

# Нормализация реляционной базы данных

# Нормализация

- **Нормализация** - формальная процедура, проверки и реорганизации отношений, в ходе которой создается оптимизированная структура БД, позволяющая избежать различных видов аномалий
- Нормализация подразумевает декомпозицию отношения на два или более, обладающих лучшими свойствами при добавлении, изменении и удалении данных
- **Задачи нормализации:**
  - Исключение из таблиц повторяющейся информации.
  - Создание структуры, в которой предусмотрена возможность ее будущих изменений и с минимальным влиянием на приложения.

# Дублирование данных

- Дублирование заключается в хранении идентичных данных
- **Необходимое**
  - важен сам факт идентичности данных - для поддержки ссылочной целостности (связей между сущностями)
- **Избыточное**
  - приводит к проблемам при обработке данных (аномалиям), которые не являются неразрешимыми в принципе, но **существенно усложняют** обработку
  - при избыточном дублировании могут возникать противоречия данных (возможность вывода двух взаимоисключающих утверждений на основе имеющихся данных)

Данные **дублированные**, но не избыточные

Книга	Студент
Книга 1	Иванов
Книга 2	Петров
Книга 3	Иванов

Данные по телефонам - **избыточные**

Книга	Студент	Телефон
Книга 1	Иванов	9631440779
Книга 2	Петров	9170550011
Книга 3	Иванов	9631440779

# Аномалии

- **Аномалия** – это такая ситуация в БД, которая приводит к противоречиям в данных, либо существенно усложняет их обработку.
  - **Аномалия модификации данных.** Необходимость изменения одного значения приводит к необходимости просмотра всей таблицы (для исключения противоречий).
  - **Аномалия удаления данных.** Состоит в том, что удаление данных из таблицы приводит к удалению информации не связанной напрямую с удаляемыми данными.
  - **Аномалия вставки данных.** Нельзя добавить одни данные без наличия других или же для вставки данных необходимо просмотреть всю таблицу (для исключения противоречий).
- **Основная причина аномалий** - хранение в одном отношении разнородной информации.

# Примеры избыточности и аномалий различных видов

Код_сотр	Фамилия	Должность	Ном_отдела	Наим_отдела
100	Иванов	Программист	10	Разработки ПС
101	Петров	Программист	10	Разработки ПС
102	Сидоров	Менеджер	10	Разработки ПС

- **Избыточность**

- Для каждого сотрудника отдела повторяются данные об отделе и должности

- **Аномалия модификации**

- Если какой-либо отдел решат переименовать необходимо изменить значение атрибута *Наим\_отдела* для многих экземпляров сущности *Сотрудник*.

- **Аномалия удаления**

- Если все сотрудники какого-либо отдела будут уволены, информация об отделе также будет утеряна.

- **Аномалия вставки**

- Если при внесении данных о каком-либо сотруднике оператор совершит ошибку и внесёт неверное наименование отдела (оставив верный номер), будет не ясно, какая из строк БД содержит правильную информацию (в базе будут содержаться противоречивые сведения)
- Вставка сведений о новом отделе компании невозможна без наличия сведений хотя бы об одном его сотруднике

# Описание предметной области

- Используется далее для иллюстрации проведения нормализации.
  - Каждая группа на кафедре учится по определённой специальности.
  - Нумерация групп - сквозная: не может быть групп с одинаковыми номерами, обучающихся на разных специальностях
  - Учебный план специальности определяет дисциплины, изучаемые в группах этой специальности
  - Одна и та же дисциплина может встречаться в учебных планах групп разных специальностей
  - Каждую дисциплину может вести как один преподаватель, так и несколько (даже в рамках одной группы)
  - Каждый преподаватель может вести несколько дисциплин, а также одну дисциплину у нескольких групп

# Исходное отношение

- Определите имеющиеся аномалии

## Cirriculum

CourseID
TeacherID
TeacherName
CourseName
SpecID
SpecName
Group1
Group2
Group3

<u>TeacherID</u>	TeacherName	<u>CourseID</u>	CourseName	SpecName	Group1	Group2	Group3
8	Иванов С.Ю.	1	Технологии программирования	Прикладная математика (ПМ), Автоматизированные системы (АС)	10	20	30
8	Иванов С.Ю.	2	Базы данных	Прикладная математика (ПМ)	10		
9, 10	Пестов О.А., Веснин Р.А.	2	Базы данных	Прикладная математика (ПМ)	20		

