**Лекция 2:**

**Основные понятия баз данных**

**Аннотация:**В лекции рассматривается общий смысл понятий базы данных (БД) и системы управления базами данных (СУБД). Даются основные понятия, относящиеся к базе данных такие, как алгоритм, кортеж, объект, сущность. Основные требования, предъявляемые к банку данных. Определения БД и СУБД.

**Аннотация:**В лекции рассматривается общий смысл понятий базы данных (БД) и системы управления базами данных (СУБД). Даются основные понятия, относящиеся к базе данных такие, как алгоритм, кортеж, объект, сущность. Основные требования, предъявляемые к банку данных. Определения БД и СУБД.

**Цель лекции:** Уяснить разницу между базой данных и системой управления базой данных. Ознакомиться с основными требованиями, которые предъявляются к банку данных и основными определениями, относящимися к БД и СУБД.

Рассмотрим общий смысл понятий базы данных (БД) и системы управления базами данных (СУБД).

С самого начала развития вычислительной техники образовались два основных направления использования ее.

Первое направление - применение вычислительной техники для выполнения численных расчетов, которые слишком долго или вообще невозможно производить вручную. Становление этого направления способствовало интенсификации методов численного решения сложных математических задач, развитию класса языков программирования, ориентированных на удобную запись численных алгоритмов, становлению обратной связи с разработчиками новых архитектур ЭВМ.

Второе направление, это использование средств вычислительной техники в автоматических или *автоматизированных информационных системах*. В самом широком смысле информационная система представляет собой программный комплекс, функции которого состоят в поддержке надежного хранения информации в памяти компьютера, выполнении специфических для данного приложения преобразований информации и/или вычислений, предоставлении пользователям удобного и легко осваиваемого интерфейса. Обычно объемы информации, с которыми приходится иметь дело таким системам, достаточно велики, а сама информация имеет достаточно сложную структуру. Классическими примерами информационных систем являются *банковские системы*, системы резервирования авиационных или железнодорожных билетов, мест в гостиницах и т.д.

На самом деле, второе направление возникло несколько позже первого. Это связано с тем, что на заре вычислительной техники компьютеры обладали ограниченными возможностями в части памяти. Понятно, что можно говорить о надежном и долговременном хранении информации только при наличии запоминающих устройств, сохраняющих информацию после выключения электрического питания. Оперативная память этим свойством обычно не обладает. В начале, использовались два вида устройств внешней памяти: магнитные ленты и барабаны. При этом емкость магнитных лент была достаточно велика, но по своей физической природе они обеспечивали *последовательный доступ* к данным. Магнитные же барабаны (они больше всего похожи на современные магнитные диски с фиксированными головками) давали возможность произвольного доступа к данным, но были ограниченного размера.

Легко видеть, что указанные ограничения не очень существенны для чисто численных расчетов. Даже если программа должна обработать (или произвести) большой объем информации, при программировании можно продумать расположение этой информации во внешней памяти, чтобы программа работала как можно быстрее.

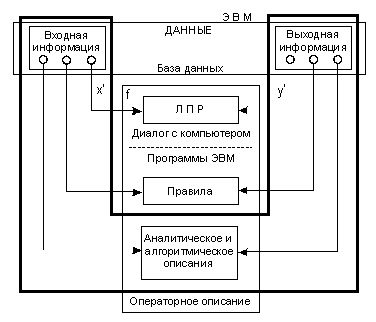
С другой стороны, для информационных систем, в которых потребность в текущих данных определяется пользователем, наличие только магнитных лент и барабанов неудовлетворительно. Представьте себе покупателя билета, который стоя у кассы должен дождаться полной перемотки магнитной ленты. Одним из естественных требований к таким системам является средняя быстрота выполнения операций.

Именно требования к вычислительной технике со стороны не численных приложений вызвали появление съемных *магнитных дисков с подвижными головками*, что явилось революцией в истории вычислительной техники. Эти *устройства внешней памяти* обладали существенно большей емкостью, чем магнитные барабаны, обеспечивали удовлетворительную скорость доступа к данным в режиме произвольной выборки, а возможность смены дискового пакета на устройстве позволяла иметь практически неограниченный архив данных.

