

# Базы данных

## А1. Введение в базы данных



Московский государственный технический университет  
имени Н.Э. Баумана

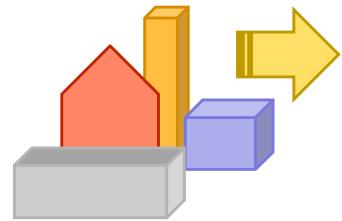
**Факультет ИБМ**

Фев 2025 года

Москва

Артемьев Валерий Иванович © 2025

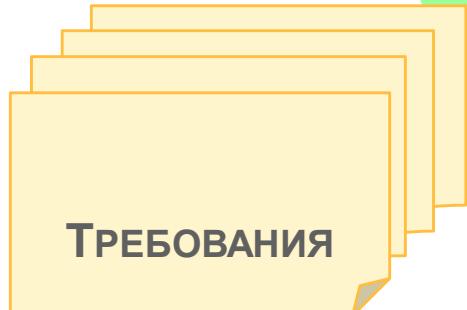
# От моделей к проектированию данных



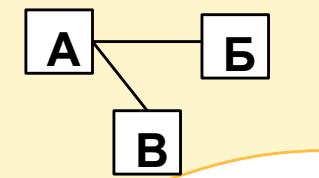
РЕАЛЬНЫЙ МИР

*Описательная  
деятельность*

документирование  
некоторых аспектов  
реального мира

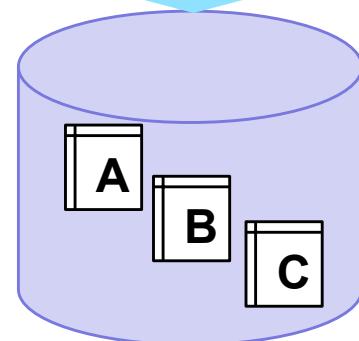


ТРЕБОВАНИЯ



МОДЕЛЬ ДАННЫХ

Автоматизация  
проектирования



СТРУКТУРЫ ДАННЫХ

# Виды информационных систем

**Информационная система (ИС)** – комплекс программно-технических средств *сбора, хранения, актуализации и обработки информации* для поддержки определённого вида деятельности (предметной области).

## Классы ИС по типу хранимой и обрабатываемой информации

**Документальные системы**  
(текст, мультимедиа, дескрипторы и полнотекстовой поиск, гипертекст): ИПС и библиотечные системы, электронный документооборот, сайты и порталы

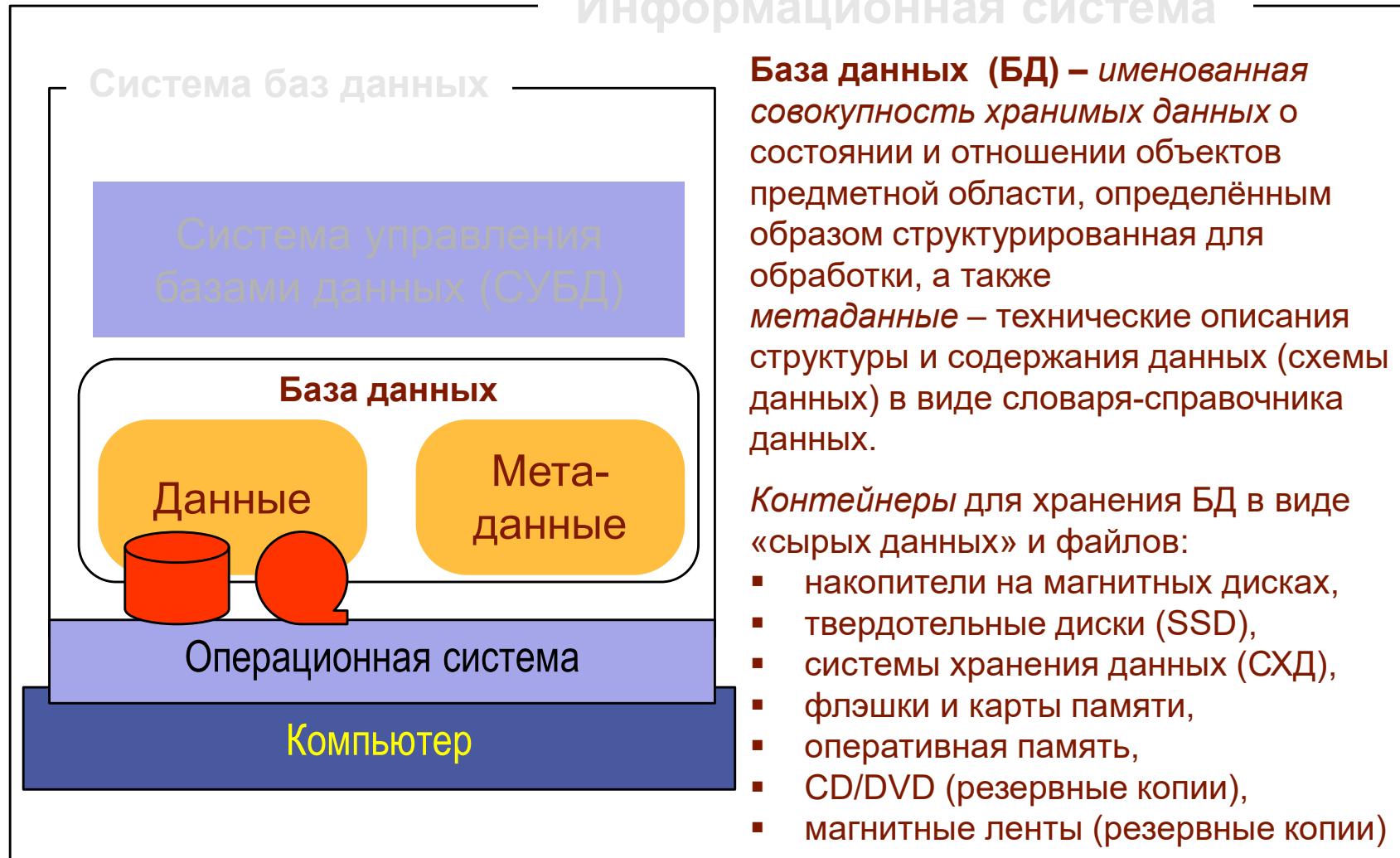
**Геоинформационные системы**  
(топология, электронная карта) – аналитические системы и ситуационные центры с пространственно-географическим компонентом (транспорт, ЖКХ, месторождения, архитектурное наследие и т.п.)

**Фактографические системы**  
(записи о фактах, БД) – учётно-операционные системы и системы поддержки принятия решений (аналитические системы)

**Системы искусственного интеллекта**  
(временные ряды, языковые модели и т.п.) прогнозные модели, рекомендательные и предписывающие системы, LLM-системы

Всё чаще начинают преобладать **гибридные ИС**

# Понятие базы данных



**База данных (БД) – именованная совокупность хранимых данных о состоянии и отношении объектов предметной области, определённым образом структурированная для обработки, а также метаданные – технические описания структуры и содержания данных (схемы данных) в виде словаря-справочника данных.**

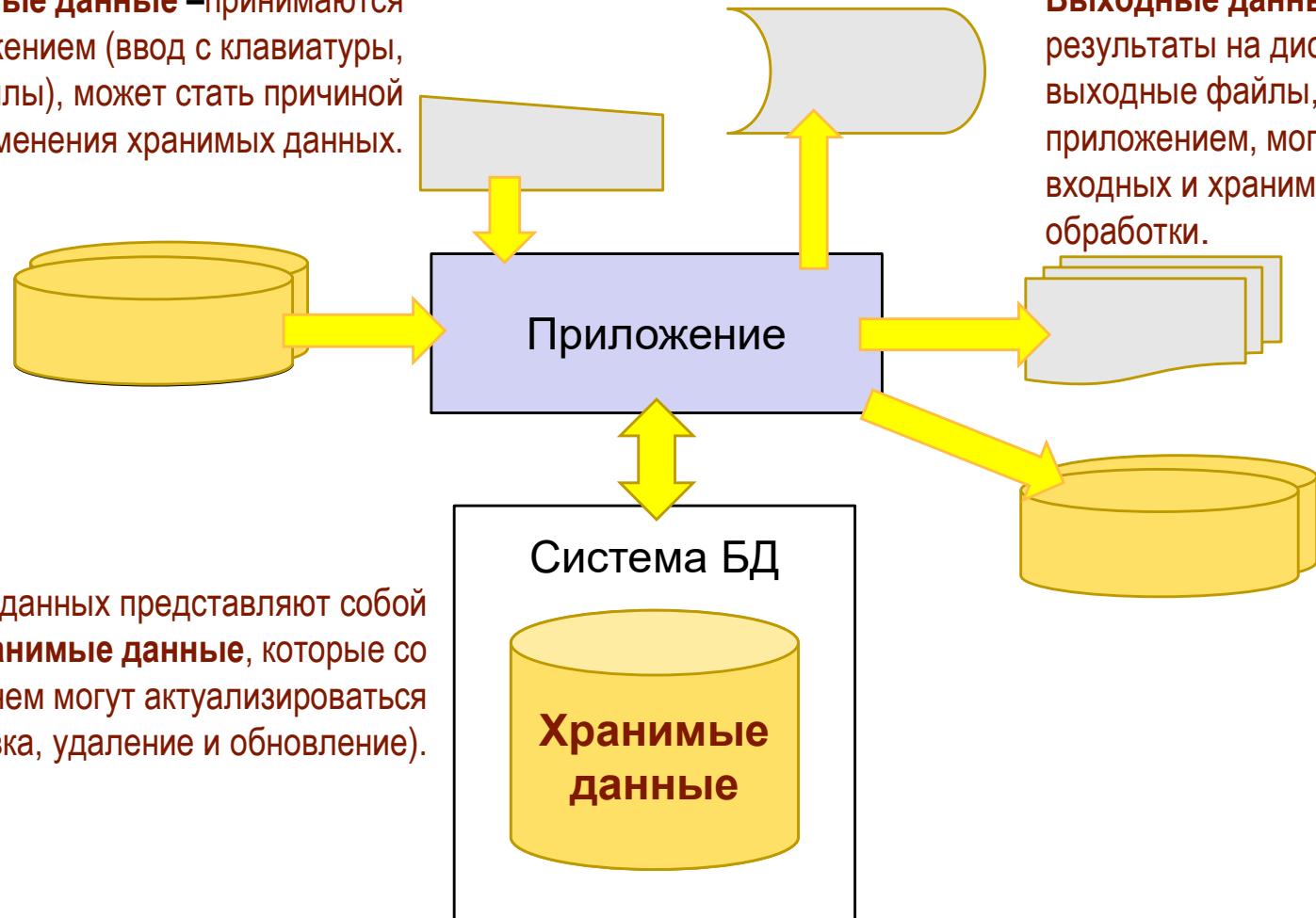
Контейнеры для хранения БД в виде «сырых данных» и файлов:

