

Нормальные формы более высокого порядка

(4НФ и 5НФ)

MV-зависимости

Семантическая связь между атрибутами может носить неоднозначный характер, что порождает многозначные зависимости (MV-зависимости).

Например, один преподаватель может преподавать несколько предметов, а один предмет может преподаваться несколькими преподавателями.

Многозначная зависимость может быть следующих типов: 1:N (один ко многим), M:1 (многие к одному) и M:N (многие ко многим).

MV-зависимости

- Пусть $R(X, Y, Z)$ - схема отношения; X, Y, Z - подмножества атрибутов R .
- Если при заданных значениях атрибутов из $\{X\}$ существует некоторое множество, состоящее из нуля или более взаимосвязанных значений атрибутов из $\{Y\}$, никак не связанных со значениями других атрибутов этого отношения $Z = R - X - Y$, то говорят о существовании многозначной зависимости между атрибутами X и Y :

MV: $X \twoheadrightarrow Y$.

Виды многозначных зависимостей

- Многозначные зависимости бывают тривиальные и нетривиальные.
- Многозначная зависимость $X \twoheadrightarrow Y$ отношения R является тривиальной, если:
 - атрибут Y является подмножеством атрибута X , т.е. $Y \subseteq X$ или
 - $X \cup Y = R$.
- Если ни одно условие не выполняется, то зависимость называется нетривиальной.
- Тривиальная МЗЗ не накладывает никаких ограничений на данное отношение, а нетривиальная – накладывает.

Еще определение

Многозначная зависимость $X \twoheadrightarrow Y \mid Z$ называется нетривиальной многозначной зависимостью, если не существует функциональных зависимостей $X \rightarrow Y$ и $X \rightarrow Z$.

Класс зависимостей, опирающихся на понятие *функционала* (обобщение понятия функции), обнаружил в 1970-е гг. Рональд Фейджин.

Он назвал такие зависимости многозначными, т.к. в них одному значению детерминанта соответствует множество значений зависимого атрибута.

Наличие многозначных зависимостей, не являющихся ФЗ, приводит к аномалиям обновления отношений.

Пример

Пусть имеем отношение, в которое входит список учебных дисциплин, рекомендованная литература и имена лекторов, читающих соответствующие курсы.

Дисциплина	Книга	Лектор
МатАн	Кудрявцев	Иванов А.
МатАн	Фихтенгольц	Петров Б.
МатАн	Кудрявцев	Петров Б.
МатАн	Фихтенгольц	Иванов А.
МатАн	Кудрявцев	Сидоров В.
МатАн	Фихтенгольц	Сидоров В.
ВМ	Кудрявцев	Иванов А.
ВМ	Кудрявцев	Петров Б.

Лекторы и рекомендованные книги друг от друга не зависят \Rightarrow данное отношение содержит многозначную зависимость.

Такое отношение обладает аномалиями. Если мы хотим порекомендовать новую книгу по курсу МатАн, придется добавить столько новых записей, сколько лекторов ведут МатАн и наоборот.

Пример

Формально, здесь две МЗЗ:

$\{\text{Дисциплина}\} \rightarrow\rightarrow \{\text{Книга}\}|\{\text{Лектор}\}.$

- 1) Это избыточно.
- 2) Для такого отношения необходимо разрабатывать дополнительный механизм контроля целостности.
- 3) Оптимальным решением проблемы будет декомпозиция отношения на два с заголовками **(Дисциплина, Книга)** и **(Дисциплина, Лектор)**. Допустимость декомпозиции устанавливает **теорема Фейджина**.

Теорема Фейджина

- Пусть X, Y, Z - непересекающиеся множества атрибутов отношения $R(X, Y, Z)$.
- Декомпозиция отношения R на проекции $R_1=R[X, Y]$ и $R_2=R[X, Z]$ будет декомпозицией без потерь тогда и только тогда, когда имеется многозначная зависимость $X \twoheadrightarrow Y \mid Z$.
- Если зависимость $X \twoheadrightarrow Y \mid Z$ является тривиальной, т.е. существует одна из функциональных зависимостей $X \rightarrow Y$ или $X \rightarrow Z$, то получаем **теорему Хеза**.

