

Лекция 2

Функциональные зависимости
и Нормальные Формы

Термины

- **Атрибут** – свойство некоторой сущности. Часто называется полем таблицы.
- **Домен атрибута** – множество допустимых значений, которые может принимать атрибут.
- **Кортеж** – конечное множество взаимосвязанных допустимых значений атрибутов, которые вместе описывают некоторую сущность (строка таблицы).

Термины

- **Отношение** — конечное множество кортежей (таблица).
- **Схема отношения** — конечное множество атрибутов, определяющих некоторую сущность. Иными словами, это структура таблицы, состоящей из конкретного набора полей.
- **Первичный ключ** — это мин. набор атрибутов, задающий уникальный идентификатор записи.

Функциональная зависимость

- **Функциональная зависимость** между атрибутами (множествами атрибутов) X и Y означает, что **для любого допустимого** набора кортежей в данном отношении: если два кортежа совпадают по значению X , то они совпадают по значению Y .

Функциональная зависимость

Пусть задан экземпляр отношения $r(R)$. Пусть $\alpha, \beta \subseteq R$ – множество атрибутов. Экземпляр удовлетворяет функциональной зависимости $\alpha \rightarrow \beta$, если для всех пар кортежей t_1 и t_2 выполнено:

$$t_1[\alpha] = t_2[\alpha] \Rightarrow t_1[\beta] = t_2[\beta]$$

Функциональная зависимость $\alpha \rightarrow \beta$ выполняется для схемы отношения $r(R)$, если она выполняется для каждой функциональной зависимости.

На примере

student_id	course_name	date	grade	name	city
123	Databases	06.01	5	Sidorov	Moscow
345	Programming	09.01	4	Petrov	Odintsovo
789	Databases	06.01	5	Ivanov	Domodedovo
234	Programming	09.01	4	Petrov	Podolsk
345	Databases	06.01	5	Petrov	Odintsovo

student_id -> name

student_id -> city

student_id -> name, city

student_id, course_name -> grade

course_name -> date ?

course_name -> grade ?

На примере

<u>student_id</u>	<u>course_name</u>	<u>date</u>	<u>grade</u>	<u>name</u>	<u>city</u>
123	Databases	06.01	5	Sidorov	Moscow
345	Programming	09.01	4	Petrov	Odintsovo
789	Databases	06.01	5	Ivanov	Domodedovo
234	Programming	09.01	4	Petrov	Podolsk
345	Databases	06.01	5	Petrov	Odintsovo
545	Databases	06.01	3	Vorontsov	Moscow

course_name -> date +

course_name -> grade -

Аномалии

Аномалией называется такая ситуация в таблице БД, которая приводит к противоречию в БД либо существенно усложняет обработку БД. Причиной является излишнее дублирование данных в таблице, которое вызывается наличием функциональных зависимостей от неключевых атрибутов.

- Аномалии-модификации (обновления)
- Аномалии-удаления
- Аномалии-добавления

Аномалии

Аномалии-модификации проявляются в том, что изменение одних данных может повлечь пересмотр всей таблицы и соответствующее изменение некоторых записей таблицы.

Аномалии-удаления — при удалении какого либо кортежа из таблицы может пропасть информация, которая не связана на прямую с удаляемой записью.

Аномалии-добавления возникают, когда информацию в таблицу нельзя поместить, пока она не полная, либо вставка записи требует дополнительного просмотра таблицы.

Аномалии

student_id	course_name	date	grade	name	city
123	Databases	06.01	5	Sidorov	Moscow
345	Programming	09.01	4	Petrov	Odintsovo
789	Databases	06.01	5	Ivanov	Domodedovo
345	Databases	06.01	3	Petrov	Odintsovo

Что произойдет, если:

- Изменилась дата экзамена Databases?
- Студент Петров отчислился?
- Нужно добавить информацию «студент Воронцов id 345 сдал БД на 3»

Нормальные формы

- **Нормальная форма** — требование, предъявляемое к структуре таблиц в теории реляционных баз данных для устранения из базы избыточных функциональных зависимостей между атрибутами (полями таблиц).

