



# Registro de Huella Durante el Procesamiento Masivo de Datos

---

Carlos Roberto Fernandez Lino

[croberto.fernandez@gmail.com](mailto:croberto.fernandez@gmail.com)

# Carlos Fernandez

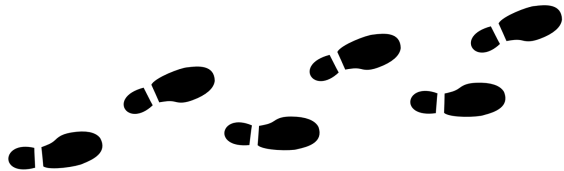
*Ingeniero en Software con más de 10 años de experiencia en desarrollo de aplicaciones Java y tecnologías relacionadas. Actualmente trabajo en Mojix. Algunas de mis áreas de interés y experiencia son computación de altas prestaciones, diseño de sistemas, internet de las cosas y linux. Además he sido docente universitario y actualmente soy coordinador del equipo de organización del Java User Group Bolivia.*



# En esta charla...

veremos un mecanismo construido en Java para registrar el camino que siguió un determinado dato dentro una cadena de proceso de eventos construida en Apache Kafka con el objetivo de depurar el estado de los datos a lo largo de toda la cadena de procesamiento.

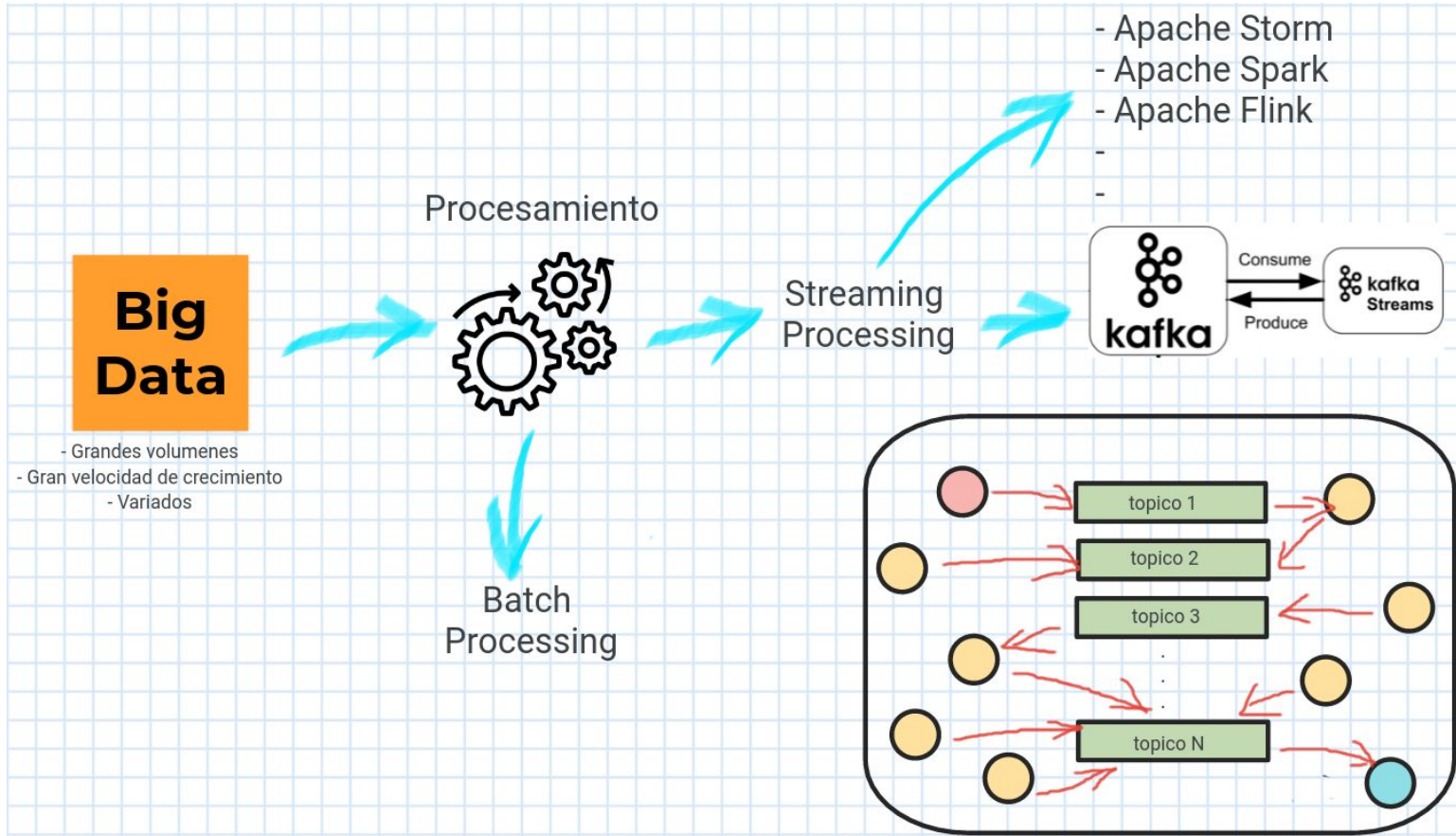
# Agenda



- Introducción
- Descripción del problema
- Prototipo planteado
- Demo
- Conclusiones



# Introducción



# Descripción del problema

## 1. Carlos...

The message is lost!!!!

Where is my message?

Questions and more questions about the messages....

There is a bug!! I cannot see my messages

This message is wrong!!!, I am reporting an issue

## 2. mmmm

I'm not sure, let's track it...

I don't know ... :(

let's review....

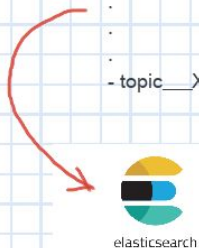
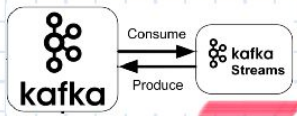
I don't think so let me review...

are you sure, can we review first?

let's track ...

