

Hadoop i Spark

Mariusz Rafało

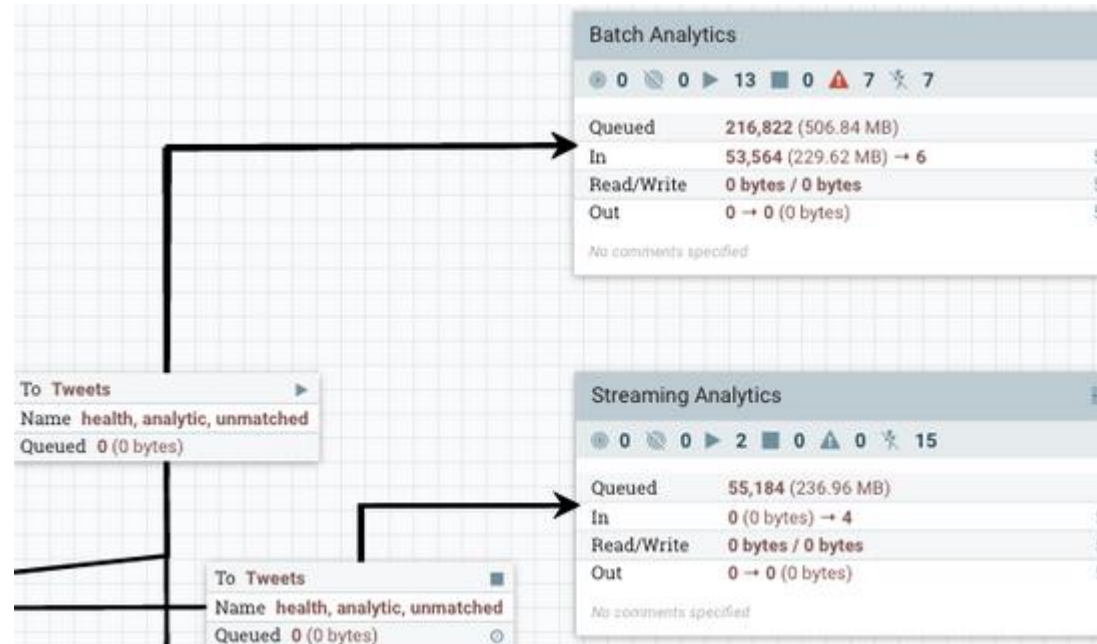
mrafalo@sgh.waw.pl

<http://mariuszrafalo.pl>

AUTOMATYZACJA PRZETWARZANIA

Zarządzanie procesem przepływu danych

- Sqoop
- Oozie
- Apache NIFI
- Airflow

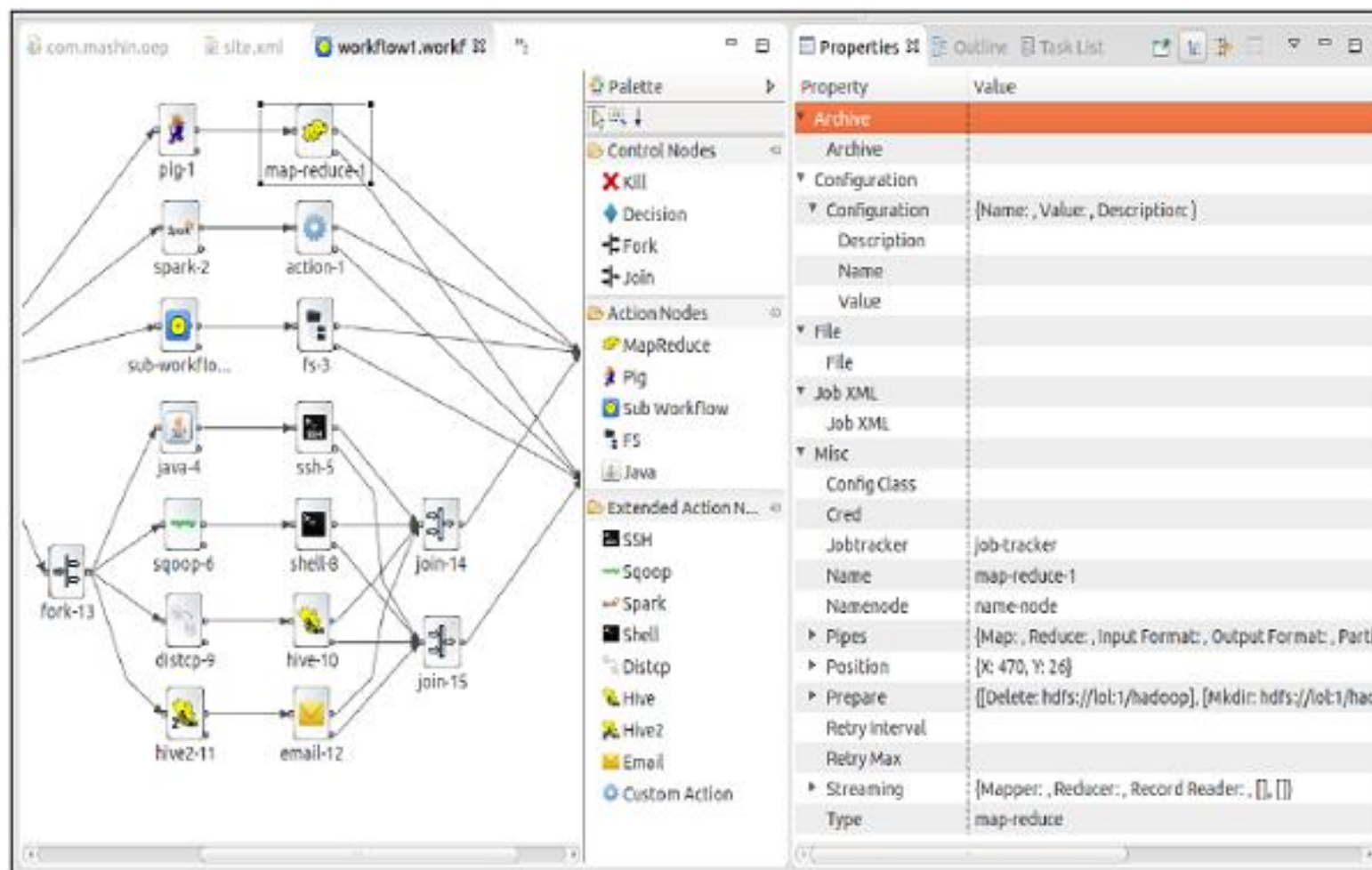


Sqoop

- 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



