

# Организация и проектирование баз данных

Курс лекций

для слушателей переподготовки специальности  
1 40 01 73 «Программное обеспечение  
информационных систем»

Версия 2017

# Базы данных и информационные системы

# Информационные системы

**Информационная система - это ...**

(з-н РБ № 455-3 Об информации, информатизации и защите информации)

**... совокупность банков данных, информационных технологий и комплекса (комплексов) программно-технических средств**

**банк данных – организационно-техническая система, включающая одну или несколько баз данных и систему управления ими**

**база данных – совокупность структурированной и взаимосвязанной информации, организованной по определенным правилам на материальных носителях**

## **Отличительные характеристики ИС**

- большие объемы хранимой информации
- простые алгоритмы обработки информации
- не критичность времени получения информации
- большое количество источников и получателей информации

# Виды информационных систем

**OLTP** (Online Transaction Processing, оперативная обработка транзакций)

БД называются БД оперативной обработки.

**OLAP** (On - Line Analytical Processing - оперативная аналитическая обработка данных) и **Data Mining** (раскопка данных).

БД называются хранилищами.

Основное назначение **OLTP**:

формирование первичных согласованных данных

Основное назначение **OLAP** и **Data Mining**:

как можно быстрее выдавать данные для построения аналитических отчетов (OLAP) и поиск скрытых закономерностей в данных и решение задач прогнозирования (Data Mining).

Основная характеристика

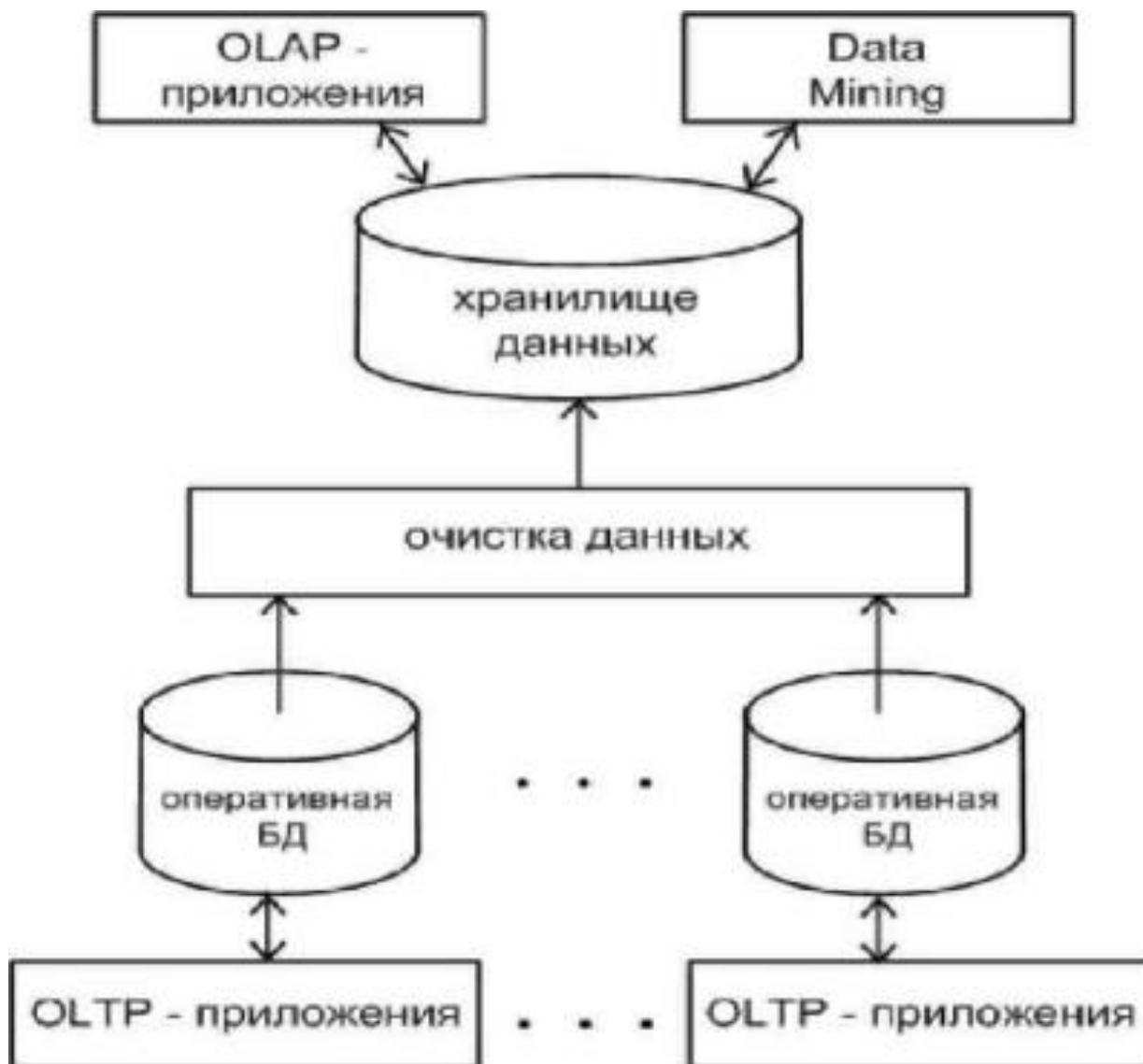
**OLTP**:

выполнение большого количества коротких транзакций на изменение данных в реальном времени.

Основная характеристика **OLAP** и **Data Mining**:

- выполнение небольшого количества сложных транзакций на выборку данных.

