



Escuela de Ingeniería de Sistemas y  
Computación  
Facultad de Ingeniería

# Práctica 1. Proceso ETL con “Pentaho Data Integration”

Oswaldo Solarte Pabón  
[oswaldo.solarte@correounivalle.edu.co](mailto:oswaldo.solarte@correounivalle.edu.co)

# Tabla de contenido

Índice de Tablas .....	2
Índice de Ilustraciones .....	2
1. Objetivos .....	5
2. Recomendaciones .....	5
3. Practica 1 .....	5
3.1. Caso de estudio .....	5
3.2. Instalación de la Base de Datos Fuente .....	6
3.3. Diseño del Data Warehouse .....	10
3.3.1. Seleccionar el proceso de negocio a modelar .....	10
3.3.2. Escoger nivel de granularidad del proceso .....	11
3.3.3. Seleccionar las dimensiones que se van a aplicar .....	12
3.3.4. Escoger los hechos medibles .....	18
3.4. Implementación del Data Warehouse .....	19
3.4.1. Implementación Dimensión IPS.....	20
3.4.2. Implementación Dimensión Médico .....	32
3.4.3. Implementación Dimensión Servicio .....	32
3.4.4. Implementación Dimensión Persona.....	36
3.4.5. Implementación Dimensión Fecha .....	48
3.4.6. Implementación (Dimensiones – Hechos) .....	65
4. Practica 2 .....	65
4.1. Diseño del Data Mart.....	65
4.2. Implementación del Data Mart .....	67
4.2.1. Dimensión Médico .....	67
4.2.2. Dimensión Medicamento.....	67
4.4.3. Implementación (Dimensiones - Hechos) .....	67

## Índice de Tablas

Tabla 1. Archivos fuente .....	6
Tabla 2 Pasos para configurar la conexión a la base de datos .....	23
Tabla 3. Vista y configuración proceso "beneficiario-cotizante" .....	43
Tabla 4. Organización de campos a unir.....	44
Tabla 5. Configuración procesos "Select values" .....	46

## Índice de Ilustraciones

Ilustración 1. Creación de base de datos desde pgAdmin .....	7
Ilustración 2. Restaurar base de datos .....	7
Ilustración 3. Restaurar base de datos .....	8
Ilustración 4. Mensaje de confirmación, base de datos restaurada.....	8
Ilustración 5. Estructura de la base de datos fuente.....	9
Ilustración 6. Información de la tabla beneficiario .....	9
Ilustración 7. Estructura del archivo de medicamentos.....	10
Ilustración 8. Atributos dimensión fecha.....	13
Ilustración 9. Atributos dimensión ubicación .....	14
Ilustración 10. Atributos dimensión persona .....	15
Ilustración 11. Estructura de la tabla IPS .....	15
Ilustración 12. Atributos de la dimensión IPS .....	16
Ilustración 13. Estructura de la tabla médico .....	16
Ilustración 14. Dimensión médico .....	16
Ilustración 15. Atributos dimensión servicio .....	17
Ilustración 16. Diseño data mart .....	17
Ilustración 17. Data mart atención de pacientes .....	19
Ilustración 18. Crear base de datos.....	20
Ilustración 19. Creación de la bodega de datos ETL-process.....	20
Ilustración 20. Proceso ETL.....	21
Ilustración 21. Creación de una transformación .....	21
Ilustración 22. Guardar dimensión fecha .....	22
Ilustración 23. Crear la conexión a la base de datos .....	22
Ilustración 24. Arrastrar herramienta "Table input" .....	23
Ilustración 25. Obtener información tabla IPS.....	24
Ilustración 26. Mensaje de confirmación para incluir campos en la consulta .....	24
Ilustración 27. Campos de la tabla IPS .....	25
Ilustración 28. Arrastrar herramienta "Select values" .....	25
Ilustración 29. Uniendo los dos procesos .....	26
Ilustración 30. Configuración proceso "Select values" .....	26
Ilustración 31. Arrastrar herramienta "Dimension lookup/update" .....	26
Ilustración 32. Configuración primera parte proceso "Dimension ips".....	27
Ilustración 33. Configuración segunda parte proceso "Dimension" .....	27

Ilustración 34. Eliminación del campo "id_ips" .....	28
Ilustración 35. Configuración tercera parte del proceso "Dimension".....	28
Ilustración 36. Ejecución del SQL para crear la dimensión ips .....	29
Ilustración 37. Tabla dim_ips creada en la base de datos .....	29
Ilustración 38. Ejecutar transformación dim_ips .....	30
Ilustración 39. Ejecución exitosa de la transformación .....	30
Ilustración 40. Visualizar los datos de la tabla dim_ips .....	31
Ilustración 41. Información ingresada en la dimensión fecha .....	31
Ilustración 42. Transformación dimensión médico .....	32
Ilustración 43. Arrastrar herramienta "Data Grid" .....	32
Ilustración 44. Configuración estructura de la dimensión servicio.....	33
Ilustración 45. Configuración información dimensión servicio .....	33
Ilustración 46, Arrastrar herramienta "Select values" .....	33
Ilustración 47. Configuración proceso "Select values" .....	34
Ilustración 48. Configuración proceso "Dimensión servicio" .....	34
Ilustración 49. Ejecución SQL dimensión servicio .....	35
Ilustración 50. Transformación dimensión ejecutada.....	35
Ilustración 51. Arrastrar herramienta "Sort rows" .....	43
Ilustración 52. Arrastrar herramienta "Merge Join" .....	44
Ilustración 53. Configuración proceso "Merge join" .....	45
Ilustración 54. Pre visualización proceso "Merge join" .....	45
Ilustración 55. Arrastrar herramienta "Select values" .....	46
Ilustración 56. Arrastrar herramienta "Dimension" .....	47
Ilustración 57. Arrastrar "Generate Rows" .....	48
Ilustración 58. Configuración paso "Generate initial date" .....	49
Ilustración 59. Vista previa proceso "Generate initial date".....	49
Ilustración 60. Arrastrar "Add sequence" .....	50
Ilustración 61. Configuración paso "Add day sequence".....	50
Ilustración 62. Unión de los procesos .....	51
Ilustración 63. Arrastrar proceso "Calculator" .....	51
Ilustración 64. Configuración proceso "Date generator" .....	51
Ilustración 65. Pre-visualización proceso "Date generator".....	52
Ilustración 66. Configuración proceso "Date information generator" .....	52
Ilustración 67. Pre-visualización proceso "Date information generator" .....	53
Ilustración 68. Arrastrar proceso "Modified Java Script Value".....	53
Ilustración 69. Configuración proceso "Modified Java Script Date Value" .....	54
Ilustración 70. Transformación dimensión fecha.....	54
Ilustración 71. Arrastrar proceso "Filter rows" .....	55
Ilustración 72. Arrastrar el proceso "Select values" .....	55
Ilustración 73. Arrastrar le proceso "Dummy".....	56
Ilustración 74. Vista previa transformación dimensión fecha .....	56
Ilustración 75. Configuración proceso "Filter rows current date" .....	56
Ilustración 76. Posibles campos a incluir en la dimensión .....	57
Ilustración 77. Selección de campos a eliminar .....	58

Ilustración 78. Eliminar campos seleccionados .....	58
Ilustración 79. Pre-visualización del proceso "Select values" .....	59
Ilustración 80. Arrastrar proceso "Dimension lookup/update" .....	59
Ilustración 81. Configuración primera parte proceso "Dimension" .....	60
Ilustración 82. Configuración segunda parte proceso "Dimension" .....	60
Ilustración 83. Eliminación del campo "date" .....	61
Ilustración 84. Configuración tercera parte del proceso "Dimension".....	61
Ilustración 85. Ejecución del SQL para crear la dimensión fecha.....	62
Ilustración 86. Tabla dim_fecha creada en la base de datos .....	62
Ilustración 87. Ejecutar transformación dim_fecha .....	63
Ilustración 88. Ejecución exitosa de la transformación .....	63
Ilustración 89. Visualizar los datos de la tabla dim_fecha .....	64
Ilustración 90. Información ingresada en la dimensión fecha .....	64
Ilustración 91. Transición Servicios .....	65
Ilustración 92. Transformación hecho atención pacientes .....	65
Ilustración 93. Diseño Data Marts proceso de negocio "formulas médicas" .....	66
Ilustración 94. Transformación dimensión médico .....	67
Ilustración 95. Transformación dimensión medicamento .....	67
Ilustración 96. Transformación hecho "medicamentos recetados" .....	67
Ilustración 97. Transformación hecho "formulas medicas".....	68

---

## 1. Objetivos

---

- Aplicar el proceso ETL, utilizando la herramienta *Pentaho Data Integration*, mediante el estudio de un ejemplo de la vida cotidiana.
- Reconocer los componentes principales de la herramienta *Pentaho Data Integration*.

---

## 2. Recomendaciones

---

Para el correcto desarrollo de esta práctica, se requiere tener previamente instalados las siguientes aplicaciones:

- ✓ Pentaho Data Integration
- ✓ PgAdmin4
- ✓ Postgresql 15

---

## 3. Practica 1

---

### 3.1. Caso de estudio

#### **Proyecto Colombia Saludable**

El gobierno nacional ha conocido de sus habilidades en el área de *Business Intelligence*. Por esta razón, teniendo en cuenta que se debe poner en marcha la nueva EPS (*Empresa Prestadora de Servicios de Salud*) *Colombia Saludable*, como sustitución del antiguo ISS (*Instituto de Seguro Social*), lo han contratado para aplicar estas técnicas y ofrecer un mejor servicio a los usuarios (cotizantes y/o beneficiarios) como lo manda la ley.

Los servicios que oferta *Colombia Saludable* son:

- ✓ Citas médicas general
- ✓ Cita médica especializada
- ✓ Hospitalización
- ✓ Cirugías
- ✓ Urgencias
- ✓ Terapias
- ✓ Solicitud de Medicamentos

Procesos administrativos:

- ✓ Vinculaciones
- ✓ Retiros
- ✓ Cambio de estado (cotizante - beneficiario)

Como fuente de datos nos brindan la base de datos del antiguo ISS y un archivo .csv que contiene la información de los medicamentos que se formulan en la EPS. Los cuales se pueden descargar en el siguiente [link](#).

Como suele ser común en Colombia no se conoce el estimado del costo de mantener esta EPS, ni se conoce lo que el estado debe proveer para la población que se envejece. Además, el estado está interesado en ofrecer el mejor servicio y por esto requiere contar con datos e información sobre:

- Tiempo promedio de espera desde asignación de cita hasta atención
- Tiempos de espera cirugías.
- Centros Hospitalarios que más atienden pacientes por región, ciudad.
- Medicinas más recetadas en un período dado.
- Medicamentos que se formulan juntos.
- Medicinas que se formulan con más frecuencia.
- Promedio de utilización de servicios por paciente, sexo, edad.
- Estudio de ocupación de hospitales y médicos dependiendo de épocas del año (navidad, vacaciones, semana santa).
- Épocas del año más atención de más espera, de más enfermos
- Días de la semana más demanda de servicio
- Costo promedio de recetas médicas.
- Regiones del país con más demanda de servicios.

## 3.2. Instalación de la Base de Datos Fuente

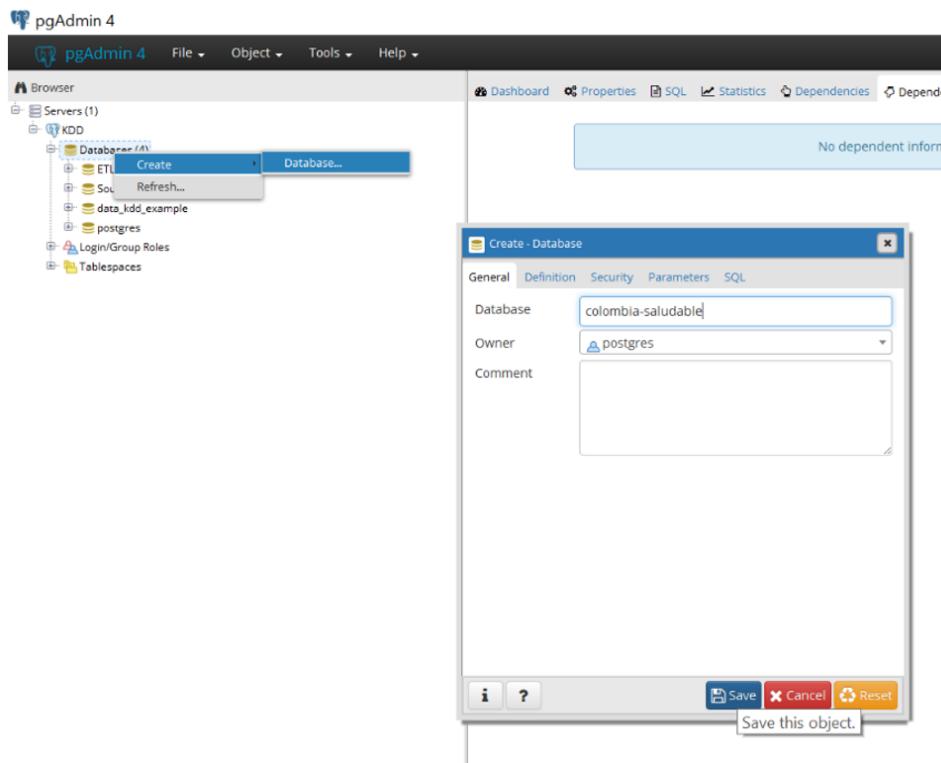
En este [link](#) encontraremos los archivos que se describen en la **Tabla 1**.

**Tabla 1. Archivos fuente**

Nombre	Descripción
<b>Source-DB</b>	Backup de la base de datos del antigua ISS
<b>Medicamentos.xls</b>	Base de datos de los medicamentos que se formulan.
<b>date_script.js</b>	Script de ayuda para crear la dimensión fecha

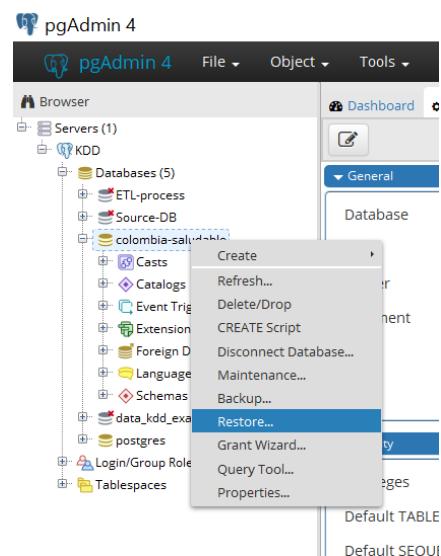
Ahora vamos a descargarlos y seguiremos los siguientes pasos para instalar la base de datos en nuestro equipo.

Desde el *Pgadmin*, crearemos una nueva base de datos con el nombre “*colombia-saludable*” como se muestra en la **ilustración1**.



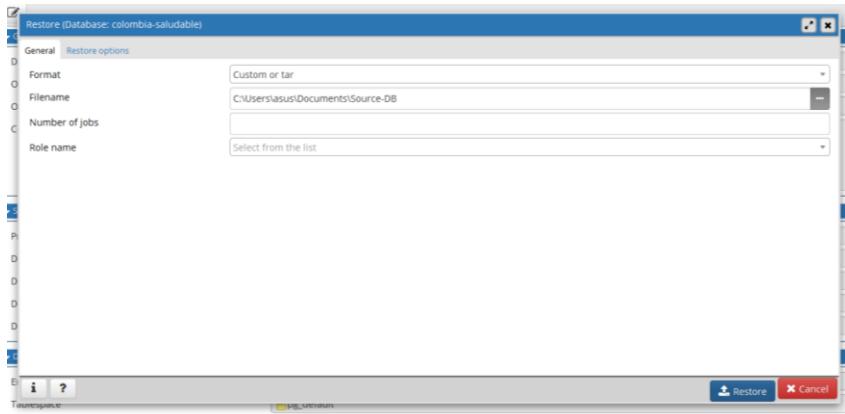
**Ilustración 1.** Creación de base de datos desde pgAdmin

Luego seleccionamos la base de datos creada, damos clic derecho y seleccionamos la opción “*Restore*” como se muestra en la **ilustración 2**.



**Ilustración 2.** Restaurar base de datos

Luego seleccionamos el backup descargado “*Source DB*”, damos clic la opción “*Restore*” como se muestra en la **ilustración 3**.



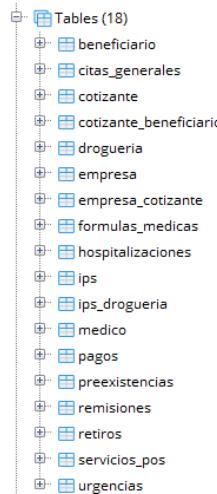
**Ilustración 3. Restaurar base de datos**

Una vez finaliza el proceso de carga de la base de datos, recibiremos un mensaje de confirmación como el que se muestra en la **ilustración 4**.



**Ilustración 4. Mensaje de confirmación, base de datos restaurada**

Verificamos desde el “*Browser*” de *pgAdmin*, que la base de datos contenga la estructura que se muestra en la **ilustración 5**.



**Ilustración 5. Estructura de la base de datos fuente.**

Además, de que las tablas contengan información como se muestra la **Ilustración 6.**

	tipo_identificacion character varying(40)	id_beneficiario character varying(40)	parentesco character varying(400)	nombre character varying(400)	fecha_nacimiento date	sexo character varying(2)	estado_civil character varying(20)	tipo_discapacidad character varying(40)
1	Tarjeta Identidad	1135438122460	Hija	Marcel Lozano Ossorio	1991-05-18	F	Soltero	Ninguna
2	Cedula	1135438123690	Hijo	Adrián Lozano Ossorio	1980-11-21	M	Vludo	Ninguna
3	Cedula	1135438126210	Hermano	Terenio Madrid Vélez	1932-11-19	M	Vludo	Ninguna
4	Cedula	1135438127420	Hijo	Santos Madrid Queda	1990-05-07	M	Divorciado	Motora
5	Cedula	1135438129880	Papa	Elias Moreno Jaramillo	1960-09-19	M	Soltero	Ninguna
6	Cedula	1135438131020	Hermano	Tomás Moreno Carreño	1967-12-21	M	Soltero	Ninguna
7	Tarjeta Identidad	1135438132210	Hija	Jorgelina Moreno Vargas	1994-01-25	F	Soltero	Ninguna
8	Tarjeta Identidad	1135438133520	Hijo	Emanuel Moreno Jaramillo	1994-09-03	M	Soltero	Ninguna
9	Cedula	1135438134430	Mama	Edna Carreño Pereira	1950-07-01	F	Soltero	Ninguna
10	Tarjeta Identidad	1135438137560	Hijo	Celso Morales Salazar	1992-09-11	M	Soltero	Ninguna
11	Cedula	1135438138800	Hermana	Albertina Morales Fernández	1950-03-04	F	Union Libre	Ninguna
12	Cedula	1135438139930	Hermano	Pancracio Morales Fernández	1937-10-03	M	Divorciado	Ninguna
13	Cedula	1135438141100	Hija	Juana Morales Correa	1980-12-24	F	Vludo	Ninguna
14	Cedula	1135438143540	Hija	Belinda Arvelao Verdugo	1990-10-16	F	Vludo	Motora
15	Tarjeta Identidad	1135438147520	Hijo	Dámaso Barrios Yáñez	1998-02-10	M	Soltero	Mental
16	Cedula	1135438150070	Hijo	Telémaco Pardo Hernandez	1984-01-08	M	Divorciado	Ninguna
17	Tarjeta Identidad	1135438153880	Hijo	Manolo Portillo Villareal	1995-02-08	M	Soltero	Ninguna
18	Cedula	1135438155030	Hermano	Rigoberto Portillo Navarrete	1946-06-16	M	Soltero	Motora
19	Cedula	1135438157450	Hermana	Egeria Rincón Correa	1975-02-22	F	Vludo	Ninguna
20	Cedula	1135438159850	Hijo	Refugio Flores Rivas	1988-01-06	M	Vludo	Ninguna
21	Cedula	1135438161090	Hija	Matierna Flores Galván	1986-10-09	F	Divorciado	Ninguna
22	Cedula	1135438162180	Hermana	Joséfa Flores Peña	1932-06-30	F	Soltero	Ninguna
23	Cedula	1135438163250	Hermano	Telmo Flores Peña	1921-03-19	M	Vludo	Ninguna
24	Cedula	1135438164380	Esposo	Mahoma Arenas Barrera	1938-10-18	M	Casado	Ninguna

**Ilustración 6. Información de la tabla beneficiario**

Por otra parte, el archivo “*medicamentos.xls*” debe tener la estructura que se muestra en la **Ilustración 7.**

Código	Nombre Genérico	Forma Farmacéutica	Presentación	Laboratorio y Registro	Precio
1012	ACICLOVIR	Crema	50mg/g Envase con 10mg	MAVER 225M2001 SSA	3500,00
1019	ACICLOVIR	Crema	50mg/g Envase con 2mg	MAVER 225M2001 SSA	3812,00
1026	ACICLOVIR	Crema	50mg/g Envase con 5mg	MAVER 225M2001 SSA	4068,60
			250mg Envase con 5 frascos ámpula con iofilizado		
1033	ACICLOVIR	Solución inyectable		PRECIMEX 220M92 SSA	4280,88
1040	ACICLOVIR	Ungüento	30mg/g Envase con 4.5mg	PHARMACOS EXAKTA 248M2002 SSA	4457,70
	ACIDO				
1047	ACETILSALICÍLICO	Tabletas	300mg Envase con 20 tabletas	FARMAQRO 219M2002 SSA	4606,16
1054	ACIDO FOLICO	Tabletas	5mg Envase con 92 tabletas	VALDECASAS 82231 SSA	4731,93
1061	ACIDO FOLICO	Tabletas	5mg Envase con 375 tabletas	VALDECASAS 82231 SSA	4839,54
1068	ACIDO FOLICO	Tabletas	400µg Envase con 30 tabletas	VALDECASAS 045V2000 SSA	4932,64
1075	ACIDO FOLICO	Tabletas	400µg Envase con 90 tabletas	VALDECASAS 045V2000 SSA	5014,11
1082	ACIDO FOLICO	Tabletas	400µg Envase con 375 tabletas	VALDECASAS 045V2000 SSA	5086,29
1089	ACIDO FOLINICO	Solución inyectable	3mg/ml Envase con 6 ampolletas	PROQUIGAMA 354M90 SSA	5151,03
1096	ACIDO FOLINICO	Tabletas	15mg Envase con 12 tabletas	TECNOFARMA 288M2001 SSA	5209,82
1103	ÁCIDO RETINOICO	Crema	0.050g/100g Envase con 20g	BIORESEARCH 490M2000 SSA	5263,86
1110	ÁCIDO RETINOICO	Crema	0.050g/100g Envase con 30g	BIORESEARCH 490M2000 SSA	5314,09
				INDUSTRIAS QUÍMICO FARMACÉUTICAS AMERICANAS 345M99 SSA	
1117	ÁCIDO RETINOICO	Crema	0.050g/100g Envase con 20g		5361,27
1124	AGUA INYECTABLE	Solución inyectable	5ml Envase con 100 ampolletas	CARBEL 325M2002 SSA	5406,02
1131	AGUA INYECTABLE	Solución inyectable	10ml Envase con 100 ampolletas	CARBEL 325M2002 SSA	10000,00
1138	ALBENDAZOL	Suspensión oral	2g/100ml Envase con 20ml	ALPHARMA 510M2000 SSA	9131,00
1145	ALBENDAZOL	Suspensión oral	2g/100ml Envase con 60ml	VALDECASAS 108M92 SSA	8442,80
1152	ALOPURINOL	Tabletas	300mg Envase con 20 tabletas	VALDECASAS 72651 SSA	7899,24

Ilustración 7. Estructura del archivo de medicamentos

### 3.3. Diseño del Data Warehouse

Para implementar nuestra solución *Business Intelligence*, crearemos nuestra bodega de datos, la cual contendrá el conjunto de *data marts* que usaremos para dar respuesta a las necesidades planteadas en el caso de estudio (**ver sección 3.1**).

Ya tenemos la fuente de información, ahora seguiremos los siguientes pasos para el diseño de cada uno de nuestros *data marts*.

1. Seleccionar un proceso de negocio a modelar
2. Escoger el nivel de granularidad del proceso.
3. Seleccionar las dimensiones que se aplicarán a los hechos.
4. Escoger los hechos medibles que poblarán cada tabla de hechos.

#### 3.3.1. Seleccionar el proceso de negocio a modelar

Para seleccionar el proceso de negocio a modelar, primero vamos a revisar, clasificar e identificar los requerimientos solicitados en el caso de estudio.

#### Preguntas...

- ✓ ¿Qué procesos de negocio identificas en los requerimientos? (**ver sección 3.1**)
- ✓ Clasifica los requerimientos de acuerdo al proceso de negocio al que pertenecen.

- 
- Tiempo promedio de espera desde asignación de cita hasta atención. (**citas médicas**)
  - Tiempos de espera en cirugías. (**cirugías**)
  - Estudio de ocupación de hospitales y médicos dependiendo de épocas del año (navidad, vacaciones, semana santa). (**hospitalizaciones**)
  - Centros Hospitalarios que más atienden pacientes por región, ciudad. (**atención de pacientes**)
  - Promedio de utilización de servicios por paciente, sexo, edad. (**atención de pacientes**)
  - Épocas del año más atención de más espera, de más enfermos. (**atención de pacientes**)
  - Días de la semana más demanda de servicio. (**atención de pacientes**)
  - Regiones del país con más demanda de servicios. (**atención de pacientes**)
- 
- Medicinas más recetadas en un período dado. (**formulas medicas**)
  - Medicamentos que se formulan juntos. (**formulas medicas**)
  - Medicinas que se formulan con más frecuencia. (**formulas medicas**)
  - Costo promedio de recetas médicas. (**formulas medicas**)
- 

Luego de la clasificación encontramos los siguientes procesos de negocio:

- Atención de pacientes (citas médicas, hospitalizaciones, cirugías)
- Formulas medicas

Vamos a seleccionar el primer proceso de negocio *Atención de pacientes*, el cual incluye los servicios *citas médicas*, *hospitalizaciones* y *cirugías*.

