

Лекция 2

Реляционная модель: нормальные
формы

Проектирование БД

Реляционные БД

В основе реляционной БД лежит реляционная модель данных, в которой все данные представляются в виде таблиц, и операции реляционной алгебры для работы с отношениями.

Таблица состоит из строк и столбцов и имеет имя, уникальное в пределах БД.

Строки таблиц содержат сведения о сущностях и называются **записями**, а в реляционной алгебре - **кортежами**.

Значения свойств объектов - атрибутов сущностей - составляют столбцы - **поля** таблицы.

Множество возможных значений атрибута сущностей составляют **домен**.

Каждое поле имеет имя, уникальное в таблице. Строки имен не имеют.

Порядок следования строк не определен, их количество не ограничено.



Данные об объектах одного класса хранят в одной таблице.

Нормальные формы

Ненормализованная форма или нулевая нормальная форма (UNF) базы данных

По реляционной теории строки в таблицах не должны быть пронумерованы, т.е. порядок строк не имеет значения, так же как не имеет значения порядок столбцов.

Т.е. например, если мы поменяем порядок столбцов, или порядок строк, ничего измениться не должно, это не должно ни на что повлиять.

Таким образом по реляционной теории мы не можем обратиться к определённой строке или столбцу по ее номеру.

И если Ваши таблицы соблюдают эти принципы, то можно переходить к нормализации базы данных.

Рассмотрим небольшой пример. Достаточно часто в Excel можно встретить таблицы следующего вида

Порядковый номер строки	A	B
1	Иван	Иванов
2	Сергей	Сергеев
3	John	Smith
4	Иван	Иванов

first_name	last_name
Иван	Иванов
Сергей	Сергеев
John	Smith
Иван	Иванов

Нормализация отношений

Разбиение (декомпозиция) таблиц на несколько более мелких представляет собой процедуру нормализации отношений.

Окончательная цель нормализации - получение такой модели данных, в которой **каждый факт появляется лишь в одном месте**, т.е. исключается избыточность информации и ее последствия.

Выделяют следующую последовательность нормальных форм: 1НФ, 2НФ, 3НФ, НФБК, 4НФ, 5НФ.

Каждой нормальной форме соответствует определенный набор ограничений на структуру данных. Каждая следующая форма отличается от предыдущей некоторым дополнительным ограничением.

Нормализация отношений - последовательный процесс приведения структуры данных к формам более высокого уровня. Цель нормализации - исключить:

- избыточность данных и как следствие:
- потенциальную противоречивость (аномалии обновления);
- аномалии включения (при дополнении неполными новыми данными);
- аномалии удаления.

Требования первой нормальной формы (1NF)

1. В таблице не должно быть дублирующих строк

2. В каждой ячейке таблицы хранится атомарное значение (одно не составное значение, отсутствуют массивы и списки в любом виде)

3. В столбце хранятся данные одного типа

4. Не должно быть полей, которые обозначают различные виды одного и того же, например, товаров.

Пример приведения таблицы к первой нормальной форме

Следующая таблица не находится даже в первой нормальной форме, так как у нас есть дублирующие строки (John Smith), а в некоторых ячейках хранятся списки значений (*каждый номер телефона — это одно значение*)

Таблица в ненормализованном виде.

Таблица в нормализованном виде:

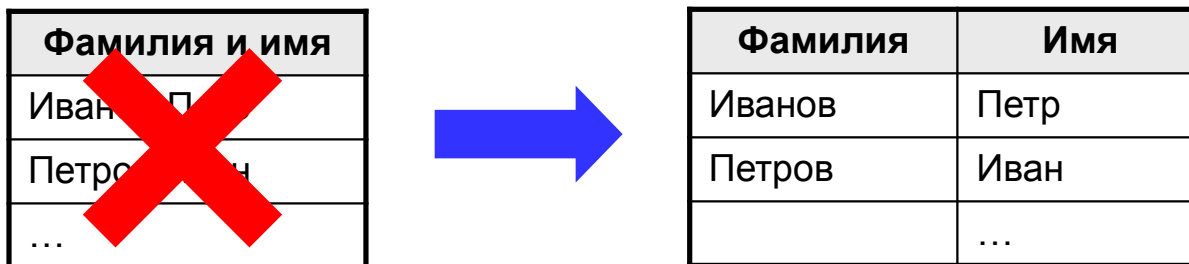
Сотрудник	Контакт	Сотрудник	Телефон	Тип телефона
Иванов И.И.	123-456-789	Иванов И.И.	123-456-789	
Сергеев С.С.	Рабочий тел. Домашний 999	Иванов И.И.	987-654-321	
		Сергеев С.С.	555-666-777	Рабочий телефон
John Smith	123-456-789	Сергеев С.С.	777-888-999	Домашний телефон
John Smith	123-456-789	John Smith	123-456-789	

