

Componentes clave de Spark en un escenario de marketing digital

Autor: Jonathan Vásquez

1. Caso de uso seleccionado

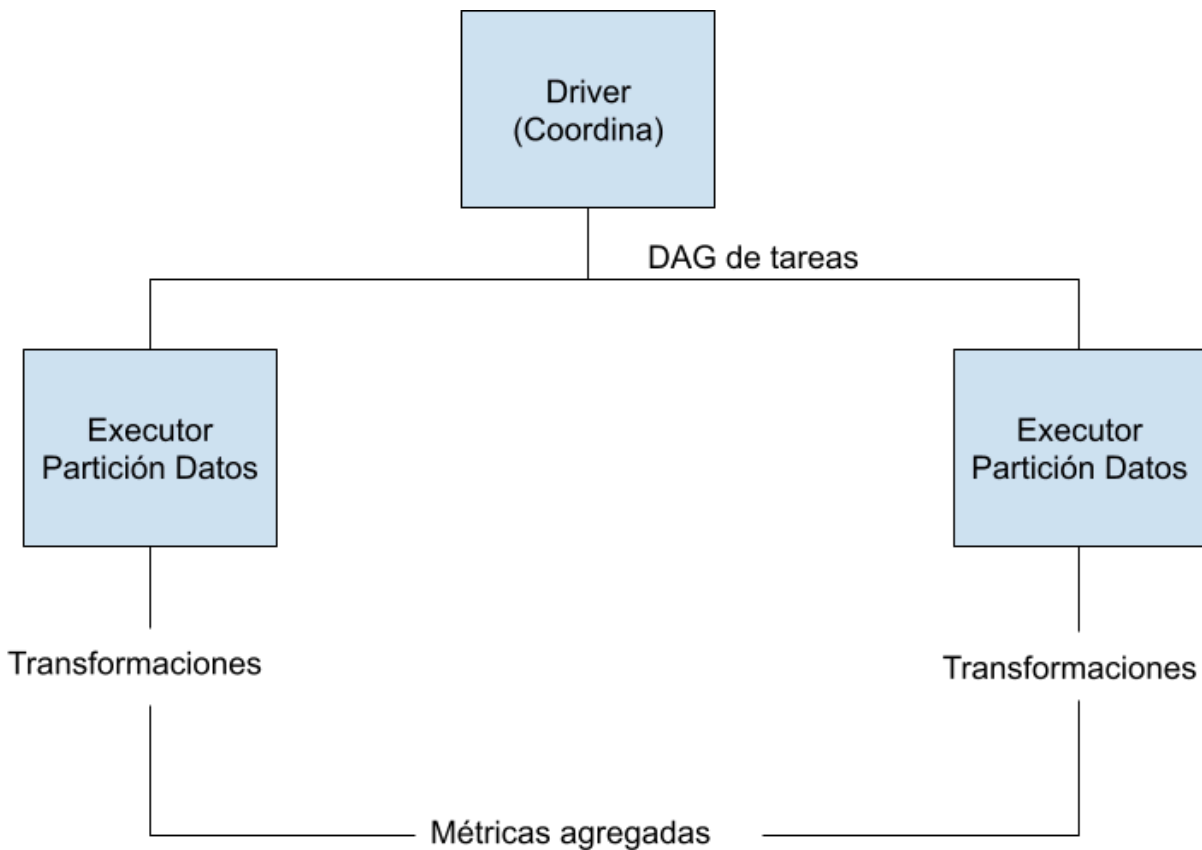
Análisis de campañas en redes sociales en tiempo real

- **Objetivo:** Monitorear y analizar en tiempo real el rendimiento de campañas publicitarias en redes sociales como Facebook e Instagram.
- **Métricas clave:** clics, impresiones, comentarios, shares, engagement por región o segmento.
- **Beneficio:** Permite ajustar estrategias publicitarias al instante y optimizar presupuesto.

2. Componentes del ecosistema Spark utilizados

Componente	Función en el caso
Driver	Coordina el trabajo de la aplicación Spark, recibe los datos de streaming, genera el DAG de tareas y envía instrucciones a los Executors.
Executors	Ejecutan las tareas enviadas por el Driver en paralelo, procesando los datos de campañas y generando métricas agregadas.
Cluster Manager	Gestiona recursos del cluster, asigna memoria y CPU a cada Executor (puede ser Spark Standalone, YARN o Kubernetes).
RDDs	Representan los datos distribuidos de las campañas (por ejemplo, logs de clics). Útil para transformaciones personalizadas.
DataFrames / Datasets	Abstracción de alto nivel sobre RDDs. Permite manipular y consultar datos estructurados de manera eficiente con Spark SQL.
Spark Streaming / Structured Streaming	Procesamiento de datos en tiempo real desde APIs de redes sociales o colas de mensajes (Kafka, Kinesis).
MLlib (opcional)	Para análisis predictivo, como identificar tendencias emergentes o predecir engagement futuro.

3. Distribución del procesamiento en el cluster



4. Justificación del uso de Spark frente a otras herramientas

- **Hadoop:** Buen manejo de grandes volúmenes, pero está más orientado a batch y no es eficiente para análisis en tiempo real.
- **Flink:** Excelente para streaming, pero Spark es más maduro, con mayor comunidad, integración con MLlib y fácil manejo de DataFrames.
- **Spark:** Permite combinar batch y streaming, procesamiento distribuido eficiente, manejo de APIs de alto nivel y escalabilidad.

Conclusión: Spark es ideal para marketing digital porque permite procesar grandes flujos de datos en tiempo real, generar métricas accionables y optimizar campañas al instante, algo fundamental para empresas como Musa Media.

