The Raw Truth about Storage Capacity

A Real-World Look at the Capacity You Buy vs. the Capacity You Can Use

When you purchase capacity for your data storage system, you may expect to use the entire amount. Unfortunatly, this is not usually the case. Usable capacity is limited by a number of factors—from the common use of RAID to the practice of vendors holding back capacity for better performance or to enable certain storage system features. This paper looks at several of those factors and how they contribute to the true cost of storage.

Roger W. Kelley
Principal Architect
Xiotech Corporation

Capacity May Seem

like a fairly straightforward thing. You buy 60 terabytes, you get to use 60 terabytes. Right? Not really. Welcome to the modern world of base-10'd, base-2'd, thin-provisioned, tiered, deduped, compressed, and metadata'd storage. A world where a terabyte bought never translates into a terabyte usable.

Where the Problem Begins

To begin with, hard drive manufacturers measure drive capacity differently than do operating systems. Hard drive manufacturers use a "base-10" or decimal unit of measure, whereas operating systems use a "base-2" or binary system (Table 1).

Put in practical terms, a 300-gigabyte hard drive as measured by the manufacturer holds 300,000,000,000 bytes. But because the operating system calculates space using a base-2 system, that 300-gigabyte hard drive shows up in the operating system as 279.4 gigabytes (300,000,000,000/1,073,741,824=279.4).

So you lose nearly 7 percent of the capacity you thought you had before you've even begun!

Getting Burned by Hot Spares

Nearly all SAN vendors utilize hot spares or reserve capacity as a way of replacing failed hard drives without disrupting LUN access. These spare drives (or spare capacity) are invariably taken from the purchased capacity, further reducing the amount of storage available for data. It is not uncommon for this capacity loss to be anywhere from 1 to 8 percent or more of available raw capacity, depending on the setup and vendor requirements.

In most cases, these "hot" spare disks consume power and produce heat for the life of your storage system but do not add to system performance.

There are a few SAN vendors now offering RAID 6, which does not require hot spares but must contribute the capacity of two drives for every RAID 6 stripe set.

The Price of Safety

All SAN vendors rely on some form of RAID technology to provide data protection. The most common RAID configurations are RAID 5 and RAID 10. Most vendors use a form of these two basic types of RAID while a few offer RAID 6 or a variation, which is similar in concept to RAID 5 but offers double the parity protection.

Each RAID technology requires a certain amount of overhead often called the "RAID penalty."

- **RAID 10** is mirroring/striping and allows 50 percent of the disk to be used for data storage; the other 50 percent is used to mirror the first 50 percent.

Unit	Base 10 (Hard Drives)		Base 2 (Operating Systems)	
1 Kilobyte	10^3 bytes	1,000 bytes	2^{10} bytes	1,024 bytes
1 Megabyte	10^6 bytes	1,000,000 bytes (1,000 kilobytes)	2^{20} bytes	1,048,576 bytes (1,024 kilobytes)
1 Gigabyte	10^9 bytes	1,000,000,000 bytes (1,000 megabytes)	2^{30} bytes	1,073,741,824 bytes (1,024 megabytes)
1 Terabyte	10^{12} bytes	1,000,000,000,000 bytes (1,000 gigabytes)	2^{40} bytes	1,099,511,627,776 bytes (1,024 gigabytes)
1 Petabyte	10^{15} bytes	1,000,000,000,000,000 bytes (1,000 terabytes)	2^{50} bytes	1,125,899,906,842,624 bytes (1,024 terabytes)

Table 1. Base-10 vs. Base-2 Capacity Measurement

- **RAID 5** is striping with parity. It can withstand failure of a single drive without losing data but loses the capacity of one drive in the RAID set due to parity.

  The RAID 5 penalty is dependent on the size of the RAID set:
  - 3-drive RAID 5: 33 percent RAID penalty
  - 5-drive RAID 5: 20 percent RAID penalty
  - 9-drive RAID 5: 11 percent RAID penalty

  Each RAID 5 type has its pros and cons with regard to safety, performance, and cost. RAID 5 parity 3 and parity 5 are most commonly seen.

