# 说明文档

为了将 kafka 中的数据抽取后写入到 elasticsearch 当中并且适用多 topic 的应用场景，需要将官方所提供的 kafka to elasticsearch 的程序功能做一定的扩展，原有的程序只支持配置单个 topic 和 index 的读写操作，扩展或可以根据配置文件中配置的多个 topic 去读对应的数据源并按照日期名称来生成 elasticsearch 当中的 index 方便后续的数据管理。当需要增加或者删除数据源的 topic 时，只需要修改配置文件中的 topic 列表，程序会定时监控配置文件，如果有新增的 topic 会新启动一个线程来进行操作，如果减少了 topic 则会将原有的任务停止。配置文件如下：

### Zookeeper properties ####################################

# Zookeeper's host:port list: <host1>:<port1>,…,<hostN>:<portN>

# default value: localhost:2181, if not specified

kafkaZookeeperList=10.13.56.52:2181

### Kafka properties ####################################

# Kafka Brokers host:port list: <host1>:<port1>,…,<hostN>:<portN>

# default: localhost:9092

#kafkaBrokersList=localhost:9092

kafkaBrokersList=10.13.56.52:9092

# Kafka Consumer group name prefix -

# default: kafka\_es\_indexer

consumerGroupName=kafka\_es\_indexer

# Kafka Topic from which the message has to be processed

# mandatory property, no default value specified.

# topics topic1,topic2

topicList=test

## below two properties define a range of partitions to be processed by this application;

## each partition will be processed by an IndexerJob in a separate Thread, so this also

## defines the number of indexer threads created by the app - please set memory requirements accordingly

# timeout in seconds before force-stopping Indexer app and all indexer jobs

appStopTimeoutSeconds=10

### ElasticSearch properties ####################################

# ElasticSearch Host and Port List for all the nodes

# Example: esHostPortList=machine\_1\_ip:9300,machine\_2\_ip:9300

esHostPortList=10.13.56.52:9300

# Name of the ElasticSearch Cluster that messages will be posted to;

# Tip: Its not a good idea to use the default name "ElasticSearch" as your cluster name.

#esClusterName=my-application\_abel-mac

esClusterName=dip-application

# Fully qualified name of a concrete message handler class

# Default: "org.elasticsearch.kafka.consumer.RawMessageStringHandler"

# Custom class should be extended from org.elasticsearch.kafka.consumer.MessageHandler class

#messageHandlerClass=org.elasticsearch.kafka.indexer.messageHandlers.RawMessageStringHandler

messageHandlerClass=org.elasticsearch.kafka.indexer.messageHandlers.TestLogMessageHandler

# Preferred message encoding to process the message before posting it to ElasticSearch.

# Default: "UTF-8"

messageEncoding=UTF-8

# number of times to try to re-connect to ES when performing batch indexing , if connection to ES fails

numberOfEsIndexingRetryAttempts=2

# sleep time in ms between attempts to connect to ES

esIndexingRetrySleepTimeMs=10000

# flag to enable/disable performance timings reporting; Set it to true when performance/load-testing your app, set to ‘false’ when deploying in production as it will affect performance of the app; defaults to ‘false'

isPerfReportingEnabled=false

#设置消息数量，当读取达到此数目的消息时向 ES 中写入

numOfWriteToES=50000