

Принципы построения файловой системы

Основные понятия

Файл – это именованный объект, который может хранить данные, программу или другую информацию.

Файловую систему можно определить как состоящую из двух составных частей:

1. Совокупность файлов и управляющей информации на диске для доступа к файлам.

2. Совокупность программных средств ОС для доступа к файлам, которые выполняют следующие операции с файлами:

создание файлов;

уничтожение файлов;

поиск файлов на диске;

чтение-запись информации из файла (в файл);

защита файлов от несанкционированного доступа;

открытие файлов;

заккрытие файлов.

виды файлов:

регулярные – обычные файлы на диске, которые состоят из блоков фиксированной длины – секторов (в MS-DOS – 512 б). При работе с регулярными файлами ОС обращается к специальной программе – драйверу блочного устройства.

специальные символьные файлы – логическое представление драйвера символьного устройства;
директории – файлы, хранящие специальную информацию о регулярных файлах и директориях.

Драйвер устройства – специальная программа, которая выполняет физическое управление ПУ.

различает два класса драйверов:

- драйверы символьных устройств (за одно обращение переносится один символ);
- драйверы блочных устройств (за одно обращение переносится один блок).

Все драйверы подразделяются на стандартные и устанавливаемые (устанавливаемые).

Символьные устройства, имеющие свои драйверы, трактуются операционной системой как файлы с уникальным именем, доступ к которым осуществляется с помощью специальных системных вызовов. Имя драйвера становится известным системе в ходе начальной загрузки. Конкретному символьному устройству соответствует только одно имя, поэтому драйвер символьного устройства управляет работой только одного устройства.

Зарезервированные имена устройств :

CON - клавиатура и экран консоли;

COM1 (AUX) - 1-й порт адаптера последовательной связи;

COM2 - 2-й порт адаптера последовательной связи;

COM3 - 3-й порт адаптера последовательной связи;

COM4 - 4-й порт адаптера последовательной связи;

LPT1 (PRN) - 1-й порт адаптера параллельной связи

(принтер);

LPT2 - 2-й порт адаптера параллельной связи;

LPT3 - 3-й порт адаптера параллельной связи;

LPT4 - 4-й порт адаптера параллельной связи;

NUL - несуществующее (фиктивное) устройство.

В отличие от драйверов символьных устройств доступ к драйверам блочных устройств осуществляется по букве логического накопителя. Один драйвер блочного устройства может отвечать за работу нескольких логических накопителей, называемых устройствами-единицами (units).

Логическая структура файла

С точки зрения программиста файл представляет собой ленту байтов, имеющую начало и конец. В файле определен указатель чтения-записи – текущая позиция ленты байтов, к которой осуществляется доступ.

При каждом переносе байтов (чтении или записи) указатель автоматически передвигается вперед на перенесенное число байтов. Достижение конца файла определяется ОС, которая сообщает об этом в программу передачей специального условия EOF – End Of File. Условие EOF регистрируется ОС только при чтении информации в тот момент, когда указатель чтения-записи достигает значения, равного размеру файла.

Для текстовых файлов условие EOF регистрируется при чтении специального символа EOF – это ASCII-символ 1Ah (символ Ctrl-Z), после которого, как правило, следует комбинация символов CR-LF (0Dh,0Ah).

Автоматическое приращение указателя записи-чтения позволяет осуществлять доступ к файлу байт за байтом. Это так называемый последовательный доступ (consecutive access). Наличие средств перемотки указателя на необходимую позицию в файле дает возможность осуществлять доступ к заданному месту в файле. Этот доступ называется прямым или произвольным (direct or random access).

Для текстовых файлов условие EOF регистрируется при чтении специального символа EOF – это ASCII-символ 1Ah (символ Ctrl-Z), после которого, как правило, следует комбинация символов CR-LF (0Dh,0Ah).

Автоматическое приращение указателя записи-чтения позволяет осуществлять доступ к файлу байт за байтом. Это так называемый последовательный доступ (consecutive access). Наличие средств перемотки указателя на необходимую позицию в файле дает возможность осуществлять доступ к заданному месту в файле. Этот доступ называется прямым или произвольным (direct or random access).

Директорий и его структура

Файловая система на диске имеет иерархическую древовидную структуру. В узлах дерева размещаются специальные файлы - директории. Самый первый в иерархии директорий называется корневым (root directory). Директорий - это специальный файл на диске, состоящий из так называемых элементов директория

Размер корневого директория ограничен и зависит от формата диска. Файл корневого директория располагается в фиксированном месте диска. Нижестоящие в иерархии директории называются субдиректориями; они могут располагаться в любом месте диска и иметь число элементов, ограниченное только физическим объемом диска.

