

The background features a dark purple gradient on the left and a vibrant, multi-colored geometric design on the right. The geometric design consists of overlapping triangles and polygons in shades of magenta, blue, and orange, separated by thin white lines. The overall aesthetic is modern and tech-oriented.

# AWS re:Invent

DECEMBER 1 - 5, 2025 | LAS VEGAS, NV



SEC327-S

# Detection Engineering at Scale

## Building a High-Fidelity Security Operation

**Nathan Pitchaikani**

Senior Security Engineer

Riot Games

**Andrew Krug**

Head of Security Advocacy and Research

Datadog



# What's a Senior Security Advocate?

Datadog

5 years

Security Advocacy and  
Research



# What's a Senior Security Engineer?

Riot Games

6 years

Detection and Response

Security Engineering



# What's a Senior Security Engineer?

Riot Games

6 years

Detection and Response

Security Engineering

3 Cats



# Detection engineering is hard





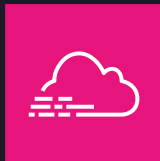


Logs as infinite as stars in the  
night sky

# Cloud practitioners are dealing with TONs of log sources



Endpoints



AWS CloudTrail



Flow logs



Amazon Bedrock

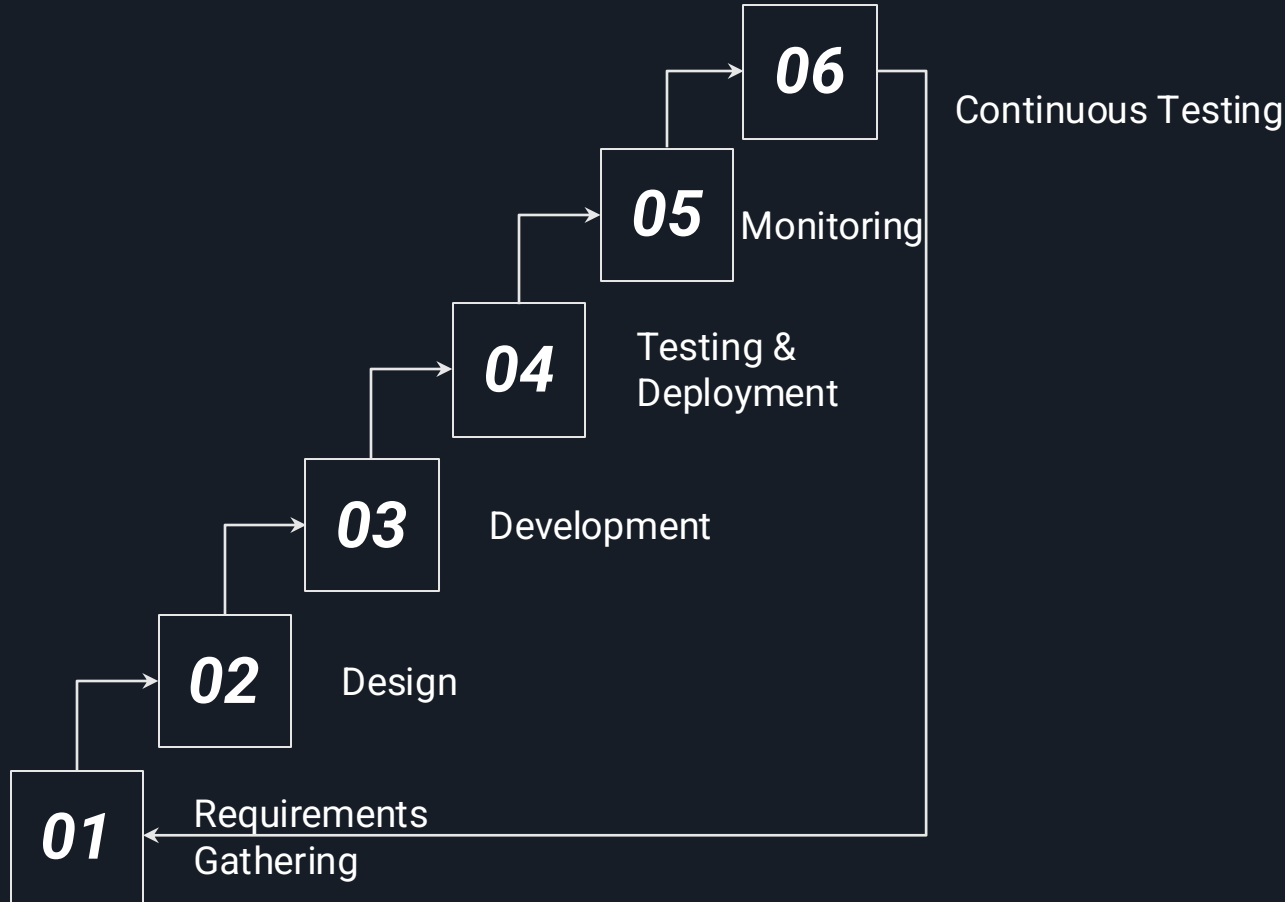


# Goal!



## Detect Bad Stuff

# Detection Development Lifecycle



**DDLC**

Detection as Code  
Principles

**SDLC**

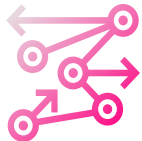
# Manual Detection Engineering Challenges



