

Технология баз данных



Любая задача обработки и принятия решений может быть представлена в виде схемы:



Рис. 1.11. Схема решения задач обработки информации и принятия решений:

Последовательность операций обработки данных называют информационной технологией (ИТ). В силу значительного количества информации в современных задачах она должна быть упорядочена. Существует два подхода к упорядочению.

1. Данные связаны с конкретной задачей (технология массивов) — упорядочение по использованию. Вместе с тем алгоритмы более подвижны (могут чаще меняться), чем данные. Это вызывает необходимость переупорядочения данных, которые к тому же могут повторяться в различных задачах.

2. Другая, более широко используемая технология баз данных, — упорядочение по хранению.

Под базой данных (БД) понимают совокупность хранящихся вместе данных при наличии такой минимальной избыточности, которая допускает их использование для одного или нескольких приложений.

Целью создания баз данных, как разновидности информационной технологии и формы хранения данных, является построение системы данных, не зависящих от принятых алгоритмов (программного обеспечения), применяемых технических средств и физического расположения данных в компьютере; обеспечивающих непротиворечивую и целостную информацию при нерегламентируемых запросах. БД предполагает многоцелевое ее использование (несколько пользователей, множество форм документов и запросов одного пользователя).

База знаний (БЗ) представляет собой совокупность БД и используемых правил, полученных от лиц, принимающих решения.

Банк данных – база данных и система управления ею (СУБД). СУБД (например, ACCESS) представляет собой приложение для создания баз данных как совокупности двумерных таблиц.

Приложение – программа или группа программ, предназначенных для выполнения стандартных работ. К приложениям относятся текстовые (например, Word), графические редакторы, электронные таблицы.



Модели данных:

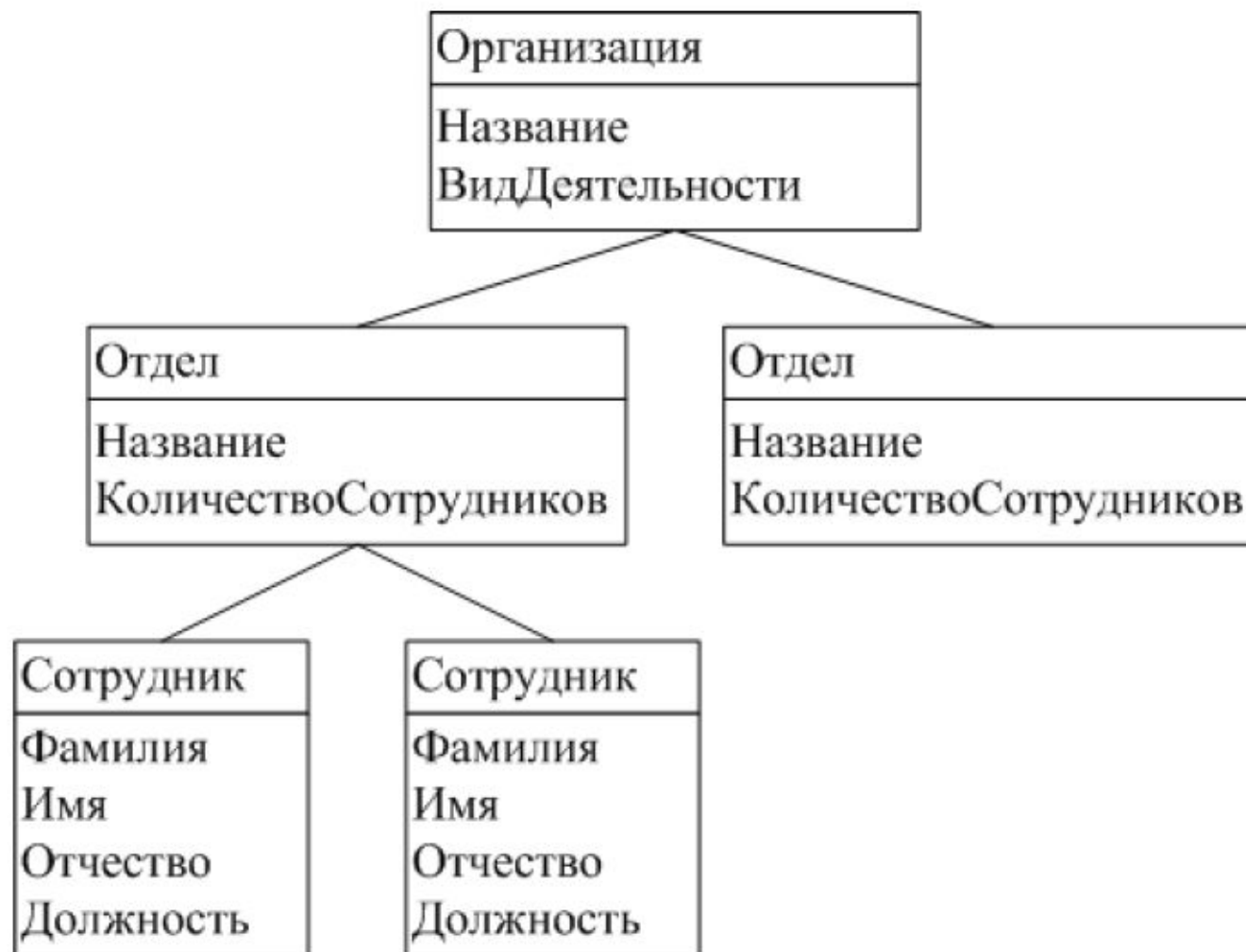


Рис. 4.1. Схема данных в иерархической модели данных

Модели данных: сетевая

Сетевая модель данных является развитием иерархической и отличается возможностью создания нескольких связей от одного потомка к разным родителям. Таким образом, схемой данных становится не дерево, а граф (рисунок 4.2).

