Adam Andre

Instructor: Nicole Martin

CINE-GT 1807: Digital Preservation

15 December 2017

RAID: History, Application, and Future in Digital Preservation

Since the advent of the Information Age, archivists have been tasked with preserving digital files. Because digital objects are so complex in nature and exist in an intangible world, it can be difficult to track, describe, and store them. Computer file systems were developed to mimic the file systems of the analog world but this method may not be the best solution for long-term preservation in a digital environment. In a rapidly evolving digital landscape, digital files can so easily be erased, altered, lost, forgotten, or corrupted. For now, the hierarchical directory storage method is a standard in computer operating systems. So how do archivists remain responsible custodians to an ever-increasing number of digital objects?

One popular way that many archivists protect their digital repositories is through a storage technology called Redundant Array of Independent Disks or RAID. As the name implies, RAID combines multiple disks so that the computer operating system recognizes the RAID array as one unit.[1] Different RAID "levels" allow for a variety of setups that utilize two unique storage techniques, striping and parity, which improves performance, reliability and access. Digital preservationists leverage these RAID characteristics to monitor and secure their digital collections. But RAID was not developed with the digital preservation community in mind. The technology grew organically from the computing industry.

[1] "A Case for Redundant Arrays of Inexpensive Disks (RAID)", www2.eecs.berkeley.edu, https://www2.eecs.berkeley.edu/Pubs/TechRpts/1987/CSD-87-391.pdf

Before RAID

There would be no RAID without the invention of hard disk drive (HDD) storage. In 1956, International Business Machines Corporation (IBM) developed the RAMAC 305, the first hard disk storage system that held 5 megabytes of data. The 305 was a refrigerator-sized machine designed to store accounting data in real time on 24-inch disks. From 1956 on, hard drives gradually reduced in size. In 1962, IBM introduced the 1311 disk drive that shrank the disk size down to 14 inches. In 1973 IBM released the Winchester, a hermetically sealed hard drive similar to the design of hard drives we know today. In 1979, Shugart (now Seagate) began producing hard disk drives for desktop computers.[2] The 1980s brought about the personal computing revolution with smaller computers that required smaller, faster drives to store data. Although they were much slower, enterprise companies still needed 14-inch disk platters to write and store large amounts of information.

The Berkeley RAIDers

In 1983, Randy H. Katz joined the University of California at Berkeley faculty. He joined with Michael Stonebreaker on the Ingres Database Project to study the upper limits of performance in database systems by tracking the number of transactions per second that could be processed. Through their experiments, Katz and Stonebreaker found that a major bottleneck was occurring at the point of writing the transaction commit log to stable storage, which at that time was one large disk with a slow input/output rate.[3]

---

[2] Frank Hayes, "The Story So Far," Computerworld, November 17, 2003, https://www.computerworld.com/article/2573180/data-center/the-story-so-far.html.

[3] Randy H. Katz, "RAID: A Personal Recollection of How Storage Became a System", October-December 2010, http://web.eecs.umich.edu/~michjc/eecs584/Papers/katz-2010.pdf

In 1986, Katz purchased an Apple MacPlus with a 10 megabyte Shoebox hard drive that opened his eyes to the possibilities of small form-factor drives. Katz, Stonebreaker, and a colleague at UC Berkeley named David Patterson began working on the XPRS Project, a project that set out to explore how an extensible database system would run on a shared-disk multiprocessor architecture.

When the XPRS Project ended in 1987, Patterson and Katz began discussing the possibilities and challenges of building large-scale multidisc storage systems using small form-factor drives. They knew that reliability would be a huge obstacle to overcome since multiple drives running at once would increase the chance of a single drive failing. Garth Gibson, then just a graduate student, joined Katz and Patterson and became the resident expert on performance-reliability tradeoffs. In the summer of 1987, Patterson and Gibson attended a disk storage technologies class at Santa Clara University. The class was taught by Al Hoagland, a titan in the disk storage community that worked on the California Digital Computer Project (CALDIC) at Berkeley and was a member of the IBM RAMAC team that built the first disk drive in 1956.

