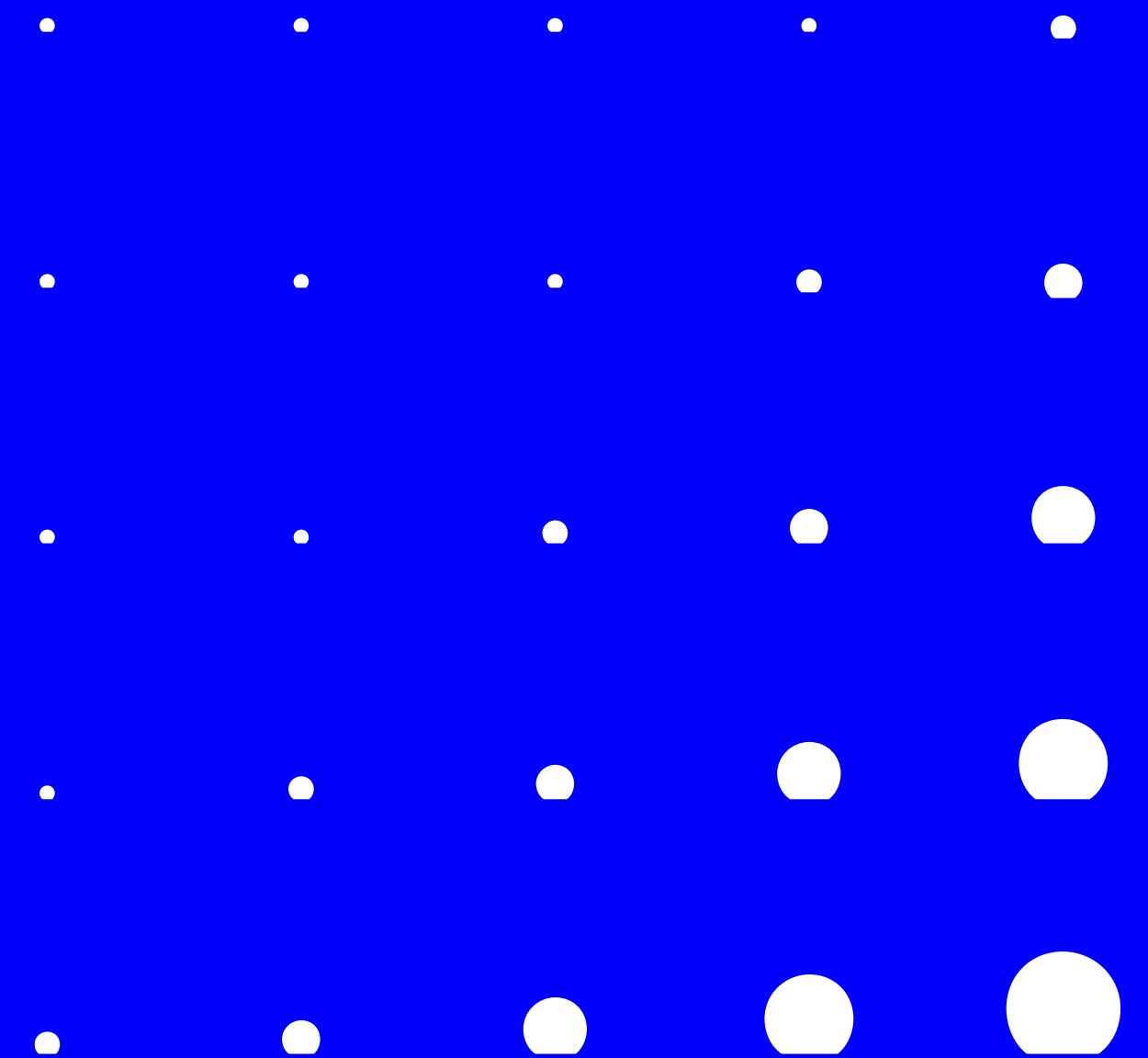


Normalisierung & Normalformen



Was ist Normalisierung?

- Unter Normalisierung einer relationalen Tabellenstruktur versteht man die Aufteilung von Attributen (Tabellenspalten) in mehrere Relationen (Tabellen) gemäß den Normalisierungsregeln.
- Ziel ist eine Form, die keine vermeidbaren Redundanzen mehr enthält.
 - Ein Datenbankschema mit Datenredundanzen kann dazu führen, dass bei Änderungen der damit realisierten Datenbank die mehrfach enthaltenen Daten nicht konsistent, sondern nur teilweise und unvollständig geändert werden.
 - Damit können sie obsolet oder widersprüchlich werden.
 - Man spricht dann von auftretenden Datenbank-Anomalien.
 - Zudem belegt mehrfache Speicherung derselben Daten unnötig Speicherplatz.

