

Secuencia de Recuperación ante Fallo de Carga de Datos de un Operacional en Data WareHouse

*Plan de Continuidad Tecnológico (PCT)*

|  |  |  |
| --- | --- | --- |
| Nombre del Documento: | Fecha: | Versión: |
| 09. HRE\_PCT\_Sec\_Recuperación ante Fallo de carga de datos de un operacional en DWH\_v2.00.docx | 30/08/2018 | 2.00 |
| Responsable: | Sr. Óscar Domínguez | |
| Elaborado por: | SERBAN | |
| Revisado por: | Sr. Javier Sánchez / Sr. Santiago Uriel | |
| Aprobado por: |  | |
| Ámbito de afectación: | HRE | |

CONTROL DE DOCUMENTACIÓN

LISTA DE DISTRIBUCIÓN

|  |  |
| --- | --- |
| Nombre del destinatario | Ámbito del destinatario |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |

CONTROL DE CAMBIOS DEL DOCUMENTO

|  |  |  |  |
| --- | --- | --- | --- |
| Ver. | Fecha | Descripción cambios | Páginas afectadas |
| 0.1 | 17-6-16 | Versión Inicial |  |
| 1.00 | 2-10-17 | Revisión 2017 |  |
| 2.00 | 30-8-18 | Revisión 2018 |  |
|  |  |  |  |

ACCIONES PENDIENTES

|  |  |  |
| --- | --- | --- |
| Descripción | Responsable | Fecha objetivo |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |

ÍNDICE

[ÍNDICE 3](#_Toc523381540)

[1 INTRODUCCIÓN 4](#_Toc523381541)

[1.1 ÁMBITO DE APLICACIÓN 4](#_Toc523381542)

[1.2 ALCANCE 4](#_Toc523381543)

[1.3 TIEMPO DE EJECUCIÓN 4](#_Toc523381544)

[1.4 DOCUMENTOS RELACIONADOS 4](#_Toc523381545)

[2 DIAGRAMA DE FLUJO 5](#_Toc523381546)

[3 DESCRIPCIÓN DE LA SECUENCIA 6](#_Toc523381547)

[4 ROLES Y RESPONSABILIDADES DE LA SECUENCIA 7](#_Toc523381548)

[5 ACTIVIDADES DE LA SECUENCIA 8](#_Toc523381549)

[5.1 MONITORIZACIÓN DE LA CARGA 8](#_Toc523381550)

[5.2 IDENTIFICAR EL ORIGEN DEL PROBLEMA 12](#_Toc523381551)

[5.3 RESOLVER EL ORIGEN DEL PROBLEMA 13](#_Toc523381552)

[5.4 RELANZAR LA RÉPLICA 14](#_Toc523381553)

[6 ANEXOS 16](#_Toc523381554)

[6.1 ANEXO I: ESQUEMA DEL SISTEMA DE INFORMACIÓN DATA WAREHOUSE 16](#_Toc523381555)

[6.2 ANEXO II: ESQUEMA DEL ORQUESTADOR DE CARGA 17](#_Toc523381556)

INTRODUCCIÓN

Ante un fallo de carga de datos de un operacional en el Data WareHouse, se procederá a ejecutar la presente secuencia de recuperación.

Esta Secuencia de Recuperación describe paso a paso como se debe proceder para la resolver una incidencia de réplica de los datos de los diferentes orígenes al sistema de información Data Warehouse.

ÁMBITO DE APLICACIÓN

La presente secuencia de recuperación afecta al sistema de información Data WareHouse y la coherencia de los datos almacenados con los de las diferentes fuentes.

ALCANCE

El alcance del documento queda determinado a la restauración de la integridad y coherencia de los datos en HRE\_DWH\_STAGE respecto de las fuentes de los mismos, a identificar y solucionar los diferentes errores en el proceso.

TIEMPO DE EJECUCIÓN

El tiempo de ejecución de la secuencia presenta una alta variación

| Inicio/Fin/Duración | Descripción |
| --- | --- |
| Inicio | Cuando se detecta que no se ha realizado la carga de datos de un operacional en el Data WareHouse |
| Fin | Cuando se comprueba que se ha realizado correctamente la carga |
| Duración | 1-4 Horas |

DOCUMENTOS RELACIONADOS

A la hora de realizar esta secuencia, se han tenido en cuenta los siguientes documentos:

| Documentación de Referencia |
| --- |
|  |
|  |
|  |

DIAGRAMA DE FLUJO

SIN ERROR

ERROR

**Resolver origen del problema**

**Identificar origen del problema**

**Lanzar recarga parcial**

**Monitorizar Carga**

**FIN**

DESCRIPCIÓN DE LA SECUENCIA

Durante la ejecución del proceso de carga diaria pueden presentarse diversos problemas por los cuales la carga réplica diaria puede fallar total o parcialmente.

