

5. НОРМАЛИЗАЦИЯ РЕЛЯЦИОННОЙ МОДЕЛИ

5.1. Функциональные зависимости

Функциональные зависимости – это особый вид ограничений целостности, а потому они предоставляют проектировщику способ отразить *семантические* зависимости между данными. Распознавание функциональных зависимостей представляет собой часть процесса выяснения смысла тех или иных данных. Например, тот факт, что поставщик проживает точно в одном городе можно отразить с помощью функциональной зависимости $\text{Номер поставщика} \rightarrow \text{Адрес поставщика}$

Адрес поставщика.

По сути, функциональная зависимость является *связью типа «многие-к-одному»* между множествами атрибутов внутри отношения. Дадим определение функциональной зависимости.

Пусть R является отношением, а X и Y – произвольные подмножества множества атрибутов отношения R . Тогда Y **функционально зависит** от X (или X функционально определяет Y) $X \rightarrow Y$ тогда и только тогда, когда каждое значение множества X отношения R связано в точности с одним значением множества Y отношения R . Иначе говоря, если два кортежа отношения R совпадают по значению X , они совпадают и по значению Y .

Например, отношение «Поставщик» удовлетворяет требованиям нескольких функциональных зависимостей:

$\{\text{Нº поставщика}\} \rightarrow \{\text{Имя поставщика}\}$

$\{\text{Нº поставщика}\} \rightarrow \{\text{Статус поставщика}\}$

$\{\text{Нº поставщика}\} \rightarrow \{\text{Адрес поставщика}\}$

$\{\text{Нº поставщика}\} \rightarrow \{\text{Имя поставщика, Статус поставщика, Адрес поставщика}\}$

$\{\text{Имя поставщика}\} \rightarrow \{\text{Адрес поставщика}\}$

$\{\text{Имя поставщика}\} \rightarrow \{\text{Статус поставщика}\}$

$\{\text{Имя поставщика}\} \rightarrow \{\text{Адрес поставщика, Статус поставщика}\}$

Левая и правая части символической записи функциональной зависимости называются *детерминантой* и *зависимой частью* соответственно. Как следует из определения, детерминант и зависимая часть являются множествами атрибутов. Если множество содержит один атрибут, оно называется *одноэлементным множеством* и тогда фигурные скобки могут быть опущены:

$\text{Нº поставщика} \rightarrow \text{Имя поставщика}$

Утверждение, что данная функциональная зависимость выполняется всегда, для всех возможных значений отношения, является *ограничением целостности* для отношения, а это значит, что при этом накладываются определенные ограничения на все *его* допустимые значения. Даже если ограничиться рассмотрением функциональных

зависимостей, имеющих место в любой момент, полный набор функциональных зависимостей, выполняющихся для всех допустимых значений заданной переменной-отношения, может быть очень большим. Необходимо минимизировать множество функциональных зависимостей до некоторых пределов. Задача поиска такого подходящего множества представляет практический интерес.

Очевидным способом сокращения существующего набора функциональных зависимостей является исключение из него **тривиальных** зависимостей, т.е. зависимостей, которые не могут не выполняться. Функциональная зависимость является **тривиальной** тогда и только тогда, когда правая часть ее символической записи является подмножеством левой части.

Одни функциональные зависимости подразумевают другие зависимости. Для данного множества зависимостей S **замыканием** S^+ называется множество всех функциональных зависимостей, подразумеваемых зависимостями множества S . Полную и исчерпывающую основу для вычисления замыкания S^+ для заданного множества S обеспечивают **правила логического вывода (аксиомы) Армстронга**.

Пусть заданы A, B, C – произвольные подмножества множества атрибутов отношения R . Тогда правила вывода Армстронга определяются следующим образом:

1. Правило **рефлексивности**: если B является подмножеством множества A , то $A \rightarrow B$.
2. Правило **дополнения**: если $A \rightarrow B$, то $AC \rightarrow BC$. (Здесь запись AC означает объединение множеств A и C).
3. Правило **транзитивности**: если $A \rightarrow B$ и $B \rightarrow C$, то $A \rightarrow C$.

Каждое из этих правил можно доказать непосредственно на основе определения функциональной зависимости. Из приведенных правил можно вывести несколько дополнительных правил, полезных для практического вычисления замыкания S^+ .

4. Правило **самоопределения**: $A \rightarrow A$.
5. Правило **декомпозиции**: если $A \rightarrow BC$, то $A \rightarrow B$ и $A \rightarrow C$.
6. Правило **объединения**: если $A \rightarrow B$ и $A \rightarrow C$, то $A \rightarrow BC$.
7. Правило **композиции**: если $A \rightarrow B$ и $C \rightarrow D$, то $AC \rightarrow BD$.

Кроме того, Дарвен доказал правило, называемое **всеобщей теоремой объединения**.

8. Если $A \rightarrow B$ и $C \rightarrow D$, то $A \setminus (C-B) \rightarrow BD$.

