The Prescription for Storage Pain:

Prescription without a Diagnosis is Malpractice

By Jerome M. Wendt – Independent Storage Analyst

TeraCloud
Storage Analytics: the value of knowing

Contents

The Storage Pain

What do I have? How is it being used? Do I need more? What should I buy next? Who should I buy it from? These questions plague corporate storage managers as they evaluate their increasingly complex heterogeneous infrastructure and try to assess what their next moves will be. For with storage an ever larger component of IT expenditures, corporate decision makers push storage managers to provide more thorough technical explanations to justify financial outlays associated with storage.

Examining and understanding their storage infrastructure is no simple feat for storage managers as major hurdles hamper their analysis. Software tools that gather this data and provide this information must first be internally justified. If justified, internal corporate walls inhibit managers from gathering the data. And if these are cleared, then the tools must be configured to gather data from various data points within the infrastructure.

But gathering data is only the beginning. For once gathered, data must be corroborated and a diagnosis made. Yet most managers likely lack the personnel, skill and time to perform the analysis necessary to transform raw data into a reliable technical recommendation. Without these hard facts, storage managers err on the side of caution and purchase storage hardware that delivers higher levels of availability, capacity and performance than application requirements may justify.

This oft employed tactic to overbuy is not without merits. Higher end hardware meets a wider breadth of application requirements and scales to higher capacities. The level of vendor support to help diagnose and troubleshoot problems is more extensive and it is simpler to manage since it generally only requires one management tool.

Yet the financial downside of over allocating server and storage resources is substantial. Forrester Research determined that in 2002 a server utilization rate of only 60 percent translated into $20 billion in wasted server resources while the largest 3,500 firms spent an average of $6.4 million in 2003 on systems integration costs. Gartner more recently determined that corporations spend 70 to 80% of their total IT budgets supporting established applications and required infrastructure components.

With the gap in prices between high end and mid-tier hardware significant but the costs of over allocation on high end equipment equally heady, management is less tolerant of buying excess capacity for low revenue applications. Yet without the means to document existing environments and establish where excess capacity and performance exists, storage managers are justifiably reluctant to downgrade to lower tier equipment.

This inability to report on and analyze their storage infrastructure is further hampered by SRM (Storage Resource Management) tools that promise analysis but primarily provide management. As a result, over purchasing, over allocating, and over provisioning storage hardware continues. Breaking this pattern requires the introduction of a new tool that ignores the management of solitary operating systems and hardware devices but rather looks to examine the storage infrastructure and give organizations the information they need to make tactical and strategic storage decisions.

"IT organizations are not looking for yet another storage report, but there is a strong desire for true analytics – insight in to the environment with actionable best practices. Solutions that provide true analytics should be well received by the marketplace."

Dave Russell, Gartner Research Director, Storage Software

The Emergence of Storage Analytics

The emergence of storage analytics is the result of the inability of existing tools and methodologies to give storage administrators what they really need – persistent point in time analysis of their environment. Unlike SRM tools, which attempt to manage the specific infrastructure components, a storage analytics tool collects data from multiple data points to form the basis for understanding the relationships that exist between the different components within the storage infrastructure and the impact that may result from changing these either the settings or devices. Examples of data points within the infrastructure that users may collect data on to analyze and understand the infrastructure include:

- z/OS Catalogs
- UNIX and Windows Volume Managers
- Datasets
- DFSMShsm
- Logical disk volumes
- Physical disk arrays
- Tape volumes

For shops with mainframe environments, a logical place to begin is with the z/OS catalog itself. Containing pointers to where the data resides in the system, storage analytics tools help to identify pointers to volumes that do not contain datasets. From a technical perspective, proactively identifying these incorrectly set pointers should help to prevent future job failures, application outages and wasted support time. This also could translate into financial benefits by avoiding potential penalties resulting from missed service level agreements (SLAs) and an improved perception with the client by being able to demonstrate how the support organization anticipates problems before they occur.

Using storage analytics tools in conjunction with volume managers on UNIX and Windows systems offer similar opportunities for identifying and reclaiming unused storage resources. Large organizations often support hundreds of servers that each may in turn control tens or hundreds of LUNs. Up-to-date information on how these servers are using their assigned storage resources, or if they are even using them at all, is knowledge storage managers often lack. Tracking this information becomes even more problematic for organizations that are using clustering technologies or are consolidating their servers and backend storage resources using virtual partitioning technologies such as AIX lpars, Sun domains or Windows VMWare where LUNs are shared between different partitions.

