

Лекция 1. Основные понятия. База данных, банк данных, информационная система

1.1 Базовые определения	1
1.2 Компоненты информационной системы	2
1.3 К чему следует стремиться	4
1.4 Классификация банков данных	6

1.1 Базовые определения

Название курса говорит о том, что предметом нашего изучения будут большие объемы данных, и программные системы для работы с данными.

Примечание: здесь и далее по тексту новые понятия будут определяться на русском и английском языках, поскольку большую роль в становлении теории баз данных сыграли именно англоязычные научные школы

Прежде всего, давайте разберемся с понятиями «данные» и «информация».

Бытуют две распространенные тенденции в трактовке этих понятий. Первая – общепринятая в повседневной жизни – «информация, данные – одно и то же». Так мы бы говорили по телефону: «У меня есть информация (данные) по этому поводу»

Вторая тенденция состоит в четком разделении сфер употребления этих понятий. Проиллюстрируем этот факт на примере:

Дана фраза: «Его рост – 180 сантиметров»

Очевидно, число «180» - данные

Остальная часть фразы подсказывает нам смысл этого числа – «значение роста некоторого мужчины в см» - информация, т.е. смысл данного

Таким образом,

Данные (data) - дискретные факты относительно явлений реального мира.

Информация (information) - приращение знаний о реальном мире, которое может быть получено на основе данных.

Представим теперь, что есть большие и очень большие объемы данных. Для нормальной работы с ними, а именно, для поиска и анализа, необходимо придумать методику хранения, которая позволила бы оперировать данными в разумные интервалы времени. Понятно, что способ хранения данных «в одной куче» не решает проблему эффективной работы. Поэтому появилось следующее понятие, которое будет центральным в данном курсе лекций, а именно - «база данных»

База данных, БД (database)- не что иное как компьютеризованная система хранения данных.

Более строгое и полное определение базы данных дает Джордж Мартин в книге [9]

База данных — совокупность взаимосвязанных, хранящихся вместе данных, при наличии такой минимальной избыточности, которая допускает их использование оптимальным образом для одного или нескольких приложений.

Обычно базы данных создаются не для решения какой-либо одной задачи для одного пользователя, а создаются как хранилища данных многоцелевого использования. Однако помимо хранения данных пользователь может захотеть (и чаще всего хочет), обработав эти данные, получить интересующую его информацию.

Поэтому, обычно любая База Данных - лишь составная часть некоторой Информационной Системы (ИС).

Существует несколько названий для информационных систем:

Информационная система, Банк Данных (data bank, database system)

Соответствие между ними такое:

ИС \approx Банк Данных = Data Bank \approx database system

Банк данных — это система специальным образом организованных данных (Система Баз Данных), программное обеспечение для управления данными (Система Управления Базами Данных; средства проектирования базы данных; генераторы отчетов), оборудование, на котором размещены данные и на котором работает персонал, и собственно, персонал.

Банк данных = Система БД + СУБД + оборудование + персонал

(Database System = Databases + DataBase Management System + hardware + users)

На рис.1.1 представлены компоненты банка данных:

