

Documentación: Script de Carga WOR Ventas

Este documento detalla el funcionamiento, uso y mantenimiento del script ETL (Extracción, Transformación y Carga) diseñado para procesar el archivo Excel "WOR Ventas.xlsx", extraer múltiples tipos de datos de ventas y cargarlos en una base de datos SQL Server con prevención de duplicados.

1. Funcionalidades Principales

El script automatiza la actualización de múltiples tablas relacionadas con ventas, forecasts y cuotas con un enfoque en la integridad de datos y prevención de duplicados.

Configuración y Conexión Segura

- **Variables de Entorno:** Utiliza un archivo `.env` para gestionar las credenciales de la base de datos de forma segura.
- **Compatibilidad PyInstaller:** Detecta automáticamente si se ejecuta como ejecutable (.exe) o en entorno de desarrollo.
- **Conexión Robusta:** Establece conexión segura con SQL Server usando SQLAlchemy con manejo de errores.

Selección y Procesamiento de Archivos

- **Interfaz de Usuario:** Proporciona una ventana de diálogo (Tkinter) para seleccionar el archivo "WOR Ventas.xlsx".
- **Carga Inteligente:** Utiliza openpyxl para cargar el archivo Excel completo con todas sus hojas y tablas.

Extracción Inteligente de Datos

El script utiliza **patrones regex generalizados** para identificar y extraer automáticamente tres tipos de tablas:

Patrones de Búsqueda:

- `Avancedeventa_Category_{Zona}_{Mes}` - Datos de avance de ventas por categoría
- `Proyeccion_Vendedor_{Zona}_{Mes}` - Proyecciones de vendedores
- `Forecast_{Zona}_{Mes}` - Datos de forecast por cliente

Zonas Soportadas:

Zone1, Zone2, Zone3, Zone4, Zone5, Zone6, Zone7, KamCentral, KamEast, E-Commerce, Outlet & Donation

Productos Soportados:

Ricky Joy Yogurt, Mellow Cones, Crazy Legs, Ricky Joy Gels, Jelly Fruits, Plis, SSC Roll On, Freeze Dried, 3D Gummies, SC Gel, Cotton Candy

Transformación y Enriquecimiento de Datos

Procesamiento Diferenciado por Tipo:

- **Tablas Category:** Renombrado de columnas, limpieza de valores nulos, mapeo de productos por ID.
- **Tablas Forecast:** Extracción dual (cuotas de zona + forecast por cliente), limpieza de filas totales, mapeo de clientes por ID.

Enriquecimiento Automático:

- **Traducción de Meses:** Convierte automáticamente nombres de meses de español a inglés.
- **Asignación de IDs:** Mapeo dinámico de clientes desde la base de datos.
- **Información Contextual:** Agrega zona, mes y año a todos los registros.

Prevención de Duplicados Avanzada

Estrategia Triple de Deduplicación:

1. **Cuotas por Zona:** Clave única = $(id_zona, mes, año)$ con $id_cliente = 0$
2. **Forecast por Cliente:** Clave única = $(id_cliente, id_zona, mes, año)$
3. **Cuotas por Categoría:** Clave única = $(id_producto, id_zona, mes, año)$

Proceso Inteligente:

- Consulta registros existentes en cada tabla de destino.
- Compara con datos nuevos usando claves únicas.
- Solo inserta registros genuinamente nuevos.

2. Arquitectura de Datos

Tablas de Destino en SQL Server:

Cuota_forecast

- **Propósito:** Almacena cuotas generales por zona (primera fila de tablas forecast)

- **Estructura:** `id_zone`, `id_cliente` (=0), `cuota`, `nombre_mes`, `mes`, `año`
- **Fuente:** Primera fila de tablas Forecast/Proyeccion_Vendedor

Forecast

- **Propósito:** Almacena proyecciones semanales por cliente
- **Estructura:** `id_cliente`, `id_zone`, `semana_1` a `semana_5`, `mes`, `año`, `nombre_mes`
- **Fuente:** Filas restantes de tablas Forecast/Proyeccion_Vendedor

Cuotas_Avance_Categoria

- **Propósito:** Almacena cuotas de dinero y volumen por producto
- **Estructura:** `id_producto`, `id_zone`, `cuota_dinero`, `cuota_volumen`, `mes`, `año`, `nombre_mes`
- **Fuente:** Tablas Avancedeventa_Category

Tablas de Referencia:

- **Clientes:** Catálogo de clientes con `id_cliente` y `nombre_cliente`
- **Mapeos estáticos:** Productos y zonas definidos en el código

3. Manejo y Uso del Script

Distribución

El script se distribuye como **archivo ejecutable (.exe)** para facilitar su uso por personal no técnico.