Storage analytic tools should provide managers insight into how SAN based LUN discovery and allocation procedure works. For in SAN environments, it is possible for LUNs to be allocated at the storage array or network level but never discovered at the operating system or volume manager level, especially if different groups manage different parts of the process as is the case in most large shops. Identifying these allocated but undiscovered LUNS removes the need to provision or purchase additional storage or allows the organization to reclaim and reuse these LUNs for other purposes. Analytics tools should also enable managers to identify improperly configured clusters servers so that if two clustered servers

are supposed to share the same back end LUNs for failover purposes and do not, this problem can be spotted and fixed prior to a failover occurring.

The ability for a storage analytic tool to report on mainframe datasets gives organizations the ability to improve the management of their allocated datasets and either reclaim or better utilize their space. Within four of the five types of allocated datasets, empty, duplicate, uncataloged and unreferenced, users should reap the following benefits by analyzing this part of their infrastructure:

▸ Isolate which datasets need idle space and how much they need to optimize performance. It also should pinpoint which datasets are empty and unopened and can be returned to the general storage pool.

▸ Detect duplicate datasets and delineate between which ones are reserved for system volumes and which ones can be deleted or archived.

▸ Highlight uncataloged datasets that otherwise would escape notice.

▸ Report on unreferenced datasets experience infrequent activity and should be moved to lower cost disk or tape.

For the fifth type of allocated dataset, those that when better managed can improve performance, a storage analytics tool should recognize those instances where CI and CA splits occur, such as when a new record does not fit into an existing CI or CA so the data is split between two CIs or CAs. By identifying the wasted space resulting from neither the original or new CI nor CA filling up with data, users can identify where these splits are and anticipate and take action to prevent potential performance slows before they occur since CI and CA splits require more I/O operations to access dispersed data.

Using storage analytics in conjunction with DFSMShsm allows organizations to spot and reclaim unused or improperly tagged resources. For instance, should DFSMShsm experience problems with a tape volume, DFSMShsm marks it INVALID and sets it aside. A storage analytics should recognize that once a tape volume is in this state, it is no longer accessible to HSM and not available for any other purpose. To correct this, the storage analytics tool should generate an exception report and ideally even produce a set of action sets that a user can take to correct the problem, in this case an inquiry of the OCDS (offline control dataset) to remove the INVALID tag placed on the volume so any active data on the volume can be consolidated on other tapes.

The tool should also have the ability to recognize when DFSMShsm creates a thrashing dataset and report the effect it has on the system. Occurring when a dataset is migrated by DFSMShsm to either virtual or real tape and then is quickly recalled, this becomes especially problematic for those datasets that are continually migrated from disk to tape and back again. The tool should be able to detect when this is occurring and report on the wasted CPU, disk and tape resources and slowed application response times. Ideally, it should even be able identify the deeper problem that may exist, such as an incorrectly assigned management class, and help to point users to a course of action to take.

Another benefit that users should realize with storage analytics is improved storage monitoring and management of both utilization and performance. By allowing volumes to grow over 80% utilization will decrease the amount of wasted space but also identifying when they exceed that preset threshold of 90 or 95% allow companies to begin to move to a just in time storage provisioning process. By also monitoring and gathering performance statistics on LUNs, it helps storage managers identify if they may move the volumes to higher capacity but lower performing LUNs either now or in the future.

Conversely, identifying hot spots on the back end storage arrays also allows managers to anticipate and react to problems before they spin out of control. Though logical volumes are generally dispersed across the backs end of a storage array, as storage arrays add capacity, performance problem can start to surface since administrators can easily lose track of which data resides on which logical volumes. The storage analytics tool allows administrators to pull this data together and spot if frequently accessed data ends up on logical volumes that reside on the same back end RAID array on the same controller.

In UNIX and Windows servers, a storage analytics tool needs to solve additional problems. First, SAN attached storage arrays support different size volumes and LUNs may be an amalgamation of smaller volumes, so an analytics tool needs to both discover these LUNs and report on their underlying volume makeup. Second, with LUN security used to control which servers can access specific volumes and multiple servers of the same OS potentially accessing the same volume, it should track which volumes are assigned to specific servers to track application usage as well as when servers are decommissioned so the LUN security may be removed and the storage recovered.