# Имеющиеся аномалии

- Если уволить преподавателя, может быть потеряна информация о дисциплине и даже группе (***аномалия удаления***)
- При открытии новой специальности, информацию о ней невозможно будет добавить, не определив какие дисциплины на ней будут читаться и кто их будет вести (***аномалия вставки***)
- При изменении названия дисциплины придётся просматривать всё отношение (***аномалия модификации***)

# ПЕРВАЯ НОРМАЛЬНАЯ ФОРМА (1НФ)

# Определение 1НФ

- Отношение находится в первой нормальной форме (1НФ) тогда и только тогда, когда:
  - определён его первичный ключ
  - все его атрибуты содержат атомарные значения
  - нет многозначных атрибутов (повторяющихся групп атрибутов)
- Для приведения отношения к 1НФ следует:
  - разделить сложные атрибуты на атомарные
  - «повторяющиеся» атрибуты совместить
  - пересмотреть состав атрибутов РК

# Приведение исходного отношения к 1НФ

## Curriculum

<u>CourseID</u>
<u>TeacherID</u>
<u>GroupNum</u>
TeacherName
CourseName
SpecID
SpecName

<u>TeacherID</u>	TeacherName	<u>CourseID</u>	CourseName	SpecID	SpecName	<u>GroupNum</u>
8	Иванов С.Ю.	1	Технологии программирования	ПМ	Прикладная математика	10
8	Иванов С.Ю.	1	Технологии программирования	ПМ	Прикладная математика	20
8	Иванов С.Ю.	1	Технологии программирования	АС	Автоматизированные системы	30
8	Иванов С.Ю.	2	Базы данных	ПМ	Прикладная математика	10
9	Пестов О.А.	2	Базы данных	ПМ	Прикладная математика	20
10	Веснин Р.А.	2	Базы данных	ПМ	Прикладная математика	20

# ВТОРАЯ НОРМАЛЬНАЯ ФОРМА (2НФ)

# Определение 2НФ

- Отношение находится во второй нормальной форме (2НФ), тогда и только тогда, когда оно находится в 1НФ, и каждый его неключевой атрибут функционально полно зависит от первичного ключа
  - Если первичный ключ отношения состоит из одного атрибута, оно автоматически находится во 2НФ
- Для приведения отношения к 2НФ следует:
  - выделить **неключевые атрибуты**, которые зависят только от части первичного ключа и перенести их в новое отношение вместе с той частью ключа, от которой они зависят

