snowflake®

THE CISO'S GUIDE TO SIEM MIGRATION

Get more out of your security data by moving from a legacy SIEM to Snowflake

TABLE OF CONTENTS

INTRODUCTION

Over the last two decades, cybersecurity teams have relied on standalone security information and event management (SIEM) systems to aggregate log data from endpoints, firewalls, servers, applications and more. However, as the volume, variety and velocity of data generated by systems and users grew exponentially, traditional SIEMs struggled to deliver the scalability, flexibility and advanced analytics needed to detect and investigate quickly.

Monitoring requirements and reporting methodologies have rapidly evolved, with an increased demand for machine learning (ML) to provide insights into data and the ability to search in real time. The demands of the business and operations have quickly moved beyond endpoint and network detection to more advanced analytics such as user behavior analytics and single pane of glass reporting to the CISO or board of executives.

Many cybersecurity teams are familiar with and frustrated by the **volume-based pricing model** that traditional SIEMs impose on customers, which may prevent them from leveraging all the data and insights to protect their organizations. There is growing momentum to architect well-rounded cybersecurity programs with a data cloud platform like Snowflake.

Why the cybersecurity industry is moving to an open architecture deployment with a modern security data lake and best-of-breed applications from Snowflake

SNOWFLAKE DATA CLOUD		
PLATFORM ADVANTAGE	**ADVANCED ANALYTICS**	**THRIVING ECOSYSTEM OF CONNECTED APPLICATIONS**
Cost	Leverage ML functions for anomaly detection and user behavior analytics	Unify data from best-of-breed application
Flexibility	Easily join business and contextual data sets for enrichment and investigation	Leverage user interfaces made for security practitioners
Freedo	Provide support for business intelligence tools such as Tableau, ThoughtSpot, Sigma, PowerBI and more	
Scalability	Consolidate security data from various sources for a holistic view of your security posture	

SLOW AND FAST MIGRATION PATHS TO MODERN SIEMS

There are two schools of thought when it comes to migrating from a traditional SIEM, such as Splunk, to a SIEM connected to Snowflake. Some customers prefer a gradual migration where they augment existing SIEM use cases and detections with Snowflake, while others prefer to onboard to Snowflake and a modern SIEM as quickly as possible. Use the guidelines below to determine which is the best fit for your organization.

IMMEDIATE REPLACEMENT — SIX MONTHS TO A YEAR

This timeline is suitable for organizations with legacy SIEM and require immediate attention. Another factor to consider is if the software license or contract is ending soon.

COMMON REASONS FOR A FASTER MIGRATION PATH:

- Detection rules on existing SIEM are not returning events
- Detection rules on existing SIEM are returning too many false positives, hindering team effectiveness
- Rigid reporting templates are not useful for executive conversations
- A recent or upcoming requirement to ingest more data or process it under shorter SLAs is not currently feasible with the existing SIEM

> Without question, we're delivering more with less in Snowflake, while saving more than 20% on our infosec costs. By leveraging four applications from Snowflake's robust ecosystem, we were able to migrate from our legacy SIEM in less than six months. These applications worked seamlessly together and helped us ingest, normalize and run detections on our data in Snowflake."
>
> **Matthew Sharp**, CISO at Xactly

WATCH THE WEBINAR

GRADUAL AUGMENTATION TO REPLACEMENT — MORE THAN A YEAR

This migration path is suitable for organizations that want to augment their existing SIEM deployment with use cases that are not yet addressed. It is also suitable for organizations that want to move over single use cases or workloads at a time, providing adequate coverage during the migration process.

COMMON REASONS FOR A SLOWER MIGRATION PATH:

- Currently have many working detections built on a legacy SIEM like Splunk, but need additional visibility into data sets that are not ingested into the SIEM (i.e., EDR telemetry)
- Want to prove out modern security data lake deployments with a single use case before migrating all workloads
- Analysts are able to get hands on with the new system sooner, allowing for more time to get comfortable prior to cutoff
- Since Splunk usage is gradually reduced, this method reduces the risk of a delay by allowing a customer to realize value and cost savings prior to cutover.

You can also build your own SIEM on top of your Snowflake security data lake.
Learn how Okta's security engineering team architected their program on Snowflake.

HOW TO MIGRATE TDIR WORKLOADS TO SNOWFLAKE

Once you have decided whether you would like to take a replacement or augmentation path, you will need to understand how to migrate data into Snowflake, build detections and enable federated search.

MIGRATING DATA AND DATA INGESTION

Whether replacing or augmenting a SIEM, data will need to be moved into Snowflake. The following provides a description of several patterns that can be used to achieve that ingestion. Customers may use one of many of the below patterns.

