

Базы данных

Шарапова Елена Витальевна
ст.преп. каф. ОРТ, к.т.н.

План курса

- 9 лекционных занятий
- 3 практических занятия
- 3 лабораторные работы
- консультации
- зачет

Литература

Дейт, К., Дж.

Введение в системы баз данных, 7-е издание.:
Пер. с англ. – М. Издательский дом
«Вильямс», 2001. – 1072 с.

ББК 32.973.26-018.2.75

ISBN 5-8459-0138-3 (рус.)

Литература

Глушаков С.В., Ломотько Д.В. Базы данных. – Харьков: Фолио, 2002. – 504 с.

Веллинг Л., Томсон Л. MySQL. Учебное пособие: М.: Изд. Дом «Вильямс», 2005. – 304 с.

Р. Стивенс. Программирование баз данных. Изд. 2-е, стереотипное. Пер. с англ. – М.: ООО «Бином-Пресс», 2007 г. – 384 с.

Методичні вказівки до лабораторних робіт з курсу "Бази даних" / Упоряд. А.І.Лучанінов – Харків: ХТУРЕ, 2002. – 40с. (електронна копія).

Лекция 1

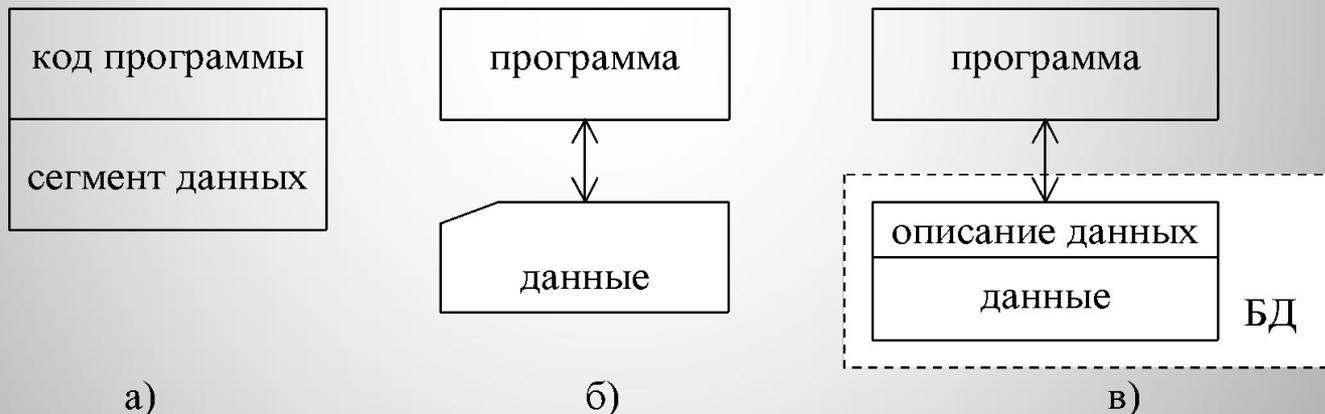
Основные сведения

Предпосылки появления баз данных

Две основные предпосылки появления баз данных:

- Необходимость хранить и обрабатывать большое количество данных.
- Разработка методов совместного использования данных.

Эволюция технологий обработки данных:



Массив данных общего пользования в системах, основанных на данных, называется *базой данных*. База данных (БД) является моделью предметной области информационной системы.

Основные термины

Информация – любые сведения о каком-либо событии, объекте или процессе, являющиеся объектом некоторых операций: восприятия, передачи, преобразования, хранения или использования.

Данные – это информация, зафиксированная в некоторой форме, пригодной для последующей обработки, передачи и хранения, например, находящаяся в памяти ЭВМ или подготовленная для ввода в ЭВМ.

Обработка данных – это совокупность задач, осуществляющих преобразование массивов данных. Обработка данных включает в себя ввод данных в ЭВМ, отбор данных по каким-либо критериям, преобразование структуры данных, перемещение данных на внешней памяти ЭВМ, вывод данных, являющихся результатом решения задач, в табличном или в каком-либо ином удобном для пользователя виде.

Система обработки данных (СОД) – это набор аппаратных и программных средств, осуществляющих выполнение задач по управлению данными.