## 3. if I build something to track messages?...

- topic \_\_A  
- topic \_\_B  
- topic \_\_C  
.  
.  
- topic \_\_X

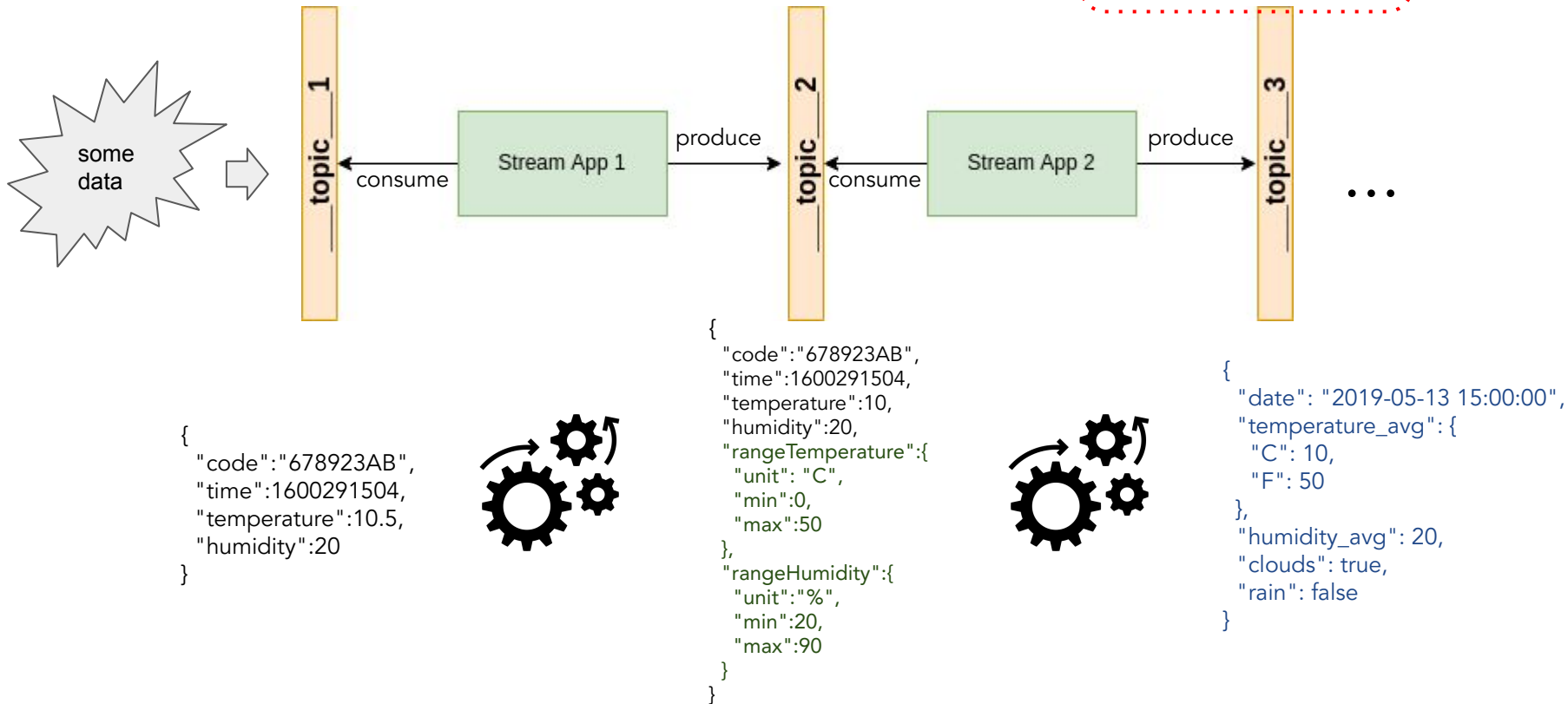
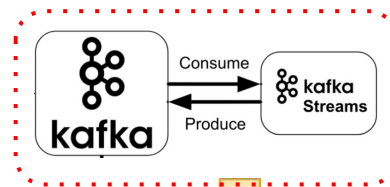


- It searches fast
- Time based indexes
- It is good for logging
- Powerful query criteria
- It has REST interface

