



Базы данных



Предпосылки появления баз данных

Две основные предпосылки появления баз данных:

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

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

Базы данных: 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. [ГОСТ 20886-85]

Базы данных: 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 (ПрО).

Предметная область (ПрО) – 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].

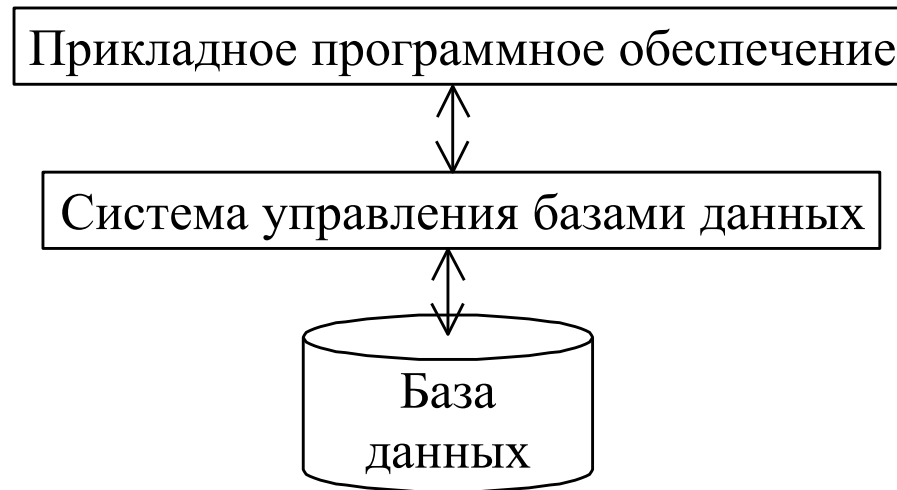
Автоматизированная информационная система (АИС) представляет собой совокупность данных, экономико-математических методов и моделей, технических, программных средств и специалистов, предназначенную для обработки информации и принятия управленческих решений.

Банк данных (БНД)

Банк данных (БНД) – это автоматизированная информационная система, включающая в свой состав комплекс специальных методов и средств (математических, информационных, программных, языковых, организационных и технических) для поддержания динамической информационной модели предметной области с целью обеспечения информационных запросов пользователей. Банк данных должен:

- Обеспечивать информационные потребности внешних пользователей.
- Обеспечивать возможность хранения и модификации больших объёмов многоаспектных данных.
- Обеспечивать заданный уровень достоверности хранимых данных и их непротиворечивость.
- Обеспечивать доступ к данным только пользователям с соответствующими полномочиями.
- 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.

