TANEJA GROUP
TECHNOLOGY ANALYSTS

T E C H N O L O G Y B R I E F

Storage Tiers and ATA

January, 2004

THE NET NET

Taneja Group sees many end users deploying multiple classes of storage devices, or "storage tiers" in order to increase the economies and overall utilization of their storage environment. In most of these cases, we see users leveraging lower costs ATA and serial ATA (SATA) arrays alongside more expensive fibre-channel storage fabrics. This is a trend being driven by a number of factors, including the overall availability of cost-effective disk-based storage, the rise of disk as a data protection medium, maturing practices regarding "platform appropriate" storage planning, and the new shape of exponential data growth rate in the enterprise. We believe that ATA-based storage tiers will bring significantly increased efficiency to the enterprise data center over the next 3 years. Accordingly, we need to understand the enablers of ATA and the storage, its implications and challenges, and take a look at what advances to expect in the vendor landscape in 2004.

What Are Storage Tiers?

Recently, Taneja Group conducted interviews with a select group of senior storage executives regarding their top concerns for the next 12 months. While we encountered a range of different priorities concerning compliance, storage security, consolidation and data protection, one trend that popped up horizontally across most discussions was a desire to add cost-effective ATA (and increasingly, serial ATA) disk arrays to augment many storage operations. The term of choice for this arrangement of multiple classes of storage within the same storage environment is "tiered storage" or "storage tiers."

In order to facilitate an industry-wide understanding, Taneja Group proposes the following definition of disk-based "storage tiers": *the deployment of two or more classes of disk storage within a primary or secondary storage environment, wherein the residence of data across these classes of storage is determined by business-driven variables.*

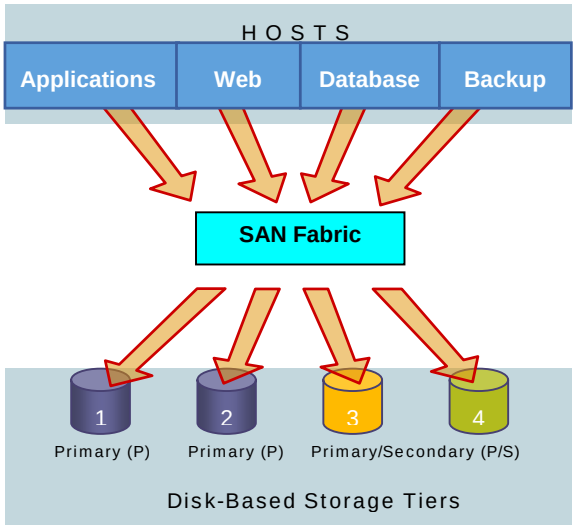
For any discussion of storage tiers, "primary storage" should be defined as stored data that directly interacts with an application environment and "secondary data" should be defined as data that provides protection to the storage environment.

Since multiple tiers of disk storage will invariably exist across primary (P) and secondary (S) environments, users have begun to enumerate these tiers as "1P, 2P, 1S, 2S, etc." Taneja Group recommends utilizing this shorthand as a common reference to explain storage tier taxonomies.

Accordingly, a model disk-based tier storage taxonomy for a SAN environment would appear as follows:

87 Elm Street, Suite 900 Hopkinton, MA 01748 Tel: 508-435-5040 Fax: 508-435-1530 www.tanejagroup.com

Example Tiered Storage Deployment



In the above conceptual diagram, we see the range of hosts, inclusive of any top-level applications, databases, web servers, and backup/restore servers, all connected to a SAN fabric. Within this SAN, the disk storage targets are classified within usage tiers numbered from 1 to 4. We define these classes as follows:

1st Tier Primary (1P): *The business critical tier*, typically the storage target for databases and associated applications at the heart of the enterprise. Accordingly, this 1st tier (1P) utilizes the most expensive, highest performing and highest availability disk in the infrastructure. In today's SAN environment, this will most likely represent a range of fibre-channel based disk arrays.

