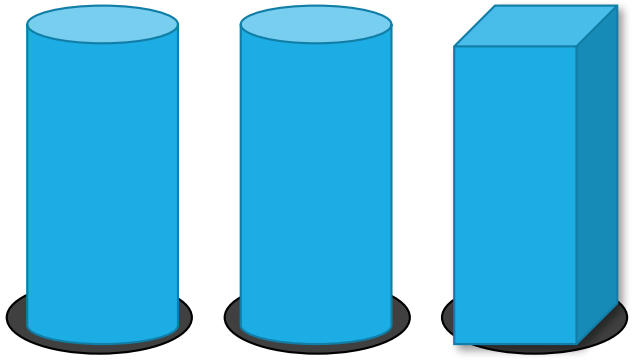
Tape in the Microsoft Datacenter:

The Good and Bad of Tape as a Target for Cloud-based Archival Storage

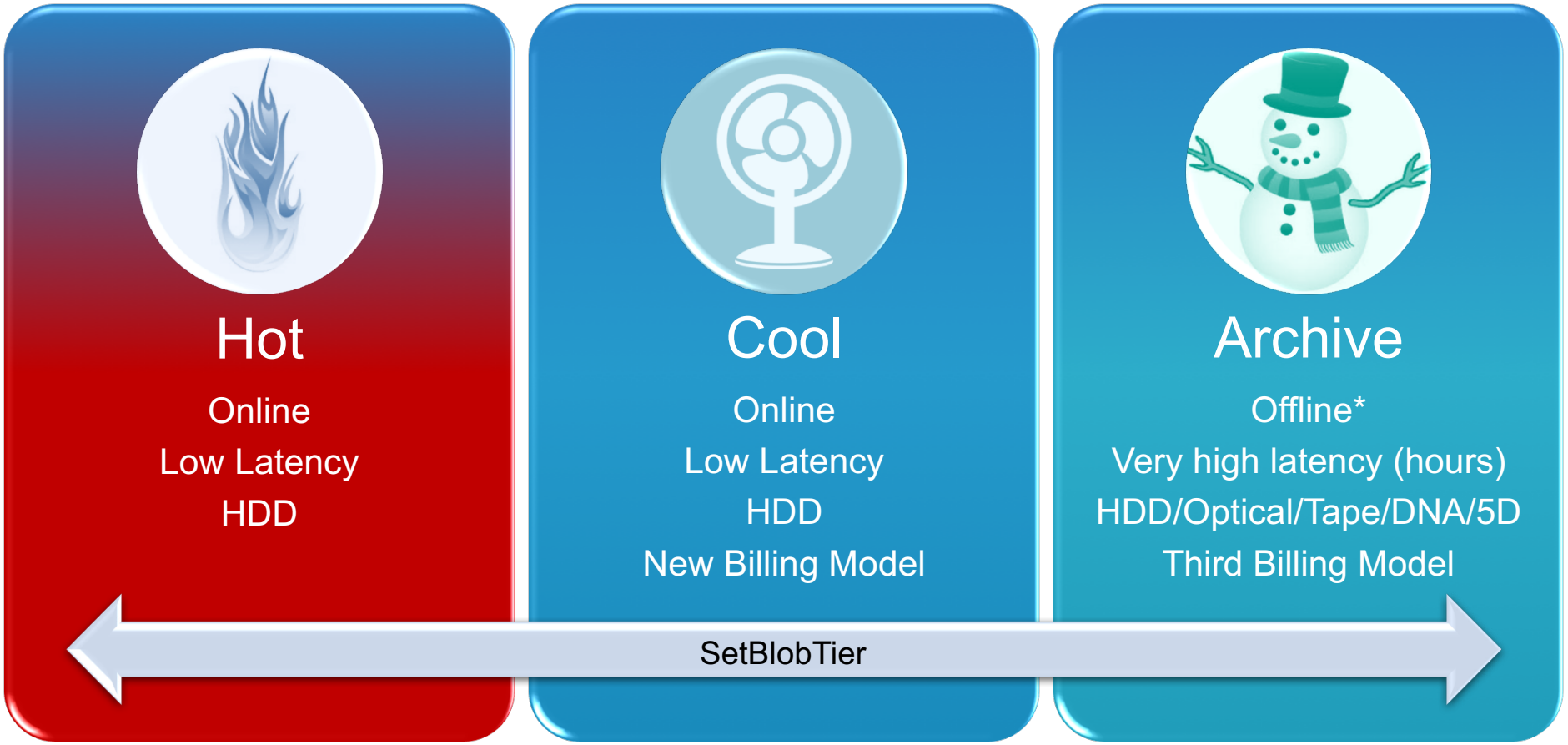
Marvin McNett
Principal Development Manager
Microsoft Azure Storage



