



A DATA ENGINEER'S JOURNEY TO MODERNIZING WITH CRIBL

Terry Mulligan

Consultant, Discovered Intelligence

www.discoveredintelligence.ca

DATA ENGINEER ROLE



TERRY MULLIGAN
Consultant,
Discovered Intelligence

- ➔ **Working within a Fortune 100** Pharmaceutical and biotechnology company
- ➔ **Originally** the 'Splunk Data Engineer'
- ➔ **Work directly with** internal customers and stakeholders
- ➔ **Responsible for** defining data requirements and collection methodologies

DATA ENGINEERING BEFORE CRIBL

- ➔ Referred to as the “BC” era (dinosaurs were common)
- ➔ Very time-consuming process
- ➔ Data could not be made easier to work with
- ➔ No tools to manage or control data volumes
- ➔ If only 5% of data was valuable, had to live with 95% bloat
- ➔ Led to tough decisions based on licensing costs
- ➔ “Set it and forget it”

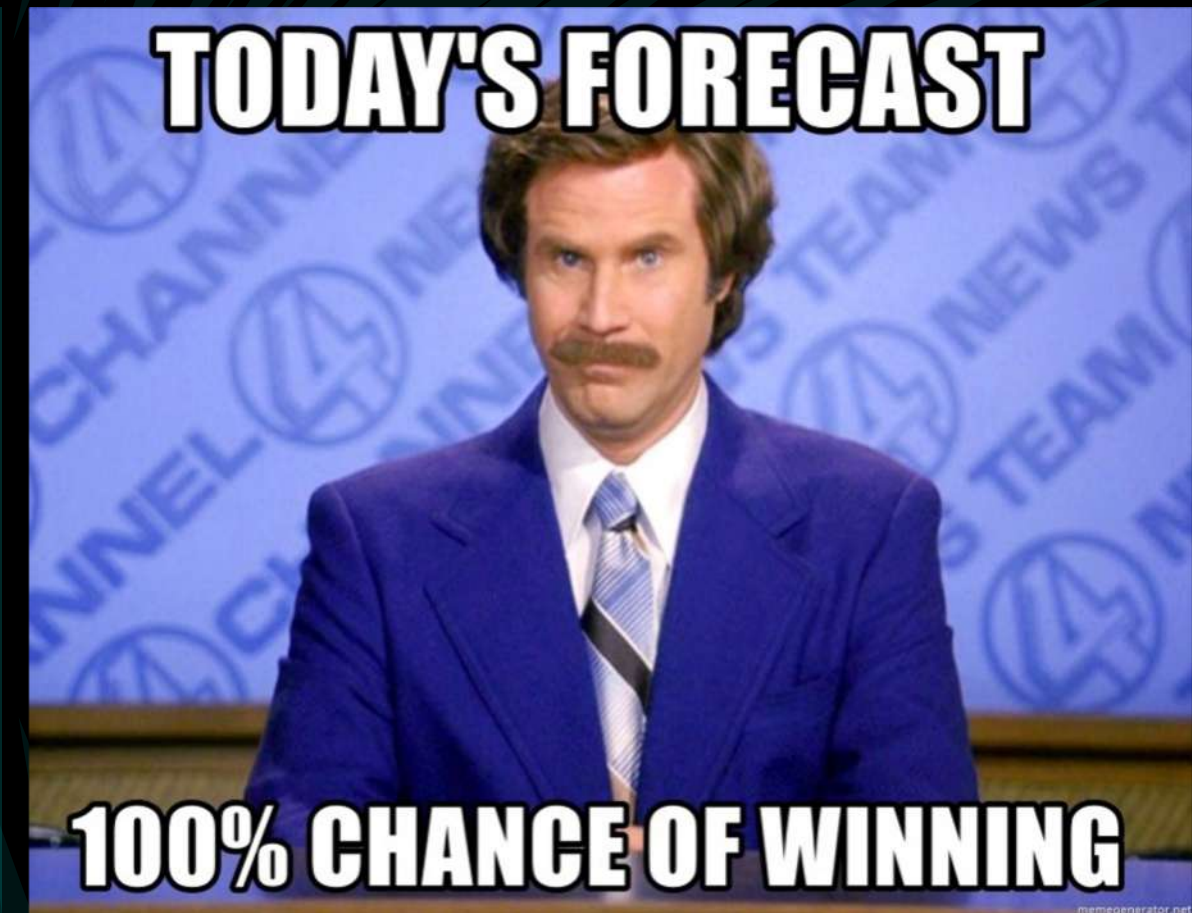


➔ ...But hold on, here comes **Cribl Stream!**

MODERNIZING DATA ENGINEERING



- ➔ Referred to as the “AC” era (After Cribl)
- ➔ Cribl has become a verb—**Cribl’ize** the data
- ➔ No longer an afterthought—Cribl is the standard ingestion point
- ➔ We are in the driver’s seat of data
 - Data can be made more usable
 - Control over data volumes to reduce licensing costs
 - Continuous optimization of data
 - Multi-destination data routing, no longer just Splunk
- ➔ Clears the data backlog
 - 0151 allows us to ‘yes’ to customers
 - Provides the ability to work with a wide array of internal customers
