



**Universitat Oberta
de Catalunya**

Máster universitario de Ciencia de Datos

Práctica 2

**Diseño y uso de bases de datos analíticas –
identificación, diseño y desarrollo de los procesos ETL.**

Autor:

Mario Ubierna San Mamés

Índice de Contenido

Índice de Contenido	3
Índice de tablas	5
Índice de ilustraciones	6
1. Introducción	13
1.1. Presentación	13
1.2. Descripción	13
2. Identificación de los procesos ETL	15
2.1. Bloque IN	15
2.2. Bloque TR.....	17
2.2.1. Dimensiones	17
2.2.2. Hechos	18
3. Diseño y desarrollo de los procesos ETL	19
3.1. Creación de tablas	19
3.1.1. Tablas del área intermedia (<i>staging area</i>)	19
3.1.2. Tablas de las dimensiones.....	23
3.1.3. Tablas de hechos	25
3.2. Bloque IN	27
3.2.1. Definición de variables de entorno	27
3.2.2. Conexión base de datos SQL Server	28
3.2.3. Transformación IN_DENUNCIAS_INFRACCIONES	29
3.2.4. Transformación IN_POBLACION.....	34
3.2.5. Transformación IN_MOVILIDAD.....	37
3.2.6. Transformación IN_AGLOMERACION	42

3.2.7.	Transformación IN_LLAMADAS112.....	49
3.2.8.	Transformación IN_FECHAS	55
3.2.9.	Transformación IN_AMBITO_GEOGRAFICO	63
3.3.	Bloque TR Dimensiones	71
3.3.1.	Transformación TR_DIM_GRUPO_EDAD	71
3.3.2.	Transformación TR_DIM_MEDICION	75
3.3.3.	Transformación TR_DIM_TIPOLOGIA.....	79
3.3.4.	Transformación TR_DIM_AMBITO_GEOGRAFICO	83
3.3.5.	Transformación TR_DIM_FECHA.....	87
3.4.	Bloque TR Hechos	93
3.4.1.	Transformación TR_FACT_LLAMADAS112	93
3.4.2.	Transformación TR_FACT_MEDICIONES	99
4.	Implementación de jobs	126
4.1.	JOB_IN	126
4.2.	JOB_TR_DIM	129
4.3.	JOB_TR_FACT.....	131
4.4.	JOB	133

Índice de tablas

Tabla 1 - Procesos ETL Bloque IN.	17
Tabla 2 - Procesos ETL Bloque TR Dimensiones.....	17
Tabla 3 - Procesos ETL Bloque TR Hechos.....	18

Índice de ilustraciones

Ilustración 1 - STG_Denuncias_Infracciones.....	20
Ilustración 2 - STG_Poblacion.	20
Ilustración 3 - STG_Llamadas112.....	21
Ilustración 4 - STG_Movilidad.....	21
Ilustración 5 - STG_Evitar_Aglomeracion.	22
Ilustración 6 - STG_Fechas.	22
Ilustración 7 - STG_Ambito_Geografico.	23
Ilustración 8 - Tablas intermedias STG.....	23
Ilustración 9 - DIM_Ambito_Geografico.....	24
Ilustración 10 - DIM_Fecha.	24
Ilustración 11 - DIM_Grupo_Edad.	24
Ilustración 12 - DIM_Medicion.	25
Ilustración 13 - DIM_Tipologia.....	25
Ilustración 14 - Tablas de dimensiones.....	25
Ilustración 15 - FACT_Llamadas112.....	26
Ilustración 16 - FACT_Mediciones.	26
Ilustración 17 - Alter table hechos.....	27
Ilustración 18 - Tablas de hechos.....	27
Ilustración 19 - Variables de entorno.	28
Ilustración 20 - Conexión a la base de datos.	29
Ilustración 21 - IN_DENUNCIAS_INFRACCIONES.	29
Ilustración 22 - Lectura IN_DENUNCIAS_INFRACCIONES.	30
Ilustración 23 - Lectura IN_DENUNCIAS_INFRACCIONES.	30
Ilustración 24 - Lectura IN_DENUNCIAS_INFRACCIONES.	31
Ilustración 25 - Mapeo Valores IN_DENUNCIAS_INFRACCIONES.....	31

Ilustración 26 - Normalización Strings IN_DENUNCIAS_INFRACCIONES.	32
Ilustración 27 - Ordenación IN_DENUNCIAS_INFRACCIONES.	32
Ilustración 28 - Guardado IN_DENUNCIAS_INFRACCIONES.	33
Ilustración 29 - Métricas IN_DENUNCIAS_INFRACCIONES.	33
Ilustración 30 - IN_POBLACION.....	34
Ilustración 31 - Lectura IN_POBLACION.....	34
Ilustración 32 - Separación Campos IN_POBLACION.....	35
Ilustración 33 - Mapeo Valores IN_POBLACION.	35
Ilustración 34 - Normalización Strings IN_POBLACION.	36
Ilustración 35 - Guardado IN_POBLACION.....	36
Ilustración 36 - Guardado IN_POBLACION.....	37
Ilustración 37 - Métricas IN_POBLACION.....	37
Ilustración 38 - IN_MOVILIDAD.....	38
Ilustración 39 - Lectura IN_MOVILIDAD.....	38
Ilustración 40 - Mapeo Valores IN_MOVILIDAD.	39
Ilustración 41 - Normalización IN_MOVILIDAD.	39
Ilustración 42 - Replace IN_MOVILIDAD.....	40
Ilustración 43 - Select Values IN_MOVILIDAD.	40
Ilustración 44 - Guardado IN_MOVILIDAD.....	41
Ilustración 45 - Guardado IN_MOVILIDAD.....	42
Ilustración 46 - Métricas IN_MOVILIDAD.	42
Ilustración 47 - IN_AGLOMERACION.....	43
Ilustración 48 - Lectura IN_AGLOMERACION.	43
Ilustración 49 - Lectura IN_AGLOMERACION.	44
Ilustración 50 - Lectura IN_AGLOMERACIONES.....	44
Ilustración 51 - Mapeo Valores IN_AGLOMERACION.	45
Ilustración 52 - Replace IN_AGLOMERACION.....	45
Ilustración 53 - Split IN_AGLOMERACION.....	46
Ilustración 54 - Normalización Strings IN_AGLOMERACION.	46
Ilustración 55 - Replace IN_AGLOMERACION.....	47
Ilustración 56 - Normalización Filas IN_AGLOMERACION.	47
Ilustración 57 - Guardado IN_AGLOMERACIONES.....	48

Ilustración 58 - Guardado IN_AGLOMERACIONES.....	49
Ilustración 59 - Métricas IN_AGLOMERACION.	49
Ilustración 60 - IN_LLAMADAS112.....	50
Ilustración 61 - Lectura IN_LLAMADAS112.....	50
Ilustración 62 - Lectura IN_LLAMADAS112.....	51
Ilustración 63 - Lectura IN_LLAMADAS112.....	51
Ilustración 64 - Mapeo Valores IN_LLAMADAS112.	52
Ilustración 65 - Mapeo Valores IN_LLAMADAS112.	53
Ilustración 66 - Normalización IN_LLAMADAS112.....	53
Ilustración 67 - Guardado IN_LLAMADAS112.....	54
Ilustración 68 - Guardado IN_LLAMADAS112.....	55
Ilustración 69 - Métricas IN_LLAMADAS112.....	55
Ilustración 70 - IN_FECHAS.	56
Ilustración 71 - Borrado IN_FECHAS.	56
Ilustración 72 - Lectura IN_FECHAS.	57
Ilustración 73 - Añadimos Constante IN_FECHAS.....	57
Ilustración 74 - Concatenación IN_FECHAS.	58
Ilustración 75 - Conversión IN_FECHAS.	58
Ilustración 76 - Guardado IN_FECHAS.	59
Ilustración 77 - Lectura IN_FECHAS.	60
Ilustración 78 - Guardado IN_FECHAS.	61
Ilustración 79 - Lectura IN_FECHAS.	62
Ilustración 80 - Guardado IN_FECHAS.	63
Ilustración 81 - Métricas IN_FECHAS.	63
Ilustración 82 - IN_AMBITO_GEOGRAFICO.....	64
Ilustración 83 - Lectura IN_AMBITO_GEOGRAFICO.....	64
Ilustración 84 - Constatnes Comarca Municipio IN_AMBITO_GEOGRAFICO.	65
Ilustración 85 - Replace IN_AMBITO_GEOGRAFICO.....	66
Ilustración 86 - Guardado IN_AMBITO_GEOGRAFICO.....	67
Ilustración 87 - Lectura IN_AMBITO_GEOGRAFICO.....	68
Ilustración 88 - Replace IN_AMBITO_GEOGRAFICO.....	69
Ilustración 89 - Guardado IN_AMBITO_GEOGRAFICO.....	70

Ilustración 90 - Métricas IN_AMBITO_GEOGRAFICO.....	70
Ilustración 91 - TR_DIM_GRUPO_EDAD.	71
Ilustración 92 - Borrado TR_DIM_GRUPO_EDAD.	72
Ilustración 93 - Grid TR_DIM_GRUPO_EDAD.....	72
Ilustración 94 - Grid TR_DIM_GRUPO_EDAD.....	73
Ilustración 95 - Normalización TR_DIM_GRUPO_EDAD.	73
Ilustración 96 - Secuenciación TR_DIM_GRUPO_EDAD.	73
Ilustración 97 - Guardado TR_DIM_GRUPO_EDAD.	74
Ilustración 98 - Guardado TR_DIM_GRUPO_EDAD.	74
Ilustración 99 - Métricas TR_DIM_GRUPO_EDAD.	75
Ilustración 100 - TR_DIM_MEDICION.	75
Ilustración 101 - Borrado TR_DIM_MEDICION.	76
Ilustración 102 - Grid TR_DIM_DIM_MEDICION.....	76
Ilustración 103 - Grid TR_DIM_MEDICION.	77
Ilustración 104 - Normalización TR_DIM_MEDICION.	77
Ilustración 105 - Secuenciación TR_DIM_MEDICION.	78
Ilustración 106 - Guardado TR_DIM_MEDICION.	78
Ilustración 107 - Guardado TR_DIM_MEDICION.	79
Ilustración 108 - Métricas TR_DIM_MEDICIONES.	79
Ilustración 109 - TR_DIM_TIPOLOGIA.....	80
Ilustración 110 - Borrado TR_DIM_TIPOLOGIA.....	80
Ilustración 111 - Lectura TR_DIM_TIPOLOGIA.....	81
Ilustración 112 - Secuenciación TR_DIM_TIPOLOGIA.....	82
Ilustración 113 - Guardado TR_DIM_TIPOLOGIA.....	82
Ilustración 114 - Guardado TR_DIM_TIPOLOGIA.....	83
Ilustración 115 - Métricas TR_DIM_TIPOLOGIA.	83
Ilustración 116 - TR_DIM_AMBITO_GEOGRAFICO.	84
Ilustración 117 - Borrado TR_DIM_AMBITO_GEOGRAFICO.	84
Ilustración 118 - Lectura TR_DIM_AMBITO_GEOGRAFICO.	85
Ilustración 119 - Secuenciación TR_DIM_AMBITO_GEOGRAFICO.	86
Ilustración 120 - Guardado TR_DIM_AMBITO_GEOGRAFICO.	86
Ilustración 121 - Guardado TR_DIM_AMBITO_GEOGRAFICO.	87

Ilustración 122 - Métricas TR_DIM_AMBITO_GEOGRAFICO.	87
Ilustración 123 - TR_DIM_FECHA.	88
Ilustración 124 - Borrado TR_DIM_FECHA.	88
Ilustración 125 - Lectura TR_DIM_FECHA.	89
Ilustración 126 - Conversión String TR_DIM_FECHA.	89
Ilustración 127 - Split TR_DIM_FECHA.	90
Ilustración 128 - Split TR_DIM_FECHA.	90
Ilustración 129 - Concatenación TR_DIM_FECHA.	91
Ilustración 130 - Conversión TR_DIM_FECHA.	91
Ilustración 131 - Secuenciación TR_DIM_FECHA.	92
Ilustración 132 - Guardado TR_DIM_FECHA.	92
Ilustración 133 - Métricas TR_DIM_FECHA.	93
Ilustración 134 - TR_FACT_LLAMADAS112.	93
Ilustración 135 - Lectura TR_FACT_LLAMADAS112.	94
Ilustración 136 - Añadimos Día TR_FACT_LLAMADAS112.	94
Ilustración 137 - Replace TR_FACT_LLAMADAS112.	95
Ilustración 138 - Búsqueda PK Fecha TR_FACT_LLAMADAS112.	96
Ilustración 139 - Búsqueda PK_Ambito_Geografico TR_FACT_LLAMADAS112.	97
Ilustración 140 - Obtención PK_Tipología TR_FACT_LLAMADAS112.	98
Ilustración 141 - Guardado TR_FACT_LLAMADAS112.	99
Ilustración 142 - Métricas TR_FACT_LLAMADAS112.	99
Ilustración 143 - TR_FACT_MEDICIONES.	100
Ilustración 144 - Lectura TR_FACT_MEDICIONES.	101
Ilustración 145 - Lectura TR_FACT_MEDICIONES.	101
Ilustración 146 - Lectura TR_FACT_MEDICIONES.	102
Ilustración 147 - Lectura TR_FACT_MEDICIONES.	102
Ilustración 148 - Añadimos Campos TR_FACT_MEDICIONES.	103
Ilustración 149 - Añadimos Campos TR_FACT_MEDICIONES.	103
Ilustración 150 - Añadimos Campos TR_FACT_MEDICIONES.	103
Ilustración 151 - Añadimos Campos TR_FACT_MEDICIONES.	104
Ilustración 152 - Búsqueda PK_Fecha TR_FACT_MEDICIONES.	104
Ilustración 153 - Búsqueda PK_Fecha TR_FACT_MEDICIONES.	105

Ilustración 154 - Búsqueda PK_Fecha TR_FACT_MEDICIONES.....	106
Ilustración 155 - Búsqueda PK_Fecha TR_FACT_MEDICIONES.....	107
Ilustración 156 - Búsqueda PK_Ambito_Geografico TR_FACT_MEDICIONES.	108
Ilustración 157 - Búsqueda PK_Ambito_Geografico TR_FACT_MEDICIONES.	109
Ilustración 158 - Búsqueda PK_Ambito_Geografico TR_FACT_MEDICIONES.	110
Ilustración 159 - Búsqueda PK_Ambito_Geografico TR_FACT_MEDICIONES.	111
Ilustración 160 - Búsqueda PK_Grupo_Edad TR_FACT_MEDICIONES.....	112
Ilustración 161 - Búsqueda PK_Grupo_Edad TR_FACT_MEDICIONES.....	113
Ilustración 162 - Búsqueda PK_Grupo_Edad TR_FACT_MEDICIONES.....	114
Ilustración 163 - Búsqueda PK_Grupo_Edad TR_FACT_MEDICIONES.....	115
Ilustración 164 - Búsquedas PK_Medicion TR_FACT_MEDICIONES.	116
Ilustración 165 - Búsqueda PK_Medicion TR_FACT_MEDICIONES.....	117
Ilustración 166 - Búsqueda PK_Medicion TR_FACT_MEDICIONES.....	118
Ilustración 167 - Búsqueda PK_Medicion TR_FACT_MEDICIONES.....	119
Ilustración 168 - Secuenciación TR_FACT_MEDICIONES.	120
Ilustración 169 - Guardado TR_FACT_MEDICIONES.	121
Ilustración 170 - Guardado TR_FACT_MEDICIONES.	122
Ilustración 171 - Guardado TR_FACT_MEDICIONES.	123
Ilustración 172 - Guardado TR_FACT_MEDICIONES.	124
Ilustración 173 - Métricas TR_FACT_MEDICIONES.	124
Ilustración 174 - Métricas TR_FACT_MEDICIONES.	125
Ilustración 175 - Métricas TR_FACT_MEDICIONES.	125
Ilustración 176 - JOB_IN.....	126
Ilustración 177 – Transformación IN_DENUNCIAS_INFRACCIONES.	127
Ilustración 178 - Transformación IN_POBLACION.	127
Ilustración 179 - Transformación IN_MOVILIDAD.	127
Ilustración 180 - Transformación IN_AGLOMERACION.	128
Ilustración 181 Transformación IN_LLAMADAS112.	128
Ilustración 182 – Transformación IN_FECHAS.	128
Ilustración 183 - Transformación IN_AMBITO_GEOGRAFICO.	129
Ilustración 184 - Métricas JOB_IN.....	129
Ilustración 185 - JOB_TR_DIM.	129

Ilustración 186 - Transformación TR_DIM_GRUPO_EDAD.....	130
Ilustración 187 - Transformación TR_DIM_MEDICION.....	130
Ilustración 188 - Transformación TR_DIM_TIPOLOGIA.	130
Ilustración 189 - Transformación TR_DIM_AMBITO_GEOGRAFICO.....	131
Ilustración 190 - Transformación TR_DIM_FECHA.	131
Ilustración 191 - Métricas JOB_TR_DIM.	131
Ilustración 192 - JOB_TR_FACT.	132
Ilustración 193 - Transformación TR_FACT_LLAMADAS112.....	132
Ilustración 194 - Transformación TR_FACT_MEDICIONES.....	132
Ilustración 195 - Métricas JOB_TR_FACT.....	133
Ilustración 196 - JOB	133
Ilustración 197 - Trabajo JOB_IN.....	133
Ilustración 198 - Trabajo JOB_TR_DIM.	134
Ilustración 199 - Trabajo JOB_TR_FACT.....	134
Ilustración 200 - Métricas JOB.	134

1. Introducción

1.1. Presentación

A partir de la solución oficial de la primera práctica (PRA1), el estudiante debe diseñar, implementar y ejecutar los procesos de extracción, transformación y carga de los datos de las fuentes de datos proporcionadas.