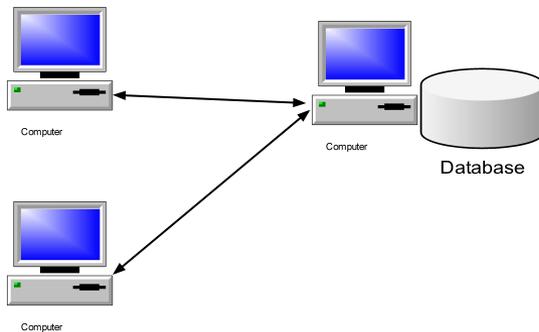
# Взаимоотношение видов ИС



# Типы информационных систем

Системы с распределенной обработкой (DDP, Distributed Data Processing)

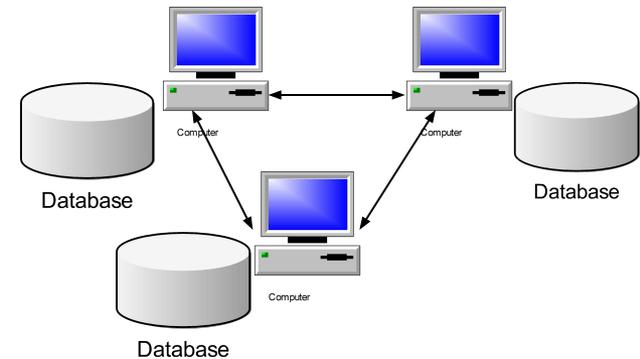
Основной признак **DDP** - централизованная база данных



Основное достоинство **DDP** – высокая согласованность данных

Системы распределенных баз данных (DDB, Distributed Data Base)

Основной признак DDB – распределенная база данных

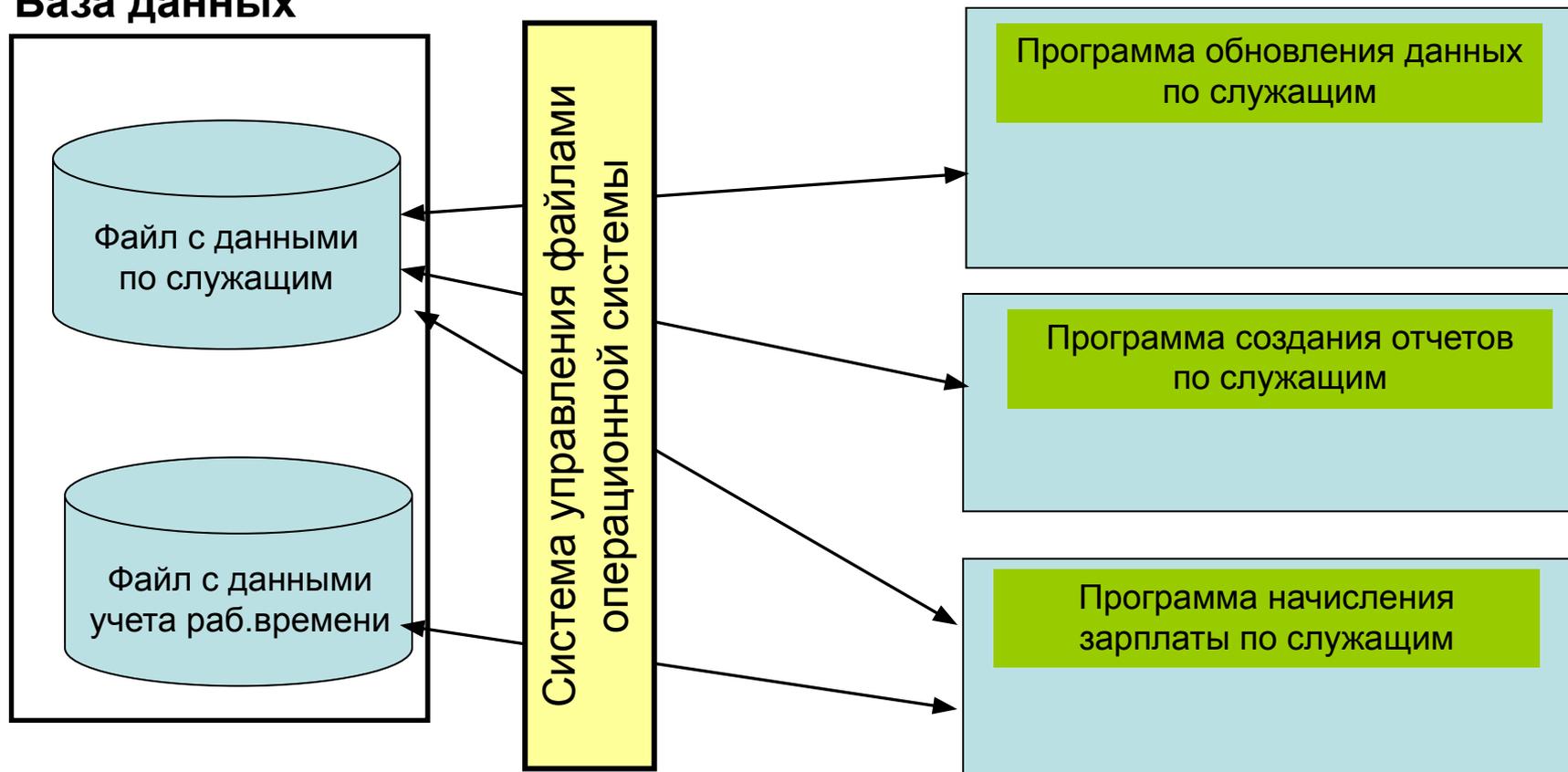


Основное достоинство DDB – масштабируемость и устойчивость

# Доступ к базе данных

База данных – совокупность структурированной и взаимосвязанной информации, отражающую состояние объектов и их отношений в предметной области

## База данных

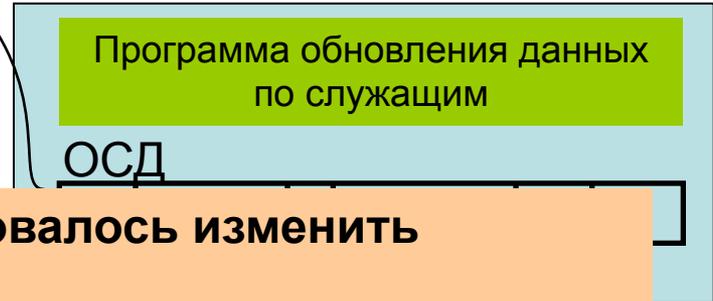


# Предпосылки появления СУБД

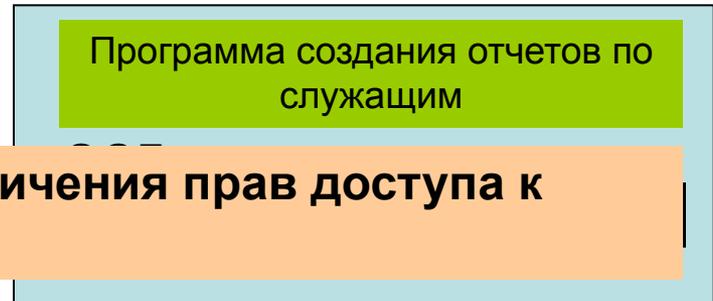
```
struct  
{ CString  s_name;  
  int     i_Status;  
  int     i_Valus ...  
}
```