- ➔ The conversation is no longer “**What can we bring in?**” It is now “**Let’s get it in!**”





SUCCESSSES OF THE JOURNEY

Current Splunk license reduction of **3TB**

On average achieving **60%** reduction in data volumes




Using **masking** to meet GDPR compliance requirements

Streamlined upgrade process: **weeks to hours**

Consolidated 14 syslog servers into **4 Cribl nodes**

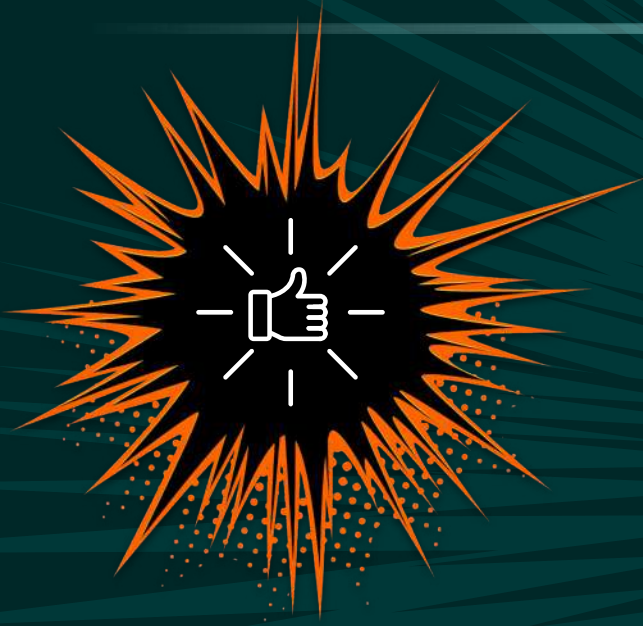
Reduced API integration times **from 2 weeks to 4 hours**

Multi-platform routing to Splunk, Chronicle, and S3 **simultaneously**

 And more to come!



BEST PRACTICES



BEST PRACTICES



Get leadership buy-in.



Clear, attainable goals for Cribl:

- Reduce license volume by X amount or consolidate hardware footprint.
- Allow Cribl to shine.



Everyone should visit Cribl University and plan their learning path.

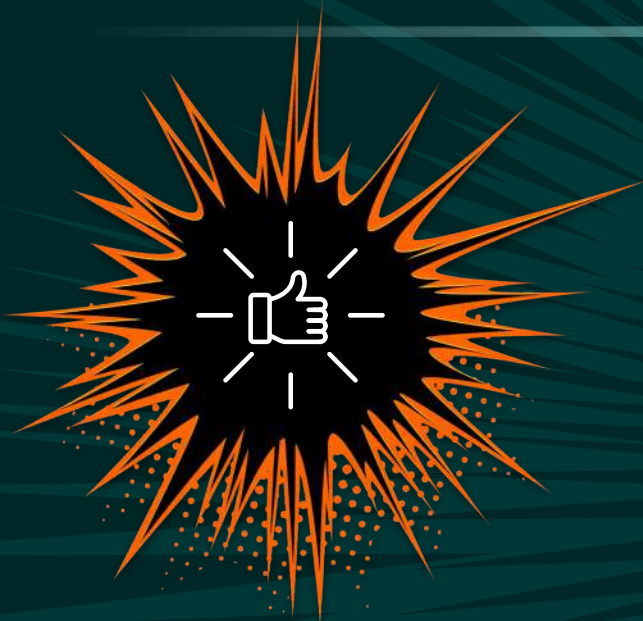


Plan your architecture:

- Single vs multiple worker groups (Push vs Pull)
- Cribl Cloud for cloud-to-cloud integrations
- DR vs HA
- Auto-scaling worker nodes



BEST PRACTICES



BEST PRACTICES



Standard naming conventions:

- Source: company_dataname_collectionmethod
- Route: company_dataname_source_destination
- Pipeline: company_pipeline_name_pre/pro/post_purpose



Worker Group naming conventions:

- Adopt a naming convention beyond the default group.
- Should be clear as to its purpose: 'Staging' 'Pull' 'Push'.



