Streamlined SIEM Migration and Daily Cost Optimization

Joe Cicero and Mike Pinch

Agenda

- The Problem we all face…
- Understanding Your Data Sources
- Understanding Data Pipeline Tools
- Understanding Data Lakes
- Brining it into SIEM

Introductions

Joe Cicero
Director of Strategic Alliances
Security Risk Advisors
Joe.Cicero@sra.io

Mike Pinch
Chief Technology Officer
Security Risk Advisors
Mike.Pinch@sra.io

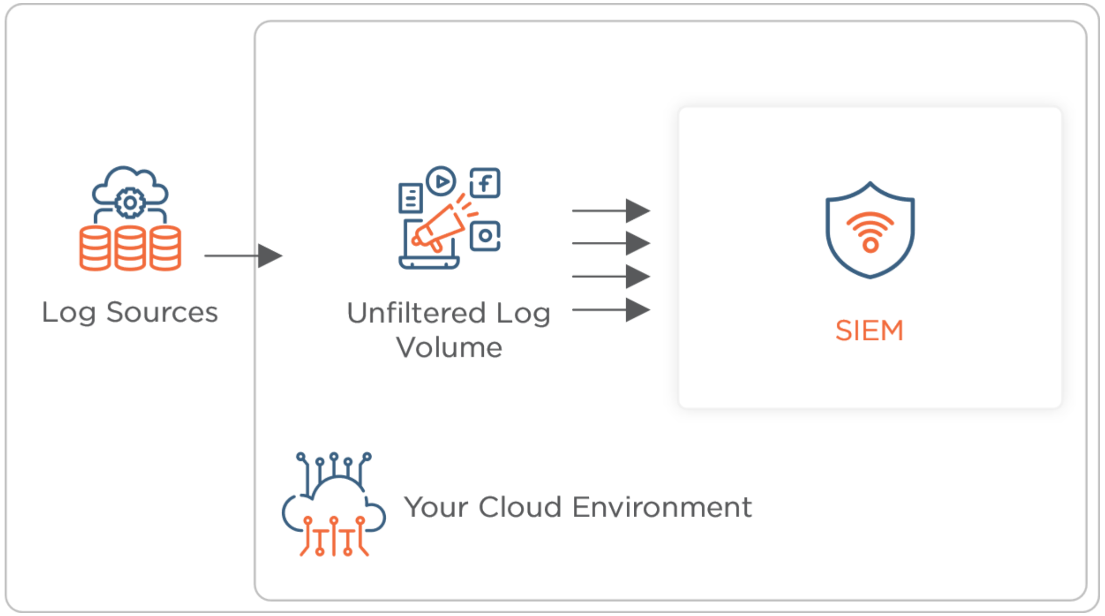
The Problem we all face…

What to collect, where to route, and what to do with it – all while controlling costs

Deconstructing a SIEM

- What does a SIEM Actually Do?
  - Data Ingestion
  - Data Storage
  - Data Search
  - Event Correlation (Sometimes Data Search)
  - Case Management
  - Data Visualization
  - Enrichment
  - Data Destruction
- SIEM Strengths
  - Case Management
  - Fast Retrieval
  - Enrichments
- SIEM Weaknesses
  - Statistical and Analytical capabilities
  - Long Term Storage
  - Cost
  - Automation
  - Interoperability with non-SIEM data/systems
  - Hunting

Stuck in a legacy model



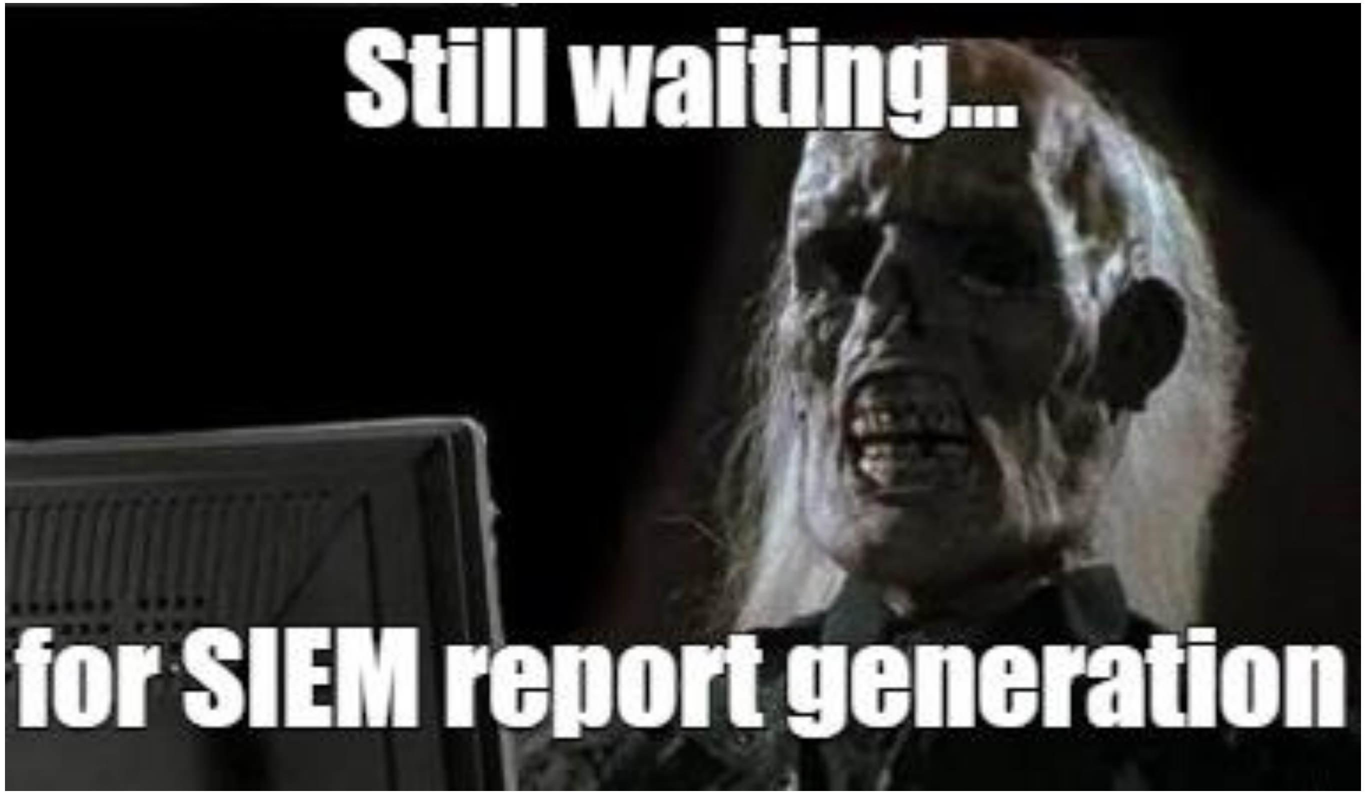
Goals:

- Reduce cost
- Increase Hot, Searchable Storage History
- Support best of breed analytics tools & AI
- Support Services / API Integration
- Never store data in more than one place
- Allow for low switching costs for new opportunities
- Allow culling, transformation, and shaping of data flows dynamically
- Support centralized management of all log data
- Provide key metrics and visibility to log source health

Transition to Data Centric Design

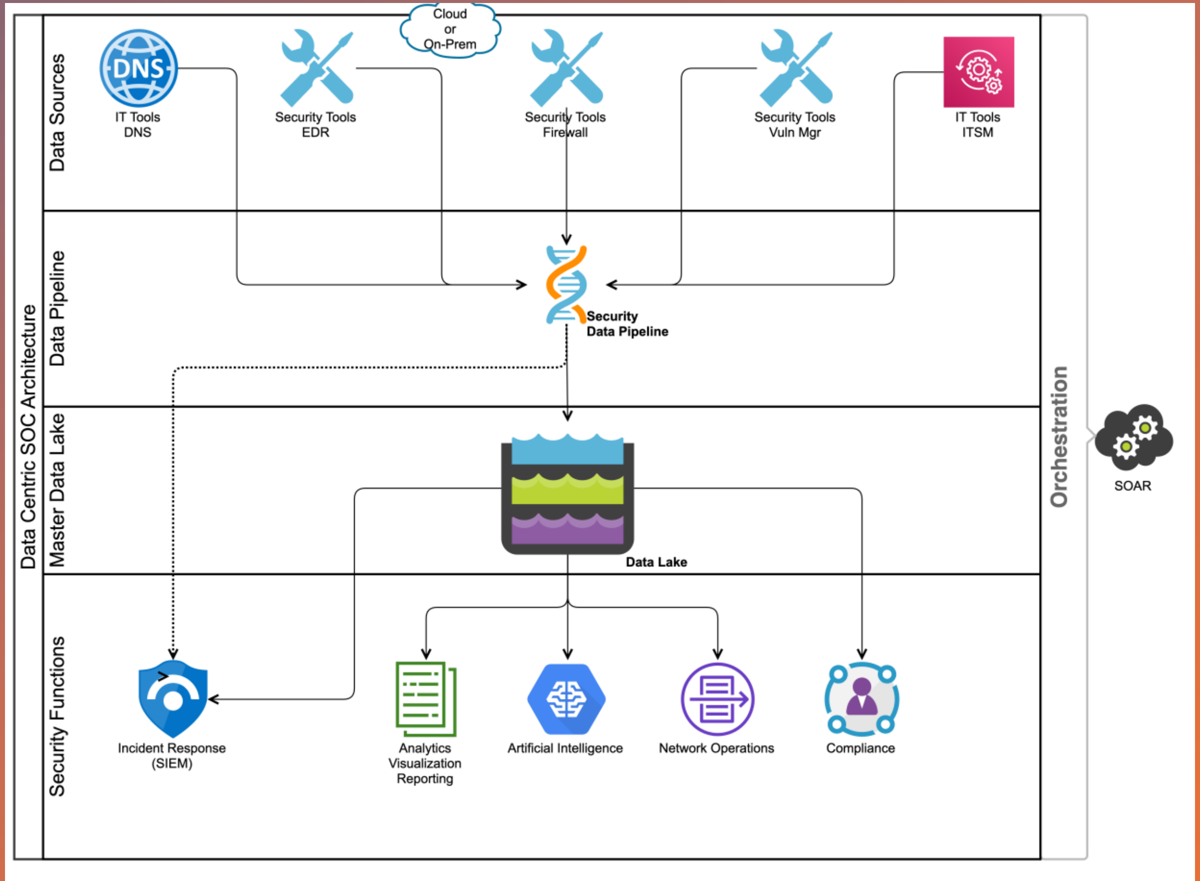
- Too many logs, too many sources, too much data – What's really necessary?
- Capacity and Cost – Events per second (EPS) vs consumption (GB/day).
- Speed
- SIEM fallacy – give me all your data…

SORRY, WE CAN'T INGEST YOUR DAILY 10 GB OF PROC CREATION EVENTS. IT'S ALREADY TOO EXPENSIVE.
200 GB OF FIREWALL LOGS
SIEM
imgflip.com



What could we do differently?

