eBook

5 Requirements for Modern DLP

netskope

In today's cloud-enabled world driven by digital transformation and hybrid work, organizations must align data security initiatives to modern trends that change at cloud speed. Moreover, every organization has specific data protection needs and unique use cases, because data protection programs are never exactly the same. That's why only a cloud-delivered Data Loss Prevention (DLP) architecture can ensure high flexibility, great scalability, and unlimited computing power. Cloud means also staying up-to-date at all times, with always-on protections and updates available in real time. A cloud DLP technology as a service is clearly the right approach to enterprise data protection, but it shouldn't be the only criteria to consider when moving to an effective data protection strategy.

A DLP technology needs to be adaptive, rich in functionality, broad in coverage, and deep in effectiveness. It also needs to provide a high degree of efficacy to guarantee accurate data protection for any type of data, across every environment and against every data loss risk.

There are useful architectural guidelines and technology recommendations that should always be considered before moving from an existing DLP deployment to one that meets modern hybrid work requirements:

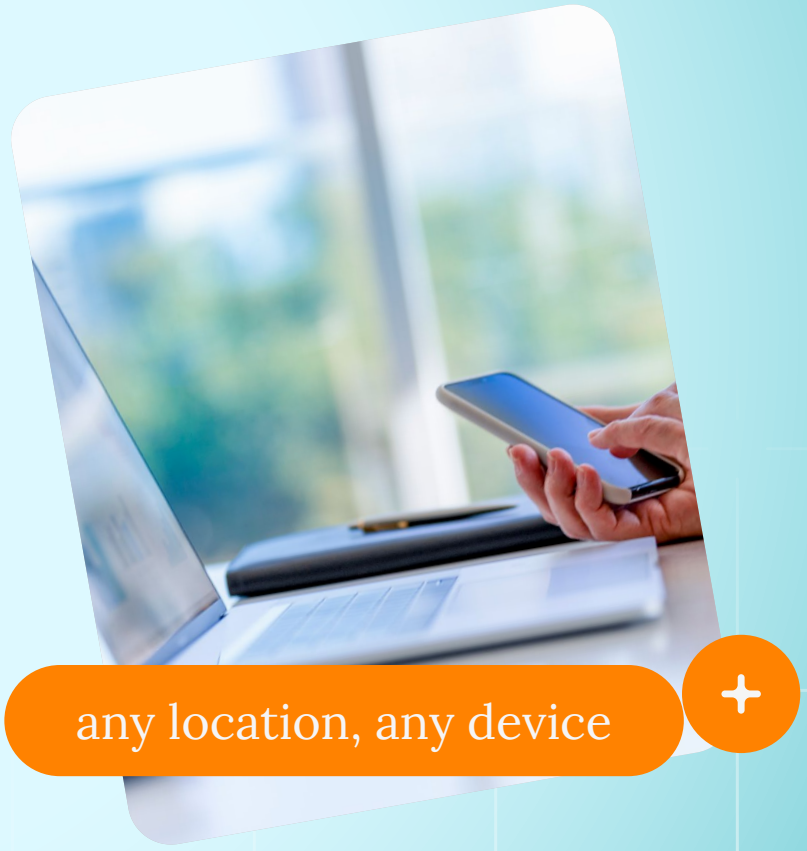
01

Comprehensiveness of coverage

You can't protect what you can't see, and data today is flowing through many more environments than before. Legacy Enterprise DLP, traditionally deployed on the physical network, provides extensive coverage of data channels on-premises, including web transmissions, email SMTP, and endpoint. It is reasonable to expect a modern cloud-based solution to extend protection to cloud-based repositories like SaaS applications, IaaS, and cloud email as well as ensuring coverage for on-premises environments like networks, endpoints, and email. It's always recommended to ensure that a cloud DLP solution provides complete enterprise coverage for both the cloud and the traditional on-premises channels. In that sense, it's also important to know that most cloud DLP solutions are architected in the cloud to solve only for the cloud use cases and don't cover certain on-premises channels such as endpoints.

