Desktop, Nearline & Enterprise Disk Drives

What’s the difference?

Willis Whittington, Seagate Technology
Feb 21st, 2008
Abstract

Desktop, Nearline & Enterprise Disk Drives
What’s the difference?

For the past twenty five years the storage marketplace has been divided into two major categories namely “Desktop” and “Enterprise”. Recently, a third player variously known as “Nearline”, “Reference” or “Business Critical” has evolved to provide a low cost, high capacity storage solution for Enterprise data that no longer needs to exist in a high availability transactional processing environment but must maintain 24 x 7 availability as a reference or backup resource.

Each of these classes of drives requires a unique and specific set of attributes to fulfill its role. This presentation will explore these differences and explain why you need to use

the right drive for the right application.
Agenda
Basic Comparisons
SAS & SATA Compatibility
The Advantages of Nearline SAS
Rotational Vibration
Data Error Rate
Error Correction Capability
Data Integrity
Performance
Annualized Failure Rate
~ Q & A along the way ~
# Comparison Table  DT / NL / MC*

<table>
<thead>
<tr>
<th>Metric</th>
<th>Desktop (DT)</th>
<th>SATA Nearline (NL)</th>
<th>SAS Nearline (NL)</th>
<th>Enterprise MC*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Capacity (GB)</td>
<td>1,000</td>
<td>1,000</td>
<td>1,000</td>
<td>450</td>
</tr>
<tr>
<td>Cost</td>
<td>Low</td>
<td>Mid</td>
<td>Mid</td>
<td>High</td>
</tr>
<tr>
<td>Power Consumption</td>
<td>1x</td>
<td>1.2x</td>
<td>1.2x+</td>
<td>1.5x</td>
</tr>
<tr>
<td>MTBF (Hrs)</td>
<td>600,000</td>
<td>1,200,000</td>
<td>1,200,000</td>
<td>1,600,000</td>
</tr>
<tr>
<td>Duty Cycle</td>
<td>Low (&lt;10%)</td>
<td>Low/Medium (&lt;20%)</td>
<td>Low/Medium (&lt;20%)</td>
<td>High (100%)</td>
</tr>
<tr>
<td>Data Integrity</td>
<td>Parity (?)</td>
<td>EDC + (ECC?)</td>
<td>EDC/ECC + Proprietary Data Integrity Protection</td>
<td>EDC/ECC + Proprietary Data Integrity Protection</td>
</tr>
<tr>
<td>Unrec Error Rate</td>
<td>10^{-14}</td>
<td>10^{-15}</td>
<td>10^{-15}</td>
<td>10^{-16}</td>
</tr>
<tr>
<td>RV Radians/sec(^2)</td>
<td>6</td>
<td>12.5</td>
<td>12.5</td>
<td>&gt;21</td>
</tr>
<tr>
<td>Error Recovery</td>
<td>SATA</td>
<td>SATA + Time Control</td>
<td>Full SCSI</td>
<td>Full SCSI</td>
</tr>
<tr>
<td>Firmware/Features</td>
<td>Standard SATA</td>
<td>SATA + Selected Nearline Features</td>
<td>SCSI + Adv. Features (Enabled by Dual CPU)</td>
<td>SCSI + Adv. Features (Enabled by Dual CPU)</td>
</tr>
<tr>
<td>Power On Hrs/Year</td>
<td>2400</td>
<td>8,760</td>
<td>8,760</td>
<td>8,760</td>
</tr>
<tr>
<td>Multi Initiator</td>
<td>No</td>
<td>No</td>
<td>16 Hosts &amp; Dual Port</td>
<td>16 Hosts &amp; Dual Port</td>
</tr>
<tr>
<td>Performance</td>
<td>1x</td>
<td>1x</td>
<td>1x+</td>
<td>1.4x / 2.5x (Seq / Rand)</td>
</tr>
<tr>
<td>T10 Data Protection</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Scalability</td>
<td>Low</td>
<td>Low</td>
<td>High + Dual Port</td>
<td>High + Dual Port</td>
</tr>
</tbody>
</table>

