Labs
« Operate an ElasticSearch cluster»

Prerquesite :

- Good Internet Connexion
- OS : Linux, MacOs, Windows 10
- YAML Editor : (VSCode, Atom, ...)
- Optional : Docker, Git

Lab0 : Installation

0.1 Getting started

Check your free disk space (you must have 20 % of your disk space free)

Download and unzip the latest release of Elastic Search

Start the server with `$ES_HOME/bin/elasticsearch`

If release is superior to 8, look at the trace and save all the information about password and enrollment token

Access to http(s)://localhost:9200

0.2 Configuration file and logs

Edit the main elasticsearch configuration file and modify the following properties:

- Cluster name
- Node name
- Listening address (put your public address there)

Try to start the server and observe the bootstrap checks traces, perform necessary fixes if necessary. (See https://www.elastic.co/guide/en/elasticsearch/reference/current/bootstrap-checks.html)

Change trace level to WARN

Set the ***node.name*** property via the command line when starting the server

0.3 Kibana Installation

Download a Kibana distribution with the same version number.

Unzip the archive and start Kibana (if 8.x +, with the enrollment token)

Access to http://localhost:5601 search for the *Dev Console*.

Execute the following queries :

```
GET /_cluster/health
GET /_search
GET /_cat/nodes
```

Lab1 : Cluster, nodes, shards, replica

1.1 Cluster, nodes, shard and replica

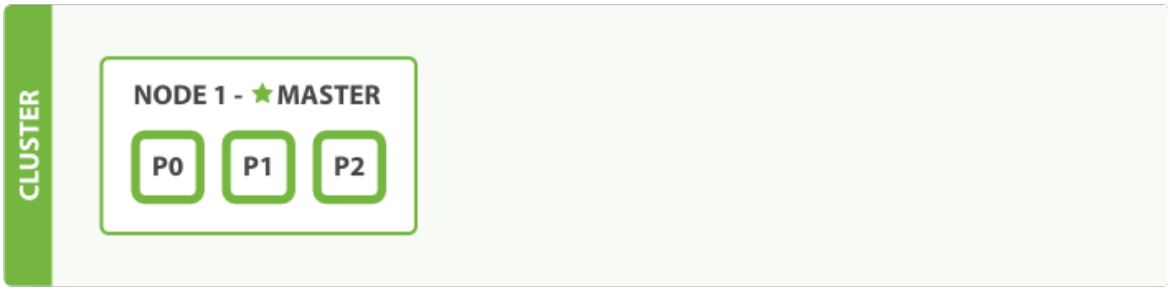
- Create an enrollment token with :
  bin/elasticsearch-create-enrollment-token -s node
- Uncomment the *transport.host* setting at the end of *config/elasticsearch.yml*.
- Restart Elasticsearch.

- Execute :
  `GET /_cluster/health?pretty`
  How many nodes, shards are available ?

- Create an index named *blogs*

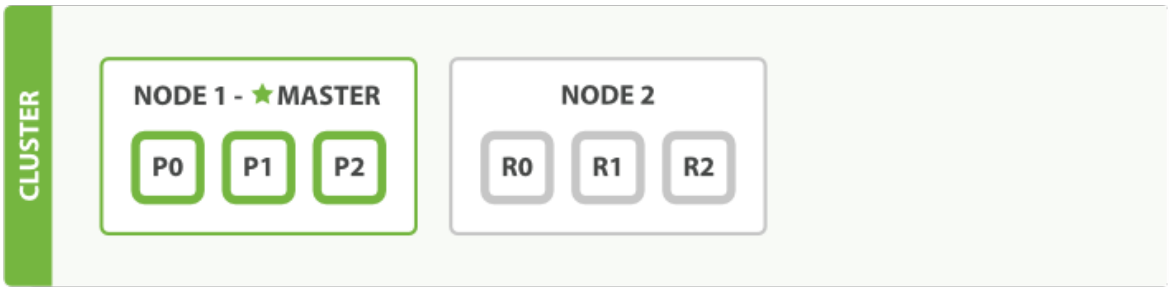
```
PUT blogs
{
 "settings" : {
 "number_of_shards" : 3,
 "number_of_replicas" : 1
 }
}
```

- Re-execute *_cluster/health?pretty*
  Health status color ? How many shards available, active ?



