Migrate data to Elastic Search

# Premise

1.Out data from data lake must follow CAP standard that according to Kafka queue.

2.Spark reads data from HDFS and pushes data to Kafka topic queue.

3.ConsumerKafka and putElasticSearchREST Processor configuration are complete and running.in workflow.

4.Elastic Search server is ready for us.

Note:

All above environments, Nifi processor and Elastic Search server must have enough resource.

eg: ES server : 3 Master Nodes and 2 data nodes.

# Operation

* **Environment**

1. Nifi env:
   * Integration Nifi env: <https://10.85.116.92:9090/nifi/>
   * Production Nifi env: <https://10.85.112.126:9090/nifi/>
2. Elastic Search:
   * AWS URL:

<https://federation-sts.accenture.com/adfs/ls/idpinitiatedsignon.aspx?loginToRp=urn:amazon:webservices>

* + Elastic Search server URL:

DEV:

<https://federation-sts.accenture.com/adfs/ls/idpinitiatedsignon.aspx?loginToRp=urn:amazon:webservices>

Staging:

<https://vpc-aliceportal-30899-stg-p5wnwhqekvy676cmkaj757ud7y.us-east-1.es.amazonaws.com>

APT:

<https://vpc-aliceportal-30899-apt-46vw4s7m43zgk6nwxbs2ssq2vy.us-east-1.es.amazonaws.com>

UAT:

<https://vpc-aliceportal-30899-stage-dmrjizgjptvhd43yb5wdgpmv5i.us-east-1.es.amazonaws.com>

PROD:

<https://vpc-aliceportal-30899-prod-bspoqyzlxjgqprzpuolxmojkpm.us-east-1.es.amazonaws.com>

Note:

Access to Elastic Search server you need Access key and Secret key

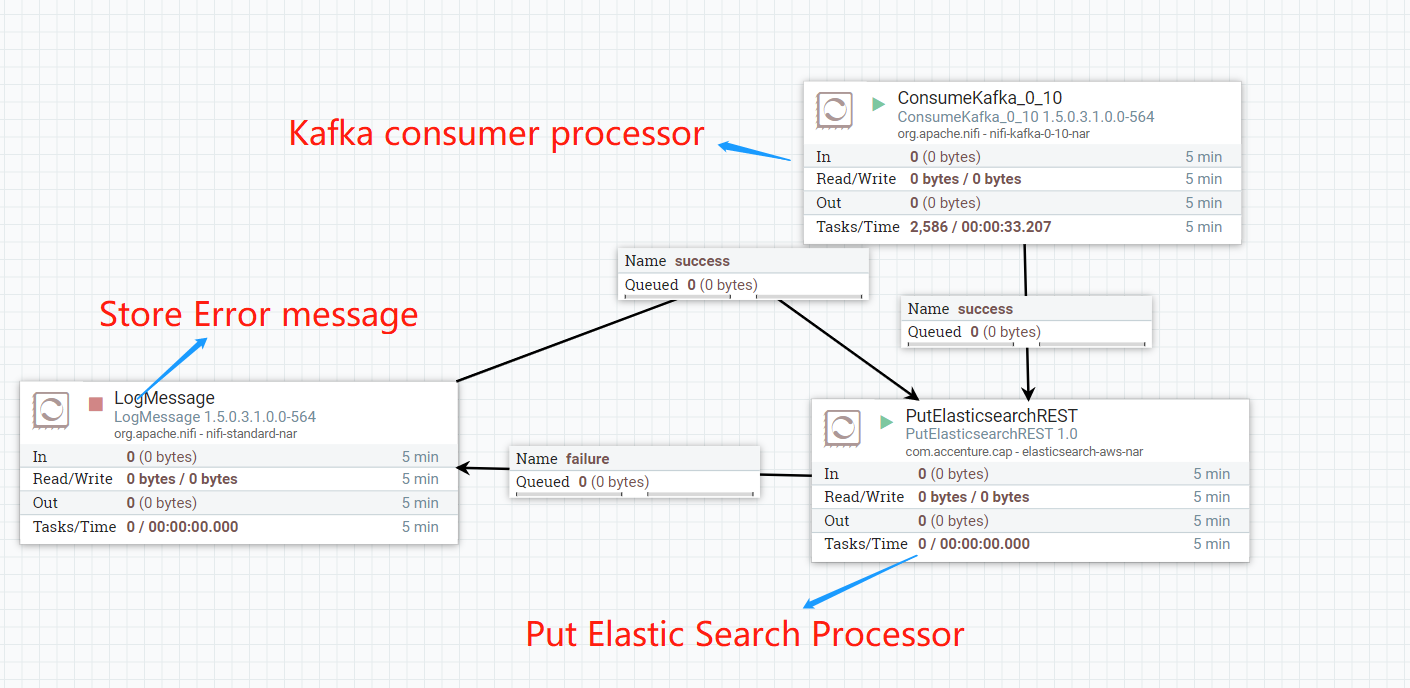
because our Elastic Search server in AWS.

