

Capturas de pantalla o logs

											Export requests	Import requests	
1	ip	heap.percent	ram.percent	cpu	load_1m	load_5m	load_15m	node.role	master	name			
2	192.199.1.59	57	97	7	0.00	0.02	0.00	dim	*	nodo-1			
3	192.199.1.60	55	90	2	0.33	0.12	0.08	dim	-	nodo-2			
4	192.199.1.61	18	97	3	0.04	0.04	0.01	dim	-	nodo-3			
5													

Status	Alerts	Nodes	Indices	JVM Heap	Total shards	Unassigned shards	Documents	Data
Green	1	3	69	2.7 GB / 6.0 GB	146	0	3,423,067	2.6 GB

Name	Alerts	Status	Roles	Shards	CPU Usage	Load Average	JVM Heap	Disk Free Space
★ nodo-1 192.199.1.59:9300	● Clear	● Online	N/A	49	↑ 1%	↓ 0	↓ 35%	↓ 32.1 GB
■ nodo-2 192.199.1.60:9300	● Clear	● Online	N/A	48	↑ 0%	↓ 0.02	↓ 35%	↓ 29.4 GB
■ nodo-3 192.199.1.61:9300	▲ 1 alert	● Online	N/A	49	↑ 0%	↓ 0.06	↑ 65%	↓ 8.9 GB

Rows per page: 20 < 1 >

Configuración de mappings y settings

Settings

```
{  
  "index": {  
    "lifecycle": {  
      "name": "noticias-policy"  
    },  
    "refresh_interval": "30s",  
    "analysis": {  
      "filter": {  
        "spanish_synonyms_filter": {  
          "type": "synonym",  
          "synonyms_path": "synonyms/spanish_synonyms.txt",  
          "expand": "true"  
        },  
        "spanish_stemmer": {  
          "type": "stemmer",  
          "language": "light_spanish"  
        },  
        "spanish_stop": {  
          "type": "stop",  
          "stopwords": "_spanish_"  
        }  
      }  
    }  
  }  
}
```

```

},
"normalizer": {
  "lowercase_normalizer": {
    "filter": [
      "lowercase"
    ],
    "type": "custom"
  }
},
"analyzer": {
  "spanish_content": {
    "filter": [
      "lowercase",
      "spanish_stop",
      "spanish_stemmer",
      "spanish_synonyms_filter"
    ],
    "type": "custom",
    "tokenizer": "standard"
  }
}
},
"number_of_shards": "3",
"number_of_replicas": "1"
}
}

```

Lifecycle: Se define una política llamada **noticias-policy** que permite que el índice gestione automáticamente su ciclo de vida, moviéndose entre fases como caliente, templada o eliminación para optimizar almacenamiento y rendimiento.

Refresh Interval: El índice se actualiza cada 30 segundos, es decir, los documentos nuevos serán visibles en las búsquedas tras ese tiempo. Esto mejora el rendimiento frente a actualizaciones muy frecuentes.

Análisis:

- **Filtros:**
 - **spanish_synonyms_filter:** aplica sinónimos definidos en un archivo externo, para que las búsquedas reconozcan palabras equivalentes.
 - **spanish_stemmer:** realiza un stemming ligero en español, reduciendo palabras a su raíz para mejorar la búsqueda (por ejemplo, "corriendo" se reduce a "corr").
 - **spanish_stop:** elimina palabras muy comunes en español como "el", "y", "de", que no aportan valor a las búsquedas.

- **Normalizador:** Convierte todo a minúsculas, útil para búsquedas insensibles a mayúsculas.
- **Analizador:** Combina los filtros anteriores con un tokenizador estándar, procesando el texto en español para búsquedas más eficientes aplicando sinónimos, stemming y eliminación de stopwords.

Shards y réplicas: El índice se divide en 3 shards para permitir paralelismo y escalabilidad, y cada shard tiene una réplica para aumentar la disponibilidad y tolerancia a fallos.