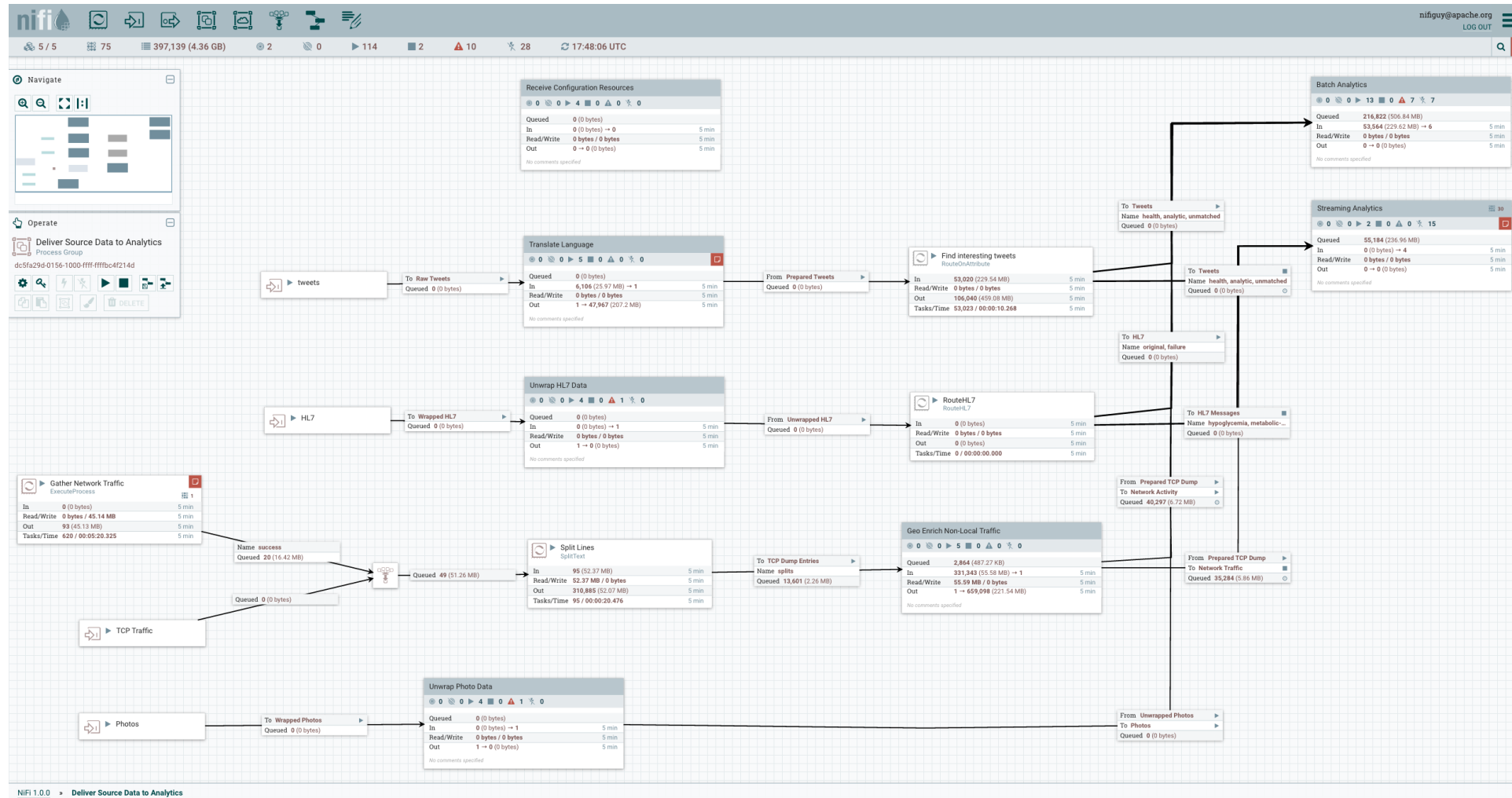
The screenshot displays the Oozie workflow editor interface. The main workspace shows a workflow diagram with various action nodes connected by arrows. The nodes include Pig (pig-1), MapReduce (map-reduce-1), Spark (spark-2), Action (action-1), Sub-workflow (sub-workflo...), File System (fs-3), Java (java-4), SSH (ssh-5), Fork (fork-13), Sqoop (sqoop-6), Shell (shell-8), Join (join-14), Distcp (distcp-9), Hive (hive-10), Join (join-15), Hive2 (hive2-11), and Email (email-12). A palette on the left lists available action types such as Kill, Decision, Fork, Join, MapReduce, Pig, Sub workflow, FS, Java, SSH, Sqoop, Spark, Shell, Distcp, Hive, Hive2, Email, and Custom Action.

