Reporte de Inconvenientes

Descripción del error:
Estamos utilizando una versión antigua de Pentaho que causa incompatibilidades con otros componentes del sistema y genera problemas con los conectores de las bases de datos.

- **Contexto**:
  El problema fue identificado al intentar integrar Pentaho con las bases de datos, y se determinó que la versión actual no es compatible con las versiones más recientes de MySQL y PostgreSQL. También se encontró que los conectores **.jar** de las versiones antiguas no están disponibles en muchos casos, al igual que el OS de base **CentOS 7**, lo que impide una conexión estable entre Pentaho y las bases de datos.

- **Acciones realizadas**:
  Se realizaron pruebas de compatibilidad entre diferentes versiones de Ubuntu, Pentaho, MySQL y PostgreSQL.
  Se identificó que las versiones más recientes (al 2024) de los sistemas de bases de datos y Pentaho (10.2) ofrecen un mejor soporte.
  Durante estas pruebas, se detectó que los **drivers .jar** necesarios para conectar Pentaho con MySQL y PostgreSQL están disponibles para estas versiones.

- **Hipótesis del problema**:
  El problema se debe a que la versión antigua de Pentaho y las bases de datos no pueden ser conectadas de manera adecuada, ya que los **.jar** de los conectores para MySQL y PostgreSQL no están disponibles para ciertas versiones anteriores, causando fallos en las conexiones.

- **Solución implementada**:
  Actualizar Pentaho a la versión 10.2 y usar versiones LTS tanto de Ubuntu como de las bases de datos:

  - **Ubuntu 24.04 LTS**
  - **Pentaho 10.2**
  - **MySQL 8.0.9**
  - **PostgreSQL 16.4**

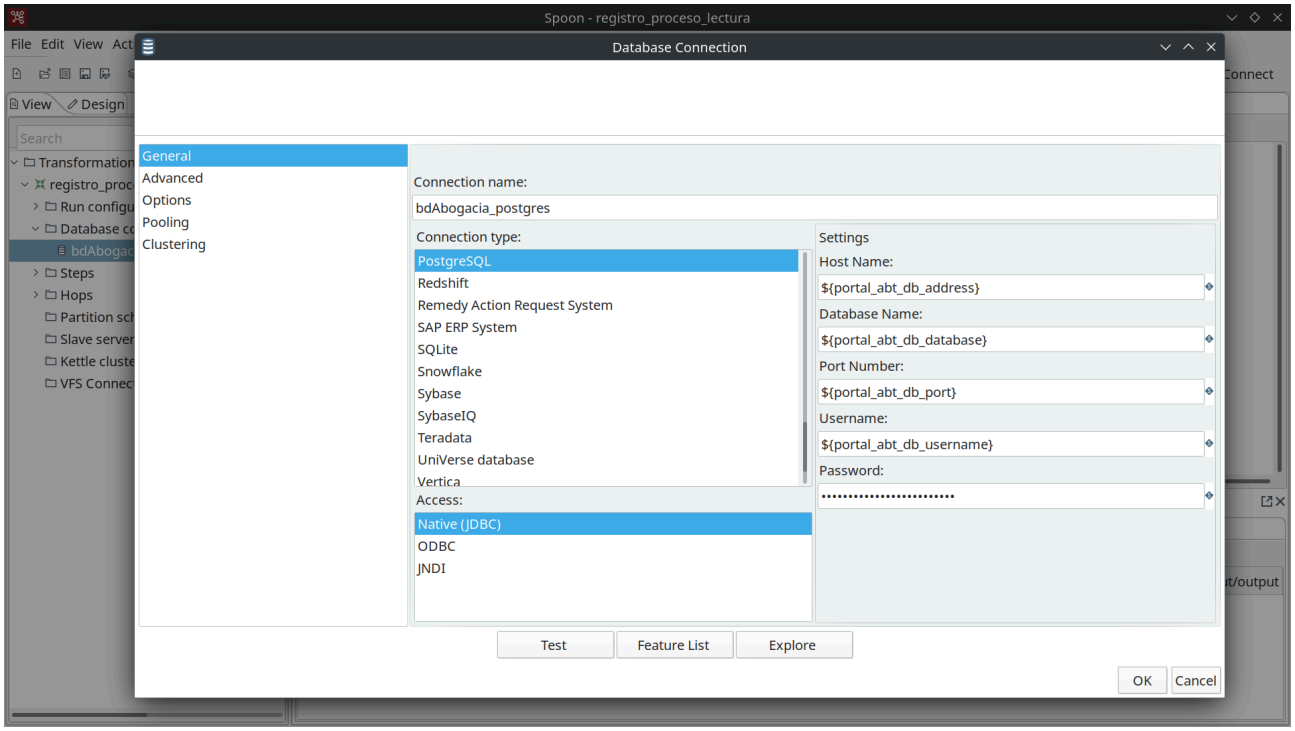
Estos cambios ofrecen las siguientes ventajas:

- Los conectores **.jar** para MySQL y PostgreSQL están disponibles y compatibles con las versiones más recientes de las bases de datos.

- Las versiones LTS (Long Term Support) garantizan actualizaciones de seguridad regulares y la disponibilidad de soluciones a largo plazo, lo que es vital para mantener la seguridad y estabilidad del sistema hasta 2029.
- **Resultados**:
  La actualización de Pentaho y los demás componentes resultó en una mayor estabilidad del sistema y en la disponibilidad de los conectores **.jar** para las bases de datos. Ahora, las conexiones se establecen sin problemas, y el sistema es compatible con las últimas versiones de los sistemas operativos y bases de datos.
- **Próximos pasos (si es necesario)**:
  Si surgen más problemas con la compatibilidad de los conectores o las bases de datos, se evaluará el uso de otros gestores de base de datos o un middleware (VMs) que facilite la compatibilidad entre los diferentes componentes.
- **Modificaciones:** contenido de la carpeta data-integration/lib con .jar necesarios

Descripción del error:
Error de autenticación de contraseña fallida al intentar conectar con las bases de datos **PostgreSQL** y **MySQL**.

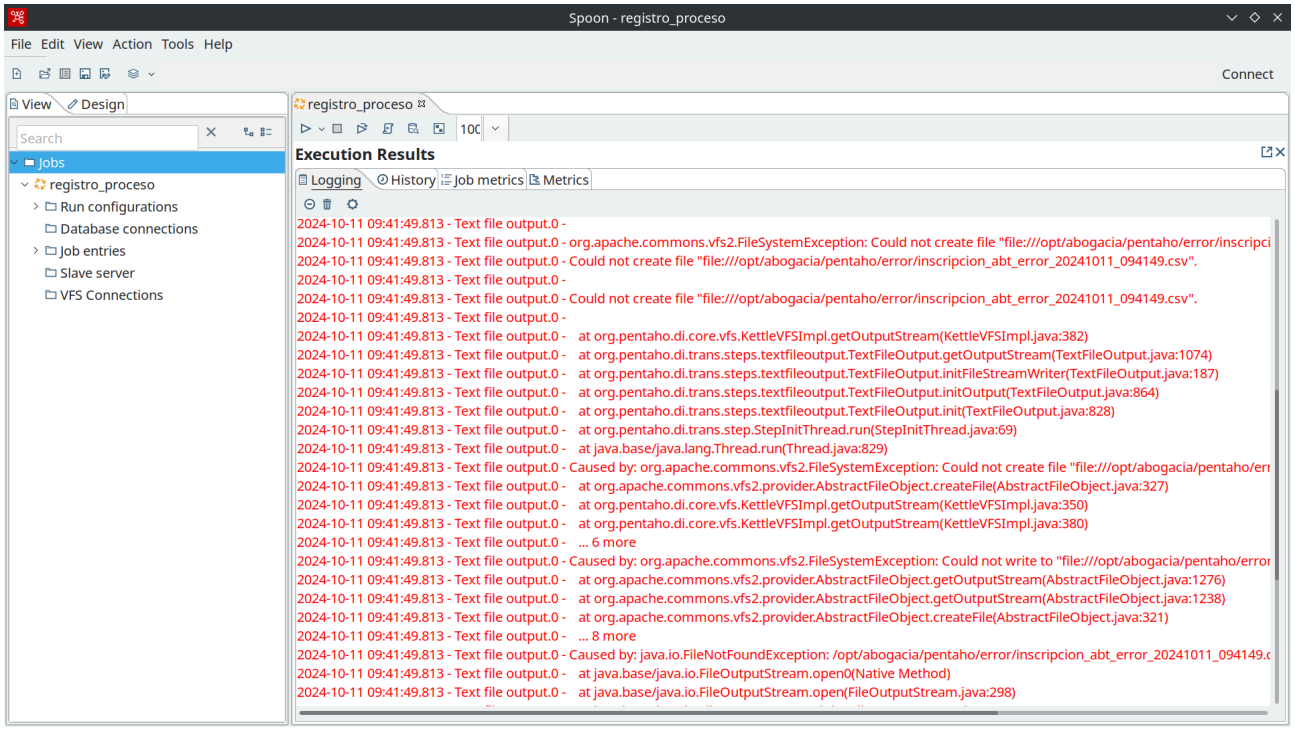
- **Contexto**:
  El problema se presentó al intentar ejecutar las transformaciones que requieren acceso a las bases de datos. Pentaho mostraba un error de autenticación, indicando que las credenciales no eran aceptadas por **PostgreSQL** y **MySQL**.