01

Today there are many use cases that are fundamental and must be properly addressed, such as transfer of sensitive data across thousands of unsanctioned risky SaaS apps or to personal instances of sanctioned SaaS apps such as a personal Gmail account or a personal OneDrive instance, or to private applications in the public cloud or in the data center. For the modern highly distributed enterprise made of multiple branches and the hybrid remote workforce, DLP must protect all data transmissions from any location and any device, including managed and unmanaged devices, and even IoT. Endpoint transfer of sensitive data to a USB is also a significant loss vehicle that must be controlled, even when these endpoints are not online. Outbound emails, sensitive in nature, are another major vector of data loss as well as confidential communications on collaboration apps like Slack and Teams.



02

Core data detection capabilities

Accurate data visibility is a tactical necessity to assess the entire operating environment and implement the optimal protection strategy. That all starts with discovery and classification of every sensitive data including structured and unstructured personally identifiable information (PII), intellectual property (IP), confidential information, and trade secrets.

Because manual data classification by the data owner can be an unreliable process, this task also needs to be automated in DLP by means of a complete, not partial, set of detection engines. Such engines define the organization's predefined detection policies or data profiles. In detail:

- **Data identifiers** have been and still are the must-have of any DLP solution. They must be able to identify thousands of different types of sensitive data based on described matching criteria that generally characterize objects like SSN, payment card information, or passport numbers, such as number of digits, text patterns, sequences, separations, and proximity keywords. Having regular expression (regex) capabilities is fundamental but this can't be a check-the-box feature. More data types and modern use cases have emerged with newer compliance requirements demanding to protect the privacy of individuals in a broader fashion. In fact, a large number of predefined data identifiers is the first element of consideration, but also the granularity of the rule customizations like severity levels, the extent of proximity checks, boolean logic, etc., must also be taken into account.

02

- The number of **file types** supported is another key element. There are thousands of data types that may contain sensitive information: Text, Presentation, Email, Images and Screenshots, Spreadsheet, Cad, Social Posts, Online Forms, Slack Messages and other Chat channels, Encapsulation, Attachments, Graphics and Pictures like JPEG and PDF, etc.
- **Artificial Intelligence/Machine Learning (AI/ML) data classifiers** aid data discovery and identification. Manually defined rules are the foundation of data detection, but in the modern world, automated engines supply invaluable assistance and make sensitive data detection and categorization more accurate. They also adapt to changing conditions and identify content similarities.
- **Exact data matching (EDM)** is a traditional yet almost infallible method designed to detect specific information that is sourced from structured data sources such as spreadsheets and databases. With EDM, a DLP solution can fingerprint and index datasets of confidential records, information that, when combined, can identify an individual such as customers' full names, Social Security numbers, addresses, identification numbers, etc., or financial records that define an individual's financial assets like credit card numbers or bank account numbers—even healthcare information or product identification and pricing databases. Such indexed information must then be tracked and found anywhere the data flows are expected to happen.

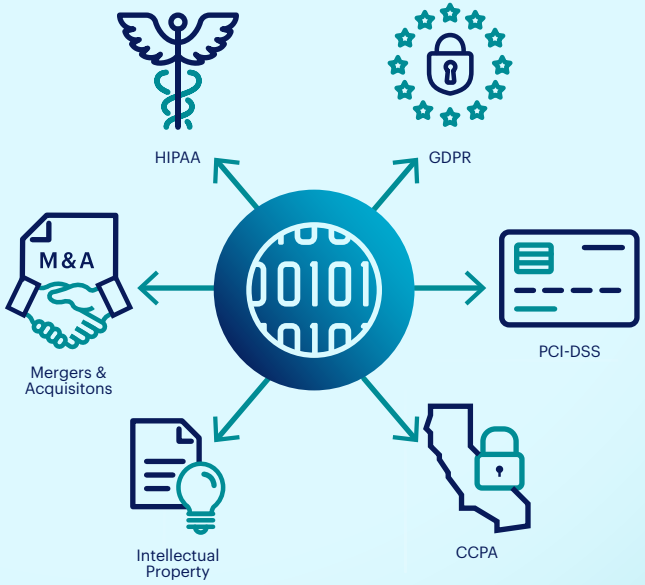
A Note on EDM

For EDM to be effective and accurate, it has to be capable of leveraging granular conditions to match various pieces of the indexed data and match combinations of data fields from a particular record, the ones that matter. EDM high scale is a very important factor, especially for large enterprises and for organizations looking for future growth. Millions or even billions of records must be supported.

