Comparing Apples to Oranges to Bricks
A Pre-Sales Engineer's Guide to
Comparing Alternative Storage Proposals

EMC Proven™ Professional Knowledge Sharing 2008

Bruce Yellin
Advisory Technology Consultant
EMC Corporation
Yellin_Bruce@EMC.com

ABSTRACT

Trying to compare competitive vendor proposals for new projects or triennial enterprise storage refreshes can be a daunting task, often beset with time and budget pressures and the fear of making the wrong choice.

Perhaps because the equipment trends towards obsolescence, is lease encumbered, has insufficient capacity or is too slow (based on our last wise decision), we are once again thrust into that evil predicament that ranks with having dental work or and purchasing a new car. Each glorious multi-colored, jargon-laden, animated vendor PowerPoint session you sit through contains self-proclaimed breakthrough paradigm shifts that challenge you to conduct a solution evaluation that soon makes you feel you are comparing apples to oranges to bricks.

Some of these storage acquisitions were logical and progressed through the needs-assessment stages to product comparisons studies, and eventually financial negotiations. Some were made with a less rigorous approach, and some were based on confusion as the staff tried (vainly) to separate the wheat from the chaff.

This article tackles the issue with a common sense approach proposing what factors are important to make a decision you and your company can happily live with until the next cycle rears its ugly head. We will explore key concepts that help the customer make the right choice for their organization.

The simple truth about performance, capacity for the clueless, financial pitfalls, service imperatives, required training and many others topics will be presented by a career pre-sales engineer.

Table of Contents

Disclaimer: The views, processes or methodologies published in this compilation are those of the authors. They do not necessarily reflect EMC Corporation's views, processes, or methodologies.

I have been working in the computer industry for almost thirty-five years, twenty of those as a pre-sales engineer. In that time, I have helped over a hundred insurance, banking, brokerage, and retail companies with their hardware and software requirements. Procurement of a centralized high-end storage infrastructure was the biggest storage issue facing these customers because of its enterprise and financial impacts.

Some of these transactions were very logical and progressed through the needs-assessment stages to product comparisons studies and, eventually, financial negotiations. Others were made with less rigorous approaches and some were based on confusion as the staff tried, sometimes vainly, to separate the wheat from the chaff. This article navigates the important concepts of evaluating complex competitive storage proposals by giving you insight into your requirements and helping you to formulate a strategy for securing the best possible deal.

Years ago, a customer in the banking business told me "...the purchase of new storage equipment conjures up dreams of extensive dental work! Why must vendors use different terms, buzzwords and superfluous complexity to describe seemingly straight-forward functionality?" He explained that the use of different words for similar disk copy functions drove him insane - EMC has TimeFinder®, IBM uses FlashCopy® and HDS calls theirs ShadowImage®. With this story in mind, I set out to simplify the storage concepts you need to consider as you map them to your business needs. Along the way, make a list of what is very important to you and what isn't, and add a dose of what your staff is capable of – i.e., are they technically strong or do they need to develop some new skills.

Capacity

This is where most customers start the vendor selection journey, and if you want to follow suit, start by taking a list of servers you have *or plan to acquire* and add up their storage requirements for usable capacity. Later in this section, we will explore the concept that usable capacity goes well beyond the typical discussion of GB_{10} versus GB_2. Less sophisticated companies have sometimes given me their requirements as "...we need the best, most reliable storage solution that can support 4,350 GB and 15 servers – oh, and we don't want to pay a lot for it." Sometimes they add "usable" to that phrase because they have sat through PowerPoint presentations that made reference to raw or usable capacity. Some customers have asked for a solution with 120 x 146 GB fibre channel drives running at 15,000 RPM – yet another approach to define a solution.

Take a step back and you see that not all “capacity” is the same. Each vendor makes a capacity claim and you need to know what it means to your business. To borrow an example from the automobile industry, you are shopping for a car that must have a 15 cu. ft. trunk and the dealer, desperate to sell you a car, tells you the trunk meets that goal. What does that really mean?

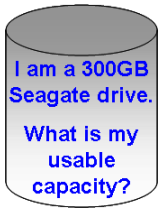
To illustrate my point, consider this anecdote. I was looking at the new 2008 Chevy Malibu a month ago and the trunk of the car looked like this photo on the left and it was listed as 15.1 cu ft. Sounded big, didn’t it? Well, I remember the tough time I had packing up my son for his college semester, so with a large Staples cardboard box in hand, I showed the salesman that the box would not fit through the opening of the trunk in the shiny new car![1]

Clearly, you can not put a 15 cu. ft. Sears refrigerator in that trunk either even though they both have the same numerical value. While you could fit 6,630 ping-pong balls in that trunk, who would really care about how many refrigerators or ping-pong balls a car can carry anyway? You should care because the same idea applies to storage! At the end of the day, it is all about data alignment and overhead, and be forewarned that misalignment and overhead can lead to wasted space, not enough space for your applications, or poor performance.

[1] I had an interesting dream about trying to get a 15 cu. ft Sears refrigerator into the 15.1 cu ft. trunk of the car – clearly comparing two numbers and concluding the refrigerator could fit into the car was crazy. Then I started to come up with something a bit more practical – how many ping-pong balls would fit into that trunk? So with perhaps too much time on my hands, the next day I calculated that I could fit over could fit 6,630 40mm ping-pong balls in that trunk. (Based on the USA Table Tennis rules, a ball is 40mm in diameter which means it measures 1.574803148 inches or a cubic foot equals 442 balls.) And speaking of funny things that could fit in the trunk of a car, in researching this white paper, I found a YouTube video of a 1.5 ton wrecking ball that does somewhat fit in the trunk of a Ford Taurus.

Buying storage because capacity numbers are aligned can be misleading. You might get close, but then again, you can easily fall short. I suggest that each application can have a different profile, require different types of data protection. The data could have alignment issues, etc., and in short, buying solely on capacity is a recipe for disaster. Another customer, obviously still confused, asked why he couldn't just take his current 25 TB environment and put it on 25 x 1 TB drives? My reply was that because "You would probably get fired!" Yes, you might be able to fit your current mixed environment of 144 x 73 GB and 146 GB drives onto 25 new giant drives, but, as I told him "...it would probably be slow as molasses!"

Completing a very demanding, yet similar project gave me the idea about this section on capacity. My customer just finalized the purchase of a multi-TB new enterprise-class storage unit without doing a proper sizing exercise. I tried many times to point out that we should all sit down and discuss this in detail and each time I was rebuked with "...we know what we are doing!" Well, the truth is they didn't know what they were doing, and in the end they bought the wrong storage and we had to redo the purchase at their expense before they could implement it.



Not all vendors use the same math when they talk about storage, and truthfully, the topic is both ambiguous and confusing! Vendors use terminology like raw, usable, engineering usable, formatted capacity and couple that with nomenclature like GB or GB_2 or GB_{10}. So is there a big difference these concepts? You bet[2]!

For those mathematically gifted, a GB can mean either 1,000,000 (8 bit) bytes or 1,073,741,824 (8 bit) bytes. Vendors use either 1024x1024x1024 or GB_2 (base 2) or 1024^3, or 1000x1000x1000 or GB_{10} (base 10) or 1000^3 bytes. Programmers and data base administrators use a GB to mean 1,073,741,824 bytes (GB_2).

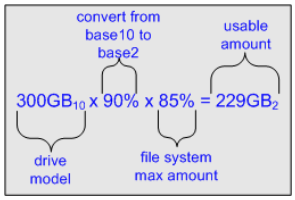
Let's discuss capacity using a real-world example of a 300 GB Seagate drive. The spec says its size is 300 GB_{10} which gives the programmer a maximum of 279.397 GB_2 (so 300x1000x1000x1000 = 279.397x1024x1024x1024 bytes) or simply 279 GB_2. I like simplicity, so what I do is take the marketing number of 300 $GB_{10,}$ multiply it by 93%, and, arrive at 279GB_2.

[2] Using the ping-pong ball example above, if we use the old International Table Tennis Federation 38MM diameter ball, we could get 17% or 1,140 more in the trunk!

STORAGE TIP: Be sure to ask how much space is needed for system functions. While it is hard to calculate how much space each vendor requires, it is clear you will be paying for it and must take it into account.

Each storage vendor needs storage for housekeeping areas for de-staging or places to store storage array logic and metadata, etc. so you could easily lose another few percent – let's assume 3%. Use my simplified 93% multiplier to convert from base 10 to base 2. This yields 90%.

Further, the old adage about never letting your file system become 100% full is still true. While this may vary from one application to another, the typical recommendation is to never let your file systems grow beyond 80-85% full or you run into the evil world of fragmentation[3] (beyond the scope of this article.) This means that you would then take 85% of the 90% number – i.e., the 300 GB_{10} drive x 90% x 85% = 229 GB_2. I like to use the simplified "3/4" math – take "3/4" of the marketing number and you arrive at the amount of space your company should use on each drive and I call this ***"useful capacity"*** (versus usable capacity.) For example, ¾ of a 300 GB_{10} drive gives you 225 GB_2. So when a vendor says that a 300 GB_{10} drive gives you 279 GB_2, they are not lying to you, but on the other hand, they are not taking into account best deployment practices either. So to tie this section up in a nice neat bow, I offer the following storage tip:



STORAGE TIP: To find out the real, useful size of a disk drive, multiply the marketing "name" by "3/4".

[3] For the best performance, you want to reduce head movement on the drive. If there is some open space here and some over there, the drive head will frantically try to fit your data into the open spaces with the result being slower than appreciated response times. You want the drive head moving to a location and then writing the data in one fell swoop. That is why you don't want fragmentation.