- накопители на магнитных дисках,
- твердотельные диски (SSD),
- системы хранения данных (СХД),
- флэшки и карты памяти,
- оперативная память,
- CD/DVD (резервные копии),
- магнитные ленты (резервные копии)

Наличие метаданных позволяет говорить о «самодокументированности» БД

# Входные, выходные и хранимые данные

**Входные данные** –принимаются приложением (ввод с клавиатуры, файлы), может стать причиной изменения хранимых данных.



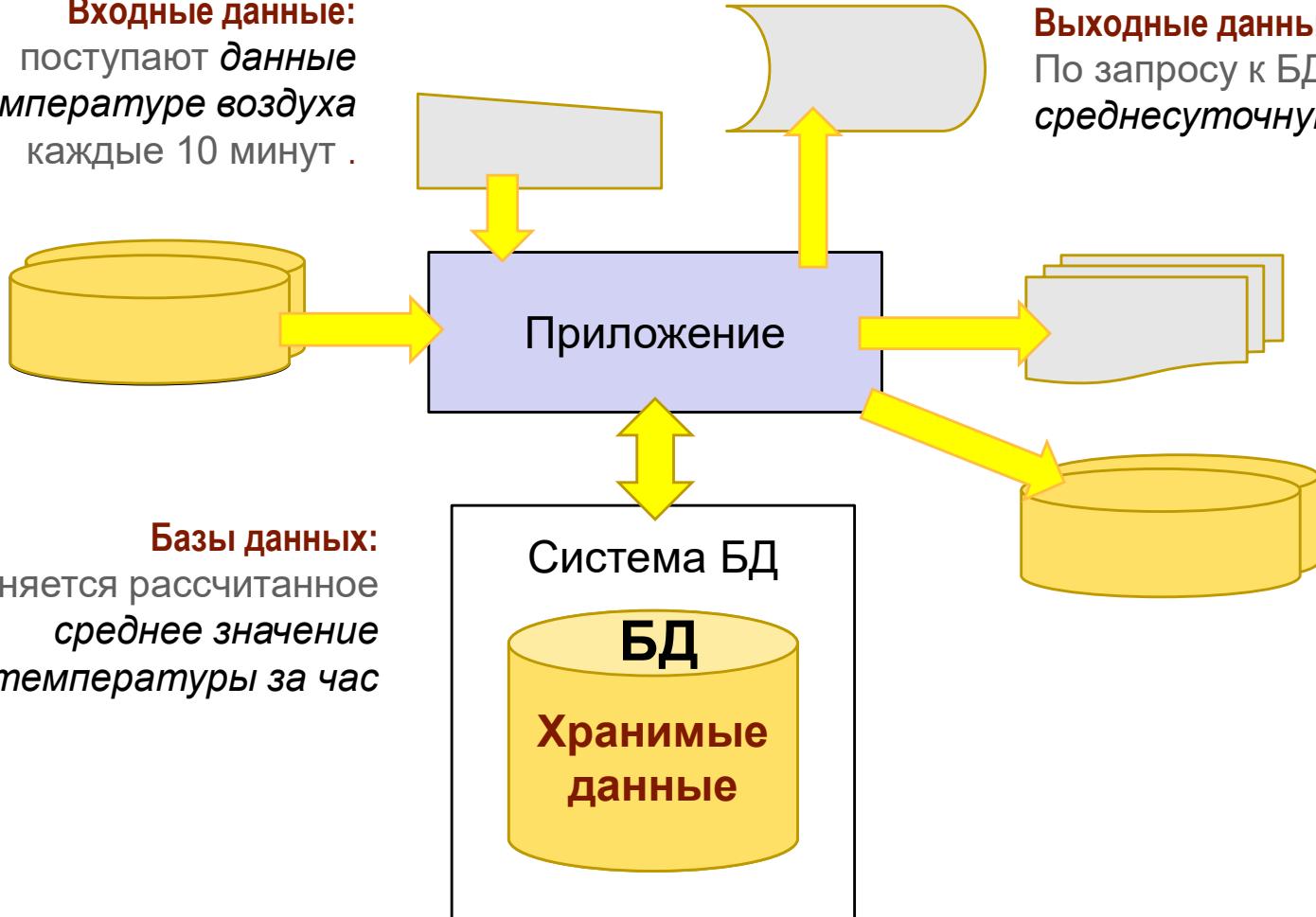
Базы данных представляют собой **хранимые данные**, которые со временем могут актуализироваться (вставка, удаление и обновление).

**Выходные данные** – сообщения, результаты на дисплей / печать, выходные файлы, выдаваемые приложением, могут получаться из входных и хранимых данных после их обработки.

В общем случае транзитные (входные и выходные) данные не являются частью Бд.

# Пример различия входных, выходных и хранимых данных

**Входные данные:**  
поступают *данные о температуре воздуха* каждые 10 минут .

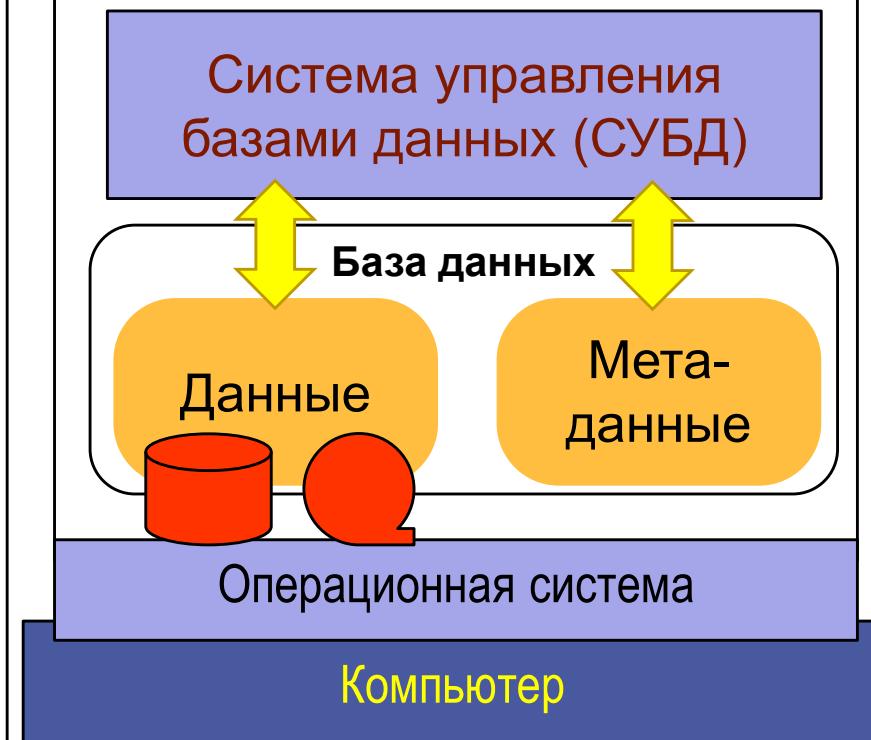


В этом случае хранимые значения будут отличаться и от входных и от выходных данных.

# Понятие СУБД и системы баз данных

## Информационная система

### Система баз данных



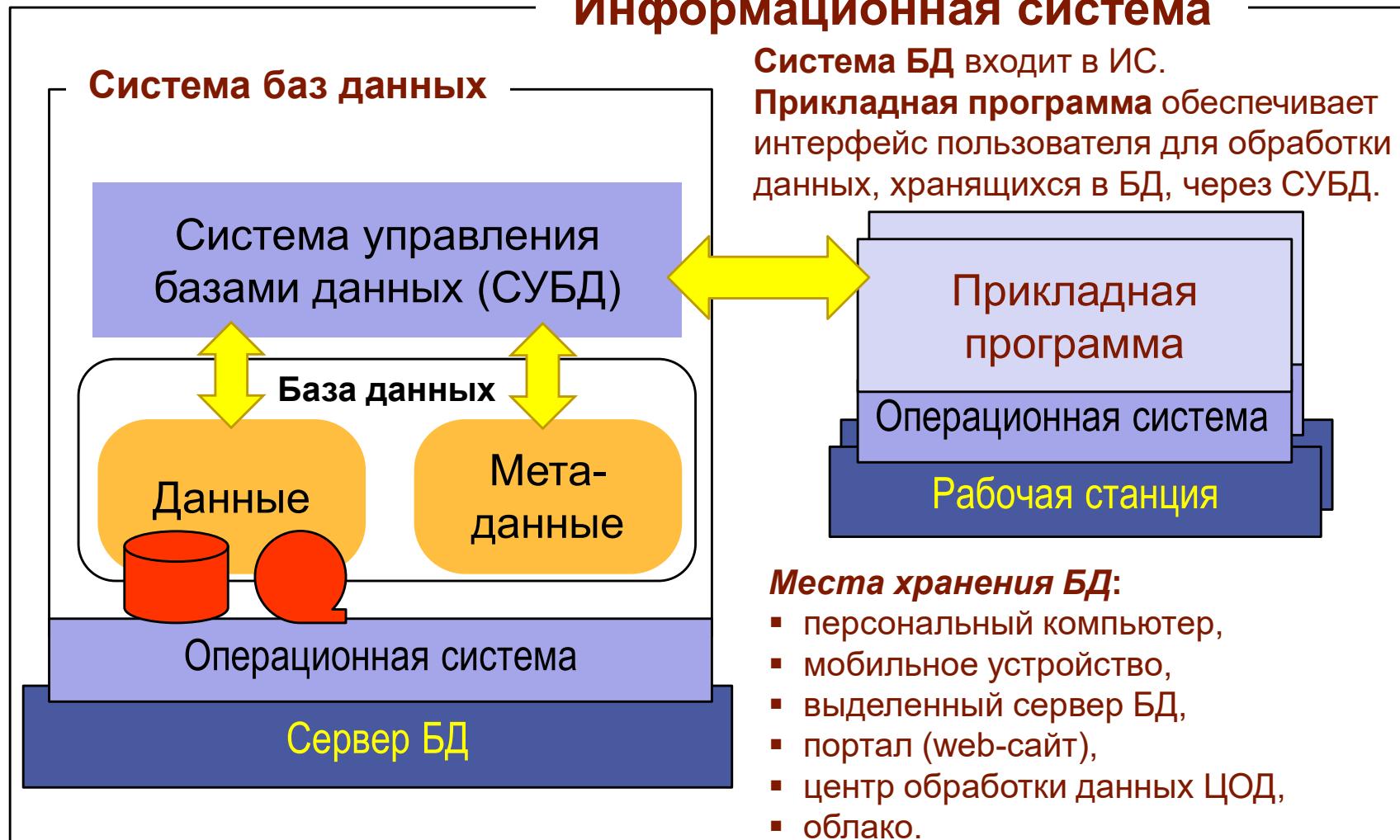
**Система управления базами данных (СУБД)** – системное ПО для создания, ведения и совместного использования БД многими пользователями. Предоставляет разработчику приложения и пользователю программный интерфейс и языковые средства для работы с БД, не вникая в детали аппаратного и даже программного обеспечения.