Las posibles causas se pueden agrupar según el subsistema donde se produzca el problema:

* Fallos en el origen: No hay datos disponibles, no tenemos acceso a ellos, cambio en la estructura de la tabla de origen, …
* Fallos en la conexión: Generalmente debidos a cortes en la red.
* Fallos en el DWH: Falta de espacio en destino, colisión con otro proceso,…

Los pasos a seguir en la secuencia de recuperación:

1. Monitorización de la carga para detectar posibles problemas.
2. Identificar el origen del problema.
3. Resolución del problema.
4. Relanzar la carga de los elementos afectados.

ROLES Y RESPONSABILIDADES DE LA SECUENCIA

A continuación, especificamos los datos de contacto del Personal Externo de HRE, ordenados jerárquicamente, que se responsabilizarían de la resolución de la incidencia, en caso de materializarse la contingencia.

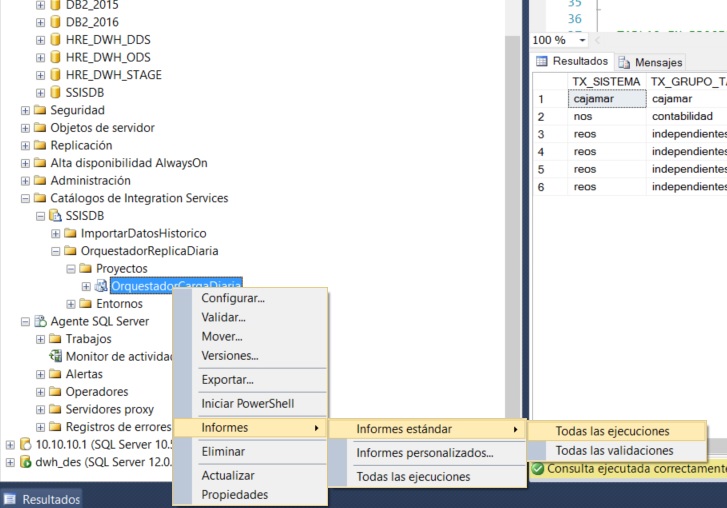
|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Personal Externo** | | | |  | **Desde dónde ejecutar la Secuencia** |
| Empresa | Nombre | Teléfono | Ubicación |  |
| SERBAN | Jaime Cano | Pendiente | Pendiente |  | Desde cualquier ubicación  Vía VPN (local) sobre DWH05PRO con cualquier herramienta de gestión de BBDD |

