

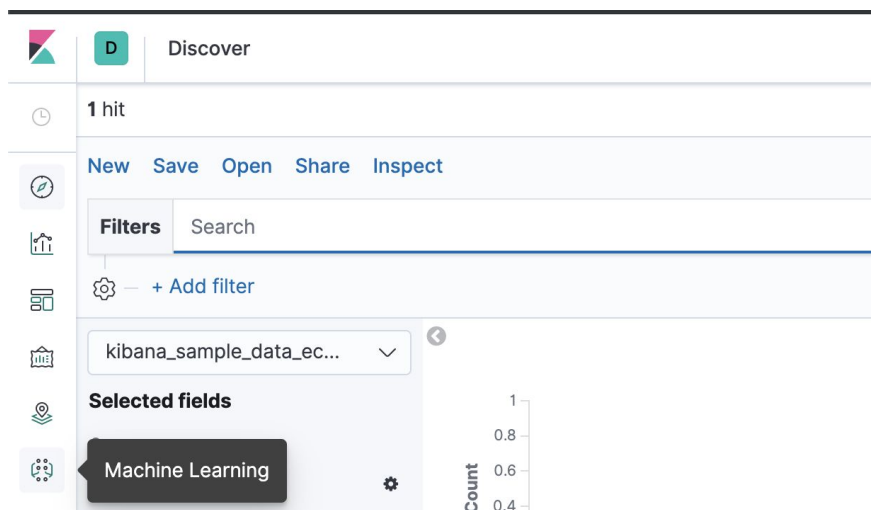
Elastic SIEM Workshop

Lab 3 - Detect Anomalies using ML

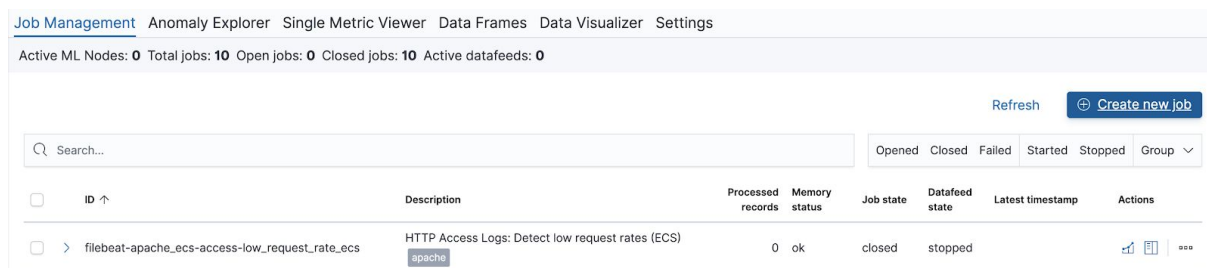
HTTP Access Logs: Detect unusual status code rates

In this lab we are going to create an Elastic ML job to detect unusual status code rates in NGIX logs, leveraging the the 'Sample Web Logs Dataset'.

1. Open the Machine Learning App



2. Click on the “Create New Job” button.



3. Select “kibana_sample_data_logs” as the dataset that the ML job you are creating will use to build the model.

Machine Learning / Create job

From a New Search, Select Index

Or, From a Saved Search

Q Filter...

8 of 8

Q Saved Searches Filter...

1-20 of 95

Name ▲

kibana_sample_data_logs

kibana_sample_data_flights

auditbeat-*

kibana_sample_data_ecommerce

packetbeat-*

filebeat-*

metricbeat-*

winlogbeat-*

Name ▲

ASA Firewall Events [Filebeat Cisco]

ASA Firewall flows [Filebeat Cisco]

Alerts [Suricata] ECS

All ASA Logs [Filebeat Cisco]

All Logs [Filebeat PostgreSQL] ECS

All logs [Filebeat Kafka] ECS

All logs [Filebeat MongoDB] ECS


Apache HTTPD ECS

4. Select the “Multi Metric” wizard

Create a job from the index pattern kibana_sample_data_logs


Use a supplied configuration


The fields in your data have been recognized as matching known configurations. Select to create a set of machine learning jobs and associated dashboards.


**Kibana sample data web logs**
Find anomalies in Kibana sample web logs data


Use a wizard

Use one of the wizards to create a machine learning job to find anomalies in your data.