Теорема Хеза (менее строгая версия теоремы Фейджина)

- Пусть X, Y, Z - множества атрибутов отношения $R(X, Y, Z)$.
- Если R удовлетворяет зависимости $X \rightarrow Y$, то R равно соединению его проекций $\{X, Y\}$ и $\{X, Z\}$.
- Обратное утверждение неверно. Почему?

4НФ

- Отношение R находится в 4НФ тогда и только тогда, когда отношение находится в НФБК и не содержит нетривиальных многозначных зависимостей.

Пример

Пусть требуется учитывать данные об абитуриентах, поступающих в ВУЗ. При анализе предметной области выделены следующие требования:

1. Каждый абитуриент имеет право подавать результаты ЕГЭ на несколько факультетов одновременно.
2. Каждый факультет имеет свой список сдаваемых предметов.
3. Один и тот же предмет может сдаваться на нескольких факультетах.
4. Абитуриент обязан сдавать все предметы, указанные для факультета, на который он поступает, хотя он уже сдавал такие же предметы на другом факультете.

Пример

Имеем отношение «Абитуриенты-Факультеты-Предметы»:

Абитуриент	Факультет	Предмет
Иванов	Математический	Математика
Иванов	Математический	Информатика
Иванов	Физический	Математика
Иванов	Физический	Физика
Петров	Математический	Математика
Петров	Математический	Информатика

Имеется нетривиальная многозначная зависимость

Факультет $\rightarrow\rightarrow$ Абитуриент | Предмет.

Из-за нетривиальности этой зависимости мы не можем

воспользоваться теоремой Хеза для декомпозиции отношения.

Факультет $\rightarrow \rightarrow$ Абитуриент | Предмет

- Для каждого факультета (для каждого значения из X) каждый поступающий на него абитуриент (значение из Y) сдает один и тот же список предметов (набор значений из Z),
- И для каждого факультета (для каждого значения из X) каждый сдаваемый на факультете предмет (значение из Z) сдается одним и тем же списком абитуриентов (набор значений из Y).
- Наличие этой зависимости не позволяет независимо вставлять и удалять кортежи.
- Кортежи обязаны вставляться и удаляться одновременно целыми наборами.

Анализ отношения

Существуют аномалии обновления из-за дублирования данных.

- Аномалия вставки. При попытке добавить в отношение новый кортеж, например (Сидоров, Математический, Математика), мы обязаны добавить также и кортеж (Сидоров, Математический, Информатика), т.к. все абитуриенты математического факультета обязаны иметь один и тот же список сдаваемых предметов.

- Аномалия удаления. Если удалим кортеж (Иванов, Физический, Математика), а вместе с ним и кортеж (Иванов, Физический, Физика), то будет потеряна информация о предметах, которые должны сдаваться на физическом факультете.

Как декомпозировать отношение?

- Декомпозиция не может быть выполнена на основе ФЗ, т.к. отношение не содержит никаких функциональных зависимостей.
- Это отношение является полностью ключевым, т.е. ключом отношения является все множество атрибутов.
- Однако, взаимосвязь между атрибутами имеется - это многозначная зависимость.

Решение

Отношение «Абитуриенты-Факультеты-Предметы» находится в НФБК, но не в 4НФ. Согласно теореме Фейджина, это отношение можно без потерь декомпозировать на два отношения:

Факультет	Абитуриент
Математический	Иванов
Физический	Иванов
Математический	Петров

Факультет	Предмет
Математический	Математика
Математический	Информатика
Физический	Математика
Физический	Физика

Полученные отношения остались полностью ключевыми, и в них по-прежнему нет ФЗ.

Причина МЗЗ

Отношения с нетривиальными МЗЗ возникают в результате естественного соединения двух отношений по общему полю, которое не является ключевым ни в одном из отношений.

Фактически это приводит к попытке хранить в одном отношении информацию о двух независимых сущностях.

Проблема разбиения отношений

Разделение установленных зависимостей по различным отношениям может привести к:

- нарушению принципа замкнутости реляционных операций,
- потере некоторых существующих кортежей,
- появлению мнимых кортежей.

Причина - еще один класс зависимостей (зависимостей проекции/соединения).

