

Файлы и файловые системы

Иерархия данных

Иерархия данных

Опр. Иерархия данных (data hierarchy) – классификация, группирующая различные последовательности битов для представления осмысленных значений.

Иерархия данных

- Последовательности битов, байты и слова содержат небольшие количества битов, интерпретируемые аппаратными устройствами и низкоуровневыми программами
- Поля, записи и файлы могут содержать множества битов, интерпретируемые операционными системами и пользовательскими приложениями

Сочетания битов

Опр. Сочетания битов (bit patterns) – нижний уровень иерархии данных. Битовые последовательности составляющие двоичные коды, используются для представления всех данных в компьютерных системах. В последовательности из n битов можно хранить 2^n различных сочетаний битов.

Байт

Опр. Байт (byte) – второй снизу уровень иерархии данных. Обычно байт состоит из 8 битов.

Слово

Опр. Слово (word) – последовательность битов, которую может одновременно обрабатывать процессор(ы) системы. В иерархии данных слова располагаются на уровень выше байтов.

Пр. Слово состоит из 4 байтов для 32-разрядного процессора и 8 байтов для 64-разрядного.

СИМВОЛ

Опр. Символ (character) – в иерархии данных – последовательность битов фиксированной длины, обычно – 8, 16 или 32 бита.

Набор символов

Опр. Набор символов (character set) – таблица, содержащая определенное конечное множество символов. К популярным наборам символов относятся ASCII, EBCDIC и Unicode.

ASCII

Опр. ASCII (American Standard Code for Information Interchange, Американский стандартный код обмена информацией) – набор символов, широко применяемый в персональных компьютерах и системах обмена данными, хранящий символы в 8-битовых байтах. В этом наборе может быть до 256 различных символов. Из-за этого ограничения в нем не поддерживаются международные наборы СИМВОЛОВ.

EBCDIC

Опр. EBCDIC (Extended Binary-Coded Decimal Interchange Code, Расширенный двоично-десятичный код обмена информацией) – восьмибитовый набор символов, используемый для представления данных в больших компьютерах, особенно производства IBM.

Unicode

Опр. Unicode – набор символов, поддерживающий международные кодировки и широко использующийся в Интернете и многоязычных приложениях. В Unicode есть 8-, 16- и 32-битные форматы представления СИМВОЛОВ.

Поле

Опр. Поле (field) – в иерархии данных – группа символов (например, имя человека, его адрес или номер телефона).

Запись

Опр. Запись (record) – в иерархии данных – группа полей (например, все поля, хранящие информацию о конкретном покупателе или студенте).

Файл

Опр. Файл (file) – именованный набор данных, который может обрабатываться как единое целое с помощью таких операций, как открытие, закрытие, чтение, запись, удаление, копирование и переименование. Отдельные элементы данных в файле могут подвергаться, например, операциям чтения, записи, обновления, вставки и удаления. Файлы могут состоять из одной или более записей.

Том

Опр. Том (volume) – часть пространства накопителя, в которой может храниться множество файлов.

Вопрос для самопроверки

- Большие наборы символов лучше маленьких? (Да/Нет)

Вопрос для самопроверки

- Большие наборы символов лучше маленьких? (Да/Нет)
- Нет. Большие наборы символов, например, Unicode, позволяют хранить и передавать данные на множестве различных языков. Однако в больших наборах символов для представления одного символа используется большое количество битов, и объем хранения данных возрастает.

Вопрос для самопроверки

- Реализованы ли 64-битовые наборы символов? (Да/Нет)

Вопрос для самопроверки

- Реализованы ли 64-битовые наборы символов? (Да/Нет)
- Нет. На сегодняшний день 64-битовые наборы символов еще не реализованы, поскольку они потребуют значительных объемов для хранения каждого символа, и позволяют представлять намного больше различных символов, чем может потребоваться в обозримом будущем.

