OPERATING SYSTEMS

Subject code : CSC 404

Subject In-charge

Nidhi Gaur

Assistant Professor

email: nidhigaur@sfit.ac.in

Module 5 : File Management

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

File Management

Define File.

- A file is a collection of related information that is mapped onto a secondary storage.

- The information stored in a file is in bits, bytes, lines or records.

- The files provide convenient mechanism to store and retrieve the data and programs.

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.

- These files are mapped to the disks by OS.

Minimal Set of Requirements

- Each user should be able to create, delete, read, and change files

- Each user may have controlled access to other user's files

- Each user may control what type of accesses are allowed to the user's files

- Each user should be able to restructure the user's files in a form appropriate to the problem

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

CONSTITUENTS OF FILE SYSTEM

Three constituents of file system

• File management:

How files are stored, shared and referenced?

• File allocation methods:

Methods to allocate files on the disk space.

• File access methods:

Provides the method to access stored files.

Terms Used with Files

File structure:

- 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

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

File Naming

A file needs to be named so as user can store and retrieve the information. The file name has two parts separated by period(.). First part is name and second is extension.

- Source code file .c
- Object file .obj
- Executable file .exe
- Text file .txt
- Batch file .bat
- Multimedia file .jpeg

- Regular file
- Directory file
- Special file (device file; character special like terminal and block special like disk)

File Access Methods

- The files stored on the disk are required to be retrieved by the user.

- There are many ways to access a file.

- The file access depends on the blocking strategy on the disk and logical structuring of records.
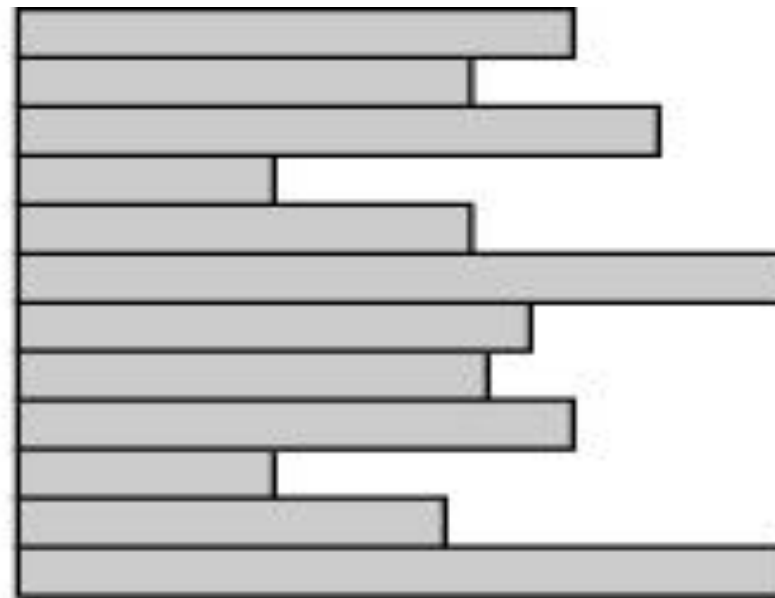
File Organization

- 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

Pile

Variable-length records
Variable set of fields
Chronological order

(a) Pile File

Figure 12.3 Common File Organizations

Sequential-access File

- ## The Sequential File

  - Fixed format used for records

  - Records are of the same length

  - Field names and lengths are attributes of the file

  - One field is the key field

    - Uniquely identifies the record

    - Records are stored in key sequence

    - Key field is usually the first field in the record by which sequential order is maintained.
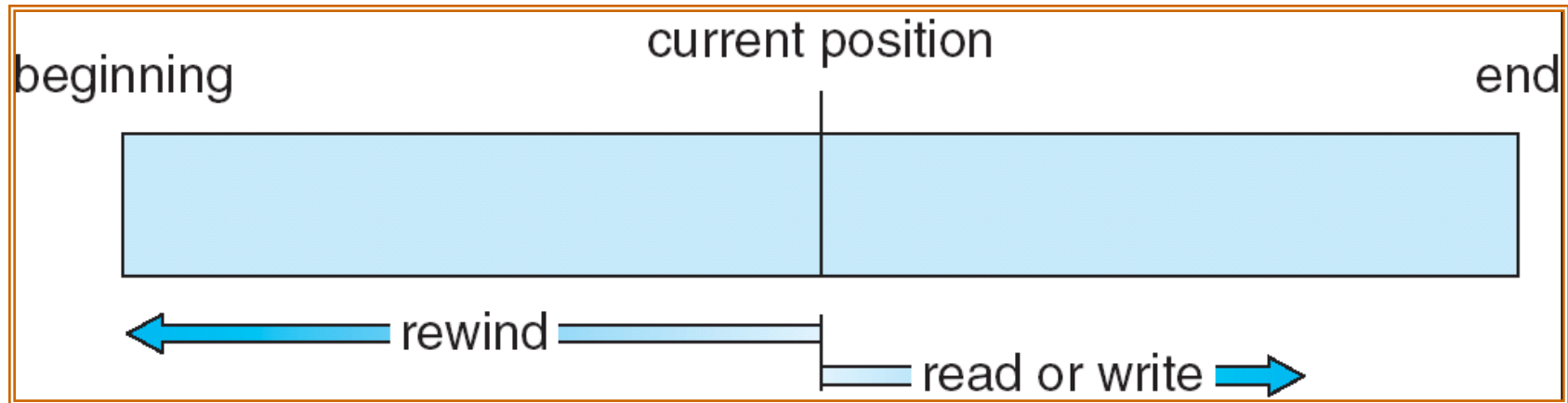
Sequential-access File

- The Sequential File

  – To read or write such file, a file pointer is maintained.

  – Whenever there is request for a read or write, current position is seen and advanced to next position after read or write.

  – The pointer can be reset to the beginning of the file by assigning value zero to it.

Sequential-access File

Sequential-Access File

- **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

Sequential File

Fixed-length records
Fixed set of fields in fixed order
Sequential order based on key field

(b) 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

  • Index is searched to find the desired key value

  • Search continues in the main file at the location indicated by the pointer

 • Instead of locating record sequentially random access approach is used.

File Organization

Indexed Sequential File

- To find a specific field, the index is searched for the highest key value, equal to or less than the desired value.

- The index takes the pointer to the vicinity of a desired record and reduces number of lookups in main file.

