File Management

Chapter 12

1

File Management

- File management system is considered part of the operating system
- Input to applications is by means of a file
- Output is saved in a file for long-term storage

2

Terms Used with Files

- Field
  - Basic element of data
  - Contains a single value
  - Characterized by its length and data type
- Record
  - Collection of related fields
  - Treated as a unit
    - Example: employee record

3

Terms Used with Files

- File
  - Collection of similar records
  - Treated as a single entity
  - Have unique file names
  - May restrict access
- Database
  - Collection of related data
  - Relationships exist among elements

4

Typical Operations

- Retrieve_All
- Retrieve_One
- Retrieve_Next
- Retrieve_Previous
- Insert_One
- Delete_One
- Update_One
- Retrieve_Few

5

File Management System

- The way a user of application may access files
- Programmer does not need to develop file management software

6

Objectives for a File Management System

- Meet the data management needs and requirements of the user
- Guarantee that the data in the file are valid
- Optimize performance
- Provide I/O support for a variety of storage device types

7

Objectives for a File Management System

- Minimize or eliminate the potential for lost or destroyed data
- Provide a standardized set of I/O interface routines
- Provide I/O support for multiple users

8

Minimal Set of Requirements

- Each user should be able to create, delete, read, and change files
- Each user may have controlled access to other users' files
- Each user may control what type of accesses are allowed to the users' files
- Each user should be able to restructure the user's files in a form appropriate to the problem

9

Minimal Set of Requirements

- Each user should be able to move data between files
- Each user should be able to back up and recover the user's files in case of damage
- Each user should be able to access the user's files by using symbolic names

10

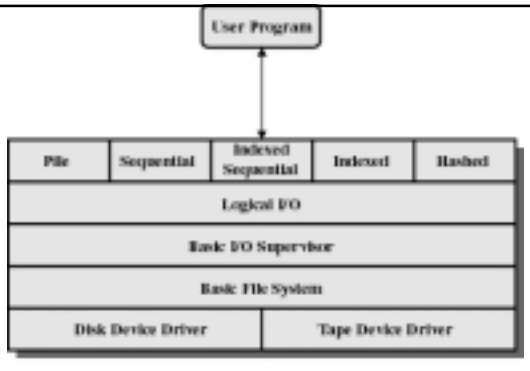


Figure 12.1 File System Software Architecture [GROS86]

Device Drivers

- Lowest level
- Communicates directly with peripheral devices
- Responsible for starting I/O operations on a device
- Processes the completion of an I/O request

12

Basic File System

- Physical I/O
- Deals with exchanging blocks of data
- Concerned with the placement of blocks
- Concerned with buffering blocks in main memory

13

Basic I/O Supervisor

- Responsible for file I/O initiation and termination
- Control structures are maintained
- Concerned with scheduling access to optimize performance
- Part of the operating system

14

Logical I/O

- Enables users and applications to access records
- Provides general-purpose record I/O capability
- Maintains basic data about file

15

Access Method

- Reflect different file structures
- Different ways to store and process data

16

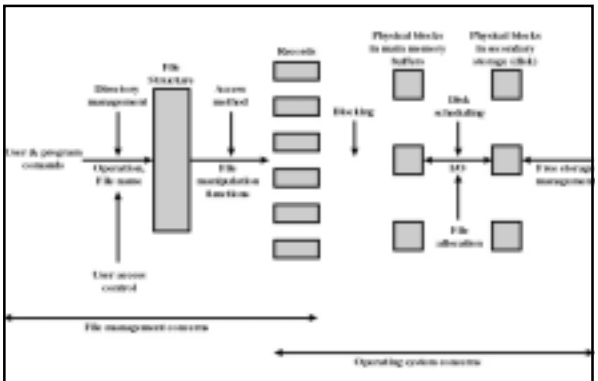


Figure 12.2 Elements of File Management

File Management Functions

- Identify and locate a selected file
- Use a directory to describe the location of all files plus their attributes
- On a shared system describe user access control
- Blocking for access to files
- Allocate files to free blocks
- Manage free storage for available blocks

18

Criteria for File Organization

- Rapid access
  - Needed when accessing a single record
  - Not needed for batch mode
- Ease of update
  - File on CD-ROM will not be updated, so this is not a concern

19

Criteria for File Organization

- Economy of storage
  - Should be minimum redundancy in the data
  - Redundancy can be used to speed access such as an index
- Simple maintenance
- Reliability

20

File Organization

- The Pile
  - Data are collected in the order they arrive
  - Purpose is to accumulate a mass of data and save it
  - Records may have different fields
  - No structure
  - Record access is by exhaustive search

