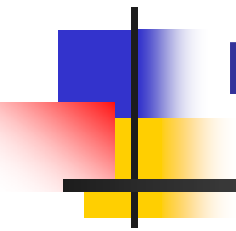
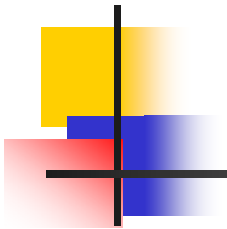
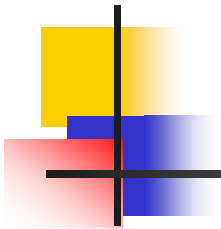


Redundant Array of Inexpensive Disks (RAID)



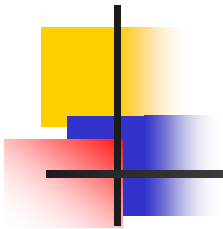
Announcement

- **A delegation of Swedish Students will be visiting the University today.**
- **The Dean, Students Affairs, has extended an invitation to our students to participate in a Lecture at 1:00 PM in Bldg. 20/103 for a duration of 1 (one) hour.**
- **There is a lunch organized in Bldg. 9 right after the lecture at 2:00 PM.**
- **Our students participating in this lecture and lunch will be given official excuses by the Student Affairs.**



Objective

- Introduction +
- Data Striping +
- Reliability and Performance Improvements +
- RAID Organization Levels +
- Choice of RAID Level +

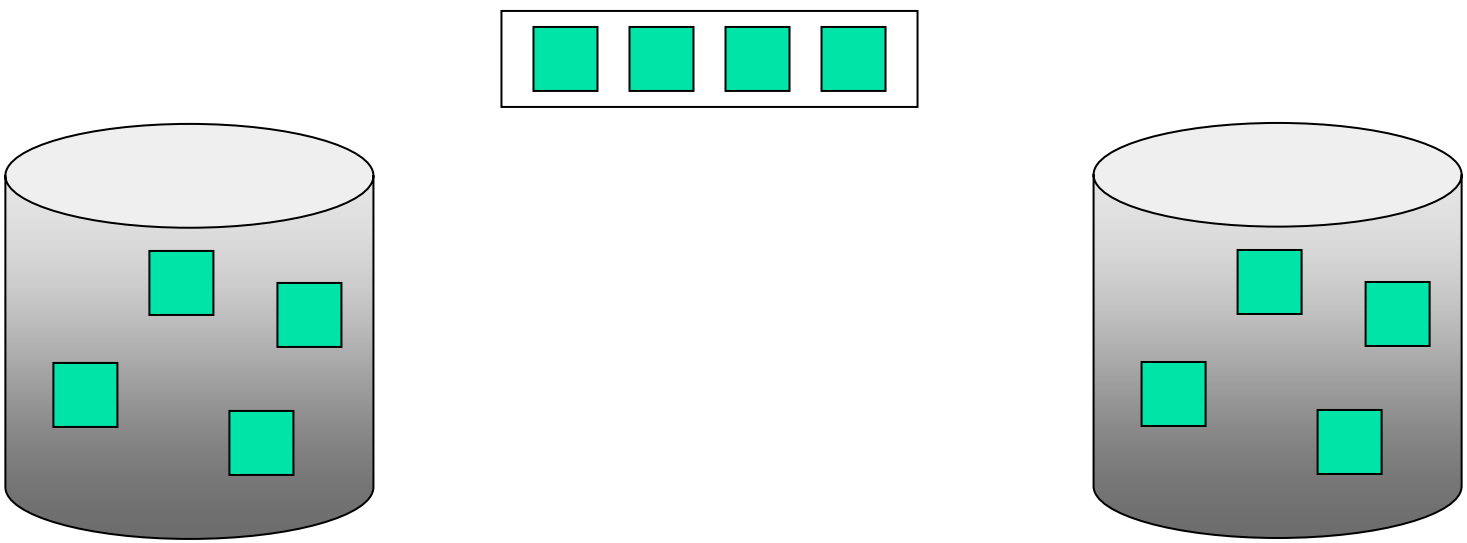


- Introduction

- RAID is a storage technology and stands for Redundant Array of Inexpensive Disks.
- RAID is a set of arrays of physical disk drives that appear to the DB users (and programmers) as if they form one large logical unit.
- RAID does not require any change in the application programs or DB use.
- The sole purpose of RAID is to improve performance and Reliability.
- RAID comes with seven levels: level 0, Level 1, ..., level 6.