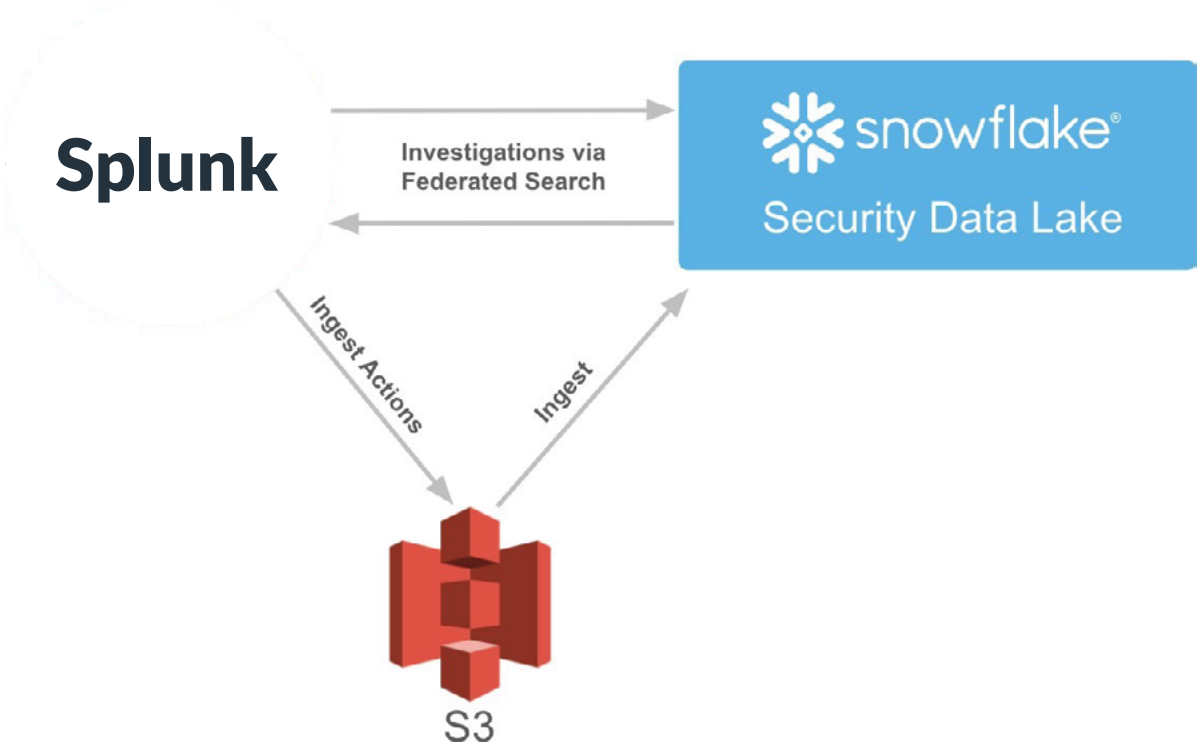
Direct ingestion

Security sources can be ingested directly into Snowflake through native means such as streaming, stages, syslog, native connectors or secure data sharing. This is the most common way to ingest security data into Snowflake. Review our **"Best Practices for Security Log Ingestion and Data Normalization in Snowflake"** white paper for more details.

Forward logs from Splunk: Ingest Actions, archives and DBConnect

The easiest and fastest way to augment Splunk is by forwarding logs directly to Snowflake. This has a quick time to value and is our recommended approach for customers looking to prove the value of Snowflake as a backend.

There are three primary methods of forwarding data from Splunk to Snowflake. Our recommended method is to use Ingest Actions, a feature from Splunk. This can forward data to an S3 bucket in addition to, or instead of, indexing. From there, data can be copied into Snowflake. Using this method provides the opportunity to pass data directly to Snowflake and not be indexed by splunk Keep in mind, this method will require Splunk to stay in your pipeline and will use some compute for sending the data.



Similar to using Ingest Actions, customers may use the Splunk archive feature, in which data is sent to S3 after a certain amount of time. We recommend using Ingest Actions as opposed to archiving since having data split between two locations will increase the complexity of analytic operations. Further, Snowflake provides cost-effective storage at similar rates to S3, so there is limited cost savings to be had by delaying ingestion.

Use of both Ingest Actions and archiving requires an S3 bucket, so customers looking to bypass this requirement may be better served by using DBConnect to transfer data to Snowflake over an JDBC connection. While some customers have found success with this method, others have expressed concerns over connection stability. We recommend you experiment with this approach yourself before making a decision.

Split-stream approach

Splunk customers using **Cribl Stream** or ApacheNifi have the option of using Snowflake as an additional destination. This method is easy to implement and can allow users to take advantage of that functionality. From there, data can be ingested into Snowflake through cloud storage or streaming, such as through Kafka or **Amazon** Kinesis.

