



Информационно-аналитические технологии государственного и муниципального управления

Домрачев С.А., доцент,
кандидат технических наук

Информационные

технологии

хранения

данных

БАЗЫ ДАННЫХ



Основные понятия

и принципы построения

Инструментальные средства хранения данных

Банк данных

Хранилище данных

База данных

Система управления БД

База знаний

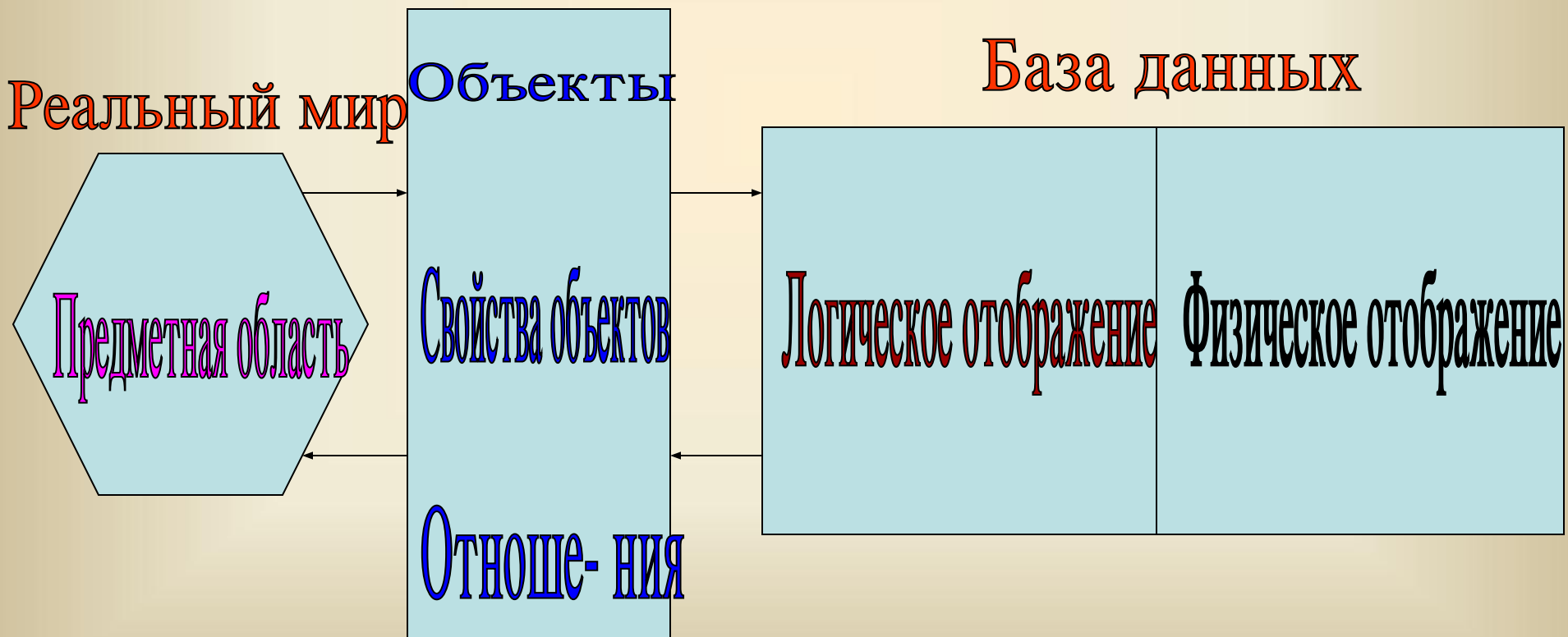
Определение базы данных

База данных

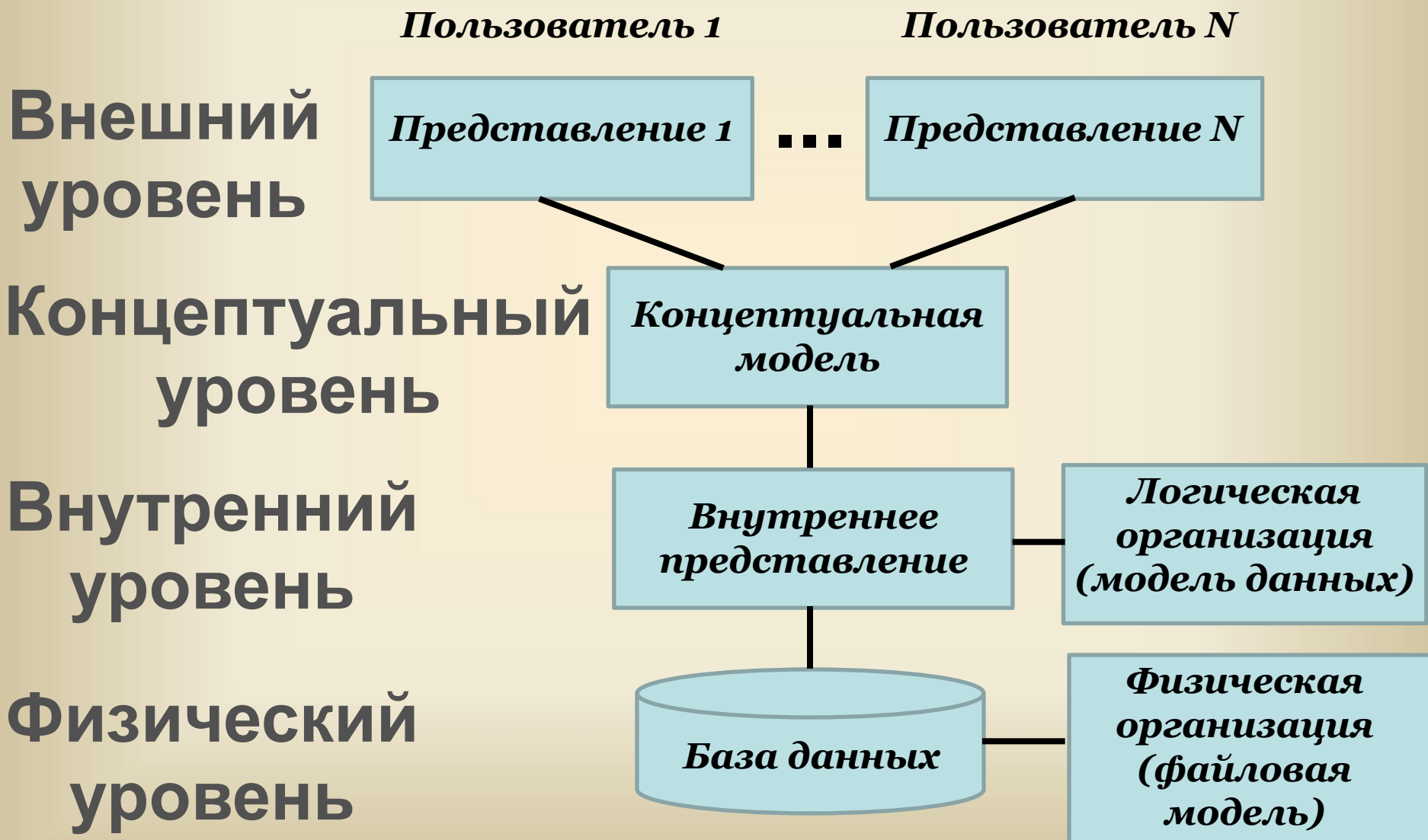
совокупность хранимых в памяти компьютера и специальным образом организованных, взаимосвязанных данных, отображающих состояние некоторой предметной области

Взаимосвязь предметной области и базы данных

Объектная система



Архитектура представления БД



Концептуальное моделирование



Отличия концептуальной и логической моделей данных

- 1 Концептуальная модель отображает отношения между сущностями, но не между экземплярами сущностей
- 2 Концептуальная модель не отображает динамику изменений предметной области
- 3 Концептуальная модель предметной области недостаточно формализована для ее представления в компьютере

Модели данных

Модель данных - средство формализованного представления данных и операций их обработки с учетом семантики ПО

Модель данных включает в себя:

- 1 Структуру данных
- 2 Операции над данными
- 3 Ограничения целостности

Операции над данными

Операции над данными - это допустимые действия над некоторой реализацией БД для перевода ее из одного состояния в другое

Любая операция состоит из двух компонент:

1. Селекция
2. Действие

Селекция определяет критерий отбора данных, над которыми должно быть произведено действие