Example of Index and Relative Files
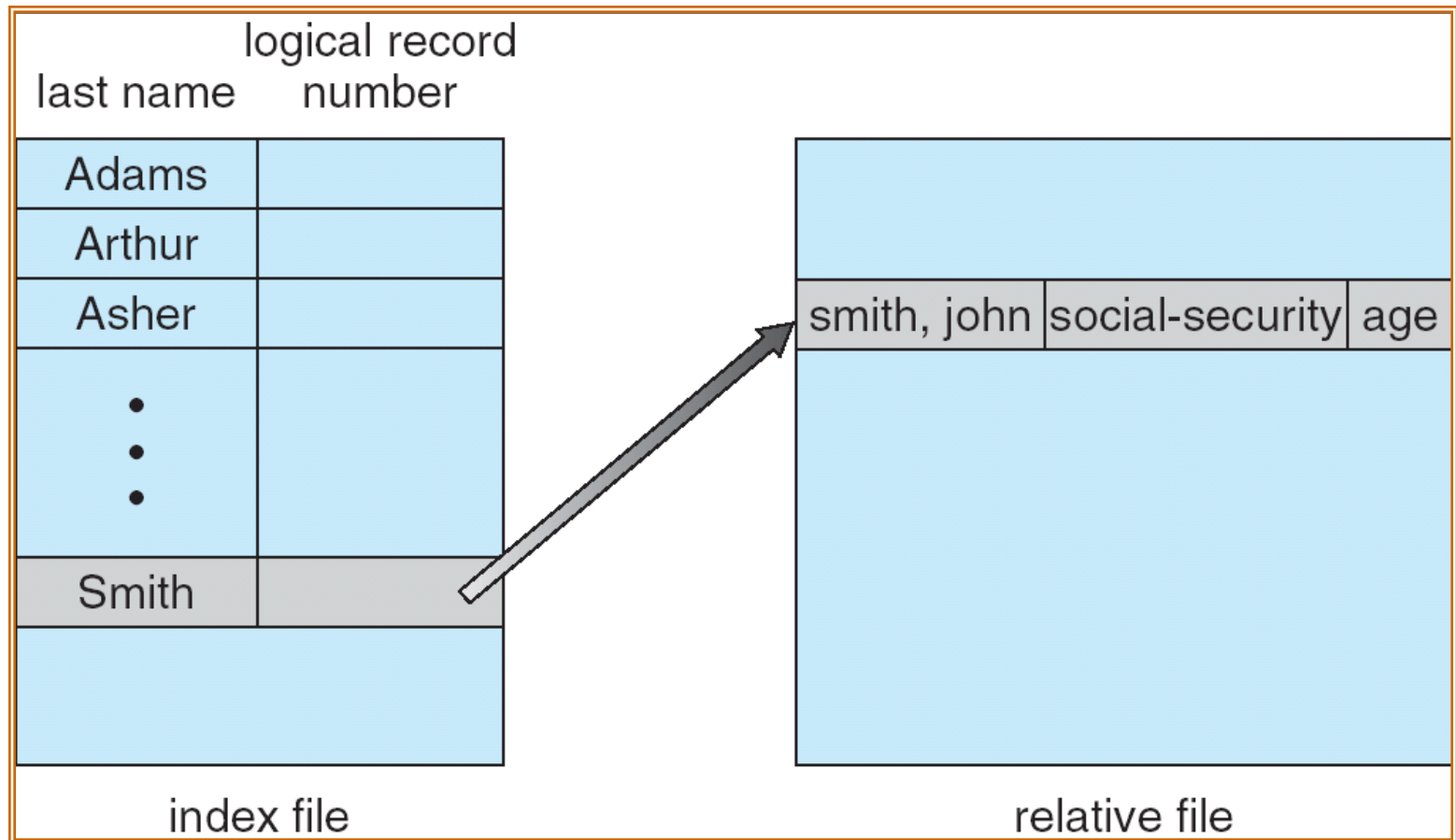
File Organization

- 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

  - The advantage is sequential nature of file and using index time to access a record is greatly reduced.

When multiple fields are combined to form a meaningful collection, it is known as_____

a) File
b) Block
c) record
d) Line

Ans: c) record

_____contains no data, but provides a mechanism that maps physical devices to file names.

a) Regular
b) Directory
c) Special
d) Compressed

Ans: c) Special

_____is a file type used to organize the list of files in a group.

a) Regular
b) Special
c) Object file
d) Directory

Ans: d) Directory

When a group of files is compressed in a single file, it is known as_____.
a) Executable file
b) Batch file
c) Regular file
d) Archive file

Ans: d) Archive file

File Organization

- The Indexed file.

  - It is not necessary to store data sequentially and search a record by its key field.

  - There may be other attributes by which user wants to find a record.

  - When sequential nature and single key is not required, indexed file access is considered.

File Organization

- Indexed File

  - May contain an exhaustive index that contains one entry for every record in the main file

  - May contain a partial index that consists of entries of the records indexed by a particular attribute.

  - Partial indexes determine the index entries based on the specified filter. It includes subset of rows that satisfy the condition.

- CREATE INDEX partial_salary ON employee(age) WHERE salary > 8000;

File Organization

Direct File Access

- Directly access a block at a known address
- By having a block number, block can be accessed directly, instead of a sequential access
- Key field required for each record
- This is meant for random structure of secondary storage like in disks.

File Organization

Direct File Access

- Not necessary to store and access records in sequence but they must be of fixed length.

  – Suitable for applications where rapid access of data from large amount of information is required.
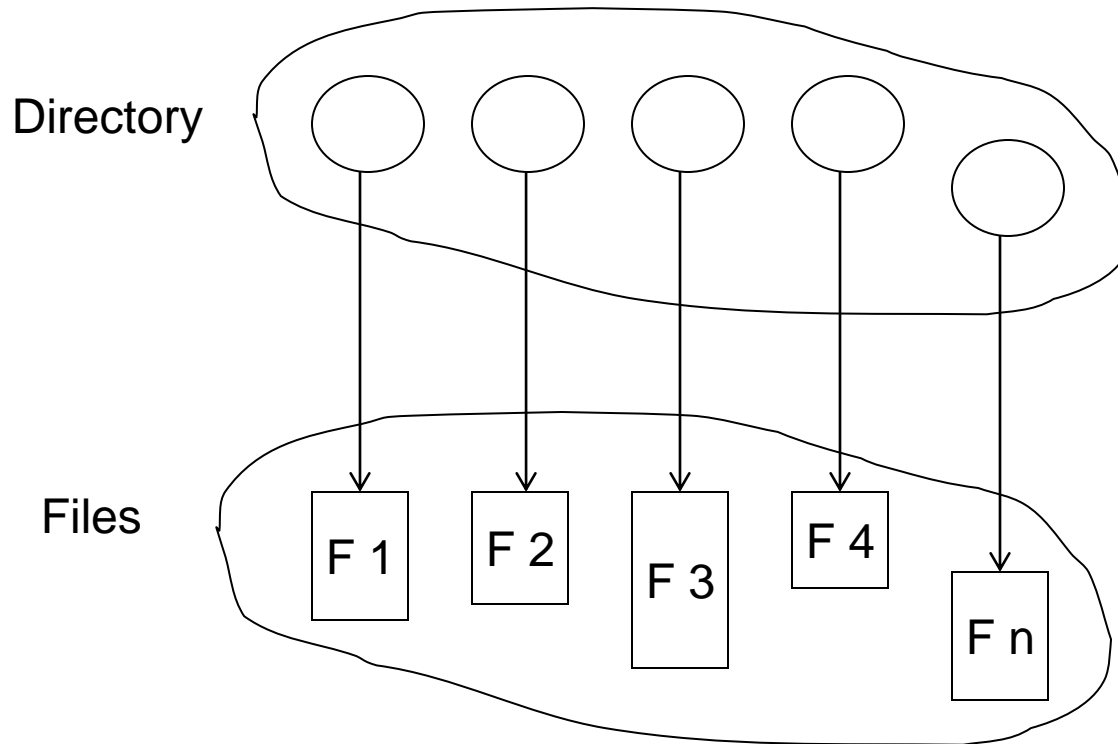
File Directories

- Directories are used to maintain the structure of a file system.

