GRUPO HORUS

Optimización y Debugging del ETL ABT

Un enfoque integral para la resolución de los problemas

Origen de problema

Origen de problema (II)

```
2024/09/10 13:50:06 - bdAbogacia_postgres - Connection to database closed!
2024/09/10 13:50:06 - Update ultimo.0 - Procesamiento finalizado (I=40, O=0, R=40, W=40, U=40, E=0)
2024/09/10 13:50:06 - inscripcion_sociedaes_abt - searching for annotations
2024/09/10 13:50:06 - inscripcion_sociedaes_abt - no annotations found
2024/09/10 13:50:06 - Spoon - La transformaci►n ha finalizado!!
2024/09/10 13:50:06 - inscripcion_sociedaes_abt - ERROR (version 9.0.0.0-423, build 9.0.0.0-423 from 2020-01-31 04.53.04 by buildguy) : Errores detectados!
2024/09/10 13:50:06 - inscripcion_sociedaes_abt - ERROR (version 9.0.0.0-423, build 9.0.0.0-423 from 2020-01-31 04.53.04 by buildguy) : Errores detectados!
2024/09/10 13:50:06 - inscripcion_sociedaes_abt - ERROR (version 9.0.0.0-423, build 9.0.0.0-423 from 2020-01-31 04.53.04 by buildguy) : Errores detectados!
2024/09/10 13:50:06 - registro_proceso - Finished job entry [inscripcion_sociedades_abt_final] (result=[false])
2024/09/10 13:50:06 - registro_proceso - Finished job entry [comunicacion_registro_administrativo_personas_estructuras_juridicas_staging] (result=[false])
2024/09/10 13:50:06 - registro_proceso - Finished job entry [comunicacion_registro_administrativo_beneficiarios_finales_staging] (result=[false])
2024/09/10 13:50:06 - registro_proceso - Finished job entry [inscripcion_sociedades_abt_staging] (result=[false])
```