Ограничения целостности

Ограничения целостности - это логические ограничения, накладываемые на данные

Ограничения вводятся в МД в целях :

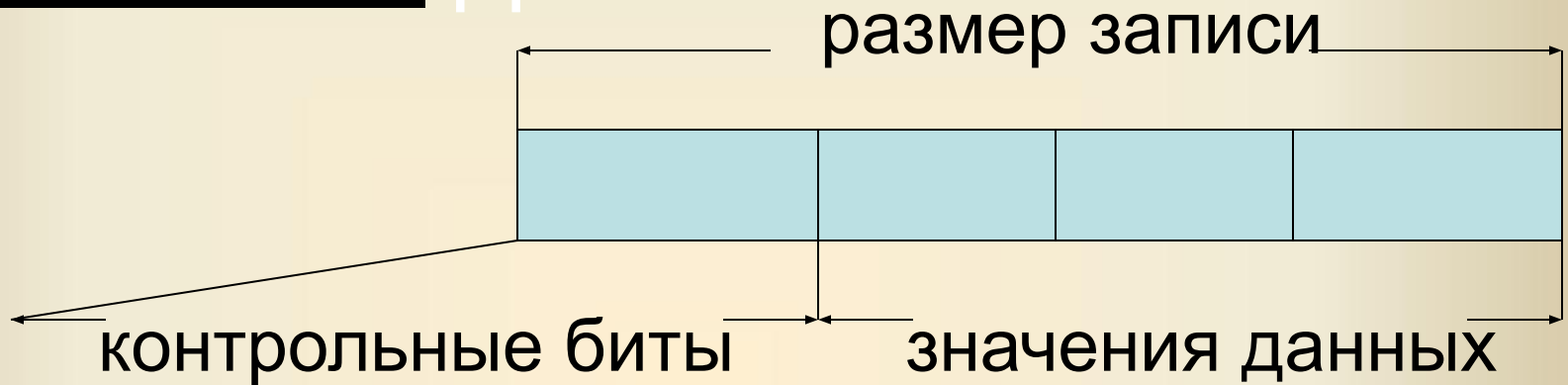
- I. Повышения ее семантичности и, как следствие, адекватности отражения ПО
- II. Поддержания целостности данных

Различают ограничения :

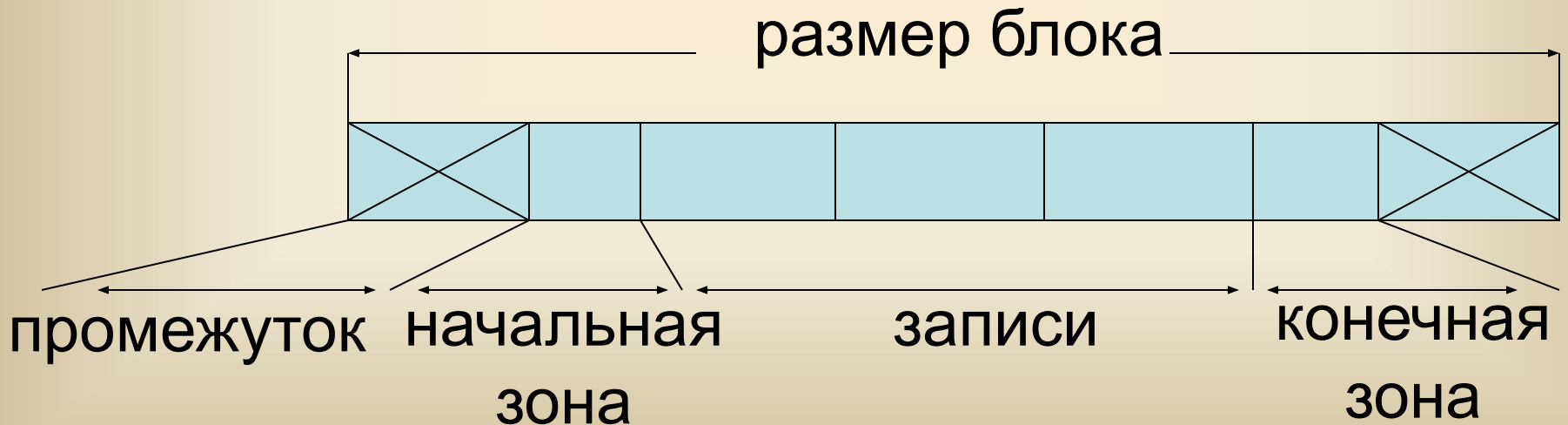
1. Внутренние
2. Явные

Физическая организация

Хранимая запись данных



Физический блок



Физическая организация данных

Способы размещения записей

- Последовательное распределение памяти
- Связанное распределение памяти
 - Ветвящаяся структура
 - Списковая структура

Физическая организация

данных Последовательное размещение



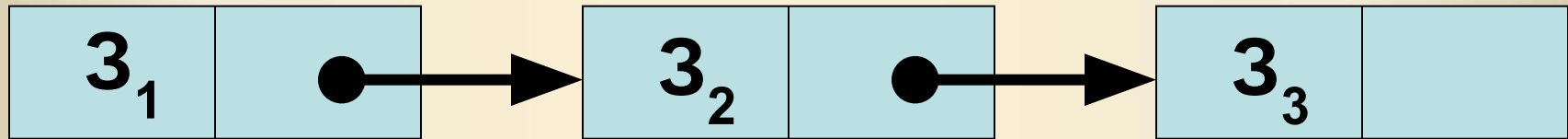
Особенности:

- Простота и естественность
- Экономия памяти (отсутствие указателей и промежутков между записями)
- Трудность модификации

