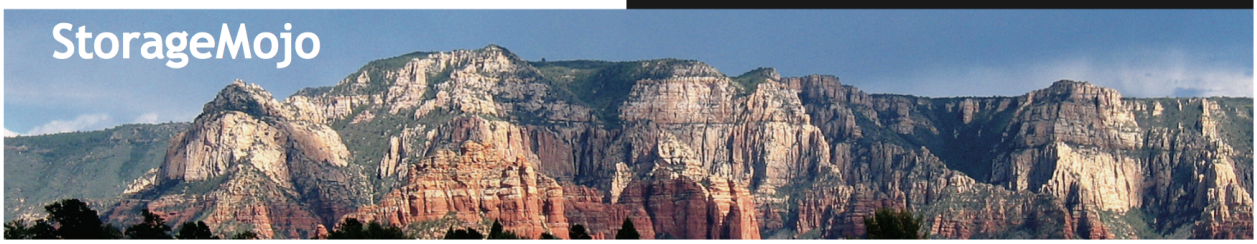
a StorageMojo whitepaper



Availability vs durability in archive solutions

Executive Summary:

Availability and durability seem similar, but play very different roles in data preservation. Availability refers to *system uptime*, while durability refers to *long-term data protection.* Durability is the key requirement for archives. RAID arrays are expensively focused on availability. Here's what you need to know.

Redundancy

Storage system **availability** is achieved through **hardware redundancy**, while **durability** is achieved by **data redundancy**. Let's look at each in turn.

RAID arrays typically support two or more of each major hardware component: controllers; I/O paths; drives; and data redundancy that will survive one (RAID 5) or two (RAID 6) drive failures. While RAID arrays are designed to achieve six nines (99.9999) availability, they often fail to do so — and that puts your data at risk.

Research has found why RAID arrays fall short of design goals.

- Drive failures are highly correlated. When one drive fails, another drive failure is much more likely. RAID 6 protects against two drive failures. But there's another problem.
- An Unrecoverable Read Error (URE) during recovery from one or two drive failures can cause recovery failure, and the larger the hard drive, the more likely there will be an URE. (See table below.)
- Today's high capacity hard drives create another problem: RAID rebuild times can now take hours, even days, during which data is at risk from another drive failure or URE, and array performance is compromised.

Another consideration is that RAID arrays rely on 24/7 field service to quickly repair hardware failures. An archive system should be able to handle multiple component failures without requiring immediate service, saving considerable expense. Failure in-place of components is perfectly acceptable in an archive environment provided that data access remains. Repairs can take place at the user's convenience.

Durable archive systems, i.e., high availability object-based storage (or object stores) also offer hardware redundancy, but the big difference is how the data is encoded and laid out on the storage. Object storage differs from block/file approaches in that each object is a single immutable entity; an object can contain most any type of data. Hardware redundancy is almost a given: a petabyte archive will typically use well over 100 disks spread across at least three servers.

Commissioned by Western Digital.

Risk of recovery failure with RAID 5 and 6 as drive capacities grow.

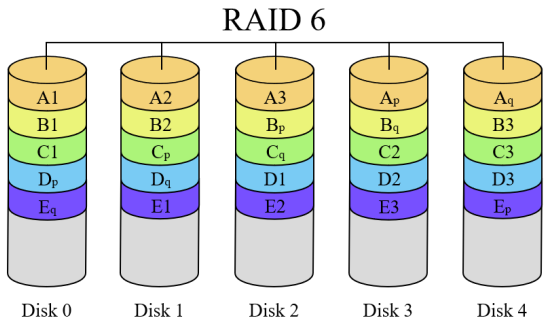
Disk Size GB	# of HDDs	Array Size GB	Bits/Array	Error Rate	RAID 5	RAID 6
500	6	2500	2.00E+13	1.00E-14	18.11%	1.74%
1000	6	5000	4.00E+13	1.00E-14	32.95%	6.13%
2000	6	10000	8.00E+13	1.00E-14	55.04%	19.07%
4000	6	20000	1.60E+14	1.00E-14	79.78%	47.44%
6000	6	30000	2.40E+14	1.00E-14	90.92%	69.10%

Comparing data layouts of RAID arrays and durable archives

All RAID systems use a form of erasure coding, where a logical operation is performed on the data to create parity data. The parity data enables data on failed drives to be recovered and written to a new or spare drive. But RAID arrays and scale out object storage use parity very differently.

