



ESCUELA POLITÉCNICA NACIONAL

ESCUELA DE FORMACIÓN DE TECNÓLOGOS



APLICACIONES DISTRIBUIDAS

ASIGNATURA:

Aplicaciones Distribuidas

PROFESOR:

Ing. Vanessa Guevara

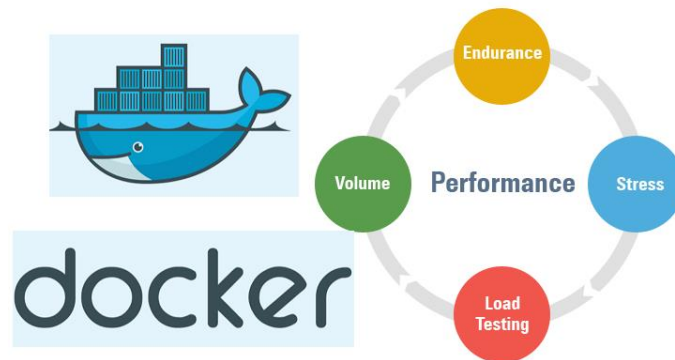
PERÍODO ACADÉMICO:

2023-B

TRABAJO EN CLASE

TÍTULO:

SISTEMAS DE BÚSQUEDA DISTRIBUIDA



Nombre: José Pinos

Fecha: 5/2/2024

1. OBJETIVOS

- Crear un entorno de búsqueda distribuida utilizando Docker para realizar consultas distribuidas.

2. INSTRUCCIONES

PARTE 1: INVESTIGAR SOBRE ELASTICSEARCH

1. ¿Qué es Elasticsearch y cuál es su propósito en los sistemas de búsqueda distribuida?

Es un motor de búsqueda y analítica de RESTful que es capaz de abordar un creciente número de casos de uso, almacena de forma central los datos para una búsqueda más rápida, los datos pueden ser no estructurados y de distintos orígenes. Permite la escalabilidad horizontal lo cual facilita agregar mayor capacidad y fiabilidad a sus nodos y clústeres [1].

2. Investigar sobre los índices, documentos y tipos en Elasticsearch [1].

2.1 Los índices son espacios de nombres lógicos que tienen colecciones de documentos, donde cada documento dentro de un índice es una colección de campos de tipo clave valor que contienen los datos.

2.2 Los documentos son las unidades básicas de información de Elasticsearch y se representan en formato JSON además contienen varios campos para almacenar información de cualquier cosa.

2.3 Los índices de Elasticsearch tenían tipos para organizar los documentos dentro de un índice en forma de campos

3. ¿Cuáles los tipos de datos admitidos y cómo se estructuran los documentos?

3.1 Los tipos de datos core de Elasticsearch son: null, boolean, byte, short, integer, long, unsigned_long, double, float, half_float, scaled_float, keyword type family, text, binary, date, ip y object [2].

3.2 Los documentos se estructuran con el formato JSON donde cada documento puede tener múltiples campos y no se requiere un esquema definido antes de importar los datos [2].

4. ¿Cuál es la sintaxis de las consultas en Elasticsearch?

4.1 Elasticsearch SQL: es una funcionalidad que permite realizar consultas usando SQL, aquí hay 4 fases que son: análisis sintáctico (SQL a AST), traducción (AST a DSL de elastic search), optimización y ejecución de la búsqueda [3].

4.2 DSL Doman Specific Language: que es un lenguaje basado en JSON para realizar consultas, estas consultas son una especie de AST, que tienen 2 tipos de cláusulas: consulta de hoja y consultas compuestas [4].

PARTE 2: IMPLEMENTAR UN SISTEMA DE BÚSQUEDA DISTRIBUIDA UTILIZANDO DOCKER

1. Utilizar Docker para crear un contenedor de Elasticsearch.
2. Utilizar la API de Elasticsearch para indexar datos que deseamos buscar.
3. Una vez que los datos están indexados en Elasticsearch, utilizar su API de búsqueda para realizar consultas simples en los datos indexados.

Comandos docker:

- 1) `docker pull docker.elastic.co/elasticsearch/elasticsearch:7.5.2`
- 2) `docker run -p 9200:9200 -p 9300:9300 -e "discovery.type=single-node"`
`docker.elastic.co/elasticsearch/elasticsearch:7.5.2`

Documento para indexar: empleados1.json

```
{
  "empleados1.json" > ...
  1 { "index": { "_index": "empleados" } }
  2 { "id": 1, "nombre": "Juan", "apellido": "Pérez", "edad": 30, "departamento": "Ventas" }
  3 { "index": { "_index": "empleados" } }
  4 { "id": 2, "nombre": "María", "apellido": "Gómez", "edad": 35, "departamento": "Marketing" }
  5 { "index": { "_index": "empleados" } }
  6 { "id": 3, "nombre": "Pedro", "apellido": "López", "edad": 28, "departamento": "Tecnología" }
  7
```

