

# **ВВЕДЕНИЕ В БАЗЫ ДАННЫХ**

*Лекция № 9.*  
**БАЗЫ ДАННЫХ И СУБД**

# БАЗЫ ДАННЫХ И СУБД

---

С самого начала развития вычислительной техники (далее – ВТ) образовались два основных направления ее использования:

- применение ВТ для сложных численных расчетов;
- применение ВТ для хранения и обработки информации.

**Первоначально** ВТ использовалась только для сложных численных расчетов. Именно потребностью обеспечить большие объемы расчетов, требовавшихся в новых отраслях науки и техники (ядерной энергетике, ракетостроении и др.), и было вызвано появление ЭВМ.

Это направление дало толчок развитию численных методов и языков программирования, ориентированных на удобную запись численных алгоритмов.

# БАЗЫ ДАННЫХ И СУБД

---

**Второе направление** возникло, когда в компьютерах появилось достаточно оперативной и внешней памяти (магнитные ленты, магнитные барабаны и, главное, магнитные диски). Оно дало развитие информационным системам (далее – ИС).

Классическими примерами ИС являются банковские системы, системы резервирования билетов, библиотечные системы и т.п.

Появление магнитных дисков, способных хранить большие объемы информации, потребовало создания систем управления данными во внешней памяти. Так появились *системы управления файлами* (далее – **СУФ**), или *файловые системы*.

С точки зрения прикладной программы **файл** – это именованная область внешней памяти, в которую можно записывать и из которой можно считывать данные.

# БАЗЫ ДАННЫХ И СУБД

---

Правила именования файлов, способ доступа к данным в файле и их структура зависят от конкретной системы управления файлами. **СУФ** берет на себя *распределение внешней памяти, отображение имен файлов в соответствующие адреса во внешней памяти и обеспечение доступа к данным.*

Файловая система обычно имеет древовидную структуру – каталоги располагаются в узлах дерева, а сами файлы являются его листьями.

Обмен данными между программой и файлами идет через *буфер оперативной памяти*, размер которого определяется либо системой, либо программистом.

# БАЗЫ ДАННЫХ И СУБД

---

Почему файловых систем недостаточно для эффективной обработки информации?

В файловой системе хранятся следующие виды файлов:

- **текстовые** (документы, программы); обрабатываются текстовыми редакторами; их структура очень проста – последовательность строк или байтов (символов);
- **объектные и исполняемые** файлы, которые также имеют простую структуру.

Таким образом, файловые системы обеспечивают хранение слабоструктурированной информации, оставляя дальнейшую структуризацию прикладным программам.

Сами файлы часто изменяются, удаляются, создаются заново.

# БАЗЫ ДАННЫХ И СУБД

---

*Информационные системы*, главным образом, ориентированы на хранение, выбор и модификацию *постоянно хранимой* информации.

Эта информация обычно имеет довольно *сложную структуру*.

С другой стороны, во всех ИС, как бы они ни отличались одна от другой, решаются одни и те же проблемы, связанные с организацией и хранением данных.

Стремление выделить и обобщить общую часть ИС, ответственную за управление сложно структурированными данными, привело к созданию **СУБД** – *систем управления базами данных*.

# БАЗЫ ДАННЫХ И СУБД

---

Система управления данными является **СУБД**, если она

- поддерживает логически согласованный набор файлов;
- предоставляет язык манипулирования данными;
- обеспечивает восстановление информации после сбоев;
- обеспечивает параллельную работу нескольких пользователей.



# Понятие БАЗЫ ДАННЫХ

---

Неформально, **база данных (БД)** – это большое по объему хранилище данных, в которое можно добавлять и из которого можно извлекать данные.

Более строгое определение БД дается Дж. Мартином в книге *«Мартин Дж. Организация баз данных в вычислительных системах. М: Мир, 1978»*.

**База данных** – это совокупность взаимосвязанных хранящихся вместе данных при наличии такой минимальной избыточности, которая допускает их эффективное использование для одного или нескольких приложений (задач);  
данные запоминаются так, чтобы они были независимы от программ, использующих эти данные;  
для добавления новых и модификации существующих данных, а также для поиска данных применяется общий управляемый способ.

# Понятие БАЗЫ ДАННЫХ

---

Базу данных иногда определяют как *неизбыточную совокупность элементов данных*.

Однако на практике во многих базах данных избыточность присутствует.

В этом случае говорят об *управляемой* или *минимальной избыточности*. Она нужна для

- уменьшения времени доступа к данным или упрощения способов адресации;
- обеспечения возможности восстановления данных при их случайной потере.

