

# Manual de Procesamiento

**DIEE 2021**

Noviembre, 2022

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

#### Información del Documento:

<b>Resumen:</b>	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
V.0.7	Incorporación de cálculos para plazas de empleo registrado para el Procesamiento del DIEE 2018 (Carolina Mina)	10 de diciembre de 2019
V.0.8	Incorporación de cálculos y ajustes para las nuevas variables de empleo y plazas de empleo registrado por rangos de edad. (Carolina Mina)	9 de noviembre de 2020
V.0.9	Incorporación de nuevos estratos y tamaños según el Código orgánico de la producción.  Cambio en el cálculo de la forma institucional.  Cambio en el cálculo de ventas totales el	21 de octubre de 2021

	sector financiero y SSFL. (Carolina Mina)	
V.1.0	Agregación de procesos primordiales en el procesamiento del DICE, desde ítem 54 (Carolina Mina)	16 de noviembre de 2022

## Contenido

.....	1
<b>INTRODUCCIÓN</b> .....	6
<b>PROCESO DE CONSTRUCCIÓN DE BASE DE DATOS DEL DIRECTORIO DE EMPRESAS</b> .....	8
<b>PROCESAMIENTO</b> .....	10
<b>DETALLES DE PROCESAMIENTO</b> .....	13
1. Start.....	15
2. Borrar_datos_esquema_paso.....	15
3. Distinción_empresasSRI. ....	16
4. Migración1_empresa.....	17
5. Migración1_ulegal.....	18
6. Migración1_ulocal.....	20
7. Valores_por_defecto. ....	20
8. Inicialización_de_variable_id_empresa.....	22
9. Inicialización_de_variable_id_unidad_legal. ....	22
10. Inicialización_de_variable_id_unidad_local.....	23
11. Inicializa_id_SRI. ....	23
12. Migración1_emp_act_económica_acto_jurídico. ....	24
13. Unidades_locales_rescatadas_actividad. ....	25
14. Migración1_ulocal_act_económica.....	25
15. Migración_i_dirección.....	26
16. Migración1_ubicación.....	27
17. Migración2_ubicación.....	27

18.	Migración_ulocal_catálogo_ok. ....	29
19.	Migración1_ulegal_clasificación_fjurídica. ....	29
20.	Migración u_legal_forma_jur_geografía.....	30
21.	Migración ulegal_id_forma_institucional. ....	30
22.	Upd_empresas_nuevas.....	32
23.	Upd_ulegal_nuevas. ....	32
24.	Upd_ulocal_nuevas. ....	33
25.	Nuevas_empresas .....	34
26.	Nuevas_ulegales.....	34
27.	Nuevas_ulocales.....	35
28.	Upd_geografía_ulegal_null.....	35
29.	Set_empresa_campos_diferentes.....	36
30.	Set_ulegal_campos_diferentes.....	37
31.	Set_ulocal_campos_diferentes.....	38
32.	Set_actualización_campos_empresa. ....	39
33.	Set_actualización_campos_ulegal. ....	40
34.	Set_actualización_campos_ulocal. ....	41
35.	Ventas_101_102_distinción.....	42
36.	Ventas_101_102.....	43
37.	Ventas_101_102_declaradas_tarde .....	44
38.	Ventas_101_102_formulario_intercambiado.....	45
39.	Ventas_101_102_sustitutivas.....	45
40.	Nuevas_direcciones. ....	46
41.	Upd_dirección_ulegal. ....	47
42.	Upd_direccion_ulocal. ....	47
43.	Ingreso_contactos.....	48
44.	Migración_medios_comunicación previo.....	48
45.	Update_medios_comunicación.....	49
46.	Migración_medios_comunicación.....	49
47.	F_empresa_empleados .....	50
48.	F_ulocal_empleados_9000.....	51
49.	F_ulocal_empleados.....	51

50.	F_empleados_trimestral_equivalente_último .....	52
51.	Remuneraciones_anual .....	53
52.	Upd_númeroUnidadesLocales.....	53
53.	Regla_juntas_riego_agua.....	54
54.	Incorporación de información del MINEDUC.....	54
55.	Actualización con otras fuentes de información .....	54
56.	Construcción de tablas consolidadas .....	55
57.	Variables Derivadas.....	55
57.1.	<i>Estratos de ventas:</i> .....	55
57.2.	<i>Estratos de empleo y plazas de empleo registrado:</i> .....	55
57.3.	<i>Estrato de ventas según el Código orgánico de la Producción:</i> .....	55
57.4.	<i>Estrato de plazas de empleo según el Código orgánico de la Producción:</i> .....	56
57.5.	<i>Tamaño Empresa:</i> .....	56
57.6.	<i>Empresas activas estratificadas:</i> .....	56
	<b>CONTEOS Y VALIDACIONES</b> .....	57
	<b>CONCLUSIONES</b> .....	58

# 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 y procesada, 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: PostgreSQL 9.2
- Sistema de Gestión de Base de Datos: PgAdmin4 6.4 y DBeaver 22.2.1
- Herramienta BI: Pentaho Data Integration 6.1
- Herramienta de análisis de información: SPSS 22
- Herramienta de análisis de datos: Lenguaje R

**PostgreSQL**, 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.

PgAdmin4 6.4 y DBeaver 22.2.1, facilitan 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.

**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.

**Lenguaje R**, R es un entorno y lenguaje de programación con un enfoque al análisis estadístico. Se utiliza en todas las fases de análisis de datos, principalmente en Adquisición de los datos de las fuentes disponibles, Preparación de los datos, Análisis de los datos.

Con esta pequeña introducción, se da una idea de cómo es la captación y el procesamiento en el DICE, 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 DICE se compone de varias fases, en el Gráfico N: 1 se las expone de manera general:

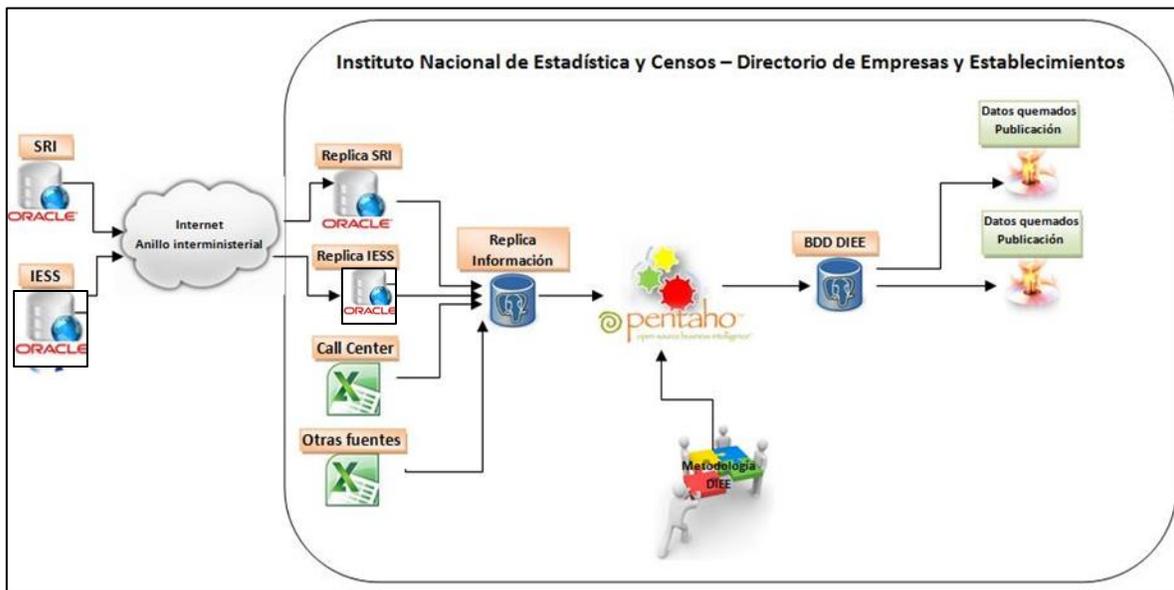


Gráfico N: 1.

El propósito de la captación de información es crear un entorno de Base de datos similar al que 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 programas con la interacción de aplicaciones como Excel.

