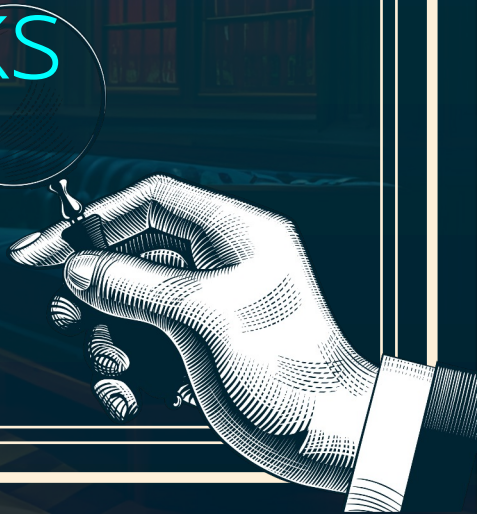
CRIBLCON 24

POWERED BY Cribl

Tune your Data Engine: How Packs and Forks Supercharge Value

JOHN LIM
Lead Systems Engineer, Cox Automotive



JOHN LIM

- Lead Systems Engineer, Cox Automotive
- 10 years of data pipelining, visualization, and reporting experience w/ Cribl Stream and Splunk
- Passionate about Technology Process and Operationalization

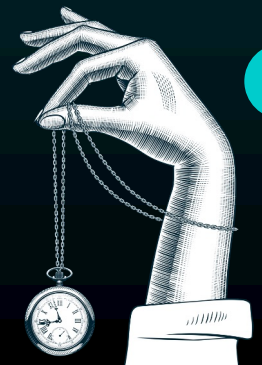
Agenda

1. **Know your company's Adoption Challenges**
   Understand the problem and plan to drive success.

2. **Identify Targets**
   Create a priority list.

3. **Establish Process and Operationalize**
   Time-boxed processes must be iterative.

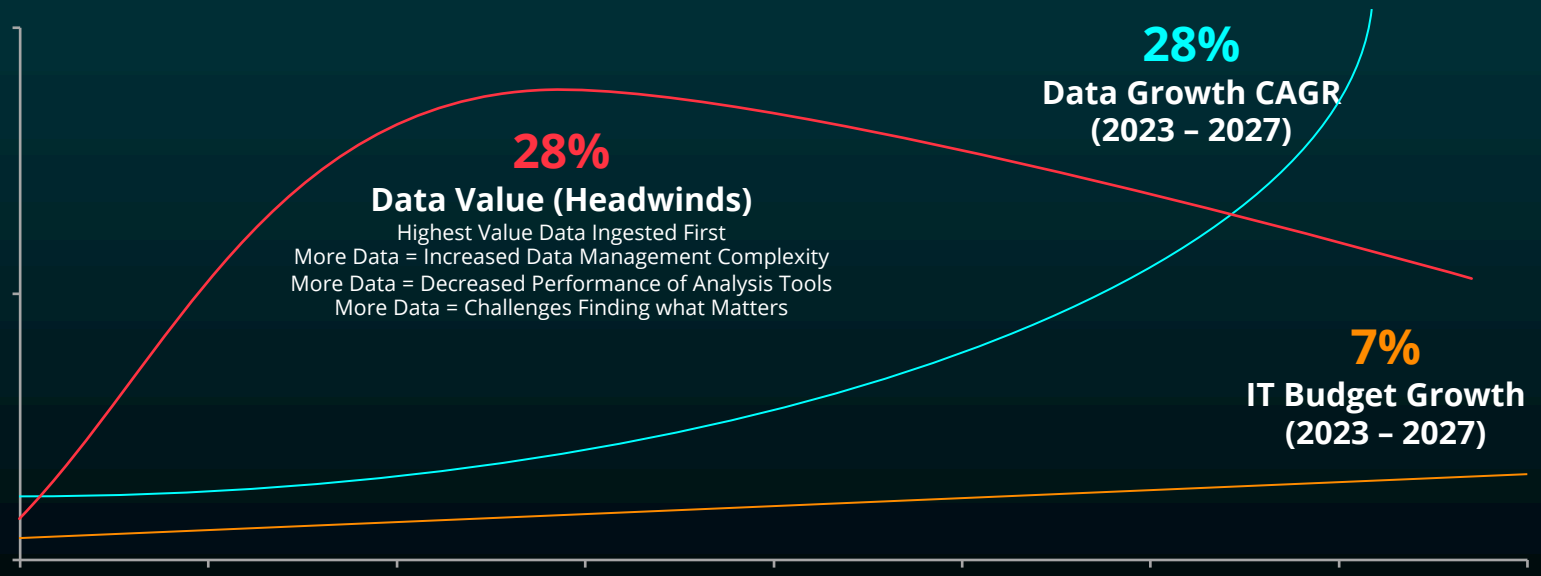
4. **Iterate and Document**
   Operations cannot be successful without robust documentation.



