CADA HECHO DE TU VIDA *Cuenta*

# Manual de Procesamiento

30 de Octubre, 2018

| | |
|---|---|
| Elaborado por: Carolina Mina. | Analista de GEERA |
| Revisado por: Libertad Trujillo. | Jefe de GEERA |
| Aprobado por: Libertad Trujillo | Jefe de GEERA |

**Información del Documento:**

| | |
|---|---|
| **Resumen:** | El presente documento tiene la finalidad de proporcionar una idea clara de cómo se realiza el proceso de captación y procesamiento de la información para la construcción de la Base de Datos (BDD) del DIEE, en el cual se explica paso a paso cada fase de transformación de la información para tener una base de datos depurada y lista para ser analizada y publicada. |

**Control e Historial de Cambios:**

| Versión | Descripción del cambio/ Autor (a): | Fecha de creación y actualización |
|---|---|---|
| V.0.4 | Actualización de: Observaciones y reglas (Carolina Mina) | 31 de noviembre de 2016 |
| V.0.5 | Actualización de: Procesamiento para el cálculo del DIEE 2016: (Carolina Mina) | 08 de noviembre de 2017 |
| V.0.6 | Actualización de: Procesamiento para el cálculo del DIEE 2017 (Carolina Mina) | 30 de octubre de 2018 |

# Contenido

# INTRODUCCIÓN.

El Directorio de Empresas (DIEE) se compone de diferentes bases de datos, entre las principales se tiene: el SRI, IESS, bases de datos con ciertas variables investigadas por el call center del Directorio de Empresas e información obtenida de ciertas encuestas internas del INEC como son: el Censo Económico (CENEC), la encuesta Exhaustiva, ACTI, Ambientales, Industriales, entre otras.

Esta información es complementada y validada en menor proporción con matrices de equivalencias de variables codificadas de diferentes maneras entre la fuente de información y el proveedor.

La información por cada fuente se obtiene de diferentes maneras; es decir que se tiene diferentes formatos o diferentes motores de bases de datos, diferentes modos de transmisión; es por eso que se hace sustancial la intervención de procesos ETL's que se encargan de transformar a toda la información y llevarla a la lógica definida en el DIEE.

Una vez conseguido que la información este consolidada, el DIEE procede a realizar análisis de la información y posteriormente se realiza una publicación.

El presente documento tiene la finalidad de proporcionar una idea clara de cómo se realiza el proceso de captación y procesamiento de la información para la construcción de la Base de Datos (BDD) del DIEE.

A través del documento se explicará paso a paso cada fase de transformación de la información para tener una base de datos depurada y lista para ser analizada y publicada.

Como se explicó anteriormente cada fuente de información viene al DIEE de diferentes maneras como por ejemplo:

- SRI: Base de datos en Oracle.
- IESS: Base de datos en Oracle.

- Call Center: Archivos Excel.
- Encuestas INEC: Archivos Excel.

Es por esto que para cada fuente de información se lleva un tratamiento diferente porque además de ser diferentes en formato son diferentes en contenido.

Las herramientas de software con las que el DIEE trabaja son:

- Motor de Base de Datos para captación: Oracle Express Edition 10g
- Motor de Base de Datos para procesamiento: PostgresSQL 9.1
- Sistema de Gestión de Base de Datos: PgAdminIII 1.5
- Heramienta BI: Pentaho Data Integration 6.1
- Herramienta DQ: SQL Power DQguru
- Herramienta de análisis de información: SPSS 22

***PostgresSQL,*** es un motor de bases de datos relacionales (RDBMS) que verifica integridad referencial, distribuida bajo licencia BSD y con su código fuente disponible libremente. Sus características técnicas la hacen una de las bases de datos más potentes y robustas del mercado.

***PgAdminIII***, facilita la gestión y administración de bases de datos ya sea mediante instrucciones SQL o con ayuda de un entorno gráfico. Permite acceder a todas las funcionalidades de la base de datos; consulta, manipulación y gestión de datos.

***Pentaho Data Integration,*** abre, limpia e integra información entre bases de datos y la pone en manos del usuario. Provee consistencia y en una sola versión se tiene todos los recursos de información, a través de la construcción de ETL's.

***SQL Power DQguru,*** es una herramienta que no solo realiza una limpieza de datos, sino que también valida y corrige direcciones, identifica y elimina duplicados y crea referencias cruzadas.

***SPSS,*** es un sistema de análisis estadístico y gestión de información que puede trabajar con datos de distintos formatos generando, análisis estadísticos complejos que nos permitirán descubrir relaciones de dependencia e interdependencia, establecer clasificaciones de sujetos y variables, predecir comportamientos, entre otros.

Con esta pequeña introducción se da una idea de cómo es la captación y el procesamiento en el DIEE, el cual tiene un orden secuencial para llegar a su objetivo final que es la base de datos depurada.

# PROCESO DE CONSTRUCCIÓN DE BASE DE DATOS DEL DIRECTORIO DE EMPRESAS

La construcción de la BDD del DIEE se compone de varias fases, en el

Gráfico N: 1 se las expone de manera general:

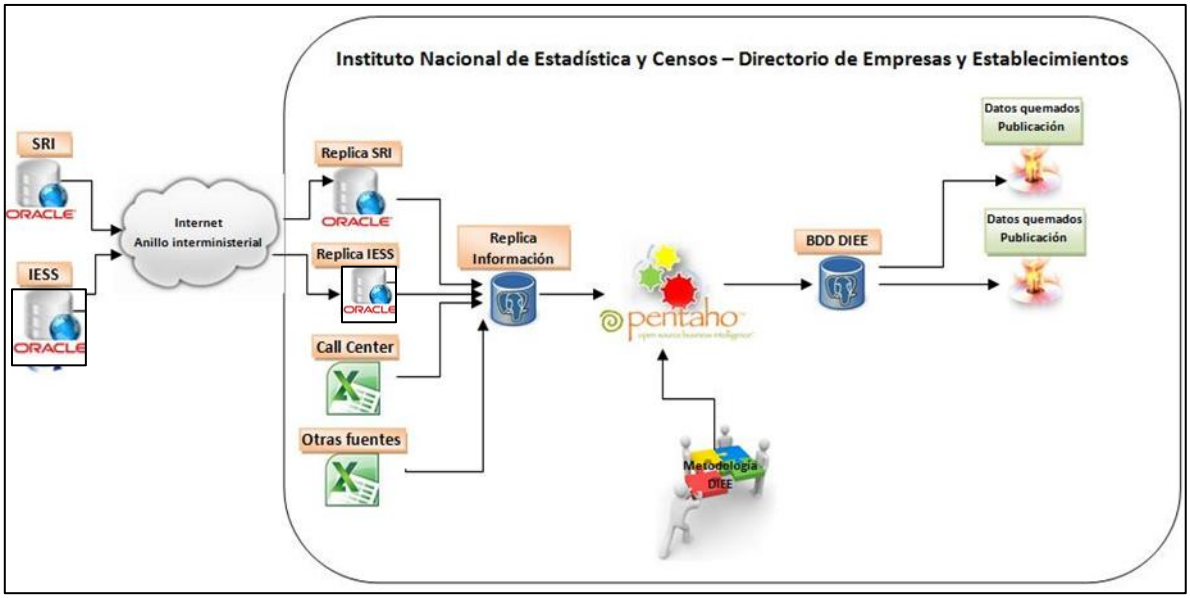


Gráfico N: 1.

El propósito de la captación de información es crear un entorno de Base de datos similar al que se tiene el proveedor de información en su Base, para ello es necesario conocer:

- Medio de comunicación a usar
- Variables a recibir

- Formato de información enviada por el proveedor
- Formato de cada variable enviada
- Volumen de información
- Frecuencia de transmisión.

Una vez identificados con claridad estos datos, se procede a diseñar y desarrollar el mecanismo de transmisión de información; sea este por uso de herramientas propias del motor de Bases de Datos, uso de Herramientas externas para tratamiento de información como ETL´s y pequeños programas con la interacción de aplicaciones como Excel.

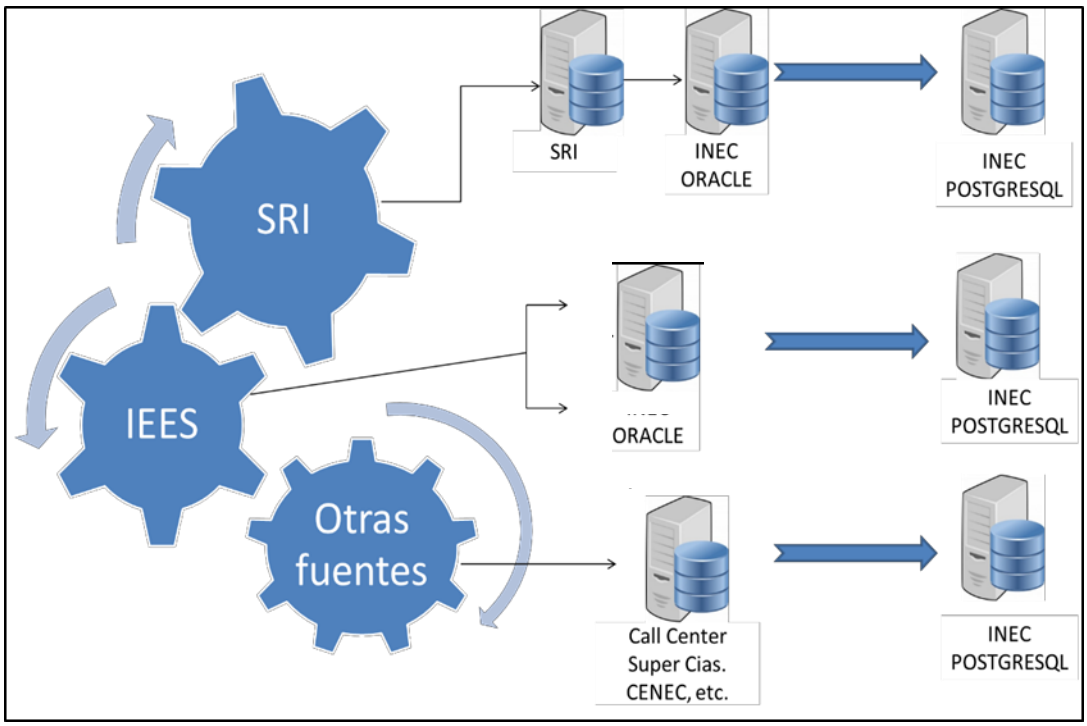


Gráfico N: 2.

Dependiendo de las herramientas usadas para la trasmisión de información generará inconvenientes en el proceso de captación. Es también importante resaltar que los problemas generados en la captación muchas veces no son identificados en el momento mismo en que este se realiza, sino en la fase de procesamiento.

IESS. Información recibida por el Directorio mensualmente, en archivos Oracle, lo que reduce la probabilidad de error.

SRI. La información llega al Directorio diariamente, mediante herramientas del Oracle (vistas materializadas), por lo que la posibilidad de error es mínima. El problema generado a partir de este modo de transmisión de información requiere de disponer el motor de BDD ORACLE, pero al usar software libre existe limitante de espacio a 5GB.

Call Center. La información es recolectada por las personas que trabajan en el call center a través de un sistema, pero también esta recolección se hace en varios archivos de Excel; al ser de esta manera implica mayor cuidado en la subida de esta información.

Otras fuentes de información. La información es obtenida en archivos de Excel, por lo que no presentan normalización o estandarización de estos datos. La complejidad de subida de información es igualmente alta.

La información captada de las fuentes principales IESS y SRI, es entregada al DIEE directamente al servidor de almacenamiento de la BDD del DIEE, en esquemas específicos para su uso.

# PROCESAMIENTO

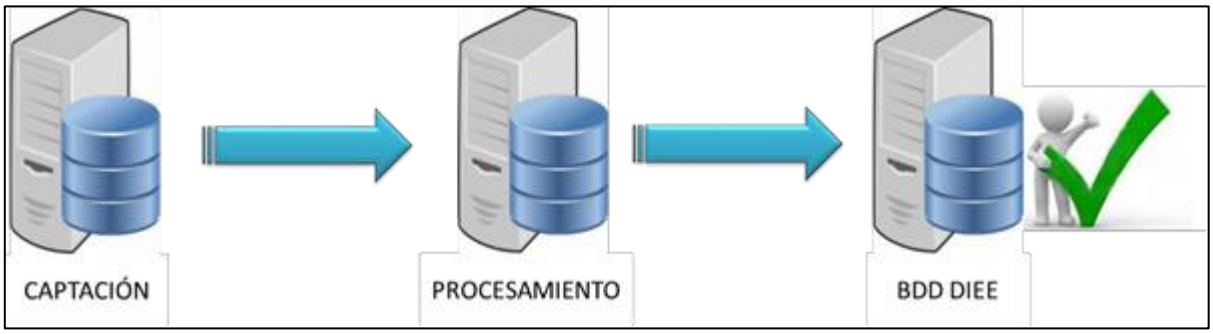


Gráfico N: 3.

