HPSS RAIT

A high performance, resilient, fault-tolerant tape data storage class

Why RAIT?

- HPSS supports striped tape without RAIT
  - Conceptually similar to RAID 0
  - Striped tape allows HPSS to stream data to multiple tape drives for exceptional performance
- Data redundancy is key in an archive
- Two basic methods of redundancy
  - Multiple copies (mirroring)
  - Redundancy through error correcting codes (such as Reed-Solomon)
- Redundancy through error correcting codes is less expensive than making multiple copies

What is HPSS RAIT?

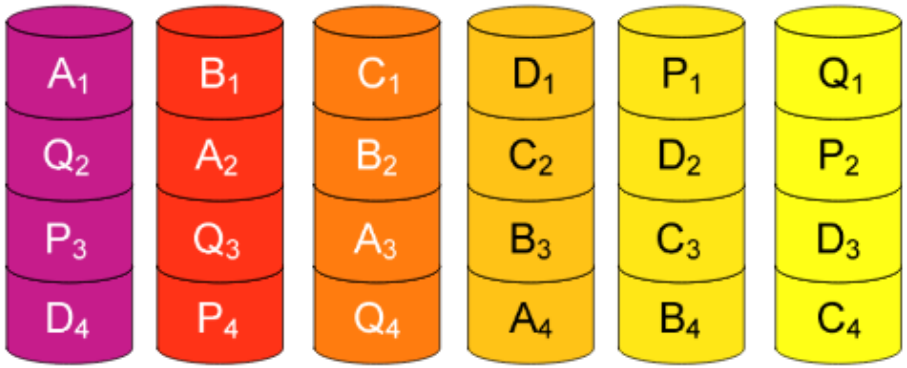
- RAIT (Redundant Array of Independent Tapes) is similar to RAID
- HPSS RAIT is a data storage class that provides a software implementation of RAIT
  - Conceptually similar to RAID 5 or 6
- Uses Reed-Solomon error correcting codes (ECC) to provide data redundancy with lower tape and drive requirements
  - Also known as forward error correction (FEC)
- Provides an advantage over storing redundant copies
  - Less hardware, less floor space, less power, etc.
- Still streams the tape drives at hardware speed
- Stripe widths (data + parity) up to 16

Value of RAIT over Mirroring

- Lower cost fault tolerance
  - Less tape drives and cartridges required
    - Faster mount times than writing a mirrored volume
  - A 8 tape stripe adding single failure redundancy:
    - Mirroring – 8 tapes/drives + 8 tapes/drives = 16 tapes/drives
    - RAIT – 8 tapes/drives + 1 tape/drive = 9 tapes/drives
- Tolerance for tape loss within the stripe
  - Transparent recovery in the event of a tape mount or media error
  - Continue to write to a stripe after the loss of one or more tapes
  - Continue to read from a tape by reconstructing the missing data from the remaining parity
- Read verification
  - Verify the data being read still matches the parity
    - Stripe-level data integrity

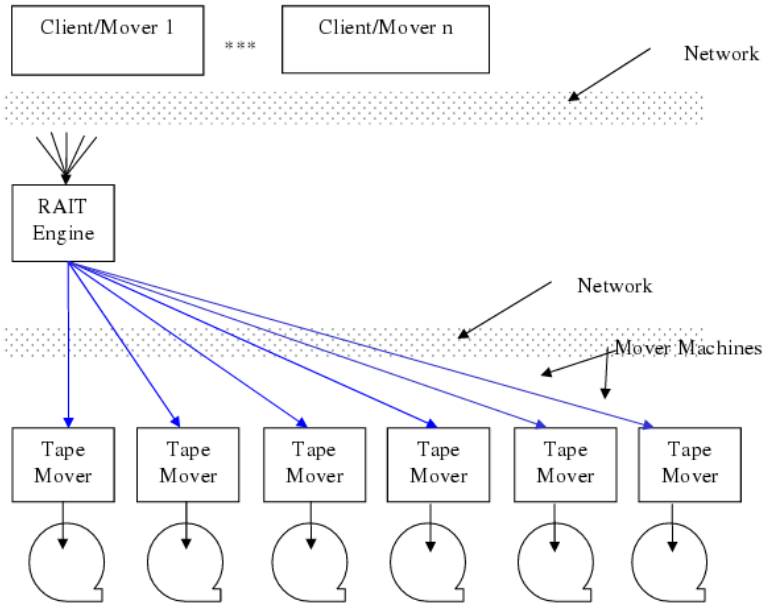
HPSS RAIT Parity Scheme

- Rotational parity
  - Similar to RAID 6
- Rotating the parity evens out the compressibility of the data
  - An equal amount of data are written to each cartridge
- Allows data to be recovered when volumes become unusable
- Parity width ranges from 1 to 7
  - Typical usage is a parity width of 1 to 3, depending on redundancy requirements
  - Parity width governs the number of volume failures that can be tolerated



Concept of Operations

- RAIT is transparent to the end client or application
- When migrating data from disk to tape, a RAIT Engine is spawned to handle RAIT operations
- The RAIT Engine sits between the client and HPSS Mover software
  - Calculates parity blocks using ECC
  - Sends the parity and/or data to the client or Mover as needed
- Multiple RAIT Engines can be configured
  - For larger stripes or for using the CPU resources of multiple systems



Configuration Options

- Supports Data Stripe Width up to 15
- Supports Parity Stripe Width up to 7
- Control the number of parity volumes that may go down before the volume may no longer be written
  - Up to the parity stripe width
- Control whether to verify the data upon reading it
  - Additional validation of the stripe data at the cost of a parity tape mount
- Control whether to mount the entire RAIT volume, or a minimal set of tapes, for reading
  - Lowers the use of tape drives when reading from a RAIT volume
  - For an 8 data, 3 parity volume, this would cause 8 or 9 tapes to be mounted rather than 11 (depending on whether the data was being verified against parity)

Proven in the Field

- NCSA
  - Pre-production testing of ingesting 5 billion files in a single namespace with constant ingest and retrieval performance
  - Ingested 426 TB and retrieved 499 TB of data in 24 hours
  - Using HPSS RAIT in production since 2013
  - Data integrity at 1/5th the cost of data mirroring
  - NCSA Press Release
- ORNL
  - Began using HPSS RAIT in January 2015
  - Faster tape mounts than writing a mirror
  - Deployed a RAIT 4+Parity, with 1GB/s throughput per RAIT volume
  - New RAIT System Keeps Large Data Sets Safe @ ORNL