




БАЗЫ ДАННЫХ



Цель курса:


**Ознакомление с технологией
проектирования баз данных и
обработки данных с
использованием средств,
предоставляемых
современными СУБД**





Информационные процессы


процессы создания, сбора, обработки, накопления, хранения, поиска, распространения и потребления информации






Информационное обеспечение

совокупность процессов сбора, обработки, хранения, анализа и выдачи информации, необходимой для обеспечения управленческой деятельности и технологических процессов






**Законодательные акты,
регламентирующие информационные
процессы и технологии:**

Федеральный закон от 27 июля 2006 г.

N 149-ФЗ


**"Об информации, информационных
технологиях и о защите информации"**





Нормативно-правовая трактовка


Информация - сведения (сообщения, данные) независимо от формы их представления





Нормативно-правовая трактовка


Информационные технологии - процессы, методы поиска, сбора, хранения, обработки, предоставления, распространения информации и способы осуществления таких процессов и методов;





Нормативно-правовая трактовка

Информационная система - совокупность содержащейся в базах данных информации и обеспечивающих ее обработку информационных технологий и технических средств;





Информация:

Это изменение объема и структуры знания о некоторой предметной области (лица, предметы, факты, события, явления, процессы) воспринимающей системой (человек, организационная структура, автоматизированная информационная система) независимо от формы и способа представления знания.




Данные :

Это информация, отражающая определенное состояние некоторой предметной области в конкретной форме представления и содержащая лишь наиболее существенные с точки зрения целей и задач сбора и обработки информации элементы образа отражаемого фрагмента действительности



Формы представления данных:

- Структурированная
 - Неструктурированная
- 

Формы представления данных:


Пример *неструктурированной* формы:

- связный текст (документ на естественном языке — на литературном, официально-деловом и т. д.);
- графические данные в виде фотографий, картинок и прочих неструктурированных изображений.

Формы представления данных:

Примеры *структурированной* формы данных:

- анкеты;
- таблицы;
- графические данные в виде чертежей, схем, диаграмм




Информационные системы
разделяются на:

- *Фактографические*
- *Документальные*
- *Геоинформационные*



Фактографические ИС


Накапливают и хранят данные и виде множества экземпляров одного или нескольких типов структурных элементов (*информационных объектов*)





Документальные ИС


Единичным элементом информации является нерасчлененный на более мелкие элементы документ






Геоинформационные ИС

Данные организованы в виде отдельных информационных объектов (с определенным набором реквизитов), привязанных к общей электронной топографической основе (электронной карте)






Информационным ядром подсистемы представления и обработки информации фактографических АИС, или, говоря иначе, внутренним носителем знаний о предметной области является ***база данных*** (БД)



Недостатки традиционных файловых систем:

- Избыточность данных
 - Слабый контроль данных
 - Недостаточные возможности управления данными
 - Большие затраты труда программиста
- 

Избыточность данных

Текущие счета	<ul style="list-style-type: none">Имя клиента – CNAME char(25) Иванов Иван Иванович
Сберегательные счета	<ul style="list-style-type: none">Имя клиента – SNAME char(20) Иванов Иван Иванович
Кредитные счета	<ul style="list-style-type: none">Имя клиента – INAME char(25) Иванов Иван Иванович

Избыточность данных

Изменили 11.01.2012

Текущие
счета

- Имя клиента – CNAME char(25)
Иванов Петр Иванович

Сберегательные
счета

- Имя клиента – SNAME char(20)
Иванов Иван Иванович

Кредитные
счета

- Имя клиента – INAME char(25)
Иванов Иван Иванович



Избыточность данных



Слабый контроль данных

Текущие
счета

- Имя владельца счета – CNAME char(25)
Иванов Петр Иванович

Сберегательные
счета

- Имя клиента – SNAME char(20)
Иванов Пётр Иванович

Кредитные
счета

- Имя клиента – INAME char(25)
Иванов Иван Иванович

Недостаточные возможности управления данными



Большие затраты труда программиста




БАЗА ДАННЫХ



База данных — это совокупность данных, организованная с определенной целью



Классификация БД по объему

- БД уровня предприятия
 - БД уровня подразделения
 - БД уровня рабочей группы
- 

ПРОГРАММНОЕ ОБЕСПЕЧЕНИЕ



Система управления базой данных (СУБД)

СУБД – это программный комплекс, обеспечивающий функционирование БД.

