

Нормализация

Таблица, отвечающая минимальному определению отношения, может иметь неэффективную или неподходящую структуру. Для некоторых отношений изменение данных может привести к нежелательным последствиям, называемым *аномалиями модификации* (modification anomalies).

Аномалии могут быть устранены путем разбиения исходного отношения на два или более новых отношения. В большинстве случаев переопределенные, или *нормализованные*, отношения являются более предпочтительными.

Аномалии модификации

- Аномалия удаления
- Аномалия вставки

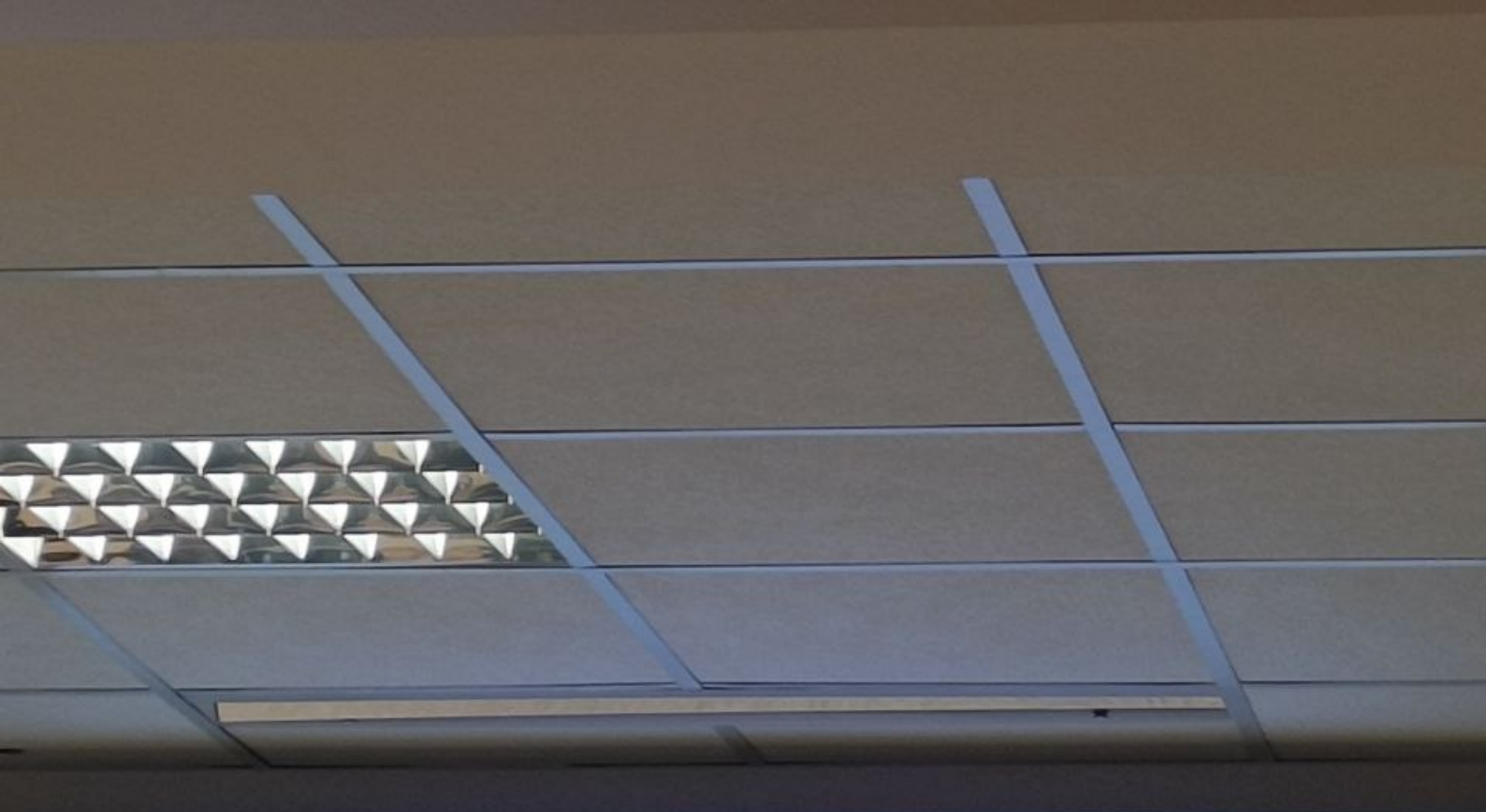
IDСтудента	Секция	Плата
100	Лыжи	2000
100	Баскетбол	3000
120	Баскетбол	3000
120	Волейбол	3000
130	Стрельба	5000

Студент-Секция

IDСтудента	Секция
100	Лыжи
110	Баскетбол
120	Баскетбол
120	Волейбол
130	Стрельба

Секция-Плата

Секция	Плата
Лыжи	2000
Баскетбол	3000
Волейбол	3000
Стрельба	5000
Плавание	2500




Ограничения, подобные этому, называются *ограничениями ссылочной целостности* (referential integrity constraints), или *ограничениями целостности по внешнему ключу* (interrelation constraints).

Суть нормализации

Проблемы возникают из-за того, что отношение СЕКЦИЯ содержит факты, относящиеся к двум различным темам.

- Кто из студентов какую секцию посещает.
- Какова плата за абонемент в каждой из секций.

При добавлении или удалении строки, информация, затрагивает две различные темы.



В своей работе, последовавшей за эпохальной статьей 1970 г., Кодд и другие определили первую, вторую и третью нормальные формы (1НФ, 2НФ и 3НФ). Позднее была введена нормальная форма Бойса-Кодда (НФБК), а затем были определены четвертая и пятая нормальные формы. Формы являются вложенными.

