

Нормализация

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)

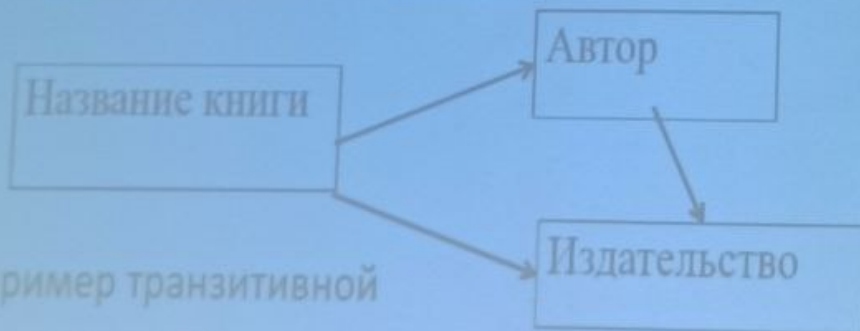
Отношение (таблица) находится во второй нормальной форме, если она находится в первой нормальной форме и функциональная зависимость полей таблицы по отношению к первичному ключу является полной функциональной зависимостью.

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

Плата зависит от части первичного ключа

Третья нормальная форма (3НФ)

первичный ключ – название книги



пример транзитивной зависимости

Третья нормальная форма (ЗНФ)

Б.Акунин	Азazelь	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

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.

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

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

Пусть студент с номером 300 отчисляется из университета. Если мы удалим строку с информацией о студенте с номером 300, мы потеряем тот факт, что Энгельс является консультантом по экономике. Это аномалия удаления.

Далее, как мы можем записать в базу тот факт, что Энгельс является консультантом по экономике? Никак, пока не появится хотя бы один студент, специализирующийся на экономике. Это аномалия вставки.

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

Далее, как мы можем записать в базу тот факт, что Энгельс является консультантом по экономике? Никак, пока не появится хотя бы один студент, специализирующийся на экономике. Это аномалия вставки.

Подобные ситуации, приводят к определению нормальной формы Бойса-Кодда (Boyce-Codd normal form, BK/NF): 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	Специальность	Руководитель
100	Коши		
150	Спиноза	Математика	Коши
200	Риман	Философия	Спиноза
250	Коши	Математика	Риман
300	Энгельс	Экономика	Энгельс
300	Риман		

Четвертая нормальная форма (4НФ)

Студент

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

- Одному и тому же значению атрибута **IDСтудента** может соответствовать много значений атрибута **Специальность**.
- Одному и тому же значению атрибута **IDСтудента** может соответствовать много значений атрибута **Секция**.

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.

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	Специальность
100	Математика
100	Философия
300	Экономика

IDСтудента	Секция
100	Плавание
100	Теннис
100	Лыжи
300	Стрельба

Четвертая нормальная форма.


Таблица (отношение) находится в четвертой нормальной форме, если она находится в Нормальная форма Бойса-Кодда (НФБК) и имеет не более одной многозначной зависимости.

Например, атрибута «В» от атрибута «А», другие атрибуты таблицы должны функционально зависеть от атрибута «А».

Пятая нормальная форма БНФ

Отдел	Сотрудник	Проект
2	Иванов	3
3	Петров	3
2	Иванов	1
1	Сидоров	2
2	Сидоров	1

- сотрудник может работать над несколькими проектами одновременно
- сотрудник может работать в нескольких отделах
- в отделе могут работать над несколькими проектами



Линейная нормальная форма (ЭНФ) связана с зависимостями, которые имеют несколько неопределенный характер. Речь здесь идет об отношениях, которые можно разделить на несколько более мелких отношений, как мы это делали выше, но затем невозможно восстановить. Условия, при которых возникает эта ситуация, не имеют ясной, интуитивной интерпретации.

Отдел	Сотрудник
2	Иванов
3	Петров
1	Сидоров
2	Сидоров

Отдел	Проект
2	3
3	3
2	1
1	2

Сотрудник	Проект
Иванов	3
Петров	3
Иванов	1
Сидоров	2
Сидоров	1

Пятая нормальная форма.

Таблица находится в пятой нормальной форме, если она находится в четвертой нормальной форме и в ней отсутствует зависимость соединения.