### 3.3.2. Escoger nivel de granularidad del proceso

Una vez escogido el proceso de negocio, ahora vamos a identificar dentro de los requerimientos el nivel de granularidad solicitado.

- Tiempo promedio de espera desde asignación de cita hasta atención. (**citas médicas**)
- Tiempos de espera en cirugías. (**cirugías**)
- Estudio de ocupación de hospitales y médicos dependiendo de épocas del año (navidad, vacaciones, semana santa). (**hospitalizaciones**)
- Centros Hospitalarios que más atienden pacientes por región, ciudad. (**atención de pacientes**)
- Promedio de utilización de servicios por paciente, sexo, edad. (**atención de pacientes**)
- Épocas del año más atención de más espera, de más enfermos. (**atención de pacientes**)
- Días de la semana más demanda de servicio. (**atención de pacientes**)
- Regiones del país con más demanda de servicios. (**atención de pacientes**)

Como nos podemos dar cuenta, el grado de granularidad que nos piden en los requerimientos para el proceso de negocio atención de pacientes, es el mismo registro de atención bien sea una cita médica, una hospitalización o una cirugía.

### 3.3.3. Seleccionar las dimensiones que se van a aplicar

Para seleccionar las dimensiones que se va a aplicar, nuevamente vamos a identificar dentro de los requerimientos cuales son:

- *Tiempo* promedio de espera desde asignación de cita hasta atención.
- *Tiempos* de espera en cirugías.
- Estudio de ocupación de hospitales y médicos dependiendo de *épocas del año* (navidad, vacaciones, semana santa).
- Centros Hospitalarios que más atienden pacientes *por región, ciudad*.
- Promedio de utilización de servicios *por paciente*, sexo, edad.
- *Épocas del año* más atención de más espera, de más enfermos.
- *Días de la semana* más demanda de servicio.
- *Regiones del país* con más demanda de servicios.

Palabras como *tiempo, épocas del año* y *días de la semana* hacen referencia a el lugar en tiempo en el cual se efectuó la atención, la cual podríamos llamar y/o identificar como *dimensión fecha*.

Palabras como *regiones del país* y *ciudad*, hacen referencia a la ubicación geográfica en la cual se prestó la atención, la cual podríamos llamar *dimensión ubicación*.

Por último, la palabra *por paciente*, hace referencia a los usuarios que se atendieron, la cual llamaremos *dimensión persona*.

Luego de nuestro análisis de requerimientos identificamos las *dimensiones fecha, ubicación* y *persona*, sin embargo, ahora debemos analizar el proceso de negocio, para identificar otras dimensiones que si bien, dentro de los requerimientos actuales no se utilizan, no quiere decir que en un futuro no puedan existir otros requerimientos que las soliciten. Analizando el proceso de negocio *atención de pacientes* podemos deducir lo siguiente:

- La atención se realiza en una IPS
- La atención la brinda un médico.
- Se requiere saber qué servicio se está prestando al paciente (cita médica, hospitalización o cirugía).

De acuerdo a lo anterior vemos la necesidad de incluir la *dimensión IPS, médico* y *servicio*.

Al finalizar tenemos las siguientes dimensiones:

- Fecha
- Ubicación
- Persona
- IPS
- Médico
- Servicio

## Pregunta...

- ✓ ¿Cómo definimos la información que debe de contener cada una de ellas?

Pues bien, ahora vamos a analizar cada una de ellas.

### Dimensión Fecha

Sabemos que la *dimensión fecha* es una de las más importantes, ya que estará presente en casi todos los *data marts*. Información como día de la semana, mes del año, año, trimestre del año, fecha en todos los formatos, etc., nos permitirá realizar análisis mucho más enriquecidos. En la **ilustración 8** se muestra un ejemplo de la información que debería tener la dimensión fecha.

The screenshot shows a database schema browser window with a table named 'fecha'. The table has the following attributes:

keyfecha	INT
date	DATE
year	INT
month	INT
day	INT
day_of_week	INT
week_of_month	INT
day_of_year	INT
week_of_year	INT
date_m_dd	VARCHAR(45)
date_yyyymm_dd	VARCHAR(45)
week_of_year_iso8601	VARCHAR(...)
year_for_week_iso8601	VARCHAR(...)
date_iso	DATE
year_for_week	VARCHAR(10)
quarter	INT
day_of_week_short_label_eng	VARCHAR(...)
day_of_week_long_label_eng	VARCHAR(...)
month_short_label_eng	VARCHAR(...)
month_long_label_eng	VARCHAR(...)
day_in_month	INT
days_in_year	INT
days_in_quarter	INT
season	VARCHAR(45)
is_leap_year	BOOLEAN
is_weekend	BOOLEAN
is_holiday	BOOLEAN
name_holiday	VARCHAR(45)

Ilustración 8. Atributos dimensión fecha

## Dimensión Ubicación

De acuerdo a los requerimientos, podríamos decir que sus atributos, son país, departamento y municipio, como se muestra en la **ilustración 9**.

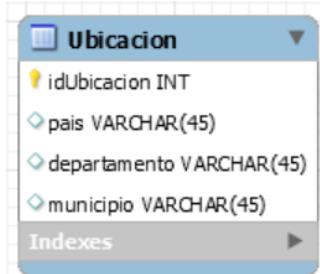


Ilustración 9. Atributos dimensión ubicación

## Dimensión Persona

De acuerdo al caso de estudio, sabemos que tenemos dos tipos de personas (cotizantes y beneficiarios), por lo cual vamos a revisar sus atributos en la base de datos fuente, y así, saber que campos debemos incluir en esta dimensión.

Tabla Beneficiario	Tabla Cotizante
<ul style="list-style-type: none"><li>beneficiario<ul style="list-style-type: none"><li>Columns (8)<ul style="list-style-type: none"><li>tipo_identificacion</li><li>id_beneficiario</li><li>parentesco</li><li>nombre</li><li>fecha_nacimiento</li><li>sexo</li><li>estado_civil</li><li>tipo_discapacidad</li></ul></li></ul></li></ul>	<ul style="list-style-type: none"><li>cotizante<ul style="list-style-type: none"><li>Columns (14)<ul style="list-style-type: none"><li>cedula</li><li>nombre</li><li>tipo_cotizante</li><li>direccion</li><li>estado_civil</li><li>sexo</li><li>fecha_nacimiento</li><li>nivel_escolaridad</li><li>extracto</li><li>proviene_otra_eps</li><li>salario_base</li><li>fecha_afiliacion</li><li>tipo_discapacidad</li><li>id_ips</li></ul></li></ul></li></ul>

Podemos verificar que campos se repiten y que campos podríamos crear para garantizar la integridad de la información por lo cual podemos seleccionar los siguientes atributos:

- tipo\_identificacion
- numero\_identificacion

- nombre
- fecha\_nacimiento
- sexo
- estado\_civil
- tipo\_discapacidad
- tipo\_usuario (cotizante o beneficiario)
- grupo\_familiar (si es beneficiario, se coloca el documento del cotizante a que pertenece, en caso contrario, se coloca el documento del cotizante)

Al finalizar, la *dimensión persona* se definiría como se muestra en la **Ilustración 10**.

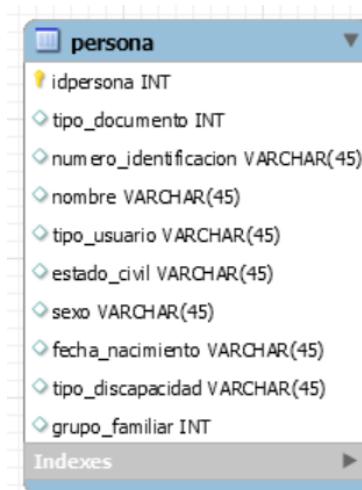


Ilustración 10. Atributos dimensión persona

## Dimensión IPS

Para definir los atributos para esta dimensión debemos revisar en nuestra base de datos fuente, la tabla IPS.

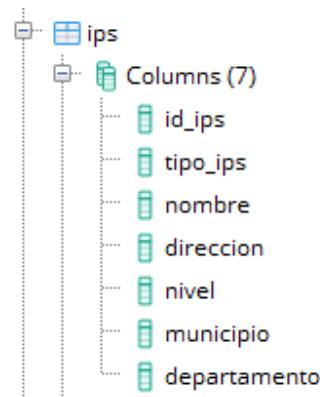


Ilustración 11. Estructura de la tabla IPS

Por lo anterior se puede definir la *dimensión IPS* como se muestra en la **Ilustración 12**.

ips	
idips	INT
id_externo_ips	INT
tipo_ips	VARCHAR(45)
nombre	VARCHAR(45)
direccion	VARCHAR(45)
nivel	VARCHAR(45)
municipio	VARCHAR(45)
departamento	VARCHAR(45)
Indexes	

Ilustración 12. Atributos de la dimensión IPS

**Nota:** Podemos ver que dentro de los atributos de esta dimensión ya se encuentran lo que definimos en la *dimensión ubicación*, por lo cual tenemos dos opciones, eliminamos estos de la *dimensión IPS* o eliminamos la *dimensión ubicación*, para ambos casos, cualquiera de las dos opciones es válida, nosotros por efectos de ahorro de tiempo y trabajo escogeremos la segunda, eliminaremos la *dimensión ubicación*.

## Dimensión Médico

Al igual que hicimos para la dimensión IPS, vamos a revisar la tabla “*medico*” en la base de datos fuente.

medico	
	Columns (7)
	cedula
	nombre
	especialidad
	subespecialidad
	licencia
	direccion_consultorio
	id_ips

Ilustración 13. Estructura de la tabla médico

Por lo anterior definimos la *dimensión médico* como se muestra en la **ilustración 14**.

medico	
idmedico	INT
id_externo_medicoo	INT
nombre	VARCHAR(45)
especialidad	VARCHAR(45)
subespecialidad	VARCHAR(45)
licencia	VARCHAR(45)
direccion	VARCHAR(45)
Indexes	

Ilustración 14. Dimensión médico

## Dimensión Servicio

Como esta dimensión surge a partir de la necesidad de lograr identificar el servicio que se está prestando (cita médica, hospitalización o cirugía), esta se podría definir como se muestra en la **Ilustración 15**.

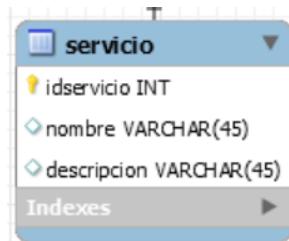


Ilustración 15. Atributos dimensión servicio

Ya podemos ir visualizando el diseño de nuestro *data mart* como se muestra en la **Ilustración 16**.

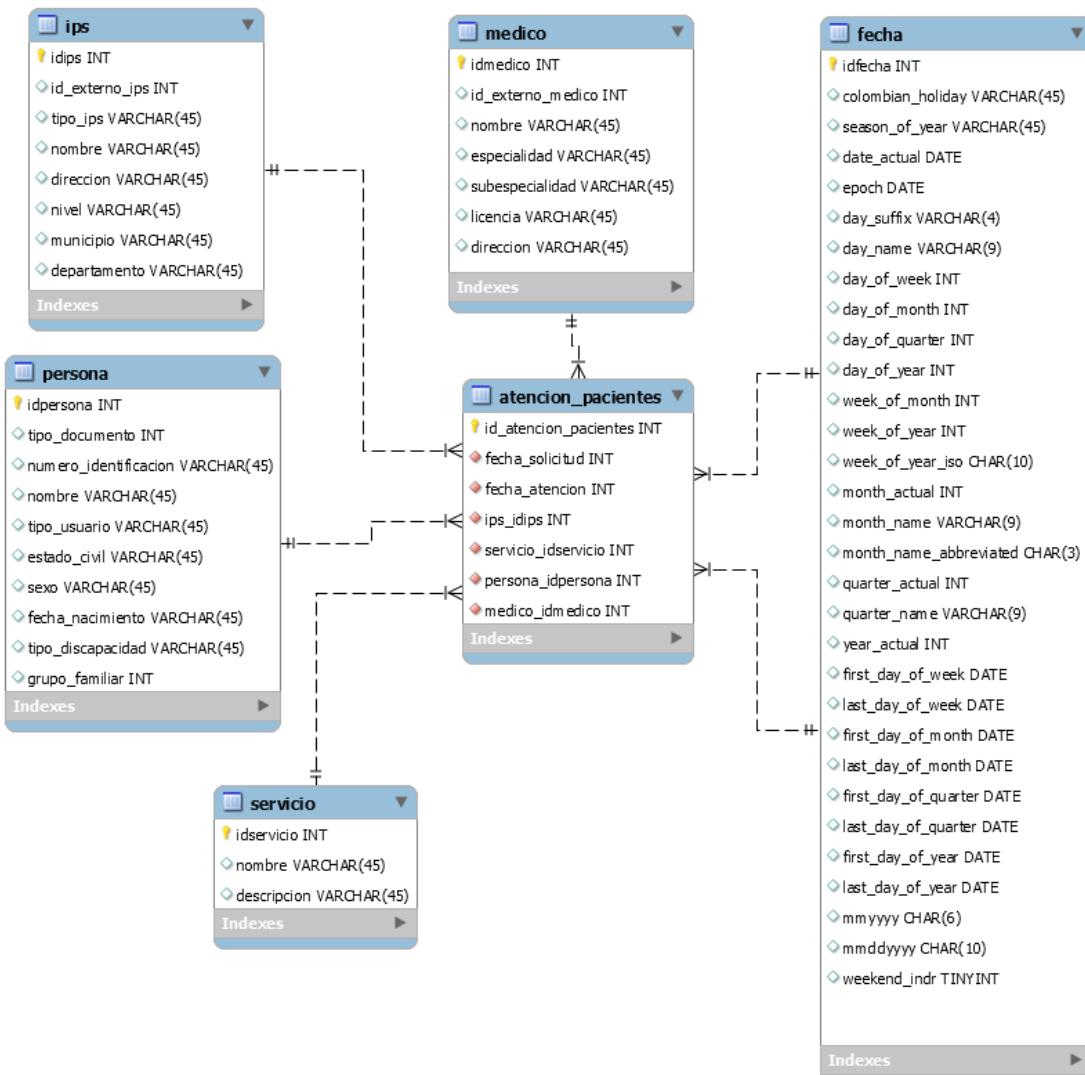


Ilustración 16. Diseño data mart

### 3.3.4. Escoger los hechos medibles

Ahora analizaremos los requerimientos para identificar los hechos medibles de nuestro proceso de negocio:

**Tiempo promedio de espera desde asignación de cita hasta atención:** para poder calcular el tiempo promedio de espera, primero debemos tener el *tiempo de espera* de la cita médica, la cual podemos calcular con la diferencia entre la fecha de atención y fecha de asignación, este tiempo lo debemos guardar, para así, poder realizar el cálculo del promedio de una manera más eficiente.

**Tiempos de espera en cirugías:** igual que el anterior

**Estudio de ocupación de hospitales y médicos dependiendo de épocas del año (navidad, vacaciones, semana santa):** para este requerimiento podemos ver que es una agrupación por el criterio de época del año, por lo cual, dentro de la dimensión fecha debemos de incluir este campo.

**Centros Hospitalarios que más atienden pacientes por región, ciudad:** este es una agrupación por hospital, región y ciudad.

**Promedio de utilización de servicios por paciente, sexo, edad:** para este requerimiento podríamos almacenar el total de utilización de servicios por paciente, para luego sacar el promedio, pero que pasa si luego nos indican que debemos calcular ese promedio de acuerdo a un rango de fecha específico, ya el cálculo almacenado no nos serviría, por lo cual, lo mejor sería realizar el cálculo desde cero.

**Épocas del año más atención de más espera, de más enfermos:** es un total de cantidad de servicios prestados por épocas del año.

**Días de la semana más demanda de servicio:** es un total de cantidad de servicios prestados por días de la semana.

**Regiones del país con más demanda de servicios:** Es un total de cantidad de servicios prestados por regiones del país.

Luego de analizar cada uno de los requerimientos, podemos concluir que el único hecho medible para nuestro data mart, sería el *tiempo de espera*, el cual lo calculamos mediante la resta de la fecha de atención y la fecha de asignación, para los demás solo con realizar la consulta tendríamos la información requerida.

Incluimos el campo indicado en la tabla de hechos, *atención pacientes*, como se muestra en la **ilustración 17.**

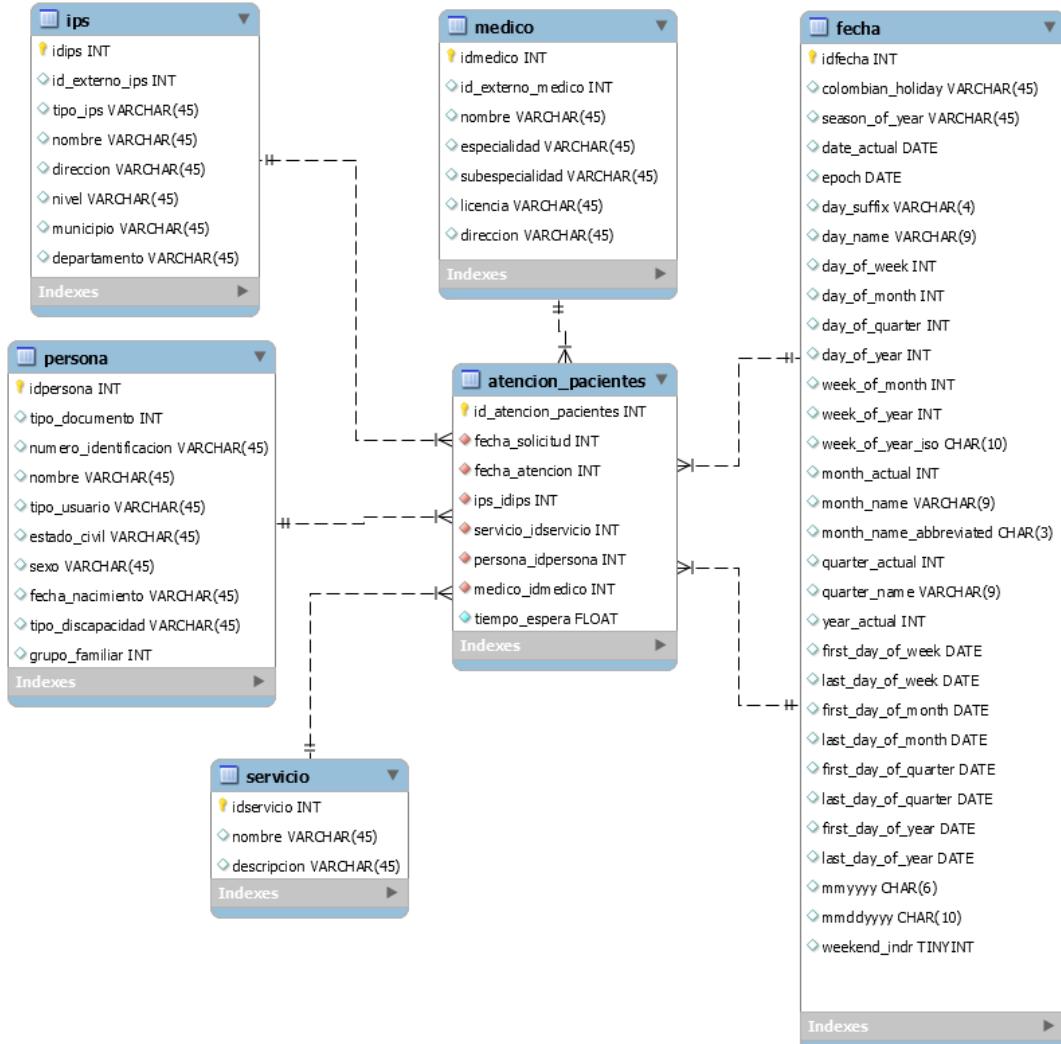


Ilustración 17. Data mart atención de pacientes

### 3.4. Implementación del Data Warehouse

Ya tenemos listo nuestro *data mart* diseñado para el proceso de negocio *atención de pacientes*, ahora vamos a implementarlo utilizando el *proceso ETL* mediante la herramienta *Pentaho Data Integration*.

Primero debemos crear la base de datos donde crearemos nuestra *bodega de datos*, para eso, ingresamos al pgAdmin, desde Databases -> clic derecho -> Create -> Database, como se muestra en la **ilustración 18**.

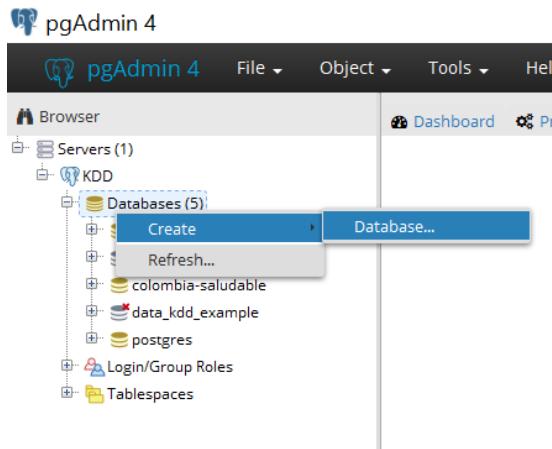


Ilustración 18. Crear base de datos

Creamos la base de datos como se muestra en la **ilustración 19**.

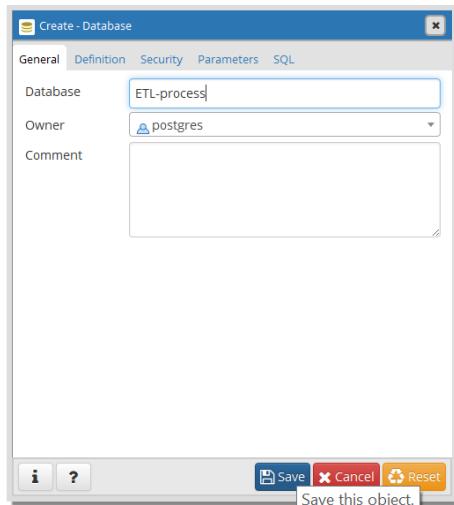
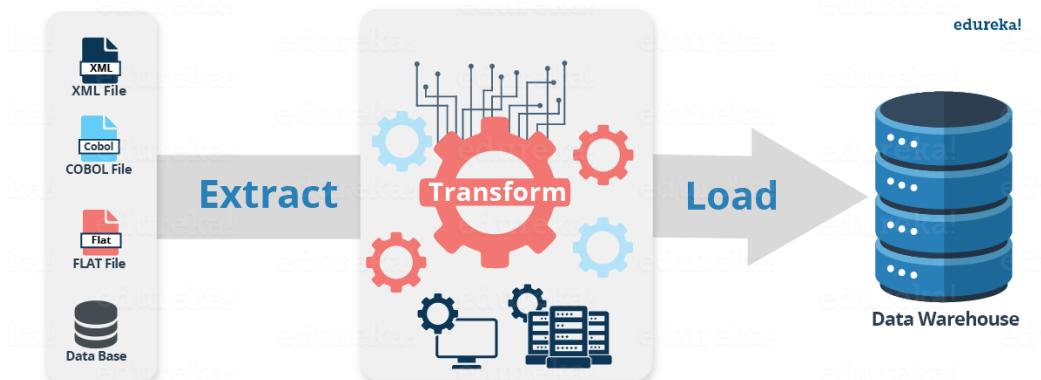


Ilustración 19. Creación de la bodega de datos ETL-process

Una vez creada nuestra bodega de datos, podemos comenzar a crear nuestro *data mart*.

### 3.4.1. Implementación Dimensión IPS

Antes de comenzar con la transformación vamos a hacer un pequeño análisis que lo que vamos a necesitar para crear la *dimensión IPS* de acuerdo al diseño planteado (**ver ilustración 12**). Para ello vamos a recordar el proceso ETL como se muestra en la ilustración.



De acuerdo a la imagen podemos identificar los pasos que debemos seguir para su implementación:

1. Identificar las fuentes de información.
2. Extraer la información de las fuentes.
3. Aplicar técnicas de trasformación para depurar la información con el fin de que coincida con la estructura de carga.
4. Cargar la información a los data mart.
5. Bodega de datos implementada.

#### 3.4.1.1. Identificar las fuentes de información.

Ingresamos a *Pentaho Data Integration* y creamos una nueva transformación como se muestra en la ilustración 21.

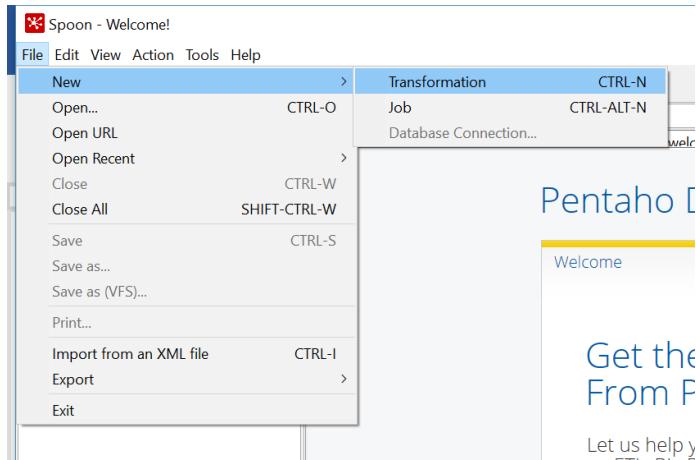
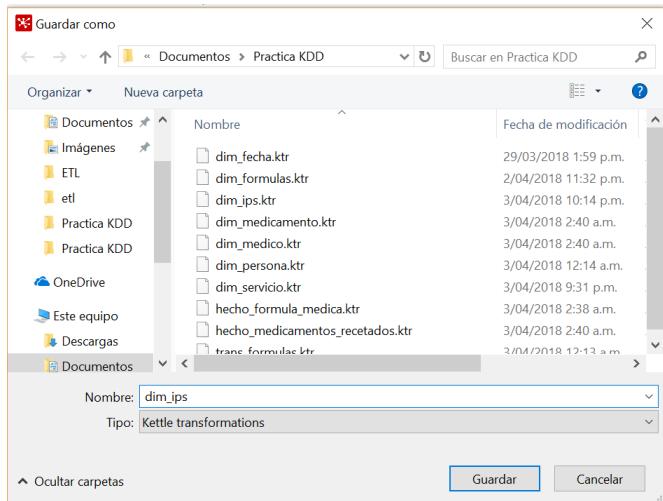


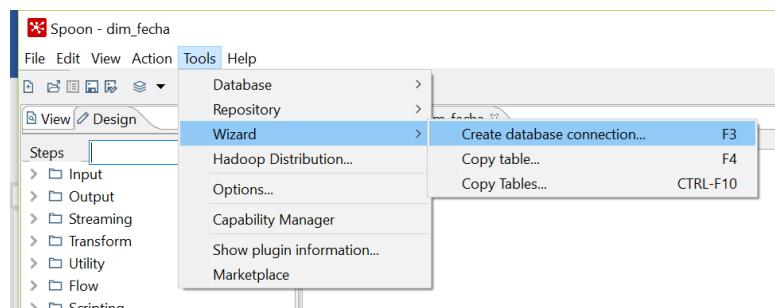
Ilustración 21. Creación de una transformación

Ahora vamos a guardarla, desde archivo-> guardar como e ingresamos el nombre *dim\_ips* como se muestra en la **ilustración 22**.



