



Reducing Enterprise Storage Costs Through Data Minimization

How Lightbeam helps organizations reduce unnecessary storage consumption across file shares, cloud object storage, and cloud infrastructure.

WHITEPAPER

Table of Contents

Executive Summary	02
The Cost of Unnecessary Enterprise Data	04
How Lightbeam Enables Data Minimization	05
Recover Capacity on SMB and Network File Shares and Reduce Amazon S3 Consumption	06
Identify Orphaned Amazon EBS Volumes and Modeling Assumptions and Supporting Research	07
Data-Deletion Assumption: 28% and Fully Loaded Annual Storage Cost: \$500 per TB	08
Annual Enterprise Data Growth: 25% and Why the Model Uses These Assumptions Together	09
Illustrative Financial Scenarios and Scenario 1: 1 PB Environment	10
Scenario 2: 5 PB Environment and Scenario 3: 20 PB Environment	11
Summary of Modeled Benefits and Delayed Infrastructure Expansion	12
Reduced Discovery and Compliance Costs	13
Conclusion	14

Executive Summary

Enterprise data environments accumulate unnecessary information over time. Employees create multiple copies of documents, cloud migrations leave source data behind, completed projects retain working files indefinitely, former employees leave unowned content, and decommissioned cloud workloads can leave storage resources active long after they are needed.

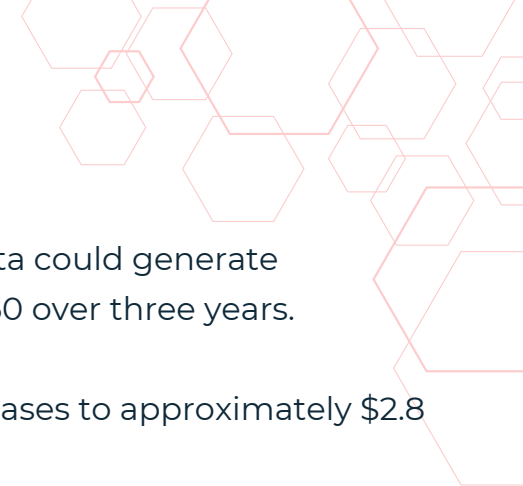
This accumulation creates continuing costs across network file shares, cloud object storage, block storage, backups, replication, infrastructure support, and administration. It also increases the amount of data organizations must secure, classify, govern, retain, and potentially produce during litigation or regulatory inquiries.

Lightbeam helps organizations discover and classify enterprise data, connect it to owners and identities, understand its business and regulatory context, and identify information that can be deleted or otherwise minimized. Opportunities include duplicate and redundant files, obsolete content on SMB shares, unnecessary Amazon S3 objects, orphaned Amazon EBS volumes, stale snapshots, and data retained beyond its legitimate business purpose.

The financial examples in this paper use the assumptions contained in the Lightbeam Data Minimization ROI model:

- 28% reduction in stored data
- \$500 fully loaded annual storage cost per TB
- 25% annual growth in enterprise data
- Three-year analysis period

These assumptions are supported by research from Veritas and IDC, corroborating industry research, public cloud-storage pricing, and enterprise storage total-cost-of-ownership practices.



Under these assumptions, an organization managing 1 PB of data could generate approximately \$140,000 in first-year storage savings and \$533,750 over three years.

For an organization managing 20 PB, the modeled benefit increases to approximately \$2.8 million in the first year and \$10.7 million over three years.

These examples focus only on storage-related economic benefits. Even so, the storage cost savings associated with data minimization alone can more than justify the investment in Lightbeam.

This analysis does not quantify the additional value of lower backup and replication consumption, delayed infrastructure expansion, reduced migration requirements, administrative efficiency, reduced data-security exposure and reduction in data breach risk, or improved regulatory compliance.

The Cost of Unnecessary Enterprise Data

Enterprise data is frequently copied, moved, exported, backed up, and replicated without a corresponding process for determining when the resulting copies are no longer required.

Common sources of unnecessary storage consumption include:

- Duplicate documents stored by multiple users or departments
- Data copied during migrations but not removed from the original system
- Multiple versions of substantially similar content
- Files associated with former employees
- Completed project data retained without an ongoing purpose
- Temporary reports, application exports, and analytical datasets
- Amazon S3 objects that are no longer actively used
- Unattached Amazon EBS volumes and stale snapshots
- Storage associated with abandoned development and testing environments
- Data that has exceeded an approved retention period
- Sensitive information retained without a continuing business need

This type of information is often described as redundant, obsolete, or trivial data, commonly abbreviated as ROT data. Veritas research found that 33% of stored enterprise data was known ROT data, while later Veritas material cited a comparable 28% benchmark.^[1]

Unnecessary data creates more than a capacity problem. Every additional copy can increase backup volumes, replication requirements, migration effort, administrative overhead, discovery costs, and the amount of sensitive information exposed during a security incident.

