

Лекция 1. Часть 2.

**Базы данных: основные
понятия**

Базы данных: основные понятия

1. История возникновения баз данных. Файловая система как их прообраз.
2. Типы задач, для решения которых необходимы базы данных.

Классификация задач, решаемых на компьютере

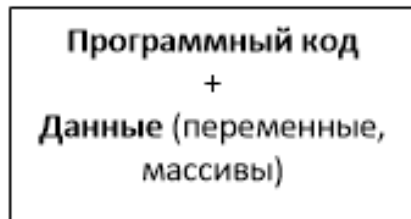
Тип	Представление данных	Алгоритм вычислений
Вычислительные задачи	Простое	Сложный
Традиционные задачи обработки данных (невычислительные)	Сложное	Простой
Современные задачи обработки данных	Сложное	Сложный

Эволюция представления данных: от файловой системы к базам данных

1950-60 годы: данные отделяются от программ и обособляются в файлы.

Первые коммерческие программы – для ведения бухгалтерии (file folder = папка для бумаг).

Вычислительные задачи



Задачи обработки данных



Концепции баз данных – это результат долгого развития файловых систем.

Логическая и физическая структура файлов

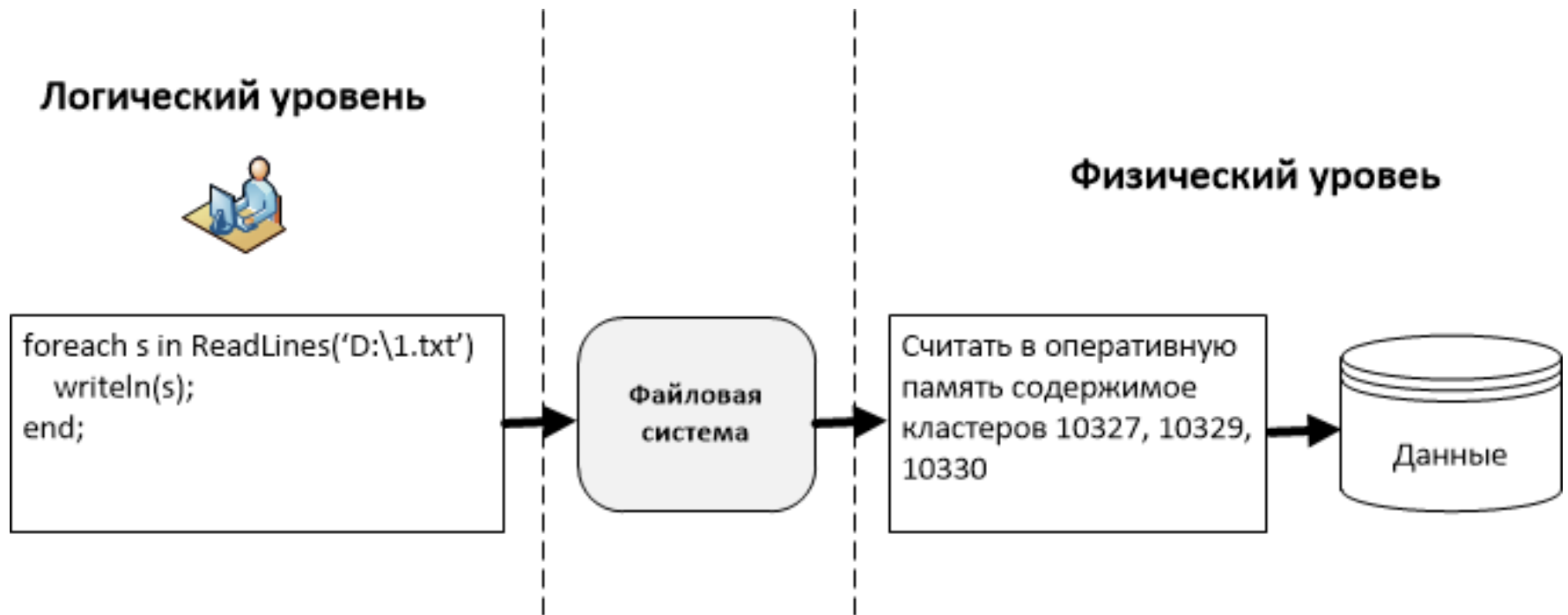
Для прикладной программы файл – это **именованная область внешней памяти**, в которую можно записывать и из которой можно считывать данные.



Файл на внешнем носителе – это **цепочка кластеров** (физических записей).

Файловая система – для абстрагирования от данных

Пользователь/программист имеет дело только с логическими данными (более удобная форма), он не касается деталей фактического низкоуровневого размещения данных.