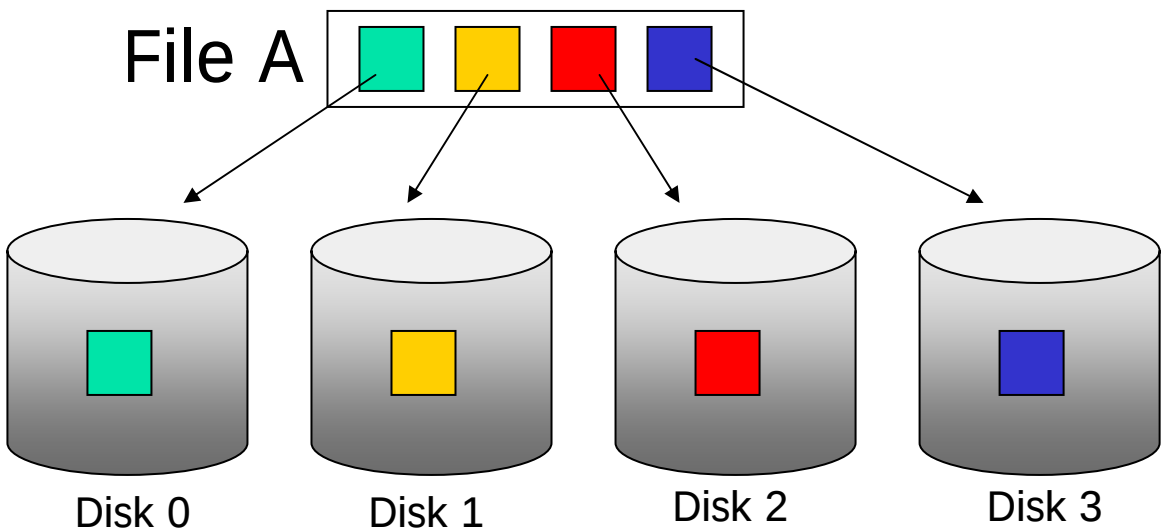
- Data Striping...

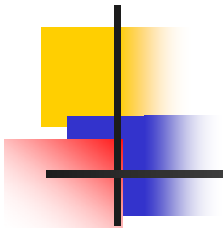
- Data striping is the saving of a file in multiple disks.
- The following figure shows the striping of file A into 4 disks.



- Data Striping...

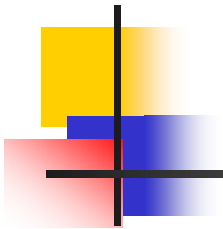
- Data striping is the saving of a file in multiple disks.
- The following figure shows the striping of file A into 4 disks.





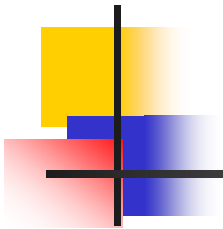
... - Data Striping...

- Data stripping uses parallelism to improve the disk performance. In the diagram on the previous slide, suppose the FILE A consists of four disk blocks. Now if FILE A is stored on one disk and we want to read the whole file, then we need four read operations to read the file. But if the file is stored in four independent disks, one block per disk, and the four disks behave like a single logical disk, then all four blocks can be read in parallel and it will take time equal to one read operation.
- Because data is stripped on to multiple disks, so some kind of parity check can be used to improve the reliability of data very similar to the parity bit used in the main memory.



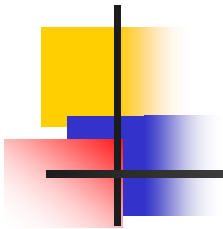
... - Data Striping

- The size of the partition in data striping is called the **striping unit**. The two extreme possible striping units can be bit or block.
- The main purpose of striping is to accomplish the balancing of workload across the disk drives.
- As an example, if striping unit is a disk block and the data is striped on to four disk drives, then reading data from these four disk drives can be done in parallel and the read can be balanced between the four disks.



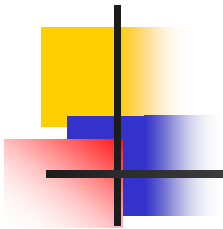
- Reliability and Performance Improvements ...