Файлы и файловые системы

Файлы

Операции с файлами

- Открытие (open file) – подготовка файла к обращениям
- Закрытие (close file) – блокирование дальнейших обращений к файлу до нового открытия
- Создание (create file) – создание нового файла
- Уничтожение (destroy file) – удаление файла

Операции с файлами

- Копирование (copy file) – копирование содержимого файла в другой файл
- Переименование (rename file) – изменение имени файла
- Отображение (list file) – вывод содержимого файла на экран или печать

Операции с элементами данных, хранящимися в файлах

- Чтение (read data) – копирование данных из файла в память процесса
- Запись (write data) – копирование данных из памяти процесса в файл

Операции с элементами данных, хранящимися в файлах

- Обновление (update data) – изменение содержимого существующего элемента данных в файле
- Вставка (insert data) – добавление в файл нового элемента данных
- Удаление (delete data) – удаление элемента данных из файла

Свойства, характеризующие файлы

- Размер (size of file) – количество данных, хранящихся в файле
- Расположение (location of file) – место, где хранится файл (на носителе или в логической структуре файлов системы)

Свойства, характеризующие файлы

- Режим доступа (accessibility of file) – ограничения на доступ к файлу
- Тип (type of file) – назначение файла (например, исполняемый файл содержит исполняемые инструкции для процесса, а для файла данных может быть указано приложение, предназначенное для работы с его содержимым)

Свойства, характеризующие файлы

- Изменчивость (volatility of file) – частота внесения изменений в данные, хранящиеся в файле
- Активность (activity of file) – процент записей в файле, к которым выполняются обращения в течение заданного периода времени

Физическая запись

Опр. Физическая запись (physical record, физический блок, physical block) – единица данных, считываемая с накопителя, или записываемая на него.

Логическая запись

Опр. Логическая запись (logical record, логический блок, logical block) – набор данных, воспринимаемый программами, как единое целое.

Неблокированные записи

Опр. Неблокированные записи (unblocked records) – записи, в которых одной физической записи соответствует одна логическая запись.

Сблокированные записи

Опр. Сблокированные записи (blocked records) – записи, в которых в одной физической может содержаться несколько логических записей. В файле с фиксированной длиной логических записей все они имеют одинаковую длину. В файле с записями произвольной длины записи могут иметь любую длину вплоть до размера физического блока.

Вопрос для самопроверки

- Может ли физическая запись содержать несколько логических записей файла?
(Да/Нет)

Вопрос для самопроверки

- Может ли физическая запись содержать несколько логических записей файла?
(Да/Нет)
- Да. В случае сблокированных записей в одной физической может содержаться несколько логических записей файла.

Вопрос для самопроверки

- При использовании записей переменной длины накладные расходы возрастают? (Да/Нет)

Вопрос для самопроверки

- При использовании записей переменной длины накладные расходы возрастают? (Да/Нет)
- Да. Потому что система должна определять длину каждой записи. Например, можно помечать конец каждой записи специальным маркером или указывать длину каждой записи в специальном поле в начале записи.

Файлы и файловые системы

Файловые системы

Файловая система

Опр. Файловая система (file system) – часть операционной системы, занимающаяся организацией файлов и обеспечением доступа к ним. Файловые системы обеспечивают как логическую, так и физическую организацию файлов. Они также управляют свободным пространством накопителей, обеспечивают безопасность, поддерживают целостность данных и т.д.

Задачи файловых систем

- Управление файлами
- Управление вспомогательными устройствами хранения
- Обеспечение целостности файлов
- Организация методов доступа к данным

Управление файлами

Опр. Управление файлами (file management) – задача файловой системы, в которую входит обеспечение возможностей хранения файлов, выполнения обращений к ним, совместного использования файлов и безопасности.

Управление вспомогательными устройствами хранения

Опр. Управление вспомогательными устройствами хранения (auxiliary storage management) – задача файловой системы, сводящаяся к выделению пространства под файлы на вторичных устройствах хранения.

Обеспечение целостности файлов

Опр. Обеспечение целостности файлов (file integrity management) – задача файловой системы, в которую входит гарантирование того, что хранимая в файлах информация не будет повреждена. Если целостность файлов гарантируется, то в файлах будет только та информация, которая должна быть.

