

Управление данными.

Часть 1.

Введение в СУБД

1.1. Информационные системы

Области использования вычислительной техники:

1. Выполнение численных расчетов
2. Автоматические или автоматизированные информационные системы

Информационная система (ИС)

это совокупность структурированных данных и комплекса программно-аппаратных средств для хранения данных и манипулирования ими.

Классификация информационных систем



Классы информационных систем

Класс 1. Информационно-поисковые системы

- ориентация на поиск данных из общего множества по определенному **поисковому критерию**.
- пользователя интересует в большей степени **извлекаемая информация**, а технология обработки этих сведений.

Классы информационных систем

Класс 2. Системы обработки данных

- ориентация на **обработку** данных;
- пользователя интересует **результат** обработки данных, а не сами данные;
- вывод информации необязателен;

Виды информационных систем

Вид 1. Фактографические системы

- регистрация **конкретных значений** данных объектов реального мира;
- информация имеет четкую **структуру** (формат);
- **однозначные** ответы на поставленные вопросы

Виды информационных систем

Вид 2. Документальные системы

- совокупность **неструктурированных** документов (текстовых и графических);
- нет однозначных ответов на поставленные вопросы. **Результат** - список документов или объектов, в какой-то мере удовлетворяющих сформулированным в запросе условиям.

Терминология

Объект

это нечто существующее и различимое, то есть имеется способ отличить один подобный объект от другого.

Данное

это определенный показатель, который характеризует объект и принимает для конкретного экземпляра этого объекта некоторое значение.

Структура данных

Структурирование информации

это введение каких-либо соглашений о способах представления данных (установка формата, т.е. определенного типа, длины значений данных).

Пример неструктурированных данных:

«Табельный номер 1234 Иванов Иван Иванович, дата рождения 10 мая 1967 года. Номер по табелю Петрова Сергея Александровича №8191, д.р. 18.10.1972 г. Табель №3451 Сидорова Алексея Петровича, родившегося 5 июля 1964 года.»

Пример структурированных данных:

| Табельный номер | Фамилия | Имя | Отчество | Дата рождения |
|-----------------|---------|---------|---------------|---------------|
| 1234 | Иванов | Иван | Иванович | 10.05.1967 |
| 8191 | Петров | Сергей | Александрович | 18.10.1972 |
| 3451 | Сидоров | Алексей | Петрович | 05.07.1964 |

1.2. Традиционный подход к организации данных. Файлы и файловые системы

Файл

это именованная область внешней памяти, в которую можно записывать и из которой можно считывать данные.

С точки зрения пользователя файл содержит линейную последовательность записей

Стандартные операции:

- создать файл (требуемого типа и размера);
- открыть ранее созданный файл;
- прочитать из файла некоторую запись (текущую, следующую, предыдущую, первую, последнюю);
- записать в файл на место текущей записи новую, добавить новую запись в конец файла.

Типы файлов

Файл последовательного доступа

Представляет собой последовательность записей данных в виде строк произвольной длины, разделенных запятыми или специальными символами, обозначающими переход на новую строку.

Особенности:

- отсутствие возможности упорядочить хранимые записи;
- размещение и извлечение записей в такой файл производится построчно в определенной последовательности.

Файл произвольного доступа

Состоит из записей фиксированной длины, которая указывается при его создании.

Особенности:

- все записи в упорядочены, каждая имеет свой номер;
- возможность быстрого перемещения на любую запись, минуя предыдущие.