2nd Tier Primary (2P): *The business operations tier* is the focal point of tiered storage interest in 2004. This "2P" level is populated with data that may be highly utilized by the organization, but does not merit or require the level of protection and availability of mission critical data. It may contain aging data or share-intensive data and reference information. The applications associated with 2nd tier storage may include email, collaboration, some CRM/ERP applications, multimedia content and departmental file data. Representative disk types at this layer include less expensive fibre-channel disk, SCSI, and ATA. Taneja Group believes that the major interest in storage tiers will be driven by the surge of ATA into this category due to its overall price/performance advantages.

3rd and 4th Tier Primary/Secondary: *The staging and archival tiers* constitute the newest areas of penetration for disk-based storage. In terms of usage models, these two tiers still show enormous amounts of fluidity. Below, we have outlined some of the various deployment methods we have seen from end users:

- *3rd tier as Secondary, No 4th tier (1S).* In many cases, this 3rd tier is the final disk-based storage tier. Thus, it constitutes the first tier of disk-based secondary storage (1S). This tier serves as a target to the backup applications, replacing a role once held exclusively by tape libraries. In this case, tape libraries are still utilized as a final archiving media.

- *3rd tier secondary, 4th tier secondary (1S, 2S).* Based on Taneja Group surveys, 5% of customers will eliminate tape-based solutions from their environment in 2004. In those few end

user cases where this has already occurred, the 3rd tier of disk-based storage typically serves as the first backup server storage target (1S), migrating off to yet another disk-based archival solution, a 4th tier (2S), for longer term storage.

- *3rd tier primary, 4th tier secondary (3P, 1S).* We expect an increasing number of end users to deploy a 3rd tier primary (3P) environment based on an even lower price/performance storage platform than their 2nd tier. Typically, this is for data that must be available to the enterprise, but for which performance metrics are not critical. Often, this might constitute an "online" disk archival for share-intensive reference information. In such scenarios, the 4th tier then acts as the first secondary disk-based storage target (1S), replacing formerly "offline" tape archival solutions altogether.

In all cases to our knowledge, ATA-based storage is the platform of choice for 3rd and 4th tier storage tiers. The sheer capacities involved in secondary storage and archiving necessitate that the lowest-cost disk-based architectures will perpetually inhabit these low but high-growth tiers of storage.

Tier Deployment Profiles

2nd Tier Deployment Profiles

Driven by customer demand throughout 2003, most of the major storage vendors positioned ATA storage as a solution for a 2nd tier of storage. In the vast majority of cases, this deployment takes the shape of an additional ATA array deployed alongside a fibre-channel based SAN. For example, a user might augment their EMC DMX1000 with one or more Clariion CX400 arrays equipped with ATA disks. Equally prevalent, we see 2nd tiers of storage deployed with enterprise class NAS filers. This may take the shape of a Network Appliance F900 series filer augmented with that same vendor's R100 series NearStore ATA appliance.

As ATA-based solutions become increasingly viable for "heavier duty" in the enterprise, we see end users both migrating 1st tier storage to this 2nd tier based on some business-based policy (age, criticality, usage rate) and also originating data on the 2nd tier of storage for non-mission critical applications that have been historically "over-deployed" on higher-end targets. The lower cost and small deployment profile of ATA storage directly translates into a more efficient storage infrastructure.

In addition to providing overall lower availability (a typical ATA tier may provide 99.9% uptime versus 99.999% in a first tier environment), 2nd tier ATA storage deployments are typically distinguished by lower RAID protection (0,5), smaller cache sizes (from 512 to 1 GB,) and, at this stage, smaller 0verall port counts versus 1st tier storage. Further, 2nd tier storage is modular in its deployment granularity versus the monolithic large capacity deployments of the 1st tier.