**Ilustración 22. Guardar dimensión fecha**

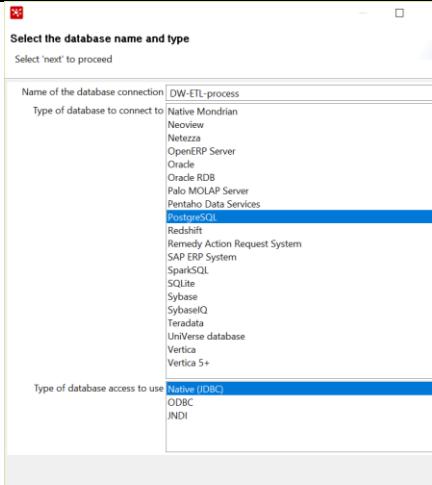
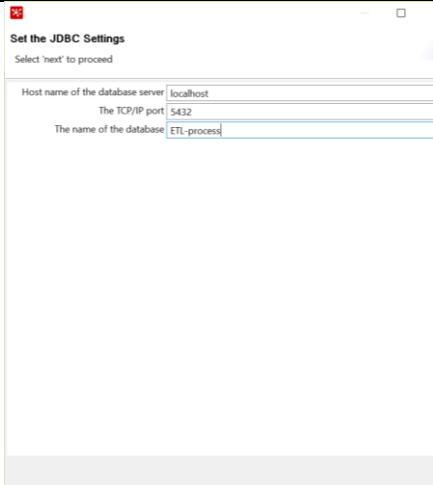
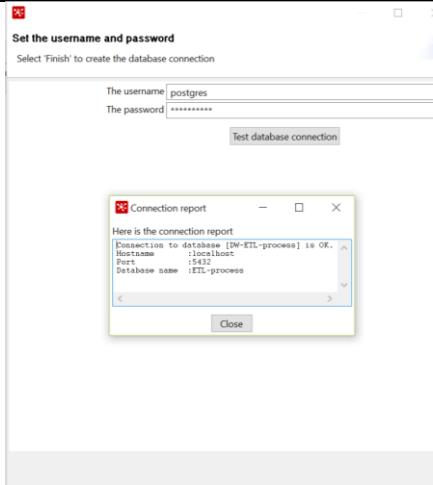
Desde esta vamos a desarrollar la transformación necesaria para crear dicha dimensión. Recordemos que los datos fuente para la creación es la tabla *"IPS"*. Para ello vamos a crear dos conexiones de base de datos, la primera referenciando la base de datos fuente *"colombia-saludable"* y la segunda referenciando nuestra bodega de datos *"ETL-process"*. Ahora vamos a crear la conexión a nuestra bodega de datos como se muestra en la **ilustración 23**.



**Ilustración 23. Crear la conexión a la base de datos**

Luego realizamos la configuración como se muestra en la **tabla 2**.

Tabla 2 Pasos para configurar la conexión a la base de datos

Paso 1	Paso 2	Paso 3
 <p>Seleccionamos el tipo de gestor de base de datos donde se encuentra creada nuestra bodega de datos, en este caso PostgreSQL</p>	 <p>Ingresamos el nombre del host, el puesto de conexión y el nombre de la bodega de datos que creamos anteriormente (<b>ver ilustración 19</b>)</p>	 <p>Para finalizar ingresamos el nombre de usuario con su respectiva contraseña. No olvides realizar el test de conexión para validar su correcto funcionamiento.</p>

### Es tu turno...

- ✓ Sigue los pasos de la **tabla 2** para crear la conexión a la base de datos fuente "*colombia-saludable*"

#### 3.4.1.2. Extraer la información de las fuentes

Ahora desde el panel de diseño en la opción "*Input*" arrastramos la herramienta "*Table input*" como se muestra en la **ilustración 24**.

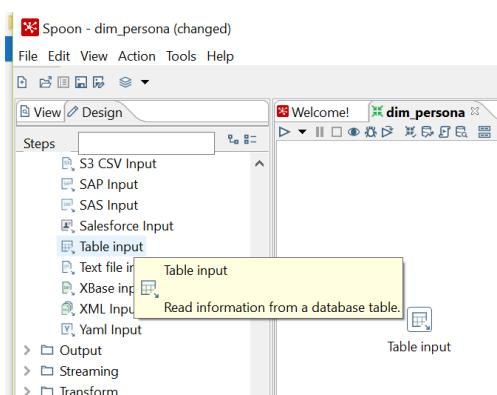
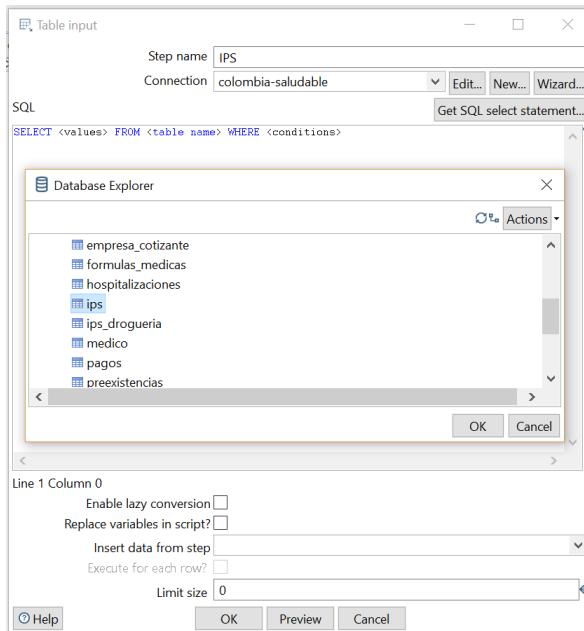


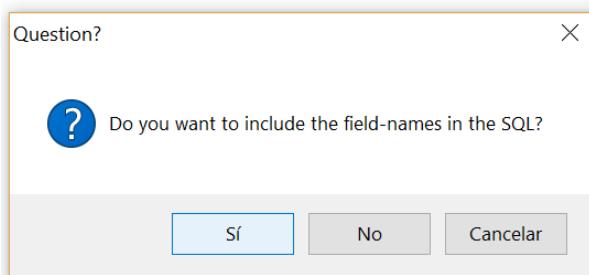
Ilustración 24. Arrastrar herramienta "Table input"

Vamos a usar esta herramienta para traer la información de la tabla fuente “*IPS*”, damos clic derecho sobre el ícono, opción editar, desde la configuración ingresamos el nombre, la conexión a la base de datos fuente, luego damos clic en el botón “*Get SQL select statement*” y seleccionamos la tabla “*IPS*” como se muestra en la **ilustración 25**.



**Ilustración 25.** Obtener información tabla IPS

Damos clic en el botón “*OK*”, nos aparecerá el siguiente mensaje.



**Ilustración 26.** Mensaje de confirmación para incluir campos en la consulta

Esto nos permite incluir los campos que contiene la tabla en la consulta sql generada, con el objetivo de incluir solo aquellos que necesitemos. Al finalizar se debe de ver como se muestra en la **ilustración 27**.

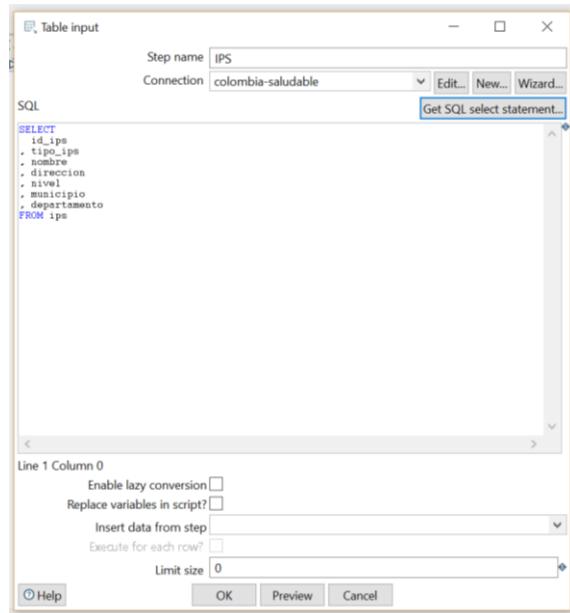


Ilustración 27. Campos de la tabla IPS.

### 3.4.1.3. Aplicar técnicas de transformación

Dese el panel de diseño, en la opción “*Transform*” vamos a arrastrar la herramienta “*Select values*”, como se muestra en la **ilustración 28**.

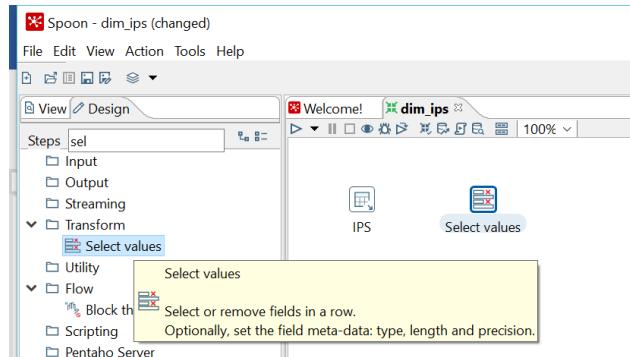
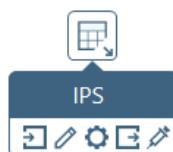
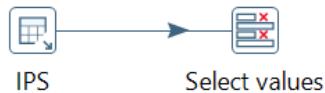


Ilustración 28. Arrastrar herramienta "Select values"

A continuación, debemos unirlos, nos posicionamos en el icono “*IPS*” y esperamos que se visualice como se muestra a continuación.

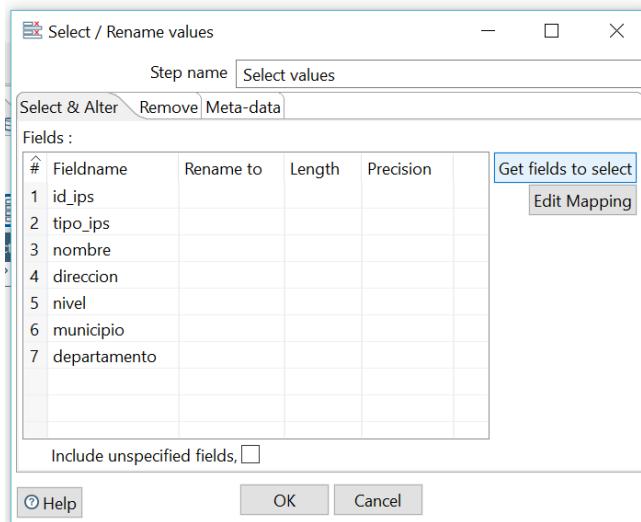


damos clic en y luego damos clic en el icono “*Select values*” al finalizar, ambos procesos deben de quedar enlazados como se muestra en la **ilustración 29**.



**Ilustración 29. Uniendo los dos procesos**

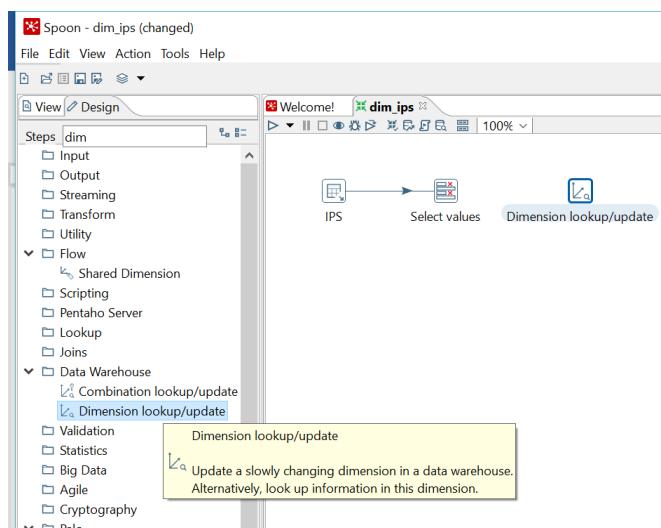
Ahora vamos a configurarla, damos clic en el botón “Get fields to select”, para incluir los campos del paso anterior, como se muestra en la **ilustración 30**.



**Ilustración 30. Configuración proceso "Select values"**

Para finalizar damos clic en el botón “OK”.

Desde el panel de diseño en la opción “*Data Warehouse*”, arrastramos la herramienta “*Dimension lookup/update*”, como se muestra en la **ilustración 31**.



**Ilustración 31. Arrastrar herramienta "Dimension lookup/update"**

Lo unimos desde el proceso anterior “*Select values*” como se muestra a continuación.



Ahora la configuramos como se muestra en la ilustración 32.

La enlazamos desde el paso “*Select values*” y la configuramos como se muestra en la **ilustración 47**.

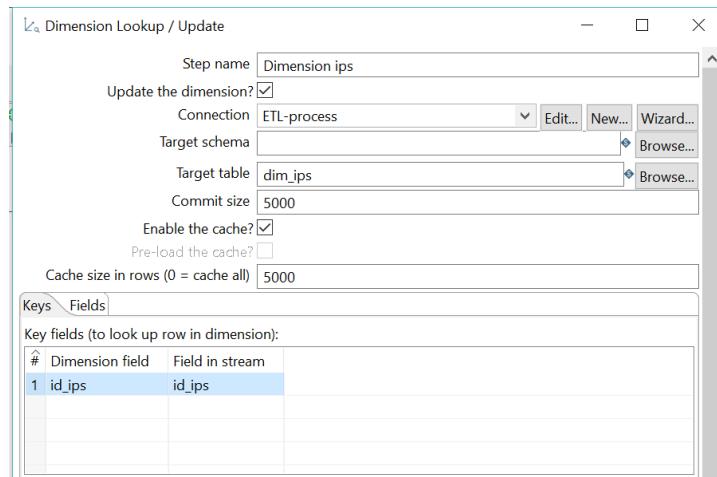


Ilustración 32. Configuración primera parte proceso “Dimension ips”

Ingresamos el nombre del proceso como “*Dimension ips*”, seleccionamos la base de datos en nuestro caso, la base de datos “*ETL-process*”, y le damos un nombre a la dimensión que vamos a crear en la base de datos “*dim\_ips*”. En la pestaña “*Keys*”, incluimos el campo “*id\_ips*”.

Desde la pestaña “*Fields*” damos clic en el botón “*Get Fields*”, como se muestra en la **ilustración 33**.

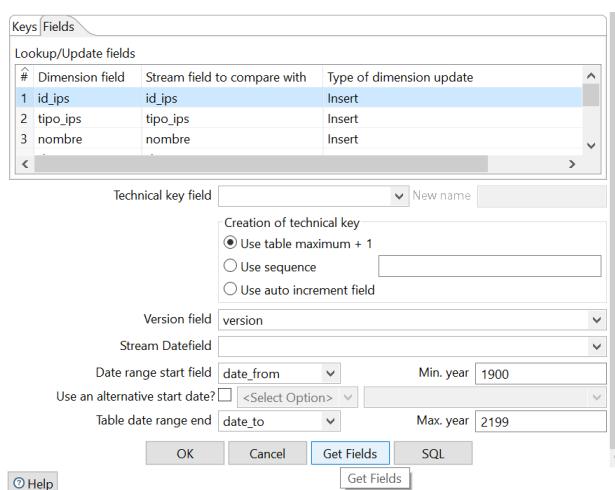
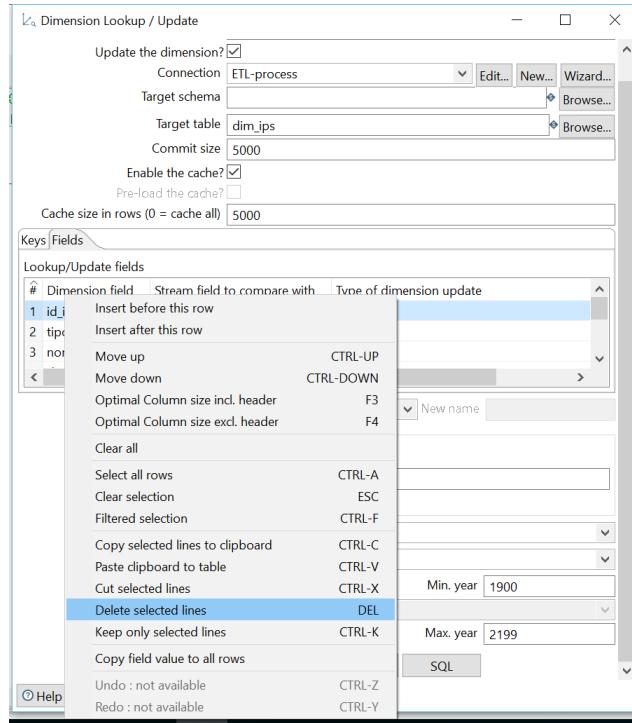


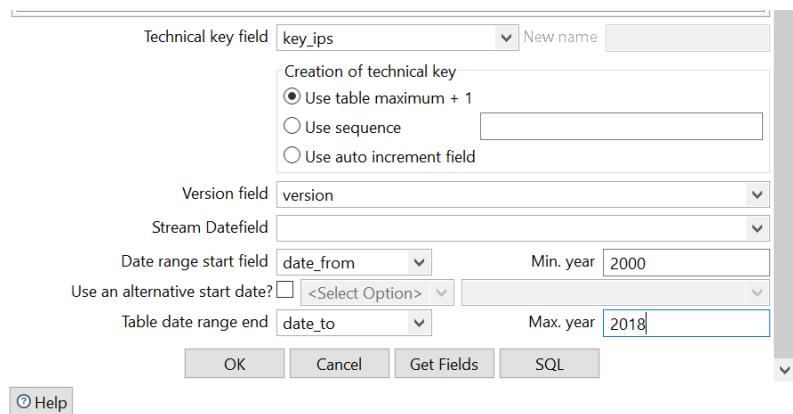
Ilustración 33. Configuración segunda parte proceso “Dimension”

Como ya incluimos el campo "*id\_ips*" en la pestaña "*Keys*", lo debemos eliminar de la pestaña "*Field*" seleccionándolo, damos clic derecho y escogemos la opción "*Delete selected lines*", como se muestra en la **ilustración 34**.



**Ilustración 34. Eliminación del campo "id\_ips"**

Seguimos configurando los campos restantes como se muestra en la **ilustración 50**.



**Ilustración 35. Configuración tercera parte del proceso "Dimension"**

Elegimos el nombre del campo que se creará como identificador de la dimensión "*key\_ips*"

Ahora necesitamos crear el esquema de la dimensión ips, con los campos seleccionados, para ello damos clic en el botón “*SQL*”, lo cual nos permite ver el código sql para crear la tabla en la base de datos, damos clic en el botón “*Execute*”, como se muestra en la **ilustración 36**.

The screenshot shows two windows side-by-side. On the left is the 'Simple SQL editor' window with the title 'SQL statements, separated by semicolon:'. It contains the following SQL code:

```

CREATE TABLE dim_ips
(
    key_ips BIGSERIAL,
    version INTEGER,
    date_from TIMESTAMP,
    date_to TIMESTAMP,
    id_ips VARCHAR(40),
    tipo_ips VARCHAR(40),
    nombre VARCHAR(400),
    direccion VARCHAR(400),
    nivel VARCHAR(4),
    municipio VARCHAR(400),
    departamento VARCHAR(400)
);
CREATE INDEX idx_dim_ips_lookup ON dim_ips(id_ips)
;
CREATE INDEX idx_dim_ips_tk ON dim_ips(key_ips)
;

```

On the right is the 'Results of the SQL statements' window with the title 'The SQL statements had the following results'. It displays the executed SQL statements and the number of statements executed:

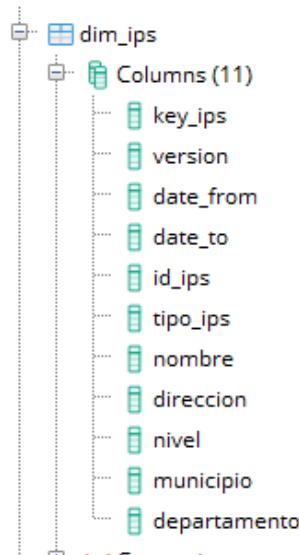
```

SQL executed: CREATE TABLE dim_ips
(
    key_ips BIGSERIAL,
    version INTEGER,
    date_from TIMESTAMP,
    date_to TIMESTAMP,
    id_ips VARCHAR(40),
    tipo_ips VARCHAR(40),
    nombre VARCHAR(400),
    direccion VARCHAR(400),
    nivel VARCHAR(4),
    municipio VARCHAR(400),
    departamento VARCHAR(400)
);
SQL executed: CREATE INDEX idx_dim_ips_lookup ON dim_ips(id_ips)
SQL executed: CREATE INDEX idx_dim_ips_tk ON dim_ips(key_ips)
3 SQL statements executed

```

**Ilustración 36.** Ejecución del SQL para crear la dimensión ips

Podemos verificar en la base de datos la tabla creada como se muestra en la **ilustración 37**.



**Ilustración 37.** Tabla dim\_ips creada en la base de datos

### 3.4.1.4. Cargar Información

Listo ahora vamos a ejecutar nuestra transformación creada, damos clic en el botón  luego damos clic en el botón **"Run"** como se muestra en la **ilustración 38**.

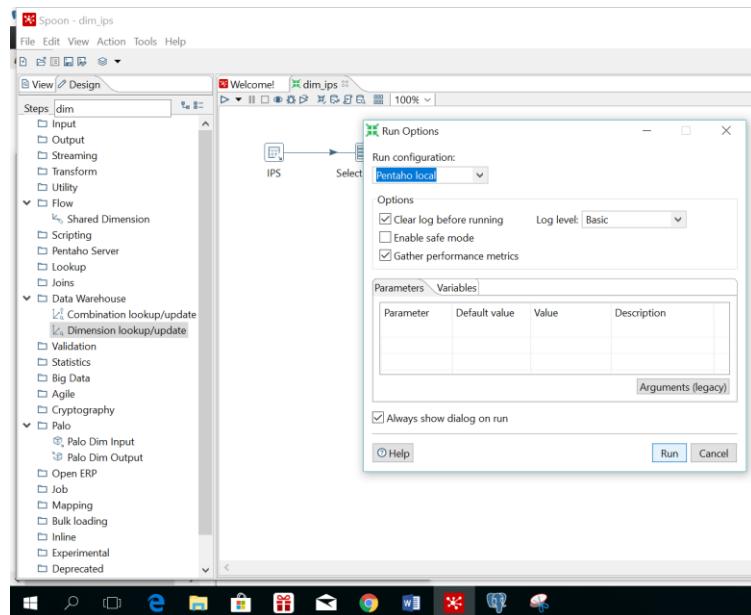


Ilustración 38. Ejecutar transformación dim\_ips

Si no se presenta ningún error, cada uno de los procesos de la transformación al finalizar la ejecución aparecerá con una imagen de chequeo de color verde, además dentro del panel de ejecución aparecerá su resultado de color negro, como se muestra en la **ilustración 39**.

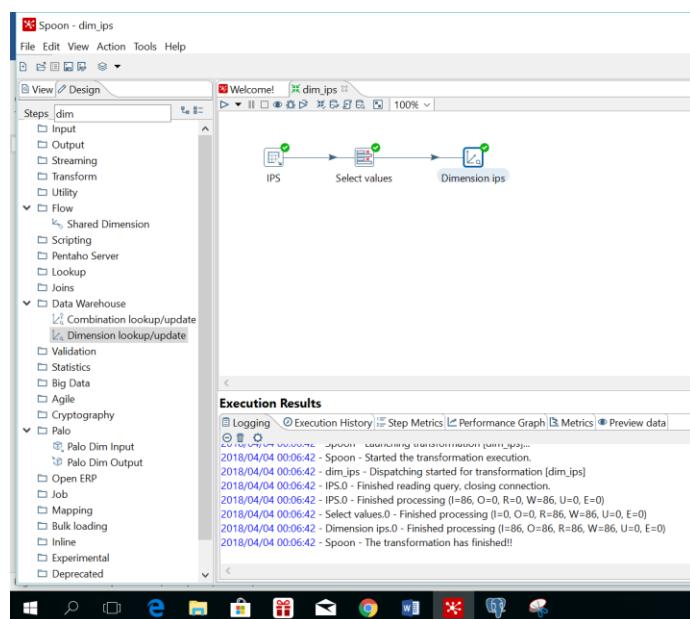
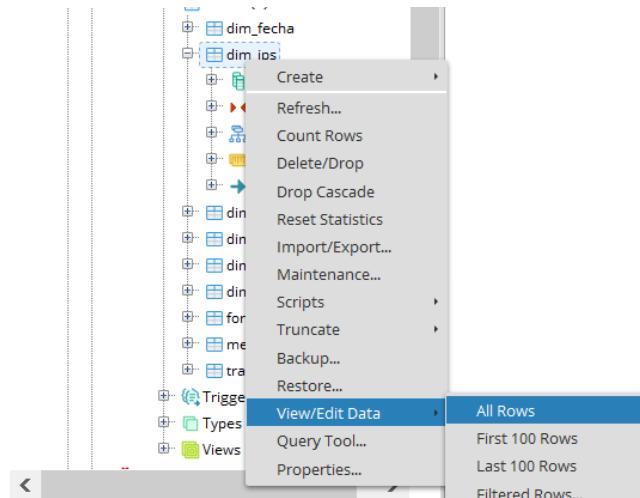


Ilustración 39. Ejecución exitosa de la transformación

### 3.4.1.5. Dimensión Implementada

Podemos verificar en la base de datos que la tabla dim\_ips se encuentra con la información ingresada como se muestra en la **ilustración 40**.