Управление данными – совокупность функций обеспечения требуемого представления данных, их накопления и хранения, обновления, удаления, поиска по заданному критерию и выдачи данных.

Основные термины

База данных (БД) – это совокупность взаимосвязанных структурированных данных, относящихся к определенной предметной области и организованных так, чтобы обеспечить независимость данных от программ обработки. Фактически база данных – это модель предметной области (ПО).

Предметная область – часть реального мира, подлежащая изучению с целью организации управления и, в конечном итоге, автоматизации.

Ведение базы данных – деятельность по обновлению, восстановлению и изменению структуры базы данных с целью обеспечения её целостности, сохранности и эффективности использования [ГОСТ 20886-85].

Система управления базами данных (СУБД) – это совокупность программ и языковых средств, предназначенных для управления данными в базе данных, ведения базы данных и обеспечения взаимодействия её с прикладными программами [ГОСТ 20886-85].

Автоматизированная информационная система (АИС) представляет собой совокупность данных, экономико-математических методов и моделей, технических, программных средств и специалистов, предназначенную для обработки информации и принятия управленческих решений.

Банк данных (БнД)

Основные задачи *банка данных*:

- Обеспечивать информационные потребности внешних пользователей.
- Обеспечивать возможность хранения данных.
- Обеспечивать заданный уровень достоверности хранимых данных.
- Обеспечивать ограниченный доступ к данным.
- Обеспечивать поиск данных по произвольной группе признаков.
- Удовлетворять заданным требованиям по производительности при обработке запросов.
- Иметь возможность реорганизации при изменении границ ПО.
- Обеспечивать выдачу пользователям данных в различной форме.
- Обеспечивать простоту и удобство обращения внешних пользователей к данным.

Автоматизированные информационные системы (АИС)

По сферам применения и правилам организации различают **информационно-поисковые (ИПС)** и **системы обработки данных (СОД)**.

В зависимости от характера информационных ресурсов АИС подразделяют на

Документальные системы
фактографические системы.

Основные компоненты документальной ИПС

· Программные средства

Поисковый массив документов

Средства поддержки информационного языка системы

Библиотеки 1.01 - [Поисковая форма]

Файл Правка Вид Вставка Формат Записи Сервис Окно Справка

Округ: Центральный | Номер: | Название: *Горького* | Класс: | Тип: | Система: | Ведение: | Найти | Очистить

	Id	Номер	Название	Тип
▶	259	43	им. Горького А.М.	
	258	43	им. Горького А.М.	
	248	23	им. Горького А.М.	Детская
*	етчик)			

Запись: 1 из 3

Компоненты системы баз данных



Основным принципом организации базы данных является совместное хранение данных и их описаний. Это отличает базу данных от любого другого набора данных, хранящихся в ЭВМ.

Описание базы данных хранится в так называемом **словаре-справочнике данных (ССД) или каталоге данных**. Хранение данных вместе с их описанием и позволяет обеспечивать независимость данных от программ, т.е. интерпретация данных определяется не программой, а описанием данных.

Уровни представления данных



Концептуальный уровень: поддерживает единый взгляд на базу данных, общий для всех её приложений и независимый от них и от среды хранения.

Внутренний уровень: схема хранения данных в среде хранения.

Внешний уровень (внешние схемы): предназначены для групп пользователей.

Схема базы данных – это описание базы данных в терминах конкретной модели данных.

Физическая и логическая независимость данных.

Предметная область. Сущности и атрибуты

Предметная область (ПО) информационной системы рассматривается как совокупность реальных процессов и объектов (**сущностей**), представляющих интерес для её пользователей.

Сущности: базовые и зависимые. Тип сущности. Экземпляр сущности. Для каждого типа сущности необходимо определить имя.

Атрибуты: характеристики сущностей. Атрибуты бывают:

1. *Идентифицирующие и описательные атрибуты.*
2. *Составные и простые атрибуты.*
3. *Однозначные и многозначные атрибуты.*
4. *Основные и производные атрибуты.*
5. *Обязательные и необязательные.*

Для каждого атрибута необходимо определить название, указать тип данных и описать ограничения целостности – множество значений, которые может принимать данный атрибут.

