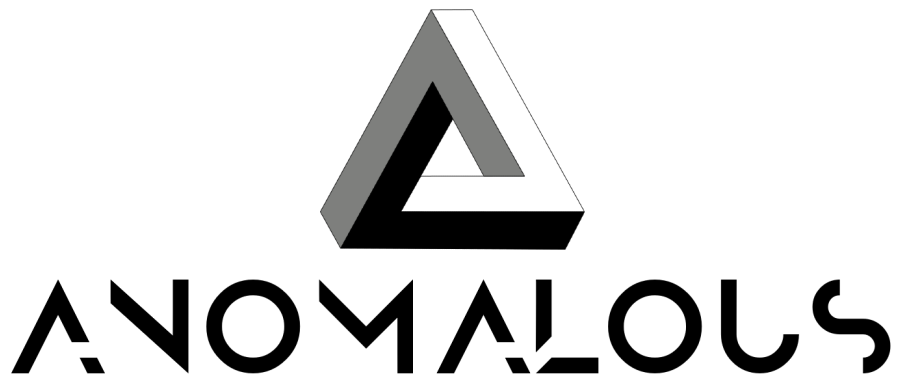


User Manual

Anomaly Detection of APT 29 Dataset using Elastic Cloud SaaS

Produced for CISA by Team Anomalous



Team Anomalous:

Wyatt Orion Child, Cameron Lawrence, John McNary Leach, Sanjeet Kumar Bagga, Kevin Liu,
Joseph Walter

Table of Contents

Logstash	2
Discover	3
Anomaly Detection	4
Data Frame Analytics	10
Kibana Dashboard	13
Credentials	16

Logstash

This section is based on a Windows operating system, but the basic structure of running Logstash should be the same. After using our [How to Configure Logstash](#) guide to set up Logstash, open Command Prompt (you don't need to be administrator). Then navigate to the Logstash folder and the bin folder within that. Once there you can run Logstash. The basic run command is "logstash". Since the target file or folder is set in the Logstash configuration file running just the basic run command will start Logstash with the first config file it can find. Logstash under our custom configurations will run until it is finished then exit. To change the input plugins refer to this [table](#). To change the output plugins refer to this [table](#). You can also add data filtering to your Logstash configuration, the available filter plugins can be found in this [table](#).

Command flags can be used to augment the Logstash run command. A full list of command flags can be found [here](#). However the most important is the "-f CONFIG_PATH" flag. This will direct Logstash to a specific config file at the CONFIG_PATH location. This will allow quick switching between multiple Logstash configs, which would allow you to for example ingest to different indices for different projects easily.

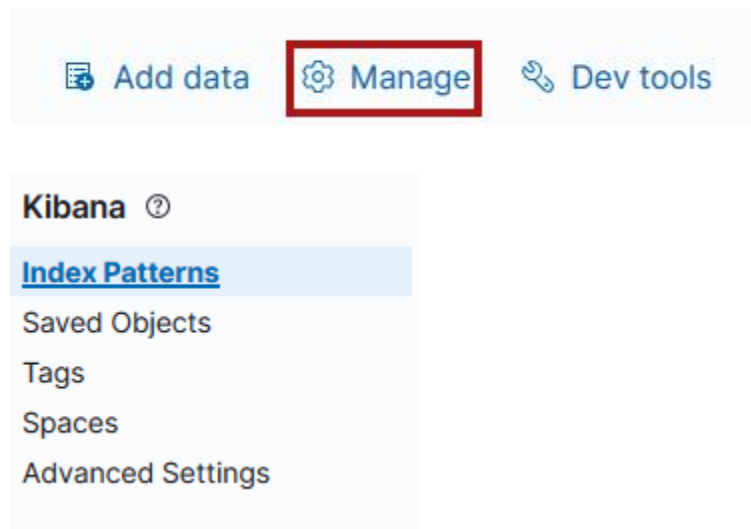
To change the target file in the config change line 6 in logstash.conf

```
6      path => "PATH/TO/LOG/FILES"
```

To change the output index pattern change the name from "logstashindex" on line 18 of logstash.conf to whatever index you want.

```
18      index => "logstashindex"
```

The final step to get the data to be usable is to go to stack management in Elastic and create an index pattern using your indices.




If you need more help going through and setting up the index pattern refer to the [How to Configure Logstash Guide](#) included in this package.

