BITS, PILANI – K. K. BIRLA GOA CAMPUS

Database Systems

(IS F243)

by

Mrs. Shubhangi Gawali

Dept. of CS and IS

RAID Levels

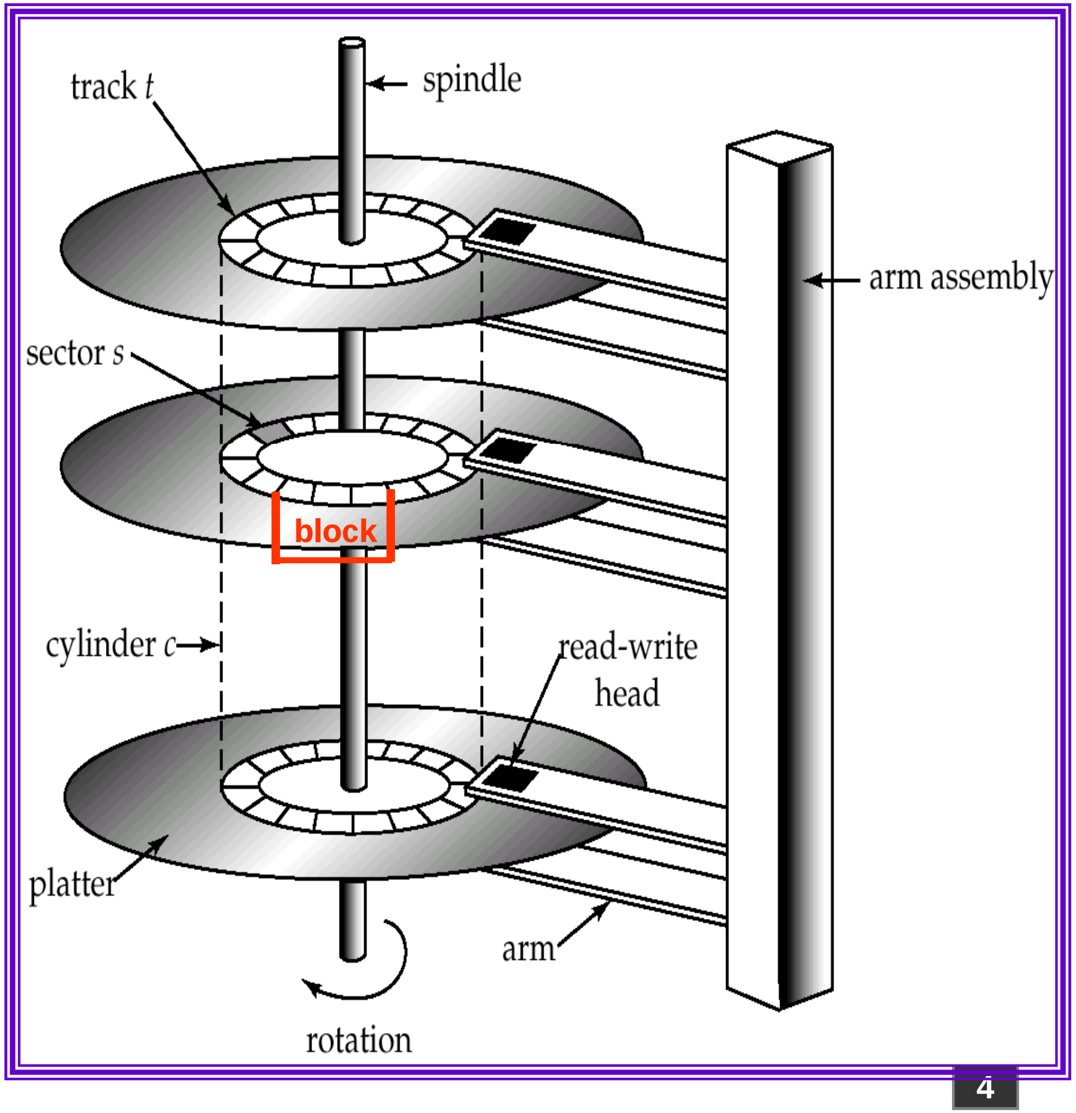
(courtesy : The University of Sydney)

Storage and Indexing

(courtesy : The University of Sydney)

Components of a Disk

- The platters spin (say, 120rps).
- The arm assembly is moved in or out to position a head on a desired track. Tracks under heads make a cylinder (imaginary!).
- Only one head reads/writes at any one time.
- Block size is a multiple of sector size (which is fixed).