- Improving reliability with RAID
  - One way to increase reliability of a DB when using RAID is by introducing redundancy, i. e., store one data item in more than one place.
  - One technique for introducing redundancy is called mirroring or shadowing, i. e. store data redundantly on two identical physical disks that are treated as one logical disk.
  - Incase of mirrored data, a data item can be read from any disk but for writing a data item, it must be written on both disks.
  - If one disk fails, the other disk is still there to continuously provide the data. It improves the reliability.



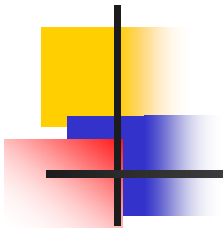
... - Reliability and Performance Improvements

- Assume data striping at block level is used and rather than storing data on one disk, it is stored on four disks. Reading one whole file in the later case will take one-fourth of the time needed to read the whole file in the former case.
- Granularity (the size of data used for stripping) can be reduced to improve the performance even more. For example, by using 8 disks, bit-level data stripping can be used. In this case, one byte of data will be read from 8-disks (one bit from each disk) in parallel.



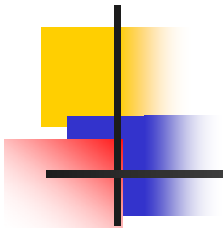
- RAID Organization and Levels ...

- Array of disk drives that implement a combination of data stripping and redundancy are called RAID.
- Several RAID organizations, referred to as RAID levels have been proposed. Each RAID level represents a different trade-off between reliability and performance.
- Most multi-user operating systems today including Microsoft NT, UNIX, and Novel NetWare support some forms of RAID.
- RAID has one important risk: the increased likelihood of a disk drive failure across the whole DB. This risk can be covered by introducing an appropriate level of redundancy, i.e. by choosing an appropriate RAID level.



... - RAID Organization and Levels

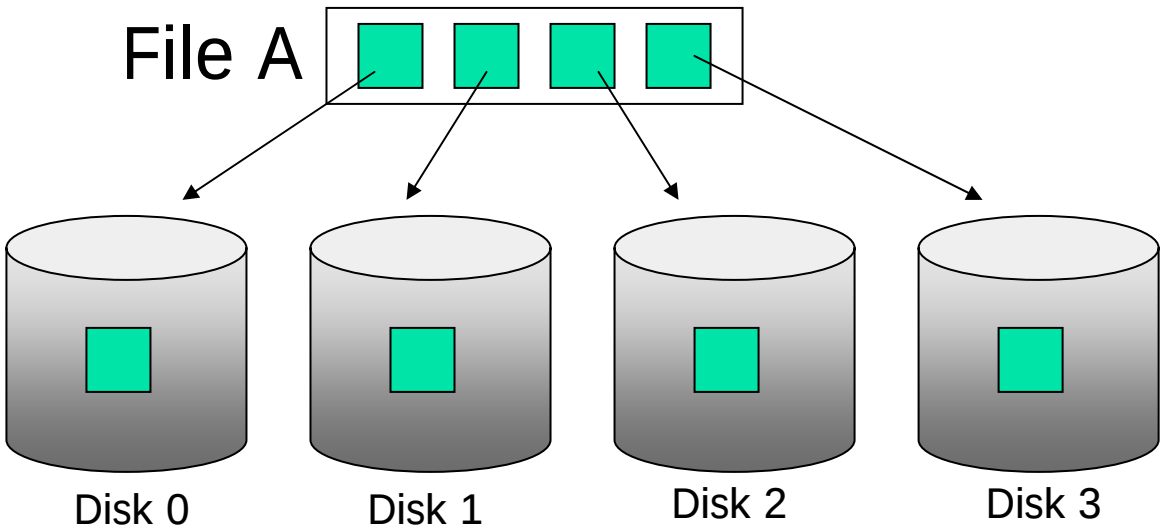
- RAID levels depend on the data redundancy introduced and correctness checking technique used in the scheme.
  - **Level 0**: No redundancy and No correctness checking.
  - **Level 1**: Redundancy through mirroring but no correctness checking.
  - **Level 2**: Mirroring or no mirroring combined with memory like correctness checking. For example, using parity bit. Various versions of Level 2 are possible.
  - **Level 3**: like Level 2 but uses single disk for parity.
  - **Level 4**: Block level data stripping and parity like Level 3.
  - **Level 5**: Block level data stripping but data and parity are distributed across all disks.
  - **Level 6**: Just two redundant disks with P + Q redundancy scheme.