С появлением магнитных дисков началась история систем управления данными во внешней памяти. До этого каждая прикладная программа, которой требовалось хранить данные во внешней памяти, сама определяла расположение каждой порции данных на магнитной ленте или барабане и выполняла обмены между оперативной и внешней памятью с помощью программно-аппаратных средств низкого уровня (машинных команд или вызовов соответствующих программ операционной системы). Такой режим работы не позволяет или очень затрудняет поддержание на одном внешнем носителе нескольких архивов долговременно хранимой информации. Кроме того, каждой прикладной программе приходилось решать проблемы именования частей данных и структуризации данных во внешней памяти.

Историческим шагом стал переход к использованию систем управления файлами. С точки зрения прикладной программы файл - это *именованная область* внешней памяти, в которую можно записывать и из которой можно считывать данные. Правила именования файлов, способ доступа к данным, хранящимся в файле, и структура этих данных зависят от конкретной *системы управления файлами* и, возможно, от типа файла. *Система управления файлами* берет на себя распределение внешней памяти, отображение имен файлов в соответствующие адреса внешней памяти и обеспечение доступа к данным.

Любая задача обработки информации и принятия решений может быть представлена в виде схемы, показанной на [рис. 1.1](https://intuit.ru/studies/courses/3439/681/lecture/14017?page=1#image.1.1).



**Рис. 1.1.**Схема решения задач обработки информации и принятия решений: x-штрих, y-штрих - входная и выходная информация; f - внутреннее операторное описание

**Определение основных терминов**

Дадим определения основных терминов. В качестве составных частей схемы выделяются информация (входная и выходная) и правила ее преобразования.

Правила могут быть в виде алгоритмов, процедур и эвристических последовательностей.

|  |  |
| --- | --- |
| **Алгоритм** | - последовательность правил перехода от исходных данных к результату. Правила могут выполняться компьютером или человеком. |
| **Данные** | - совокупность объективных сведений. |
| **Информация** | - сведения, неизвестные ранее получателю информации, пополняющие его знания, подтверждающие или опровергающие положения и соответствующие убеждения. Информация носит субъективный характер и определяется уровнем знаний субъекта и степенью его восприятия. Информация извлекается субъектом из соответствующих данных. |
| **Знания** | - совокупность фактов, закономерностей и эвристических правил, с помощью которых решается поставленная задача. |

Последовательность операций обработки данных называют информационной технологией (ИТ). В силу значительного количества информации в современных задачах она должна быть упорядочена. Существует два подхода к упорядочению.

1. Данные связаны с конкретной задачей (технология массивов) - упорядочение по использованию. Вместе с тем алгоритмы более подвижны (могут чаще меняться), чем данные. Это вызывает необходимость переупорядочения данных, которые к тому же могут повторяться в различных задачах.
2. В связи с этим предложена другая, широко используемая технология баз данных, представляющая собой упорядочение по хранению.

|  |  |
| --- | --- |
| **КОДАСИЛ (CODASYL)** | - набор стандартов для *сетевых БД*. |
| **Кортеж** | - совокупность полей или запись. |
| **Объект** | - термин, обозначающий факт, лицо, событие, предмет, о котором могут быть собраны данные. |
| **Сущность** | - примитивный объект данных, отображающий элемент предметной области (человек, место, вещь и т.д.). |

Под **базой данных (БД)** понимают совокупность хранящихся вместе данных при наличии такой минимальной избыточности, которая допускает их использование оптимальным образом для одного или нескольких приложений. Целью *создания баз данных*, как разновидности информационной технологии и формы хранения данных, является построение системы данных, не зависящих от принятых алгоритмов (программного обеспечения), применяемых технических средств и физического расположения данных в ЭВМ; обеспечивающих непротиворечивую и целостную информацию при нерегламентируемых запросах. БД предполагает многоцелевое ее использование (несколько пользователей, множество форм документов и запросов одного пользователя).