Indexar empleados1.json

```
C:\Users\APP\DISTIBUIDAS\Desktop\SistemaArchivosDistribuidos>curl.exe -X POST "localhost:9200/empleados/_bulk" -H "Content-Type: application/json" --data-binary "@empleados1.json"
{"took":9884,"errors":false,"items":[{"index":{"_index":"empleados","_type":"doc","_id":"E88ge408taWkb2sA86h1","_version":1,"result":"created","_shards":{"total":2,"successful":1,"failed":1}},{"index":{"_index":"empleados","_type":"doc","_id":"FM8ge408taWkb2sA86h1","_version":1,"result":"created","_shards":{"total":2,"successful":1,"failed":0},"_seq_no":1,"primary_term":1}},{"index":{"_index":"empleados","_type":"doc","_id":"Fc8ge408taWkb2sA86h1","_version":1,"result":"created","_shards":{"total":2,"successful":1,"failed":0},"_seq_no":2,"primary_term":1,"status":201}}]}
```

Buscar un empleado Pedro

```
C:\Users\APP DISTRIBUIDAS\Desktop\SistemaArchivosDistribuidos>curl.exe -X GET "localhost:9200/empleados/_search" -H "Content-Type: application/json" '{"query":{"match":{"nombre":"Pedro"}}}'
{"took":31,"timed_out":false,"shards":{"total":1,"successful":1,"skipped":0,"failed":0},"hits":{"total":{"value":3,"relation":"eq"},"max_score":1.0,"hits":[{"_index":"empleados","_type":"doc","_id":"E88ge408taKb2sA86h1","_score":1.0,"_source":{"id":1,"nombre":"Juan","apellido":"Perez","edad":30,"departamento":"Ventas"}},{"_index":"empleados","_type":"doc","_id":"FW8ge408taKb2sA86h1","_score":1.0,"_source":{"id":2,"nombre":"Maria","apellido":"Gomez","edad":35,"departamento":"Marketing"}},{"_index":"empleados","_type":"doc","_id":"Fc8ge408taKb2sA86h1","_score":1.0,"_source":{"id":3,"nombre":"Pedro","apellido":"Lopez","edad":28,"departamento":"Tecnologia"}]}]}]}curl: (3) nested brace in URL position 11:
{"query":{"match":{"nombre":"Pedro"}}}
```

Buscar un empleado Juan

```
C:\Users\APP DISTRIBUIDAS\Desktop\SistemaArchivosDistribuidos>curl.exe -X GET "localhost:9200/empleados/_search" -H "Content-Type: application/json" '{"query":{"match":{"nombre":"Juan"}}}'
{"took":11,"timed_out":false,"shards":{"total":1,"successful":1,"skipped":0,"failed":0},"hits":{"total":{"value":3,"relation":"eq"},"max_score":1.0,"hits":[{"_index":"empleados","_type":"doc","_id":"E88ge408taKb2sA86h1","_score":1.0,"_source":{"id":1,"nombre":"Juan","apellido":"Perez","edad":30,"departamento":"Ventas"}},{"_index":"empleados","_type":"doc","_id":"FW8ge408taKb2sA86h1","_score":1.0,"_source":{"id":2,"nombre":"Maria","apellido":"Gomez","edad":35,"departamento":"Marketing"}},{"_index":"empleados","_type":"doc","_id":"Fc8ge408taKb2sA86h1","_score":1.0,"_source":{"id":3,"nombre":"Pedro","apellido":"Lopez","edad":28,"departamento":"Tecnologia"}]}]}]}curl: (3) nested brace in URL position 11:
{"query":{"match":{"nombre":"Juan"}}}
```

4. CONCLUSIONES

- 1) Indexar la información es una manera de hacer que la búsqueda de archivos o datos sea más rápida al estar organizada de acuerdo a un criterio
- 2) Elasticsearch es un motor de búsqueda que trabaja de manera eficiente con los documentos de tipo json y peticiones HTTP
- 3) Elasticsearch también permite trabajar con el lenguaje de consulta SQL y su adaptación Elasticsearch SQL

BIBLIOGRAFÍA

- [1] Elasticsearch, «El corazón del Elastic Stack, gratuito y abierto,» Elasticsearch, [En línea]. Available: <https://www.elastic.co/es/elasticsearch>. [Último acceso: 5 febrero 2024].
- [2] Elasticsearch, «Data Types,» Elasticsearch, [En línea]. Available: <https://www.elastic.co/es/elasticsearch>. [Último acceso: 5 Febrero 2024].
- [3] Medium, «Cómo realizar consultas en ELASTICSEARCH,» Medium, [En línea]. Available: <https://medium.com/nिकासource/c%C3%B3mo-realizar-consultas-en-elasticsearch-92ad5e3dd460>. [Último acceso: 5 Febrero 2024].
- [4] Elasticsearch, «Características de Elasticsearch,» Elasticsearch, [En línea]. Available: <https://www.elastic.co/es/elasticsearch/features>. [Último acceso: 5 Febrero 2024].