

# What is Stream?

---

Cribl Stream is an observability pipeline tool living between any data source and any destination. These destinations can be systems of analysis (Splunk, Elastic, etc.) or systems of retention (S3 buckets, Data Lakes, etc.). Stream routes, reduces, replays and redacts data sources in flight.

## Basic Elements

---

Stream uses Sources, Destinations, Routes and Pipelines to move data coming from its sources, processes through its pipelines (group of functions) and send the results to one or many destinations in the format that is best for the environment and its use cases.

This Labs has 3 parts:

**Part 1 – Introduction to Stream elements**

**Part 2 – Routes and Pipelines**

**Part 3 – Answering to Use Cases**

Access your Cribl Stream instance from the jump box provided using the Chrome web browser pointing to: [Cribl](#) Login as Admin with the password of Go2atc4labs!

## Part 1 – Introduction to Stream elements

### 1. Review Sources and Destinations

Sources LogStream can receive continuous data input from various Sources, including Splunk, HTTP, Elastic Beats, Kinesis, Kafka, TCP JSON, and many others. Sources can be of a Push or Pull type where Push sources can forward data into Cribl Stream as any other receiving system (Syslog, Splunk, etc.). Data from these Sources is normally sent to a set of LogStream Workers through a load balancer. Some Sources, such as Splunk forwarders, have native load-balancing capabilities, so you should point these directly at LogStream. Pull sources can interact with Stream via REST APIs or other methods available within Stream integrations. For sources not necessarily integrated with Stream Scripts may also be used to collect data from Pull sources.

Destinations Stream can send data to multiple destinations. These destinations may have native interactions as in sources or being sent at the protocol level such as Syslog, TCP, TCP/Json etc. There may be Streaming, No Streaming and Other Destinations. These destinations can have integrations (as in Sources) or Streaming Destination will forward a continuous flow of data out of Stream to systems or connoting points such as Splunk HEC Non-Stream Destination wills send data to retention endpoints such as S3 buckets, Azure Blob Storage and others. Other Destinations will serve special purposes within the data flow. [Destinations](<https://docs.cribl.io/logstream/destinations>)

- Select Data/Sources > Sources from the top menu
- Explore all sources available, filter at the bottom selected top menu for Collectors, Push, Pull, System and Internal sources.

The screenshot shows the Splunk Data Sources interface. At the top, there are tabs for Data / Sources, Routing, Processing, Monitoring, and Notifications. Below the tabs, a search bar and a filter button labeled "Configured only" are present. A toolbar with icons for All, Collectors, Push, Pull, System and Internal, and a search bar are also visible.

**Collectors:**

- Collectors Azure Blob
- Collectors Filesystem
- Collectors Google Cloud Storage
- Collectors REST
- Collectors S3
- Collectors Script
- Collectors Splunk Search

**Push:**

- Syslog (1)
- TCP JSON (1)
- Splunk TCP (1)
- Splunk HEC (1)
- Amazon Firehose
- Prometheus Remote Write
- Grafana
- Loki
- HTTP (1)
- Raw HTTP
- Elasticsearch API (1)
- Metrics
- SNMP Trap (1)
- TCP (1)
- AppScope (1)
- OpenTelemetry
- Datadog Agent

- Select Push > Syslog Your sources are listed on the left and under Manage Syslog Sources you may configure several Syslog sources. These sources will act as a Syslog server receiving data from any syslog capable device sending data to any existing Syslog server (SyslogNG, Rsyslog, etc.).
- Click on the pre-configured source `in_syslog` and observe the configuration components available for this Source. Note, on the left menu, the out of the box TLS capability.

The screenshot shows the configuration page for the `in_syslog` source. The top navigation bar includes links for Sources, Syslog, `in_syslog`, Configure, Status, Charts, Live Data, Logs, and Help. The main configuration area is divided into sections: General Settings, TLS Settings (TCP Only), Processing Settings, Fields (Metadata), Pre-Processing, Advanced Settings, and Connected Destinations. The `General Settings` section contains fields for Input ID\*, Address\*, UDP Port, TCP Port, Fields to Keep, and Tags. The `TLS Settings (TCP Only)` section has an `Enabled` checkbox set to No. The `Processing Settings` section includes a `Fields (Metadata)` dropdown. The `Pre-Processing` section has fields for UDP Port and TCP Port. The `Advanced Settings` section has a `Connected Destinations` dropdown.

## 2. Configure a Source

- From the top menu select Data/Sources > Sources then from Push select Syslog
- Click on Add New From the top right button
- Enter the following values:
- Input ID: Syslog\_Source
- Address: 0.0.0.0
- UDP Port: 9514
- TCP Port: 9514
- Click Save and wait until the Live Status becomes green.

Manage Syslog Sources [Help](#)

ID	Address	UDP Port	TCP Port	Routes/QC	Enabled	Status
in_syslog	0.0.0.0	9514	9514	Routes	No	Live
Syslog_Source	0.0.0.0	9514	9514	Routes	Yes	Live

Your source has been configured, in this case Syslog, you may now send data from devices (firewalls, routers, servers, etc.) to the IP address for your Worker Node(s) and start receiving data.

*Note: no data has been stored on Cribl Stream.*

### 3. Configure a Splunk Destination

- Select Data/Sources > Destinations
- Click on Splunk Single Instance (Tile)
- Click on Add New From the top right button
- Enter the following values:
- Output ID: Splunk\_Lab
- Address: 10.253.33.249
- Port: 9997
- Backpressure behavior: Block
- Click Save and wait for the Live status to become green
- Click on the Splunk\_Lab destination
- Select Test in the context top menu and for Select Sample chose syslog.log
- Click Run Test

Observe the Test Results confirming your sample data was sent and received by the configured destination (Splunk\_Lab)

The screenshot shows the Splunk Test Input interface. At the top, there are tabs for Configure, Status, Charts, Live Data, Logs, and Test. The Test tab is selected. Below the tabs, there's a "Test Input" section with a code editor containing two JSON log entries. The first entry is from a host named "wintheiser6653" with severity 7, facility 11, and appname "consequuntur". The second entry is from a host named "torp3545" with severity 4, facility 0, and appname "minus". Both entries have a timestamp of 1574729435.447. To the right of the code editor, there are buttons for "Run Test", "Select sample: syslog.log", and "Select events". Below the code editor, it says "Test Results" and "Last run: 2022-03-18 03:14:18 UTC". A green box indicates a "Success" status.

```
[{"_raw": "<95>Nov 25 18:50:35 wintheiser6653 consequuntur[8196]: The THX circuit is down, generate the optical panel so we can connect the JSON application!", "host": "wintheiser6653", "severity": 7, "facility": 11, "severityName": "debug", "facilityName": "ftp", "appname": "consequuntur", "procid": "8196", "message": "The THX circuit is down, generate the optical panel so we can connect the JSON application!", "source": "/var/log/syslog", "sourcetype": "syslog", "_time": 1574729435.447}, {"_raw": "<4>Nov 25 18:50:35 torp3545 minus[9617]: Use the redundant JSON transmitter, then you can bypass the open-source firewall!", "host": "torp3545", "severity": 4, "facility": 0, "severityName": "warning", "facilityName": "kern", "appname": "minus", "procid": "9617", "message": "Use the redundant JSON transmitter, then you can bypass the open-source firewall!", "source": "/var/log/syslog", "sourcetype": "syslog", "_time": 1574729435.447}],
```

Test Results (Last run: 2022-03-18 03:14:18 UTC)

Success

## 4. Configure an Elastic Destination

- Select Data/Sources > Destinations
- Click on Elasticsearch (Tile)
- Click on Add New From the top right button

**Enter the following values:**

- Output ID: Elasticsearch
- Bulk API URL\*: 10.253.33.250
- Index: elastic\_lab
- Type: \_doc
- Authentication Enabled toggle set to Yes
- Authentication Method button set to Manual
- Username: admin
- Password: Go2atc4labs!
- Backpressure behavior: Block
- Click Save and wait for the Live status to become green
- Click on the Elasticsearch destination

- Select Test in the context top menu and for Select Sample chose syslog.log
- Click Run Test

Observe the Test Results confirming your sample data was sent and received by the configured destination (Elasticsearch).

The screenshot shows the Cribl interface with the 'Data / Destinations' tab selected. Under 'Destinations', 'Elasticsearch' is chosen. The 'Test' tab is active. In the 'Test Input' section, there are two log entries shown as JSON objects:

```

{
  "_raw": "<@Nov 25 18:50:35 wintheiser6653 consequuntur[8196]: The THX circuit is down, generate the optical panel so we can connect the JSON application!",
  "host": "wintheiser6653",
  "severity": 7,
  "facility": 11,
  "severityName": "debug",
  "facilityName": "ftp",
  "appname": "consequuntur",
  "procid": "8196",
  "message": "The THX circuit is down, generate the optical panel so we can connect the JSON application!",
  "source": "/var/log/syslog",
  "sourcetype": "syslog",
  "_time": 1574729435.447
},
{
  "_raw": "<@Nov 25 18:50:35 torp3545 minus[9617]: Use the redundant JSON transmitter, then you can bypass the open-source firewall!",
  "host": "torp3545",
  "severity": 4,
  "facility": 0,
}

```

The 'Test Results' section at the bottom shows a green bar with a checkmark and the word 'Success'.

## 5. Configure a S3 Bucket as Destination

For this task we will use an internal object storage solution to represent the S3 behavior or Simple Storage Service (Amazon S3).

- From the top menu select Data / Destinations.
- From the list of integrations select the MinIO tile
- Click on Add New From the top right button

**Enter the following values:**

- Output ID: S3\_Minio
- MinIO Endpoint\*: `http://192.168.2.52:9000`
- MinIO Bucket Name\*: 's3-syslog'
- Staging Location\*: `$/CRIBL_HOME/state/outputs/staging`
- Key Prefix\*: Cribl
- Partitioning Expression: `C.Time.strftime(_time ? _time : Date.now())/1000, '%Y/%m/%d'`
- Data Format: json
- File Name Prefix Expression: `CriblOut`
- File Name Suffix Expression:
 

```
.${C.env["CRIBL_WORKER_ID"]}.${__format}${__compression === "gzip" ? ".gz" : ""}
```
- Compress: none
- Backpressure behavior: Block
- Tags: <LEAVE\_EMPTY>

- Click Save

**You will be back on the Destinations list. Wait until the Status becomes green:**

- Click on S3\_Minio
- On the top menu (in the context window) click Test
- On Select sample chose syslog.log and click Run Test

## 6. Create a S3 Source

Now that we have a local S3 destination configured receiving data in your S3 bucket, lets configure a S3 collector to use the Replay function within Cribl Stream. We will configure a regular S3 bucket as a Source, the same way you would do if you wanted to read data from an AWS S3 bucket.

