

Setup Guide:

PEACH Core Analytics System

1. Download the following packages onto your machine:

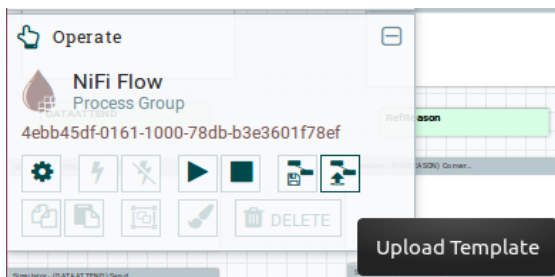
NiFi (<https://nifi.apache.org/>)
Kafka (<https://kafka.apache.org/>)
Spark (<http://spark.apache.org/>)
Druid (<http://druid.io/>)
Metabase (<http://www.metabase.com/start/docker.html>)
Zookeeper (<http://zookeeper.apache.org>)

2. Setup NiFi:

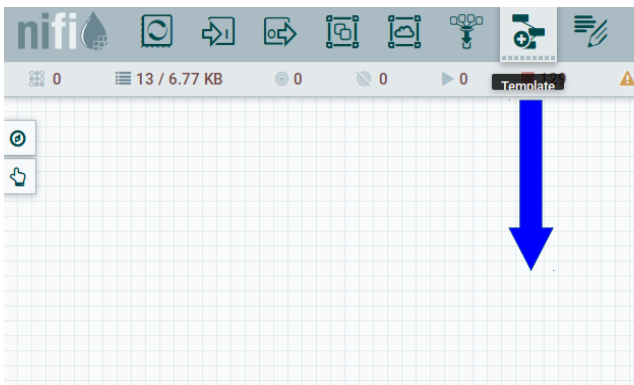
- Run NiFi on your machine: Open the NiFi directory on terminal and input commands:
(It may take a few minutes to start)

```
$ bin/nifi.sh run
```

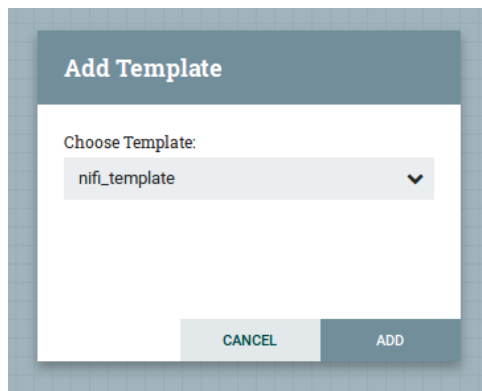
- Access NiFi GUI via <http://localhost:8080/nifi/>
- Import a template by clicking the icon from the components toolbar:
(Select template stored in /core-analytics-deliverables/NiFi-Kafka/)



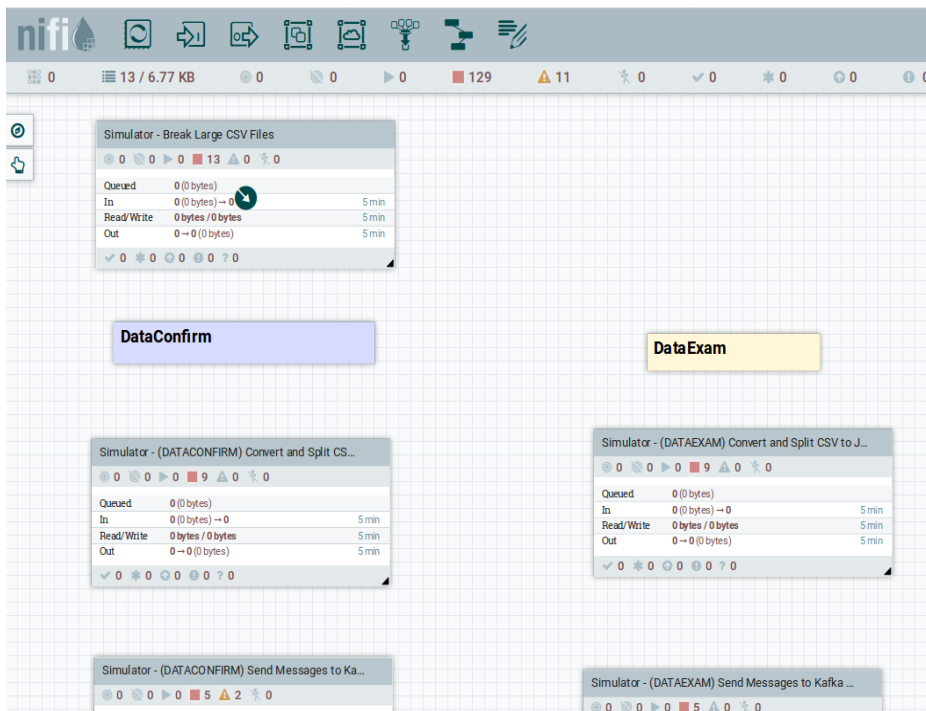
- Add the template by dragging the template icon onto your main screen:



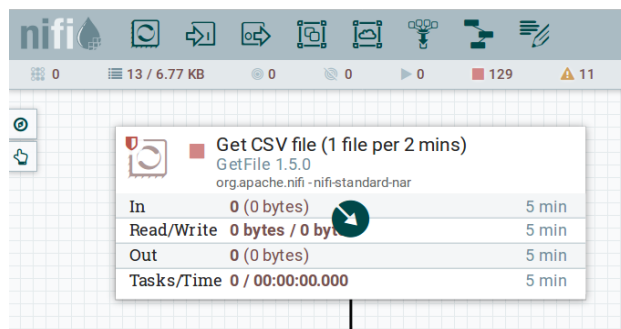
- Choose the uploaded template and then press add:



- Double click on the group you wish to configure:



- Double click the processor relevant to input or output:



- Click the properties tab and then change the input/output directory to the relevant file path in your machine:

Configure Processor

SETTINGS

SCHEDULING

PROPERTIES

COMMENTS

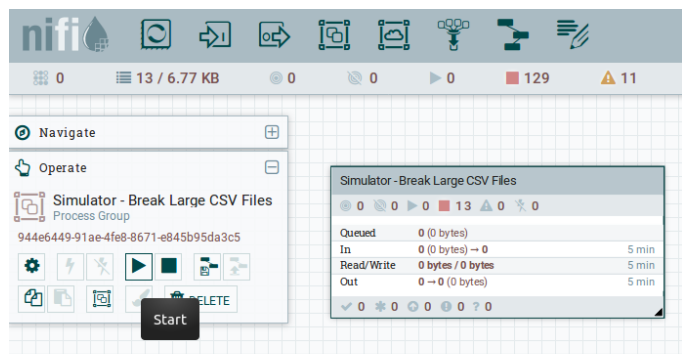
Required field +

Property	Value	
Input Directory	? Empty string set	
File Filter	? [\\].*.csv	
Path Filter	? No value set	
Batch Size	? 1	
Keep Source File	? false	
Recurse Subdirectories	? false	
Polling Interval	? 0 sec	
Ignore Hidden Files	? true	
Minimum File Age	? 0 sec	
Maximum File Age	? No value set	
Minimum File Size	? 0 B	
Maximum File Size	? No value set	

CANCEL

APPLY

- Apply any changes you make. Right-click and select 'Leave Group' to return to the main screen.
- Click any group of processes that you wish to start and press the 'Play' icon on the components toolbar:



3. Setup NiFi to Kafka:

- Open the Zookeeper directory on terminal and input command:

```
$ cp conf/zoo_sample.cfg conf/zoo.cfg  
$ ./bin/zkServer.sh start
```

- Open the Kafka directory on terminal and input command:

```
$ bin/kafka-server-start.sh config/server.properties
```

- Add relevant files to the folders you assigned in the NiFi processors. Transmit data to Kafka by pressing play on the NiFi process groups.

4. Setup Kafka to Spark:

- Copy the file path of "sparkCS.jar" that is in folder: /core-analytics-deliverables/Kafka-Spark/. Submit the job to Spark by changing Path/To in the following commands:

```
$ zip -d Path/To/SparkCS.jar META-INF/*.RSA META-INF/*.DSA META-INF/*.SF  
$ ./bin/spark-submit --class com.ipponusa.SparkConsumer --master local[8] Path/To/SparkCS.jar
```

- You can access <http://localhost:4040> to check the status.

5. Setup Druid and Metabase:

- You can use these instructions to start Druid and import data from Kafka:

<http://druid.io/docs/0.10.0/tutorials/quickstart.html>

<http://druid.io/docs/0.10.0/tutorials/tutorial-kafka.html>

- Run Metabase using the following commands:

```
$ docker run -d -p 3000:3000 --name metabase metabase/metabase
```