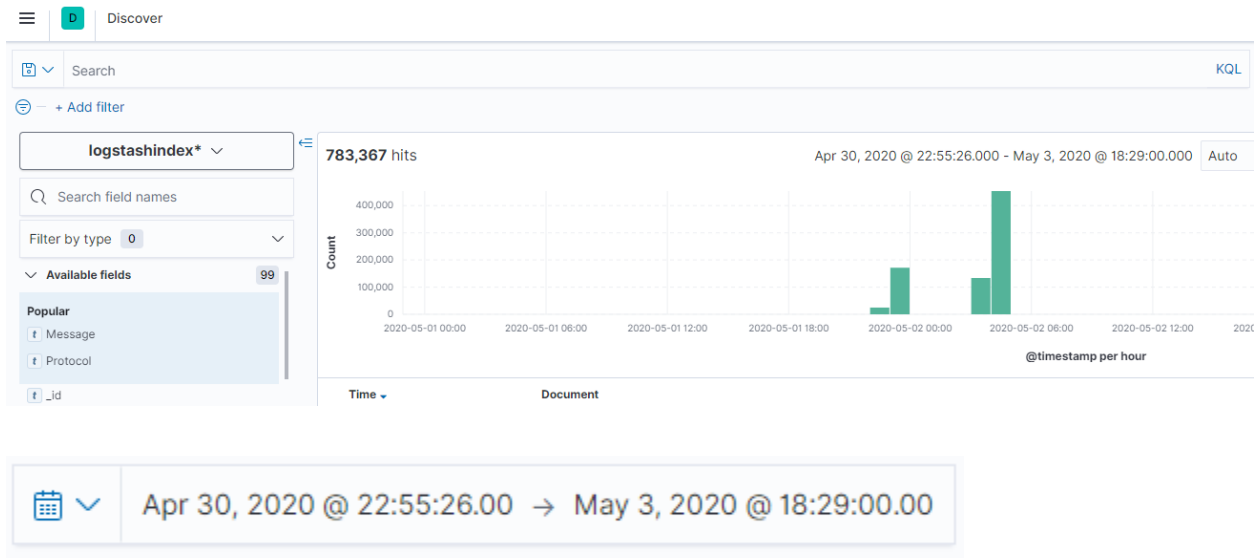
Once you have done all these steps the index pattern you created will be available to use across Elastic.

Discover

To navigate to Discover:

- 1) Login to Elastic environment using provided credentials
- 2) Click the  icon in the top-left corner of the page
- 3) Under the “Analytics” section, click “Discover”

There you will find the Discover module. Make sure that you have the correct index selected on the top-left, below the search and filter bars. You will also probably have to update the specified start and end times to correspond with the timeframe that you are looking for in your chosen index.




On the left side of the module, you will find a list of fields that are contained within the log messages. At the top of the screen, you can use KQL (Kibana's Query Language) to filter the log messages by a chosen value of a specific field (for example, "Application : *" to filter for all messages where the Application field exists.)

Discover is useful for getting an overview of the fields present in the given dataset, as well as for finding specific instances of anomalies within the dataset.

Anomaly Detection

To navigate to the Anomaly Explorer:

- 1) Login to Elastic environment using provided credentials
- 2) Click the  icon in the top-left corner of the page
- 3) Under the "Analytics" section, click "Machine Learning"

You will be introduced to the Anomaly Detection and Analytics page.

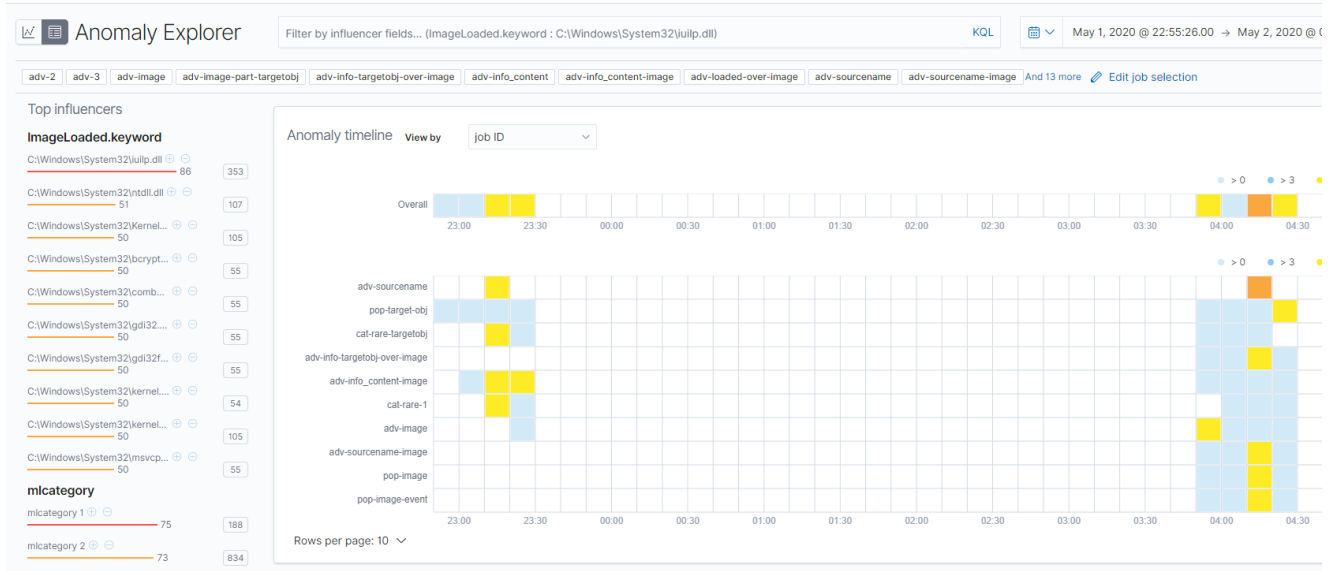
Anomaly Detection						Active ML nodes: 0 Total jobs: 23 Open jobs: 0 Closed jobs: 23 Active datafeeds: 0	
Group ID ↑	Max anomaly score ^①	Jobs in group	Latest timestamp	Docs processed	Actions		
dhs-cisa	50	23	May 2nd 2020, 04:29:23	14,101,966	View		
Rows per page: 10 ▾						< 1 >	
						Refresh Manage jobs	