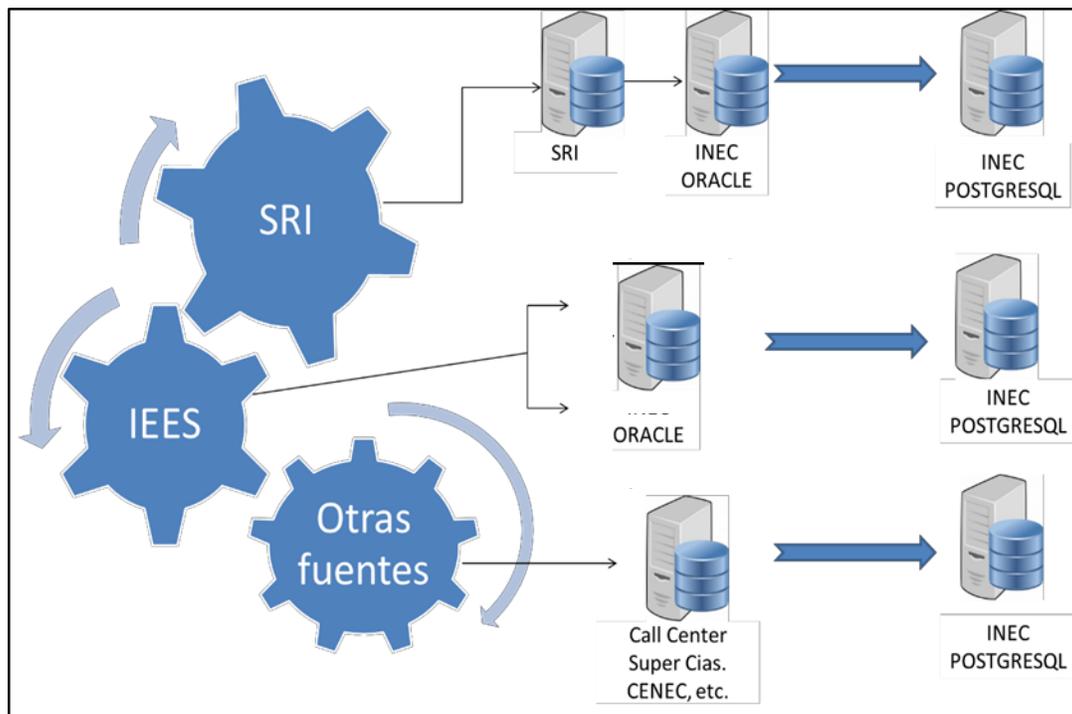


Gráfico N: 2.

Dependiendo de las herramientas usadas para la transmisión 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.

Las principales fuentes de información son:

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 DICE directamente al servidor de almacenamiento de la BDD del DICE, en esquemas específicos para su uso.

## PROCESAMIENTO

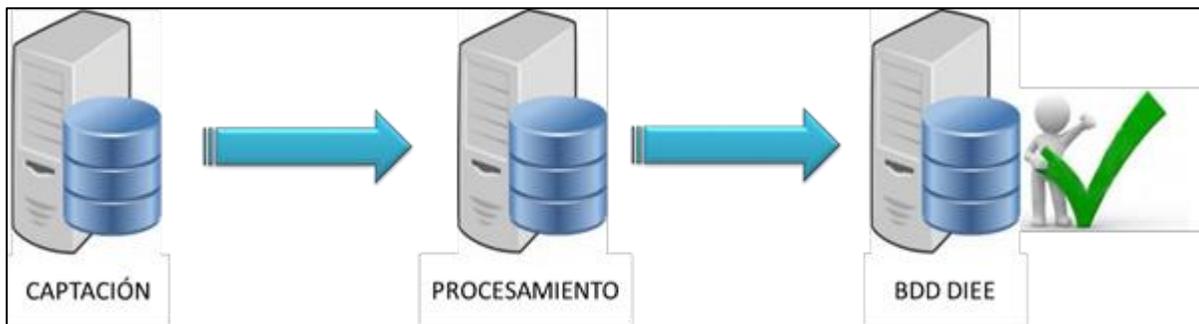


Gráfico N: 3.

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 CIU, hasta el 2014 se trabajaba con CIU3, a partir del 2015 ya se ha homologado al CIU4, sin embargo, existen códigos del SRI que no tienen un equivalente directo en CIU4, debido a que el SRI maneja un formato propio del CIU4 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 DICE, y muchas de ellas a partir del caso que se ha presentado.

- 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 CIU4.
- Se ha definido pasar la actividad económica principal de empresa a establecimiento, para el caso de contribuyentes con establecimientos únicos.
- Luego, 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 DICE, 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, en el que se distingue el estado, la fuente, los puntos de georeferencia.
- 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 medios de comunicación (números de teléfonos, celulares, correo electrónico, web) 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:

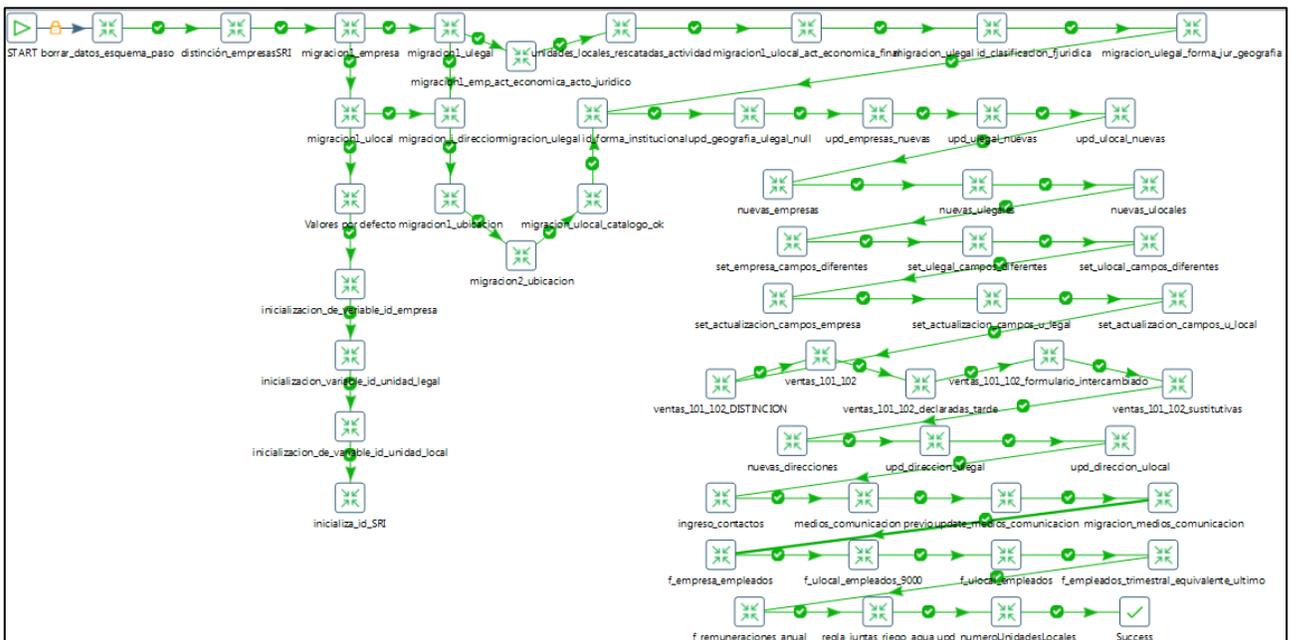


Gráfico N: 5.

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 los datos reales y sin fallos.

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

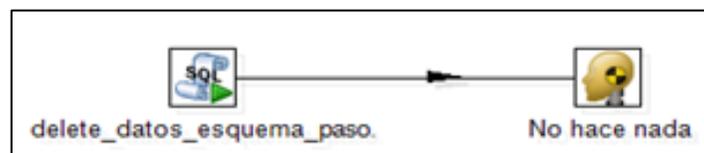


Gráfico N: 8.

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:

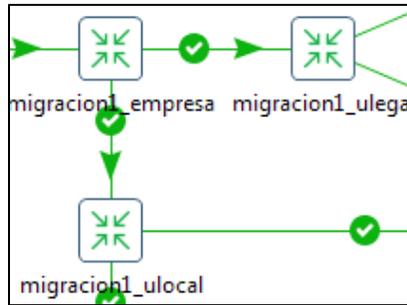


Gráfico N: 11.

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 **pasos**.

#### 4. Migración1\_empresa.



Gráfico N: 12.

Dentro de esta transformación existen los siguientes objetos:

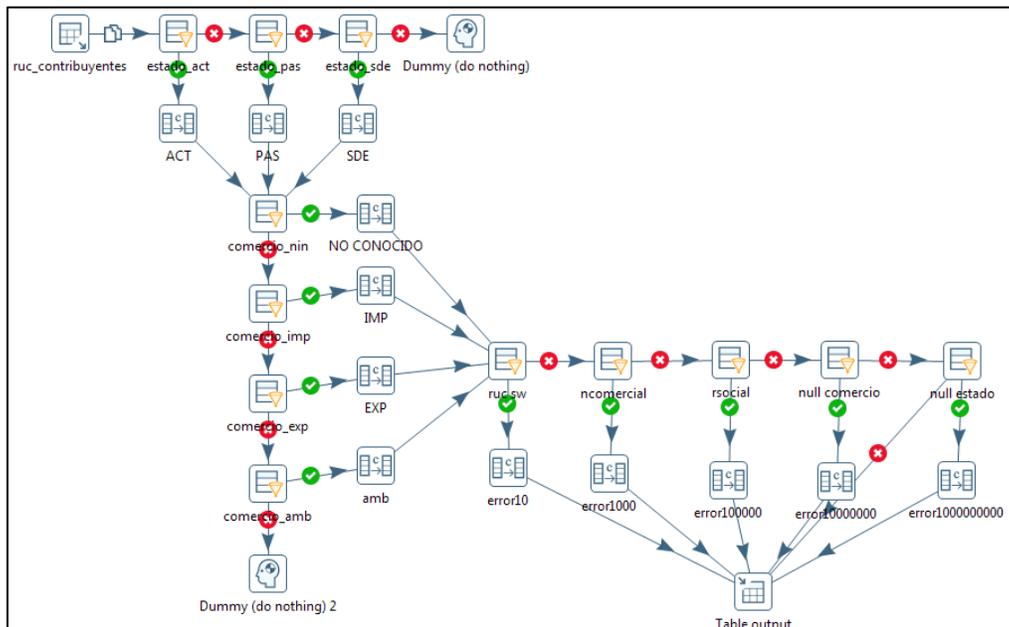


Gráfico N: 13.