RAID systems lay out data and parity across a fixed bank of disks, typically in a single shelf. In a RAID 6 layout there might be 4 data disks and 2 parity disks, which means that a third of the drive capacity is dedicated to redundancy, while protecting against only two drive failures. RAID is about protecting disk drives. Object stores that employ advanced erasure codes are about protecting data, not disk drives.

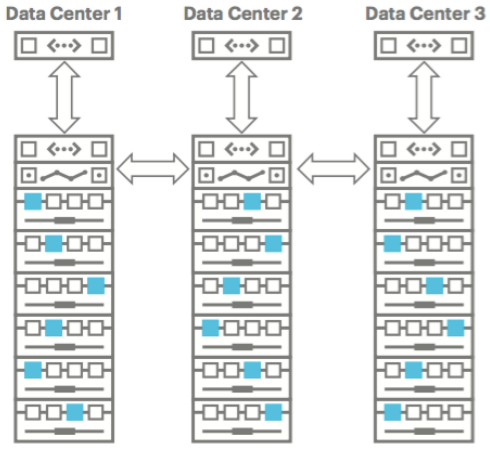
Durable archive systems also use parity to protect data, but use algorithms that did not exist when RAID was proposed in the late 1980s. These modern, advanced erasure codes — often called *rateless* or *fountain* codes — offer higher redundancy and greater efficiency, perfect for economical, long-term, data protection. RAID rebuilds a physical device: the HDD, object storage rebuilds a logical construct: an object. Data and parity are not separate in objects like they are in RAID.



Data and parity are built into objects, unlike RAID volumes. Objects are broken into shards of data and distributed across the available storage. In an 18/8 three-geo spread configuration, the object is broken into 18 shards, and up to 8 shards can become unavailable without harm to the object. This allows an entire data center to become unavailable while keeping data accessible. For a single rack configuration the encoding might be 18/5, where the object is broken into 18 shards and can lose any 5 shards and still maintain the object.

No RAID array offers the same level of data protection and capacity efficiency. In addition, a single namespace object store can be spread across a single server, rack, datacenter, or as the diagram to the right shows — across multiple geographies.

Put into numbers: RAID arrays aim for 99.999% availability. That means, on average, less than six minutes of unscheduled downtime each year, but typically there is no specification for data loss. Scale out object storage archives aim for 99.9999999999999% durability. That means that if you stored 1,000 trillion objects, only one would not be readable. How many organizations have even a trillion files?



A high-availability, multi-datacenter, archive solution.

The practical difference

For data centers uptime is the most important metric. That's the economic justification for costly, uptime-oriented storage arrays.

Archives, on the other hand, are measured on data durability, because that is an archive's key mission. A non-trivial archive will have plenty of hardware redundancy. Data durability, on the other hand, is critical when an enterprise has to respond an e-discovery request and under threat of severe sanctions if documents can't be produced.

Scale out object storage archives have other advantages. It's difficult, if not impossible, to economically back up large archives for data migration, it is vital that the archive support rolling technology updates, something RAID systems rarely offer.

For example, one of the advantages of object stores is that, unlike RAID arrays, they do not necessarily require that all drives offer the same capacity. That makes it possible to replace an existing — perhaps failing — drive with a newer and larger capacity drive, which also enables rolling upgrades as components wear out. Your archive vendor can tell you more.

Conclusion

The days when tape silos were the automatic answer for archiving are long past. Modern archive systems based upon object storage, using technologies pioneered by hyper-scale data centers, enable enterprises to archive their data at costs below public clouds, and with greater security than public clouds can offer.

I encourage data center planners and architects to review the research literature appended here to fully appreciate the limitations of RAID arrays for archive use. The rate of change in storage technologies is accelerating, and IT professionals are wise to adapt to remain competitive against out-sourced options.

About The Author

Robin Harris is the president and chief analyst of TechnoQWAN LLC, publisher of StorageMojo.com. He has over 30 years experience in the IT industry in product management and marketing, business development and strategic planning at companies large and small. He earned degrees from Yale University and the Wharton School of the University of Pennsylvania.