

Проектирование реляционных баз данных на основе принципов нормализации

Лекция 10

Проектирование БД

- При разработке БД обычно выделяется несколько уровней моделирования, при помощи которых происходит переход от предметной области к конкретной реализации БД средствами конкретной СУБД. Выделяют следующие уровни:
 - Сама предметная область
 - Модель предметной области
 - Логическая модель данных
 - Физическая модель данных
 - Собственно БД и приложения

Проектирование БД

- *Предметная область* - это часть реального мира, данные о которой мы хотим отразить в БД.

Проектирование БД

- ***Модель предметной области*** - это наши знания о предметной области. Знания могут быть как в виде неформальных знаний в мозгу эксперта, так и выражены формально при помощи каких-либо средств.

Проектирование БД

- *Логическая модель данных* описывает понятия предметной области, их взаимосвязь, а также ограничения на данные, налагаемые предметной областью.
- *Примеры понятий* - "сотрудник", "отдел", "проект", "зарплата".
- *Примеры взаимосвязей* между понятиями - "сотрудник числится ровно в одном отделе", "сотрудник может выполнять несколько проектов", "над одним проектом может работать несколько сотрудников".
- *Примеры ограничений* - "возраст сотрудника не менее 16 и не более 60 лет".

Проектирование БД

- ***Физическая модель данных*** описывает данные средствами конкретной СУБД.
 - Отношения, разработанные на стадии формирования логической модели данных, преобразуются в таблицы, атрибуты становятся столбцами таблиц, для ключевых атрибутов создаются уникальные индексы, домены преобразуются в типы данных, принятые в конкретной СУБД.
 - Ограничения, имеющиеся в логической модели данных, реализуются различными средствами СУБД, например, при помощи индексов, декларативных ограничений целостности, триггеров, хранимых процедур.

Критерии оценки качества логической модели данных

- Адекватность БД предметной области
- Легкость разработки и сопровождения БД
- Скорость выполнения операций обновления данных (вставка, обновление, удаление кортежей)
- Скорость выполнения операций выборки данных

предметной области.

Условия

- Состояние БД в каждый момент времени должно соответствовать состоянию предметной области.
- Изменение состояния предметной области должно приводить к соответствующему изменению состояния БД.
- Ограничения предметной области, отраженные в модели предметной области, должны некоторым образом отражаться и учитываться БД.

Проектирование БД

- Проектирование схемы БД может быть выполнено двумя путями:
 - путем декомпозиции (разбиения), когда исходное множество отношений, входящих в схему БД заменяется другим множеством отношений (число их при этом возрастает), являющихся проекциями исходных отношений;
 - путем синтеза, то есть путем компоновки из заданных исходных элементарных зависимостей между объектами предметной области схемы БД.

Проектирование БД

- Классическая технология проектирования реляционных баз данных связана с *теорией нормализации*, основанной на анализе функциональных зависимостей между атрибутами отношений.
- Понятие *функциональной зависимости* является фундаментальным в теории нормализации реляционных баз данных.

Основные понятия

- **Функциональные зависимости** определяют устойчивые отношения между объектами и их свойствами в рассматриваемой предметной области.
- **Функциональной зависимостью** набора атрибутов **B** отношения **R** от набора атрибутов **A** того же отношения, обозначаемой как **R.A**→**R.B** или **A**→**B** называется такое соотношение проекций **R[A]** и **R[B]**, при котором в каждый момент времени любому элементу проекции **R[A]** соответствует только один элемент проекции **R[B]**, входящий вместе с ним в какой-либо кортеж отношения **R**.

