

Информационные системы обработки данных

- **Системы управления базами данных**
- **Хранилища данных**

Системы управления базами данных

Система управления базами данных (СУБД) — это комплекс программных средств, предназначенный для создания, ведения и совместного использования БД многими пользователями.

База данных (БД) представляет собой совокупность специальным образом организованных данных, хранимых в памяти вычислительной системы и отображающих состояние объектов и их взаимосвязей в рассматриваемой предметной области.

**В основе любой БД лежит модель данных ,
включающая в себя:**

1. множество **формальных объектов**, с помощью которых описывается данная ПрО (предметная область);
2. **набор операций** для манипулирования объектами модели;
3. **методологии**, позволяющие получить формальное описание из реальной ситуации и наоборот (описание ситуации на естественном языке из некоторого формального представления).
4. **способы задания ограничений целостности** (набор условий, предназначенный для выявления ошибки при описании модели данных).

По типу используемой модели можно выделить базы данных :

- Иерархическая
- Сетевая
- Реляционная

В последние годы активно внедряются :

- Объектно-ориентированные
- Объектно-реляционные

Модель представления данных - логическая структура хранимых в базе данных.

Иерархическая модель данных



Если в модели каждый порожденный элемент имеет не более одного исходного, то такая модель называется **иерархической**.

Иерархическая модель данных

Тип данных «**дерево**» (составной из *подтипов*, каждый из которых является, в свою очередь, типом «дерево»).

Корневым называется тип, который имеет подчиненные типы и сам не является *подтипом*.

Подчиненный тип (подтип) является потомком по отношению к типу, который выступает в роли предка (родителя).

Близнецы – потомки одного и того же типа.

Иерархическая БД представляет собой иерархически организованный набор типов «запись»

Отдел

Отд_номер	Отд_размер	Отд_размер
-----------	------------	------------

Начальник

Нач_номер	Нач_имя	Нач_телефон
-----------	---------	-------------

Сотрудник

Сотр_номер	Сотр_имя	Сотр_зарплата
------------	----------	---------------

Для организации данных могут использоваться следующие группы методов:

- представление **линейным списком** с последовательным распределением памяти (адресная арифметика, левосписковые структуры);
- представление **связными линейными списками** (методы, использующие указатели и справочники).

Операции манипулирования иерархически организованными данными :

- поиск указанного экземпляра БД
- переход от **одного дерева к другому**;
- переход от **одной записи к другой внутри дерева** (например, к следующей записи типа Сотрудники);
- **вставка** новой записи в указанную позицию;
- **удаление текущей записи** и т. д.

Правило контроля целостности:

*потомок не может существовать без родителя,
а у некоторых родителей может не быть потомков.*

Достоинства иерархической модели :