Partners

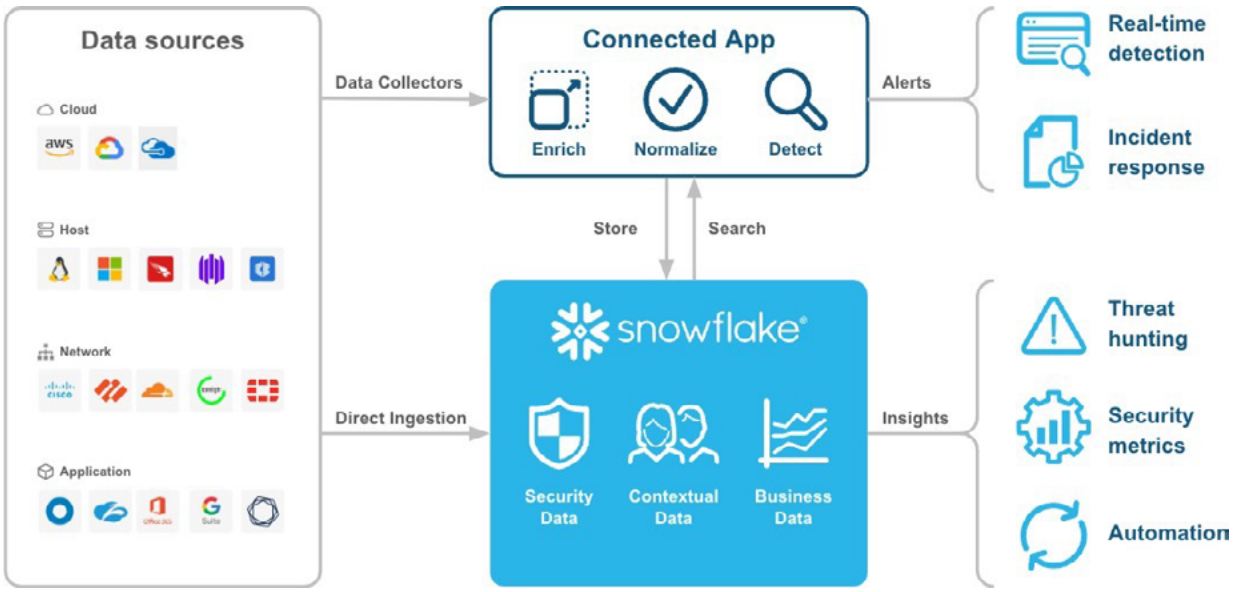
Snowflake partners with several companies offering products that can help with security ingestion. Modern SIEMs and standalone security ETL tools provide a managed option for ingesting data. More information can be found in our **"Best Practices for Security Log Ingestion and Data Normalization in Snowflake"** white paper.

DETECTIONS

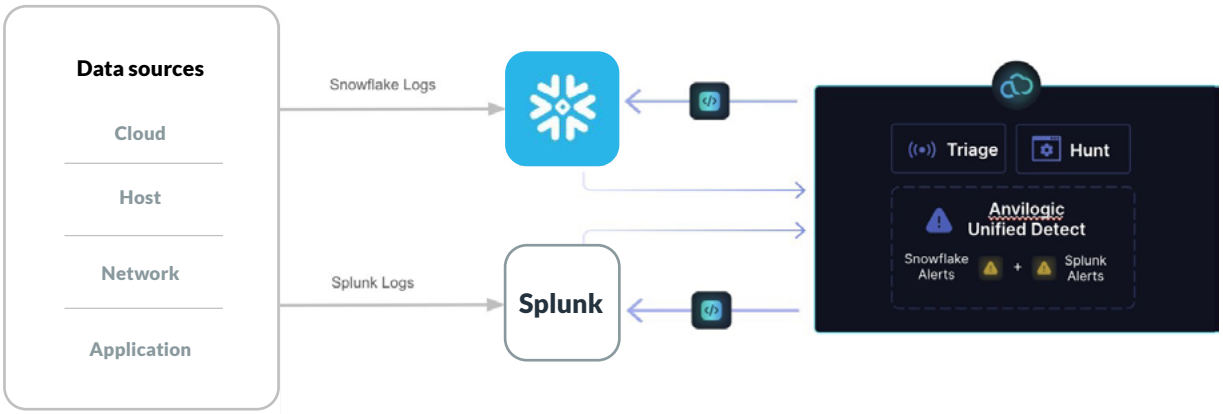
Snowflake supports detection strategies across the buy-to-build spectrum. Most customers opt to use a connected application partner. A connected application is SasS that operates directly on your Snowflake account. This means all data resides in your security perimeter, you maintain control, and you then have access to use that data for other purposes such as analytics, machine learning or even with other connected applications in **our ecosystem**.

Connected applications can perform a variety of functions supporting everything from email security to risk quantification or even scanning Snowflake itself. For a SIEM augmentation or replacement, customers will generally choose a connected app that provides a detection library and incident-response functionality. Here we'll get into a list of companies that can help.

Anvilogic is ideal for organizations seeking autonomy over their data ingestion pipelines. It differentiates itself by facilitating deployment of detections within Snowflake and Splunk. This is a technical benefit for organizations reluctant to fully replace their current SIEM solutions. With Anvilogic, these organizations can integrate Snowflake alongside their existing SIEM, enabling a smooth transition and coexistence that can be maintained during the initial migration phase or indefinitely.



SPLUNK DETECTION AUGMENTATION



Gem's agentless cloud detection and response (CDR) platform is used by SecOps and IR teams to rapidly detect cloud-native threats and automate triage, investigations, containment, and forensics. The platform is a fit for organizations looking to offload ingestion and correlation of their cloud logs (AWS, Azure, GCP, Okta, etc.) into Gem's data lake architecture, while integrating with their existing SIEMs as the centralized point for all alerts in the SOC.

The Hunters SOC Platform is a SIEM alternative that takes advantage of Snowflake's backend for scalability, performance and cost effectiveness, and has been used successfully for rapid migrations. Hunters automates the entire TDIR process, replacing repetitive human work with machine-powered detection, enrichment, correlation, prioritization, triage and investigation.