**Система баз данных** (устар.: *банк данных*) – комплекс программно-технических средств, содержащих базу данных и систему управления данными для их коллективного многоцелевого применения.

Не путайте БД, СУБД и систему БД!

БД без СУБД и прикладного программного обеспечения «не живут»!

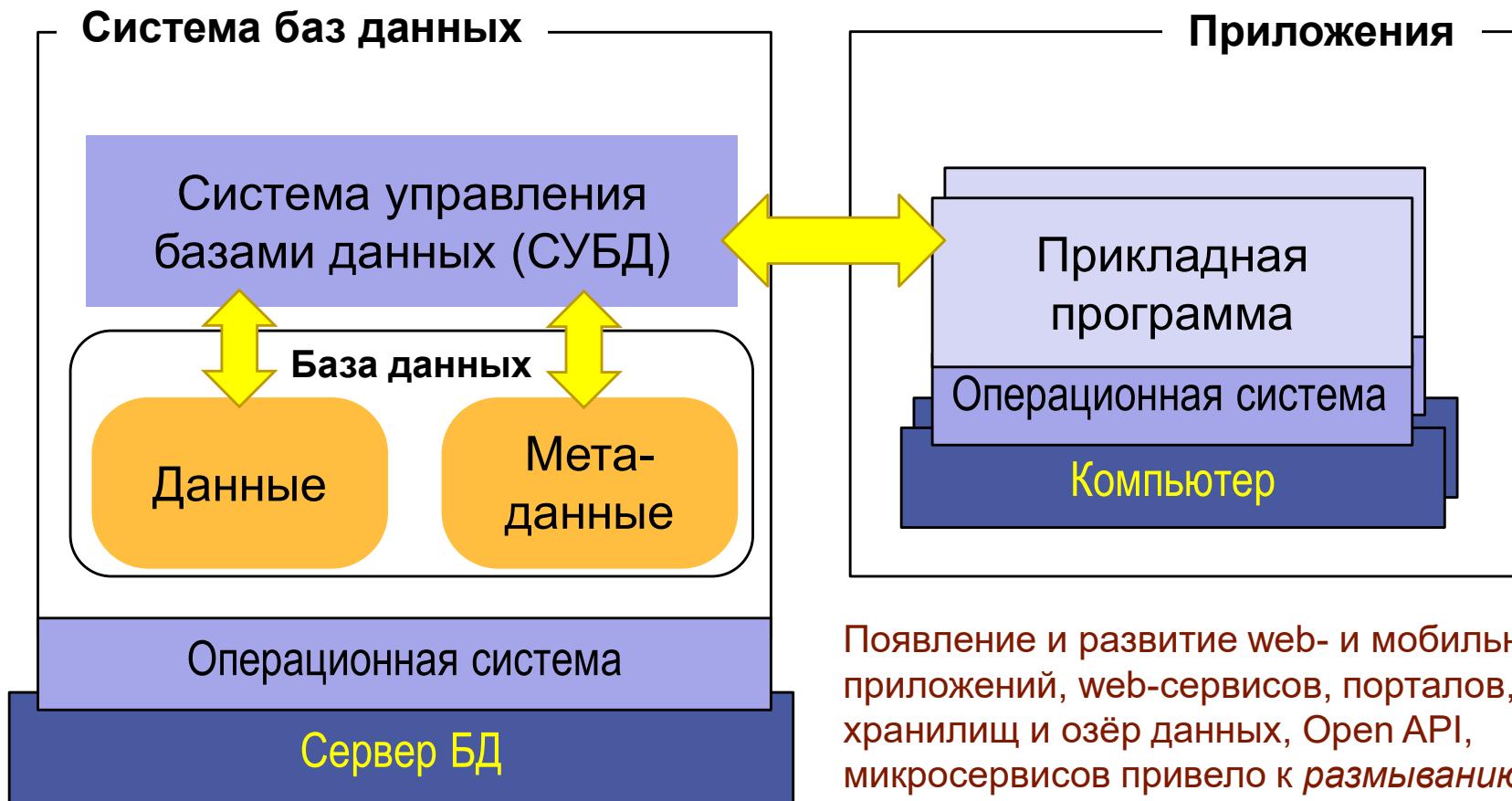
# Место системы баз данных в ИС



**Ориентация на приложение:** раньше система БД являлась частью ИС

# Современный взгляд на отношение приложений и системы БД

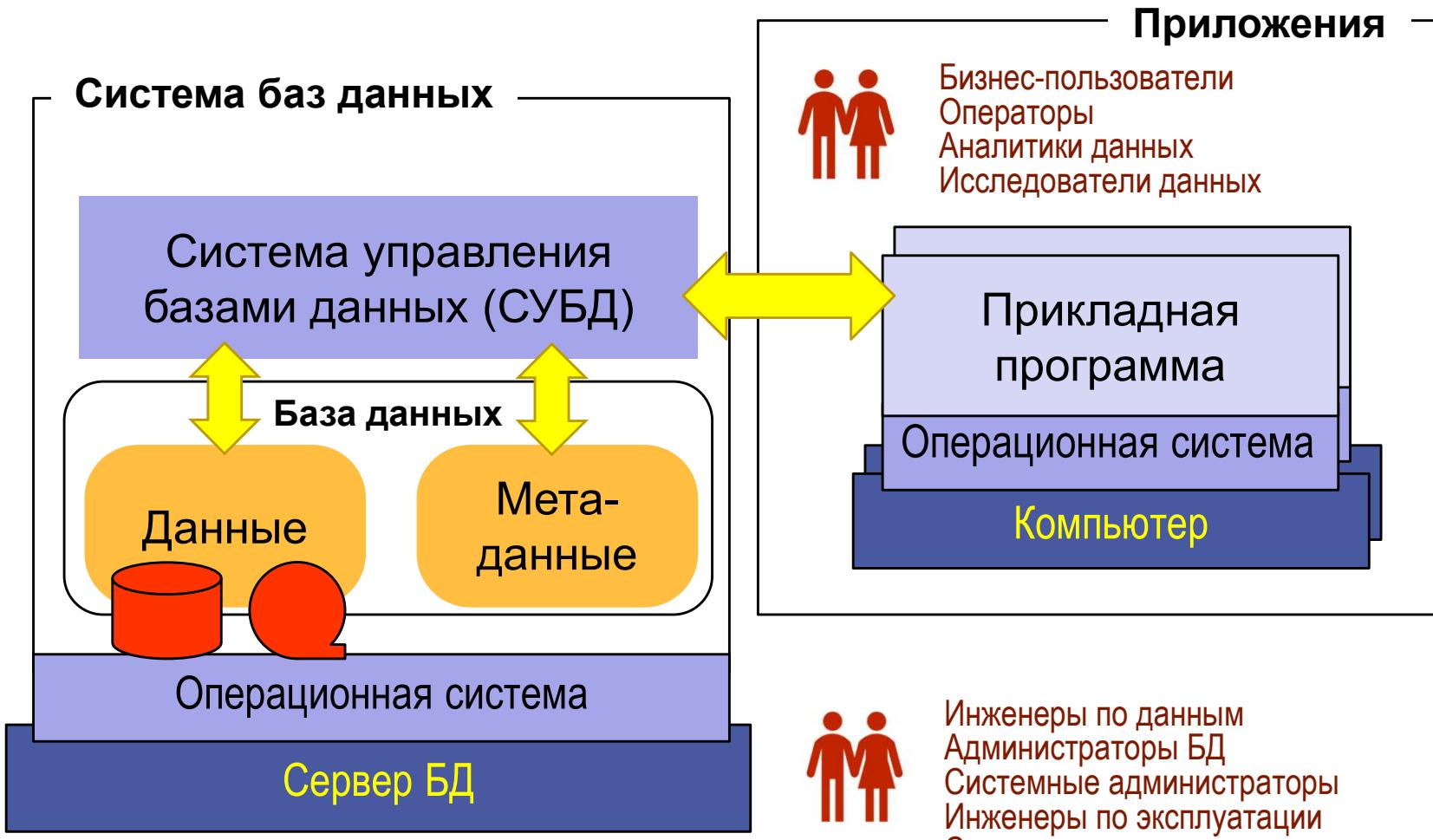
## Информационная система



Появление и развитие web- и мобильных приложений, web-сервисов, порталов, хранилищ и озёр данных, Open API, микросервисов привело к размытию понятия ИС и отделению систем БД.

