Enterprise Strategy Group | **Getting to the bigger truth.™**

Market Report

Defining Tier-1 Storage in the Modern Data Center

Tier-1 Storage Criteria in the Age of Virtualization, Clouds, & IT Flexibility

By Mark Peters

August 2011

Contents

The Purpose & Nature of this Paper

Storage in the New IT World

The purpose of this paper is to define what constitutes "tier-1" storage in the modern IT world and in the data centers and services that support it. The term seems to be implicitly understood when discussing traditional environments, but what constitutes the norm in data centers has evolved rapidly over the last few years. What was a simple environment just a few years ago with mainframes or a few large servers to be supported has evolved into a complex web of virtual machines, clouds, and expanding user expectations. These three factors demand, and also create, flexibility, but they do so in a way that pushes a lack of predictability upon the storage infrastructure.

Part of a Broader Categorization Initiative

Of course, the impacts of the structural, application, and environmental changes are not just felt in the storage infrastructure. This paper is just one of a series in which ESG will attempt to define high-level criteria, attributes, and categorization for the infrastructure required to operate a "next-generation data center" be it on premises, in the cloud, or as some hybrid implementation.[1] These categorization exercises are necessary to help delineate those vendors and technologies that are only suited for past/current data center requirements from those that have what it takes to support a true next-generation data center.

Partake in the Effort

ESG's objective is to run the overall undertaking with the support of the industry as these categorizations are certainly not specific to ESG and as such input from other professionals—users, vendors, commentators—can only be beneficial. Each document will frame the debate and, for lack of anything better, outline a proposed methodology and approach to the definitional conundrum.ESG will act as a clearing house and compositor and will update this paper (and the others as they become available) regularly. Please address comments and suggestions to the author (contact information available on the ESG website), as dialogue is encouraged to continually improve the definitions.

A Line in the Sand

Clearly, having tier-1 abilities in anything should be a function of delivering the requisite product functions that meet necessary attributes. This paper is a "line in the sand" for what those criteria should be for tier-1 storage. It does not purport to be *the* answer, although it may be! It is, however, *an* answer; a "straw horse" of definition based upon numerous industry conversations. Within each criterion, there will also be some variation of completeness of delivery and ESG has also provided guidelines for the levels of maturity within each criterion where possible: basic, moderate, and advanced. The intent is to use both the general criteria and the specific maturity/completeness as an objective gauge by which the whole industry can measure the suitability of products for the next-generation data center.

[1] Note: ESG has already worked to produce some standard criteria and attributes across the three main categories of infrastructure—server, networks, and storage—and would be pleased to discuss and work upon with any interested parties.

Tier-1 Storage in the New Data Center

The Beauty of Storage Tiers

Human beauty, so the saying goes, is in the eye of the beholder. It defies description, but we know it when we see it. Is that really so? Recent scientific investigation has attempted to deconstruct what defines beauty and there are rigorous mathematical ratios (such as the dimension of facial features, length of nose compared to width of eyebrows, etc.) that can be shown to be standard across subjects that multiple beholders perceive as beautiful. Now, of course this work and these relationships do not mean that all beautiful people look alike. But it does show that they have certain attributes that could be considered to be a "core beauty foundation" upon which their own individual facial nuances are layered.

What on earth does this have to do with tier-1 storage? The parallels are extremely pertinent.Tier-1 storage is something that people nod sagely about, and they know what it is when they are confronted by it. But, when asked for a definition, things get somewhat hazier. After all, one user's tier-1 storage system may be another's data dump depending on situations, budgets, applications, and industries. But, as with beauty, does this have to be the case? Surely there must be some absolutes for which nothing but the best will do? Is there a set of core tier-1 storage system attributes? As with the beautiful people, it would *not* mean that all tier-1 storage systems are the same, but rather it *would* mean that all such storage has certain prescribed attributes.

As we enter a new era of virtualized and "clouded" IT (with its burgeoning complexity, yet strident focus on flexibility and economic value) a term like "tier-1 storage" would be useful, provided it had actual meaning (as opposed to just a positive marketing nicety) attached to it. This is for two reasons:

1. There is a wave of recent, emerging, and brand-new storage providers (themselves built on waves of storage virtualization, scale up- and out- architectures, and [relatively] new technologies such as solid state) that might well be tier-1 storage, but which are implicitly precluded from joining the self-anointed "tier-1 club" today by assumption rather than fact.
2. It is very easy to throw around a descriptor such as "tier-1" without a definition. It is hard for tier-1 providers to push back on those who would unfairly seek their mantle.