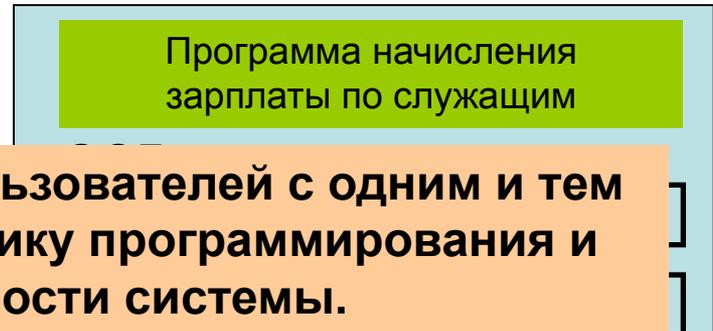
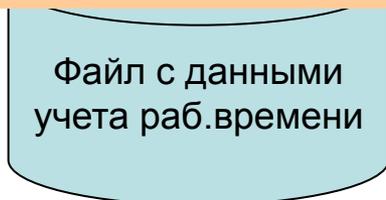
**1. При изменении модели данных требовалось изменить программы всех пользователей.**



**2. Невозможность обеспечения разграничения прав доступа к данным разных пользователей.**



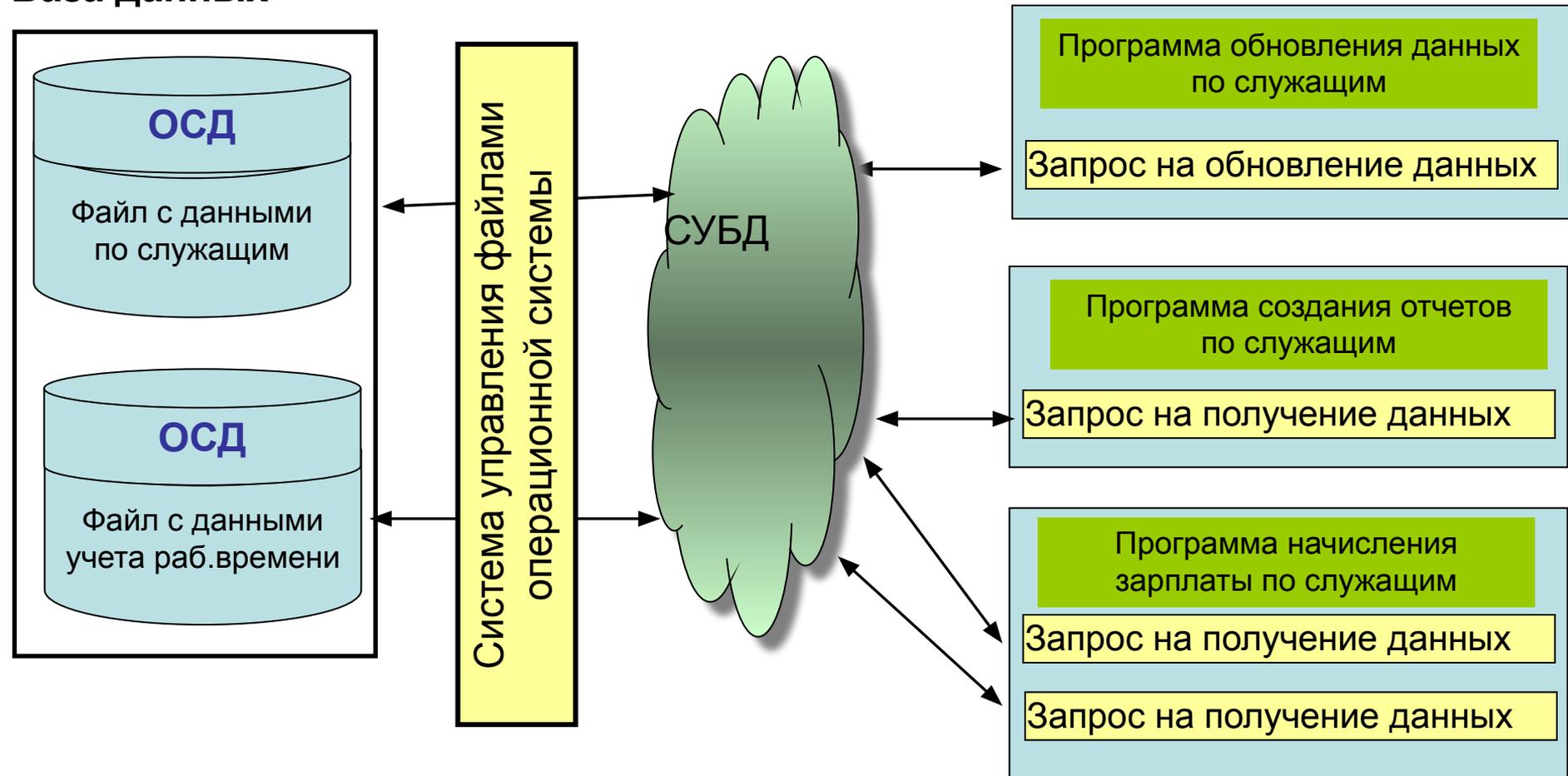
**3. Параллельная работа нескольких пользователей с одним и тем же файлом значительно затрудняла логику программирования и приводила к снижению производительности системы.**



# СУБД

СУБД – совокупность языковых и программных средств, предназначенных для управления совместным использованием базы данных многими пользователями