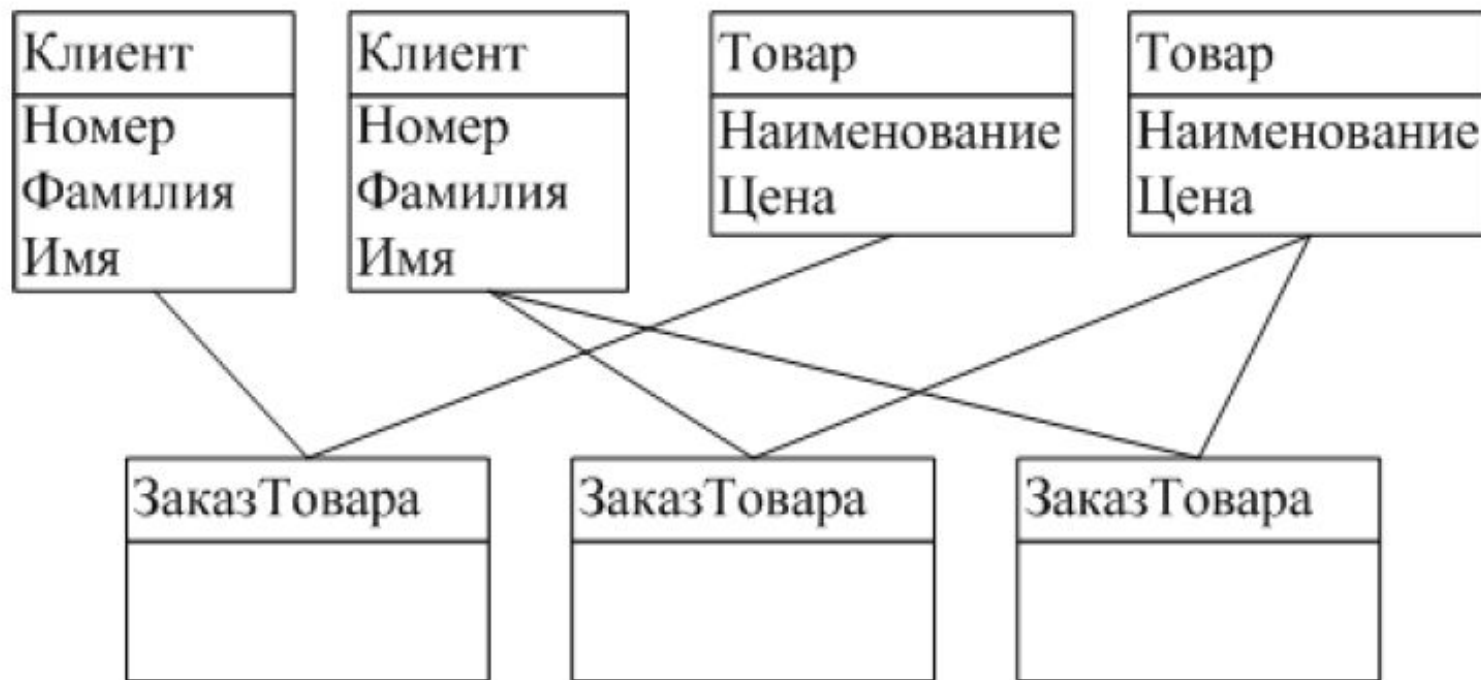


Рис. 4.2. Схема данных в сетевой модели данных

Модели данных: реляционная

Отношение (Таблица) Сотрудники

КодСотрудника	Фамилия	Имя	НомерОтдела
5	Попов	Артем	3
6	Иванов	Николай	5
7	Петров	Денис	2
8	Соколов	Дмитрий	4
9	Иванова	Анна	1
10	Кузнецов	Михаил	5
11	Павлов	Егор	1
12	Орлов	Иван	5

Атрибут (Столбец)

Фамилия

Попов
Иванов
Петров
Соколов
Иванова
Кузнецов
Павлов
Орлов

КодСотрудника	Фамилия	Имя	НомерОтдела
---------------	---------	-----	-------------

Заголовок отношения

10	Кузнецов	Михаил	5
----	----------	--------	---

Кортеж (Строка)

