

TP-531 • From: Seagate Global Product Marketing • October 2004



# Technology Paper

## Simplifying Tiered Storage

### Introduction

Two powerful forces are shaping today's enterprise storage landscape: the pressure for greater efficiency and cost-effectiveness, and the inexorable demand for more storage capacity. Reconciling these rival imperatives falls to IT managers, who must balance upper management's fiscal goals with a stringent mandate that IT deliver seamless performance, capacity and reliability.

Conceptually, the path to more efficient storage is straightforward: Deploy multiple levels, or tiers, of storage to optimize price/performance based on the specific characteristics (quantity, required availability, and so forth) of the data. Such a tiered storage approach demands multiple device types, each cost-effectively performing its respective storage duties.

Driven by dramatic growth in areal densities, the emergence of affordable disc drives offering vast capacities and high performance was a watershed event in the evolution of enterprise storage. The low cost-per-Gbyte of desktop-class Serial ATA (SATA) drives is spurring development of a new class of nearline, or secondary, storage to serve low-I/O, high-capacity applications. Filling the price/performance gap between online (high-availability, transactional) and offline (archival, tape-centric) storage, nearline disc drives are key to making cost-effective tiered storage a reality.

But for one glaring problem: most SATA drives are designed for the desktop, not the data center. While effective in some low-end server environments with usage profiles that resemble those of desktop computers, these drives are incongruous with external storage duties in the enterprise. Lacking enterprise-class reliability and easy integration with existing enterprise infrastructures (thus imposing the expense and management complexity of a separate SATA infrastructure), such desktop drives are clearly not the tiered storage panacea some hoped they might be.

Enter the new genre of high-capacity/low-cost enterprise disc drives—engineered for the rigorous workloads of enterprise use, and purpose-built to seamlessly integrate with high-performance [Fibre Channel and Serial Attached SCSI (SAS)] drives in enterprise applications. These innovative hybrid drives enable straightforward management of tiered storage solutions, with no compromise in application performance or reliability.

### Taking Stock of Storage

In the aftermath of the dot-com implosion, and still smarting from the fiscal excesses of that era, the enterprise redoubled its efforts to keep a tight rein on storage costs. To that end, many IT managers took inventory of both the quantity and nature of their stored data—and quickly identified a significant opportunity for cost savings.

These data audits merely confirmed the inefficiency of the traditional storage model, which acknowledges only two types of data storage: online and offline. *Online* describes high-availability, transactional data that demands the performance and reliability of enterprise-class disc drives, while *offline* refers to archival data that is infrequently accessed and stored in libraries of high-capacity tapes.

But what of the vast quantity of data that falls between these two categories? Applications such as file serving, fixed-content data, disc-to-disc backup, bulk storage and short-term archiving all entail data that is below the threshold of high-availability, mission-critical storage, but must still be readily accessible to multiple users. Furthermore, regulatory compliance requirements (Sarbanes-Oxley, HIPAA) stipulate rapid retrieval of enormous quantities of financial and medical data. Now growing even more rapidly than critical online data, this *in-between* information will soon comprise the majority of enterprise data.

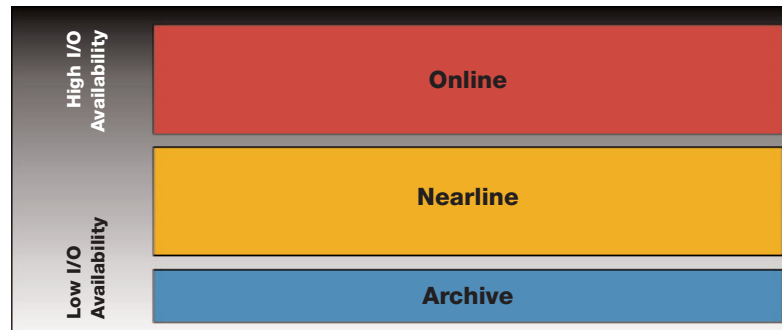
Under the simple online/offline dichotomy noted above, there were only two options for this plethora of data: online storage (optimized for performance, not capacity) with its relatively high cost-per-Gbyte; or offline storage (optimized for capacity, not accessibility) with its slow, labor-intensive data retrievals. In both cases, the net result was gross inefficiency, wasting either costly hardware or overtaxed IT staff resources. An updated storage model was long overdue....

**Tiered storage** offers the ideal solution. Complementing established online and offline storage platforms, a third tier of *nearline* storage can house the vast quantities of rarely accessed but fundamentally essential data that floods the enterprise. By melding high capacity, low cost-per-Gbyte and easy integration into established enterprise infrastructures, a new class of disc drives can fill the gap between performance and archival applications, enabling comprehensive tiered storage strategies that ensure optimal cost/performance for every type of data.