Para realizar este proceso es necesario contar con documentos como:

- Matriz de prioridades.
- Matriz de reglas.
- Plan de validación y tabulación.

Estos documentos proporcionan la guía para el desarrollo de los procesos de Extracción de información desde la BDD replicada, Transformación de

acuerdo a la lógica que se maneje internamente en la institución y Carga de datos a la estructura de BDD del Directorio de Empresas.

En los procesos desarrollados en Pentaho, los tres documentos no son plasmados de forma explícita o separada, más si es posible separar los ETL's que gobiernan la migración de cada fuente de información.

Catálogos de información.

En los procesos ETL plasmados en la herramienta de Software Pentaho, existen diferentes grados de dificultad.

Estos a su vez adquieren más complejidad si la variable ha sido revisada por el call center o validada por algún otro medio, ya que en los procesos debe también considerarse actualizaciones solo para registros de menor prioridad de fuente de información.

Actividad Económica.

Una de las variables que presenta mayor complicación es la actividad económica, esto debido a algunas consideraciones:

- Anteriormente el SRI manejaba diferente versión de CIIU, hasta el 2014 se trabajaba con CIIU3, a partir del 2015 ya se ha homologado al CIIU4, sin embargo, existen códigos del SRI no tienen un equivalente directo en CIIU4, debido a que el SRI maneja un formato propio del CIIU4 a 9 dígitos.
- El repositorio en el que reporta el SRI, la información de actividad económica a nivel de Establecimiento, no permite la identificación de una actividad económica principal por cada establecimiento.

Para tratar estos inconvenientes se han creado varias reglas en el DIEE, y muchas de ellas a partir del caso que se ha presentado.

- Se ha definido pasar la actividad económica principal de empresa a establecimiento, para el caso de contribuyentes con establecimientos únicos.
- Se ha creado una matriz de equivalencia a diferentes niveles, para algunos de los códigos del SRI que no se ha encontrado equivalente directo en CIIU4.

- Al no existir un campo que señale la actividad económica por cada uno de los establecimientos, se ha definido un valor ordinal para las actividades económicas y asignando a la primera actividad como la actividad del establecimiento.

Direcciones

La estructura que al momento presenta la BDD del DIEE, adaptada de acuerdo a sus necesidades; es así que tenemos 2 repositorios para este fin, en los cuales se almacena la siguiente información:

- Datos generales de la dirección y se distingue el estado, la fuente, los puntos de georefencia.
- Detalles de la dirección: calle principal, calle secundaria, número, etc.

Estos datos igualmente son validados y actualizados dependiendo de la fuente de información.

Fechas de actualización o cambios.

Se realiza actualización de las variables recibidas del SRI dependiendo de la fecha de actualización y tomando en cuenta la matriz de prioridades. Tomando en cuenta que en contribuyentes es el único que indica la fecha de actualización, esta fecha es tomada para la actualización de establecimientos, es importante mencionar que se desconoce qué variable o variables han sido actualizadas en la fecha proporcionada.

Con la actualización de variables, se actualiza también las variables de control.

Existen también registros con fechas de cierre y/o fechas de cese superior a la fecha de corte, pero estos datos no son considerados, ya que a la fecha de corte estos tienen aún el estado anterior.

Empleados, Remuneraciones y Ventas

Esta información es particular, ya que existe de varios años por cada empresa en el caso de ventas; y de empleados y remuneraciones, existe tanto a nivel de empresas como de establecimientos.

Para el caso de empleados, al reportar la información mensualmente e indicar los datos por cada establecimiento, estos deben ser promediados previo a la subida de información a la BDD de empresa anual. En lo

referente a remuneraciones se sube solo como una sumatoria de lo que cada empresa destina en remuneraciones al año.

En el caso de ventas solo se realiza un filtro de la información de ventas de la última declaración registrada, para subirla cada año.

Medios de comunicación

Los medios de comunicación han recibido tratamiento (transformación) de acuerdo a lo expuesto en el manual de validación y tabulación.

La BDD de medios de comunicación principalmente recibe información del SRI y en mínima proporción del CALL CENTER. En este punto ha sido necesario descartar registros al no acogerse a las reglas que deben cumplir los teléfonos para ser tomados en cuenta.

Registros nuevos

Los registros nuevos obtenidos del corte anual, se ingresan con normalidad, sin recibir actualización alguna; excepto la actualización solicitada bajo demanda (fuente de validación CALL CENTER).

# DETALLES DE PROCESAMIENTO

La herramienta Pentaho es la que juega uno de los roles más importantes en esta fase debido a que aquí se trabaja con procesos ETL's.

ETL: Siglas en inglés que significan: Extraer, Transformar y Cargar, por ello se dice que un ETL es el proceso que permite mover datos desde múltiples fuentes, limpiarlos y cargarlos en otra base de datos, data mart, o data warehouse para apoyar un proceso de negocio.

Job: Es el conjunto de objetos que conforman una transformación.


Gráfico N: 4.

El siguiente gráfico muestra de manera general como se realiza las "transformaciones" en Pentaho. Es necesario indicar que se tienen:

- Job. Puede tener "n" cantidad de transformaciones
- Transformaciones. Puede tener "n" cantidad de scripts y objetos para realizar cambios y adaptaciones de información.
- Scripts. Líneas de código creadas con un propósito especial

A continuación se indica el aspecto que presenta el JOB en la herramienta de software:

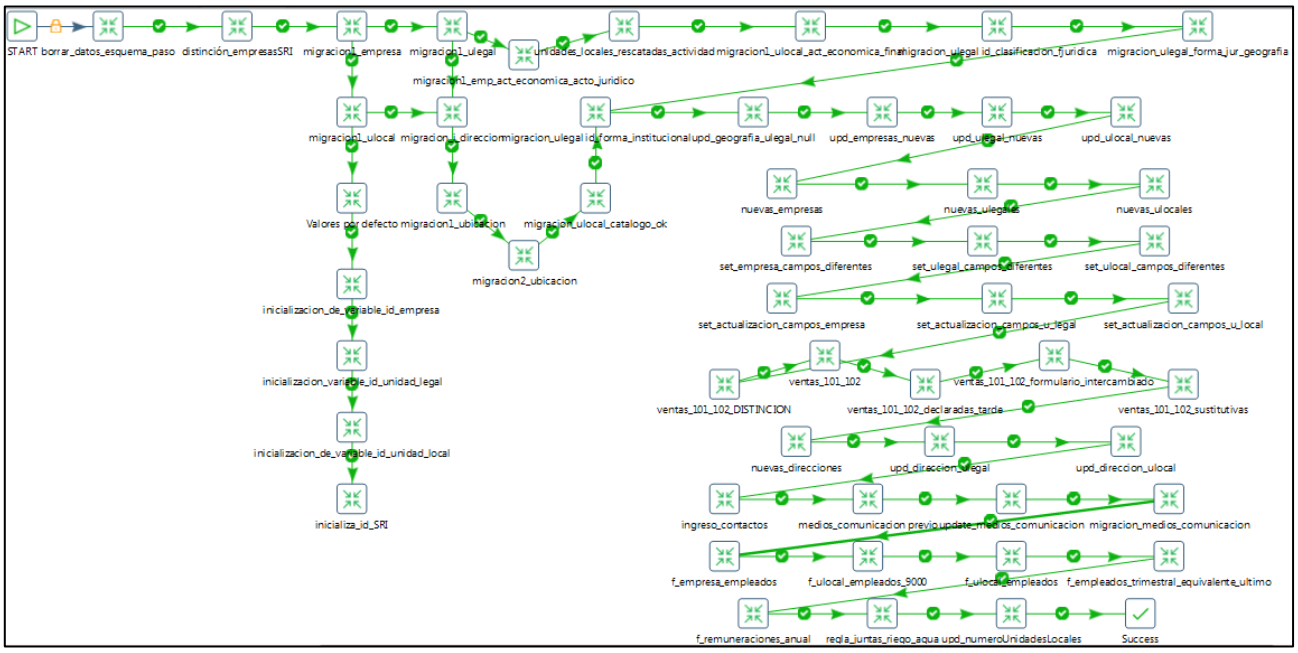


Gráfico N: 5.

En esta sección se irá explicando que se ejecuta en cada transformación y cuál es su finalidad.

Para comenzar a trabajar con las transformaciones, previamente se crea dentro de la base **diee_201401**, un esquema alterno ***paso*** que contiene las tablas principales de DIEE como son f_empresa llamada en "paso" como i_empresa, i_unidad_local, i_unidad_legal, etc, su objetivo es actuar como puente de la información antes de llegar a la base final, ya que existen transformaciones que no se pueden ejecutar directamente en la base final, que se encuentra almacenada en el esquema ***diemp***.

1. **Start.**

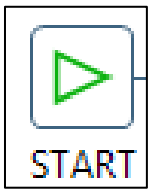


Gráfico N: 6.

El objeto “START” tiene como objetivo darle comienzo a la ejecución de todas las transformaciones.

2. **Borrar_datos_ esquema_paso.**


Gráfico N: 7.

En el DIEE se crea una base de datos alterna PASO donde se va a trabajar y se ejecutarán todos los cambios, esto con la finalidad de pasar a la base de datos oficial ya los datos reales y sin fallos.

Dentro de la transformación (Grafico N: 7) existen los siguientes objetos:

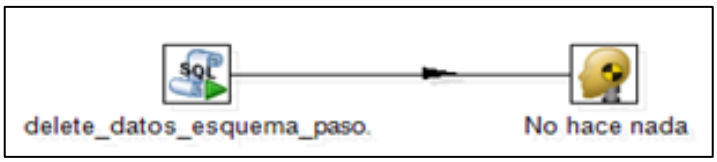


Gráfico N: 8.

El objeto “delete_datos_esquema_paso ejecuta script´s donde borra la información de las tablas:

- paso.i_empresa.
- paso.i_unidad_legal.
- paso.i_unidad_local.
- paso.i_direccion.
- paso.i_ubicacion_direccion.

### 3. Distinción_empresasSRI.


Gráfico N: 9.

Dentro de esta transformación existen los siguientes objetos:

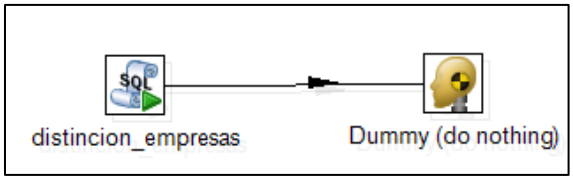


Gráfico N: 10.

Con el objeto "distincion_empresas" se crea una nueva variable llamada directorio_o_tipo, que mediante la asignación de un código nos indicará si una empresa es nueva, antigua, o tuvo un cambio de RUC, esto luego de la comparación entre lo que se tiene en la BDD del DIEE y la nueva información captada del SRI, esta variable nos servirá más adelante dentro del procesamiento a saber cómo proceder dependiendo del código que tenga la misma.

Migración de tablas principales

En la migración de las tablas principales intervienen las transformaciones:

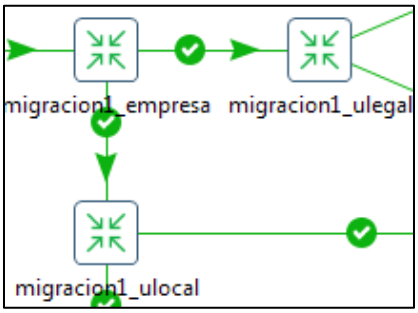


Gráfico N: 11.

Las que se encargan de migrar determinadas variables de las tablas: ruc_contribuyentes y ruc_establecimientos de la fuente SRI a las tablas: i_empresa, i_unidad_local, i_unidad_legal del esquema ***paso***.

## 4. Migración1_empresa.

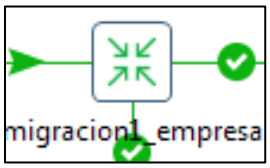


Gráfico N: 12.

