

Operating Systems

Assignment Project Exam Help

<https://powcoder.com>

Lecture 9b

Add WeChat powcoder

Previously

1

File systems

- File systems implementation
- Kernel data structures
- Block allocation

Assignment Project Exam Help

<https://powcoder.com>

Add WeChat powcoder

File System Management

- File systems
- Data consistency
- Performance
- Virtual file systems

Assignment Project Exam Help

<https://powcoder.com>

Add WeChat powcoder

Recap: Questions

3

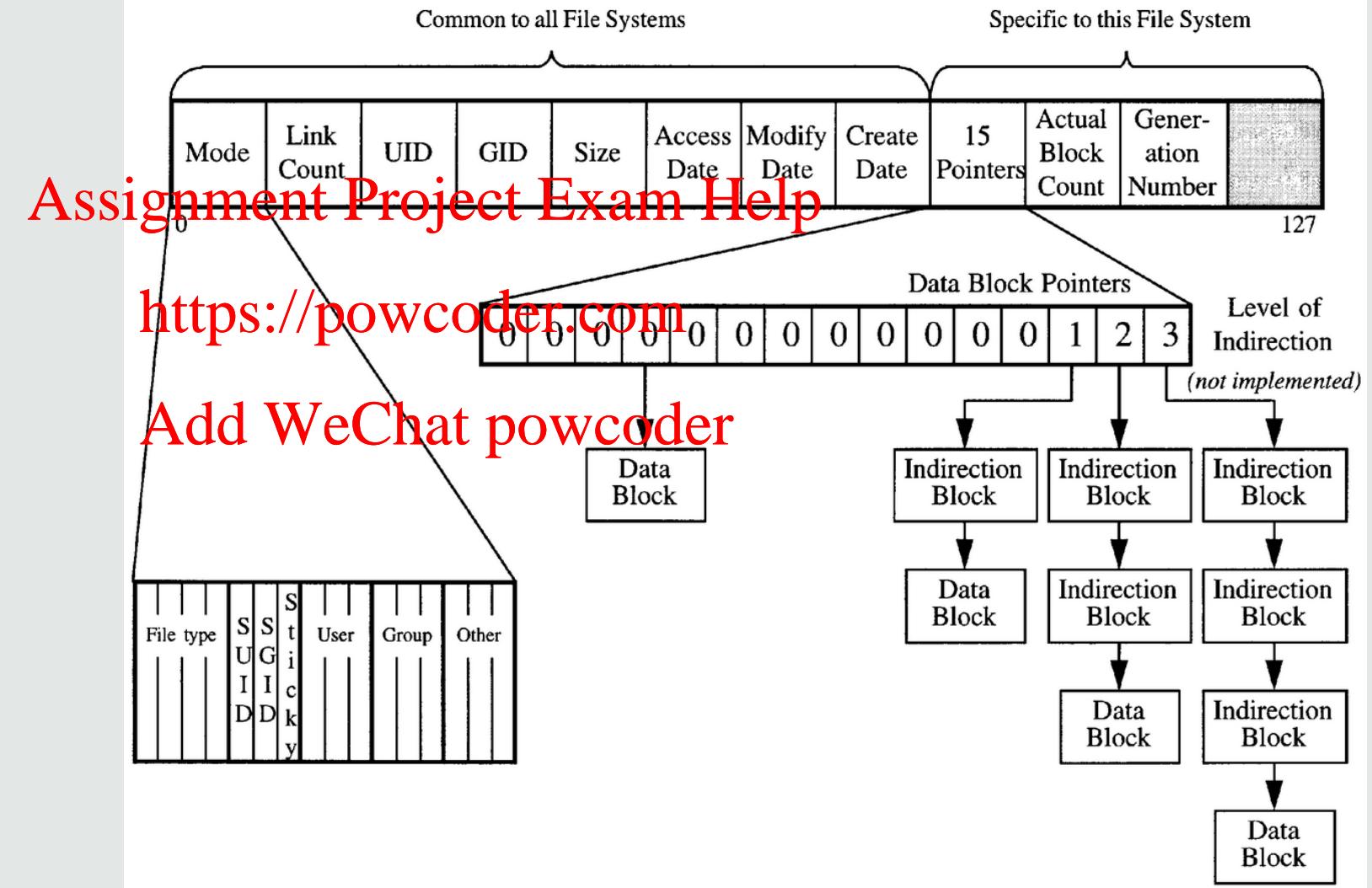
Recap questions

1. What does the mount operation do?
Assignment Project Exam Help
<https://powcoder.com>
2. What information does the kernel maintain for files used by processes?
3. What is reference counting in resource management?
4. What is the unit of storage allocation for files?
Add WeChat powcoder
5. What is a file allocation table?
6. What is an i-node?
7. What is the difference between a hard and a soft link?

Recap: File abstraction

4

UNIX i-nodes



Recap: File abstraction

UNIX i-nodes

Accessing /usr/ast/mbox

Assignment Project Exam Help
<https://powcoder.com>
Add WeChat powcoder

Root directory		I-node 6 is for /usr		Block 132 is /usr directory		I-node 26 is for /usr/ast		Block 406 is /usr/ast directory	
1	.			6	.			26	.
1	..			1	..			6	..
4	bin	Mode size times	132	19	dick	30	erik	64	grants
7	dev			51	jim	26	ast	92	books
14	lib			45	bal			60	mbox
9	etc							81	minix
6	usr							17	src
8	tmp								

