



Andrei Miu

# GESTIÓN DE EVENTOS EN TIEMPO REAL

## Tarea Unidad 5

Analizar y plantear las distintas necesidades, retos y soluciones al gestionar y transformar datos en tiempo real, así como explorar las opciones para unificar estos flujos con los procesos por lotes en una plataforma común

## Índice

Identificación de necesidades y retos .....	3
Análisis de soluciones para datos en tiempo real .....	3
Integración con procesos por lotes .....	4
Investigación de plataformas .....	4
Ventajas y desventajas .....	4
Comparativa: Tiempo Real vs Lotes .....	4
Esquema conceptual .....	4
Conclusiones .....	4
Bibliografía .....	5

## Identificación de necesidades y retos

### Escenarios donde se necesitan datos en tiempo real:

- **Monitorización de dispositivos IoT:** Sensores en una fábrica que reportan fallas en tiempo real.
- **Análisis del comportamiento del cliente:** Plataformas de e-commerce que muestran productos recomendados al instante.
- **Detección de fraudes:** Bancos que analizan transacciones sospechosas mientras ocurren.

### Principales retos técnicos y organizativos:

- **Latencia:** Minimizar el retraso entre el evento y el análisis.
- **Escalabilidad:** Manejar un creciente volumen de datos.
- **Integración con sistemas heredados:** Conectar nuevas tecnologías con infraestructura antigua.

## Análisis de soluciones para datos en tiempo real

### Herramientas para procesar datos en tiempo real:

- **Apache Kafka:** Sistema distribuido de mensajería con baja latencia.
- **Apache Flink:** Procesamiento en tiempo real con tolerancia a fallos.
- **Spark Streaming:** Extensión de Apache Spark para flujos de datos.

### ¿Cómo abordan los retos y cuáles son los criterios de selección de cada una?

Los criterios al afrontar los retos suelen ser:

- **Latencia:** Cuánto tiempo tarda en procesar eventos.
- **Coste:** Recursos necesarios para operar la herramienta.
- **Facilidad de uso:** Curva de aprendizaje y compatibilidad con sistemas actuales.

**Apache Kafka:** Es Open Source, es decir es gratuito si lo configuras y mantienes tú mismo en tus propios servidores o máquinas virtuales.

**Apache Flink:** Es gratuito si lo ejecutas en tu propia infraestructura. Necesitaras un clúster con nodos, también un técnico para configurarlo y mantenerlo. Consume CPU y RAM significativos para procesar datos en tiempo real.

**Spark Streaming:** Es gratuito si lo instalas y gestionas tú mismo. Requiere bastantes recursos para procesar grandes volúmenes de datos.

<u>Herramienta</u>	<u>Servicio</u>	<u>Costes</u>
<b>Kafka</b>	Confluent Cloud	Infraestructura, discos
<b>Flink</b>	Amazon Kinesis	Recursos de computación
<b>Spark Streaming</b>	AWS EMR	HDFS, Clúster

## Integración con procesos por lotes

### Investigación de plataformas

Algunas plataformas unificadas:

- **Apache Beam:** Modelo único para flujos por lotes y en tiempo real.
- **Databricks:** Integración nativa de Spark para ambos enfoques.
- **Snowflake:** Soporte híbrido para consultas en tiempo real y por lotes.

### Ventajas y desventajas

Las ventajas y desventajas de las plataformas unificadas:

- **Ventajas:** Reducción de complejidad, reducción de costos, acceso de datos centralizado y unificación de herramientas.
- **Desventajas:** Alto coste inicial de implementación, problemas de escalabilidad y curva de aprendizaje inicial.

### Comparativa: Tiempo Real vs Lotes

- **Tiempo real:** Baja latencia, datos inmediatos, volúmenes pequeños pero continuos.
- **Por lotes:** Procesamiento periódico de grandes volúmenes de datos.

## Esquema conceptual

**Arquitectura:**

- Ingesta de datos con Kafka.
- Procesamiento en tiempo real con Flink.
- Almacenamiento de datos históricos en un Data Lake o Warehouse.
- Procesos por lotes con Spark.

**Cómo aborda los retos:**

- **Coherencia de datos:** Uso de un sistema centralizado para eventos y lotes.
- **Escalabilidad:** Herramientas distribuidas.
- **Mantenimiento:** Uso de plataformas estándar

## Conclusiones

Los procesos en tiempo real requieren procesamiento inmediato para proporcionar resultados al instante, mientras que los procesos por lotes agrupan datos para ser procesados de forma periódica.

¿Cuándo usar una **plataforma unificada**? Cuando se busca:

- Centrar las operaciones
- Prioriza la eficiencia
- Si se necesita una experiencia de usuario

¿Cuándo mantener los **sistemas separados**? Cuando:

- Existe un riesgo de dependencia alto
- Tu presupuesto es ajustado inicialmente
- Cuando las tareas de la empresa son específicas

Los **servicios administrados** son más caros, pero reducen la complejidad operativa. Esto es ideal para empresas que buscan rapidez en la implementación.

## Bibliografía

<https://www.datacentermarket.es/mercado/cuatro-consejos-para-el-procesamiento-de-datos-en-tiempo-real/>

<https://www.solvexia.com/blog/real-time-analytics>

[https://www.simplilearn-com.translate.goog/tutorials/data-analytics-tutorial/applications-of-data-](https://www.simplilearn-com.translate.goog/tutorials/data-analytics-tutorial/applications-of-data-analytics? x tr sl=en& x tr tl=es& x tr hl=es& x tr pto=rq#:~:text=From%20healthcare%20and%20finance%20to,impact%20on%20businesses%20and%20society.)

[analytics? x tr sl=en& x tr tl=es& x tr hl=es& x tr pto=rq#:~:text=From%20healthcare%20and%20finance%20to,impact%20on%20businesses%20and%20society.](https://www.simplilearn-com.translate.goog/tutorials/data-analytics-tutorial/applications-of-data-analytics? x tr sl=en& x tr tl=es& x tr hl=es& x tr pto=rq#:~:text=From%20healthcare%20and%20finance%20to,impact%20on%20businesses%20and%20society.)

<https://www.confluent.io/es-es/confluent-cloud/>

<https://aws.amazon.com/es/kinesis/>

[https://docs.aws.amazon.com/es\\_es/emr/latest/ManagementGuide/emr-what-is-emr.html](https://docs.aws.amazon.com/es_es/emr/latest/ManagementGuide/emr-what-is-emr.html)

<https://www.vailos.com/etl-tiempo-real-etl-por-lotes-mejor-opcion/>

<https://www.avatech.nz/post/real-time-vs-batch-processing>

<https://flink.apache.org/>

<https://spark.apache.org/docs/latest/index.html>

<https://www.ibm.com/es-es/topics/hdfs>

<https://aprenderbigdata.com/stream-processing/#:~:text=en%20el%20momento.-,%C2%BFes%20Flink%20m%C3%A1s%20r%C3%A1pido%20que%20Spark%20Streaming%3F,latencias%20y%20el%20posible%20throughput.>

<https://blog.nfon.com/es/ventajas-comunicaciones-unificadas>