21

Pile

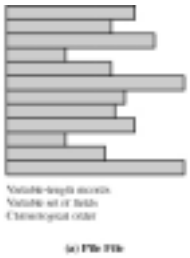


Figure 12.3 Common File Organizations

22

File Organization

- The Sequential File
  - Fixed format used for records
  - Records are the same length
  - All fields the same (order and length)
  - Field names and lengths are attributes of the file
  - One field is the key filed
    - Uniquely identifies the record
    - Records are stored in key sequence

23

File Organization

- The Sequential File
  - New records are placed in a log file or transaction file
  - Batch update is performed to merge the log file with the master file

24

Sequential File

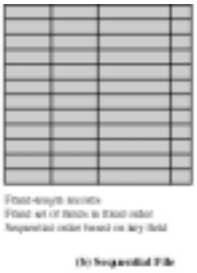


Figure 12.3 Common File Organizations

File Organization

- Indexed Sequential File
  - Index provides a lookup capability to quickly reach the vicinity of the desired record
    - Contains key field and a pointer to the main file
    - Indexed is searched to find highest key value that is equal or less than the desired key value
    - Search continues in the main file at the location indicated by the pointer

26

File Organization

- Comparison of sequential and indexed sequential
  - Example: a file contains 1 million records
  - On average 500,00 accesses are required to find a record in a sequential file
  - If an index contains 1000 entries, it will take on average 500 accesses to find the key, followed by 500 accesses in the main file. Now on average it is 1000 accesses

27

File Organization

- Indexed Sequential File
  - New records are added to an overflow file
  - Record in main file that precedes it is updated to contain a pointer to the new record
  - The overflow is merged with the main file during a batch update
  - Multiple indexes for the same key field can be set up to increase efficiency

28

Indexed Sequential File

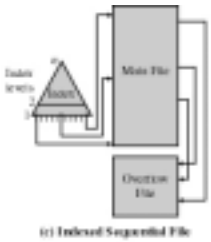


Figure 12.3 Common File Organizations

File Organization

- Indexed File
  - Uses multiple indexes for different key fields
  - May contain an exhaustive index that contains one entry for every record in the main file
  - May contain a partial index

30

Indexed File

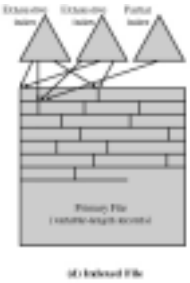


Figure 12.3 Common File Organizations

31

File Organization

- The Direct, or Hashed File
  - Directly access a block at a known address
  - Key field required for each record

32

File Directories

- Contains information about files
  - Attributes
  - Location
  - Ownership
- Directory itself is a file owned by the operating system
- Provides mapping between file names and the files themselves

33

Simple Structure for a Directory

- List of entries, one for each file
- Sequential file with the name of the file serving as the key
- Provides no help in organizing the files
- Forces user to be careful not to use the same name for two different files

34

Two-level Scheme for a Directory

- One directory for each user and a master directory
- Master directory contains entry for each user
  - Provides address and access control information
- Each user directory is a simple list of files for that user
- Still provides no help in structuring collections of files

35

Hierarchical, or Tree-Structured Directory

- Master directory with user directories underneath it
- Each user directory may have subdirectories and files as entries

36

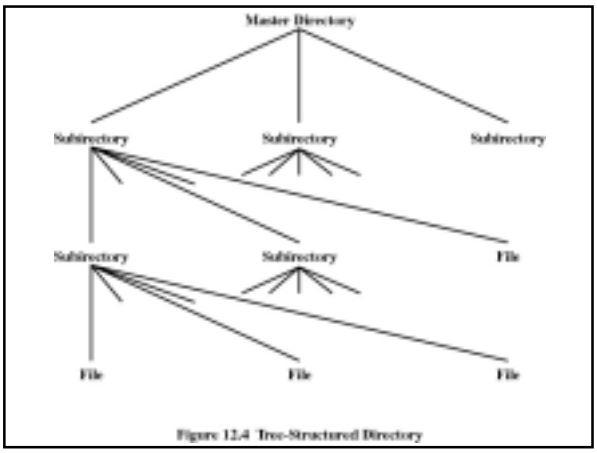
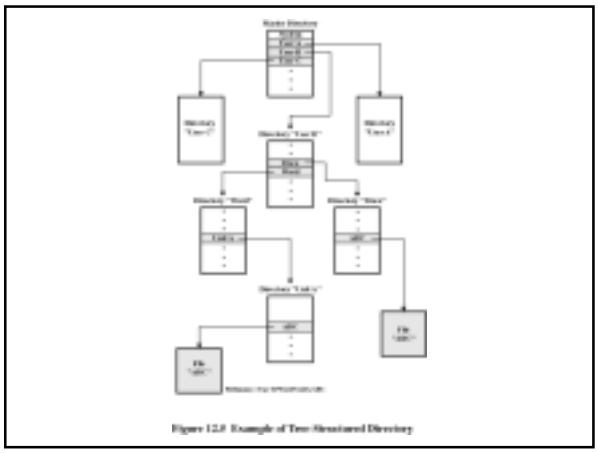