Looking up usr yields i-node 6

I-node 6 says that /usr is in block 132

/usr/ast is i-node 26

I-node 26 says that /usr/ast is in block 406

/usr/ast/mbox is i-node 60

Recap: File system implementation

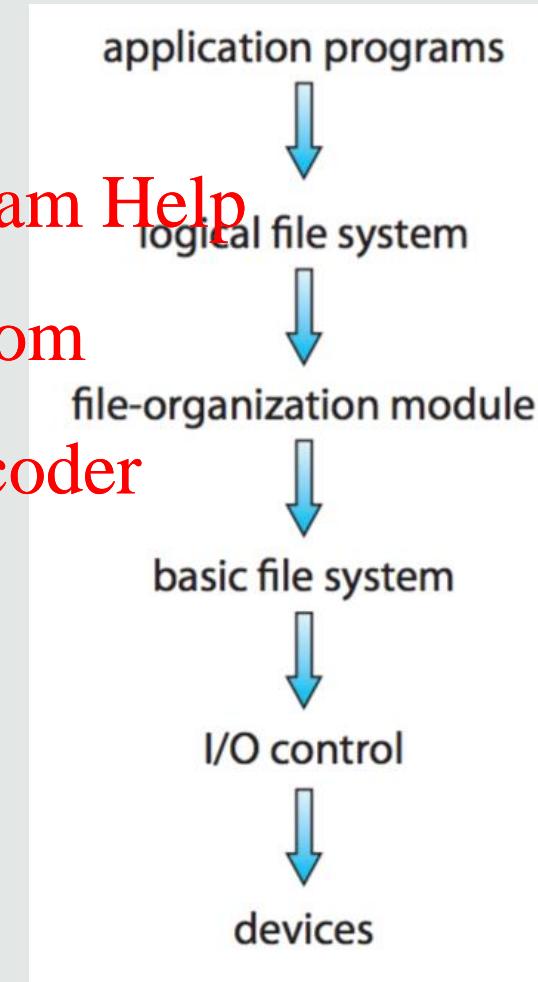
6

- Logical file system
 - Provides file interface (e.g., system calls)
- File organisation module
 - Manages block allocation to files
E.g. i-nodes in UNIX
- Basic file system
 - Block-based storage format
E.g. FAT-32, NTFS, ext4
- I/O device drivers
 - Physical access to storage media
E.g. HD, DVD, USB

Assignment Project Exam Help

<https://powcoder.com>

Add WeChat powcoder



Storage devices

7

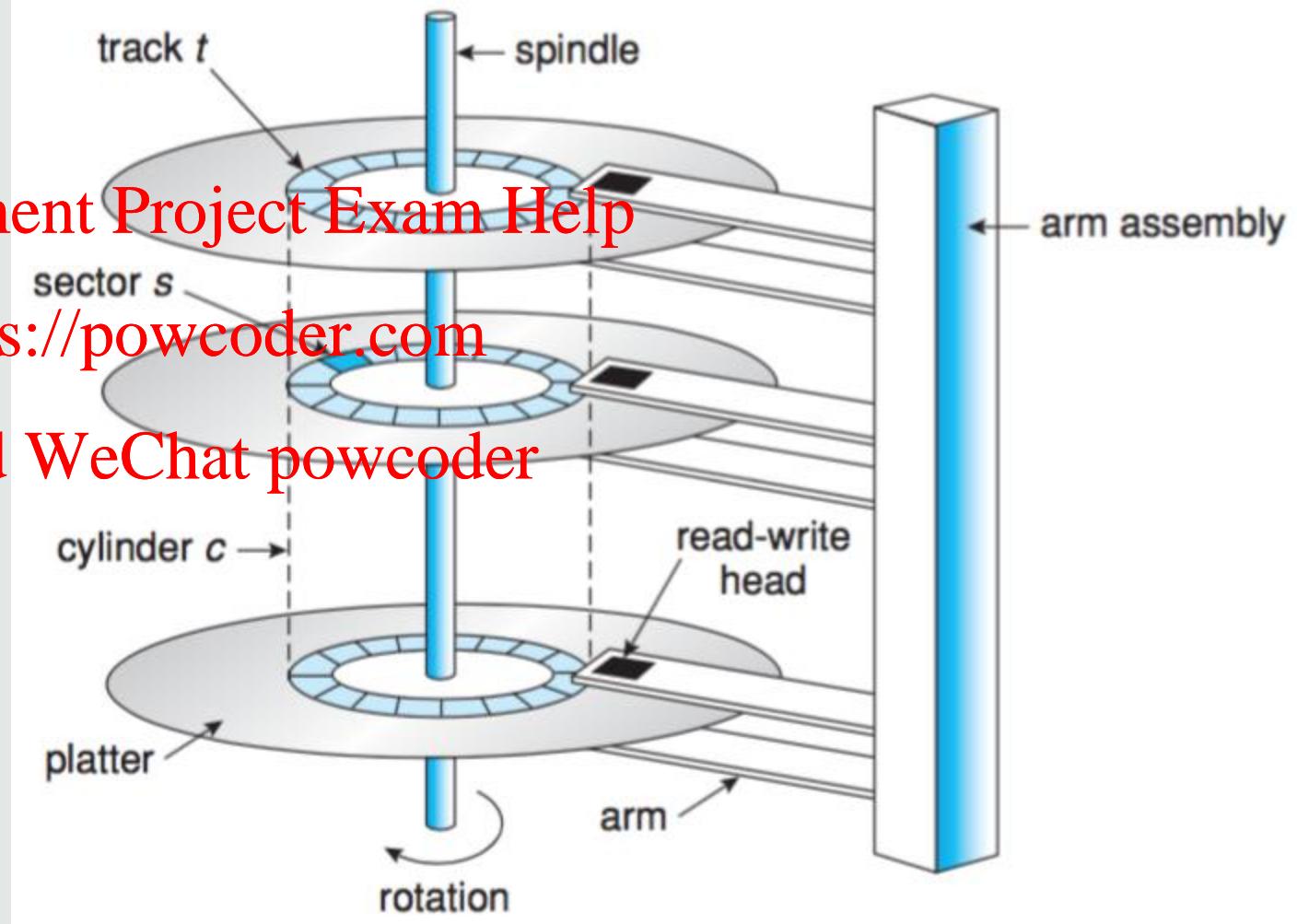
E.g., a hard disk

- Low-level formatting: info for disk controller, e.g. tracks, sectors, ECCs for fixing soft errors, bad blocks