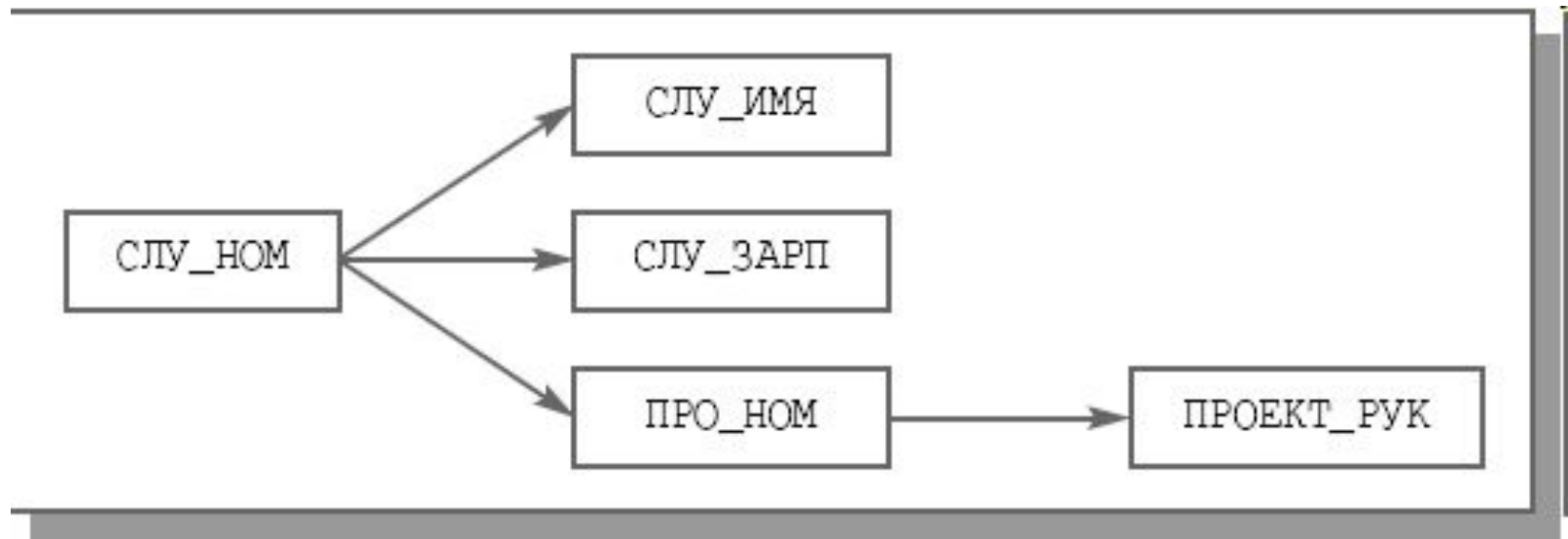
Пояснение

- Функциональные зависимости определяют не текущее состояние БД, а все возможные ее состояния, то есть они отражают те связи между атрибутами, которые присущи реальному объекту, который моделируется с помощью БД.
- Поэтому определить функциональные зависимости по текущему состоянию БД можно только в том случае, если экземпляр БД содержит абсолютно полную информацию (то есть никаких добавлений и модификации БД не предполагается).

Пример диаграммы FD

Для иллюстрации минимального множества используют диаграммы FD.

Например на рисунке приведена диаграмма FD отношения СЛУЖАЩИЕ_ПРОЕКТЫ.



ОСНОВНЫЕ ПОНЯТИЯ

- **Функциональная зависимость $R.A \rightarrow R.B$** называется *полной*, если набор атрибутов **B** функционально зависит от **A** и не зависит функционально от любого подмножества **A**,
 - то есть **$R.A \rightarrow R.B$** называется *полной*, если для любого **A1**, являющегося подмножеством **A**, **R.B** функционально не зависит от **R.A**, в противном случае зависимость **$R.A \rightarrow R.B$** называется *неполной*.

ОСНОВНЫЕ ПОНЯТИЯ

- **Функциональная зависимость $R.A \rightarrow R.B$**
называется *транзитивной*, если существует набор атрибутов C такой, что:
 - C не является подмножеством A ,
 - C не включает в себя B ,
 - Существует функциональная зависимость $R.A \rightarrow R.C$,
 - Не существует функциональной зависимости $R.C \rightarrow R.A$,
 - Существует функциональная зависимость $R.C \rightarrow R.B$.

Основные понятия

- **Ключом отношения** называется набор атрибутов отношения, который полностью и однозначно (*функционально полно*) определяет значения всех остальных атрибутов отношения, то есть **возможный ключ** — это набор атрибутов, однозначно определяющий кортеж отношения, и при этом при удалении любого атрибута из этого набора его свойство однозначной идентификации кортежа теряется.