- Unzip the distribution in another location
- Start a second node with the previous enrollment token :
  bin/elasticsearch --enrollment-token <token>

- Re-execute `_cluster/health?pretty`
  Health status color ? How many shards available, active ?



- Start a third node
  Health status color ? How many shards available, active ?
- Increase the number of replica
  ```
  PUT /blogs/settings
  { "number_of_replicas" : 2 }
  ```

- Stop the first node
- Status health of the cluster ?
- Restrart the first node

1.2 Disabling security

For the following labs, we are disabling security. In 8.x+, security is enabled by default.

To disable it, you have to :

- Set xpack.security.enabled: false
- Comments all the properties relative to *xpack.security* in *elasticsearch.yml*
- Remove properties stored in *elasticsearch* keystore.
  You can do it with :
  ```
  bin/elastisearch-keystore remove <name-of-the-setting>
  ```

In *kibana.yml* comment all properties related to *xpack.security*

Lab2 : Ingestion of office documents

2.1 Ingestion with dynamic mapping

Installation of ingest-attachment plugin

- Install the ***ingest-attachment*** plugin
  `./elasticsearch-plugin install ingest-attachment`

- Restart Node

Pipeline creation

- With the API, create a pipeline with a single processor *attachment*

Indexing

- Use the provided program to index all the provided documents :

```
$ES_HOME/jdk/bin/java -jar ingest.<version>.jar
<directory_to_ingest> <index> <pipeline> [<http_host> <http_port>]
```

- Check the number of indexed documents

2.1 Ingestion with predefined mapping

Create a second index with the following configuration (settings and mappings) :

```
PUT /office2
{
  "settings": {
    "number_of_shards": 3,
    "number_of_replicas": 1
  },
  "mappings": {
    "properties": {
      "attachment": {
        "properties": {
          "content": {
            "type": "text",
            "analyzer": "french",
            "fields": {
              "en": {
                "type": "text",
                "analyzer": "english"
              }
            }
```

```
            },
            "content_length": {
              "type": "long"
            },
            "content_type": {
              "type": "keyword"
            },
            "date": {
              "type": "date"
            },
            "language": {
              "type": "keyword"
            },
            "title": {
              "type": "text",
              "analyzer": "french",
              "fields": {
                "en": {
                  "type": "text",
                  "analyzer": "english"
                },
                "keyword": {
                  "type": "keyword",
                  "ignore_above": 256
                }
              }
            },
            "name": {
              "type": "keyword"
            },
            "author" : {
                "type" : "text"
            }
          }
        }
      }
    }
}
```

Reindex the previous documents into this index

Lab3 : Queries

3.1 Search Lite

Perform the following queries :

• Documents responding to "Java"

```
GET /<index>/_search?q=java
```

• Documents not responding to "Java"

```
GET /<index>/_search?q=-java
```

• Limit the documents returned from the first request

```
GET /<index>/_search?q=-java&size=5
```

• Documents whose content match "Java", excluding attachment.content in the response

```
GET /<index>/_search?
q=attachment.content:java&_source_excludes=attachment.content
```

• PDF documents with content responding to "Java"

```
GET /<index>/_search?q=attachment.content:java AND
attachment.content_type:pdf&_source_excludes=attachment.content
```

• Documents created after a particular date and whose content matches "Java Elastic Search" but not "Administration"

```
GET /<index>/_search?q=attachment.date:%3E2016-10-01+%2B(Java
Elastic Search)&_source=name,attachment.date&pretty
```

