



RAID Whitepaper

Celeros
Bringing SANity to Storage Costs

Redundant Arrays of Inexpensive (Independent) Disks or RAID was first coined by researchers at the University of California at Berkeley. The fundamental principle behind RAID is that it allows a collection of individual disks to behave as one larger, faster, and more reliable disk. Thus capacity, performance, security, and reliability of the disk subsystems exceeds that of its constituents. Once almost exclusively the province of expensive SCSI disks, RAID has gained in popularity with the steady increase in the affordability and performance of ATA drives. How disks are accessed, written to and read from, results in many different implementations of RAID, referred to as RAID Levels that each have advantages and also their associated costs. While this brief is not meant to be an exhaustive dissertation on RAID, we will cover the levels supported by Celeros and explain the cost benefits of these different implementations. Celeros uses a dedicated controller with separate processor with up to 256MB of cache to offload the main CPU and memory. Remember that differentiation between the levels comes in the trade offs they make in the three dimensions of reliability/fault tolerance, performance/capacity, and cost. Please note also that no system is totally and utterly fool proof. Backups remain critical even when RAID is used.

RAID Definitions and Techniques

First, let us define **Logical Arrays** as a split or combination of **Physical Arrays**, which in turn are one or more **Physical Drives** that are simply the individual hard disks that comprise these arrays. **Logical Drives** are then made of one or more Logical Arrays.

Mirroring refers to complete redundancy of data on identical disks. The data that is being written on one Logical Array is completely duplicated on a similar array thereby providing 100% data redundancy. The cost associated with mirroring is that the amount of available storage is reduced by 50%, writes are slightly slower albeit reads are faster in some situations.

Striping refers to a technique that allows Physical Drives in a Logical Array to be used in parallel in order to gain in performance. In this technique, data is broken down in Byte or Block levels or **stripes**, where every Byte or Block is written to a separate disk in the array. Byte level can at times be a 512-byte sector, while Block size can be selected from variety of choices. The gain in performance is similar between Reads and Writes.

In some RAID levels, striping is combined with a technique called **Parity** to enhance fault tolerance. Parity, similar to parity in memory, is simply adding a Block (Byte) of calculated parity data to several Blocks (Bytes) in such a way that any one of the Blocks (Bytes) can be reconstructed in case of loss, from the remainder of the Blocks (Bytes) and the parity Block (Byte). While Parity gains from performance of striping, its disadvantages are more complexity and loss of some disk space (taken up by parity information.)

RAID Levels

There are many ways to combine RAID techniques. Some standardized combinations have been defined and are referred to as RAID Levels, even though 'Level' in this context does not denote any hierarchy or advantage. Levels are independent and different. Some RAID levels combine multiple other levels to achieve certain aims. RAID Advisory Board (RAB) has been active since 1992 in education and standardization of RAID technology. See <http://www.raid-advisory.com/>.

Techniques discussed above are used in different levels. Mirroring is used in levels 1, 0+1, 10 (aka 1+0). Striping without parity is used in level 0, 0+1, and 10. Striping with Byte level parity is used in level 3 and with Block level parity is used in level 5.

While the minimum number of drives required at each level are noted, there is no inherent maximum to number of drives in arrays other than the one imposed by controllers.

Level 0: Simple **striping** is used in this level to gain in performance. This level does not offer any redundancy. Data is broken into stripes of user-defined size and written to a different drive in the array. Minimum of two disks are required. It uses 100% of the storage capacity since no redundant information is written.

Recommended use for this level is when your data changes infrequently and is backed up regularly and you require high-speed access. Web servers, graphics design, audio and video editing, and online gaming are some example applications that might benefit from this level.

Level 1: This level uses **mirroring** and data is duplicated on two drives. If either fails, the other continues to function until the failed drive is replaced. At the cost of 50% of available capacity, this level provides very high availability. Rebuild of failed drives is relatively fast. Read performance is good and write performance is fair compared to single drive read and write. A minimum of 2 drives is required. Whenever the need for high availability and vital data are involved, this level is a good candidate for use.

Level 5: One of the most popular RAID techniques, it uses Block Striping of data along with parity and writes them to all drives. In contrast to the RAID levels that write the parity information to a single drive and use the rest of the drives for data blocks, RAID 5 distributes the parity blocks amongst all drives, keeping parity separate from the data blocks generating it. It requires a minimum of 3 disks. The impact on capacity is equivalent to removing one drive from the array. If any one drive fails, the array is said to be *degraded*, and the data blocks residing on that drive can be derived from parity and data on remainder of the drives. RAID controllers usually allow a spare drive to be configured that is used when the array is degraded and the array can be rebuilt in the background while normal operation continues. RAID 5 combines good performance, good fault tolerance, with high efficiency. It is best suited for transaction processing and is often used for "general purpose" service, as well as for relational database applications, enterprise resource planning and other business systems.

Level 10: RAID level 10 is an example of combining two RAID levels to achieve more targeted results. RAID level 10 is often referred to as '**Stripe of Mirrors.**' It is often confused with its brethren level 0+1 that is referred to as "Mirrored Stripes." While in each case drives are mirrored and blocks are striped to these drives, in level 10, Blocks are striped to N/2 sets of mirrored drives (N being number of drives in the array) while in level 0+1, blocks are striped to 2 mirrored sets each containing N/2 drives. Due to mirroring, the storage efficiency is at 50%. This level offers excellent fault tolerance and availability. It is recommended for applications requiring high performance and high reliability that are willing to sacrifice the efficiency (twice the number of drives to achieve the capacity). These include enterprise servers and moderate size database systems.

Level 50: Also referred to as level 5+0. It combines **Block Striping with distributed parity** with straight Block Striping of level 0. Another words it uses Block Stripe of level 0 on Level 5 elements. Minimum number of drives is 6. Resulting capacity can be derived from subtracting one drive for each set of Level 5 elements. As an example, in a 6-drive array that can only be configured as two sets of 3, resulting capacity is equivalent to 4 drives. Level 50 is recommended when high fault tolerance, large capacity, and random read/writes are required. It is sometimes used for large databases.

	<i>Level 0</i>	<i>Level 1</i>	<i>Level 5</i>	<i>Level 10</i>	<i>Level 50</i>
Efficiency	100%	50%	Good (~75%)	50%	Good (~75%)
Fault Tolerance	None	Very Good	Good	Excellent	Good-Very Good
Availability	Low	Very Good	Good-Very Good	Excellent	Very Good/Exc.
Random Read	Very Good	Good	Very Good/Exc.	Very Good/Exc.	Very Good/Exc.
Random Write	Very Good	Good	Fair	Good-Very Good	Good
Sequential Read	Very Good/Exc.	Fair	Good-Very Good	Very Good/Exc.	Very Good
Sequential Write	Very Good	Good	Fair to Good	Good-Very Good	Good

Table 1: Comparing attributes of different RAID levels.

References:

[RAID: High-Performance, Reliable Secondary Storage](#)
P M Chen, E K Lee, G A Gibson, R H Katz and D A Patterson

[A Case for Redundant Arrays of Inexpensive Disks \(RAID\)](#)
D. Patterson, G. Gibson, and R. Katz.

www.arstechnica.com

www.pcguide.com

About Celeros

At Celeros our mission is to make *reliable, high performance* storage solutions that are *easy* to operate and *affordable*. While we do not have any religion with respect to the type of technology that can help our customers, we are ardent believers in choosing appropriate technologies that cost effectively solve today’s problems and scale to address tomorrow’s needs.

To learn more, please visit www.celeros.com or email us at info@celeros.com