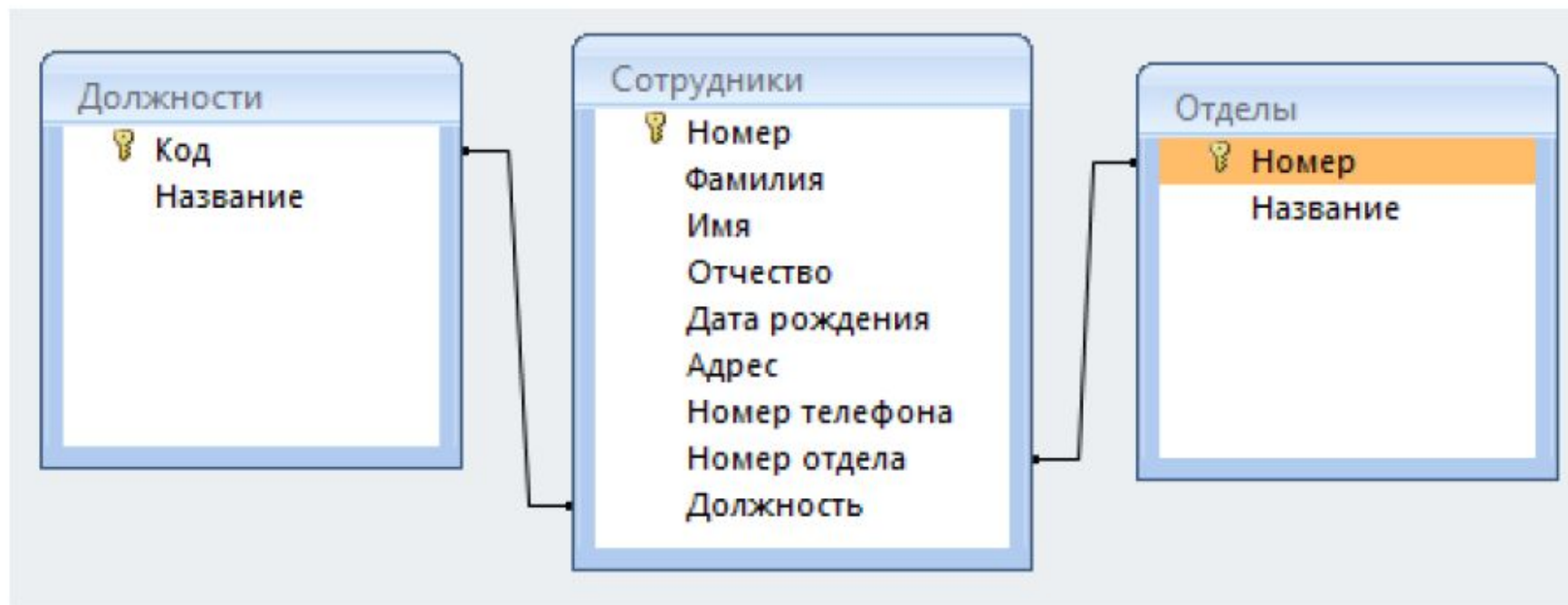
Фамилия	Наименование атрибута
---------	-----------------------

КодСотрудника	Первичный ключ
---------------	----------------

Рис. 4.3. Основные элементы реляционной модели данных

Внешний ключ используется для установления связи между таблицами. Рассмотрим следующий пример, в таблице «Сотрудники» содержится информация о сотрудниках: номер, фамилия, имя, отчество, оклад, должность, номер отдела. Атрибуты «Должность» и «Номер отдела» в таблице «Сотрудники» являются внешними ключами, в то время как в таблицах «Должности» и «Отделы» эти атрибуты являются первичными ключами.

Схема данных иллюстрирующая данный пример представлена на рисунке 4.4.