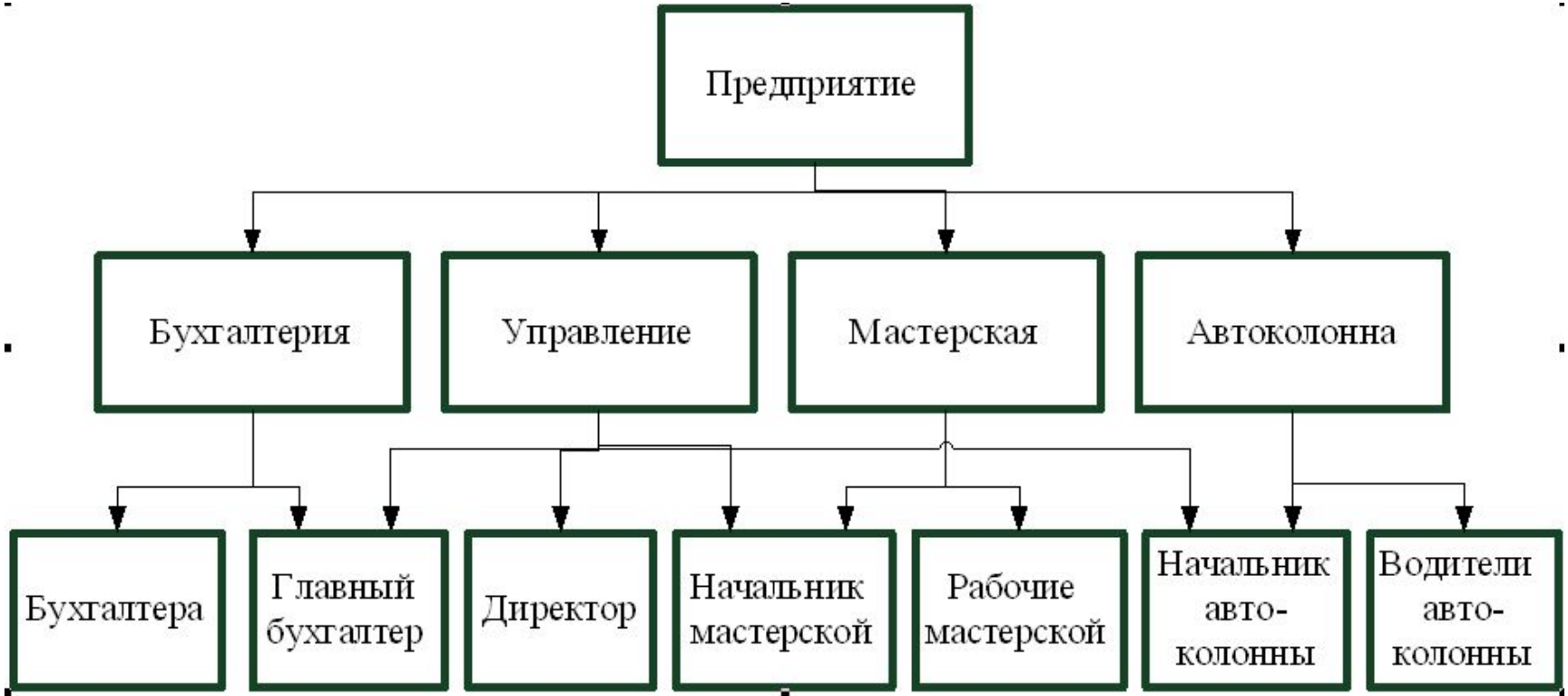
- эффективное использование памяти ЭВМ ;
- неплохие показатели времени выполнения основных операций над данными;
- модель удобна для работы с иерархически упорядоченной информацией.

Недостатком является:

- громоздкость для обработки информации с достаточно сложными логическими связями;
- сложность понимания для обычного пользователя.

Примеры: IMS, PC/Focus, Ока, ИНЭС и МИРИС.

Сетевая модель



Если в модели каждый порожденный элемент может иметь более одного исходного, то такая модель называется **сетевой**



Для описания схемы сетевой БД используется две группы типов: *«запись»* и *«связь»*.

Тип *«связь»* определяется для двух типов *«запись»*: *предка* и *потомка*.

Переменные типа *«связь»* являются *экземплярами связей*.

Операции манипулирования данными :

- поиск записи в БД;
- переход от предка к первому потомку;
- переход от потомка к предку;
- создание новой записи;
- удаление текущей записи;
- обновление текущей записи;
- включение записи в связь;
- исключение записи из связи;
- изменение связей и т. д.

Достоинством сетевой модели данных является:

- возможность эффективной реализации по показателям **затрат памяти и оперативности**
- сетевая модель предоставляет большие возможности в смысле **допустимости образования произвольных связей** (в сравнении с иерархической моделью)

Недостатком сетевой модели данных является:

- высокая **сложность и жесткость схемы БД**, построенной на ее основе;
- **сложность для понимания и выполнения обработки информации** в БД обычным пользователем;
- **ослаблен контроль целостности связей**

Сетевые СУБД: *IDMS, db_Vista III, СЕТЬ, СЕТОР и КОМПАС* X

Реляционная модель данных



Операции работы с данными: объединение, пересечение, разность, произведение, ограничение и соединение.

Таблица (*отношение*) имеет строки (*записи*) и столбцы (колонки).

Каждая строка таблицы имеет одинаковую структуру и состоит из ***полей***.

Строкам таблицы соответствуют ***кортежи***, а столбцам — ***атрибуты отношения***.

