



Universitat Rovira i Virgili



Integración de un problema de IoT con soluciones de BigData

María José López Osa

TRABAJO FINAL DE MÁSTER

Iván Rodero Castro

MÁSTER EN INGENIERÍA COMPUTACIONAL Y
MATEMÁTICAS

Bilbao
2017

Estructura de la presentación:

- Objetivo
- Tecnologías involucradas
- Integración en una plataforma
- Demostración
- Conclusiones

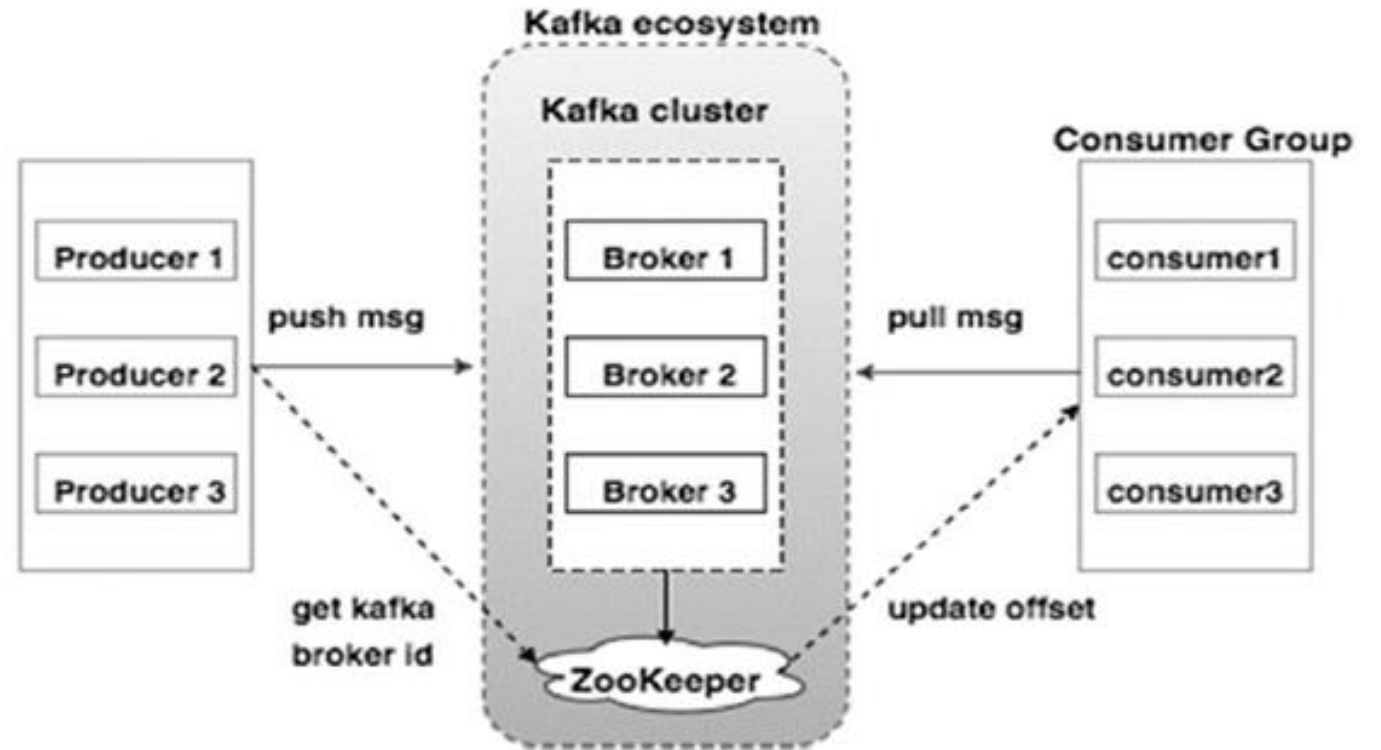
Objetivos

1. Integrar en un mismo sistema, soluciones de IoT y de BigData, junto con sistema de almacenamiento distribuidos.
2. Implementar un caso de uso para medir velocidades del viento en distintos puntos y ofrecer información en base a estas mediciones.

Tecnologías involucradas (I)



Kafka es una plataforma open-source para el procesamiento de datos en streaming o transmisión continua realizada por la Apache Foundation e implementada en Scala y Java.





