

2 Vorlauf: Datenbanken

2.1 Einleitung

- Eine DB ist eine selbständige und auf Dauer ausgelegte Datenorganisation, welche einen Datenbestand sicher und flexibel verwalten kann.
- Der Datenentwurf bestimmt, welche Daten maschinell verarbeitet und gespeichert werden
- Ziele des Datenentwurfs: Integration, Korrektheit, Generalisierung (Offenheit gegenüber Entwicklungen), Wirtschaftlichkeit
- Relationale contra hierarchische und objektorientierte DB

2.2 Datenbanksysteme

2.2.1 Eigenschaften einer DB

- Keine oder nur kontrollierte Redundanz
- Applikationen DB-unabhängig
- Flexibilität bezüglich neuer Applikationen
- Mehrbenutzerfähig
- Datenintegrität (korrekt und vollständig, konsistent)
- Datensicherheit (Backup, Recovery)
- Datenschutz (Zugriffsberechtigungen)
- Effizienz

2.2.2 Bestandteile einer DB

- DBMS: Kern der Datenbank, alle für die Datenverwaltung notwendigen Systemroutinen (Basisfunktionen wie Suchen, Lesen, Schreiben, Erzeugen, Löschen). Zugriff nur über definierte Schnittstellen, häufig ODBC als (technische) Standard-Schnittstelle.
- DB-Sprache: (Logische) Schnittstelle zwischen DB und Benutzer. Häufig SQL (Structured Query Language)
- Administrationsprogramme
- Programmiermöglichkeiten: Masken (Formulare), Berichte (Reports), Menüs, Import-/Export-Funktionen für Daten

2.3 Datenmodellierung

2.3.1 Entity-Relationship-Modell (Chen)

- Es gibt Regeln für die Strukturierung von Daten, damit sie bestimmte Kriterien erfüllen
- Die globale Datennormalisierung bezweckt, daß Redundanzen eliminiert werden und damit die Datenkonsistenz gewährleistet werden kann. Redundanzfreie Datenspeicherung bedeutet, daß kein Teil eines Datenbestandes weggelassen werden kann, ohne daß dies zu Informationsverlusten führt. Und auch: Redundanzfreiheit bedeutet, daß bei Datenänderungen nur an einer einzigen Stelle eingegriffen werden muß.
- Grundbegriffe

2.3.2 Relationen

- Vorteile sauberer Relationen (Ziele)
- Nachteile sauberer Relationen
- 16 theoretisch mögliche Beziehungstypen, von denen aber nur vier als zulässig übrig bleiben
- Rekursive Relationen: Vermeidung durch Transformation
- Mehrfachbeziehungen: Klarheit durch neue Primärschlüssel

- Spezialisierte Tabellen: zur Vermeidung von leeren Attributen
- Transaktion: Mehrere zusammenhängende Datenänderungen, die nur ganz oder gar nicht ausgeführt werden dürfen
- Referentielle Integrität: Prüfung auf zulässige Fremdschlüssel (Referenzierung)
- Indizierung: Prüfung auf Eindeutigkeit einfacher oder zusammengesetzter Primärschlüssel

Fazit

- Es sind durch Transformationen nur die hierarchischen Relationen als relevant übrig geblieben
- Bestimmte Relationen müssen durch eine Transaktionssteuerung abgesichert werden
- Schlüssel können redundant sein (als Fremdschlüssel) → unvermeidlich

2.3.3 Normalisierung

Wie erreicht man das Ziel der Redundanzfreiheit praktisch? Welche Attribute sind in welchen (ggfs. neuen) Tabellen zu speichern? Wie kann man die sog. Mutationsanomalien vermeiden?