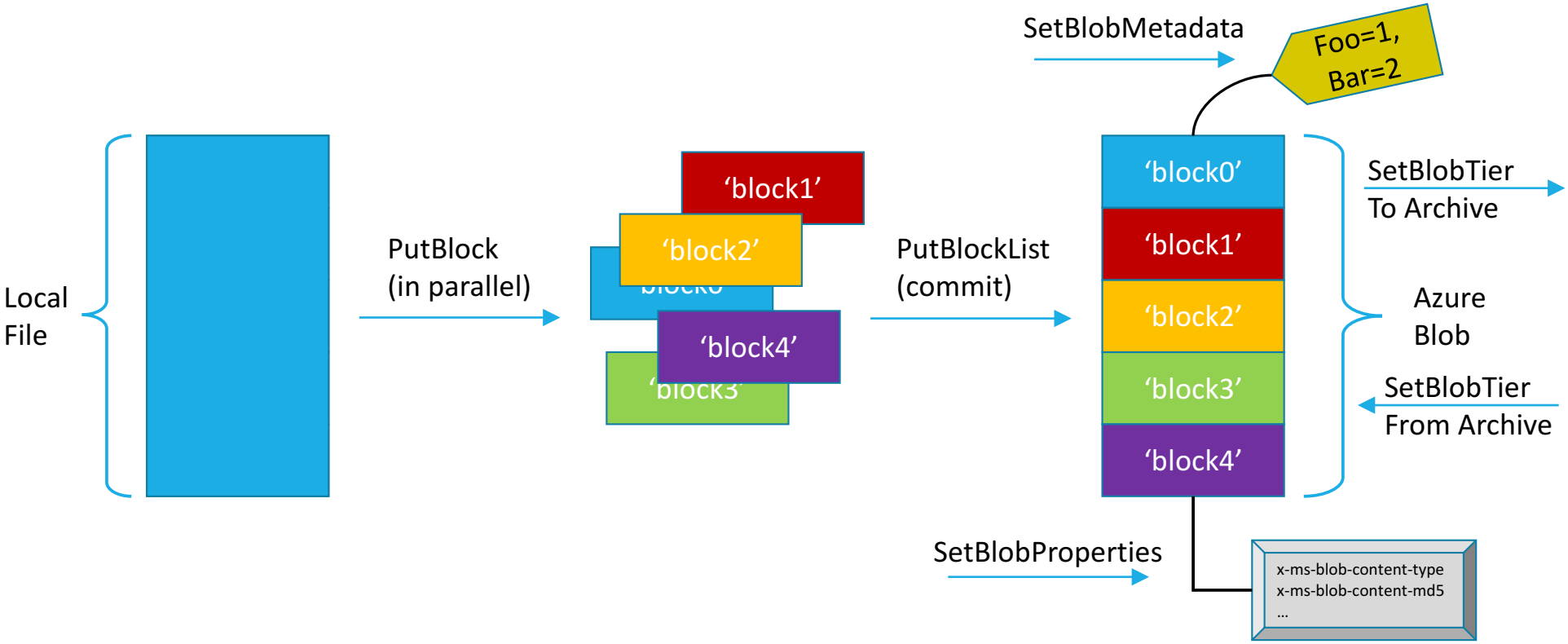
Content

I. Azure Archival Storage
II. A Case for Tape
III. Complications to Large-scale Adoption
IV. Potential Mitigations

Azure Blob Storage



Block Blob Storage



A Case for Tape

as a backing store for Azure Blob Service's archival tier

Economics

