



Базы данных

***Гаврилов Александр Викторович
к.т.н., доцент***



Тема 1. Введение в курс «Базы данных»

Лекция 1

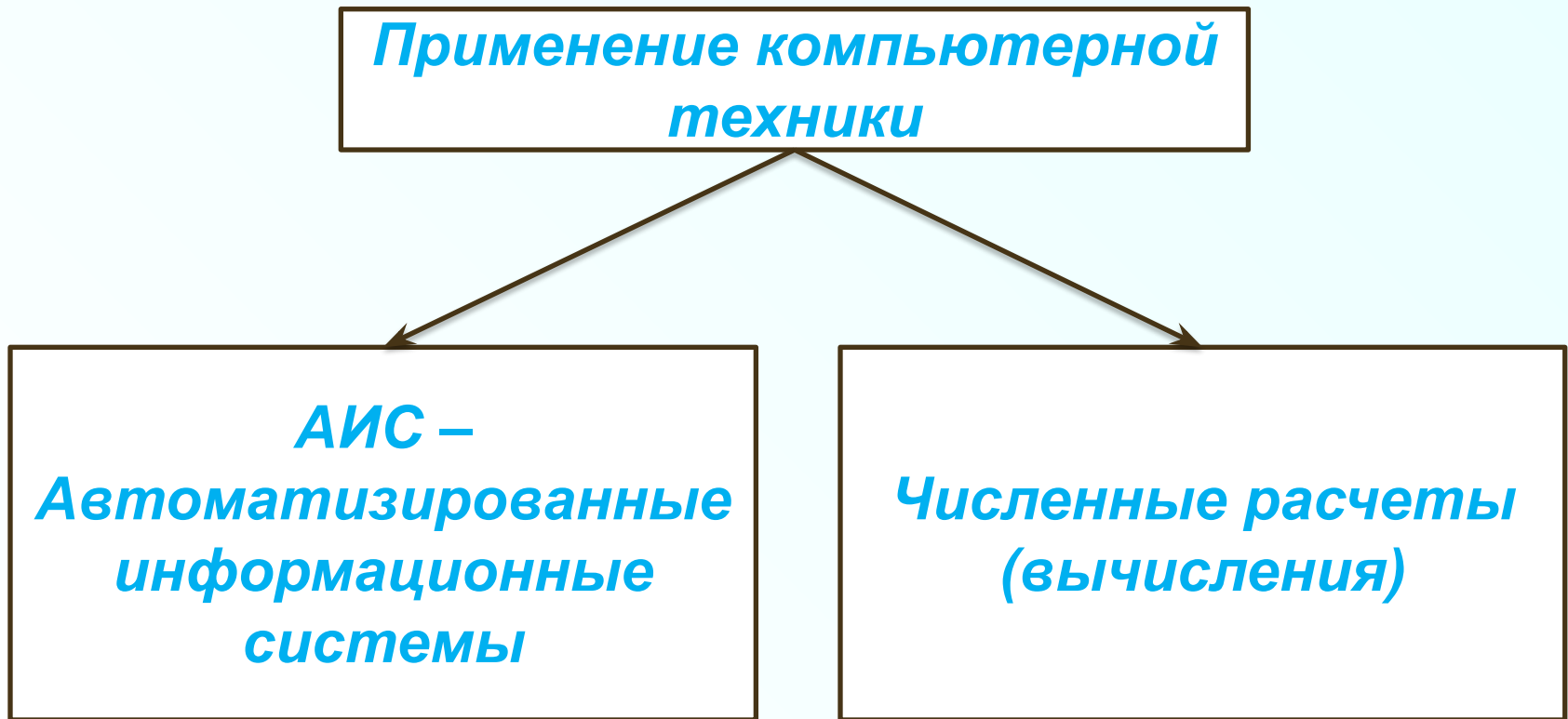
Вопросы лекции:

1. История возникновения баз данных
2. Основные термины и определения
3. Классификация СУБД
4. Банки данных

История возникновения баз данных



Использование средств вычислительной техники в автоматических или автоматизированных информационных системах является одним из двух основных направлений ее применения



Этапы развития БД.

Этап 0. Файловые системы

Магнитные диски впервые были реализованы в 1956 году в исследовательской лаборатории корпорации IBM, расположенной в Сан-Хосе (Калифорния), где был выпущен серийный дисковый накопитель IBM 350 — первое устройство с подвижной головкой для чтения и записи.

Важным шагом в развитии баз данных явился переход к использованию централизованных систем управления файлами, или, используя общепринятый в данный момент термин, **файловым**

системам

Этапы развития БД.

Этап 0. Файловые системы


Магнитные диски впервые были реализованы в 1956 году в исследовательской лаборатории корпорации IBM, расположенной в Сан-Хосе (Калифорния), где был выпущен серийный дисковый накопитель IBM 350 — первое устройство с подвижной головкой для чтения и записи.



Важным шагом в развитии баз данных явился переход к использованию централизованных систем управления файлами, или, используя общепринятый в данный момент термин, **файловым системам**

Этапы развития БД.

Этап 0. Файловые системы

 **Файловая система** — порядок, определяющий формат содержимого и способ физического хранения информации, которую принято группировать в виде файлов. Конкретная файловая система определяет размер имен файлов и (каталогов), максимальный возможный размер файла и раздела, набор атрибутов файла.