### Emergence of Tiered Storage

Born of necessity and fueled by skyrocketing areal densities, the new concept of nearline storage is quickly gaining traction in the enterprise because it holds the promise of simultaneously cutting storage costs and improving data accessibility. The burgeoning wave of high-capacity, inexpensive ATA drives can serve low-I/O, bulk data applications at significantly lower cost-per-Gbyte than online (Fibre Channel, SCSI/SAS) drives, while delivering far faster data access than offline tape libraries.

**Tiered Storage Hierarchy**



By spanning the full gamut of enterprise storage duties, the new triumvirate of online, nearline and offline solutions has made tiered storage a practical reality. To complement the efficiency of this hierarchical storage model, many enterprises are adopting Information Lifecycle Management (ILM). As the term suggests, ILM recognizes that data has a lifecycle, during which its need for accessibility and security will vary. ILM promotes optimal storage efficiency by dynamically migrating data across storage tiers as its value and relevance change. Data is thus always housed on the most appropriate, cost-effective storage device.

Initial forays into nearline storage have been met with mixed results, largely due to the unavoidable reliance on parallel ATA drives. Hamstrung by the interface's inherent limitations (poor throughput with multiple users, master/slave and termination issues), IT professionals have nevertheless managed to achieve a modicum of success with low-end servers and RAIDs equipped with ATA drives. But it wasn't until the arrival of Serial ATA (SATA) that nearline storage has taken the enterprise market by storm.

## SATA at Center Stage

With its modern serial architecture, superior throughput (1.5 Gbit/sec) and lack of master/slave headaches, SATA has quickly superseded its parallel predecessor both on the desktop and in some low-end server applications. A number of IT managers, under growing pressure to reduce costs, have contemplated the combination of performance and low cost-per-Gbyte and quickly concluded that SATA-based disc drives are ideally suited for nearline use in tiered storage solutions.

However, many experienced IT professionals have been more circumspect, troubled by factors less immediately obvious but nevertheless crucial in determining SATA's enterprise value proposition. Despite myriad technological advancements over their venerable forebearers, typical SATA disc drives remain at their core desktop-class devices. As such, they are designed for eight hour-per-day, five days-per-week deployment with light workloads, rather than the demands of enterprise-ready nearline storage.

By definition, nearline disc drives store less frequently accessed data, and thus needn't meet the demanding continuous-use standards of online drives. Nevertheless, they are expected to provide dependable availability to data when so required. To ensure this in an enterprise environment, **all nearline drives** must meet the benchmark enterprise reliability standard of 1.0 million hours MTBF.

Within this same 1.0 million hours MTBF rating, two different workload standards are actually employed, corresponding to two discrete enterprise environments. Just as SCSI/SAS and Fibre Channel address unique needs within online storage, so too do these infrastructures play distinct roles in nearline settings. SAS-based tiered storage typically supports point applications (characterized by intermittent workloads and fewer users), and thus requires a less exacting standard. FC-based tiered storage, however, warrants a more rigorous standard because it must support infrastructure applications (entailing near-continuous workloads, 24x7 availability and many users) without compromising application performance or availability. In either setting, nearline storage duties demand greater reliability than desktop-class SATA drives (600,000 hours MTBF) can provide.

With this last point in mind, would low initial expenditures on SATA drives be counterbalanced by higher rates of drive failure and the attendant costs of RAID volume rebuilds, downtime and diminished productivity?

Further, SATA requires deployment of its own infrastructure, separate and discrete from an enterprise's existing Fibre Channel or parallel SCSI environment. This entails the redundancy (and corresponding expense) of qualifying, purchasing, inventorying and maintaining a range of SATA parts/components. Existing FC shelves, cabinets and cabling cannot be reused. Moreover, such enterprise infrastructure hardware is specifically designed to meet the density requirements (for example, vibration and temperature management) of enterprise-class drives. In contrast, SATA cabinets/enclosures are not suitable for such demanding use. In addition, the SCSI-based storage management, diagnostic and application software that so effectively drives the day-to-day functionality of enterprise Fibre Channel and SCSI infrastructures would be of limited value in a SATA-based environment. Finally, how would managing an additional storage infrastructure impact IT department workloads and efficiency?

In light of such cautionary factors, the fiscal benefits of deploying an all SATA-based nearline storage could well be substantially less than anticipated. Lacking enterprise-class reliability and ignoring the enormous investment in existing enterprise storage infrastructure, a desktop SATA-based tiered storage strategy can require significant additional expense for hardware, application software and IT management. Such a scheme contradicts tiered storage's fundamental goal of increased efficiency...and demands a better alternative.

