



BEUTH HOCHSCHULE FÜR TECHNIK BERLIN
University of Applied Sciences

Wirtschaftsinformatik 2

LE 10 – Normalformen

Prof. Dr. Thomas Off

<http://www.ThomasOff.de/lehre/beuth/wi2>

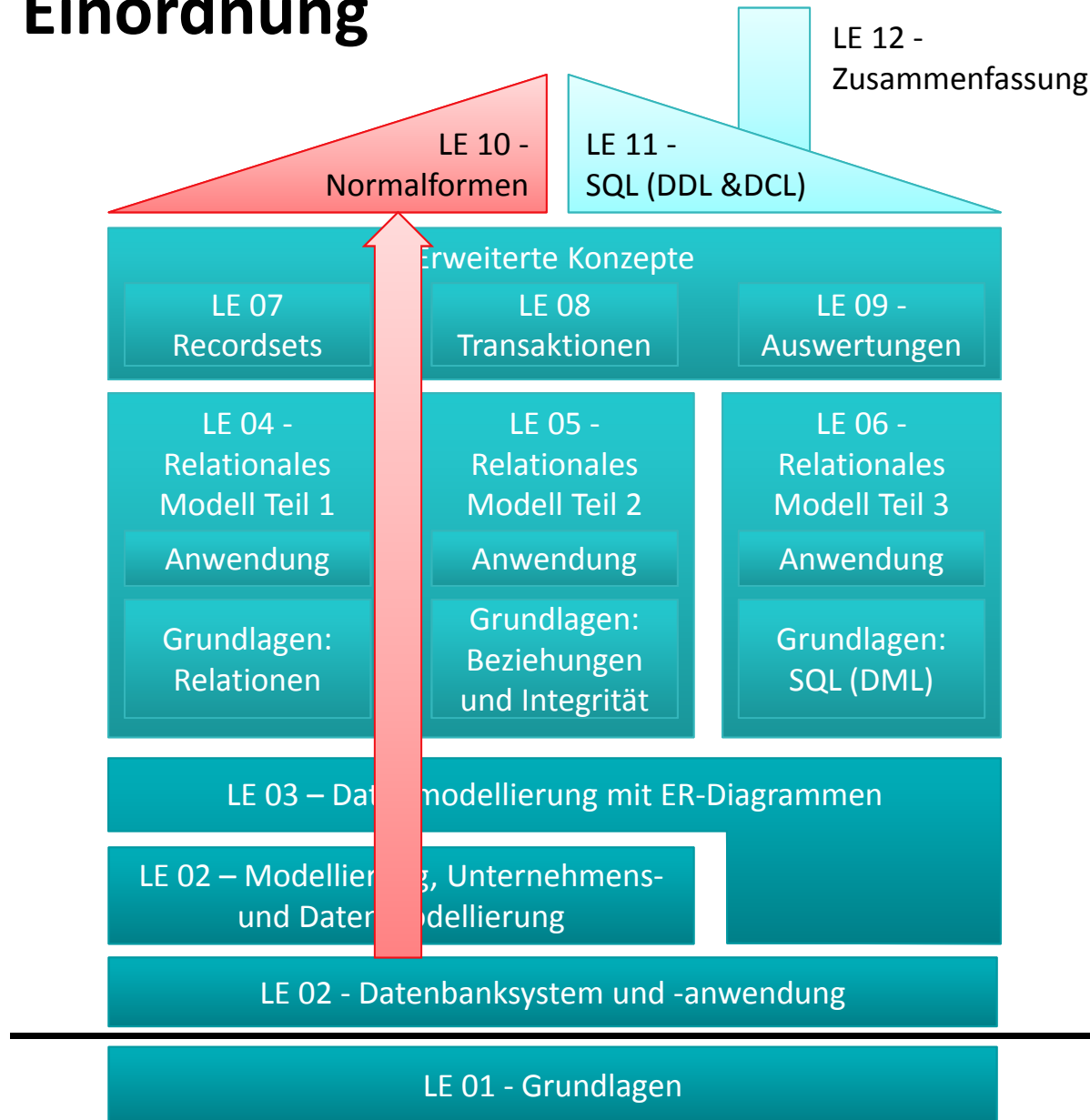
Ziel



Ziel dieser Lehreinheit

- Normalformen und Normalisierung kennenlernen
- Verstehen der Herausforderungen und Probleme die sich beim Entwurf von Datenbanken aus den an sie gestellten Anforderungen ergeben
- Anwenden der Normalisierung auf vorhandene Relationen, um die Anforderungen besser zu erreichen
- Analysieren und Beurteilen eines Datenbankentwurfs hinsichtlich des Grades der Normalisierung

Einordnung





Inhalt

Ziel und Einordnung

Rückblicke

Einführung

Normalformen

- Ausgangssituation
- Erste Normalform
- Zweite Normalform und ihr formaler Hintergrund
- Dritte Normalform und ihr formaler Hintergrund
- Weitere Normalformen

Normalisierung

- Gewählter Ansatz
- Geplanter Einsatzzweck

Zusammenfassung

Ausblick

Rückblick (LE09 – Auswertungen)



Berichte in MS Access

- werden in aufbereiteter Form als Dokument mit relevanten Daten, mit Gruppierungs- und Summenfunktionen und mit einem Layout erzeugt und zum Export bereitgestellt

Exemplarische Vorgehensweise zur Erstellung von Berichten

- Berichtsstruktur entwerfen
- Erstellung von Abfragen und Teilberichten
 - Erstellen geeigneter Abfragen, die notwendige Angaben liefern
 - Erstellung eines einzelnen Berichtes pro Abfrage
- Zusammenfassung der einzelnen Berichte zu einem Gesamtbericht mit Sortierung, Summen, Gruppierungen
- Einbettung der Berichtserzeugung in den Programmcode (DoCmd.OpenReport)

Rückblick (LE09 – Auswertungen)



SQL bietet Statistikfunktionen, z.B. AVG(), COUNT(), MAX(), MIN(), SUM()

Unterabfragen: Mittel zur Auswertung von Zusammenhängen

```
SELECT COUNT(kndOrt) AS AnzahlWohnorte
FROM
  (SELECT DISTINCT kndOrt
   FROM tblKunden) AS Unterabfrage;
```

Gruppierung von Daten

- GROUP BY: Spalten, anhand derer Tupel mit gleichen Werten gruppiert werden
- HAVING: Filterung anhand von Werten, die sich durch die Gruppenbildung ergeben

```
SELECT kndOrt, COUNT(*) AS KundenImOrt FROM tblKunden
GROUP BY kndOrt
HAVING COUNT(*) > 7;
```

Rückblick



LE 10 - Normalformen



Rückblick (LE04 – Relationales Modell Teil 1)



Definition "Schlüsselkandidat"

- besteht aus einem oder mehreren Attributen
- identifiziert eindeutig jedes Tupel (Datensatz)
- ist minimal, d.h. beim Weglassen eines Attributes geht Eindeutigkeit verloren

Definition "Primärschlüssel" (syn. "Primary Key")

- besitzt eine Relation mehr als einen Schlüsselkandidaten, wird einer als Primärschlüssel ausgewählt
- andere Schlüsselkandidaten sind "Alternativschlüssel"

Rückblick



LE 02 - BRDialgen



Rückblick (LE 02 – Grundlagen DBS)



Anforderungen an relationale Datenbanken

- vermeiden die Mehrfachspeicherung (Redundanz) gleicher Daten
 - keine unnötige Belegung von Speicherplatz
 - Vermeidung von Inkonsistenzen/Widersprüchen wenn nicht alle Vorkommen gleich und gleichzeitig geändert werden
- einfache und sichere Änderungsmöglichkeiten der Daten
 - Reduzierung von Abhängigkeiten
 - Reduzierung des Risikos von Inkonsistenzen/Widersprüchen
- ...



