Crawly - User Guide

Published on April 30, 2020

Originally written by Suyash Srivastava

Introduction

- ○ Introduction
  - ○ Use Cases
  - ○ Updates
  - ○ Workflow
    - Crawling by Akamai IP and f/r log lines
    - Crawling using f/r log lines New
    - Crawling by Reference Number
    - Crawling by Reference Number whose Request Id = 0 New
    - Crawling by Response Headers
    - Crawling by Cloudmon ID
    - Crawling by giving C2S and NS4 (NSDS_UL/NSOS_UL) LogLine New
    - Crawling by giving input in query-param New
    - Results Page
    - Results Page
    - Information about IP Tool, Edgescape and Region360
    - Other Information New
    - Recrawl
  - ○ Reference

Crawly is the new and improved version of the existing log crawler. It traces the complete request recursively till the last hop and

provides the entire request logs to the engineers within few seconds. The output is then parsed displaying the significant fields that a TSE usually looks for in the course of troubleshooting.

Crawly is undergoing continuous iterations for improvements and enhancements. The CHANGELOG - Crawly contains all the details of every iteration in a chronological order which is aimed to help users track any modification / enhancement made in the tool.

Use Cases

Prominent features of Crawly are as follows:

- Performs the crawling of entire request log like the existing crawler by accepting an Akamai IP and a f/r/S log line.
- Accepts various other inputs like reference number, response header, cloudmon ID and NS4 log lines(NSOS_UL, NSDS_UL, C2S) to fetch the logs for any request.
- Parses the logs and displays the commonly used fields in a table.
- Provides raw log lines as well.
- Forwards the log lines to Log parser.
- Fetches media logs for VOD Translator, SMTCORE, Archiver and SMM
- Provides end to end logs for any logline entered. Can provide a log-line input from parent as well. It fetches the logs in both the direction.
- Provides the functionality to provide input through a query string.
- Fetches the Image Manager Pristine logs of image manager.
- Fetch nsds_ul / nsos_ul logs from c2s loglines.
- Provides shareable link for sharing the crawly result.
- Fetches "dbg_asis.log" / "dbg_asis.log.old.*" logs to assist with ESI troubleshooting
- Fetches ghost.ddc-local.log also.
- Provides you with the visualization for the Ghost logs and Image Manager logs.

Updates

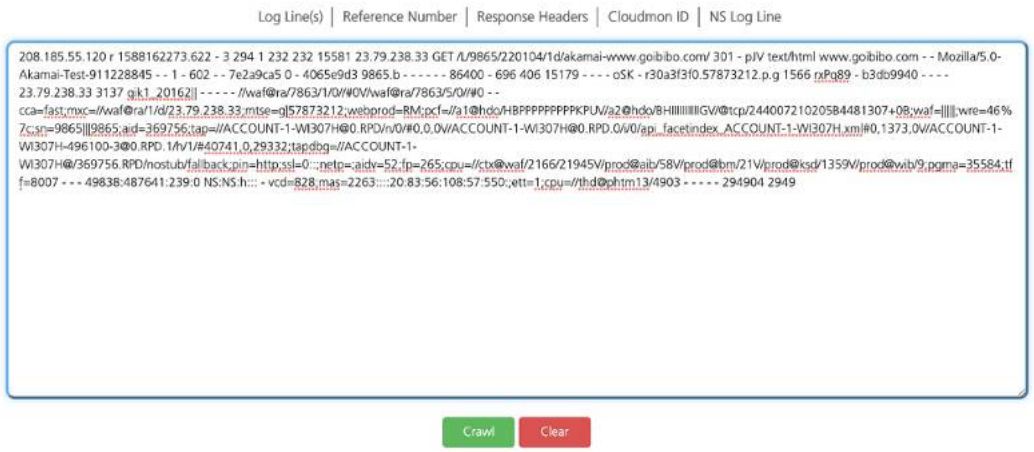
All the updates can be seen in the Change Log.