## Tighter Integration is the Key

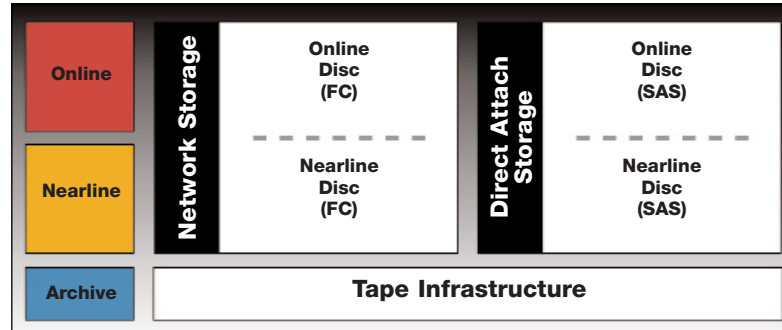
The solution is high-capacity, low-cost disc drives *purpose-built for enterprise use*. As one would expect, such drives should deliver the exceptional reliability (1.0 million hours MTBF) that enterprise duty demands. But more importantly, these nearline-optimized hybrid drives must enable tighter integration of multiple storage tiers (online, nearline and offline) by **utilizing the same interfaces** (Fibre Channel, SCSI) already ubiquitous in the enterprise. Such interface rationalization can streamline storage infrastructures, reducing costs, easing management and improving performance.

Fibre Channel has firmly cemented its position as the enterprise's premier network storage interface. Beyond their significant expenditures on FC infrastructure itself, enterprises have invested sizable quantities of money and man-hours into optimizing their SCSI-based storage management and application software solutions. Taking full advantage of these valuable assets, Fibre Channel nearline disc drives optimize tiered storage efficiency by:

- Reducing FC infrastructure total cost-per-Tbyte by optimizing storage cost/performance
- Adding nearline capacity to existing FC infrastructures with no hardware (cabinets, carriers, power supply), software (drivers, diagnostics) or management changes

- Improving infrastructure performance, availability and simplifying management with 100 percent FC data paths
- Incorporating dual ports to enable failover in high-availability environments

### Tiered Storage Device Hierarchy



Long rooted in parallel SCSI technology, Direct Attach Storage (DAS) is poised to enter the serial age with the arrival of Serial Attached SCSI (SAS). SAS retains the proven strengths of its predecessor (rock-solid reliability, a rich and mature command set) while achieving blazing throughput (3.0 Gbit/sec) and remarkable scalability (over 16,000 devices through the use of expanders). Significantly enhancing its value proposition, SAS adds compatibility with SATA to ensure unprecedented freedom to specify and consolidate the most appropriate, cost-effective storage solutions for a broad range of applications.

Fully leveraging this compatibility, enterprise-optimized SATA drives seamlessly integrate into SAS infrastructures, enabling both performance (SAS) and capacity (nearline SATA) solutions to coexist in the same SAS domain, even the same enclosure. Beyond the significant cost savings of utilizing a common infrastructure, enterprise-class SATA drives promote further efficiencies by enabling IT managers to:

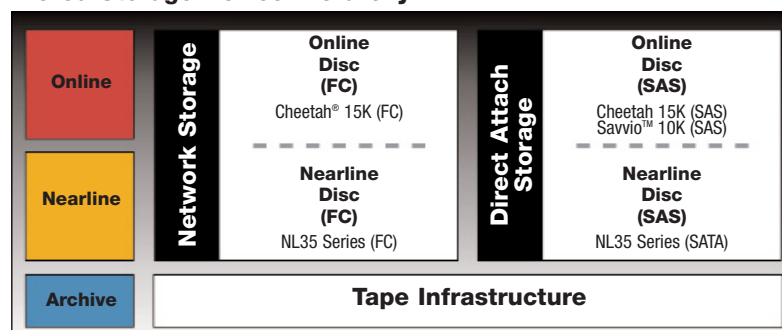
- Fine-tune storage cost/performance for a given application with the optimal blend of SAS and SATA drives
- Fully exploit SATA's cost-per-Gbyte economies without compromising enterprise management and reliability requirements
- Increase storage capacity while maintaining application requirements

As can be seen, deploying high-capacity, low-cost-per-Gbyte enterprise drives for nearline storage in both network (Fibre Channel) and direct-attach (SAS) tiered storage environments can deliver a multitude of benefits. Seagate® has created the NL35 Series disc drives to make those benefits a reality.

### Seagate: Simplifying Tiered Storage

NL35 Series is a new family of high-capacity, low-cost-per-Gbyte enterprise disc drives that fills the price/performance gap between online storage and archive storage. Purpose-built to seamlessly integrate with high-performance disc drives in enterprise applications, NL35 Series drives enable manageable tiered storage solutions that don't compromise application performance or reliability.

### Tiered Storage Device Hierarchy



It's no secret that tiered storage schemes add complexity that can negate their benefits. NL35 Series drives easily integrate with existing storage infrastructures, enabling data centers to efficiently combine and manage multiple tiers of storage to better match application needs.