Manufacturer's Model	Formatted Capacity GB_{10}	Formatted Capacity GB_2	**Using the 3/4 rule, useful amount GB_2**
73GB	73	68	55
146GB	147	137	110
300GB	300	280	225
500GB	500	465	375
750GB	750	698	563
1TB	1,000	931	750

Lastly, we lose useful capacity when we use RAID protection. As we will discuss later, every pair of drives protected using RAID-1 means your application gets only one "useful" drive to write data on. Here is a useful chart that you can easily expand upon:

Manufacturer's Model	Formatted Capacity GB_{10}	Formatted Capacity GB_2	Using the 3/4 rule, useful amount GB_2	# RAID-1 drives	Useful capacity with RAID-1	# RAID 5 (3+1) drives	Useful capacity with RAID-5 (3+1)	# RAID 5 (4+1) drives	**Useful capacity with RAID-5 (4+1)**
73GB	73	68	55	2	55	4	165	5	**220**
146GB	147	137	110	2	110	4	330	5	**440**
300GB	300	280	225	2	225	4	674	5	**899**
500GB	500	465	375	2	375	4	1,124	5	**1,499**
750GB	750	698	563	2	563	4	1,688	5	**2,250**
1TB	1,000	931	750	2	750	4	2,250	5	**3,000**

So assuming a 300 GB_{10} drive and RAID-5 (4+1), we see that the useful capacity is 899 GB_2 for the 5 drives or 899/(5 physical drives) = 180 GB_2 for each drive. You can see why I emphasize that you should never ask your vendor for just "25 TB of capacity".

With all that we have discussed in this section, I would also suggest that you let your vendor know your growth projections for each server for at least the next 12 months. Put it all into a spreadsheet and discuss this with your sales engineer. With that information, your storage solution can be correctly sized so that you are ready to go.

Disk Failures, Data Protection and RAID

Most drives are mechanical with moving parts that contribute to the likelihood of component failure. But if you back up your system nightly or weekly to tape or even disk, do you really care if a drive fails? You care! The desire to guard against data loss drove the industry to invent ways to prevent it from occurring and as a result, concepts like Redundant Array of Inexpensive Disks (RAID) were developed.

Since disk drives fail, we pay a lot of attention to statistical concepts such as mean-time between failures (MTBF) and concepts like annualized failure rate (AFR). For example, the Seagate Cheetah drive has an AFR of 0.62%[4] and while that sounds high, the MTBF is an impressive 1,420,000 hours, or about 161 years! That sounds reassuring, but when you read the fairly recent study published by Google[5] involving over 100,000 hard drives, you find their disturbing information about failure rates ranging from 2.5% in the first 3 months to 8% in the first 2 years[6]. So who do you trust and what does this mean to you? The Google experience shows statistics can lie and that purchasing a 100+ drive system will insure the following will come true:

OBSERVATION: You are going to experience a drive failure in the first year, and perhaps more than one!

A disk failure may not necessarily mean a total failure – the combination of RAID and "predictive" counters for disk events are associated with terms like "seek errors", "CRC errors", "bad spot remapping", and others should come to your rescue. When the counter reaches a threshold, the system will alert the administrator that a limit is reached and the drive should be replaced.

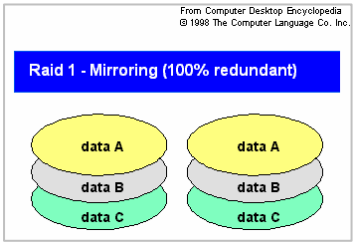
[4] Seagate Cheetah Data Sheet http://www.seagate.com/docs/pdf/datasheet/disc/ds_cheetah10k.7.pdf
[5] "Failure Trends in a Large Disk Drive Population", http://research.google.com/archive/disk_failures.pdf
[6] Equally important, they noted that "Temperature is often quoted as the most important environmental factor affecting disk drive reliability." "...after their first scan error, drives are 39 times more likely to fail within 60 days than drives with no such errors. First errors in reallocations, offline reallocations, and probational counts are also strongly correlated to higher failure probabilities."

While each drive does have some reserved space for bad block remapping, you would probably be hoping you have a "hot spare". The hot spare's job is to automatically replace a drive. After the hot spare receives all the copied data from the failing drive or from the surviving members of a parity protected group, the failing drive can be removed from the system while the system continues to run. In my years, I often enjoy hearing the story of a service engineer who arrives at a customer site with a replacement drive before the customer gets a chance to open a service call – automated call-home/email-home functions in the storage array have sent a distress "Batman Signal" and with the proper diagnostic capability, a spare part is pulled from inventory and dispatched with an engineer to proactively repair an array.

The concept of RAID[7] was conceived by Norman Ken Ouchi of IBM and documented in a 1978 patent[8]. The first time it was called RAID is in a paper titled "A Case for Redundant Arrays of Inexpensive Disks (RAID)[9]" by USC Berkeley Professors Patterson, Gibson and Katz in 1987. The trio made a case for achieving higher performance with a bunch of "cheap" disks rather than a single expensive drive. But as they realized, more drives – let's say five – meant five times the failure rate of a single drive.

Consider the time it would take to remove a bad drive and reload the data from last nights' backup tape. Meanwhile, a server - or many servers - that relied on that drive may be inoperable. You would only get the image that existed when you backed it up yesterday.

RAID is a must-have feature since most customers cannot tolerate data loss. There are three main levels of RAID protection in the market: RAID-1, RAID-5[10] and RAID-6.



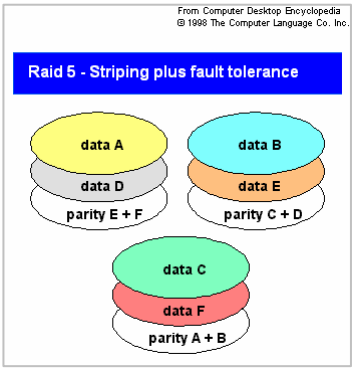
RAID-1 uses an extra drive to protect another drive; your primary storage on the "left" drive (data A in the picture) and its exact copy on the "right" drive. You will find RAID-1 used by applications requiring the best performance and highest level of protection. Because of the 1:1 relationship, this also is the most costly scheme. Some vendors,

[7] The RAID acronym has represented many phrases, including Redundant Arrays of Independent Disks, Redundant Arrays of Independent Drives, and Redundant Arrays of Inexpensive Drives
[8] http://patft.uspto.gov/netacgi/nph-Parser?Sect1=PTO1&Sect2=HITOFF&d=PALL&p=1&u=%2Fnetahtml%2FPTO%2Fsrchnum.htm&r=1&f=G&l=50&s1=4092732.PN.&OS=PN/4092732&RS=PN/4092732
[9] http://www.eecs.berkeley.edu/Pubs/TechRpts/1987/CSD-87-391.pdf
[10] There are many implementations of RAID-5 – please consult your vendor for details.

but not all, implement RAID-1 to allow an application to retrieve data from either side of the pair based on which disk head is closer to the correct data. From a performance perspective, the storage array can achieve twice as many READS as RAID-5 because RAID-1 means two identical copies. WRITEs are faster because there is no need to calculate parity across the drive pair, just a "simple" second disk WRITE. Be sure to ask your vendor how their RAID-1 implementation works.

STORAGE TIP: Use RAID-1 where the highest performance and greatest protection is needed, and you can live with a slightly higher cost (example - for 100 source disks, you need 100 additional disks for protection.) *Recommended for portions of demanding databases.*

RAID-5 takes a more rational approach with a "best of both worlds" scheme. It delivers good performance and good protection at a reasonable cost. The protection comes from spreading the parity across alternate data drives. Should a single drive in the group have problems, the parity data and surviving disks are used to reconstruct the data on a spare drive. RAID-5 is used by most applications. The nomenclature is sometimes written as RAID-5 (4+1) representing 4 data drives and 1 parity drive.



STORAGE TIP: RAID-5 provides good protection against data loss, but it requires another disk to protect a group of original disks. RAID-5 (4+1) protection for 100 source disks requires 25 additional disks. I have seen protection schemes using anywhere from a 3:1 (3+1) to a 15:1 (15+1) grouping of source-to-protection disks. *Recommended for most situations.*

With RAID-5 and large 750 GB or 1 TB drives, a drive failure means a rebuild which can take double-digit hours or even days depending on system load and other factors (there are only so many Input/Output operations per second IOPS[11] to go around). A second failure during this rebuild could be catastrophic to your business.

[11] Input/Output Operations Per Second – this is covered in more detail later in the white paper.

To reduce the risk, customers are starting to look at RAID-6[12] which can deal with two simultaneous drive failures in a group. It offers this capability through the use of a second parity drive. With RAID-6, you might have 6 data drives and 2 parity drives. So this presents an interesting economics problem. Do you go with higher density drives and tie up two for parity? Or use less dense drives and a single parity drive? Over time, RAID-6 should become an attractive choice as 1 TB+ drives become economically viable and environmentally popular.

STORAGE TIP: RAID-0 offers you no protection against data loss. *This is not recommended* unless the data is easily replaceable or is of low value.

If we use 300 GB drives, what is the approximate usable capacity and cost per GB?

RAID Definition	# of drives	Usable # of drives	Capacity based on a $300GB_{10}$ Drive	Usable Space GB_2 based on the 3/4 rule previously described	If a 300GB drive cost $656*, what is the cost per usable GB?
RAID-0	1	1	300	225	$2.92
RAID-1	2	1	300	225	$5.84
RAID-5	4	3	899	674	$3.89
RAID-5	5	4	1,199	899	$3.65
RAID-6	8	6	1,799	1,349	$3.89
RAID-6	16	12	3,597	2,698	$3.89

* NOTE: From a price survey on 1/10/08 at www.CDW.com.