Workflow

Crawly 4.x automatically detects the type of input mentioned. The user doesn't have to pre-select an option before giving a request.

Crawling by Akamai IP and f/r log lines

User can provide an Akamai IP followed by the log line. Note that the IP and the log lines are separated by a single space as shown below



Crawling using f/r log lines New

User can also fetch log lines by directly providing r/f log line.

Note:

- Can enter only the logline without leading ghost IP. r line of the first hop will not be accepted.
- if a logline from the middle of the hop is mentioned, crawly will fetch the entire trace's logs

Crawling by Reference Number

The log can be crawled for a given reference number as well.

Crawling by Reference Number whose Request Id = 0 New

User can also fetch logs for the reference number whose request_id is 0.

Crawling by Response Headers

User can also fetch log lines by providing response headers as well. User can paste all response headers* as follows:

Date e.g., Date: Tue, 01 Nov 2016 12:26:36 GMT

...

...
X-Akamai-Request-ID e.g., X-Akamai-Request-ID: 2ca28bbf.725ff9f
X-Cache e.g, X-Cache: TCP_MISS from a72-247-150-199.deploy.akamaitechnologies.com (AkamaiGHost/8.1.4.1.2-18493594)
X-Cache-REMOTE(optional) e.g., X-Cache-Remote: TCP_MISS from a204-2-187-37.deploy.akamaitechnologies.com (AkamaiGHost/8.1.4.1.2-18493594)

...

...

---

*Note: Date, X-Akamai-Request-ID and X-Cache headers are mandatory. X-Cache-REMOTE is an optional one. There is no fixed order of occurrence for response headers as input.

Crawling by Cloudmon ID

User can also fetch log lines by providing a Cloudmon ID. Crawly currently accepts only 24 character Cloudmon IDs. Any Cloudmon ID with less than 24 characters has the hex part of the IP truncated to drop the 0 leading it not possible to reverse calculate the true IP. This is a limitation in Cloudmonitor and not something Crawly can address.

Crawling by giving C2S and NS4 (NSDS_UL/NSOS_UL) LogLine New

User can also provide log lines for C2S/NSOS_UL/NSDS_UL log lines. The service types acceptable are - ["cobrascp", "cobraftp", "cobrasftp", "cobrasync", "cobracms", "cmsp_asp", "nsos_ul", "nsds_ul"].

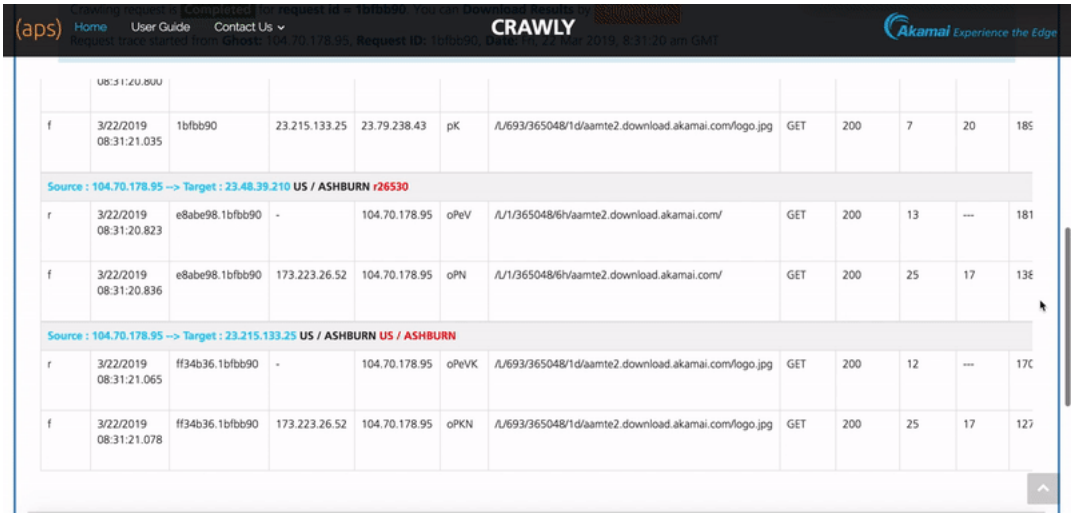
Crawling by giving input in query-param New

User can also provide the input with the query param. However, the input type should be the one that is supported by crawly and it needs to be URL encoded to pass the validation.