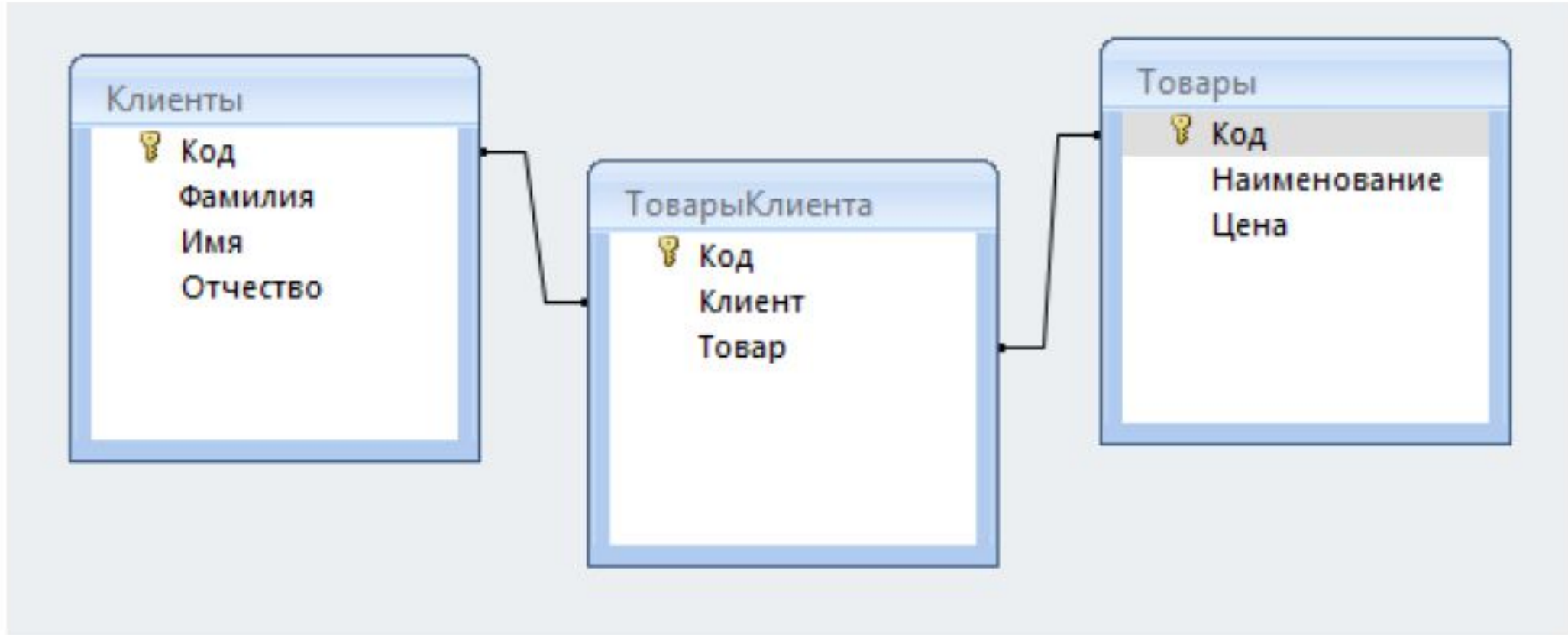


Рис. 4.5. Связь многие ко многим

Пользователь	
Ключ (Key)	Атрибуты (Attributes)
1	Имя: Иван Фамилия: Иванов Отчество: Иванович Пол: Мужской Дата рождения: 23.03.1985
2	Имя: Анна Фамилия: Петрова Отчество: Ивановна Пол: Женский Дата рождения: 18.05.1987 Телефон: 123456789

Рис. 4.6. Записи в хранилище типа ключ-значение

Основные понятия и определения

Поле – столбец документа (таблицы). Имя поля часто называют атрибутом.

Домен – совокупность значений одного поля.

Универсум – совокупность значений всех полей.

Запись – строка таблицы (в реляционной модели данных); в сетевой модели данных - элемент структуры, аналогичный таблице в реляционной модели данных.

Ключ – поле с уникальными (неповторяющимися) записями, используемое для определения места расположения записи. Ключ может состоять из совокупности полей (составной ключ). Иногда ключ называют идентификатором.

Основные понятия и

определения

Администратор базы данных (АБД) — лицо, отвечающее за выработку требований к БД, её проектирование, реализацию, эффективное использование и сопровождение.

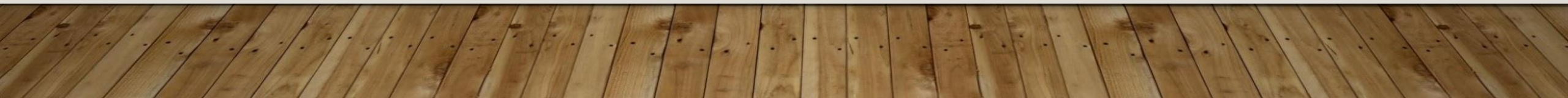
Время доступа — промежуток времени между выдачей команды записи (считывания) и фактическим получением (записью) данных.

Время отклика — промежуток времени от момента запроса к БД до фактического получения данных.

Кортеж — совокупность полей или запись (строка).

Метаданные — данные о данных, описание информационных ресурсов, их характеристик, местонахождения, способов использования и т.д.

Например, перечень таблиц с характеристиками каждой из них.



Исходным элементом базы данных является таблица, структурные составляющие которой – поле и запись.

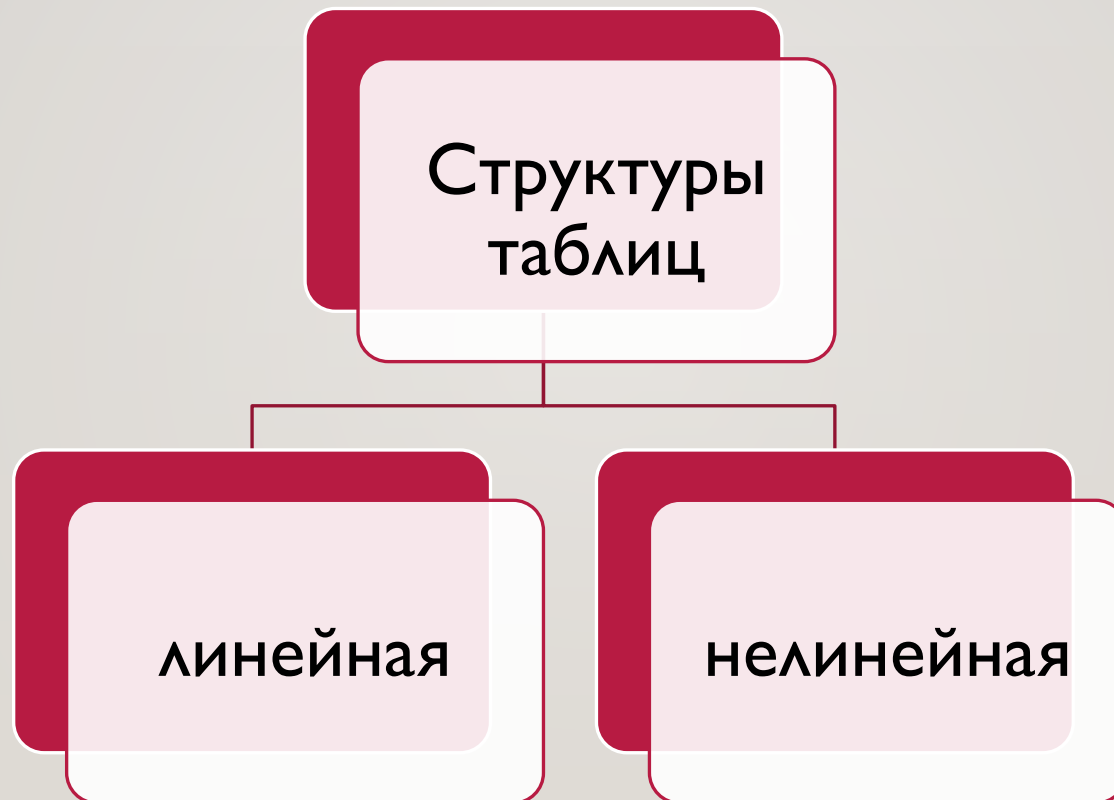


Таблица 1.11

Таблица данных о кафедре

N_преподавателя	Фамилия	Кафедра	Дисциплина	Число_часов
115	Козлов А. И.	Информатика	Базы данных	64
...
59	Серов О. В.	Информатика	Базы данных	64