**Single metric**
Detect anomalies in a single time series.

**Multi metric**
Detect anomalies in multiple metrics by splitting a time series by a categorical field.

**Population**
Detect activity that is unusual compared to the behavior of the population.

**Advanced**
Use the full range of options to create a job for more advanced use cases.

5. Select “Count” as the aggregation for “event rate”

Job settings

Fields

☒ event rate

Count ▼

☐ clientip

Distinct count ▼

☐ ip

Distinct count ▼

☐ bytes

Mean ▼

☐ machine.ram

Mean ▼

☐ memory

Mean ▼

5. Split Data as “response.keyword”

Split Data [Remove split](#)

↑ response.keyword

6. Add “client ip” as a Key Field (influencers)

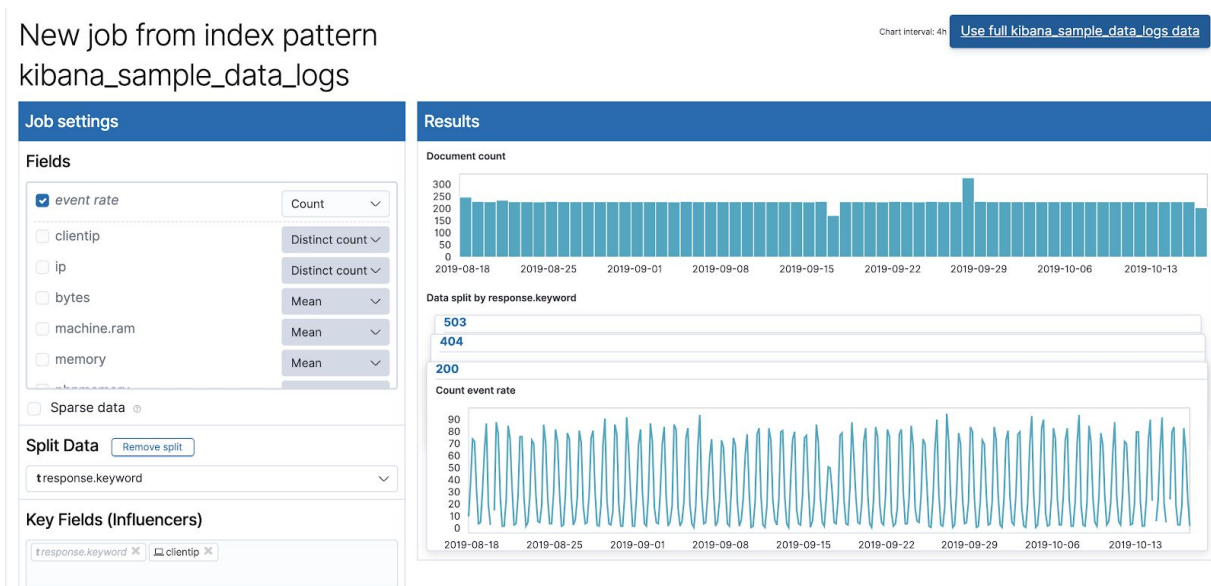
Key Fields (Influencers)

↑ response.keyword

ip

clientip

7. Click on “Use full kibana_sample_data_logs_data” button on the top right.



8. Use 60m as the “Bucket Span”

9. Name your ML job (e.g., "response-code-split) and click on the "Create Job" button.

Job Details

Name ⓘ

response-code-split

Description ⓘ

Job description

Job Groups ⓘ


Job Group

▶ **Advanced** ⓘ

[Move to advanced job configuration](#)

[Validate Job](#) ⓘ [Create Job](#)

10. Once the job is created, click on “View Results” to open the “Anomaly Explorer” view.

Job response-code-split
created 

[Reset](#) [View Results](#)

11. From the “Anomaly Explorer” look at the anomaly with highest severity, related to a client ip with a high rate of 404 requests. What’s the client ip in question?

time	severity	detector	found for	influenced by	actual	typical	description	actions
October 5th 2019	98	count	404	clientip: 30.156.16.164 response.keyword: 404 	101	1.03	98x higher	
<div><div>Description</div><div>critical anomaly in count found for response.keyword 404</div><div>Details on highest severity anomaly</div><div>response.keyword 404</div><div><div>time</div><div>October 5th 2019, 05:00:00 to October 5th 2019, 06:00:00</div></div><div><div>function</div><div>count</div></div><div><div>actual</div><div>101</div></div><div><div>typical</div><div>1.03</div></div><div><div>job ID</div><div>multi</div></div><div><div>probability</div><div>1.0857897151588303e-39</div></div><div><div>Influencers</div><div>clientip 30.156.16.164 </div><div>response.keyword 404 </div></div></div>								