

Работу выполнил:
студент 3 курса
Информационных систем и технологий,
Бушуев Никита Федорович
Вопросы по теме «Нормализация»

1. Какие ограничения должны быть наложены на таблицу, чтобы она могла считаться отношением?

Во-первых, каждая строка такой таблицы содержит данные, описывающие некоторую сущность или ее часть.

Во-вторых, каждый столбец таблицы представляет один из атрибутов данной сущности.

В-третьих, каждая ее ячейка должна содержать одно и только одно значение; повторяющиеся элементы в ячейке недопустимы.

В-четвертых, необходимо, чтобы все записи в одном и том же столбце принадлежали к одному и тому же типу.

В-пятых, Каждый столбец отношения имеет уникальное имя, и порядок расположения столбцов несуществен.

В-шестых, в отношении не может быть двух идентичных строк.

2. Дайте определение термина функциональная зависимость. Приведите пример двух атрибутов, имеющих функциональную зависимость, и двух атрибутов, не имеющих функциональной зависимости.

Функциональная зависимость описывает связь между атрибутами отношения. Например, если атрибут В функционально зависит от атрибута А ($A \rightarrow B$), то каждое значение атрибута А связано только с одним значением атрибута В. Причем атрибуты А и В могут состоять из одного или нескольких атрибутов. То есть, если две строки имеют одно и то же значение атрибута А, то они обязательно имеют одно и то же значение атрибута В. При этом для одного значения атрибута В могут существовать несколько различных значений атрибута А.

Например, возьмем следующую таблицу, которая представляет университетские курсы:

Course	Teacher	Position
Математика	Смит	Профессор

Алгоритмы	Адамс	Ассистент
JavaScript	Адамс	Ассистент

Здесь атрибут Teacher функционально зависит от атрибута Course (Course \rightarrow Teacher). То есть зная название курса, мы можем определить его преподавателя. И в этом случае можно говорить, что между атрибутами Course и Teacher есть связь 1:1, а между Teacher и Course связь 1:N, так как есть несколько курсов, которые может вести один преподаватель. При этом атрибут Course функционально не зависит от атрибута Teacher.

3. Если атрибут НомерСтудента функционально определяет атрибут Секция, означает ли это, что значение атрибута НомерСтудента может быть только одно? Обоснуйте свой ответ.

Зная номер студента, мы можем сказать, в какую секцию он ходит. То есть секция функционально зависит от номера студента. Их отношение – один ко многим, так как в одну секцию может ходить множество студентов. Поэтому ответ **НЕТ**

4. Дайте определение термина детерминант.

Детерминант - любой атрибут, от которого полностью функционально зависит некоторый другой атрибут.

5. Приведите пример отношения с функциональной зависимостью, в которой детерминант состоит из двух или более атрибутов.

Student Id	Name	Course Id	Course	Date	Teacher Id	Teacher	Position
1	Том	1	Математика	11/06/2017	1	Смит	Профессор
1	Том	2	JavaScript	14/06/2017	2	Адамс	Ассистент
2	Сэм	3	Алгоритмы	12/06/2017	2	Адамс	Ассистент
3	Боб	1	Математика	13/06/2017	1	Смит	Профессор

На данный момент эта таблица имеет составной первичный ключ StudentId+CourseId. Какие функциональные зависимости от ключевых атрибутов здесь можно выделить:

StudentId, CourseId \rightarrow Date

StudentId \rightarrow Name

CourseId \rightarrow Course, TeacherId, Teacher, Position

От обеих частей составного ключа StudentId+CourseId зависит только атрибут Date - дата, в которую студент с идентификатором StudentId поступил на курс с идентификатором CourseId.

Атрибут Name зависит только от части составного ключа - от атрибута StudentId, так как зная идентификатор студента, можно сказать, какое у него имя. В данном случае имеет факт частичной зависимости.

Атрибуты Course, TeacherId, Teacher, Position зависят от другой части ключа - от атрибута CourseId. Зная значение CourseId, можно сказать, как называется курс, какой у курса преподаватель, какую должность он занимает. Опять же здесь частичная зависимость.

6. Дайте определение термина ключ.

Ключ или потенциальный ключ – это минимальный набор атрибутов, по значениям которых можно однозначно выбрать требуемый экземпляр сущности. Минимальность означает, что исключение из набора любого атрибута не позволяет идентифицировать сущность по оставшимся. Каждая сущность должна, но не обязана обладать хотя бы одним возможным ключом.

