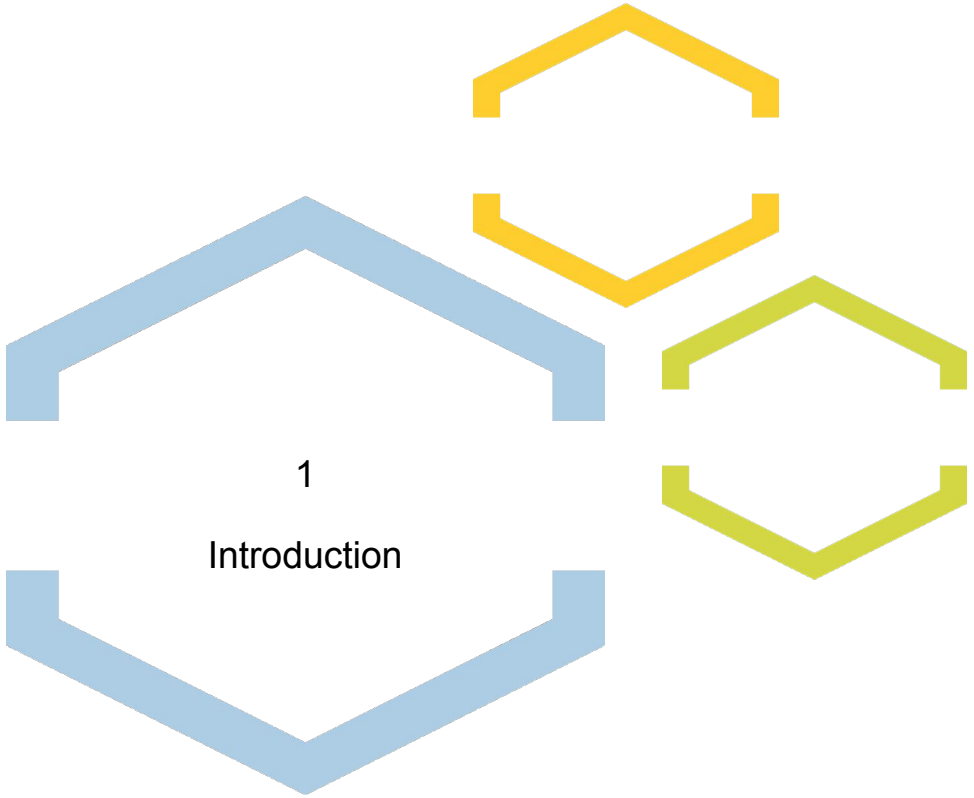
Raúl Reguillo Carmona

Apache NiFi

A General Purpose Tool for Big Data

BEEVA

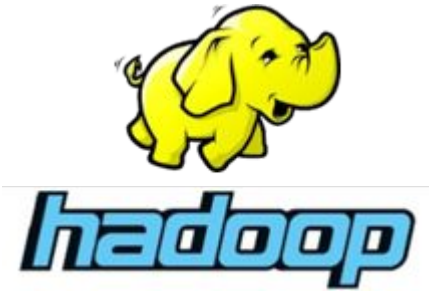
Table of Contents



1

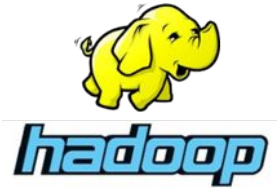
Introduction

- What NiFi is…
  - General purpose Big Data Tool
  - Basic operations (extract, transform and load)
  - Based on processors and flowfiles
  - Drag and drop
  - **Integration**

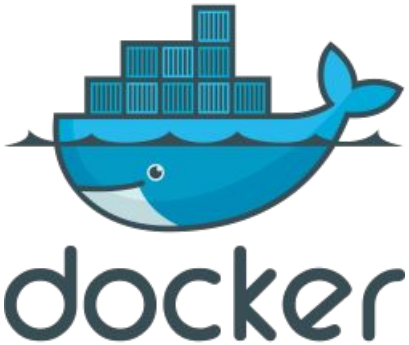


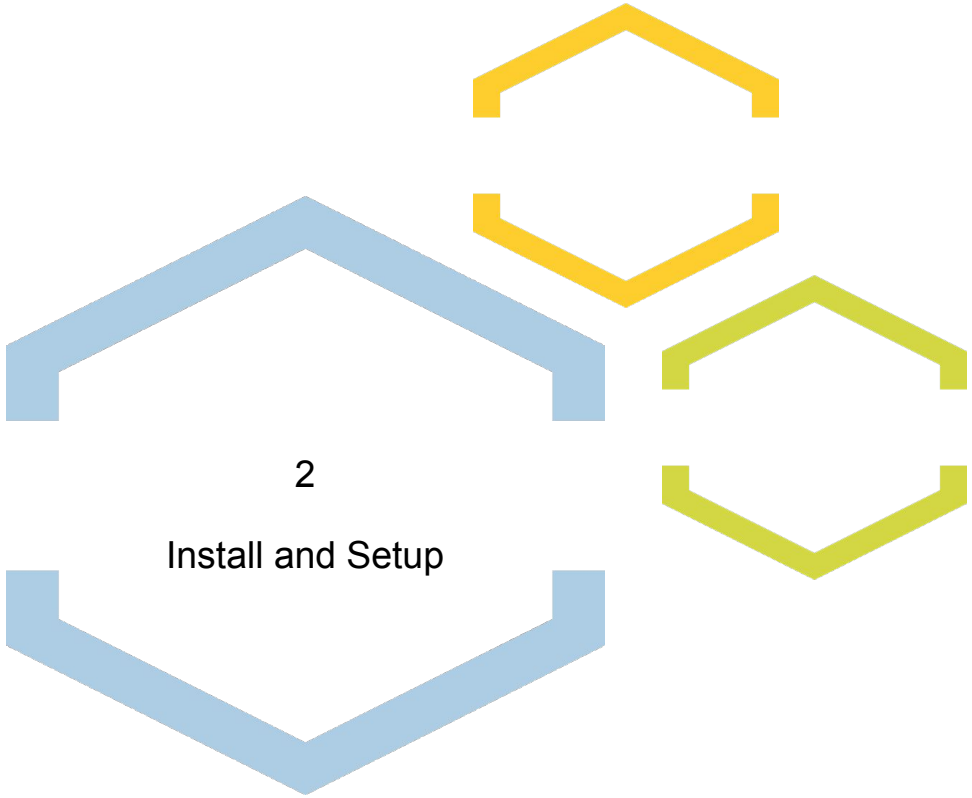
BEEVA

- Features
  - Clear interface
  - Changes at runtime
  - API
  - Kerberos/SSL/SSH/HTTPS/Multitenant friendly
  - Lots of processors
    - but also extensible
  - Expression language
    - dynamic configurations