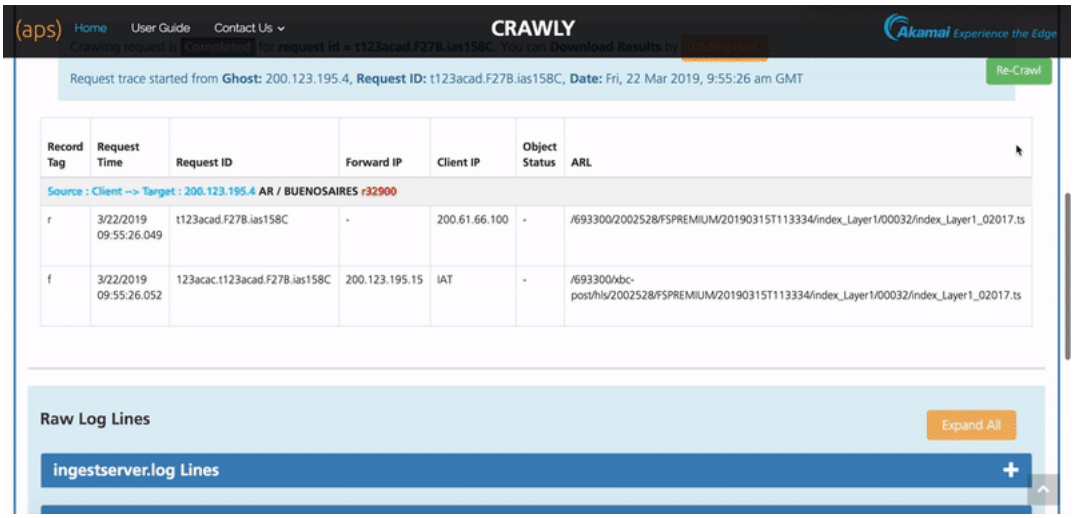
The syntax for it: https://crawly-aps.akamai.com/#!/?crawl-input={LOGLINE/REFCODE/HEADERS/CLOUDMON_ID/NS_LOGLINE}

Results Page

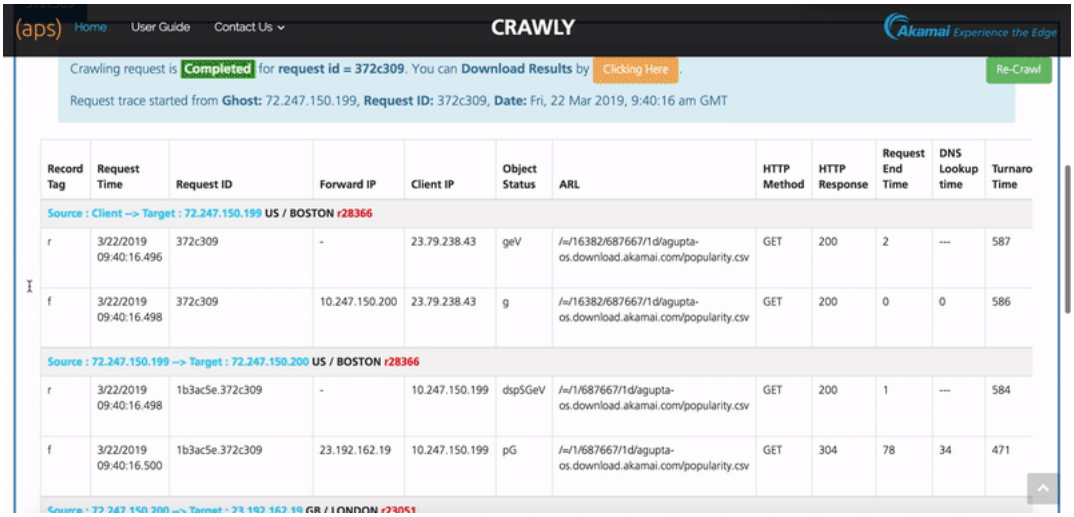
The results page consists of raw log lines and table containing commonly used fields. We can also more fields that we want to display in the table using the filter on the top.