 **Первая** развитая **файловая система** была разработана фирмой **IBM** для ее серии **System/360** в **1964** году

Недостатки применение файловых систем для хранения и обработки данных в информационных системах:

- ✓ **Избыточность данных.** Из-за дублирования данных в разных файлах память на внешних запоминающих устройствах используется неэкономно
- ✓ **Несогласованность данных.** Учитывая, что одна и та же информация может размещаться в разных файлах, технологически тяжело проследить за внесением изменений одновременно во все файлы.
- ✓ **Зависимость структур данных и прикладных программ.** Этот недостаток файловых систем приводит к значительному увеличению стоимости сопровождения программных средств. Иногда стоимость сопровождения программных средств может достигать близко 70 % стоимости их разработки.

Этапы развития БД.

Этап 1. Базы данных на больших ЭВМ 1960–1980 гг.



Получаемые в результате библиотеки, реализующие дополнительные индивидуальные средства управления данными, являлись существенной частью информационных систем и практически повторялись от одной системы к другой.



Стремление выделить и обобщить общую часть информационных систем, ответственную за управление сложно структурированными данными, вылилось в реализацию новых программных систем, названных впоследствии **Системами Управления Базами Данных (СУБД)**, а сами хранилища информации, которые работали под управлением данных систем, назывались **базами данных**.

Этапы развития БД.

Этап 1. Базы данных на больших ЭВМ 1960–1980 гг.

На сегодняшний день история развития СУБД насчитывает уже более 45 лет. В 1968 году компания IBM разработала первую промышленную СУБД:

IBM IMS (Information Management System)

Главным архитектором СУБД был **Верн Уоттс**. Начав работу в IBM в 1956 году, он непрерывно работал над IMS начиная от времени её первоначального проектирования вплоть до своей кончины 4 апреля 2009 года.

В задачу IMS входила обработка спецификации изделия для ракеты Сатурн-5 и кораблей Аполлон.

В качестве носителя информации использовалась **магнитная лента**, а в качестве структуры данных – **иерархическая модель**.

IBM System/360



Этапы развития БД.

Этап 2. Настольные (desktop) СУБД 1975 - 1995 гг.



Звание первого персонального компьютера принадлежит модели 5100 производства фирмы IBM, выпущенной в 1975 году. Он был более компактным, чем мэйнфреймы, имел встроенные монитор, клавиатуру и накопитель на магнитной ленте, и предназначался для решения научно-инженерных задач.



Первым же массовым **персональным компьютером** производства фирмы IBM, выпущенным в 1981 году, стал **IBM PC** модели **5150**, положивший начало семейству наиболее распространённых современных персональных компьютеров.

Этапы развития БД.

Этап 2. Настольные (desktop) СУБД 1975 - 1995 гг.



Спрос на развитые удобные программы обработки данных заставлял поставщиков программного обеспечения поставлять все новые системы, которые принято называть **настольными (desktop) СУБД.**



Наличие на рынке большого числа СУБД, выполняющих сходные функции, потребовало разработки методов экспорта, импорта и открытых форматов хранения данных. Так появились первые коммерческие **СУБД с реляционной моделью данных.**


Этапы развития БД.

Этап 2. Настольные (desktop) СУБД 1975 - 1995 гг.

- ✓ Основанные на реляционном подходе СУБД для персональных компьютеров принято считать **системами второго поколения.**
- ✓ В 80-х годах были созданы различные коммерческие реляционные СУБД - например, DB2 или SQL/DS корпорации IBM, Oracle и др. Большинство СУБД имели развитый и удобный пользовательский интерфейс, предлагающий интерактивный режим работы с БД, как в рамках описания БД, так и в рамках проектирования запросов.

Этапы развития БД.

Этап 2. Настольные (desktop) СУБД 90-е - 1995 гг.

 Главное **ограничение** при работе с настольными СУБД накладывалось **монопольным доступом**, поскольку первое время персональные компьютеры не были подключены к вычислительным сетям. Базы данных на них создавались для работы одного пользователя.

Этапы развития БД.

Этап 3. Распределенные базы данных с 1985 по наст. вр.



Третий этап развития СУБД связывают с распространением локальных и глобальных компьютерных сетей.



На сегодняшний день третий этап можно считать незавершённым.



К этому этапу можно отнести разработку ряда стандартов в рамках языков описания и манипулирования данными:

SQL-89, SQL-92, SQL-99, SQL:2003, SQL:2006, SQL:2008



Представителями современных СУБД можно считать серверы баз данных **Oracle, MS SQL, Informix, DB2, MySQL** и другие.

РЕЗЮМЕ

История развития баз данных насчитывает более 50 лет.

Условно выделяют три этапа. При этом между ними нет жестких временных ограничений, этапы плавно переходили из одного в другой и существовали параллельно:

- Этап 1. Базы данных на больших ЭВМ 1960–1980 гг.
- Этап 2. Настольные (desktop) СУБД 1975–1995 гг.
- Этап 3. Распределенные базы данных 1985–... гг.

Основные термины и определения

Основные термины и определения

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.

Основные термины и определения

В части четвертой **«Гражданского кодекса Российской Федерации»** дается

следующее определение базы данных:

«Базой данных является представленная в объективной форме совокупность самостоятельных материалов (статей, расчетов, нормативных актов, судебных решений и иных подобных материалов), систематизированных таким образом, чтобы эти материалы могли быть найдены и обработаны с помощью электронной вычислительной машины (ЭВМ)».

Основные термины и определения

Данные (data – факт) – это совокупность сведений, зафиксированных на определенном носителе в форме, пригодной для постоянного хранения, передачи и обработки.

Преобразование и обработка данных позволяет получить **информацию**.

Основные термины и определения

Вычислительная система (ВС)

представляет собой совокупность взаимосвязанных и согласованно действующих ЭВМ или процессоров и других устройств, обеспечивающих автоматизацию процессов приема, обработки и выдачи информации потребителям.