- From the list of integrations select the Collectors S3 tile
- Click on Add New From the top right button

**Enter the following values:**

- Output ID: S3\_collect
- Auto-populate from: <LEAVE\_BLANK>
- S3 bucket\*: 's3-syslog'
- Region: <LEAVE\_BLANK>
- Path: /Cribl/\${\_time:%Y}/\${\_time:%m}/\${\_time:%d}/
- Extend AUTHENTICATION and Select Manual
- Access key: admin
- Secret key: Go2atc4labs!
- Click Save

**Back to the Sources list**

- Click on the S3\_collect source configured
- On the bottom left click Run
- On the new context window click Run

On the result context window, you should see the content from the S3 destination (syslog.log sample file) played back as result

The screenshot shows a 'Previewing' window for a collection named '53\_collect'. The filter expression is set to '\_inputID=="collection:1648518251.2.adhoc.53\_collect"'. The preview pane displays four log entries (1, 2, 3, 4) with the following details:

- Entry 1:** Raw log: {"\_raw": "2021-11-25T19:50:35.000Z", "host": "criblout\_02", "severity": "info", "facility": "local0", "appname": "criblout", "time": "1637687835", "cribl\_breaker": "fallback", "host": "Cribl-Server", "source": "Cribl/2022/03/18/CriblOut\_02\_EB0078.0.json"}.
- Entry 2:** Raw log: {"\_raw": "2021-11-25T19:50:35.000Z", "host": "criblout\_02", "severity": "info", "facility": "local0", "appname": "criblout", "time": "1637687835", "cribl\_breaker": "fallback", "host": "Cribl-Server", "source": "Cribl/2022/03/18/CriblOut\_02\_EB0078.0.json"}.
- Entry 3:** Raw log: {"\_raw": "2021-11-25T19:50:35.000Z", "host": "criblout\_02", "severity": "info", "facility": "local0", "appname": "criblout", "time": "1637687835", "cribl\_breaker": "fallback", "host": "Cribl-Server", "source": "Cribl/2022/03/18/CriblOut\_02\_EB0078.0.json"}.
- Entry 4:** Raw log: {"\_raw": "2021-11-25T19:50:35.000Z", "host": "criblout\_02", "severity": "info", "facility": "local0", "appname": "criblout", "time": "1637687835", "cribl\_breaker": "fallback", "host": "Cribl-Server", "source": "Cribl/2022/03/18/CriblOut\_02\_EB0078.0.json"}.

## Quick Connect

Within Cribl Stream you can send data from sources to destinations already configured with a drag and drop action. You may also add a pipeline to your quick connection and process data independently of existing routes (if any configured)

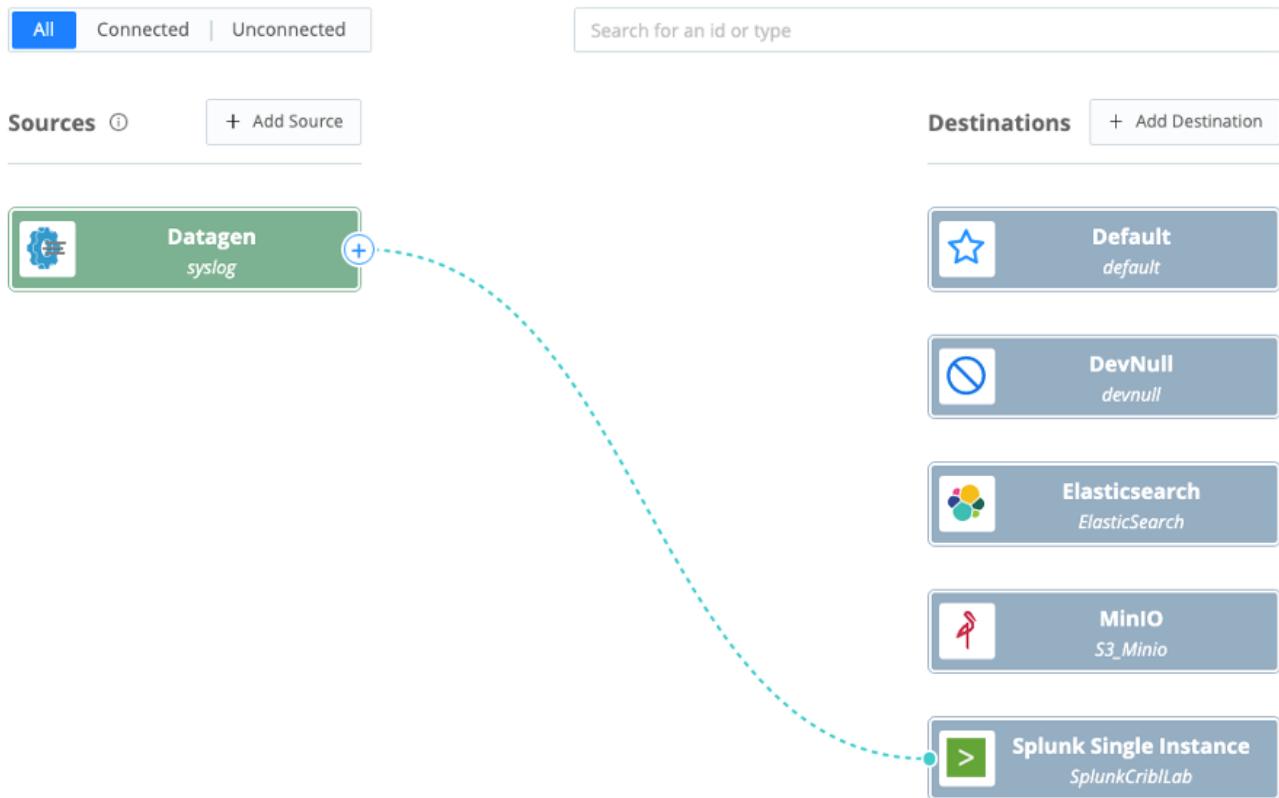
### 7. Use QuickConnect to send data to Splunk and Elastic

- From the top main menu select Routing/QuickConnect
- From Sources (left side) click +New Source
- From the new context window browse to System and Internal and mouse over Datagen then click 'Select Existing'
- From the list of available datagen sources, select 'syslog'
- On the new context window click Yes.
- Back on the Quick Connect panel click and drag the mouse connecting the Datagen/syslog (+ sign) to the Splunk Single Instance (SplunkCriblLab) connecting both objects.

### 8. Apply a passthrough and 1 pipeline to the QuickConnect route.

A new context window will present you a selection to choose how to process the data in this connected route.

- Click on Passthru (sending raw data to the destination, no pipelines applied) click save.



9. Analyze the results in Splunk or Elastic (Splunk and Elastic dashboards being created for richer visualization and value realization)

- Mouse over the Single Instance (SplunkCribLab) Destination and click 'Capture' to display if data is being sent to the selected destination.
- From the new context window validate if syslog data (your datagen source in this case) is being sent to the Splunk Single Instance configured.

The screenshot shows the Splunk Single Instance interface under the 'Live Data' tab. A filter expression `__outputId == 'splunk:SplunkCribLab'` is applied. The log entries listed are:

```

1  _raw: <4>Nov 25 18:50:35 reichelli146 nostrum[6664]: Try to hack the SAS program, maybe it will synthesize the 1080p pixel!
2019-11-25 # _time: 1574729435.447
19:50:35.447 appname: nostrum
-05:00 facility: 0
facilityName: kern
host: reichelli146
message: Try to hack the SAS program, maybe it will synthesize the 1080p pixel!
procid: 6664
severity: 4
severityName: warning
source: /var/log/syslog
sourcetype: syslog

2  _raw: <4>Nov 25 18:50:35 reichelli146 nostrum[6664]: Try to hack the SAS program, maybe it will synthesize the 1080p pixel!
2019-11-25 # _time: 1574729435.447
19:50:35.447 appname: nostrum
-05:00 facility: 0
facilityName: kern
host: reichelli146
message: Try to hack the SAS program, maybe it will synthesize the 1080p pixel!
procid: 6664
severity: 4
severityName: warning
source: /var/log/syslog
sourcetype: syslog

3  _raw: <4>Nov 25 18:50:35 boehm1m1712 asperiores[9419]: Backing up the alarm won't do anything, we need to override the digital
2019-11-25 SDD system!
# _time: 1574729435.447
19:50:35.447 appname: asperiores
-05:00 facility: 0

```

Buttons at the bottom include 'Create A Datagen File' and 'Save as Sample'.

Follow the same process to add Elasticsearch as a second destination receiving the same source (Datagen syslog) already sending data to Splunk Single Instance.

- Mouse over Elasticsearch and click on 'Capture' to validate if data is being sent to the configured destination.
- Select from the top menu Data/Sources and click the Datagen tile.
- In the syslog datagen source, click on 'Connected Destinations' from the left menu.
- On the right pane click on 'Send to Routes'
- From the new context window click Yes
- Click Save.

## Part 2 – Routes and Pipelines

Now that we have successfully sent data via QuickConnect, lets use Routes to send data to the same destinations. First let's enable another Source from our Datagen (windows\_xml).

### 1. Enable a new source

- From the top menu click on Data/Source
- Click on the Datagen tile
- From the Manage Datagen Sources pane click on the no toggle under the Enable column on the windows\_xml source.
- On the new context window click Yes.

- On the same source (windows\_xml) click on Live under the Status column and validate the proper sources is being generated.
- Click on the top X on the opened context window.