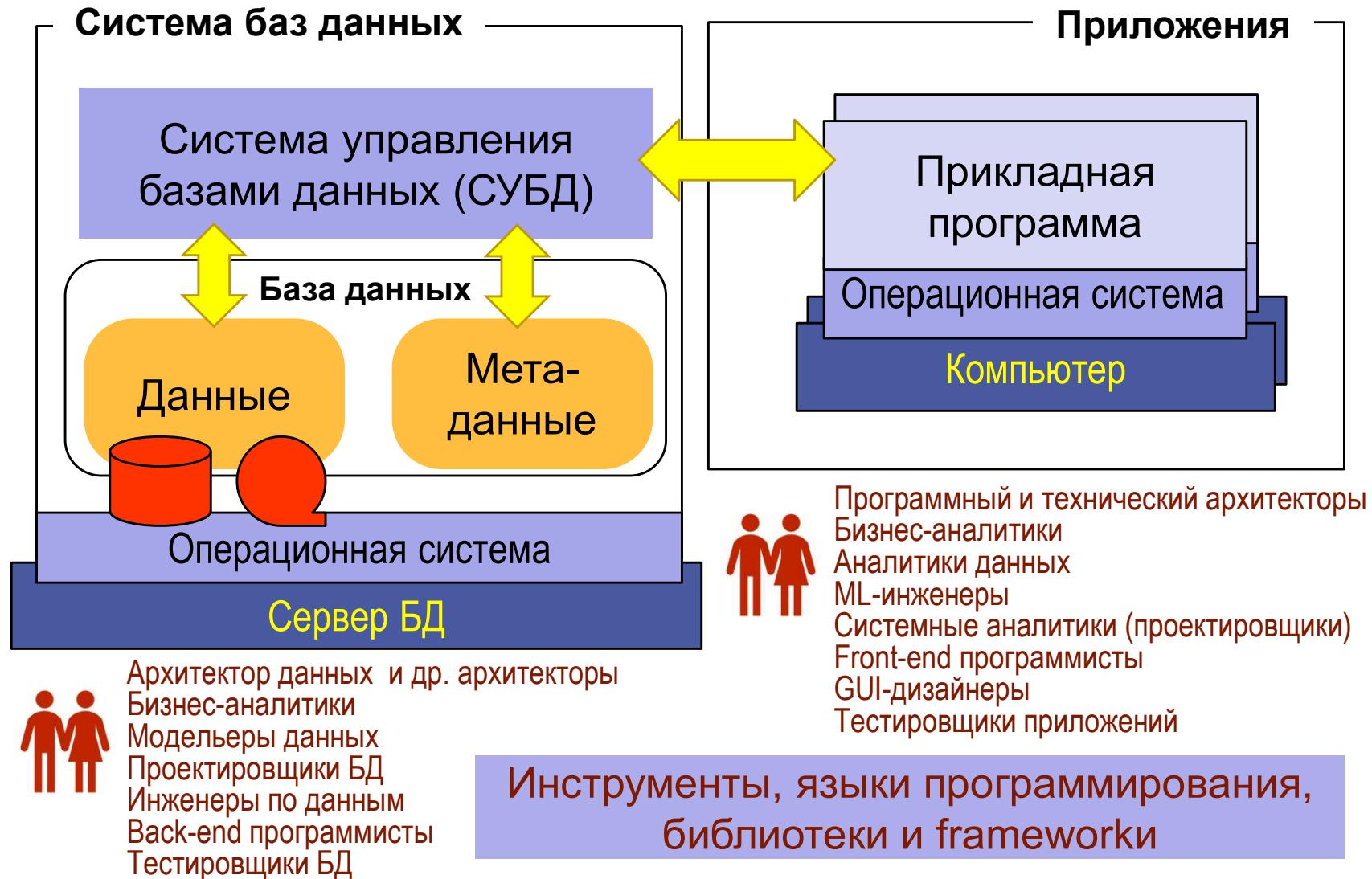
**Датацентрический подход:** обособление и обобществление систем БД

# Роли в системе БД и приложении



Новые роли: аналитики и исследователи данных, инженеры по данным

# Роли при создании БД и приложения



Новые роли: аналитики и исследователи данных, инженеры по данным

# Бумажные носители данных



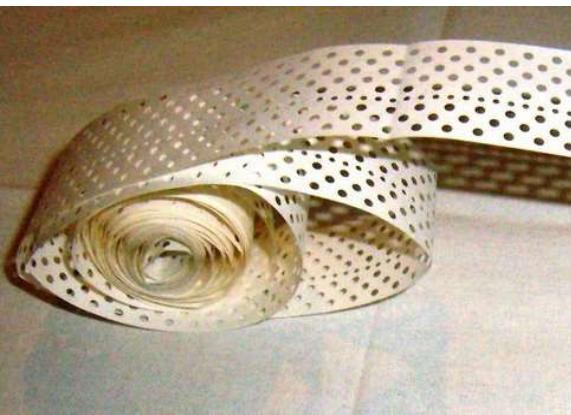
# Перфокарты: Ж.М. Жаккард, ткацкий станок

## **Ленты перфокарт:**

Б. Бушон,  
Ж.Б.Фалькон,  
Ж. де Вокансон  
ткацкие станки и  
игрушки-роботы

XVIII B.

1804



# Перфоленты телефон, ограниченное применение для ЭВМ

## Транспортные карты, учёт работы, шарманка

1846

**Табуляторы  
перфокарт:  
Г. Холлерит,  
перепись  
населения США,  
22 т карт, 1 Гб**

1890

# Устройства подготовки, ввода-вывода на перфокартах: колоды карт для хранения данных и программ на больших ЭВМ

# Основание фирмы IBM

## **Счётные, сортирующие и суммирующие табуляторы: бухучёт, инвентаризация, расчёт зарплаты, планирование и т.п**

1960-е – 1990-е годы

## **1950-е – 1980-е годы**

# Электронные носители данных



Дисковые  
массивы:  
SAN/FC, NAS

Дисковые  
массивы:  
RAID

CD-диски



DVD-диски  
Электронные  
архивные  
системы  
(кассеты 12 Тб)

SSD, Израиль  
Флэш-память,  
карты памяти



Магнитная  
кассета:  
Phillips

Жёсткие  
магнитные диски:  
IBM

5 млн символов  
сейчас 20 Тб

Магнитная  
лента:  
ЭВМ UNIVAC I

300 Мб

Магнитный  
барабан:  
расширение ОП

Магнитная  
плёнка:  
Ф. Пфлюмер

1928

1932

1951

1956

1963

1969

1970-е

1980-е

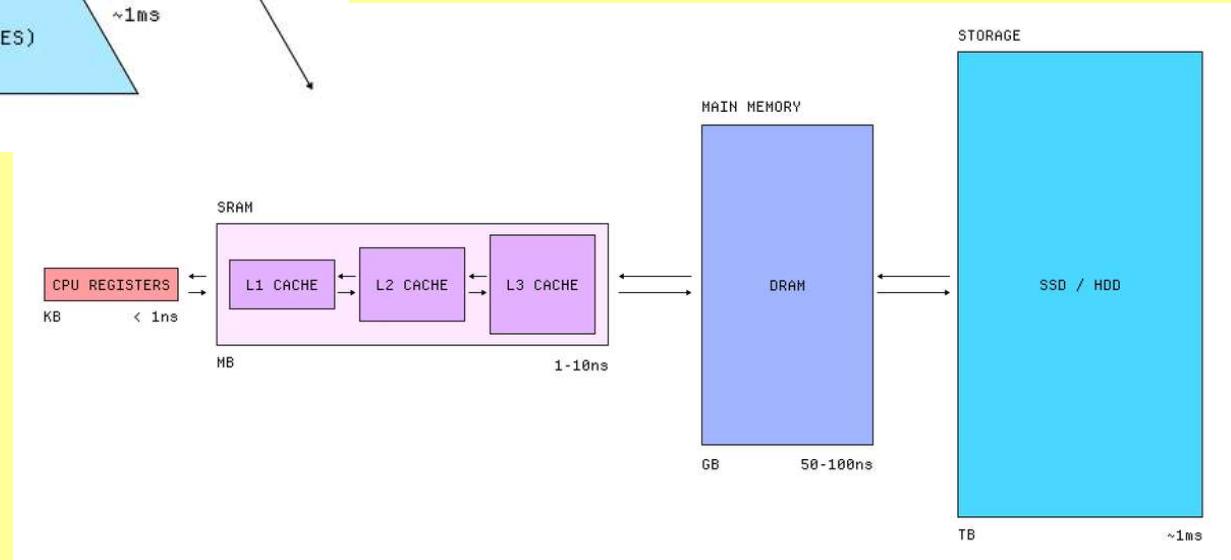
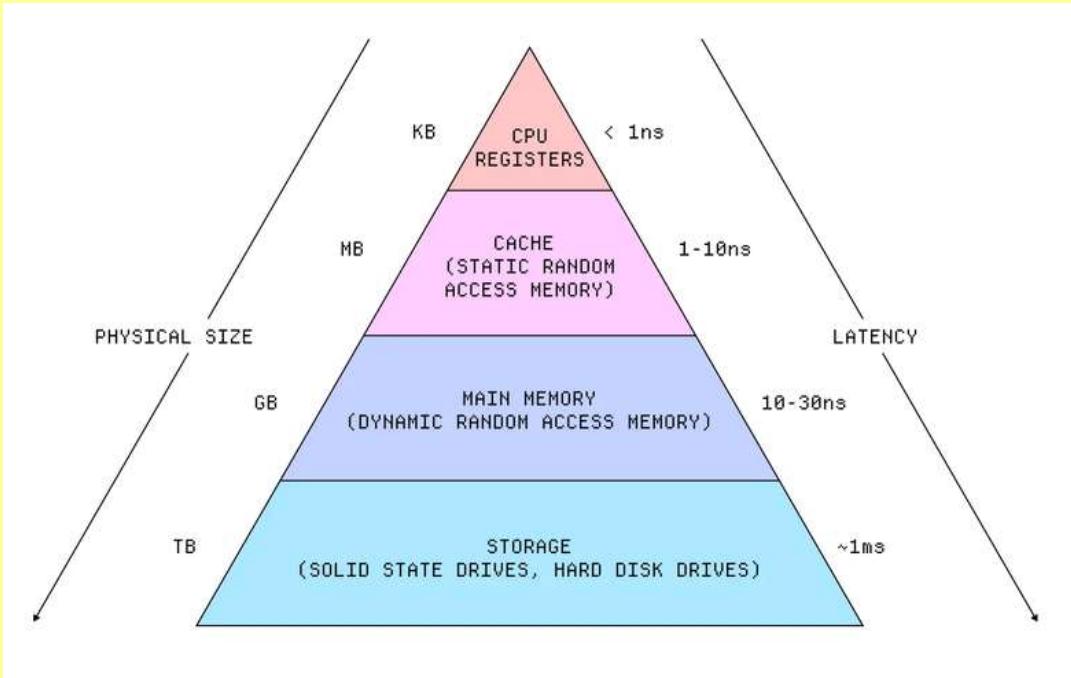
1984