Нормализация базы данных

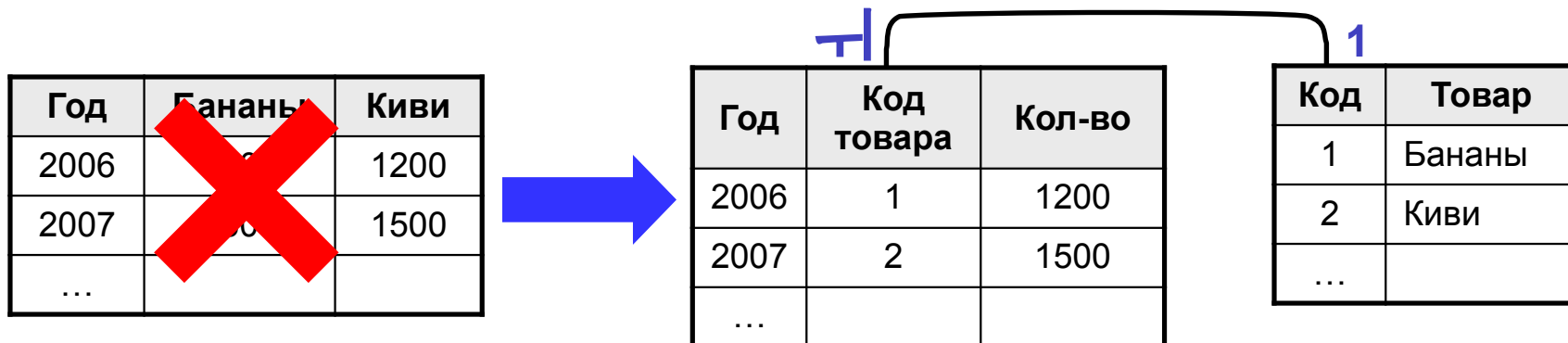
Нормализация (нормальные формы) – это разработка такой структуры БД, в которой нет избыточных данных и связей.

Основные принципы:

- ❑ Любое поле должно быть **неделимым** – 1-я НФ.



- ❑ Не должно быть полей, которые обозначают различные виды одного и того же, например, товаров.



Требования второй нормальной формы (2NF)

Чтобы база данных находилась во второй нормальной форме (2NF), необходимо чтобы ее таблицы удовлетворяли следующим требованиям:

- Таблица должна находиться в первой нормальной форме
- Таблица должна иметь ключ
- Все неключевые столбцы таблицы должны зависеть от полного ключа (*в случае если он составной*)

Ключ – это столбец или набор столбцов, по которым гарантировано можно отличить строки друг от друга, т.е. ключ идентифицирует каждую строку таблицы.

По ключу мы можем обратиться к конкретной строке данных в таблице.

Если ключ составной, т.е. состоит из нескольких столбцов, то все остальные неключевые столбцы должны зависеть от всего ключа, т.е. от всех столбцов в этом ключе.

Если какой-то атрибут (столбец) зависит только от одного столбца в ключе, значит, база данных не находится во второй нормальной форме.

Приведение к второй нормальной форме

ФИО	Должность	Подразделение	Описание подразделения
Иванов И.И.	Программист	Отдел разработки	Разработка и сопровождение приложений и сайтов
Сергеев С.С.	Бухгалтер	Бухгалтерия	Ведение бухгалтерского и налогового учета финансово-хозяйственной деятельности
John Smith	Продавец	Отдел реализации	Организация сбыта продукции

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 продукции

Первая нормальная форма - 1NF

Данные о посещении поликлиники, сгруппированные по пациентам, представляют ненормализованное отношение:

Пациент	Г/р	Соц/полож	Отделение	Врач	Кат.	Оклад	Диагноз
Иванов И.И.	1940	Пенсионер	Терапевтическое	Петрова	2	2200	Радикулит
			Физиотерапевтич	Горячева	1	2100	Радикулит
			Хирургическое	Ножиков	3	2300	Ушиб колен
Васильев В.В.	1951	Служащий	Терапевтическое	Петрова	2	2200	ОРВИ
			Офтальмологиче	Зрачкова	4	2400	Соринка

Таблица находится **в первой нормальной форме** (1НФ), если значения полей во всех записях заданы однозначно, т.е. значения всех атрибутов каждой сущности **атомарны** :

Пациент	Г/р	Соц/полож	Отделение	Врач	Кат.	Оклад	Диагноз
Иванов И.И.	1940	Пенсионер	Терапевтическое	Петрова	2	2200	Радикулит
Иванов И.И.	1940	Пенсионер	Физиотерапевтич	Горячева	1	2100	Радикулит
Иванов И.И.	1940	Пенсионер	Хирургическое	Ножиков	3	2300	Ушиб колен
Васильев В.В.	1951	Служащий	Терапевтическое	Петрова	2	2200	ОРВИ
Васильев В.В.	1951	Служащий	Офтальмологиче	Зрачкова	4	2400	Соринка