Figure 12.4 Tree-Structured Directory



Hierarchical, or Tree-Structured Directory

- Files can be located by following a path from the root, or master, directory down various branches
  - This is the pathname for the file
- Can have several files with the same file name as long as they have unique path names

39

Hierarchical, or Tree-Structured Directory

- Current directory is the working directory
- Files are referenced relative to the working directory

40

File Sharing

- In multiuser system, allow files to be shared among users
- Two issues
  - Access rights
  - Management of simultaneous access

41

Access Rights

- None
  - User may not know of the existence of the file
  - User is not allowed to read the user directory that includes the file
- Knowledge
  - User can only determine that the file exists and who its owner is

42

Access Rights

- Execution
  - The user can load and execute a program but cannot copy it
- Reading
  - The user can read the file for any purpose, including copying and execution
- Appending
  - The user can add data to the file but cannot modify or delete any of the file's contents

43

Access Rights

- Updating
  - The user can modify, deleted, and add to the file's data. This includes creating the file, rewriting it, and removing all or part of the data
- Changing protection
  - User can change access rights granted to other users
- Deletion
  - User can delete the file

44

Access Rights

- Owners
  - Has all rights previously listed
  - May grant rights to others using the following classes of users
    - Specific user
    - User groups
    - All for public files

45

Simultaneous Access

- User may lock entire file when it is to be updated
- User may lock the individual records during the update
- Mutual exclusion and deadlock are issues for shared access

46

Fixed Blocking

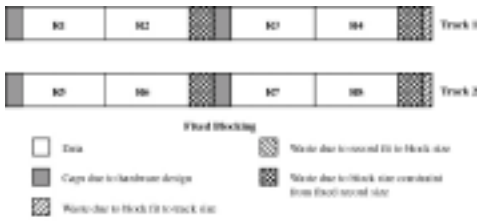


Figure 12.6 Record Blocking Methods [WIED87]

47

Variable Blocking: Spanned

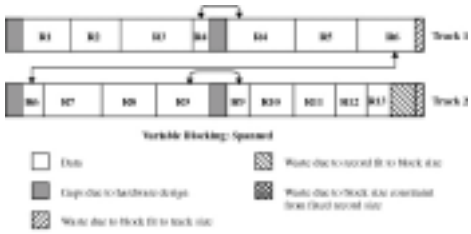


Figure 12.6 Record Blocking Methods [WIED87]

48

Variable Blocking Unspanned

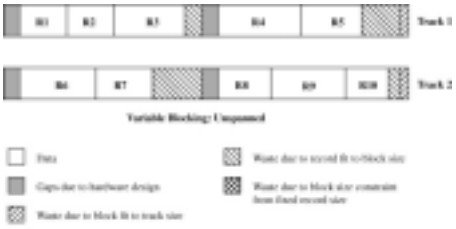


Figure 12.6 Record Blocking Methods [WIED87]

49

Secondary Storage Management

- Space must be allocated to files
- Must keep track of the space available for allocation

50

Preallocation

- Need the maximum size for the file at the time of creation
- Difficult to reliably estimate the maximum potential size of the file
- Tend to overestimated file size so as not to run out of space

51

Methods of File Allocation

- Contiguous allocation
  - Single set of blocks is allocated to a file at the time of creation
  - Only a single entry in the file allocation table
    - Starting block and length of the file
- External fragmentation will occur