**Lack of version control and audit history**



**Siloed development**



**Inconsistent quality and output**

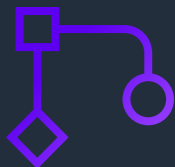


**Hard to manage at scale**

# Why Detection as Code? (DaC)



# Detection as Code (DaC) Benefits



Version Control &  
Peer Review



Testing & Validation



CI / CD Deployment

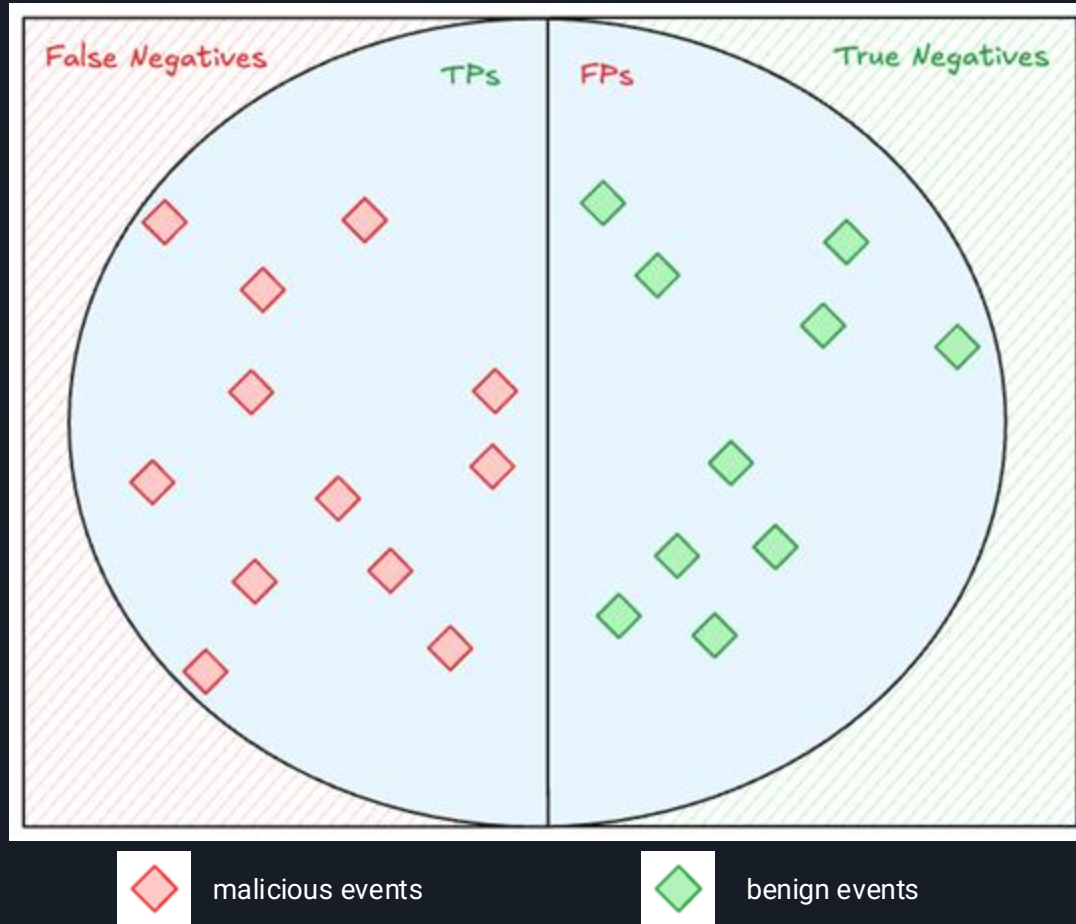


Reusability &  
Modularity

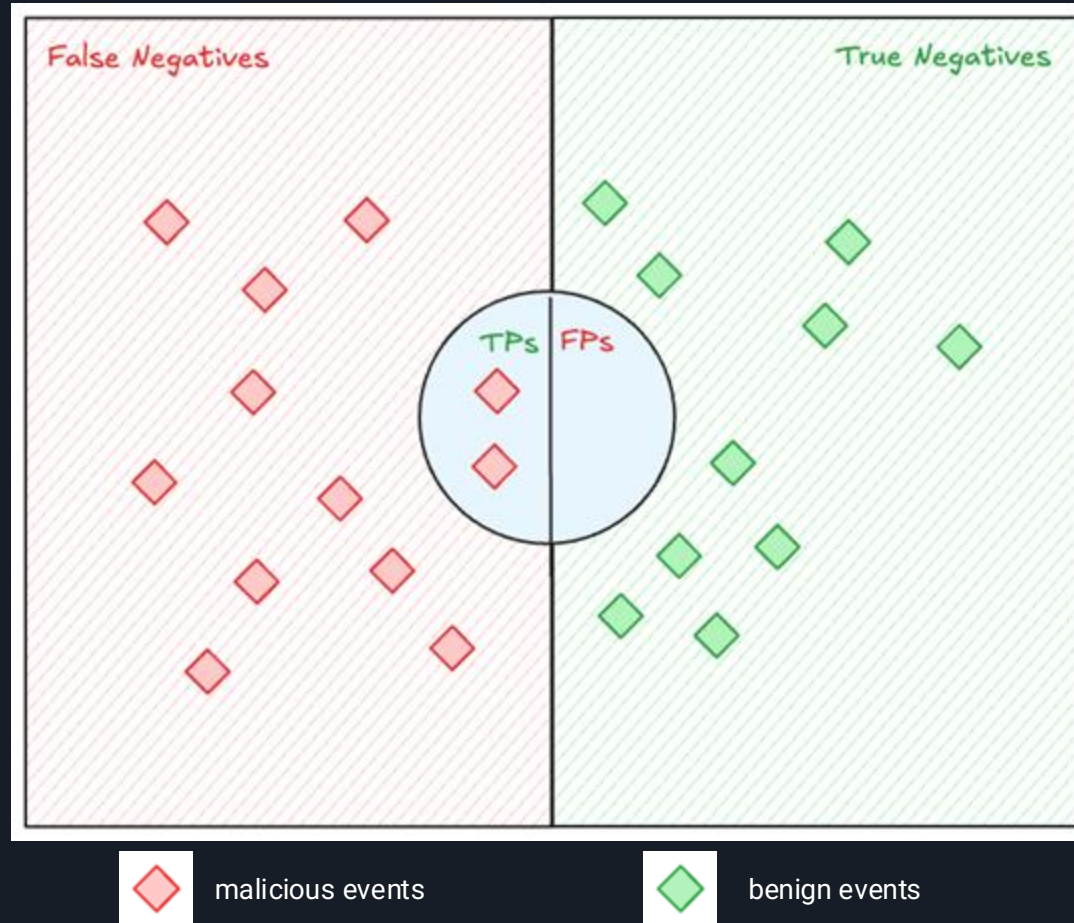
**The goal of a detection  
engineer is to write a perfectly  
accurate detection...**  
**but this is practically  
unachievable**



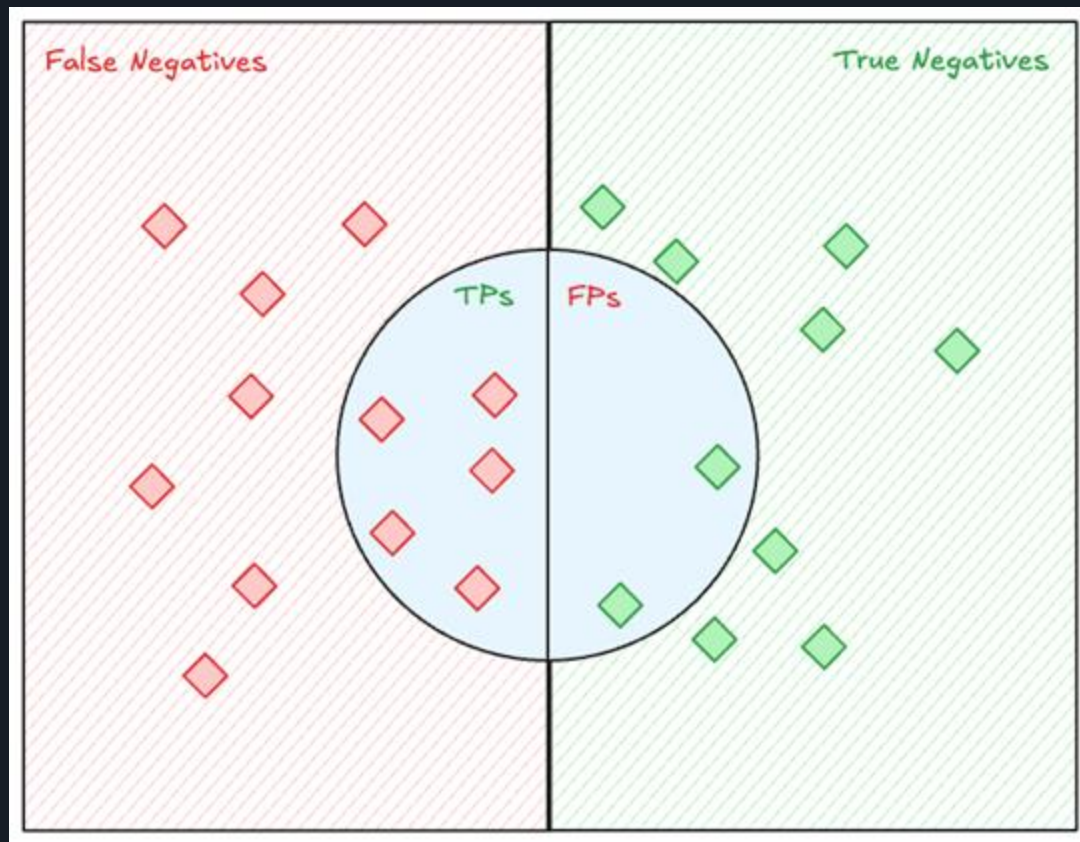
# Perfect Recall



# Perfect Precision



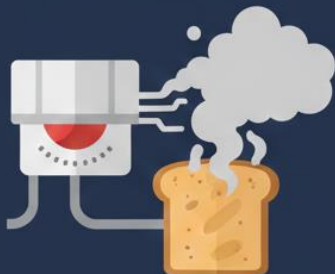
# The Middle Ground



# The precision recall tradeoff



## MAXIMIZE RECALL



Analagy: Smoke Detector  
(Better safe than sorry)