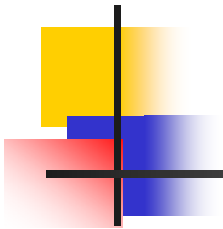


-- RAID Level 0

- Has no redundant data.
- Has the best write performance since updates do not have to be duplicated.
- Its read performance is better than non-RAID storage but poorer than all other RAID levels.
- The following diagram shows RAID Level 0 with array of four disks.

--- RAID Level 0

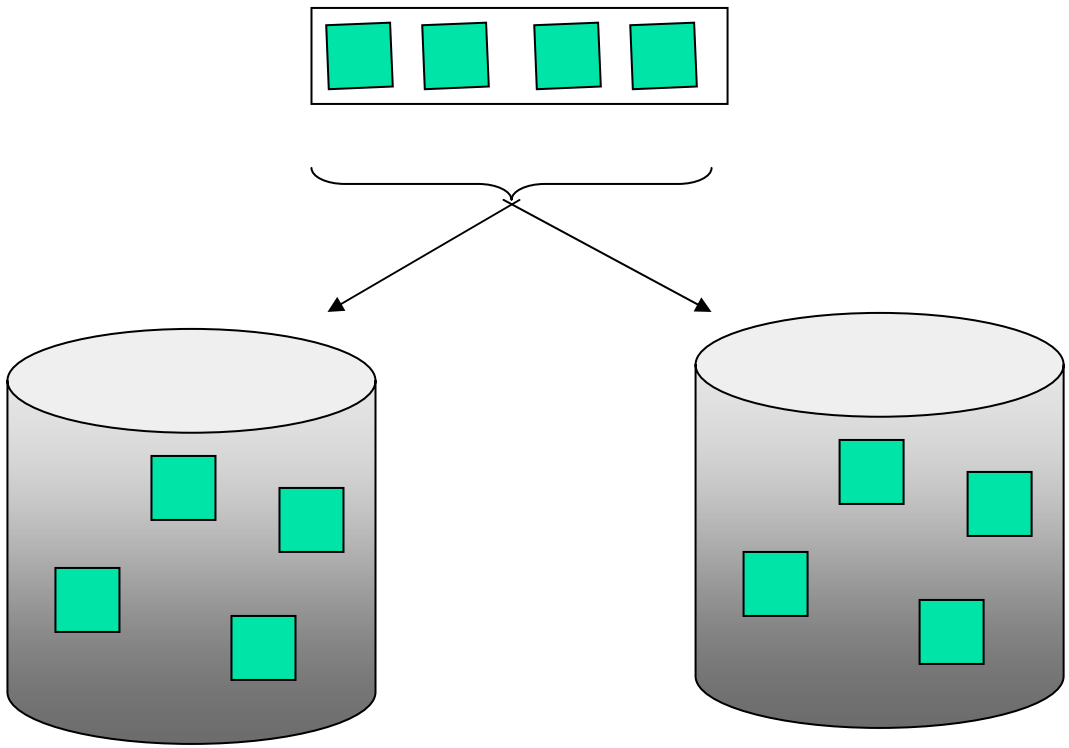


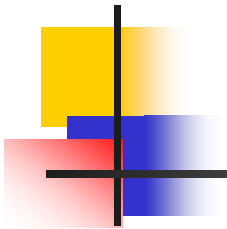


-- RAID Level 1

- Has fully redundant data, i.e., the two identical copies of the data on two different arrays of disks are maintained. This type of redundancy is called mirroring.
- The following diagram shows RAID Level 1 with array of four disks and the data is mirrored on the other four disks
- In RAID Level 1, every write of a disk block involves a write on both disks.
- In RAID Level 1, reading of data can be done from either disk. Incase of multiple blocks data can be read from both disks in parallel.

