ALBERT-LUDWIGS-
UNIVERSITÄT FREIBURG

Algorithms and Methods for Distributed Storage Networks

4: Volume Manager and RAID

Christian Schindelhauer

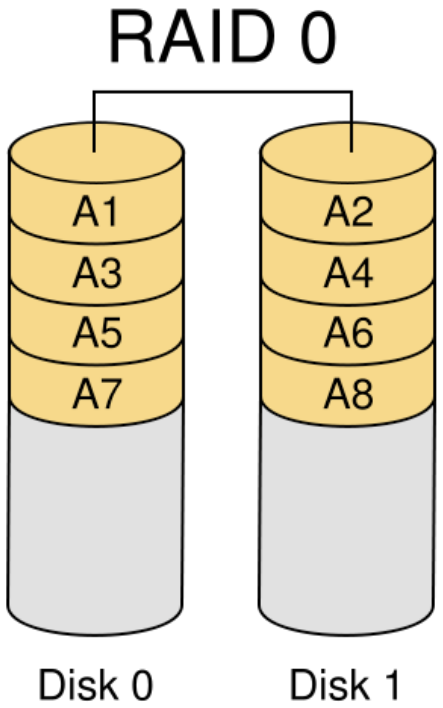
Albert-Ludwigs-Universität Freiburg
Institut für Informatik
Rechnernetze und Telematik
Wintersemester 2007/08

RAID

- **Redundant Array of Independent Disks**
  - Patterson, Gibson, Katz, „A Case for Redundant Array of Inexpensive Disks“, 1987
- **Motivation**
  - Redundancy
    - error correction and fault tolerance
  - Performance (transfer rates)
  - Large logical volumes
  - Exchange of hard disks, increase of storage during operation
  - Cost reduction by use of inexpensive hard disks

Raid 0

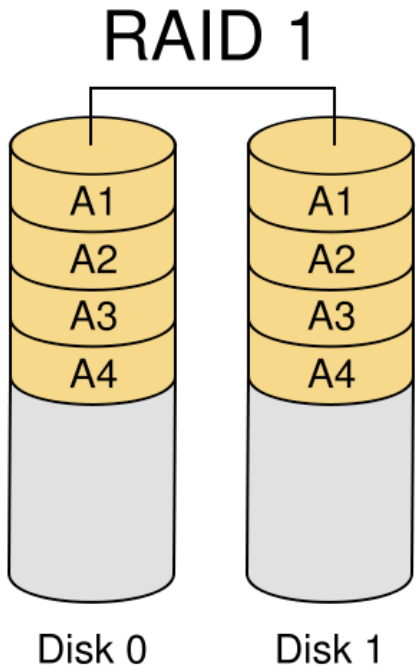
- **Striped set without parity**
  - Data is broken into fragments
  - Fragments are distributed to the disks
- **Improves transfer rates**
- **No error correction or redundancy**
- **Greater disk of data loss**
  - compared to one disk
- **Capacity fully available**



http://en.wikipedia.org/wiki/RAID

Raid 1

- **Mirrored set without parity**
  - Fragments are stored on all disks
- **Performance**
  - if multi-threaded operating system allows split seeks then
  - faster read performance
  - write performance slightly reduced
- **Error correction or redundancy**
  - all but one hard disks can fail without any data damage
- **Capacity reduced by factor 2**



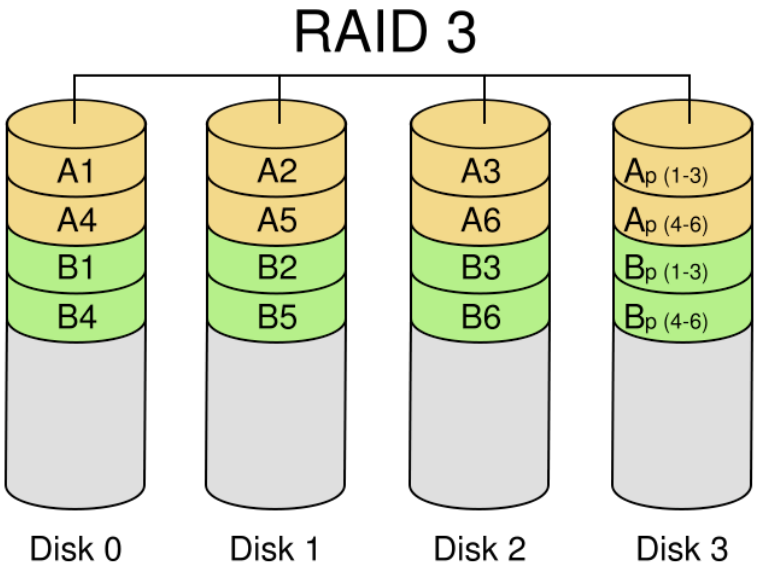
http://en.wikipedia.org/wiki/RAID

RAID 2

- **Hamming Code Parity**
- **Disks are synchronized and striped in very small stripes**
- **Hamming codes error correction is calculated across corresponding bits on disks and stored on multiple parity disks**
- **not in use**