- Goal: Don't miss ANY threats  
(Minimize False Negatives)
- Cost of Error: HIGH  
(Breach, Data Loss)

## MAXIMIZE PRECISION

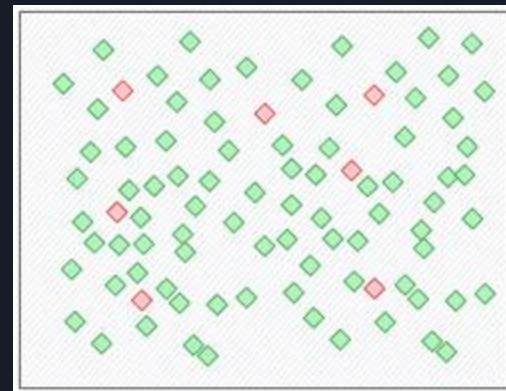


Analagy: Spam Filter  
(Avoid false alarms)

- Goal: Only show TRUE threats  
(Minimize False Positives)
- Cost of Error: HIGH  
(Analyst Burnout, Wasted Time)

Prioriize Recall      Precision

REALITY: You can't have both perfectly.  
Choose based io risk.



benign events



malicious events

# Finding Your Middle Ground

Understand your **tolerance for false positives**

Use **historical data** to gauge how a rule will perform

**Continuously monitor** rule performance (false positive rates)

Determine the **impact** of tuning a rule

# Welcome Riot Games







# Challenge #1

**Too much  
logging breaks  
the bank**



**Too little loses  
the signal**

# SIEM OPTIMIZATION

Launch Your Security Into Orbit



## Filter early, save big.

Drop noise before it hits the SIEM.



## Adapt on the fly.

Update filters without breaking the pipeline.



## Normalize for clarity.

Clean, consistent data = better detections.



## Stay resilient.

Buffer logs and prevent data loss during spikes.



## Scale with confidence.

Handle growing log volumes without choking the SIEM.



## Enforce data hygiene.

Tag, sanitize, and structure logs before they ever leave the source.

### MISSION CONTROL PIPELINE

Raw Logs



Smart Filtering

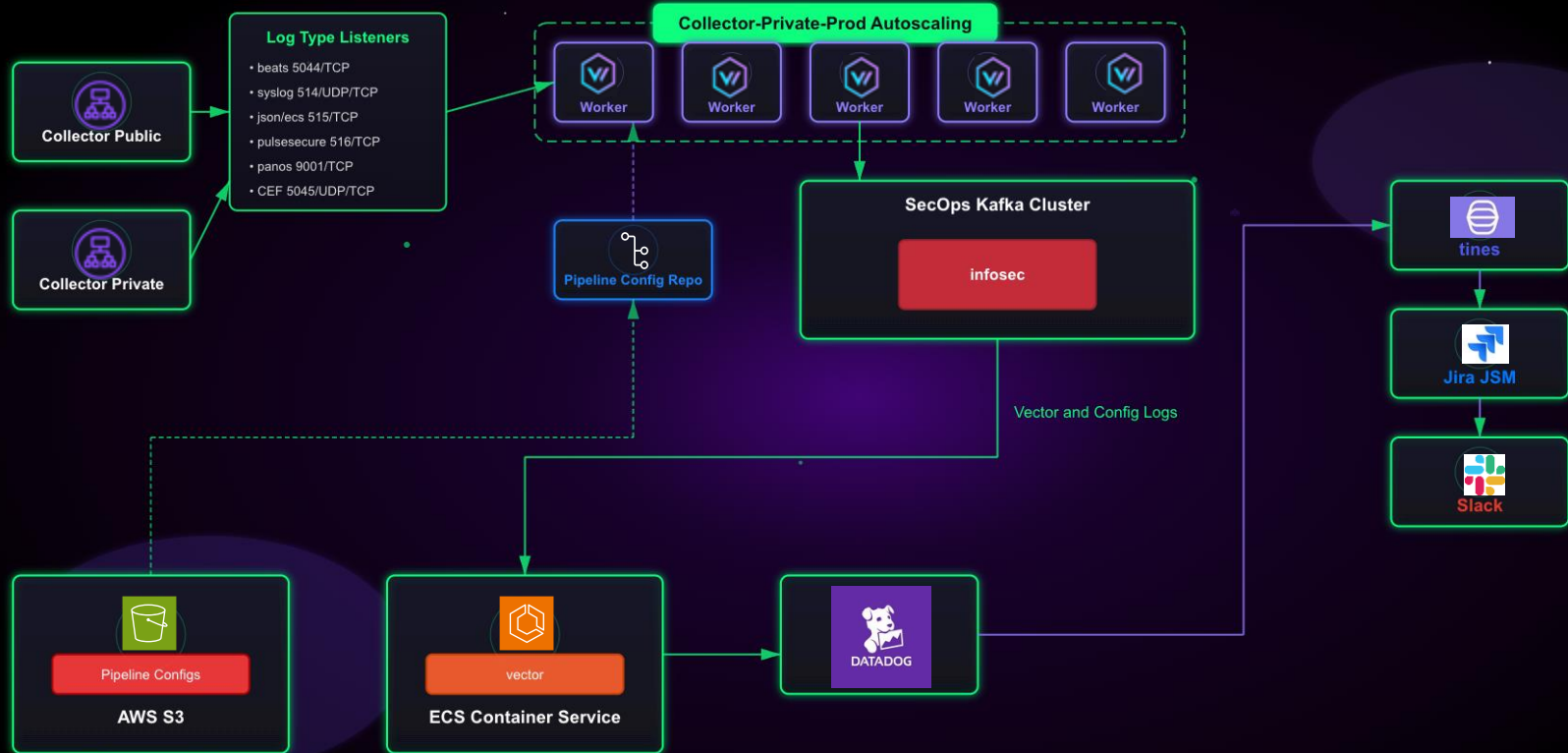


Normalized Data



SIEM

# Log Pipeline Architecture



vector.yaml

```
# vector.yaml - Simple: Windows Event Log -> Datadog
```

```
data_dir: "C:\\ProgramData\\Vector"
```

```
sources:
```

```
  win_events:
```

```
    type: windows_event_log
```

```
    channels: ["Security"]
```

```
transforms:
```

```
  add_owner:
```

```
    type: remap
```

```
    inputs: [win_events]
```

```
source: |
```

```
  .log_source_owner = "Security Team"
```

```
sinks:
```

```
  datadog:
```

```
    type: datadog_logs
```

```
    inputs: [remove_unused]
```

```
    api_key: "${DATADOG_API_KEY}"
```

```
    site: "datadoghq.com"
```

```
    encoding:
```

```
      codec: json
```

```
      compression: gzip
```