Mappings

```
{
  "dynamic": "strict",
  "properties": {
    "date": {
      "format": "strict_date_optional_time||epoch_millis",
      "type": "date"
    },
    "@timestamp": {
      "format": "strict_date_optional_time||epoch_millis",
      "type": "date"
    },
    "author": {
      "normalizer": "lowercase_normalizer",
      "type": "keyword"
    },
    "fuente": {
      "normalizer": "lowercase_normalizer",
      "type": "keyword"
    },
    "embedding": {
      "dims": 768,
      "similarity": "cosine",
      "index": true,
      "type": "dense_vector"
    },
    "title": {
      "analyzer": "spanish_content",
      "type": "text",
      "fields": {
        "raw": {
          "type": "keyword"
        }
      }
    },
    "body": {
      "analyzer": "spanish_content",
      "type": "text"
    }
  }
}
```

```
"url": {  
  "normalizer": "lowercase_normalizer",  
  "type": "keyword"  
},  
"body_length": {  
  "type": "integer"  
}  
}
```

Dynamic: "strict"

- Esto significa que **solo se pueden indexar los campos definidos** en `properties`.
- Si se intenta guardar un documento con un campo no definido, Elasticsearch lo rechazará.

Properties (campos definidos):

- **date y @timestamp:**
 - Tipo `date`, con formatos `strict_date_optional_time` o `epoch_millis`.
 - Se usan para almacenar la fecha del artículo y la marca de tiempo de ingestión.
- **author y fuente:**
 - Tipo `keyword` (no analizado), ideal para filtros exactos y agregaciones.
 - Se normalizan a minúsculas con `lowercase_normalizer` para que la búsqueda no distinga mayúsculas.
- **embedding:**
 - Tipo `dense_vector` de 768 dimensiones.
 - Permite almacenar vectores de características (por ejemplo, de modelos de lenguaje) para **búsquedas semánticas**.
 - La similitud se calcula con `coseno` y se indexa para poder buscar por proximidad vectorial.
- **title:**

- Tipo `text`, se analiza con el analizador `spanish_content` que aplica stemming, stopwords y sinónimos en español.
 - También tiene un subcampo `raw` de tipo `keyword`, útil para búsquedas exactas o agregaciones.
- **body:**
 - Tipo `text`, también analizado con `spanish_content` para búsquedas de texto completo en español.
 - **url:**
 - Tipo `keyword`, normalizado a minúsculas.
 - Útil para búsquedas exactas de URLs sin importar mayúsculas.
 - **body_length:**
 - Tipo `integer`.
 - Permite almacenar la longitud del contenido, por ejemplo, para filtrar artículos según su tamaño.

Explicación de las políticas de vida seleccionadas.

noticias-policy

Hot phase



■ Hot phase

■ Delete phase

Move data into phase when

30d old

Delete searchable snapshot

Yes

Se establece esta política de vida en todos los índices de noticias para que no se almacenen noticias antiguas que ya no aportan nada en la actualidad.

Hot phase: Los datos recién indexados permanecen activos y buscables en esta fase. Aquí se concentran las operaciones de lectura/escritura frecuentes.

Delete phase: Los datos se eliminan automáticamente cuando cumplen **30 días** de antigüedad.

Delete searchable snapshot: Se confirma que se borran también los snapshots que se podrían haber creado para conservar los datos antes de su eliminación.

Listado de API Keys generadas

Se han generado 3 api keys diferentes.

```
[  
  {  
    "id": "gFMZwJoB_ng_JLRryej",  
    "name": "metricbeat-key",  
    "api_key": "hh01Bw7gE8r7mPPsFPBZxA",  
    "encoded": "z0ZNWndKb0JfbmdfsKxScnllYWo6aGgwMUJ3N2dFOHI3bVBQc0ZQQlp4QQ=="  
  },  
  {  
    "id": "jAbx3zoB8_buqiHcbWPg",  
    "name": "logstash-key",  
    "api_key": "3ynGrlurs3CySNbLdXXDRg",  
    "encoded": "akFieDNab0I4X2J1cWlIY2JXUGc6M3luR3JsdXJzM0N5U05iTGRYWERSZw=="  
  },  
  {  
    "id": "8weo3poB8_buqiHcR9gy",  
    "name": "script_querys_key",  
    "api_key": "SN6PYZxz7rRFbsfxDOAngQ",  
    "encoded": "OHDlbzNwb0I4X2J1cWlIY1I5Z3k6U042UFlaeHo3c1JGYNnmeERPQW5nUQ=="  
  },  
  {  
    "id": "odWT7zoB8FUZJiYYgY6s",  
    "name": "chatbot-noticias-key",  
    "api_key": "kbAPUZh01R_E8nOUedLyUg",  
    "encoded": "b2RXVDDab0I4R1VaSm1ZWwdzN1M6a2JBUFVaaE8xU19FOG5PVwVktH1VZw=="  
  }  
]
```

La primera api key **metricbeat-key** se usa para que metricbeat pueda monitorizar diferentes parámetros como la cpu de las máquinas, disco, ram..., además de la salud del cluster de los 3 nodos de elasticsearch. Y tiene los siguientes permisos.

```
"privileges": [
    "write",
    "create_index",
    "manage",
    "read"
],
```

Con el fin de poder monitorizar de forma correcta.

