# Demos
# « ElasticSearch»

## Demo1 : Cluster, nodes, shards, replica

With a cluster of 3 nodes, we can tolerate one failure

Start only 2 nodes

```
 GET /_cat/nodes
```

```
GET /_cluster/health?pretty
```

How many nodes, shards are available ?

Create an index

```
PUT /blogs
{
 "settings" : {
 "number_of_shards" : 3,
 "number_of_replicas" : 2
 }
}
```

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

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

## Demo2 : metricbeats and predefined Kibana dashboards

Start metricmeat, look at the creation of the index

Access the predefined dashboard : Overview of system metrics

http://localhost:5601/app/dashboards#/view/Metricbeat-system-overview-ecs?_g=(filters:!(),refreshInterval:(pause:!t,value:0),time:(from:now-15m,to:now))

# Demo3 : XML Ingestion with logstash

3.1 Ingest of one XML document similar to EUR-LEX

3.2 Ingesting Apache logs

Creation of an index template

Develop a pipeline :config_apach.conf

# Demo4 : Ingestion of office documents

## *Installation of ingest-attachment plugin*

On each node :

```
bin/elasticsearch-plugin install ingest-attachment
```

Then restart each node

## *Pipeline creation*

```
PUT /_ingest/pipeline/attachment
{
  "description" : "Extract attachment information",
  "processors" : [
    {  "attachment" : { "field" : "data" } },
    {  "remove" : { "field" : "data" } }
  ]
}
```

## *Indexing*

- Use the provided program to index all the provided documents :
- Check the number of indexed documents

# Demo4 : Analyzers

4.1 Testing predefined analyzers

4.2 A custom analyzer on office4

Par exemple :

```
PUT /docs4/
{
  "settings": {
    "number_of_shards": 2,
    "number_of_replicas": 1,
    "analysis": {
      "filter": {
        "french_elision": {
          "type": "elision",
          "articles_case": true,
          "articles": [
            "l",
            "m",
            "t",
            "qu",
            "n",
            "s",
            "j",
            "d",
            "c",
            "jusqu",
            "quoiqu",
            "lorsqu",
            "puisqu"
          ]
        },
        "french_stop": {
          "type": "stop",
          "stopwords_path": "myFrenchStop.txt"
        },
        "french_synonym": {
          "type": "synonym",
          "synonyms_path": "mySynonym.txt"
        },
        "french_minimal_stemmer": {
          "type": "stemmer",
          "language": "minimal_french"
        }
      },
      "analyzer": {
        "my_french": {
          "tokenizer": "standard",
          "filter": [
            "french_elision",
            "lowercase",
            "french_stop",
            "french_synonym",
            "french_minimal_stemmer"
          ]
        }
      }
    }
  },
  "mappings": {
    "properties": {
      "attachment": {
        "properties": {
          "content": {
            "type": "text",
            "analyzer": "my_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": "my_french",
            "fields": {
              "en": {
                "type": "text",
                "analyzer": "english"
              },
              "keyword": {
                "type": "keyword",
                "ignore_above": 256
              }
            }
          },
          "name": {
            "type": "keyword"
          }
        }
      }
    }
  }
}
```

## Demo5: DSL Syntax

Perform DSL queries based on office index :

- PDF documents sorted by date
- Documents whose content field responds to "administration"
- Documents whose content or title field responds to "administration"
- Documents whose content or title field responds to "administration" and whose creation date falls within a range
- PDF documents whose content or title field responds to "administration" and whose creation date falls within a range
- Documents whose content field responds to "Administration" or "Oracle"
- Documents whose content field responds to "Administration" and optionally "Oracle"

### *5.2 Control relevance*

Use the ***explain*** parameter:

Influence the score with boosting and function_score :

```
GET /office2/_search
{
  "query": {
    "function_score": {
      "query": {
        "match": {
          "attachment.content": "elasticsearch"
        }
      },
      "script_score": {
        "script": {
          "source": "Math.log(2 + 
doc['attachment.content_length'].value)"
        }
      }
    }
  }
}

GET /_search
{
  "query": {
    "function_score": {
      "query": {
        "match": {
          "attachment.content": "elasticsearch"
        }
      },
      "field_value_factor": {
        "field": "attachment.content_length",
        "factor": 1.2,
        "modifier": "sqrt",
        "missing": 1
      }
    }
  }
}
```

### 5.3 Partial matching

- Prepare a new index that uses multiple indexing for the field *attachment.title* :
  - *keyword* DataType
  - *text* with *standard* analyzer
  - *text* with the *edge_ngram* analyzer
- Perform partial matching requests using one of the 3 mapping fields. Compare results and response times

### 5.4 Phrases

Perform queries with phrases:

- Retrieve documents containing the phrase "java framework", allow 5 word distance
- Retrieve documents whose title begins with "administration j"

### 5.5  Fuzzy, Natural language

- Perform fuzzy searches with typos
- Optional : Prepare a new index with a phonetic filter, perform searches with misspellings

### 5.6 Highlighting

Perform previous queries by adding highlight on content field (limit fragment size)

## Demo6 : Agregations

- Execute agregation query :
  - By type of documents
  - By language
  - By both

- Define buckets by size of documents

- Find the average size of a document
- Find the averages size of document by type sorted with the highest value
- Average size by year
- Average size of PDF and .odp docs

# Demo7 : Kibana Dashboard

## *7.1 Creation from scratch of a Kibana Dashboard to exploit Apache Data*

- Les KPI
  - Total of hits
  - Distinct Ips
  - Average size of a request
- Histogram of frequentation with a sub-bucket of return code
- A map

## 7.2 Timelion visualizations

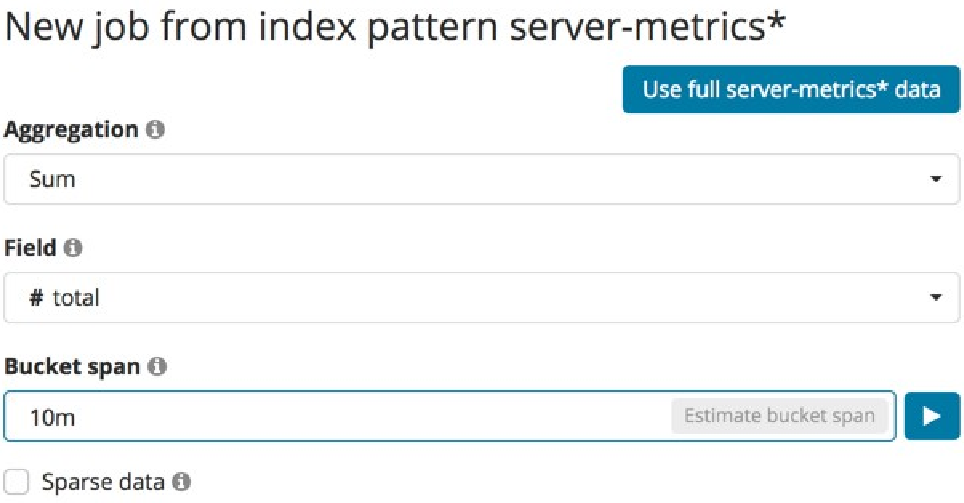
*https://www.elastic.co/guide/en/kibana/current/timelion.html*

# Demo8 : Machine Learning Job

Activate License : *Management* → *License management*

Machine Learning → Job → Create job

Select Data View server-metrics and *Single Metric Job* puis :



New job from index pattern server-metrics*

Use full server-metrics* data

Aggregation

Sum

Field

# total

Bucket span

10m

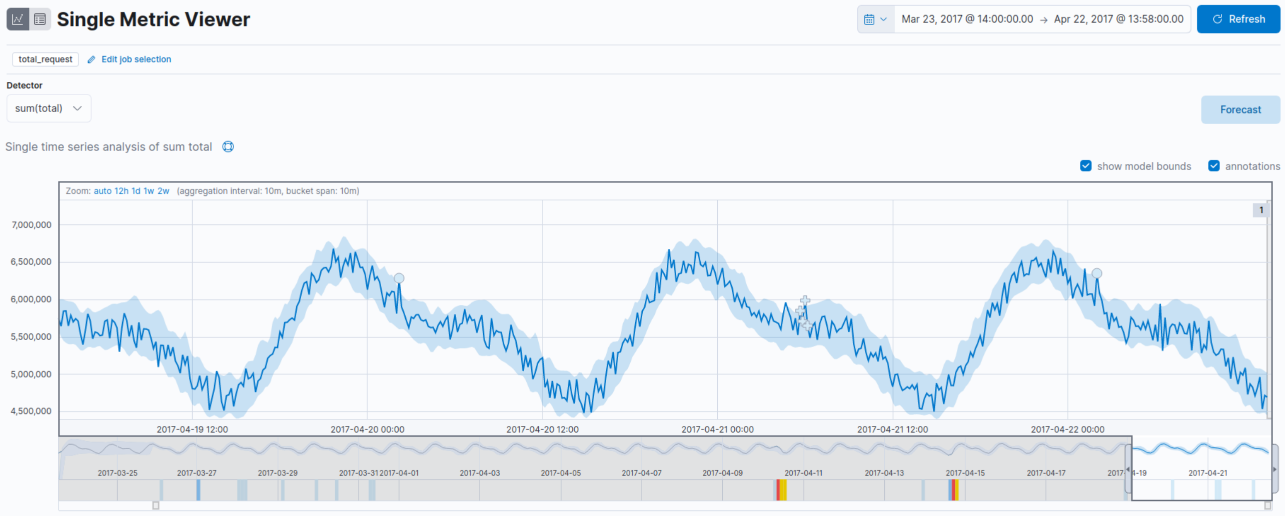
Estimate bucket span

□ Sparse data

Use all available data for analysis
Give a name to the job, for example ***total_request*** , a group for example ***training***