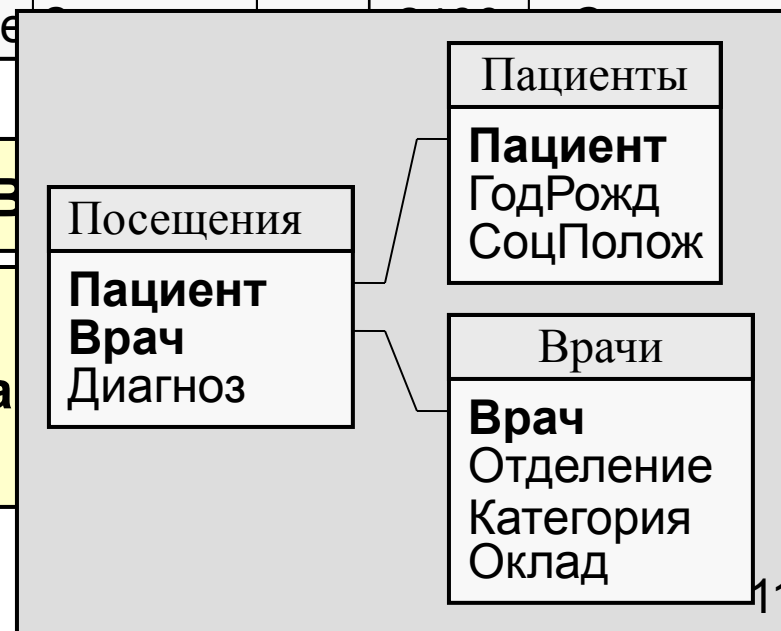
Вторая нормальная форма - 2NF

Таблица находится **во второй нормальной форме** (2НФ), если она удовлетворяет определению 1НФ и не содержит неполных зависимостей неключевых полей от возможного первичного ключа.

Пациент	Г/р	Соц/полож	Отделение	Врач	Кат.	Оклад	Диагноз
Иванов И.И.	1940	Пенсионер	Терапевтическое	Петрова	2	2200	Радикулит
Иванов И.И.	1940	Пенсионер	Физиотерапевтич	Горячева	1	2100	Радикулит
Иванов И.И.	1940	Пенсионер	Хирургическое	Ножилов	3	2300	Ушиб колен
Васильев В.В.	1951	Служащий	Терапевтическое	Петрова	2	2200	ОРВИ
Васильев В.В.	1951	Служащий	Офтальмологичес				

Посещения
Пациент ГодРожд СоцПолож Отделение Врач Категория Оклад Диагноз

Возможный ключ: Пациент-В
Неполные зависимости: ГодРожд, СоцПолож от Па Отделен., Катег., Оклад от



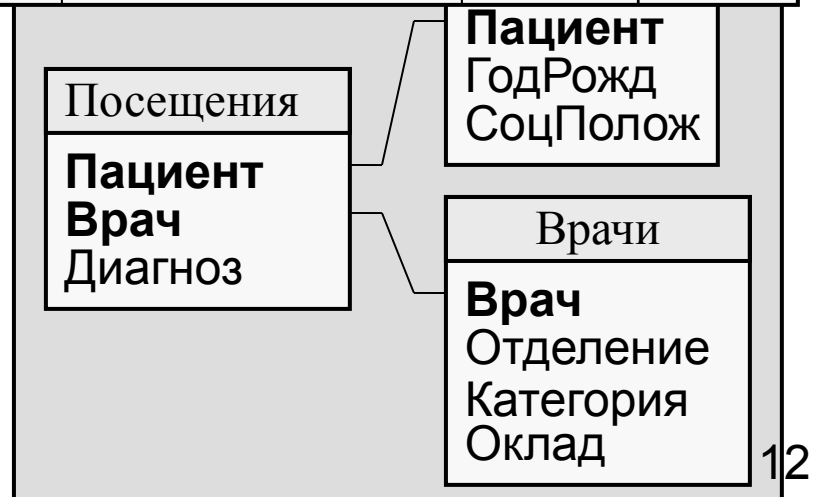
Приведение структуры данных к 2NF

Пациент	Г/р	Соц/полож	Отделение	Врач	Кат.	Оклад	Диагноз
Иванов И.И.	1940	Пенсионер	Терапевтическое	Петрова	2	2200	Радикулит
Иванов И.И.	1940	Пенсионер	Физиотерапевтич	Горячева	1	2100	Радикулит
Иванов И.И.	1940	Пенсионер	Хирургическое	Ножиков	3	2300	Ушиб колен
Васильев В.В.	1951	Служащий	Терапевтическое	Петрова	2	2200	ОРВИ
Васильев В.В.	1951	Служащий	Офтальмологиче	Зрачкова	4	2400	Соринка

Пациенты		
Пациент	ГодРожд	СоцПолож
Иванов И.И.	1940	Пенсионер
Васильев В.В.	1951	Служащий

Посещения		
Пациент	Врач	Диагноз
Иванов И.И.	Петрова	Радикулит
Иванов И.И.	Горячева	Радикулит
Иванов И.И.	Ножиков	Ушиб колен
Васильев В.В.	Петрова	ОРВИ
Васильев В.В.	Зрачкова	Соринка

Врачи			
Врач	Отделение	Катег.	Оклад
Петрова	Терапевтическое	2	2200
Горячева	Физиотерапевтич	1	2100
Ножиков	Хирургическое	3	2300
Зрачкова	Офтальмологиче	4	2400



Требования третьей нормальной формы (3NF)

Требование третьей нормальной формы (3NF) заключается в том, чтобы в таблицах отсутствовала транзитивная зависимость.