In the fall of 1987, Patterson, Katz and Gibson began developing the taxonomy for what they called RAID: Redundant Array of *Inexpensive Disks ("Inexpensive" would later be changed to "Independent" by marketing teams because the technology was not actually inexpensive at the time). They structured the RAID taxonomy in levels based on the type of storage technologies that were already being used in the industry. Below are the original RAID levels:[4]

---

[4] "A Case for Redundant Arrays of Inexpensive Disks (RAID)", www2.eecs.berkeley.edu, https://www2.eecs.berkeley.edu/Pubs/TechRpts/1987/CSD-87-391.pdf

- RAID 1 – Mirrored Disks
- RAID 2 – Stripes Data at Bit Level across all drives with Hamming Code for error correction
- RAID 3 – Stripes Data at Byte Level with 1 Parity disk
- RAID 4 – Stripes Data at Block Level with 1 Parity disk
- RAID 5 – Stripes Data at Block Level with Parity distributed across all drives

RAID 1 was influenced by Tandem Computers, a company focused on high availability systems. They mirrored their data so that it could be read from the drives twice as fast with the tradeoff of half of the capacity. RAID 2 had a proprietary use at Thinking Machines, an early developer of parallel high-performance computers with a bit oriented single construction. Storage companies Maxstor and Micropolis announced products that appeared to use one disk but were actually multiple disks with one for parity that would act as one logical unit. At the time, the Berkeley RAIDers thought that they invented RAID 4 and RAID 5 but those technologies were already being implemented at IBM unbeknownst to them.

Katz, Patterson, and Gibson compiled all of this data into a technical report titled "A Case for Redundant Arrays of Inexpensive Disks (RAID)" in November 1987. They sent the report to several reviewers in the storage industry including Al Hoagland. Hoagland sent copies of the report to his contacts in the industry. The report went viral and served as the catalyst that brought the idea of multidisk storage into the limelight. The RAIDers also submitted it to the 1988 Association for Computing Machinery's Special Interest Group on Management of Data Conference where it was officially published to the rest of the academic world. From 1990 to 1993, NASA and DARPA funded a 192-disk bay RAID 5 prototype at UC Berkeley configured to support network attachment. In 1999, the RAIDers received the Institute of Electronical and

Electronics Engineers (IEEE) Reynold B. Johnson Information Storage Systems Award. RAID is now a multi-billion dollar international industry sector.

Striping and Parity in RAID 5 and RAID 6

The two most advanced data storage techniques that a RAID can employ are striping and parity. Striping is writing sequential data either at the bit, byte, or block level across multiple drives.[5] A single strand of data on one drive is called a "strip." All of the strips combined across multiple drives are called a "stripe".[6] This method of dispersing data across drives allows multiple drives to share the workload of what would normally be on one drive without RAID. It also increases data read and write speeds because it can leverage the input/output rates of a cluster of drives rather than just one. The trade-off of only using striping though is that it increases the chance of losing data. With multiple drives working at once, the chance that one drive in the cluster fails is high. If one drive is lost, all the data is lost because there is no redundancy of information.

That is where parity comes in. In RAID, parity is the repetition of data to provide redundancy and error correction in the case of a failed drive(s).[7] But parity does not create an exact duplication of the data like mirroring would in RAID 1. Parity works through a logical computing function called Exclusive/Or (XOR) to produce a "parity bit". This function will

---

[5] "What Is Disk Striping? - Definition from WhatIs.Com," SearchStorage, http://searchstorage.techtarget.com/definition/disk-striping.

[6] "RAID-5 (Striping with Parity)," accessed December 15, 2017, https://sort.symantec.com/public/documents/sf/5.0/hpux/html/vxvm_admin/ag_ch_intro_vm17.html.