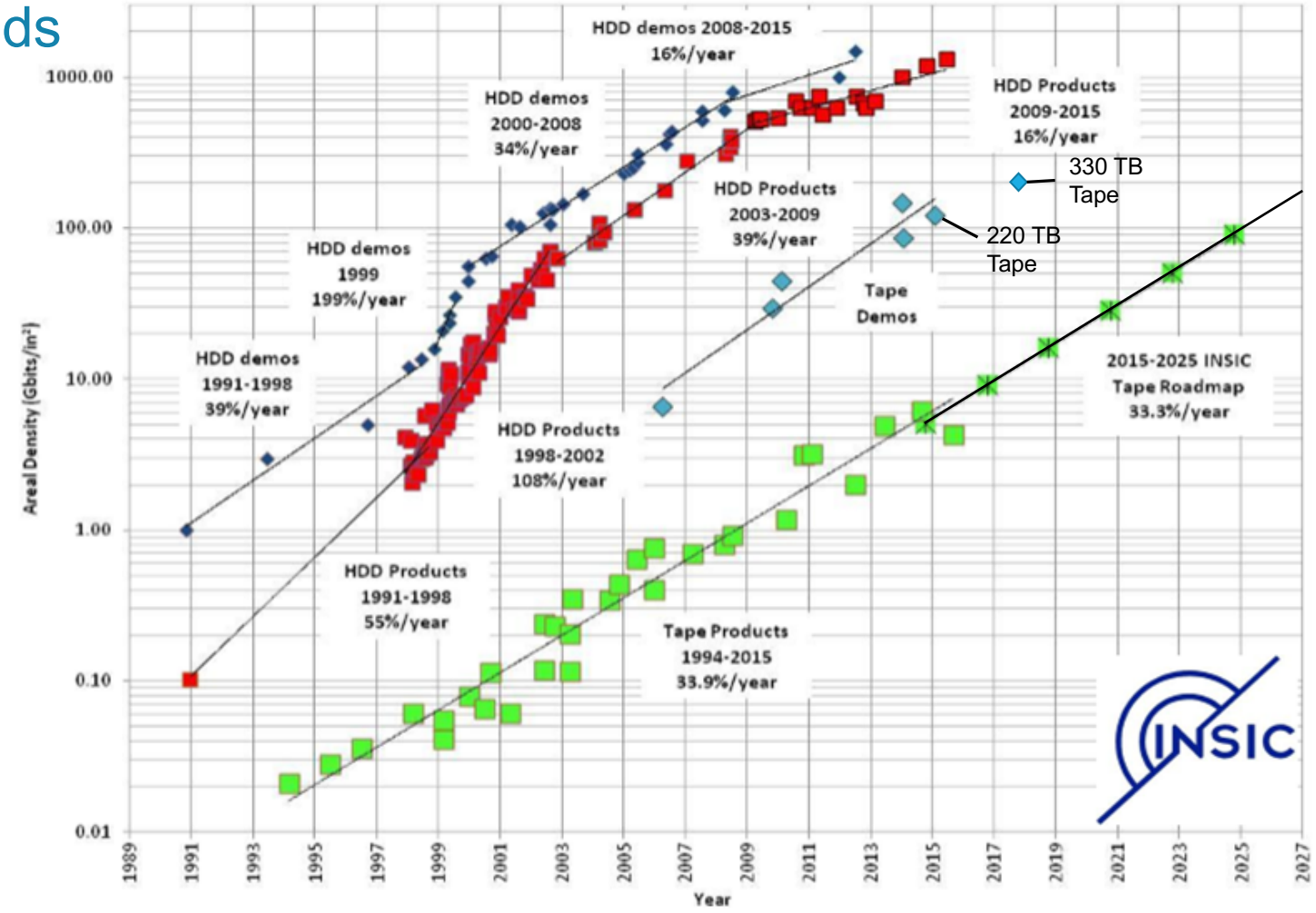
Per Year Cost to Store a Zettabyte

	2015	2020
HDD	$41 Billion	$17 Billion
Optical	$27 Billion	$4 Billion
Tape	$8 Billion	$1.9 Billion

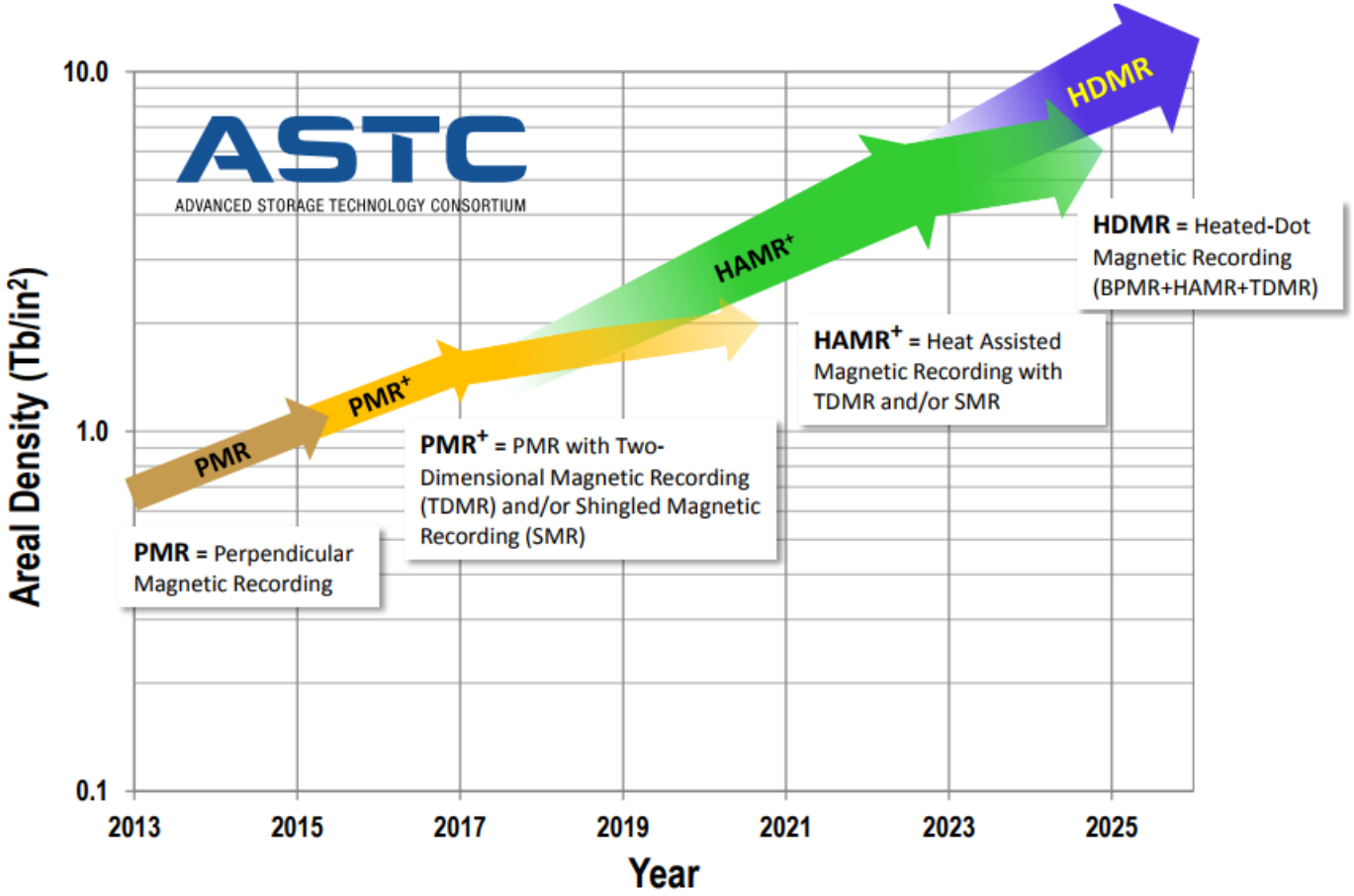
Source: Ogus, A, “How to Store a Zettabyte on a Budget”, TapePower 2015

