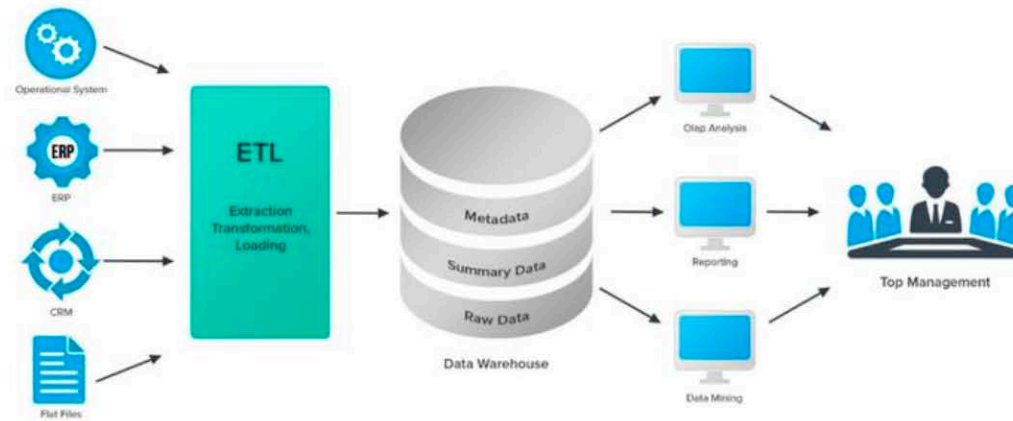


# FACULTAD DE INGENIERÍA ESCUELA DE COMPUTACIÓN

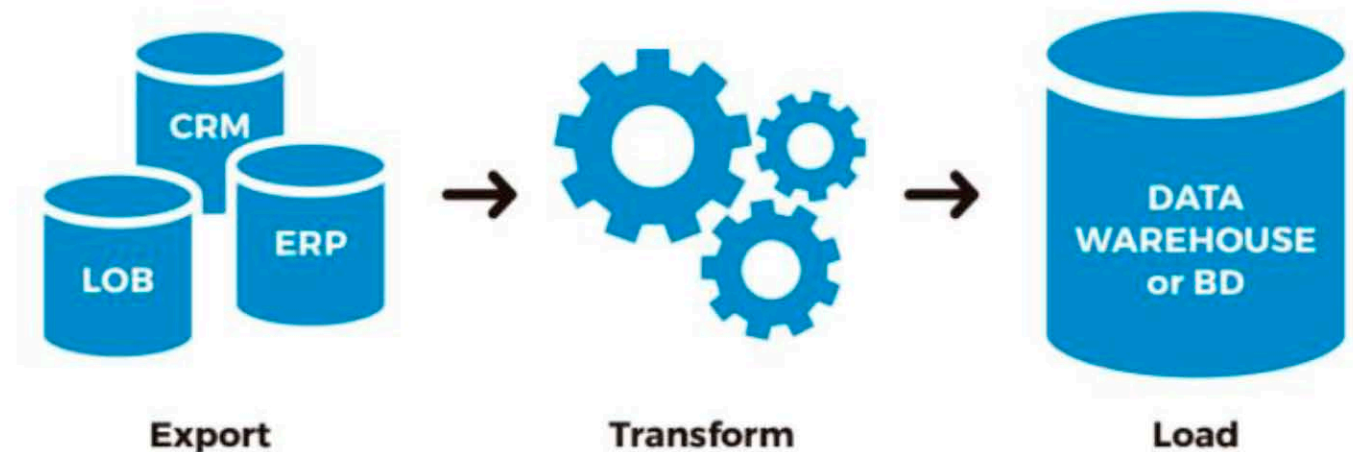


## Datawarehouse y Minería de Datos

### Guía #4: Herramientas ETL - Manejo de Excepciones y Errores

## Recordando que es ETL

Un proceso ETL es un sistema que permite la movilización de información alojada en diferentes fuentes de datos (ERP, CRM, Excel, archivos de texto plano, bases de datos de sistemas transaccionales, etc.), con el fin de integrarlos en un solo lugar. Este almacenamiento de datos puede ser en una base de datos nueva o incluso en un Datawarehouse.



# Las fases del proceso ETL

El proceso ETL se compone de tres fases:

## ➤ Proceso de extracción de los datos

En esta etapa tenemos claridad de donde provienen nuestros datos. Como lo hemos comentado en clases anteriores la fuente de datos pueden ser desde sistemas CRM, ERP, DB, archivos de texto plano, hojas Excel entre otros. Es importante aclarar que en esta etapa se debe de analizar los datos extraídos y verificar si cumplen con la estructura que nosotros hemos definido para nuestro destino o almacén de datos. Uno de los objetivos de la extracción es convertir los datos a un solo formato y prepararlos para su transformación.

