Desktop, Nearline & Enterprise Disk Drives

Willis Whittington, Seagate Technology
The material contained in this tutorial is copyrighted by the SNIA.

Member companies and individuals may use this material in presentations and literature under the following conditions:

- Any slide or slides used must be reproduced without modification
- The SNIA must be acknowledged as source of any material used in the body of any document containing material from these presentations.

This presentation is a project of the SNIA Education Committee.
Desktop, Nearline & Enterprise Disc Drives
- Deltas by Design -

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
- SATA Native Command Queuing
- Error Recovery Limitations
- Error Correction Capability
- Data Integrity
- Performance
- Annualized Failure Rate

~ Q & A along the way ~
Basic Comparisons
## Comparison Table  DT / NL / MC

<table>
<thead>
<tr>
<th>Feature</th>
<th>Desktop (DT)</th>
<th>Nearline (NL)</th>
<th>Enterprise MC*</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Capacity</strong></td>
<td>1,000 GB</td>
<td>1,000 GB</td>
<td>450 GB</td>
</tr>
<tr>
<td><strong>Cost</strong></td>
<td>Low</td>
<td>Mid</td>
<td>High</td>
</tr>
<tr>
<td><strong>Power Consumption</strong></td>
<td>1x</td>
<td>1.2x</td>
<td>1.5x</td>
</tr>
<tr>
<td><strong>Reliability (MTBF)</strong></td>
<td>600,000 Hours</td>
<td>1,200,000 Hours</td>
<td>1,600,000 Hours</td>
</tr>
<tr>
<td><strong>Duty Cycle</strong></td>
<td>Low (&lt;10%)</td>
<td>Low/Medium (&lt;20%)</td>
<td>High (100%)</td>
</tr>
<tr>
<td><strong>Data Integrity</strong></td>
<td>Parity (?)</td>
<td>EDC + (ECC?)</td>
<td>EDC/ECC + Proprietary Protection</td>
</tr>
<tr>
<td><strong>Tolerance to RV</strong></td>
<td>6 Radians/sec²</td>
<td>12.5 Radians/sec²</td>
<td>&gt;21 Radians/sec²</td>
</tr>
<tr>
<td><strong>Error Recovery</strong></td>
<td>Standard SATA</td>
<td>SATA + Time Control</td>
<td>Sense Keys, Codes, FRUs</td>
</tr>
<tr>
<td><strong>Firmware/Features</strong></td>
<td>Standard SATA</td>
<td>SATA + Selected Nearline Features</td>
<td>SCSI + Advanced Features (Enabled by Dual CPU Power)</td>
</tr>
<tr>
<td><strong>Power On Hours</strong></td>
<td>2400 Hours/year</td>
<td>8,760 Hours/year</td>
<td>8,760 Hours/year</td>
</tr>
<tr>
<td><strong>Multi Initiator</strong></td>
<td>No</td>
<td>No</td>
<td>16 Concurrent Hosts</td>
</tr>
<tr>
<td><strong>Performance</strong></td>
<td>1x</td>
<td>1x</td>
<td>1.4x / 2.5x (Seq / Random)</td>
</tr>
<tr>
<td><strong>T10 Data Protection</strong></td>
<td>No</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td><strong>Scalability</strong></td>
<td>Low</td>
<td>Low</td>
<td>High</td>
</tr>
</tbody>
</table>

*Mission Critical*
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
- High rpm control
- Command scheduling
- Superior error protection
- Superior error correction
- Smart servo algorithms
- Perform. optimization
- Data integrity checks
- Sequential h/w assist
SAS & SATA Compatibility
Connectors and cables are required to:

- Support 3.0 Gbps speed with headroom for 6 Gbps
- Be suitable for both 3½” and 2½” storage devices
- Be blind-mateable
- Be hot-pluggable (staggered pins)
SAS/SATA Connector Compatibility

Connector Flip Side

Port B

Keyway

Pluggable

SAS

SATA
SAS/SATA Compatibility

- SATA drives are plug compatible with SAS backplanes
- SAS Topology allows SATA device addressing
- This is done with Serial Tunneling Protocol (STP)
SAS/SATA Compatibility

- SATA drives are **plug compatible** with SAS backplanes
- SAS Topology allows SATA device addressing
- This is done with Serial Tunneling Protocol (STP)

SAS Drives

SATA Drives

Interposer Cards boost transceiver power and add an active/active mux

Pseudo Dual Port

Transceivers
Port Selector

Transceivers
Port Selector

Dual Mode (SAS and SATA) Edge Expander
SATA to SAS Interposer Card

- SATA drives are designed for a desktop/cabled interconnect
- Onus is on the Integrator to assure use in Backplane applications
- Interposer card
  - Provides additional Transceiver drive capability in SAS cabinets
    - Necessary to preserve signal integrity in medium to large backplanes
  - Provides additional electronics for “Port Selector” feature
    - Buffer and Mux for active-active access

Drive Carrier Assy

Interposer Board

Standard or custom connector to host Back Panel
The Advantages of Nearline SAS
SATA Compared to SAS

Stepping up to SAS provides Mission Critical Compatibility

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

NEARLINE SATA
- Mission Critical Features

SATA Compared to SAS
- Mission Critical Compatibility
Physical Differences Between SAS & SATA

SAS Electronics

SAS Port “B”

Nearline Head/Disc Assy.

