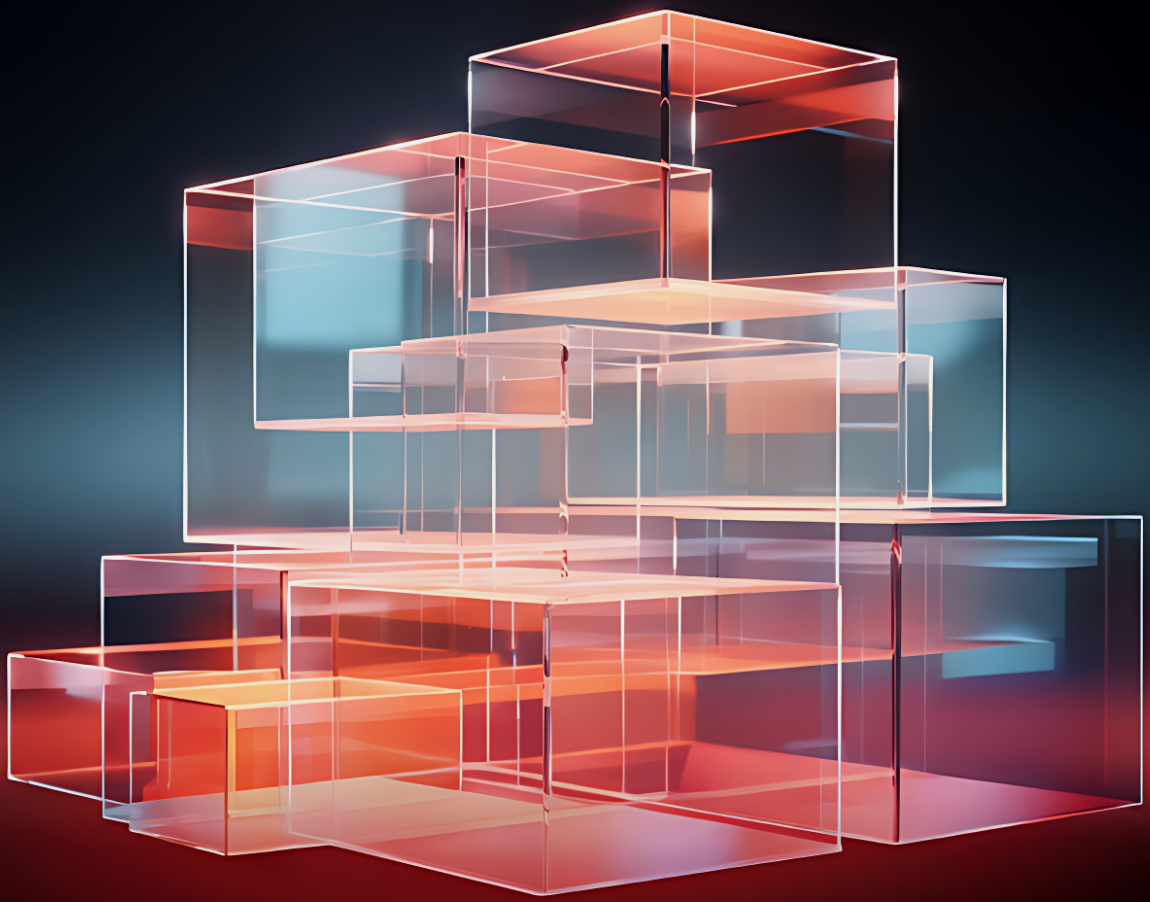
cyberhaven

DLP Buyer's Guide

11 Criteria for Evaluating Data Loss Prevention Solutions



The technology we use is evolving, and DLP is evolving too

The technology that people use in the workplace is changing, and with that change comes new risks to corporate data. Today, data is constantly moving between SaaS systems of record, computers, messaging applications, and collaboration tools. Data often spreads to people who wouldn't have access to it at the source, and there are new ways people can exfiltrate it beyond the reach of the company. The problem of "data loss prevention" is changing.

And in response to this evolving risk landscape, data loss prevention technology is changing as well. Whether you're buying a DLP product for the first time, or replacing an older solution, this guide summarizes what to look for. There are new DLP capabilities designed to help with the risks to data that didn't exist a few years ago, and also improvements to DLP technology that fix the shortcomings of the products that came to market last decade.

> "Today, the data loss prevention market is evolving to address the limitations of traditional approaches, which relied heavily on content inspection capabilities that are resource-intensive and often lead to performance issues with high numbers of false positives."
>
> Ravisha Chugh and Andrew Bales
> Gartner, "Market Guide for Data Loss Prevention", September 4, 2023

New data loss vectors are emerging

Data is leaving your company in ways that didn't exist a few years ago. Direct device-to-device transfer technology like Apple AirDrop and Windows Nearby Share make it possible to move large amounts of data quickly from a work laptop to a personal computer. Generative AI tools like ChatGPT are incredibly powerful for rewriting or summarizing content, but because they learn from input they can also expose whatever you paste into them to other users.