Пример традиционного подхода к организации данных

ПОЛЬЗОВАТЕЛИ

ПРИЛОЖЕНИЯ

ФАЙЛЫ ДАННЫХ



Пример традиционного подхода к организации данных

ПОЛЬЗОВАТЕЛИ

ПРИЛОЖЕНИЯ

ФАЙЛЫ ДАННЫХ

ОТДЕЛ
СНАБЖЕНИЯ

Учет материалов

Состав шихтовых материалов

Расход шихтовых материалов

Учет топливно-
энергетических
ресурсов

Расход топливно-
энергетических ресурсов

Состав топливно-
энергетических ресурсов

ТЕХНОЛОГИЧЕ
СКАЯ
ГРУППА

Расчет шихтовых
материалов

Дутьевые параметры

Состав шихтовых материалов

Расход шихтовых материалов

Оптимальное
распределение
комбинированного
дутья между
печами цеха

Режимные и конструктивные
параметры всех печей цеха

Расход топливно-
энергетических ресурсов по
всем печам цеха

Недостатки традиционного подхода

1. Избыточность данных
2. Проблемы непротиворечивости данных
3. Ограниченная доступность данных
4. Сложности в организации и управлении
5. Недостаточность средств защиты хранимых данных
6. Низкопроизводительная работа в многопользовательской среде
7. Отсутствие процедур восстановления данных после возникновения отказов;
8. Отсутствие средств манипулирования данными;
9. Высокая стоимость программирования и сопровождения;
0. Негибкость к изменениям и др.

1.3. Организация данных с использованием БД

Данные

это любая информация об объектах окружающего мира, представленная в формализованном виде, пригодном для ее передачи, хранения и обработки при помощи некоторого процесса с использованием средств вычислительной техники.

База данных (БД)

это именованная совокупность данных, организованных по определенным правилам, предусматривающим общие принципы описания, хранения и манипулирования данными, не зависящая от прикладных программ.

Предметная область

это часть реального мира, подлежащего изучению для организации управления и, в конечном счете, автоматизации.

Организация данных с использованием БД

Система управления базами данных (СУБД)

это совокупность языковых и программных средств, предназначенных для создания, ведения и совместного использования БД многими пользователями.

Банк данных (БнД)

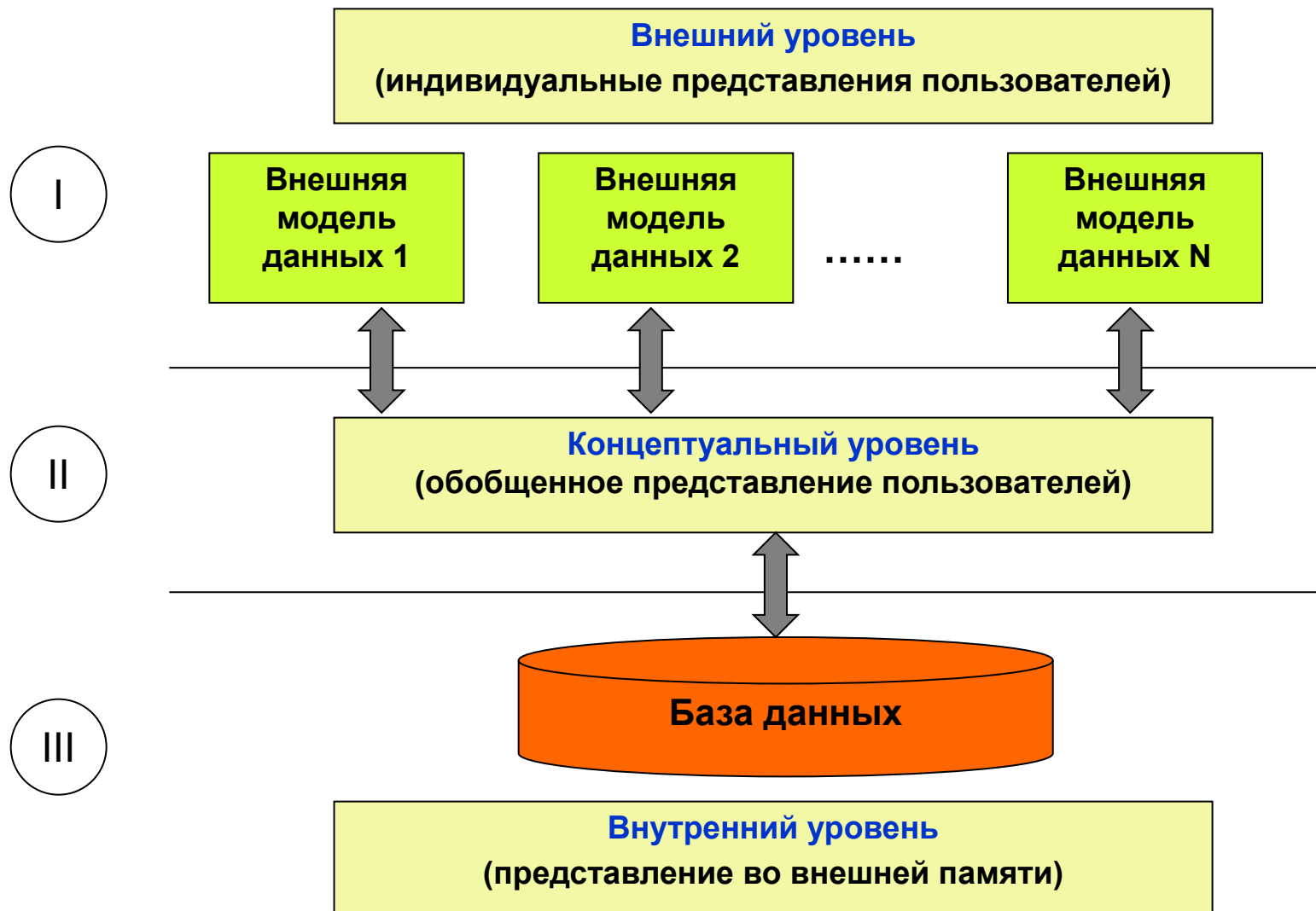
это система специальным образом организованных данных – баз данных, программных, технических, языковых, организационно-методических средств, предназначенных для обеспечения централизованного накопления и коллективного многоцелевого использования данных.

