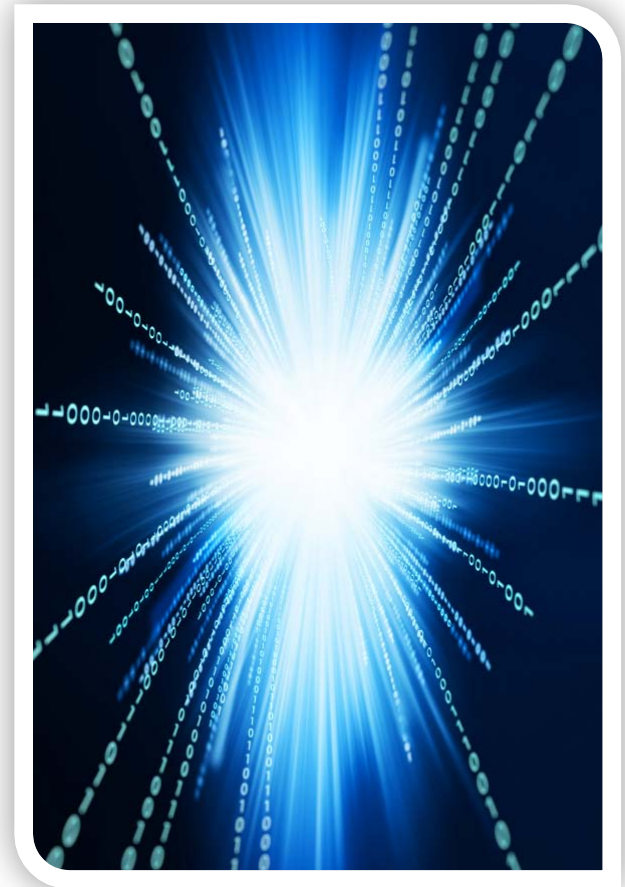


# Big Data and the Evolution of Tape Technologies

**Kevin Dudak**  
**Spectra Logic**

- ❑ The explosion of Big Data is breaking the traditional ways we store data.
- ❑ Evolutions in tape technology have opened up new ways to store and manage these massive data sets.
- ❑ This session will look at new hardware and software that opens up new possibilities for massive data storage

- ❑ Data sets are getting more massive
- ❑ Most of the data growth is in unstructured data
- ❑ The growth will never end



# Analytics changes the growth curve

- ❑ Analytics is driving
  - ❑ New Value to old data
  - ❑ Data is now being used for unintended purposes



# What should we save

- ❑ It isn't as easy to tell anymore what data has value
  - ❑ Server logs
  - ❑ Surveillance videos
  - ❑ What's the next thing?



# Scale causes new problems

- ❑ How do you effectively store Petabytes for years?
  - ❑ How do you protect it?
  - ❑ How do you move it?
    - ❑ 1 PB of data will take 10 days to move over a single 10 GbE link
  - ❑ How do you insure it's integrity?
  - ❑ How do you afford to store it all?

# Tape as a storage location

- ❑ Tape provides cost effective bulk digital storage
- ❑ Traditionally hidden behind a backup application
  - ❑ This made it difficult to use for storage needs outside of backup
- ❑ New Technologies allow you to use and integrate tape into non backup systems

- ❑ Most scalable storage systems in the world
  - ❑ Multiple library systems can scale to Exabytes
- ❑ Lowest energy consumption of any storage type
- ❑ Better Bit Error Rates than Disk





# Bit Error Rates

- Number of bits written before the failure of a bit

| Device                 | Hard error rate in bits | PB equivalent |
|------------------------|-------------------------|---------------|
| SATA consumer          | 1.00E+14                | .11           |
| SATA Enterprise        | 1.00E+15                | 1.11          |
| Enterprise SAS/FC      | 1.00E+16                | 11.10         |
| LTO                    | 1.00E+17                | 111.02        |
| Enterprise Tape Drives | 1.00E+19                | 11,102.22     |

- Error rate for disk is number of bits before a failure of a sector

512 bytes today, 4096 soon

- Error rate for tape is number of bits before the failure of a bit

What difference does a storage device make?

The number of hours of operation (per 100 devices) at maximum rated throughput before a write error occurs.

- Consumer SATA disk: **2.3 Hours**
- Enterprise SATA disk: **21.6 Hours**
- Enterprise SAS/FC disk: **6.7 Days**
- LTO-5 tape: **96.2 Days**
- Enterprise tape: **15 Years**

# Common Tape Features

- ❑ LTO and TS1140 both support
  - ❑ AES-256 Encryption
  - ❑ Speed matching drives
  - ❑ Read back 2 generations
  - ❑ Write back 1 generation
  - ❑ LTFS
  - ❑ Compression
  - ❑ ECC



- ❑ Linear Tape-Open
  - ❑ Multiple drive and library manufactures
- ❑ Multi Vendor Open Format
- ❑ Over 80,000 PB of data protected so far
- ❑ Costs pennies per GB to store data in an automated system
- ❑ Roadmap through Generation 8

- ❑ LTO 6 will be shipping soon
  - ❑ Up to 6.25 TB compressed per tape\*
  - ❑ Up to 400MB/s compressed transfer rate\*
  
- ❑ LTO 6 more than doubles the density of LTO5
  - ❑ LTO5 density in enterprise library 352 TB/Sq Ft
  - ❑ LTO6 density in enterprise library 733 TB/Sq/Ft
  