Ключи можно разделить по признаку общности как:

- Простой – сформирован из значений единственного поля, которые однозначно определяют каждую запись
- Составной – сформирован из значений нескольких полей, применяется

По ограничению целостности данных:

- Первичный ключ – служит как ограничение целостности в рамках одной таблицы для однозначной идентификации, конкретно поле первичного ключа не может повторяться или быть пустым.
- Внешний ключ(вторичный) – служит как ограничение целостности связей нескольких таблиц, конкретно подчиненная таблица не может ссылаться на несуществующие записи главной таблицы (что позволяет строить целостные модели данных).

7. Если атрибут НомерСтудента является ключом отношения, является ли он детерминантом? Может ли конкретное значение этого атрибута появиться в отношении более одного раза?

Каждый первичный Ключ – детерминант. Нет, не может.

8. Что такое аномалия удаления? Приведите пример, отличный от того, который дан в тексте.

Принцип аномалии удаления состоит в том, что при удалении некоторых данных из отношения может произойти потеря других данных. Например, у

нас есть БД о детях, где содержится также ФИО и телефон родителей, по достижении 18 лет – ребенок удаляется из базы, а вместе с ним, теряется и информация о родителей.

9. Что такое аномалия вставки? Приведите пример, отличный от того, который дан в тексте.

В отношении СОТРУДНИКИ_ОТДЕЛЫ_ПРОЕКТЫ (Н_СОТР, ФАМ, Н_ОТД, ТЕЛ, Н_ПРО, ПРОЕКТ, Н_ЗАДАН)

где

Н_СОТР - табельный номер сотрудника

ФАМ - фамилия сотрудника

Н_ОТД - номер отдела, в котором числится сотрудник

ТЕЛ - телефон сотрудника

Н_ПРО - номер проекта, над которым работает сотрудник

ПРОЕКТ - наименование проекта, над которым работает сотрудник

Н_ЗАДАН - номер задания, над которым работает сотрудник

нельзя вставить данные о сотруднике, который пока не участвует ни в одном проекте. Действительно, если, например, во втором отделе появляется новый сотрудник, скажем, Пушкинов, и он пока не участвует ни в одном проекте, то мы должны вставить в отношение кортеж (4, Пушкинов, 2, 33-22-11, null, null, null). Это сделать невозможно, т.к. атрибут Н_ПРО (номер проекта) входит в состав потенциального ключа, и, следовательно, не может содержать null-значений.

Точно также нельзя вставить данные о проекте, над которым пока не работает ни один сотрудник.

Причина аномалии - хранение в одном отношении разнородной информации (и о сотрудниках, и о проектах, и о работах по проекту).

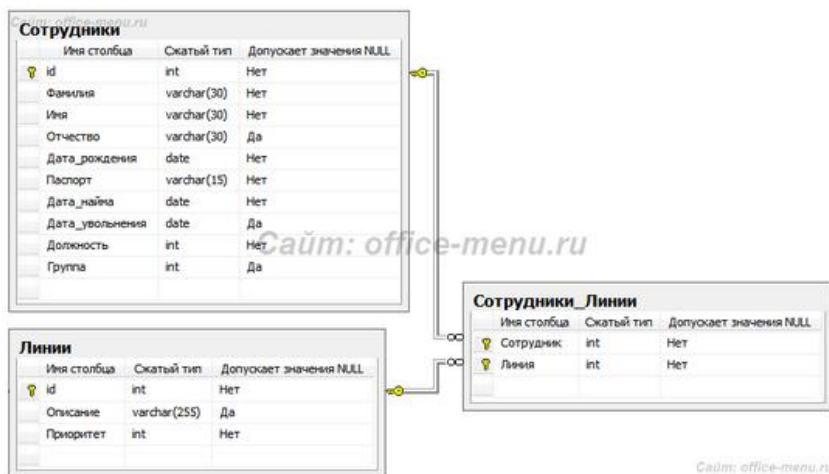
Вывод - логическая модель данных неадекватна модели предметной области. База данных, основанная на такой модели, будет работать неправильно.

10. Объясните, как соотносятся между собой первая, вторая, третья нормальная формы, нормальная форма Бойса-Кодда, четвертая, пятая и доменно-ключевая нормальные формы.

Каждая следующая нормальная форма «лучше» предыдущей нормальной формы. При переходе к следующей нормальной форме ограничения предыдущих нормальных форм сохраняются. То есть нельзя перейти к следующей нормальной форме, не достигнув предыдущей.

11. Дайте определение термина вторая нормальная форма. Приведите пример отношения, которое находится в первой нормальной форме, но не находится во второй нормальной форме. Преобразуйте это отношение в отношения, находящиеся во второй нормальной форме.

Вторая нормальная форма (в этом определении предполагается, что единственным ключом отношения является первичный ключ). Отношение R находится во второй нормальной форме (2NF) в том и только в том случае, когда находится в 1NF, и каждый неключевой атрибут полностью зависит от первичного ключа.



