



# Introduction to Elasticsearch3

s1270232 Yoshino Tomoaki

2021.07.28



# Agenda

1. Last week...

2. Full-text search

3. Demo



# Notice



# Environment

2021/07

OS : Windows 10 Pro

Env: Java 8

Elastic: Elasticsearch 7.13

Kibana 7.13

MongoDB 4.2



**Elasticsearch**



Last week...

# Last week...

## overview

Goal: insert English wiki & compare query response with MongoDB

Data:

origin: enwiki-20110901-pages-articles.xml.bz2

use : files of spited for each page (s1170078)

size : 2.61GB

page: 159,535

How: insert data into DB with python



Elasticsearch

# Last week...

## Python script

insert xml to elasticsearch

flow: loop for each 2000 pages

1. read xml
2. convert dictionary
3. insert

### Terminal

```
pip install elasticsearch, xmltodict
```

<https://elasticsearch-py.readthedocs.io/en/v7.13.4/index.html>



Elasticsearch

```
from elasticsearch import Elasticsearch, helpers
import xmltodict
import json
import glob

def main():
    with Elasticsearch(timeout=180) as es:
        for _, _ in helpers.streaming_bulk(es, get_page(), request_timeout=60*3):
            pass

def get_page():
    for file in glob.glob('Data/**/*.xml', recursive=True):
        try:
            with open(file, mode='rt', encoding='utf-8') as f:
                xml_string = f.read()

                p = json.loads(json.dumps(xmltodict.parse(xml_string)))
                yield {
                    '_op_type': 'create',
                    '_index': 's1270232',
                    'doc': p
                }
        except Exception as e:
            print(e)
            continue

if __name__ == "__main__":
    main()
```

# Last week...

## Result

---

## Error:

```
object mapping for [doc.page.revision.text.template.field] tried to parse  
field [field] as object, but found a concrete value
```



Elasticsearch

0

4



# Last week...

## Reason

### Elasticsearch make mapping

- mapping = Scheme
- Automatically create & update (default)

※ So Elasticsearch is not perfectly NoSQL.

```
"properties": {
  "page": {
    "properties": {
      "id": {
        "type": "text",
        "fields": {
          "keyword": {
            "type": "keyword",
            "ignore_above": 256
          }
        }
      }
    }
  },
  "revision": {
    "properties": {
      "catlinks": {
        "properties": {
          "catlink": {
            "properties": {
              "@href": {
                "type": "text",
                "fields": {
                  "keyword": {
                    "type": "keyword",
                    "ignore_above": 256
                  }
                }
              }
            }
          }
        }
      }
    }
  }
}
```



Elasticsearch

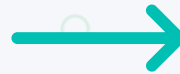
# Last week...

## Reason

---

Object type in json

```
{ "i": { "link": { "#text": "mongo" } } }
```



Object type in Elasticsearch

```
"i": {  
  "properties": {  
    "link": {  
      "properties": {  
        "#text": {  
          "type": "text",  
          "fields": {  
            "keyword": {  
              "type": "keyword",  
              "ignore_above": 256  
            }  
          }  
        }  
      }  
    }  
  }  
}
```

<https://www.elastic.co/guide/en/elasticsearch/reference/current/object.html>

# Last week...

## Reason

---

### Object type in json

```
{ "i": { "link": { "#text": "mongo" } } }
```

```
{ "i": "elastic" }
```



### Object type in Elasticsearch

```
"i": {  
  "properties": {  
    "link": {  
      "properties": {  
        "#text": {  
          "type": "text",  
          "fields": {  
            "keyword": {  
              "type": "keyword",  
              "ignore_above": 256  
            }  
          }  
        }  
      }  
    }  
  }  
}
```