Key:  
- **Good**
- **Better**
- **Best**

*Mission Critical
Anatomy of an Enterprise Drive

Motor
Higher rpm than NL or DT
Tighter specifications
Less runout
More expensive
Anatomy of an Enterprise Drive

Discs
- Four platter design
- Smaller diameter than NL/DT
- Full media certification
- Fully characterized
- Variable sector format
Anatomy of an Enterprise Drive

Head Stack
Eight head design
Low mass, high rigidity
Voice coil designed for
  - optimal performance
  - 100% duty cycle
Higher cost design
Anatomy of an Enterprise Drive

Environmental Control Module
Humidity Control
Chemical Absorption
Multi-point filtration
Windage Design
Anatomy of an Enterprise Drive

Misc Mechanical
- Powerful Voice Coil Motor
- Stiffer Covers
- Air Control Devices
- Faster Seeks
- High Servo Sample Rate
- Low RV susceptibility
Anatomy of an Enterprise Drive

Electronics
Dual processors
Multi host queuing
Dual port
 Twice the memory of NL/DT
High rpm control
Command scheduling
Superior error protection
Superior error correction
Smart servo algorithms
Perform. optimization
Data integrity checks
Sequential h/w assist
I/O Connectors for SAS & SATA

For SAS, the key-way is filled in and its flip side is used for the 2nd Port.

This prevents a SAS drive from being plugged into a SATA cabinet slot.
The Advantages of Nearline SAS
NL SATA Compared to NL SAS

Stepping up to SAS provides Mission Critical Compatibility

NEARLINE SATA

- Full Duplex (Bidirectional) I/O
- Concurrent Data Channels
- Enterprise Command Queuing
- Full SCSI Command Set
- Multiple Host Support
- Variable Sector Size
- End-to-End Data Integrity
- 100% Phy Compatible
- Full EDC & ECC
- Dual Port

Mission Critical Features
SAS/SATA NL Physical Differences

SAS Electronics

SAS Port “B”

Nearline Head/Disc Assy.

SATA Electronics
SAS/SATA NL Differences

Full Duplex
Dual Port

SAS

Full Duplex
In both directions at one time

Two Data Channels
2 Concurrent Writes
OR
2 Concurrent Reads
OR
1 Write + 1 Read

One Data Channel
1 Write
OR
1 Read

SATA

Half Duplex

Xmits
In one direction at one time

Xmits
In both directions at one time
Multi Host Command Queuing

Drive Queue Supports 16 Hosts

Drive NCQ* supports a single Host

*Native Command Queuing
NL SAS/SATA Summary

Both SATA and SAS Nearline drives are designed for use in Enterprise Mission Critical environments.

SAS Nearline drives have additional advantages which are made possible by the Serial SCSI interface and enterprise electronics:

- Full system interface compatibility at the protocol, physical ("phy"), and command level
- Enterprise error recovery and performance optimization controls
- Full data integrity protection both within the drive and at the system level with DIF support.
Rotational Vibration
PS drives are not designed for backplane (JBOD/SBOD) use and are not equipped to cope with the effects of RV HDA subjected to rotational forces.

**Neighboring Drive’s Servo needs to compensate for externally induced RV**

**RV is Proportional to Seek Current**

'Scope Picture, Seagate Prod. Dev.
Rotational Vibration

Impact on Performance*

![Graph showing impact on performance of different types of drives under rotational vibration. The graph plots % degradation on the y-axis against RV level (rad/sec²) on the x-axis. Three categories are shown: Desktop, Nearline, and Enterprise. The graph illustrates how different levels of RV affect the percentage of degradation for each category.]

*Source: STX Competitive Analysis.
RV in 33 Different Cabinets

Rotational Vibration

- More stringent RV spec. needed for SATA cabinets
- RV aggravated by system fans, random access and “bursty” workloads
UER* on High Capacity RAID Sets

The UER for SATA desktop is 1 in $10^{14}$ bits transferred
- $10^{14}$ bits = $12\frac{1}{2}$ terabytes

A 500 Gbyte drive has $1/25 \times 10^{14}$ bits

Rebuilding a SATA drive in a RAID 5 set of 5 drives means transferring $5/25 \times 10^{14}$ bits = $1/5$ of UER spec.
- 20% probability of an Unrecoverable Error during the rebuild.

