ATEMPO MIRIA · HPC DATA MANAGEMENT

HPC Data Management

Orchestrating unstructured data in the exascale era

How HPC organizations can take control of unstructured data growth: from Lustre to LTO tape, object storage and cloud.

01 · THE PRESSURE

A data volume that is rewriting the rules

$37.6B	18.3%	+30%/yr
HPC DATA MGMT MARKET	MARKET CAGR	HPC DATA GROWTH
in 2025, projected at $202B by 2035	2025–2035, driven by AI and exascale	for a decade, with no sign of slowing

For a decade, the volume of data stored in scientific research labs and high-performance computing centres has grown at over 30% per year. This trajectory has been driven by a 1,000-fold increase in system performance and a 100-fold increase in system memory. HPC environments now generate unstructured data at a scale that far exceeds what conventional data management tools were designed to handle.

Parallel file systems, Lustre above all, form the backbone of these infrastructures. Lustre runs on over 65% of the world's Top 100 supercomputers, including six of the ten fastest. The Frontier supercomputer at Oak Ridge National Laboratory, the first operational exascale system, runs a 700 PB Lustre filesystem. Research-intensive HPC systems operate at over 95% storage utilisation on average.

A structural shift is compounding this pressure: AI, including model training and large-scale inference, now shares the same infrastructure as classical scientific simulation. Cloud providers have taken note: Google Cloud, Oracle and AWS each offer managed Lustre services at petabyte scale. Unstructured HPC data no longer lives exclusively on-premises; it must flow between local clusters, tape archives, object storage and cloud endpoints.

The challenge is not simply storing this data. It is moving, protecting, archiving and synchronising it across heterogeneous tiers, without stalling production, and while keeping costs under control.

02 · THE ARCHITECTURE

HPC storage tiers: data always in motion

To absorb data growth while managing budgets and serving multiple user communities, HPC organisations structure their storage in tiers. Each tier has a dedicated purpose, distinct hardware characteristics and specific I/O profiles. Data must move between them fluidly and automatically.

- **Buffer / Burst / Scratch** — Ultra-high-performance NVMe or SSD storage where compute executes. Small volumes, extremely high I/O throughput. Data here is transient and must be evacuated quickly to free capacity for the next job.
- **Project / Home Space** — Mid-tier Lustre or GPFS storage, quota-controlled by team. Researchers store active datasets, code and intermediate results here. This tier requires regular protection and granular restore capabilities.
- **Archive / Cold Store** — Long-term storage on LTO tape, S3-compatible object storage or cloud. Many petabytes retained indefinitely. For most HPC compute results, rerunning the calculation is prohibitively expensive. Archiving is the only viable strategy.

The operational complexity lies precisely in managing these flows: moving the right data at the right time to the right tier, without interrupting running calculations, without losing context, and without exhausting primary storage budgets. Universities, for example, bill users on data retention, availability and sharing, all of which require fine-grained orchestration across tiers.

03 · THE OBSTACLE

The problem that general-purpose tools cannot solve

Managing unstructured data on Lustre runs into a fundamental technical obstacle: how do you efficiently identify which files have changed since the last backup or archive job?

On conventional NAS platforms, a change journal answers this question automatically. On Lustre, no such native mechanism exists. Without it, conventional solutions have no choice but to walk the entire filesystem tree, recursively traversing every directory until every file has been enumerated, before any data movement can begin.

This duration routinely exceeds available backup windows. Storage administrators are left with an impossible choice: protect only a portion of the data, or block production for the duration of the scan. For institutions managing hundreds of millions to billions of files, including national laboratories, research universities and major media organisations, this constraint makes complete backup and archiving operationally unworkable with non-specialised tools.

Lustre's distributed architecture compounds the difficulty further: metadata is stored separately from data, spread across MDS and OSS components on different nodes. Even a parallelised tree walk remains a lengthy, resource-intensive process that degrades cluster performance throughout its execution.