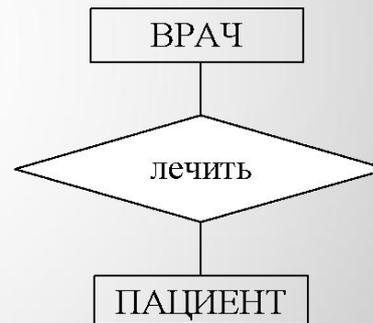
Предметная область. Связи

Связь – это осмысленная ассоциация между сущностями. Для связи указывается название, вид (факультативная или обязательная), степень (1:1, 1:n или m:n) и кардинальность (унарная, бинарная, тернарная или n-арная).

Тип связи. Экземпляр связи.



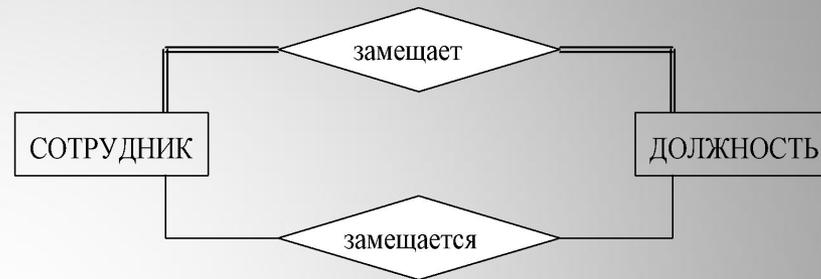
а) унарная связь



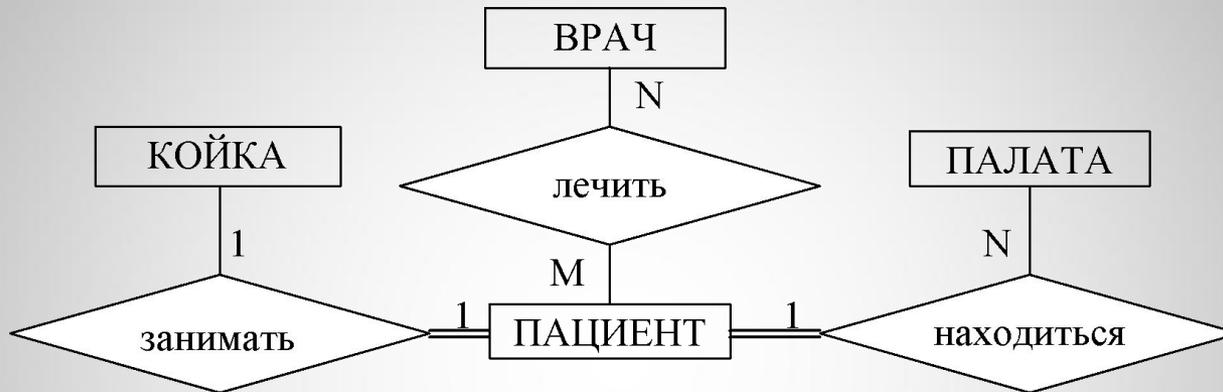
б) бинарная связь



в) тернарная связь



Предметная область. Связи



Связи, приведённые на рисунке, с учётом семантики означают следующее:

- ✓ пациент–койка (1:1);
- ✓ палата–пациент (1:n);
- ✓ пациент–врач (n:m).

Актуализация данных в БД

Каждому моменту времени можно сопоставить некоторое **состояние предметной области**.

Ограничения целостности.

Для того чтобы обеспечить соответствие базы данных текущему состоянию предметной области, база данных *динамически обновляется* (периодически или в режиме реального времени). Это обновление называется **актуализацией данных**.

Актуализация может проводиться:

- вручную,;
- автоматизировано;
- автоматически.

Контроль правильности вносимых изменений.

ОСНОВНЫЕ МОДЕЛИ ДАнных

Модель данных является инструментом
моделирования произвольной
предметной области.

Понятие модели данных

Модель данных – это совокупность правил порождения структур данных в базе данных, операций над ними, а также ограничений целостности, определяющих допустимые связи и значения данных, последовательность их изменения.

Модель данных состоит из трёх частей:

1. **Набор типов структур данных.**
2. **Набор операторов или правил вывода.**
3. **Набор общих правил целостности.**

Типы структур данных.