Better odds would be available with RAID 1 or 6
- RAID 1 rebuilds from a single mirror drive
- RAID 6 can tolerate a second error during the rebuild.

Risks can be reduced with good error management
- Intelligent rebuild (ignore unused capacity)
- Background media scan (dynamic certification)

*Unrecoverable Error Rate
**DT / NL / MC UER**

Probability of Unrecoverable Errors during RAID Rebuild

- **Desktop Drives UER = 10^{-14}**
- **Nearline UER = 10^{-15}**
- **Enterprise UER = 10^{-16}**

*Unrecoverable Error Rate*
Error Correction Capability
Standard vs Reverse ECC

(Write Command)

**Standard ECC**
- User Data → User
- User → ECC Generator
- ECC Generator → + → User ec
- User ec → Randomizer
- Randomizer → randomized
- randomized → RLL Encoder
- RLL Encoder → encoded
- encoded → ECC Generator
- ECC Generator → + → User Data

**Reverse ECC**
- User Data → User
- User → Randomizer
- Randomizer → randomized
- randomized → RLL Encoder
- RLL Encoder → encoded
- encoded → ECC Generator
- ECC Generator → + → User Data
Standard vs Reverse ECC

(Read Command)

Standard ECC

- User Data
  - ECC Correction
  - De-randomize
  - 120 Bit Error (Propagates in Decoder)
    - RLL Decoder
      - 20 Bit Error
        - Error Corrected on the fly

Reverse ECC

- User Data
  - De-randomize
  - RLL Decoder
  - No Error Propagation
  - Error Corrected on the fly
Sync Mark Errors on SATA Drives

- The Sync Field is used to get the read channel in frequency sync with the data recorded on the media.
- The Sync Mark is used to define the beginning of the User Data Field.
- Failure to recognize the Sync Mark (due to a thermal asperity or a grown media defect) means the User Data Field is not delineated and the data is lost.
Sync Mark Errors on SAS Drives

Read Operation

Sync Field → Media Flaw → Sync Mark 2 embedded in the data field

Sync Mark

User Data Field

Read Channel finds SM2 and reads the data following it into the Buffer, starting at location n+1

Sector Format

Sync Mark

Media Flaw

Sync Mark 2 embedded in the data field

n bytes

ECC

0011011000011110101011000011010100110110001010101000111

Before ECC

DATA BUFFER

0000000000000000000110110000111101011000011010100110110001010101000111

After ECC

The missing n bytes are recovered using the ECC

1101011101001110011011000011110101100011010100110110001010101000111

Read Channel realizes the first Sync Mark is missing:
- Loads Buffer with n zeroes
- Starts searching for SM2

Read Channel finds SM2 and reads the data following it into the Buffer, starting at location n+1

The missing n bytes are recovered using the ECC
Performance
Performance Comparison

Enterprise Sequential Access

Desktop / NL Sequential Access

Vendor Range

Sequential Transfer Rate $\propto$ RPM x Disc Dia x Bit Density

Enterprise Transfer Rate = $\frac{RPM \times Disc \, Dia \times Bit \, Density}{RPM \times Disc \, Dia \times Bit \, Density}$

= $\frac{15000 \times 65}{7200 \times 95}$

= 142%

(independent of seek time & Latency)
Performance Comparison

Enterprise Sequential

Desktop / NL Sequential

Mbytes/sec

100
90
80
70
60
50
40
30
20
10
0

1 4 16 32

Operations/sec

400
360
320
280
240
200
160
120
80
40
0

Enterprise Random

Desktop / NL Random

Seagate
SPC-1C comprises I/O operations demonstrating small storage subsystem performance (1-16 drives) while performing the typical functions of a business critical application.
AFR
(Annualized Failure Rate)
SATA drives in Enterprise applications run hotter, at higher duty cycle, and for more Power-On-Hours than in desktop applications.
In Conclusion.....

Although technological advances, driven by Enterprise research, will be leveraged into SATA products, there will continue to be functional limitations imposed on these devices by the overriding metric of Low $/GB Storage.