--- RAID Level 1: Mirroring

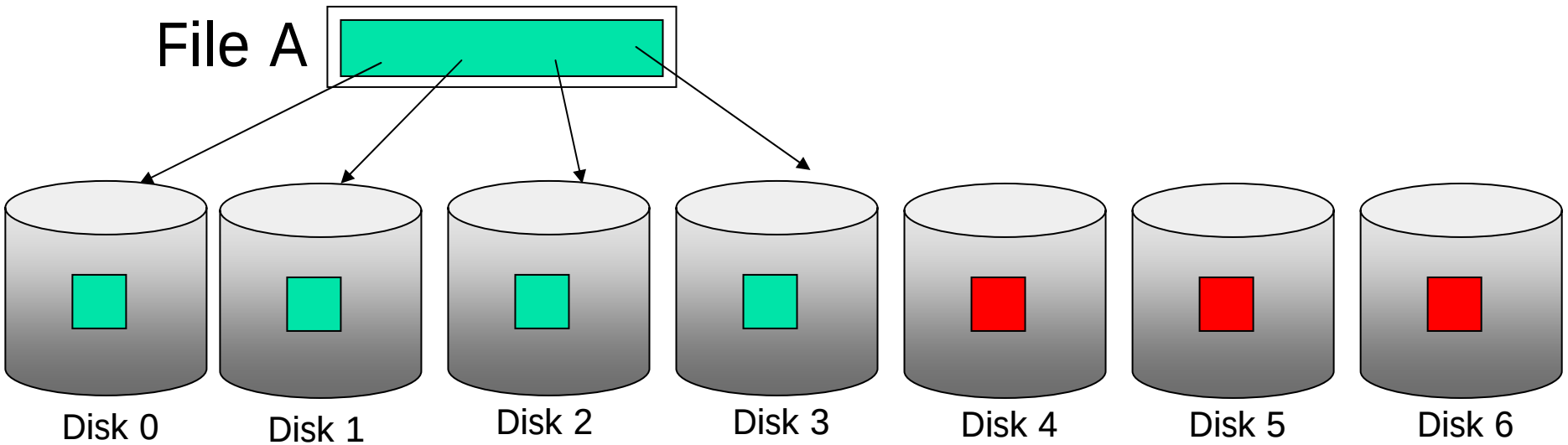


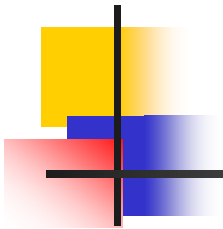


-- RAID Level 2

- In RAID Level 2, some disks are used with data stripping for parallel data processing and other disks are dedicated for stirring error correction codes. Data is stored without any redundancy.
- RAID Level 2 bit-level data stripping.
- Error correction codes are used to detect errors in the stripes and to reconstruct data when pages within the data stripes are damaged.
- The following diagram shows RAID Level 2 with array of four disks used for storing data and three disks used for storing error correction codes.