Пример организации баз данных

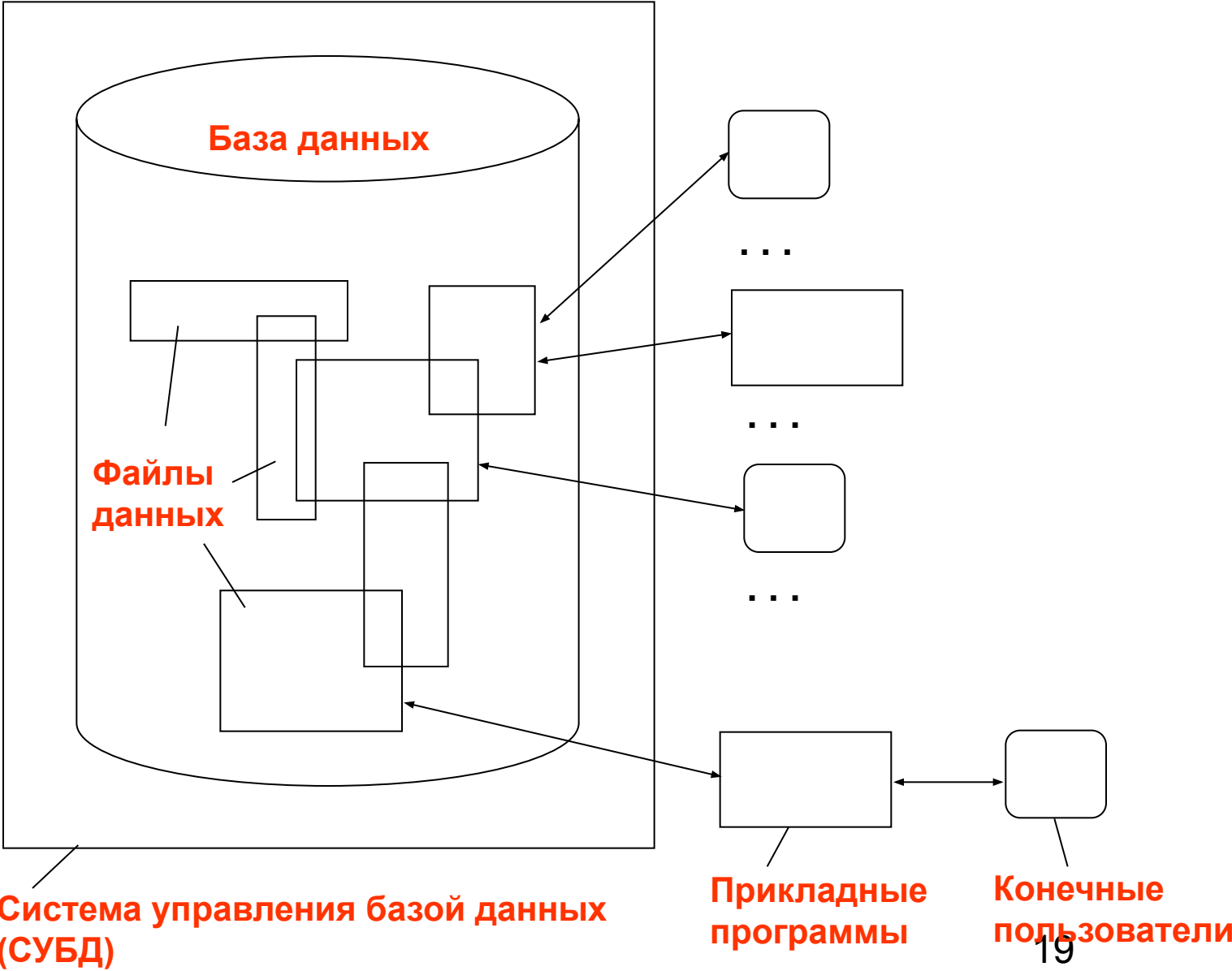


1.4. Архитектура базы данных