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

Основные функции СУБД:

- ✓ **Создание БД**
- ✓ **Создание пользователей и указание привилегий**
- ✓ **Обеспечение работы пользователей с БД с учетом привилегий**
- ✓ **Поддержание целостности данных**
- ✓ **Поддержание механизма транзакций**
- ✓ **Журналирование**
- ✓ **Управление оперативной памятью**

Основные термины и определения

Структурирование – это введение соглашений о способах представления данных.

Создавая базу данных, пользователь стремится упорядочить информацию по различным признакам и быстро извлекать выборку с произвольным сочетанием признаков. Сделать это возможно, только если данные структурированы.

Неструктурированными называют данные, записанные, например, в текстовом файле.

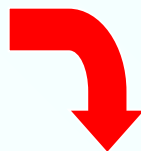
Личное дело № 16493, Сергеев Петр Михайлович, дата рождения 1 января 1976 г.;

Л/д № 16593, Петрова Анна Владимировна, дата рожд. 15 марта 1975 г.;

№ личн. дела 16693, д.р. 14.04.76, Анохин Андрей Борисович.

Структурированные
данные

Неструктурированные
данные



№ личного дела	Фамилия	Имя	Отчество	Дата рождения
16493	Сергеев	Петр	Михайлович	01.01.76
16593	Петрова	Анна	Владимировна	15.03.75
16693	Анохин	Андрей	Борисович	14.04.76

Основные термины и определения

Целостность БД – непротиворечивость информации, хранящейся в БД.

Транзакция – законченная совокупность действий над БД, которая переводит БД из одного целостного состояния в другое целостное состояние. Совокупность простых операций над БД, объединенных в единое целое, и выполняемых по принципу «все или ни одной». Т.е. в случае возникновения ошибок при выполнении какой-либо операции, входящей в транзакцию, БД возвращается в состояние до выполнения транзакции.

Привилегия пользователя – права пользователя на выполнение операций с данными (запись, корректировка, чтение, удаление), а также выполнение других действий над БД.

Классификация СУБД и БД

Классификация СУБД и БД

По сфере возможного применения:

✓ **универсальные**

Пример: MS Access, PostgreSQL.

✓ **специализированные (проблемно-ориентированные) СУБД и БД**

Примеры ИС, в которых необходимо использовать специализированные СУБД:

- отдельные АСУТП, где нужна СУБД реального времени, обладающая полной функциональностью универсальной СУБД;
- биометрические системы;
- системы военного назначения;
- государственные информационные системы и т.д.

Пример: В АС военного назначения используются СУБД ЛИНТЕР и Линтер-ВС.

Классификация СУБД

По «мощности» СУБД делятся на:

- ✓ **«Настольные»** – невысокие требования к техническим средствам, ориентация на конечного пользователя («дружелюбность» интерфейса, простота создания БД и обработки информации), низкая стоимость.
Пример: MS Access, Visual FoxPro.
- ✓ **Корпоративные** – обеспечивают работу в распределенной среде, высокую производительность, имеют развитые средства администрирования и более широкие возможности поддержания целостности. Системы сложны, дороги, требуют значительных вычислительных мощностей.
Примеры: Oracle, DB2, Sybase, MS SQL Server, Progress

Корпоративные СУБД, как правило, реализуют архитектуру клиент-сервер.

Помимо хранения централизованной базы данных центральная машина – **сервер базы данных**, должна обеспечивать выполнение основного объема обработки данных. Запрос на данные, выдаваемый клиентом (рабочей станцией), порождает поиск и извлечение данных на сервере. Извлеченные данные транспортируются по сети от сервера к клиенту. Спецификой архитектуры клиент-сервер является использование языка запросов SQL.



Классификация СУБД

По характеру использования СУБД делятся на:

- ✓ **Персональные.** Обеспечивают возможность создания персональных БД и недорогих приложений, работающих с ними. Персональные СУБД или разработанные с их помощью приложения зачастую могут выступать в роли клиентской части многопользовательской СУБД
Пример: MS Access, Visual FoxPro, Paradox.
- ✓ **Многопользовательские.** Включают в себя **сервер БД** и **клиентскую часть** и, как правило, могут работать в неоднородной вычислительной среде (с разными типами ЭВМ и операционными системами).
Примеры: Oracle, MySQL.

Классификация СУБД

По степени доступности БД выделяют:

✓ Общедоступные БД.

Примеры: Банк документов на сайте Президента Российской Федерации (<http://kremlin.ru/>), Информационно-правовая система «Законодательство России» (<http://pravo.gov.ru/>).

✓ БД с ограниченным доступом пользователей.

В качестве примера можно привести БД, используемые в системе органов внутренних дел (криминалистические учеты, розыскные учеты, оперативно-справочные учеты, автоматизированные банки данных дактилоскопической информации (АДИС "Папилон")

Классификация СУБД

По способу доступа к БД выделяют:

✓ Файл-серверные СУБД.

Примеры: Microsoft Access, Paradox, dBase, FoxPro, Visual FoxPro.

✓ Клиент-серверные СУБД.

Примеры: Oracle, Firebird, Interbase, IBM DB2, Informix, MS SQL Server, Sybase Adaptive Server Enterprise, PostgreSQL, MySQL, Caché, ЛИНТЕР

Классификация СУБД

В **файл-серверных СУБД** файлы данных располагаются централизованно на файл-сервере. СУБД располагается на каждом клиентском компьютере (рабочей станции). Доступ СУБД к данным осуществляется через локальную сеть. Синхронизация чтений и обновлений осуществляется посредством файловых блокировок.

Преимуществом этой архитектуры является низкая нагрузка на процессор файлового сервера.