Inhalt

Ziel und Einordnung

Rückblicke

Einführung

Normalformen

- Ausgangssituation
- Erste Normalform
- Zweite Normalform und ihr formaler Hintergrund
- Dritte Normalform und ihr formaler Hintergrund
- Weitere Normalformen

Normalisierung

- Gewählter Ansatz
- Geplanter Einsatzzweck

Zusammenfassung

Ausblick

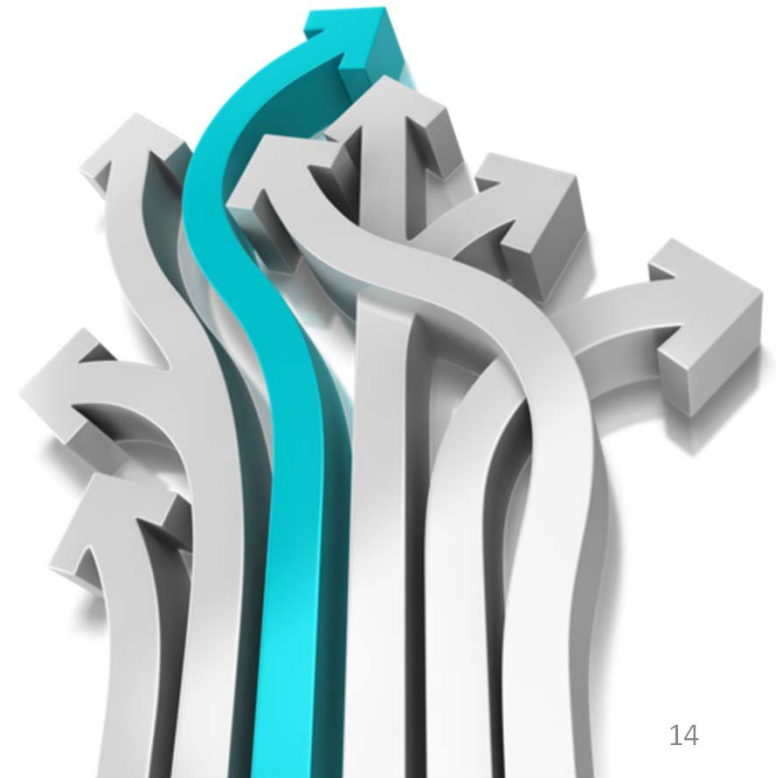
Einführung der Normalisierung



Verschiedene Ansätze

Relationen entwerfen, Bsp.:

- Strukturierter Ansatz durch Ableitung aus ER-Modellen
- Sammlung zu speichernder Informationen und deren Zusammenfassung in Tabellen
- ...



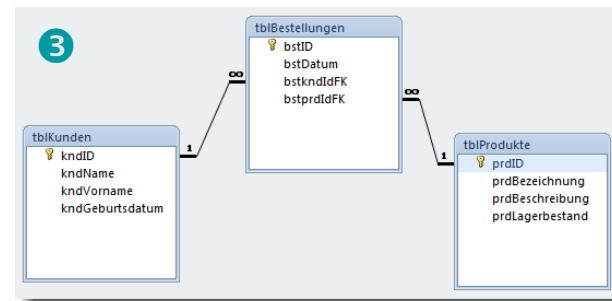
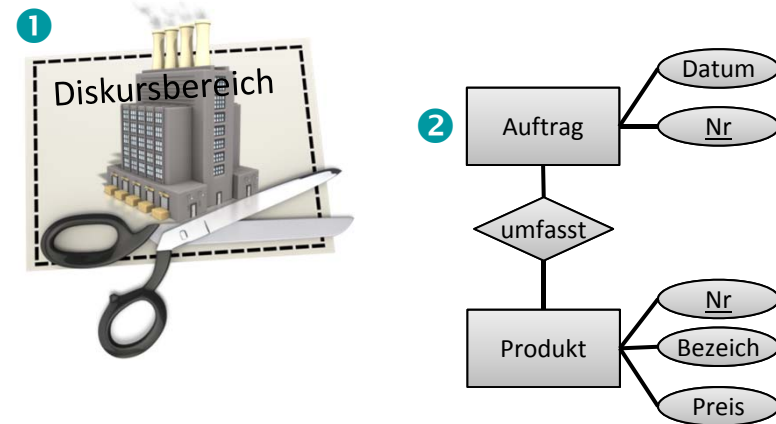
Einführung der Normalisierung



Verschiedene Ansätze

Relationen entwerfen, Bsp.:

- Strukturierter Ansatz durch Ableitung aus ER-Modellen
 - führt zu mehreren Tabellen, die
 - einzelne Sachverhalte des Diskusbereichs abbilden und
 - in Beziehung stehen.
- Sammlung zu speichernder Informationen und deren Zusammenfassung in Tabellen



4

```
CREATE TABLE [tblBestellungen]
([bstID] COUNTER, [bstDatum]
DATE, [bstkndIdFK] LONG,
[bstprdIdFK] LONG)

CREATE UNIQUE INDEX bstID ON
[tblBestellungen]([bstID])

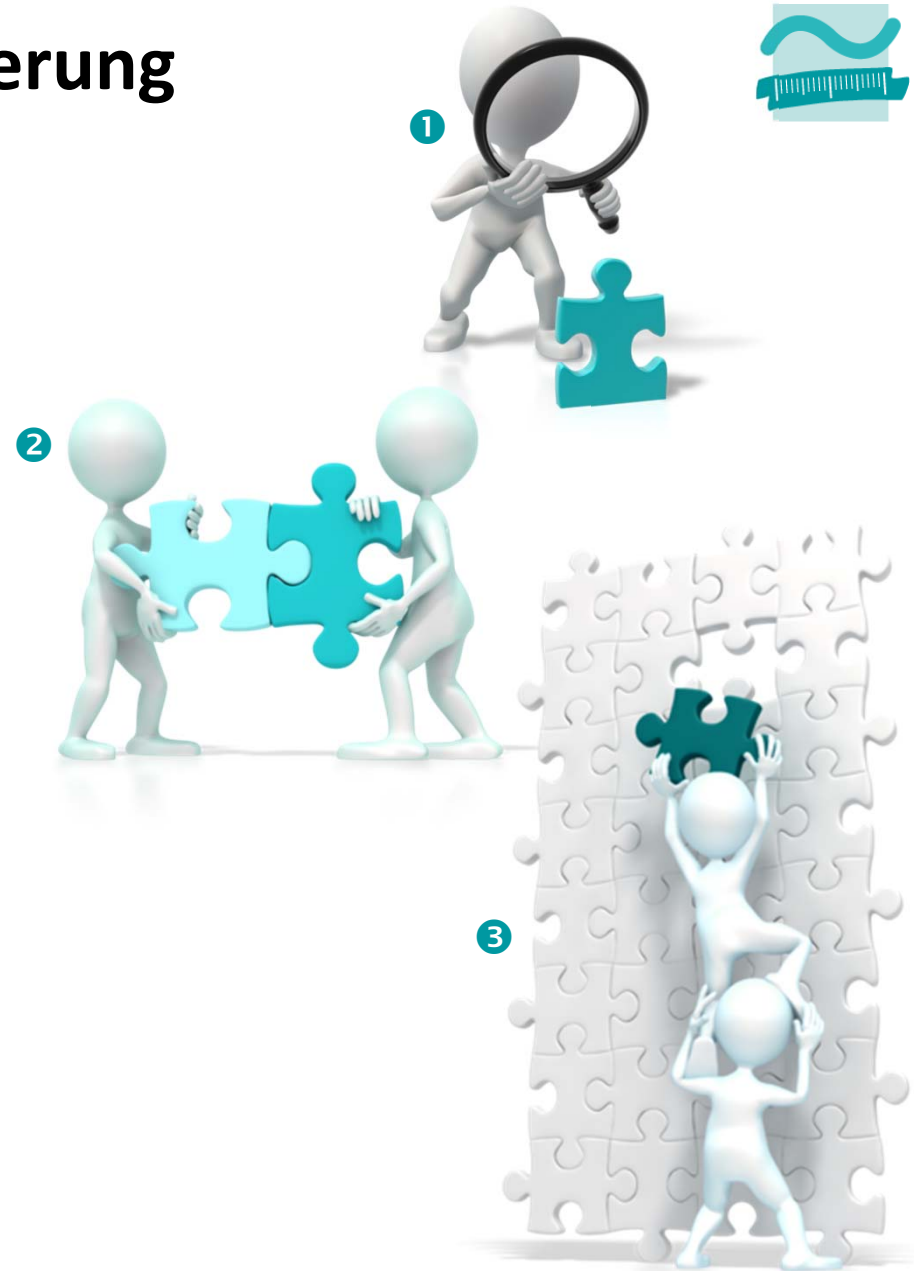
-- ...
```

