# Ejercicio 2 – Configuración de índices

|  |  |
| --- | --- |
| Apellidos, Nombre: | Batista, Tania |
| Puntuación Final: |  |

## Pregunta 1 [Puntuación 2]

Crear un índice y su mapping para documentos que lleven el formato mostrado en el siguiente ejemplo

Asegurarse que las siguientes condiciones

* El campo host no es analizado
* Es un índice válido para un cluster de ElasticSearch de un solo nodo

*for documents that have the format shown in the example*

* *Create an index and its mapping*
* *Make sure the “host” field is not analyzed*
* *Make sure the index is valid for a single-node ElasticSearch cluster*

{

"source": "MySource-10.134.4.94-DBName",

"lastExecDate": "2015-04-22 10:45:45",

"currentDate": "2015-04-22 08:52:25",

"instant": "2015-04-22T08:52:42Z",

"host": "EMEOE3402.dir.svc.accenture.com",

"serviceName": "System Center Core Monitoring",

"counterName": "agent processor utilization",

"category": "agent processor utilization",

"component": "",

"value": "0"

}

* Only set “host” fields on the mapping
* you can just index a document and the index, type, and fields will spring to life automatically

PUT logs

{

"mappings": {

"log": {

"properties": {

"host": {

"type": "keyword"

}

}

}

},

"settings": {

"index": {

"number\_of\_shards": "1",

"number\_of\_replicas": "1"

}

}

}

**

*Source:*

*https://www.elastic.co/guide/en/elasticsearch/reference/current/dynamic-mapping.html*

*https://www.elastic.co/guide/en/elasticsearch/reference/current/mapping.html*

## Pregunta 2 [Puntuación 1]

##### Configurar el índice para que elimine todo documento al cabo de 1 minuto de haberlo dado de alta.

*Configure the index so that it deletes all documents after 1 minute of being registered.*

* Set the curators to delete the documents.
* For this you separate the index by minutes, and create a curator.

*Source:*

*https://www.elastic.co/guide/en/elasticsearch/client/curator/current/about.html*

*https://www.elastic.co/guide/en/elasticsearch/client/curator/current/ex\_delete\_indices.html*

## Pregunta 3 [Puntuación 1]

##### Obtener los settings actuales del índice que acabamos de crear

*Get the current settings of the index that we just created*

* GET logs/\_settings

## Pregunta 4 [Puntuación 1]

### Haciendo uso del índice creado en el apartado anterior,

### Definir dos alias en una sola llamada sobre el índice [0.2]: indiceAlias1, indiceAlias2

*Define two aliases in a single call on the index: index1, index2*

POST \_aliases

{

"actions" : [

{ "add" : { "index" : "logs", "alias" : "indexAlias1" } },

{ "add" : { "index" : "logs", "alias" : "indexAlias2" } }

]

}

*Source: https://www.elastic.co/guide/en/elasticsearch/reference/6.2/indices-aliases.html*

### Obtener todos los alias del servidor de ES [0.1]

*Get all the aliases of the ES server*

* GET \_aliases

### ¿Para qué crees que son útiles los alias sobre los índices en ElasticSearch? Comenta algún escenario que se te ocurra para su uso en producción. [0.8]

*Why do you think aliases are useful over indexes in ElasticSearch? Discuss some scenario that you can think of for use in production.*

* Yes, aliases are very useful features in ElasticSearch.
* They are useful if we want to switch an index (like indexAlias1) with a new one (such as indexAlias2), without changing our application source code or the index mapping

*Source: https://www.elastic.co/guide/en/elasticsearch/guide/current/index-aliases.html*

## Pregunta 5 [Puntuación 1]

Mostrar todos los índices que tiene el ElasticSearch actual. Proporcionar, al menos, dos soluciones a la pregunta.

*Show all the indexes that the current ElasticSearch has. Provide 2 or more solutions to the question.*

* GET \_cat/indices
* GET /\_stats?filter\_path=indices

## Pregunta 6 [Puntuación 2]

### Implementar un analizador que permita eliminar cualquier rastro de etiquetas HTML de un documento y nos retorne el contenido del mismo

#### Eliminar todos los tags HTML del documento suministrado.

#### Particionar el texto haciendo uso de espacios en blanco

#### Eliminar los caracteres de puntuación

#### Mover todos los tokens a lowercase.

#### Eliminar stopwords tradicionales

#### El analyzer debe ser capaz de retornar el texto, palabra a palabra, sin los tags HTMLs proporcionados en un documento como el siguiente

*Implement an analyzer to eliminate any traces of HTML tags from a document and return the content of the same*

* *Partition the text using blank spaces*
* *Eliminate punctuation characters*
* *Move all tokens to lowercase.*
* *Eliminate traditional stopwords*
* *Remove all HTML tags from the supplied document*
* *The analyzer must be able to return the text, word by word, without the HTML tags)*

PUT my\_index

{

"settings": {

"analysis": {

"analyzer": {

"my\_analyzer": {

"tokenizer": "standard",

"char\_filter": ["htmlstrip"],

"filter": ["lowercase", "englishstopwords"]

}

},

"char\_filter": {

"htmlstrip": {

"type": "html\_strip",

"escaped\_tags": ["b"]

}

},

"filter": {

"englishstopwords": {

"type": "stop",

"stopwords": "\_english\_"

}

}

}

}

}

*Source: https://www.elastic.co/guide/en/elasticsearch/reference/current/analysis-custom-analyzer.html*

## Pregunta 7 [Puntuación 1]

### ¿Qué hace el comando *http://<server>/<indexName>/\_optimize* cuando lo ejecutamos contra uno o varios índices?

*What does the command do when we execute it against one or more indexes?*

* It optimizes the indexes so the search can be performed a lot faster.
* It reduces the number of segments by merging them.
* This feature is deprecated for ES 2.1.0

*Source: https://www.elastic.co/guide/en/elasticsearch/reference/2.2/indices-optimize.html*

## Pregunta 8 [Puntuación 1]

### Si estamos trabajando en un entorno Linux, que consideraciones de configuración de ElasticSearch deberíamos contemplar para mejorar el performance de múltiples índices simultáneos. Explica los efectos de cualquier cambio que sugieras.

*If we are working in a Linux environment, what ElasticSearch configuration considerations we should contemplate to improve the performance of multiple simultaneous indexes? Explain the effects of any change you suggest.*

* There are many things to consider.
* We need to check the limitations of the file systems, network, memory, etc.
* You may need to write scripts for scheduled cleaning, otherwise ElasticSearch may grow too much, and the server would not be able to reload

*Source: https://www.elastic.co/guide/en/elasticsearch/guide/current/indexing-performance.html*