Goals for Today

- Learning Objective:
  - Learn about directory structures
  - Run through some disk performance exercises
- Announcements, etc:
  - C4: I removed 2 papers from your reading list
    - No longer need to read Multics Security Eval or SELinux
  - MP3 is out! Due **April 18th**.

Reminder: Please put away devices at the start of class

CS 423
Operating System Design: Directory Structures & Disk Performance

Professor Adam Bates

Spring 2018

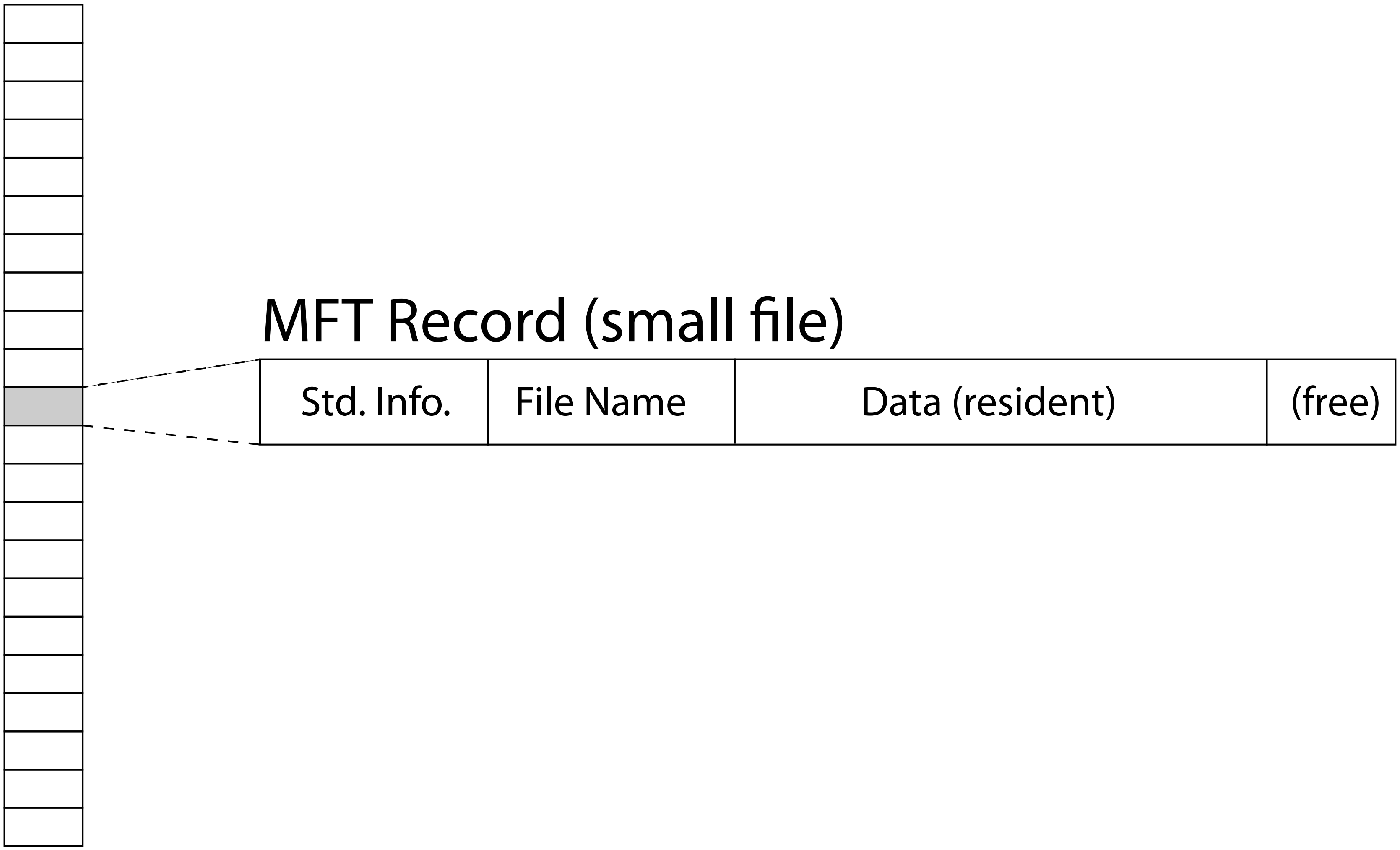
NTFS

- Master File Table
  - Flexible 1KB storage for metadata and data
- Extents
  - Block pointers cover runs of blocks
  - Similar approach in linux (ext4)
  - File create can provide hint as to size of file
- Journalling for reliability