Einführung der Normalisierung

Verschiedene Ansätze

Relationen entwerfen, Bsp.:

- Strukturierter Ansatz durch Ableitung aus ER-Modellen
- Sammlung zu speichernder Informationen und deren Zusammenfassung in Tabellen
 - führt (initial) zu Tabellen mit vielen Attributen, ohne dass einzelne Sachverhalte oder Beziehungen erkennbar sind



Einführung der Normalisierung



Verschiedene Ansätze

Relationen entwerfen, Bsp.:

- Strukturierter Ansatz durch Ableitung aus ER-Modellen
- Sammlung zu speichernder Informationen und deren Zusammenfassung in Tabellen
- ...



*Gibt keine Regeln
für einen guten
Relationen-
entwurf?*



Normalformen

Normalformen sind Regeln

- die helfen, Fehler im Entwurfsprozess zu vermeiden
- mit denen die Ziele relationaler Datenbanken erreicht werden können

Normalisierung als Prozess, diese Regeln anzuwenden

- um schrittweise Redundanzen und (nicht gewünschte) Abhängigkeiten aufzuspüren
- der in definierten Stufen (Normalformen) durchlaufen wird und
- in dem Anpassungen der Relationenstruktur vorgenommen werden, ohne dass Informationsverluste entstehen.

Normalformen

Ausgangssituation



Normalformen

Ausgangssituation



"Verkäufer (Nr, Name) verkaufen immer mindestens ein Produkt. Es gibt eine Priorität, in der ein Verkäufer sich um den Verkauf von Produkten (Bezeichnung) kümmert. Der Verkäufer erzielt Umsatz durch den Verkauf von Produkten. Jeder Verkäufer gehört zu genau einer Filiale (Nr, PLZ, Ort)."

Normalformen

Ausgangssituation



Verkäufer ¹	VkNr	Name	Filiale	Plz	Ort	Produkte	Umsatz
	123	Albers	1	12345	Berlin	1. Kühlschrank, 2. Waschmaschine	120.000€
	234	Boehrs	1	12345	Berlin	1. TV, 2. Kühlschrank, 3. DVD	78.000€
	345	Dinkel	2	13456	Berlin	1. TV	56.000€
	456	Dinkels	2	13456	Berlin	1. Kühlschrank	12.000€

1) vgl. hier und im Folgenden [1], S. 64 ff.

Normalformen

Ausgangssituation



Verkäufer ¹	VkNr	Name	Filiale	Plz	Ort	Produkte	Umsatz
	123	Albers	1	12345	Berlin	1. Kühlschrank, 2. Waschmaschine	120.000€
	234	Boehrs	1	12345	Berlin	1. TV, 2. Kühlschrank, 3. DVD	78.000€
	345	Dinkel	2	13456	Berlin	1. TV	56.000€
	456	Dinkels	2	13456	Berlin	1. Kühlschrank	12.000€

1) vgl. hier und im Folgenden [1], S. 64 ff.

Normalformen



Ausgangssituation: Wie lautet die SQL-Abfrage ...

- aller Produkte des Sortiments?
- zum Hinzufügen einer Filiale ohne dass es schon Verkäufer gibt (Neueröffnung)?
- zur Änderung von "TV" in "Fernseher"?



Verkäufer	VkNr	Name	Filiale	Plz	Ort	Produkte	Umsatz
	123	Albers	1	12345	Berlin	1. Kühlschrank, 2. Waschmaschine	120.000€
	234	Boehrs	1	12345	Berlin	1. TV, 2. Kühlschrank, 3. DVD	78.000€
	345	Dinkel	2	13456	Berlin	1. TV	56.000€
	456	Dinkels	2	13456	Berlin	1. Kühlschrank	12.000€

Hintergrund



Anomalien in Datenbanken

- im Allgemeinen ein "Unregelmäßigkeit", Abweichung von üblichen Regeln
- Anomalien können in Datenbanken Auslöser von Inkonsistenzen/Problemen sein



Hintergrund

Einfügeanomalie

- Entität kann nicht oder nur schwer zu einer Tabelle hinzugefügt werden,
- weil für den Datensatz noch nicht bekannte Werte (z.B. als Teil des Primärschlüssels) benötigt werden.

Änderungsanomalie

- tritt auf, wenn eine Entität mehrfach gespeichert wurde (Redundanz) und dadurch die mehrfache Änderung gleicher Werte notwendig ist, obwohl sich nur ein Sachverhalt geändert hat und damit das Risiko von Inkonsistenzen steigt.

Löschanomalie

- beim Löschen einer Entität gehen Informationen im zugehörigen Datensatz verloren, die noch benötigt werden und nicht hätten gelöscht werden sollen.

Hintergrund



Anomalien in Datenbanken

- Hinweis: Beispiele für weitere Arten von Anomalien, die durch mehrere Benutzer verursacht werden können, wurden in LE 08 vorgestellt (Lost Update, Dirty Read, Non-Repeatable Read, Phantom Read).

Normalformen



Ausgangssituation: Wie lautet die SQL-Abfrage ...

- aller Produkte des Sortiments?
- zum Hinzufügen einer Filiale ohne dass es schon Verkäufer gibt (Neueröffnung)?
- zur Änderung von "TV" in "Fernseher"?



Verkäufer	VkNr	Name	Filiale	Plz	Ort	Produkte	Umsatz
	123	Albers	1	12345	Berlin	1. Kühlschrank, 2. Waschmaschine	120.000€
	234	Boehrs	1	12345	Berlin	1. TV, 2. Kühlschrank, 3. DVD	78.000€
	345	Dinkel	2	13456	Berlin	1. TV	56.000€
	456	Dinkels	2	13456	Berlin	1. Kühlschrank	12.000€

Normalformen



Ausgangssituation: Weitere Probleme...

- Einfügeanomalie: Neue Filiale kann erst eingefügt werden, wenn alle Informationen vorliegen (z.B. wenn auch Verkäufer der Filiale zugeordnet wurde)
- Löschanomalie: Löschen von Verkäufern führt u.U. zum Löschen der Filiale gesamten Filiale
- Änderungsanomalie: Wenn eine Filiale umzieht, müssen alle betroffenen Datensätze gefunden und gleich geändert werden

Verkäufer	VkNr	Name	Filiale	Plz	Ort	Produkte	Umsatz
	123	Albers	1	12345	Berlin	1. Kühlschrank, 2. Waschmaschine	120.000€
	234	Boehrs	1	12345	Berlin	1. TV, 2. Kühlschrank, 3. DVD	78.000€
	345	Dinkel	2	13456	Berlin	1. TV	56.000€
	456	Dinkels	2	13456	Berlin	1. Kühlschrank	12.000€



Normalformen

1. Normalform

- Ziel: Vereinfachung der Aktualisierung und des Zugriffs
- Regel
 - Mehrere Datenwerte in einer Zelle sind nicht zulässig.
 - Alle Spalten dürfen in ihren Zellen jeweils nur atomare Werte enthalten.
- Ausgangssituation (Beispiel): **Ist das 1. Normalform?**


Verkäufer	VkNr	Name	Filiale	Plz	Ort	Produkte	Umsatz
	123	Albers	1	12345	Berlin	1. Kühlschrank, 2. Waschmaschine	120.000€
	234	Boehrs	1	12346	Berlin	1. TV, 2. Kühlschrank, 3. DVD	78.000€
	345	Dinkel	2	13456	Berlin	1. TV	56.000€
	456	Dinkels	2	13456	Berlin	1. Kühlschrank	12.000€