Así pues, esta actividad tiene como objetivo identificar y desarrollar los procesos de carga del almacén de datos y que esta sea efectiva.

1.2. Descripción

Si nos centramos en los subobjetivos, esta segunda parte del caso práctico consiste en lo siguiente:

- Identificar los procesos de extracción, transformación y carga de datos (ETL) hacia el almacén de datos.
- Diseñar y desarrollar los procesos ETL mediante las herramientas de diseño proporcionadas.
- Implementar con los trabajos (*jobs*) los procesos ETL para que su carga planificada sea efectiva.

Además del documento con la solución de la PRA2 que se debe entregar, también se tendrá en consideración la implementación sobre la máquina virtual proporcionada en el curso.

En resumen, el documento de la solución de la PRA2 debe incluir los siguientes aspectos:

- Descripción de todas las acciones que se han realizado.

- Capturas de pantalla que muestren todas las partes significativas del ETL, sus características y su correspondiente explicación.
- Capturas de pantalla que demuestren la correcta ejecución de la ETL y el tiempo de ejecución.
- Capturas de pantalla que demuestren la correcta carga de los datos (cargados en la base de datos).

2. Identificación de los procesos ETL

A la hora de diseñar los procesos de carga de una base de datos analítica no hay una única estrategia. Es habitual estructurar los procesos ETL sobre la base de las entidades de datos que se deben actualizar, ya que existen diferencias conceptuales en la actualización de una dimensión con respecto a la de una tabla de hechos. La división del proceso de carga inicial en diferentes bloques de actualización facilitará el diseño de un orden de ejecución y la gestión de las dependencias. Cada uno de estos bloques de actualización se dividirá en las correspondientes etapas de extracción, transformación y carga.

Se identifican los dos bloques siguientes:

- **Bloque IN:** procesos de carga de los datos desde las fuentes a las tablas intermedias en el área de maniobras (*staging area*). Estos procesos se distinguen por el prefijo «IN_» en el nombre.
- **Bloque TR:** procesos de transformación para cargar los datos desde las tablas intermedias hasta nuestro almacén, según el modelo multidimensional diseñado. Así pues, son diferentes los procesos ETL de transformación para cargar las dimensiones de aquellos que se realizan para cargar las tablas de hechos. Estos procesos se distinguen con el prefijo «TR_» en el nombre.

2.1. Bloque IN

Respecto al bloque In, el cual nos va a permitir almacenar la información en el staging area, tenemos los siguientes procesos:

Nombre ETL	Descripción	Orígenes de los datos	Tabla de destino (stage)
------------	-------------	-----------------------	--------------------------

IN_DENUNCIAS_INFRACCIONES	Carga de los datos correspondientes a las estadísticas sobre los expedientes incoados por el artículo 36.6 LOPSC de desobediencia durante el estado de emergencia sanitaria COVID-19 en la comunidad de Euskadi.	ACUMULADO-DENUNCIAS-INFRACCIONES.xlsx	STG_Denuncias_Infracciones
IN_POBLACION	Carga los datos respectivos a las cifras de la población española.	población_9687bsc.csv	STG_Poblacion
IN_MOVILIDAD	Movilidad de la población durante el estado de alarma.	35167bsc.csv	STG_Movilidad
IN_AGLOMERACION	Porcentaje de la población que evitaba las aglomeraciones con motivo del coronavirus, por grupo de edad y provincia.	statistic_id1104235_covid19_-poblacion-que-evitabalas-aglomeraciones-seguridad-en-espana-2020.xlsx	STG_Evitar_Aglomeracion
IN_LLAMADAS_112	Llamadas al 112 por ámbito geográfico y tipología (accidentes de tráfico, civismo, incendios, asistencia sanitaria, seguridad...)	rows.xml	STG_Llamadas112
IN_FECHAS	Almacenamos todas las fechas de todos los ficheros de datos proporcionados.	STG_Llamadas112 STG_Denuncias_Infracciones STG_Movilidad	STG_Fechas
IN_AMBITO_GEOGRAFICO	Almacenamos todas las localización de todos los	STG_Evitar_Aglomerationes	STG_Ambito_Geografico

ficheros de datos proporcionados.	STG_Poblacion
	STG_Llamadas112

Tabla 1 - Procesos ETL Bloque IN.

2.2. Bloque TR

Respecto al bloque TR tenemos tanto los procesos para dotar de datos a las dimensiones como a los hechos.

2.2.1. Dimensiones

Los procesos ETL que se encargan de añadir la información a las dimensiones son los siguientes:

Nombre del ETL	Descripción	Tabla de origen	Tabla de destino (dimensión)
TR_DIM_FECHA	Carga y transformación de la dimensión temporal.	STG_Fechas	DIM_Fecha
TR_DIM_AMBITO_GEOGRAFICO	Carga y transformación de la dimensión con los datos de los ámbitos geográficos.	STG_Ambito_Geografico	DIM_Ambito_Geografico
TR_DIM_GRUPO_EDAD	Carga y transformación de la dimensión con los datos de los grupos de edad.	Manual, a partir de un grid.	DIM_grupo_Edad
TR_DIM_MEDICION	Carga y transformación de la dimensión con los datos de las mediciones.	Manual, a partir de un grid.	DIM_Medicion
TR_DIM_TIPOLOGIA	Carga y transformación de la dimensión con los datos de la tipología.	STG_Llamadas112	DIM_Tipologia

Tabla 2 - Procesos ETL Bloque TR Dimensiones.

2.2.2. Hechos

Respecto a los hechos tenemos los siguientes procesos de carga:

Nombre del ETL	Descripción	Tabla de origen
TR_FACT_LLAMADAS112	Carga y transformación de la tabla de hechos Fact_Llamadas112.	STG_Llamadas112
TR_FACT_MEDICIONES	Carga y transformación de la tabla de hechos Fact_Mediciones	STG_Denuncias_infracciones STG_Evitar_Aglomeracion STG_Movilidad STG_Poblacion

Tabla 3 - Procesos ETL Bloque TR Hechos.

3. Diseño y desarrollo de los procesos ETL

En este apartado, se deben diseñar los procesos de carga identificados en el punto anterior con la herramienta de diseño proporcionada. En este caso es Pentho Data Integration (PDI).

3.1. Creación de tablas

El primer paso para la implementación de los procesos ETL consiste en la creación de las tablas. Esto se llevará a cabo una única vez, mediante *scripts*, sobre la base de datos proporcionada (en nuestro caso: SQL Server). Se deberán crear las tablas intermedias y las tablas del modelo dimensional de la solución oficial, es decir, las dimensiones y las tablas de hechos. Para hacerlo, deben utilizarse los *scripts* facilitados junto a la solución de la PRA1.

3.1.1. Tablas del área intermedia (*staging area*)

Lo primero que vamos a hacer es la creación de las tablas intermedias:

Tabla intermedia STG_Denuncias_Infracciones

```

USE [DB_marioum]
GO

/***** Object: Table [dbo].[STG_Denuncias_Infracciones]
SET ANSI_NULLS ON
GO

SET QUOTED_IDENTIFIER ON
GO

CREATE TABLE [dbo].[STG_Denuncias_Infracciones](
    [provincia] [varchar](100) NULL,
    [identificados_ertzaintza] [float] NULL,
    [detenidos_ertzaintza] [float] NULL,
    [denuncias_ertzaintza] [float] NULL,
    [vehic_intercept_ertzaintza] [float] NULL,
    [identificados_ppll] [float] NULL,
    [detenidos_ppll] [float] NULL,
    [denuncias_ppll] [float] NULL,
    [vehic_intercept_ppll] [float] NULL,
    [fecha] [datetime] NULL
) ON [PRIMARY]
GO

```

Ilustración 1 - STG_Denuncias_Infracciones.

Tabla intermedia STG_Poblacion

```

USE [DB_marioum]
GO

/***** Object: Table [dbo].[STG_Denuncias_Infracciones]
SET ANSI_NULLS ON
GO

SET QUOTED_IDENTIFIER ON
GO

CREATE TABLE [dbo].[STG_Poblacion](
    [provincia_codigo] [varchar](2) NULL,
    [provincia_nombre] [varchar](100) NULL,
    [poblacion] [bigint] NULL,
    [periodo] [varchar](25) NULL
) ON [PRIMARY]
GO

```

Ilustración 2 - STG_Poblacion.

Tabla intermedia STG_Llamadas112

```

USE [DB_marioum]
GO

/***** Object: Table [dbo].[STG_Denuncias_Infracciones]
SET ANSI_NULLS ON
GO

SET QUOTED_IDENTIFIER ON
GO

CREATE TABLE [dbo].[STG_Llamadas112](
    [anio] [int] NULL,
    [mes] [int] NULL,
    [provincia] [varchar](100) NULL,
    [comarca] [varchar](100) NULL,
    [municipio] [varchar](100) NULL,
    [tipo] [varchar](100) NULL,
    [llamadas] [int] NULL
) ON [PRIMARY]

```

*Ilustración 3 - STG_Llamadas112.***Tabla intermedia STG_Movilidad**

```

USE [DB_marioum]
GO

/***** Object: Table [dbo].[STG_Denuncias_Infracciones]
SET ANSI_NULLS ON
GO

SET QUOTED_IDENTIFIER ON
GO

CREATE TABLE [dbo].[STG_Movilidad](
    [zonas_movilidad] [varchar](27) NULL,
    [periodo] [datetime] NULL,
    [total] [decimal](5, 2) NULL
) ON [PRIMARY]
GO

```

Ilustración 4 - STG_Movilidad.

Tabla intermedia STG_Evitar_Aglomeracion

```

USE [DB_marioum]
GO

/***** Object: Table [dbo].[STG_Denuncias_Infracciones]
SET ANSI_NULLS ON
GO

SET QUOTED_IDENTIFIER ON
GO

CREATE TABLE [dbo].[STG_Evitar_Aglomeracion](
    [provincia] [varchar](100) NULL,
    [comunidad_autonoma] [varchar](100) NULL,
    [grupo_edad] [varchar](7) NULL,
    [porc_poblacion] [float] NULL
) ON [PRIMARY]
GO

```

Ilustración 5 - STG_Evitar_Aglomeracion.

Tabla intermedia STG_Fechas

```

USE [DB_marioum]
GO

/***** Object: Table [dbo].[STG_Fechas]
SET ANSI_NULLS ON
GO

SET QUOTED_IDENTIFIER ON
GO

CREATE TABLE [dbo].[STG_Fechas](
    [fecha] [date] NOT NULL
) ON [PRIMARY]
GO

```

Ilustración 6 - STG_Fechas.

Tabla intermedia STG_Ambito_Geografico

```
USE [DB_mariouism]
GO

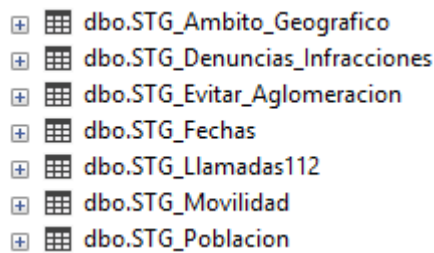
/***** Object: Table [dbo].[STG_Ambito_Geografico]
SET ANSI_NULLS ON
GO

SET QUOTED_IDENTIFIER ON
GO

CREATE TABLE [dbo].[STG_Ambito_Geografico](
    [provincia_codigo] [varchar](2) NOT NULL,
    [provincia_nombre] [varchar](100) NOT NULL,
    [comunidad_autonoma] [varchar](100) NULL,
    [comarca] [varchar](100) NULL,
    [municipio] [varchar](100) NULL
) ON [PRIMARY]
GO
```

Ilustración 7 - STG_Ambito_Geografico.

Comprobamos que todas las tablas intermedias se han creado de forma correcta:



A screenshot of the SQL Server Enterprise Manager interface showing a list of tables in the 'dbo' schema. The list includes: dbo.STG_Ambito_Geografico, dbo.STG_Denuncias_Infracciones, dbo.STG_Evitar_Aglomeracion, dbo.STG_Fechas, dbo.STG_Llamadas112, dbo.STG_Movilidad, and dbo.STG_Poblacion. Each table name is preceded by a small icon representing a table.

*Ilustración 8 - Tablas intermedias STG.***3.1.2. Tablas de las dimensiones**

Lo segundo que debemos de hacer es la creación de las tablas de dimensiones:

Tabla dimensión DIM_Ambito_Geografico

```
USE [DB_marioum]
GO

/***** Object: Table [dbo].[STG_Denuncias_Infracciones] *****/
SET ANSI_NULLS ON
GO

SET QUOTED_IDENTIFIER ON
GO

CREATE TABLE [dbo].[DIM_Ambito_Geografico](
    [pk_ambito_geografico] [int] NOT NULL,
    [provincia_codigo] [varchar](2) NOT NULL,
    [provincia_nombre] [varchar](100) NOT NULL,
    [comunidad_autonoma] [varchar](100) NULL,
    [comarca] [varchar](100) NULL,
    [municipio] [varchar](100) NULL,
    CONSTRAINT [PK_DIM_Ambito_Geografico] PRIMARY KEY CLUSTERED
(
    [pk_ambito_geografico] ASC
)WITH (PAD_INDEX = OFF, STATISTICS_NORECOMPUTE = OFF, IGNORE_DUP_KEY = OFF, ALLOW_ROW_LOCKS = ON, ALLOW_PAGE_LOCKS = ON) ON [PRIMARY]
) ON [PRIMARY]
GO
```

Ilustración 9 - DIM_Ambito_Geografico.

Tabla dimensión DIM_Fecha

```
USE [DB_marioum]
GO

/***** Object: Table [dbo].[STG_Denuncias_Infracciones] *****/
SET ANSI_NULLS ON
GO

SET QUOTED_IDENTIFIER ON
GO

)CREATE TABLE [dbo].[DIM_Fecha](
    [pk_fecha] [int] NOT NULL,
    [anyo] [int] NOT NULL,
    [mes] [int] NOT NULL,
    [dia] [int] NOT NULL,
    [fecha] [date] NOT NULL,
    CONSTRAINT [PK_DIM_Fecha] PRIMARY KEY CLUSTERED
(
    [pk_fecha] ASC
)WITH (PAD_INDEX = OFF, STATISTICS_NORECOMPUTE = OFF, IGNORE_DUP_KEY = OFF, ALLOW_ROW_LOCKS = ON, ALLOW_PAGE_LOCKS = ON) ON [PRIMARY]
) ON [PRIMARY]
GO
```

Ilustración 10 - DIM_Fecha.

Tabla dimensión DIM_Grupo_Edad

```
USE [DB_marioum]
GO

/***** Object: Table [dbo].[STG_Denuncias_Infracciones] *****/
SET ANSI_NULLS ON
GO

SET QUOTED_IDENTIFIER ON
GO

CREATE TABLE [dbo].[DIM_Grupo_Edad](
    [pk_grupo_edad] [int] NOT NULL,
    [nombre] [varchar](20) NOT NULL,
    [intervalo] [varchar](20) NOT NULL,
    CONSTRAINT [PK_DIM_Grupo_Edad] PRIMARY KEY CLUSTERED
(
    [pk_grupo_edad] ASC
)WITH (PAD_INDEX = OFF, STATISTICS_NORECOMPUTE = OFF, IGNORE_DUP_KEY = OFF, ALLOW_ROW_LOCKS = ON, ALLOW_PAGE_LOCKS = ON) ON [PRIMARY]
) ON [PRIMARY]
GO
```

Ilustración 11 - DIM_Grupo_Edad.

Tabla dimensión DIM_Medicion

```
USE [DB_marioum]
GO

/***** Object: Table [dbo].[STG_Denuncias_Infracciones] *****/
SET ANSI_NULLS ON
GO

SET QUOTED_IDENTIFIER ON
GO

CREATE TABLE [dbo].[DIM_Medicion](
    [pk_medicion] [int] NOT NULL,
    [nombre] [varchar](100) NOT NULL,
    [unidad_medida] [varchar](20) NOT NULL,
    CONSTRAINT [PK_DIM_Medicion] PRIMARY KEY CLUSTERED
(
    [pk_medicion] ASC
)WITH (PAD_INDEX = OFF, STATISTICS_NORECOMPUTE = OFF, IGNORE_DUP_KEY = OFF, ALLOW_ROW_LOCKS = ON, ALLOW_PAGE_LOCKS = ON) ON [PRIMARY]
) ON [PRIMARY]
GO
```

Ilustración 12 - DIM_Medicion.