Analytics						Total analytics jobs: 4 Running: 0 Stopped: 4	
ID ↑	Type	Status	Progress	Creation time	Actions		
dfa-1	outlier_detection	stopped	Phase 4/4 <div></div>	April 3rd 2021, 16:24:14	View		
dfa-many-queries	outlier_detection	stopped	Phase 3/4 <div></div>	April 8th 2021, 21:46:56	View		
dfa-psexec	outlier_detection	stopped	Phase 4/4 <div></div>	April 16th 2021, 11:09:31	View		
dfa-sourcename-powershell	outlier_detection	stopped	Phase 4/4 <div></div>	April 23rd 2021, 13:52:44	View		
Rows per page: 10 ▾						< 1 >	
						Refresh Manage jobs	

The Anomaly Detection section contains an aggregate score of all machine learning jobs created for this project. This section will provide information on all jobs that were created for this module. Clicking “Manage Jobs” will provide a list of all available jobs.

<input type="checkbox"/>	>	adv-2	dhs-cisa	783,367	hard_limit	closed	stopped	2020-05-02 04:29:23			...
<input type="checkbox"/>	>	adv-3	dhs-cisa	783,367	hard_limit	closed	stopped	2020-05-02 04:29:23			...
<input type="checkbox"/>	>	adv-image	dhs-cisa	783,367	ok	closed	stopped	2020-05-02 04:29:23			...
<input type="checkbox"/>	>	adv-image-part-targetobj	dhs-cisa	783,367	hard_limit	closed	stopped	2020-05-02 04:29:23			...
<input type="checkbox"/>	>	adv-info-targetobj-over-image	dhs-cisa	783,367	ok	closed	stopped	2020-05-02 04:29:23			...
<input type="checkbox"/>	>	adv-info_content	dhs-cisa	783,367	ok	closed	stopped	2020-05-02 04:29:23			...
<input type="checkbox"/>	>	adv-info_content-image	dhs-cisa	783,367	ok	closed	stopped	2020-05-02 04:29:23			...
<input type="checkbox"/>	>	adv-loaded-over-image	dhs-cisa	783,367	ok	closed	stopped	2020-05-02 04:29:23			...
<input type="checkbox"/>	>	adv-sourcename	dhs-cisa	783,367	ok	closed	stopped	2020-05-02 04:29:23			...
<input type="checkbox"/>	>	adv-sourcename-image	dhs-cisa	783,367	ok	closed	stopped	2020-05-02 04:29:23			...

Clicking the “View” button () will take you to the anomaly explorer of that job(s).