Staging environment:

- Critical for testing and building pipelines before moving into production.
- An environment for new people to learn and be QA'd before working in a production setting.
- Teachers can become students by evaluating other pipelines/packs.



TIPS & TRICKS



TIPS & TRICKS



Quick Reference Guide is your friend.

- Includes the basic information a rookie needs to get started.
- Will save you time.
- <https://cribl.io/resources/cribl-quick-reference-guide/>

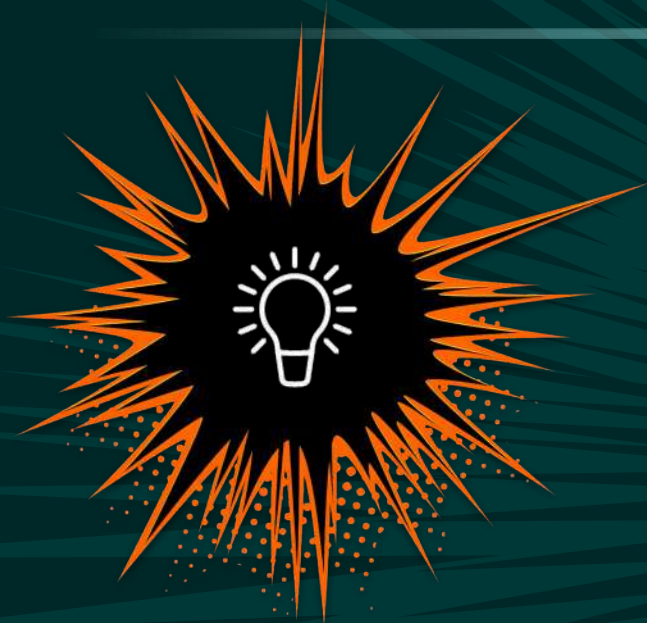


Develop a default set of data transformation rules:

- Remove null fields.
- Remove timestamps since they will be in `_time` in Splunk.
- Shorten field names and alias back in Splunk.
- Remove headers from syslog data sources.
- Flatten JSON events to make it simpler to work with—greatest thing since sliced bread!
- For Syslog sources, stuff severity, priority, and application into the Splunk source field.



TIPS & TRICKS



TIPS & TRICKS



Leverage Lookups to assist with event processing.

- Field renaming—one line instead of multiple renames.
- Applying time zones to hosts.



Develop guidelines for pipeline format/style. Ours includes:

- First function is a comment where we have a brief description and document changes.
- Second function, if required, is a dedicated EVAL function for index rewriting.
- Last function is a dedicated EVAL function that cleans up the event before ingestion.
- Descriptions field in functions are populated.
- Add comment functions if more detail is required.
- Group functions to make the pipeline more readable.



TIPS & TRICKS



TIPS & TRICKS

- ➔ Take advantage of the Pack Dispensary.
 - Powerful, free, and an excellent learning tool
- ➔ Learn the ins & outs of Cribl event breakers—Stick around for [Pipeline Lab at 3:00pm!](#)
 - Timestamps use capture groups.
 - Truncation behaves differently than Splunk.
- ➔ Cribl uses a different flavor of regex.
 - Cribl uses ECMAScript, and Splunk uses PCRE2.
 - They are similar, but there are differences.
- ➔ Cribl uses a different flavor of strtptime.
 - Very similar to Splunk but with a few differences.



LEARNING MOMENTS: LESSONS LEARNED



LEARNING MOMENTS

- ➔ Take the training.
- ➔ New data sources are easier to **Criblize** than existing ones.
- ➔ Make Cribl your standard ingestion point. Update your SOPs.
- ➔ Educate your users that Cribl is now the standard ingestion point.
- ➔ Increase your engagement with the customer:
 - Understand their needs so you can tailor the data for them.
 - Educate them on the power of Cribl.
- ➔ Naming standards are harder to implement after starting.



LEARNING MOMENTS:

THINGS I WISH I DID OR KNEW BEFORE STARTING



LEARNING MOMENTS

- ➔ Took advantage of the training opportunities.
- ➔ Reviewed the available resources.
- ➔ Read the documentation—especially the Quick Start Guide.
- ➔ Used a sandbox instead of using the production worker group.
- ➔ Joined Slack sooner.
- ➔ Embraced Packs sooner.



CRIBLCON!
2023