- About this PoC
  - NiFi properties
  - Some use cases
  - Integration between services
  - NiFi as *man in the middle*
  - Use along with
    - Docker
    - ELK stack
    - Custom microservices

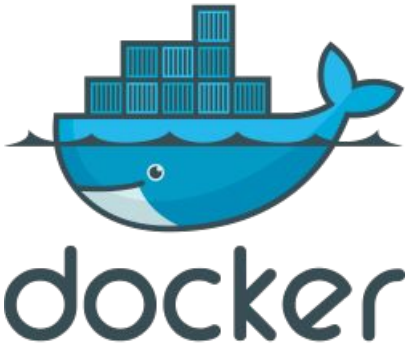




2

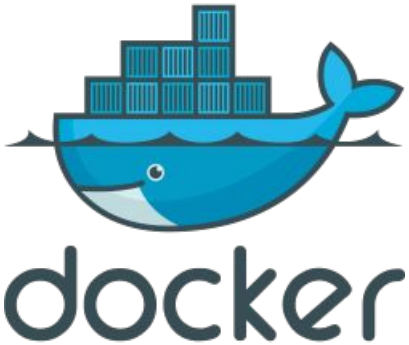
Install and Setup

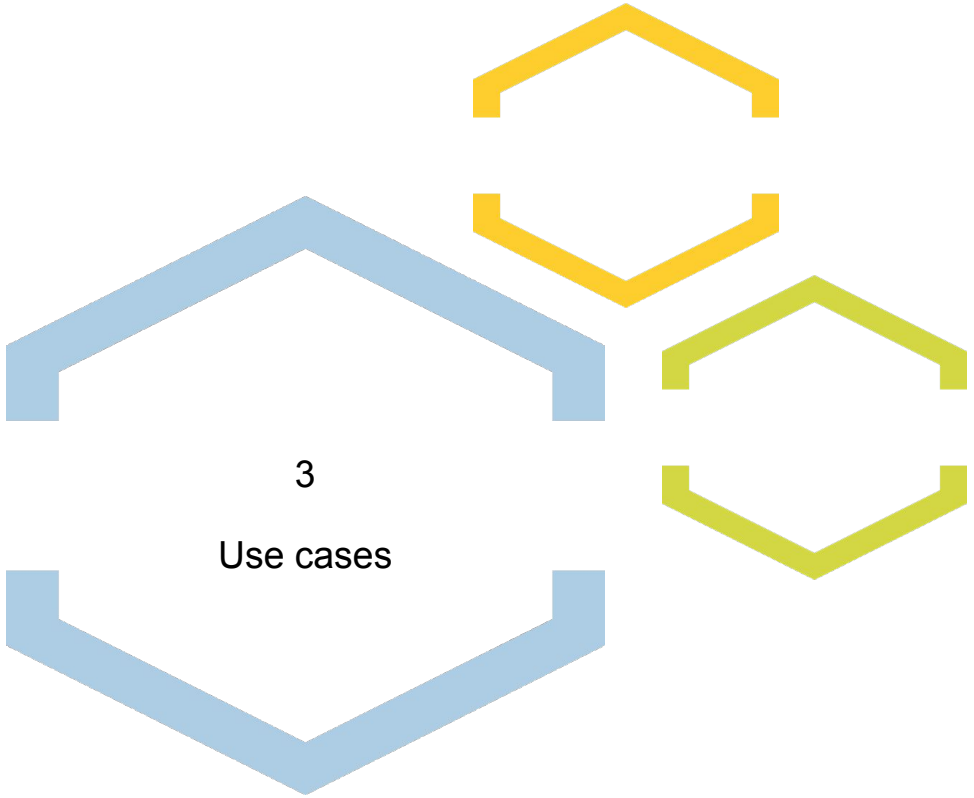
- PoC
  - Dockerized NiFi
  - Dockerized ELK Stack
  - Dockerized Microservice
- Easy to use and test!



```
git clone <repo>
cd <repo>/submodules/ms.semtweet/ && mvn clean package
cd <repo>/submodules/ && docker-compose up
```

1. NiFi
   - http://localhost:9090/nifi/
2. semtweet (microservice)
3. ElasticSearch
4. Kibana
   - http://localhost:5601/

