OPSWAT.

DATASHEET

Proactive DLP™

Stop Potential Data Breaches and Regulatory Compliance Violations Detect & Block Sensitive Data in Files & Emails

How It Works

Proactive DLP detects and blocks sensitive information across 125+ file types through AI-powered recursive inspection and content-based policy enforcement, helping protect organizations from data breaches and supporting regulatory compliance. Files are analyzed using content-based detection to identify and classify sensitive elements—such as credit card numbers or Social Security numbers—which are then processed according to defined policies, enabling actions like blocking, redaction, or customized workflows before content is delivered or stored.

Advanced Sensitive Data Detection

- Proactive Detection: Identifies and blocks sensitive data across 125+ file types, including embedded layers, cropped images, and obscured text.
- Secret & Credential Detection: Detects and blocks API keys, passwords, tokens, and other embedded credentials.
- Document Identification: Uses AI to recognize regulated documents such as IDs, passports, licenses, and contracts.
- NSFW & Offensive Content Detection: Detects explicit or offensive text and imagery.
- OCR: Extracts text from scanned documents, images, and PDFs for analysis.
- NER: Classifies unstructured text into entities like names, credit cards, and government IDs.
- Metadata Detection & Removal: Identifies and removes hidden metadata (e.g., author, GPS, version history).
- Redaction & Anonymization: Automatically redacts or anonymizes sensitive data in Office files and PDFs.
- Watermarking & Tagging: Applies watermarks or metadata tags for content traceability.

Key Features

- AI-Powered Content Detection
- Recursive File Inspection
- Content-Based Policy Enforcement
- Regulatory Compliance Alignment
- Integration with Multi-Layered Security Ecosystem
- 125+ Supported File Types

Types of Sensitive Data OPSWAT Proactive DLP Detects

- Social Security numbers
- Credit card numbers
- IPv4 addresses and CIDR (Classless Inter-Domain Routing)
- Custom RegEx (regular expressions)
- Secrets in text files (AWS, Microsoft Azure, IBM Cloud, and Google Cloud Platform)
- PHI (protected health information) and PII in DICOM (Digital Imaging and Communications in Medicine) files

Performance

Windows System Info	
RAM	32GB
CPU	Intel® Core™ i7-6700 CPU @ 3.40GHz × 8
OS	Windows Server 2019
Disk Drive	256GB SSD

Linux System Info	
RAM	32GB
CPU	Intel® Core™ i7-4790 CPU @ 3.60GHz × 8
OS	Ubuntu 20.04.1 LTS
Disk Drive	256GB SSD

Resources	
Windows	MetaDefender Core v4.19.0 with 8 engines
Linux	MetaDefender Core v4.19.0 with 5 engines

Test Results

File Type	Average File Size (KB)	Total Files	Sensitive Files	Detection (s/file)		Detection + Redaction (s/file)	
				Windows	Linux	Windows	Linux
Text	170	1456	728	0.04	0.05	0.04	0.05
Text	3174	864	592	0.25	0.21	0.25	0.22
Pdf	392	1572	785	0.28	0.28	0.44	0.47
Pdf	6553	846	242	2.5	2.72	3.0	4.9
Word	224	629	314	0.14	0.16	0.15	0.16
Word	2969	808	128	0.26	0.26	0.36	0.35
Excel	191	2942	1471	0.51	0.43	0.52	0.46
Excel	1904	840	192	3.3	2.8	3.3	2.7

OPSWAT.
Protecting the World's Critical Infrastructure