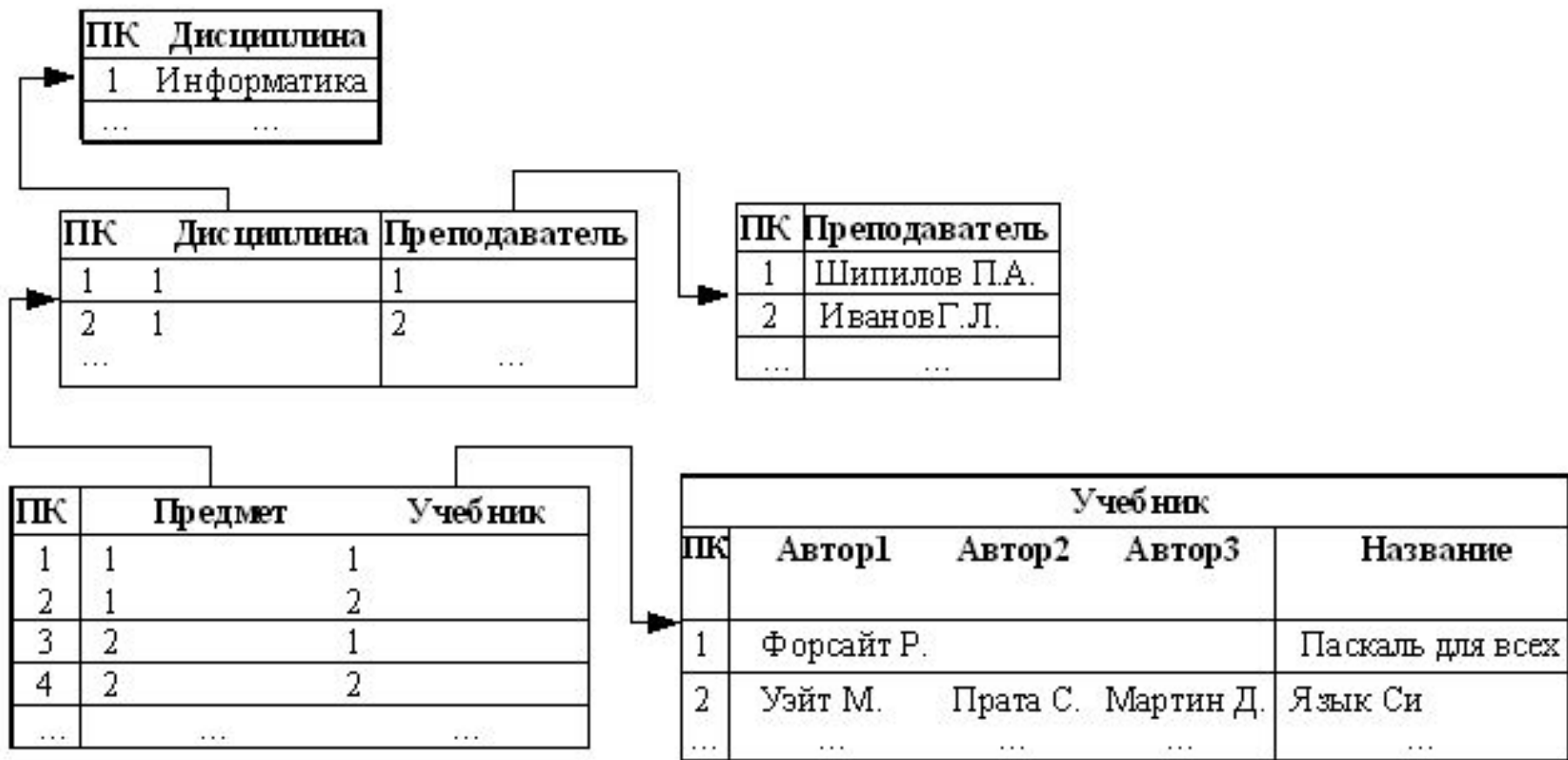
Нормализация – это разбиение таблицы на две или более, обладающих лучшими свойствами при включении, изменении и удалении данных.

Цель – получение такого проекта базы данных, в котором в котором ***исключена избыточность информации***.

