Redundant Array of Independent Disks (RAID) Technology Overview

What is RAID?
The basic idea behind RAID is to combine multiple small, inexpensive disk drives into an array to accomplish performance or redundancy goals not attainable with one large and expensive drive. This array of drives will appear to the computer as a single logical storage unit or drive.

RAID is a method in which information is spread across several disks, using techniques such as *disk striping* (RAID Level 0), *disk mirroring* (RAID level 1), and *disk striping with parity* (RAID Level 5) to achieve redundancy, lower latency and/or increase bandwidth for reading or writing to disks, and maximize the ability to recover from hard disk crashes.

The underlying concept of RAID is that data may be distributed across each drive in the array in a consistent manner. To do this, the data must first be broken into consistently-sized "chunks" (often 32K or 64K in size, although different sizes can be used). Each chunk is then written to a hard drive in RAID according to the RAID level used. When the data is to be read, the process is reversed, giving the illusion that multiple drives are actually one large drive.

Who Should Use RAID?
Anyone who needs to keep large quantities of data on hand (such as an average system administrator) would benefit by using RAID technology. Primary reasons to use RAID include:

- Enhanced speed
- Increased storage capacity using a single virtual disk
- Lessening the impact of a disk failure

Hardware RAID versus Software RAID

There are two possible RAID approaches: Hardware RAID and Software RAID.

Hardware RAID
The hardware-based system manages the RAID subsystem independently from the host and presents to the host only a single disk per RAID array.

An example of a Hardware RAID device would be one that connects to a SATAx or SCSI controller and presents the RAID arrays as a single SCSI drive. An external RAID system moves all RAID handling "intelligence" into a controller located in the external disk subsystem. The whole subsystem is connected to the host via a normal SATAx or SCSI controller and appears to the host as a single disk.
RAID controllers also come in the form of cards that act like a SATAx or SCSI controller to the operating system but handle all of the actual drive communications themselves. In these cases, you plug the drives into the RAID controller just like you would a SCSI controller, but then you add them to the RAID controller's configuration, and the operating system never knows the difference.

Software RAID
Software RAID implements the various RAID levels in the kernel disk (block device) code. It offers the cheapest possible solution, as expensive disk controller cards or hot-swap chassis are not required. Software RAID also works with cheaper IDE disks as well as SCSI disks. With today's fast CPUs, Software RAID performance can excel against Hardware RAID.

Hardware vs. Software RAID

- Software RAID
  - Software RAID: run on the server's CPU
  - Directly dependent on server CPU performance and load
  - Occupies host system memory and CPU operation, degrading server performance
- Hardware RAID
  - Hardware RAID: run on the RAID controller's CPU
  - Does not occupy any host system memory. Is not operating system dependent
  - Host CPU can execute applications while the array adapter's processor simultaneously executes array functions: true hardware multi-tasking

Notes
A hot-swap chassis allows you to remove a hard drive without having to power-down your system.

RAID Levels and Linear Support
RAID supports various configurations, including levels 0, 1, 4, 5, and linear. These RAID types are defined as follows:

- ***Level 0 —*** RAID level 0, often called "striping," is a performance-oriented striped data mapping technique. This means the data being written to the array is broken down into strips and written across the member disks of the array, allowing high I/O performance at low inherent cost but provides no redundancy. The storage capacity of a level 0 array is equal to the total capacity of the member disks in a Hardware RAID or the total capacity of member partitions in a Software RAID.

- ***Level 1 —*** RAID level 1, or "mirroring," has been used longer than any other form of RAID. Level 1 provides redundancy by writing identical data to each member disk of the array, leaving a "mirrored" copy on each disk. Mirroring remains popular due to its simplicity and high level of data availability. Level 1 operates with two or more disks that may use parallel access for high data-transfer rates when reading but more commonly operate independently to provide high I/O transaction rates. Level 1 provides very good data reliability and improves performance for read-intensive applications but at a relatively high cost. The storage capacity of the level 1 array is equal to the capacity of one of the mirrored hard disks in a Hardware RAID or one of the mirrored partitions in a Software RAID.

- ***Level 4 —*** Level 4 uses parity concentrated on a single disk drive to protect data. It is better suited to transaction I/O rather than large file transfers. Because the dedicated parity disk represents an inherent bottleneck, level 4 is seldom used without accompanying technologies such as write-back caching. The storage capacity of Hardware RAID level 4 is equal to the capacity of member disks, minus the capacity of one member disk. The storage capacity of Software RAID level 4 is equal to the capacity of the member partitions, minus the size of one of the partitions if they are of equal size.