The final component that a storage analytic tool should monitor is tape media – virtual or real. With virtual tape loosing the unlimited capacity typically associated with tape libraries, storage analytics need to assist managers in maximizing the utilization of space on the tape libraries by identifying data that is expiring or old and noting what should be moved off to real. And with real tape capacities increasing, these tools need to enable managers to take advantage of this additional capacity by both identifying partially full tapes and tapes with expired data on them so their capacity can be either utilized or reclaimed.

So with all of this information from all of these different data points, organizations can for the first time begin to get a big picture of what their environment truly looks like and what is going on in their storage environment from multiple different viewpoints. Yet for organizations to understand the full benefits of storage analytics, they also need to understand how it differentiates itself from traditional SRM products.

Separation from SRM

How beneficial the storage analytics tool ultimately will be to the organization will hinge on its ability to differentiate itself from other SRM tools they may already have in their environment. The degree of separation from the two will be determined by the storage analytic's tool to produce reports that provide increasingly higher levels of information and insight into a user's storage environment.

LISTINGS	Producing listings of the information collected from the storage environment or an inventory of hardware devices and at this level storage analytic tools may be virtually indistinguishable from SRM tools depending on the size of the environment.
TECHNICAL ANALYTICS	Storage analytics and straight SRM tools start to separate at this level. Storage analytics tools pull together technical data from multiple points and multiple vendors products and correlate them to highlight specific problems or opportunities for saving to provide managers with the next level of storage insight. Using more sophisticated tracking tools and algorithms, they should identify technical problems such as thrashing datasets or servers where capacity could be reclaimed or better utilized.
APPLICATION ANALYTICS	At this level, storage analytic tools begin to distinguish themselves from SRM tools. With this level of insight, users can document how changes to the storage infrastructure will impact an application and determine what new storage components to introduce into their environment.
BUSINESS ANALYTICS	By adding financial data into the analysis, organizations can associate revenues or costs with the application and diagnose how profitable or unprofitable a certain application is. This allows storage managers to justify the purchase of the right storage resources to maintain the application's profitability or help identify what changes need to be made to stem the losses.

These increasing levels of analysis give storage managers the higher levels of justifications and rationale they need to take not just the right technical actions, but the right business actions based upon solid and verifiable technical, application and financial data. From a technical perspective, it documents what is occurring in the storage environment and what resources are already available that may rectify the problem and avoids the canned, ready made technical justification for new storage hardware. From an application perspective, it allows storage departments to work more closely with the application owner and help ensure the stability of the application. And from a business perspective, organizations can more easily identify what applications are making or losing money and make plans on how to keep them profitable, make them profitable or cut their losses.

The Prescription for Storage Pain

Storage managers can no longer afford to make million dollar storage decisions using hundred dollar spreadsheets. With storage environments rapidly growing, changing and evolving, storage managers and administrators are taxed just keeping up with changes to their storage environment and have little time to maintain and update static documents that are out of date as soon as they are created.

Yet buying SRM tools that provide simple listings of storage environments or provide questionable abilities to manage storage hardware are not viable options either. Knowing what your storage environment looks like and contains is a vital first step but insufficient for storage managers who need solid analysis and hard facts to justify changes to their storage environments. And promoting SRM tools as replacements for vendor supplied storage management tools that only work part or most of the time is worse than having no tools at all.

The necessity for storage analytics tools lies in their ability to unobtrusively and easily collect data from multiple sources in the storage environment and turn that into data into understandable and actionable information. Doing that requires storage analytics tools that include algorithms that marry and analyze collected and supplied data to provide the technical, application and financial justifications that storage managers need to use to justify storage purchases.

TeraCloud's Storage Analytics is one such service storage managers should consider using for these justifications. Designed to gather data from AIX, Red Hat Linux, Sun Solaris, Windows and z/OS operating systems as well as from major storage arrays, tape libraries and SAN devices, it touches and correlates data from major points within the storage infrastructures. And giving managers the ability to document and understand their storage infrastructure is the proper first step towards optimizing the infrastructure they already own and justifying the components for the storage infrastructure they need to build.

About Us

TeraCloud storage-management solutions provide a comprehensive space-management, capacity planning, and data-management solution for mainframe and heterogeneous storage environments including consolidated mainframe and enterprise-wide monitoring, detection, analysis, and automated resolutions for DAS, NAS, and SAN environments.

This helps our customers centrally manage and control their data center storage as a strategic asset. Our mission is to help you optimize your data management practices and environment by delivering storage inventory solutions that are easy to use. With comprehensive storage management solutions to monitor, detect, analyze, and tune your storage inventory, TeraCloud keeps data and applications available in the most economical way, letting you save time, save money and regain control.