A description allows everyone, existing and new vendors alike, to have a level semantic and technical field upon which to be measured, replacing glib assumption with specific assertion. Now, that would be a thing of beauty, wouldn't it?

The Storage "Tier-archy:" What and Why?

Describing some storage as tier-1 is by, self-definition, to acknowledge that there are at least two tiers of storage. In reality, there are more, of course. The goal of this paper is not to focus to any great extent on the merits of those various tiers or of tiering in general, but rather to define the criteria necessary for tier-1 storage in a modern data center. Some brief references are necessary as background, but that is all.

For the purpose of this report, tier-1 does not necessarily mean the highest tier that a user needs or can afford since their particular needs might be met by some other technology even if they refer to it as "tier-1." Instead, tier-1 storage is storage designed for mission-critical applications in extremely high performing, extremely highly available, and extremely well-protected data environments. Put another way, the tier-1 storage that this paper seeks to define is about **quality tiering** (which is desired or dogmatically required for an application) as opposed to **forced or economic tiering** (born of lesser needs or pragmatic choice).

New IT World = New Storage Needs

Traditionally, the top tier of storage has been very much about performance and reliability: IOPS and "five nines." Of course, these requirements do not go away in a new world of virtualization and clouds! What makes the "new tier-1" for the "new data center" such a tough challenge is that it requires vendors to combine traditional attributes with exceptional agility and efficiency. As we've moved through the various eras of IT, from mainframe, to distributed, to internet-focus, to virtualized/cloud, the storage infrastructure to support each type of IT had to change, too. Just adding capacity to a "string" of tier-1 mainframe disk in the 1980s or 1990s could require weeks of planning and multiple visits to the change management meetings. Today, that sort of thing had better happen automatically and non-disruptively. We used to employ armies of specialists who did capacity planning and implementation. There were even guidelines as to how many terabytes each individual might conceivably manage. Today, we manage petabyte systems and expect dynamic provisioning tools and automated Quality of Service (QoS) software to manage things for us. Today, predictability and order are replaced with flexibility and the unexpected, and yet somehow the storage infrastructure has to flex and adapt accordingly. Virtualization, flexible business needs, and all sorts of clouds have made for extremely active, unpredictable, and variable demands upon storage. With so many changes in terms of what's expected from a tier-1 storage system, we cannot afford to assume modern capabilities based on old criteria.

High Level Storage Needs[2]

How We Got Here

There are clearly some immediate areas within storage that need to be addressed for it to be a compliant and valuable contributor to this "new IT." In order to understand exactly why, a little history is needed.

Commercial computing took hold when one single infrastructure stack executed one specific application for one specific purpose. The original mainframe was a glorified calculator. Centralized computing was predictable and controllable, albeit expensive. But it could be managed: one processor system and one IO subsystem.

Decentralized (or distributed) computing was developed largely to try to solve the economic challenges of centralized computing (essentially CAPEX) and yielded low-cost, commodity servers which we promptly plugged into proprietary, large, expensive, monolithic storage boxes. Servers became cheaper and more interoperable while storage has remained proprietary and expensive. In the old days, the server was the thing that cost all the money. You picked your server by your OS. You picked your OS by your application. Storage was a "peripheral."

Today, servers are cheap and interoperable while storage is outlandishly expensive, complex, incompatible, and difficult. In many respects, it is the last bastion of IT awkwardness: the peripheral tail wagging the purposeful dog!

Where to Next?

Let's take for granted that we want to virtualize (which could include clouds as an element) in general because it can bring efficiencies in asset utilization, take advantage of the commoditization of hardware, leverage common infrastructures, provide seamless mobility options, etc. If we can do all this, then we are set up to drive the next (higher) level of value where we can then aspire to provide infrastructure that:

- **Self-optimizes:** boxes that tune/reconfigure themselves for the workloads that are presented and change as those requirements change.
- **Self-heals:** infrastructure deals with fault scenarios autonomously, remapping/rebuilding itself so that the application is not affected.
- **Scales dynamically:** up or down, in or out; infrastructure that extends—virtually—to whatever requirements the workload(s) presents.
- **Self-manages:** adapts to changing scenarios based on policy and enforces those policies via automation.

