

TP-540 • From: Global Product Marketing • June 2005



Technology Paper

Nearline Storage Requirements: Choosing the Optimal Disc Drive

Introduction

With the growing availability of tiered storage solutions, the enterprise is entering a golden age of storage performance and efficiency. Never before has there been such a comprehensive range of disc drives to meet every level of enterprise storage applications. Things have come a long way in just a few short years...

In the late nineties enterprise data was largely transactional, so it was only natural to employ costly high-performance drives for virtually all disc-based enterprise storage applications. However, as time passed and the quantity of reference data (including safety copies of mission-critical files) mushroomed, it became clear that more targeted, cost-effective storage was needed.

Happily, such high-efficiency storage has now become a reality. Tiered storage solutions utilize purpose-built disc drives to ensure every application—from online transactional duties to nearline bulk storage tasks—can be addressed with the optimal blend of performance, reliability and cost-effectiveness.

Making this possible is the introduction of a new class of nearline disc drives that provide high capacity at a low cost-per-gigabyte. These capacity-optimized drives bridge the gap between mainstream performance-oriented, enterprise-class disc drives and cost-focused tape systems to meet a data center's full range of storage requirements. Available in both Fibre Channel and SATA versions, nearline-ready disc drives rely on custom hardware/firmware features—not just their interfaces—to meet their particular application requirements.

Despite this fact, speculation has arisen regarding this latter category of SATA-based nearline storage solutions. Unlike conventional enterprise disc drives with SCSI, Serial Attached SCSI (SAS) and Fibre Channel (FC) interfaces, SATA-based disc drives are designed for a range of applications, including low-cost desktop computers. Not surprisingly, desktop-class drives are cheaper than the new class of nearline-ready SATA drives specifically engineered for nearline applications in enterprise environments. Might these desktop drives be...well...*good enough*? Many IT pros are left scratching their heads and asking: "Just what *is* the right drive for nearline storage?"

Overview: Enterprise Storage Roster

Before delving into the details of nearline storage requirements, it's instructive to briefly review the disc drive storage landscape as it stands today. Enterprise storage applications can be broken down into three basic categories: **low-cost server**, **nearline** and **online**. The first two offer significant opportunities for SATA disc drives to deliver value. Understanding the distinctions between these discrete environments will help you determine the most appropriate SATA disc drive for your nearline applications.

Of course, there is another layer of storage that lies beneath these three enterprise storage applications. Representing the least demanding storage environment in most businesses, **desktop** computers naturally rely on SATA—the serial successor to parallel ATA. Desktop storage generally involves a single user, primarily storing application software and a smattering of data files on the PC’s internal drive (most data files are housed on a network file server). Unburdened by multiple concurrent users or the need to safeguard mission-critical data files, desktop-class SATA drives are typically engineered to deliver appropriate performance and reliability at low cost. This “bang for the buck” approach is ideal for the desktop, but can become problematic as SATA storage moves into more demanding low-cost server and nearline environments.

Low-cost servers can be seen as the economical jack-of-all-trades in enterprise computing. Easy to configure and deploy, they’re general-use solutions that often support a relatively small group or subset of users (for example, a specific department within a large corporation). In such environments, overall user counts are relatively small and the number of concurrent users modest. In some less demanding low-cost server environments, desktop-class SATA disc drives can be economical alternatives to high-performance SCSI drives, particularly when Native Command Queuing (NCQ) capabilities are incorporated. So equipped, moderate-capacity desktop drives can deliver respectable throughput, typically a higher priority than capacity in low-cost server applications.

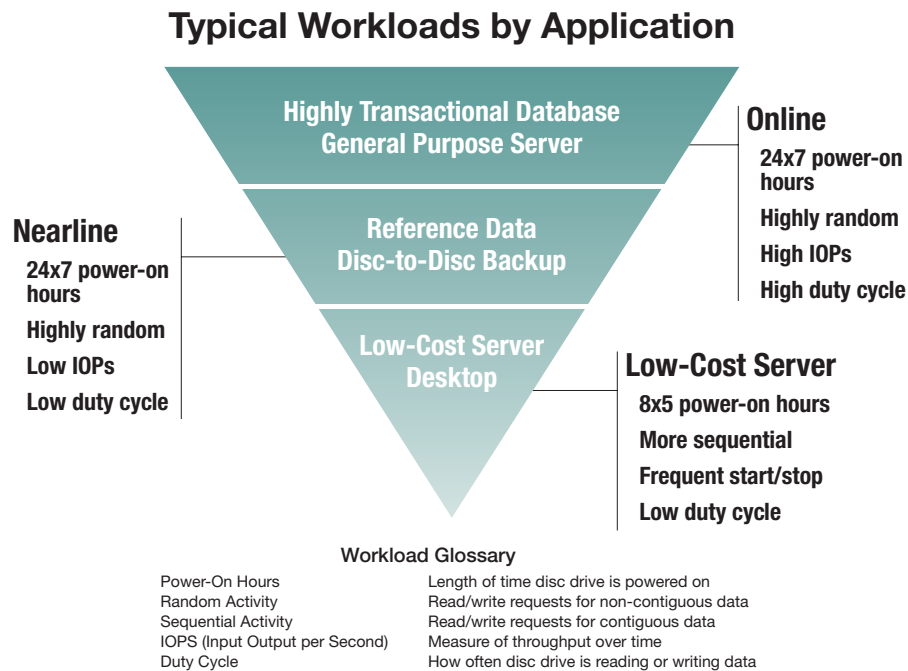
Enterprise Storage Application Profiles		
Application	Application Profile	Storage Profile
Desktop	E-mail, Web browsing, document reading/writing, printing	Economical single-user storage; least demanding storage environment
Low-Cost Server	Departmental file/print/application serving; also small Web or e-mail servers	Internal server storage; pushes desktop-class SATA drives to edge of operational envelope
Nearline	Storing infrequently accessed information, including backup/restore data and reference data, such as medical records	Low-cost bulk storage; requires higher reliability, seamless integration
Online	Transactional databases for companies of all sizes, contact management, other large-scale applications; also Web, application or e-mail serving and small databases employing RAID	Separately attached storage, internal storage; most demanding storage environments, requires highest levels of performance and reliability