Файловая система – часть операционной системы, включающая:

1. Совокупность всех файлов на диске с их **физической организацией**.

Файловая система – часть операционной системы, включающая:

1. Совокупность всех файлов на диске с их **физической организацией**.
2. Структуры данных управления файлами (каталоги файлов, дескрипторы файлов, таблицы распределения файлов, ...), т. е. **логическая организация файловых структур**.

Файловая система – часть операционной системы, включающая:

1. Совокупность всех файлов на диске с их **физической организацией**.
2. Структуры данных управления файлами (каталоги файлов, дескрипторы файлов, таблицы распределения файлов, ...), т. е. **логическая организация файловых структур**.
3. Комплекс системных **программных средств**, реализующих управление файлами (создание, уничтожение, чтение, запись, поиск и другие операции над файлами).

C:\Документы\Мой файл.doc

Каталог файлов

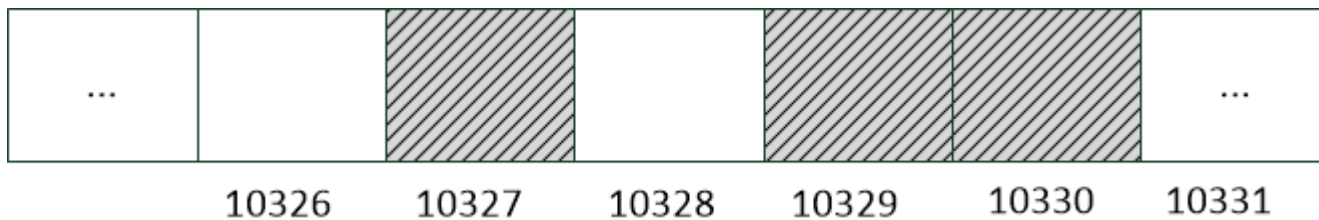
Дескриптор файла 1
...
Дескриптор файла N

Дескриптор файла

Имя	<i>Мой файл.doc</i>
Дата создания	<i>01.02.2017</i>
Атрибуты	<i>Archive</i>
Первый кластер	<i>10327</i>
Размер	<i>9326</i>
...	...

Таблица размещения файлов

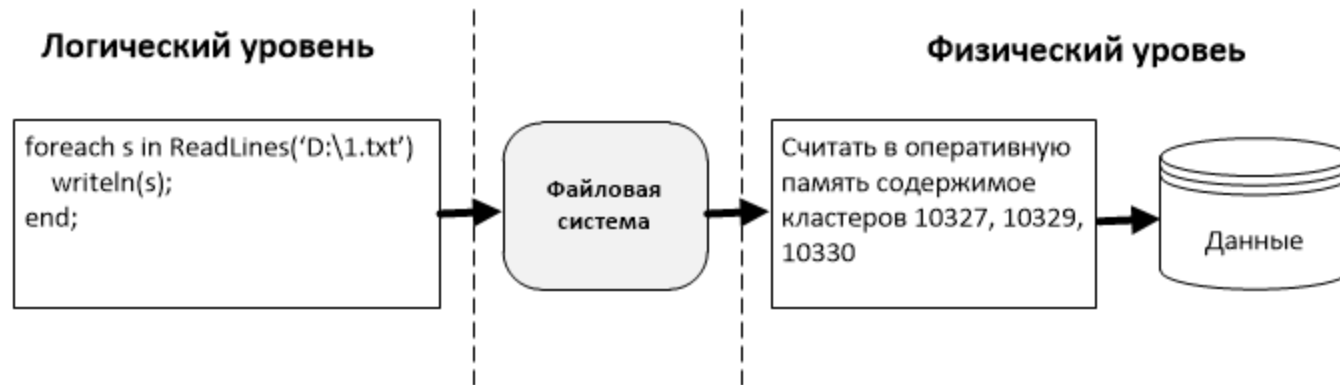
№ кластера	Статус
10326	<i>Сбойный</i>
10327	<i>10329</i>
10328	<i>Свободный</i>
10329	<i>10330</i>
10330	<i>Конец цепочки</i>
10331	<i>Свободный</i>
...	...