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 **pasos**, con los debidos cambios como por ejemplo:

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

SRI	DICE
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 transformaciones encuentran que hay valores que no existen como por ejemplo en los catálogos o el nombre tiene menos de tres caracteres, los 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:

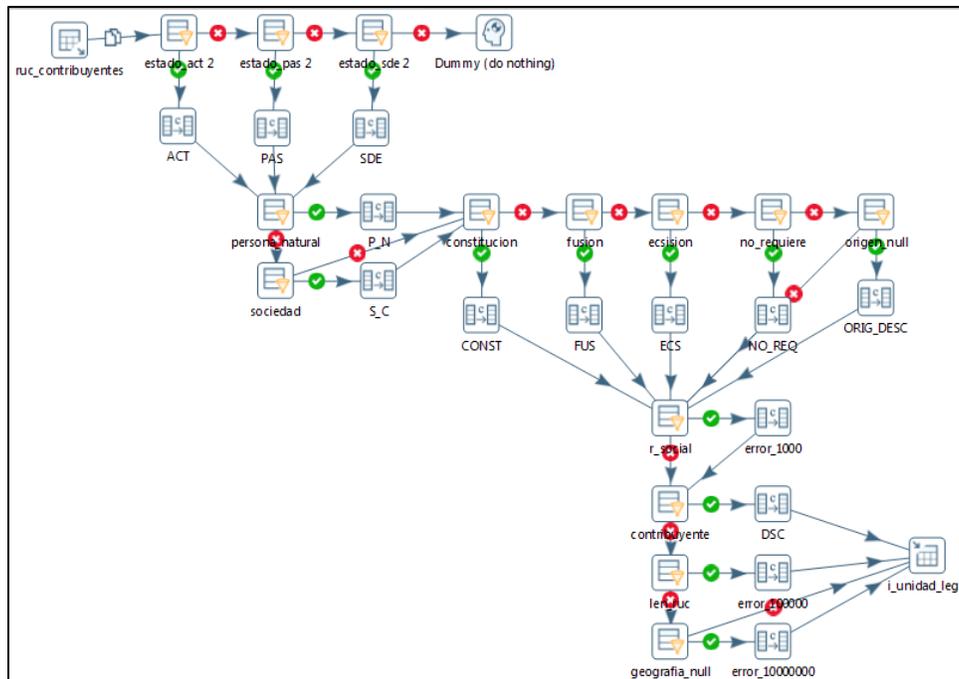


Gráfico N: 15.

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 **passo**.

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		DICE	
<code>ESTADO_PERSONA_NATURAL</code>	<code>ESTADO_SOCIEDAD</code>	<code>ID_UNIDAD_LEGAL_ESTADO</code>	
<b>ACT</b>	ACT		1
<b>PAS</b>	PAS		2
<b>SDE</b>			3

## 6. Migración1\_ulocal.

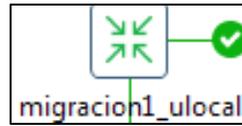


Gráfico N: 16.

La transformación de unidad local tiene los siguientes objetos:

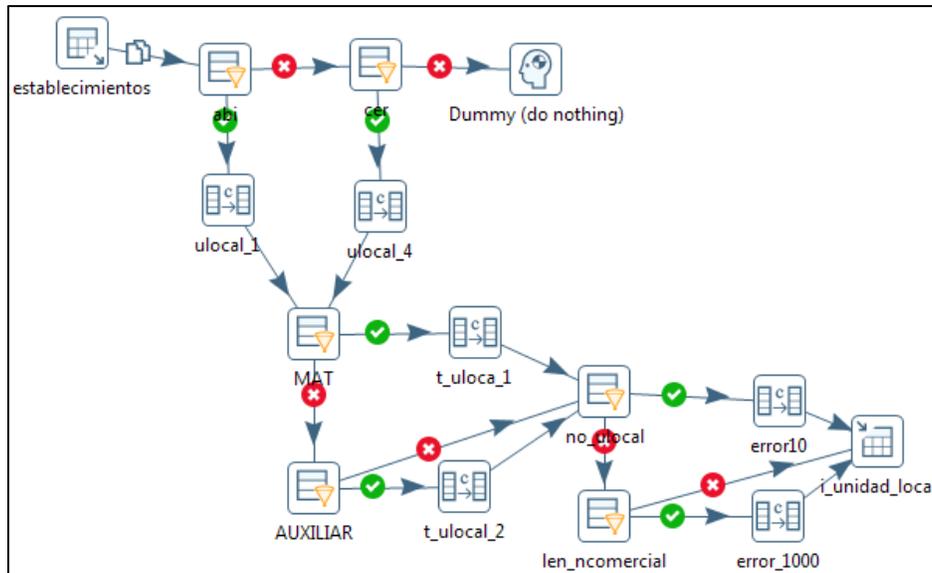


Gráfico N: 17.

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 reporte la transformación se los procederá a corregir.

## 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:

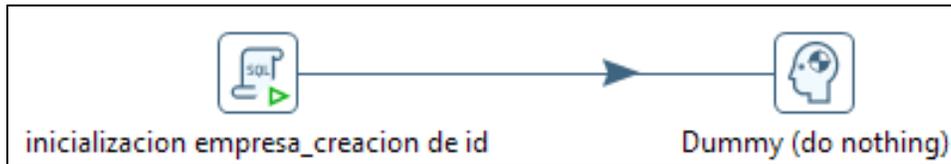


Gráfico N: 22.

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.

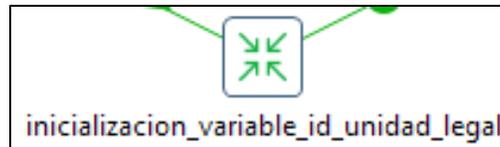


Gráfico N: 23.

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

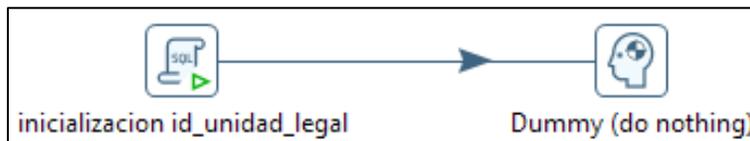


Gráfico N: 24.

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.

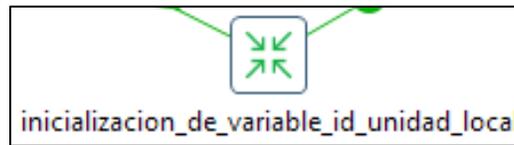


Gráfico N: 25.

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.

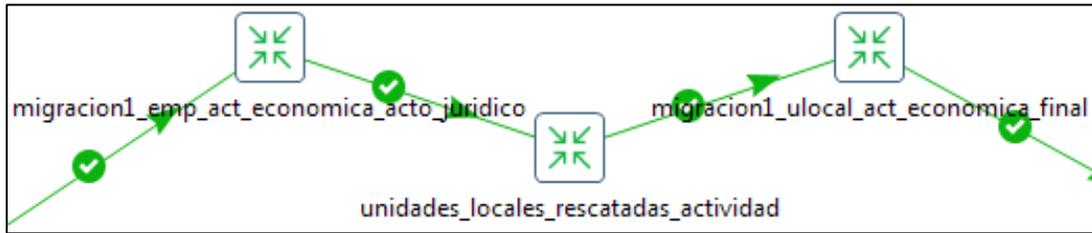


Gráfico N: 29.

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.

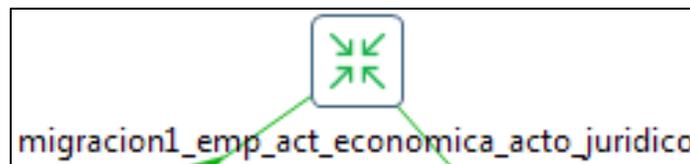


Gráfico N: 30.

