

ГАПОУ "Камский строительный колледж  
имени Е.Н.Батенчука"

**Учебная дисциплина**  
**СИСТЕМЫ УПРАВЛЕНИЯ БАЗЫ ДАННЫХ**

**Лекция 5. Проектирование баз данных  
фактографических автоматизированных  
информационных систем (АИС).  
Нормализация.**

Руководитель занятия:  
преподаватель спецдисциплин  
Еремина И.И.

Проектирование БД осуществляется на основе формализации структуры и процессов предметной области и включает **концептуальное и схемно-структурное** (ориентированное на конкретную СУБД) проектирование.

В группе разработчиков БД выделяют специалистов по

1. формализации предметной области,
2. специалистов по программному обеспечению СУБД,
3. технических дизайнеров
4. специалистов по эргономике.

# Специалисты по формализации предметной области

возглавляют, как правило, весь проект создания АИС и обеспечивают функции взаимодействия с заказчиком



# Специалисты по программному обеспечению СУБД

относятся к категории профессиональных программистов, определяют выбор СУБД и обеспечивают построение ее средствами построения БД по разработанной постановщиком задачи концептуальной схеме.



# Технические дизайнеры СУБД



Database Design – это набор процессов, которые облегчают проектирование, разработку, внедрение и обслуживание корпоративных систем управления данными предприятия. Это помогает создавать системы баз данных:

- соответствующие требованиям пользователей
- высокой производительности.

Правильно спроектированная база данных, помимо улучшения производительности, проста в обслуживании, улучшает согласованность данных и экономически выгодна с точки зрения дискового пространства. Обратите внимание, гениальность базы данных в ее дизайне.

# Специалисты по эргономике СУБД

обеспечивают эстетическую и эргономическую сторону интерфейса с пользователем в АИС при вводе, обработке и поиске данных.





# Концептуальное проектирование

Концептуальное проектирование БД является в значительной степени эвристическим процессом, и адекватность построенной модели предметной области проверяется в большинстве случаев эмпирически по анализу и проверке удовлетворения информационных потребностей пользователей.

В процедуре концептуального проектирования можно выделить следующие этапы:

- ❖ Обзор и изучение области использования АИС для формирования общего представления о предметной области;
- ❖ Формирование и анализ круга функций и задач АИС;
- ❖ Определение основных субъектов-сущностей предметной области и отношений между ними;
- ❖ Формализованное описание предметной области.

# Обзор и изучение области использования АИС для формирования общего представления о предметной области

Происходит только в тесном контакте разработчика и заказчика. При этом изучается **необходимая документация**. На этой основе определяются **основные процессы, участники и информационные потоки в предметной области**. Принципиальным моментом является **фрагментирование** предметной области, т.е. ее **разделение на организационные, технологические, функциональные фрагменты**.



# Фрагментирование предметной области

- руководители организации;
- подразделения организации, их руководители;
- сотрудники, исполняющие документы;
- мероприятия;
- документы, обработка которых или подготовка которых реализует управленческие решения ;
- служба документационного обеспечения управления;
- руководители и работники, ведущие регистрацию, учет, обработку и хранение документов.

# Итоговый результат концептуального проектирования

**Определение основных объектов-сущностей предметной области и отношений между ними.**

Выделение основных информационных объектов начинается с анализа документов и регламентаций (положения, инструкции, бланки, формализованные карточки, отчеты, формы, журналы).

Анализ **“бумажной” документации** позволяет сформировать **перечень атрибутов**, характеризующих те или иные объекты-сущности и отношения между ними. При определении перечня атрибутов руководствуются соображениями **минимальной достаточности**. Перечень самих объектов-сущностей и их атрибутов должен **быть достаточным** для решения всех частных задач системы и удовлетворять информационным потребностям пользователей, но он также **не должен быть избыточным**, чтобы минимизировать расходы.

