

Тема 11

Базы данных

11.1. Понятие базы данных. Системы управления базами данных

Слово «данные» определяется как диалектическая составная часть информации в виде зарегистрированных сигналов. Регистрация данных может быть осуществлена любым физическим методом (механическое перемещение физических тел, изменение их формы или параметров качества поверхности, изменение электрических, магнитных, оптических характеристик, химического состава или характера химических связей, изменение состояние электронной системы и т. д.). Изначально при создании баз данных применялись следующие типы данных:

- 1) числовые (например, 17; 0,27; 2E-7);
- 2) символьные или алфавитно-цифровые (в частности, «потолок», «стол»);
- 3) даты, которые задаются с помощью специального типа «Дата» или как обычные символьные данные (например, 12.02.2005, 12/02/2005).

Позднее были определены другие типы данных, в том числе:

- 1) временные и дата-временные, которые применяются для хранения информации о времени и/или дате (например, 5.02.2005, 7:27:04, 23.02.2005 16:00);
- 2) символьные данные переменной длины, предназначенные для хранения текстовой информации большой длины;
- 3) двоичные, которые используются для хранения графических объектов, аудио– и видеоинформации, пространственной, хронологической и другой специальной информации;
- 4) гиперссылки, позволяющие хранить ссылки на различных ресурсах, располагающихся вне базы данных.

База данных – это совокупность определенным образом взаимосвязанных данных, хранящихся в памяти ЭВМ для отображения

структуры объектов и их связей в изучаемой предметной области. Она является основной формой организации хранения данных в информационных системах.

Система управления базами данных представляет собой комплекс символьных и программных средств, предназначенных для создания, ведения и организации совместного доступа к базам данных множества пользователей.

Первые СУБД были разработаны фирмами *IBM – IMS* (1968 г.) и *Software AG – ADABA* (1969 г.). В настоящий момент существует большое число различных систем управления базами данных (более нескольких тысяч), и их количество постоянно растет.

Среди основных функций СУБД (функций высшего уровня) можно выделить хранение, изменение и обработку информации, а также разработку и получение различных выходных документов.

К функциям СУБД более низкого уровня относятся:

- 1) управление данными во внешней памяти;
- 2) управление буферами ОП;
- 3) управление транзакциями;
- 4) ведение журнала изменений в базе данных;
- 5) обеспечение целостности и безопасности баз данных.

11.2. Иерархическая, сетевая и реляционная модели представления данных

Информация в базе данных некоторым образом структурирована, т. е. ее можно описать моделью представления данных (моделью данных), которые поддерживаются СУБД. Эти модели подразделяют на иерархические, сетевые и реляционные.

При использовании *иерархической модели* представления данных связи между данными можно охарактеризовать с помощью упорядоченного графа (или дерева). В программировании при описании структуры иерархической базы данных применяют тип данных «дерево».

Основными достоинствами иерархической модели данных являются:

- 1) эффективное использование памяти ЭВМ;
- 2) высокая скорость выполнения основных операций над данными;
- 3) удобство работы с иерархически упорядоченной информацией.

К недостаткам иерархической модели представления данных относятся:

- 1) громоздкость такой модели для обработки информации с достаточно сложными логическими связями;
- 2) трудность в понимании ее функционирования обычным пользователем.

Незначительное число СУБД построено на иерархической модели данных.

Сетевая модель может быть представлена как развитие и обобщение иерархической модели данных, позволяющее отображать разнообразные взаимосвязи данных в виде произвольного графа.

Достоинствами сетевой модели представления данных являются:

- 1) эффективность в использовании памяти компьютера;
- 2) высокая скорость выполнения основных операций над данными;
- 3) огромные возможности (большие, чем у иерархической модели) образования произвольных связей.

К недостаткам сетевой модели представления данных относятся:

- 1) высокая сложность и жесткость схемы базы данных, которая построена на ее основе;
- 2) трудность для понимания и выполнения обработки информации в базе данных непрофессиональным пользователем.

Системы управления базами данных, построенные на основе сетевой модели, также не получили широкого распространения на практике.

Реляционная модель представления данных была разработана сотрудником фирмы *IBM* Э. Коддом. Его модель основывается на понятии «отношения» (relation). Простейшим примером отношения служит двумерная таблица.

Достоинствами реляционной модели представления данных (по сравнению с иерархической и сетевой моделями) являются ее понятность, простота и удобство практической реализации реляционных баз данных на ЭВМ.

К недостаткам реляционной модели представления данных относятся:

- 1) отсутствие стандартных средств идентификации отдельных записей;
- 2) сложность описания иерархических и сетевых связей.