**Ilustración 40.** Visualizar los datos de la tabla dim\_ips

Podemos ver que la tabla se encuentra con la información generada a partir de la trasformación creada para la dimensión fecha. Se debe visualizar la información como se muestra en la **ilustración 41**.

time zone	date_to	id_ips	tipo_ips	nombre	direccion	nivel	municipio	departamento
	timestamp without time zone	[null]	character varying(40)	[null]	character varying(40)	[null]	[null]	[null]
00	2018-12-31 23:59:59.999	IPS_1	Clinica	Clinica de Occidente	Kra 76 # 38-102	[null]	Cali	Valle del Cauca
00	2018-12-31 23:59:59.999	IPS_2	Clinica	Clinica Valle del Lili	Diagonal 98 # 50-59	[null]	Cali	Valle del Cauca
00	2018-12-31 23:59:59.999	IPS_3	Hospital	Hospital Infantil Clinica Noel	Kra 72 # 85-17	3	Cali	Valle del Cauca
00	2018-12-31 23:59:59.999	IPS_4	Hospital	Hospital Universitario del V...	Transversal 94 # 75-74	4	Cali	Valle del Cauca
00	2018-12-31 23:59:59.999	IPS_5	Hospital	Hospital San Juan de Dios	Diagonal 105 # 66-119	1	Cali	Valle del Cauca
00	2018-12-31 23:59:59.999	IPS_6	Hospital	Hospital La Buena Esperanza...	Calle 127 # 79-78	3	Yumbo	Valle del Cauca
00	2018-12-31 23:59:59.999	IPS_7	Hospital	Hospital Isalias Duarte Cencio	Calle 23 # 25-131	2	Cali	Valle del Cauca
00	2018-12-31 23:59:59.999	IPS_8	Centro de Salud	Centro Medico Carlos Ardi...	Diagonal 107 # 87-109	[null]	Cali	Valle del Cauca
00	2018-12-31 23:59:59.999	IPS_9	Centro de Salud	Centro Medico San Ignacio	Kra 27 # 2-131	[null]	Palmita	Valle del Cauca
00	2018-12-31 23:59:59.999	IPS_10	Centro de Salud	Centro Medico Las Americas	Calle 32 # 74-31	[null]	Jamundi	Valle del Cauca
00	2018-12-31 23:59:59.999	IPS_11	Clinica	Clinica La Sabana	Calle 57 # 35-55	[null]	Bogota	Cundinamarca
00	2018-12-31 23:59:59.999	IPS_12	Clinica	Clinica La Inmaculada	Calle 77 # 49-6	[null]	Bogota	Cundinamarca
00	2018-12-31 23:59:59.999	IPS_13	clinica	Clinica Vida	Avenida 117 # 86-45	[null]	Facatativá	Cundinamarca
00	2018-12-31 23:59:59.999	IPS_14	Hospital	Hospital Isalias Duarte Cencio	Kra 3 # 62-58	4	Bogota	Cundinamarca
00	2018-12-31 23:59:59.999	IPS_15	Hospital	Hospital Chapinero	Avenida 110 # 52-81	3	Bogota	Cundinamarca
00	2018-12-31 23:59:59.999	IPS_16	Hospital	Hospital San Cristobal	Kra 100 # 83-29	2	Bogota	Cundinamarca
00	2018-12-31 23:59:59.999	IPS_17	Hospital	Hospital Suba	Transversal 49 # 13-12	1	Bogota	Cundinamarca
00	2018-12-31 23:59:59.999	IPS_18	Hospital	Hospital Pablo Tobon Uribe	Avenida 19 # 102-33	2	Girardot	Cundinamarca
00	2018-12-31 23:59:59.999	IPS_19	Centro de Salud	Centro Medico Meissen	Diagonal 105 # 36-1	[null]	Facatativá	Cundinamarca
00	2018-12-31 23:59:59.999	IPS_20	Centro de Salud	Centro Medico San Rafael	Calle 104 # 83-89	[null]	Bogota	Cundinamarca
00	2018-12-31 23:59:59.999	IPS_21	Centro de Salud	Centro Medico Le Misericord...	Transversal 17 # 31-86	[null]	Bogota	Cundinamarca
00	2018-12-31 23:59:59.999	IPS_22	Clinica	Clinica El Rosario	Kra 63 # 108-102	[null]	Medellin	Antioquia

**Ilustración 41.** Información ingresada en la dimensión fecha

### 3.4.2. Implementación Dimensión Médico

#### Es tu turno...

- ✓ Crea la "*dimensión médico*" a partir del diseño propuesto (**ver ilustración 14**) siguiendo los pasos vistos en la implementación anterior.

Al finalizar debes de tener una transformación similar a la siguiente:



Ilustración 42. Transformación dimensión médica

### 3.4.3. Implementación Dimensión Servicio

Continuamos con la dimensión servicio, sabemos que la vamos a utilizar para diferenciar el tipo de servicio prestado los cuales se encuentran definidos (citas médicas, hospitalizaciones y urgencias). Como su información no proviene de una tabla específica, tendremos que construirla siguiendo el diseño planteado (**ver ilustración 15**).

Desde el panel de diseño, en la opción "*Input*" arrastramos la herramienta "*Data Grid*" como se muestra en la **ilustración 43**.

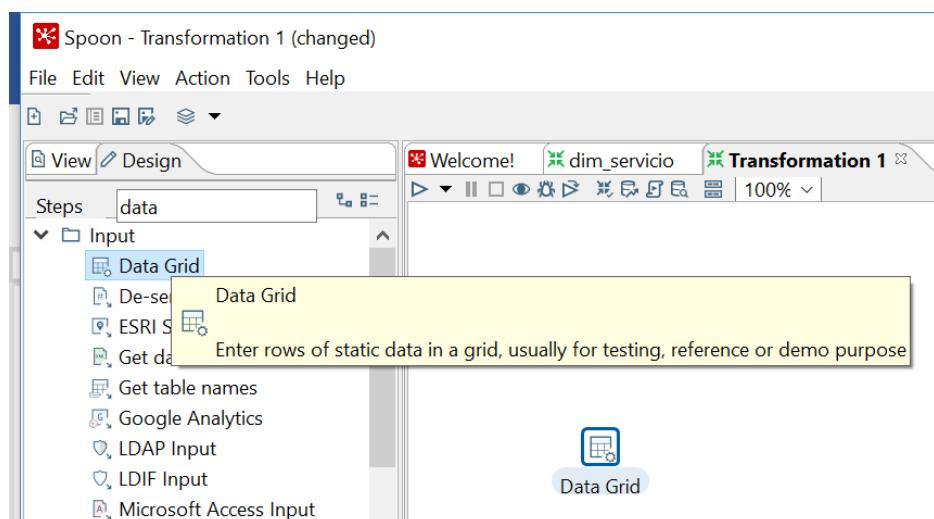
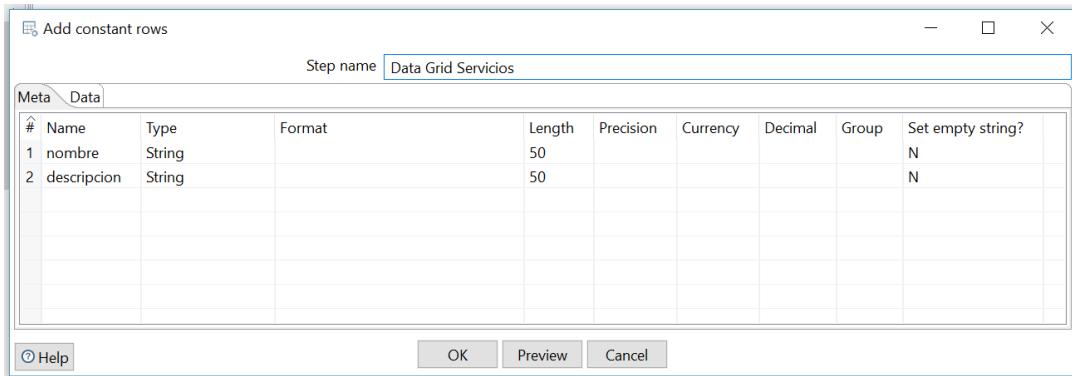


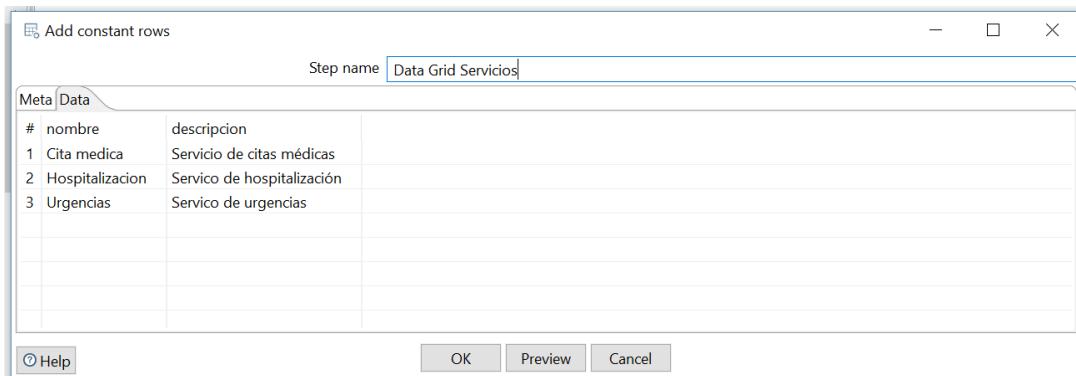
Ilustración 43. Arrastrar herramienta "Data Grid"

Ahora lo configuramos como se muestra en la **ilustración 44**.



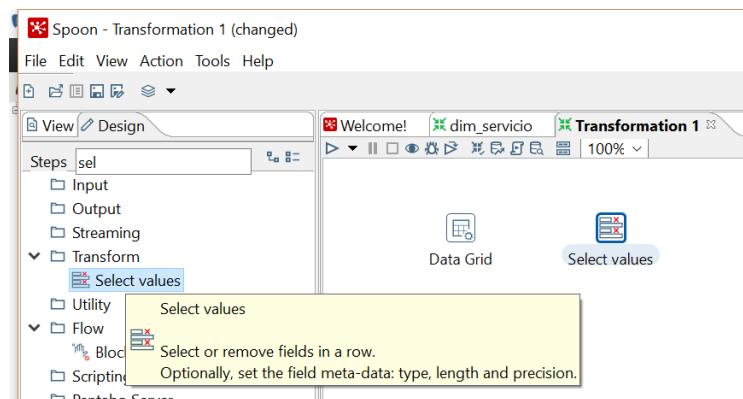
**Ilustración 44. Configuración estructura de la dimensión servicio**

En este paso estamos partiendo de una estructura vacía, la cual estamos definiendo con la estructura de la dimensión que queremos crear con sus atributos "*nombre*" y "*descripcion*" desde la pestaña "*Meta*". En la pestaña "*Data*" vamos a incluir la información requerida como se muestra en la **ilustración 45**.



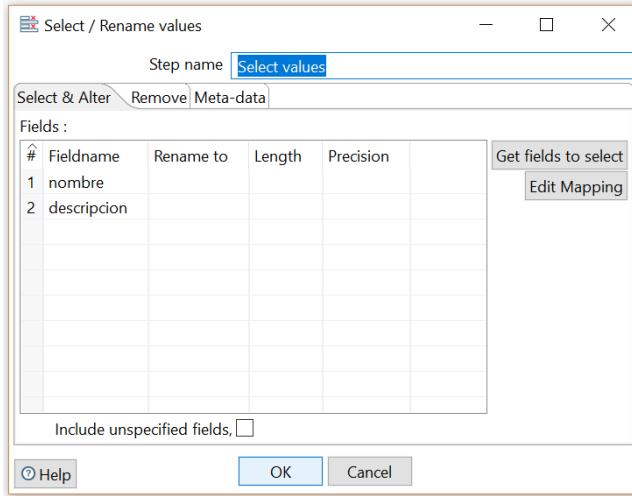
**Ilustración 45. Configuración información dimensión servicio**

Ahora desde el panel de diseño en la opción "*Transform*", arrastramos la herramienta "*Select values*", como se muestra en la **ilustración 46**.



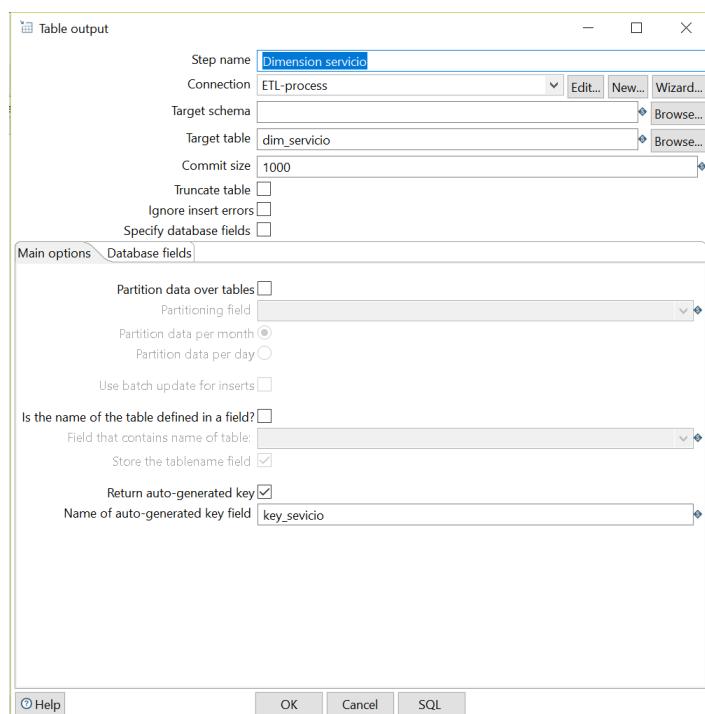
**Ilustración 46. Arrastrar herramienta "Select values"**

Lo unimos desde el paso *"Data Grid"* y lo configuramos desde el botón *"Get fields to select"* como se muestra en la **ilustración 47**.



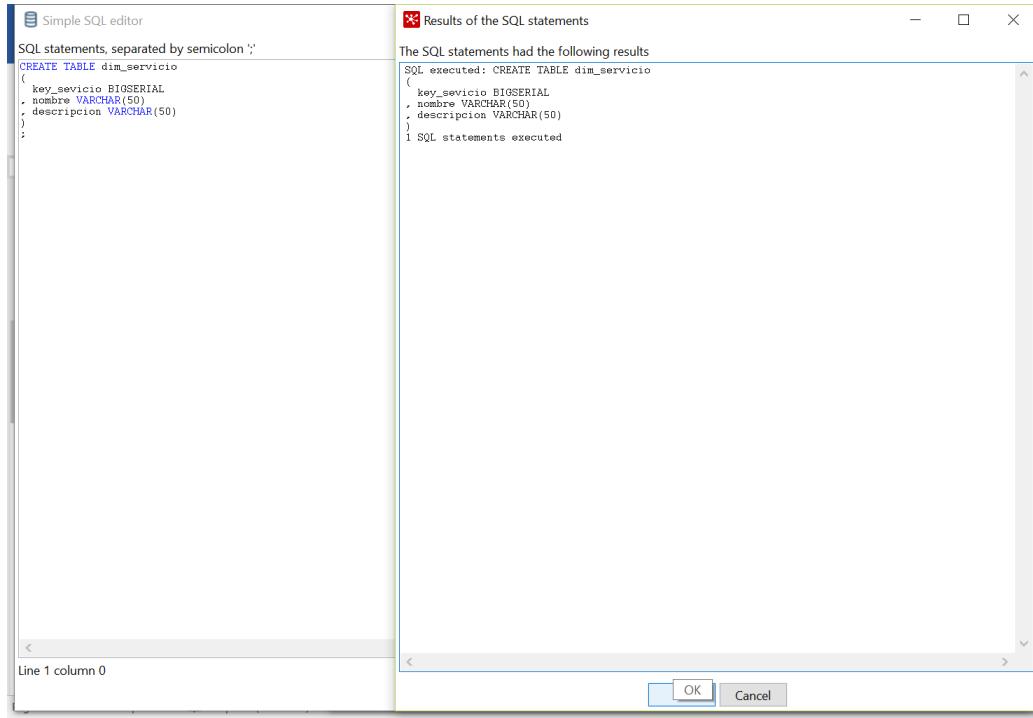
**Ilustración 47. Configuración proceso "Select values"**

Ahora solo nos falta crear la dimensión, para ello desde el panel de diseño en la opción *"Data Warehouse"*, arrastramos la herramienta *"Dimisión lookup/update"*, lo unimos desde el paso *"Select values"* y lo configuramos como se muestra en la **ilustración 48**.



**Ilustración 48. Configuración proceso "Dimensión servicio"**

Ahora generamos y ejecutamos el **"SQL"** como se muestra en la ilustración.



The screenshot shows a 'Simple SQL editor' window on the left and a 'Results of the SQL statements' window on the right. The SQL code in the editor is:

```
SQL statements, separated by semicolon ;
CREATE TABLE dim_servicio
(
    key_servicio BIGSERIAL
    , nombre VARCHAR(50)
    , descripcion VARCHAR(50)
);
;
```

The results window shows the output of the executed SQL statement:

```
SQL executed: CREATE TABLE dim_servicio
(
    key_servicio BIGSERIAL
    , nombre VARCHAR(50)
    , descripcion VARCHAR(50)
)
1 SQL statements executed
```

Buttons at the bottom of the results window are 'OK' and 'Cancel'.

Ilustración 49. Ejecución SQL dimensión servicio

Ya tenemos lista nuestra transformación, ahora la ejecutamos.

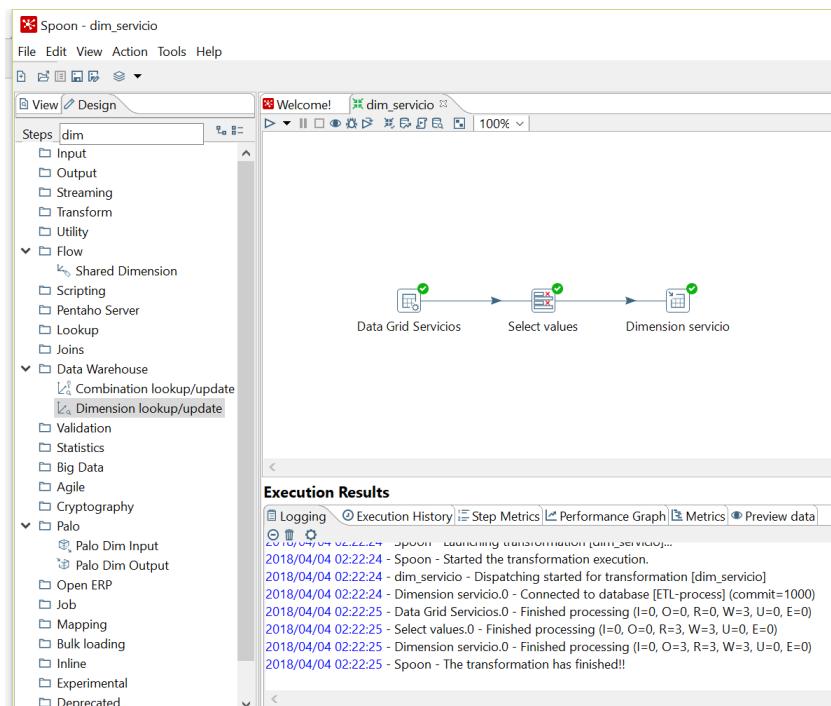


Ilustración 50. Transformación dimensión ejecutada.

### 3.4.4. Implementación Dimensión Persona

Antes de comenzar con la transformación vamos a hacer un pequeño análisis que lo que vamos a necesitar para crear la dimensión persona de acuerdo al diseño planteado (**ver ilustración 10**). No olvidemos los pasos anteriores.

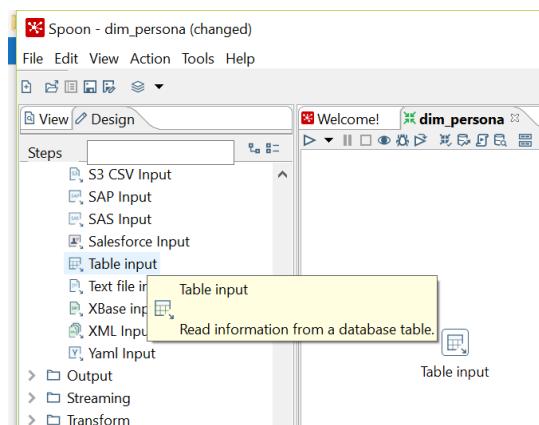
#### 3.4.4.1. Identificar las fuentes de información.

Desde *pentaho* creamos una nueva transformación con el nombre de *dim\_persona*, Recordemos que los datos fuente para la creación de esta dimensión con las tablas "*Cotizante*" y "*Beneficiario*". Para ello vamos a crear dos conexiones de base de datos, la primera referenciando la base de datos fuente "*colombia-saludable*" y la segunda referenciando nuestra bodega de datos "*ETL-process*".

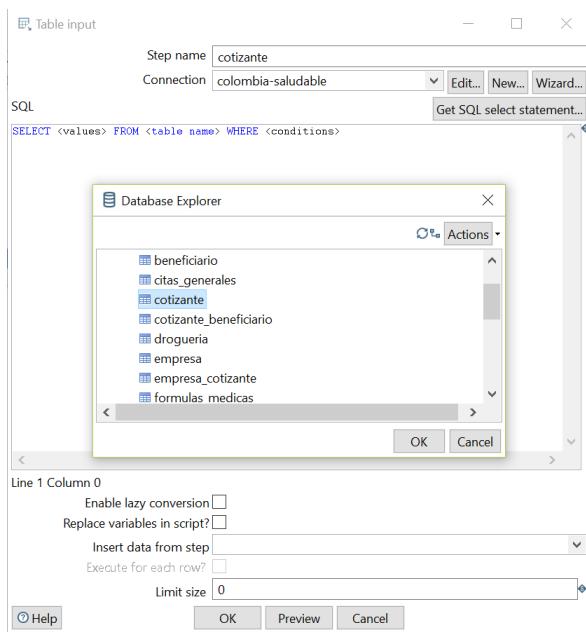
**Nota:** Para crear las conexiones a las bases de datos indicadas, recuerde los pasos vistos en la **Tabla 2**.

#### 3.4.4.2. Extraer la información de las fuentes

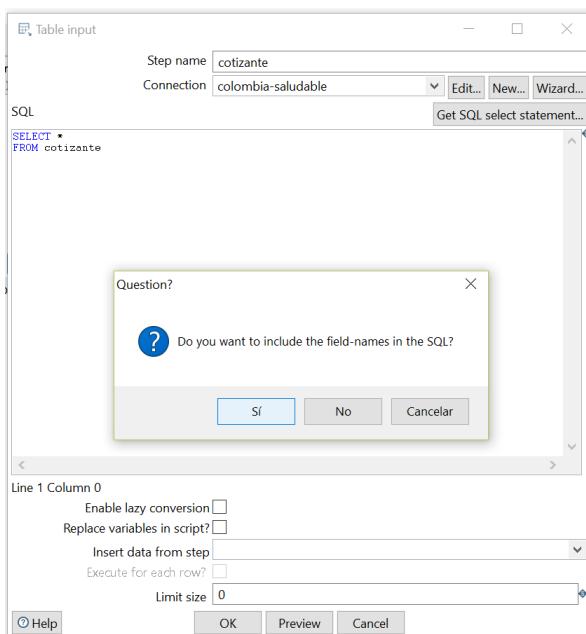
Ahora desde el panel de diseño en la opción "*Input*" arrastramos la herramienta "*Table input*" como se muestra en la ilustración.



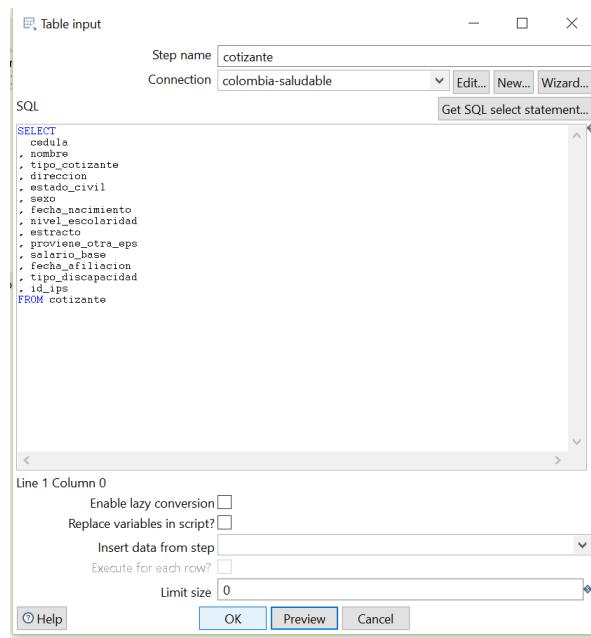
Vamos a usar esta herramienta para traer la información de la tabla fuente cotizante, desde la configuración ingresamos el nombre, la conexión a la base de datos fuente, luego damos clic en el botón "Get SQL select statement" y seleccionamos la tabla cotizante como se muestra en la ilustración.