# Формализованное описание концептуальной схемы БД

- 1. Семантические модели данных**, т.е. происходит преобразование **концептуальной схемы** БД в структуру соответствующей реляционной базы данных. Такую структуру еще называют **дatalogической моделью БД**. В этом случае разработку концептуальной схемы БД называют семантическим моделированием данных.
- 2. Для графического представления** структуры данных используют **ER-модели**. Основными компонентами семантической модели по Чену (*сущности, наборы сущностей, атрибуты сущностей, наборы значений атрибутов, ключевые атрибуты сущностей, связи, виды связей, атрибуты связей, наборы связей*).

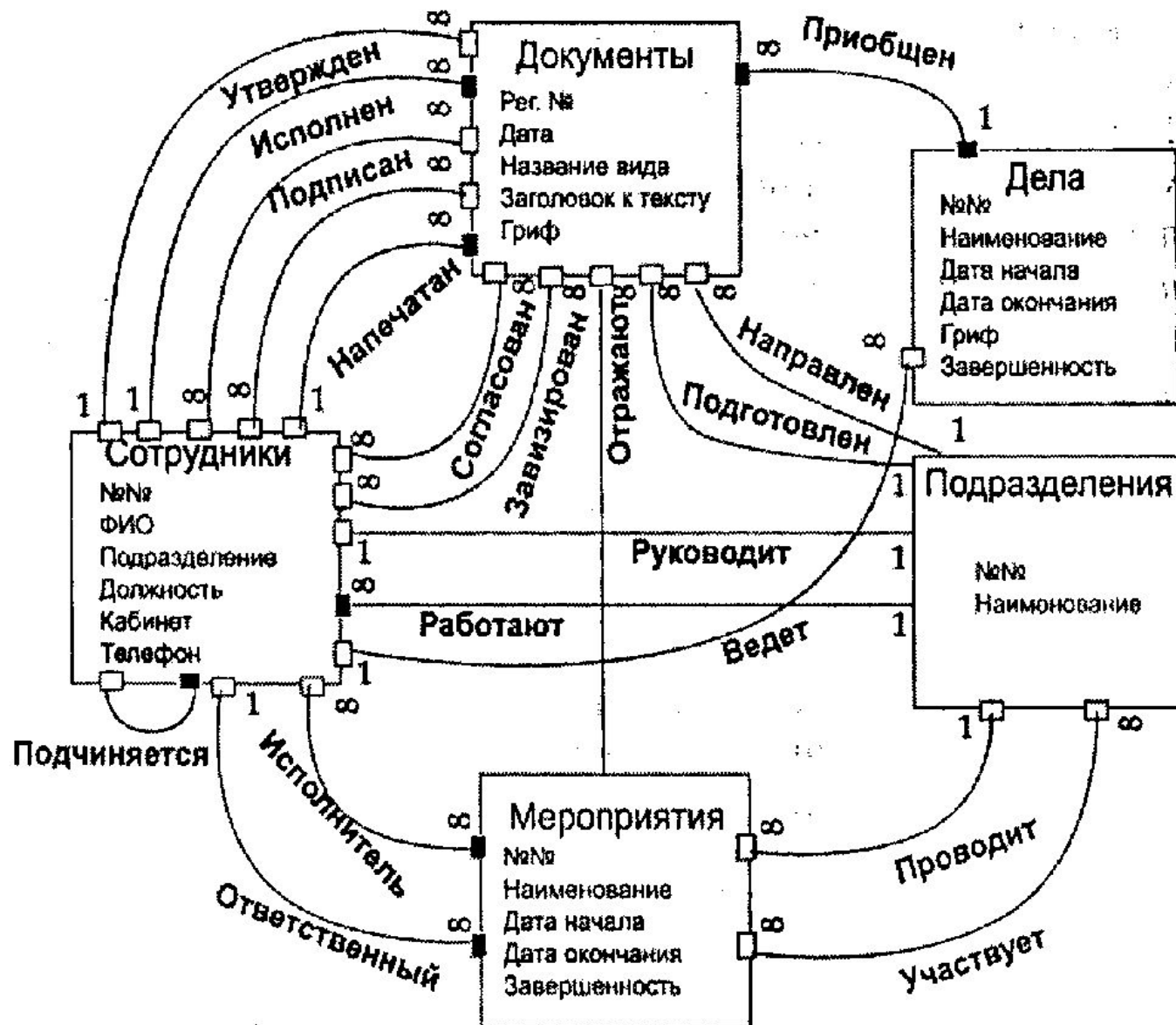
# Формализованной описание концептуальной схемы БД

- осуществляется на бумаге и служит *основой эскизного проекта* создания БД.

