

Лекция 2

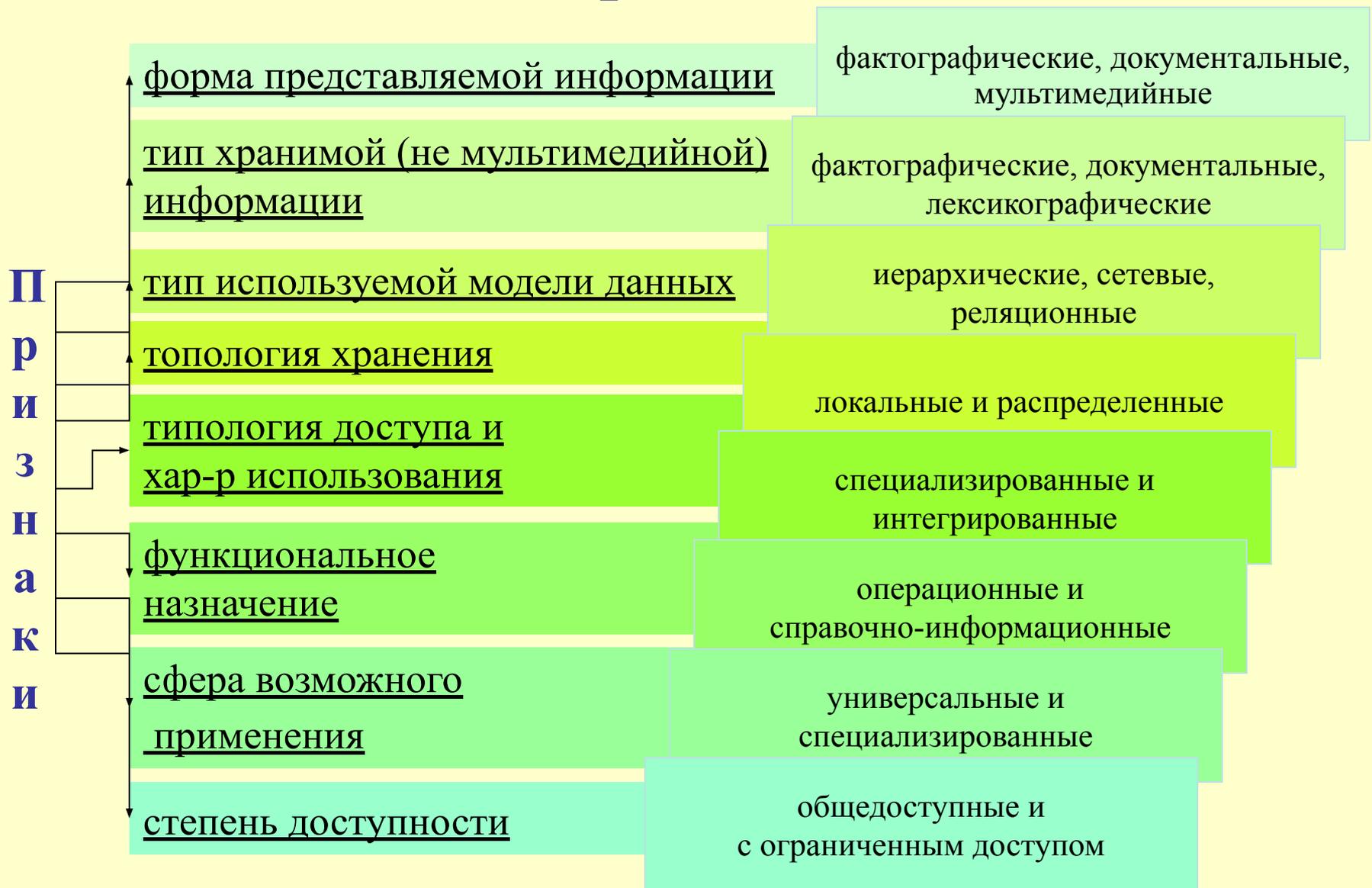
**Классификация БД. Фактографические
и документальные БД.**

**БД оперативной и ретроспективной
информации. Хранилища данных.**

Локальные и распределенные БД.