# Приведение отношения ко 2НФ

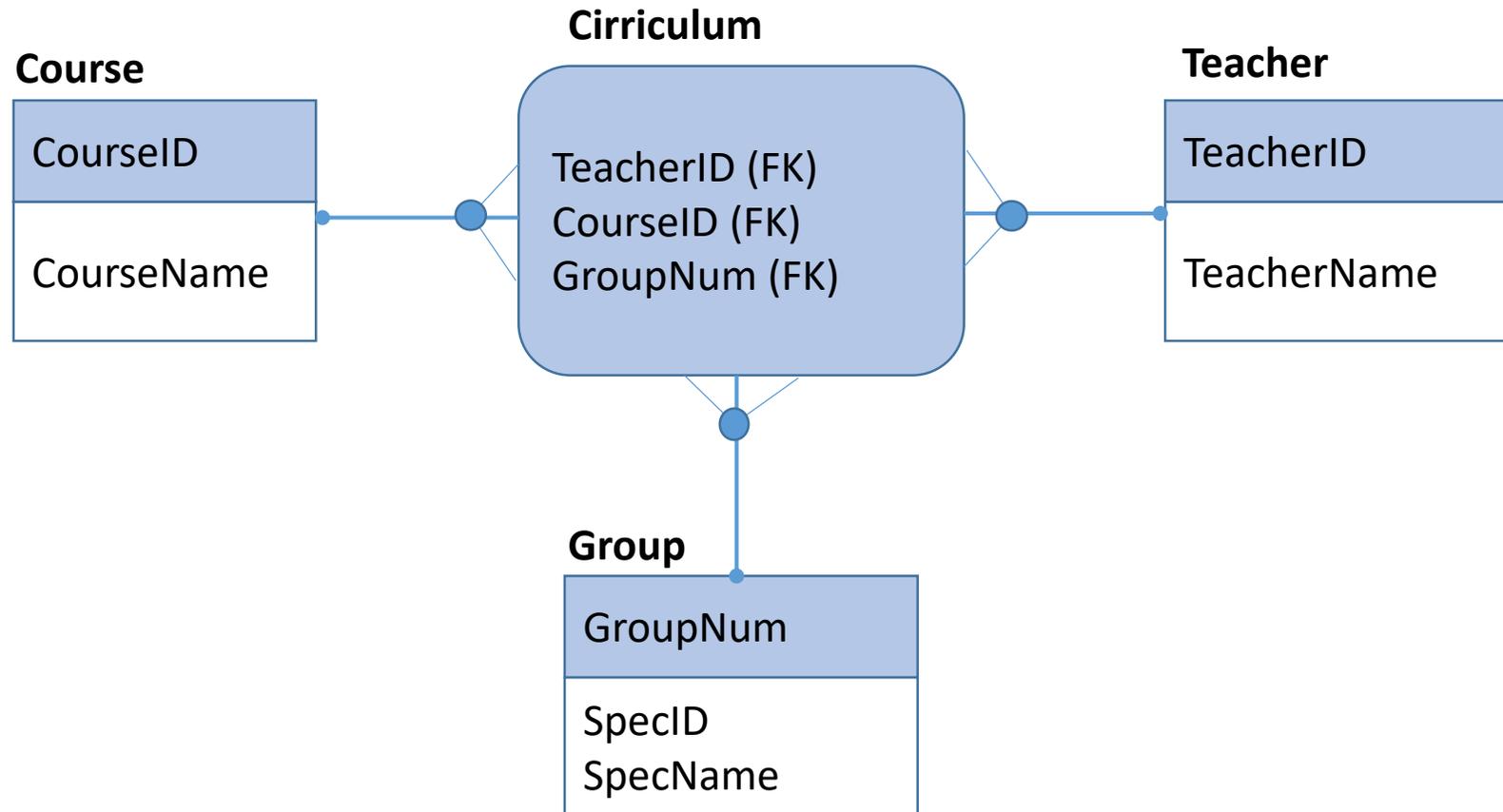
- Имеющиеся функциональные зависимости от части ключа:
  - TeacherID -> Teacher
  - CourseID -> Course
  - Group -> { SpecID, SpecName }

<u>TeacherID</u>	<u>CourseID</u>	<u>GroupNum</u>
8	1	10
8	1	20
8	1	30
8	2	10
9	2	20
10	2	20

<u>TeacherID</u>	TeacherName	<u>CourseID</u>	CourseName
8	Иванов С.Ю.	1	Технологии программирования
9	Пестов О.А.		
10	Веснин Р.А.	2	Базы данных

<u>GroupNum</u>	SpecID	SpecName
10	ПМ	Прикладная математика
20	ПМ	Прикладная математика
30	АС	Автоматизированные системы

# Приведение сущности ко 2НФ



# Аномалии

- Были устранены приведённые ранее аномалии удаления, вставки и модификации
  - Удаление информации о преподавателе не влияет на информацию о дисциплине и группе
  - Можно добавить информацию о новой специальности, не зная какие дисциплины на ней будут читаться и кто их будет вести
  - Название дисциплины в случае необходимости обновится только в одном месте
- Осталась аномалия модификации
  - При изменении названия специальности придётся просматривать всё отношение (менять его для каждой группы, обучающейся по этой специальности)

ТРЕТЬЯ НОРМАЛЬНАЯ ФОРМА  
(ЗНФ)

# Определение 3НФ

- Отношение находится в третьей нормальной форме (3НФ) тогда и только тогда, когда оно находится в 2НФ и не содержит транзитивных функциональных зависимостей между неключевыми атрибутами и первичным ключом
  - ни один **неключевой атрибут** не должен функционально зависеть от другого **неключевого атрибута**
- Для приведения сущности к 3НФ следует:
  - создать новое отношение и перенести в него атрибуты с одной и той же зависимостью от неключевого атрибута