# Построение схемы БД

Создается структура БД, соответствующая концептуальной схеме.

**Адекватность реализации концептуальной схемы определяется эмпирически и эвристически в ходе отладки и дальнейшей эксплуатации БД.**



# Проектирование схем реляционных баз данных

## Поддержка целостности данных

Целостность данных необходима для поддержки внутреннего логического соответствия данных, хранящихся в базе. Обычно СУБД требуют, чтобы целостности данных поддерживалась самим приложением.

Например, каждое приложение после выполнения какой-либо операции с данными проверяет, а не удалил ли пользователь из справочника организацию, которая имеет сотни заказов от вашей фирмы.

Часто бывает, что несколько десятков приложений используют одну и ту же базу данных. Например, предприятие ведет БД, которая хранится на файл-сервере. БД состоит из сотен таблиц.



# Проектирование схем реляционных БД

## продолжение

### Три типа целостности данных

- **Целостность сущности (или таблицы)** требует, чтобы все записи в таблице имели уникальный идентификатор – **первичный ключ** (связи между первичным и внешним ключами).
- **Целостность столбцов** достигается ограничениями на данные, которые могут быть занесены в столбец. Целостность столбцов часто обеспечивается правилами проверки ввода, заданием типов данных и указанием списка возможных значений
- **Ссылочная целостность обеспечивает связь между первичным** (в родительской таблице) и **внешним** (в дочерней таблице) ключом. Запись в родительской таблице не может быть удалена, а первичный ключ изменен, если в дочерней таблице существует связанная запись. Для дочерней таблицы нельзя ввести внешний ключ, отсутствующий среди первичных ключей родительской таблицы.