La transformación de migracion1\_emp\_act\_economica tiene los siguientes objetos:

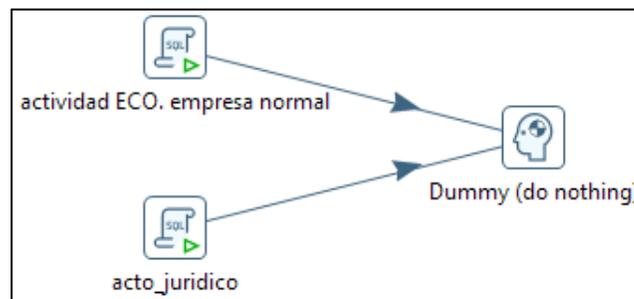


Gráfico N: 31.

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 CIU4, 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.

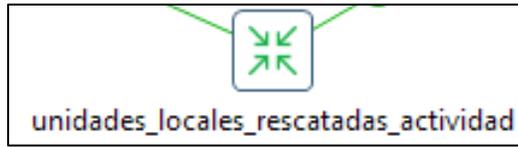


Gráfico N: 32.

La transformación tiene los siguientes objetos:

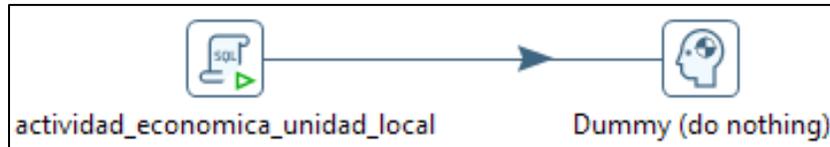


Gráfico N: 33.

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.

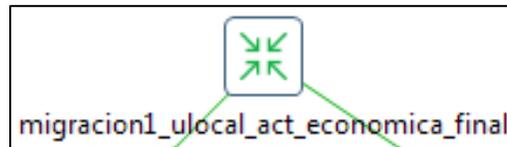


Gráfico N: 34.

La transformación de migracion1\_ulocal\_act\_economica tiene los siguientes objetos:

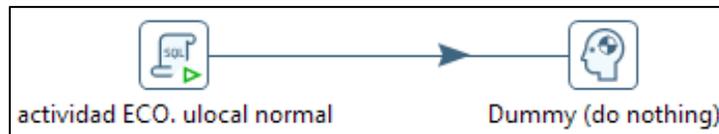


Gráfico N: 35.

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

Migración de Dirección.

En esta parte intervienen las transformaciones:

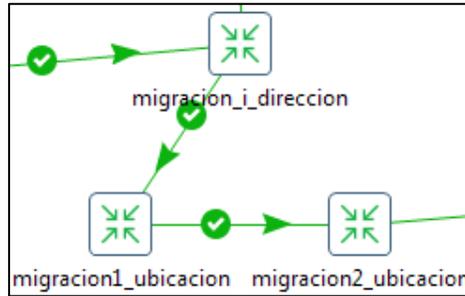


Gráfico N: 36.

**15. Migración\_i\_dirección.**



Gráfico N: 37.

La transformación de migración\_i\_direccion se compone de los siguientes objetos.

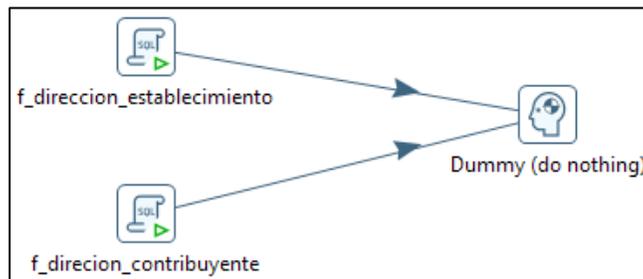


Gráfico N: 38.

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 con éstos datos la tabla de *i\_direccion* del esquema **paso**.

## 16. Migración1\_ubicación.

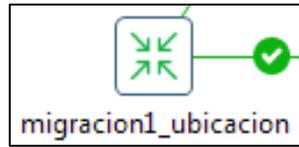


Gráfico N: 39.

La transformación de migración1\_ubicación se compone de los siguientes objetos.

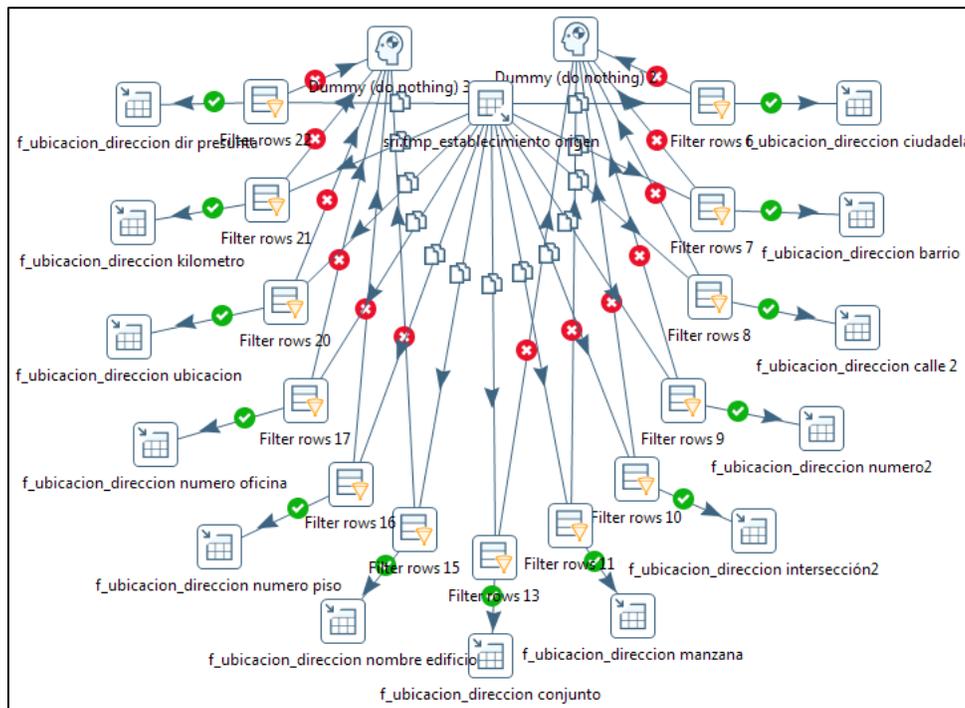


Gráfico N: 40.

## 17. Migración2\_ubicación.



Gráfico N: 41.

La transformación de migración2\_ubicación se compone de los siguientes objetos:

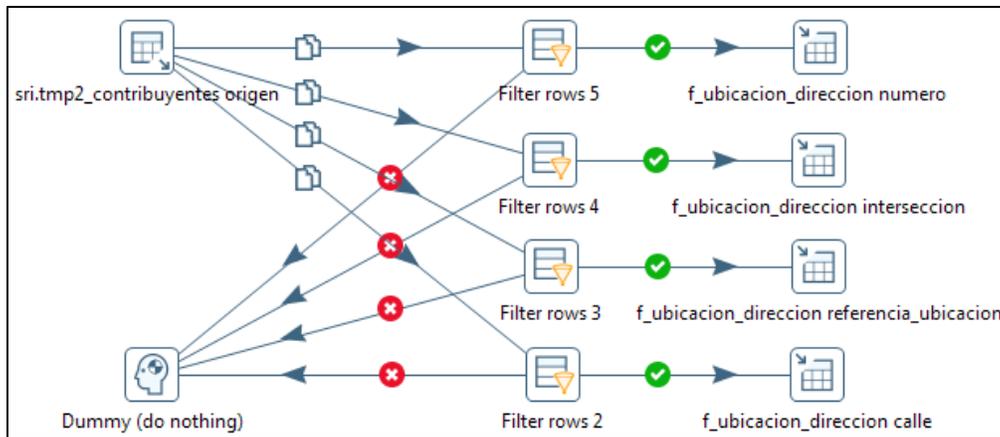


Gráfico N: 42.

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

En 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 2 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.

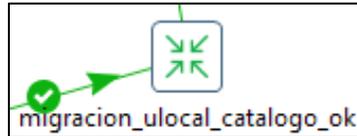


Gráfico N: 43.

La transformación de migración\_ulocal\_catálogo\_ok se compone de los siguientes objetos.

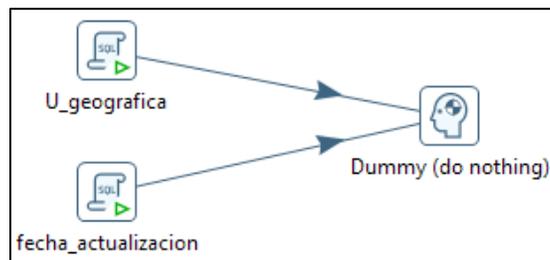


Gráfico N: 44.