The Properties panel on the right is open for the selected 'map-reduce-1' action. It shows a table of properties and their values:

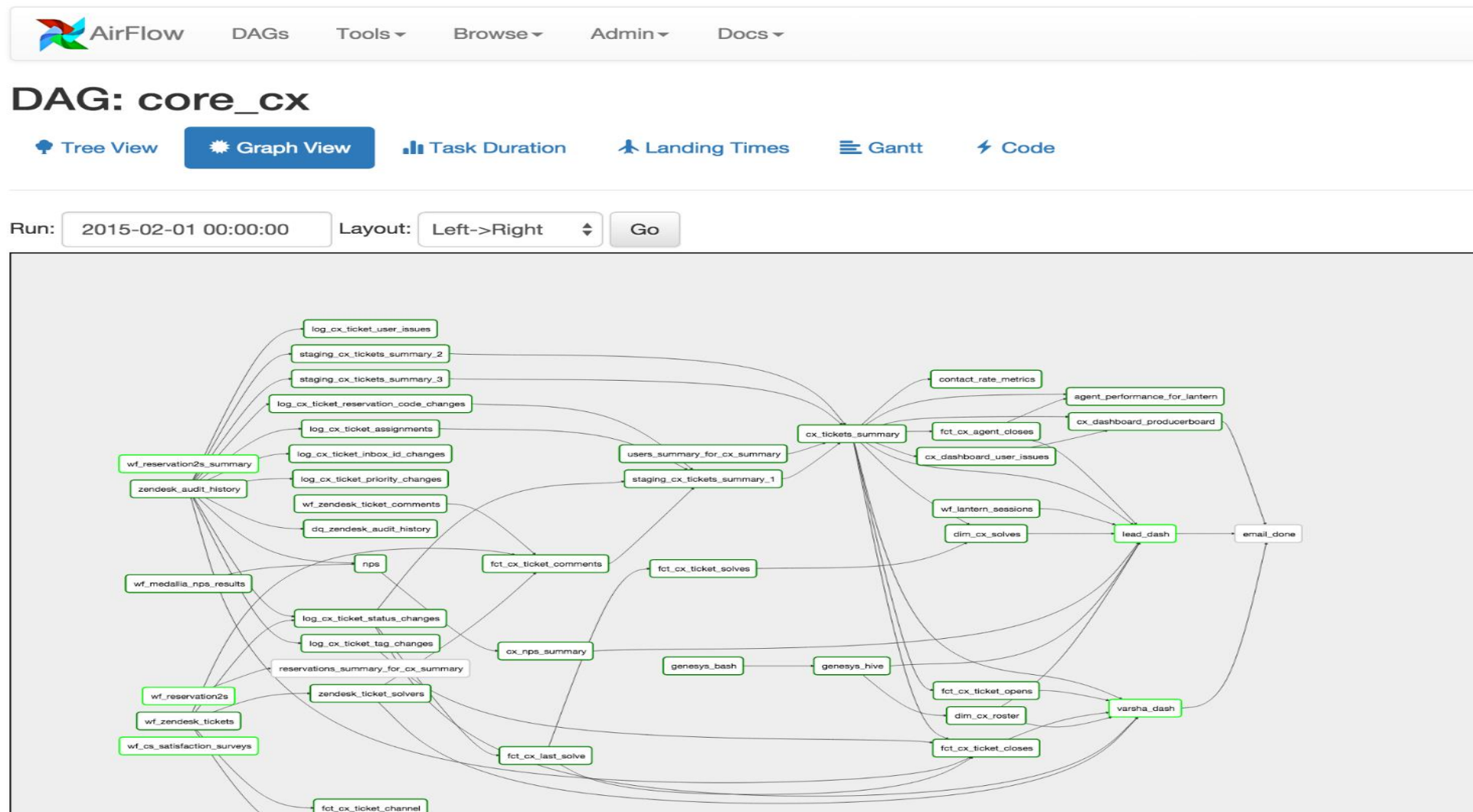
Property	Value
Archive	
Archive	
Configuration	
Configuration	{Name: , Value: , Description: }
Description	
Name	
Value	
File	
File	
Job XML	
Job XML	
Misc	
Config Class	
Cred	
Jobtracker	job-tracker
Name	map-reduce-1
Namenode	name-node
Pipes	{Map: , Reducer: , Input Format: , Output Format: , Partit
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

Automatyzacja przetwarzania: Apache NiFi



Automatyzacja przetwarzania: Airflow



Dziękuję za uwagę