Damos clic en el botón "OK", nos aparecerá el siguiente mensaje



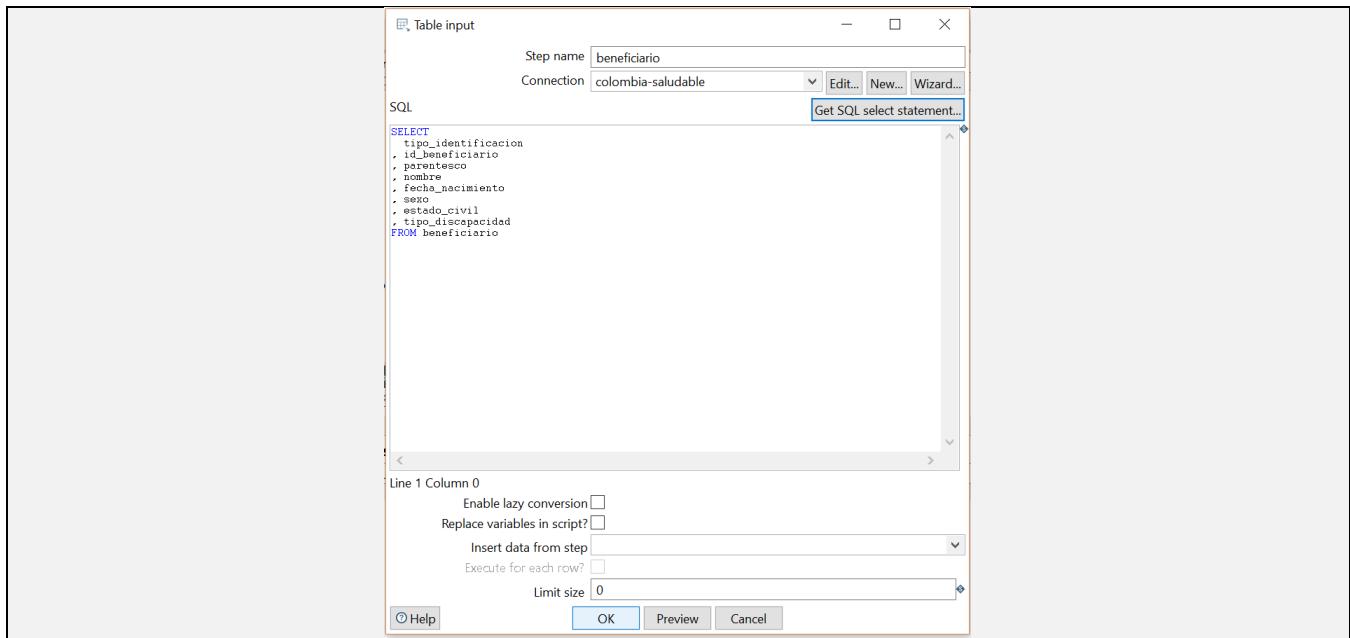
Esto nos permite incluir los campos que contiene la tabla en la consulta sql generada, con el objetivo de incluir solo aquellos que necesitemos, para nuestro caso, de acuerdo al diseño, sabemos que no vamos a utilizar todos, por lo cual vamos a dar clic en el botón "Sí", para en un paso posterior poder configurarlos. Al finalizar de ver como se muestra en la **ilustración**.



**Nota:** No olvide que puede verificar el paso configurado desde el botón "*Preview*".

### Es tu turno...

- ✓ Sigue los pasos anteriores para extraer la información de la tabla "*beneficiario*"



### 3.4.4.3. Aplicar técnicas de trasformación para depurar la información.

Vamos a empezar por la tabla “*cotizante*”, comparando cada uno de sus campos con los del diseño propuesto, luego para cada campo plantearemos una posible acción y/u operación a efectuar para lograr llegar al diseño.

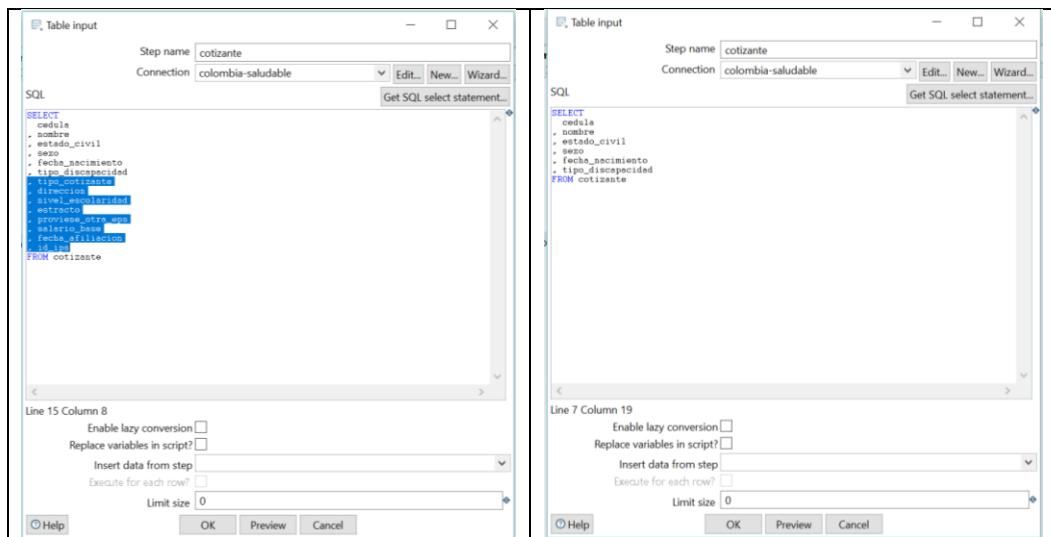
Tabla Cotizante	Dimensión Persona	Operación
-----	key_persona	Se crea al generar la dimensión
-----	tipo_documento	Crear atributo “ <i>tipo_documento</i> ” con valor constante <b>Cedula</b> .
cedula	numero_identificacion	Renombrar atributo “ <i>cedula</i> ” por “ <i>numero_identificacion</i> ”.
nombre	nombre	Permanece igual
-----	tipo_usuario	Crear atributo “ <i>tipo_usuario</i> ” con valor constante <b>cotizante</b>
estado_civil	estado_civil	Permanece igual
sexo	sexo	Permanece igual
fecha_nacimiento	fecha_nacimiento	Permanece igual
tipo_discapacidad	tipo_discapacidad	Permanece igual
-----	grupo_familiar	Crear atributo “ <i>grupo_familiar</i> ” con el valor del campo “ <i>cedula</i> ” de la tabla “ <i>cotizante</i> ”
tipo_cotizante	-----	No incluir campo
dirección	-----	No incluir campo
nivel_escolaridad	-----	No incluir campo
extracto	-----	No incluir campo
proviene_otra_eps	-----	No incluir campo
salario_base	-----	No incluir campo
fecha_afiliacion	-----	No incluir campo
tipo_ips	-----	No incluir campo

De acuerdo a la comparación de la **tabla**, podemos encontrar tres operaciones diferentes:

1. Renombrar campo “cedula” a “numero\_identificacion”
2. Crear los siguientes atributos:
  - a. tipo\_documento: con valor constante “Cedula”
  - b. tipo\_usuario: con valor constante “cotizante”
  - c. grupo\_familiar: con el valor del atributo “cedula” de la tabla “cotizante”
3. Eliminar los atributos:
  - a. tipo\_cotizante
  - b. dirección
  - c. nivel\_escolaridad
  - d. extracto
  - e. proviene\_otra\_eps

- f. salario\_base
- g. fecha\_afiliacion
- h. tipo\_ips

Ya tenemos una idea clara de los procesos que debemos construir para transformar la tabla cotizante, ahora vamos a crearlos en *pentaho*. Primero vamos a eliminar los atributos que no necesitamos para ello, desde la edición de la herramienta “*cotizante*” vamos a quitar los campos no requeridos, como se muestra en la tabla.



Para crear los nuevos atributos, como dos de ellos son valores constantes definidos por nosotros, utilizaremos la herramienta “*Add constant*” y la configuraremos como se muestra en la ilustración.



#	Name	Type	Format	Length	Precision	Currency	Decimal	Group	Value	Set empty string?
1	tipo_usuario	String							cotizante	N
2	tipo_documento	String							Cedula	N

Para el atributo “grupo\_familiar”, como su valor proviene del campo “cedula”, utilizaremos la herramienta “calculator” y lo configuraremos como se muestra en la ilustración.



Calculator												
Step name: grupo_familiar												
Fields:												
#	New field	Calculation	Field A	Field B	Field C	Value type	Length	Precision	Remove	Conversion mask	Decimal symbol	Grouping symbol
1	grupo_familiar	Create a copy of field A	cedula			String	15		N			

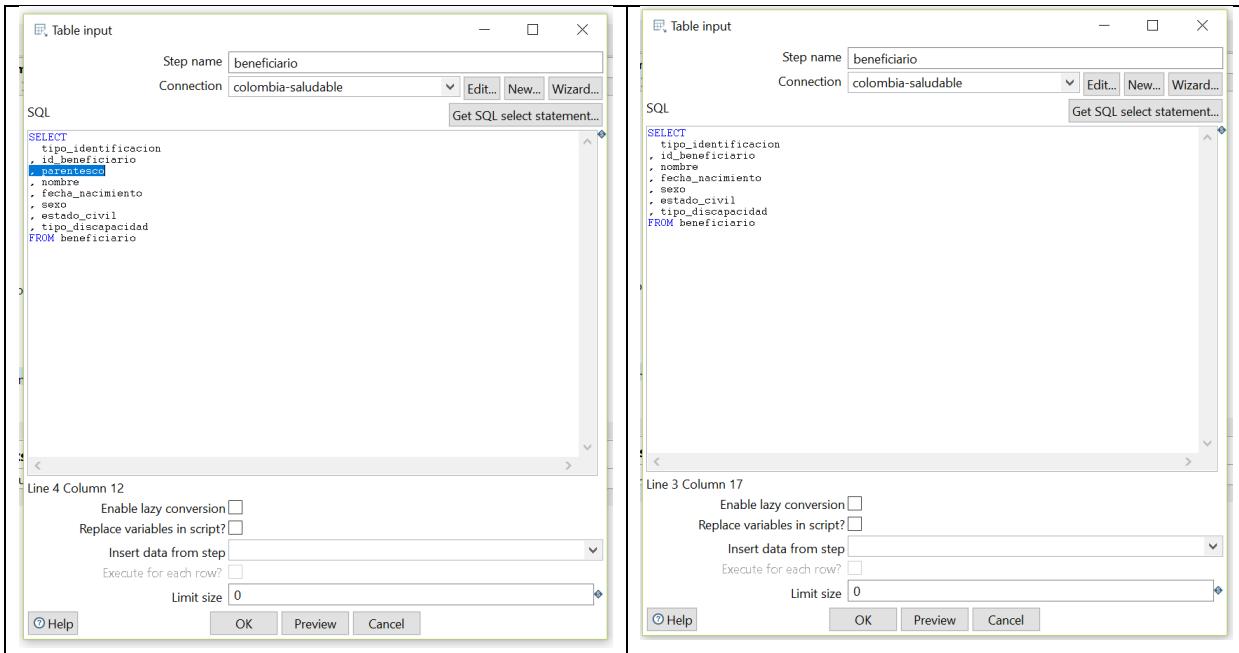
Hasta ahora nuestra transformación se ve como se muestra en la ilustración



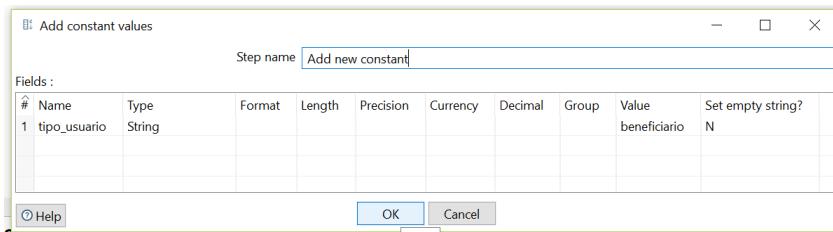
Ahora hacemos aplicamos el mismo proceso para la tabla "*beneficiario*".

Tabla Beneficiario	Dimensión Persona	Operación
-----	key_persona	Se crea al generar la dimensión
tipo_identificacion	tipo_documento	Renombrar atributo " <i>tipo_identificación</i> " por " <i>tipo_documento</i> ".
id_beneficiario	numero_identificacion	Renombrar atributo " <i>id_beneficiario</i> " por " <i>numero_identificacion</i> ".
nombre	nombre	Permanece igual
-----	tipo_usuario	Crear atributo " <i>tipo_usuario</i> " con valor constante <b>beneficiario</b>
estado_civil	estado_civil	Permanece igual
sexo	sexo	Permanece igual
fecha_nacimiento	fecha_nacimiento	Permanece igual
tipo_discapacidad	tipo_discapacidad	Permanece igual
-----	grupo_familiar	Crear atributo " <i>grupo_familiar</i> " mediante un join con la tabla " <i>beneficiario_cotizante</i> " y obtener el atributo " <i>cedula</i> " de la tabla " <i>cotizante</i> "
parentesco	-----	No incluir campo

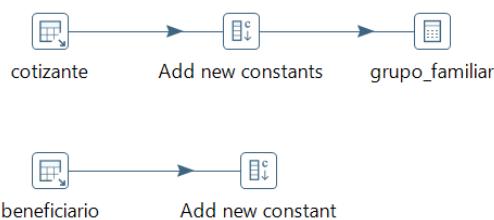
Primero vamos a eliminar los atributos que no necesitamos.



Agregamos el atributo con valor constante



Hasta ahora nuestra transformación se debe de ver como se muestra en la ilustración



Recordemos que para el atributo "*grupo\_familiar*" necesitamos traer la relación "*beneficiario\_cotizante*" de la base de datos fuente, luego debemos unir su información con la información de la tabla beneficiario, y así, obtener la cedula del cotizante.

## Es tu turno...

- ✓ Trae la información de la tabla "cotizante-beneficiario" a la transformación.

Al finalizar se debe de ver y configurar como se muestra en la **tabla 3**.

Tabla 3. Vista y configuración proceso "beneficiario-cotizante"

Vista desde ventana de diseño de Pentaho	Configuración Proceso
<p>cotizante      Add new constants</p> <p>beneficiario    Add new constant</p> <p>beneficiario-cotizante</p>	<p>Table input</p> <p>Step name: beneficiario-cotizante</p> <p>Connection: colombia-saludable</p> <p>SQL:</p> <pre>SELECT     cotizante,     beneficiario FROM cotizante_beneficiario</pre> <p>Line 1 Column 0</p> <p>Enable lazy conversion <input type="checkbox"/></p> <p>Replace variables in script? <input type="checkbox"/></p> <p>Insert data from step <input type="button"/></p> <p>Execute for each row? <input type="checkbox"/></p> <p>Limit size: 0</p> <p>OK Preview Cancel</p>

Ahora vamos a unir la información de las tablas "*beneficiario*" y "*cotizante\_beneficiario*", de la siguiente forma. Desde el panel de diseño en la opción "Transform" vamos a arrastrar la herramienta "*Sort rows*" dos veces y las vamos a unir, como se muestra en la **ilustración 51**.

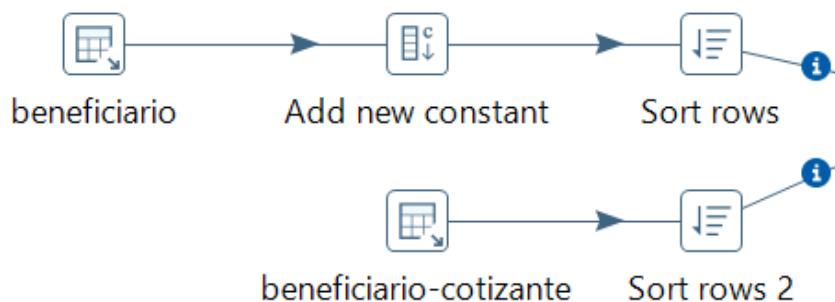


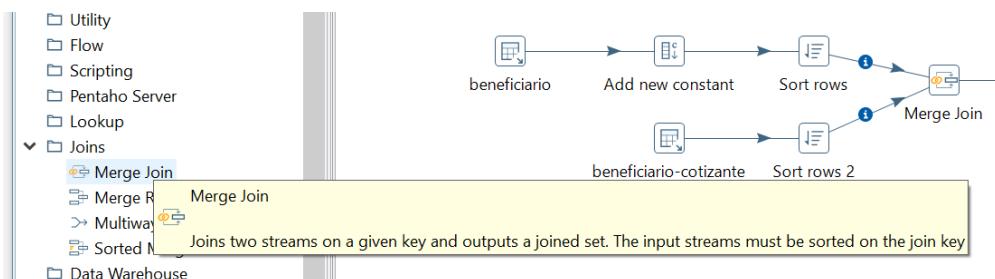
Ilustración 51. Arrastrar herramienta "Sort rows"

En *Pentaho* antes de hacer un *join* (unir), debemos organizar la información por el atributo de la tabla a la cual vamos a hacer la unión de las mismas, en nuestro caso el campo "*id\_beneficiario*" de la relación beneficiario y el campo "*beneficiario*" de la relación "*beneficiario-cotizante*", como se muestra en la **tabla 4.**

**Tabla 4. Organización de campos a unir**

Organizar tabla beneficiario	Organizar tabla beneficiario-cotizante

Ahora ya podemos realizar el join, para ellos vamos desde el panel de diseño en la opción "*Joins*" arrastramos la herramienta "**Merge Join**", como se muestra en la **ilustración 52.**



**Ilustración 52. Arrastrar herramienta "Merge Join"**

A continuación, configuraremos como se muestra en la **ilustración 53.**

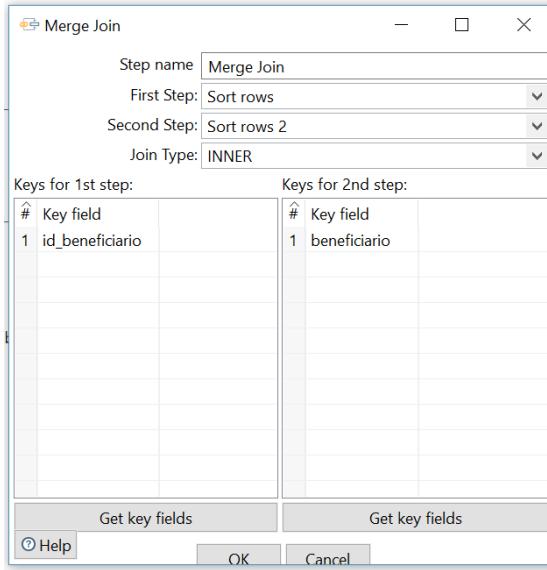


Ilustración 53. Configuración proceso "Merge join"

Podemos hacer una pre-visualización de la unión de ambas tablas, clic derecho, opción "*Preview*", como se muestra en la **ilustración 54**.

#	tipo_identificacion	id_beneficiario	nombre	fecha_nacimiento	sexo	estado_civil	tipo_discapacidad	tipo_usuario	cotizante	beneficiario
1	Tarjeta Identidad	1135438122460	Marcel Lozano Osorio	1991/05/18 00:00:00,000	F	Soltero	Ninguna	beneficiario	899113543	1135438122460
2	Cedula	1135438123690	Adrián Lozano Osorio	1980/11/21 00:00:00,000	M	Viudo	Ninguna	beneficiario	899113543	1135438123690
3	Cedula	1135438126210	Terencio Madrid Vélez	1932/11/19 00:00:00,000	M	Viudo	Ninguna	beneficiario	900113543	1135438126210
4	Cedula	1135438127420	Santos Madrid Ojeda	1990/05/07 00:00:00,000	M	Divorciado	Motora	beneficiario	900113543	1135438127420
5	Cedula	1135438129880	Elías Moreno Jurado	1960/09/19 00:00:00,000	M	Soltero	Ninguna	beneficiario	901113543	1135438129880
6	Cedula	1135438131020	Tomás Moreno Carreño	1967/12/21 00:00:00,000	M	Soltero	Ninguna	beneficiario	901113543	1135438131020
7	Tarjeta Identidad	1135438132210	Jorgelina Moreno Vargas	1994/01/25 00:00:00,000	F	Soltero	Ninguna	beneficiario	901113543	1135438132210
8	Tarjeta Identidad	1135438133320	Emanuel Moreno Jaramillo	1994/09/03 00:00:00,000	M	Soltero	Ninguna	beneficiario	901113543	1135438133320
9	Cedula	1135438134430	Edna Carreño Pereira	1950/07/01 00:00:00,000	F	Soltero	Ninguna	beneficiario	901113543	1135438134430
1	Tarjeta Identidad	1135438137560	Celso Morales Salazar	1992/09/11 00:00:00,000	M	Soltero	Ninguna	beneficiario	902113543	1135438137560
1	Cedula	1135438138800	Albertina Morales Fernández	1950/03/04 00:00:00,000	F	Union Libre	Ninguna	beneficiario	902113543	1135438138800
1	Cedula	1135438139930	Pancracio Morales Fernández	1937/10/03 00:00:00,000	M	Divorciado	Ninguna	beneficiario	902113543	1135438139930
1	Cedula	1135438141100	Juana Morales Correa	1980/12/24 00:00:00,000	F	Viudo	Ninguna	beneficiario	902113543	1135438141100
1	Cedula	1135438143540	Belinda Arévalo Verdugo	1990/10/16 00:00:00,000	F	Viudo	Motora	beneficiario	903113543	1135438143540
1	Tarjeta Identidad	1135438147520	Dámaso Barrios Yáñez	1998/02/10 00:00:00,000	M	Soltero	Mental	beneficiario	905113543	1135438147520
1	Cedula	1135438150070	Telémaco Pardo Hernandez	1984/01/08 00:00:00,000	M	Divorciado	Ninguna	beneficiario	906113543	1135438150070
1	Tarjeta Identidad	1135438153880	Manolo Portillo Villarreal	1995/02/08 00:00:00,000	M	Soltero	Ninguna	beneficiario	908113543	1135438153880
1	Cedula	1135438155030	Rigoberto Portillo Navarrete	1946/06/16 00:00:00,000	M	Soltero	Motora	beneficiario	908113543	1135438155030
1	Cedula	1135438157450	Egeria Rincón Correa	1975/02/22 00:00:00,000	F	Viudo	Ninguna	beneficiario	909113543	1135438157450
2	Cedula	1135438159850	Refugio Flores Rivas	1988/01/06 00:00:00,000	M	Viudo	Ninguna	beneficiario	910113543	1135438159850

Ilustración 54. Pre visualización proceso "Merge join"

Ahora vamos a seleccionar los campos que necesitamos para crear la dimensión. Desde el panel de diseño en la opción "*Transform*", arrastramos dos herramientas "*Select values*", como se muestra en la **ilustración 55**.

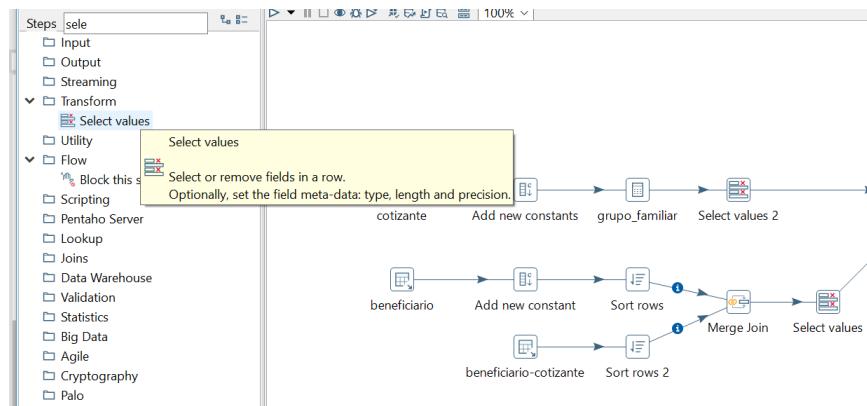


Ilustración 55. Arrastrar herramienta "Select values"

Realizamos la configuración que se muestra en la **tabla 5**.

Tabla 5. Configuración procesos "Select values"

Configuración "select values" proceso "merge join"	Configuración "select values" proceso "grupo familiar"																																																												
<p>hecho_atencion_bacientes Transformation 1 dim_perso</p> <table border="1"> <thead> <tr> <th>#</th> <th>Fieldname</th> <th>Rename to</th> </tr> </thead> <tbody> <tr> <td>1</td> <td>tipo_identificacion</td> <td>tipo_documento</td> </tr> <tr> <td>2</td> <td>idBeneficiario</td> <td>numero_identificacion</td> </tr> <tr> <td>3</td> <td>nombre</td> <td></td> </tr> <tr> <td>4</td> <td>sexo</td> <td></td> </tr> <tr> <td>5</td> <td>estado_civil</td> <td></td> </tr> <tr> <td>6</td> <td>fecha_nacimiento</td> <td></td> </tr> <tr> <td>7</td> <td>tipo_discapacidad</td> <td></td> </tr> <tr> <td>8</td> <td>tipo_usuario</td> <td></td> </tr> <tr> <td>9</td> <td>cotizante</td> <td>grupo_familiar</td> </tr> </tbody> </table>	#	Fieldname	Rename to	1	tipo_identificacion	tipo_documento	2	idBeneficiario	numero_identificacion	3	nombre		4	sexo		5	estado_civil		6	fecha_nacimiento		7	tipo_discapacidad		8	tipo_usuario		9	cotizante	grupo_familiar	<p>hecho_atencion_bacientes Transformation 1 dim_perso</p> <table border="1"> <thead> <tr> <th>#</th> <th>Fieldname</th> <th>Rename to</th> </tr> </thead> <tbody> <tr> <td>1</td> <td>tipo_documento</td> <td></td> </tr> <tr> <td>2</td> <td>cedula</td> <td>numero_identificacion</td> </tr> <tr> <td>3</td> <td>nombre</td> <td></td> </tr> <tr> <td>4</td> <td>sexo</td> <td></td> </tr> <tr> <td>5</td> <td>estado_civil</td> <td></td> </tr> <tr> <td>6</td> <td>fecha_nacimiento</td> <td></td> </tr> <tr> <td>7</td> <td>tipo_discapacidad</td> <td></td> </tr> <tr> <td>8</td> <td>tipo_usuario</td> <td></td> </tr> <tr> <td>9</td> <td>grupo_familiar</td> <td></td> </tr> </tbody> </table>	#	Fieldname	Rename to	1	tipo_documento		2	cedula	numero_identificacion	3	nombre		4	sexo		5	estado_civil		6	fecha_nacimiento		7	tipo_discapacidad		8	tipo_usuario		9	grupo_familiar	
#	Fieldname	Rename to																																																											
1	tipo_identificacion	tipo_documento																																																											
2	idBeneficiario	numero_identificacion																																																											
3	nombre																																																												
4	sexo																																																												
5	estado_civil																																																												
6	fecha_nacimiento																																																												
7	tipo_discapacidad																																																												
8	tipo_usuario																																																												
9	cotizante	grupo_familiar																																																											
#	Fieldname	Rename to																																																											
1	tipo_documento																																																												
2	cedula	numero_identificacion																																																											
3	nombre																																																												
4	sexo																																																												
5	estado_civil																																																												
6	fecha_nacimiento																																																												
7	tipo_discapacidad																																																												
8	tipo_usuario																																																												
9	grupo_familiar																																																												

**Nota:** No olvides que el nombre de los atributos debe coincidir, además del orden de los campos seleccionados.

Por último, solo nos falta crear la dimensión. Desde la opción "*Data Warehouse*", arrastramos la herramienta "*Dimension lookup/update*" y la unimos desde ambos *selects*, como se muestra en la **ilustración 56**.

