Storage Environment Update 2017

Carl Watts
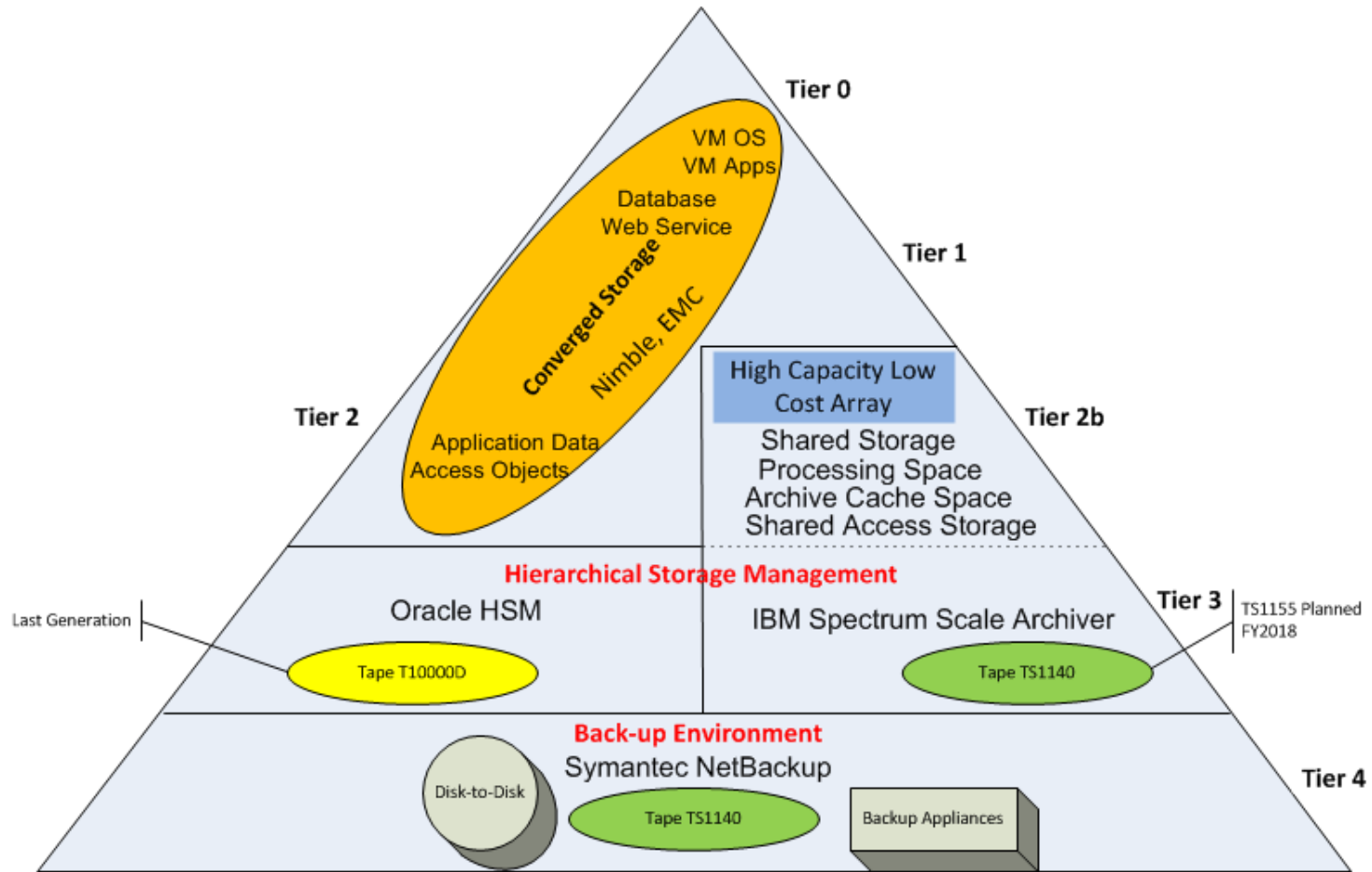
Information Technology Specialist

Enterprise System Engineering – System Development

[cwat at loc dot gov]

The Converging of the Storage Tiers

"Working on the cure for those suffering with digital Disposophobia"

Content Storage

- Content is equal to single copy of a file and it's associated derivatives
- Number of copies:
  - Preservation –
    - minimum of two copies in two different data centers
    - current maximum is two different platforms and media with two copies, but still in the same two data centers
  - Presentation –
    - moving to two online copies across two different data centers

"Working on the cure for those suffering with digital Disposophobia"

Continues Content Growth

▸ Access Content
  ◦ 1.08PB – 2015
  ◦ 1.61PB – 2016 → 203M unique files
  ◦ 2.00PB – July 2017
▸ Preservation Content
  ◦ 8.30PB – 2015
  ◦ 10.96PB – 2016 → 215M unique files
  ◦ 13.49PB – July 2017

"Working on the cure for those suffering with digital Disposophobia"

Move to a Combined POSIX,REST and Longterm Storage Environment

- Spectrum Scale
  - POSIX
    - NFS
    - SMB
  - REST
    - OpenSwift
    - S3
  - Long-term
    - LTFS

"Working on the cure for those suffering with digital Disposophobia"

Continues Migration

"Working on the cure for those suffering with digital Disposophobia"

Migration completed since last year

- Oracle HSM
  - Tape migration T10000K-B to T10000K-D
  - Data restructuring
    - Broke-up one large file system into several smaller file systems
    - Associated specific tapes to each file system to reduce resource contention
- IBM HPSS
  - Full environment migration to Spectrum Scale Archiver

"Working on the cure for those suffering with digital Disposophobia"

Migration Issue Found

- Lesson Learned from migration efforts
  - Issue
    - Over 2,000+ files failed to verify
  - Reviewed
    - Content was verified as good via workflows (post transfer)
    - Both tape copies were the same via OHSM – BAD
  - Possible cause
    - Data path with the disk controller
    - Possible issue with earlier version of OHSM

"Working on the cure for those suffering with digital Disposophobia"

Migration Issue Found

- Lesson Learned from migration efforts (cont.)
  - Resolution
    - Recovered from online and alternate longterm environment
  - Changes needed at workflows layer
    - Fixity check of each copy on long-term device is needed after the write is accomplished
    - All long-term copies need to be verified prior to deletion or removal of source copy

Migration Scheduled

- Content Environment
  - xSTOR environment between data centers (3+PB)
  - Processing environment (160+TB)
  - Transitory environment (50+TB)
- Long-term Storage Environment
  - Spectrum Scale Archive between data centers (970TB)
  - Move OHSM archive between (4.7+PB)

"Working on the cure for those suffering with digital Disposophobia"

Thank you

"Working on the cure for those suffering with digital Disposophobia"