Begriff der Abhängigkeit

- Funktionale Abhängigkeit: Aus dem Attribut A ergibt sich eindeutig das Attribut B
- Volle Abhängigkeit: Aus der Attribut**kombination** A, C ergibt sich eindeutig das Attribut B
- Transitiv Abhängigkeit: Aus A ergibt sich B und aus B ergibt sich C

Begriff des lokalen/globalen Attributes

- Lokales Attribut: Nur in einer Tabelle vorhanden und nicht Teil des Primärschlüssels. Darf nicht in anderen Tabellen auftauchen
- Globales Attribut: In mindestens einer Tabelle Teil des Primärschlüssels (und kann damit auch in anderen Tabellen auftauchen, aber nur als Teil eines Fremdschlüssels)

1. Normalform

Eine **Tabelle** befindet sich in der 1. Normalform, wenn alle Attribute nur einfache Attributwerte aufweisen, wobei auch Nullwerte zulässig sind.

2. Normalform (Voraussetzung: 1. Normalform)

Eine **Tabelle** mit zusammengesetztem Primärschlüssel (also mindestens zwei Schlüsselemente) befindet sich in der 2. Normalform, wenn die Nichtschlüssel-Attribute nur vom **gesamten** Primärschlüssel voll abhängig sind (und nicht schon von einem Teil des Primärschlüssels).

3. Normalform (Voraussetzung: 2. Normalform)

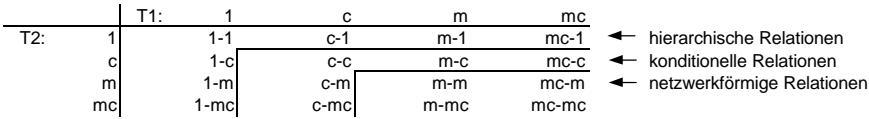
Eine **Tabelle** befindet sich in der 3. Normalform, wenn kein Nichtschlüssel-Attribut vom Primärschlüssel transitiv abhängig ist.

4. Normalform (Voraussetzung: 3. Normalform)

Eine **Datenbasis** befindet sich in der 4. Normalform, wenn nur noch lokale und globale Attribute existieren.

Relationen zwischen Tabellen

Mögliche Relationstypen



1-1



PNr	Name	BNr
1	Meier	2
2	Müller	4
3	Schulze	3
4	Schmidt	1
5	Lehmann	5

BNr	Titel
1	Schneller essen
2	Besser leben
3	Schöner schlafen
4	Lauter lachen
5	Höher sitzen

PNr	Name
1	Meier
2	Müller
3	Schulze
4	Schmidt
5	Lehmann

BNr	Titel	PNr
1	Schneller essen	4
2	Besser leben	1
3	Schöner schlafen	3
4	Lauter lachen	2
5	Höher sitzen	5

PNr	Name	Titel
1	Meier	Besser leben
2	Müller	Lauter lachen
3	Schulze	Schöner schlafen
4	Schmidt	Schneller essen
5	Lehmann	Höher sitzen

1-c



PNr	Name	BNr
1	Meier	2
2	Müller	
3	Schulze	3
4	Schmidt	1
5	Lehmann	

BNr	Titel
1	Schneller essen
2	Besser leben
3	Schöner schlafen

PNr	Name
1	Meier
2	Müller
3	Schulze
4	Schmidt
5	Lehmann

BNr	Titel	PNr
1	Schneller essen	4
2	Besser leben	1
3	Schöner schlafen	3

PNr	Name
1	Meier
2	Müller
3	Schulze
4	Schmidt
5	Lehmann

BNr	Titel
1	Schneller essen
2	Besser leben
3	Schöner schlafen

PNr	BNr
1	2
3	3
4	1

1-m



PNr	Name
1	Meier
2	Müller
3	Schulze
4	Schmidt
5	Lehmann

BNr	Titel	PNr
1	Schneller essen	4
2	Besser leben	1
3	Schöner schlafen	3
4	Lauter lachen	5
5	Höher sitzen	2
6	Tiefer fallen	1
7	Weiter springen	3
8	Frecher reden	1