How Lightbeam Enables Data Minimization

Lightbeam provides the discovery, classification, identity, ownership, and business context required to make informed decisions about enterprise data.

Rather than treating every old or inactive file as a deletion candidate, Lightbeam helps organizations answer questions such as:

- What does the data contain, and is it sensitive or regulated?
- Who owns it, and which users, groups, applications, or systems can access it?
- Is it a duplicate of information stored elsewhere?
- When was it last used?
- Is it subject to a retention requirement or legal hold?
- Does it still support an active business process?
- Can it be deleted, archived, or moved to a lower-cost tier?

Lightbeam Smart Templates can identify specific context within documents and use that context to trigger retention or deletion policies. For example, a policy can be based on a contract end date, an application date, the age of a data subject, or another business event contained within the data itself. This enables organizations to move beyond broad rules based only on file age and apply more precise, context-aware minimization policies.

This context enables security, privacy, infrastructure, cloud, data-governance, and business teams to minimize data through a governed and auditable process.

Recover Capacity on SMB and Network File Shares

SMB shares and other network file systems commonly contain years of accumulated departmental and user-generated data. Folder structures may no longer reflect the organization, ownership may be unclear, and files may remain long after the projects or employees associated with them are gone.

Lightbeam can help identify exact and redundant file copies, stale departmental folders, data owned by inactive identities, files that have not been accessed within an approved period, data that has exceeded its retention period, and sensitive information stored without an active business requirement.

Removing unnecessary data can reclaim usable capacity and delay additional purchases or infrastructure expansion. It can also reduce the quantity of information included in backups, replication, migrations, security reviews, and disaster-recovery processes.

Reduce Amazon S3 Consumption

Amazon S3 is billed in part according to the amount and storage class of data retained. For S3 Standard in the U.S. East region, published pricing declines to \$0.021 per GB per month for storage above 500 TB. Removing unnecessary objects can therefore reduce ongoing storage charges when those objects are no longer needed.^[2]

Potential S3 minimization opportunities include duplicate objects across buckets, old application exports, data copied during cloud migrations, abandoned analytical datasets, development and testing data no longer in use, and objects retained after their business purpose has ended.

The financial benefit depends on storage class and contractual pricing. Removing a terabyte from frequently accessed storage generally creates a larger direct benefit than removing the same volume from deep archive storage.

Identify Orphaned Amazon EBS Volumes

Amazon EBS volumes can continue to incur charges after their associated EC2 instances, applications, or projects have been decommissioned.

Potential opportunities include EBS volumes left after EC2 instances are terminated, storage associated with abandoned applications, volumes without an identifiable business owner, oversized or underutilized volumes, and snapshots retained beyond approved lifecycle requirements.

Lightbeam helps connect these resources to owners, identities, applications, sensitivity, and business context so the organization can determine whether they remain necessary.

Modeling Assumptions and Supporting Research

The financial scenarios in this paper are based on three inputs from the Lightbeam Data Minimization ROI model. These assumptions provide a consistent, high-level method for illustrating the potential economic value of data minimization.

Assumption	Modeled value
Data deletion rate	28%
Fully loaded annual storage cost	\$500 per TB
Annual enterprise data growth	25%

Data-Deletion Assumption: 28%

Lightbeam uses a 28% data-deletion assumption based primarily on Veritas research concerning redundant, obsolete, and trivial enterprise data. Veritas' Global Databerg research found that 33% of enterprise data was known ROT data, and later Veritas material cited approximately 28% ROT data.^[1]

The model uses 28% because it is the lower, more conservative point within the Veritas benchmark range and is tied specifically to redundant, obsolete, and trivial information. Other industry research has also found substantial levels of duplicate, rarely used, and unused data, reinforcing that the problem remains material.

Lightbeam's opportunity is broader than traditional storage-system deduplication. The platform can help identify not only identical files but also stale, obsolete, unowned, expired, or unnecessary information across repositories and cloud resources.

Fully Loaded Annual Storage Cost: \$500 per TB

Enterprise storage costs extend beyond the price of raw disk or cloud capacity. A fully loaded storage-cost model can include primary capacity, backups, replication, disaster-recovery copies, snapshots, hardware and software support, storage administration, monitoring, power, cooling, data-center space, reserved capacity, and infrastructure refresh.