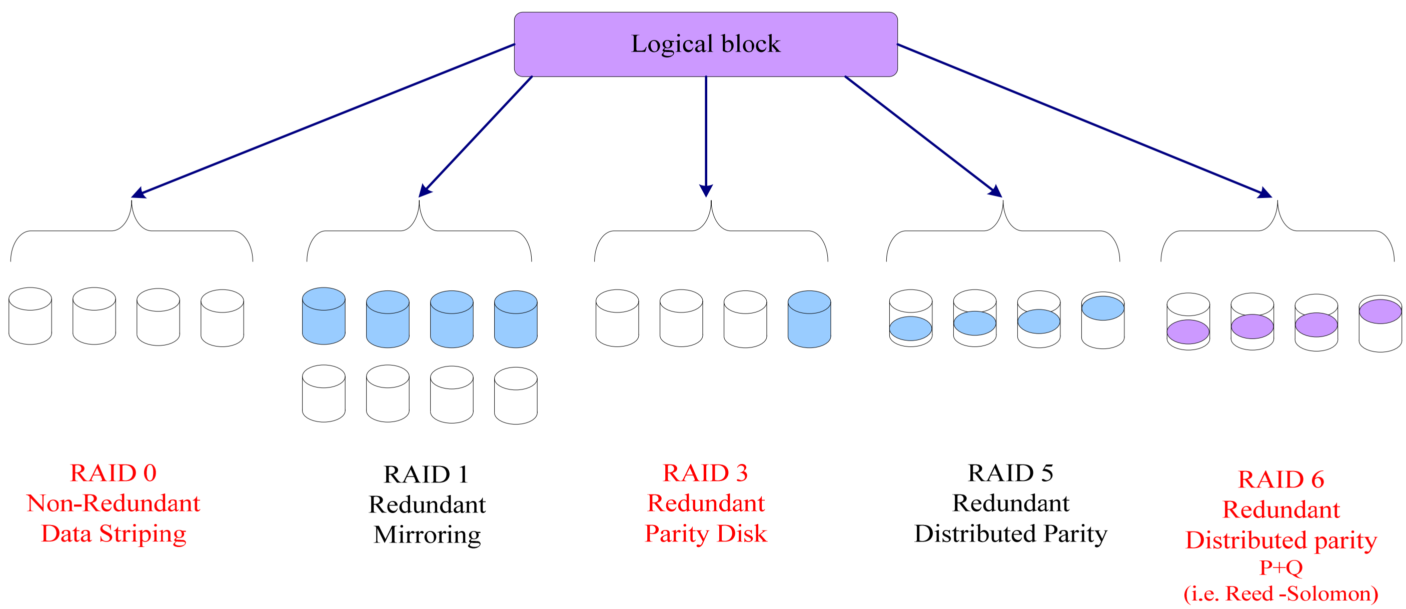
- ***Level 5*** — This is the most common type of RAID. By distributing parity across some or all of an array's member disk drives, RAID level 5 eliminates the write bottleneck inherent in level 4. The only performance bottleneck is the parity calculation process. With modern CPUs and Software RAID, that usually isn't a very big problem. As with level 4, the result is asymmetrical performance, with reads substantially outperforming writes. Level 5 is often used with write-back caching to reduce the asymmetry. The storage capacity of Hardware RAID level 5 is equal to the capacity of member disks, minus the capacity of one member disk. The storage capacity of Software RAID level 5 is equal to the capacity of the member partitions, minus the size of one of the partitions if they are of equal size.
- ***Linear RAID* —** Linear RAID is a simple grouping of drives to create a larger virtual drive. In linear RAID, the chunks are allocated sequentially from one member drive, going to the next drive only when the first is completely filled. This grouping provides no performance benefit, as it is unlikely that any I/O operations will be split between member drives. Linear RAID also offers no redundancy and, in fact, decreases reliability — if any one member drive fails, the entire array cannot be used. The capacity is the total of all member disks.

