CHINA LINUX KERNEL
中国Linux内核开发者大会

華中科技大學
网络空间安全学院
School of Cyber Science and Engineering, HUST

第19届中国 Linux内核开发者大会

赞助单位

華中科技大學 | HUAWEI | 阿里云 | FUJITSU | 腾讯云
OPPO | 火山引擎 | vivo | 龙芯中科 LOONGSON TECHNOLOGY | 蚂蚁集团 ANT GROUP

支持单位

清華大學 Tsinghua University | 迪捷软件 DIGIPROTO | OpenEuler | OpenAnolis 龙蜥社区 | OpenCloudOS

支持社区&媒体

CSDN | 云巅论剑 | sf 思否 | 阅码场 yomocode | InfoQ 极客传媒
51CTO | 开源江湖

2024年10月 湖北·武汉

華中科技大學

F2FS : Next-Generation Compression

Yangtao Li

frank.li@vivo.com

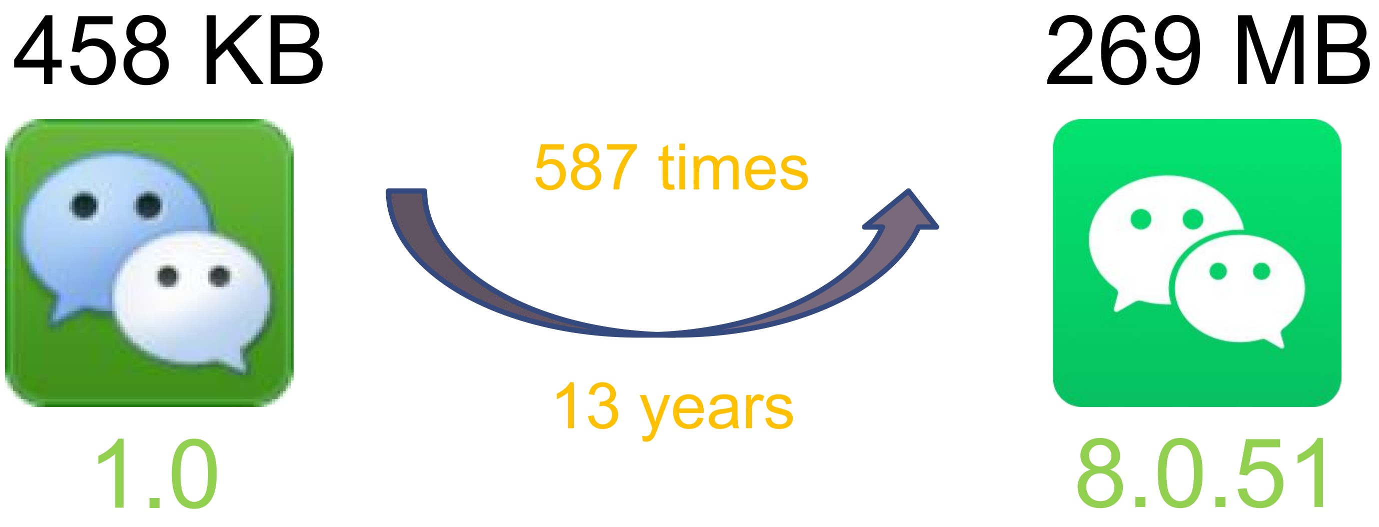
CONTENT

- Background
- F2FS Compression
- Next-Generation Compression

Part One

Backgroud

1. Background - App size keeps growing



458 KB

269 MB

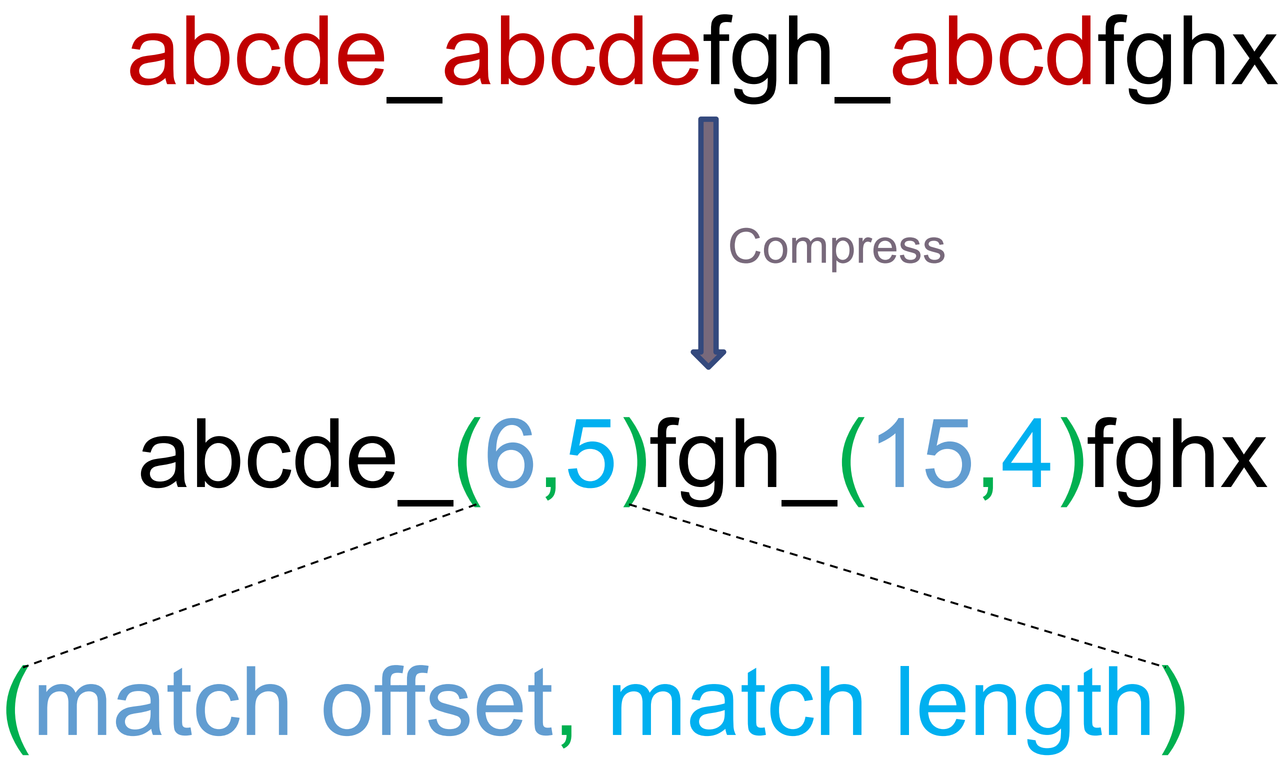
587 times

13 years

1.0

8.0.51

1. Background - Compression principle



abcde_abcdefgh_abcdfghx

Compress

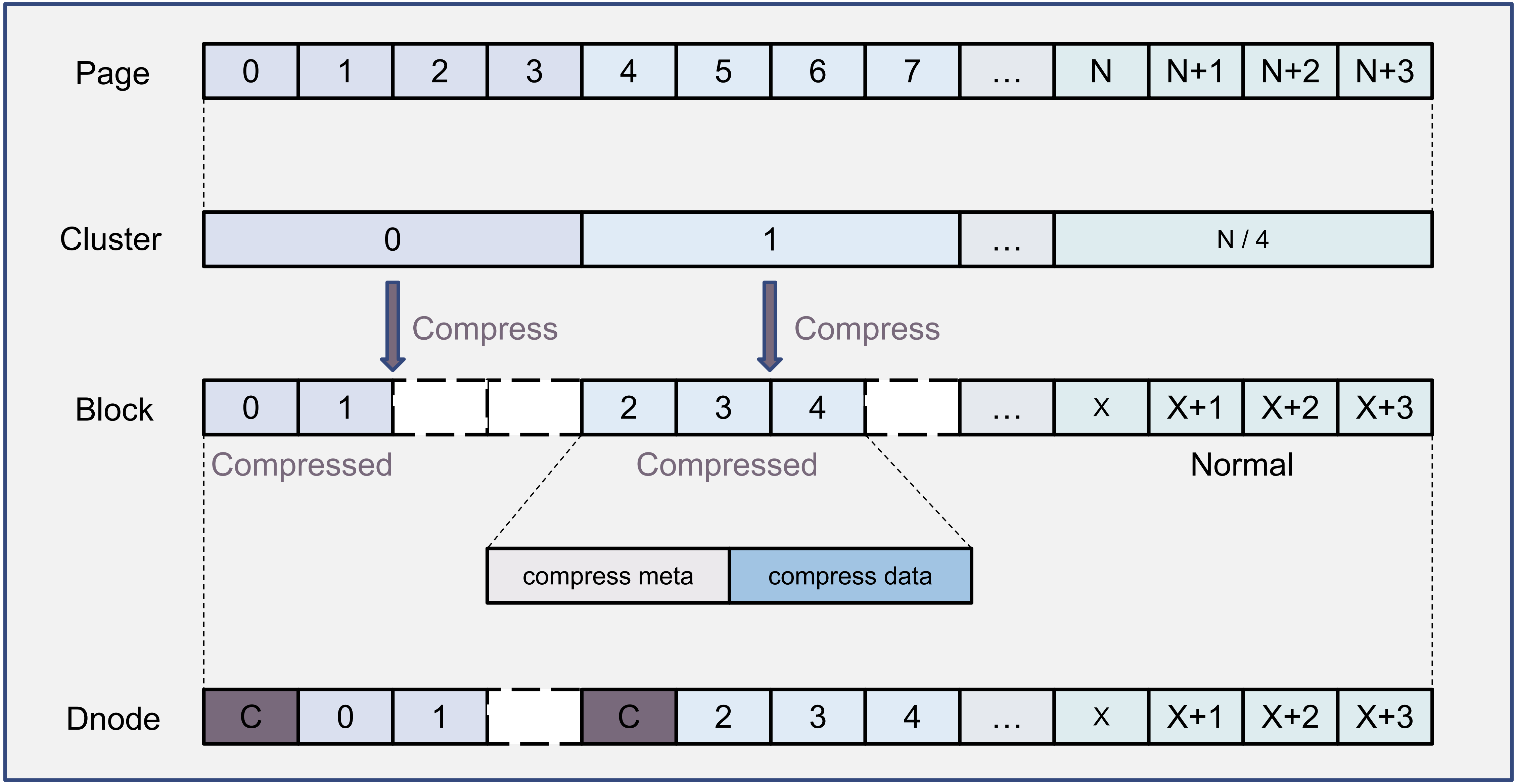
abcde_(6,5)fgh_(15,4)fghx

(match offset, match length)

