**ГЛАВА 3
БАЗЫ ДАННЫХ
§13. Основные понятия**

Первоначально вычислительные машины разрабатывались для выполнения расчётных задач — решения систем уравнений, определения траекторий полёта снарядов и спутников и т. п. Однако сейчас значительную часть времени компьютеры заняты обработкой нечисловых данных — текстов, изображений, звука и видео. В нашу жизнь вошли информационные системы, с помощью которых мы узнаём прогноз погоды и расписание поездов, определяем маршруты путешествий, заказываем билеты на самолёты, бронируем номера в гостиницах и т. п.

 **Информационная система (ИС) в широком смысле — это аппаратные и программные средства, предназначенные для того, чтобы своевременно обеспечить пользователей нужной информацией.**

У информационной системы две основные задачи — она должна обеспечивать:

• хранение данных;
• доступ к данным, т. е. возможность искать и изменять данные.

Массивы данных, с которыми работают информационные системы, обычно имеют большой объём (нередко несколько гигабайтов и даже терабайтов) и размещаются во внешней памяти компьютера. Данные хранятся в таком виде, чтобы их было легко искать и изменять. Такие наборы данных называются базами данных.

 **База данных (БД) — это специальным образом организованная совокупность 1 данных о некоторой предметной области, хранящаяся во внешней памяти компьютера.**

**1** Термин «совокупность» обозначает множество элементов, обладающих общими свойствами.

Данные сами по себе бесполезны, если мы не умеем с ними работать. Поэтому необходимо специальное программное обеспечение, которое позволяет искать и изменять данные.

 **Система управления базой данных (СУБД) — это программные средства, которые позволяют выполнять все необходимые операции с базой данных.**

Хотя термины «база данных» и СУБД обозначают различные понятия, они неразрывно связаны: свойства базы данных определяются СУБД, которая ею управляет, и наоборот. Комплекс "БД + СУБД" называется системой базы данных (англ. database system) или информационной системой в узком смысле.

Вы знаете, что данные в компьютерном формате — это двоичные коды, которые могут обозначать всё, что угодно. Поэтому СУБД должна «знать» формат файлов (что где записано). В первых информационных системах каждая база данных имела свой собственный формат, который придумывал её автор. Это очень неудобно, потому что для каждой ИС нужно разрабатывать специальную программу для работы с базой данных. Более того, при любых изменениях в БД (например, при увеличении размера данных) надо переделывать все работающие с ней программы и переводить старые данные в новый формат.

Поэтому постепенно перешли к использованию специальных программ (они и получили название СУБД), которые занимались только хранением и обработкой данных, независимо от их содержания, и могли применяться в самых разных задачах. При этом вместе с основными данными нужно хранить описание их структуры — так называемые метаданные, «сведения о данных», которые использует СУБД для обращения к данным.

СУБД решают все задачи, связанные с управлением данными, в том числе:

• поиск данных;
• редактирование данных;
• выполнение несложных расчётов;
• обеспечение целостности (корректности, непротиворечивости) данных;
• восстановление данных после сбоев.

Как правило, пользователь работает с СУБД не напрямую, а через прикладную программу, в которой предусмотрен удобный ввод данных и оформление результатов (рис. 3.1).



***Рис. 3.1***

Иногда функции СУБД и прикладной программы объединяются в одной программе (например, в OpenOffice.org Base или Microsoft Access).

В простейшем случае база данных и СУБД находятся на одном компьютере. Такая информационная система называется локальной, с ней работает один пользователь. В современных браузерах (Google Chrome, Safari, Mozilla Firefox) есть встроенные средства, позволяющие создавать и использовать локальные ИС.

Преимущество локальных ИС — автономность, т. е. независимость от работы локальных и глобальных сетей. Их недостатки проявляются тогда, когда с БД должны работать несколько пользователей:

• базу данных нужно обновлять на каждом компьютере;
• невозможно «стыковать» изменения, вносимые пользователями.

В локальных ИС разработчики иногда применяют свои собственные форматы хранения данных, однако в этом случае теряется переносимость — возможность использования базы данных в других информационных системах.

Как правило, в современных информационных системах используют удалённые базы данных, расположенные на серверах (специально выделенных компьютерах) локальной или глобальной сети. В этом случае несколько пользователей могут одновременно работать с базой и вносить в неё изменения.

СУБД, работающие с удалёнными базами данных, можно разделить на два типа по способу работы с файлами:

• файл-серверные СУБД;
• клиент-серверные СУБД.

Файл-серверные СУБД (например, Microsoft Access) работают на компьютерах пользователей (они называются рабочими станциями) (рис. 3.2). Это значит, что сервер только хранит файлы, но не участвует в обработке данных. Когда пользователь вносит изменения в базу, СУБД с его рабочей станции блокирует файлы на сервере, чтобы их не могли изменить другие пользователи.



***Рис. 3.2***

При большом количестве пользователей проявляются недостатки файл-серверных ИС:

• обработку данных выполняют СУБД на рабочих станциях, поэтому компьютеры пользователей должны быть достаточно мощными;
• при поиске данных вся БД копируется по сети на компьютер пользователя, это создает значительную лишнюю нагрузку на сеть;
• слабая защита от неправомерного доступа к данным (защита устанавливается на рабочих станциях, а не в едином центре);
• ненадёжность при большом количестве пользователей, особенно если они вносят изменения в базу данных.

Чтобы избавиться от этих недостатков, нужно переместить СУБД на сервер.