Приведенные правила являются *полными* в том смысле, что для заданного множества функциональных зависимостей S минимальный набор функциональных зависимостей, которые подразумевают все зависимости из множества S , может выведен из функциональных зависимостей множества S на основе этих правил. Они также являются

исчерпывающими, поскольку никакие дополнительные функциональные зависимости с их помощью не могут быть выведены. Иначе говоря, все эти правила могут быть использованы для получения замыкания S^+ .

Пусть задано отношение R и его некоторое множество функциональных зависимостей. Удобнее всего представлять это множество функциональных зависимостей в виде диаграммы функциональных зависимостей. Например диаграмма функциональных зависимостей отношения «Поставщик» представлена на Рис.11.

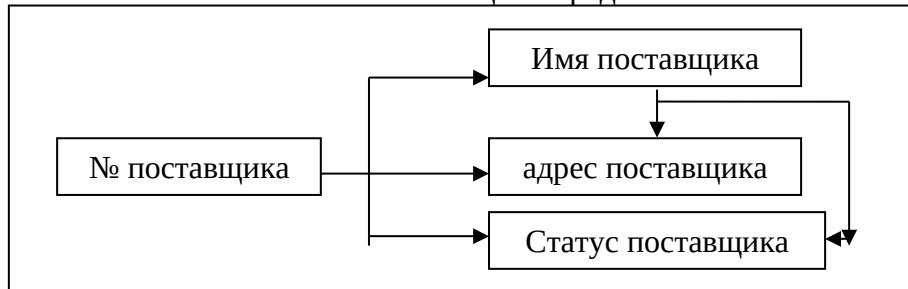


Рис.11. Диаграмма функциональных зависимостей отношения «Поставщик»

Каждая стрелка начинается с детерминанта, а ведет к атрибутам зависимой части отношения.

5.2. Суть процесса нормализации

Реляционная схема отношений содержит как структурную, так и семантическую информацию. Структурная информация связана с объявлением отношения, а семантическая выражается множеством известных функциональных зависимостей между атрибутами отношений. Однако некоторые функциональные зависимости могут быть нежелательными из-за побочных эффектов или аномалий, которые они вызывают при модификации базы данных. В связи с этим возникает вопрос о корректности представленной схемы отношений.

Корректной считается схема отношений, в которой отсутствуют нежелательные функциональные зависимости. В противном случае приходится прибегать к процедуре, называемой *декомпозицией* (разложением), при которой данное множество отношений заменяется другим множеством отношений, являющихся проекциями первых. Цель этой процедуры – устранение нежелательных функциональных зависимостей, а, значит, аномалий операций обновления данных. В этом заключается *суть процесса нормализации*. Иначе говоря, **нормализация** – это пошаговый обратимый процесс замены данной схемы (совокупности) отношений другой схемой, в которой эти отношения имеют более простую, регулярную форму. В теории нормальных форм определяются различные нормальные формы, которые ограничивают типы допустимых функциональных зависимостей.

5.3. Нормальные формы

Нормализация – это формальный метод анализа отношений на основе их первичного ключа (или потенциальных ключей) и существующих функциональных зависимостей. Нормализация включает ряд правил, которые могут использоваться для проверки отдельных отношений таким образом, чтобы вся база данных могла быть нормализована до желаемой степени. Если некоторое требование не удовлетворяется, то нарушающее данное требование отношение должно быть *декомпозировано* на отношения, каждое из которых (в отдельности) удовлетворяет всем требованиям нормализации.

Нормализация осуществляется в несколько последовательно выполняющихся этапов, каждый из которых соответствует некоторой нормальной форме, обладающей известными свойствами. В ходе нормализации структура отношений становится все более строгой и менее уязвимой по отношению к аномалиям обновления. Схема взаимосвязей между отдельными нормальными формами приведена на