Enter the data lake...



Understanding Your Data Sources

Where and what logs...

Security Data Pipeline

A critical feature of a modern SOC operation

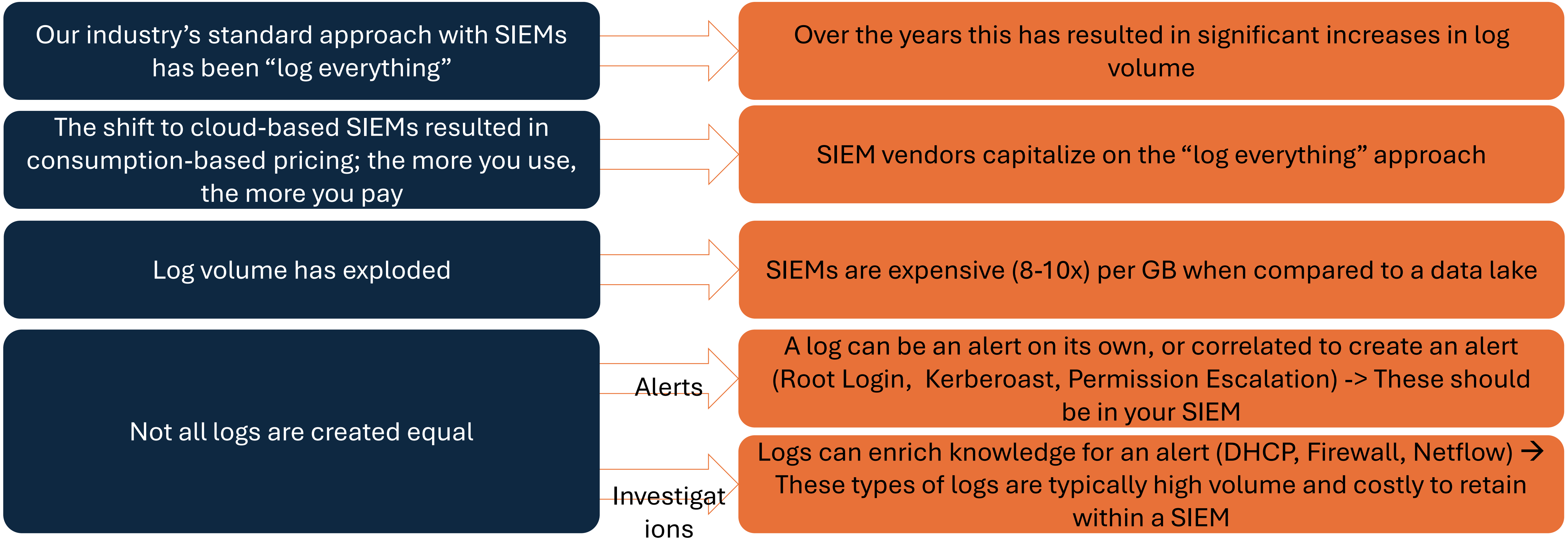
Channeling your data and taking full advantage of dynamically being able to transform, enrich, reduce, mask, and monitor your log data allows for little waste or duplication and more visibility into your environment

Enhances your agility to migrate tools and platforms by reducing the switching cost and complexity

Allows for managing data storage tiering, and only sending logs to your SIEM that are needed for detections

Put investigation and compliance logs into a lower cost, hot searchable solution

Security Data Pipeline Management



Our industry's standard approach with SIEMs has been "log everything"

Over the years this has resulted in significant increases in log volume

The shift to cloud-based SIEMs resulted in consumption-based pricing; the more you use, the more you pay

SIEM vendors capitalize on the "log everything" approach

Log volume has exploded

SIEMs are expensive (8-10x) per GB when compared to a data lake

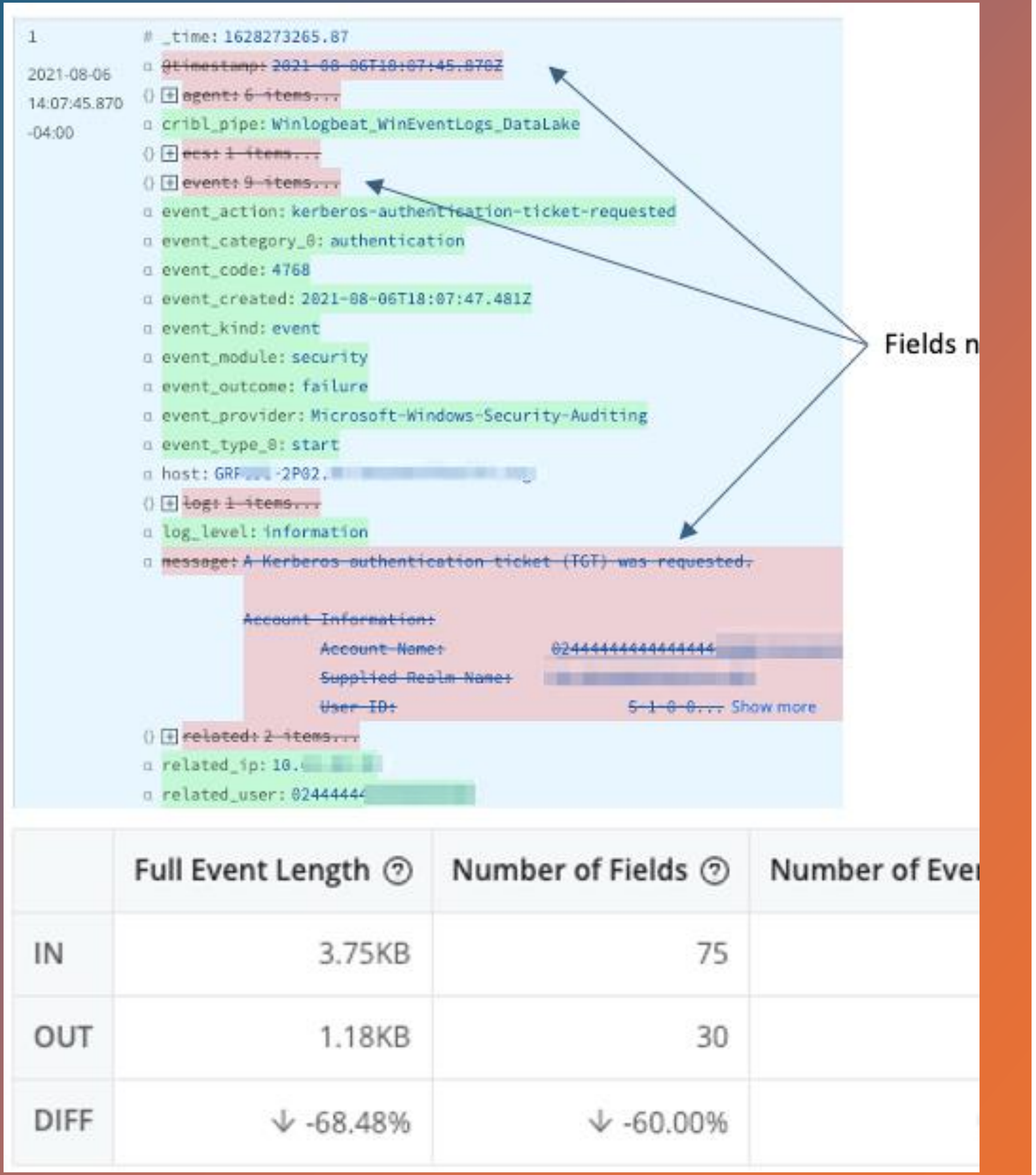
Not all logs are created equal

Alerts

A log can be an alert on its own, or correlated to create an alert (Root Login, Kerberoast, Permission Escalation) -> These should be in your SIEM

Investigations

Logs can enrich knowledge for an alert (DHCP, Firewall, Netflow) → These types of logs are typically high volume and costly to retain within a SIEM



```
1                 # _time: 1628273265.87
2021-08-06        ~~@timestamp: 2021-08-06T18:07:45.870Z~~
14:07:45.870      {} ⊞ ~~agent: 6 items...~~
-04:00            cribl_pipe: Winlogbeat_WinEventLogs_DataLake
                  {} ⊞ ~~ecs: 1 items...~~
                  {} ⊞ ~~event: 9 items...~~
                  event_action: kerberos-authentication-ticket-requested
                  event_category_0: authentication
                  event_code: 4768
                  event_created: 2021-08-06T18:07:47.481Z
                  event_kind: event
                  event_module: security
                  event_outcome: failure
                  event_provider: Microsoft-Windows-Security-Auditing
                  event_type_0: start
                  host: GRP[redacted]-2P02.[redacted]
                  {} ⊞ ~~log: 1 items...~~
                  log_level: information
                  ~~message: A Kerberos authentication ticket (TGT) was requested.~~

                          ~~Account Information:~~
                                 ~~Account Name:~~           ~~02444444444444444~~[redacted]
                                 ~~Supplied Realm Name:~~    [redacted]
                                 ~~User ID:~~                       ~~S-1-0-0...~~ Show more
                  {} ⊞ ~~related: 2 items...~~
                  related_ip: 10.[redacted]
                  related_user: 024444[redacted]
```

	Full Event Length ⓘ	Number of Fields ⓘ	Number of Eve
IN	3.75KB	75	
OUT	1.18KB	30	
DIFF	↓ -68.48%	↓ -60.00%	