Таблица 1.12

Таблица данных о студентах

Шифр студента			Оценки по семестрам			
№	Фамилия	Дисциплина	1	2	3	4
57	Скоков П. Н.	Информатика	4	5	5	4
...
12	Петров А. М.	Математика	3	5	5	4

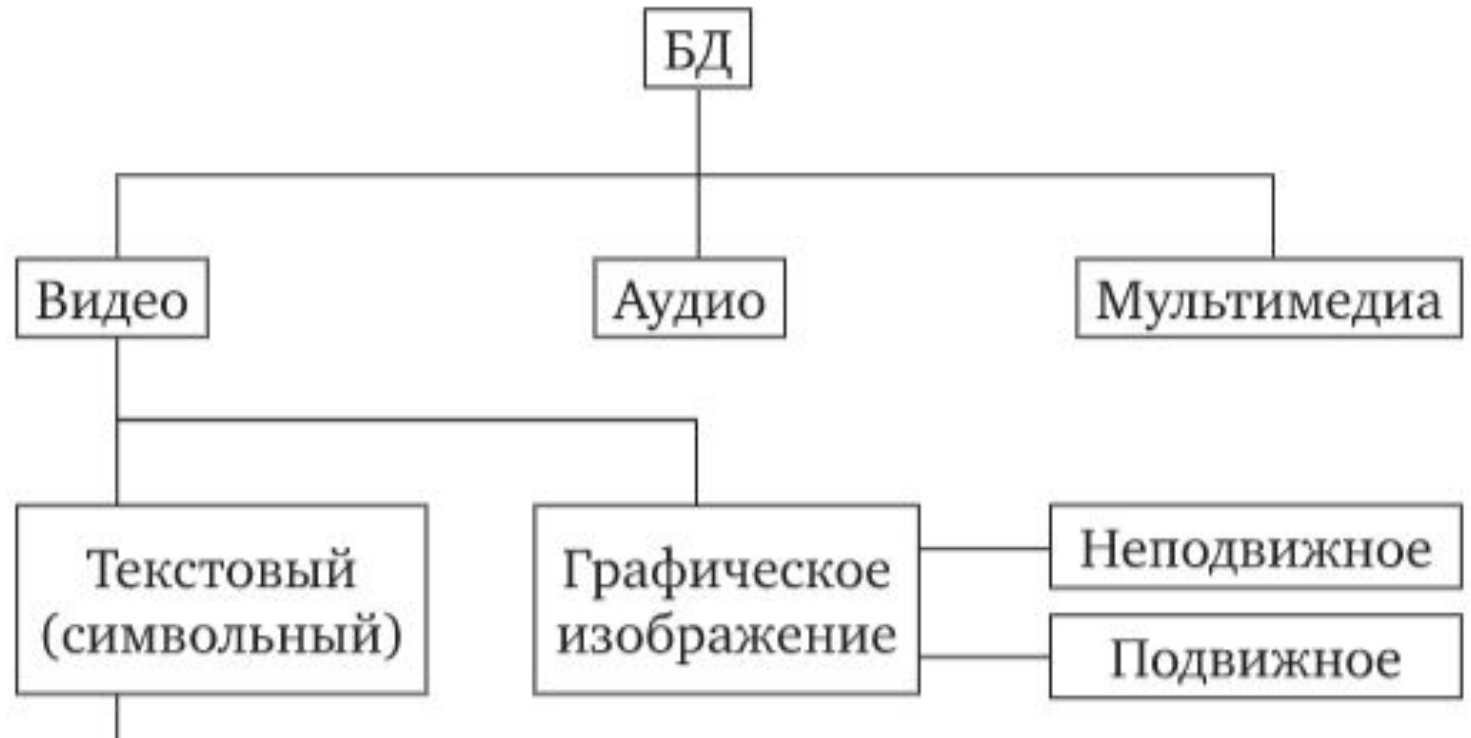
Реляционная и иерархическая модели данных реализуют только линейную структуру, сетевая и объектно-ориентированная модели позволяют использовать и нелинейную структуру.

