LightningIQ

QUICK START · 2-PAGE WORKSHEET

DATA CENTER REFRESH · TACTICAL COMPANION

Identifying ROT & PII in Your Environment

A do-this-now playbook for storage and security teams. Find the redundant, obsolete, trivial, and sensitive data hiding in your estate — and start cutting footprint and risk before you spend a dollar on new infrastructure.

KNOW IT What you're looking for

ROT — Redundant, Obsolete, Trivial

- **Redundant:** duplicate copies, repeated backups, near-identical versions of the same file
- **Obsolete:** files past retention, ex-employee data, abandoned projects, stale archives
- **Trivial:** system artifacts, caches, temp files, personal media with no business value

PII & Sensitive Data

- **PII / PHI:** names, SSNs, financials, health records, customer & employee data
- **Confidential:** IP, source code, contracts, M&A, board material, credentials
- **Regulated:** anything in scope for GDPR, HIPAA, SOX, CCPA, PCI, or legal hold

DO IT 5 moves to find & reduce ROT

1 Scan everything in place — don't move data first

Inventory every repository: file shares, SAN, NAS, object, cloud buckets, archive, backup. **Tip:** in-place scanning avoids copy cost, risk, and production impact, and finishes in hours/days.

2 Profile by age, owner, type & last-access

Surface dark data, ownerless shares, and files untouched in 3+ years. These are your fastest, lowest-risk reduction wins.

3 Quantify duplicates & near-duplicates

Hash-match exact copies and cluster near-identical versions. Report ROT as a % of total footprint and as projected $ saved.

4 Tag a defensible disposition for each class

Delete, archive, or tier — with the rule and approver recorded. **Tip:** keep an audit log of every decision for compliance.

5 Right-size new infrastructure to the cleaned footprint

Size hardware and cloud against post-cleanup volume, not today's. Most teams land 30–60% smaller.

DO IT 5 moves to find & remediate PII / sensitive data

1 **Run automated content classification**

Pattern + context detection for PII, PHI, financial, and credentials across structured and unstructured data — not just file names.

2 **Map exposure: who can access what**

Flag open shares, over-permissioned folders, and externally-shared links holding sensitive content. **Tip:** prioritize the highest-sensitivity, widest-access combinations first.

3 **Triage ownerless & orphaned sensitive data**

Assign an owner or quarantine. Ownerless PII is the most common source of avoidable breach scope.

4 **Remediate before you migrate**

Tighten permissions, encrypt, redact, or relocate to a governed tier. Never carry unresolved exposure onto new infrastructure.

5 **Preserve holds & chain of custody**

Verify legal-hold items are retained and tracked through the refresh. Keep the audit trail intact for DSARs and regulators.

AIM FOR What good looks like

30–70%	**100%**	**0**	**30–60%**
of unstructured data identified as ROT	of sensitive data located & owner-assigned	open shares holding PII at cutover	smaller hardware / cloud purchase

WATCH FOR Red flags worth an immediate look

- ! Shares no one will claim ownership of
- ! "Everyone" or domain-wide access on sensitive paths
- ! Ex-employee home directories still live
- ! Folders untouched for 3+ years still on hot storage
- ! Backups of backups (and copies of those)
- ! Spreadsheets full of customer PII outside systems

Use this worksheet as a starting point and adapt it to your industry and refresh scope. The data intelligence you generate here keeps paying off — for compliance, security, AI readiness, and the next refresh.