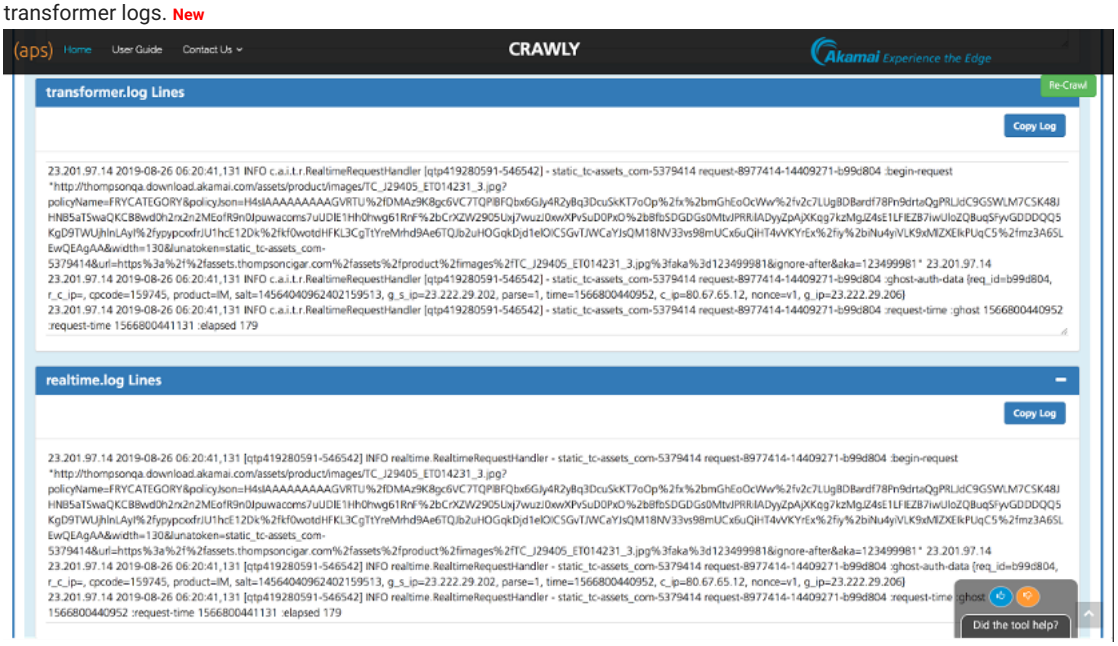
Crawly fetches **MSL4** (Ingest Server and XBC) logs for a given log line input. It also fetches 'ingestserver.log', 'smtcore.access.log', 'smtcore.client.log' ,
'smtcore.trace.log' and 'xbc hotbackup' for any backup machines. New



Crawly fetches **NSOS** and **NSDS logs**.

