



Securosis

# THE EXECUTIVE GUIDE TO DATA LOSS PREVENTION



*Technology Overview, Business Justification, and Resource Requirements*

# Introduction to Data Loss Prevention

## Intelligent Protection for Digital Assets



In the information age, “loss prevention” is no longer limited to reducing theft along the supply chain and in retail or office locations. Our most sensitive assets are more likely to be found in databases, email, and files — rather than in boxes, filing cabinets, and warehouses. From customer lists, payment information, and corporate financials, to product and engineering plans — the lifeblood of any enterprise is found as much in electronic bits on networks and laptops as in offices, retail stores, manufacturing plants, and shipments.

## Defining Data Loss Prevention

(DLP) suites combine an array of technologies to protect information throughout its lifecycle, with minimal impact on business process. They accomplish this by ‘understanding’ both the content and context of information, matching them against central policies, and enforcing business rules. At the core, DLP uses deep content analysis to peer inside files and network communications to identify sensitive content, rather than relying on manual processes such as tagging, watermarking, and hand-classification. In other words, DLP actually recognizes the information it is looking at (within limits), and matches it to policies set for acceptable use of the content.

Although we’ve long recognized the importance of our digital assets, in recent years it has become increasingly difficult to effectively protect them. We can’t simply lock the data away behind walls and guards — much of its value is directly tied to our ability to easily and seamlessly use it throughout our organizations, and any security that interferes with business process is likely as damaging as the loss it’s intended to prevent.

To meet these challenges we have seen the emergence of a new technology: Data Loss Prevention. DLP is designed to protect information assets with minimal interference in business processes. It provides new insights into how information is used, and enforces protective controls to prevent unwanted incidents. It’s not perfect, and no technology can completely eliminate information loss, but in combination with appropriate security processes DLP can reduce risk, improve data management practices, and even lower certain compliance costs.

In practice you define what information you want to find or protect, then use DLP to automatically locate where it’s stored, where it’s being communicated, and where it’s being moved (e.g., to laptops or USB sticks). You can merely audit and alert if something violates policy, or implement a variety of protective actions.

### *We define DLP as:*

*Products that, based on central policies, identify, monitor, and protect data at rest, in motion, and in use through deep content analysis.*



This definition encapsulates the core components of a DLP solution: centralized management, identification of defined information, monitoring of usage, and protection from policy violations. A DLP solution does this in storage, on networks, and on employee computers, using advanced analysis techniques. While there are tools that just solve parts of the problem (such as endpoint-only tools) we generally recommend full suites which include all the major components, since they provide better protection and are more cost effective in the long term. We’ll describe how it all works later in this report, after we outline the business justifications for investing in DLP.

# Reasons to Invest in Data Loss Prevention



## Data Loss Prevention is Typically Justified Using Five Drivers:

- **Risk Reduction:** By knowing where your data is stored and how it's being used, you can reduce your overall exposure to potential loss.
- **Cost Savings:** DLP may help reduce other costs associated with data management and security.
- **Compliance Support:** DLP helps reduce the direct costs associated with some regulatory compliance, can ease audits, and reduces the risks of certain compliance-related incidents.
- **Policy Enforcement:** Many data management policies in enterprises are difficult or impossible to enforce. DLP supports enforcement of acceptable use of information, not just security controls.
- **Data Security and Threat Management:** While no security tool stops all threats, DLP reduces the risk of certain malicious activity.

## Example Use Cases for DLP

- A retailer uses DLP to inventory their entire storage infrastructure and employee laptops to identify unencrypted credit card numbers which could be in violation of the Payment Card Industry (PCI) Data Security Standard. These reports are first used to remove as much of the data as possible, and then provided to PCI auditors to demonstrate compliance. This cut 3 months off their compliance cycle and reduced PCI audit costs by 20%.
- A financial institution uses a full-suite DLP deployment to protect customer financial information. They link the DLP system to their primary customer database, and then monitor network communications and generate alerts if customer records are transmitted without encryption. They use data-at-rest scanning to make sure customer records aren't stored on employee laptops or on servers except where explicitly permitted, and block the transfer of files with customer records to portable storage (USB thumb drives). They also allow employees to use personal email, like Hotmail, but block outbound messages with customer records. Since implementation, they've seen an 80% reduction in data misuse and prevented 2 potential data breaches.
- An engineering firm uses DLP to track the movement of sensitive engineering plans for new products. The plans are normally stored on a central server, so the DLP solution monitors that server and identifies any of those documents, or parts of them, being sent outside the organization.

# How DLP Works

## A Non-Technical Overview

Rather than talking about specific technology components, we categorize DLP architectures based on where they protect information: data-in-motion for email and other network communications, data-at-rest for stored data, and data-in-use as for interactions with the data on computers, such as copying it to USB drives. Behind all this is the central management server where we define policies and manage incidents.