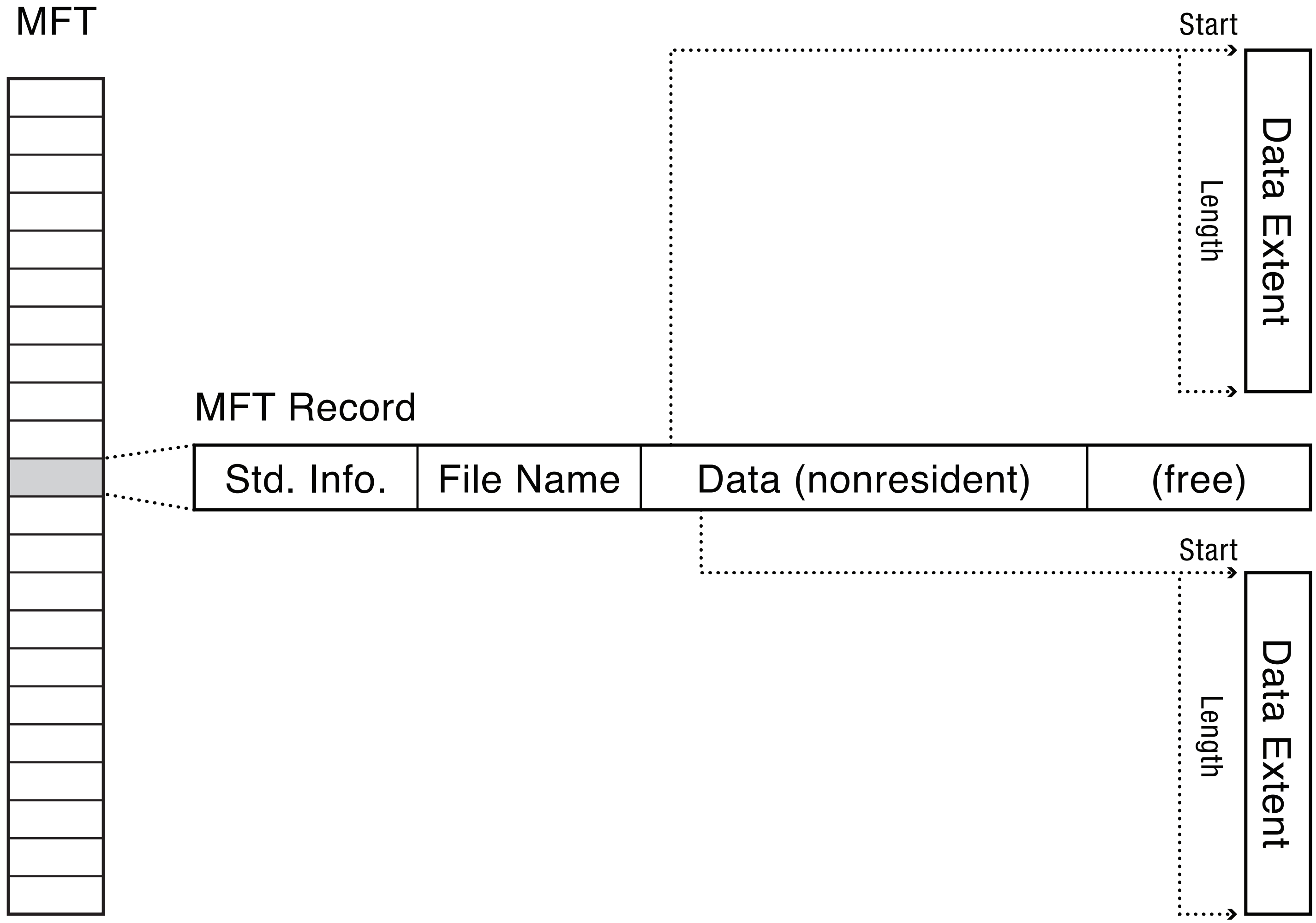
NTFS

Master File Table

MFT Record (small file)