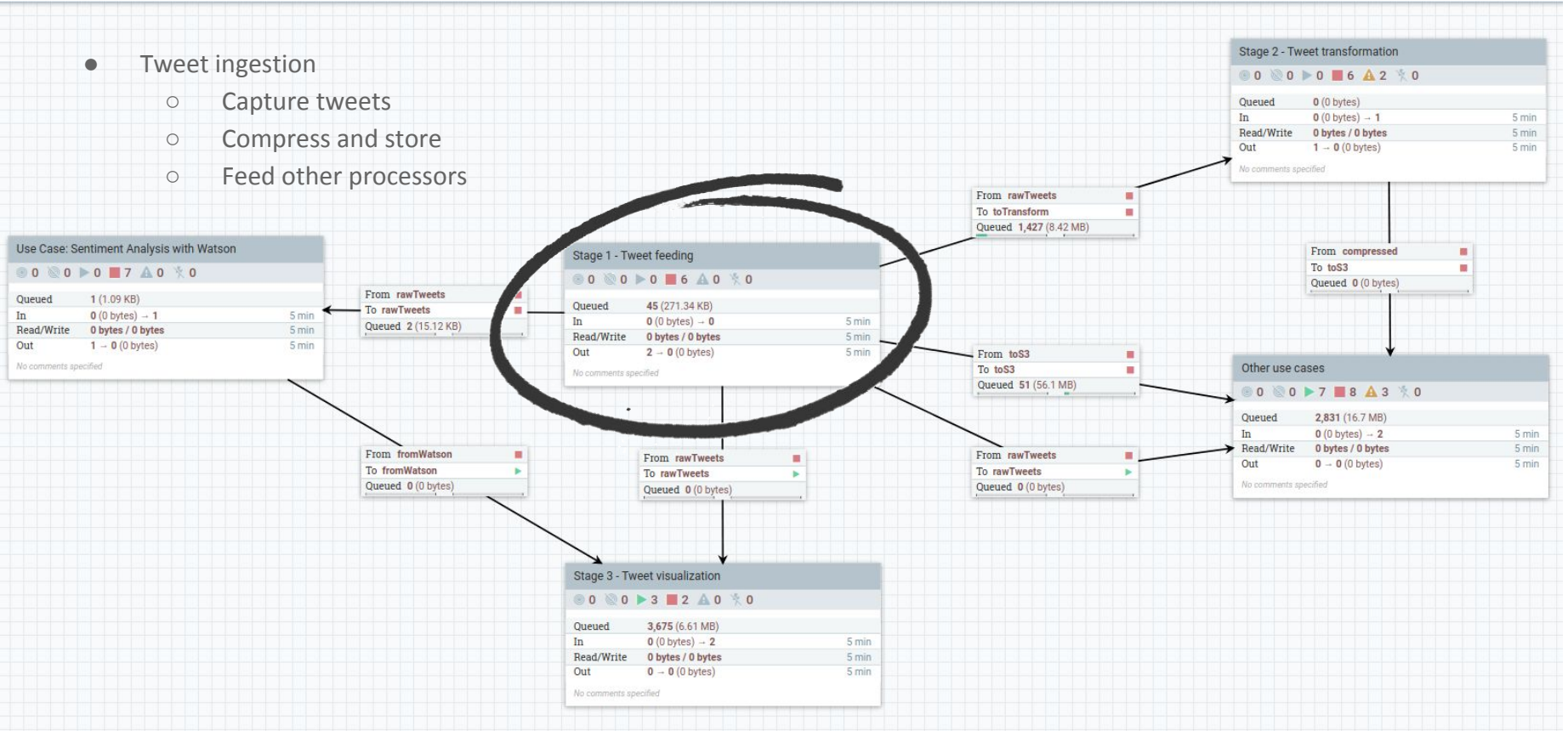




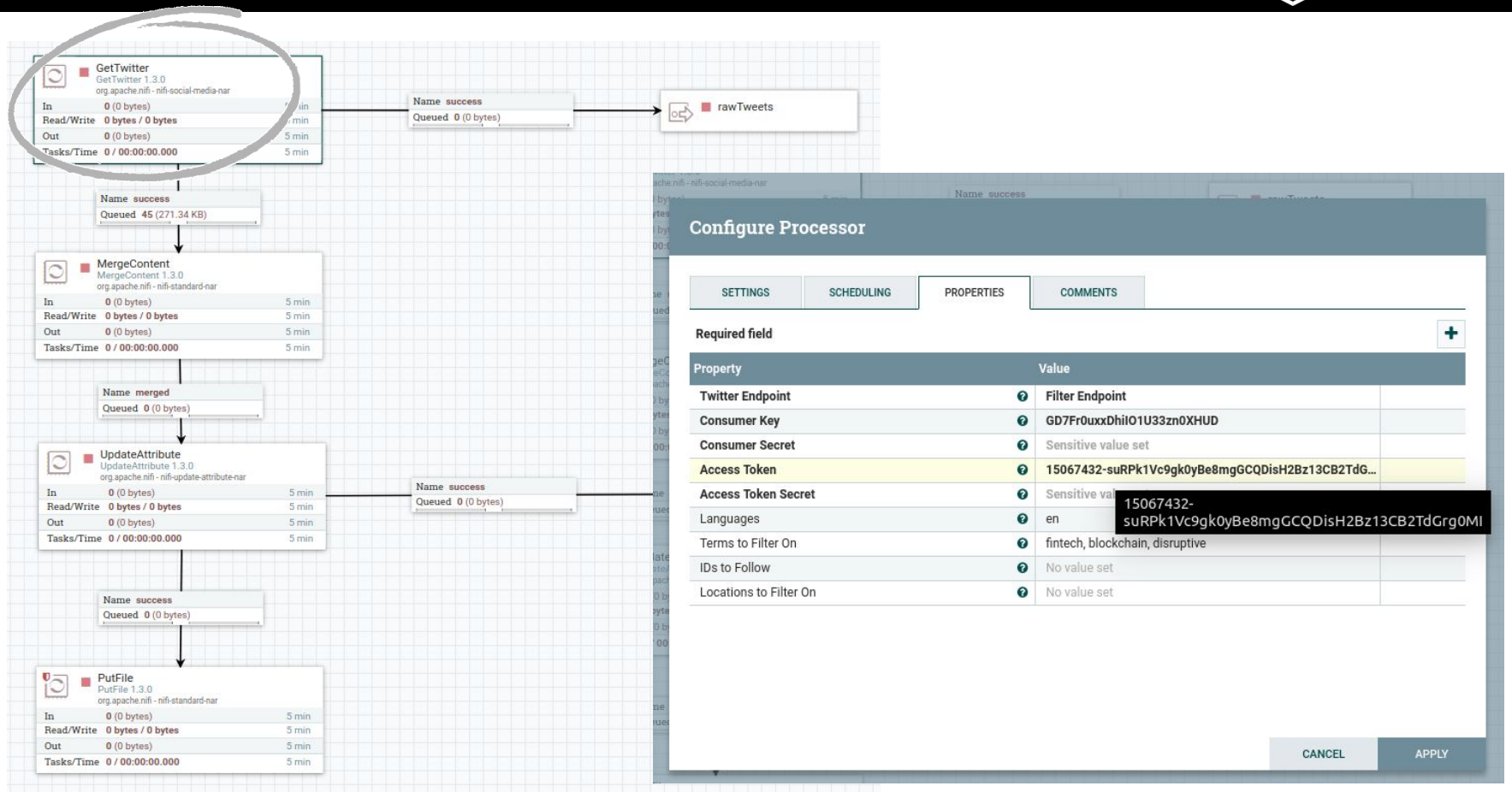
3

Use cases