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

Если в первой нормальной форме наше внимание было нацелено на соблюдение реляционных принципов, во второй нормальной форме в центре нашего внимания был первичный ключ, то в третьей нормальной форме все наше внимание уделено столбцам, которые не являются первичным ключом, т.е. неключевым столбцам.

Чтобы нормализовать базу данных до третьей нормальной формы, необходимо сделать так, чтобы в таблицах отсутствовали неключевые столбцы, которые зависят от других неключевых столбцов.

Иными словами, неключевые столбцы не должны пытаться играть роль ключа в таблице, т.е. они действительно должны быть неключевыми столбцами, такие столбцы не дают возможности получить данные из других столбцов, они дают возможность посмотреть на информацию, которая в них содержится, так как в этом их назначение.

Главное правило третьей нормальной форме (3NF) звучит следующим образом: **таблица должна содержать правильные неключевые столбцы**

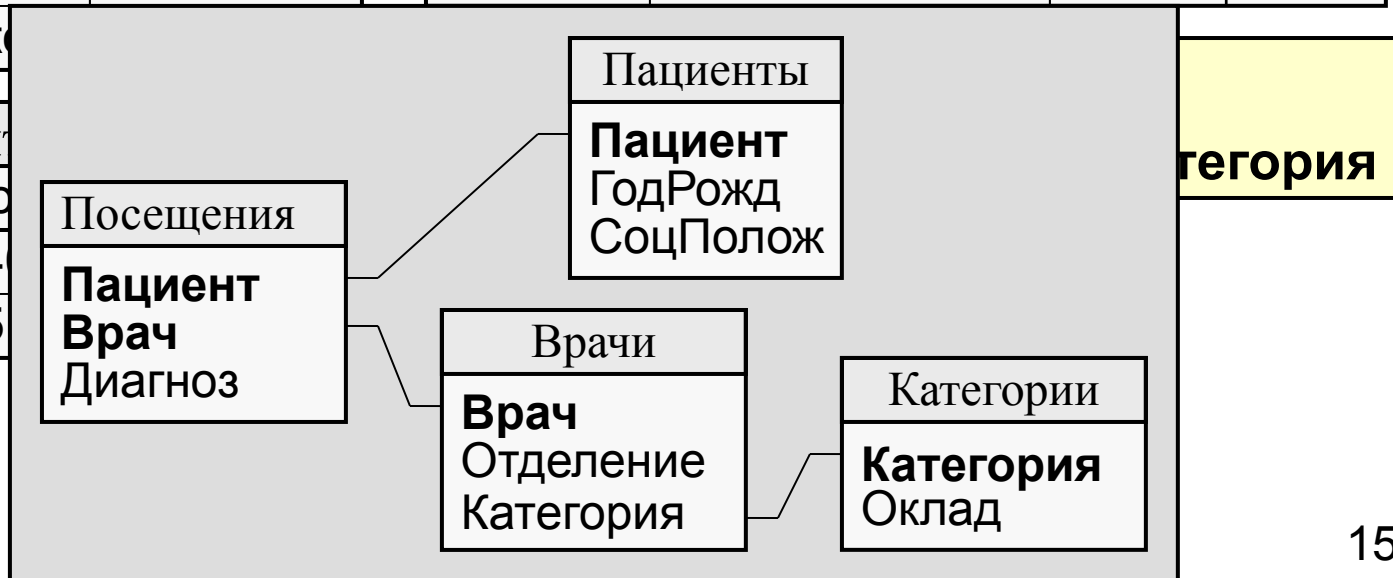
Третья нормальная форма - 3NF

Таблица находится **в третьей нормальной форме (3НФ)**, если она удовлетворяет определению 2НФ и ни одно из ее неключевых полей не зависит функционально (однозначно) от любого другого неключевого поля.