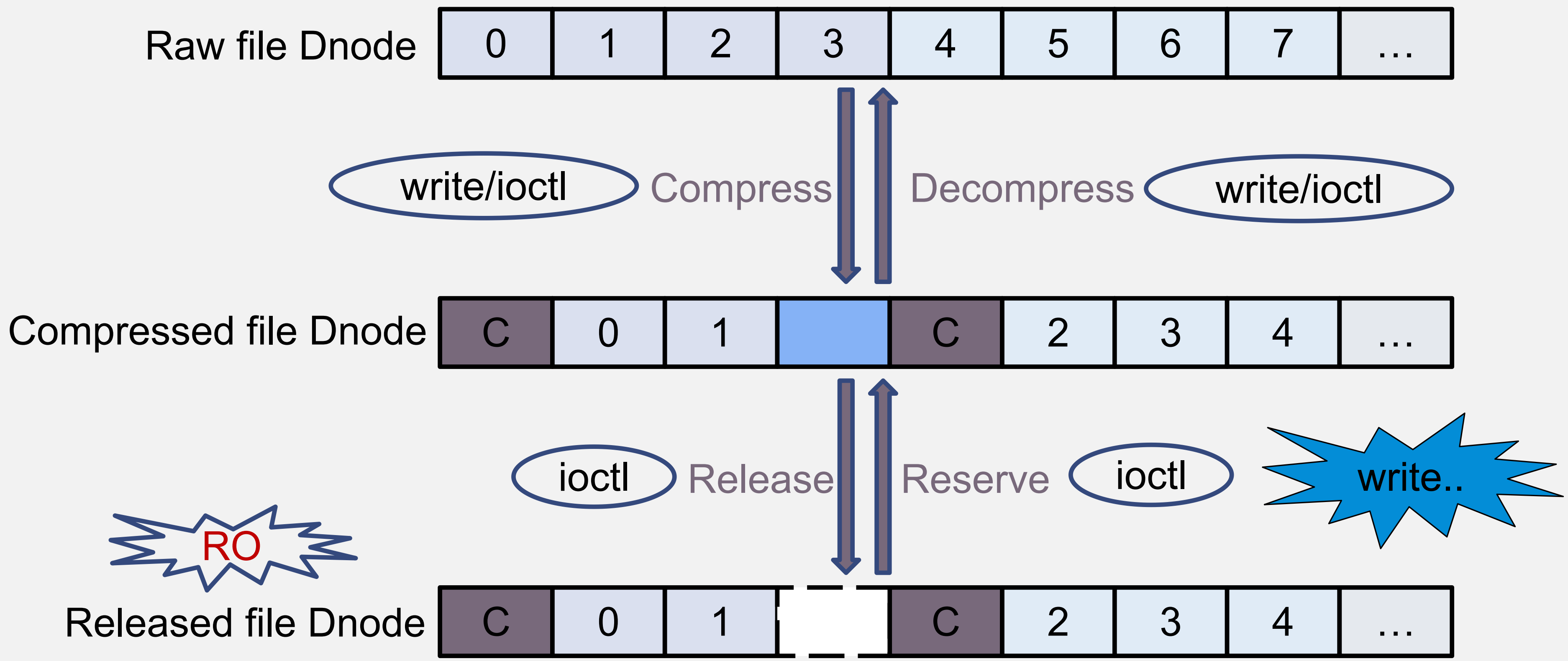
Part Two

F2FS Compression

2. F2FS Compression - F2FS compression implementation



2. F2FS Compression - Feature Extensions



Raw file Dnode | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | …

write/ioctl Compress

Decompress write/ioctl

Compressed file Dnode | C | 0 | 1 | | C | 2 | 3 | 4 | …

ioctl Release

Reserve ioctl

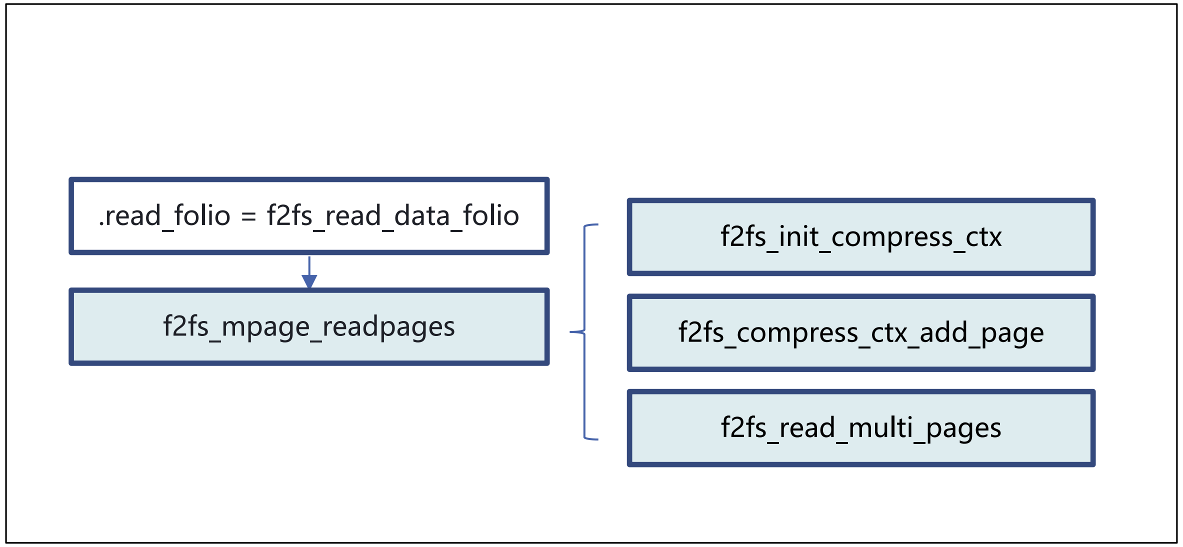
write..

RO

Released file Dnode | C | 0 | 1 | | C | 2 | 3 | 4 | …

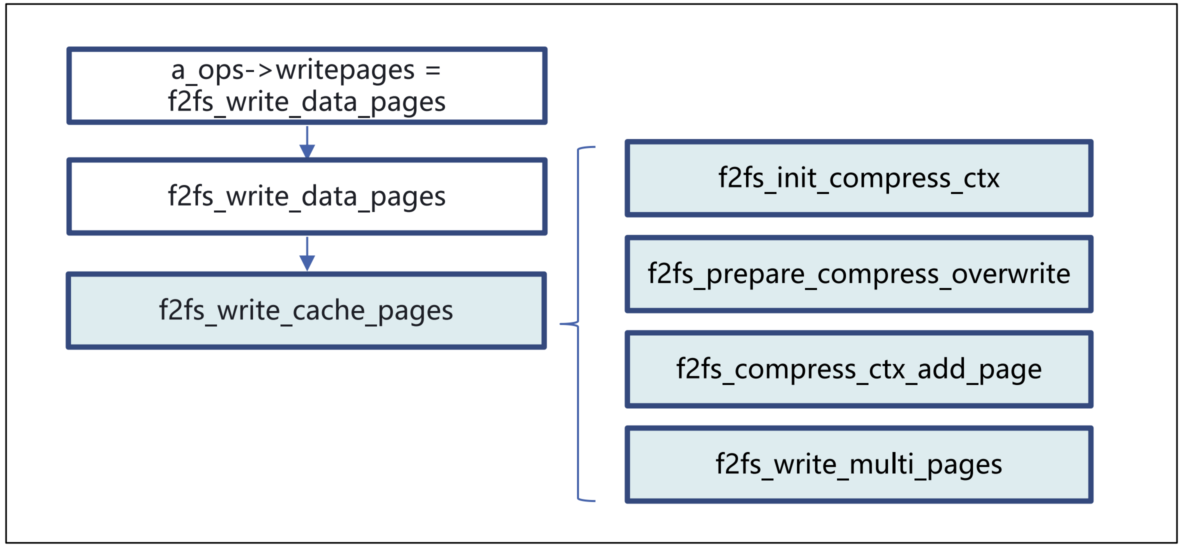
2. F2FS Compression - Performance Optimization

f2fs buffer read



- avoid duplicate counting of valid blocks when read compressed file
- reduce memory allocation in f2fs_mpage_readpages once
- support compress cache
- ……

f2fs buffer write



a_ops->writepages = f2fs_write_data_pages
f2fs_write_data_pages
f2fs_write_cache_pages
f2fs_init_compress_ctx
f2fs_prepare_compress_overwrite
f2fs_compress_ctx_add_page
f2fs_write_multi_pages

- remove unneeded read when rewrite whole cluster
- reduce one page array alloc and free when write compressed page
- use onstack pages
- supports writing cluster-aligned IO in Direct IO mode
- ……

2. F2FS Compression - Upstream Contribution

- reserve blocks on released compress inode while writing
- move the conditional statement to hold the inode lock in f2fs_reserve_compress_blocks()
- do not allow to defragment files have FI_COMPRESS_RELEASED
- introduce f2fs_set_compress_level()
- fix to check lz4hc compression when CONFIG_F2FS_FS_LZ4HC is not enabled
- add F2FS_IOC_GET_COMPRESS_OPTION_V2 ioctl
- intorduce f2fs_all_cluster_page_ready
- remove redunant invalidate compress pages
- support POSIX_FADV_DONTNEED drop compressed page cache
- reduce one page array alloc and free when write compressed page
- add nocompress extensions support
- fix to wait page writeback in f2fs_write_raw_pages()
- fix to release compress file for F2FS_IOC_RESERVE_COMPRESS_BLOCKS when has no space
- fix inconsistent update of i_blocks in release_compress_blocks and reserve_compress_blocks
- fix compressed file start atomic write may cause data corruption
- fix remove page failed in invalidate compress pages
- fix overwrite may reduce compress ratio unproperly
- don't force buffered io when in COMPR_MODE_USER mode
- ....