Areal Density Trends

- Projected areal density increases of 33.3%/yr until 2025
  - Proven record of delivery
- Often adopt technologies from previous HDD generations
- HDD approaching physical limits
  - Tough road map going forward – new technologies



ASTC HDD Technology Roadmap



Independent Scaling of Bandwidth and Capacity

Capacity and Bandwidth
in one package

VS

Bandwidth

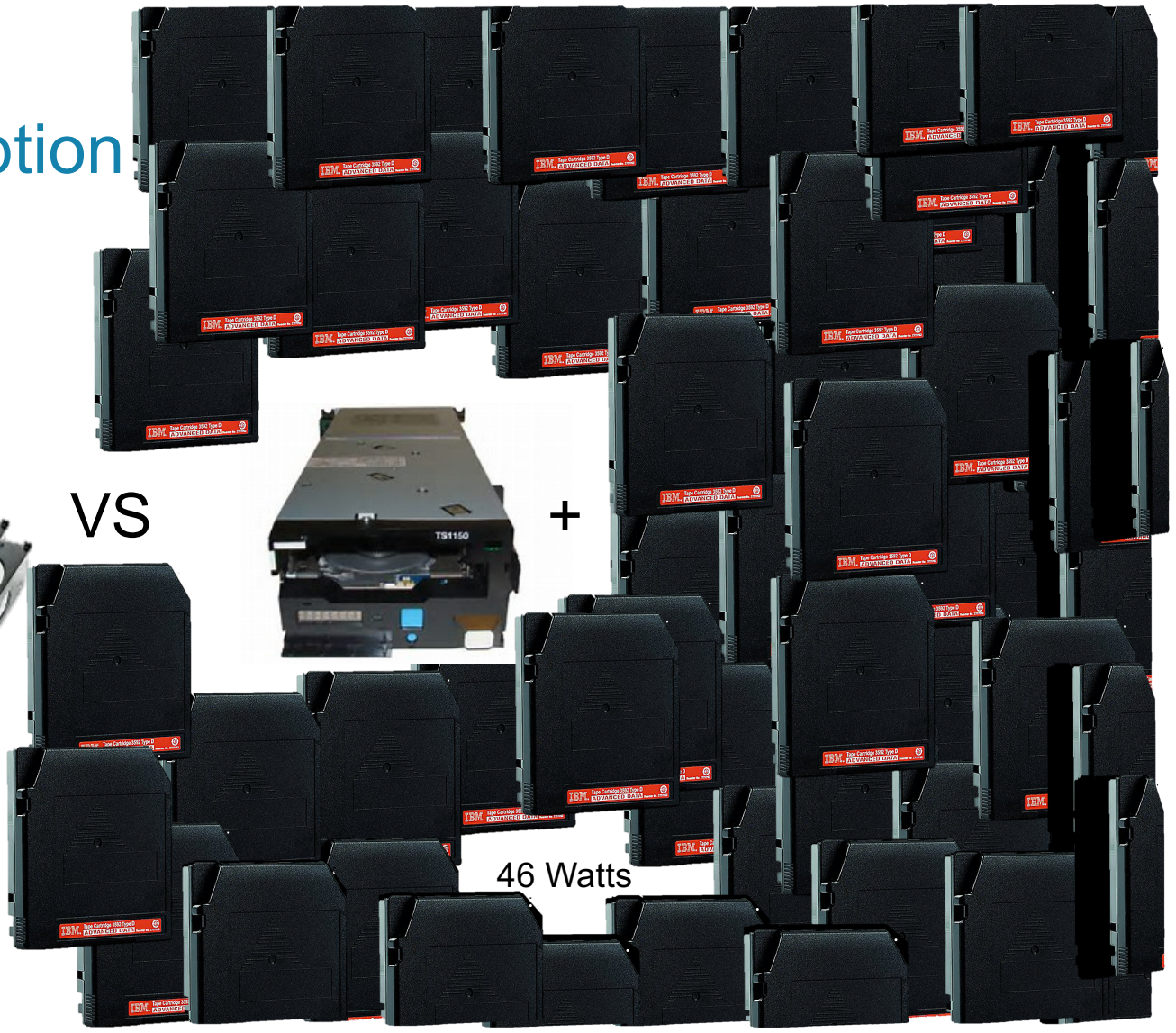
+

Capacity

Power Consumption

40-55 Watts

VS

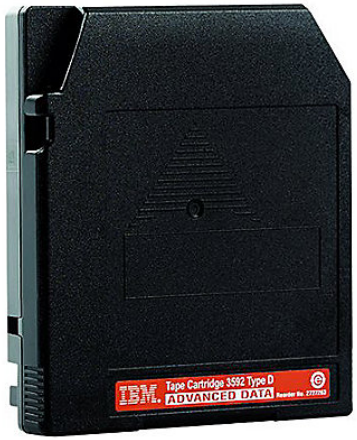


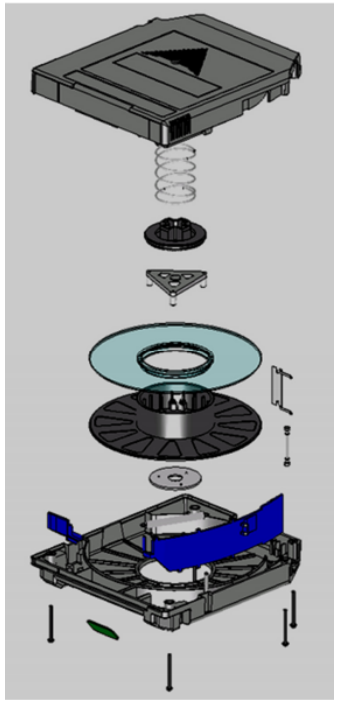
46 Watts

Media Lifetime

3592 Cartridge (JD)

- 30 year storage life*
- 38,400 E2E passes