--- RAID Level 2

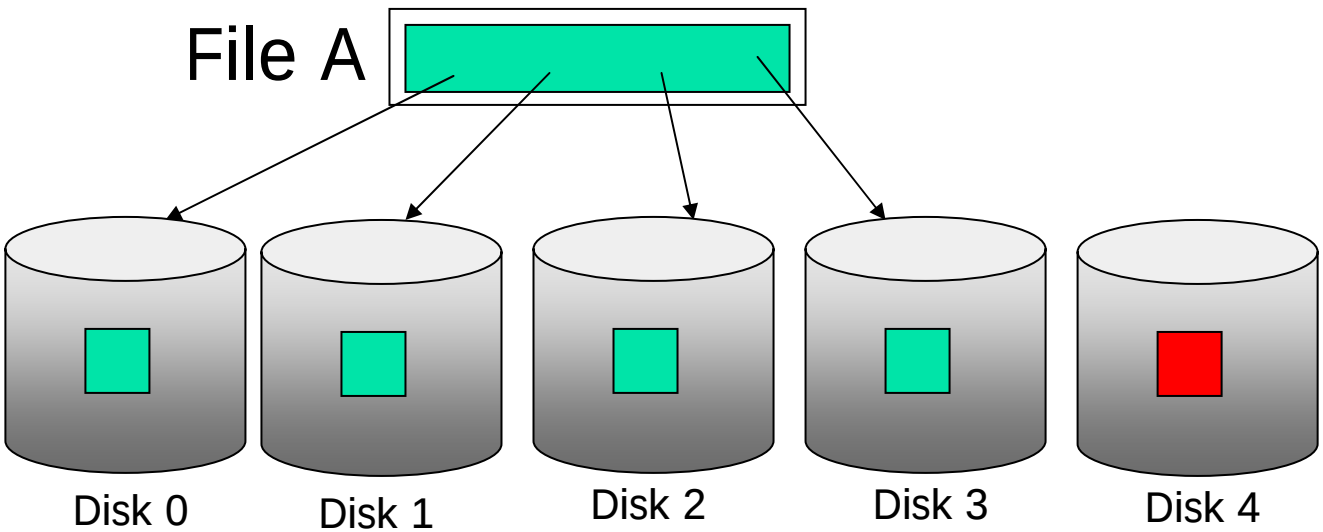


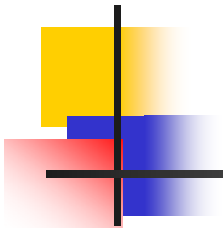


-- RAID Level 3

- In RAID Level 3, data stripes are stored across drives, and only one drive is used to store stripe level error correction codes.
- RAID level 3 uses bit-level stripping.
- The following diagram shows RAID Level 3 with an array of four disks used for storing data and one disk used for storing error correction codes.

--- RAID Level 3

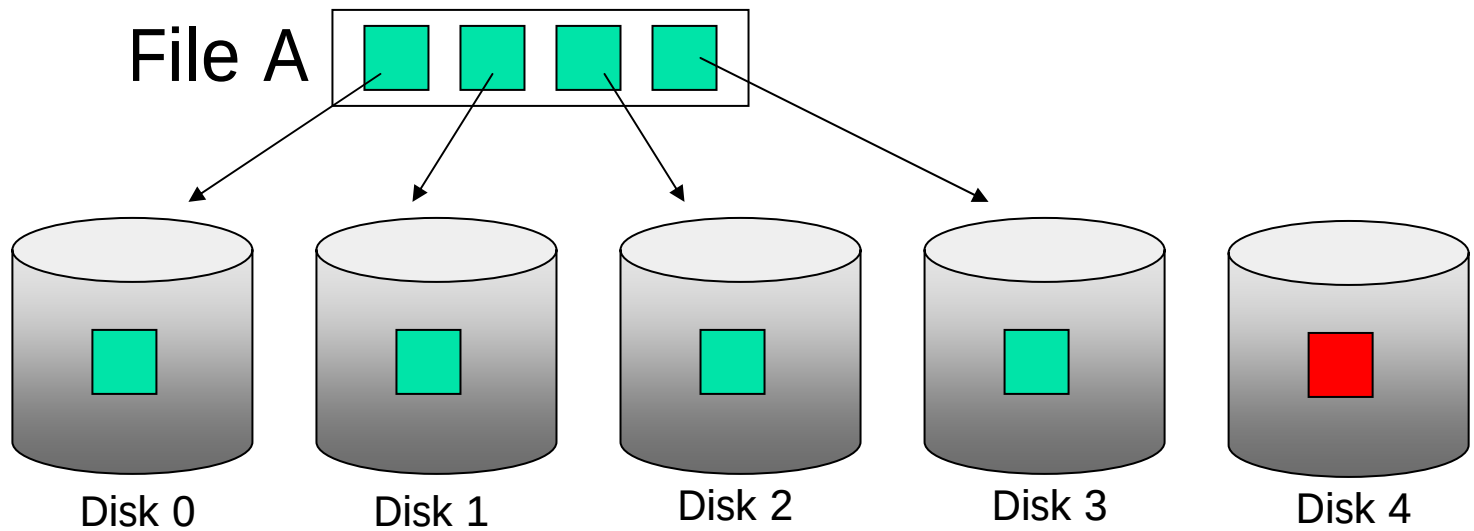


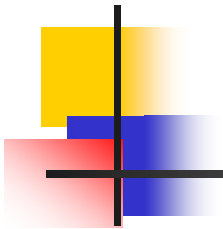


-- RAID Level 4

- In RAID Level 4, data stripes are stored across drives, and only one drive is used to store stripe level error correction codes.
- RAID level 3 uses lock-level stripping.
- The following diagram shows RAID Level 4 with an array of four disks used for storing data and one disk used for storing error correction codes.