## База данных



# Основные функции СУБД

**1. Определение данных**

**Описание структур данных**

**2. Обработка данных**

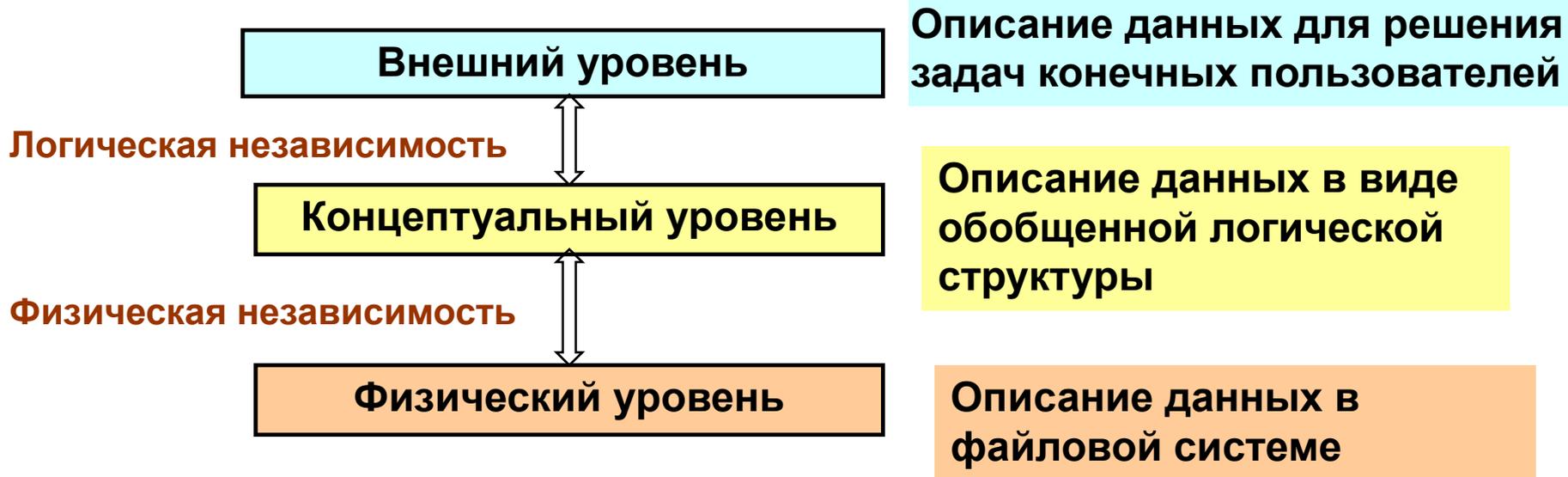
**Добавление, изменение и удаление данных, сортировка и поиск.**

**3. Управление данными**

**Определение пользователей и их прав доступа к различным данным, блокировки доступа к данным, управление данными на внешних устройствах, ...**

# Архитектура БД

ANSI (American National Standards Institute) предложена организации БД в виде 3-х уровней абстракций описания данных



**Логическая независимость** – возможность изменения одного приложения без корректировки другого

**Физическая независимость** – перенос хранения информации с одного носителя на другой при сохранении работоспособности всех приложений

Каждый уровень архитектуры БД представляется моделью данных

# Модели данных

# Объект БД - информация

Термин «информация» происходит от латинского слова *«informatio»* - разъяснение, изложение, сведения.

информация – сведения о лицах, предметах, фактах, событиях, явлениях и процессах независимо от формы их представления;

Информация отличается от понятия *«данные»* и *«сообщение»*.

6032106011720325

60321

06

01

17

2

03

25

Код группы

день

месяц

год

период

кол. предметов

кол.слуш.

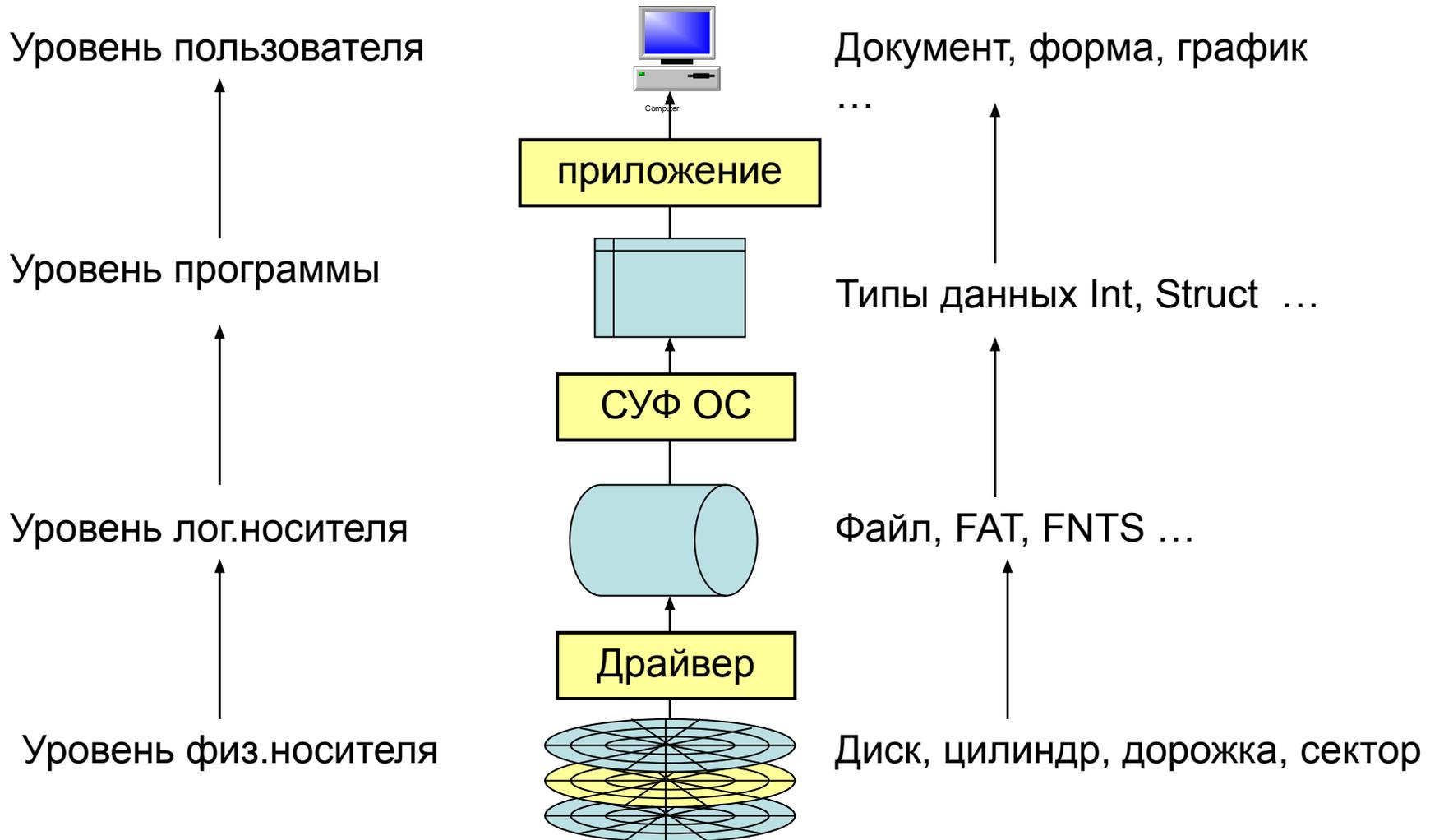
Информация извлекается из данных и зависит от объекта воспринимающего (обрабатывающего) это сообщение.