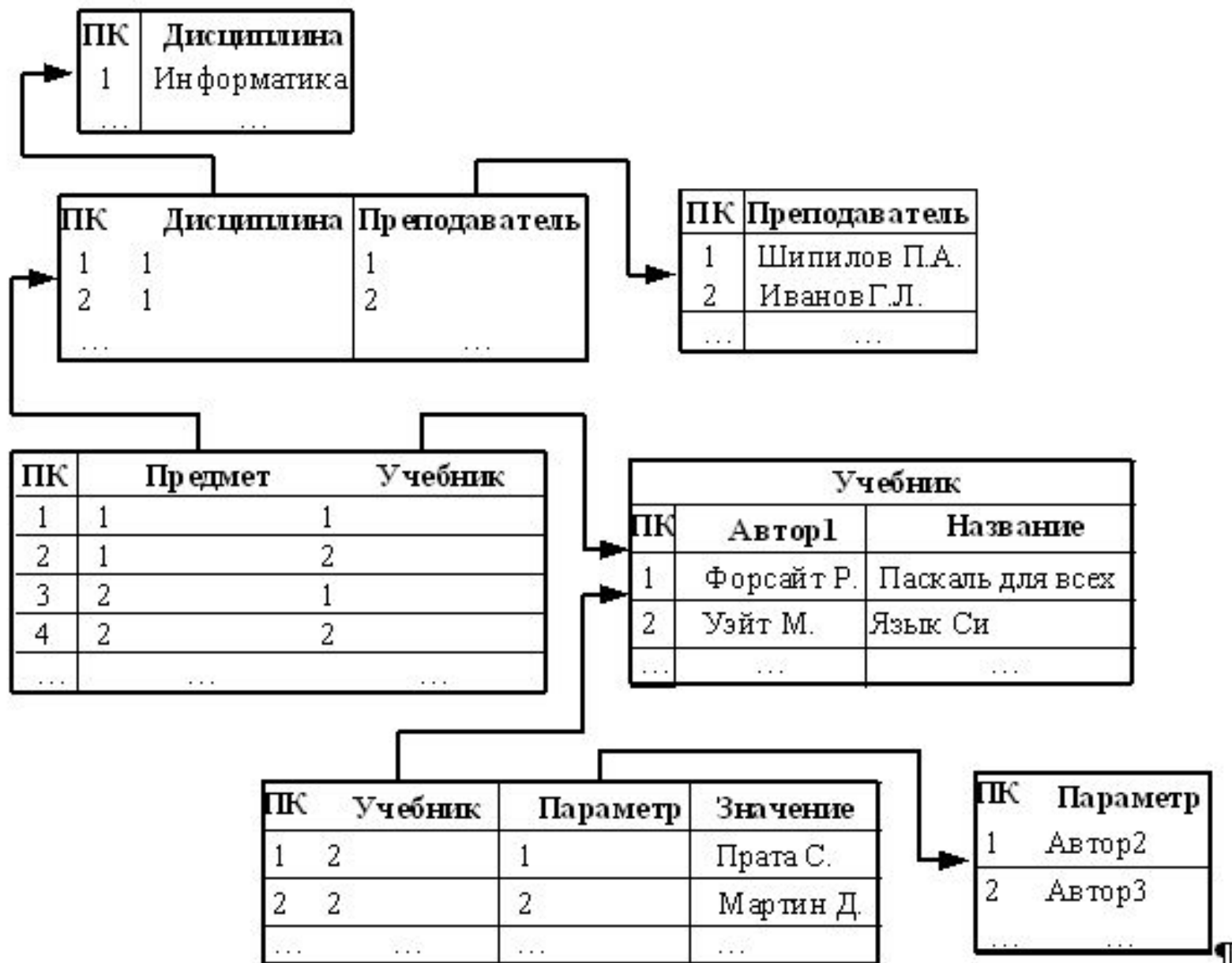
Пример реляционной базы (начальная таблица)

Дисциплина	Преподаватель	Учебник
Информатика	<u>Шипилов П.А.</u>	<u>Форсайт Р. Паскаль для всех</u>
Информатика	<u>Шипилов П.А.</u>	<u>Уэйт М., Прата С., Мартин Д. Язык Си</u>
Информатика	Иванов Г.Л.	<u>Форсайт Р. Паскаль для всех</u>
Информатика	Иванов Г.Л.	<u>Уэйт М., Прата С., Мартин Д. Язык Си</u>
...

III нормальная форма базы данных



IV нормальная форма базы данных



Достоинства реляционной модели

1. Развитая теория реляционной модели данных.
2. Возможность сведения иерархической и сетевой модели данных к реляционному способу организации данных.
3. Поддержка механизмов ускоренного доступа к данным.
4. Возможность манипулирования данными без необходимости знания физической организации БД.
5. Наличие языка запросов к базе данных SQL.
6. Система берет на себя вопросы по эффективности доступа к данным, а также контроля за целостностью информации.

Недостатки реляционной модели

1. Разработка программного обеспечения приходится **проектировать** свою задачу не в терминах ПрО (самой по себе достаточно сложной), а в **терминах реляционных таблиц**, что затрудняет процесс разработки.
2. НЕТ жесткой методики преобразования элементов ПрО в реляционные таблицы.
3. При сохранении информации объекты раскладываются на простые элементы, а затем при чтении – собирать заново (реляционная модель лишается одного из главных своих преимуществ – гибкости к изменению структуры БД).

Объектно-ориентированная модель

ODMG-93 (Object Database Management Group)

Структура графически представима в виде дерева, узлами которого являются *объекты*.