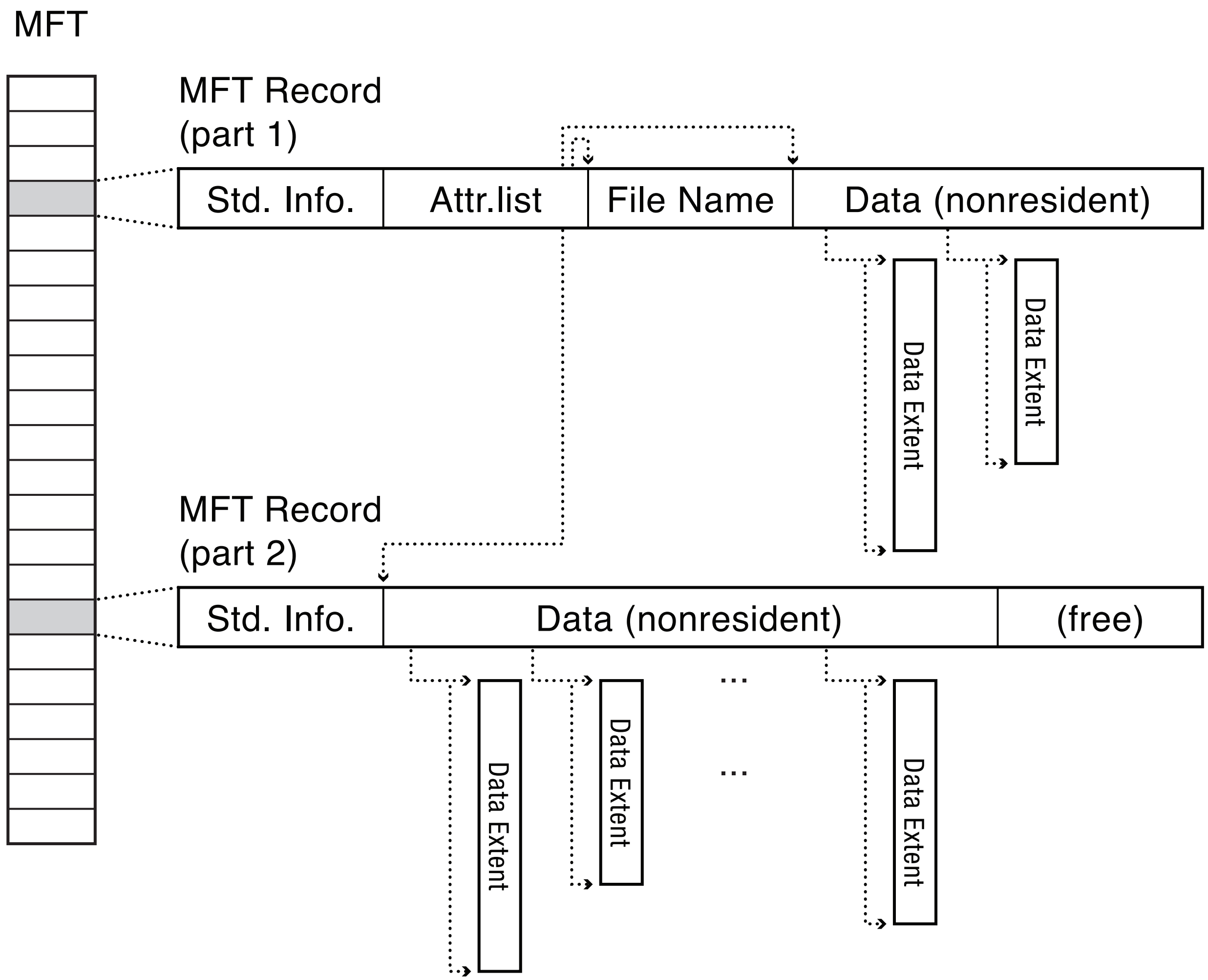


NTFS

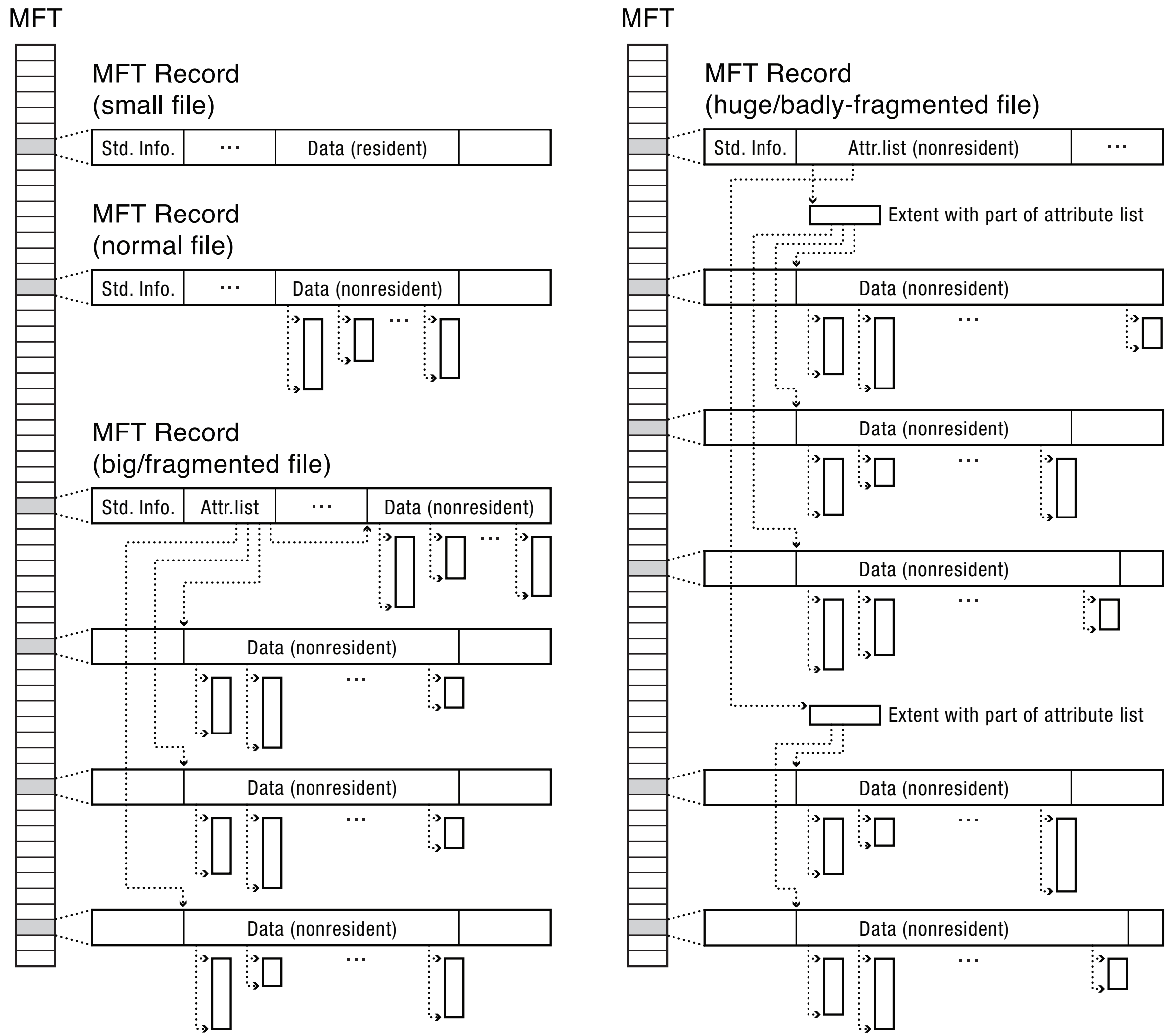


MFT
MFT Record
Std. Info.
File Name
Data (nonresident)
(free)
Start
Length
Data Extent
Start
Length
Data Extent

NTFS Indirect Block



NTFS

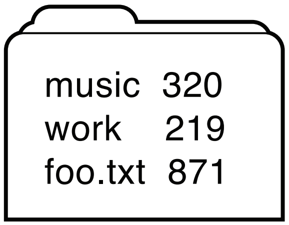

MFT
MFT Record
(small file)
Std. Info.
...
Data (resident)
MFT Record
(normal file)
Std. Info.
...
Data (nonresident)
MFT Record
(big/fragmented file)
Std. Info.
Attr.list
...
Data (nonresident)
Data (nonresident)
Data (nonresident)
Data (nonresident)
MFT
MFT Record
(huge/badly-fragmented file)
Std. Info.
Attr.list (nonresident)
...
Extent with part of attribute list
Data (nonresident)
Data (nonresident)
Data (nonresident)
Extent with part of attribute list
Data (nonresident)
Data (nonresident)

Named Data

file name offset $\xrightarrow{\text{directory}}$ *file number offset* $\xrightarrow{\text{index structure}}$ *storage block*