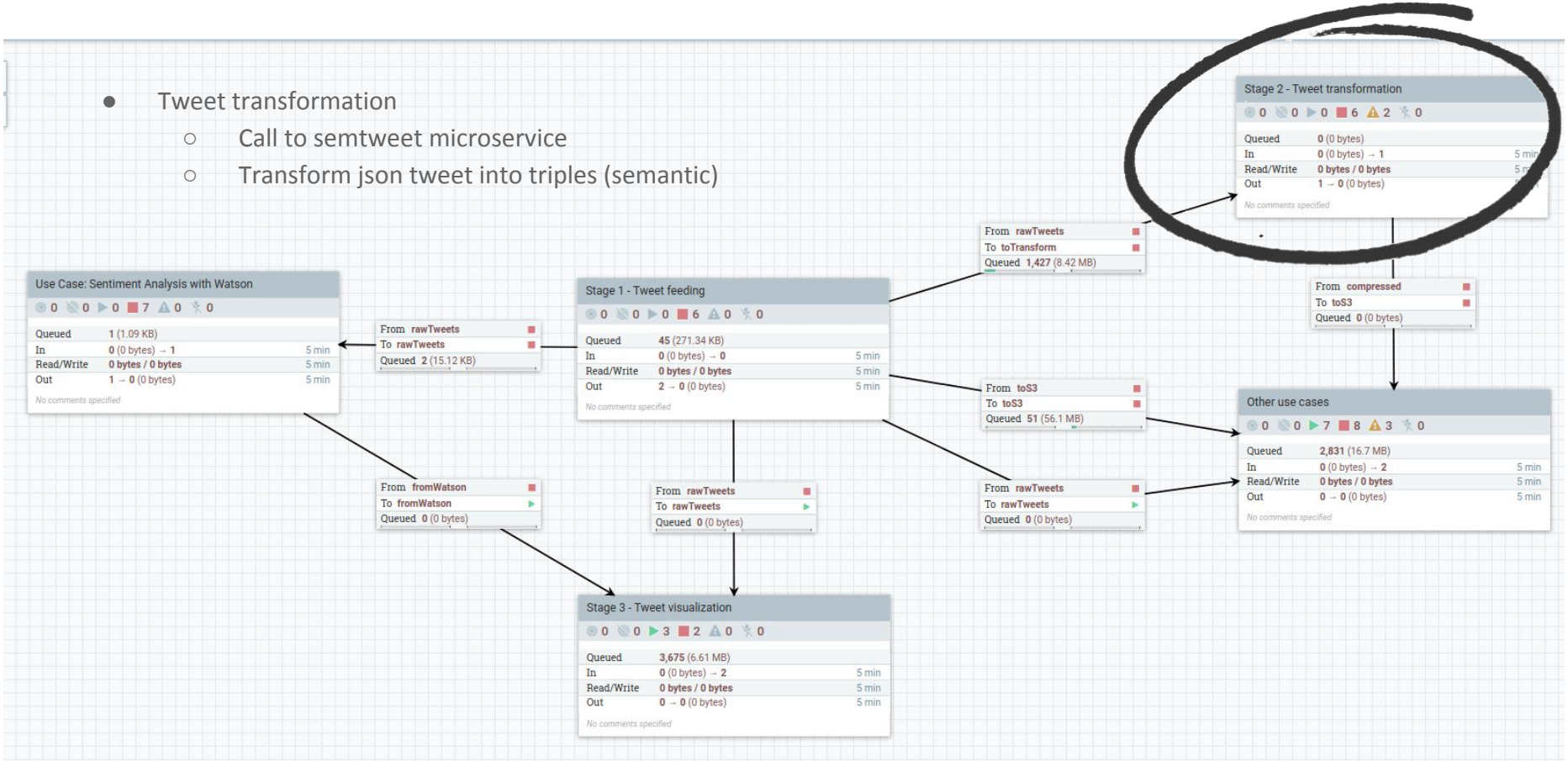
- Tweet ingestion
  - Capture tweets
  - Compress and store
  - Feed other processors