Log Cleansing

- **Reducing the noise by eliminating unnecessary fields within log files**

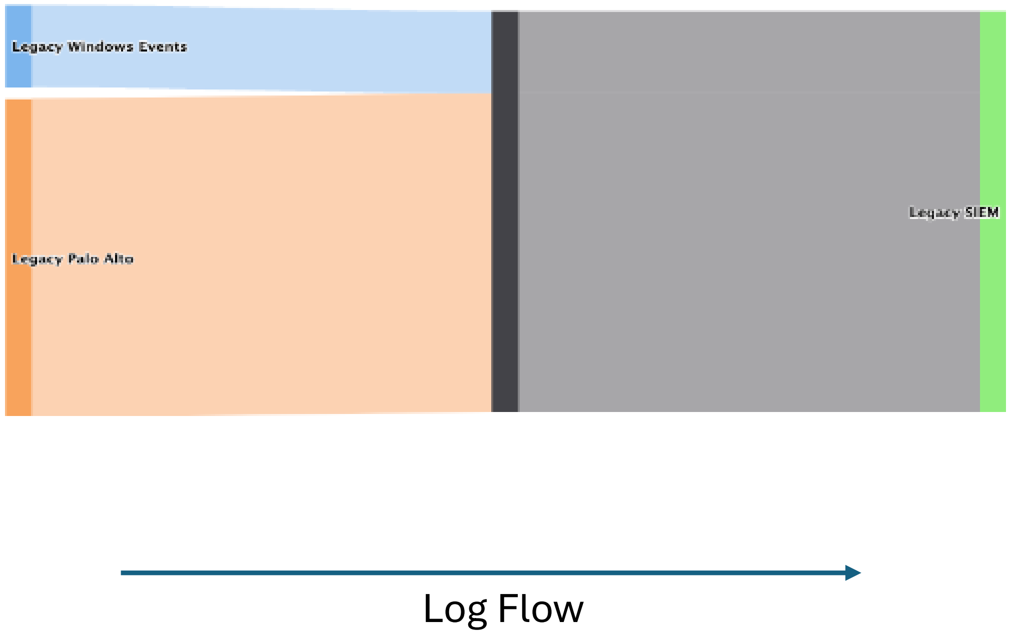
- **Native Windows Event Log**
  - 75 Fields Per Log
  - 3.75kb per log
  - Redundant Fields
  - Unneeded Fields
  - Mix of Critical and Non-Critical Event Types

- **Same Event After Processing**
  - 30 Fields per log
  - 1.18kb per log
  - Removed Redundancy
  - Removed Noise
  - Intelligent Routing to Data Lake or Sentinel

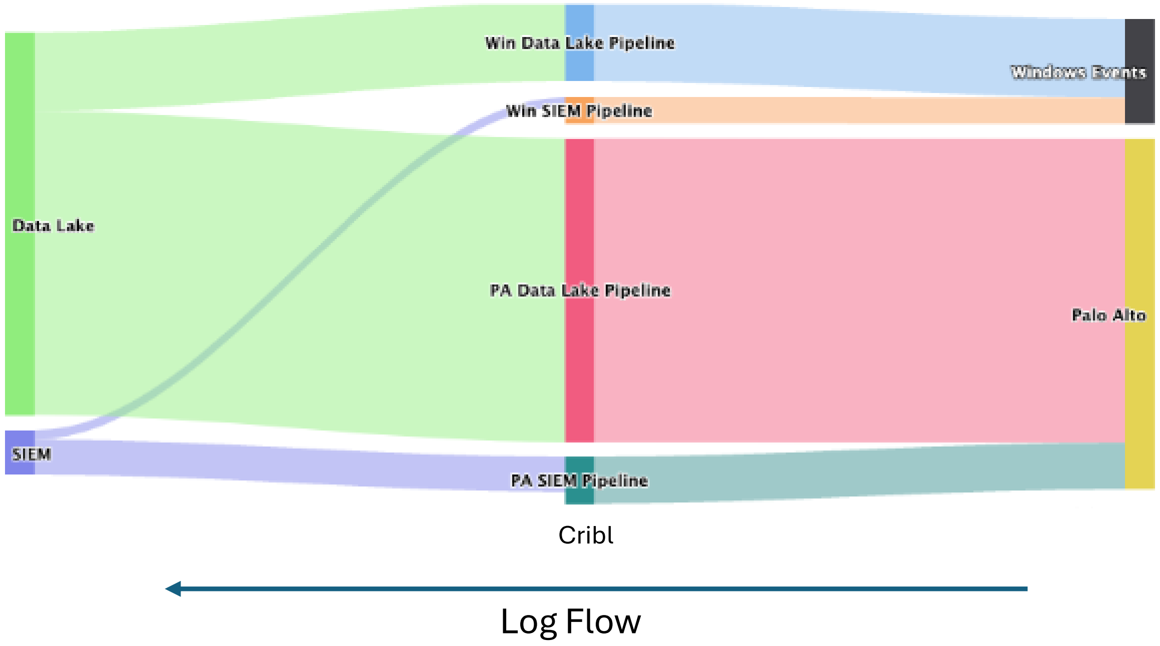
- **Outcome**
  - 68.5% log size reduction (log cleansing)
  - 90%+ cost reduction (log routing – next slide)

Log Routing

Traditional Log Routing

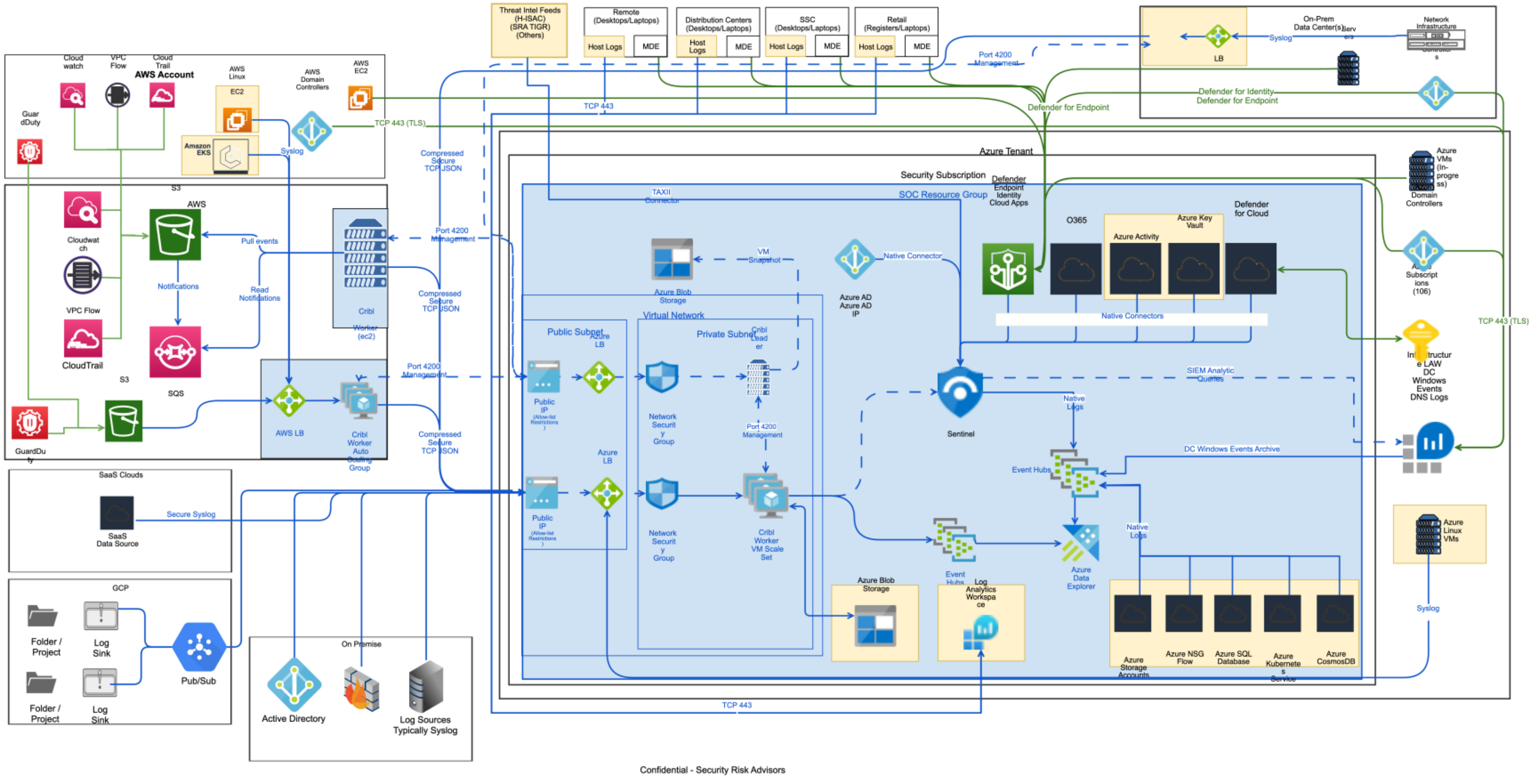


Security Data Pipeline Routing



Prioritized log storage results in stronger detections, lower costs, and longer-term hunting capabilities

Data pipelines in the wild



- The preferred architecture we've seen most successful is
  - Data Pipeline
    - Cribl
  - Data Lake
    - Azure Data Explorer
  - SIEM
    - Azure Sentinel

Pipeline as a cost saver

Fortune 200 Company

- Splunk Cloud with ES
- Ingestion rate of 1.2TB/day
- Modernized pipeline using Cribl to reduce ingestion volume
  - Reduction of ~430GB/day
  - Splunk Licensing Cost Reduction/Repurposed
  - Scope included firewall and VPN logs
  - Additional savings to be realized through data lake implementation