Tabla dimensión DIM_Tipologia

```
USE [DB_marioum]
GO

/***** Object: Table [dbo].[STG_Denuncias_Infracciones] *****/
SET ANSI_NULLS ON
GO

SET QUOTED_IDENTIFIER ON
GO

CREATE TABLE [dbo].[DIM_Tipologia](
    [pk_tipologia] [int] NOT NULL,
    [nombre] [varchar](100) NOT NULL,
    CONSTRAINT [PK_DIM_Tipologia] PRIMARY KEY CLUSTERED
(
    [pk_tipologia] ASC
)WITH (PAD_INDEX = OFF, STATISTICS_NORECOMPUTE = OFF, IGNORE_DUP_KEY = OFF, ALLOW_ROW_LOCKS = ON, ALLOW_PAGE_LOCKS = ON) ON [PRIMARY]
) ON [PRIMARY]
GO
```

Ilustración 13 - DIM_Tipologia.

Comprobamos que todas las tablas de dimensiones se han creado correctamente:

```
+  dbo.DIM_Ambito_Geografico
+  dbo.DIM_Fecha
+  dbo.DIM_Grupo_Edad
+  dbo.DIM_Medicion
+  dbo.DIM_Tipologia
```

Ilustración 14 - Tablas de dimensiones.

3.1.3. Tablas de hechos

Finalmente creamos las diferentes tablas de los hechos:

Tabla hecho FACT_Llamadas112

```
USE [DB_marioum]
GO

/***** Object: Table [dbo].[STG_Denuncias_Infracciones] *****/
SET ANSI_NULLS ON
GO

SET QUOTED_IDENTIFIER ON
GO

CREATE TABLE [dbo].[FACT_Llamadas112](
    [pk_fk_fecha] [int] NOT NULL,
    [pk_fk_ambito_geografico] [int] NOT NULL,
    [pk_fk_tipologia] [int] NOT NULL,
    [llamadas] [int] NULL,
    CONSTRAINT [PK_FACT_Llamadas112] PRIMARY KEY CLUSTERED
(
    [pk_fk_fecha] ASC,
    [pk_fk_ambito_geografico] ASC,
    [pk_fk_tipologia] ASC
)WITH (PAD_INDEX = OFF, STATISTICS_NORECOMPUTE = OFF, IGNORE_DUP_KEY = OFF, ALLOW_ROW_LOCKS = ON, ALLOW_PAGE_LOCKS = ON) ON [PRIMARY]
) ON [PRIMARY]
GO
```

Ilustración 15 - FACT_Llamadas112.

Tabla hecho FACT_Mediciones

```
USE [DB_marioum]
GO

/***** Object: Table [dbo].[STG_Denuncias_Infracciones] *****/
SET ANSI_NULLS ON
GO

SET QUOTED_IDENTIFIER ON
GO

CREATE TABLE [dbo].[FACT_Mediciones](
    [pk_id] [int] NOT NULL,
    [fk_fecha] [int] NOT NULL,
    [fk_ambito_geografico] [int] NOT NULL,
    [fk_grupo_edad] [int] NOT NULL,
    [fk_medicion] [int] NOT NULL,
    [valor] [decimal](17, 2) NULL,
    CONSTRAINT [PK_FACT_Mediciones] PRIMARY KEY CLUSTERED
(
    [pk_id] ASC
)WITH (PAD_INDEX = OFF, STATISTICS_NORECOMPUTE = OFF, IGNORE_DUP_KEY = OFF, ALLOW_ROW_LOCKS = ON, ALLOW_PAGE_LOCKS = ON) ON [PRIMARY]
) ON [PRIMARY]
GO
```

Ilustración 16 - FACT_Mediciones.

Realizamos los alter table de las tablas de hechos:

```

ALTER TABLE [dbo].[FACT_Llamadas112] WITH CHECK ADD CONSTRAINT [FK_FACT_Llamadas112_DIM_Ambito_Geografico] FOREIGN KEY([pk_fk_ambito_geografico])
REFERENCES [dbo].[DIM_Ambito_Geografico] ([pk_ambito_geografico])
GO

ALTER TABLE [dbo].[FACT_Llamadas112] CHECK CONSTRAINT [FK_FACT_Llamadas112_DIM_Ambito_Geografico]
GO

ALTER TABLE [dbo].[FACT_Llamadas112] WITH CHECK ADD CONSTRAINT [FK_FACT_Llamadas112_DIM_Fecha] FOREIGN KEY([pk_fk_fecha])
REFERENCES [dbo].[DIM_Fecha] ([pk_fecha])
GO

ALTER TABLE [dbo].[FACT_Llamadas112] CHECK CONSTRAINT [FK_FACT_Llamadas112_DIM_Fecha]
GO

ALTER TABLE [dbo].[FACT_Llamadas112] WITH CHECK ADD CONSTRAINT [FK_FACT_Llamadas112_DIM_Tipologia] FOREIGN KEY([pk_fk_tipologia])
REFERENCES [dbo].[DIM_Tipologia] ([pk_tipologia])
GO

ALTER TABLE [dbo].[FACT_Llamadas112] CHECK CONSTRAINT [FK_FACT_Llamadas112_DIM_Tipologia]
GO

ALTER TABLE [dbo].[FACT_Mediciones] WITH CHECK ADD CONSTRAINT [FK_FACT_Mediciones_DIM_Ambito_Geografico] FOREIGN KEY([fk_ambito_geografico])
REFERENCES [dbo].[DIM_Ambito_Geografico] ([pk_ambito_geografico])
GO

ALTER TABLE [dbo].[FACT_Mediciones] CHECK CONSTRAINT [FK_FACT_Mediciones_DIM_Ambito_Geografico]
GO

ALTER TABLE [dbo].[FACT_Mediciones] WITH CHECK ADD CONSTRAINT [FK_FACT_Mediciones_DIM_Fecha] FOREIGN KEY([fk_fecha])
REFERENCES [dbo].[DIM_Fecha] ([pk_fecha])
GO

ALTER TABLE [dbo].[FACT_Mediciones] CHECK CONSTRAINT [FK_FACT_Mediciones_DIM_Fecha]
GO

ALTER TABLE [dbo].[FACT_Mediciones] WITH CHECK ADD CONSTRAINT [FK_FACT_Mediciones_DIM_Grupo_Edad] FOREIGN KEY([fk_grupo_edad])
REFERENCES [dbo].[DIM_Grupo_Edad] ([pk_grupo_edad])
GO

ALTER TABLE [dbo].[FACT_Mediciones] CHECK CONSTRAINT [FK_FACT_Mediciones_DIM_Grupo_Edad]
GO

ALTER TABLE [dbo].[FACT_Mediciones] WITH CHECK ADD CONSTRAINT [FK_FACT_Mediciones_DIM_Medicion] FOREIGN KEY([fk_medicion])
REFERENCES [dbo].[DIM_Medicion] ([pk_medicion])
GO

ALTER TABLE [dbo].[FACT_Mediciones] CHECK CONSTRAINT [FK_FACT_Mediciones_DIM_Medicion]
GO

```

Ilustración 17 - Alter table hechos.

Comprobamos que se han creado todas las tablas correspondientes:

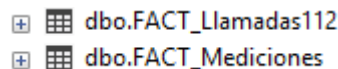


Ilustración 18 - Tablas de hechos.

3.2. Bloque IN

En este bloque se van a realizar las transformaciones para que la información en forma bruta se pase a las tablas intermedias, y luego haremos uso de éstas para crear las transformaciones de dimensiones y hechos.

3.2.1. Definición de variables de entorno

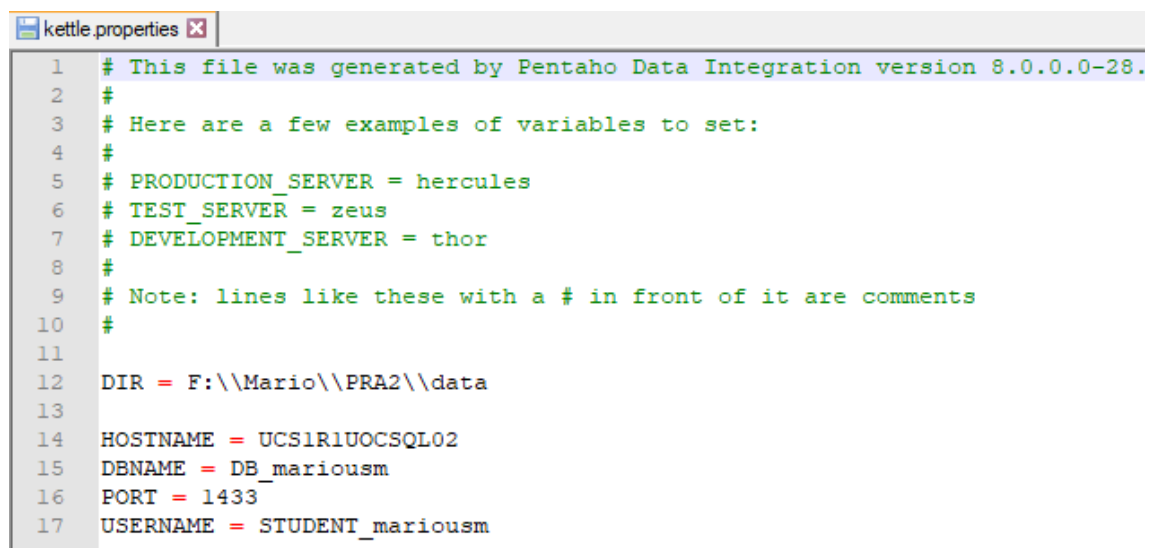
Es una buena práctica utilizar variables de entorno para así poder evitar errores en el definiciones futuras. Para ello accedemos a *kettle.properties* y definimos las siguientes variables:

Para el origen en el que se encuentran todos los archivos definimos la variables DIR_ENT:

- Nombre: DIR
- Valor: F:\Mario\PRA2\data

Para la cadena de conexión a la base de datos vamos a usar:

- Nombre: HOSTNAME
- Valor: UCS1R1UOCSQL02
- Nombre: DBNAME
- Valor: DB_mariouism
- Nombre: PORT
- Valor: 1433
- Nombre: USERNAME
- Valor: STUDENT_mariouism

A screenshot of the 'kettle.properties' window in Pentaho Data Integration. The window has a title bar with a close button. The text inside is as follows:

```
1  # This file was generated by Pentaho Data Integration version 8.0.0.0-28.
2  #
3  # Here are a few examples of variables to set:
4  #
5  # PRODUCTION_SERVER = hercules
6  # TEST_SERVER = zeus
7  # DEVELOPMENT_SERVER = thor
8  #
9  # Note: lines like these with a # in front of it are comments
10 #
11
12 DIR = F:\\Mario\\PRA2\\data
13
14 HOSTNAME = UCS1R1UOCSQL02
15 DBNAME = DB_mariouism
16 PORT = 1433
17 USERNAME = STUDENT_mariouism
```

Ilustración 19 - Variables de entorno.

3.2.2. Conexión base de datos SQL Server

El siguiente paso es crear la conexión a la base de datos que va a ser usada tanto por las transformaciones como por los jobs que se realicen en esta práctica.

Para ello creamos la nueva conexión y establecemos los valores definidos en las variables de entorno:

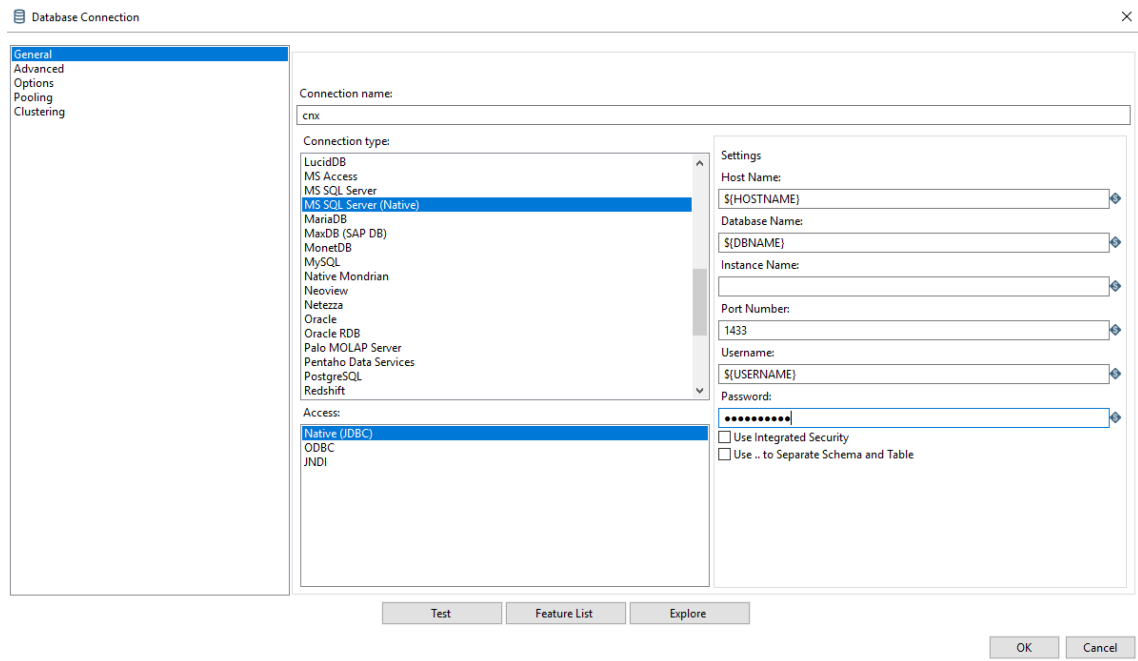


Ilustración 20 - Conexión a la base de datos.

3.2.3. Transformación IN_DENUNCIAS_INFRACCIONES

Una vez que ya hemos definido las variables de entorno y la conexión podemos proceder a realizar todas las transformaciones y trabajos.

La primera transformación que vamos a realizar se llama “IN_DENUNCIAS_INFRACCIONES”, su objetivo es leer todos los datos del archivo “ACUMULADO-DENUNCIAS-INFRACCIONES.xlsx” en la tabla intermedia “STG_Denuncias_Infracciones”.

En este caso no hemos hecho ninguna modificación en el Excel original, por lo que la transformación nos queda de la siguiente forma:

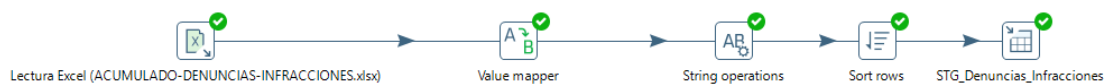


Ilustración 21 - IN_DENUNCIAS_INFRACCIONES.

Ahora vamos a explicar paso a paso lo que hemos hecho:

Lectura del Excel

Lo primero de todo es leer el fichero Excel que se nos proporciona, y para ello usamos el componente “Microsoft Excel Input”, una vez hecho eso escribimos el nombre del paso, le indicamos el fichero que va a utilizar, y le indicamos que el formato del fichero Excel es XLSX:

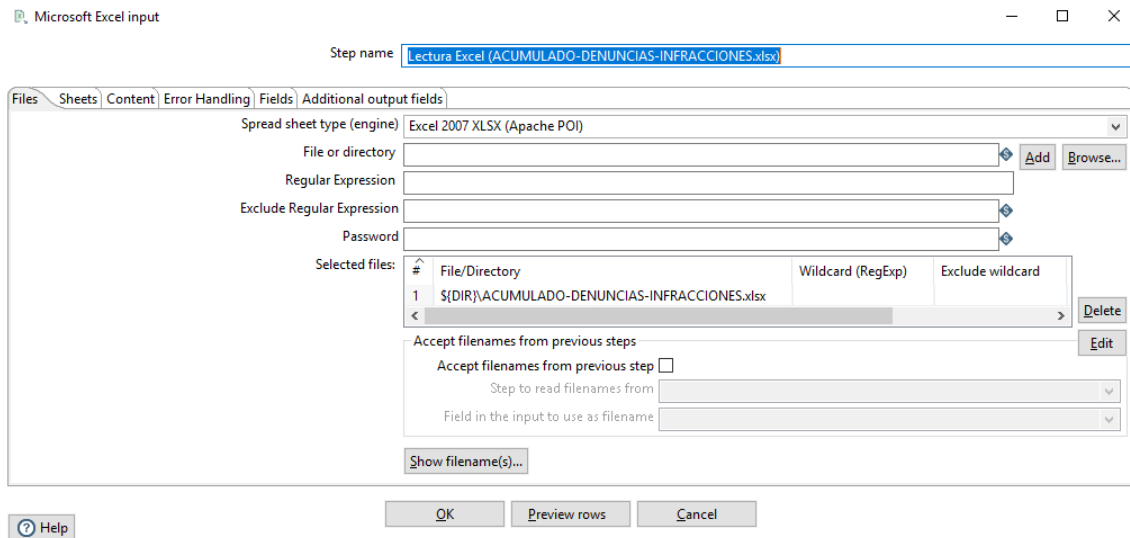


Ilustración 22 - Lectura IN_DENUNCIAS_INFRACCIONES.