## KEY CONFIGURATION POINTS

- 1 Data Directory**  
Persistent storage location for Vector's internal state and buffer data
- 2 Source Type**  
Native Windows Event Log integration for direct event collection
- 3 Security Channel**  
Collects Windows Security events including logins, permissions, and audit logs
- 4 Data Enrichment**  
Adds custom metadata field to identify log ownership for filtering
- 5 Datadog Sink**  
Direct integration with Datadog's log ingestion API

vector.yaml

```
# vector.yaml - CloudTrail (S3) -> remove field -> Datadog
data_dir: "/var/lib/vector"

sources:
  cloudtrail_s3:
    type: aws_s3
    bucket: "my-cloudtrail-logs" # CHANGE: your bucket name
    region: "us-east-1" # CHANGE if needed
    key_prefix: "AWSLogs/" # typical CloudTrail prefix
    compression: auto
    encoding:
      codec: json

transforms:
  remove_unneeded_field:
    type: remap
    inputs: [cloudtrail_s3]

source: |
  del(.eventVersion) # example: remove verbose field

sinks:
  datadog_logs:
    type: datadog_logs
    inputs: [remove_unneeded_field]
    api_key: "${DATADOG_API_KEY}"
    site: "datadoghq.com" # e.g., us5.datadoghq.com if applicable
    encoding:
      codec: ison
```

## KEY CONFIGURATION POINTS

- 1 AWS S3 Source**  
Native integration for reading CloudTrail logs directly from S3 buckets
- 2 S3 Bucket Configuration**  
Specify your CloudTrail logs bucket with automatic JSON parsing
- 3 Field Removal Transform**  
Reduces data volume by removing unnecessary fields before forwarding
- 4 Datadog Destination**  
Sends processed CloudTrail logs to Datadog for analysis and alerting
- 5 Batching Configuration**  
Optimizes throughput with 5000 events per batch or 2-second timeout

# DATA PIPELINE EXCELLENCE

Open Source High-Performance Observability at Warp Speed



## Built for speed.

Rust-powered and lightweight for high performance.



## Collect once, send anywhere.

Unified pipeline for logs, metrics, and traces.



## Filter at the edge.

Cut noise and cost before data hits your SIEM.



## Enrich on the fly.

Parse, tag, and transform logs in motion.



## Scale without stress.

Handles massive throughput with built-in buffering.



## Stay in control.

Real-time visibility into pipeline health and flow.



## Flexible by design.

Integrates with any source or destination seamlessly.

### VECTOR PIPELINE ARCHITECTURE

Sources



Transform



Vector Core



Route



Sinks

LEARN MORE & CONTRIBUTE

[vector.dev](https://vector.dev)



SCAN TO VISIT



# SIEM LOG INDEXING: EXPECTATIONS VS REALITY

A tale of two clusters...

## WITHOUT PROPER INDEXING



```
source="firewall" AND dest_ip="10.0.0.1"
```

⌚ Search Time: 47 minutes

⚡ TIMEOUT

```
2024-01-15 10:23:45 scanning bucket 1 of 9,847...
2024-01-15 10:23:46 scanning bucket 2 of 9,847...
2024-01-15 10:23:47 scanning bucket 3 of 9,847...
2024-01-15 10:23:48 still searching...
2024-01-15 10:23:49 warning: taking for update...
2024-01-15 10:23:50 error: QUERY FAILED
2024-01-15 10:23:51 range...
```

🔥 HIGH MEMORY

❌ QUERY FAILED

🔥 CPU 99%

💣 OOM KILLER

📀 10TB scanned

💰 \$500/query

😴 Analysts sleeping

VS

## WITH SMART INDEXING



```
source="firewall" AND dest_ip="10.0.0.1"
```

⚡ Search Time: 0.3 seconds

```
✓ Found 1,234 matching events
✓ Indexed fields: source, dest_ip, timestamp
✓ Query optimized using bloom filters
✓ Scanned only relevant partitions
✓ Docs is segmented
✓ Getting promotion
✓ Life is good
```

📀 50MB scanned

💰 \$0.05/query

😊 Analysts happy

"My SIEM without indexing is like a library where all the books are thrown in a pile"

...and the librarian is on fire 🔥📖

Why think  
about filtering  
log sources?

Cost Optimization

Signal to Noise Ratio

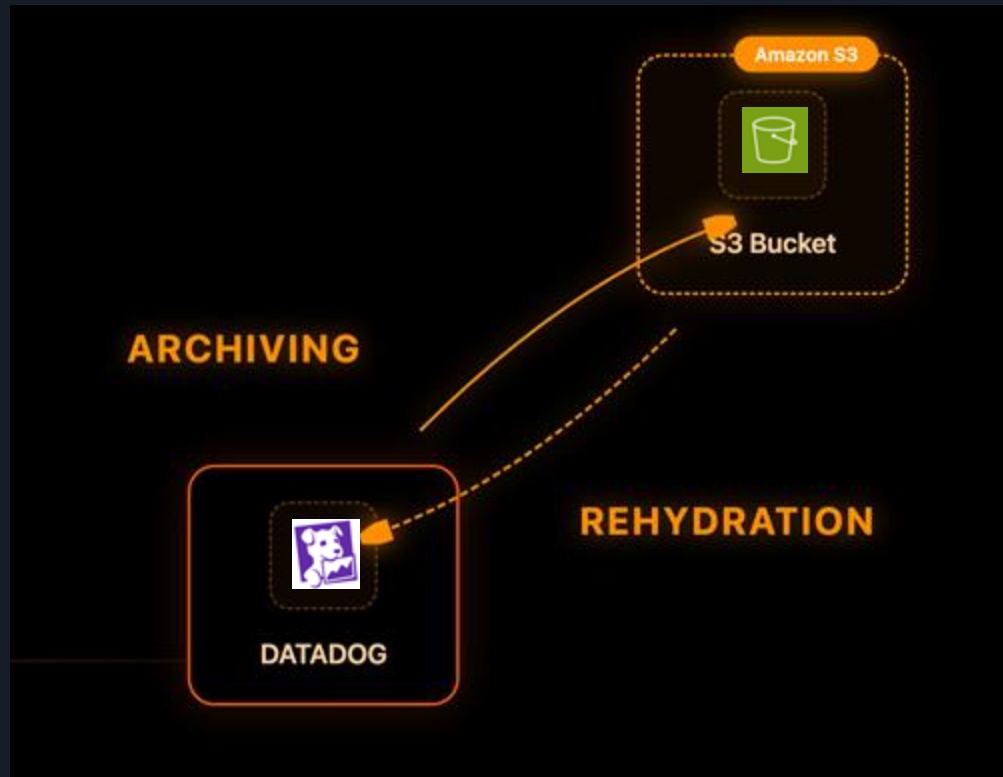
Data Quality Control



# Storage + Rehydration

Jobs handled by the Cloud SIEM:

Storage  
Indexing  
Rehydration



# Storage

- **Cloud-based log storage solution** Automatically forwards all logs to your AWS S3 bucket in compressed JSON format with date-based organization
- **Cost-optimized with flexible storage classes** Supports S3 storage tiers (Standard to Deep Archive) with lifecycle policies, reducing costs up to 68%
- **Rehydration capability for on-demand analysis** Retrieves archived logs back into Datadog within hours for compliance audits or historical investigations

## Documentation link:

[https://docs.datadoghq.com/logs/log\\_configuration/archives/](https://docs.datadoghq.com/logs/log_configuration/archives/)

Add an archive

- 1 Set Archive Name  
Set Archive Name  
VPN Logs
- 2 Define Which Data To Forward [View Full Documentation](#)  
Note: Data is routed to the first matching archive. Data is not duplicated across multiple archives.  
Filter  
source:pan.firewall @type:GLOBALPROTECT
- 3 Select Archive Type  
Amazon S3 Google Cloud Storage Azure Storage
- 4 Configure Bucket  
  - Log in to your AWS account
  - Grant your Datadog Role write access on the S3 Bucket and Path that will contain your archive (detailed instructions here)
  - Input Account, Bucket, and Path details below
  - Select a Storage Class (supported Storage Classes here)

AWS Account  
Select value

S3 Bucket  
bucket-name

Path (optional)  
/datadog/logs

Storage Class  
Standard

☐ I can confirm that the selected role has an IAM policy that grants the correct write-access on the above S3 Bucket and Path.
- 5 (Optional) Tags

Cancel Save



# Use Case

- 1 Questionable Activity from X user
- 2 Backtrack to who had the IP Address
- 3 Was this user allow to do that?