Посещения		
Пациент	Врач	Диагноз
Иванов И.И.	Петрова	Радикулит
Иванов И.И.	Горячева	Радикулит
Иванов И.И.	Ножилов	Ушиб колен
Васильев В.В.	Петрова	ОРВИ
Васильев В.В.	Зрачков	

Врачи			
Врач	Отделение	Катег.	Оклад
Петрова	Терапевтическое	2	2200
Горячева	Физиотерапевтич	1	2100
Ножилов	Хирургическое	3	2300
Зрачкова	Офтальмологиче	4	2400

Пациент	
Пациент	ГодРод
Иванов И.И.	194
Васильев В.В.	195



Приведение структуры данных к 3NF

Посещения

Пациент	Врач	Диагноз
Иванов И.И.	Петрова	Радикулит
Иванов И.И.	Горячева	Радикулит
Иванов И.И.	Ножилов	Ушиб колен
Васильев В.В.	Петрова	ОРВИ
Васильев В.В.	Зрачкова	Соринка

Пациенты

Пациент	ГодРожд	СоцПолож
Иванов И.И.	1940	Пенсионер
Васильев В.В.	1951	Служащий

Врачи

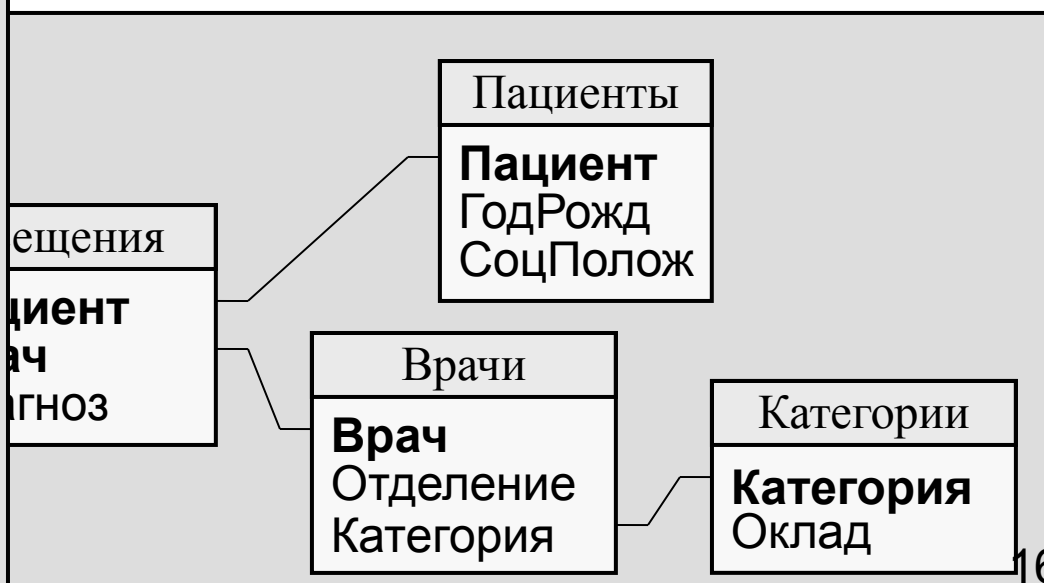
Врач	Отделение	Катег.	Оклад
Петрова	Терапевтическое	2	2200
Горячева	Физиотерапевтич	1	2100
Ножилов	Хирургическое	3	2300
Зрачкова	Офтальмологиче	4	2400

Врачи

Врач	Отделение	Катег.
Петрова	Терапевтическое	2
Горячева	Физиотерапевтич	1
Ножилов	Хирургическое	3
Зрачкова	Офтальмологиче	4

Категории

Категория	Оклад
1	2100
2	2200
3	2300
4	2400



Создание таблиц в СУБД

На практике процесс нормализации заканчивается приведением структуры данных к 3НФ.

3НФ является компромиссом между полной нормализацией структуры данных.

Сотрудники

N	ФИО	Должность	Оклад	Ставка	Кабинет	Вн.тел.	Отделен.
14	Сидорова С.С.	Ст.мед.сестра	1800	1,5	23	3-25	Терапевт.
16	Ножиков Н.Н.	Хирург	2300	1	16	1-24	Хирургич.
17	Петрова П.П.	Терапевт	2200	1,5	26	2-67	Терапевт.

нормализации структуры данных.

Создание таблиц состоит в задании имени таблицы, имен полей, указании типов данных в полях и выделении ключевых полей. В некоторых СУБД автоматически создается поле для неизменяемого порядкового номера записи, которое может быть использовано в качестве ключа.

Сотрудники

Поле	Тип	Ключ
КодСотрудника	Счетчик	*
ФИО	Текст	
Должность	Текст	
Оклад	Денежный	
Ставка	Веществ	
Кабинет	Целый	
ВнутрТелефон	Текст	
Отделение	Текст	17

Ключевые поля и связи

Таблицы, создаваемые в СУБД, обычно соответствуют 2НФ, т.е. значения атрибутов атомарны и все поля определяются первичным ключом.

Сотрудники							
№	ФИО	Должность	Оклад	Ставка	Кабинет	Вн. тел.	Отделен.
14	Сидорова С.С.	Ст.мед.сестра	1800	1,5	23	2-54	2
16	Ножиков Н.Н.	Хирург	2300	1	15	1-44	3
17	Петрова П.П.	Терапевт	2200	1,5	26	2-67	2

Для однозначного соответствия данных в таблицах устанавливаются связи между ними путем задания ссылок на значения ключевых полей.

Поле, с помощью значений которого однозначно определяется подстановка данных из записей другой таблицы, называется **внешним ключом**.