Una vez hecho eso, le indicamos qué hoja tiene que leer y desde qué fila y columna, en nuestro caso la hoja “Datos_tratados” y la fila 5 columna 0:

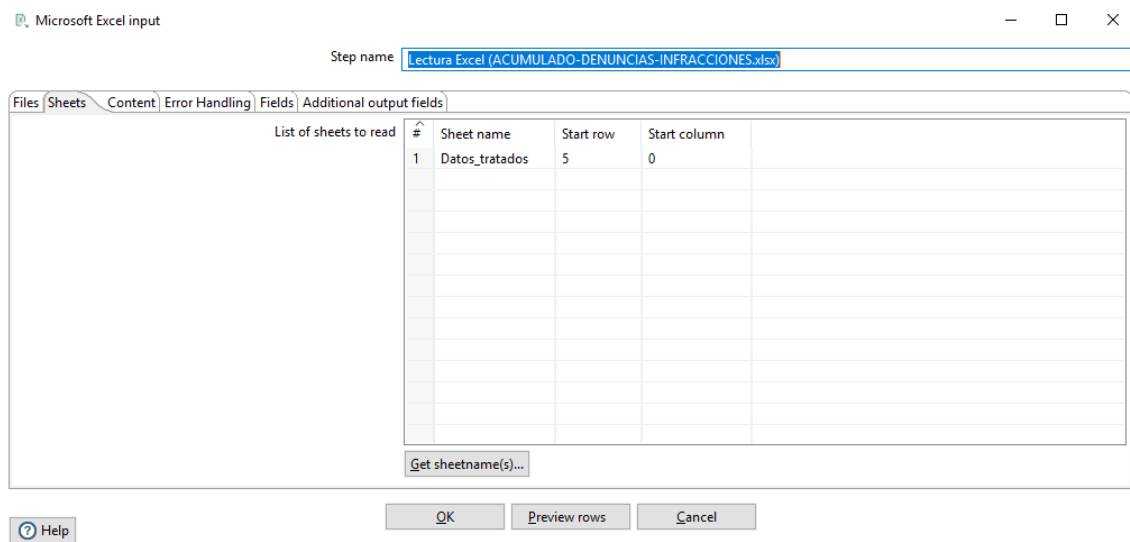


Ilustración 23 - Lectura IN_DENUNCIAS_INFRACCIONES.

Posteriormente obtenemos los campos leídos en la pestaña “Field”:

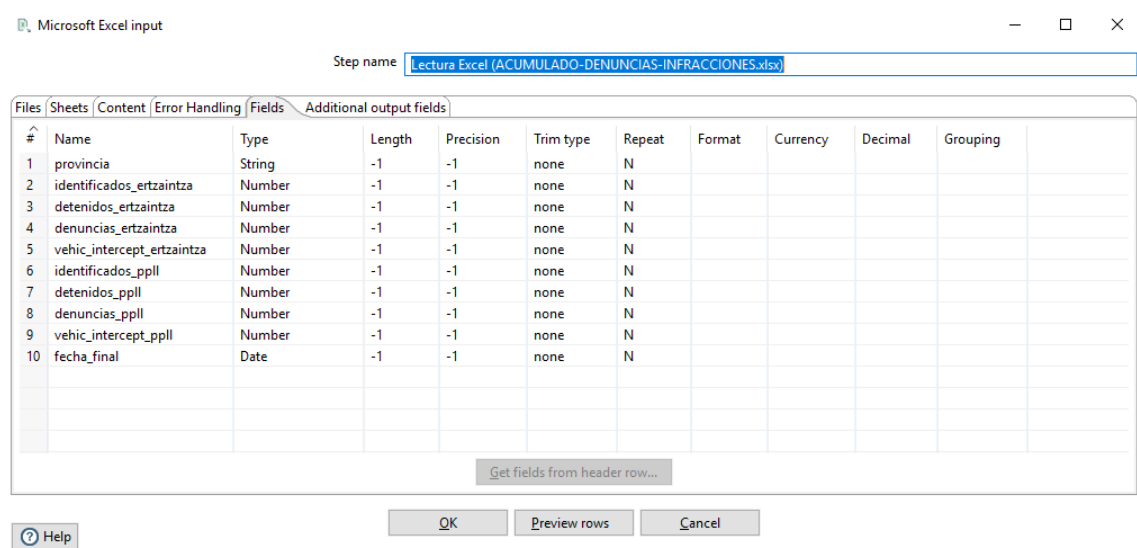


Ilustración 24 - Lectura IN_DENUNCIAS_INFRACCIONES.

Mapeo

Una vez leídos los datos vemos que las provincias están escritas en euskera, por lo que para homogeneizar los datos hemos decidido convertirlas al castellano. Por lo tanto, hacemos la traducción tal y como vemos en la siguiente captura:

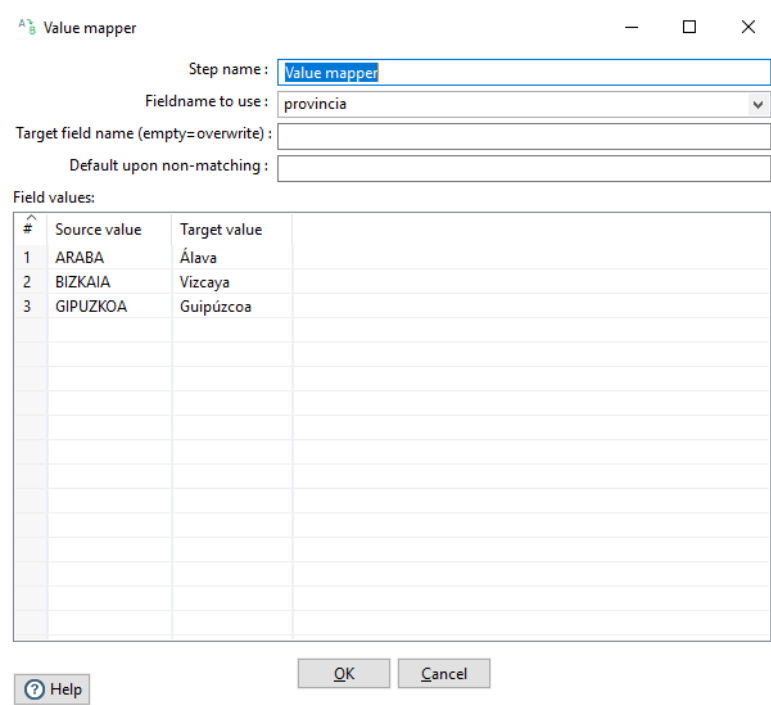


Ilustración 25 - Mapeo Valores IN_DENUNCIAS_INFRACCIONES.

Normalización

Posteriormente hacemos una normalización de los campos que son de tipo “string”, ya que en éstos vamos a convertir los valores a mayúsculas y sin espacios, tal y como vemos en la siguiente ilustración:

String operations

Step name: String operations

The fields to process:

#	In stream field	Out stream field	Trim type	Lower/Upper	Padding	Pad char	Pad Length	InitCap	Escape	Digits	Remove Special character
1	provincia		both	upper	none			N	None	none	none

Help OK Get fields Cancel

Ilustración 26 - Normalización Strings IN_DENUNCIAS_INFRACCIONES.

Ordenación

Posteriormente, ordenamos todos los campos de forma ascendente:

Sort rows

Step name: Sort rows

Sort directory: %java.io.tmpdir% Browse...

TMP-file prefix: out

Sort size (rows in memory): 1000000

Free memory threshold (in %):

Compress TMP Files? ☒

Only pass unique rows? (verifies keys only) ☐

Fields:

#	Fieldname	Ascending	Case sensitive compare?	Sort based on current locale?	Collator Strength	Presorted?
1	provincia	Y	N	N	0	N
2	identificados_ertzaintza	Y	N	N	0	N
3	detenidos_ertzaintza	Y	N	N	0	N
4	denuncias_ertzaintza	Y	N	N	0	N
5	vehic_intercept_ertzaintza	Y	N	N	0	N
6	identificados_ppll	Y	N	N	0	N
7	detenidos_ppll	Y	N	N	0	N
8	denuncias_ppll	Y	N	N	0	N
9	vehic_intercept_ppll	Y	N	N	0	N
10	fecha_final	Y	N	N	0	N

Help OK Cancel Get Fields

Ilustración 27 - Ordenación IN_DENUNCIAS_INFRACCIONES.

Guardado

Finalmente, introducimos todos los valores en la base de datos, es decir, en la tabla intermedia STG_Denuncias_Infracciones, indicamos que haga un truncate de la tabla y con la conexión definida guardamos los valores en la tabla correspondiente:

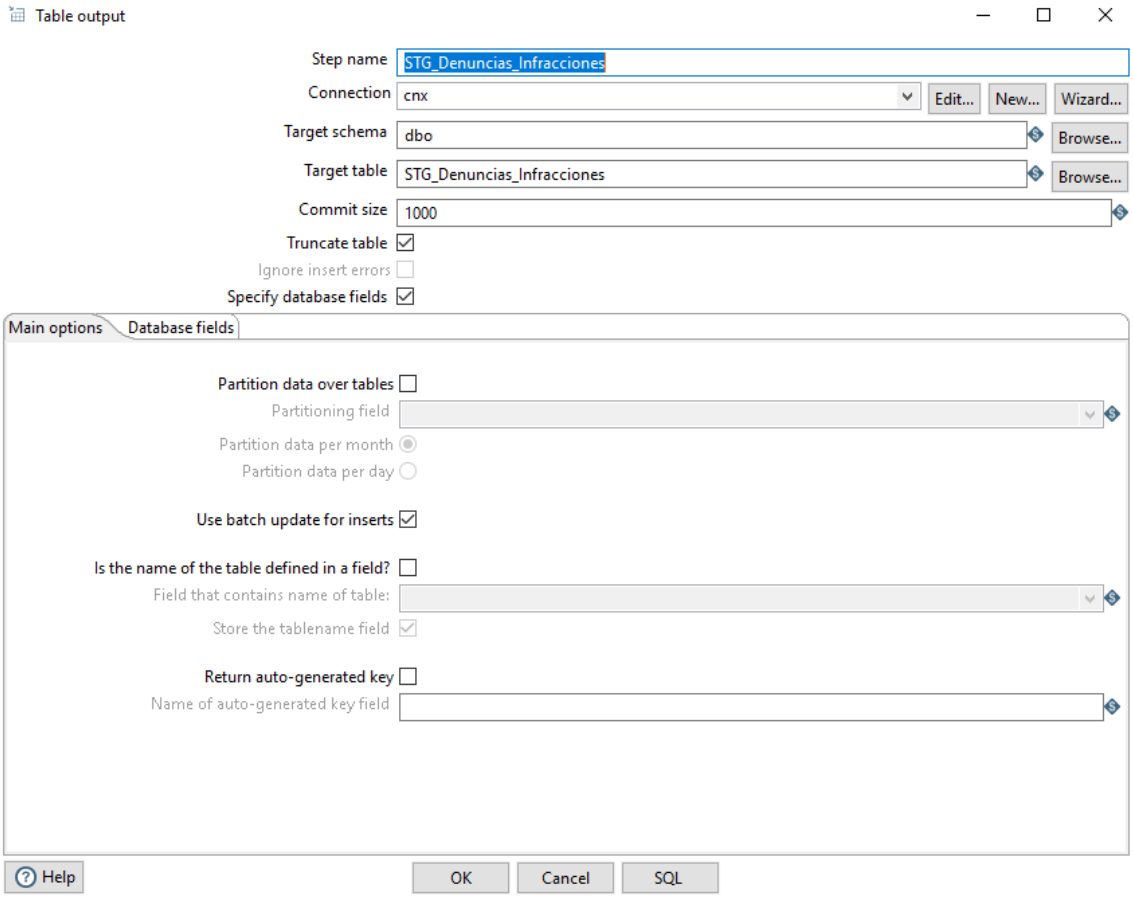


Ilustración 28 - Guardado IN_DENUNCIAS_INFRACCIONES.

Al ejecutar la anterior transformación obtenemos las siguiente métricas:

Execution Results

Stepname	Copynr	Read	Written	Input	Output	Updated	Rejected	Errors	Active	Time	Speed (r/s)	input/output
1 Lectura Excel (ACUMULADO-DENUNCIAS-INFRACCIONES.xlsx)	0	0	219	219	0	0	0	0	Finished	0.2s	894	-
2 Value mapper	0	219	219	0	0	0	0	0	Finished	0.3s	826	-
3 String operations	0	219	219	0	0	0	0	0	Finished	0.3s	782	-
4 Sort rows	0	219	219	0	0	0	0	0	Finished	0.3s	702	-
5 STG_Denuncias_Infracciones	0	219	219	0	219	0	0	0	Finished	0.3s	637	-

Ilustración 29 - Métricas IN_DENUNCIAS_INFRACCIONES.

Observamos que tenemos 219 registros leídos y en nuestra base de datos se han almacena también 219 registros, por lo que la información es correcta.

3.2.4. Transformación IN_POBLACION

La segunda transformación que vamos a realizar se llama “IN_POBLACION”, su objetivo es leer todos los datos del archivo “poblacion_9687bsc.csv” y almacenarlos en la taba intermedia “STG_Poblacion”.

En este caso no hemos hecho ninguna modificación al fichero original, por lo que la transformación nos queda de la siguiente forma:



Ilustración 30 - IN_POBLACION.

Lectura CSV

Lo primero que debemos de hacer es cargar la información que se nos proporciona a partir del fichero CSV correspondiente. Por lo tanto, lo primero escribimos el nombre del paso, indicamos el fichero y el delimitador del CSV, en nuestro caso “;”:

CSV file input

Step name: Lectura CSV (poblacion_9687bsc.csv)

Filename: \${DIR}\poblacion_9687bsc.csv Browse...

Delimiter: ; Insert TAB

Enclosure: "

NIO buffer size: 50000

Lazy conversion? ☒

Header row present? ☒

Add filename to result ☐

The row number field name (optional):

Running in parallel? ☐

New line possible in fields? ☐

Format: mixed

File encoding:

#	Name	Type	Format	Length	Precision	Currency	Decimal	Group	Trim type
1	Edad Simple	String		5		\$,	.	none
2	Provincias	String		25		\$,	.	none
3	Sexo	String		11		\$,	.	none
4	Periodo	String		18		\$,	.	none
5	Total	Integer	#,###,###,##	9	3	\$,	.	none

Ilustración 31 - Lectura IN_POBLACION.

Cabe destacar que hemos tenido que modificar el tipo del campo “Total” ya que lo reconocía como decimal cuando realmente es un entero.

Split

Luego separamos el campo “Provincias”, para así obtener tanto el código como la provincia correspondiente, para ello indicamos que el campo que queremos separar es “Provincias”, y luego en el grid establecemos los nuevos campos:

Split fields

Step name: Split fields

Field to split: Provincias

Delimiter:

Enclosure:

#	New field	ID	Remove ID?	Type	Length	Precision	Format	Group	Decimal	Currency	Nullif	Default	Trim typ
1	provincia_codigo		N	String	2						99		both
2	provincia_nombre		N	String	2						NA		both

Help OK Cancel

Ilustración 32 - Separación Campos IN_POBLACION.

Mapeo

Una vez hecho el paso anterior tenemos que mapear valores, esto se debe a que los nombres de las provincias no son del todo correctos (aparecen en gallego, euskera, catalán y valenciano). Además, al hacer la separación algunos nombres de provincias compuestas han desaparecido, es por ello que necesitamos de este paso para solventar los problemas:

Value mapper

Step name: Value mapper

Fieldname to use: provincia_nombre

Target field name (empty=overwrite):

Default upon non-matching:

Field values:

#	Source value	Target value	
1	Alicante/Alacant	Alicante	
2	Araba/Álava	Álava	
3	Balears,	Baleares	
4	Bizkaia	Vizcaya	
5	Castellón/Castelló	Castellón	
6	Coruña,	La Coruña	
7	Gipuzkoa	Guipúzcoa	
8	Girona	Gerona	
9	Lleida	Lérida	
10	Palmas,	Las Palmas	
11	Rioja,	La Rioja	
12	Santa	Santa Cruz de Tenerife	
13	Valencia/València	Valencia	
14	Ciudad	Ciudad Real	
15	Ourense	Orense	

Help OK Cancel

Ilustración 33 - Mapeo Valores IN_POBLACION.

Normalización

Antes de almacenar los datos en la base de datos, vamos a normalizar los strings para que todos estén en mayúsculas y no tengan espacios ni al principio ni al final:

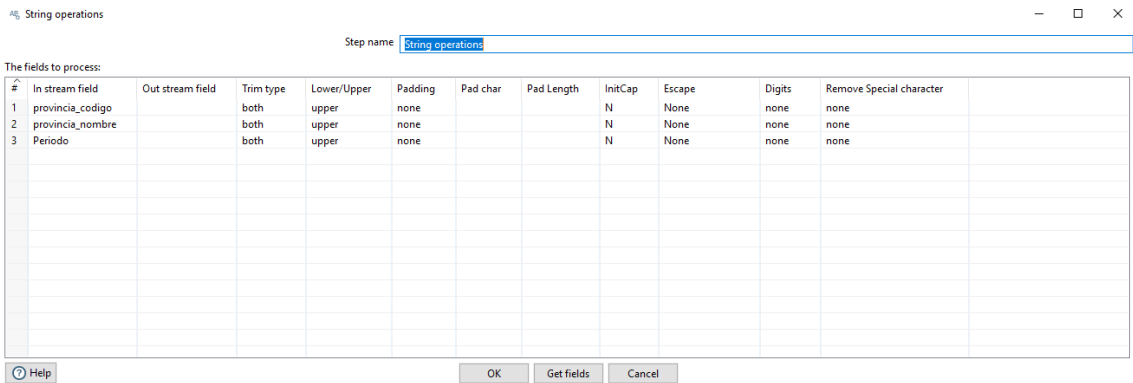


Ilustración 34 - Normalización Strings IN_POBLACION.