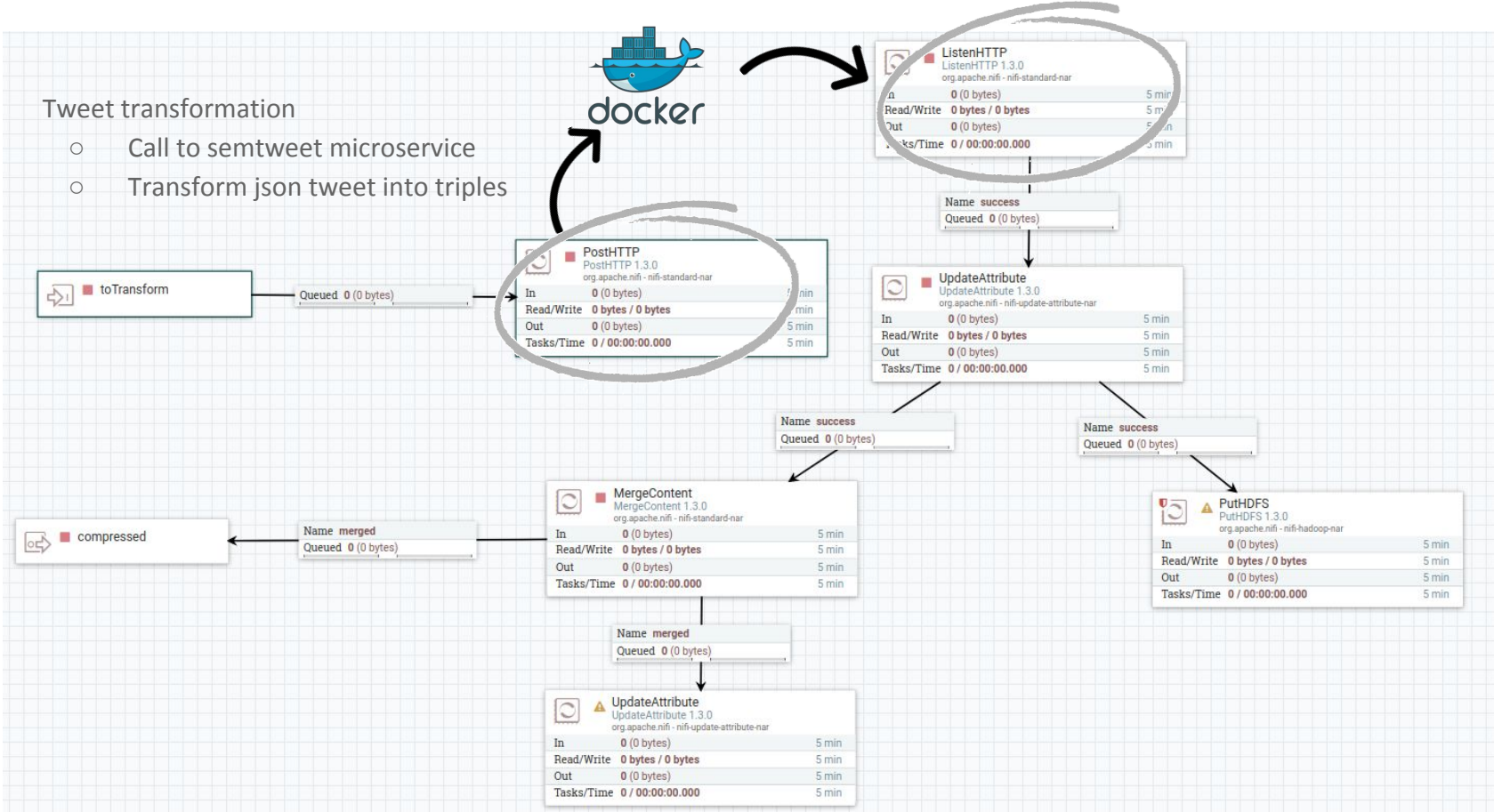
3. General topology - a) Tweet feeding



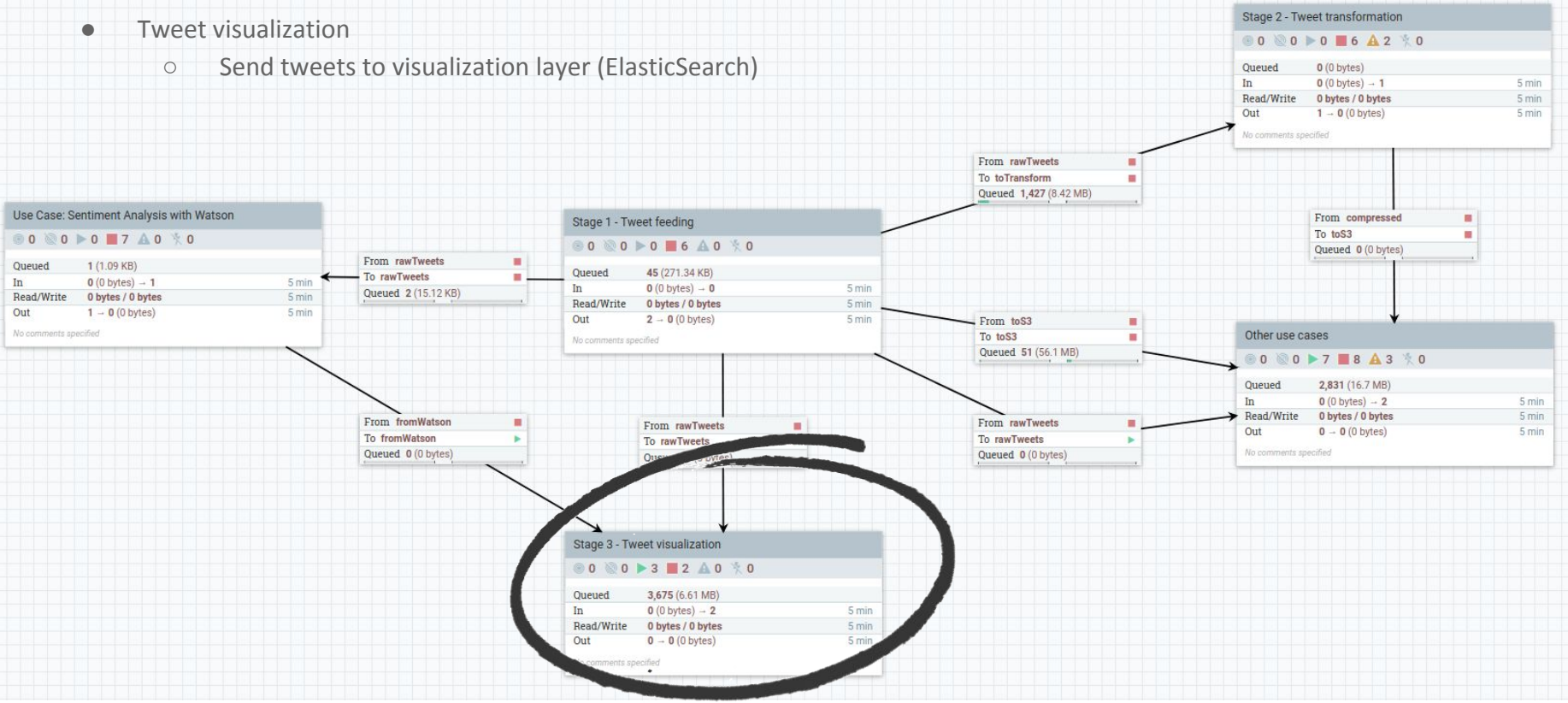
- Tweet transformation
  - Call to semtweet microservice
  - Transform json tweet into triples (semantic)

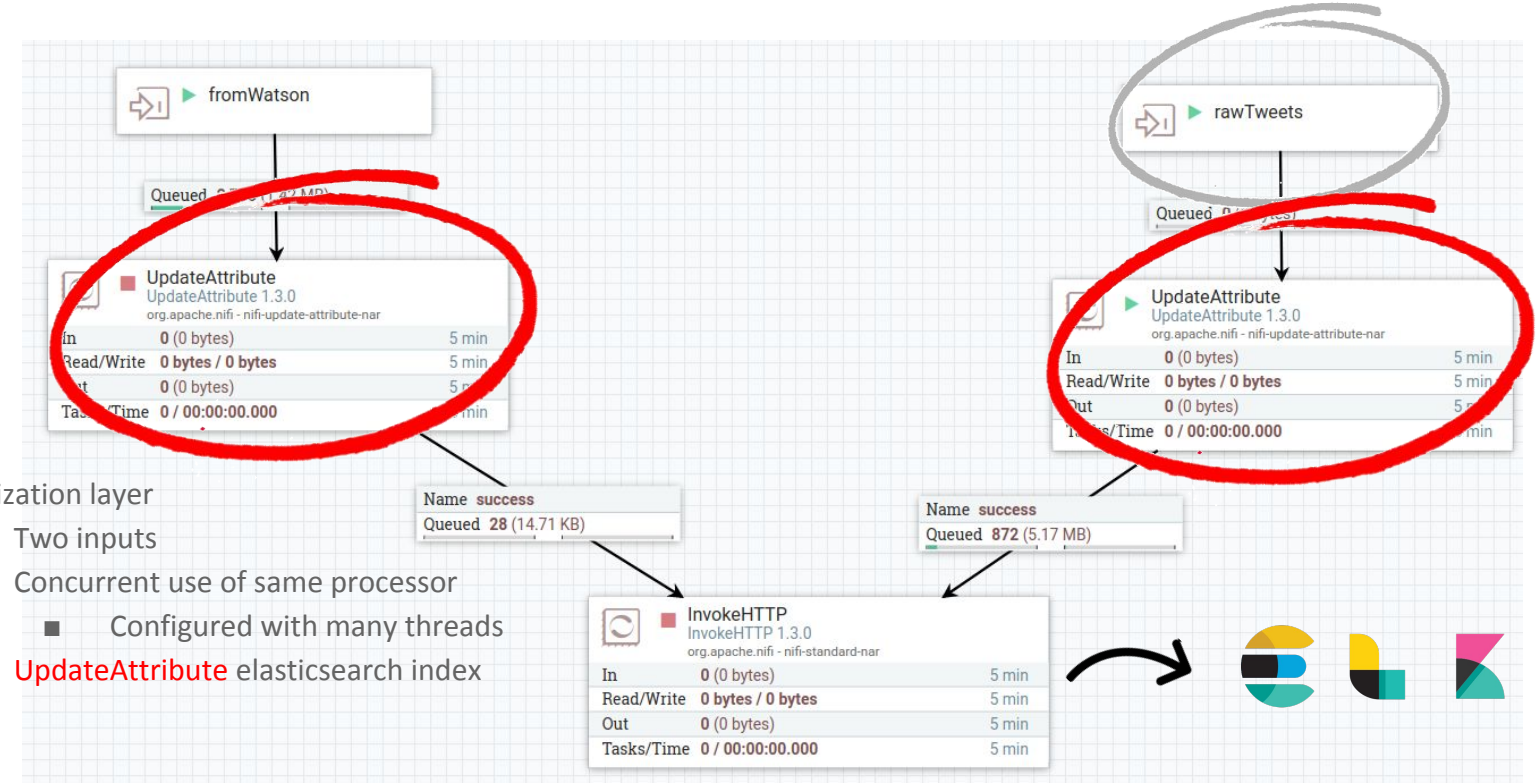


- Tweet transformation
  - Call to semtweet microservice
  - Transform json tweet into triples