## 2. Create a new Route

- From the top menu select Routing/Data Routes
- In the Routes panel (left) click the + Route to add a new Route to the existing ones.

The screenshot shows the Splunk interface for managing routes. At the top, there's a search bar labeled 'Search routes'. Below it, a toolbar with icons for search, refresh, and settings. To the right are buttons for '+ Route', '...', and a gear icon. Underneath the search bar are several filter options: 'Route', 'Filter', 'Pipeline/Output', 'Events' (with 'In', 'Out', and 'Dropped' sub-options), and 'All' (with 'In', 'Out', and 'Dropped' sub-options). There are also some other small icons like a magnifying glass and a trash can.

**On the newly create Route enter the following values:**

- Route Name\*: to\_Splunk
- Filter: select \_\_inputId=='datagen:windows\_xml'
- Pipeline\*: passthru
- Enable Expression: No
- Output: select splunk:SplunkCribLab
- Final: No

The screenshot shows the 'Edit Route' dialog for a route named 'to\_Splunk'. The top header bar shows the route name, filter, pipeline, and status (0.000% complete). Below the header are several input fields and controls:
 

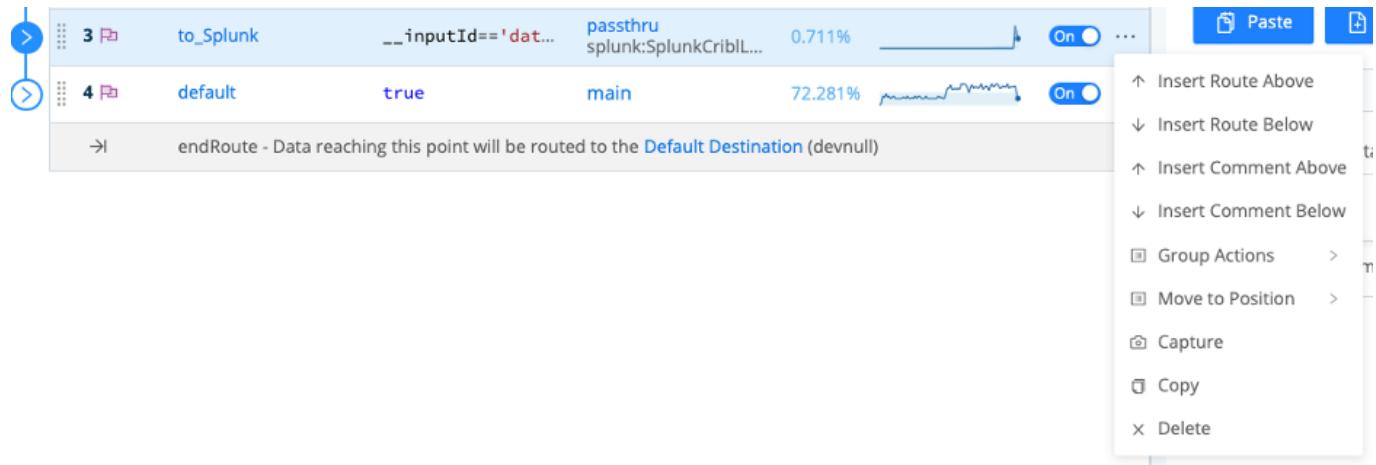
- Route Name\***: to\_Splunk
- Filter**: \_\_inputId=='datagen:windows\_xml'
- Pipeline\***: passthru
- Enable Expression**: No (checkbox)
- Output**: splunk:SplunkCribLab
- Description**: First route to Splunk
- Final**: No (checkbox)

 At the bottom left is a 'Clones' section with a '+ Add Clone' button. At the bottom right are 'Cancel' and 'Save' buttons. A note at the bottom states: '→ endRoute - Data reaching this point will be routed to the Default Destination (devnull)'.

- Click Save

## 3. Capture sample data from the route created

- Make sure your new route is not below any other routes with the Final toggle set to yes, if it is drag it above that final route.
- From the to\_Splunk Route click on the 3 dots and select Capture



On the new context window validate you are capturing samples from your configured Route

The screenshot shows the 'Capture Sample Data' dialog box. It displays a list of XML event samples. The filter expression is set to '\_inputId==`datagen:windows\_xml`'. The list includes several events with fields like \_raw, \_time, and \_id. At the bottom, there are buttons for 'Capture...', 'Cancel', 'Create A Dataname File', and 'Save as Sample File'.

- At the bottom right click on Save as Sample File

**On the new context window enter the following values:**

- File Name\*: windows\_xml\_sample.log
- Description: <LEAVE\_EMPTY>
- Expiration (hours): <LEAVE\_EMPTY>
- Tags: <LEAVE\_EMPTY>
- From the right pane validate if your sample file has been created (if not refresh your browser)
- Under Preview click Simple and validate your sample data from your Route.

Sample Data    Preview Simple ?    Preview Full ?    Quick Stats

Sample data file    Pipeline

windows\_xml\_sample.log    Select a pipeline    Run

**IN OUT**    Select Fields (3 of 3)

```

1   @ _raw: <Event xmlns='http://schemas.microsoft.com/win/2004/08/events/event'><System><Provider Name='Microsoft-Wind
2022-03-28      ows-Security-Auditing' Guid='{54849625-5478-9876-A1B2-3E3B0328C30D}'/><EventID>4672</Event... Show more
23:11:22.327    # _time: 1648523482.327
-04:00          @ cribl_pipe: time_adjustment

2   @ _raw: <Event xmlns='http://schemas.microsoft.com/win/2004/08/events/event'><System><Provider Name='Microsoft-Wind
2022-03-28      ows-Security-Auditing' Guid='{54849625-5478-9876-A1B2-3E3B0328C30D}'/><EventID>4624</Event... Show more
23:11:22.328    # _time: 1648523482.328
-04:00          @ cribl_pipe: time_adjustment

3   @ _raw: <Event xmlns='http://schemas.microsoft.com/win/2004/08/events/event'><System><Provider Name='Microsoft-Wind
2022-03-28      ows-Security-Auditing' Guid='{54849625-5478-9876-A1B2-3E3B0328C30D}'/><EventID>4634</Event... Show more
23:11:22.330    # _time: 1648523482.33
-04:00          @ cribl_pipe: time_adjustment

4   @ _raw: <Event xmlns='http://schemas.microsoft.com/win/2004/08/events/event'><System><Provider Name='Microsoft-Wind
2022-03-28      ows-Security-Auditing' Guid='{54849625-5478-9876-A1B2-3E3B0328C30D}'/><EventID>4625</Event... Show more
23:11:22.332    # _time: 1648523482.332
-04:00          @ cribl_pipe: time_adjustment

5   @ _raw: <Event xmlns='http://schemas.microsoft.com/win/2004/08/events/event'><System><Provider Name='Microsoft-Wind
2022-03-28      ows-Security-Auditing' Guid='{54849625-5478-9876-A1B2-3E3B0328C30D}'/><EventID>4634</Event... Show more
23:11:24.326    # _time: 1648523484.326
-04:00          @ cribl_pipe: time_adjustment

6   @ _raw: <Event xmlns='http://schemas.microsoft.com/win/2004/08/events/event'><System><Provider Name='Microsoft-Wind
2022-03-28      ows-Security-Auditing' Guid='{54849625-5478-9876-A1B2-3E3B0328C30D}'/><EventID>4625</Event... Show more
23:11:24.326    # _time: 1648523484.326
-04:00          @ cribl_pipe: time_adjustment

7   @ _raw: <Event xmlns='http://schemas.microsoft.com/win/2004/08/events/event'><System><Provider Name='Microsoft-Wind
2022-03-28      ows-Security-Auditing' Guid='{54849625-5478-9876-A1B2-3E3B0328C30D}'/><EventID>4625</Event... Show more
23:11:24.336    # _time: 1648523484.336
-04:00          @ cribl_pipe: time_adjustment

```

## 4. Create a Pipeline

Let's create a simple pipeline to process our Syslog source.

- Go to the top menu and select Processing / Pipelines
- On the right pane click on the Syslog\_sample.log sample file (we will use the captured sample with this Pipeline)
- On the left pane at the top click the '+ Pipeline' button and select 'Create Pipeline'

Search pipelines		Show All	All Processing   Pre/Post	+ Pipeline			
Pipeline...	Description...	Route ...	Functions	Output	Stats		Create Pipeline
cisco_asa	Filter an...	None	11	None	IN 0	OUT 0	ERR
cisco es...		None	3	None	IN 0	OUT 0	ERR

- On the left pane, enter only the ID as Syslog\_test and click Save.

## 5. Add functions to the pipeline

- Click on the '+ Function' button on the right most side within the newly create Pipeline

Build out your pipeline. Use **Preview** to help shape your data.

- Add Function(s) and Preview**  
Click on Add Function on top right. Use Preview to ensure events are processed correctly.
- Attach to a Route**  
Associate this pipeline with a route (top left).

+ Function    Sample Data    Preview Simple    Preview Full

Search

- Advanced
- Beta Functions
- Deprecated Functions
- Formatters
- Standard

a message: You can't synthesize the bus without generating the wireless ADP driver

- Mouse click Standard / Eval or type Eval on the mini search bar and click Eval.
- In the Function values enter the word 'message' (no quotes) in the Remove Fields field and click Save.
- Make sure your Syslog\_sample.log sample file is loaded on the right pane and validate the results by clicking on the OUT button on the top bar within the Sample data loaded.

Syslog\_test Attach to Route

+ Function    Sample Data    Preview Simple    Preview Full

All

Sample data file: syslog\_sample.log

Pipeline: Syslog\_test

IN OUT

1

\_raw: <@Nov 25 18:50:35 ullrich5633 itaque[5396]: You can't synthesize the bus without generating the wireless ADP driver.  
\_time: 1574729455.448  
appname: itaque  
crthlpipes: Syslog\_test  
-05:00  
facility: 0  
facilityName: kern  
host: ullrich5633  
level: 4  
levelName: warning  
source: /var/log/syslog  
sourcetype: syslog

2

\_raw: <@Nov 25 18:50:35 kemmer5135 nostrum[9259]: Parsing the protocol won't do anything, we need to calculate the auxillary SMS protocol!  
\_time: 1574729455.448  
appname: nostrum  
crthlpipes: Syslog\_test  
-05:00  
facility: 0  
facilityName: kern  
host: kemmer5135  
level: 4  
levelName: warning  
source: /var/log/syslog  
sourcetype: syslog

Now that we have excluded the field message from the processing logs, lets use another Function "Drop" to reduce our data even further.

