Лекция 2

Табличные структуры (реляционные БД): будем использовать в качестве единицы хранения не отдельные значения, а таблицу (отношение) в целом. Данные могут группироваться в таблицы различными способами.

Простейший способ, используем одно универсальное отношение, в которое включаются все необходимые атрибуты. Проблемы при использовании универсальных отношений: избыточность данных, потенциальная противоречивость (аномалия). Существуют три вида аномалий:

1. Аномалия вставки: если ввести данные, противоречащие существующим в таблице, то это приведет к разрушению всей базы.
2. Аномалия удаления: если все сотрудники увольняются в один и тот же день, информация об отделе исчезает.
3. Аномалия модификации: если какой-либо отдел переименуется, то в универсальном отношении нужно будет вручную вводить все переименования.

Нормализация БД – решение проблем аномалий состоит в разделении данных и связей. Это разделение обеспечивается процедурой нормализации.

Цель нормализации: получение такого проекта БД, в которой каждый факт появляется только в одном месте. Процесс нормализации состоит в декомпозиции исходных отношений на более простые. Теория нормализации основана на наличии зависимостей между компонентами базы. Существуют три вида зависимостей:

1. Функциональные. Атрибут Б функционально зависит от атрибута А, если каждому значению А соответствует ровно одно значение Б.
2. Многозначные. Один атрибут отношения многозначно определяет другой атрибут отношения, если для каждого значения первого атрибута имеется множество значений второго атрибута.
3. Транзитивные. Функциональная зависимость X -> Y называется транзитивной, если существует такой атрибут Z, что имеются функциональные зависимости X -> Z, Z -> Y и отсутствует функциональная зависимость Z -> X.

|  |  |  |
| --- | --- | --- |
| A | B | Комментарий |
| 1 | 1 | Один к одному |
| 1 | M | Один ко многим |
| M | 1 | Многие к одному |
| M | M | Многие ко многим |

Каждая ступень процесса нормализации (их всего 5) приводит схему отношений в последовательные нормальные формы. Они называются 0нф (ненормализованная), 1нф, …, 5нф.

Отношение находится в 1нф, если все атрибуты являются атомарными. Чтобы привести схему к 1нф, надо

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| № сотрудника | ФИО | Должность | № отдела | Наименование | Квалификация |
| 7513 | Иванов И.И. | Программист | 128 | Отдел проектирования | С |
| 7513 | Иванов И.И. | Программист | 128 | Отдел проектирования | Java |
| 9842 | Сергеева С.С. | Администратор БД | 42 | Финансовый отдел | DBZ |
| 6513 | Петров П.П. | Программист | 128 | Отдел проектирования | VB |
| 6513 | Петров П.П. | Программист | 128 | Отдел проектирования | Java |
| 9063 | Сидоров С.С. | Системный администратор | 128 | Отдел проектирования | Linux |

Чтобы исключить дублирование данных, в отношении не должно быть повторяющихся кортежей, т.е. должна быть возможность уникальной идентификации каждого отдельного кортежа по значениям одного или нескольких атрибутов. Эти атрибуты называются реляционными ключами. Потенциальный ключ – атрибут или множество атрибутов, которые единственным образом идентифицируют кортеж. Primary key – потенциальный ключ, который выбран для уникальной идентификации кортежей внутри отношения. Первичный ключ назначается осмысленно с учетом возможных расширений БД.

Отношение находится во 2нф, если оно уже находится в 1нф и каждый неключевой атрибут полностью функционально зависит от всех составляющих первичного ключа. Переход от 1нф ко 2нф называется декомпозицией.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| ID сотрудника | ФИО | Должность | № отдела | Наименование |
| 7513 | Иванов И.И. | Программист | 128 | Отдел проектирования |
| 9842 | Сергеева С.С. | Администратор БД | 42 | Финансовый отдел |
| 6513 | Петров П.П. | Программист | 128 | Отдел проектирования |
| 9063 | Сидоров С.С. | Системный администратор | 128 | Отдел проектирования |

|  |  |
| --- | --- |
| ID сотрудника | Квалификация |
| 7513 | C |
| 7513 | Java |
| 9842 | DBZ |
| 6513 | VB |
| 6513 | Java |
| 9063 | Linux |

Отношение находится в 3нф, если уже находится во 2нф и ни один из его неключевых атрибутов не связан функциональной зависимостью с каким-либо другим неключевым атрибутом. Для приведения к 3нф выделяем отдельную таблицу.

|  |  |  |  |
| --- | --- | --- | --- |
| ID сотрудника | ФИО | Должность | № отдела |
| 7513 | Иванов И.И. | Программист | 128 |
| 9842 | Сергеева С.С. | Администратор БД | 42 |
| 6513 | Петров П.П. | Программист | 128 |
| 9063 | Сидоров С.С. | Системный администратор | 128 |

|  |  |
| --- | --- |
| ID сотрудника | Квалификация |
| 7513 | C |
| 7513 | Java |
| 9842 | DBZ |
| 6513 | VB |
| 6513 | Java |
| 9063 | Linux |

|  |  |
| --- | --- |
| № отдела | Наименование |
| 128 | Отдел проектирования |
| 42 | Финансовый отдел |

Видно, что наибольший интерес при анализе базы представляет не содержимое строк, а заголовок таблицы и первичные ключи. Именно такая нотация используется в данном анализе.

|  |  |
| --- | --- |
| Отделы | |
| PK | № Отдела |
|  | Наименование |

|  |  |
| --- | --- |
| Сотрудники | |
| PK | ID сотрудника |
|  | ФИО |
|  | Должность |
| FK | № отдела |

|  |  |
| --- | --- |
| Квалификация | |
| PK | Квалификация |
| PK, FK2 | ID сотрудника |

Полученная нами схема называется ER – диаграммой. Первая (левая) таблица – родительская. Вторая и третья таблицы – дочерние. В дочерних таблицах появились новые атрибуты (foreign key), которых не было в концептуальной модели. Это ключевые атрибуты родительской таблицы, которые мигрировали в дочерние таблицы для обеспечения связи между таблицами посредством внешних ключей (FK). Схема веер-диаграммы находится в 3нф. При анализе ER-модели нужно содержательно читать смысл полей.

Отношение находится в нормальной форме Бойса – Кодда (иначе – в усиленной третьей нормальной форме) тогда и только тогда, когда каждая его нетривиальная и неприводимая слева функциональная зависимость имеет в качестве своего детерминанта некоторый [потенциальный ключ](http://ru.wikipedia.org/wiki/%D0%9F%D0%BE%D1%82%D0%B5%D0%BD%D1%86%D0%B8%D0%B0%D0%BB%D1%8C%D0%BD%D1%8B%D0%B9_%D0%BA%D0%BB%D1%8E%D1%87).

Отношение находится в 4нф, если оно находится в нормальной форме Бойса – Кодда и не содержит нетривиальных [многозначных зависимостей](http://ru.wikipedia.org/wiki/%D0%9C%D0%BD%D0%BE%D0%B3%D0%BE%D0%B7%D0%BD%D0%B0%D1%87%D0%BD%D0%B0%D1%8F_%D0%B7%D0%B0%D0%B2%D0%B8%D1%81%D0%B8%D0%BC%D0%BE%D1%81%D1%82%D1%8C).

Отношение находится в 5нф (иначе – в проекционно-соединительной нормальной форме) тогда и только тогда, когда каждая нетривиальная зависимость соединения в нем определяется потенциальным ключом (ключами) этого отношения.