# Понятие БАЗЫ ДАННЫХ

---

Под *независимостью данных от программ* и наоборот, программ от данных, подразумевается, что изменение одних не приводит к изменению других.

В действительности можно говорить только об *относительной независимости данных*, т.е. небольшие изменения одних не приводят к изменению других.

Важным моментом в БД является понятие *целостности данных*. Если БД содержит данные, используемые различными пользователями, очень важно, чтобы элементы данных и связи между ними не разрушались.

Вычислительная система должна гарантировать *целостность* хранимых в ней данных, т.е. *связей между данными* и *корректность* самих данных.

# Понятие БАЗЫ ДАННЫХ

---

Есть два способа поддержания целостности данных:

- 1) *аппаратный* или **системный**, т.е. защита от разрушения при сбоях и некорректном доступе;
- 2) *программный*, т.е. задание ограничений целостности.

Второй способ (программный) предполагает проверку соответствия значений данных некоторым критериям.

Например: находятся ли значения данных в определенном диапазоне (возраст, год рождения, зарплата) или удовлетворяют ли значения некоторых элементов данных какому-либо отношению, уравнению, функции, другим ограничениям (фамилии, например, должны состоять только из букв и начинаться с большой буквы).

# Основные функции СУБД

---

**Основными функциями СУБД являются:**

- 1) управление данными во внешней памяти;
- 2) управление буферами оперативной памяти;
- 3) управление транзакциями;
- 4) журнализация и восстановление базы данных после сбоев;
- 5) поддержка языков баз данных.

# Основные функции СУБД

---

## 1. Управление данными во внешней памяти.

Заключается в обеспечении необходимых структур внешней памяти как для хранения данных, непосредственно входящих в базу данных, так и для служебных целей.

(Как это делается, пользователь не должен знать, так как все заботы берет на себя СУБД.)

## 2. Управление буферами оперативной памяти.

Обмен данными между центральным процессором и внешними устройствами остается все еще довольно медленным.

Для увеличения скорости этого обмена используется *буферизация данных в оперативной памяти* (далее – ОП).

Обычно той буферизации, которую дает операционная система, недостаточно и СУБД имеют и поддерживают собственный набор буферов ОП с собственной дисциплиной их замены.

# Основные функции СУБД

---

## 3. Управление транзакциями.

**Транзакция** – это последовательность операций над БД, рассматриваемая СУБД как единое целое.

Понятие транзакции необходимо для *поддержания логической целостности базы данных*. Каждая транзакция начинается при целостном состоянии БД и оставляет это состояние целостным после своего завершения.

Если транзакция успешно выполняется, то новое состояние БД фиксируется во внешней памяти, иначе ни одно из изменений никак не отразится в состоянии БД.

Транзакция может рассматриваться как *минимальная единица активности пользователя* по отношению к СУБД. Это понятие очень важно в многопользовательских СУБД, когда с БД параллельно работает несколько человек. В этом случае рассматривают серии или смеси транзакций.

# Основные функции СУБД

---

## 4. Журнализация и восстановление базы данных после сбоев.

Одним из основных требований к СУБД является надежное хранение данных во внешней памяти.

**Надежность СУБД** – это способность восстановить последнее целостное состояние базы данных после любого *аппаратного или программного сбоя*.

**Аппаратные сбои** бывают *мягкие* (внезапное выключение или остановка компьютера) и *жесткие* (потеря информации на носителях внешней памяти).

**Программный сбой** – это аварийное завершение работы СУБД или прикладной программы. Программный сбой эквивалентен мягкому аппаратному сбою.



# Основные функции СУБД

---

Для восстановления БД нужно хранить избыточную информацию, причем та информация, которая используется для восстановления, должна храниться особенно надежно. Наиболее распространенным способом поддержания такой избыточности информации является ведение *журнала изменений БД*.

*Журнал* – это особая часть базы данных, не доступная пользователям СУБД и поддерживаемая с особой тщательностью, в которую поступают записи обо всех изменениях основной части БД.

Кроме того, в «серьезных» СУБД ведется *полная архивная копия базы данных*.

*Журнал и архивная копия* используются для восстановления БД после *жесткого сбоя*.

При *мягких сбоях* достаточно одного *журнала изменений БД*.

# Основные функции СУБД

---

## 5. Поддержка языков баз данных.

Для работы с БД используются специальные языки – *языки баз данных*.

В ранних СУБД выделялось два разных по функциональному назначению языка: *SDL* – язык определения схемы БД (*Schema Definition Language*) и *DML* – язык манипулирования данными (*Data Manipulation Language*).

*SDL* служил для определения логической структуры БД, т.е. той структуры, которая представляется пользователю.