# Этапы проектирование

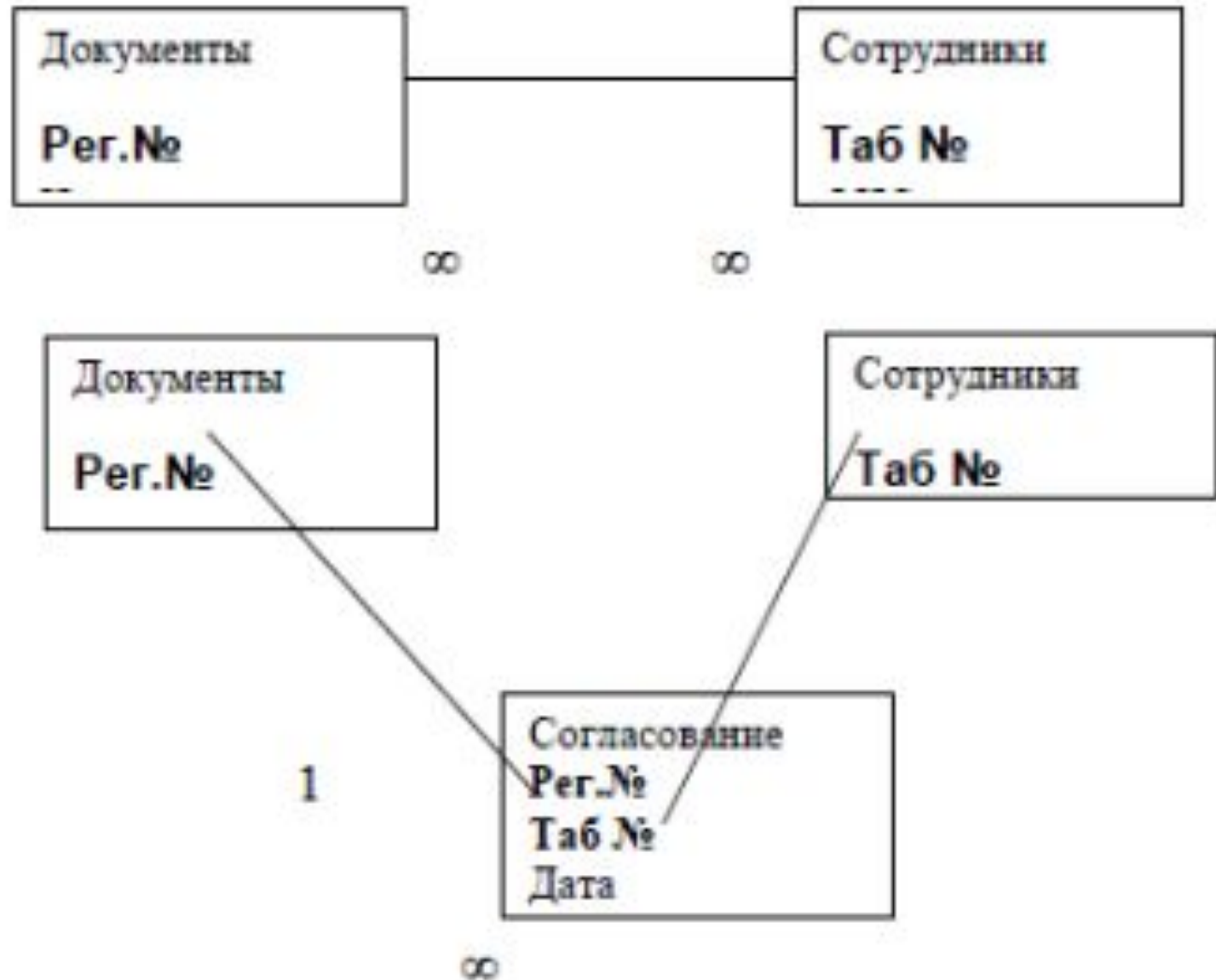
1. определение перечня таблиц и связей;
2. определение перечня полей, типов полей, ключевых полей каждой таблицы, установление связей между таблицами через внешние ключи;
3. определение или установление индексов для полей в таблицах;
4. разработка списков (словарей) для полей с перечислительным характером значений данных;
5. установление ограничений целостности по полям таблиц и связей;
6. нормализация таблиц, доработка перечня таблиц и их связей.

# Проектирование и создание таблиц

Поля таблиц определяются на основе первоначально отработанных атрибутов информационных объектов концептуальной схемы БД.

Установление ограничений целостности по полям и связям обеспечивается:

- назначением первичного ключа (уникальность записи);
- назначением индексированного поля в режиме “совпадения не допускаются”;
- установления свойства “обязательное поле”;
- установления допустимого диапазона значений и др. свойств.



# Нормализация таблиц

Нормализация представляет процесс разделения данных по отдельным связанным таблицам. Нормализация устраняет избыточность данных (data redundancy) и тем самым избежать нарушения целостности данных при их изменении, то есть избежать аномалий изменения (update anomaly).

**Нормализация** — это процесс организации данных в базе данных, включающий создание таблиц и установление отношений между ними в соответствии с правилами, которые обеспечивают защиту данных и делают базу данных более гибкой, устраняя избыточность и несогласованные зависимости.

# Нормализация таблиц

В ненормализованной форме таблица может хранить информацию о двух и более сущностях. Также она может содержать повторяющиеся столбцы. Также столбцы могут хранить повторяющиеся значения. В нормализованной же форме каждая таблица хранит информацию только об одной сущности.

Нормализация предполагает применение нормальных форм к структуре данных. **Существуют 7 нормальных форм.** Каждая нормальная форма (за исключением первой) подразумевает, что к данным уже была применена предыдущая нормальная форма. Например, прежде чем применить третью нормальную форму к данным должна быть применена вторая нормальная форма. И строго говоря, база данных считается нормализованной, если к ней применяется третья нормальная форма и выше.