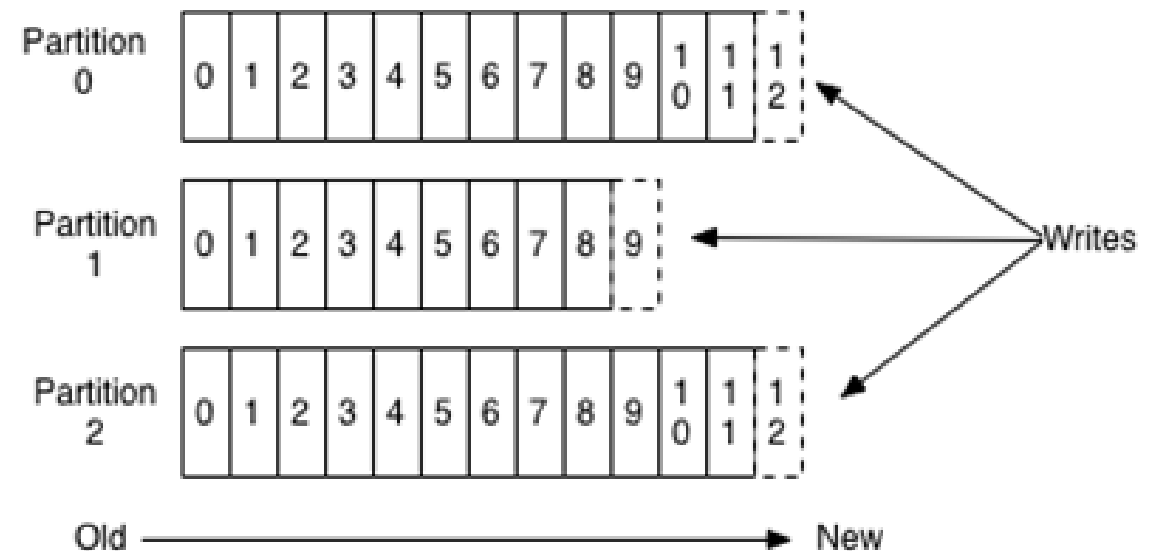
Tecnologías involucradas (II)

El elemento clave en Kafka es el **topic**.

Los producers publican sus registros en un topic para que los consumers se suscriban a ellos.

Los topics se dividen en particiones. Para cada topic, Kafka mantiene un log de particiones.