La segunda api key **logstash-key** se usa para la ingestión de datos en los diferentes índices a través de logstash. Tiene los siguientes permisos.

```
,  
"privileges": [
    "write",
    "create_index"
],
```

La tercera api key **script_querys_key** sirve para el script de querys de búsqueda en los índices de noticias de elasticsearch. Tiene los siguientes permisos.

```
"privileges": [
    "read"
],
```

Por último la cuarta api key **chatbot-noticias-key** sirve para el chatbot de noticias para su consulta de datos. Tiene los siguientes permisos.

```
"privileges": ["read", "view_index_metadata",
    "monitor"]
```

Script queries

Se ha desarrollado un script interactivo con diferentes queries pudiendo filtrar por diferentes campos como título, fechas, autor y body además se puede aplicar fuzziness en cada una de las consultas.

```
Opciones de filtro:  
1. Buscar por título  
2. Buscar por autor  
3. Buscar por fecha (rango)  
4. Buscar por body  
5. Ejecutar búsqueda  
6. Salir sin buscar  
Elige una opción (1-6):
```

Si elegimos por ejemplo buscar por título. Introducimos la palabra clave a buscar y nos pregunta si queremos aplicar fuzziness. En este ejemplo ponemos juntos como palabra clave y aplicamos fuzziness.

```
Elige una opción (1-6): 1  
Introduce palabra clave para el título: junts  
¿Usar fuzziness automática? (s/n): s  
Filtro de título añadido.
```

También vamos a aplicar un filtro por body. En este ejemplo ponemos constitución como palabra clave, en este caso no aplicamos fuzziness.

```
Elige una opción (1-6): 4  
Introduce palabra clave para buscar en body: constitucion  
¿Usar fuzziness automática? (s/n): n  
Filtro de body añadido.
```

Cuando tengamos los filtros que queramos ejecutamos la búsqueda.

```
Opciones de filtro:  
1. Buscar por título  
2. Buscar por autor  
3. Buscar por fecha (rango)  
4. Buscar por body  
5. Ejecutar búsqueda  
6. Salir sin buscar  
Elige una opción (1-6):
```

Y vemos cómo nos devuelve una noticia que contiene la palabra junts en el título y constitución en el body.

```

--- Documento 1 ---
date: 2025-12-03T11:21:27.000Z
@timestamp: 2025-12-04T11:26:08.339419023Z
title: Page, sobre las nuevas cesiones de Sánchez a Junts: "Pensar que de rodillas Puigdemont nos va a respetar más es absurdo"
body_length: 2887
fuente: www.elmundo.es
author: EFE (FOTO)
url: https://www.elmundo.es/espana/2025/12/03/69301679fdddf8458b4591.html
body: No es una novedad la posición en contra del miliano García-Pagea las cesiones de Pedro Sánchez a Junts cambio de sus votos en el Congreso, pero este miércole s la ha expuesto de una forma muy gráfica. "[Carles] Puigdemont y los suyos no han respetado al Gobierno cuando les hablamos de pie. Pensar que de rodillas nos va a respetar más es un poco absurdo", ha afirmado el presidente de Castilla-La Mancha.

Así, el barón del PSOE más crítico con la política de alianzas del partido a nivel nacional muestra su disconformidad con la estrategia de acelerar la tramitación e algunas de las exigencias de la formación independentista para tratar de rearmar la mayoría parlamentaria de la investidura y, de paso, cortocircuitar el intento de acercamiento del PPA ella a través de la patronal catalana para que se sume a una eventual moción de censura.

"Nosotros vamos a hacer todo lo posible para no poner en el mercado el concepto de igualdad bajo ningún concepto. De ninguna manera la igualdad y menos por un socialista se puede poner en el mercado de la transacción política", ha recalcado Page antes de intervenir en un acto por el 47º aniversario de la aprobación de la Constitución Española.

Según el dirigente castellanomanchego, "ya se han hecho demasiadas cesiones" a lo largo de la legislatura y considera que "todo el mundo está hartos de los privilegios que se buscan desde el independentismo". A lo que apostilla: "No ha traído nada más que dolor a Cataluña y dolor al conjunto de España".