ACTIVIDADES DE LA SECUENCIA

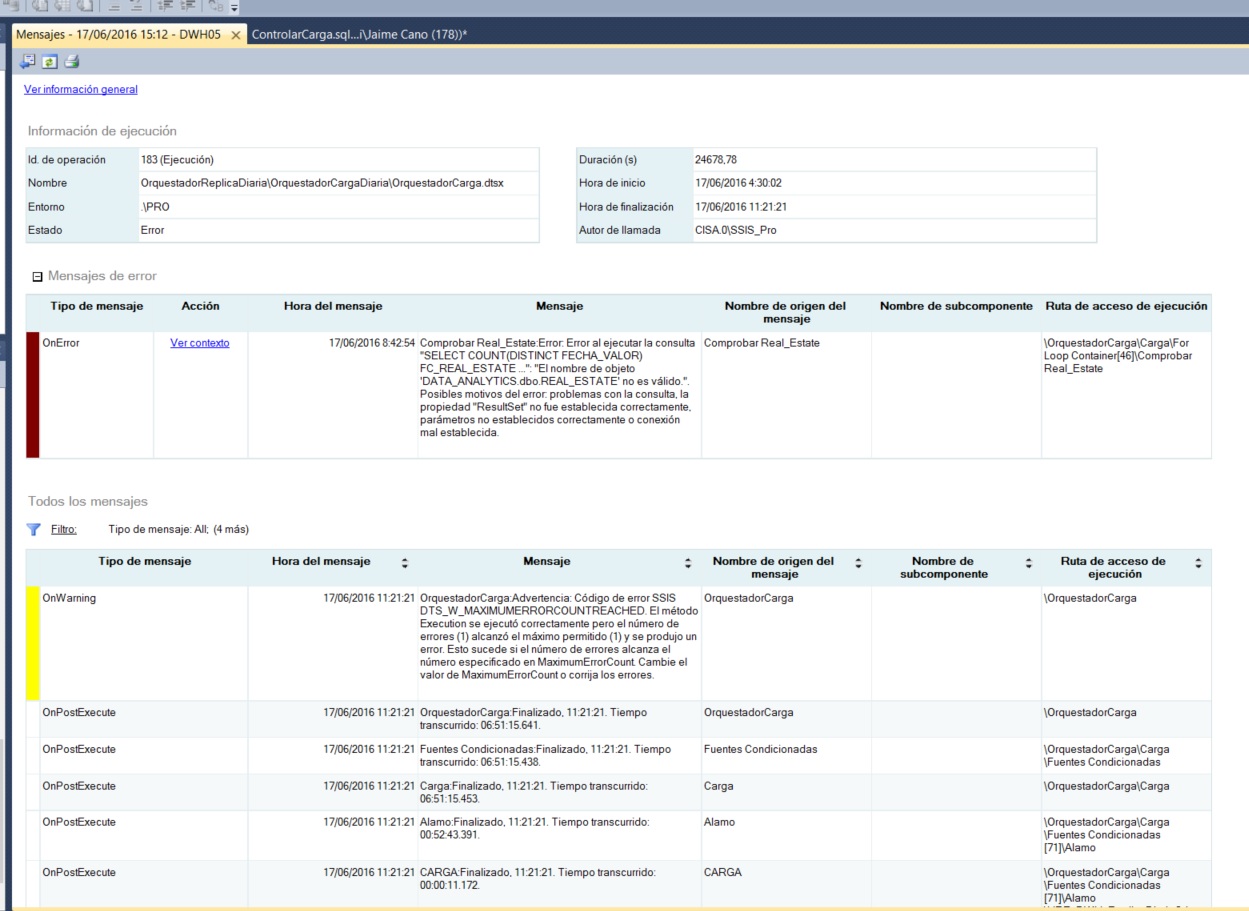
MONITORIZACIÓN DE LA CARGA

El proceso de carga diaria está dirigido por un proceso de SSIS llamado OrquestadorCargaDiaria. Este proceso gestiona el orden y el tiempo al que se lanza cada proceso. Éste a su vez lanza una vez por cada grupo de tablas el proceso SSIS ReplicaDiariaOrigenes que está integrado en el mismo proyecto de Visual Studio. En caso de error este proceso se relanza una segunda vez 15 minutos después de la primera finalización errónea.

El progreso de este proceso puede ser monitorizado a través del informe de ejecución del Catálogo de SSIS al que se puede acceder desde SSMS:



Con esto podemos acceder a los errores más graves, que afectan a un subsistema entero o a la carga por completo. El informe tiene este aspecto:

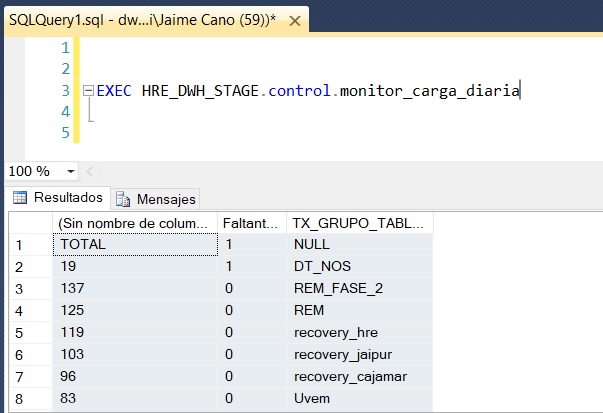


La pestaña de mensajes nos muestra todos los mensajes del log de SSIS, destacando en primer lugar los de error. Éstos no tienen porqué detener el proceso, ya que algunos están controlados y la propia lógica del proceso los maneja (repitiendo la operación después de un tiempo, por ejemplo).