**База знаний (БЗ)** представляет собой совокупность БД и используемых правил, полученных от лиц, принимающих решения (*ЛПР*).

Наряду с понятием "база данных" существует термин "*банк данных*", который имеет две трактовки.

1. В настоящее время данные обрабатываются *децентрализовано* (на рабочих местах) с помощью персональных компьютеров (ПК). Первоначально же использовалась централизованная обработка на больших ЭВМ. В силу централизации базу данных называли банком данных и потому часто не делают различия между базами и банками данных.
2. **Банк данных** - база данных и система управления ею (СУБД). СУБД (например, FoxPro) представляет собой приложение для *создания баз данных* как совокупности двумерных таблиц.

|  |  |
| --- | --- |
| **Банк данных (БнД)** | - это система специально организованных данных, программных, языковых, организационных и технических средств, предназначенных для централизованного накопления и коллективного многоцелевого использования данных. |
| **Базы данных (БД)** | - это именованная совокупность данных, отображающая состояние объектов и их отношения в рассматриваемой предметной области. Характерной чертой баз данных является постоянство: данные постоянно накапливаются и используются; состав и структура данных, необходимы для решения тех или иных прикладных задач, обычно постоянны и стабильны во времени; отдельные или даже все элементы данных могут меняться - но и это есть проявления постоянства - постоянная актуальность. |
| **Система управления базами данных (СУБД)** | - это совокупность языковых и программных средств, предназначенных для создания, ведения и совместного использования БД многими пользователями. |

Иногда в составе банка данных выделяют архивы. Основанием для этого является особый режим использования данных, когда только часть данных находится под оперативным управлением СУБД. Все остальные данные обычно располагаются на носителях, оперативно не управляемых СУБД. Одни и те же данные в разные моменты времени могут входить как в базы данных, так и в архивы. Банки данных могут не иметь архивов, но если они есть, то в состав банка данных может входить и система управления архивами.

*Эффективное управление* внешней памятью являются основной *функцией СУБД*. Эти обычно специализированные средства настолько важны с точки зрения эффективности, что при их отсутствии система просто не сможет выполнять некоторые задачи уже по тому, что их выполнение будет занимать слишком много времени. При этом ни одна из таких специализированных функций не является видимой для пользователя. Они обеспечивают независимость между логическим и физическим уровнями системы: прикладной программист не должен писать программы индексирования, распределять память на диске и т. д.

**Основные требования, предъявляемые к банкам данных**

Развитие теории и практики создания информационных систем, основанных на концепции баз данных, создание унифицированных методов и средств организации и поиска данных позволяют хранить и обрабатывать информацию о все более сложных объектах и их взаимосвязях, обеспечивая многоаспектные информационные потребности разных пользователей. Основные требования, предъявляемые к банкам данных, можно сформулировать так:

* **Многократное использование данных:** пользователи должны иметь возможность использовать данные различным образом.
* **Простота:** пользователи должны иметь возможность легко узнать и понять, какие данные имеются в их распоряжении.
* **Легкость использования:** пользователи должны иметь возможность осуществлять (процедурно) простой доступ к данным, при этом все сложности доступа к данным должны быть скрыты в самой системе управления базами данных.
* **Гибкость использования:** обращение к данным или их поиск должны осуществляться с помощью различных методов доступа.
* **Быстрая обработка запросов на данные:** запросы на данные должны обрабатываться с помощью высокоуровневого *языка запросов*, а не только прикладными программами, написанными с целью обработки конкретных запросов.
* **Язык взаимодействия** конечных пользователей с системой должен обеспечивать конечным пользователям возможность получения данных без использования прикладных программ.

База данных - это основа для будущего наращивания прикладных программ: базы данных должны обеспечивать возможность быстрой и дешевой разработки новых приложений.

