Chapter 8

File Management

At a Glance

Instructor's Manual Table of Contents

Lecture Notes

Overview

The File Manager controls every file in the system. The chapter begins with an introduction to this file management system. This is followed by a presentation on definitions for commonly used terms. The chapter explains how files are organized logically, how they are stored physically, how they are accessed, and who is allowed to access them. The chapter also describes the interaction between the File Manager and the Device Manager. The efficiency of the File Manager depends on how the system's files are organized (sequential, direct, or indexed sequential); how they are stored (contiguously, noncontiguously, or indexed); how each file's records are structured (fixed-length or variable-length); and how access to these files is controlled. The chapter concludes with a discussion on data compression.

Learning Objectives

After completing this chapter, the student should be able to describe:

- The fundamentals of file management and the structure of the file management system
- File-naming conventions, including the role of extensions
- The difference between fixed-length and variable-length record format
- The advantages and disadvantages of contiguous, noncontiguous, and indexed file storage techniques
- Comparisons of sequential and direct file access
- The security ramifications of access control techniques and how they compare
- The role of data compression in file storage

Teaching Tips

The File Manager

1. Provide students with an overview of the File Manager, outlining its responsibilities. Note that the File Manager is also called the file management system. Explain that it is software responsible for creating, deleting, modifying and controlling access to files. Point out that it also manages the resources used by the files.

2. Explain how the File Manager provides support for libraries of programs and data to online users for spooling operations and for interactive computing.

3. Explain that the File Manager performs its functions in collaboration with the Device Manager.

Teaching Tip	Refer to the following Web site in order to learn more about File Manager tasks: http://publib.boulder.ibm.com/iseries/v5r2/ic2924/index.htm?info/ddm/rbae5mst02.htm

Responsibilities of the File Manager

1. Discuss how the File Manager's responsibilities are quite complex and involve the physical components and information resources of the system, as well as its policies for storing and distributing files.

2. Outline the four tasks that the File Manager follows to carry out its responsibilities:
   a. Keep track of where each file is stored.
   b. Use a policy that will determine where and how the files will be stored, making sure to efficiently use available storage space and provide efficient access to files.
   c. Allocate each file when a user has been cleared for access to it, and then record its use.
   d. Deallocate the file when the file is to be returned to storage, and communicate its availability to others who may be waiting for it.

3. Point out that the File Manager's policy determines where each file is stored and how the system and users will be able to access them through commands that are independent from device details.

4. Discuss the two factors the File Manager's policy utilizes to determine who may have access to what material: flexibility of access to the information and its subsequent protection. Explain that the File Manager allows access to shared files, provides distributed access, and allows users to browse through public directories. Additionally, the operating system must protect its files against system malfunctions and provide security checks through account numbers, passwords, and lockwords to preserve the integrity of the data and safeguard against tampering.

5. Describe the steps involved in the file allocation and deallocation processes.

Definitions

1. Define the term field: a group of related bytes that can be identified by the user with a name, type, and size.

2. Define the term record: a group of related fields.

3. Define the term file: a group of related records that contains information used by specific application programs to generate reports.

4. Use Figure 8.1 on page 253 in the text to illustrate these terms.

5. Discuss how a database appears to the File Manager to be a type of file. Emphasize that databases are more complex, because they are actually groups of related files that are interconnected at various levels to give users flexibility of access to the stored data.

6. Explain that a program file contains instructions and a data file contains data.

7. Discuss how directories are special files with listings of filenames and their attributes.

Teaching Tip	Point out that the File Manager treats all files exactly the same way as far as storage is concerned.

Interacting with the File Manager

1. Explain how the user communicates with the File Manager. There are two methods. The first is to embed commands in the user's program. The second is to submit them interactively.

2. Provide examples of embedded commands (OPEN, CLOSE, READ, WRITE MODIFY). Note that OPEN and CLOSE pertain to the file availability for the program working on it. READ and WRITE are Input/Output (I/O) commands. MODIFY is a specialized command.

3. Provide examples of interactive commands (OPEN, DELETE, RENAME, COPY). Note that files may be created with other system-specific terms and provide examples.

4. Use Figure 8.2 on page 255 in the text to demonstrate common interactive commands.

5. Emphasize that these commands are designed to be simple and devoid of detailed instructions required to run the device. Note that the details are found in the device driver so the commands are device independent.

6. Discuss how due to device independence and command simplicity, the user does not need to know a file's exact physical location on the disk pack, the medium type, or network specifics.