Network tools are blind to new encryption

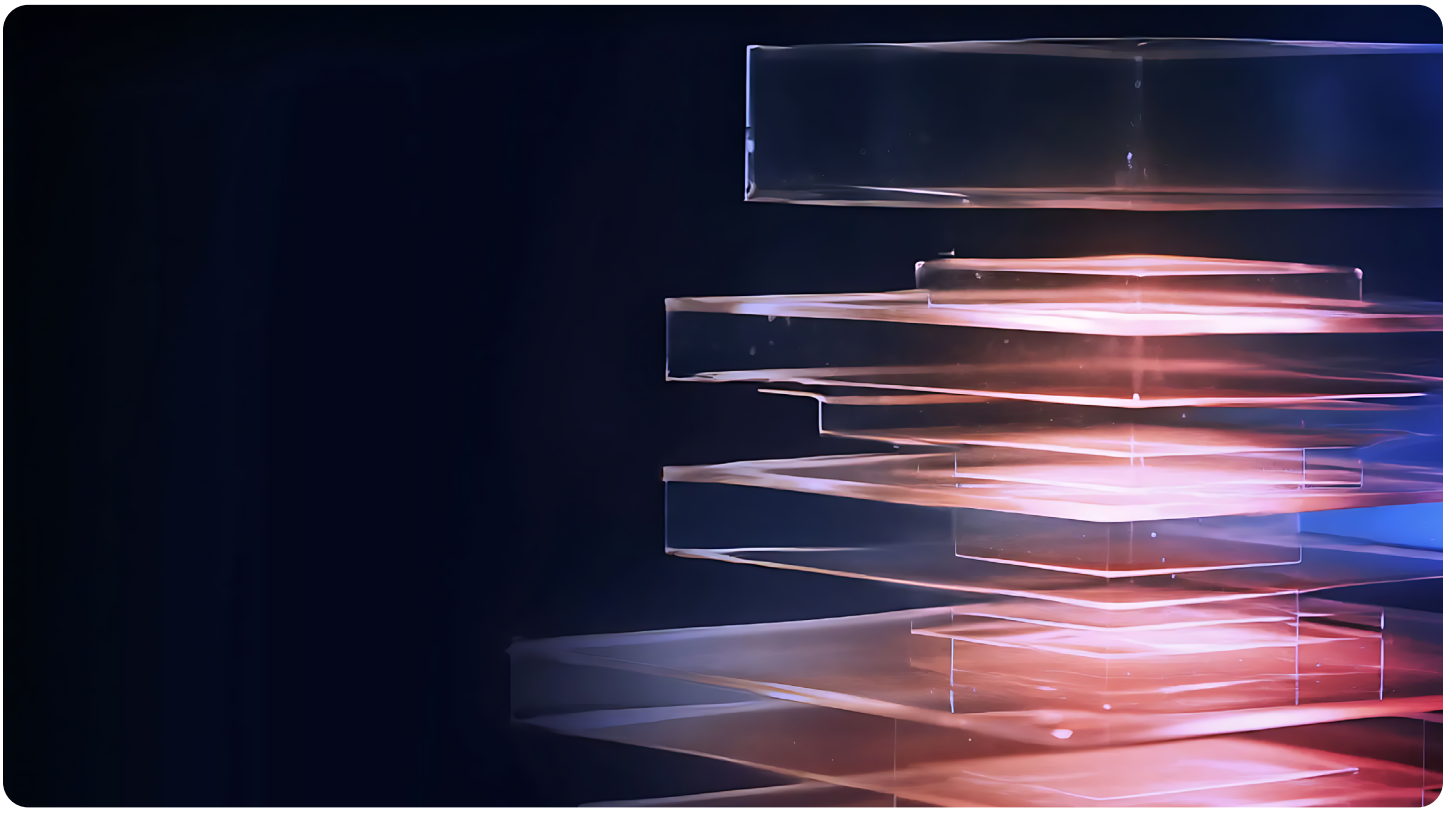
Web proxies and cloud access security brokers (CASB) were designed to inspect data leaving a device to the cloud by decrypting SSL/TLS traffic. But an increasing number of cloud apps are adding certificate pinning and end-to-end encryption to prevent man-in-the-middle attacks and surveillance. While improving one area of security, these technologies have the side effect of making it impossible for network-based security products to decrypt their traffic.

Personal instances of corporate apps

In many cases, the collaboration tools that companies allow (even encourage) employees to use are also available as consumer products. Companies want employees to work using their corporate Google Drive, OneDrive, or Gmail accounts, but don't want employees putting corporate data in their personal accounts for these same services. There's a need to enforce distinct policies for personal and corporate accounts for the same app.

Data security is spread across multiple security tools

Some companies use multiple DLP products to cover different exfiltration vectors–for example an email DLP product and a separate endpoint DLP product. While sold by one vendor, the products may have been built by two startups acquired by the vendor and have different policy engines and management interfaces. The situation can get even more complicated once you add in a CASB for cloud DLP and data security features built into a platform like Office 365.



The challenges with traditional DLP

- Data classification based on keywords or RegEx patterns leads to false positives
- Important data without a content pattern or text content is not classified or protected
- Only shows you data and repositories you know to look for (and create a policy for)
- User-applied tags and labels are inconsistent and don't follow data copied out of a file
- Software agents installed on end-user devices cause poor performance and system crashes
- Policies are complicated to create and manage, sometimes requiring JSON and custom code

> "The false positive rate of using RegEx is through the roof. You look for the word 'classified' in a document and get news articles containing that word. False positives sink many DLP programs."
>
> Arlan McMillan
> Chief Security Officer, Kirkland and Ellis

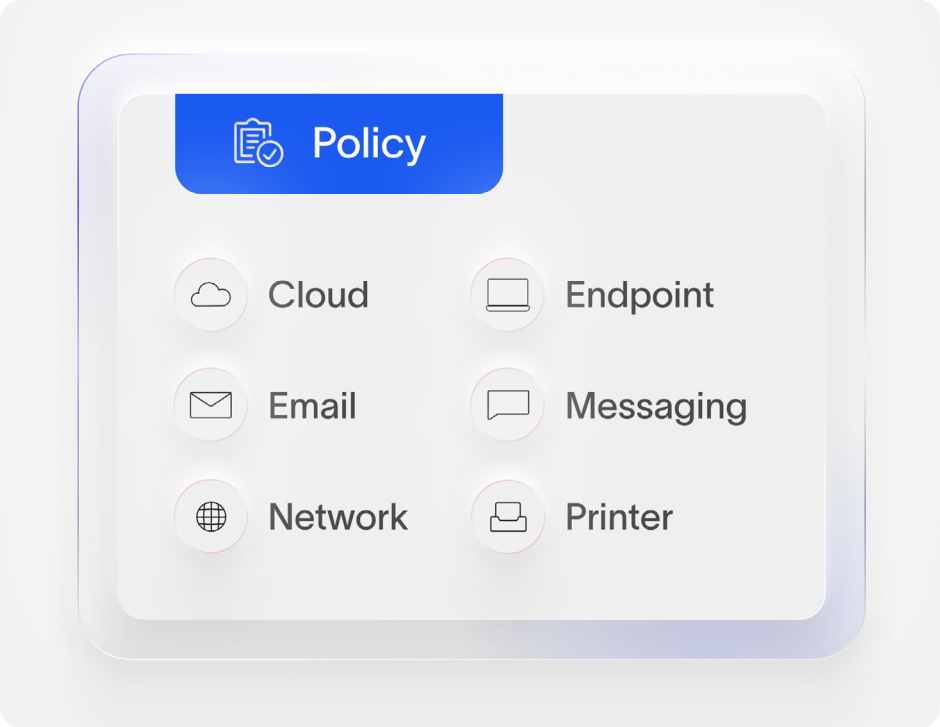
1

Criteria 1: One product, one interface, one policy engine for all exfiltration channels