Композиция структур данных

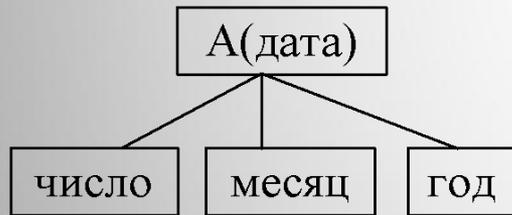
Элемент данных – наименьшая поименованная единица данных, к которой СУБД может обращаться непосредственно и с помощью которой выполняется построение всех остальных структур.

Для каждого элемента данных должны быть определены название и тип.

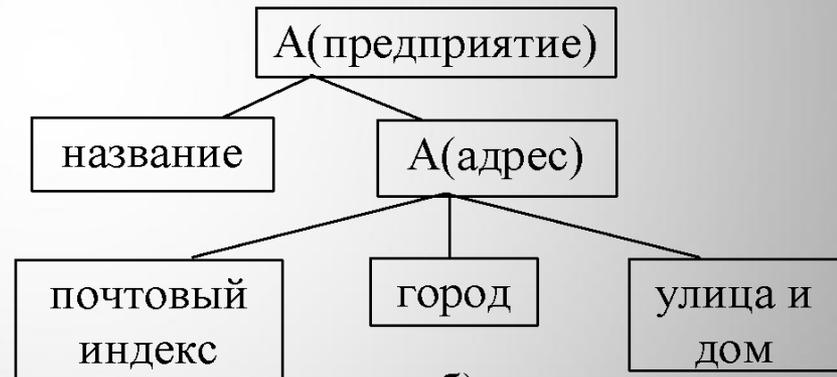
Агрегаты

Агрегат данных – поименованная совокупность элементов данных внутри записи, которую можно рассматривать как единое целое. Агрегат может быть *простым* (рис. а) и *составным* (рис. б).

A(<название>) – агрегат данных



а)



б)

Для каждого агрегата должны быть определены название и структура.

Запись

Запись – поименованная совокупность элементов данных или элементов данных и агрегатов. Различают тип записи (её структуру) и экземпляр записи, т.е. запись с конкретными значениями элементов данных. Одна запись описывает свойства одной сущности ПО (экземпляра).

Иногда термин "запись" заменяют термином "группа".

Пример записи типа "Сотрудник":



Набор

Набор (или *групповое отношение*) – поименованная совокупность записей, образующих двухуровневую иерархическую структуру. Каждый тип набора представляет собой связь между двумя или несколькими типами записей. Для каждого типа набора один тип записи объявляется владельцем набора, остальные типы записи объявляются членами набора. Для группового отношения также различают тип и экземпляр.

Фрагмент диаграммы Бахмана для БД "Город":



База данных – поименованная совокупность экземпляров групп и групповых отношений.

Операции над данными

Модель данных определяет множество действий, которые допустимо производить над некоторой реализацией БД для её перевода из одного состояния в другое. Это множество соотносят с **языком манипулирования данными** (Data Manipulation Language, DML). Любая операция над данными включает в себя **селекцию данных** (select).

По типу производимых действий различают следующие операции:

- идентификация данных ;
- выборка данных;
- включение данных в БД;
- удаление данных;
- модификация данных .

Транзакцией называют упорядоченное множество операций, переводящих БД из одного согласованного состояния в другое.

Ограничения целостности

Ограничения целостности – это правила, которым должны удовлетворять значения элементов данных. Ограничения целостности делятся на:

- **явные**
- **неявные**

Также различают **статические** и **динамические** ограничения целостности. Статические ограничения присущи всем состояниям ПО, а динамические определяют возможность перехода ПО из одного состояния в другое.

За выполнением ограничений целостности следит СУБД в процессе своего функционирования. Она проверяет ограничения целостности каждый раз, когда они могут быть нарушены (например, при добавлении данных, при удалении данных и т.п.), и гарантирует их соблюдение.

Таким образом, ограничения целостности обеспечивают логическую непротиворечивость данных при переводе БД из одного состояния в другое.

Сетевая модель данных (СМД)

Сетевая модель позволяет организовывать БД, структура которых представляется графом общего вида.