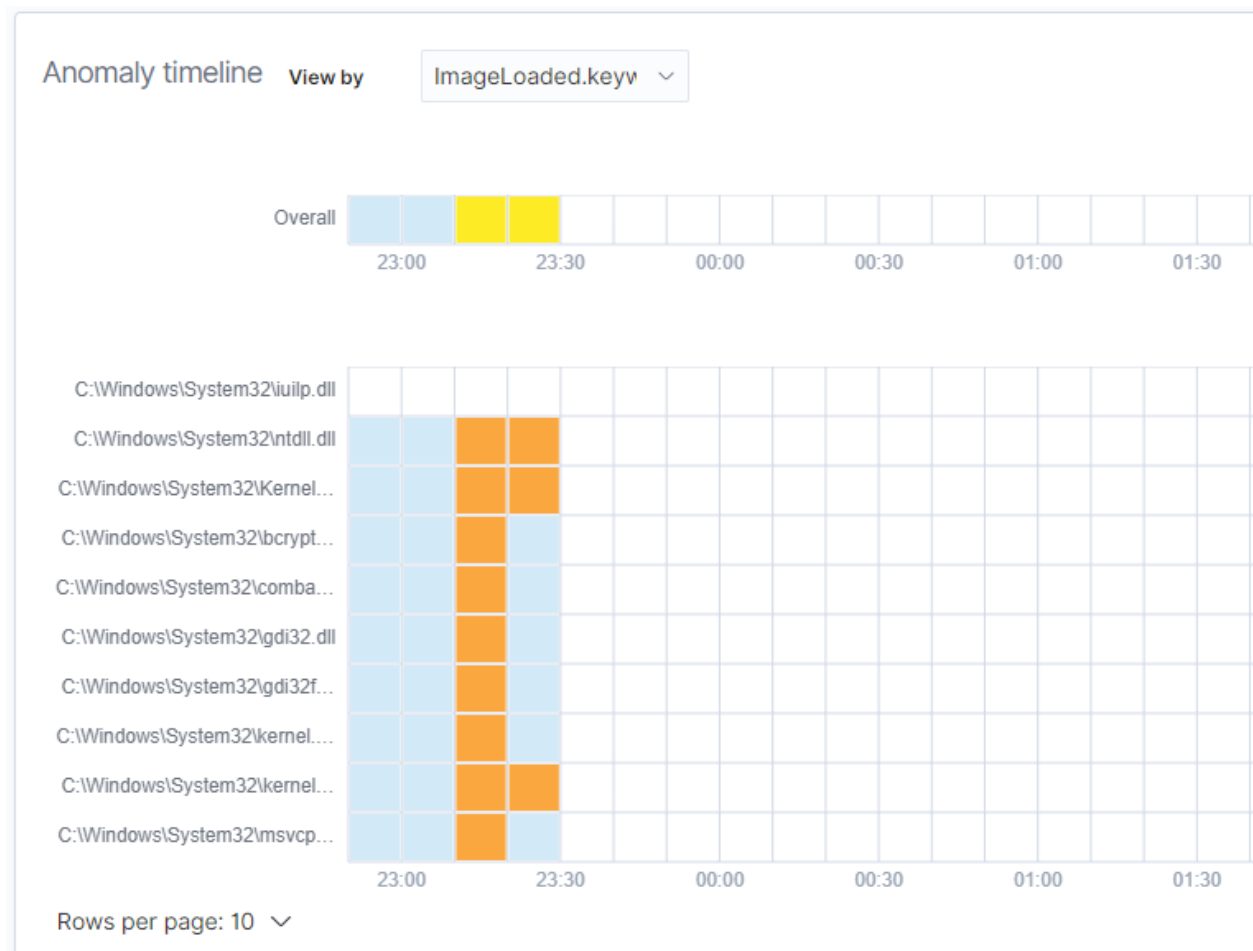


The Anomaly Explorer allows you to visualize all of the events that are found to be “anomalistic” by our machine learning jobs. The anomaly score is calculated by the detector field and the bucket span and is further modified by the influencer fields if they are present.



In the Anomaly Explorer, the “Top Influencers” panel will show you the values for the queried fields where the anomaly score is highest, sorted by the particular field’s score. If the job is an advanced job, the “Top Influencers” column will contain multiple fields and will show the effect they had on the data in other features.

Hovering around the specific influencer discovered will provide the full name of the data point as well as the anomaly scores. The “+” button will allow you to filter for that specific feature, and the “-” will remove the feature from being queried in the results output.



The anomaly timeline describes when anomalies in a specific feature were discovered in the dataset. Hovering around the colored boxes will show the name of the data point it discovered as anomalistic, as well as the anomaly score. Clicking it will filter out subsequent data for that specific metric. Clicking the “View by” dropdown menu will allow you to change what data is being shown on the timeline. If you are using the Anomaly Explorer on a group of jobs, you will also be able to list which job has the most anomalistic data.

Severity threshold		Interval								
<div><div>warning</div><div></div></div>		<div><div>Auto</div><div></div></div>								
	time	severity ↓	detector	found for	influenced by	actual	typical	description	job ID	category examples
>	May 2nd 2020, 03:00	93	count by "Image.keyword" partitionfield="TargetObject.keyword"	C:\windows\system32\svchost.exe	Image.keyword: C:\windows\system32\svchost.exe <div><div></div><div></div></div> TargetObject.keyword: <div><div></div><div></div></div>	430	0.243	<div>↑</div> More than 100x higher	adv-image-part-targetobj	
>	May 2nd 2020, 04:00	86	count by "Image.keyword" over "ImageLoaded.keyword"	C:\Windows\System32\lulpl.dll	Image.keyword: C:\Windows\System32\svchost.exe <div><div></div><div></div></div> ImageLoaded.keyword: C:\Windows\System32\lulpl.dll <div><div></div><div></div></div>	111	1.41	<div>↑</div> 79x higher	adv-image	
>	May 2nd 2020, 04:00	79	count by mcategory	mcategory 1	mcategory: 1 <div><div></div><div></div><div></div></div>	2754	2.24	<div>↑</div> More than 100x higher	categorization-image	C:\Windows\System32\conhos... C:\Windows\System32\dns.exe C:\Windows\System32\lsass.exe C:\Windows\System32\svchos...
>	May 2nd 2020, 04:00	79	count by mcategory	mcategory 1	mcategory: 1 <div><div></div><div></div><div></div></div>	2754	2.24	<div>↑</div> More than 100x higher	categorization-image-2	C:\Windows\System32\conhos... C:\Windows\System32\dns.exe C:\Windows\System32\lsass.exe C:\Windows\System32\svchos...
>	May 2nd 2020, 03:00	76	count by mcategory	mcategory 2	mcategory: 2 <div><div></div><div></div><div></div></div>	2182	0.578	<div>↑</div> More than 100x higher	categorization-image	C:\windows\System32\svchost...

Anomalies

Severity threshold

warning

Interval

Auto

time	severity ↓	detector	found for	influenced by	actual
<div> <div></div> May 1st 2020, 23:00 </div>	<div> < 1 </div>	info_content("Image.key word") over "ImageLoaded.keyword"	C:\Windows\System32\ msasn1.dll	Image.keyword: C:\Windows\PSEXESVC. exe <div></div> <div></div> ImageLoaded.keyword: C:\Windows\System32\ msasn1.dll <div></div> <div></div>	31

Description

warning anomaly in info_content("Image.keyword") over "ImageLoaded.keyword" found for ImageLoaded.keyword C:\Windows\System32\msasn1.dll

Details on highest severity anomaly

ImageLoaded.keyword

C:\Windows\System32\msasn1.dll

time

May 1st 2020, 23:11:30 to May 1st 2020, 23:11:45

function

info_content

fieldName

Image.keyword

actual

31

typical

51.6

job ID

adv-info_content-image

probability

0.027188364145061036

Influencers

Image.keyword

C:\Windows\PSEXESVC.exe

ImageLoaded.keyword

C:\Windows\System32\msasn1.dll

Clicking the “Dropdown” arrow on each anomalistic instance will show more details of the data itself. This data is limited, and it will not provide more context as to what was happening during the session. We recommend the usage of this feature to get the timestamp to cross-reference with the Data Frame Analytics jobs demonstrated below and/or the Discover module demonstrated above.