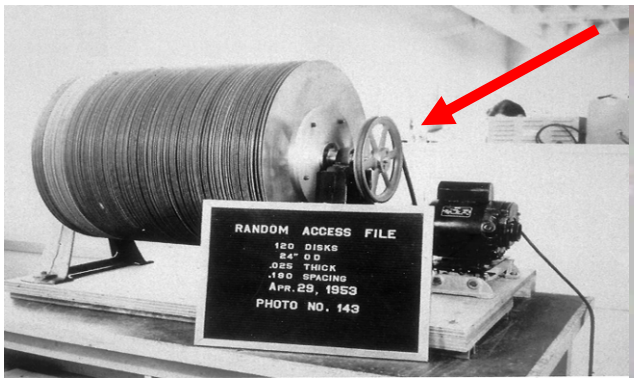
So there are really no big surprises; the cost per GB with RAID-1 is 60% higher than RAID-5 (4+1) ($5.84 vs. $3.65). RAID-6 costs about the same per GB as RAID-5. And remember, the IOPS on a 1 TB drive are about the same as a 500 GB, so while you could store twice as much as the 500 GB drive, replacing two 500 GB drives with a single 1 TB drive would give you 50% less performance.

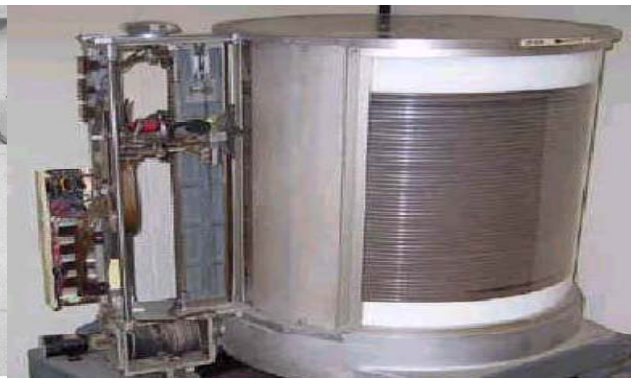
The largest drive you can currently buy is 1 TB_{10}, but vendors are working on promising ideas that could produce 3 TB_{10} drives in 2010 and 12 TB_{10} in 2014[13]. So those of you who think you could store 25 TB of company data on 25 x 1 TB_{10} drives (25 TB_{10}) and back it up to a tape drive, please raise your hand! Sorry to disappoint you – while drive density is off the charts, drive

[12] Network Appliance has a proprietary variant called RAID-DP for double parity.
[13] HDD Performance Trends http://www.gizmag.com/go/6176/

speeds and access times remain relatively flat. Drive vendors realize there is a large market for truly faster drives and may employ hybrid memory-disk concepts in future products. Sad to say, I have not read any serious reports about commercially available drives running 20,000 RPM.

The disk drive has a fascinating background. The first drive pictured on the left is called RAMAC[14] (Random Access Memory Accounting) and was built by IBM in 1956. It weighed 550 pounds and held 5 MB worth of data (a single MP3 song!) Notice the belt driven spindle mechanism on the side! On the right is a restoration of the original drive[15]. It is hard to imagine the first drive just turned 50 years old, but it is equally hard to imagine us carrying an iPod based on this technology.





While I was doing some research, I came upon an article where someone built a floppy-disk RAID-5 (4+1) with USB drives for his Apple under OS X.[16] It also appears that he came up with a similar concept, but this time he used Sony memory sticks!

[14] http://gallery.mac.com/ashoagland#100111/RAMAC_horiz_stack&bgcolor=black
[15] Computer History Museum photo archive, http://www.magneticdiskheritagecenter.org/
[16] http://ohlssonvox.8k.com/fdd_raid.htm

Top-notch Performance at the Lowest Cost

Over the years, I have found this to be one of the most difficult concepts for my customers to grasp. Unfortunately, vendors are really not in favor of test drives. But just like the magic of cooking a great meal starts with using fresh ingredients, designing a great storage solution starts with a thorough understanding of the application profile. So let's try an example or two. Assume your server runs an Oracle database and services hundreds of users. Oracle Corporation created a design profile they called the Optimal Flexible Architecture[17] (OFA) and one of its tenets is to place "write intensive" log volumes on RAID-1 protection and data tables on RAID-5. Why? As discussed previously, RAID-1 is superior when an application needs a lot of READs or WRITEs. The reason you wouldn't necessarily use RAID-1 for the entire Oracle storage layout is that it is expensive[18] and table spaces generally are not WRITE-intensive. The other OFA premise is to spread your database activity over as many hard drives as possible to allow the application to benefit from more drives and thus more IOPS.

Microsoft has a similar position and advocates the same basic strategy of using RAID-1 for key SQL and Exchange files, RAID-5 for other less intensive areas, and to spread the disk load out as wide as practical. Microsoft views the hard drive as a slow device since it has a mechanical arm that moves to retrieve data. The technical terms are seek time[19] and access time[20] and translate into key criteria for how fast a drive can respond to a READ or WRITE request. Rotations per minute (RPM) is the other key criteria and the faster the RPM, the faster the drive can respond. For example, a 1 TB drive that spins at 7,200 RPM has a significantly lower performance profile when compared to ten 100 GB 7,200 RPM drives (equals a 1 TB drive in capacity.) Hence, the endorsement to involve as many disks as possible.

STORAGE TIP: Use RAID-1 when the application vendor requires it or there is intense READ/WRITE activity. Use RAID-5 for general use and RAID-6 for backup or archive. Use 15,000 RPM drives for RAID-1, mid-speed 10,000 RPM for RAID-5 and 7,200 RPM for RAID-6.

[17] http://download.oracle.com/docs/cd/B28359_01/install.111/b32006/ofa.htm
[18] In our discussion of RAID-1 versus RAID-5, we noted the number of drives needed to store the same amount of data is much higher with RAID-1 than RAID-5.
[19] The average time it takes for the disk head to physically move to the correct place on the drive.
[20] http://en.wikipedia.org/wiki/Access_time

Microsoft has also made IOPS statements about drives used with Exchange and asserts that in certain systems, a 10,000 RPM drive could deliver 100-150 IOPS and a 15,000 RPM drive would be 30% faster with 150-200 IOPS[21]. It did not matter if the 10 K RPM drive was a 73 GB, 146 GB or a 300 GB drive – likewise for the 15 K drive. Similar studies indicate that a Serial Advanced Technology Attachment (SATA) drive runs 40-50% slower. While the numbers vary a bit, this basic concept is true for all major storage providers including EMC[22], IBM[23], HDS and NetApp as documented by Microsoft ESRP[24]. For example, in a test by NetApp on a FAS3050[25], they found that the performance of 32 x 144 GB 15 K drives was the same as 44 x 144 GB 10 K drives or 96 x 250 GB 7,200 RPM drives. The good news is that all major vendors seem to be on the Microsoft bandwagon when it comes to testing their systems with Exchange. This gives you a great comparison of system performance but remember, your applications will not have the exact same performance profile as Exchange.

So what do you do if there isn't a vendor comparison you can rely upon? How do you decode all the terminology and buzzwords and claims the storage vendor makes when describing the efficiencies of the path the application data must travel through before it reaches the disk drive? There are host bus adapters, Storage Area Networks (SANs), switching layers, storage microcode logic, high-speed cache buffers, cache hits and misses, 2 Gb or 4 Gb (gigabits), complex pre-fetch algorithms, etc. Sooner or later, the data must reside on a disk drive, and that's where the specifications quoted by disk manufacturers mathematically state that a drive has certain performance characteristics.

STORAGE TIP: Vendors tell you not to look at just drive specs because of value added cache and algorithms, etc. While your workload differs from any benchmark, you still need to compare systems. Ask for their Microsoft ESRP results. In addition, review the math behind the drives they use.

[21] "Best Practices and Technical Considerations for Deploying Microsoft Exchange 2003 on a Storage Area Network (SAN) " http://www.culminisconnections.com/sites/nyexug/Meeting%20Presentations/NYExUG_2006.10%20EMC_Exchange%20Group%20Presentation.pdf

[22] "ESRP Storage Program EMC CLARiiONCX3-80 24,000 User Storage Solution for Microsoft Exchange Server 2007", page

[23] "IBM® System Storage™ DS4800 12000 Mailbox Exchange Server 2003 Storage Solution", page 10

[24] The Exchange Solution Reviewed Program was developed by Microsoft Corporation to provide a common storage testing framework for vendors to provide information on its storage solutions for Microsoft Exchange Server software.

[25] "Comparison of Storage Protocol and Disk Drive Performance with Microsoft® Exchange Server 2003 Workloads Lee Dorrier, Network Appliance, Inc. November 2006 | TR-3514", page 12. http://www.netapp.com/library/tr/3514.pdf

Example: The specifications of a Seagate Cheetah ST373207FC[26] list it is a 73 GB 10,000 RPM drive with an average READ seek time of 4.7 ms and an average rotational latency[27] of 3.0 ms. Divide 1000 by the sum of these two numbers and you arrive at 129 READ IOPS. A WRITE calculation, as expected, takes a bit longer at 5.3 ms, so with the same rotational latency of 3.0 ms, we calculate this drive to be capable of 120 WRITE IOPS. $\frac{1000}{(5.3+3.0)} = 120 IOPS$

IOPS FORMULA: $$\frac{1000}{(avg.seek + avg.rotational\ latency)} = IOPS$$

STORAGE TIP: A 300 GB drive has roughly the capacity of four 73 GB drives, but four 73 GB drives have 4x the IOPS of a 300 GB drive. Which applications need higher IOPS drives and what is their READ/WRITE ratio?