- Contains information about files
  - Attributes (name, type, last modified)
  - Location
  - Ownership

- Directory itself is a file owned by the operating system

Directory Structure

- A collection of nodes containing information about all files

Directory

Files

Both the directory structure and the files reside on disk

Organize the Directory (Logically) to Obtain

- Efficiency – locating a file quickly

- Naming – convenient to users
  - Two users can have same name for different files
  - The same file can have several different names

- Grouping – logical grouping of files by properties, (e.g., all Java programs, all games, …)
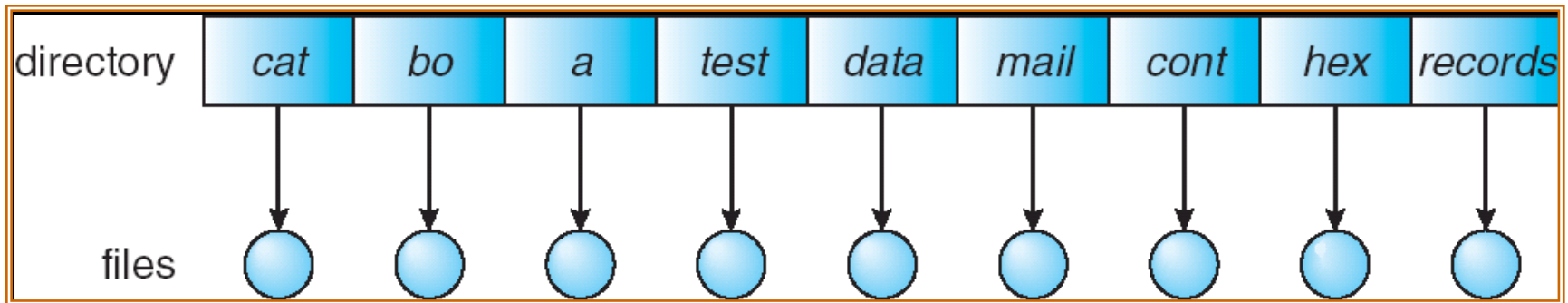
Simple Structure for a Directory

- 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

Single-Level Directory

- A single directory for all users


Naming problem
Grouping problem

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

- When a user starts the master directory is searched for the entry.

- The address of user directory is retrieved, its own directory is searched in master directory.

- Different users may have file with the same name.

Two-Level Directory

- Separate directory for each user
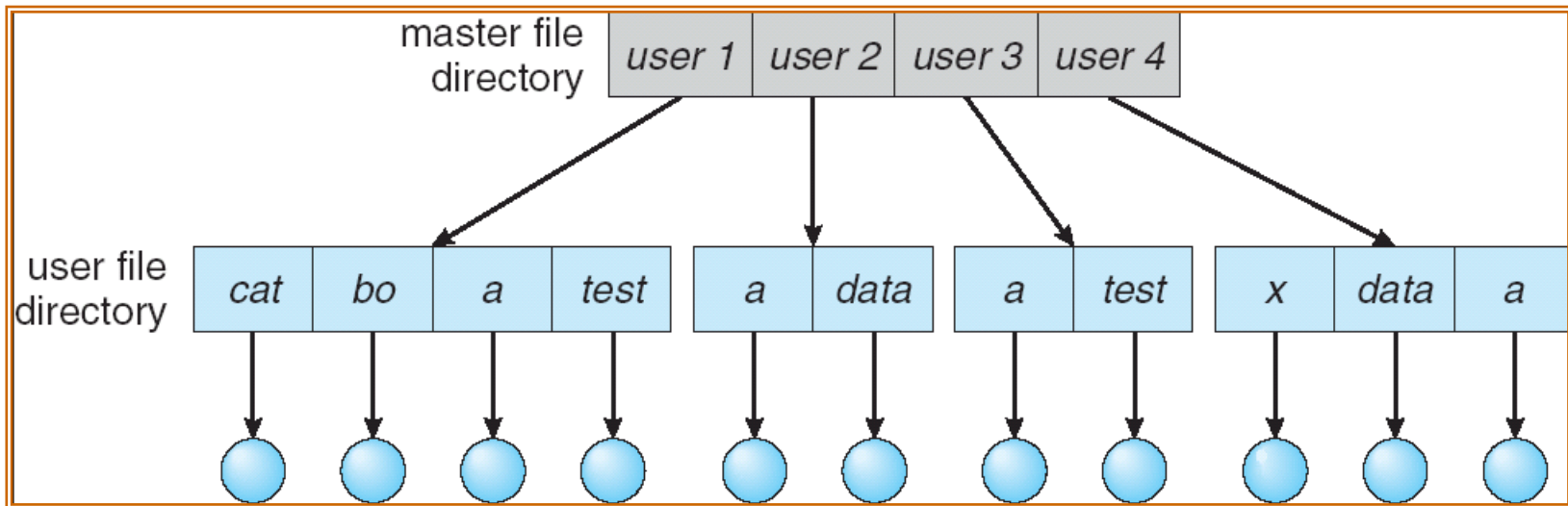
☐ Path name

☐ Can have the same file name for different user

☐ Efficient searching

☐ No grouping capability

Hierarchical, or Tree-Structured Directory

- A directory where there is no limit on subdirectories of a single user is known as tree structure directory.

- Master directory with user directories underneath it

- Each user directory may have subdirectories and files as entries.

- A large set of files can be further divided and grouped.

- A directory has both subdirectory and files.

