

Нормализация БД

План:

1. Понятие нормализации
2. Первая нормальная форма
3. Вторая нормальная форма
4. Третья нормальная форма
5. Высшие нормальные формы

При проектировании БД могут появиться нежелательные свойства, такие как избыточность, аномалии обновления, аномалии включения, аномалии удаления и др. Для уменьшения нежелательных характеристик БД к схемам отношений применяют процедуры нормализации.

Нормализация - это разбиение таблицы на две или более, обладающие лучшими свойствами при включении, изменении и удалении данных.

Окончательная цель нормализации сводится к получению такого проекта базы данных, в котором каждый факт появляется лишь в одном месте, т.е. исключена избыточность информации. Это делается не столько с целью экономии памяти, сколько для исключения возможной противоречивости хранимых данных.

Нормализация – это процесс разделения информации на структурные единицы, т.е. таблицы.

Нормализация БД должна быть выполнена с учётом следующего правила: таблицы, которые содержат повторяющуюся информацию, для устранения дублирования значений должны быть разделены на отдельные таблицы, что приводит к сокращению размеров БД.

В теории реляционных баз данных вводятся понятия так называемых “нормальных форм” — требований к организации данных в таблицах.

Нормальные формы нумеруются последовательно, по мере ужесточения требований. В правильно спроектированной БД таблицы находятся как минимум в третьей нормальной форме.

Так же можно сказать, что процесс нормализации представляет собой приведение таблиц к требуемому уровню нормальности: первый, второй и третий. Каждый уровень нормальности соответствует определённой нормальной форме.

Теория нормализации основана на концепции нормальных форм. Говорят, что таблица находится в данной нормальной форме, если она удовлетворяет определённому набору требований. Теоретически существует пять нормальных форм, но на практике обычно используются только первые три. Первые две нормальные формы являются промежуточными шагами для приведения базы данных к третьей нормальной форме.

Первая нормальная форма (1НФ)

Первая нормальная форма предписывает, что все данные, содержащиеся в таблице, должны быть атомарными (неделимыми). Перечень соответствующих атомарных типов данных определяется СУБД. Требование 1НФ совершенно естественное. Оно означает, что в каждом поле каждой записи должна находиться только одна величина, но не массив и не какая-либо другая структура данных.

Вторая нормальная форма (2НФ)

Говорят, что таблица находится во второй нормальной форме, если она находится в 1НФ и каждый не ключевой столбец полностью зависит от первичного ключа. Другими словами, значение каждого поля должно полностью определяться значением первичного ключа. Важно отметить, что

зависимость от первичного ключа понимается именно как зависимость от ключа целиком, а не от отдельной его составляющей (в случае составного ключа).

Чтобы перейти от первой нормальной формы ко второй, нужно выполнить следующую последовательность действий:

- Определить, на какие части можно разбить первичный ключ, так чтобы некоторые из неключевых полей зависели от одной из этих частей (эти части не обязаны состоять из одной колонки).
- Создать новую таблицу для каждой такой части ключа и группы, зависящих от нее полей и переместить их в эту таблицу. Часть бывшего первичного ключа станет при этом первичным ключом новой таблицы.
- Удалить из исходной таблицы поля, перемещенные в другие таблицы, кроме тех, которые станут внешними ключами.

Третья нормальная форма (3НФ)

Говорят, что таблица находится в 3НФ, если она соответствует 2НФ и все не ключевые столбцы взаимно независимы.

Взаимную зависимость столбцов удобно понимать следующим образом: столбцы являются взаимно зависимыми, если нельзя изменить один из них, не изменяя другой.

Чтобы перейти от второй нормальной формы к третьей, нужно выполнить следующую последовательность действий:

- Определить все поля, от которых зависят другие поля.
- Создать новую таблицу для каждого такого поля или группы полей и группы зависящих от него полей и переместить их в эту таблицу. Поле или группа полей, от которого зависят перемещенные поля, станет при этом первичным ключом новой таблицы.
- Удалить из исходной таблицы поля, перемещенные в другие таблицы, кроме тех, которые станут внешними ключами.

Высшие нормальные формы

В теории реляционных баз данных рассматриваются и формы высших порядков — нормальная форма Бойса — Кодда, 4НФ, 5НФ и даже выше. Большого практического значения эти формы не имеют, и разработчики, как правило, всегда останавливаются на 3НФ.

Вопросы для самоконтроля:

1. Назовите цели нормализации.
2. Чем опасно избыточное дублирование информации?
3. Назовите основные свойства нормальных форм.
4. Какие ограничения таблиц относят к 1НФ, 2НФ и 3НФ?
5. Приведите примеры таблиц, соответствующих и не соответствующих требованиям нормальных форм.