Tabla de contenidos

Versión

Consideraciones para la actualización del software

Variables de entorno

Variables de ejecución con inicio en $ causando errores

Permisos

El ETL intenta crear archivos durante la ejecución

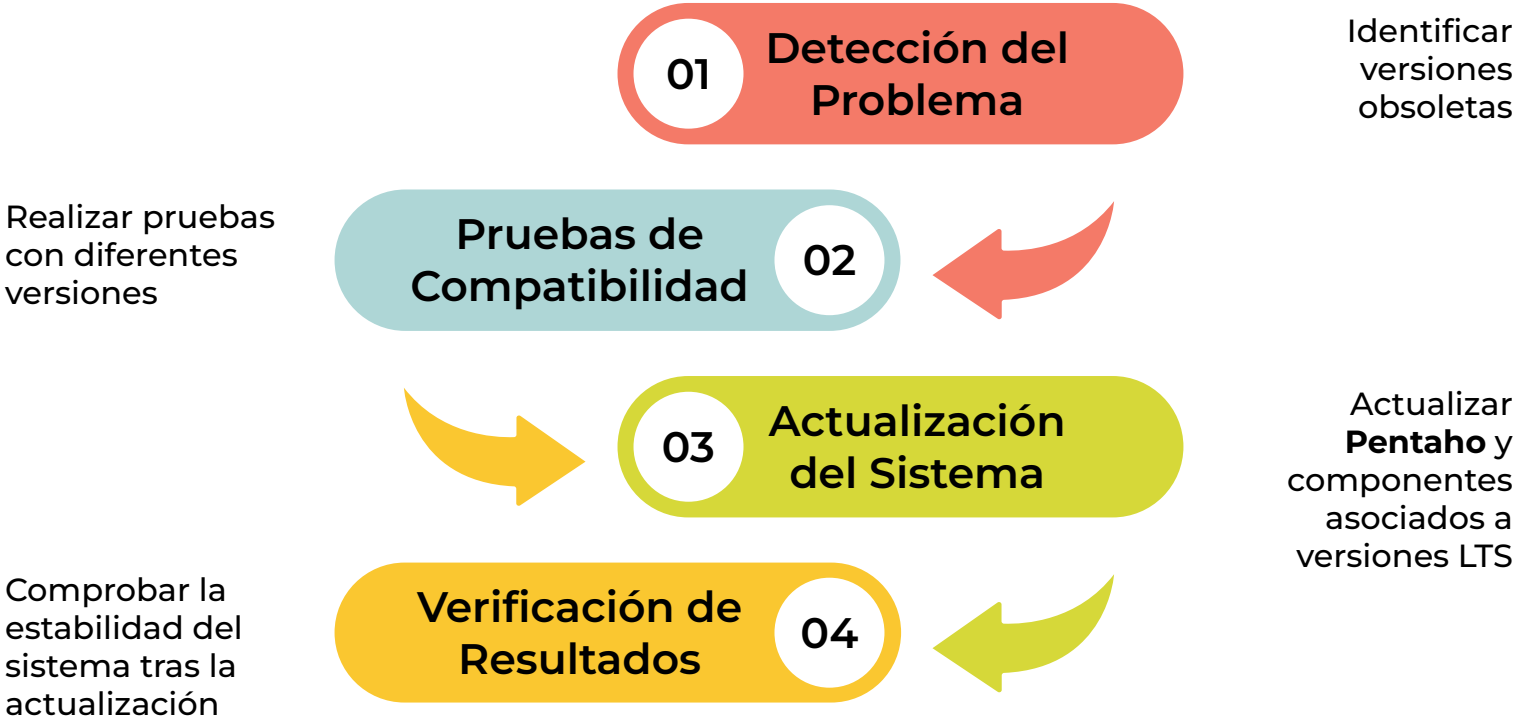
Formato

Pentaho carga jobs como .ktr, que están en formato XML

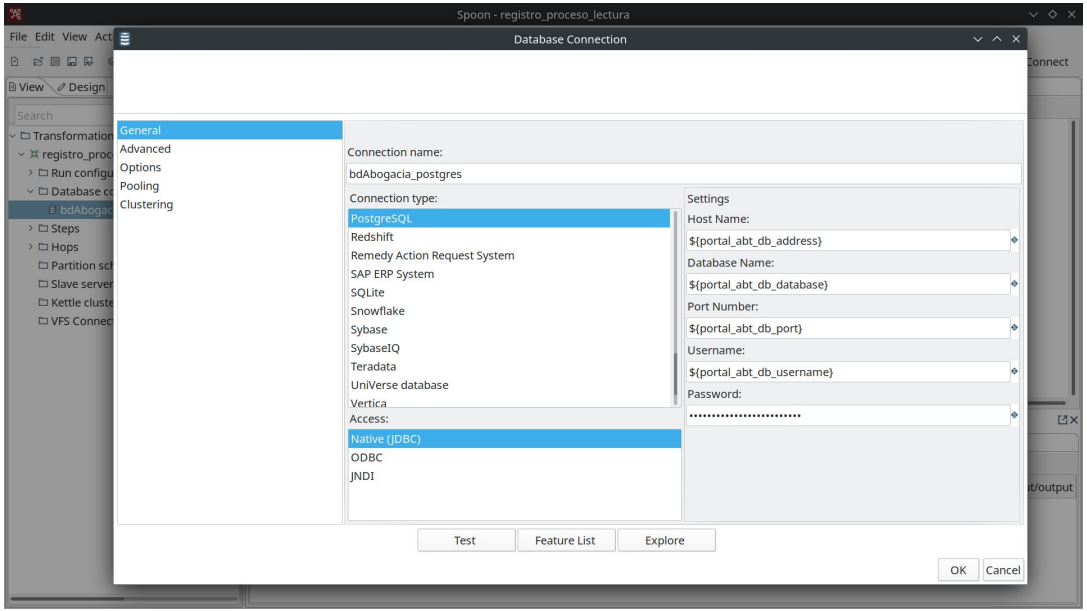
Filtrado

Error al insertar filas en las tables de la DB

Versión


01
Detección del Problema
Identificar versiones obsoletas
02
Pruebas de Compatibilidad
Realizar pruebas con diferentes versiones
03
Actualización del Sistema
Actualizar **Pentaho** y componentes asociados a versiones LTS
04
Verificación de Resultados
Comprobar la estabilidad del sistema tras la actualización

Variables de entorno



Ejemplo de la pantalla de configuración de variables de entorno del Pentaho

GRUPO HORUS

MINISTERIO DE ECONOMÍA Y FINANZAS PARAGUAY

Bug de los credenciales

Step 01

Identificación del Error

Revisar las configuraciones de conexión en Pentaho y uso adecuado de variables

Step 02

Corrección de Credenciales

Sustituir las variables $ por credenciales reales y verificar el uso correcto de argumentos

Step 03

Verificación de Acceso

Confirmar acceso correcto y ajustar kettle.properties para futuras implementaciones