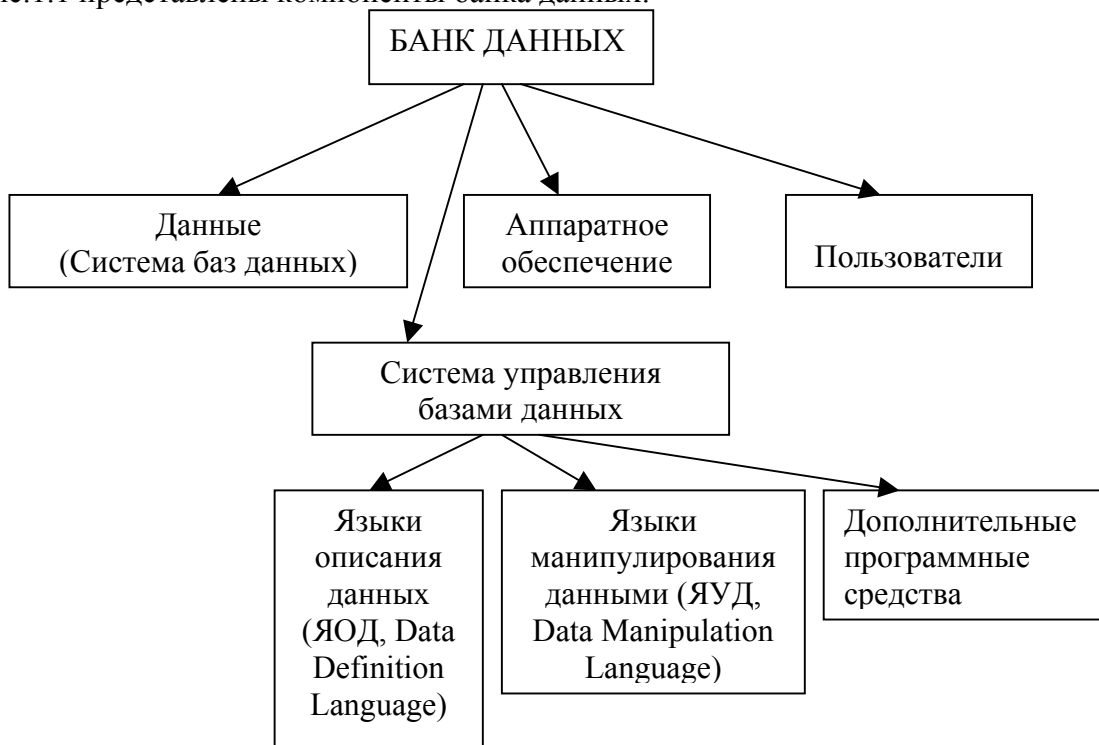


Рисунок 1.1 Компоненты банка данных.

Примеры информационных систем:

1. **Продажа авиабилетов.** Основные характеристики этой ИС: постоянные обновления данных, части банка данных находятся в разных городах, требуется постоянный круглосуточный доступ
2. **Электронная библиотека.** Основные характеристики: разнообразные типы данных (текст, графика, фото, видео, музыка), связанные огромным количеством разнообразных связей, требуется "изоциренный" поиск.

1.2 Компоненты информационной системы

Рассмотрим теперь каждую компоненту ИС отдельно.

1) Система баз данных

В общем случае *данные в БД являются интегрированными и общими*.

- Под понятием “*интегрированные*“ имеется в виду возможность представить БД как объединение нескольких отдельных *файлов данных*, полностью или частично не перекрывающихся.
- под понятием “*общие*“ имеется в виду возможность использования отдельных частей данных в БД различными пользователями, т.е. несколько пользователей могут работать с одной и той же областью данных (причем каждый по-своему, и даже одновременно).

Поскольку возможности аппаратного обеспечения бывают разными, различают *одно- и многопользовательские СБД*.

Поэтому

Система баз данных – один или несколько файлов данных, хранящихся на одном или нескольких (возможно, разнесенных территориально) накопителях, допускающих (возможно, одновременную) работу одного или нескольких пользователей.

2) Оборудование

К необходимому оборудованию относят:

- устройства хранения информации (накопители);
- процессоры и память (ОЗУ).

3) Программное обеспечение

Между собственно данными, которые сохраняются в СБД, и пользователями системы располагается уровень программного обеспечения – диспетчер баз данных, который еще называют «система управления базами данных», или СУБД (database management system, DBMS).

Основная функция СУБД – предоставление пользователю БД возможности работать с данными, не вникая в детали на уровне аппаратного обеспечения. Более подробно функции СУБД рассмотрены в этой лекции ниже.

4) Пользователи

В применении к банкам данных различают три группы пользователей:

1. *Прикладные программисты*, которые пишут прикладные программы, использующие базу данных, используя языки программирования высокого уровня (C, Pascal,...);
2. *Пользователи*, которые получили доступ к СБД непосредственно с терминала или рабочей станции. Язык обращения к СБД – тот, который поддерживает СУБД (например, для доступа к реляционным базам данных используется язык SQL). Эти языки обычно называются языками запросов. Пользователь также имеет возможность получить доступ к БД через интерфейс, основанный на меню и формах.
3. *Администраторы БД (АБД)*, которые создают саму СБД, и занимаются поддержкой ее функционирования.

Возможно, у вас возникли вопросы вроде: почему база данных, а не обычные бумажные картотеки? Почему, наконец, база данных, а не, например, электронные таблицы Excel? Зачем нужен этот посредник в виде СУБД?