Esta transformación tiene dos scripts que ejecutan lo siguiente:

- “U\_geografia”: se actualiza el id\_geografía para unidad local desde el catálogo de geografía.
- Se actualiza la fecha de actualización, para el respectivo llenado de variables de control.

### 19. Migración1\_ulegal\_clasificación\_fjurídica.

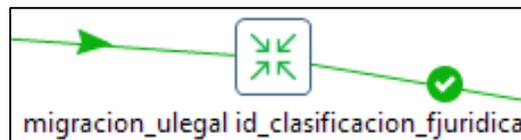


Gráfico N: 45.

La transformación tiene los siguientes objetos:



Gráfico N: 45.1.

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:

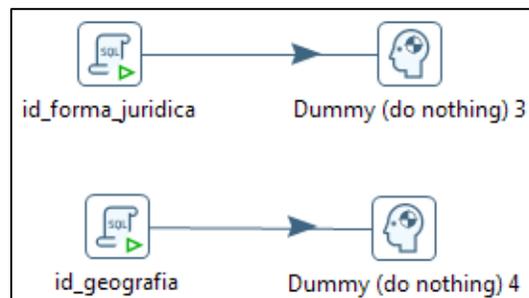


Gráfico N: 47.

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.

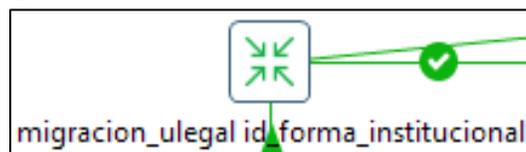


Gráfico N: 48.

La transformación tiene los siguientes objetos:

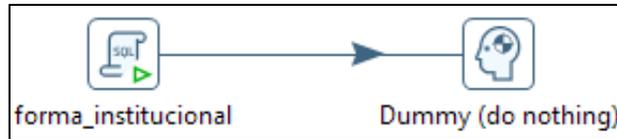


Gráfico N: 49.

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

	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:

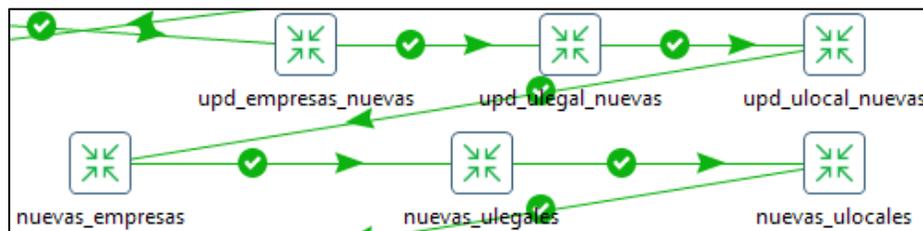


Gráfico N: 51.

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:

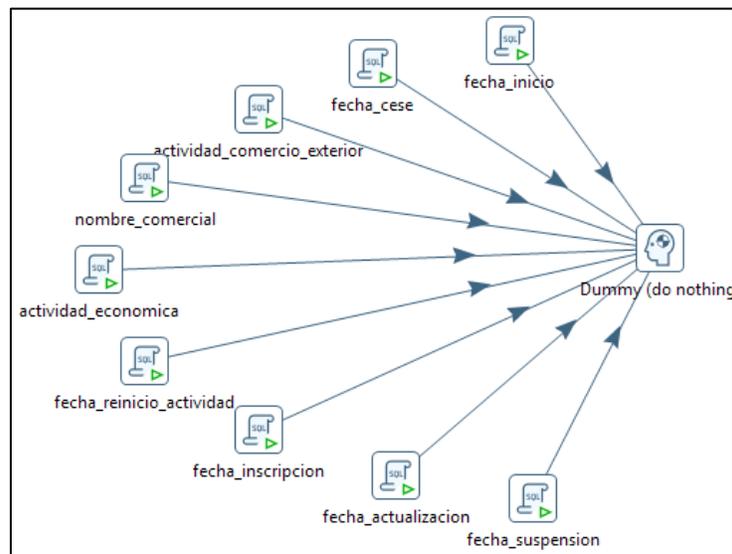


Gráfico N: 53.

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.

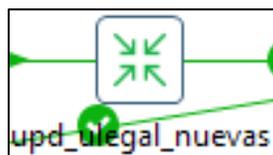


Gráfico N: 54.

La transformación tiene los siguientes objetos:

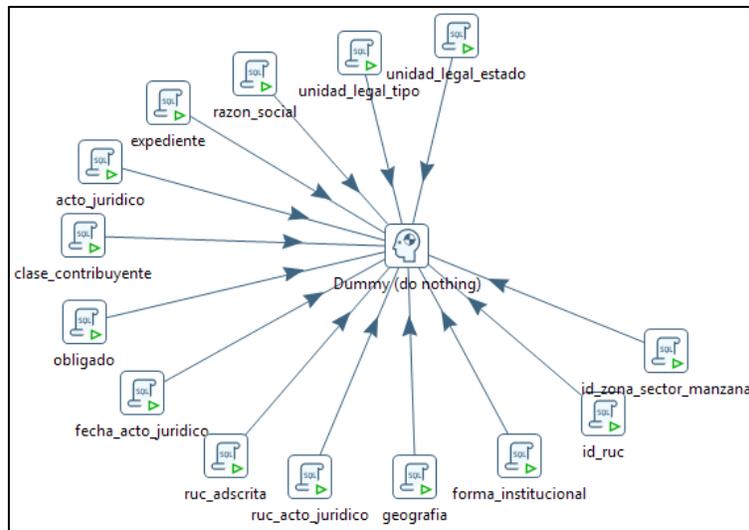


Gráfico N: 55.

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

#### 24. Upd\_ulocal\_nuevas.

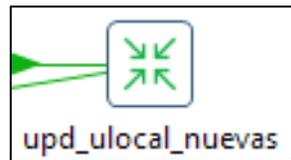


Gráfico N: 57.

La transformación tiene los siguientes objetos:

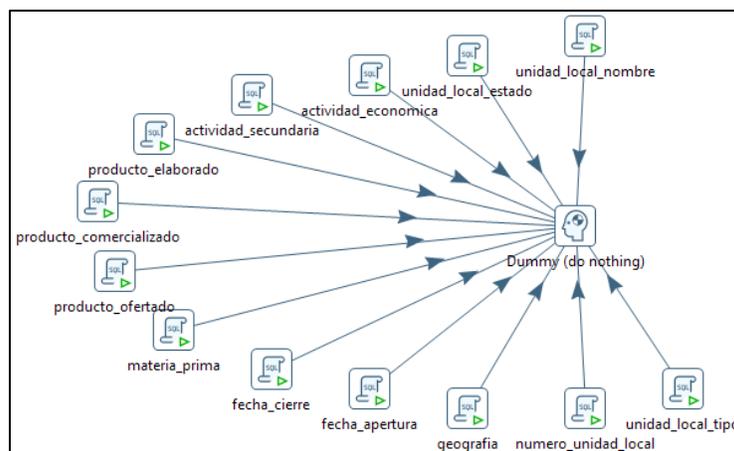


Gráfico N: 58.

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

## 25. Nuevas\_empresas

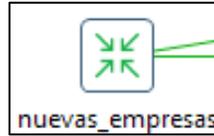


Gráfico N: 59.

La transformación contiene los siguientes objetos:



Gráfico N: 60.

Ya que en las 3 anteriores transformaciones se 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

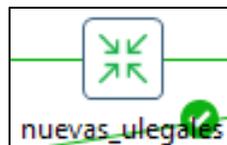


Gráfico N: 61.

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

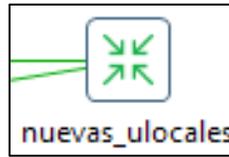


Gráfico N: 63.

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:

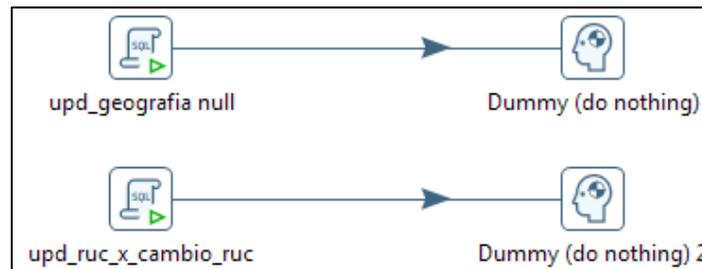


Gráfico N: 66.

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 el RUC, para los casos en los que existe un cambio de RUC.

Actualización de campos con diferencias.

Para esta actualización intervienen las transformaciones:

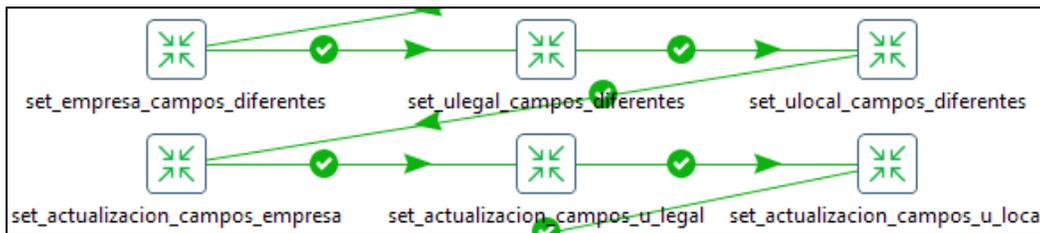


Gráfico N: 67.