3rd and 4th Tier Deployment Profiles
Taneja Group believes that by the end of 2004, the 3rd and 4th tiers of storage will become recognized as areas of massive growth due to their fluid roles in both primary and secondary disk-based storage.

In 2003, a comprehensive Taneja Group end user survey of the data protection industry revealed that fully *53% of respondents intended to deploy disk-based backup tiers in 2004*. In our interviews, we saw a particularly high interest in leveraging ATA solutions as a disk-based target for backup servers. In this scenario, the ATA array becomes a "staging tier" for some duration of time before moving data off to tape-based archival. A number of purpose-built ATA-based backup and archival solutions are now entering the storage market, promising to greatly increase the compression and restore speeds of existing ATA-based secondary storage tiers. This compression factor will be of critical importance in order to reduce the burgeoning ratio of secondary to primary storage that characterizes most enterprises. We expect ATA solutions to play a major role in increasing the capacity efficiency of secondary storage capacities over the next 36 months.

Storage Tier Market Drivers

Taneja Group projects that within 36 months, tiered storage deployments utilizing ATA and SATA will be a common fixture in over 50% of enterprise data centers, encompassing both primary and secondary storage environments. Below, we've summarized several of the key drivers that will fuel this tiered storage adoption.

Driver 1: ATA Achieves Viability
Taneja Group end user discussions and research confirm that customers have become increasingly comfortable in technical evaluations of ATA disk platforms for both primary and secondary storage. While still not a "perfect" disk-platform in terms of speed or reliability, customers generally understand its benefits to less-mission critical storage. In addition, the growing range of major vendors' hardware and software support, continually falling disk prices, and the advances in serial ATA have all come together to create a surge of perceived technical viability for ATA in the past 12 months. We believe this adoption trend will only accelerate in 2004.

The adoption of ATA-based solutions is unprecedented in the speed of its rise to enterprise-centric functionality. In the past 18 months, we have witnessed ATA technologies go from nearly non-existent external storage capacities to the multiple petabytes in deployment. Most startlingly, we see ATA deployed across primary and secondary usage models with equal prevalence. We believe this rapid and widespread, generalized industry adoption to be strongly indicative of a major ground-shift in enterprise storage media over the next 36 months.

Driver 2: Platform Economics
When we ask users why they deploy ATA storage tiers, we typically hear a strong economic rationale such as, "*it's a magnitude less expensive than my high-*

end disk." While this is true, when we dig deeper, we discover that this economic argument is tied in with an emerging data management strategy. Specifically, due to the expense of high-end storage, customers increasingly want a 2nd or 3rd tier of cost-effective disk to ensure "platform appropriate" residency for their various data types.

In fact, Taneja Group has found that many shops intend to curtail the acquisition of arrays in their 1st storage tier by aggressively migrating or originating non-mission critical data on 2nd or 3rd tiers of ATA storage. We expect this manner of "platform appropriate" capacity planning to become a standard practice throughout the enterprise over the next 24 months. Clearly, many major vendors have already anticipated this shift in capacity planning mentality and are showcasing their ATA-based storage solutions as efficiency enhancements to their flagship 1st tier storage offerings.

Driver 3: New Shape of Data Growth

Taneja Group believes that users are moving to this "platform appropriate" deployment strategy for a simple reason: Data growth rates, while perpetually high, are changing in shape such that the traditional economics of 1st tier storage capacity planning do not make sense. As indicated above, our end user research confirms that the largest portion of data growth in the enterprise now resides in reference information and non-mission critical data pools. We see a "vicious circle" of exponential data growth as more and more of this data migrates to "online" disk architectures, thereby increasing ease of access and further compounding the growth rates of disk-based storage. As such, continuing to deploy large monolithic storage platforms as "catch-all" storage tiers no longer makes economic sense. It has become clear that the non-mission critical data *must* migrate to more cost-effective platforms.