- On the left pane click on the **+ Function** button and select Standard/Drop or type Drop on the mini search bar and click on the result.
- With the Function loaded, enter the following values in the field **Filter: appname=='itaque'**

Note the events matching the filter in the Drop function are greyed out and will not be sent to the destination thus reducing the number of events.

Let's add another function to change/redact our data

- From the top bar on the left pane click on the “+ Function” button and select Standard/Rename or type Rename in the mini search bar and click on the result.
- In the Function click the “+ Add field” button
- Within the Rename fields group type facilityName in the Current name and NEW\_facility\_Name in the New Name fields
- Click Save

Observe the results on the right pane with the syslog\_sample.log file selected and the OUT button enabled

**Now that we have defined a Pipeline, we need to attach it to a Route.**

## 6. Add a Pipeline to the Route

- On the left pane within your Splunk\_test Pipeline, click on the top left link “Attach to Route”

You will be brought to the Routes list.

- Select the Syslog\_to\_Splunk Ruote (created by you earlier)
- In Pipeline select Splunk\_test

## 7. Apply the destination to the Route

Now lets apply the destination that will receive the process stream from this Route.

- From Output select splunk:SplunkCribLab
- In Description enter: "Sending Syslog data to Splunk"
- Final toggle set to No
- Click Save

Syslog\_to\_splunk

Route Name\* Syslog\_to\_splunk

Filter `__inputId=='datagen:syslog'`

Pipeline\* passthru

Enable Expression No

Output splunk:SplunkCribLab

Description Sending Syslog data to Splunk

Final No

Clones Add Clone

→ endRoute - Data reaching this point will be routed to the Default Destination (devnull)

Cancel Save

## 8. Analyze results on Splunk or Elastic (no dashboards provided for this part)

- Mouse over the Single Instance (SplunkCribLab) Destination and click 'Capture' to display if data is being sent to the selected destination.
- From the new context window validate if the syslog data (your datagen source in this case) is being sent to the Splunk Single Instance configured.

The screenshot shows the Splunk CriblLab interface with the 'Live Data' tab selected. A filter expression is set to `__outputId == 'splunk:SplunkCriblLab'`. The left sidebar lists various fields like \_raw, \_time, appname, facility, etc., with checkboxes. The main pane displays three log entries:

```

1  _raw: <4>Nov 25 18:50:35 reichelli146 nostrum[6664]: Try to hack the SAS program, maybe it will synthesize the 1080p pixel!
2019-11-25 # _time: 1574729435.447
19:50:35.447 appname: nostrum
-05:00 facility: 0
facilityName: kern
host: reichelli146
message: Try to hack the SAS program, maybe it will synthesize the 1080p pixel!
procid: 6664
severity: 4
severityName: warning
source: /var/log/syslog
sourcetype: syslog

2  _raw: <4>Nov 25 18:50:35 reichelli146 nostrum[6664]: Try to hack the SAS program, maybe it will synthesize the 1080p pixel!
2019-11-25 # _time: 1574729435.447
19:50:35.447 appname: nostrum
-05:00 facility: 0
facilityName: kern
host: reichelli146
message: Try to hack the SAS program, maybe it will synthesize the 1080p pixel!
procid: 6664
severity: 4
severityName: warning
source: /var/log/syslog
sourcetype: syslog

3  _raw: <4>Nov 25 18:50:35 boehm1712 asperiores[9419]: Backing up the alarm won't do anything, we need to override the digital
2019-11-25 SDD system!
19:50:35.447 # _time: 1574729435.447
-05:00 appname: asperiores
facility: 0

```

Buttons at the bottom include 'Create A Datagen File' and 'Save as Sample'.

**Follow the same process to add Elasticsearch as a second destination to receive the same source (Datagen syslog) as the Splunk Single Instance.**

- Mouse over Elasticsearch and click on 'Capture' to validate if data is being sent to the configured destination.

## Disconnecting the QuickConnect Routes

Now select from the top menu Data/Sources and click the Datagen tile.

- In the syslog datagen source click on 'Connected Destinations' from the left menu.
- On the right pane click on 'Send to Routes'
- From the new context window click Yes
- Click Save.

## Part 3 - Cribl Stream Use Cases

### Reduction

Let's put in practice some of the techniques we learned so far. Reduction can be done through a series of functions in a pipeline, we will use 2 simple examples on different sources.

#### 1. Use the Json reduction elements

**Check if your data source is active and running:** Data / Source / DataGen

- Big\_JSON source should be active and live. (take a look by clicking in the live button)
- After data stops flowing click on the bottom right button "Save as Sample File"

**On the Sample Files Settings use the following values:**

- File Name: big\_json.log
- Description: <LEAVE\_BLANK>
- Expiration (hours): delete any values and leave in blank, if you add any number (hours) the sample will be removed after this time.
- Tags: <LEAVE\_BLANK>
- Click Save

**Create a new pipeline from Processing / Pipelines:** (note your sample file "big\_json\_data.log" listed on the right panel)

- Click in the + Pipeline button (top right of the pipelines panel) / Create Pipeline
- Enter Reduction for ID Leave all other fields as they are.
- Click on the top right button (+ Function) to add a function to the Reduction Pipeline
- From the list select Standard / Parser
- Load your sample file (big\_json\_data.log) from the right panel and expand the \_raw field to see its contents.

**For this usecase we will remove the multiValueHeaders field (which has the same value as the headers) and eliminate any fields that contain the value "null"**

**Enter the following values for the Parser function you've just created:**

- Filter: true
- Descripgion: <LEAVE\_BLANK>
- Final: No
- Operation Mode: Reserialize
- Type: JSON Object
- Source Field: \_raw
- Destination Field: <LEAVE\_BLANK>
- List of Fields: <LEAVE\_BLANK>
- Fields To Keep: <LEAVE\_BLANK>
- Fields To Remove: multiValueHeaders.\*
- Fields Filter Expression: value != null

You should see the fields removed on the right panel with the big\_json\_data.log sample loaded and by clicking the top left OUT button (in the sample panel)

The screenshot shows the Cribl Stream Processing interface. On the left, there's a configuration panel for a 'Parser' function. It includes fields for 'Filter' (set to 'true'), 'Description' (empty), 'Operation Mode' (set to 'Reserialize'), 'Type' (set to 'JSON Object'), 'Source Field' (set to '\_raw'), and 'Destination Field' (empty). Below these are sections for 'List of Fields', 'Fields To Keep', 'Fields To Remove' (containing 'multiValueHeaders.\*'), and a 'Fields Filter Expression' (set to 'value != null'). On the right, the 'Sample Data' panel shows a preview of the input and output data. The input data is a log entry from April 23, 2020, at 09:37:40.000. The output data is a reduced version of this log entry, showing only the '\_raw' field and the '\_cribl\_pipe: Reduction' field.

Visualize results on Cribl Stream basic statistics (no system of analysis required)

- At the top right click on the Basic Statistics icon (to the left of the Select Fields) drop down within the Sample panel:

This screenshot shows the basic statistics visualization panel. It displays three rows of data: IN (34.48KB, 2 events, 10 total), OUT (19.24KB, 3 events, 10 total), and DIFF (-44.19%, -43.14%, 50.00% reduction). Below this, a detailed view of the reduced event is shown, highlighting fields like '\_time' and '\_cribl\_pipe'. A 'Select Fields (3 of 3)' button is visible on the right.

	_raw Length	Full Event Length	Number of Fields	Number of Events
IN	34.48KB	34.75KB	2	10
OUT	19.24KB	19.76KB	3	10
DIFF	↓ -44.19%	↓ -43.14%	↑ 50.00%	0.00%

You reduced 44.19% of this data source.

## 2. Create a new S3 destination for the Big Json data reduced.

Just as we did in the **Configure a S3 Bucket as Destination** task lets configure a new destination to our S3 storage using a new bucket for this data source (big-json)

- From the top menu select Data / Destinations.
- From the list of integrations select the MinIO tile
- Click on Add New From the top right button

## Enter the following values:

- Output ID: Big\_Json\_reduced
- MinIO Endpoint\*: http://192.168.2.52:9000
- MinIO Bucket Name\*: 'big-json'
- Staging Location\*: \$CRIBL\_HOME/state/outputs/staging
- Key Prefix\*: Cribl
- Partitioning Expression: C.Time.strftime(\_time ? \_time : Date.now()/1000, '%Y/%m/%d')
- Data Format: json
- File Name Prefix Expression: CriblOut
- File Name Suffix Expression:  

```
.${C.env["CRIBL_WORKER_ID"]}.${__format}${__compression === "gzip" ? ".gz"
: ""}
```
- Compress: none
- Backpressure behavior: Block
- Tags: <LEAVE\_EMPTY>
- Click save

Following the same process as in the **Configure a S3 Bucket as Destination** task, validate the configuration for the new Destination (MinIO\_Big\_Json)

- Clicking on it and selecting Test from the top menu.
- From Select Sample: choose big\_json.log
- Click the Run Test button (at the top right)