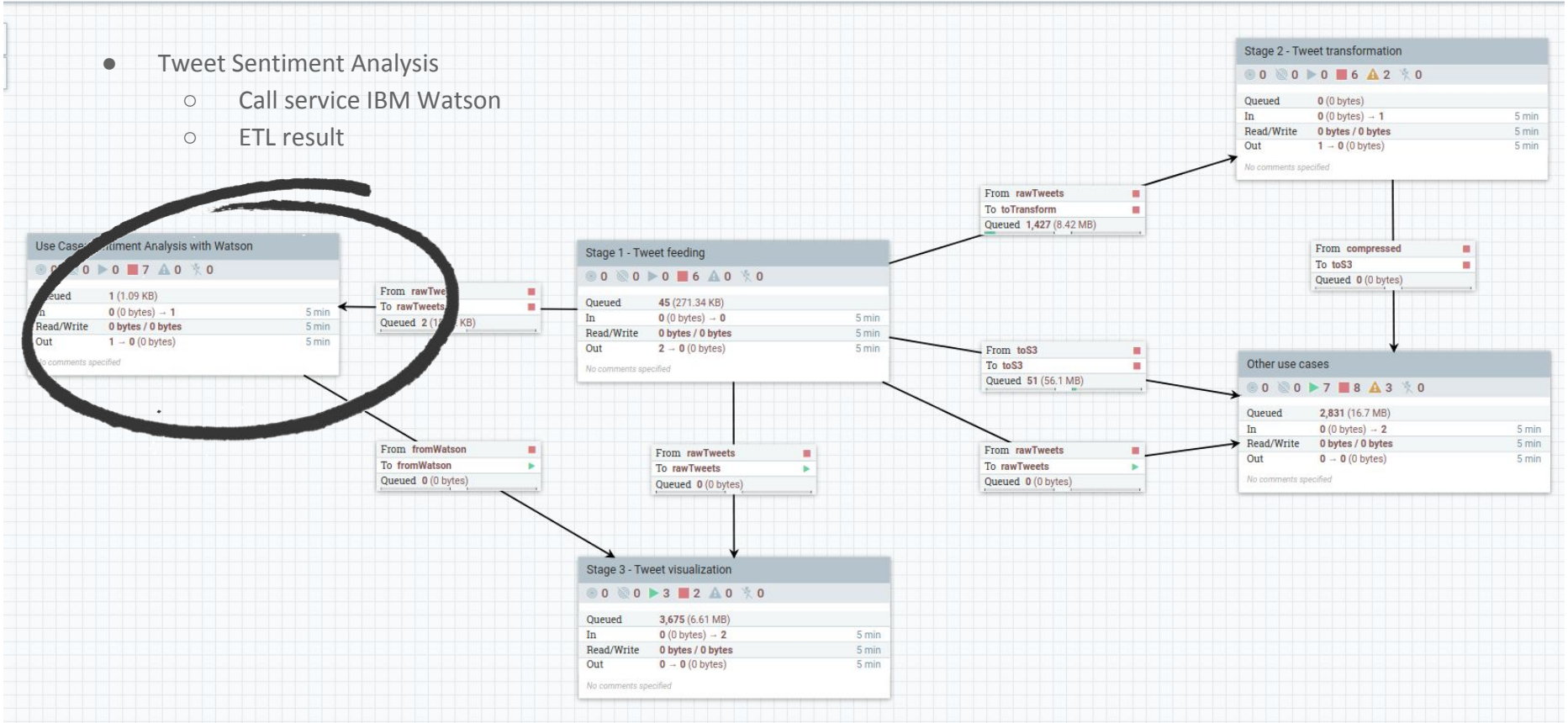
- Tweet visualization
  - Send tweets to visualization layer (ElasticSearch)