03

Future-state data profiling

Don’t look just for the data identifiers you need now. Look for several thousands of predefined data identifiers, including localized patterns like country-based identification cards, as your future needs most likely will expand along with your organization’s size and data protection maturity. Look for the presence of regulatory compliance templates that you need to support to verify that the latest are all there—GDPR, CCPA, PII, PCI, PHI, to name a few of the better-known regulations and types. Understand the level of commitment of the vendor to keep up with the most recent compliance requirements, and determine if the vendor will likely expand them in the future. The ability to edit existing regexes or identifiers or to create custom data identifiers with granular controls is critical as every organization has different needs such as a specific type of information that is sensitive just for that particular organization.



04

Advanced data detection capabilities

Over the years, data has also evolved significantly, growing in volume, variety, and velocity, and is more unstructured than ever. The introduction of newer data types and modern ways of sharing and transmitting data, the massive growth of data volumes, and new compliance requirements demand advanced ways of detecting sensitive information. Legacy DLP solutions have introduced advanced detection capabilities in the past but have started failing to produce accurate detection results due to lack of computing power and scale. Therefore, they generate more and more false positives that hinder business flows and overwhelm incident response teams.

On the other hand, the majority of recent cloud DLP solutions may still be immature and unproven in terms of efficacy. It's important to verify the presence and the level of sophistication of the following advanced detection capabilities:

- In the present world, users find it very convenient to snap photos of documents, forms, ID cards, whiteboards, and even pictures of other pictures. For example, screenshots are a very common vehicle to quickly capture information and immediately share it with a colleague. As a consequence, **Optical Character Recognition (OCR)** and AI-based image recognition are becoming more and more instrumental for a modern and future-looking data protection strategy. With OCR, a DLP solution can extract textual information from an image and can then apply data classification based on the detection policies that are in place.
- **AI and ML image classification** is fundamental to recognize common file and document types like SSN cards, patents and M&A documents, tax forms, source code, desktop screenshots, passports and other IDs, etc., without necessarily extracting the content that such assets contain. Such detection methods must provide a high level of sophistication to be able to recognize images through variations like blurry, crumbled, and damaged pieces of content, with information that may be hard to read clearly. This is because pictures and screenshots may be taken quickly and with poor or too strong light conditions or because a document may be damaged and aged.

04

- **File and document fingerprinting** is another advanced capability that many organizations find vital. Certain mission-critical documents, intellectual property, and highly confidential files must be protected at all costs from complete or partial exfiltration and duplicate copies. File fingerprinting can index entire documents and then detect exact or even partial copies of the information that they contain with certain degrees of similarity, when this content is found across environments and transmission channels that are considered risky, such as an upload to a private instance of an email application.

