

Тема 1. Основы инфологического моделирования баз данных

1. Основные понятия проектирования баз данных

В широком смысле база данных – это совокупность сведений о конкретных объектах реального мира и их отношениях в какой-либо предметной области.

Под **предметной областью** (ПО) базы данных понимается часть реального мира, сведения о которой представляют интерес с точки зрения пользователя.

БД отличается от любого другого набора записей тем, что наряду с данными *содержит своё собственное описание*.
Пользователями базы данных могут быть различные прикладные программы, а также люди.

Создание баз данных, их поддержка и обеспечение доступа пользователей к ним осуществляется с помощью специального программного средства – системы управления базами данных.

Система управления базами данных (СУБД) – это комплекс программных и языковых средств, необходимых для создания баз данных, поддержания их в актуальном состоянии и организации поиска в них необходимой информации.

Проект базы данных следует начинать с *анализа предметной области и выявления требований к ней отдельных пользователей* (например, сотрудников организации, для которых создается база данных).

Проектирование обычно поручается человеку (или группе лиц) – **администратору базы данных (АБД)**. Им может быть как специально выделенный сотрудник организации, так и будущий пользователь базы данных, достаточно хорошо знакомый с машинной обработкой данных.

Объединяя частные представления о содержимом базы данных, полученные в результате опроса пользователей, и свои представления о данных, которые могут потребоваться в будущих приложениях, ***АБД сначала создает обобщенное неформальное описание создаваемой базы данных.***

Это описание, выполненное с использованием естественного языка, математических формул, таблиц, графиков и других средств, понятных всем людям, работающим над проектированием базы данных, называют ***инфологической моделью данных.***

Анализ предметной область

Требования
пользователей

Нормативные
документы

Существующая
практика
представления
информации

Анализ
аналогов



Инфологическая модель данных

Описание, не привязанное к конкретной СУБД, выполненное с использованием естественного языка, математических формул, таблиц, графиков и других средств.



Даталогическая модель данных

Описание на языке конкретной СУБД.



Физическая модель данных

Внутреннее описание хранимых в СУБД данных.



База данных

2. Инфологическая модель данных "Сущность-связь"

2.1. Основные понятия

Модель «сущность-связь» (entity-relationship model) предложена американским исследователем в области баз данных Питером Ченом в 1976 году.

В различных вариантах она вошла в состав многих автоматизированных средств поддержки проектирования информационных систем.

В настоящее время нет единого стандарта этой модели, но есть набор общих конструкций, лежащих в основе большинства её вариантов. Эти общие конструкции мы и изучим здесь.

Цель инфологического моделирования – обеспечение наиболее естественных для человека способов сбора и представления той информации, которую предполагается хранить в создаваемой базе данных.

Поэтому инфологическую модель данных пытаются строить по аналогии с естественным языком (последний не может быть использован в чистом виде из-за сложности компьютерной обработки текстов и неоднозначности любого естественного языка).

2.2. Элементы ER – модели

Базовыми элементами инфологических моделей являются:

- 1)сущности;
- 2)связи между сущностями;
- 3)свойства сущностей (атрибуты);
- 4)ключевые атрибуты.

2.2.1. Сущность

Сущность (entity) – это некоторый объект, выделяемый (идентифицируемый) пользователем в предметной области.

Например

СТУДЕНТ,
ПРЕПОДАВАТЕЛЬ,
УЧЕБНИК,
АУДИТОРИЯ,
УЧЕБНЫЕ ЗАНЯТИЯ для группы
и т.п.

Видно, что сущностями могут быть люди, предметы, места, события и т.д.

Необходимо различать такие понятия, как **тип сущности** (класс сущности) и **экземпляр сущности**.

Понятие **тип сущности** относится к набору однородных объектов - личностей, предметов, событий или понятий, выступающих как целое.

Экземпляр сущности относится к конкретной вещи в наборе. Например, типом сущности может быть ГОРОД, а экземпляром – Москва, Киев и т.д.

В реляционных базах данных сущности реализуются с помощью таблиц.

2.2.2. Атрибут

Атрибут – это поименованная характеристика сущности (свойство типа сущности), значимая с точки зрения пользователя.

Его **наименование должно быть уникальным** для конкретного типа сущности, но может быть одинаковым для различного типа сущностей (например, ЦВЕТ может быть определен для многих сущностей: АВТОМОБИЛЬ, ТЕКСТ и т.д.).

В реляционных базах данных атрибуты реализуются с помощью столбцов.

2.2.3. Ключ

Ключ – минимальный набор атрибутов, по значениям которых можно однозначно найти (идентифицировать) требуемый экземпляр сущности. Минимальность означает, что исключение из набора любого атрибута не позволяет идентифицировать сущность по оставшимся.

ПРИМЕР. Для сущности *Расписание* ключом является атрибут **Номер_рейса** или набор: **{Пункт_отправления, Время_вылета и Пункт_назначения}**.

Ключом может быть не любой атрибут сущности. Например, **ДатаНайма** или **Должность** преподавателя вряд ли могут использоваться для идентификации преподавателей.

Другими словами атрибут нельзя назначить ключом сущности. Он либо является таковым, либо не является.

В реляционных базах данных ключи реализуются с помощью первичных ключей.

2.2.4. СВЯЗЬ

Связь – это характеристика отношений между двумя или более сущностями.

В реляционных базах данных связи реализуются с помощью механизма первичных и внешних ключей.

Описание сущностей и их связей – это и есть (с точки зрения проектировщика БД) основная часть модели базы данных.

2.3. Классификация сущностей и связей.

2.3.1. Язык ER-диаграмм. Виды связей

Одной из систем инфологического моделирования является язык ER-диаграмм (от англ. **Entity-Relationship**, т.е. сущность-связь).

В нем сущности изображаются *прямоугольниками*, ассоциации — *ромбами* или *шестиугольниками*, атрибуты — овалами, а связи между ними — ненаправленными ребрами, над которыми может проставляться степень связи (1 или буква, заменяющая слово "много") и необходимое пояснение.

