

windows下logstash同步mysql数据到elasticsearch中

- 1.在官网 (<https://www.elastic.co/cn/downloads/past-releases#elasticsearch>) 下载
 - elasticsearch
 - logstash
 - 下载mysql-connector-java(应该是驱动架包吧 不懂java不知道说得对不对)。地址 <https://mvnrepository.com/artifact/mysql/mysql-connector-java>
- 2.配置文件
 - logstash.conf

```
input {
  jdbc {
    # 设置 MySQL/MariaDB 数据库url以及数据库名称
    jdbc_connection_string => "jdbc:mysql://127.0.0.1:3306/foodie-shop-dev?
useUnicode=true&characterEncoding=UTF-8&autoReconnect=true"
    # 用户名和密码
    jdbc_user => "root"
    jdbc_password => "363626256"
    # 数据库驱动所在位置，可以是绝对路径或者相对路径
    jdbc_driver_library => "../mysql-connector-java-5.1.41.jar"
    # 驱动类名
    jdbc_driver_class => "com.mysql.jdbc.Driver"
    # 开启分页
    jdbc_paging_enabled => "true"
    # 分页每页数量，可以自定义
    jdbc_page_size => "1000"
    # 执行的sql文件路径
    statement_filepath => "../foodie-items.sql"
    # 设置定时任务间隔 含义：分、时、天、月、年，全部为*默认含义为每分钟跑一次任务
    schedule => "* * * * *"
    # 索引类型
    type => "_doc"
    # 是否开启记录上次追踪的结果，也就是上次更新的时间，这个会记录到
last_run_metadata_path 的文件
    use_column_value => true
    # 记录上一次追踪的结果值
    last_run_metadata_path => "../track_time"
    # 如果 use_column_value 为true， 配置本参数，追踪的 column 名，可以是自增id或者
时间
    tracking_column => "updated_time"
    # tracking_column 对应字段的类型
    tracking_column_type => "timestamp"
    # 是否清除 last_run_metadata_path 的记录，true则每次都从头开始查询所有的数据库记
录
    clean_run => false
    # 数据库字段名称大写转小写
    lowercase_column_names => false
  }
}
output {
  elasticsearch {
```

```

# es地址
hosts => ["127.0.0.1:9200"]
# 同步的索引名
index => "foodie-items"
# 设置_docID和数据相同
document_id => "%{id}"
# document_id => "%{itemId}"
}
# 日志输出
stdout {
  codec => json_lines
}
}

```

- foodie-items.sql

```

SELECT
  i.id as itemId,
  i.item_name as itemName,
  i.sell_counts as sellCounts,
  ii.url as imgUrl,
  tempSpec.price_discount as price,
  i.updated_time as updated_time
FROM
  items i
  LEFT JOIN items_img ii ON i.id = ii.item_id
LEFT JOIN
  (SELECT item_id, MIN(price_discount) as price_discount FROM items_spec GROUP
BY item_id) tempSpec on i.id = tempSpec.item_id
  WHERE ii.is_main = 1
  and i.updated_time >= :sql_last_value

```

logstash

- Logstash 是免费且开放的服务器端数据处理管道，能够从多个来源采集数据，转换数据，然后将数据发送到您最喜欢的“存储库”中。
- Logstash 能够动态地采集、转换和传输数据，不受格式或复杂度的影响。利用 Grok 从非结构化数据中派生出结构，从 IP 地址解码出地理坐标，匿名化或排除敏感字段，并简化整体处理过程。

logstash输入端

- elasticsearch
- jdbc
- kafka
- log4j
- rabbitmq
- redis
- websocket
- file
- http

logstash输出端

- CSV
- elasticsearch
- email
- kafka
- mongodb
- rabbitmq
- redis
- websocket
- file