Note: Anomaly Detection will not be able to find anomalous instances of every feature identified. For example, it was difficult to find anomalistic instances of python.exe being executed. Instead, being able to determine supplemental services can also be useful in determining anomalistic data. For example, webclnt.dll has a couple of minor anomalies discovered but svchost.exe also manages web client services on Windows-based systems, so determining anomalistic data based on that service is also very useful.

A screenshot of a search bar with a light blue border and a light gray background. The text inside the search bar is "Image.keyword : *PSEXESVC*" in a dark gray font.

For this dataset, there may be a need to filter for specific influencer data and data instances. By using the Kibana Query Language (KQL), a query can be created and applied within the Anomaly Explorer. To run the query, click the “Filter by” search bar that is located near the “Anomaly Explorer” title.


Fields that should be prioritized in filtering are those that deal with running processes/services (Image, ImageLoaded, TargetObject, Application.) Features that should be prioritized in filtering include files that can allow a change in security policies (gpi, gpd, lsass, lsarpc), web client services and processes (webclnt.dll, svchost.exe), and malicious payloads that relate to the APT 29 “Duke” set of software (psexesvc, Python, PowerShell).

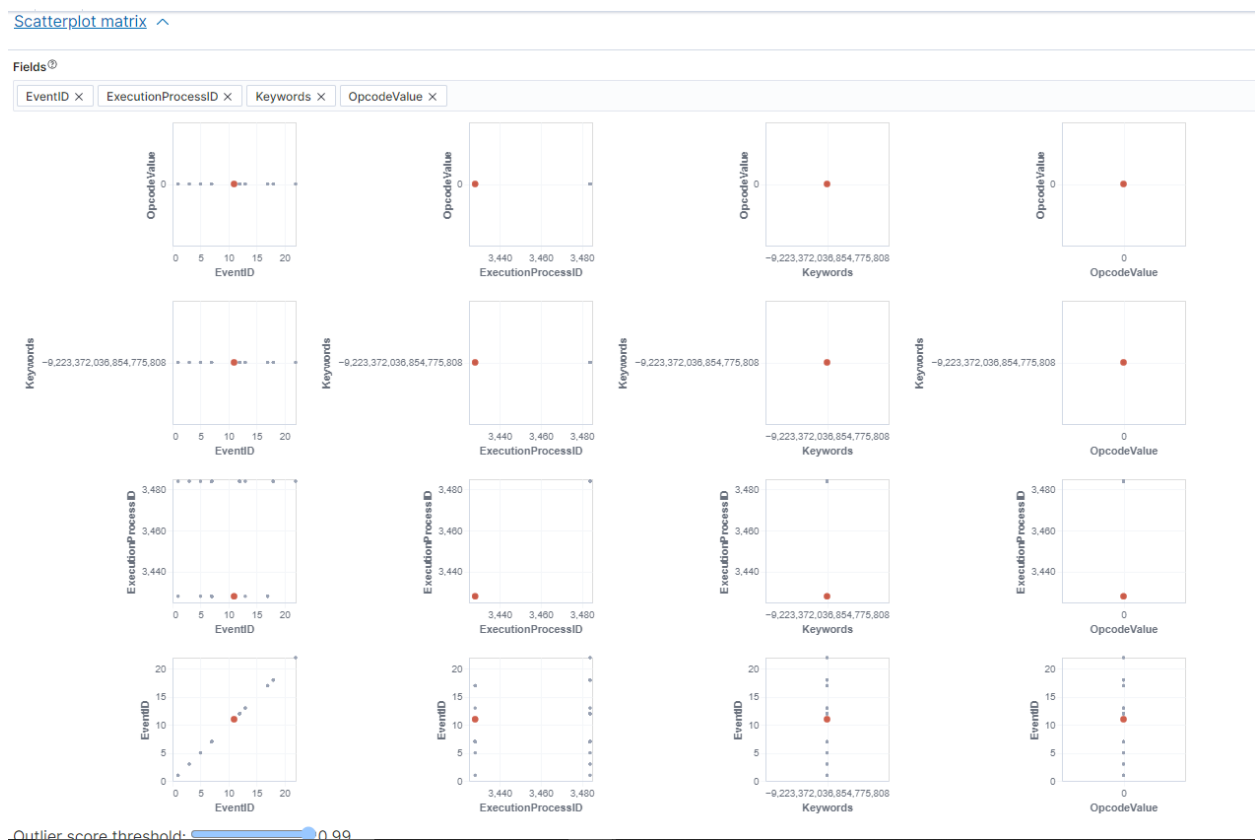
Data Frame Analytics

The Data Frame Analytic jobs are important for finding anomalistic data within individual categories of features (Ex: Querying for the python.exe feature specifically within the Image.keyword field). They also allow you to view other fields and features for that specific data point. This is very important, as it allows for us to gain a better understanding of what is going on with the data and/or cross-reference results with Anomaly Detection jobs to validate anomalies.

To create a DFA job:

Analytics						Total analytics jobs: 4 Running: 0 Stopped: 4
ID ↑	Type	Status	Progress	Creation time	Actions	
dfa-1	outlier_detection	stopped	Phase 4/4	April 3rd 2021, 16:24:14	View	
dfa-many-queries	outlier_detection	stopped	Phase 3/4	April 8th 2021, 21:46:56	View	
dfa-psexex	outlier_detection	stopped	Phase 4/4	April 16th 2021, 11:09:31	View	
dfa-sourcename-powershell	outlier_detection	stopped	Phase 4/4	April 23rd 2021, 13:52:44	View	
Rows per page: 10						< 1 >
						Refresh Manage jobs

Clicking the “View” button () will take you to the Data Frame Analyzer of that job.



