The Business Case for Cloud Data Protection

Nightfall AI
www.nightfall.ai

Table of Contents

About Nightfall AI

Our Vision

We stand for a modern, holistic approach to cloud data protection - understanding and mitigating data-related risks across cloud applications should be simple. Nightfall makes safeguarding sensitive data for cloud applications easy and seamless.

We're reimagining data security and compliance through a cloud-native, accurate, and performant platform. Our mission is to bring efficiency and efficacy to data protection. We partner with organizations and software platforms to ensure the sensitive data that people entrust them with every day is secure and found only in places where it should be.

We envision a future where Nightfall powers data protection for every app. When data security and compliance just work, both organizations and consumers are safer. Nightfall's one trusted platform provides organizations with a simpler way to discover, classify, and protect sensitive data, while accelerating innovation and minimizing risk.

Who We Are

Nightfall is a cloud-native data protection platform that integrates directly with cloud apps and data infrastructure to discover, classify, and protect sensitive information. Typical use cases for Nightfall include data classification, data leak prevention, and content moderation. Nightfall has pre-built integrations for popular cloud apps like Slack, GitHub, Google Drive, Confluence, and Jira, and also offers the ability to build custom integrations with any product via our Developer Platform.

With Nightfall:

- Identify where you have risks related to sensitive information exposure in the cloud.
- Protect sensitive information, for example by removing it from where it doesn't belong.
- Add content inspection and remediation capabilities to any third-party application, without agents or proxies.
- Leverage machine learning (ML)-based detection for smarter results, with high accuracy, easy tuning, and fewer false positives.

- Configure and customize data detection policies that meet your organization's unique needs, such as allowing data to exist in one location but not another.

Nightfall provides 150+ ML-based detectors out of the box to identify and remove a wide range of sensitive data, helping organizations to implement a holistic approach to data stewardship across their cloud ecosystem, and to comply with HIPAA, PCI, SOC, CCPA, and client requirements. Nightfall classifies 10M+ instances of sensitive data per month, and works with high-growth SaaS companies like Amount, Prove, UserTesting, and Sisense. Nightfall is backed by Bain Capital Ventures, Venrock, Webb Investment Network, and a cadre of high-profile operators, including CEO/executives of Okta, Splunk, FireEye, and Salesforce.

Why now?

Organizations are rapidly adopting cloud SaaS and infrastructure, and this is putting a strain on information technology and security teams. The average organization leverages 1,935 cloud services (McAfee Cloud Adoption and Risk Report, 2019), many of which connect with each other to allow users to collaborate and share information easily. Even traditionally on-prem vendors like Atlassian are beginning to migrate customers to the cloud - Atlassian has stopped selling new licenses for their on-prem server products and announced they will end support for existing customers by February 2024.

Due to the ease of information-sharing, data sprawls rapidly across the cloud ecosystem, and security teams are scrambling to keep up with implementing holistic data stewardship policies that account for this new frontier. What many organizations don't realize is that sensitive information is commonly shared (even if accidentally) within these third-party apps - things like personally identifiable information (PII), protected health information (PHI), financial identifiers, passwords, and more. Without internal policies and processes in place, organizations must rely solely on third-party services to protect this sensitive information, even as users continue to inadvertently proliferate the information.

At the same time, the lines segmenting work and personal IT are blurring - end users are now bringing their own devices and connecting over their own networks, especially in a post-pandemic world. This shift to a distributed workforce has rendered legacy network- or endpoint-based data security solutions ineffective. And many modern companies would prefer not to engage in cumbersome, invasive, resource-intensive deployments to begin with.

Sensitive data in the cloud should be considered at risk

In today's increasingly digital world, regulators are ramping up the focus on data privacy, with growing regulation and pressure to identify how data is collected, processed, and stored. Regulations like CCPA and HIPAA, and compliance certifications like SOC are setting the tone for more regulation and standardization relating to data privacy and security. Additionally, regulatory bodies are making clear that fines and penalties aren't empty threats. It is your organization's responsibility to steward your customer and employee data responsibly, and ensure that sensitive data is not stored where it doesn't belong and where it could potentially be exposed. Every day without a holistic cloud information management process in place is another day that sensitive information is left at the mercy of your various cloud vendors' security policies.