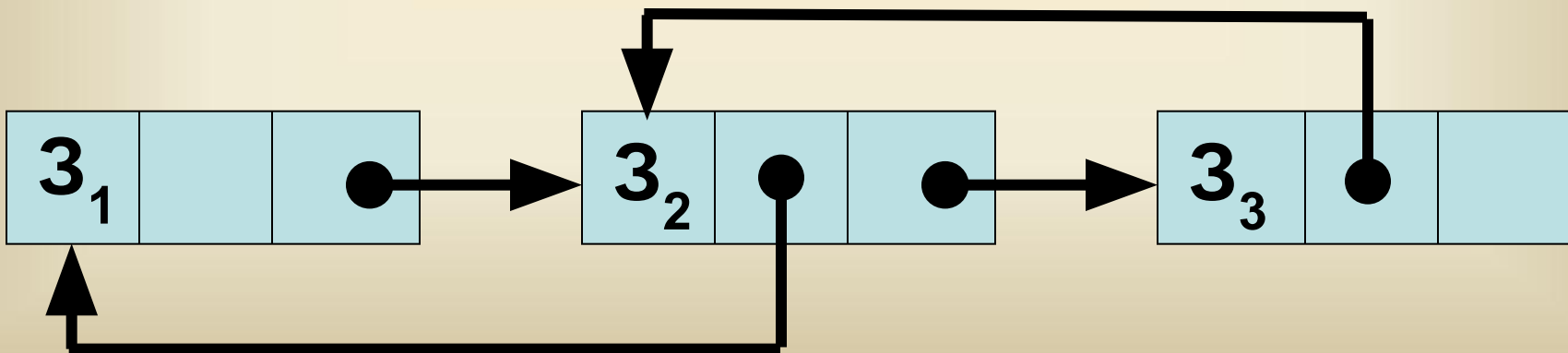
Физическая организация

Связанное размещение данных

Цепная структура



Структура с двунаправленными указателями



Физическая организация ветвящихся данных

ФИО	№ зачетки	Дисциплина
Иванов И.И.	1252	R+1

Список дисциплин	R+1	15	16	17

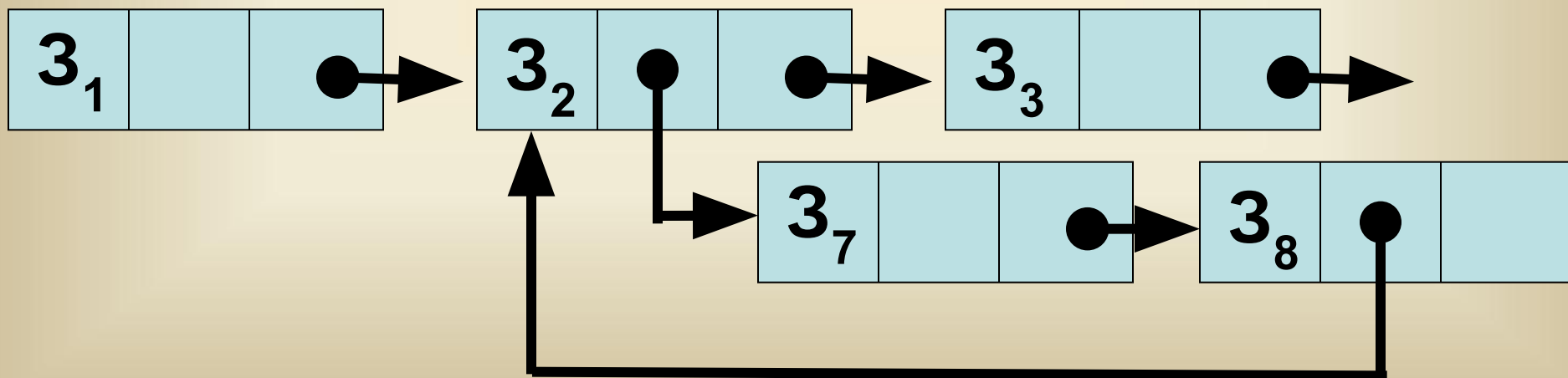
15	Информационные системы
16	Моделирование систем

Физическая организация

Списковая структура данных

Состав списка:

- Имя списка (является адресом или указателем на адрес первой позиции списка)
- Элементы данных
- Указатели на элементы данных



Типы

моделей

данных

Сетевая модель данных

Основные положения:

1. ПО рассматривается как совокупность элементов, имеющих между собой бинарные связи различных типов
2. Каждый элемент допускает аналогичную декомпозицию
3. Допускаются связи между различными уровнями записей
4. Ориентация связей в сетевой структуре не учитывается

Структурная схема сетевой модели данных

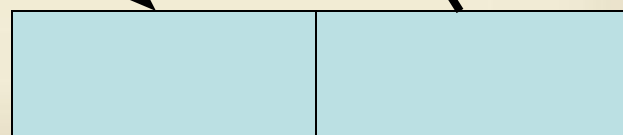
Деталь



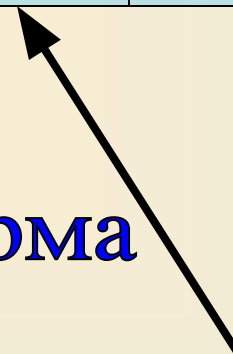
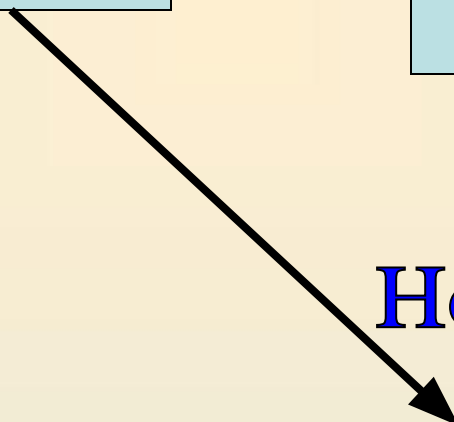
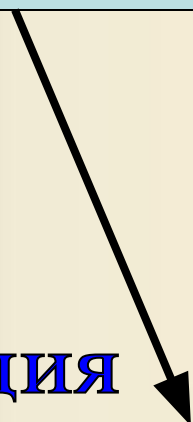
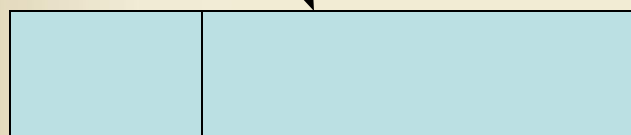
Материал



Норма



Операция



Сравнительная характеристика сетевой модели данных

Достоинства:

1. Универсальность. Выразительные возможности являются наиболее обширными в сравнении с остальными моделями
2. Возможность доступа к данным через значения нескольких отношений

Недостатки:

1. Сложность, т.е. обилие понятий, вариантов их взаимосвязей и особенностей реализации
2. Допустимость только навигационного принципа доступа к данным

Иерархическая модель данных

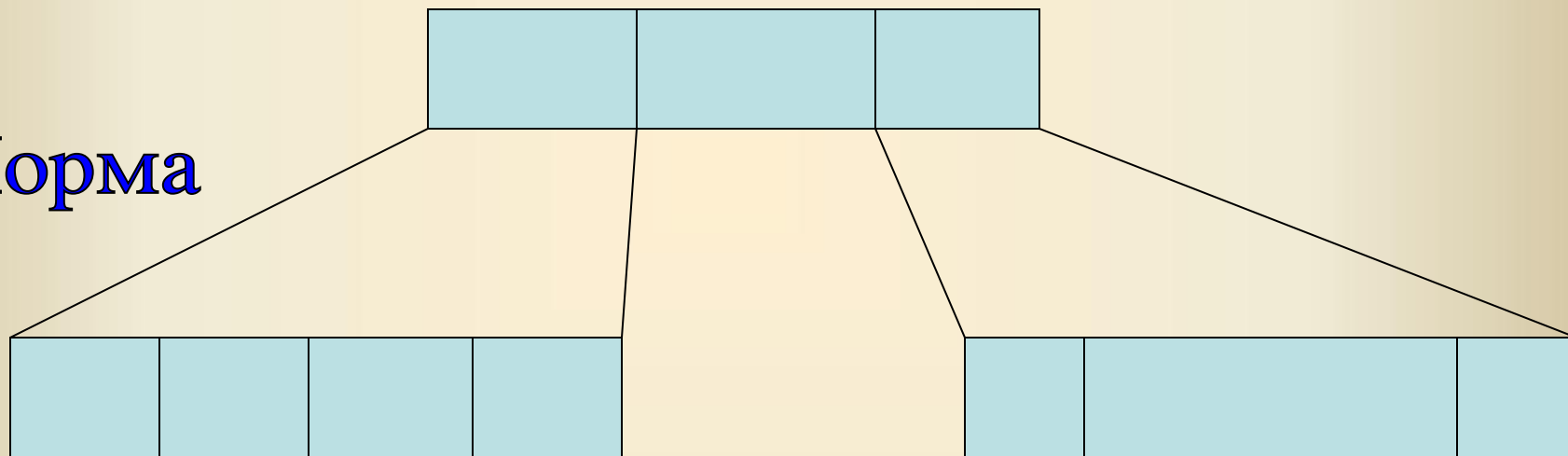
Основные положения:

1. Иерархия всегда начинается с корневого узла
2. На первом уровне иерархии находится только корневой узел
3. На нижних уровнях находятся порожденные (зависимые) узлы
4. Каждый порожденный узел (уровень i) связан только с одним исходным узлом (уровень $i-1$)
5. Каждый исходный узел может иметь один или несколько порожденных узлов, называемых подобными
6. Доступ к подобному узлу выполняется через его исходный узел
7. Существует единственный иерархический путь доступа к искомому узлу, начиная от корня дерева

Структурная схема иерархической модели данных

Деталь

Норма



Операция

Материал

Сравнительная характеристика иерархической модели данных

Достоинства:

1. Простота. Иерархический принцип соподчиненности понятий является естественным для многих практических задач
2. Минимальный расход памяти. Позволяет получить представление с наименьшим расходом физической памяти

Недостатки:

1. Неуниверсальность. Некоторые варианты связей невозможно реализовать или реализация сопряжена с повышенной избыточностью в БД
2. Допустимость только навигационного принципа доступа к данным

Реляционная модель данных

«Отношение» определяется как подмножество декартова произведения доменов.