Недостатки: потенциально высокая загрузка локальной сети; затруднённость или невозможность централизованного управления; затруднённость или невозможность обеспечения таких важных характеристик, как высокая надёжность, высокая доступность и высокая безопасность. Применяются чаще всего в локальных приложениях, которые используют функции управления БД; в системах с низкой интенсивностью обработки данных и низкими пиковыми нагрузками на БД.

На данный момент файл-серверная технология считается устаревшей, а её использование в крупных ИС — недостатком.³⁵

Классификация СУБД

Клиент-серверная СУБД располагается на сервере вместе с БД и осуществляет доступ к БД непосредственно, в монопольном режиме. Все клиентские запросы на обработку данных обрабатываются клиент-серверной СУБД централизованно.

Недостаток клиент-серверных СУБД состоит в повышенных требованиях к серверу.

Достоинства: потенциально более низкая загрузка локальной сети; удобство централизованного управления; удобство обеспечения таких важных характеристик, как высокая надёжность, высокая доступность и высокая безопасность.

Классификация СУБД

По форме представляемой информации выделяют фактографические, документальные, мультимедийные БД.

- ✓ фактографические БД, в которых хранится информация об интересующих пользователя объектах предметной области в виде «фактов» (facts – «данные» (англ.)). Например, данные о сотрудниках, данные о поставщиках и поставках продукции и т.п. При этом в качестве факта рассматривается неделимый по смыслу информационный элемент, отражающий значение какого-либо свойства объекта. Примеры: Oracle, PostgreSQL, MySQL, Microsoft Access.

Классификация СУБД

По форме представляемой информации выделяют:

- ✓ **документальные БД.** Единицей хранения является какой-либо документ (например, текст закона или статьи), и пользователю в ответ на его запрос выдается либо ссылка на документ, либо сам документ, в котором он может найти интересующую его информацию. К документальным БД относятся, например, базы данных научного цитирования: РИНЦ, Web of Science, Scopus и т.д. Большое распространение получили документальные справочные правовые системы КонсультантПлюс, ГАРАНТ, Кодекс.
- ✓ **мультимедийные БД,** содержащие мультимедийную информацию: картографические, видео-, аудио-, графические и др. Пример: банк данных дактилоскопической информации (АДИС "Папилон"³⁸).

Классификация СУБД

По типологии хранения – **локальные и распределенные**.

По функциональному назначению (характеру решаемых задач и, соответственно, характеру использования данных) – **операционные и справочно-информационные**. К последним относятся ретроспективные БД (электронные каталоги библиотек), которые используются для информационной поддержки основной деятельности и не предполагают внесения изменений в уже существующие записи, например, по результатам этой деятельности. **Операционные БД предназначены для управления различными технологическими процессами.**

Классификация СУБД и БД

По способу распространения:

- ✓ **Commercial Software** — коммерческие (с ограниченными лицензией возможностями на использование), разрабатываемое для получения прибыли.
Примеры: Oracle, Microsoft Access.
- ✓ **Freeware** — свободные, распространяемые без ограничений на использование, модификацию и распространение.
Примеры: MySQL.
- ✓ **Shareware** — условно-бесплатные.

Классификация СУБД

По характеру преобладающей обработки информации:

- ✓ **OLTP (On-Line Transaction Processing)** – системы оперативной обработки информации
- ✓ **OLAP (On-Line Analytical Processing)** – системы для сложной аналитической обработки информации

Характеристика	OLTP	OLAP
Преобладающие операции	Ввод данных, поиск	Анализ данных
Характер запросов	Преобладают простые транзакции и запросы	Сложные запросы
Характеристика хранимых данных	Оперативные, детализированные	Агрегированные, охватывающие большой период времени

Классификация СУБД

По используемой Модели данных

- ✓ Иерархические
- ✓ Сетевые
- ✓ Реляционные
- ✓ Объектно-ориентированные
- ✓ Объектно-реляционные
- ✓ NoSQL:
 - “Ключ-значение”
 - Документные
 - Поколоночные (столбцовые)

Банки данных

Основные термины и определения

Банк данных – система БД, программных, технических, языковых, организационно-методических средств, предназначенных для обеспечения централизованного накопления и коллективного многоцелевого использования данных.

Банк данных (Бнд) является разновидностью ИС, в которой реализованы функции централизованного хранения и накопления обрабатываемой информации, организованной в одну или несколько баз данных.

Предпосылки широкого использования БД

✓ объекты реального мира находятся в сложной взаимосвязи между собой. Это приводит к необходимости, чтобы их информационное отражение также представляло единое взаимоувязанное целое;

✓ функции создания и ведения информационного фонда и предоставления нужных данных тем или иным процессам являются универсальными общими при решении разнообразных задач;

✓ информационные потребности различных пользователей существенно пересекаются, что делает целесообразным использование единых баз данных и обеспечение доступа к ним разных пользователей;

✓ современный уровень развития технического и программного обеспечения, а также теории и практики построения информационных систем позволяют

Требования к БНД

- ✓ **адекватность** отображения предметной области;
- ✓ возможность **взаимодействия** пользователей разных категорий и в разных режимах;
- ✓ **дружелюбность интерфейсов** и малое время на освоение системы, особенно для конечных пользователей;
- ✓ обеспечение **секретности и конфиденциальности** для некоторой части данных: определение групп пользователей и их полномочий;
- ✓ обеспечение взаимной **независимости** программ и данных;
- ✓ обеспечение надежности функционирования БНД;
- ✓ приемлемые характеристики функционирования БНД (стоимость обработки, время реакции системы на