Organizations that leak sensitive information are not let off the hook just because a cloud application was involved in the leak, as demonstrated by the following case studies.

Capital One, 2019
Misconfigured S3 bucket exposes 106 million customers' data and results in $80 million in fines

The 2019 Capital One data breach was the coming together of several unfortunate developments. The first was a misconfigured web access firewall (WAF) in Capital One's AWS environment. The second was that the attack was carried out by an AWS employee, which might have given them a pretty good understanding of how exactly to carry out the attack.The WAF, which monitors traffic between web applications and the company's environment, was running in an EC2 instance that had read access to any files in the S3 bucket storing Capital One's data. It's possible that the firewall might have not been configured properly or the hacker simply bypassed it, allowing for a server-side forgery request to be used to access resources in Capital One's S3 bucket. Capital One paid an $80 million fine for the breach. Misconfigurations like this are commonly responsible for breaches. These can be hard to detect without tools like Cloud DLP which provide visibility into cloud environments with sensitive data.

Uber, 2016
$148 million in fines resulting from leaked GitHub credentials.

In the fall of 2016, Uber was revealed to have covered up a data breach impacting 57 million users and 300,000 drivers for over a year. Uber was targeted by two hackers who schemed to

hack tech companies and extort money from them through their bug bounty programs. Using a custom-built tool the hackers tested a stolen cache of user credentials to validate if they could be used to access GitHub accounts used by corporate employees of target companies. The hackers had success accessing the codebases of several companies, including LinkedIn's Lynda.com. Once in these codebases, they searched for hardcoded credentials that would let them access sensitive backend systems. In the case of Uber, they were able to successfully access the company's AWS account and reached out to Uber shortly after. Uber paid the hackers $100,000 for a promise to delete the data. Breaches like this one illustrate the difficulty of ensuring that secrets don't surface within repositories, even ones that are private. A tool that can scan code repositories for credentials and PII is important to mitigate such risks.

Modern-day problems: content moderation.

In today's modern organization, the risks of information proliferation aren't simply limited to "traditional" sensitive data. Employees now collaborate more than ever via written chat messages and email, and these indirect interactions, unfortunately, increase the likelihood of toxic behavior such as harassment, cyberbullying, hate speech, and profanity, which could erode your organizational culture or lead to employee attrition. The ability to perform real-time cloud content inspection is essential in order to automate content moderation processes at your organization.

Traditional solutions are ineffective.

Once organizations realize they should be protecting information in the cloud, they typically start to wonder whether they can leverage their existing endpoint and network solutions to accomplish cloud DLP. However, these legacy solutions are ineffective, and not up to the challenge, for the reasons we describe below.

Not Built for the Cloud. Data loss prevention has traditionally been addressed by endpoint DLP technologies, network DLP technologies, and cloud access security brokers (CASBs). As the world shifts to the cloud and the web of connection and transit points becomes increasingly complex, it no longer makes sense to manage data via every possible node it may pass through. A far simpler approach is to identify and remediate content directly at the source - at the cloud application layer, which legacy solutions cannot effectively do.

Endpoint agents were a great solution for bygone times, but nowadays only provide limited visibility in the Bring Your Own Device era. But even if you could install an endpoint agent on

every device possibly accessible to your teams, endpoint technologies rely on end-user agents that require installs, updates, patches, and ongoing maintenance, not to mention processing overhead on the end-user's device. DLP on the endpoint is non-comprehensive and should be viewed as the last point of defense. Oftentimes, by the time data has reached the endpoint, the data is already compromised and could be in the hands of a bad actor. Personal devices are a blindspot. If a bad actor knows an endpoint is being monitored, they'll be less likely to use that endpoint day today.