$400K Annual Savings Realized

Large Healthcare Provider

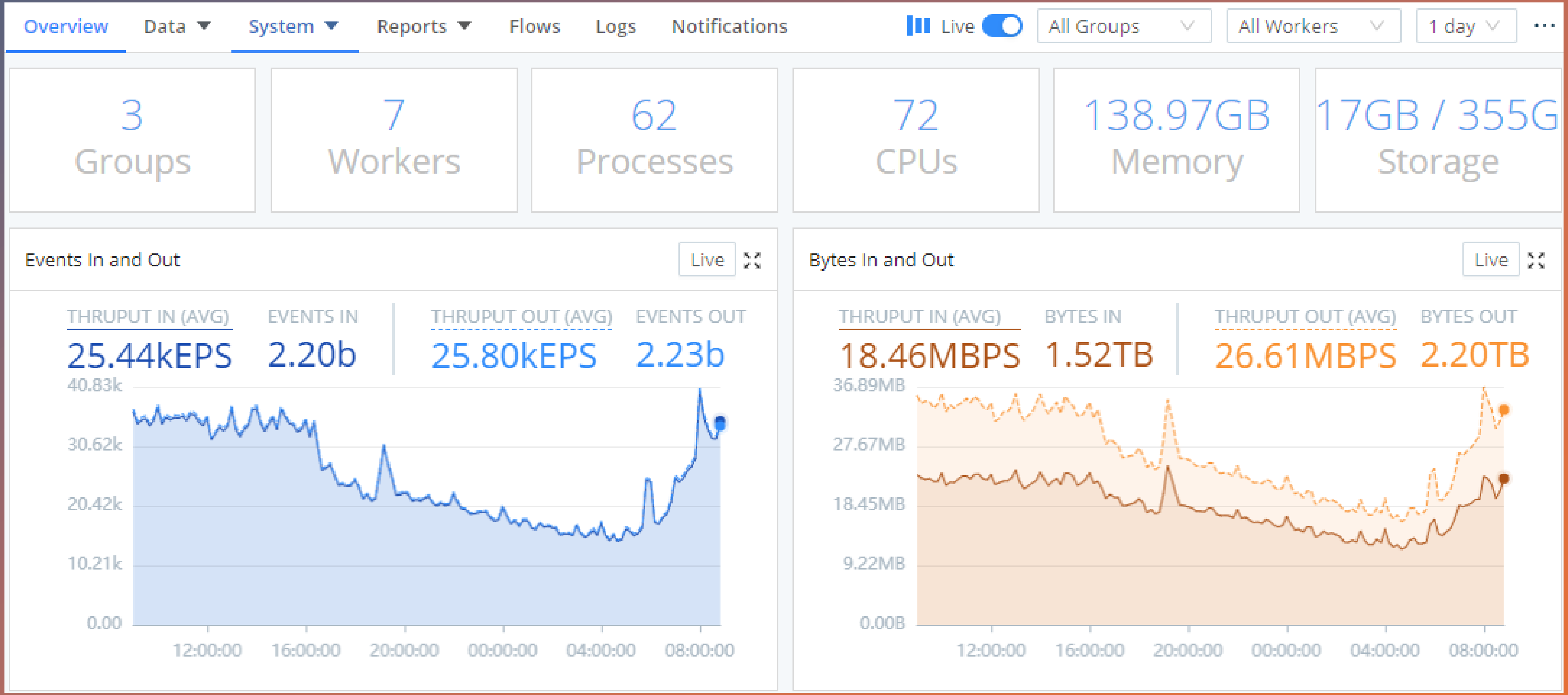
- Migrated from Splunk Cloud to Azure Sentinel SIEM
- Sentinel and Azure Data Explorer, Cribl for pipeline management
  - Splunk costs ~$900k/yr (Before)
  - Cribl & Azure Sentinel / ADX Costs ~$200k/yr (After)
  - Increased data retention timeline 4x
  - Improved detection capabilities
  - Estimated savings >80% in projected SIEM costs

$700K Annual Savings Realized

Understanding Data Pipeline Tools

A look at Cribl's FREE Version

Cribl (free) Pipeline Management



3 Groups

7 Workers

62 Processes

72 CPUs

138.97GB Memory

17GB / 355G Storage

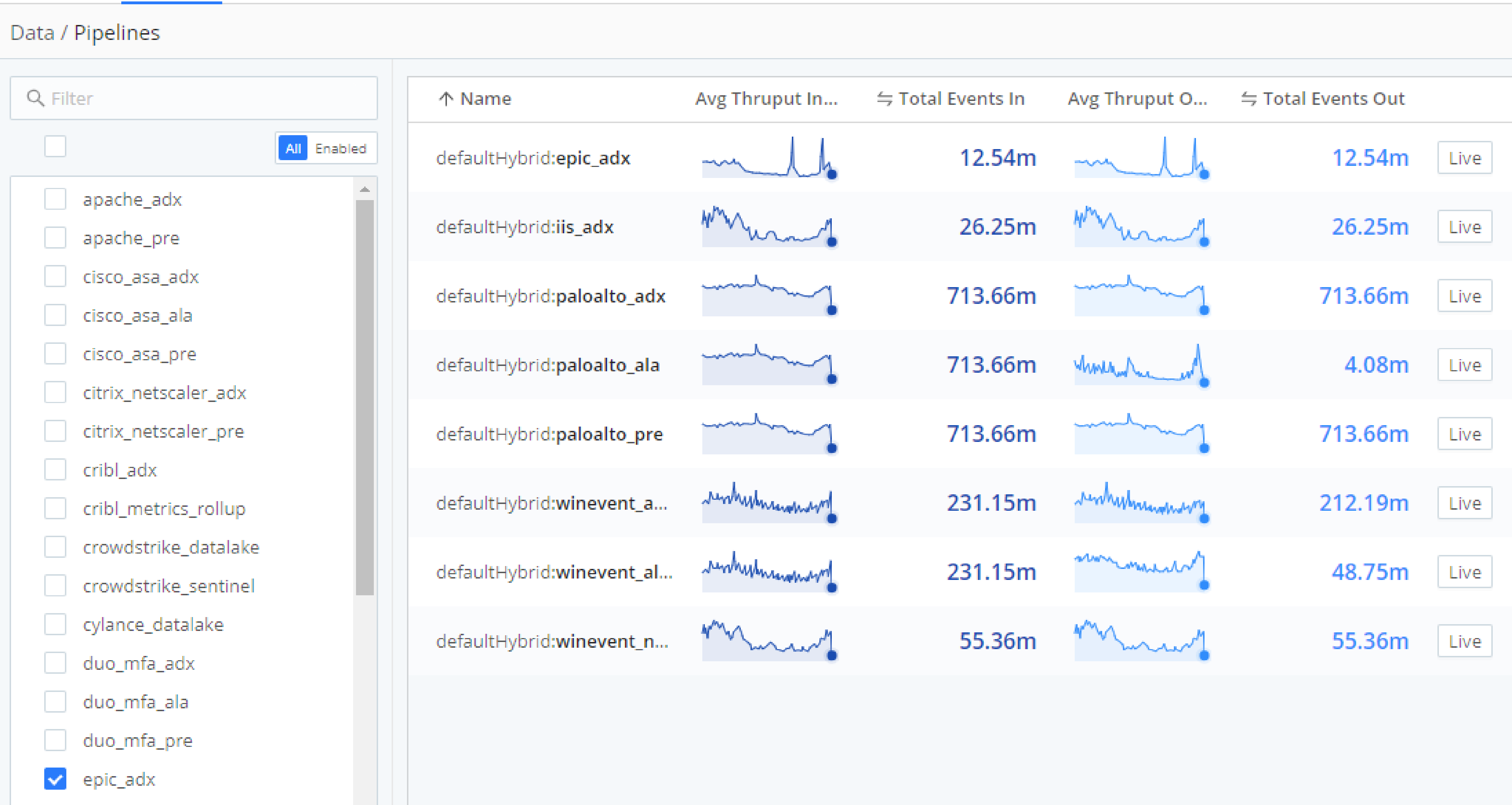
Events In and Out

THRUPUT IN (AVG)	EVENTS IN	THRUPUT OUT (AVG)	EVENTS OUT
25.44kEPS	2.20b	25.80kEPS	2.23b

Bytes In and Out

THRUPUT IN (AVG)	BYTES IN	THRUPUT OUT (AVG)	BYTES OUT
18.46MBPS	1.52TB	26.61MBPS	2.20TB

Cribl (free) Observability



Data / Pipelines

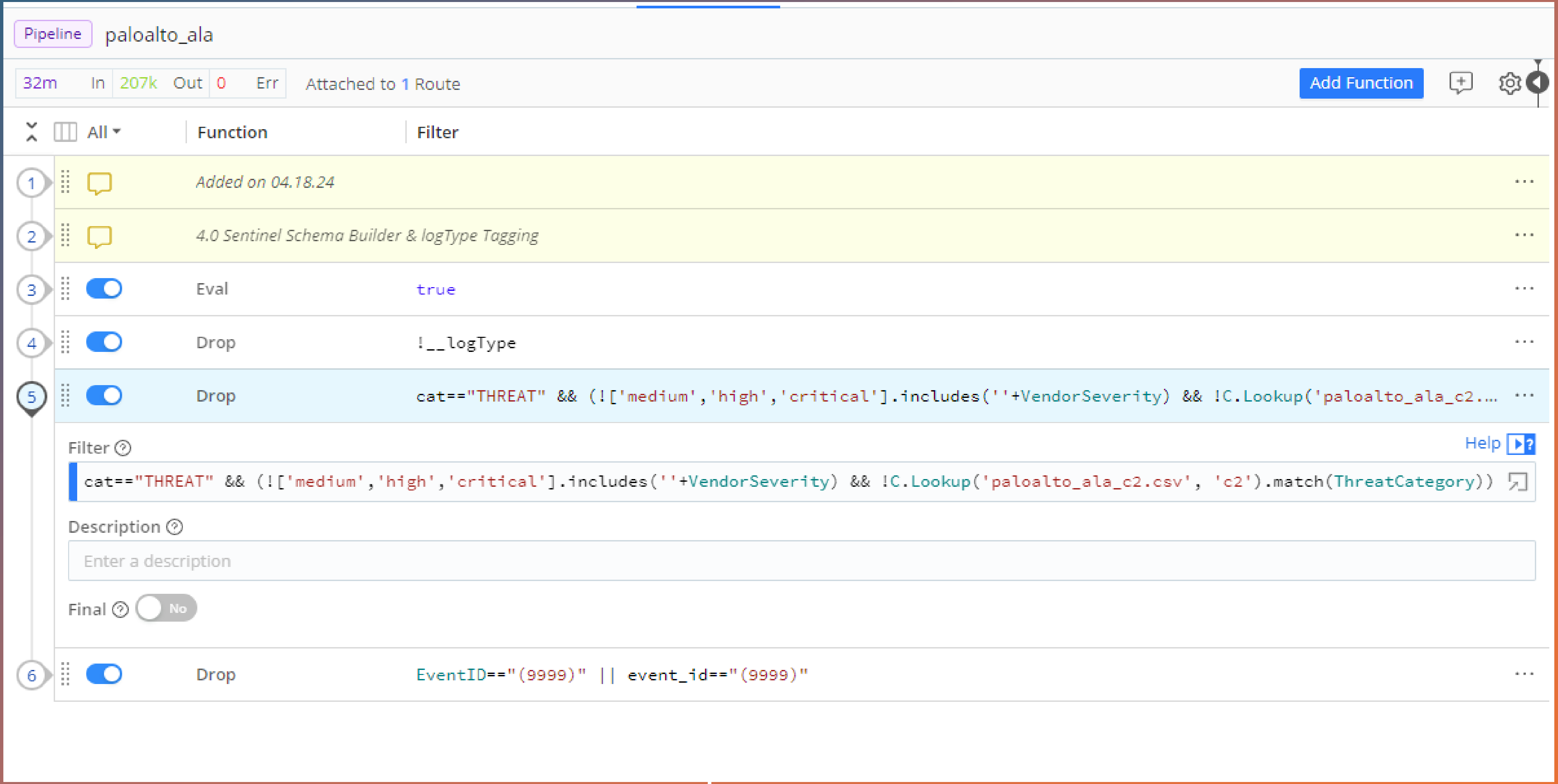
Filter

☐ All Enabled

- ☐ apache_adx
- ☐ apache_pre
- ☐ cisco_asa_adx
- ☐ cisco_asa_ala
- ☐ cisco_asa_pre
- ☐ citrix_netscaler_adx
- ☐ citrix_netscaler_pre
- ☐ cribl_adx
- ☐ cribl_metrics_rollup
- ☐ crowdstrike_datalake
- ☐ crowdstrike_sentinel
- ☐ cylance_datalake
- ☐ duo_mfa_adx
- ☐ duo_mfa_ala
- ☐ duo_mfa_pre
- ☑ epic_adx