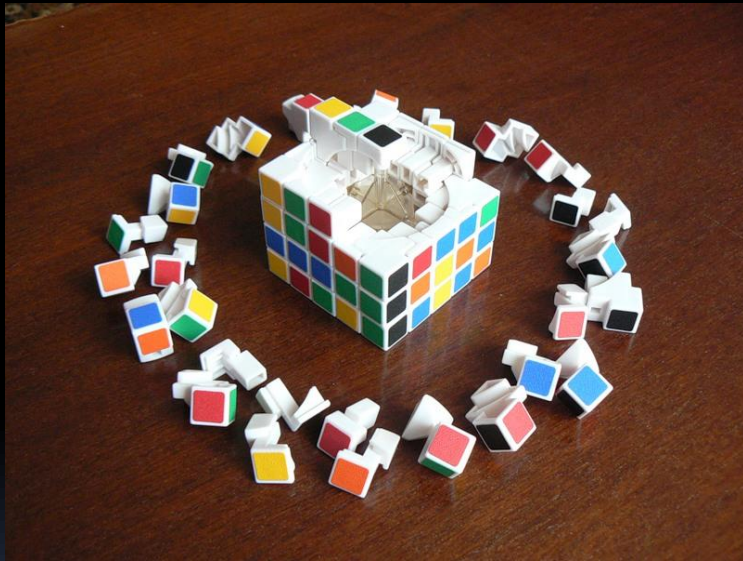
Она отвечает за их **сохранность, безопасность, целостность, взаимное соответствие** и обеспечивает **доступ пользователей к данным**.

Безопасность



1. Обязательная регистрация и последующее распознавание пользователей до предоставления им доступа.
2. Разделение пользователей на группы с различными привилегиями доступа.

Целостность данных



Целостность - взаимная согласованность отдельных фрагментов данных и их корректность.

Согласованность означает, что все порции данных в БД должны быть единообразно смоделированы и включены в систему. Квалифицировать данные как **корректные** можно в том случае, если они достоверны, точны и значимы.

Взаимное соответствие




Взаимное соответствие или **непротиворечивость** - отсутствие в базах данных либо частях одной и той же базы сведений, которые бы противоречили друг другу.

Общение пользователя с БД





Преимущества СУБД

- Контроль за избыточностью данных
 - Непротиворечивость данных
 - Больше полезной информации при том же объеме хранимых данных
 - Совместное использование данных
 - Поддержка целостности данных
 - Повышенная безопасность
 - Применение стандартов
 - Повышение эффективности с ростом масштабов системы
 - Возможность нахождения компромисса при противоречивых требованиях
 - Повышение доступности данных и их готовности к работе
 - Улучшение показателей производительности
 - Упрощение сопровождения системы за счет независимости от данных
 - Улучшенное управление параллельной работой
 - Развитые службы резервного копирования и восстановления
- 


Недостатки СУБД

- Сложность
- Размер
- Стоимость СУБД
- Дополнительные затраты на аппаратное обеспечение
- Затраты на преобразование
- Производительность
- Более серьезные последствия при выходе системы из строя



Система базы данных


СОСТОИТ ИЗ

- базы данных
 - СУБД
 - соответствующего оборудования
 - людей
- 



МОДЕЛИ ДАННЫХ НА ОСНОВЕ ЗАПИСЕЙ

В модели на основе записей база данных состоит из нескольких записей фиксированного формата, которые могут иметь разные типы. Каждый тип записи определяет фиксированное количество полей, каждое из которых имеет фиксированную длину.



МОДЕЛИ ДАННЫХ НА ОСНОВЕ ЗАПИСЕЙ

Существуют три основных типа логических моделей данных на основе записей:

- *реляционная модель данных* (relational data model)
- *сетевая модель данных* (network data model)
- *иерархическая модель данных* (hierarchical data model)

МОДЕЛИ ДАННЫХ НА ОСНОВЕ ЗАПИСЕЙ

Реляционная модель данных основана на понятии *математических отношений*. В реляционной модели данные и связи представлены в виде таблиц, каждая из которых имеет несколько столбцов с уникальными именами.