0NF bzw. keine Normalform

- Die 0NF ist dann gegeben, wenn alle Informationen in einer Tabelle vorhanden sind und noch unnormalisiert vorliegen.
- Die 0NF liegt in vielen Fällen während der Anforderungsanalyse einer Datenbank vor.
- Die Anforderungsanalyse in der Datenbankentwicklung beginnt mit dem Sammeln von unstrukturierten und unsortierten Informationen aus den verschiedenen Fachbereichen, zum Beispiel zur Entwicklung eines Warenwirtschaftssystems.

ONF bzw. keine Normalform

- Def.: „Eine Relationstyp (Tabelle) befindet sich in der nullten Normalform, wenn alle Datenelemente der realen Welt in einer Tabelle zusammen gefasst und aufgelistet sind.“
 - Die Tabelle ist dann in der „Nullten Normalform“, d. h. unnormalisiert.
- Ein häufiges Beispiel für eine Tabelle in der Nullten Normalform ist eine unnormalisierte Rechnungsinformation, die Stück für Stück normalisiert werden muss.
 - Während der Anforderungsanalyse in der Datenbankentwicklung ist folgende Rechnungsinformation aufgenommen worden...

0NF: Datenbeispiel

	1	2	3
Rechnungsnummer	101	101	102
Rechnungsdatum	01.01.2025	01.01.2025	02.01.2025
Zahlungsart	Visa	PayPal	PayPal
Kundennummer	1	1	2
Kudentyp	Privatkunde	Privatkunde	Geschäftskunde
Name	Julia Meier	Julia Meier	NULL
Firmenname	NULL	NULL	IT Solutions GmbH
Kundenanschrift	Gartenweg 3 50667 Köln	Gartenweg 3 50667 Köln	Hauptstraße 12 80331 München
Kreditkartennummer	4567123412345678	4567123412345678	NULL
Gültig bis	Aug 26	Aug 26	NULL
Produktnummer	1001	1002	1001
Produktname	Laptop	Maus	Laptop
Einzelpreis	1200.00	25.00	1200.00
Menge	2	3	1
Lieferanten-ID	1	2	1
Lieferantenname	TechSupply	GadgetWorld	TechSupply
Rabatt	5.00	0.00	5.00
Lieferantenanschrift	Holzweg 1 14232 Leipzig	Am Weiher 4b 51580 Reichshof	Holzweg 1 14232 Leipzig

1NF

- Die Erste Normalform (1NF) ist dann gegeben, wenn alle Informationen in einer Tabelle atomar vorliegen.
 - Es bedeutet, dass jede Information innerhalb einer Tabelle eine eigene Tabellenspalte bekommt.
 - Informationen, die vorher unstrukturiert und unsortiert vorlagen, werden nun einheitlich und klar strukturiert.

0NF -> 1NF

	1	2	3
Rechnungsnummer	101	101	102
Rechnungsdatum	01.01.2025	01.01.2025	02.01.2025
Zahlungsart	Visa	PayPal	PayPal
Kundennummer	1	1	2
Kudentyp	Privatkunde	Privatkunde	Geschäftskunde
Name	Julia Meier	Julia Meier	NULL
Firmenname	NULL	NULL	IT Solutions GmbH
Kundenanschrift	Gartenweg 3 50667 Köln	Gartenweg 3 50667 Köln	Hauptstraße 12 80331 München
Kreditkartennummer	4567123412345678	4567123412345678	NULL
Gültig bis	Aug 26	Aug 26	NULL
Produktnummer	1001	1002	1001
Produktname	Laptop	Maus	Laptop
Einzelpreis	1200.00	25.00	1200.00
Menge	2	3	1
Lieferanten-ID	1	2	1
Lieferantenname	TechSupply	GadgetWorld	TechSupply
Rabatt	5.00	0.00	5.00
Lieferantenanschrift	Holzweg 1 14232 Leipzig	Am Weiher 4b 51580 Reichshof	Holzweg 1 14232 Leipzig

