

ElasticSearch 说明文档

ElasticSearch

地址：111.229.131.214:9200

无用户名、密码认证

数据库配置

Elasticsearch版本为8.0.0，使用docker部署。关闭了身份验证、SSL加密等安全配置。

使用ik分词器，版本为8.0.0。 <https://github.com/infinilabs/analysis-ik/releases/tag/v8.0.0>

索引配置

项目导入依赖spring-boot-starter-data-elasticsearch。

数据持久化对象

```
1  @Data
2  @Document(indexName = "news")
3  @AllArgsConstructor
4  @NoArgsConstructor
5  @Builder
6  public class NewsDoc {
7      @Id ()
8      private Long id;
9
10     @Field(type = FieldType.Text, searchAnalyzer = "ik_smart", analyzer =
        "ik_max_word")
11     private String title;           //标题
12
13     @Field(type = FieldType.Date)
14     private Date date;             //日期
15
16     @Field(type = FieldType.Text, searchAnalyzer = "ik_smart", analyzer =
        "ik_max_word")
17     private String content;        //内容
18
19     @Field(type = FieldType.Text)
20     private String source;         //来源
21
22     @Field(type = FieldType.Keyword)
```

```

23     private Category category; //类别
24
25     @Field(type = FieldType.Double)
26     private Double standPoint; //立场 保存的是小数
27
28     @Field(type = FieldType.Text)
29     private List<String> keywords; //关键词
30 }

```

添加同义词配置

PUT 111.229.131.214:9200/news/_settings?pretty

```

1 {
2   "settings": {
3     "analysis": {
4       "filter": {
5         "my_synonym_filter": {
6           "type": "synonym",
7           "synonyms_path" : "analysis/synonym.txt"
8         }
9       },
10      "analyzer": {
11        "my_synonyms": {
12          "tokenizer": "ik_smart",
13          "filter": [
14            "lowercase",
15            "my_synonym_filter"
16          ]
17        }
18      }
19    }
20  }
21 }

```

更改索引mapping

```

1 {
2   "properties" : {
3     "_class" : {
4       "type" : "keyword",

```

```

5         "index" : false,
6         "doc_values" : false
7     },
8     "category" : {
9         "type" : "keyword"
10    },
11    "content" : {
12        "type" : "text",
13        "analyzer" : "ik_max_word",
14        "search_analyzer" : "my_synonyms"
15    },
16    "date" : {
17        "type" : "date",
18        "format" : "date_optional_time||epoch_millis"
19    },
20    "id" : {
21        "type" : "long"
22    },
23    "keywords" : {
24        "type" : "text"
25    },
26    "source" : {
27        "type" : "text"
28    },
29    "standPoint" : {
30        "type" : "double"
31    },
32    "title" : {
33        "type" : "text",
34        "analyzer" : "ik_max_word",
35        "search_analyzer" : "my_synonyms"
36    }
37 }
38 }

```

索引信息

GET 111.229.131.214:9200/news

response如下:

```

1 {
2     "news": {
3         "aliases": {},
4         "mappings": {
5             "properties": {

```

```
6         "_class": {
7             "type": "keyword",
8             "index": false,
9             "doc_values": false
10        },
11        "category": {
12            "type": "keyword"
13        },
14        "content": {
15            "type": "text",
16            "analyzer": "ik_max_word",
17            "search_analyzer": "my_synonyms"
18        },
19        "date": {
20            "type": "date",
21            "format": "date_optional_time||epoch_millis"
22        },
23        "id": {
24            "type": "long"
25        },
26        "keywords": {
27            "type": "text"
28        },
29        "source": {
30            "type": "text"
31        },
32        "standPoint": {
33            "type": "double"
34        },
35        "title": {
36            "type": "text",
37            "analyzer": "ik_max_word",
38            "search_analyzer": "my_synonyms"
39        }
40    }
41},
42"settings": {
43    "index": {
44        "routing": {
45            "allocation": {
46                "include": {
47                    "_tier_preference": "data_content"
48                }
49            }
50        },
51        "refresh_interval": "1s",
52        "number_of_shards": "1",
```

```

53         "provided_name": "news",
54         "creation_date": "1713201242492",
55         "store": {
56             "type": "fs"
57         },
58         "analysis": {
59             "filter": {
60                 "my_synonym_filter": {
61                     "type": "synonym",
62                     "synonyms_path": "analysis/synonym.txt"
63                 }
64             },
65             "analyzer": {
66                 "my_synonyms": {
67                     "filter": [
68                         "lowercase",
69                         "my_synonym_filter"
70                     ],
71                     "tokenizer": "ik_smart"
72                 }
73             }
74         },
75         "number_of_replicas": "1",
76         "uuid": "0PojdOnRQ6u-2czZXCMHqQ",
77         "version": {
78             "created": "8000099"
79         }
80     }
81 }
82 }
83 }