SUS MUCH?



# Rehydration

- **Retrieve archived logs on-demand** Log Rehydration pulls logs from your S3 archive back into Datadog's Log Explorer by creating "Historical Views" with specific time ranges and search queries
- **Fast scanning at scale** Scans and reindexes terabytes of archived logs within hours, with scan size estimation to help manage AWS data transfer costs
- **Flexible use cases** Ideal for compliance audits, investigating past incidents, analyzing historical trends, or accessing logs excluded from indexing to control costs

## Documentation link:

[https://docs.datadoghq.com/logs/log\\_configuration/rehydrating/](https://docs.datadoghq.com/logs/log_configuration/rehydrating/)

Rehydrate from Archive

1 Set Scanning Range

Select Time Period  
May 27, 10:54 pm - May 28, 10:54 pm

Select Archive  
PAN Traffic (Corp)

Archive Size  
Scan size is unknown - Estimate it to forecast scanning costs.  
Estimate

2 Set Historical View

Name Historical Index  
ipm-epitchaikand

Set Indexing Query  
source:ipm:firewall @type:GLOBALPROTECT @usr.name:ipitchaikandhrisganes.com

Stop Rehydration if Volume Exceeds  
300 million

Retain Logs for  
3 days

3 Notify Team on Rehydration Completion

Use this field to notify your teammates as soon as the historical view is ready for exploration or if an error occurs with this rehydration. Trigger notifications through integrations with the @handle syntax.

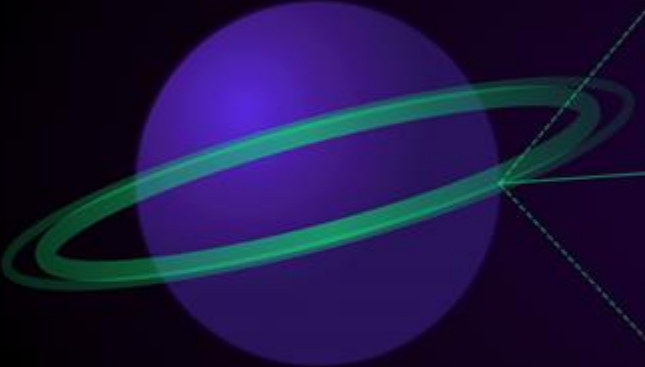
Scan size is unknown - Estimate it above to forecast scanning costs.

Cancel Rehydrate From Archive



# Datadog Archive Search

Search Through Space and Time



## 1 Search Without Rehydration

Archive Search enables you to preview and search archived logs in Flex Frozen or customer-owned S3 storage without rehydrating or moving the data

## 2 Unified Search Experience

Search archived logs using the same Datadog query language and interface, with full context including log messages, timestamps, attributes, and tags

## 3 Cost-Effective Compliance

Reduces costs and operational overhead by eliminating data movement, while accelerating security audits, compliance reviews, and historical investigations

[View Documentation](#)

<https://www.datadoghq.com/blog/archive-search/>

Archive Search is currently available in Preview for Flex Frozen tier and customer-owned S3 storage



## Challenge #2

**Detection quality doesn't come from more rules — it comes from better process**



# Detection as Code – Modernizing Threat Detection

## What is Detection as Code (DaC)?

Detection as Code is a modern security approach that treats threat detection logic—such as SIEM alerts and security rules—as software code. Instead of manually creating rules in a UI, DaC applies software development lifecycle (SDLC) principles to build, test, and deploy detections in a structured, automated, and reliable way.

## ⚡ Core Principles



### Version Control

- All detection logic stored in Git
- Complete, auditable history of every change



### Automated Testing

- Validated against historical data
- Ensures accuracy, minimizes false positives



### Collaboration & Review

- Pull/merge request workflow
- Improves quality and effectiveness



### Automation (CI/CD)


- Automated testing and deployment
- Rapid rollout at scale



*"If you wouldn't deploy a microservice without tests or version control, why would you deploy a detection rule that alerts your SOC at 2AM without the same rigor? Detection as Code is not about adding complexity – it's about making detection engineering sustainable."*

— Zack Allen, Sr. Director of Security Detection & Research

# Demo: Managing a Rule in GUI

**DATADOG**

Q Go to... Ctrl + K

Recent

Bits AI

Dashboards

Monitors

Service Mgmt

Actions

Infrastructure

Cloud Cost

APM

Digital Experience

Software Delivery

Security

AI Observability

Errors

Metrics

Logs

Integrations

5gxlpn7ndn@d...  
sxg2xw25ynpe-WS

Invite

Support

Help

Welcome, 5gxlpn7ndn! 

Get Started

You are 33% done setting up

You have 14 days left in your trial. 

Upgrade

Cloud SIEM

Overview

Content Packs

Signals

Detections

Investigate

Settings

Detection Rules

Historical Jobs

## Detection Rules

Manage and define the rules scanning your applications. Learn more about detection rules [in our documentation](#).

Rules List

MITRE ATT&CK Map NEW

1

enabled:true

X

Group by

None

Sort by


Creation Date

Show Controls

0 rules found

Show deprecated rules ☐

Select Rules



## No rules found

Refine your search or create new detection rules to generate custom Security Signals

+ New Rule

# Demo: Managing a Rule using IDE

The screenshot displays an IDE interface with a dark theme. The top bar shows the project name 'rg\_datadog\_cicd' and a search bar. The left sidebar contains a 'Project' view showing a file tree. The main editor area is currently empty, displaying search shortcuts. The bottom panel shows a terminal window with a command prompt.