Assignment Project Exam Help

<https://powcoder.com>

Add WeChat powcoder



Partition Tables

MBR (Master Boot Record)

- Cylinder-Head-Sector (CHS) addressing: 512MB limit
- Maximum 15 partitions (with logical partitions)
- MBR-partition size limited to 2TB
- Logical Block Addressing (LBA): Today 48 or 64bit

Assignment Project Exam Help

<https://powcoder.com>

Add WeChat powcoder

Address	Description		Size (bytes)
+000 _{hex}	+0	Bootstrap code area	446
+10E _{hex}	+446	Partition entry №1	16
+1CE _{hex}	+462	Partition entry №2	16
+1DE _{hex}	+478	Partition entry №3	16
+1EE _{hex}	+494	Partition entry №4	16
+1FE _{hex}	+510	55 _{hex}	2
+1FF _{hex}	+511	AA _{hex}	
Total size: 446 + 4×16 + 2			512

Partition Tables

GPT (Globally unique identifiers PT)

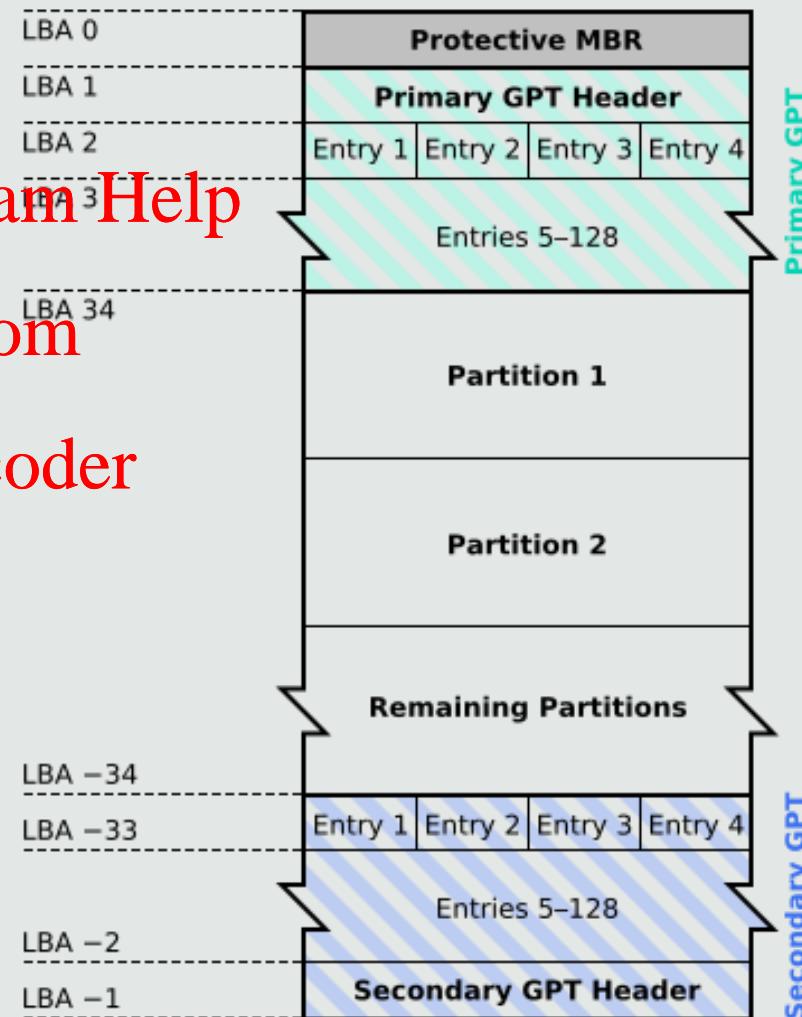
- Supported by all modern operating systems

Assignment Project Exam Help

<https://powcoder.com>

Add WeChat powcoder

GUID Partition Table Scheme



File System Formats

10

Raw format

- Just blocks, no file system
- E.g. Linux swap space

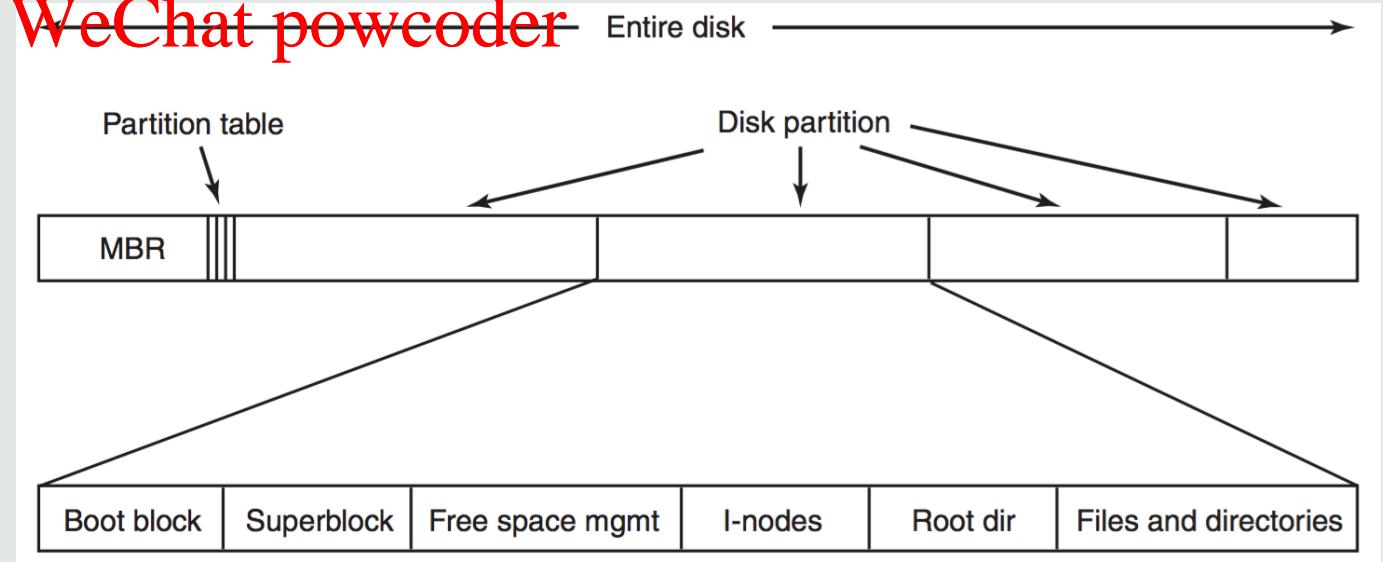
Assignment Project Exam Help

<https://powcoder.com>

High-level formatting:

- Initialise volume with file system data structures

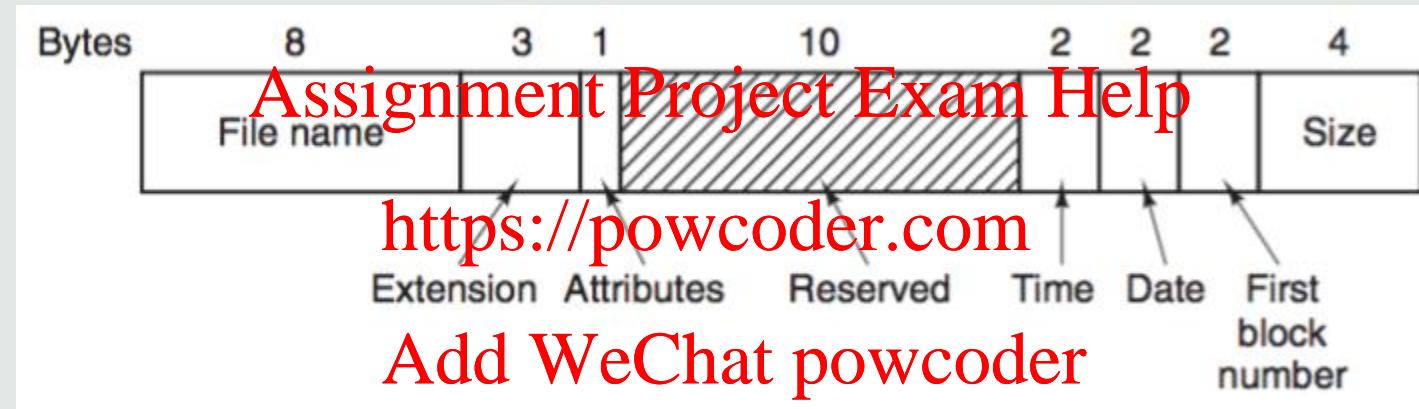
Add WeChat powcoder



File System Examples: FAT

11

FAT-12 (MS-DOS) directory entry

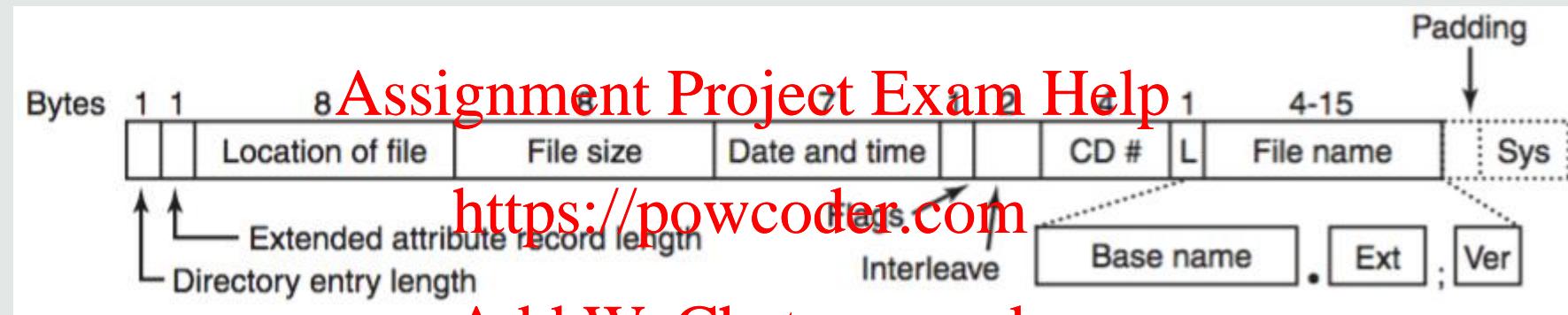


- 12 bit block number, 512B blocks size → partition size 2MB
- FAT-16: 16 bit cluster numbers, 32KB cluster size → 2GB
- FAT-32: 28 bit cluster numbers, 64KB cluster size → 16TB
- but $2^{31} = 2\text{GB}$ file size (systems that support large files: $2^{32} = 4\text{GB}$)
- No file permissions