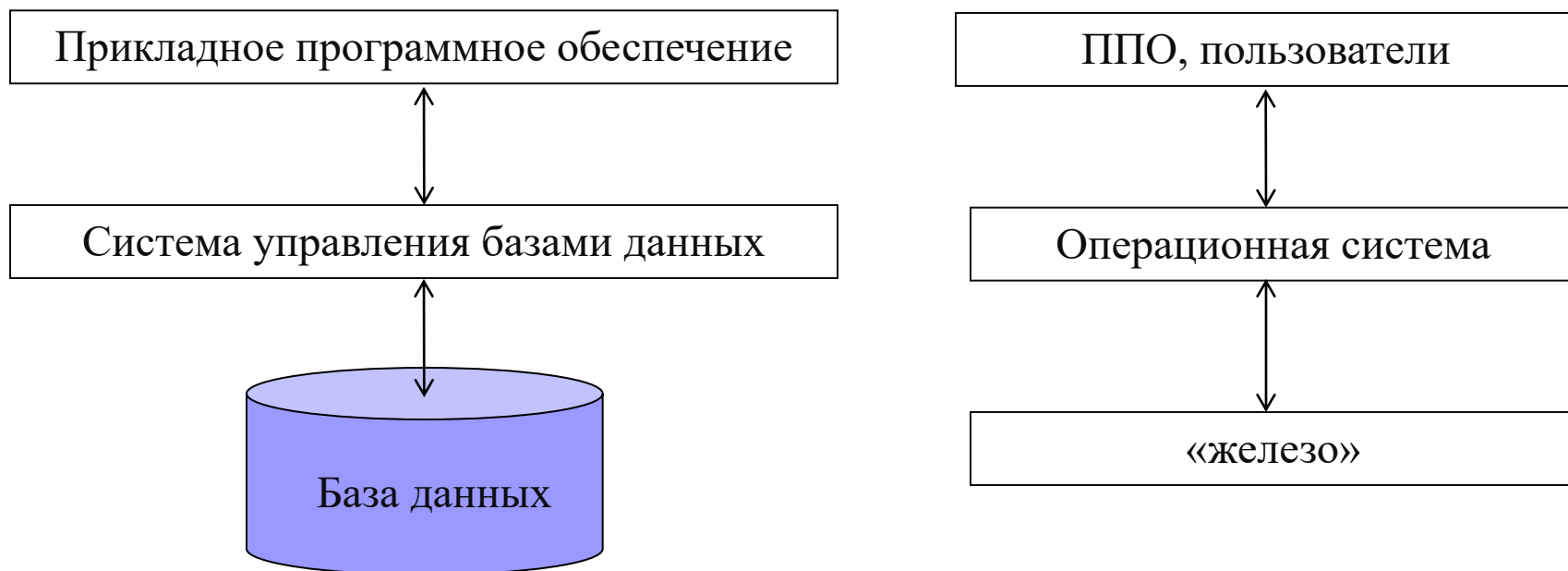
Компоненты системы баз данных



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

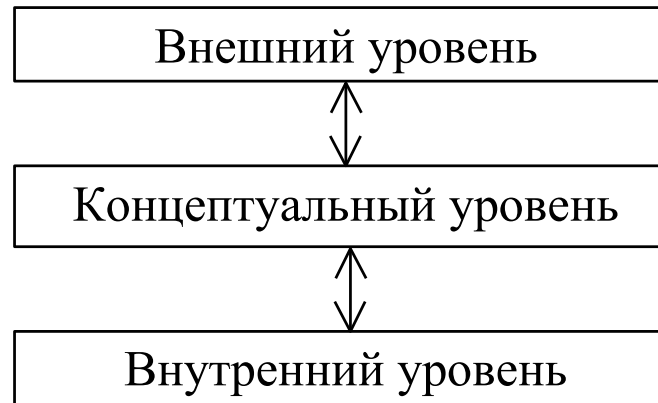
Описание базы данных хранится в так называемом **словаре-справочнике данных (ССД) или каталоге данных**. Хранение данных вместе с их описанием и позволяет обеспечивать независимость данных от программ, т.е. интерпретация данных определяется не программой, а описанием данных.

Основные функции СУБД



- Обеспечение доступа ППО (пользователей) к базе данных
- Управление базой данных

Уровни представления данных (архитектура ANSI/SPARC)



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.

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

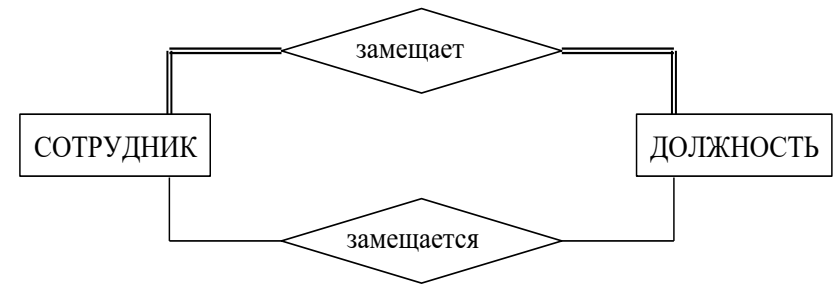
Связь – это осмысленная ассоциация между сущностями. Для связи указываются:

- название,
- вид (факультативная или обязательная),
- степень (унарная, бинарная, тернарная или n-арная),
- кардинальность (1:1, 1:n или m:n).

Степень определяется количеством типов сущностей, входящих в связь.

Семантика связи: определяется предметной областью. Например: студент *учится в* группе.

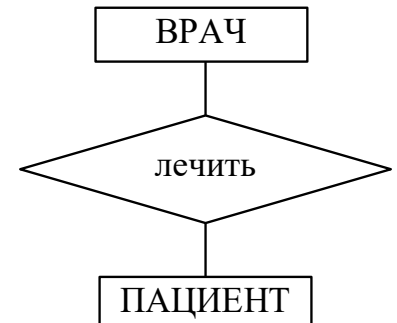
Экземпляр связи: относится к конкретным экземплярам сущностей. Например: студент Попов А.Н. учится в группе СТ-101.



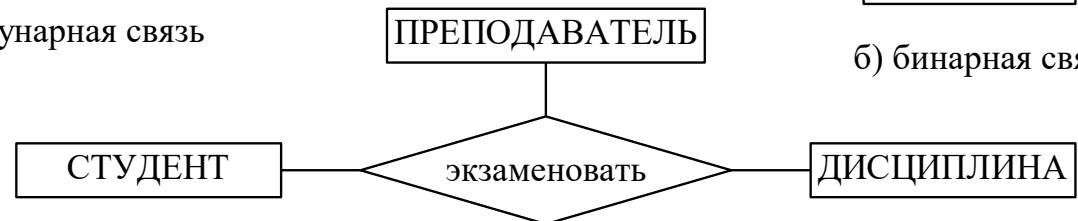
(Сущность базовая и зависимая, родительская и подчиненная.)