Directory Structures

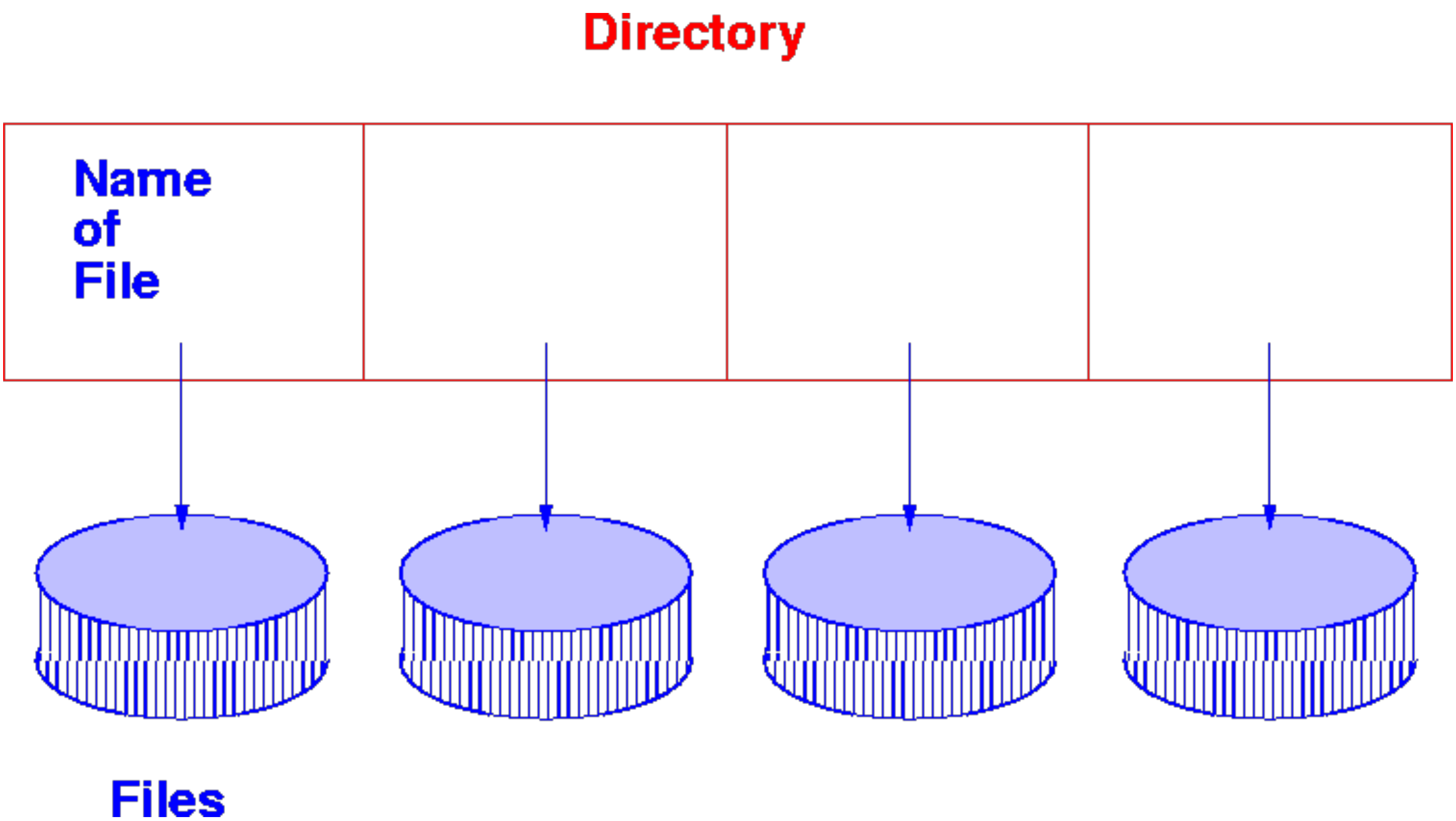
- maps symbolic names into logical file names
  - search
  - create file
  - list directory
  - backup, archival, file migration
- Directories are also files!



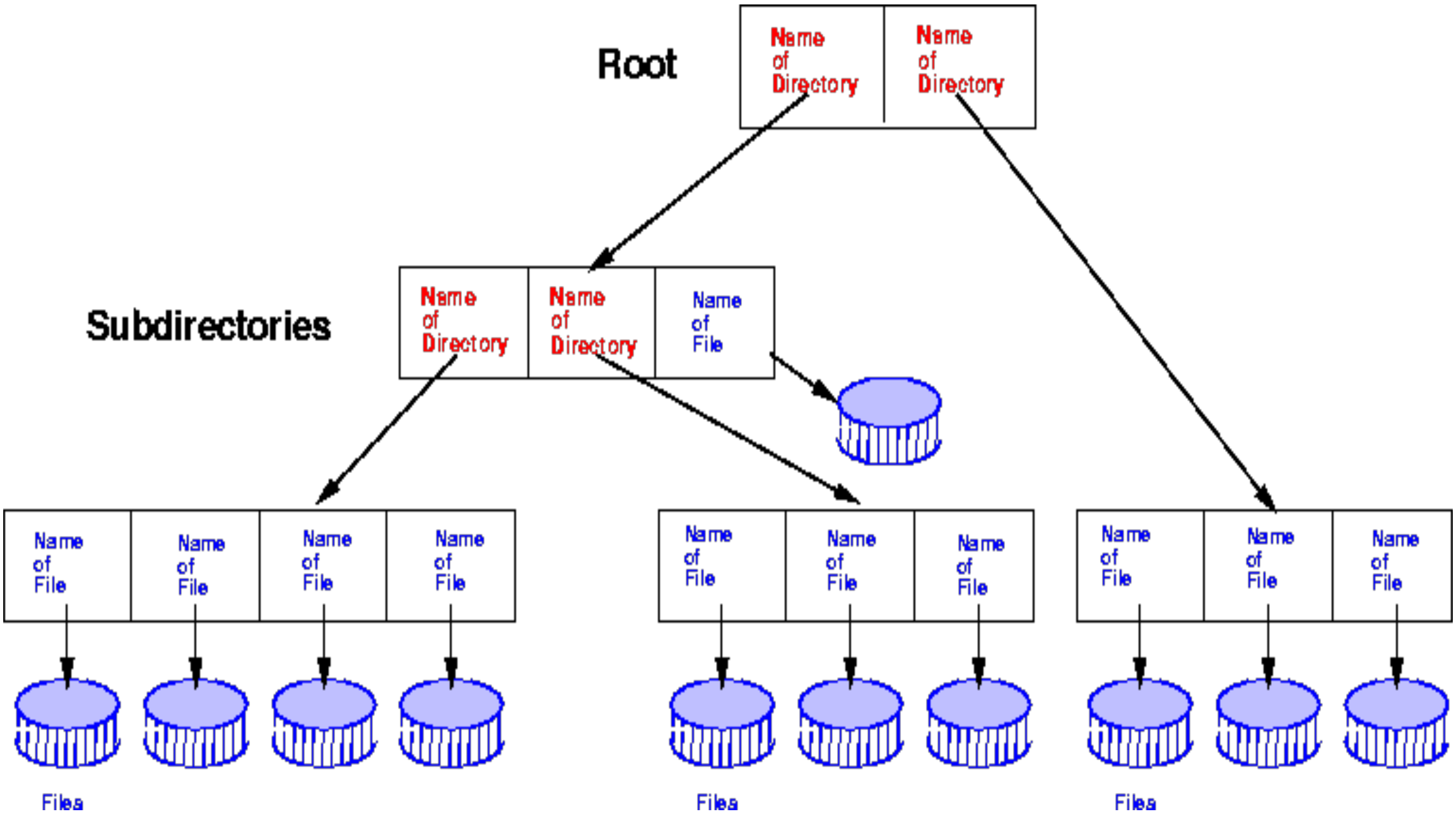
Directory Internals

- Directories are also files! Special files
  - Denoted by “File Type” in inode
  - Hold access paths to the files in the file system
- They have data blocks (inodes) on disk
- Enables support for very large directories
- The partition superblock points to the inode of the root directory.