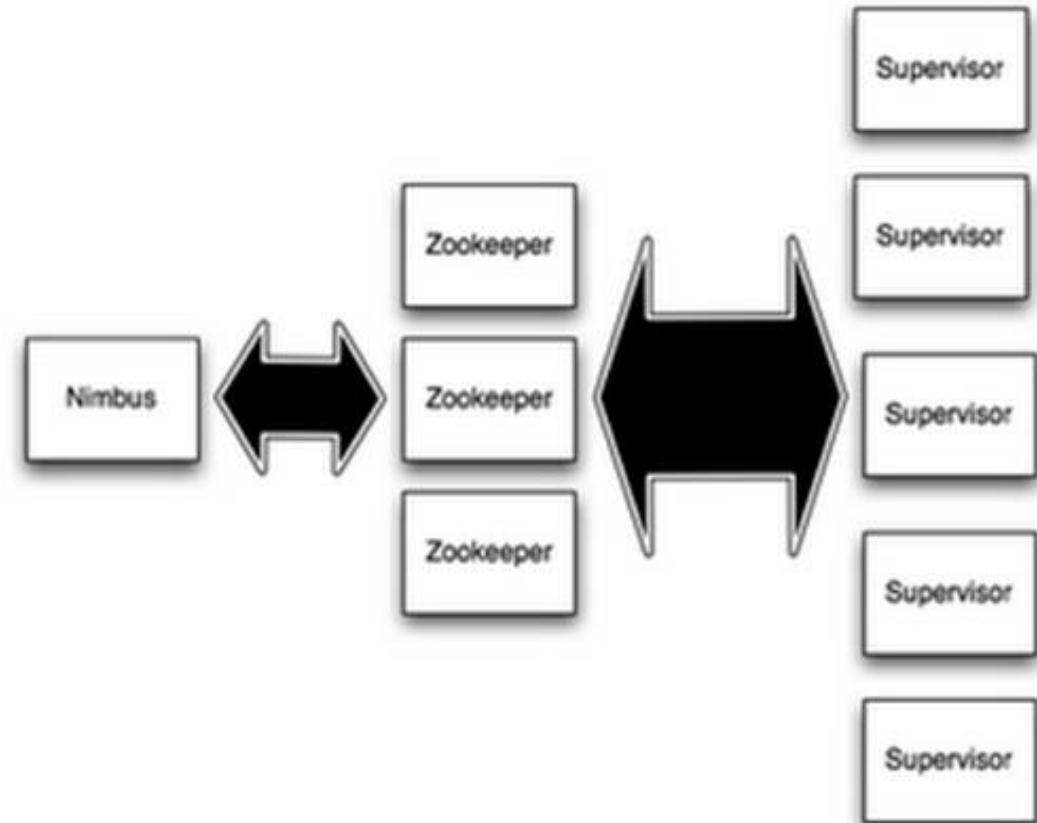
Anatomy of a Topic



Tecnologías involucradas (IV)



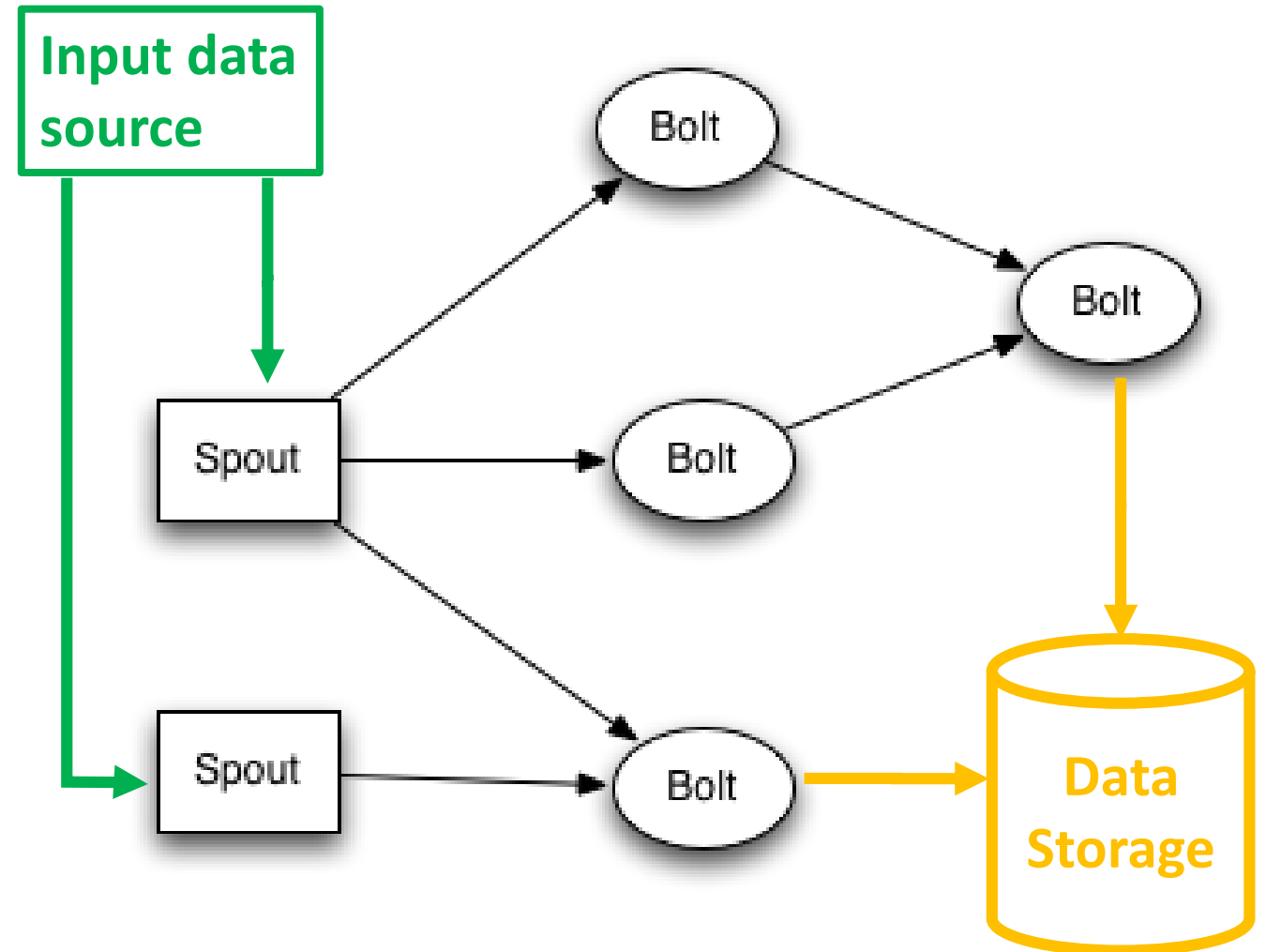
Apache Storm es un sistema open Source de computación en tiempo real distribuido y libre. Storm facilita el procesamiento de forma fiable de flujos de datos en cantidades sin límite, de una forma simple, y utilizable con cualquier lenguaje de programación



