## Тема 5. Нормализация отношений в базе данных

*Цель: Дать понятие и назначение нормализации отношений. Выявить сложности обработки отношений, находящихся в первой и второй нормальной формах. Освоить алгоритм нормализации отношений.*

План

1. Понятие нормализации отношений
2. Первая нормальная форма
3. Вторая нормальная форма
4. Третья нормальная форма.
5. Алгоритм нормализации (приведение к 3нф)
6. Анализ критериев для нормализованных и ненормализованных моделей данных
7. **Понятие нормализации отношений**

Одни и те же данные могут группироваться в таблицы (отношения) различными способами, т.е. возможна организация различных наборов отношений взаимосвязанных информационных объектов. Группировка атрибутов в отношениях должна быть рациональной, т.е. минимизирующей дублирование данных и упрощающей процедуры их обработки и обновления.

Определенный набор отношений обладает лучшими свойствами при включении, модификации, удалении данных, чем все остальные возможные наборы отношений, если он отвечает требованиям нормализации отношений.

**Нормализация отношений** — формальный аппарат ограничений на формирование отношений (таблиц), который позволяет устранить дублирование, обеспечивает непротиворечивость хранимых в базе данных, уменьшает трудозатраты на ведение (ввод, корректировку) базы данных.

Е.Коддом выделены три *нормальные формы отношений* и предложен механизм, позволяющий любое отношение преобразовать к третьей (самой совершенной) нормальной форме.

**2. Первая нормальная форма (1нф)**

***Первая нормальная форма*** (1НФ) - это обычное отношение. Согласно нашему определению отношений, любое отношение автоматически уже находится в 1НФ. Напомним кратко свойства отношений (это и будут свойства 1НФ):

* В отношении нет одинаковых кортежей.
* Кортежи не упорядочены.
* Атрибуты не упорядочены и различаются по наименованию.
* Все значения атрибутов атомарны.

**3. Вторая нормальная форма (2 нф)**

Чтобы рассмотреть вопрос приведения отношений ко второй нормальной форме, необходимо дать пояснения к таким понятиям, как функциональная зависимость и полная функциональная зависимость.

Описательные реквизиты информационного объекта логически связаны с общим для них ключом, эта связь носит характер *функциональной зависимости* реквизитов.

***Функциональная зависимость*** реквизитов — зависимость, при которой в экземпляре информационного объекта определенному значению ключевого реквизита соответствует только одно значение описательного реквизита.

***Функционально полная зависимость*** неключевых атрибутов заключается в том, что каждый неключевой атрибут функционально зависит от ключа, но не находится в функциональной зависимости ни от какой части составного ключа.

Т.е. если часть описательных реквизитов полностью зависит от составного ключа, а другие описательные реквизиты этого же объекта, зависят только от части составного ключа, то такая зависимость не является функционально-полной.

Отношение будет находиться во ***второй нормальной форме***, тогда и только тогда, когда оно находится в первой нормальной форме, и нет неключевых атрибутов, зависящих от части составного ключа. (т.е. каждый неключевой атрибут функционально полно зависит от составного ключа.)

**4. Третья нормальная форма (3 нф)**

Понятие третьей нормальной формы основывается на понятии *нетранзитивной* зависимости.

***Транзитивная зависимость*** наблюдается в том случае, если один из двух описательных реквизитов зависит от ключа, а другой описательный реквизит зависит от первого описательного реквизита.

Описательные атрибуты называются ***взаимно независимыми***, если ни один из них не является функционально зависимым от другого.

Отношение будет находиться в третьей нормальной форме, если оно находится во второй нормальной форме, и каждый неключевой атрибут нетранзитивно зависит от первичного ключа (т.е. все описательные атрибуты взаимно независимы).

**5. Алгоритм нормализации (приведение к 3нф)**

Итак, алгоритм нормализации (т.е. алгоритм приведения отношений к 3НФ) описывается следующим образом.

Шаг 1 (Приведение к 1НФ). На первом шаге задается одно или несколько отношений, отображающих понятия предметной области. По модели предметной области (не по внешнему виду полученных отношений!) выписываются обнаруженные функциональные зависимости. Все отношения автоматически находятся в 1НФ.

Шаг 2 (Приведение к 2НФ). Если в некоторых отношениях обнаружена зависимость атрибутов от части сложного ключа, то проводим декомпозицию этих отношений на несколько отношений следующим образом: те атрибуты, которые зависят от части сложного ключа выносятся в отдельное отношение вместе с этой частью ключа. В исходном отношении остаются все ключевые атрибуты и остаток описательных.

Шаг 3 (Приведение к 3НФ). Если в некоторых отношениях обнаружена зависимость некоторых неключевых атрибутов от других неключевых атрибутов, то проводим декомпозицию этих отношений следующим образом: те неключевые атрибуты, которые зависят других неключевых атрибутов выносятся в отдельное отношение. В новом отношении ключом становится детерминант функциональной зависимости.

**6. Анализ критериев для нормализованных и ненормализованных моделей данных**

Соберем воедино результаты анализа критериев (таблица 10), по которым мы хотели оценить влияние логического моделирования данных на качество физических моделей данных и производительность базы данных:

Таблица 10 – Результаты анализа критериев для нормализованных и ненормализованных моделей данных

|  |  |  |
| --- | --- | --- |
| **Критерий** | **Отношения слабо нормализованы**  **(1НФ, 2НФ)** | **Отношения сильно нормализованы**  **(3НФ)** |
| Адекватность базы данных предметной области | ХУЖЕ (-) | ЛУЧШЕ (+) |
| Легкость разработки и сопровождения базы данных | СЛОЖНЕЕ (-) | ЛЕГЧЕ (+) |
| Скорость выполнения вставки, обновления, удаления | МЕДЛЕННЕЕ (-) | БЫСТРЕЕ (+) |
| Скорость выполнения выборки данных | БЫСТРЕЕ (+) | МЕДЛЕННЕЕ (-) |

Как видно из таблицы, более сильно нормализованные отношения оказываются лучше спроектированы (три плюса, один минус). Они больше соответствуют предметной области, легче в разработке, для них быстрее выполняются операции модификации базы данных. Правда, это достигается ценой некоторого замедления выполнения операций выборки данных.

**Контрольные вопросы**

1. Что такое нормализация отношений?
2. Какое отношение находится в первой нормальной форме?
3. Какие сложности возникают при обработке данных, если отношение находится в первой нормальной форме?
4. Дайте определение функциональной и функционально-полной зависимости атрибутов (реквизитов).
5. В каком случае можно говорить, что отношение находится во второй нормальной форме?
6. Что такое декомпозиция отношений? С какой целью она проводится?
7. Какие сложности возникают при обработке данных, если отношение находится во второй нормальной форме?
8. Что такое транзитивная зависимость атрибутов (реквизитов)?
9. В каком случае можно говорить, что отношение находится в третьей нормальной форме?

Опишите последовательность действий приведения отношения к третьей нормальной форме.