md-RAID

complex partitioning scheme in Linux

«md» name is used for

- Linux RAID driver
- Linux RAID metadata

md-RAID can consist of:

- physical devices
- partitions
- any Linux block devices

md-RAID is very common in NASes

md-RAID metadata

Superblock – the only MD metadata

- occupies 1 KB
- is stored in the beginning or in the end of each member disk
- several versions: 0.9 or 1.x

Superblock contains

- RAID information – the same on all the disks
  - Number of disks in RAID
  - RAID level
  - RAID layout for RAID5/6 (left/right, symmetric/asymmetric)
  - Block size
- Disk information – unique to each disk
  - Column height
  - Superblock generation number (later updates have higher number)
  - Disk role (disk number in RAID)

md-RAID failure scenario

Common NAS failure scenario

1. One disk has failed
2. Rebuild starts
3. Some other disk develops a bad block or transient failure
4. Rebuild stops; faulty disk is dropped from the array
    - some NASes spoil MD superblocks on the good disks
    - some NASes can destroy even GPT partitions on the good disks
    - a client trying to make the NAS do the rebuild may worsen the situation
5. Now the array is missing two disks, RAID5 offline

md-RAID Recovery

Basic considerations

- Sort all the superblocks by age

  earlier superblocks provide more information about RAID

- Search for more superblocks

  if NAS was reset, previous superblocks can still exist

Recovery approaches

1. Editing superblocks and GPT on clones
   - copy a healthy superblock and GPT
   - edit the role
   - try to read data
2. RAID recovery first, then file recovery
   - create regions based on MBR/GPT partitions
   - take RAID block size and RAID level from md superblock