Лекция 6

Нормализация отношений

Построение схемы БД может быть выполнено двумя путями:

* путем декомпозиции (разбиения), когда исходное множество отношений, входящих в схему

БД заменяется другим множеством отношений (их число при этом возрастает), являющихся проекциями исходных отношений

* путем синтеза, то есть путем компоновки из заданных исходных элементарных

зависимостей между объектами предметной области схемы БД

Процесс проектирования с использованием декомпозиции представляет собой процесс последовательной нормализации схем отношений, при этом каждая последующая итерация соответствует нормальной форме более высокого уровня и обладает лучшими свойствами по сравнению с предыдущей.

Каждой нормальной форме соответствует определенный набор ограничений, и отношение находится в некоторой нормальной форме, если удовлетворяет свойственному ей набору ограничений.

Функциональные зависимости

В процессе нормализации рассматриваются различные функциональные зависимости. Функциональные зависимости определяют не текущее состояние БД, а все возможные ее состояния, то есть они отражают те связи между атрибутами, которые присущи реальному объекту, моделируемые в БД.

Функциональная зависимость. Атрибут Y некоторого отношения функционально зависит от X (атрибуты могут быть составными), если в любой момент времени каждому значению X соответствует одно значение Y. Функциональная зависимость обозначается X ^ Y.

Избыточная функциональная зависимость - это зависимость, заключающая в себе такую информацию, которая может быть получена на основе других зависимостей, имеющихся в базе данных.

Полная функциональная зависимость. Неключевой атрибут функционально полно зависит от составного ключа если он функционально зависит от всего ключа в целом, но не находится в функциональной зависимости от какого-либо из входящих в него атрибутов.

Транзитивная функциональная зависимость. Пусть X, Y, Z - три атрибута некоторого отношения. При этом X ^ Y и Y ^ Z, но обратное соответствие отсутствует, т.е. Z -/-> Y и Y -/-> X. Тогда Z транзитивно зависит от X.

Многозначная зависимость. Пусть X, Y, Z - три атрибута отношения R. В отношении R существует многозначная зависимость R.X ->> R.Y только в том случае, если множество значений Y, соответствующее паре значений X и Z, зависит только от X и не зависит от Z.

В общем случае необходимо проводить нормализацию к пятой нормальной форме (5НФ). На практике зачастую оказывается достаточным приведение к третьей нормальной форме (3НФ).

Первая нормальная форма

Первая нормальная форма (1НФ): отношение находится в 1НФ, если значения всех его атрибутов атомарны.

Иначе можно сказать, что в каждой позиции пересечения столбца и строки таблицы расположено в точности одно значение, а не набор значений. Отношения в 1НФ часто называются просто нормализованными отношениями.

Под атомарностью понимается степень структурирования и детализации информации в БД. Глубина структурирования определяется практической необходимостью при манипулировании данными. Примером является глубина структурирования адреса. Можно хранить в одном поле весь адрес (город, улица, дом, квартира). Данный атрибут будет атомарным, если нет необходимости манипулировать отдельными городами или улицами, в противном случае этот атрибут не является атомарным и необходимо его дальнейшее разбиение на отдельные атрибуты (город), (улица, дом, квартира).

Пример ненормализованного и нормализованного (в 1НФ) отношений приведен на Рис. 2-16.

Вторая нормальная форма

Вторая нормальная форма (2НФ): Отношение (таблица) находится во 2НФ, если оно находится в 1НФ, и каждый неключевой атрибут функционально полно зависит от всего ключа.