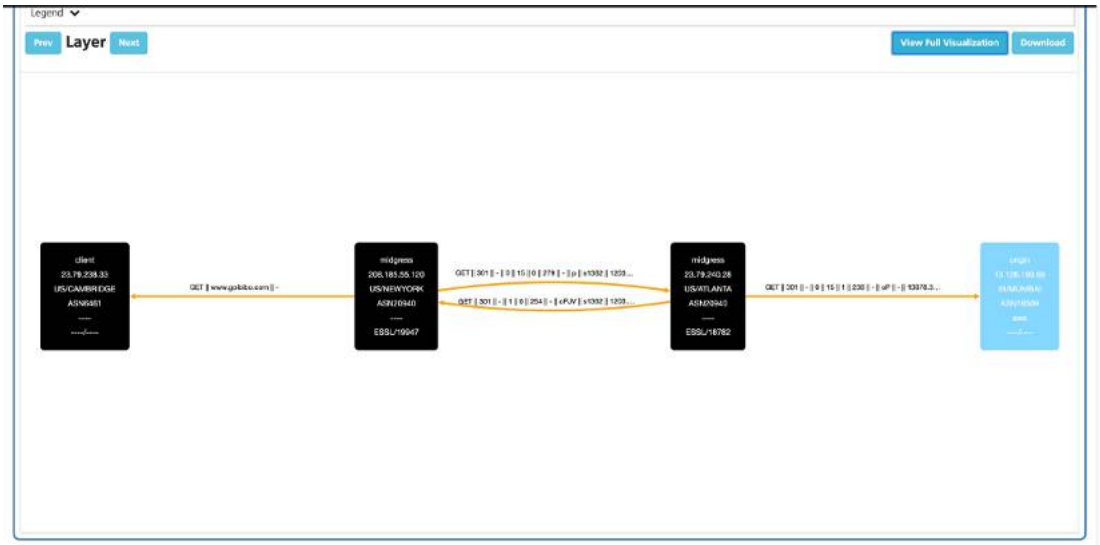


Crawly fetches **transformer.log** and **realtime.log** of **Image Manager**. It can also fetch Image Pristine Manager logs for transformer logs. **New**



Crawly fetches **NetStorage** and **C2S** Logs. **New**

Crawly also draws visualization for **Ghost** and **ImageManager** logs. **New**



1. Features
    1. Lines are color coded as per their status code (500, 400 etc)
    2. Right-clicking on a tile allows
        1. Download logs for that tile (IP)
        2. Ability to view tile information in greater detail in Netarch and Region360
        3. Hovering over lines gives a more verbose representation of the data present. Distance is color coded with distance exceeding 500km colored in red
        4. Tiles are placed vertically
            1. If they belong to the same hop
            2. If they are siblings ie object_status in f line contains g
    3. Tiles are colored based on their provider

4. Can be viewed in layers

Results Page

Information about IP Tool, Edgescape and Region360

Clicking on the IP after '**Region360 link**' will take you to Region360.

Other Information New

You can also download the results for later use using the 'Download Raw Logs' button.

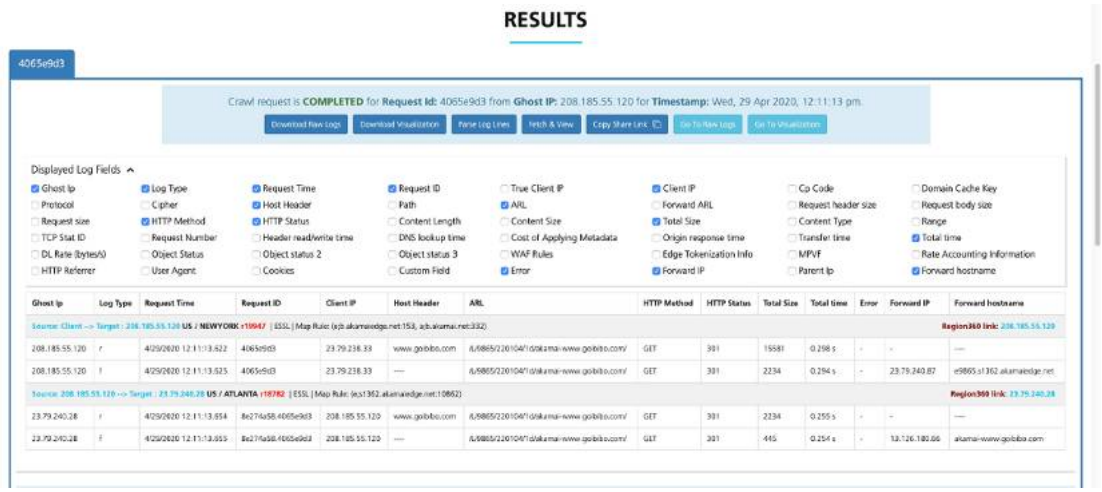
You can also download the visualization using the 'Download Visualization' button.