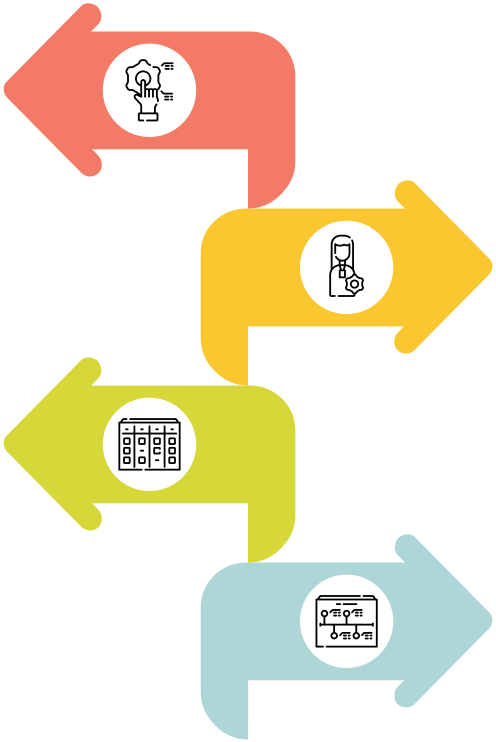
Errores de permisos

Contexto del Problema

Problemas de permisos durante la creación de archivos en el proceso ETL.

Creación de Carpeta

Uno de los trabajos en el proceso ETL requiere la creación de una carpeta específica



Ejecutar con Sudo

Uso de sudo para otorgar permisos elevados al script spoon.sh.

Verificación de Resultados

Confirmar que la ejecución con sudo permite la correcta creación de la carpeta

Error: Permisos para crear carpetas

Spoon - registro_proceso

File Edit View Action Tools Help

Connect

View | Design

Search

- Jobs
  - registro_proceso
    - Run configurations
    - Database connections
    - Job entries
    - Slave server
    - VFS Connections