The following is an example of a 75% READ to 25% WRITE ratio of popular Seagate drives:

Seagate Model	Drive Type	RPM Speed	Capacity GB_{10}	Price **	READ ms	WRITE ms	Avg. Rotational Latency ms	IOPS with 75/25% R/W Ratio	Price per IOP
Cheetah 10K.7 ST373207FC	FC	10K	73	$240	4.7	5.3	3.00	127	$1.89
Cheetah 10K.7 ST3146707FC	FC	10K	146	$330	4.7	5.3	3.00	127	$2.60
Cheetah 10K.7 ST3300007FC	FC	10K	300	$656	4.7	5.3	3.00	127	$5.17
Cheetah 15K.5 ST373455FC	FC	15K	73.4	$254	3.5	4.0	2.00	177	$1.44
Cheetah 15K.5 ST3146855FC	FC	15K	146.8	$366	3.5	4.0	2.00	177	$2.07
Cheetah 15K.5 ST3300655FC	FC	15K	300	$740	3.5	4.0	2.00	177	$4.18
Cheetah NS ST3400755SS	SAS Low Power	10K	400	$814	3.9	4.2	2.98	143	$5.69
Barracuda ES ST3250820NS	SATA	7.2K	250	$93	8.5	9.5	4.16	77	$1.21
Barracuda ES ST3500630NS	SATA	7.2K	500	$155	8.5	9.5	4.16	77	$2.01
Barracuda ES ST3750640NS	SATA	7.2K	750	$273	8.5	9.5	4.16	77	$3.55
Barracuda ES.2 ST3750640NS	SATA	7.2K	1,000	$392	8.5	9.5	4.16	77	$5.09

** NOTE: From a price survey on 1/10/08 at www.CDW.com. If unavailable, prices from www.NEWEGG.com are used

This table shows the fastest 15,000 RPM drive delivers 177 IOPS while the slowest 7,200 RPM SATA offers 77 IOPS (almost 60% less). If you purchased two 250 GB SATA drives @$93 each for a total of $186, you would have 21 more IOPS than a single $240 73 GB FC (Fibre Channel) drive and approximately 7x the amount of available space! But since you really care about the environment, you realize that the pair of 250 GB drives uses more electricity for the same

[26] "Cheetah 10K.7" http://www.seagate.com/docs/pdf/datasheet/disc/ds_cheetah10k.7.pdf
[27] The average time it takes for the disk platter to rotate ½ turn.

relative IOPS performance as a single drive[28]. You may want to ask your application supplier for IOPS suggestions. The question comes down to performance or capacity.

Tiers of Storage and Protection

Tiering is a hot topic among storage vendors these days. They claim that data has different corporate value throughout its lifetime, and that it should be placed on the proper tier. A tier can be based on drive performance, drive size, price point, protection level, etc. Coupled with tools for automated or manual migration between tiers, vendors claim you can achieve an optimal design.

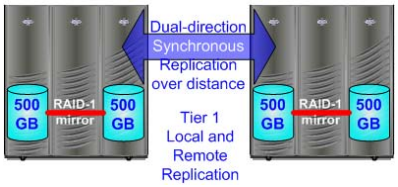
I normally see drives arranged in tiers 1-3 as summarized by chart below. Interestingly, EMC just introduced a "tier 0" solid-state disk and claim it sets a new performance record and has a superior "green" profile as well. While I have yet to have any hands-on testing with these new drives, I can appreciate the idea of no moving parts.

Drive Type	Type	Speed	Tier	Application profile
73GB	solid-state	~150-450K	0	Critical Performance
146GB	solid-state	~150-450K	0	Critical Performance
73GB	Fibre Channel	10K	1	High Performance
73GB	Fibre Channel	15K	1	High Performance
146GB	Fibre Channel	15K	1	High Performance
146GB	Fibre Channel	10K	1	General Purpose
300GB	Fibre Channel	10K	1	General Purpose
400GB	Fibre Channel	10K	1	General Purpose
500GB	Fibre Channel or SATA or SAS	7.2K	2	Non-critical
750GB	SATA or SAS	7.2K	2	Non-critical
1TB	SATA or SAS	7.2K	2	Non-critical

Tiered physical drives make sense, but remember that you need a way to seamlessly move your data from one tier to another without interrupting your application. Why? Many companies have cycles or business seasonality. For example, the retail world is very busy the last 2-3 months of the year and it makes sense to have key applications access data on the fastest tier of drives, and vice-versa. In the financial world, there are also key business cycles, so the ability to either automatically or via schedule move applications from slow to fast drives is important.

[28] The 250GB SATA drive is rated at 10.6 watts, so two would equal 21.2 watts. A 73GB 10K FC drive is rated at 11.7 watts. On a Watt/IOPS basis, the 73GB drive wins while on a Watt/GB basis, the 250GB drive wins!

As promised, we can also tier by RAID protection levels. As discussed, RAID-1 is faster with READs and WRITEs than RAID-5. You may also select tiers based on the performance profiles of Fibre Channel[29] drives, SAS or SATA drives, or possibly connection methods between the host and storage such as Fibre Channel switches[30], iSCSI[31]/LAN[32] or even NFS[33]/CIFS[34]/LAN. You should select a storage platform that gives you many or all of these options.



Lastly, data protection is another basis for creating a tier. Tier 1 protection could be both local and remote protection (e.g., EMC's SRDF, IBM's Metro Mirror, or HDS's TrueCopy), while tier 2 could be just local protection. A tier could also be created for types of local-remote replication (synchronous or asynchronous), or even frequency of backups. There are many permutations, so I suggest you settle on some basic criteria after you have a thorough understanding of your application and business requirements.

Footprints, Racks and Cabinets

While often not as critical as performance or cost, space requirements can be essential when considering a vendor's proposal if data center floor space is limited. Therefore, rack density is key and the goal is to fit the greatest capacity in the fewest number of floor tiles. While vendors offer high density drives as a way to reduce the overall storage footprint, you should balance this with our discussion on performance, drive density, and quantity.

Cabinets or racks facilitate the use of vertical space, enable consolidation, ensure proper airflow, organize wiring, provide physical security and insure serviceability – i.e., access to the front or rear should components need to be installed or replaced.

STORAGE TIP: Make sure there is enough room to open the cabinet's door when service is required – ask your vendor for more information.

[29] http://en.wikipedia.org/wiki/FATA_(hard_drive)
[30] http://en.wikipedia.org/wiki/Fibre_Channel_switch
[31] http://en.wikipedia.org/wiki/Iscsi
[32] http://en.wikipedia.org/wiki/LAN
[33] http://en.wikipedia.org/wiki/Network_File_System_%28protocol%29
[34] http://searchcio-midmarket.techtarget.com/sDefinition/0,,sid183_gci213851,00.html

STORAGE TIP: In your data center, use blue painters tape to lay out the footprint of each proposed solution and any door swing patterns. With a raised computer room floor, vendors require precise floor cutouts for cables, so make sure you mark that out with tape as well.

The cooling airflow through the system must meet the vendor's requirements – these are usually measured in cubic feet per minute (cfm). The cabinet's doors must allow for this air flow, and air going out is just as important as air coming in. It is smart to select the vendor's cabinet because you are guaranteed to meet airflow requirements.

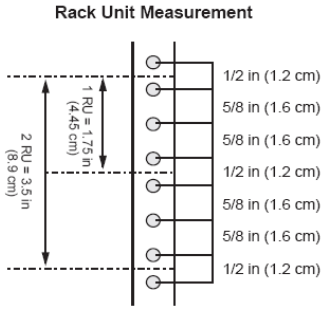
STORAGE TIP: Don't overlook ventilation - some cabinets vent from front-to-back or back-to-front or bottom-to-top. With front-to-back ventilation, make sure you stagger rows of very hot equipment so the heat coming out of one rack is not brought into another rack that needs cooler air. Some equipment vents to the sides, so if you are using your own racks for this gear, please ask for advice from your supplier.

STORAGE TIP: If you use your own cabinets, minimize floor space by using split rear perforated doors (typically 11-12 inches wide) or use easily removable doors.

Another suggestion is to run your cables through the top of the rack (cable raceway or tray or trough) rather than through the bottom (raised floor) because access to floor tiles can be tricky when things become crowded.

STORAGE TIP: Equipment may need to be serviced from the left or right side, so it is not wise to place your cabinet or rack against a wall.

Racks use a unit of measure called a “U” which stands for a “rack unit”. A single rack unit is “1U” and equals 44.5 mm or 1.75 in. It refers to the distance between mounting holes in the side rails of a rack. Manufacturers use racks that have multiple holes every 1U in height and sometimes the holes are numbered. The rails can either be “universal” with square holes fitted with cage nuts or Electronic Industries Alliance (EIA) standard rails with 10/32” tapped holes. Be sure you know which kind you need. While researching this section, I learned the now defunct Russian measurement of a vershok[35] was also 1.75 inches and dates back to Peter the Great, who ruled over 300 years ago!



COMPONENT HEIGHT: Multiply the number of U by 1.75 to get the number of inches. Example - a 42U rack = 42 x 1.75 inches = 73.5 inches tall, plus add a few inches for the thickness of the base and top of the cabinet.

STORAGE TIP: While the design has evolved, you should also know that not all EIA-310 racks are created equal - EIA specification applies to horizontal placement, it does not imply a standard depth or a total height, or even if the mounting holes are round or square!

New data centers require an architect who needs input to make sure the design is correctly sized. Too large and you pay higher real estate and cooling costs. Too small and you may sacrifice on your storage requirements or have to knock down walls to expand (hopefully not an outside wall!). You will have your data center for as long as you occupy the building, so build it with the idea of expanding it in phases. One option is to put office cubicles on the other side of an inside wall so if growth is necessary, it will be easy to move the cubicles.