You can parse all the log lines at once using the 'Parse Log Lines' button.

You can get the shareable link by clicking on the 'Copy Share Link' button.

You can add more fields to view in the table by clicking on the fields.

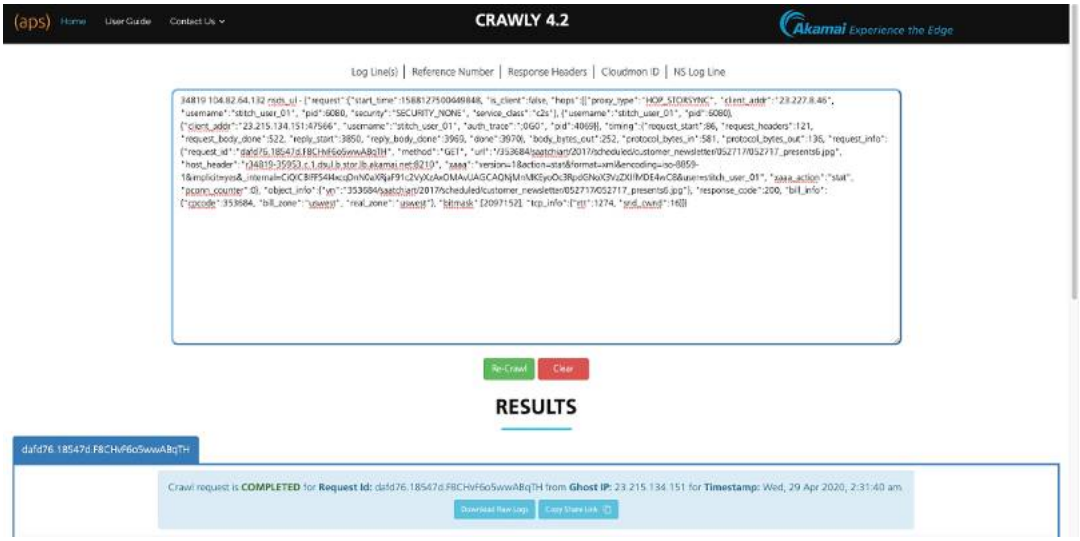
Hovering over the object_status cell provides additional information regarding each character in the object status



