



Hoard

Controlled, Verifiable Data Orchestration

- **Massive Data Replication:** to anywhere Hoard can access and has privileges to read and write
- **Packaging and Staging:** for exporting and transporting data, and filesystem capacity reduction
- **Tape archive management across libraries**
- **Retrieve to anywhere and recover from any Hoard target**
- **Data tagging, tracking, and metrics reporting down to the file level**
- **File integrity verification, metadata tracking, and audit logging**
- **Developed specifically for air-gapped systems (no internet required)**
- **Designed with High-performance and Scientific Computing dataset management in mind**
- **Visibility into everything that has ever happened and where everything has ever been placed.**



Hoard can be used to replicate individual files and entire directory trees to an unlimited number of locations.

- Creates one or more additional copies of data to designated replication targets
- Preserves metadata (permissions, ownership, timestamps)
- Tracks replication state and copy locations in the Hoard database
- Supports incremental operation (only new/changed files are replicated)
- Maintains verifiable file integrity

Benefits

- Protects against primary storage failure
- Reduces risk of data loss during maintenance or outages
- Enables geographic or infrastructure redundancy
- Supports compliance requirements for multi-copy retention
- Improves operational continuity in HPC and enterprise environments

Use Cases

- Protecting research datasets on secondary disk arrays
- Maintaining redundant copies before performing destructive operations
- Replicating data to high-availability storage tiers
- Meeting regulatory requirements for duplicate retention
- Safeguarding active projects prior to archiving
- Data ingest from external devices such as canisters, portable RAIDs, and external drives



Hoard can be used to package and stage data for other activities or reclaim storage capacity.

- Packages files into structured tar archives for efficient handling
- Optimizes large file trees into fewer, manageable objects
- Prepares data for long-term storage tiers (e.g., tape)
- Records file-to-archive mapping in the database
- Preserves metadata and ownership information

Benefits

- Reduces filesystem inode pressure
- Simplifies large-scale dataset handling
- Reduces operational overhead of small files
- Enables deterministic, trackable archive sets

Use Cases

- Packaging data for transport over a WAN and via platforms like Vcinity
- Consolidating millions of small files before archival
- Replicating project datasets into other storage tiers before removing them from performance storage
- Managing storage pressure on high-performance filesystems
- Reducing long-term storage costs



Hoard can use tape libraries to manage data archives.

- Writes staged archives to LTO tape libraries
- Tracks tape pool, barcode, and location metadata
- Verifies writes and records tape placement
- Supports policy-driven archival workflows
- Integrates with changers and multiple tape pools

Benefits

- Data/Storage Administrators can use a single tool to archive, retrieve, and track data across disparate libraries
- Scriptable interface makes it easy to automate and use as a backend for other software and workflows
- Does not use proprietary formats, so tape archives can be shared with others not using Hoard

Use Cases

- Archive distinct types or classes of data to their own logical libraries, while keeping them physically segregated
- Generate non-proprietary tape archives for remote sites (sharing with partners, and cold storage)
- Local tape archives
- Disaster recovery



Hoard can retrieve individual files, entire directory trees, and complete tagged data.

- Locates the best available copy to retrieve (disk, stage archive, or tape)
- Restores files to origin or alternate location
- Reconstructs original directory structure
- Restores ownership and permissions
- Supports verified restore operations

Benefits

- Transparent recovery regardless of storage tier
- Eliminates manual tape search and file tracking
- Preserves operational continuity
- Enables selective file recovery (not just full datasets)
- Provides full audit trail of restore operations

Use Cases

- Recovering deleted files
- Restoring datasets for re-analysis
- Rehydrating archived projects for active use
- Disaster recovery scenarios



Hoard remembers everywhere it has ever put a file and provides tools to query the status of it all.

Hoard may be used to segregate data during ingest, or at any time, by using separate locations (destination targets), while maintaining centralized visibility into status.

Hoard provides full visibility into audible actions

- Always know **Who**, **What**, and **When**!

Hoard provides full visibility into current and past jobs of all types.

- **Status:** list jobs and view their details, including processing stages
- **Control:** cancel, kill, requeue, and rerun jobs
- **Troubleshoot:** built-in diagnostic utilities to identify causes of any problem, and tools to provide information that you may by your storage vendor
- Did your datacenter take a power or cooling hit while Hoard was running replicating, packaging, ingesting, or archiving jobs? No worries, Hoard will continue when it comes back up.

Hoard remembers everything, and it will happily tell you about it in various formats like: columnar, JSON, and CSV. *Great for scripting!*



For more information, feel free to reach out to me directly.

Charlie D Whitehead III

Industry: charlie.whitehead@aptecho.net

Government: charlie@gov.aptecho.com