## 4. Goal

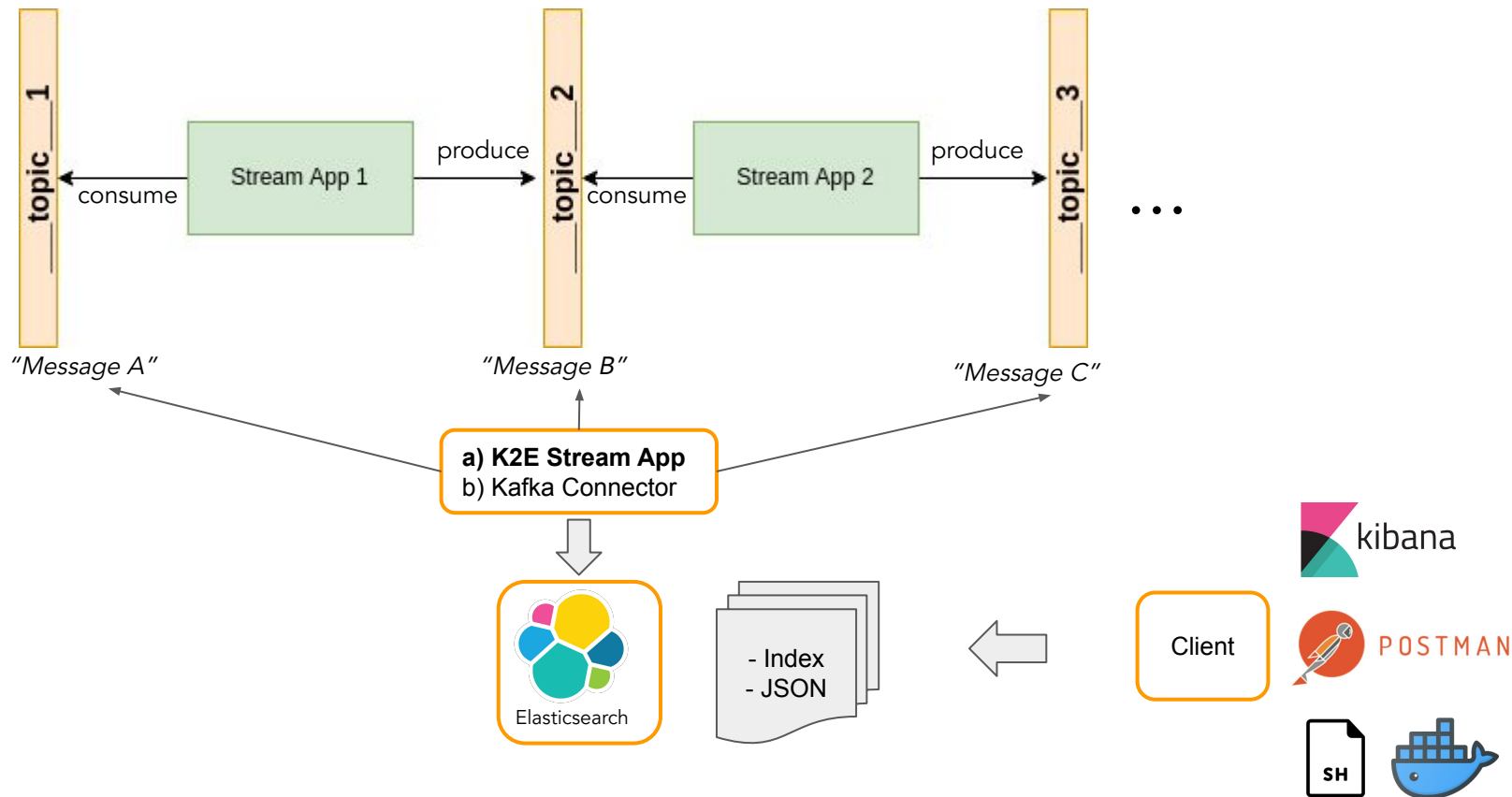
Track the entire pipeline and know the route of a message.

# Descripción del problema





# Prototipo planteado





# iDemo!

<https://github.com/carfernandez16/streams-footprint>

# Conclusiones

## - ¿Funcionó la solución?

*Si funciona, pero hay problemas a considerar.*

## - ¿Cuales son los principales problemas?

*El transformación que sufren los mensajes adiciona complejidad a la identificación de la ruta que siguió el mensaje. El tamaño y formato de mensaje influyen bastante.*

## - ¿Es posible conocer la ruta del mensaje en toda la cadena de procesamiento?

*No es posible siguiendo una manera simple y no intrusiva.*

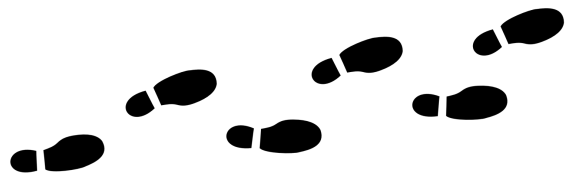
# Conclusiones

- **¿Además de Elasticsearch que otras opciones podría usar?**

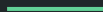
*Si, lo mejor es usar una herramienta que se ajuste a tu escenario. (manejo, implementación, costos)*

- **¿Como solucionarías el problema del polimorfismo de la información?**

*Una alternativa sería la incorporación estratégica de metadata en los mensajes para toda la cadena de procesamiento.*



# Preguntas



# ¡Gracias!

---

Carlos Roberto Fernandez Lino

<https://twitter.com/carfernandez16>

<https://github.com/carfernandez16>