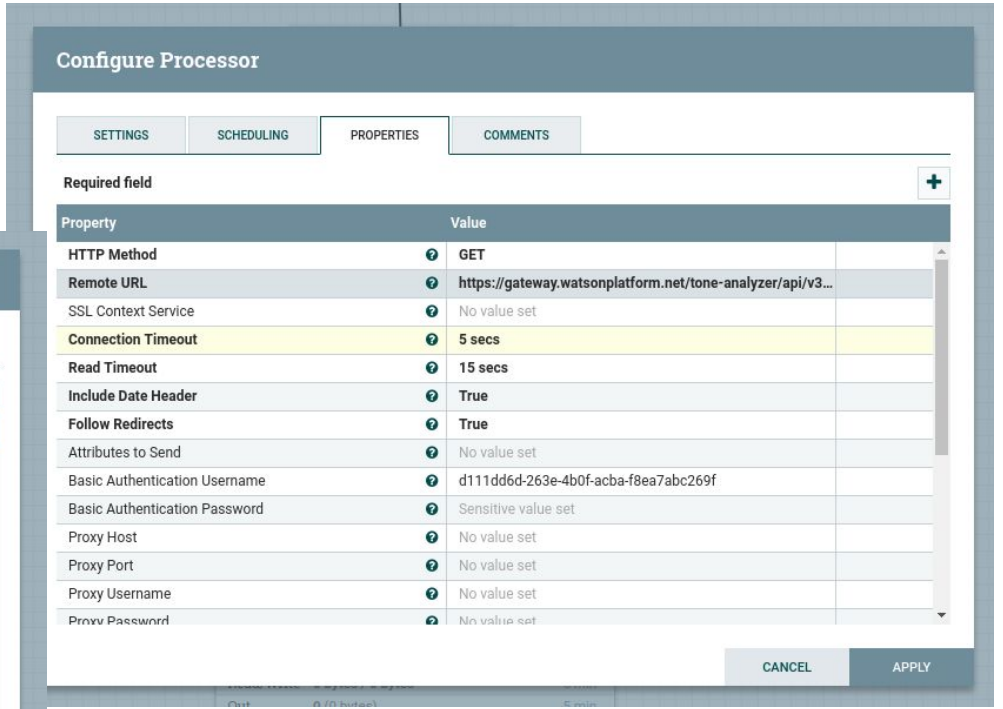


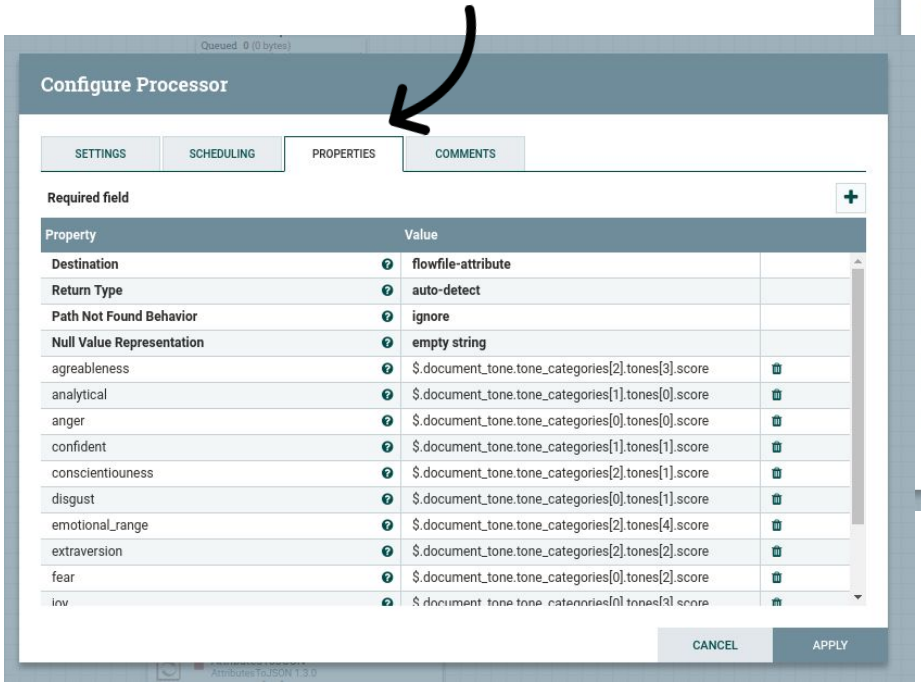
- Visualization layer
  - Two inputs
  - Concurrent use of same processor
    - Configured with many threads
  - UpdateAttribute elasticsearch index

- Tweet Sentiment Analysis
  - Call service IBM Watson
  - ETL result

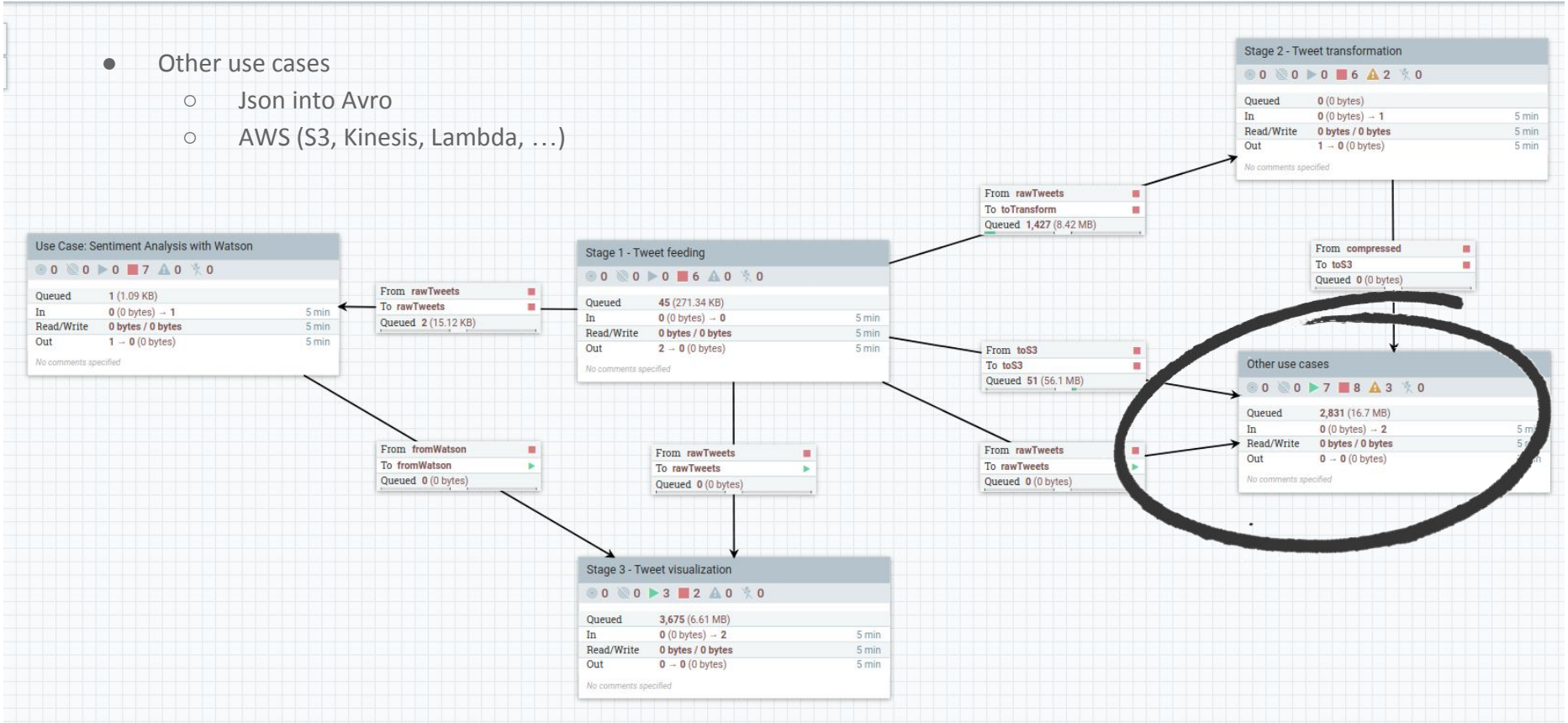


- Invoke HTTP (if you can *curl* it, you can use NiFi)
- Extract info from returned json



