This lab will give you an introduction to Cribl Stream as well as an opportunity to interact with the system and understanding the power of Observability complemented by Cribl elegance.

What is Stream?

Cribl Stream is an observability pipeline tool living in 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

Streams users 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: [http://10.233.36.248:9000](http://10.233.36.248:9000" \o "http://10.233.36.248:9000/)

Login as Admin with the password of Go2atc4labs!

Part 1

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.

[https://docs.cribl.io/logstream/destinations](https://nam12.safelinks.protection.outlook.com/?url=https%3A%2F%2Fdocs.cribl.io%2Flogstream%2Fdestinations&data=04%7C01%7CSteve.Peck%40wwt.com%7C34de392a86a74a26d48308da12936a84%7Ca2d8e6b4e26e44218f3dec288c827c7d%7C1%7C0%7C637842721892965094%7CUnknown%7CTWFpbGZsb3d8eyJWIjoiMC4wLjAwMDAiLCJQIjoiV2luMzIiLCJBTiI6Ik1haWwiLCJXVCI6Mn0%3D%7C3000&sdata=mJpQ64nQr5GD1i6lRP6kV1iQgefwn5J0%2BDexKh0%2FAGw%3D&reserved=0" \o "https://nam12.safelinks.protection.outlook.com/?url=https%3A%2F%2Fdocs.cribl.io%2Flogstream%2Fdestinations&data=04%7C01%7CSteve.Peck%40wwt.com%7C34de392a86a74a26d48308da12936a84%7Ca2d8e6b4e26e44218f3dec288c827c7d%7C1%7C0%7C637842721892965094%7CUnknown%7CTWFpbGZsb3d8eyJWIjoiMC4wLjAwMDAiLCJQIjoiV2luMzIiLCJBTiI6Ik1haWwiLCJXVCI6Mn0%3D%7C3000&sdata=mJpQ64nQr5GD1i6lRP6kV1iQgefwn5J0%2BDexKh0%2FAGw%3D&reserved=0)

Task 01

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.

A screenshot of a computer

Description automatically generated with medium confidence

Select Push > Syslog

Your sources are listed on the left and under Manage Syslog Sources you may configure several Syslog sources. These sources wil act as a Syslog server recieing data from any syslog capable device sending data to any existing Syslog server (SyslogNG, Rsylog, 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.

Graphical user interface, application

Description automatically generated

Task 02

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.

Graphical user interface, application, Teams

Description automatically generated

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 storage on Cribl Stream.

Task 03

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)

Graphical user interface, application, Teams

Description automatically generated

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)

Graphical user interface, text, application, Teams

Description automatically generated

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](http://192.168.2.52:9000" \o "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

Graphical user interface, application

Description automatically generated with medium confidence

You will be back on the Destinations list. Wait until the Status becomes green and 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 want to read data from a 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

Application

Description automatically generated with medium confidence

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 chose how to process the data in this connected route. Click on Passthru (sending raw data to the destination, no pipelines applied) click save.

Graphical user interface, application

Description automatically generated

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 (SplunkCriblLab) 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.

Graphical user interface, text, application, email

Description automatically generated

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.

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 2

Now we have successfully sent data via QuickConnect lets use Routes to send data to the same destinations.

First lets 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 Managed 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.

Cick on the top X on the opened context window.

1. Create a new Route

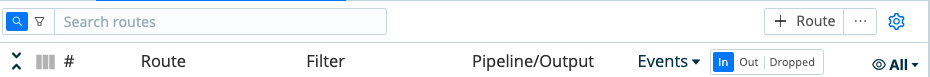
We’ll create 2 new routes for this part of our lab.

We need to connect existing sources (using datagen) and capture sample data from each route to be used on our future “Pipelining” process.

These sources are: syslog and windows\_xml

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.



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   default:default

Description: First route to Splunk

Final: No

Graphical user interface, text, application, email

Description automatically generated

Click Save

1. 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 most left 3 dots and select Capture

Graphical user interface, text, application, email

Description automatically generated

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

Table

Description automatically generated with medium confidence

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 an validate your sample data from your Route.

Graphical user interface, text, application, email

Description automatically generated

1. 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’

Graphical user interface, application

Description automatically generated

One the left pane enter only the ID as Syslog\_test and click Save.

      Add functions to the pipeline

Click on the ‘+ Function’ button on the right most side within the newly create Pipeline

Graphical user interface, application

Description automatically generated

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 load on the right pane and validate the results by clicking on the OUT button on the top bar within the Sample data loaded.

Graphical user interface, text, application, email

Description automatically generated

Now that we have excluded the field message from the processing logs, lets use another select and use another Function “Drop” and 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’

Graphical user interface, text, application, email

Description automatically generated

Note the events matching the filter in the Drop function are greyed out and not send to the destination 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

Graphical user interface, application

Description automatically generated

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

5.      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

    Apply the destination to the Route

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

From Output select splunk:SplunkCriblLab

In Description enter: “Sending Syslog data to Splunk”

Final toggle set to No

Click Save

Graphical user interface, text, application, email

Description automatically generated

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

Part 3

Cribl Stream Use Cases

Reduction

1.       Use the Json reduction elements

2.       Send to routes

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

Routing

1.       Create a S3 destination

2.       Send data to a S3 bucket (WWT will use an internal MinIO)

3.       Check if data has arrived on the S3 bucket.

Replay

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

2.       Replay data as preview within Cribl Stream

3.       Validate event breakers

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

5.       Apply the selected destination to the created route

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

7.       Analyze data in Splunk or Elastic

Redact

1.       Use existing sources

2.       Create a function (Mask) on the already created pipeline

3.       Change/mask/hash value in one field (TBD)

4.       Validate on the Sample file within Cribl Stream.