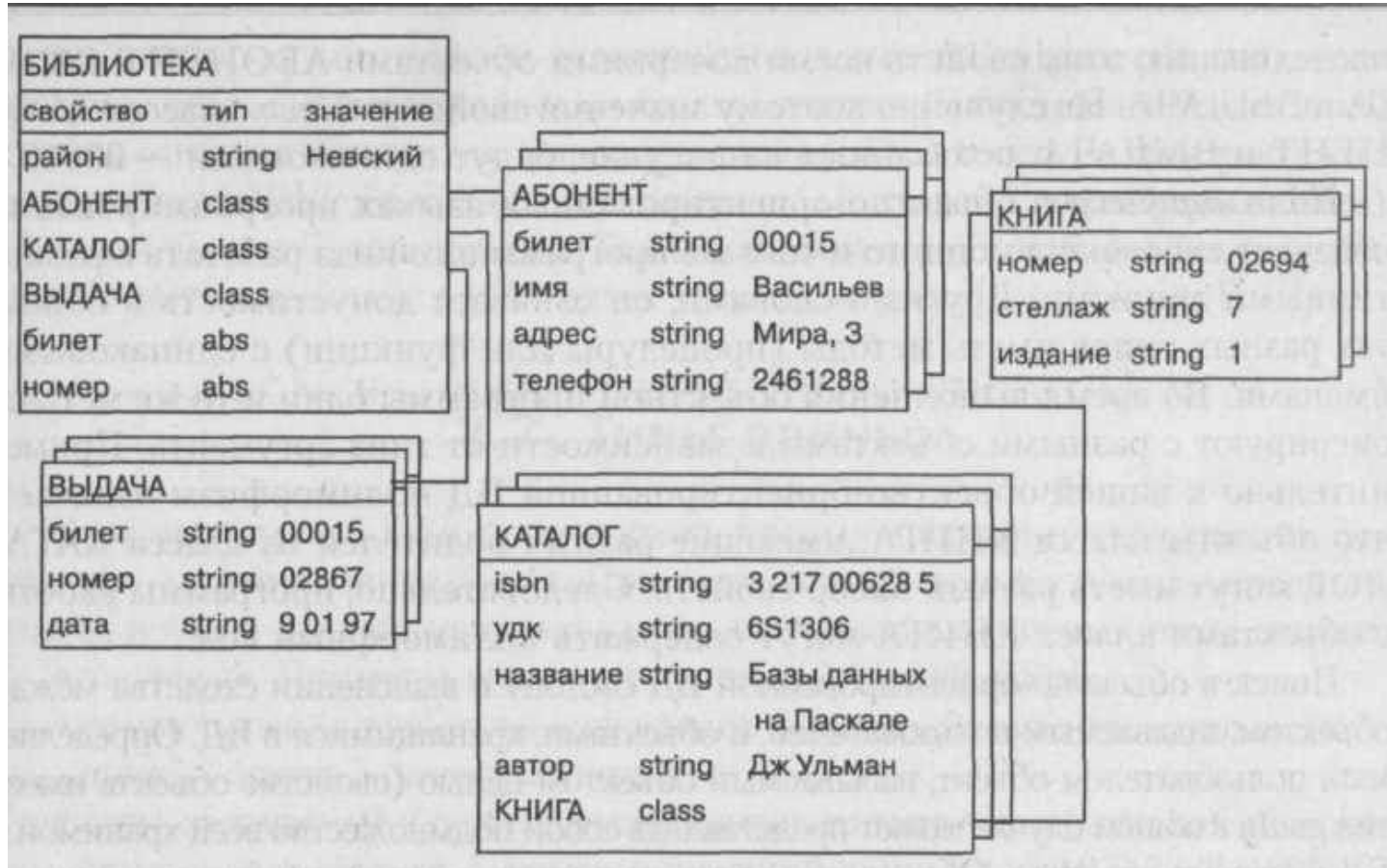
Свойства объектов описываются:

- стандартным типом (**string**—строка символов)
- типом конструируемым пользователем (**class**— *объект*, являющийся *экземпляром* соответствующего *класса*).

Каждый *объект-экземпляр класса* считается потомком объекта, в котором он определен как свойство.

Он принадлежит своему классу и имеет одного родителя.

Объектно-ориентированная БД



Методы манипулирования данными:

- применяются *логические операции*, усиленные механизмами **инкапсуляции, наследования и полиморфизма**
- создание и модификация БД сопровождается *автоматическим формированием* и последующей *корректировкой индексов* (индексных таблиц), содержащих информацию для *быстрого поиска данных*.

Достоинство объектно-ориентированной модели в сравнении с реляционной :

- возможность отображения информации о сложных взаимосвязях объектов;
- позволяет идентифицировать отдельную запись базы данных и определять функции их обработки.