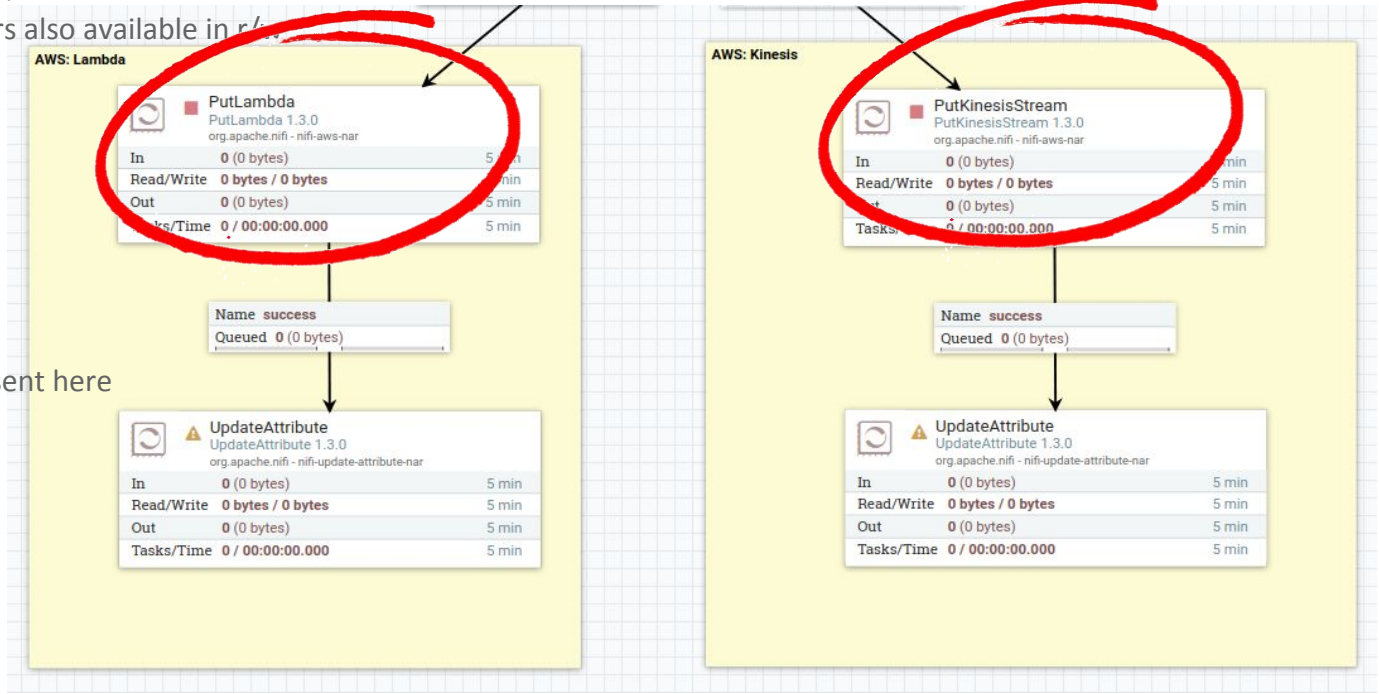


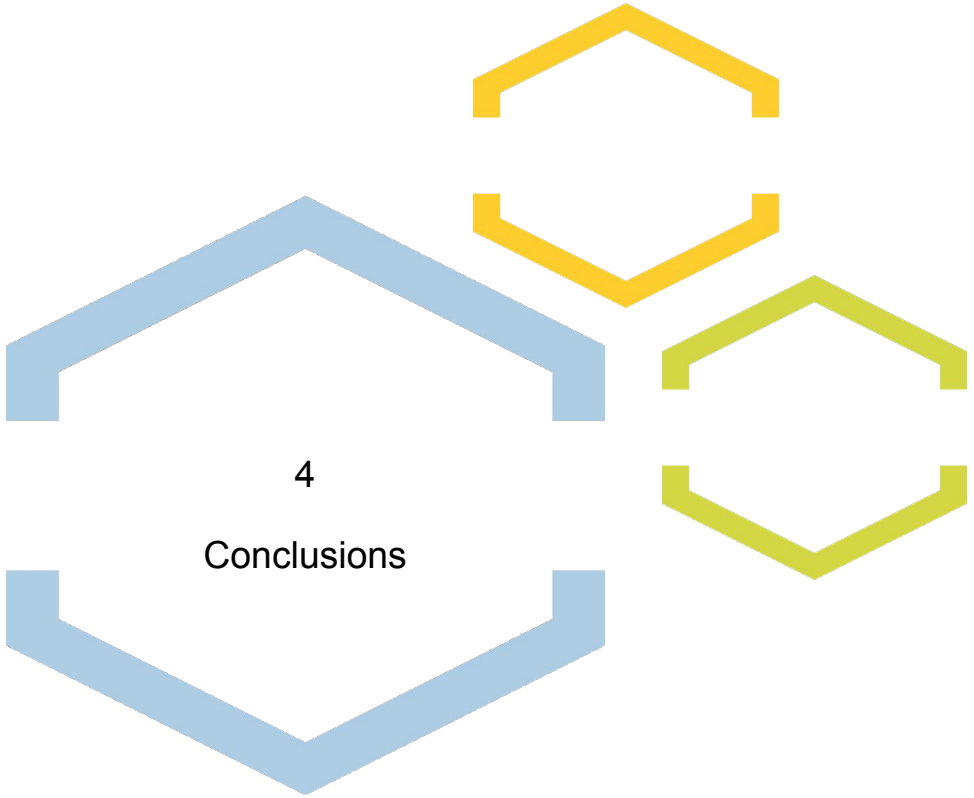
- Other use cases
  - Json into Avro
  - AWS (S3, Kinesis, Lambda, …)



- Used with
  - AWS Lambda: counting number of items in the json
  - AWS Kinesis (just insert)
    - Kafka processors also available in r/
  - AWS S3
    - List
    - Fetch
    - Put
    - HDFS friendly
  - Avro/Orc/Parquet/…

- Other tested features not present here
  - Flume integration
  - MongoDB
  - Cassandra
  - HBase
  - SFTP
  - …



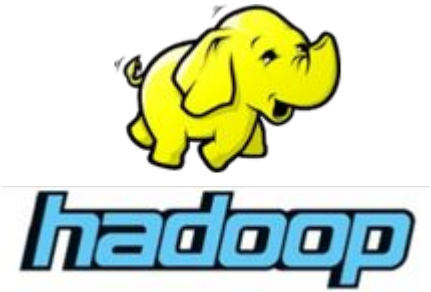


4

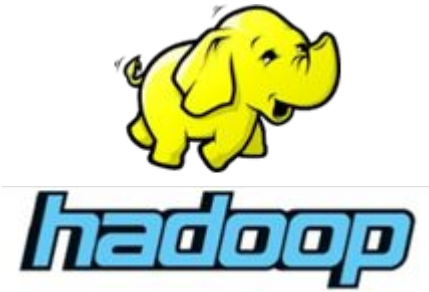
Conclusions

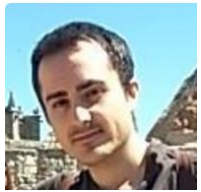
BEEVA

- What NiFi is…
  - General purpose Big Data Tool
  - **Integration**
  - **Versatile**
  - **Easy to use**



- **Beware**!
  - Generates so many dependencies
  - Tends to bring everything
  - **Addictive**





Raúl Reguillo Carmona

Software Engineer. Jack of all trades.

raul.reguillo@beeva.com |@Einath_Kelaven

BEEVA

hablemos@beeva.com

www.beeva.com