Public-cloud object storage provides a transparent baseline for raw capacity. At volumes above 500 TB, Amazon S3 Standard is priced at \$0.021 per GB per month, equivalent to approximately \$252 per TB annually before backup, replication, administration, support, and other operating expenses.^[2]

Lightbeam uses an illustrative fully loaded annual cost of \$500 per TB, approximately twice the raw annual cost of high-volume standard cloud object storage. The additional amount represents the broader cost of backups, replication, support, infrastructure, administration, capacity overhead, and future expansion.

The value created by reclaiming storage can appear as direct cloud-cost reduction, avoided future capacity purchases, or capacity reclaimed for another business purpose.

Annual Enterprise Data Growth: 25%

Lightbeam uses a 25% annual data-growth assumption to represent the continuing expansion of enterprise data environments. IDC forecasts cited by Micron indicate that the volume of data created each year will increase at a compound annual growth rate of 24.4% from 2023 through 2028. Lightbeam rounds this forecast to 25% for the model.^[3]

The forecast is consistent with growth driven by cloud adoption, collaboration and file-sharing platforms, application telemetry, analytics and AI workloads, IoT and edge systems, regulatory retention, and copies created through backup, replication, and migration.

The model assumes that the economic opportunity associated with unnecessary data grows with the overall environment. Year 2 savings therefore increase by 25%, and Year 3 savings increase by an additional 25%.

Why the Model Uses These Assumptions Together

Veritas provides the primary basis for the percentage of data addressed: 28% ROT data. Public cloud pricing establishes a raw storage-cost baseline: approximately \$250 per TB annually for high-volume, frequently accessed object storage.

Enterprise TCO practices support including broader operating costs such as backup, replication, administration, infrastructure, and support.

IDC provides the basis for annual growth: approximately 24.4%, rounded to 25%.

Illustrative Financial Scenarios

The following examples use the assumptions contained in the Lightbeam Data Minimization ROI model. They focus exclusively on gross storage-related benefits and do not include Lightbeam licensing costs, staffing costs, backup savings, replication savings, administrative savings, risk reduction, or other economic benefits.

Modeling assumption	Value
Data-deletion rate	28%
Fully loaded annual storage cost	\$500 per TB
Annual data growth	25%
Analysis period	3 years
Petabyte conversion	1 PB = 1,000 TB

Scenario 1: 1 PB Environment

An organization manages 1 PB, or 1,000 TB, of enterprise data. Applying the modeled 28% deletion rate results in 280 TB of storage removed or avoided in the first year.

Measure	Year 1	Year 2	Year 3
Modeled storage environment	1,000 TB	1,250 TB	1,562.5 TB
Data reduced at 28%	280 TB	350 TB	437.5 TB
Annual storage benefit	\$140,000	\$175,000	\$218,750

Three-year gross storage benefit: \$533,750

Scenario 2: 5 PB Environment

An organization manages 5 PB, or 5,000 TB, across network file shares, cloud object storage, block storage, and enterprise applications. Applying the modeled 28% deletion rate results in 1,400 TB, or 1.4 PB, of storage removed or avoided in the first year.

Measure	Year 1	Year 2	Year 3
Modeled storage environment	1,000 TB	1,250 TB	1,562.5 TB
Data reduced at 28%	280 TB	350 TB	437.5 TB
Annual storage benefit	\$140,000	\$175,000	\$218,750

Three-year gross storage benefit: \$2,668,750

Scenario 3: 20 PB Environment

A large enterprise manages 20 PB, or 20,000 TB, across file shares, cloud object storage, block storage, backup systems, and enterprise applications. Applying the modeled 28% deletion rate results in 5,600 TB, or 5.6 PB, of storage removed or avoided in the first year.

Measure	Year 1	Year 2	Year 3
Modeled storage environment	20,000 TB	25,000 TB	31,250 TB
Data reduced at 28%	5,600 TB	7,000 TB	8,750 TB
Annual storage benefit	\$2,800,000	\$3,500,000	\$4,375,000

Three-year gross storage benefit: \$10,675,000

Summary of Modeled Benefits

Initial environment	First-year data reduced	Year 1	Year 2	Year 3	3-year benefit
1 PB	280 TB	\$140,000	\$175,000	\$218,750	\$533,750
5 PB	1,400 TB	\$700,000	\$875,000	\$1,093,750	\$2,668,750
20 PB	5,600 TB	\$2,800,000	\$3,500,000	\$4,375,000	\$10,675,000

Additional Benefits Beyond Primary Storage

Reduced Backup and Disaster-Recovery Consumption

An organization manages 5 PB, or 5,000 TB, across network file shares, cloud object storage, block storage, and enterprise applications. Applying the modeled 28% deletion rate results in 1,400 TB, or 1.4 PB, of storage removed or avoided in the first year.