Подразделения		
№	Наименование	
1	Дирекция	
2	Терапевтическое отд.	
3	Хирургическое отд.	
4	Физиотерапевтич.каб.	

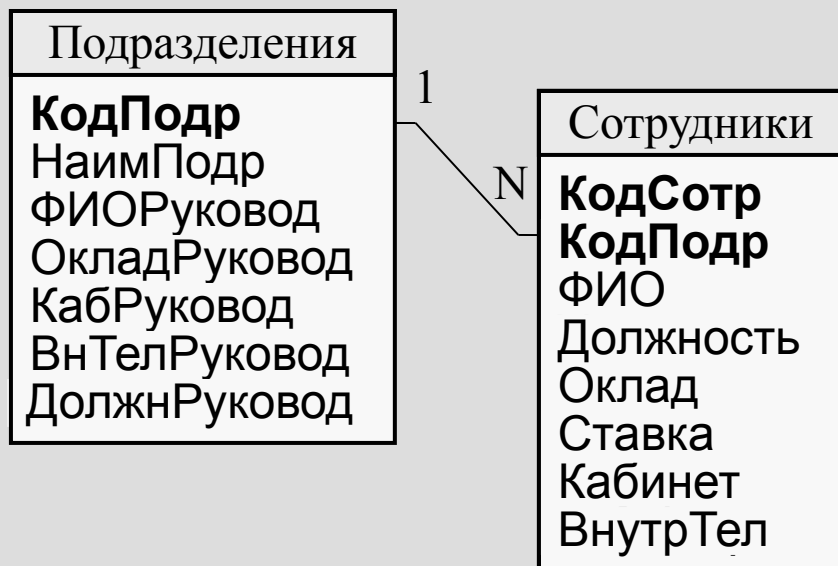
Тип данных внешнего ключа должен совпадать с типом данных первичного ключа.

Установление связей между таблицами

Для установления связей между таблицами и задания их типов в разных СУБД используются разные приемы.

Сотрудники							
N	ФИО	Должность	Оклад	Ставка	Кабинет	Вн.тел.	Отделен.
14	Сидорова С.С.	Ст.мед.сестра	1800	1,5	23	3-25	2
16	Ножиков Н.Н.	Хирург	2300	1	16	1-24	3
17	Петрова П.П.	Терапевт	2200	1,5	26	2-67	2

Структура данных



Подразделения

N	Наименование
1	Дирекция
2	Терапевтическое отд.
3	Хирургическое отд.
4	Физиотерапевтич.каб.

Обычно поля внешних ключей располагают сразу после полей первичного ключа. Часто их включают в состав первичного ключа

Этапы проектирования базы данных

Проект реляционной базы данных - это набор взаимосвязанных отношений, для которых определены все атрибуты, первичные ключи и ограничения целостности.

Процесс проектирования БД представляет собой последовательность переходов от неформального словесного описания предметной области к формализованному описанию объектов предметной области в терминах некоторой модели.

Выделяют четыре этапа проектирования реляционной базы данных:

- системный анализ предметной области;
- построение инфологической модели предметной области;
- даталогическое проектирование базы данных;
- физическое проектирование базы данных.

Между вторым и третьим этапами принимается решение об использовании конкретной СУБД для реализации проекта.

Системный анализ предметной области

На этапе системного анализа предметной области выполняется подробное словесное описание ее объектов и связей между ними.

Существуют два подхода к выбору состава и структуры предметной области:

- функциональный подход;
- предметный подход.

Функциональный подход - принцип движения «от задач». Он применяется, когда известны функции группы лиц и комплексов проблем, для обслуживания информационных потребностей которых создается БД или ИС.

Предметный подход - когда невозможно выделить минимальный набор объектов предметной области, и информационные потребности будущих пользователей БД или ИС жестко не фиксируются. В описание включаются только наиболее существенные для предметной области объекты и взаимосвязи. В этом случае БД наращивается по мере ее эксплуатации.

На практике часто применяется компромиссный вариант.

Итог: описание информации об объектах, которая должна храниться; формулировки задач, которые будут решаться, описание алгоритмов их решения, описание генерируемых выходных документов, описание входных документов.²¹

Автоматизация этапа системного анализа

На этапе системного анализа выполняется семантическое (смысловое, концептуальное) моделирование БД с помощью CASE-технологий (Computer Aided Software Engineering).

Цель - организация интерфейса проектировщика и конечного пользователя с ИС на уровне представлений предметной области, а не на уровне структур данных.

Пакет **ERwin 7.2** - имеет поддержку 25 СУБД. Использует два уровня представления модели: логический и физический. Процесс построения БД состоит из трех этапов.

1. Определение сущностей, связей между ними, задание атрибутов и ключей.
2. Назначение соответствий имя сущности - имя таблицы, атрибут сущности - поле таблицы, задание ограничений.
3. Генерация базы данных.

ERwin имеет интерфейс обычной программы.