7. Explain how logical commands are broken down into a sequence of lower-level signals. These signals trigger the systematic actions performed by the device. These signals also supervise the progress of the operation by testing the device status.

8. Review the example on page 255 of the decomposed steps taken when a user program issues a READ command.

9. Describe the need for error checking and how the File Manager handles this on behalf of the user for every device. Emphasize that this is important because users do not have to include these error checking instructions for every possible device (terminal, keyboard, printer, disk drive, etc.) within their own programs.

Typical Volume Configuration

1. Provide students with an overview of the concept of a volume, which is a secondary storage unit (removable or non-removable). Explain the concept of multifile volumes. Explain the concept of multivolume files.

2. Explain the volume descriptor using Figure 8.3 on page 256 in the text. Note that it is stored at the beginning of each volume and includes the volume name and other vital information about the storage unit.

3. Discuss briefly the master file directory (MFD), which is stored immediately after the volume descriptor. Outline what it lists. Point out the fact that early operating systems supported only a single directory per volume.

4. Outline the disadvantages of a single directory per volume as supported by early operating systems, using the bullet points listed on pages 256 and 257 in the text as a guide.

About Subdirectories

1. Discuss how in newer File Manager components, the MFD for each volume contains entries for both files and subdirectories. Explain that a subdirectory is created when the user opens an account to access a computer. Note that this subdirectory is treated as a file in the MFD and the entry is flagged to indicate that this file is really a subdirectory and has unique properties. Specifically, its records are filenames pointing to files.

2. Emphasize that this is an improvement from a single directory scheme but still precludes users from grouping their files in a logical order to improve the accessibility and efficiency of the system.

3. Provide an overview of the subdirectories created by today's File Managers. Point out that this structure is an extension of the previous two-level directory organization and is implemented as an upside-down tree. Use Figure 8.4 on page 258 in the text to illustrate this concept.

4. Outline the information that is included in the file descriptor following the bullet points listed on page 258 in the text.

Teaching Tip	Today's sophisticated File Managers allow their users to create their own subdirectories so related files can be grouped together. Many computer users call these subdirectories "folders."

File-Naming Conventions

1. Explain to the students that there are two components comprising a filename: a relative filename and an extension. Review the difference between the complete filename (absolute filename) and the relative filename.

2. Discuss different types of extensions. Explain that an extension identifies a file type or its contents. Provide examples. Note that if an extension is not known, the system asks the user for intervention. Use Figure 8.5 on page 259 in the text to demonstrate this.

3. Point out that every operating system has specific rules that affect the length of the relative name and the types of characters allowed. Illustrate with the operating systems presented on page 260, and note how the names tend to grow in length as the File Managers grow in flexibility.

File Organization

1. Provide students with an overview of file organization. Remind students that all files are composed of records and that when a user gives a command to modify the contents of a file, it is actually a command to access the records within the file.

Record Format

1. Explain to students that within each file, records are all presumed to have the same format. They may be (1) fixed length or (2) variable length. Use Figure 8.6 on page 261 in the text to illustrate these two types pictorially.

2. Discuss the two types of record formats and the characteristics of each. Note that these records, regardless of their format, can be blocked or not blocked. Use examples as shown in Figure 8.4 on page 258 in the text to explain both fixed-length and variable-length records.

3. Discuss the advantages and disadvantages of fixed-length and variable-length records. Provide examples to show where these are useful. For example, fixed-length records are ideal for data files, whereas the variable-length formats are used in files accessed sequentially (text files, program files), or in files that use an index to access records.

Physical File Organization

1. Discuss the physical organization of a file, which involves how records are arranged, and the characteristics of the medium used to store it.

2. Remind students of the ways in which files can be stored on magnetic disks, such as sequential, direct, or indexed sequential. Discuss practical characteristics that the programmer or analyst usually considers when selecting the ideal file organization.

3. Discuss sequential record organization, pointing out how records are stored and retrieved in this scheme. Explain how selecting a key field speeds up the record searching process. Discuss the advantages and disadvantages of this scheme. For example, it is easy to implement; however, it complicates maintenance algorithms, as the original order must be preserved every time records are added or deleted.

4. Discuss the direct record organization, which uses direct access files that can be implemented only on direct access storage devices. Explain how records are identified in direct record organization. Discuss the concepts of relative address, logical address, and hashing algorithms. Note that the problem with hashing algorithms is that several records with unique keys (such as customer numbers) may generate the same logical addresses, and then there's a collision. Use Figure 8.7 on page 264 in the text to illustrate the collision problem with the hashing algorithm.