Data exfiltration can happen in any manner: an employee can copy a CAD file from a laptop to a USB drive, print a strategy document in Google Drive, or transfer source code to an unsanctioned cloud application. Regardless of where data exposure occurs or how it happens, you want eyes on this activity and a consistent set of policies in place to ensure that the incident doesn't escalate into a data leak.

In its current state, DLP is fragmented. You deploy different tools to protect data in cloud applications, email, and on endpoints. Even assuming you have a SIEM to bring incident alerts together under one view, multiple tools mean disparate policy engines with different features that don't consistently detect security incidents. The resulting set of piecemeal solutions means each of your environments can have differing levels of policy coverage, ultimately increasing the overhead and costs of your security program while hampering its effectiveness.

When it comes to modern DLP, look for solutions that are comprehensive, providing a centralized location and a single source of truth for reviewing all the activity happening within your organization. You likely won't be abandoning your SIEM, but you will have a single place to apply policies across your organization, with the only thing that matters being the incidents impacting your data, and not where these incidents took place.



Features to look for:

- One DLP policy engine that can enforce a standardized set of policies across all exfiltration channels.
- One interface to create and manage policies, data classification categories, exceptions, users, and watchlists.

How to leverage this functionality:

- Set up policies to stop data egress everywhere, for example create a policy to prevent customer data from leaving via these exfiltration channels:
  - Corporate cloud environments to external users
  - Corporate email accounts to external users
  - Web uploads
  - Personal SaaS accounts or non sanctioned applications
  - Copy/paste
  - USB storage devices
  - Printing
  - Bluetooth or peer to peer protocols like AirDrop
  - Encrypted communication channels like Signal and WhatsApp

2

Criteria 2: Augment content analysis with data lineage for more accurate classification

Until recently, DLP products have relied on content inspection to automatically classify data. But using keywords and regular expressions (RegEx), DLP products are only able to accurately classify a limited range of data types. Outside of those data types, content-based classification has two problems:

1. Common RegEx patterns trigger on sensitive data but also lead to false positives

2. Many forms of sensitive data cannot be identified at all using content

Content inspection excels at identifying data with a very specific alphanumeric pattern like a credit card number. Not only are credit card numbers always 16 digits long, they can further be validated using a Luhn algorithm. Other alphanumeric patterns like Social Security numbers that don't have a checksum are harder to detect without false positives. And then there are content patterns that are sensitive entirely depending on context. Names, phone numbers, and emails are sometimes customer data and sometimes not sensitive at all.

The key piece of information that legacy data classification schemes relying on content inspection are missing is context. Simply knowing that a file contains names and phone numbers, for example, doesn't tell you whose names and numbers these are. But knowing that contact information originated from a customer record table in Snowflake, or in Salesforce, where the organization keeps such records suddenly changes the equation. With this additional detail, a piece of data flagged because of its content can more accurately be classified.

Data lineage is the technology that provides this additional context. Data lineage tracks data within your company to build an understanding of the history of the data including where it originated, how it traveled within the organization, who edited it, and more. This context not only reduces the false positives of content-based classification, but it also enables the classification of data that is traditionally much harder for content inspection to meaningfully identify.

Some types of sensitive data cannot be identified using content patterns, either because the nature of their content means they don't contain a consistent pattern (e.g. go-to-market strategy documents, investment plans, and financial reports) or because they contain no text content whatsoever (e.g. recorded internal meetings, unreleased TV episodes, product design files, marketing images). Here, data lineage can accurately classify sensitive data based on where it originated, how it was handled or stored, and who edited it.

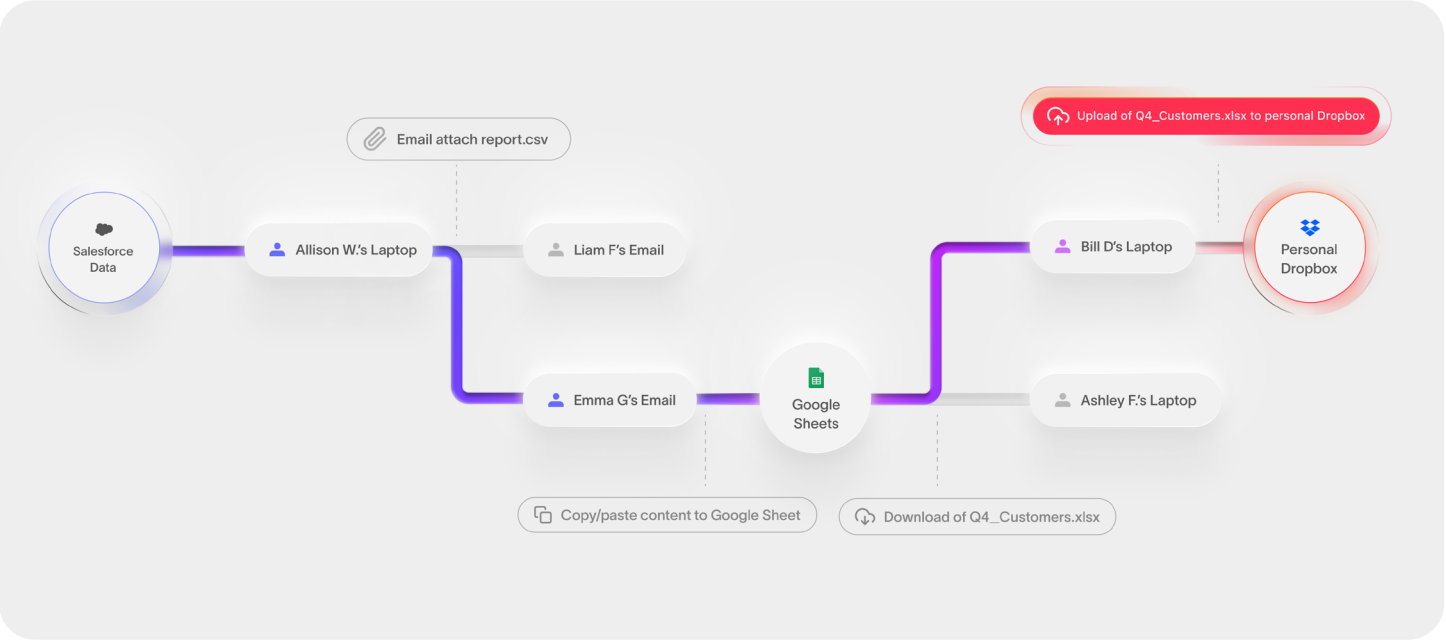
What to look for:

- A DLP platform that integrates data lineage as a core functionality, tracking data from its origin through each step it takes within the organization.
- Policies that support combining content and context-based rules in order to better classify data and reduce false positives.
- Ability to classify data purely based on data lineage attributes to cover data that contains no consistent pattern or no text content whatsoever.

How to leverage this functionality:

- Classify data by combining content-based detections with contextual factors:
  - Where the data originates
  - How it moved throughout the organization
  - Who edited or added to the data

> "Invest in a DLP solution that can understand the full lineage of the data, identify baseline activities for the user, and compare subsequent actions to the baseline activity by gathering contextual clues about the who, what, when and where of the data."
>
> Ravisha Chugh and Andrew Bales
> Gartner, "Market Guide for Data Loss Prevention",
> September 4, 2023



3

Criteria 3: Rapidly calibrate policies by testing on historical events

Data loss prevention projects have suffered from a long time to value, in part because they required your security team to first learn what data you have, where it's stored, and the ways in which your organization is currently using it. Only then can you craft what your policies should be and build these policies within the DLP tool to flag risks to data or stop data exfiltration. DLP products further lengthened this process in the way they applied new or updated policies.

With older DLP products, it can take weeks or months to create and tune a policy because policies only apply to new user actions. After you change a policy you have to wait to see how well the policy works as new events come into the system. If you set up the policy in a way that leads to false positives or misses risky events altogether, it takes awhile to see this, then update the policy, then wait to see how the updated policy works as new events happen.

But today, DLP products eliminate this lengthy testing period by showing you a preview of the incident alerts that would be generated from a policy change. This preview works by applying a new policy across historical user actions, and shortens the iteration that may be needed with crafting a policy from weeks down to minutes. This workflow is made possible by DLP products that keep a record of all events for each piece of data.

Beyond making it easier to test new policies, storing all these events also opens up the ability for you to browse or explore what data exists at your company and how and where it flows through the organization. Armed with this visibility, you can more effectively work with the business to craft better policies for data and risks you may not otherwise know about.

> "I've used traditional DLP technologies in the past and sometimes the noise-to-signal ratio can be a lot. The context Cyberhaven gives us has significantly improved our data protection, monitoring of data movement, and insider risk."
>
> Prabhath Karanth
> VP and Global Head
> of Security & Trust
> Navan (formerly TripActions)

What to look for:

- → A DLP platform that stores all events for every piece of data at the organization.
- → Ability to preview the results of a policy against historical events.
- → Visibility into all data, where it's stored, who accesses it, in order to better craft policies.

How to leverage this functionality:

- → To develop an understanding of what data exists and risks to that data, explore:
  - Who is using data?
  - What purpose are they using data for?
  - What does business-need look like for acceptable data use policies?
- → Testing policies against historical events to preview what alerts or preventative actions would have been taken and quickly iterate on the configuration of a policy.

4

Criteria 4: Combine data classification with behavior analysis to better detect insider threats

> "DLP vendors are increasingly converging with insider risk management platforms. This convergence enables better detection of data exfiltration as it enriches DLP events with anomalous user behaviors, improved risk scoring and real-time monitoring capabilities."
>
> Ravisha Chugh and Andrew Bales
> Gartner, "Market Guide for Data Loss Prevention",
> September 4, 2023

Data loss prevention and insider risk management are two security products that have approached the same problem in different ways. They both aim to limit risk to a company's important data from careless or malicious insiders. DLP products focused on classifying data and controlling its movement but didn't understand user behavior. Insider risk products analyzed behavior like data upload volume but didn't understand what type of data was being uploaded.

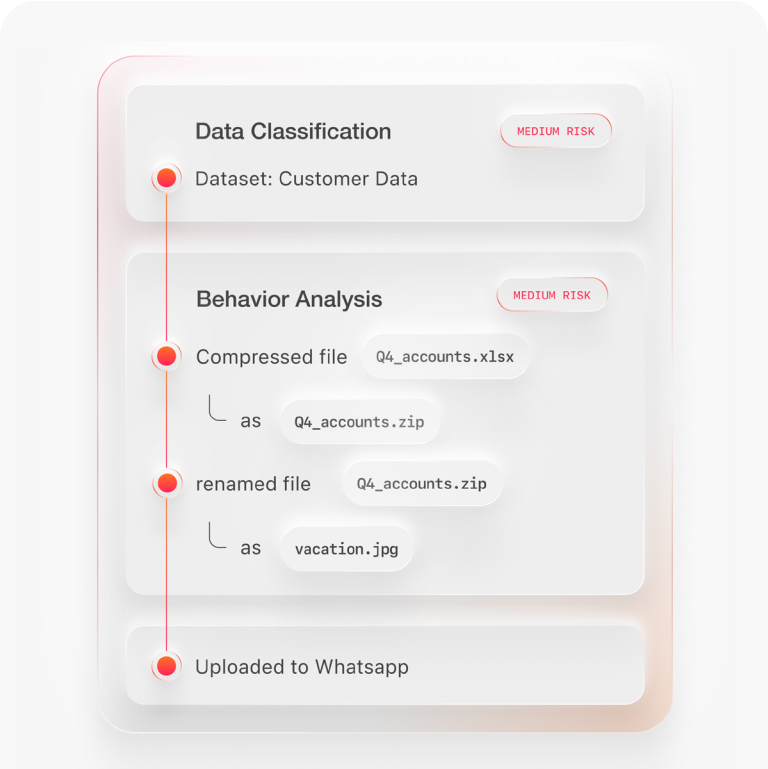
These two product categories are converging into one. While it's possible to merge signals from separate DLP and IRM tools together in a SIEM, doing so only improves the alerts you receive. In order to take action to protect data when it's at risk, a single product and policy engine needs to correlate both the data and behavior together, along with other risk factors like the user's risk score, in order to take action and prevent exfiltration.