Домен - некоторое множество элементов.

$$D = D_1 \times D_2 \times \dots \times D_k \quad , \text{ где}$$

$$D_1 = \{d_{11}, d_{12}, d_{13}, \dots, d_{1N}\};$$

$$D_2 = \{d_{21}, d_{22}, d_{23}, \dots, d_{2N}\};$$

...

$$D_k = \{d_{k1}, d_{k2}, d_{k3}, \dots, d_{kN}\};$$

Структурная схема реляционной модели данных

Поле_1	Поле_2	Поле_3	- - -	Поле_k
Запись_1				
Запись_2				
- - -				
Запись_N				

Свойства:

1. Каждая строка представляет собой кортеж из k значений
2. Порядок столбцов фиксирован
3. Любые две строки различаются хотя бы одним элемен-м
4. Строки и столбцы могут обрабатываться в любой последовательности

Сравнительная характеристика реляционной модели данных

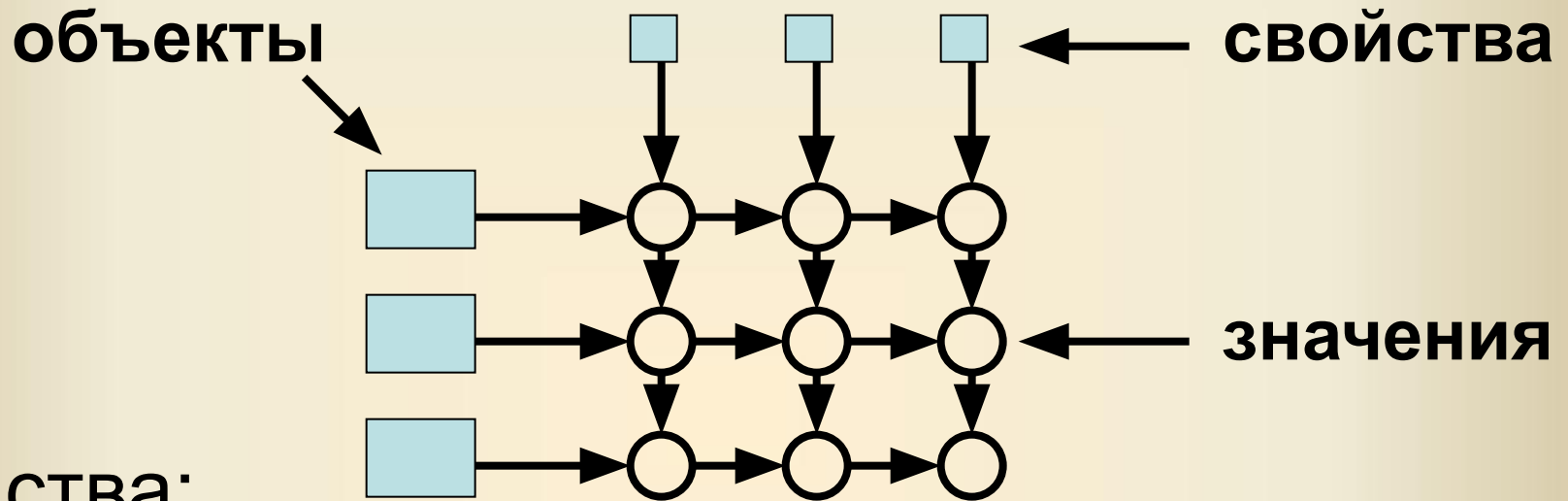
Достоинства:

1. Теоретическое обоснование. Наличие теоретически обоснованных методов нормализации отношений и проверки ацикличности структуры позволяет получать БД с заданными характеристиками
2. Независимость данных. Изменение структуры приводит к незначительным изменениям в ПП

Недостатки:

1. Низкая скорость при выполнении операций соединения
2. Достаточно большой расход физической памяти при реализации БД

Объектная модель данных



Свойства:

1. Состоит из совокупности объектов, характеризующихся некоторым набором свойств
2. Между объектами и характеристиками могут быть различные связи и отношения
3. При описании свойств фиксируют их название и значение

Классификация

баз данных

Классификация баз данных

По технологии обработки данных:

Базы данных

```
graph TD; A[Базы данных] --> B[Централизованные]; A --> C[Распределенные]
```

Централизованные

Распределенные

Классификация баз данных

По характеру обрабатываемой информации:

Базы данных

Документальны

е

Фактографически

е

Отличительные особенности

Документальные БД

- 2 слабоструктурированная информация
- 2 информационно-поисковый язык
- 2 критерии смыслового соответствия

Фактографические БД

- 2 жесткоструктурированная информация
- 2 язык описания данных
- 2 язык манипулирования данными

Документальные БД

Информационная потребность – потребность пользователя в определенной информации, возникающая в процессе его практической деятельности

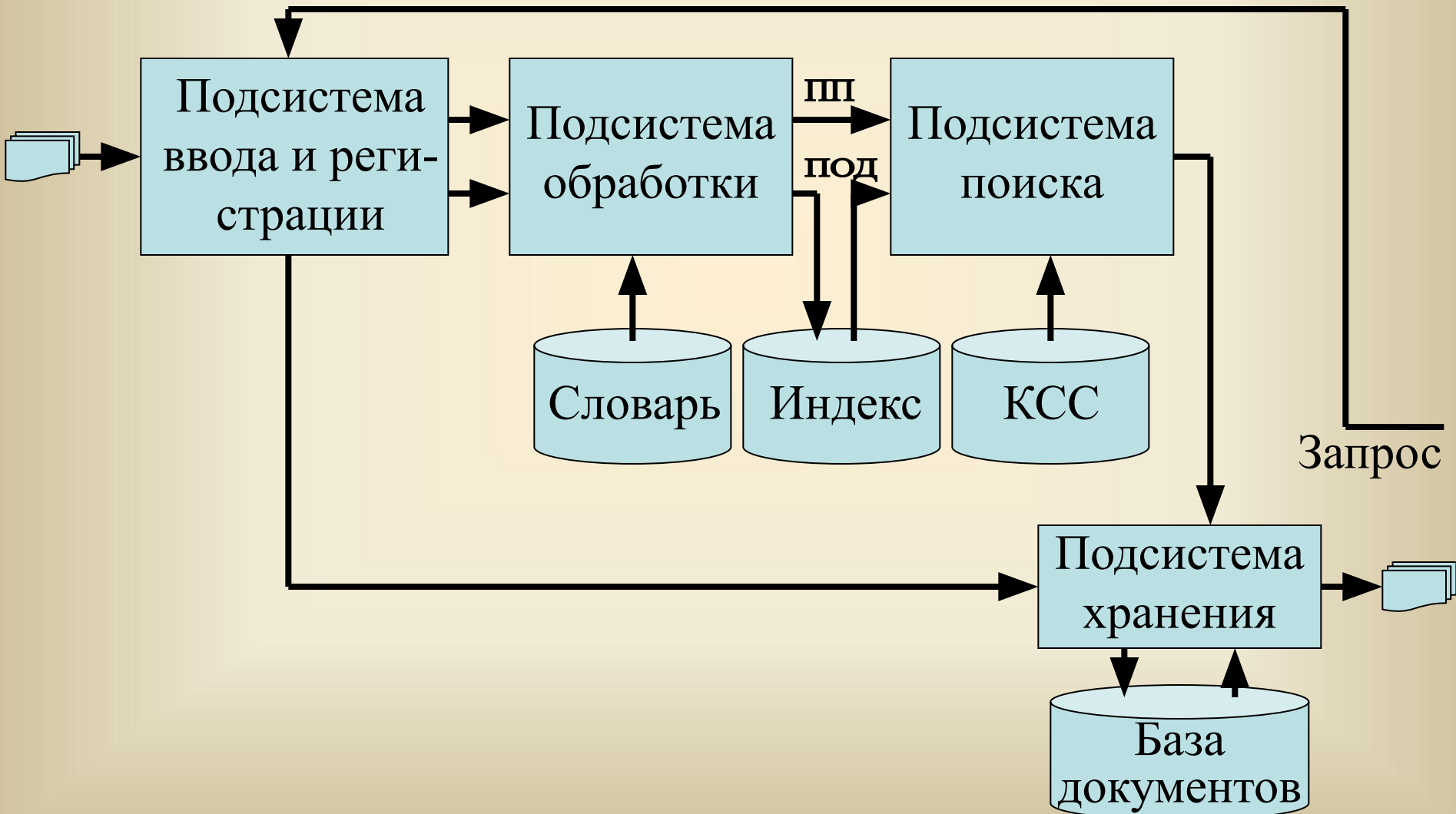
Информационный запрос – частное значение информационной потребности пользователя в определенный момент времени, выраженное на естественном языке

Пертинентность – соответствие смыслового содержания документа информационной потребности пользователя

Релевантность – соответствие содержания документа информационному запросу в том виде, в каком он сформулирован

