THE CLIPPER GROUP
Navigator™
Navigating Information Technology Horizons℠
Published Since 1993 Report #TCG2010038 August 20, 2010

NetApp's Storage Efficiency — Why Less is More

Analyst: Anne MacFarland

Management Summary

Buying a camper's backpack presents serious challenges. Not only do you have to worry about the fit and bells and whistles that make it distinctive, you really want to pick one of the *right size*. One natural tendency is to find the largest capacity pack that comfortably fits, so that you are prepared for tan extended trip into the woods. Unfortunately, most folks rarely need that extreme capacity. Most of the time, something smaller would be better. Why? Ignoring the fact that big packs always cost more, you are carrying around a lot of unused capacity. This is a burden on your back and it may mean that the pack isn't as balanced as it might be if more fully loaded. **Finding the *right size* for your projected needs is important. This requires more forethought than just looking at capacity.** Additionally, you might need some help in selecting what you really need. The same is true for sizing your next storage procurement.

Recently, NetApp was asked to bid on 65 TBs of storage for a data center transformation project. It came back with a suggested bid of 27 TBs – all that it said would be needed. NetApp won the bid – and the final implementation was 23 TBs. How could they do this? Why would they do this? The answer to the first question is a set of capabilities, which NetApp collectively calls *Storage Efficiency* that comes with its storage arrays. They all leverage NetApp's Data *ONTAP*, *WAFL*[1], or both. Some must be licensed. *All must be turned on* to create the efficiencies that reduce the storage capacities needed to support applications and business processes. Fewer disks draw less energy and reduce data center cooling expense. This benefit keeps on giving.

Why would NetApp do this? In an era of quickly growing data, NetApp's compelling efficiency attracts customers, and puts other less efficient storage management scenarios to shame. **There is a need for a fresh look at storage management, because the use of virtual machines, together with scale-out applications, has wreaked havoc with traditional storage practices.** Even careful calculations no longer can predict where hot spots will occur. **Since hot spots cannot be predicted, storage system must provide optimal performance under all conditions and efficiency should not compromise performance.**

To date, storage efficiencies have been focused on secondary storage or backup. It is time to expand that focus to primary storage. By leveraging multi-core processors and adding NetApp's *Flash Cache*, high-performance requirements can be met while efficiencies keep costs under control. Storage efficiency and performance *can* be synergistic if the system is designed that way. Read on for more details on how NetApp has done just that.

IN THIS ISSUE

➢ Foundations of Storage Efficiency 2
➢ Conclusion 4

[1] NetApp claims WAFL *(Write Anywhere File Layout)* is the first file system optimized for random access (i. e., disk.). It inherently stripes data opportunistically, creating a layer of virtualization that is the basis of NetApp's Storage Efficiency.

The Clipper Group, Inc. - Technology Acquisition Consultants ◆ Internet Publisher
One Forest Green Road ◆ Rye, New Hampshire 03870 ◆ U.S.A. ◆ 781-235-0085 ◆ 781-235-5454 FAX
Visit Clipper at ***www.clipper.com*** ◆ Send comments to ***editor@clipper.com***

Foundations of Storage Efficiency

Efficiency is a matter of habit. Good habits, in the context of a data center, become consistent practices. Automating these practices can lead to more scalable IT operations. NetApp has endowed its infrastructure with the following good habits that meet the needs of their customers who are struggling to cope with the growth of their business information. Some are well-established practices, while others are new. Using all of them has a cumulative effect that is evident in both capacity and management savings.

Create Less

Making copies – full copies – of data comes naturally to most of us. It supports independent use. It assures *anytime* access to the data. Repositories are a comfort, just like all the equipment you put in your backpack, just in case you might need it. However, in data center terms, surplus copies of data are a bad habit that creates numerous problems. They do not stay in sync with official data and may not be backed up. They fill disk capacities. It is hard to know which is the correct version to use.

Incremental copies of only changed data were probably the earliest examples of storage efficiency. They started traditional backup down a path away from repetitive full backups and toward a continuous approach to data protection.[2] Development, testing, application cloning, and collaborative initiatives all call for solutions that support data use without creating an unnecessary *more.*

NetApp offers the following mitigations to the many problems of shadow data. All are part of the *Data ONTAP* operating system, and should be leveraged as a matter of course.

NetApp Flex Clones

NetApp *FlexClones* are writable snapshots. This makes them useful for testing and many kinds of collaborative endeavors. FlexClones can be synced or they may be used as separate versions. They journal all changes, and take up less space than traditional versioning. This is a far less wasteful approach than when groups of people each save the versions they deem useful. FlexClones have been around since 2006.

Thin Provisioning

With NetApp *Thin Provisioning,* storage space is provisioned only when it is written. In addition, reserve capacity is set, as a default, to zero, not to the maximum expectations dictated by the application or administrator. The *less* that is supported is that of allocated space. This approach allows provisioning beyond actual physical capacities, and alerts when there is an actual capacity problem.