05

Risk-aware data protection, a zero trust-ready model

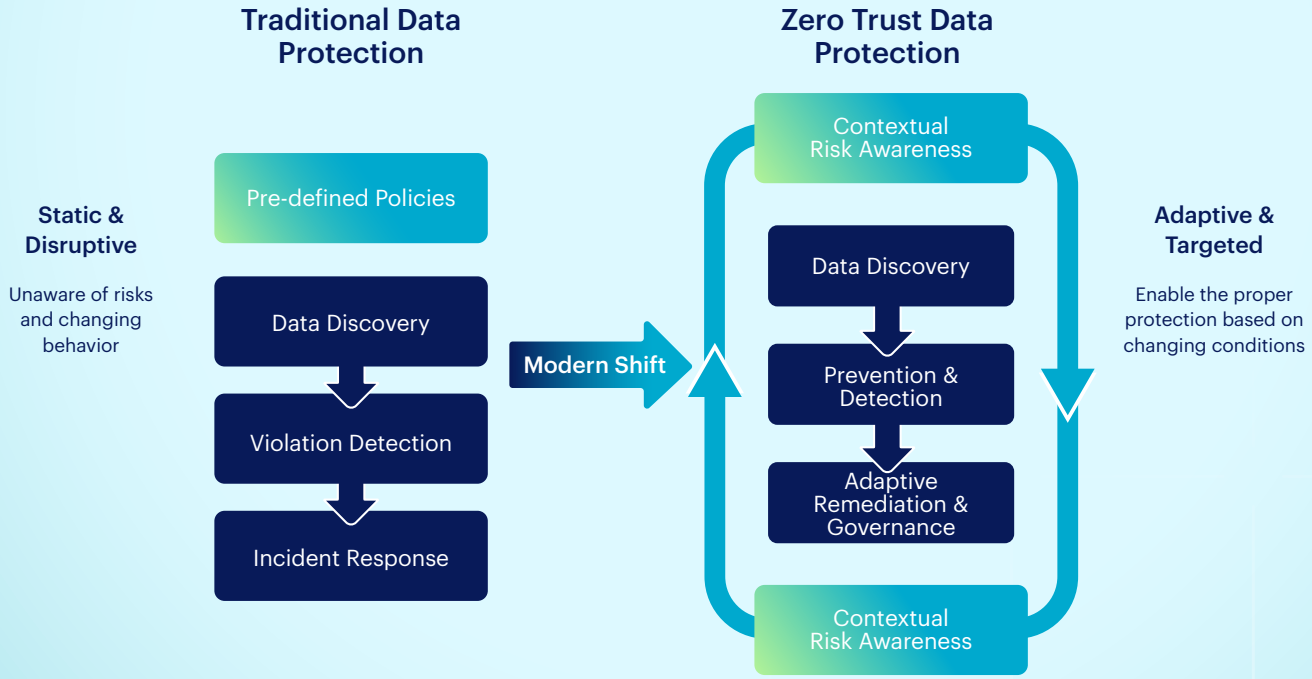
Digital transformation has forever changed our operating paradigm, and needs a model to match. Zero trust is a modern strategy that brings security controls to the data itself as a new perimeter, and replaces implicit trust with continuous and adaptive risk assessment in order to constantly adapt to changing risk conditions. Data controls of the past have proven to cause operational friction and hinder value creation because they lacked context. This is why traditional DLP has failed to be effective: There was not enough business context and risk awareness to provide confidence for preventing data movements. Most of the incident remediation decisions in DLP had to be manually made by the incident response team, which also lacked enough risk and behavioral context. Because of that, traditional DLP is today perceived as a business inhibitor especially when blocking mode is turned on, rather than an effective data protection solution. In fact, most organizations are using it as a data discovery and compliance tool, working rather in monitoring mode in order to avoid problems.

With zero trust applied, these challenges are solved. Data protection technology is required to shift from a static model made of predefined fixed policies, lacking context and unaware of risks and changing behaviors, to a dynamic and adaptive zero trust approach that can leverage security context and continually enable the proper protection action automatically based on changing conditions.

Automated data protection response requires defined processes and granular policies along with clear rules of engagement, what actions to take under what conditions with what degree of confidence. DLP must integrate with the most number of security control points, continually ingest their logs and findings, and leverage them dynamically. A zero trust-ready DLP must take into account organizational risks from users, devices, data, networks, and applications in order to gain rich risk awareness and always provide the right remediation action. For example, behavioral monitoring for users, devices, and applications gives valuable insight into anomalous user activity, potentially malicious actions, risky apps, unsafe connecting locations, unsecure postures, and indicators of compromise.

05

To be truly effective, a zero trust data protection solution needs to monitor what's taking place and who's doing what across the entire corporate infrastructure, including clouds, remote users, and unmanaged devices.



For More Information

Netskope, a global cybersecurity leader, is redefining cloud, data, and network security to help organizations apply Zero Trust principles to protect data. The Netskope Intelligent Security Service Edge (SSE) platform is fast, easy to use, and secures people, devices, and data anywhere they go.

Learn how Netskope helps customers be ready for anything on their data protection journey.

netskope