Результат зависит от свойств этого объекта.

В зависимости от целей и интересов из одного и того же сообщения можно извлечь совершенно разную информацию.

# Иерархия информации

Информация в ИС имеет много уровней абстрагирования



# Модель данных

Модель – это абстракция, с помощью которой человек описывает суть проблемы или явления без акцента на существенных деталях, с целью понимания и возможности исследования проблемы или явления.

Модель данных – это набор концепций, содержащих сведения об организации данных, их отношениях (взаимосвязях), ограничениях и допустимых операциях.



# Модель данных

Каждый уровень архитектуры БД представляется моделью данных

Модели внешнего уровня являются подсхемами даталогических моделей или структурами данных, используемых в системах программирования (объектно-ориентированная модель)

**Даталогическая модель определяет тип конкретной СУБД**

Модели концептуального уровня называют даталогическими. Они представляет собой отображение логических связей между элементами данных безотносительно к их содержанию и среде хранения. В них отражены особенности предметной области, для которой создаётся БД.

Модели физического уровня называются физическими моделями данных и определяют способы размещения и доступа данных в файлах.

# Классификация концептуальных моделей данных

## Даталогические модели

Описывают слабоструктурированную информацию

описывают информацию в виде жесткой структуры

### Документальные

Формат документа

Дискрипторные

Тезаурусные

### Фактографические

Графовые

Множественные

Объектно-ориентированные

Иерархические

Сетевые

Реляционные

Бинарных ассоциаций

Объектно-реляционные

Модели, ориентированные на формат документа

Тезаурусные модели содержат определенные

Дескрипторные модели основаны на

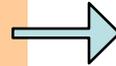
Графовые модели данных отражают совокупность

Множественные модели основаны на теории

Модели бинарных ассоциаций используются в системах искусственного интеллекта. К этим моделям относятся модели, например, Data Semantics (автор Абриал) и DIAM II (автор Сенко).

# Модель данных основа СУБД

Документальные модели – определяют семейство документно-ориентированных БД.



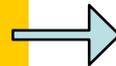
Lotus Notes ( [IBM](#))/ Notes/Domino;  
CouchDB, MongoDB

Реляционная модель данных – определяет семейство реляционных СУБД



DB2; Informix; Oracle; Teradata Database; Microsoft SQL Server; SyBase; Microsoft Access; PostgreSQL; MySQL; FoxPro ...

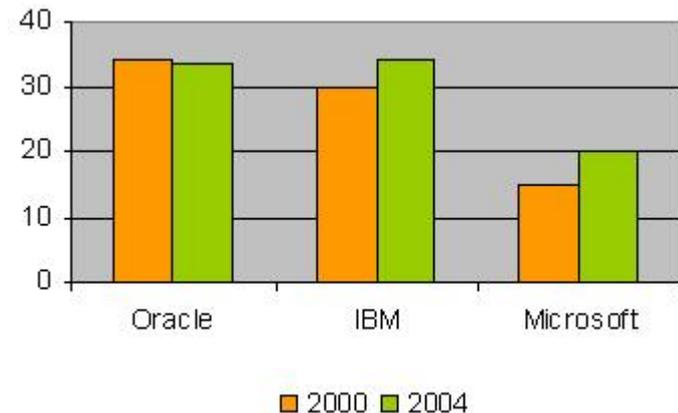
Объектно-ориентированные модели данных – определяют семейство объектно-ориентированных СУБД