**Project Structure:**

- rg\_datadog\_cicd C:\Users\...\Documents\Work\_Content\rg\_datadog\_cicd
  - .github
  - bin
  - detection\_rules
    - cloudtrail
  - docs
  - siem\_analyzed
  - suppression\_rules
  - venv library root
  - .gitignore
  - .yamllint
  - compose.yaml
  - Dockerfile
  - Jenkinsfile
  - README.md
- External Libraries
- Scratches and Consoles

**Search Everywhere** Double Shift

**Go to File** Ctrl+Shift+N

**Recent Files** Ctrl+E

**Navigation Bar** Alt+Home

Drop files here to open them

**Terminal** Local x

```
(venv) PS C:\Users\...\Documents\Work_Content\rg_datadog_cicd\bin>
```

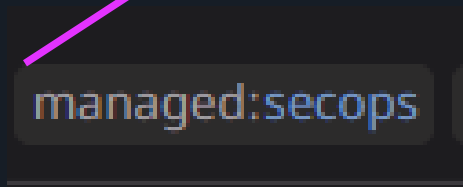
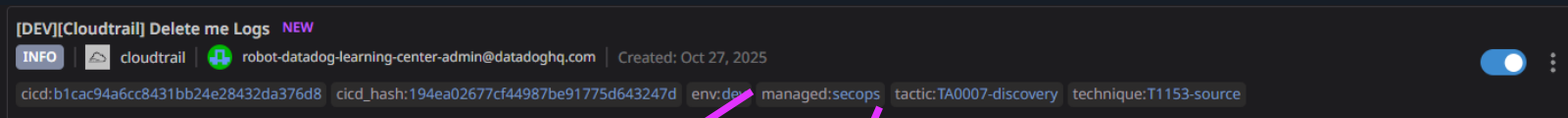
rg\_datadog\_cicd > detection\_rules > cloudtrail

Python 3.13 (rg\_datadog\_cicd)

# Enforcing CI/CD

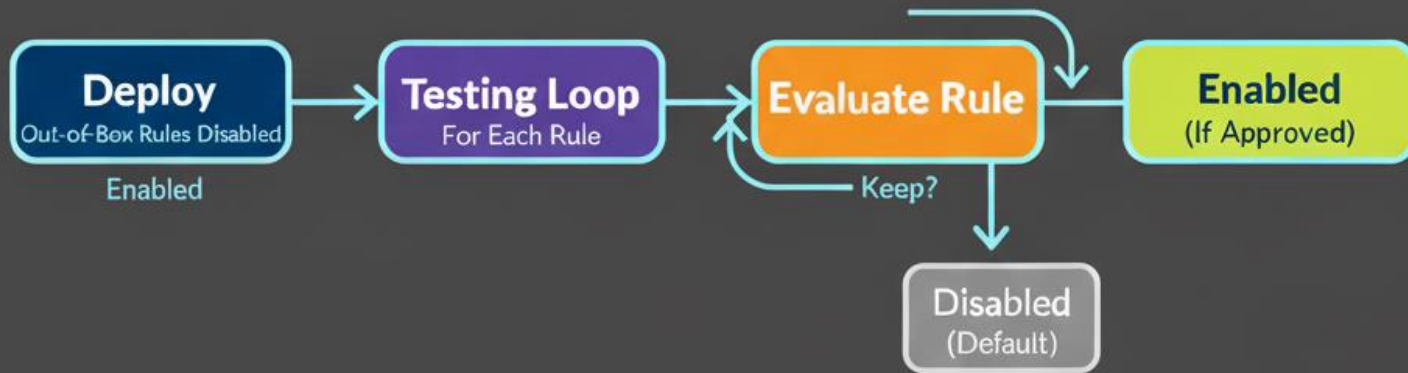


# Break Glass Situation



Manual Tagging Bypasses  
CI/CD Control

## OOTB Rules are great ... but





# Challenges in Out of Box Rules

Alert Fatigue 

Normalization Mismatches 

Coverage Gaps 

Maintenance and Lifecycle 

# Advice

OOB  
Rules  
have a lot  
of value

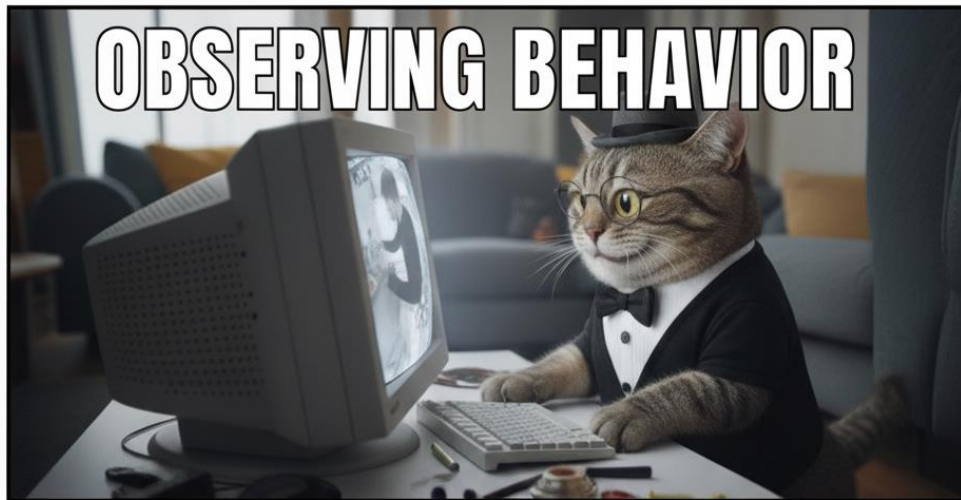
# Advice

OOB Rules  
have a lot of  
value

When paired  
with an  
activation plan

# Day 1:

## As a behavioral analyst



# Correlation: Behavior

Very useful but very noisy

Lots of malicious activities are caught by these detections



# OKTA Unusual Behavior

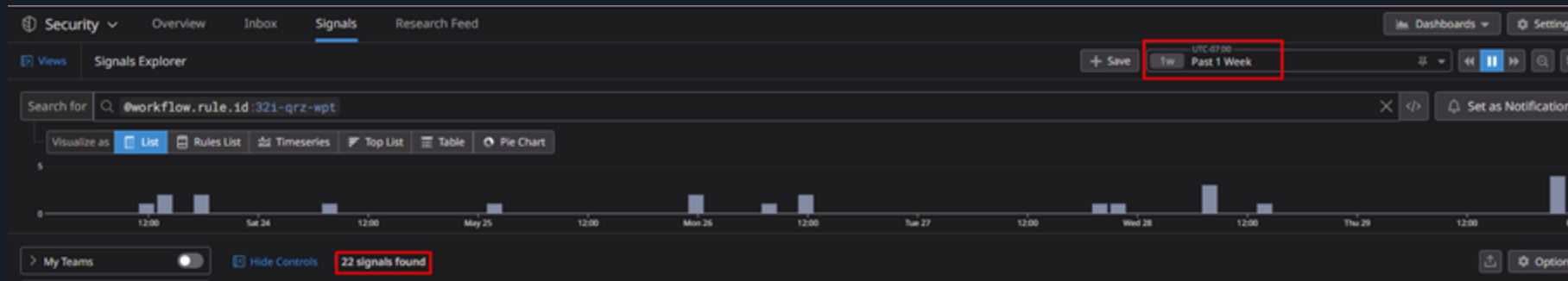
OKTA has inbuilt behavior logic

```
logOnlySecurityData {  
  behaviors {  
    New City          POSITIVE  
    New Country       NEGATIVE  
    New Device        POSITIVE  
    New Geo-Location  POSITIVE  
    New IP            POSITIVE  
    New State         POSITIVE  
    Velocity          NEGATIVE  
  }  
  risk {  
    level            MEDIUM  
    reasons          Anomalous Location, Anomalous Device  
  }  
}
```

# Example - Unusual Behavior OKTA

Is this a noisy alert?