Условием второй нормальной формы является отсутствие зависимости неключевых полей от части составного ключа. Так как составной ключ в учебной базе наблюдается только в таблице «Сотрудники_Линии», то рассмотрим пример на ней.

Имя столбца	Сжатый тип	Допускает значения NULL
Сотрудник	int	Нет
Линия	int	Нет
Описание	varchar(255)	Да
Приоритет	int	Нет

На представленной диаграмме столбцы описания и приоритета зависят от столбца «Линия», входящего в составной ключ. Это значит, что для каждой линии, подключенной разным сотрудникам, потребуется повторно указывать описание и приоритетность. Подобная структура приводит к **избыточности данных**.

Если соблюдены правила первой нормальной формы, то создание таблицы «Линии» и перенос в нее зависимых столбцов удовлетворяет второй нормальной форме.

12. Дайте определение термина третья нормальная форма. Приведите пример отношения, которое находится во второй нормальной форме, но

не находится в третьей нормальной форме. Преобразуйте это отношение в отношения, находящиеся в третьей нормальной форме.

Отношение находится в 1НФ и каждый из неключевых атрибутов зависит от всего ключа.

Student Id	Name	Course Id	Course	Date	Teacher Id	Teacher	Position
1	Том	1	Математика	11/06/2017	1	Смит	Профессор
1	Том	2	JavaScript	14/06/2017	2	Адамс	Ассистент
2	Сэм	3	Алгоритмы	12/06/2017	2	Адамс	Ассистент
3	Боб	1	Математика	13/06/2017	1	Смит	Профессор

На данный момент эта таблица имеет составной первичный ключ StudentId+CourseId. Какие функциональные зависимости от ключевых атрибутов здесь можно выделить:

StudentId, CourseId → Date

StudentId → Name

CourseId → Course, TeacherId, Teacher, Position

От обеих частей составного ключа StudentId+CourseId зависит только атрибут Date - дата, в которую студент с идентификатором StudentId поступил на курс с идентификатором CourseId.

Атрибут Name зависит только от части составного ключа - от атрибута StudentId, так как зная идентификатор студента, можно сказать, какое у него имя. В данном случае имеет факт частичной зависимости.

Атрибуты Course, TeacherId, Teacher, Position зависят от другой части ключа - от атрибута CourseId. Зная значение CourseId, можно сказать, как называется курс, какой у курса преподаватель, какую должность он занимает. Опять же здесь частичная зависимость.

Наличие частичных зависимостей говорит о том, что таблица не находится во второй нормальной форме. И для перехода к этой форме необходимо переместить атрибуты, которые не входят в первичный ключ, в новую таблицу вместе с копией части первичного ключа, от которой они функционально зависят.

В нашем случае из одной таблицы получатся три. Таблица Students:

StudentId	Name
1	Том
2	Сэм
3	Боб

Таблица Courses:

CourseId	Course	TeacherId	Teacher	Position
1	Математика	1	Смит	Профессор
2	JavaScript	2	Адамс	Ассистент
3	Алгоритмы	2	Адамс	Ассистент

И таблица StudentCourses:

StudentId	CourseId	Date
1	1	11/06/2017
1	2	14/06/2017
2	3	12/06/2017
3	1	13/06/2017

Итогом стало образование связи многие ко многим (много студентов - много курсов) между таблицами Students и Courses через таблицу StudentCourses.

Таким образом, база данных перешла во вторую нормальную форму.

13. Дайте определение термина нормальная форма Бойса—Кодда. Приведите пример отношения, которое находится в ЗНФ, но не находится в НФБК. Преобразуйте это отношение в отношения, находящиеся в НФБК.

Отношение R находится в нормальной форме Бойса-Кодда (BCNF) в том и только в том случае, если каждый детерминант является возможным ключом.

Предположим, рассматривается отношение, представляющее данные о бронировании теннисных кортов на день:

Бронирования кортов

Номер корта	Время начала	Время окончания	Тариф
1	09:30	10:30	«Корт 1 для членов клуба»
1	11:00	12:00	«Корт 1 для членов клуба»
1	14:00	15:30	«Корт 1 для не членов клуба»
2	10:00	11:30	«Корт 2 для не членов клуба»
2	11:30	13:30	«Корт 2 для не членов клуба»
2	15:00	16:30	«Корт 2 для членов клуба»

Отношение соответствует второй (2NF) и третьей (3NF) нормальной форме. Требования второй нормальной формы выполняются, так как все атрибуты входят в какой-то из потенциальных ключей, а неключевых атрибутов в отношении нет. Также нет и транзитивных зависимостей, что соответствует требованиям третьей нормальной формы. Тем не менее, существует функциональная зависимость **Тариф** → **Номер корта**, в которой левая часть (детерминант) не является потенциальным ключом отношения, то есть отношение *не находится в нормальной форме Бойса — Кодда*.