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**

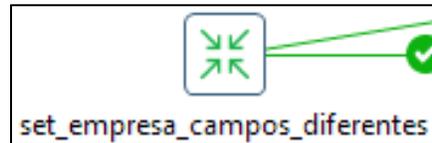


Gráfico N: 68.

La transformación contiene los siguientes objetos:

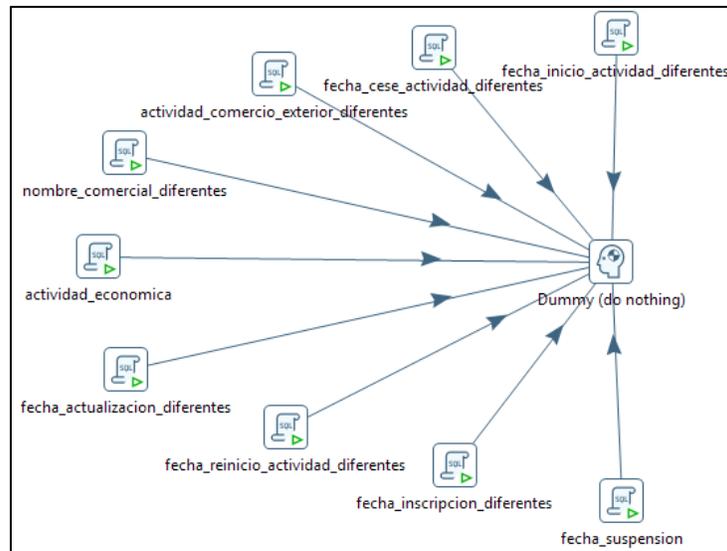


Gráfico N: 69.

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.

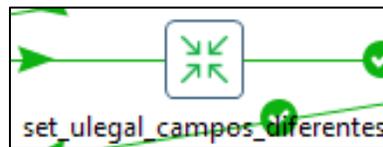


Gráfico N: 70.

La transformación contiene los siguientes objetos:

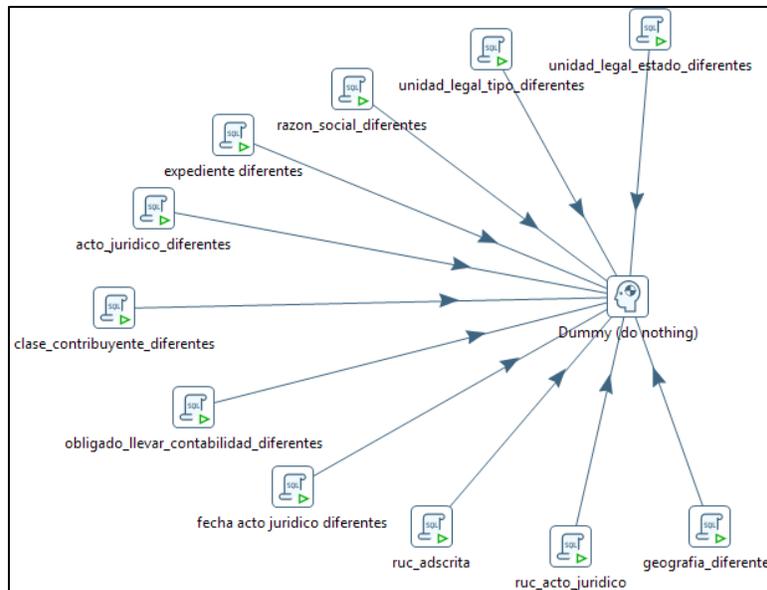


Gráfico N: 71.

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.

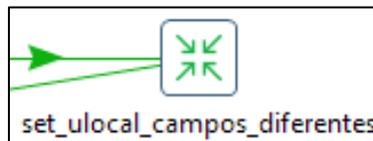


Gráfico N: 72.

La transformación contiene los siguientes objetos:

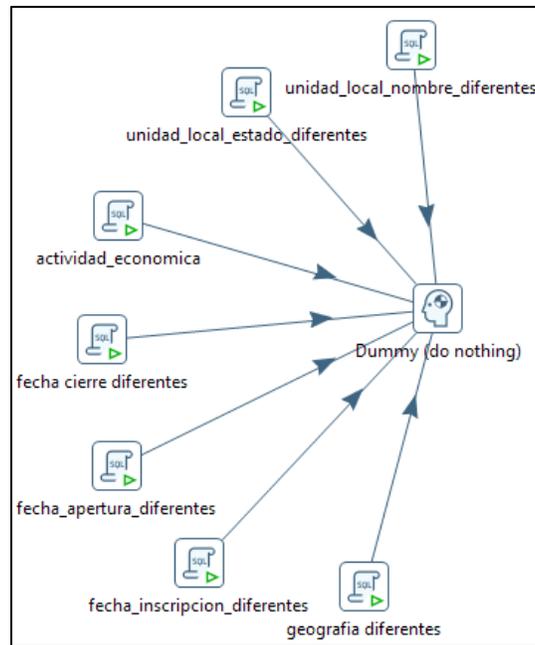


Gráfico N: 73.

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:

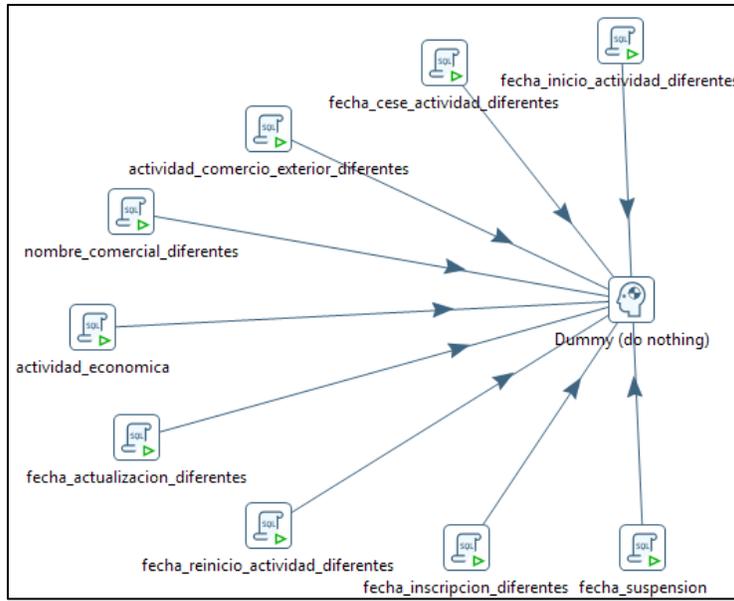


Gráfico N: 75.

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.

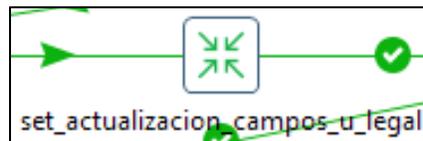


Gráfico N: 76.

La transformación contiene los siguientes objetos:

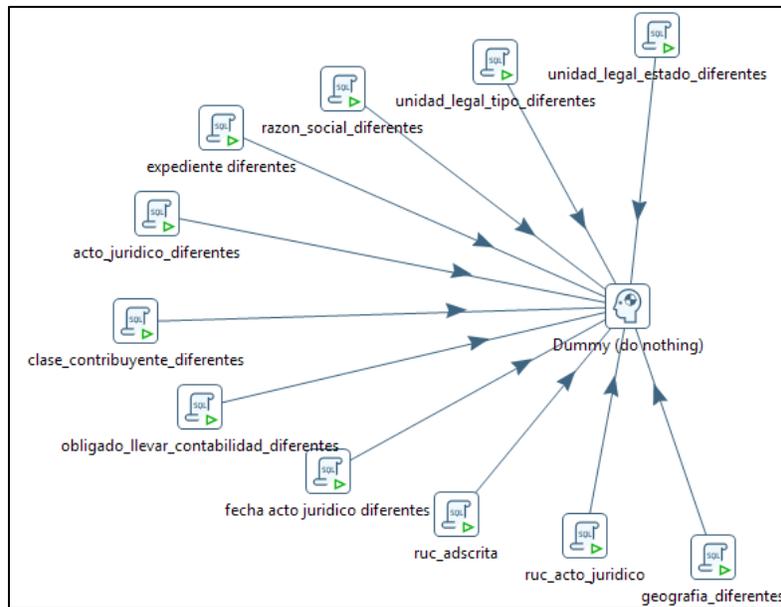


Gráfico N: 77.