Normalformen

1. Normalform

- Ziel: Vereinfachung der Aktualisierung und des Zugriffs
- Regel
 - Mehrere Datenwerte in einer Zelle sind nicht zulässig.
 - Alle Spalten dürfen in ihren Zellen jeweils nur atomare Werte enthalten.
- Ausgangssituation (Beispiel): **Nicht in 1. Normalform!**



Verkäufer	VkNr	Name	Filiale	Plz	Ort	Produkte	Umsatz
	123	Albers	1	12345	Berlin	1. Kühlschrank, 2. Waschmaschine	120.000€
	234	Boehrs	1	12346	Berlin	1. TV, 2. Kühlschrank, 3. DVD	78.000€
	345	Dinkel	2	13456	Berlin	1. TV	56.000€
	456	Dinkels	2	13456	Berlin	1. Kühlschrank	12.000€

Normalformen



1. Normalform

- Ergebnis der Normalisierung (Beispiel):
 - noch immer nicht nur atomare Werte in allen Spalten (Kombination aus Priorität und Produktbezeichnung)
 - Schlüssel Verkäufersnummer (VkNr) nicht mehr eindeutig!

Verkäufer	<u>VkNr</u>	Name	Filiale	Plz	Ort	Produkt	Umsatz
	123	Albers	1	12345	Berlin	1. Kühlschrank	61.000€
	123	Albers	1	12345	Berlin	2. Waschmaschine	59.000€
	234	Boehrs	1	12345	Berlin	1. TV	12.500€
	234	Boehrs	1	12345	Berlin	2. Kühlschrank	55.000€
	234	Boehrs	1	12345	Berlin	3. DVD	10.500€
	345	Dinkel	2	13456	Berlin	1. TV	56.000€
	456	Dinkels	2	13456	Berlin	2. Kühlschrank	12.000€

Normalformen



1. Normalform

- Ergebnis der Normalisierung (Beispiel):
 - nur atomare Werte in allen Spalten
 - Schlüssel Verkäufersnummer (VkNr) nicht mehr eindeutig
 - neue Schlüsselkandidaten als zusammengesetzte Schlüssel aus
 - VkNr + Priorität des von ihm verkauften Produktes (Prio)
 - VkNr + Bezeichnung des verkauften Produktes (Produkt)

Verkäufer	VkNr	Name	Filiale	Plz	Ort	Prio	Produkt	Umsatz
	123	Albers	1	12345	Berlin	1	Kühlschrank	61.000€
	123	Albers	1	12345	Berlin	2	Waschmaschine	59.000€
	234	Boehrs	1	12345	Berlin	1	TV	12.500€
	234	Boehrs	1	12345	Berlin	2	Kühlschrank	55.000€
	234	Boehrs	1	12345	Berlin	3	DVD	10.500€
	345	Dinkel	2	13456	Berlin	1	TV	56.000€
	456	Dinkels	2	13456	Berlin	1	Kühlschrank	12.000€

Normalformen



1. Normalform

- Ergebnis der Normalisierung (Beispiel):
 - nur atomare Werte in allen Spalten
 - Schlüssel Verkäufersnummer (VkNr) nicht mehr eindeutig
 - neue Schlüsselkandidaten als zusammengesetzte Schlüssel aus
 - VkNr + Priorität des von ihm verkauften Produktes (Prio)
 - VkNr + Bezeichnung des verkauften Produktes (Produkt)
 - Wahl von VkNr + Prio als Primärschlüssel

Verkäufer	<u>VkNr</u>	Name	Filiale	Plz	Ort	<u>Prio</u>	Produkt	Umsatz
	123	Albers	1	12345	Berlin	1	Kühlschrank	61.000€
	123	Albers	1	12345	Berlin	2	Waschmaschine	59.000€
	234	Boehrs	1	12345	Berlin	1	TV	12.500€
	234	Boehrs	1	12345	Berlin	2	Kühlschrank	55.000€
	234	Boehrs	1	12345	Berlin	3	DVD	10.500€
	345	Dinkel	2	13456	Berlin	1	TV	56.000€
	456	Dinkels	2	13456	Berlin	1	Kühlschrank	12.000€

Normalformen



... ist das ein guter Entwurf? Bezogen auf die Ziele

- komfortable und schnelle Abfragemöglichkeiten
- leichte und einfache Änderbarkeit
- Redundanzfreiheit (um z.B. Speicherplatz zu sparen, Änderungen zu vereinfachen)

Verkäufer	<u>VkNr</u>	Name	Filiale	Plz	Ort	<u>Prio</u>	Produkt	Umsatz
	123	Albers	1	12345	Berlin	1	Kühlschrank	61.000€
	123	Albers	1	12345	Berlin	2	Waschmaschine	59.000€
	234	Boehrs	1	12345	Berlin	1	TV	12.500€
	234	Boehrs	1	12345	Berlin	2	Kühlschrank	55.000€
	234	Boehrs	1	12345	Berlin	3	DVD	10.500€
	345	Dinkel	2	13456	Berlin	1	TV	56.000€
	456	Dinkels	2	13456	Berlin	1	Kühlschrank	12.000€

Normalformen



... ist das ein guter Entwurf?

Fortschritte erkennbar, aber Redundanzen erschweren z.B.:

- Namensänderung (mehrfache Änderung nötig)

Verkäufer	<u>VkNr</u>	Name	Filiale	Plz	Ort	<u>Prio</u>	Produkt	Umsatz
	123	Albers	1	12345	Berlin	1	Kühlschrank	61.000€
	123	Albers	1	12345	Berlin	2	Waschmaschine	59.000€
	234	Boehrs	1	12345	Berlin	1	TV	12.500€
	234	Boehrs	1	12345	Berlin	2	Kühlschrank	55.000€
	234	Boehrs	1	12345	Berlin	3	DVD	10.500€
	345	Dinkel	2	13456	Berlin	1	TV	56.000€
	456	Dinkels	2	13456	Berlin	1	Kühlschrank	12.000€

Normalformen



... ist das ein guter Entwurf?

Fortschritte erkennbar, aber Redundanzen erschweren z.B.:

- Namensänderung (mehrfache Änderung nötig)
- Umzug einer Filiale (mehrfache Änderung nötig)

Verkäufer	<u>VkNr</u>	Name	Filiale	Plz	Ort	<u>Prio</u>	Produkt	Umsatz
	123	Albers	1	12345	Berlin	1	Kühlschrank	61.000€
	123	Albers	1	12345	Berlin	2	Waschmaschine	59.000€
	234	Boehrs	1	12345	Berlin	1	TV	12.500€
	234	Boehrs	1	12345	Berlin	2	Kühlschrank	55.000€
	234	Boehrs	1	12345	Berlin	3	DVD	10.500€
	345	Dinkel	2	13456	Berlin	1	TV	56.000€
	456	Dinkels	2	13456	Berlin	1	Kühlschrank	12.000€

Normalformen



... ist das ein guter Entwurf?

Fortschritte erkennbar, aber Redundanzen erschweren z.B.:

- Namensänderung (mehrfache Änderung nötig)
- Umzug einer Filiale (mehrfache Änderung nötig)
- Löschen des Produkts "TV" aus dem Sortiment (würde auch Verkäuferin Dinkel löschen)

Verkäufer	VkNr	Name	Filiale	Plz	Ort	Prio	Produkt	Umsatz
	123	Albers	1	12345	Berlin	1	Kühlschrank	61.000€
	123	Albers	1	12345	Berlin	2	Waschmaschine	59.000€
	234	Boehrs	1	12345	Berlin	1	TV	12.500€
	234	Boehrs	1	12345	Berlin	2	Kühlschrank	55.000€
	234	Boehrs	1	12345	Berlin	3	DVD	10.500€
	345	Dinkel	2	13456	Berlin	1	TV	56.000€
	456	Dinkels	2	13456	Berlin	1	Kühlschrank	12.000€

Hintergrund



Funktionale Abhängigkeit¹

- Attribut b ist funktional abhängig vom Attribut a der gleichen Relation R , wenn zu jedem Wert von a höchstens ein Wert von b möglich ist
- oder anders formuliert: wenn zwei Tupel der Relation den gleichen Wert für a haben, dann haben Sie auch den gleichen Wert für b
- Hinweis: a und b nicht nur als einzelne Attribute, sondern auch zusammengesetzt möglich
- Notation

$$a \rightarrow b$$

1) vgl. hier und im Folgenden [1], S. 121 f.