NL35 Series disc drives **reduce storage infrastructure costs** by providing a high-capacity/low-cost storage option for data applications that don't require the performance of mainstream enterprise drives. Despite their cost-effective design, NL35 Series drives meet the stringent requirements of the data center, ensuring that application performance and availability are maintained.

Because nearline storage requirements can vary widely depending upon the application, Seagate (unlike other drive manufacturers) offers both **Fibre Channel** and **SATA** drives in its NL35 Series to meet such diverse needs.

### Seagate NL35 Series FC (Fibre Channel)

NL35 Series Fibre Channel disc drives increase the productivity of Fibre Channel infrastructures by simplifying deployment of multiple tiers of storage in a single SAN. These native FC drives can be directly inserted into existing FC SANs, eliminating the need for any changes to storage software or additional hardware to bridge between SATA drives and FC infrastructures.

The NL35 Series enables Fibre Channel infrastructures to seamlessly support complementary tiered-storage solutions: Cheetah 15K drives for high-availability online applications and economical NL35 Series drives for storage of less frequently accessed data.

Although highly cost-efficient, NL35 Series Fibre Channel disc drives are designed to operate around the clock under the rigorous workloads typical of networked storage applications.

Seagate NL35 Series FC (Fibre Channel) Overview	
FEATURE	BENEFIT TO ENTERPRISE
<b>400-Gbyte capacity</b>	Increases capacity density, lowers cost-per-Gbyte for nearline storage applications
<b>Dual-port, 2 Gbit/sec FC interface</b>	Integrates seamlessly in existing FC environments; no hardware or storage software changes needed
<b>Full support of FC feature set</b>	Interoperates within existing FC network architectures and management schemes
<b>1.0 million hours MTBF</b>	Operates in networked enterprise nearline environments without compromising application performance or availability
<b>Background Media Scan (BGMS)</b>	Maximizes media integrity, minimizes data errors
<b>Enhanced Error Control and Correction (IOEDC)</b>	Full-path error correction improves data integrity, reduces data error rates
<b>Rotational Vibration (RV) attenuation</b>	NL35 Series drives can be integrated in highly dense configurations, enabling multi-terabyte capacities in a single tray
<b>Advanced servo performance</b>	High servo sample rates, tighter servo control maintains data integrity in high-workload applications

### Seagate NL35 Series SATA

NL35 Series SATA disc drives leverage SAS compatibility with SATA to enable tiered storage in a single, enterprise-class SAS infrastructure. Purpose-built for the rigors of enterprise duty, these drives carry a 1.0 million hours MTBF rating and a host of enterprise-friendly features (see table below).

SAS/SATA compatibility produces unprecedented storage synergies and efficiencies, allowing enterprises to meet both performance and capacity requirements by combining SAS and SATA drives within a single SAS infrastructure.

Furthermore, deploying SATA drives in a SAS infrastructure vastly improves SATA's scalability. Hundreds (even thousands) of NL35 Series SATA drives can be deployed and managed within a single SAS domain, enabling highly scalable, cost-effective storage solutions for capacity-intensive nearline applications.

Seagate NL35 Series SATA (Serial ATA) Overview	
FEATURE	BENEFIT TO ENTERPRISE
<b>250- and 400-Gbyte capacities</b>	Increases capacity density, lowers cost-per-Gbyte for nearline storage applications
<b>Single-port, 1.5 Gbit/sec SATA II 1.0 interface</b>	Interoperates within high-reliability SAS network architectures and management schemes
<b>1.0 million hours MTBF</b>	Operates in intermittent workload nearline environments without compromising application performance or availability
<b>16-Mbyte cache</b>	More efficient processing of peak nearline enterprise workloads
<b>Native Command Queuing</b>	Optimizes processing of multi-threaded or concurrent workloads in enterprise environment
<b>Deterministic Error Recovery</b>	Improves data integrity, reduces data error rates
<b>Rotational Vibration (RV) attenuation</b>	NL35 Series drives can be integrated in highly dense configurations, enabling multi-terabyte capacities in a single tray
<b>Advanced servo performance</b>	High servo sample rates, tighter servo control maintains data integrity in high-workload applications

### Conclusion

For many enterprises, the reality of tiered storage has fallen well short of its promise. Intended to increase efficiency and reduce storage costs, tiered storage instead often adds expense and complexity by addressing nearline storage duties with desktop-class drives that are incompatible with enterprise infrastructures. The advent of high capacity, low-cost-per-Gbyte enterprise drives that complement, rather than complicate, such infrastructures heralds a new era of efficiency and cost-effectiveness in tiered storage. With the introduction of the Seagate NL35 Series disc drives, the promise of tiered storage is finally fulfilled.