2. F2FS Compression - Result

App data saves 10%-20% space

Part There

Next-Generation Compression

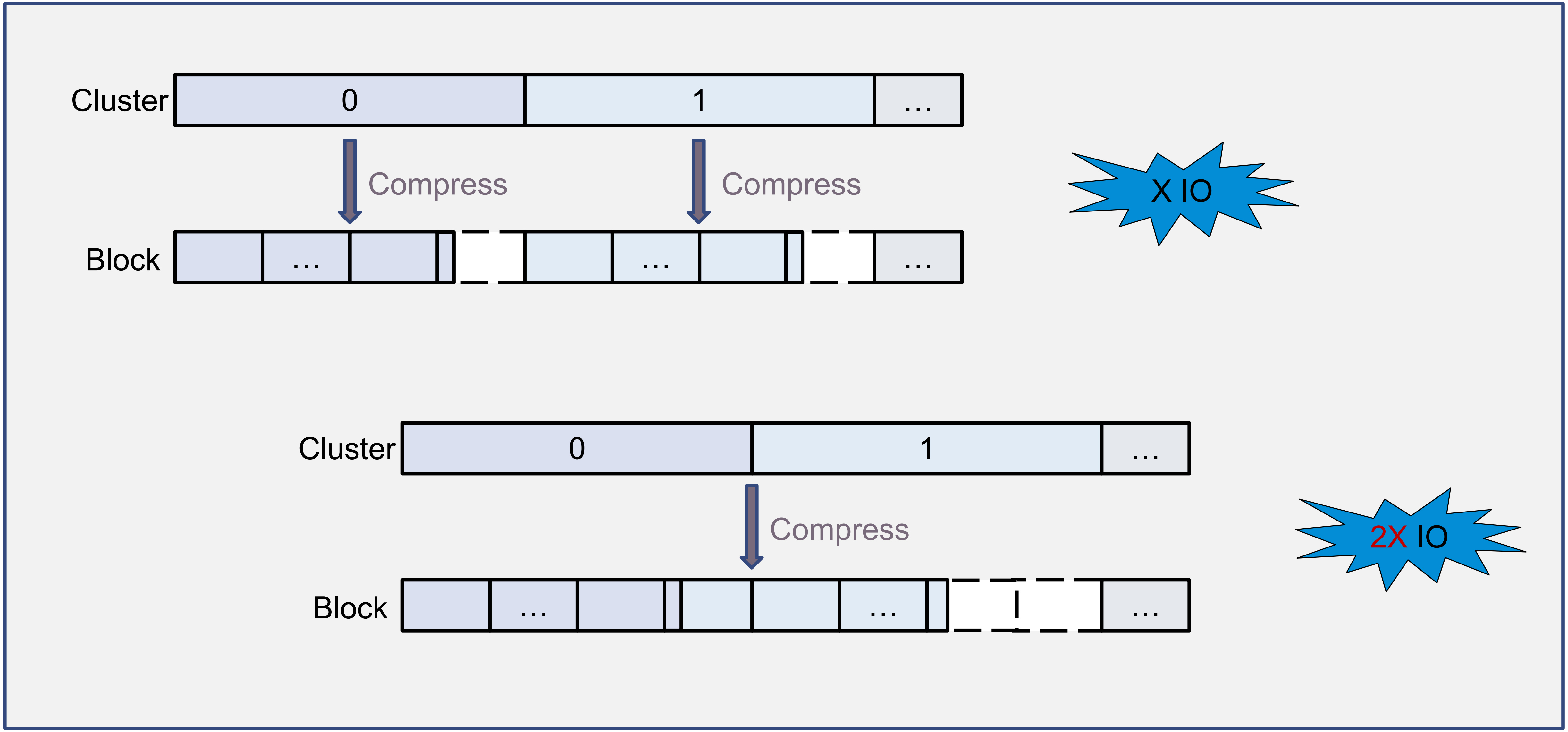
3. Next-Generation Compression - Dynamic Cluster