0NF -> 1NF

	1	2	3
Rechnungsnummer	101	101	102
Rechnungsdatum	01.01.2025	01.01.2025	02.01.2025
Zahlungsart	Visa	PayPal	PayPal
Kundennummer	1	1	2
Kudentyp	Privatkunde	Privatkunde	Geschäftskunde
Name Vorname, Nachname	Julia, Meier	Julia, Meier	NULL
Firmenname	NULL	NULL	IT Solutions GmbH
Kundenanschrift Strasse, Hausnr, PLZ, Ort	Gartenweg, 3, 50667, Köln	Gartenweg, 3, 50667, Köln	Hauptstraße, 12, 80331, München
Kreditkartennummer	4567123412345678	4567123412345678	NULL
Gültig bis Monat, Jahr	Aug 26	Aug 26	NULL
Produktnummer	1001	1002	1001
Produktname	Laptop	Maus	Laptop
Einzelpreis	1200.00	25.00	1200.00
Menge	2	3	1
Lieferanten-ID	1	2	1
Lieferantenname	TechSupply	GadgetWorld	TechSupply
Rabatt	5.00	0.00	5.00
Lieferantenanschrift Strasse, Hausnr, PLZ, Ort	Holzweg, 1, 14232, Leipzig	Am Weiher, 4b, 51580, Reichshof	Holzweg, 1, 14232, Leipzig

2NF

- Def.: „Ein Relationstyp (Tabelle) befindet sich genau dann in der zweiten Normalform (2NF), wenn
 - er sich in der 1NF befindet und
 - wenn jedes einzelne Attribut vom vollen Primärschlüssel dieses Relationstyps abhängig ist.“
(...und nicht nur von einem Teil davon)
- Die 2NF ist ein wichtiger Schritt zu einer voll normalisierten relationalen Datenbank.
 - Sie wird meistens schon indirekt erreicht, wenn der Datenbankentwickler mit der Erstellung eines ER-Diagramms beschäftigt ist.
 - Gute Datenbankentwickler brauchen für die 2NF kein Modell auf dem Papier, sondern können Geschäftsprozesse direkt mit dem Kunden besprechen und zeitnah in einer Datenbankapplikation implementieren.

Schritte zur 2NF

- Identifiziere Primärschlüssel und ordne die restlichen Attribute diesen Primärschlüsseln zu.
 - Ein Primärschlüssel kann auch aus mehreren Attributen bestehen!
 - Die restlichen Attribute dürfen nicht nur von einem Teil eines zusammengesetzten Primärschlüssels abhängen.
 - Wenn Abhängigkeiten von nur einem Teil des Schlüssels bestehen, wird die Tabelle weiter zerlegt.

1NF -> 2NF: Primärschlüssel & Datengruppen

	1	2	3
Rechnungsnummer	101	101	102
Rechnungsdatum	01.01.2025	01.01.2025	02.01.2025
Zahlungsart	Visa	PayPal	PayPal
Kundennummer	1	1	2
Kudentyp	Privatkunde	Privatkunde	Geschäftskunde
Name Vorname, Nachname	Julia, Meier	Julia, Meier	NULL
Firmenname	NULL	NULL	IT Solutions GmbH
Kundenanschrift Strasse, Hausnr, PLZ, Ort	Gartenweg, 3, 50667, Köln	Gartenweg, 3, 50667, Köln	Hauptstraße, 12, 80331, München
Kreditkartennummer	4567123412345678	4567123412345678	NULL
Gültig bis Monat, Jahr	Aug 26	Aug 26	NULL
Produktnummer	1001	1002	1001
Produktname	Laptop	Maus	Laptop
Einzelpreis	1200.00	25.00	1200.00
Menge	2	3	1
Lieferanten-ID	1	2	1
Lieferantenname	TechSupply	GadgetWorld	TechSupply
Rabatt	5.00	0.00	5.00
Lieferantenanschrift Strasse, Hausnr, PLZ, Ort	Holzweg, 1, 14232, Leipzig	Am Weiher, 4b, 51580, Reichshof	Holzweg, 1, 14232, Leipzig

1NF -> 2NF: Primärschlüssel & Datengruppen

- Rechnung
 - Rechnungsnummer
 - Rechnungsdatum
 - Zahlungsart
 - Kundennummer

1NF -> 2NF: Primärschlüssel & Datengruppen

- Kunde
 - **Kundennummer**
 - Kundentyp
 - Name: Vorname | Nachname
 - Firmenname
 - Kundenanschrift: Straße | Hausnummer | PLZ | Ort

1NF -> 2NF: Primärschlüssel & Datengruppen

- **Kreditkartendaten**
 - **Kreditkartennummer**
 - Kundennummer
 - Gültig bis: Monat, Jahr
- **Produkt**
 - **Produktnummer**
 - Produktname
 - Einzelpreis