Accessing a Disk Page

- Time to access (read/write) a disk block:
  - seek time (moving arms to position disk head on track)
  - rotational delay (waiting for block to rotate under head)
  - transfer time (actually moving data to/from disk surface)
- Seek time and rotational delay dominate.
  - Seek time varies from about 1 to 20msec
  - Rotational delay varies from 0 to 10msec
  - Transfer rate is about 1msec per 4KB page
- Key to lower I/O cost: reduce seek/rotation delays! Hardware vs. software solutions?

RAID

- Disk Array: arrangement of several disks to increase performance and improve reliability of storage system.

- RAID: Redundant Arrays of Independent Disks
  - Data striping + redundancy

- Data striping
  - distribute data over several disks
    - High capacity and high speed
  - the more disk,, the lower reliability
    - e.g., a system with 100 disks, each with MTTF of 100,000 hours (approx. 11 years), will have a system MTTF of 1000 hours (approx. 41 days)

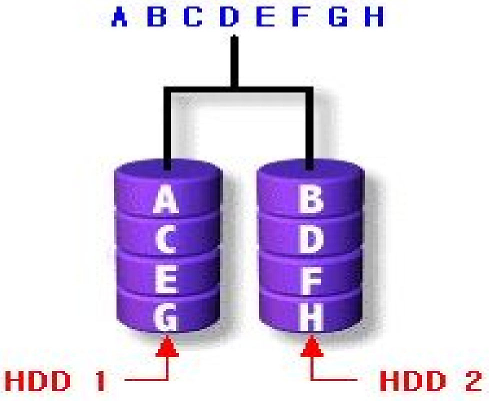
- Redundancy
  - redundant information is maintained
    - high reliability by storing data redundantly, so that data can be recovered even if a disk fails

Wrap-Up

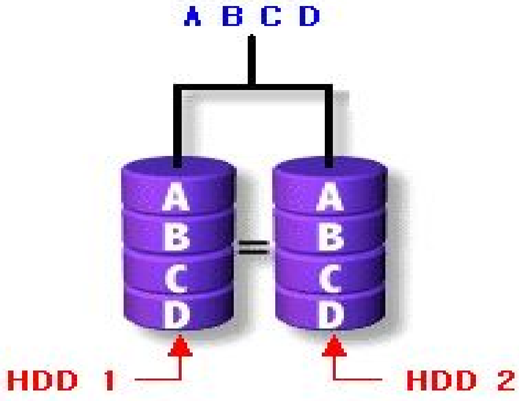
- Storage
  - Disk
  - Buffer management
  - File organization
- Indexing
  - Tree-structured Indexing
  - Hash-based Indexing

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 lost 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.