[35] http://mamchenkov.net/wordpress/2004/08/11/units-of-measurement/

STORAGE TIP: With sparsely filled cabinets, be careful placing heavy equipment on slide-out rails near the top – ask about stabilizing legs to prevent the cabinet from tipping over. Rack mounting storage equipment by yourself is asking for trouble – be sure you have at least one other person to help you with the weight, or use a lifting device.

Moveable cabinets have caster wheels that accidentally could all face each other if the installation is not done properly. Your vendor should advise you if all four can be on the same tile. By the way, a raised floor tile is 2 ft x2 ft (.609 m X .609 m) and can be made from a variety of materials including wood/particle board, steel with or without a fill, concrete with a smooth surface, and aluminum. Some tiles have a high pressure laminate surface while others come with vinyl or carpeted surfaces. As you can imagine the tiles filled with concrete hold the most weight. Be sure to check with your vendor for any computer room floor tiling requirements. Some equipment can weigh more than a Volkswagen!

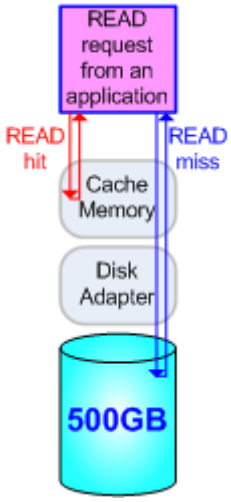
The Importance of Cache Memory

I believe the single most important factor that differentiates each vendor's high-end storage arrays from mid-tier arrays is the amount and "*intelligence*" of their cache architecture. Cache is a level removed from the physical disk and acts as an intermediary between your host looking to perform a READ or WRITE and the physical spinning drives in your storage array.

Cache memory makes your applications run faster. It takes advantage of the typical behavior of an application. As it turns out, applications frequently access data in bunches or patterns and this *locality of reference*[36] really makes cache shine! For example, if your word processor retrieves this white paper from the storage unit, you might find it occupies several adjacent disk sectors[37]. With an intelligent cache, the system will sense you are trying to access these sectors sequentially and will retrieve one or more of the subsequent sectors and place them into the cache. In that way, when you do request the next sector, it will already be in cache and access time to retrieve it could be reduced by 80% or more!

[36] http://en.wikipedia.org/wiki/Locality_of_reference
[37] http://en.wikipedia.org/wiki/Disk_sector

When an application requests data from a storage array, it believes it is communicating directly with the hard drive. In reality, cache forms a virtual, invisible layer to the host and is smart enough to know that it should provide the data to the host whenever it can – it is sometimes called a "*READ hit*". As you see from this illustration, when the "READ hit" is successful, data is returned from memory which can be as fast as 1-2 ms. When the data is not in memory, it is a "READ miss" and access can be 5-10 ms or more.



When an application writes data to an array, it can be stored in cache and later placed on the disk when time permits, thereby allowing the host to continue its operations without having to wait for a drive to rotate or heads to move. Where the physical characteristics of a drive are important is when the data must be retrieved from the drive itself (READ miss) or the cache fills up and older data must be written to the actual drive (*destaged*).

STORAGE TIP: There are many nuances to intelligent cache systems, but the most important is that your vendor sizes it correctly. Too little cache memory and your applications can run slowly, while too much and you are going to over-pay for perhaps only a small gain. Correctly caching data can dramatically improve disk-intensive applications.

Is Tape Dead (yet)?

If you think the storage business isn't exciting, then what about tape? Magnetic tape as a storage device came into vogue with the Univac I in 1951[38], long before the first disk drive was invented.

Your storage vendors may be saying that backup using tape is a dead, passé technology and suggest that backup to disk or a disk appliance using SATA drives is the best option. I believe there is some truth here because in an informal survey of many of my customers, they've told me that 80% of their restored data was backed up in the past 7 days. So these customers would

[38] http://en.wikipedia.org/wiki/Magnetic_tape

see an immediate benefit doing restores from a 7-day disk image. After that period, the number of requests seemed to drop off significantly. What is your experience?

A prudent "hybrid" strategy is to back up your data to disk and make a tape copy. Perform the "7 day" restores from disk and after that, from tape. The disk image would completely roll-over every 7 days. Clearly, this scheme is not for everyone and each case should be examined. For example, if you use your backups for archival purposes, then a case can be made to archive on near-line[39] storage versus off-line storage. I am sure there are cases in every field, such as medical images, that go against tape technology. Tape is inexpensive, even compared to 1 TB SATA drives, when you look at the complete total cost of ownership (TCO). And if you strive for a "green" environment (covered in another section), then off-line tape is very green!

Let's look at tape technology today. The promised capacities and speeds of LTO-5 and LTO-6, or DLT-S5 through DLT-S7 are amazing, as shown in this partial tape roadmap:

Tape Technology	Shipping?	Native GB	Performance MB/s	Compressed GB	Performance MB/s
LTO-4	yes	800	120	1,600	240
DLT-S4	yes	800	60	1,600	120
SAIT-2	yes	800	45	2,000	117
LTO-5	2008-2009?	1,600	180	3,200	360
LTO-6	2010-2011?	3,200	270	6,400	540
DLT-S5	unknown	1,500-1,750	100-125	3,000-3,500	200-250
DLT-S6	unknown	3,000-3,500	200-250	6,000-7,000	400-500
DLT-S7	unknown	6,000-7,000	400-500	12,000-14,000	800-1,000
SAIT-3	2008-2009?	2,000	120	5,200	312
SAIT-4	2010?	4,000	240	10,400	624

TAPE TIPS: When schedules permit, tape backup can make sense. If not, move to faster disk, disk snapshots or other disk-tape hybrid solutions.

[39] EMC coined a phrase called Content Addressable Storage and uses a separate near on-line unit called Centera.

 **TAPE TIPS: Doing many restores? Do your Help Desk representatives deal with angry callers for taking so long? Backup-to-disk helps since it uses direct access (versus sequential access). If restores are infrequent, tape may be your solution.**

 **TAPE TIPS: If you need to back up large chunks of data that are not accessed often, then you are in the sweet spot for tape.**

 **TAPE TIPS: How long do you need to keep it? Experts believe tape can last for 10-20 years, if maintained. While there aren't many 20 year old disk drives still running, what drive will you use to read a 20 year old tape?**

 **TAPE TIPS: Off-site storage? It is easy to transport tapes by truck or van, which is also the bad news. How many articles have you read about sensitive tape data that went missing? Use tape encryption!**

Going Green

Green is today's "in" color. If you believe what scientists have told us about greenhouse gases, then you also feel an obligation to help save our planet. The storage business, perhaps for the same reasons, wants your new purchase to be green. The main focus of each vendor's green strategy is conserving power and reducing cooling requirements.

This subject is a bit too complex for it to be a win-win situation for most customers. There are tradeoffs when you want to save power and reduce BTUs[40], so let's review a few.

[40] The British thermal unit (BTU) is a measurement of heat energy.

Instead of using RAID-1, consider RAID-5. We know that RAID-0 can be a recipe for data loss, but it is a very efficient use of spinning media. RAID-1 offers a great deal of protection and great performance, but to store 73 GB of data, you need 2 x 73 GB drives and thereby twice the power and cooling as RAID-0. RAID-5 also offers protection with acceptable performance, is more efficient in storing data than RAID-1, and the amount of extra power and cooling is quite acceptable. For example, if your RAID-5 storage used 5 total drives, you would be paying to power/cool 4 data drives and 1 parity drive.

RAID Level	# of drives	# of "data" drives needing power & cooling	# of "overhead" drives needing power & cooling
RAID-0	1	1	0
RAID-1	2	1	1
RAID-5	4	3	1
RAID-5	5	4	1
RAID-6	8	6	2
RAID-6	16	12	4

Can you use higher capacity drives? The amount of power and cooling required for a 73 GB drive is roughly the same as a 146 GB or 300 GB drive of the same characteristics. As an example, if you copied 4 x 73 GB drives onto 2 x 146 GB drives, you would cut the power and cooling requirement by half. Likewise, if you could get the 4 x 73 GB drives to fit on a single 300 GB drive, power and cooling would be reduced by 75%. The following chart shows that 4 x 73 GB drives would require 56 watts and the same amount of storage using a single 300 GB drive would take only 18 watts. This would also save you (191-61) 130 BTUs.

Manufacturer's Model	Typical (watts)	# drives	Total Watts	Heat created in BTUs per kWH	Total IOPS with 75/25% R/W Ratio
73 GB 10 K	14.00	4	56	191	508
146 GB 10 K	15.50	2	31	106	254
300 GB 10 K	18.00	1	18	61	127

GOING GREEN TIP: A 1 TB drive uses less power and produces less heat than 73 GB 10K drive (11.6 W vs. 14 W). Use larger drives to save on power and cooling only if your application performs properly.

Great idea, right? Well, if you remember our discussion earlier on IOPS, the table above also clearly shows that the performance measured in theoretical Total IOPS available to the 300 GB drive is ¼ the number of IOPS available to the 4 x 73 GB set of drives. So if your application can perform well with fewer IOPS, then it may be a candidate to reside on larger capacity drives.