Преимущества и недостатки БНД

□ К числу преимуществ использования БНД относятся:

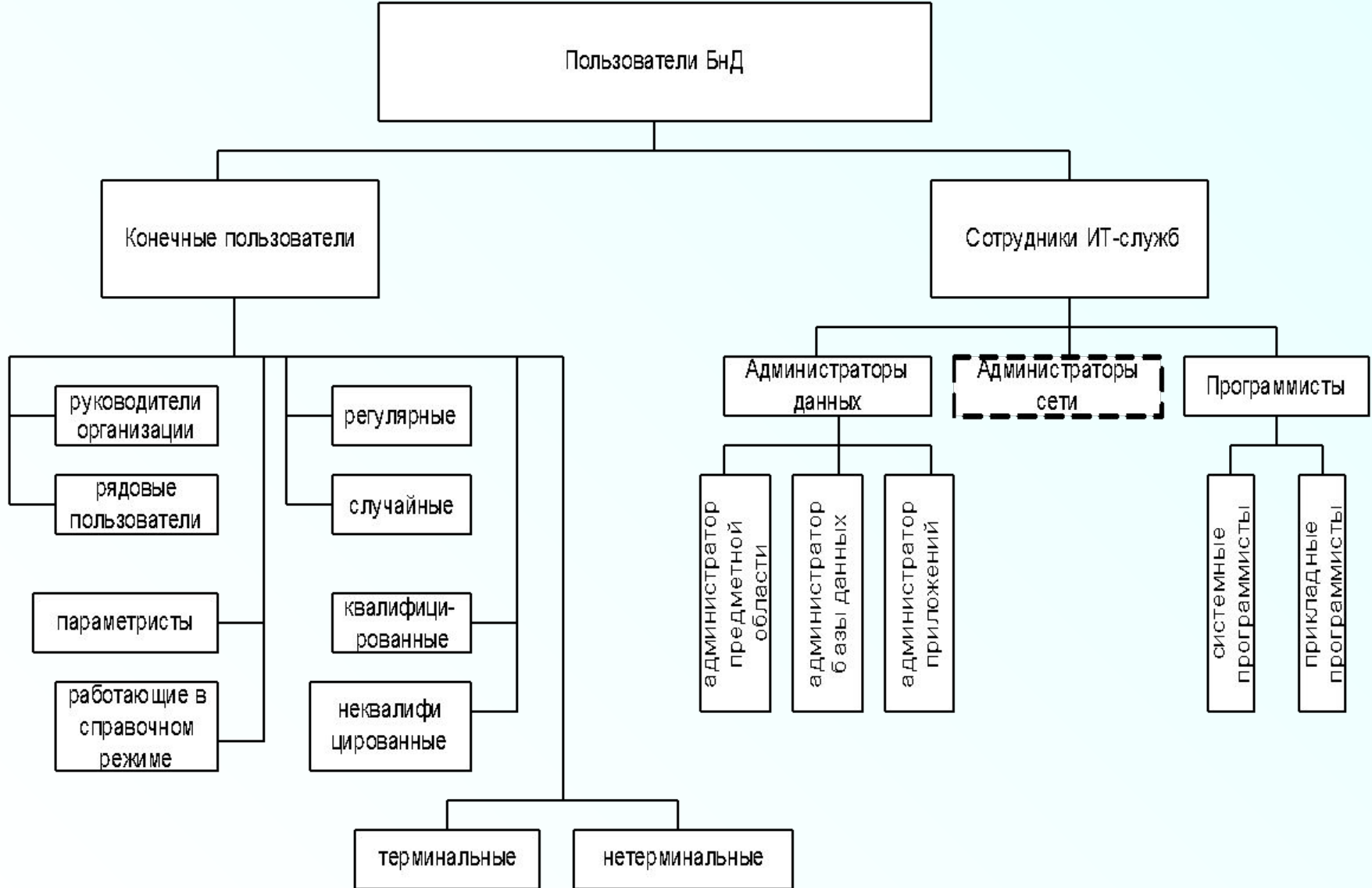
- ✓ непротиворечивость и целостность информации;
- ✓ возможность обращаться к БД не только при решении заранее определенных задач, но и с нерегламентированными запросами;
- ✓ сокращение избыточности хранимых данных;
- ✓ сокращение трудоемкости ведения БД;
- ✓ высокое качество управления данными.