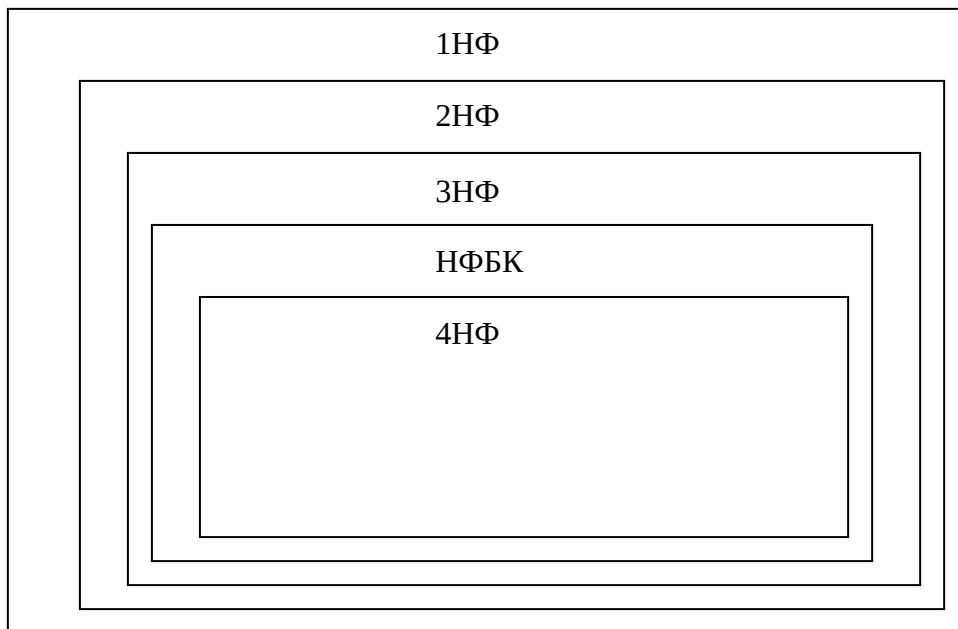


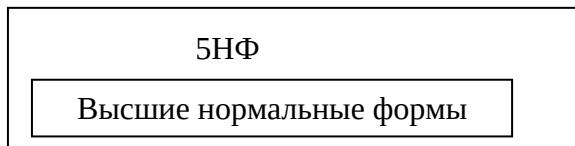
Рис. 10.

Рис. 10. Уровни нормализации

Из схемы следует, что все нормализованные отношения находятся в 1НФ, некоторые из них также находятся в 2НФ, а некоторые отношения в 2НФ также находятся в 3НФ и т.д.

Первые три нормальные формы (1НФ, 2НФ, 3НФ) были предложены Коддом (Codd), в определении последующих принимали также участие Бойсе (Boyce) (НФБК) и Фейгин (Fagin) (4НФ, 5НФ).

5.3.1. Первая нормальная форма (1НФ)



Отношение находится в **1НФ**, если значения всех его атрибутов простые, атомарные, т.е. не являются множеством или повторяющейся группой.

При проектировании осуществляется преобразования данных из формата источника в формат ненормализованной таблицы. **Ненормализованной** таблицей является таблица, содержащая повторяющиеся группы данных. Повторяющейся группой называется группа, состоящая из одного или более атрибутов таблицы, в которой возможно наличие нескольких значений для единственного значения ключа. Например, ненормализованной является таблица Табл.4.

№ рейса	Пункт отправления	Пункт назначения	День недели	Время вылета
A-45	Краснодар	Москва	понедельник	8-30
			среда	8-30
			пятница	9-00
A-67	Краснодар	Санкт-Петербург	суббота	17-50
			воскресенье	17-50

Табл.4. Ненормализованная таблица «Рейсы»

Существует простой способ приведения ненормализованной таблицы к 1НФ путем «размножения» значений атрибутов, т.е. ввода данных в каждую строку и каждый столбец отношения. В результате в полученное отношение вносится некоторая избыточность данных, которая в ходе дальнейшей нормализации будет устранена. Результат приведения отношения «Рейсы» к 1НФ отражен в Табл.5.

№ рейса	Пункт отправления	Пункт назначения	День недели	Время вылета
A-45	Краснодар	Москва	понедельник	8-30
A-45	Краснодар	Москва	среда	8-30
A-45	Краснодар	Москва	пятница	9-00
A-67	Краснодар	Санкт-Петербург	суббота	17-50
A-67	Краснодар	Санкт-Петербург	воскресенье	17-50

Табл.5. Нормализованная таблица «Рейсы» (1НФ)

5.3.2. Вторая нормальная форма (2НФ)

Вторая нормальная форма (2НФ) основана на понятии полной функциональной зависимости.

В некотором отношении множество атрибутов **В** называется **полностью функционально зависимым** от множества атрибутов **А**, если множество **В** функционально зависит от полного значения множества **А** и не зависит ни от какого подмножества значений атрибутов **А**.

Отношение является **отношением** во **2НФ**, если оно находится в первой нормальной форме и каждый неключевой атрибут **функционально полно** зависит от первичного ключа.

Рассмотрим пример отношения «Поставки», приведенного в Табл.6.

<i>№ поставщика</i>	<i>Товар</i>	<i>Цена</i>
П1	лимоны	35
П1	апельсины	22
П2	лимоны	35
П2	мандарины	27

Табл. 6. Отношение «Поставки»

Известно, что поставщики могут поставлять разные товары, при этом один и тот же товар могут поставлять разные поставщики. Цена на один и тот же товар фиксирована. Семантические ограничения можно отразить при помощи функциональных зависимостей. Представим их диаграммой функциональных зависимостей, приведенной на Рис. 11.