МОДЕЛИ ДАННЫХ НА ОСНОВЕ ЗАПИСЕЙ

В **сетевой** модели данные представлены в виде коллекций **записей**, а связи — в виде **наборов**. В отличие от реляционной модели, связи здесь явным образом моделируются наборами, которые реализуются с помощью указателей. Сетевую модель можно представить как граф с записями в виде узлов графа и наборами в виде его *ребер*.


МОДЕЛИ ДАННЫХ НА ОСНОВЕ ЗАПИСЕЙ

Иерархическая модель является ограниченным подтипом сетевой модели. В ней данные также представлены как коллекции *записей*, а связи — как *наборы*. Однако в иерархической модели узел может иметь только одного родителя. Иерархическая модель может быть представлена как древовидный граф с записями в виде узлов (которые также называются *сегментами*) и множествами в виде ребер.

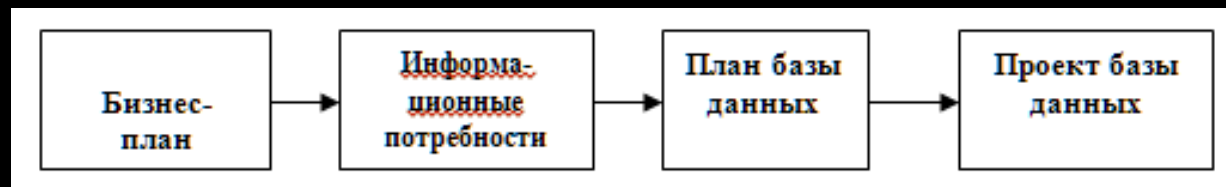


Планирование базы данных

стратегическая попытка определить
информационные потребности организации
на продолжительный период времени в
будущем



Планирование базы данных



Планирование базы данных

Преимущества:

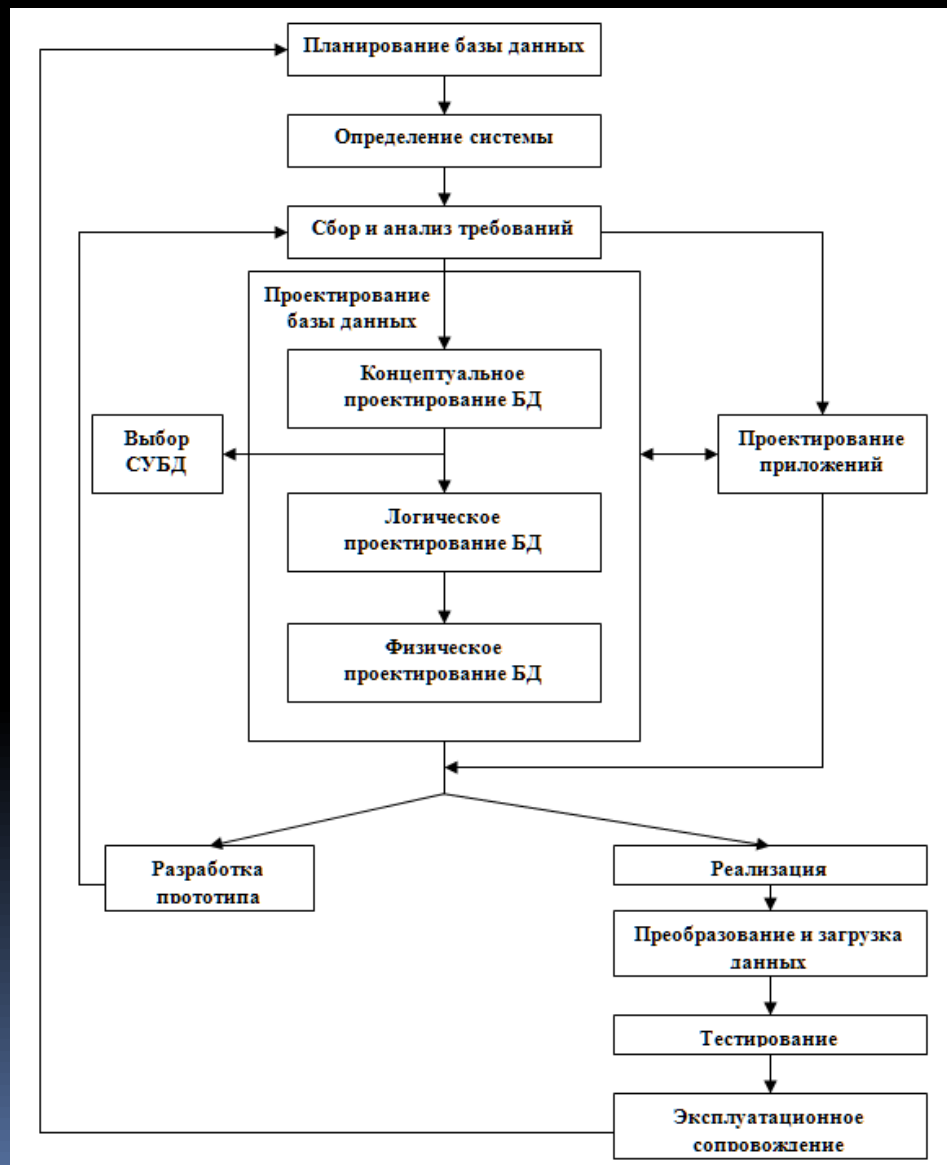
- отражает текущее понимание менеджерами информационных ресурсов;
- определяет и уточняет требования к ресурсам, помогая убедиться в том, что ресурсы будут доступны;
- определяет возможности эффективного управления ресурсами, включая сотрудничество между разными отделами организации;
- определяет планы действий для достижения намеченных целей;
- может обеспечить мощный стимул и понимание направления развития работниками всех уровней, сосредоточить их усилия, повысить производительность и заставить их в полной мере почувствовать себя частью предприятия;