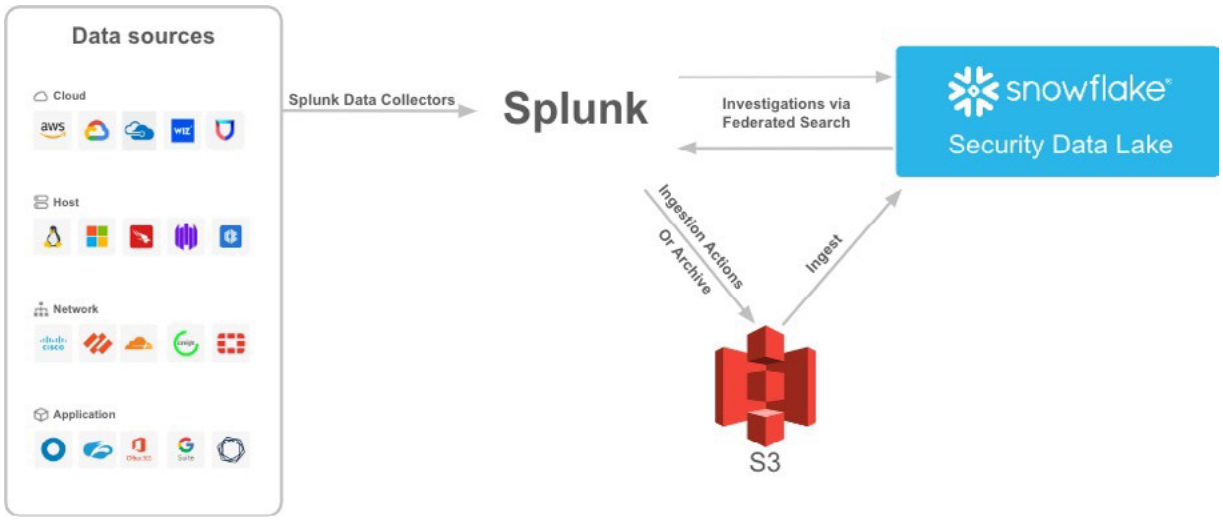
Panther provides detection as code capabilities, allowing customers to write and edit their detections as Python. They provide data connectors to assist customers with ingest and have the ability to export export alerts fired on Snowflake stored data sources to Splunk.

Some customers will opt to build their own SIEM directly on Snowflake. These customers are typically more advanced in their data engineering capabilities or have very specific requirements or scale. Organizations may also choose to write native detections when they need to augment a very specific source but don't need to detect or separate incident response capabilities.

Customer examples: **Okta** and **Comcast**

FEDERATED SEARCH

Snowflake offers the ability to connect via a JDBC driver, exposed to Splunk over DBConnect. Analysts can run investigations and power dashboards where data processing takes place in Snowflake. In this model, only the result set is brought back into Splunk. This means data brought back does not get indexed or counted against ingest limit, and there's only a small amount of compute used to bring the result set — not the raw data — back.



A demonstration of how federated queries can be used in a real-life scenario can be found here:

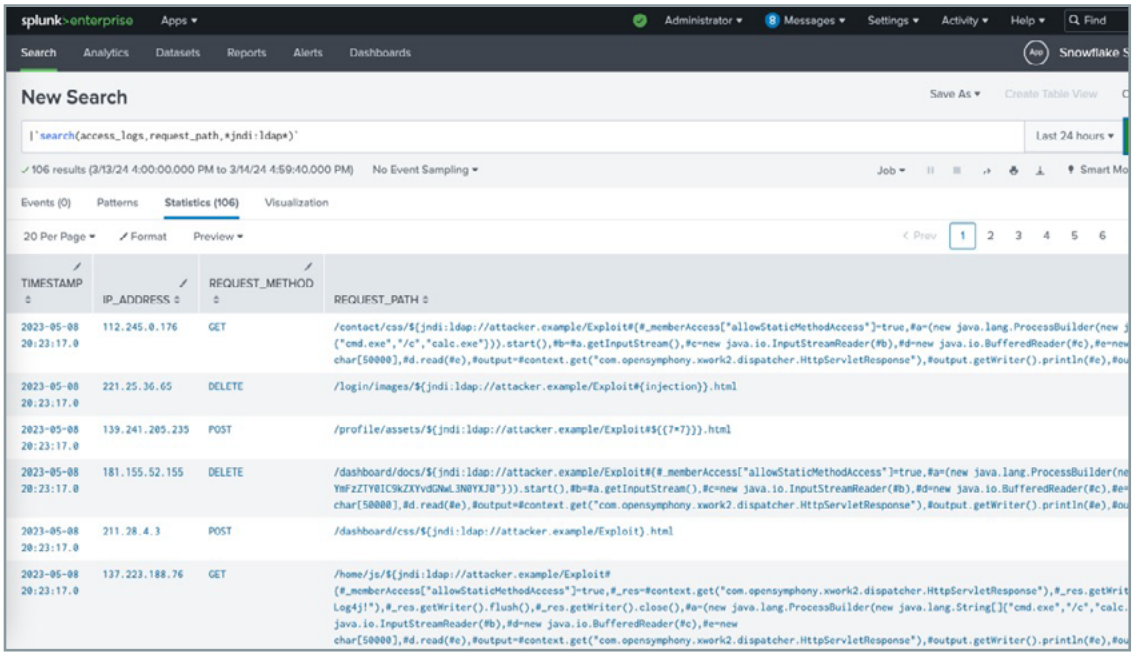


Figure: Search access logs for Log4Shell attack string.