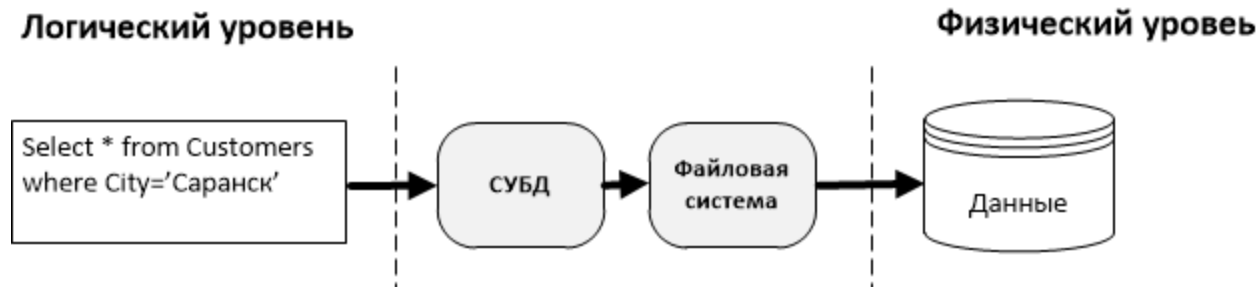
Кластеры на диске

Базы данных – следующий шаг в абстрагировании пользователя/программиста от данных.

Работа с файлом



Работа с базой данных



Структурированные данные

Языки программирования позволяют определять составные типы данных (записи, структуры), создавать переменные таких типов и сохранять их в файле.

```
Type Person = Record
  Id: LongInt;
  Lastname: string[30];
  Firstname: string[30];
  Surname: string[30];
  Birthday: string[10];
end;
var f: file of Person;
```

Файл содержит логические записи, состоящие из полей.

Id	Lastname	Firstname	Surname	Birthday
1376	Сидоров	Александр	Григорьевич	13.05.1958
12	Петров	Сергей	Валерьевич	21.09.1978
...

Плоские структурированные файлы

Два способа хранения логических записей в файле:

- Список полей с символами-разделителями.
- Список из полей фиксированной длины, равной ширине соответствующих столбцов таблицы.

1376;Сидоров;Александр;Григорьевич;13.05.1958					12;Петров;Сергей;Ва	
1376	Сидоров	Александр	Григорьевич	13.05.1958		
12	Петров	Сергей	Валерьевич	21.09.1978	427	

Плоские файлы – первый прообраз баз данных.

- В файле содержатся только данные, информации о записях нет.
- Структура записей задается в прикладной программе, работающей с этими данными.

Файлы с метаданными

Файл состоит из:

- заголовка, где хранится информация о структуре записей (имена и размерность полей) и их количестве;
- области данных из записей фиксированной длины.

<i>Записи из пяти полей. 1-е поле: Id, целое, длина 10 байт. 2-е поле: Lastname, символьное, длина 30 байт. ...</i>	<i>Запись 1</i>	<i>Запись 2</i>	<i>...</i>
Заголовок	Данные		

Преимущество: структура данных описывается в самих файлах с данными, а не в прикладных программах.

Файлы DBF – стандартный формат для ранних баз данных на персональных компьютерах.

Программы для обработки данных в файлах

Задача 1. Кадровый учет. Файл:

Фамилия	Имя	Отчество	Дата рожд-я	Место жит-ва	Должность	Оклад
Иванов	Иван	Иванович	01.02.1985	г. Саранск	Инженер	30000

Программы для обработки данных в файлах

Задача 1. Кадровый учет. Файл:

Фамилия	Имя	Отчество	Дата рожд-я	Место жит-ва	Должность	Оклад
Иванов	Иван	Иванович	01.02.1985	г. Саранск	Инженер	30000

Задача 2. Начисление заработной платы. Файл:

Фамилия	Имя	Отчество	Оклад	Месяц	Отработано, дней	Сумма
Иванов	Иван	Иванович	30000	Февраль	10	15000

Программы для обработки данных в файлах

Задача 1. Кадровый учет. Файл:

Фамилия	Имя	Отчество	Дата рожд-я	Место жит-ва	Должность	Оклад
Иванов	Иван	Иванович	01.02.1985	г. Саранск	Инженер	30000

Задача 2. Начисление заработной платы. Файл:

Фамилия	Имя	Отчество	Оклад	Месяц	Отработано, дней	Сумма
Иванов	Иван	Иванович	30000	Февраль	10	15000

Задача 3. Учет больничных. Файл:

Фамилия	Имя	Отчество	Оклад	Месяц	На больн-м, дней	Сумма
Иванов	Иван	Иванович	30000	Февраль	6	8500

Программы для обработки данных в файлах

Задача 1. Кадровый учет. Файл:

Фамилия	Имя	Отчество	Дата рожд-я	Место жит-ва	Должность	Оклад
Иванов	Иван	Иванович	01.02.1985	г. Саранск	Инженер	30000

Задача 2. Начисление заработной платы. Файл:

Фамилия	Имя	Отчество	Оклад	Месяц	Отработано, дней	Сумма
Иванов	Иван	Иванович	30000	Февраль	10	15000

Задача 3. Учет больничных. Файл:

Фамилия	Имя	Отчество	Оклад	Месяц	На больн-м, дней	Сумма
Иванов	Иван	Иванович	30000	Февраль	6	8500

Информация дублируется – это плохо!