Compression Ratio

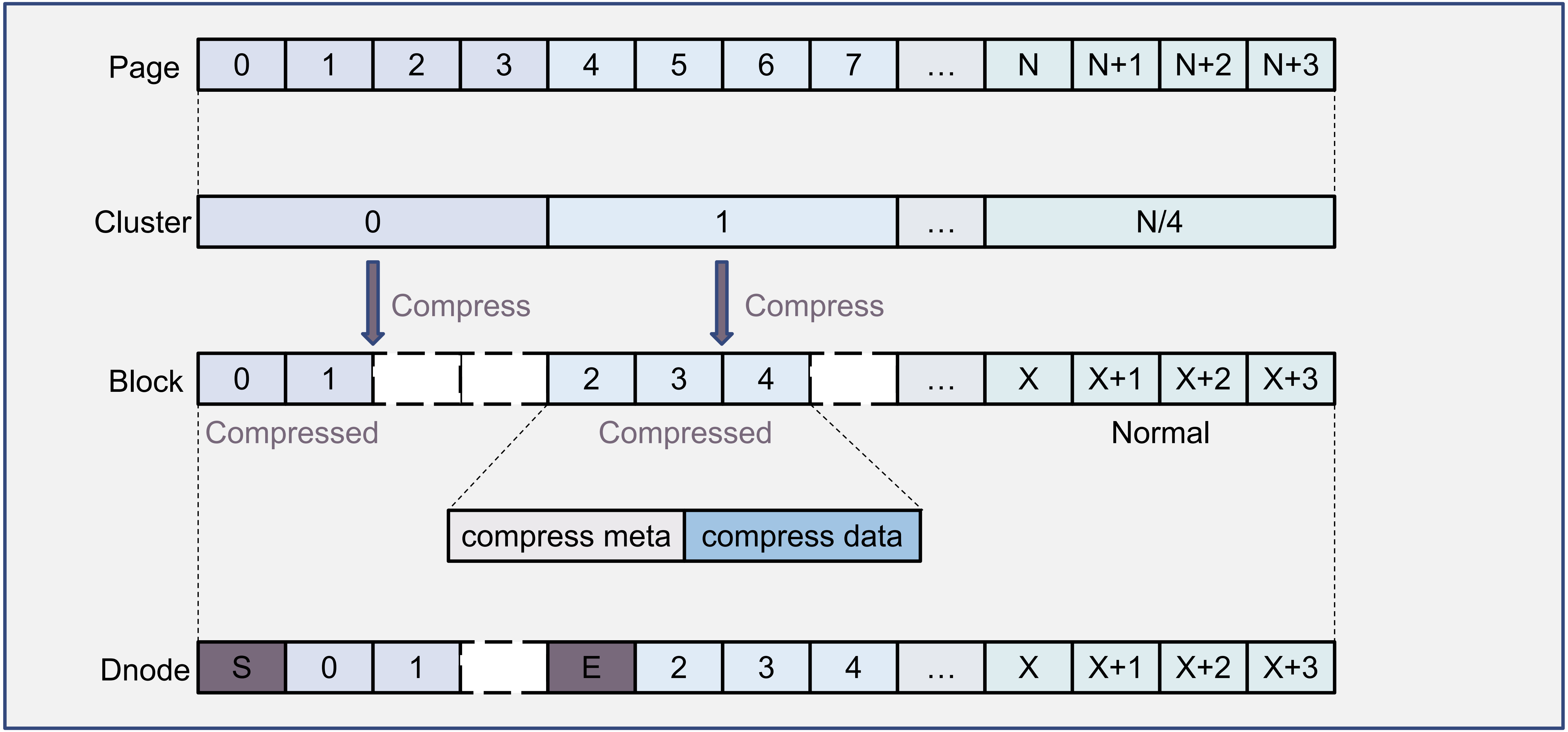
vs

IO Amplification

3. Next-Generation Compression - Dynamic Cluster



3. Next-Generation Compression - Dynamic Cluster



3. Next-Generation Compression - Dynamic Cluster

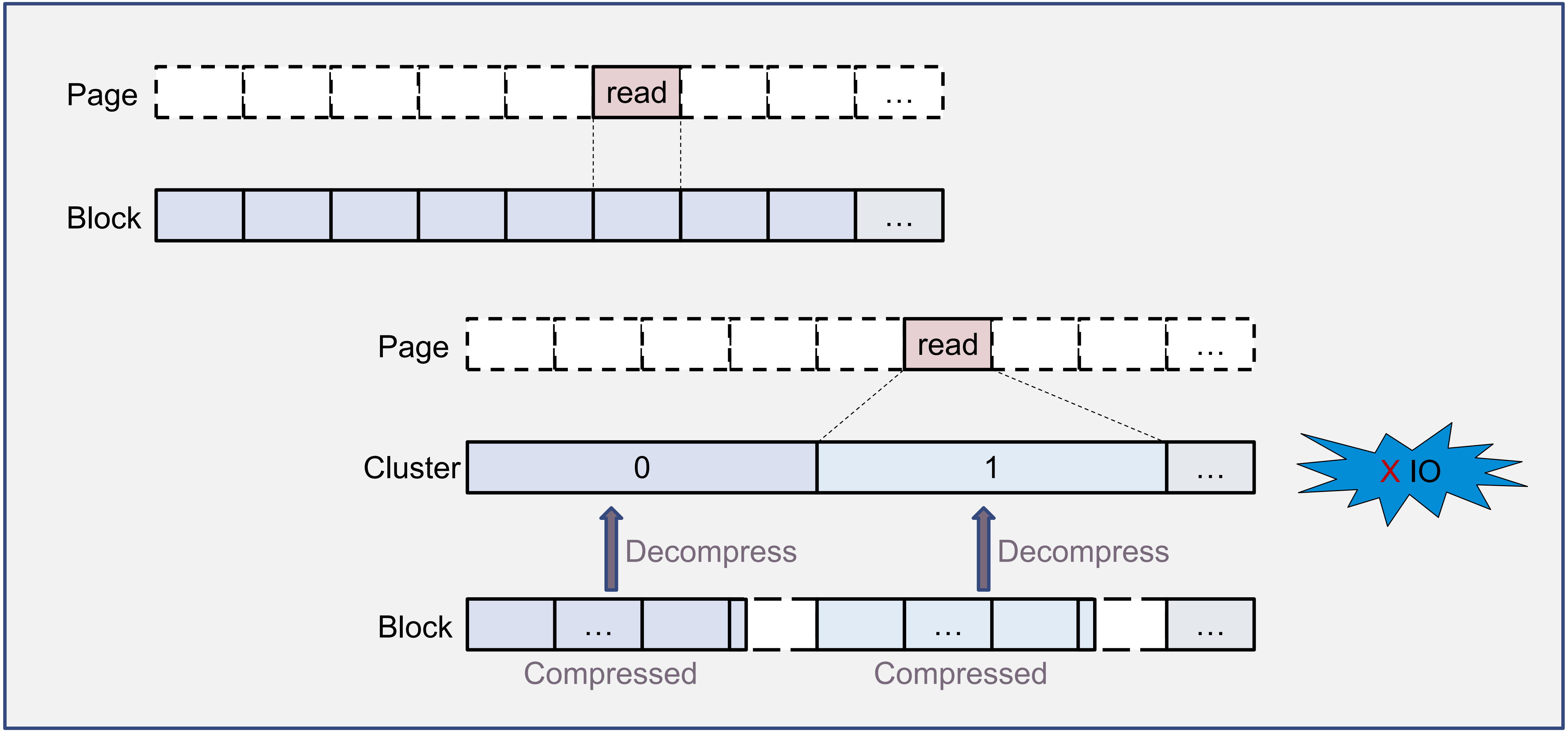
Dynamic Cluster Compression

- Compression rate increased by up to 24%
- Reduced read and write IO by up to 98%
- Suitable for read and write files that prioritize compression rate

