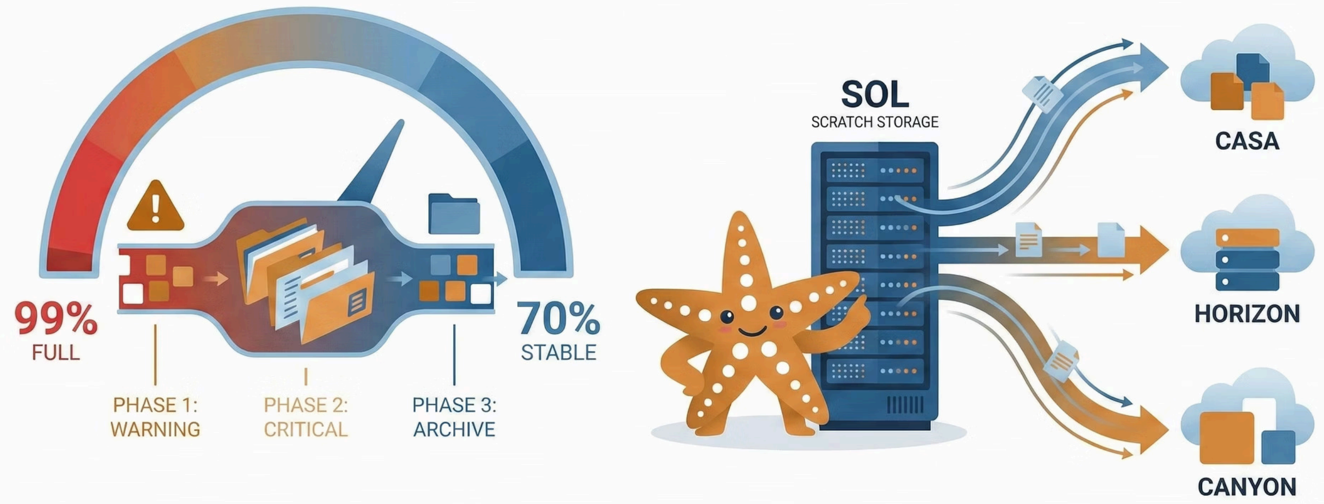
Revolutionizing Scratch Storage Management in High-Performance Computing



Client

Arizona State University (ASU) Research Technology Office (RTO)

Challenge

Chronic misuse of 4PB scratch storage, leading to 99% utilization, system instability, and crashes, combined with failed manual cleanup attempts.

Solution

Starfish-powered automated workflow to identify, warn users, and archive aging data at the directory level before deletion.

Results

Freed 3 PB of data; stabilized scratch utilization at 70%; generated new revenue streams; and fostered a cultural shift in how scratch storage is used.

The Critical Challenge: A System at Breaking Point

Arizona State University's Research Technology Office (RTO) faced a storage crisis common to high-performance computing (HPC) facilities. The scratch storage on their Sol supercomputer was designed as a temporary, high-speed workspace, but researchers were treating the 4-petabyte (PB) system as a de facto long-term archive.

This fundamental misuse led to catastrophic instability. The scratch file system repeatedly filled to **99% capacity**, resulting in system crashes that impacted the workflows of thousands of active researchers.

> "We had quotas, and we had a 90-day retention policy on paper, but in practice we weren't deleting anything," explained **Josh Burks, Senior HPC Systems Analyst at ASU Research Computing**. "The scratch system kept filling, and we were stuck begging users to clean up their data."

Previous manual and script-based attempts to notify users of aging files proved ineffective. The resulting email reports, sometimes containing unwieldy CSV lists with millions of file entries, were ignored or flagged as spam.

The Starfish Solution: Precision Automation

In 2024, the solution was found by turning inward. Josh Burks leveraged Starfish, a data management platform ASU already owned, to create a system that was smarter, faster, and less disruptive than the previous attempts.

The breakthrough used directory tree statistics (known to Starfish users as **recursive aggregate statistics**), which identified **entire directories** of inactive data. This approach proved to be much more workable than identifying individual files.

The Automated, Three-Phase Enforcement Workflow

Josh's team established a weekly automated process powered by direct Starfish database queries. Directories with no file access, modification, or change activity were categorized into three time-based buckets:

Phase 1: Warning
(45 to 75 Days)

Initial alert and advisory.

