

LO ESENCIAL · Módulo 2 · Almacenamiento y NAS

Dónde viven los documentos y los backups. Por qué el NAS no computa y por qué la prueba de un backup es restaurar.

2.1 · NAS comercial contra autoconstruido

Objetivos

- Comparar con costo total a cinco años
- Decidir según control, mantenimiento y garantía

Qué es un NAS y qué no es

Un **NAS** (*network attached storage*) es un archivador en red: discos, un sistema de archivos con integridad, permisos, y protocolos para que los demás equipos lean y escriban. **No es un servidor de IA**, por mucho «AI» que diga la caja: su CPU está pensada para mover bytes, no para multiplicar matrices.

En la topología del curso, el NAS es el **almacenamiento**: documentos originales, índice de respaldo, copias de seguridad. El nodo de servicios lo monta por red; el de cómputo ni lo toca directamente.

Comercial contra autoconstruido

	NAS comercial (2–4 bahías)	Autoconstruido (PC + sistema de NAS libre)
Costo inicial	medio	bajo–medio
Costo a 5 años	medio: discos + eventual reemplazo de la unidad	bajo–medio: discos; la placa dura más
Control	limitado: lo que el fabricante permita	total
Mantenimiento	actualizaciones automáticas; soporte del fabricante	usted o quien deje entrenado
Garantía	sí, de la unidad completa	por pieza
Integridad de datos	según modelo (los baratos no tienen checksums)	ZFS/btrfs con checksums y snapshots
Riesgo para el cliente	que el fabricante descontinúe la línea	que el implementador desaparezca (módulo 14)
Cuándo	cliente sin técnico, PyME pequeña, prioridad: que funcione solo	hay alguien técnico en la empresa, o el implementador da mantenimiento por contrato

Lo que no se negocia, sea cual sea

1. **Sistema de archivos con checksums** (ZFS o btrfs). Un bit corrupto en una factura escaneada no se nota hasta que se busca.
2. **Snapshots** automáticos: cada hora, retenidos 7 días; cada día, 30; cada mes, 12.
3. **RAID no es backup**. RAID sobrevive a un disco dañado; no a un borrado, un ransomware ni a un incendio. Lección siguiente.
4. **UPS** con apagado limpio.
5. **Discos de la serie para NAS**, no de escritorio: vibración, 24×7.

TCO a cinco años (plantilla)

```

compra unidad + discos (año 0)
+ discos de reemplazo (asuma 1 de cada 4 en 5 años)
+ electricidad (W × 24 × 365 × 5 × tarifa)
+ horas de mantenimiento × tarifa (comercial: ~2 h/año; propio: ~8 h/año)
+ copia externa (segundo equipo o almacenamiento de objetos)
    
```

Ese total, dividido en 60 meses, es lo que se le dice al cliente. Casi siempre sale **más barato que un mes de una persona buscando un documento perdido**.

PROCEDIMIENTO · Laboratorio 2.1 · la comparación para la ferretería

labs/02-01/tco.md: dos columnas (comercial / propio), las cinco líneas del TCO, y la decisión justificada en tres frases. Commit: «Capa 2.1: decisión de almacenamiento».

Qué se degrada en cada perfil

A — El NAS es igual: no depende del cómputo

B — Igual. El NAS se dimensiona por documentos y retención, no por modelo

Cuándo NO usar esto NAS comercial contra autoconstruido

- No use el NAS como nodo de servicios si hace más que archivar: los NAS con contenedores aguantan un proxy y una base pequeña, no un RAG con reranking.
- No use un disco USB como «NAS»: sin checksums, sin snapshots, sin red.

PRÁCTICA · Práctica

1. ¿Por qué RAID 1 no protege contra un ransomware?
2. Calcule el TCO a 5 años de un NAS de 2 bahías con dos discos de 8 TB en su ciudad.

Referencias

1. OpenZFS — documentación: <https://openzfs.github.io/openzfs-docs/>
2. TrueNAS — documentación: <https://www.truenas.com/docs/> [VOLÁTIL – verificado: 2026-09]
3. Backblaze — estadísticas anuales de fallo de discos: <https://www.backblaze.com/cloud-storage/resources/hard-drive-test-data>

2.2 · 3-2-1 y la prueba es restaurar**Objetivos**

- Montar backup cifrado con versiones

- Demostrar un restore desde cero en máquina limpia

La regla y la trampa

3-2-1: tres copias de los datos, en dos medios distintos, una fuera del sitio. Es de 2005 (Peter Krogh, para fotografías) y sigue siendo el mínimo razonable.