Большинство СУБД, применяемых как профессиональными, так и непрофессиональными пользователями, построены на основе реляционной модели данных (Visual FoxPro и Access фирмы *Microsoft*, Oracle фирмы *Oracle* и др.).

11.3. Постреляционная, многомерная и объектно-ориентированная модели представления данных

Постреляционная модель представления данных является расширенной версией реляционной модели данных и позволяет устранить ограничение неделимости данных, хранящихся в записях таблиц. Именно поэтому хранение данных в постреляционной модели по сравнению с реляционной считается более эффективным.

Достоинство постреляционной модели состоит в том, что она дает возможность формирования совокупности связанных реляционных таблиц через одну постреляционную таблицу, что обеспечивает высокую наглядность представления информации и эффективность ее обработки.

Недостаток такой модели заключается в сложности решения проблемы обеспечения целостности и непротиворечивости хранимых данных.

Примерами постреляционных СУБД являются системы UniVers, Buddha и Dasdb.

В 1993 г. в свет вышла статья Э. Кодда, в которой он сформулировал 12 основных требований к системам класса OLAP (On-line Analytical Processing – оперативная аналитическая обработка). Главные из описанных принципов были связаны с возможностями концептуального представления и обработки

многомерных данных. Этот момент стал отправной точкой роста интереса к многомерным моделям представления данных.

Многомерные модели – это узкоспециализированные СУБД, которые используются для интерактивной аналитической обработки информации. Многомерная организация данных отличается более высокой наглядностью и информативностью в сравнении с реляционной моделью.

Основным недостатком многомерной модели данных является ее громоздкость для решения простейших задач обычной оперативной обработки информации.

Примерами СУБД на основе таких моделей являются Ess-base фирмы *Arbor Software*, Oracle Express Server фирмы *Oracle* и др.

Объектно-ориентированные модели представления данных позволяют идентифицировать отдельные записи базы. Между записями базы данных и функциями их обработки формируются определенные взаимосвязи с помощью механизмов, похожих на соответствующие средства в объектно-ориентированных языках программирования.

Достоинствами объектно-ориентированной модели данных являются:

- 1) возможность показа информации о сложных взаимосвязях объектов;
- 2) способность идентификации отдельной записи базы данных и определения функции ее обработки.

К недостаткам объектно-ориентированной модели данных относятся:

- 1) трудность в понимании ее деятельности непрофессиональным пользователем;
- 2) неудобство обработки данных;
- 3) небольшая скорость выполнения запросов.

Среди объектно-ориентированных СУБД можно выделить системы POET фирмы *POET Software*, Versant фирмы *Versant Technologies* и др.

11.4. Классификации систем управления базами данных

Под определение СУБД может попасть любой программный продукт, способный поддерживать процессы проектирования, администрирования и

использования базы данных, поэтому была разработана классификация СУБД по видам программ:

1) полнофункциональные – самые многочисленные и мощные по своим возможностям программы, например Microsoft Access, Microsoft FoxPro, Clarion Database Developer и др.;

2) серверы баз данных – применяются для организации центров обработки данных в сетях ЭВМ. Среди них программы Microsoft SQL Server, NetWare SQL фирмы *Novell*;

3) клиенты баз данных – различные программы (полнофункциональные СУБД, электронные таблицы, текстовые процессоры и т. д.), обеспечивающие большую производительность вычислительной сети, если клиентская и серверная части базы данных будут произведены одной фирмой, но такое условие не является обязательным;

4) средства разработки программ работы с базами данных – предназначены для разработки таких программных продуктов, как клиентские программы, серверы баз данных и их отдельные приложения, а также пользовательские приложения. Средствами разработки пользовательских приложений служат системы программирования, библиотеки программ для различных языков программирования, пакеты автоматизации разработок. Самыми часто используемыми средствами разработки пользовательских приложений являются инструментальные средства Delphi фирмы *Borland* и Visual Basic фирмы *Microsoft*.

По виду применения СУБД подразделяются на персональные и многопользовательские.

Персональные СУБД (например, Visual FoxPro, Paradox, Access) используются при проектировании персональных баз данных и недорогих приложений, работающих с ними, которые, в свою очередь, могут применяться в качестве клиентской части многопользовательской СУБД.

Многопользовательские СУБД (например, Oracle и Informix) состоят из сервера баз данных и клиентской части и способны работать с различными типами ЭВМ и ОС различных фирм-производителей.

Чаще всего информационные системы строятся на основе архитектуры *клиент-сервер*, в которую входят вычислительная сеть и распределенная база данных. *Вычислительная сеть* используется для организации научной работы на ПК и в сетях. *Распределенная база данных* состоит из многопользовательской базы данных, размещенной на компьютере-сервере, и персональной базы данных, находящейся на рабочих станциях. Сервер базы данных осуществляет выполнение основного объема обработки данных.