Delayed Infrastructure Expansion

Reclaiming capacity can delay storage hardware purchases, expansion of cloud commitments, data-center upgrades, and infrastructure-refresh projects.

Lower Migration Costs

Removing unnecessary information before a cloud migration, application consolidation, or data-center transition can reduce project duration, bandwidth requirements, and professional-services costs.

Reduced Security Exposure

Every unnecessary copy of sensitive information expands the potential attack surface. Data minimization reduces both the probability of a breach and its potential blast radius by limiting the volume of sensitive data available to compromised users, malicious insiders, ransomware, and unauthorized AI tools or agents. According to IBM's 2025 Cost of a Data Breach Report, researched by the Ponemon Institute, the average cost of a data breach in the United States reached \$10.22 million.^[4]

Reduced Discovery and Compliance Costs

Redundant and obsolete information can increase the volume of data that must be searched during litigation, investigations, audits, privacy requests, and regulatory reviews.

Improved AI and Analytics Readiness

Eliminating redundant, stale, and low-value content can improve data quality and reduce the likelihood that AI systems retrieve outdated or inappropriate information.

A Governed Approach to Deletion

Data minimization should not be treated as an indiscriminate cleanup exercise. An old, inactive, or duplicate file may still be subject to a retention schedule, legal hold, regulatory requirement, contractual obligation, or legitimate business need.

A defensible data-minimization process should include:

- Discovering and classifying the data.
- Identifying duplicate, stale, obsolete, and orphaned information.
- Connecting the data to owners, identities, systems, and business processes.
- Applying retention, privacy, legal-hold, and security policies.
- Routing proposed actions to the appropriate owners.
- Documenting approvals and exceptions.
- Deleting, archiving, or moving approved data.
- Continuously repeating the process as the environment grows.

Lightbeam enables this governed approach by providing a continuously updated understanding of enterprise data, its sensitivity, ownership, access, usage, and business context.

Conclusion

Enterprise storage costs are driven not only by the data an organization actively uses, but also by the unnecessary information it continues to retain.

Lightbeam helps organizations identify and address duplicate files, stale SMB content, unnecessary S3 objects, orphaned EBS volumes, expired data, and information that no longer has a legitimate business purpose. By connecting data to identity, ownership, sensitivity, usage, and retention context, Lightbeam helps enterprises translate data visibility into defensible minimization actions.

Using the assumptions in the Lightbeam model, a 28% reduction in stored data can produce substantial gross storage benefits:

- \$533,750 over three years for a 1 PB environment
- \$2,668,750 over three years for a 5 PB environment
- \$10,675,000 over three years for a 20 PB environment

The value grows as enterprise data volumes grow, making data minimization not only a one-time storage optimization initiative but an ongoing discipline for controlling data costs, reducing risk, and improving the quality of enterprise information.

Sources

- 1.Veritas Global Databerg Report: Veritas reported that 33% of stored data was redundant, obsolete, or trivial. Later Veritas materials cited a comparable 28% ROT benchmark. <https://www.veritas.com/news-releases/2016-03-15-veritas-global-databerg-report-finds-85-percent-of-stored-data>
- 2.Amazon S3 Pricing: AWS lists S3 Standard storage above 500 TB per month at \$0.021 per GB in U.S. East pricing. <https://aws.amazon.com/s3/pricing/>
- 3.IDC Global DataSphere forecast, cited by Micron: IDC forecasts that the volume of data created each year will grow at a 24.4% CAGR from 2023 through 2028. <https://www.micron.com/about/blog/applications/data-center/idcs-insight-on-todays-flexible-server-architecture>
- 4.IBM Cost of a Data Breach Report 2025: IBM and the Ponemon Institute report that the average cost of a data breach in the United States reached \$10.22 million. <https://www.ibm.com/reports/data-breach>



Turn identity into your strongest data control.

Ready to turn insight into action? Get a free demo of the Lightbeam Data Identity Graph and start enforcing data policy with confidence.

[Get Demo](#)

About Lightbeam

Lightbeam provides AI Guardrails for Enterprise Data, helping organizations control what sensitive data AI can reach, use, and expose. Powered by the patented Lightbeam Data Identity Graph, Lightbeam maps sensitive data to identity, access, business context, policy, and risk so teams can stop shadow AI, enforce least privilege access, reduce data exposure, automate privacy workflows, and prove controls are working. Lightbeam helps enterprises safely adopt GenAI, copilots, SaaS AI, and agents while protecting the people and businesses behind the data.