Phase 2: Critical
(76 to 89 Days)

Heightened notification that action is imminent.

Phase 3: Remove/Archive
(90+ Days)

An automated **archive tag** is applied, flagging the data for removal from scratch storage.

This approach replaced massive file lists with concise, user-specific reports. Emails were sent directly to the users, with a crucial copy sent to their Principal Investigators (PIs), successfully enforcing group-level data responsibility.

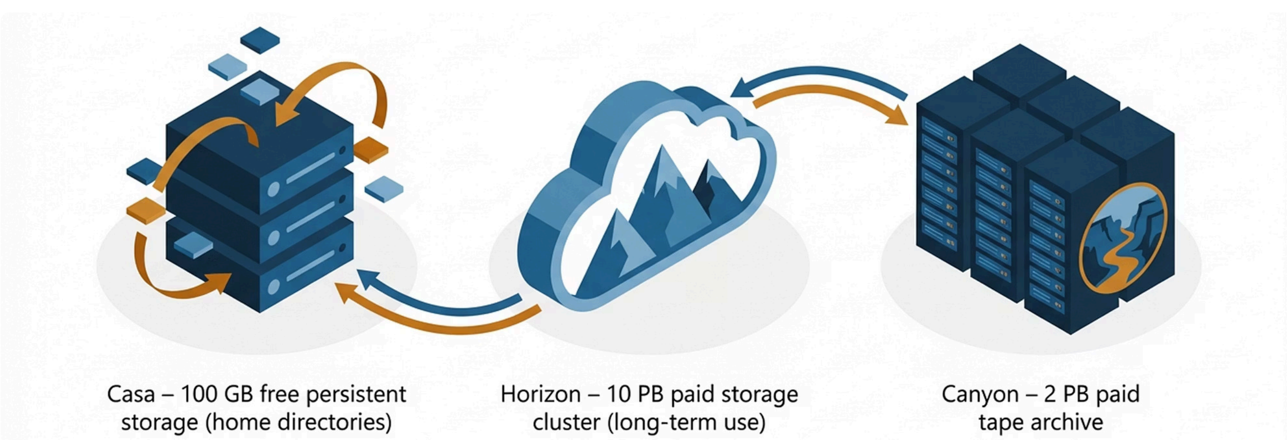
The emails are, as Josh describes them, **"just annoying enough"** to get attention without being ignored, and for the first time, users are now responding and taking action.

Building Trust: Archive First, Delete Later

Recognizing that outright deletion would cause significant researcher backlash, ASU implemented a policy of **"archive first, delete later."** This soft approach was critical to building user trust.

A robust path for data migration was established:

- **Temporary Holding:** An older storage cluster, nicknamed Void, was repurposed as a temporary archive for data tagged for removal.
- **Persistent Storage Options:** Researchers were provided pathways to migrate their data to paid, long-term options, integrating the process with ASU's existing storage chargeback system:



Starfish facilitated the easy, efficient movement of data to these targets, ensuring that researchers received warnings before migration and could effortlessly restore data if needed.

Transformative Results and Cultural Shift

Since implementing the Starfish-powered workflow, ASU has transformed its storage environment and research culture.

Quantitative Outcomes

3 PB

Data Freed

Approximately 3 PB of data were successfully archived or moved from scratch storage, all without a single critical data loss.

70%

System Stability

Scratch file system utilization is now stable, hovering at a healthy **70% capacity**, eliminating previous crashes and service interruptions.

$$$

Revenue Generation

Increased adoption of ASU's paid storage offerings has created new chargeback revenue streams for future infrastructure investment.

Qualitative Outcomes

Faculty and students now understand and adhere to the policy that scratch is temporary storage.

Governance support from Deans and PIs reinforced expectations and policy compliance.

Transparent communication and reversible archiving translated into minimal resistance.

> "Surprisingly, there was very little pushback," Josh noted. "The warnings, the archive step, and the ability to recover data if needed made all the difference."

Conclusion: A Sustainable Blueprint

ASU's experience offers a sustainable, repeatable blueprint for any research computing facility facing scratch storage headaches. By leveraging Starfish to combine automated directory-level identification, targeted communication, and an "archive-first" trust strategy, ASU stabilized its critical HPC resources.

As Josh Burks states: "This approach helped us avoid crashes, prevent data loss, and change how researchers think about storage. **It's a solution any research computing facility facing similar issues can adopt.**"