1NF -> 2NF: Primärschlüssel & Datengruppen

- **Lieferant**
 - **Lieferanten-ID**
 - Lieferantennamen
 - Lieferantenanschrift: Straße | Hausnummer | PLZ | Ort
- ...übrig bleiben „Menge“ und „Rabatt“ -> Gespräch mit dem Auftraggeber:
 - Eine Rechnung besteht aus Rechnungspositionen!
 - Jede Rechnungsposition besteht aus einem Produkt in einer gewissen Menge!
 - Ein Rabatt kann ein Lieferant geben auf einzelne Produkte ab einer gewissen Menge!
 - Ein Lieferant kann mehrere Produkte liefern und ein Produkt kann von mehreren Lieferanten geliefert werden!

1NF -> 2NF: Primärschlüssel & Datengruppen

- **Rechnungsposition**
 - Rechnungsnummer
 - Produktnummer (besser: laufende Nummer)
 - Menge [nur vom ganzen Schlüssel abhängig, daher 2NF]
- **ProduktRabatt**
 - Produktnummer
 - Lieferanten-ID
 - Menge [nur vom ganzen Schlüssel abhängig, daher 2NF]
 - Rabatt [nur vom ganzen Schlüssel abhängig, daher 2NF]

- Def.: „Ein Relationstyp befindet sich genau dann in der dritten Normalform (3NF), wenn
 - er sich in der 2NF befindet und
 - kein Nichtschlüsselattribut transitiv von einem anderen Nichtschlüsselattribut abhängt.
- Die 3NF ist das Ziel einer erfolgreichen Normalisierung in einem relationalen Datenbankmodell.
 - Sie verhindert einerseits Anomalien und Redundanzen in Datensätzen und andererseits bietet sie genügend Performance für SQL-Abfragen.
 - Die 3NF ist oft ausreichend, um die perfekte Balance aus Redundanz, Performance und Flexibilität für eine Datenbank zu gewährleisten.
 - Es existieren sogar eine 4NF und 5NF, die aber in der Praxis unbedeutend sind!

2NF -> 3NF: transitive Abhängigkeiten - Kunde

- Kunde
 - Kundennummer
 - Kundentyp
 - Name: Vorname | Nachname
 - Firmenname
 - Kundenanschrift: Straße | Hausnummer | PLZ | Ort
- Ob ein Name oder ein Firmenname existiert, ist vom Kundentyp abhängig und nicht vom Primärschlüssel!
 - Die jeweils anderen Werte sind NULL!

2NF -> 3NF: transitive Abhängigkeiten - Kunde

- **Kunde**
 - **Kundennummer**
 - Kundenanschrift: Straße | Hausnummer | PLZ | Ort
- **Privatkunde**
 - **Kundennummer**
 - Vorname
 - Nachname
- **Geschäftskunde**
 - **Kundennummer**
 - Firmenname

2NF -> 3NF: transitive Abhängigkeiten - Adressen

- **Kunde**
 - **Kundennummer**
 - **Kundenanschrift: Straße | Hausnummer | PLZ | Ort**
- **Lieferant**
 - **Lieferanten-ID**
 - **Lieferantenname**
 - **Lieferantenanschrift: Straße | Hausnummer | PLZ | Ort**
- Die Adressdaten sind voneinander abhängig und nicht von den Primärschlüsseln, sie tauchen sogar gebündelt in mehreren Relationstypen auf!

2NF -> 3NF: transitive Abhängigkeiten - Adressen

- Kunde
 - **Kundennummer**
 - Adress-ID
- Lieferant
 - **Lieferanten-ID**
 - Lieferantennamen
 - Adress-ID

2NF -> 3NF: transitive Abhängigkeiten - Adressen

- Adresse
 - ID
 - Straße
 - Hausnummer
 - PLZ
 - Ort
- Bei genauer Betrachtung gibt es noch eine Abhängigkeit zwischen PLZ und Ort, sodass man nur die PLZ in der Adresse speichern könnte und den PLZ-/Ort-Bezug in einer weiteren, separaten Tabelle!

Auswirkungen der Normalisierung

- Die Normalisierung hat ihre absolute Berechtigung darin,
 - redundante Datenspeicherung zu vermeiden und damit
 - Anomalien zu vermeiden.
- Die Normalisierung führt zu einer erhöhten Anzahl an Tabellen, die über Primär-/Fremdschlüssel miteinander verbunden sind.
 - Welche Auswirkung hat dieser Effekt, wenn Sie Suchanfragen auf den Datenbestand stellen müssen?