Ejercicio N° 1: Diseño de flujo de ingesta batch para consolidación de ventas regionales

**1. Objetivo del flujo de ingesta**

El objetivo es consolidar diariamente la información de ventas proveniente de múltiples sucursales en archivos .csv, que presentan formatos heterogéneos. Se busca unificar estos datos en una estructura común para facilitar el análisis por parte del equipo de Business Intelligence. Esto responde a la necesidad de contar con datos confiables, actualizados y estandarizados para la toma de decisiones estratégicas.

**2. Estructura del flujo ETL**

**🔹 Extracción**

* **Tipo de archivos**: .csv con estructuras variables.
* **Frecuencia**: diaria, durante la noche.
* **Origen**: carpetas compartidas, FTP, o buckets S3 regionales.

**🔹 Transformación**

* Limpieza de registros incompletos o corruptos.
* Conversión de formatos de fecha, moneda y codificación.
* Estandarización de columnas (nombres, tipos de datos).
* Enriquecimiento con datos de referencia (e.g., códigos de sucursal).

**🔹 Carga**

* Destino: base de datos relacional (PostgreSQL) o Data Warehouse (Redshift, BigQuery).
* Organización por fecha y región.
* Generación de logs de carga y resumen de registros procesados

**3. Herramienta ETL recomendada**

**Apache NiFi**  
**Justificación**:

* Permite diseñar flujos visuales sin necesidad de escribir código.
* Soporta múltiples protocolos de entrada (FTP, S3, HTTP).
* Tiene componentes nativos para transformación de archivos CSV.
* Ofrece trazabilidad, manejo de errores y programación de tareas.

**4. Buenas prácticas**

* **Validación de datos**: aplicar reglas para detectar registros inválidos antes de la carga.
* **Programación fuera del horario de carga**: ejecutar el proceso en la madrugada para no afectar el rendimiento del sistema.
* **Trazabilidad**: registrar cada archivo procesado con fecha, origen y cantidad de registros.
* **Manejo de errores**: configurar rutas alternativas para archivos fallidos y alertas automáticas.

5. Esquema técnico del flujo

[1] Recepción de archivos CSV desde sucursales

     ↓

[2] Extracción con Apache NiFi (GetFile / ListFTP / ListS3)

     ↓

[3] Transformación:

     - ConvertRecord (CSV → JSON)

     - UpdateAttribute (normalización)

     - ReplaceText (formato)

     ↓

[4] Validación y filtrado de registros

     ↓

[5] Carga en base de datos (PutDatabaseRecord)

     ↓

[6] Registro de logs y errores

     ↓

[7] Notificación de éxito/falla