Если какой-либо атрибут зависит от части составного первичного ключа, то необходимо:

* создать новое отношение, атрибутами которого будут:
* часть составного ключа (первичный ключ нового отношения)
* атрибут, зависящий от нового ключа
* из исходного отношения исключить атрибут, включенный в новое отношение

То есть, если имеется отношение R(k1, k2, a1, a2), находящееся в 1НФ,

где k1, k2 - составной первичный ключ, а а1 и а2 - неключевые атрибуты отношения R, и имеются функциональные зависимости:

k1, k2 ^ а1 (атрибут а1 функционально полно зависит от первичного ключа k1, k2), k1 ^ а2 (атрибут а2 зависит от части первичного ключа k1, т.е. имеется неполная функциональная зависимость)

Для приведения отношения R к 2НФ, это отношение декомпозируется на два отношения: R1(k1, а2) и R2(k1, k2, а1). Отношения R1 и R2 будут иметь связь один-ко-многим по атрибуту k1.

Пример: Дано отношение Поставки(КодПоставщика, КодПродукта, ЕдиницаИзмерения). Поставщик может поставлять различные продукты, один и тот же продукт может поставляться разными поставщиками. Тогда первичным ключом отношения будут атрибуты КодПоставщика и КодПродукта. Значит, существует функциональная зависимость:

КодПоставщика, КодПродукта ^ ЕдиницаИзмерения

С другой стороны, какой бы поставщик не поставит продукт, единица измерения от этого не изменится (например, цельное молоко измеряется литрами независимо от поставщика, а соль - килограммами). Т.е. существует еще одна функциональная зависимость (неключевой атрибут зависит от части первичного ключа):

КодПродукта ^ ЕдиницаИзмерения

После исключения неполной функциональной зависимости получим отношения: Поставки(КодПоставщика, КодПродукта) и Продукты(КодПродукта, ЕдиницаИзмерения)

При неполной функциональной зависимости возникают аномалии:

* включения (пока поставщиком не будет поставлен продукт, нельзя указать единицу

измерения)

* удаления (исключение поставщика может привести к потере единицы измерения продукта)
* обновления (при изменении единицы измерения продукта, приходится менять данные везде,

где встречается данный продукт)

Данные виды аномалий возникают при любой избыточной функциональной зависимости.

Третья нормальная форма

Третья нормальная форма (3НФ): Отношение находится в 3НФ, если оно находится во 2НФ и каждый неключевой атрибут нетранзитивно зависит от первичного ключа.

То есть, если имеется отношение R(k1, а1, а2), находящееся в 2НФ, где k1 - первичный ключ, а а1 и а2 - неключевые атрибуты отношения R, и имеются функциональные зависимости: k1 ^ а1 а1 ^ а2

тогда атрибут а2 транзитивно зависит от k1.

Для приведения отношения R к ЗНФ, это отношение декомпозируется на два отношения: R1(k1, а1) и И2(а1, а2). Отношения R1 и R2 будут иметь связь многие-к-одному по атрибуту а1.

Пример: Дано отношение Группы(Группа, Специальность, Факультет) с первичным ключом Группа. Группа однозначно определяет специальность, а специальность однозначно определяет факультет. Т.е. существуют следующие функциональные зависимости:

Группа ^ Специальность (и наоборот, Специальность -/-> Группа)

Специальность ^ Факультет (Факультет -/-> Специальность)

После исключения транзитивной функциональной зависимости получим отношения: Группы(Группа, Специальность) и Специальности(Специальность, Факультет)

Нормальная форма Бойса-Кодда

Ситуация, когда отношение будет находиться в 3НФ, но не в нормальной форме Бойса-Кодда (НФБК), возникает при условии, что отношение имеет два (или более) возможных ключа, которые являются составными и имеют общий атрибут. Заметим, что на практике такая ситуация встречается достаточно редко, для всех прочих отношений 3НФ и НФБК эквивалентны.

То есть, если имеется отношение R(a1, a2, a3, a4), находящееся в 3НФ, где a1, a2 - возможный ключ, a2, a3 - возможный ключ, а a4 - неключевой атрибут отношения R, и имеются функциональные зависимости: a1 ^ a3 a3 ^ a1 a1, a2 ^ a4 a2, a3 ^ a4

Для приведения отношения R к НФБК, это отношение декомпозируется на два отношения: R1(a1, a3) и R2(a1, a2, a4) или R1(a3, a1) и R2(a2, a3, a4).