Документальные БД

Структурно-функциональная схема документальной БД



Системы

управления

базами данных

Системы управления БД

СУБД - программная система, предназначенная для создания общей совокупности данных в памяти компьютера, поддержания ее в актуальном состоянии и обеспечения эффективного доступа пользователей к данным в рамках предоставленных им полномочий

Системы управления БД

Преимущества использования:

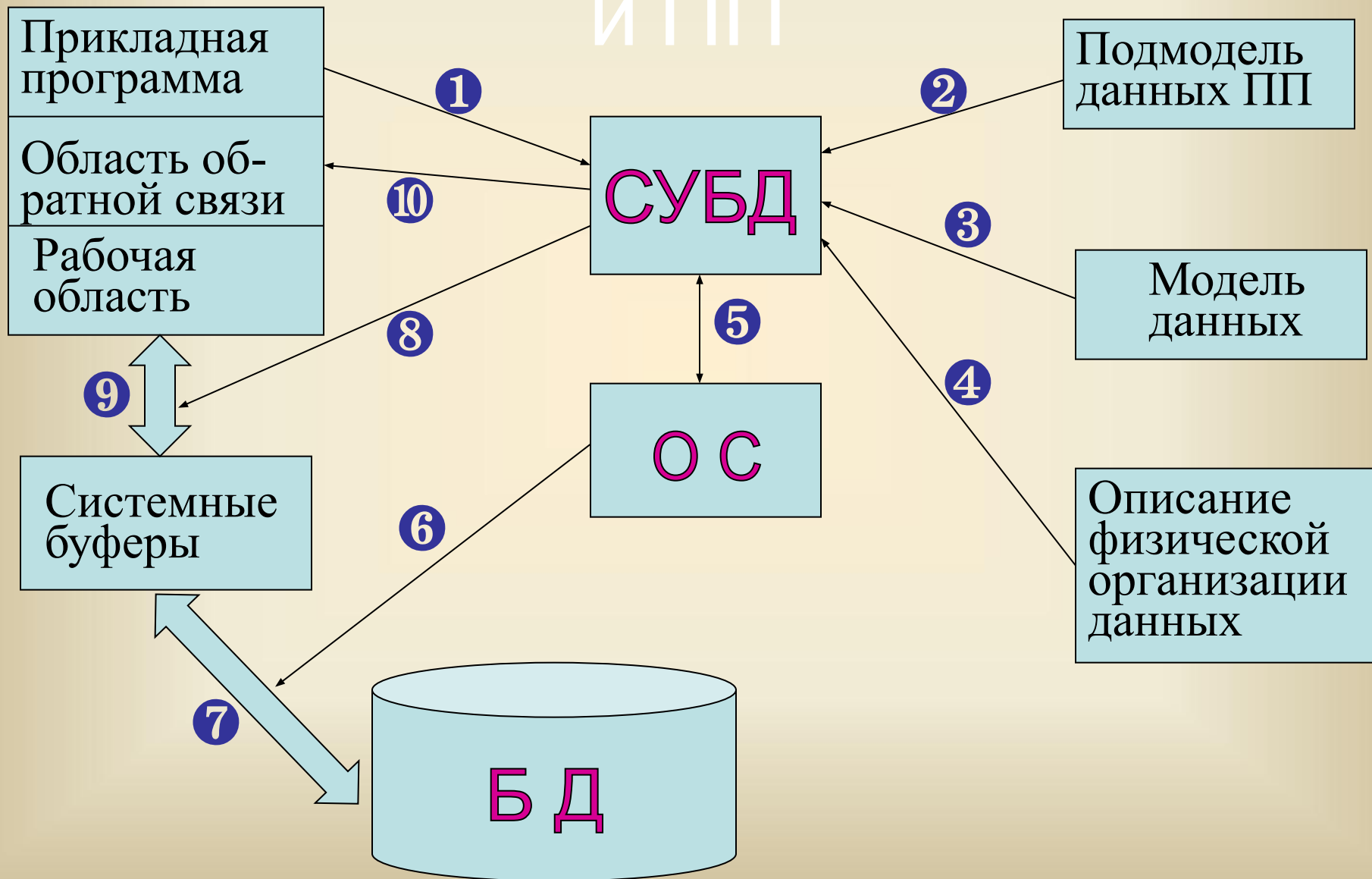
- а Совместное использование данных
- а Контроль избыточности данных
- а Обеспечение непротиворечивости данных
- а Поддержка целостности данных
- а Повышение безопасности данных
- а Возможность резервного копирования и восстановления данных

Системы управления БД

Основные функции:

- а Управление данными
- а Управление буферами ОП
- а Управление транзакциями
- а Восстановление БД после сбоев

Схема взаимодействия СУБД, ОС и ПП



Методология проектирования баз данных

Жизненный цикл БД

К основным этапам ЖЦ можно отнести:

- 1 Планирование разработки БД
- 2 Определение требований
- 3 Проектирование базы данных
- 4 Реализация базы данных
- 5
Тестирование

Требования к концептуальной модели

- 1 Адекватное отображение предметной области
- 2
- 3 **Непротиворечивость**
3 Однозначность трактовки всех компонентов модели
- 4 Конечность модели
- 5 Легкость модификации
- 6 Возможность композиции и декомпозиции

Методы разработки концептуальной модели

Модель «сущность-связь» (ER):

формализованное семантическое
описание предметной области

Сущность – концептуальное средство моделирования объекта ПО, о котором накапливается информация в БД

Атрибут – поименованная характеристика сущности

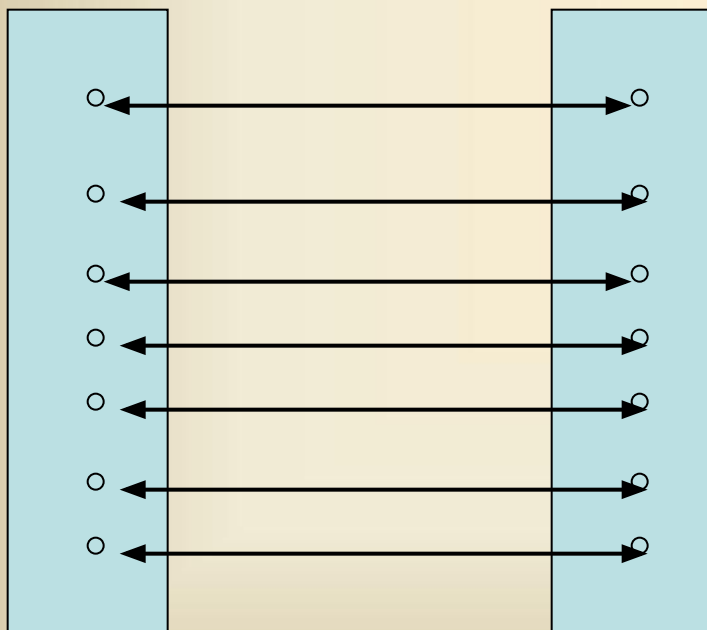
Связь – средство представления отношений между сущностями