Недостатками

- высокая понятийная сложность;
- неудобство обработки данных и низкая скорость выполнения запросов.

Объектно-реляционный подход (ORM – *Object Relation Mapping*)

Объектно-ориентированный подход	Реляционный подход
Класс	Отношение (таблица)
Объект	Кортеж (запись)
Атрибут	Атрибут
Иерархия классов	Система реляционных связей
Связи между объектами	

Отличие между *объектно-ориентированными* и *объектно-реляционными БД* заключается в том, что при введении объекта в качестве атрибута другого объекта в объектно-реляционных БД сохраняется **не сам объект, а ссылка на него.**

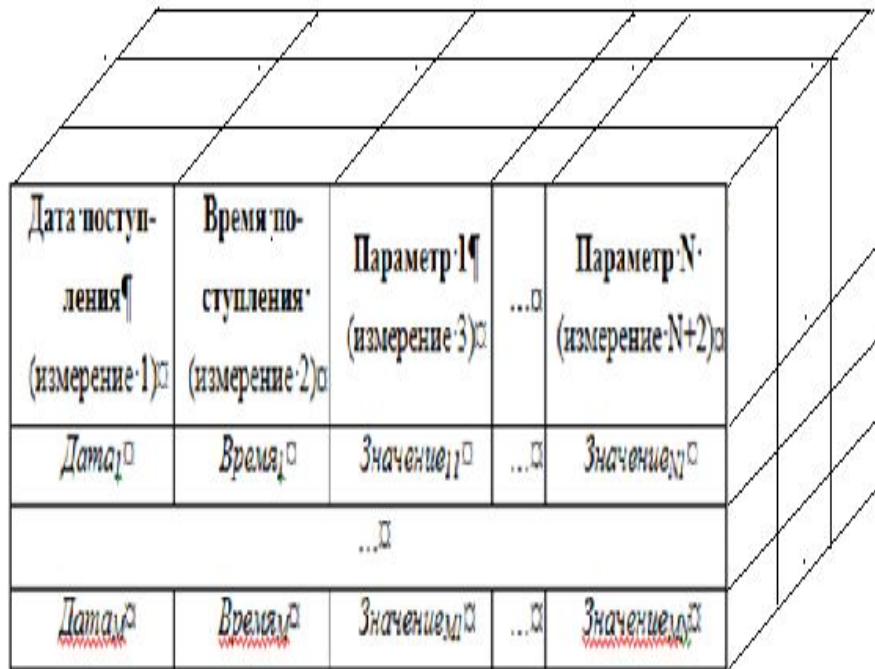
Хранилища данных

- Структура хранилища данных
- Свойства хранилищ данных
- Область применения хранилищ данных
- Data Mining – технология аналитической обработки данных
- Системы поддержки принятия решений(СППР)

Хранилище данных (ХД) — это предметно-ориентированное, интегрированное, привязанное ко времени и неизменяемое собрание данных для поддержки процесса принятия управляющих решений.

Хранилище данных (ХД) — представляет собой банк данных определенной структуры, содержащий информацию в историческом контексте.

Структура гиперкуба



Измерения			Факты	
Дата	Наименование	Покупатель	Сумма	Количество
09.01.01	Эуфиллин 0.5 N30	ТПЛ "ФАРМАЦИЯ"	44019	20100
21.01.01	СТРЕПТОМИЦИНА СУЛЬ	ОБЛ.КОЖВЕНДИСПАН	29601	14300
21.01.01	Эуфиллин 0.5 N30	ОБЛ.КОЖВЕНДИСПАН	27594	12600
28.01.01	ГУМИЗОЛЬ 1МЛ N10	МКП ФИРМА "ЭПРОН"	2296	350
20.05.01	Эуфиллин 0.5 N30	"ЯМАНУЧИ ФАРМА"	1372	700

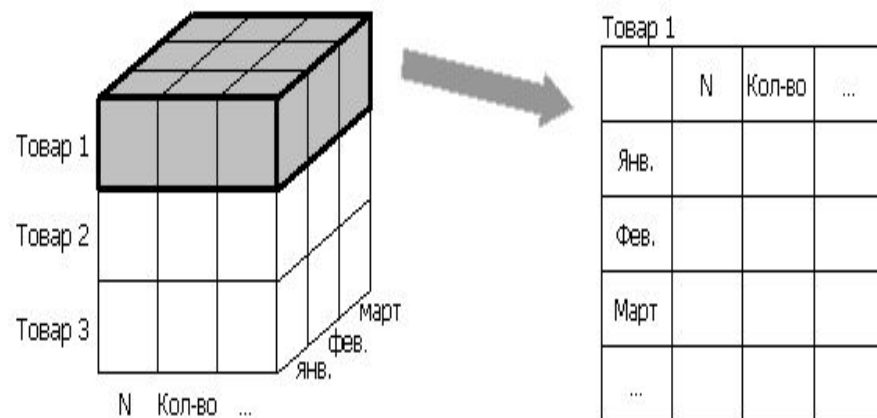


Измерение — это множество, образующее одну из граней гиперкуба.

