

Complementaria Guía 4

Autor: BA1819 ,GM181938

2020

Universidad Don Bosco



Datawarehouse y Minería de Datos

Complementaria Guía 4

Integrantes:

Bonilla Avilés, David Alejandro BA1819

Grande Menjivar, Willian Adonis GM181938

Catedrático: Ing. Alexander Siguenza

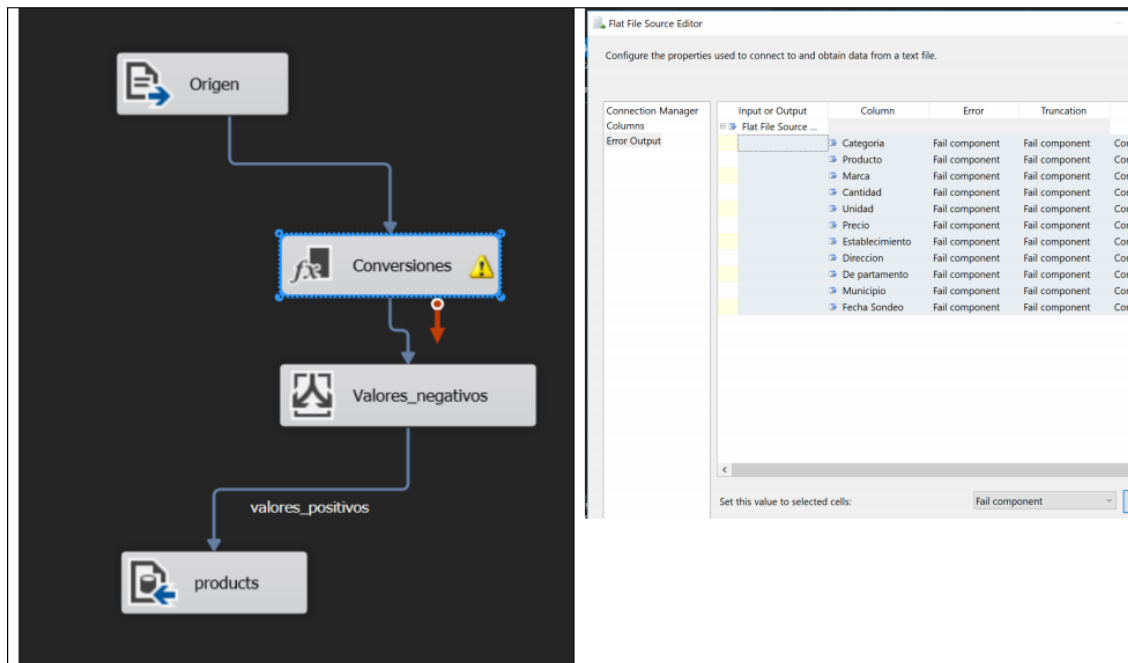
Soyapango, 19 de septiembre del 2020

Contenido

Ejercicio Complementario 1	4
• ¿Se ejecuta correctamente el ejercicio?	4
• ¿Qué sucede si cambia el error output por Ignore failure?.....	4
• ¿Qué sucede si configura el Error de Conversion para que realice Ignore failure?	4
Ejercicio Complementario 2	5

Ejercicio Complementario 1

Cree un nuevo ETL como se muestra en el diagrama indicando que el Flat File Source “Origen” para en el Error Output no gestione errores como se ve a continuación:



Ejecute el paquete y responda:

- **¿Se ejecuta correctamente el ejercicio?**

No se ejecuta correctamente, se puede decir que esto sucede por el fail component

- **¿Qué sucede si cambia el error output por Ignore failure?**

Funciona nuevamente, ya que omite cuando sucede un error y continúa corriendo el etl. Esto significa que ese registro no es tratado.

- **¿Qué sucede si configura el Error de Conversion para que realice Ignore failure?**

Significa que omite el error que se está dando, por ello el registro ya no es devuelto aun archivo donde se guardan los registros que dieron algún tipo de error, en el caso que la conversión arrojará una percepción por supuesto.

Ejercicio Complementario 2

Al utilizar el For Each Loop Container el flat file error se vuelve muy extenso, lo cual significa que perdemos muchos registros. ¿Puede recuperarse esa DATA?

Si, se puede optar por varios caminos para poder recuperar estos registros, lo primero y mas fácil seria ver los registros que dan error, analizar porqué dan error y modificar el ETL de tal manera que estos datos sean los menos posibles, un ejemplo serio:

Si arroja un error porque una conversión es a un numero entero positivo, se pudiera manipular la información para obligar a los datos a ser convertidos a números positivos, esto claro si fuese un error en la toma de dichos datos. Otro punto serio cambiar el tipo de dato para que aceptase positivos también.

Otra solución seria obtener los archivos con datos que dieron error, analizarlos y hacer un ETL exclusivo para ellos, puede ser que necesiten una conversión especial o algo por el estilo, inclusive puede ser datos faltantes en un supuesto campo de código postal, por lo que se pueden sustituir los registros que tienen un guion en el código postal por el código postal 0000.

En nuestro caso en especifico hay bastantes registros en los archivos donde se encuentras los datos de origen que en la casilla precio tienen un registro como: "Sin Precio"

✎ xaa: Bloc de notes

Archivo Edición Formato Ver Ayuda

Misceláneos,Ábaco,Sin Marca,1,Unidades,\$1.75,"Zona Centro, calle Ruben Dario, Hula Hula","Zona Centro, calle Ruben Dario, Hula Hula",San Salvador,San Salvador,5/1/2018
 Misceláneos,Ábaco,Sin Marca,1,Unidades,\$1.75,"Zona Centro, calle Ruben Dario, Hula Hula","Zona Centro, calle Ruben Dario, Hula Hula",San Salvador,San Salvador,5/1/2018
 Misceláneos,Ábaco,Sin Marca,1,Unidades,\$1.75,"Zona Centro, calle Ruben Dario, Hula Hula","Zona Centro, calle Ruben Dario, Hula Hula",San Salvador,San Salvador,5/1/2018
 Misceláneos,Ábaco,Sin Marca,1,Unidades,\$1.75,"Zona Centro, calle Ruben Dario, Hula Hula","Zona Centro, calle Ruben Dario, Hula Hula",San Salvador,San Salvador,5/1/2018
 Misceláneos,Ábaco,Sin Marca,1,Unidades,\$1.50,"Zona Centro, calle Ruben Dario, Hula Hula","Zona Centro, calle Ruben Dario, Hula Hula",San Salvador,San Salvador,5/1/2018
 Misceláneos,Ábaco,Sin Marca,1,Unidades,sinprecio,"Zona Centro, calle Ruben Dario, Hula Hula","Zona Centro, calle Ruben Dario, Hula Hula",San Salvador,San Salvador,5/1/2018
 Misceláneos,Bárra de silicon delgada (unidad),Sin Marca,1,Unidades,\$0.20,Almacen y Librería SÓLO,"Calle Rubén Darío #418, San Salvador",San Salvador,San Salvador,5/1/2018
 Misceláneos Bárra de silicon delgada (unidad) Sin Marca 1 Unidades \$0.15 "Bazar la Costura Casa Matriz" "2a Calle Delante No.4 Chalchihuala" Santa Ana Chalchihuala 5/1/2018

Por lo que se puede modificar el ETL, para en primer lugar ubicar el campo precio como algún tipo de String en el flat file source, después sustituyendo los "sin precio" como 0.00 luego hacer la conversión al tipo Currency[DT_CY].