Типы отношений

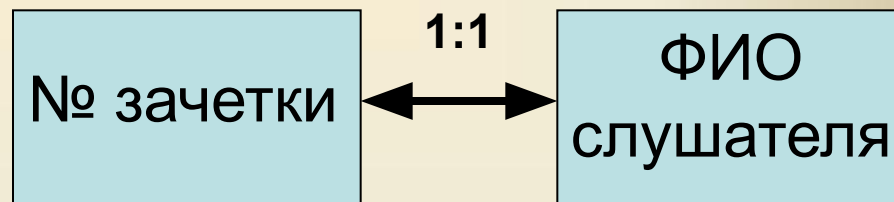
Отображение 1:1

Тип А

Тип В



Пример

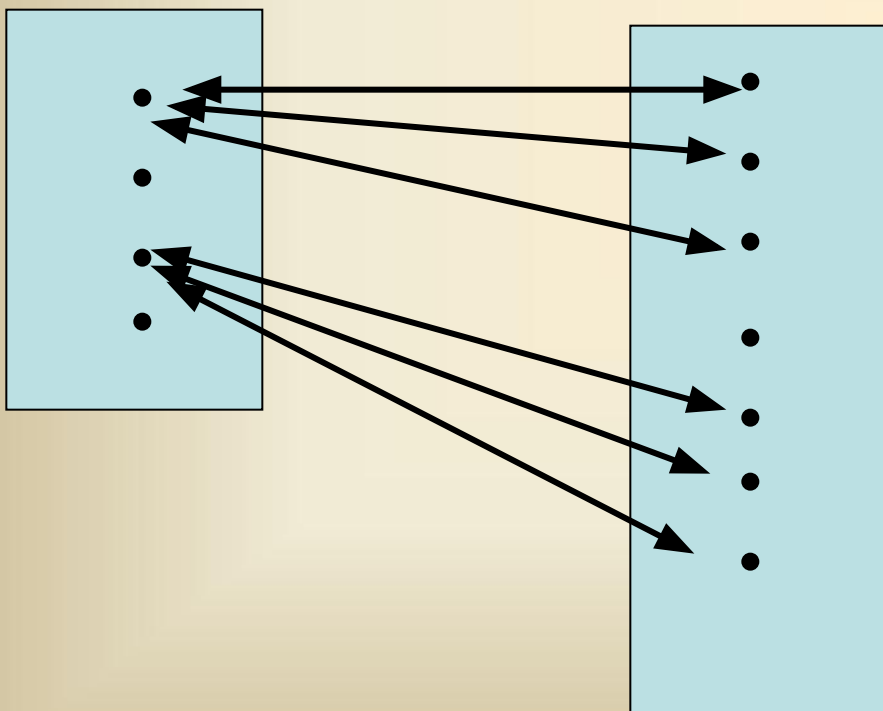


Типы отношений

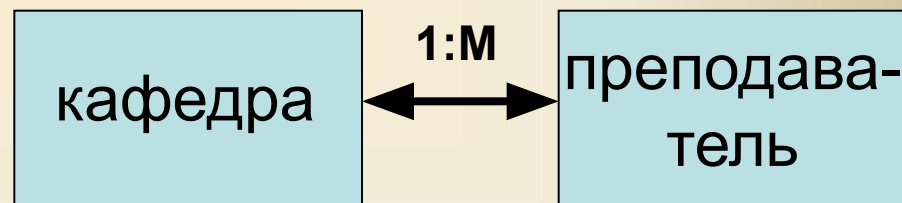
Отображение 1:M

Тип А

Тип В



Пример

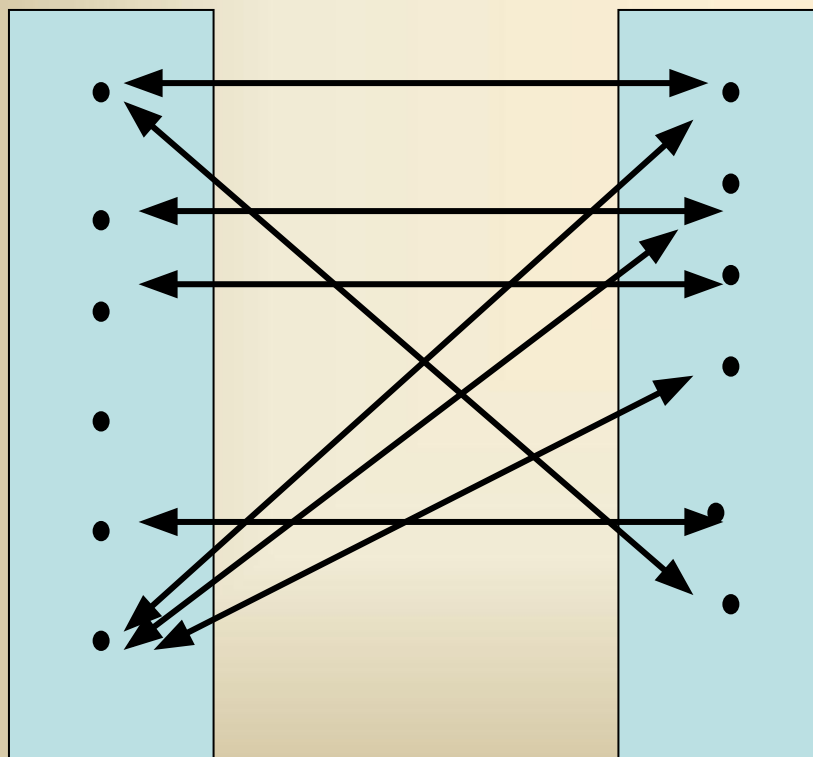


Типы отношений

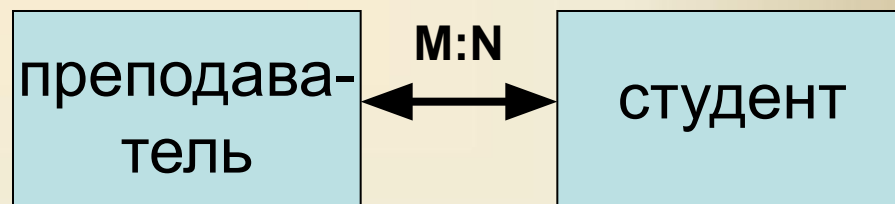
Отображение M:N

Тип А

Тип В



Пример



Методология

IDEF 1X

Определение сущностей

EntityName /1

EntityName /2

Независимая сущность - экземпляр сущности, который может быть однозначно идентифицирован без определения его отношений с другими сущностями

Зависимая сущность - экземпляр сущности, однозначная идентификация которого зависит от его отношения к другой сущности

Правила:

1. Каждая сущность должна иметь уникальное имя
2. Сущность обладает одним или несколькими атрибутами, которые принадлежат сущности, либо наследуются через отношения
3. Сущность обладает одним или несколькими атрибутами, которые однозначно идентифицируют экземпляр сущности
4. Каждая сущность может обладать любым количеством отношений

Определение атрибутов

имя_атрибута
[имя_атрибута]

[имя_атрибута]
[имя_атрибута]

• • •

} Атрибуты первичного ключа

Ключи:

Первичный ключ - один (или несколько атрибутов), чьи значения однозначно определяют каждый экземпляр сущности

Вторичный ключ - собственные атрибуты сущности, характеризующие те или иные ее

Внешний ключ - атрибуты первичного ключа сущности-родителя, наследуемые сущностями-потомками

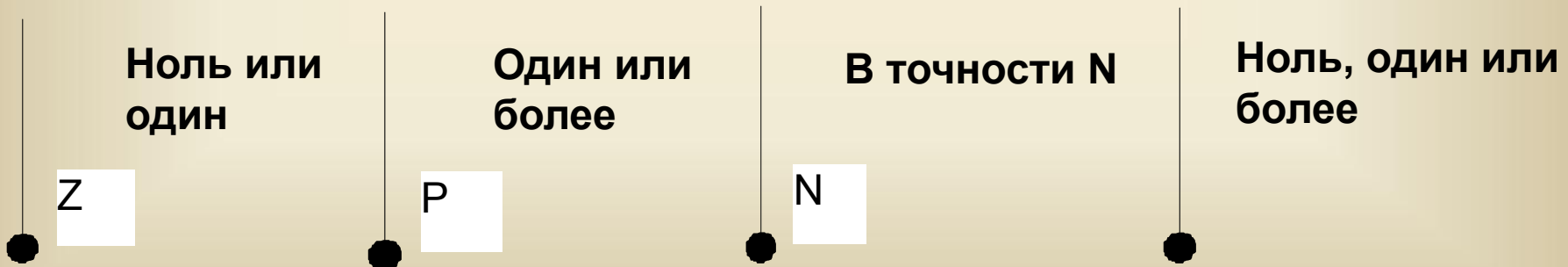
Определение атрибутов

Правила атрибутов

1. Каждый атрибут должен иметь уникальное имя с уникальным смыслом
2. Сущность может обладать любым количеством атрибутов
3. Сущность может обладать любым количеством наследуемых атрибутов, но наследуемый атрибут должен быть частью первичного ключа
4. Для каждого экземпляра сущности должно существовать значение ее каждого атрибута
5. Ни один из экземпляров сущности не может обладать более чем одним значением связанного с ней атрибута

Определение отношений

Отношение «родитель-потомок» : связь между сущностями, при которой каждый экземпляр одной сущности (родитель), ассоциирован с произвольным количеством экземпляров другой сущности (потомки), а каждый экземпляр сущности-потомка ассоциирован в точности с одним экземпляром сущности-родителя



Определение отношений

Типы отношений

Идентифицирующее отношение – сущность-потомок однозначно определяется своей связью с сущностью-родителем

Неидентифицирующее отношение – сущность-потомок однозначно определяется без относительно его связи с сущностью-родителем