Простые связи

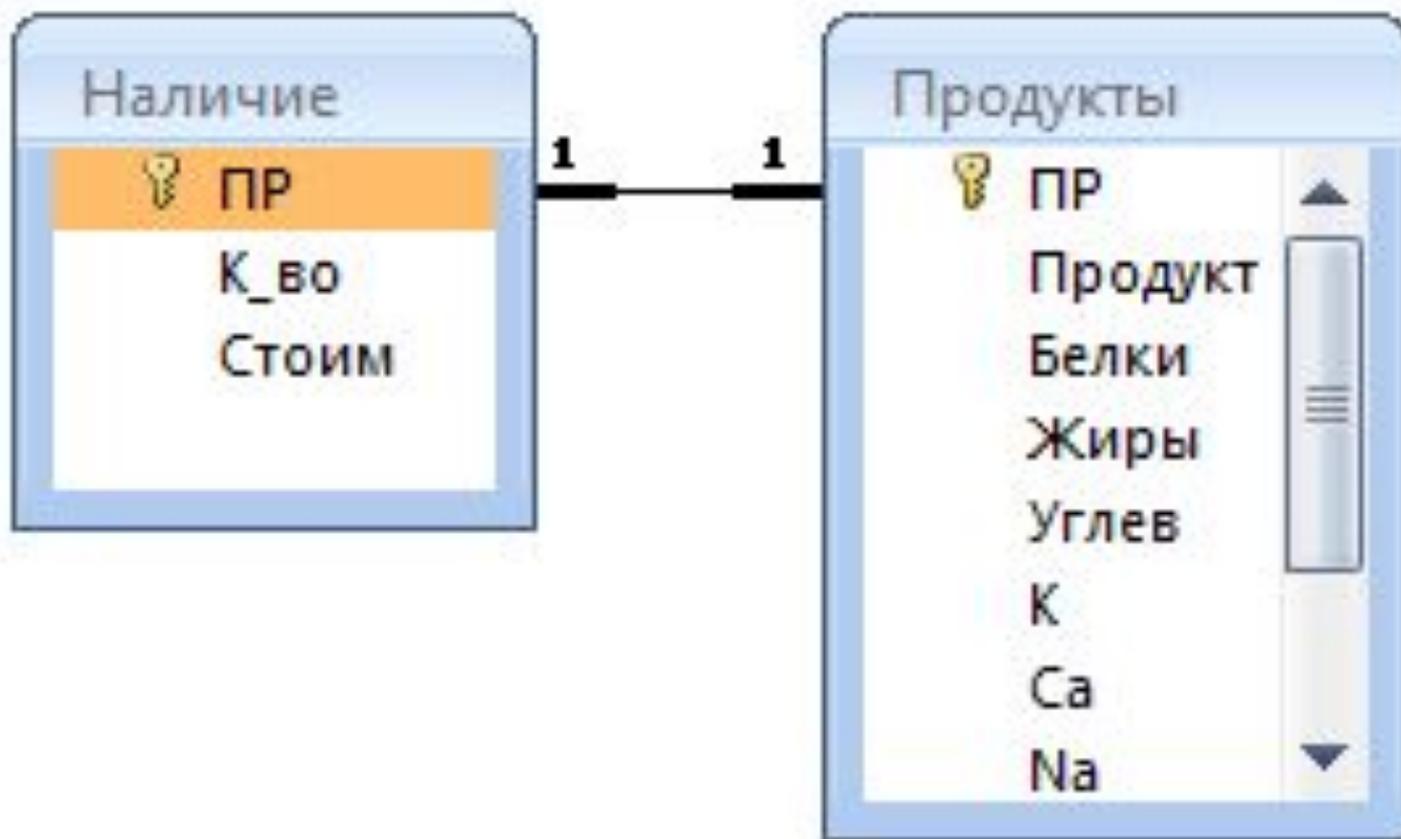
Между двумя сущностями, например, **A** и **B** возможны четыре вида связей.

1. Связь ОДИН-К-ОДНОМУ (1:1).

Каждому представителю (экземпляру) сущности **A** соответствует 1 или 0 представителей сущности **B**:

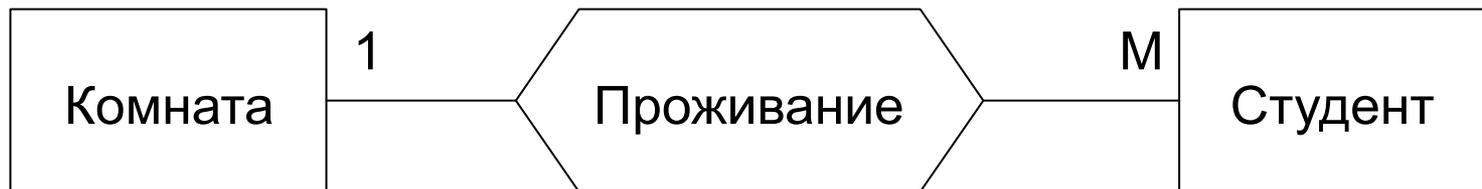


Примечание. В реляционных базах данных связь ОДИН-К-ОДНОМУ, как правило, реализуется с помощью двух таблиц.



2. Связь ОДИН-КО-МНОГИМ (1:М).

Каждому представителю сущности А соответствуют 0, 1 или несколько представителей сущности В.

