

Propuesta Datos Grandes

Integrantes:

- Fabián
- Fernando
- Jorge

Datos Sintéticos

Campo	Tipo	Descripción
transaction_id	string	ID único de la transacción
store_id	string	Identificador de la tienda
timestamp	datetime	Fecha y hora de la compra
product_id	string	ID del producto
category	string	Categoría (bebidas, snacks, limpieza, etc.)
price	float	Precio unitario
quantity	int	Cantidad vendida
payment_method	string	Efectivo, tarjeta, app
customer_age	int	Edad estimada del cliente

Pipelines

- Data producer -> Kafka -> Samza -> PowerBI
- Data producer -> Kafka -> HDFS -> Spark -> PowerBI/HTML

Dashboard en tiempo real

- Ventas por tienda / minuto.
- Productos más vendidos.
- Método de pago más utilizado.

Reportes batch:

- Evolución de ventas diarias.
- Promedio de ticket por cliente.
- Análisis de comportamiento de compra por edad y categoría.