По сравнению с бумажным способом хранения компьютеризованный способ более интересен по таким причинам:

- Компактность: Все-таки один компьютер – это меньше, чем ящик с папками.
- Скорость обработки и поиска
- Низкие трудозатраты при работе
- Применимость: Одни и те же данные (например, справочники городов, улиц, списки сотрудников предприятия, и т.п.) могут быть полезны при выполнении запросов из разных информационных систем.

Возможно, эти преимущества лучше проявляются в многопользовательской системе. Более того, в последнем случае появляется еще одно дополнительное преимущество:

- Возможность централизованного управления данными предприятия: Это освобождает всех пользователей кроме администраторов БД от функций управления.

А поскольку функции создания и ведения БД, выдачи результатов запросов являются универсальными, общими для разных задач, соответственно это и привело к идее создания пакета прикладных программ для обработки массивов данных, или СУБД.

1.3 К чему следует стремиться

Какой должна быть идеальная база данных?

Для ответа на этот вопрос обратимся к истории. Из истории известно, что впервые понятия базы данных и банка данных появились в конце 60х г.г. в работах *Эдгара Кодда*, сотрудника фирмы IBM. До этого в сфере обработки данных говорили о файлах данных и наборах данных. Первые банки данных развились на основании файловых систем. Файловые системы поддерживают только хранение и, частично, описание структуры данных (на уровне создания разных каталогов для разных данных). Понятия о языке запросов не было. Контроль целостности данных и разграничение доступа к ним также отсутствовал.

Те хранилища данных, которые были созданы до этого времени, были жестко связаны с программами обработки данных, с аппаратным обеспечением, и изменения требований к выдаваемой информации, изменения в структуре самих данных и т.п. влекли за собой огромные издержки по корректному преобразованию всего хранилища, программного обеспечения.

Решению этой проблемы безболезненной модификации структуры систем БД были посвящены усилия многих ученых и исследовательских групп фирм производителей ЭВМ, в частности:

- CODASYL,
- X3 ANSI/SPARK/DBMS - Американский Научный Институт Стандартов,
- Association for Computer Machinery (ACM)

Одним из результатов совместной работы этих учреждений были требования к *идеальному банку данных*.

Банк данных должен обеспечивать:

1. Адекватное целостное представление определенной части реального мира
2. Оптимальную избыточность данных за счет интегрированного хранения.
Причина: информационные потребности пользователей банка данных существенно пересекаются, что делает целесообразным использование общих данных и обеспечение доступа к ним различных пользователей
3. Конфиденциальность, разграничение прав доступа и разделение полномочий.

4. Надежность, устойчивость к повреждениям, возможность полного восстановления данных, приемлемые соотношения производительности.
5. Дружелюбный интерфейс и легкость в обучении пользователя.
6. Независимость программ от данных.
Цель: любое изменение в структуре данных или в методе доступа к ним влечет за собой минимальные изменения в программном коде.

Комментарий к требованию независимости данных от программного кода:

Приложения, реализованные на старых системах, во многом зависели от данных, т.е. способ организации данных в памяти и способ доступа к ним диктуются требованиями приложения. Сведения об организации данных и способе доступа были встроены в логику и код приложения, а значит, невозможно было изменить структуру хранения данных или метод доступа без того, чтобы не изменить приложения. Следовательно, существует, по крайней мере, две причины поддержки независимости программного кода:

1. Для разных приложений могут потребоваться разные представления одних и тех же данных.
2. Администратор БД должен иметь возможность менять структуру данных.

Наряду с требованиями к идеальной СБД были получены требования и к идеальной системе управления базами данных:

СУБД должна обеспечивать:

1. Определение данных (с помощью Языка Определения Данных, DDL).
2. Управление данными и их обработка (с помощью Языка Манипулирования Данными, DML).
3. Поддержка целостности данных, т.е. соответствие данных текущему состоянию некоторой части реального мира.
4. Обеспечение безопасности, т.е. разграничение прав доступа пользователей к данным.
5. Восстановление данных и резервное дублирование.
6. Предоставление, создание и ведение словаря данных, который сохраняет метаданные, или описание структуры базы данных (в виде схемы)

Укрупненная архитектура СУБД в общем виде показана на рис. 1.2.