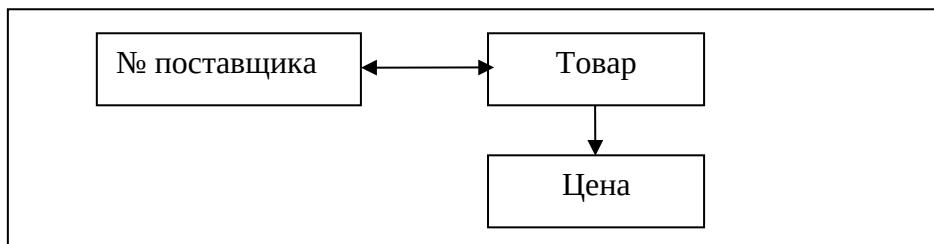


Рис.11. Диаграмма функциональных зависимостей отношения «Поставки»

В этом отношении первичным ключом является комбинация атрибутов «№ поставщика», «Товар». Очевидно, что в отношении «Поставки» имеет место неполная функциональная зависимость атрибута «Цена» от первичного ключа, поскольку атрибут «Цена» зависит только от атрибута «Товар». В отношении «Поставки» имеют место аномалии, выявляемые при выполнении операций манипулирования данными:

- ◆ **добавления** – при добавлении данных о поставке некоторым поставщиком товара, необходимо указать фиксированную цену товара;
- ◆ **обновления** – если меняется цена на некий товар, необходимо просмотреть все кортежи отношения, поскольку один и тот же товар может поставляться многими поставщиками;
- ◆ **удаления** – при удалении кортежа с некоторой поставкой можно потерять информацию о цене товара, если этот товар поставлялся единственным поставщиком.

Очевидной причиной аномалий в отношении является неполная функциональная зависимость атрибута «Цена» от первичного ключа. Чтобы привести отношение «Поставки» ко **2НФ** необходима декомпозиция на два отношения, приведенные на Рис 12.

Функциональные зависимости, описанные диаграммой (Рис. 11), очевидно, сохранены. Т.е. декомпозиция выполнена без потерь. Каждое полученное отношение находится во 2НФ, поскольку первичные ключи в отношении – простые.

Поставки товаров

№ поставщика	Товар
П1	лимоны
П1	апельсины
П2	лимоны
П2	мандарины

Цена товара

Товар	Цена
лимоны	35
апельсины	22
мандарины	27

Рис.12. Отношения «Поставки товаров» и «Цена товара»

5.3.3. Третья нормальная форма (ЗНФ)

Третья нормальная форма (ЗНФ) основана на понятии транзитивной функциональной зависимости.

Если для атрибутов А, В и С некоторого отношения существуют зависимости вида $A \rightarrow B$ и $B \rightarrow C$, то говорят, что атрибут С **транзитивно** зависит от атрибута А через атрибут В (при этом А функционально не зависит ни от В, ни от С).

Отношение **находится** в ЗНФ, если оно находится во 2НФ и каждый неключевой атрибут **нетранзитивно** зависит от первичного ключа.

Рассмотрим пример приведения отношения «Хранение товаров» (Табл. 7) к ЗНФ.

Фирма	Склад	Объем
Сонар-система	C1	100000
Пионер	C2	20000
Цель	C1	100000
Эльдорадо	C2	20000

Табл.7. Отношение «Хранение товаров»

Известно, что каждая фирма может хранить товары только на одном складе, но на одном и том же складе могут храниться товары, принадлежащие нескольким фирмам. Естественно, каждый склад имеет определенный объем (m^3). Семантические ограничения описанного примера можно отразить при помощи функциональных зависимостей. Представим их диаграммой функциональных зависимостей, приведенной на Рис. 13.

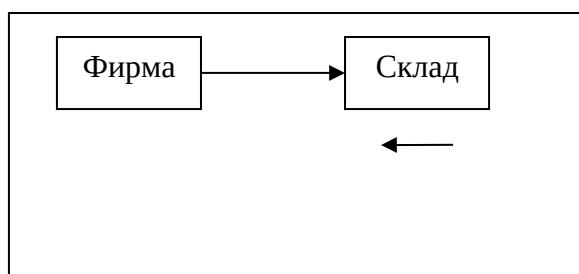




Рис.13. Диаграмма функциональных зависимостей отношения «Хранение товаров»

Атрибут «Фирма» является первичным ключом отношения «Хранение товаров». В данном отношении имеется транзитивная зависимость атрибута «Объем» от атрибута «Склад». В связи с этим имеют место аномалии, выявляемые при выполнении операций манипулирования данными:

- ◆ **добавления** – при добавлении данных о хранении некоторой фирмой товаров на складе требуется также указать фиксированный объем данного склада;
- ◆ **обновления** – если меняется объем некоторого склада (например, в результате реконструкции), необходимо просмотреть все кортежи отношения, поскольку на одном и том же складе могут храниться товары многих фирм;
- ◆ **удаления** – в случае, когда некоторая фирма перестала хранить товары на складе, вместе с удалением соответствующего кортежа можно потерять информацию об объеме склада, если на данном складе хранились товары только этой фирмы.