Raid 3

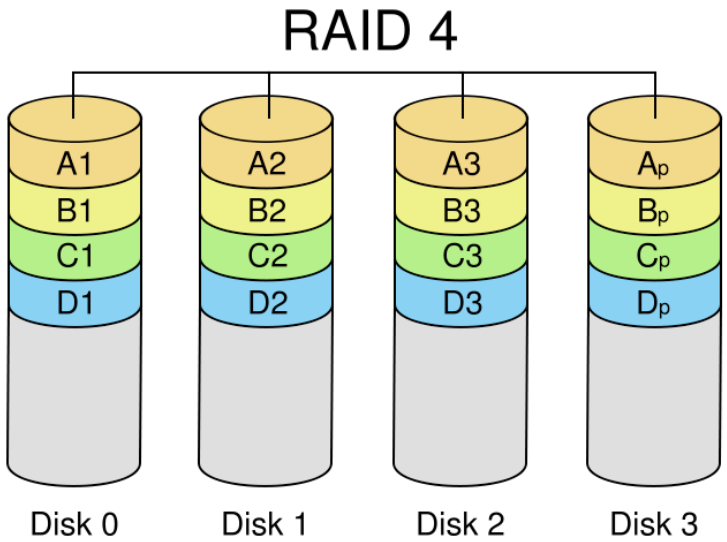
- **Striped set with dedicated parity (byte level parity)**
  - Fragments are distributed on all but one disks
  - One dedicated disk stores a parity of corresponding fragments of the other disks
- **Performance**
  - improved read performance
  - write performance reduced by bottleneck parity disk
- **Error correction or redundancy**
  - one hard disks can fail without any data damage
- **Capacity reduced by 1/n**



http://en.wikipedia.org/wiki/RAID

Raid 4

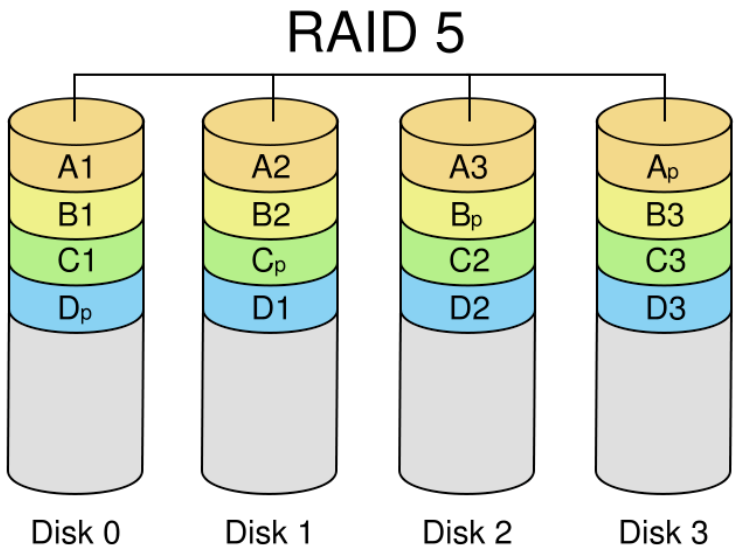
- **Striped set with dedicated parity (block level parity)**
  - Fragments are distributed on all but one disks
  - One dedicated disk stores a parity of corresponding blocks of the other disks on I/O level
- **Performance**
  - improved read performance
  - write performance reduced by bottleneck parity disk
- **Error correction or redundancy**
  - one hard disks can fail without any data damage
- **Hardly in use**



http://en.wikipedia.org/wiki/RAID

Raid 5

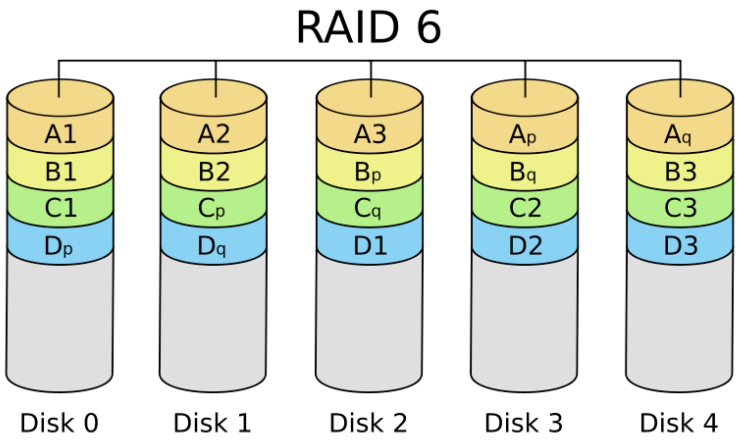
- **Striped set with distributed parity (interleave parity)**
  - Fragments are distributed on all but one disks
  - Parity blocks are distributed over all disks
- **Performance**
  - improved read performance
  - improved write performance
- **Error correction or redundancy**
  - one hard disks can fail without any data damage
- **Capacity reduced by 1/n**