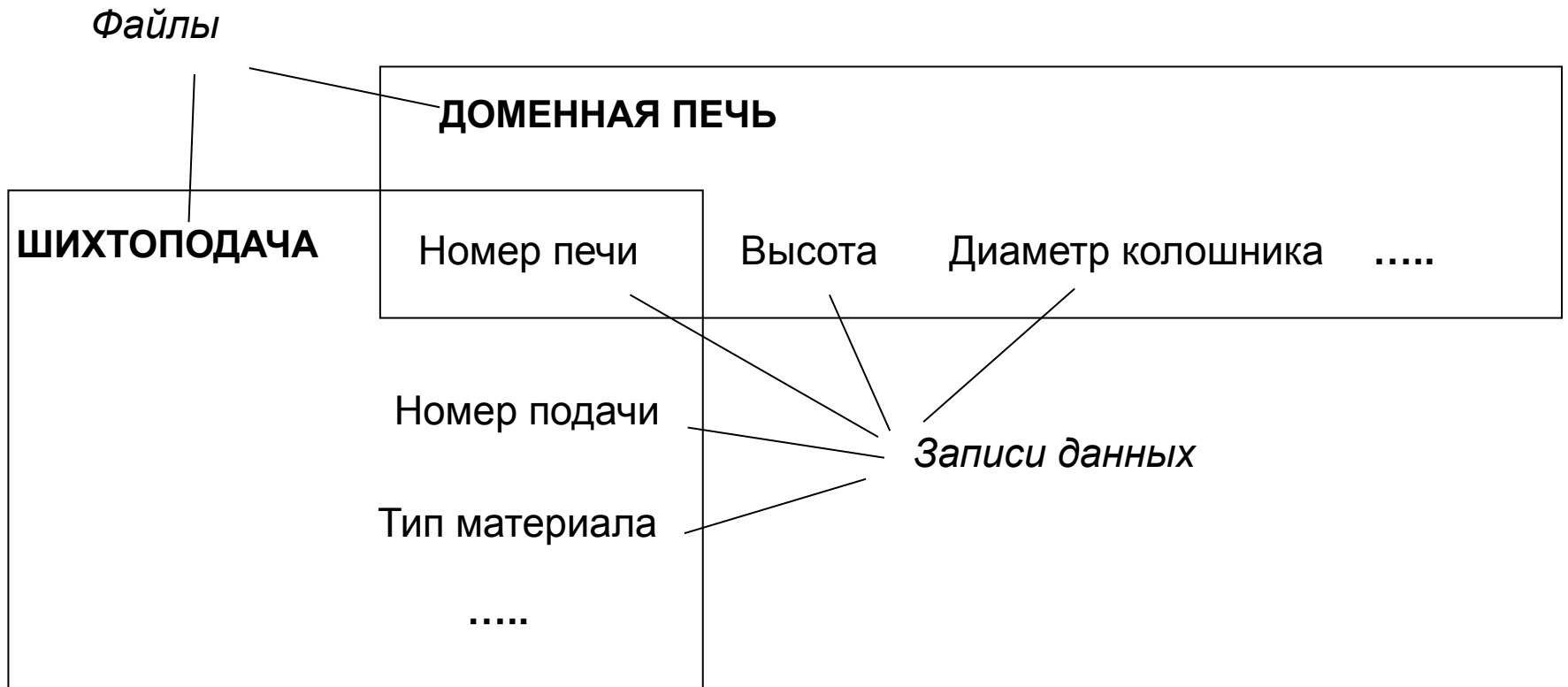
Архитектура базы данных,
предложенная ANSI (American National Standards Institute)