A Note on Finances

One thing that has not changed, and indeed which is becoming more crucial by the day because of the well known gap between the demand for IT and the resources to supply it, is the need for all storage to make financial sense. We may couch our conversation in terms of the "right data" and the "right place," but those are really symptoms;

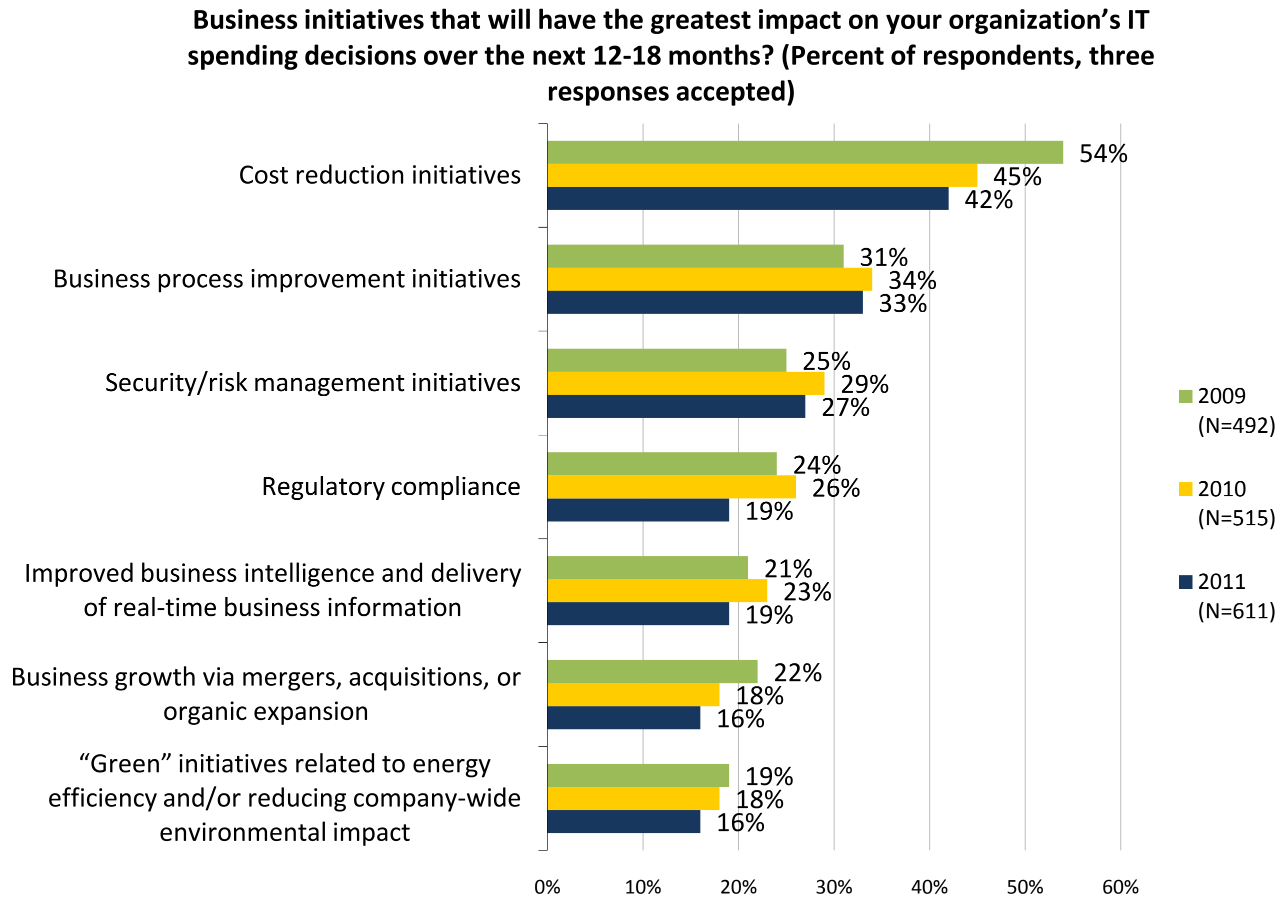
---

[2] Note: this section is an extract from the ESG Market Report, *The Future of Storage in a Virtualized Data Center*, January 2011.

the underlying cause is the simple fact that no one can afford to have everything on main memory, which is what would occur (putting things like back up, replication, and volatility to one side) if storage were free. A tier or level of storage can be built from a wide variety of storage products and the number of choices has never been greater. Choices range from ultra-high capacity, low cost, lower performance storage devices to subsystems with advanced data management functionality, scalable capacity, and very high levels of performance and data protection.

The need for financial efficiency is clear. At the same time, the needs of the "new data center" (that is, some mix of clouds, virtualization, maybe ITaaS, and definitely burgeoning needs and expectations) are for extreme flexibility and agility. Balancing the two (and, of course, doing so with high levels of reliability and availability) is the job of the IT organization and ESG research (see Figure 1) shows that the needs for efficiency (cost control) and agility (business process improvement) are indeed the top two business initiatives that impact IT spending decisions.