Вопрос для самопроверки

- Файловые системы работают только с данными на вторичных устройствах хранения? (Да/Нет)

Вопрос для самопроверки

- Файловые системы работают только с данными на вторичных устройствах хранения? (Да/Нет)
- Нет. Файловые системы работают с файлами, представляющими собой именованные наборы данных, которые могут храниться на любом носителе, включая оперативную память.

Вопрос для самопроверки

- Имеют ли сходство файловые системы и системы управления виртуальной памятью? (Да/Нет)

Вопрос для самопроверки

- Имеют ли сходство файловые системы и системы управления виртуальной памятью? (Да/Нет)
- Да. Файловые системы должны управлять выделением пространства на накопителях и контролировать доступ к накопителю. Во многих системах виртуальная память реализована в виде файла обмена.

Файлы и файловые системы

Директории

Директория

Опр. Директория (directory) – файл, хранящий ссылки на другие файлы. В записях в директориях обычно содержатся имена файлов, их типы, размеры и другие данные.

Пример содержимого записи в файле директории

Имя	Символьная строка, содержащая имя файла
Местоположение	Физический блок или логический адрес файла в файловой системе
Размер	Количество байтов, занимаемых файлом
Тип	Описание назначения файла
Время создания	Время создания файла

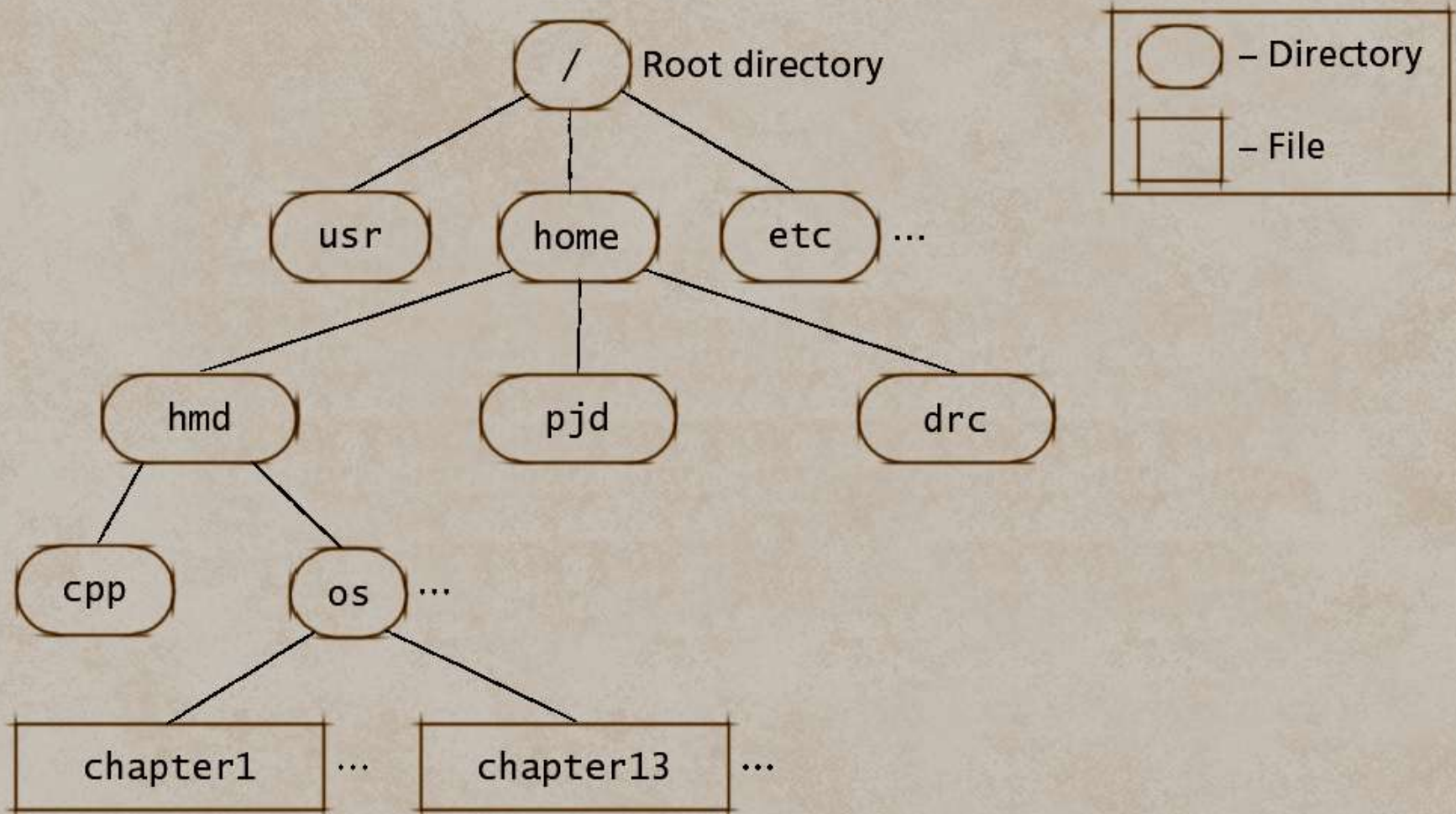
Плоская структура директорий

Опр. Плоская структура директорий (flat directory structure) – одноуровневая файловая система, содержащая только одну директорию.

Иерархически структурированная файловая система

Опр. Иерархически структурированная файловая система (hierarchically structured file system) – файловая система, в которой у каждой директории может быть много дочерних, но только одна родительская директория.

Иерархическая файловая система



Корень

Опр. Корень (root) – начало иерархической структуры файловой системы.

Корневая директория

Опр. Корневая директория (root directory) – директория, содержащая указатели на пользовательские директории.

Пользовательские директории