[7] "What Is Parity? - Definition from WhatIs.Com," SearchStorage, http://searchstorage.techtarget.com/definition/parity.

analyze each bit of data and assign an additional bit to that byte based on how many 0s and 1s are in the byte. The XOR table is convenient for visualizing this function:

INPUT		OUTPUT
A	B	A XOR B
0	0	0
0	1	1
1	0	1
1	1	0

[8]

In basic terms, the RAID technology will consult this table when bytes of data are being written to a disk. For example, in one type of RAID 5 configuration if byte A "11010010" is written to one disk and byte B "00110110" is written to a second disk, the RAID will XOR each digit in those two bytes based on their position in the byte. Take the first digit from A=1 and first digit of B=0. That will give you an output of P=1. The RAID will perform this for each second digit, and third digit, and so on until it has produced a byte of parity information.

- A – 11010010
- B – 00110110
- P – 11100100

Depending on how a RAID is configured, it will also store this parity bit on the tail-end of each seven digits to create a full byte. This method of "even" or "odd" parity will attach a parity bit to a string of seven bits based on the number of 1s. In even parity, if there are an odd number of 1s in a string of seven, then a parity bit of 1 is attached to create an even number of 1s. In odd

[8] "XOR Gate," *Wikipedia*, November 29, 2017, https://en.wikipedia.org/w/index.php?title=XOR_gate&oldid=812700840.

parity, if there are an odd amount of 1s, then a parity bit of 0 is attached to keep the odd amount of 1s. Again, a table is the simplest way to understand this concept:

How to Adjust Parity Bits for Even or Odd Parity

ORIGINAL DATA	EVEN PARITY	ODD PARITY
00000000	0	1
01011011	1	0
01010101	0	1
11111111	0	1
10000000	1	0
01001001	1	0

SOURCE: UNIVERSITY OF LONDON BIRKBECK

TechTarget [9]

Even and odd parity is not completely fault-resistant. Since it only attaches one parity bit to a string of seven bits, it is possible that multiple bits in a string could be corrupted and the parity information would be false and useless, resulting in a loss of data. However, in personal computer transmissions, it is extremely unlikely that multiple bits in a string of seven are corrupted. In larger enterprise systems however, 3 parity bits are attached to a string of seven to prevent data loss in the case of multiple-bit corruption.

In mid to small-level digital repositories, like in the archiving community, RAID 5 (striping and parity with the ability to lose one disk without data loss) and RAID 6 (striping and parity with the ability to lose two disks without data loss) are implemented because it is secure, reliable, highly accessible, and affordable. For a 6-bay disk RAID 6 array on a QNAP network attached storage (NAS) device, the Cinema Studies Department Archive at New York University

[9] What Is Parity? - Definition from WhatIs.Com," SearchStorage, http://searchstorage.techtarget.com/definition/parity.

was able to get 18 TB of total disk space with 11 TB of data storage and two drives for parity for less than $1,200. That should ensure the safety of the digital repository for 3-5 years. And with the prices of hard drive storage getting cheaper and bigger every year, that $1,200 amount will drop. IndieCollect uses a 12-bay Synology server configured as RAID 6 for their large digital film scans. This set-up, although slightly more expensive than NYU's, allows for very quick read speeds for editors and secure storage in the short term.

The unfortunate fact is that these hard disk drives eventually fail. The mechanisms of the hard drive work fast and precisely (nanometers) to read and write data onto a disc. These failures are relatively unpredictable but as a general rule of thumb, you should replace hard disk drives every 3-5 years. When one drive fails and the data has to be re-written to a new drive using the parity information, the likelihood of an additional driving failing during the rebuild increases. To mitigate the risk, most digital archivists use RAID 6 to protect themselves from two drives failing at once. Some digital preservationists use disks from different manufacturers to reduce the chance that multiple drives fail at once which is the nightmare of any RAID configuration. In the end, archivists do have to migrate their data to new drives so that the data is stored safely for another 3-5 years.