Доменно-ключевая нормальная форма (ДКНФ)

Все описанные нормальные формы были выделены исследователями, выявившими аномалии у некоторых отношений, которые находились в нормальной форме более низкого порядка.

Доменно-ключевая нормальная форма (ДКНФ)

Понятие ДКНФ весьма просто: отношение находится в доменно-ключевой нормальной форме, если каждое ограничение, накладываемое на это отношение, является логическим следствием определения доменов и ключей.

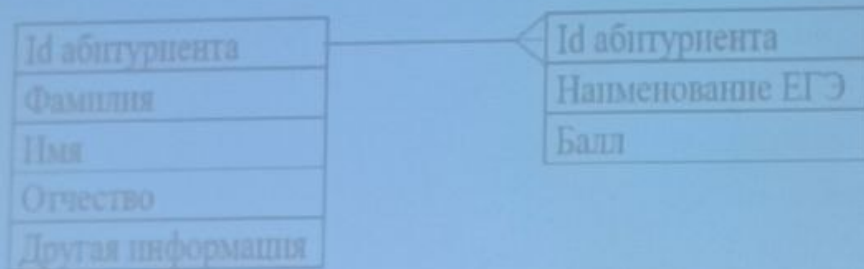
Домен — это описание допустимых значений атрибута. Он состоит из двух частей: физического описания и семантического, или логического, описания.

Физическое описание — это множество значений, которые может принимать атрибут, а логическое описание — это смысл данного атрибута. Нахождение и создание отношений в ДКНФ является более искусством, чем наукой.

Тип системы базы данных

- **OLTP** – Online Transaction Processing
(интерактивная обработка транзакций)
- **OLAP** - Online Analytical Processing
(интерактивная аналитическая обработка)

Денормализация



CONSTRAINT

CONSTRAINT

Задаёт свойства ограничений PRIMARY KEY, UNIQUE, FOREIGN KEY или CHECK, а также определений DEFAULT, которые можно добавлять в таблицу при помощи ALTER TABLE.

Синтаксис

```
[ CONSTRAINT constraint_name ]  
{  
  { PRIMARY KEY | UNIQUE }  
  [ CLUSTERED | NONCLUSTERED ]  
  (column [ ASC | DESC ] [ ,...n ] )  
  [ WITH FILLFACTOR = fillfactor  
  [ WITH ( <index_option>[ , ...n ] ) ] ]  
  [ ON { partition_scheme_name ( partition_column_name ... )  
  [ filegroup | "default" ] } ]
```


Аргументы

constraint_name

Имя ограничения. Если аргумент `constraint_name` не указан, то ограничению присваивается имя, формируемое системой.

PRIMARY KEY

Ограничение, выполняющее принудительную проверку целостности сущностей для указанного столбца или столбцов при использовании уникального индекса. Для каждой таблицы может быть создано только одно ограничение PRIMARY KEY.

Аргументы

UNIQUE

Ограничение, обеспечивающее целостность сущностей для указанного столбца или столбцов при использовании уникального индекса.

Аргументы

CLUSTERED | NONCLUSTERED

Указывает, что для ограничения PRIMARY KEY или UNIQUE создается кластеризованный или некластеризованный индекс. По умолчанию ограничения PRIMARY KEY имеют значение CLUSTERED. По умолчанию ограничения UNIQUE имеют значение NONCLUSTERED.

Аргументы

column

Столбец или список столбцов в скобках, используемые в новом ограничении.

[ASC | DESC]

Указывает порядок сортировки столбца или столбцов, участвующих в ограничениях таблицы. Значение по умолчанию — ASC.

Аргументы

WITH FILLFACTOR =*fillfactor*

Указывает, насколько полно должна заполняться каждая страница индекса, используемая для хранения индексных данных.

Аргументы

ON

{ *partition_scheme_name*(*partition_column_name*) | *filegroup* | "default" }

Указывает место хранения индекса, созданного для ограничения.

Синтаксис (продлжение)

| FOREIGN KEY

(column [,...n])

REFERENCES referenced_table_name [(ref_column [,...n])]

[ON DELETE { NO ACTION | CASCADE | SET NULL | SET DEFAULT }]

[ON UPDATE { NO ACTION | CASCADE | SET NULL | SET DEFAULT }]

]

[NOT FOR REPLICATION]

Аргументы

FOREIGN KEY REFERENCES

Ограничение, обеспечивающее ссылочную целостность данных в столбце. Ограничения FOREIGN KEY требуют, чтобы любое значение в столбце обязательно существовало в указанном столбце ссылочной таблицы. *referenced_table_name*

Таблица, на которую ссылается ограничение FOREIGN KEY.