What to look for:

- Ability to combine data classification and user behavior signals within a single policy.
- A user-centric view of risk and mishandling of data, and patterns over time.
- Risk scores that incorporate behavior and data attributes, and not just behavior alone.

How to leverage this functionality:

- Review risk scores to understand how users are affecting data exfiltration risk over time.
- Identify high risk users and add them to a watchlist with stepped up policy enforcement.



5

Criteria 5: Distinguish between personal and corporate instances of cloud apps

As cloud adoption has taken off, companies have had to address the risk of employees putting corporate data in unsanctioned applications that they don't control. Either by using a CASB to limit data movement to unsanctioned applications, or blocking access to personal cloud applications altogether with a web proxy or firewall, many companies have limited this movement. A trickier issue is employees using personal accounts of sanctioned applications.

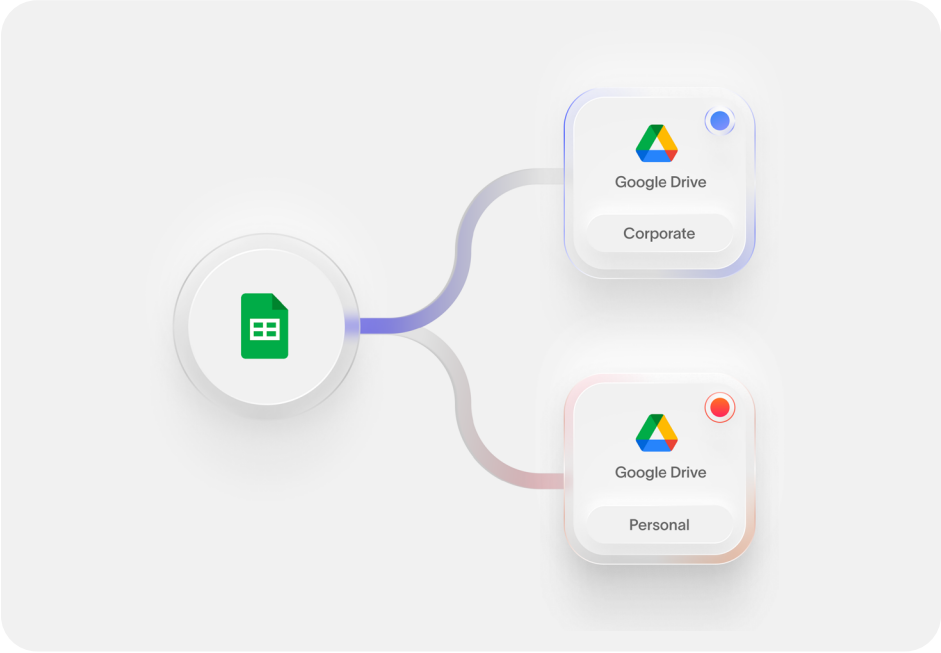
Many cloud applications that companies use today also have consumer versions; OneDrive, Google Drive, Gmail, Slack, and Box to name a few. Sensitive company data that can (and should) be stored in an employee's corporate account should not find its way to an employee's personal account for these same services. Prioritize DLP solutions with the ability to detect which type of account a user is logged in with and enforce policies based on that.

What to look for:

- Ability to distinguish between the corporate and personal accounts users are logging into to access a website or application.
- Ability to create and enforce distinct policies based on whether the account is personal or corporate.

How to leverage this functionality:

- Compile a list of corporate sanctioned applications and the types of data that can be stored in them.
- Define policies that limit movement of company data to personal accounts of sanctioned applications.



6

Criteria 6: Monitoring data across cloud and on-prem environments

Data tends to flow within organizations between different systems, devices, and users. But until recently, data security tools only covered data across a limited number of assets–CASB for cloud applications, DSPM for cloud infrastructure, DLP for on-premises file servers, DLP for endpoints. Whatever controls these products have, they lose visibility and control once data moves somewhere else.

Consider the example of data moving from Snowflake to a data scientist's local machine for analysis. Most cloud-focused data protection solutions don't see what happens to data once it leaves Snowflake. And for a standalone endpoint solution, this data effectively comes into existence the moment it's downloaded onto the local machine, mostly detached from its context as a copy of data created, managed, and owned within Snowflake.

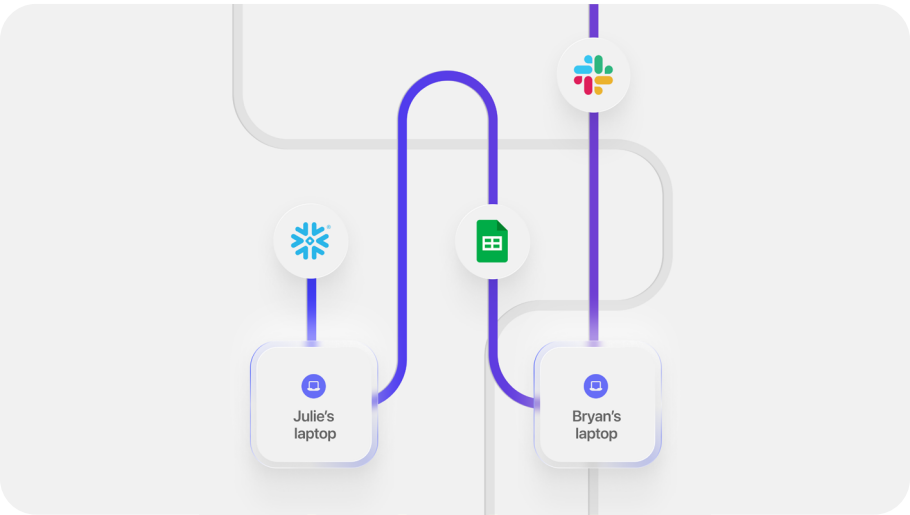
Prioritize DLP solutions that monitor movement of data between different environments within your extended enterprise and apply the same set of security measures to data, regardless of where it is (or even based on context for where it was previously).

