

Тема 9. Базы данных.

Вопрос 1. Понятие баз данных.

- **База данных (БД)** — совокупность определенным образом организованной информации в рамках некоторой предметной области.
- **База данных** — организованная совокупность данных, предназначенная для длительного хранения во внешней памяти ЭВМ и постоянного применения.

- Базы данных бывают **фактографическими и документальными.**
- В фактографических БД содержатся краткие сведения об описываемых объектах, представленные в строго определенном формате. (база данных книжного фонда библиотеки; база данных кадрового состава учреждения).
- Документальная БД содержит обширную информацию разного типа: текстовую, графическую, звуковую, мультимедийную.

- **Информационная система** — это совокупность базы данных и всего комплекса аппаратно-программных средств для ее хранения, изменения и поиска информации, для взаимодействия с пользователем.
- Примерами информационных систем являются системы продажи билетов на пассажирские поезда и самолеты.
- WWW — это тоже пример глобальной информационной системы.

- Если различные части одной базы данных хранятся на множестве компьютеров, объединенных между собой сетью, то такая БД называется **распределенной базой данных**.
- Модель данных строится по принципу взаимосвязанных таблиц - **реляционной**.
- Один тип объекта является главным, все нижележащие – подчиненными - **иерархической**.
- Любой тип данных одновременно может быть главным и подчиненным - **сетевой**.

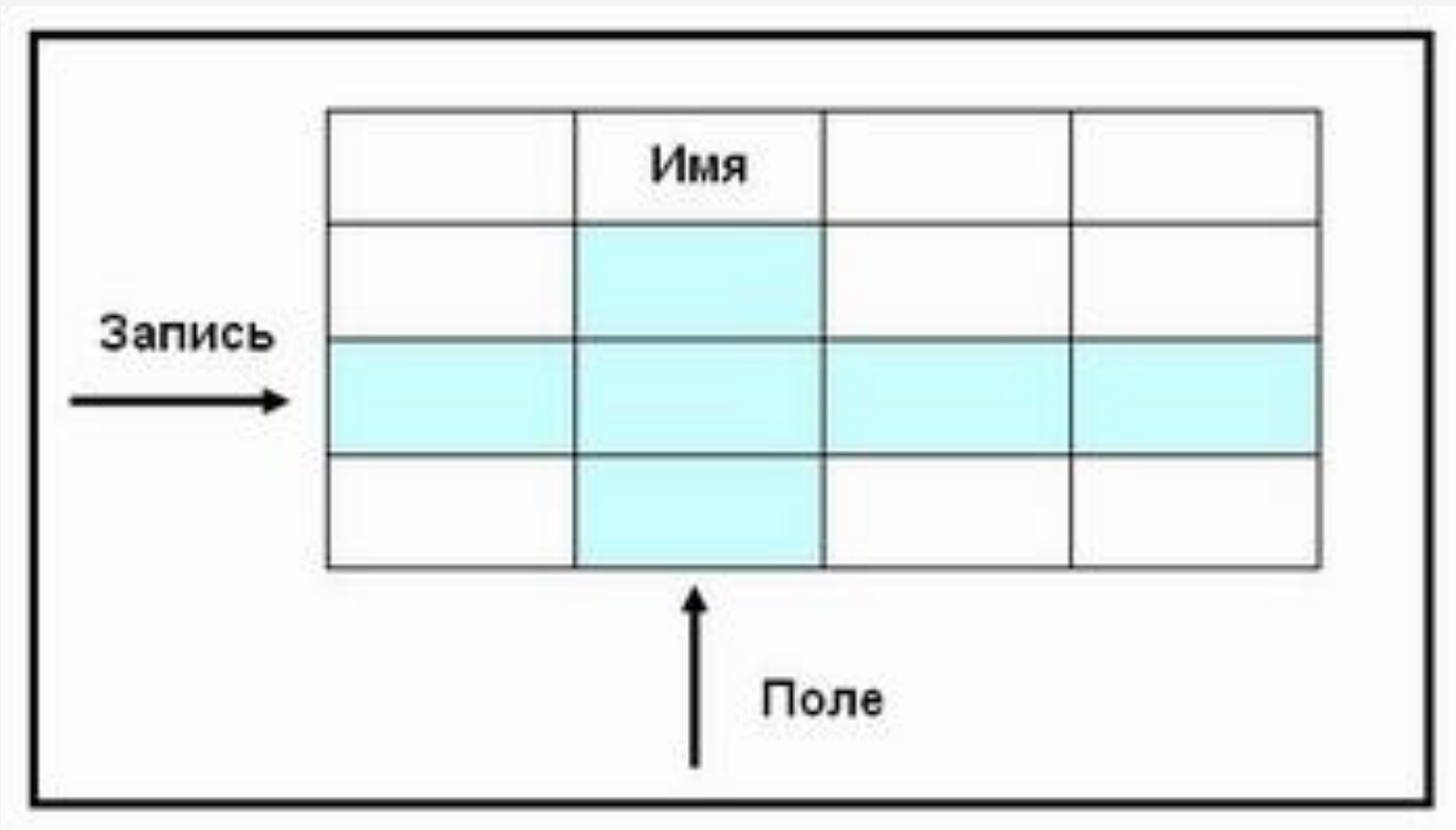
Функции баз данных:

1. Возможность совместного доступа к данным.
2. Сокращение избыточности данных;
3. Возможность поддержки транзакций;
Транзакция – логическая единица работы, обычно включающая несколько операций базы данных;
4. Обеспечение целостности данных;
5. Организация защиты данных;
6. Возможность балансировки противоречивых требований
7. Независимость данных.

Реляционные базы данных

- Базы данных с табличной формой организации называются реляционными БД.
- В реляционных БД строка таблицы называется **записью**, а столбец — **полем**.
- Каждое поле таблицы имеет имя.
- Одна запись содержит информацию об одном объекте той реальной системы, модель которой представлена в таблице.

- Общий вид таблицы реляционной базы данных:




- **Поля** — это различные характеристики (атрибуты) объекта.
- Значения полей в одной строчке относятся к одному объекту.
- Разные поля отличаются именами.
- Записи различаются значениями ключей.
- Главным **ключом** в базах данных называют поле (или совокупность полей), значение которого не повторяется у разных записей.

- В БД «Домашняя библиотека» разные книги могут иметь одного автора, могут совпадать названия книг, год издания, полка.
- Но **инвентарный номер** у каждой книги свой (поле НОМЕР).
- Он-то и является **главным ключом** для записей в этой базе данных.
- Ключ может состоять и более чем из двух полей., такой ключ называется **составным**.

- Пример простого ключа:

Простой ключ



Номер	Автор	Название	Год	Полка
001	Беляев А.Р.	Звезда КЭЦ	1990	3
002	Олеша Ю.К.	Избранное	1987	5
003	Беляев А.Р.	Избранное	1994	1

В БД «Домашняя библиотека» у разных книг могут совпадать значения полей, но инвентарный номер у каждой книги свой

Пример составного ключа:

- База данных, которая хранится в компьютере управления образования области.
- В ней содержатся сведения о всех средних школах районных центров в виде такой таблицы:
- В такой таблице у разных записей не могут совпасть только одновременно два поля ГОРОД и НОМЕР ШКОЛЫ.
- Эти два поля вместе образуют составной ключ: ГОРОД-НОМЕР ШКОЛЫ.

Составной ключ



Город	№ школы	Директор	Адрес	Телефон
Крюков	1	Иванов А.П.	Пушкина, 5	2-33-56
Шадринск	1	Строев С.С.	Лесная, 14	4-33-11
Шадринск	2	Иванов А.П.	Мира, 34	4-23-24

В БД «Школы области» у разных записей одновременно не могут совпадать только сочетание двух полей:
город и номер школы (это составной ключ)

- С каждым полем связано еще одно очень важное свойство — **тип поля**.
- *Тип определяет множество значений, которые может принимать данное поле в различных записях.*
- В реляционных базах данных используются четыре основных типа полей:
 - числовой;
 - символьный;
 - дата;
 - логический.

- **Числовой тип** имеют поля, значения которых могут быть только числами.
- **Символьный тип** имеют поля, в которых будут храниться символьные последовательности (слова, тексты, коды и т.п.).
- **Тип «дата»** имеют поля, содержащие календарные даты в форме «день/месяц/год» (в некоторых случаях используется американская форма: месяц/день/год).
- **Логический тип** соответствует полю, которое может принимать всего два значения: «да» — «нет» или «истина» — «ложь».