Tecnologías involucradas (V)



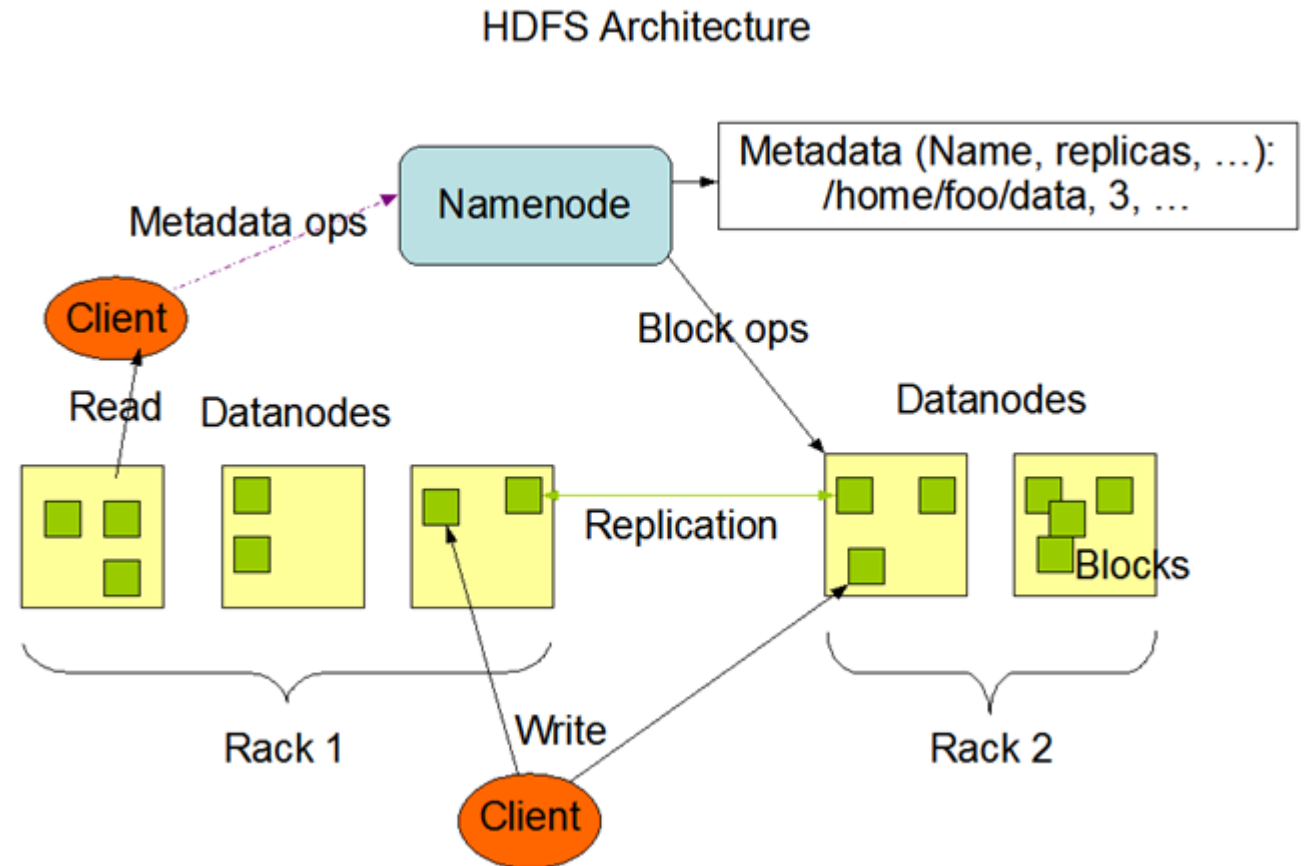
En Storm encontramos con el concepto de “Topologías” que se despliegan en un cluster. Una topología es un grafo de computación, donde cada nodo contiene lógica de proceso y donde los enlaces entre los nodos indican cómo pueden ser transmitidos los datos entre los nodos que forman esa topología.