Преимущества и недостатки БНД

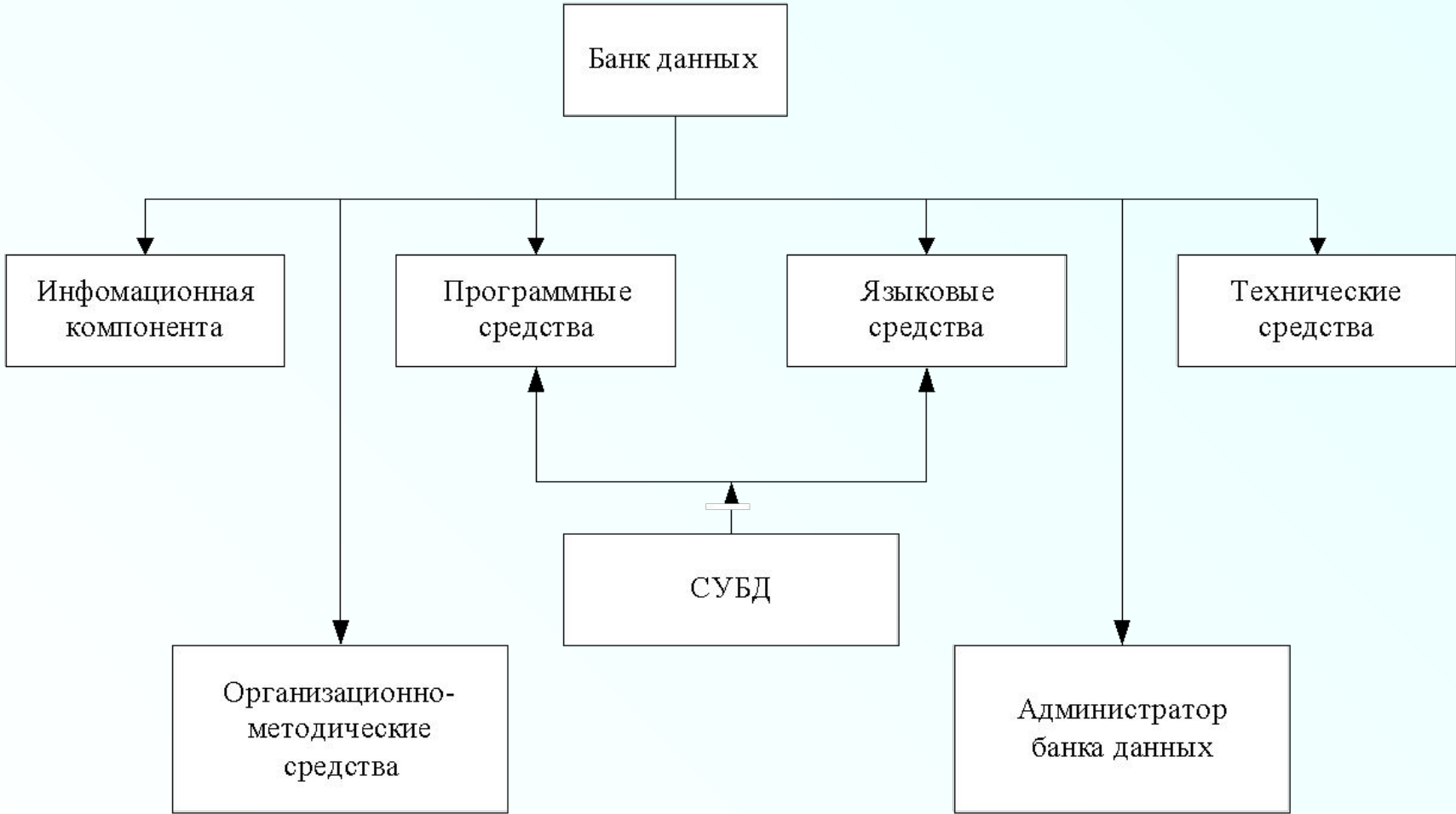
❑ Недостатки БНД:

- ✓ сложность;
- ✓ высокие требования к квалификации разработчиков БНД;
- ✓ в результате интеграции возможна некоторая потеря эффективности отдельных приложений (но общая эффективность всей системы будет выше);
- ✓ требуются специализированное программное обеспечение, которое, в зависимости от класса системы, может быть сравнительно дорогим, предъявляющим повышенные требования к техническим средствам;
- ✓ эксплуатация распределенных корпоративных БНД – процесс сложный и дорогостоящий.

Пользователи БД



Компоненты банка данных



Информационная компонента



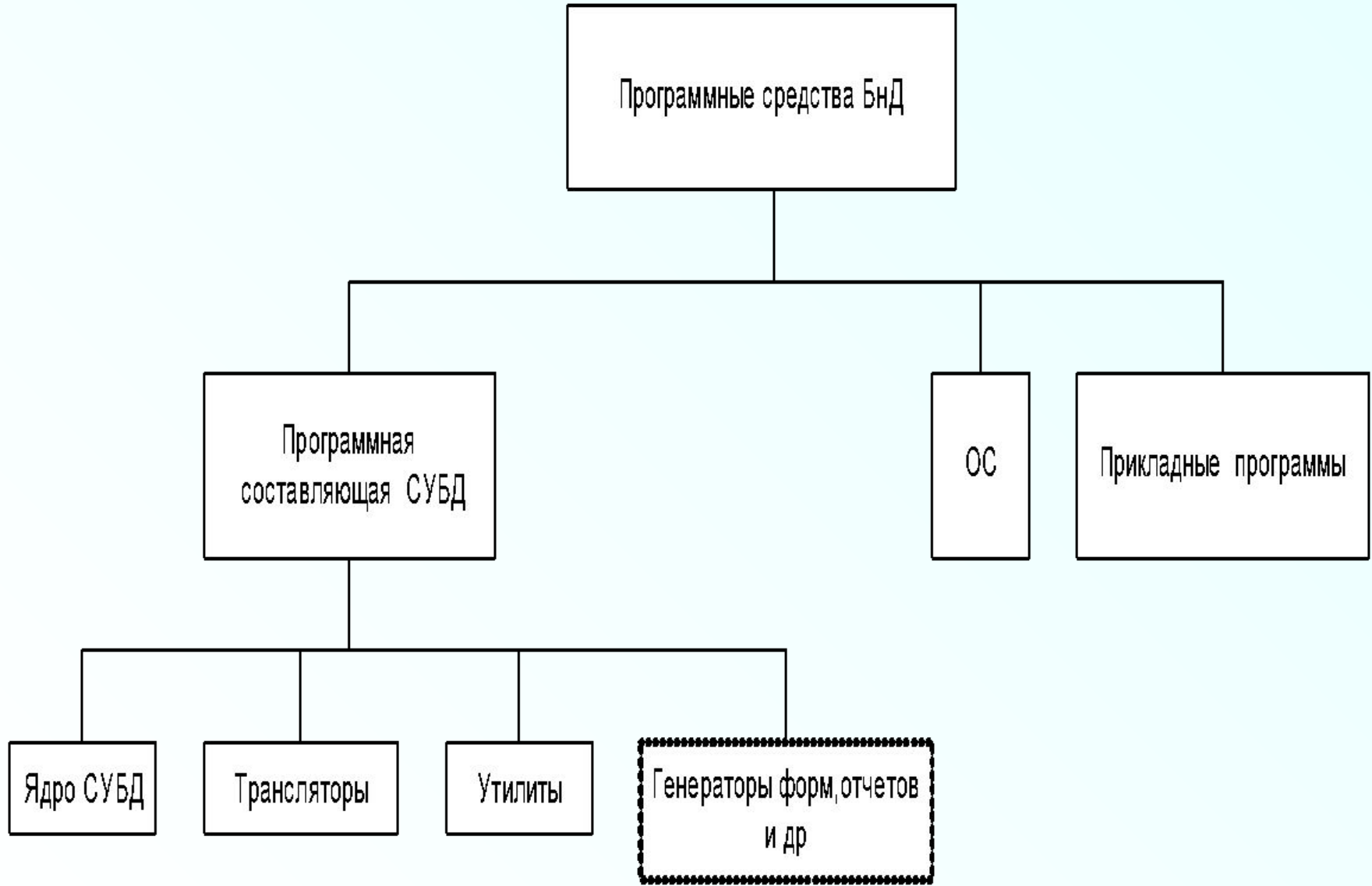
База данных.