Классификация БД

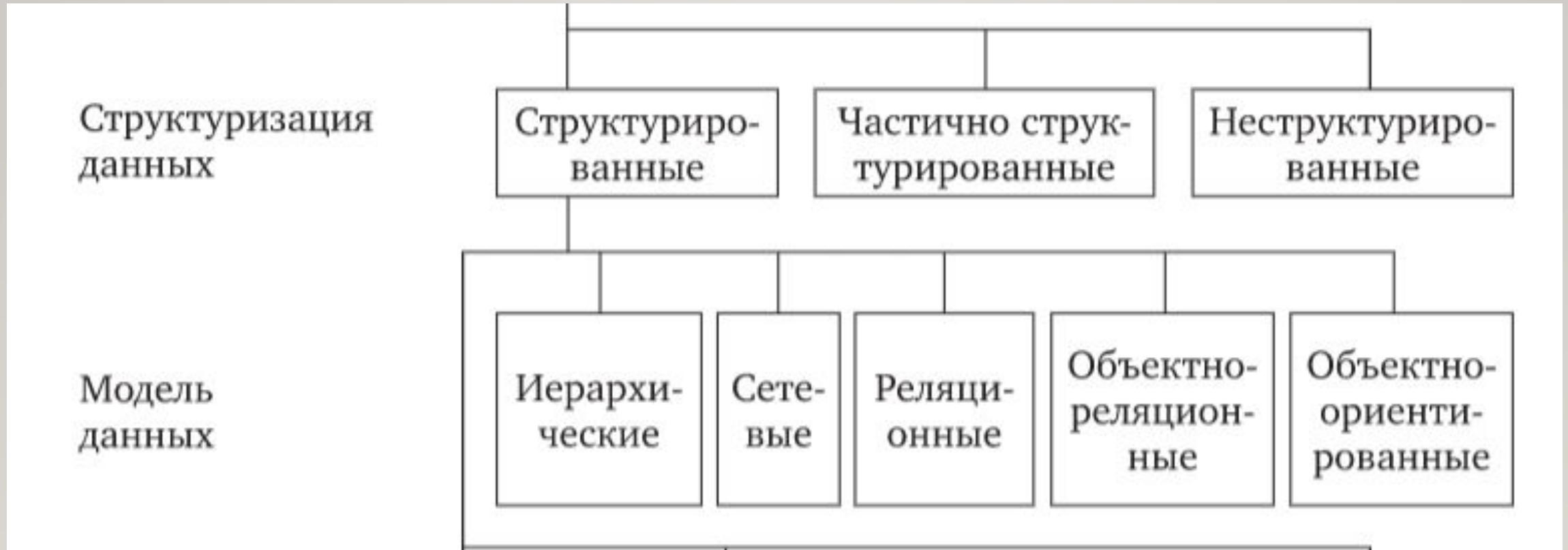
Классификационный признак

Форма представления

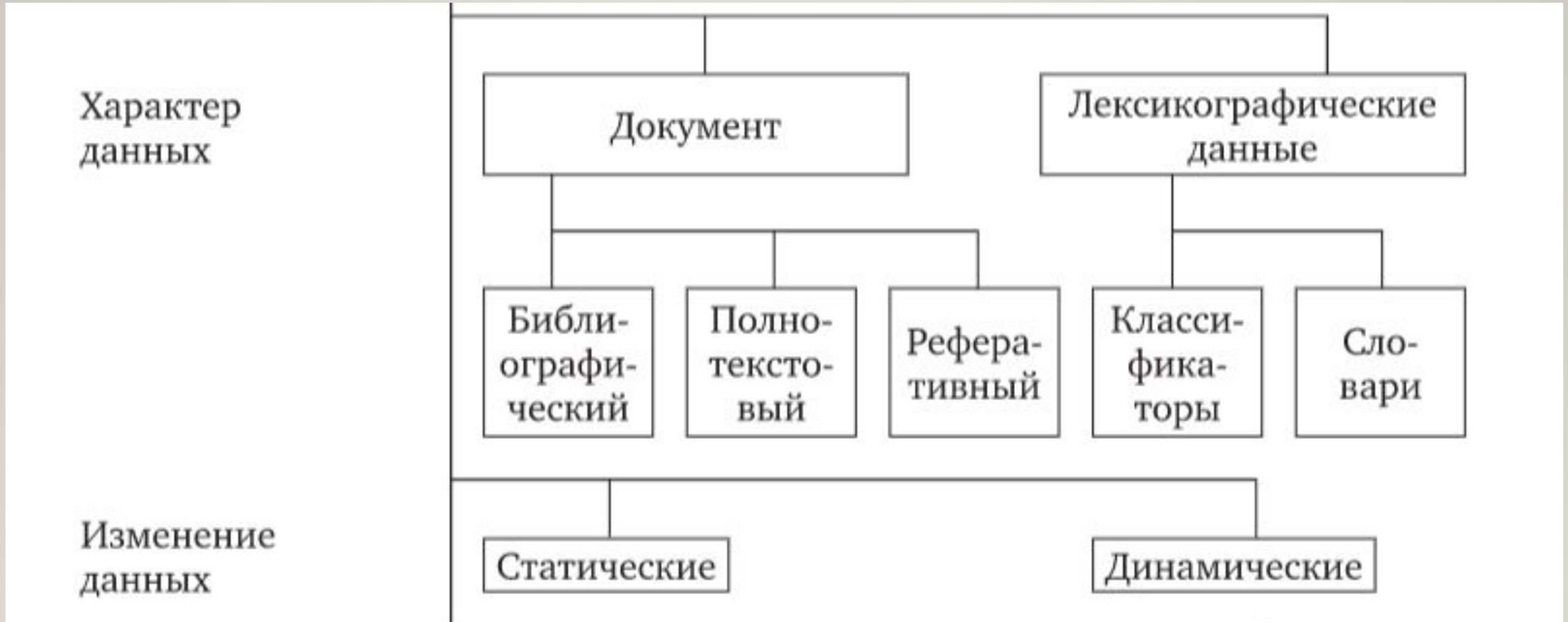
Вид данных



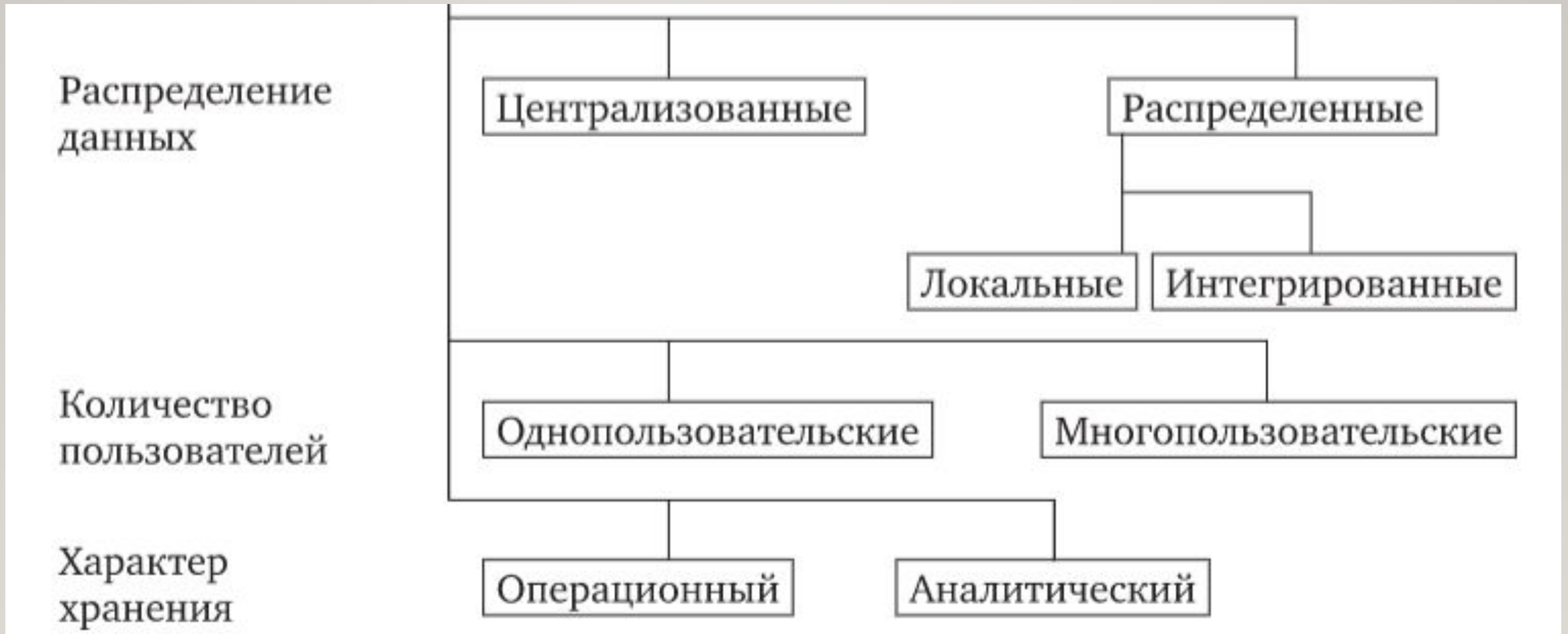
Классификация БД



Классификация БД



Классификация БД



Классификация СУБД



Рис. 1.13. Классификация СУБД

Транзакция — процесс изменения файла или БД, вызванный передачей одного входного сообщения. Это сообщение (команду) часто тоже называют транзакцией.

В **статических** БД частота обновления данных много ниже частоты их считывания. Данные напрямую не связаны со временем. Например, анкетные данные, которые используются гораздо чаще, чем изменяются. Именно для статических БД строилась изначально теория баз данных.

В статических БД данные чаще изменяются, чем добавляются.

В последнее время все чаще обращаются к **динамическим** БД, в которых частоты считывания и обновления данных соизмеримы. В динамических БД время выступает явно в виде понятий момента времени (дата) или интервала времени (семестр, месяц, год). Например, данные об успеваемости студентов групп за время обучения (по семестрам).

До середины 1990-х гг. под базами данных понимали статические БД, которые впоследствии получили название операционных (транзакционных) БД, а за рубежом — OnLine Transaction Processing (OLTP).

К середине 1990-х гг. в базах данных класса OLTP скопилось столько хронологической информации, что объем БД резко возрос, а быстродействие начало падать. Например, в работе деканата чаще всего требуются детальные данные о текущем учебном годе. В то же время в БД хранятся ретроспективные данные и за предыдущие годы. Такие данные необходимы значительно реже, и чаще всего — в агрегированном виде (например, выдать фамилии студентов, которые три последних семестра получали только отличные оценки).

Стало ясно, что ретроспективную информацию следует периодически передавать в отдельную БД. К тому же выяснилось, что ретроспективные данные обладают новым качеством: они позволяют вырабатывать стратегические решения. Возникла возможность формирования систем поддержки принятия стратегических решений (СППР). Такие системы получили за рубежом название OnLine Analytic Processing (OLAP).

Связь OLTP и OLAP показана на рис. 1.14. OLAP служит как бы дополнением OLTP.

