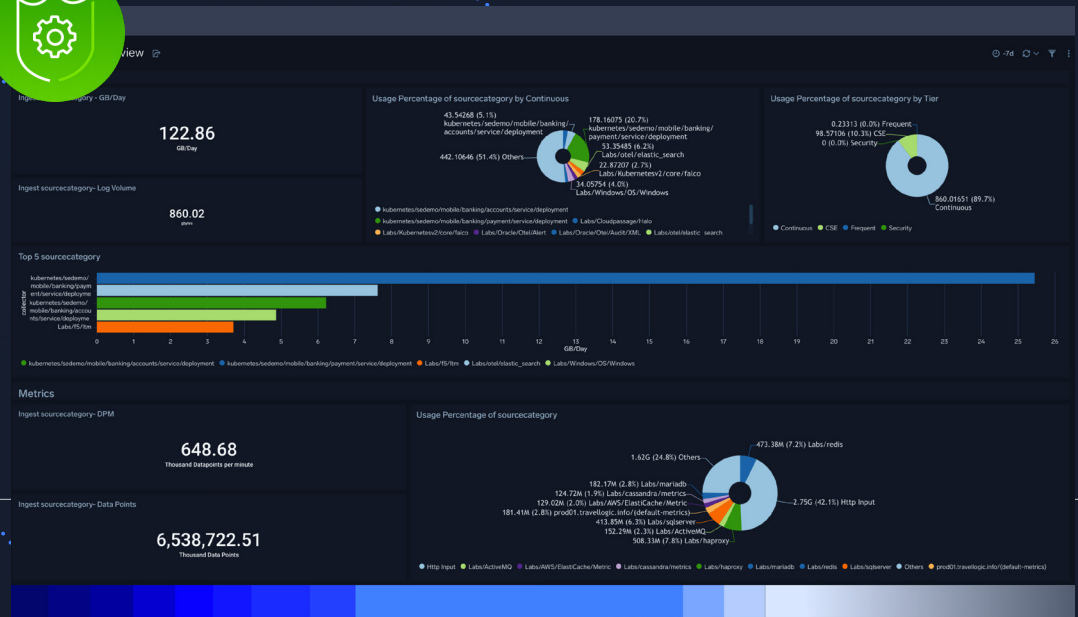


How to collect, store, search and analyze your data

sumo logic



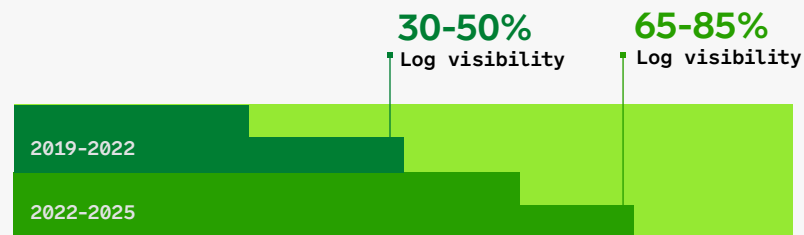
Scattered data is a security risk

When it comes to quickly addressing security threats, it's common for data to be spread throughout different tools, clouds and functions, making it hard to access, see and use effectively. Under these circumstances, identifying potential threats is an arduous task.

With businesses going all-in on cloud-based infrastructure and apps, safeguarding digital assets is getting even trickier. Recent [McKinsey](#) research confirms this reality.

Cybersecurity trends

- An increase in cybersecurity attacks on small-and medium-sized businesses
- Loss of digital trust, as more and more customers cease doing business with affected parties
- The crucial need for log processing and visibility. Just three years ago, the average enterprise only had visibility into 30% of their operations (now up to 50% and growing). Businesses are moving even more aggressively to increase visibility via log data ingest.



Over the past three years, companies have boosted their share of total log volume visibility from about 30 percent to about 50 percent on average and are pushing toward 65 to 80 percent over the next three years.

N=173

Source: McKinsey Cyber Market Map 2022

DEFINING A

Security data lake

To get ahead of potential threats, your organization needs a wealth of detailed security data that you can act on — in one easy-to-access location. Enter the [security data lake](#): your source for digitally diving into security insights.

A security data lake is a centralized repository that collects and analyzes large amounts of security data from various sources, offering a complete view of an organization's security posture. This tool allows security teams to identify potential threats, investigate security incidents and respond proactively to potential risks.

Security data lakes provide a historical overview of security events. Your team can proactively use this information to identify patterns and anomalies, ensuring that your data serves a valuable purpose in protecting your organization's assets.



Retaining data in its original, raw format is also a benefit of security data lakes. This feature provides valuable information for investigation and forensic purposes. The result? Easier threat hunting and investigation through standardized querying and visualization of all data.

You've probably heard the terms "database" and "data lake" being thrown around in similar contexts, but they aren't as interchangeable as they may seem.