Network technologies require network agents/proxies that can introduce latency and lead users to find workarounds. Network proxies have blunt actions, like blocking certain traffic or websites. Blocking Box means employees will re-route around IT to use Dropbox, Google Drive, or any number of other unsanctioned tools. While it's possible to block these sanctioned tools, there are hundreds of thousands of SaaS applications, websites, and tools that can be discovered by someone who is motivated to share data. This sprays data even further.

High False Positive Rates. Legacy technologies rely on traditional methods like regular expressions and fingerprints instead of machine learning because they do not have the ability to centralize data labeling and training in the cloud scalably while maintaining performance inline or on-device.

Because endpoint and network solutions require processing or network bandwidth, the focus is on simple algorithms that reduce overhead, instead of focusing on advanced machine learning techniques. Data fingerprinting works for known-knowns, machine learning is required for unknown-unknowns - the biggest risks are those you don't yet know about.

These solutions also lack application-specific context, which leads to lower accuracy (read: alert fatigue). Too many false alerts not only render data protection unmanageable for your team, but it also means true positives may slip through the cracks For example, a file in Google Drive that's shared publicly is at much higher risk than a restricted file that is only shared internally. Without application-context, a traditional solution has no way to discern this - you'll get alerts based on policy, but not actually in accordance with the risk level.

Limited Scanning Scope. Cloud silos have been aggregating sensitive data in your organization since they were first adopted. Traditional solutions look at data on a go-forward basis, but there's significant risk in what is already stored in these environments. Effective solutions will scan

existing data in those environments and provide insight and capabilities to remediate any findings.

Costly Deployment. Traditional solutions can take months to deploy. There are significant IT costs associated with this installation, as well as the gap in coverage - the organization is unprotected during this time, all the while employees are interrupted in their day-to-day work with communications and instructions from IT for proper implementation.

Lack of User Friendliness. 23% of respondents in the 2020 Insider Threat Report (Cybersecurity Insiders) acknowledge that traditional DLP initiatives "impede employee productivity and collaboration." These initiatives can often be time-consuming, involving data tagging, manual remediation, and employee training. They often have a negative impact on internal culture ("big brother is watching me") rather than getting the teams engaged and participating in security. In addition, they often have a negative impact on system performance - for instance, DLP endpoint agents typically result in high latency and unresponsive applications and endpoints - causing frustration and decreased productivity across the employee base.

Cloud data security is a real issue.

The urgency of this business need coupled with the inefficacy of traditional solutions is resulting in real data security and compliance risks for organizations.

Here are a few recent examples:

Twilio, 2021 Extensive supply-chain attack exposes secrets of dozens of companies, including Twilio
Codecov, a cloud-based tool that assesses code coverage by software tests, confirmed that they had been breached by a malicious attacker, causing code to be exported to a third-party server. The impacted script was widely used, including within Github Actions and by Twilio in a number of projects and CI pipelines. Twilio found that a number of private GitHub repositories, which held both secrets and a small number of email addresses, had been cloned. Twilio addressed the risks "by thoroughly reviewing and rotating any potentially exposed credentials."

Cloud DLP solutions can help in similar situations to immediately identify where secrets or PII live in repositories so that companies at risk can not only act quickly to remediate sensitive findings, but also monitor these tools and systems in real-time to proactively address issues before a breach occurs.

Twitter, 2020
Threat actor accesses passwords stored in Slack to hijack Twitter backend

In the summer of 2020, high profile Twitter accounts began posting tweets promoting a Bitcoin address. It was found that the accounts were not individually compromised, but that a threat actor had gained access to Twitter's backend. The intrusion began with a "vishing" or voice phishing scam in which the attackers pretended to be Twitter IT support staff helping employees troubleshoot VPN access. This let the attackers move through Twitter's IT systems.