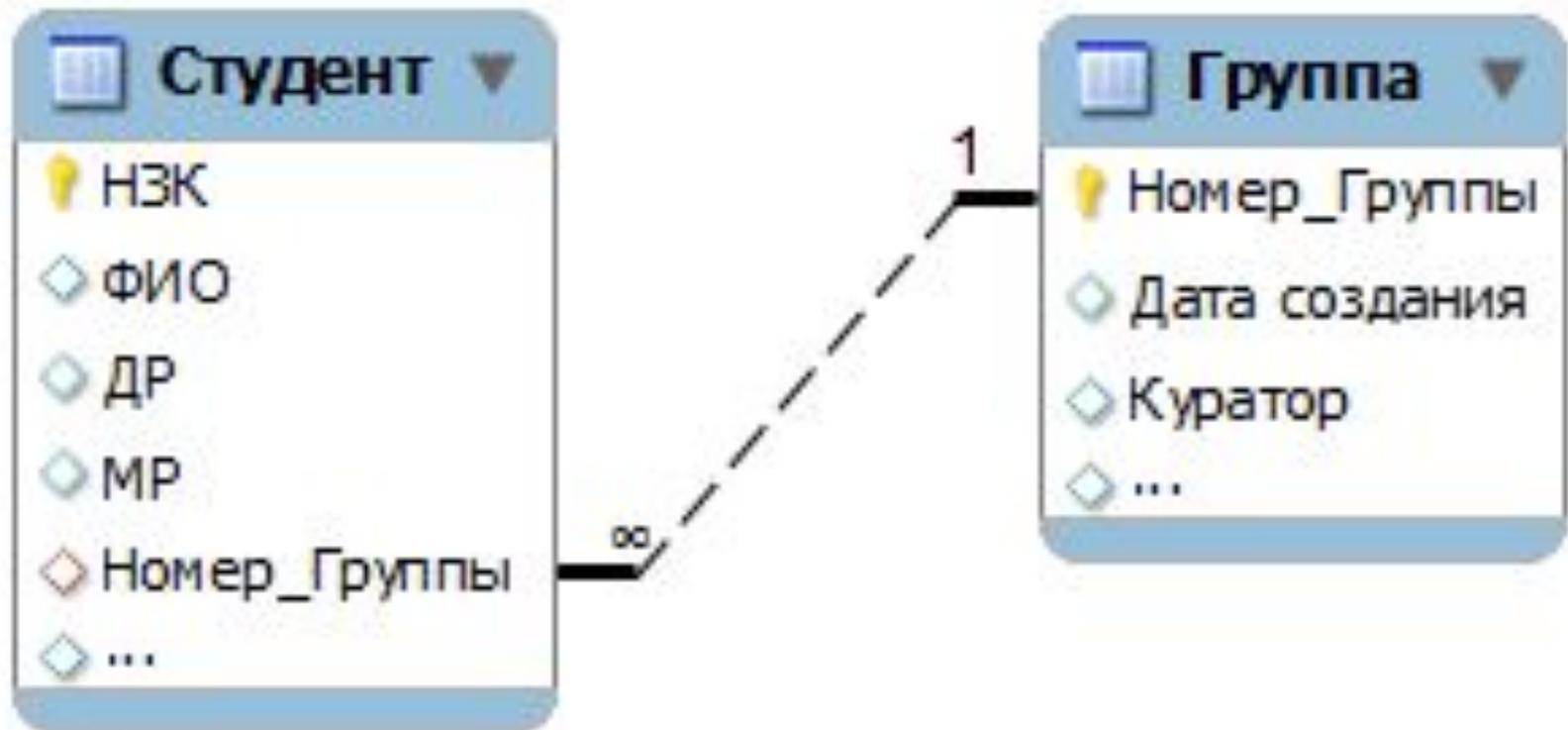


Комната может пустовать, в ней может жить один или несколько студентов.



Каждый студент может быть зачислен ровно в одну учебную группу. Поэтому степень этой связи со стороны сущности СТУДЕНТ равна единице.

Примечание. В реляционных базах данных связь ОДИН-КО-МНОГИМ, как правило, реализуется с помощью двух таблиц.



Так как между двумя сущностями возможны связи в обоих направлениях, то существует еще один тип связи:

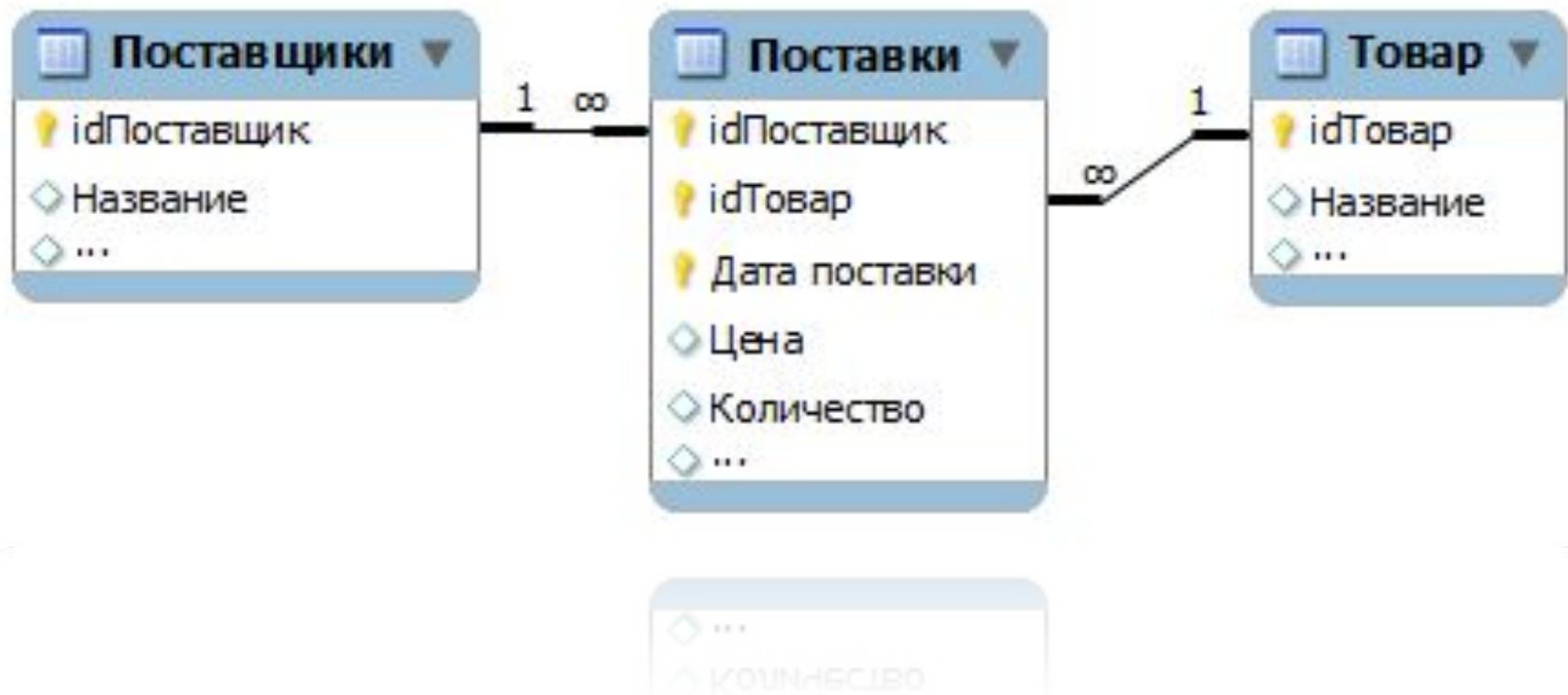
3. Связь МНОГИЕ-К-ОДНОМУ (М:1)

которая полностью подобна связи ОДИН-КО-МНОГИМ.

4. СВЯЗЬ МНОГИЕ-КО-МНОГИМ (M:N).



Примечание. В реляционных базах данных связь МНОГИЕ-КО-МНОГИМ как правило, реализуется с помощью трех таблиц.



Сложные связи

1. Множество связей между одними и теми же сущностями

Пример.