```

Redundancia y disponibilidad de los datos en el cluster ante la caída de un nodo

Hemos tirado el nodo uno como podemos ver en las siguientes capturas vistas desde el dev tools y el metricbeat.

											Export requests	Import requests
ip	heap.percent	ram.percent	cpu	load_1m	load_5m	load_15m	node.role	master	name			
192.199.1.61	37	97	39	1.19	0.41	0.16	dim	*	nodo-3			
192.199.1.60	22	97	29	0.39	0.21	0.14	dim	-	nodo-2			

Name	Alerts	Status
nodo-1 192.199.1.59:9300	Clear	Offline
nodo-2 192.199.1.60:9300	Clear	Online
nodo-3 192.199.1.61:9300	1 alert	Online

Al tirar el primer nodo y tenerlo configurado de forma que se hagan todo alrededor de los 3 nodos vemos con las siguientes pruebas la redundancia de los datos.

Ejecutamos una consulta que nos de 5 documentos de un indice para ver que se ven los datos.

```
GET noticias-2025.12.02/_search
```

```
{  
  "size": 5,  
  "query": {  
    "match_all": {}  
  }  
}
```

```
max_score : 1,  
"hits": [  
  {  
    "_index": "noticias-2025.12.02",  
    "_id": "DZDb3poB0fjpQZ1tumqr",  
    "_score": 1,  
    "_source": {  
      "url": "https://www.mundodeportivo.com/futbol/atletico-madrid/20251108/1002565950/griezmann-confirmo-tesoro-atletico-banquillo.html",  
      "body_length": 1946,  
      "title": "Griezmann confirmó el tesoro que tiene el Atlético en el banquillo",  
      "body": """Salir y besar el Santo. Eso es lo que hizo Antoine Griezmann para poner el 2-1 en el partido entre el Atlético de Madrid y el Levante. Hasta ese momento, el equipo de Simeone había chocado ante un conjunto granota muy bien plantado, ante el que habían forzado más de una decena de saques de esquina, pero poco más. Ninguno de ellos tuvo consecuencia."}  
    }  
  }  
]
```

Era el minuto 61 cuando el Principito salió al terreno de juego en lugar de Alexander Sorloth. El partido parecía que se le podía atragantar a los colchoneros, y en el primer balón que tocó llegó el gol. Antes de su remate se pudo ver un excelente pase en profundidad de Giuliano Marcos Llorente. Y este puso un centro para que el francés rompiera el partido.

Podemos ver también los índices disponibles. Y vemos como están totalmente accesibles aun con el nodo 1 fuera de servicio.

health	status	index	uuid	pri	rep	docs.count	docs.deleted	store.size
green	open	noticias-2025.12.01	mnEhVJrtQeigX2Rf4iVMNw	3	1	166	0	2.
green	open	noticias-2025.12.02	p3rscZ-HSWWTMyYWRxx9iQ	3	1	12273	0	88.

Vemos también cómo al usar el script de querys con el filtro de título con la palabra clave sanchez sigue funcionando correctamente.

```
--- Documento 8 ---
url: https://elpais.com/espagna/2025-11-09/sanchez-vende-en-colombia-un-mundo-alternativo-al-de-trump.html
body_length: 6724
author: None
date: 2025-11-09T18:15:47.000Z
@timestamp: 2025-12-02T09:23:48.463019056Z
title: Sánchez vende en Colombia un mundo alternativo al de Trump
fuente: elpais.com

--- Documento 9 ---
url: https://www.abc.es/espagna/sanchez-acera-declara-conserva-wasaps-lobato-20251105122031-nt.html
body_length: 7530
author: None
date: 2025-11-05T12:43:39.000Z
@timestamp: 2025-12-02T09:24:11.602476218Z
title: Sánchez Acera declara que no conserva los wasaps con Lobato
fuente: www.abc.es

--- Documento 10 ---
url: https://www.mundodeportivo.com/futbol/betis/20251111/1002567648/joaquin-sanchez-nuevo-consejero-real-betis-balompie.html
body_length: 1533
title: Joaquín Sánchez, nuevo consejero del Real Betis Balompié
date: 2025-11-11T17:51:10.000Z
author: None
fuente: www.mundodeportivo.com
```