Future of RAID for Digital Archivists

Although multiple drive failures at once are rare, it does happen. Digital archivists want to reduce the risk as much as possible with the finances and resources that they have available. RAID has given lower-to-mid level institutions the opportunity to protect their digital objects at an affordable price, but there can always be more security to ensure safe storage in the long run. So what are some options for archivists in institutions with a limited budget? There are two different options, one that is more attainable in the near future and one that is already in action for large big data repositories.

Solid State Drives (SSD)

Solid state drives seem to be the future of small form-factor digital storage. As opposed to hard disk drives (HDD), there are no moving parts inside of a solid state drive. SSDs are composed of connected memory chips that read and write data, similar to a flash thumb drive. This type of architecture drastically outpaces hard disk drives in read and write speeds. In one study of hard disk drives versus solid state drives in RAID 6, the SSD beat the HDD by over 34,000 input/outputs per second.[10] According to a plethora of studies, results have shown that SSDs have an annual fail rate of 0.01% while HDDs have an annual failure rate of 4-6%.[11] SSDs seem to be the obvious choice for reliable, high-performance storage. This SSD RAID method is being used by some vendors, so why isn't it being used universally for digital storage? As of this writing, a 1 TB hard disk drive costs approximately $50 and a solid state drive of the same capacity costs $250. That translates to 5 cents per gigabyte for HDDs and 25 cents per gigabyte for SSDs. This is a difference that cannot be overcome by most collecting institutions at the moment so they go for the cheaper HDD option. But according to Moore's Law, chip performance will (and has) doubled every 18 months. It is only a matter of a few years before SSDs become as cheap as HDDs. Digital archivists will be able to make the easy transition to SSDs and slot them directly into their RAID 6 workflows.

---

[10] Karl Paulsen, "Comparing IOPS for SSDs and HDDs," TV Technology, http://www.tvtechnology.com/expertise/0003/comparing-iops-for-ssds-and-hdds/276487.
[11] Chuck Brown Intel SSD business development manager, "Debunking SSD Myths," Network World, January 22, 2015, https://www.networkworld.com/article/2873551/data-center/debunking-ssd-myths.html.

Object-Based Storage

Object storage, or Object-based storage (OBS), is a system by which each individual data object is stored on its own with any and all associated metadata instead of in data blocks.[12] This type of storage is especially useful for objects that only need to be read and not written to. If an object needs to be written to, the whole object has to be recalled, and when the writing is done, the object has to be completely rewritten to storage. OBS is used to solve rapid data growth issues which is why it is currently being used only by big data storage companies such as Amazon, Facebook, and Spotify. It is ideal for "dark" or "cold" storage situations where the data objects only need a place to be stored safe and not accessed. OBS is obviously not ideal for everyday storage, but is already being used by digital preservationists for an added layer of security and geographic separation. Maybe someday in the future, a system like this will be able to exist on local storage for archivists since the package structure resembles that of a SIP, AIP, or DIP in the Open Archival Information System (OAIS) Reference Model.

Conclusion

The RAID technology has existed for nearly 30 years for a reason. Since 1988, the technology has stayed relatively the same but it has grown to be used by nearly every type of collecting organization. As data objects grow in size and scope though, RAID's limitations become more apparent. RAID is currently constrained by the hard disk drive storage devices that it runs on. Digital archivists are always looking for the best ways to ensure the protection of their

[12] Yadin Porter de León, Tony Piscopo, and Druva Admin – 08.14.14, "Object Storage versus Block Storage: Understanding the Technology Differences," *Druva* (blog), August 14, 2014, https://www.druva.com/blog/object-storage-versus-block-storage-understanding-technology-differences/.

data. It appears that archivists will continue to use RAID as a digital storage technique but soon switch over to solid state drives. It will be worth watching what big data companies do with object-based storage and if that technology can be translated into an afforadable, smaller form-factor environment. For now, RAID will continue to endure and be the go-to storage strategy to ensure the safety and security of digital objects.