Dentro de esta transformación existen los siguientes objetos:

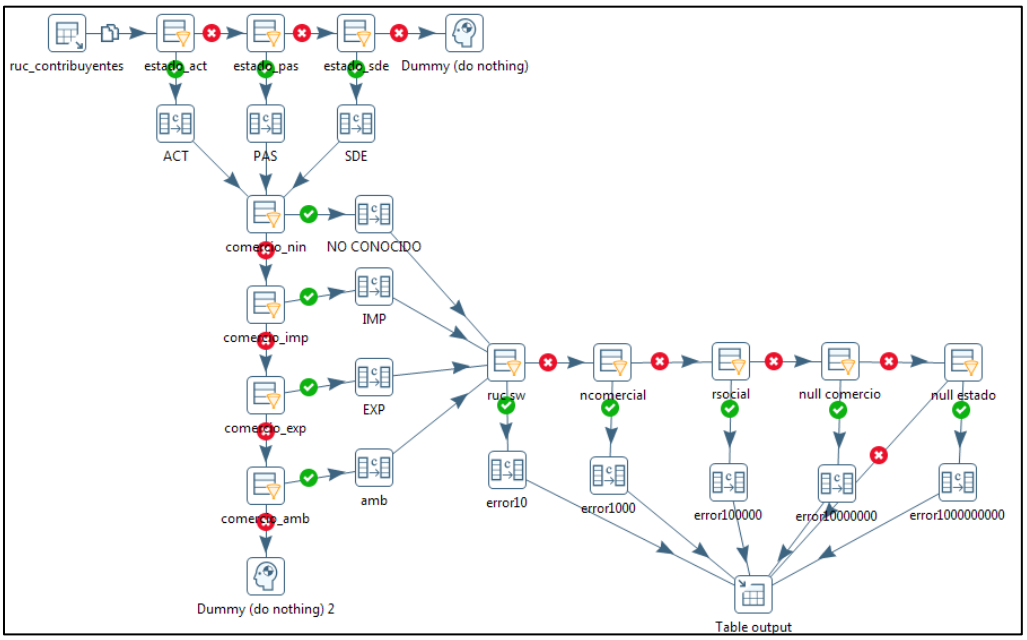


Gráfico N: 13.

Esta transformación tiene como principal objetivo el migrar los datos desde la fuente SRI a partir de la tabla ruc_contribuyentes hacia la tabla i_empresa del esquema ***paso***, con los debidos cambios como por ejemplo:

Se transforma el catálogo de estados de empresa que tiene el SRI al catálogo que tiene el DIEE y de la misma manera se hace para comercio exterior. Quedando de la siguiente manera:

| SRI | DIEE |
|---|---|
| **COMERCIO_EXTERIOR** | ID_ACTIVIDAD_COMERCIO_EXTERIOR |
| **NULL** | 99 |
| **IMP** | 01 |
| **EXP** | 02 |
| **AMB** | 03 |

- Se valida que nombre comercial tenga una longitud de mínimo tres caracteres.
- Cuando las trasformaciones encuentran que hay valores que no existen como por ejemplo en los catálogos o el nombre tiene menos de tres caracteres les cataloga con error a las empresas y se las tiene identificadas, se emite un reporte para su posterior análisis.

## 5. Migración1_ulegal.


Gráfico N: 14.

Dentro de esta transformación existen los siguientes objetos:

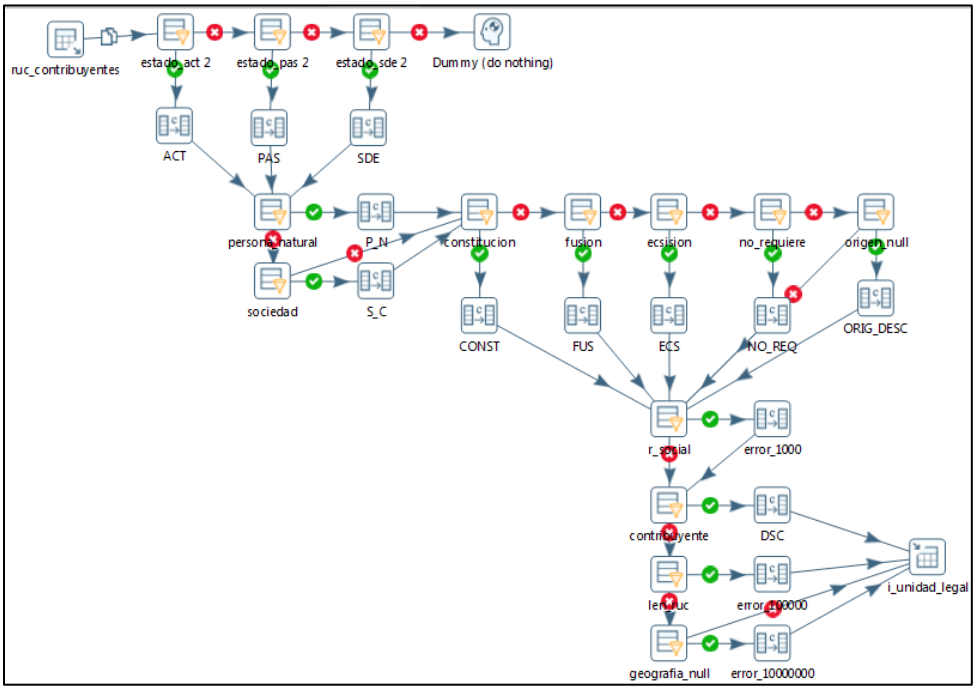


Gráfico N: 15.

Al igual que la migración de empresa esta busca pasar los datos que corresponden a la parte legal que está en la tabla ruc_contribuyentes del SRI, a la tabla i_unidad_legal del esquema ***paso***.

En este proceso ETL tenemos objetos que transforman y validan información:

- Transforma los estados de las unidades legales.
- Transforma el acto jurídico.
- Valida la clase de contribuyente.
- Pasa la información del expediente.
- Valida el largo de la razón social y del ruc.
- De la misma manera se marca a las empresas cuando tengan algún tipo de error y se procederá a reportar cuales son los errores encontrados.

| SRI | | DIEE |
|---|---|---|
| **ESTADO_PERSONA_NATURAL** | ESTADO_SOCIEDAD | ID_UNIDAD_LEGAL_ESTADO |
| **ACT** | ACT | 1 |
| **PAS** | PAS | 2 |
| **SDE** | | 3 |

## 6. Migración1_ulocal.

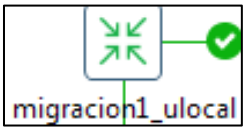


Gráfico N: 16.

La transformación de unidad local tiene los siguientes objetos:

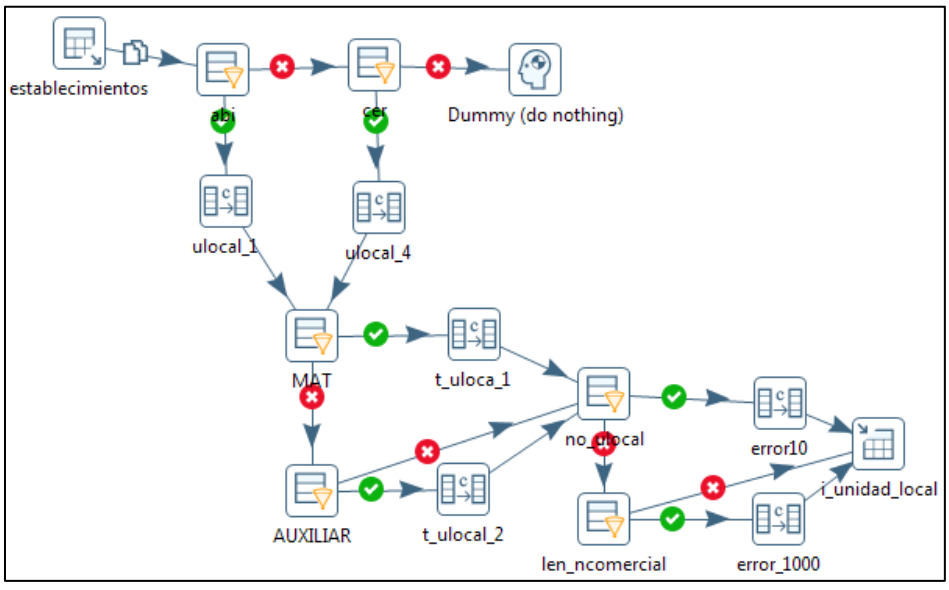

Gráfico N: 17.

Lo que se busca en unidad local es:

- Validar el número de la unidad local.
- Transformar el estado de la unidad local.
- Transformar el tipo de unidad local.
- Validar el largo del nombre comercial.
- De la misma manera los errores que bote la transformación se los procederá a reportar.

## 7. Valores_por_defecto.


Gráfico N: 18.

La transformación de valores por defecto tiene los siguientes objetos:


Gráfico N: 19.

La transformación busca llenar los valores que se han definido en el Plan de Validación y tabulación como los "valores por defecto", es decir:

Por ejemplo si en la variable "Nombre comercial", el valor es nulo, esta transformación automáticamente llena ese campo con el valor "-1", así para variables como: fechas, actividad secundaria, producto elaborado, etc. Todas dependiendo del valor que se haya definido en el Plan de Validación y Tabulación.

Inicialización de variables.

En la inicialización de variables intervienen las transformaciones:

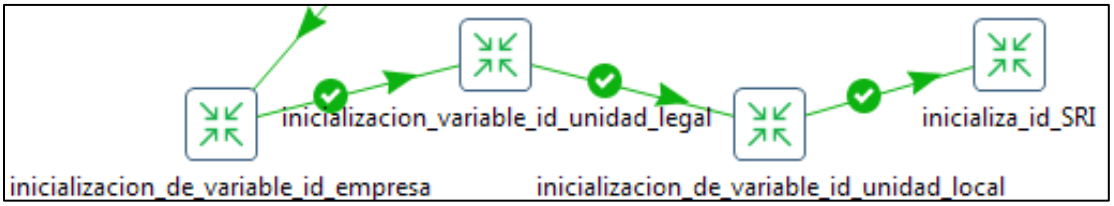


Gráfico N: 20.

Las que se encargan de llenar la información de los códigos (id) de las tablas principales del esquema ***paso*** y de la fuente SRI:

- i_empresa.
- i_unidad_legal.
- i_unidad_local.
- ruc_contribuyentes.
- ruc_establecimientos.

Respectivamente, a partir de los datos de la base del DIEE.

## 8. Inicialización_de_variable_id_empresa.

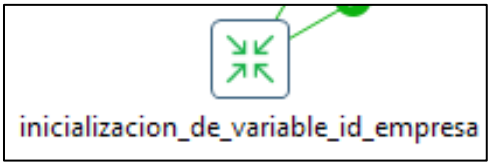


Gráfico N: 21.

La transformación de inicialización de variables id empresa se compone de los siguientes objetos:

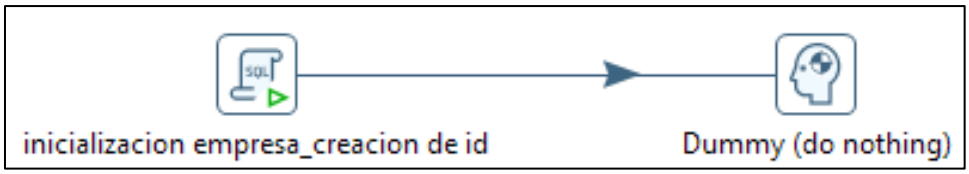


Gráfico N: 22.

Su objetivo principal como lo dice en el nombre es inicializar las variables en el id propio de empresa.

## 9. Inicialización_de_variable_id_unidad_legal.

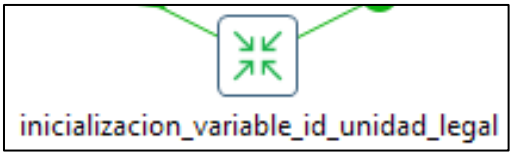


Gráfico N: 23.

La transformación de inicialización de variables id unidad_legal se compone de los siguientes objetos:

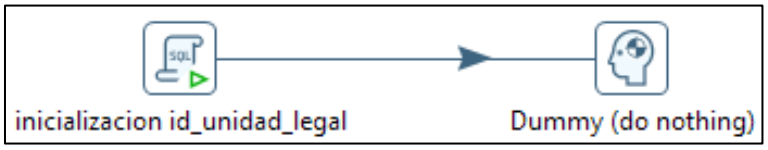


Gráfico N: 24.

Su objetivo principal como lo dice en el nombre es inicializar las variables en el id propio de unidad legal.

## 10. Inicialización_de_variable_id_unidad_local.

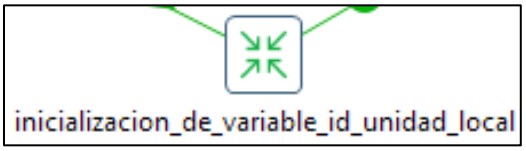


Gráfico N: 25.

La transformación de inicialización de variables id_unidad_local se compone de los siguientes objetos:

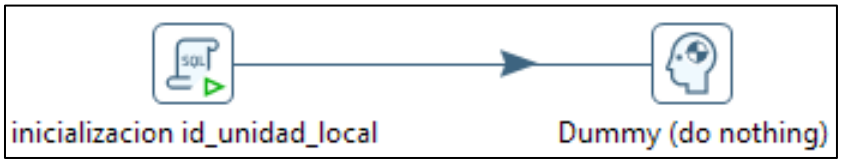


Gráfico N: 26.