↑ Name	Avg Thruput In...	⇆ Total Events In	Avg Thruput O...	⇆ Total Events Out	
defaultHybrid:**epic_adx**		12.54m		12.54m	Live
defaultHybrid:**iis_adx**		26.25m		26.25m	Live
defaultHybrid:**paloalto_adx**		713.66m		713.66m	Live
defaultHybrid:**paloalto_ala**		713.66m		4.08m	Live
defaultHybrid:**paloalto_pre**		713.66m		713.66m	Live
defaultHybrid:**winevent_a...**		231.15m		212.19m	Live
defaultHybrid:**winevent_al...**		231.15m		48.75m	Live
defaultHybrid:**winevent_n...**		55.36m		55.36m	Live

Cribl (free) Pipeline Management



Pipeline paloalto_ala

32m In 207k Out 0 Err

Attached to 1 Route

	Function	Filter
1	*Added on 04.18.24*	
2	*4.0 Sentinel Schema Builder & logType Tagging*	
3	Eval	`true`
4	Drop	`!__logType`
5	Drop	`cat=="THREAT" && (!['medium','high','critical'].includes(''+VendorSeverity) && !C.Lookup('paloalto_ala_c2.…`
6	Drop	`EventID=="(9999)" || event_id=="(9999)"`

Filter

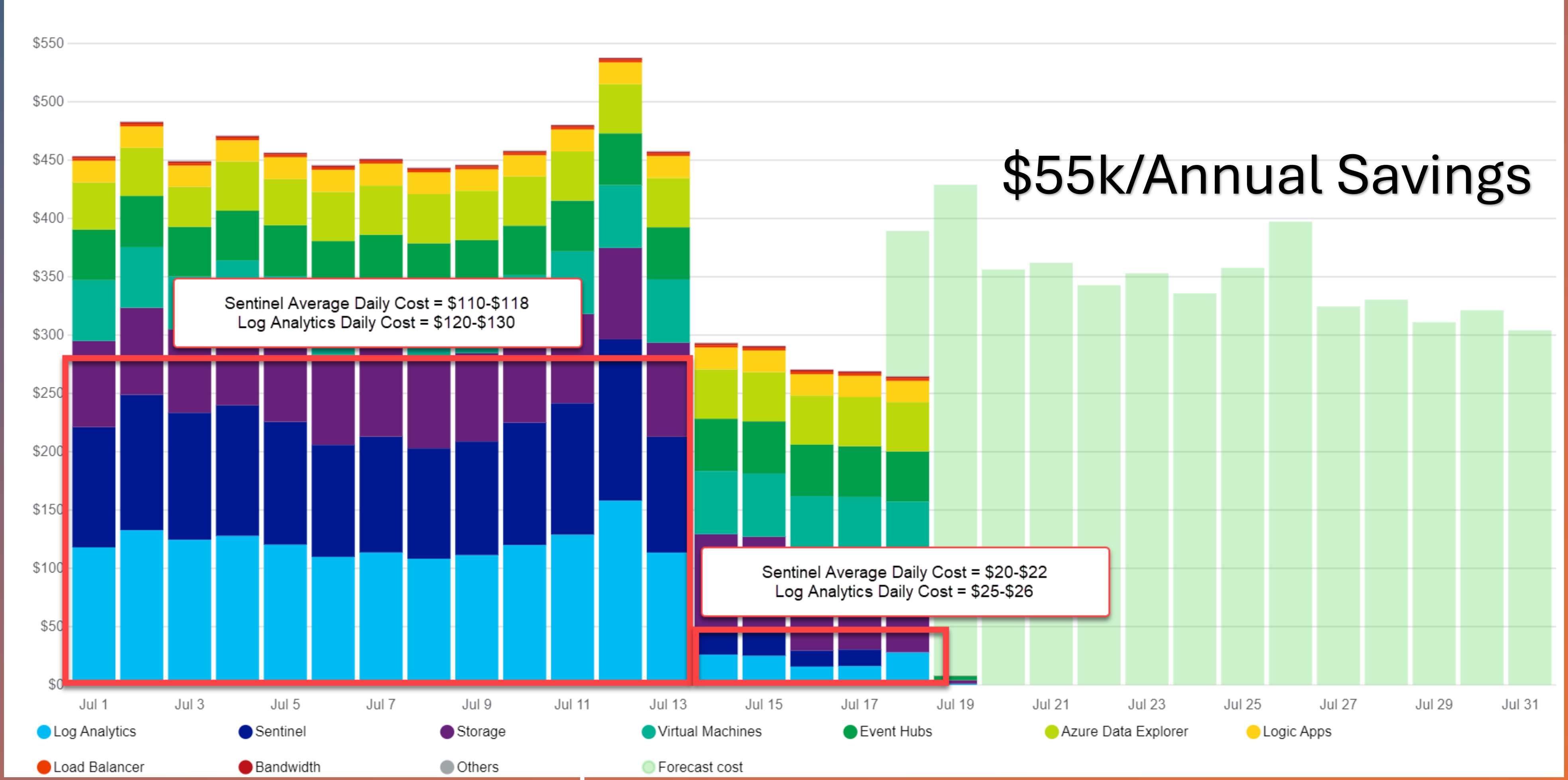
```
cat=="THREAT" && (!['medium','high','critical'].includes(''+VendorSeverity) && !C.Lookup('paloalto_ala_c2.csv', 'c2').match(ThreatCategory))
```

Description

Enter a description

Final No

Cribl (free) Pipeline Management



$55k/Annual Savings

Sentinel Average Daily Cost = $110-$118
Log Analytics Daily Cost = $120-$130

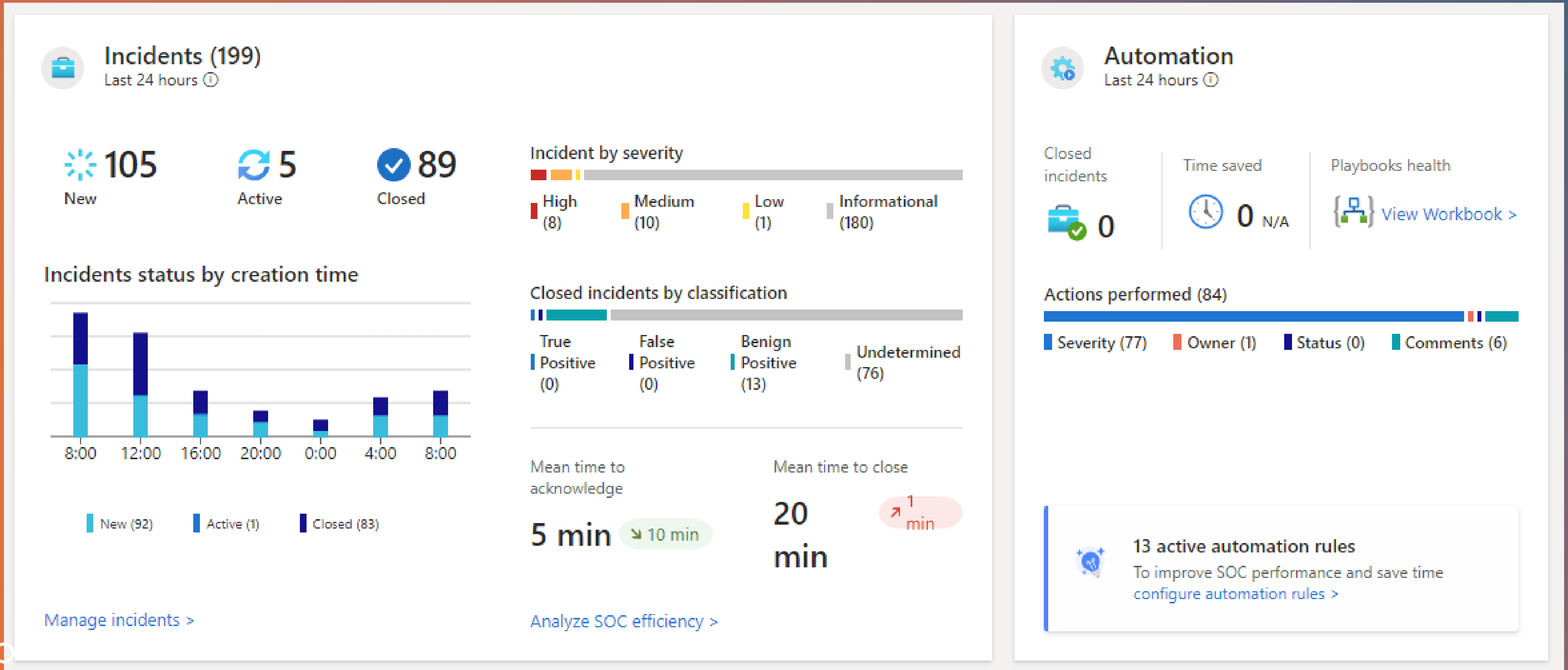
Sentinel Average Daily Cost = $20-$22
Log Analytics Daily Cost = $25-$26

DCSync Detections require event ID 4662. With Cribl, we can further filter on specific Object GUIDs while still maintaining detection fidelity

Understanding Data Lakes

A look at Microsoft Azure ADX but could also work with AWS or similar cloud solutions.