The graph shows outliers found in certain fields in the dataset. Fields can be added through the “Fields” bar. Enabling “Random Scoring” will highlight anomalistic data points, and hovering over those data points will provide more information about the Event ID and Outlier Score of those individual points.

Results ^

Total docs
>10000

Histogram charts
817 columns hidden
1 fields sorted

ml.outlier_score	@timestamp	SourceName	ContextInfo
0.999	May 2, 2020 @ 04:16:32.3...	Microsoft-Windows-PowerShell	
0.999	May 1, 2020 @ 23:21:26.6...	Microsoft-Windows-PowerShell	
0.999	May 1, 2020 @ 22:59:06.4...	Microsoft-Windows-PowerShell	
0.999	May 1, 2020 @ 23:00:15.4...	Microsoft-Windows-PowerShell	
0.999	May 1, 2020 @ 23:21:32.2...	Microsoft-Windows-PowerShell	Severity = Informational Host Name = ConsoleHost Host Version = 5.1.18362.628 Host ID = 0299cb0e-483f-46b0-a86c-302479840076 Host Application = powershell.exe -nop -w hidden...
0.999	May 1, 2020 @ 23:21:32.2...	Microsoft-Windows-PowerShell	Severity = Informational Host Name = ConsoleHost Host Version = 5.1.18362.628 Host ID = 0299cb0e-483f-46b0-a86c-302479840076 Host Application = powershell.exe -nop -w hidden...
0.999	May 1, 2020 @ 23:21:32.2...	Microsoft-Windows-PowerShell	Severity = Informational Host Name = ConsoleHost Host Version = 5.1.18362.628 Host ID = 0299cb0e-483f-46b0-a86c-302479840076 Host Application = powershell.exe -nop -w hidden...
0.999	May 1, 2020 @ 23:21:32.2...	Microsoft-Windows-PowerShell	Severity = Informational Host Name = ConsoleHost Host Version = 5.1.18362.628 Host ID = 0299cb0e-483f-46b0-a86c-302479840076 Host Application = powershell.exe -nop -w hidden...
0.999	May 1, 2020 @ 23:21:32.2...	Microsoft-Windows-PowerShell	Severity = Informational Host Name = ConsoleHost Host Version = 5.1.18362.628 Host ID = 0299cb0e-483f-46b0-a86c-302479840076 Host Application = powershell.exe -nop -w hidden...
0.999	May 1, 2020 @ 23:21:32.2...	Microsoft-Windows-PowerShell	Severity = Informational Host Name = ConsoleHost Host Version = 5.1.18362.628 Host ID = 0299cb0e-483f-46b0-a86c-302479840076 Host Application = powershell.exe -nop -w hidden...
0.999	May 1, 2020 @ 23:21:32.2...	Microsoft-Windows-PowerShell	Severity = Informational Host Name = ConsoleHost Host Version = 5.1.18362.628 Host ID = 0299cb0e-483f-46b0-a86c-302479840076 Host Application = powershell.exe -nop -w hidden...
0.999	May 1, 2020 @ 23:21:32.2...	Microsoft-Windows-PowerShell	Severity = Informational Host Name = ConsoleHost Host Version = 5.1.18362.628 Host ID = 0299cb0e-483f-46b0-a86c-302479840076 Host Application = powershell.exe -nop -w hidden...
0.999	May 1, 2020 @ 23:21:32.2...	Microsoft-Windows-PowerShell	Severity = Informational Host Name = ConsoleHost Host Version = 5.1.18362.628 Host ID = 0299cb0e-483f-46b0-a86c-302479840076 Host Application = powershell.exe -nop -w hidden...
0.999	May 1, 2020 @ 23:21:32.2...	Microsoft-Windows-PowerShell	Severity = Informational Host Name = ConsoleHost Host Version = 5.1.18362.628 Host ID = 0299cb0e-483f-46b0-a86c-302479840076 Host Application = powershell.exe -nop -w hidden...
0.999	May 1, 2020 @ 23:21:32.2...	Microsoft-Windows-PowerShell	Severity = Informational Host Name = ConsoleHost Host Version = 5.1.18362.628 Host ID = 0299cb0e-483f-46b0-a86c-302479840076 Host Application = powershell.exe -nop -w hidden...
0.999	May 1, 2020 @ 23:21:32.2...	Microsoft-Windows-PowerShell	Severity = Informational Host Name = ConsoleHost Host Version = 5.1.18362.628 Host ID = 0299cb0e-483f-46b0-a86c-302479840076 Host Application = powershell.exe -nop -w hidden...
0.999	May 1, 2020 @ 23:21:32.2...	Microsoft-Windows-PowerShell	Severity = Informational Host Name = ConsoleHost Host Version = 5.1.18362.628 Host ID = 0299cb0e-483f-46b0-a86c-302479840076 Host Application = powershell.exe -nop -w hidden...
0.999	May 1, 2020 @ 23:21:32.2...	Microsoft-Windows-PowerShell	Severity = Informational Host Name = ConsoleHost Host Version = 5.1.18362.628 Host ID = 0299cb0e-483f-46b0-a86c-302479840076 Host Application = powershell.exe -nop -w hidden...
0.999	May 1, 2020 @ 23:16:32.0...	Microsoft-Windows-PowerShell	

Scrolling down to the results, information will be visualized by ml.outlier score (column on the far left). The closer the score is to 1, the greater the possibility of it being anomalous. By default, the columns will be listed in alphabetical order and only a specific number of columns will be shown. For different Data Frame jobs, different columns will have to be unhidden. For example, the PowerShell DFA job shown here provides the most information about the data point when “SourceName” and “ContextInfo” are unhidden.