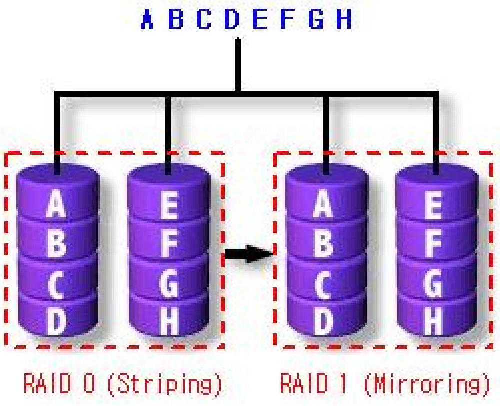
RAID 0: nonredundant striping



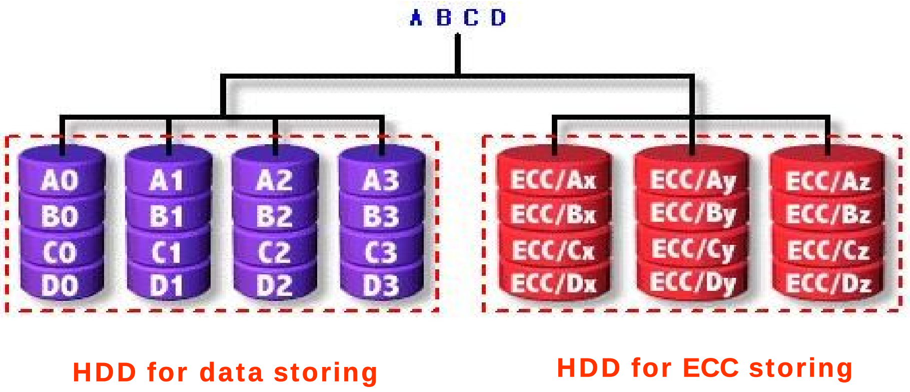
RAID 1: mirrored disks

RAID Levels (Cont.)

- RAID Level 0+1: Striping and Mirroring
  - Parallel reads, a write involves two disks.
- RAID Level 2: Memory-Style Error-Correcting-Codes (ECC) with bit striping.
  - Striping unit is single bit
  - Store code for error correcting



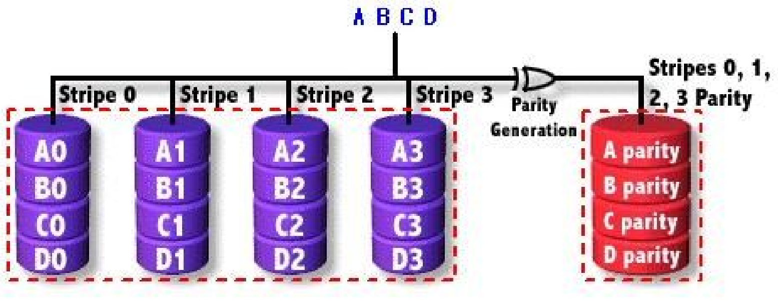
RAID 0+1: striping and mirroring



RAID 2: error correcting codes

RAID Levels (Cont.)

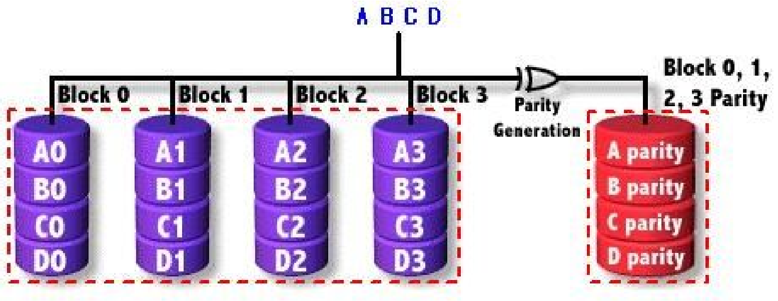
- RAID Level 3: Bit-Interleaved Parity
  - a single parity bit is enough for error correction, since we know which disk has failed
    - When writing data, corresponding parity bits must also be computed and written to a parity bit disk
- RAID Level 4: Block-Interleaved Parity;
  - uses block-level striping, and keeps a parity block on a separate disk for corresponding blocks from N other disks.