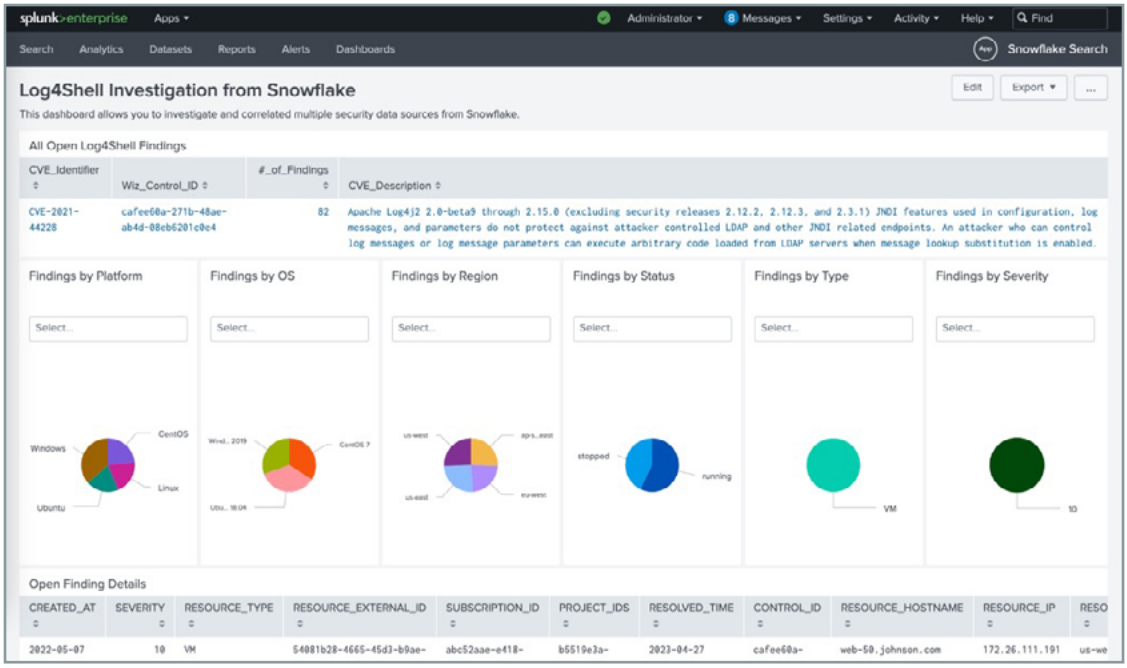


Figure: Look for specific vulnerabilities in Snowflake via the Splunk user interface in SPL.

Watch the full Splunk augmentation demo here.

CONVERTING QUERIES FROM SPL TO SQL

The fundamental issue of translation stems from an innate difference in paradigms. SPL is a streaming language, while SQL is not. Although there's some overlap — the allowance of non-streaming commands in SPL and streaming-type syntax in SQL — fundamentally they are built differently. In theory, streaming languages are associated with real-time data processing and event-driven architectures, while languages like SQL are associated with historical data analysis and complex queries involving joins, aggregations, subqueries and window functions. In practice, analysts will use the tools they have to accomplish what they need, regardless of which language is theoretically more appropriate.

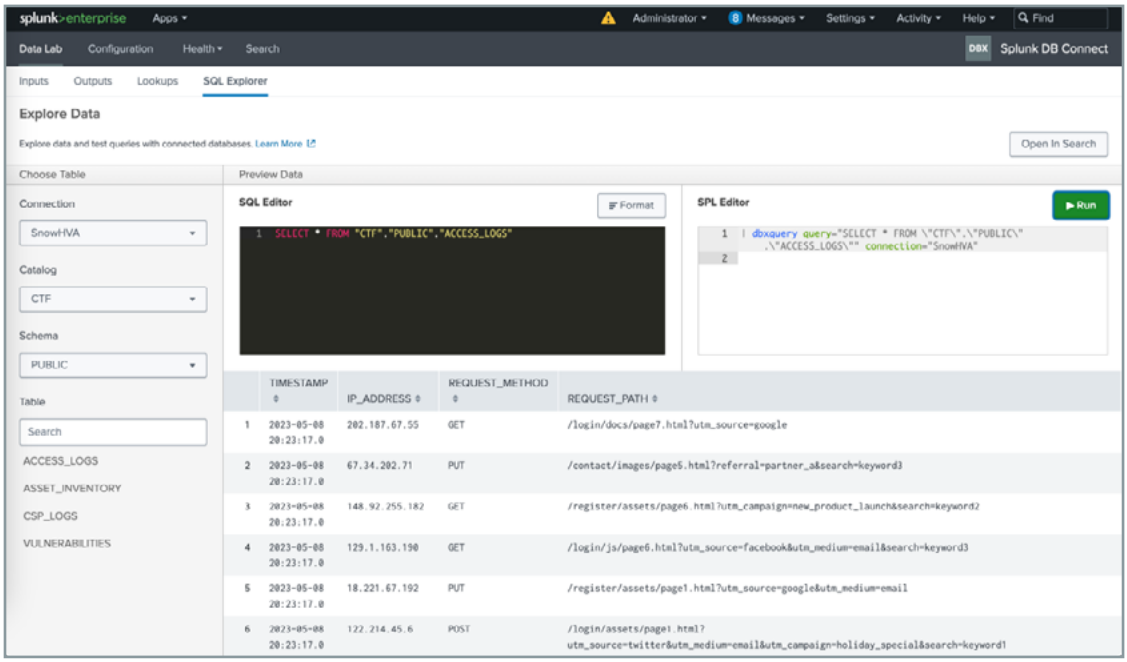


Figure. DbConnect data exploration tool helps format SPL queries from SQL.