- **RAID 6** is striping with dual parity. It can withstand failure of up to two drives in the RAID set without losing data, but it can take a long time to rebuild should the two drives be lost. RAID 6 requires the capacity of two drives per RAID set.

  As a percentage of the total capacity, RAID 6 requires a larger RAID penalty on small RAID sets than RAID 5 (two drives' worth of capacity instead of one), giving you less capacity to store data with smaller configurations. However, this penalty percentage eases dramatically with larger RAID 6 configurations.

RAID Type	Required Disks	Capacity Penalty
RAID 5	3 Disks	1 Disk
RAID 5 Parity 3	3 Disks	33%
RAID 5 Parity 5	5 Disks	20%
RAID 10	2 Disks	50%
RAID 6	4 Disks	2 Disks

Table 2. RAID and Capacity Penalty

The Real World Is Not a Perfect World

In a perfect world how full a hard drive is should not adversely affect its overall performance. All "hot" or active data would exist next to each other on the outer edge of the hard drive platters, and all "cold" or inactive data would rest on the inner portion of the platters. Thus, filling up a hard drive to nearly 100 percent would not compromise its performance since all data being accessed would be on the fastest part of the drive and would be next to each other, reducing head travel and seek times.

Unfortunately, this type of data layout requires an intimate knowledge of your specific data access patterns. It also depends on consistency in those data access patterns. And this is only the first challenge.

The second challenge is once the data is properly positioned, how do you keep it properly positioned going forward? Theoretically, a person could be used to maintain your data layout, but this would be a poor use of your organization's IT dollars. As the saying goes: "disk is cheap; people are expensive."

Better yet, software could be and has been written to assess data access patterns and arrange the data automatically. But this doesn't fix the inconsistency of data access over time. Software that rearranges data spends an inordinate amount of time moving data, creating an enormous, process-intensive overhead that limits array performance.

Worse yet, existing software that only looks at data access times fails to assign business value to data

Short Stroking for Performance

Storage vendors often run benchmark tests against drives that are only 5 to 10 percent full (called "short stroking" the drive), which can result in impressive performance numbers.

This would be a legitimate benchmark test if customers would simply cooperate and not fill their hard drives beyond 5 or 10 percent full. Unfortunately, most customers expect to utilize as much of their capacity as possible, so these benchmark tests have little value in the real world.

blocks. Thus, **mission-critical data that is utilized infrequently is moved to slower drives and RAID sets**. When this data is needed, it is in the wrong place or on the wrong hard drives from a performance perspective.

In the real world, it often is very difficult to establish data access patterns and even more difficult to extrapolate those patterns over time. As a consequence, most users don't go beyond a fundamental exercise of initially arranging high utilization LUNs next to each other. The inability to concentrate hot or active data to a specific part of the hard drive, and easily keep it there, can result in greater head travel and higher seek times as drives get full.

Thus, a typical 15,000 rpm drive at 5 percent full will perform at about 300 I/O per second (IOPS). That same drive will perform at roughly 150 IOPS at 50 percent full and 116 IOPS at 80 percent full (Fig. 1).

SAN capacity utilization, therefore, is an exercise in weighing the cost of slower performance versus the cost of leaving storage capacity unutilized. Most SAN vendors have best practices for maximum capacity utilization—after removing the RAID penalty common to all systems. A common anecdotal maximum recommended in the industry is 75 to 80 percent after RAID. Go beyond this point and performance/cost benefit suffers unacceptably.

This area also is influenced by SAN features, such as snapshot, thin provisioning, data realignment, etc. All of these require some threshold of available capacity for your protection and system functionality. This available capacity is often called "hold back."

Instead of hovering around 75 to 80 percent utilization after RAID, some SAN vendors' best practices leave you at levels approaching an unbelievable 45 to 50 percent after RAID and hold backs. It's not that the actual space is missing, but it is held back or reserved for other uses. Where it gets difficult is that you may see the capacity (and think you have plenty of capacity available), but you are unable to use it for storage without violating best practices, generating email alerts, etc.