Клиент-серверная СУБД (рис. 3.3) расположена на том же компьютере, где находится база данных. Она полностью берёт на себя всю работу с данными, т. е. читать и изменять данные в базе можно только с помощью этой СУБД.



***Рис. 3.3***

На компьютере пользователя работает прикладная программа- клиент, которая по сети обращается к СУБД для выполнения операций с данными.

Задачи клиента:
• отправить серверу команду (запрос) на специальном языке;
• получив ответ сервера, вывести результат на экран пользователя или на печать.

В современных клиент-серверных СУБД для управления данными чаще всего используют язык SQL (англ. Structured Query Language — язык структурных запросов). Он содержит все команды, необходимые работы с данными, включая получение нужной информации, создание и изменение базы данных.

Задачи сервера:

• ожидать запросы клиентов по сети;
• при поступлении запроса поставить его в очередь на выполнение;
• выполнить запрос;
• передать результаты клиенту.

Отметим, что серверная и клиентская части могут быть установлены на одном компьютере.

Самые известные среди коммерческих клиент-серверных СУБД — Microsoft SQL Server и Oracle. Существуют бесплатные клиент-серверные СУБД: Firebird (www.firebirdsql.org), PostgreSQL (www.postgresql.org), X\MySQL (www.mysql.com). Свободная СУБД MySQL часто используется для управления базами данных на небольших веб-сайтах.

Достоинства клиент-серверных СУБД:

• основная обработка данных выполняется на сервере, поэтому рабочие станции могут быть маломощными;
• проще модернизация (достаточно увеличить мощность сервера);
• надёжная защита данных (устанавливается на сервере);
• снижается нагрузка на сеть, так как передаются только нужные данные (запросы и результаты выполнения запросов);
• надёжная работа при большом количестве пользователей (запросы ставятся в очередь).

Их недостатки — повышенные требования к мощности сервера и высокая стоимость коммерческих СУБД (Microsoft SQL Server и Oracle).

Многие современные информационные системы (например, поисковые системы в Интернете) работают с огромными объёмами данных, которые невозможно разместить на одном компьютере. Поэтому появились распределённые базы данных, расположенные на множестве компьютеров, и соответствующие СУБД для управления ими. Пользователь работает с распределённой базой данных точно так же, как и с обычной (нераспределённой). Главная проблема в этой области — обеспечение целостности и непротиворечивости данных, особенно при разрыве связи между компьютерами.

Сейчас в базах данных нередко хранится очень важная информация, прежде всего финансовая. Поэтому необходимо обеспечить максимально возможную защиту данных от различных сбоев (например, при отключении питания).

Предположим, что в банке нужно перевести 100 000 рублей со счёта 12345 на счёт 54321. Эта банковская операция включает несколько действий с базой данных:

1) прочитать сумму на счету 12345;
2) уменьшить её на 100 000 рублей и записать результат обратно;
3) прочитать сумму на счету 54321;
4) увеличить её на 100 000 рублей и записать результат обратно.

Представьте себе, что случится, если после выполнения первых двух дей
ствий произойдет сбой питания и действия 3 и 4 не будут выполнены. Ответ ясен — первый клиент просто потеряет 100 000 рублей со своего счёта, т. е. данные станут некорректными.

Чтобы этого не произошло, либо все шаги операции перевода денег должны быть выполнены, либо ни один шаг не должен быть выполнен. Кроме того, между отдельными шагами операции никакие другие действия не должны выполняться. Такие сложные многошаговые операции называются транзакциями (от англ. transaction — сделка).

 **Транзакция — это группа операций, которая представляет собой одно законченное действие. Транзакция должна быть выполнена целиком или не выполнена вообще.**

Как же обеспечить выполнение транзакций? Для этого часто используется журналирование по тому же принципу, что и в файловых системах. Перед внесением изменений в базу данных СУБД создаёт копии всех данных, которые будут изменяться, и записывает в специальный файл (журнал) все операции, которые нужно выполнить. Затем эти операции выполняются фактически и, если всё завершено успешно, запись удаляется из журнала. Если произошёл сбой, в журнале будет найдена информация о тех операциях, которые уже были завершены, и база данных восстанавливается в исходное состояние (транзакция не выполнена).

Некоторые СУБД, например Firebird, не используют журнал, а при внесении изменений создают в базе данных новые записи, которые отмечаются как «рабочие» только тогда, когда транзакция завершена.

1. Что такое информационная система? Из каких компонентов она состоит?
2. Что такое база данных? Какими свойствами она должна обладать?
3. Является ли базой данных бумажная картотека в библиотеке? Ответ обоснуйте.
4. Какие функции выполняет СУБД?
5. Почему произошёл переход от множества разнообразных форматов хранения данных к использованию универсальных СУБД? Приведите примеры и обоснуйте.
6. Что такое метаданные?
7. Объясните схему работы пользователя с базой данных.
8. Назовите достоинства и недостатки локальных ИС.
9. Назовите достоинства и недостатки файл-серверных СУБД.
10. В каких ситуациях вы могли бы рекомендовать использование файл-серверных СУБД? Ответ обоснуйте. Подготовьте сообщение.
11. Назовите достоинства и недостатки клиент-серверных СУБД.
12. Как разделяются функции между клиентской и серверной программами?
13. Что такое SQL?
14. Что такое распределённые базы данных?
15. Что такое транзакция?
16. Как обеспечивается защита данных в случае сбоев при использовании механизма транзакций?

**Подготовьте сообщение**

а) «Информационные системы вокруг нас»
б) «Технология клиент — сервер»
в) «Бесплатные СУБД»
г) «Коммерческие и бесплатные СУБД: плюсы и минусы»