- 20,000 loads/unloads
- 40,000 short run passes



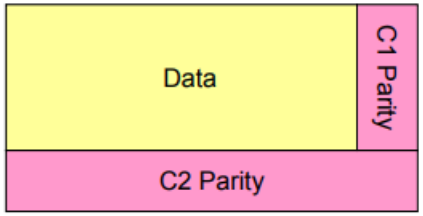


Picture: IBM

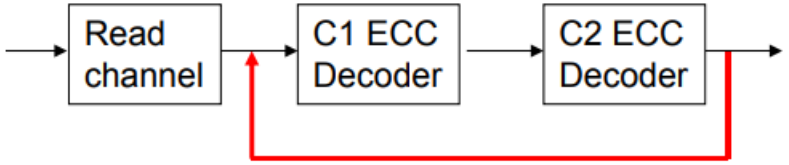
* Under the correct environment

Data Integrity

- Read after write integrity checking
- E2E CRC integrity checking (Logical Block Protection)
- Two level ECC handles both burst (C1) and media errors (C2)
- User Byte Error Rate (UBER) of 10^{-20} (1 in 100EB)



Source: IBM Research - Zurich



Different Attack Vectors

- Offline
- Software Stack
- Firmware

Complications to Large-scale Adoption

of tape in Microsoft datacenters

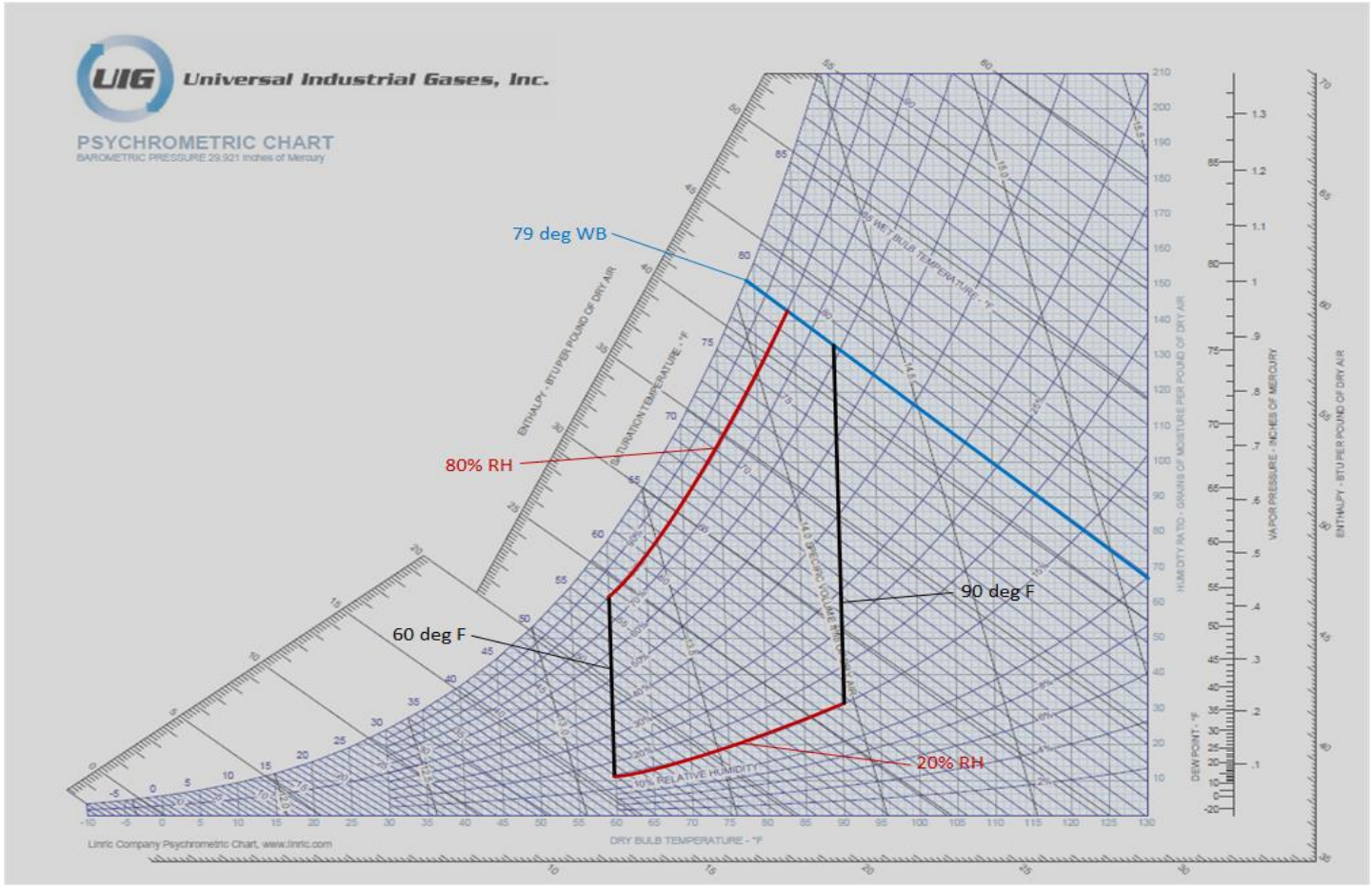
Misalignment with Hyperscale Datacenter Norms