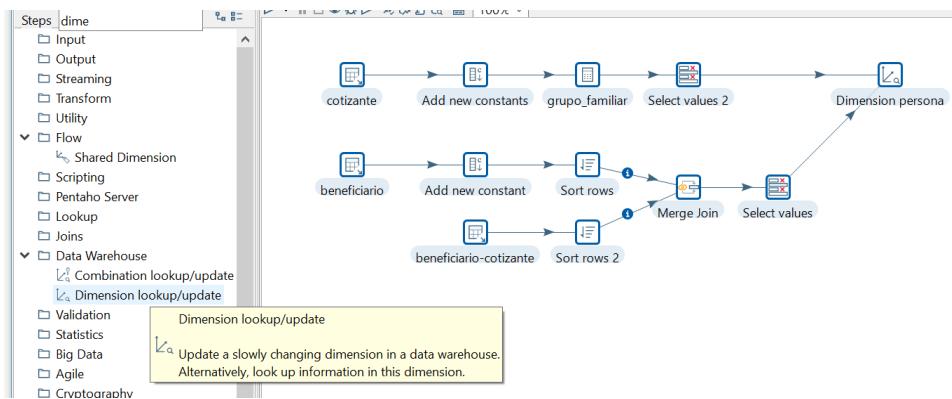
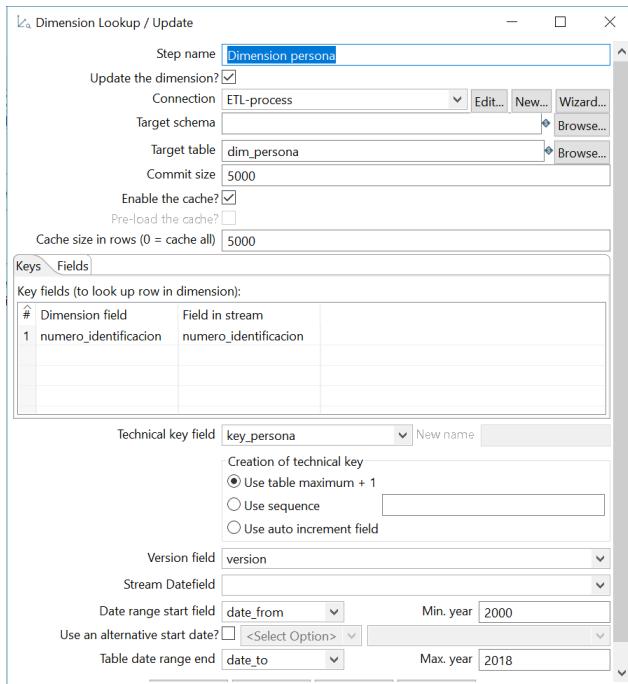
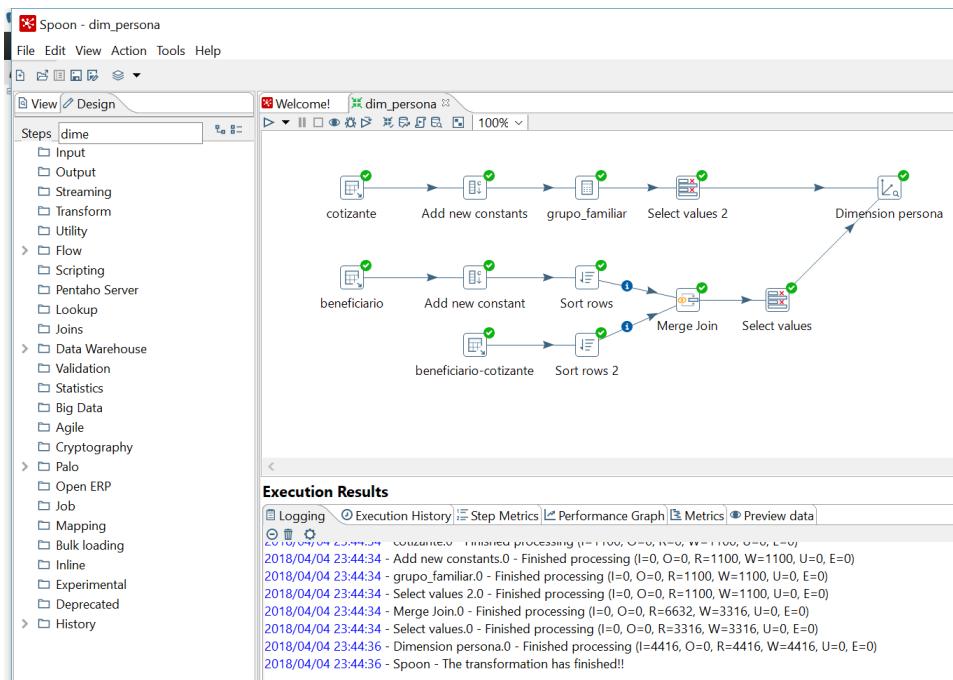


Ilustración 56. Arrastrar herramienta "Dimension"

Realizamos la siguiente configuración, no olvides traer los campos en la pestaña "*Fields*" y ejecutar el SQL para crear el esquema de la dimensión en nuestra bodega de datos.

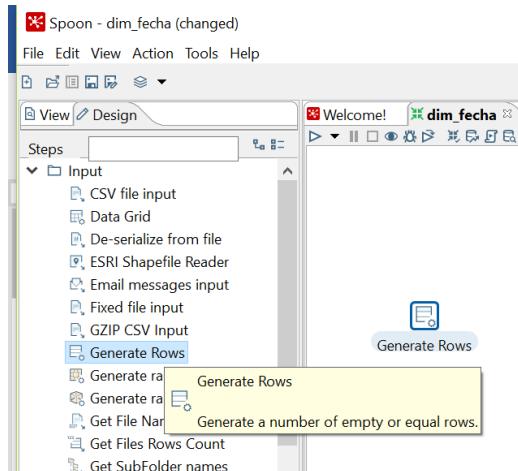


Al finalizar ejecutamos y verificamos en la base de datos la transformación.



### 3.4.5. Implementación Dimensión Fecha

Desde el panel de diseño, en la opción “**Input**” vamos a buscar y arrastrar la herramienta “**Generate Rows**” como de muestra en la **ilustración 57**.



**Ilustración 57. Arrastrar "Generate Rows"**

Ahora vamos a configurarla, damos clic derecho y editar, configuramos los campos como se muestra en la **ilustración 58**.

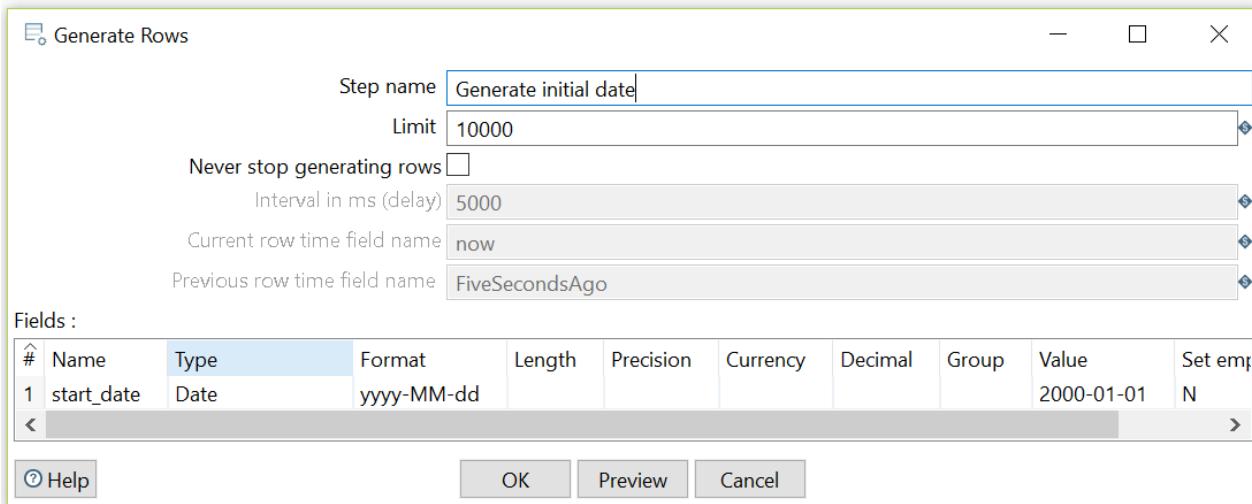


Ilustración 58. Configuración paso "Generate initial date"

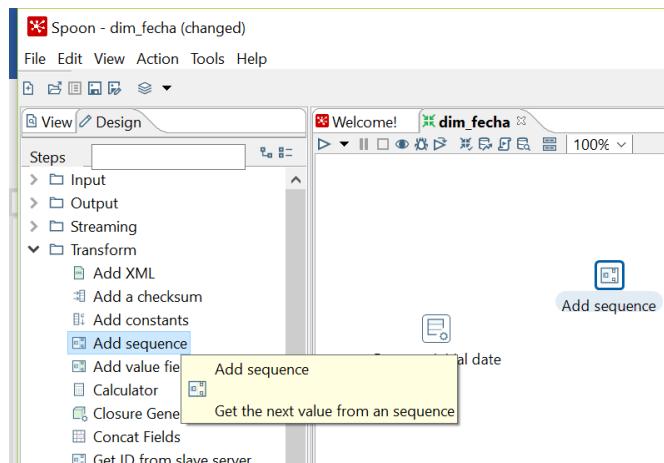
Utilizamos esta herramienta para crear 10000 registros de tipo fecha, con el valor inicial 2000-01-01, podemos visualizarlo con la opción "*Preview*" con una cantidad de 100 registros, como se muestra en la **ilustración 59**.

Rows of step: Generate initial date (100 rows)	
#	start_date
1	2000-01-01
2	2000-01-01
3	2000-01-01
4	2000-01-01
5	2000-01-01
6	2000-01-01
7	2000-01-01
8	2000-01-01
9	2000-01-01
10	2000-01-01
11	2000-01-01
12	2000-01-01
13	2000-01-01
14	2000-01-01
15	2000-01-01
16	2000-01-01
17	2000-01-01
18	2000-01-01
19	2000-01-01
20	2000-01-01

Close Show Log

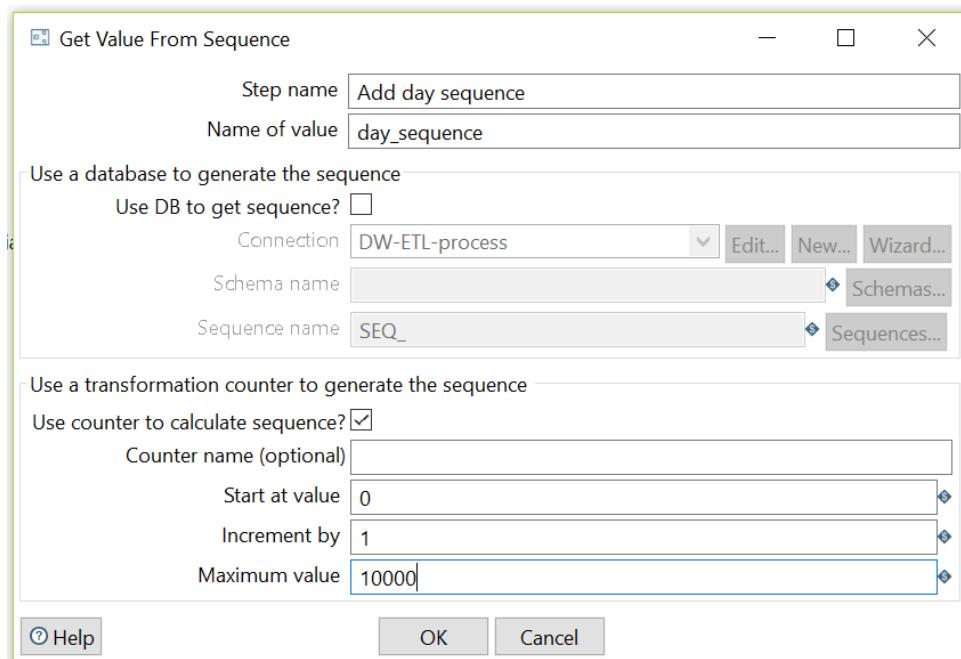
Ilustración 59. Vista previa proceso "Generate initial date"

Desde el panel de diseño buscamos la opción "*Transform*" y arrastramos la herramienta "*Add sequence*" como se muestra en la **ilustración 60**.



**Ilustración 60. Arrastrar "Add sequence"**

En seguida, vamos a configurarla como se muestra en la **ilustración 61.**



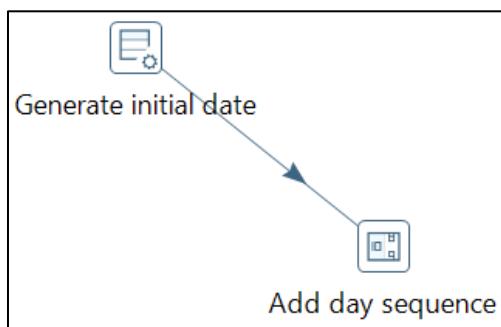
**Ilustración 61. Configuración paso "Add day sequence"**

En este paso estamos creando una secuencia de días desde 0 hasta 10000 que es el total de registros creados en el paso anterior.

A continuación, debemos unirlos, nos posicionamos en el ícono "*Generate initial date*" y esperamos que se visualice como se muestra a continuación.

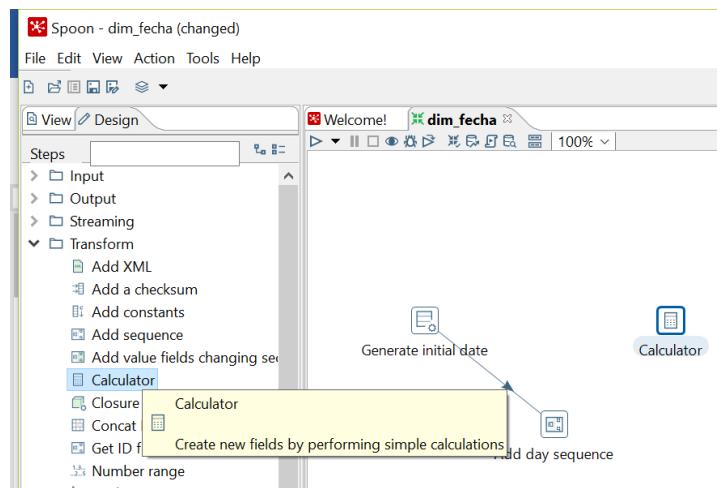


damos clic en  y luego damos clic en el ícono "*Add secuence days*" al finalizar, ambos procesos deben de quedar enlazados como se muestra en la **ilustración 62**.



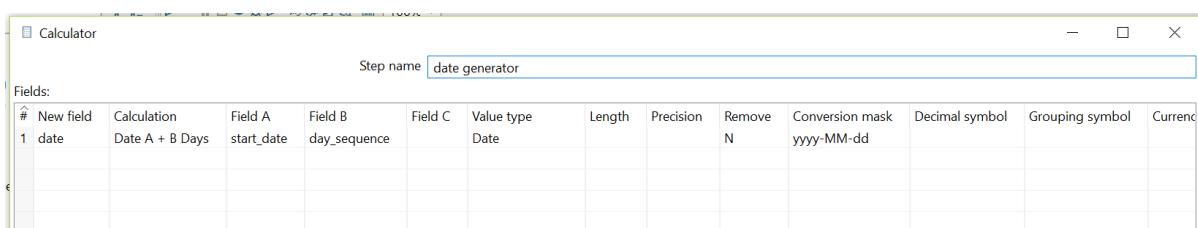
**Ilustración 62. Unión de los procesos**

Desde el panel de diseño en la opción "*Trasnform*" vamos a arrastrar la herramienta "*Calculator*", como se muestra en la **ilustración 63**.



**Ilustración 63. Arrastrar proceso "Calculator"**

Lo unimos desde el paso anterior "*Add day secuence*" y lo configuramos como se muestra en la **ilustración 64**.



**Ilustración 64. Configuración proceso "Date generator"**

En este paso lo que estamos haciendo es sumar los registros de fecha creados en el paso "*Generate initial date*" la secuencia de días generados en el paso "*Add day secuence*" como muestra en la **ilustración 65**.

Examine preview data

Rows of step: date generator (1000 rows)

#	start_date	day_sequence	date
1	2000-01-01	0	2000-01-01
2	2000-01-01	1	2000-01-02
3	2000-01-01	2	2000-01-03
4	2000-01-01	3	2000-01-04
5	2000-01-01	4	2000-01-05
6	2000-01-01	5	2000-01-06
7	2000-01-01	6	2000-01-07
8	2000-01-01	7	2000-01-08
9	2000-01-01	8	2000-01-09
10	2000-01-01	9	2000-01-10
11	2000-01-01	10	2000-01-11
12	2000-01-01	11	2000-01-12
13	2000-01-01	12	2000-01-13
14	2000-01-01	13	2000-01-14
15	2000-01-01	14	2000-01-15
16	2000-01-01	15	2000-01-16
17	2000-01-01	16	2000-01-17
18	2000-01-01	17	2000-01-18
19	2000-01-01	18	2000-01-19
20	2000-01-01	19	2000-01-20

Ilustración 65. Pre-visualización proceso "Date generator"

Hasta este punto ya tenemos el proceso para generar las fechas desde 2000-01-01, pero esto no es suficiente para nuestra dimensión fecha, recordemos que debemos crear los registros definidos en el diseño (**ver ilustración 8**).

Ahora vamos a usar las operaciones que permite la función *Calculator* para generar la información adicional que requerimos. Arrastramos otra herramienta "*Calculator*" la unimos desde "*Date generator*" y la configuramos como se muestra en la **ilustración 66**.

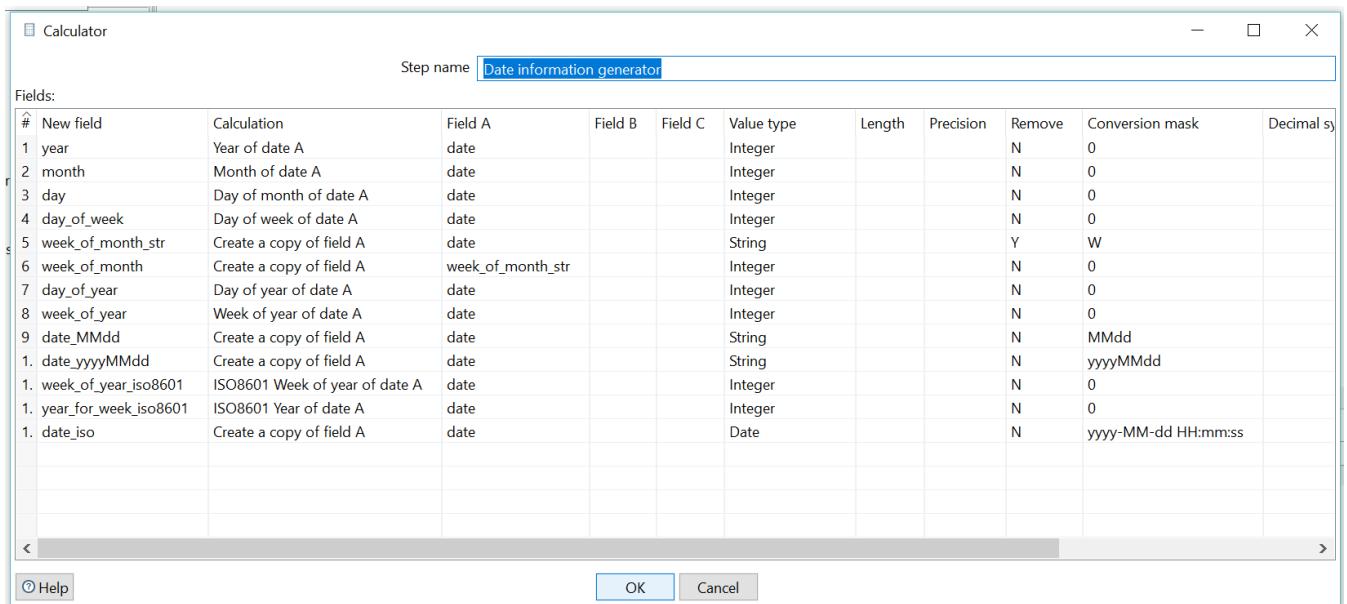


Ilustración 66. Configuración proceso "Date information generator"

Como podemos ver estamos usando las operaciones definidas, para segmentar la información de la fecha. Si queremos ver cómo va quedando la información, damos clic derecho en el icono de “*Date information generator*” y seleccionamos la opción “*Preview*”, elegimos ver los primeros 100 registros damos clic en “*Quick lauch*” como se muestra en la **ilustración 67**.

#	start_date	day_sequence	date	year	month	day	day_of_week	week_of_month	day_of_year	week_of_year	date_MMdd	date_YYYYMMdd	week_of_year_iso8601	year_for_week_iso8601	date_iso
1	2000-01-01	0	2000-01-01	2000	1	1	7	1	1	1	0101	20000101	52	1999	2000-01-01 00:00:00
2	2000-01-01	1	2000-01-02	2000	1	2	1	1	2	1	0102	20000102	52	1999	2000-01-02 00:00:00
3	2000-01-01	2	2000-01-03	2000	1	3	2	2	3	2	0103	20000103	1	2000	2000-01-03 00:00:00
4	2000-01-01	3	2000-01-04	2000	1	4	3	2	4	2	0104	20000104	1	2000	2000-01-04 00:00:00
5	2000-01-01	4	2000-01-05	2000	1	5	4	2	5	2	0105	20000105	1	2000	2000-01-05 00:00:00
6	2000-01-01	5	2000-01-06	2000	1	6	5	2	6	2	0106	20000106	1	2000	2000-01-06 00:00:00
7	2000-01-01	6	2000-01-07	2000	1	7	6	2	7	2	0107	20000107	1	2000	2000-01-07 00:00:00
8	2000-01-01	7	2000-01-08	2000	1	8	7	2	8	2	0108	20000108	1	2000	2000-01-08 00:00:00
9	2000-01-01	8	2000-01-09	2000	1	9	1	2	9	2	0109	20000109	1	2000	2000-01-09 00:00:00
10	2000-01-01	9	2000-01-10	2000	1	10	2	3	10	3	0110	20000110	2	2000	2000-01-10 00:00:00
11	2000-01-01	10	2000-01-11	2000	1	11	3	3	11	3	0111	20000111	2	2000	2000-01-11 00:00:00
12	2000-01-01	11	2000-01-12	2000	1	12	4	3	12	3	0112	20000112	2	2000	2000-01-12 00:00:00
13	2000-01-01	12	2000-01-13	2000	1	13	5	3	13	3	0113	20000113	2	2000	2000-01-13 00:00:00
14	2000-01-01	13	2000-01-14	2000	1	14	6	3	14	3	0114	20000114	2	2000	2000-01-14 00:00:00
15	2000-01-01	14	2000-01-15	2000	1	15	7	3	15	3	0115	20000115	2	2000	2000-01-15 00:00:00
16	2000-01-01	15	2000-01-16	2000	1	16	1	3	16	3	0116	20000116	2	2000	2000-01-16 00:00:00
17	2000-01-01	16	2000-01-17	2000	1	17	2	4	17	4	0117	20000117	3	2000	2000-01-17 00:00:00
18	2000-01-01	17	2000-01-18	2000	1	18	3	4	18	4	0118	20000118	3	2000	2000-01-18 00:00:00
19	2000-01-01	18	2000-01-19	2000	1	19	4	4	19	4	0119	20000119	3	2000	2000-01-19 00:00:00
20	2000-01-01	19	2000-01-20	2000	1	20	5	5	20	5	0120	20000120	2	2000	2000-01-20 00:00:00

Ilustración 67. Pre-visualización proceso “Date information generator”

**Nota 1:** Para salir del modo “*Preview*” debemos dar clic en el botón “*Stop*”.

**Nota 2:** No olvides ir guardando los cambios realizados.

Podemos ver cómo va tomando forma nuestra dimensión fecha, ahora, que podemos hacer para calcular la información restante, como días festivos, época del año, nombre del día de la semana, nombre del mes, entre otros. Para ello vamos a utilizar el siguiente [script](#) ([descargar](#))

Desde la opción “*Scripting*” arrastramos la herramienta “*Modified Java Script Value*” como se muestra en la **ilustración 68**.