Su objetivo principal como lo dice en el nombre es inicializar las variables en el id propio de unidad local.

## 11. Inicializa_id_SRI.


Gráfico N: 27.

La transformación de inicialización de variables id se compone de los siguientes objetos:

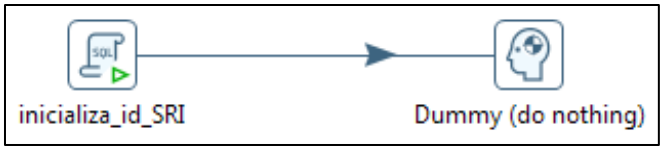


Gráfico N: 28.

Esta transformación tiene la finalidad de llenar los campos de: id_empresa, id_unidad_legal, id_unidad_local que previamente se han creado en las tablas: ruc_contribuyentes y ruc_establecimientos del SRI, el llenado se hace con los datos de las tablas de ***paso***.

Actividad Económica.

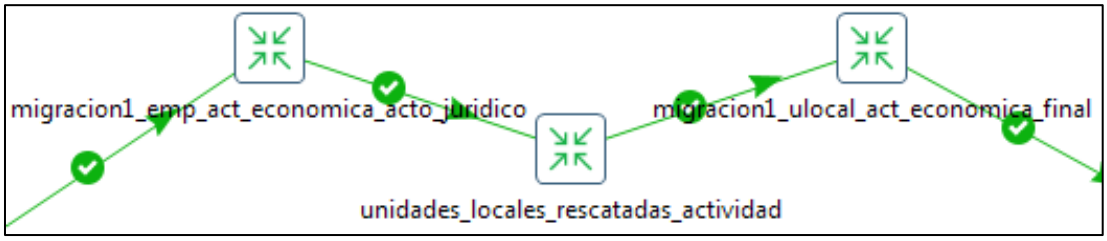


Gráfico N: 29.

Estas transformaciones se encargan de ingresar el id_actividad_economica según las matrices de correspondencias que se tienen en el DIEE.

## 12. Migración1_emp_act_económica_acto_jurídico.


Gráfico N: 30.

La transformación de migracion1_emp_act_economica tiene los siguientes objetos:

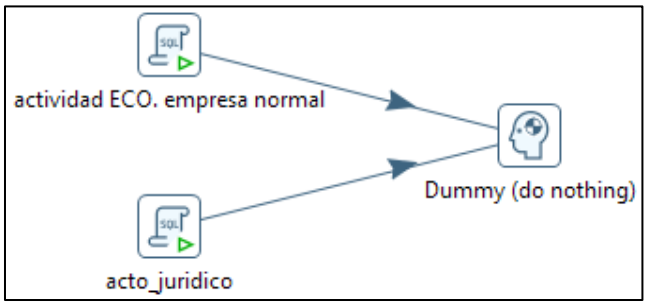


Gráfico N: 31.

La transformación se compone de objetos que ejecutan script's, cada script tiene su objetivo:

- "actividad ECO.empresa normal": se realiza mapeo automático de actividad económica CIIU4, adicionalmente se utiliza la matriz de equivalencia para los códigos sin equivalente directo.
- "Id_acto_juridico": hace una actualización del ruc_acto_juridico cuando es constitución, escisión o fusión.

## 13. Unidades_locales_rescatadas_actividad.


Gráfico N: 32.

La transformación tiene los siguientes objetos:


Gráfico N: 33.

Aquí se busca identificar a los establecimientos únicos, únicos por estado y bajar la actividad de empresa directo al establecimiento.

Se identifican los establecimientos que tienen una sola actividad para poder convertir su actividad directamente con las matrices de conversión.

## 14. Migración1_ulocal_act_económica.


Gráfico N: 34.

La transformación de migracion1_ulocal_act_economica tiene los siguientes objetos:

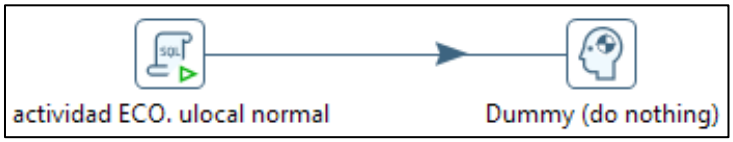


Gráfico N: 35.

Al igual que la anterior transformación este se compone básicamente del mapeo normal de actividades económicas, la aplicación de la matriz propia del DIEE.

Migración de Dirección.

En esta parte intervienen las transformaciones:

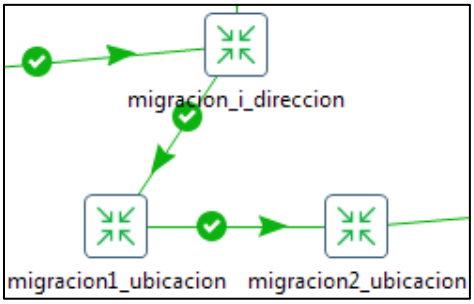


Gráfico N: 36.

## 15. Migración_i_dirección.


Gráfico N: 37.

La transformación de migración_i_direccion se compone de los siguientes objetos.

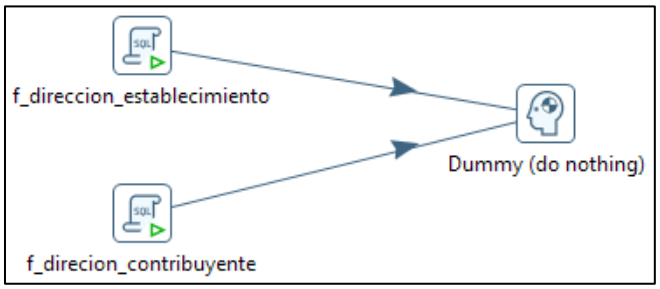


Gráfico N: 38.

Estos tienen la función de extraer los datos desde las tablas de ruc_contribuyentes y ruc_establecimientos del SRI, en lo referente a dirección y se asocia el tipo de vía y el tipo de zona a la cual pertenece la dirección de cada empresa y establecimiento, para llenar para llenar con éstos datos la tabla de i_dirección del esquema ***paso***.

## 16. Migración1_ubicación.

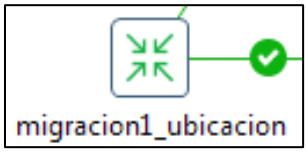


Gráfico N: 39.

La transformación de migración1_ubicacion se compone de los siguientes objetos.

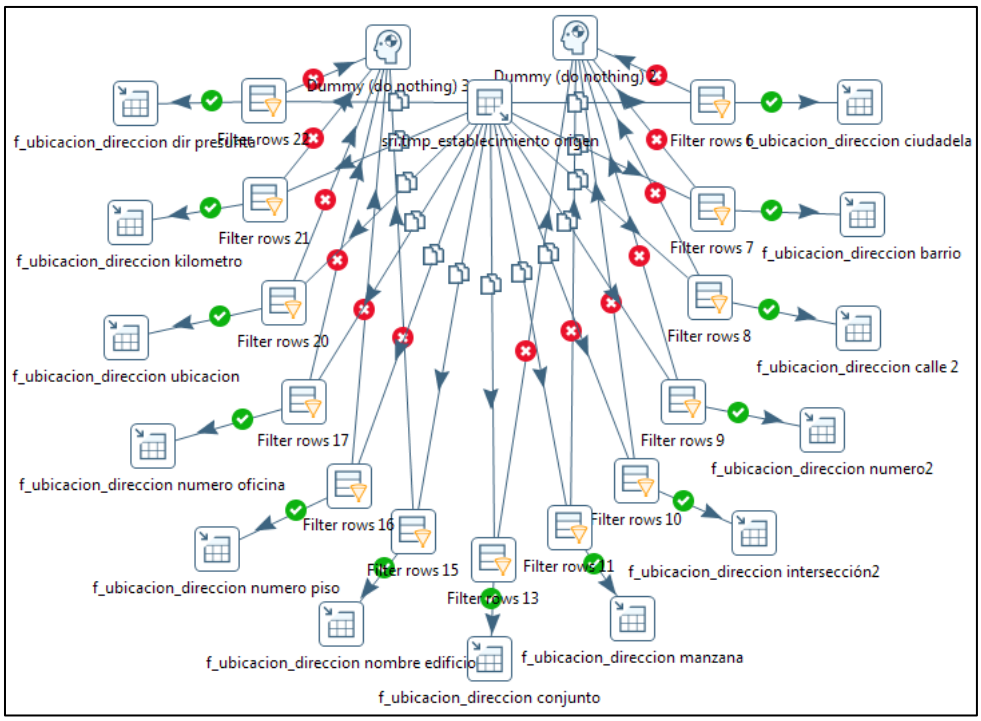


Gráfico N: 40.

## 17. Migración2_ubicación.

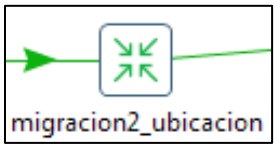


Gráfico N: 41.

La transformación de migración2_ubicacion se compone de los siguientes objetos:

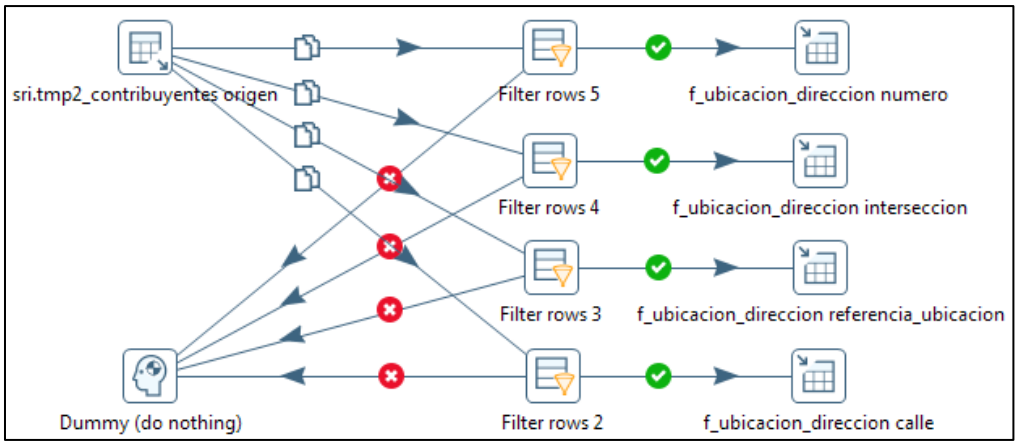


Gráfico N: 42.

Las transformaciones de los gráficos 40 y 42 se compones de varios objetos pero con objetivos en común, pasar la información de las direcciones de contribuyentes a empresa y de establecimientos a unidad local.

La fuente SRI se tiene toda la información referente a direcciones en una sola tabla, en donde se detallan los siguientes datos;

**Para empresas:** calle, número, intersección, referencia_ubicacion.

**Para establecimientos:** barrio, ciudadela, conjunto, bloque, calle, interseccion, nombre_edificio, numero, numero_oficina, manzana, supermanzana, kilometro, carretero, camino, numero_piso, direccion_presunta, referencia_ubicacion.

De los cuales se han agrupado en 12 variables para el mejor manejo de las direcciones, por ello se manejan las siguientes variables: calle_final, numero, interseccion_final, kilometro, conjunto, nombredificio_bloque, numero_piso, numero_oficina, ciudadela, barrio, referencia_ubicacion, manzana_supermanzana, según los nombres se puede apreciar que las variables:

nombredificio_bloque está concatenando la información de nombre_edificio y bloque; manzana_supermanzana, concatena las variables manzana y supermanzana; y las variables: calle_final e interseccion_final se componen adicionalmente de la información de los datos de las variables de camino y carretero, dependiendo de la información que se tenga en estas 4 variables. (Ver documento: ANEXO 3 al Plan de Validación y Tabulación)

A su vez, las variables del SRI anteriormente detalladas, en la base del DIEE son almacenadas en una sola variable, en la tabla f_ubicacion_direccion, con el nombre de: descripción.

## 18. Migración_ulocal_catálogo_ok.

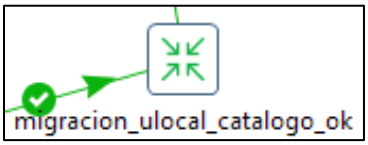


Gráfico N: 43.

La transformación de migración_ulocal_catalogo_ok se compone de los siguientes objetos.

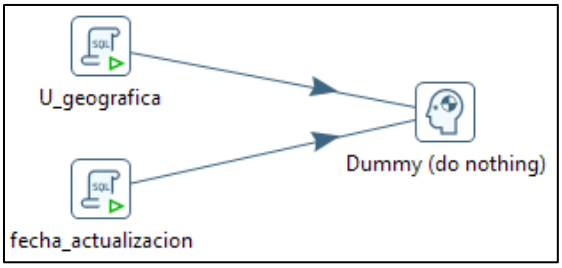


Gráfico N: 44.