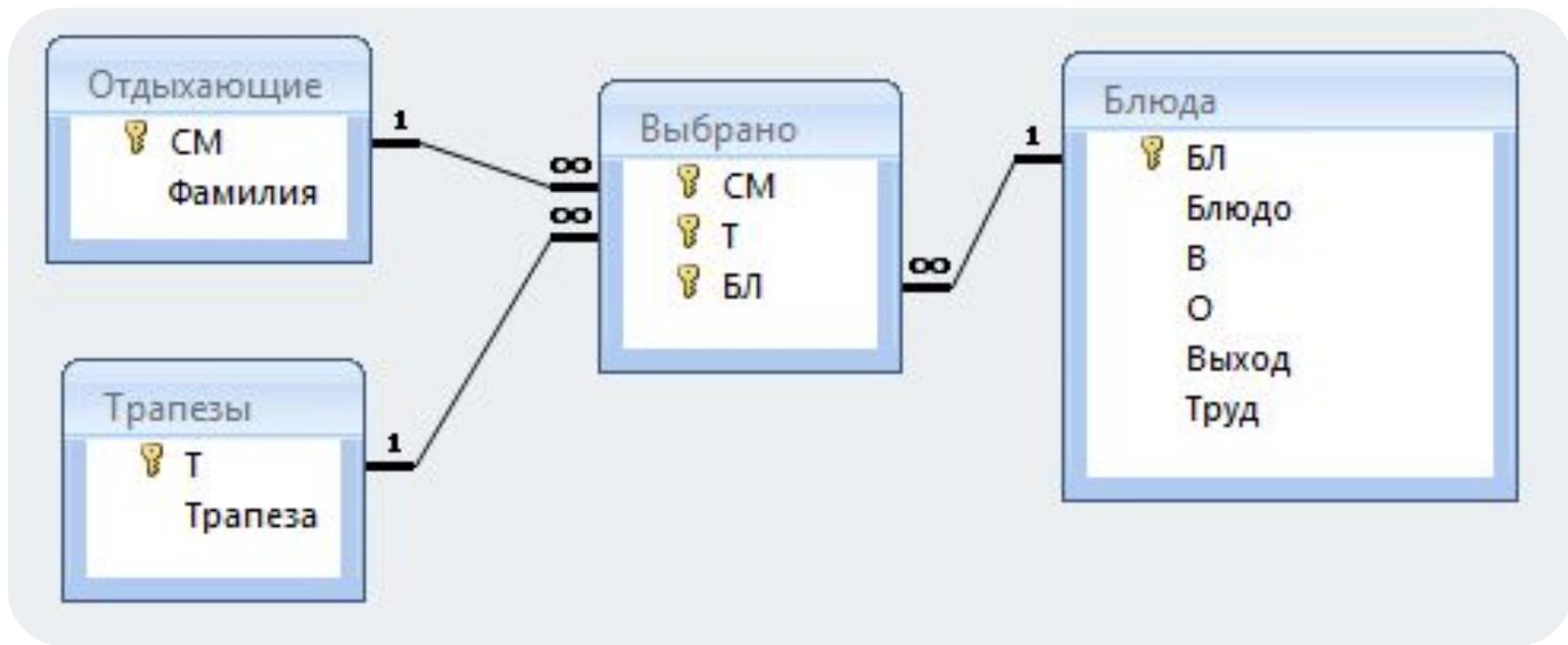
(пациент, имея одного лечащего врача, может иметь также несколько врачей-консультантов; врач может быть лечащим врачом нескольких пациентов и может одновременно консультировать несколько других пациентов).

2. Тренарные связи



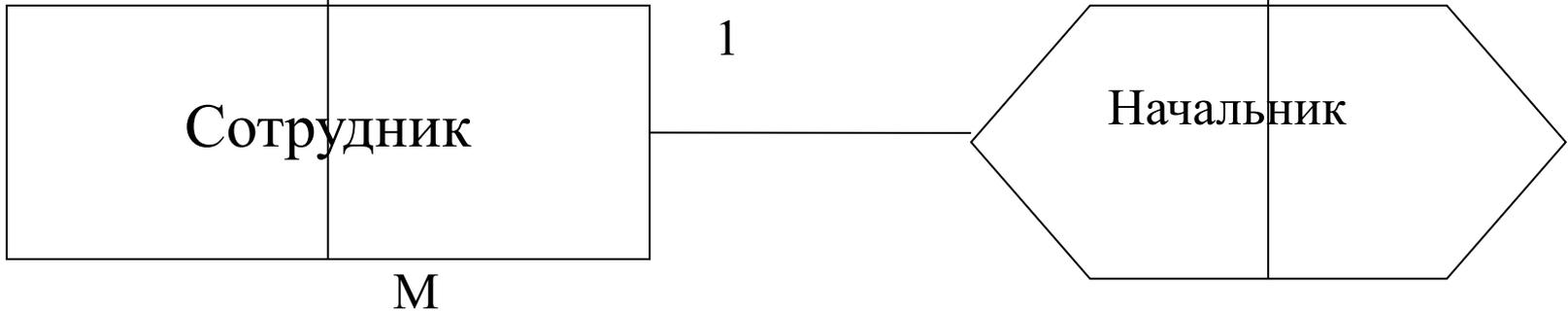
(врач может назначить несколько пациентов на несколько анализов, анализ может быть назначен несколькими врачами нескольким пациентам и пациент может быть назначен на несколько анализов несколькими врачами);

Пример из базы данных Пансион



1b9u639

3. Рекурсивная связь. Сущность связывается с собой.



4. Связи более высоких порядков, смысл которых иногда очень сложен.

2.3.2. Классификация сущностей.

Определены четыре основные класса

сущностей:

1) стержневые,

2) ассоциативные,

3) характеристические,

4) обозначающие.

Язык инфологического моделирования "Таблица-связь".

В нем все сущности изображаются **одностолбцовыми таблицами с заголовками, состоящими из имени сущности и типа сущности (в скобках)**.