Планирование базы данных

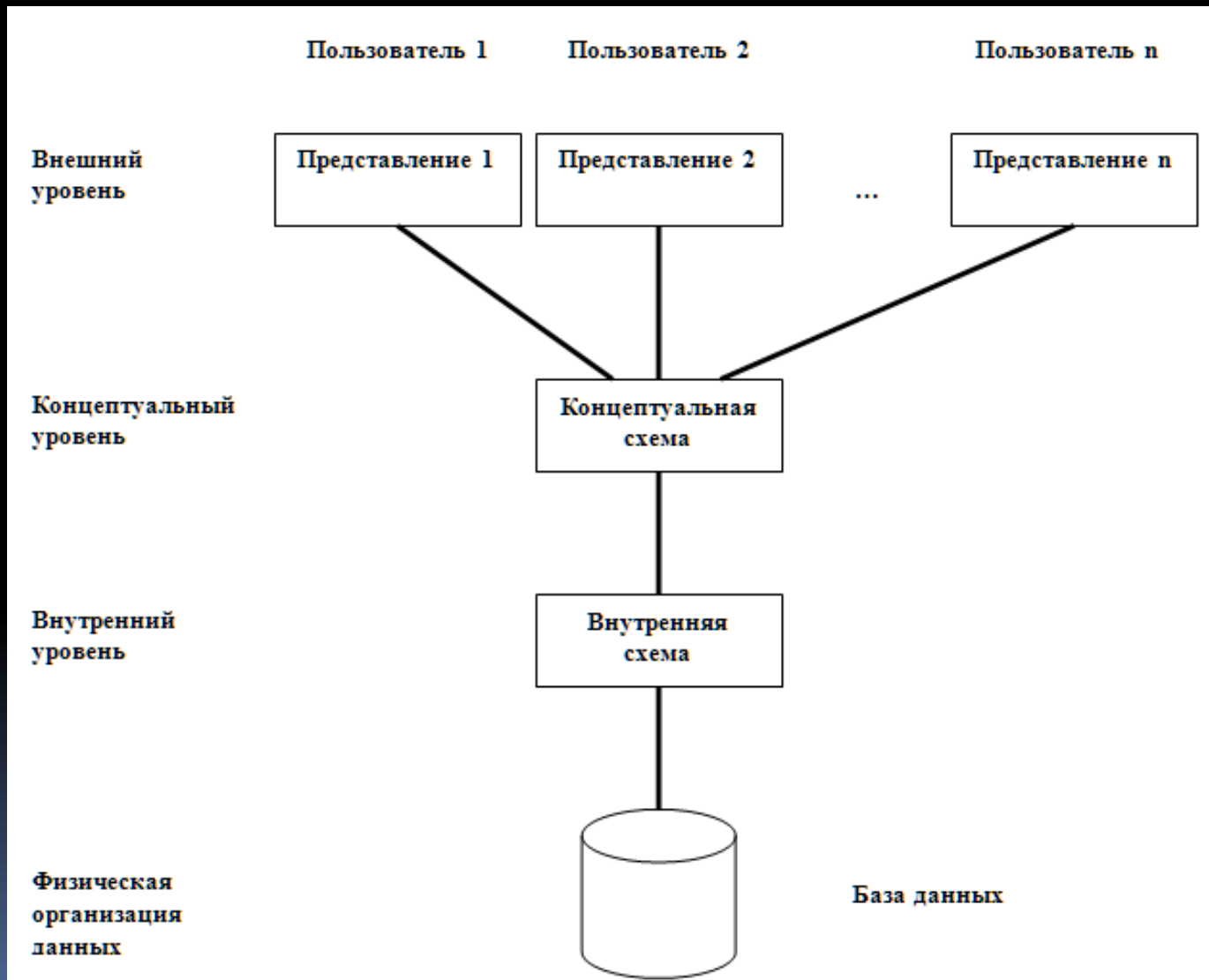
Результат должен содержать:

- Информационные потребности различных отделов организации.
- Информационные потребности руководства разного уровня.
- Информационные потребности филиалов организации.
- Информационная модель, отвечающая этим потребностям.
- Предполагаемый объем данных в различных регионах в изучаемый период.
- Предварительные оценки стоимости наращивания возможностей системы.
- Рекомендации по подробному проектированию новой системы или расширения старой с календарными планами.

Жизненный цикл базы данных




ТРЕХУРОВНЕВАЯ АРХИТЕКТУРА ANSI-SPARC





ТРЕХУРОВНЕВАЯ АРХИТЕКТУРА ANSI-SPARC

Цель трехуровневой архитектуры заключается в отделении пользовательского представления базы данных от ее физического представления



ТРЕХУРОВНЕВАЯ АРХИТЕКТУРА ANSI-SPARC

Причины разделения на уровни

- Каждый пользователь должен иметь возможность обращаться к одним и тем же данным, реализуя свое собственное представление о них. Каждый пользователь должен иметь возможность изменять свое представление о данных, причем это изменение не должно оказывать влияния на других пользователей

ТРЕХУРОВНЕВАЯ АРХИТЕКТУРА ANSI-SPARC

Причины разделения на уровни

- Пользователи не должны непосредственно иметь дело с такими подробностями физического хранения данных в базе, как индексирование и хеширование.

ТРЕХУРОВНЕВАЯ АРХИТЕКТУРА ANSI-SPARC

Причины разделения на уровни

- Администратор базы данных (АБД) должен иметь возможность изменять структуру хранения данных в базе, не оказывая влияния на пользовательские представления.

ТРЕХУРОВНЕВАЯ АРХИТЕКТУРА ANSI-SPARC

Причины разделения на уровни

- Внутренняя структура базы данных не должна зависеть от таких изменений физических аспектов хранения информации, как переключение на новое устройство хранения.

ТРЕХУРОВНЕВАЯ АРХИТЕКТУРА ANSI-SPARC

Причины разделения на уровни

- АБД должен иметь возможность изменять концептуальную структуру базы данных без какого-либо влияния на всех пользователей.



ТРЕХУРОВНЕВАЯ АРХИТЕКТУРА ANSI-SPARC

Внешний уровень – представление базы данных с точки зрения пользователей.



ТРЕХУРОВНЕВАЯ АРХИТЕКТУРА ANSI-SPARC

Концептуальный уровень – обобщающее представление базы данных. Этот уровень описывает то, какие данные хранятся в базе данных, а также связи, существующие между ними.

ТРЕХУРОВНЕВАЯ АРХИТЕКТУРА ANSI-SPARC


На концептуальном уровне представлены следующие компоненты:

- все сущности, их атрибуты и связи;
- накладываемые на данные ограничения;
- семантическая информация о данных;
- информация о мерах обеспечения безопасности и поддержки целостности данных.