File System Examples: ISO 9660

12

ISO 9660 (CD-ROM) directory entry



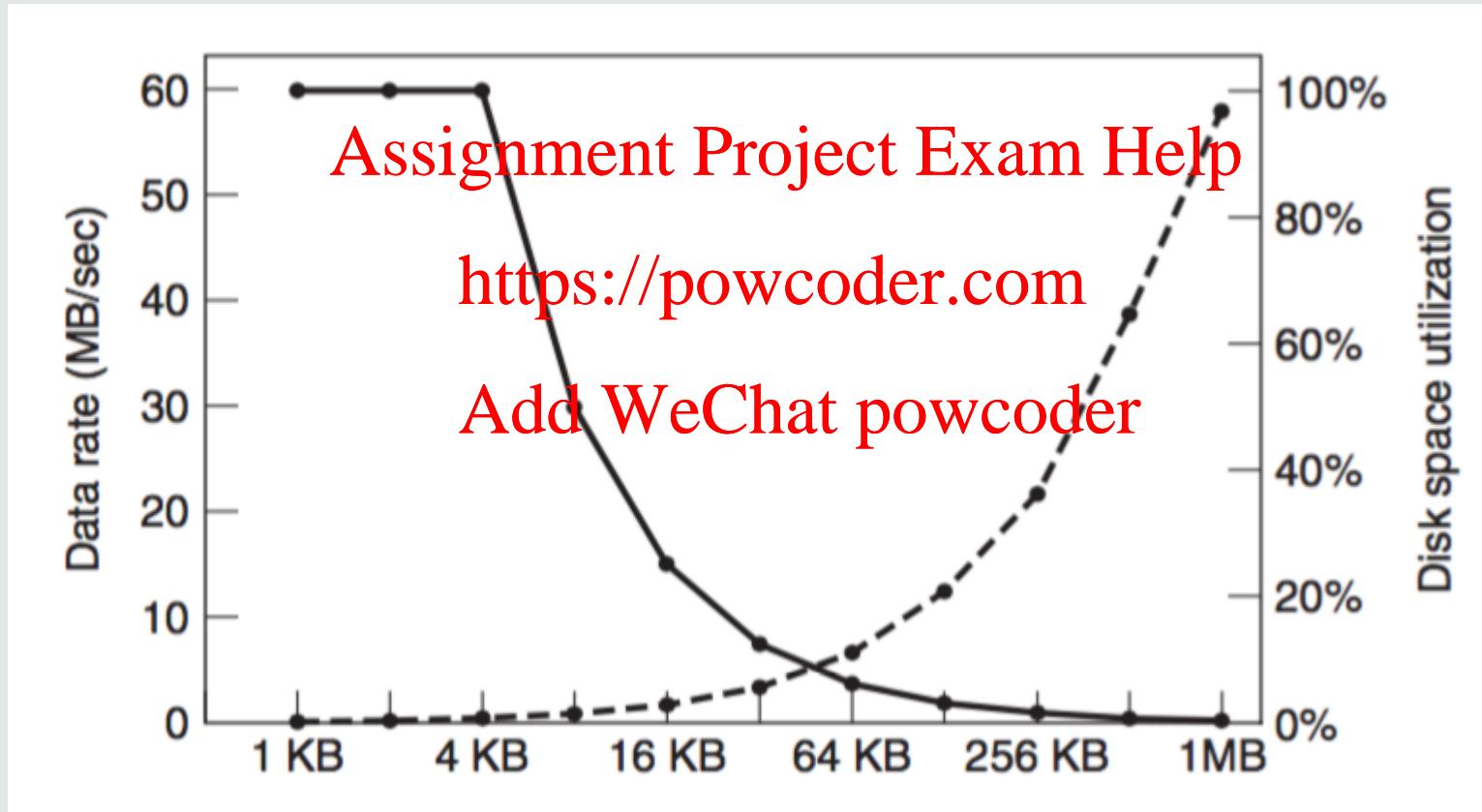
- 2KB blocks
- Contiguous allocation (mostly)

Limitations:

- Length of file names, file attributes, directory nesting, . . .
- Rock Ridge (UNIX) and Joliet (Windows) extensions

How to choose the block size?

13



Mean file size

- In 1984: 1.1KB
- In 2005: 2.4KB
- in 2017: ?

Assignment Project Exam Help

<https://powcoder.com>

Internal fragmentation

- 4KB blocks: 10% largest files occupy 93% of utilised blocks
- 8KB blocks: 10% largest files occupy 90% of utilised blocks

Add WeChat powcoder

Disk access time

- Rotation speed
- Seek time to position head
- Small block size + non-contiguous allocation = unreasonably long access times
- → „defragmentation“

File system operations involve many writes (e.g. appending data in a file)

1. Update free blocks list to get a new block
2. Add pointer to that block and increase size of file in inode
3. Write data into block

System crash → can lead to inconsistent data on storage device

<https://powcoder.com>
Add WeChat powcoder

How to detect inconsistencies:

- E.g. UNIX fsck
- Mark problems and try to recover
- Drawback: time-consuming; data loss when recovery fails

Approaches: Log-Structured File Systems, RAID

Data Storage Consistency

16

Reusing ideas from database systems

- Transactions
- Log: Store changes on hard disk
- Commit them in the background
- On a crash, inconsistencies are limited to interrupted transactions

Assignment Project Exam Help

<https://powcoder.com>

Add WeChat powcoder

Some operations appear to be much faster

- Also known as “journaling”
- Examples: NTFS, ext4, . . .

Data Storage Consistency

17

Journaling FS example: ext4

- 48 bit block numbers, 4KB block size → 2⁴⁸ / 4096 = 2³⁸ blocks
- 32 bit i-node numbers; 64 bit for file size
- 128MB journal