## ➤ Proceso de transformación de los datos

Aquí es donde se hacen las siguientes tareas: validación de las reglas de negocio, validaciones técnicas (duplicados, integridad de los datos, verificar valores nulos, etc.), obtención nuevos valores calculados, unir los datos de múltiples fuentes, normalización de los datos, ordenar, realizar cruce de datos, filtrar entre otras.

## ➤ Proceso de carga de datos

Los datos transformados se comienzan a cargar en el destino definido (BD o Datawarehouse). En el caso del datawarehouse los datos son registrados para después auditarse, es decir que en este punto se puede tener datos históricos y actuales.

# Controles a utilizar en la práctica

## ➤ **Flat File Source (origen de archivo plano)**

El origen de archivo plano lee los datos por medio de un archivo de texto. El archivo de texto debe tener formato delimitado de ancho fijo y mixto.

- ✓ El formato delimitado utiliza columnas y delimitadores de filas para definir columnas y filas.
- ✓ El formato de ancho fijo utiliza el ancho para definir columnas y filas. Este formato también incluye un carácter para rellenar los campos hasta alcanzar el ancho máximo.
- ✓ El formato derecho irregular utiliza el ancho para definir todas las columnas, excepto la última, que se delimita mediante el delimitador de filas.





# Controles a utilizar en la práctica

## ➤ ***Flat File Destination (Destino de archivo plano)***

El destino de archivo plano escribe datos en un archivo de texto. El archivo de texto puede tener formato delimitado, de ancho fijo, de ancho fijo con delimitador de filas o desigual a la derecha.

Se puede configurar el destino de archivo plano de las maneras siguientes:

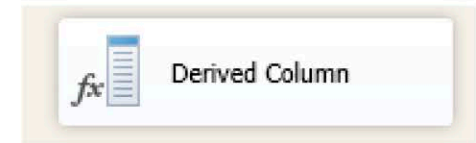
- ✓ Proporcionar un bloque de texto que se inserta en el archivo antes de escribir cualquier dato. El texto puede proporcionar información tal como encabezados de columna.
- ✓ Especificar si se deben sobrescribir datos en un archivo de destino que tenga el mismo nombre.



# Controles a utilizar en la práctica

## ➤ ***Derived Column Transformation (Transformación de columna derivada)***

La transformación Columna derivada crea nuevos valores de columna aplicando expresiones a las columnas de entrada de la transformación. Una expresión puede contener cualquier combinación variables, funciones, operadores y columnas de la entrada de transformación. El resultado puede agregarse como una nueva columna o insertarse en una columna existente como un valor de reemplazo. La transformación Columna derivada puede definir varias columnas derivadas, y cualquier variable o columna de entrada puede aparecer en varias expresiones.

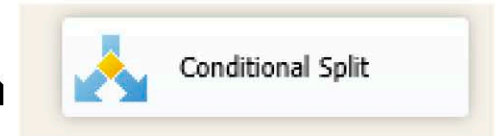


# Controles a utilizar en la práctica

## ➤ ***Conditional Split (Division condicional, transformación)***

La división condicional puede dirigir filas de datos a salidas diferentes en función del contenido de los datos. La implementación de la transformación división condicional es similar a una estructura de decisión CASE en los lenguaje de programación.

Evalúa expresiones y, en función de los resultados, dirige la fila de datos a la salida especificada. Esta transformación también proporciona una salida predeterminada, por lo que si una fila no coincide con ninguna expresión, se dirige a la salida predeterminada.



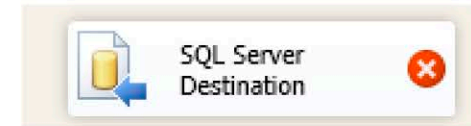
# Controles a utilizar en la práctica

## ➤ **SQL Server Destination (SQL Server, destino)**

El SQL Server destino, se conecta a una base de datos local de SQL Server y realiza una carga masiva de datos en tablas y vistas de SQL Server.

***No se puede usar el controlador SQL Server destino en paquetes con acceso a una base de datos de SQL Server en un servidor remoto.*** En su lugar, los paquetes deben utilizar el destino de OLE DB.

Los usuarios que ejecutan paquetes que incluyen el control SQL Server destino, requieren de permisos globales. Para otorgar este permiso a los usuarios, se debe utilizar la Directiva de seguridad local del Sistema Operativo.





# Comencemos con nuestra práctica

---