We can raise a ticket to get Access key and Secret key from CAP team

(Gates: [gates.yuxiang.chi@accenture.com](mailto:gates.yuxiang.chi@accenture.com))

* **How can we use them?**

1. The whole pipeline workflow as bellow screenshot:



* + **Kafka consumer processor(ConsumeKafka processor)**:

This processor consumes Kafka topic data and pushes data to Nifi queue.

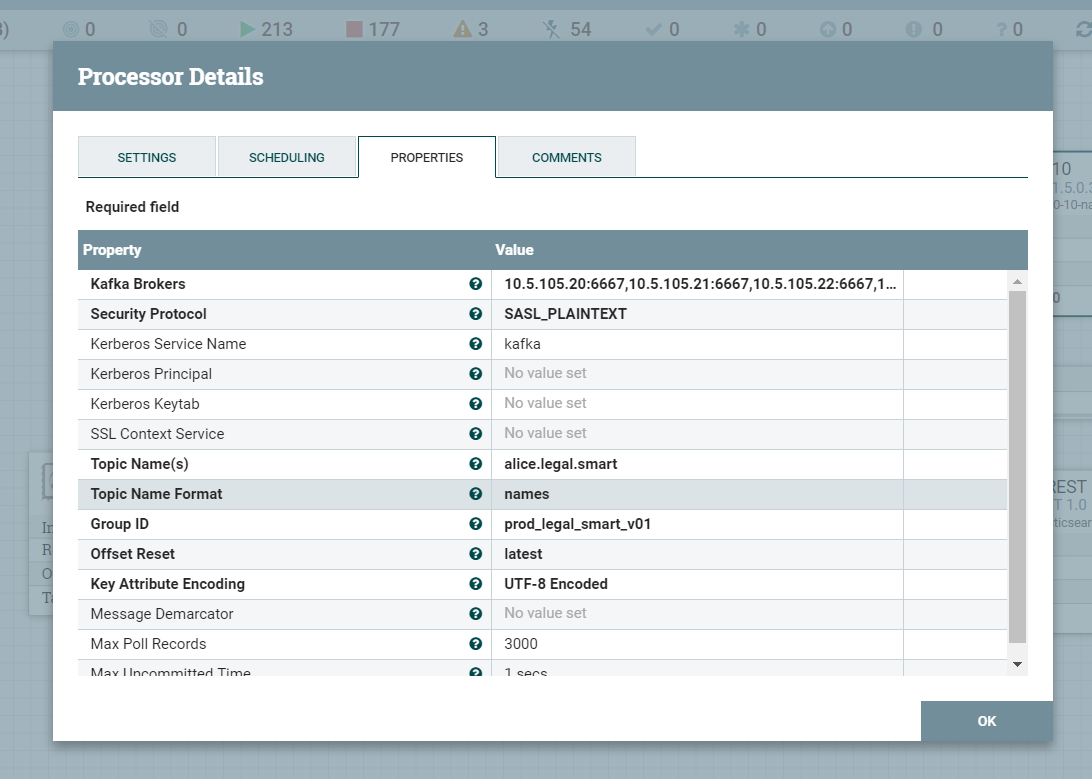
* + **Put Elastic Search Processor(putElasticSearchREST processor)**:

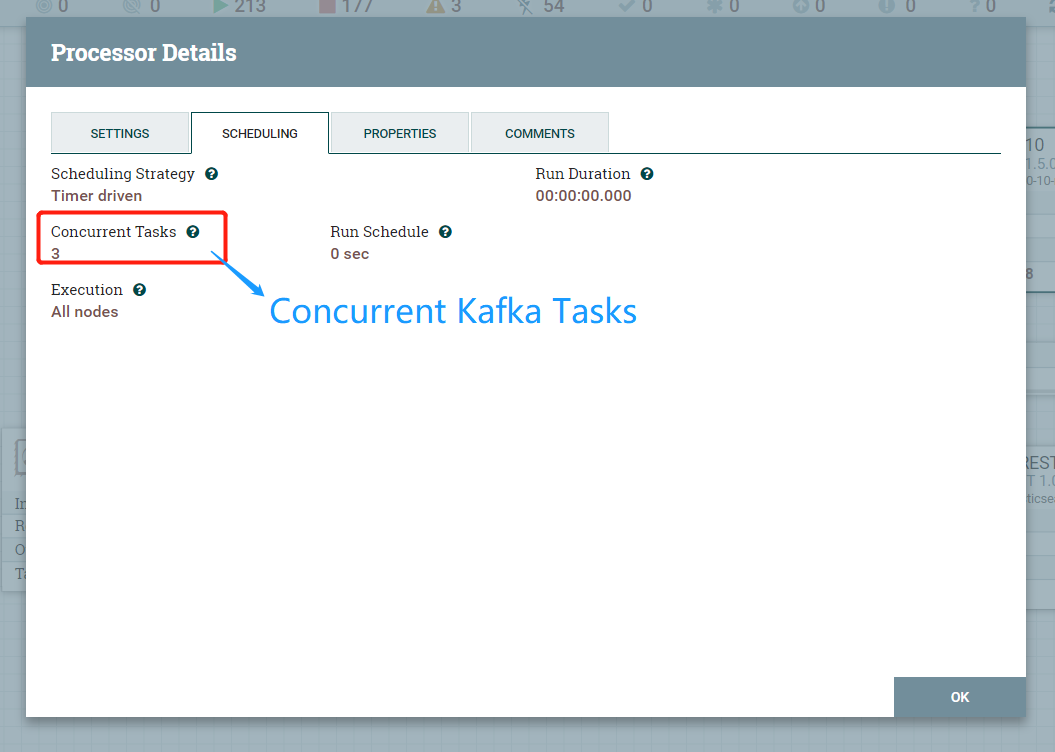
This processor gets data from Nifi queue and puts every record to Elastic Search.

* + **Store Error message(LogMessage processor)**:

This processor stores error message when putting data to Elastic Search server is failure.

1. Kafka consumer processor configuration, please see bellow screenshot:





Tab **PROPERTIES**:

[Kafka Brokers]:

This is the Kafka server brokers configuration.

[Security Protocol]:

This is security setting for access to Kafka server

[Kerberos Service Name]:

This is the Kafka Kerberos service name that is setting in kafka\_jaas.conf.

[Topic Name]:

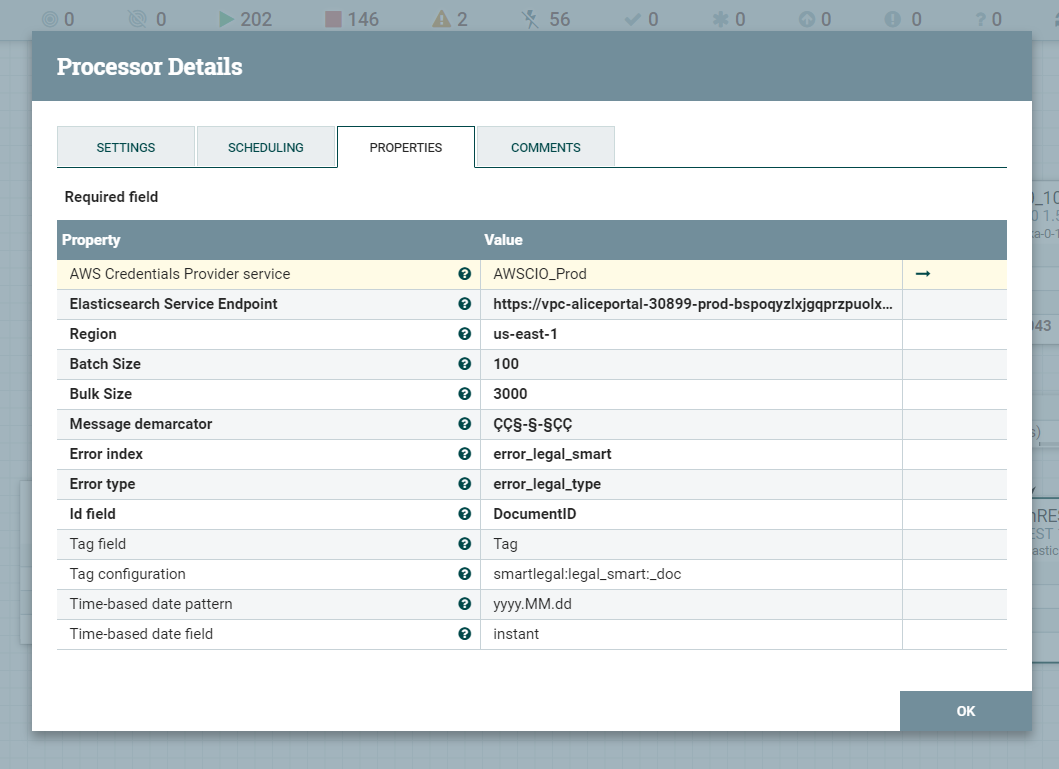
Kafka topic name.