**Соотношение основных требований и
свойств СУБД: система компромиссов**

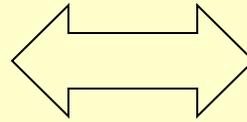
Классификация БД



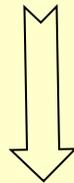
**Уровни
информационных
процессов**

**Основные
характеристики БД**

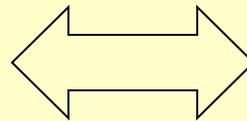
**Уровень
информационных
технологий (ИТ)**



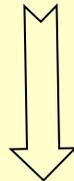
*физическая
структура БД*



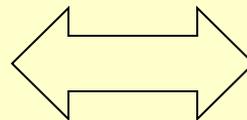
**Уровень
системы (ИС)**



*логическая
структура БД*



**Уровень
Информационных
ресурсов (ИР)**



содержание БД

Фактографические БД	Документальные БД
ПрО декомпозируется на объекты и связи	ПрО представляет собой совокупность в общем не взаимодействующих объектов
Объект характеризуется фиксированным набором атрибутов, большая часть из которых носит обязательный характер	Набор атрибутов для объекта конечен, но не фиксирован
Атрибут имеет атомарное значение	Атрибут может иметь множество значений
Значение атрибута не зависит от контекста	Значение атрибута может определяться контекстом

Поиск информации и поиск данных

Свойство	Поиск данных	Информационный поиск
Совпадение компонентов запроса и единицы хранения информации при поиске	полное	неполное
Метод вывода результата поиска	индукция	дедукция
Тип модели поиска	детерминированная	вероятностная
Язык запросов	искусственный	естественный
Спецификация запроса	полная	неполная
Поиск единиц хранения информации	путем совпадения	расчетом релевантности
Наличие «шума» в результатах поиска	не допускается	допускается

Хранилище данных –

предметно-ориентированный,
интегрированный,
привязанный ко времени,
неизменяемый набор данных,
предназначенный для поддержки принятия решений.

Сетевое хранилище данных –

распределенное хранилище данных,
реализованное в среде Web и не имеющее
центрального репозитария данных.

Система OLTP

Хранилище данных

Содержит текущие данные

Содержит исторические данные

Хранит подробные сведения

Хранит подробные сведения, а также частично и полностью обобщенные данные

Данные являются динамическими

Данные в основном являются статическими

Повторяющийся способ обработки данных

Нерегламентированный, неструктурированный и эвристический способ обработки данных

