RAID and File System.

Ortiz Vega Angelo
Course Code: CE2103
Name: Algorithms and Data Structures II,
Academic Area of Computer Engineering.
Cartago, Costa Rica.

-Keywords: RAID, FileSystem, Ext2, Ext3, Fat, Fat32, NTFS

-Content:

RAID:

RAID (redundant array of independent disks; originally redundant array of inexpensive disks) is a way of storing the same data in different places on multiple hard disks to protect data in the case of a drive failure. However, not all RAID levels provide redundancy.

RAID use mirroring. Mirroring is the simplest way to give redundant storage. This is a technique that can only be used on two disks unless combined with striping. When data is written to one disk, it is simultaneously written to the other disk, so in a mirrored array the two drives are always an exact copy of each other. If one of the drives fails, service can continue uninterrupted and without data loss as the other drive simply takes over. This method of redundancy doesn't require any fancy calculations so is usually a part of onboard RAID solutions as it's quite cheap to implement. The down side of mirroring is the inefficient use of space. In an array that uses mirroring, half of the total capacity of the disks goes to redundancy.

RAID CONTROLLER:

A RAID controller can be used as a level of abstraction between the OS and the physical disks, presenting groups of disks as logical units. Using a RAID controller can improve performance and help protect data in case of a crash.

A RAID controller can be used in both hardware- and software-based RAID arrays. In a hardware-based RAID product, a physical controller manages the array. When in the form of a Peripheral Component Interconnect or PCI Express card, the controller can be designed to support drive formats such as SATA and SCSI. A physical RAID controller can also be part of the motherboard.

With software-based RAID, the controller uses the resources of the hardware system. While it performs the same functions as a hardware-based RAID controller, software-based RAID controllers may not enable as much of a performance boost.

If a software-based RAID implementation isn't compatible with a system's boot-up process, and

hardware-based RAID controllers are too costly, firmware- or driver-based, RAID is another implementation option.

PARITY:

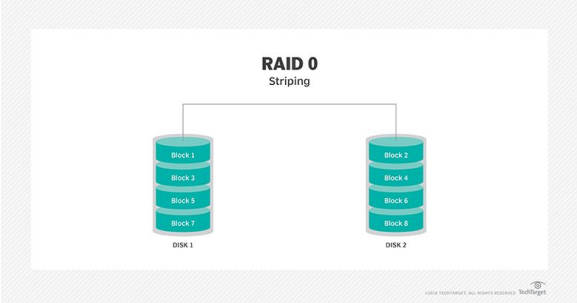
Parity is a technique that checks whether data has been lost or written over when it is moved from one place in storage to another or when it is transmitted between computers.

How Parity works?

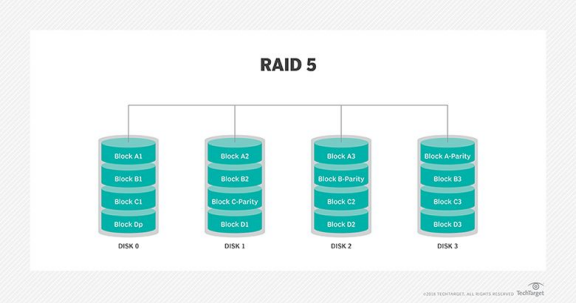
Because data transmission is not entirely error-free process, data is not always received in the same way as it was transmitted. A parity bit adds checksums into data that enable the target device to determine whether the data was received correctly.

An additional binary digit, the *parity bit,* is added to a group of bits that are moved together. This bit, sometimes referred to as a *check bit*, is used only to identify whether the moved bits arrived successfully.

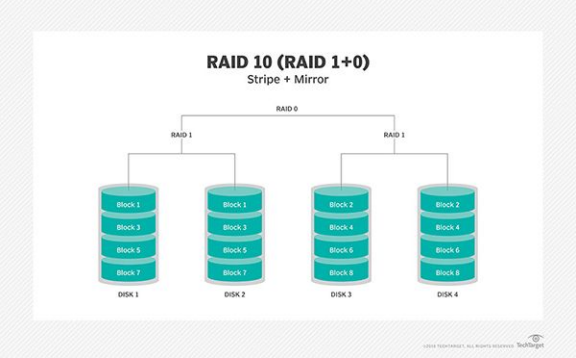
RAID 0



RAID 5



RAID 10



DATA RECOVERY:

RAID works by placing data on multiple disks and allowing input/output (I/O) operations to overlap in a balanced way, improving performance. Because the use of multiple disks increases the mean time between failures (MTBF), storing data redundantly also increases fault tolerance.