Hintergrund

Beispiele für funktionale Abhängigkeit

- VKNr → Name
- VKNr → Ort
- (VKNr, Prio) → Produkte
- Filiale → Plz
- Filiale → Ort

Negativbeispiele – keine funktionale Abhängigkeit

- Prio ↛ Produkte
- VKNr ↛ Umsatz

Verkäufer	<u>VKNr</u>	Name	Filiale	Plz	Ort	<u>Prio</u>	Produkt	Umsatz
	123	Albers	1	12345	Berlin	1	Kühlschrank	61.000€
	123	Albers	1	12345	Berlin	2	Waschmaschine	59.000€
	234	Boehrs	1	12345	Berlin	1	TV	12.500€
	234	Boehrs	1	12345	Berlin	2	Kühlschrank	55.000€
	234	Boehrs	1	12345	Berlin	3	DVD	10.500€
	345	Dinkel	2	13456	Berlin	1	TV	56.000€
	456	Dinkels	2	13456	Berlin	1	Kühlschrank	12.000€



Hintergrund

Volle funktionale Abhängigkeit¹

- Attribut b ist **voll funktional abhängig** von Attribut a der selben Relation R , wenn
 - es funktional abhängig ist von a
 - nicht aber funktional abhängig nur von einem Teil von a
- Hinweis: wenn a nicht zusammengesetzt ist, bedeutet funktionale Abhängigkeit immer auch volle funktionale Abhängigkeit
- Notation

$$a \Rightarrow b$$

¹) vgl. hier und im Folgenden [1], S. 121 f.



Hintergrund

Beispiel für volle funktionale Abhängigkeit

- (VkNr, Prio) \Rightarrow Produkt
- (VkNr, Prio) \Rightarrow Umsatz
- Filiale \Rightarrow Plz
- Filiale \Rightarrow Ort

Negativbeispiel – keine volle funktionale Abhängigkeit

- (VkNr, Prio) $\not\Rightarrow$ Name
- (VkNr, Prio) $\not\Rightarrow$ Filiale

Verkäufer	<u>VkNr</u>	Name	Filiale	Plz	Ort	<u>Prio</u>	Produkt	Umsatz
	123	Albers	1	12345	Berlin	1	Kühlschrank	61.000€
	123	Albers	1	12345	Berlin	2	Waschmaschine	59.000€
	234	Boehrs	1	12345	Berlin	1	TV	12.500€
	234	Boehrs	1	12345	Berlin	2	Kühlschrank	55.000€
	234	Boehrs	1	12345	Berlin	3	DVD	10.500€
	345	Dinkel	2	13456	Berlin	1	TV	56.000€
	456	Dinkels	2	13456	Berlin	1	Kühlschrank	12.000€

Normalformen



... ist das ein guter Entwurf?

Fortschritte erkennbar, aber Redundanzen erschweren z.B.:

- Namensänderung (mehrfache Änderung nötig)
- Umzug einer Filiale (mehrfache Änderung nötig)
- Löschen des Produkts "TV" aus dem Sortiment (würde auch Verkäuferin Dinkel löschen)

Verkäufer	VkNr	Name	Filiale	Plz	Ort	Prio	Produkt	Umsatz
	123	Albers	1	12345	Berlin	1	Kühlschrank	61.000€
	123	Albers	1	12345	Berlin	2	Waschmaschine	59.000€
	234	Boehrs	1	12345	Berlin	1	TV	12.500€
	234	Boehrs	1	12345	Berlin	2	Kühlschrank	55.000€
	234	Boehrs	1	12345	Berlin	3	DVD	10.500€
	345	Dinkel	2	13456	Berlin	1	TV	56.000€
	456	Dinkels	2	13456	Berlin	1	Kühlschrank	12.000€

Normalformen



2. Normalform

- Ziele
 - Nur zusammengehörige Daten sind in einer Relation enthalten.
 - Jede Relation stellt nur einen Sachverhalt der Realität dar.
- Regel: in 1. Normalform und alle Nicht-Schlüsselattribute vom gesamten Schlüssel abhängig (volle funktionale Abhängigkeit, vom Primärschlüssel gehen Doppelpfeile aus)
- Ausgangssituation (Beispiel): **Ist das 2. Normalform?**

Verkäufer	<u>VkNr</u>	Name	Filiale	Plz	Ort	<u>Prio</u>	Produkt	Umsatz
	123	Albers	1	12345	Berlin	1	Kühlschrank	61.000€
	123	Albers	1	12345	Berlin	2	Waschmaschine	59.000€
	234	Boehrs	1	12345	Berlin	1	TV	12.500€
	234	Boehrs	1	12345	Berlin	2	Kühlschrank	55.000€
	234	Boehrs	1	12345	Berlin	3	DVD	10.500€
	345	Dinkel	2	13456	Berlin	1	TV	56.000€
	456	Dinkels	2	13456	Berlin	1	Kühlschrank	12.000€

Normalformen



2. Normalform

- Ziele
 - Nur zusammengehörige Daten sind in einer Relation enthalten.
 - Jede Relation stellt nur einen Sachverhalt der Realität dar.
- Regel: in 1. Normalform und alle Nicht-Schlüsselattribute vom gesamten Schlüssel abhängig (volle funktionale Abhängigkeit, vom Primärschlüssel gehen Doppelpfeile aus)
- Ausgangssituation (Beispiel): **Nicht in 2. Normalform!**

Verkäufer	<u>VkNr</u>	Name	Filiale	Plz	Ort	<u>Prio</u>	Produkt	Umsatz
	123	Albers	1	12345	Berlin	1	Kühlschrank	61.000€
	123	Albers	1	12345	Berlin	2	Waschmaschine	59.000€
	234	Boehrs	1	12345	Berlin	1	TV	12.500€
	234	Boehrs	1	12345	Berlin	2	Kühlschrank	55.000€
	234	Boehrs	1	12345	Berlin	3	DVD	10.500€
	345	Dinkel	2	13456	Berlin	1	TV	56.000€
	456	Dinkels	2	13456	Berlin	1	Kühlschrank	12.000€



Normalformen

2. Normalform

- Ergebnis der Normalisierung
 - Aufteilung auf zwei Relationen
 - Verkäufer mit Primärschlüssel VkNr, nicht zusammengesetzt
 - Produkte
 - mit zusammengesetztem Primärschlüssel VkNr, Prio
 - VkNr ist gleichzeitig Fremdschlüssel für Zuordnung zu Verkäufer

Verkäufer	<u>VkNr</u>	Name	Filiale	Plz	Ort
	123	Albers	1	12345	Berlin
	234	Boehrs	1	12345	Berlin
	345	Dinkel	2	13456	Berlin
	456	Dinkels	2	13456	Berlin

Produkte	<u>VkNr</u>	<u>Prio</u>	Produkt	Umsatz
	123	1	Kühlschrank	61.000€
	123	2	Waschmaschine	59.000€
	234	1	TV	12.500€
	234	2	Kühlschrank	55.000€
	234	3	DVD	10.500€
	345	1	TV	56.000€
	456	1	Kühlschrank	12.000€



Normalformen

... ist das ein guter Entwurf? Bezogen auf die Ziele

- komfortable und schnelle Abfragemöglichkeiten
- leichte und einfache Änderbarkeit
- Redundanzfreiheit (um z.B. Speicherplatz zu sparen, Änderungen zu vereinfachen)

Verkäufer	<u>VkNr</u>	Name	Filiale	Plz	Ort
	123	Albers	1	12345	Berlin
	234	Boehrs	1	12345	Berlin
	345	Dinkel	2	13456	Berlin
	456	Dinkels	2	13456	Berlin

Produkte	<u>VkNr</u>	<u>Prio</u>	Produkt	Umsatz
	123	1	Kühlschrank	61.000€
	123	2	Waschmaschine	59.000€
	234	1	TV	12.500€
	234	2	Kühlschrank	55.000€
	234	3	DVD	10.500€
	345	1	TV	56.000€
	456	1	Kühlschrank	12.000€

Normalformen



... ist das ein guter Entwurf?