## You should have the following results:

```
_raw": "{\"resource\": \"/done\", \"path\": \"/done\", \"httpMethod\": \"POST\", \"headers\": {\"Accept\": \"application/json, text/javascript, */*; q=0.01\", \"Accept-Encoding\": \"gzip, deflate, br\", \"Accept-Language\": \"en-US,en;q=0.9\", \"CloudFront-Forwarded-Proto\": \"https\", \"CloudFront-Is-Desktop-Viewer\": \"true\", \"CloudFront-Is-Mobile-Viewer\": \"false\", \"CloudFront-Is-SmartTV-Viewer\": \"false\", \"CloudFront-Is-Tablet-Viewer\": \"false\", \"CloudFront-Viewer-Country\": \"US\", \"Content-Type\": \"application/json\", \"Host\": \"jef18742ajw.execute-api.us-west-2.amazonaws.com\", \"origin\": \"https://cdn.foo.com\", \"Referer\": \"https://cdn.foo.com/bugger/index.html\", \"User-Agent\": \"Mozilla/5.0 (Linux; U; Android 2.3.4; ko-kr; HTC_X515E Build/GRJ22) AppleWebKit/533.1 (KHTML, like Gecko) Version/4.0 Mobile Safari/533.1\", \"Via\": \"2.0 0776SH00yVunTP5M766C1N2yGTHX0lp.cloudfront.net (CloudFront)\", \"X-Amz-Cf-Id\": \"1jsk2VQF96vKeckdvt1YbrB5MJP13q6a9JkbSPg7N26PzN9mUD0kCD\", \"X-Amzn-Trace-Id\": \"Root=0ImGwfkveoel17e062x61gClay9ai6K0j\", \"X-Forwarded-For\": \"12.0.0.1\", \"X-Forwarded-Port\": \"443\", \"X-Forwarded-Proto\": \"https\", \"MultiValueHeaders\": {\"Accept\": [\"application/json, text/javascript, */*; q=0.01\"], \"Accept-Encoding\": [\"gzip, deflate, br\"], \"Accept-Language\": [\"en-US,en;q=0.9\"], \"CloudFront-Forwarded-Proto\": [\"https\"], \"CloudFront-Is-Desktop-Viewer\": [\"true\"], \"CloudFront-Is-Mobile-Viewer\": [\"false\"], \"CloudFront-Is-SmartTV-Viewer\": [\"false\"], \"CloudFront-Is-Tablet-Viewer\": [\"false\"], \"CloudFront-Viewer-Country\": [\"US\"], \"Host\": [\"jef18742ajw.execute-api.us-west-2.amazonaws.com\"], \"origin\": [\"https://cdn.foo.com\"], \"Referer\": [\"https://cdn.foo.com/bugger/index.html\"], \"User-Agent\": [\"Mozilla/5.0 (Linux; U; Android 2.3.4; ko-kr; HTC_X515E Build/GRJ22) AppleWebKit/533.1 (KHTML, like Gecko) Version/4.0 Mobile Safari/533.1\"], \"Via\": [\"2.0 0776SH00yVunTP5M766C1N2yGTHX0lp.cloudfront.net (CloudFront)\", \"X-Amz-Cf-Id\": [\"4FfvP694Z17bfadH66nyt1ts3Cw4S8.cloudfront.net (CloudFront)\"], \"X-Forwarded-For\": [\"12.0.0.1\"], \"X-Forwarded-Port\": [\"443\"], \"X-Forwarded-Proto\": [\"https\"], \"queryStringParameters\": null, \"multiValueQueryStringParameters\": null, \"pathParameters\": null, \"stageVariables\": null, \"requestContext\": {\"resourceId\": \"vleewyl\", \"resourcePath\": \"/done\"}}}}
```

Test Results (1) Last run: 2022-04-11 19:44:12 UTC

<span style="color: green;">✓</span> Success
--

## 3. Send data to a S3 bucket (In this lab will use an internal MinIO server)

To send data to the S3 bucket, we will need to create a new route.

- From Routing / Data Routes Click the + Route at the top right button within the Routes panel (left)

## Enter the following values:

- Route Name\*: Big\_json\_to\_S3
- Filter: \_\_inputId=='datagen:Big\_JSON'
- Pipeline\*: Reduction
- Enable Expression: No
- Output: minio:Big\_Json\_reduced
- Description: <LEAVE\_BLANK>
- Final: No

Route Name\* Big\_json\_to\_S3

Filter `__inputId=='datagen:Big_JSON'`

Pipeline\* Reduction

Enable Expression No

Output minio:Big\_Json\_reduced

Description Enter a description

Final Yes

→ endRoute - Data reaching this point will be routed to the [Default Destination \(devnull\)](#)

[Cancel](#) [Save](#)

- Click Save

## 4. Check if data has arrived on the S3 bucket.

From the Routes panel, observe that the recently created Route now is sending data to the S3 bucket.



- Click on the three dots next to the **On** toggle and select **Capture**

You should see data coming from your Source, being processed by your Pipeline and sent to your Destination

The screenshot shows the Cribl Capture Sample Data interface. At the top, there's a filter expression bar with the value '\_InputId!=#datagen:Big\_JSON'. Below it is a 'Fields' dropdown set to 'All'. The main area displays a table of captured logs:

	_raw	_time
1	... raw: {"resource": "/done", "path": "/done", "httpMethod": "POST", "headers": {"Accept": "application/json, text/javascript, */*; q=0.81", "Accept-Encoding": "gzip, deflate, br", "Accept-Language": "en-US,en;q=0.81"}, ... Show more 2020-04-23 09:37:40.000 #_time: 1587649868 -04:00	
2	... raw: {"resource": "/done", "path": "/done", "httpMethod": "POST", "headers": {"Accept": "application/json, text/javascript, */*; q=0.81", "Accept-Encoding": "gzip, deflate, br", "Accept-Language": "en-US,en;q=0.81"}, ... Show more 2020-04-23 09:37:40.000 #_time: 1587649868 -04:00	
3	... raw: {"resource": "/done", "path": "/done", "httpMethod": "POST", "headers": {"Accept": "application/json, text/javascript, */*; q=0.81", "Accept-Encoding": "gzip, deflate, br", "Accept-Language": "en-US,en;q=0.81"}, ... Show more 2020-04-23 09:37:41.000 #_time: 1587649861 -04:00	
4	... raw: {"resource": "/done", "path": "/done", "httpMethod": "POST", "headers": {"Accept": "application/json, text/javascript, */*; q=0.81", "Accept-Encoding": "gzip, deflate, br", "Accept-Language": "en-US,en;q=0.81"}, ... Show more 2020-04-23 09:37:40.000 #_time: 1587649868 -04:00	
5	... raw: {"resource": "/done", "path": "/done", "httpMethod": "POST", "headers": {"Accept": "application/json, text/javascript, */*; q=0.81", "Accept-Encoding": "gzip, deflate, br", "Accept-Language": "en-US,en;q=0.81"}, ... Show more 2020-04-23 09:37:40.000 #_time: 1587649868 -04:00	
6	... raw: {"resource": "/done", "path": "/done", "httpMethod": "POST", "headers": {"Accept": "application/json, text/javascript, */*; q=0.81", "Accept-Encoding": "gzip, deflate, br", "Accept-Language": "en-US,en;q=0.81"}, ... Show more 2020-04-23 09:37:40.000 #_time: 1587649868 -04:00	

At the bottom right, there are buttons for 'Cancel', 'Create A Datagen File', and 'Save as Sample File'.

- Click Cancel (we already have a sample file for this data captured at the Source)

## Replay

### 1. Configure a S3 Collector Source utilizing the S3 bucket created above

Now let's replay the data sent into the S3 bucket configured above (Big\_JSON\_reduced)

- Go to the main menu at the top and select Sources
- Click on the S3 Collector tile
- Click on the + Add New button at the top right and enter the following values:
  - Collector ID: Big\_JSON\_Collect
  - Auto-populate from: <LEAVE\_BLANK>
  - S3 Bucket: 'big-json'
  - Region: <LEAVE\_BLANK>
  - Path: /Cribl/\${\_time:%Y}/\${\_time:%m}/\${\_time:%d}/
  - Path Extractors: <LEAVE\_BLANK>
  - Recursive: Yes
  - Max Batch Size (objects): 10
- Click on AUTHENTICATION expansion
- Authentication Method: Manual
- Access Key: admin
- Secret Key: Go2atc4labs!

From the menu on the left select:

- Results Settings / Fields
- Click on Event Breakers
- On the right click on + Add Ruleset
- Select Cribl (ndjson breaker)
- Click on Fields on the left
- Click on the + Add Field and enter the following values:
- Name: source
- Value: 'big-json-from-collector'
- Click on Result Routing on the left
- Enter the following values on the right:
- Send to Routes: Yes
- Pre-Processing Pipeline: time\_adjustment
- Throttling: 0

Leave all other fields unchanged

- Click Save

The screenshot shows the Cribl Stream interface with the 'New Collector' configuration dialog open. The 'Collector Settings' tab is active. Key configuration details include:

- Collector ID:** Big\_json\_Collect
- S3 bucket:** big-json
- Path:** /Cribl\${\_time:N} \${\_time:N} \${\_time:N}
- Result Routing:** Set to 'Yes' (selected)
- Pre-Processing Pipeline:** time\_adjustment
- Throttling:** 0
- Advanced Settings:**
  - Recursive:** Enabled (radio button)
  - Authentication Method:** Manual (selected)
  - Access key:** admin
  - Secret key:** (redacted)
  - Tags:** (empty)

At the bottom of the dialog are buttons for 'Save & Run' and 'Save & Schedule', and a 'Cancel' button.

## 2. Replay data in preview mode within Cribl Stream

- Back to the Collectors Source from the Big.Json\_Collect item in Actions column click Run
- Select Preview and leave all fields unchanged
- Click Run

You should have data coming from the S3 Bucket configured with data already reduced by Cribl on the Reduction pipeline earlier.