Tecnologías involucradas (VI)

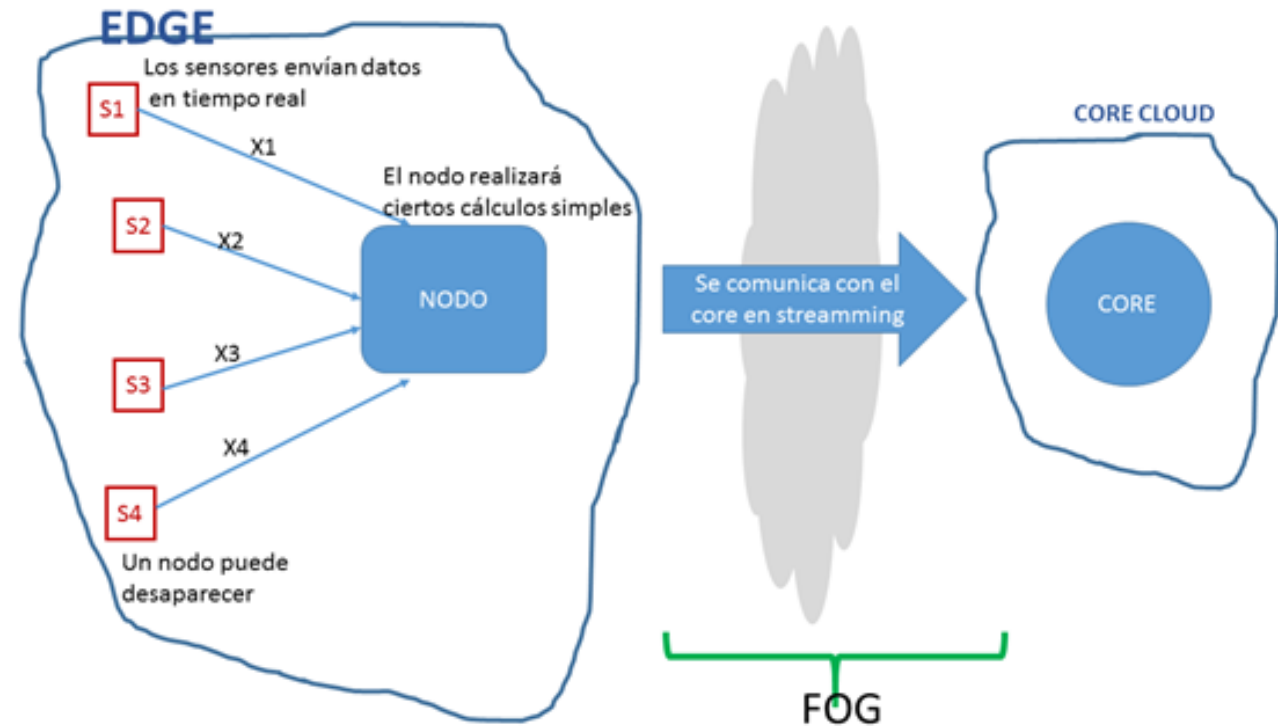


Apache Hadoop es un entorno de trabajo open-source que se utiliza para almacenaje distribuido y procesamiento de conjuntos de datos de BigData, usando el modelo de programación MapReduce. La parte básica de Apache Hadoop consiste en una parte de almacenaje, conocida como HDFS (Hadoop Distributed File System).



Integración en una plataforma (I)

Diseño de una plataforma que incluyera sensores que envían datos a un nodo o sistema, que realiza un procesamiento y lo traslada a otro sistema para que realice los procesamientos necesarios para obtener la información requerida.



