ELK: Elasticsearch (conceptos)

indice

- colección de documentos con características similares
- identificado por un nombre (minúsculas)

documento

- unidad básica de información
- formato JSON
- en un índice se almacenan todos los documentos que queramos

shards

- paquetizar y distribuir la información (indice) entre los nodos
    - permite subdividir/escalar la información almacenada
    - paralelización operaciones

réplica

- permite réplicas de los shards
- HA
- paralelización operaciones en las réplicas
- cada índice tiene shard primario y réplicas
- 2.174.000.000 documentos por shard
- los shards no se pueden modificar, las réplicas si

BDD distribuida

- información almacenada entre los nodos
- transparente para el usuario
- independencia OS
- procesamiento distribuido de consultas
- información fragmentada
- HA

tipos de nodos

- instancia elasticsearch = nodo
- interconectados, cluster
- peticiones HTTP REST API
- comunicación interna

Master Node

- crear/borrar índices
- nodos que forman parte del cluster
- decide que nodo aloja cada shard
- indexado de búsqueda de datos: CPU, RAM, I/O
- buena práctica: master dedicado.

Data Node

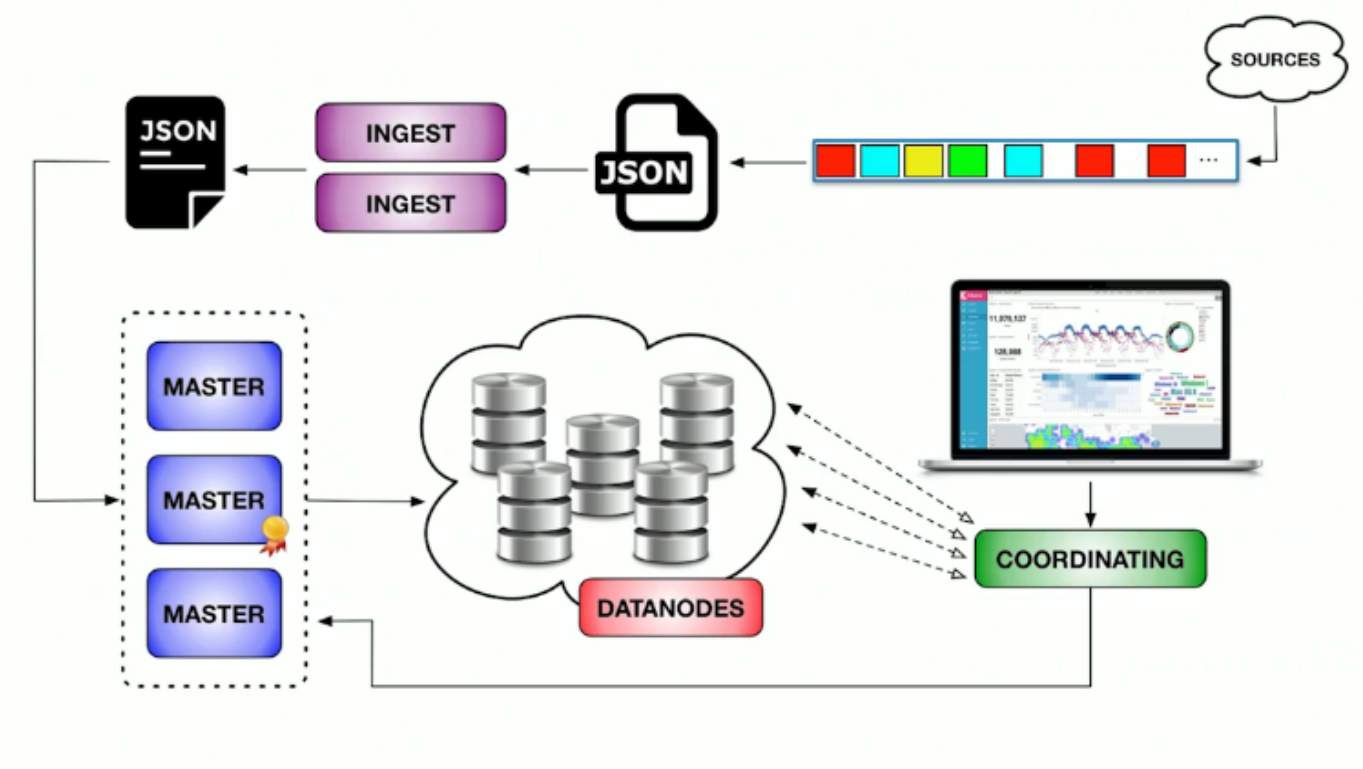
- almacenan los shards
- operaciones: CRUD, búsquedas, agregaciones
- importante monitorizar CPU, RAM, I/O
- añadir más datanodes si se sobrecargan para balancear procesamientos y reparto de datos

Ingest Node

- pre-procesamiento de la información entrante antes de su indexado
- buena prácica: ingest dedicados
- equivalente logstash

Coordinating Node

- encaminar peticiones, respuestas de búsquedas, distribuir indexados → balanceador de carga inteligente
- ayudan a descargar la tarea de coordinación de los Master Nodes
- usan el estado del cluster para encaminar búsquedas y peticiones allá donde toque



1.1_introduccion_elasticsearch.pdf