а) унарная связь



б) бинарная связь

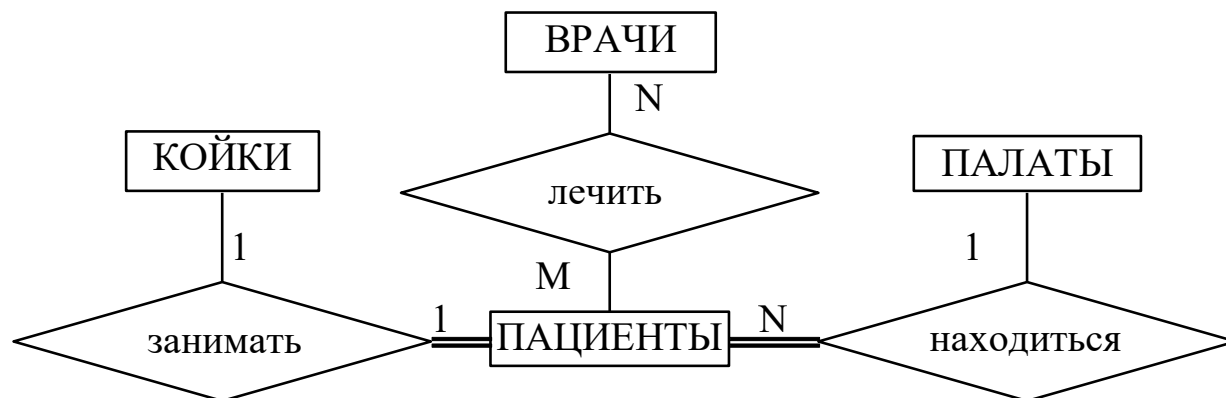


в) тернарная связь

Предметная область. Связи

Кардинальность

связи определяется количеством экземпляров сущностей, входящих в связь.



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

- пациент–койка (1:1) – каждый пациент занимает одну койку, каждая койка в каждый момент времени может быть занята только одним пациентом;
- палата–пациент (1:n) – каждый пациент находится в одной палате, в каждой палате могут находиться несколько пациентов;
- пациент–врач (n:m) – каждый пациент может лечиться у нескольких врачей, каждый врач может лечить несколько пациентов.

Операции над данными

Любая операция над данными включает в себя **селекцию данных** (select).

Условие селекции – это некоторый критерий отбора данных, в котором могут быть использованы логическая позиция элемента данных, его значение и связи между данными.

По типу производимых действий различают следующие операции:

- идентификация данных и нахождение их позиции в БД;
- выборка (чтение) данных из БД;
- включение (запись) данных в БД;
- удаление данных из БД;
- модификация (изменение) данных БД;
- формирование связей с другими данными (записями).

Обработка данных в БД осуществляется с помощью процедур базы данных – транзакций. **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 и связей между ними.

Ограничения целостности делятся на:

- **явные** (включаются в структуру базы данных с помощью средств языка контроля данных (DCL, Data Control Language))
- **неявные** (определяются самой структурой данных).

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.

Актуализация данных в БД

Каждому моменту времени можно сопоставить некоторое **состояние предметной области**. Состояния ПО должны подчиняться совокупности правил, которые характеризуют семантику предметной области.

Ограничения целостности: правила, которым подчиняются данные в предметной области.

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, если изменения в данные вносит пользователь;
- автоматизировано, если изменения инициируются пользователем, но выполняются программно;
- автоматически, если данные поступают в электронном виде и обрабатываются программой без участия человека.

Контроль правильности вносимых изменений: все изменения данных должны соответствовать ограничениям целостности предметной области. Контроль может осуществляться: вручную, программно, с помощью дополнительных средств СУБД и на уровне базы данных (т.н. **ограничения целостности базы данных**).

MySQL

Система управления базами данных (СУБД), которая позволяет хранить, управлять и извлекать данные.



Основные команды MySQL

- `USE [имя БД]` – установить БД с указанным именем в качестве текущей
- `SHOW TABLES` – показать список таблиц текущей БД
- `SHOW TABLES FROM [имя БД]` – показать список таблиц из указанной БД
- `SHOW FIELDS FROM [имя таблицы]` – показать список полей указанной таблицы
- `DESC [имя таблицы]` – альтернативный вариант для `SHOW TABLES FROM`
- `SELECT COUNT(*) FROM [имя таблицы]` – показать количество строк в указанной таблице