registro_proceso

100

Execution Results

Logging | History | Job metrics | Metrics

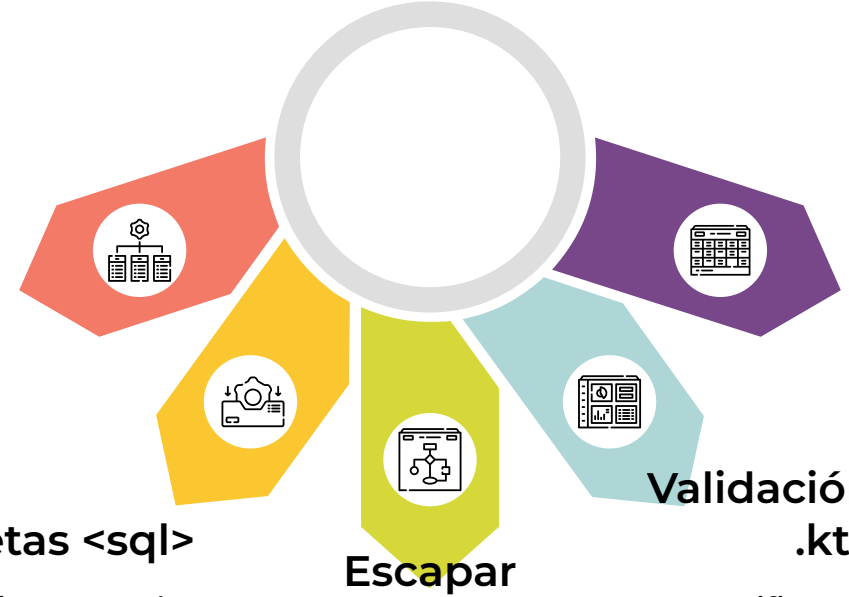
```
2024-10-11 09:41:49.813 - Text file output.0 -
2024-10-11 09:41:49.813 - Text file output.0 - org.apache.commons.vfs2.FileSystemException: Could not create file "file:///opt/abogacia/pentaho/error/inscripci
2024-10-11 09:41:49.813 - Text file output.0 - Could not create file "file:///opt/abogacia/pentaho/error/inscripcion_abt_error_20241011_094149.csv".
2024-10-11 09:41:49.813 - Text file output.0 -
2024-10-11 09:41:49.813 - Text file output.0 - Could not create file "file:///opt/abogacia/pentaho/error/inscripcion_abt_error_20241011_094149.csv".
2024-10-11 09:41:49.813 - Text file output.0 -
2024-10-11 09:41:49.813 - Text file output.0 -    at org.pentaho.di.core.vfs.KettleVFSImpl.getOutputStream(KettleVFSImpl.java:382)
2024-10-11 09:41:49.813 - Text file output.0 -    at org.pentaho.di.trans.steps.textfileoutput.TextFileOutput.getOutputStream(TextFileOutput.java:1074)
2024-10-11 09:41:49.813 - Text file output.0 -    at org.pentaho.di.trans.steps.textfileoutput.TextFileOutput.initFileStreamWriter(TextFileOutput.java:187)
2024-10-11 09:41:49.813 - Text file output.0 -    at org.pentaho.di.trans.steps.textfileoutput.TextFileOutput.initOutput(TextFileOutput.java:864)
2024-10-11 09:41:49.813 - Text file output.0 -    at org.pentaho.di.trans.steps.textfileoutput.TextFileOutput.init(TextFileOutput.java:828)
2024-10-11 09:41:49.813 - Text file output.0 -    at org.pentaho.di.trans.step.StepInitThread.run(StepInitThread.java:69)
2024-10-11 09:41:49.813 - Text file output.0 -    at java.base/java.lang.Thread.run(Thread.java:829)
2024-10-11 09:41:49.813 - Text file output.0 - Caused by: org.apache.commons.vfs2.FileSystemException: Could not create file "file:///opt/abogacia/pentaho/err
2024-10-11 09:41:49.813 - Text file output.0 -    at org.apache.commons.vfs2.provider.AbstractFileObject.createFile(AbstractFileObject.java:327)
2024-10-11 09:41:49.813 - Text file output.0 -    at org.pentaho.di.core.vfs.KettleVFSImpl.getOutputStream(KettleVFSImpl.java:350)
2024-10-11 09:41:49.813 - Text file output.0 -    at org.pentaho.di.core.vfs.KettleVFSImpl.getOutputStream(KettleVFSImpl.java:380)
2024-10-11 09:41:49.813 - Text file output.0 -    ... 6 more
2024-10-11 09:41:49.813 - Text file output.0 - Caused by: org.apache.commons.vfs2.FileSystemException: Could not write to "file:///opt/abogacia/pentaho/error
2024-10-11 09:41:49.813 - Text file output.0 -    at org.apache.commons.vfs2.provider.AbstractFileObject.getOutputStream(AbstractFileObject.java:1276)
2024-10-11 09:41:49.813 - Text file output.0 -    at org.apache.commons.vfs2.provider.AbstractFileObject.getOutputStream(AbstractFileObject.java:1238)
2024-10-11 09:41:49.813 - Text file output.0 -    at org.apache.commons.vfs2.provider.AbstractFileObject.createFile(AbstractFileObject.java:321)
2024-10-11 09:41:49.813 - Text file output.0 -    ... 8 more
2024-10-11 09:41:49.813 - Text file output.0 - Caused by: java.io.FileNotFoundException: /opt/abogacia/pentaho/error/inscripcion_abt_error_20241011_094149.c
2024-10-11 09:41:49.813 - Text file output.0 -    at java.base/java.io.FileOutputStream.open0(Native Method)
2024-10-11 09:41:49.813 - Text file output.0 -    at java.base/java.io.FileOutputStream.open(FileOutputStream.java:298)
```

Ejemplo de error del ETL debido a permisos insuficientes, imposible crear la carpeta en **opt/abogacia/pentaho/error**

GRUPO HORUS

MINISTERIO DE ECONOMÍA Y FINANZAS PARAGUAY

Error en la lectura de los jobs



Archivos .ktr (formato XML)

Problema con la carga de trabajos en Pentaho debido a caracteres NO escapados

Etiquetas <sql>

Identificar que el problema proviene de la interpretación errónea de los signos < y > en las consultas SQL