5. Discuss the advantages of direct record organization over sequential record organization. For example, it allows accessing of any record in any order without having to begin the search from the beginning of the file, and the files can be updated more quickly than sequential files.

6. Discuss the indexed sequential record organization, which is created and maintained through an Indexed Sequential Access Method (ISAM) software package. Point out the advantages of using this software package.

7. Explain how records are retrieved in the indexed sequential record organization. Discuss its advantages over the previous two organizations, including that it does not create collisions, the overflow area can be used to store records added during the lifetime of the file, and it allows both direct access to a few requested records and sequential access to many records.

Teaching Tip	Refer to the following Web site to learn more about file organization: www.cim.mcgill.ca/~franco/OpSys-304-427/lecture-notes/node53.html

Quick Quiz 1

1. Which of the following is an embedded command?
   a. Create
   b. Delete
   c. Copy
   d. Open
   Answer: d

2. Which of the following information is contained in the volume descriptor? (Choose all that apply.)
   a. Creation date
   b. User's name
   c. File system code
   d. Volume length
   Answer: a and c

3. Which of the following organization schemes uses a hashing algorithm?
   a. Direct record organization
   b. Sequential record organization
   c. Indexed sequential record organization
   d. Indexed direct record organization
   Answer: a

Physical Storage Allocation

1. Provide students with an overview of physical storage allocation. Remind students that the File Manager works with files not just as whole units, but also as logical units or records. Be sure to mention that records within a file must have the same format; however, they can vary in length. Also mention that file storage actually refers to record storage.

2. Use Figure 8.8 on page 266 in the text to explain the most common record formats.

Contiguous Storage

1. Explain the contiguous storage scheme, which was used in early operating systems. Discuss its advantages and disadvantages, such as that it is easy to implement but files cannot be expanded easily, and that it leads to fragmentation.

2. Use Figure 8.9 on page 266 in the text to illustrate contiguous file storage.

Noncontiguous Storage

1. Explain the noncontiguous storage scheme, outlining its advantages over a contiguous storage scheme.

2. Use Figure 8.10 on page 267 and Figure 8.11 on page 269 in the text to explain the two ways through which file extents are usually linked. These include linking occurring at the storage level and at the directory level. Discuss the advantages and disadvantages of these two methods.

Indexed Storage

1. Explain the indexed storage scheme, which allows direct record access by bringing together the pointers linking every extent of that file into an index block.

2. Use Figure 8.12 on page 269 in the text to illustrate the concept of an index storage allocation with one block.

3. Discuss the advantages and disadvantages of this scheme. Point out that it supports both sequential and direct record access; however, it does not necessarily improve the use of storage space because each file must have an index block.

Teaching Tip	Refer to the following Web site in order to learn more about physical storage allocation: www.unet.univie.ac.at/aix/aixprggd/genprogc/file_space_allocation.htm

Access Methods

1. Provide students with an overview of access methods. Point out that access methods are dictated by a file's organization. That is, a file that has been organized in sequential fashion can support only sequential access to its records and these records can be of either fixed or variable length.

2. Explain how the File Manager uses the address of the last byte read to access the next sequential record. Therefore, the current byte address (CBA) must be updated every time a record is accessed.

3. Use Figure 8.13 on page 270 to illustrate the difference between the storage of fixed-length and of variable-length records.

Sequential Access

1. Explain how the CBA is updated for sequential access of fixed-length records and sequential access of variable-length records.

Direct Access

1. Explain how the CBA is updated for direct access with fixed-length records.

2. Explain why it is virtually impossible to access a record directly if the file is organized for direct access with variable-length records. Discuss the alternatives.

3. Explain how an indexed sequential file can be accessed. Point out that the index file must be searched for the pointer to the block where the data is stored.

Levels in a File Management System

1. Explain that the efficient management of files cannot be separated from the efficient management of the devices housing them. Note that for an I/O system to perform efficiently, there are a wide range of functions that have to be organized and managed. Point out that hierarchies are used to perform these functions.

2. Discuss the levels in a file management system, and explain how information is passed from the File Manager at the top of the hierarchy to the Device Manager at the bottom.

3. Use Figure 8.14 on page 272 in the text to illustrate the typical modules of the file management system. Point out that the highest-level module is called the “basic file system,” and it passes information to the logical file system, which, in turn notifies the physical file system, which works with the Device Manager.