ADX - Datalake Demo / Screenshots



Incidents (199)
Last 24 hours ⓘ

105 New

5 Active

89 Closed

Incident by severity

High (8)
Medium (10)
Low (1)
Informational (180)

Incidents status by creation time

Closed incidents by classification

True Positive (0)
False Positive (0)
Benign Positive (13)
Undetermined (76)

Mean time to acknowledge
5 min ↘ 10 min

Mean time to close
20 min ↗ 1 min

Manage incidents >

Analyze SOC efficiency >

Automation
Last 24 hours ⓘ

Closed incidents
0

Time saved
0 N/A

Playbooks health
View Workbook >

Actions performed (84)

Severity (77)
Owner (1)
Status (0)
Comments (6)

13 active automation rules
To improve SOC performance and save time
configure automation rules >

ADX - Datalake Demo / Screenshots

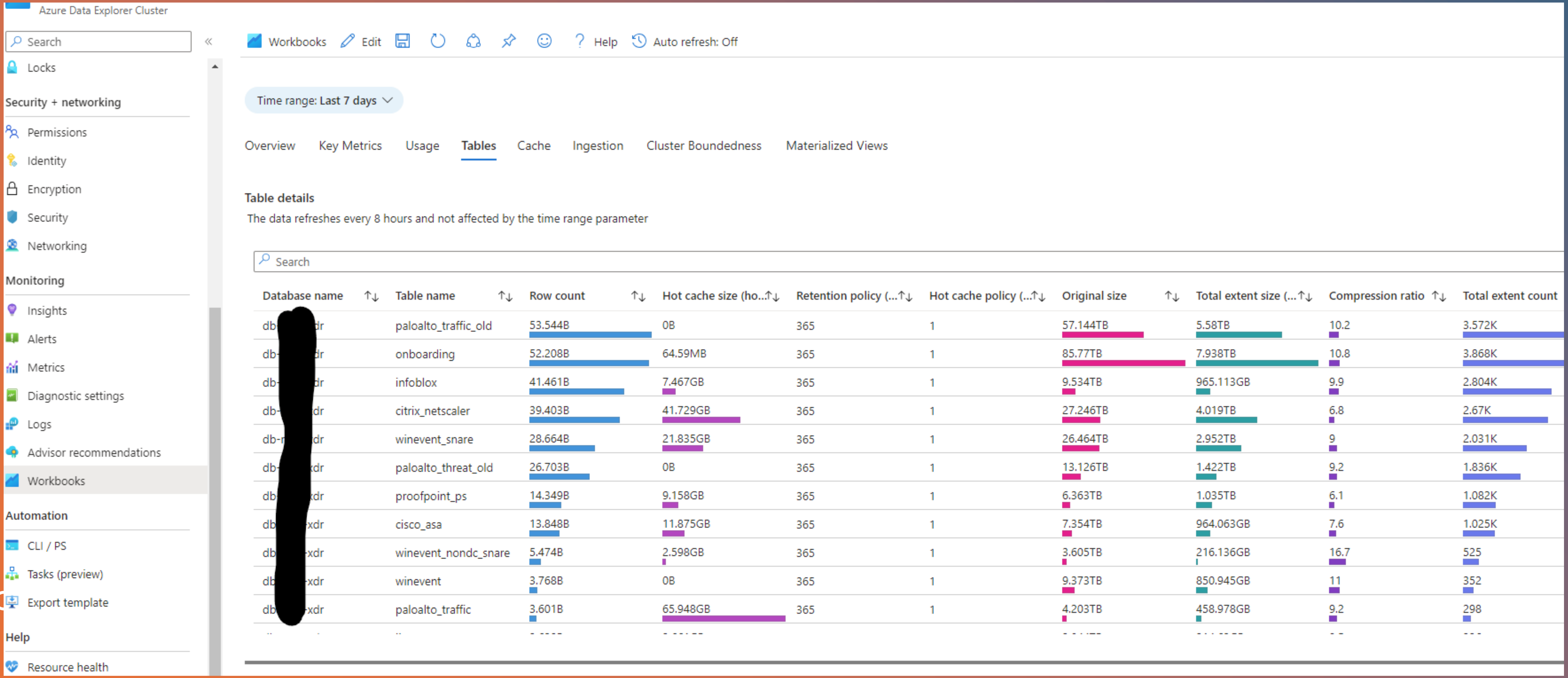
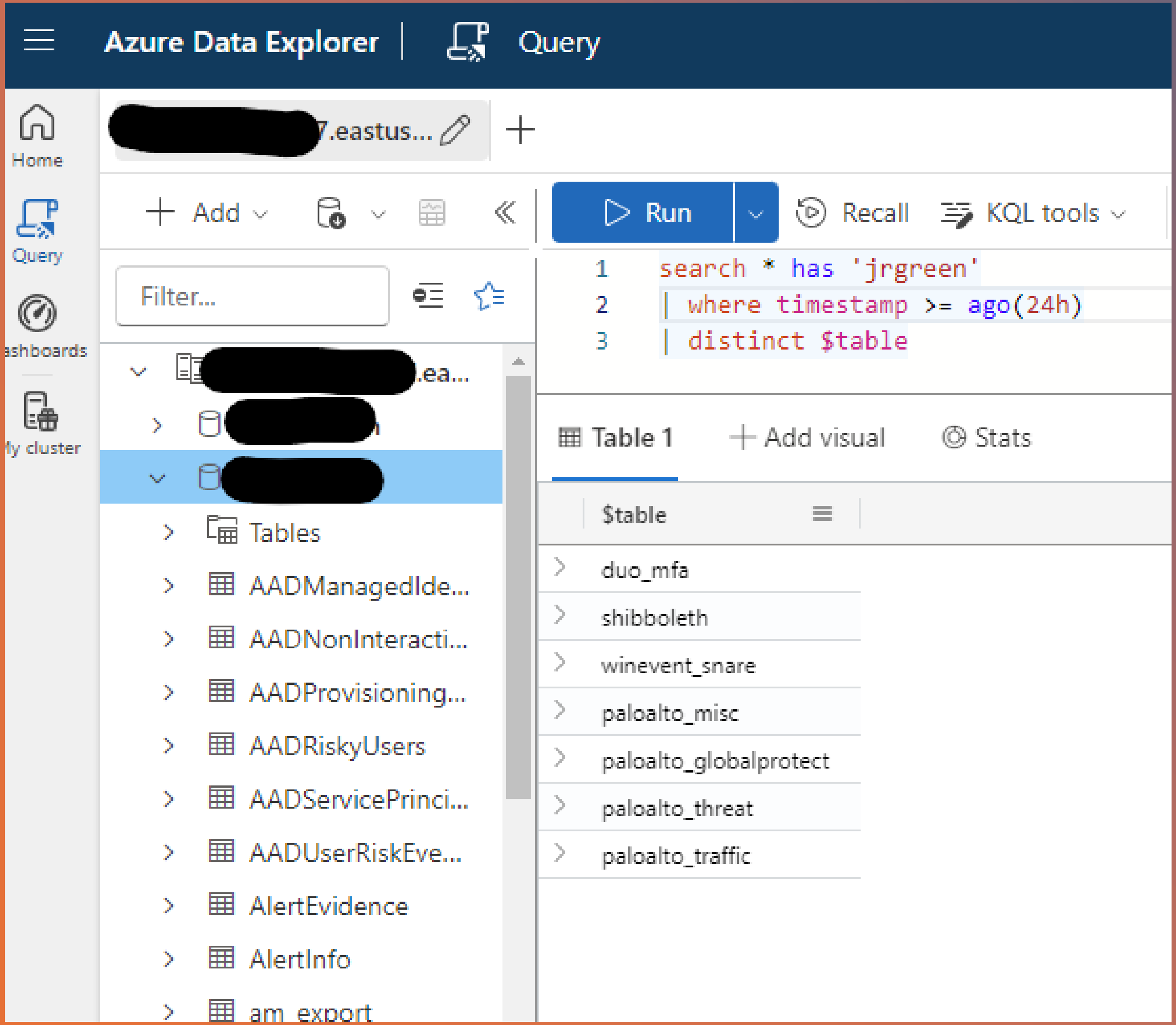


Table details

The data refreshes every 8 hours and not affected by the time range parameter

Database name ↑↓	Table name ↑↓	Row count ↑↓	Hot cache size (ho...↑↓	Retention policy (...↑↓	Hot cache policy (...↑↓	Original size ↑↓	Total extent size (...↑↓	Compression ratio ↑↓	Total extent count
db[redacted]dr	paloalto_traffic_old	53.544B	0B	365	1	57.144TB	5.58TB	10.2	3.572K
db-[redacted]dr	onboarding	52.208B	64.59MB	365	1	85.77TB	7.938TB	10.8	3.868K
db-[redacted]dr	infoblox	41.461B	7.467GB	365	1	9.534TB	965.113GB	9.9	2.804K
db-[redacted]dr	citrix_netscaler	39.403B	41.729GB	365	1	27.246TB	4.019TB	6.8	2.67K
db-[redacted]dr	winevent_snare	28.664B	21.835GB	365	1	26.464TB	2.952TB	9	2.031K
db-[redacted]xdr	paloalto_threat_old	26.703B	0B	365	1	13.126TB	1.422TB	9.2	1.836K
db[redacted]xdr	proofpoint_ps	14.349B	9.158GB	365	1	6.363TB	1.035TB	6.1	1.082K
db[redacted]-xdr	cisco_asa	13.848B	11.875GB	365	1	7.354TB	964.063GB	7.6	1.025K
db[redacted]-xdr	winevent_nondc_snare	5.474B	2.598GB	365	1	3.605TB	216.136GB	16.7	525
db[redacted]-xdr	winevent	3.768B	0B	365	1	9.373TB	850.945GB	11	352
db[redacted]-xdr	paloalto_traffic	3.601B	65.948GB	365	1	4.203TB	458.978GB	9.2	298