Escapar caracteres

Reemplazar los signos < y > con sus versiones escapadas en XML: < y >

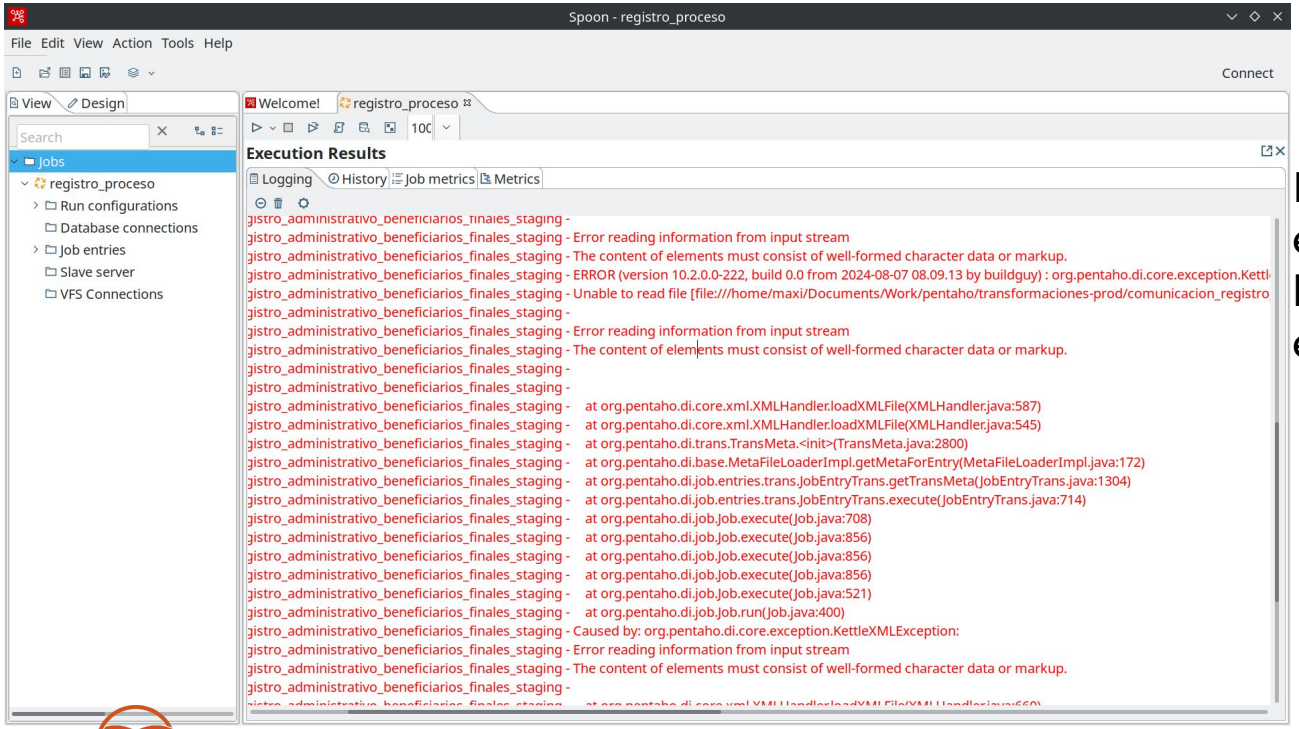
Validación de los .ktr

Verificar que los archivos .ktr están en el formato XML correcto antes de cargarlos en Pentaho

Resultados y Próximos Pasos

Resultados positivos tras aplicar las modificaciones.

Error en la lectura de los archivos .ktr



Ejemplo de otro error, fallo en la lectura de los **.ktr** en format **XML**

GRUPO HORUS

MINISTERIO DE ECONOMÍA Y FINANZAS PARAGUAY

Arrow Infographics

Columnas vacias

Problemas al insertar filas en la tablas en la base de datos PostgreSQL debido a valores nulos

Problemática

El pipeline ETL no anticipó correctamente la posibilidad de valores vacíos en la columnas

Estrategias de Mitigación

Implementación de una solución para evitar la interrupción del pipeline ETL.