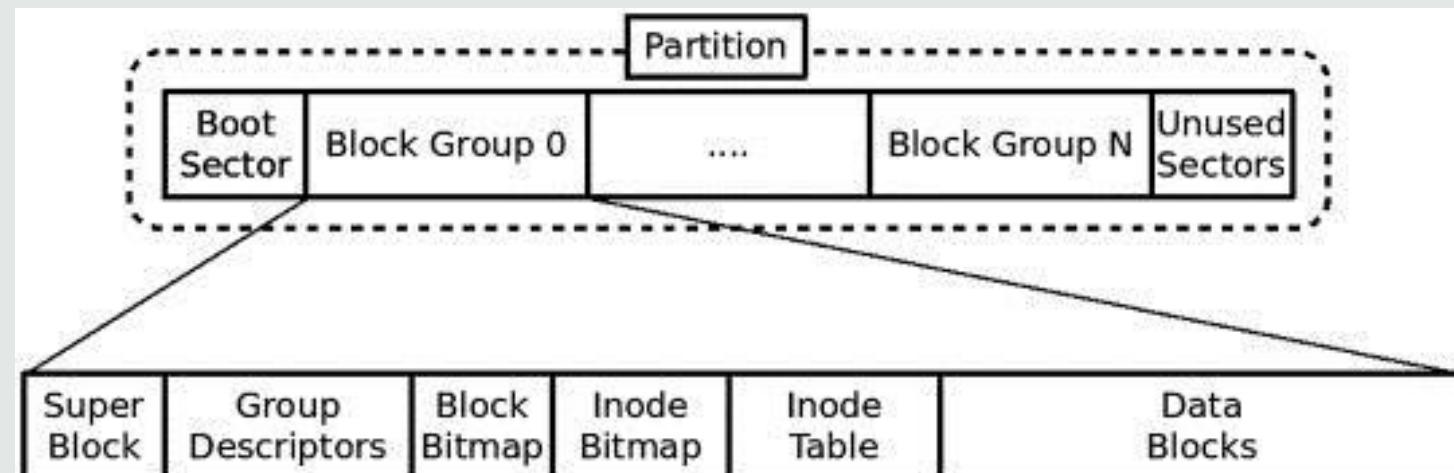
Assignment Project Exam Help

<https://powcoder.com>

Pre-allocation support

Disk partitioned into block groups

Add WeChat powcoder



Data Storage Consistency

18

Journaling FS example: ext4

- Journaling Block Device (JBD) generic journaling layer
- Assignment Project Exam Help

- Journal entries:

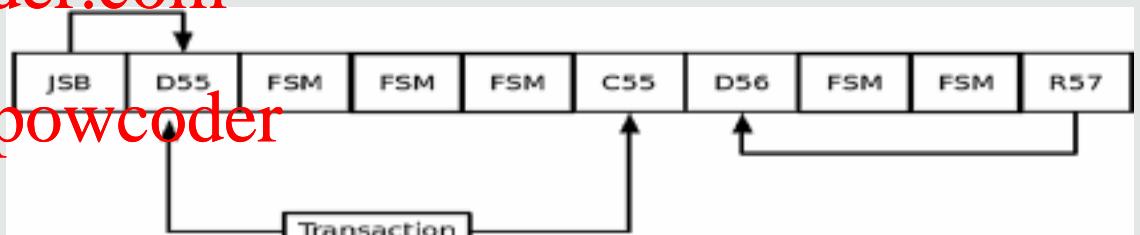
- Descriptor
(contains destination block number)
- Data (data / metadata)
- Commit
- Revocation

- Journaling modes:
journal / ordered / writeback

- If interrupted, transaction is replayed

<https://powcoder.com>

Add WeChat powcoder



Key

Abbreviation	Meaning
JSB	Journal Super Block
D[X]	Descriptor Block X
FSM	File System Metadata
C[X]	Commit Block X
R[X]	Revoke Block X

Data Storage Consistency

19

Redundant Arrays of Independent Disks (RAID)

Objectives

- Increase reliability by redundant data storage
- Increase performance by parallel data access

Mirroring

- Assume disks fail independently
- Example: MTBF 100000h, MTTR 10h
→ mean time to data loss $100000^2 = (2 * 10) \approx 57000$

Data striping

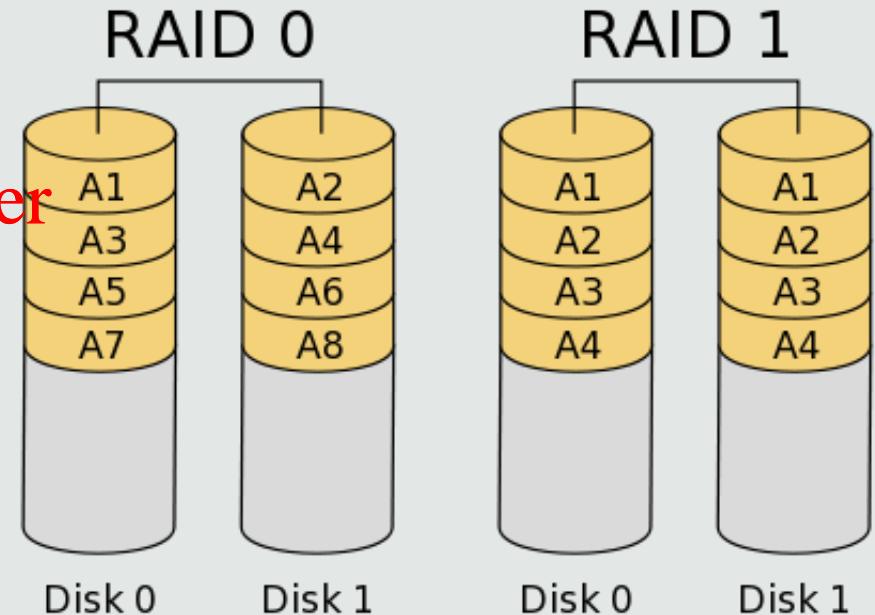
- Distribute data across n disks
- Write every $(k \bmod n)^{\text{th}}$ chunk of m bits onto disk $k \rightarrow n$ times faster

Assignment Project Exam Help

<https://powcoder.com>

Add WeChat powcoder

Error correcting codes
RAID Levels (see also textbook)



Performance Optimisation

20

- Block Caching
 - Typical policies: Clock, LRU, etc
 - Group blocks (superblock, i-nodes, data, etc)
 - Consistency: write dirty blocks as soon as possible, (e.g. when block full, periodically)
<https://powcoder.com>
- Write-Through Cache
 - Data written immediately
 - Requires more disk I/O
 - Data loss more unlikely for removable disks
- Block Read Ahead
 - Speculatively read blocks ahead of time, e.g. when file is read sequentially