```

已完成的迭代

2024.4.8迭代

重构search方法，使用ik分词器，配置同义词文件，以达到模糊搜索的要求。

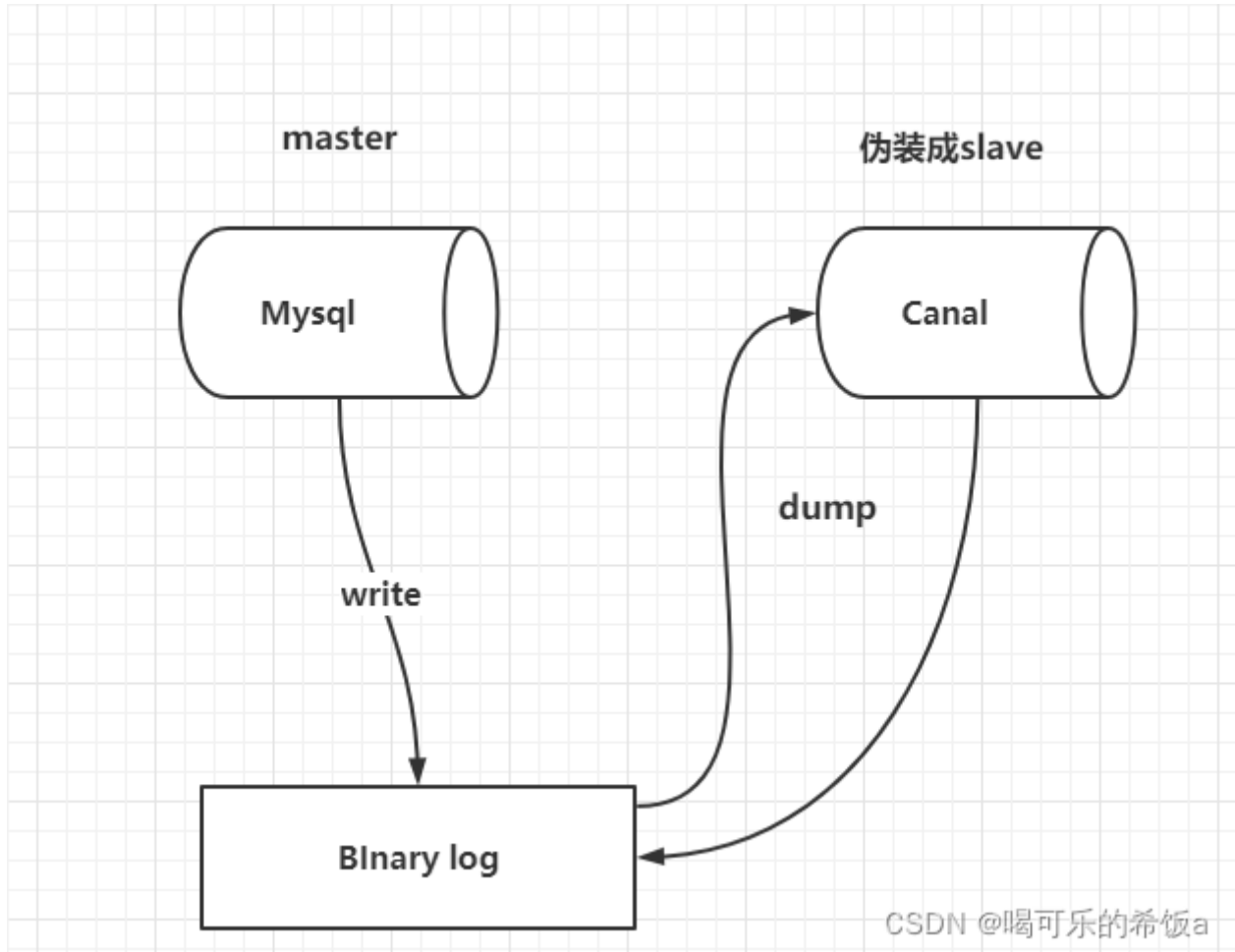
2024.4.11迭代

高亮显示关键词

接下来的任务

Mysql+Canal+ElasticSearch实现数据同步：

https://blog.csdn.net/H_porridge/article/details/123992421



大致原理

- canal模拟mysql master slave的交互协议，把自己伪装mysql slave，向mysql master发送dump协议（dump用于备份）
- mysql master接收到dump请求，向canal推送binlog
- canal通过解析binlog，将数据同步到其他的数据源

参考

ElasticSearch下载：<https://www.elastic.co/cn/downloads/elasticsearch>

ElasticSearch配置教程：https://blog.csdn.net/aibo_yyds/article/details/120081333