Single-level Directory

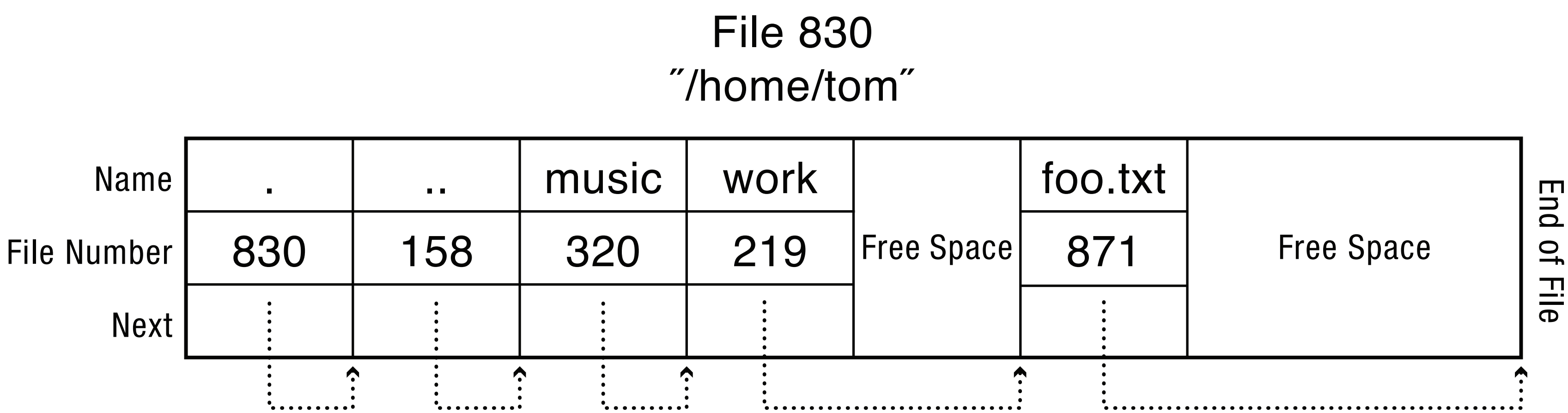


Tree-Structured Directories



Directory Layout

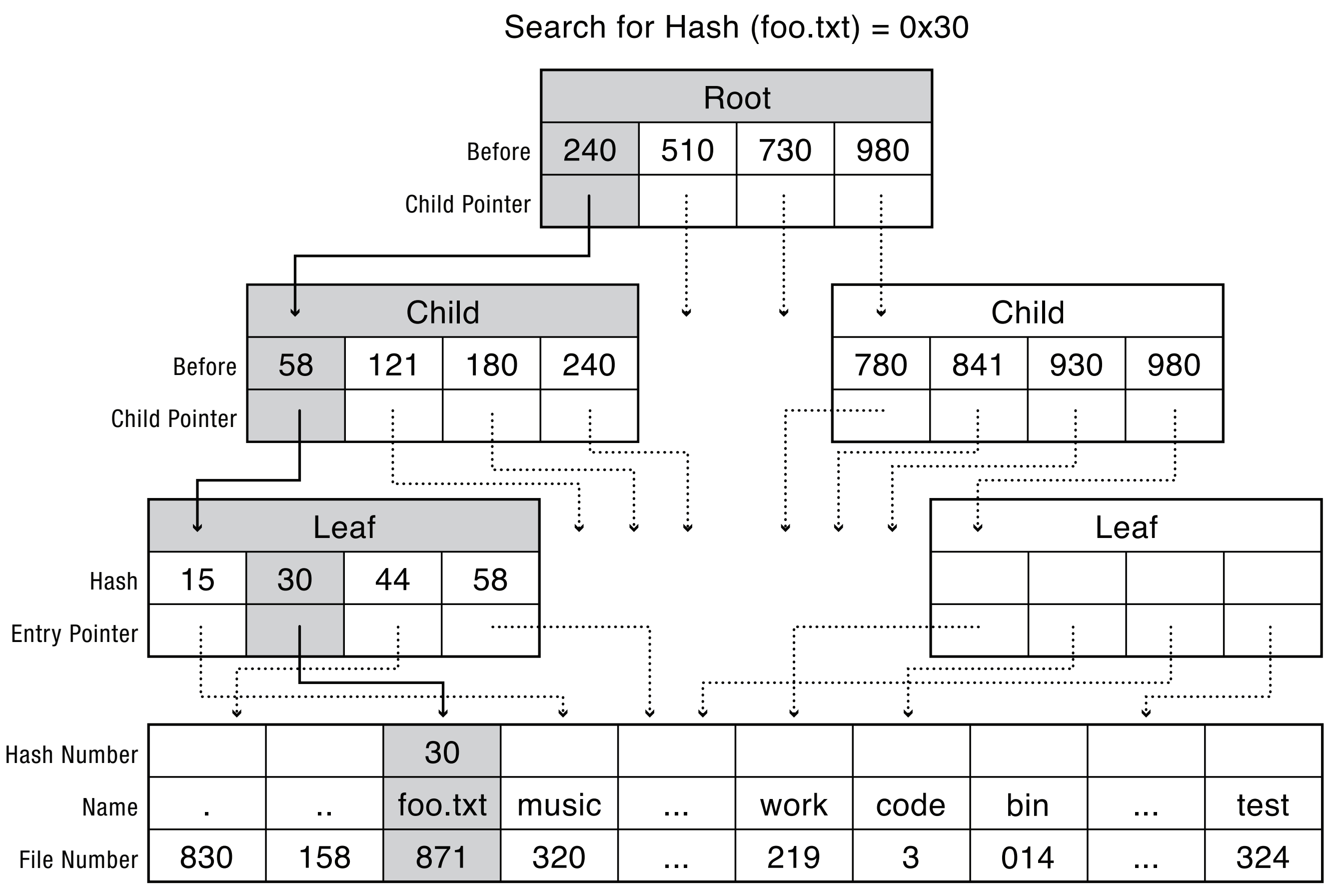
- Represent directory as a list of files
- Linear search to find filename
- Suitable for small directories



What can we do to improve efficient in large directories?

