



Нормализация отношений

*"Сложная система, спроектированная наспех,
никогда не работает, и исправить её,
чтобы заставить работать,
невозможно".*

Законы Мерфи. 16-й закон системантики

Различают следующие проблемы в БД:

- избыточность данных;
- аномалии обновления;
- аномалии удаления;
- аномалии ввода.

Избыточность данных характеризуется наличием в кортежах отношений повторяющейся информации. Многократное дублирование данных приводит к неоправданному увеличению занимаемого объема внешней памяти

Аномалии удаления могут возникать при удалении записей из ненормализованных таблиц и характеризуются вероятностью удаления не всех дублированных кортежей.

Аномалии ввода возникают при добавлении в таблицу новых записей, обычно в поля с ограничениями NOT NULL (не пустые). Когда в отношение на данный момент времени невозможно ввести однозначную информацию.

Аномалии обновления, прежде всего, связаны с избыточностью данных, что приводит к проблемам при их изменении. При изменении повторяющихся данных придется многократно изменять их значения, однако, если изменения будут внесены не во все кортежи, возникнет несоответствие информации, которое называется аномалией обновления.

Кафедра	Факультет/Институт
Информационных технологий	ПТиИП
Информационной безопасности	ПТиИП

Нормализация схемы отношения

Нормализация схемы отношения выполняется путём декомпозиции схемы.

Декомпозиция отношения не должна приводить к потере зависимостей между атрибутами сущностей.

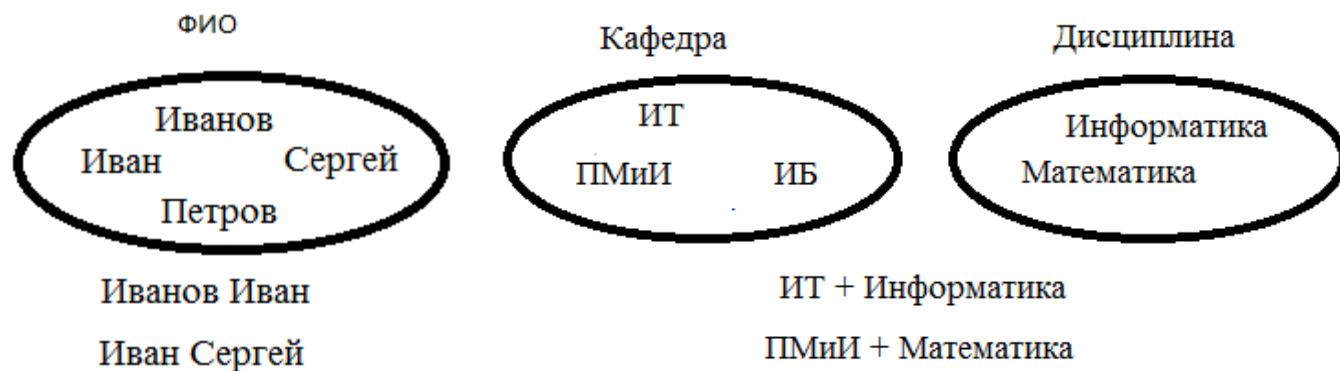
Для декомпозиции должна существовать операция реляционной алгебры, применение которой позволит восстановить исходное отношение.

Первая нормальная форма

Введём понятие простого и сложного атрибута:

Простой атрибут – это атрибут, значения которого атомарны (т.е. неделимы).

Сложный атрибут может иметь значение, представляющее собой конкатенацию нескольких значений одного или разных доменов. Аналогом сложного атрибута может быть агрегат или повторяющийся агрегат данных.



Первая нормальная форма (1НФ).

Отношение приведено к 1НФ, если все его атрибуты простые и каждая ее строка содержит только одно значение для каждого атрибута.

Первая нормальная форма



Шифр	ФИО
1222	Иванов Сергей Николаевич
2326	Достоевский Федр Михайлович

Шифр	Фамилия	Имя	Отчество
1222	Иванов	Сергей	Николаевич
2326	Достоевский	Федр	Михайлович

СТУДЕНТ (№_зачетки, Фамилия, Группа, Факультет, Семестр, Предмет, Преподаватель, Вид_Работы, Оценка)

№ зачетки	Семестр	Дисциплина	Фамилия	Группа	Факультет	Преподаватель	Вид_работы	Оценка
01	1	Химия	Панов	Г1	Ф1	Сомов	Экз	Отл
01	1	Физика	Панов	Г1	Ф1	Петров	Экз	Отл
01	1	История	Панов	Г1	Ф1	Львов	Экз	Отл
02	1	Химия	Туров	Г2	Ф1	Сомов	Экз	Хор
02	1	Физика	Туров	Г2	Ф1	Петров	Экз	Отл
02	1	История	Туров	Г2	Ф1	Львов	Экз	Хор

Первичным ключом в отношении СТУДЕНТ является группа атрибутов:

{№_зачетки, Семестр, Дисциплина}.