THERE ARE TWO MAIN MOTIVATORS FOR CONVERTING QUERIES IN THE FIRST PLACE:

1. **Teams migrating a workload to an SQL-based system wish to transfer their dashboards and queries as effectively as possible:** In this case, migration is a one-time effort, and these queries will likely be in production for an extended period and run repeatedly. This means performance and optimization can have a large impact, so time spent validating and optimizing queries is arguably justified. Not that you have to do it totally by hand, of course. Snowflake's experts have heard good feedback from customers using a large language model (LLM) to do the initial conversion, then having a SME review and tweak.

Strategy for gradual adoption
Using custom macros to emulate functionalities like wild-card searches or writing less-efficient but simpler queries.

2. **Leads want to soften the transition for analysts and threat hunters accustomed to SPL who may find it challenging to switch paradigms abruptly:** In this approach, we suggest a "crawl, walk, run" strategy in which analysts gradually become more familiar with SQL, utilizing LLM-powered SPL-to-SQL query translators from modern SIEM vendors. This method minimizes disruption and enables experts to adapt to the new paradigm while mitigating change risks.

Strategy for gradual adoption
Leverage LLMs to assist in query writing or, if using an SPL-supported UI, funnel SQL results into SPL functions. Eventually, security analysts will grow comfortable with SQL and can natively craft queries in it.

Regardless of which technical migration pattern is best for your organization, translating SPL queries into SQL requires the understanding of and adaptation to fundamentally different paradigms. While automated tools like LLMs offer a starting point, they cannot fully capture the nuances and efficiencies needed when shifting to a new language. For organizations and individuals enacting this transition, a balanced approach is key — using technology to facilitate initial conversions, and then gradually immersing analysts into SQL through a structured learning path. Just as mastering a new language takes time and practice, so does the transition from SPL to SQL.

The good news is that this transition is already being commonly undertaken, and experience has demonstrated that employing the right tools and mindset are key to not just achieving objectives, but also galvanizing a more experienced and well-rounded security team through the process.

HOW TO MIGRATE SECURITY METRICS TO SNOWFLAKE

You may migrate either TDIR or security metrics to Snowflake first — it does not matter which use case you lead with. However, it is important to plan ahead to ensure the architecture you choose can support additional use cases in the future. Consider whether or not you want to own your data pipelines to remove specialized data models.

DATA INGESTION

Find details above in the section on data ingestion, and check out the **"Best Practices for Security Log Ingestion and Data Normalization in Snowflake"** white paper. As a review, these are the methods of ingestion:

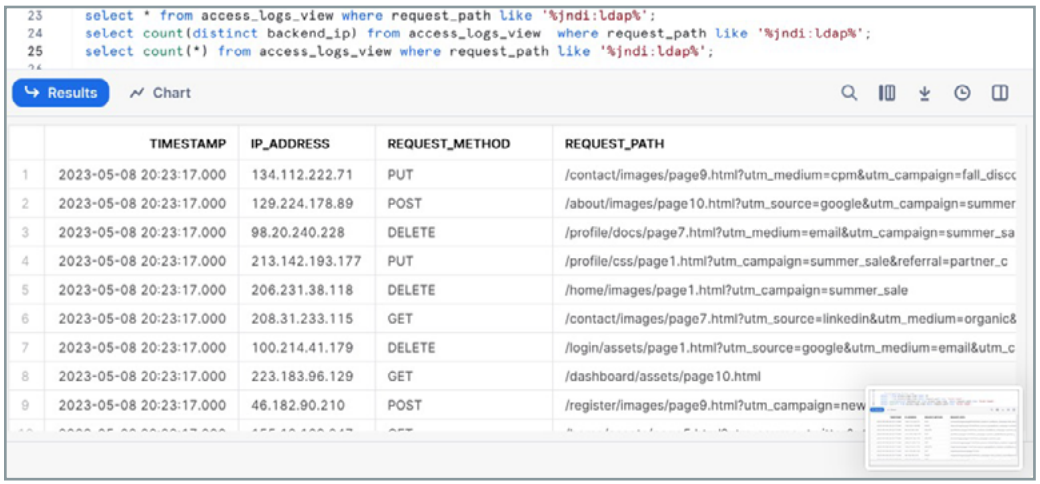
- Direct ingestion via Snowflake features: Snowpipe from S3, secure data sharing, Snowflake native connectors
- Split-stream ingestion using Cribl or ApacheNifi
- Forward logs from Splunk: Ingest Actions, archives, and DBConnect
- Ecosystem partners: SIEMs connected to Snowflake or security ETL vendors

DATA NORMALIZATION

Snowflake considers it a best practice to store logs "raw" whenever possible. This prevents data loss and allows for greater freedom and flexibility in a cost-effective and performant manner.

Because of Snowflake's storage and compression design, storing raw data is typically cost-effective and predictable. In addition, Snowflake's handling of semi-structured data allows users the same performance as with structured data. Customers often use multiple ingestion and transformation methods to prepare data for their use cases.
Transforming data in Snowflake

Depending on the use case, logs can often be normalized natively in an ELT fashion using views, materialized views, and custom user-defined functions that can be written in various languages, such as Python, Java, JavaScript and SQL. Dynamic Tables allow for transformations to happen automatically on one or more tables, and are similar to DAGs. Finally, features like schema detection, inference and evolution allow for easier creation and maintenance of views from structured and semi-structured data.