В состав БД включаются не только собственно хранимые данные о предметной области, но и описания БД. Описания баз данных относятся к метайнформации, т.е. информации об информации. Описание баз данных часто называют схемой. Централизованное хранилище метайнформации называется *словарем данных (репозиторием)*.

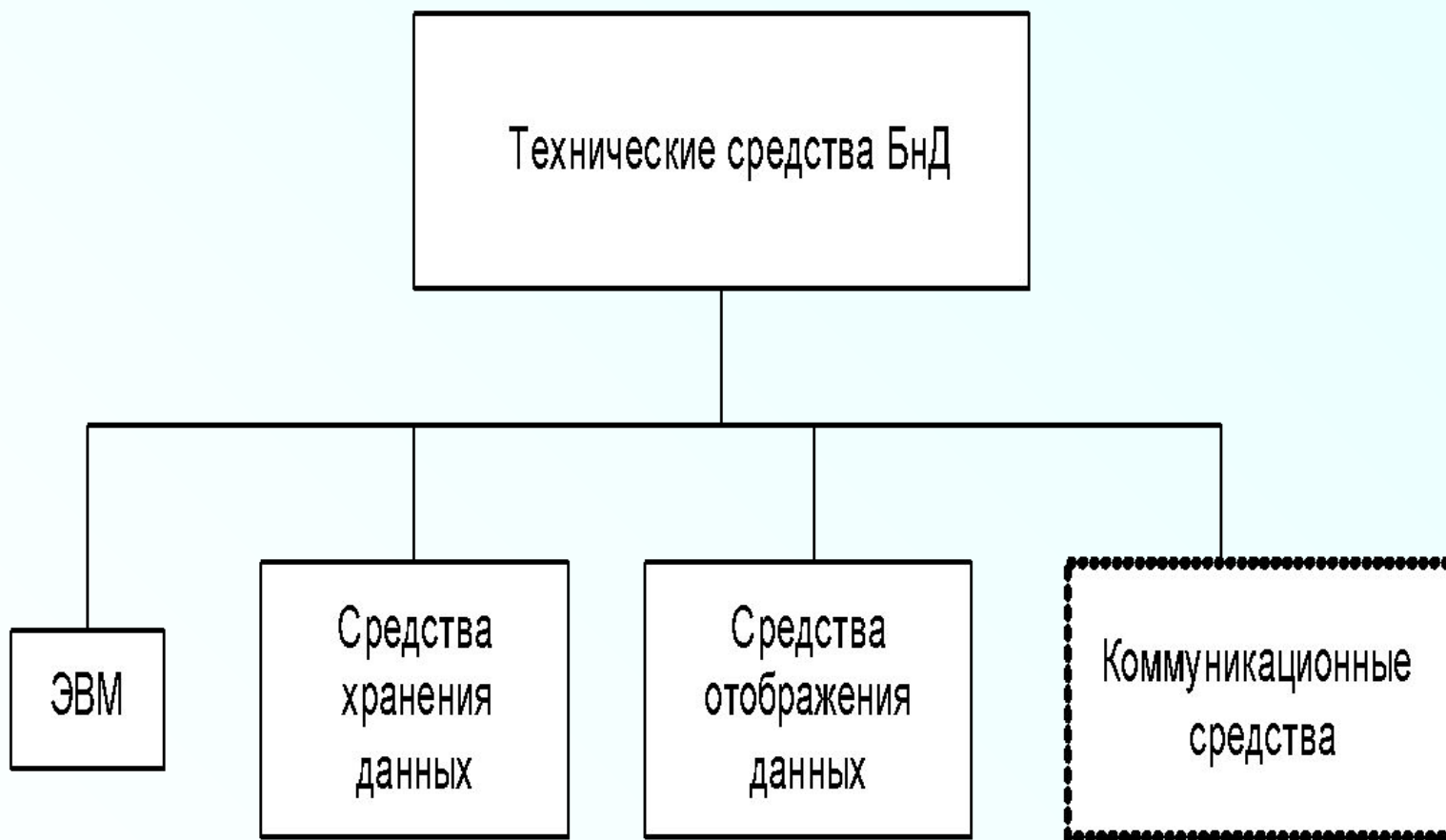


Словарь данных (СД) представляет собой подсистему БД, предназначенную для централизованного хранения информации о структурах данных, взаимосвязях файлов БД друг с другом, типах данных и форматах их представления, принадлежности данных пользователям, кодах защиты и разграничения доступа и т.п.