Вторая нормальная форма

Вторая нормальная форма (2НФ).

Отношение находится во 2НФ, если оно приведено к 1НФ и каждый неключевой атрибут функционально полно зависит от составного первичного ключа.

Для того чтобы привести отношение ко 2НФ, нужно:

1. построить его проекцию, исключив атрибуты, которые не находятся в функционально полной зависимости от составного первичного ключа;
2. построить дополнительные проекции на часть составного ключа и атрибуты, функционально зависящие от этой части ключа.

Вторая нормальная форма

Функциональная зависимость. Поле В таблицы функционально зависит от поля А той же таблицы в том и только в том случае, когда в любой заданный момент времени для каждого из различных значений поля А обязательно существует только одно из различных значений поля В. Отметим, что здесь допускается, что поля А и В могут быть составными.

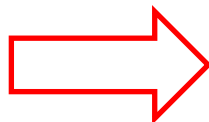
Блюда		
БЛ	Блюдо	Вид
1	Лобио	Закуска
2	Харчо	Суп
3	Шашлык	Горячее
4	Кофе	Десерт
...

Полная функциональная зависимость. Поле В находится в полной функциональной зависимости от составного поля А, если оно функционально зависит от А и не зависит функционально от любого подмножества поля А.

Приведение отношения к 2НФ

Серия паспорта	Номер паспорта	Фамилия
1200	5464	Иванов
1200	7777	Петров

1200 + 5464 = Иванов
1200 = ???



Полная функциональная зависимость

Для предметной области справедливы следующие функциональные зависимости:

$F_1 = \text{№_зачетки} \rightarrow \text{Фамилия, Группа, Факультет}$

$F_2 = \text{№_зачетки, Семестр, Дисциплина} \rightarrow \text{Преподаватель, Вид_Работы, Оценка}$

$F_3 = \text{№_зачетки, Семестр, Дисциплина} \rightarrow \text{Фамилия, Группа, Факультет}$

$F_4 = \text{№_зачетки, Семестр, Дисциплина} \rightarrow \text{Оценка}$

$F_5 = \text{Дисциплина} \rightarrow \text{Преподаватель}$

$F_6 = \text{Семестр, Дисциплина} \rightarrow \text{Вид_Работы}$

$F_7 = \text{Группа} \rightarrow \text{Факультет.}$

Приведение отношения к 2НФ

Функциональная зависимость F_3 является неполной, т.к. набор атрибутов {Фамилия, Группа, Факультет}, в соответствии с F_1 , функционально зависит от атрибута №_зачетки, входящего в состав атрибутов левой части функциональной зависимости F_3 .

$F_1 = \text{№_зачетки} \rightarrow \text{Фамилия, Группа, Факультет}$

$F_3 =$

$\text{№_зачетки, Семестр, Дисциплина} \rightarrow \text{Фамилия, Группа, Факультет}$

Функциональные зависимости F_1, F_2, F_4, F_5, F_6 являются полными функциональными зависимостями.

Отметим, что зависимости F_2, F_3, F_4 являются зависимостями набора атрибутов схемы отношения от ключа отношения.

Следовательно, отношение СТУДЕНТ не находится во второй нормальной форме. Используя правило декомпозиции, приведем отношение СТУДЕНТ ко второй нормальной форме.

Приведение отношения к 2НФ

F_1 = №_зачетки -> Фамилия, Группа, Факультет

F_3 = №_зачетки, Семестр, Дисциплина -> Фамилия, Группа, Факультет

Проекции в результате декомпозиции:

R1 (№_зачетки, Фамилия, Группа, Факультет)

R2 (№_зачетки, Семестр, Дисциплина, Преподаватель, Вид_Работы, Оценка).

№ зачетки	Фамилия	Группа	Факультет
01	Панов	Г1	Ф1
02	Туров	Г2	Ф1

№ зачетки	Семестр	Дисциплина	Преподаватель	Вид работы	Оценка
01	1	Химия	Сомов	Экз	Отл
01	1	Физика	Петров	Экз	Отл
01	1	История	Львов	Экз	Отл
02	1	Химия	Сомов	Экз	Хор
02	1	Физика	Петров	Экз	Отл
02	1	История	Львов	Экз	Хор

Приведение отношения к 2НФ

Отношение R2 не находится во 2НФ:

1. неполная функциональной зависимости
непервичного атрибута *Преподаватель* от
ключа отношения
2. неполная функциональной зависимости
атрибута *Вид_Работы* от ключа отношения

Проекции в результате декомпозиции отношения R2:

R3 (№ зачетки, Семестр, Дисциплина, Оценка),

R4 (Дисциплина, Преподаватель),

R5 (Семестр, Дисциплина, Вид_Работы).