Сущность-родитель

Сущность-потомок

Сущность А / 1

Ключевые_атрибуты_А

Имя отношения

Сущность В / 2

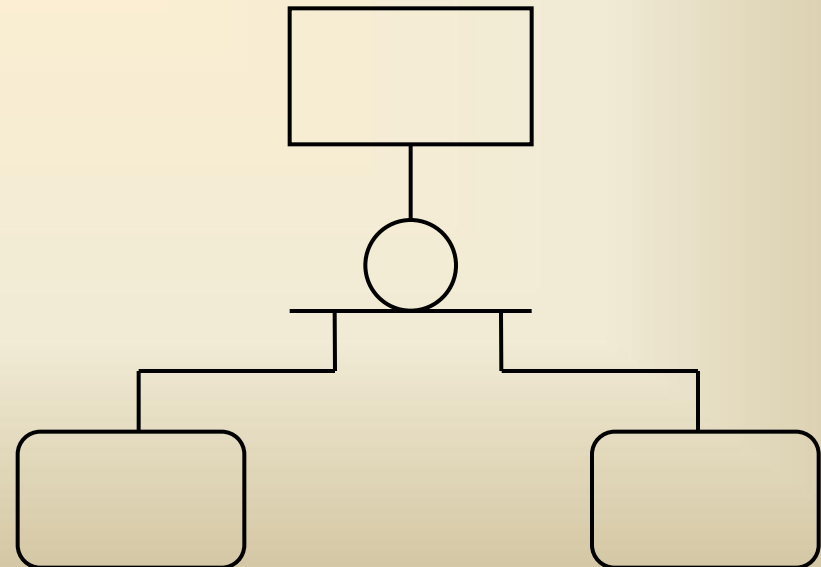
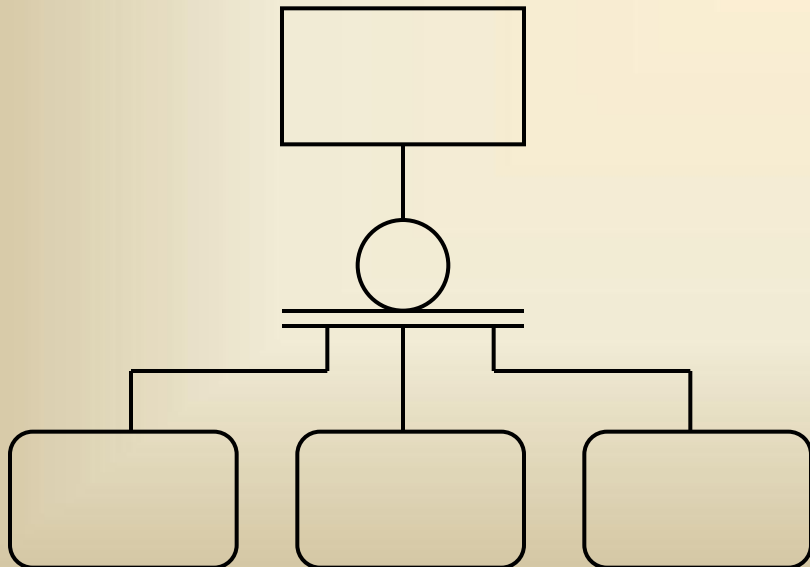
Ключевые_атрибуты_А
Ключевые_атрибуты_В

Определение отношений

Типы отношений

Отношение категоризации – отношение между сущностью, называемой общей-сущностью, в которой каждый экземпляр связан в точности с одним экземпляром одной и только одной из других сущностей, называемых сущностями-категориями

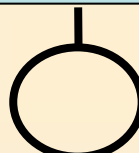
Полное множество категорий Неполное множество категорий



Пример отношения категоризации

Лектор / 1

Идентифик_лектора
ФИО_лектора Адрес_лектора Ученая_степень



Штатное_время /1

Идентифик_лектора
Месячный_оклад

Совместительство /2

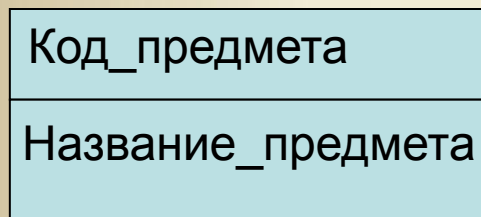
Идентифик_лектора
Часть_месячн_оклд

Внештатное_время/3

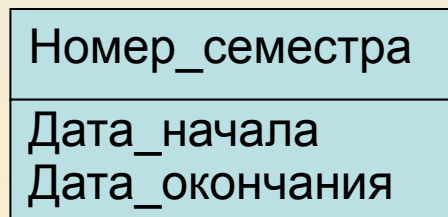
Идентифик_лектора
Почасовая_оплата

Пример информационной модели

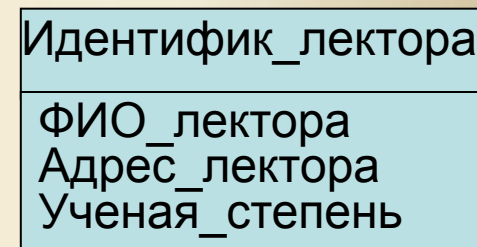
Предмет / 1



Семестр / 2



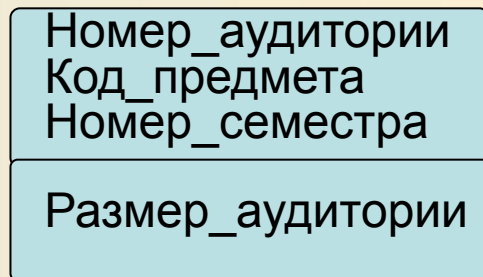
Лектор / 3



имеет

Р

Аудитория / 4

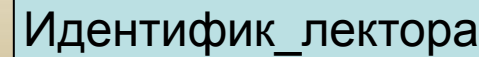
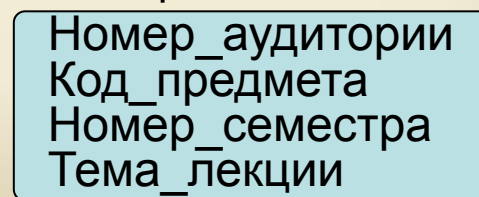


имеет

Неидентифицирующее отношение

читает

Лекция / 5



используется для

Идентифицирующее отношение



Инструментальные средства разработки концептуальных моделей

При разработке ER-моделей могут быть использованы следующие CASE средства:

- ¿ ProKit WORKBENCH
- ¿ Design/IDEF
- ¿ Designer 2000
- ¿ Power Designer
- ¿ All Fusion (ERWin)
- ¿ ERStudio