Filter Expression\* ⓘ  
\_\_inputId='collection:1649728246.33.adhoc.Big\_Json\_Collect'

Fields All None

	_raw	_time	cribl_breaker	host	source
1	... raw: {"resource": "/done", "path": "/done", "httpMethod": "POST", "headers": {"Accept": "application/json, text/javascript, */*; q=0.01", "Accept-Encoding": "gzip, deflate, br", "Accept-Language": "en-US, en;q=0.9"}, "query": "S_wmpB... Show more", "time": 1587549960, "cribl_breaker": "Cribl:ndjson", "host": "Cribl-Server", "source": "big_json-from-collector"}	2020-04-23 09:37:40.000	-0:00	Cribl-Server	big_json-from-collector
2	... raw: {"resource": "/done", "path": "/done", "httpMethod": "POST", "headers": {"Accept": "application/json, text/javascript, */*; q=0.01", "Accept-Encoding": "gzip, deflate, br", "Accept-Language": "en-US, en;q=0.9"}, "query": "S_wmpB... Show more", "time": 1587549961, "cribl_breaker": "Cribl:ndjson", "host": "Cribl-Server", "source": "big_json-from-collector"}	2020-04-23 09:37:41.000	-0:00	Cribl-Server	big_json-from-collector
3	... raw: {"resource": "/done", "path": "/done", "httpMethod": "POST", "headers": {"Accept": "application/json, text/javascript, */*; q=0.01", "Accept-Encoding": "gzip, deflate, br", "Accept-Language": "en-US, en;q=0.9"}, "query": "S_wmpB... Show more", "time": 1587549960, "cribl_breaker": "Cribl:ndjson", "host": "Cribl-Server", "source": "big_json-from-collector"}	2020-04-23 09:37:40.000	-0:00	Cribl-Server	big_json-from-collector
4	... raw: {"resource": "/done", "path": "/done", "httpMethod": "POST", "headers": {"Accept": "application/json, text/javascript, */*; q=0.01", "Accept-Encoding": "gzip, deflate, br", "Accept-Language": "en-US, en;q=0.9"}, "query": "S_wmpB... Show more", "time": 1587549960, "cribl_breaker": "Cribl:ndjson", "host": "Cribl-Server", "source": "big_json-from-collector"}	2020-04-23 09:37:40.000	-0:00	Cribl-Server	big_json-from-collector
5	... raw: {"resource": "/done", "path": "/done", "httpMethod": "POST", "headers": {"Accept": "application/json, text/javascript, */*; q=0.01", "Accept-Encoding": "gzip, deflate, br", "Accept-Language": "en-US, en;q=0.9"}, "query": "S_wmpB... Show more", "time": 1587549960, "cribl_breaker": "Cribl:ndjson", "host": "Cribl-Server", "source": "big_json-from-collector"}	2020-04-23 09:37:40.000	-0:00	Cribl-Server	big_json-from-collector
6	... raw: {"resource": "/done", "path": "/done", "httpMethod": "POST", "headers": {"Accept": "application/json, text/javascript, */*; q=0.01", "Accept-Encoding": "gzip, deflate, br", "Accept-Language": "en-US, en;q=0.9"}, "query": "S_wmpB... Show more", "time": 1587549961, "cribl_breaker": "Cribl:ndjson", "host": "Cribl-Server", "source": "big_json-from-collector"}	2020-04-23 09:37:41.000	-0:00	Cribl-Server	big_json-from-collector

Cancel Create A Datanode File Save as Sample File

### 3. Create another route or utilize an existing one (Splunk or Elastic)

- From the top menu select Routing / Data Routes
- From the Existing **to\_Splunk** route select **Filter** and change from `__inputId=='datagen:windows_xml'` to `__inputId=='datagen:Big_JSON'`
- Keep the Pipeline as passthru
- Change the default Destination Output to: minio:Big\_Json\_reduced
- Leave all other fields and toggles unchanged
- Click Save

Route Name\* Big\_Json\_to\_S3

Filter ⓘ \_\_inputId=='datagen:Big\_JSON'

Pipeline\* ⓘ passthru

Enable Expression ⓘ No

Output ⓘ minio:Big\_Json\_reduced

Description ⓘ Enter a description

Final ⓘ No

Clones ⓘ

Field Name	Field Value
+ Add Field	
+ Add Clone	

Validate if data is being sent to the S3 bucket.

- Within the Big\_Json\_to\_S3 route configured above, click on the 3 dots to the right of the **On** toggle:



- Select Capture

You should see data going out to the S3 bucket configured

Stream > Destinations > MinIO > Big.Json_reduced						
Configure	Status	Charts	Live Data	Logs	Test	Help
Filter Expression* <input type="text" value="\$_outputId == 'minio:Big.Json_reduced'"/>						
Fields	All	None				
<input checked="" type="checkbox"/> _raw						
<input checked="" type="checkbox"/> _time						
1	0	o [\$_raw: {"resource": "/done", "path": "/done", "httpMethod": "POST", "headers": {"Accept": "application/json, text/javascript, */*; q=0.81", "Accept-Encoding": "gzip, deflate, br", "Accept-Language": "en-US,en;q=0.9"}, "body": "..."}]	2020-04-23 08:... Show more			
			09:37:40.000 # _time: 1587649869 -0400			
2	0	o [\$_raw: {"resource": "/done", "path": "/done", "httpMethod": "POST", "headers": {"Accept": "application/json, text/javascript, */*; q=0.81", "Accept-Encoding": "gzip, deflate, br", "Accept-Language": "en-US,en;q=0.9"}, "body": "..."}]	2020-04-23 08:... Show more			
			09:37:41.000 # _time: 1587649881 -0400			
3	0	o [\$_raw: {"resource": "/done", "path": "/done", "httpMethod": "POST", "headers": {"Accept": "application/json, text/javascript, */*; q=0.81", "Accept-Encoding": "gzip, deflate, br", "Accept-Language": "en-US,en;q=0.9"}, "body": "..."}]	2020-04-23 08:... Show more			
			09:37:41.000 # _time: 1587649881 -0400			
4	0	o [\$_raw: {"resource": "/done", "path": "/done", "httpMethod": "POST", "headers": {"Accept": "application/json, text/javascript, */*; q=0.81", "Accept-Encoding": "gzip, deflate, br", "Accept-Language": "en-US,en;q=0.9"}, "body": "..."}]	2020-04-23 08:... Show more			
			09:37:44.000 # _time: 1587649884 -0400			
5	0	o [\$_raw: {"resource": "/done", "path": "/done", "httpMethod": "POST", "headers": {"Accept": "application/json, text/javascript, */*; q=0.81", "Accept-Encoding": "gzip, deflate, br", "Accept-Language": "en-US,en;q=0.9"}, "body": "..."}]	2020-04-23 08:... Show more			
			09:37:41.000 # _time: 1587649861 -0400			
6	0	o [\$_raw: {"resource": "/done", "path": "/done", "httpMethod": "POST", "headers": {"Accept": "application/json, text/javascript, */*; q=0.81", "Accept-Encoding": "gzip, deflate, br", "Accept-Language": "en-US,en;q=0.9"}, "body": "..."}]	2020-04-23 08:... Show more			
			09:37:44.000 # _time: 1587649884 -0400			
7	0	o [\$_raw: {"resource": "/done", "path": "/done", "httpMethod": "POST", "headers": {"Accept": "application/json, text/javascript, */*; q=0.81", "Accept-Encoding": "gzip, deflate, br", "Accept-Language": "en-US,en;q=0.9"}, "body": "..."}]	2020-04-23 08:... Show more			
			09:37:42.000 # _time: 1587649862 -0400			
8	0	o [\$_raw: {"resource": "/done", "path": "/done", "httpMethod": "POST", "headers": {"Accept": "application/json, text/javascript, */*; q=0.81", "Accept-Encoding": "gzip, deflate, br", "Accept-Language": "en-US,en;q=0.9"}, "body": "..."}]	2020-04-23 08:... Show more			
			09:37:40.000 # _time: 1587649862 -0400			
9	0	o [\$_raw: {"resource": "/done", "path": "/done", "httpMethod": "POST", "headers": {"Accept": "application/json, text/javascript, */*; q=0.81", "Accept-Encoding": "gzip, deflate, br", "Accept-Language": "en-US,en;q=0.9"}, "body": "..."}]	2020-04-23 08:... Show more			
			09:37:41.000 # _time: 1587649881 -0400			

#### 4. Send data from the S3 bucket in Full mode to route

So far, we have configured a new route capable of sending data to a S3 bucket (using MinIO as destination) and collect data from a S3 bucket (using S3 collector as source), we will now collect that data, apply a Pipeline for a Reduction use case and send that data (reduced) to our Splunk Destination.

Let's use the existing `to_Splunk` route and modify its filter, `Pipeline`.

**Expand the to\_Splunk route and enter/modify the following values:**

- Route Name\*: <NO\_CHANGE>
  - Filter: source == 'big-json-from-collector'
  - Pipeline: Reduction
  - Output: <NO\_CHANGE>
  - Description: <NO\_CHANGE>
  - Click Save

Route Name\* `to_Splunk`

Filter `source == 'big-json-from-collector'`

Pipeline\* `Reduction`

Enable Expression `No`

Output `splunk:SplunkCribLab`

Description `First route to Splunk`

Final `No`

Clones `+ Add Field`  
`+ Add Clone`

- Select Data / Sources from the top menu
- Click on the S3 tile
- In the Big\_Json\_Collect source click on Run under the Actions column.
- In the context window select Full Run
- Click Run

Mode\* `Full Run`

Time range `Absolute` `Relative` Earliest `Enter earliest` Latest `Enter latest` Range Timezone `UTC`

Filter `true`

> ADVANCED SETTINGS

Cancel Run

**Note: for this exercise we used the ad-hoc Full Run function, normally S3 Collectors utilize a scheduler to collect and send the results through the available routes that will forward the data to one or many destinations.**

Now let's check Splunk and validate the data is arriving at its destination after being reduced by Cribl Stream.