Причиной аномалий в отношении «Хранение товаров» является транзитивная зависимость атрибута «Объем» от не ключевого атрибута «Склад». Чтобы привести отношение «Хранение товаров» к ЗНФ необходима декомпозиция на два отношения, приведенные на Рис 14. Функциональные зависимости, описанные диаграммой (Рис. 13), очевидно, сохранены. Т.е. декомпозиция выполнена без потерь. Каждое полученное отношение находится в ЗНФ, поскольку в каждом отношении имеется только два атрибута, один из которых – первичный ключ.

Склады

Фирма	Склад
Сонар-система	C1
Пионер	C2
Цель	C1
Эльдорадо	C2

Объемы складов

Склад	Объем
C1	100000
C2	20000

Рис.14. Отношения «Склады» и «Объемы складов»

5.3.4. Нормальная форма Бойса-Кодда (НФБК)

Нормальная форма Бойса-Кодда (НФБК) учитывает функциональные зависимости, в которых участвуют все потенциальные ключи отношения, а не только его первичный ключ. Для отношения с единственным потенциальным ключом его ЗНФ и НФБК являются эквивалентными.

Отношение находится в **НФБК** тогда и только тогда, когда каждый его **дeterminант** является потенциальным ключом.

Нарушения требований НФБК происходят крайне редко, поскольку это возможно в следующих случаях:

- ◆ имеются два или более составных потенциальных ключа;
- ◆ эти потенциальные ключи перекрываются, т.е. ими совместно используется, по крайней мере, один общий атрибут.

Рассмотрим пример приведения отношения «Проекты» (Табл. 8) к НФБК.

Наименование проекта	№ детали	№ поставщика
Сонар-система	C1	10
Сонар-система	C2	20
Пионер	C2	30
Пионер	C1	40
Эльдорадо	C2	50

Табл.8. Отношение «Проекты»

Известно, что в проекте используется несколько деталей, но каждая деталь проекта при этом поставляется только одним поставщиком. Каждый поставщик обслуживает единственный проект, но проекты могут обеспечиваться несколькими поставщиками (разных деталей). Семантические ограничения отношения можно отразить при помощи функциональных зависимостей. Представим их диаграммой функциональных зависимостей, приведенной на Рис. 15.

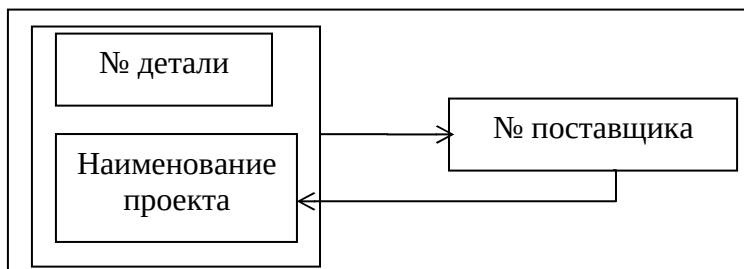


Рис.15. Диаграмма функциональных зависимостей отношения «Проекты»

В качестве первичного ключа отношения «Проекты» принята комбинация атрибутов «Наименование проекта», «№ детали». Данное отношение находится в 1НФ, 2НФ, 3НФ и не находится в НФБК. В связи с этим имеют место аномалии, выявляемые при выполнении операций манипулирования данными, например аномалии:

- ◆ **добавления** – при добавлении данных о некотором поставщике необходимо создавать фиктивный проект, а также деталь для этого поставщика

- ◆ **изменения** – при изменения наименования проекта в некотором кортеже, также необходимо просмотреть все отношение и поменять наименование проекта для всех используемых деталей.
- ◆ **удаления** – при удалении информации о некотором поставщике, удаляется также информация о деталях, которые он поставлял.

Причиной аномалий в отношении «Проекты» является то, что не все его детерминанты являются потенциальными ключами (это относится к атрибуту «№ поставщика». Для приведения отношения «Проекты» к НФБК необходима декомпозиция на два отношения, приведенные на Рис 16. Функциональные зависимости, описанные диаграммой (Рис. 15), очевидно, сохранены. Т.е. декомпозиция выполнена без потерь. Каждое полученное отношение находится в НФБК, поскольку в каждом отношении имеется только два атрибута, один из которых – первичный ключ.

Поставщики-детали

№ поставщика	№ детали
10	C1
20	C2
30	C2
40	C1
50	C2

Проекты –поставщики

№ поставщика	Наименование Проекта
10	Сонар-система
20	Сонар-система
30	Пионер
40	Пионер
50	Эльдараадо

Рис.16. Отношения «Поставщики-детали» и «Проекты-поставщики»

5.3.5. Четвертая нормальная форма (4НФ)

Многозначные зависимости (multi-valued dependency) возникают, например, когда необходимо привести к первой нормальной форме отношение, с независимыми многозначными атрибутами. Пусть имеется два таких атрибута Y и Z. Тогда для получения 1НФ необходимо для каждого набора значений остальных атрибутов X повторить эту строку для каждого сочетания атомарного значения Y с каждым атомарным значением Z, что называется выравниванием таблицы.