> **!** On a Lustre cluster with 500 million files, a full tree walk can take 3 to 4 weeks, just for change detection, before a single byte of data has moved.

04 · THE APPROACH

FastScan: detecting changes without scanning the filesystem

Atempo developed a fundamentally different approach for Miria, called FastScan. Rather than walking the filesystem after the fact, FastScan integrates directly with the internal components of Lustre and GPFS at the node level.

FastScan continuously collects new, modified and deleted files at node level, enriches each entry with its metadata, and consolidates everything into a database that is available to the Miria server as a real-time service call. When a backup or archive job is triggered, Miria queries this database directly: the list of changed files is ready immediately, with no waiting.

WITHOUT FASTSCAN

Full filesystem tree walk. 3 to 4 weeks for 500M files. MDS saturation. Backup window exceeded before any real data movement begins, forcing an impossible choice between protecting part of the data and blocking production.

WITH FASTSCAN (MIRIA)

Changed-file list available in real time. Data movement starts immediately. Multiple Data Movers run in parallel. Full tape speed from the first instant. No impact on operational access to the cluster.

This architecture separates change detection from data movement into two independent streams, a design that allows Miria to reach throughput levels that centralised-agent solutions cannot structurally match. FastScan is available for both Lustre and IBM Spectrum Scale (GPFS), covering the two dominant parallel filesystems in production HPC environments.

05 · THE PLATFORM

Miria: four use cases across the HPC data lifecycle

Unstructured HPC data follows a lifecycle that demands four distinct types of operations. Miria is designed to orchestrate all four from a single platform, with a common engine and a centralised administration interface.

Archiving · offloading primary storage without losing the data

Miria for Archiving moves research data from primary Lustre or GPFS storage to long-term destinations such as LTO tape libraries, on-premises object storage or cloud, saturating tape drives at full speed with a single Data Mover instance. Researchers have access to a self-service web portal to launch their own archive and restore operations, selecting the logical file tree structure of their choice. All operations run over HTTPS, enabling SSH tunnel access for researchers working from anywhere in the world.

Backup · protecting Project and Home Space day to day

Above a certain threshold of roughly 100 TB or one million files, conventional backup approaches become unworkable on Lustre. Snapshots do not guarantee disaster recovery for large datasets. Replication propagates ransomware as fast as it propagates legitimate data. Miria for Backup uses FastScan to drive weekly full backups and daily incrementals, without ever penalising cluster performance, with granular restore capabilities for both day-to-day incidents and full disaster recovery scenarios.

Migration · changing platforms without service interruption

Storage refresh cycles, vendor changes and data centre consolidations require moving hundreds of terabytes to multiple petabytes of unstructured data between heterogeneous platforms. Miria for Migration runs the transfer in the background, maintaining operational access to the source throughout. Supported platforms include Lustre, GPFS/Spectrum Scale, Dell PowerScale/Isilon, Qumulo, StorNext, NFS/CIFS NAS, S3-compatible object storage, LTO tape and public cloud, with automatic format conversion and full metadata preservation.

Synchronisation · keeping large file sets consistent across sites

Miria offers four levels of synchronisation: one-way replication with or without deletion propagation, selective subset replication, and full bidirectional replication. FastScan-powered change detection means sync cycles can run frequently without triggering a full rescan. This is particularly valuable for organisations maintaining live mirrors across geographically distributed sites.

06 · WHAT'S NEW

Native Erasure Coding: resilient archiving with a minimal storage footprint

NEW IN MIRIA NATIVE ERASURE CODING FOR LONG-TERM ARCHIVAL

Long-term data archiving in HPC environments has traditionally forced a stark trade-off: full replication across multiple copies delivers strong resilience but multiplies storage costs proportionally; a single copy minimises footprint but leaves organisations exposed to catastrophic loss. Erasure coding breaks this binary.

Miria now includes native erasure coding capabilities, built directly into the archiving engine. Rather than duplicating entire datasets, erasure coding divides archived data into fragments and computes a configurable number of redundancy shards. Any subset of the original fragments is sufficient to reconstruct the complete dataset, meaning organisations can tolerate the failure of one or more storage nodes, drives or even entire sites, while maintaining a total storage footprint significantly smaller than full replication.

