

# Hadoop i Spark

Mariusz Rafało

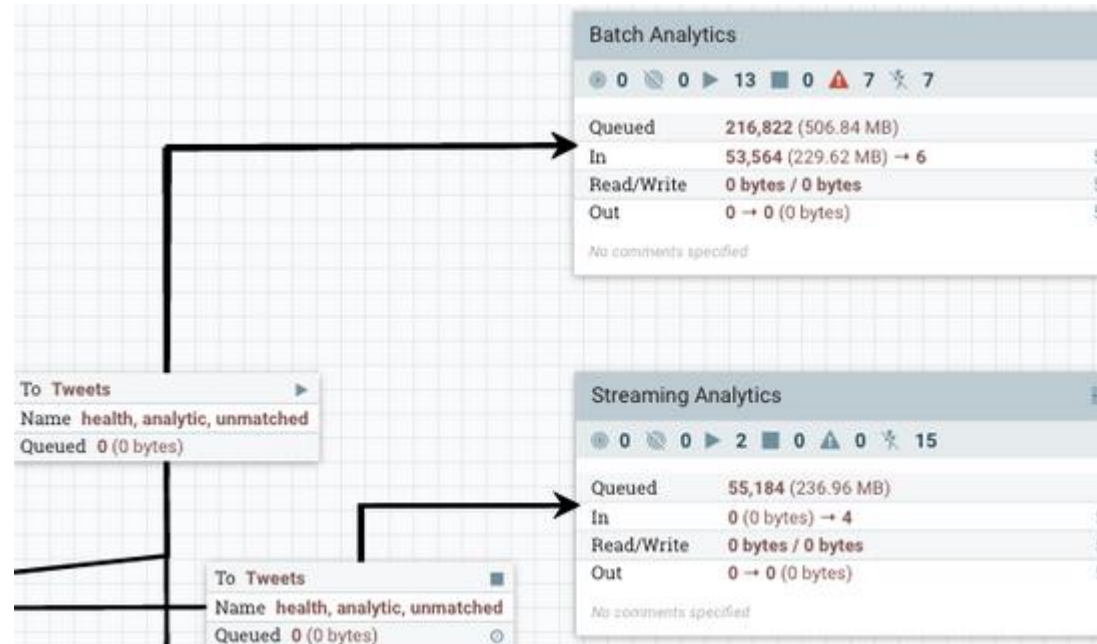
[mrafalo@sgh.waw.pl](mailto:mrafalo@sgh.waw.pl)