Views can make it easier to analyze variant data without needing to pre-normalize

Transforming data with partner tools

Snowflake's security ETL partners like Dassana and Monad provide out-of-the-box ingestion and normalization for security teams. Modern SIEMs that ingest directly into Snowflake generally provide normalization as part of their product.

VISUALIZATION TOOLS

Once the data is ingested and normalized, we can finally build visualizations and surface key insights. There are four common methods security teams use for data visualization.

Modern SIEMs

SIEMs connected to Snowflake can query Snowflake data and surface insights into the partner application interface. If the security team prefers the dashboards and templates from the partner, they can leverage the interface to build dashboards.

Business intelligence tools

Teams with BI tools that are leveraged across the entire organization — such as Tableau, Power BI, Thoughtspot, Sigma, etc. — can collaborate with their data teams to build dashboards quickly.

Snowsight

Snowsight is the native dashboard capability on Snowflake. Security or data teams familiar with SQL can build dashboards on Snowsight quickly and effectively.

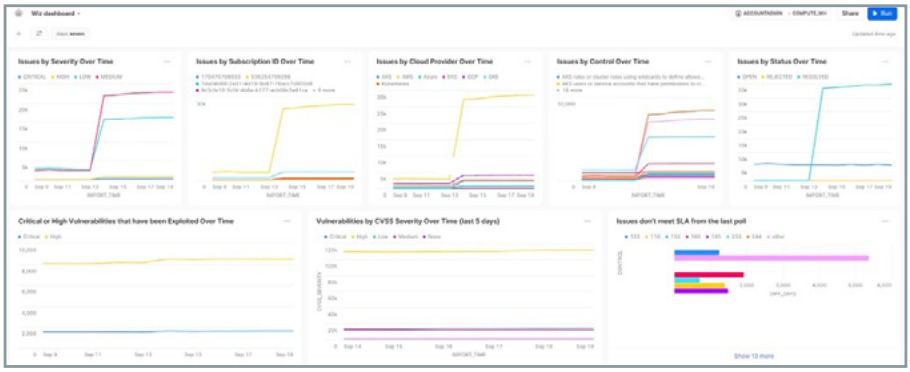


Figure: Ingest Wiz data via the Wiz-Snowflake integration can be used to build dashboards in Snowsight with SQL.

Streamlit

With Streamlit in Snowflake, you can turn data and ML models into interactive apps with Python. This is a quick and easy way to build apps for internal business users to access insights.

Example:

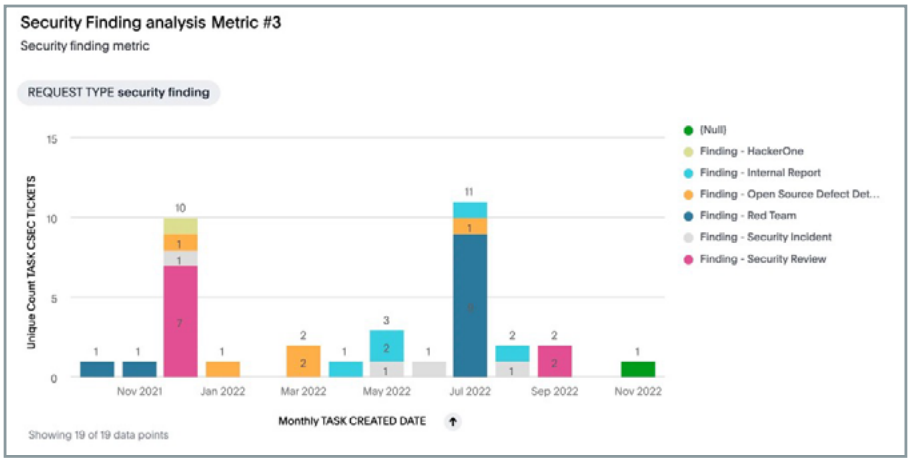
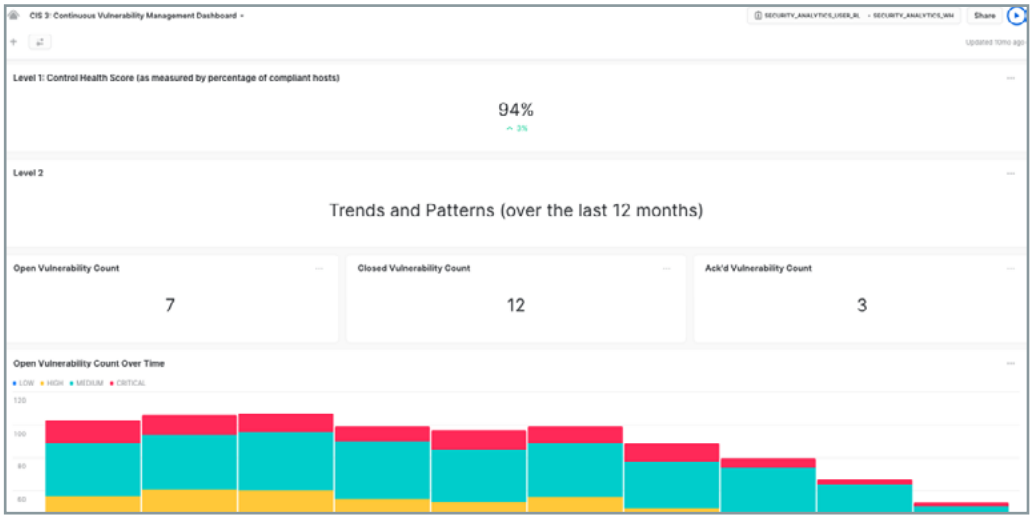


Figure: An interactive app showcases types of open security tickets.