Назначение отдельных полей элемента директория:

Байты 0-7. Имя файла.

Имя файла может состоять не более чем из 8 ASCII-символов: латинских букв, цифр, символов подчеркивания, +, - и других специальных символов. Если используется менее 8 символов, остальные позиции дополняются пробелами.

Все директории ниже корневого обязательно содержат два зарезервированных имени, помещаемых в файл директория при его создании. Это имена "." и "..". Имя ".." соответствует файлу директория-предка (parent directory), т. е. того, в котором зарегистрирован данный директорий. Имя "." соответствует файлу данного директория, т.е. поле номера кластера такого элемента содержит номер первого кластера файла данного директория.

В случае, когда выполняется уничтожение файла, ОС в качестве первого символа имени такого файла записывает в элементе директория символ ? (ASCII E5h), и кластеры, занятые файлом, могут использоваться для других файлов. Однако ОС "подстраховывает" пользователя на случай ошибочного уничтожения файлов, оставляя, пока это возможно, кластеры стертого файла незанятыми. При необходимости выделить кластер под новый или "растущий" существующий файл, выделяются ранее не использовавшиеся файлами кластеры. (Естественно, что это возможно только до тех пор, пока такие, совершенно чистые кластеры есть на диске.) Эту особенность поведения ОС используют многие программы, выполняющие "восстановление" стертых с диска файлов. Ясно, что восстановление будет невозможным, если кластеры ранее удаленного файла уже заняты другим файлом. Если по каким-либо причинам необходимо иметь файл с именем, начинающимся с символа ?, первый байт рассматриваемого поля будет равен 05h.

Байты 8-10. Расширение имени файла.

Расширение имени может содержать не более 3 ASCII-символов: латинских букв, цифр, символов подчеркивания, '+', '-' и других специальных символов. Если используется менее трех символов, остальные позиции дополняются пробелами. Как и для имени файла, MS-DOS автоматически преобразует символы расширения к верхнему регистру.

Байт 11. Атрибут файла.

Атрибут указывает тип файла. Отдельные его биты при установке в единицу означают следующее:

- бит 0 - указывает на то, что файл является только читаемым (read only), т.е. при попытке его открытия будет фиксироваться ошибка нарушения прав доступа к файлу;
- бит 1 - указывает на скрытый (hidden) файл. Такой файл не будет показываться в листинге директория, получаемом, например, командой DIR;
- бит 2 - указывает на системный (system) файл. Такой файл не будет показываться в листинге директория, получаемом, например, командой DIR;
- бит 3 - указывает на то, что данный элемент директория соответствует метке тома. Первые 11 байт элемента директория содержат саму метку, а остальные поля элемента игнорируются. Метка тома может располагаться только в корневом директории;
- бит 4 - указывает на то, что данный файл - это субдиректорий и поле номера задает первый кластер файла субдиректория на диске;
- бит 5 - архивный бит. Устанавливается в 1 при создании файла и выполнении операции записи. Это позволяет обнаружить все файлы, изменившиеся с момента последнего сброса бита в 0. Сброс битов осуществляет, например, специальная утилита BACKUP, выполняющая резервные копии файлов.

– **Байты 12-21.** Зарезервированы и содержат обычно нули.

Байты 22 - 23. Время создания (последней модификации) файла.

Биты в этих байтах используются следующим образом:

15 14 13 12 11 10 9 8 7 6 5 4 3 2 1 0

ч ч ч ч ч м м м м м м с с с с с,

где ччччч - двоичное представление часов (0 - 23); ммммммм - двоичное представление минут (0 - 59); ссссс - двоичное представление числа секундных единиц. Часто поле времени создания файла используется программами-вирусами для индикации того, что данный файл уже "заражен". Например, если поле ммммммм задает заведомо некорректное значение 62, это, как правило, результат работы вируса.

Байты 24 - 25. Дата создания (последней модификации) файла

Биты в этих байтах используются следующим образом:

15 14 13 12 11 10 9 8 7 6 5 4 3 2 1 0

Г Г Г Г Г Г Г М М М М Д Д Д Д Д,

где гgggg - двоичное представление года (0-119, что соответствует годам с 1980 по 2099); мmmm - двоичное представление месяца (1 - 12) ;ддддд-двоичное представление дня (1-31).

Байты 26 - 27. Номер первого кластера файла.

Байты 28-31. Размер файла в байтах.

Так как данное поле состоит из 4 байтов, максимальный размер файла равен 2^{32} или 4Г байт.