Опр. Пользовательские директории (user directories) – директории, содержащие записи о пользовательских файлах; каждая запись указывает размещение файла на накопителе.

Рабочая директория

Опр. Рабочая директория (working directory) – директория, содержащая непосредственно доступные пользователю файлы.

Путь

Опр. Путь (pathname) – строка, обозначающая файл или директорию их логическими именами, разделяя директории символами-разделителями (например, «\»). Абсолютный путь указывает размещение файла или директории, начиная с корневой директории; относительный путь указывает их размещение, начиная с текущей рабочей директории.

Мягкая ссылка

Опр. Мягкая ссылка (soft link) – файл, содержащий путь к другому файлу, на который он ссылается.

Пр. Ярлык в Windows

Жесткая ссылка

Опр. Жесткая ссылка (hard link) – запись в директории, указывающая физическое размещение файла на носителе (обычно номер начального блока).

Вопрос для самопроверки

- Одноуровневые файловые системы подходят для большинства систем?
(Да/Нет)

Вопрос для самопроверки

- Одноуровневые файловые системы подходят для большинства систем? (Да/Нет)
- Нет. В большинстве систем нужно хранить множество файлов с совпадающими именами, что невозможно в одноуровневых файловых системах.

Вопрос для самопроверки

- Директория – это файл? (Да/Нет)

Вопрос для самопроверки

- Директория – это файл? (Да/Нет)
- Да. Директория – это файл, хранящий ссылки на другие файлы.

Вопрос для самопроверки

- Жесткая ссылка – это файл? (Да/Нет)

Вопрос для самопроверки

- Жесткая ссылка – это файл? (Да/Нет)
- Нет. Жесткая ссылка – это запись в директории, указывающая на размещение файла на устройстве хранения. В виде файлов реализуются мягкие ссылки.

Файлы и файловые системы

Метаданные

Метаданные

Опр. Метаданные (metadata) – недоступные непосредственно пользователям данные, с помощью которых файловая система управляет файлами. Например, данные о свободных блоках накопителя (чтобы гарантировать, что новые данные не будут записаны поверх записанных ранее), о времени последнего изменения файлов (для целей учета) и пр.