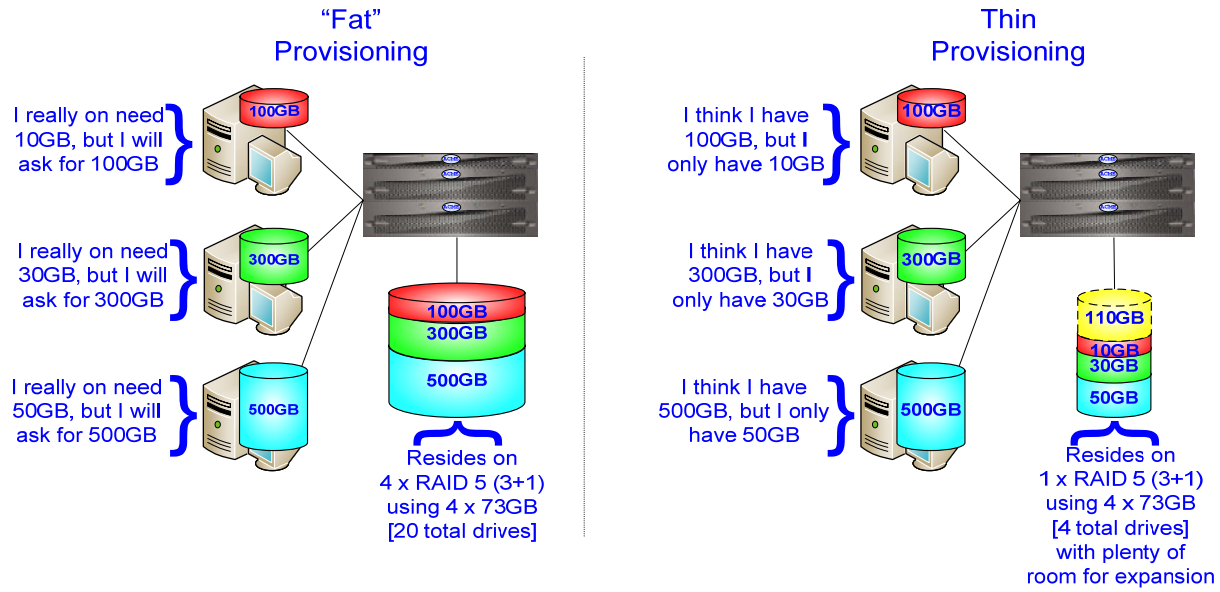
GOING GREEN TIP: You can easily over-do drive consolidation and create a poorly performing system. The best candidates for consolidation are those that are infrequently accessed. Examples: archived data, disk backups, medical and photographic images, etc.

If your application doesn't run at the same time as other applications (no IOPS conflict), then it may also be a candidate to reside on larger capacity drives. You would think of this as time shifting, or aligning applications so they do not conflict with each others disk usage.

Most higher-end storage frames allow for copies of your primary storage for use with backups, reporting, destructive testing, etc. Many vendors offer the ability to make an exact full volume copy as well as a partial snapshot. With a full volume copy, your energy profile goes up dramatically. For example, you need another 73 GB drive to make an exact copy of a 73 GB RAID-1 set. Should you want to protect that copy with RAID-1 protection, you would have used a total of 4 drives. A partial snapshot captures the changes to a volume, so you can save a great deal of space (and energy), but there can also be performance drawbacks. Ask your vendor about them.

GOING GREEN TIP: Instead of "full volume copies," snapshots or snaps can save as much as 90% of your disk space.

The relatively new concept of “thin provisioning” can save energy by increasing utilization rates. In practice, applications are allocated more physical storage then they really require (because of the bureaucracy involved in asking for more, etc.). Thin provisioning lets the application “think” it has the amount of storage originally requested but it may be given less to reduce unused capacity. The application runs fine but physical storage is kept to a minimum, resulting in substantial savings in power and cooling. In the following example, the “fat” provisioning model leads to a physical set of 20 x 73 GB drives arranged in a RAID-5 (3+1) set while the thin provisioning model results in an 80% reduction in physical drives. When an application needs more space, it is automatically given it from a reserved space area.



GOING GREEN TIP: Thin provisioning can greatly reduce the number of disk drives in your system and dramatically reduce your energy profile.

Another way to help the green movement is to continue using tape backup instead of the new of backup to disk trend. If you find the argument intriguing, I suggest that you explore the idea that a tape library requires significantly less power and cooling than a disk backup farm. Once data is written to tape and dismounted, it does not have a substantial energy profile while the same could not be said for hundreds of disks. There is a baseline energy profile for the library and there would have to be powered tape drives, but the media is off-line. A disk unit probably still draws more energy even if the drives are in a low power mode.

De-duplication stores data only once. Think about an office environment where I email a PowerPoint presentation to 30 co-workers and they each store it on central storage. Twenty-nine copies would be stored redundantly. With the file de-duplicated, the 29 users merely have a pointer to a single copy. If you need to make a change to it, such as changing the name from Bruce Yellin to yours, then a unique copy is created. Some sophisticated de-duplication schemes go beyond this by just saving the change. In our example this would require only a few additional bytes.

GOING GREEN TIP: Your system will need fewer drives and consume less energy if you use de-duplication strategies.

Another interesting idea to save energy, championed by Copan Systems, is Massive Array of Idle Disks (MAID). You store data on a MAID system that is not constantly being accessed, or accessed during set periods of time. SATA drives are powered down when not being accessed (similar to a laptop's hard drive when running on battery power). The tradeoff involves added latency when the data is first requested because the drive must spin-up; this could take seconds rather than milliseconds for a typical hard drive. MAID could be a good fit with backup-to-disk tasks or with applications storing static, infrequently accessed data such as x-rays. Ask your storage vendor about this feature.

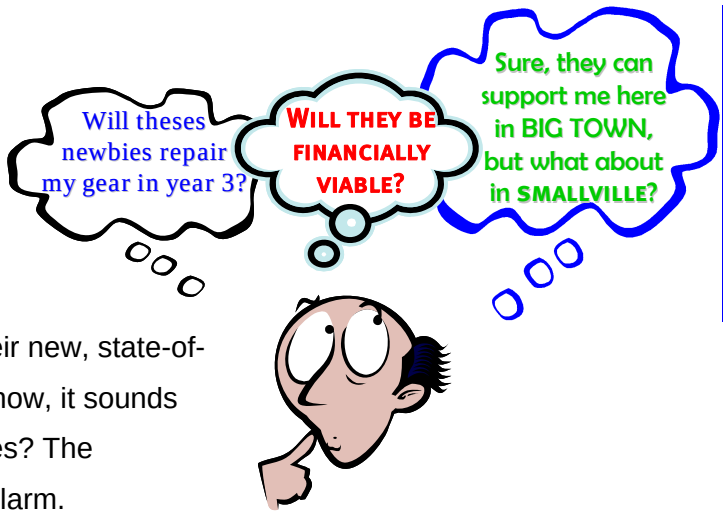
GOING GREEN TIP: Look into MAID technology if a lot of disk data is not accessed for long periods of time.

Can you use SATA or SAS drives instead of Fibre-Channel drives? On average, SATA drives consume a lot less power. From the chart below, a Seagate FC 300 GB 10 K drive consumes 18 watts and a Seagate SAS LP 400 GB 10 K drives consumes 12.1 watts. Again, you want to consider the tradeoffs in IOPS. SATA and SAS are very common in small-medium business, and in the enterprise space, it is often found in Tier 3 environments (although with the right application mix, it could just as easily be found in more mission critical applications.)

Seagate Model	Drive Type & Size GB_{10}	Typical (watts)	# drives (3/4 of manuf amount)	# disk trays @15 drives per tray	totals KW (drives + trays)	electric cost per year to run the drives	Heat created in BTUs per kWH	energy to cool the BTUs created per kWH	annual cost to cool the heat	electric cost for 3 years to power and cool the drives	energy costs to store with RAID-0 for 3 years	energy costs to store with RAID-1 for 3 years	energy costs to store with RAID-5 (4+1) for 3 years
Cheetah 10K.7 ST373207FC	FC 73 10K	14.0	457	31	9.5	$8,316	32.4	3.1	$2,744	$33,179	$33,179	$66,358	$41,474
Cheetah 10K.7 ST3146707FC	FC 146 10K	15.5	228	16	5.1	$4,502	17.5	1.7	$1,486	$17,961	$17,961	$35,923	$22,452
Cheetah 10K.7 ST3300007FC	FC 300 10K	18.0	111	8	2.8	$2,453	9.6	0.9	$809	$9,787	$9,787	$19,573	$12,233
Cheetah 15K.5 ST373455FC	FC 73 15K	14.7	457	31	9.8	$8,596	33.5	3.2	$2,837	$34,296	$34,296	$68,593	$42,871
Cheetah 15K.5 ST3146855FC	FC 146 15K	16.5	228	16	5.4	$4,702	18.3	1.8	$1,552	$18,759	$18,759	$37,519	$23,449
Cheetah 15K.5 ST3300655FC	FC 300 15K	18.8	111	8	2.9	$2,531	9.9	1.0	$835	$10,097	$10,097	$20,195	$12,622
Cheetah NS ST3400755SS	SAS LP 400 10K	12.1	83	6	1.6	$1,409	5.5	0.5	$465	$5,622	$5,622	$11,243	$7,027
Barracuda ES ST3250820NS	SATA 250 7.2K	13.0	133	9	2.6	$2,307	9.0	0.9	$761	$9,204	$9,204	$18,408	$11,505
Barracuda ES ST3500630NS	SATA 500 7.2K	13.0	67	5	1.4	$1,197	4.7	0.5	$395	$4,777	$4,777	$9,554	$5,971
Barracuda ES ST3750640NS	SATA 750 7.2K	13.0	44	3	0.9	$769	3.0	0.3	$254	$3,068	$3,068	$6,136	$3,835
Barracuda ES.2 ST3750640NS	SATA 1,000	11.6	33	3	0.7	$602	2.3	0.2	$199	$2,400	$2,400	$4,800	$3,000

The table above shows the electricity consumed to power the device and cool the heat generated by it. With assumptions such as needing 25 TB of usable capacity (our ¾ of raw capacity calculation discussed earlier) and drive trays (required for power and logic) requiring 100 watts and holding 15 drives, the table reveals some startling trends.