ТРЕХУРОВНЕВАЯ АРХИТЕКТУРА ANSI-SPARC

Концептуальный уровень поддерживает каждое внешнее представление, в том смысле, что любые доступные пользователю данные должны содержаться (или могут быть вычислены) на этом уровне



ТРЕХУРОВНЕВАЯ АРХИТЕКТУРА ANSI-SPARC

Внутренний уровень – физическое представление базы данных в компьютере. Этот уровень описывает, как информация храниться в базе данных.

ТРЕХУРОВНЕВАЯ АРХИТЕКТУРА ANSI-SPARC

На внутреннем уровне хранится следующая информация:

- распределение дискового пространства для хранения данных и индексов;
- описание подробностей сохранения записей (с указанием реальных размеров сохраняемых элементов данных);
- сведения о размещении записей;
- сведения о сжатии данных и выбранных методах их шифрования.

ТРЕХУРОВНЕВАЯ АРХИТЕКТУРА ANSI-SPARC

Ниже внутреннего уровня находится *физический уровень*, который контролируется операционной системой, но под управлением СУБД.

ТРЕХУРОВНЕВАЯ АРХИТЕКТУРА ANSI-SPARC

Общее описание базы данных называется
схемой базы данных.


ТРЕХУРОВНЕВАЯ АРХИТЕКТУРА ANSI-SPARC

На самом высоком уровне имеется несколько *внешних схем* или *подсхем*, которые соответствуют разным представлениям данных.



ТРЕХУРОВНЕВАЯ АРХИТЕКТУРА ANSI-SPARC


На концептуальном уровне описание базы данных называют *концептуальной схемой*





ТРЕХУРОВНЕВАЯ АРХИТЕКТУРА ANSI-SPARC


На самом низком уровне абстракции описание базы данных называют *внутренней схемой*





ТРЕХУРОВНЕВАЯ АРХИТЕКТУРА ANSI-SPARC


Для каждой базы данных существует только одна концептуальная и одна внутренняя схема.



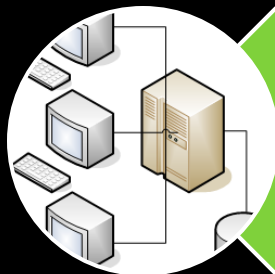


ТРЕХУРОВНЕВАЯ АРХИТЕКТУРА ANSI-SPARC

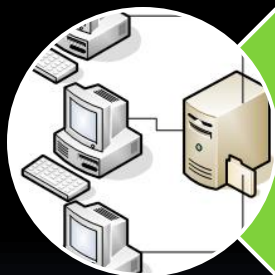
СУБД отвечает за установление
соответствия между тремя типами схем, а
также за проверку их непротиворечивости



