**Тема 1. Введение в предмет, основные понятия и термины**

***Основные понятия.***

Исходное представление о базах данных массового пользователя будет, скорее всего, построено на аналогии: база данных - база, склад, хранилище, что близко к действительности. Проводя дальнейшую аналогию, можно сказать, что база данных – это хранилище, где принимают взаимосвязанные данные, сортируют по темам и хранят в некотором порядке в одном месте, а также частично перерабатывают. Эти данные выдают строго по назначению и по требованию, оформленному согласно определенным правилам. Потребителю не обязательно знать, в каком виде хранятся данные, при этом он решает какие данные, в каком количестве и в какой форме ему их нужно выдать. Это образное и, в общем, верное представление.

Основные идеи современных информационных технологий опираются на концепцию баз данных. Следовательно, базовым элементом информационных технологий являются данные.

В основе концепции лежит механизм предоставления обрабатывающей программе из всех хранимых данных только тех, которые ей необходимы, и в форме, требуемой именно этой программе. Сама же форма описывается на логическом уровне, т.е. уровне, видимом из программы.

Одним из важнейших понятий в теории баз данных является понятие информации. ***Информация*** – это любые сведения о каком-либо событии, процессе, объекте. Другими словами, это все, что может интересовать пользователя.

***Данные*** – это информация, представленная в определенном виде. Для компьютерных технологий данные представлены в дискретном, фиксированном виде, удобном для хранения и обработки на компьютере, а также для передачи по каналам связи.

Таким образом, ***база данных*** – это именованная совокупность данных, отражающая состояние объектов и их отношений в рассматриваемой предметной области.

Характерной чертой баз данных является постоянство:

данные постоянно накапливаются и используются;

состав и структура данных, необходимых для решения тех или иных прикладных задач, обычно постоянны во времени;

отдельные или все элементы данных могут меняться, что является признаком постоянной актуальности.

Т.е. речь идет о ***признаке персистентности*** (устойчивости) данных в базе данных.

Таким образом, можно привести более точное определение базы данных, а именно, ***база данных*** – это совокупность взаимосвязанных устойчивых данных при наличии такой минимальной избыточности, которая допускает их независимое использование оптимальным образом для одного или нескольких приложений в определенной предметной области. БД состоит из множества связанных файлов, причем связи (отношения) между ними хранятся вместе данными.

Поскольку одни и те же данные могут использоваться для решения многих задач, то и приложений (программ) к одной и той же базе данных может быть много. Все приложения, работающие с одной и той же базой данных, должны функционировать корректно, не мешать друг другу и учитывать все изменения, которые вносятся другими приложениями. Работу всех приложений координирует СУБД.

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

Примеры СУБД: MS SQL, MS Access, Oracle.

***Классификация БД.***

По технологии обработки данных базы данных подразделяются на распределенные и централизованные.

Распределенная база данных состоит из нескольких, возможно, пересекающихся или даже дублирующих друг друга частей, которые хранятся в различных компьютерах одной вычислительной сети. Т.е. это набор логически связанных между собой совокупностей разделяемых данных (и их описаний), которые физически распределены в некоторой компьютерной сети. Работа с такой базой осуществляется с помощью распределенной системы управления базой данных.

***Распределенная СУБД*** – это программный комплекс, предназначенный для управления распределенными базами данных и обеспечивающий прозрачный доступ пользователей к распределенной информации.

Из определения следует, что СУБД должна сделать распределение данных прозрачным (незаметным) для конечного пользователя. Другими словами, от пользователей должен быть полностью скрыт тот факт, что распределенная база данных состоит из нескольких фрагментов, которые могут размещаться на различных компьютерах. Цель обеспечения прозрачности состоит в том, чтобы распределенная система внешне выглядела как централизованная. Это требование называют основнымпринципом создания распределенных СУБД.



Рис. 1. Топология распределенной СУБД

Централизованная база данных хранится в памяти одной вычислительной системы. Доступ к данным в этом случае обеспечивается для любого пользователя по сети. СУБД просто поддерживает распределенную обработку, но не может рассматриваться как распределенная СУБД.



Рис. 2. Топология системы с распределенной обработкой данных

По способу доступа базы данных подразделяются на базы данных с локальным доступом и базы данных с сетевым доступом.

***Архитектура СУБД: двухуровневая и трехуровневая.***

Централизованные базы данных с сетевым доступом могут иметь двухуровневую или трехуровневую архитектуру.

Двухуровневая система может представлять собой технологию файлового сервера или технологию «клиент-сервер», причем первая технология является частным случаем второй.