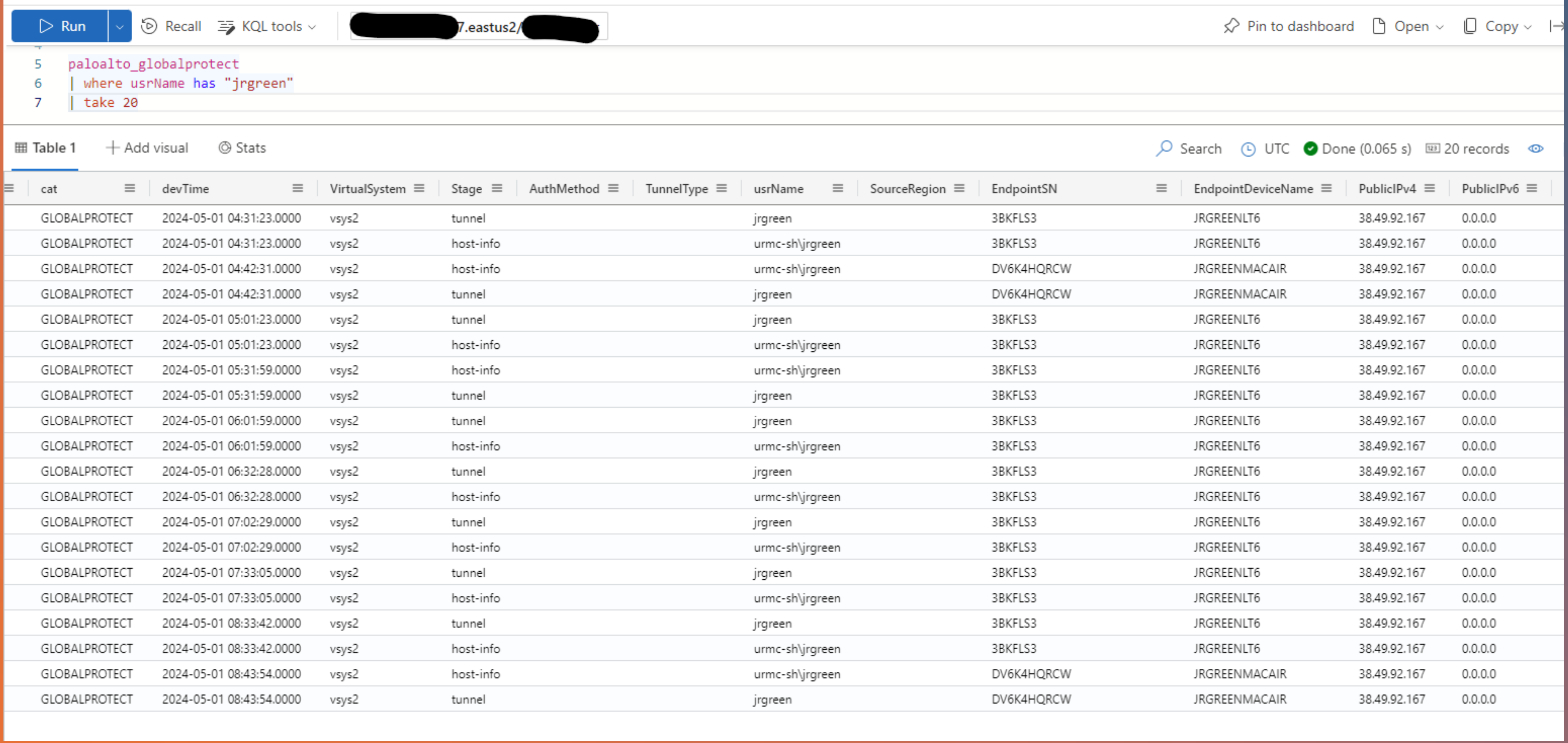
ADX - Datalake Demo / Screenshots



```
search * has 'jrgreen'
| where timestamp >= ago(24h)
| distinct $table
```

$table
duo_mfa
shibboleth
winevent_snare
paloalto_misc
paloalto_globalprotect
paloalto_threat
paloalto_traffic

ADX - Datalake Demo / Screenshots



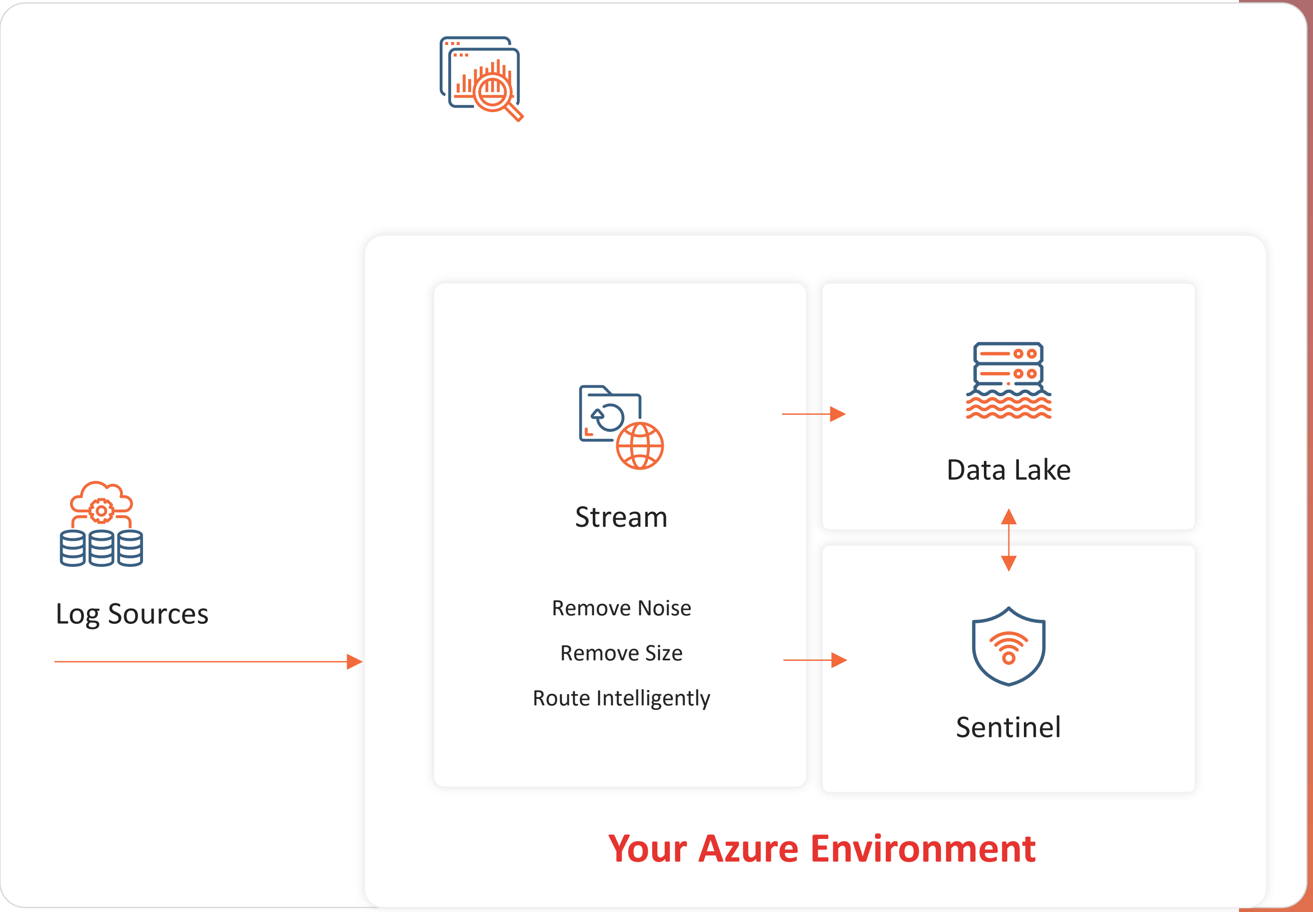
```
paloalto_globalprotect
| where usrName has "jrgreen"
| take 20
```

cat	devTime	VirtualSystem	Stage	AuthMethod	TunnelType	usrName	SourceRegion	EndpointSN	EndpointDeviceName	PublicIPv4	PublicIPv6
GLOBALPROTECT	2024-05-01 04:31:23.0000	vsys2	tunnel			jrgreen		3BKFLS3	JRGREENLT6	38.49.92.167	0.0.0.0
GLOBALPROTECT	2024-05-01 04:31:23.0000	vsys2	host-info			urmc-sh\jrgreen		3BKFLS3	JRGREENLT6	38.49.92.167	0.0.0.0
GLOBALPROTECT	2024-05-01 04:42:31.0000	vsys2	host-info			urmc-sh\jrgreen		DV6K4HQRCW	JRGREENMACAIR	38.49.92.167	0.0.0.0
GLOBALPROTECT	2024-05-01 04:42:31.0000	vsys2	tunnel			jrgreen		DV6K4HQRCW	JRGREENMACAIR	38.49.92.167	0.0.0.0
GLOBALPROTECT	2024-05-01 05:01:23.0000	vsys2	tunnel			jrgreen		3BKFLS3	JRGREENLT6	38.49.92.167	0.0.0.0
GLOBALPROTECT	2024-05-01 05:01:23.0000	vsys2	host-info			urmc-sh\jrgreen		3BKFLS3	JRGREENLT6	38.49.92.167	0.0.0.0
GLOBALPROTECT	2024-05-01 05:31:59.0000	vsys2	host-info			urmc-sh\jrgreen		3BKFLS3	JRGREENLT6	38.49.92.167	0.0.0.0
GLOBALPROTECT	2024-05-01 05:31:59.0000	vsys2	tunnel			jrgreen		3BKFLS3	JRGREENLT6	38.49.92.167	0.0.0.0
GLOBALPROTECT	2024-05-01 06:01:59.0000	vsys2	tunnel			jrgreen		3BKFLS3	JRGREENLT6	38.49.92.167	0.0.0.0
GLOBALPROTECT	2024-05-01 06:01:59.0000	vsys2	host-info			urmc-sh\jrgreen		3BKFLS3	JRGREENLT6	38.49.92.167	0.0.0.0
GLOBALPROTECT	2024-05-01 06:32:28.0000	vsys2	tunnel			jrgreen		3BKFLS3	JRGREENLT6	38.49.92.167	0.0.0.0
GLOBALPROTECT	2024-05-01 06:32:28.0000	vsys2	host-info			urmc-sh\jrgreen		3BKFLS3	JRGREENLT6	38.49.92.167	0.0.0.0
GLOBALPROTECT	2024-05-01 07:02:29.0000	vsys2	tunnel			jrgreen		3BKFLS3	JRGREENLT6	38.49.92.167	0.0.0.0
GLOBALPROTECT	2024-05-01 07:02:29.0000	vsys2	host-info			urmc-sh\jrgreen		3BKFLS3	JRGREENLT6	38.49.92.167	0.0.0.0
GLOBALPROTECT	2024-05-01 07:33:05.0000	vsys2	tunnel			jrgreen		3BKFLS3	JRGREENLT6	38.49.92.167	0.0.0.0
GLOBALPROTECT	2024-05-01 07:33:05.0000	vsys2	host-info			urmc-sh\jrgreen		3BKFLS3	JRGREENLT6	38.49.92.167	0.0.0.0
GLOBALPROTECT	2024-05-01 08:33:42.0000	vsys2	tunnel			jrgreen		3BKFLS3	JRGREENLT6	38.49.92.167	0.0.0.0
GLOBALPROTECT	2024-05-01 08:33:42.0000	vsys2	host-info			urmc-sh\jrgreen		3BKFLS3	JRGREENLT6	38.49.92.167	0.0.0.0
GLOBALPROTECT	2024-05-01 08:43:54.0000	vsys2	host-info			urmc-sh\jrgreen		DV6K4HQRCW	JRGREENMACAIR	38.49.92.167	0.0.0.0
GLOBALPROTECT	2024-05-01 08:43:54.0000	vsys2	tunnel			jrgreen		DV6K4HQRCW	JRGREENMACAIR	38.49.92.167	0.0.0.0