Связи между записями в СМД выполняются в виде указателей, т.е. каждая запись хранит ссылку на другую однотипную запись (или признак конца списка) и ссылки на списки подчинённых записей, связанных с ней групповыми отношениями.



СМД. Основные характеристики

1. Способ упорядочения подчинённых записей.

Поддерживаются три способа упорядочения:

- Очередь – добавление в конец списка (FIFO – first input, first output).
- Стек – добавление в начало списка (LIFO – last input, first output).
- Сортировка по значению ключа.

2. Режим включения подчинённых записей.

Режим включения бывает **автоматический** и **ручной**.

При автоматическом режиме подчинённая запись связана с записью-владельцем обязательной связью, поэтому она включается в групповое отношение и прикрепляется к записи-владельцу в момент внесения в БД.

При ручном режиме включения подчинённая запись может находиться в БД и не быть прикрепленной к записи-владельцу. Она вручную включается в групповое отношение тогда, когда это отношение (связь) возникает.

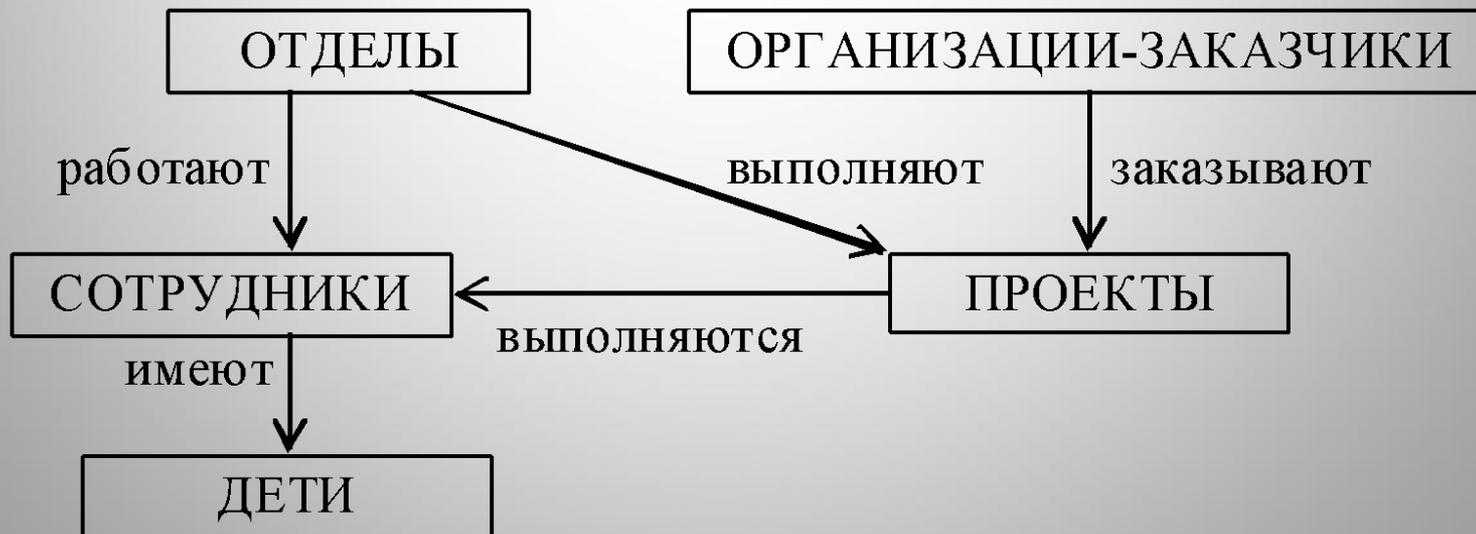
СМД. Основные характеристики

3. Режим исключения подчинённых записей.

Режим исключения определяется **классом членства**.

Различают три класса членства:

- **фиксированный**;
- **обязательный**;
- **необязательный**.

