PRODUCT EXPLORER

Gestión segura de datos de prueba mediante descubrimiento y de-identificación

---

CLOUDATA | JULIO 2026

tdm.cloudata.pe

01 | DOVATDM

Producto en una mirada

DovaTDM transforma archivos reales en datos de prueba utilizables, trazables y con menor exposición de información sensible.

Propósito

DovaTDM es una aplicación web de Test Data Management para equipos que necesitan validar software, integraciones y analítica sin distribuir copias directas de datos productivos. Detecta campos sensibles, aplica políticas repetibles y entrega un archivo de-identificado junto con su reporte de ejecución.

> **PRINCIPIO** El archivo fuente se procesa en memoria. La aplicación no lo persiste en el sistema de archivos; almacena en PostgreSQL el resultado de-identificado, el trabajo y el reporte como documentos cifrados.

Capacidades actuales

Capacidad	Cobertura
Formatos	CSV, JSON, TXT, XML y LOG en UTF-8
Carga	Un archivo por trabajo, hasta 15 MB
Detección	Nombres de campo, patrones, validaciones y algoritmo de Luhn
Transformación	Masking, redacción, tokenización, seudonimización y desplazamiento de fechas
Persistencia	PostgreSQL con cifrado pgcrypto y campos planos nulos
Trazabilidad	Reporte y auditoría sin exponer valores sensibles originales

Resultado para el negocio

- Menor propagación de datos productivos hacia QA, UAT y desarrollo.
- Datos consistentes entre ejecuciones para pruebas repetibles.
- Evidencia centralizada de la política aplicada y del resultado generado.

02 | DOVATDM

Flujo de de-identificación

Un proceso guiado, determinista y auditable desde la recepción hasta la descarga.

1. **Intake.** El operador selecciona un archivo compatible. DovaTDM valida formato, tamaño, codificación y contenido antes de procesarlo.
2. **Discovery.** El motor perfila columnas y valores para identificar categorías sensibles mediante nombres, patrones y controles especializados.
3. **Policy.** El usuario elige una política que define la acción por categoría: enmascarar, redactar, tokenizar, seudonimizar o desplazar fechas.
4. **Transform.** Las reglas se ejecutan en memoria. Las transformaciones deterministas conservan coherencia cuando el escenario de prueba lo requiere.
5. **Protect.** El resultado, el reporte y el trabajo se guardan cifrados en PostgreSQL. El archivo fuente no se escribe en disco por la aplicación.
6. **Deliver.** El operador descarga el archivo de-identificado y revisa el reporte en Audit sin visualizar valores sensibles originales.

Categorías detectables

Identificadores	Datos operativos	Contenido no estructurado
Email, teléfono, documento nacional y nombre	Tarjeta, cuenta, dirección, fecha e IP	Texto libre con redacción por expresiones regulares

Controles de entrada

- Rechazo de contenido binario y archivos que no sean UTF-8.
- Límite de 15 MB y validación previa al trabajo.
- Neutralización de fórmulas peligrosas al producir archivos CSV.

IMPORTANTE La calidad del resultado depende del formato, los encabezados y los patrones presentes. Discovery debe revisarse antes de utilizar el archivo en un entorno de prueba.

03 | DOVATDM

Módulos del producto

Cada módulo representa una etapa concreta del ciclo de gestión de datos de prueba.

Módulo	Función principal	Salida
Intake	Carga controlada y validación inicial del archivo.	Archivo aceptado para análisis
Discovery	Perfilado y detección de categorías sensibles.	Hallazgos por campo y categoría
Policies	Selección de reglas de protección según el uso previsto.	Política versionada para la ejecución
Masking	Aplicación de transformaciones y generación del resultado.	Archivo de-identificado
Audit	Consulta del trabajo y reporte sin valores sensibles originales.	Evidencia operativa y descarga
Configuration	Visibilidad del estado de infraestructura requerida y opcional.	Estado técnico del servicio

Acceso por responsabilidad

Perfil	Acceso esperado
Administrador	Acceso integral y administración de usuarios.
Operador	Intake, Discovery, Masking y Audit para ejecutar el proceso.

DISEÑO OPERATIVO La separación por módulos permite revisar la detección y la política antes de generar un resultado, reduciendo ejecuciones accidentales con una configuración incorrecta.

Entregables por trabajo

- Archivo de-identificado listo para su uso controlado.
- Reporte con categorías, acciones y conteos de transformación.
- Registro de auditoría asociado al usuario y a la ejecución.

04 | DOVATDM

Catálogo de políticas

Perfiles predefinidos para equilibrar protección, formato y utilidad de prueba.

Política	Uso recomendado	Enfoque
Safe Harbor Default	Pruebas generales con datos heterogéneos.	Protección determinista y no reversible.
Strict Redaction	Escenarios donde importa más minimizar exposición que conservar formato.	Redacción amplia de identificadores.
GDPR Pseudonymization	Pruebas que requieren relaciones estables entre registros.	Seudónimos y tokens deterministas.
PCI Payment Data	Flujos de pago y validación de integraciones financieras.	Redacción reforzada de tarjetas y cuentas.
Analytics High Utility	Analítica y tendencias con fechas consistentes.	Seudónimos estables y date shifting.
QA Format Preserving	Pruebas funcionales repetibles.	Estructura útil y valores estables.
Irreversible Tokenization	Identificadores y cuentas que no deben reconstruirse.	Tokenización no reversible prioritaria.