Tape Libraries	Microsoft Datacenters
Fibre Channel interconnect	Ethernet Connected
Low Power	Power provisioning based on racks of servers
Non-standard footprint …and requires special clearances	Space provisioning based on standard rack size
Specific operational and archival environmental envelopes	Adiabatically Cooled

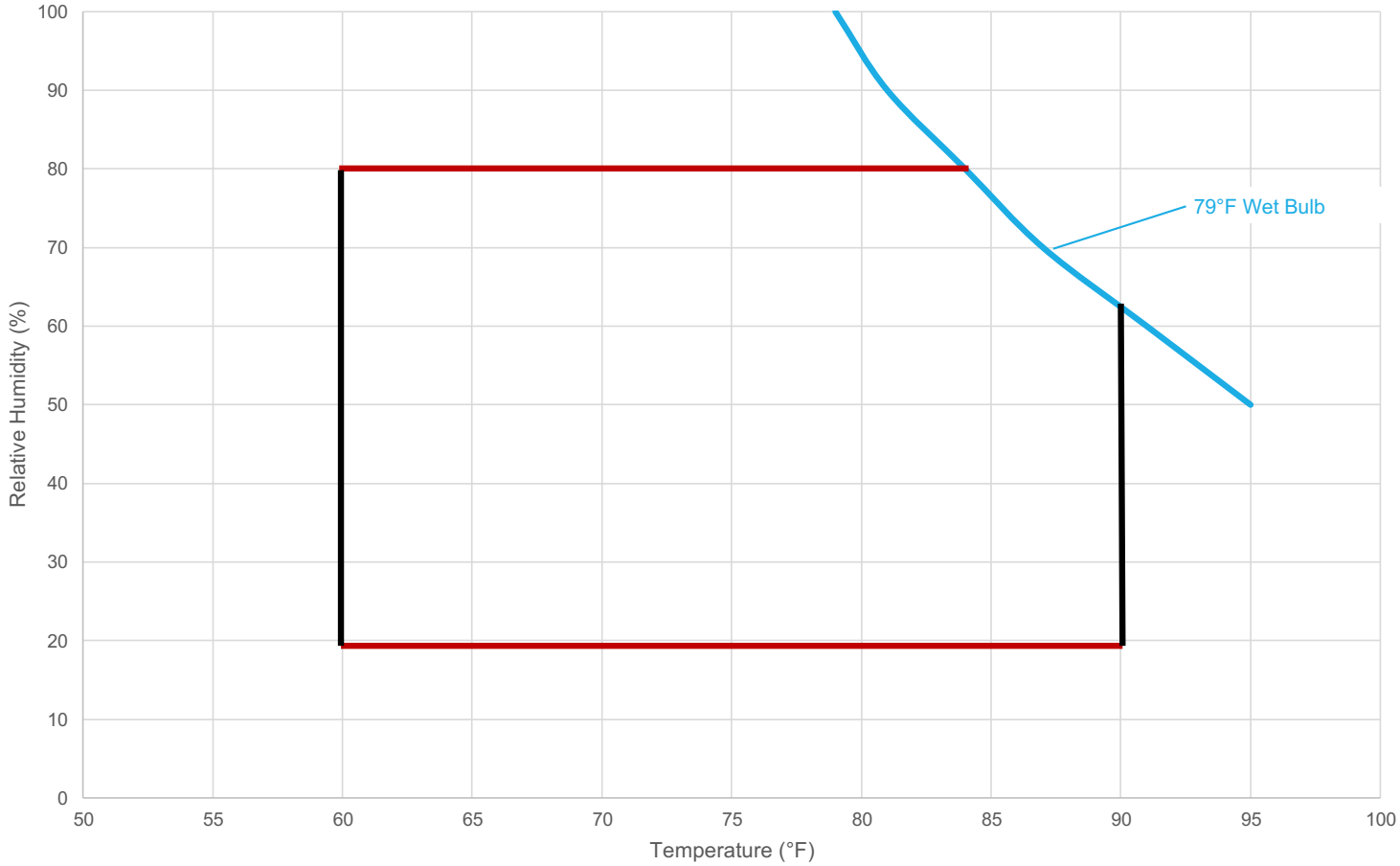
Environmental Requirements

- Current DCs designed for servers and HDDs
  - Adiabatically cooled
- Variation among library and media vendors
  - Temperature ranges
  - Relative Humidity ranges
  - Wet-bulb, Dew-point
  - Rates of change
  - Operational or Archival specifications?