Esta transformación tiene dos script que ejecutan lo siguiente:

- "U_geografia": se actualiza el id_geografia para unidad local desde el catálogo de geografia.
- Se actualiza la fecha de actualización, para el respectivo llenado de variables de control.

## 19. Migración1_ulegal_clasificación_fjurídica.


Gráfico N: 45.

La transformación tiene los siguientes objetos:

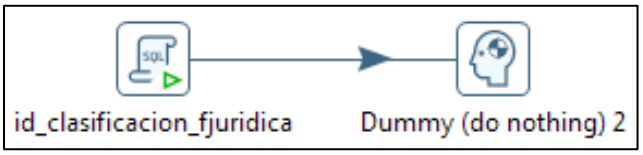


Gráfico N: 46.

Llena el campo "id_clasificacion_fjuridica": asigna el id clasificación de forma jurídica de SRI a la catalogación del DIEE.

## 20. Migración u_legal_forma_jur_geografía.

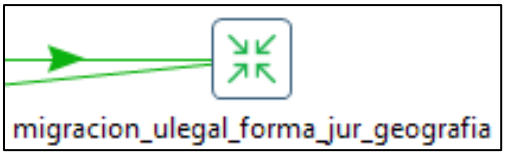


Gráfico N: 46.

La transformación tiene los siguientes objetos:

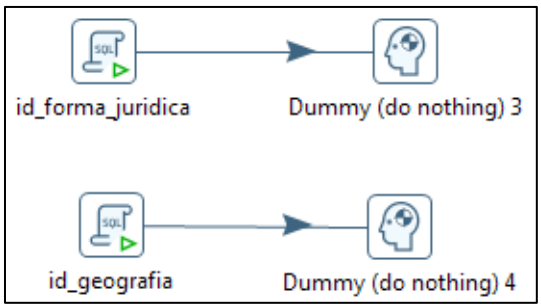


Gráfico N: 47.

Como se puede observar en el Gráfico N: 47 tiene también objetos de ejecución de script´s que se detallan a continuación:

- "id_forma_juridica": asigna el id forma jurídica de SRI a la catalogación del DIEE.
- "id_geografia": asigna el id de geografía de SRI a la catalogación del
- DIEE.

## 21. Migración ulegal_id_forma_institucional.

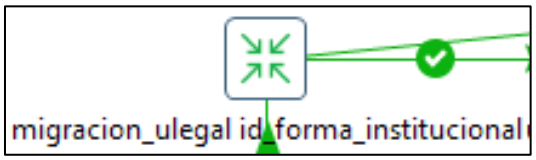


Gráfico N: 48.

La transformación tiene los siguientes objetos:


Gráfico N: 49.

Forma institucional: El propósito de esta variable es obtener información más desagregada de las empresas y establecimientos y obtener mejores resultados en la fase de análisis de la información. La catalogación de la variable en mención es la siguiente:

Edit Data - PostgreSQL 9.2 (x86) (localhost:5432) - diee_201401 - catalogo.d_forma_institucional

File Edit View Tools Help

No limit

| | id_forma_ins [PK] serial | descripcion character varying(100) |
|---|---|---|
| 1 | 1 | Régimen simplificado RISE |
| 2 | 2 | Persona natural no obligada a llevar contabilidad |
| 3 | 3 | Persona natural obligada a llevar contabilidad |
| 4 | 4 | Sociedad con fines de lucro |
| 5 | 5 | Sociedad sin fines de lucro |
| 6 | 6 | Empresa Pública |
| 7 | 7 | Institución Pública |
| 8 | 8 | Economía Popular y Solidaria |

Gráfico N: 50.

Actualización de Empresas o Establecimientos nuevos.

En esta parte intervienen las transformaciones:

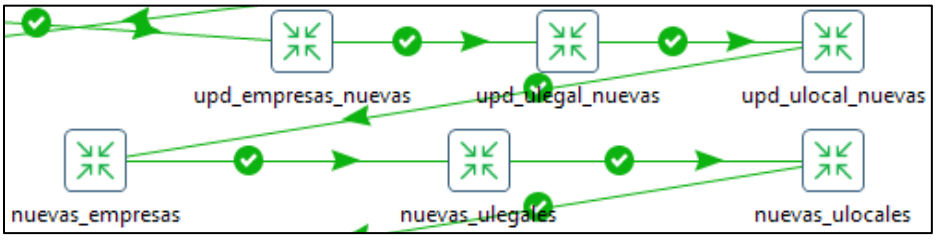


Gráfico N: 51.

En las que se establecen las variables de control tanto para empresas como para establecimientos en las tablas de: i_empresa, i_unidad_legal, i_unidad_local del esquema ***paso***.

## 22. Upd_empresas_nuevas.


Gráfico N: 52.

Actualización de empresas que ingresan al directorio.

La transformación contiene los siguientes objetos que se encargan del siguiente proceso:

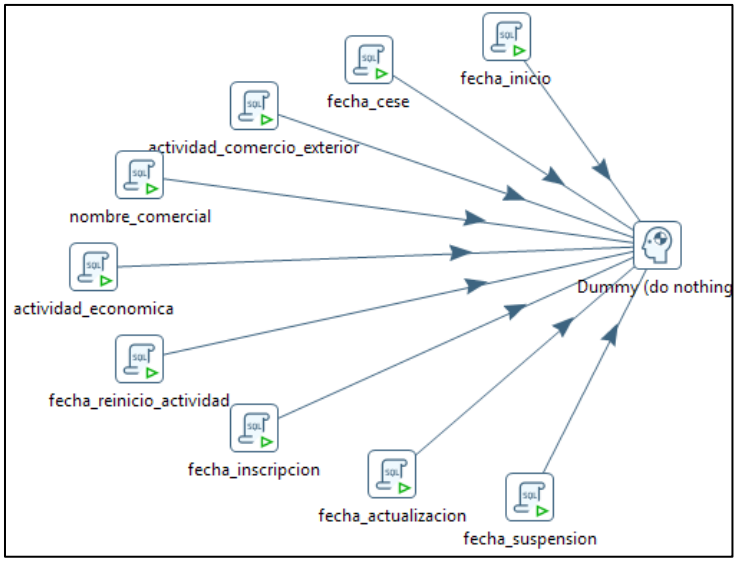


Gráfico N: 53.

En cada objeto se definen las diferentes variables de control, como son: registro, registro_fecha, fuente y fuente_fecha, fecha_desde para las variables de empresas que ingresan al directorio.

## 23. Upd_ulegal_nuevas.


Gráfico N: 54.

La transformación tiene los siguientes objetos:

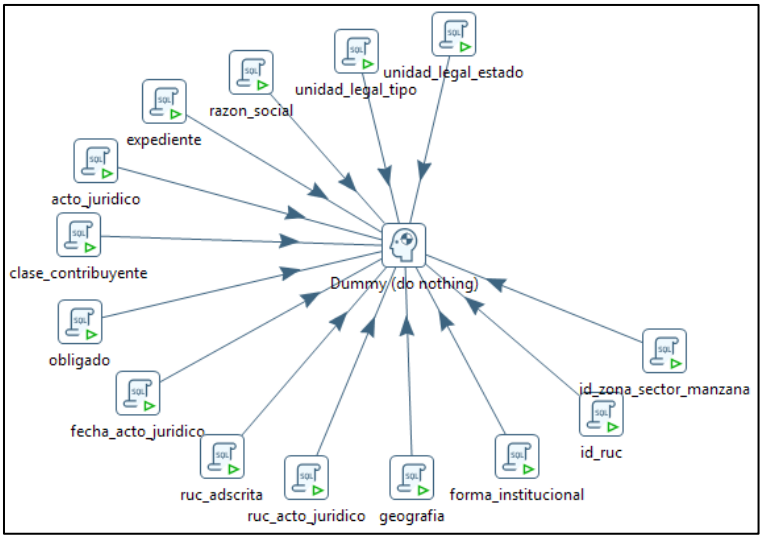


Gráfico N: 55.

Al igual que la anterior transformación se definen las diferentes variables de control de unidad legal que ingresan al directorio.

## 24. Upd_ulocal_nuevas.

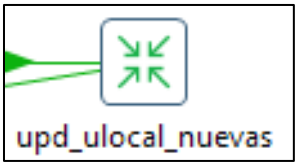


Gráfico N: 57.

La transformación tiene los siguientes objetos:

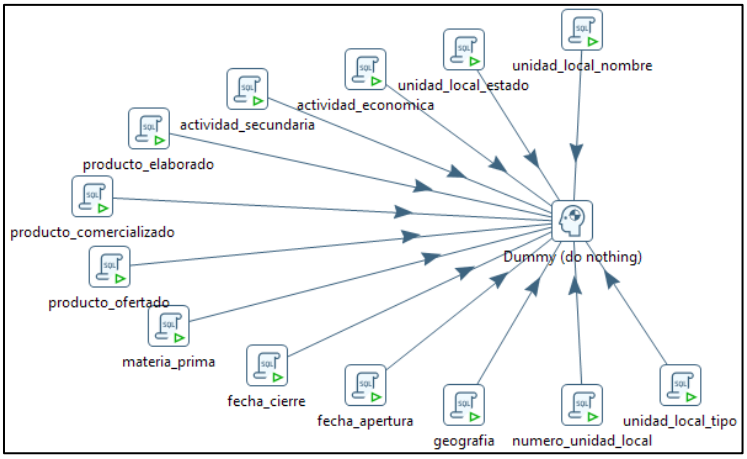


Gráfico N: 58.

En esta transformación se definen las diferentes variables de control de las unidades locales que ingresan al directorio.

## 25. Nuevas_empresas


Gráfico N: 59.

La transformación contiene los siguientes objetos:

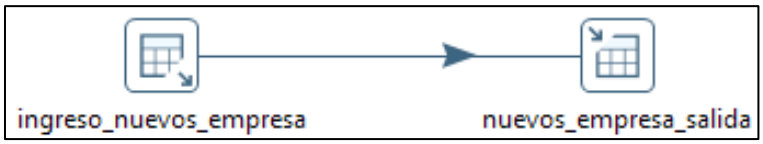


Gráfico N: 60.

Ya que en las 3 anteriores transformaciones de ingresaron las variables de control, aquí se realiza la inserción de las variables de los RUC's nuevos en las tablas principales, en este caso en f_empresa desde la tabla i_empresa.

## 26. Nuevas_ulegales


Gráfico N: 61.

La transformación contiene los siguientes objetos:


Gráfico N: 62.

De la información transferida en la transformación anterior se toma el código creado para empresa (id_empresa), que servirá para asociar a la información a transferir a la tabla f_unidad_legal, esto se realiza para las nuevas empresas a agregarse al directorio.

## 27. Nuevas_ulocales

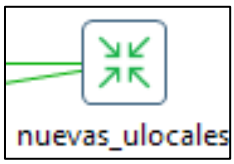


Gráfico N: 63.

La transformación contiene los siguientes objetos:


Gráfico N: 64.

Esta transformación traslada la información únicamente de los nuevos establecimientos a agregarse al directorio, para lo cual, de igual manera que la transformación anterior, se toma el código de la empresa para asociar a la información de cada empresa, para poder pasar la información de los establecimientos a la tabla f_unidad_local.

## 28. Upd_geografía_ulegal_null.


Gráfico N: 65.

La transformación contiene los siguientes objetos:

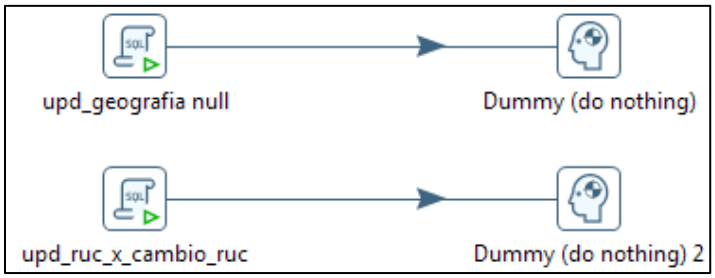


Gráfico N: 66.

Cuando no se tiene dato de geografía en unidad legal, la primera parte de la transformación se encarga de extraer dicha información, a partir de la geografía existente en el establecimiento matriz de la empresa en cuestión.

En la segunda  parte se actualiza en RUC, para los casos en los que existe un cambio de RUC.

Actualización de campos con diferencias.