RAID-5 (4+1) uses 2/3 the energy of RAID-1. The cost savings for 25 TB on 73 GB 10 K drives amounts to $24,884 ($66,358 - $41,474) over 3 years. If a RAID-5 (4+1) design used 500 GB 7200 RPM drives instead of a 146 GB 10 K drives, a savings of $16,481 ($22,452 - $5,971) is achieved. Again, this is just energy cost; there are other non-green savings due to lower initial costs. Keep in mind what we mentioned earlier; the tradeoff is energy savings and lower costs versus performance. Study your application's profile before moving to lower cost or lower performing drives.



Service Imperatives

Here comes the brand new vendor touting why their new, state-of-the-art stuff is faster, better or cheaper and, you know, it sounds great! After all, who wouldn't want all these features? The something-for-nothing syndrome is setting off an alarm.

While you don't always need to select a top vendor, you do need to balance the newbie's great price with the risk that they might not survive as a company. While we all appreciate the "Catch-22" that a newbie can't survive if customers always pick the top tier vendors, you want to remember you are trying to solve a business problem in a responsible way. My recommendation is to protect your investment by selecting a vendor who will be there to service your equipment, especially if you intend to keep it 3 or 4 years. According to IDC, as of December 6, 2007, the table to the right lists the revenue and market share of the top storage providers.

	3Q07 Revenue	Market Share
EMC	$964	21.90%
IBM	$617	14.00%
HP	$589	13.40%
Dell	$409	9.30%
Hitachi	$374	8.50%
Others	$1,450	32.90%
Total	$4,403	100.00%

Customer support is a key area for consideration. Since a great deal of vendor support is provided by human beings, the quality of the assistance you get is largely reliant on a company's investment in their workforce.

SERVICE TIP: Larger vendors tend to provide their people with better training and higher pay. Ask for references – don't you want to know about the responsiveness and attitude of service personnel?

Some newbie companies have not gone public, which raises all sorts of funding questions about how soon they will have a positive cash flow. This could have an impact of their ability to provide quality, timely service. Many years ago, when I worked in the software side of the business, I used to hear phrases from new, upstart companies (who are no longer around) like,

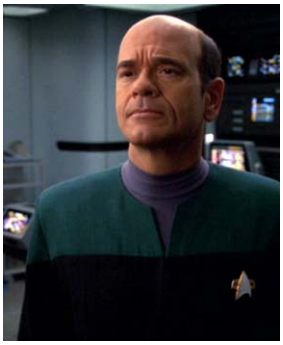
"We will put a copy of our source code in escrow so that if we do go belly-up, you, Mr. Customer, are protected." The thought of trying to deal with another company's source code when they are bankrupt is frightening. Be careful when a newbie says they will stock pile extra spare parts at your site "just in case" because I have found that service is more than a pile of parts, it requires expertise and many storage systems rely on software microcode – you are staffed to debug microcode, right? Lastly, they might not have a lot of employees, so find out if your equipment will be serviced by their personnel or outsourced to a 3rd party.

Regarding standard support concepts, proactive support is a great alternative to suffering an outage. Does the prospective storage solution include methods to predict a failure? For example, we have already learned that a disk drive rarely dies all at once; they have signatures of impending failure such as bad block remapping. Does your storage system alert you and your storage provider to this issue so parts can be replaced ***before*** they fail? They may also use heuristics to predict failure based on seemingly unrelated issues – e.g., if an application is experiencing a slowdown, it might be inferred that a fibre channel switch port was overheating, causing retries, and likely to fail in a few days or weeks.

With any support system, the key concept involves communication – how do you and your vendor find out that something is wrong? Some systems allow for paging, Internet-based notification, generated email, automated phone calls, etc. Part of the communication needs to involve diagnostic codes that allow other systems and professionals to take corrective action (before the situation gets worse). These capabilities allow for remote assistance, possible remote repair and, when necessary, dispatching local service personnel along with the correct part for a repair. Your company's help desk needs to be tied into this entire process, so be sure to involve them in this support discussion.

What level of support do you need? This can either be an enterprise-wide decision or localized if your storage system is compartmentalized – e.g., if you separate production from development, the immediacy of support may be an area where you can save money. We hear about 7x24 coverage, but vendors may also offer 5x8 coverage and concepts such as self-help, response times in hours or days, etc. Each offering will have price points and give you the chance to save money.

SERVICE TIP: If your production environment is segregated from development, the development environment may only need 5x8 coverage. Do you need 2-hour response time or will 4-hour or even next-day support be adequate? By managing your risk, you could save a lot of money.



To offer 7x24 support, your vendor will likely need to have multiple support centers around the world. Your support issue may actually travel around the world as each support center's operations correspond with daylight hours. Be sure to ask your vendor how they can access your equipment remotely in a secure manner. To me, remote support brings thoughts about the holographic doctor on Star Trek Voyager who appears in sick bay when summoned. That is what great customer service is all about!

With service inevitable, can your staff do the work or do you need support assistance? The immediate benefits of doing this yourself are a justifiable cost savings, the ability to do the work as soon as practical without having to wait for an available vendor resource. Replacing a disk drive has been simplified on many systems with software wizards walking you through the job. Ask about keeping a drive on hand rather than having to wait for a delivery service.

When a bad drive is replaced, what happens to your company's data on that drive? Perhaps your company (or clients) has a policy designed to shield them from bad publicity for losing data, perhaps not. These days, this is a big issue, so you may want to ask your vendor about data erasure services.

It is critical to avoid assigning blame when you are in the middle of multivendor complex problems. Trying to keep everyone focused on solving your business problem is your primary goal, so you want to select a vendor who behaves like a business partner and will help you quickly resolve a situation. You want your vendor to take ownership of an issue until it is clear to all parties that the ownership belongs somewhere else. For example, if an application is running slow, until you know for certain that it is a server problem or a SAN problem, etc., you might want your storage provider to take ownership of the issue as though it was theirs.

Occasional performance problems are unavoidable, yet they might not be covered under basic service agreements because they are typically not break-fix issues. Nonetheless, it is an area that could bring your company to its knees if systems start running slowly. You want to ask about the vendor's policy when it comes to system health checks and performance analysis.

SERVICE TIP: Have you tried calling your prospective vendor's support 800 #? Did you call during the day? Nights? Weekends? How long were you on hold? How was your experience? What are the guidelines for severity 1 or 2 problems?

Self-maintainable and self-healing equipment is a new trend in the storage business. All vendors have hot-spare technology to address failed hard drives while some vendors offer performance optimizers that automatically balance disk performance between drives that are over-utilized with those that are under-utilized. What do you need? Make a list of each vendor's capabilities and decide what you want to pay for.

Is there an installation cost for your new equipment? You might be able to save some money if your staff can do some of the work. What is included in the installation service? Does it include data center planning? Does it include the costs to remove any older equipment or parts?

Required Education

I have always thought it was foolish to invest $500,000 or more in equipment that houses a company's data and not invest in the people who need to interact with that equipment. You should have, at minimum, a primary individual who is fully versed in the features you purchased and a secondary person to act as back-up.

TRAINING TIP: Training not only teaches you how products work, but also helps you plan and coordinate activities. Vendors will appreciate your professionalism in being able to communicate with a common vocabulary.

Modern vendor education comes in various forms - self-paced computer-based training, remote classrooms, and instructor-led training. Many courses also include hands-on labs that walk you through various administrative tasks and other aspects of properly running and managing your equipment. The materials may also serve as the basis for “run books” that allow your organization to have repeatable, documented practices. Be sure to ask your vendor for a customized education program based on the equipment and features you are purchasing and negotiate it as part of your purchase.

Financial Leasing Pitfalls

Congratulations! You’ve once again navigated the complex world of buying new storage equipment. The jungle of vendor’s vocabulary has been cleared and you are ready to say “yes” and sign on the dotted line. But wait, you still have one bridge left to cross, and that is how to pay for it. While I won’t make you an expert on leasing, and my obvious recommendation is to involve your company’s finance person in this discussion, the old adage of the more you know the better off you are certainly applies. So let’s look at storage leasing.

If your company purchases the equipment, they are going to pay cash (or get a loan) up front and will own the gear along with (hopefully) a beneficial hardware and software warranty of 36-48 months after which maintenance fees are due.

Some companies can manage a $2M purchase, but others would rather pay $50K for 48 months. If your company decides to lease the equipment, they make payments over a period of time, but never own the equipment, ownership belongs to the lessor[41]. After the last payment, either you will return the equipment, purchase it (at a fair market price), or simply extend the lease. There are also tax and other financial benefits – some of which we will touch upon.

As I’ve said, when you purchase equipment, you own it and in the end, you likely dispose of it at your own expense either though donation, or perhaps recycling it as scrap metal, selling it to another party, etc. With leased equipment, the company that owns the “paper” (lessor) owns the equipment and at the end of the 3 or 4 year lease term is responsible for disposal.

[41] The “lessor” is the person who does the leasing. The “lessee” is the person to whom something is leased

MIGRATION CHALLENGES: Remember, whether you own it or lease it, your data is on the storage and it presents a migration challenge. This is something you need to account for either way.

- **What will you migrate to and will you need any extra equipment?**
- **How much time will it take and what will it cost?**
- **Will you need a vendor's "Professional Services" assistance?**
- **What is the risk to your business if something fails?**
- **Do you want to erase your data on the storage before returning it?**

At a high level, there are two types of leases, a capital lease and an operational lease. A capital lease[42] is in many ways the same as purchasing the equipment. The equipment's value becomes part of your company's balance sheet and is treated like both a long-term asset (it adds value to your company) and a long-term liability (you have to pay for it.) Most customers select the capital lease because the depreciation expense is included. You should call your company's finance person and chat at this point!