ОСНОВНЫЕ ПОНЯТИЯ

- ***Неключевым атрибутом*** называется любой атрибут отношения, не входящий в состав ни одного возможного ключа отношения.
- ***Взаимно-независимые атрибуты*** — это такие атрибуты, которые не зависят функционально один от другого.
- Если в отношении существует несколько функциональных зависимостей, то каждый атрибут или набор атрибутов, от которого зависит другой атрибут, называется ***детерминантом отношения***.

Основные понятия

- Схемы БД называются *эквивалентными*, если содержание исходной БД может быть получено путем естественного соединения отношений, входящих в результирующую схему, и при этом не появляется новых кортежей в исходной БД.

Пояснение

- 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 некоторый определенный *набор ограничений*, и отношение находится в некоторой нормальной форме, если *удовлетворяет* свойственному ей *набору ограничений*.

Классификация НФ

- В теории реляционных БД обычно выделяется следующая последовательность нормальных форм:
 - первая нормальная форма (1NF);
 - вторая нормальная форма (2NF);
 - третья нормальная форма (3NF);
 - нормальная форма Бойса—Кодда (BCNF);
 - четвертая нормальная форма (4NF);
 - пятая нормальная форма, или форма проекции-соединения (5NF или PJNF).

Свойства НФ

- Основные свойства нормальных форм:
 - каждая следующая нормальная форма в некотором смысле улучшает свойства предыдущей;
 - при переходе к следующей нормальной форме свойства предыдущих нормальных форм сохраняются.

1НФ

- Отношение находится в первой нормальной форме тогда и только тогда, когда на пересечении каждого столбца и каждой строки находятся только элементарные значения атрибутов.

Свойства 1НФ

- Свойства 1НФ:
 - В отношении нет одинаковых кортежей.
 - Кортежи не упорядочены.
 - Атрибуты не упорядочены и различаются по наименованию.
 - Все значения атрибутов атомарны.

Пояснение 1НФ

- Отношения, находящиеся в первой нормальной форме, часто называют просто *нормализованными отношениями*.
- Соответственно, *ненормализованные отношения* могут интерпретироваться как таблицы с неравномерным заполнением.

Таблица с неравномерным заполнением

Наим.	Город	Адрес	Эл. почта	WWW	Вид	Конт. лица
Поршневой 3-д	Владимир	Ул. 2-я Кольцевая, 17	info@plunger.ru	www.plunger.ru	Поставщик	Иванов И.И., зам. дир., тел (3254)76-15-95 Петров П.П., нач. отд. сбыта, тел (3254)76-15-35
ООО Вымпел	Курск	Ул. Гоголя, 25	pennon@mail.ru		Клиент	Сидоров С.С., директор, тел. (7634)66-65-38
ИЧП Альфа	Владимир	Ул. Пушкинская, 37, оф. 565	alpha323@list.ru		Клиент	Васильев В.В., директор, тел (3254)74-57-45

Пример 1НФ

Наим.	Город	Адрес	Эл. почта	WWW	Вид	Конт. лица
Поршневой з-Владимир д		Ул. 2-я Кольцевая, 17	info@plunger.ru	www.plunger.ru u	Поставщик	Иванов И.И., зам. дир., тел (3254)76-15-95
Поршневой з-Владимир д		Ул. 2-я Кольцевая, 17	info@plunger.ru	www.plunger.ru u	Поставщик	Петров П.П., нач. отд. сбыта, тел (3254)76-15-35
ООО Вымпел Курск		Ул. Гоголя, 25	pennon@mail.ru		Клиент	Сидоров С.С., директор, тел. (7634)66-65-38
ИЧП Альфа	Владимир	Ул. Пушкинская, 37, оф. 565	alpha323@list.ru u		Клиент	Васильев В.В., директор, тел (3254)74-57-45

Пример 1НФ

Наим.	Город	Адрес	Эл. почта	WWW	Вид	Должность	Ф.И.О.	Код города	Тел.
Поршневы 3-д	Владимир	Ул. 2-я Кольцевая, 17	info@pler.ru	www.pler.ru	Поставщик	зам. дир.	Иванов И.И.	3254	76-1595
Поршневы 3-д	Владимир	Ул. 2-я Кольцевая, 17	info@pler.ru	www.pler.ru	Поставщик	нач. отд. сбыта	Петров П.П.,	3254	7615-35
ООО Вымпел	Курск	Ул. Гоголя, 25	peo@mail.ru		Клиент	директор	Сидоров С.С.	7634	6665-38
ИЧП Альфа	Владимир	Ул. Пушкинская, 37, оф. 565	al33@list.ru		Клиент	директор	Васильев В.В.	3254	745745