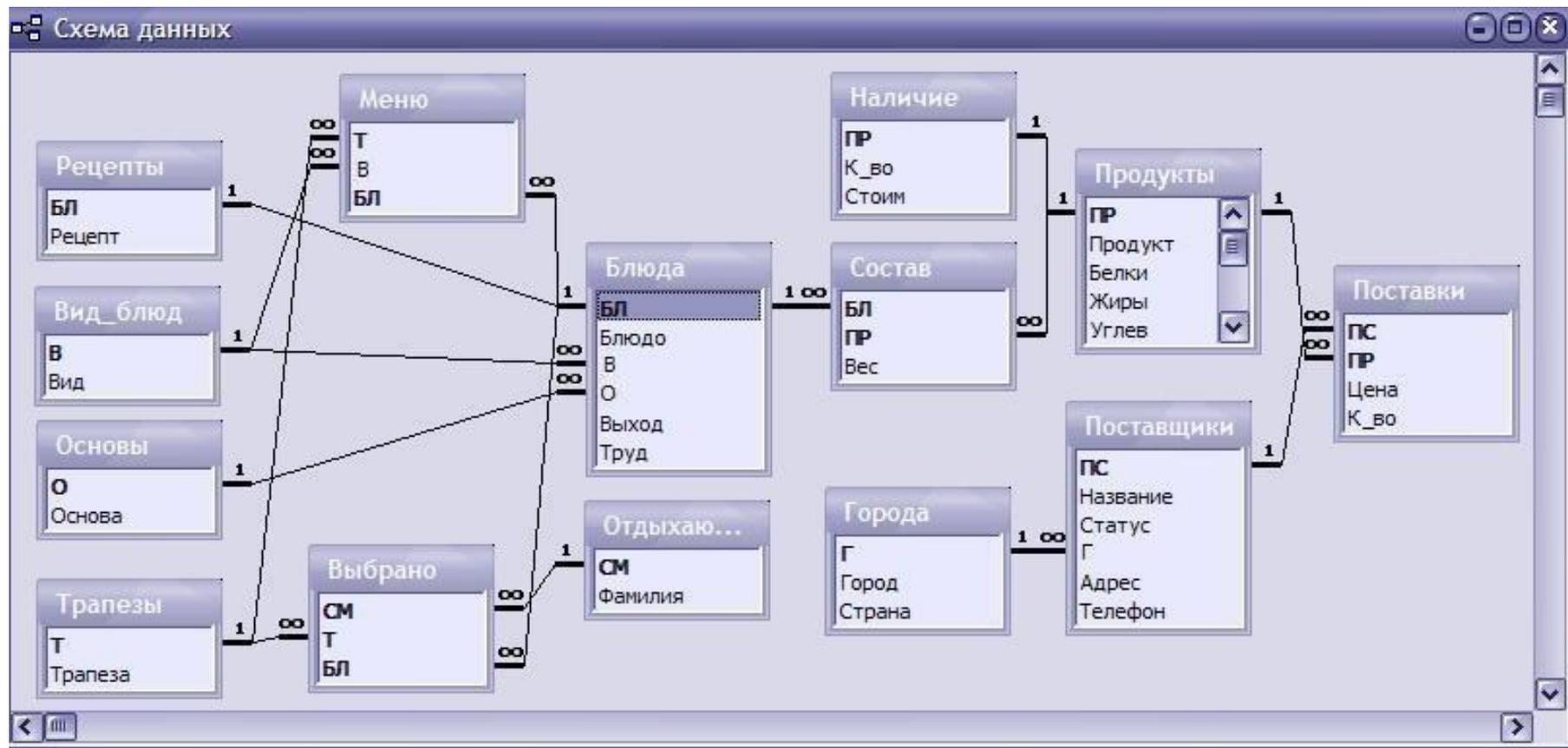
Строки таблицы – это перечень атрибутов сущности, а те из них, которые составляют первичный ключ, располагаются рядом и выделяются.

Связи между сущностями указываются стрелками, направленными от первичных ключей или их составляющих, с указанием типа связи (одна ко многим и т.д.)

ПРИМЕРЫ

Т.е. это схема данных ACCESS, в которой по другому обозначается степень связи и указывается вид сущности.

Схема базы данных ПАНСИОН



1. Стержневая сущность (стержень) – это сущность, не зависящая от других сущностей.

В реляционных базах данных стержневая сущность не содержит внешних ключей.

Язык инфологического моделирования (ЯИМ)

На ЯИМ стержневые сущности представляются предложениями вида:

СУЩНОСТЬ (атрибут_1, атрибут_2 , ..., атрибут_n)

где атрибуты, входящие в ключ, должны быть выделены с помощью подчеркивания или жирным шрифтом.

Пример. Сущности ВРАЧ и ПАЦИЕНТ можно описать на ЯИМ следующим образом:

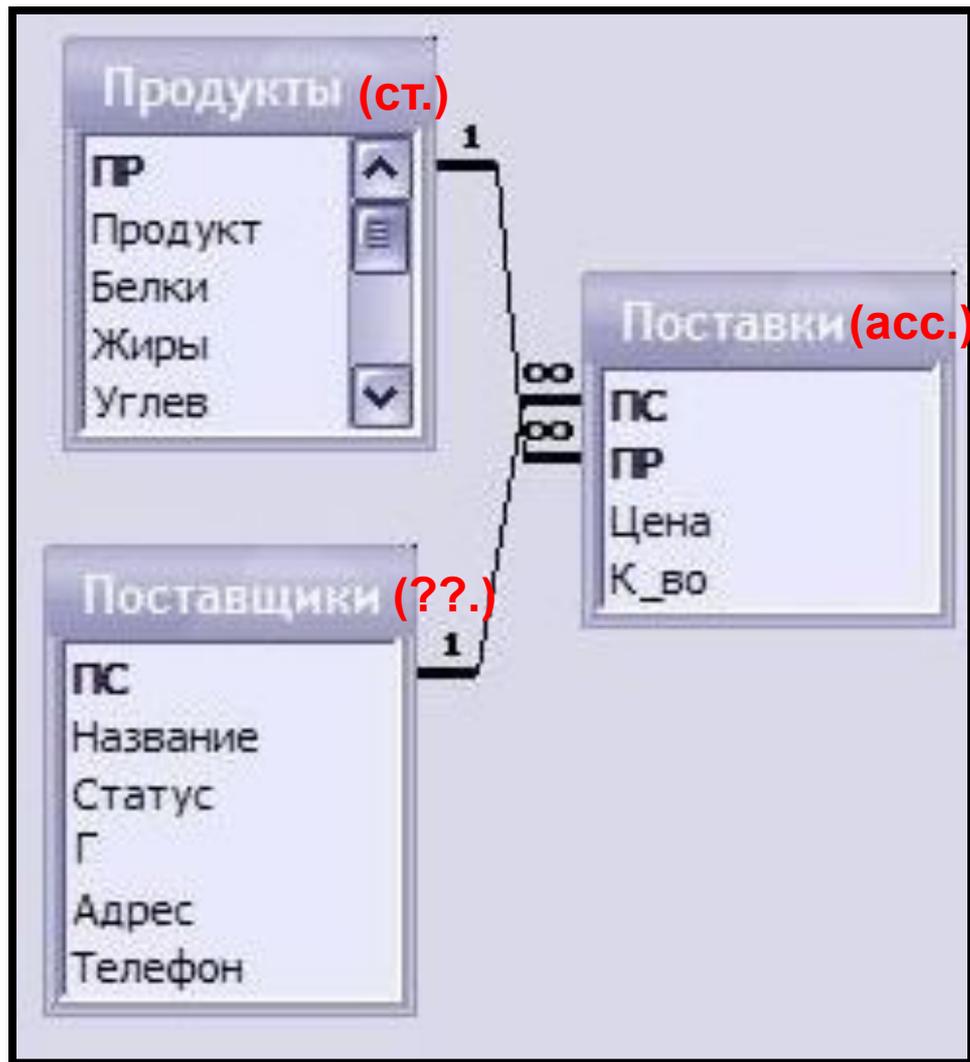
Врач (**Номер_врача**, Фамилия, Имя, Отчество,
Специальность)

Пациент (**Регистрационный_номер**, Фамилия, Номер койки,
Имя, Отчество, Адрес, Дата рождения, Пол)

В базе данных ПАНСИОН стержневые сущности:

- ❖ ВИД_БЛЮД,
- ❖ ОСНОВЫ,
- ❖ ТРАПЕЗЫ,
- ❖ ГОРОДА,
- ❖ ОТДЫХАЮЩИЕ,
- ❖ ПРОДУКТЫ

2. Ассоциативная сущность (ассоциация) – это связь вида "многие-ко-многим" между двумя или более сущностями.