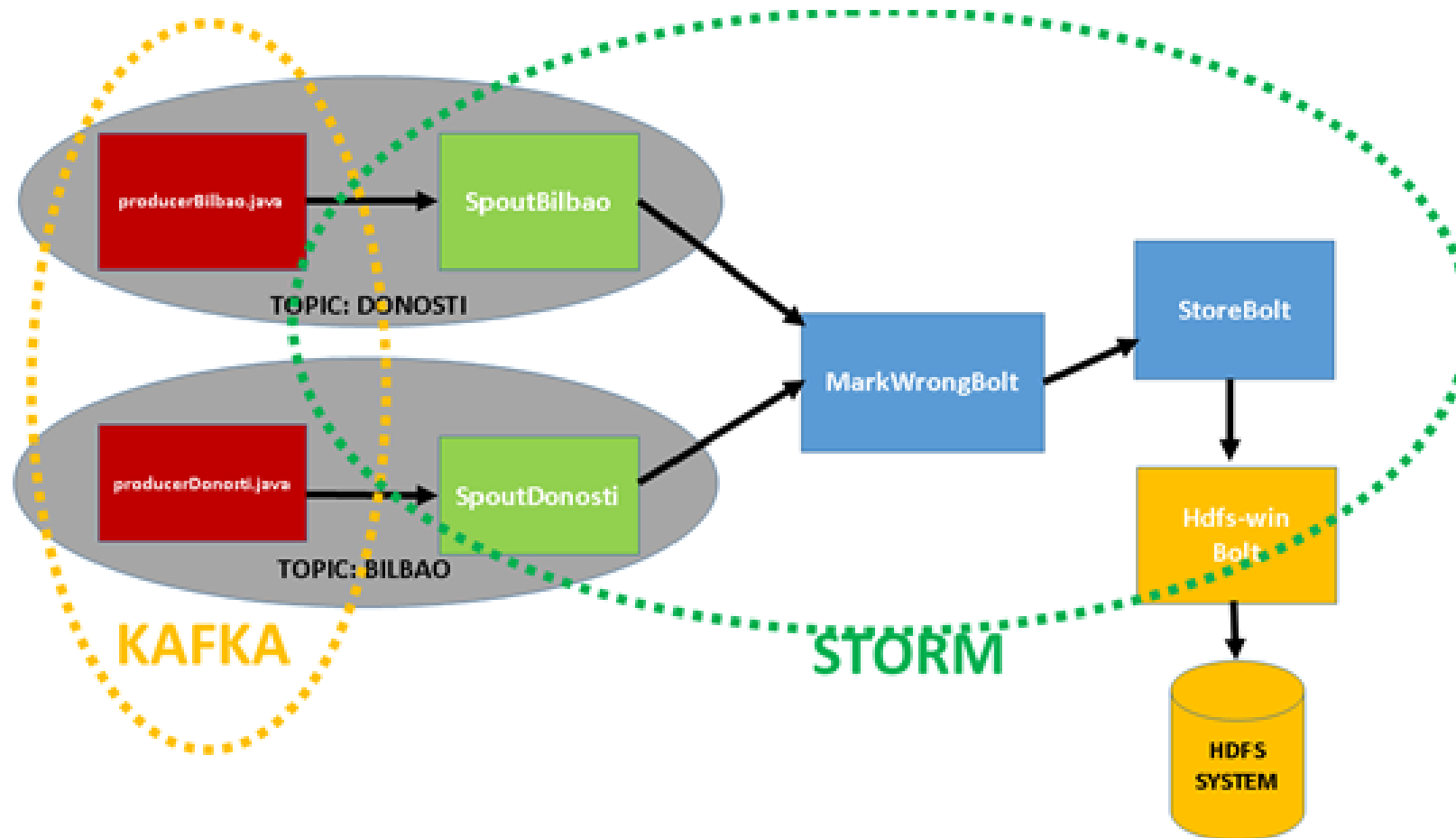
Integración en una plataforma (II)

La implementación se centra en recoger información de unos nodos, inyectar esos datos en el sistema kafka, procesar los datos recibidos en el sistema Storm.



Integración en una plataforma (III)

La arquitectura del caso de uso incluye el desarrollo central, más la producción de datos simulados por producers kafka, y un almacenaje en un sistema hdfs.



Demonstración

Conclusiones

Es posible desplegar un problema de IoT, integrando plataformas asociadas al mundo del BigData y el almacenamiento distribuido, y llevando los datos de un extremo a otro de la cadena de forma automática e integrada.

Los tutoriales y las páginas de “gettingStarted” de muchas tecnologías existentes actualmente, aportan información incompleta.

Maven es una herramienta muy valiosa para la gestión de dependencias en un proyecto de eclipse y java.

El método seguido para la realización del trabajo ha resultado ser el idóneo.