

Storage Solutions

Beating back 'bad' data



History may very well show that what saved the day during a tumultuous financial period in 2008 and 2009 was not a silver bullet but, rather, hard work and strong leadership. A similar view can absolutely be taken when discussing data storage and the market reaction to the proliferation of data. That data is growing exponentially isn't news any longer, and one can easily find any of a number of different references regarding grains of sand, grains of rice, stars in the sky and so on to illustrate how quickly data is growing and how much data we now store. The fact is, customers create data every day some of which is useful but most of which will become redundant and/or duplicate. Don't believe me? Watch how quickly a 'joke' email is forwarded to a distribution list. Innocent enough, but now you've sent a 30MB file to 30 people who will open it, laugh, and never open it again...although probably not before forwarding it on to another 30 people. So, if you're counting, that's 30 original copies, each forwarded to 30 additional mates, that's 930 copies at 27.9GB of the exact same file not likely to ever be accessed again.

And how to find that 'bad' data whilst looking after the 'good' data such as sales spreadsheets, customer orders, maintenance agreements, and so on?

Gartner estimates that acquisition of storage accounts for 20% or less of the total cost of ownership for data over a three year period. The remaining 80% is comprised of operating expenditure; power, cooling, management, backup/recovery, business continuity, datacentre floor space, and so on. Many of our customers are running out of space, with rising power and cooling costs thanks to high oil prices. Storing shedloads of joke email and 'bad' data now seems a lot less attractive, but many corporate data strategies are 'flat' and store data on high cost Tier 1 storage. What to do?

Vendors have been addressing the very real expenses associated with data sprawl through the development of information lifecycle management tools and products. HP, one of Computacenter's vendor partners, certainly has an ILM strategy and product portfolio which merits serious investigation.

That said, this is where the hard work and strong leadership come in to play.

It would be wonderful if there were a silver bullet available in the form of a packaged storage product, but my view is that storage assessments which include data classification is a critical start to identifying where best to store data. The sad fact is that data has grown so quickly that many customers have not classified their data. Is that a problem? It depends on your perspective, but in the absence of classifying your data how could a customer develop a storage strategy which increases manageability, increases utilisation, and decreases the managed storage footprint all without increasing overhead?

Once we have classified and identified the data we can implement strategic information lifecycle management. Data deduplication, for example, is a great way to rid a corporate storage estate of 'bad' data, and 10:1 ratios are conservative when talking about data dedupe in a backup environment. Put another way, were we to deploy data dedupe in a customer environment we could expect to actively store 20TB of 'good' data for every 200TB of total 'bad' data volumes. And storage vendors are rapidly bringing data deduplication into primary storage arrays, but this is a future versus a 'ready to go now'.

Another example is to archive dormant data. If it hasn't been accessed in six months or more, moving it to an archive can significantly reduce OPEX costs. How? Well, most archive systems use SATA drives which consume 97% less power and cooling than primary storage array fibre channel drives and compression gives a 7:1 ratio or better a single archive can store as much data as seven primary storage arrays. Yes, access may take a few milliseconds more but think of the cost savings!

Ultimately we must provide sound counsel and strong leadership when working with our customers, and storage assessments which include data classification should be the starting point of any information lifecycle management project.

Computacenter's Storage Assessment service seeks to leverage the best of the existing Storage Infrastructure, Data Classification, and Backup Assessment offerings to provide a complete customer estate storage assessment at an attractive price point. The assessment collects data at a customer site for a thirty day period after which the data is collected, analysed, and a report with recommendations written and presented to the customer.