Figure 1. Business Initiatives That Will Impact IT Spending Decisions, Three-year Trend



Source: Enterprise Strategy Group, 2011.

New Data Center Tier-1 Storage Criteria Defined

For the purposes of brevity, we shall assume that the general concepts of tiers and tiering are understood. The trade-off decisions among price, performance, availability and so on for various applications and users have been made; data classification has been done (where it can be in today, since automated and policy-based solutions that tier based on age and access patterns are taking over in the new data center); and now it's time to evaluate which tier-1 storage to use for your mission critical needs.

As a side note, the next-generation storage infrastructure model will eventually need to be folded into the historical tiering model, or indeed supplant it. We need to take the manual labor out of managing storage. An effective policy-based classification and tiering requires a system that can accommodate the automation and orchestration of resources—you can't effectively tier with stove-piped, fixed systems of old. As we shall see, this calls for tiers that can expand dynamically, absorb and return capacity, etc., in order to be truly effective in the new world.

Here are the criteria that should exist for a tier-1 next-generation storage infrastructure designed to address the demands of a virtualized, clouded, variable world, with flexible and unpredictable storage demands. A few general notes apply to the criteria:

1. Basic (traditional) attributes such as high performance, high reliability and an easy to use management GUI are taken as givens.
2. Here, the focus is on the criteria that separate and distinguish the "new" tier-1 storage (whether block or file at the system view) from the traditional.
3. The sub-bullets shown under each criterion are the maturity levels (where applicable and where they have been defined to date).
4. There is a deliberately pedantic aspect to the terminology and, at times, a very specific delineation of functions. This is simply so that aspects that might all "hang together" automatically in one vendor's implementation are not inadvertently assumed to do the same in all products. For instance configuration and provisioning might be integrated into a set up wizard, but that is not always the case; plus it helps to highlight the distinction of the new criteria from the traditional.
5. A less technically precise, but far more colloquially appealing, description of the main impact/benefit follows each of the criteria; a kind of "Cliffs Notes" version to glean the key points.

- **Self-Configuration**
    - **This is about various levels of automated set up, and is designed to speed and simplify system adoption and growth. No more spreadsheets and whiteboards for system administrators to figure things out!**

  1. Automatic initialization.
  2. Automatic configuration of new disk/controller/cache resources.
  3. Auto load balancing and optimization of new resources.

- **Dynamic Volume/File Provisioning**
    - **Flexible provisioning (by attribute and/or over time) accelerates and simplifies one of the most common storage administration tasks. The lower maturity levels are very common while the true full dynamic ability drives enhanced optimization and is operationally very valuable.**

  1. Can configure and present basic LUNs/volumes/files.
  2. Ability to provision resources dynamically, without pre-planning, based upon policies/SLAs which can have differing performance and/or economic attributes.
  3. Ability to dynamically and non-disruptively reconfigure/reprovision storage system resources to address new service level requirements.

- **Self-optimizing and Tuning**
    - **Think of this as an engine management system that automatically helps to balance $/IOPS, $/GB and service levels in a user environment.**

1. At the basic level, all things are treated identically.
2. Automatic tiering is based on pre-determined schedules which may include some on-the-fly reaction and some meritocracy such as time-based auto-tiering.
3. Can dynamically and perpetually configure/reconfigure the storage system to optimize the use of available resources based on QoS, performance and/or economic parameters.

- **Sub-LUN Optimization (of Capacity and Other Resources)**
    - **An enhanced, more granular management technique that improves the ability to balance $/IOPS, service levels, and $/GB with the aim of delivering satisfactory service levels at the lowest cost. Multiple workloads with varying QoS levels can thereby reside on one storage system while retaining optimum efficiency.**

1. Multiple [sub] LUNs can share a single disk.
2. Multiple [sub] LUNs of varying QoS can share a single disk, reducing the need to provision and manage separate pools for each QoS.
3. Sub-LUN/volume elements are perpetually optimized at granularly, each with QoS, capabilities, and autonomous attributes derived from the master volume. In other words, multiple sub-LUN elements that have different QoS needs can share common disk resources without prescribed pools or reservations.

- **Secure Multi-tenancy**
    - **This capability is about securing the visibility to, and interaction with, data belonging to any particular user on a shared system.**

1. Basic level is just physical domains.
2. Permission-based access to [virtual] independent storage domains securely implemented on shared physical assets.
3. Dynamic support for varying QoS across virtual domains.