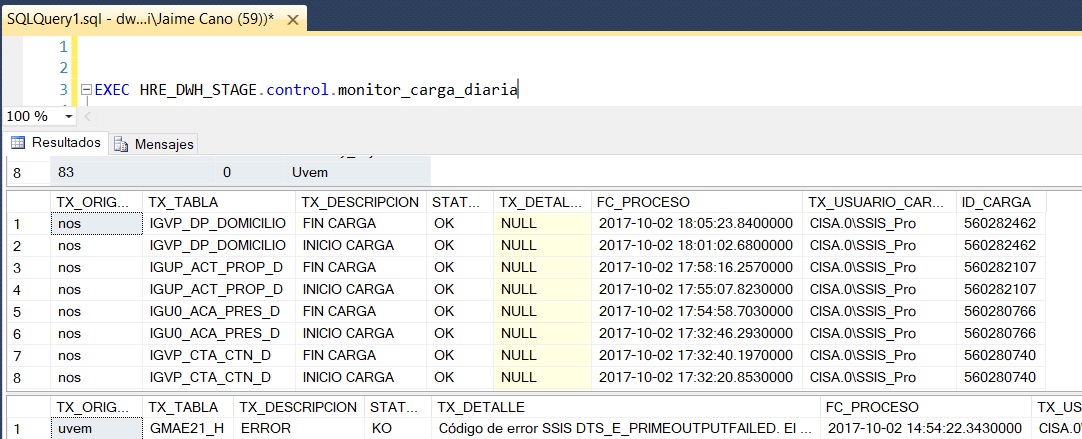
Además del catálogo de SSIS, todas las operaciones de copia de datos en una tabla se loguean en la tabla control.LOG\_PROCESO\_CARGA. Todas crean dos entradas en el log como mínimo, una cuando inicia la carga y otra cuando finaliza bien sea correctamente o se produzca algún error.

El procedimiento almacenado *HRE\_DWH\_STAGE.control.Monitor\_Carga\_Diaria* devuelve la información relativa al estado de la carga a través de 5 ventanas de resultados:

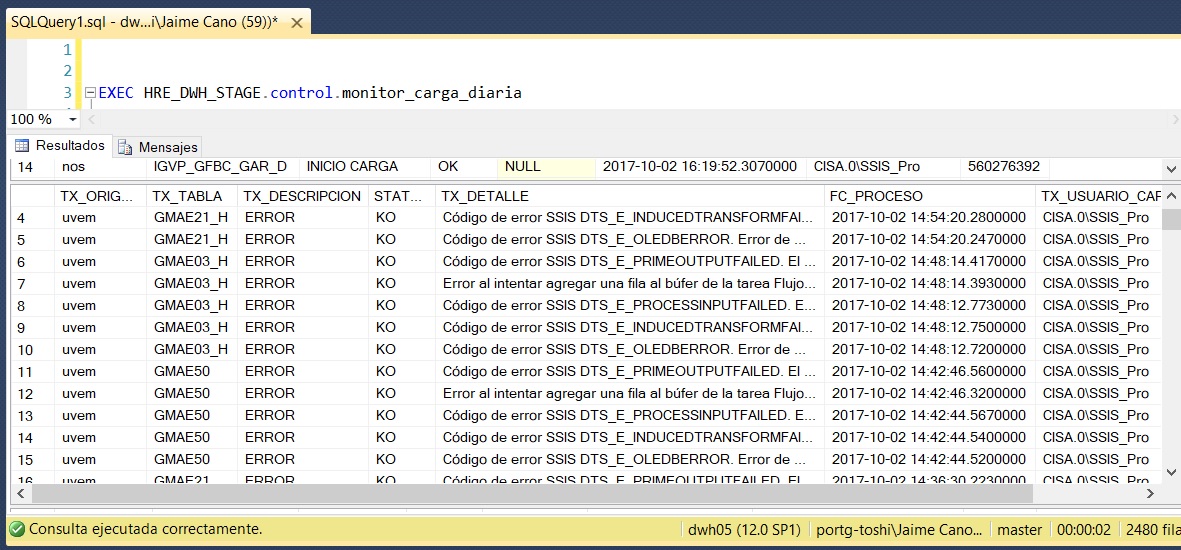
Ventana de resumen de carga. Tablas cargadas y faltantes por grupo.