--- RAID Level 4

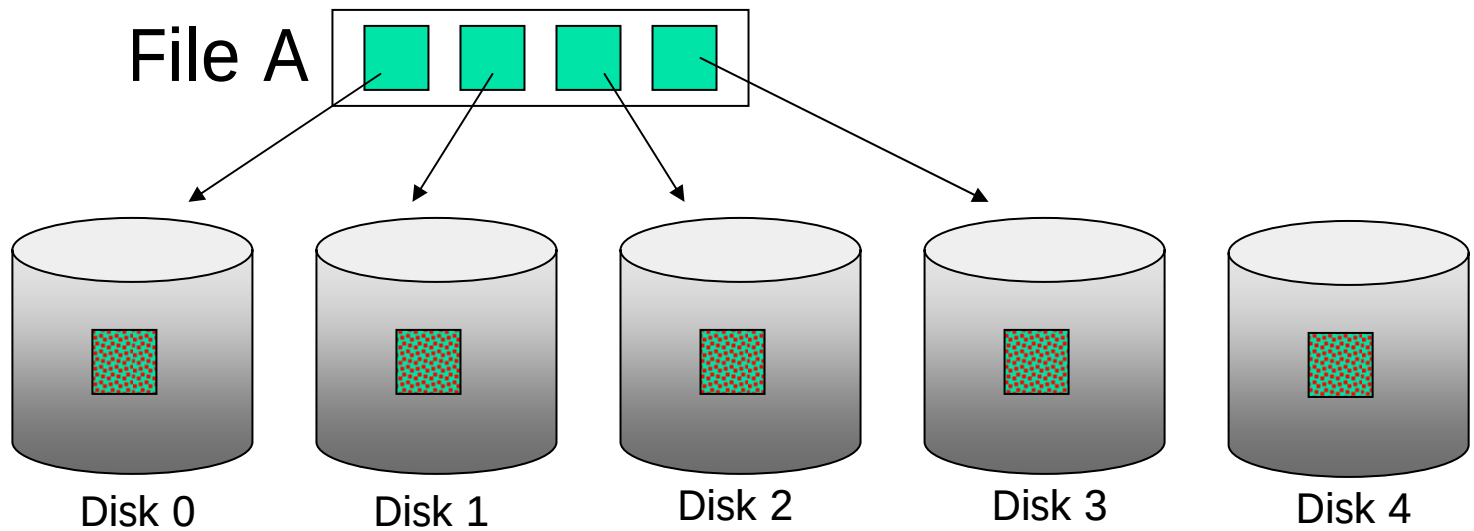


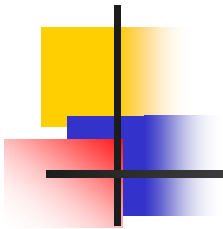


-- RAID Level 5

- In RAID Level 5, each drive contains both data and error correction codes.
- RAID level 5 uses lock-level stripping.
- In RAID Level 5 read operation can be done in parallel in all drives. A write operation will access one or more drives to write the new record plus the error-correction code.
- The following diagram shows RAID Level 5 with an array of five disks used for storing data and error correction codes.

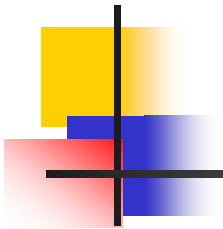
--- RAID Level 5





-- RAID Level 6

- Requires two check disks, but also uniformly distributes redundant data using block-level stripping.
- It has the ability to recover from two simultaneous disk failures.
- RAID level 6 applies the P + Q redundancy scheme using Reed-Soloman codes to protect against up to two disk failures.
- The following diagram shows RAID Level 6 with an array of six disks used for storing data and error correction codes.



- Choice of RAID Level

- **Level 0**: Improves the overall system performance at the lowest cost.
- **Level 1**: Is best for fault-tolerant, DB applications that require high percentage of uptime. It can be costly, but is best for variety of workloads.
- **Level 3**: Is appropriate for workloads consisting mainly for large transfer requests of several contiguous blocks.
- **Level 5**: Is a good general purpose solution. It provides high performance for large as well as small requests.
- **Level 6**: Appropriate if high level of reliability is required. It is the most expensive choice.