# Нормальные формы таблиц

- 1. Первая нормальная форма (1NF)** предполагает, что сохраняемые данные на пересечении строк и столбцов должны представлять скалярное значение, а таблицы не должны содержать повторяющихся строк.
- 2. Вторая нормальная форма (2NF)** предполагает, что каждый столбец, не являющийся ключом, должен зависеть от первичного ключа.
- 3. Третья нормальная форма (3NF)** предполагает, что каждый столбец, не являющийся ключом, должен зависеть только от первичного ключа.

**Нормальная форма Бойса-Кодда (BCNF)** является немного более строгой версией третьей нормальной формы.

- 4. Четвертая нормальная форма (4NF)** применяется для устранения многозначных зависимостей (multivalued dependencies) - таких зависимостей, где столбец с первичным ключом имеет связь



# Нормальные формы таблиц

- 5. Пятая нормальная форма (5NF)** разделяет таблицы на более малые таблицы для устранения избыточности данных. Разбиение идет до тех пор, пока нельзя будет воссоздать оригинальную таблицу путем объединения малых таблиц.
- 6. Шестая нормальная форма (domain key normal form / 6NF).** Каждое ограничение в связях между таблицами должно зависеть только от ограничений ключа и ограничений домена, где домен представляет набор допустимых значений для столбца. Эта форма предотвращает добавление недопустимых данных путем установки ограничения на уровне отношений между таблицами, но не на уровне таблиц или столбцов. Данная форма, как правило, не применима на уровне СУБД, в том числе и в SQL Server.

Наиболее простой нормальной формой является первая, суть которой определяется требованием неделимости полей и единственности значений по полям. Приведен пример ненормализованной таблицы “Сотрудники”, имеющей составное (делимое) поле “Мероприятия” с множественными значениями по полям “Условное наименование” и “Награда”.

Таб. номер	Фамилия	Звание	Мероприятия, в которых участвовал сотрудник		Кабинет	Служ. телефон
			Условное наименование	Награда		
001	Пронин	Майор	Операция Ы	Премия	110	112233
			Бриллиантовая рука	Отпуск		
002	Исаев	Полковник	Операция «Берлин»	-	110	112233
007	Бонд	Капитан	Операция «Золотой глаз»	Ягуар	С-110	332211
			Операция «Багамы»	Феррари		

Приведение таких таблиц к первой нормальной форме осуществляется путем образования составных ключей, при которых устраняются ситуации с множественными значениями полей. Показана такая таблица.

<i>Таб. номер</i>	<i>Условное наименование</i>	Награда	Фамилия	Звание	Кабинет	Служ. телефон
001	Операция Ы	Премия	Пронин	Майор	110	112233
001	Бриллиантовая рука	Отпуск	Пронин	Майор	110	112233
002	Операция «Берн»	-	Исаев	Полковник	110	112233
007	Операция «Золотой глаз»	Ягуар	Бонд	Капитан	С-110	332211
007	Операция «Багамы»	Феррари	Бонд	Капитан	С-110	332211

Таблица в первой нормальной форме **содержит многочисленные ситуации дублирования данных**. Кроме того, например, удаляя запись о сотруднике, участвующем в какой-либо операции, можно удалить информацию о том, что он вообще работает. При переводе сотрудника в другое подразделение или другой кабинет придется изменять все записи – кортежи с данным сотрудником по различным операциям.

Поэтому **Коддом** был разработан специальный механизм разбиения таблиц для приведения к более совершенным нормальным формам. Этот **механизм основан на понятии функциональной зависимости полей-атрибутов**.

Поле-атрибут  $Y$  функционально зависит от поля-атрибута  $X$ , если любому значению  $X$  всегда соответствует только одно значение  $Y$ . К примеру, ФИО всегда соответствует одному значению “Таб №”. В таблице, находящейся в первой нормальной форме, все неключевые атрибуты функционально зависят от ключа таблицы.

Для перевода таблицы из первой нормальной формы во вторую необходимо:

- образовать проекцию (вертикальное подмножество) исходной таблицы на составной ключ и на поля, находящиеся в полной функциональной зависимости от составного ключа;
- построить еще одну или несколько проекций на часть составного ключа с полями, функционально зависимыми только от этой части ключа.

