

Keep AWS Logs From Running Wild

BY PUTTING PANTHER IN CHARGE



The AWS Security Challenge

Amazon Web Services (AWS) is the world's most comprehensive and broadly adopted cloud platform. Millions of customers worldwide are using AWS to lower costs, become more agile, and innovate faster. AWS is architected to be the most flexible and secure cloud computing environment available today, and its core infrastructure is built to satisfy security requirements for high-profile organizations across each industry and geography.

AWS is not immune to challenges, however. One of the top issues every security professional is faced with as part of their daily responsibilities is collecting and normalizing all of the data produced by AWS and then making sense of that information to confirm security controls are being followed properly. This vast environment of cloud resources forces security teams to switch back-and-forth between tools to deal with hundreds to thousands of security alerts every day.

AWS logs are “noisy” and are often voluminous, and this drives the need for a robust architecture to optimize speed, flexibility, and scale. AWS requires significant configuration out of the box, which can be taxing on resources because they might not have the necessary skill set and time to implement this properly. This adds to unnecessary operational and security risks. Eliminating or redirecting logs out of AWS into your SIEM is technically challenging, unnecessarily costly, and will most likely have significant ETL challenges. Most organizations do not have the skill or the will to effectively and efficiently implement, operate, and maintain an unnecessarily complex environment like AWS as provided to its customers.

Security teams need a single platform for aggregating, organizing, and prioritizing security-relevant data from AWS accounts and combining signals from multiple hybrid platform environments across the organization.

What is Panther?




Panther takes vast amounts of security logs and provides normalization, real-time analysis, and a scalable data warehouse to store and query them.

Panther is the core of detection and response for modern, cloud-focused teams, especially built on AWS. It's a SaaS solution designed for speed, scale, and flexibility while removing the operational burden of managing data. With Panther and AWS, security teams get a centralized store of all activity across security-relevant logs like CloudTrail, Application Load Balancers, VPC Flow, Guard Duty, and more, queryable with SQL and processed with Python.

Panther's utilization of detection-as-code, security data lakes, and real-time alerting gives teams powerful capabilities to meet their unique needs, discover attacker behavior quickly, and answer difficult questions during a breach.



Key Highlights

 Speed Analyze attacker behaviors in real-time and quickly investigate activity across all of your data at a Petabyte-scale	 Scale Transform unstructured security logs into a well structured and robust security data lake	 Flexibility Use Python-based Detection-as-Code to build automation and unique, business-specific logic for alerts
---	--	--

Speed: Detection and Real-Time Alerting at the Speed of Now

With traditional SIEMs, queries over several weeks or months of data can take multiple hours and even up to a day to return results. In an investigation, every second counts, which is why Panther's architecture is built on a scalable and serverless data warehouse that removes all DevOps overhead. By leveraging an architecture that separates storage (where the data lives) and compute (how the data is loaded and queried), Panther is uniquely designed to provide fast searches across extremely large amounts of data, so security teams can get the answers they need quickly.

Additionally, Panther also analyzes all log data streamed into the platform in real-time to provide security teams immediate signals right when new suspicious activity happens. This complementary mechanism is designed to save money and time by bypassing the data warehouse and greatly decreasing the latency for receiving an alert. Customized alert grouping and thresholds enable the ability to analyze up to 24 hours worth of data for a given detection.

Thanks to the elastic nature of the cloud, security teams can now quickly load all of the data they need, scale up on-demand, and bypass all ops overhead to drive better security outcomes and improve visibility

Scale: Ingest All of Your Data

Security teams are too often holding back all of the data they need to send for storage and analysis due to ops or licensing costs. Panther alleviates these pains with a fully serverless and cloud-native architecture along with native support for Snowflake, allowing for all security data to be sent, regardless of size. The separation of storage and compute enables both data ingestion at a massive scale and fast queries.

Panther's data lake is built on parsed and normalized data, which is configured automatically for common log types. This works by comparing incoming data to a collection of schemas, conforming them into a consistent structure, and extracting common atomic indicators such as IPs, domains, ARNs, and more. This critical step ensures that data is searchable at scale, enables fast queries over massive amounts of security data, and enables complex analytics such as joins across log types. Additionally, automatically structured data means no guessing field names or fussing around with regular expressions while in the middle of an investigation.