Построение инфологической модели

Инфологическая модель должна включать такое формализованное описание предметной области, которое легко читается неспециалистами в области информационных технологий.

Это описание должно быть емким, представлять глубину и корректность проработки проекта и не быть привязанным к конкретной СУБД.

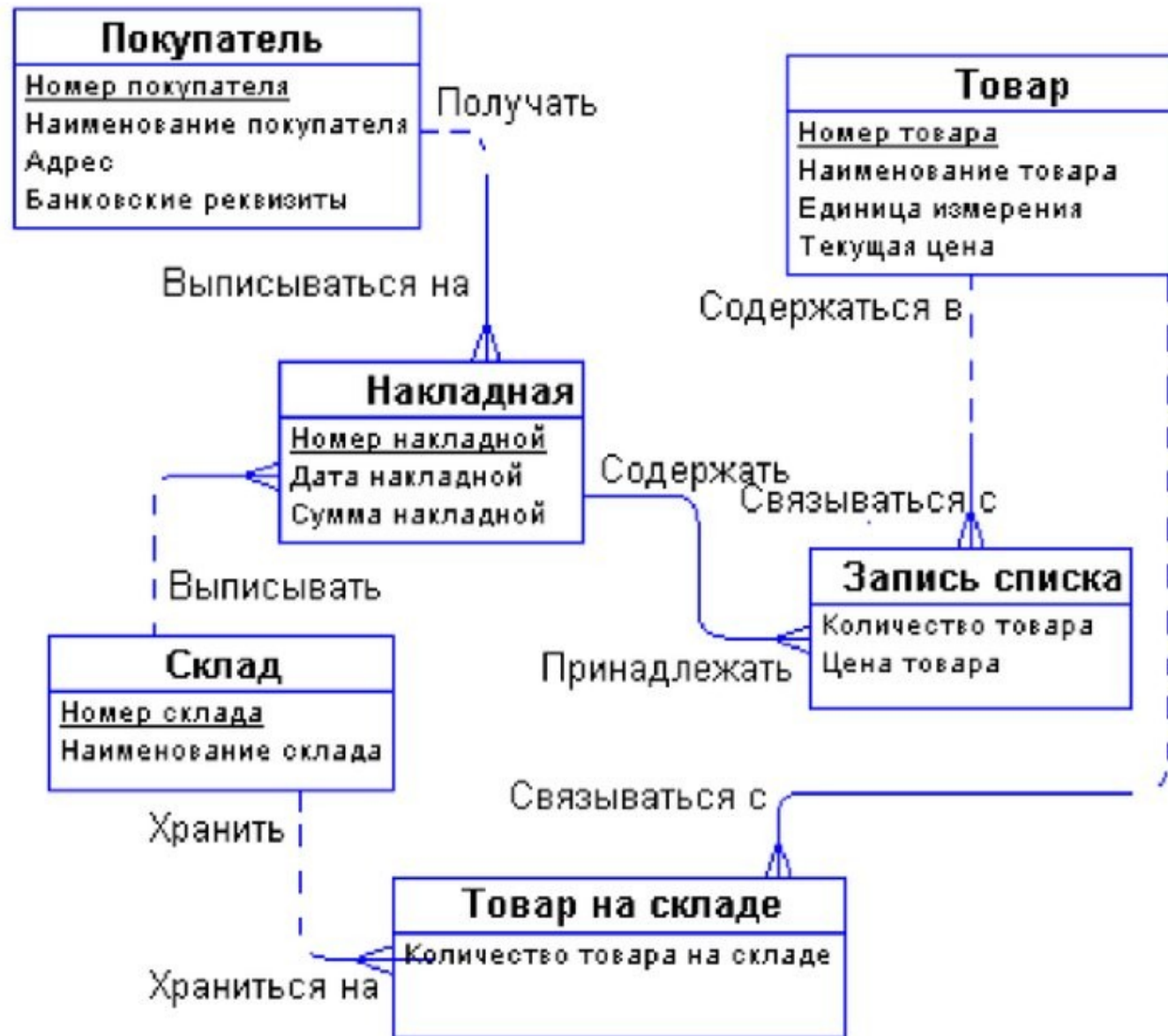
Для реализации этапа построения инфологической модели стандартом является модель «сущность-связь» (ER-модель).

Существуют программные CASE-средства преобразования проекта БД из ER-модели в реляционную даталогическую модель, соответствующую конкретной СУБД. Однако, не существует единой общепринятой системы обозначений для ER-модели, и разные CASE-системы используют разные графические нотации.

При создании относительно несложных проектов (до 30 таблиц) и заранее очевидном выборе конкретной СУБД для реализации проекта этапы построения инфологической модели и даталогического проектирования могут быть совмещены. В этом случае при построении информационной модели предметной области используется терминология и графические примитивы выбранной СУБД.²³

Модель «сущность-связь» (ER-модель)

20.09.2022



Даталогическое проектирование

На этапе даталогического проектирования разрабатывается схема БД в терминах конкретной СУБД.

Состав таблиц, их наименования.

