Altair - Elasticsearch Plug-in Configuration Guide

Contents

1 Supported versions of Elasticsearch and Opensearch

In principle, The Plug-in works with the different versions of Elasticsearch and Opensearch. The plugin tested with the following versions to ensure its proper functioning. Grafana v10.0.4 is used for visualizing the data in dashboard for that test.

1.1 Supported Elasticsearch versions

- 7.16.2
- 8.6.0
- 8.8.2

1.2 Supported Opensearch versions

Elasticsearch's major version is given as **es-version** in Mistral config, which is needed to support the different releases of Elasticsearch. Opensearch follows different versioning from Elasticsearch, but it has to be provided with a compatible Elasticsearch's major version to work accordingly. The supported Opensearch version and its compatibility with Plug-in is provided as below.

Opensearch Version	Plug-in version	es-version in Mistral config
1.3.14	7.1.4 or latest	7 or 8
2.11.1	7.1.4 or latest	8

2 Installing the Elasticsearch Plug-in

Extract the Mistral plug-ins archive that has been provided to you somewhere sensible. Please make sure that you use the appropriate version of the plug-ins for the architecture of the machine on which the plug-in will run.

In addition, if the Mistral Plug-ins package was obtained separately from the main Mistral product, please ensure that the version of the plug-in downloaded is compatible with the version of Mistral in use. Version v7.2.8 of the Elasticsearch Plug-in as described in this document is compatible with all versions of Mistral compatible with plug-in API version 7. At the time of writing this is Mistral 2023.1.0 to 2025.2.0.

The Elasticsearch plug-in can be found in, for example:

`<installation directory>/output/mistral_elasticsearch_v7.2.8/x86_64/`

for the 64 bit Intel compatible version, and in:

`<installation directory>/output/mistral_elasticsearch_v7.2.8/aarch64/`

for the 64 bit ARM compatible version.

The plug-in must be available in the same location on all execution hosts in your cluster.

2.1 Installing the Elasticsearch Index Mapping Template

Prior to using the Mistral Elasticsearch plug-in the index datatype mapping template should be configured within Elasticsearch. The files mappings_<n>.x.json contain the appropriate configuration for Elasticsearch installations where <n> corresponds to the major version of Elasticsearch in use.

The provided templates ensure that date fields are correctly identified. The template for Elasticsearch version 2.x also specifies some key fields that should not be analysed to make working with Grafana simpler. Other, site specific, configuration can be added at the user's discretion.

We recommend using date based indicies for storing Mistral data, but we also support numbered indicies. If you choose date based indices, the plug-in will attempt to create this automatically for you when it starts up. If you choose to use numbered indicies then you will need to run mistral_create_elastic_template.sh manually.

If you have access to curl the file mistral_create_elastic_template.sh can be run to create the template from the command line. This script will attempt to detect the Elasticsearch version in use and create the template using the appropriate mapping file.

The script takes the following command line options:

`-?, --help`

Display usage instructions

`-i idx_name, --index=idx_name`

The basename of the index. If not specified the template will be created for indexes called mistral. If a custom value is used here a matching option must be provided to the plug-in.

`-c cacert-path, --cacert=cacert-path`

CA certificate file to use to verify the peer, in pem format.

`-h hostname, --host=hostname`

The name of the machine on which Elasticsearch is hosted. If not specified the script will use localhost.

`-N "host1,[host2,...]", --noproxy="host1,[host2,...}"`

Comma separated list of hosts that do not require a proxy.

`-p filename, --password=filename`

The name of a file containing the password to be used when creating the index if needed. The password must be on a single line.

`-P n, --port=n`

The port to be used when connecting to Elasticsearch. If not specified the script will use port 9200.

`-s, --ssl`

Use HTTPS protocol instead of HTTP to connect to Elasticsearch.

`-k, --skip-ssl-validation`

Disable SSL certificate validation when connecting to Elasticsearch

`-u user, --username=user`

The username to be used when connecting to Elasticsearch if needed.

`-d, --date`

Use date based index naming. By default the indexes are named idx_name-00000N, if you would prefer to use idx_name-YYYY-MM-DD then specify this option.