Flexibility: Detections-as-Code and Data Normalization (extracting ARNs, account IDs, etc.)

Detection and alerting logic in traditional SIEMS is often limited and often even impossible to customize. With Panther, security data is analyzed with expressive Python-based detection-as-code, enabling teams to build robust and tailored security alerting pipelines with version control, unit tests, reusable code, and CI/CD.

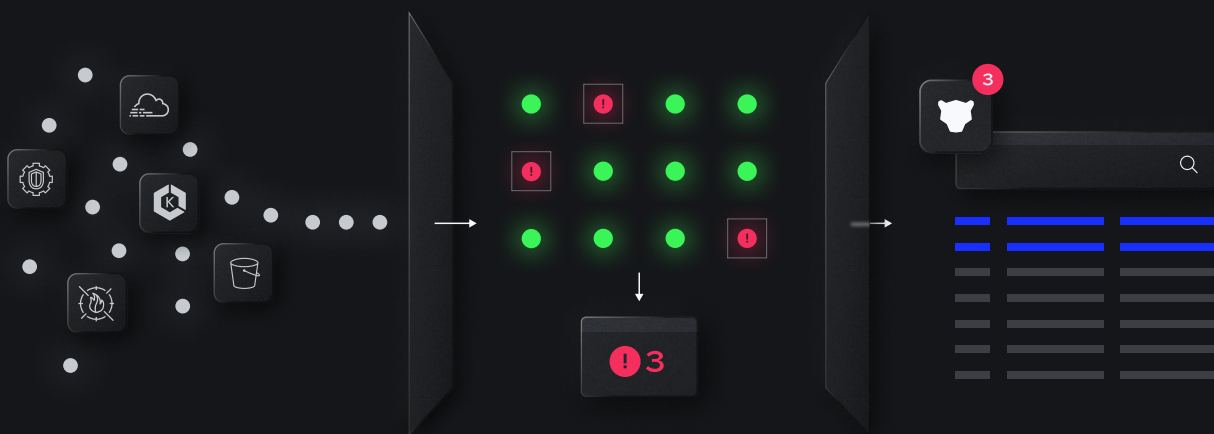
Python is the most popular programming language in the world today and is used extensively in cybersecurity by analysts, engineers, and many experts in the field. By supporting common programming paradigms and best practices such as the use of helper functions, reusable code, existing libraries, peer reviews and comments, and dedicated functions to decorate alerts like alert titles, deduplication logic, and alert destinations, Python enables significantly more flexibility than traditional SIEM domain-specific languages (DSLs) for writing custom, business specific detections.

Panther also ships with AWS detection packs that cover common attacks, vulnerabilities, and security best practices. These detections are mapped to frameworks such as MITRE ATT&CK and CIS, and are continually improved with a community-based approach. As more experts collaborate to improve the industry as a whole, the 'Githubfication' of Infosec has been emerging as a trend and Panther's support for detection-as-code enables teams to embrace an open, shareable, and contributor-friendly model for speeding up infosec learning and collaboration (OSS detections can be found [here](#)).

Finally, Panther seamlessly fits into a modern, best-of-breed detection and response stack, and can easily be integrated with other security tools to enable automation and collaboration across teams. Examples include SOAR, SIEM, endpoint security tools (e.g., EDR), IDS, Firewalls, Active Directory, LDAP, and more.

HOW IT WORKS

Cloud Scale SIEM



The Magic Equation: Panther + AWS

Panther brings disparate security logs from multiple AWS accounts together into a single view and makes them usable with speed, scale, and flexibility while operating as a robust security data platform.

Panther's data pipeline is built on the idea of "Streaming ETL (Extract, Transform, and Load)," where realtime security data is parsed, normalized, and stored in an efficient and compressed format at machine speed. By using a method of micro-batching, latency is measured in a few minutes versus hours or more. This brings structure to security data and enables teams to connect the dots during an investigation by querying the extracted fields, such as common IOCs, IOAs, and other telemetry across all data. This provides an extremely scalable operational environment that enables security teams to process, analyze, and retain exabytes of security data at unprecedented low costs when they need it, which is right now.