Guardado

Finalmente, guardamos los datos en la tabla “STG_Poblacion”, indicando que haga un truncate de la tabla y asociamos los campos:

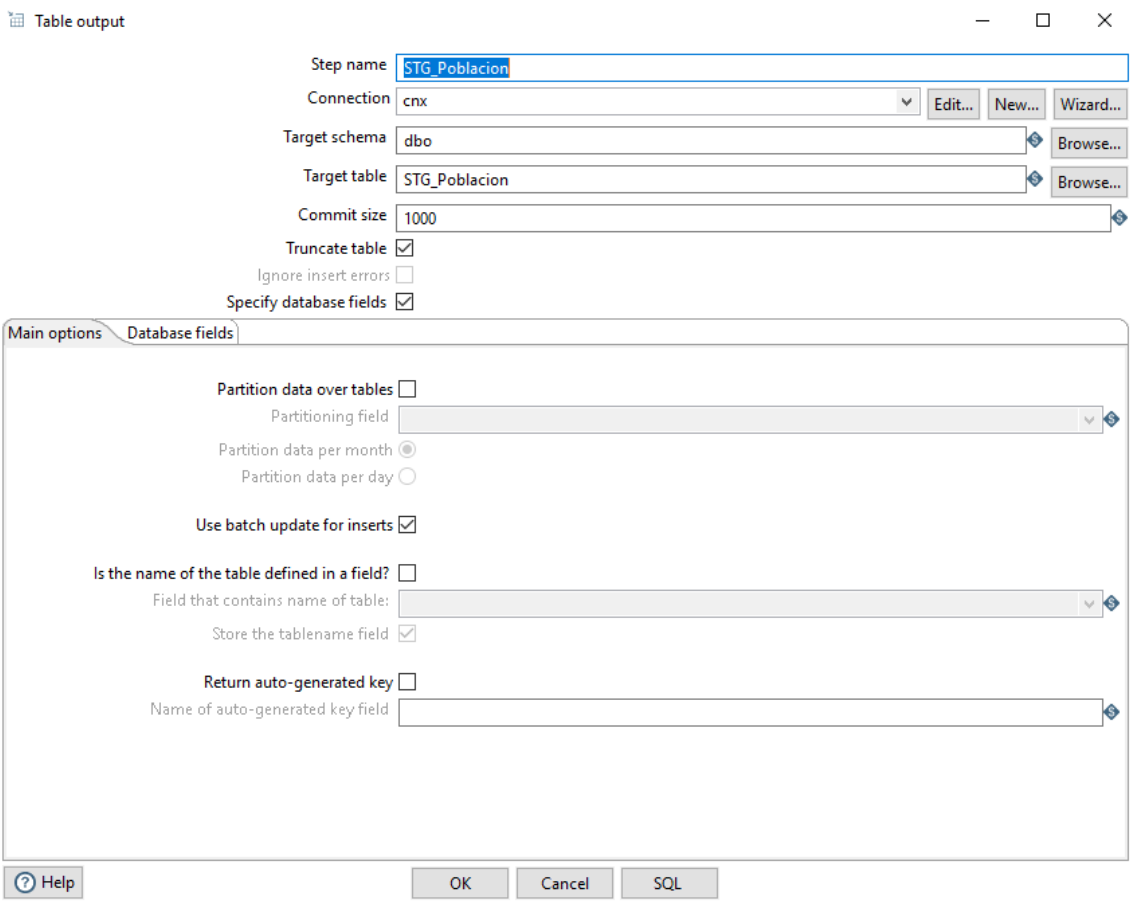


Ilustración 35 - Guardado IN_POBLACION.

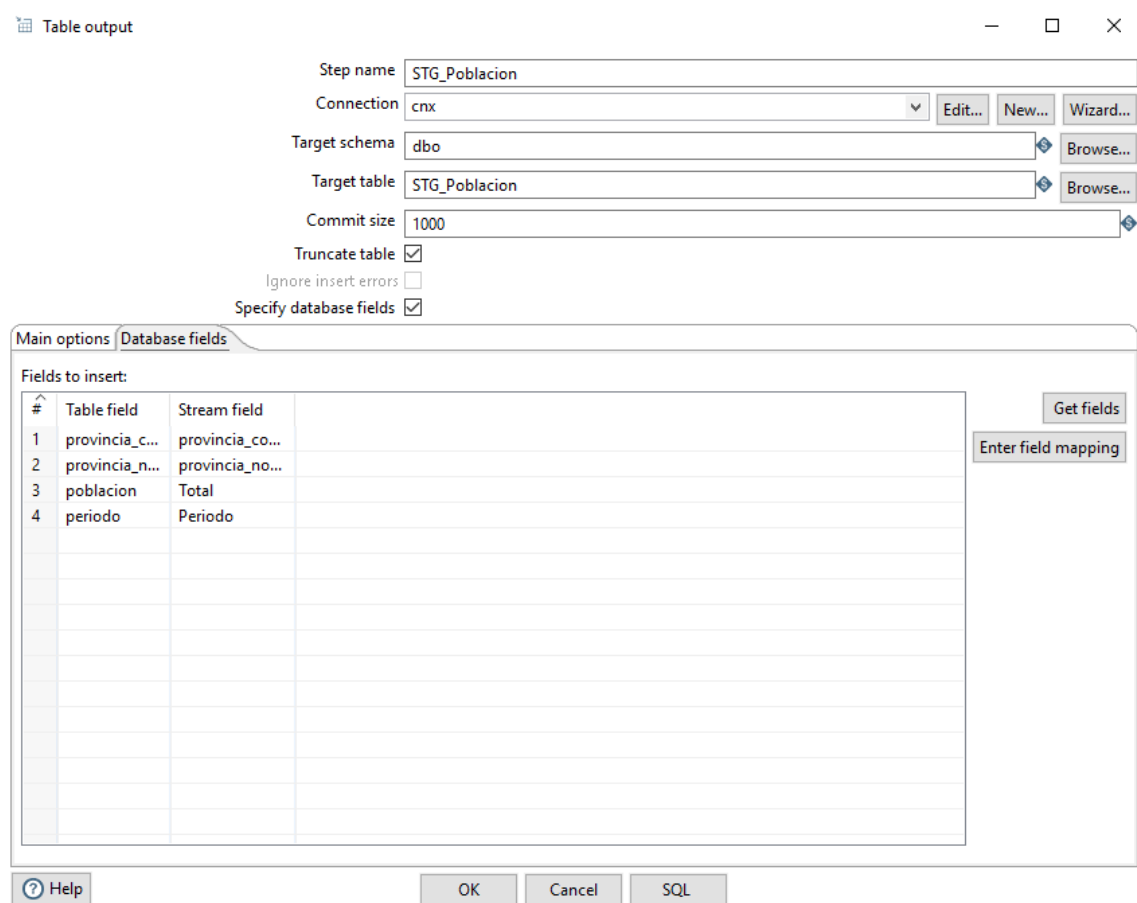


Ilustración 36 - Guardado IN_POBLACION.

Al ejecutar la anterior transformación obtenemos las siguientes métricas:

Execution Results

#	Stepname	Copynr	Read	Written	Input	Output	Updated	Rejected	Errors	Active	Time	Speed (r/s)	input/output
1	Lectura CSV (poblacion_9687bsc.csv)	0	0	52	53	0	0	0	0	Finished	0.0s	1,325	-
2	Split fields	0	52	52	0	0	0	0	0	Finished	0.1s	426	-
3	Value mapper	0	52	52	0	0	0	0	0	Finished	0.1s	406	-
4	String operations	0	52	52	0	0	0	0	0	Finished	0.1s	397	-
5	STG_Poblacion	0	52	52	0	52	0	0	0	Finished	0.2s	292	-

Ilustración 37 - Métricas IN_POBLACION.

Observamos que tenemos 53 registros (52 registros + 1 cabecera) y se han almacenado 52 registros, por lo que la información es correcta.

3.2.5. Transformación IN_MOVILIDAD

La tercera transformación que vamos a realizar se llama “IN_MOVILIDAD”, su objetivo es leer todos los datos del archivo “35167bsc.csv” y guardarlos en la tabla intermedia “STG_Movilidad”.

En este caso no hemos hecho ninguna modificación en el fichero CSV original, por lo que la transformación nos queda de la siguiente manera:

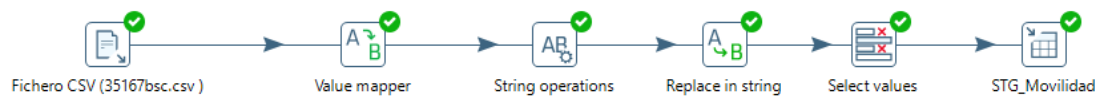


Ilustración 38 - IN_MOVILIDAD.

Ahora vamos a explicar paso a paso lo que hemos hecho:

Lectura CSV

Lo primero que tenemos que hacer es leer el fichero CSV que se nos proporciona, luego escribimos el nombre del paso, indicamos el fichero y el delimitador del CSV, en nuestro caso “,”:

CSV file input

Step name: Fichero CSV (35167bsc.csv)

Filename: \${DIR}\35167bsc.csv

Delimiter: ;

Enclosure: "

NIO buffer size: 50000

Lazy conversion? ☒

Header row present? ☒

Add filename to result ☐

The row number field name (optional):

Running in parallel? ☐

New line possible in fields? ☐

Format: mixed

File encoding:

#	Name	Type	Format	Length	Precision	Currency	Decimal	Group	Trim type
1	Zonas de movilidad	String		27		\$,	.	both
2	Periodo	Timestamp	dd/MM/yyyy			\$,	.	both
3	Total	String	#,##	15	0	\$,	.	both

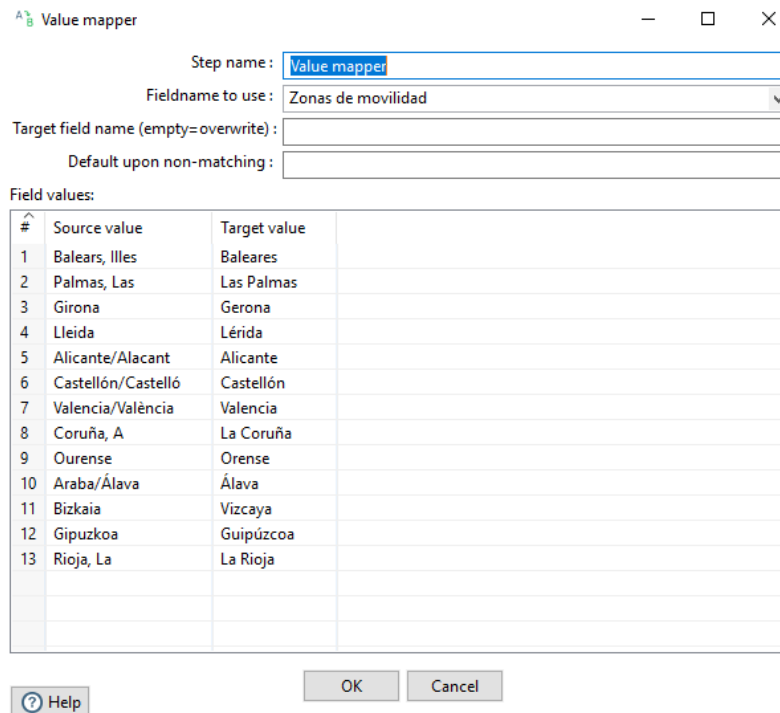
Buttons: Help, OK, Get Fields, Preview, Cancel

Ilustración 39 - Lectura IN_MOVILIDAD.

Cabe destacar que el atributo “Total” hemos indicado que sea de tipo string, ya que si considerábamos que fuera numérico a la hora de introducirlo en la base de datos no guardaba los decimales.

Mapeo

Una vez leídos todos los datos, tenemos que realizar un mapeo del campo “Zonas de movilidad”, esto se debe a que los nombres de las provincias vienen en (euskera, gallego, catalán, valenciano y balear), sin embargo para homogeneizar todas las provincias las traducimos al castellano.



Value mapper

Step name: Value mapper

Fieldname to use: Zonas de movilidad

Target field name (empty=overwrite):

Default upon non-matching:

Field values:

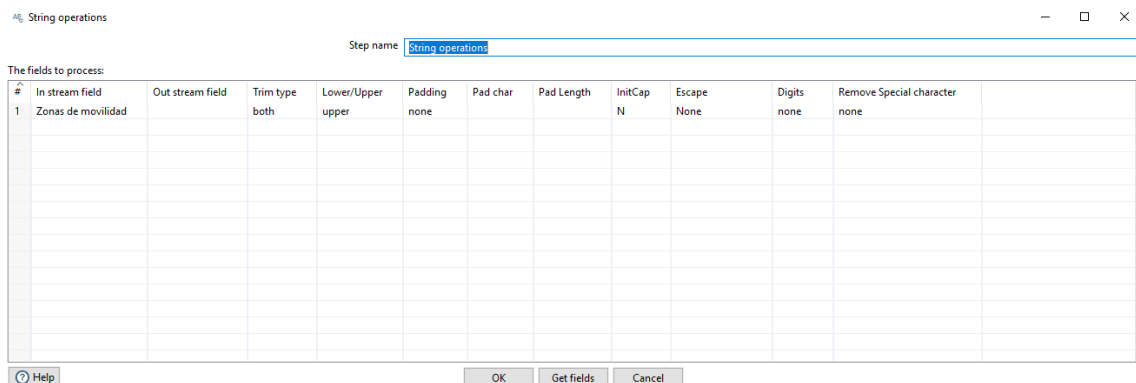
#	Source value	Target value
1	Balears, Illes	Baleares
2	Palmas, Las	Las Palmas
3	Girona	Gerona
4	Lleida	Lérida
5	Alicante/Alacant	Alicante
6	Castellón/Castelló	Castellón
7	Valencia/València	Valencia
8	Coruña, A	La Coruña
9	Ourense	Orense
10	Araba/Álava	Álava
11	Bizkaia	Vizcaya
12	Gipuzkoa	Guipúzcoa
13	Rioja, La	La Rioja

Help OK Cancel

Ilustración 40 - Mapeo Valores IN_MOVILIDAD.

Normalización

Una vez que tenemos ya los datos de forma correcta, normalizamos las provincias para que no haya un espacio al principio o al final de la cadena, y establecemos que todas las cadenas estén en mayúsculas:



String operations

Step name: String operations

The fields to process:

#	In stream field	Out stream field	Trim type	Lower/Upper	Padding	Pad char	Pad Length	InitCap	Escape	Digits	Remove Special character
1	Zonas de movilidad		both	upper	none			N	None	none	none

Help OK Get fields Cancel

Ilustración 41 - Normalización IN_MOVILIDAD.

Replace

Posteriormente reemplazamos en el string de “Total” la coma por el punto, ya que de esta manera cuando luego introduzcamos el valor en la base de datos sí que no va a aparecer con decimales:

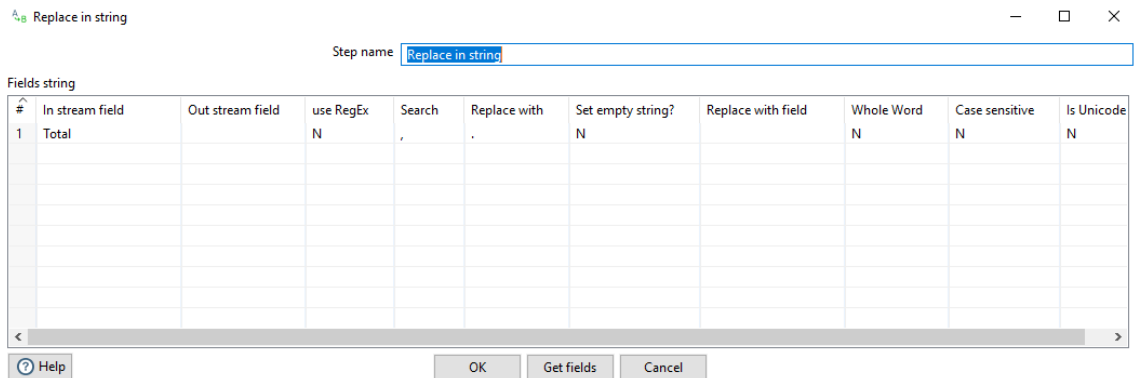


Ilustración 42 - Replace IN_MOVILIDAD.