В реляционных базах данных *ассоциативная сущность* имеет составной первичный ключ, в который входит не менее двух внешних ключей.

Язык инфологического моделирования (ЯИМ)

На ЯИМ ассоциация обозначается следующим образом:

АССОЦИАЦИЯ [СУЩНОСТЬ N1, СУЩНОСТЬ N2, ...]
(атрибут_1, атрибут_2, ..., атрибут_n)

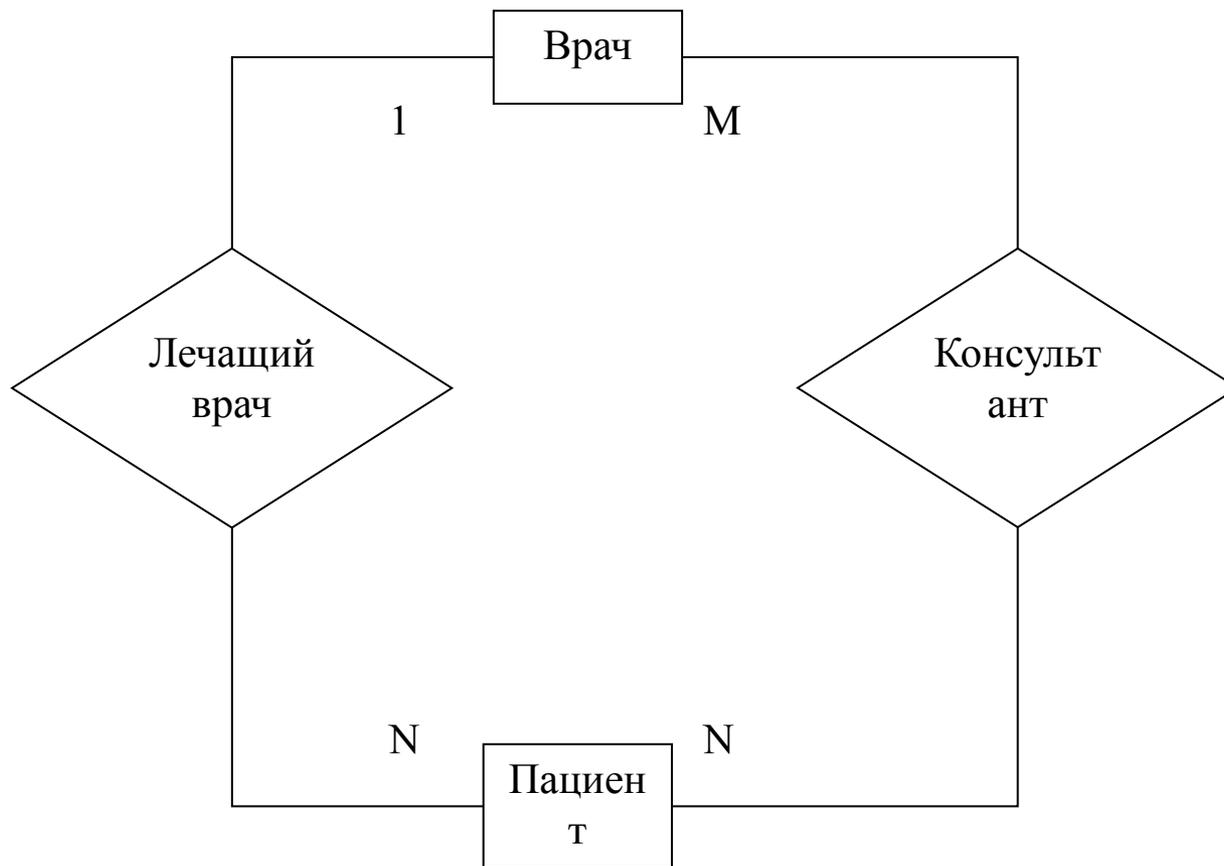
Здесь N – степень связи, а атрибуты, входящие в ключ, должны быть выделены с помощью подчеркивания или жирным шрифтом. Например, *взаимоотношения врачей и пациентов* можно отобразить следующим образом:

Лечащий_врач [Врач 1, Пациент M]
(Номер_врача, Регистрационный_номер)

Консультант [Врач M, Пациент N]
(Номер_врача, Регистрационный_номер).

ER - диаграмма





В базе данных

ПАНСИОН

ассоциации:

МЕНЮ,

ВЫБРАНО,

СОСТАВ,

ПОСТАВКИ.

Ассоциации рассматриваются как полноправные сущности:

- 1)они могут участвовать в других ассоциациях и обозначениях точно так же, как стержневые сущности;
- 2)могут обладать свойствами, т.е. иметь не только набор ключевых атрибутов, необходимых для указания связей, но и любое число других атрибутов, характеризующих связь. Например, ассоциация "Состав" содержит ключевые атрибуты "БЛ", "ПР", а также уточняющий атрибут "Вес".