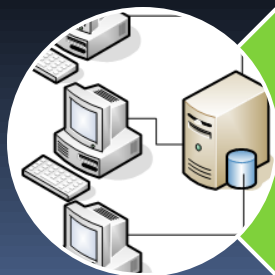
АРХИТЕКТУРА МНОГОПОЛЬЗОВАТЕЛЬСКИХ СУБД



Телеобработка

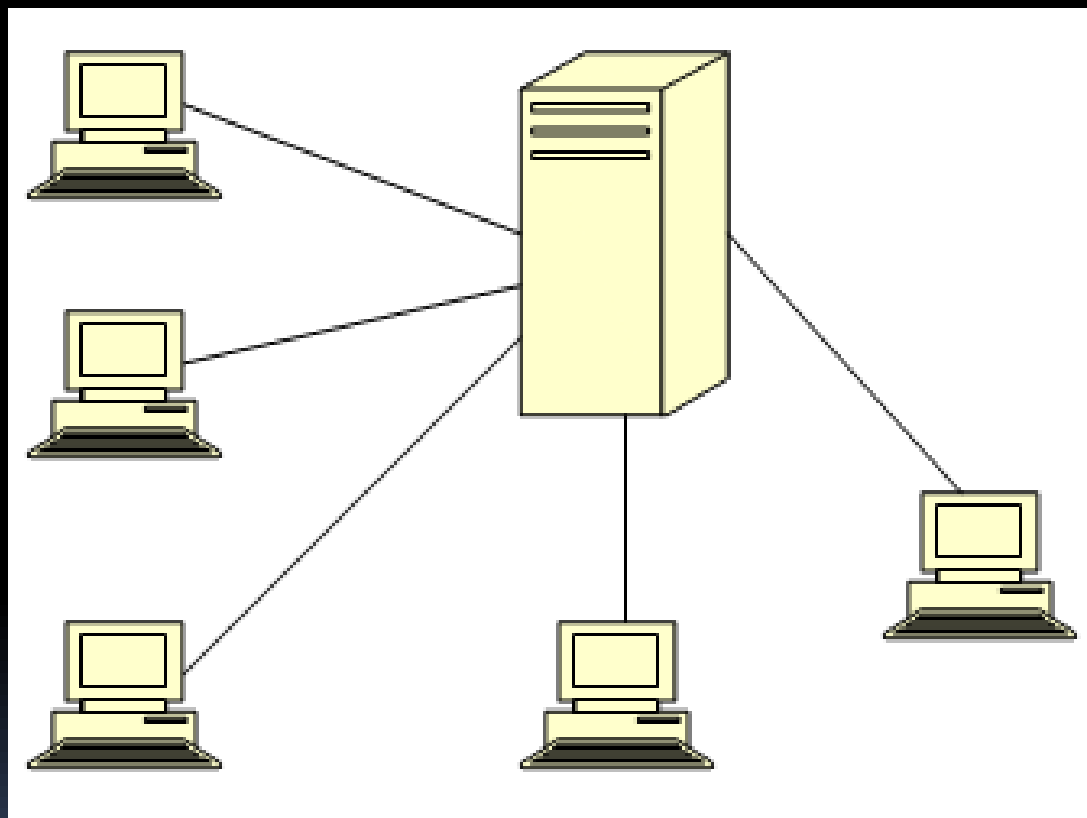


Файловый сервер

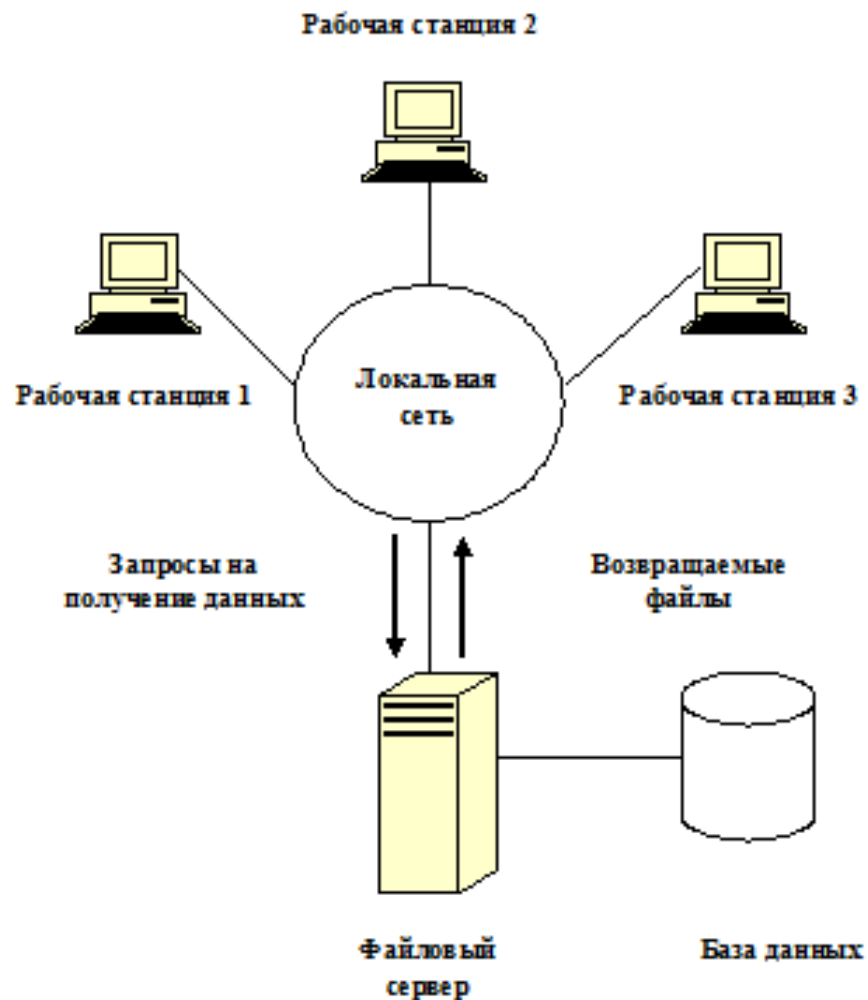


Технология «клиент-сервер»