- **Acciones realizadas**:
  Se revisaron las configuraciones de conexión en el administrador de conexiones de Pentaho.
  Se detectó que las variables de ejecución utilizadas para las credenciales de la base de datos que comenzaban con un **$** lo que ocasiona problemas al resolverlas correctamente durante la ejecución.
  Para solucionar este problema, se decidió reemplazar estas variables con los valores reales de las credenciales directamente en la configuración de conexión.
  Se consideró otra opción: modificar el archivo **kettle.properties** para establecer las variables de entorno correctas, lo cual también permitiría que las credenciales se carguen adecuadamente.
- **Hipótesis del problema**:
  El error se debía a que Pentaho no estaba resolviendo correctamente las variables de ejecución para las credenciales en la configuración de conexión, lo que provocaba una autenticación fallida al intentar acceder a las bases de datos.
- **Solución implementada**:
  Se reemplazaron las variables que comienzan con **$** en la configuración de conexión por las credenciales reales de la base de datos.
  Otra opción que se evaluó y puede implementarse en futuras configuraciones es modificar el archivo **kettle.properties** para establecer correctamente las variables de entorno, garantizando que estas sean interpretadas correctamente durante la ejecución.
- **Resultados**:
  Después de realizar los cambios en las credenciales dentro de la configuración de conexión, la autenticación se realizó correctamente, permitiendo que Pentaho accediera a las bases de datos sin errores. También se estableció un plan para revisar la configuración de **kettle.properties** en futuros despliegues.
- **Modificaciones:** .bash de ejecución modificado, modificados los credenciales

Descripción del error:
Error relacionado con la creación de archivos durante la ejecución del ETL, lo que causaba fallos en el proceso debido a problemas de permisos.

- **Contexto**:
  Durante la ejecución del proceso ETL, se requiere la creación de archivos en el sistema. Sin embargo, Pentaho no pudo crear dichos archivos debido a que el script **spoon.sh** se estaba ejecutando sin los permisos necesarios, lo que resultó en un error de permisos y la interrupción de la ejecución del proceso.



- **Acciones realizadas**:
  Se revisaron los logs de error para determinar que el problema estaba relacionado con los permisos de escritura de archivos en el sistema.
  Se comprobó que el script **spoon.sh** que inicia Pentaho se estaba ejecutando sin privilegios de **sudo**, lo que impedía la creación de archivos.
  Se decidió ejecutar el script **./spoon.sh** con privilegios de **sudo** para evitar futuros problemas de permisos durante la creación de archivos en el sistema.

- **Hipótesis del problema**:
  El problema se origina porque Pentaho, al intentar crear archivos durante la ejecución del proceso ETL, no tenía los permisos suficientes en el sistema para hacerlo. Esto se debía a que el script **spoon.sh** se ejecutaba sin privilegios de **sudo**.