**22 signals** in 1 week



# AWS Unusual Detections



Amazon GuardDuty

	Unusual IAM Policy Creation	72	
	Unusual IAM Role Created	44	
	API Calls From Unusual Country	31	
	Unusual Key Pair Creation	30	



# OBSERVING BEHAVIOR



# Datadog's Signal Explorer

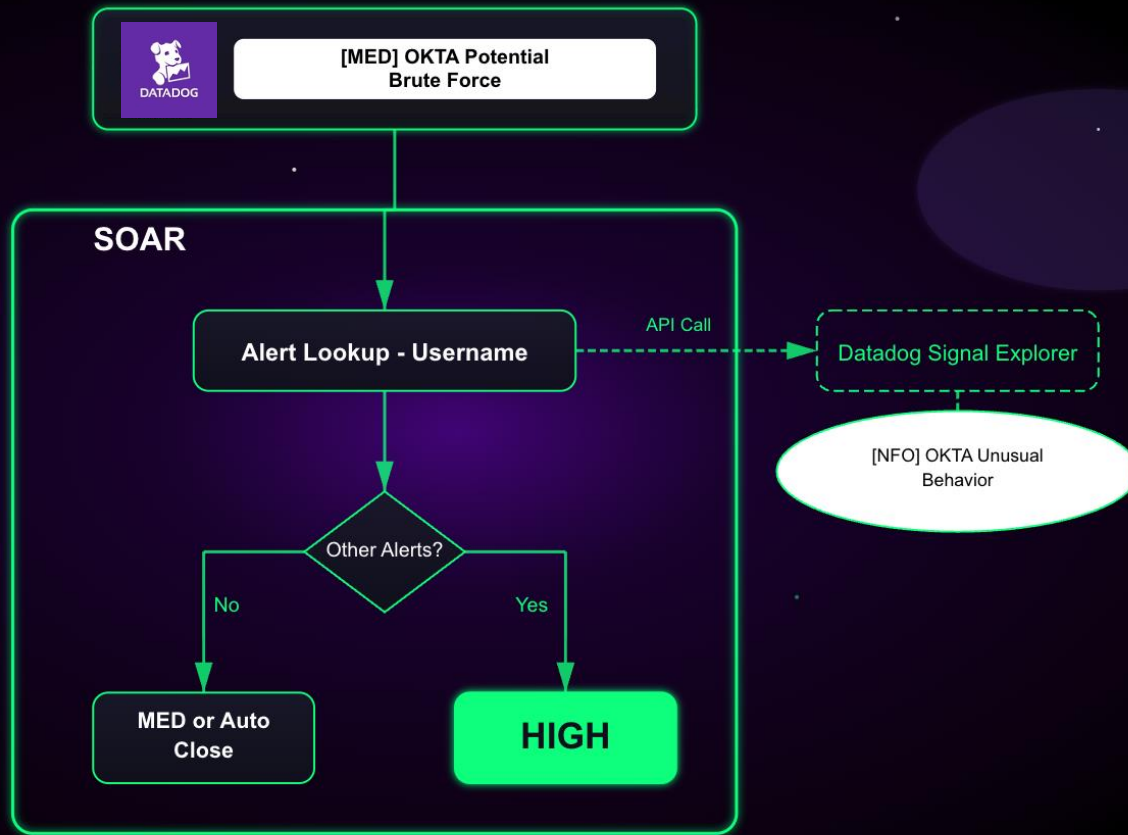


Logs for each alerts which has triggered

Lets make our Behavior alerts → NFO Alerts

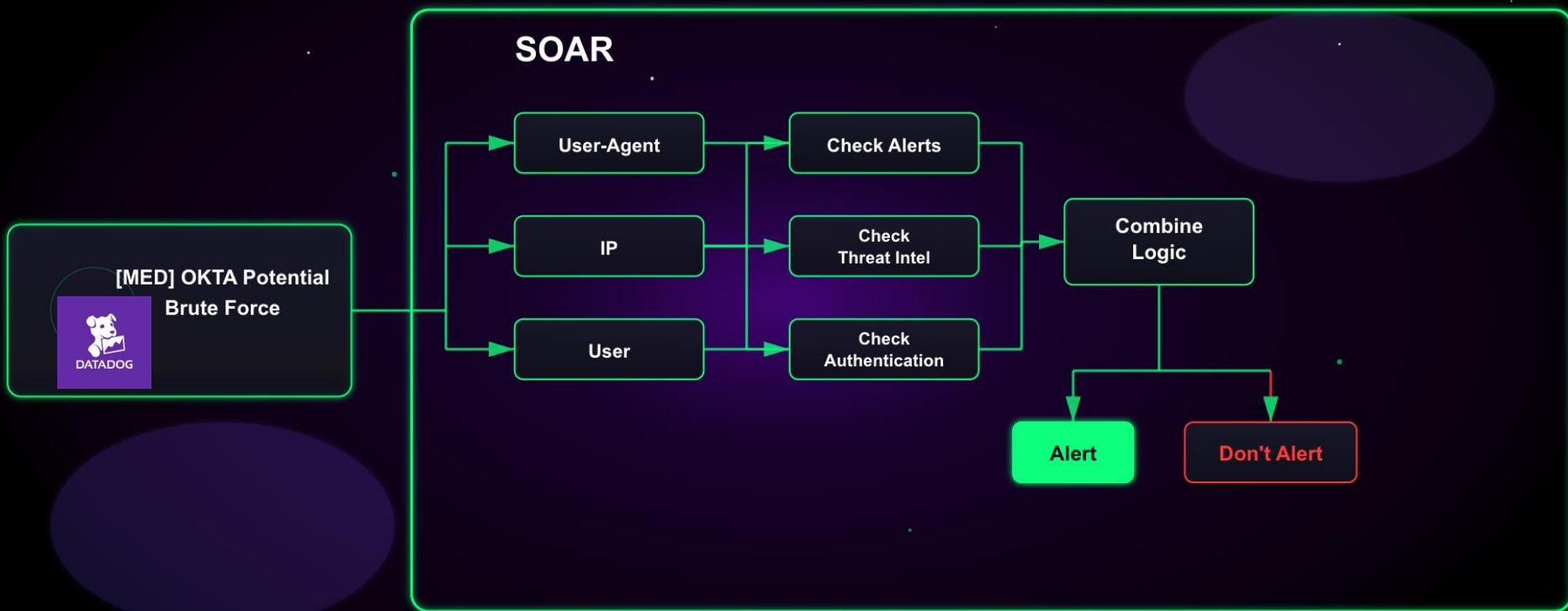
# Example

- All alerts have a few common fields
  - Hostname
  - IP
  - username
- Query Datadog Signal API for any past alerts



# Putting it all together

Automations via SOAR Tool



# Next Phase of Detection Engineering



# Benefits of OCSF

Open Cybersecurity Schema Framework

## Unified Schema

Consistent data format across all security tools eliminates translation errors

## Reduced Effort

Eliminates custom parsing pipelines, cutting integration time significantly

## Community Driven

Schema evolves with emerging threats through collaborative development

## Better Correlation

Common taxonomy enables cross-platform threat correlation and faster response

## Vendor Neutral

Open standard backed by AWS, Datadog, and 100+ organizations

## Interoperability

Seamless data exchange between SIEM, XDR, SOAR, and cloud platforms



# Use Datadog to standardize your security logs

**Remap Security Logs to OCSF**  
Created Nov 21, 2024, 11:31am

Overview Workers Latest Deployment & Setup

PROCESSORS (5) + Add

- 1. Filter
- 2. Grok Parser
- 3. Sensitive Data Scanner
- 4. GeoIP
- 5. Remap to OCSF **RECOMMENDED**

You've added 5 mappings.  
[Manage mappings](#)