Метод нормальных форм

- **Метод нормальных форм (НФ)** состоит в сборе информации об объектах решения задачи в рамках одного отношения и последующей *декомпозиции* этого отношения на несколько взаимосвязанных отношений на основе процедур нормализации отношений.
- **Цель нормализации:** исключить избыточное дублирование данных, которое является причиной аномалий, возникших при добавлении, редактировании и удалении кортежей (строк таблицы).

Декомпозиция

A diagram illustrating the decomposition of a relation. A large red rectangle encloses the top-left portion of the slide, containing the `Students` table and its description. A red arrow points from the bottom-left corner of this red box to the `Courses` table below.

<u>student_id</u>	<u>name</u>	<u>city</u>
123	Sidorov	Moscow
345	Petrov	Odintsovo
789	Ivanov	Domodedovo

`Students(student_id, course_name, room, grade, name, city)`

A diagram illustrating the decomposition of a relation. A large red rectangle encloses the top-right portion of the slide, containing the `courses_students_link` table and its description. A red arrow points from the bottom-left corner of the red box to the `Courses` table below.

<u>student_id</u>	<u>course_name</u>	<u>grade</u>
123	Databases	5
345	Programming	4
789	Databases	5
345	Databases	5

`courses_students_link (student_id, course_name, grade)`

A diagram illustrating the decomposition of a relation. A large red rectangle encloses the bottom portion of the slide, containing the `Courses` table and its description. A red arrow points from the bottom-left corner of the red box to the `Courses` table below.

<u>course_name</u>	<u>date</u>
Databases	06.01
Programming	09.01

`Courses(course_name, room)`

Декомпозиция

Декомпозициями без потерь называются те, в результате которых имеется возможность собрать исходное отношение из декомпозированных отношений без потери информации.

СЛУЖАЩИЕ_ПРОЕКТЫ

СЛУ_НОМ	СЛУ_ИМЯ	СЛУ_ЗАРП	ПРО_НОМ	ПРО_РУК
2934	Иванов	22000	1	Иванов
2941	Иваненко	22000	2	Иваненко

Декомпозиция 1

СЛУ_НОМ	ПРО_НОМ	ПРО_РУК
2934	1	Иванов
2941	2	Иваненко

Декомпозиция 2

СЛУ_ЗАРП	ПРО_НОМ	ПРО_РУК
22000	1	Иванов
22000	2	Иваненко

СЛУ_НОМ	СЛУ_ИМЯ	СЛУ_ЗАРП
2934	Иванов	22000
2941	Иваненко	22000

СЛУ_НОМ	СЛУ_ИМЯ	СЛУ_ЗАРП
2934	Иванов	22000
2941	Иваненко	22000

Декомпозиция

Декомпозиция 1

СЛУ_НОМ	ПРО_НОМ	ПРО_РУК
2934	1	Иванов
2941	2	Иваненко

Декомпозиция 2

СЛУ_ЗАРП	ПРО_НОМ	ПРО_РУК
22000	1	Иванов
22000	2	Иваненко

СЛУ_НОМ	СЛУ_ИМЯ	СЛУ_ЗАРП
2934	Иванов	22000
2941	Иваненко	22000

СЛУ_НОМ	СЛУ_ИМЯ	СЛУ_ЗАРП
2934	Иванов	22000
2941	Иваненко	22000

Естественное соединение декомпозиции 1

СЛУ_НОМ	СЛУ_ИМЯ	СЛУ_ЗАРП	ПРО_НОМ	ПРО_РУК
2934	Иванов	22000	1	Иванов
2941	Иваненко	22000	2	Иваненко

Естественное соединение декомпозиции 2

СЛУ_НОМ	СЛУ_ИМЯ	СЛУ_ЗАРП	ПРО_НОМ	ПРО_РУК
2934	Иванов	22000	1	Иванов
2934	Иванов	22000	2	Иваненко
2941	Иваненко	22000	1	Иванов
2941	Иваненко	22000	2	Иваненко

Декомпозиция

Декомпозиция отношения $R\{A,B,C\}$ на проекции $R1 = R\{A,B\}$ и $R2 = R\{A,C\}$ называется декомпозицией без потерь, если
 $R = R1 \text{ NATURAL JOIN } R2.$

Теорема Хита:

Пусть A,B,C - непересекающиеся множества атрибутов отношения $R\{A,B,C\}$. Декомпозиция отношения R на проекции $R1 = R\{A,B\}$ и $R2 = R\{A,C\}$ будет декомпозицией без потерь, если выполняется функциональная зависимость $A \rightarrow B$.