3 Configuring Mistral to use the Elasticsearch Plug-in

Please see the Plug-in Configuration section of the main Mistral User Guide for full details of the plug-in configuration specification. Where these instructions conflict with information in the main Mistral User Guide please verify that the plug-in version available is compatible with the version of Mistral in use and, if so, the information in the User Guide should be assumed to be correct.

3.1 Mistral Plug-in Configuration

The Mistral plug-in configuration is in YAML and goes in the same file as the main Mistral Configuration File. The plug-in is declared with the `plugin` mapping and requires at minimum a `path` key-value pair. All of the specified settings are members of the `plugin` mapping.

```
plugin:
    path: plugins/mistral_elasticsearch/x86_64/mistral_elasticsearch
```

This section describes the specific settings required to enable the Elasticsearch Plug-in.

3.1.1 Path

The `path` key must be set to the path of the elasticsearch plguin-in executable. This must be either absolute or relative to the `MISTRAL_INSTALL_DIRECTORY` environment variable and needs to be accessible and the same on all hosts. Environment variables are not supported in the value.

```
path: plugins/mistral_elasticsearch/x86_64/mistral_elasticsearch
```

3.1.2 Interval

The `interval` key takes a single integer value parameter. This value represents the time in seconds the Mistral application will wait between calls to the specified plug-in e.g.

```
interval: 300
```

The value chosen is at the discretion of the user, however care should be taken to balance the need for timely updates with the scalability of the Elasticsearch installation and the average length of jobs on the cluster.

3.1.3 Options

The `options` mapping is optional and lists all options to be passed to the plug-in as command line arguments to the executable. A full list of valid options for this plug-in can be found in the section Plug-in Configuration File Options. The order of options is not preserved. These values are passed to the plug-in executable as `--key=value`. For example,

```
options:
    host: hostname
    error: filename
```

will pass to the plug-in executable the command line arguments `--host=hostname` and `--error=filename`.

3.1.4 Switches

The `switches` mapping is optional and lists all switches to be passed to the plug-in as command line arguments to the executable. A full list of valid switches for this plug-in can be found in the section Plug-in Configuration File Switches. The order of switches is not preserved. Switches not present are presumed to be off. These switches are passed to the plug-in executable as `--key`. For example,

```
switches: ["date", "ssl"]
```

will pass to the plug-in executable the command line arguments `--date` and `--ssl`.

3.2 Plug-in Configuration File Options

The following configuration file key-value options are supported by the Elasticsearch plug-in.

`error: file`

Specify location for error log. If not specified all errors will be output on `stderr` and handled by Mistral error logging.

`cacert: file`

CA certificate file to use to verify the peer, in pem format.

`host: hostname`

The hostname of the Elasticsearch server with which to establish a connection. If not specified the plug-in will default to `localhost`.

`index: index_name`

Set the index to be used for storing data. This should match the index name provided when defining the index mapping template (see the section Supported Elasticsearch Versions). The plug-in will create indexes named `<idx_name>-yyyy-MM-dd`. If not specified the plug-in will default to `mistral`.

`mode: octal-mode`

Permissions used to create the error log file specified by the -e option.

`noproxy: "host1,[host2,...]"`

Comma separated list of hosts that do not require a proxy.

`password: secret`

The password required to access the Elasticsearch server if needed.

`port: number`

Specifies the port to connect to on the Elasticsearch server host. If not specified the plug-in will default to `9200`.

`username: user`

The username required to access the Elasticsearch server if needed.

`es-version: num`

The major version of the Elasticsearch server to connect to. If not specified the plug-in will default to “5”.

3.3 Plug-in Configuration File Switches

The following configuration file switches are supported by the Elasticsearch plug-in.

`ssl`

Connect to the Elasticsearch server via secure HTTP.

`skip-ssl-validation`

Causes the plug-in to not validate any self-signed CA certificates.

`date`

Use date based index naming. By default the indexes are named idx_name-00000N, if you would prefer to use idx_name-YYYY-MM-DD then specify this option. Recommended.