Суперблок

Опр. Суперблок (superblock) – блок, содержащий жизненно важные для обеспечения целостности файловой системы метаданные (например, количество блоков в файловой системе, список или битовый массив свободных блоков, идентификатор файловой системы, расположение корневой директории).

Форматирование накопителя

Опр. Форматирование накопителя (formatting of storage device) – подготовка накопителя для файловой системы, которая обычно включает проверку накопителя на наличие неработающих областей, стирание ранее хранившихся на нем данных, создание корневой директории. Многие файловые системы при форматировании также создают суперблок.

Таблица открытых файлов

- При открытии файла операционная система сначала находит информацию о нем, просматривая структуру директорий
- Чтобы избежать многократных просмотров система хранит в оперативной памяти таблицу, ведущую учет открытых файлов
- Таблица открытых файлов обычно содержит
 - Дескрипторы файлов
 - Блоки управления файлами

Дескриптор файла

Опр. Дескриптор файла (file descriptor) – неотрицательное целое число, являющееся индексом в таблице открытых файлов. Процесс обращается к дескриптору вместо имени файла, чтобы получить доступ к данным файла без необходимости перемещаться по структуре директорий.

Блок управления файлом

Опр. Блок управления файлом (file control block) – метаданные, содержащие необходимую файловой системе информацию о файле.

Обычное содержимое блока управления файлом

- Символьное имя файла
- Данные о расположении файла на накопителе
- Организационная структура (например, файл последовательного доступа или произвольного доступа)
- Сведения о типе накопителя (например, жесткий диск или компакт-диск)

Обычное содержимое блока управления файлом

- Данные управления доступом (например, о том, какие пользователи могут обращаться к файлу и какие они могут выполнять операции)
- Данные о типе файла (например, файл данных, программа на языке С или исполняемый файл)

Обычное содержимое блока управления файлом

- Характер файла (постоянный или временный)
- Счетчики обращений к файлу (например, количество операций чтения)
- Дата и время создания файла
- ...

Вопрос для самопроверки

- Желательно ли хранить избыточные копии суперблоков? (Да/Нет)

Вопрос для самопроверки

- Желательно ли хранить избыточные копии суперблоков? (Да/Нет)
- Да. Если суперблок будет поврежден, файловая система будет не в состоянии обратиться к файлам. Поэтому большинство файловых систем хранит избыточные копии суперблоков на накопителях. Если суперблок будет поврежден, его можно заменить одной из копий.

Вопрос для самопроверки

- Могут ли пользователи непосредственно обращаться к метаданным? (Да/Нет)

Вопрос для самопроверки

- Могут ли пользователи непосредственно обращаться к метаданным? (Да/Нет)
- Нет. Если доступ к метаданным не ограничивать, то случайное неправильное их использование может привести к потере целостности данных или их уничтожению.

Файлы и файловые системы

Монтирование

Монтирование

Опр. Монтирование (mount operation) – операция, объединяющая две отдельные файловые системы в одно пространство имен, делая их доступными из общей корневой директории. Применяется, например, для доступа к данным, хранящимся на переносном винчестере, DVD, или на другой рабочей станции в сети.

Пространство имен

Опр. Пространство имен (namespace) – набор файлов, которые могут идентифицироваться файловой системой.

Точка монтирования

Опр. Точка монтирования (mount point) – заданная пользователем директория в иерархии родной файловой системы, в которую команда монтирования помещает корневую директорию монтируемой файловой системы.