Декомпозиция на несколько проекций

Будем называть отношение n -декомпозируемым, если оно может быть декомпозировано без потерь на n проекций.

СЛУ_ПРО_ЗАДАН

СЛУ_НОМ	ПРО_НОМ	СЛУ_ЗАДАН
2934	1	A
2934	1	B
2934	2	A
2941	1	A

СЛУ_ПРО

СЛУ_НОМ	ПРО_НОМ
2934	1
2934	2
2941	1

ПРО_ЗАДАН

ПРО_НОМ	СЛУ_ЗАДАН
1	A
1	B
2	A

СЛУ_ЗАДАН

СЛУ_НОМ	СЛУ_ЗАДАН
2934	A
2934	B
2941	A

Декомпозиция на несколько проекций

СЛУ_ПРО NATURAL JOIN СЛУ_ЗАДАН

СЛУ_НОМ	ПРО_НОМ	СЛУ_ЗАДАН
2934	1	A
2934	1	B
2934	2	A
2941	1	A
2934	2	B

СЛУ_ПРО NATURAL JOIN ПРО_ЗАДАН

СЛУ_НОМ	ПРО_НОМ	СЛУ_ЗАДАН
2934	1	A
2934	1	B
2934	2	A
2941	1	A
2941	1	B

СЛУ_ЗАДАН NATURAL JOIN ПРО_ЗАДАН

СЛУ_НОМ	ПРО_НОМ	СЛУ_ЗАДАН
2934	1	A
2934	1	B
2934	2	A
2941	1	A
2941	2	A

Лишний кортеж

Первая нормальная форма

Отношение находится в 1НФ, если все его атрибуты являются простыми, все используемые домены должны содержать только скалярные значения. Не должно быть повторений строк в таблице.

Первая нормальная форма

Фирма	Модели
BMW	M5, X5M, M1
Nissan	GT-R

1НФ:

Фирма	Модели
BMW	M5
BMW	X5M
BMW	M1
Nissan	GT-R

Вторая нормальная форма

Отношение находится во 2НФ, если оно находится в 1НФ и каждый не ключевой атрибут неприводимо зависит от Первичного Ключа (ПК).

Неприводимость означает, что в составе потенциального ключа отсутствует меньшее подмножество атрибутов, от которого можно также вывести данную функциональную зависимость.

Вторая нормальная форма

Модель	Фирма	Цена	Скидка
M5	BMW	5500000	5%
X5M	BMW	6000000	5%
M1	BMW	7500000	5%
GT-R	Nissan	5000000	10%
M1	Nissan	5000000	10%

2НФ:

Модель	Фирма	Цена
M5	BMW	5500000
X5M	BMW	6000000
M1	BMW	7500000
GT-R	Nissan	5000000

Фирма	Скидка
BMW	5%
Nissan	10%

Третья нормальная форма

Переменная отношения находится в третьей нормальной форме тогда и только тогда, когда она находится во второй нормальной форме, и отсутствуют транзитивные функциональные зависимости неключевых атрибутов от ключевых.

$A \rightarrow B$, $B \rightarrow C$, $A \rightarrow C$ - транзитивность

Третья нормальная форма

<u>Фирма</u>	<u>Магазин</u>	<u>Телефон</u>
BMW	Риал-авто	87-33-98
Audi	Риал-авто	87-33-98
Nissan	Некст-Авто	94-54-12

ЗНФ:

<u>Фирма</u>	<u>Магазин</u>
BMW	Риал-авто
Audi	Риал-авто
Nissan	Некст-Авто

<u>Магазин</u>	<u>Телефон</u>
Риал-авто	87-33-98
Некст-Авто	94-54-12

Нормальная форма Бойса-Кодда (НФБК)

Переменная отношения находится в нормальной форме Бойса — Кодда (иначе — в усиленной третьей нормальной форме) тогда и только тогда, когда каждая её нетривиальная и неприводимая слева функциональная зависимость имеет в качестве своего детерминанта некоторый потенциальный ключ.