52

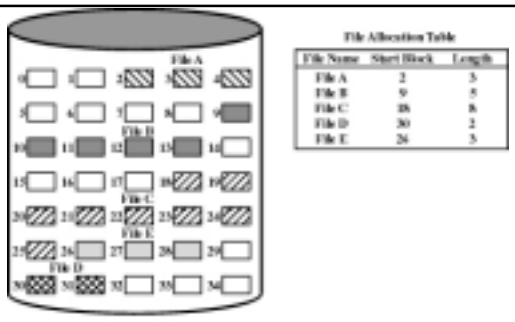


Figure 12.7 Contiguous File Allocation

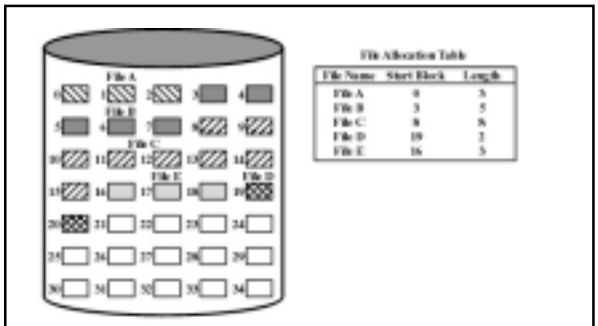


Figure 12.8 Contiguous File Allocation (After Compaction

Methods of File Allocation

- Chained allocation
  - Allocation on basis of individual block
  - Each block contains a pointer to the next block in the chain
  - Only single entry in the file allocation table
    - Starting block and length of file
- No external fragmentation
- Best for sequential files
- No accommodation of the principle of locality

55

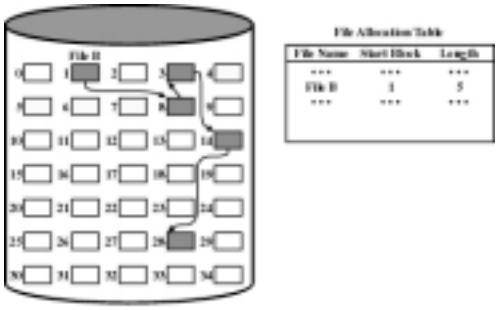


Figure 12.9 Chained Allocation

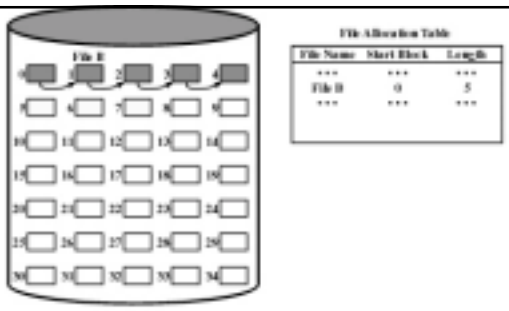


Figure 12.10 Chained Allocation (after consolidation)

Methods of File Allocation

- Indexed allocation
  - File allocation table contains a separate one-level index for each file
  - The index has one entry for each portion allocated to the file
  - The file allocation table contains block number for the index

58

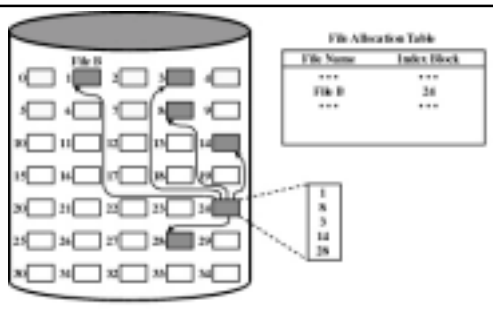


Figure 12.11 Indexed Allocation with Block Portions

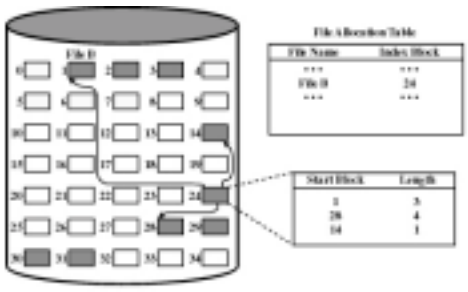


Figure 12.12 Indexed Allocation with Variable-Length Portions

UNIX File Management

- Types of files
  - Ordinary
  - Directory
  - Special
  - Named

61

Figure 12.13 UNIX Block Addressing Scheme

Windows 2000 File System

- Key features of NTFS
  - Recoverability
  - Security
  - Large disks and large files
  - Multiple data streams
  - General indexing facility

63

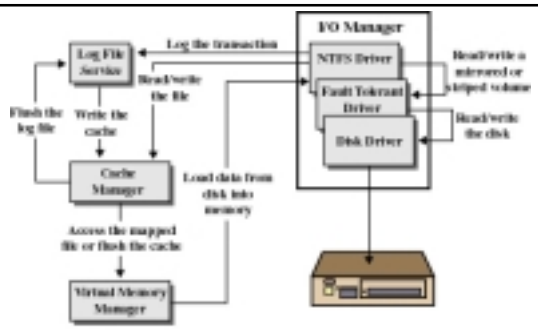


Figure 12.15 Windows NTFS Components [CUST94]