Образуется многозначная зависимость в которой:

- каждому значению X соответствует набор значений Y;
- каждому значению X соответствует набор значений Z;
- значения атрибутов Y и Z не зависят один от другого.

Многозначную зависимость принято обозначать $X \rightarrow\rightarrow Y|Z$, хотя можно было бы указать наличие двух существующих одновременно обычных функциональных зависимостей $X \rightarrow Y$ и $X \rightarrow Z$. Иногда обозначают многозначную зависимость $X \rightarrow\rightarrow Y$ или $X \rightarrow\rightarrow Z$.

Определение 1

MV-зависимость $X \rightarrow\rightarrow Y$ называется тривиальной если $X \subseteq Y$, либо $X \cup Y = \{X, Y, Z\}$.

Определение 2:

Пусть r – отношение со схемой $R(S)$, а X, Y, Z - непересекающиеся множества его атрибутов, такие, что $X \cup Y \cup Z = S$. Атрибуты Y и Z многозначно зависят от X (обозначение $X \rightarrow \rightarrow Y|Z$) если из того, что в отношении r содержатся кортежи $r_1 = (x, y, z_1)$ и $r_2 = (x, y_1, z)$ следует, что в отношении r содержится также кортеж $r_3 = (x, y, z)$.

Замечание 1: По симметрии определения в r содержится и кортеж $r_4 = (x, y_1, z_1)$. Атрибуты Y и Z как бы симметричны по отношению к X .

Замечание 2: При наличии MV-зависимости кортежи обязаны вставляться и удаляться одновременно **целыми наборами**.

Определение 3.

Переменная отношения R находится в четвёртой нормальной форме, если она находится в НФБК и все нетривиальные многозначные зависимости фактически являются функциональными зависимостями от её потенциальных ключей.

Эквивалентная формулировка определения:

Переменная отношения R находится в четвёртой нормальной форме тогда и только тогда, когда в случае существования подмножеств A и B атрибутов этой переменной отношения R , для которых выполняется нетривиальная многозначная зависимость $A \rightarrow \rightarrow B$, все атрибуты переменной отношения R также функционально зависят от A .

Пример:

Предположим, что рестораны производят разные виды пиццы, а службы доставки ресторанов работают только в определенных районах города. Составной первичный ключ соответствующей переменной отношения включает три атрибута: {Ресторан, Вид пиццы, Район доставки}.

Такая переменная отношения не соответствует 4НФ, так как существует следующая многозначная зависимость:

$$\begin{aligned}\{\text{Ресторан}\} &\rightarrow \rightarrow \{\text{Вид пиццы}\} \\ \{\text{Ресторан}\} &\rightarrow \rightarrow \{\text{Район доставки}\}\end{aligned}$$

То есть, например, при добавлении нового вида пиццы придется внести по одному новому кортежу для каждого района доставки. Возможна логическая аномалия, при которой определенному виду пиццы будут соответствовать лишь некоторые районы доставки из обслуживаемых рестораном районов.

Для предотвращения аномалии нужно декомпозировать отношение, разместив независимые факты в разных отношениях. В данном примере следует выполнить декомпозицию на {Ресторан, Вид пиццы} и {Ресторан, Район доставки}.

Однако, если к исходной переменной отношения добавить атрибут, функционально зависящий от потенциального ключа, например

цену с учётом стоимости доставки (<{Ресторан, Вид пиццы, Район доставки} → Цена), то полученное отношение будет находиться в 4НФ и его уже нельзя подвергнуть декомпозиции без потерь (смотри ниже). Указанные выше многозначные зависимости в данном случае называются внедрёнными зависимостями.

5.3.6. Пятая нормальная форма (5НФ)

Декомпозиция без потерь

Декомпозицией отношения R называется замена R на совокупность отношений $\{R_1, R_2, \dots, R_n\}$ такую, что каждое из них есть проекция R , и каждый атрибут R входит хотя бы в одну из проекций декомпозиции.

Например, для отношения R с атрибутами $\{a, b, c\}$ существуют следующие основные варианты декомпозиции:

$$\begin{aligned} &\{a\}, \{b\}, \{c\} \\ &\{a\}, \{b, c\} \\ &\{a, b\}, \{c\} \\ &\{b\}, \{a, c\} \\ &\{a, b\}, \{b, c\} \\ &\{a, b\}, \{a, c\} \\ &\{b, c\}, \{a, c\} \\ &\{a, b\}, \{b, c\}, \{a, c\} \end{aligned}$$

Рассмотрим теперь отношение R' , которое получается в результате операции естественного соединения, применённой к отношениям, полученным в результате декомпозиции R .