Плоская структура монтирования

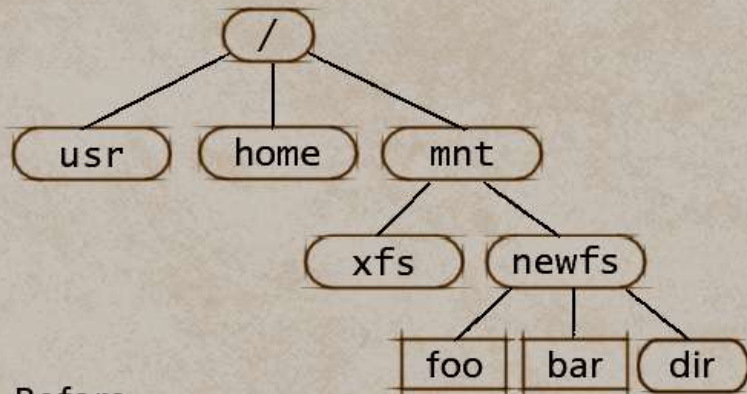
- Использовалась в ранних версиях файловых систем OS Windows
- Каждая монтируемая файловая система обозначалась буквой и размещалась на одном и том же уровне структуры директорий
- Например, файловая система, содержащая файлы операционной системы обозначалась C:, а следующая файловая система – D:

Произвольное монтирование

- Используется в NTFS и файловых системах UNIX
- Точки монтирования можно размещать в файловой системе где угодно
- Содержимое системной директории родной файловой системы в точке монтирования временно скрывается, пока другая файловая система монтирована в этой директории

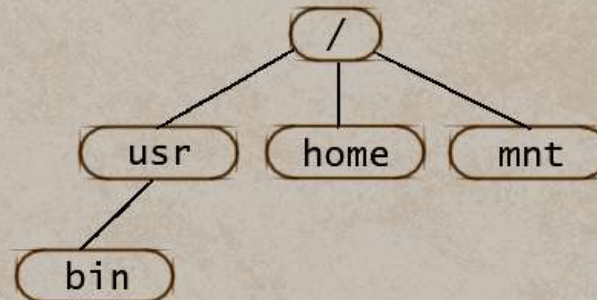
Произвольное монтирование файловой системы

File system A

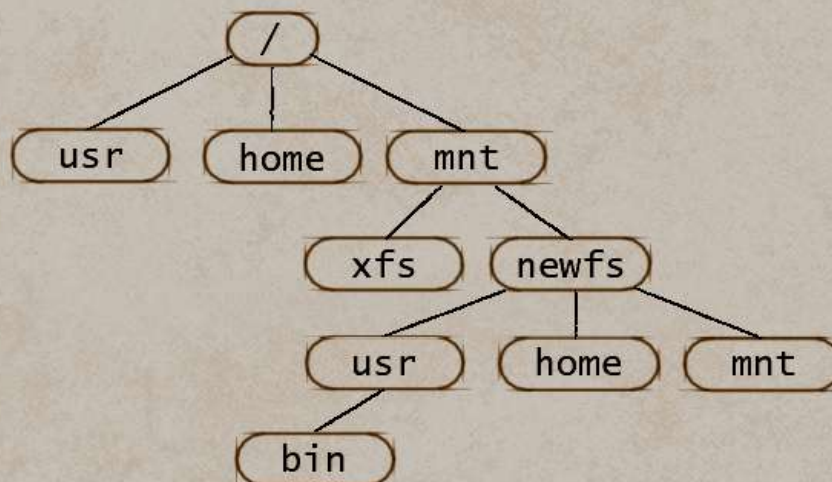


Before

File system B



File system B mounted at directory /mnt/newfs in file system A



After

Вопрос для самопроверки

- Смонтированная файловая система и родная должны быть одного типа?
(Да/Нет)

Вопрос для самопроверки

- Смонтированная файловая система и родная должны быть одного типа?
(Да/Нет)
- Нет. Основное преимущество монтирования файловых систем состоит в том, что оно позволяет нескольким различным файловым системам быть доступными через единый интерфейс.

Вопрос для самопроверки

- Может ли файловая система создать жесткие ссылки на файлы в смонтированной файловой системе?
(Да/Нет)

Вопрос для самопроверки

- Может ли файловая система создать жесткие ссылки на файлы в смонтированной файловой системе? (Да/Нет)
- Нет. Жесткие ссылки указывают специфические для устройства номера блоков, соответствующие файловой системе, в которой они хранятся, и они не могут использоваться для указания физического размещения данных в других файловых системах.