- **Solución implementada**:
  Se resolvió ejecutar el script **./spoon.sh** con privilegios de **sudo**. De esta forma, Pentaho tendrá los permisos necesarios para crear los archivos requeridos durante la ejecución del ETL sin problemas de permisos.

- **Resultados**:
  Ejecutar el **spoon.sh** con **sudo** resolvió el problema de permisos, permitiendo que el proceso ETL se ejecute correctamente y que los archivos necesarios se creen sin errores.

- **Modificaciones:** .bash de ejecución modificado, ejecución con sudo

Descripción del error:
Error al cargar trabajos en Pentaho debido a problemas con la interpretación de los signos **<** y **>** dentro de las etiquetas **<sql>** en los archivos **.ktr** (que están en formato XML).

- **Contexto**:
  Los trabajos en Pentaho se cargan como archivos **.ktr**, que están en formato **XML**. En estos archivos, hay etiquetas **<sql>** que contienen consultas SQL. El problema surge cuando las consultas dentro de estas etiquetas incluyen los signos **<** o **>**, los cuales son interpretados por Pentaho como un cierre de etiqueta XML en lugar de parte de la consulta SQL. Esto provoca que el trabajo no se cargue correctamente y genere tiempos de ejecución extremadamente cortos (indicando que no se está ejecutando nada).

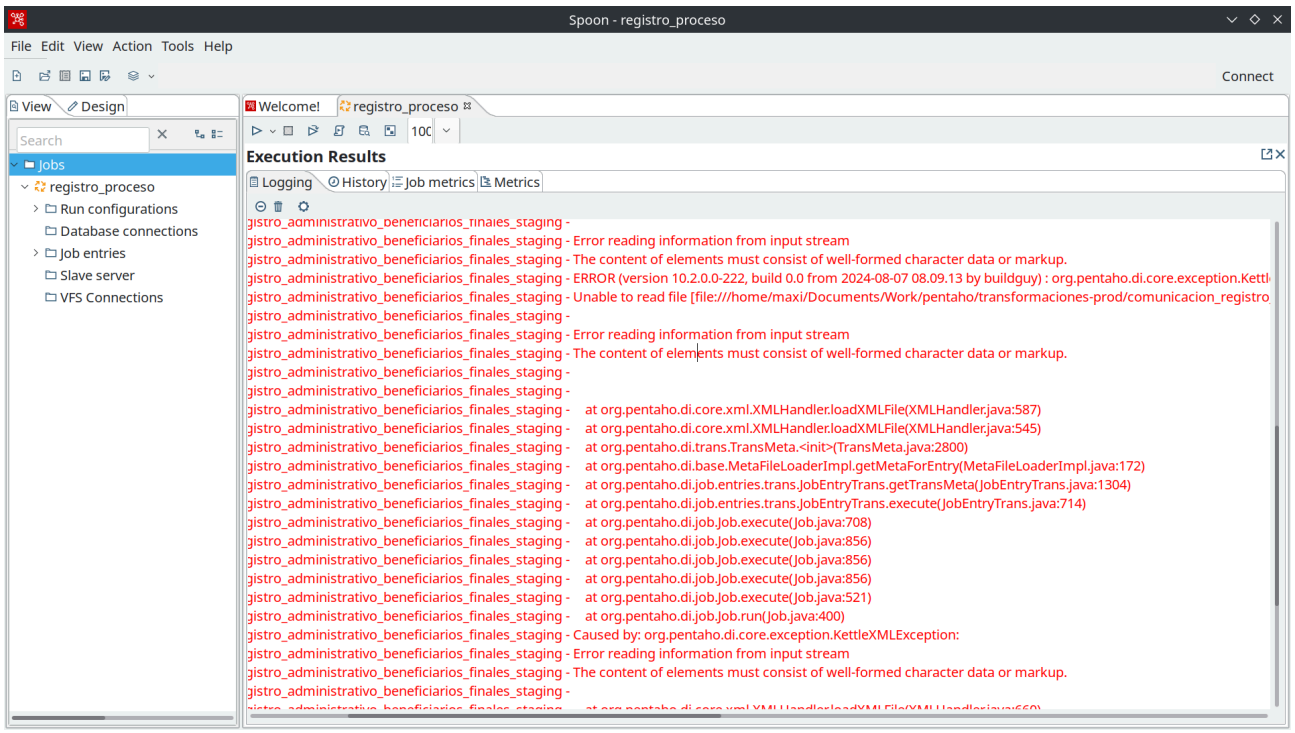


Tabla de caracteres escapados:

Carácter Original	Carácter Escapado en XML
<	<
>	>
&	&
"	"
'	'

- **Acciones realizadas**:
  Se revisaron los archivos **.ktr** y se identificó que el problema estaba relacionado

con las consultas SQL dentro de las etiquetas **<sql>** que contenían signos **<** y **>**. Pentaho interpretaba estos signos como etiquetas de cierre de XML, lo que causaba que los archivos **.ktr** estuvieran mal formateados y no se pudieran ejecutar correctamente.
Se utilizó un **validador XML** para confirmar que los archivos **.ktr** no estaban en el formato correcto debido a la inclusión incorrecta de los caracteres **<** y **>** en las consultas SQL.

- **Hipótesis del problema**:
  El problema surgía porque Pentaho no podía interpretar correctamente las consultas SQL dentro de las etiquetas **<sql>** debido a que los signos **<** y **>** se interpretaban como parte de la estructura XML, lo que generaba archivos mal formados y fallos en la carga del trabajo.

- **Solución implementada**:
  La solución fue reemplazar los signos **<** y **>** dentro de las consultas SQL por sus **caracteres escapados** en XML:

  - El signo **<** se reemplazó por <.
  - El signo **>** se reemplazó por >.

También se recomendó utilizar un **validador XML** para verificar que los archivos **.ktr** estén en el formato correcto antes de cargarlos en Pentaho.

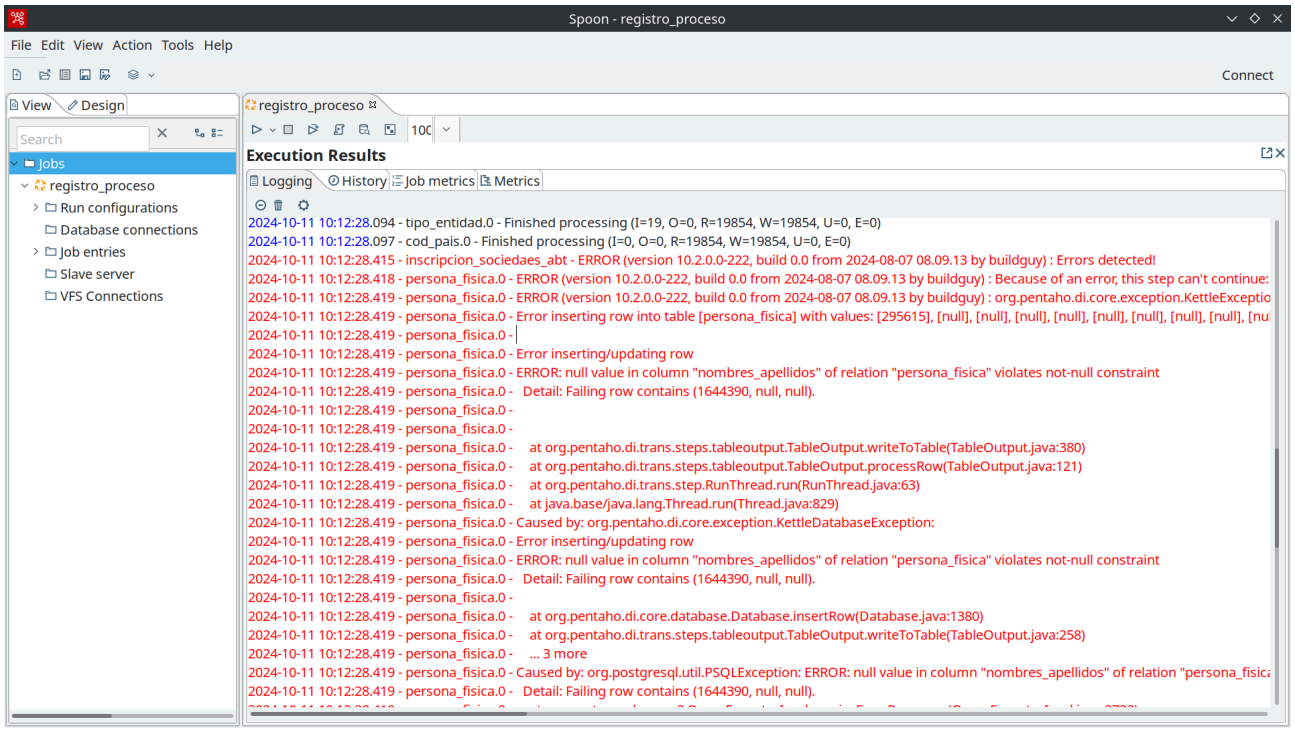
- **Resultados**:
  Después de reemplazar los signos **<** y **>** por los caracteres escapados correspondientes, los archivos **.ktr** se cargaron correctamente en Pentaho, y los trabajos se ejecutaron sin problemas. Esto también soluciona los tiempos de ejecución extremadamente cortos, lo que indicaba que el trabajo ahora se estaba ejecutando correctamente.