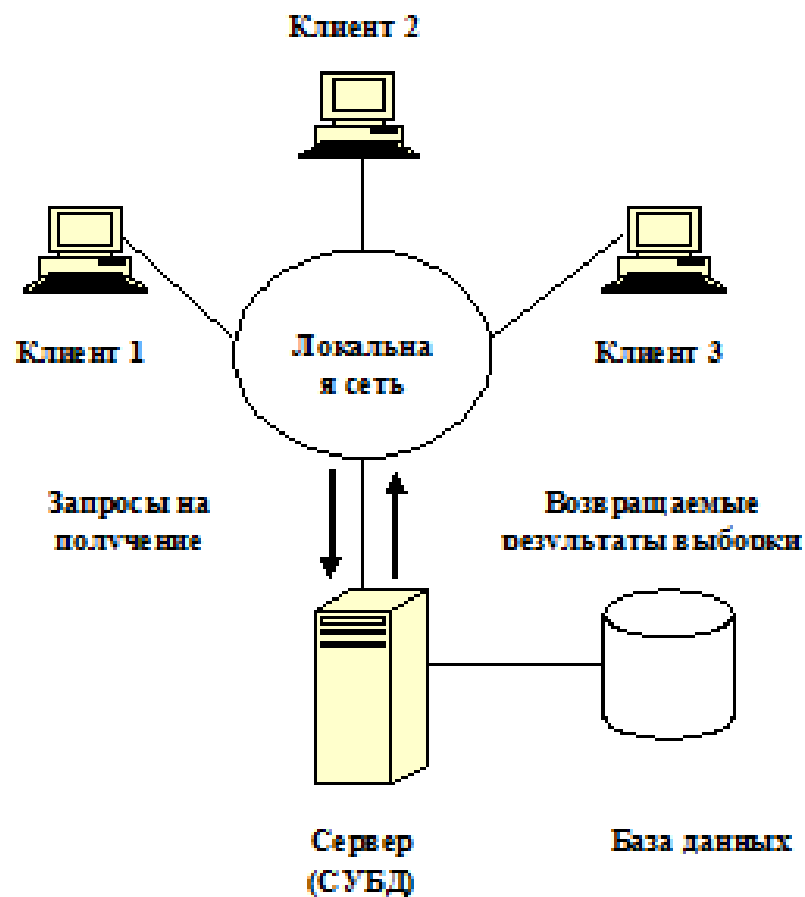
Телеобработка



Файловый сервер



Технология «клиент/сервер»



ЭТАПЫ ПРОЕКТИРОВАНИЯ БАЗЫ ДАННЫХ



Концептуальное
проектирование



Логическое
проектирование



Физическое
проектирование

1100



Процесс создания модели используемой на предприятии информации, не зависящей от любых физических аспектов ее представления



Логическое проектирование



Процесс создания модели используемой на предприятии информации на основе выбранной модели организации данных, но без учета типа целевой СУБД и других физических аспектов реализации

Физическое проектирование




Процесс подготовки
описания реализации
базы данных на
вторичных
запоминающих
устройства



Транзакции

Транзакция – одно или последовательность действий, выполняемых одним и тем же пользователем (или прикладной программой), которые получают доступ к базе данных или изменяют ее содержимое.



Транзакции

Транзакция может состоять из нескольких операций.


С точки зрения пользователя эти операции представляют собой единое задание.

С точки зрения проектировщика СУБД каждая транзакция переводит базу данных из одного непротиворечивого состояния в другое.



Транзакции


Цель проектирования транзакций заключается в определении и документировании высокоуровневых характеристик всех транзакций, которые должны будут выполняться в разрабатываемой базе данных





Транзакции


Существуют три основных типа транзакций:

- транзакции извлечения,
 - транзакции обновления,
 - смешанные транзакции
- 



Транзакции


Транзакции извлечения используются для выборки некоторых данных с целью отображения их на экране или подготовки отчета.





Транзакции


Транзакции обновления используются для вставки новых, удаления старых или же изменения уже существующих записей базы данных.





Транзакции

Смешанные транзакции включают как операции извлечения, так и операции обновления данных.





Вопросы?