Nearline storage is the linchpin of tiered storage in midrange and high-end network environments, filling the gap between online and offline (archival) enterprise storage. Nearline applications can be divided into two basic categories: data protection and reference data. Both entail data that doesn’t justify the cost of high-availability, mission-critical storage, but must still be readily accessible at any time. This last point is key to understanding the rigorous disc drive environments of nearline applications: In order to ensure uninterrupted data access for a broad variety of users, nearline systems are **always** on. By contrast, desktop-class drives store data for only a single user, and are typically powered on only during normal business hours (for instance, eight hours a day, five days a week). Being an “always-on” repository for enterprise data significantly raises nearline-ready drives’ reliability requirements and explains why desktop-class SATA drives are unsuitable for nearline duty.

Online storage application requirements can be neatly summed up with an age-old aphorism: Time is money. Online data-powering everything from highly critical transactional databases to local e-mail servers—is the very lifeblood of many enterprises, and online storage is all about keeping that data flowing. In the online world, diminished throughput and bouts with downtime are far more than inconvenient; they often translate into substantial lost revenues. That’s why the enterprise demands high-performance Fibre Channel and Serial Attached SCSI (SAS) disc drives. Specifically engineered for the rigors of constant access by a multitude of concurrent users, these drives employ sophisticated (and expensive) technology to deliver maximum performance and reliability. In this unforgiving environment, SATA drives are simply out of place.

Different Applications, Different Needs

To better understand the distinction between these enterprise storage applications, consider their real-world workloads. As illustrated below, while the basic elements that define workload (power-on hours, random vs. sequential activity, IOPS and duty cycle) are present throughout, the specific mix of these elements is unique to each application:



Given the diverse workloads of these enterprise storage applications, it's no surprise each application requires disc drives purpose-built to meet its unique operational requirements. Engineered for the new class of nearline storage solutions, nearline-ready drives take full advantage of the latest technologies to ensure optimal reliability and efficiency.

Meeting the Nearline Challenge

While nearline applications don't require the high level of data availability and IOPS demanded by online applications, they do share the need for around-the-clock data accessibility. And though nearline data activity is far less frequent than online activity, both are highly random in nature. These random reads/writes force drive heads to rapidly and repeatedly traverse a drive's discs.

To deliver the enterprise-class reliability standard of 1.0 million hours MTBF, nearline-ready SATA drives are specifically designed to withstand the rigors of random reads/writes and 24x7, always-on operation. In contrast, the typical 600,000 hours MTBF rating of desktop-class SATA drives is obtained in the mild environment of sequential reads/writes and 8x5 power-on hours, and thus has no relevance when considering the use of such drives in nearline applications.

But nearline reliability goes beyond MTBF ratings. Nearline-ready SATA drives also incorporate a Workload Management to dynamically protect them from excessive peak workloads. To further safeguard reliability, nearline-ready drives perform "offline scans" during drive idle time to periodically test the media surface for defects.

Nearline-Ready Firmware

Complementing their emphasis on nearline-class reliability, nearline-ready drives also incorporate firmware features that simplify integration into multi-disc nearline applications, including backup/restore and archiving. These tools boost drive performance, ease integration and enhance data integrity in nearline storage environments.

Error Recovery Control, for example, uses the ATA Smart Command Transport protocol to allow servers using nearline-ready drives to adapt error recovery timing to match the application. Conventional desktop-class SATA drives have a fixed error recovery timing that can cause problems such as timeout errors in nearline system configurations.

Another nearline-ready feature that leverages Smart Command Transport is Write Same, which allows multiple identical writes to be sent with a single command, simplifying system integration and operation.

Nearline-ready drives have enhanced firmware download capabilities to streamline the process. These features simplify firmware upgrades and eliminate the disruption of individually rebooting tens or hundreds of drives in a nearline system.

SATA and the SAS Connection

SAS compatibility with SATA enables tiered storage in a single, enterprise-class SAS infrastructure. A SAS/SATA tiered storage strategy elevates the value of SATA-based nearline disc drives and deserves closer examination.

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 Gbits/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, nearline-optimized SATA drives seamlessly integrate into SAS infrastructures, enabling both performance (SAS) and capacity (nearline SATA) solutions to coexist in the same SAS domain—indeed, even the same enclosure. Beyond the significant cost savings of utilizing a common infrastructure, nearline-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/Gbyte economies without compromising enterprise management and reliability requirements
- Increase storage capacity while maintaining application requirements

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

Conclusion

Tiered storage enables greater performance, reliability and efficiency by utilizing storage devices optimized for their specific applications. As previously discussed, nearline applications present unique workloads to disc drive storage, workloads well beyond the capabilities of desktop-class SATA drives.

It's important to remember that the fundamental value of nearline-ready disc drives does not derive from their interfaces per se. Depending on the specific storage infrastructure and/or nearline applications involved, Fibre Channel drives may be more appropriate than their SATA counterparts. But whatever their interface, nearline-ready drives are engineered to deliver a common core of key capabilities: ease of integration, reliability, capacity and performance.

Nearline-ready SATA drives exemplify a fundamental rule of tiered storage: Use the right tool for the job. Purpose-built to deliver optimal reliability and integration in nearline storage applications, nearline-ready SATA drives ensure you'll reap the full rewards of your tiered storage investment.