<https://www.elastic.co/guide/en/elasticsearch/reference/current/object.html>



Elasticsearch

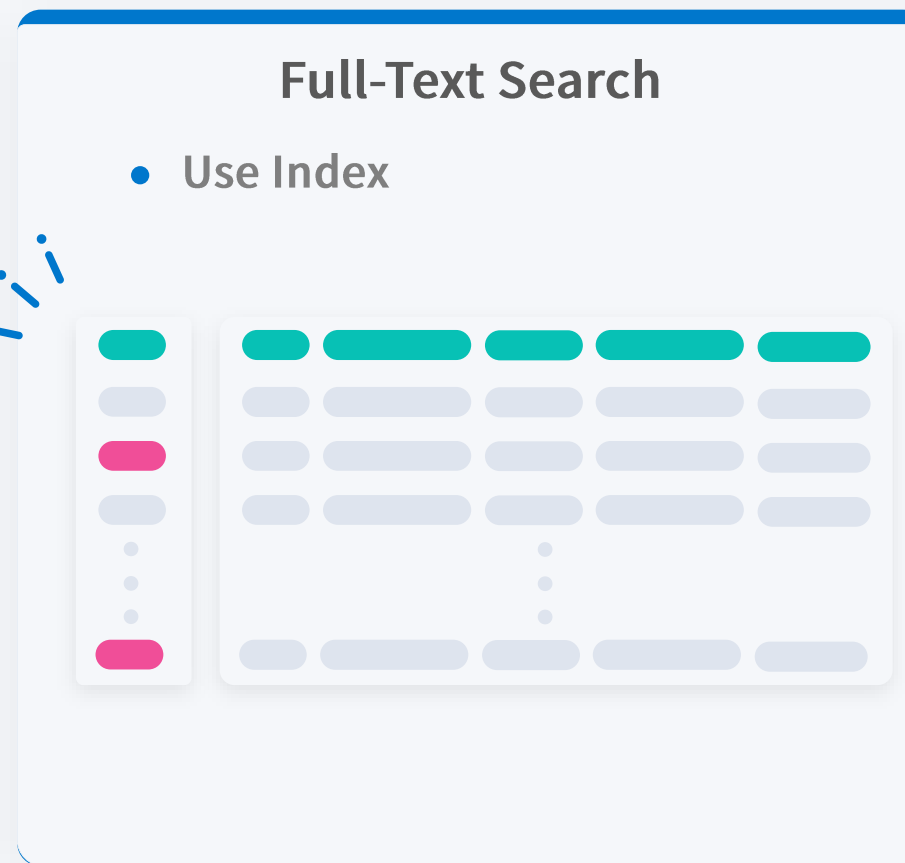


# Full-text search

# Full-text search

## Full-text search

- search by index (inverted)
- fast



<https://github.com/ozlerhakan/mongodb-json-files>

# Full-text search

## in Elasticsearch

- create inverted index for each field

condition: text field

※ keywords field stored raw contents

```
"#text": {  
  "type": "text",  
  "fields": {  
    "keyword": {  
      "type": "keyword",  
      "ignore_above": 256  
    }  
  }  
}
```

<https://www.elastic.co/guide/en/elasticsearch/reference/current/text.html>

# Full-text search

## in Elasticsearch

### Query:

- match
- match\_phrase
- multi\_match
- query\_string

etc.

### Search:

- full text search
- fuzzy matching
- phrase queries

<https://www.elastic.co/guide/en/elasticsearch/reference/current/full-text-queries.html>



**Elasticsearch**

# Full-text search

## in Elasticsearch

### multi\_match

- multi-field queries
- base: match
- can use **wildcard**
- fields default: **all (\*)**

```
{
  "query": {
    "multi_match": {
      "query": "this is a test",
      "fields": ["subject", "message"]
    }
  }
}
```

```
{
  "query": {
    "multi_match": {
      "query": "this is a test",
      "fields": ["subject", "*_name"]
    }
  }
}
```

<https://www.elastic.co/guide/en/elasticsearch/reference/current/full-text-queries.html>





# Demo

# query performance

Database: wiki

Table: ja

Documents: 1276008

1. title which include “会津” in all field.

2. title which include {“会津” AND “猪苗代” in all field}.



Elasticsearch

1

2

# Additional query performance

## Query

1.

Elastic: `{"size": 10000, "query": {"multi_match": {"query": "会津"}}, "_source": ["title"]}`

2.

Elastic: `{"size": 10000, "query": {"multi_match": {"query": "会津 猪苗代", "operator": "AND"}}, "_source": ["title"]}`



Elasticsearch

1  
3

result

	Elasticsearch
1	20720ms
2	9997ms

→ maybe, can practical use



Elasticsearch

# References

- Python Elasticsearch Client &mdash; Elasticsearch 7.13.4 documentation  
(<https://elasticsearch-py.readthedocs.io/en/v7.13.4/index.html>)
- Full text queries | Elasticsearch Guide [7.13] | Elastic  
(<https://www.elastic.co/guide/en/elasticsearch/reference/current/full-text-queries.html>)
- Multi-match query | Elasticsearch Guide [7.13] | Elastic  
(<https://www.elastic.co/guide/en/elasticsearch/reference/current/query-dsl-multi-match-query.html>)
- Text field type | Elasticsearch Guide [7.13] | Elastic  
(<https://www.elastic.co/guide/en/elasticsearch/reference/current/text.html>)



# Thank you!

s1270232 Yoshino Tomoaki