Para esta actualización intervienen las transformaciones:

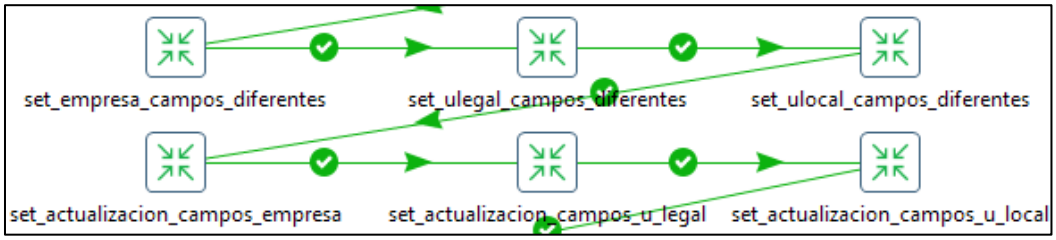


Gráfico N: 67.

En las que se actualiza determinado campo y sus respectivas variables de control, se actualiza con la información de la fuente SRI si los datos son diferentes.

## 29. Set_empresa_campos_diferentes

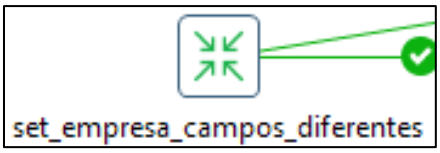


Gráfico N: 68.

La transformación contiene los siguientes objetos:

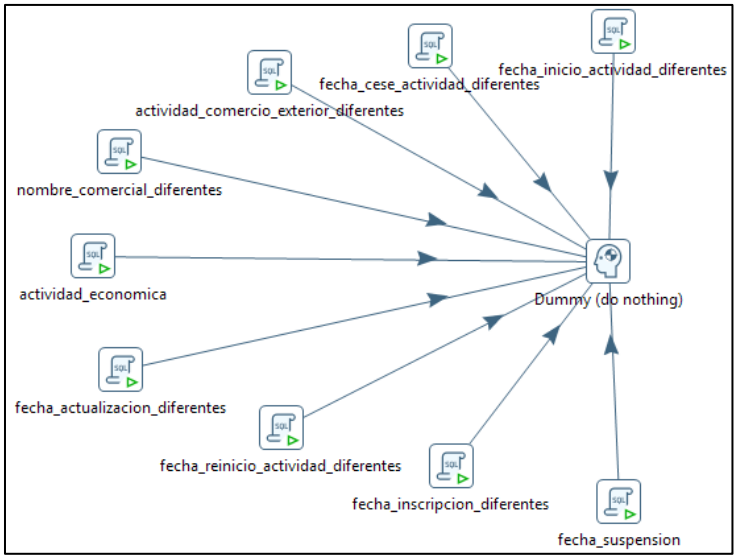


Gráfico N: 69.

Esta transformación actualiza los datos de las variables de control cuando haya existido una actualización de cualquier dato que posea variable de control.

## 30. Set_ulegal_campos_diferentes.

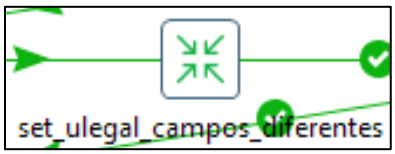


Gráfico N: 70.

La transformación contiene los siguientes objetos:

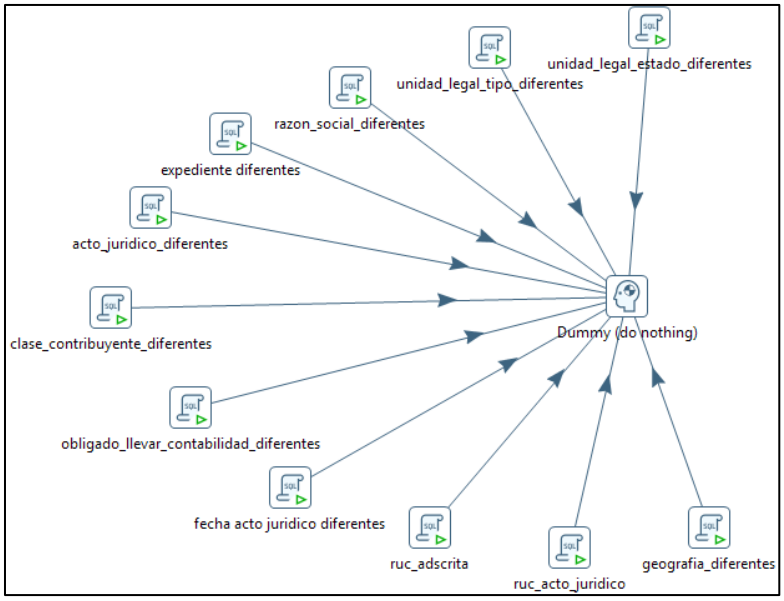


Gráfico N: 71.

Al igual que la anterior transformación esta actualiza los datos de las variables de control cuando haya existido una actualización que provenga de la fuente SRI.

## 31. Set_ulocal_campos_diferentes.


Gráfico N: 72.

La transformación contiene los siguientes objetos:

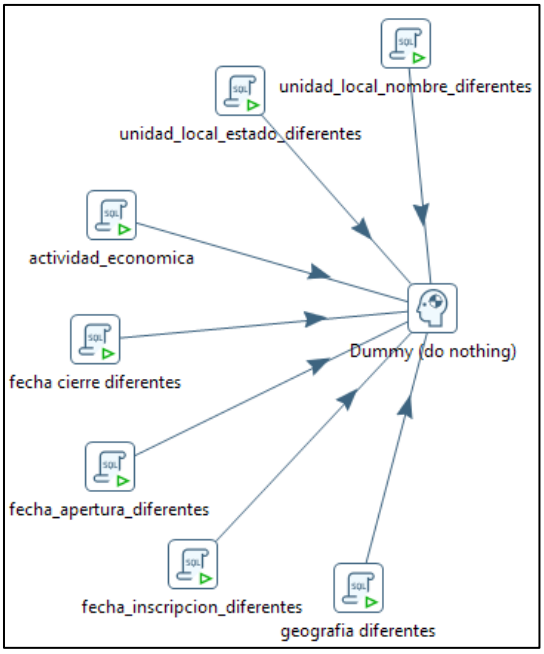


Gráfico N: 73.

Al igual que la anterior transformación esta actualiza los datos de unidad local en sus variables de control cuando haya existido una actualización que provenga de la fuente SRI.

## 32. Set_actualización_campos_empresa.


Gráfico N: 74.

La transformación contiene los siguientes objetos:

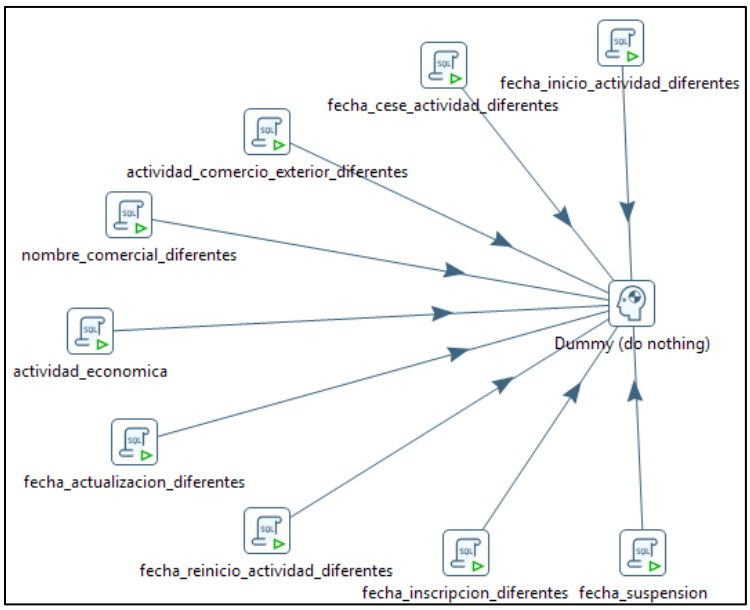


Gráfico N: 75.

Esta transformación actualiza los datos en la tabla de empresa cuando haya existido una actualización, para los casos que pertenecen a la fuente SRI.

## 33. Set_actualización_campos_ulegal.

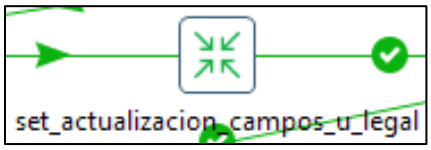


Gráfico N: 76.

La transformación contiene los siguientes objetos:

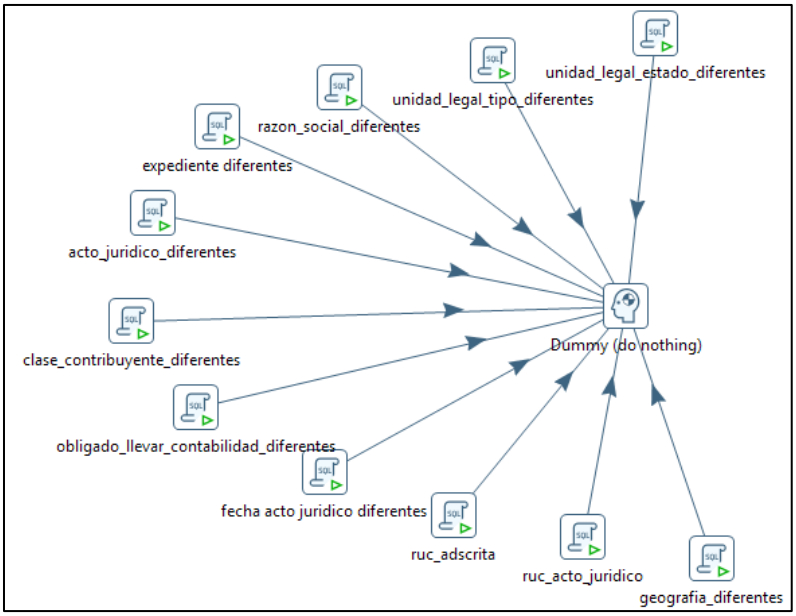


Gráfico N: 77.

Al igual que la anterior transformación esta actualiza los datos en la tabla de unidad_legal cuando haya existido una actualización, para los casos que pertenecen a la fuente SRI.

## 34. Set_actualización_campos_ulocal.


Gráfico N: 78.

La transformación contiene los siguientes objetos:

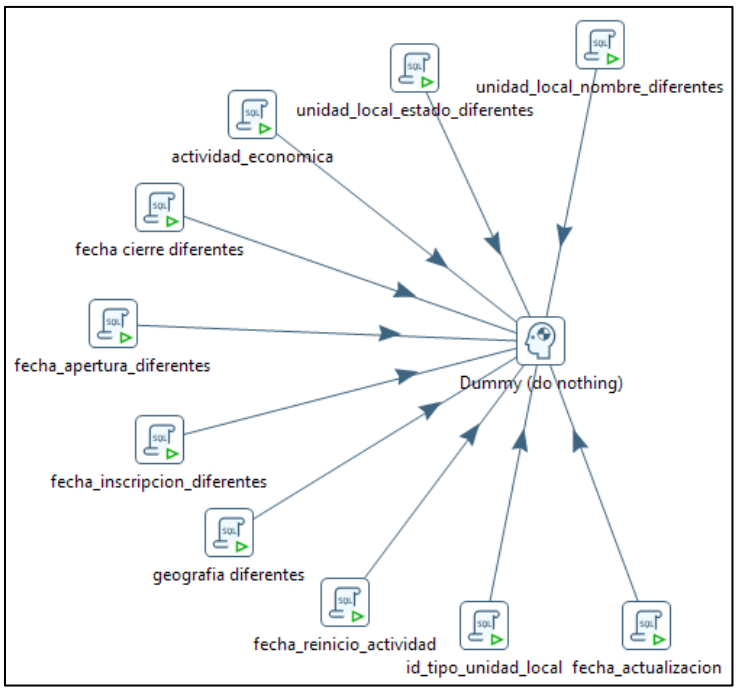


Gráfico N: 79.

De la misma manera la transformación actualiza los datos en la tabla de unidad_local cuando haya existido una actualización, para los casos que pertenecen a la fuente SRI.

Ventas

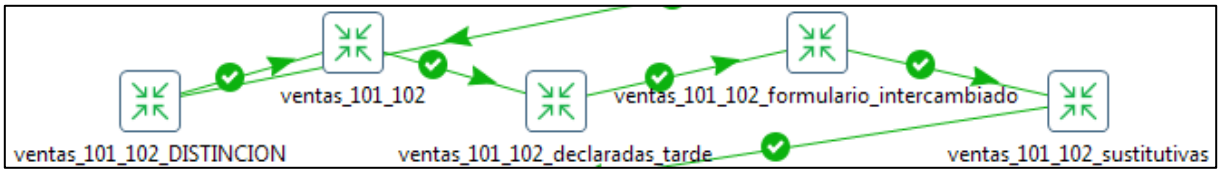


Gráfico N: 80.