Select Values

Cuando ya tenemos el string modificado, hay que convertirlo de decimal, ya que en la tabla de la base de datos el “Total” lo almacenamos como un número. Para ello, hacemos uso del componente “Select_Values” y en “Meta-data” establecemos que cree un nuevo campo que sea de tipo numérico con el campo original “Total”, el resultado de esta operación lo guardamos en “totalSQL”:

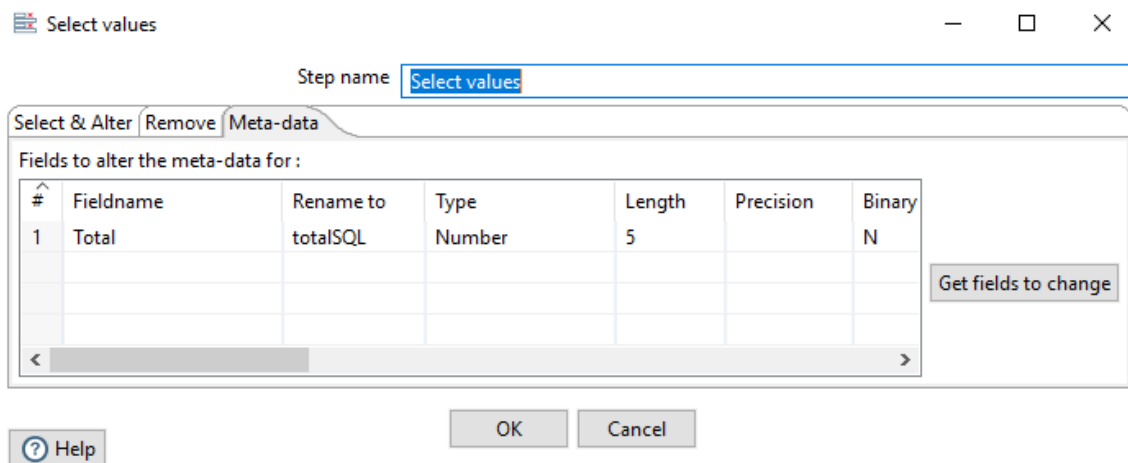


Ilustración 43 - Select Values IN_MOVILIDAD.

Guardado

Finalmente, guardamos todo el proceso realizado en la tabla “STG_Movilidad”, indicando que hay un truncate de la tabla y asociamos los campos obtenidos con los de la tabla:

Table output

Step name

STG_Movilidad

Connection

cnx

Edit...

New...

Wizard...

Target schema

dbo

Browse...

Target table

STG_Movilidad

Browse...

Commit size

1000

Truncate table

☒

Ignore insert errors

☐

Specify database fields

☒

Main options

Database fields

Partition data over tables

☐

Partitioning field

Partition data per month

☒

Partition data per day

☐

Use batch update for inserts

☒

Is the name of the table defined in a field?

☐

Field that contains name of table:

Store the tablename field

☒

Return auto-generated key

☐

Name of auto-generated key field

Help

OK

Cancel

SQL

Ilustración 44 - Guardado IN_MOVILIDAD.

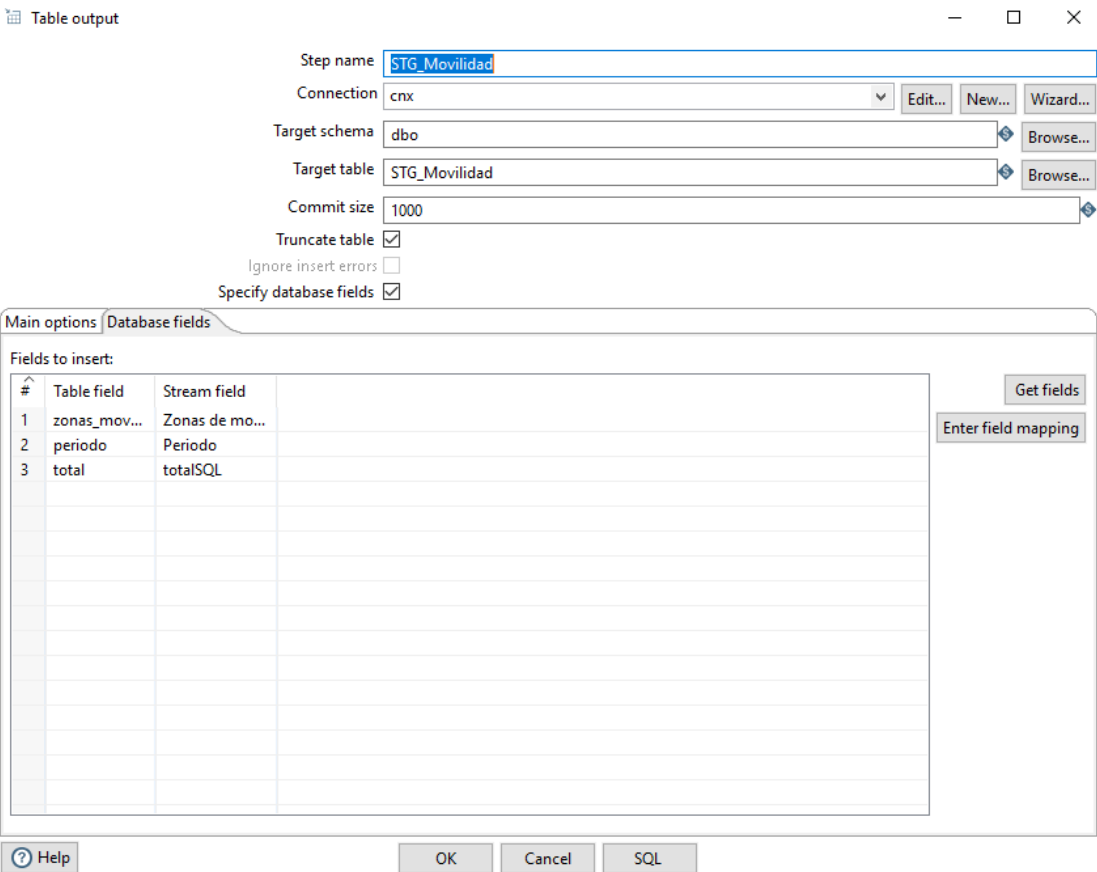


Ilustración 45 - Guardado IN_MOVILIDAD.

Para terminar con esta transformación obtenemos las métricas de su ejecución:

Execution Results

#	Stepname	Copynr	Read	Written	Input	Output	Updated	Rejected	Errors	Active	Time	Speed (r/s)	input/output
1	Fichero CSV (35167bsc.csv)	0	0	4732	4733	0	0	0	0	Finished	0.0s	189,320	-
2	Value mapper	0	4732	4732	0	0	0	0	0	Finished	0.1s	58,420	-
3	String operations	0	4732	4732	0	0	0	0	0	Finished	0.1s	48,286	-
4	Replace in string	0	4732	4732	0	0	0	0	0	Finished	0.1s	43,413	-
5	Select values	0	4732	4732	0	0	0	0	0	Finished	0.1s	33,560	-
6	STG_Movilidad	0	4732	4732	0	4732	0	0	0	Finished	0.3s	15,933	-

Ilustración 46 - Métricas IN_MOVILIDAD.

Como podemos observar, leemos 4733 registros (4732 observaciones + 1 cabecera) y almacenamos 4732, por lo que la información es correcta.

3.2.6. Transformación IN_AGLOMERACION

La cuarta transformación que vamos a realizar se llama “IN_AGLOMERACION”, su objetivo es leer todo los datos del Excel “statistic_id1104235_covid-19_-poblacion-que-evitaba-las-aglomeraciones-segun-edad-en-espana-2020.xlsx” y guardarlos en la tabla intermedia “STG_AGLOMERACION”.

En este caso sí hemos hecho una modificación en el fichero Excel, no sabemos por qué motivo determinadas provincias tenían un espacio o carácter especial que no era visible y al hacer la lectura, independientemente de si hacíamos un “trim” o usábamos un “string operations” no eliminaba ese “espacio/carácter especial”, es por ello que hemos eliminado de forma manual dicho espacio en el campo “provincia” de los registros afectados.

La transformación nos queda de la siguiente manera:

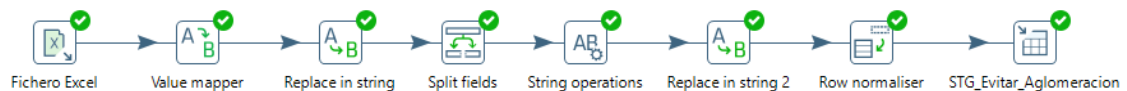


Ilustración 47 - IN_AGLOMERACION.

Lectura

Lo primero que tenemos que hacer es leer el fichero Excel que se nos ha proporcionado, para ello escribimos el nombre del paso, indicamos el fichero y su formato correspondiente a XLSX:

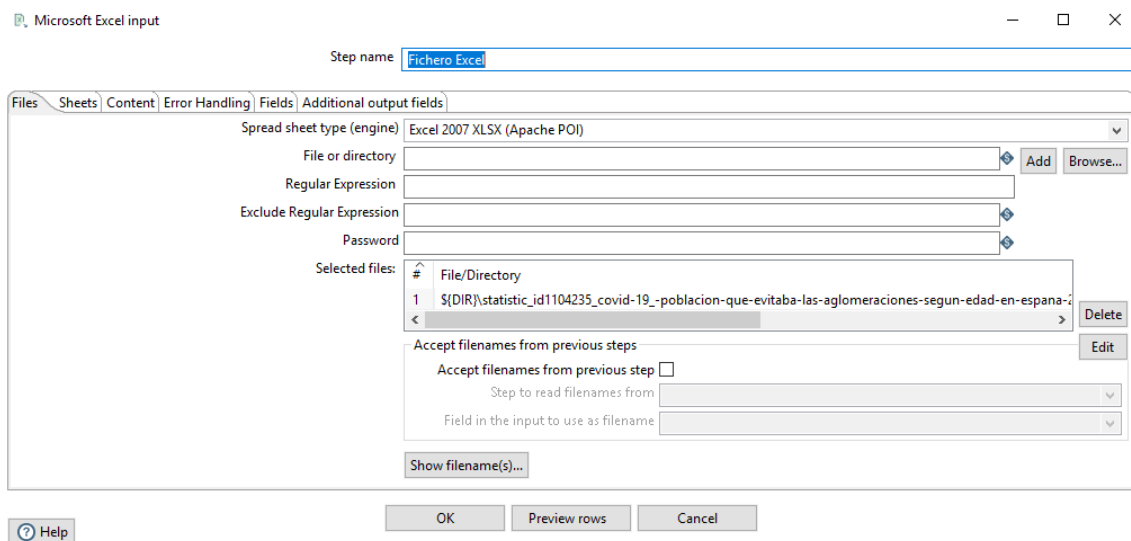


Ilustración 48 - Lectura IN_AGLOMERACION.

Una vez hecho eso, le indicamos qué hoja tiene que leer y desde qué fila y columna, en nuestro caso la hoja “Datos_provincias” y la fila 5 columna 2:

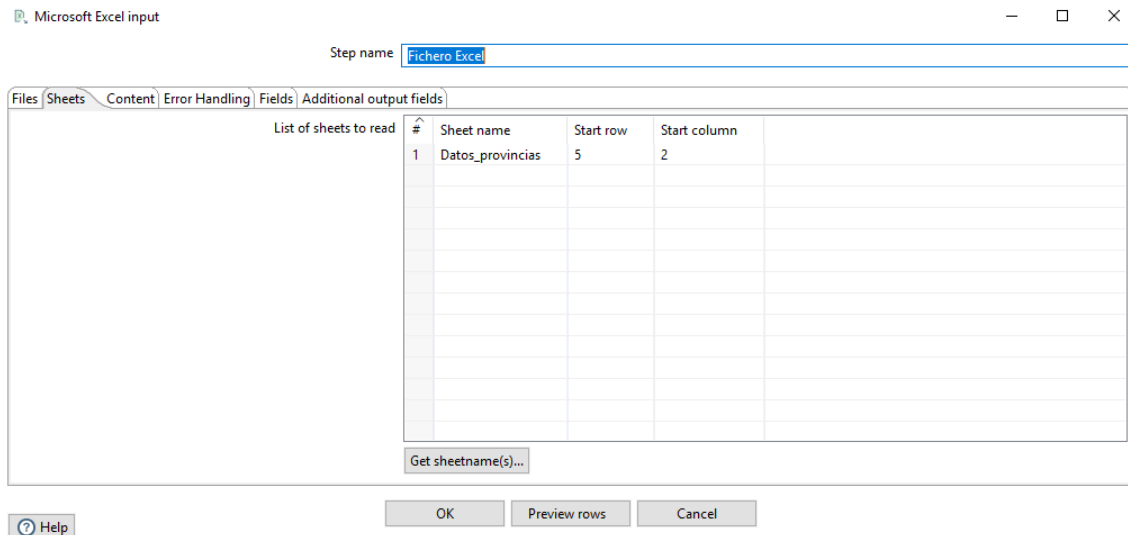


Ilustración 49 - Lectura IN_AGLOMERACION.

Posteriormente obtenemos los campos leídos en la pestaña “Fields”, los nombres de los campos han sido definidos de forma manual:

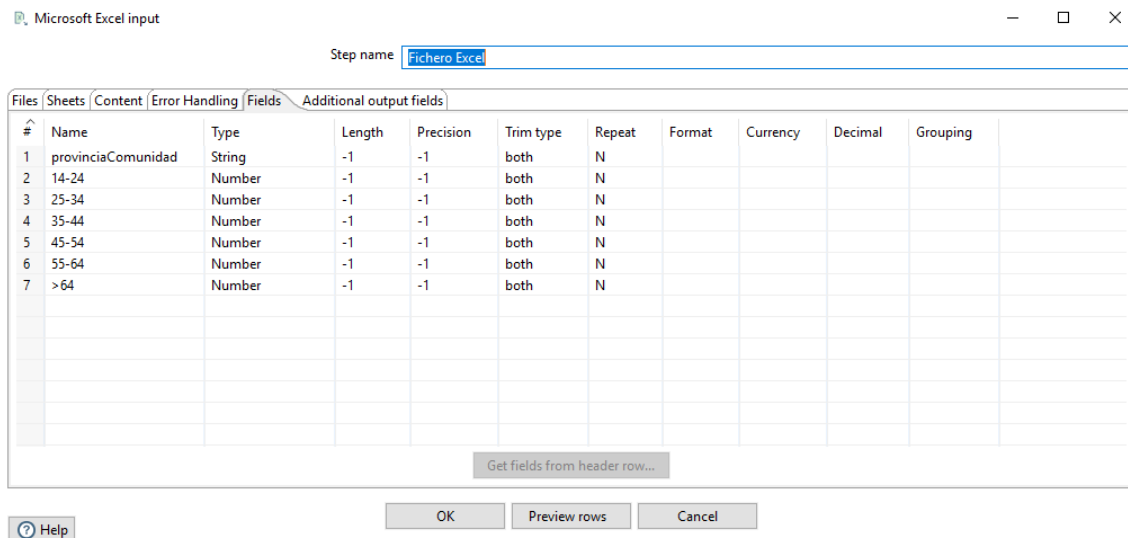


Ilustración 50 - Lectura IN_AGLOMERACIONES.

Mapeo

Al igual que ha sucedido con transformaciones anteriores, muchas provincias vienen también con su nombre en catalán/gallego/euskera/valenciano... Es por ello que hemos decido mantener el nombre en castellano, mapeando así los valores de las provincias que venían en otro idioma:

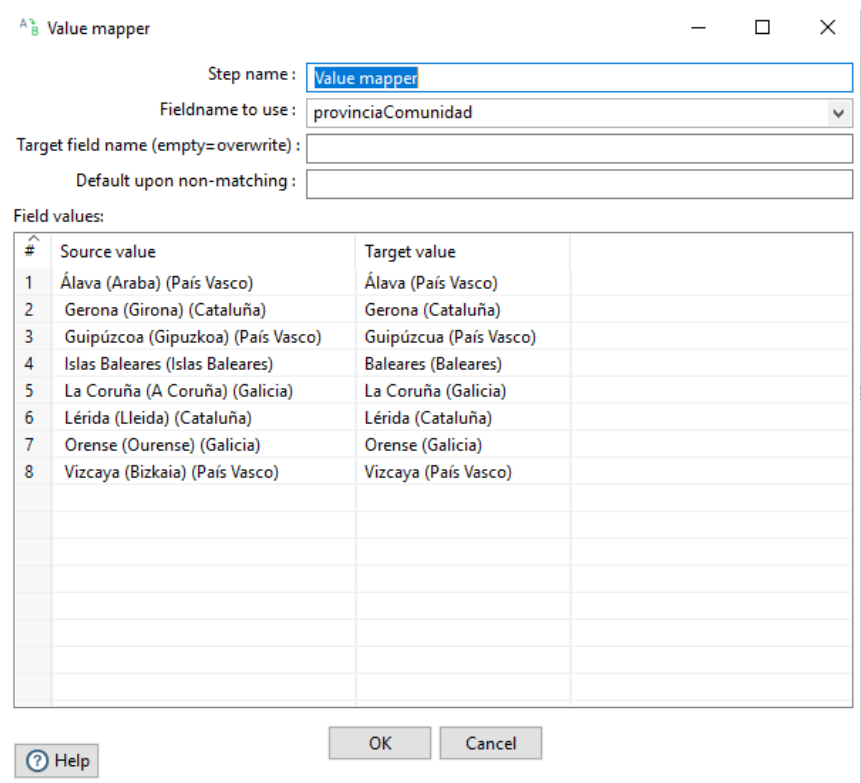


Ilustración 51 - Mapeo Valores IN_AGLOMERACION.