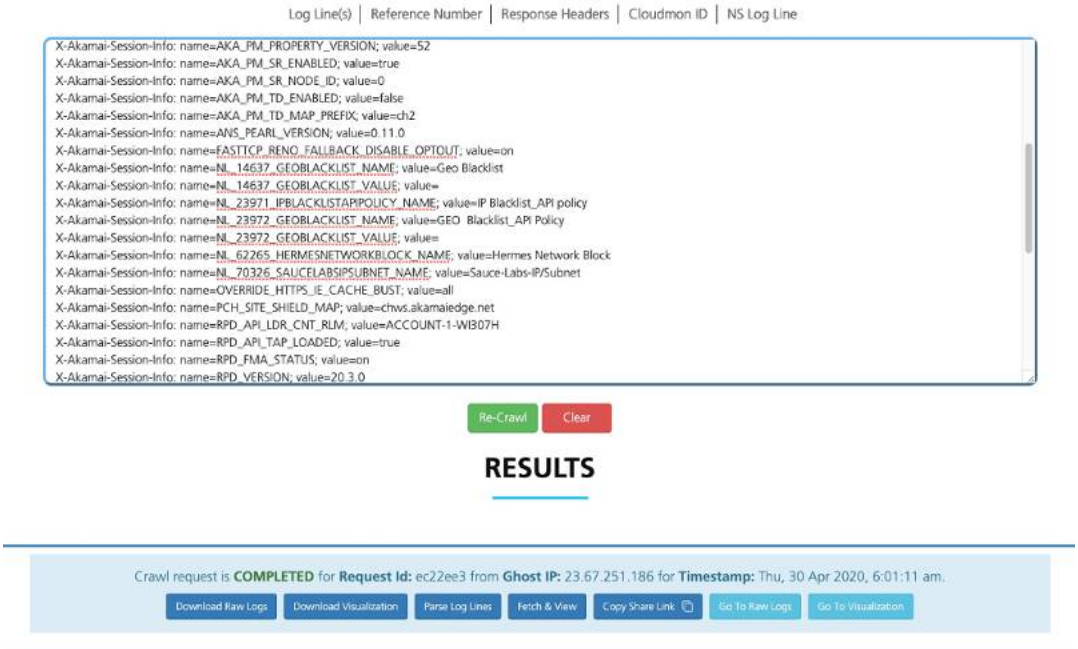
Recrawl

You can now retry recrawl for the input that has been already been crawled and has the same request-id, ip and timestamp.

For NS4 inputs:



For Other Inputs (Log Line/Request Header/Cloudmon Id/Reference num):



Reference

https://tools.aps.akamai.com/crawly/

Comments

B

Add a comment...

Comment

MOST RECENT

V

Victor Velechovsky

Hi Crawly team, is there an API available that I can use to get ddc logs for a request programatically (by passing in the response headers from the request)?

In other words is the Crawly API meant to be interacted with directly and are there docs for it?

Translate into English

June 17, 2025

1 Like Reply

John Payne

When this tool works, it works great

Looking in the console when this happens, it seems that https://crawly-aps.akamai.com/api/crawly/crawl is giving a 400 bad request.

Translate into English

February 13, 2019

0 Like Reply

Niranjan Balasubramani

Hello John,

Kindly clear your cache and retry again.

Translate into English

February 13, 2019

0 Like Reply

M

_lumapps_mig Mig

[redacted]

Show more

Translate into English

November 14, 2018

0 Like Reply

_lumapps_mig Mig

Originally written by Salem Artin

Thanks for your reply. Is it possible that I/we/someone host Crawly within the lab/locally and do a bit of tweaking so that it works?

Translate into English

November 14, 2018

0 Like Reply

Scott Lerner

We charge by the line, so it could be expensive :-).

You can work with **Ravi Ranjan**, our engineering manager to explore code access. I don't think that will be an issue to grant you the ability to pull.

Translate into English

November 14, 2018

0 Like Reply

_lumapps_mig Mig

[redacted]

Show more

Translate into English

November 14, 2018

0 Like Reply

_lumapps_mig Mig

[redacted]

Show more

Translate into English

November 14, 2018

0 Like Reply

Eli Boxx

[redacted]

Show more

Translate into English

November 14, 2018

0 Like Reply

Display 1 more reply

E

Eli Boxx

[redacted]

Show more

Translate into English

August 9, 2017

0 Like Reply

M

_lumapps_mig Mig

[redacted]

Show more

Translate into English

August 10, 2017

0 Like Reply

Display 3 more replies

E

Eli Boxx

[redacted]

Show more

Translate into English

June 26, 2017

0 Like Reply

R

Ravi Ranjan

Thank you **Eli Boxx** for the feedback. We will get this added in the next release of Crawly (yet to schedule the release).

Translate into English

June 27, 2017

0 Like Reply

M

_lumapps_mig Mig

[redacted]

Show more

Translate into English

June 14, 2017

0 Like Reply

Francisco Vargas

I'm so happy to read this kind of feedback! Keep the awesome work up APS!

Scott Lerner, any way you are measuring the impact of APS tools in terms of productivity? I heard Tejinder Singh saying that you actually collect data of how much time is spent on each tool which is awesome to measure the ROI.

Translate into English

June 14, 2017

0 Like Reply

Display 4 more replies

M

_lumapps_mig Mig

[redacted]

Show more

Translate into English

April 3, 2017

0 Like Reply

M

_lumapps_mig Mig

[redacted]

Show more

Translate into English

April 3, 2017

0 Like Reply

Show more