Приведение отношения к 2НФ

№ зачетки	Фамилия	Группа	Факультет
01	Панов	Г1	Ф1
02	Туров	Г2	Ф1

№_зачетки	Семестр	Дисциплина	Оценка
01	1	Химия	Отл
01	1	Физика	Отл
01	1	История	Отл
02	1	Химия	Хор
02	1	Физика	Отл
02	1	История	Хор

Дисциплина	Преподаватель
Физика	Петров
История	Львов
Химия	Сомов

Семестр	Дисциплина	Вид_работы
1	Физика	Экз
1	История	Экз
1	Химия	Экз

1. R1 (№ зачетки, Фамилия, Группа, Факультет),
2. R3 (№ зачетки, Семестр, Дисциплина, Оценка),
3. R4 (Дисциплина, Преподаватель),
4. R5 (Семестр, Дисциплина, Вид_Работы).

Третья нормальная форма

Отношение R находится в 3НФ тогда и только тогда, когда выполняются следующие условия (Кодд):

- R находится во второй нормальной форме;
- в таблицах отсутствует транзитивная зависимость.

Транзитивная зависимость – это когда неключевые столбцы зависят от значений других неключевых столбцов.



Третья нормальная форма

Таблица сотрудников во второй нормальной форме.

Табельный номер	ФИО	Должность	Подразделение	Описание подразделения
1	Иванов И.И.	Программист	Отдел разработки	Разработка и сопровождение приложений и сайтов
2	Сергеев С.С.	Бухгалтер	Бухгалтерия	Ведение бухгалтерского и налогового учета финансово-хозяйственной деятельности
3	John Smith	Продавец	Отдел реализации	Организация сбыта товаров
4	Романов Р.Р.	Директор	Отдел реализации	Организация сбыта товаров

Атрибут «*Описание подразделения*» содержит детальные сведения того подразделения, в котором работает сотрудник.

Атрибут «*Описание подразделения*» не зависит напрямую от первичного ключа.

Третья нормальная форма

Столбец «*Описание подразделения*» не связан напрямую с сотрудником, он связан напрямую со столбцом «*Подразделение*», который напрямую связан с сотрудником, ведь сотрудник работает в каком-то конкретном подразделении (**транзитивная зависимость**).

Таблица сотрудников в третьей нормальной форме.

Табельный номер	ФИО	Должность	Подразделение
1	Иванов И.И.	Программист	1
2	Сергеев С.С.	Бухгалтер	2
3	John Smith	Продавец	3

Таблица подразделений в третьей нормальной форме.

Идентификатор подразделения	Подразделение	Описание подразделения
1	Отдел разработки	Разработка и сопровождение приложений и сайтов
2	Бухгалтерия	Ведение бухгалтерского и налогового учета финансово-хозяйственной деятельности
3	Отдел реализации	Организация сбыта товаров

Приведение отношения к 3НФ

Определение **3НФ** запрещает наличие транзитивных зависимостей между первичными атрибутами, поэтому анализу подлежит только отношение

R1 (№ зачетки, Фамилия, Группа, Факультет),

в котором присутствует несколько первичных атрибутов.

Функциональная зависимость $F_7 = \text{Группа} \rightarrow \text{Факультет}$ в отношении R1, приводит к транзитивной зависимости $\text{№_зачетки} \rightarrow \text{Группа} \rightarrow \text{Факультет}$, от которой следует избавиться, оставив атрибут *Группа* в качестве внешнего ключа в отношении R1. В результате получатся две проекции:

$R1_1 (\text{№_зачетки}, \text{Фамилия}, \text{Группа})$

$R1_2 (\text{Группа}, \text{Факультет}).$

Результаты приведения к 3НФ

1. R1₁ (№ зачетки, Фамилия, Группа)
2. R1₂ (Группа, Факультет)
3. R3
(№ зачетки, Семестр, Дисциплина, Оценка)
4. R4 (Дисциплина, Преподаватель)
5. R5 (Семестр, Дисциплина, Вид_Работы)

Первая нормальная форма

- Каждая строка содержит данные, относящиеся к одному объекту или его части
- Каждый столбец должен иметь уникальное имя
- Каждый столбец содержит данные одного атрибута объекта. Все элементы столбца должны быть одного типа
- Две строки таблицы не могут быть идентичны
- Порядок строк и столбцов не имеет значение

Вторая нормальная форма

- Таблицы должны быть в первой нормальной форме
- Все неключевые атрибуты должны зависеть от всех ключевых атрибутов

Третья нормальная форма

- Таблица должны быть во второй нормальной форме
- Таблица не должна иметь неключевые атрибуты, находящиеся в транзитивной зависимости от первичного ключа

Приведение к БКНФ

Требования нормальной формы Бойса-Кодда следующие:

- Таблица должна находиться в третьей нормальной форме.
- Ключевые атрибуты составного первичного ключа не должны зависеть от неключевых атрибутов.