Databases enable teams to efficiently access and manage large amounts of structured or semi-structured data electronically. The database design supports Online Transaction Processing (OLTP), allowing for real-time data processing and record-keeping. However, with the International Data Corporation (IDC) [projecting](#) that 80% of the world's data will be unstructured by 2025 — this type of storage is simply unsustainable.

Additionally, databases are not a preferred option for analyzing cloud security data. They function primarily for operational and transactional workloads rather than analytical workloads. That's where data lakes come into play.

Data lakes support Online Analytical Processing (OLAP). Security teams use these to collect data from multiple sources to power analytical insights. Data lakes offer solutions to fill in the gaps in database functionality, specifically:

- The ability to store structured, semi-structured and unstructured data
- Ingest data without having to define schema
- Optimize performance and efficiency due to separated storage and compute functions

A security data lake allows you to store and access various data types and formats, making it easier to process and analyze data from multiple tools and technologies. Traditional solutions cannot handle such significant volumes of data, leaving valuable security logs and event data scattered across various systems and tools.

What types of security logs can data lakes store?



Firewall logs

Firewalls deliver valuable information to identify potential threats, including malware, application types and command and control activities.



Network security products

For standalone systems like intrusion protection or network data loss prevention, centralizing these logs with the rest of your data is the best practice.



Proxy and web filtering logs

In the absence of proxy and web filtering logs in your firewall, businesses must scrutinize IP, URL and domain data to identify potential links to malicious locations. User-agent logs are also valuable in unraveling complex breaches and resolving issues.



User access

Tracking a user's Windows authentication, single sign-on and Active Directory are great sources to tie one user to the event in the system, even if they change IP addresses in the middle of their activity.



Endpoint security solutions

Implementing endpoint security solutions can be a game-changer in network security. By collecting data from each device connected to your network, security experts can easily filter out false positives and focus on the real threats.



Threat intelligence

Accessing logs and data from recent threats at other organizations can help your algorithm recognize similar patterns or behaviors faster in the future.

Centralizing security data, protecting business

Let's look at Medidata, a cloud-based provider for clinical trials and studies. Medidata wanted to quickly identify any potential signs of attacks and promptly take action to mitigate them. This is crucial for protecting their data and helps them maintain a high level of trust with their clients' centralized location, enabling them to derive valuable insights from the volume of their machine data.

However, Medidata generated terabytes of log files each month, which made it impossible to manually spot attacks. To address this, they needed a solution that could aggregate raw data in a centralized location, enabling them to derive valuable insights from the volume of their machine data.

Medidata sought a security partner that could work both on-premises and within AWS. Most providers also required additional hardware or software, which Medidata did not want.

“ Users throughout our organization benefit from the insights we're continually extracting from Sumo Logic's centralized machine data repository.”

Isaac Wong | VP Platform Architecture and User Experience | Medidata



When Medidata contacted Sumo Logic, it was smooth sailing.

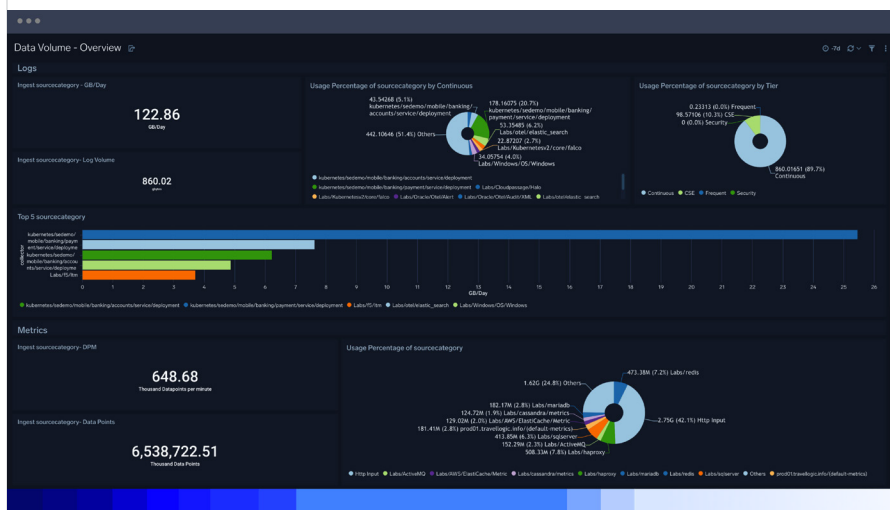
Sumo Logic's SaaS analytics platform provides threat visibility across an organization's entire infrastructure and application stack — no extra hardware or software required. Additionally, Sumo Logic offered vendor-agnostic flexibility, meaning their platform could easily ingest data from on-premises and the AWS data lake to its centralized security data lake.