Понятие зависимости по соединению

- Функциональные и многозначные зависимости позволяют произвести декомпозицию исходного отношения без потерь на две проекции.
- Однако есть отношения, которые нельзя декомпозировать без потерь на две проекции.
- При обратном соединении этих проекций с помощью естественного соединения возможно появление ложных строк.

Пример

Рассмотрим отношение R и ее проекции, включающие по два атрибута

X	Y	Z
1	1	2
1	2	1
2	1	1
1	1	1

Проекция $R1=R[X,Y]$

X	Y
1	1
1	2
2	1

Проекция $R3=R[Y,Z]$

X	Z
1	2
1	1
2	1

Проекция $R2=R[X,Z]$

Y	Z
1	2
2	1
1	1

Пример

Соединение $R1 \text{ JOIN } R2$ имеет вид

X	Y	Z
1	1	2
1	1	1
1	2	2
1	2	1
2	1	1

Проверьте попарные соединения: $R1 \text{ JOIN } R3$ и $R2 \text{ JOIN } R3$

Отношение R не восстанавливается ни по одному из попарных соединений $R1 \text{ JOIN } R2$, $R1 \text{ JOIN } R3$ или $R2 \text{ JOIN } R3$.

Однако отношение R восстанавливается соединением всех трех проекций: $R1 \text{ JOIN } R2 \text{ JOIN } R3 = R$.

Зависимость соединения

- Пусть R - отношение; A, B, \dots, Z - произвольные (возможно пересекающиеся) подмножества атрибутов отношения R .
- Отношение R удовлетворяет зависимости соединения $*(A, B, \dots, Z)$ тогда и только тогда, когда оно равносильно соединению всех своих проекций с подмножествами атрибутов A, B, \dots, Z , т.е.

$$R = R[A] \text{ JOIN } R[B] \text{ JOIN } \dots \text{ JOIN } R[Z]$$

Отношение R в примере удовлетворяет зависимости соединения $*(XY, XZ, YZ)$.

Теорема Фейджи́на (другая формулировка)

Ранее: По теореме Фейджи́на, отношение $R(X, Y, Z)$ может быть декомпозировано без потерь на проекции $R_1 = R[X, Y]$ и $R_2 = R[X, Z]$ тогда и только тогда, когда имеется многозначная зависимость $X \twoheadrightarrow Y \mid Z$.

Согласно определению зависимости соединения, теорема Фейджи́на может быть переформулирована.

Отношение $R(X, Y, Z)$ удовлетворяет зависимости соединения $*(XY, XZ)$ тогда и только тогда, когда имеется многозначная зависимость $X \twoheadrightarrow Y \mid Z$.

Многозначная зависимость является частным случаем зависимости соединения, т.е. если в отношении имеется многозначная зависимость, то имеется и зависимость соединения. Обратное неверно.

Виды зависимостей соединения

Зависимости соединения бывают тривиальные и нетривиальные.

Зависимость соединения $*(A, B, \dots, Z)$ называется нетривиальной, если выполняются два условия:

- 1) одно из множеств атрибутов A, B, \dots, Z не содержит потенциального ключа отношения;
- 2) ни одно из множеств атрибутов не совпадает со всем множеством атрибутов отношения R .

Если одно из условий не выполняется, то зависимость называется тривиальной.

5НФ

- Отношение находится в 5НФ, если оно находится в 4НФ и любая зависимость соединения в ней является тривиальной.
- 5НФ является теоретическим исследованием и практически не применяется при реальном проектировании БД.
- Очень редко таблица, находящаяся в 4НФ, не соответствует 5НФ. Это те ситуации, в которых реальные правила, ограничивающие допустимые комбинации атрибутов, никак не выражены в структуре таблицы (*пример ниже*).

Пример ситуации

- Предположим, что продавец может торговать продукцией нескольких фирм, ассортимент у фирм различен, причем продавец может предлагать только часть товаров конкретной фирмы.
- Отношение (Продавец, Фирма, Вид товара) соответствует 4НФ, однако не отражает ограничения, связанного с ассортиментом продукции фирм. Может возникнуть кортеж, в котором фирме будет соответствовать вид товара, который она не выпускает.
- В данном случае (для приведения к 5НФ) отношение должно быть разбито на три: (Продавец, Фирма), (Фирма, Вид товара), (Продавец, Вид товара).

Спасибо за внимание!