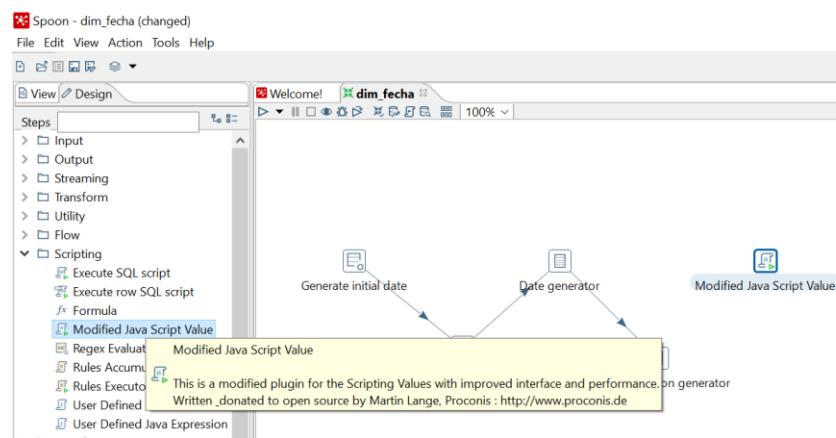
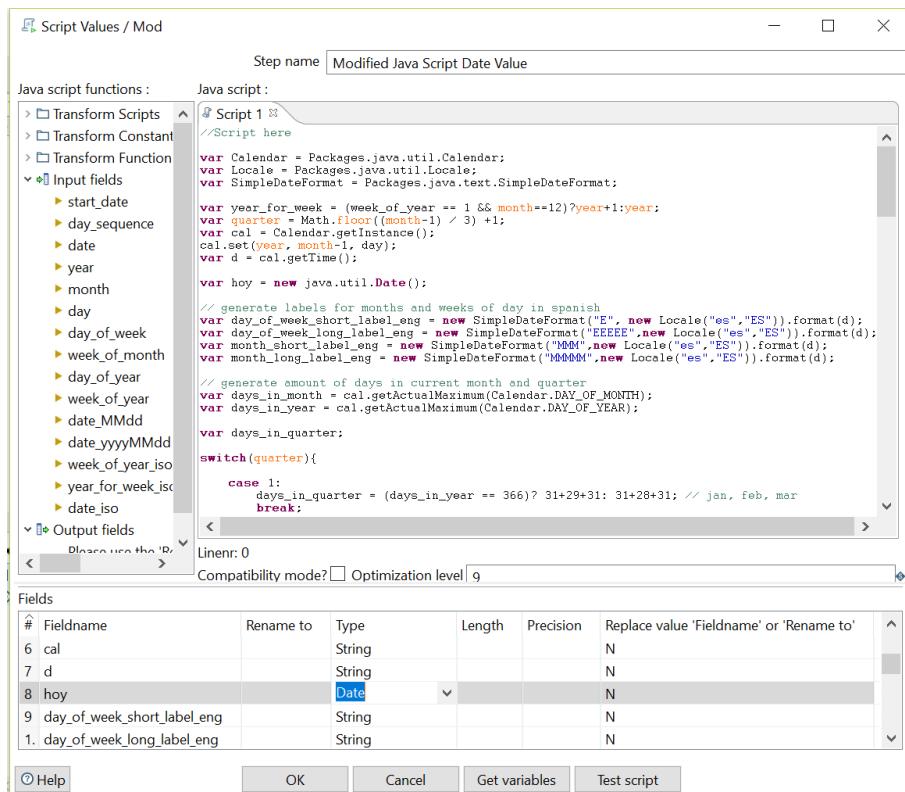


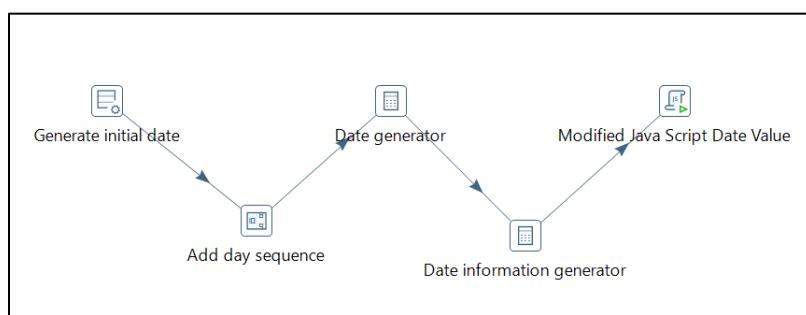
Ilustración 68. Arrastrar proceso “Modified Java Script Value”

Lo unimos desde el paso anterior "*Date information generator*", copiamos la información del script descargado y lo pegamos en el cuadro script, luego damos clic en el botón "*Get variables*", las cuales aparecerán en el panel inferior, buscamos la variable "*hoy*" y le cambiamos el tipo de "*String*" a "*Date*" como se muestra en la **ilustración 69**.



**Ilustración 69. Configuración proceso "Modified Java Script Date Value"**

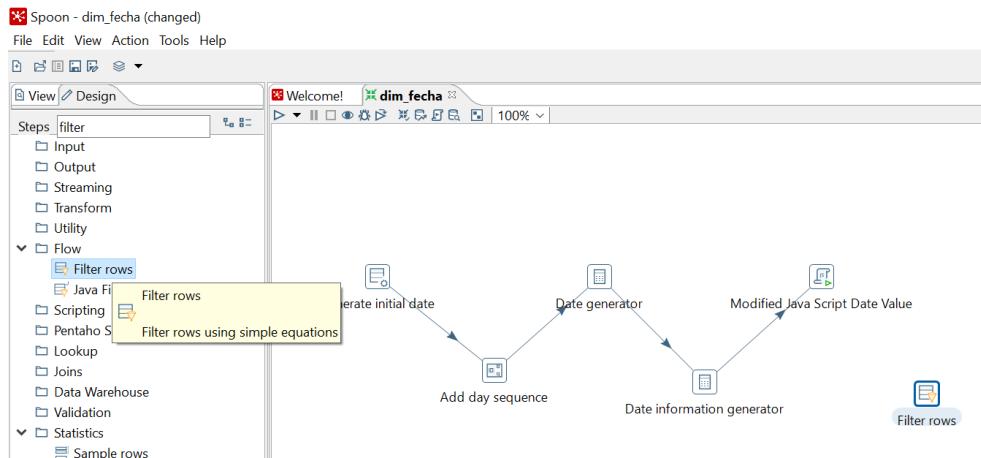
Hasta ahora nuestra transformación se debe de ver como se muestra en la **ilustración 70**.



**Ilustración 70. Transformación dimensión fecha.**

Los pasos siguientes nos van a ayudar filtrar y seleccionar la información que vamos a incluir en nuestra dimensión.

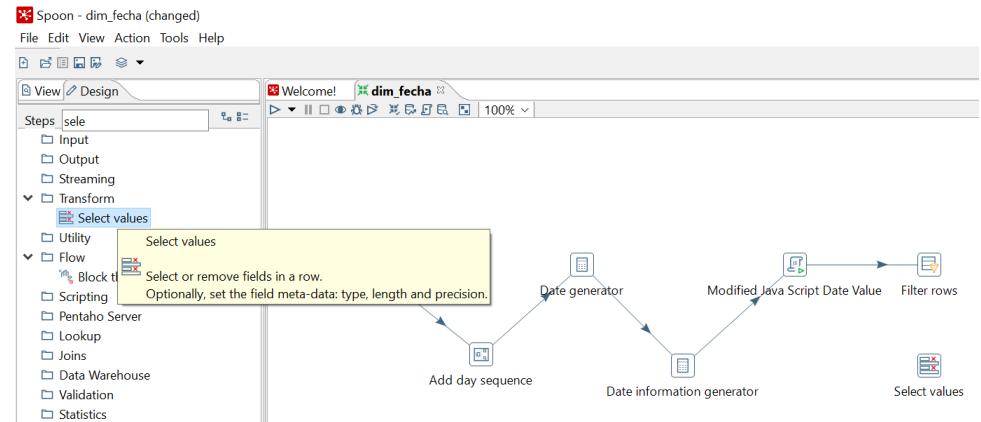
Arrastramos desde la opción “*Flow*” del panel de diseño la herramienta “*Filter rows*” como se muestra en la **ilustración 71**.



**Ilustración 71. Arrastrar proceso “Filter rows”**

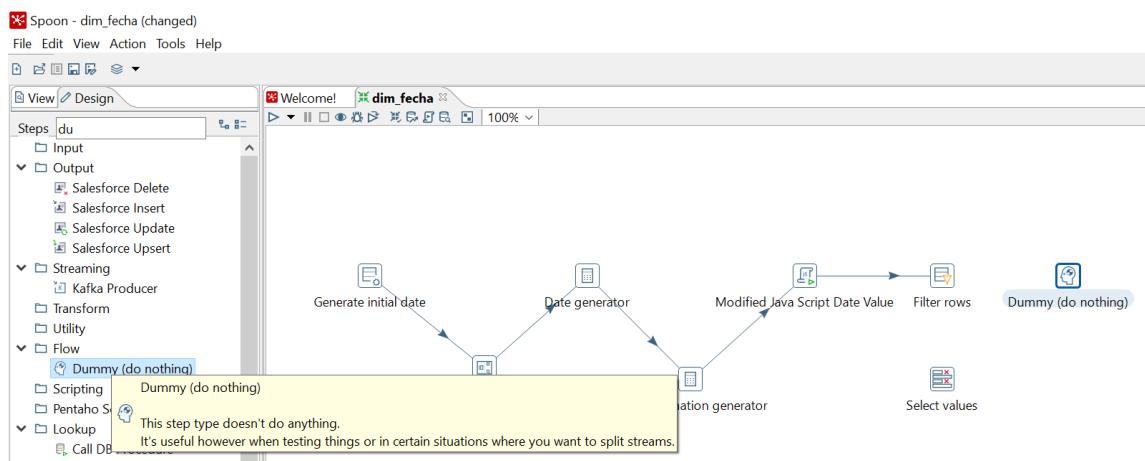
La unimos desde el paso anterior “*Modified Java Script Date Value*”, pero antes de configurarlo, primero debemos arrastrar dos elementos más.

Desde la opción “*Transform*” arrastramos la herramienta “*Select values*”, como se muestra en la **ilustración 72**.



**Ilustración 72. Arrastrar el proceso “Select values”**

Por otra parte, desde la opción “*Flow*” arrastramos la herramienta “*Dummy (do nothing)*” como se muestra en la **ilustración 73**.



Ahora debemos unirlas desde el paso "*Filter row*", con el paso "*Dummy*" para el resultado falso y "*Select value*" el resultado verdadero, como se muestra en la **ilustración 74**.

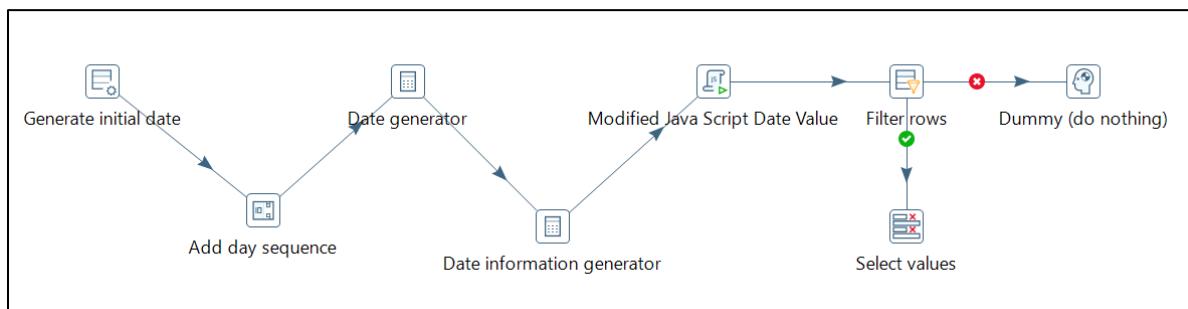


Ilustración 74. Vista previa transformación dimensión fecha

Ahora vamos a configurar el paso "*Filter rows*", como se muestra en la **ilustración 75**.

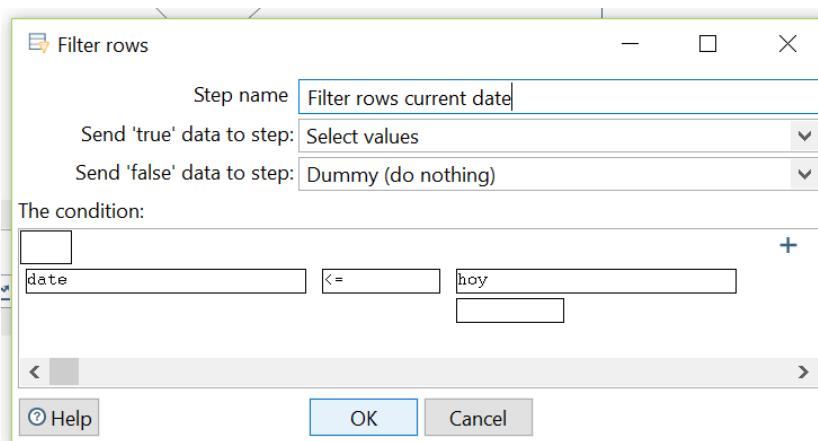
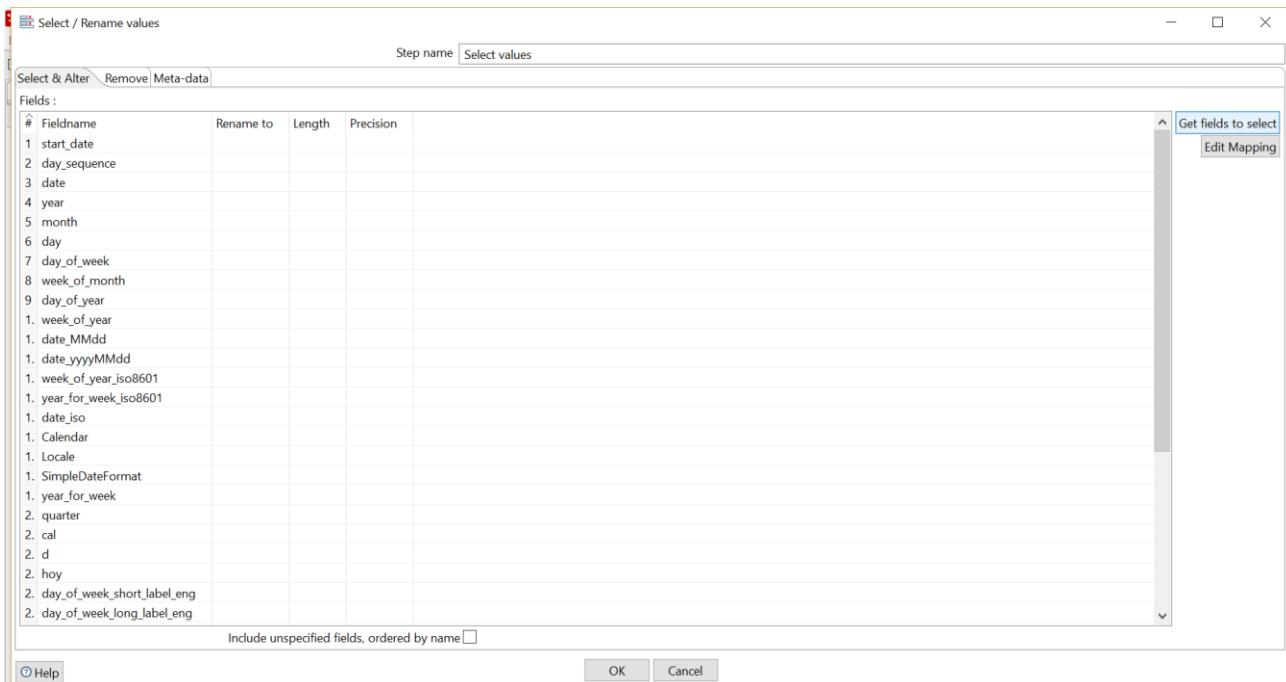


Ilustración 75. Configuración proceso "Filter rows current date"

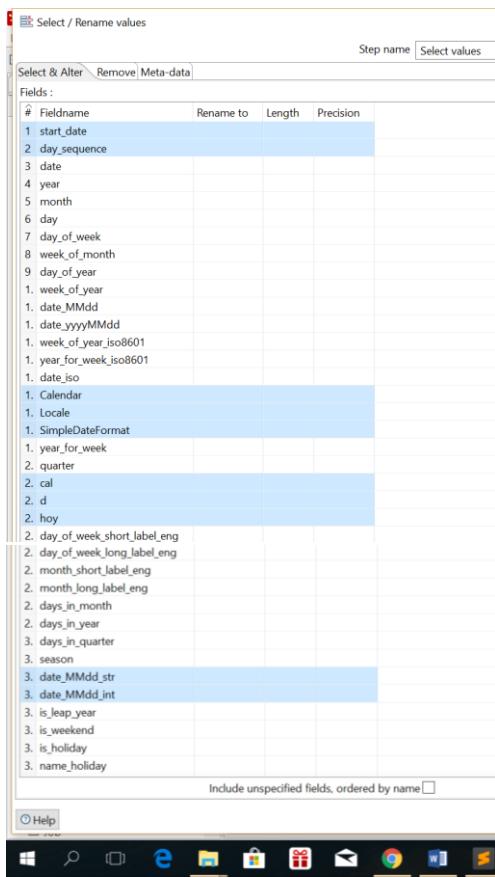
Lo que estamos indicando en este paso es que si los registros de la columna fecha, son menores o iguales a la fecha actual, los vamos a seleccionar, en caso contrario no se hace nada. Dicho en otras palabras, solo vamos a seleccionar las fechas generadas hasta la fecha actual.

Seguimos con la configuración del proceso "*Select values*" para seleccionar las columnas que vamos a incluir en nuestra dimensión, damos clic derecho seleccionamos "*editar*", luego damos clic en el botón "*Get fields to select*" como se muestra en la **ilustración 76**.



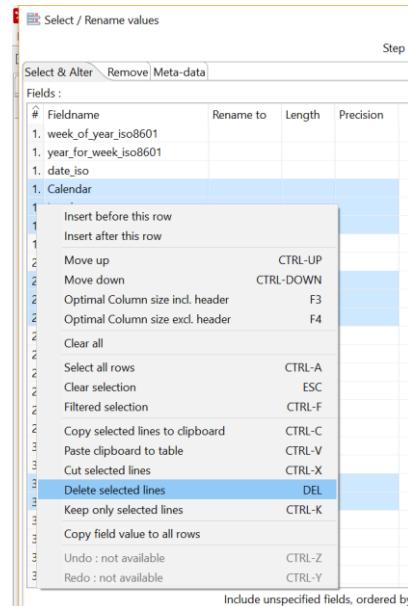
**Ilustración 76. Posibles campos a incluir en la dimensión**

Como podemos ver, no solo aparecen los campos que hemos definido en los pasos anteriores, sino que también las variables definidas en el proceso del script, vamos seleccionar y eliminar aquellas que solo usamos para ejecutar un cálculo específico y no representan información valiosa para nuestra dimensión, como se muestra en la **ilustración 77**.



**Ilustración 77. Selección de campos a eliminar**

Luego, sobre una de las filas seleccionadas damos clic derecho y escogemos la opción "*Delete selected lines*", como se muestra en la **ilustración 78**.



**Ilustración 78. Eliminar campos seleccionados**

Para finalizar damos clic en “**OK**”. Vamos a hacer un “**Preview**” para validar como va quedando la información desde el paso “**Select values**”, como se muestra en la **ilustración 79**.

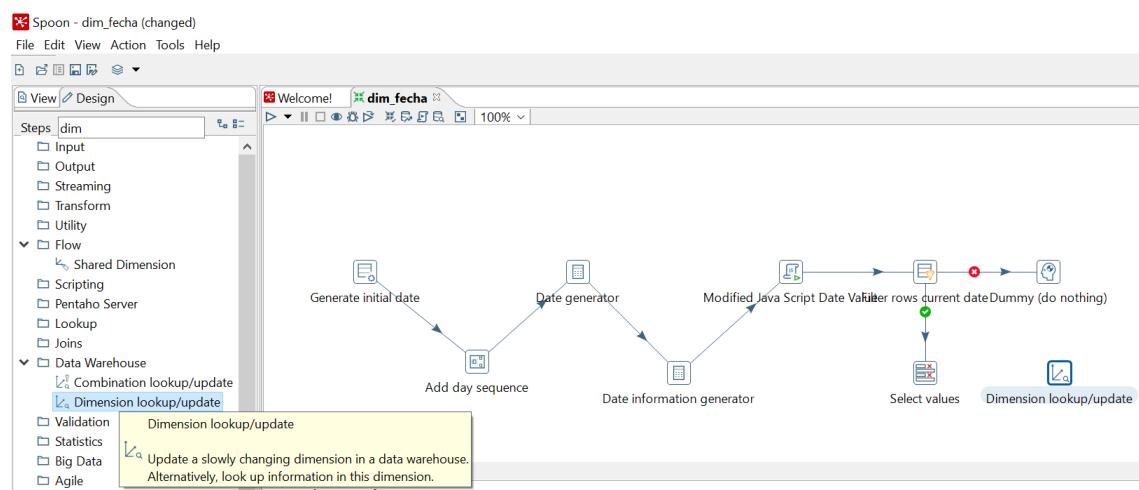
The screenshot shows a JBoss ETL (Spoon) interface. At the top, there's a menu bar with File, Edit, View, Action, Tools, Help, and a Connect button. Below the menu is a toolbar with various icons. A preview window titled "Examine preview data" displays a table with 1000 rows of date-related data. The columns include date, year, month, day, day\_of\_week, week\_of\_month, day\_of\_year, week\_of\_year, date\_MMdd, date\_yyyyMMdd, week\_of\_year\_iso8601, year\_for\_week\_iso8601, date\_iso, and year\_fc. The data spans from January 1, 2000, to January 20, 2000. At the bottom of the preview window, there are buttons for Close, Stop, and Get more rows. Below the preview window, the status bar shows "Cryptography" and "Palo". The log area at the bottom right shows "2018/03/29 13:36:41 - Date information generator.0 - Finished processing [Stop the preview transformation] =0, E=0".

**Ilustración 79. Pre-visualización del proceso “Select values”**

**Nota 1:** Para terminar la pre-visualización no olvides dar clic en el botón “**Stop**”

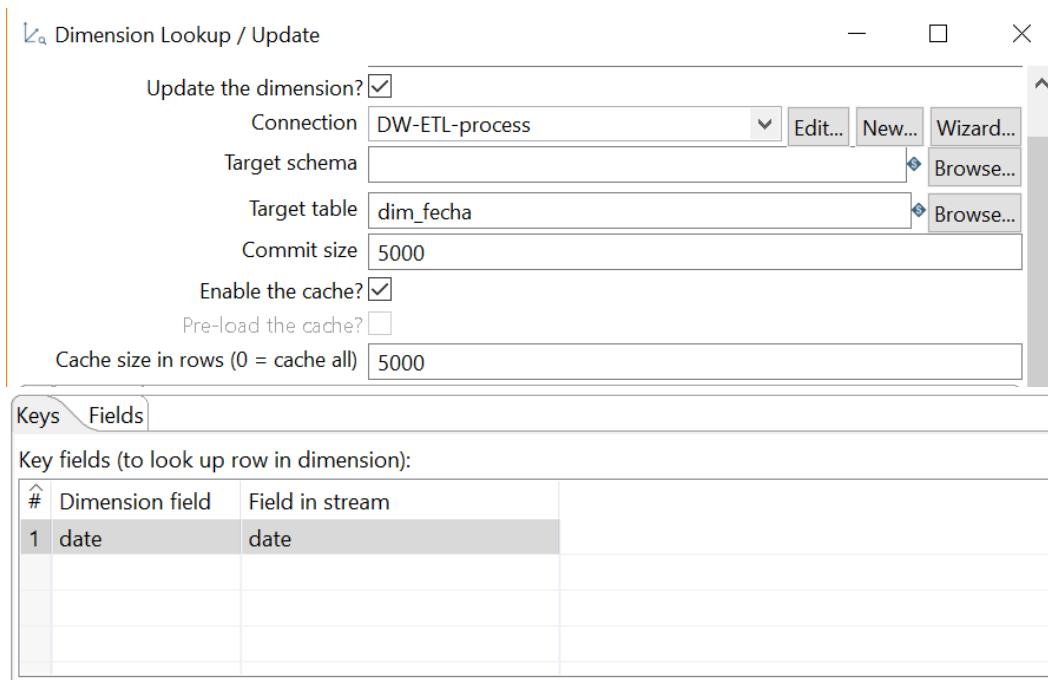
**Nota 2:** No olvides guardar los cambios

Ahora solo nos falta crear la dimensión con los campos seleccionados en nuestra bodega de datos, para ello, desde la opción “**Data Warehouse**” vamos a arrastrar la herramienta “**Dimension lookup/update**”, como se muestra en la **ilustración 80**.



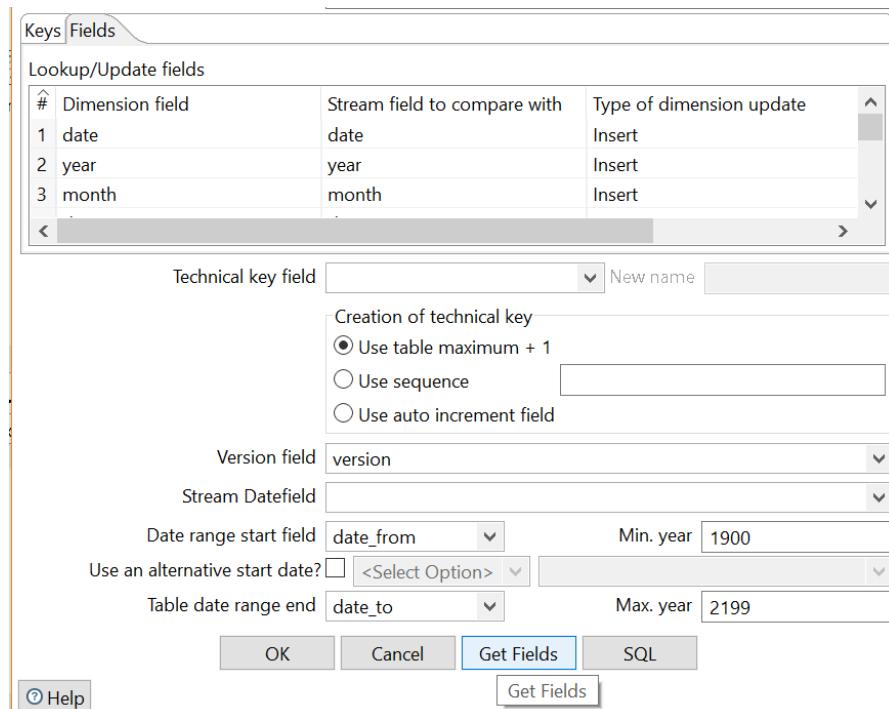
**Ilustración 80. Arrastrar proceso “Dimension lookup/update”**

La enlazamos desde el paso “*Select values*” y la configuramos como se muestra en la **ilustración 81**.