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:

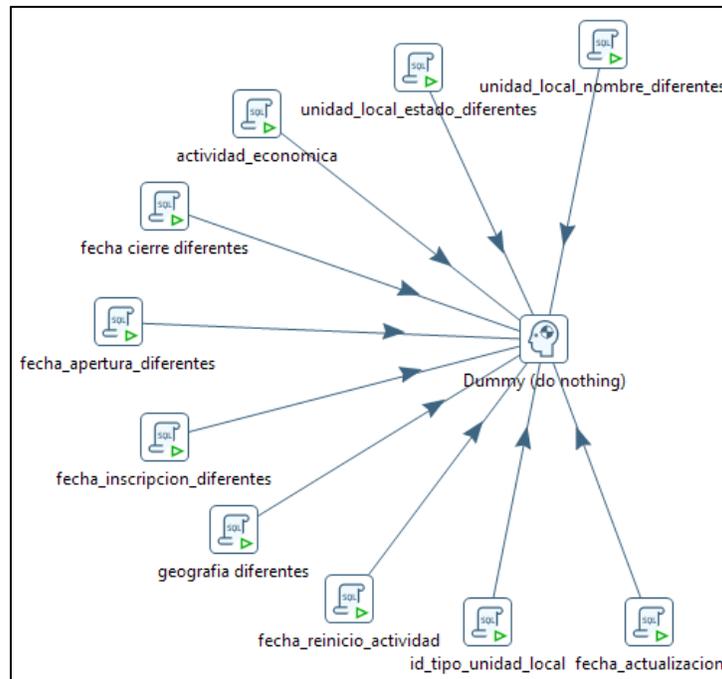


Gráfico N: 79.

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

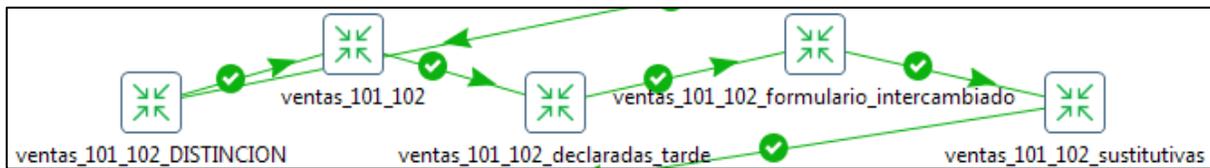


Gráfico N: 80.

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

### 35. Ventas\_101\_102\_distinción.

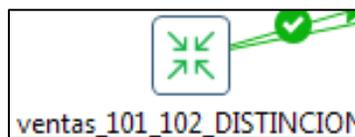


Gráfico N: 81.

La transformación contiene los siguientes objetos:



Gráfico N: 82.

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

### 36. Ventas\_101\_102.



Gráfico N: 83.

La transformación contiene los siguientes objetos:

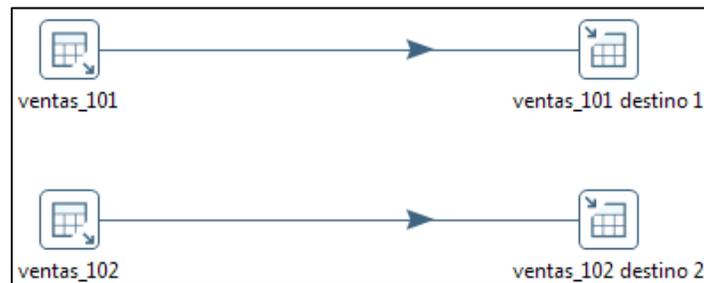


Gráfico N: 84.

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
37	CTO_CIM_MATERIA_PRIMA_2030	COSTO IMPORTACIONES MATERIA PRIMA	708	708

Gráfico N: 85.

Los campos que se utiliza para nutrir a la base del DICE 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\_mv\_ine\_anexo3\_estruc\_f101
- owb\_mv\_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

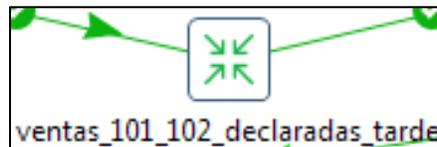


Gráfico N: 86.

La transformación contiene los siguientes objetos:

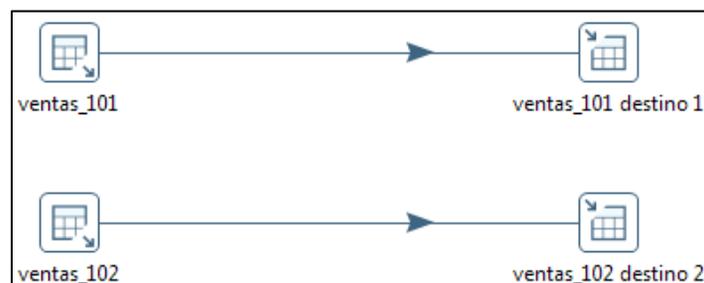


Gráfico N: 87.

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

### 38. Ventas\_101\_102\_formulario\_intercambiado

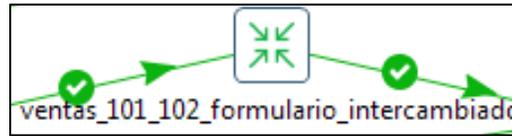


Gráfico N: 88.

La transformación contiene los siguientes objetos:

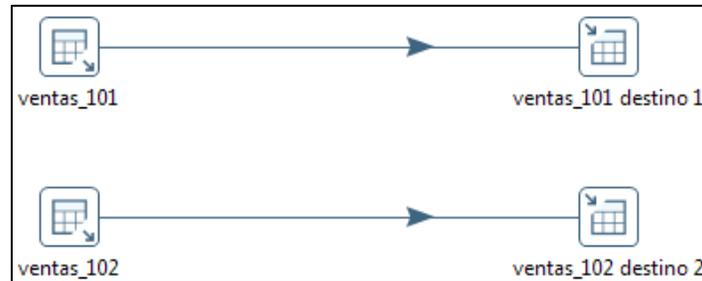


Gráfico N: 89.

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

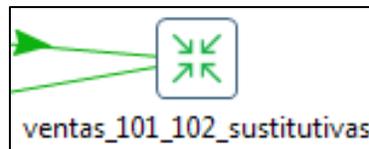


Gráfico N: 90.

La transformación contiene los siguientes objetos:

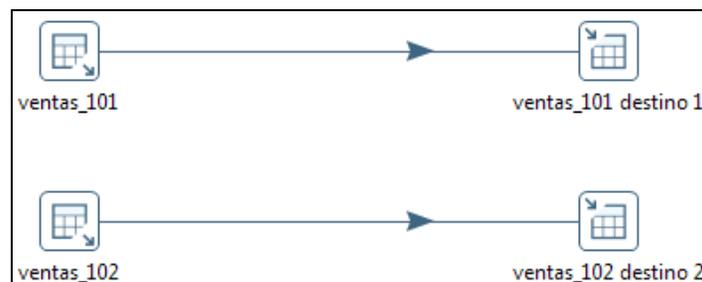


Gráfico N: 91.

Se actualizan las declaraciones que tienen diferente valor de total de ventas declaradas, comparando entre lo que se tiene en la BDD DICE 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:

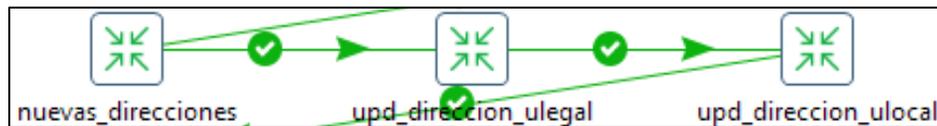


Gráfico N: 92.

Donde se realiza una migración de la información de las tablas de dirección del esquema **paso** a las tablas definitivas de la base del DICE 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 DICE 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 DICE, 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\_dirección\_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:

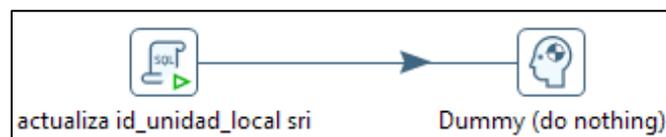


Gráfico N: 102.

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

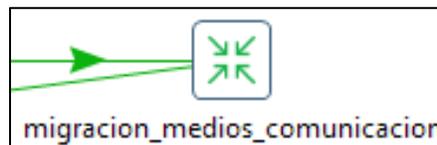


Gráfico N: 105.

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:

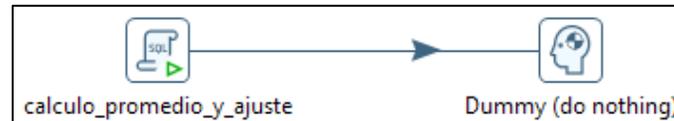


Gráfico N: 108.

En la primera parte de este objeto se realiza la sumatoria de los afiliados de todos los establecimientos de la empresa de cada mes, para afiliados 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 registraron afiliados, por último se realiza un ajuste a los afiliados 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.