What to look for:

- Ability to follow data between cloud and on-premises environments and endpoints.
- Tracking of derivatives or exports of data from one system to another.
- Ability to enforce consistent controls even as data moves between different assets.

How to leverage this functionality:

- Look up where copies of data from one asset have spread–e.g. Show all copies and derivatives of data from the customer table in Snowflake.
- Create policies that treat data the same across systems—e.g. Users who cannot access data in one system should not have a copy of that data in another system or device.



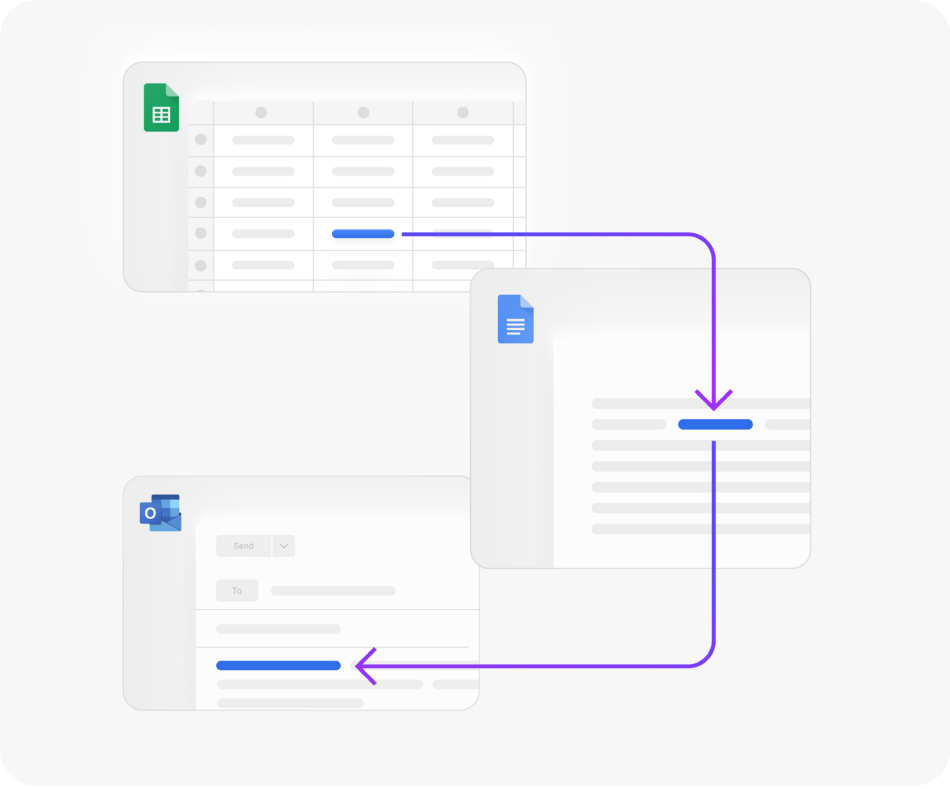
7

Criteria 7: Tracking and securing data as it moves between apps and files

Ultimately, the purpose of DLP is to protect not just files but sensitive data in whatever form it takes. DLP solutions that tag or label files only partially cover important data. A lot of data within organizations today isn't stored in a file—it's in SaaS applications, messaging applications, or structured databases. A user can copy/paste content from one application to another, or they can copy/paste data out of a file, in which case any tag or label doesn't follow the data.

For this reason, look for DLP solutions that have data classification not at the file level, but at the level of individual data. For example, if someone copies a paragraph from a company press release under embargo, and pastes it into WhatsApp, this should immediately set off red flags, even if the incident doesn't involve the movement of a file. Similarly, if a customer name and address are copied from a record in Salesforce and pasted into a spreadsheet that can be viewed by the entire company, this should be part of the key context surrounding the incident.

Moreover, DLP solutions that leverage data lineage as part of their data classification scheme need to do so granularly, rather than applying this data solely at the file layer. Doing so is critical because data can move more freely than files, and there are a variety of incidents that can expose sensitive data without involving the movement or sharing of files.



What to look for:

- Data lineage that tracks data's origin and history at a granular level, like when data moves from a file to another file or SaaS application, or from one application to another.
- Ability to persistently track and apply classification based on origin or tags/labels for data that is stored in a format that's not a file, like in SaaS applications or messaging.

How to leverage this functionality:

- Create policies with data classification based on origin or user-applied tag/label, even after the data is copied out of the file.
- Create policies that block the pasting of sensitive information into messaging applications like Slack, or web apps like ChatGPT.

8

Criteria 8: Uniform coverage for all files and data types

Another limitation that security teams can encounter with DLP tools is that they don't provide uniform coverage for all types of data and files. This is especially true of complimentary, out-of-the-box DLP services offered by SaaS providers like Google and Microsoft, via Workspace DLP and Microsoft Purview license, respectively.

These types of solutions are best suited to scan file types native to their respective platform and cannot apply scanning or classification to hundreds of other file types. For example, Microsoft's DLP solution can't classify file types like CAD files, Figma files, source code, and more. This limitation ultimately hampers the comprehensiveness of these solutions and creates coverage gaps as you cannot apply policies consistently across file types.

What to look for:

- A DLP platform that can classify and protect a comprehensive set of file types, as opposed to just standard Microsoft Office files.

How to leverage this functionality:

- Create and enforce policies across a comprehensive set of file types, not just some file types that support proprietary labels.

