HAVE WE TAKEN DATA REDUNDANCY TOO FAR?

During a recent conversation about disk configuration and data redundancy on a storage array I began to think about everything we put into data redundancy. The question that came to mind is the title of this post *'Have we taken data redundancy too far?'*

Now don't get me wrong, I love data as much as the next fellow, and I definitely understand it's importance to the business and to compliance. I'm not advocating tossing out redundancy or data protection, etc etc. My question is when is enough enough, and or is there a better way?

To put this in perspective let's take a look at everything that stacks up to protect enterprise data:

Disks:

We start with the lowly disk, which by itself has no redundancy. While disks themselves tend to have one of the highest failure rates in the data center they have definitely come a long way. Many have the ability to self protect and warn of impending failure at a low level, and they can last for years without issue.

Disks alone are a single point-of-failure in which all data on the disk is lost if the drive fails. Because of this we've worked to come up with better ways to apply redundancy to the data. The simplest form of this is RAID.

RAID:

RAID is 'Redundant Array of Inexpensive Disks', it's also correct to say it 'Redundant Array of Independent Disks.' No matter what you call it RAID allows what would typically be a single disk on its own to act as part of a group of disks for the purposes of redundancy, performance or both. You can think of this like disk clustering.

Some common RAID types used for redundancy are:

- RAID 0 – Disk striping, data is striped across 2 disks to improve performance, each disk becomes a single point of failure for the entirety of the stored data.

- RAID 1 – Disk mirroring, data is written to both disks simultaneously as an exact copy. This method allows either disk to fail with no data loss.

- RAID 5 – Raid 5 is striping with parity. What this mean is that using three or more disks all data is written in stripes across available disks and additional parity data is striped across the disks to provide redundancy. Because of the parity data one disk can be lost from the group without data loss. Raid 5 is N-1 capacity, meaning you lose 1 disk worth of space to the parity data.

- RAID 6 – Disk striping with double parity. Think RAID 5 with an extra disk lost for parity but the ability to lose two disks without data loss. This is N-2 capacity.

In many cases 'hot-spares' will also be added to the RAID groups. The purpose of a hot'-spare is to have a drive online but not participating in the RAID for failure events. If a RAID disk fails the hot-spare can be used to replace it immediately until an administrator can swap the bad drive.

Snapshots:

Another level of redundancy many enterprise storage arrays will use is snapshots. Snapshots can be used to perform point-in-time recoveries. Basically when a snapshot is taken it locks the associated blocks of data ensuring they are not modified without copying them. If a block needs to be changed it will be written in a new location without effecting the original. In order to revert to a snapshot the change data is simply removed leaving the original locked blocks. While snapshots are not a backup or redundancy feature on their own they can be used as part of other systems, and are excellent for development environments where testing is required on various data sets, etc. Snapshots consume additional space as two copies are kept of any locked block that is changed.

Primary/Secondary replication:

Another method for creating data redundancy is tiered storage. In a tiered redundancy model the primary storage serving the applications is held on the highest performing disk and data is backed up or replicated to lower performance less expensive disk or disk arrays.

Virtual Tape Libraries (VTL):

Virtual tape libraries are storage arrays that present themselves as standard tape libraries for the purposes of backup and archiving. VTL is typically used in between primary storage and actual tape backups as a means of decreasing the backup window.

Tape backups:

In most cases the last stop for backup and archiving is still tape. This is because tape is cheap, high density, and ultra-portable. Large amounts of data can be streamed to tape libraries which can store the data and allow tapes to be sent to off-site storage facilities.

Adding it up:

When you put these redundancy and recovery systems together and start layering them on top of one another you end up with high ratios of storage media being purposed for redundancy and recovery compared to the actual data being served. 10:1, 20:1, 100:1 or more is not uncommon when considering archive/redundancy space compared to usable space.

Source: http://www.definethecloud.net/have-we-taken-data-redundancy-too-far/