Replace

Posteriormente tenemos que hacer un replace del símbolo “)” por nada, de esta forma luego podemos dividir el campo en dos, para así obtener el nombre de la provincia y su comunidad autónoma:

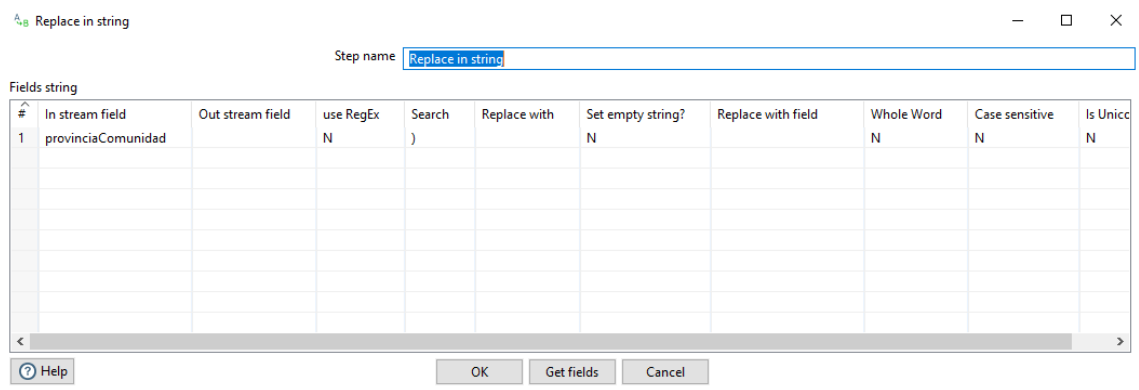


Ilustración 52 - Replace IN_AGLOMERACION.

Split

Una vez que ya tenemos la información que queremos, la podemos separar estableciendo como separado “[espacio](“, de esta forma creamos dos nuevos campos: uno para la provincia y otro para la comunidad.

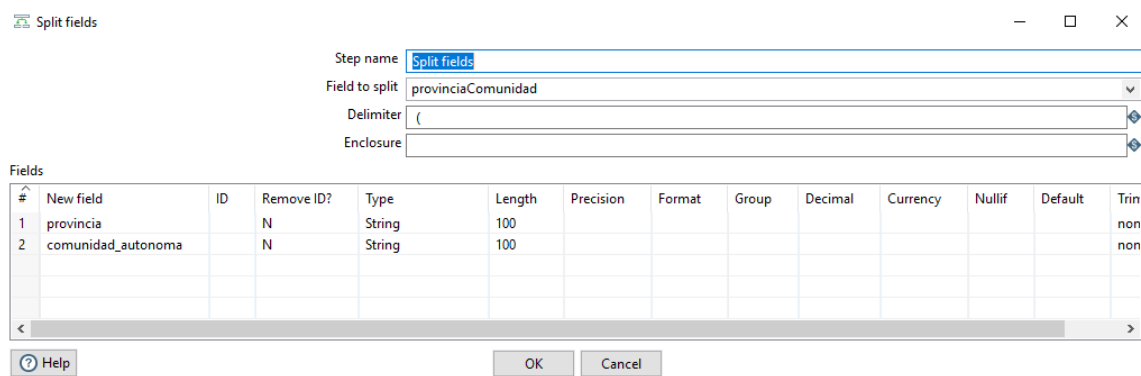


Ilustración 53 - Split IN_AGLOMERACION.

Normalización

Cuando ya tenemos la información separada, podemos hacer uso de un “string operations” para normalizar todos los strings, es decir, establecer mayúsculas y eliminar espacios:

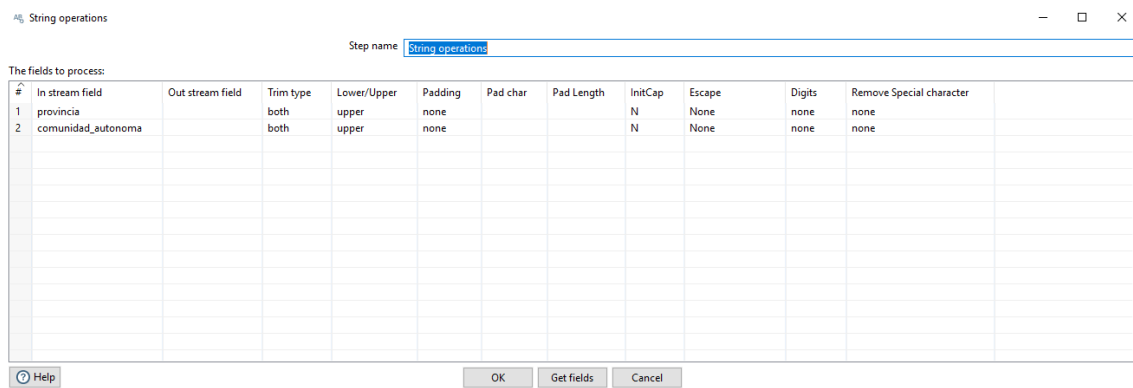


Ilustración 54 - Normalización Strings IN_AGLOMERACION.

Replace

Analizando los datos hemos visto que todas las comunidades y provincias cumplían las reglas ortográficas, sin embargo, la comunidad Aragón la escribían sin tilde. Es por ello que para mantener la misma lógica en todas las transformaciones hemos corregido dicho problema:

Replace in string

Step name: Replace in string 2

#	In stream field	Out stream field	use RegEx	Search	Replace with	Set empty string?	Replace with field	Whole Word	Case sensitive	Is
1	comunidad_autonoma		N	ARAGON	ARAGÓN	N		N	N	N

Buttons: Help, OK, Get fields, Cancel

Ilustración 55 - Replace IN_AGLOMERACION.

Normalización filas

Posteriormente, hemos tenido que normalizar filas para que las columnas respectivas al grupo de edad fueran filas y no columnas. Para ello establecemos el nuevo campo que vamos a crear y los valores que va a tener dicho campo:

Row normaliser

Step name: Row normaliser

Type field: grupo_edad

#	Fieldname	Type	new field
1	14-24	14-24	porc_poblacion
2	25-34	25-34	porc_poblacion
3	35-44	35-44	porc_poblacion
4	45-54	45-54	porc_poblacion
5	55-64	55-64	porc_poblacion
6	>64	>64	porc_poblacion

Buttons: Help, OK, Cancel, Get Fields

Ilustración 56 - Normalización Filas IN_AGLOMERACION.

Guardado

Finalmente, una vez que tenemos ya todos los datos normalizados podemos proceder al guardado de los mismo en la tabla intermedia "STG_AGLOMERACION". Tenemos que marcar el truncate table y asociar los campos:

Table output

Step name

STG_Evitar_Aglomeracion

Connection

cnx

Edit...

New...

Wizard...

Target schema

dbo

Browse...

Target table

STG_Evitar_Aglomeracion

Browse...

Commit size

1000

Truncate table

☒

Ignore insert errors

☐

Specify database fields

☒

Main options

Database fields

Partition data over tables

☐

Partitioning field

Partition data per month

☒

Partition data per day

☐

Use batch update for inserts

☒

Is the name of the table defined in a field?

☐

Field that contains name of table:

Store the tablename field

☒

Return auto-generated key

☐

Name of auto-generated key field

Help

OK

Cancel

SQL

Ilustración 57 - Guardado IN_AGLOMERACIONES.

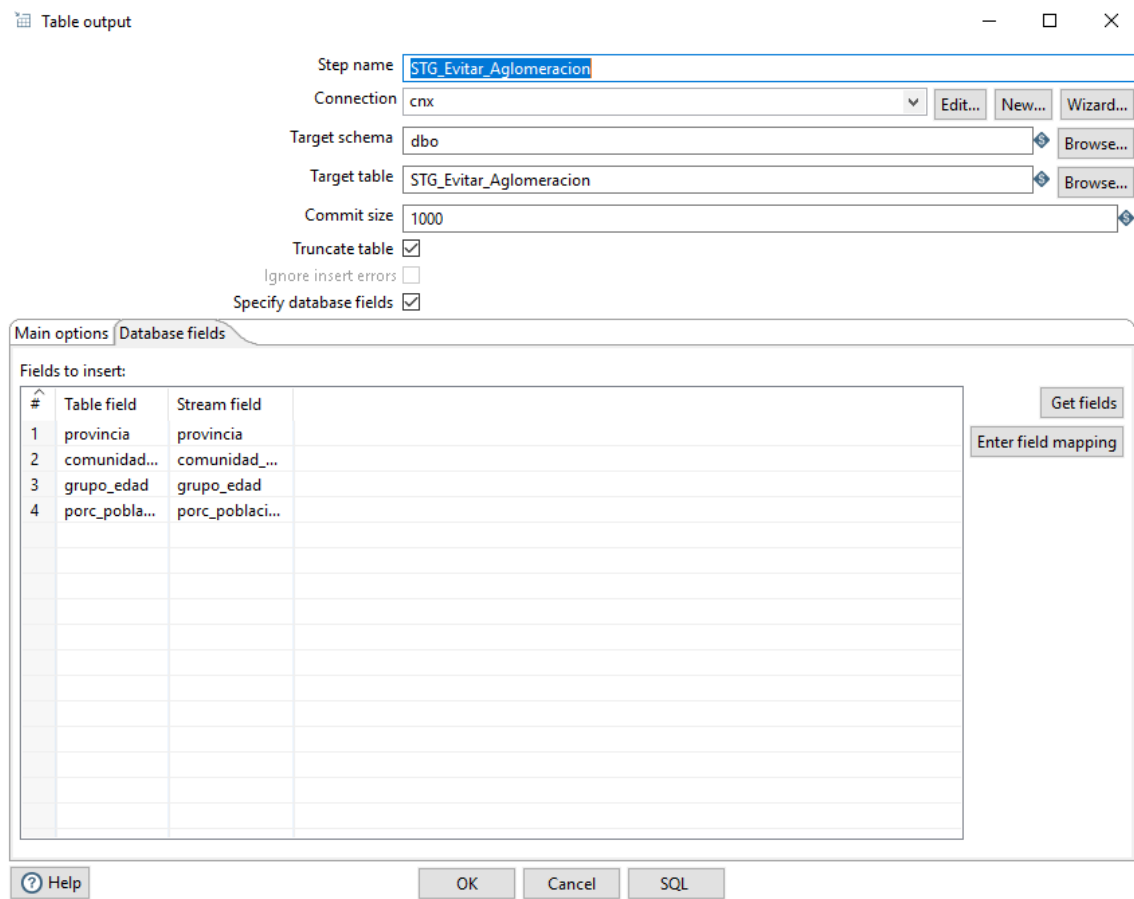


Ilustración 58 - Guardado IN_AGLOMERACIONES.

Al ejecutar la anterior transformación obtenemos las siguientes métricas:

Execution Results

#	Stepname	Copynr	Read	Written	Input	Output	Updated	Rejected	Errors	Active	Time	Speed (r/s)	input/output
1	Fichero Excel	0	0	50	50	0	0	0	0	Finished	0.5s	102	-
2	Value mapper	0	50	50	0	0	0	0	0	Finished	0.5s	101	-
3	Replace in string	0	50	50	0	0	0	0	0	Finished	0.5s	100	-
4	Split fields	0	50	50	0	0	0	0	0	Finished	0.5s	99	-
5	String operations	0	50	50	0	0	0	0	0	Finished	0.5s	98	-
6	Replace in string 2	0	50	50	0	0	0	0	0	Finished	0.5s	96	-
7	Row normaliser	0	50	300	0	0	0	0	0	Finished	0.5s	574	-
8	STG_Evitar_Aglomeracion	0	300	300	0	300	0	0	0	Finished	0.6s	540	-

Ilustración 59 - Métricas IN_AGLOMERACION.

Como podemos observar leemos 50 registros y almacenamos 300, esto se debe a la normalización de las filas para el atributo “grupo_edad”.

3.2.7. Transformación IN_LLAMADAS112

La quinta transformación respecto al bloque IN es “IN_LLAMADAS112”, ésta se encarga de hacer la lectura del archivo “rows.xml” el cual contiene todas las llamadas, y las vamos a guardar en la tabla intermedia “STG_Llamadas112”.

En este caso no hemos hecho ninguna modificación en el fichero XML original, por lo que la transformación nos queda de la siguiente forma:

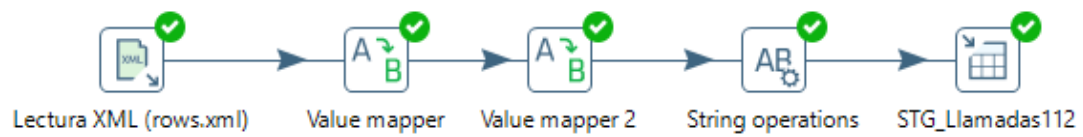


Ilustración 60 - IN_LLAMADAS112.

Lectura XML

Lo primero que tenemos que hacer es leer la información que se nos proporciona en el fichero XML. Para ello escribimos el nombre del paso e indicamos el fichero:

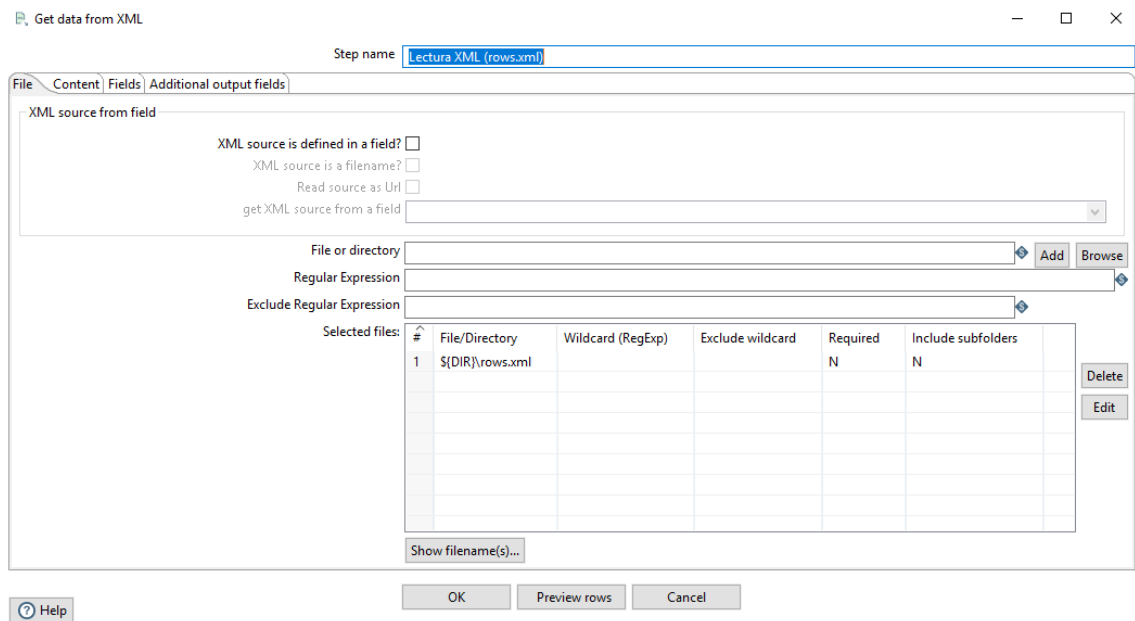


Ilustración 61 - Lectura IN_LLAMADAS112.

Luego nos dirigimos a la pestaña "Content" y definimos desde qué loop tiene que empezar a leer nuestro fichero XML:

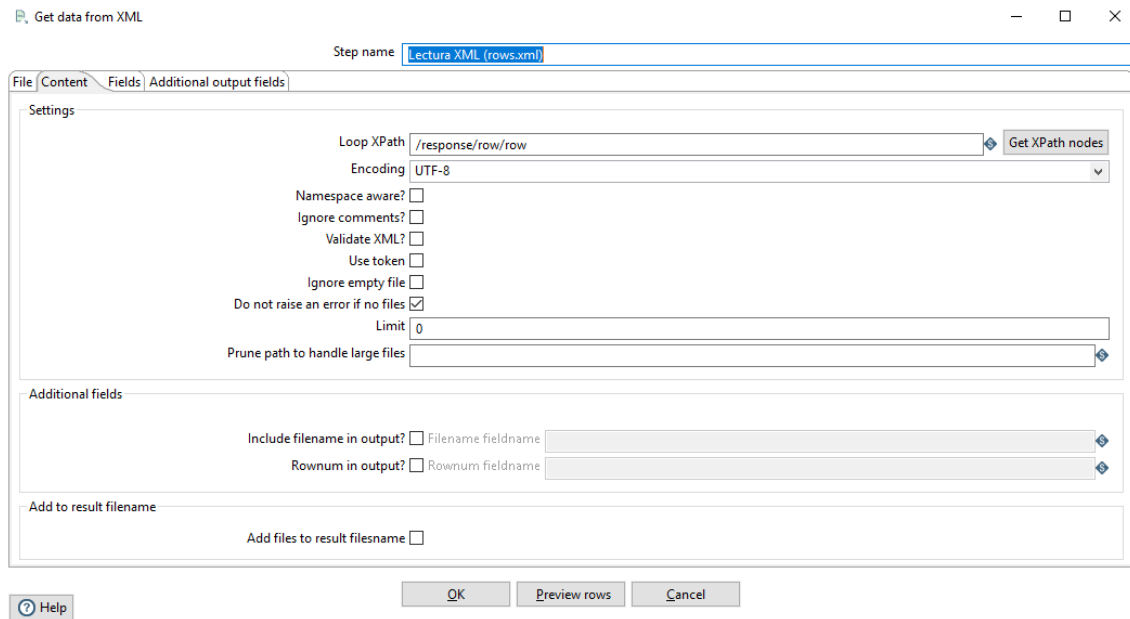


Ilustración 62 - Lectura IN_LLAMADAS112.