2 НФ

- Схема отношения **R** находится во 2НФ относительно множества функциональных зависимостей **F**, если она находится в 1НФ и каждый неключевой атрибут **полностью зависит** от каждого ключа для **R**.
- Пояснение.
 - Отношение находится во 2НФ, если оно находится в 1НФ, и при этом все неключевые атрибуты зависят только от ключа целиком, а не от какой-то его части.

Пример 2 НФ

Наим.	Город	Адрес	Эл. почта	WWW	Вид	Должност ть	Ф.И.О.	Тел.
Поршнево й 3-д	Владимир	Ул. 2-я Кольцева я, 17	info@plunger.ru	www.plunger.ru	Поставщи к	зам. дир.	Иванов И. И.	76-15-95
Поршнево й 3-д	Владимир	Ул. 2-я Кольцева я, 17	info@plunger.ru	www.plunger.ru	Поставщи к	нач. отд. сбыта	Петров П. П.,	76-15-35
ООО Вымпел	Курск	Ул. Гоголя, 25	pennon@mail.ru		Клиент	директор	Сидоров С.С.	66-65-38
ИЧП Альфа	Владимир	Ул. Пушкинск ая, 37, оф. 565	alpha323@list.ru		Клиент	директор	Васильев В.В.	74-57-45

Город Код города

Владимир

3254

Курск

7634

3 НФ

- Схема отношения **R** находится в 3НФ относительно множества функциональных зависимостей **F**, если она находится в 2НФ и ни один из непервичных атрибутов в **R** не является **транзитивно зависимым** от ключа для **R**.
- Пояснение.
 - Чтобы привести отношение к 3НФ, необходимо устранить функциональные зависимости между неключевыми атрибутами отношения. Другими словами, факты, хранимые в таблице, должны зависеть только от ключа.

Пример ЗНФ

Наим.	Город	Адрес	Эл. почта	WWW	Вид
Поршневой з-д	Владимир	Ул. 2-я Кольцевая, 17	info@plunger.ru	www.plunger.ru	Поставщик
ООО Вымпел	Курск	Ул. Гоголя, 25	pennon@mail.ru		Клиент
ИЧП Альфа	Владимир	Ул. Пушкинская, 37, оф. 565	alpha323@list.ru		Клиент

Наим.	Город	Ф.И.О.	Должность	Тел.
Поршневой з-д	Владимир	Иванов И.И.	зам. дир.	76-15-95
Поршневой з-д	Владимир	Петров П.П.,	нач. отд. сбыта	76-15-35
ООО Вымпел	Курск	Сидоров С.С.	директор	66-65-38
ИЧП Альфа	Владимир	Васильев В.В.	директор	74-57-45

НФ Бойса-Кодда

- Отношение находится в *нормальной форме Бойса—Кодда*, если оно находится в третьей нормальной форме и каждый детерминант отношения является **ВОЗМОЖНЫМ КЛЮЧОМ** отношения.
 - Примечания.
 - Если в отношении существует несколько функциональных зависимостей, то каждый атрибут или набор атрибутов, от которого зависит другой атрибут, называется *детерминантом отношения*.
 - *Возможный ключ* — это набор атрибутов, однозначно определяющий кортеж отношения, и при этом при удалении любого атрибута из этого набора его свойство однозначной идентификации кортежа теряется.

4НФ

- Отношение **R** находится в *четвертой нормальной форме* (4NF) в том и только в том случае, если в случае существования многозначной зависимости **A** \twoheadrightarrow **B** все остальные атрибуты **R** функционально зависят от **A**.

5 НФ

- Отношение \mathbf{R} находится в пятой нормальной форме (нормальной форме проекции-соединения — $\mathbf{PJ/NF}$) в том и только в том случае, когда любая зависимость соединения в \mathbf{R} следует из существования некоторого возможного ключа в \mathbf{R} .