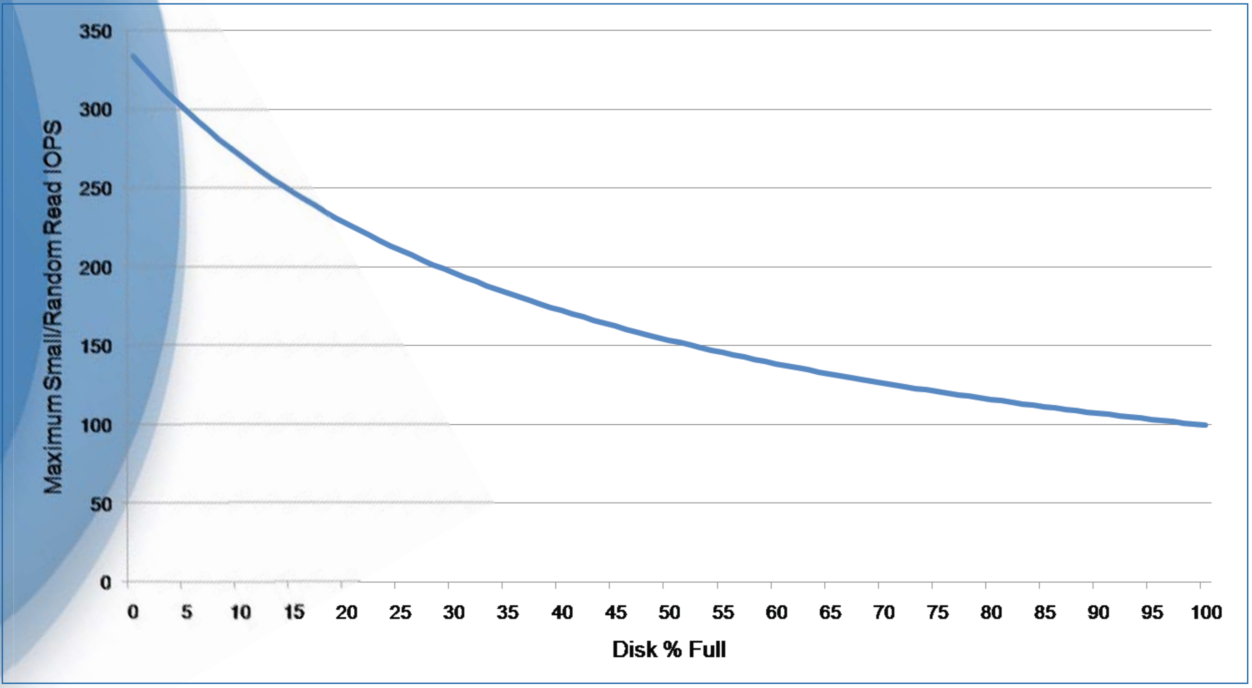

Fig. 1. Disk Full Percent vs. Maximum IOPS

How All This Affects You

To see how all these factors contribute to actual usable capacity, consider a SAN with the following characteristics:

- 60 terabytes starting raw capacity
- 200 drives: 300 gigabyte, 15,000 rpm
- 4 hot spare disks
- RAID 5 parity 5

Take the 60 terabytes of raw capacity and subtract for hot spares, base-2 operating system conversion, RAID penalty, and typical vendor hold back (Table 3).

	Spindle Capacity	Remaining Capacity
Starting Capacity	200 disks x (300GB x 10^9)	60TB x 10^{12} B or 60.00TB
Capacity After Hot Spares	196 disks x (300GB x 10^9)	58.8TB x 10^{11} B or 58.80TB
Capacity OS Sees (Base-2)	58.80TB x 10^{11}/1,073,741,824B	54,761.77GB or 53.48TB
Capacity After 20% RAID 5/5 Penalty	54,761.77TB - 20%	43,809.42GB or 42.78TB
Capacity After 20% Vendor Hold Back	43,809.42TB - 20%	35,047.54GB or 34.23TB

Table 3. Spindle Capacity vs. Usable Capacity

	System Cost/Capacity	Cost per Usable Terabyte
Starting Cost	$480,000/60.00TB	$8,000
Cost After Hot Spares	$480,000/58.80TB	$8,163
Cost After OS (Base-2)	$480,000/53.48TB	$8,976
Cost After 20% RAID 5/5 Penalty	$480,000/42.78TB	$11,220
Cost After 20% Vendor Hold Back	$480,000/34.23TB	$14,024

Table 4. Starting Cost per Terabyte vs. Cost per Usable Terabyte

You are left with only 34.23 terabytes of usable capacity, down from the 60 terabytes you bought!