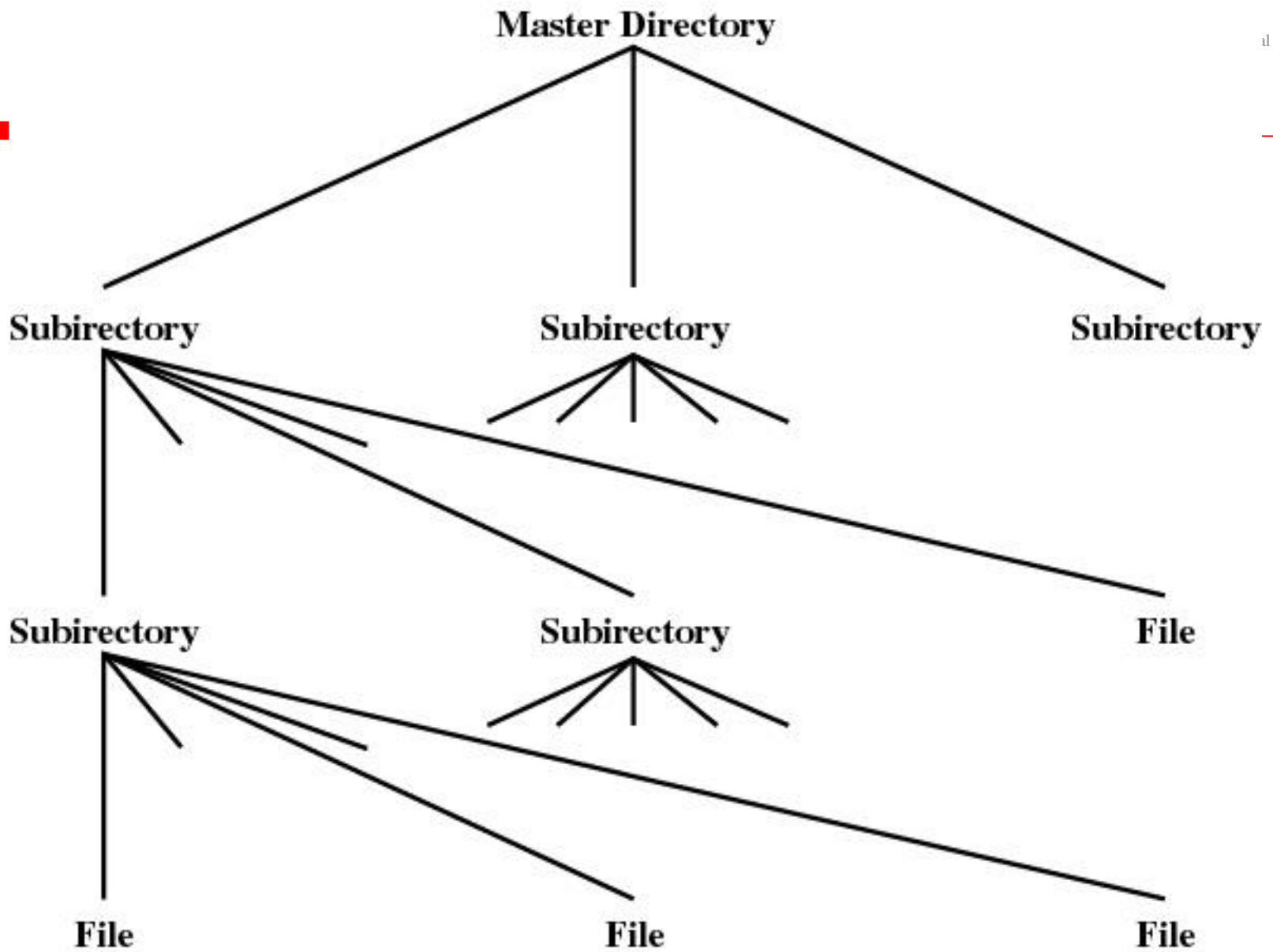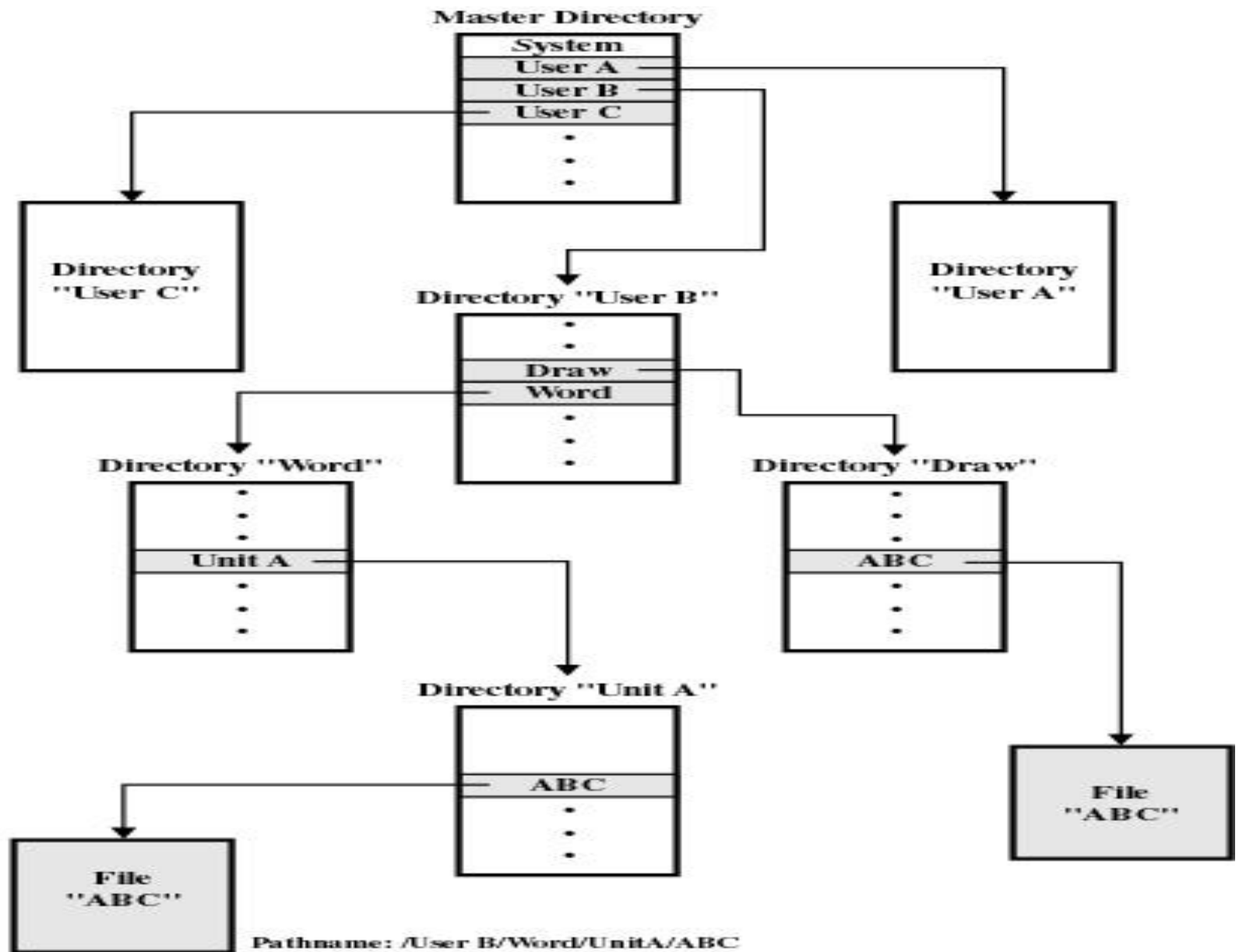- The advantage is a user can access another user's file by using pathname

Figure 12.4 Tree-Structured Directory

Master Directory

System
User A
User B
User C
•
•
•

Directory "User C"

Directory "User B"

•
•
Draw
Word
•
•
•

Directory "User A"

Directory "Word"

•
•
•
Unit A
•
•
•

Directory "Draw"

•
•
•
ABC
•
•
•

Directory "Unit A"

— ABC
•
•
•

File "ABC"

File "ABC"

Pathname: /User B/Word/UnitA/ABC

St. F
Depa **Figure 12.5 Example of 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.