Medidas Preventivas para Errores de Datos

Error insertando valores NULL

Ejemplo de mensajes de error

```
2024/09/10 13:50:06 - bdAbogacia_postgres - Connection to database closed!
2024/09/10 13:50:06 - Update ultimo.0 - Procesamiento finalizado (I=40, O=0, R=40, W=40, U=40, E=0)
2024/09/10 13:50:06 - inscripcion_sociedaes_abt - searching for annotations
2024/09/10 13:50:06 - inscripcion_sociedaes_abt - no annotations found
2024/09/10 13:50:06 - Spoon - La transformaci►n ha finalizado!!
2024/09/10 13:50:06 - inscripcion_sociedaes_abt - ERROR (version 9.0.0.0-423, build 9.0.0.0-423 from 2020-01-31 04.53.04 by buildguy) : Errores detectados!
2024/09/10 13:50:06 - inscripcion_sociedaes_abt - ERROR (version 9.0.0.0-423, build 9.0.0.0-423 from 2020-01-31 04.53.04 by buildguy) : Errores detectados!
2024/09/10 13:50:06 - inscripcion_sociedaes_abt - ERROR (version 9.0.0.0-423, build 9.0.0.0-423 from 2020-01-31 04.53.04 by buildguy) : Errores detectados!
2024/09/10 13:50:06 - registro_proceso - Finished job entry [inscripcion_sociedades_abt_final] (result=[false])
2024/09/10 13:50:06 - registro_proceso - Finished job entry [comunicacion_registro_administrativo_personas_estructuras_juridicas_staging] (result=[false])
2024/09/10 13:50:06 - registro_proceso - Finished job entry [comunicacion_registro_administrativo_beneficiarios_finales_staging] (result=[false])
2024/09/10 13:50:06 - registro_proceso - Finished job entry [inscripcion_sociedades_abt_staging] (result=[false])
```

```
2024-10-11 10:12:28.097 - cod_pais.0 - Finished processing (I=0, O=0, R=19854, W=19854, U=0, E=0)
2024-10-11 10:12:28.415 - inscripcion_sociedaes_abt - ERROR (version 10.2.0.0-222, build 0.0 from 2024-08-07 08.09.13 by buildguy) : Errors detected!
2024-10-11 10:12:28.418 - persona_fisica.0 - ERROR (version 10.2.0.0-222, build 0.0 from 2024-08-07 08.09.13 by buildguy) : Because of an error, this step can't continue:
2024-10-11 10:12:28.419 - persona_fisica.0 - ERROR (version 10.2.0.0-222, build 0.0 from 2024-08-07 08.09.13 by buildguy) : org.pentaho.di.core.exception.KettleExceptio
2024-10-11 10:12:28.419 - persona_fisica.0 - Error inserting row into table [persona_fisica] with values: [295615], [null], [null], [null], [null], [null], [null], [null], [nu
2024-10-11 10:12:28.419 - persona_fisica.0 -
2024-10-11 10:12:28.419 - persona_fisica.0 - Error inserting/updating row
2024-10-11 10:12:28.419 - persona_fisica.0 - ERROR: null value in column "nombres_apellidos" of relation "persona_fisica" violates not-null constraint
2024-10-11 10:12:28.419 - persona_fisica.0 -   Detail: Failing row contains (1644390, null, null).
2024-10-11 10:12:28.419 - persona_fisica.0 -
2024-10-11 10:12:28.419 - persona_fisica.0 -
2024-10-11 10:12:28.419 - persona_fisica.0 -     at org.pentaho.di.trans.steps.tableoutput.TableOutput.writeToTable(TableOutput.java:380)
2024-10-11 10:12:28.419 - persona_fisica.0 -     at org.pentaho.di.trans.steps.tableoutput.TableOutput.processRow(TableOutput.java:121)
2024-10-11 10:12:28.419 - persona_fisica.0 -     at org.pentaho.di.trans.step.RunThread.run(RunThread.java:63)
2024-10-11 10:12:28.419 - persona_fisica.0 -     at java.base/java.lang.Thread.run(Thread.java:829)
2024-10-11 10:12:28.419 - persona_fisica.0 - Caused by: org.pentaho.di.core.exception.KettleDatabaseException:
2024-10-11 10:12:28.419 - persona_fisica.0 - Error inserting/updating row
2024-10-11 10:12:28.419 - persona_fisica.0 - ERROR: null value in column "nombres_apellidos" of relation "persona_fisica" violates not-null constraint
2024-10-11 10:12:28.419 - persona_fisica.0 -   Detail: Failing row contains (1644390, null, null).
```

Identificación del origen del error, columna nombres_apellidos en la tabla persona_fisica