SOURCE  
Syslog-ng  
90.3k events/s in | 850.1k bytes/s in

DESTINATION  
Google Chronicle  
20.7k events/s in | 194.7k bytes/s in

PROCESSORS (4) + Add

- 1. Filter
- 2. Grok Parser
- 3. Sensitive Data Scanner
- 4. GeoIP

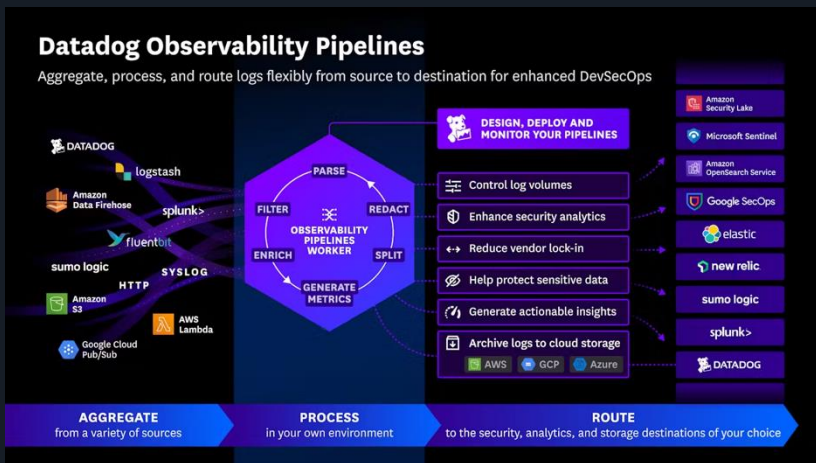
DESTINATION  
Splunk HEC  
63.2k events/s in | 583.2k bytes/s in

**Remap to OCSF: Manage Mappings**

Search log type... + Add Mapping


LOG TYPE	MAPPING TYPE	FILTER	OCSF EVENT CLASS
GCP Cloud Audit SetiamPolicy	Library	protoPayload.methodName:SetiamPolicy	Account Change (3001)
Okta System Log Authentication	Library	eventType:usersession.start	Authentication (3002)
Palo Alto Networks Firewall Traffic	Library	message:"TRAFFIC"	Network Activity (4001)
Azure Audit Change User Password	Custom	activityDisplayName:"Change user password"	Account Change (3001)
CloudTrail Account Change	Library	eventName:CreateUser	Account Change (3001)

# Two Options



Observability Pipelines

<https://bit.ly/4oo8TBz>



# VECTOR

BY DATADOG

A lightweight, ultra-fast tool for building observability pipelines

QUICKSTART VECTOR REMAP LANGUAGE COMPONENTS

Vector Open Source

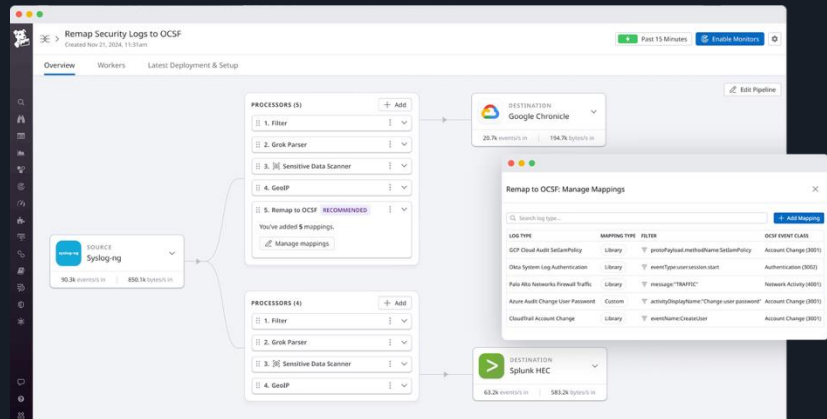
<https://bit.ly/4apRDsg>






# OCSF Processor Available Today



Stream, transform, and standardize log data in the OCSF format to improve threat detection, speed investigations, and simplify SIEM integration—without increasing cost or complexity




# Embrace new detection methods






**DATADOG**


Go to...  

- Recent
- Bits AI
- Dashboards
- Monitors
- Service Mgmt
- Actions
- Infrastructure
- Cloud Cost
- APM
- Digital Experience
- Software Delivery
- Security


 **Detection Rules** > Create a New Detection

 Enabled  Import  Export


## Create a New Rule

**1**  **Define your Real-time rule**


Choose a detection method for creating signals:




**✓ Threshold**  
Detect when events exceed a user-defined threshold.  
[Learn more](#)




**New Value**  
Detect when an attribute changes to a brand new value.  
[Learn more](#)




**Anomaly**  
Detect when a behavior deviates from its historical baseline.  
[Learn more](#)




**Content Anomaly**  
Detect when an event's content is an anomaly compared to the historical baseline.  
[Learn more](#)




**Impossible Travel**  
Detect if impossible speed is detected in user activity logs.  
[Learn more](#)




**Third Party**  
Map third-party security logs to signals, setting severity based on log attributes.  
[Learn more](#)



**Signal Correlation**  
Chain multiple rules to create higher fidelity signals.  
[Learn more](#)

**2**  **Define Search Queries**

Specify search queries to detect relevant signals. [Learn More](#) 

# Bits AI Security Analyst

Autonomous Cloud SIEM investigations powered by AI

Preview



## EVIDENCE-BASED CONCLUSIONS

- Benign — No concern
- Suspicious — Action needed
- Inconclusive — Review

## Autonomous Triage

Investigates SIEM signals by analyzing detection rules. Signals are marked with facets for filtering and notifications.

## MITRE ATT&CK Framework

Plans and executes investigations, pivoting between IOCs and querying historical signals and logs.

## Integrated SOAR Actions

Recommends remediations and enables direct execution via Action Interface with access controls.

## Part of Bits AI Agent Family

One of three AI agents supporting security, SRE, and development teams across Datadog.



# Key takeaways

## Detection as Code is essential

Bringing SDLC concepts to the way you build detections is critical to scalability

## Data pipeline is everything

Having one model for ingest that you apply to extract, transform, and load force multiplies

## Process, process, process

Process is more important than tooling when it comes to rigor

## Embrace Open Standards

Great platforms support open standards and increase interoperability



**The goal of a detection  
engineer is to write a perfectly  
accurate detection...**

**but this is ~~practically~~  
~~unachievable~~ mostly achievable**

# Resources for more

Datadog Security Labs – <https://securitylabs.datadoghq.com>

Get these slides – <https://bit.ly/sec327-s>

Stratus Red Team - <http://github.com/datadog/stratus-red-team>

OCSF Framework - <https://schema.ocsf.io/>



# Datadog for Startups

*Starting or scaling? We have you covered*



Error Tracking



Bits AI SRE\*



Product Analytics



LLM Observability



Metrics



MCP Server\*



Log Management



Real User Monitoring

*... and lots more*

## Free for a year

up to \$100k in credits



For Series A or earlier startups  
New to Datadog customers  
Referred by Partner Network

*Thank you for joining us at AWS Re:invent*

# Check out the latest re:Invent Launches



<https://bit.ly/43WleG0>



# Thank you



[linkedin.com/in/andrewkrug](https://www.linkedin.com/in/andrewkrug)



[linkedin.com/in/nathan-pitchaikani-221177b3](https://www.linkedin.com/in/nathan-pitchaikani-221177b3)

Meet Datadog's at re:Invent



<https://bit.ly/484ix7M>

Subscribe to the  
Security Labs Newsletter



<https://bit.ly/security-newsletter>

DDFS Program



<https://bit.ly/ddfs>