Ситуация, когда отношение будет находиться в 3NF, но не в BCNF, возникает при условии, что отношение имеет два (или более) возможных ключа, которые являются составными и имеют общий атрибут. Заметим, что на практике такая ситуация встречается достаточно редко, для всех прочих отношений 3NF и BCNF эквивалентны.

Главное правило нормальной формы Бойса-Кодда (BCNF) звучит следующим образом:

Часть составного первичного ключа не должна зависеть от неключевого столбца.

Приведение к БКНФ

Таблица проектов и кураторов.

Проект	Направление	Куратор
1	Разработка	Иванов И.И.
1	Бухгалтерия	Сергеев С.С.
2	Разработка	Иванов И.И.
2	Бухгалтерия	Петров П.П.
2	Реализация	John Smith
3	Разработка	Андреев А.А.

Существует
функциональная
зависимость

Направление →
Куратор

Таблица не находится в нормальной форме **Бойса-Кодда**, дело в том, что зная куратора, мы можем четко определить, какое направление он курирует, иными словами, часть составного ключа, т.е. «*Направление*», зависит от неключевого атрибута, т.е. «*Куратора*».

Недостатком данной структуры является то, что, например, по ошибке можно приписать куратора к одному направлению, хотя он может относиться к другому.

Приведение к БКНФ

Чтобы привести данную таблицу к нормальной форме Бойса-Кодда, необходимо, как всегда сделать декомпозицию данного отношения, т.е. разбить эту таблицу на несколько таблиц.

Таблица кураторов.

Идентификатор куратора	ФИО	Направление
1	Иванов И.И.	Разработка
2	Сергеев С.С.	Бухгалтерия
3	Петров П.П.	Бухгалтерия
4	Захар Петрович	Реализация
5	Андреев А.А.	Разработка

Таблица связи кураторов и проектов.

Проект	Идентификатор куратора
1	1
1	2
2	1
2	3
2	4
3	5

Приведение к БКНФ

Исходная таблица:

Номер клиента	Дата собеседования	Время собеседования	Номер комнаты	Номер сотрудника
C345	13.10.2003	13:00	103	A138
C355	13.10.2003	13:05	103	A136
C368	13.09.2003	13:00	102	A154
C366	13.09.2003	13:30	105	A207

В результате приведения к форме Бойса—Кодда получаются две таблицы:

Номер клиента	Дата собеседования	Время собеседования	Номер Сотрудника
C345	13.10.2003	13:00	A138
C355	13.10.2003	13:05	A136
C368	13.09.2003	13:00	A154
C366	13.09.2003	13:30	A207

Дата собеседования	Номер сотрудника	Номер комнаты
13.10.2003	A138	103
13.10.2003	A136	103
13.09.2003	A154	102
13.09.2003	A207	105

Исходное отношение

Шифр студента	Руководитель	Должность рук.	К1	К2	К3
1022	Иванов	доцент	01.09	10.09	20.09
4123	Петров	профессор	02.09	15.09	22.09

Первая нормальная форма

Шифр студента	Руководитель	Должность рук.	Консультации
1022	Иванов	доцент	01.09
1022	Иванов	доцент	10.09
1022	Иванов	доцент	20.09
4123	Петров	профессор	02.09
4123	Петров	профессор	15.09
4123	Петров	профессор	22.09

Вторая нормальная форма

Таблица «Студенты»

Шифр студента	Руководитель	Должность рук.
1022	Иванов	доцент
4123	Петров	профессор

Таблица «Консультации»

Шифр студента	Консультации
1022	01.09
1022	10.09
1022	20.09
4123	02.09
4123	15.09
4123	22.09

Третья нормальная форма

Таблица «Студенты»

Шифр студента	Руководитель
1022	Иванов
4123	Петров

Таблица «Консультации»

Шифр студента	Консультации
1022	01.09
1022	10.09
1022	20.09
4123	02.09
4123	15.09
4123	22.09

Таблица «Руководители»

Руководитель	Должность рук.
Иванов	доцент
Петров	профессор