OLAP сразу была оценена по достоинству менеджерами, поскольку позволяла: существенно повысить эффективность труда руководителей различного ранга; повысить конкурентоспособность фирм; получать дополнительно высокую прибыль.

Появление конкурентных систем OLAP привело к...

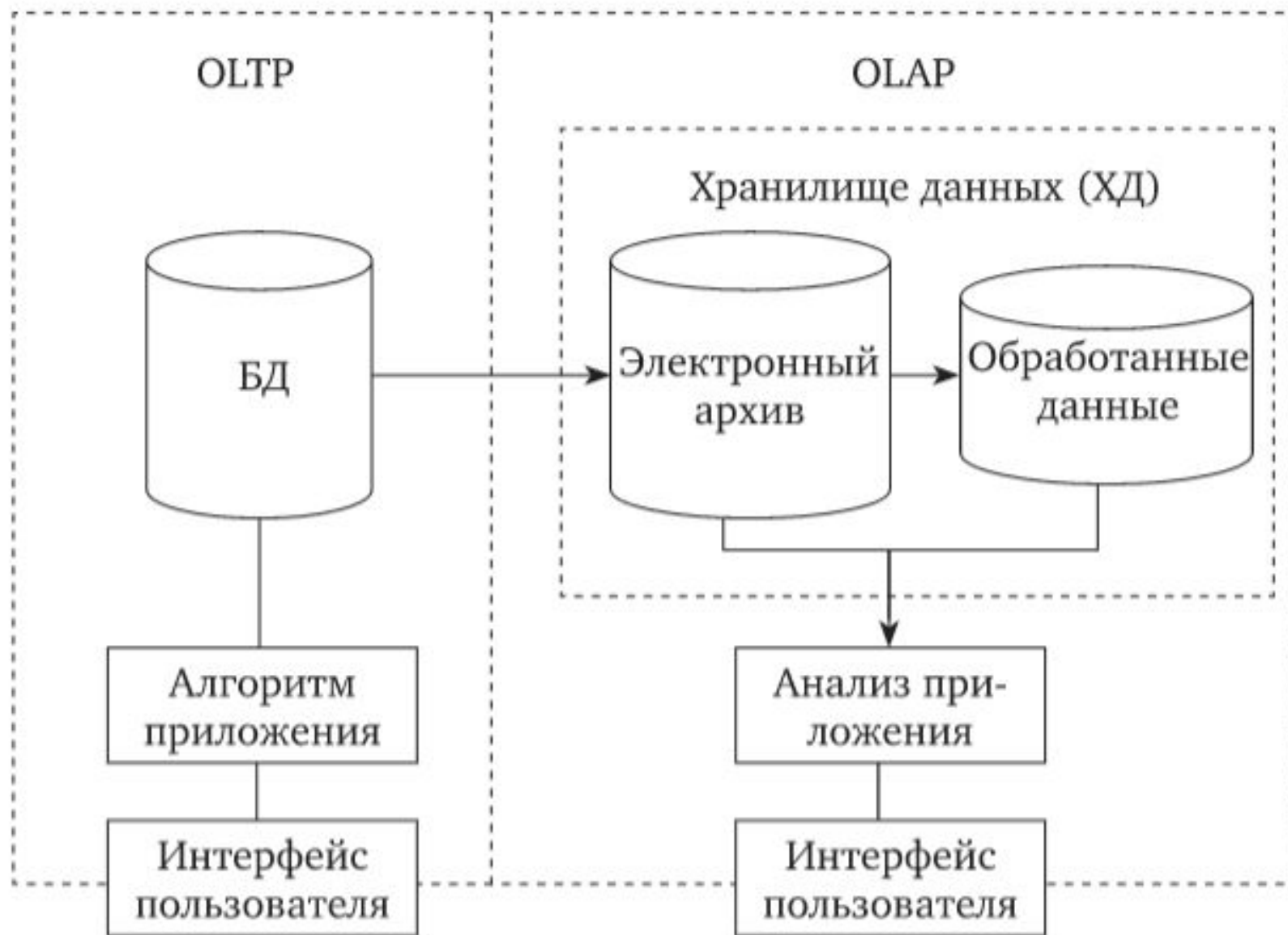


Рис. 1.14. Соотношение OLTP и OLAP

Таблица 1.13

Свойства данных в OLTP и OLAP

Свойство	OLTP	OLAP
Назначение данных	Оперативный поиск, несложная обработка	Аналитическая обработка: прогнозирование, моделирование, анализ и выявление связей, выявление статистических закономерностей
Уровень агрегации данных	Детальные данные	Агрегированные данные
Период хранения данных	До года	До нескольких десятков лет
Изменчивость данных	Изменяются	Добавляются
Упорядочение данных	По любому полю	По хронологии
Объем обрабатываемой информации	Небольшой	Очень большой

Окончание табл. 1.13

Свойство	OLTP	OLAP
Скорость обработки	Средняя	Очень высокая
Критерий эффективности работы	Количество транзакций в единицу времени	Скорость выполнения сложных запросов
Загрузка	Часто и небольшими порциями	Редко и очень большими порциями

Состав СУБД и работа БД

СУБД представляет собой оболочку, с помощью которой после построения структуры таблиц, задания связей между таблицами и заполнения таблиц данными получается соответствующая база данных. В связи с этим полезно поговорить о системе программно-технических, организационных и «человеческих» составляющих (рис. 1.15).



Рис. 1.15. Состав СУБД