Можно улучшить структуру с помощью декомпозиции отношения на два, получив отношения, удовлетворяющие BCNF (подчёркнуты атрибуты, входящие в первичный ключ). Для большей наглядности к информации о тарифах добавлен атрибут **Для членов клуба**:

Тарифы			Бронирования		
Тариф	Номер корта	Для членов клуба	Тариф	Время начала	Время окончания
«Корт 1 для членов клуба»	1	Да	«Корт 1 для членов клуба»	09:30	10:30
«Корт 1 для не членов клуба»	1	Нет	«Корт 1 для членов клуба»	11:00	12:00
«Корт 2 для членов клуба»	2	Да	«Корт 1 для не членов клуба»	14:00	15:30
«Корт 2 для не членов клуба»	2	Нет	«Корт 2 для не членов клуба»	10:00	11:30
			«Корт 2 для не членов клуба»	11:30	13:00
			«Корт 2 для членов клуба»	15:00	16:30

14. Дайте определение термина многозначная зависимость. Приведите пример.

Многозначная зависимость. Это случай, когда одному значению одного атрибута соответствует несколько значений другого атрибута.

15. Почему многозначные зависимости не являются проблемой в отношениях, имеющих только два атрибут.

-

16. Дайте определение термина четвертая нормальная форма. Приведите пример отношения, которое находится в НФБК, но не находится в 4НФ. Преобразуйте это отношение в отношения, находящиеся в 4НФ.

Переменная отношения R находится в четвёртой нормальной форме, если она находится в НФБК и все нетривиальные многозначные зависимости фактически являются функциональными зависимостями от её потенциальных ключей. Предположим, что рестораны производят разные виды пиццы, а службы доставки ресторанов работают только в определенных районах города. Составной первичный ключ соответствующей переменной отношения включает три атрибута: **{Ресторан, Вид пиццы, Район доставки}**.

Такая переменная отношения не соответствует 4НФ, так как существует следующая многозначная зависимость:

- **{Ресторан} \twoheadrightarrow {Вид пиццы}**
- **{Ресторан} \twoheadrightarrow {Район доставки}**

То есть, например, при добавлении нового вида пиццы придется внести по одному новому кортежу для каждого района доставки. Возможна логическая аномалия, при которой определенному виду пиццы будут соответствовать лишь некоторые районы доставки из обслуживаемых рестораном районов.

Для предотвращения аномалии нужно декомпонировать отношение, разместив независимые факты в разных отношениях. В данном примере следует выполнить декомпозицию на **{Ресторан, Вид пиццы}** и **{Ресторан, Район доставки}**.

Однако, если к исходной переменной отношения добавить атрибут, функционально зависящий от потенциального ключа, например цену с учётом стоимости доставки (**{Ресторан, Вид пиццы, Район доставки} \rightarrow Цена**), то полученное отношение будет находиться в 4НФ и его уже нельзя подвергнуть декомпозиции без потерь.

17. Дайте определение термина доменно-ключевая нормальная форма. В чем состоит важность этой формы?

Отношение находится в доменно-ключевой нормальной форме, если каждое ограничение, накладываемое на это отношение, является логическим следствием определения доменов и ключей.

18. Преобразуйте следующее отношение к ДКНФ. Сделайте и сформулируйте соответствующие предположения о функциональных зависимостях и доменах.

ОБОРУДОВАНИЕ (Производитель, Модель, ДатаПриобретения, ИмяПокупателя, ТелефонПокупателя, АдресЗавода, Город, Штат, Индекс)

19. Преобразуйте следующее отношение в ДКНФ. Сделайте и сформулируйте соответствующие предположения о функциональных зависимостях и доменах.

СЧЕТ (Номер, ИмяПокупателя, НомерПокупателя, АдресПокупателя, КодТовара, ЦенаТовара, КоличествоТовара, НомерПродавца, ИмяПродавца, ПромежуточныйИтог, Налог, ВсегоКОплате)

20. Снова ответьте на вопрос 20, но теперь добавьте атрибут НалоговыйСтатусПокупателя (0, если покупатель не освобожден от налога, и 1, если освобожден). Добавьте также ограничение, что налог не должен включаться в счет, если НалоговыйСтатусПокупателя равен 1.

21. Приведите пример (отличный от того, который дан в тексте) ситуации, в которой, как вы считаете, нормализацию производить не стоило бы. Изобразите отношения и обоснуйте свое решение.

22. Укажите две ситуации, в которых проектировщики базы данных могут преднамеренно прибегнуть к дублированию данных. В чем заключается риск, связанный с подобными решениями?