Принципы построения систем управления баз данных следуют из **требований, которым должна удовлетворять организация баз данных:**

- **Производительность и готовность.** Запросы от пользователя базой данных удовлетворяются с такой скоростью, которая требуется для использования данных;
- **Минимальные затраты;**

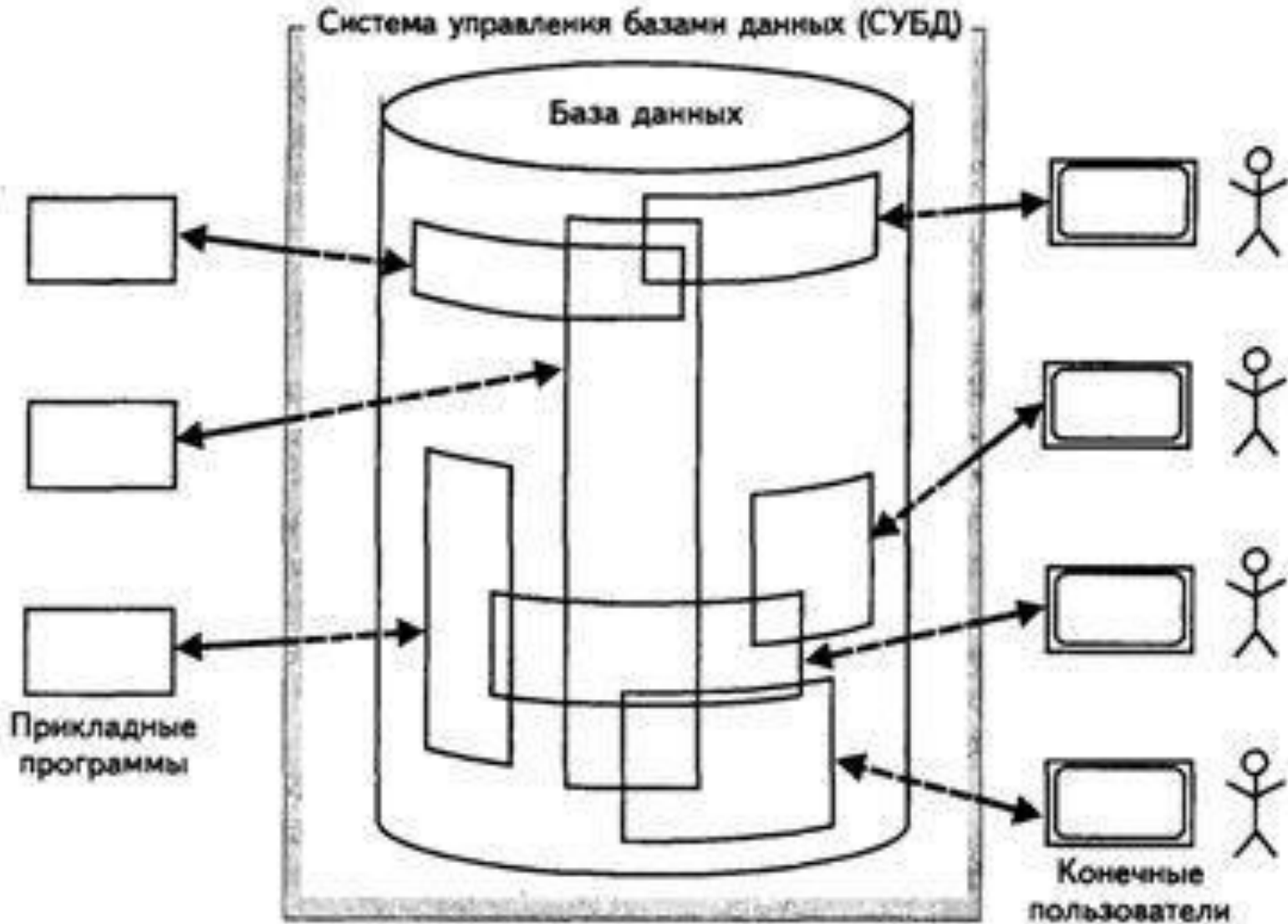
- **Простота и легкость использования;**
- **Простота внесения изменений;**
- **Возможность поиска;**
- **Целостность.** Базы данных могут содержать данные, используемые многими пользователями. Очень важно, чтобы в процессе работы элементы данных и связи между ними не нарушались.;
- **Безопасность и секретность.**

Вопрос 2. Системы управления баз данных.

- **Система управления баз данных (СУБД)** Программный комплекс, предназначенный для управления распределенными базами данных и обеспечивающий прозрачный доступ пользователей к распределенной информации.

