

Session 8 | File Management

Learning Outlines:

1. Describe different types of files
2. Describe different techniques of File allocation method
3. Explain the concept of file management
4. Compare different file allocation techniques

Files

Definition --> Data collections created by user.

Desirable properties of files:

- Long-term Existence:
Doesn't disappear after user logged off, stored on disk/secondary storage
- Sharable between processes: Sharable with controlled permit.
- Structure: Can be organized into hierarchical structure to represent the relationships between files

File Management System:

- Must support multiple storage devices
- Must have standardized set of I/O
- Make sure files are valid
- I/O Support for multiple users

Minimal User Requirements:

There should be a method to back up files.

Access files by names, names are easier to remember compared to numbers.

Reliable

File Organization Types:

1. The Indexed Files

This organization type uses an index that contains pointers to the actual data stored in specific locations in the file system, allowing for quick access to records by key. Unlike other methods, it facilitates both sequential and random access, improving search performance but at the cost of additional storage for the index.

2. Direct or hashed file:

Direct files utilize a hash function that converts a record's key into a unique address, enabling immediate access to records. This method differs from others by providing the fastest access for non-sequential keys but can suffer from issues like hashing collisions and requires a good hash function to distribute records evenly.

3. The pile:

A pile file is a simple unstructured file where records are collected in no particular order. This is in stark

contrast to indexed or sequential methods, as it offers the fastest write times with no overhead but is inefficient for read operations unless the entire file is processed.

4. Sequential Pile:

Records are stored in a specific sequential order based on a key field, making it efficient for processing large volumes of records in order. Unlike the pile or direct files, sequential files are optimized for batch processing and require reorganization if new records are to be inserted in order.

5. Indexed sequential file:

Combines the sequential access of sequential files with an index to allow faster searching, thus supporting both sequential and direct access. This dual approach differs from pure sequential files by allowing quicker individual record retrieval and from indexed files by maintaining a sequential order, usually requiring periodic re-indexing to maintain efficiency.

Two-Level Scheme

Two-Level means there is one directory for each user and a master.

Master Directory has an entry for each user directory providing address and access control information.

Each user directory is a simple list of the files of that user.

Names must be unique only within the collection of files of a single user.

File system can easily enforce access restriction on directories.

PreAllocation vs Dynamic Allocation, and Contiguous

PreAllocation --> not really efficient, therefore it is not used anymore. YOU allocate certain amount of space before installing an application.

Dynamic Allocation

Allocates space accordingly, as needed.

Contiguous Allocation:

Gampang kena Fragmentation. Files load more faster.

Chained Allocation:

Blocks doesn't need to be contiguous anymore. Slower, have to read the first block to find the next block. Robust to fragmentation.

Indexed Allocation:

No fragmentations. Much faster in terms of searching. Setbacks: Sacrifice one block for an "index block" (e.g, need 6 blocks, for 5-blocks length files --> 5 for files, 1 for index block)

Free Space Management

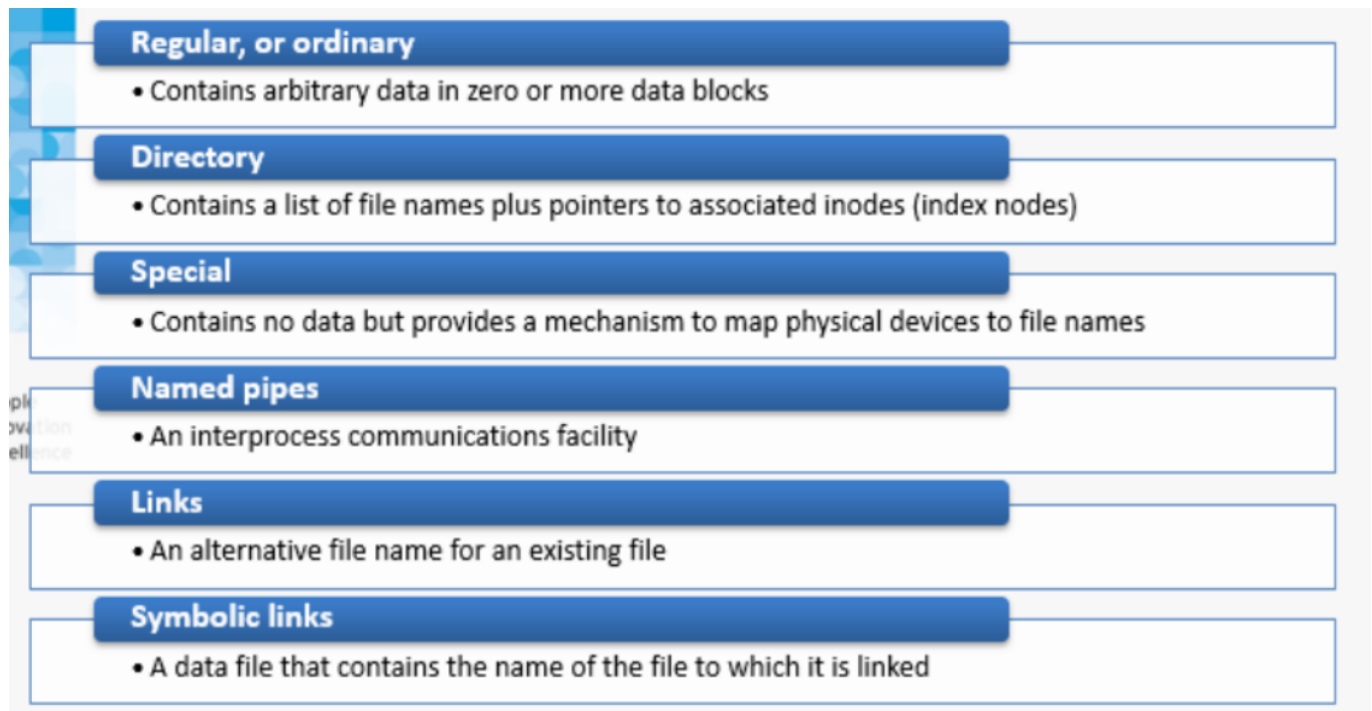
Keep track where the blocks are empty, ready to be saved with data.