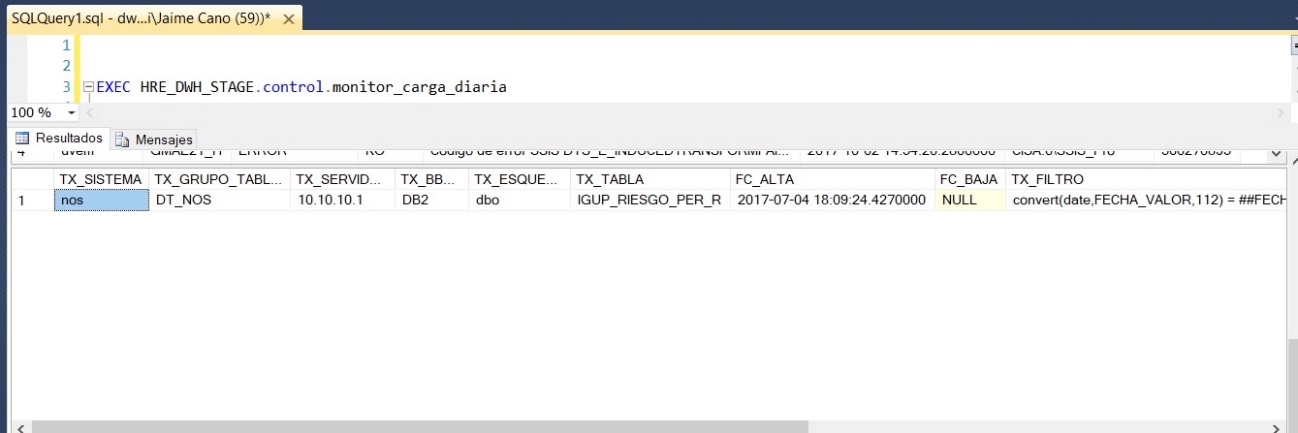


Ventana de Log carga. Todas las operaciones en orden temporal decreciente.

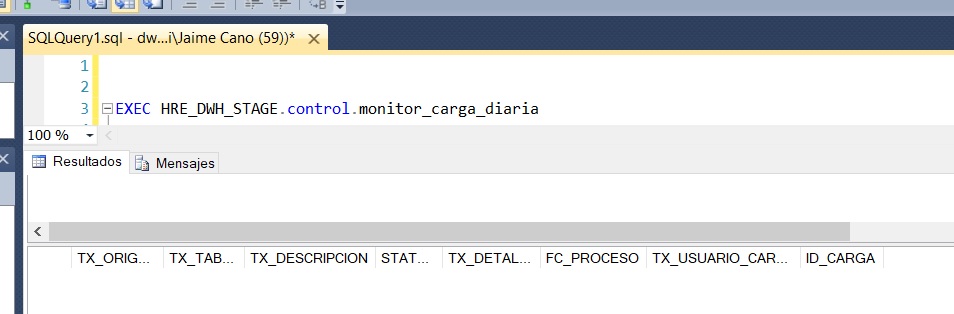


Ventana de Errores: Todos los errores ordenados por grupo y hora del proceso.



Ventana de tablas restantes.

Ventana de tablas en proceso de carga.



IDENTIFICAR EL ORIGEN DEL PROBLEMA

Generalmente el log de LOG\_PROCESO\_CARGA es suficientemente descriptivo a la hora de identificar estos errores.

Ejemplos:

Cambios de esquema en origen:

* *La tabla no existe*
* *El nombre de columna o los valores especificados no corresponden a la definición de la tabla.*
* *El nombre de columna 'fecha\_captura' no es válido.*

Error de los datos en origen:

* *Infracción de la restricción PRIMARY KEY 'PK\_\_FUHPVP\_\_27EA49CA2B7146BA'. No se puede insertar una clave duplicada en el objeto 'prinex\_gesnova.FUHPVP'. El valor de la clave duplicada es (000, 000).*
* *No se puede capturar una fila del proveedor OLE DB "OraOLEDB.Oracle" para el servidor vinculado "RECOVERY".*

Fallo de conexión:

* *No se puede inicializar el objeto de origen de datos del proveedor OLE DB "OraOLEDB.Oracle" para el servidor vinculado "PRINEX\_GESNOVA".*
* *Proveedor de TCP: Se agotó el tiempo de espera del semáforo.*
* *Interfaces de red SQL Server: Error al buscar el servidor/instancia especificado [xFFFFFFFF].*

RESOLVER EL ORIGEN DEL PROBLEMA

Cambios de esquema en Origen

El caso más común es que se hayan añadido columnas, o se hayan cambiado la definición de las mismas, en la tabla de origen. En este caso lo que hay que hacer es comparar el esquema y añadir las columnas nuevas o modificar las existentes en la tabla de réplica.

Error en los datos de Origen.

Generalmente es debido a que en una tabla en la que no debería haber duplicados lo hay. En este caso se debe analizar si debemos cambiar nuestra definición de la tabla y quitar la restricción de unicidad generalmente cambiando la Primary Key por un Índice Clúster con los mismos campos.

Otros problemas pueden ser que los datos de origen tengan una codificación distinta a la de nuestra base de datos o el proveedor OLEDB(no es una codificación ISO UTF-8) y el campo no sea UNICODE, en este caso debemos implementar una vista intermedia que convierta los datos en origen a lo que espera nuestra BBDD, por ejemplo:

USE [HRE\_DWH\_STAGE]

GO

SET ANSI\_NULLS ON

GO

SET QUOTED\_IDENTIFIER ON

GO

CREATE VIEW [RECOVERY\_HRE\_HAYA02].[TAR\_TAREAS\_NOTIFICACIONES] as select

\* from openquery(RECOVERY,'select TAR\_ID

,CLI\_ID

(...)