RAID Organization

- RAID : Redundant Array of Independent (Inexpensive) Disks
- Concept formalized by a research team at the University of Berkeley

Principles:

- Distributing the data across several disks (data striping)
- Redundancy (except for RAID 0) based upon XOR operation



RAID Architectures

- MAID (Massive Array of Inactive Disks): collection of SATA or PATA disks only active when accessed. Low acquisition cost and reduced electrical consumption. Can be used wherever access time is not a key factor (e.g. cache to a tape library)

Use cases for RAID:

- **RAID 0:** Performance without redundancy
- **RAID 1:** Performance and expensive redundancy (2 x disks)
- **RAID 3:** Cost effective redundancy (1 parity disk for N data disks) and high performance for large data transfers
- **RAID 5:** Cost effective redundancy (1 parity disk for N data disks) and high performance data transfers. Compared with RAID 3, the distributed parity removes contention on parity disk and performance on data updates is improved
- **RAID 6:** Same basic characteristics as RAID 5, but with the ability to survive the concurrent failure of two disks rather than just one

Storage Virtualization

Naive implementation of RAID leads to a very large (but reliable) disks

- Storage virtualization brings the vision of small virtual disks
- Advantages:
  - Offers several virtual disks of capacity and RAID levels according to their usage
  - Hide physical differences in disk sizes and technologies
  - Optimize the use of installed configurations

Comparing the various RAID levels

RAID Type	Name	Storage cost	Relative data availability	Large sequential read speed	Large sequential write speed	Random read bandwidth
0	Data striping	> 1	Lower than that of a conventional organization	Higher - Depends on the number of parallel disks	Higher - Depends on the number of parallel disks	Higher
1	Mirroring (of order M; M=2	x M (see note[1])	> RAID 3 & RAID 5 < RAID 6	Up to M times a single disc	Lower than a single disc	Up to M times a single disc
0 + 1	Striped mirror (of order M; M=2 usually)	M x N	> RAID 3 & RAID 5 < RAID 6	Up to M times a RAID 0 equivalent	Can be higher than that of the single disc as a function of N	Up to M times a RAID 0 equivalent
3	Parity disk	N + 1	>> conventional disc	Higher Depends on the number of parallel disks and the need to compute parity (< RAID 0)	Higher Depends on the number of parallel disks and the need to compute parity (< RAID 0)	Higher
5	“spiral” parity	N + 1	>> conventional disc ~ RAID 3	< RAID 0 because of the parity check	< RAID 0	Higher > RAID 3
6	Double “spiral” parity	N + 2	Higher than all the other types	Slightly > RAID 5	< RAID 5 (2 parity ‘blocks’)	Slightly > RAID 5

Note: If redistributing the data across a number of disks is unnecessary, use of the simple mirror results in a doubling of the number of disks required for data storage. It is also possible to have more than 2 mirrors (the number M is used in the table to specify the number of mirrors)

Comparing System Architecture Options for RAID

	RAID supported by the server	RAID supported by controller in server	RAID supported within a specialized subsystem
Advantages	• Low cost	• Moderate cost	• Connectivity usually high (constrained by the subsystem) along with the possibility of connecting multiple subsystems
	• High connectivity (i.e., the server's innate connectivity)	• Good execution times and good bandwidth (specialized hardware)	• High bandwidth (specialized hardware)
	• Scalability (increasing server performance increases RAID performance)		• Good write performance, if a write cache is available
	• High availability (no extra hardware elements involved)		• High availability (doubling internal controllers and multiple access paths)
			• Independence between host interconnect (e.g. FC-AL) and disk (e.g. SATAx and SCSI).
Disadvantages	• Server performance is impacted by the extra load of implementing RAID functionality	• Number of disks supported constrained by the connectivity capabilities of the controller	• Specialist hardware (redundant secure cache)
	• Data availability demands mean that the disks must have dual access interfaces to allow recovery after failure of the server	• Data availability demands mean that the disks must have dual access interfaces to allow recovery after failure of the server or the controller	• Higher cost
			• Better response time than a pure server-based solution thanks to the server / subsystem interconnect

Architectural options for storage virtualization

- System Architecture Options for Storage Virtualization
- Comparison of System Architecture Options for Storage Virtualization

	Client systems	Storage Subsystem	SAN Subsystem	Specialist hardware
Advantages	• Narrow integration with File Systems and databases • Virtualization based on proven principles	• Allows the support of heterogeneous storage (technology and vendor independence)	• Ability to connect diverse clients	• Centralized control • High performance due to separation of control and data transfer • Supports heterogeneous clients
Disadvantages	• Administrative complexity • Global visibility of storage means that cauterization techniques must be used	• Solution proprietary to each vendor • Global visibility of storage means that cauterization techniques must be used • Qualification of the solution • Cost of the various subsystems • Multiple points of administration	• Need to use cauterization techniques for availability • Limits choice to hardware that can support virtualization • Interoperability concerns between different vendors • Global visibility of storage means that cauterize techniques must be used	• Difficulty of qualifying the solution in a heterogeneous environment • High availability requires redundant equipment • Complexity of connection • Requires special drivers in the clients