3.2 DSL

Perform DSL queries based on office index :

• PDF documents sorted by date

```
GET /doc/_search
{
  "query": {
    "bool": {
      "filter": {
        "term": {
          "attachment.content_type": "application/pdf"
        }
      }
    }
  },"sort": [
    {
      "attachment.date": {
        "order": "desc"
      }
```

```
    }
  ]
}
```

• Documents whose content or title field responds to "administration"

```
GET <index>/_search?_source=name,attachment.title
{
  "query" : {"multi_match" : { "query" : "administration", "fields" :
["attachment.title", "attachment.content"] } }
}
```

• Documents whose content or title field responds to "administration" and whose creation date falls within a range

```
GET <index>/_search?_source=name,attachment.title
{
  "query" : {
    "bool" : {
      "must" : { "multi_match" : { "query" : "administration", "fields" :
["attachment.title", "attachment.content"] } },
      "filter" : {
        "range" : {  "attachment.date": {   "gte": "2016-01-01",    "lt": "2016-
12-31" } }
      }
    }
  }
}
```

• Documents whose content field responds to "Administration" and optionally "Oracle"

```
GET <index>/_search?_source=name,attachment.title
{
  "query" : {
    "bool" : {
      "must" : { "match" : { "attachment.content" :"Administration"} },
      "should" : { "match" : { "attachment.content" : "Oracle" } }
    }
  }
}
```

• Document whose title starts with Adm*

```
GET /<index>/_search?_source=name,attachment.title
{
  "query" : {
    "prefix": {
      "attachment.title": {
        "value": "adm"
      }
    }
  }
}
```

• Retrieve documents containing the phrase "java framework", allowing 5 word distance

```
GET /office2/_search?_source=name,attachment.title
{
  "query" : {
    "match_phrase": {
      "attachment.content.en": {
        "value": "java framework",
        "slop": 5
      }
    }
  }
}
```

- Fuzzy search with typo

```
GET /office2/_search?_source=name,attachment.title
{
  "query" : {
    "fuzzy": {
      "attachment.author": {
        "value": "java framework",
        "slop": 5
      }
    }
  }
}
```

- Highlighting

```
GET <index>/_search?_source=name,attachment.title
{
  "query": {
    "bool": {
      "must": {
        "match": {
          "attachment.content": "Administration"
        }
      },
      "should": {
        "match": {
          "attachment.content": "Oracle"
        }
      }
    }
  },
  "highlight": {
    "fields": {
      "attachment.content": {}
    }
  }
}
```

3.3 Aggregations

Execute agregation query :

- By type of documents

```
GET <index>/_search
{
  "aggs": {
    "type_doc": {
      "terms": {
        "field": "attachment.content_type"
      }
     }
  }
}
```

- By language

```
GET <index>/_search
{
  "aggs": {
    "langue": {
      "terms": {
        "field": "attachment.language"
      }
    }
```

```
  }
}
```

- By both

```
GET <index>/_search
{
  "aggs": {
    "type_doc": {
      "terms": {
        "field": "attachment.content_type"
      },
      "aggs": {
        "langue": {
          "terms": {
            "field": "attachment.language"
          }
        }
      }
    }
  }
}
```

- Average size by year

```
GET <index>/_search
{
  "aggs": {
    "date": {
      "date_histogram": {
        "field": "attachment.date",
        "interval": "year",
        "format": "yyyy-MM-dd",
        "min_doc_count": 1
      },
    "aggs": {
      "moyenne": {
        "avg": {
          "field": "attachment.content_length"
        }
      }
    }
    }
  }
}
```

Lab4 : Single-node architecture

Write a basic script that allows to start elastic search from a different config location.