,convert(TAR\_DESCRIPCION,''US7ASCII'',''WE8ISO8859P1'')

(...)

,RPR\_REFERENCIA

,T\_REFERENCIA

,SYS\_GUID from HAYA02.TAR\_TAREAS\_NOTIFICACIONES');

GO

El código *convert(TAR\_DESCRIPCION,''US7ASCII'',''WE8ISO8859P1'')* es el que soluciona el problema aquí.

Fallos de conexión

Generalmente se deben a cortes en el servicio de red o que la base de datos de origen no está disponible momentáneamente. En estos casos una simple recarga es suficiente.

Si el problema persiste hay que escalar el problema para que o bien nos restauren los permisos en la BBDD de origen o nos abran puertos en el Firewall o lo que sea que nos esté cortando el acceso, antes de recargar.

RELANZAR LA RÉPLICA

Réplica completa

Si hay que recargar todas las tablas esto se hace mediante el Job *HRE\_DWH\_STAGE\_ReplicaDiariaOrigenes\_COMPLETO* que es el que se lanza cada día para realizar el proceso.

Réplica Parcial

Si el problema sólo ha afectado a un subconjunto de las tablas o a una en concreto hay que lanzar sólo un conjunto de tablas. Si la recarga afecta a un grupo de tablas completo hay Jobs de réplica ya creados para estos grupos, como por ejemplo *HRE\_DWH\_ReplicaDiariaOrigenes\_Recovery* que recargará todas las tablas de *Recovery*.

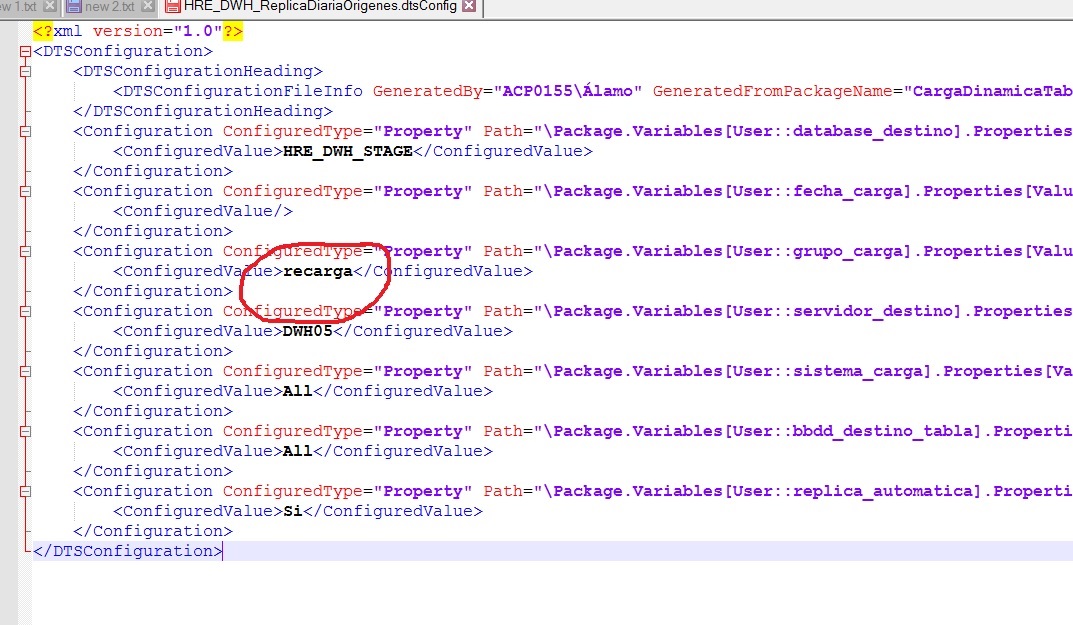
Si no queremos recargar un grupo completo o los errores afectan a grupos diferentes podemos crearnos un grupo ad-hoc para la recarga, asignándosele a las tablas afectadas por los errores:

UPDATE control.fuentes\_dwh

SET TX\_GRUPO\_TABLAS = 'recarga'

