MAXIMIZING YOUR STORAGE COST CUTTING EFFORTS

Storage Switzerland, LLC

In 2010 an obvious theme is cutting costs. In IT and storage in particular that cost cutting will be popularized by data archiving and data reduction vendors. Both are very legitimate steps to take.

In the typical data center a good rule of thumb is that 80% of the data on primary storage can be migrated to an archive platform. The user buys the archive solution of their choice, then migrates all of that data, but where is the ROI? Clearly after archiving there are acres of free disk space on the primary storage platforms and some ROI can be derived if this process delays the purchase of additional storage, but the storage administrator can not start turning off empty drives. There aren't any. The way most storage arrays work, the remaining data is still scattered across all the drives in each LUN.

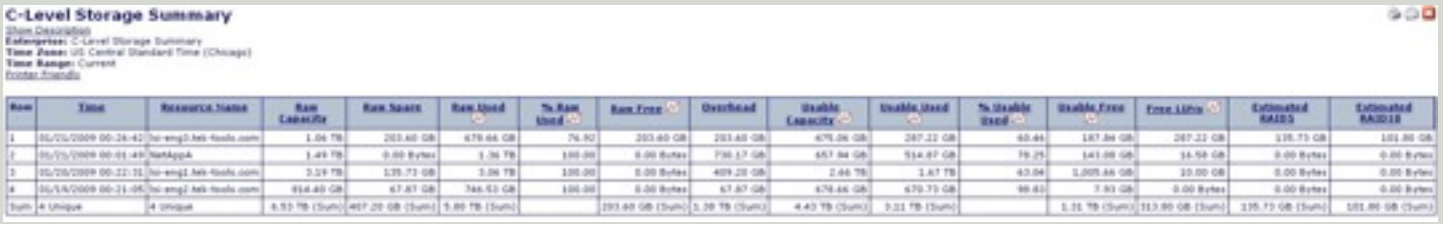
The problem with the archive first and ask questions later is that no analysis was done to determine what data is where and what data can be relocated somewhere else in the system. Tools exist today from companies like SolarWinds that provide independent third party analysis of the storage environment and how applications interact with that environment. In short to get maximum benefit when driving out costs you need to first drive in efficiency through a proper analysis of the environment.

Inventory of Data Set

The first step in maximizing your storage cost cutting efforts is to understand what those storage assets are and what they contain. This will not only include a simple file view of data that has not been accessed and can be moved to secondary storage but will also provide an analysis of what that data is and what type of optimization should be applied to it during movement.

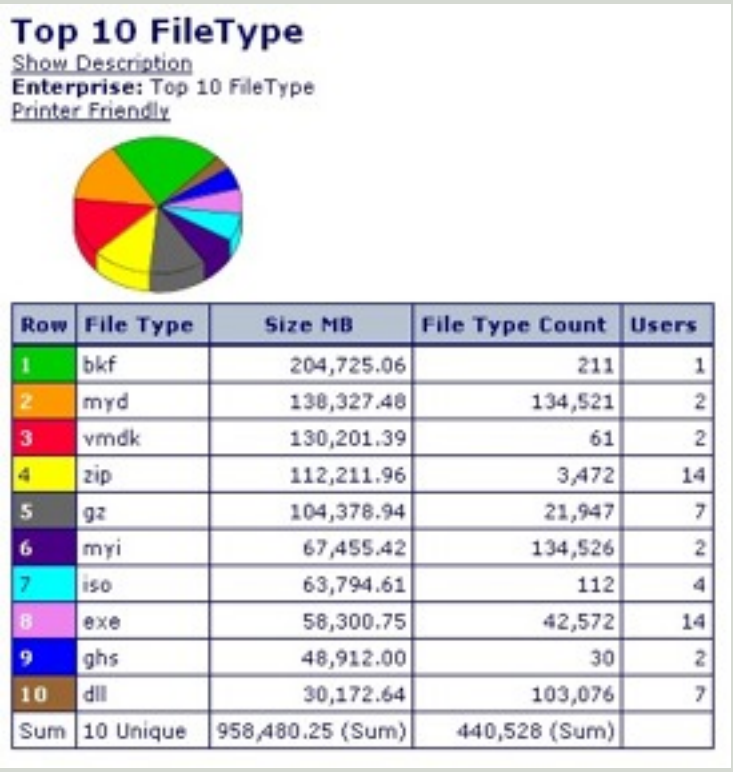
For example is that data an ideal candidate for some form of compression or deduplication? If so which method should be used? There are solutions that are focused on providing optimization of standard files like VMware images and then there are those that specialize in optimizing .jpgs and PDF files. An analysis tool will indicate what type of data is in the archival candidate set. Most data reduction technologies cost about the same, the ROI that an analysis tool brings to this step is selecting the technique that is going to provide the maximum reduction for your investment.

A third and critical parameter is how fast will data be added to the archive. Again with a proper analysis tool those factors can be examined and even trended for future growth. The ROI here is derived from knowing how large of an archive to invest in upfront and to properly budget for not only archive but also primary storage investments in the future.



Mapping the Remains

Cutting storage costs can be as simple as eliminating shelves of unused capacity.



Top 10 FileType

Show Description
Enterprise: Top 10 FileType
Printer Friendly

Row	File Type	Size MB	File Type Count	Users
1	bkf	204,725.06	211	1
2	myd	138,327.48	134,521	2
3	vmdk	130,201.39	61	2
4	zip	112,211.96	3,472	14
5	gz	104,378.94	21,947	7
6	myi	67,455.42	134,526	2
7	iso	63,794.61	112	4
8	exe	58,300.75	42,572	14
9	ghs	48,912.00	30	2
10	dll	30,172.64	103,076	7
Sum	10 Unique	958,480.25 (Sum)	440,528 (Sum)	

The challenge is that in most data centers this capacity is not represented by racks of storage shelves in an array that have no data on them. More likely it is available capacity that is scattered throughout the environment, none of it in a contiguous block and certainly not on one contiguous shelf.