Severity = Informational Host Name = ConsoleHost Host Version = 5.1.18362.628 Host ID = 83b76c30-796d-4656-89d7-6d973ff51182 Host Application = powershell.exe -c Get-ItemPropertyValue 'HKLM:\SOFTWARE\Javasoft\value Supplement' | Invoke-Expression Engine Version = 5.1.18362.628 Runspace ID = 2740b122-8c97-4c21-b66b-41a191d78cdc Pipeline ID = 1 Command Name = Command Type = Script Script Name = Command Path = Sequence Number = 18 User = DMEVALS\pbeesly Connected User = Shell ID = Microsoft.PowerShell

4a6e-8714-a0d0029781f8 Host A...
4a6e-8714-a0d0029781f8 Host A...
-4656-89d7-6d973ff51182 Host A...

ConsoleHost Host Version = 5.1.18362.628 Host ID = 83b76c30-796d-4656-89d7-6d973ff51182 Hos...

Clicking the bottom right part of any field in these columns will provide the complete data listed.

If you need more help going through and setting up the machine learning jobs refer to the [How to Setup Machine Learning](#) included in this package.

Kibana Dashboard

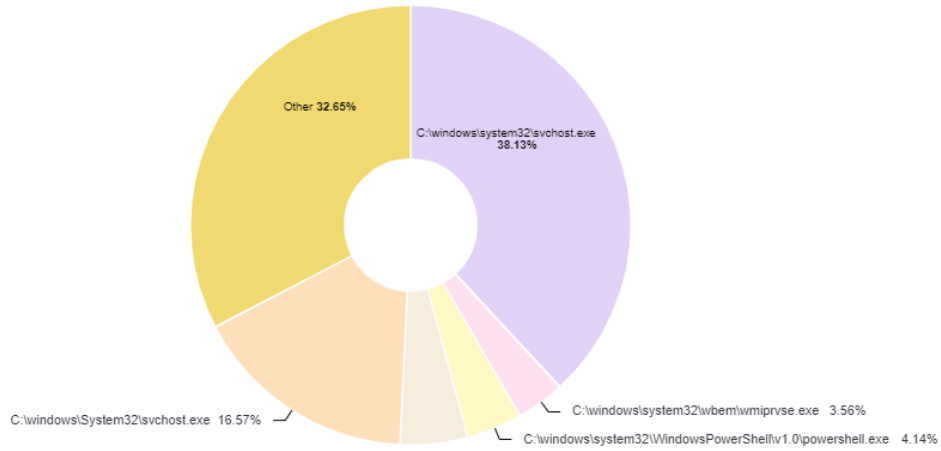
Kibana's dashboard module is a great tool to visualize your data by creating visualizations of various fields within your index (dataset). After you have created all of the visualizations you need, you now can compile them into a dashboard. ***Note:** If you have not created visualizations for your index yet, then reference our supplemental document [Kibana Suggestions](#), where there is a section describing a step-by-step guide.

Begin by clicking **Create Dashboard**, this will redirect you to the 'Editing Dashboard' section within Kibana. You now have two different options, **Create Panel** or **Add From Library**. Clicking - Create Panel - a pop-up appears where you can select from the following options: Lens, Maps, TSVP, Custom Visualization, and Aggregation Based. Select the best choice, from our project we only utilized the option Lens. Selecting - Add From Library - opens your Visualization Library which consists of all of the visualizations you have created thus far. Clicking a visualization within your library automatically adds that visualization to your dashboard (Make sure to set a proper time range). You also can resize a visualization within your dashboard by clicking and holding the arrow in the bottom right-hand corner. Additionally, if you wish to change the color scheme or the size of information presented, you can do this by selecting the cogwheel next to a particular visualization and clicking 'Edit Lens'. This will take you to the Visualization Library where you can edit those for this particular visualization.

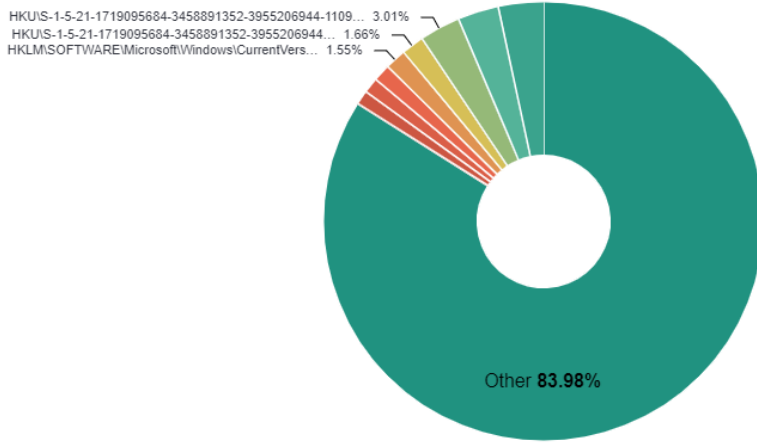
Kibana allows users to conduct queries within the Dashboard to hone in and see visualizations of a particular field. This can be done by either: manually conducting a query search or you can select a part of a visualization which will then automatically conduct a query on what was selected. The screenshots below cover the following material: Sample Visualizations, Overview of a dashboard, Overview of a dashboard with a search query, Filter section for dashboards, and Suggested searches.