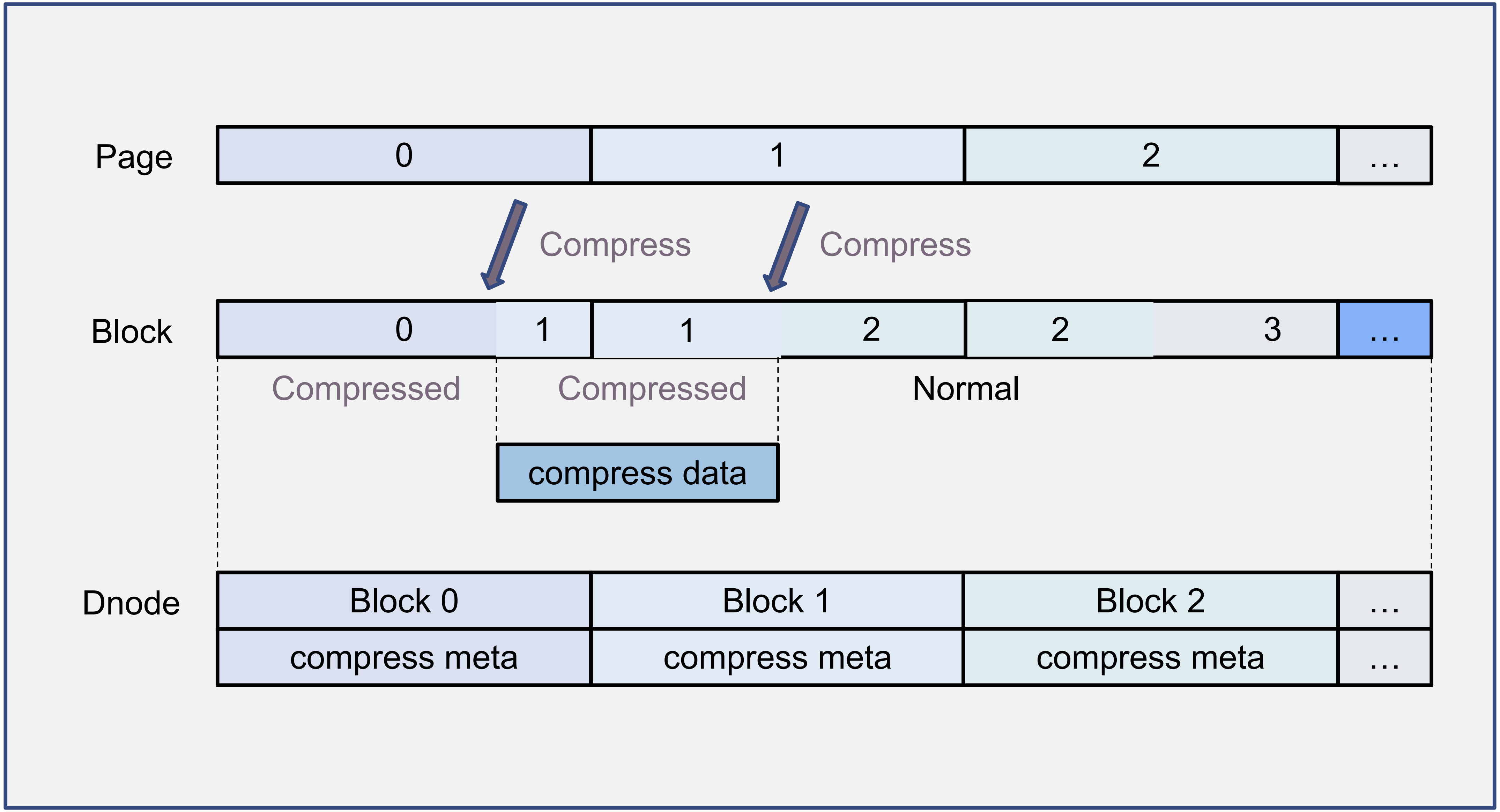
3. Next-Generation Compression - Block-based Fixed length output

Read amplification?

3. Next-Generation Compression - Block-based Fixed length output



3. Next-Generation Compression - Block-based Fixed length output



3. Next-Generation Compression - Block-based Fixed length output

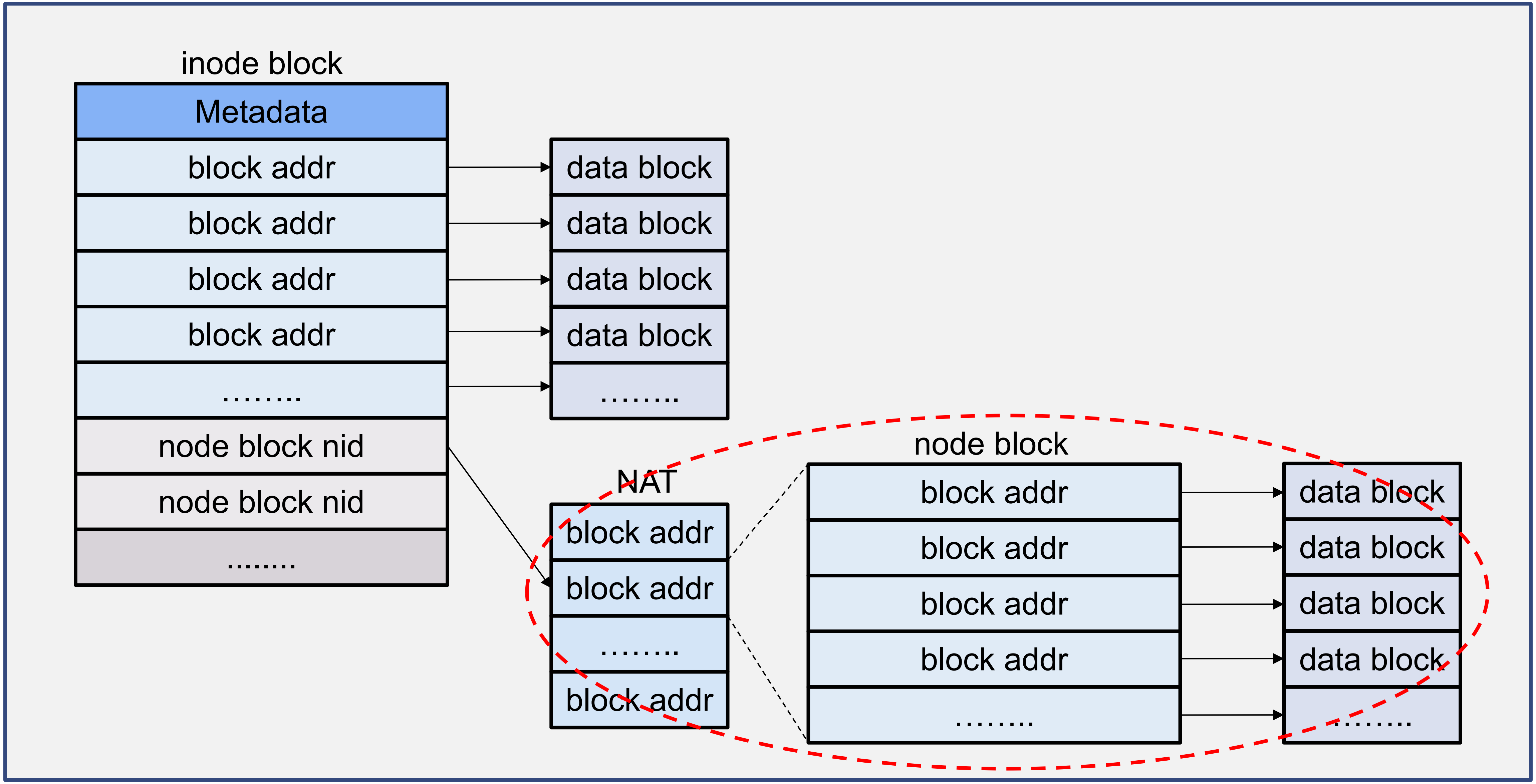
Block-based Fixed length output Compression