<http://mariuszrafalo.pl>

# AUTOMATYZACJA PRZETWARZANIA

# Zarządzanie procesem przepływu danych

- Sqoop
- Oozie
- Apache NIFI
- Airflow

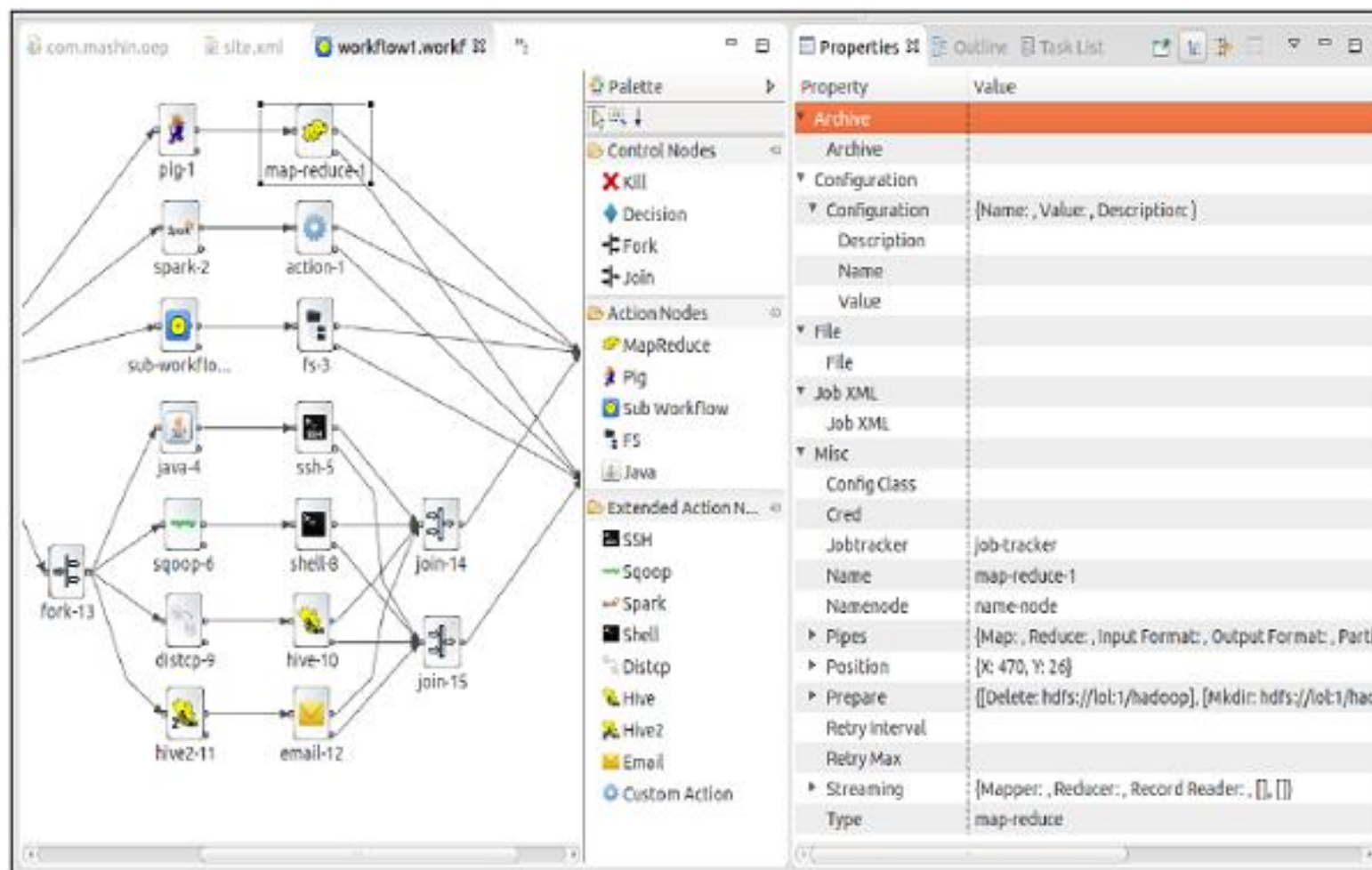


# Sqoop

- Narzędzie do wymiany danych pomiędzy bazami SQL a HDFS
- Obsługuje równoległość odczytu danych
- Przekazuje metadane podczas działania skryptu (możliwość śledzenia postępu pracy)
- Pracuje z HDFS, HBase i Hive

```
sqoop import --libjars /usr/hdp/current/sqoop-client/lib/ojdbc7.jar --connect  
jdbc:oracle:thin:@10.106.26.44:1530/P030T --username USER --password PASS --  
table CON.V_ENERGY_EXCH_POB_SE --split-by ODS_ID --target-dir  
/hadoop/01/dane/sas1
```

# Automatyzacja przetwarzania: Oozie



The screenshot displays the Oozie workflow editor interface. The main workspace shows a workflow diagram with various action nodes connected by arrows. The nodes include:

- Control Nodes: Kill, Decision, Fork, Join
- Action Nodes: MapReduce, Pig, Sub workflow, FS, Java
- Extended Action Nodes: SSH, Sqoop, Spark, Shell, Distcp, Hive, Hive2, Email, Custom Action