Нормальная форма Бойса-Кодда (НФБК)

<u>Номер стоянки</u>	<u>Время начала</u>	<u>Время окончания</u>	<u>Тариф</u>
1	09:32:10	10:30:01	Эконом
1	11:02:43	12:02:12	Эконом
1	14:20:15	15:30:32	Стандарт
2	10:30:25	12:10:45	Премиум-В
2	12:34:11	14:02:11	Премиум-В
2	15:20:56	18:02:42	Премиум-А

Тарифы:

- «Эконом»: стоянка 1 для льготников
- «Стандарт»: стоянка 1 для не льготников
- «Премиум-А»: стоянка 2 для льготников
- «Премиум-В»: стоянка 2 для не льготников.

Нормальная форма Бойса-Кодда (НФБК)

Тариф	Номер стоянки	Имеет льготы
Эконом	1	Да
Стандарт	1	Нет
Премиум-А	2	Да
Премиум-В	2	Нет

Тариф	Время начала	Время окончания
Эконом	09:32:10	10:30:01
Эконом	11:02:43	12:02:12
Стандарт	14:20:15	15:30:32
Премиум-В	10:30:25	12:10:45
Премиум-В	12:34:11	14:02:11
Премиум-А	15:20:56	18:02:42

Четвертая нормальная форма

Переменная отношения находится в четвёртой нормальной форме, если она находится в нормальной форме Бойса — Кодда и не содержит нетривиальных многозначных зависимостей.

В отношении $R(A, B, C)$ существует многозначная зависимость $R.A \rightarrow R.B$ в том и только в том случае, если множество значений B , соответствующее паре значений A и C , зависит только от A и не зависит от C .

Пример многозначной зависимости

Дисциплина	Книга	Лектор
МатАн	Кудрявцев	Иванов А.
МатАн	Фихтенгольц	Петров Б.
МатАн	Кудрявцев	Петров Б.
МатАн	Фихтенгольц	Иванов А.
МатАн	Кудрявцев	Смирнов В.
МатАн	Фихтенгольц	Смирнов В.
ВМ	Кудрявцев	Иванов А.
ВМ	Кудрявцев	Петров Б.

{Дисциплина} → {Книга}

{Дисциплина} → {Лектор}

Пятая нормальная форма

Отношения находятся в 5НФ, если оно находится в 4НФ и отсутствуют сложные зависимые соединения между атрибутами.

Если «Атрибут_1» зависит от «Атрибута_2», а «Атрибут_2» в свою очередь зависит от «Атрибута_3», а «Атрибут_3» зависит от «Атрибута_1», то все три атрибута обязательно входят в один кортеж.

Доменно-ключевая нормальная форма

Переменная отношения находится в ДКНФ тогда и только тогда, когда каждое наложенное на неё ограничение является логическим следствием ограничений доменов и ограничений ключей, наложенных на данную переменную отношения.

Доменно-ключевая нормальная форма

Ограничение домена – ограничение, предписывающее использовать для определённого атрибута значения только из некоторого заданного домена. Ограничение по своей сути является заданием перечня (или логического эквивалента перечня) допустимых значений типа и объявлением о том, что указанный атрибут имеет данный тип.

Ограничение ключа – ограничение, утверждающее, что некоторый атрибут или комбинация атрибутов является потенциальным ключом.

Любая переменная отношения, находящаяся в ДКНФ, обязательно находится в 5НФ. Однако не любую переменную отношения можно привести к ДКНФ.

Шестая нормальная форма

Переменная отношения находится в шестой нормальной форме тогда и только тогда, когда она удовлетворяет всем нетривиальным зависимостям соединения.

Из определения следует, что переменная находится в 6НФ тогда и только тогда, когда она неприводима, то есть не может быть подвергнута дальнейшей декомпозиции без потерь.

Итоги

Процесс проектирования реляционной базы на основе метода нормализации преследует две основных цели:

- избежать избыточности хранения данных
- устраниить аномалии отношений