How does that translate to dollars? Assume the SAN above has a "fully burdened" cost of $480,000. This includes everything connected with the array, including maintenance, switches, HBAs, and cables. Now take that fully burdened cost and divide it first by the starting capacity (60 terabytes) and then by the remaining capacity after accounting for all the reductions mentioned above (Table 4).

What started as $8,000 per terabyte ends up being more than $14,000 per terabyte!

Storage vendors who hold back capacity for feature sets or performance drive up the true cost of their capacity—in the example above, more than $2,800 of the extra cost per terabyte is due solely to vendor hold back. That adds up to a whopping $168,240 extra for a 60 terabyte SAN! Conversely, vendors who allow you to utilize more of the capacity offer a better value per dollar spent.

Is "Thin" the Answer?

Some vendors tout thin provisioning as a way to increase storage utilization value. But thin provisioning doesn't really enhance utilization efficiency since you still must provide one block of storage for one block of data. What it potentially offers is server management efficiency.

And because of the catastrophic nature of running out of actual usable capacity when using thin provisioning, those vendors who offer it require you to keep a safe distance from the "end of your actual storage." Even though it may seem you are reducing costs, it plays into the same problem that has been discussed above.

You may not have to initially purchase all the storage you think you will need for your servers up front, but the storage you do purchase will have a hold back to keep you from running out of actual space during normal (and, more importantly, abnormal) operations. This decreases maximum utilization of physical capacity and effectively drives up the cost of your purchased capacity in the same way described in this white paper.

Whereas it might have been a good idea when expanding LUNs was difficult, with modern operating systems this is no longer the case. And now, with operating systems starting to offer dynamic LUN shrink and expand, thin provisioning may be a technology whose time has come and gone.

Conclusion

As you can tell from the above exercise, how much capacity a particular SAN vendor "wastes" directly impacts the value of the system. A particular system may seem to be low cost at first, but if it provides 50 percent less usable capacity, it actually is twice as expensive as it appears to be. Another SAN that wastes less capacity may actually be a better value even though the purchase price is higher. In the above case the initial price per terabyte almost doubles when factoring actual usable capacity.

Since vendor waste varies widely in the SAN and NAS world, you should consider utilizing a formula similar to that used above to compare the best practices of competing vendors. Only then can you truly gain insight into what you are buying.

A vendor that wastes less capacity can logically provide less raw capacity up front while achieving similar storage capacity goals—resulting in lower acquisition costs as well as lower operational costs for your organization in the form of cooling, power, and rack/floor space.

It is critical in these tough economic times to make the best use of your budgetary dollars. Buying value is and always has been a fiscally prudent utilization of precious resources. When reviewing the offerings from the many storage vendors, you should look beyond the quoted price and ascertain how much of the purchased capacity will not be available for storing the data for which it was purchased.

Important Notice

By accepting, reviewing, or using this document, you, as the recipient, agree to be bound by the terms of this notice. Information in this document is subject to change without notice. Names and data used in examples are fictitious unless otherwise noted. No association with any real company, organization, product, domain name, email address, logo, person, place, or event is intended or should be inferred. Xiotech and/or its licensors may have patents or pending patent applications, trademarks, copyrights, or other intellectual property rights covering the subject matter in this document. The configuration(s) tested or described in this document may not be the only available solution(s). This document is not intended (nor may it be construed) as an endorsement of any product(s) tested, as a determination of product quality or correctness, or as assurance of compliance with any legal, regulatory, or other requirements. This document provides no warranty of any kind, including, but not limited to, any express, statutory, or implied warranties, whether this document is considered alone or in addition to any product warranty (limited warranties for Xiotech products are stated in separate documentation accompanying or relating to each product). No direct or indirect damages or any remedy of any kind shall be recoverable by the recipient or any third party for any claim or loss in any way arising or alleged to arise from or as a result of this document, whether considered alone or in addition to any other claim.

Xiotech®

6455 Flying Cloud Drive : Eden Prairie, MN 55344-3305 : 1.866.472.6764 : www.xiotech.com