En los últimos tiempos, la demanda de procesamiento de flujos continuos de datos (data streams) se ha incrementado considerablemente. Esto se debe a que ya no es suficiente con procesar grandes volúmenes de datos. Los datos, además, deben ser procesados rápidamente permitiendo a los sistemas reaccionar ante los eventos lo antes posible. Ejemplos de sistemas que necesitan éste nivel de procesamiento son los sistemas de detección de fraude, monitoreo de recursos, comercio, etcétera.

# Big Data

El término Big Data, muy utilizado en la actualidad, hace referencia a lo que se conoce como las tres V, Volumen, Variedad y Velocidad. Con ello, se quiere indicar que un sistema Big Data no solo implica trabajar con grandes volúmenes de datos, sino que estos datos pueden ser muy variados y se deben procesar rápidamente.

# Procesamiento de Flujos de Datos (Stream Processing)

Es un sistema diseñado para analizar y actuar en tiempo real un flujo continuo de datos. En contraste a los modelos de procesamiento de datos tradicionales en los cuales los datos son primero almacenados y luego procesados y analizados, cuando se procesa un flujo de datos, los datos son procesados y analizados mientras entran en el sistema. Esto permite lo que se llama ***procesar datos en movimiento***. Esto permite conectar a los procesadores de datos a fuentes de datos externas introduciendo a los mismos al flujo de procesamiento.

Una solución de procesamiento de datos en tiempo real debe ser capaz de:

* Procesar cantidades enormes de datos permitiendo filtrado, agregación, predicción, alertas, reglas, etcétera.
* Respuesta en tiempo real a los mensajes/eventos recibdos.
* Asegurar rendimiento y escalabilidad cuando el volumen de datos crece en tamaño y/o complejidad.
* Integración fácil y rápida con la infraestructura y fuentes de datos existentes.
* Rápida implementación y puesta en producción de nuevos requisitos de procesamiento.