Data lake architecture enables the ability to collect data up and down the stack to get as much context as possible to include but not be limited to cloud, network, database, host, and application data. By prioritizing the collection of logs that have security value, ensure that time and resources are used efficiently. This can be a particularly important challenge to overcome early on as analyzing a number of data sources can quickly become very noisy, irrelevant, and can take up unnecessary space and cycles. In addition, organizations can ingest, parse, normalize, and analyze their security data and store it for long-term retention, creating a well structured and scalable security data lake.

There are many benefits to going serverless that cannot be met with alternative approaches



Elastic and Scalable

Use what you need when you need it at machine speed.



Cost-Effective

Extremely high return on investment due to low license and administrative costs.



Security

Eliminates the need for OS patching of EC2 servers that are no longer needed to run your SIEM infrastructure.



Ease of Use

By eliminating infrastructure constraints, operating teams can reprioritize resources and focus on other priorities.



Visibility

Many legacy approaches with on-prem infrastructure have strict limits on ingestion and retention.



AWS Tools of the Trade

PRIORITIZED BY SECURITY USEFULNESS

In June 2021, Amazon published their AWS Security Reference Architecture which shows AWS account structure, security services, and features to maximize security workloads in their infrastructure. This guidance is broken into 4 major components.

Security Foundations

- ① AWS Organizations, Accounts, and IAM Guardrails
- ② The AWS Security Reference Architecture
- ③ IAM Resources
- ④ Code Repository for AWS Security Reference Architecture Examples

Below is a list of the most critical AWS security tools and features as well as useful third-party tools that operate seamlessly with Panther - in a simple and straightforward manner. This then becomes a synergistic combination that can turn a traditional formula of $1+1=2$ into $1+1=50x$ or more.

The key benefit of putting Panther in charge is that it can effortlessly collect all security-relevant AWS log types into a centralized and normalized single view, such as:

- Application Load Balancers
- RDS MySQL Instances
- CloudTrail Events
- CloudWatch Logs/Events
- GuardDuty Alerts
- S3 Access Logs
- Firewall Events
- VPC Network Traffic Flow

AWS use creates a myriad of disparate logs, with differing data types, data structures and many alerts. It is virtually impossible to parse, normalize and understand all of this data without the right automated solution.

Panther quickly and easily centralizes all of your AWS data in a data lake, normalizes the data, and enables you a single view, providing greater situational awareness



Integrations and Data Store Connectors

In addition, Panther has out-of-the-box native integrations that make it easy to analyze your data, triage alerts quickly, and remediate incidents using the tools your teams love. Custom workflow integrations can be implemented with PagerDuty, Slack, Jira, Asana, Microsoft Teams, and many others. For more complex users, integrating SOAR can produce unique use cases and can be incorporated into innovative workflows.

Snowflake Benefits

With Panther and Snowflake, you can:

- Collect terabytes of normalized security data in Snowflake for affordable, long-term retention
- Scale up your warehouse with the click of a button when you need to query months or years of data during an investigation
- Join Panther data (e.g. alerts) with other data sources in your Snowflake in a single interface to assess the security posture of your organization.
- Take advantage of Snowflake's rich ecosystem of integrations to gain new insights from your security data.

Recommended best practice is to have Panther manage this environment to ensure environments are always operating at optimal levels 24x7x365.

See [Panther Integrations](#) for a full list.



Security Use Cases

Detection

Traffic monitoring of Netflow, IDS (traffic mirroring), ALB logs, WAF

Analyzing network traffic is often very difficult because the volume of data is so high and most SIEMs cannot handle that much load. Luckily, because of Panther's Data Warehouse and ETL design, teams can onboard all helpful network traffic, such as netflow, IDS/IPS, or web traffic firewall logs. The benefit of having this data in Panther is that it can be joined to other context from hosts or applications inside of VPCs and detect behaviors like exfiltration, vulnerability scanning, access to sensitive administrative ports, and more.

Detect access to sensitive S3 data sources

Sensitive data stores should only be accessed by a small and known list of trusted entities. This rule monitors for access to sensitive data (in this case S3 buckets) and ensures only certain entities (in this case predefined IAM roles) are accessing that data store. Here we're using some minorly complex python to handle a variety of cases in just a few lines of code.

Aggregation of other AWS security sensors

GuardDuty, Macie, Config, and TrustedAdvisor. Most security sensors provide their own analysis and generate alerts that can be normalized into a single place for additional filtering, analysis, and incident response purposes. Detection-in-depth is an approach all security teams should take to build confidence that some activity is in fact bad, and Panther is the place where teams can centralize this data and detection logic.