Bit Tables

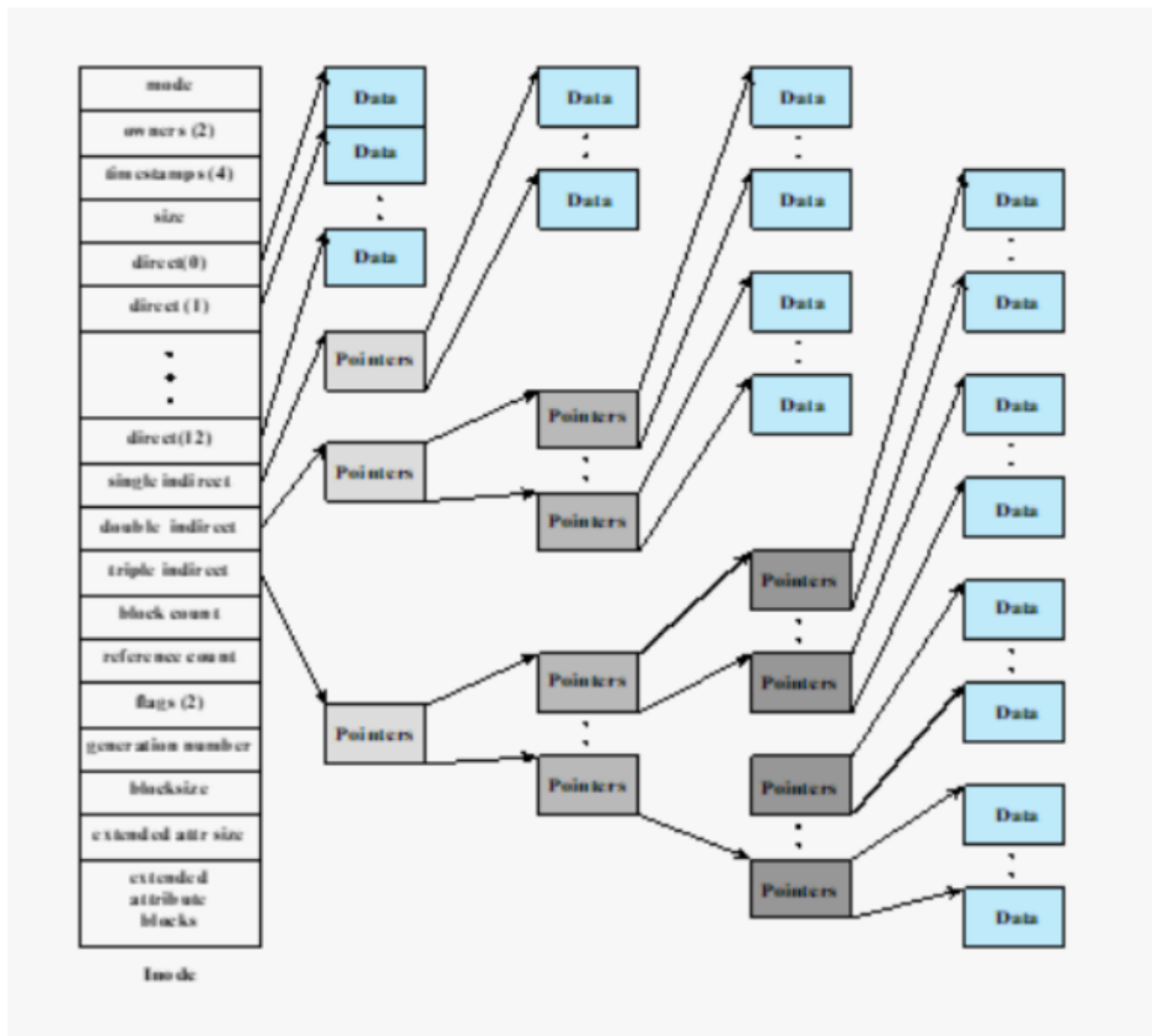
faster to allocate which blocks are empty.

Example of Bit Tables: 11001100111 --> 1 Occupied, 0 is empty

UNIX File Management System



Pipes refers to the output of a program which later be used as an input in another programs (See semaphore for more explanations).



Single Indirect -> pointer to files.

Double Indirect -> Pointer > Pointer > Files

Triple Indirect -> Pointer > Pointer > Pointer > Files

Windows File System

- NTFS -> New Technology File System
- Sector -> Smallest physical storage unit on the disk
- Cluster -> One or more contiguous sectors
- Volume -> Consists one or more cluster