- Reduce read IO by 50% ~ 99%
- Suitable for high-frequency read-only files

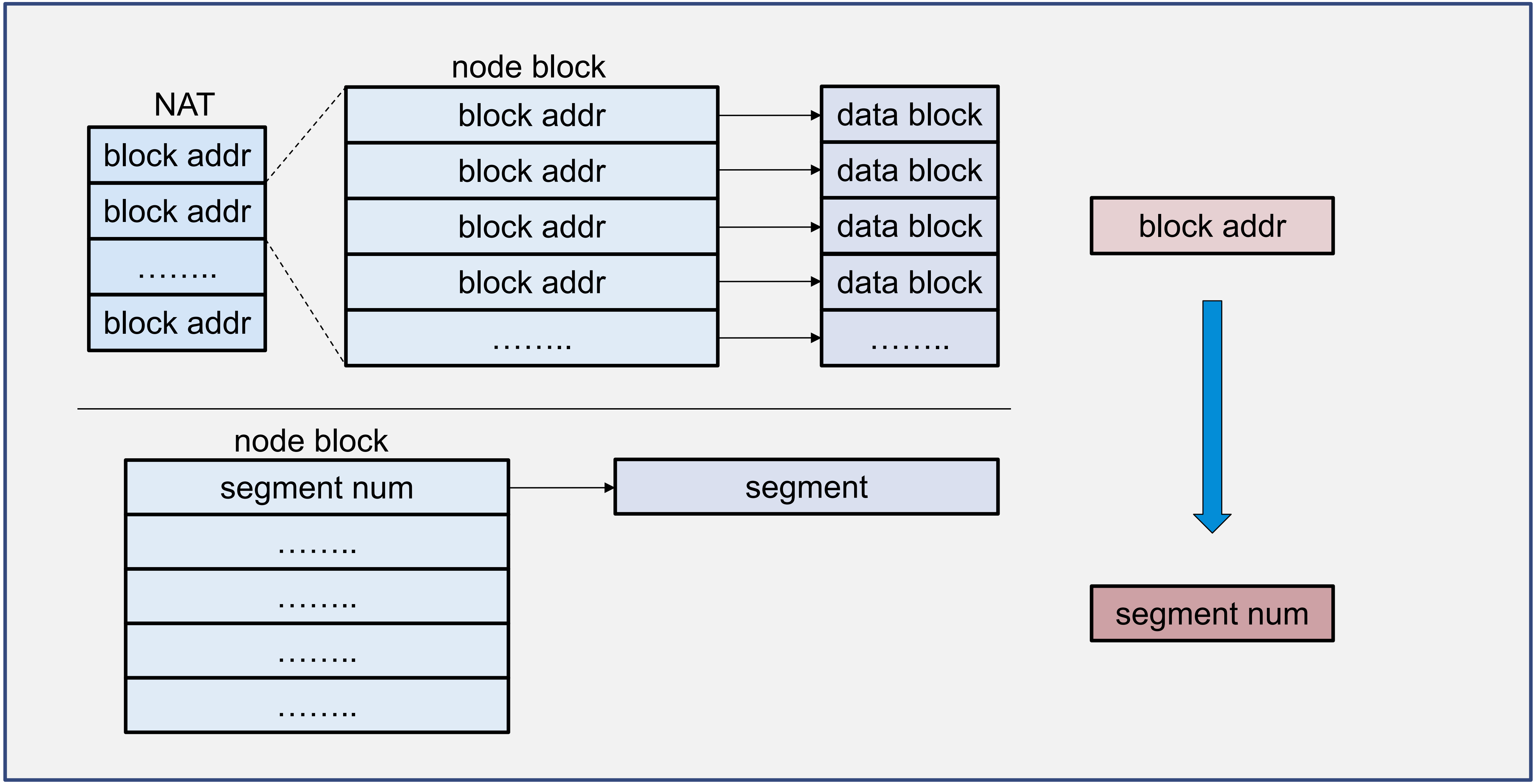
3. Next-Generation Compression - Segment-based Fixed length output