*Пример*: Дано отношение Экзамен(№ зачетки, № паспорта, Дисциплина, Дата, Оценка). Возможными ключами будут атрибуты: № зачетки, Дисциплина, Дата и № паспорта, Дисциплина, Дата. Имеются следующие функциональные зависимости:

№ зачетки, Дисциплина, Дата ^ Оценка № паспорта, Дисциплина, Дата ^ Оценка № зачетки ^ № паспорта № паспорта ^ № зачетки

После приведения отношения к НФБК могут быть получены отношения:

Студент(№ зачетки, № паспорта), Экзамен(№ зачетки, Дисциплина, Дата, Оценка) или

Студент(№ паспорта, № зачетки), Экзамен(№ паспорта, Дисциплина, Дата, Оценка)

Четвертая нормальная форма

Четвертая нормальная форма (4НФ): Отношение находится в 4НФ, если оно находится в НФБК, и в нем отсутствуют многозначные зависимости, не являющиеся функциональными зависимостями. или

Отношение R находится в 4НФ в том случае, если в случае существования многозначной зависимости A ->> B все остальные атрибуты R функционально зависят от A.

То есть, если имеется отношение R(a1, a2, a3), находящееся в НФБК и имеются функциональные зависимости:

* зависимость множества значений атрибута а2 от множества значений атрибута a1 (a1 ->>

a2)

* зависимость множества значений атрибута а3 от множества значений ключевого атрибута

а1 (a1 ->> a3)

Для приведения отношения R к 4НФ, это отношение декомпозируется на два отношения: R1(a1, a2) и R2(a1, a3).

Пример: Дано отношение Книги(18ВК, Название, Автор, Область знаний). Книга имеет уникальный идентификатор ISBN, книга может быть написана коллективом авторов, книга может относиться к нескольким областям знаний (Таблица 1).

Таблица 1

|  |  |  |  |
| --- | --- | --- | --- |
| ISBN | Название | Автор | Область знаний |
| 5-123-12345-1 | Информатика для экономистов | Иванов А.В. | Информатика |
| 5-123-12345-1 | Информатика для экономистов | Иванов А.В. | Экономика |
| 5-123-12345-1 | Информатика для экономистов | Петров С.М. | Информатика |
| 5-123-12345-1 | Информатика для экономистов | Петров С.М. | Экономика |

Существуют следующие функциональные зависимости: ISBN ^ Название ISBN ->> Автор ISBN ->> Область знаний

После приведения отношения к 4НФ будут получены отношения:

Книги(18ВЫ. Название)

АвторыКниг(18ВЫ. Автор)

ОбластиЗнанийКииг(18ВЫ. Область знаний)

2.3.7. Пятая нормальная форма (нормальная форма проекции-соединения)

Зависимость соединения. Отношение R (X. Y. .... Z) удовлетворяет зависимости соединения \*(X, Y. .... Z) в том и только в том случае. когда R восстанавливается без потерь путем соединения своих проекций на X. Y. .... Z. Где X. Y. .... Z - наборы атрибутов отношения R.

Пятая нормальная форма (5НФ): Отношение R находится в 5НФ в том и только в том случае, когда любая зависимость соединения в R следует из существования некоторого возможного ключа в R.

То есть. если имеется отношение R(k1. k2. k3). находящееся в 4НФ. где k1. k2. k3 - составной первичный ключ. и имеется зависимость соединения:

\*({k1. k2}. {k1. k3}. {k2. k3})

Для приведения отношения R к 5НФ. это отношение декомпозируется на три отношения: R1(k1. k2). R2(k1. k3) и R3(k2. k3).

5НФ редко используется на практике. Очень тяжело определить само наличие зависимостей «проекции-соединения». потому что утверждение о наличии такой зависимости делается для всех возможных состояний БД. а не только для текущего экземпляра отношения R.