Brining it into SIEM

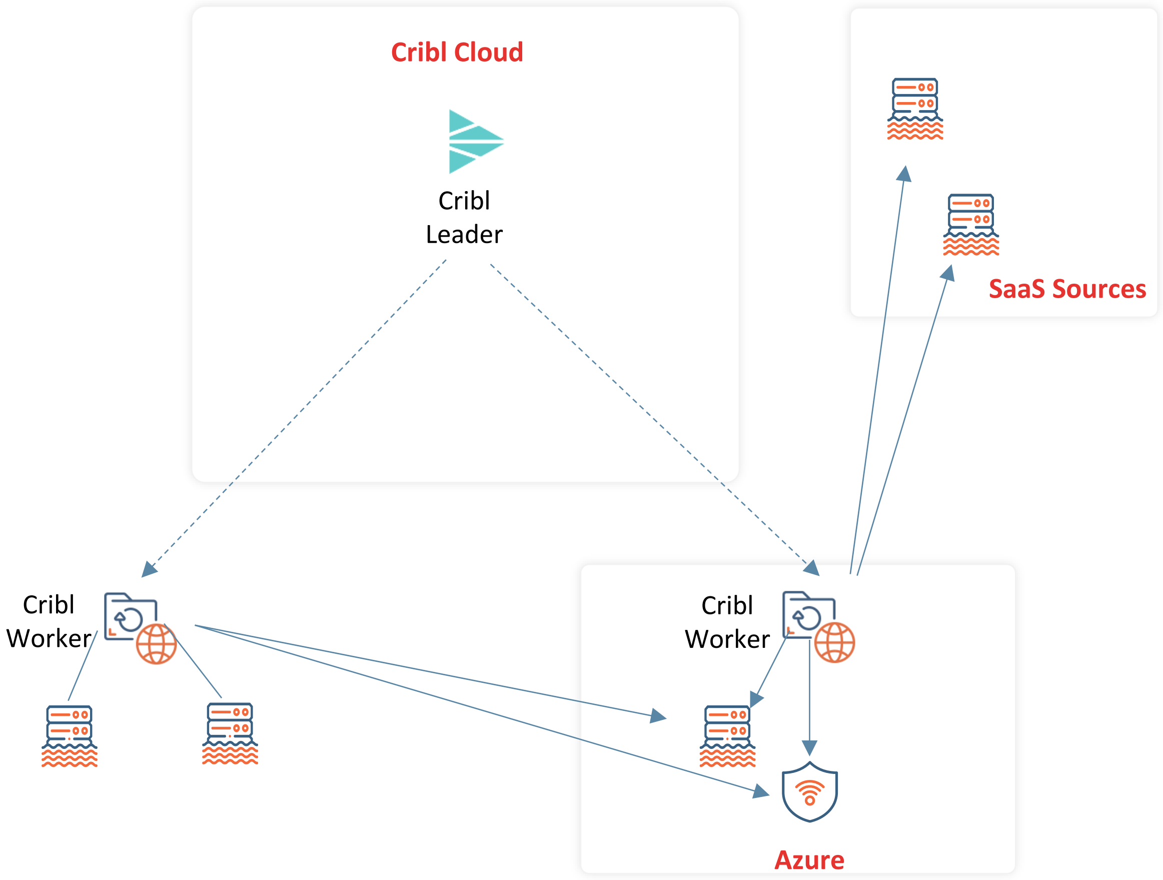
A look at Microsoft Sentinel when ADX is utilized

Brining it all together:

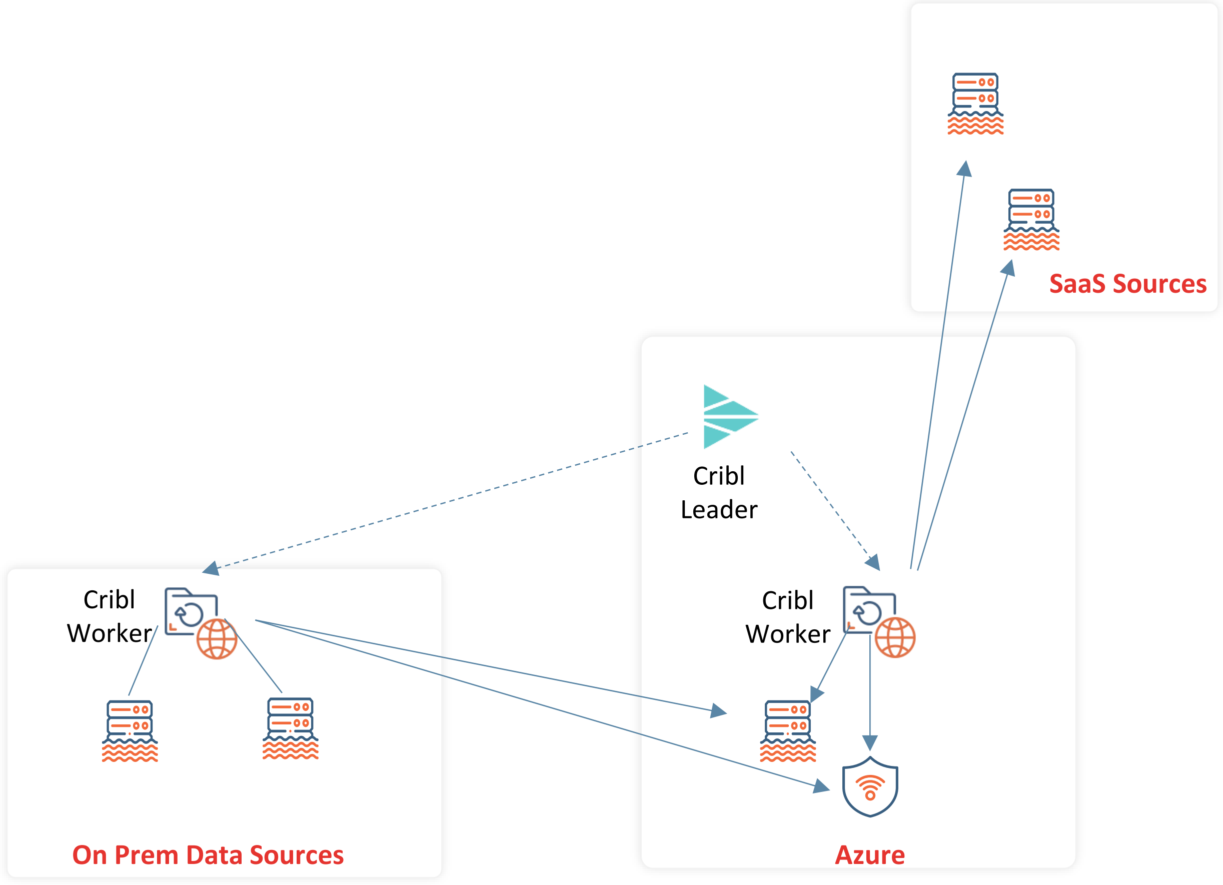


Connecting Logs using Cribl

Cloud:



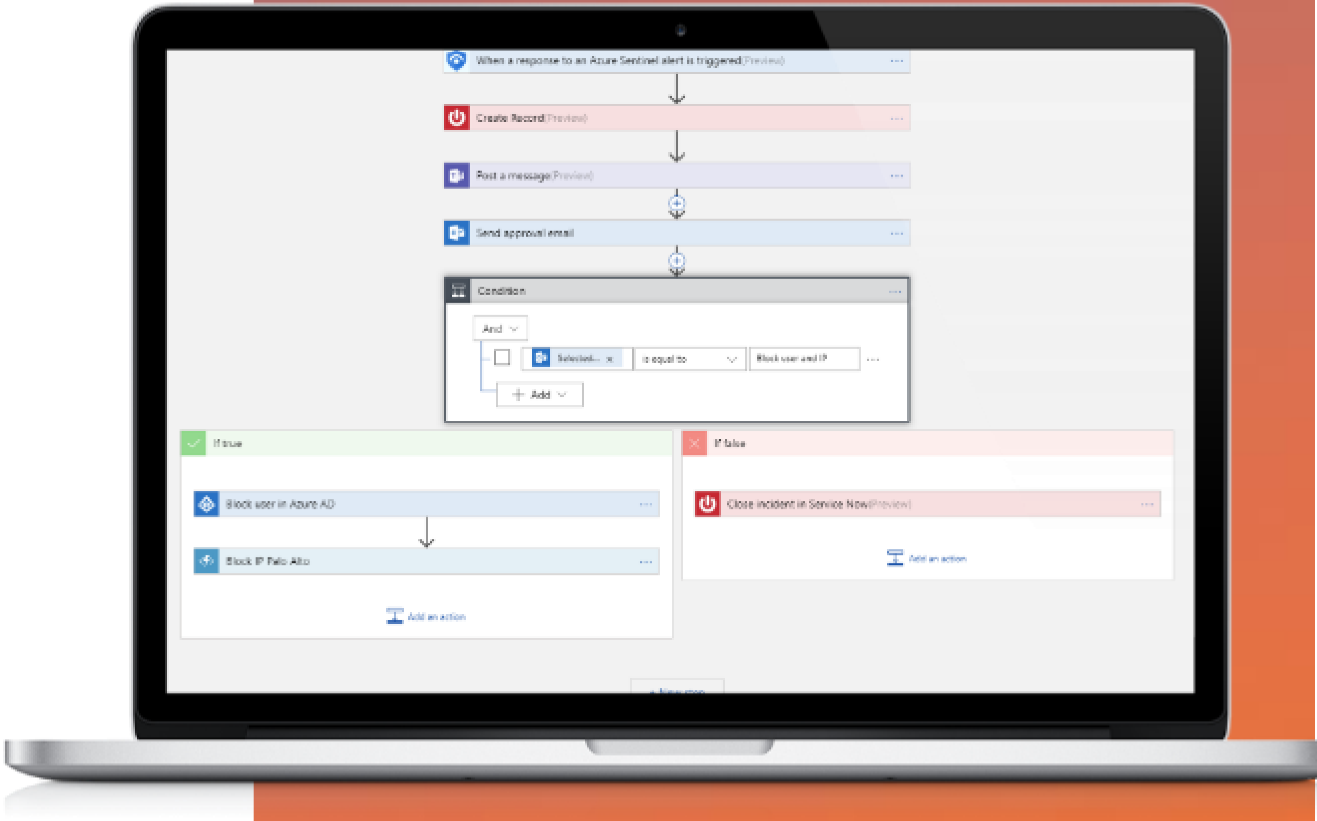
On-Prem:



Security Automation as a First-Class Feature of your SIEM

FEATURES

- Sentinel SOAR capabilities allow creation of custom and limitless response capabilities, with little to no coding
- Massive library of SOAR integrations available out of the box and the open-source community.
- No additional investment or significant costs to utilize SOAR, unlike purchasing a 3rd party bolt-on tool



Achieving cost efficiencies while harnessing all that a SIEM can do.

Any questions?