http://en.wikipedia.org/wiki/RAID

Raid 5

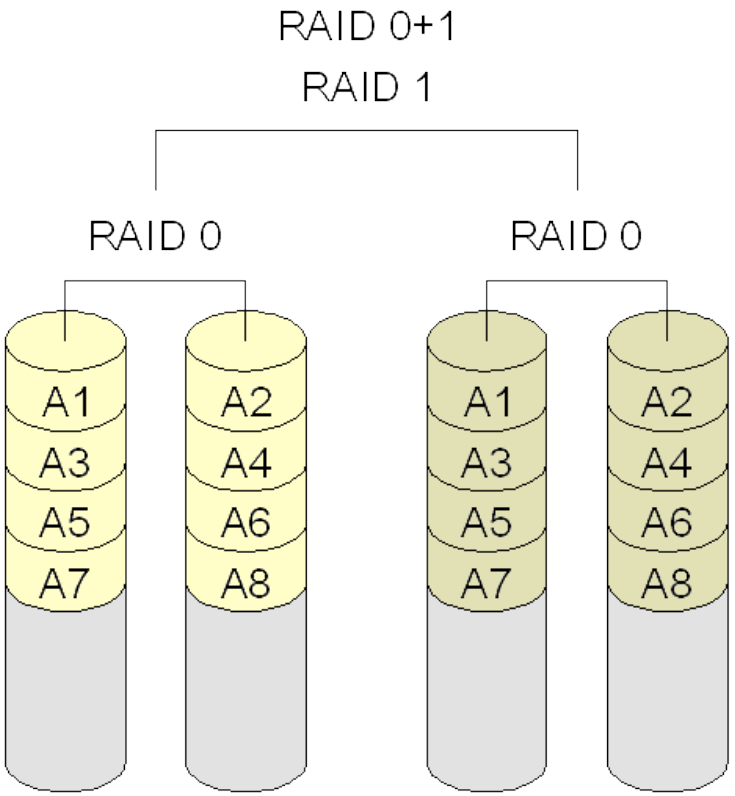
- **Striped set with dual distributed parity**
  - Fragments are distributed on all but two disks
  - Parity blocks are distributed over two of the disks
    - one uses XOR other alternative method
- **Performance**
  - improved read performance
  - improved write performance
- **Error correction or redundancy**
  - two hard disks can fail without any data damage
- **Capacity reduced by 2/n**



http://en.wikipedia.org/wiki/RAID

RAID 0+1

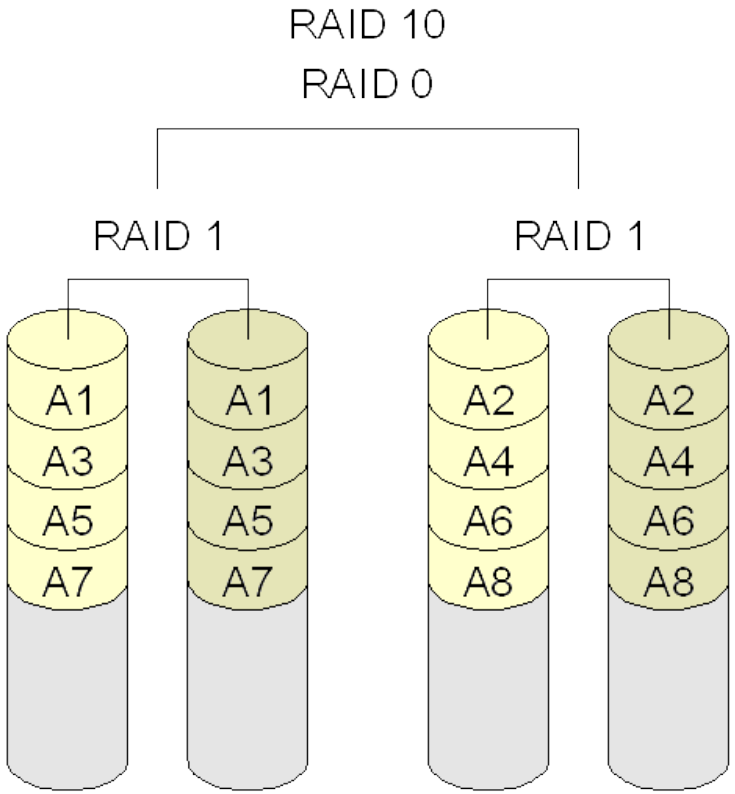
- **Combination of RAID 1 over multiple RAID 0**
- **Performance**
  - improved because of parallel write and read
- **Redundancy**
  - can deal with any single hard disk failure
  - can deal up to two hard disk failure
- **Capacity reduced by factor 2**