> ! Erasure coding delivers enterprise-grade resilience at a fraction of the storage overhead of full replication: typically 1.25x to 1.5x the original data volume, versus 2x or 3x for mirroring.

Flexible scheme configuration

Miria's erasure coding implementation is fully configurable. Organisations can select the erasure coding scheme that matches their specific resilience requirements and storage budget, and adjust it as their infrastructure or risk tolerance evolves.

PROFILE	EC SCHEME	OVERHEAD	SITES	BEST SUITED FOR
Efficiency-first	8+2	+25%	1-2	Maximum storage efficiency. Tolerates 2 simultaneous failures. Ideal for on-prem tape or object archival with tight capacity budgets.
Balanced resilience	6+3	+50%	1-3	Strong fault tolerance with moderate overhead. Suitable for multi-node or multi-rack archive pools.
High resilience	4+2	+50%	2-3	Compact scheme, tolerates 2 failures across fewer fragments. Good fit for geo-distributed small-to-mid archives.
Multi-site DR	6+2+2	+33%	3+	Fragments distributed across 3 or more geographically separated sites. Full site loss tolerance with low overhead.
Full replication (ref.)	3× copy	+200%	1-3	Traditional mirroring for reference. High resilience, very high cost. Shown here for comparison only.

Multi-site distribution

Miria's erasure coding engine is designed to distribute fragments across multiple storage locations, whether on-premises storage nodes, separate data centres or cloud endpoints. This means a single archive job can simultaneously write erasure-coded fragments to geographically separated sites, providing site-level disaster recovery without the cost of full duplication.

The number of sites and the distribution of fragments across them are configurable per archive policy. An organisation running a 6+3 scheme across three sites, for example, retains full data recoverability even in the event of a complete site outage, using only 50% more storage than the original dataset. The scheme can be adjusted as infrastructure evolves, with no need to re-architect the archiving pipeline.

> "Erasure coding in Miria gives us the ability to define exactly the level of resilience we need for each data tier, without paying the penalty of full replication. It is a fundamental shift in how we think about long-term archive economics."

07 · THE CASE FOR PURPOSE-BUILT

Why a purpose-built platform makes the difference

Managing unstructured data at HPC scale raises challenges that general-purpose data protection tools were not designed to address. Deep integration with parallel filesystems, the ability to operate across billions of files without a scan window, support for the full diversity of HPC storage protocols and destinations: each of these demands a solution built specifically for HPC environments.

01 **Native Lustre and GPFS integration**

FastScan integrates with internal MDS/OSS components, not through an NFS mount. The list of changed files is available in real time, with no filesystem traversal required.

02 **Full tape speed, linearly scalable throughput**

A single Data Mover saturates an LTO tape drive. Multiple Data Movers running in parallel scale archive and backup throughput proportionally, with no architectural ceiling.

03 **Native erasure coding for resilient, efficient archiving**

Configurable EC schemes and multi-site distribution deliver enterprise-grade resilience at a fraction of full-replication storage costs, with the flexibility to adjust the scheme as infrastructure evolves.

04 **One platform across four use cases**

Archive, backup, migration and synchronisation in a single engine. A coherent pipeline and a unified interface, with no assembly of third-party products required.

05 **Full heterogeneity, zero vendor lock-in**

Lustre, GPFS, NAS, S3 object storage, LTO tape, public cloud. Miria is the stable orchestration layer above the infrastructure, regardless of how it evolves.

ABOUT ATEMPO

Independent European data movement, for over 25 years.

Atempo is an independent European software vendor with over 25 years of experience in data protection and movement. Its three flagship solutions Lina, Miria and Tina are deployed across thousands of organisations worldwide, and headquartered in Paris and present in Europe, the United States and Asia through a network of over 100 integrators and managed service providers.

- LINA
- MIRIA
- TINA