Fortschritte erkennbar, aber Redundanzen erschweren z.B.:

- Umzug einer Filiale (mehrfache Änderung nötig)
- Änderung der Produktbezeichnung (mehrfache Änderungen nötig)

Verkäufer	VkNr	Name	Filiale	Plz	Ort
	123	Albers	1	12345	Berlin
	234	Boehrs	1	12345	Berlin
	345	Dinkel	2	13456	Berlin
	456	Dinkels	2	13456	Berlin

Produkte	VkNr	Prio	Produkt	Umsatz
	123	1	Kühlschrank	61.000€
	123	2	Waschmaschine	59.000€
	234	1	TV	12.500€
	234	2	Kühlschrank	55.000€
	234	3	DVD	10.500€
	345	1	TV	56.000€
	456	1	Kühlschrank	12.000€

Hintergrund



Determinante

- Attribut ***d*** ist eine Determinante, wenn ein anderes Attribut der gleichen Relation voll funktional abhängig von ***d*** ist
- bedeutet vereinfacht, dass alle Attribute, von denen ein Doppelpfeil ausgeht, Determinanten sind
- Hinweis: ***d*** kann zusammengesetztes Attribut sein

1) vgl. hier und im Folgenden [1], S. 126 f.



Hintergrund

Beispiel für Determinanten

- Relation Verkäufer
 - VKNr \Rightarrow Name; VKNr ist Determinante für Name
 - Filiale \Rightarrow Plz; Filiale ist Determinante für Plz
 - Filiale \Rightarrow Ort; Filiale ist Determinante für Ort

Negativbeispiel – keine Determinanten

- Relation Verkäufer:
 - Ort \nrightarrow Filiale; Ort ist keine Determinante für Filiale

Verkäufer	VKNr	Name	Filiale	Plz	Ort
	123	Albers	1	12345	Berlin
	234	Boehrs	1	12345	Berlin
	345	Dinkel	2	13456	Berlin
	456	Dinkels	2	13456	Berlin

Produkte	VKNr	Prio	Produkt	Umsatz
	123	1	Kühlschrank	61.000€
	123	2	Waschmaschine	59.000€
	234	1	TV	12.500€
	234	2	Kühlschrank	55.000€
	234	3	DVD	10.500€
	345	1	TV	56.000€
	456	1	Kühlschrank	12.000€

Normalformen



... ist das ein guter Entwurf?

Fortschritte erkennbar, aber Redundanzen erschweren z.B.:

- Umzug einer Filiale (mehrfache Änderung nötig)
- Änderung der Produktbezeichnung (mehrfache Änderungen nötig)

Verkäufer	VkNr	Name	Filiale	Plz	Ort
	123	Albers	1	12345	Berlin
	234	Boehrs	1	12345	Berlin
	345	Dinkel	2	13456	Berlin
	456	Dinkels	2	13456	Berlin

Produkte	VkNr	Prio	Produkt	Umsatz
	123	1	Kühlschrank	61.000€
	123	2	Waschmaschine	59.000€
	234	1	TV	12.500€
	234	2	Kühlschrank	55.000€
	234	3	DVD	10.500€
	345	1	TV	56.000€
	456	1	Kühlschrank	12.000€

Normalformen



3. Normalform

- Ziel: Nur unmittelbar zusammengehörige Daten in einer Relation enthalten, die genau einen Sachverhalt ausdrückt
- Regel
 - Relation ist in 2. NF
 - vereinfacht: es gibt kein Nicht-Schlüsselattribut, das von anderen Nicht-Schlüsselattributen abhängig ist (Doppelpfeile gehen nur von Schlüsselkandidaten aus)
 - formal: jede Determinante ist ein Schlüsselkandidat
- Ausgangssituation (Beispiel): **Ist das 3. NF?**

Verkäufer	VkNr	Name	Filiale	Plz	Ort
	123	Albers	1	12345	Berlin
	234	Boehrs	1	12345	Berlin
	345	Dinkel	2	13456	Berlin
	456	Dinkels	2	13456	Berlin

Produkte	VkNr	Prio	Produkt	Umsatz
	123	1	Kühlschrank	61.000€
	123	2	Waschmaschine	59.000€
	234	1	TV	12.500€
	234	2	Kühlschrank	55.000€
	234	3	DVD	10.500€
	345	1	TV	55.000€
	456	1	Kühlschrank	12.000€

Normalformen



3. Normalform

- Ziel: Nur unmittelbar zusammengehörige Daten in einer Relation enthalten, die genau einen Sachverhalt ausdrückt
- Regel
 - Relation ist in 2. NF
 - vereinfacht: es gibt kein Nicht-Schlüsselattribut, das von anderen Nicht-Schlüsselattributen abhängig ist (Doppelpfeile gehen nur von Schlüsselkandidaten aus)
 - formal: jede Determinante ist ein Schlüsselkandidat
- Ausgangssituation (Beispiel): **Warum Verkäufer nicht 3. NF?**

Verkäufer	<u>VkNr</u>	Name	Filiale	Plz	Ort
	123	Albers	1	12345	Berlin
	234	Boehrs	1	12345	Berlin
	345	Dinkel	2	13456	Berlin
	456	Dinkels	2	13456	Berlin

Produkte	<u>VkNr</u>	<u>Prio</u>	Produkt	Umsatz
	123	1	Kühlschrank	61.000€
	123	2	Waschmaschine	59.000€
	234	1	TV	12.500€
	234	2	Kühlschrank	55.000€
	234	3	DVD	10.500€
	345	1	TV	56.000€
	456	1	Kühlschrank	12.000€

Ist 3. Normalform!

Normalformen



3. Normalform

- Ziel: Nur unmittelbar zusammengehörige Daten in einer Relation enthalten, die genau einen Sachverhalt ausdrückt
- Regel
 - vereinfacht: es gibt kein Nicht-Schlüsselattribut, das von anderen Nicht-Schlüsselattributen abhängig ist (Doppelpfeile gehen nur von Schlüsselkandidaten aus)
 - formal: jede Determinante ist ein Schlüsselkandidat
- Ausgangssituation (Beispiel): **Verkäufer ist nicht 3. NF!**

PLZ und Ort
abhängig von
Nummer der Filiale

Verkäufer	VkNr	Name	Filiale	Plz	Ort
	123	Albers	1	12345	Berlin
	234	Boehrs	1	12345	Berlin
	345	Dinkel	2	13456	Berlin
	456	Dinkels	2	13456	Berlin

Produkte	<u>VkNr</u>	<u>Prio</u>	Produkt	Umsatz
	123	1	Kühlschrank	61.000€
	123	2	Waschmaschine	59.000€
	234	1	TV	12.500€
	234	2	Kühlschrank	55.000€
	234	3	DVD	10.500€
	345	1	TV	56.000€
	456	1	Kühlschrank	12.000€

Ist 3. Normalform!



