002 | George Lucas – Star Wars | 1944 | 1977 | {1. Krieg der Sterne, 2. Das Imperium schlägt zurück, 3. Die Rückkehr der Jedi-Ritter} |
| 0003 | Peter Jackson – Der Hobbit | 1961 | 2012 | {1. Eine unerwartete Reise, 2. Smaugs Einöde, 3. Die Schlacht der fünf Heere } |

(Geb.Jahr Reg: Geburtsjahr des Regisseurs; Ersch.Jahr: Erscheinungsjahr – und zwar Ersch.Jahr des ersten Titels in der Trilogie!)

* Zur I. NF:

Film-Trilogien

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| ID | Regisseur | Trilogie | Geb.Jahr Reg. | Ersch.Jahr | Titelnr. | Titel |
| 0001 | Peter Jackson | Der Herr der Ringe | 1961 | 2001 | 1 | Die Gefährten |
| 0001 | Peter Jackson | Der Herr der Ringe | 1961 | 2001 | 2 | Die zwei Türme |
| 0001 | Peter Jackson | Der Herr der Ringe | 1961 | 2001 | 3 | Die Rückkehr des Königs |
| 0002 | George Lucas | Star Wars | 1944 | 1977 | 1 | Krieg der Sterne |
| 0002 | George Lucas | Star Wars | 1944 | 1977 | 2 | Das Imperium schlägt zurück |
| 0002 | George Lucas | Star Wars | 1944 | 1977 | 3 | Die Rückkehr der Jedi-Ritter |
| 0003 | Peter Jackson | Der Hobbit | 1961 | 2012 | 1 | Eine unerwartete Reise |
| 0003 | Peter Jackson | Der Hobbit | 1961 | 2012 | 2 | Smaugs Einöde |
| 0003 | Peter Jackson | Der Hobbit | 1961 | 2012 | 3 | Die Schlacht der fünf Heere |

(evtl. könnte man den Regisseur auch noch aufteilen in Vor- und Nachname; nachgeholt in der II.NF)

* Zur II. NF:

Film-Trilogien

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| ID | Reg.Name | Reg.Vorname | Trilogie | Geb.J.Reg. | Ersch.J. |
| 0001 | Jackson | Peter | Der Herr der Ringe | 1961 | 2001 |
| 0002 | Lucas | George | Star Wars | 1944 | 1977 |
| 0003 | Jackson | Peter | Der Hobbit | 1961 | 2012 |

Trilogie-Titel

|  |  |  |
| --- | --- | --- |
| ID | Titelnr. | Titel |
| 0001 | 1 | Die Gefährten |
| 0001 | 2 | Die zwei Türme |
| 0001 | 3 | Die Rückkehr des Königs |
| 0002 | 1 | Krieg der Sterne |
| 0002 | 2 | Das Imperium schlägt zurück |
| 0002 | 3 | Die Rückkehr der Jedi-Ritter |
| 0003 | 1 | Eine unerwartete Reise |
| 0003 | 2 | Smaugs Einöde |
| 0003 | 3 | Die Schlacht der fünf Heere |

* Zur III.NF:

Film-Trilogien

|  |  |  |  |
| --- | --- | --- | --- |
| ID | Reg.ID | Trilogie | Ersch.J. |
| 0001 | 9999 | Der Herr der Ringe | 2001 |
| 0002 | 9998 | Star Wars | 1977 |
| 0003 | 9999 | Der Hobbit | 2012 |

(Reg.ID: ID für den Regisseur)

Regisseur

|  |  |  |  |
| --- | --- | --- | --- |
| Reg.ID | Reg.Name | Reg.Vorname | Geb.J. |
| 9999 | Jackson | Peter | 1961 |
| 9998 | Lucas | George | 1944 |

Trilogie-Titel

|  |  |  |
| --- | --- | --- |
| ID | Titelnr. | Titel |
| 0001 | 1 | Die Gefährten |
| 0001 | 2 | Die zwei Türme |
| 0001 | 3 | Die Rückkehr des Königs |
| 0002 | 1 | Krieg der Sterne |
| 0002 | 2 | Das Imperium schlägt zurück |
| 0002 | 3 | Die Rückkehr der Jedi-Ritter |
| 0003 | 1 | Eine unerwartete Reise |
| 0003 | 2 | Smaugs Einöde |
| 0003 | 3 | Die Schlacht der fünf Heere |