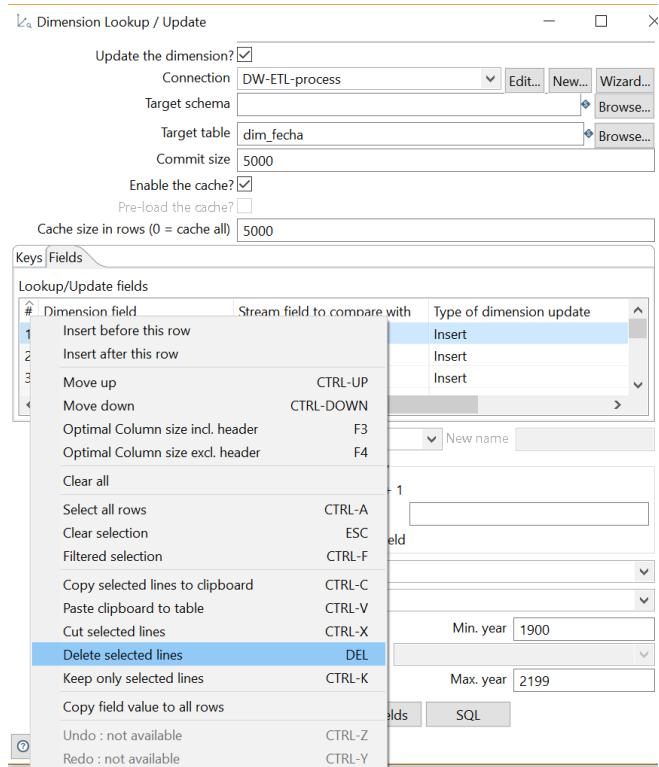
**Ilustración 81.** Configuración primera parte proceso "Dimension"

Desde la pestaña “*Fields*” damos clic en el botón “*Get Fields*”, como se muestra en la **ilustración 82**.



**Ilustración 82.** Configuración segunda parte proceso "Dimension"

Como ya incluimos el campo "*date*" en la pestaña "*Keys*", lo debemos eliminar de la pestaña "*Field*" seleccionándolo, damos clic derecho y escogemos la opción "*Delete selected lines*", como se muestra en la **ilustración 83**.



**Ilustración 83. Eliminación del campo "date"**

Seguimos configurando los campos restantes como se muestra en la **ilustración 84**.

Technical key field	key_date	New name
Creation of technical key		
<input checked="" type="radio"/> Use table maximum + 1 <input type="radio"/> Use sequence <input type="radio"/> Use auto increment field		
Version field	version	
Stream Datefield		
Date range start field	date_from	Min. year 2000
Use an alternative start date?	<input type="checkbox"/> <Select Option>	
Table date range end	date_to	Max. year 2199
<input type="button" value="OK"/> <input type="button" value="Cancel"/> <input type="button" value="Get Fields"/> <input type="button" value="SQL"/>		

**Ilustración 84. Configuración tercera parte del proceso "Dimension"**

Ahora necesitamos crear el esquema de la dimensión fecha, con los campos seleccionados, para ello damos clic en el botón “**SQL**”, lo cual nos permite ver el código sql para crear la tabla en la base de datos, damos clic en el botón “**Execute**”, como se muestra en la **ilustración 85**.

The screenshot shows two windows from Oracle SQL Developer. The left window is titled "Simple SQL editor" and contains the SQL code for creating the "dim\_fecha" table and its indexes. The right window is titled "Results of the SQL statements" and displays the results of the executed SQL, showing three statements: CREATE TABLE, CREATE INDEX idx\_dim\_fecha\_lookup, and CREATE INDEX idx\_dim\_fecha\_tk.

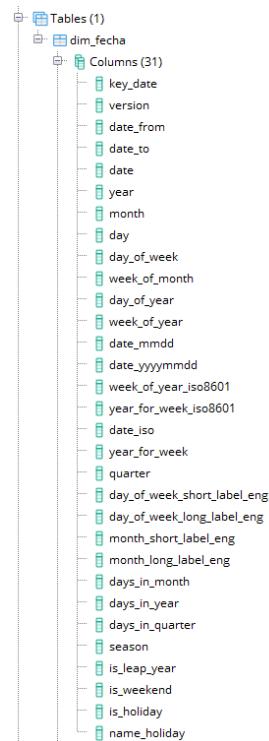
```

CREATE TABLE dim_fecha
(
    key_date BIGSERIAL
    ,version INTEGER
    ,date_from TIMESTAMP
    ,date_to TIMESTAMP
    ,date_ "TIMESTAMP"
    ,year DOUBLE PRECISION
    ,year_ "DOUBLE PRECISION
    ,day_ "DOUBLE PRECISION
    ,day_of_week DOUBLE PRECISION
    ,month_short DOUBLE PRECISION
    ,day_of_year DOUBLE PRECISION
    ,week_of_year DOUBLE PRECISION
    ,date_mmdd TEXT
    ,date_yyyyymmdd TEXT
    ,week_of_year_iso8601 DOUBLE PRECISION
    ,year_for_week_iso8601 DOUBLE PRECISION
    ,date_iso TEXT
    ,quarter NUMBER(10, 2)
    ,year_for_week TEXT
    ,quarter_label_eng TEXT
    ,day_of_week_short_label_eng TEXT
    ,day_of_week_long_label_eng TEXT
    ,month_short_label_eng TEXT
    ,month_long_label_eng TEXT
    ,days_in_month INTEGER
    ,days_in_year INTEGER
    ,days_in_quarter INTEGER
    ,session TEXT
    ,is_leap_year TEXT
    ,is_weekend TEXT
    ,is_holiday TEXT
    ,name_holiday TEXT
)
CREATE INDEX idx_dim_fecha_lookup ON dim_fecha("date")
;
CREATE INDEX idx_dim_fecha_tk ON dim_fecha(key_date)
;

```

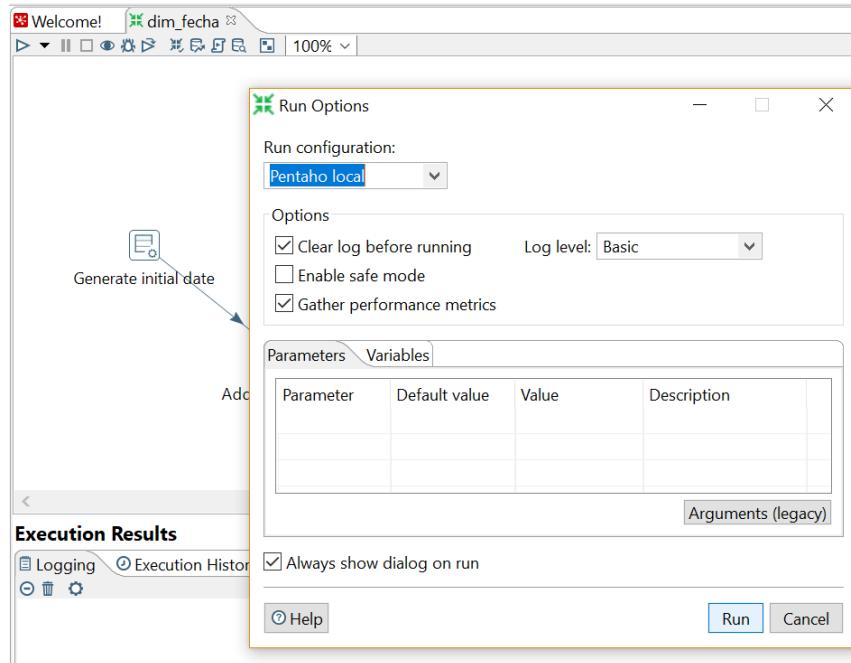
**Ilustración 85. Ejecución del SQL para crear la dimensión fecha**

Podemos verificar en la base de datos la tabla creada como se muestra en la **ilustración 86**.



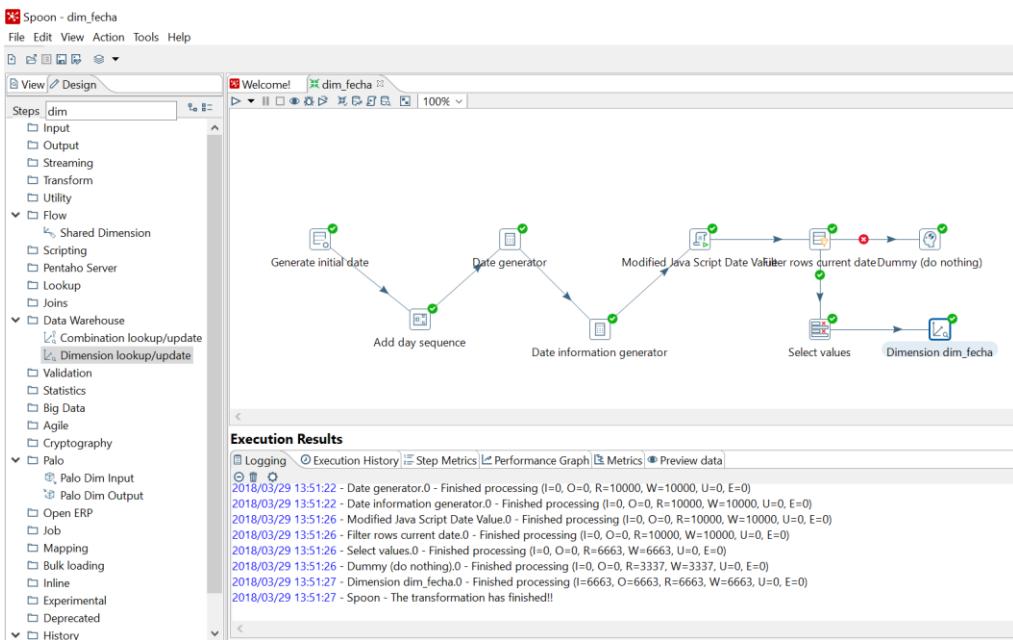
**Ilustración 86. Tabla dim\_fecha creada en la base de datos**

Listo ahora vamos a ejecutar nuestra transformación creada, damos clic en el botón  luego damos clic en el botón **"Run"** como se muestra en la **ilustración 87**.



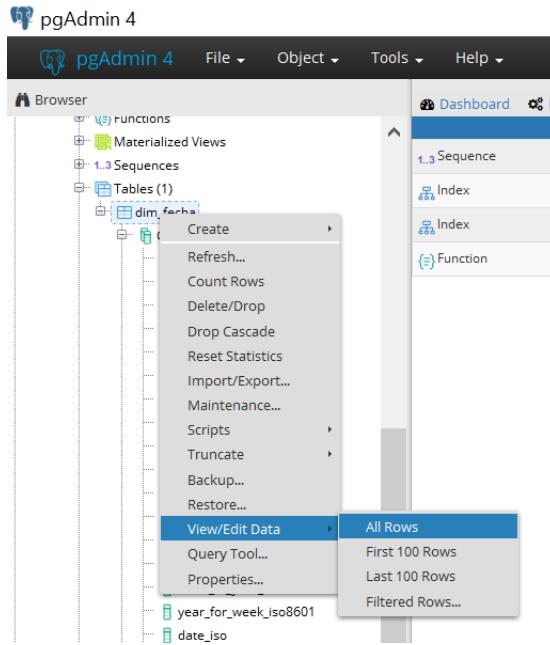
**Ilustración 87. Ejecutar transformación dim\_fecha**

Si no se presenta ningún error, cada uno de los procesos de la transformación al finalizar la ejecución aparecerá con una imagen de chequeo de color verde, además dentro del panel de ejecución aparecerá su resultado de color negro, como se muestra en la **ilustración 88**.



**Ilustración 88. Ejecución exitosa de la transformación**

Podemos verificar en la base de datos que la tabla dim\_fecha se encuentra con la información ingresada como se muestra en la **ilustración 89**.



**Ilustración 89.** Visualizar los datos de la tabla dim\_fecha

Podemos ver que la tabla se encuentra con la información generada a partir de la transformación creada para la dimensión fecha. Se debe visualizar la información como se muestra en la **ilustración 90**.

	key_date	version	date_from	date_to	date	year	month	day	day_of_week	week_of_month
1	0	1	[null]	[null]	[null]	[null]	[null]	[null]	[null]	[null]
2	1	1	2000-01-01 00:00:00	2199-12-31 23:59:59.999	2000-01-01 00:00:00	2000	1	1	7	1
3	2	1	2000-01-01 00:00:00	2199-12-31 23:59:59.999	2000-01-02 00:00:00	2000	1	2	1	
4	3	1	2000-01-01 00:00:00	2199-12-31 23:59:59.999	2000-01-03 00:00:00	2000	1	3	2	
5	4	1	2000-01-01 00:00:00	2199-12-31 23:59:59.999	2000-01-04 00:00:00	2000	1	4	3	
6	5	1	2000-01-01 00:00:00	2199-12-31 23:59:59.999	2000-01-05 00:00:00	2000	1	5	4	
7	6	1	2000-01-01 00:00:00	2199-12-31 23:59:59.999	2000-01-06 00:00:00	2000	1	6	5	
8	7	1	2000-01-01 00:00:00	2199-12-31 23:59:59.999	2000-01-07 00:00:00	2000	1	7	6	
9	8	1	2000-01-01 00:00:00	2199-12-31 23:59:59.999	2000-01-08 00:00:00	2000	1	8	7	
10	9	1	2000-01-01 00:00:00	2199-12-31 23:59:59.999	2000-01-09 00:00:00	2000	1	9	1	
11	10	1	2000-01-01 00:00:00	2199-12-31 23:59:59.999	2000-01-10 00:00:00	2000	1	10	2	
12	11	1	2000-01-01 00:00:00	2199-12-31 23:59:59.999	2000-01-11 00:00:00	2000	1	11	3	
13	12	1	2000-01-01 00:00:00	2199-12-31 23:59:59.999	2000-01-12 00:00:00	2000	1	12	4	
14	13	1	2000-01-01 00:00:00	2199-12-31 23:59:59.999	2000-01-13 00:00:00	2000	1	13	5	
15	14	1	2000-01-01 00:00:00	2199-12-31 23:59:59.999	2000-01-14 00:00:00	2000	1	14	6	
16	15	1	2000-01-01 00:00:00	2199-12-31 23:59:59.999	2000-01-15 00:00:00	2000	1	15	7	
17	16	1	2000-01-01 00:00:00	2199-12-31 23:59:59.999	2000-01-16 00:00:00	2000	1	16	1	
18	17	1	2000-01-01 00:00:00	2199-12-31 23:59:59.999	2000-01-17 00:00:00	2000	1	17	2	
19	18	1	2000-01-01 00:00:00	2199-12-31 23:59:59.999	2000-01-18 00:00:00	2000	1	18	3	
20	19	1	2000-01-01 00:00:00	2199-12-31 23:59:59.999	2000-01-19 00:00:00	2000	1	19	4	
21	20	1	2000-01-01 00:00:00	2199-12-31 23:59:59.999	2000-01-20 00:00:00	2000	1	20	5	
22	21	1	2000-01-01 00:00:00	2199-12-31 23:59:59.999	2000-01-21 00:00:00	2000	1	21	6	
23	22	1	2000-01-01 00:00:00	2199-12-31 23:59:59.999	2000-01-22 00:00:00	2000	1	22	7	

**Ilustración 90.** Información ingresada en la dimensión fecha

### 3.4.6. Implementación (Dimensiones – Hechos)

#### Es tu turno...

- ✓ Ahora vamos a intentar implementar la transformación que se muestra en la **ilustración 91 y 92**.

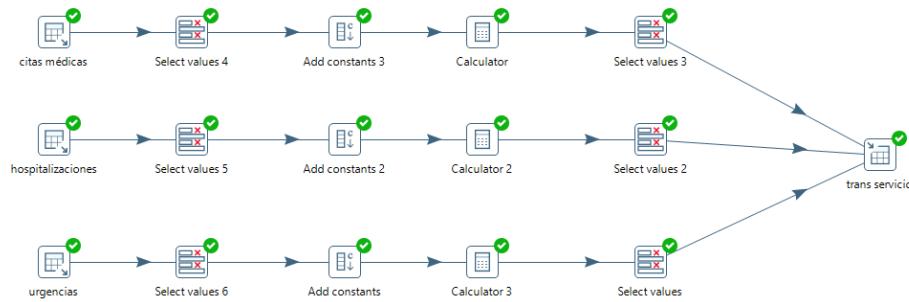


Ilustración 91. Transición Servicios

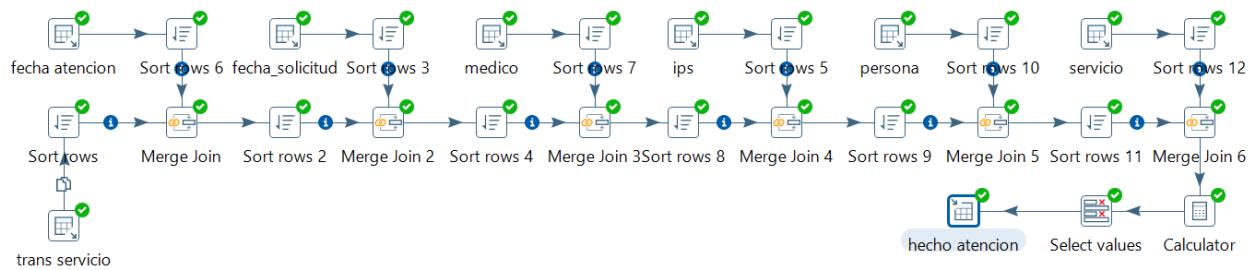


Ilustración 92. Transformación hecho atención pacientes

## 4. Practica 2

Ahora vamos a continuar con el proceso de negocio *formulas médicas*, desde su diseño hasta su implementación.

### 4.1. Diseño del Data Mart

#### Es tu turno...

Proponga un diseño de *data mart* para el proceso de negocio *formulas médicas* del caso de estudio descrito en la **sección 3.1**, de acuerdo a los siguientes requerimientos:

- Medicinas más recetadas en un período dado.
- Medicamentos que se formulan juntos.
- Medicinas que se formulan con más frecuencia.
- Costo promedio de recetas médicas.

No olvides seguir los pasos vistos en la **sección 3**:

**1. Seleccionar el proceso de negocio a modelar**

formulas médicas

**2. Escoger el nivel de granularidad del proceso.**

Encontramos dos niveles de granularidad, a nivel de medicamento y a nivel de formula.

**3. Seleccionar las dimensiones que se aplicarán a los hechos**

Dimensión fecha, persona, médico y medicamento.

**4. Escoger los hechos medibles que poblaran cada tabla de hechos**

Para el nivel de granularidad de medicamento el código de la formula.

Para el nivel de granularidad de formula el precio el cual se debe calcular a partir del precio de cada medicamento.

Al finalizar debemos tener un diseño similar al que se muestra en la **ilustración 93**.

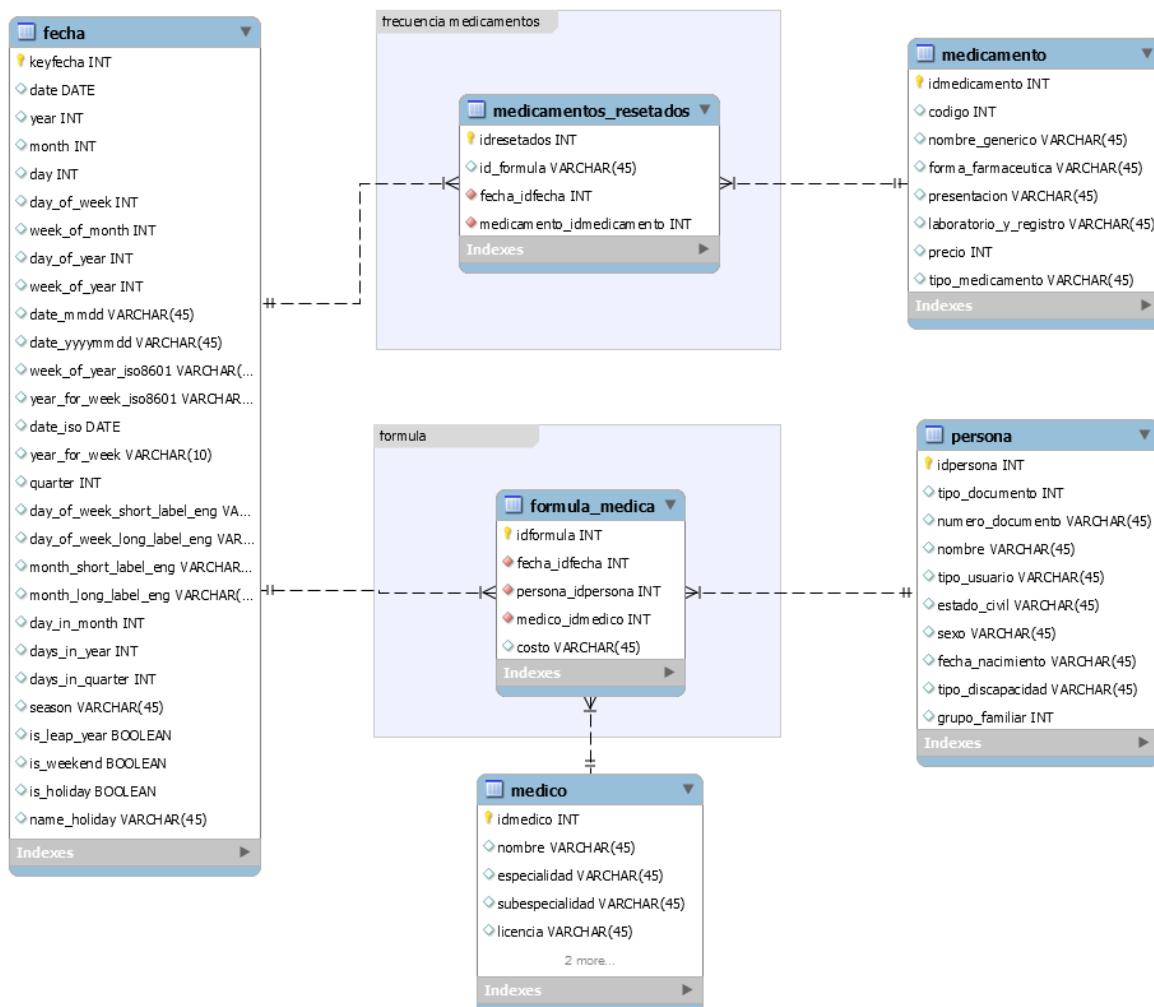


Ilustración 93. Diseño Data Marts proceso de negocio "formulas médicas"

## 4.2. Implementación del Data Mart

Como las dimensiones fecha y persona ya se encuentran creadas, solo crearemos las que hacen falta “dimensión médico”, “dimensión medicamento”.

### 4.2.1. Dimensión Médico



Ilustración 94. Transformación dimensión médico

### 4.2.2. Dimensión Medicamento



Ilustración 95. Transformación dimensión medicamento

### 4.4.3. Implementación (Dimensiones - Hechos)

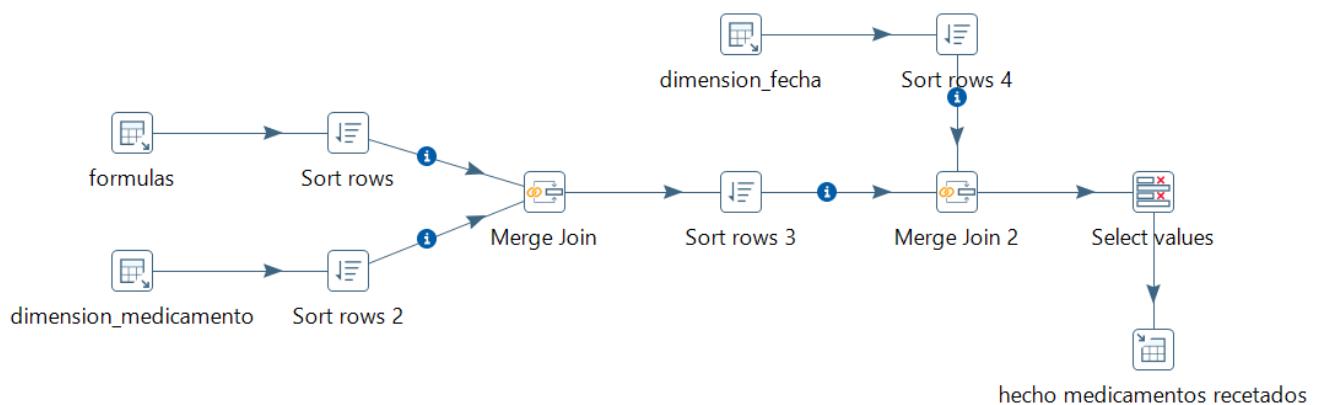


Ilustración 96. Transformación hecho "medicamentos recetados"

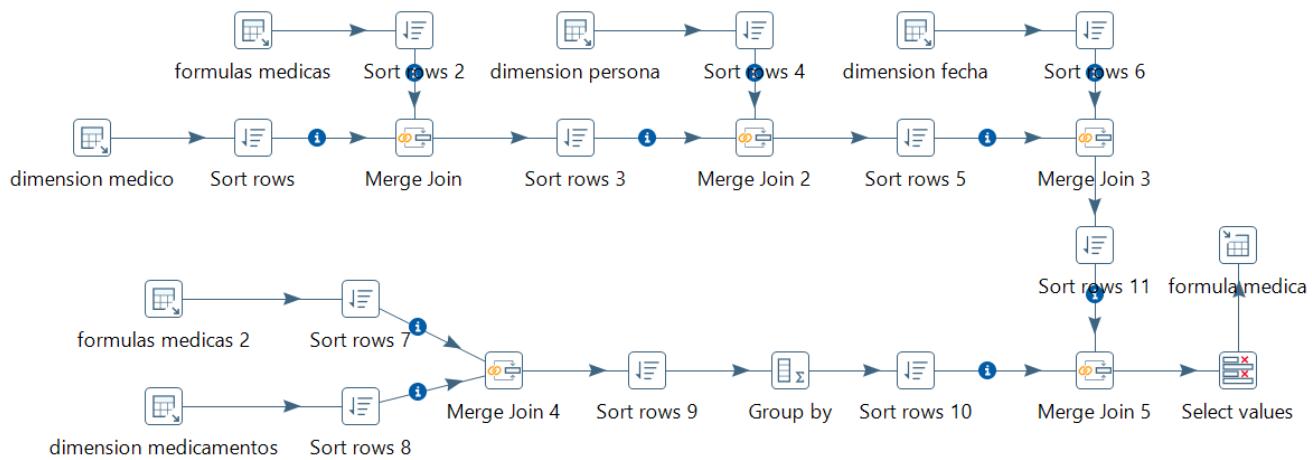


Ilustración 97. Transformación hecho "formulas medicas"