GRUPO HORUS

MINISTERIO DE ECONOMÍA Y FINANZAS PARAGUAY

Consideraciones extra

Un agregado al documento original para soporte adicional

Versiones recientes compatibles

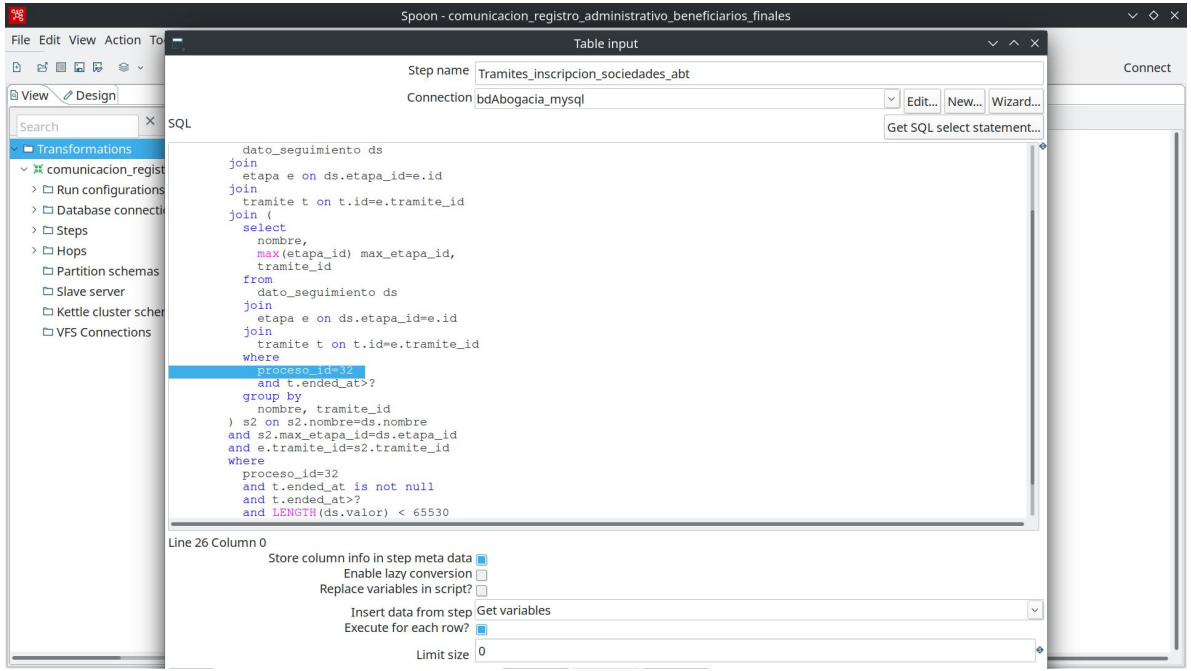
OS: Ubuntu 24.04 LTS

ETL: Pentaho 10.2

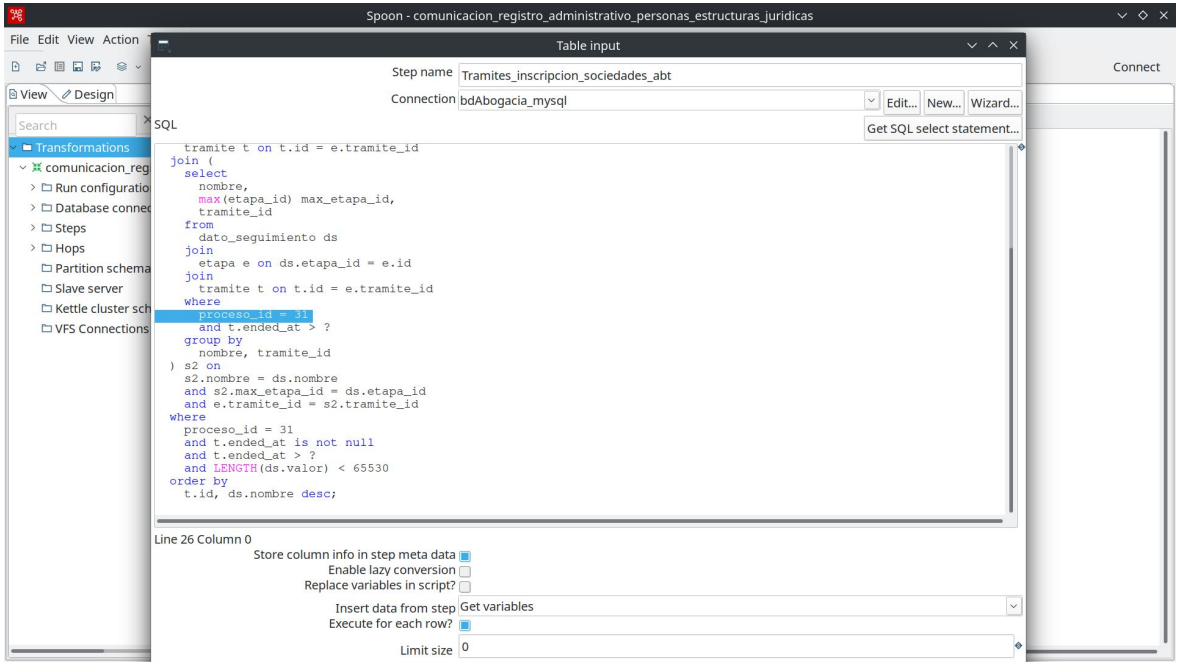
DBs:

- MySQL 8.0.9
- PostgreSQL 16.4

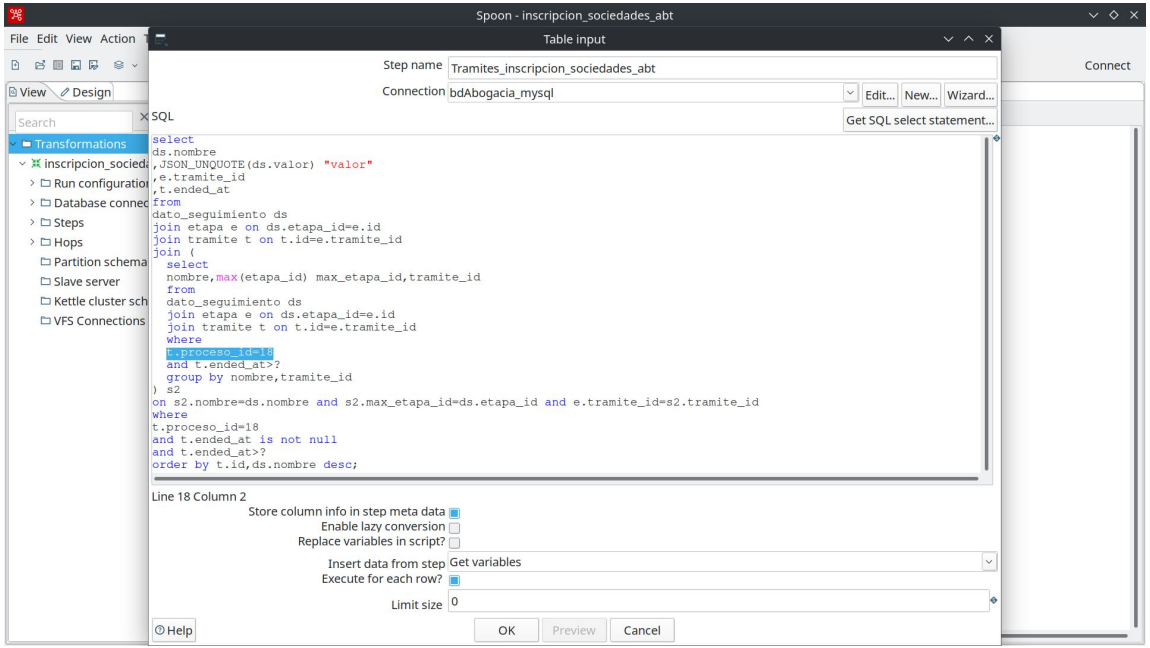
Comunicacion registro administrativo beneficiarios finales



Comunicacion registro administrativo personas estructuras juridicas



Inscripcion sociedades abt



Niveles de logging para debugging

Cuando ejecutas un trabajo o una transformación con el registro habilitado, tienes las siguientes opciones de nivel de detalle en el registro en la ventana de Opciones de Ejecución:

Nada	No registrar ninguna salida de log.
Error	Mostrar solo errores.
Mínimo	Usar solo un registro mínimo.
Básico	Este es el nivel predeterminado.
Detallado	Proporcionar un registro detallado.
Depuración	Para propósitos de depuración, salida muy detallada.
Nivel de Fila	Registro a nivel de fila. Esto generará una gran cantidad de datos en el log.

Resumen de los 5 Errores Corregidos

1. **Incompatibilidad de Conectores de Base de Datos**:
   - *Corrección*: Actualización de Pentaho y componentes a versiones más recientes compatibles con MySQL y PostgreSQL.
   - *Recomendación*: Mantener siempre actualizado el software y los conectores para evitar problemas de compatibilidad.
2. **Error de Autenticación de Contraseña**:
   - *Corrección*: Se resolvió el problema de variables de ejecución mal interpretadas, reemplazándolas por credenciales directas.
   - *Recomendación*: Asegurar de que las variables de entorno estén correctamente configuradas para evitar errores de autenticación.
3. **Problemas de Permisos con `sudo`**:
   - *Corrección*: Ejecutar el script `spoon.sh` con privilegios de `sudo` para permitir la creación de archivos necesarios.
   - *Recomendación*: Utilizar `sudo` cuando sea necesario para evitar errores de permisos en procesos críticos del sistema.

Resumen de los 5 Errores Corregidos (II)

1. **Error al Cargar Archivos `.ktr` por Signos `<` y `>` en SQL**:
   - *Corrección*: Reemplazo de los signos `<` y `>` por sus caracteres escapados (`<` y `>`) en las consultas SQL dentro de los archivos `.ktr`.
   - *Recomendación*: Usar validadores XML y asegúrate de que las consultas SQL estén correctamente formateadas antes de cargarlas.
2. **Inserción de Valores Nulos en Base de Datos**:
   - *Corrección*: Filtrado de filas con valores nulos en la columnas `nombres_appellidos, cargo y vigencia` en el flujo ETL de los jobs `sociedad_abt_finales y inscripcion_ sociedades,` respectivamente
   - *Recomendación*: Implementa validaciones tanto en el front-end como en el ETL para manejar correctamente los datos incompletos o vacíos antes de la inserción.

CLARA YNES OZUNA ROTELA

Digitally signed by CLARA YNES OZUNA ROTELA
Date: 2024.11.25 16:38:39 -03'00'