Программные средства БД

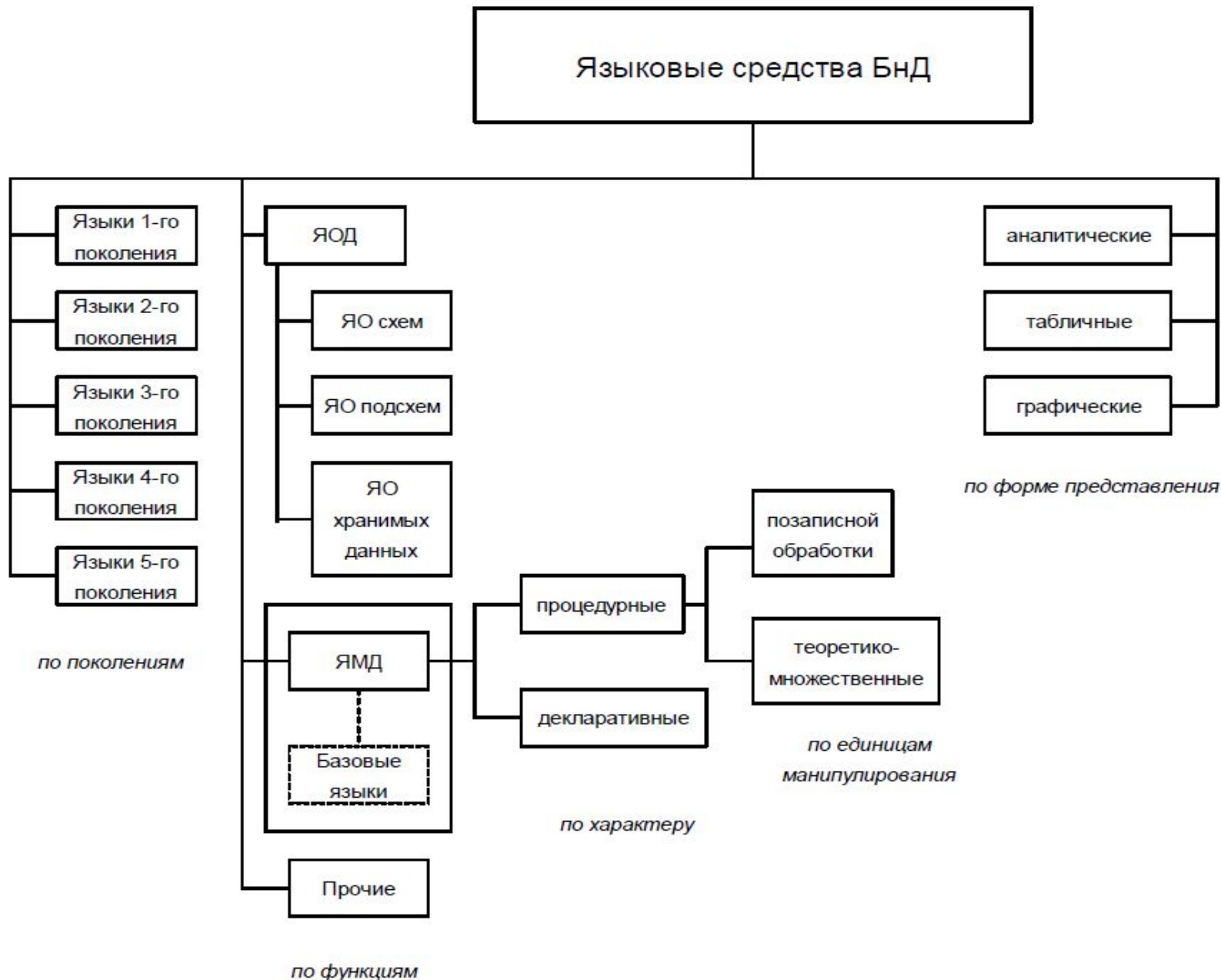


Технические средства банка данных



Технические средства БД

Языковые средства банка данных



Компоненты языка четвертого поколения



Языковые средства банка данных



В зависимости от функционального назначения различают **языки описания данных (ЯОД)** и **языки манипулирования данными (ЯМД)**. Иногда в особую группу выделяют языки запросов (ЯЗ).



ЯОД позволяет определять схемы базы данных, характеристики хранимых и виртуальных данных и параметры организации их хранения в памяти и может включать в себя средства поддержки целостности базы данных, ограничения доступа, секретности.



ЯМД обычно включает в себя средства запросов к базе данных и поддержания базы данных (добавление, удаление, обновление данных, создание и уничтожение БД, изменение определений БД, обеспечение запросов к справочнику БД).

Языковые средства банка данных



Языки манипулирования данными разделяются на две большие группы: **процедурные** и **непроцедурные**. При пользовании процедурными языками надо указать, какие действия и над какими объектами необходимо выполнить, чтобы получить результат. В непроцедурных языках указывается, что надо получить в ответе, а не как этого достичь.



Примерами непроцедурных языков являются языки, основанные на реляционном исчислении. Представителем языков, основанных на реляционном исчислении кортежей, является широко используемый язык запросов **SQL (Structured Query Language, язык структурированных запросов)**. Табличный язык **QBE (Query by Example, запрос по образцу)** также является непроцедурным языком⁷.

Спасибо за внимание!