1995

Кон. 1990-х

# Иерархия систем хранения данных





# Исторически сложились 2 направления для работы с данными на ЭВМ

	Виды обработки	Объёмы данных	Структуры данных	Области применения
1	Ввод-вывод и временное хранение данных, сложные расчёты, двоичная арифметика	небольшие	простые	Дешифрация, вооружения, физика, космос, математика, инженерия и т.п.
2	Постоянное хранение, поиск, вставка, обновление, удаление данных и простые расчёты, десятичная арифметика	относительно большие	средние по сложности и сложные	Перепись населения, бухгалтерский учёт, кадровый учёт, платежи, страхование, библиотеки и т.п.
3	Накопление и аналитическая обработка данных	большие	сложные	Принятие решений, анализ и исследование данных, прогнозирование, машинное обучение, ИИ

# Программы–предшественники СУБД



Операции ввода-вывода в **FORTRAN**  
PUNCH / PRINT,  
OPEN / CLOSE  
READ/ WRITE TAPE,  
REWIND, BACKSPACE

1957



**COBOL – язык программирования**  
обработки бизнес-данных

1959



**BSAM, BISAM, BDAM –**

последовательный,  
индексно-последовательный  
и прямой методы доступа  
к наборам данных в OS/360

**IBM RPG –**  
**генератор отчётов**

Цикл обработки записей  
данных, переквалификация  
операторов табуляторов



1964

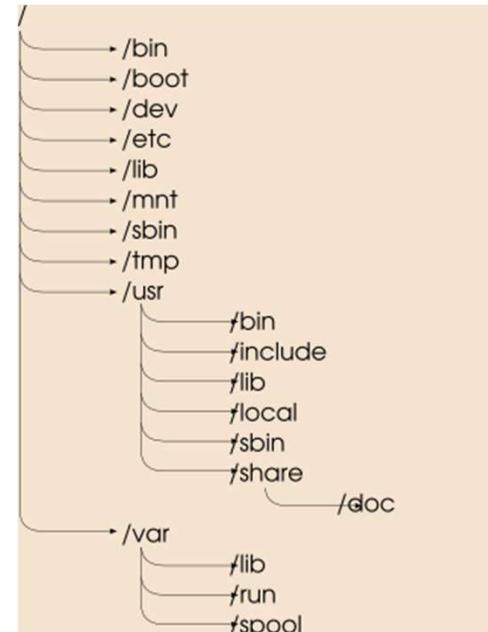
1960-е

**Файловые системы**

ОС CTSS и UNIX

**Утилиты** копирования

и сортировки данных



1970-е

# Язык программирования обработки бизнес-данных COBOL



1959 г.

«Естественный  
английский» язык  
программирования.

Описание структуры  
данных, десятичные числа.

Основной язык разработки приложений для сфер экономики,  
финансов, гос. управления до начала 1970-х годов.

До появления магнитных дисков преобладала  
последовательная обработка сортированных  
наборов данных на магнитных лентах.

Развивается и используется по сей день.  
Много унаследованных приложений.  
70% банковских транзакций в США.

Создатели: Комитет **CODASYL**, группа фирм,  
рук. **Грейс Хоппер**, контр-адмирал ВМС США.

```
LOOP.  
    READ IN-FILE, AT END GOTO EOD.  
    ADD AMOUNT TO SUM.  
    GOTO LOOP.  
EOD.  
    MOVE SUM TO TOTAL-AMOUNT.  
    CLOSE IN-FILE.
```



# Методы доступа к данным и наборы данных в IBM OS/360



**BSAM** – базисный последовательный метод доступа к данным

**QSAM** – последовательный метод доступа к данным с очередями

**BISAM** – базисный индексно-последовательный метод доступа к данным

**QISAM** – индексно-последовательный метод доступа к данным с очередями

**BRAM** – библиотечный метод доступа

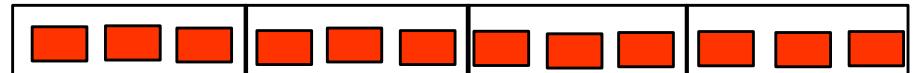
**BDAM** – прямой (произвольный) метод доступа к данным

Были заложены основы обработки данных  
на внешних носителях

Развитие этих методов использовано в первых СУБД

# Последовательный метод доступа к данным SAM

## Последовательная обработка данных



Не сортированные / сортированные записи

Не сблокированные / сблокированные записи В

Форматы фиксированной F и переменной длины V

блок

$b$  записей

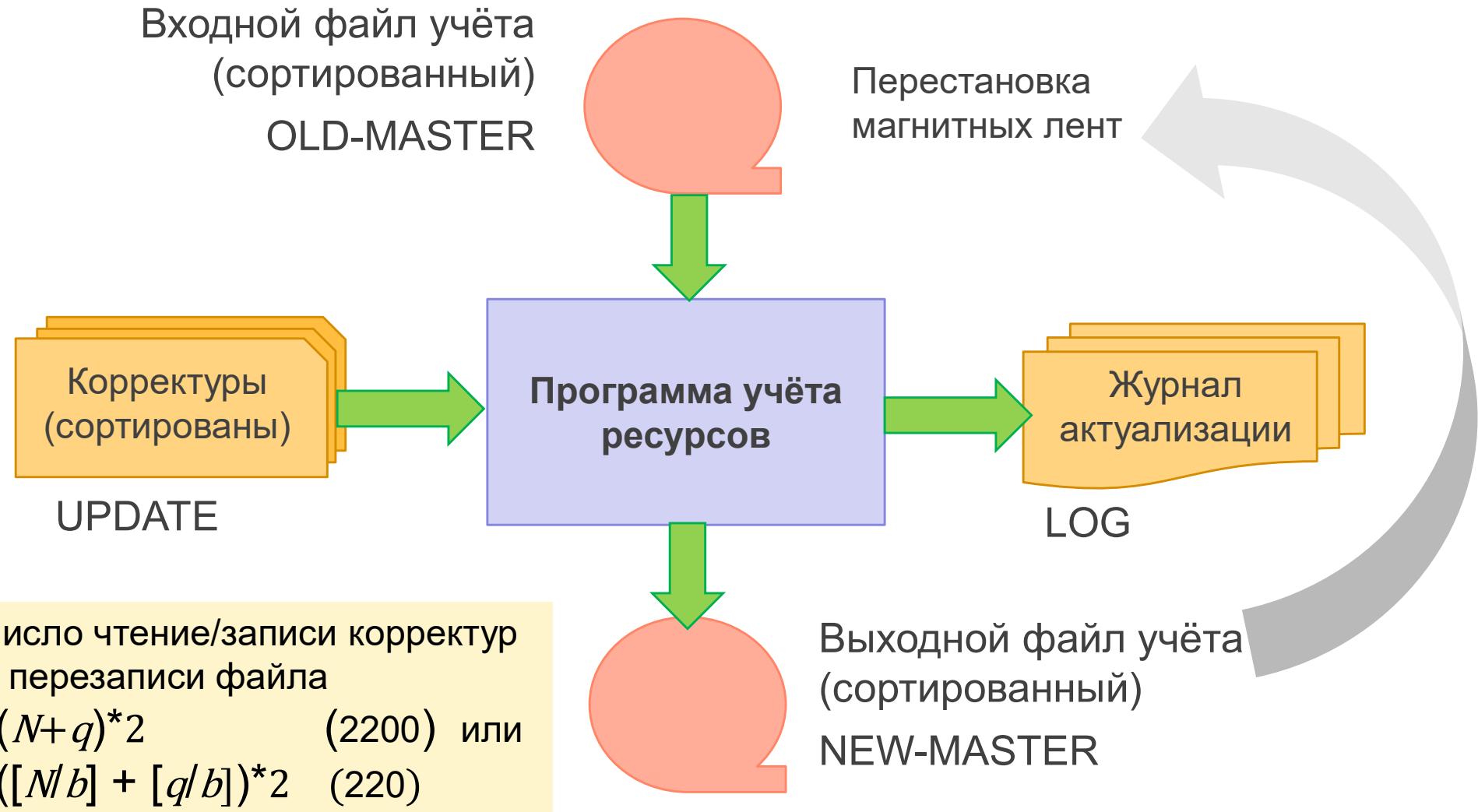
Поиск записей ( $N=1000$ ;  $b=10$ ;  $q=100$ )

Записи сортированы?	Записи сблокированы?	Ср. число чтений при поиске		Пример
		одной записи	$q$ записей	
нет	нет	$N/2$	$q^* N/2$	500; $50^*10^3$
да	нет	$N/2$	$(N+q)/2$	500; 550
нет	да	$[N/b]/2$	$q^*[N/b]/2$	50; $5^*10^3$
да	да	$[N/b]/2$	$([N/b] + [q/b])/2$	50; 55

Сортировка была важна при последовательном пакетном поиске.