***Архитектуру «файл-сервер»*** еще называют технологию удаленного управления данными. Она предполагает выделение одной из машин сети в качестве центральной, которая и будет файловым сервером. На нем хранится база данных, а также файлы, обеспечивающие доступ к данным. Все другие компьютеры сети выполняют функции рабочих станций и являются, строго говоря, клиентами. Данные в соответствии с пользовательским запросом передаются на рабочую станцию блоками. На клиенте СУБД анализирует полученную информацию, и если в полученном блоке не содержится ответ на запрос, то сервер передает следующую порцию данных. Процесс передачи данных с сервера на клиент продолжается до тех пор, пока не будет выполнен запрос или пока не будут переданы все данные.



Рис. 3. Технология «файл-сервер»

В качестве примера такой технологии можно привести АБИС «Ирбис 32».

***«Клиент-серверная» архитектура*** – это модель удаленного доступа к данным. Здесь также база данных хранится на сервере. Но СУБД делится на 2 части: клиентскую и серверную. Серверная часть СУБД, т.е. ее ядро находится на сервере, тогда как на клиенте располагаются части приложения, отвечающие за ввод и отображение данных. В основе работы сервера БД лежит использование языка запросов SQL (structured query language). Клиент обращается к серверу БД с SQL-запросами. Сервер же принимает запросы, обрабатывает их и возвращает клиенту только результат обработки. Таким образом, количество передаваемой информации по сети уменьшается во много раз.



Рис. 4. Технология «клиент-сервер»

Пример «клиент-серверной» модели – АБИС «Ирбис 64».

***Трехуровневая модель*** является расширением двухуровневой модели. В ней вводится промежуточный уровень между клиентом и сервером, который называется сервером приложений. Клиентская часть взаимодействует с пользователем и содержит функции ввода и отображения данных, а также коммуникационные функции, которые обеспечивают доступ клиенту в локальную или глобальную сеть. Это может быть HTML-страница в Web-браузере, либо Windows-приложение, взаимодействующее с Web-сервисами. Сервер баз данных в этой модели только управляет информационными ресурсами БД. Вся программная логика вынесена на сервер приложений, который обеспечивает формирование запросов к базе данных, передаваемых на выполнение серверу баз данных. Сервер приложений может быть Web-сервером или специализированной программой.

Поскольку функции клиента облегчены переносом части прикладных функций на сервер, то он называется «тонким клиентом».



Рис. 5. Трехуровневая архитектура СУБД.

Примером такой модели является АБИС «Opac-Global».

***Структурные элементы базы данных: файл, запись, поле. Атрибуты, ключи.***

Логическая структура баз данных не зависит от физического представления. Любая база данных состоит из одной или нескольких подбаз, а именно файлов, таблиц, массивов. Каждая такая подбаза включает в себя агрегаты данных, т.е. записи, документы.

***Запись (документ)*** – это совокупность разнотипных и разноструктурных данных, описывающих объект реального мира. Запись состоит из полей.

***Поле*** – это именованный элементарный или составной фрагмент записи (документа), содержащий информацию об определенном аспекте объекта.

Например, база данных, состоящая из таблицы с данными о сотрудниках библиотеки. Запись – это совокупность разнотипных сведений о сотруднике, т.е. его фамилия, имя (строковые данные), дата рождения (тип даты), стаж работы (числовой тип данных), взятые вместе. Поле – это фрагмент записи, например, дата рождения, которая показывает возрастной аспект объекта (сотрудника).

Каждый объект можно охарактеризовать с помощью ***атрибутов***. Атрибуты бывают простыми и составными, однозначными и многозначными.

***Простые атрибуты*** не могут быть разбиты на более мелкие компоненты, т.е. сами являются атомарными. В нашем примере, простые атрибуты – это фамилия, имя, стаж. ***Составные атрибуты*** делятся на составляющие. Например, дату рождения можно разбить на год, месяц и число.

Если атрибут отдельного объекта может иметь только одно значение, то его называют ***однозначным***. В нашем примере к таким атрибутам относятся фамилия, имя. Если же атрибут может иметь несколько значений, то его называют ***многозначным***. Например, каждый сотрудник может иметь более одного номера телефона.

Среди атрибутов выделяют такие, с помощью которых можно идентифицировать экземпляр объекта. Такие атрибуты называют ***ключами***. Ключи могут быть простыми и составными. В нашем примере, однозначно идентифицировать сотрудника можно с помощью составного ключа, т.е. фамилии, имени и даты рождения.