APT 29 - Image (ML)



APT 29 - Target Object (ML)



Search

KQL

May 1, 2020 @ 22:55:26.000 → Apr 8, 2021 @ 18:29:00.000

+ Add filter

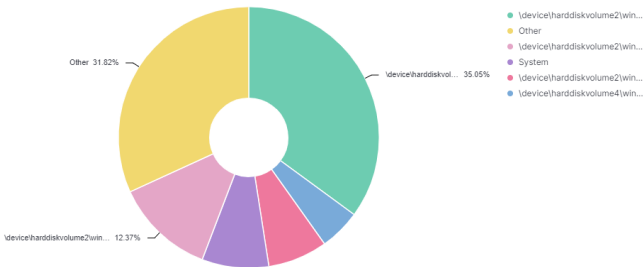
APT 29 - Record Count

783,367
Count of records

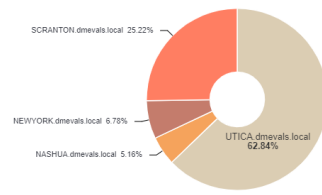
APT 29 - Timeline



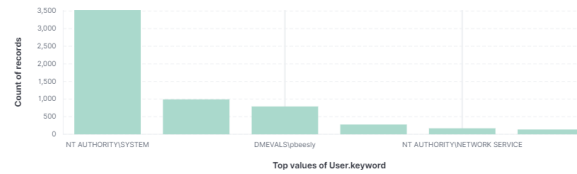
APT 29 - Top Applications

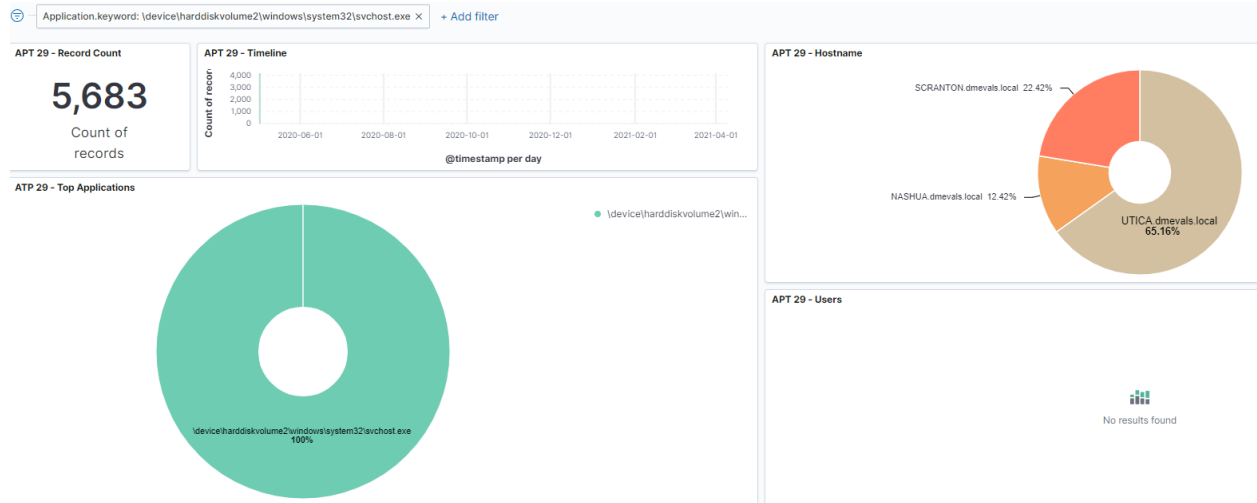


APT 29 - Hostname



APT 29 - Users





EDIT FILTER

[Edit as Query DSL](#)

Field

Select a field first

Operator

Waiting

☐ ☒ Create custom label?

Cancel

Save

Search

	<code>_id</code>	Filter results that contain <code>_id</code>
	<code>_index</code>	Filter results that contain <code>_index</code>
	<code>_type</code>	Filter results that contain <code>_type</code>
	<code>@timestamp</code>	Filter results that contain <code>@timestamp</code>
	<code>@version.keyword</code>	Filter results that contain <code>@version.keyword</code>
	<code>@version</code>	Filter results that contain <code>@version</code>
	<code>AccessList.keyword</code>	Filter results that contain <code>AccessList.keyword</code>
	<code>AccessList</code>	Filter results that contain <code>AccessList</code>
	<code>AccessMask.keyword</code>	Filter results that contain <code>AccessMask.keyword</code>
	<code>AccessMask</code>	Filter results that contain <code>AccessMask</code>
	<code>AccessReason.keyword</code>	Filter results that contain <code>AccessReason.keyword</code>

Credentials

<https://9e8fe26ad3d44ab1a18df87e9046c493.us-east-1.aws.found.io:9243/>

Profile

Provide personal details.


Username

 cisa

Password

Protect your data with a strong password.

Password

 CISA!Rocks



Password must be at least 6 characters.