* **Сохранение затрат умственного труда:** существующие программы и *логические структуры* данных не должны переделываться при внесении изменений в базу данных.
* **Наличие интерфейса прикладного программирования:** прикладные программы должны иметь возможность просто и эффективно выполнять запросы на данные; программы должны быть изолированными от расположения файлов и *способов адресации* данных.
* **Распределенная обработка данных:** система должна функционировать в условиях вычислительных сетей и обеспечивать эффективный доступ пользователей к любым данным распределенной БД, размещенным в любой точке сети.
* **Адаптивность и расширяемость:** база данных должна быть настраиваемой, причем настройка не должна вызывать перезаписи прикладных программ. Кроме того, поставляемый с СУБД набор предопределенных типов данных должен быть расширяемым - в системе должны иметься средства для определения новых типов и не должно быть различий в использовании системных и определенных пользователем типов.
* **Контроль целостности данных:** система должна осуществлять контроль ошибок в данных и выполнять проверку взаимного логического соответствия данных.
* **Восстановление данных после сбоев:** автоматическое восстановление без потери данных транзакции. В случае аппаратных или программных сбоев система должна возвращаться к некоторому согласованному состоянию данных.
* **Вспомогательные средства** должны позволять разработчику или *администратору базы данных* предсказать и оптимизировать производительность системы.
* **Автоматическая реорганизация и перемещение:** система должна обеспечивать возможность перемещения данных или автоматическую реорганизацию *физической структуры*.

**Компоненты банка данных**

Определение банка данных предполагает, что с функционально-организационной точки зрения *банк данных* является сложной человеко-машинной системой, включающей в себя все подсистемы, необходимые для надежного, эффективного и продолжительного во времени функционирования.

В структуре банка данных выделяют следующие компоненты:

* Информационная база;
* Лингвистические средства;
* Программные средства;
* Технические средства;
* Организационно-административные подсистемы и нормативно-методическое обеспечение.

**Организационно-методические средства** - это совокупность инструкций, методических и регламентирующих материалов, описаний структуры и процедуры работы пользователя с СУБД и БД.

**Пользователи БД и СУБД**

Пользователей (СУБД) можно разделить на две основные категории: **конечные пользователи**; **администраторы баз данных**.

Особо следует поговорить об администраторе базы данных (АБД). Естественно, что база данных строится для конечного пользователя (КП). Однако первоначально предполагалось, что КП не смогут работать без специалиста-программиста, которого назвали администратором базы данных. С появлением СУБД они взяли на себя значительную часть функций АБД, особенно для БД с небольшим объемом данных. Однако для крупных централизованных и распределенных баз данных потребность в АБД сохранилась. В широком плане под АБД понимают системных аналитиков, проектировщиков структур данных и информационного обеспечения, проектировщиков технологии процессов обработки, системных и прикладных программистов, операторов, специалистов в предметной области и по техническому обслуживанию. Иными словами, в крупных базах данных это могут быть коллективы специалистов. В обязанности АБД входит:

1. анализ предметной области, статуса информации и пользователей;
2. проектирование структуры и модификация данных;
3. задание и обеспечение целостности;
4. загрузка и ведение БД;
5. защита данных;
6. обеспечение восстановления БД;
7. сбор и статистическая обработка обращений к БД, анализ эффективности функционирования БД;
8. работа с пользователем.

**Краткие итоги**

Базы данных (БД) - это именованная совокупность данных, отображающая состояние объектов и их отношения в рассматриваемой предметной области.

Система управления базами данных (СУБД) - это совокупность языковых и программных средств, предназначенных для создания, ведения и совместного использования БД многими пользователями.

Основные требования, предъявляемые к банкам данных: многократное использование данных, простота, легкость использования, гибкость использования, быстрая обработка запросов на данные, язык взаимодействия.

Пользователей (СУБД) можно разделить на две основные категории: конечные пользователи; администраторы баз данных.

**Вопросы для самопроверки**

* Дайте определение базы данных.
* Дайте определение банка данных.
* Назовите две трактовки банка данных.
* Что такое система управления базой данных?
* Основные требования, предъявляемые к банку данных.
* Что такое данные, информация, знания?
* Пользователи СУБД и БД?
* Основные функции администратора БД.
* Что обеспечивает возможность быстрой и дешевой разработки новых приложений?