*DML* содержал набор операторов манипулирования данными – добавление, удаление, модификацию и выборку данных.

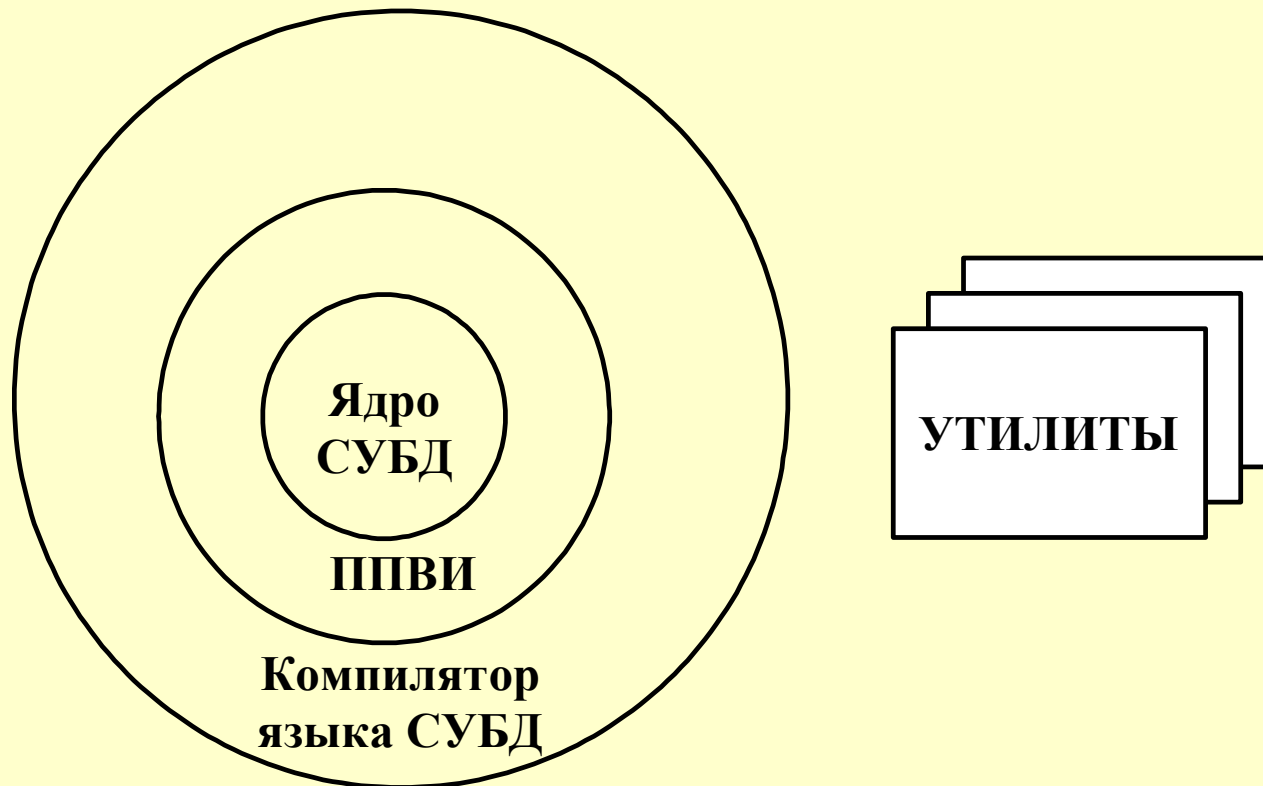
В современных СУБД поддерживается единый интегрированный язык, включающий все необходимые средства для работы с базой данных, начиная от ее создания до реализации пользовательского интерфейса. (Это язык SQL (*Structured Query Language*)).

# Типовая организация СУБД

---

Современная СУБД включает следующие подсистемы:

- ядро СУБД;
- компилятор языка БД;
- подсистему поддержки времени исполнения;
- набор утилит.



# Типовая организация СУБД

---

**Ядро** отвечает за управление данными во внешней памяти, управление буферами оперативной памяти, управление транзакциями и журнализацию БД.

**Компилятор языка БД** транслирует операторы языка баз данных в некоторую исполняемую программу в машинном или промежуточном машинно-независимом коде. (Современные компиляторы также выполняют оптимизацию запросов к БД.)

**Подсистема поддержки времени исполнения (ППВИ)** обеспечивает реальное выполнение программы, являясь фактически интерпретатором языка БД.

**Утилиты БД** – это служебные программы, выполняющие такие процедуры, которые слишком дорого выполнять с использованием языка БД. Например: загрузка и выгрузка БД, глобальная проверка целостности, сбор статистики и т.д.