



RAID: Redundant Arrays of Independent Disks

- disk organization techniques that manage a large numbers of disks, providing a view of a single disk of
 - high capacity and high speed by using multiple disks in parallel,
 high reliability by storing data redundantly, so that data can be recovered even if a disk fails

Improvement of Reliability via Redundancy

- Redundancy store extra information that can be used to rebuild information lost in a disk failure
- E.g., Mirroring (or shadowing)
 - Duplicate every disk. Logical disk consists of two physical disks.
 - Every write is carried out on both disks
 - Reads can take place from either disk
 - If one disk in a pair fails, data still available in the other
 - Data loss would occur only if a disk fails, and its mirror disk also fails before the system is repaired
 - Probability of combined event is very small
 - Except for dependent failure modes such as fire or building collapse or electrical power surges
- Mean time to data loss depends on mean time to failure, and mean time to repair
 - E.g. MTTF of 100,000 hours, mean time to repair of 10 hours gives mean time to data loss of 500*10⁶ hours (or 57,000 years) for a mirrored pair of disks (ignoring dependent failure modes)

Improvement in Performance via Parallelism

- Two main goals of parallelism in a disk system:
 - 1. Load balance multiple small accesses to increase throughput
 - 2. Parallelize large accesses to reduce response time.
- Improve transfer rate by striping data across multiple disks.
- Bit-level striping split the bits of each byte across multiple disks
 - In an array of eight disks, write bit i of each byte to disk i.
 - Each access can read data at eight times the rate of a single disk.
 - But seek/access time worse than for a single disk
 - Bit level striping is not used much any more
- Block-level striping with n disks, block i of a file goes to disk (i mod n) + 1



RAID Levels

- Schemes to provide redundancy at lower cost by using disk striping combined with parity bits
 - Different RAID organizations, or RAID levels, have differing cost, performance and reliability characteristics
 - RAID Level 0: Block striping; non-redundant.
 - Used in high-performance applications where data loss is not critical.
 - RAID Level 1: Mirrored disks with block striping
 - Offers best write performance.
 - Popular for applications such as storing log files in a database system.



(a) RAID 0: nonredundant striping



(b) RAID 1: mirrored disks

- RAID Level 2: Memory-Style Error-Correcting-Codes (ECC) with bit striping.
- RAID Level 3: Bit-Interleaved Parity
 - a single parity bit is enough for error correction, not just detection, since we know which disk has failed
 - When writing data, corresponding parity bits must also be computed and written to a parity bit disk
 - To recover data in a damaged disk, compute XOR of bits from other disks (including parity bit disk)



(c) RAID 2: memory-style error-correcting codes



(d) RAID 3: bit-interleaved parity

10,25

CSilberschatz, Korth and Sudarshan

Database System Concepts - 6th Edition



RAID Levels (Cont.)

RAID Level 3 (Cont.)

- Faster data transfer than with a single disk, but fewer I/Os per second since every disk has to participate in every I/O.
- Subsumes Level 2 (provides all its benefits, at lower cost).
- RAID Level 4: Block-Interleaved Parity; uses block-level striping, and keeps a parity block on a separate disk for corresponding blocks from N other disks.
 - When writing data block, corresponding block of parity bits must also be computed and written to parity disk
 - To find value of a damaged block, compute XOR of bits from corresponding blocks (including parity block) from other disks.



(e) RAID 4: block-interleaved parity

CSilberschatz, Korth and Sudarshan

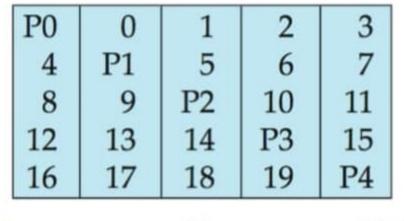


RAID Levels (Cont.)

- RAID Level 5: Block-Interleaved Distributed Parity; partitions data and parity among all N + 1 disks, rather than storing data in N disks and parity in 1 disk.
 - E.g., with 5 disks, parity block for nth set of blocks is stored on disk (n mod 5) + 1. with the data blocks stored on the other 4 disks.



(f) RAID 5: block-interleaved distributed parity



Database System Concepts - 6th Edition



RAID Levels (Cont.)

10.29

RAID Level 5 (Cont.)

- Higher I/O rates than Level 4.
 - Block writes occur in parallel if the blocks and their parity blocks are on different disks.
- Subsumes Level 4: provides same benefits, but avoids bottleneck of parity disk.
- RAID Level 6: P+Q Redundancy scheme; similar to Level 5, but stores extra redundant information to guard against multiple disk failures.
 - Better reliability than Level 5 at a higher cost; not used as widely.



⁽g) RAID 6: P + Q redundancy