Декомпозиция называется декомпозицией без потерь, если R' в точности совпадает с R .

Неформально говоря, при декомпозиции без потерь отношение «разделяется» на отношения-проекции таким образом, что из полученных проекций возможна «сборка» исходного отношения с помощью операции естественного соединения.

Далеко не всякая декомпозиция является декомпозицией без потерь. Проиллюстрируем это на примере отношения R с атрибутами $\{a, b, c\}$, приведённом выше. Пусть отношение R имеет вид:

R

a	b	c
a_1	b_1	c_1
a_1	b_2	c_2
a_2	b_3	c_1
a_2	b_2	c_2

Декомпозиция $R_1 = \{a, c\}$, $R_2 = \{b, c\}$ имеет вид:

R_1

R_2

a	b
-----	-----

a_1	c_1
a_1	c_2
a_2	c_1
a_2	c_2

b	c
b_1	c_1
b_2	c_2
b_3	c_1

Результат операции соединения этих отношений:

$$R' = \text{join}_{ic}(R_1, R_2)$$

a	b	c
a_1	b_1	c_1
a_1	b_2	c_2
a_2	b_3	c_1
a_2	b_2	c_2
a_1	b_3	c_1
a_2	b_1	c_1

Очевидно, что R' не совпадает с R , а значит такая декомпозиция не является декомпозицией без потерь. Рассмотрим теперь декомпозицию $R_1 = [a, b], R_2 = \{b, c\}$:

$$R_1 R_2$$

a	b
a_1	b_1
a_1	b_2
a_2	b_3
a_2	b_2

b	c
b_1	c_1
b_2	c_2
b_3	c_1

Такая декомпозиция является декомпозицией без потерь, в чём читатель может убедиться самостоятельно.

В некоторых случаях отношение вовсе невозможно декомпозировать без потерь. Существуют также примеры отношений, для которых нельзя выполнить декомпозицию без потерь на две проекции, но которые можно подвергнуть декомпозиции без потерь на три или большее количество проекций.

Зависимость соединения.

Пусть R — переменная отношения, а A, B, \dots, Z — некоторые подмножества множества её атрибутов.

Если декомпозиция любого допустимого значения R на отношения, состоящие из множеств атрибутов A, B, \dots, Z , является декомпозицией без потерь, говорят, что переменная отношения R удовлетворяет зависимости соединения $*\{A, B, \dots, Z\}$.

Иными словами, переменная отношения R удовлетворяет зависимости соединения $*\{A, B, \dots, Z\}$ тогда и только тогда, когда любое допустимое значение переменной отношения R эквивалентно соединению её проекций по подмножествам A, B, \dots, Z множества атрибутов.

Подобно тому, как функциональная зависимость есть частный случай многозначной зависимости, многозначная зависимость является частным случаем зависимости соединения. Зависимость соединения является предельным обобщением понятий многозначной и функциональной зависимости, то есть это наиболее общая форма зависимости между атрибутами отношения.

Важно понимать, что зависимость соединения определяется не для конкретного значения переменной отношения в тот или иной момент времени, а по всем возможным значениям. Поэтому понятие зависимости соединения определено не для отношения (конкретного значения), а для переменной отношения. Зависимость соединения определяется не механически по текущим значениям, а следует из внешнего знания о природе и закономерностях данных, которые могут находиться в переменной отношения. То же самое относится к многозначной и функциональной зависимостям.

Зависимость соединения $*\{A, B, \dots, Z\}$ является тривиальной тогда и только тогда, когда по крайней мере одно из подмножеств A, B, \dots, Z является множеством всех атрибутов отношения (включает все атрибуты). В противном случае зависимость соединения является нетривиальной.

Формулировка определения

Отношение находится в пятой нормальной форме (иначе — в проекционно-соединительной нормальной форме) тогда и только тогда,

когда каждая нетривиальная зависимость соединения в нём определяется потенциальным ключом (ключами) этого отношения.

Зависимость соединения $*\{A, B, \dots, Z\}$ определяется потенциальным ключом (ключами) тогда и только тогда, когда каждое из подмножеств A, B, \dots, Z множества атрибутов является суперключом отношения.

Условие «каждое из подмножеств A, B, \dots, Z множества атрибутов является суперключом отношения» можно эквивалентно сформулировать так: «каждое из подмножеств A, B, \dots, Z множества атрибутов включает некоторый потенциальный ключ отношения».

Свойства 5НФ

Любое отношение в 5НФ автоматически находится также в 4НФ и, следовательно, во всех других нормальных формах. 5НФ является окончательной нормальной формой (по крайней мере в контексте операций проекции и соединения).