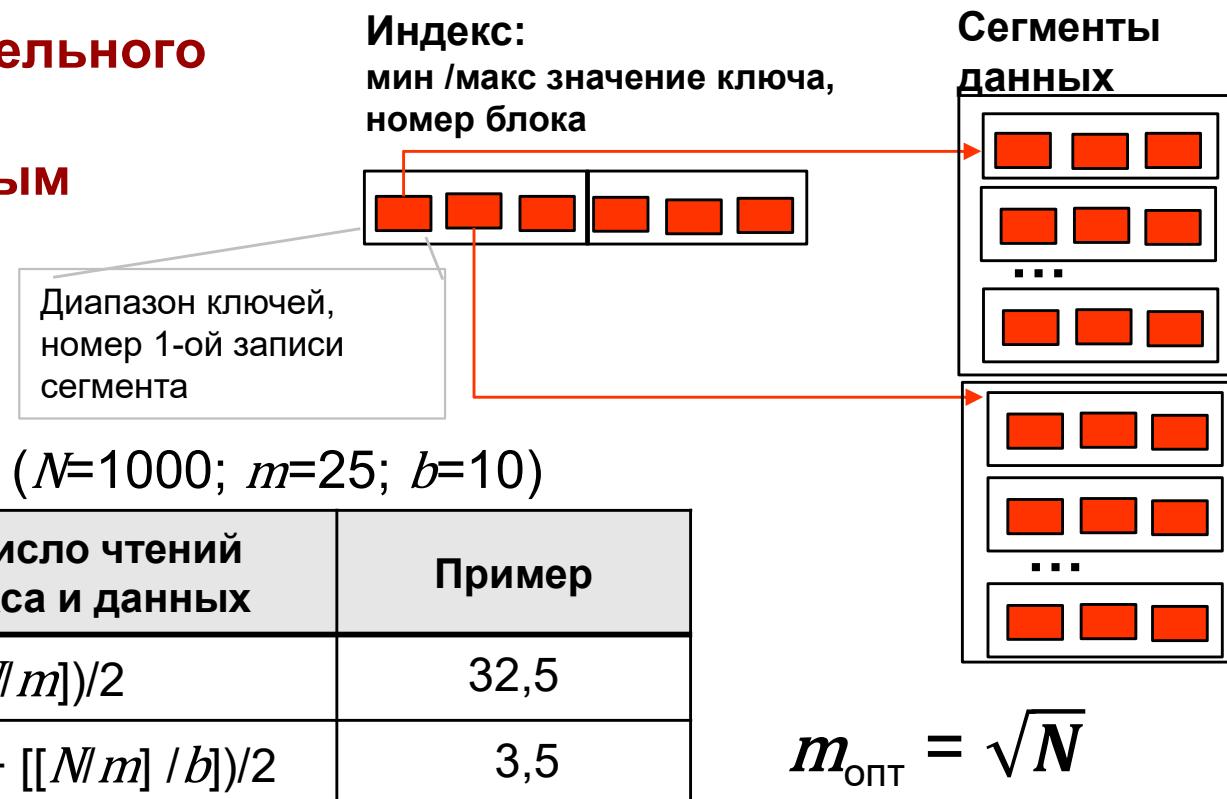
Сблокированные записи увеличивают производительность

# Последовательное (пакетное) обновление записей



# Индексно-последовательный метод доступа к данным ISAM

Комбинация последовательного и прямого доступа к сортированным данным



Изменение данных сложнее: от обновления по месту до вставки/ удаления с возможной реорганизацией сегмента и корректировкой индекса

Разработан ещё для пула МЛ, важна сортировка записей.  
Актуален для дисков, его модификации используется в СУБД.

# Прямой метод доступа к данным

- К **данным на дисках** возможен **прямой доступ**
- Проблема соответствия ключа номеру записи
- При наличии такого прямого соответствия – получаем **разреженные наборы записей**
- Для **плотных** наборов записей используют **разные способы преобразования:**
  - Хеширование индексы (кластеры)
  - Древовидные индексы, как правило, В-деревья
  - Битовые индексы (по столбцам)
  - Индекс-таблица = индекс + данные



# Появление систем управления БД



Иерарх. СУБД IMS,  
язык DL/1 – комитет  
CODASYL, IBM

Сетевая СУБД IDS,  
GE, Ч. Бахман

1960-е

1967

Реляционный  
подход, Э. Кодд

РСУБД System R, язык SEQUEL-2 –  
основа SQL, Д. Чемберлин  
РСУБД SQL/DS и DB2 (IBM),  
РСУБД Ingres, язык Quel, UCB,  
М. Стоунбрекер,



Сетевая СУБД IDMS,  
Cullinane

Язык запросов  
QBE, М. Злуф



РСУБД ORACLE –  
1-ая коммерческая,  
Л. Эллисон



Сетевая СУБД TOTAL, связанные списки, Cincom,  
СУБД ADABAS, инвертированные списки, Software AG