1-mc

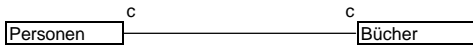


PNr	Name
1	Meier
2	Müller
3	Schulze
4	Schmidt
5	Lehmann

BNr	Titel	PNr
1	Schneller essen	4
2	Besser leben	1
3	Schöner schlafen	1
4	Lauter lachen	5
5	Höher sitzen	2
6	Tiefer fallen	1
7	Weiter springen	5
8	Frecher reden	1

Relationen zwischen Tabellen

c-c



PNr	Name	BNr
1	Meier	2
2	Müller	
3	Schulze	3
4	Schmidt	5
5	Lehmann	

BNr	Titel
1	Schneller essen
2	Besser leben
3	Schöner schlafen
4	Lauter lachen
5	Höher sitzen

PNr	Name
1	Meier
2	Müller
3	Schulze
4	Schmidt
5	Lehmann

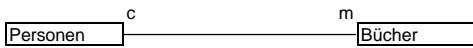
BNr	Titel	PNr
1	Schneller essen	1
2	Besser leben	
3	Schöner schlafen	3
4	Lauter lachen	
5	Höher sitzen	4

PNr	Name
1	Meier
2	Müller
3	Schulze
4	Schmidt
5	Lehmann

BNr	Titel
1	Schneller essen
2	Besser leben
3	Schöner schlafen
4	Lauter lachen
5	Höher sitzen

PNr	BNr
1	2
3	3
4	5

c-m



PNr	Name
1	Meier
3	Schulze
4	Schmidt

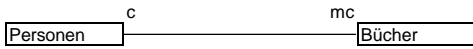
BNr	Titel	PNr
1	Schneller essen	
2	Besser leben	1
3	Schöner schlafen	3
4	Lauter lachen	
5	Höher sitzen	4
6	Tiefer fallen	1
7	Weiter springen	
8	Frecher reden	1

PNr	Name
1	Meier
3	Schulze
4	Schmidt

BNr	Titel
1	Schneller essen
2	Besser leben
3	Schöner schlafen
4	Lauter lachen
5	Höher sitzen
6	Tiefer fallen
7	Weiter springen
8	Frecher reden

PNr	BNr
1	2
1	6
1	8
3	3
4	5

c-mc



PNr	Name
1	Meier
2	Müller
3	Schulze
4	Schmidt
5	Lehmann

BNr	Titel	PNr
1	Schneller essen	
2	Besser leben	1
3	Schöner schlafen	3
4	Lauter lachen	
5	Höher sitzen	4
6	Tiefer fallen	1
7	Weiter springen	
8	Frecher reden	1

PNr	Name
1	Meier
2	Müller
3	Schulze
4	Schmidt
5	Lehmann

BNr	Titel
1	Schneller essen
2	Besser leben
3	Schöner schlafen
4	Lauter lachen
5	Höher sitzen
6	Tiefer fallen
7	Weiter springen
8	Frecher reden

PNr	BNr
1	2
1	6
1	8
3	3
4	5

m-m



PNr	Name
1	Meier
3	Schulze
4	Schmidt

BNr	Titel	PNr
1	Schneller essen	
2	Besser leben	1
3	Schöner schlafen	3
3	Schöner schlafen	4
4	Lauter lachen	
5	Höher sitzen	4
5	Höher sitzen	1

PNr	Name
1	Meier
3	Schulze
4	Schmidt

BNr	Titel
1	Schneller essen
2	Besser leben
3	Schöner schlafen
4	Lauter lachen
5	Höher sitzen

PNr	BNr
1	2
1	5
3	3
4	3
4	5

Normalisierung

Rohtabelle "Geschäftsdaten"

Kundenname	Adresse	Automarke	Typ	Seriennummer	Verkäufer	Datum
Meier	Planetenweg 7	VW	Golf	123456	Schmid	23.04.92
		Opel	Kadett	345678	Plüss	07.08.92
Müller	Altstadt 12	VW	Golf	388721	Frey	17.06.92
Steffen	Gartenstr. 7	VW	Polo	222245	Schmid	15.07.92
Steffen	Augasse 12	Audi	Quattro	122154	Frey	13.11.92
		Opel	Manta	445321		
					Schenk	