http://en.wikipedia.org/wiki/RAID

RAID 10

- **Combination of RAID 0 over multiple RAID 1**
- **Performance**
  - improved because of parallel write and read
- **Redundancy**
  - can deal with any single hard disk failure
  - can deal up to two hard disk failure
- **Capacity reduced by factor 2**



http://en.wikipedia.org/wiki/RAID

More RAIDs

- **More:**
  - RAIDn, RAID 00, RAID 03, RAID 05, RAID 1.5, RAID 55, RAID-Z, ...
- **Hot Swapping**
  - allows exchange of hard disks during operation
- **Hot Spare Disk**
  - unused reserve disk which can be activated if a hard disk fails
- **Drive Clone**
  - Preparation of a hard disk for future exchange indicated by S.M.A.R.T

Volume Manager

- **Volume manager**
  - aggregates physical hard disks into virtual hard disks
  - breaks down hard disks into smaller hard disks
  - Does not provide operating system, but enables it
- **Can provide**
  - resizing of volume groups by adding new physical volumes
  - resizing of logical volumes
  - snapshots
  - mirroring or striping, e.g. like RAID1
  - movement of logical volumes

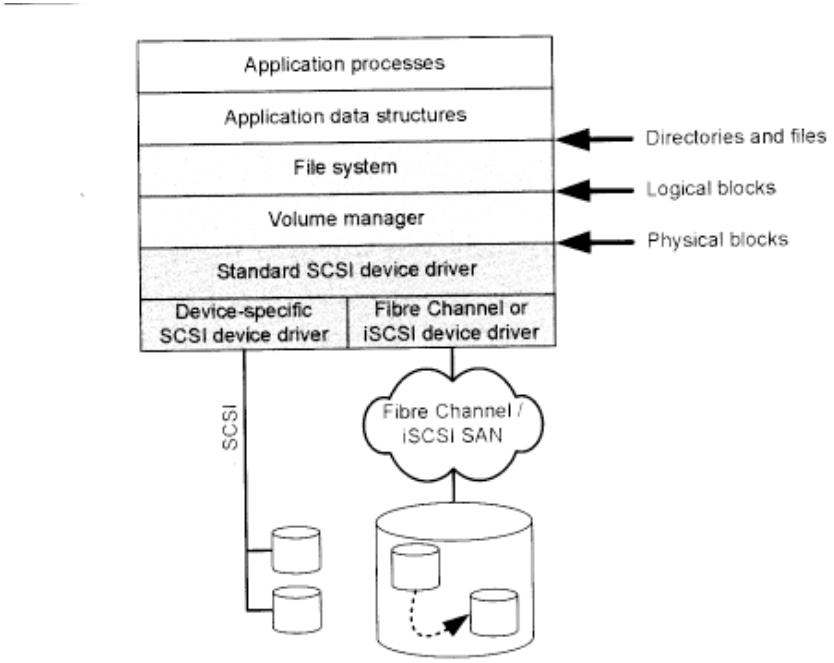


Figure 4.1 File system and volume manager manage the blocks of the block-oriented hard disks. Applications and users thus use the storage capacity of the disks via directories and files

From: Storage Networks Explained, Basics and Application of Fibre Channel SAN, NAS, iSCSI and InfiniBand, Troppens, Erkens, Müller, Wiley

Overview of Terms

- **Physical volume (PV)**
  - hard disks, RAID devices, SAN
- **Physical extents (PE)**
  - Some volume managers splite PVs into same-sized physical extents
- **Logical extent (LE)**
  - physical extents may have copies of the same information
  - are addresed as logical extent
- **Volume group (VG)**
  - logical extents are grouped together into a volume group
- **Logical volume (LV)**
  - are a concatenation of volume groups
  - a raw block devices
  - where a file system can be created upon

ALBERT-LUDWIGS-
UNIVERSITÄT FREIBURG

Algorithms and Methods for Distributed Storage Networks

4: Volume Manager and RAID

Christian Schindelhauer

Albert-Ludwigs-Universität Freiburg
Institut für Informatik
Rechnernetze und Telematik
Wintersemester 2007/08