Normalformen

3. Normalform

- Ergebnis der Normalisierung
 - Weitere Relation Filiale
 - Verkäufer mit Fremdschlüssel für Filiale

Filiale	<u>FNr</u>	Plz	Ort
	1	12345	Berlin
	2	13456	Berlin

Verkäufer	<u>VkNr</u>	Name	<u>FNr</u>
	123	Albers	1
	234	Boehrs	1
	345	Dinkel	2
	456	Dinkels	2

Produkte	<u>VkNr</u>	<u>Prio</u>	Produkt	Umsatz
	123	1	Kühlschrank	61.000€
	123	2	Waschmaschine	59.000€
	234	1	TV	12.500€
	234	2	Kühlschrank	55.000€
	234	3	DVD	10.500€
	345	1	TV	56.000€
	456	1	Kühlschrank	12.000€



Normalformen

... ist das ein guter Entwurf? Bezogen auf die Ziele

- komfortable und schnelle Abfragemöglichkeiten
- leichte und einfache Änderbarkeit
- Redundanzfreiheit (um z.B. Speicherplatz zu sparen, Änderungen zu vereinfachen)

Filiale	<u>FNr</u>	Plz	Ort
	1	12345	Berlin
	2	13456	Berlin

Verkäufer	<u>VkNr</u>	Name	<u>FNr</u>
	123	Albers	1
	234	Boehrs	1
	345	Dinkel	2
	456	Dinkels	2

Produkte	<u>VkNr</u>	<u>Prio</u>	Produkt	Umsatz
	123	1	Kühlschrank	61.000€
	123	2	Waschmaschine	59.000€
	234	1	TV	12.500€
	234	2	Kühlschrank	55.000€
	234	3	DVD	10.500€
	345	1	TV	56.000€
	456	1	Kühlschrank	12.000€

Normalformen



... ist das ein guter Entwurf? Fortschritte erkennbar, aber Redundanzen erschweren z.B.:

- Änderung der Produktbezeichnung (mehrfache Änderung nötig)

Filiale	<u>FNr</u>	Plz	Ort
	1	12345	Berlin
	2	13456	Berlin

Verkäufer	<u>VkNr</u>	Name	<u>FNr</u>
	123	Albers	1
	234	Boehrs	1
	345	Dinkel	2
	456	Dinkels	2

Produkte	<u>VkNr</u>	<u>Prio</u>	Produkt	Umsatz
	123	1	Kühlschrank	61.000€
	123	2	Waschmaschine	59.000€
	234	1	TV	12.500€
	234	2	Kühlschrank	55.000€
	234	3	DVD	10.500€
	345	1	TV	56.000€
	456	1	Kühlschrank	12.000€

Normalformen



... ist das ein guter Entwurf?

Ja.



Filiale	<u>FNr</u>	Plz	Ort
	1	12345	Berlin
	2	13456	Berlin

Produkte	<u>PNr</u>	Produkt
	1	Kühlschrank
	2	Waschmaschine
	3	TV
	4	DVD

Verkäufer	<u>VkNr</u>	Name	<u>FNr</u>
	123	Albers	1
	234	Boehrs	1
	345	Dinkel	2
	456	Dinkels	2

Umsätze	<u>VkNr</u>	<u>Prio</u>	<u>PNr</u>	Umsatz
	123	1	1	61.000€
	123	2	2	59.000€
	234	1	3	12.500€
	234	2	1	55.000€
	234	3	4	10.500€
	345	1	3	56.000€
	456	1	1	12.000€



Normalformen

Weiterführende Normalformen

- 4. Normalform
- 5. Normalform

sind in diesem Kurs nicht relevant.



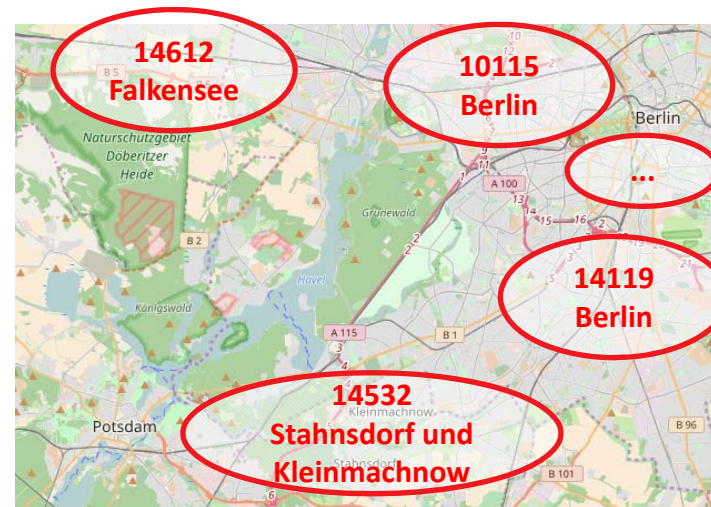
Hinweise

Häufig wird die Frage nach der Abhängigkeit von PLZ und Ort gestellt...

- Es gibt PLZ, zu denen genau ein Ort gehört (14612 Falkensee)
- Es gibt PLZ, zu denen mehrere Orte gehören (14532)
- Es gibt Orte, zu denen mehrere PLZ gehören (Berlin)

... deshalb gibt es keine funktionale Abhängigkeit des Ortes von der PLZ.

Gebiet	Plz	Ort
	10115	Berlin
	14119	Berlin
	14532	Stahnsdorf
	14532	Kleinmachnow
	14612	Falkensee



© OpenStreetMap-Mitwirkende



Hinweise

Ist die Tabelle Filiale wirklich in 3. Normalform?

- Zu jeder Filialnummer existiert höchstens eine Wertekombination aus PLZ und Ort, d.h. es liegt funktionale Abhängigkeit vor
- Filialnummer ist nicht zusammengesetzt, d.h. es liegt volle funktionale Abhängigkeit vor
- Es gibt keine funktionale Abhängigkeit des Ortes von der PLZ, d.h. FNr ist die einzige Determinante

Filiale	<u>FNr</u>	Plz	Ort
	1	10115	Berlin
	2	10115	Berlin
	3	14476	Potsdam
	4	14482	Potsdam
	5	14552	Michendorf
	6	14552	Saarmund
	7	14552	Fresdorf

Verschiedene Filiale in verschiedenen PLZ-Gebieten des gleichen Ortes

Verschiedene Filialen in einem PLZ-Gebiet des gleichen Ortes.

Verschiedene Filialen in einem PLZ-Gebiet verschiedener Ortes.

Hinweise



Funktionale Abhängigkeit

- kann anhand von Beispieldaten in einer Tabelle nur widerlegt werden
- kann nur anhand des Wissens über den Anwendungsbereich abgeleitet werden

Normalisierung unabhängig vom Ansatz



Exemplarischer Ausgangspunkt hier

- Sammlung zu speichernder Informationen und
- deren Zusammenfassung in Tabellen (Universalrelationen)

... und dennoch

- Normalisierung unabhängig vom gewählten Ansatz und auch bei strukturierter Ableitung aus ER-Modellen sinnvoll



Normalisierung abhängig vom Einsatzzweck

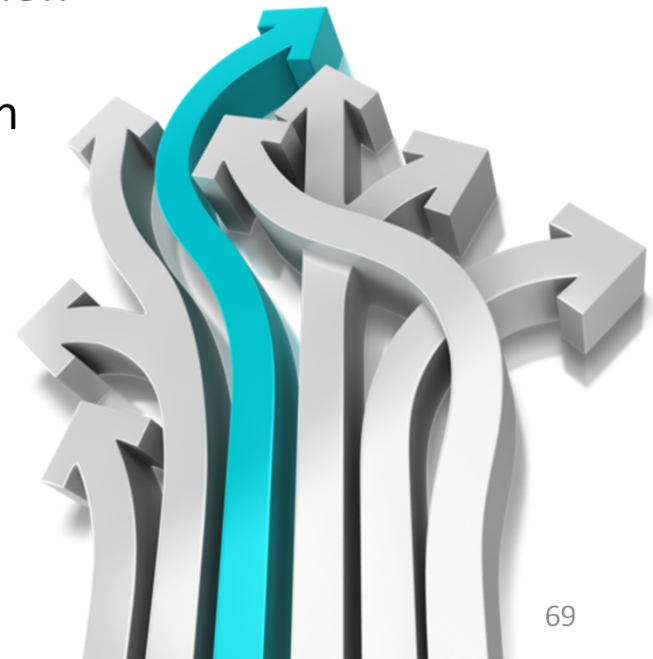


Normalisierung

- Ziel: Beseitigung von Anomalien (bei Einfügen, Löschen und Ändern) insbesondere durch Redundanzfreiheit
- führt zu Redundanzfreiheit durch Relationen mit wenigen Attributen
- Auswertung erfordert JOIN-Operationen über verschiedene normalisierte Relationen