Значение — данные, которые подвергаются анализу в ячейках куба

Основные операции над гиперкубом:

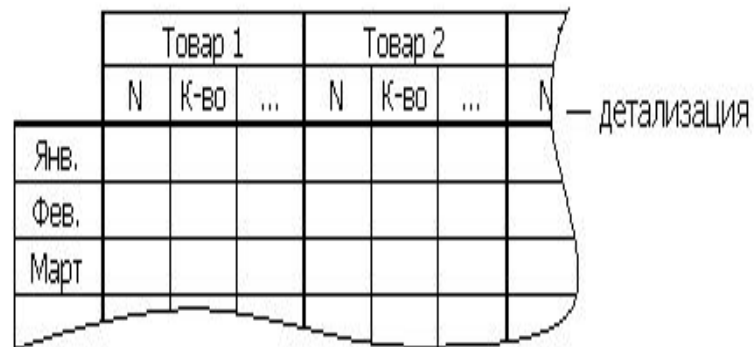
Сечение — формируется подмножество гиперкуба, в котором значение одного или более измерений фиксировано.



Вращение — изменение порядка представления измерения (применяется к двумерным таблицам).

Свертка — группирование(объединение) данных в каком-либо измерении.

Детализация — переход от обобщенных к детализированным данным (операция, обратная свертке).



ХД в зависимости от размера делятся :

- Малые (до 10^6 ячеек данных)
- Средние (до 10^8)
- Крупные ($\sim 10^8$)
- Сверхбольшие ($\sim 10^9$)

Подходы к построению ХД:

- Многомерная модель хранилища (**MOlap**).
- Реляционная (**ROlap**).
- Гибридная (**HOlap**).

MOLAP

Используют при небольшой базе данных и стабильном наборе измерений.

Преимущество: быстрое чтение и поиск данных

Недостатки : нерациональное использование памяти (все измерения и все аргументы функции хранятся в многомерном виде, много пустых ячеек).

ROLAP

Преимущество — база данных может быть большой.

Недостаток — медленное выполнение аналитических запросов.

HOLAP

Основные данные хранятся в реляционной базе, а агрегированные — в многомерной структуре (кубе).

Основные свойства хранилищ данных:

- ***Ориентация на ПрО:***

данные в хранилище организованы вокруг существенных аспектов прикладной деятельности;

- ***Интегрированность:***

информация в хранилище должна интегрироваться в целостную структуру, обеспечивающую возможности анализа данных;

- ***Агрегированность.***

Чтобы при выполнении аналитических запросов избежать выполнения операций группирования, данные должны обобщаться (агрегироваться) при загрузке хранилища;

- ***Поддержка хронологии:***

хранилище можно рассматривать как набор моментальных снимков состояния данных так, что атрибут времени всегда явно присутствует в структурах данных хранилища;