Первая нормальная форма (1НФ)

Значения полей должны быть атомарными.

НомЗач	Фамилия	Адрес
100753	Макаров	Снежинск, ул. Щепкина, д. 17
100730	Петрова	Екатеринбург, ул. Крауля, д. 56

Вторая нормальная форма (2НФ)

Отношение (таблица) находится во второй нормальной форме, если она находится в первой нормальной форме и функциональная зависимость полей таблицы по отношению к первичному ключу является полной функциональной зависимостью.

IDСтудента	Секция	Плата
100	Лыжи	2000
100	Баскетбол	3000
120	Баскетбол	3000
120	Волейбол	3000
130	Стрельба	5000

Плата зависит от части первичного ключа

IDСтудента	Секция
100	Лыжи
110	Баскетбол
120	Баскетбол
120	Волейбол
130	Стрельба

Секция	Плата
Лыжи	2000
Баскетбол	3000
Волейбол	3000
Стрельба	5000
Плавание	2500

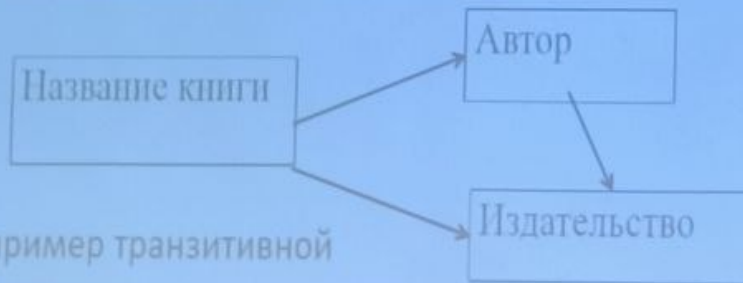
Третья нормальная форма (ЗНФ)

Б.Акунин	Азазель	ЗАХАРОВ
Б.Акунин	Квест	ЗАХАРОВ
А.Бушков	Охота на Пиранию	ОЛМА-ПРЕСС
Т.Устинова	Гений пустого места	ЭКСМО

- Автор может сотрудничать только с одним издательством
- Не может быть книг с одинаковым названием

Третья нормальная форма (ЗНФ)

первичный ключ – название книги



пример транзитивной зависимости

после декомпозиции

Название книги	PK
Автор	

Издательство	PK
Автор	

Третья нормальная форма.

Таблица находится в третьей нормальной форме, если она находится во второй нормальной форме и каждый ее атрибут, не входящий в первичный ключ, нетранзитивно зависит от первичного ключа.

Замечание. Таблица, первичный ключ которой состоит из всех полей таблицы, автоматически оказывается в третьей нормальной форме.

Нормальная форма Бойса-Кодда(НФБК) «Руководитель»

IDСтудента	Специальность	Руководитель
100	Математика	Коши
150	Философия	Спиноза
200	Математика	Риман
250	Математика	Коши
300	Экономика	Энгельс
300	Математика	Риман

Требования к этому отношению (таблице):

- Студент может иметь одну или несколько специальностей.
- Консультантами по одному и тому же предмету могут быть несколько преподавателей.

В отношении $R(A, B, C)$ существует
многозначная зависимость, если A
многозначным образом определяет B и C , а
сами B и C не зависят друг от друга.

- IDСтудента не может быть ключом.
- Руководитель определяет специальность , но не может быть ключом.
- Специальность не определяет руководителя.
- (IDСтудента, Специальность) не может быть ключом.
- (Руководитель, Специальность) не может быть ключом.
- (IDСтудента, Руководитель) может быть ключом.

Два или более атрибута или группы атрибутов, которые могут быть ключом, называются *ключами-кандидатами* (candidate keys).

Кроме ключей-кандидатов, есть еще одна функциональная зависимость, которую следует рассмотреть: атрибут **Руководитель** определяет атрибут **Специальность** (любой из преподавателей является консультантом только по одному предмету; следовательно, зная имя преподавателя, мы можем определить специальность). Таким образом, **Руководитель** является детерминантом.

Подобные ситуации, приводят к определению *нормальной формы Бойса-Кодда* (Boyse-Codd normal form, ВК/НФ): *отношение находится в НФБК, если каждый детерминант является ключом-кандидатом.*

Нормальная форма Бойса-Кодда (НФБК)

Таблица находится в НФБК, если каждый ее атрибут, от которого его атрибуты зависят функционально полно, будет являться возможным ключом данной таблицы.

ID Студента	Руководитель		
100	Коши		
150	Спиноза	Специальность	Руководитель
200	Риман	Математика	Коши
250	Коши	Философия	Спиноза
300	Энгельс	Математика	Риман
300	Риман	Экономика	Энгельс

Четвертая нормальная форма (4НФ)

Студент

IDСтудента	Специальность	Секция
100	Математика	Плавание
100	Философия	Теннис
100	Математика	Теннис
100	Философия	Плавание
300	Экономика	Стрельба

Такая зависимость атрибутов называется *многозначной зависимостью* (multi-value dependency). Многозначные зависимости приводят к аномалиям модификации.

- Избыточность данных

IDСтудента	Специальность	Секция
100	Математика	Плавание
100	Философия	Теннис
100	Математика	Теннис
100	Философия	Плавание
100	Математика	Лыжи
100	Философия	Лыжи
300	Экономика	Стрельба

Многозначная зависимость существует, когда отношение имеет минимум три атрибута, причем два из них являются многозначными, а их значения зависят только от третьего атрибута.