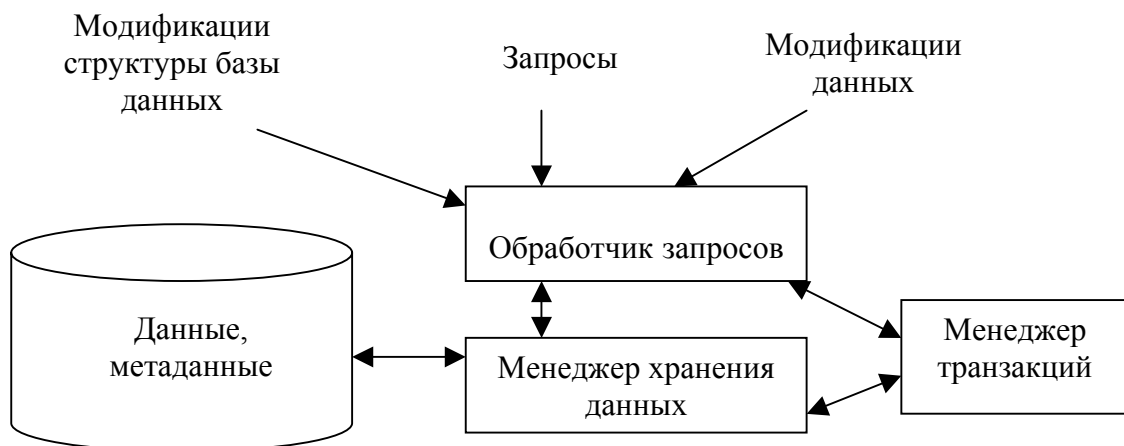


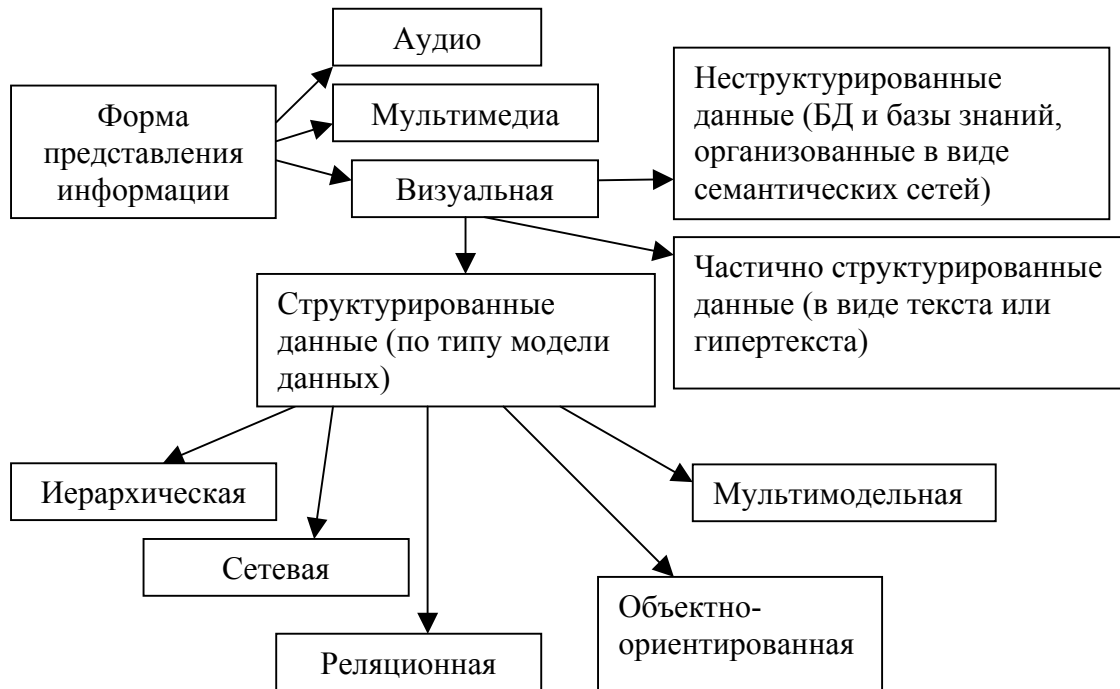
Рисунок 1.2 Архитектура СУБД

Остается отметить, что в настоящее время не существует ни одной СУБД, позволяющей создавать «идеальные» системы баз данных произвольных масштабов.

4 Классификация банков данных

Способов классификации много, поэтому приводятся лишь наиболее очевидные.

1) По форме представления информации



2) По типу хранимой информации



3) По характеру организации хранения