En esta sección se trabaja sobre las ventas que reporta el SRI

## 35. Ventas_101_102_distinción.

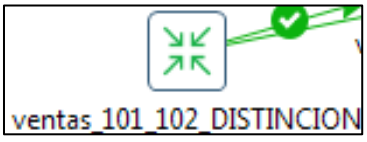


Gráfico N: 81.

La transformación contiene los siguientes objetos:

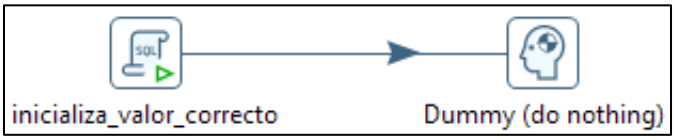


Gráfico N: 82.

Esta transformación marca que empresas deben estar en el formulario 101 y las que deben estar en el 102, para que al momento de pasar la información al DIEE no exista duplicidad de datos.

## 36. Ventas_101_102.

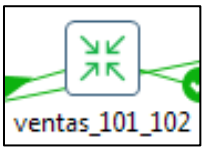


Gráfico N: 83.

La transformación contiene los siguientes objetos:

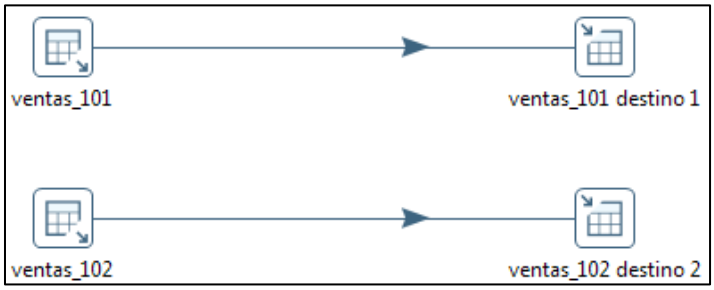


Gráfico N: 84.

En primera instancia, la información proporcionada por el SRI llega de la siguiente manera:

| | | | | |
|---|---|---|---|---|
| 26 | UTILIDAD_EJE_PATRIMONIO_1760 | UTILIDAD DEL EJERCICIO PATRIMONIO | 517 | 517 |
| 27 | PERDIDA_EJE_PATRIMONIO_1770 | PERDIDA DEL EJERCICIO PATRIMONIO | 519 | 519 |
| 28 | VLN_EAF_TDC_1800 | VENTAS NETAS LOCALES EXCLUYE ACTIVOS FIJOS TARIFA DIFERENTE DE CERO | 601 | 601 |
| 29 | VLN_EAF_TCE_1810 | VENTAS NETAS LOCALES EXCLUYE ACTIVOS FIJOS TARIFA CERO | 602 | 602 |
| 30 | EXPORTACIONES_NETAS_1820 | EXPORTACIONES NETAS | 603 | 603 |
| 31 | OTR_RENTAS_EXENTAS_100_3460 | OTRAS RENTAS GRAVADAS | 606 | 606 |
| 32 | UTILIDAD_VTA_ACT_FIJOS_1860 | UTILIDAD VENTA ACTIVOS FIJOS | 607 | 607 |
| 33 | DIV_PERCIBIDOS_LOCALES_1870 | DIVIDENDOS PERCIBIDOS LOCALES | 608 | 608 |
| 34 | VENTA_NETA_ACTIVOS_FIJOS_1940 | VENTA NETA ACTIVOS FIJOS | 691 | 691 |
| 35 | CTO_IVI_MATERIA_PRIMA_2010 | COSTO INVENTARIO INICIAL MATERIA PRIMA | 706 | 706 |
| 36 | CTO_CLN_MATERIA_PRIMA_2020 | COSTO COMPRAS LOCALES NETAS MATERIA PRIMA | 707 | 707 |

Gráfico N: 85.

Los campos que se utiliza para nutrir a la base del DIEE son los que están subrayados con colores: amarillo y lila, los últimos 4 han sido incrementados el último año, al directorio. La información que se utiliza para extraer los registros de ventas que el SRI proporciona, corresponde a las siguientes tablas:

- owb_mvw_ine_anexo3_estruc_f101
- owb_mvw_ine_anexo3_estruc_f102

Que refieren a la información del formulario 101 y 102 respectivamente, donde el formulario 101 contiene las ventas de Personas Jurídicas y el 102 las ventas de las Personas Naturales.

Para la extracción de esta información, la transformación mostrada en el Gráfico N:84 se encarga de pasar de las dos tablas mencionadas de la fuente, la información de ventas tanto del formulario 101 como del 102 por cada año registrado, a la tabla f_empresa_ventas.

## 37. Ventas_101_102_declaradas_tarde

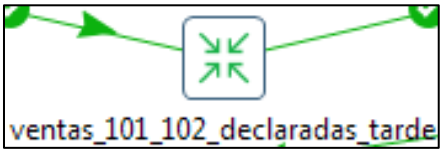


Gráfico N: 86.

La transformación contiene los siguientes objetos:

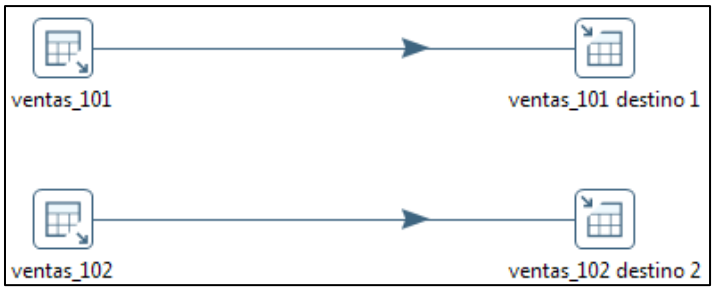


Gráfico N: 87.

Se incorporan todas las declaraciones tardías, tanto del formulario 101 como del 102.

## 38. Ventas_101_102_formulario_intercambiado

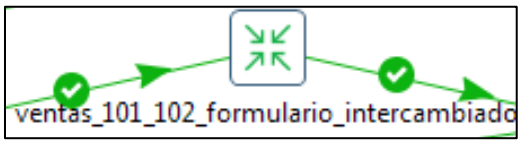


Gráfico N: 88.

La transformación contiene los siguientes objetos:

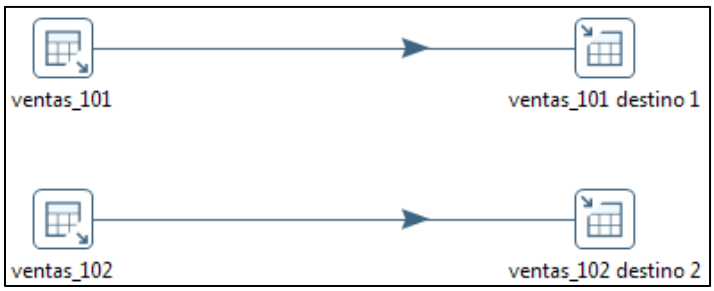


Gráfico N: 89.

Se incorporan todas las declaraciones que se han registrado en un formulario incorrecto, es decir: que siendo persona natural ha declarado en el formulario 101 o que siendo persona jurídica ha declarado en el formulario 102, toda esta información es recuperada.

## 39. Ventas_101_102_sustitutivas

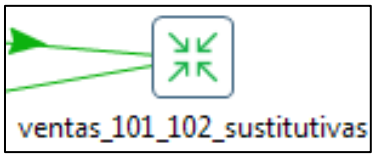


Gráfico N: 90.

La transformación contiene los siguientes objetos:

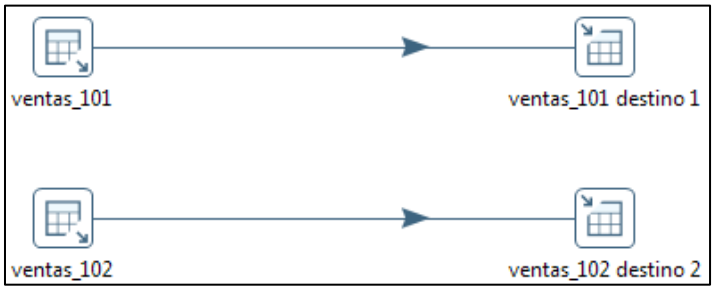


Gráfico N: 91.

Se actualizan las declaraciones que tienen diferente valor de total de ventas declaradas, comparando entre lo que se tiene en la BDD DIEE y la nueva información proporcionada por el SRI, lo que implica que se ha realizado una declaración sustitutiva.

Nuevas direcciones

En esta parte están las transformaciones:


Gráfico N: 92.

Donde se realiza una migración de la información de las tablas de dirección del esquema ***paso*** a las tablas definitivas de la base del DIEE en el esquema ***diemp***, solamente de los registros nuevos referentes a dirección.

## 40. Nuevas_direcciones.

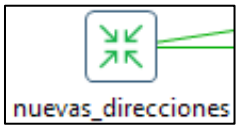


Gráfico N: 93.

La transformación contiene los siguientes objetos:

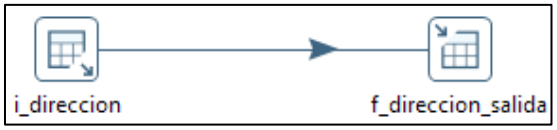


Gráfico N: 94.

Los objetos del Gráfico N: 94, buscan pasar la información de las nuevas direcciones almacenadas en la tabla de i_direccion del esquema ***paso*** a la tabla f_direccion en la base del DIEE en el esquema ***diemp***.

## 41. Upd_dirección_ulegal.

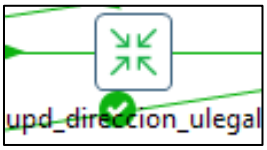


Gráfico N: 95.

La transformación contiene los siguientes objetos:

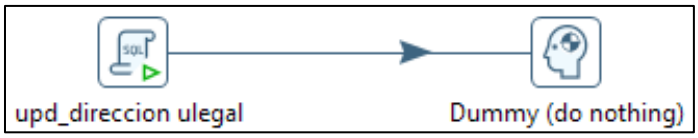


Gráfico N: 96.

En esta transformación se identifican las empresas que hayan cambiado de dirección, lo cual se determina mediante cambios en: calle, número e intersección, cuando existen estos cambios se analiza si la fecha de actualización de la fuente SRI es mayor a la última actualización de la dirección en la base del DIEE, si se tiene este caso se procede a comparar entre la base actual y la del año pasado del SRI para analizar si han existido cambios entre ambos años, caso en el que se procederá a actualizar con estado de "0" las direcciones pertenecientes a las empresas que tuvieron cambios, luego de esto se procede a insertar las nuevas direcciones provenientes del SRI, éstas se registran con estado "1".

## 42. Upd_direccion_ulocal.


Gráfico N: 97.

La transformación contiene los siguientes objetos:

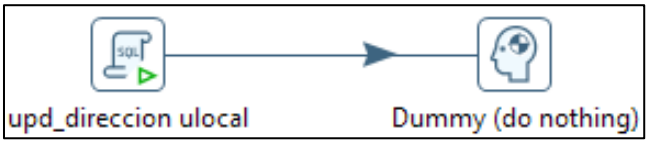


Gráfico N: 98.

Esta transformación realiza el mismo proceso que la transformación upd_direccion_ulegal pero lo realiza para las direcciones de los establecimientos.

Esta transformación transfiere toda la información de la tabla de i_empresa del esquema ***paso***, de las empresas que ingresan al directorio, a la tabla de f_empresa de la base del DIEE en el esquema ***diemp***.

## 43. Ingreso_contactos


Gráfico N: 99.

La transformación contiene los siguientes objetos:

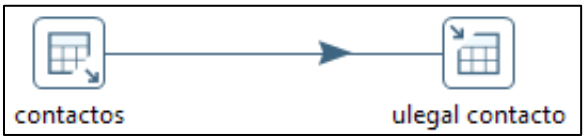


Gráfico N: 100.

## 44. Migración_medios_comunicación previo.


Gráfico N: 101.

La transformación contiene los siguientes objetos:

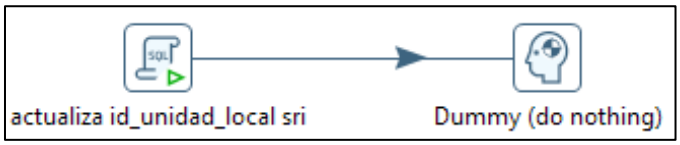


Gráfico N: 102.

El Gráfico N: 102 se encarga de actualizar el código de unidad local (id_unidad_local) de contactos de la fuente SRI con el código de la tabla

f_unidad_local de la base del DIEE, para ser utilizados en la siguiente transformación.

## 45. Update_medios_comunicación


Gráfico N: 103.

La transformación contiene los siguientes objetos


Gráfico N: 104.