Программы для обработки данных в файлах

Задача 1. Кадровый учет. Файл:

Фамилия	Имя	Отчество	Дата рожд-я	Место жит-ва	Должность	Оклад
Иванов	Иван	Иванович	01.02.1985	г. Саранск	Инженер	30000

Задача 2. Начисление заработной платы. Файл:

Фамилия	Имя	Отчество	Оклад	Месяц	Отработано, дней	Сумма
Иванов	Иван	Иванович	30000	Февраль	10	15000

Задача 3. Учет больничных. Файл:

Фамилия	Имя	Отчество	Оклад	Месяц	На больн-м, дней	Сумма
Иванов	Иван	Иванович	30000	Февраль	6	8500

Информация **дублируется** – данные могут стать противоречивыми.

Данные нужно **интегрировать**, хранить в одном месте.

Общая информационная база

Вариант 1. Объединить все в одном файле:

ФИО	Дата рожд-я	Место жит-ва	Должность	Оклад	Месяц	Отработано, дней	На больн-м, дней	Зарплата	Больнич.
-----	-------------	--------------	-----------	-------	-------	------------------	------------------	----------	----------

Общая информационная база

Вариант 1. Объединить все в одном файле:

ФИО	Дата рожд-я	Место жит-ва	Должность	Оклад	Месяц	Отработано, дней	На больн-м, дней	Зарплата	Больнич.
-----	-------------	--------------	-----------	-------	-------	------------------	------------------	----------	----------

Недостатки:

- Останется дублирование данных внутри файла.
- Сильно возрастет время решения задачи 3.

Общая информационная база

Вариант 1. Объединить все в одном файле:

ФИО	Дата рожд-я	Место жит-ва	Должность	Оклад	Месяц	Отработано, дней	На больн-м, дней	Зарплата	Больнич.
-----	-------------	--------------	-----------	-------	-------	------------------	------------------	----------	----------

Недостатки:

- Останется дублирование данных внутри файла.
- Сильно возрастет время решения задачи 3.

Вариант 2. Два файла:

Фамилия	Имя	Отчество	Дата рожд-я	Место жит-ва	Должность	Оклад
---------	-----	----------	-------------	--------------	-----------	-------

ФИО	Оклад	Месяц	Отработано, дней	На больн-м, дней	Зарплата	Больничный
-----	-------	-------	------------------	------------------	----------	------------

Общая информационная база

Вариант 1. Объединить все в одном файле:

ФИО	Дата рожд-я	Место жит-ва	Должность	Оклад	Месяц	Отработано, дней	На больн-м, дней	Зарплата	Больнич.
-----	-------------	--------------	-----------	-------	-------	------------------	------------------	----------	----------

Недостатки:

- Останется дублирование данных внутри файла.
- Сильно возрастет время решения задачи 3.

Вариант 2. Два файла:

Фамилия	Имя	Отчество	Дата рожд-я	Место жит-ва	Должность	Оклад
---------	-----	----------	-------------	--------------	-----------	-------

ФИО	Оклад	Месяц	Отработано, дней	На больн-м, дней	Зарплата	Больничный
-----	-------	-------	------------------	------------------	----------	------------

Вариант 3. Два файла:

Табельный номер	Фамилия	Имя	Отчество	Дата рождения	Место жит-ва	Должность	Оклад
-----------------	---------	-----	----------	---------------	--------------	-----------	-------

Табельный номер	Оклад	Месяц	Отработано, дней	На больн-м, дней	Зарплата	Больничный
-----------------	-------	-------	------------------	------------------	----------	------------

База данных как новый вид данных

База данных – совокупность взаимосвязанных хранящихся вместе данных при наличии такой минимальной избыточности, которая допускает их использование оптимальным образом для одного или нескольких приложений.

База данных как новый вид данных

База данных – совокупность взаимосвязанных хранящихся вместе данных при наличии такой минимальной избыточности, которая допускает их использование оптимальным образом для одного или нескольких приложений.

Базы данных нужны при использовании общих данных несколькими задачами (программами).

Основной критерий оптимальности функционирования базы данных – время выполнения запросов пользователей к данным.

Резюме

- Базы данных — результат развития файловых систем.
- Информация о структуре данных должны храниться в базе данных, а не в приложении.
- Дублирование информации в системе – **зло**, т. к. можно легко нарушить логическую целостность (непротиворечивость) данных.
- Информация в базе данных разделяется на связанные друг с другом части. Сделать это можно по-разному. Нужно минимизировать дублирование данных.