As we outline the architecture, keep in mind that we're showing you every possible DLP component. You won't always need every piece, and in practice some pieces are commonly combined together.



## Data-In-Motion

When discussing DLP, we use data-in-motion to describe content moving around our networks, including email, web traffic, instant messaging, and other communications. We use three components to monitor and protect data in motion:

- **Network Monitor:** An appliance or server that sniffs your network to identify and monitor your data. Typically these are used in a passive mode, where they just scan all data passing by, but some tools also proxy traffic to directly block policy violations. Many DLP tools can also work with existing web gateways to block unacceptable traffic (including some encrypted traffic), such as someone trying to send a customer list to their personal site.
- **Email:** Email is a bit different than other network traffic, and the most important channel for protecting information. The way email works

## Data-At-Rest

One of the most valuable capabilities of a DLP solution is its ability to search deep into stored data to identify sensitive information. Consider the value of knowing every location where customer records are stored, even if someone's moved them onto a laptop or an unapproved server. We use three components for data-at-rest:

1. **Remote Scanning:** The DLP solution connects to the storage repository, just like a user looking for a file. It then scans all files for sensitive data.
2. **Local Software Agent:** Remote scanning uses network bandwidth, so while it can reach anywhere users can, it isn't always the most efficient method. For major repositories, we can install a more efficient software agent to scan information locally and then send results back to the management server.

offers additional options for protection — such as automatically routing emails with sensitive data to a manager for approval before letting them out. To achieve this, we just pass email traffic through a DLP email component.

- **Endpoint Software Agent:** All organizations have mobile workers who connect to the Internet from outside the corporate network, such as remote workers at home and sales staff at wireless-enabled coffee houses. A software agent on their computers can monitor and control their network traffic.

In practice, the network monitor and email agent are usually combined on the central management server, but can be split out for distributed organizations and larger businesses. The software agents on laptops protect both data-at-rest and data-in-use, so they don't require multiple packages.

Agents are also the best option for monitoring and protecting files on employee laptops and desktops.

3. **Application Integration:** Some enterprises use document management solutions such as Microsoft SharePoint and EMC Documentum. DLP solutions integrate directly with these tools to take advantage of their capabilities.

Just as we can block network traffic to protect the data, we have enforcement options for data-at-rest. We can automatically move files to a protected server to quarantine them, delete them, or sometimes even encrypt them.

# How DLP Works

## Data-In-Use

As we've already discussed, we use software agents on employee systems to identify, monitor, and protect data-in-motion and data-at-rest. These same agents can also protect data-in-use, such as employees trying to move files to portable storage drives, using sensitive data in unapproved applications, or even printing the data to take it home.

Since laptops and desktops aren't as powerful as servers, and we use them differently, we can't always enforce all the same DLP policies on our endpoints, and not all endpoint DLP tools provide the same features. Usually we need to make trade-offs in the kinds of policies we enforce due to content analysis limitations. Some DLP solutions improve security and performance by using different policies if the endpoint is on the corporate network versus remote.

## Central Management

The central management server is the brains of the DLP solution; it's where we define policies and manage workflow, and in many cases it performs some or most of the monitoring and protection functions. Since DLP is focused on a business problem (protecting sensitive information) as opposed to a technical problem (network attacks) it's important for DLP products to support both non-technical and technical users. Policy creation shouldn't demand a degree in computer science, and incident management should be an activity a compliance, risk, or even human resources manager can perform.

Other key features include robust reporting; integration with user directories to tie policies to users and groups; hierarchical management for multiple systems in large environments; and standard performance, reliability, and management features common to most security tools.

## Understanding Content Analysis

Deep content analysis is the distinguishing and most important feature of Data Loss Prevention. It's how the DLP solution identifies sensitive information, which it then matches against policies for enforcement. Two steps occur in content analysis — breaking apart files to get to the information, and then the analysis itself. There are six main content analysis techniques in use, although their availability and effectiveness vary greatly across different products.

- **Rules/Expressions:** The DLP solution looks for information that matches a defined pattern, such as the structure of a credit card or Social Security Number.
- **Partial Document Matching:** Documents are submitted to the DLP solution, which then looks for complete or partial matches. For example, partial document matching can identify a single paragraph from a protected document pasted into an email, and is very effective for protecting intellectual property.