Change the following properties :

- Names of the cluster and of the node
- Data and log directories
- Network host
- The discovery (single-node)

Check that all the boorstrap checks are passed and no discovery process has been started

Check the heap of the JVM and fix it if needed

Lab5 : Fault-tolerant architectures

For the followsing labs, you can use different machines or docker-compose

5.1 3 interchangeable nodes

Set up an architecture with interchangeable nodes and one instance of kibana. Start the cluster and access folowing URLs :

```
GET _cat/nodes
GET _cluster/health
<kibana>:5601
```

5.2 Specialized nodes

Set up an architecture with :

- 3 eligible master nodes
- 3 data nodes
- One coordination nore
- One instance of kibana

Lab6 : Monitoring

6.1 Logs

See the different log files generated
Update the configuration to enable DEBUG trace
Use the provided JMeter file to load the cluster. Lok at the slow log

6.2 Low-level Cluster API

Execute request to monitor

- indices
- nodes

Without a restart increase the level of log

6.3 Monitoring Cluster

6.3.1 Set up a clusters
Set up a start-up script which starts a single-node monitoring cluster on another port.
Disable moniotiring and security for this cluster.
Set up a start up script which starts an instance of kibana connected to this single-node

Enable monitoring for the production cluster
xpack.monitoring.collection.enabled

6.3.2 Set up a metricbeat
Download the same release number of ***mericbeat***

Edit the ***metricmeat.yml*** and specify the adress of the monitoring cluster for the *elasticsearch.output* property
Enable elasticsearch-xpack module and configure the IP adress of the production cluster in modules.d/elasticsearch-xpack.yml
Disable system module

Start metricbeat and checks :

- The logs of metricbeat
- The indices created in the monitoring cluster
- The data in the monitoring kibana instance

6.3.3 Set up a filebeat
Download the same release number of ***filebeat***

Edit the ***filebeat.yml*** and specify the adress of the monitoring cluster for the *elasticsearch.output*

property and the kibana adress dedicated to monitoring
Enable ***elasticsearch*** module and load dashboards in kibana with : *filebeat setup -e*
Check dashboards in kibana

Configure the elasticsearch module located in *config.d/elasticsearch.yml*

Start filebeat and checks :

- The logs of filebeat
- The indices created in the monitoring cluster
- The data in the monitoring kibana instance

6.3.4 Load the cluster

You can use the provided *JMeter* file to load the cluster and see something in the kibana's dashboards

Lab 7 Operation

7.1 Operations

BackUp and Restore

Perform an index backup then restore

Update index configuration

Create an index aliases named « *office* » which point to one your office index, say ***office1***

Access documents by this index, you can use the previous JMeter file

Reindexing *office1* into *office2*

Update your alias in order to point to *office2*

7.2 Resizing and restarting

On a multi-node cluster perform a rolling restart

Lab 8 Security

8.1 Manual configuration

8.1.1 Authenticating users
Turn on security and reset the passwords of the built-in users.

Store the kibana password in a keystore

Restart kibana and access to the Administration console in order to create new users

8.1.2 TLS between nodes
Genrate the certificate and private key that will be used for the nodes and copy it on each node of your cluster.

Enable transport ssl in node's configuration and update the following properties :
xpack.security.transport.ssl.verification_mode: certificate
xpack.security.transport.ssl.client_authentication: required
xpack.security.transport.ssl.keystore.path: elastic-certificates.p12
xpack.security.transport.ssl.truststore.path: elastic-certificates.p12

Store optional passwords into a keystore

Restart the cluster and check the logs

8.1.3 https for elasticsearch
Use ***elasticsearch-certutil*** to generate an archive for https containing a self-signed certificate

Use this archive to copy certificate on each configuration directories of node(s)

Enable https and configure the following property in elasticsearch.yml :
xpack.security.http.ssl.keystore.path: http.p12

Add the optional password in the keystore

Copy the *elasticsearch-ca.pem* in the conf of Kibna update the following property :
elasticsearch.ssl.certificateAuthorities
elasticsearch.hosts

8.1.4 https for kibana
See slides