- Если прикладная информационная система опирается на некоторую систему управления данными, обладающую этими свойствами, то эта система управления данными является *системой управления базами данных (СУБД)*.
- **Система управления баз данных** – компьютеризированная система хранения записей, основное назначение которой – хранить данные, предоставляя пользователям средства ее извлечения и модификации.

- Упрощенная схема СУБД



- **Функции СУБД:**

- 1. Непосредственное управление данными во внешней памяти***
- 2. Управление буферами оперативной памяти***
- 3. Управление транзакциями***
- 4. Журнализация**
- 5. Поддержка языков БД**

Непосредственное управление данными во внешней памяти

- Эта функция включает обеспечение необходимых структур внешней памяти как для хранения данных, непосредственно входящих в БД, так и для служебных целей, например, для ускорения доступа к данным в некоторых случаях (обычно для этого используются индексы).

Управление транзакциями

- С управлением транзакциями в многопользовательской СУБД связаны важные понятия
- *сериализации транзакций* и
- *сериального плана выполнения смеси транзакций.*

- Под **сериализаций параллельно выполняющихся транзакций** понимается такой порядок планирования их работы, при котором суммарный эффект смеси транзакций эквивалентен эффекту их некоторого последовательного выполнения.
- **Сериальный план выполнения смеси транзакций** - это такой план, который приводит к сериализации транзакций.

Журнализация

- Журнал - это особая часть БД, недоступная пользователям СУБД, в которую поступают записи обо всех изменениях основной части БД.
- Для восстановления БД после сбоя используют журнал и **архивную копию БД**.
- **Архивная копия** - это полная копия БД к моменту начала заполнения журнала

Поддержка языков БД

- Для работы с базами данных используются специальные языки, в целом называемые языками баз данных.
- В современных СУБД обычно поддерживается единый интегрированный язык, содержащий все необходимые средства для работы с БД, начиная от ее создания, и обеспечивающий базовый пользовательский интерфейс с базами данных.

- Чаще всего выделялись два языка - **язык определения схемы БД** (*SDL - Schema Definition Language*) и **язык манипулирования данными** (*DML - Data Manipulation Language*).

- **SDL** служил главным образом для определения логической структуры БД, т.е. той структуры БД, какой она представляется пользователям.
- **DML** содержал набор операторов манипулирования данными, т.е. операторов, позволяющих заносить данные в БД, удалять, модифицировать или выбирать существующие данные.

- Стандартным языком наиболее распространенных в настоящее время реляционных СУБД является язык SQL (Structured Query Language).
- Язык SQL сочетает средства SDL и DML, т. е. позволяет определять схему реляционной БД и манипулировать данными.

Вопрос 3. Типовая организация СУБД

Организация типичной СУБД и состав ее компонентов соответствует рассмотренному нами набору **функций**.

Логически в современной реляционной СУБД можно выделить:

- ядро СУБД (Data Base Engine);
- компилятор языка БД (обычно SQL);
- подсистему поддержки времени выполнения;
- набор утилит.

- Ядро СУБД отвечает за управление данными во внешней памяти, управление буферами оперативной памяти, управление транзакциями и журнализацию.
- Можно выделить такие компоненты ядра как:
 - менеджер данных,
 - менеджер буферов,
 - менеджер транзакций и
 - менеджер журнала.

- Ядро СУБД обладает собственным интерфейсом, не доступным пользователям напрямую и используемым в программах, производимых компилятором SQL (или в подсистеме поддержки выполнения программ) и утилитах БД.
- Ядро СУБД является основной резидентной частью СУБД.

- Основной функцией компилятора языка БД является компиляция операторов языка БД в некоторую выполняемую программу.
- Результатом компиляции является выполняемая программа, представляемая в некоторых системах в машинных кодах, но более часто в выполняемом внутреннем машинно-независимом коде.

- В последнем случае реальное выполнение оператора производится с привлечением подсистемы поддержки времени выполнения, представляющей собой, по сути дела, интерпретатор этого внутреннего языка.

- В отдельные утилиты БД обычно выделяют такие процедуры как, например:
 - загрузка и выгрузка БД,
 - сбор статистики,
 - глобальная проверка целостности БД и т.д.
- Утилиты программируются с использованием интерфейса ядра СУБД, а иногда даже с проникновением внутрь ядра.