Within the company's Slack Workspace, the hackers found credentials that allowed them to access the Twitter platform's backend to reset and access Twitter user accounts. SaaS applications like Slack allow for the proliferation of PII and other types of sensitive information. Without the ability to enforce security best practices among employees, such platforms can increase the exfiltration risk of sensitive data. That's where solutions like cloud data loss prevention come in. Cloud DLP is useful in determining when sensitive data is shared or posted in channels where it doesn't belong. Incidents that violate policy can be easily remediated through redaction or quarantine of files and messages containing offending content.

Report: 8 healthcare orgs leak over 150k patient's PHI on GitHub, 2020

As a high volume collaboration tool, GitHub can be a platform where coding best practices—such as hardcoding tokens and credentials—aren't always followed. However, another major risk that teams might not be aware of is that GitHub can be an environment where customer PII and PHI can proliferate as well. In August 2020, a security researcher examined databases and repositories for data leakage for over a year and a half. Among the discoveries made by the researcher was that PHI was publicly searchable on GitHub with simple search parameters. The researcher specifically found eight entities leaking PHI from 9

repositories. While it's important for organizations to enforce coding best practices and remediate the exposure of sensitive information, it can be difficult to do without dedicated tools. Solutions like cloud DLP which can scan repositories on push events can prevent the accumulation of credentials, PII, or PHI in your repos.

Report: Major public and private sector orgs leak business-critical data through Jira, 2019

Security researcher Avinash Jain was able to publicly access Jira dashboards for organizations like NASA and even tech giants like Google and Yahoo. In some cases, Jain was able to access project details, usernames, and other internal data. This occurred due to the improper permissions for dashboard filters within JIra. Mispermissions of this nature aren't that uncommon across SaaS systems. For example, similar stories for services like Trello and Google Drive have emerged. Such misconfigurations can go undetected for months, leading to massive sensitive data leakage over time. Cloud DLP can instantly discover such occurrences, rapidly remediate them, and inform relevant parties.

The ROI of Cloud Data Protection

A comprehensive security program should have multiple layers of protection in place. Rather than attempting to prevent or detect breaches (a role filled by plenty of other security vendors), cloud DLP takes a different approach - alerting you to instances of sensitive data in your cloud applications, so that you can remove it from where it doesn't belong. This way, there is no sensitive data there to be stolen or exposed. By keeping your cloud ecosystem clear of sensitive data that doesn't belong there, the actual repercussions of a breach are far reduced. And with cloud DLP, this added peace of mind is available with minimal effort, overhead, and maintenance. It's a simple way to add protection without overburdening your security team.

Below, we explore some examples of how Nightfall cloud DLP can benefit your organization.

Reduce the severity of a potential breach

Limit the damage of a potential data breach by limiting the sensitive data that's at risk.

The Problem: Cybersecurity dominates headlines and affects companies of all sizes around the world. In 2015 alone, there were thousands of known incidents impacting hundreds of millions of identities and costing millions of dollars per incident. The average total cost of a data breach in 2020 was $3.86 million (IBM Cost of a Data Breach Report, 2020). According to the 2019 Cloud Adoption and Risk report, the average enterprise organization experiences 31.3 cloud-related security threats each month.

How Nightfall cloud DLP can help: Nightfall reduces the likelihood of a sensitive data breach by detecting and protecting sensitive data across cloud assets, adding application-level, context-aware content inspection, and providing a centralized control plane for managing sensitive data policies and remediation across your environment.

There are a few factors to consider when evaluating the severity of a potential breach:

- Are there external records to protect? How many?
- Are there internal secrets to protect?
- Is compliance required?
- How many employees do you have?

The estimated cost of a breach can be estimated based on industry benchmarks, including:

- Base US cost per breached record
- Cloud adjustment if the data is in the cloud
- Base US industry cost per breached record
- Cost per employee

Nightfall dampens the cost of a data breach by providing DLP protection, removing sensitive data from cloud applications where it is improperly stored or shared. Nightfall alerts also provide the potential for end-user education, to reduce the likelihood of accidental data sharing moving forward. Nightfall DLP is an essential part of a layered security strategy, and our customers view Nightfall as a crucial line of defense against internal or external data exposure.