11.5. Языки доступа к базам данных

Выделяют два типа языков доступа к базам данных:

- 1) язык описания данных – высокоуровневый язык, предназначенный для описания логической структуры данных;
- 2) язык манипулирования данными – совокупность конструкций, обеспечивающих выполнение базовых операций по работе с данными: ввод, модификацию и выборку данных по запросам.

Самыми распространенными языками доступа являются два стандартизированных языка:

- 1) QBE (Query by Example) – язык запросов по образцу, характеризующийся свойствами языка манипулирования данными;
- 2) SQL (Structured Query Language) – структурированный язык запросов, состоящий из свойств языков обоих типов.

Язык QBE был разработан на основе реляционного исчисления с переменными-доменами. Он помогает формировать сложные запросы к базе данных с применением заполнения предлагаемой системой управления базой данных запросной формы. Любая из реляционных СУБД обладает своим вариантом языка QBE. Достоинствами подобного способа задания запросов к БД являются:

- 1) высокая наглядность;
- 2) отсутствие необходимости указания алгоритма выполнения операции.

Структурированный язык запросов SQL появился на основе реляционного исчисления с переменными кортежами. Было разработано несколько стандартов данного языка, самыми известными из которых являются SQL-89 и SQL-92. Язык SQL применяется для выполнения операций над таблицами и данными, которые содержатся в этих таблицах, и некоторых сопутствующих операций. Он не применяется как отдельный язык и чаще всего служит частью встроенного языка программирования СУБД (например, FoxPro СУБД Visual FoxPro, ObjectPAL СУБД Paradox, Visual Basic for Applications СУБД Access).

Язык SQL ориентирован только на доступ к данным, поэтому его относят к средствам разработки программ и называют *встроенным*. Выделяют два основных метода использования встроенного SQL:

1) *статический* – характеризуется тем, что в текст программы помещены вызовы функций языка SQL, жестко включаемые в выполняемый модуль после компиляции. Изменения в вызываемых функциях могут производиться на уровне отдельных параметров вызовов с помощью переменных языка программирования;

2) *динамический* – отличается динамическим построением вызовов SQL-функций и интерпретацией этих вызовов при выполнении программы. Чаще всего применяется в тех случаях, когда в приложении вид SQL-вызова заранее неизвестен, и он выстраивается в диалоге с пользователем.

11.6. Базы данных в сети Интернет

Основой публикации баз данных во всемирной сети Интернет является простое расположение информации из баз данных на Web-страницах сети.

Публикация баз данных в Интернет призвана решить целый ряд задач, среди которых можно выделить следующие:

1) организация взаимосвязи систем управления базами данных, которые работают на различных платформах;

2) построение информационных систем в сети Интернет на основе многоуровневой архитектуры баз данных;

3) построение локальных Интранет-сетей с помощью технологий публикации баз данных в Интернет;

4) применение в Интернет информации из имеющихся локальных сетевых баз данных;

5) использование баз данных для упорядочения информации, представленной в сети Интернет;

6) использование обозревателя Web как доступной клиентской программы для доступа к базам данных в Интернет.

Для публикации баз данных на Web-страницах используются два основных способа формирования Web-страниц, содержащих информацию из баз данных:

1) статическая публикация – Web-страницы создают и хранят на Web-сервере до момента поступления запроса пользователя на их получение (в виде файлов на жестком диске в формате Web-документа). Данный способ используется при публикации информации, редко обновляемой в базе данных. Основными достоинствами такой организации публикации баз данных в сети Интернет является ускоренный доступ к Web-документам, которые содержат информацию из баз данных, и уменьшение нагрузки на сервер при обработке клиентских запросов;

2) динамическая публикация – Web-страницы создают при поступлении запроса пользователя на сервер. Сервер пересылает запрос на генерацию таких страниц программе – расширению сервера, формирующей требуемый документ. После этого сервер отсылает готовые Web-страницы обратно обозревателю. Данный способ формирования Web-страниц используется тогда, когда содержимое базы данных часто обновляется, например в режиме реального времени. Данным способом публикуется информация из баз данных

для интернет-магазинов и информационных систем. Динамические страницы образуются с помощью различных средств и технологий, например ASP (Active Server Page – активная серверная страница), PHP (Personal Home Page tools – средства персональных домашних страниц).

Среди программных средств, позволяющих получить информацию из сети Интернет, выделились Web-приложения (интернет-приложения), которые представляют собой набор Web-страниц, сценариев и других программных средств, расположенных на одном или нескольких компьютерах и предназначенных для выполнения прикладной задачи. Приложения, которые публикуют базы данных в сети Интернет, выделены в отдельный класс Web-приложений.