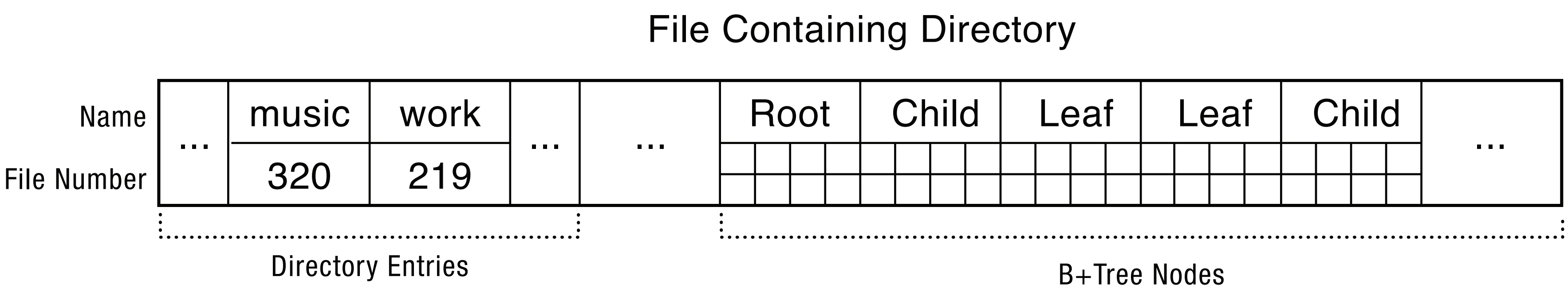
B Trees

- Logarithmic search to find filename
- Suitable for large directories

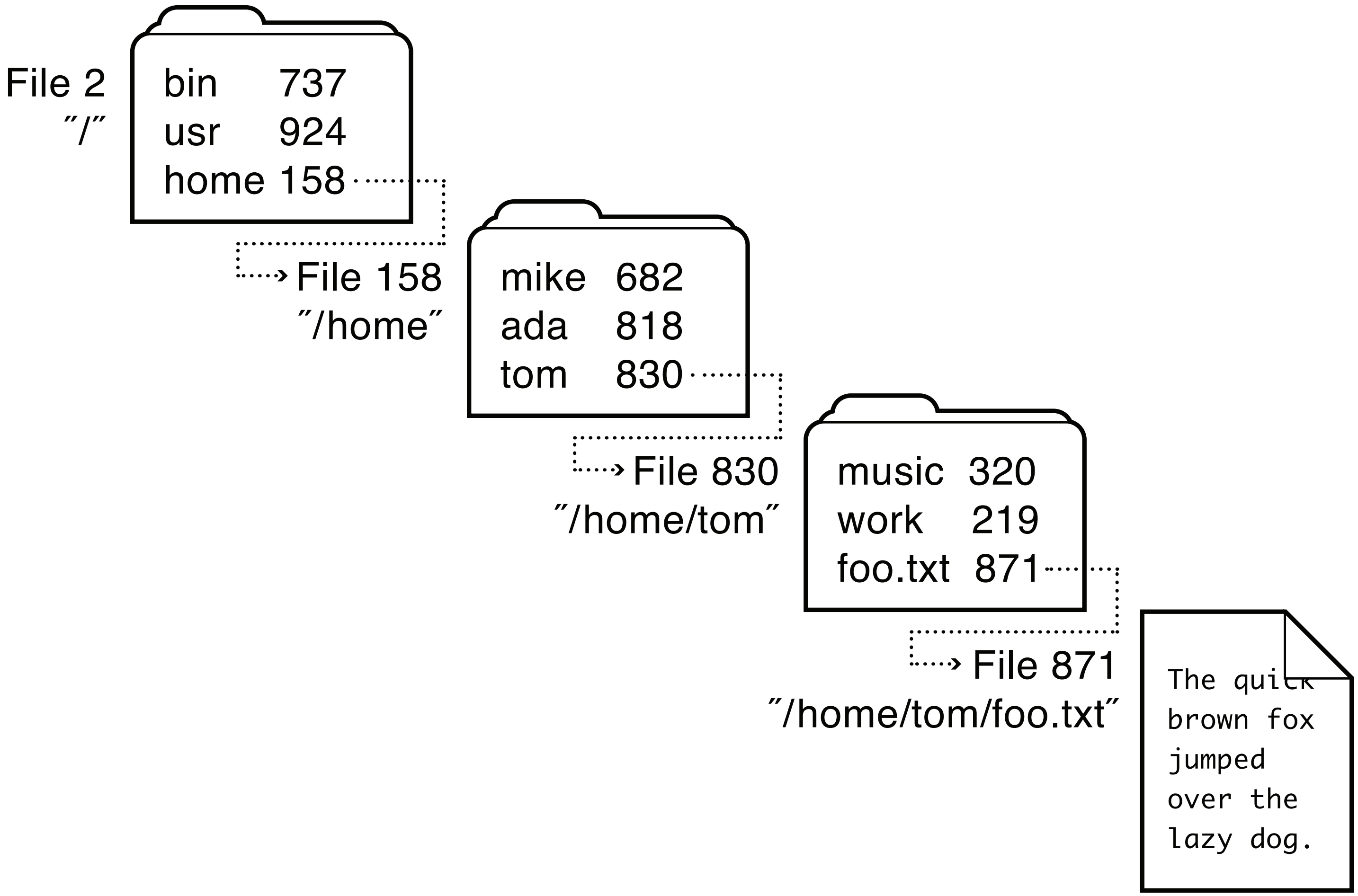


B Trees

- Logarithmic search to find filename
- Suitable for large directories



Recursive Filename Lookup