- A pathname is a sequence of one or more path components in the tree structure that uniquely identifies a file.

- There are two types of path: Absolute and Relative

Hierarchical, or Tree-Structured Directory

- Absolute path is from root directory

- To define Relative path One need to know the current directory

- Current directory is the working directory

- Files are referenced relative to the working directory

Methods of File Allocation
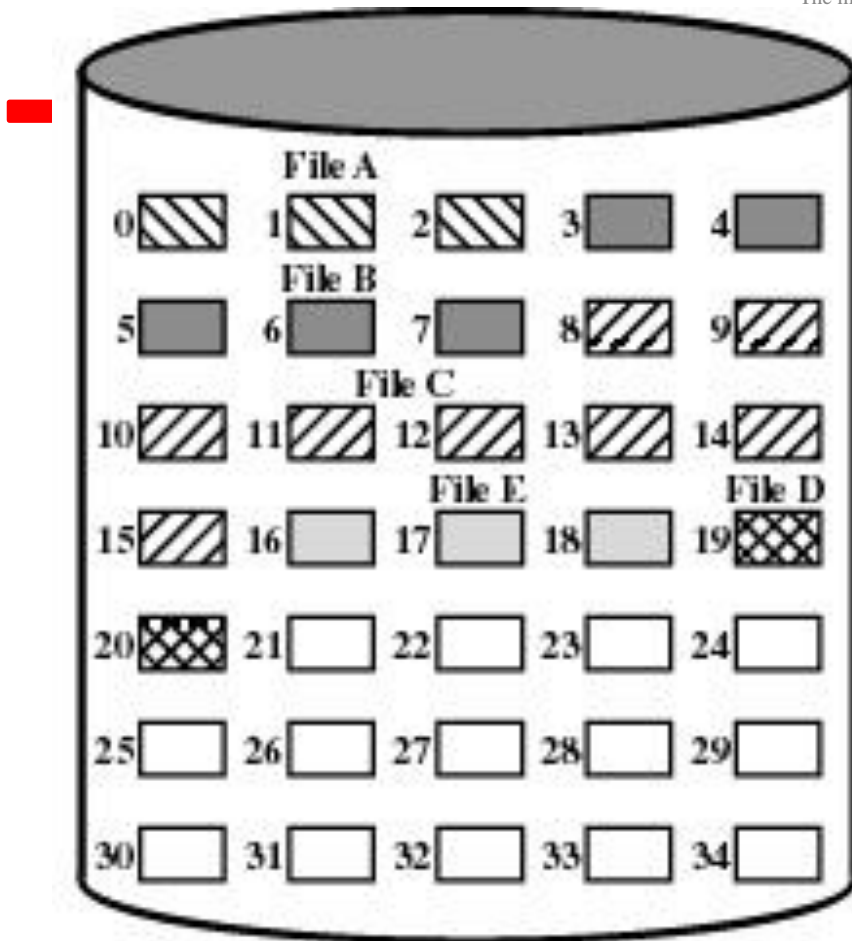
- Contiguous allocation

  – Single set of blocks 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
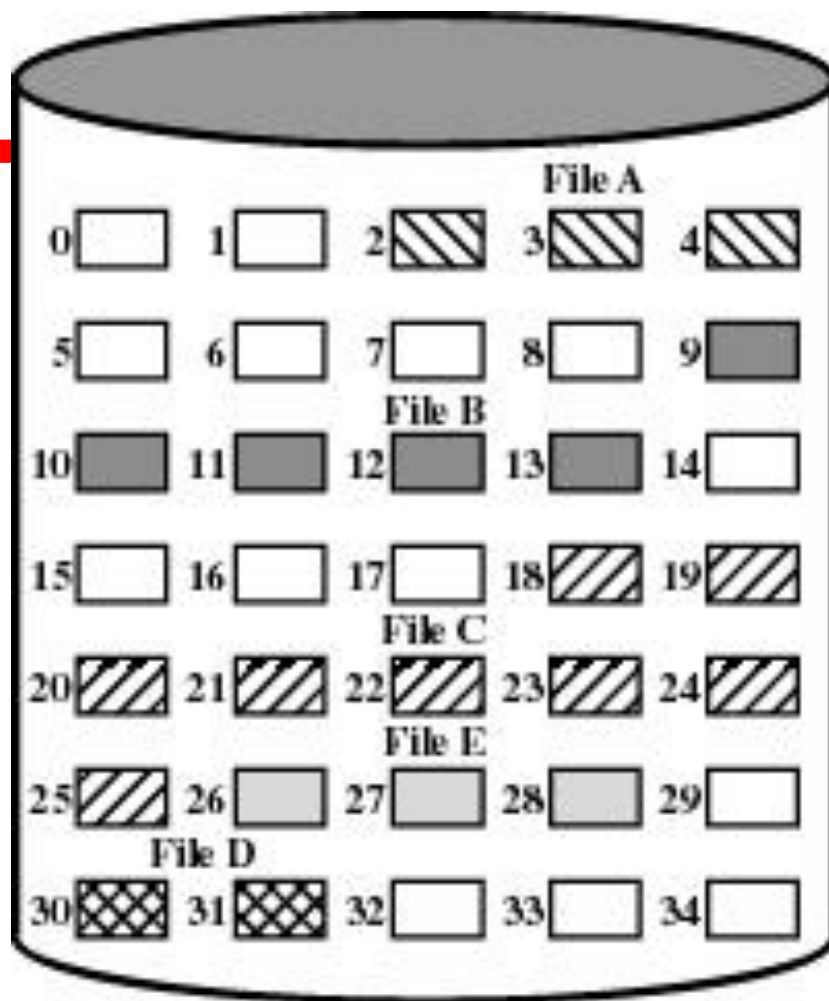
File Allocation Table

File Name	Start Block	Length
File A	2	3
File B	9	5
File C	18	8
File D	30	2
File E	26	3

File A

File B

File C

File E

File D

0 1 2 3 4
5 6 7 8 9
10 11 12 13 14
15 16 17 18 19
20 21 22 23 24
25 26 27 28 29
30 31 32 33 34

