

IA local para PyMEs · 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.

Qué se construye en este módulo

 **NAS comercial contra autoconstruido**

Comparar con costo total a cinco años

 **3-2-1 y la prueba es restaurar**

Montar backup cifrado con versiones

 **Retención y habeas data (Ley 1581 de 2012)**

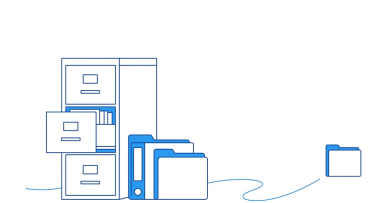
Escribir la política de datos del cliente



2.1 · NAS comercial contra autoconstruido

NAS comercial contra autoconstruido

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.



En una tabla

Perfil	Qué pasa
A	El NAS es igual: no depende del cómputo
B / C	Igual. El NAS se dimensiona por documentos y retención, no por modelo

El comando, copiable

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

Si no corre desde cero con un comando, no entra al curso.

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.

— —

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

3-2-1 y la prueba es restaurar

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.

En una tabla

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

El comando, copiable

```
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
```

Si no corre desde cero con un comando, no entra al curso.

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 posi
- No confíe en la «papelera» del NAS como copia: no es una copia.



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

Retención y habeas data (Ley 1581 de 2012)

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.



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.

En una tabla

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 Titular	el que está en registros públicos la persona dueña del dato	el NIT de una empresa el empleado, el cliente, el proveedor

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.



Entregables del módulo

- `restaurar.sh`, `restauraciones.md` con la primera prueba.
- la política.

**Dónde viven los
documentos y los backups.**