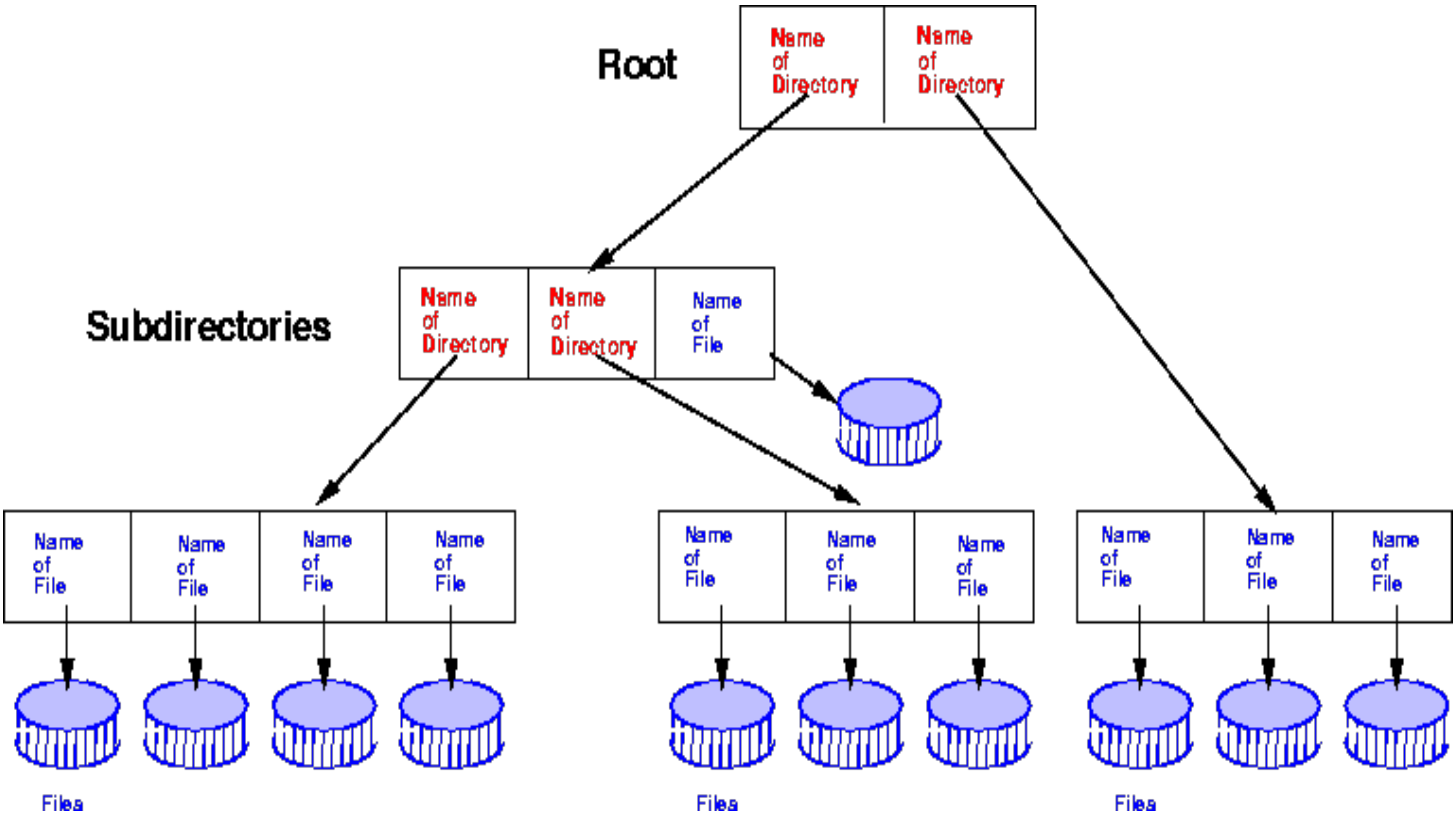


Tree-Structured Directories

- arbitrary depth of directories
- leaf nodes are files
- interior nodes are directories
- path name lists nodes to traverse to find node
- use absolute paths from root
- use relative paths from current working directory pointer