It can be a bit difficult to quantify the avoidance of hypothetical data breaches. However, the numbers do show that an astonishing amount of sensitive data tokens occur unprotected in cloud applications in the wild. Nightfall customers have found critical sensitive data that required

remediation at an average rate of ~1% - how many findings would that be for the amount of cloud data you're storing?

Demonstrate compliance

Check the box on compliance needs with a simple yet effective solution.

The Problem: Various compliance regimes such as HIPAA, PCI, SOC, ISO, CCPA, GLBA, and many others relate to and dictate how sensitive data should be collected, stored, processed, used, and more. This web of compliance can be difficult to keep track of, especially given that some regimes are somewhat open to interpretation. In addition to formal regulations, end-customers such as banks often have stringent compliance requirements per their contractual terms.

How Nightfall cloud DLP can help: Nightfall provides a simple way to demonstrate application-level DLP controls, to meet compliance or customer requirements. Nightfall customers can easily manage data detection settings, track sensitive data findings, and even take remediation actions from within the Nightfall solution.

Enhance internal culture

Automate content moderation to keep your digital workplace positive.

The Problem: The shift to a digital, cloud-based workplace has occurred, and employees now spend much of their time engaging and collaborating with each other online. Not only can online content spread broadly and quickly, but it's also more at risk of violating company standards and expectations - because online interaction can tend to veer from the social norms that reign when people are face-to-face.

How Nightfall cloud DLP can help: Nightfall can be used for content moderation, detecting harassment, bullying, hate speech, profanity, toxicity, and more. This reduces the risk of HR compliance issues, lawsuits, and degraded company culture. With Nightfall, organizations can proactively identify risks that are beyond the scope of a code of conduct, acceptable use policy, or employee handbook. This allows HR to be more proactive and address potential problems before they balloon and ripple across the organization, which can have long-term negative effects on culture.

IT Savings

Data classification and DLP solutions can yield cost savings by reducing IT overhead.

The Problem: Without a dedicated solution, organizations often need to manually monitor cloud applications for sensitive data (or choose not to manage the risk). This involves significant IT resources and can lead to employee disengagement and turnover due to tedious and repetitive work.

How Nightfall cloud DLP can help: High-quality detection yields lower false positives, so the security team can be effective and productive with their time, triaging real issues. Nightfall also offers options to automate remediation, to keep the process running in the background. As a result, IT and security can invest more time on moving the needle forward, such as furthering proactive and preventative controls.

General Productivity Benefits

Cloud DLP can help improve employee productivity and reduce training overhead.

The Problem: An organization's security posture is reliant on its people. Traditional security solutions can impede and block end-users from doing their work, leading them to spend time and energy routing around IT. A big brother approach can also have a negative impact on overall workplace experience and culture.

How Nightfall cloud DLP can help: Cloud DLP yields higher employee productivity because employees spend less time working around blocked application access. With end-user alerts, Nightfall also assists with employee education and helps end-users correct poor data hygiene efficiently. Increased productivity and a shared security culture can lead to a happier workforce with lower attrition rates.

Build vs. Buy

Cloud DLP has a significantly lower total cost of ownership than building and maintaining an internal DLP solution.

The Problem: Traditional DLP solutions have failed to meet the needs of the modern enterprise. As a result, organizations have turned to building their own solutions. These homegrown

solutions have significant costs and overhead associated with them that can have negative consequences and slow down the business.

How Nightfall cloud DLP can help: Re-allocating teams to focus on cloud DLP means taking them away from progressing your core business. This has a significant opportunity cost, so it's important to consider the number of months/years that it would take to prototype an in-house solution.

Cloud DLP provides a plug-and-play solution with fast time to value. The ongoing cost of ownership is also lower - no need to install updates or patches, and Nightfall's team of data scientists maintains the machine-learning detection engine for you.

Of course, you'll also reduce the risk of potentially massive costs associated with data breaches.