- From your Jump Box, open a new browser window
- Navigate to: http://10.0.53.55:8000
- Username: admin
- Password: Go2atc4labs!
- From the Apps menu select Search & Reporting
- In the Search field enter the following search: index=big\_json
- At the end of the search command line select the time picker for last 15 minutes

You should see the following results:

Time	Event
4/11/22 9:13:25.879 PM	<pre>{   "body": {"done": "true"},   "headers": {     "Accept": "application/json, text/javascript, */*; q=0.01",     "Accept-Encoding": "gzip, deflate, br",     "Accept-Language": "en-US,en;q=0.9",     "CloudFront-Forwarded-Proto": "https",     "CloudFront-Is-Desktop-Viewer": "true",     "CloudFront-Is-Mobile-Viewer": "false",     "CloudFront-Is-SmartTV-Viewer": "false",     "CloudFront-Is-Tablet-Viewer": "false",     "CloudFront-Viewer-Country": "US",     "Host": "jeff1874.execute-api.us-west-2.amazonaws.com",     "Referer": "https://cdn.foo.com/burger/index.html",     "User-Agent": "Mozilla/5.0 (iPad; CPU iPhone OS 5_0_1 like Mac OS X) AppleWebKit/534.46 (KHTML, like Gecko) Version/5.1 Mobile/9A405 Safari/7534.48.3",     "Via": "2.0 viaf0a6e33ed0f1c1f7fb9e0e3ub.cloudfront.net (CloudFront)",     "X-Amz-CF-Id": "OjwApNkSDFzqgYMEUHfYbQUNLJH1B1tpAJ1Imvov5tW9MbGANG",     "X-Amzn-Trace-Id": "Root=7abe03jRgfsf149Yeoe1HyqoMnJfW9mgX",     "X-Forwarded-For": "27.182.11.11",     "X-Forwarded-Port": 443,     "X-Forwarded-Proto": "https",     "Content-Type": "application/json",     "Origin": "https://cdn.foo.com"   },   "httpMethod": "POST",   "isBase64Encoded": "false",   "path": "/done",   "requestContext": {     "accountId": "12345678901234567890",     "identity": "12345678901234567890",     "stage": "prod"   },   "resource": "/done" }</pre> <p>Show as raw text</p> <p>host = Cribl-Server   source = big-json-from-collector   sourcetype = tcp-cooked</p>

## Go back to Routes in Cribl Stream

- On the **to\_Splunk** route Click on the **On/Off** toggle placing in the **Off** position

## 7. Send data to Elastic

Now that we have sent data to MinIO (S3), replayed that data with our S3 Collector and sent the replay to an Splunk instance, we will now send a different source using Syslog data (previously configured in this Lab) to ElasticSearch and validate the results in Kibana.

- From the top menu in Cribl Stream click on Routing / Data Sources
- Expand the **to\_Elastic** route (previously configured in this lab)

### Enter the following values:

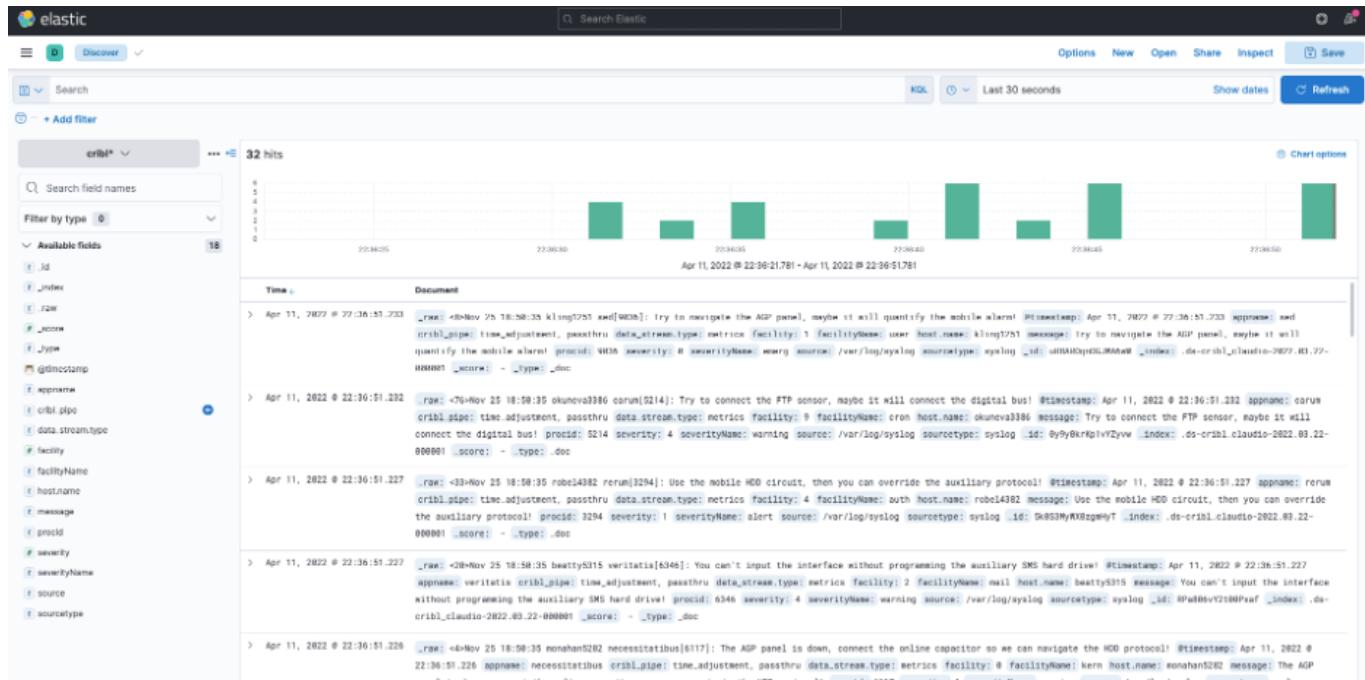
- Route Name\*: <NO\_CHANGE>
- Filter: \_\_inputId=='datagen:syslog'
- Pipeline: passthru
- Output: <NO\_CHANGE>
- Description: <NO\_CHANGE>

- Final: No

Leave all other fields unchanged

- Click Save
- From your Jump Box, open a new browser window
- Navigate to: <http://10.0.53.59:5601>
- No username or password used with Kibana
- In Kibana on the top click on the hamburger menu
- Select Analytics / Discover
- Change the Index Pattern to cribl\* under the + Add filter link
- Click on the Time Picker to the left of the Search command line
- Select Last 30 Minutes
- Click Apply

You should see the Syslog data ingested into ElasticSearch:



- Go back to Routes in Cribl Stream
- On the to\_Elastic route Click on the On/Off toggle placing in the Off position

## Redact

For the Redacting usecase we will use a different data source (Business Events). The data source is already configured as a Data Gen source but we need to capture a sample file intoorder to work in our redaction before we send any data into any system of analysis.

- In Cribl Stream top menu select Data / Sources
- Click on the Datagen tile
- Select the Business\_Events Datagen source
- Click on Live Data
- When the capture ends click on the **Save as Sample File** on the bottom right
- Enter the following values:
- File Name\*: Business\_events.log
- Description: <LEAVE\_BLANK>
- Expiration (hours): delete any number in this field or the sample file will be removed after this value in hours
- Tags: <LEAVE\_BLANK>

The screenshot shows the Cribl Stream interface with the 'Live Data' tab selected for the 'Business\_events' source. The left pane lists several log entries with columns for raw data, timestamp, and \_time. The right pane contains 'SAMPLE FILE SETTINGS' for saving the data as a sample file, including fields for File Name\*, Description, Expiration (hours), and Tags, along with a 'Save' button.

## 1. Use existing sources

### Lets create a new Pipeline and attach it to an existing route

- From the main menu at the top select Processing / Pipelines
- On the top button at the right click on + Pipeline and select Create Pipeline
- Name the Pipeline Redact

## 2. Create a Mask Function In the Redact Pipeline

- From the top click on the + Function and select Mask

### Add the following values to the fields in the Mask Function:

- Filter: true
- Description: <LEAVE\_BLANK>
- Final: No

In the Masking Rules add 2 rules:\*

## Social Security Masking

- Match Regex: (social)=([0-9]{3}-?[0-9]{2}-?[0-9]{4})
- Replace Expression: \${g1}\${C.Mask.md5(g2, g2)}

## Credit Cards Identify and Hash

- Match Regex: ([0-9]{14,16}) Note: Make sure to click on the flags icon and select /g for Global
- Replace Expression: C.Mask.isCC(g1) ? C.Mask.md5(g1, g1.length) : g1
- Apply to Fields: \*
- Click Save

## 3. The Mask Function

The Mask Function has redacted your Social Security and identified a Credit Card number from a common regex pattern distinguishing real Credit Card from regular values and added a hash to both using the same size from the second groups captured in the regex. This function can be used for multiple redacting use cases (I.E.: PII data)

Let's add a Parser Function and extract fields from the Businessevents.log and validate if the **social** and **cardNumber** fields have been redacted/masked as desired. We can validate that by reading from the `_raw` but lets exercise what we've learned in this Lab.

- From the top right in the left panel (in the Redact Pipeline) click the **+ Function**
- Select **Parser**

## Enter the following values:

- Filter: true
- Description: <LEAVE\_BLANK>
- Final: No

- Operation Mode\*: Extract
- Type\*: Key=Value Pairs
- Library: <LEAVE\_BLANK>
- Source Field: \_raw
- Desination Field: <LEAVE\_BLANK>
- Clean Fields: Yes
- List of Fields: <LEAVE\_BLANK>
- Fields to Keep: <LEAVE\_BLANK>
- Fields to Remove: <LEAVE\_BLANK>
- Fields Filter Expression: <LEAVE\_BLANK>
- Click Save

The screenshot shows the Cribl Stream interface with a configuration for a Parser function. The configuration includes:

- Parser ID:** 2
- Parser Name:** Parser
- Enabled:** true
- Filter:** true
- Description:** Enter a description
- Final:** No
- Operation Mode\***: Extract
- Type\***: Key=Value Pairs
- Library:** Select from Library
- Source Field:** \_raw
- Destination Field:** Destination field name
- Clean Fields:** Yes
- List of Fields:** Enter field names
- Fields To Keep:** Enter field names
- Fields To Remove:** Enter field names
- Fields Filter Expression:** Enter field filter expression

#### 4. Validate on the Sample file within Cribl Stream.

- On the right panel (Sample files) make sure the Bussinevents.log is loaded
- Click on the **OUT** at the top left in the Sample Data panel (right)

You should see several fields extracted (Parser function), verify the **social** and **cardNumber** fields have been properly Masked

The screenshot shows the Splunk Redact Pipeline configuration. The pipeline consists of three main stages:

- Stage 1:** A Mask function with a Filter set to true. The description is left blank. The final operation mode is Extract.
- Stage 2:** A Parser function with a Filter set to true. The description is left blank. The final operation mode is Extract.
- Stage 3:** An Eval function with a Filter set to true. The description is left blank. The final operation mode is Extract.