4. Explain that each level of the file management system is implemented using structured and modular programming techniques that set up the hierarchy. Explain that in the hierarchy, higher-level modules pass information to the lower-level modules, and the process continues down the chain until the lowest module that communicates with the physical device and interacts with the Device Manager is reached.

5. Note that each module is subdivided into tasks that are more specific. Walk through the example provided on page 273 with the students to follow an I/O instruction through the file management system. Refer to Table 8.1 on page 273 of the text.

6. Discuss the various levels at which verification occurs in the file management system. First, verification occurs in the directory level. Second, verification occurs in the access control verification module. Third, it happens in the logical file system. Fourth, verification occurs in the device interface module. Emphasize the coordinated effort required by every part of the file management system for correct operation.

Access Control Verification Module

1. Provide students with an overview of the need for an access control verification module, discussing early operating systems and today’s operating systems from the file-sharing capabilities point of view. Note that all files may be shared. Point out the advantages and disadvantages of file sharing and the need for access control.

2. Outline five possible actions that can be performed on a file. They include the ability to READ only, WRITE only, EXECUTE only, DELETE only, or some combination of the four. Point out that each file management system has its own method to control file access.

3. Outline the four methods to control file access. They include the access control matrix, access control lists, capability lists, and lockword control.

Access Control Matrix

1. Explain the access control matrix method using Table 8.2 on page 275 and Table 8.3 on page 276 in the text. Discuss the advantages of this method, such as the fact that it is easy to implement and works well for systems with few files and few users. Discuss the disadvantages, noting that as the number of files and users increases, the matrix becomes larger, perhaps too large to fit into main memory. In addition, it results in wasted space because of null entries.

Access Control Lists

1. Explain the access control lists method using Table 8.4 on page 276 in the text. Point out that this method requires less storage space than an access control matrix and that it explicitly names each user allowed access to each file.

2. Explain how some systems shorten the access control list even more by putting every user into different categories. Examples of categories include system, owner, group, and world. Explain each of these categories briefly as described on page 277 in the text.

Capability Lists

1. Explain the capability lists method using Table 8.5 on page 277 in the text. Point out its advantages. For example, it requires less storage space than an access control matrix, and it is easier to maintain than an access control list when users are added or deleted from the system. Moreover, it can control access to devices as well as to files.

Lockwords

1. Provide students with an overview of the use of a lockword, a protection technique similar to a password. Note that it protects a single file. Discuss its advantages and disadvantages. For example, it requires the smallest amount of storage for file protection; however, it can be guessed by hackers or passed on to unauthorized users. Moreover, it does not control the type of access to a file. That is, anyone who knows the lockword can read, write, execute, or delete the file.

Data Compression

1. Explain data compression techniques briefly, which are used to save space in files.

Text Compression

1. Discuss different methods that are used in text compression. These include compressing text records with repeated characters, with repeated terms, or by using a front-end compression scheme. Provide examples as shown in the text on page 278 to clarify each of these methods.

2. Point out that there is always a trade-off in data compression. For example, storage space is gained but processing time is lost. Be sure to mention that for all data compression schemes, the system must be able to distinguish between compressed and uncompressed data.

Other Compression Schemes

1. Inform the students that there are other compression standards for large files such as video, music, and photograph files. Explain to the students that the International Organization for Standardization has a mission to issue standards that all systems and users may abide by when working with these files.

Teaching Tip	Refer to the following Web site to learn more about data compression techniques: www.vectorsite.net/ttdcmp.html

Teaching Tip	For more on the International Organization for Standardization, go to www.iso.org.

Quick Quiz 2

1. Which of the following storage allocation schemes results in the problem of fragmentation?
   a. Contiguous storage
   b. Noncontiguous storage
   c. Indexed storage
   d. Direct storage
   Answer: a

2. Which of the following methods can control access to files as well as to devices?
   a. Capability list
   b. Access control list
   c. Access control matrix
   d. All of the above
   Answer: a

3. Which of the following data compression techniques compresses text, using symbols to represent most commonly used words?
   a. Records with repeated characters
   b. Repeated terms
   c. Front-end compression
   d. All of the above
   Answer: b

Class Discussion Topics

1. Have students discuss different types of file organization schemes. If they were approached by a hospital to maintain the data records of the patients, which file organization would they prefer? Why?

2. Have students discuss different access control verification methods. What are the advantages and disadvantages of each method? Which of these methods would they prefer for a highly secure system with several files and several users? Have students substantiate their choices with reasons.

