



dataSec

Data Classification- Part One by Donald Poorman, CTO Evolution

A Lesson Learned

My family and I moved two years ago from a cramped townhouse to a larger single family house for, as my wife described, "...more room to stretch out into." Being the neurotic Type-A individual I am, I became excited with the organizational possibilities that laid before me with the new two-car garage that came with the house. I had BIG plans – shelves, cabinets, hand tool pegboards, and other organizational doodads were going to adorn my panacea for household organization. My strategy was to impress family, friends, and win bar bets by being able to describe, with laser-pinpoint accuracy where you could find my red handled pruning shears or keyhole saw without even being in the garage...I'm proud to be Type-A in that way.

Interestingly enough, a funny thing happened once we moved in. Between the unending chores of "settling in" and unpacking, combined with the obvious necessity that we decorate the entire house within the first six months, my plans for the garage went somewhat awry. After not being able to get my car inside one night, I took a survey of what had accumulated these past years and found some interesting things among the rubble on the floor: three paint cans with one brush stroke worth of paint left in them, two brand new water filters for our kitchen sink (I was oblivious that I had any and was planning on buying a new one), two boxes of infant baby clothes (my daughter is now three years old), and other sundry items that I'm sure mattered at some point in time. Needless to say, my next household project will be to install shelves and cabinets in the garage and keep the stuff that matters in there or the kitchen, and move the rest to the basement. At least I'll be able to park my car inside again. Not my wife's, but who's cleaning out the garage here?

At any rate, if you think about it, storing data in an enterprise can be much like how things ended up in my garage. Without a proper understanding what databases, documents, e-mail, and other elements make up your data, it's impossible to organize a way of providing proper availability and access to end users.

Data storage management is a rarely adhered-to practice. After all, storage has never been considered "sexy" – system administrators are generally more preoccupied with system stability, security, and keeping end users happy than worrying about how the data produced is being stored.

The Need to Clean House

With exponential increase in the amount of data, combined with the ever growing capacity of hard drives, it's very easy to duck, cover, and say something like, "Just add more disk space!" The fundamental flaw with this philosophy is the data growth causing the commotion generally represents roughly 10% of the total data in the enterprise – the rest more than likely lies unused in the dark, dusty regions of the hard drives where nobody visits anymore. To add to the mess, chances are good that there are also multiple copies of the same files, databases, or other sundry items lurking around as well. For example, think about how many people have that "Dancing Baby" video from a few years back still stored in their home directories on expensive RAID-5 disk that may be getting mirrored to another server via volume replication for quick recovery and you'll see where I'm headed...





dataSec

Data Classification- Part One by Donald Poorman, CTO Evolution

To make life easier and keep the bosses at bay, consider classifying your data and handling it accordingly once you get a handle on what you're storing.

Classifying Data

Every business that uses information systems to get them through their day to day operations has data that can be classified 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.

You may have noticed that my ties to data criticality are tied to tangible business impact like money or safety, not the "whine" factor that generally prevails when data is not available. You know what I mean here – the level of end user whining associated with data when it is unavailable. Don't let this factor fool you into thinking that an application like e-mail must be a business critical application because a lot of users will complain when it's down. Compared to patient record data or accounting data, e-mail's bearing on a business may not be as great as you may initially think. However, as a somewhat general example, you may find yourself categorizing your data as follows:

- Mission Critical: Patient records, accounting data, CRM data, ERP data
- Mission Essential: Commonly stored files (by department like accounting or operations), E-mail
- Mission Relevant: User home directories (usually the repository for the "Dancing Baby" films and such, legacy data, archived e-mail)

Dealing with Data Once It's Categorized

So, you've taken a closer look at what comprises your stored data - what's next? Remember, a data availability plan is about providing maximum accessibility to end users for the categories you've selected. Though you should match the appropriate data category to the appropriate storage media, not everything in your enterprise needs to reside on a storage area network (SAN) or a network attached storage (NAS) device!!!! This is not to preclude the fact that implementing these solutions for certain data categories is valid, but don't also buy the hype that storage-specific solutions like SANs or NAS are "silver bullets" that will cause your problems to go away as much as they are the magical cure that will heal your grandmother's gout or solve world hunger. For instance, SANs with real-time, block level replication are great for 99.999% uptime environments, but there are also volume replication software solutions that are just as viable and more cost effective.

The key to a better tomorrow for your data is to get to know your what you're storing, come up with a plan, and implement a storage infrastructure that makes technical (and fiscal) sense.





dataSec

Data Classification- Part One by Donald Poorman, CTO Evolution

Why Go Through the Trouble?

Besides having bragging rights and impressing your friends at parties before they shove you in a corner alone, proper planning and execution of a data classification endeavor has the following advantages:

- It's about economics. As we've already explored, thinking that dancing baby videos and patient records deserving equal time on the same replicated RAID-5 set on a \$120,000 disk array does not make obvious financial sense. Plus, I'd sure hate to be the patient whose records couldn't be entered or updated because my care provider ran out of disk space due to everybody downloading and saving their own copies of an animated "Dancing Santa" cartoon.

- It's about being proactive. OK, OK, OK. I know my vision of anything in the IT world being proactive is about as crazy as hoping the world will gather on a hilltop, drink Coke, and sing some hippy song, but having a good understanding of, and adhering to the information lifecycle can make planning for data growth a lot more bearable. How many times have you added disk space, or even worse, a brand new server because you were paged at 3:00 AM because all available system partitions were at 99% capacity? Even worse, how often have you frantically looked for files to delete to stave off a system failure due to full disk space?

- Because the word "recovery" is mostly ignored in the phrase, "Disaster Recovery". In part two of my series, Taming the Data Storage Beast, I will explore some of the misconceptions about disaster recovery, mainly the idea that if everything is backed up to tape, we can relax at night with a big bowl of nachos and watch "Survivor – The Hamptons" because the data is easily available and recoverable from another form of media. Don't be fooled into thinking that it is acceptable to wait 12 hours while an important database or large file structure gets recovered from tape...

Get to know your data – you'll be happy you did when the time comes to figure out what to do with it.