9

Criteria 9: Real-time user coaching while allowing employees to override

Most employees are not deliberately malicious, with the lion's share of data exposure incidents in many organizations occurring through unintentional mistakes. If you see an employee putting company data in an unsanctioned app, like a free PDF converter website, it might be that they're simply trying to be productive and there either isn't a company-provided option or they don't know about the tools the company has made available to employees.

There are limits to quarterly or annual security training for employees. Many people learn better when they get real-time feedback. To address this reality, look for DLP solutions that coach users on appropriate behavior with just-in-time popup notifications as incidents happen. Providing employees with these real-time lessons is more impactful than recurring security training sessions that often fail to contextualize security risks in a tangible way for employees.

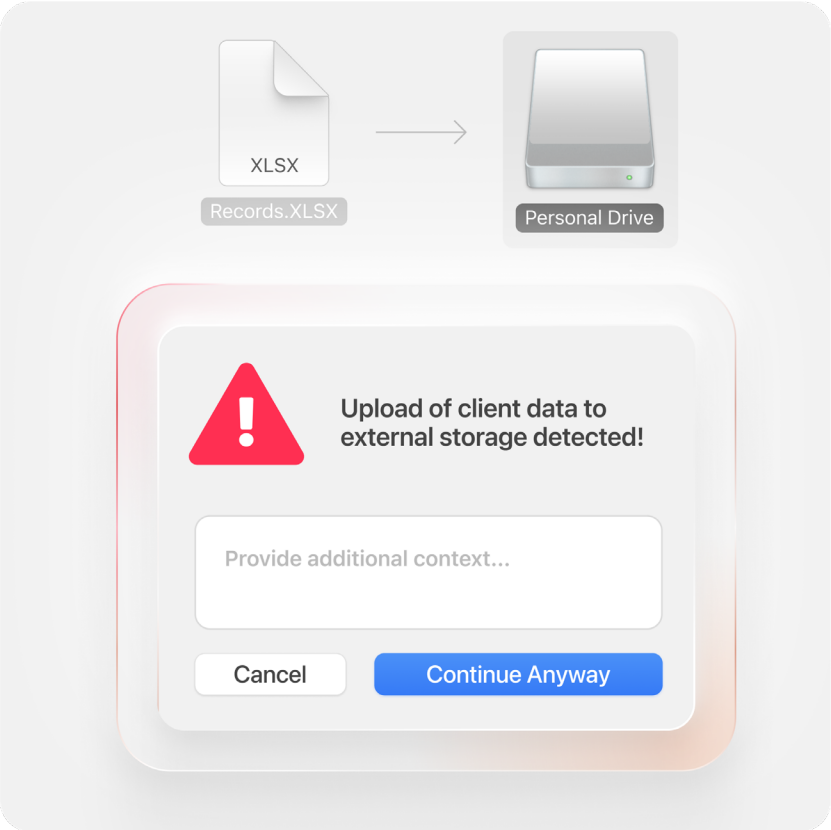
Alternatively, if users can provide a justification for their actions, you may want to allow them to do so and continue their work uninterrupted. Simply blocking the user without giving them any recourse can introduce friction into employees' workflows. And there are always approved exceptions—like when your client requests you upload documents to their Dropbox account (even though Dropbox is unsanctioned by your company).

What to look for:

- Support for just-in-time notifications to end-users that occur on the device to cover all exfiltration cases, not just within an application like Slack or in the web browser.
- Ability to have notifications with or without blocking enabled, so that even if you don't prevent an action the employee can be coached.
- Ability to give employees the option to provide a business justification and override the policy, based on policy and risk of their action.

How to leverage this functionality:

- Create policies for notifications based on risk. For example:
  - For low severity events just display a notification
  - For medium severity, block the action but in the popup allow user override
  - For high severity, block the action without option to override
- Track changes to user behavior over time to report on risk reduction. Real-time notifications can quickly shift user behavior even if the action isn't blocked.
- Review exceptions requested by employees and identify gaps in IT technology; tools the company should invest in to better enable employees.



10

Criteria 10: Full picture of incidents to accelerate investigations

Most DLP solutions have surfaced alerts to security teams without context. You might get an alert showing that a user uploaded data they weren't supposed to, but to understand the incident you needed to search for more events from the DLP solution or another tool. But today, DLP solutions can provide much more context that helps security teams quickly understand user intent and reduce the time for conducting incident investigations.

Look for DLP solutions that show all of the actions a user took against a piece of data before and after the incident. This is invaluable for determining if a user deliberately tried to take data and obfuscate their activity. For example, if an employee attempted multiple times to exfiltrate sensitive data, trying different exfiltration channels and being stopped each time, this shows a pattern of behavior to remove sensitive data from the company. A user who changes the file extension of a CSV to a JPEG to make it look like they were sending an image, or putting sensitive data in a ZIP file, could indicate malicious intent to hide their activity.

This context can also help you determine if the individual triggering the alert should even have access to the data. If they aren't supposed to, then you could see how they obtained the data in the first place. Maybe an employee uploaded a copy of a Workday export to Google Drive, where its permissions were too broad and another employee who wouldn't have access to the data downloaded it. Understanding the data's journey through the company can also reveal a second person colluding to steal sensitive data by passing it to someone else.