RAID-DP

RAID was developed to mitigate the unreliability of disk drives that is inherent in their mechanical nature. The ease of data replication does not obsolesce the first line of defense that is RAID. Replication protects against data corruption or loss of service, while RAID protects against device failure. Mirroring, or RAID 1, requires a full doubling of capacity. Double Parity RAID is far more cost effective than mirroring, though it comes with a slight performance penalty on writes. With the increased capacity of disk drives, many vendors, including NetApp, have moved to double parity RAID. Because of the nature of NetApp's WAFL, NetApp's double parity RAID carries less performance overhead, since WAFL has never updated blocks in place, and does not have to accommodate that scenario.

Shrink What You Have Created

Think of the stuff sack[3] that can compress a sleeping bag to fit more compactly in the backpack described on page one. Compression and de-duplication each provide capacity shrinkage.

Compression and Deduplication

For NetApp, both compression and deduplication work at the same level. Both compression and deduplication can be used together. Either order will work, though usually whatever is in-line goes first. The key is that they be used in a way that is automated and consistent, both on reduction and on retrieval. NetApp makes both features of *Data ONTAP* that comes with the arrays. They just have to be licensed and turned on.

The recent popularity of application scale-out by a clone-and-tweak process using hypervisors, such as those from VMware presents many instances of redundant information – and thus many opportunities to shrink it. In near-clone situations, like virtual desktop applications, the shrinkage can be up to 90%, because most of the data to be stored (the operating environment, for instance) is identical for each desktop. In other situations where an application is replicated as

[2] Of course, how far a business goes down that path is a trade-off between mitigating risk and incurring cost.

[3] A compression stuff sack is a bag with radial straps that can compress the volume of bulky items, such as sleeping bags.

multiple, parallelized instances, deduplication also is conspicuously effective. By contrast, storage of images is a situation where deduplication is not very effective.

Leverage Storage Efficiency to Optimize Data Services, Such As Protection and Archiving

Storage efficiency is a benefit that keeps on giving. Every storage utility that involves moving, replicating, or even presenting data across a network benefits from having to deal with less capacity. Take a minute to review all the NetApp Utilities that benefit from storage efficiency.

Data Protection

- *SnapMirror* works at the volume or LUN level on the array (not on a virtual machine or server). Thus, the SnapMirror replication does not affect the application environment.
- *SnapVault* handles disk-to-disk backup.
- *Data Motion* supports transparent migration of stored volumes for load balancing or for archiving.

For all of these processes, the use of NetApp storage efficiency elements reduces demands on the storage resources. The same is true for all the utility processes invoked by applications. The effect is cumulative. You get efficiencies by creating less, more by deduplication and compression, and *STILL MORE* every time data when is replicated or migrated.

The Proof – NetApp AutoSupport

NetApp surfaces and visualizes these efficiencies through something called *My AutoSupport* (currently v1.8). Any customer with a maintenance contract can, at no cost, call up their own data center efficiency profile, and compare it to what it would have been without the storage efficiency features. It is an easy way to assess the effectiveness of an efficiency strategy.

Because a good benefit needs to be celebrated, NetApp customers that use these Storage Efficiency tools like to brag in the positive, not the negative. Being 300% more efficient sounds better than saving a third. At the last NetApp Analyst conference, one NetApp customer from Microsoft, Cory McKee, said that he had achieved over 1,000% efficiency from deploying the NetApp storage efficiency strategy. [4]

Accelerate Response Time

All the efficiencies in the world will not be satisfactory, if they negatively affect business operations. Years ago, all of these storage efficiency strategies would have had such a negative impact on getting work done. What has changed is that processors now have the multiple cores and multiple threads that allow them to operate independently and in parallel. They can support multiple processes without one process affecting another.

This leads to a need for load balancing and further optimization. Classically, in networked storage, IOPS are spread across spindles using striping. The old way of upping response time was to leave spindles lightly loaded – which, of course, increased storage sprawl. Now, with server virtualization, requests from multiple virtual servers are intermingled in the same data stream. The predictable patterns of an application's data use can be interrupted and the IOPS demand can be intense and unpredictable. Therefore, in addition to a multi-core controller, NetApp figured it would make sense to have a tier of flash or Solid State Disks (SSDs) as a cache. While such a tier is expensive, it gives much faster response time. Used with good caching algorithms, there is no wasted space. So, NetApp developed a *Performance Assist Module (PAM)*[5], now known as *Flash Cache.*

Flash Cache

Flash Cache is a random read assist. It is a plug-in board that also requires a software license. Here are some of the ways it can be used.

- *Priority-based caching* is the classic caching of *hot* data. If it is known to be hot, it can be immediately cached on write.
- *Single instance caching* is the caching of data used by multiple processes, applications, or users. Having it in a cache layer eliminates the locking contention that disk reads can produce.
- *Predictive caching* (focused on random reads) is a well-honed art for large, unvirtualized applications.
- *Metadata cache* is a strategy used in where the application leverages directories or catalogs. Many Web applications work this way. In the time a user takes to log in and select an option, something which is handled in SSD cache, the rest of the related content can be staged for immediate access. In consumer-facing service

[4] Yes, virtual desktops were a part of that spectacular result, as well as faster builds and other deployment processes.