## 2.3 Visualization of results

View results in the ***Single Metric Viewer***



Single Metric Viewer

Mar 23, 2017 @ 14:00:00.00 → Apr 22, 2017 @ 13:58:00.00

Refresh

total_request

Edit job selection

Detector

sum(total)

Forecast

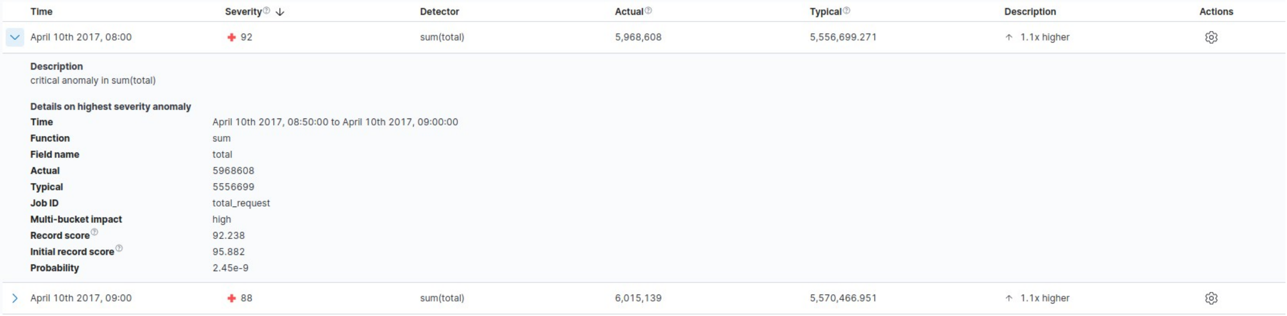
Single time series analysis of sum total

☑ show model bounds ☑ annotations

Zoom: auto 12h 1d 1w 2w (aggregation interval: 10m, bucket span: 10m)

Drag the time selector to select a section containing a critical anomaly.

View information related to the anomaly, in the table at the bottom



| Time | Severity | Detector | Actual | Typical | Description | Actions |
|---|---|---|---|---|---|---|
| April 10th 2017, 08:00 | ✚ 92 | sum(total) | 5,968,608 | 5,556,699.271 | ↑ 1.1x higher | ⚙ |
| April 10th 2017, 09:00 | ✚ 88 | sum(total) | 6,015,139 | 5,570,466.951 | ↑ 1.1x higher | ⚙ |

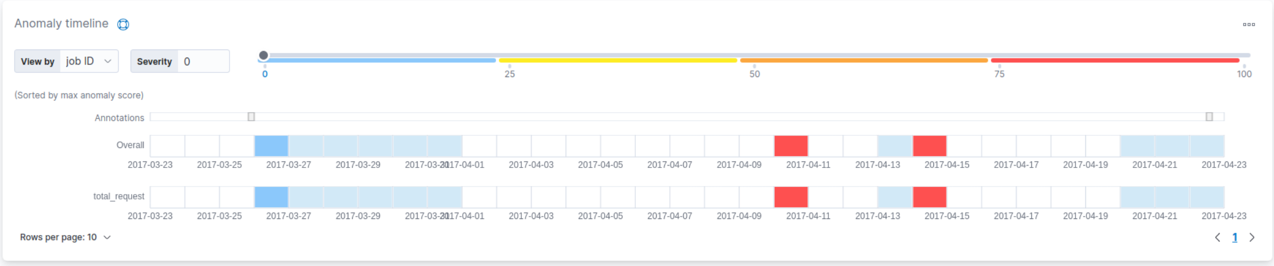
**Description**
critical anomaly in sum(total)

**Details on highest severity anomaly**

| | |
|---|---|
| **Time** | April 10th 2017, 08:50:00 to April 10th 2017, 09:00:00 |
| **Function** | sum |
| **Field name** | total |
| **Actual** | 5968608 |
| **Typical** | 5556699 |
| **Job ID** | total_request |
| **Multi-bucket impact** | high |
| **Record score** | 92.238 |
| **Initial record score** | 95.882 |
| **Probability** | 2.45e-9 |

Then view the results with the ***Anomaly Explorer***

Select a critical anomaly



Anomaly timeline

View by: job ID

Severity: 0

(Sorted by max anomaly score)

Rows per page: 10