

Informe Técnico – Integración de Datos en FastTrack Logistics

1. Diagnóstico del entorno actual

- CRM: gestiona clientes, necesita integración vía APIs REST.
- ERP: operaciones internas, mejor con integración batch/ETL periódica.
- Sensores IoT: datos en tiempo real de ubicación y temperatura, requieren integración streaming.
- Google Sheets: reportes manuales diarios, se recomienda integración mediante conectores o APIs.

Problema principal: datos dispersos, duplicados e inconsistentes, lo que genera decisiones poco confiables y reportes fragmentados.

2. Arquitectura propuesta

- Capa de ingesta:
 - Apache NiFi para orquestar flujos de datos (batch + streaming).
 - Kafka para la captura en tiempo real desde IoT.
- Capa de integración/transformación:
 - Talend o AWS Glue para ETL/ELT (limpieza, normalización, enriquecimiento).
- Capa de almacenamiento:
 - Data Lake (S3/HDFS) para datos crudos.
 - Data Warehouse (Snowflake, Redshift o BigQuery) para datos consolidados.
- Capa de consumo:
 - Herramientas BI (Power BI, Tableau) para dashboards unificados.

3. Tecnologías seleccionadas y justificación

- Apache NiFi: flexible para ingesta multi-formato y control de flujos.
- Kafka: soporta alta velocidad y resiliencia en datos de sensores IoT.
- AWS Glue: ETL serverless, reduce costos de infraestructura.
- Snowflake: escalabilidad y consultas analíticas rápidas.
- Power BI/Tableau: facilidad de uso y adopción empresarial.

4. Estrategia de calidad y seguridad

Calidad:

- Eliminación de duplicados y validación de reglas de negocio (ej. rangos de temperatura válidos).
- Auditoría de transformaciones y trazabilidad completa.

Seguridad:

- Autenticación con IAM y control de accesos por rol.
- Cifrado en tránsito (TLS) y en reposo (AES-256).
- Monitoreo y alertas ante accesos no autorizados.

5. Flujo de integración (ejemplo extremo a extremo)

1. Sensor IoT envía datos a Kafka en tiempo real.
2. NiFi consume los datos, aplica reglas de validación (ej. formato JSON).
3. Los datos limpios se almacenan en Data Lake (S3).
4. AWS Glue transforma y carga al Data Warehouse (Snowflake).
5. Power BI conecta al Data Warehouse y muestra un dashboard con ubicación de camiones y alertas de temperatura.

6. Beneficios esperados

- Consolidación de datos → fin de silos de información.
- Trazabilidad completa de la cadena logística.
- Decisiones más rápidas y confiables gracias a dashboards en tiempo real.
- Mejor atención al cliente al disponer de datos integrados de pedidos y entregas.
- Escalabilidad para soportar el crecimiento futuro.