

Лекция 3

Нормальные формы в реляционных БД. 1NF, 2NF, 3NF. Целостность баз данных

Цель

Изучить теорию нормализации реляционных баз данных, освоить нормальные формы 1NF, 2NF, 3NF и понять их практическое значение для проектирования эффективных и непротиворечивых баз данных.

Основные вопросы

1. Понятие нормализации и аномалий баз данных;
2. Первая нормальная форма (1NF);
3. Вторая нормальная форма (2NF);
4. Третья нормальная форма (3NF);
5. Целостность данных и ее виды;

Лекция

Введение в нормализацию

Нормализация - процесс организации данных в базе данных для уменьшения избыточности и улучшения целостности данных. Основная цель - устранение аномалий.

Аномалии баз данных:

Аномалии вставки:

- Невозможность добавить данные без наличия связанных данных
- Необходимость дублирования информации

Аномалии обновления:

- Необходимость изменения данных в нескольких местах
- Риск несогласованности при частичном обновлении

Аномалии удаления:

- Потеря связанной информации при удалении записей
- Непреднамеренное удаление необходимых данных

Пример ненормализованной таблицы:

ЗаказID	Клиент	Товары	Сумма	Дата
1	Иванов	Книга(2), Ручка(5)	1500	2024-01-15
2	Петров	Тетрадь(3)	300	2024-01-16

Функциональные зависимости

Функциональная зависимость - ограничение между двумя множествами атрибутов в отношении.

Обозначение: $X \rightarrow Y$

- X - детерминант
- Y - зависимый атрибут

Свойства:

- Рефлексивность: если $Y \subseteq X$, то $X \rightarrow Y$
- Пополнение: если $X \rightarrow Y$, то $XZ \rightarrow YZ$
- Транзитивность: если $X \rightarrow Y$ и $Y \rightarrow Z$, то $X \rightarrow Z$

Первая нормальная форма (1NF)

Определение: Отношение находится в 1NF, если:

- Все атрибуты атомарны (не содержат составных значений)
- Все значения атрибутов принадлежат одному домену
- Каждая запись уникальна

Приведение к 1NF:

Исходная таблица (нарушает 1NF):

СтудентID	Имя	Курсы
1	Иван	Математика, Физика
2	Петр	Химия

После приведения к 1NF:

СтудентID	Имя	Курсы
1	Иван	Математика
1	Иван	Физика
2	Петр	Химия

Вторая нормальная форма (2NF)

Определение: Отношение находится в 2NF, если:

- Оно находится в 1NF;
- Все неключевые атрибуты полностью функционально зависят от первичного ключа;

Частичная зависимость - когда неключевой атрибут зависит только от части составного ключа.

Пример нарушения 2NF:

Таблица "Заказы_Товары":

ЗаказID	ТоварID	Название_товара	Количество	Цена
1	101	Книга	2	500

1	102	Ручка	5	100
2	101	Книга	1	500

Проблемы:

- Название_товара зависит только от ТоварID (части ключа);
- Избыточность данных о товарах;
- Аномалии обновления;

Приведение к 2NF:

Таблица "Заказы_Детали":

ЗаказID	ТоварID	Количество	Цена
1	101	2	500
1	102	5	100
2	101	1	500

Таблица "Товары":

ТоварID	Название_товара
101	Книга
102	Ручка

Третья нормальная форма (3NF)

Определение (Кодда): Отношение находится в 3NF, если:

- Оно находится в 2NF
- Ни один неключевой атрибут не находится в транзитивной зависимости от первичного ключа

Определение (Бойса-Кодда): Для каждой функциональной зависимости $X \rightarrow A$ выполняется одно из:

- X содержит ключ
- A является частью ключа

Транзитивная зависимость - когда неключевой атрибут зависит от другого неключевого атрибута.

Пример нарушения 3NF:

Таблица "Сотрудники":

СотрудникID	Имя	ОтделID	Название_отдела	Город
1	Иван	10	IT	Москва
2	Петр	10	IT	Москва
3	Анна	20	HR	СПб

Проблемы:

- Название_отдела и Город зависят от ОтделID (неключевой атрибут)
- Избыточность данных об отделах
- Аномалии обновления

Приведение к 3NF:

Таблица "Сотрудники":

СотрудникID	Имя	ОтделID
1	Иван	10
2	Петр	10
3	Анна	20

Таблица "Отделы":

ОтделID	Название_отдела	Город
10	IT	Москва
20	HR	СПб

Нормальные формы Бойса-Кодда (BCNF)

Определение: Отношение находится в BCNF, если для каждой функциональной зависимости $X \rightarrow A$:

- X является суперключом

Отличие от 3NF: BCNF устраняет все зависимости от неключевых атрибутов.

Целостность баз данных

Целостность данных - свойство базы данных, означающее корректность и непротиворечивость хранимых данных.

Виды целостности

Целостность сущностей (Entity Integrity)

- Первичный ключ не может содержать NULL;
- Первичный ключ должен быть уникальным;

Пример:

```
CREATE TABLE Students (
  StudentID INT PRIMARY KEY, -- не может быть NULL
  Name VARCHAR(50) NOT NULL
);
```

Целостность ссылок (Referential Integrity)

- Внешний ключ должен ссылаться на существующее значение первичного ключа;
- Или быть NULL;

Правила поддержания ссылочной целостности:

- CASCADE - каскадное удаление/обновление;
- RESTRICT - запрет операций, нарушающих целостность;
- SET NULL - установка NULL в зависимых таблицах;
- SET DEFAULT - установка значения по умолчанию;

Пример:

```
CREATE TABLE Orders (
    OrderID INT PRIMARY KEY,
    CustomerID INT,
    FOREIGN KEY (CustomerID) REFERENCES Customers(CustomerID)
        ON DELETE CASCADE
        ON UPDATE RESTRICT
);
```

Пользовательская целостность (User-defined Integrity)

- Ограничения, определяемые разработчиком
- Проверочные ограничения (CHECK)
- Уникальность (UNIQUE)
- Ограничения NOT NULL

Пример:

```
CREATE TABLE Employees (
    EmployeeID INT PRIMARY KEY,
    Salary DECIMAL(10,2) CHECK (Salary > 0),
    Email VARCHAR(100) UNIQUE,
    DepartmentID INT NOT NULL
);
```

Практические аспекты нормализации

Когда нормализовать:

- Транзакционные системы (OLTP);
- Системы с частыми обновлениями;
- Требования высокой целостности данных;

Когда денормализовать:

- Системы отчетности (OLAP);
- Системы с преимущественно чтением;
- Требования высокой производительности запросов;

Компромиссы:

- Нормализация \leftrightarrow Производительность;
- Целостность \leftrightarrow Гибкость;
- Простота разработки \leftrightarrow Оптимизация;

Контрольные вопросы

1. Что такое аномалии базы данных и какие виды аномалий существуют?
2. Сформулируйте требования для 1NF, 2NF и 3NF.
3. Что такое функциональная зависимость? Приведите примеры.
4. В чем разница между частичной и транзитивной зависимостью?
5. Какие виды целостности данных обеспечиваются в реляционных БД?
6. Когда целесообразно проводить денормализацию базы данных?

Литература

1. Дейт К. Дж. Введение в системы баз данных. - Глава 10;
2. Коннолли Т., Бегг К. Базы данных: проектирование, реализация и сопровождение. - Глава 13-14;
3. Кузнецов С.Д. Основы баз данных. - Глава 8;
4. Кренке Д. Теория и практика построения баз данных. - Глава 5.