Figure 12.7 Contiguous File Allocation

File Allocation Table

File Name	Start Block	Length
File A	0	3
File B	3	5
File C	8	8
File D	19	2
File E	16	3

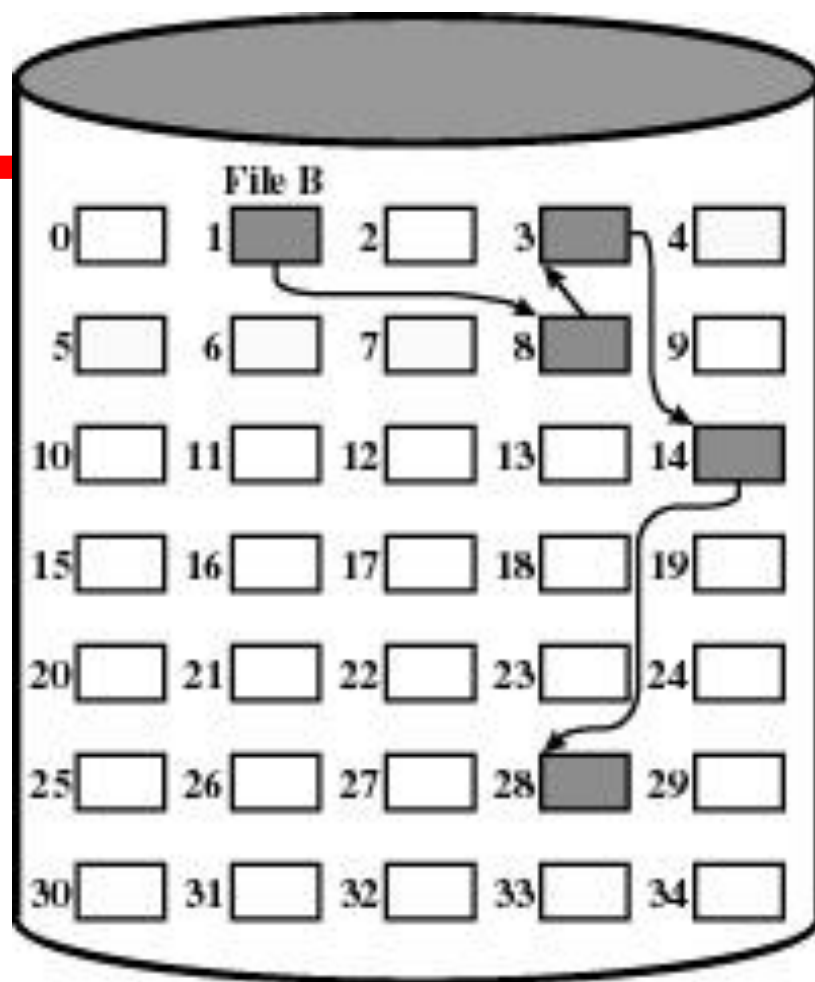**Figure 12.8   Contiguous File Allocation (After Compaction**

Methods of File Allocation

Chained allocation

–Allocation on basis of individual block

–Each block contains a pointer to the next block in the chain

–entry in the file allocation table

- Starting block and length of file

- No external fragmentation

- Best for sequential files

File Allocation Table

File Name	Start Block	Length
...	...	...
File B	1	5
...	...	...

Figure 12.9   Chained Allocation

Nidhi Gaur

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

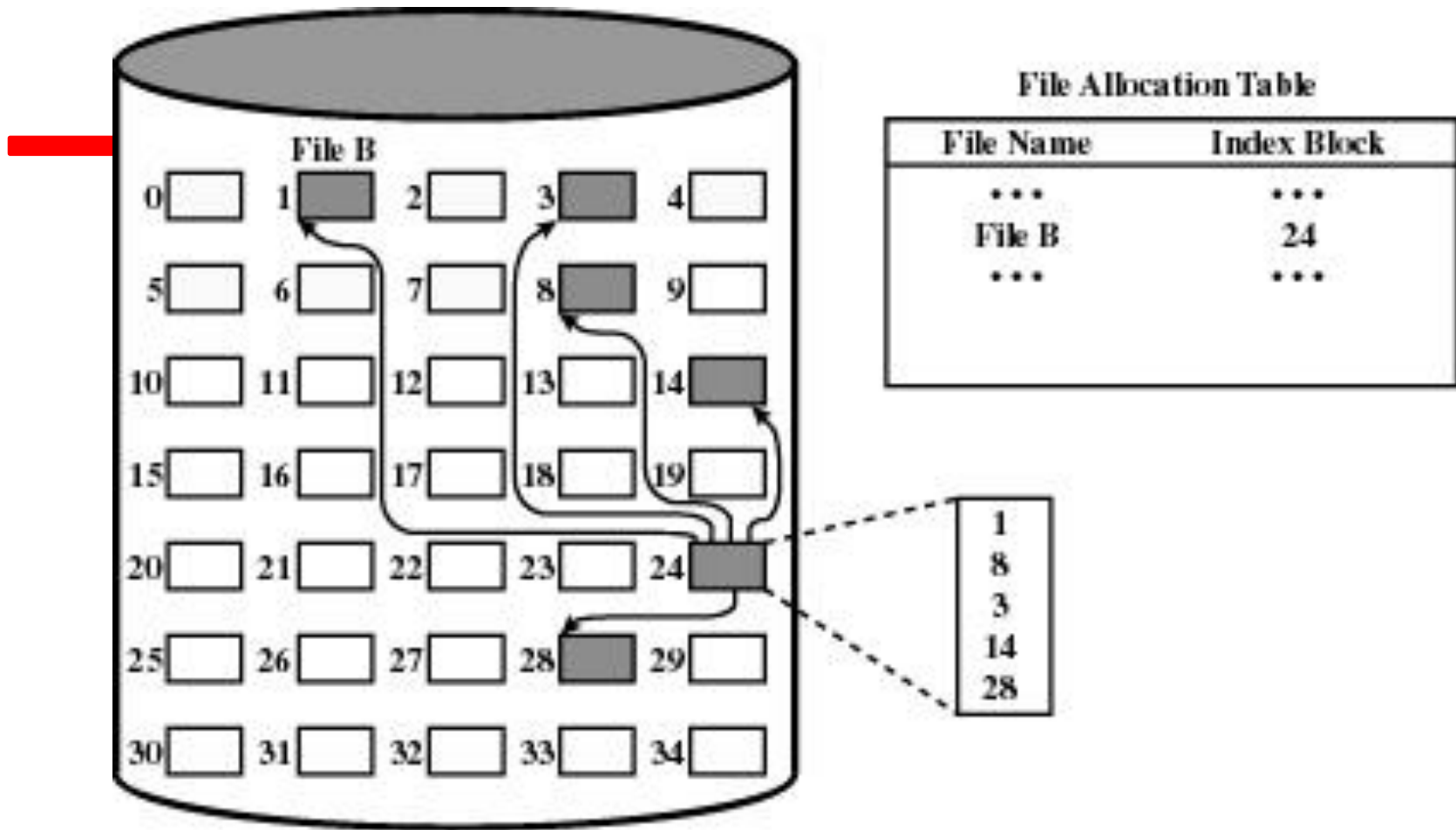– The file allocation table contains block number for the index