Requisitos Previos

1. Archivo .env

Debe estar en la misma carpeta que el ejecutable:

```
SERVER_NAME=tu_servidor
PORT=tu_puerto
DATABASE_NAME=tu_base_de_datos
DB_USERNAME=tu_usuario
DB_PASSWORD=tu_contraseña
```

2. Estructura de Base de Datos

- Tabla `Clientes` con columnas `id_cliente` y `nombre_cliente`
- Tablas de destino: `Forecast`, `Cuota_forecast`, `Cuotas_Avance_Categoria`

- Permisos de escritura en todas las tablas

3. Archivo Excel

- Nombre: "WOR Ventas.xlsx"
- Contiene hojas con tablas nombradas según los patrones establecidos
- Datos organizados por zona y mes

Pasos para Ejecución

1. Ejecutar el Script

Hacer doble clic en el archivo ejecutable.

2. Seleccionar Archivo

En la ventana de diálogo, buscar y seleccionar "WOR Ventas.xlsx".

3. Monitorear Progreso

La consola mostrará:

- Archivo cargado exitosamente
- Tablas encontradas y extraídas
- Traducción de meses
- Proceso de carga por tabla
- Registros insertados vs. duplicados evitados

4. Verificar Resultados

Al finalizar, revisar:

- Número total de tablas procesadas
- Registros insertados en cada tabla
- Mensajes de error (si los hay)

4. Funcionalidades Avanzadas

Implementación Dual de Cuotas

El script mantiene **dos implementaciones** para flexibilidad:

Implementación Actual (Activa):

```
python
```

```
ingest_zone_quotas_data(total_zone_quotas) # Cuotas por zona
```

Implementación Anterior (Comentada):

```
python
```

```
# ingest_cuota_forecast_data(total_Forecast) # Cuotas por cliente
```

Manejo Robusto de Errores

- Try-catch en todas las operaciones críticas
- Limpieza automática de conexiones
- Mensajes de error descriptivos
- Continuidad del proceso ante errores menores

Optimización de Rendimiento

- Inserción en lotes de 1000 registros
- Consultas optimizadas para deduplicación
- Manejo eficiente de memoria con pandas
- Disposición automática de conexiones

5. Mantenimiento y Solución de Problemas

Errores Comunes y Soluciones

"No se seleccionó ningún archivo"

- **Causa:** Usuario canceló selección de archivo
- **Solución:** Reiniciar script y seleccionar archivo válido

"Error al cargar el archivo de Excel"

- **Causa:** Archivo corrupto o formato incorrecto
- **Solución:** Verificar que el archivo es Excel válido (.xlsx)

"No se encontró columna 'TOTAL'"

- **Causa:** Estructura de tabla forecast incorrecta

- **Solución:** Verificar formato de tablas en Excel

"Error de conexión a base de datos"

- **Causa:** Credenciales incorrectas o servidor inaccesible
- **Solución:** Verificar archivo .env y conectividad

Logs y Monitoreo

El script proporciona logging detallado en consola:

- **INFO:** Progreso normal del proceso
- **WARNING:** Situaciones no críticas (ej. clientes no encontrados)
- **ERROR:** Problemas que impiden la carga de datos

6. Versión Actual y Roadmap

Versión Actual: 2.0

Características Principales:

- **Distribuido como ejecutable (.exe)**
- **Extracción automática** de múltiples tipos de tablas
- **Prevención robusta de duplicados** con múltiples estrategias
- **Carga incremental** en tres tablas distintas
- **Implementación dual** para flexibilidad operativa

Mejoras vs. Versión 1.0:

- Soporte para archivos Excel (vs. solo CSV)
- Extracción automática de múltiples tablas (vs. una tabla)
- Mapeo dinámico de clientes (vs. estático)
- Traducción automática de meses
- Implementación flexible con rollback capability

Roadmap Futuro

Próximas Mejoras:

- **Logging a Archivo:** Sistema de logs persistente para auditoría
- **Dashboard de Monitoreo:** Interfaz web para visualizar cargas históricas

- **Validación de Datos:** Reglas de negocio más estrictas pre-carga
- **Notificaciones:** Alertas automáticas por email sobre el estado de carga

Mejoras de Usuario:

- **GUI Avanzada:** Interfaz gráfica con barras de progreso
- **Configuración Visual:** Editor gráfico para mapeos y configuraciones
- **Reportes Automáticos:** Generación de reportes post-carga

Mejoras Técnicas:

- **Paralelización:** Procesamiento simultáneo de múltiples tablas
- **Cache Inteligente:** Reutilización de mapeos entre ejecuciones
- **Versionado de Datos:** Control de versiones para rollbacks

7. Consideraciones de Seguridad

Protección de Credenciales

- Uso obligatorio de archivo .env
- No hardcoding de credenciales
- Compatibilidad con sistemas de secrets management

Integridad de Datos

- Transacciones para operaciones críticas
- Validación de tipos de datos
- Prevención de SQL injection via SQLAlchemy

Auditoría

- Logging completo de operaciones
- Rastreabilidad de cambios por ejecución
- Identificación de registros insertados vs. duplicados