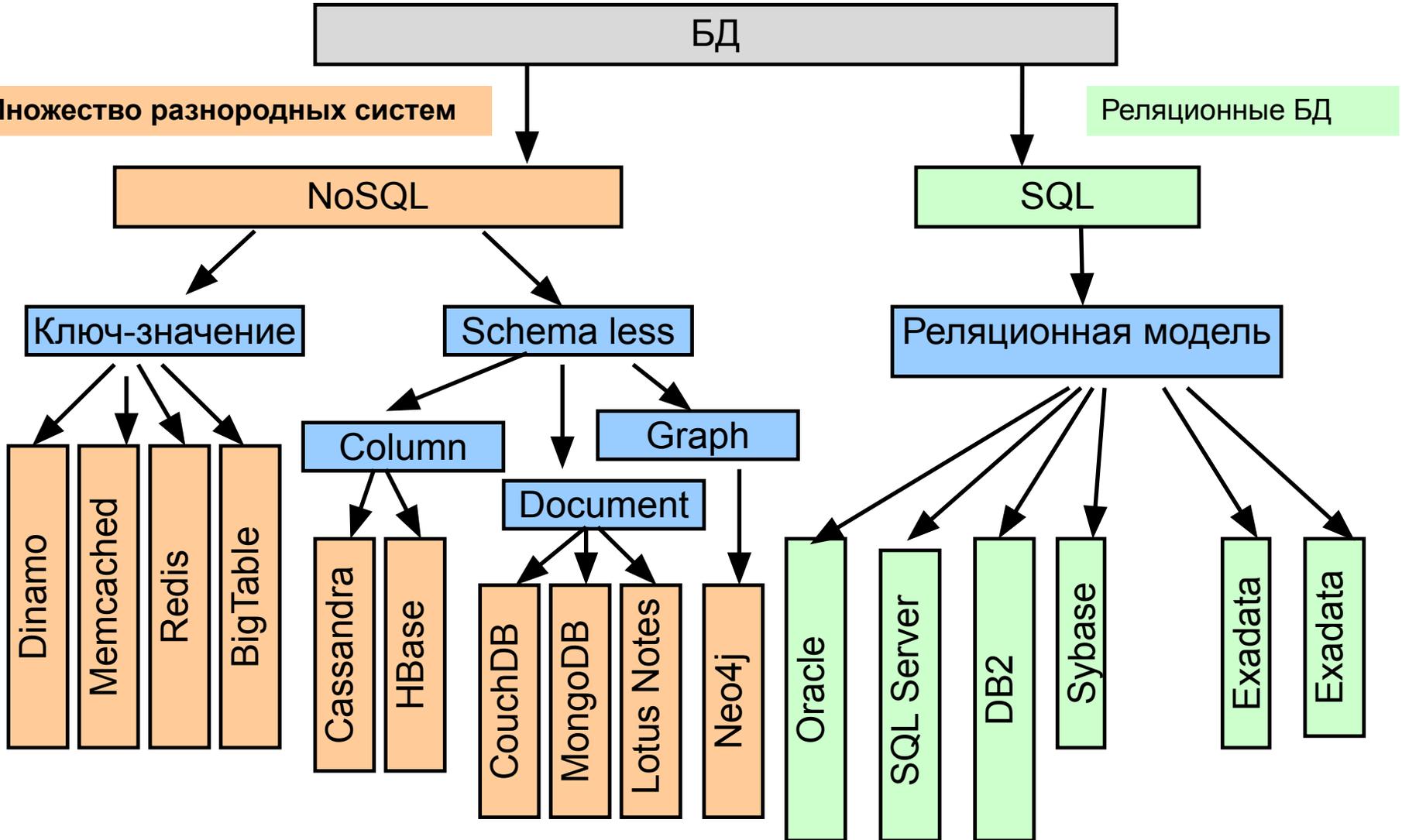
Db4o; ObjectStore; Caché

Реляционные СУБД составляют значительную часть рынка (порядка 80% от всего рынка).

Значительная часть рынка реляционных СУБД принадлежит Oracle, DB2, MS SQL Server, SyBase, MySQL (более 90%).



# Ещё одна классификация базы данных



# Характеристика NoSQL БД

1. Не используется SQL

2. Неструктурированные (schemaless)

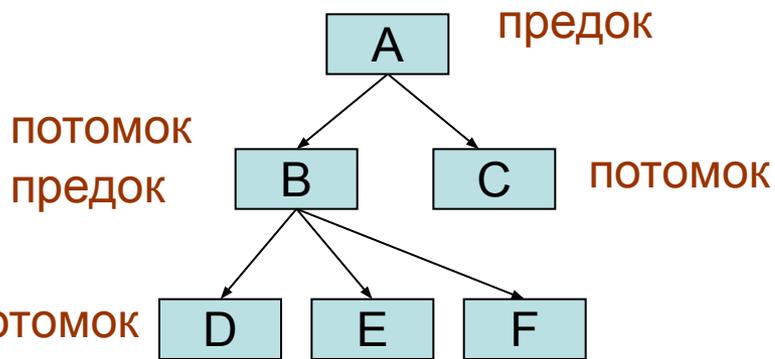
3. Представление данных в виде агрегатов (aggregates).

4. Слабые ACID свойства.

5. Применяются в распределенных системах, без совместно используемых ресурсов (share nothing).

# Иерархическая модель

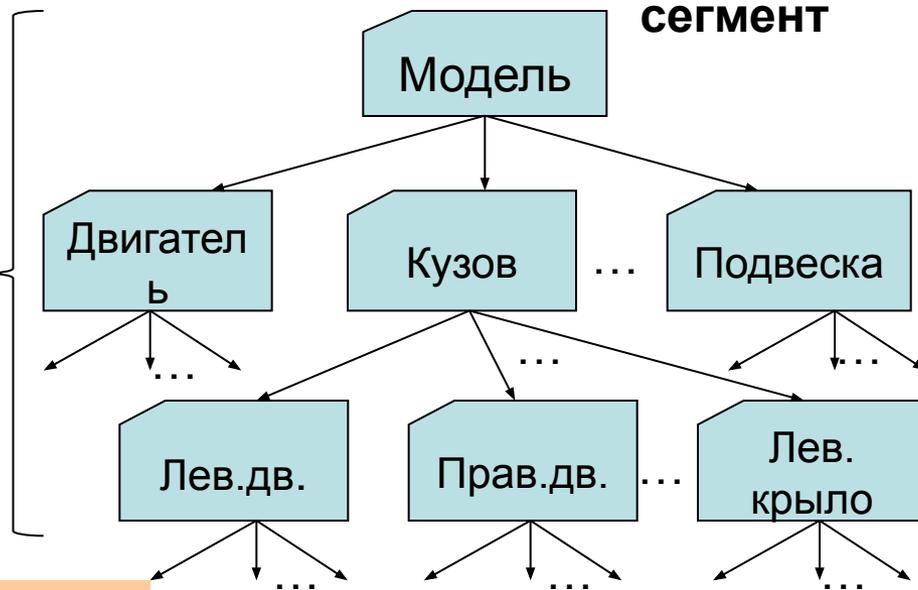
Модель описывает структуру данных в виде направленного графа



