### **1. Diagnóstico del Entorno Actual**

La empresa FastTrack Logistics utiliza múltiples sistemas que no están integrados, lo que ha creado un entorno de datos fragmentado. Este escenario genera ineficiencias operativas y dificulta la toma de decisiones estratégicas.

El análisis de los sistemas actuales y sus necesidades de integración es el siguiente:

| Sistema | Datos generales | Integración |
| --- | --- | --- |
| ERP | Ventas, inventario, etc. | Batch |
| CRM | Gestión clientes | Batch |
| IoT | Sensores (por ejemplo GPS camiones) | Streaming |
| Sheets | Reportes diarios/comunes | Batch |

### 2. Arquitectura Propuesta

La propuesta se basa en una arquitectura de Data Lake con capas de procesamiento diferenciadas:

Fuentes de Datos → Capa de Ingesta → Capa de Procesamiento → Capa de Almacenamiento → Capa de Consumo.

Componentes principales:

* Data Lake: Almacenamiento central unificado
* ETL/ELT Pipeline: Procesamiento y transformación de datos
* Data Warehouse: Datos estructurados para análisis
* API Gateway: Acceso unificado a los datos
* Dashboard/BI: Visualización y reportes
* 2.2 Tipos de Integración por Sistema
* CRM: Integración híbrida (batch)
* Datos maestros: sincronización diaria
* Transacciones críticas: cada 15 minutos
* ERP: Integración batch
* Extracciones nocturnas de datos transaccionales
* Sincronización de maestros cada 4 horas
* Sensores IoT: Integración streaming
* Ingesta continua de telemetría
* Procesamiento en tiempo real para alertas
* Google Sheets: Integración batch programada
* Extracción automática diaria
* API de Google Sheets para automatización

### 3. Herramientas seleccionadas

* Ingestión de datos para streaming: AWS Spark -> herramienta optimizada para datos streaming
* Ingestión de datos para Batch: AWS Glue -> Herramienta ETL que puede conectarse a DB y archivos
* Repositorio: Amazon S3 -> Permite almacenar grandes volúmenes de datos
* Visualización: Amazon QuickSight -> Herramienta de Amazon para BI

### 4. Flujo de Integración y Estrategia

1. Fuente: Por ejemplo el sensor del camión que envía paquetes o pedidos
2. Ingesta: Eld ato llega a un stream de AWS Spark
3. Almacenamiento: En repositorio de Amazon S3 Data Lake
4. ETL: Amazon Redshift para pasar la info. Del data Lake a Data Warehouse para estructurarla
5. Destino Final: Visualización de Data Warehouse a QuickSight

5. Beneficios Esperados

* Fuente única de datos: Centraliza la info. Operativa, eliminando errores, duplicados y otros
* Eficiencia Operativa: Automatiza la recolección de la data
* Toma de decisiones rápida: Los stakeholders podrán acceder a la info. Actualizada y precisa para tomar decisiones