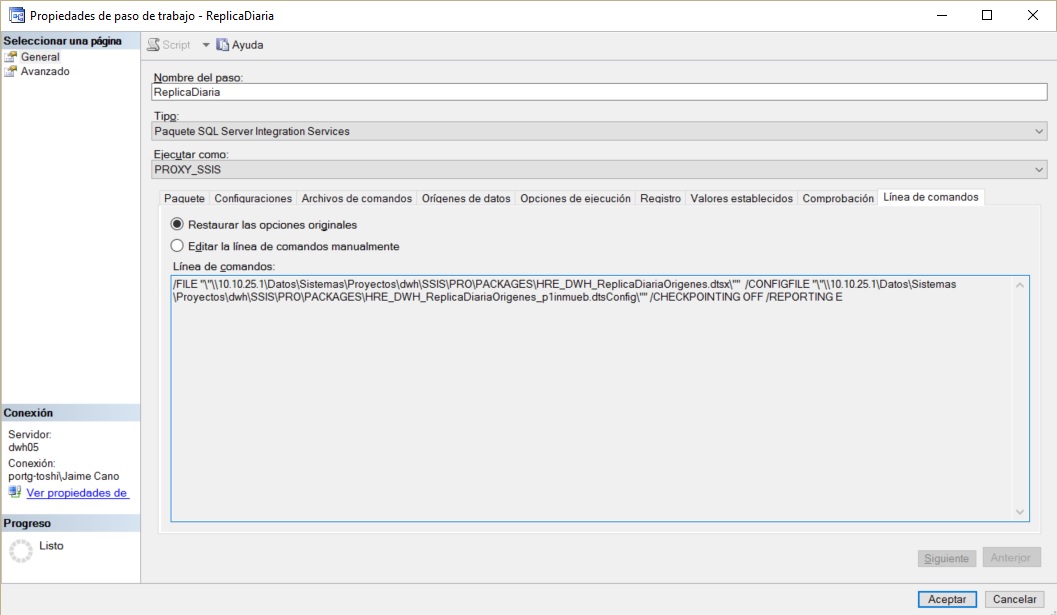
where tx\_tabla IN ('tabla1','tabla2','tabla3')

Luego configuramos el archivo de configuración para que cargue este grupo:



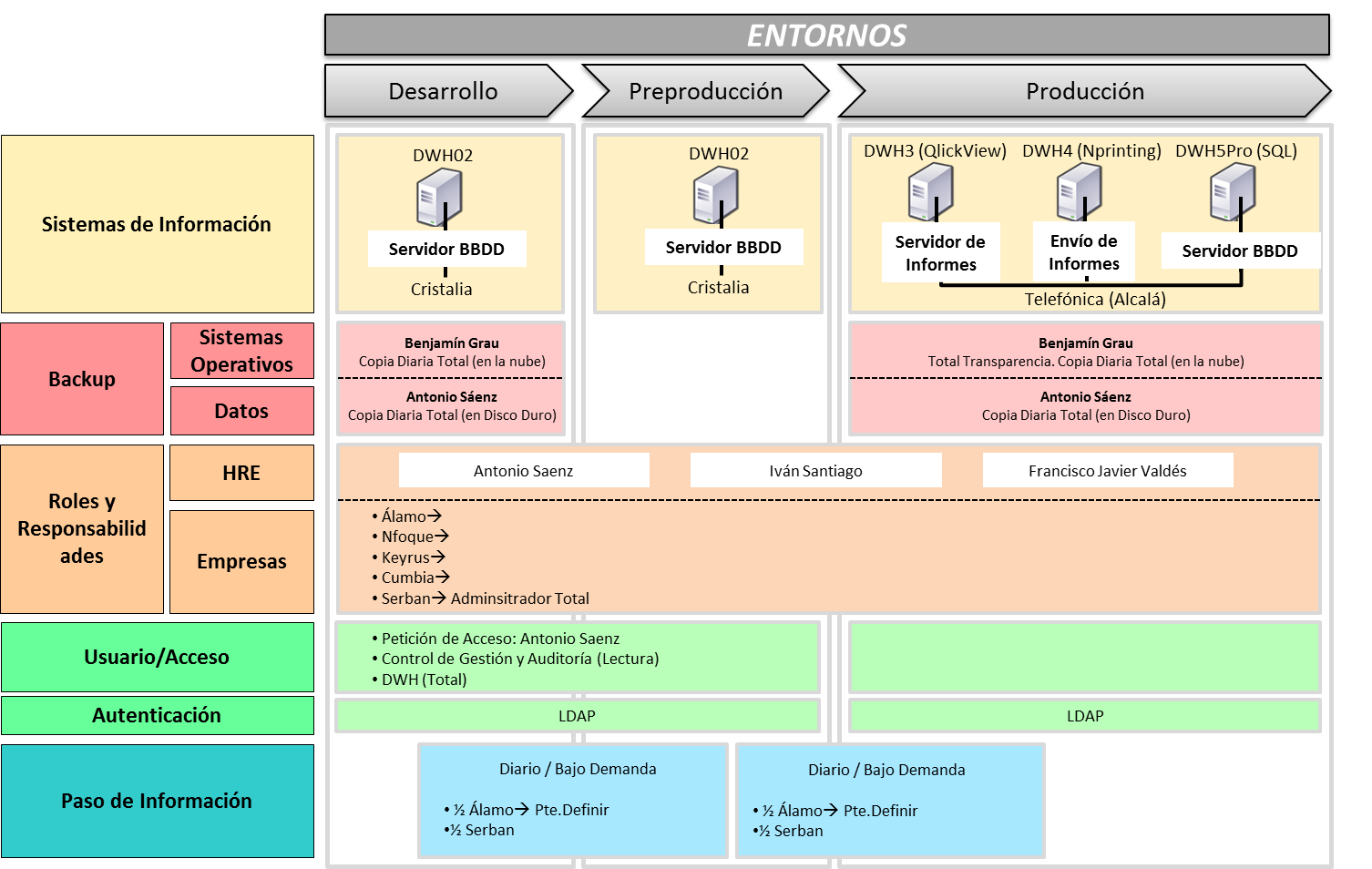
Posteriormente a la carga hay que restablecer todas las tablas a sus grupos de carga originales para que se carguen en días sucesivos con normalidad. Si no se hace las tablas quedarán descolgadas del proceso de carga.