HDD for data storing

HDD for parity storing

RAID 3: bit-interleaved parity



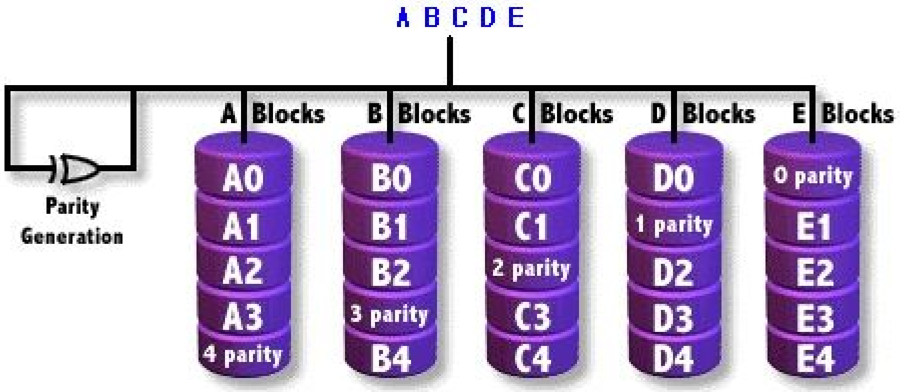
HDD for data storing

HDD for parity storing

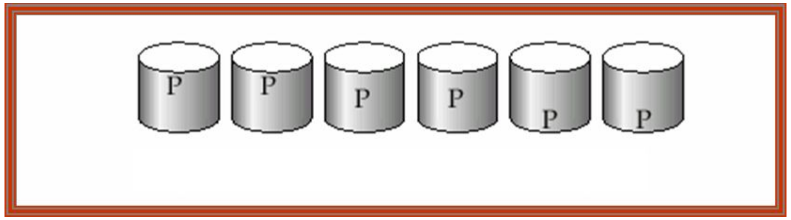
RAID 4: block-interleaved parity

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.
- 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.



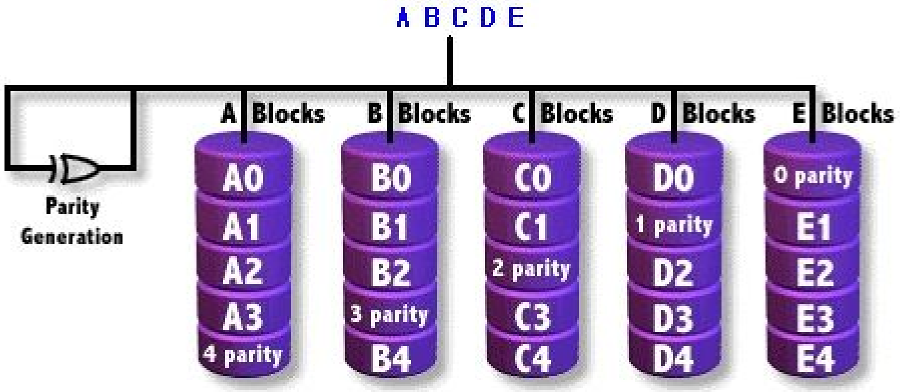
RAID 5: block-interleaved distribute parity



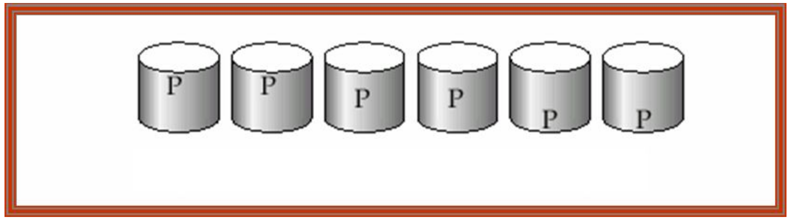
RAID 6: P+Q redundancy schem

Example of RAID Levels

- 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.
- 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.



RAID 5: block-interleaved distribute parity



RAID 6: P+Q redundancy schem

Choice of RAID Level

- Factors in choosing RAID level
  - Monetary cost
  - Performance: # of I/Os per second and bandwidth during normal operation
  - Performance during failure
  - Performance during rebuild of failed disk / time to rebuild failed disk
- RAID 0 is used only when data safety is not important
  - e.g. data can be recovered quickly from other sources
- Level 2 and 4 never used since they are subsumed by 3 and 5
- Level 3 is not used anymore since bit-striping forces single block reads to access all disks, wasting disk arm movement, which block striping (level 5) avoids
- Level 6 is rarely used since levels 1 and 5 offer adequate safety for almost all applications
- So competition is between 1 and 5 only
  - Level 5 is preferred for applications with low update rate, and large amounts of data
  - Level 1 is preferred for all other applications