Tape Data Integrity and Fixity

Michael O’Donnell
Director, HW Development

May 2013
The Importance of Data Integrity
Current Data Protection Options Are Costly
Cost Often Outweighs Risk of Data Loss

1 error can be expected in ~1 EB written to enterprise tape*

How do I know my data is safe?

OPTION 1: REPLICATE DATA
› Doubles hardware costs

OPTION 2: PERIODICALLY READ BACK ALL THE DATA
› Adds stress to datacenter infrastructure, increased hardware and people costs

OPTION 3: TRUST AND DO NOTHING

* 1 error in ~10 PB on LTO tape and 1 error in ~10 TB on disk
Integrity for the Life of Your Data

Continuous Archive Validation

1. Data Write
2. Mount/Exchange
3. Data Read
4. Data Integrity Validation

Tape Analytics
Health Analysis

Periodic Media Scan

Data Integrity Validation

Media Validation
5 Ways to Initiate Media Validation
Choose the Path that Best Meets Your Needs

- Simple > SLC (library controller)
- Automated Policies > StorageTek Tape Analytics 2.0
- StorageTek Storage Archive Manager (SAM) 5.3
- Tivoli Storage Manager 6.3.3
- Custom Application
Archive End to End Data Integrity

StorageTek Data Integrity Validation

- User creates a CRC (T10 ANSI standard) for each record
- StorageTek T10000C checks CRC as each record is received
- The CRC of each record is written to tape with that record
- When a record is read from tape the CRC is always checked
  - The SCSI Verify command can be used to check each record without transferring data to the application. (i.e. internally verified by the StorageTek T10000C)
First, an Example...
Data is Corrupted Inside the Server and Unknowingly Written to Tape
Conventional Data Validation
Inefficient Process Consumes Time and Server Resources

Write Step 1:
Server pulls file from disk, generates and catalogues a 256-bit CRC over entire file

Verification Step 1:
Server software recreates CRC over entire file; compares to catalogued CRC

Verification Step 2:
If CRCs do not match, server re-sends entire file

Write Step 2:
Server sends file to tape

Corrupted file is written to tape inside server

File is corrupted

Server reads back corrupted file

Disk

Server

Tape Drive

Cartridge

(4GB file, 2:1 compression)
Data Validation with StorageTek T10000C
Discover Corrupted Records *Before* They Are Written to Tape

**Write Step 1:**
Server pulls file from disk, and generates 32-bit CRC over each record within file.

**Write Step 3:**
Server re-sends only the failing record for comparison, alerts server of corrupted record.

**Verification Step 1:**
Server reads back file with 32-bit CRCs.

**Verification Step 2:**
Server hardware compares re-created CRCs to tape CRCs.

**Verification Time:** 14 seconds
(4GB file, 2:1 compression)
Media Validation

Periodic Archive Media Scans
SAM QFS + T10000C Means Fast, Secure Audits
Data is Quickly Verified Within T10000C, NOT an External Server

Step 1:
Verification Time: 9 seconds
Server sends SCSI Verify command to T10000C
on tape to re-created CRCs

Step 2:
T10000C reads record(s) while comparing CRCs on tape to re-created CRCs

Step 3:
T10000C sends SCSI Verify status to server
The Importance of Tape Complex Monitoring
Tape Management Challenges
Drive & Media Exchange Combinations Add Complexity
StorageTek Tape Analytics
Proactively Manage the Entire Tape Storage System from a Single Interface

✓ Simplify Tape Management
  Monitors your drives and media so you can focus your resources elsewhere

✓ Leverage Intelligent Analytics
  Provides proactive health indicators that can be trusted, using Oracle’s proprietary algorithms

✓ Worry Free Deployment
  Collects performance data via the library without entering your live data path

✓ Grow with Peace of Mind
  Scales easily and monitors globally dispersed libraries from a single interface

<table>
<thead>
<tr>
<th>Flexible Hardware Support</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
</tr>
<tr>
<td>Tape Drives</td>
</tr>
<tr>
<td>StorageTek T10000A,B,C</td>
</tr>
<tr>
<td>StorageTek LTO3,4,5,6</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
</tbody>
</table>
Automated Health Monitoring
Tape Analytics Monitors Every Tape Exchange, Simplifying Management
StorageTek Tape Analytics
Monitoring Every Exchange in the Archive

Tape Analytics Analyzes Over 150 Exchange Health Attributes
Introducing StorageTek Library Media Validation

StorageTek SL8500 or SL3000 + StorageTek Tape Analytics + StorageTek T10000C
Library Media Validation

3 Easy Steps

**Flexible Drive Selection**
Select any drive in library to perform validations

**Flexible Media Selection**
Select any cartridge in library to verify

**Flexible Verify Selection**
Select type of media verify to perform
- Basic Verify
- Standard Verify
- Complete Verify
- Complete Verify Plus

MEDIA VERIFY OPTIONS

TAPE DRIVES

Drive Pool

MEDIA

Media Queue
Designed for Simplicity, Security, and Speed

**Simplicity**
- The MOST initiation options:
  - Initiate with Tape Analytics, library, and other applications
- Keep validation inside the T10000C
  - No extra fibre
  - No extra servers
  - No key management

**Security**
- Data stays secure within the tape drive
  - No need to move the data to an external server
- Cartridges stay secure within the library

**Speed**
- T10000C offers the FASTEST data validation speeds
- No keys required to run scans
- No speed limit with compressed data
Integrity for the Life of Your Data

Continuous Archive Validation

1. Data Write
2. Mount/Exchange
3. Data Read
4. Media Validation

Data Integrity Validation

Periodic Media Scan

Tape Analytics Health Analysis