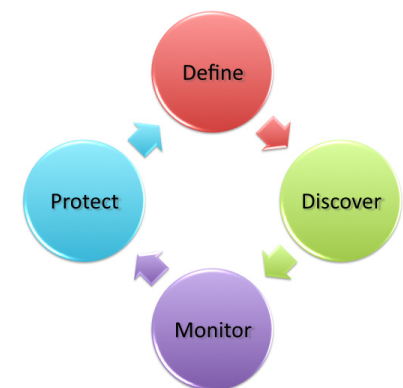
- **Database Matching:** The DLP solution connects to a database, and only looks for content from that database, such as customer records.
- **Statistical:** A large volume of documents is loaded into the system, and the DLP solution uses statistical analysis to identify new content that resembles the sources. This is much more prone to false positives, but may allow you to protect new content automatically, instead of relying on direct matching.
- **Conceptual:** The system uses predefined dictionaries and rules to find content that matches an “idea” — such as sexual harassment, job seeking, or industrial espionage.
- **Categories:** Pre-defined categories for common types of sensitive data, such as financial records or health care information. Often kept up-to-date with current regulatory requirements via subscription.

Most DLP solutions allow you to combine different analysis techniques into a single policy.

## The Data Loss Prevention Cycle

All these features are tied together in the Data Loss Prevention Cycle:

1. **Define:** Build a policy that defines the information to protect, and how to protect it.
2. **Discover:** Use the DLP solution to find the defined information throughout the organization. Relocate or remove information where it shouldn't be.
3. **Monitor:** Track usage of the defined information at rest, in motion, and in use. Generate alerts on policy violations.
4. **Protect:** Quarantine emails, relocate files, block copies to portable storage, and other enforcement actions.



# Pricing Models and Resource Expectations

## DLP Pricing

Data Loss Prevention solution pricing varies by vendor, but tends to fall into one of a few models. Most DLP solutions are priced as per-user annual subscriptions, with a flat rate for the initial hardware/software tiered based on performance. Many DLP providers offer perpetual licensing for the equivalent of a 3-year subscription. Support and maintenance fees are normally calculated at 10-20% of overall contract value per year, depending on the level of support desired. Some vendors split pricing based on which DLP components are deployed (network, endpoint, and discovery) for organizations which wish to phase in their deployments. All vendors offer volume discounts.

Street pricing varies widely, and the following examples are just rough estimates to give you an idea of what to expect. For a mid-sized organization of 5,000 employees, pricing for only network monitoring usually costs in the neighborhood of \$100,000, while a full-suite deployment (with endpoint, discovery, multiple systems, and infrastructure integration) can grow to \$300,000-\$500,000. Larger organizations can expect the customary significant volume discounts.

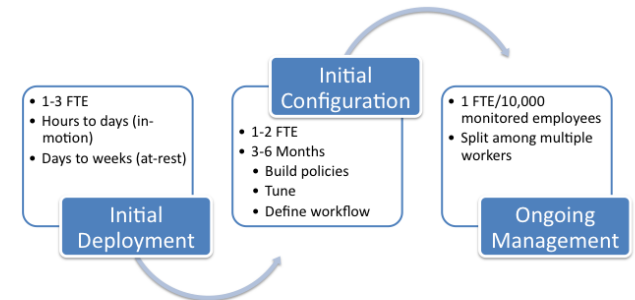


## Resource Expectations

As with pricing, resource requirements vary greatly based on the nature and scale of deployment.

- **Initial Deployment:** Network monitoring and remote scanning (for data-at-rest) deploy very rapidly, typically in a matter of hours for smaller deployments, and no more than a few days or weeks for larger deployments with multiple monitors, using 1-3 Full Time Equivalent (FTE) employees. Very complex networks with additional integration requirements require additional time. Endpoint agent deployment on employee systems and servers varies more, depending on existing system infrastructure and management. This typically demands no more effort than any other piece of software, with little initial configuration, since tuning occurs later on the central management server.
- **Initial Configuration:** This includes development of initial policies, integration of the system with existing infrastructure such as user directories, and testing and deployment of initial policies. Although basic policies can be deployed in days, expect to spend 3-6 months tuning the initial set, defining workflow, and building processes to manage incidents. Mid-size organizations typically require a single FTE to manage the system and develop policies, with additional part-time FTEs to manage incidents.

- **Ongoing Management:** A mid-sized organization with a basic deployment typically requires only 1-2 FTEs for ongoing system management, policy creation, and incident handling. This is often split across a group of employees working part time with the system, with one person handling system management and policy creation, and a group of security, compliance, risk, legal, and/or human resources employees dealing with incidents. As deployments scale, we typically see .25-1 FTE per 10,000 employees, again depending on the number and nature of policies deployed. Some mid-sized organizations find they use considerably less than a single FTE.



## Summary

Data Loss Prevention is a powerful tool for protecting information assets. DLP identifies sensitive information, monitors its use, and protects it from abuse. DLP tools accomplish this by monitoring your network, scanning your storage infrastructure, and tracking data use on endpoints such as laptops

and desktops through deep content analysis, which allows you to define what information to protect and how to protect it. DLP isn't perfect, and can't stop all misuse of information or malicious attacks, but is a powerful tool for reducing compliance costs and the risks of information loss and data breaches.