Acciones disponibles

Acción	Comportamiento
Mask	Oculta parte del valor manteniendo una representación reconocible.
Redact	Sustituye el contenido detectado por un marcador protegido.
Tokenize	Genera un token estable sin conservar el valor original.
Pseudonymize	Crea un sustituto determinista útil para relaciones de prueba.
Date shift	Desplaza fechas de forma consistente para preservar intervalos.
Regex redact	Elimina patrones sensibles encontrados dentro de texto libre.

ALCANCE Los nombres GDPR y PCI describen perfiles técnicos de protección. No constituyen una certificación ni garantizan cumplimiento regulatorio por sí solos.

05 | DOVATDM

Arquitectura y seguridad

Controles de defensa en profundidad para la aplicación, la base de datos y la operación.

NAVEGADOR > NGINX / TLS > DovaTDM > POSTGRESQL / PGCRYPTO

Capa	Control implementado
Perímetro	HTTPS con certificado Let's Encrypt y proxy inverso Nginx.
Aplicación	Servicio enlazado a 127.0.0.1:8021; no expuesto directamente a Internet.
Sesión	Cookies Secure, HttpOnly y SameSite; protección CSRF y control de roles.
Navegador	CSP, protección contra framing y encabezados de seguridad.
Abuso	Rate limiting y límite de carga de 15 MB.
Base de datos	PostgreSQL en 127.0.0.1:5432; cifrado pgcrypto y columnas planas nulas.
Integridad	Digest HMAC del contenido para validación sin almacenar el original.
Recuperación	Backups diarios cifrados, permisos 0600 y retención de siete días.
Host	Servicio sin swap y con core dumps deshabilitados para reducir rastros sensibles.

Ciclo de datos

- Fuente: transitoria en memoria durante el procesamiento.
- Resultado, trabajo y reporte: cifrados en PostgreSQL.
- Respaldo: cifrado, acceso restringido y eliminación por retención.

ARQUITECTURA ACTUAL El alcance productivo no requiere Redis, RabbitMQ ni almacenamiento de objetos. Estos componentes permanecen opcionales para evoluciones futuras.

06 | DOVATDM

Casos de uso y alcance

DovaTDM está orientado a archivos de texto estructurados o semiestructurados dentro de un proceso controlado.

Casos de uso prioritarios

- Preparar datos para QA y UAT sin distribuir identificadores reales.
- Reproducir defectos en desarrollo con relaciones de datos consistentes.
- Validar migraciones e integraciones usando conjuntos de prueba protegidos.
- Ejecutar analítica de prueba conservando tendencias y desplazando fechas.
- Compartir evidencia de soporte sin incluir valores sensibles del cliente.

Alcance operativo actual

Incluido	Fuera del alcance actual
CSV, JSON, TXT, XML y LOG	Excel, PDF y archivos binarios
Carga web de un archivo por trabajo	Procesamiento batch y planificación automática
Políticas predefinidas	Diseñador visual de reglas personalizadas
Usuarios y roles locales	SSO empresarial y federación de identidades
PostgreSQL cifrado	Conectores directos a bases de datos externas
Descarga manual y auditoría	API pública de orquestación

BUENA PRÁCTICA El archivo de-identificado debe conservarse solo durante el tiempo necesario, con acceso restringido y una revisión de utilidad y riesgo antes de usarlo fuera del equipo autorizado.

Principio de adopción

Comenzar con un conjunto representativo y pequeño, validar Discovery y comparar el resultado con los criterios de prueba. Luego ampliar el volumen sin cambiar la política aprobada.

07 | DOVATDM

Adopción y criterios de aceptación

Una puesta en uso verificable para equipos de desarrollo, QA, datos y seguridad.

Inicio rápido

1. **Ingresar.** Autenticarse en la aplicación con un perfil autorizado.
2. **Seleccionar.** Cargar un archivo compatible y elegir la política apropiada.
3. **Revisar.** Confirmar los hallazgos de Discovery antes de ejecutar.
4. **Generar.** Procesar el trabajo y descargar el resultado de-identificado.
5. **Auditar.** Verificar el reporte, la política aplicada y los conteos de transformación.

Checklist de aceptación

- El acceso corresponde al rol asignado y las sesiones se cierran correctamente.
- El archivo fuente no aparece en almacenamiento de aplicación ni en auditoría.
- Los valores sensibles detectados no permanecen en el resultado.
- Los valores deterministas se mantienen consistentes cuando corresponde.
- El reporte y la descarga están disponibles para el trabajo autorizado.
- La validación productiva y el respaldo cifrado terminan sin errores.

> **ESTADO DEL PRODUCTO** Operativo en producción. La validación automatizada cubre 12 pruebas del motor, almacenamiento cifrado, autenticación, CSRF, políticas, trabajos, reportes, descargas y cierre de sesión, además de un smoke test productivo.

Acceso y soporte

Producto: https://tdm.cloudata.pe

Soporte: help@cloudata.pe

Alcance basado en el release productivo 20260721-05 y el repositorio DovaTDM, verificados el 22 de julio de 2026.