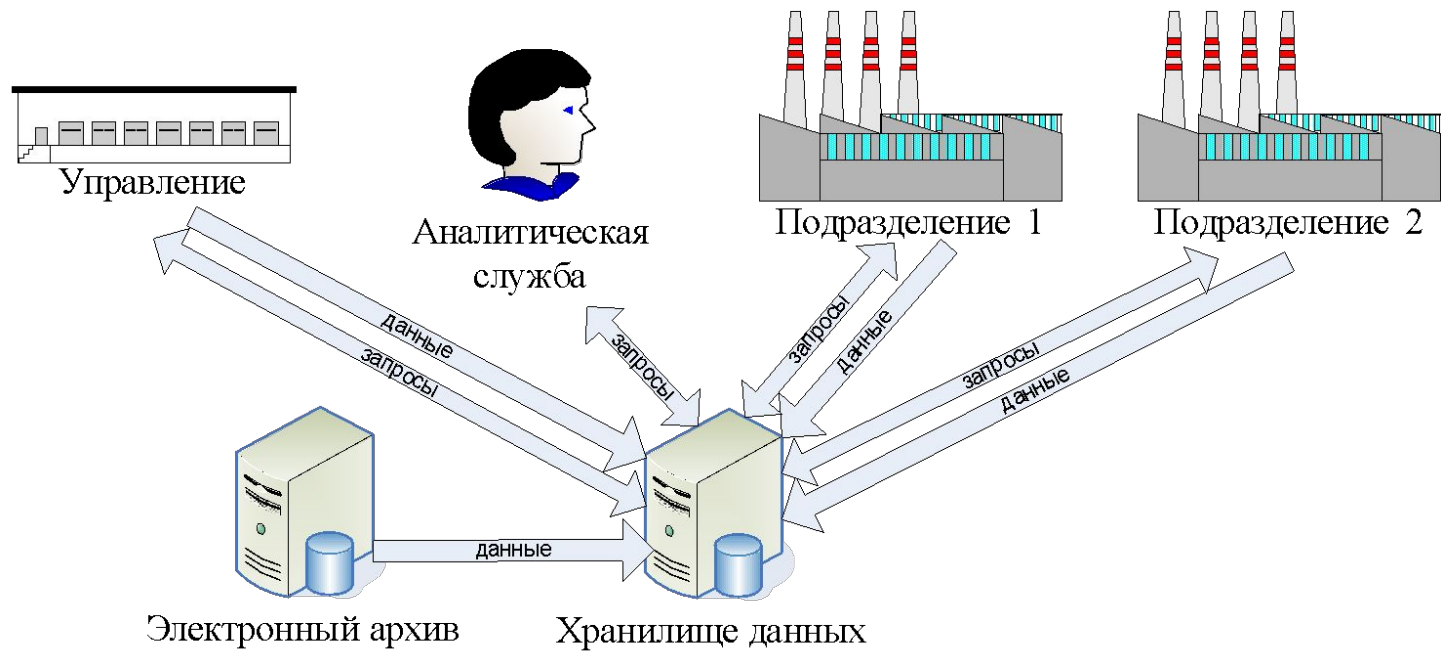
- ***Неизменяемость:***

данные после загрузки в ХД остаются неизменными, внесения каких-либо изменений, кроме добавления записей, не разрешается.

Область применения хранилищ данных

- для своевременного обеспечения аналитиков всей информацией, необходимой для выработки решений;
- для создания единой модели данных организации;
- для создания интегрированного источника данных, предоставляющего удобный доступ к разнородной информации (единый «источник истины»).

Схема взаимодействия хранилища данных с клиентскими приложениями



Data Mining – это *технология выявления скрытых (ранее неизвестных) взаимосвязей внутри больших объемов данных.*

Data Mining выполняет следующие виды анализа:

- классификация;
- регрессионный анализ;
- прогнозирование временных последовательностей (рядов);
- кластеризация;
- выявление ассоциаций и последовательностей.

Системы поддержки принятия решений(СППР)

СППР – являются человеко-машинными объектами, которые позволяют лицам, принимающим решения (ЛПР), использовать данные, знания, объективные и субъективные модели для анализа и решения слабоструктурированных и неструктурированных проблем.

СППР – это компьютерная система, позволяющая *ЛПР* сочетать собственные субъективные предпочтения с компьютерным анализом ситуации при выработке рекомендаций в процессе принятия решения.

Функции СППР :

- помощь ЛПР при анализе обстановки (ситуации) и ограничений, накладываемых внешней средой;
- выявление предпочтений ЛПР или выявление и ранжирование приоритетов, учет неопределенности в оценках ЛПР и формирование его предпочтений;
- генерацию возможных решений ,т.е. формирование списка альтернатив;
- оценку возможных альтернатив, исходя из предпочтений ЛПР и ограничений, накладываемых внешней средой;
- моделирование принимаемых решений (когда это возможно);
- компьютерный анализ последствий принимаемых решений;
- сбор данных о результатах реализации принятых решений и выбор лучшего результата.

Выработка решений в этих системах происходит в результате итерационного процесса, в котором участвуют:

- система ППР в роли вычислительного звена;
- человек как управляющее звено, задающее исходную информацию и оценивающее полученный результат.



Основные компоненты информационной технологии поддержки принятия решений



Отличие информационной технологии ППР:

- ориентация на решение слабоформализованных (плохо структурированных) задач;
- сочетание традиционных методов доступа и обработки компьютерных данных с возможностью использования математических моделей решения задач;
- ориентация на непрофессионального пользователя компьютера;
- высокая адаптивность, обеспечивающая возможность приспособления к требованиям пользователя.