



SCENE SAVERS

Making History Modern

How Safe is my Data?

As an archivist, what keeps you awake at night? When you are dealing with physical media, such as videotape, audiotapes and motion picture films, you can pick them up. You can inspect them. You can note any problems, such as unusual smells, like vinegar syndrome, or uneven pack on the reels or other physical problems. But, with the requirement to move into the digital world to preserve and provide access to content, what about your data? Now, rather than using visual and tactile senses to determine if there is a problem, you have to rely on sometimes mysterious methods to make sure your digital assets aren't disappearing right before your eyes.

I was prompted to write this article by something that recently happened at our facility. We digitized a large videotape collection for a client. All of the data was being copied to a RAID protected network attached storage (NAS) system. I'm not going to name names, to protect the innocent... but let's just say the NAS was from a well-respected global supplier of hard drive technology. The system was a 16 TB NAS unit configured to RAID 5, which theoretically meant that the data should still be safe, even if a complete hard drive was lost. Well... the worst happened and a hard drive failed.

This failure occurred only about four months after putting the NAS into use... which is a bit strange. I'll explain why below. We received a replacement drive from the manufacturer and worked with their technical support to recover the files... but it didn't work. All of the data was lost. Fortunately, we had the foresight to have a second NAS created simultaneously with the first... and had all of the files backed up to LTO data tapes... but if we didn't, all the digitizing work would have been for nothing. Again, I ask... how safe is your data?

There is a recent article out from a cloud based backup and storage company called Backblaze. Backblaze analyzed hard drive failure rates over the last four years based on more than 25,000 spinning disks used in their storage facility. They've come up with some interesting information, including brands and model numbers that are more reliable and less reliable over a four year life span. Again, I'm not going to name names... to read the article [CLICK HERE](#).

The gist of the article is that hard drives last longer than expected, but their expected life is three to four years when disks are spinning. This is but a blip in time for an archive, who would typically measure expected life in terms of generations. According to Backblaze, 78% of the drives they purchase are lasting longer than expected, but 22% are failing within their first four years. Backblaze uses mainly consumer type hard drives... similar to what most archives would use... so this information provides a pretty good starting point for us to look at how safe data is on spinning disks.

I mentioned earlier that I felt it was unusual for a hard drive to fail after four months. The reason I believe this is strange is because of something called a bathtub curve. This curve anticipates a high

424 SCOTT ST. COVINGTON, KY 41011
859.291.5100

www.scenesavers.com





failure rate when a drive is new due to manufacturing defects, then predicts a low failure rate for a period of time, then a high failure rate as internal components wear out. The shape of the curve is similar to a bathtub (See Figure 1). Well... our drive didn't fail right away... and it didn't fail after several years of service. The drive was being used almost on a daily basis during a normal work day for about four months... then it failed. This isn't likely attributable to a manufacturing defect, as we would have expected failure sooner. We don't believe it was wear and tear... the drive wasn't old enough. It was a completely unexpected failure... simply due to being "unlucky".

Getting back to the Backblaze article, their results show that within the first 1.5 years, drives fail at about 1.5% per year. For the next 1.5 years, failure rates go down, to about 1.4% per year, but then after three years, failures increase significantly to 11.8% per year (See Figure 2). Granted archives aren't using their drives the same way Backblaze is. Within Backblaze, the drives are spinning almost constantly, 24x7. Most of their failures are "soft" failures, where the drive hasn't completely failed but is taking longer to read/write data. One challenge archives might face is "hard" failures, which seem to occur more often during powering up, or going through cycles of spinning down and then spinning up again. I know that several years ago, we were seeing a significantly higher failure rate with one particular brand of external drives because the internal control software was causing the drives to cycle up and down a lot. This was competing with the Windows operating system which was also attempting to control when the drives spun down. This constant cycling appeared to cause premature failure. While it is a small sample, we had 2 drives fail out of 10 within the first month of being put into service.

Little or no data exists on hard drive failure after drives were allowed to sit not being powered up for years, so it is practically impossible to determine if you are safer leaving drives turned off and sitting on a shelf. Some of you might be wondering about Mean Time Between Failure (MTBF) as reported by the manufacturers, and why I left it out of this article. According to a [Study by CMU](#) in 2006, MTBF data doesn't mean much in real world applications, and often overstates reliability of hard drives. Gee, doesn't this sound familiar? Especially when you think of other "solutions" in the archival world with