- **Scale-out Infrastructure**
    - **This is the ability to deploy new resources fast and as needed, and thereby to grow flexibly. A scale-out infrastructure is obviously one of the keys to supporting new, dynamic, and unpredictable capacity demands.**

1. Basic level is limited to two controllers.
2. Ability to add controllers, cache, and storage devices horizontally.
3. Ability to add performance and capacity resources independently and dynamically, and for the system to automatically absorb and reconfigure new assets as part of the resource pool in real-time.

- **Scalable Multi-tenancy**
    - **The combination of scaling and multi-tenancy provides additional benefits in terms of greater consolidation per physical asset, which can reduce hardware sprawl (and hence both CAPEX and associated OPEX) and management**

1. Basic level only supports two controllers per system
2. Ability to scale transactions or workloads (themselves being transactional and/or sequential) by adding device, controller, cache and/or IO capabilities. This can be characterized as predictable scaling.
3. Dynamically integrate new assets into the overall resource pool and redistribute/reconfigure to automatically optimize the performance/SLAs of diverse and high-service-level workloads based on QoS policies. This can be characterized as the dynamic ability to address and manage unpredictable workload changes.

- **High Availability (HA) and Planned Resilience**
    - **The benefit here is clear: it is about having tools to preclude the escalating negative impacts of prolonged downtime which is especially harmful to heavily-trafficked shared storage such as that which is found in highly virtualized and cloud environments.**

1. Failure resilience, which maintains QoS even if major system components (cache, controller, etc.) fail.
2. Multi-site (anything more than two) remote data replication is crucial to ensuring a rapid return to service should an entire array or site fail (and even tier-1 systems can do so).
3. "Planned Resilience" concerns minimizing the real operational impact of issues. A "graceful" failure and a rapid return to service are more important than the number of "9s." Whether it is a cloud or a "regular" data center operation, the idea of staying down for days is untenable.

- **Optimized Asset Utilization**
    - **The main intent and benefit here is financial impact, including TCO and ROI; better utilization means buying less hardware and also aligns with a pay-as-go model which suits many modern budget cycles well.**

1. Thin provisioning, so that real capacity is only consumed when data is actually written.
2. Ability to convert under-utilized ("fat") assets to optimized ("thin") utilization; other capacity reduction tools such as deduplication and compression can also apply.
3. Reclamation, dynamically and perpetually, releases under-utilized assets back to the overall pool.

- **Virtualized and Federated Resource Pooling**
    - **This extends the ability to pool resources and serve workloads/applications across systems and distance, enabling additional economic improvement, flexibility, and even security.**

1. Simply applies to LUNs/volumes
2. All LUNs or volumes can span all resources within the system
3. Virtual storage instances can move transparently and non-disruptively across physical systems within the broader/federated pool, and do so while retaining their QoS/policy/SLA attributes.

The goal is to achieve automated operational flexibility and scalability (a.k.a., IT agility) that is combined with the optimum use and re-use of all the resources (a.k.a., business efficiency). When used for mission-critical applications (those demanding top-notch performance and RAS), the criteria above constitute the range of elements that can be combined to produce varying levels of tier-1 storage in new world data centers.

The Bigger Truth

Often a paragon of progress in general, the IT industry can be surprisingly conservative or assumptive at times. A set of clear criteria for the prerequisite attributes of tier-1 storage in the new era of computing is a prime example. We don’t have an agreed set of such criteria even though this would help vendors and users alike. Instead, we typically simply assert “tier-1ness” (if you’re a vendor) or you’ll “know it when you see it” (if you’re an IT user).

With so much change occurring in IT and data centers—virtualization, clouds, and a necessary fixation on economic efficiency—this is neither a sensible nor a sustainable state of affairs. You would never think to measure the attributes of a car by the expectations or design specs of the mid 1970s, nor would you seek to procure wireless service for your iPhone or BlackBerry based on the dial-up Internet standards of even a decade ago. Things have changed. And yet those sorts of approaches are how we still tend to judge the uber-premium, absolutely positively, high end, tier-1 storage. We not only need criteria, we need the appropriate criteria for the emerging data center where uncertainty reigns supreme for storage demands and tier-1 storage had better be able to cope and respond so that users can meet their IT and business needs. Hopefully this paper will be cause for the industry to stop and to think. Hopefully, it provides at least some of the answer.

Enterprise Strategy Group | **Getting to the bigger truth.**

20 Asylum Street | Milford, MA 01757 | Tel:508.482.0188 Fax: 508.482.0218 | www.enterprisestrategygroup.com

4AA3-6569ENW