Аргументы

FOREIGN KEY REFERENCES

Ограничение, обеспечивающее ссылочную целостность данных в столбце. Ограничения FOREIGN KEY требуют, чтобы любое значение в столбце обязательно существовало в указанном столбце ссылочной таблицы. *referenced_table_name*

Таблица, на которую ссылается ограничение FOREIGN KEY.

Аргументы

ref_column

Столбец или список столбцов в скобках, на которые ссылается новое ограничение FOREIGN KEY.

Аргументы

`ON DELETE { NO ACTION | CASCADE | SET NULL | SET DEFAULT }`

Указывает, какое действие производится над строками изменяемой таблицы, если эти строки имеют ссылочную связь, и строка, на которую имеется ссылка, удаляется из родительской таблицы. Параметр по умолчанию — `NO ACTION`.

Аргументы

NO ACTION

Компонент SQL Server Database Engine формирует ошибку, и выполняется откат операции удаления строки из родительской таблицы.

CASCADE

Если из родительской таблицы удаляется строка, соответствующие ей строки удаляются и из ссылающейся таблицы.

Аргументы

SET NULL

Все значения, составляющие внешний ключ, при удалении соответствующей строки родительской таблицы устанавливаются в значение NULL.

Аргументы

`ON UPDATE { NO ACTION | CASCADE | SET NULL | SET DEFAULT }`

Указывает, какое действие совершается над строками в изменяемой таблице, когда эти строки имеют ссылочную связь и строка родительской таблицы, на которую указывает ссылка, обновляется. Параметр по умолчанию — `NO ACTION`.

Аргументы

NOT FOR REPLICATION

Может быть определено для ограничений FOREIGN KEY и CHECK. Если ограничение сопровождается этим предложением, оно не выполняется, когда агенты репликации выполняют операции вставки, обновления или удаления.

Синтаксис (окончание)

```
| DEFAULT constant_expression FOR column [ WITH VALUES ]  
| CHECK [ NOT FOR REPLICATION ] ( logical_expression )  
|
```

Аргументы

constant_expression

Литеральное значение или лчанию для столбца. Определения DEFAULT
качестве значения столбца указания значений по умолчанию для новых
FOR column * вставляющих строках данных

Определяет столбец, соответствующий с определением
таблицы.

Аргументы

CHECK

Ограничение, обеспечивающее целостность домена путем ограничения возможных значений, которые могут быть введены в столбец или столбцы.

Аргументы

logical_expression

Логическое выражение, используемое в ограничении CHECK и возвращающее значения TRUE или FALSE. Аргумент *logical_expression*, используемый в ограничениях CHECK, не может содержать ссылок на другие таблицы, но может ссылаться на другие столбцы той же таблицы в той же строке. Выражение не может ссылаться на псевдоним типа данных.

Аргументы

logical_expression

Логическое выражение, используемое в ограничении CHECK и возвращающее значения TRUE или FALSE. Аргумент *logical_expression*, используемый в ограничениях CHECK, не может содержать ссылок на другие таблицы, но может ссылаться на другие столбцы той же таблицы в той же строке. Выражение не может ссылаться на псевдоним типа данных.

Аргументы

ON DELETE { NO ACTION | CASCADE | SET NULL | SET DEFAULT }

Указывает, какое действие производится над строками изменяемой таблицы, если эти строки имеют ссылочную связь, и строка, на которую указывает ссылка, удаляется из родительской таблицы. Параметр по умолчанию — NO ACTION.

Если из родительской таблицы удаляется строка, соответствующая ссылке, то строка удаляется и из связанной таблицы.

Индексы

Индекс в SQL Server определяется таким образом, чтобы можно было обнаружить строки, необходимые для выполнения запроса к базе данных.

Простой индекс (simple index).

Комбинированный индекс (compound index).

Типы индексов:

- кластеризованные (*только один на таблицу*)
- некластеризованные
- первичные и вторичные индексы XML

Уникальный (unique) и
неуникальный (nonunique) индексы