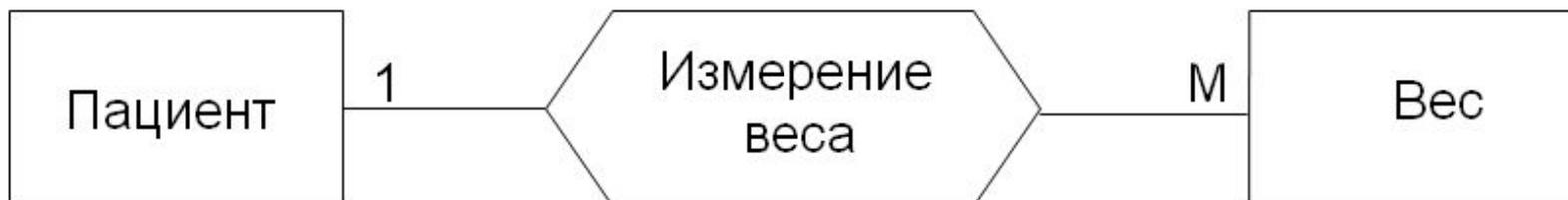
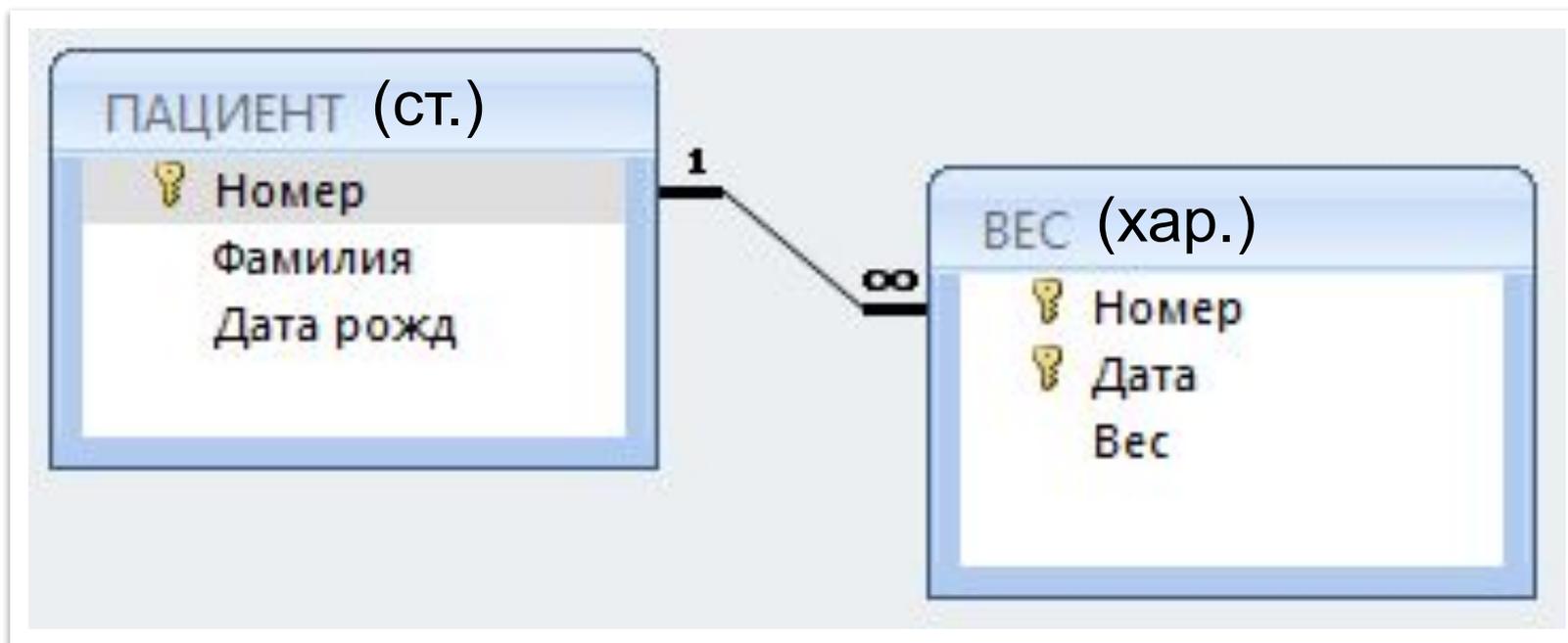
3. Характеристическая сущность (характеристика) –

это связь вида "многие-к-одной" или "одна-к-одной" между двумя сущностями. Единственная цель характеристики состоит в описании или уточнении некоторой другой сущности.

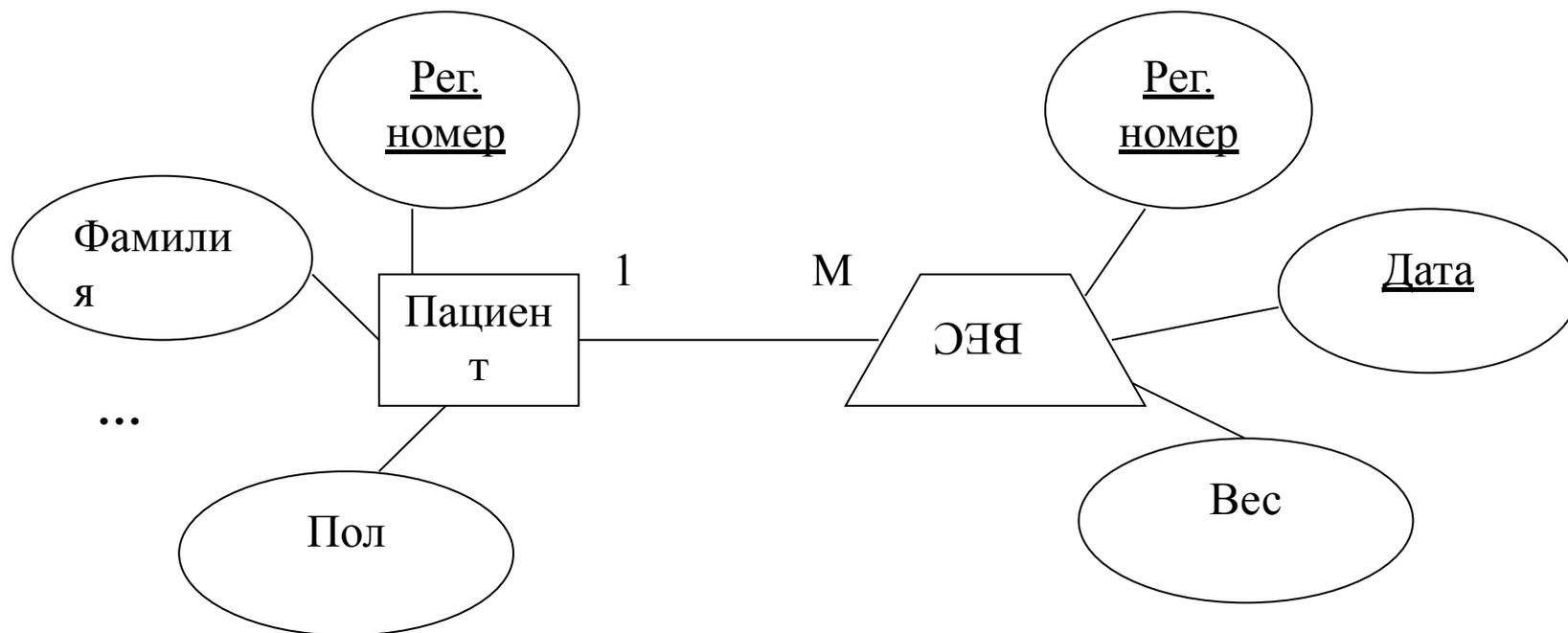
Пример. Предположим, у пациента регулярно измеряется вес. Это можно реализовать введением характеристической сущности **ПАЦИЕНТ**

Фамилия	Дата	Вес	Дата рожд.
Иванов	11.03.08	72	13.05.70
Иванов	12.03.08	73	13.05.70
...
Петров	1.05.08	54	22.01.85
Петров	2.05.08	55	22.01.85

Дублирование можно исключить созданием дополнительной сущности **ВЕС**



Расширенная ER-диаграмма:



Язык инфологического моделирования (ЯИМ)

ХАРАКТЕРИСТИКА (атрибут 1, атрибут 2, . . .) { **ХАРАКТЕРИЗУЕМАЯ СУЩНОСТЬ** }.