Модель данных для БД  
автомобильной компании

Корневой  
сегмент

сегменты

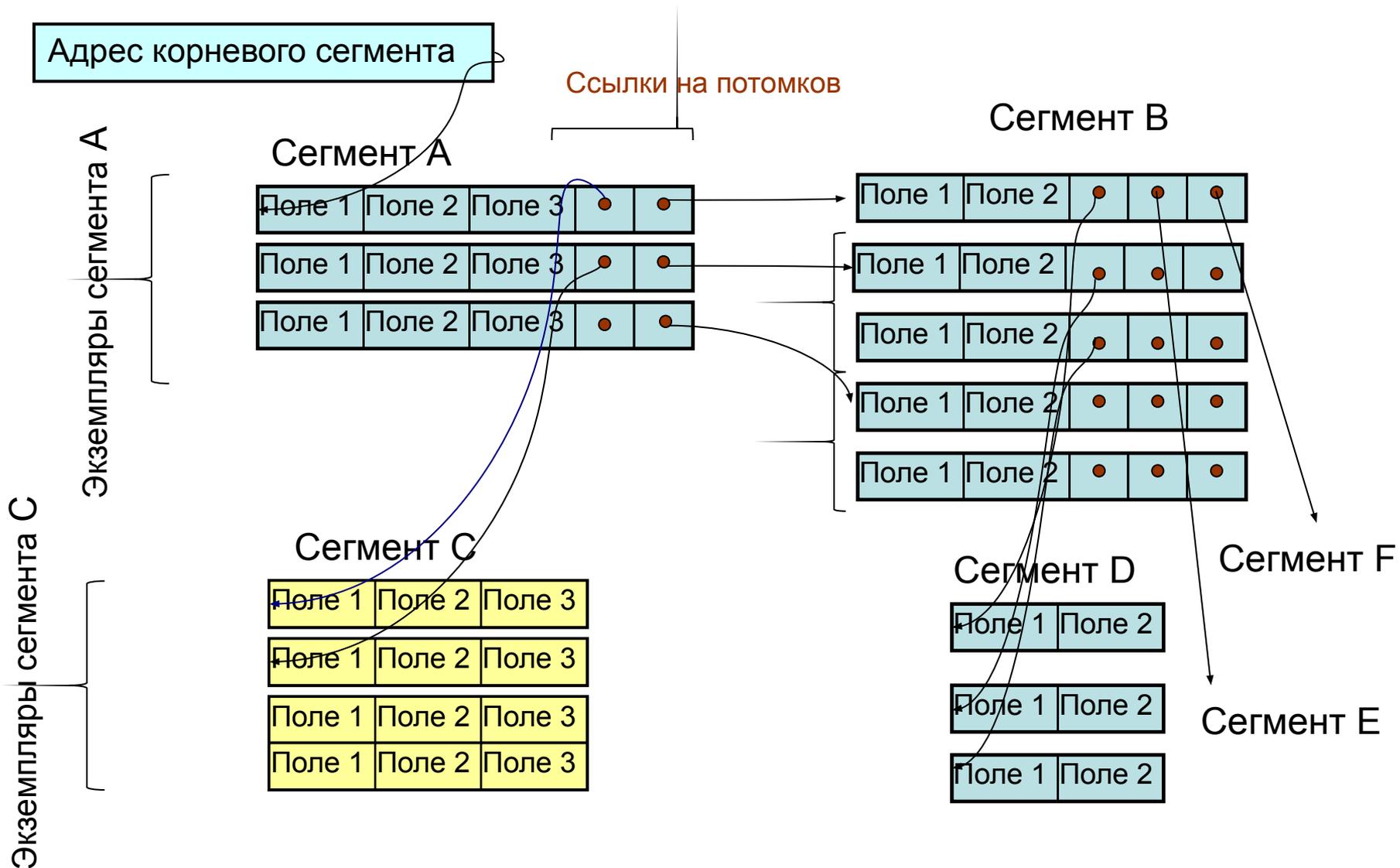


Сегмент – это набор  
однородных записей

Записи состоят из полей

Поле – это минимальная неделимая  
структура данных, доступная  
пользователю как единое целое

# Иерархическая модель



# Иерархическая модель

Описание данных иерархической модели включает описание сегментов и полей для каждого сегмента