Add WeChat powcoder

Performance Optimisation

21

Hard disks: reduce seek time

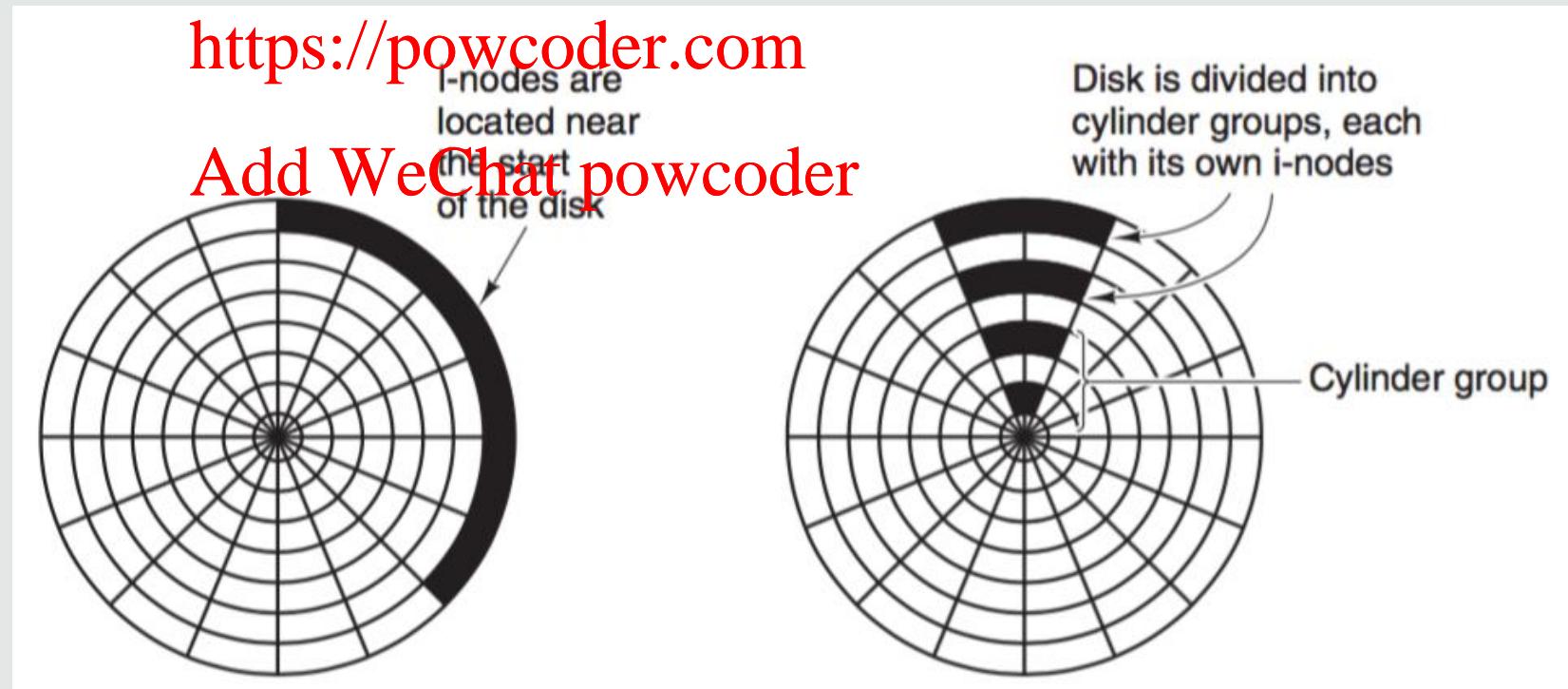
- Clustering, block groups
- Disk scheduling
(next week)