Атрибуты, входящие в ключ, должны быть выделены с помощью подчеркивания или жирным шрифтом.

ВЕС (Рег. номер, Дата, Вес) { Пациент }

В реляционных базах данных характеристическая сущность имеет простой или составной первичный ключ, который содержит ровно один внешний ключ.

В базе данных

ПАНСИОН

характеристики:

РЕЦЕПТЫ,

НАЛИЧИЕ.

Характеристика *полностью* *зависит* от
характеризуемой сущности, она предполагают наличие
некоторой другой сущности, которая будет
"характеризоваться".

4. Обозначающая сущность или обозначение – это связь вида "многие-к-одной" или "одна-к-одной" между двумя или более сущностями, которая отличается от характеристики тем, что не зависит от обозначаемой сущности.

Рассмотрим пример, связанный с зачислением сотрудников в различные отделы организации. Здесь возможны 2 варианта:

1. Сотрудник может одновременно зачисляться в несколько отделов или не зачисляться ни в один отдел.

2. Каждый из сотрудников должен быть обязательно зачислен только в один из отделов.

Язык инфологического моделирования (ЯИМ)

ОБОЗНАЧЕНИЕ (атрибут 1, атрибут 2, ...) [СПИСОК ОБОЗНАЧАЕМЫХ СУЩНОСТЕЙ]

Здесь атрибуты, входящие в ключ, должны быть выделены с помощью подчеркивания или жирным шрифтом.

Отделы (**Номер отдела**, Название отдела, ...)

1

Служащие (**Табельный номер**, Фамилия, ...)

Зачисление [Отделы M, Служащие N]

(**Номер отдела**, **Табельный номер**, Дата зачисления).

Отделы (**Номер отдела**, Название отдела, ...)

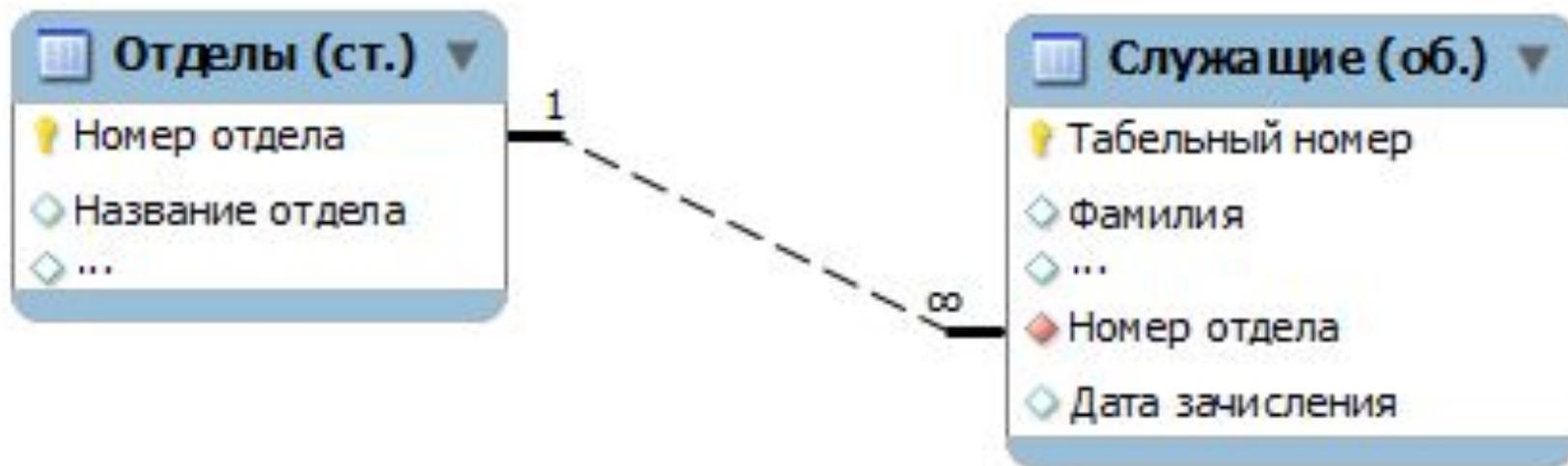
2

Служащие (**Табельный номер**, Фамилия, ... , **Номер отдела**, Дата зачисления)[Отделы]

1. Сотрудник может одновременно зачисляться в несколько отделов или не зачисляться ни в один отдел. В каждом отделе может быть ни одного, один или несколько сотрудников.

Нарисовать схему ТАБЛИЦА-СВЯЗЬ

2. Каждый из сотрудников должен быть обязательно зачислен только в один из отделов.



В реляционных базах данных обозначающая сущность имеет первичный ключ, который не содержит внешний ключ. Среди полей, не входящих в первичный ключ, имеется хотя бы один внешний ключ

В базе данных ПАНСИОН обозначения:
БЛЮДА, ПОСТАВЩИКИ.

Обозначения не являются полностью независимыми сущностями, поскольку они предполагают наличие некоторой другой сущности на которую они ссылаются, которая будет "обозначаться".

Пример

Предметная область – **Географический справочник**

В справочнике должна храниться информация об:

- Континенты.
- Страны.
- Площадь страны.
- Численность населения страны.
- Государственный строй.
- Исторический очерк страны.
- Государственный (е) язык (и).
- Основные национальности с указанием процентного состава.
- Соседние страны.
- Омывающие моря.

В результате анализа предметной области были выделены следующие объекты и их характеристики.

- Континент.
- Страна.
 - Площадь.
 - Численность населения.
 - Исторический очерк
- Государственный строй.
- Государственный язык.
- Национальность.
- Море.

В создаваемой базе данных данным объектам и характеристикам соответствуют сущности, их атрибуты и связи между ними.

