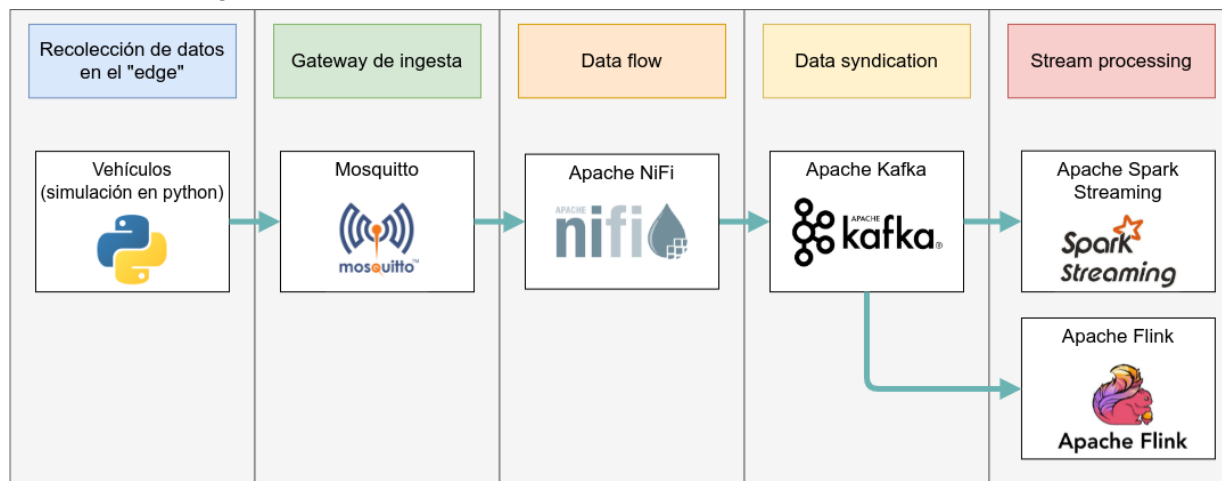


Introducción

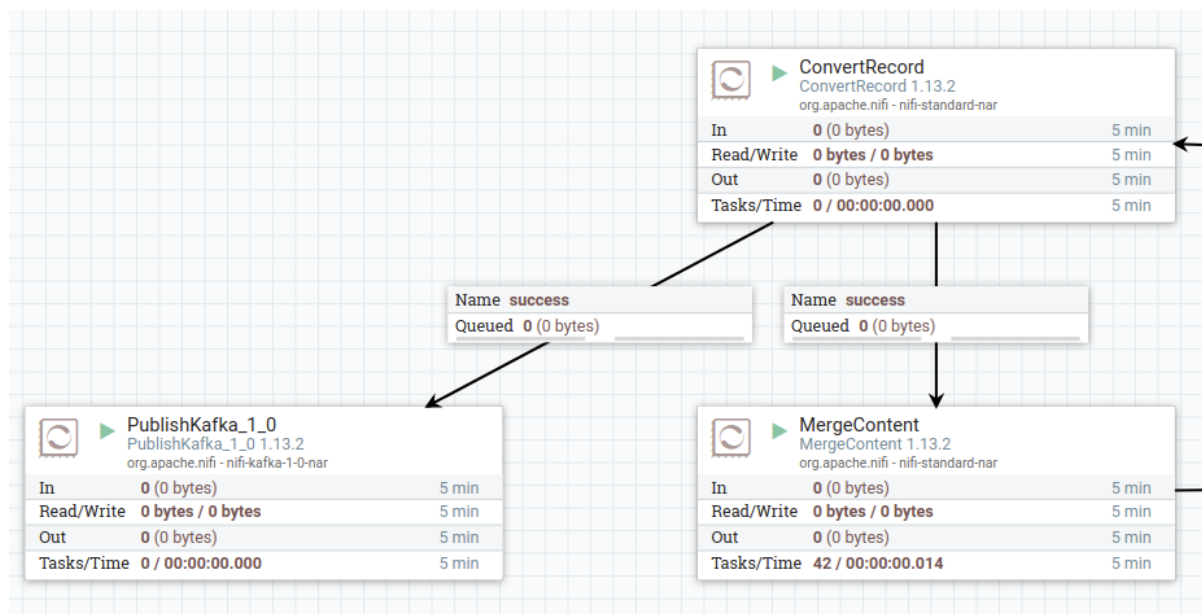
En esta práctica se continuará con una implementación de una arquitectura de referencia para IoT de la siguiente forma:



La implementación de la recolección de datos hasta el data flow fue realizada en prácticas anteriores. En esta práctica se realizarán cambios desde data flow en adelante, dejando Apache Flink para una siguiente práctica.

Cambios en NiFi

Vamos a bifurcar los datos del flujo de ingesta de la práctica de NiFi, para que sean enviados a Kafka. Los datos a enviar serán los resultantes del cambio de formato a CSV.



1. Agregar un processor PublishKafka_1_0 y configurarlo de la siguiente manera:
 - Conectarlo desde el ConvertRecord con success.
 - En la pestaña Properties, configurar:
 - 'Use Transactions': false.

- 'Topic Name': gps.
- En la pestaña Settings, seleccionar para auto terminar en failure y success.

2. Iniciar el processor nuevo y los processors anteriores. No es necesario que se inicie el flujo desde el MergeContent en adelante.

Creación del topic en Kafka

En Kafka deberemos crear el topic 'gps' al que se le enviarán los datos desde NiFi.

Ejecutar:

```
cd kafka
bin/kafka-topics.sh --create --zookeeper localhost:2181
--replication-factor 1 --partitions 1 --topic gps
```

Procesamiento

En Jupyter Notebook iniciar el notebook Consumer-Kafka-Structured-Streaming. Este va a consumir la información de Kafka y procesarla.

Para iniciar todo el flujo (Mosquitto -> NiFi -> Kafka -> Spark streaming) se deberá ejecutar el simulador de información. En el directorio del simulador ejecutar:

```
/home/azureuser/conda/bin/python reproductor.py
```

Ver en el notebook la información entrante y procesada.

Al finalizar la práctica, detener el notebook, el envío de información del reproductor y el flujo en NiFi.