Tape Storage for Warm Data

Shigeki Misawa
Scientific Data and Computing Center

March 18, 2022

Building a Warm Data Repository

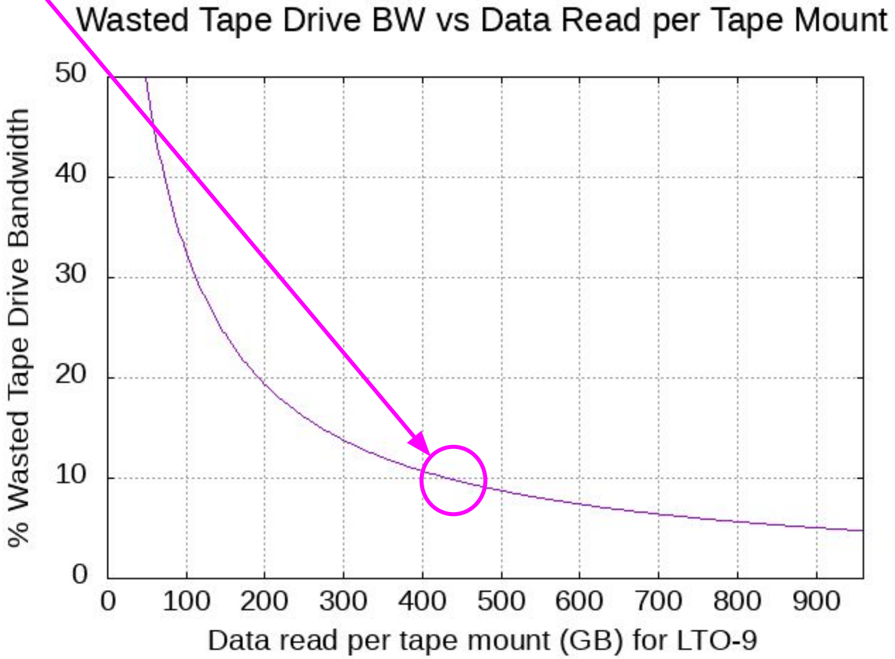
- Increased data volumes are driving the use of tape systems as "warm" archives.
- Intrinsic limitations of tape must be overcome
  - Non-trivial tape media mount times (up to 2 minutes)
  - Long seek times (can be seconds)
- Cost effective, high performance warm data repositories are much harder to design compared to cold data archives
- Performance highly dependent on access patterns

Mitigating Tape Limitations

- Maximize data retrieved per tape mount
  - Read at least 480 GB (for LTO-9) per mount
  - Group files into clusters for read back in a single mount
    - Require identification of file clusters when writing
    - Requires coalescing individual file requests when reading
  - Write files in cluster contiguously on tape, read files in order written
  - Management of clusters of clusters may be need for writes and reads
- Single access pattern needs to dominate
  - Analogy - dB table access either by column or row, not both

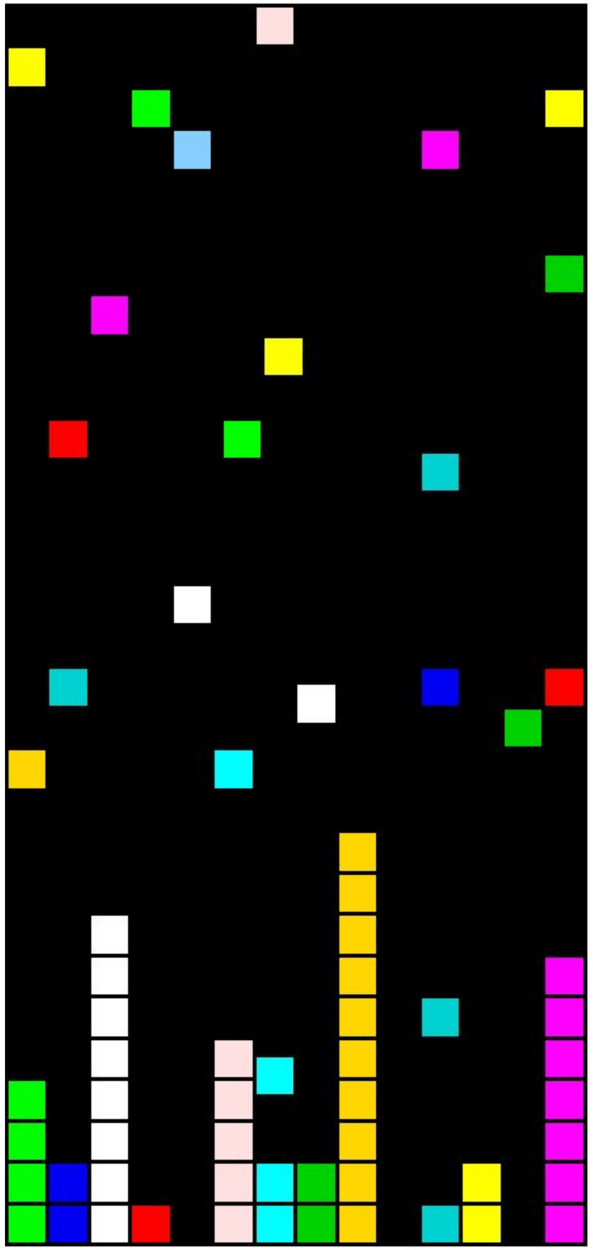
$$F_{Waste} = \frac{1}{1 + T_{Read} / T_{Move}}$$

F_{Waste} = Fraction of wasted BW
T_{Move} = Mount/Seek time
T_{Read} = Time Reading



Real World Data Ingest

- Visualize ingested data as a spreadsheet or dB table
  - Columns = File clusters
  - Cells = Files
- Cells
  - Arriving at random intervals
  - Vary in size
- Columns
  - New columns appear at random intervals
  - Columns vary in length (# files and total size)
  - Last cell in column not identified
  - Time to receive complete columns varies
  - Multiple columns are being written concurrently



Brookhaven National Laboratory

Optimizing Writing in the Real World

- Data source to tape drive data management pipelines
  - Typically works in units of files
  - File clustering information not necessarily captured and if captured, not necessarily communicated to the tape system.
- Optimal use of file clustering information requires:
  - Selected buffering capabilities to stage complete clusters before writing to tape
  - Indication of receipt of complete cluster
- Presence and use of cluster of cluster optimizations lags cluster optimization
- Capabilities and limitations of pipeline components dictates where optimizations can be applied and ease with which they can be developed

Brookhaven National Laboratory

Optimizing Reading in the Real World

- Read optimization requires aggregating file request by tape
  - Simpler than write optimization as required information is local to the tape system
  - Requires requests for files be submitted en mass in a relatively short period of time.
    - In most cases, submission by cluster not directly supported by systems, so cluster request are reconstructed from individual file requests
  - Optimized reading greatly enhanced when coupled with optimized writes.
- Read optimization can affect downstream data consumers
  - Latencies will be higher as more file requests are submitted upfront
  - Re-ordering of file requests may have an impact on downstream systems
    - e.g., data consumer may not be ready (or scheduled, in case of processing jobs) when file becomes available

Summary

- Building a high performance warm archive requires adding capabilities to software systems in various parts of the data management pipeline
- Operation of a robust warm archive with tape involve careful management of the full access (I/O) lifecycle of the data and the data management pipeline
- Participation of data providers and consumers is critical to the optimal operation of the archive

Brookhaven National Laboratory

Tape Storage for Warm Data

Shigeki Misawa
Scientific Data and Computing Center

March 18, 2022