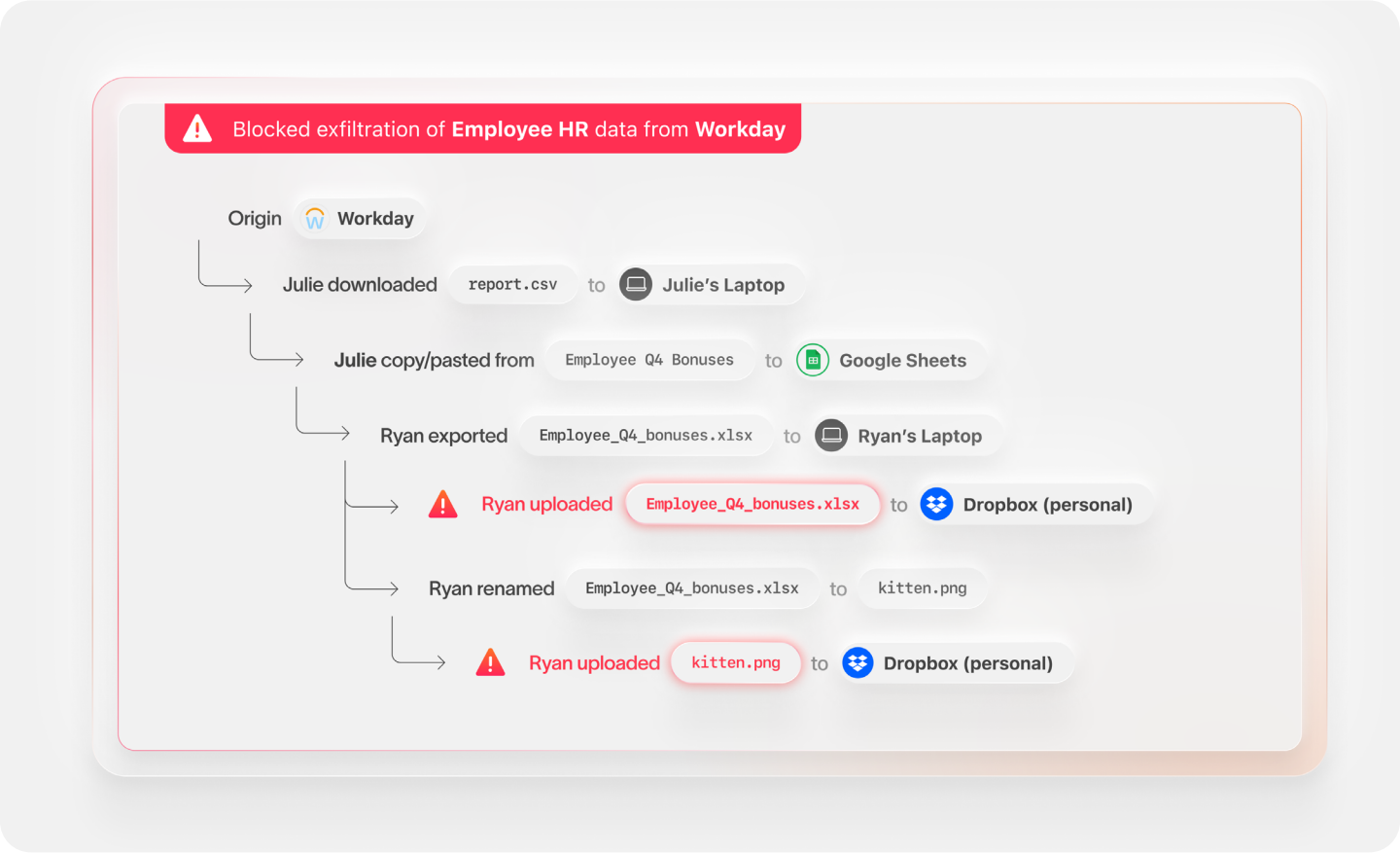
> "Data lineage helps tremendously when doing a security investigation and we need to understand the events leading up to attempted exfiltration, if there were more people involved and how data was sent between them, and how someone got ahold of data to begin with."
>
> Dan Walsh
> CISO
> VillageMD

What to look for:

- An incident view that automatically shows the events for a piece of data before and after the triggering event that generated the alert.
- Ability to see all the way back to where the data originated and its path through to company to the individual who attempted to exfiltrate it.

How to leverage this functionality:

- Review the events before and after the triggering event to understand the user's intent—look for attempts to obfuscate or repeatedly exfiltrate data.
- If the person in question shouldn't have access to this data, look for how they got ahold of a piece of data. This could reveal things like incorrectly set permissions.
- If the user was sent the data from someone else, look for patterns where the same person is sharing sensitive information that could indicate collusion.



11

Criteria 11: Modern, cloud-based platform

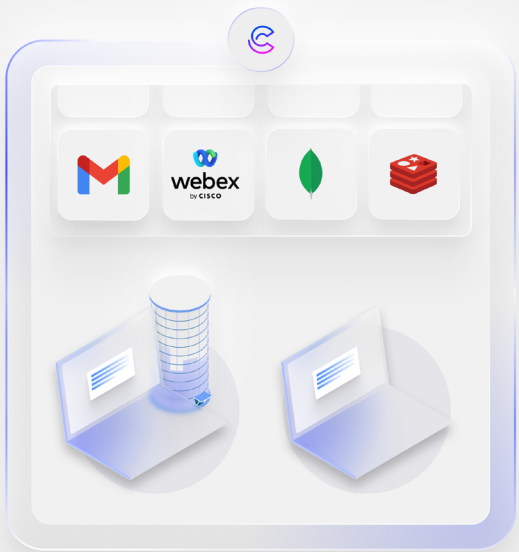
If you're upgrading from a DLP solution, you may be happy to leave behind the day-to-day management of application servers and database servers that power your old DLP tool. Today, DLP products are built on modern, cloud-based architectures that simplify management. Moreover, they come with powerful privacy and security features to give confidence that the data you protect using the DLP product is secure and protected in the cloud.

What to look for:

- DLP platform where the processing, policy management, and interface are delivered via the cloud.
- Option to host copies of data stored for rendering match highlighting and forensic file capture within your own cloud infrastructure, not the vendor's.

How to leverage this functionality:

- Offload the day-to-day management of your DLP application to your vendor's cloud operations team.
- Configure storage in your own cloud account for files and data stored for forensics and incident review.



Cyberhaven is a 2023 Gartner Cool Vendor in Data Security

Cyberhaven's approach to detecting and mitigating both insider risks and data loss events, and its provision of broad visibility into user behavior, positions its product as a next-generation alternative to traditional DLP solutions.

Gartner, "Cool Vendors in Data Security", August 8, 2023

About Cyberhaven

Cyberhaven is the data security company revolutionizing how companies protect their most important information from theft and misuse. Until now, security products only recognized and protected a limited range of data types because they relied on finding patterns in the content itself. Our data tracing technology analyzes billions of events surrounding every piece of data to better understand and classify it, allowing for protection of a much broader range of sensitive data in any form, anywhere it goes.

To learn more about Cyberhaven, visit cyberhaven.com.