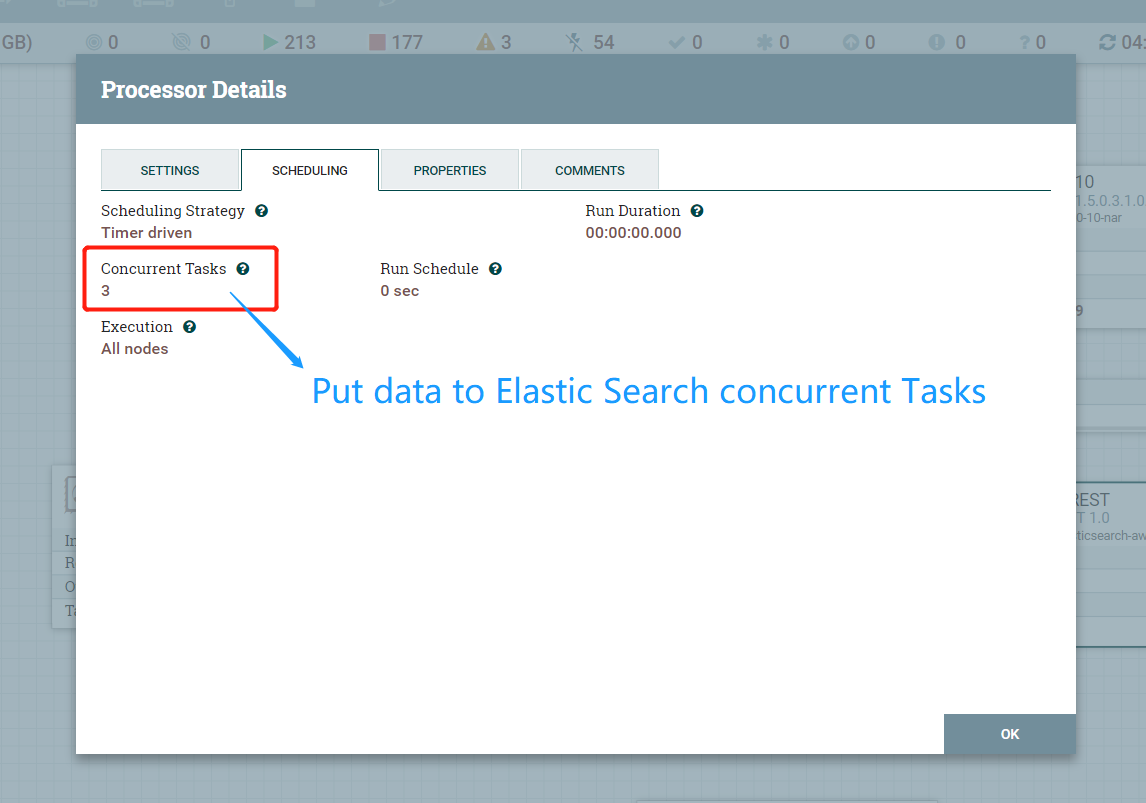
Tab **SHCEDULING**:

Concurrent Tasks:

This is setting concurrent Kafka tasks number.

1. Put data to Elastic Search processor configuration, please see bellow screenshot:





Tab **PROPERTIES**:

[AWS Credentials Provider service]:

This is the AWS service. You will fill Access key and Secret key in this service to get credential for AWS.

[Elasticsearch Service Endpoint]:

This is the Elastic Search server you will access to. The endpoint address you can find in AWS Elastic Search service or get this from CAP team(Gates: gates.yuxiang.chi@accenture.com)

[Region]:

This property is setting “us-east-1”.

[Message demarcator]:

This is a special demarcator if you use it to combine your data. This processor will follow it to split your combination data.

Tab **SHCEDULING**:

Concurrent Tasks:

This is setting concurrent tasks number to put data to Elastic Search.

1. Reference document
   1. Nifi:



* 1. Elastic Search:



1. Query and Test Elastic Search

We can query and test Elastic Search by Postman tool.

Please refer to 4 Reference document *Elastic Search Blog.docx*.