<i>Таб. номер</i>	<i>Условное наименование</i>	Награда
001	Операция Ы	Премия
001	Бриллиантовая рука	Отпуск
002	Операция «Берн»	-
007	Операция «Золотой глаз»	Ягуар
007	Операция «Багамы»	Феррари

Таб. номер	Фамилия	Звание	Кабинет	Служ. телефон
001	Пронин	Майор	110	112233
001	Пронин	Майор	110	112233
002	Исаев	Полковник	110	112233
007	Бонд	Капитан	С-110	332211
007	Бонд	Капитан	С-110	332211

В таблицах во второй нормальной форме также могут присутствовать аномалии. Например, происходит дублирование информации о служебном телефоне “112233”, т.к. атрибут “Сл. Телефон” фактически зависит от атрибута “кабинет”, а не от атрибута “Таб.№”.

В таких случаях говорят о транзитивной зависимости атрибута “Сл. Телефон” от “Таб.№”.

Таблица в третьей нормальной форме характеризуется независимостью неключевых атрибутов и их полная функциональная зависимость от первичного ключа.

Для преобразования из второй в третью нормальную форму таблицу разделяют на две или более проекций так, чтобы конечные поля-атрибуты в цепочках транзитивной зависимости вынести в отдельные таблицы, связав разделившиеся части таблицы внешними ключами по полям-атрибутам, находящимся внутри цепочек транзитивной зависимости.



Таб. номер	Фамилия	Звание	Кабинет
001	Пронин	Майор	110
001	Пронин	Майор	110
002	Исаев	Полковник	110
007	Бонд	Капитан	С-110
007	Бонд	Капитан	С-110

Кабинет	Служ. телефон
110	112233
С-110	332211

# **Тесты для самоконтроля**

# 1. Модель данных - это

- a) взаимосвязанные данные;
- b) совокупность структур данных и операций их обработки;
- c) поименованная совокупность данных, отображающая состояние объектов и их отношений в рассматриваемой предметной области;
- d) совокупность языковых и программных средств, предназначенных для создания, введения и совместного ИС-пользования БД многими пользователями

## **2. Предметная область информационной системы рассматривается как**

- a) совокупность абстрактных объектов
- b) совокупность реальных объектов
- c) совокупность некоторых объектов
- d) совокупность всех объектов

### **3. Информационный объект образуется совокупностью**

- a) логически взаимосвязанных сущностей
- b) логически взаимосвязанных атрибутов
- c) логически невзаимосвязанных атрибутов
- d) логически невзаимосвязанных сущностей

## 4. Кортеж - это

- a) заполненные строки
- b) заполненные таблицы
- c) количество строк в таблице
- d) заполненные столбцы

## 5. Какие типы отношений существуют?

- a) первичные, базовые, результативные
- b) первичные, основные, производные
- c) базовые, производные, промежуточный результат
- d) производные, основные, результативные

## 6. К специальным реляционным операциям относят

- a) объединение, пересечение, разность и произведение
- b) деление, выборка, проекция и соединение
- c) расширение, подведение итогов и обновление
- d) обновление, выборка и проекция



**7. Операция, результатом которой является отношение, содержащее кортежи общие для сравниваемых отношений - это**

- a) Объединение
- b) Разность
- c) Произведение
- d) Пересечение

**8. Результатом этой операции является вставка нового кортежа, изменение или удаление кортежа**

- a) подведение итогов
- b) Расширение
- c) Обновление
- d) Выборка

**9. Дано отношение Преподаватели (КодПреподавателя, Фамилия, Имя, Отчество, ДатаПринятияНаРаботу, Телефон) после выполнения некоторой реляционной операции получили отношение, содержащее преподавателей у которых нет телефона. Какая операция реляционной алгебры была выполнена?**

- a) Проекция
- b) Выборка
- c) Подведение итогов
- d) Обновление

# 10. Нормализация - это

- a) объединение таблиц
- b) установление связей
- c) удаление таблиц
- d) разбиение таблиц