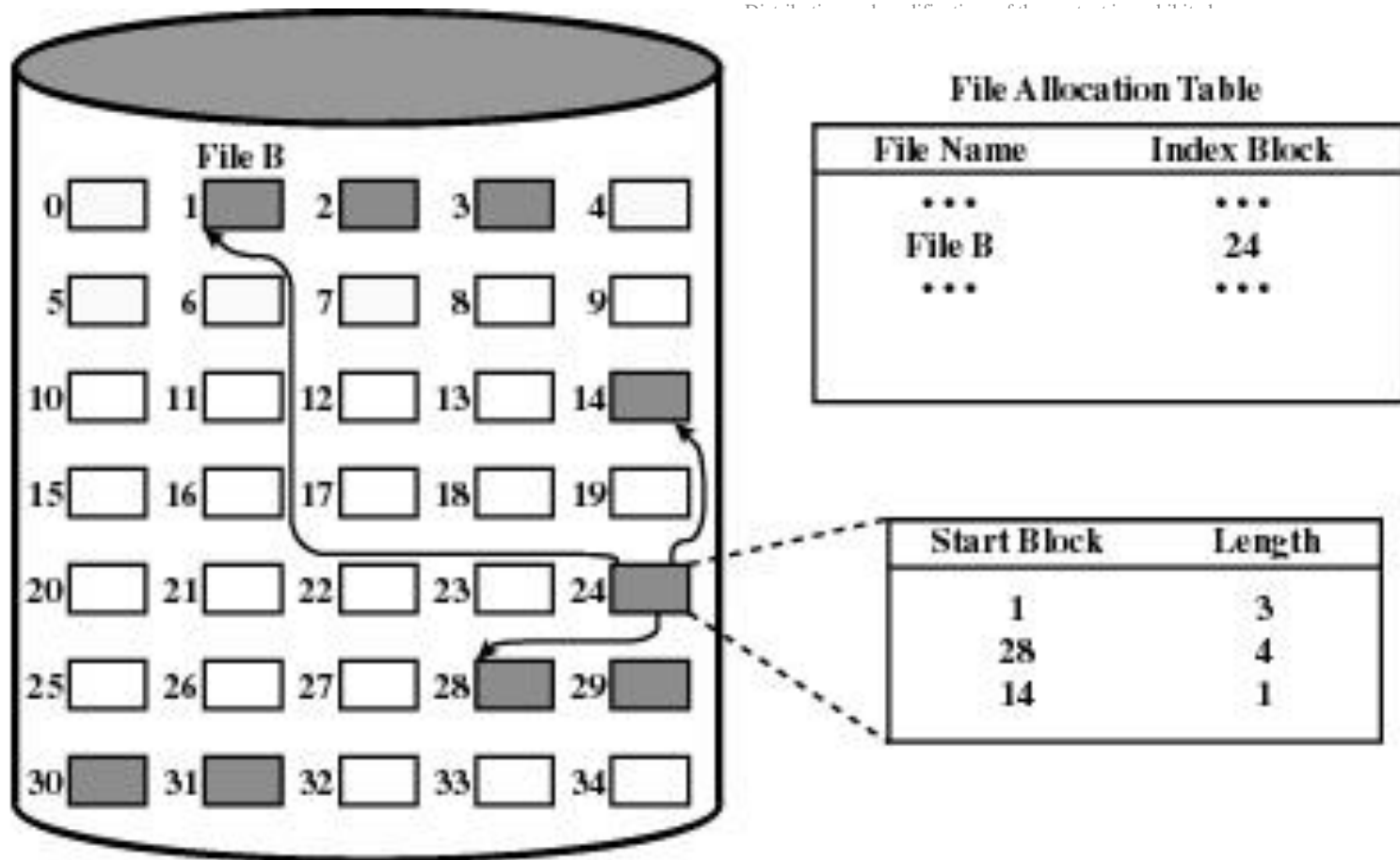**Figure 12.11 Indexed Allocation with Block Portions**

Figure 12.12 Indexed Allocation with Variable-Length Portions

Q: The projects A and B need to have a file named Utility, but with different contents. Can both projects use the same file name? If yes, which directory structure would be best for this case?

Ans: Two level or Tree structure

Q: Can a two level directory structure be considered as a tree structure directory?
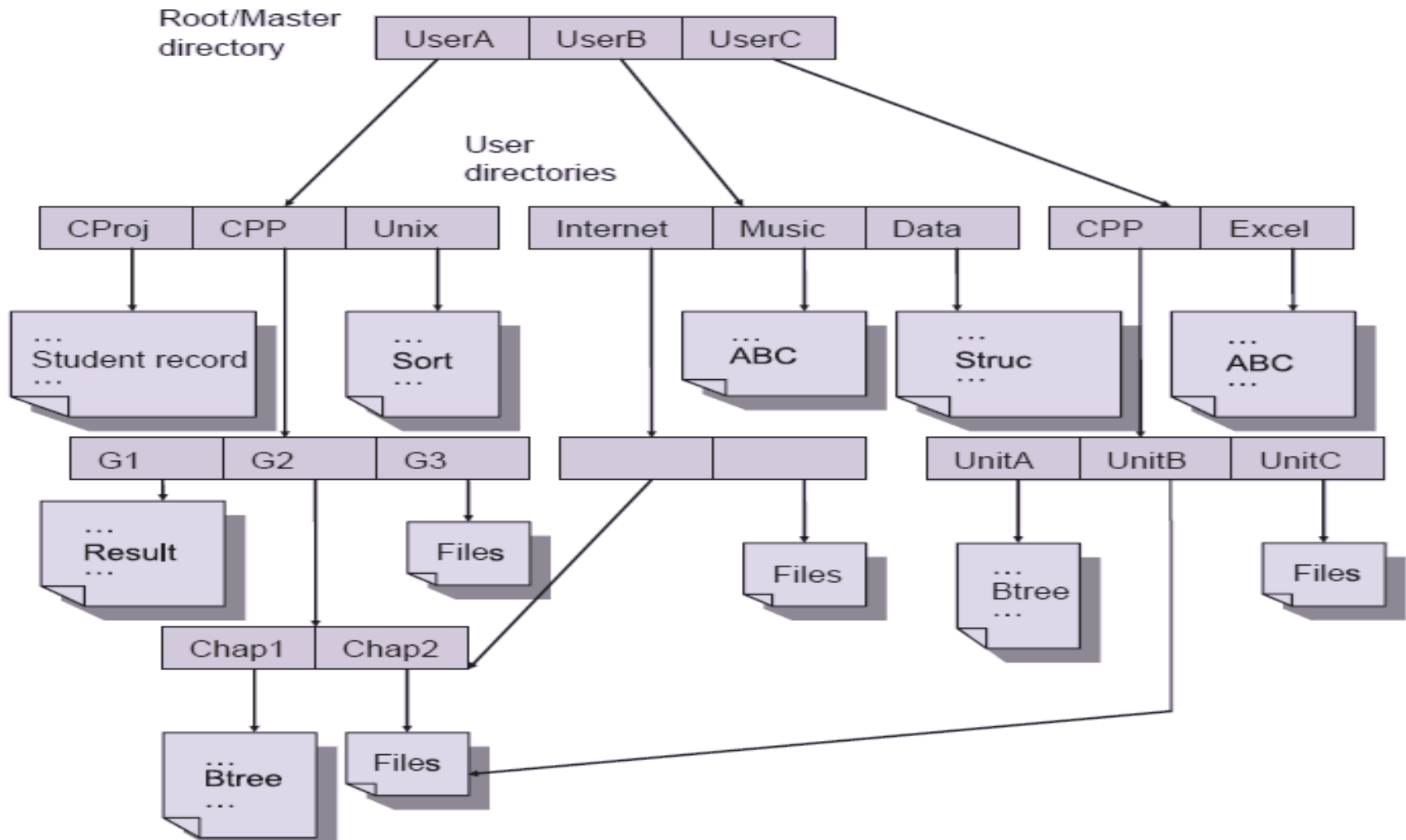
Ans: Yes with only two levels