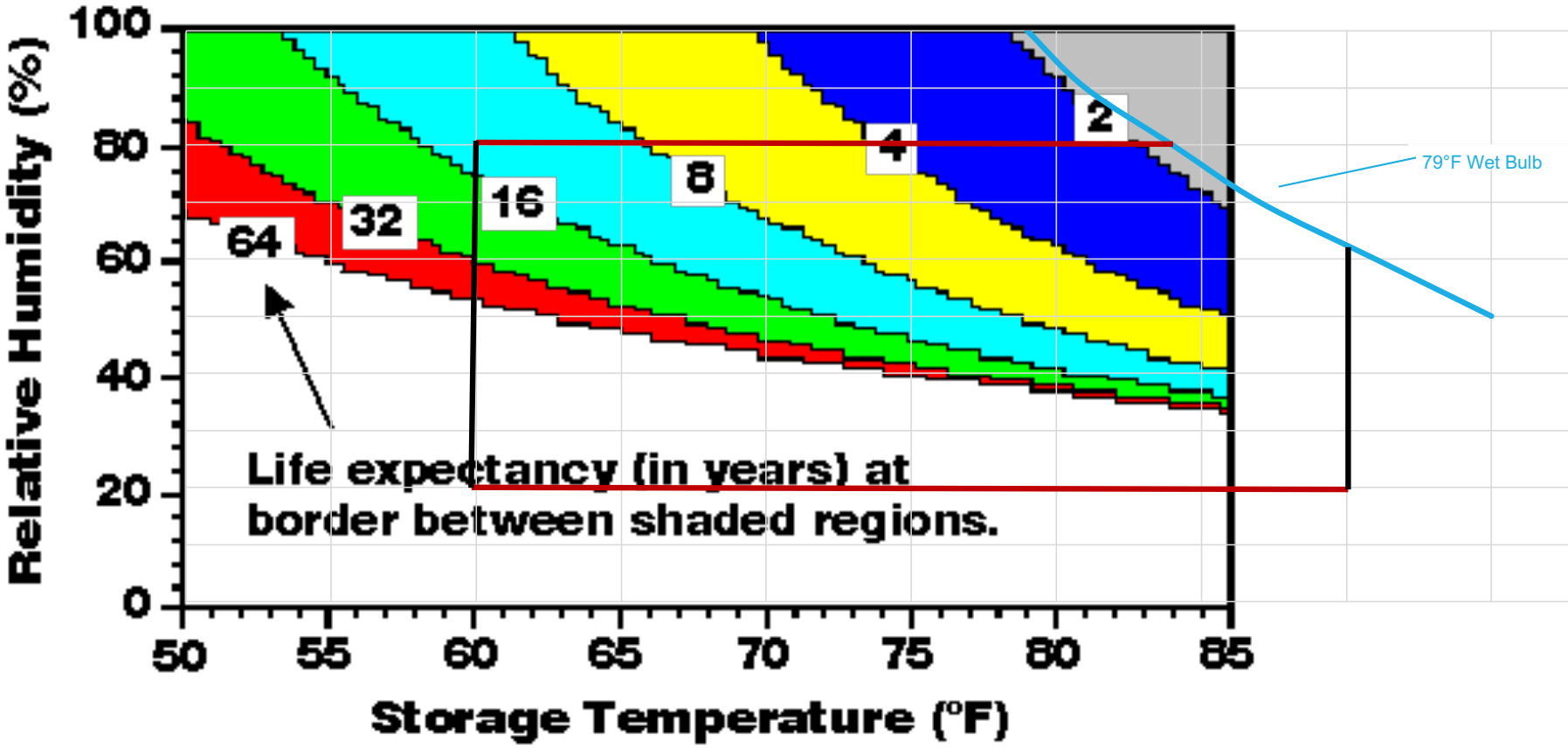
Typical Operational Envelope



Typical Operational Envelope

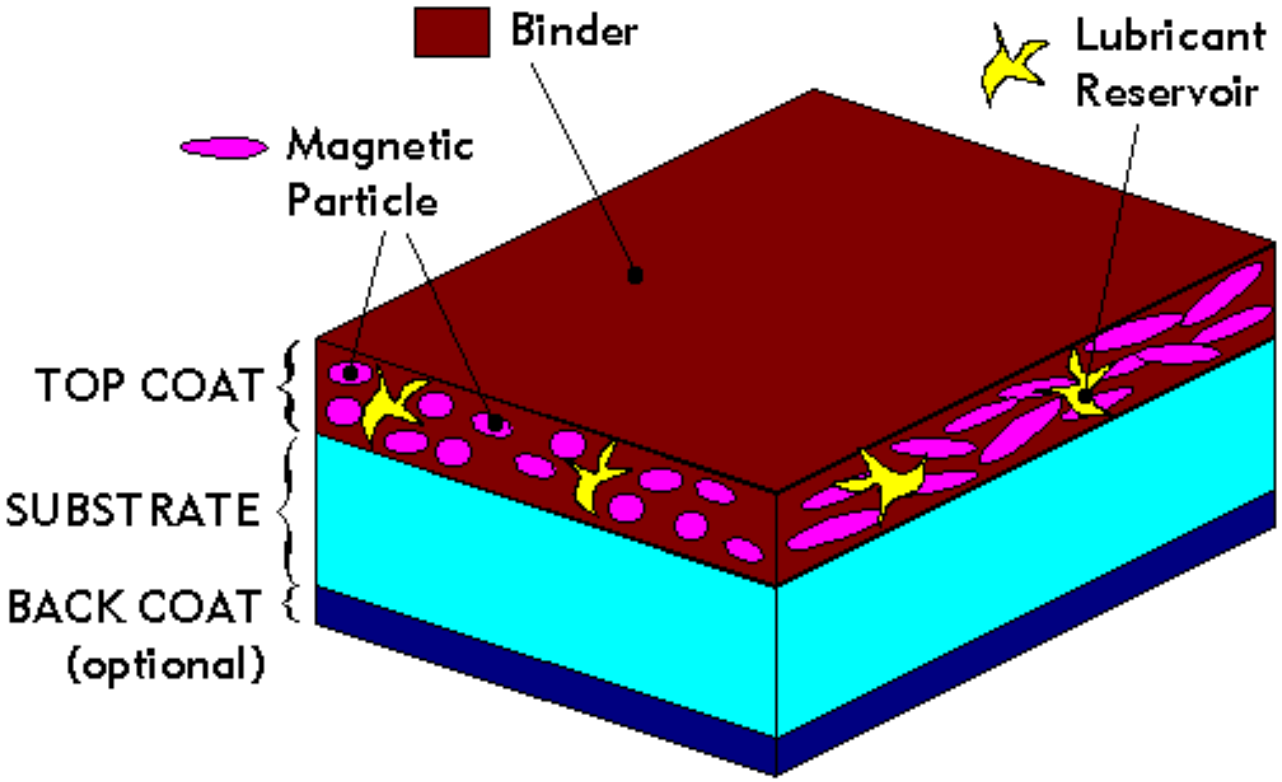


Magnetic Tape Life Expectancies



Source: Bogart, John Van, “Magnetic Tape Storage and Handling: A Guide for Libraries and Archives”, 1995

Magnetic Tape Anatomy



Source: Bogart, John Van, “Magnetic Tape Storage and Handling: A Guide for Libraries and Archives”, 1995

Limitations

- Workload support
  - Ideal workloads are large files (GBs) accessed very infrequently
  - More Readers (drives) helps -- up to a point
    - For smaller reads, accessors become the bottleneck
      - can’t scale number of accessors in most libraries
        - need to deploy smaller libraries
- Decommissioning and Disaster Recovery Time
  - Weeks to months to read all data
  - Larger tapes means this takes even longer if data rates don’t keep up
    - Or need to decrease tape to drive ratio