[5] PAM was first released in August 2008, implemented on a PCI card on the storage controller. The first generation of PAM had 30 instances per card of 16 GB DRAM. Now, NetApp compact flash elements are 512 GB or 256 GB.

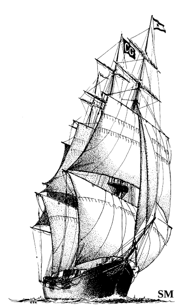
sites, this kind of tiered immediacy of response worked well. Many health care applications also can use this strategy.

NetApp prefers compact flash to SSDs because it is more cost-effective. Fronting SATA disks, flash memory can give the response time near that of disks using the more expensive Fibre Channel protocol.[6] This evolved management mode also provides an alternative to conventionally-defined tiered storage[7], which can be difficult for administrators in smaller data centers to manage.

This strategy fits well with NetApp management, which is focused on service levels. Administrators now manage policies. The tasks are automated. *My AutoSupport* provides the detailed statistics to close the information loop and provide the basis for policy tweaks. The metrics given by *My AutoSupport* clearly indicate where flash would be useful. When combined with policy-based automation, *and particularly with NetApp Storage Efficiency elements,* it makes storage capacities much more effective in generating business value at a lower cost.

Conclusion

The good habits of storage efficiency aggregate the more you use them – and the more your data is used. NetApp's storage efficiencies, together with Flash Cache, shrink the bulk and restore the performance. If you would like your data center to be less weighed down with excess baggage, consider NetApp's storage efficiencies.



[6] It also removes the hot spots referred to earlier in this paper.

[7] Conventional tiered storage involves multiple kinds of media at multiple price points. By contrast, having a single kind of disks that can be enhanced *where needed* reduces sparing and the skill set that must be maintained. This economy is not the focus of this paper but it inevitably reflects in the bottom line.

About The Clipper Group, Inc.

The Clipper Group, Inc., is an independent consulting firm specializing in acquisition decisions and strategic advice regarding complex, enterprise-class information technologies. Our team of industry professionals averages more than 25 years of real-world experience. A team of staff consultants augments our capabilities, with significant experience across a broad spectrum of applications and environments.

- ***The Clipper Group can be reached at 781-235-0085 and found on the web at www.clipper.com.***

About the Author

***Anne MacFarland* is a Senior Contributing Analyst for The Clipper Group.** Ms. MacFarland specializes in strategic business solutions offered by enterprise systems, software, and storage vendors, in trends in enterprise systems and networks, and in explaining these trends and the underlying technologies in simple business terms. She joined The Clipper Group in 2001 after a long career in library systems, business archives, consulting, research, and freelance writing. Ms. MacFarland earned a Bachelor of Arts degree from Cornell University, where she was a College Scholar, and a Masters of Library Science from Southern Connecticut State University.

- ***Reach Anne MacFarland via e-mail at Anne.MacFarland@clipper.com or at 781-235-0085 Ext. 128. (Please dial "128" when you hear the automated attendant.)***

Regarding Trademarks and Service Marks

The Clipper Group Navigator, **The Clipper Group Explorer**, **The Clipper Group Observer**, **The Clipper Group** *Captain's Log*, **The Clipper Group Voyager**, **Clipper Notes**, and *"clipper.com"* are trademarks of The Clipper Group, Inc., and the clipper ship drawings, *"Navigating Information Technology Horizons"*, and *"teraproductivity"* are service marks of The Clipper Group, Inc. The Clipper Group, Inc., reserves all rights regarding its trademarks and service marks. All other trademarks, etc., belong to their respective owners.

Disclosures

Officers and/or employees of The Clipper Group may own as individuals, directly or indirectly, shares in one or more companies discussed in this bulletin. Company policy prohibits any officer or employee from holding more than one percent of the outstanding shares of any company covered by The Clipper Group. The Clipper Group, Inc., has no such equity holdings.

After publication of a bulletin on *clipper.com*, The Clipper Group offers all vendors and users the opportunity to license its publications for a fee, since linking to Clipper's web pages, posting of Clipper documents on other's websites, and printing of hard-copy reprints is not allowed without payment of related fee(s). Less than half of our publications are licensed in this way. In addition, analysts regularly receive briefings from many vendors. Occasionally, Clipper analysts' travel and/or lodging expenses and/or conference fees have been subsidized by a vendor, in order to participate in briefings. The Clipper Group does not charge any professional fees to participate in these information-gathering events. In addition, some vendors sometime provide binders, USB drives containing presentations, and other conference-related paraphernalia to Clipper's analysts.

Regarding the Information in this Issue

The Clipper Group believes the information included in this report to be accurate. Data has been received from a variety of sources, which we believe to be reliable, including manufacturers, distributors, or users of the products discussed herein. The Clipper Group, Inc., cannot be held responsible for any consequential damages resulting from the application of information or opinions contained in this report.