- **Próximos pasos (si es necesario)**:
  Como prevención futura, se recomienda revisar siempre las consultas SQL dentro de los archivos **.ktr** y asegurarse de utilizar los caracteres escapados cuando sea necesario. Además, se sugiere usar un validador XML como parte del proceso de desarrollo para confirmar que los archivos **.ktr** están en el formato correcto antes de ser ejecutados en Pentaho.

- **Modificaciones:** jobs registro_proceso_lectura, comunicacion_registro_adminitrativo_beneficiarios_finales_staging, comunicacion_registro_administrativo_beneficiarios_finales_finales

Descripción del error:
Error al insertar filas en la tabla **persona_fisica** en la base de datos PostgreSQL, específicamente al intentar insertar valores nulos en la columna **nombres_appellidos**.

- **Contexto**:
  El problema surge durante la etapa de carga de datos en la tabla **persona_fisica** de la base de datos PostgreSQL. Algunas filas generan un error debido a la inserción de valores nulos en la columna **nombres_appellidos**, lo que no es permitido por las restricciones de la tabla. Este error ocurre de manera intermitente y no afecta a todas las filas.



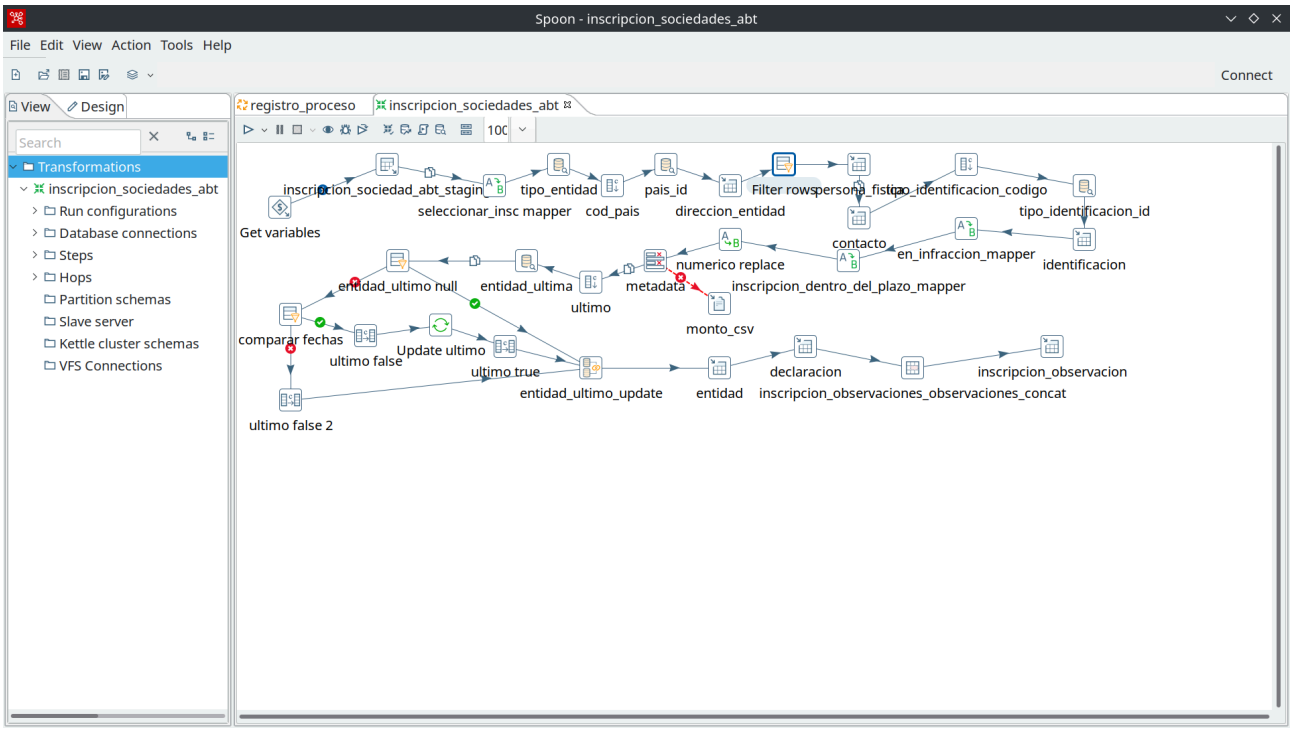
- **Acciones realizadas**:
  Se identificó que el error ocurre cuando algunas filas intentan insertar valores nulos en la columna **nombres_appellidos**.
  Como solución temporal, se forzó la ejecución del siguiente paso en el flujo ETL, a pesar de que algunas filas retornaban errores. Esto permitió que el proceso continúe sin interrumpir completamente el pipeline.
  Se identificó que el problema está relacionado con cómo el pipeline ETL está manejando los valores de la columna **nombres_appellidos**, y se sospecha que la transformación de esta columna no está cubriendo correctamente ciertos casos.

- **Hipótesis del problema**:
  - **Transformación incorrecta**: Es posible que el pipeline ETL no esté cubriendo ciertos casos donde la columna **nombres_appellidos** debería recibir un valor, y en su lugar se está intentando insertar un valor nulo.
  - **Datos de origen incompletos**: Otra posible causa es que los datos de origen que alimentan esta columna contienen valores nulos o vacíos que no están siendo manejados adecuadamente durante la transformación.

- **Problemas de mapeo**: Podría haber un error en el mapeo de los datos entre las fuentes de datos y la tabla de destino en PostgreSQL, lo que lleva a la inserción de valores incorrectos o nulos.
- **Reglas de negocio no implementadas**: Falta de validación o reglas de negocio que deberían llenar o corregir los valores de esta columna antes de la inserción.

- **Solución implementada** :
  Se forzó el uso del siguiente paso en el pipeline ETL, filtrando las filas que generaban el error de valores nulos en **nombres_appellidos** en desde la columna **reponsable_de_la_solicitud.**

- **Resultados**:
  La solución ha permitido que el pipeline ETL continúe ejecutándose sin interrupciones, pero algunas filas siguen generando errores debido a los valores nulos en la columna **nombres_appellidos** son menos de 10 filas. Se necesita una solución definitiva para evitar estos errores y garantizar que todas las filas se inserten correctamente.



- **Próximos pasos**:
  Considerar otras posibles causas, como problemas en los datos de origen, para asegurar que los datos sean consistentes y correctos antes de ser cargados en la base de datos.

- **Modificaciones:** proceso ‘Filter rows’ agregado después de direccion_entidad, filtrando así las menos de 10 columnas que provocaba el error

GEORGE TEMPERLEY OVALLE

Firmado digitalmente por GEORGE TEMPERLEY OVALLE
Fecha: 2024.11.04 16:34:16 -06'00'