The pipeline is set to process data from the 'IN' source 'Business\_events.log' and output to the 'OUT' destination. The 'Sample Data' pane shows a single event entry:

```

eID:ID:ESP-PO-1C738E128738F;1F7B086F
@time:1581426269.021
accountNumber: 3000101412
accountType: PrePaid
adapterName: UpdateBillingProvQuote
autoBillCycle: T
billingCycle: 28
cardNumber: 833d850d9ea3698
conversationId: E58-845C2939998C03C3:9967DE31:A22798608205:5598
credits: NA
cribl_pipe: Redact
esn: FBF04208212733
Event: UpdateBillingProvQuote
EventConversationID: NA
index: businessevents
JMSCorrelationID: NA
JMSMessageId: 13:ESP-PO-1C738E128738F;1F7B086F
JMSReplyTo: pub-esb-genericasync-response
marketCity: WASHINGTON
marketState: DC
marketZip: 20002
min: 7035426326
mid: NA
methodName: InternalEvent
networkProviderName: Splunkel1]]
orderNumber: 9000000000000013
orderType: ChargeEM
phoneCode: ZIP-64
phoneNumber: iPhone 6 Plus 64 GIG
phoneType: iPhone
planCode: ULTR50
planDescription: Nationwide Prepaid Unlimited Minutes, Unlimited Mobile to Mobile, Unlimited Data
planName: Unlimited Prepaid
planPrice: 50
planType: PrePaid
properties: 
quoteNumber: 64727137
quotePriority: NORMAL
ReplyTo: NA
serviceLane: UpdateBillingProvisioning
social: 94dc54931
timestamp: 1581426269
timeTolLevel: 1
userName: oGuedes

```

## 5. Add an index name and send data to Splunk using a Pipeline

Let's add an index name in this Pipeline and send this data to Splunk (could be Elastic as well)

- In the same Pipeline (Redact) click the + Function button on the top right within the left panel (the Redact Pipeline)
- Select Eval
- Enter the following values:
  - Filter: true
  - Description: <LEAVE\_BLANK>
  - Final: No
  - Evaluate Fields: + Add Field
  - Name: index
  - Value Expression: 'businessevents'
  - Keep Fields: <LEAVE\_BLANK>
  - Remove Fields: <LEAVE\_BLANK>
- Click Save

The screenshot shows the Splunk search builder interface. At the top, there's a search bar with the query "true". Below it, a "Filter" section contains the filter "true". A "Description" section has a placeholder "Enter a description". Under "Evaluate Fields", there's a table with one row: "Name" is "index" and "Value Expression" is "'businessevents'". There are buttons for "+ Add Field", "Keep Fields", "Remove Fields", and "Final" (set to "No").

## Send data to Splunk and validate if the fields and values were indexed correctly for the Redact Pipeline

- From your Jump Box, open a new browser window
- Navigate to: http://10.0.53.55:8000
- Username: admin
- Password: Go2atc4labs!
- From the Apps menu select Search & Reporting
- In the Search field enter the following search: index=businessevents
- At the end of the search command line select the time picker for last 24 hours

You should see the following results (Note the cardNumber and social values are hashed)

The screenshot shows the Splunk search results for the query "index=businessevents". It displays 8 events from 4/22/2021 10:00:00 PM to 4/22/2015 02:00 PM. The results are as follows:

Time	Event
2020-02-11 07:04:33.292	Event [Event<updateBillingProvQuote, timestamp=1581426274, properties={JMSCorrelationID=NA, JMSMessageID=ID:ESP-PD_0CE12502A464:A12B8A7C, orderType=ChangeESN, quotaPriority=NORMAL, conversationID=ESP-107CA430802E45 JMSSenderName=internalEvent, AdapterName=quoteNumber=44738495, ReplyToNA, username=nikolasol, EventConversationID=NA, adv=765826313, accountType=PrePaid, marketCity="COLUMBUS", marketState=GA, marketZip=31997, bill id, phoneNumber="Samsung GALAXY S5", planCode=ULPRE59, planType=PostPaid, planPrice=59, planName="Unlimited Prepaid", planDescription="Nationwide Prepaid Unlimited Minutes, Unlimi te9, networkProviderName=Splunktel}]
2020-02-11 07:04:33.293	Event [Event<updateBillingProvQuote, timestamp=1581426274, properties={JMSCorrelationID=NA, JMSMessageID=ID:ESP-PD_0CE12502A464:A12B8A7C, orderType=ChangeESN, quotaPriority=NORMAL, conversationID=ESP-107CA430802E45 JMSSenderName=internalEvent, AdapterName=quoteNumber=44738495, ReplyToNA, username=nikolasol, EventConversationID=NA, adv=765826313, accountType=PrePaid, marketCity="COLUMBUS", marketState=GA, marketZip=31997, bill id, phoneNumber="Samsung GALAXY S5", planCode=ULPRE59, planType=PostPaid, planPrice=59, planName="Unlimited Prepaid", planDescription="Nationwide Prepaid Unlimited Minutes, Unlimi te9, networkProviderName=Splunktel}]
2020-02-11 07:04:33.294	Event [Event<updateBillingProvQuote, timestamp=1581426274, properties={JMSCorrelationID=NA, JMSMessageID=ID:ESP-PD_0CE12502A464:A12B8A7C, orderType=ChangeESN, quotaPriority=NORMAL, conversationID=ESP-107CA430802E45 JMSSenderName=internalEvent, AdapterName=quoteNumber=44738495, ReplyToNA, username=nikolasol, EventConversationID=NA, adv=765826313, accountType=PrePaid, marketCity="COLUMBUS", marketState=GA, marketZip=31997, bill id, phoneNumber="Samsung GALAXY S5", planCode=ULPRE59, planType=PostPaid, planPrice=59, planName="Unlimited Prepaid", planDescription="Nationwide Prepaid Unlimited Minutes, Unlimi te9, networkProviderName=Splunktel}]
2020-02-11 07:04:33.295	Event [Event<updateBillingProvQuote, timestamp=1581426274, properties={JMSCorrelationID=NA, JMSMessageID=ID:ESP-PD_0CE12502A464:A12B8A7C, orderType=ChangeESN, quotaPriority=NORMAL, conversationID=ESP-107CA430802E45 JMSSenderName=internalEvent, AdapterName=quoteNumber=44738495, ReplyToNA, username=nikolasol, EventConversationID=NA, adv=765826313, accountType=PrePaid, marketCity="COLUMBUS", marketState=GA, marketZip=31997, bill id, phoneNumber="Samsung GALAXY S5", planCode=ULPRE59, planType=PostPaid, planPrice=59, planName="Unlimited Prepaid", planDescription="Nationwide Prepaid Unlimited Minutes, Unlimi te9, networkProviderName=Splunktel}]
2020-02-11 07:04:33.296	Event [Event<updateBillingProvQuote, timestamp=1581426274, properties={JMSCorrelationID=NA, JMSMessageID=ID:ESP-PD_0CE12502A464:A12B8A7C, orderType=ChangeESN, quotaPriority=NORMAL, conversationID=ESP-107CA430802E45 JMSSenderName=internalEvent, AdapterName=quoteNumber=44738495, ReplyToNA, username=nikolasol, EventConversationID=NA, adv=765826313, accountType=PrePaid, marketCity="COLUMBUS", marketState=GA, marketZip=31997, bill id, phoneNumber="Samsung GALAXY S5", planCode=ULPRE59, planType=PostPaid, planPrice=59, planName="Unlimited Prepaid", planDescription="Nationwide Prepaid Unlimited Minutes, Unlimi te9, networkProviderName=Splunktel}]
2020-02-11 07:04:33.297	Event [Event<updateBillingProvQuote, timestamp=1581426274, properties={JMSCorrelationID=NA, JMSMessageID=ID:ESP-PD_0CE12502A464:A12B8A7C, orderType=ChangeESN, quotaPriority=NORMAL, conversationID=ESP-107CA430802E45 JMSSenderName=internalEvent, AdapterName=quoteNumber=44738495, ReplyToNA, username=nikolasol, EventConversationID=NA, adv=765826313, accountType=PrePaid, marketCity="COLUMBUS", marketState=GA, marketZip=31997, bill id, phoneNumber="Samsung GALAXY S5", planCode=ULPRE59, planType=PostPaid, planPrice=59, planName="Unlimited Prepaid", planDescription="Nationwide Prepaid Unlimited Minutes, Unlimi te9, networkProviderName=Splunktel}]
2020-02-11 07:04:33.298	Event [Event<updateBillingProvQuote, timestamp=1581426274, properties={JMSCorrelationID=NA, JMSMessageID=ID:ESP-PD_0CE12502A464:A12B8A7C, orderType=ChangeESN, quotaPriority=NORMAL, conversationID=ESP-107CA430802E45 JMSSenderName=internalEvent, AdapterName=quoteNumber=44738495, ReplyToNA, username=nikolasol, EventConversationID=NA, adv=765826313, accountType=PrePaid, marketCity="COLUMBUS", marketState=GA, marketZip=31997, bill id, phoneNumber="Samsung GALAXY S5", planCode=ULPRE59, planType=PostPaid, planPrice=59, planName="Unlimited Prepaid", planDescription="Nationwide Prepaid Unlimited Minutes, Unlimi te9, networkProviderName=Splunktel}]
2020-02-11 07:04:33.299	Event [Event<updateBillingProvQuote, timestamp=1581426274, properties={JMSCorrelationID=NA, JMSMessageID=ID:ESP-PD_0CE12502A464:A12B8A7C, orderType=ChangeESN, quotaPriority=NORMAL, conversationID=ESP-107CA430802E45 JMSSenderName=internalEvent, AdapterName=quoteNumber=44738495, ReplyToNA, username=nikolasol, EventConversationID=NA, adv=765826313, accountType=PrePaid, marketCity="COLUMBUS", marketState=GA, marketZip=31997, bill id, phoneNumber="Samsung GALAXY S5", planCode=ULPRE59, planType=PostPaid, planPrice=59, planName="Unlimited Prepaid", planDescription="Nationwide Prepaid Unlimited Minutes, Unlimi te9, networkProviderName=Splunktel}]