La trampa: casi todo el mundo tiene backups; casi nadie ha **restaurado** uno. Un backup que no se ha restaurado no es un backup: es una esperanza.

ATENCIÓN · Ojo

La prueba no es «se copió». La prueba es «en una máquina limpia, con este comando, el sistema volvió a funcionar». Esa prueba se hace **antes** de necesitarla y se repite cada trimestre.

Qué se respalda y qué no

Se respalda	No se respalda (se reconstruye)
documentos originales del cliente	pesos de los modelos (se vuelven a descargar)
base de datos (metadatos, trazas, usuarios)	índice vectorial (se reconstruye desde los documentos)
configuración: docker-compose, .env cifrado , esquemas	cachés, logs rotados
el repositorio del cliente	contenedores (se reconstruyen desde el compose)

Esa tabla decide el tamaño del backup: **gigas, no teras**. Los modelos y el índice son reproducibles; lo que no se puede reproducir es lo del cliente.

Herramienta

Cualquier herramienta de backup con **deduplicación, cifrado y versiones** sirve. En el laboratorio se usa restic [VOLÁTIL – verificado: 2026-09] porque es una sola binaria, cifra por defecto y escribe igual en un disco, un NAS o un almacenamiento de objetos. Alternativas: borg (similar, más antigua), kopla.

```
cd ~/labs/el-tornillo && mkdir -p labs/02-02 && cd labs/02-02
export RESTIC_REPOSITORY=/mnt/nas/backups/el-tornillo      # copia 2 (NAS)
export RESTIC_PASSWORD_FILE=$HOME/.el-tornillo-backup.key # NUNCA en el repo git
restic init
restic backup ~/labs/el-tornillo/datos ~/labs/el-tornillo/config --tag diario
restic forget --keep-daily 30 --keep-weekly 12 --keep-monthly 12 --prune
restic snapshots
```

La **copia 3 (fuera del sitio)** es el mismo comando con otro RESTIC_REPOSITORY: un segundo NAS en la casa del gerente, un SBC con un disco, o un almacenamiento de objetos. El módulo 3

explica cómo llega allá sin abrir puertos.

El script que prueba

```
#!/usr/bin/env bash
# restaurar.sh — reconstruye el sistema en una máquina LIMPIA. Si esto no corre, no hay backup.
set -euo pipefail
DESTINO=${1:-/tmp/restauracion-$(date +%F)}
restic restore latest --target "$DESTINO"
cd "$DESTINO/home/$USER/labs/el-tornillo"
docker compose --env-file config/.env up -d      # los contenedores se reconstruyen
sleep 20
curl -fsS http://localhost:8080/salud && echo "RESTAURADO OK: $DESTINO"
```

Lo que debe ver: RESTAURADO OK y el servicio respondiendo **en otra máquina** (o una VM limpia). Se anota fecha, duración y tamaño en labs/02-02/restauraciones.md. Es el documento que el cliente puede pedir cuando quiera.

Entregable: restaurar.sh, restauraciones.md con la primera prueba. Commit: «Capa 2.2: backup 3-2-1 probado».

Qué se degrada en cada perfil

A — Igual. El backup no usa GPU. La restauración del índice tarda más (se recalcula en CPU)

B — Igual; el índice se reconstruye en minutos

Cuándo NO usar esto 3-2-1 y la prueba es restaurar

- No incluya los pesos de los modelos en el backup: son gigas que se descargan en minutos.
- No cifre con una clave que solo tenga el implementador: **entregada al cliente en sobre cerrado** (módulo 14). Un backup cuya clave se perdió es el peor backup posible.
- No confíe en la «papelera» del NAS como copia: no es una copia.

PRÁCTICA · Práctica

1. ¿Qué pasa si el ransomware cifra el NAS y el backup está en el mismo NAS? ¿Cómo lo evita 3-2-1?