Перечень, порядок следования и типы полей.

Ключи и индексы, их состав и свойства.

Связи, их типы и свойства.

Свойства, параметры и форматы данных для полей.

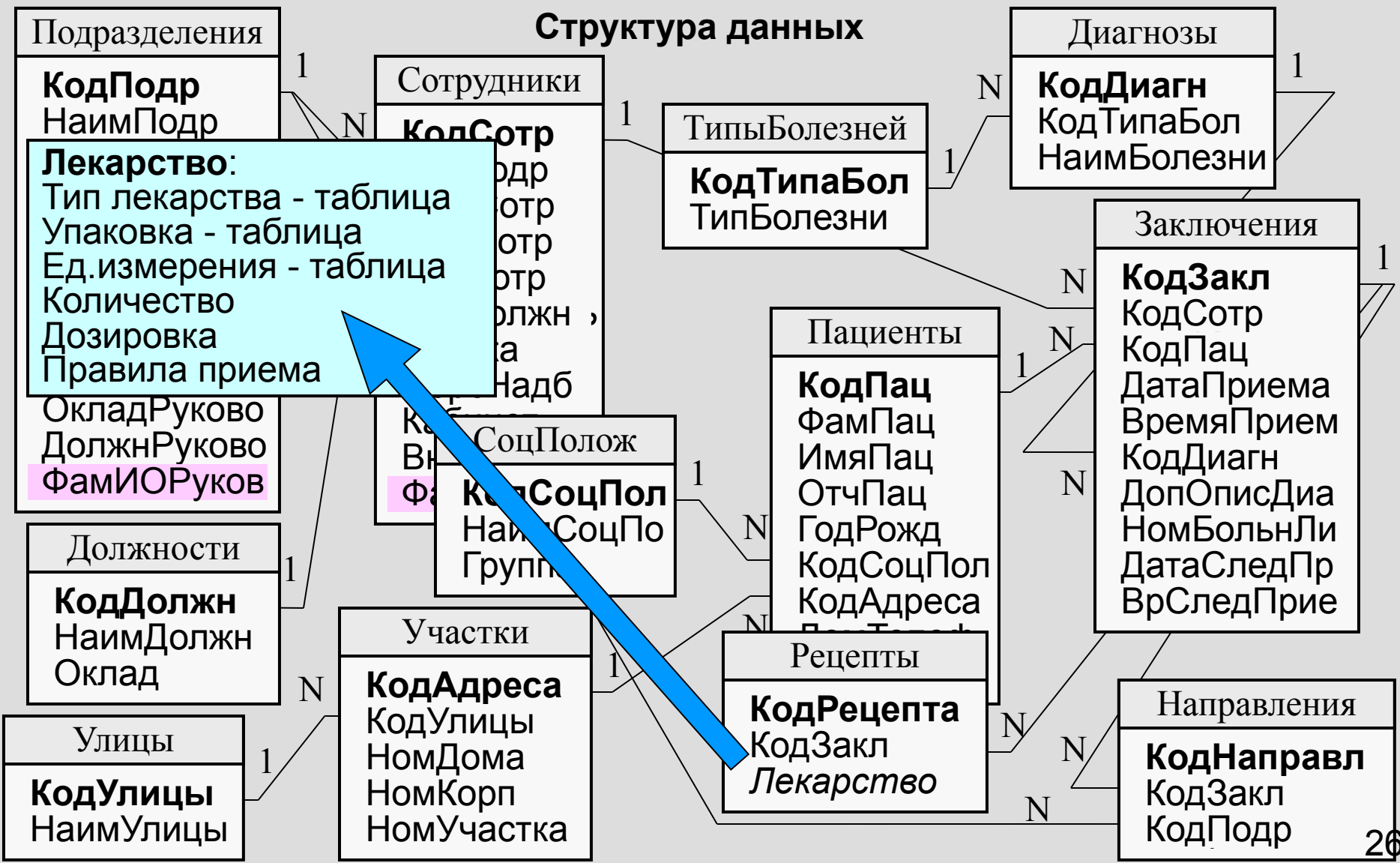
Два пути проектирования схемы БД:

- **декомпозиция** отношений на основе анализа зависимостей между атрибутами - замена отношений их проекциями методом нормализации - классический вариант;
- **синтез** схемы из исходных элементарных зависимостей между объектами предметной области.

Пример даталогического проектирования

Схема базы данных «Поликлиника»

Структура данных



Понятие транзакции

Под **транзакцией** понимается воздействие на БД, переводящее ее из одного целостного состояния в другое.

Воздействие на БД выражается в изменении данных в таблицах: изменение района проживания пациента - удаление всех его данных или списание в архив.

При успешном завершении транзакции выполняется одновременное подтверждение всех изменений.

Если одно из изменений завершается неуспешно, то не подтверждается ни одно из изменений.

В результате должен быть произведен **откат** к прежнему состоянию БД.

Корректное выполнение транзакций обеспечивает **семантическую (смысловую) целостность** БД.