Высокая интенсивность обработки транзакций

Средняя и низкая интенсивность обработки транзакций

Предсказуемый способ использования данных

Непредсказуемый способ использования данных

Предназначена для обработки транзакций

Предназначено для проведения анализа

Ориентирована на прикладные области

Ориентировано на предметные области

Поддержка принятия повседневных решений

Поддержка принятия стратегических решений

Обслуживает большое количество работников

Обслуживает относительно малое количество работников руководящего звена

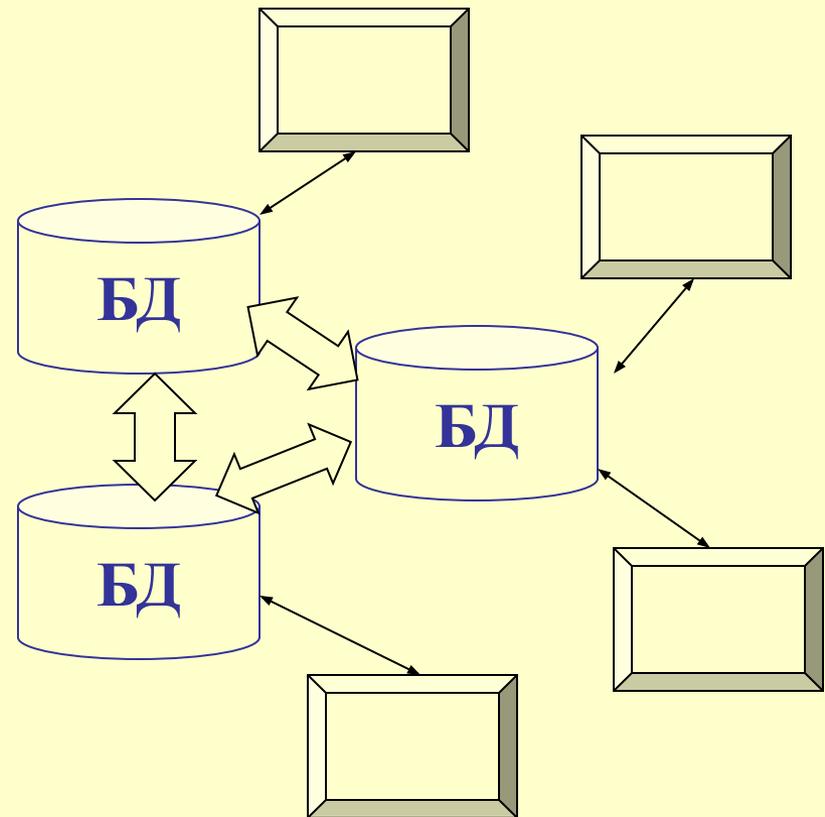
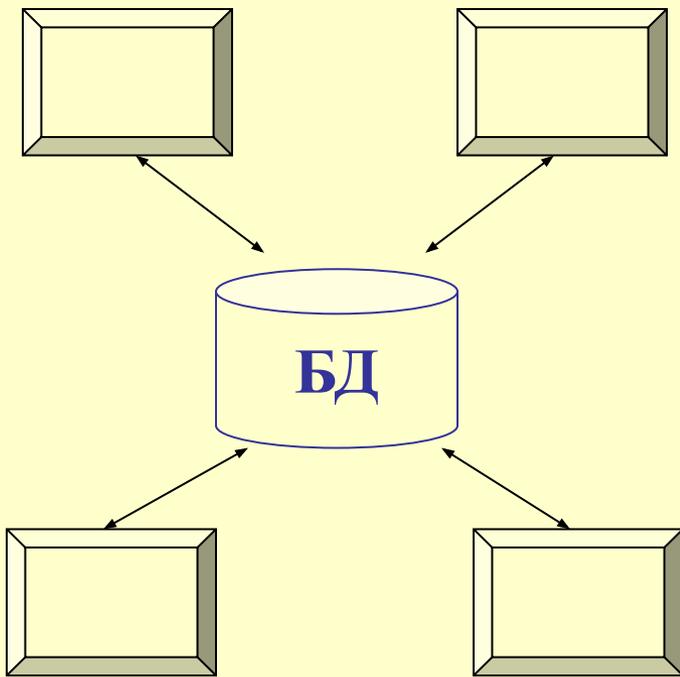
Проблемы разработки и сопровождения хранилищ данных

- Недооценка ресурсов, необходимых для загрузки данных
- Скрытые проблемы источников данных
- Отсутствие требуемых данных в имеющихся архивах
- Повышение требований конечных пользователей
- Унификация данных
- Высокие требования к ресурсам
- Владение данными
- Сложное сопровождение
- Долговременный характер проектов
- Сложности интеграции

Технологии обработки данных

*Системы распределенной
обработки данных*

*Системы распределенных
баз данных*



- ? Каким образом сложные нелинейные структуры данных представить в виде линейных – наиболее соответствующих принципу последовательного представления (хранения) в машинной памяти.
- ? Каким образом организовать данные, чтобы была возможность эффективного внесения, удаления и редактирования данных.
- ? Как организовать данные, чтобы использование пространства памяти (плотность данных) было достаточно рациональным, а скорость доступа к записям данных – высокой.
- ? Каким образом организовать данные, чтобы поиск был эффективным и позволял отыскивать записи по нескольким ключам.

Компромиссы создания БД