Q: A and B are working on project X and Y by maintaining separate directories. However there is a utility which may be used by both. Which type of directory structure would be best for this case?

Ans: Acyclic Graph structure

Acyclic Graph Structure/ File Sharing

Acyclic Graph structure

- The tree structure does not allow sharing of file or directory among various users.

- Two users cannot have same file in their directory.

- Sharing must not be confused with accessing.

- Sharing means having the same file or sub directory in the directory of concerned user.

- If there is change in contents of shared file then it must be visible to all shared users.

Acyclic Graph structure

- Implementation of sharing violates the tree structure.

- This type of structure is an acyclic graph.

- An acyclic graph is implemented through a pointer, known as link.

- A link is a pointer that points to a file or directory.

File Sharing

File sharing two types:

- Sequential

- Concurrent

File System in Linux

- The data held in files is kept in data blocks.

- These data blocks are all of the same length and the block size of a particular EXT2 file system is set when it is created
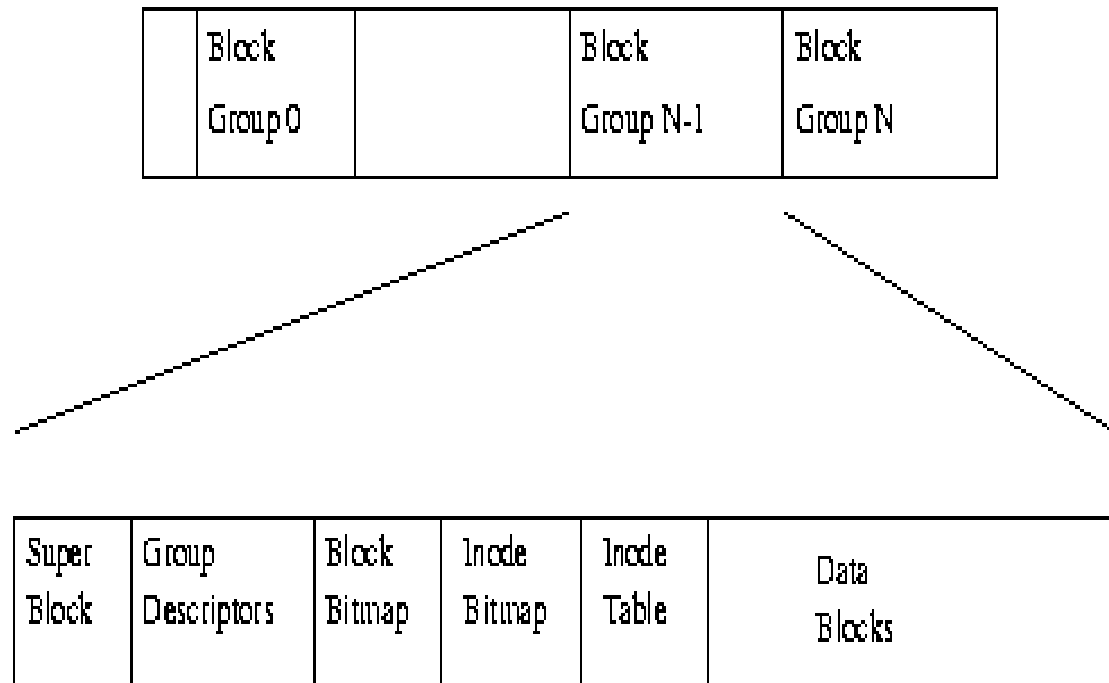
File System in Linux

- EXT2 defines the file system topology by describing each file in the system with an inode data structure.

- An inode describes which blocks the data within a file occupies as well as the access rights of the file, the file's modification times and the type of the file.

- Every file in the EXT2 file system is described by a single inode and each inode has a single unique number identifying it.

- The inodes for the file system are all kept together in inode tables.

EXT2 FILE SYSTEM (LINUX)

	Block Group 0		Block Group N-1	Block Group N

Super Block	Group Descriptors	Block Bitmap	Inode Bitmap	Inode Table	Data Blocks

EXT2 FILE SYSTEM (LINUX)

- The layout of the EXT2 file system is occupying a series of blocks in a block structured device.

- So far as each file system is concerned, block devices are just a series of blocks that can be read and written.

- A file system does not need to concern itself with where on the physical media a block should be put, that is the job of the device's driver.

- Whenever a file system needs to read information or data from the block device containing it, it requests that its supporting device driver reads an integral number of blocks.

- The EXT2 file system divides the logical partition that it occupies into Block Groups.

St. Francis Institute of Technology
Department of Computer Engineering

Operating System
Nidhi Gaur

File Types – Name, Extension

file type	usual extension	function
executable	exe, com, bin or none	ready-to-run machine-language program
object	obj, o	compiled, machine language, not linked
source code	c, cc, java, pas, asm, a	source code in various languages
batch	bat, sh	commands to the command interpreter
text	txt, doc	textual data, documents
word processor	wp, tex, rtf, doc	various word-processor formats
library	lib, a, so, dll	libraries of routines for programmers
print or view	ps, pdf, jpg	ASCII or binary file in a format for printing or viewing
archive	arc, zip, tar	related files grouped into one file, sometimes com-pressed, for archiving or storage
multimedia	mpeg, mov, rm, mp3, avi	binary file containing audio or A/V information

End of Chapter