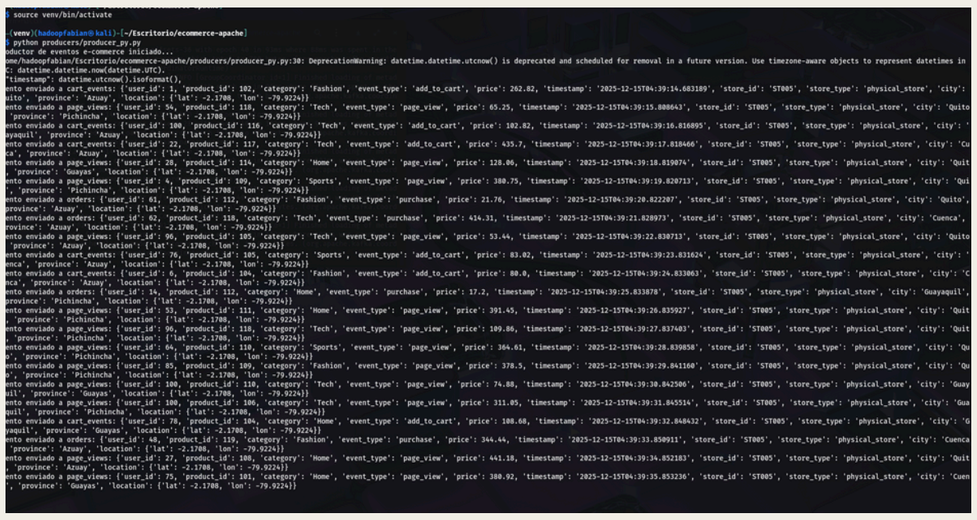
PROCESAMIENTO E-COMMERCE

1. GENERACIÓN DE EVENTOS

Un script simula usuarios navegando, agregando productos al carrito y realizando compras.
Cada acción se envía como un evento **JSON** hacia **Apache Kafka**.

DATOS GENERADOS



producer_py.py

2. INGESTA CON KAFKA

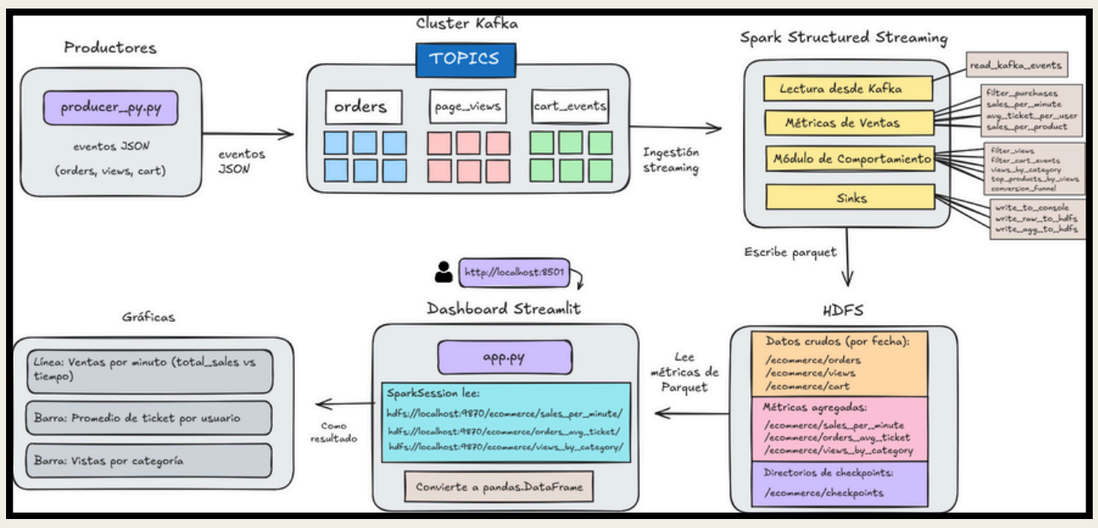
Los eventos se distribuyen en tres **topics**: **page_views, cart_events, orders** para mejor organización y flujo continuo de mensajes.

3. STRUCTURED STREAMING

Spark Structured Streaming consume los eventos desde Kafka y ejecuta:

- Filtrado por tipo de evento
- Agregaciones por ventanas de tiempo
- Cálculo de métricas clave (ventas por minuto, vistas por categoría, ticket promedio)
- Join entre streams para analizar la conversión vista → carrito → compra

ARQUITECTURA



Fabián Rodas y Paula Benalcázar

4. ALMACENAMIENTO HDFS

Tanto los datos crudos como las métricas agregadas se guardan en HDFS en formato **Parquet**, organizados por **fecha**.

/orders, /views, /cart, /sales_per_minute

5. VISUALIZACIÓN

Un **dashboard** en **Streamlit** lee los datos procesados desde HDFS y muestra métricas como:

- Ventas por minuto
- Actividad por categoría
- Productos más vistos.

6. CONCLUSIONES

- Kafka y Spark procesan eventos en tiempo real eficazmente.
- HDFS posee almacenamiento histórico organizado y accesible.
- La arquitectura demuestra alta escalabilidad y tolerancia a fallos.
- Las métricas permiten comprender mejor el comportamiento del usuario.