- \*LTO Consortium Press Release
- <http://www.lto-technology.com/pdf/LTO%20PR%20Maintains%20Top%20Momentum.pdf>

- ❑ IBM manufactured drive – supported by 2 library manufacturers
- ❑ Higher Capacity, Speed and better Bit Error Rate than LTO
  - ❑ TS1140 density in enterprise library 1,074 TB/Sq/Ft
  - ❑ 4 TB compressed per tape
  - ❑ 650 MB/sec compressed transfer rate
- ❑ Up formats TS1130 tapes to higher capacity
- ❑ Roadmap through Generation 6

# Linear Tape File System (LTFS)

- ❑ File interface to tape
- ❑ Introduced with LTO5
- ❑ Open source and specification
  - ❑ Stand alone tape drive looks like a USB memory key
  - ❑ Currently, the specification does not address tape libraries
- ❑ Makes the tape self describing

LTFS 2.0 Specification

[http://www.ultrium.com/technology/LTFS\\_Format\\_To%20Print.pdf](http://www.ultrium.com/technology/LTFS_Format_To%20Print.pdf)

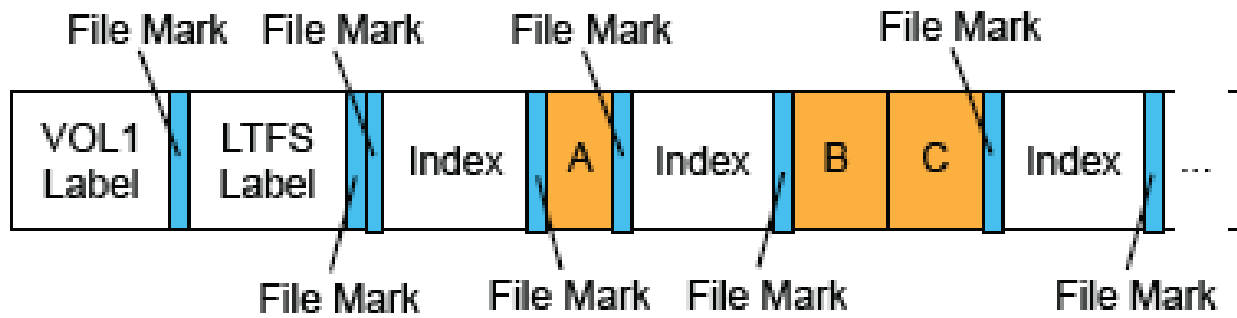
- ❑ Enhances Data Portability: Physical and Logical
  - ❑ Move data easily between sites
  - ❑ Move data between applications





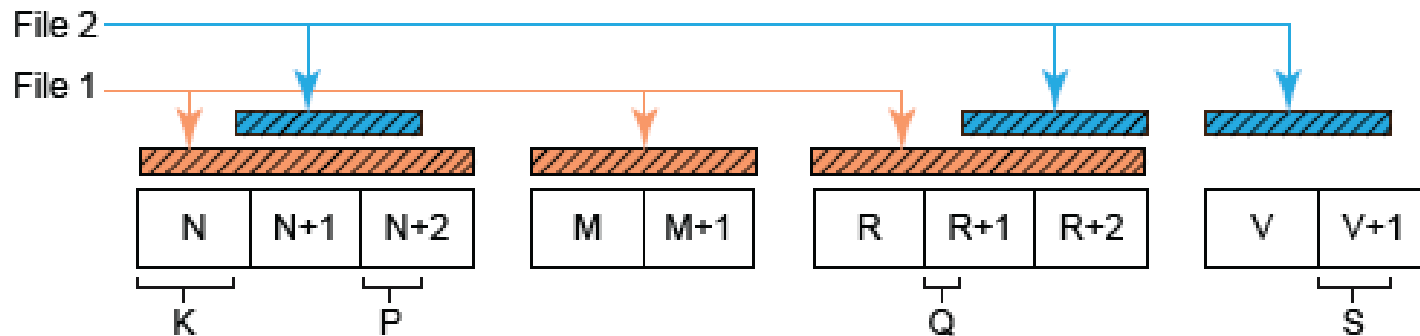
- ❑ How does it work?
  - ❑ Tape is divided into 2 partitions
    - ❑ One Index partition
    - ❑ One data partition
      - ❑ Data partition has indexes as well
  - ❑ Indexing is XML based

# LTFS Data Partition



Source: LTO Consortium ([http://www.ultrium.com/technology/LTFS\\_Format\\_To%20Print.pdf](http://www.ultrium.com/technology/LTFS_Format_To%20Print.pdf))

- The specification supports data reduction or deduplication



Source: LTO Consortium ([http://www.ultrium.com/technology/LTFS\\_Format\\_To%20Print.pdf](http://www.ultrium.com/technology/LTFS_Format_To%20Print.pdf))

## □ NAS Tape

- Leverage a NAS interface and LTFS to create an infinite file storage system
- Easy to integrate into numerous external data systems

- ❑ Lowest cost storage available
- ❑ LTFS changes things
  - ❑ Easier than ever to use tape
  - ❑ You can create your own
  - ❑ Makes it easy to move massive data sets

# Next Steps

- ❑ Consider long term storage needs
- ❑ Should you buy or build?
- ❑ LTFS Format Specification Document

# Questions?