Why is High Velocity Adoption Needed?

"Get more value from the same budget."

IT AND SECURITY FACE TIDAL WAVE OF DIVERSE DATA



Adoption Challenges

Success relies on identifying points of friction.

Legacy State

- Existing methodologies of data ingestion, processing and output may be lacking.
- How will you handle the data stream migration to minimize impact?

Reliance on Current Data

- Endpoints like Splunk may already be parsing subpar data (regex, etc.)
- How will you handle downstream knowledge objects?

Data Attribution

- Are your datasets properly tagged prior to ingestion?
- How do you know who owns each dataset that will be migrated to Cribl Stream?

"Problems worthy of attack prove their worth by fighting back."

PIET HEIN,

Danish Polymath

Dataset Targeting

So much data! What should I target first?

TOE-IN

~10%

Enhance custom data for specific use cases. Small volume and app generated.

PROBLEMATIC

~20%

Fix specific problems with parsing – KV pairs, timestamps, line breaks, aggregation.

HIGH IMPACT

~70%

Reduce ingestion for large datasets, typically in the Network space (syslog).

Create a priority list and schedule!

Establish the Linear Process

An effective process has “time-boxed” steps.

TARGET

Pick a target dataset and evaluate existing methodologies.

PACK

Identify the Cribl Stream pack to use. If it doesn’t exist, find sample data and prepare a custom pipeline.

FORK

Fork the data to two unique destinations for evaluation and minimize downstream impact.

EVALUATE

Work with the customer to review data format changes and clone/modify existing knowledge objects.

CUT-OVER

Switch to the new and improved data stream. Turn off legacy data stream.

Target

Cisco Syslog – FTD (Firewall Threat Defense) and ASA (Adaptive Security Appliance)

Potentially High Impact Syslog

- Reduce ingestion.
- Only retain required fields.
- Suppress noisy messages.

Pipeline diagnostics - cisco_asa_cleanup

Statistics | Pipeline Profile | Advanced CPU Profile

	_raw Length	Full Event Length
IN	4.63KB	20.30KB
OUT	153.00B	574.00B
DIFF	↓ -96.77%	↓ -97.24%

Pipeline diagnostics - cisco_ftd_cleanup

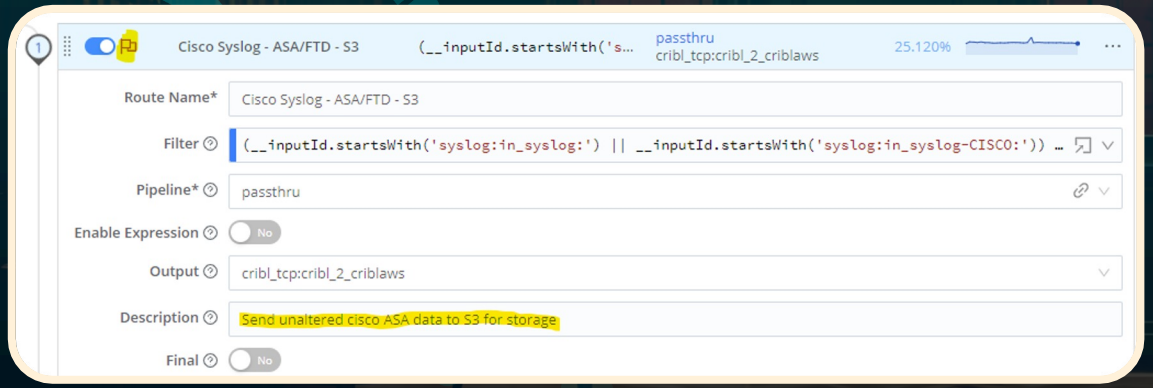
Statistics | Pipeline Profile | Advanced CPU Profile

	_raw Length	Full Event Length
IN	73.20KB	288.85KB
OUT	60.01KB	225.14KB
DIFF	↓ -18.01%	↓ -22.05%

Fork

Send original, unaltered Cisco Syslog to a different destination

Use the Pipeline and Route Filters to fork data. This keeps the original dataset prior to Pack application intact.