How Sumo Logic empowered Medidata:

- Provided a centralized location to monitor data and discover valuable insights
- Delivered real-time visibility across Medidata's hybrid infrastructure and application stack
- Automatically generated audit-ready compliance reports from both its on-premises and AWS event logs
- Simplified cloud and on-premises audits
- Strengthened Medidata's security posture with a composite view across the network, server and application stack

With Sumo Logic, Medidata utilized its wealth of security data to confidently provide services to 24 of the top 25 global pharmaceutical companies developing life-saving medical treatments and diagnostics.

To learn more about how Medidata uses Sumo Logic to support its security analytics find the details [here](#).



Sumo Logic provides maximum flexibility with vendor-agnostic data collection and storage of security logs, combined with domain-agnostic analytics.

Ready to dive in?

Your step-by-step guide to building a security data lake

The increasing complexity of IT environments and a lack of threat visibility and expertise, can leave security teams feeling overwhelmed. Security data lakes are currently the best solution for managing the increasingly large amounts of security data we produce globally.

Sumo Logic is a powerful and scalable log analytics platform that secures your data and provides centralized access for security data analysis.



What are the basic process steps involved in building a security data lake? On the following pages we share several key steps involved in building a [security data lake](#).

1 Defining your security data goals

Determine what data you need to analyze. Data may come from your security detection tools, network devices, applications, servers and endpoints. For example, collecting data on user activity and access logs is a great place to start if you're concerned about insider threats.

2 Choosing a data lake solution

Choose a storage solution that manages vast amounts of data efficiently and adheres to compliance requirements. Certain platforms may require saving data in cold storage, which hinders your investigation capabilities. Sumo Logic offers [data tiering](#), allowing you to store data affordably without compromising accessibility.

If you're not sure where to start, consider asking these questions:

- What kind of data sources can they ingest from?
- How is the data searchable?
- How much storage capacity do you need?
- What is the event-per-second throughput?

3 Implementing data ingestion

Set up a process for collecting data into your data lake. With Sumo Logic, you can begin data ingestion in a secure and compliant manner out of the box. Our high event-per-second throughput ensures you can handle the most demanding workloads and data ingestion without issue.

4 Enforcing data security

Security controls are necessary to protect your data lake from unauthorized access and breaches. At Sumo Logic, we use state-of-the-art encryption techniques to safeguard your data in transit and at rest.

But it's not just about protection. You must also prioritize compliance and monitoring to ensure your data lake meets all regulatory requirements. That's why we hold numerous [certifications and attestations](#).

5 Searching your security data

With the vast amounts of data generated by security tools, it can be overwhelming to sift through it all. Our Search Query Language helps identify anomalous behavior and flags potential security threats in real-time. Query your data lake to uncover patterns, investigate alerts and identify potential security risks. Start connecting the dots between seemingly disparate pieces of information.

6 Establishing data governance

Sumo Logic makes this easy with role-based access control (RBAC), data classification, retention, archiving, auditing and compliance dashboards.

For example, our data classification tools allow you to easily identify and manage sensitive information, ensuring that only authorized users have access. And you can keep your data for as long as you need to meet regulatory requirements while also ensuring that it is easily accessible.

7 Analyzing security data

Organizations today face a multitude of challenges when it comes to securing their data.

Sumo Logic Cloud Security Analytics is built on a security-first principle. With real-time threat detection, compliance reporting and dashboards, our cloud-native architecture unifies security events and investigations across multiple cloud platforms, including AWS, Azure and Google Cloud Platform.

8 Monitoring your data

Continuously monitor and analyze your security data to identify potential threats and respond proactively.

Our platform is also updated with the latest security content from our [Sumo Logic Threat Labs team](#). This team moves quickly to support customers in crisis, like when the [Log4j vulnerability](#) surfaced. They continuously monitor and analyze data to identify threats, both known and unknown.

The power of centralized security data

Security data lake is an important use case of Sumo Logic [Cloud Security Analytics](#), the ideal operating posture for managing data.

In addition to security data lake discussed here, Sumo Logic covers other critical use cases that include:

- **Audit and compliance** — To meet security regulations and follow best practices
- **Threat detection and investigation** — To identify problems quickly
- **Application security** — To embed security throughout the application lifecycle

Each of these use cases is built on top of the Cloud Security Analytics solution, making it a foundational part of the Sumo Logic platform to monitor, troubleshoot and secure your applications.

Sumo Logic's platform delivers complete visibility across your public cloud, hybrid cloud and on-prem environments, all while providing real-time insights and intelligence. By embracing the infinite power of log analytics, Sumo Logic offers an ideal balance of simplicity and flexibility, making it a top choice for security teams looking to stay ahead of potential threats by centralizing their data collection.

When you're ready to take your security data to the next level, [contact us](#) or start your [free 30-day trial](#).

Sumo Logic.
The infinite power of log analytics.