SATA Electronics
Full Duplex, Dual Port & 2 Active Channels

**Full Duplex, Dual Port**
- Two Data Channels
  - 2 Concurrent Writes
  - 2 Concurrent Reads
  - 1 Write + 1 Read

**Half Duplex**
- One Data Channel
  - 1 Write
  - OR
  - 1 Read

**Full Duplex**
- Xmits in both directions at one time

**Half Duplex**
- Xmits in one direction at one time
Multi Host Command Queuing

- Drive Queue Supports 16 Hosts
  - SAS
  - Interposer to handle Q’ing for 2\textsuperscript{nd} host

- Drive NCQ* supports a single Host
  - SATA

*SAS EXPANDER*
In Conclusion …

- 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.
PS drives are not designed for backplane (JBOD/SBOD) use and are not equipped to cope with the effects of RV. Neighboring Drive’s Servo needs to compensate for externally induced RV.

HDA subjected to rotational forces

RV is Proportional to Seek Current

'Scope Picture, Seagate Prod. Dev.
Impact on Performance*

*Source: STX Competitive Analysis.
More stringent RV spec. needed for SATA cabinets
RV aggravated by system fans, random access and “bursty” workloads
Data Error Rate
The UER for SATA desktop products is 1 in $10^{14}$ bits read
- $10^{14}$ 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
- This means there is a 20% probability of an Unrecoverable Error during the rebuild.

Better odds would be available with RAID 1 or 6
- Higher redundancy, faster Rebuild, more expensive

Risks can be reduced with good error management
- Intelligent rebuild (e.g. ignore unused capacity)
- Background media scan

*Unrecoverable Error Rate
Nearline UER = 10^{-15} (2007)

Enterprise UER = 10^{-16} (2006)

*Unrecoverable Error Rate
SATA Native Command Queuing
Error Recovery Limitations
Creating the Command Queue

Command 6
Command 5
Command 4
Command 3
Command 2
Command 1
Optimize Queue and Execute

Commands will not be executed

Execute
Execute
Execute

Command 6
Command 3
Command 4

Status = Good
Status = Error

Commands will not be executed
Consequences of an Error

- Command Queue processing halts on an ERROR
- All commands remaining in the Queue are cleared
Consequences of an Error

- Command Queue processing halts on an ERROR
- All commands remaining in the Queue are cleared
- Host must reissue outstanding commands

Complicates Dual Port Interposer Applications
Error Correction Capability
Standard vs Reverse ECC

(Write Command)

**Standard ECC**

- User
  - User Data
  - +
  - ECC Generator
  - ec

**Reverse ECC**

- User
  - User Data
  - randomized
  - Randomizer
  - ec
  - RLL Encoder
  - encoded

Deltas by Design
© 2007 Storage Networking Industry Association. All Rights Reserved.
Sync Mark Errors on Desktop/NL 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 MC Enterprise drives

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

Before ECC
0000000000000000110110000111101010110000110100110100100110010100100111

After ECC
11010111101001100110110000111101010110000110100110100100110010100100111

Deltas by Design
© 2007 Storage Networking Industry Association. All Rights Reserved.
Data Integrity
End to End Data Checking

**Data Frame**

- **SOF**
- **Header**
- **User Data**
- **End to End Data Check**
- **CRC**
- **EOF**

- **Logical Block Guard**
- **Logical Block Application Tag**
- **Logical Block Reference Tag**

- 2 Bytes
- 2 Bytes
- 4 Bytes

*Host and drive can check the data*

Not available on Desktop and SATA Nearline drives with fixed 512 byte sectors
Data Integrity Protection in Enterprise Drives

- Data Frame from Host
  - Data checked against CRC
  - CRC Discarded, temporary Parity added
  - LBA Seeded Error Detection Code added to Data
  - Error Correction Code added to Data & IOEDC
  - Protected Data now processed by drive

- DDR SDRAM
  - Data checked against IOECC
  - IOECC Discarded
  - Data & IOEDC are RLL encoded
  - Reverse ECC added
  - Final EDC added before writing

- Buffer
  - Data checked against IOECC
  - IOECC Discarded
  - Data & IOEDC are RLL encoded
  - Reverse ECC added
  - Final EDC added before writing

- Disc Drive Block Diagram
  - Eliminates the risk of unreported data miscompare errors

I/O Connector

- I/O Controller

- Data Frame from Host

- DDR SDRAM

- Buffer

- Disc Drive Block Diagram

- Formatter

- Data Integrity Protection in Enterprise Drives

Deltas by Design  
© 2007 Storage Networking Industry Association. All Rights Reserved.
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

\[
\frac{\text{Enterprise Transfer Rate}}{\text{Desktop Transfer Rate}} = \frac{\text{RPM x Disc Dia x Bit Density}}{\text{RPM x Disc Dia x Bit Density}}
\]

\[
= \frac{15000 \times 65}{7200 \times 95}
\]

\[
= 142\%
\] (independent of seek time & Latency)
SPC-1C Performance Comparison

- SPC-1C comprises I/O operations designed to demonstrate 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.
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.
Q&A / Feedback

Please send any questions or comments on this presentation to SNIA: trackstorage@snia.org

Many thanks to the following individuals for their contributions to this tutorial.

Craig Parris
Daniel Dummer
Willis Whittington
Wolfgang Rosner

SNIA Education Committee