more？

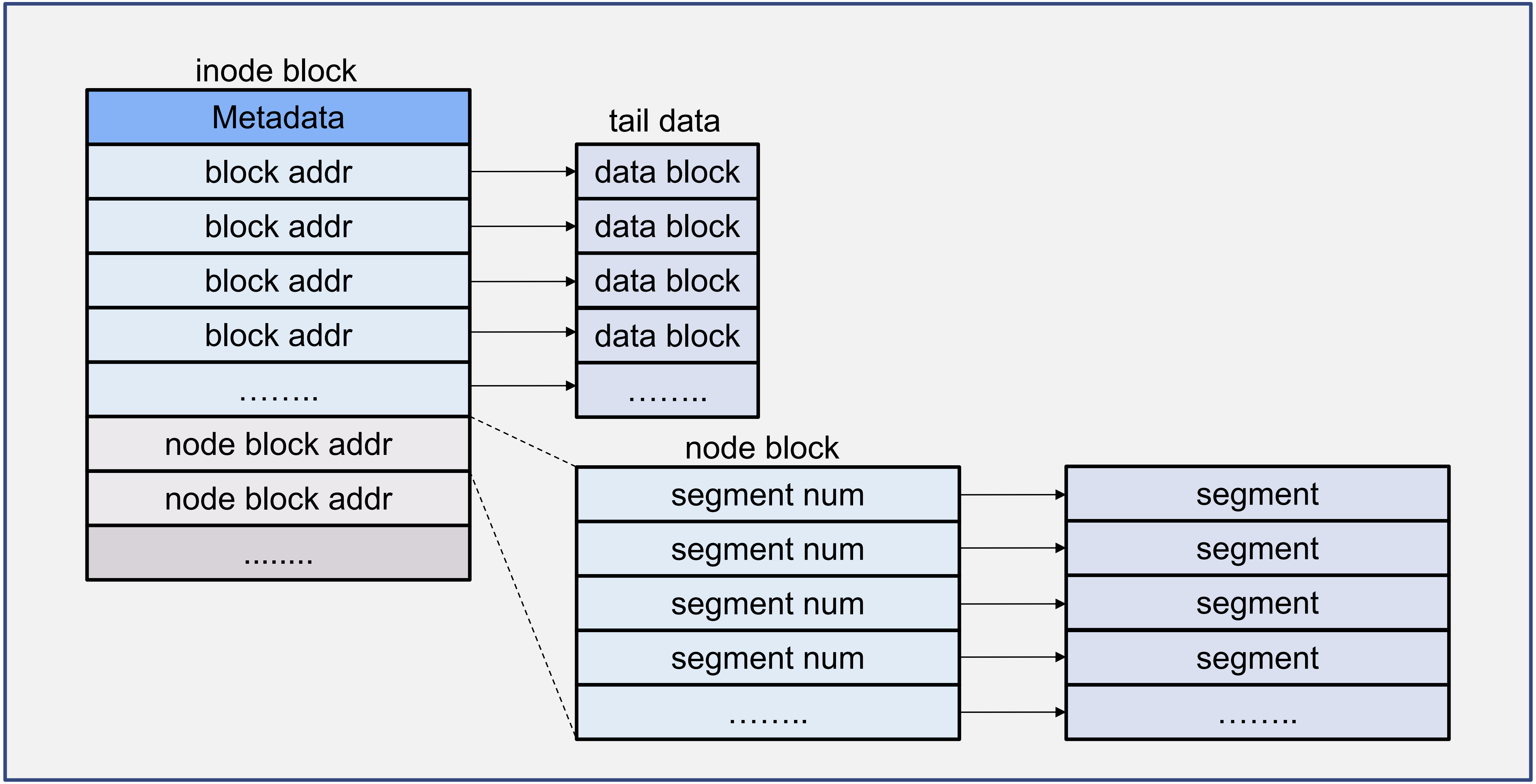
3. Next-Generation Compression - Segment-based Fixed length output



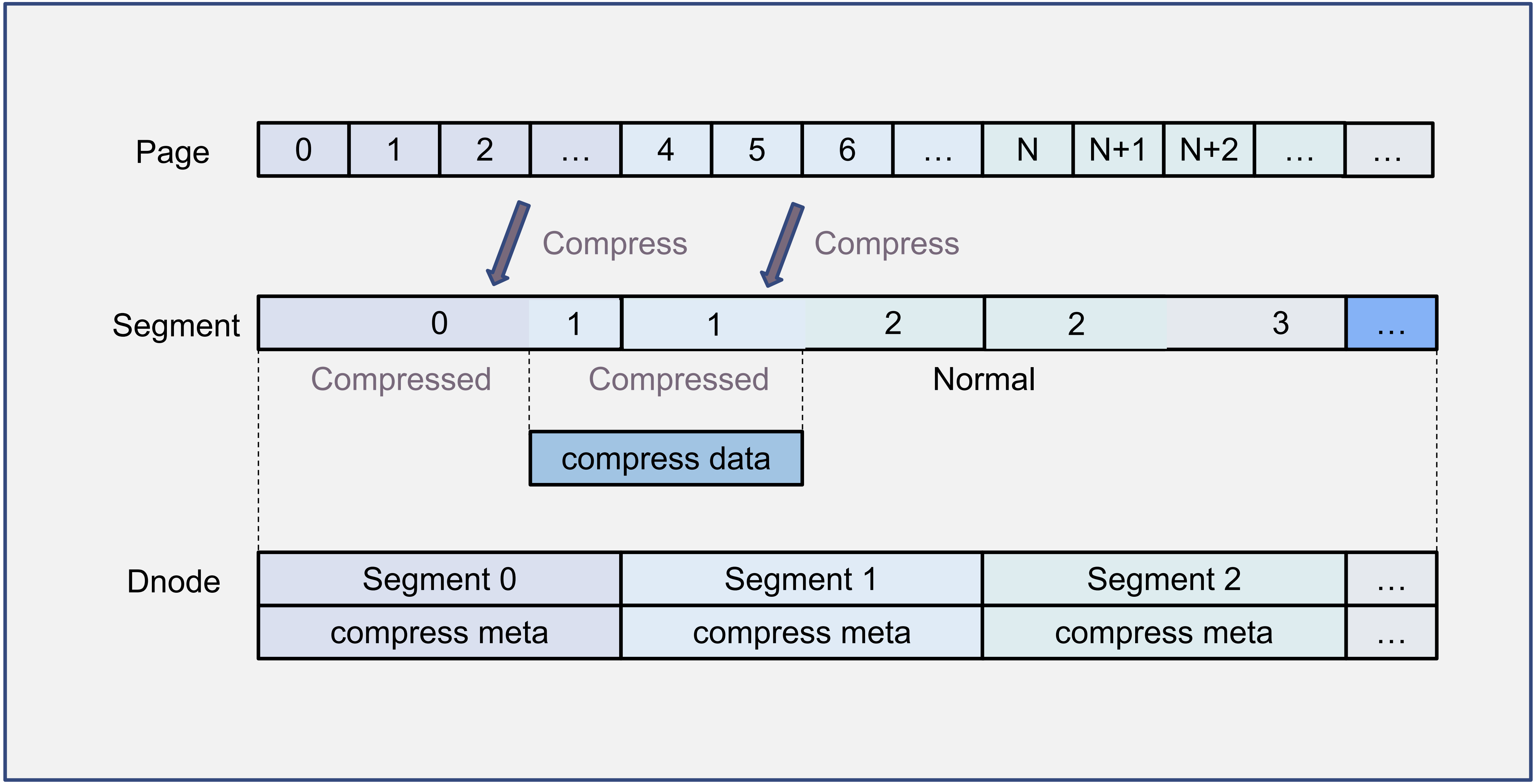
3. Next-Generation Compression - Segment-based Fixed length output



3. Next-Generation Compression - Segment-based Fixed length output



3. Next-Generation Compression - Segment-based Fixed length output



3. Next-Generation Compression - Segment-based Fixed length output

Segment-based Fixed length output Compression

- Significantly reduce the number of node blocks and improve space utilization
- Reduce file metadata IO by up to 99%
- No NAT table consumption
- Significantly reduce the amount of GC relocation data
- Suitable for large files read sequentially

THANK YOU