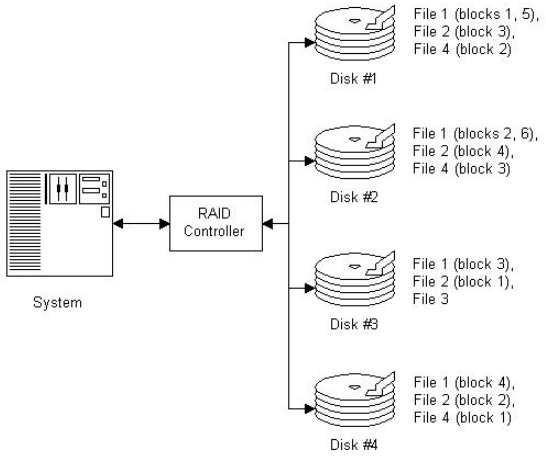
RAID arrays appear to the operating system (OS) as a single logical hard disk. RAID employs the techniques of disk mirroring or disk striping. Mirroring copies identical data onto more than one drive. Striping partitions each drive's storage space into units ranging from a sector (512 bytes) up to several megabytes. The stripes of all the disks

are interleaved and addressed in order.

In a single-user system where large records, such as medical or other scientific images, are stored, the stripes are typically set up to be small (perhaps 512 bytes) so that a single record spans all disks and can be accessed quickly by reading all the disks at the same time.

In a multiuser system, better performance requires that you establish a stripe wide enough to hold the typical or maximum size record. This allows overlapped disk I/O across drives.

Disk mirroring and disk striping can be combined on a RAID array. Mirroring and striping are used together in RAID 01 and RAID 10.



File System:

EXT2:

How does it work?

Also called second extended file system. It works for Linux Kernel, in the following way: It contains a table with i-nodes, which contain, the size of the block, the permissions, and the location to a sector of the disk that contains by blocks the data of the file, these blocks have a specific size.

How are the files / directories organized?

This has a table where the i-nodes are stored, which contain the information of the file. The location of this is a reference to the sector of the disk where the data is stored per block, thus fragmenting the file.

Maximum file size

The size of each individual file varies from 16 GB to 2 TB.

File system size: from 4TiB to 32TiB

On the contrary: The size of the file system varies from 2TB to 32TB.

Pros:

Its simplicity, it stands out, since it is easy to understand and "manipulate", it is also emphasized that the changes in this type of files are made at a considerably higher speed, based on its simplicity and fluidity.

Cons:

It is evident that it is not a file system registered by newspapers. It is also evident that his successor has many more possibilities and improvements, so that, although it is simple, fluid and

without errors, his successor came as a slightly more complex improvement but with many improvements and more possibilities.

FAT:

How does it work?

The FAT system stores the specific position of the beginning of each file (and this means sector, cylinder and disk, if there are several) on the hard disk. From there, the file uses as many "basic" blocks (clusters, in English) as you need, to write the entire file to disk.

Pros

Easy to attach to a file, Easy to delete a file.
FAT is very fast if the FAT table manages to be loaded in memory, since obtaining the necessary data from it will be very fast.

Cons

Access to small files is slow, Random access is very slow,Fragmented.

NTFS

How does it work?

The "New Technology File System" stores everything that has to do with files, it does so in the form of metadata. These are stored as Unicode (UTF-16).

How are the files / directories organized?

As for the structure of the files, it has a structure of Tree-B, this accelerates all types of search and access to files, reducing fragmentation, which was the most criticized in the FAT.

Maximum file size 16TiB

File system size 5MB

Pros

This allows to define the size of the cluster, from 512 bytes, in this way it is possible to manipulate at will the memory spaces in which a file will be stored.

Cons

Its main drawback is that it needs a considerable amount of memory on its hard disk for itself, so that, unlike the others, it defines a minimum amount of memory that it can have.