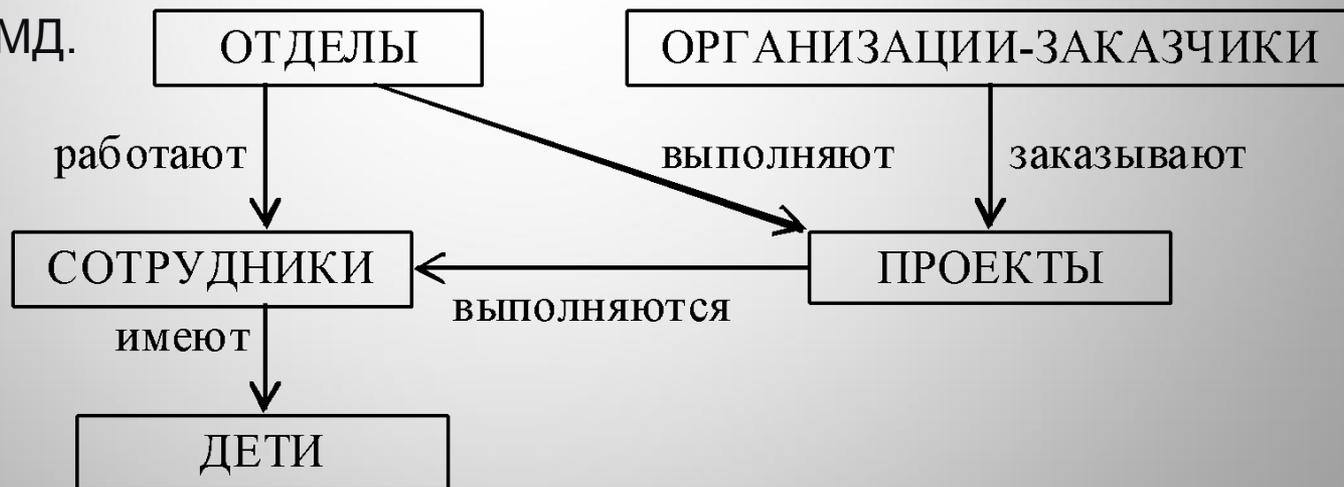


СМД. Основные характеристики

В СМД применяются следующие операции над данными:

- *запомнить*;
- *включить в групповое отношение*;
- *переключить*;
- *обновить*;
- *извлечь*;
- *удалить*;
- *исключить из группового отношения*.

Навигация в СМД.



СМД. Реализации. Достоинства и недостатки

Наиболее распространенной и стандартизованной из реализаций СМД является модель CODASYL.

Примером сетевой СУБД является система Integrated Database Management System (IDMS).

СМД является **наиболее полной** с точки зрения реализации различных типов связей и ограничений целостности, но она является достаточно **сложной** для проектирования и поддержки.

Также в СМД **не обеспечивается независимость данных от программ**. Из-за этих недостатков эта модель не получила широкого распространения.

Иерархическая модель данных (ИМД)

Иерархическая модель позволяет строить БД с иерархической древовидной структурой. В основе ИМД лежит понятие дерева.

Дерево – это связный неориентированный граф, который не содержит циклов. В случае когда дерево становится ориентированным, ориентация определяется от корня. Дерево как ориентированный граф определяется так:

- имеется единственная особая вершина, называемая корнем, в которую не заходит ни одно ребро;
- во все остальные вершины заходит только одно ребро, а исходит произвольное количество ребер;
- граф не содержит циклов.

Конечные вершины, то есть вершины, из которых не выходит ни одной дуги, называются *листьями* дерева. Количество вершин на пути от корня к листьям в разных ветвях дерева может быть различным.

ИМД. Основные характеристики

Графическая диаграмма концептуальной схемы базы данных называется *деревом определения*. Пример иерархической базы данных:



ИМД не
поддерживает:

**необязательный класс членства и ручной
режим включения записей.**

ИМД. Основные характеристики

Полный сцепленный ключ. Операции над данными. Навигация в ИМД.



Основной недостаток: дублирование данных.

Модели данных

- Иерархическая модель данных (ИМД).
 - Сетевая модель данных (СМД).
 - Реляционная модель данных (РМД).
 - Объектно-реляционная модель данных (ОРМД).
- } I поколение
- II поколение
- III поколение
- Стандарт SQL-3 (SQL-2003). Oracle (с версии 8.0), DB2, Informix, PostgreSQL, SQL Server 2008 и др.)
- Объектно-ориентированная модель данных (ООМД). O2, GemStone, Iris и др.
- Стандарт ODMG 3.0 (Object Database Management Group).
- Многомерные базы данных.
 - Поточковые базы данных.
 - ...
- 