

Elastic Search: Mapping[1]

How Data are Typed, Stored, and Indexed

Weimao Ke

Drexel University

Table of contents

1. Mapping Basics
2. How to create a Mapping?
3. Major Field Data Types
4. Examples

Mapping Basics

What is Mapping

Mapping is to define how a document and its fields are stored and indexed:

- string fields to be treated as full text and analyzed;
- data types of other fields such as numbers, dates, and geolocations;
- formats of the data;
- custom rules for dynamic fields.

Field Data Types

Each field has a data type:

- a simple type like text, keyword, date, long, double, boolean or ip;
- a type of JSON such as object or nested;
- a specialized type like geo_point and geo_shape.

A field can be indexed in different ways for different purposes, e.g.:

- a string field to be indexed as text and keyword;
- or analyzed using english and french analyzers;
- multi-fields with the *fields* parameter

A field without a predefined mapping:

- Dynamic mapping to automatically add the field;
- Dynamic mapping rules can be used to customize the mapping;
- However, potential mapping explosion...

Mapping explosion and limits:

- `index.mapping.total_fields.limit`: the max number of fields in an index;
- `index.mapping.depth.limit`: the max depth of a field (for its inner objects);
- `index.mapping.nested_fields.limit`: the max number of distinct nested fields;
- `index.mapping.field_name_length.limit`: the max length of a field name.

How to create a Mapping?

Mapping

Create an index with a mapping:

```
PUT /my-index
{
  "mappings": {
    "properties": {
      "age":    { "type": "integer" },
      "email":  { "type": "keyword" },
      "name":   { "type": "text"  }
    }
  }
}
```

Add a field to an existing mapping:

```
PUT /my-index/_mapping
{
  "properties": {
    "employee-id": {
      "type": "keyword",
      "index": false
    }
  }
}
```

What about updating/changing a mapping?

- Short answer: You cannot do that!
- Because data have been indexed and stored.
- Instead, create a new index with the new mapping and reindex your data.

What about renaming a field?

- Short answer: You cannot do that, once data are in there.
- You can, however, create an alias as an alternate name,
- Or rebuild your index with data.

Mapping

To view the mapping of an index:

```
GET /my-index/_mapping
```

Result:

```
{
  "my-index" : {
    "mappings" : {
      "properties" : {
        "age" : {
          "type" : "integer"
        },
        "email" : {
          "type" : "keyword"
        },
        ...
      }
    }
  }
}
```

Mapping

To view the mapping of a specific field:

```
GET /my-index/_mapping/field/employee-id
```

Result:

```
{
  "my-index" : {
    "mappings" : {
      "employee-id" : {
        "full_name" : "employee-id",
        "mapping" : {
          "employee-id" : {
            "type" : "keyword",
            "index" : false
          }
        }
      }
    }
  }
}
```

Major Field Data Types

Field Data Types

Core Data Types:

- Text: full-text values such as a message or product description;
- Keyword: string values such as ID, email address, zip code;
- Numeric: numbers such as integer, float, etc.
- boolean: true or false
- Date and Date nanoseconds: date/time values in different precisions;
- Range: numeric ranges, date ranges

`text` , `keyword` ..

`long` , `integer` , `short` , `byte` , `double` , `float` ..

`date` , `date_nanos` , `boolean` , `binary` ..

`integer_range` , `float_range` , `double_range` , `date_range` ..

Advanced Data Types:

- Complex data types: object, nested (JSON)
- Geo data types: geo_point, geo_shape
- Specialized types: ip, completion, token_count, ..
- Arrays: any field can contain 0 or more values by default
- Multi-fields (*fields*): one field to be indexed in multiple ways

Examples

Field Data Types

Keyword data type:

```
PUT my_index
{
  "mappings": {
    "properties": {
      "tags": {
        "type": "keyword"
      }
    }
  }
}
```

```
PUT /my_index/_doc/1
{tags:"information"}
```

```
PUT /my_index/_doc/2
{tags:["information","retrieval"]}
```

Field Data Types

Text data type:

```
PUT my_index
{
  "mappings": {
    "properties": {
      "full_name": {
        "type": "text"
      }
    }
  }
}
```

```
PUT /my_index/_doc/1
{"full_name": "John P. Smith"}
```

```
PUT /my_index/_doc/2
{"full_name": "Price Waterhouse Cooper"}
```

Field Data Types

Multi-fields:

```
PUT my_index
```

```
{
  "mappings": {
    "properties": {
      "city": {
        "type": "text",
        "fields": {
          "raw": {
            "type": "keyword"
          }
        }
      }
    }
  }
}
```

```
PUT my_index/_doc/1
{"city": "New_York"}
```

```
PUT my_index/_doc/2
{"city": "York"}
```

- The city.raw field is a keyword version of the city field.
- The city field can be used for full text search.
- The city.raw field can be used for sorting and aggregations

References

- [1] elastic.co. Elasticsearch reference [7.5]: Mapping.
<https://www.elastic.co/guide/en/elasticsearch/reference/7.5/mapping.html>. Accessed: 2020-1-16.