2. Programe restaurar.sh en un calendario trimestral. ¿Quién lo corre si usted no está?

Referencias

1. restic — documentación: <https://restic.readthedocs.io/> [VOLÁTIL – verificado: 2026-09]
2. Krogh, P. (2005/2009). *The DAM Book*. O'Reilly — origen de la regla 3-2-1.
3. CISA — *Data Backup Options* y guías de ransomware: <https://www.cisa.gov/>

2.3 · Retención y habeas data (Ley 1581 de 2012)

Objetivos

- Escribir la política de datos del cliente
- Distinguir dato personal, sensible y público

Por qué un curso técnico tiene una lección legal

Un asistente documental de una PyME colombiana va a ingerir **cédulas, nóminas, direcciones, historias de crédito y firmas**. En Colombia eso está regulado por la **Ley 1581 de 2012** (régimen general de protección de datos personales), reglamentada por el **Decreto 1377 de 2013**, y vigilada por la **Superintendencia de Industria y Comercio (SIC)**. Ignorarlo no es un riesgo técnico: es una sanción al cliente. Y el diseño del sistema es donde se cumple o se incumple.

Las cuatro palabras que hay que distinguir

Término	Qué es	Ejemplo en la ferretería
Dato personal	cualquier información que identifique o haga identificable a una persona	nombre y cédula de un proveedor persona natural
Dato sensible	el que afecta la intimidad o puede causar discriminación: salud, origen, orientación, biometría, sindical	el certificado médico de un empleado
Dato público	el que está en registros públicos	el NIT de una empresa
Titular	la persona dueña del dato	el empleado, el cliente, el proveedor

Regla del sistema: los datos sensibles **no entran** al asistente salvo autorización expresa y un caso de uso justificado. Es más fácil no ingerirlos que protegerlos.

Los principios que se traducen a diseño

Principio (Ley 1581, art. 4)	En el sistema
Finalidad	cada colección de documentos tiene un propósito escrito; el asistente no cruza colecciones sin permiso
Libertad	hay autorización del titular (los contratos y formatos del cliente la incluyen; se verifica)
Acceso y circulación restringida	permisos por grupo antes de recuperar (módulo 6.3); el vendedor no ve nómina
Seguridad	cifrado en reposo (backups) y en tránsito (malla); sin puertos abiertos
Confidencialidad	el implementador firma acuerdo; su acceso es revocable (módulo 3.3)
Transparencia	las trazas permiten decir qué se hizo con qué dato (módulo 11)

Retención: cuánto se guarda y cuándo se borra

No hay una cifra única: depende del tipo de documento y de otras normas (tributarias, laborales).

La política se escribe **por tipo de documento**, con el cliente y su contador:

Documento	Retención orientativa	Nota
Facturas y soportes contables	según norma tributaria vigente (años)	verificar con el contador
Contratos laborales y nómina	según norma laboral (años tras terminación)	verificar
Hojas de vida no contratadas	meses	borrar; son datos de terceros sin relación
Trazas del asistente	12 meses	anonimizar después
Backups	mismas reglas que el original	la política de forget de la lección anterior las implementa

ATENCIÓN · Ojo

Borrar significa borrar también del índice, de los backups vencidos y de las trazas. Un sistema que no puede borrar un dato no cumple la ley aunque lo tenga cifrado.

PROCEDIMIENTO · Laboratorio 2.3 · la política de una página

labs/02-03/politica-datos.md con: responsable del tratamiento (el cliente), encargado (usted, si aplica), finalidades por colección, qué datos sensibles quedan **fuera**, retención por tipo, canal para que un titular pida acceso o supresión, y cómo se ejecuta esa supresión en el sistema (comando o procedimiento). Una página. Firmada por el gerente.

Entregable: la política. Commit: «Capa 2.3: política de datos».

Nada. La política es la misma en A, B y C — y es de las pocas piezas que **no cambia** cuando cambia el modelo.

Cuándo NO usar esto Retención y habeas data (Ley 1581 de 2012)

- No es asesoría legal: esta lección da el mapa; los plazos de retención concretos los fija el contador o el abogado del cliente. Escríbalo así en la política.
- No copie una política de internet: la SIC revisa que corresponda al tratamiento **real**.

PRÁCTICA · Práctica

1. Un cliente quiere que el asistente «recuerde todo lo que cada empleado le pregunta, para siempre». Escriba la respuesta en tres líneas con la ley en la mano.
2. Un empleado pide que borren sus datos. Liste cada lugar del sistema donde están.

Referencias

1. Ley 1581 de 2012 — Congreso de Colombia. Texto en Función Pública: <https://www.funcionpublica.gov.co/eva/gestornormativo/norma.php?i=49981>
2. Decreto 1377 de 2013 — reglamentario de la Ley 1581.
3. Superintendencia de Industria y Comercio — Protección de datos personales: <https://www.sic.gov.co/>
4. Circular Externa 002 de 2015 de la SIC (Registro Nacional de Bases de Datos) — verificar vigencia y umbrales aplicables a la PyME.