1970-е

1970

1973

1975

1979

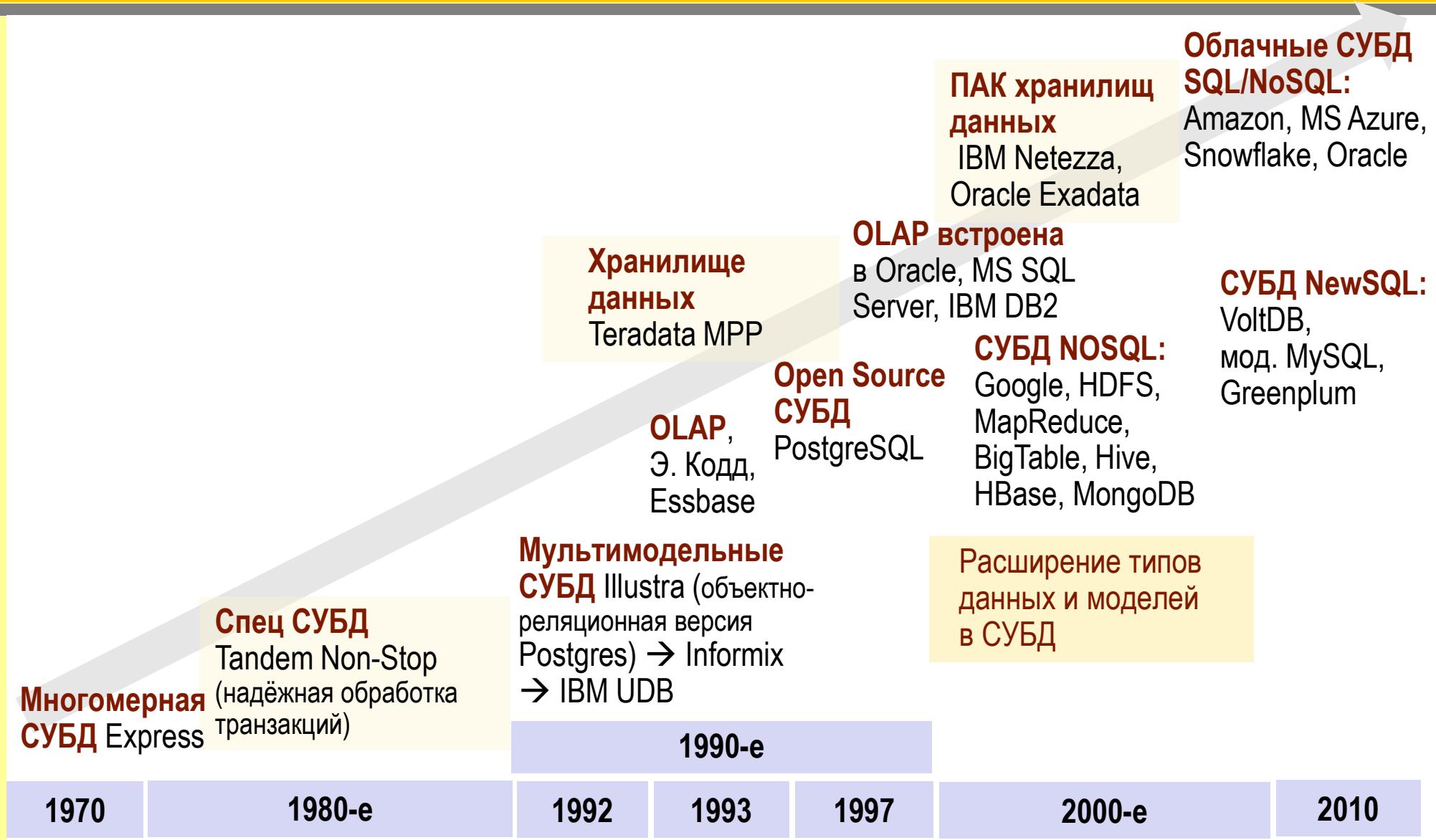
1980

1992

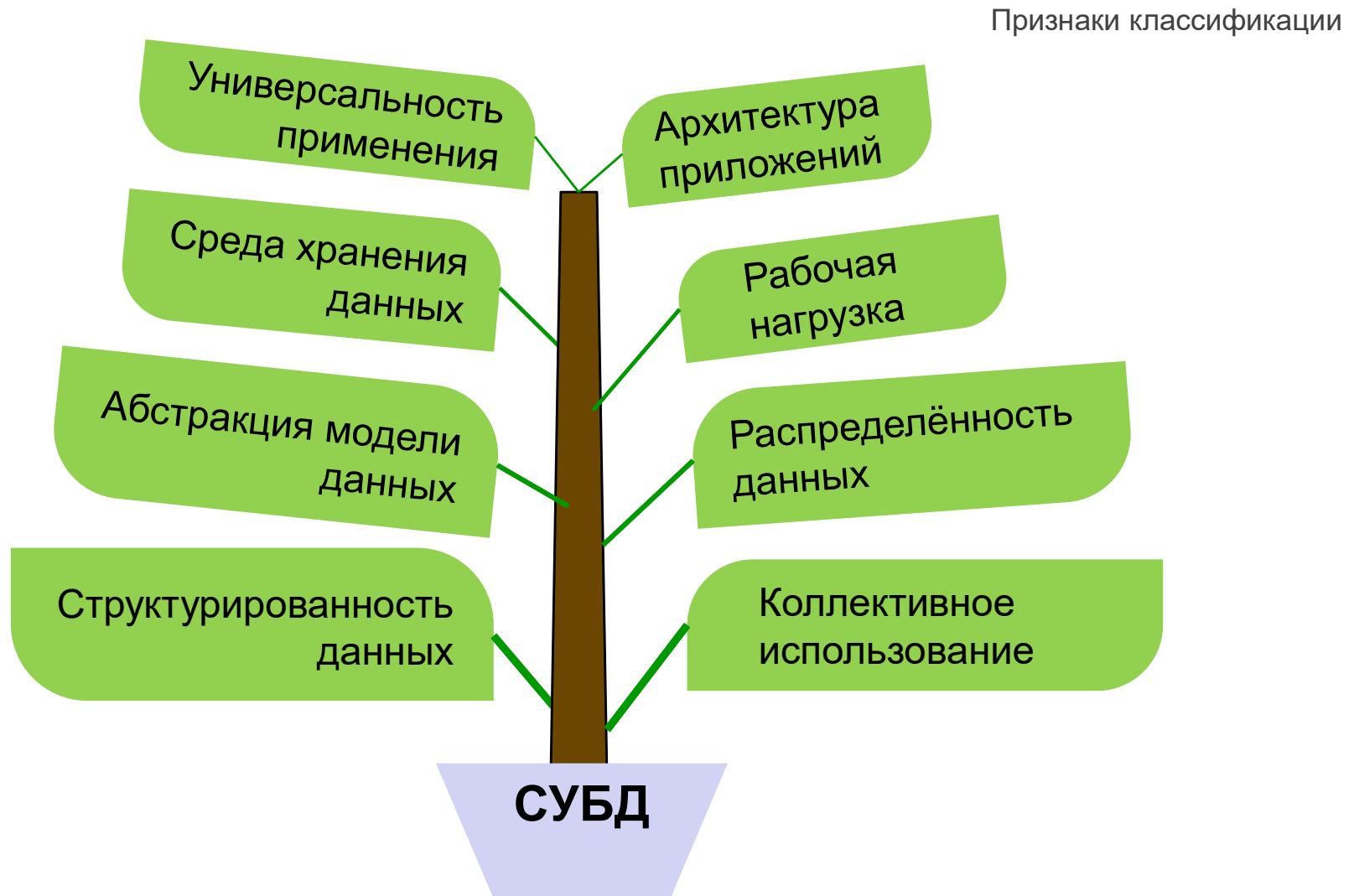
Персональная  
СУБД MS Access

Персональная  
СУБД dBase

# Развитие СУБД



# Классификация СУБД



# Данные с разной степенью структурированности

	Структурированные данные	Полуструктурированные данные	Неструктурированные данные
Форматы	Таблицы, строки, столбцы, ключи	XML, JSON, CSV, log	Текст, картинки, аудио, видео, гипертекст, геокарты
Абстракция модели данных	Реляционная, многомерная, иерархия, сетевая	Иерархия, граф, плоская таблица	Нет модели данных
Схема данных	Хорошо определённая фиксированная схема	Слабая схема данных	Нет схемы
Хранение данных	Реляционные СУБД, классические хранилища данных (data warehouse)	Нереляционные СУБД (NOSQL)	Файловые системы, озёра данных (data lake), облачные ХД
Методы обработки	SQL-запросы, QBE, BI, data mining, AI	Языки запросов, data mining, AI	NLP, анализ текста, LLM, распознавание образов, анализ аудио и видео

# Абстракции модели данных

Иерархические БД	IBM IMS
Сетевые БД	Cullinane IDMS
Реляционные БД ( <b>SQL</b> и <b>NewSQL</b> , построчные и колоночные)	Oracle, MS SQL Server, PostgreSQL, MySQL, Clickhouse, MS Access
Темпоральные (изменения во времени)	MS SQL Server 2016+, PostgreSQL расш.
Объектные и объектно-реляционные БД	Db4o, ObjectStore, ObjectDB

## NoSQL

БД «ключ-значение»

Документные БД

Графовые БД

БД с семейством колонок

Поисковые (полнотекстовые) БД

Redis, Riak KV, OrientDB, Tarantool

MongoDB, YDB, Енисей

Neo4J, Virtuoso, Memgraph

Cassandra, HBase, Azure Cosmos DB

Elasticsearch, Solr, Splunk, Sphinx

Oracle, MS SQL, PostgreSQL+PostGIS,  
Hiperion, Oracle OLAP, MS Analytical Services

Sedna, BaseX, MarkLogic

InfluxDB, Prometheus, Graphite, Riak TS

Kdb, Pinecone, Chroma, Milvus

Jackrabbit, ModeShape

Пространственные БД

Многомерные БД (OLAP, BI)

Native XML

БД временных рядов

Векторные БД

Контентные БД (изображения, аудио, видео,  
гипертекст)

## Мультимодельные БД

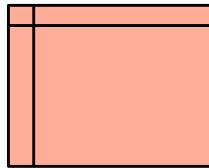
# Возможности СУБД с разными моделями данных

Возможность \ Модель	Реляционная	Графовая	Объектная	Семейство колонок	Документная	Ключ-значение
Поиск по ключу	+	+	+	+	+	+
Извлечение/ обновление части атрибута	+	+	+	+	+	-
Поиск по неключевым атрибутам	+	+	+	+	+	-
Составной ключ	+	+	+	+	-	-
Связи между сущностями	+	+	+	-	-	-
Стандартный интерфейс	+	±	-	-	-	-
Слабоструктурированные данные	+	-	-	+	+	+

# Мультимодельные СУБД

СУБД	Изначальная модель	Дополнительные модели
Oracle	Реляционная	Многомерная, графовая, документная, многомерная, пространственная
MS SQL Server	Реляционная	Графовая, документная, многомерная
PostgreSQL	Реляционная	Графовая, документная
MongoDB	Документная	Ключ-значение, графовая
DataStax	Семейство колонок	Документная, графовая
Redis	Ключ-значение	Документная, графовая
Tarantool	—	Реляционная, ключ-значение, документная, пространственная
OrientDB	—	Графовая, документная
Azure CosmosDB	—	Графовая, документная, реляционная

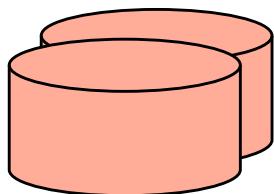
# Среды хранения баз данных



## Основная память

**базы данных in-memory**

Redis, SAP HANA, Oracle Database in-memory,  
Memcached, Apache Ignite

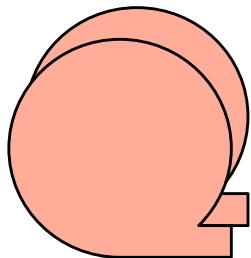


## Вторичная память:

магнитные диски, SSD, СХД –

**базы данных на дисках**

Oracle, MS SQL Server, PostgreSQL,  
MySQL, MS Access



## Третичная память:

магнитные кассеты, оптические диски –

**архивы и резервные копии**

*не являются БД и с ними СУБД не работают,  
но они сопутствуют БД и СУБД*

# Универсальность применения СУБД

Определяется

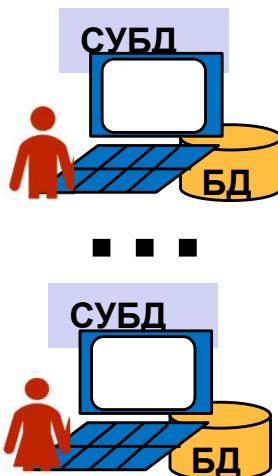
- возможностями
- мультимодельностью
- характером решаемых задач
- требованиями к безопасности и надёжности
- требованиями к производительности

Различают

- универсальные (общего назначения) СУБД
- специализированные СУБД

# Коллективное использование БД

## Персональные однопользовательские базы данных и СУБД

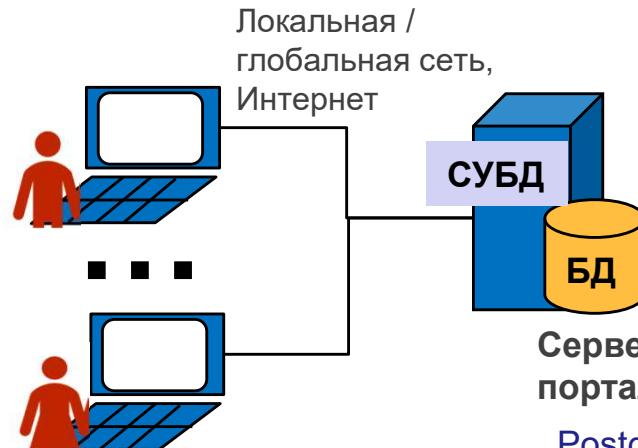


ПК, ноутбуки,  
планшеты,  
терминалы,  
рабочие станции

MS Access  
FileMaker  
Libre Office Base

Авторизация и права  
доступа к файлу БД,  
блокировка файла при  
изменениях

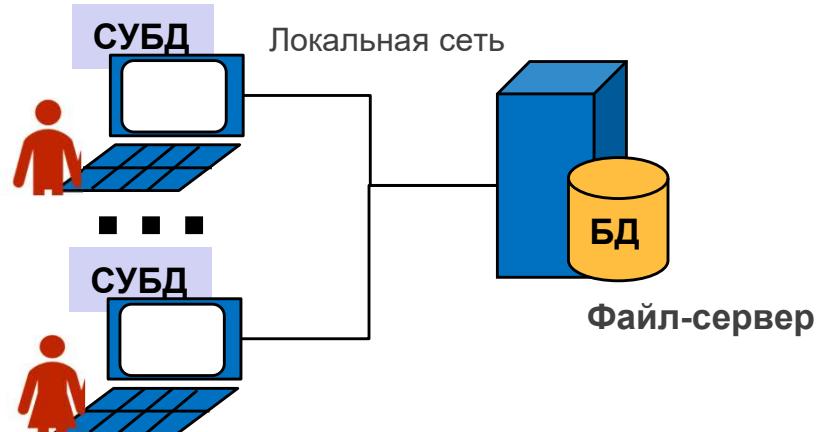
## Коллективные многопользовательские базы данных и СУБД



Локальная /  
глобальная сеть,  
Интернет

Сервер БД,  
портал, сайт

PostgreSQL  
MySQL  
MS SQL Server  
Oracle



Авторизация и права  
доступа к БД, таблицам,  
строкам и столбцам.  
Целостность транзакций.  
Изоляция обновлений.

