Let’s Rethink Our Views of Storage Before It Is Too Late.

Don MacVittie, 2010-09-11

When I was in Radiographer (X-Ray Tech) training in the Army, we were told the cautionary tale of a man who walked into an emergency room with a hatchet in his forehead and blood everywhere. As the staff of the emergency room rushed to treat the man's very serious head injury, his condition continued to degrade. Blood everywhere, people rushing to and fro, the XRay tech with a portable XRay machine trying to squeeze in while nurses and doctors are working hard to keep the patient alive. And all the frenzied work failed. If you've ever been in an ER where a patient dies – particularly one that dies of traumatic injuries rather than long-term illness – it is difficult at best. You want to save everyone, but some people just don't make it. They're too injured, or came to the ER too late, or the precise injury is not treatable in the time available. It happens, but no one is in a good mood about it, and everyone is wondering if they could have done something different. In US emergency rooms at least, it is *very* rare that a patient dies and the reason lies in failure of the staff to take some crucial step. There are too many people in the room, too much policy and procedure built up, to fail at that level. And part of that policy and procedure was teaching us the cautionary tale. You see, the tale wasn't over with the death of the patient. The tale goes on to say that the coroner's report said the patient died not of a head injury, but of bleeding to death through a knife wound in his back. The story ends with the warning not to focus on the obvious injury so exclusively that you miss the other things going on with the patient. It was a lesson well learned, and I used it to good effect a couple of times in my eight years in Radiography.

Since the introduction of Hierarchical Storage Management (HSM) many years ago, the focus of many in the storage space is on managing the amount of data that is being stored on your system, optimizing access times and insuring that files are accessible to those who need them, when they need them. That's important stuff, our users count upon us to keep their files safe and serve up their unstructured data in a consistent and reliable manner. At this stage of the game we have automated tiering such as that offered by F5's ARX platform, we have remote storage for some data, we have cloud storage if there is overflow, there are backups, replications, snapshots, and even some cases of Continuous Data Protection… And all of these items focus on getting the data to users when they want in the most reliable manner possible.

But, like our cautionary tale above, it is far too easy to focus on one piece of the puzzle and miss the rest. The rest is that tons of your unstructured data is chaff. Yes indeed, you've got some fine golden grains of wheat that you are protecting, but to do so, today it is a common misperception to feel that you have to protect the chaff too. It's time for you to start pushing back, perhaps past time. The buildup of unnecessary files is costing the organization money and making it more difficult to manage the files that really are important to the day-to-day running of your organization.

My proposal is simple. Tell business leaders to clean up their act. Only keep what is necessary, stop hoarding files that were of marginal use when created, and negligible or no use today. We have treated storage as an essentially unlimited resource for long enough, time to say "well yes, but each disk we add to the storage hierarchy increases costs to the organization". Meet with business leaders and ask them to assign people to go through files. If your organization is like ones I've worked at, when someone leaves their entire user folder is kept, almost like a gravestone. Not necessarily touched, just kept. Most of those files aren't needed at all, and it becomes obvious after a couple of months which those are. So have your business units clean up after themselves. I've said it before, I'll say it again, IT is not in a position to decide what stays and what goes, only those deeply involved in the running of that bit of the business can make those calls.

The other option is to use whatever storage tiering mechanism you have to shuffle them off to neverland, but again, do you want a system making permanent delete decisions about a file that may not have been touched in two years but (perhaps) the law requires you keep for seven? You can do it, but it will always much better to have users police their own area, if you can.

While focused on availability of files, don't forget to deal with deletion of unneeded information. And there is a lot of it out there, if the enterprises I'm familiar with are any indication. Recruit business leaders, maybe take them a sample that shows them just how outdated or irrelevant some of their unstructured data is "the football pool for the 1997 season… Is that necessary?" is a good one. Unstructured storage needs are going to continue to grow, mitigated by tiering, enhanced resource utilization, compression, and dedupe, but why bother deduping or even saving a file that was needed for a short time and is now just a waste of space?

No, no it won't be easy to recruit such help. The business is worried about tomorrow, not last year. But convincing them that this is a necessary step to saving money for more projects tomorrow is part of what IT management does. And if you can convince them, you'll see a dramatic savings in space that might put off more drastic measures. If you can't convince them, then you'll need a way to "get rid of" those files without getting rid of them. Traditional archival storage or a Cloud Storage Gateway are both options in that case, but best to just recruit the help cleaning up the house.