`skip-index-templates`

Causes the plug-in to not create any index templates. This should not be used during an upgrade of Mistral to ensure new index templates are defined. It is provided in the case that connections to ElasticSearch are slow.

4 Mistral's Elasticsearch Document Model

This section describes how the Mistral Elasticsearch Plug-in stores data within Elasticsearch.

The Mistral Elasticsearch Plug-in will create indexes with a trailing index number `mistral-00000N`. This allows for easy managing of the indexes in Kibana using lifecycle policies and rollover. Previously the default was to create indexes with a date appended, by default these would be named `mistral-YYYY-MM-DD`. This allows for easy manual management of historic data. The old behaviour can be restored with the `-d` option if preferred.

Documents are inserted into these indexes in the same format as output by Mistral, please refer to the Mistral Manual for details. By default, the only explicit type mapping defined is for `timestamp` which is set to be a date field.

There are five indices created in total, each for different types of data, and therefore with different recommended retention policies. In the below table describing each index and the data within, the default names are used.

Index name	Description	Amount of Data	Lifecycle Policy considerations
mistral	Time series data for file use and resources. This is the historical index in previous versions of Mistral.	One record per mount point used, per job, per timeframe + one memory/CPU record per job.	This index can be filled with a lot of data very quickly, so should be moved to long term storage or deleted after not very long.
network-mistral	time series network statistics similar to the above, will only exist if network output is turned on	One record per network IP, per job, per timeframe.	As above.
total-mistral	Written at the end of the job if totals are enabled, totals of the time series data above. This data was previously in the main index.	There is one record per mount point used + one memory/CPU record per job.	This index generally fills up much slower than the previous two, but may be just as quick when many short jobs are run. This data tends to be easier to use when comparing jobs as a whole but can be safely kept around for a while.
total-network-mistral	As above, but for the network data. Will only be output if network is turned on.	There is one record per network IP used, per job.	As above.

Index name	Description	Amount of Data	Lifecycle Policy considerations
summary-mistral	Written at the end of job, if totals are enabled. This data used to be called the traffic light report and previously only went to CSV. This divides the I/O into good/medium/bad categories and reports the counts and durations (which are sampled) of those operations.	One record per mountpoint per job + one more per job.	This index fills up the slowest, being proportional to the number of jobs run. This is also useful for comparing jobs as a whole and can be safely kept for some time.

4.1 Lifecycle Policies

Below are some examples of index lifecycle policies when considering date based indices. For date based indices, each index is no longer written to after the day is over and therefore is only used for searching from that point on. For further information about lifecycles, please consult the official Elasticsearch documentation.

This first policy is suitable for the time series indices as they are filled in very quickly. It moves data out of fast storage when the day is over, slower storage after a week and deletes it adter a year.

```
{
  "policy": {
    "phases": {
      "hot": {
        "min_age": "0ms",
        "actions": {}
      },
      "warm": {
        "min_age": "1d",
        "actions": {
          "set_priority": {
            "priority": 50
          }
        }
      },
      "cold": {
        "min_age": "7d",
        "actions": {
          "set_priority": {
            "priority": 0
          }
        }
      },
      "delete": {
        "min_age": "30",
```

```
        "actions": {
          "delete": {
            "delete_searchable_snapshot": true
          }
        }
      }
    }
  }
}
```

This second policy is more suitable for the total indices, due to their greater use for comparisons between jobs. It moves data out of fast storage after a week, into slow storage after thirty days and deletes it after year.

```
{
  "policy": {
    "phases": {
      "hot": {
        "min_age": "0ms",
        "actions": {
          "set_priority": {
            "priority": 100
          }
        }
      },
      "warm": {
        "min_age": "7d",
        "actions": {
          "set_priority": {
            "priority": 50
          }
        }
      },
      "cold": {
        "min_age": "30d",
        "actions": {
          "set_priority": {
            "priority": 0
          }
        }
      },
      "delete": {
        "min_age": "365d",
        "actions": {
          "delete": {
            "delete_searchable_snapshot": true
          }
        }
      }
    }
  }
}
```