AUTOMATING ACTIONS BASED ON DATA

Having dynamic dashboards for leaders and team members will significantly increase your mean time to detect and respond. However, we recommend to automate processes wherever you can. For example, let's say you've created these self-patching vulnerability dashboards that enable the team to understand which vulnerability is out of SLA and needs immediate attention. We'll review the two levels of automation you can achieve with Snowflake.

Level 1 automation: Analysts can review dashboards at the beginning of their workday so they know which vulnerabilities they must tackle first.



Level 2 Automation:

To take things to the next level, you can set up alerts for other teams within your organization to take action with security orchestration automation and response (SOAR) tools that integrate with Snowflake. Here's an example of how you can automate role changes based on usage with Tines. By doing so, you can ensure that the right people have the right levels of access to different applications and data.

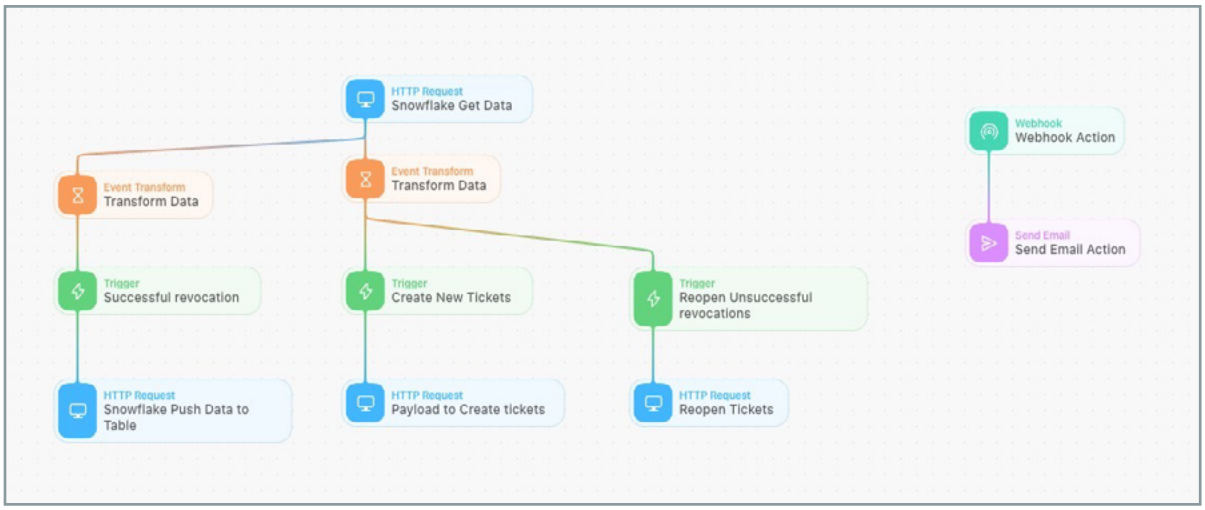


Figure: User Access Review automation playbook

One such example is the User Access Reviews, where security analysts can use an automation tool from Snowflake's application ecosystem (Tines) with Snowflake to create the access Revocation tickets in our ticketing system. The playbook workflow is set up to take relevant steps based on the user data and ticket state. As and when the ticket state updates to "closed," the logic checks for presence of the user in the user access data, and if the data doesn't reflect the same, we would reopen the ticket. In this case, the data behaves as the feedback loop, which provides an additional layer of testing while saving a lot of manual time and effort.

OTHER SECURITY USE CASES TO CONSIDER

Detection and response migrations are typically more cumbersome than other cybersecurity use cases. Consider using Snowflake to power these other augmentation use cases alongside your existing SIEM deployment.

Example augmentation use cases:

1. Threat hunting
2. User behavior analytics
3. Anomaly detection
4. Ad-hoc or breach investigations

Migration timelines will vary depending on the business objective and scope of work. Keep in mind the amount of data that needs to be migrated along with the time it may take to build use cases on top of that data.

Download our **O'Reilly Ebook** to learn how to easily deploy a modern security data lake, so you can decouple the data platform from SIEM capabilities for streamlined log aggregation, and out-of-the-box detection and response capabilities for greater visibility.

Reach out to your Snowflake representative to learn how we can support you in your cybersecurity journey.

ABOUT SNOWFLAKE

Snowflake enables every organization to mobilize their data with Snowflake's Data Cloud. Customers use the Data Cloud to unite siloed data, discover and securely share data, and execute diverse artificial intelligence (AI) / machine learning (ML) and analytic workloads. Wherever data or users live, Snowflake delivers a single data experience that spans multiple clouds and geographies. Thousands of customers across many industries, including 691 of the 2023 Forbes Global 2000 (G2K) as of January 31, 2024, use the Snowflake Data Cloud to power their businesses.

Learn more at **snowflake.com**