1.5. Компоненты системы баз данных



Фрагмент организации данных в базе данных



Языковые средства СУБД

- **DDL** – Data Definition Language,
Язык Определения Данных, ЯОД
- **DML** – Data Management Language,
Язык Манипулирования Данными, ЯМД
- Язык запросов (**SQL**, Structured Query Language)

Категории пользователей

- Конечные пользователи
- Прикладные программисты
- Администраторы данных (АД),
администраторы базы данных (АБД),

1.6. Основные функции группы администратора БД

1. Анализ предметной области.
2. Проектирование структуры БД.
3. Задание ограничений целостности при описании структуры БД и процедур обработки БД.
4. Первоначальная загрузка и ведение БД.
5. Защита данных.
6. Обеспечение восстановления БД.
7. Анализ обращений пользователей БД.
8. Анализ эффективности функционирования БД.
9. Работа с конечными пользователями.
10. Подготовка и поддержание системных средств.

1.7. Преимущества и недостатки современного подхода к организации данных

Преимущества

1. Сокращение избыточности данных.
2. Устранение противоречивости данных.
3. Общий доступ к данным.
4. Соблюдение стандартов.
5. Введение ограничений для обеспечения безопасности.
6. Обеспечение целостности данных.

Недостатки современного подхода к организации данных

1. Потеря пользователями права единоличного владения данными.
2. Повышение вероятности нарушений защиты данных.
3. Повышенная угроза секретности хранимой информации.

Требования к современной СУБД

- Эффективное выполнение различных функций предметной области
- Минимизация и контроль избыточности хранимых данных
- Предоставление для принятия решений непротиворечивой (согласованной) информации
- Обеспечение возможности одновременного доступа к базе данных нескольких уполномоченных пользователей
- Обеспечение управления безопасностью
- Простая физическая реорганизация, т.е. изменение структуры данных в базе данных
- Возможность централизованного управления базой данных

1.8. Классификация систем баз данных

По характеру использования

Однопользовательские системы

— это системы, в которых в одно и то же время к базе данных может получить доступ **не более одного** пользователя, так называемые базы данных с локальным доступом

Многопользовательские системы

— это системы, в которых к базе данных могут получить доступ одновременно **несколько** пользователей, так называемые базы данных с удаленным (сетевым) доступом

Классификация систем баз данных

По технологии обработки данных

Централизованные системы

— база данных **физически** хранится в памяти одного компьютера.