Рональд Фейгин в 1979 г. показал, что любая переменная отношения может быть подвергнута декомпозиции без потерь на эквивалентный набор переменных отношения в 5НФ, то есть 5НФ всегда достижима. Однако Кристофер Дейт отмечает, что процедура определения того, что некоторая переменная отношения находится в 4НФ, а не в 5НФ, и, таким образом, существует возможность её дальнейшей выгодной декомпозиции, всё ещё остаётся не вполне ясной. Это связано с тем, что задача определения всех зависимостей соединения для отношения может оказаться очень сложной, а по поводу отношения можно утверждать, что оно находится в 5НФ, только при условии известности всех его потенциальных ключей и всех его зависимостей соединения.

Очень редко отношение, находящееся в 4НФ, не соответствует 5НФ. Это те ситуации, в которых реальные правила, ограничивающие допустимые комбинации атрибутов, никак не выражены в структуре отношения (см. пример ниже). В таком случае, если отношение не приведено к 5НФ, бремя обеспечения логической целостности данных отчасти перекладывается на приложение, отвечающее за добавление, удаление и изменения данных. В этом случае существует риск возникновения ошибок. Пятая нормальная форма исключает возникновение таких аномалий.

Пример: схема из трёх сущностей со связями многие-ко-многим, а именно

- автомобиль;
- цвет кузова;
- модель.

Ассоциативные сущности: модель – цвет, автомобиль – цвет, автомобиль – модель.

Добавляем ассоциативную сущность модель – цвет – автомобиль

5.4. Задания по нормализации для аудиторной и самостоятельной работы

Задано отношение и множество функциональных зависимостей. Определить ключ, выявить текущую нормальную форму и преобразовать это отношение к 2НФ, 3НФ, НФБК, 4НФ, подтверждая на каждом шаге правильность результата.

Задача 1.

Задано отношение Рабочий:

Рабочий

Код_рабочего	Имя_рабочего	Специальность	Код_руководителя	Почасовая_ставка	Код_здания
1235	М. Фарадей	Электрик	1311	3,50	{312,515}
1412	К Немо	Штукатур		3,00	{312,460,435, 515}
1311	Х. Колумб	Электрик		3,50	435

и множество функциональных зависимостей:

Код_рабочего → Имя_рабочего, Специальность, Код_руководителя

Код_рабочего → → Код_здания

Специальность → → Код_рабочего

Код_здания → → Код_рабочего

Специальность → Почасовая_ставка.

Задача 2.

Задано отношение Рабочий:

Рабочий

Код_рабочего	Имя_рабочего	Специальность	Почасовая_ставка	Дата_начала_работы	Код_здания
1235	М. Фарадей	Электрик	3,50	10.10.99	312
1412	К Немо	Штукатур	3,00	07.10.99	312
1235	М. Фарадей	Электрик	3,50	17.10.99	515
1412	К. Немо	Штукатур	3,00	08.12.99	460
1412	К. Немо	Штукатур	3,00	15.10.99	435

и множество функциональных зависимостей:

Код_рабочего → Имя_рабочего, Специальность

Код_рабочего → → Код_здания

Специальность → → Код_рабочего

Код_здания, Код_рабочего → Дата_начала_работы

Специальность → Почасовая_ставка.

Задача 3.

Дано отношение в 1НФ и множество функциональных зависимостей:

Студент(№_студента, предмет, повт_предмет, возраст, курс)

№_студента → возраст

№_студента → курс

№_студента → → предмет

№_студента → → повт_предмет

возраст → курс

Задача 4.

Дано отношение в 1НФ и множество функциональных зависимостей:

Поставки(№_поставщика, №_детали, №_проекта, объем_поставки, статус, город)

№_поставщика → №_проекта

№_поставщика → → №_детали

№_поставщика, №_детали, №_проекта → объем_поставки

№_поставщика → статус

№_поставщика → город

город → статус

Задача 5.

Дано отношение в 1НФ и множество функциональных зависимостей:

Преподаватель(№_преподавателя, ФИО_преподавателя, уч_степень, должность, возможный_предмет, предмет_в_нагрузке, вид_занятий, количество_часов)

№_преподавателя → ФИО, уч_степень, должность

№_преподавателя → → возможный_предмет

№_преподавателя → → предмет_в_нагрузке

предмет_в_нагрузке → → вид_занятий

№_преподавателя, предмет_в_нагрузке, вид_занятий →

количество_часов

уч_степень → должность

Задача 6.

Дано отношение в 1НФ и множество функциональных зависимостей:

Фильм (Название_фильма, Год_выхода, Длительность_фильма,

Тип_фильма, Доходность_фильма, Название_киностудии,

Адрес_киностудии, Имя_кинозвезды, Лучший_фильм_кинозвезды)

Название_фильма, Год_выпуска → Длительность_фильма, Тип_фильма,
Доходность_фильма, Название_киностудии, Адрес_киностудии
Название_киностудии → Адрес_киностудии
Название_фильма → → Имя_кинозвезды
Название_фильма → → Год_выпуска
Название_киностудии → → Название_фильма, Год_выпуска
Имя_кинозвезды → Лучший_фильм_кинозвезды