Denormalisierung

- Umkehr der Normalisierung und Zusammenfassung von mehreren Relationen in einer gemeinsamen Relation
- sinnvoll zur Leistungsoptimierung, wenn Einfüge-, Lösch- und Änderungsoperationen selten
- Unterstützt insbesondere die Auswertung von Datenbeständen
- Beispiel: Data Warehouse
 - Faktentabellen mit Redundanzen
 - Dimensionstabellen denormalisiert als Sternschema





Inhalt

Ziel und Einordnung

Rückblicke

Einführung

Normalformen

- Ausgangssituation
- Erste Normalform
- Zweite Normalform und ihr formaler Hintergrund
- Dritte Normalform und ihr formaler Hintergrund
- Weitere Normalformen

Normalisierung

- Gewählter Ansatz
- Geplanter Einsatzzweck

Zusammenfassung

Ausblick

Zusammenfassung



Normalisierung als Prozess der anhand definierter Regeln, der Normalformen der verlustfreien Optimierung dient

Normalformen 1 bis 3¹⁾

- In Anlehnung an:
"Die Wahrheit, die ganze Wahrheit und nichts als die Wahrheit.
So wahr mir Gott helfe!"
- Bedeutet Normalisierung:
"Der Schlüssel, der ganze Schlüssel und nichts als der Schlüssel.
So wahr mir Codd helfe!"
 - alle atomaren Attribute beziehen sich auf den Schlüssel (1. NF)
 - sie beziehen sich auf den gesamten Schlüssel (2. NF)
 - die Attribute hängen nur vom Schlüssel ab und nicht von weiteren Attributen (3. NF)

1) nach Wikipedia: Normalisierung (Datenbanken). http://de.wikipedia.org/wiki/Normalisierung_%28Datenbank%29#Merkspruch

Zusammenfassung



Wichtig

- Relation mit nicht zusammengesetztem Primärschlüssel, die sich in 1. NF befindet, befindet sich auch in 2. NF
- Relation, die sich nicht in 2. NF befindet, hat einen zusammengesetzten Primärschlüssel

Antworten auf Frage: "In welcher NF ist Relation R?"

- Ausgehend von 1. NF prüfen
- Ist sie in 1. NF, dann weiter prüfen ob 2. NF
- Ist sie in 2. NF, dann weiter prüfen, ob 3. NF
- ...
- Abschließende Argumentation: Sie ist in n. NF, weil Sie ist aber nicht in n+1. NF, weil



Inhalt

Ziel und Einordnung

Rückblicke

Einführung

Normalformen

- Ausgangssituation
- Erste Normalform
- Zweite Normalform und ihr formaler Hintergrund
- Dritte Normalform und ihr formaler Hintergrund
- Weitere Normalformen

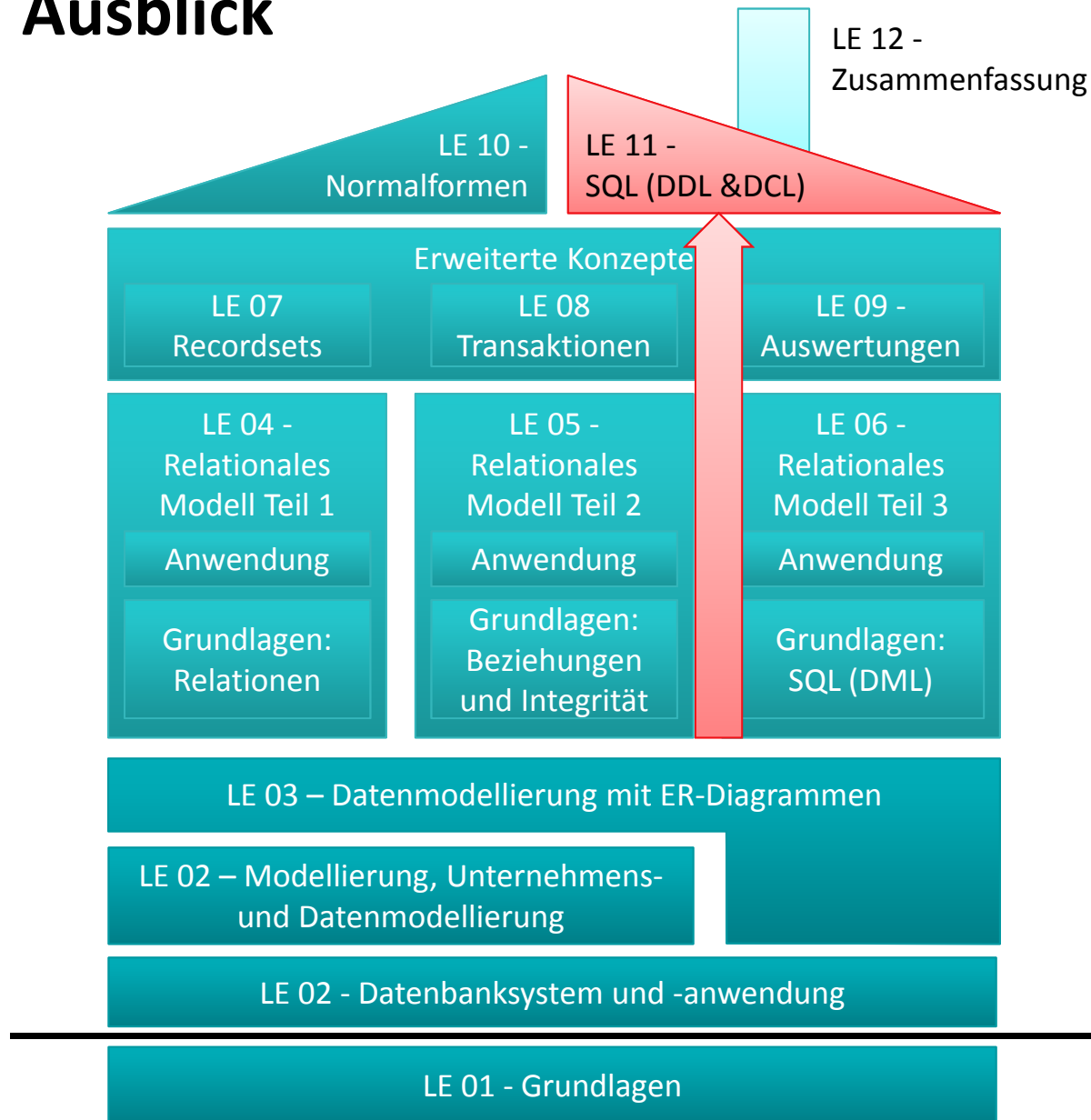
Normalisierung

- Gewählter Ansatz
- Geplanter Einsatzzweck

Zusammenfassung

Ausblick

Ausblick



Quellen



- [1] E. Schicker, E.: Datenbanken und SQL. Teubner, Stuttgart, 1996.
- [2] A. Fink, G. Schneiderreit, S. Voß: Grundlagen der Wirtschaftsinformatik. Physika (Springer), Heidelberg, 2001.
- [3] H. Sauer: Relationale Datenbanken. 5. Aufl., Addison-Wesley, München, 2002.
- [4] G. Vossen: Datenmodelle, Datenbanksprachen und Datenbankmanagementsysteme. 5. Aufl., Oldenbourg, München, 2008.
- [5] Wikipedia: Normalisierung (Datenbanken).
[http://de.wikipedia.org/wiki/Normalisierung %28Datenbank%29#Merkspruch](http://de.wikipedia.org/wiki/Normalisierung_%28Datenbank%29#Merkspruch)



BEUTH HOCHSCHULE FÜR TECHNIK BERLIN
University of Applied Sciences

Wirtschaftsinformatik 2

LE 10 – Normalformen

Prof. Dr. Thomas Off

<http://www.ThomasOff.de/lehre>