Todos los Jobs de réplica diaria excepto el que está integrado en el Orquestador, funcionan a través del mismo paquete SSIS *HRE\_DWH\_ReplicaDiariaOrigenes.dtsx* que se configura para cada job mediante ficheros de configuración como por ejemplo *HRE\_DWH\_ReplicaDiariaOrigenes\_Aspro.dtsConfig* o *HRE\_DWH\_ReplicaDiariaOrigenes\_Reos.dtsConfig*. Hay que comprobar que el Job que vamos a lanzar se va a ejecutar con el archivo de configuración que hemos modificado. Esto lo haremos modificando el paso de Job y comprobando la ruta del paquete y del archivo de configuración:



ANEXOS

ANEXO I: ESQUEMA DEL SISTEMA DE INFORMACIÓN DATA WAREHOUSE



ANEXO II: ESQUEMA DEL ORQUESTADOR DE CARGA

El orquestador de carga divide ésta en tres bloques:

1. Fuentes no Condicionadas: Está formada por fuentes independientes a las de Data Type y REOS.REAL\_ESTATE. Está formada por los siguientes grupos(entre paréntesis el grupo tal y como está configurado en la tabla control.FUENTES\_DWH):
   1. SarebWorkflowExpedientes(bi\_expedientes)
   2. Aspro (aspro)
   3. Prinex\_gesnova (prinex\_gesnova)
   4. Mini\_recovery (mini\_recovery\_hre)
   5. Recovery (recovery\_hre)
   6. Recovery Cajamar (recovery\_cajamar, recovery\_cajamar\_2)
   7. Recovery Jaipur (recovery\_jaipur)
   8. P1inmueb (p1inmueb)
   9. Avales (avales)
2. Fuentes Condicionadas por horario: Son las fuentes que tienen como condicionante una determinada hora para empezarse a cargar por temas de disponibilidad en origen.
   1. Maestro de Operaciones (5:00) (Maestro\_Operaciones)
   2. Prinex Cajamar (8:30) (Prinex\_cajamar, Prinex\_cajamar\_2, Prinex\_cajamar\_3)
3. Fuentes Condicionadas: Dependen que haya datos en las tablas de datatape en [10.10.10.1] en Valencia actualizados al día de la carga, el proceso lo chequea cada 5 minutos hasta un máximo de 12 horas después de que se inicie la carga, cuando desiste el proceso. Grupos:
   1. Alamo (alamo)
   2. Maestro\_Activos (mactivos)
4. Real \_Estate: Dependen que haya datos en la tabla REAL\_ESTATE en [10.10.10.1] en Valencia actualizados al día de la carga, el proceso lo chequea cada 5 minutos hasta un máximo de 12 horas después de que se inicie la carga, cuando desiste el proceso. Grupos:
   1. REOS (reos)

Al finalizar estos cuatro bloques el sistema envía los correspondientes correos de *Carga Finalizada* si el proceso finalizó correctamente o *Carga Desistida* si después de doce horas no hubo datos en REAL\_ESTATE.