Now that feasible alternatives are available, smart customers are exploiting those storage options to their full advantage. Taneja Group believes that vendors with differentiated solutions delivered via cost-effective storage platforms will benefit greatly from this trend. In fact, EMC and Network Appliance have amply demonstrated the fundamental truth of this trend in 2003 with their respective successes selling ATA-based offerings.

Driver 4: Storage User Maturity

A fourth and critical driver in the adoption of storage tiers is increasing end user maturity with regards to the benefits of centralized storage management. For example, after a successful initial SAN adoption in a given organization, many end users are seeking to bring as much data as possible into their networked storage management framework. In 2003, Taneja Group spoke with many end users across all verticals who openly discussed a desire to extend their SAN to non-mission critical applications. "*Just because a given application does not require 'five nines' does not mean we should manage it with less integrity or granularity than our mission critical data,*" is a common theme echoed by many SAN managers. It is often

in this context that a storage tier strategy first becomes articulated by an end user.

We know that end users desire flexible, centralized tools to expand the SAN into lower tiers. Accordingly, the onus is on major SAN component and software vendors to make it feasible for users to deploy cost-effective, interoperable ATA storage with their existing storage management tools.

What's Still Needed In Tiers?

While ATA solutions have certainly opened the door to a new era for more efficient data center computing, we believe there are several major areas that require improvement in order to maximize this trend. We expect vendors to answer these needs in the coming 12 months.

Higher Utilization and Efficiency

Most customers who deploy ATA storage tiers today are not approaching the storage utilization rates that they achieved on their 1st tier of storage. The reason for this is straightforward: Despite their desire as articulated above, most users have not found a cost-effective or easy way to deploy their fabric-wide storage management tools (e.g. advanced volume management and virtualization) on their ATA storage tiers. As a result, ATA capacities are not often sliced according to the appropriate levels of granularity for the various business groups or applications. This leads to lower utilization and overall efficiency rates that are markedly lower than in the 1st tier of storage.

Most of this phenomenon can be attributed to the incremental deployment of ATA arrays as isolated "islands" in many customer shops. In 2004, we expect that leading ATA solutions will work to provide integrated ATA-based offerings that leverage existing storage management tools. This is the key to improving the overall "ROI" of the ATA storage tier investment.

Tier-Centric Software Management

A key to gaining maximal returns from ATA storage tiers is tight integration with 1st tier storage. Few customers today have achieved this level of coordination between tiers. From device discovery to partitioning, volume management and on to capacity planning, we know that the lower tiers of storage should be treated as a part of the entire storage environment in order to bring the highest possible return. Accomplishing this task is not a technological hurdle for vendors, but rather an exercise in thoughtful product integration and an awareness of real workflow issues that users face between storage tiers.

Taneja Group believes that tier-centric management requires vendors to begin asking themselves several serious questions that will influence product design: How do their customers originate their data? How do they migrate data between tiers? How do they want to leverage storage tiers in a capacity planning context? Vendors who answer these questions and translate it into value-added storage management solutions will win customer mindshare.

Platform Interoperability

From our end user discussions, we know that ATA capacities have a way of proliferating throughout the enterprise as independent, non-networked data islands. By their very nature as cost-effective, easy to deploy solutions, ATA lends itself to "fire and forget" management. As a result, strategic vendors will find legacy ATA environments are not uncommon, and any integrated storage tier opportunity will have to address end users' desires to maintain multiple vendors. Taneja Group believes that customers will absolutely demand platform interoperability and will demonstrate resistance to vendor lock-in more profoundly than they have in the past with 1st tier storage.

For this reason, we believe that vendors who can demonstrate heterogeneous fabric-based storage management in their storage tier strategy will become very attractive to many end users. Perhaps more importantly, vendors who tout their platform interoperability will be perceived by end users as having "got it" that array based competition has become, quite literally, a commodity sale.

Candera ATA Appliance and Storage Tiers

Based on both the high level of interest and the challenges that we know end users are facing in deploying tiered storage, we are very happy to see the recent product announcement from Candera Inc. regarding its new Candera ATA Appliance.