A capital lease includes one or more of the following phrases[43]:

- The lessor transfers ownership of the asset to the lessee.
- A bargain purchase option is given to the lessee.
- The life of the lease is greater than 75% of the economic life of the asset.
- The present value of the minimum lease payments is => 90% of the fair market value.

Based on this definition, if the lease is not a capital lease (fails all these tests), it is an operating lease. With the operating lease, you use the equipment for an "easy monthly payment" but it does not appear on your company's books as either a liability or an asset and can not be depreciated. Some customers like operating leases because it avoids hitting the books and it can result in lower payments (because the residual value of the equipment has been factored in.)

[42] The definition comes from how it is treated on a balance sheet – it is an obligation that has to be capitalized.
[43] http://en.wikipedia.org/wiki/Accounting_for_leases_in_the_United_States

A coterminous lease uses the existing end date for any new equipment you add to the lease. (A rolling lease has a new end date). You need to be careful when adding (upgrading) equipment to a coterminous lease because the full cost of the equipment needs to be spread out over the remaining term.

Here is an example. It is January, 2008 and your high-end storage lease started January, 2006 and ends January, 2009 (you are 24 months into a 36 month lease). You buy some new drives for a cost of $10,000 and want it financed conterminously with your January, 2006 lease. The new drives go on the existing 36 month lease starting in January, 2008 with a lease end date of January, 2009. Your outlay will be $10,000 (plus finance charges) for drives you will only be using for 12 months – i.e., $10,000/12 months. These new drives and the rest of the system are returned to the lessor in January, 2009. If you were to finance this upgrade with a 36-month rolling lease, you would still have the upgrade through January 2011 or long after the original base unit was returned – what value does that have? My recommendation is to discuss any substantial upgrades with your vendor and see if they will make you a better deal if you are willing to buy a newer system with the added capacity sooner than the stated lease end-date.

LEASING TIP: You need to pay attention to system upgrades when the base equipment is leased. Your company's finance department will help you decide whether to use a coterminous lease or a rolling lease.

A lease has three main components – the equipment cost, the lease factor, and the (easy) monthly payment. These equate to the following math terms:

Leasing terms:

$$CostOfEquipment = \frac{MonthlyPayment}{RateFactor}$$ **$100,000 = $3,000/.03**

$$MonthlyPayment = CostOfEquipment \times RateFactor$$ **e.g., $3,000=$100,000x.03**

$$RateFactor = \frac{MonthlyPayment}{CostOfEquipment}$$ **e.g., .03 = $3,000/$100,000**

Let's try a simple leasing example using Excel. The equipment you want lease costs $1,000,000, the interest rate is 7%, you want a 48 month lease, and the residual value is 10% (or $100,000). You could put the following into Excel and learn your easy monthly payment is $21,427.

Rate	Term	Cost	Residual Value	Easy Monthly Payment
7.00%	48	$1,000,000	$100,000	$21,427

Formula

=PMT(a2/12, b2, c2-d2, 0, 1)*-1

=PMT(Rate%/12, Term, Cost-Residual, Future Value, Payment due at the beginning of the period) *-1

At the end of a lease, you will likely have these options[44]:

- Upgrade – Payoff remaining obligation and write a new lease.
- Purchase FMV – Pay asset's Fair Market Value and own the asset.
- Return – Satisfy the lease payments and return the asset to the lessor.
- Renew – Continue to lease the asset.

What is the useful life of equipment? An easy answer is based on how long you plan to use it, after that, it has no useful life (at least not to you). Seriously, there are IRS definitions that define equipment depreciation, but your storage unit could run for 10 years. Interestingly enough, 10 years ago your brand new hard drive could have been a Seagate SCSI 7.2K holding 2 GB or 4 GB or a top-of-the-line 10 K holding 9 GB, and you would be using it with Windows 95 or the new Windows NT 4.0 operating system.

In practice, storage has a useful life of 36-48 months. You might give a little thought to using production-quality equipment in a less critical role – for example, a high-speed SAN switch with a production life of 48 months could still be useful in your development environment for another 24-36 months. You need to weigh the support costs against new equipment or pay for time and materials when it breaks. Older equipment tends to break down more frequently and could cost more to repair or require more energy than newer equipment.

[44] http://www.alanyc.org/docs/455,41,End of Lease Options

How to get the BEST $DEAL$

We tend to buy goods and services from people we like doing business with. This is especially true in the storage business since storage plays such a significant role in safeguarding corporate information. But even the nicest people have to offer you a good deal to get your business. Understanding your environment and knowing what to ask are the best places to start. No one is suggesting you need a degree in finance or negotiating skills, but just like everything else, knowledge is power. And remember, costs exist in many categories. Ask about the "total cost", a tactic I personally use when shopping for a car, "What's my drive away cost?" which means the car, finance charges, trade-in value, sales tax, dealer prep, license plates, everything!

BEST DEAL TIP: Equipment "costs" can include shipping, installation, wiring, union labor requirements, taxes, training, insurance, hardware and software support on-site, after-hour support, warranty extension, tiered costs (based on raw capacity), and others.

BEST DEAL TIP: Can you make your purchase at the end of a calendar quarter or a year? Many vendors have financial deadlines that you can leverage to negotiate higher discounts.

BEST DEAL TIP: Volume purchases - the more gear you purchase in a single transaction, the higher the discount your vendor may be able to get you. It may be easier to get discounts on software than hardware.

BEST DEAL TIP: A storage vendor may also sell other equipment such as servers and communication gear, so ask for a bundled price.

BEST DEAL TIP: Special offers can save a lot of money. Ask about:

- **Free upgrades, such as faster drives or software titles.**
- **Are there any promotions? e.g., buy 3 SAN switches, get 1 free?**
- **Trade-ins on old equipment or sweep-the-floor[45] opportunities.**
- **How about free maintenance? Annual hardware maintenance costs can run as high as 12% to 18% of the equipment cost.**

BEST DEAL TIP: Can you get a discount for pre-paying hardware and software maintenance contracts for your ownership period?

BEST DEAL TIP: How about a four-year lease? Some equipment has a longer useful life than others. A longer lease could be to your advantage.

BEST DEAL TIP: Ask about upgrade pricing. You will be using this equipment for 36-48 months, so upgrades are likely. Given hardware prices are dropping, look for a percentage-off list price rather than a cost per GB.

BEST DEAL TIP: Buy enough storage or equipment for the next 12 months. Constant upgrades can be a nuisance, could cost more, and might not lead to optimal designs. Negotiate the price of upgrades, especially if you have a coterminous lease in mind. Please see the section on leasing.

BEST DEAL TIP: Can your business run with 5x8 services instead of 7x24? Ask about service options.

[45] Many vendors find it desirable for them to be your sole supplier and as a result might offer you higher discounts if the deal removes competitive equipment from your data center.

BEST DEAL TIP: Thin provisioning can help if your end-users ask for more storage than they need. You allocate the storage that is really required from a common storage pool and purchase storage to keep the pool stocked. This keeps utilization very high with little waste.

BEST DEAL TIP: Fixing it yourself is not crazy. Can you reduce maintenance costs by having a trained staffer do simple hardware repairs? You may stock common vendor-supplied parts such as disk drives. Major repairs still require vendor customer service experts.

BEST DEAL TIP: Some vendors resell other vendors equipment. Can you get a better deal from alternate suppliers? (This is not a complete list.)

OEM Relationships	Brocade McData	Cisco	Dell	EMC	HDS	HP	IBM	NetApp	Sun
Brocade/McData			✓	✓	✓	✓	✓	✓	✓
Cisco			✓	✓	✓	✓	✓	✓	✓
Dell				✓					
EMC			✓						
HDS						✓			✓
HP					✓				
IBM									
NetApp							✓		
Sun					✓				

I have found that some customers like to have multiple vendors offering the same class of storage product. The belief is that this can give you significant leverage during negotiations by pitting one vendor against the other. If you are a very large IT shop with dozens of people on your staff and they are trained, then you may have a case for this approach. But if you are a small-to-medium sized shop, it may be a staff/training burden to support multiple vendors.

There can also be unintended consequences of multi-vendor bidding. When vendors are forced to be overly-aggressive in their bids, you may find that “best practices” (which may cost more) may go out the window and leave you with an underperforming or under-featured solution. For example, instead of using 100 x 146 GB 15 K drives to get great IOPS performance, competition may force them to use significantly cheaper 50 x 300 GB 10 K drives. The vendor who thinks they have a slim chance of getting your business may offer a “low ball” configuration, thereby forcing the vendor you would like to select to do the same thing.

Lastly, be wary of low cost bids and ask yourself “how can two offers be so far apart?” Are you getting less hardware or software, is the service and support the same, will upgrades cost a lot of money, and even ponder if this vendor will be able to sustain their business if they constantly win deals with very low prices.

Conclusion

I have outlined some ideas to support your decision to purchase new storage equipment. I suggest starting with a list of your current environment and features. A list could easily have “points” associated to these categories. If points aren’t appealing, consider the Consumer Reports approach of different colored and partially colored circles. For example, if performance is a top priority and service isn’t important (I doubt it), top performance could be worth 10 points and service could be worth 5 points, or full green circles and half circles. This would give a vendor and their proposal a score.

In the end, you need to balance the technical analysis and cost with a feel for the type of people you are doing business with. Are they truly interested in being a business partner? Since you need to buy a new system every 3-4 years, I have found there is nothing better than to be working with a vendor who has a smart pre-sales engineer who understands both your business and your personal goals. These professionals can help steer you in the right direction from their vantage point of working with dozens of other customers facing similar business decisions.