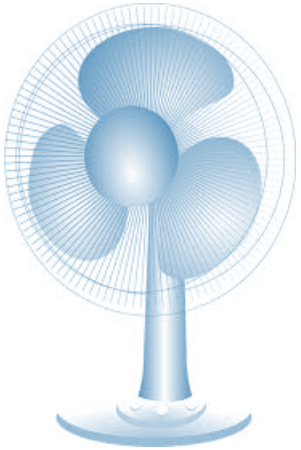
Potential Mitigations

Ethernet Connected

- More widely used, understood, available
- Modern DCs use consolidated infrastructure
- IBM TS1155E
  - 10 Gb Ethernet host attachment interface
  - iSCSI or iSER

Library Maintained Environment

- Self-cooled/conditioned
  - Library internally maintains proper environment
- Halt operations if/when excursions occur
  - Currently, onus is on customer to do the right thing
  - Should apply to non-cooled libraries as well
- MS has requested this feature from library vendors
  - Required in DCs that cannot support operational envelope
    - 90% of our DCs



New Library Designs

- Scale accessors, drives, tapes independently
- Modular design to quickly add bandwidth
  - quick attach drive frames
- Exa-Zettabyte scale
  - Container sized libraries
  - Fitted into purpose-built buildings
  - Rooms of tapes/drives/robots
- RFIs coming

Thank You