# Приведение отношения к 3НФ

- Имеющиеся зависимости между неключевыми атрибутами:
  - SpecID -> SpecName

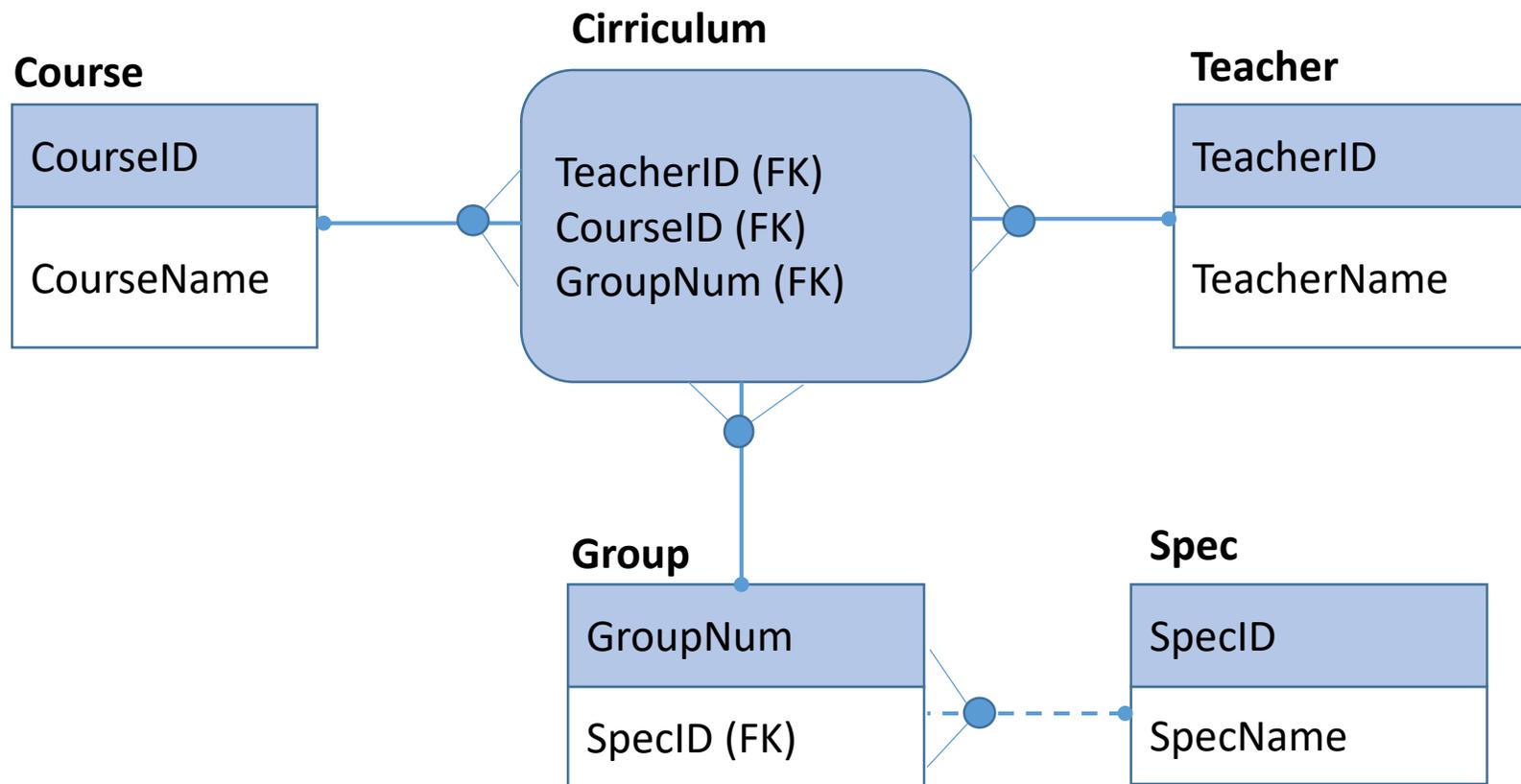
<u>TeacherID</u>	TeacherName
8	Иванов С.Ю.
9	Пестов О.А.
10	Веснин Р.А.

<u>CourseID</u>	CourseName
1	Технологии программирования
2	Базы данных

<u>TeacherID</u>	<u>CourseID</u>	<u>GroupNum</u>
8	1	10
8	1	20
8	1	30
8	2	10
9	2	20
10	2	20

<u>GroupNum</u>	SpecID	<u>SpecID</u>	SpecName
10	ПМ	ПМ	Прикладная математика
20	ПМ	ПМ	Прикладная математика
30	АС	АС	Автоматизированные системы

# Приведение сущности к 3НФ



# Результаты нормализации

- **Плюсы:**

- Снижается избыточность данных
  - Таблицы содержат только один элемент избыточных данных – это поле связи, присутствующее одновременно у родительской и дочерних таблиц.

- **Минусы:**

- Сильно увеличивается число таблиц:
  - Теряется восприятие системы в целом.
  - Усложняются запросы, приходится объединять данные из разных таблиц (скорость выполнения падает)

Результат «нормализации» автомобиля