# Распределённые данные и СУБД

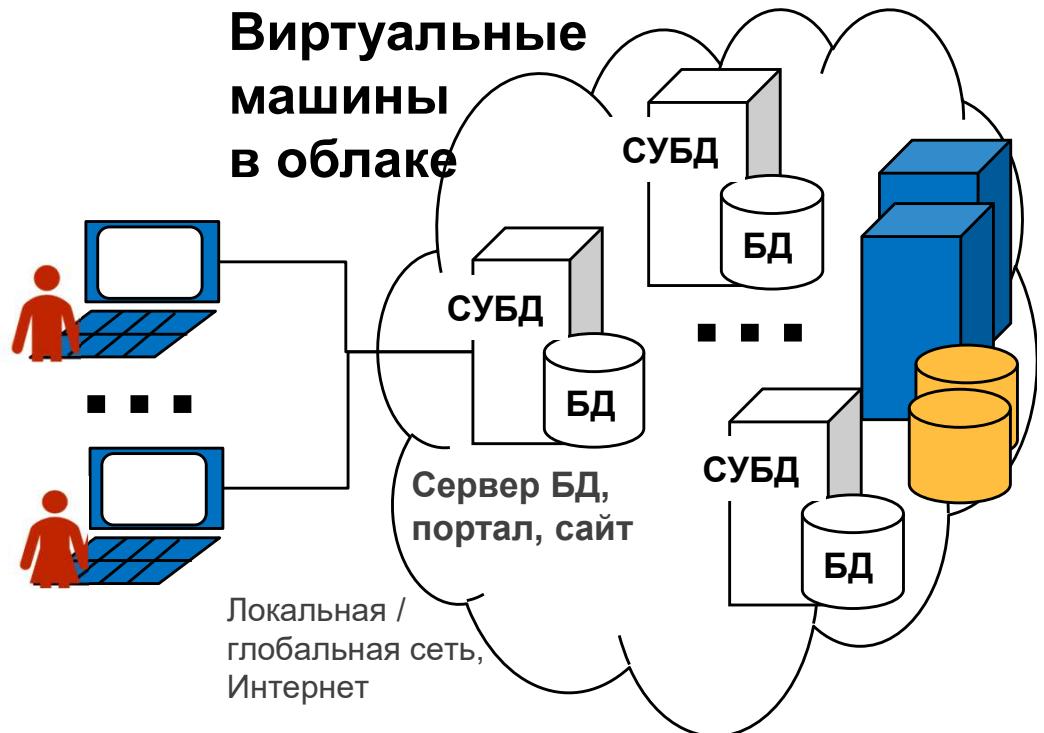
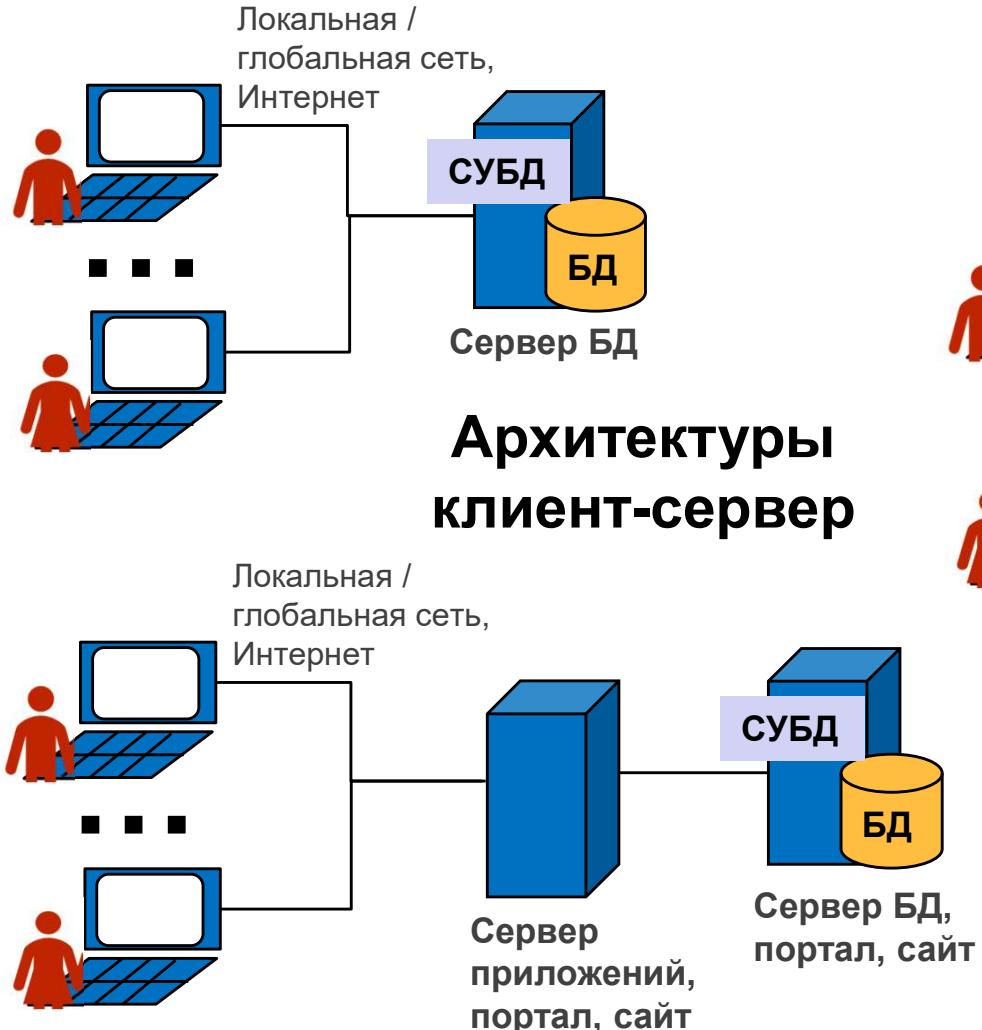
## Сосредоточенные БД

- Локальные
- Централизованные

## Распределённые БД

- Однородные
- Неоднородные
- Фрагментированные (шардинг)
- Тиражируемые (репликации)

# Архитектура приложений



**Облачные сервисы:**  
**IaaS** – инфраструктура  
**PaaS** – платформа (СУБД)  
**DbaaS** – СУБД + БД

# Рабочая нагрузка СУБД

**Транзакционная нагрузка OLTP**

**Высокопроизводительные НР OLTP**

**Высоконадёжные НА OLTP**

**Хранилища данных DWH**

**Поддержка принятия решений OLAP/DSS**

**Системы искусственного интеллекта AI / ML**

**Мультимедийная нагрузка**

# Рейтинг современных СУБД

## DB-Engines Ranking

			429 systems in ranking, January 2026							
Rank			DBMS	Database Model	Score			Jan 2026	Dec 2025	Jan 2025
Jan 2026	Dec 2025	Jan 2025			Jan 2026	Dec 2025	Jan 2025			
1.	1.	1.	Oracle	Relational, Multi-model 	1237.34	+2.94	-21.42			
2.	2.	2.	MySQL	Relational, Multi-model 	867.52	-0.97	-130.63			
3.	3.	3.	Microsoft SQL Server	Relational, Multi-model 	706.26	-16.26	-92.30			
4.	4.	4.	PostgreSQL	Relational, Multi-model 	666.27	+6.84	+2.86			
5.	5.	5.	MongoDB	Multi-model 	376.74	+4.46	-25.77			
6.	6.	6.	Snowflake 	Relational	207.79	+5.34	+53.89			
7.	7.	7.	Redis	Key-value, Multi-model 	144.16	+1.68	-9.20			
8.	8.	↑ 13.	Databricks 	Multi-model 	141.55	+3.53	+53.70			
9.	9.	9.	IBM Db2	Relational, Multi-model 	112.72	-3.04	-10.25			
10.	10.	↓ 8.	Elasticsearch	Multi-model 	107.15	-2.68	-27.78			

# 12 Open Source СУБД в 2025 году

Database Management system	Data model	Requirements	Best for
<b>MySQL</b>	SQL	High read/write throughput	Web apps, CMS, eCommerce
<b>PostgreSQL</b>	SQL	Complex queries, advanced indexing	GIS, financial services, analytics
<b>MariaDB</b>	SQL	Transactional + analytical workloads	MySQL replacement
<b>MongoDB</b>	NoSQL	Real-time queries, flexible schema	IoT, content management
<b>SQLite</b>	SQL	Lightweight, embedded use	Mobile apps, desktop tools
<b>CockroachDB</b>	SQL	Strong consistency	Multi-region SaaS
<b>Redis</b>	NoSQL	Sub-millisecond latency	Caching, real-time analytics
<b>CouchDB</b>	NoSQL	Offline sync, RESTful API	IoT, distributed mobile apps
<b>Neo4j</b>	NoSQL	Relationship traversal, deep joins	Social networks, fraud detection
<b>FirebirdSQL</b>	SQL	OLTP + OLAP concurrency	Embedded systems, SMB apps
<b>OrientDB</b>	SQL	Fast traversal, flexible schema	Identity management, recommendation engines
<b>Cassandra</b>	NoSQL	High write throughput, linear scalability	Time-series, event

**Терпения и удачи всем, кто  
связан с базами данных**

# **Спасибо за внимание!**

**Валерий Иванович Артемьев**

**МГТУ имени Н.Э. Баумана, кафедра ИУ-5**

**Банк России**

**Департамент *данных*, проектов и процессов**

**Тел.: +7(495) 753-96-25  
e-mail: viart@bmstu.ru**