SSD issues: Relocate blocks that are often written to even out wear

Assignment Project Exam Help

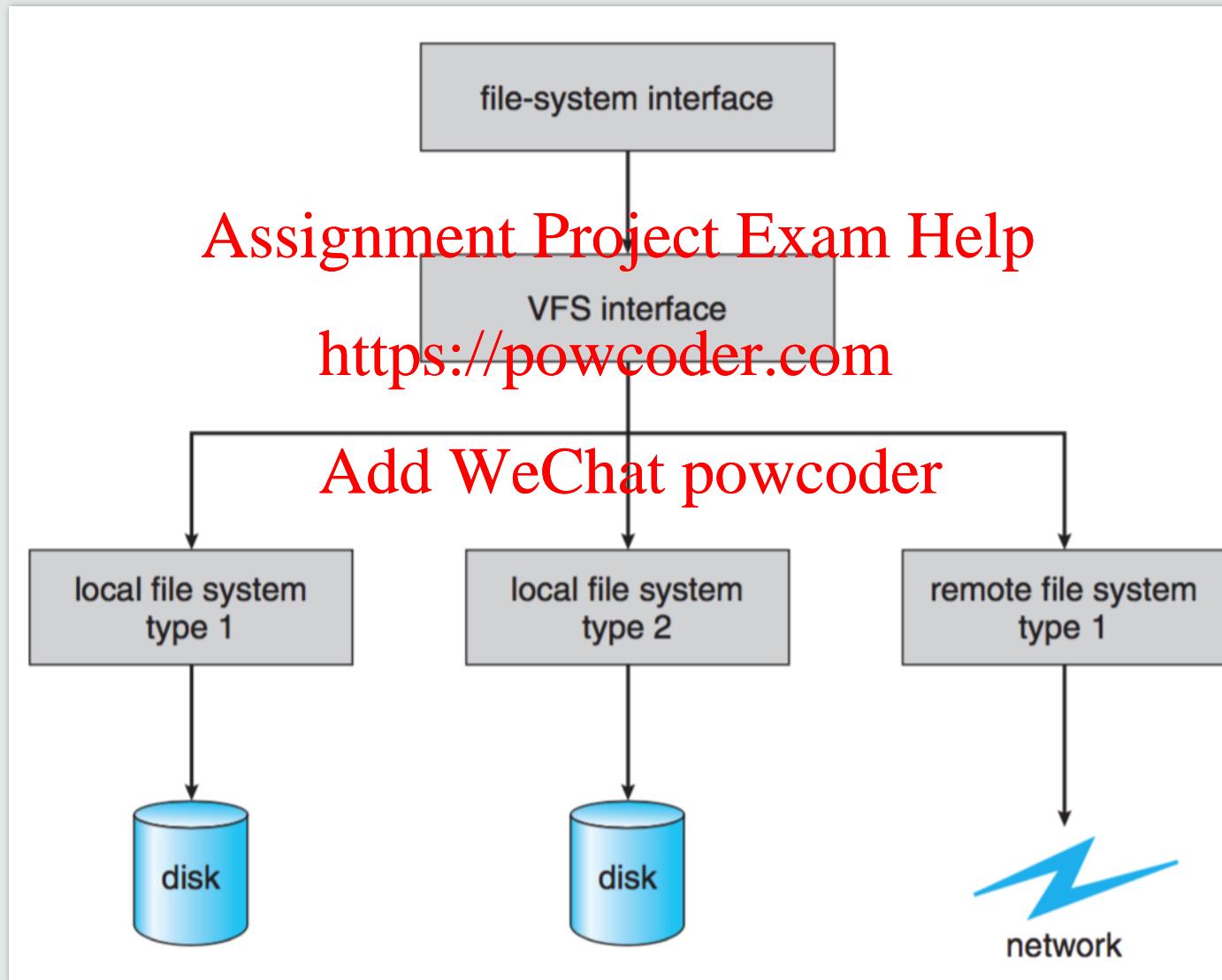
<https://powcoder.com>

Add WeChat powcoder



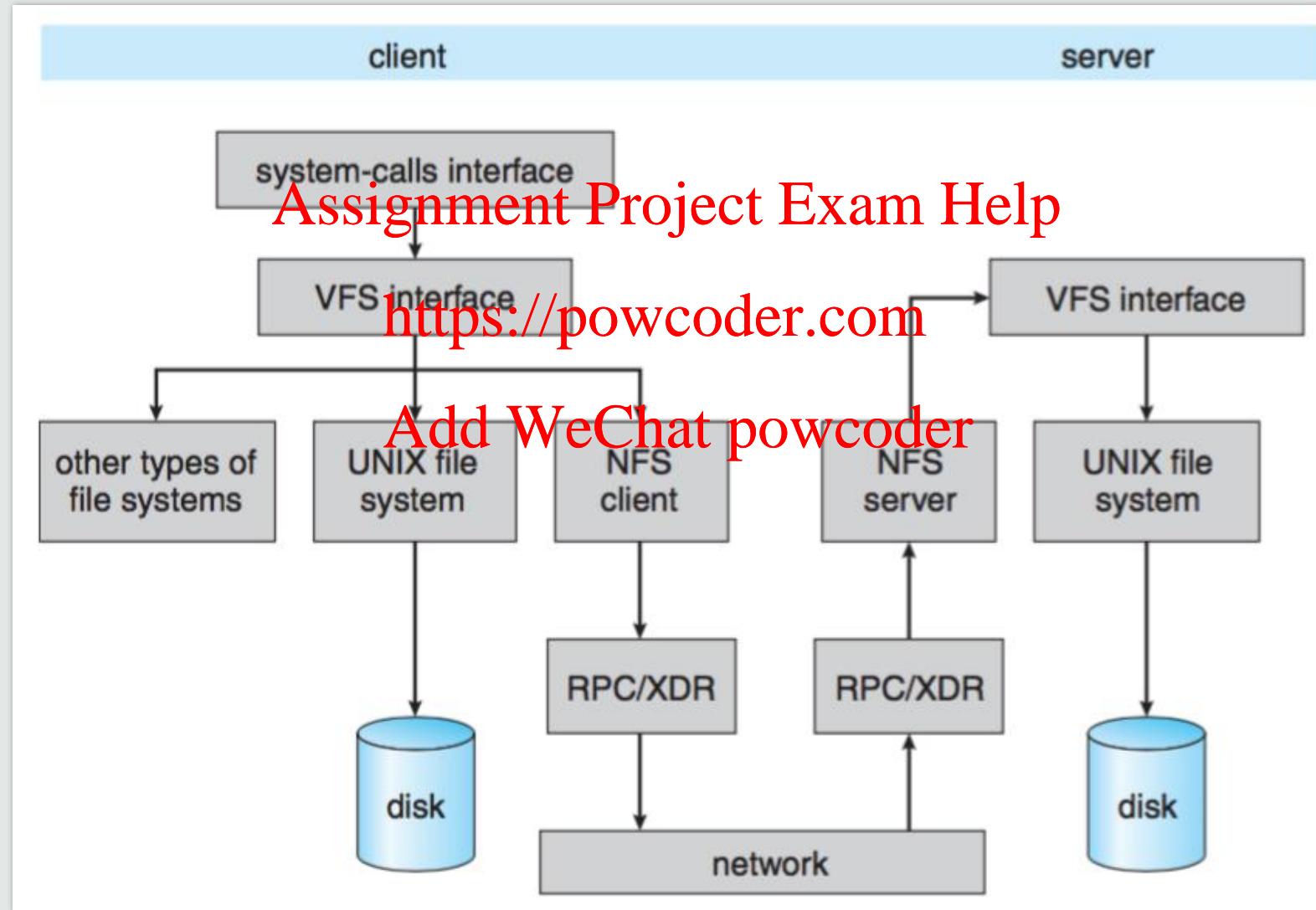
Virtual File Systems

22



Virtual File Systems

23



File abstraction

- Unified interface to various devices
- System calls:
open(), read(), write(), close(), ...

File System Management

- File system formats and block size
- Kernel data structures for file management
- Consistency: Journaling, RAID, ...
- Optimisations: caches, clusters, ...

Assignment Project Exam Help
<https://powcoder.com>
Add WeChat powcoder

Directory structure and Block allocation

- Contiguous allocation
- Linked list allocation, e.g. FAT
- Indexed allocation, e.g. i-nodes

- Tanenbaum & Bos., Modern Operating Systems

- Chapter 4

Assignment Project Exam Help

- Silberschatz et al., Operating System Concepts

- Chapters 10, 11, 12

Add WeChat powcoder

- Introduction
 - Operating System Architectures
 - Processes
 - Threads - Programming
 - Process Scheduling - Evaluation
 - Process Synchronisation
 - Deadlocks
 - Memory Management
 - File Systems
 - Input / Output
 - Security and Virtualisation
- Assignment Project Exam Help**
<https://powcoder.com>
Add WeChat powcoder