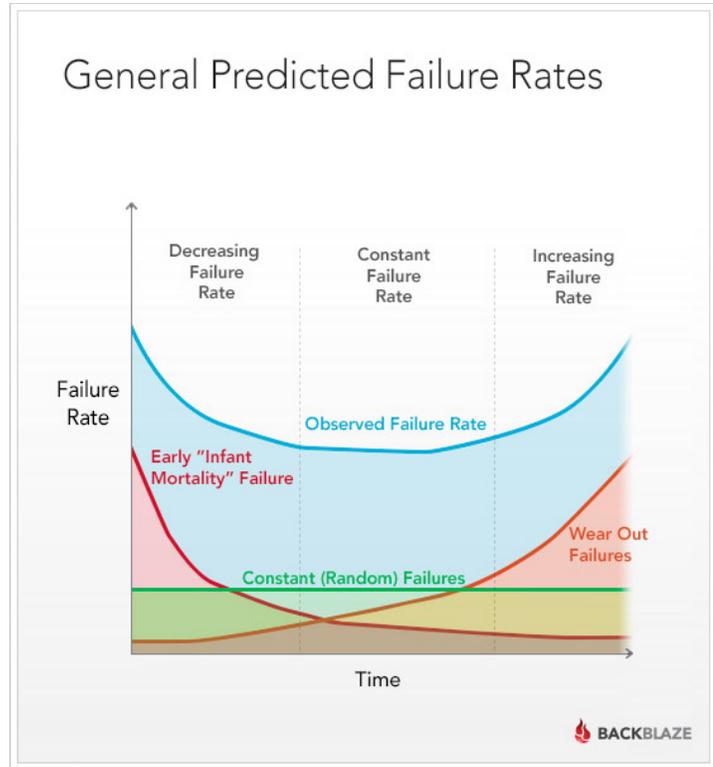


Figure 1. Bathtub curve depicting rate of drive failure over time.



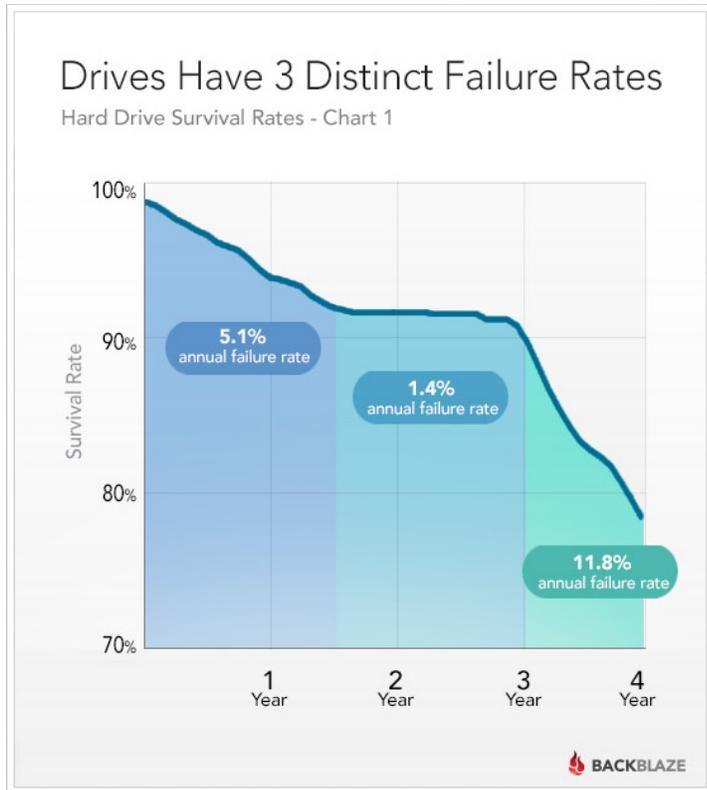


Figure 2. Hard drive survival rate over four year period.

“claimed” life expectancies (remember the 100 year DVD?).

Backblaze has an even more recent blog posting about specific brands of hard drive, which can be accessed [Here](#). This post discusses which manufacturers appear to have more reliable drives. I bring it up here because you might be thinking “I buy Brand-X, and they claim to have better hard drives”. The reality is that there are only three main drive manufacturer’s remaining, Seagate, Western Digital and Toshiba... and whatever brand you are buying, chances are it is a re-branded version of one of these manufacturers. So, you might say, well what about LaCie? Nope, purchased by Seagate in 2012. Ok, what about G-Tech? Again, no, purchased by Western Digital. Ok, what about Hitachi? Sorry, purchased by Western Digital in 2012.

So, what does all of this mean? Well... it means that hard drives are guaranteed to do one thing... fail. It might be sooner, or it might be later... but they will fail. All archivists know that good archival practice dictates that your data should be backed up, and stored on three different pieces of media, in three different locations. Ideally, you would use a mix of hard drives and robust storage technology, like LTO data tapes. You are all doing this, aren't you? The unfortunate reality is that many times we think we'll get to it, and we don't. When I have more budget or when I have more time... I'll get to it. Don't wait until it is too late. Please use this article as a reminder and make sure your data is safe.

Links:

[Google Study on Hard Drive Failure](#)

[CMU Study of Accuracy of Mean Time Between Failure Data](#)