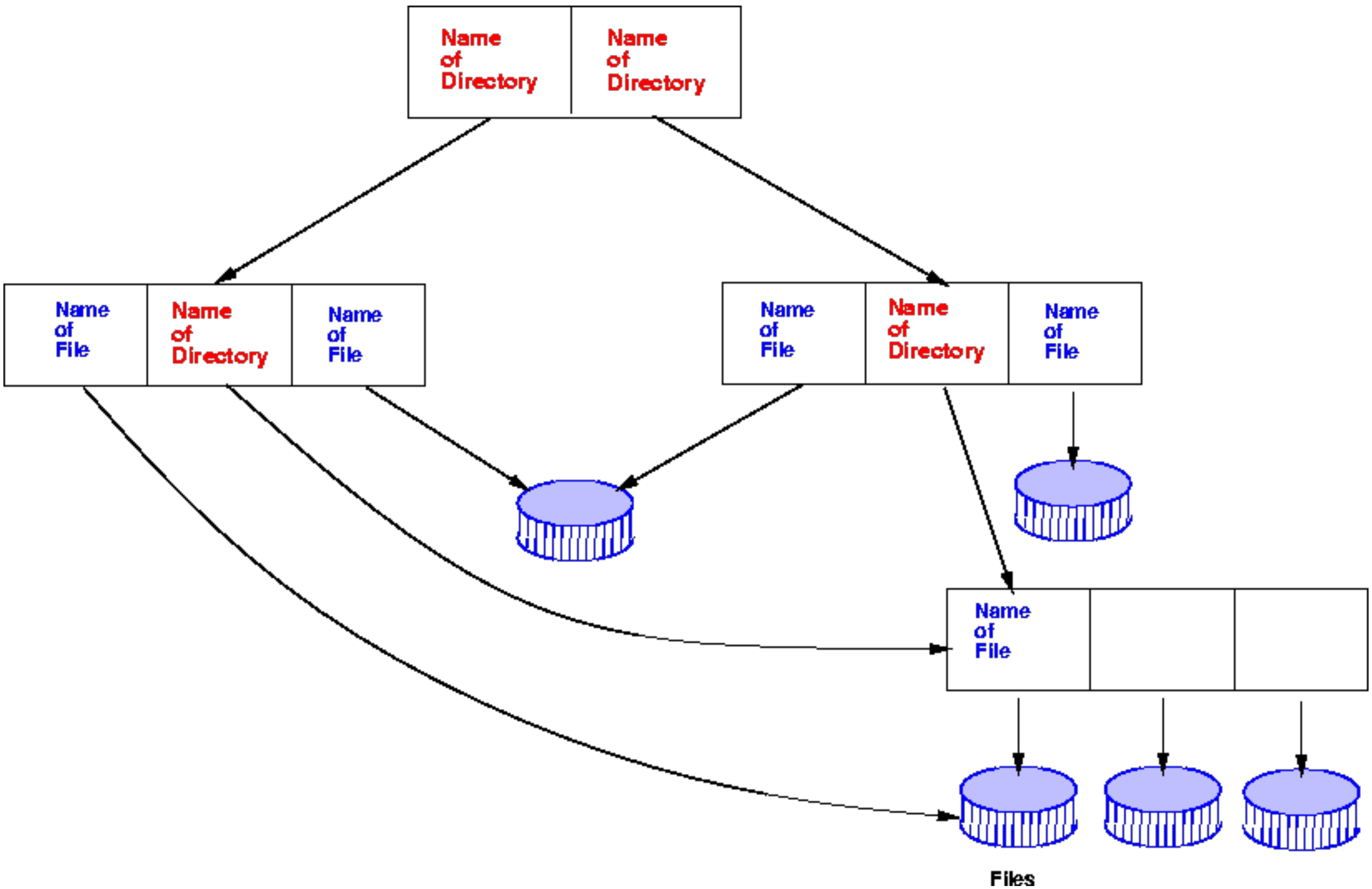
Tree-Structured Directories

- We usually think of directories as trees…



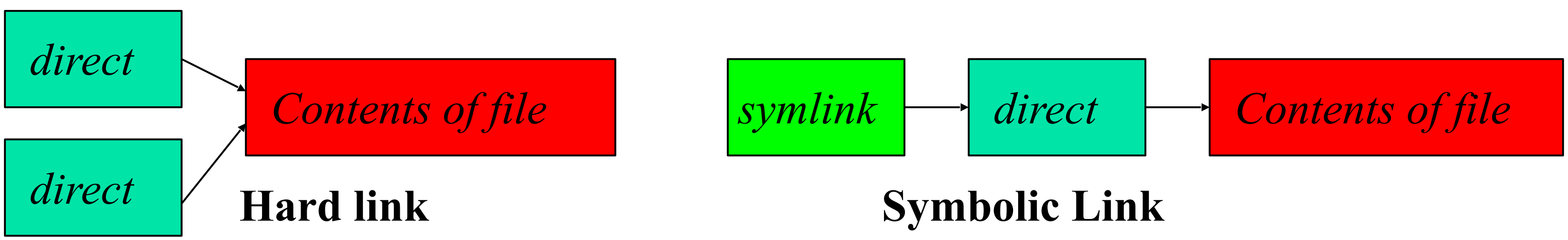
Acyclic Graph Structured Dir's

- But in practice they're actually acyclic graphs!



Symbolic Links

- **Symbolic** links are different than regular links (often called **hard links**). Created with **ln -s**
- Can be thought of as a directory entry that points to the name of another file.
- Does not change link count for file
  - When original deleted, symbolic link remains
- They exist because:
  - Hard links don’t work across file systems
  - Hard links only work for regular files, not directories



Changing Gears

Some notes on disk performance…

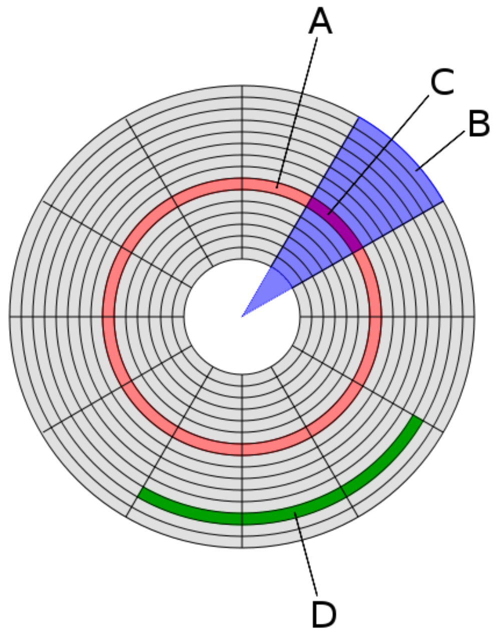
Typical Modern Spec's

- Disk rotation speed: 5,000-15,000 RPM
- Number of sectors per track: 500-2000
- Number of tracks: 100,000-200,000
- Average seek latency: 2ms
- Block size: 0.5K-4K
- Multiple zones (layout is different for inner and outer tracks)
- Multiple bays (stacked drives)

Estimated Sustained Average Transfer Rate

- Suppose that a disk drive spins at 7200 RPM (revolutions per minute), has a sector size of 512 bytes, and holds 160 sectors per track.
- What is sustained average transfer rate of this drive in megabytes per second?

A: Track.
B: Sector.
C: Sector of Track.
D: File



Estimated Sustained Average Transfer Rate

- Suppose that a disk drive spins at 7200 RPM (revolutions per minute), has a sector size of 512 bytes, and holds 160 sectors per track.
- What is sustained average transfer rate of this drive in megabytes per second?

Disk spins 120 times per second (7200 RPM/60)

Each spin transfers a track of 80 KB (160 sectors x0.5K)

Sustained average transfer rate is 120x80 = 9.6MB/s.

Average Performance of Random Access

- Suppose that a disk drive spins at 7200 RPM (revolutions per minute), has a sector size of 512 bytes, and holds 160 sectors per track.
- Average seek time for the drive is 8 milliseconds
- Estimate # of random sector I/Os per second that can be done and the effective average transfer rate for random-access of a sector?

Disk spins 120 times per second
Average rotational cost is time to travel half track: 1/120 * 50%=4.167ms

Transfer time is 8ms to seek
+ 4.167 ms rotational latency
+ 0.052 ms (reading one sector takes 0.0005MB/ 9.6MB).
= 12.219 ms

# of random sector access/second = 1/0.012219=81.8

Effective transferring rate: 0.5 KB * 81.8= 0.0409 MB/s.

Flash Memory

- Electronically Erasable Programmable Read Only Memory (EEPROM)
- Example specifications (NAND Flash):
  - Page size: 2KB (approx.)
  - Block size: 64 pages (128KB)
  - Device size: 16K blocks
- Random READ: 25μs, Sequential READ: 25ns
- WRITE performance
  - PROGRAM PAGE: 220μs, BLOCK ERASE: 1.5ms
- Endurance: 100,000 PROGRAM/ERASE cycles