Additional Projects

1. Have students research multiple file systems for hard disks supported in Windows, such as MS DOS's FAT file system, the 32-bit FAT file system, and the NTFS file management system.

2. Have students perform research to compile a list of various data-compression techniques that are commonly employed in today's computer systems. Which of these techniques provides the best efficiency – that is, the maximum gain in storage space with the least effect on processing time?

Additional Resources

1. Microsoft.com:
   www.microsoft.com

2. IBM.com:
   www.ibm.com

3. A History of Operating Systems:
   www.osdata.com/kind/history.htm

4. Intel.com:
   www.intel.com

5. Freescale.com:
   www.freescale.com

6. Sun Microsystems:
   www.sun.com/

7. International Organization for Standardization:
   www.iso.org

Key Terms

- **Absolute filename:** a file's name, as given by the user, preceded by the directory (or directories) where the file is found and, when necessary, the specific device label.
- **Access control list:** an access control method that lists each file, the names of the users who are allowed to access it, and the type of access each is permitted.
- **Access control matrix:** an access control method that uses a matrix with every file, every user, and the type of access each user is permitted on each file.
- **Capability list:** an access control method that lists every user, the files to which each has access, and the type of access allowed to those files.
- **Contiguous storage:** a type of file storage in which all the information is stored in adjacent locations in a storage medium.

- **Current byte address (CBA):** the address of the last byte read. It is used by the File Manager to access records in secondary storage and must be updated every time a record is accessed.
- **Current directory:** the directory or subdirectory in which the user is working.
- **Data compression:** a procedure used to reduce the amount of space required to store data by reducing, encoding, or abbreviating repetitive terms or characters.
- **Data file:** a file that contains only data.
- **Database:** a group of related files that are interconnected at various levels to give users flexibility of access to the data stored.
- **Device independent:** programs that are devoid of the detailed instructions required to interact with any I/O device present in the computer system.
- **Direct record organization:** files stored in a direct access storage device and organized to give users the flexibility of accessing any record at random, regardless of its position in the file.
- **Directory:** a storage area in a secondary storage volume (disk, disk pack, etc.) containing information about files stored in that volume.
- **Extension:** in some operating systems, it's the part of the filename that indicates which compiler or software package is needed to run the files. In UNIX and Linux, it is optional and called a suffix.
- **Extents:** any remaining records and all other additions to the file that are stored in other sections of the disk.
- **Field:** a group of related bytes that can be identified by the user with a name, type, and size. A record is made up of fields.
- **File:** a group of related records that contains information to be used by specific application programs to generate reports.
- **File descriptor:** information kept in the directory to describe a file or file extent.
- **Fixed-length record:** a record that always contains the same number of characters.
- **Hashing algorithm:** the set of instructions used to perform a key-to-address transformation in which a record's key field determines its location.
- **Indexed sequential record organization:** a way of organizing data in a direct access storage device. An index is created to show where the data records are stored. Any data record can be retrieved by consulting the index first.
- **Key field:** (1) a unique field or combination of fields in a record that uniquely identifies that record; or (2) the field that determines the position of a record in a sorted sequence.
- **Lockword:** a sequence of letters and/or numbers provided by users to prevent unauthorized tampering with their files.
- **Logical address:** the result of a key-to-address transformation.
- **Master file directory (MFD):** a file stored immediately after the volume descriptor. It lists the names and characteristics of every file contained in that volume.
- **Noncontiguous storage:** a type of file storage in which the information is stored in nonadjacent locations in a storage medium.
- **Program file:** a file that contains instructions for the computer.
- **Record:** a group of related fields treated as a unit. A file is a group of related records.
- **Relative address:** in a direct organization environment, it indicates the position of a record relative to the beginning of the file.
- **Relative filename:** a file's name and extension that differentiates it from other files in the same directory.
- **Sequential record organization:** the organization of records in a specific sequence. Records in a sequential file must be processed one after another.

- **Subdirectory:** a directory created by the user within the boundaries of an existing directory. Some operating systems call this a folder.
- **Variable-length record:** a record that isn't of uniform length, doesn't leave empty storage space, and doesn't truncate any characters, thus eliminating the two disadvantages of fixed-length records.
- **Volume:** any secondary storage unit, such as hard disks, disk packs, CDs, DVDs, removable disks, or tapes.
- **Working directory:** the directory or subdirectory in which the user is currently working.