En esta transformación se analiza los medios de comunicación que han cambiado y se actualizan si: la fecha de la última actualización es menor que la fecha de actualización del dato del SRI por el cual se va a actualizar la información, se procede a cambiar a estado "0" a los datos anteriores, para no tener duplicidad en la información, y finalmente se procede a agregar la nueva información a la tabla de *f_medio_comunicacion.*

## 46. Migración_medios_comunicación


Gráfico N: 105.

La transformación contiene los siguientes objetos:


Gráfico N: 106.

Estos objetos se encargan de insertar los nuevos medios de comunicación que se han encontrado en la base de la fuente SRI, en la tabla principal de *f_medio_comunicacion*.

## 47. F_empresa_empleados


Gráfico N: 107.

La transformación contiene los siguientes objetos:

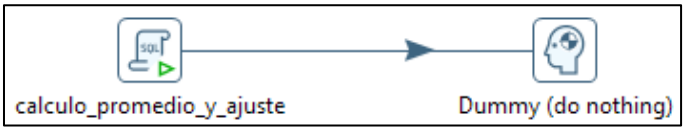


Gráfico N: 108.

En la primera parte de este objeto se realiza la sumatoria de los empleados de todos los establecimientos de la empresa de cada mes, para empleados hombres, mujeres y el total de la suma entre ambas variables, en la segunda parte se obtiene el promedio anual de todos los meses donde se registró empleados, esto como información a cargar en la tabla de f_empresa_afiliados_anual del esquema ***diemp***, por último se realiza un ajuste a los empleados hombres y mujeres, ya que como producto del redondeo en el promedio existen casos con diferencias entre la suma de hombres más mujeres con respecto al total.

## 48. F_ulocal_empleados_9000


Gráfico N: 109.

La transformación contiene los siguientes objetos:

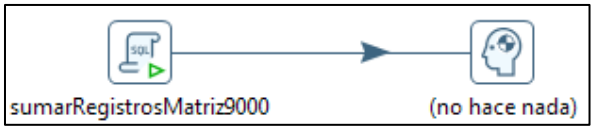


Gráfico N: 110.

En el Gráfico N: 110 se ejecutan una serie de scripts, donde inicialmente se encera la variable id_tipo_unidad_local, para ser nuevamente llenado con la información de la tabla f_unidad_local de la base del DIEE, con el fin de almacenar toda la información que no pertenece a ninguna unidad local en una tabla temporal. Estos datos vienen de la fuente del IESS generalmente con número de unidad local superior o igual a 9000, los empleados que reportan estos casos, primero se enlaza entre la geografía que tiene el IESS y la que tiene la BDD del DIEE, y se recuperan los casos coincidentes, los restantes son sumados a los empleados afiliados de la matriz de la empresa respectiva.

Si se presenta el caso que no existe matriz en la información proporcionada por el IESS, se suma esta la información de establecimientos, con unidad local de 9000, al establecimiento con mayor número de empleados.

## 49. F_ulocal_empleados


Gráfico N: 111.

La transformación contiene los siguientes objetos:

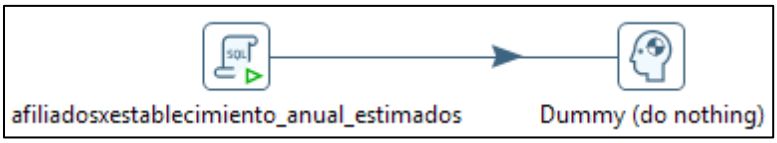


Gráfico N: 112.

Aquí se obtienen empleados por establecimiento estimados por año, esto con respecto al total obtenido del promedio anual y se distribuye según el porcentaje de afiliados que haya acumulado en el año cada establecimiento.

## 50. F_empleados_trimestrales_equivalente_último

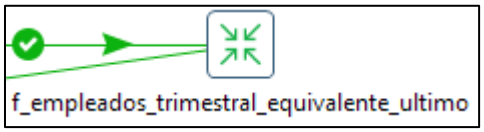


Gráfico N: 113.

La transformación contiene los siguientes objetos:

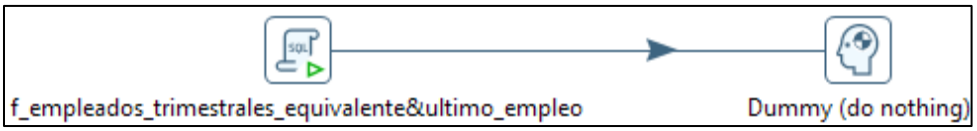


Gráfico N: 114.

El script del Gráfico N: 114 se encarga de realizar un promedio *trimestral* de los empleados tanto de hombres como mujeres para obtener el total de empleados por trimestre de cada empresa. Se calcula también el *empleo equivalente* que es el total de empleados reportados en el año divididos para 12 y finalmente se obtiene el cálculo de último empleo, que son los empleados registrados en el mes de noviembre o caso contrario del último mes registrado.

## 51. Remureraciones_anual

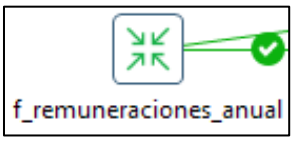


Gráfico N: 115.

La transformación contiene los siguientes objetos:


Gráfico N: 116.

En el script del Gráfico 116 se realiza la sumatoria de los remuneraciones de hombres y mujeres obteniéndose con éstos la remuneración total, que han sido registrados en el IESS en todos los meses para todos los empleados de cada empresa, obteniéndose así las remuneraciones

acumuladas en todo el año, sin realizar ningún tipo de promedio, como si se lo realiza para empleados.

## 52. Upd_númeroUnidadesLocales.

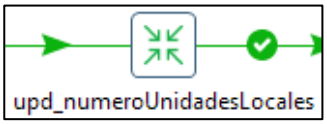


Gráfico N: 117.

La transformación contiene los siguientes objetos:

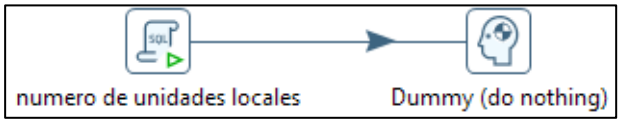


Gráfico N: 118.

En el Gráfico N: 118 se ejecuta el script que realiza un conteo del número de unidades locales activas, para con esta información llenar el campo de numero_unidades_locales en la tabla f_empresa de la base del DIEE, este valor corresponde al número total de establecimientos abiertos que tiene cada empresa.

## 53. Regla_juntas_riego_agua

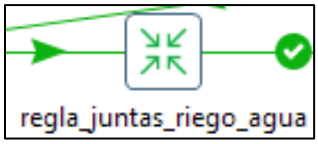


Gráfico N: 119.

La transformación contiene los siguientes objetos:


Gráfico N: 120.

Con este script se corrige la información de la BDD DIEE por un elemento de Matriz de Reglas, en la cual se cambia la actividad económica de la tabla f_empresa y la forma institucional de la tabla f_unidad_legal, de las empresas que tengan parámetros de Junta de Agua o Riego dentro de la razón social, tal como se detalla en la matriz de Reglas.

## 54. Clasificacion Empleados Ventas


Gráfico N: 121.

La transformación contiene los siguientes objetos:

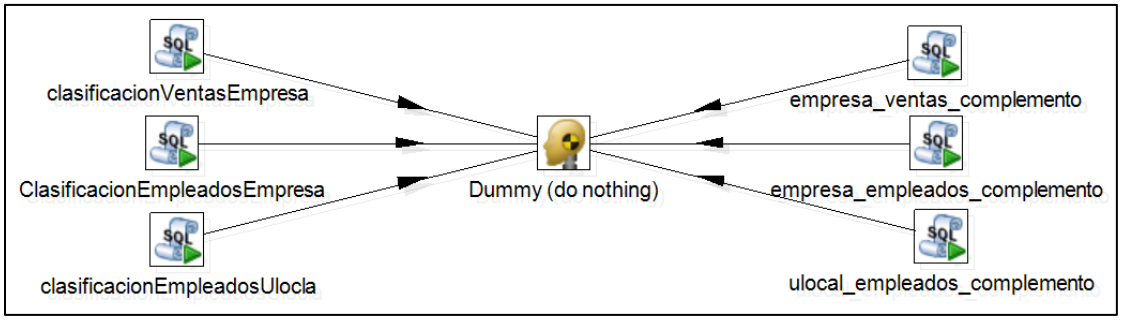


Gráfico N: 122.

En esta transformación se asignan los estratos para empleados y ventas en las empresas según los catálogos de clase de ventas y empleados, que se muestran en las siguientes tablas:

Estratos de ventas:

| codigo | estrato | valor_inferior | valor_superior |
|---|---|---|---|
| 1 | ESTRATO I | 0 | 100000 |
| 2 | ESTRATO II | 100000 | 1000000 |
| 3 | ESTRATO III | 1000000 | 2000000 |
| 4 | ESTRATO IV | 2000000 | 5000000 |
| 5 | ESTRATO V | 5000000 | 99999999999 |

Para el Estrato I se considera a todos los casos que tengan forma institucional diferente de 7 (Institución Pública), la clase contribuyente sea igual a RISE y el personal afiliado esté entre 1 y 10.

Estratos de empleados:

| codigo | estrato | valor_mímimo | valor_máximo |
|---|---|---|---|
| 1 | ESTRATO I | 1 | 9 |
| 2 | ESTRATO II | 10 | 49 |
| 3 | ESTRATO III | 50 | 99 |

| | | | |
|---|---|---|---|
| 4 | ESTRATO IV | 100 | 199 |
| 5 | ESTRATO V | 200 | 10000000 |

## 55. Clasificación Tamaño Empresa

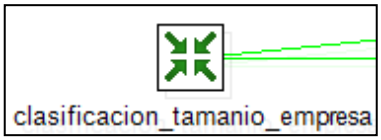


Gráfico N: 123.

La transformación contiene los siguientes objetos

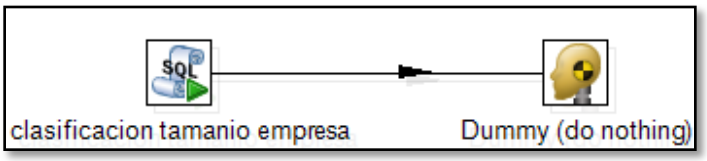


Gráfico N: 124.

Con esta transformación obtenemos la variable tamaño, esta variable se determina en base a los estratos de ventas y empleados.

## 56. Clasificación empresas activas estratificadas


Gráfico N: 125.

La transformación contiene los siguientes objetos

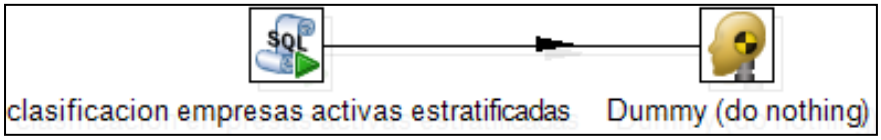


Gráfico N: 126.

En esta trasformación se obtiene la variable situación, que es un código para identificar a todas las empresas activas y que poseen estratos.

# CONTEOS Y VALIDACIONES.

Para verificar y validar la información que se ha obtenido a partir del procesamiento se procede con conteos establecidos que se tienen en el Plan de Validación y Tabulación del DIEE, si estos conteos están correctos se puede continuar con el congelamiento de la base de datos, caso contrario se analiza cual es el error y se hace un reprocesamiento de la base de datos con la finalidad de corregir el error.

Dentro de los conteos y validaciones se realizaron las siguientes actividades:

- Se realizó el cálculo de estratos de ventas y personal afiliado, así como el tamaño y situación, para los años del 2009 al 2017.
- Se mantienen para este año las tablas de empleo equivalente y último empleo donde se obtiene información más acertada de empleados.
- Para las validaciones se emplearon mallas de validación construidas en R, para asegurarnos que la información entregada al personal de análisis sea consistente y no tenga errores.

# CONCLUSIONES.

- El contar con procesos de calidad de datos ayuda a la fase de procesamiento de información, así como a la definición de reglas para el tratamiento de los mismos.
- Los procesos ETL's han sido levantados en algunos puntos de forma paralela a la documentación, esto ha minimizado su facilidad de ser plasmados en la herramienta de Software, por lo que en futuros procesamientos claramente pueden evidenciarse mejoras en los ETL.
- Existen scripts que han sido creados independientemente de los ETL para las fuentes de datos como la superintendencia de bancos y el call center, pero esto puede ser fácilmente automatizado si se genera un sistema apegado a las reglas que rigen la BDD del DIEE.
- Al procesamiento, en cada una de sus fases, se lo ha ido perfeccionando año tras año, con el fin de que los procesos sean automatizados y así obtener cada vez un producto de mejor calidad, siendo esa la meta de este proceso.