**SEGM NAME** = <имя сегмента>

**BYTES** = <размер сегмента в байтах>

**FREQ** = <среднее количество экземпляров>

**PARENT** = <имя корневого сегмента>

**FIELD NAME** = (<имя поля>[,SEQ],{U|M})

**START** = <размер поля в байтах>

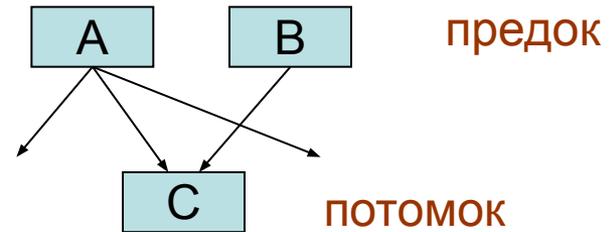
**TYPE** = {X|P|C}

**FIELD NAME** = ...

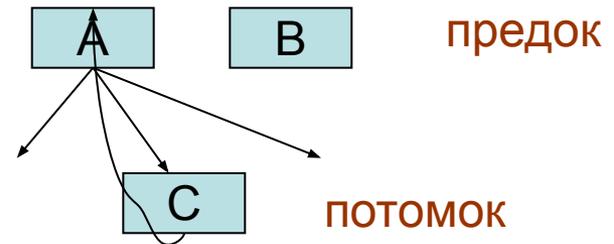
# Иерархическая модель

## Недостатки иерархической модели

**1. Невозможность простой реализации связей, основанных на потомке, имеющем несколько предков**



**2. Сложность реализации связей М:М, основанных на потомке, имеющем несколько предков**

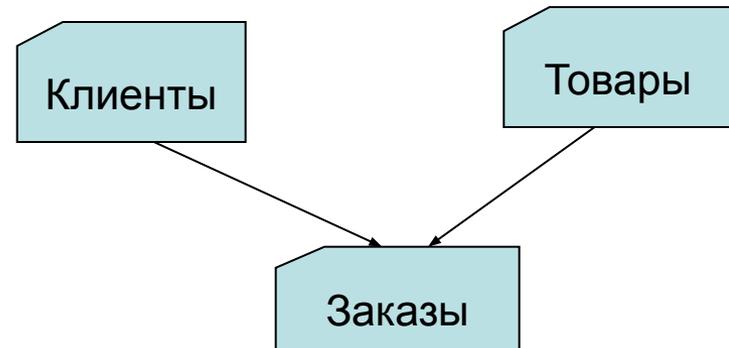


# Сетевая модель

Модель описывает структуру данных в виде наборов двухуровневых графов

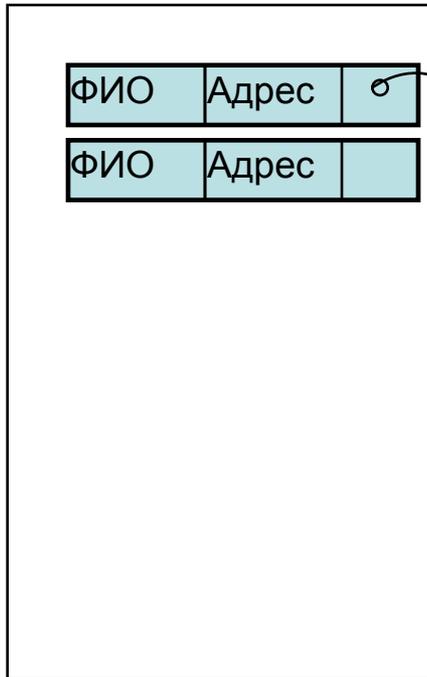
Набором называется двухуровневый граф, связывающий отношением «один-ко-многим» два типа записей

Между двумя типами записей может быть определено любое количество наборов

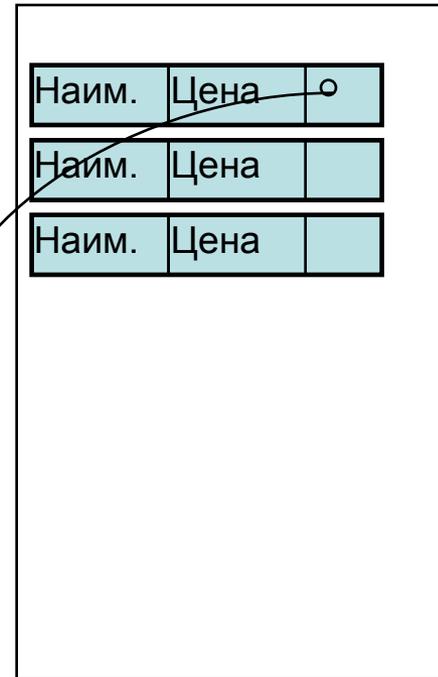


# Сетевая модель

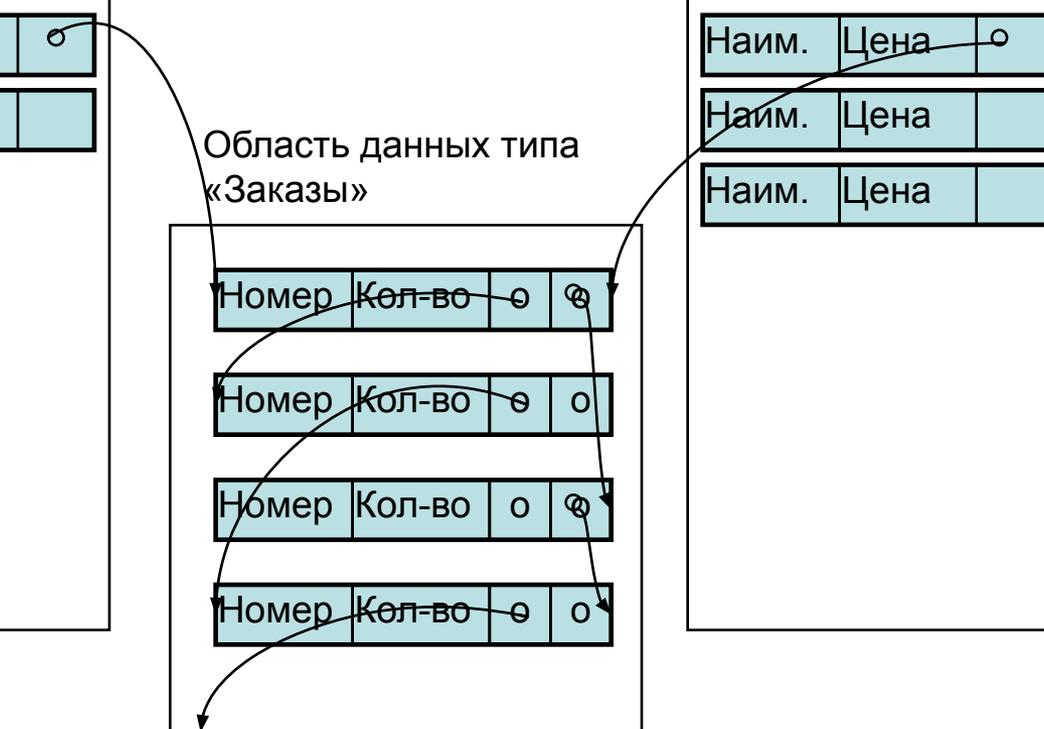
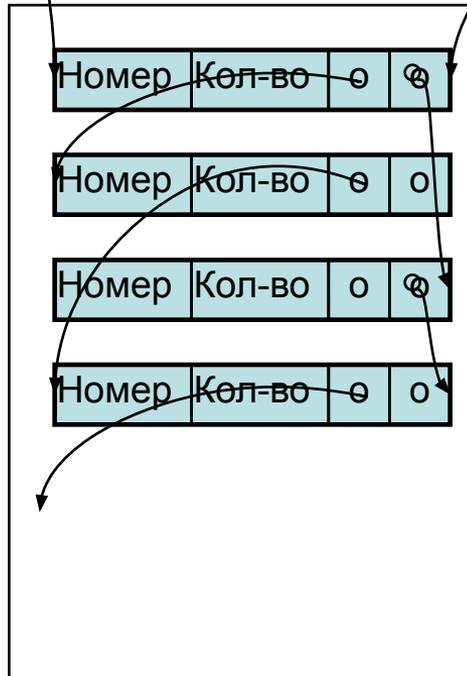
Область данных типа  
«Клиент»



Область данных типа  
«Товары»



Область данных типа  
«Заказы»



# Недостатки графовых моделей

**1. Жесткая структура, требующая распределения памяти между структурными элементами этих моделей**

**2. Запросы к БД реализовывались как навигационные программы.**