dataSec

Data Classification- Part Two by Donald Poorman, CTO Evolution

When I last left you in part one of my series, "Taming the Data Storage Beast," I was prattling on towards the end about remembering that the word "Recovery," though largely ignored, makes up 50% of the phrase, "Disaster Recovery." In this paper, let's explore what can be done regarding data recovery once you have a good handle on the different categories your stuff falls into.

A Refresher in Data Classification

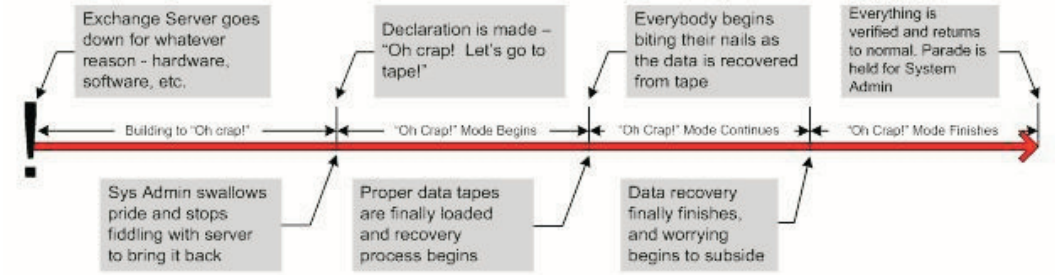
As you may recall my point from part one, not all data is created equal. We may think that our bad haiku poems or letters we've written home to our moms asking for more money rank in storage importance right up there with mission critical CRM data, but know deep down inside this should not be the case. Because of this, my idea was that every business that uses information technologies in its day to day operations produces data that falls into three categories:

- Mission Critical: Data that, if missing for anywhere from zero to five minutes, would cost the business money or harm to clients. For healthcare companies this would be patient data, for management companies this would be accounting data, etc.
- Mission Essential: Data that, if missing for anywhere from one to four hours, would degrade the business's operations and potentially cost the business money or harm to clients.
- Mission Relevant: Data that, if missing for more than four hours, would have some marginal affect on the business's operations.

Remember, each category is directly correlated to the amount of time and effort it takes to recover the information should a disaster strike. Because of this, cost for data availability goes up as the classification importance increases.

Evaluating the Typical Recovery Process and the "Oh Crap!" Factor

In order to better understand data criticality as it relates to ensuring its availability, let's take a look at a typical scenario of recovering data from a critical server outage, utilizing what I like to call the, "Oh crap!" factor (many people who know me know I'd use another term, but I will refrain from using it for more sensitive readers):



As you can see, there are other "soft" factors that guide the length of the timeline to bring data back from the dead that go beyond raw numbers tied to data transfer speed.

evolution
change for the better

dataSec

Data Classification- Part Two by Donald Poorman, CTO Evolution

OK, so recovery time REALLY matters with data classification. Tell me more!

Since the most predictable part of the "Oh crap!" disaster recovery process is the time it takes to get data from backup/alternate media to the primary storage device, data transfer speed is a big factor in determining the overall , let's first take a look at a very standard approach to recovery – getting it all back from tape. Take a look at these data transfer rates, based on some of the popular tape media types:

- DDS-4: 2.5 MB/Sec
- DLT-80: 6.0 MB/Sec
- DLT-VS160: 8.0 MB/Sec
- Super DLT-220: 11 MB/Sec
- AIT-1: 4 MB/Sec
- AIT-2: 6 MB/Sec
- LTO-1: 15-16MB/Sec

Don't forget - these transfer speeds are raw device-level numbers and don't take into account network latency and other anomalies that may slow the whole process down when recovering what you need.

With this in mind, let's figure out how much time it would take to recover the server from the above example, operating under the assumption that it had a total of 60 GB of storage available to it and the required tapes to recover everything are onsite and ready to go. Bear in mind, your mileage will vary in the first parts of the timescale based on system administrator pride, tape availability, etc.:

Buildup to "Oh Crap!":	**2 Hours**
"Oh Crap!", Part One (Finding the right tapes, etc.):	**1 Hour**
"Oh Crap!", Part Two (60 GB Data Recovery from a DLT-80 Tape):	**3 Hours**
Post-Recovery Actions (Verifying data, synchronizing logs, etc.):	**1 Hour**
Total Approximate Recovery Time:	*7 Hours*

Seven hours - almost an entire working day to recover what could be considered a more mission critical function like e-mail, CRM, etc.

Wow. I never realized tape recovery could be so slow! What are my alternatives?

First, an important disclaimer before you read on!!! At no point in time in this white paper am I advocating businesses discard tape backup altogether for disaster recovery…I merely think there are better backup options that can be used as a first line of defense, with tape as the final place to go for getting lost data back.

evolution
change for the better

dataSec

Data Classification- Part Two by Donald Poorman, CTO Evolution

First, an important disclaimer before you read on!!! At no point in time in this white paper am I advocating businesses discard tape backup altogether for disaster recovery…I merely think there are better backup options that can be used as a first line of defense, with tape as the final place to go for getting lost data back.

Alright, so we know tape can be slow, especially when the boss is calling every five minutes asking when he or she can get back to e-mailing jokes of the month or cute pictures of his or her kids to friends abroad. What can be done to speed up the most predictable part of data recovery (the transfer rate)? Again, many different options abound, with their prices commensurate with how quickly they work, so your choices should be in sync with the data classifications you've set. Bearing that in mind, let's explore some suggestions for recovery based on the classifications we've established above:

- **Mission Critical:** For this classification, "instant" recovery is the best option and can only be addressed with shared storage options like a SAN or a shared disk array because it involves utilizing hardware-level technologies like real-time mirrored LUNs or software-based solutions like volume replication. This approach should only be reserved for the most important data in your environment because of extremely elevated costs and complexities.

- **Mission Essential:** The first line of defense for recovering data in this category is to consider technologies like shadow copying or snapshot technologies. For shadow copying in the Microsoft Windows 2003 world, an added benefit to this method of data recovery for user files is that the end users can be empowered to recover their own data. For snapshot technologies, you'll have to utilize special packages for database-based data (i.e. Veritas for Microsoft Exchange or SQL 2000).

- **Mission Relevant:** At the bottom of the list, this data classification group can be handled in the traditional sense of recovery through alternate media means like tape. The only additional consideration for this category is to handle day to day backups with some form of hard disk, whether it's a USB 2.0 external hard drive, space on a direct-attached disk array, or a LUN on a SAN (if you're feeling really crazy). After all, giant wads of available disk space have gotten cheaper with the introduction of SATA technology to the storage array world.

Not to sound repetitious, but remember again that the above suggestions are first lines of defense for data recovery, and tape or other removable media should never be discounted as the second or third line of defense in preparation for a catastrophe like a giant anvil from a Road Runner cartoon squashing your primary data center and you having to rebuild everything from scratch.

If you use better first lines of defense for data recovery, you can be a better hero to your organization when things go wrong.

evolution
change for the better