A partir del procesamiento del periodo 2018, este cálculo se realiza tanto para empleo registrado como para plazas de empleo registrado, de acuerdo a las reglas 97, 98, 108 y 109 del documento de *Matriz de Reglas*.

A partir del procesamiento del periodo 2019, se han agregado las variables de empleo y plazas de empleo registrado por rangos de edad, de acuerdo a la siguiente clasificación:

1. Entre 15 y 17 años
2. Entre 18 y 29 años
3. Entre 30 y 44 años
4. Entre 45 y 64 años
5. 65 años y más.

La sumatoria de todos los rangos debe ser igual al total del empleo o plazas de empleo registrado, según corresponda, para los casos donde no

sucede esta igualdad se realiza un ajuste, en el que se debe sumar o restar la diferencia, según sea el caso, en el rango de edad que mayor cantidad de plazas o empleo registre.

#### 48. F\_ufocal\_empleados\_9000

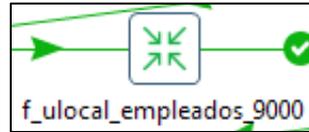


Gráfico N: 109.

La transformación contiene los siguientes objetos:

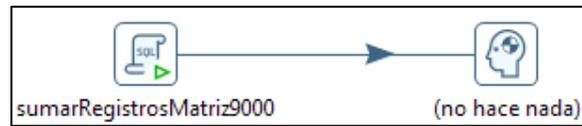


Gráfico N: 110.

En el Gráfico N: 110 se ejecutan una serie de scripts, donde inicialmente se encera la variable `id_tipo_unidad_local`, para ser nuevamente cargado con la información de la tabla `f_unidad_local` de la base del DICE, 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 afiliados que reportan estos casos, primero se enlaza entre la geografía que tiene el IESS y la que tiene la BDD del DICE, y se recuperan los casos coincidentes, los casos que no coinciden son agregados al establecimiento donde se reporte el mayor número de afiliados. Si no existe información de otros establecimientos, se crea el establecimiento matriz y se agrega los afiliados restantes.

A partir del procesamiento del periodo 2018, este cálculo se realiza tanto para empleo registrado como para plazas de empleo registrado, de acuerdo a las reglas 112 y 113 del documento de *Matriz de Reglas*.

#### 49. F\_ufocal\_empleados

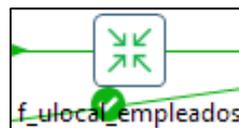


Gráfico N: 111.

La transformación contiene los siguientes objetos:

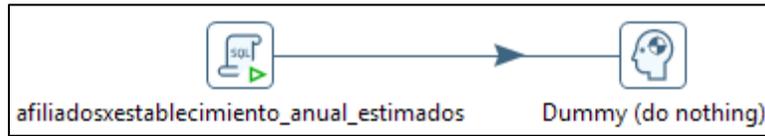


Gráfico N: 112.

Aquí se obtienen afiliados 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\_trimestral\_equivalente\_último

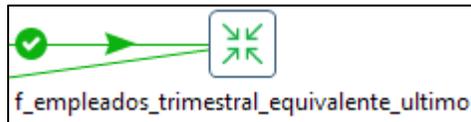


Gráfico N: 113.

La transformación contiene los siguientes objetos:

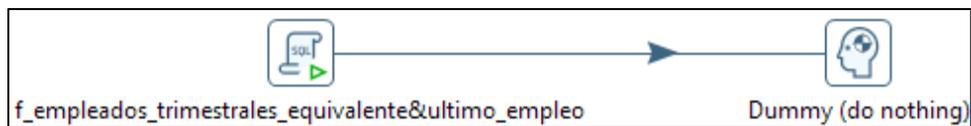


Gráfico N: 114.

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

A partir del procesamiento del periodo 2018, este cálculo se realiza tanto para empleo registrado como para plazas de empleo registrado.

## 51. Remuneraciones\_anual



Gráfico N: 115.

La transformación contiene los siguientes objetos:



Gráfico N: 116.

En el script del Gráfico N: 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:

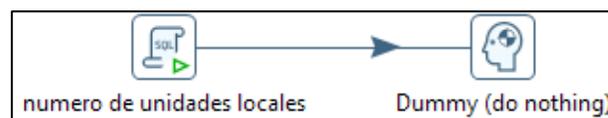


Gráfico N: 118.

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 DICE, 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:

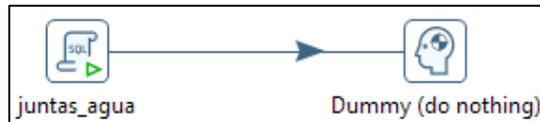


Gráfico N: 120.

Con este script se corrige la información de la BDD DICE 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. Incorporación de información del MINEDUC

Se realiza la incorporación de las nuevas unidades educativas entregadas por el MINEDUC en la base de datos del DICE, se actualizan las variables que hayan presentado cambios del resto de unidades educativas. Con la información de afiliados y remuneraciones, se realiza una redistribución, para lo cual; se distribuye la afiliación que haya reportado el IESS de cada distrito provincial a las diferentes unidades educativas pertenecientes al distrito, esto, según el peso de afiliación que haya asignado el MINEDUC, por lo cual, la información distribuida será totalmente consistente con lo reportado inicialmente por el IESS.

### 55. Actualización con otras fuentes de información

Mediante la entrega de matrices de actualización (archivos excel) se realizan procesos de actualización en la base de datos del DICE sobre variables prioritarias como son: actividad económica, geografía, direcciones, medios de comunicación, actos jurídicos, entre otras. Estos de fuentes que complementan y nutren la información del DICE como son: Call center (información recuperada por personal de GEERA), encuestas industriales, entre otras.

## 56. Construcción de tablas consolidadas

Terminada esta etapa de procesamiento, se crea una tabla que consolida toda la información relevante procesada de las diferentes tablas del modelo ER de la BDD del DICE, de todos los años donde se dispone de información.

## 57. Variables Derivadas

Se asignan los estratos para empleo registrado, plazas de empleo registrado y ventas en las empresas según los catálogos de clase de ventas y empleados, que se muestran en las siguientes tablas:

### 57.1. 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	9999999999

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 plazas de empleo registrado esté entre 1 y 10.

### 57.2. Estratos de empleo y plazas de empleo registrado:

codigo	Estrato	valor_mínimo	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

### 57.3. Estrato de ventas según el Código orgánico de la Producción:

codigo	Estrato	valor_inferior	valor_superior
1	ESTRATO I	1	300000
2	ESTRATO II	300001	1000000
3	ESTRATO III	1000001	5000000

4	ESTRATO IV	5000001	999999999999
---	------------	---------	--------------

#### 57.4. Estrato de plazas de empleo según el Código orgánico de la Producción:

codigo	Estrato	valor_mínimo	valor_máximo
1	ESTRATO I	1	9
2	ESTRATO II	10	49
3	ESTRATO III	50	199
4	ESTRATO IV	200	10000000

#### 57.5. Tamaño Empresa:

Esta variable se determina en base a los estratos de ventas y plazas o empleo registrado, según las siguientes condiciones:

- Si tiene estrato de ventas el tamaño será igual a este valor
- Si no tiene estrato de ventas y posee estrato de plazas o empleo, el tamaño será igual al estrato de plazas y empleo según corresponda.
- Si tiene estrato de ventas y posee estrato de plazas o empleo, pero la forma institucional es igual a 7, el tamaño será igual al estrato de plazas y empleo según corresponda.

#### 57.6. Empresas activas estratificadas:

Aquí se obtiene la variable situación, donde se asigna con código igual a uno, para identificar a todas las empresas activas o que poseen estratos, es decir, que tienen ventas o plazas de empleo o son RISE. De este grupo se excluyen las empresas que tengan actividad inválida, codificada con sección: X, Z, T o U.

## 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 DICE, 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 en adelante.
- Se mantienen para este año las tablas de afiliados equivalente y último plazas y empleo donde se obtiene información más acertada de empleados.
- Se modificó el cálculo de la variable forma institucional, debido a una reestructura del RUC según el SRI, por lo que ahora se consideran otras variables como es tipo\_contribuyente que es ahora el elemento principal para el cálculo de la misma.
- Se modificó el cálculo de ventas para el sector financiero y sociedades sin fines de lucro (SSFL). Para este proceso, según la metodología establecida por el equipo de Análisis, se incorporó en la BDD, en la tabla de ventas, las variables: TOTAL\_REX\_DON\_APO\_1910, para agregar al cálculo total de ventas para las SSFL y TOTAL\_INGRESOS\_1930, que es ahora el total de ventas de Bancos y cooperativas.
- 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.
- 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.

<b>Fecha:</b> 16 de noviembre de 2022	
<b>Elaborado por:</b> Carolina Mina	<b>Aprobado por:</b> Libertad Trujillo