The next step in maximizing storage cost cutting efforts is to understand the data sets that are left on primary storage. That 20% may be small but it is active, likely critical and as stated earlier, scattered all over the available disk shelves. Ideally the goal is to consolidate this data on to a smaller number of storage platforms and a smaller number of disk shelves being careful to balance cost savings vs. data vulnerability. A data analysis tool like SolarWinds Profiler will provide a complete inventory of where the remaining data set actually is. From there the storage administrator can use the tools often provided by the array vendor to consolidate LUNs and reduce the number of shelves required. These actions will which eventually will result in turning off the shelves, possibly some storage controllers, and will lead to dramatic hard dollar savings in power consumption and the soft dollar savings of reduced management overhead.

The ability to consolidate this data is going to require not only knowing that there is excess capacity but also where to place the actual data for maximum utilization.

Inventory of Performance

During the process of LUN consolidation it is also valuable to understand what performance the storage systems can deliver and what performance the applications need. This understanding represents another ROI point for analysis by making sure that applications are operating at peak performance and thereby utilizing the proper amount of storage. An application that can't or does not need to drive data at the full speed of the storage architecture will gain nothing from being on that storage architecture.

Most storage arrays provide similar if not identical data protection like snapshots and replication across all supported performance tiers. The remaining differentiator is the performance those tiers provide and the cost to acquire and power those tiers. For example, faster drives that are smaller in capacity require more physical units, often consume more power on a per unit basis and as a result will require more shelves for the same capacity as larger capacity, slower tiers. Again, acquiring and powering shelves costs money.

A tool that can analyze both the performance aspects of the storage and the performance requirements of the application will allow the storage administer to correctly match workloads with performance capabilities and optimize the cost reduction.

This step goes hand in hand with the consolidation of LUNs because relocating data onto fewer storage shelves is also going to impact the performance capabilities of the storage that is receiving the relocated data.

File Age Categories Summary

Count

Size

Resources: 19 Root dirs : 43

Days Range	Count	Count %	Size (GB)	Size %
0 - 60 days :	200,399	10.45	333.51	27.47
60 - 120 days :	139,888	7.29	386.03	31.79
120 - 180 days :	94,587	4.93	109.23	9.00
180 - 365 days :	452,683	23.60	224.80	18.51
> 365 days :	1,030,764	53.73	160.64	13.23
Total:	1,918,321		1,214.22	

Maximum Resource Utilization Requires Constant Monitoring

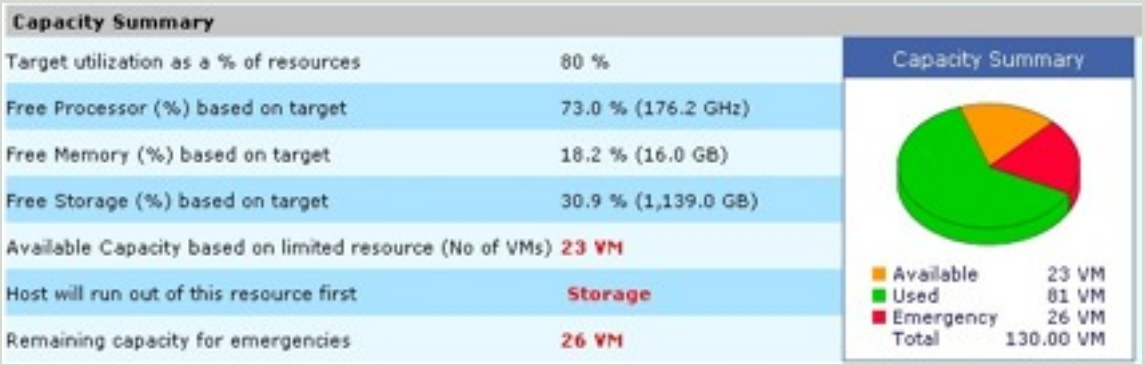
Manual analysis of just the inventory components of this process could be overwhelming to the IT staff, especially one whose levels have more than likely been stretched thinner than ever. Automating this process is a key requirement.

Once the commitment is made to use every IT resource to its safe maximum, the environment needs to be actively monitored and trended. Even if the time and resources were available to manually inventory and analyze the above parameters, which is unlikely since it would take weeks if not months in a data center where all the components are being used to their full potential without excess, a spike in utilization needs to be quickly identified or alerted to (even predicted) so adjustments to the environment can be made.

Server Virtualization

Of course as part of any cost savings effort the subject of server virtualization (or expanding the server virtualization effort) comes up. The tools used to analyze and provide real time monitoring of the environment, as described above, can also provide real world analysis of servers that are ideal candidates for virtualization. For example SolarWinds Profiler uses real world statistics and then maps those into a simulator so that performance impact to both the virtual host and the newly virtualized application can now be measured. For more details on this, see our article on Maximizing your Virtualization Project.

In the current economic situation excess capacity, performance, computing power or even archive areas are all expenses that cannot be tolerated. Most likely, even as the economy turns around, data centers will not be allowed to return to wasting budget dollars in excess. The optimized IT environment is a permanent condition. Using realtime analysis tools not only make that an achievable goal, they do so while increasing the overall efficiency of the IT staff, returning to it the most valuable asset of all tim;. time to do their job and time to go home on time.



About Storage Switzerland

Storage Switzerland, is an analyst firm focused on the virtualization and storage marketplaces. Storage Switzerland provides strategic consulting and analysis to storage users, suppliers and integrators.