Detect sensitive API calls in CloudTrail

Often, organizations want to reuse base images for EC2 instances in the form of AMIs, and one easy way to do this is just to make the image public. But this makes the image accessible to anyone in the world, which is often (almost always) not a good idea because these base images may contain sensitive company IP or secrets like API keys.

Threat Hunting and Investigations

Panther's log processing extracts a set of standard fields (atomic IPs, domains, etc.) in all log records, enabling fast and easy data correlation from multiple data sources.

As an example, Panther's Indicator Search allows teams to take an IP address and search for all related activity between logs like CloudTrail and S3 Access Logs to map the history of an attack.

Additionally, Panther has a high emphasis on AWS specific indicators, such as AWS instances, ARNs, AWS Account IDs and more.

Incident Response

Rules and alerts have Criticality levels often numbered from 1 to 5 (e.g. Criticality levels: INFO, LOW, MEDIUM, HIGH, CRITICAL), which determines the response velocity and vigilance that is required to mitigate the incident.



Analyzing Custom Application Logs

Most Enterprises have production application workloads running on AWS either on EC2 or systems like Kubernetes, that contain context about how users interact with applications. Most security teams want to ingest this data for the purpose of understanding access to sensitive user data, whether it's personal, financial, or otherwise.

Typically, these logs are sent to CloudWatch logs and are difficult to ingest into other SIEMs. Luckily, Panther can ingest and normalize this data for correlation, detection, and investigation. In comparison to built-in log types, users can define a YAML-based data schema to describe the field structure, type, and other metadata to parse the log into structured, searchable data.

Combining Use Cases

Joining AWS Cloud Security Posture to Log Data

- Panther scans AWS infrastructure and applies policies to it to detect misconfigurations
- This data in AWS can be joined with the collected and normalized log data to tell the full story

Utilize VPC Flow Logs With Additional Meta-Data

VPC Flow+ALB

- Correlates AWS requests to client IP addresses
- Combines both network logs in the AWS network stack
- Provides proactive load balancing in AWS

CloudTrail + VPC Flow Logs

- Provides the ability to create and maintain custom rules from AWS telemetry data
- Allows the ability to provision new instances to your existing AWS infrastructure
- Enhances Crypto Mining detection

Okta + CloudTrail logs

- Increases visibility into AWS and provides a bigger picture



Best Practices and Takeaways

- ① Log everything into a centralized AWS account
- ② Prioritize implementing and monitoring S3 access logs
- ③ Leverage Organizational CloudTrails
- ④ Use VPC logs to monitor selective traffic
- ⑤ Implement redundancy and high availability failover
- ⑥ Mandate CIS AWS Foundations Benchmark Controls
- ⑦ Implement AWS-as-Security Boundaries
- ⑧ Recommended Integrations

Platforms such as Tines are highly complementary to modern solutions to take action on generated alerts. Analysts and Engineers should not be performing manual and repetitive tasks and it is recommended to prioritize the ability to easily configurable alerts on to automation for taking action. This helps scale detection programs by pinging users, opening cases, or preventing unnecessary alerts from reaching your security team. When an alert is generated from the detection engines (or other sources), a modern solution will dispatch a notification to security teams to triage the alert in systems like PagerDuty, Slack, Jira, and Microsoft Teams.

With the swivel seat removed from the equation, security teams can better streamline their process workflows. Organizations are now able to implement complex tasks such as instantly mitigating vulnerabilities as well as implement automated actions at will and as needed. The use cases and examples for this critical capability are endless.



Get Started with Panther + AWS Today

Panther brings logs from all of your AWS accounts together into a single view and makes them searchable and digestible with speed, scale, and flexibility while operating as a robust security data platform. Panther's utilization of detection-as-code, security data lakes, and real-time alerting gives teams powerful capabilities to meet their unique needs, discover attacker behavior quickly, and answer difficult questions during a breach. Built fully on cloud-native technology, Panther will be there to meet the most demanding needs of teams.

Learn how to secure your cloud, network, applications, and endpoints with Panther Enterprise. Revolutionize your security operation.

[Request a Demo](#) →

