Deployment Guide

Protecting Personally Identifiable Information

5/7/2025

Protecting Personally Identifiable Information

Web Engagement allows you to protect personally identifiable information (PII) for all direct requests made to Elasticsearch by way of the HTTP interface, which is normally on port 9200. In particular, you can:

- Prevent changes to the Elasticsearch Cassandra index
- Restrict access to sensitive data

Note: This protection does not extend to access through the binary interface on port 9300, which is used both by Java clients and for communication between Cassandra nodes.

Preventing Changes to the Elasticsearch Cassandra Index

You can prevent modifications to the Elasticsearch index by using a read-only filter that prohibits REST calls (POST, PUT, DELETE) on the HTTP interface for all URLs used by Elasticsearch. However, even if you use the filter, Elasticsearch still allows you to make POST or PUT read requests containing JSON content. Here is an example:

```
POST http://server:9200/indexName/someType/_search
{ JSON }
```

You can also make read-only requests to known resources such as `/*.kibana/` and the Web Engagement Reporting Server collector configuration parameter.

The filter uses a regular expression to limit access. The default regex looks like this:

```
/_msearch|/_mget|/_search|/_search_shards|/_suggest|/_count|/_validate|/_explain|\.kibana|
garp\.collector|tenantConfiguration|\.garp|systemConfiguration|/conversion|/webtraffic
```

If a request is rejected, the server will respond with `403 "Access Forbidden"`.

Application Options

The following options belong to the [elasticsearch] section.

Key	Value	Default	Comment
http-read-only	Boolean	false	Enables read-only mode
read-only-urls-regexp	String	See the default expression mentioned above	Value must be a valid regular expression

Restricting Access to Sensitive Data

You can also filter all responses from Elasticsearch queries that return data. If one of the top level keys of the JSON source response matches the filter it will be replaced by ********. It is important to note that even a JSON object or a JSON array will be replaced by this string. The length of the string is constant.

Here is the default expression, which is case insensitive:

```
password
```

Unfiltered Response

```
{
   "took":90,
   "timed_out":false,
   "_shards":{
      "total":5,
      "successful":5,
      "failed":0
   },
   "hits":{
      "total":1,
      "max_score":1.0,
      "hits":[
         {
            "_index":"cointr",
            "_type":"user",
            "_id":"1",
            "_score":1.0,
            "_source":{
               "postDate":"2015-06-24",
               "password":"secret",
               "user":"jh"
            }
         }
      ]
   }
}
```

Filtered Response

```
{
   "took":1,
   "timed_out":false,
   "_shards":{
      "total":5,
      "successful":5,
      "failed":0
   },
   "hits":{
      "total":1,
      "max_score":1.0,
      "hits":[
         {
            "_index":"cointr",
            "_type":"user",
            "_id":"1",
            "_score":1.0,
```

```
            "_source":{
               "postDate":"2015-06-24",
               "password":"********",
               "user":"jh"
            }
         }
      ]
   }
}
```

Application Options

The following options belong to the [elasticsearch] section.

Key	Value	Default	Comment
http-filter	Boolean	false	Enables response filter mode
reponse-filter-regexp	String	password	Value must be a valid regular expression and is not case sensitive
reponse-filter-regexp-type	String	REGEXP	Regular expression type: **REGEXP**—standard Java regular expression **PATH**—JSON path expression