Options:

- Send to original destination during evaluation.
- Send to S3 for long-term storage after Pack is in place.

Evaluate

Review the Cisco FTD / ASA Pack with Stakeholders

Stakeholder Approval

- Use the fork to present new version of the data
- Obtain feedback and make any needed adjustments to pipeline / pack.

Sample Data | Simple Preview

Sample data file: FTD_20240331-104841.log → Pipeline: cisco_ftd_cleanup | Run

IN | OUT | Select Fields (17 of 17)

#	Event
1 2024-03-31 10:48:41.606 -04:00	b _no_matches: false a _raw: %FTD-4-419002: Duplicate TCP SYN from [redacted] to [redacted] with different initial sequence number # _time: 1711896521.606 # cribl_bytes_in: 160 [] cribl_pipe: a generic_syslog-v1 a prep_for_splunk a cisco_ftd_cleanup a dest_ip: [redacted] # ~~facility: 20~~ a ~~facilityName: local4~~ a ftd_code: 419002 a host: raa_ftd_active a index: cai_network_platform a ~~message: Mar 31 2024 14:48:41: %FTD-4-419002: Duplicate TCP SYN from [redacted] to [redacted] with different initial sequence number~~ # ~~severity: 4~~ a ~~severityName: warning~~ a source: cribl:man_dd:syslog:in_syslog:udp a sourcetype: cisco:asa

Cut-Over

Send Cisco FTD and ASA data thru Pack

3 | Cisco Syslog - FTD - PACK | (__inputId.startsWith('s... | PACK cribl-cisco-ftd-cleanup (... splunk_lb:prod_splunkcloud | 24.954%

Route Name* Cisco Syslog - FTD - PACK

Filter (__inputId.startsWith('syslog:in_syslog:') || __inputId.startsWith('syslog:in_syslog-CISCO:')) ...

Pipeline* PACK cribl-cisco-ftd-cleanup (Cisco FTD)

Enable Expression No

Output splunk_lb:prod_splunkcloud

Description Process Cisco FTD data through pack

6 | Cisco Syslog - ASA - PACK | (__inputId.startsWith('s... | PACK cribl-cisco-asa-cleanup ... splunk_lb:prod_splunkcloud | 0.141%

Route Name* Cisco Syslog - ASA - PACK

Filter (__inputId.startsWith('syslog:in_syslog:') || __inputId.startsWith('syslog:in_syslog-CISCO:')) ...

Pipeline* PACK cribl-cisco-asa-cleanup (Cisco ASA)

Enable Expression No

Output splunk_lb:prod_splunkcloud

Description Process Cisco ASA data through pack

Final Yes

The Data Engine

Build your new migration process into a continuous lifecycle.

1. Pick your Target and Identify Correct Pack

Pick the target dataset and evaluate existing onboarding methodologies. Identify the Cribl Stream pack to use. If a pack doesn't exist, find sample data and prepare a custom pipeline.

2. Fork the Data and Evaluate New Pipeline

Fork the data to two unique destinations for evaluation and minimize downstream impact. Work with the customer to review data format changes and clone/modify existing knowledge objects (if any).

3. Cut Over

Switch to the new and improved Cribl Stream data pipeline by turning off the legacy data stream.

4. Re-valuate Your Targets

Based on your inventory of data and the three types (Toe-In, Problematic, and High Impact), pick the next target for Stream onboarding.

Iterate and Document

Each new migration must be documented and reviewed.

Iterate for Scale

Every dataset migration can be unique.

- A well-written process can be handed off to multiple engineers to execute at scale.
  - Engineers can typically handle one end-to-end migration per 2-week sprint.

Document Thoroughly

Both the process and information about the migration itself must be documented.

- Customers must be identified.
  - Every migrated dataset must have an owner, ideally baked into the data through attribution.

Establish Customer Cadence

Quarterly is recommended but can be limited to when a change in data pattern is detected.

- Data monitoring (e.g., volume drops)
  - Integrate new data stream into existing alert processes.

"Trust the process."

SAM HINKIE,

American Basketball Executive

Thank you!