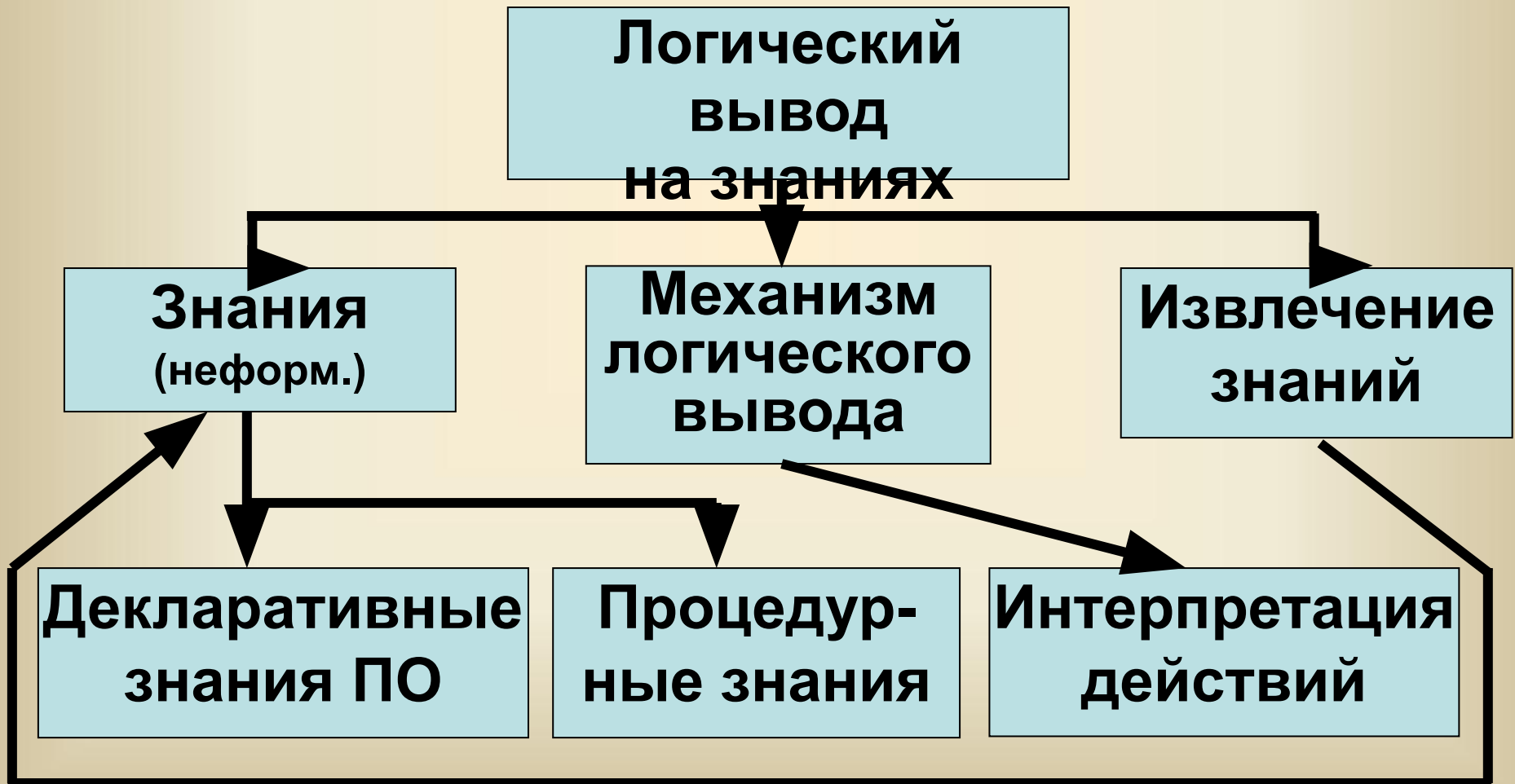
Инструментальные средства разработки концептуальных моделей

Требования к системам разработки

- 1 Число и перечень поддерживаемых СУБД
- 2 Поддержка коллективной работы при проектировании
- 3 Построение модели по описанию структуры БД (ремоделирование)
- 4 Поддерживаемые виды моделей
- 5 Открытость системы
- 6 Наличие средств для моделирования хранилищ данных

Базы знаний

Представление неформализованных задач



Модели представления знаний

Все модели представления знаний можно свести к следующим основным классам:

1. Продукционные модели
2. Формальные логические модели
3. Фреймы

Фреймы

Фрейм - это абстрактный образ для представления некоего стереотипа восприятия

Модель фрейма является универсальной, т.к. позволяет отобразить все многообразие знаний о мире через:

1. Фреймы-*структуры*
2. Фреймы-*роли*
3. Фреймы-*сценарии*
4. Фреймы-*ситуации*

Структура фрейма

Имя фрейма			
Имя слота	Значение слота	Способ получения значения	Присоединенная процедура

Способы получения значения слотом:

1. По умолчанию от фрейма-образца
2. Через наследование свойств
3. По формуле указанной в слоте
4. Через присоединенную процедуру
5. Явно из диалога с пользователем
6. Из базы данных

Хранилища данных

Определение хранилища данных

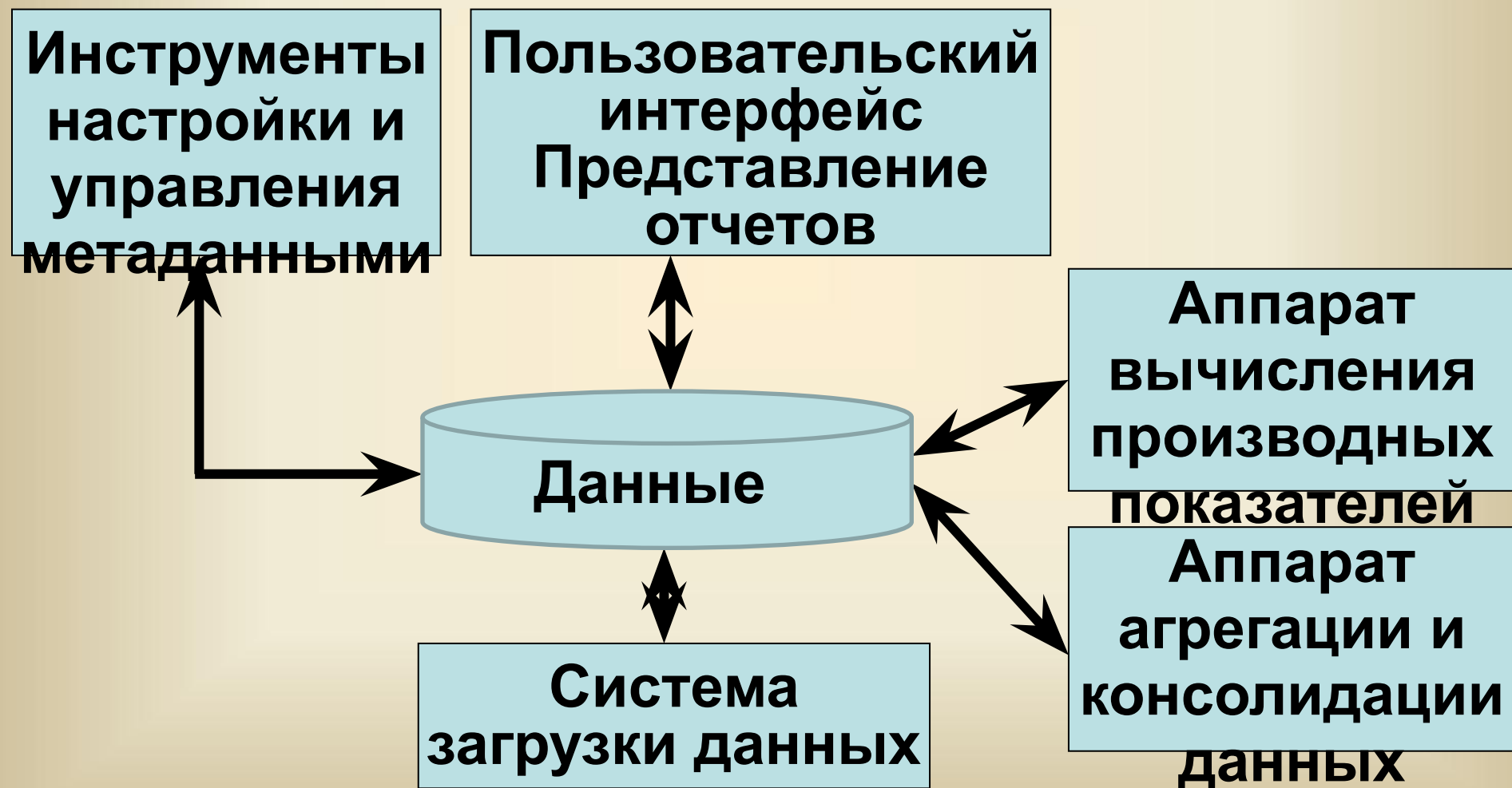
Концепция хранилища данных – подготовка данных для последующего анализа

Основные положения концепции:

- 1 Интеграция и согласование данных из различных источников
- 2 Разделение наборов данных

Хранилище данных – это предметно-ориентированный, интегрированный, неизменяемый и поддерживающий хронологию набор данных, предназначенный для обеспечения принятия управленческих решений

Архитектура хранилища данных



Свойства хранилищ данных

Свойства присущие хранилищам данных:

- 1 Ориентация на специфику предметной области
- 2 Интегрированность
- 3 Неизменяемость данных
- 4 Поддержка хронологии

OLAP технология

***OLAP** (On Line Analytical Processing) – инструмент оперативного анализа данных, содержащихся в хранилище в режиме реального времени*

Основные особенности технологии OLAP:

- I. Многомерное представление данных
- II. Интуитивное манипулирование данными
- III. Детализация и агрегирование данных
- IV. Пакетное извлечение данных (отчеты)
- V. Различные модели анализа
- VI. Неограниченное число измерений и уровней представления