Распределенная система баз данных

— состоит из нескольких, возможно, пересекающихся или даже дублирующих друг друга частей БД, хранимых **в различных компьютерах-серверах**, которые в общем случае могут быть **удалены географически** друг от друга на значительные расстояния, т.е. территориально распределены

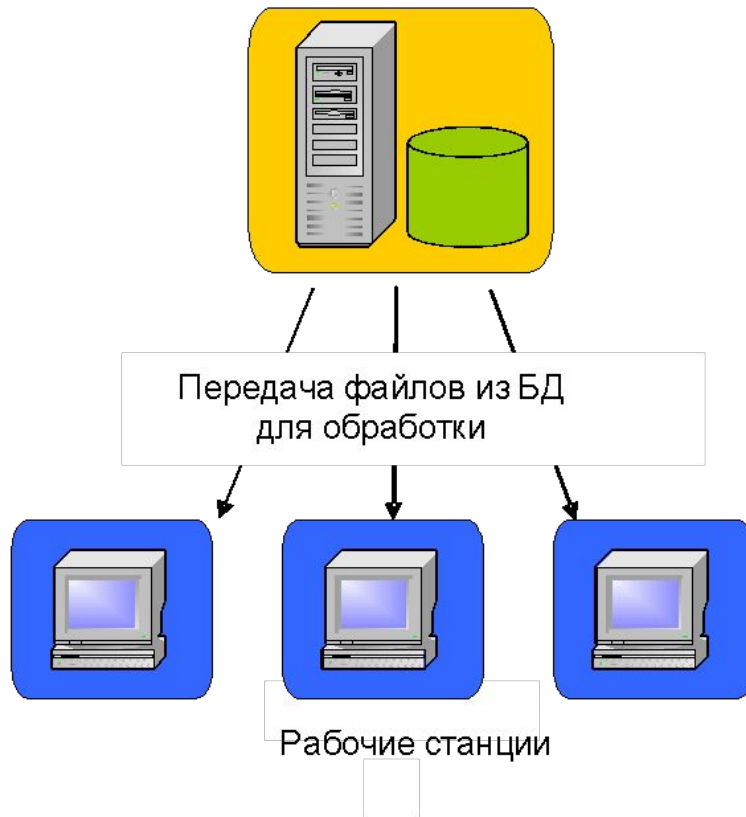
Архитектуры построения централизованных систем

Схема обработки информации в БД

по принципу файл/сервер (а) и клиент/сервер (б)

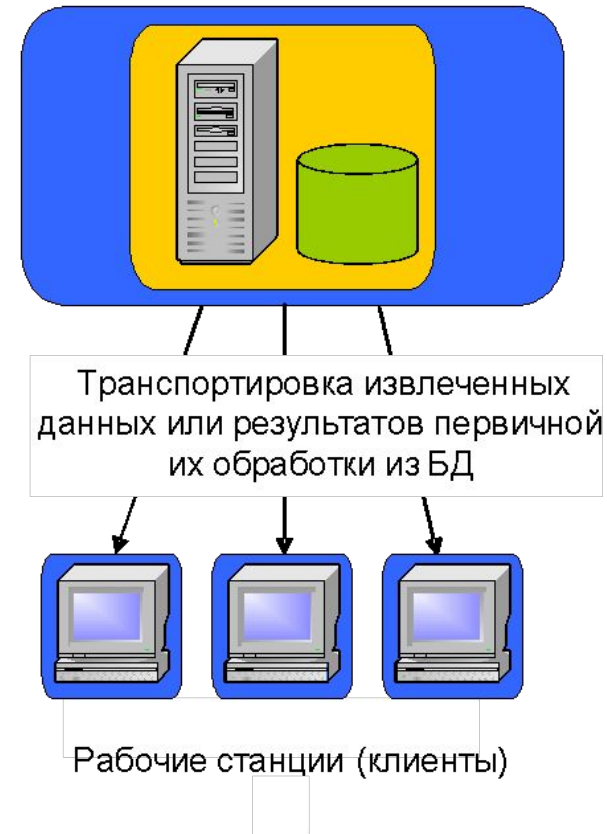
а



Файл/сервер



б

Клиент/сервер



-  хранение данных
-  обработка данных

Фрагмент архитектуры построения распределенных систем

Полная распределенная система обработки БД