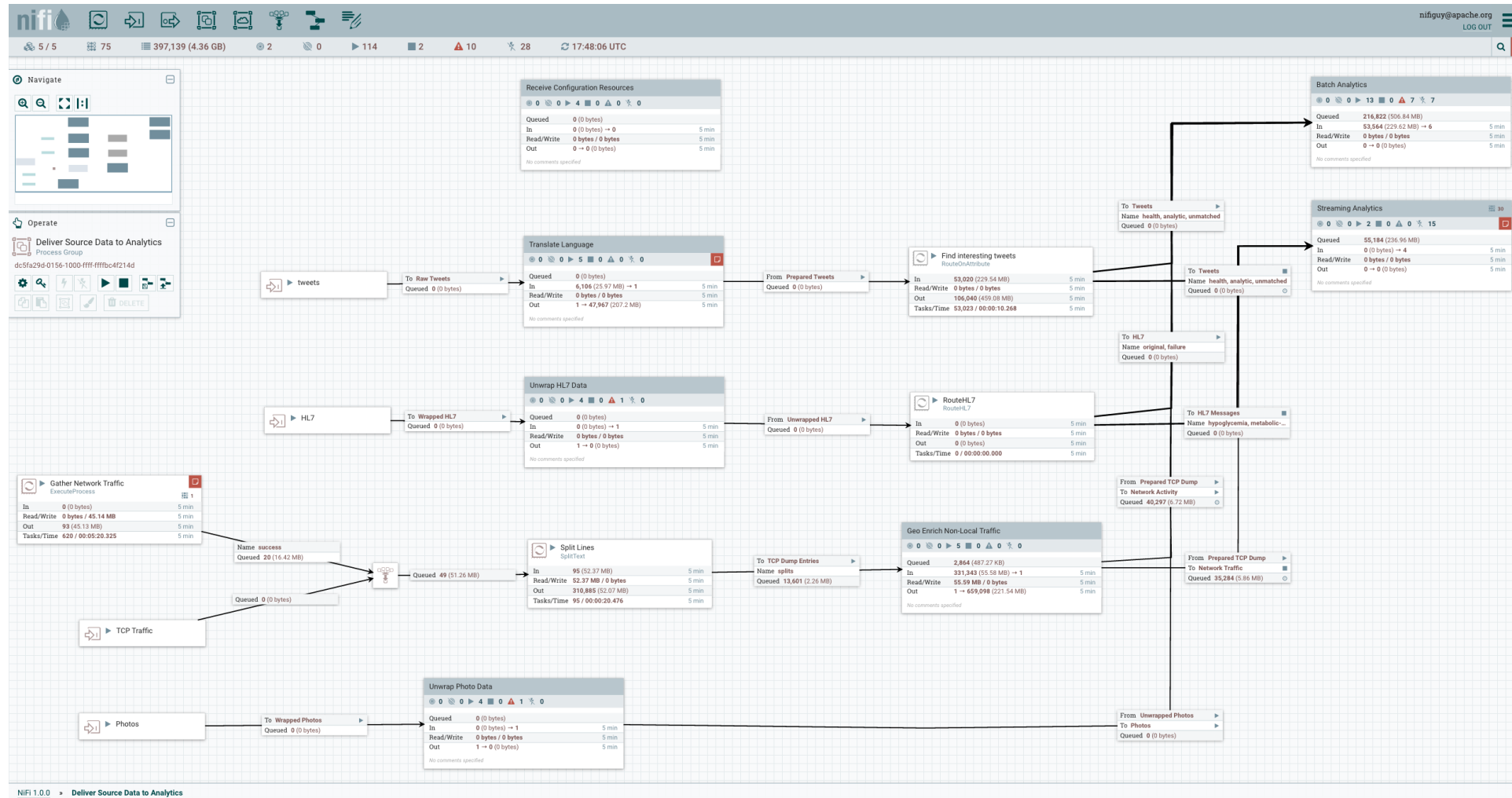
The workflow diagram shows a sequence of actions: pig-1, spark-2, sub-workflo..., java-4, sqoop-6, distcp-9, hive2-11, map-reduce-1, action-1, fs-3, ssh-5, shell-8, hive-10, email-12, join-14, and join-15. The 'map-reduce-1' node is highlighted in the diagram.

The Properties panel on the right shows the configuration for the selected 'map-reduce-1' action:

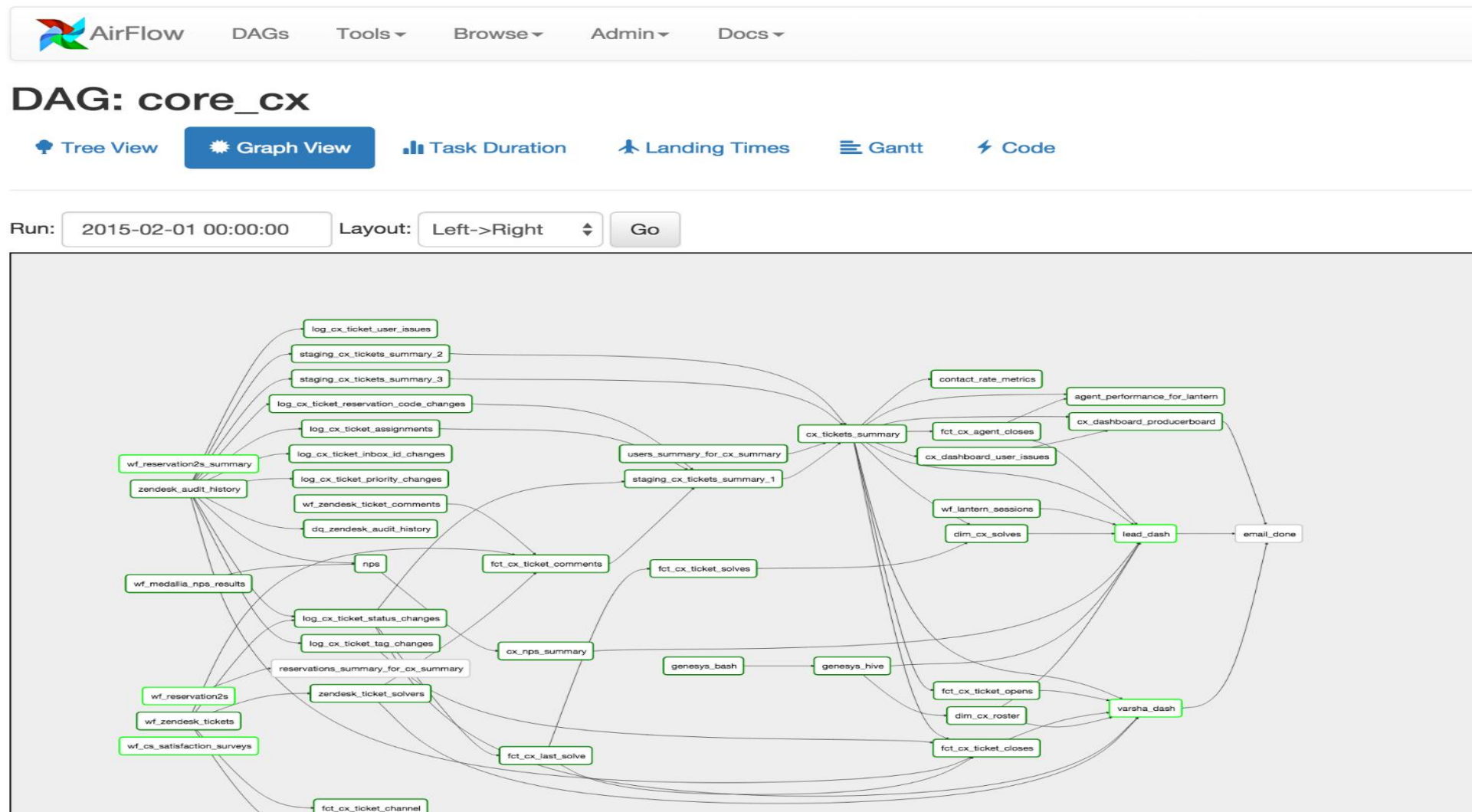
Property	Value
Archive	
Archive	
Configuration	
Configuration	{Name: , Value: , Description: }
Description	
Name	
Value	
File	
File	
Job XML	
Job XML	
Misc	
Config Class	
Cred	
Jobtracker	job-tracker
Name	map-reduce-1
Namenode	name-node
Pipes	{Map: , Reducer: , Input Format: , Output Format: , Partit
Position	{X: 470, Y: 26}
Prepare	{Delete: hdfs://lol:1/hadoop}, {Mkdir: hdfs://lol:1/had
Retry Interval	
Retry Max	
Streaming	{Mapper: , Reducer: , Record Reader: , [], []}
Type	map-reduce

Źródło: [https://www.tutorialspoint.com/apache\\_oozie/apache\\_oozie\\_introduction.htm](https://www.tutorialspoint.com/apache_oozie/apache_oozie_introduction.htm)

# Automatyzacja przetwarzania: Apache NiFi



# Automatyzacja przetwarzania: Airflow



Dziękuję za uwagę