Kommentar:

Datensätze inhaltlich in mehrere Datensätze aufteilen, falls mehrfache Attributwerte auftreten, die (sozusagen durch Komma getrennt) in einer einzigen Tabellenzelle des ursprünglichen Datensatzes stehen (Beispiel Meier). Außerdem müssen die Sachverhalte klargestellt sein, die in der Tabelle beschrieben sind, die identifizierenden Attribute müssen eindeutig sein (wenn ich mehrere Kunden Steffen habe, dann ist der Name nicht eindeutig, also zur Identifikation nicht geeignet). Sind sie das nicht, müssen eindeutige Attribute eingeführt werden (neue Primärschlüssel). Hier: Einführung der KundenId notwendig, Autold und VerkäuferId sinnvoll.

1. Normalform

KundenId	Kundenname	Adresse	Autold	Automarke	Typ	Seriennummer	VerkäuferId	Verkäufer	Datum
1	Meier	Planetenweg 7	1	VW	Golf	123456	1	Schmid	23.04.92
1	Meier	Planetenweg 7	2	Opel	Kadett	345678	2	Plüss	07.08.92
2	Müller	Altstadt 12	3	VW	Golf	388721	3	Frey	17.06.92
3	Steffen	Gartenstr. 7	4	VW	Polo	222245	1	Schmid	15.07.92
4	Steffen	Augasse 12	5	Audi	Quattro	122154	3	Frey	13.11.92
			6	Opel	Manta	445321			
							4	Schenk	

Wenn ein Nicht-Schlüsselattribut von einem Teil des Primärschlüssels abhängig ist, dann läßt sich offensichtlich eine Tabelle auslagern, die mindestens dieses Nicht-Schlüsselattribut und den genannten Teil des Primärschlüssels enthält, von dem es abhängig ist. Der ursprüngliche Primärschlüssel ist also für manche Attribute zu umfassend definiert. Ziel ist also, eine Tabelle inhaltlich in mehrere Tabellen aufteilen, die unterschiedliche Entitätstypen beschreiben. Hier: Kunden und Autos auslagern, deren Attribute hängen immer nur von der KundenId bzw. der Autold ab.

2. Normalform

Kunden

KundenId	Kundenname	Adresse
1	Meier	Planetenweg 7
2	Müller	Altstadt 12
3	Steffen	Gartenstr. 7
4	Steffen	Augasse 12

Autos

Autold	Automarke	Typ	Seriennummer
1	VW	Golf	123456
2	Opel	Kadett	345678
3	VW	Golf	388721
4	VW	Polo	222245
5	Audi	Quattro	122154
6	Opel	Manta	445321

Verkäufe

KundenId	Autold	Datum	VerkäuferId	Verkäufer
1	1	23.04.92	1	Schmid
1	2	07.08.92	2	Plüss
2	3	17.06.92	3	Frey
3	4	15.07.92	1	Schmid
4	5	13.11.92	3	Frey

Hier werden die Beziehungen der Attribute einer Tabelle untereinander untersucht. Wenn es eine transitive Abhängigkeit gibt, dann kann eine weitere Tabelle ausgelagert werden.

Ziel ist also genauso wie bei der 2. Normalform, eine Tabelle inhaltlich in mehrere Tabellen aufteilen, die unterschiedliche Entitätstypen beschreiben, allerdings ist das Entscheidungskriterium ein anderes als bei der 2. Normalform. Hier: Der Verkäufername hängt von der VerkäuferId ab, die VerkäuferId hängt vom Primärschlüssel (KundenId, Autold) ab.

3. Normalform

Verkäufe

KundenId	Autold	Datum	VerkäuferId
1	1	23.04.92	1
1	2	07.08.92	2
2	3	17.06.92	3
3	4	15.07.92	1
4	5	13.11.92	3

Verkäufer

VerkäuferId	Verkäufer
1	Schmid
2	Plüss
3	Frey
4	Schenk

+ Kunden

+ Autos

wie in 2. Normalform