Leveraging the company's U*Star software architecture as deployed in the existing SCE510 Network Storage Controller, Candera has delivered a new ATA appliance targeted specifically at the tiered storage opportunity.

Candera Specifications

The appliance specifications should meet most customer needs: Candera will ship in active-active configurations from 4TB to 16TB of SATA RAID capacity in rack-mount form, with up to 4GB cache and 8 FC ports capable of supporting up to 48 HBA host connections.

The existing Candera offering is a fabric-resident, in-band storage controller that provides a range of heterogeneous discovery, virtualization and planning capabilities across the SAN fabric. With this new offering, Candera extends their existing storage software capabilities into the ATA storage arena with an easy to deploy appliance that promises a bridge to the wider SAN infrastructure of 1st tier storage.

In our opinion, this is an offering that should garner the attention of both customer and major vendors, coming to market with enterprise SAN-class features at a lower price point than existing ATA offerings. For example, we believe that if Candera can sell their ATA appliance at 50% to 60% of an equivalently outfitted EMC Clariion CX400, heads will definitely turn.

Candera's Differentiation

Taneja Group believes this announcement is significant because it highlights the value fabric-based software should play in the enablement of storage tiers. The following are what we believe to be the salient points of differentiation in Candera's offering and its impact on ATA storage tiers

- **Heterogeneous support**. As a heterogeneous solution deployed in the fabric, Candera brings ATA storage tiers into play with the rest of the SAN infrastructure. As indicated above, we believe this has been a principle stumbling block in early ATA storage tier adoption outside some of the major vendors' proprietary offerings. Delivering this kind of flexibility is a critical component to enabling storage tiers in larger enterprises.

- **Enterprise intelligence**. Simply purchasing an ATA array is not, *defacto*, a "storage tier". The ATA solution must be able to interact with higher storage classes using common software tools. By enabling traditional "enterprise class" storage intelligence on ATA class storage, Candera is pointing in the direction we believe the industry will follow in 2004 to extend the ATA storage tier market.

- **Easy scale and migration.** Given the pain of extreme growth rates that accompany ATA storage tiers, we believe solutions like Candera's will win because they bring easily managed deployment without disruption. Equally important, Candera allows easy capacity addition and migration of legacy data across heterogeneous storage pools. By comparison, we believe solutions that do not solve this "data island" problem will lose in competitive purchasing scenarios. Customers are away that the long term management costs are simply too high.

- **Enterprise reliability.** As we have discussed, ATA storage is fully expected to be generally less reliable than its 1st tier cousin. However, this does not mean that ATA solutions should be void of reliable service level measurement or performance visibility, as many commodity offerings are today. Candera's active-active configuration and SATA RAID disks are fully in agreement with the trend of ATA storage tiers becoming increasingly "enterprise-like" in their performance expectations.

Taneja Group Opinion

Candera appears to be one of the first storage tier players in the industry to identify several of the key factors that we have outlined in this report: end users want to integrate tiered storage with existing SAN environments, they demand interoperable platforms, and they want to extend the network intelligence they have become accustomed to in their existing SAN management experience.

To the extent that Candera and its future competitors in this space continue to emphasize these salient aspects of ATA storage enablement, we believe they will win customer attention. Vendors who

ignore these larger trends and continue to sell "vanilla" ATA arrays will find themselves commoditized as early as the second half of 2004.

Taneja Group believes that the disk storage tiering trend, while in its infancy today, will have a massive impact on storage acquisitions, management, and planning for years to come. Further, we believe that within 24 months, the current "trade-offs" associated with ATA features and reliability will no longer be an issue.

We believe that a key contributor to the advancement of ATA in storage tiers will be fabric-centric software intelligence that seamlessly integrates ATA storage tiers within the larger storage framework. Vendors that provide these solutions will garner customer attention and become recognized leaders in tiered storage enablement.