Finalmente, obtenemos los campos y los definimos nosotros de forma manual:

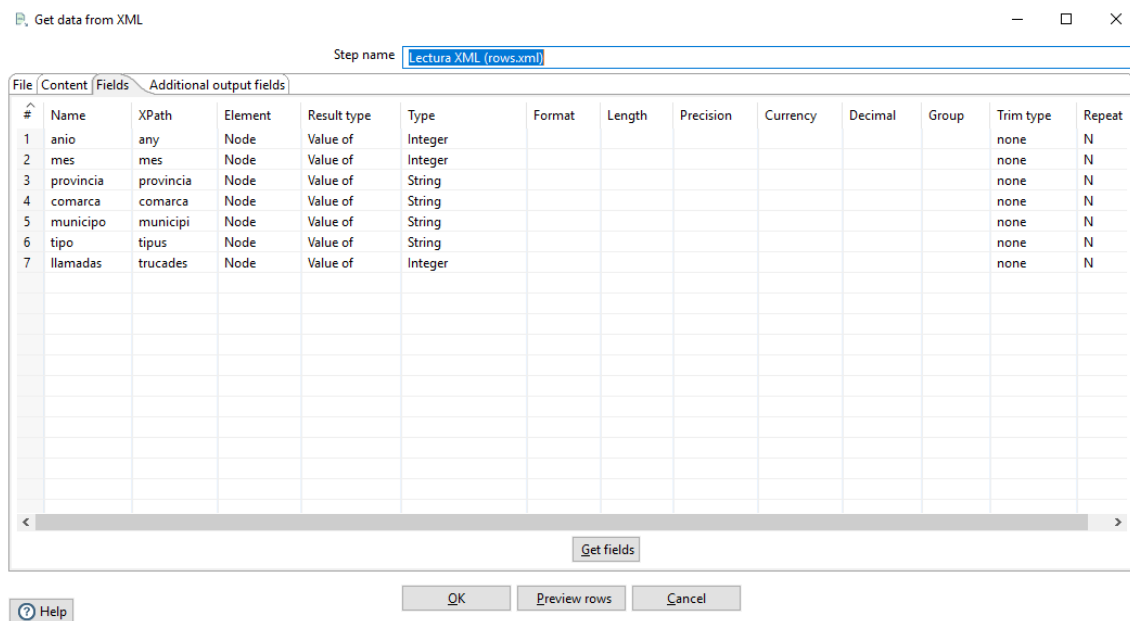


Ilustración 63 - Lectura IN_LLAMADAS112.

Mapeo

Una vez leído los registros tenemos que hacer un cambio de valor de algunas provincias, ya que éstas aparecen en catalán y vamos a mantener en la base de datos solamente la traducción al castellano:

[illegible]

Ilustración 64 - Mapeo Valores IN_LLAMADAS112.

Mapeo

Necesitamos hacer otro mapeo de valores para el campo “tipo” ya que éste representa el motivo de la llamada y sus valores están en catalán. Por lo tanto, realizamos lo mismo que en el paso anterior:

Value mapper

Step name :

Value mapper 2

Fieldname to use :

tipo

Target field name (empty=overwrite) :

Default upon non-matching :

Field values:

#	Source value	Target value
1	Medi ambient	MEDIO AMBIENTE
2	Assistència sanitària	ASISTENCIA SANITARIA
3	Seguretat	SEGURIDAD
4	Civisme	CIVISMO
5	Altres incidències	OTRAS INCIDENCIAS
6	Incendi	INCENDIO
7	Fuita (aigua, gas, altres)	FUGA
8	Trànsit	TRÁFICO
9	Accident	ACCIDENTE
10	Meteorologia	METEOROLOGÍA

Help

OK

Cancel

Ilustración 65 - Mapeo Valores IN_LLAMADAS112.

Normalización

Antes de introducir todos los datos a la base de datos, tenemos que normalizar las cadenas de valores, es decir, establecer los campos string a mayúscula y sin espacios al comienzo ni al final:

String operations

Step name :

String operations

The fields to process:

#	In stream field	Out stream field	Trim type	Lower/Upper	Padding	Pad char	Pad Length	InitCap	Escape	Digits	Remove Special character
1	provincia		both	upper	none			N	None	none	none
2	comarca		both	upper	none			N	None	none	none
3	municipio		both	upper	none			N	None	none	none
4	tipo		both	upper	none			N	None	none	none

Help

OK

Get fields

Cancel

Ilustración 66 - Normalización IN_LLAMADAS112.

Guardado

Finalmente, guardamos los datos en la tabla “STG_Llamadas112”, indicando que haga un truncate de la tabla y asociamos los campos:

The screenshot shows the 'Table output' dialog box with the following configuration:

- Step name: STG_Llamadas112
- Connection: cnx
- Target schema: dbo
- Target table: STG_Llamadas112
- Commit size: 1000
- Truncate table: ☒
- Ignore insert errors: ☐
- Specify database fields: ☒

The 'Main options' tab is selected, showing the following options:

- Partition data over tables: ☐
- Partitioning field: [empty]
- Partition data per month: ☒
- Partition data per day: ☐
- Use batch update for inserts: ☒
- Is the name of the table defined in a field? ☐
- Field that contains name of table: [empty]
- Store the tablename field: ☒
- Return auto-generated key: ☐
- Name of auto-generated key field: [empty]

Buttons at the bottom: Help, OK, Cancel, SQL.

Ilustración 67 - Guardado IN_LLAMADAS112.

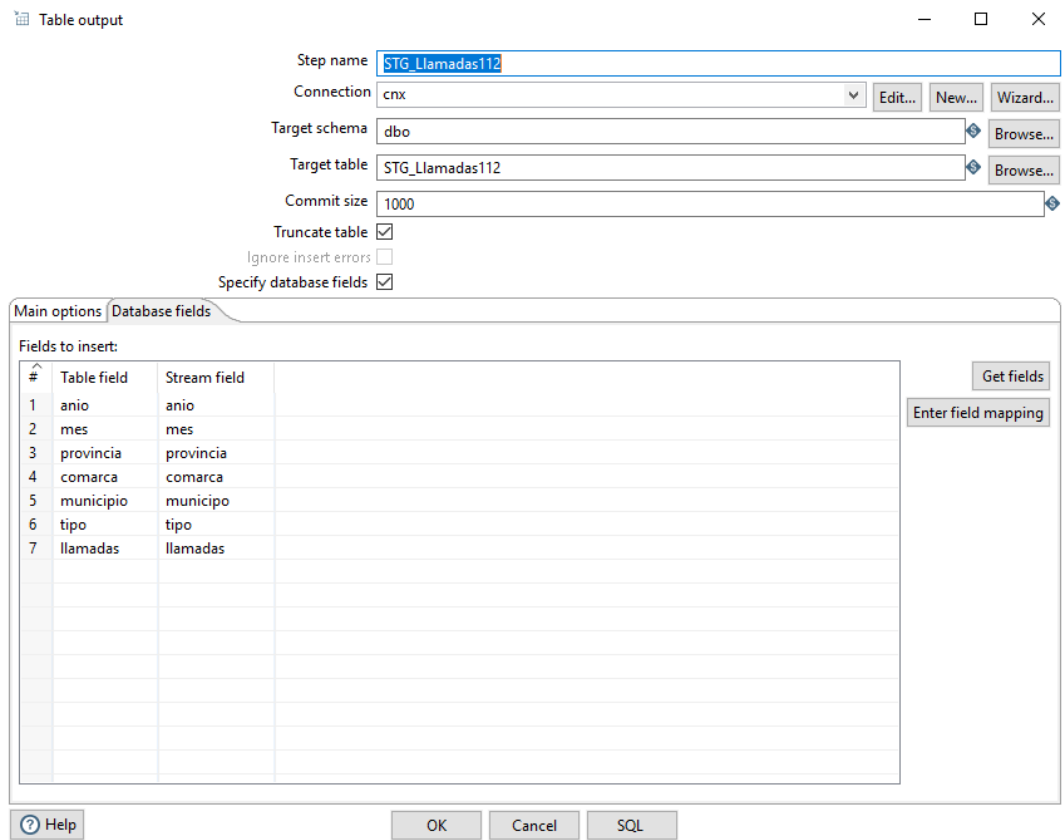


Ilustración 68 - Guardado IN_LLAMADAS112.

Al ejecutar la anterior transformación obtenemos las siguientes métricas:

Execution Results													
Logging Execution History Step Metrics Performance Graph Metrics Preview data													
#	Stepname	Copynr	Read	Written	Input	Output	Updated	Rejected	Errors	Active	Time	Speed (r/s)	input/output
1	Lectura XML (rows.xml)	0	0	340307	340307	0	0	0	0	Finished	32mn 44s	173	-
2	Value mapper	0	340307	340307	0	0	0	0	0	Finished	32mn 44s	173	-
3	Value mapper 2	0	340307	340307	0	0	0	0	0	Finished	32mn 45s	173	-
4	String operations	0	340307	340307	0	0	0	0	0	Finished	32mn 45s	173	-
5	STG_Llamadas112	0	340307	340307	0	340307	0	0	0	Finished	32mn 45s	173	-

Ilustración 69 - Métricas IN_LLAMADAS112.

Observamos que tenemos 340307 registros leídos y almacenamos el mismo número de registros, por lo que la información es correcta.

3.2.8. Transformación IN_FECHAS

La penúltima transformación que vamos a realizar respecto a este bloque es “IN_FECHAS”, su objetivo es leer todas las fechas que hay en todos los ficheros fuente y almacenarlas en una tabla intermedia llamada “STG_Fechas”.

La transformación nos ha quedado de la siguiente forma:

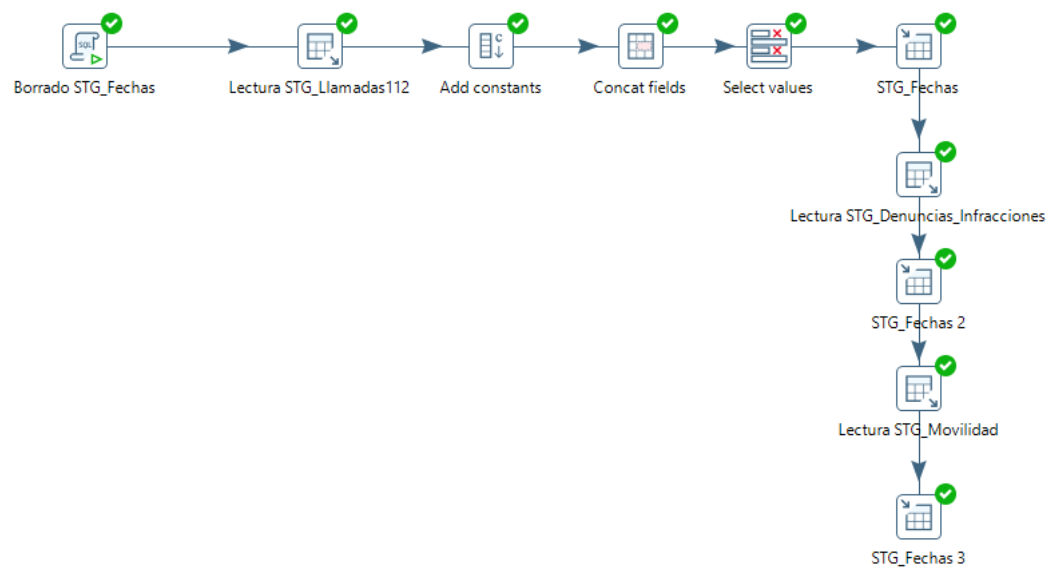


Ilustración 70 - IN_FECHAS.

Borrado

Lo primero de todo es hacer un borrado de la tabla, ya que al no obtener la información directamente de los ficheros puede darse el caso de que ya tengamos información en dicha tabla, por lo tanto borramos todos los registros de forma manual a partir de una sentencia SQL:

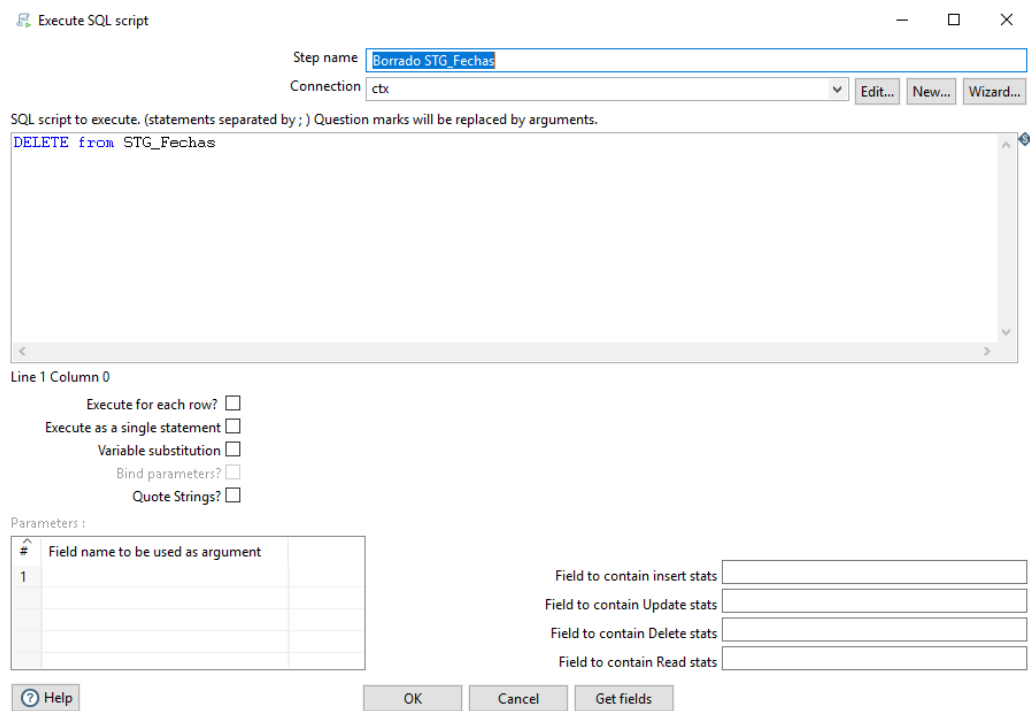


Ilustración 71 - Borrado IN_FECHAS.

Lectura

Una vez eliminados todos los registros leemos todas las fechas que se encuentran en la tabla intermedia “STG_Llamadas112”, para ello cargamos la información del mes y año tal y como se muestra en la siguiente ilustración:

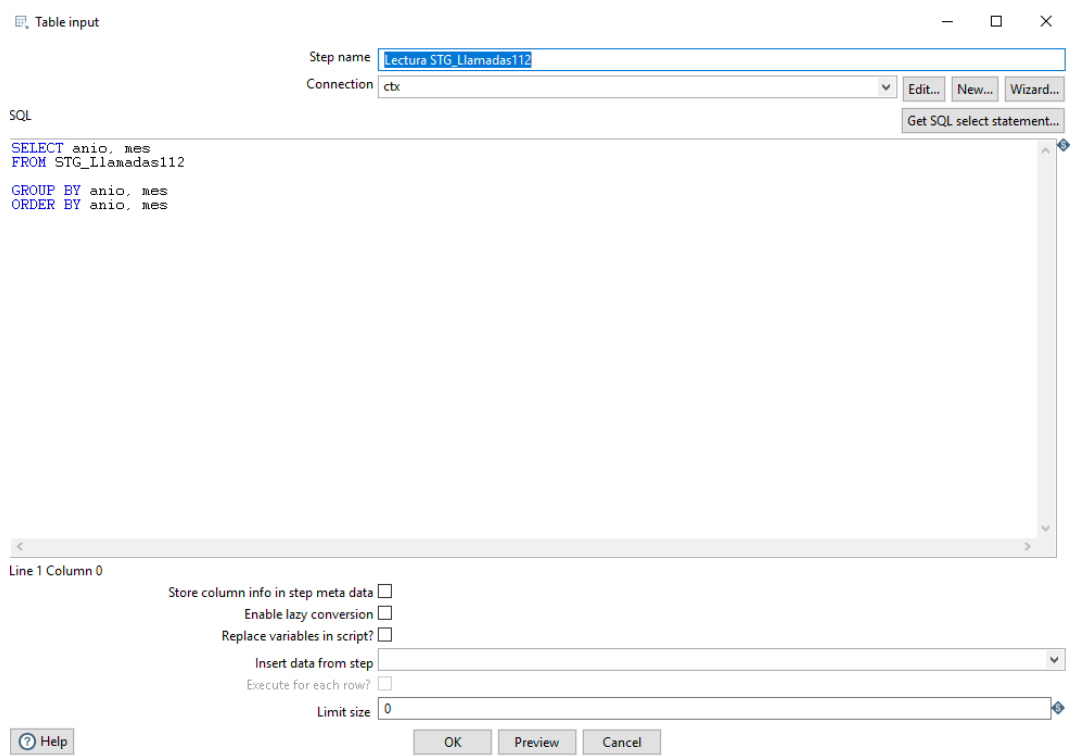


Ilustración 72 - Lectura IN_FECHAS.

Añadimos el día

Una característica de la información de “STG_Llamadas112” es que sí que se nos proporciona el año y mes pero no el día, es por ello que creamos un nuevo campo para el día cuyo valor va a ser siempre 1:

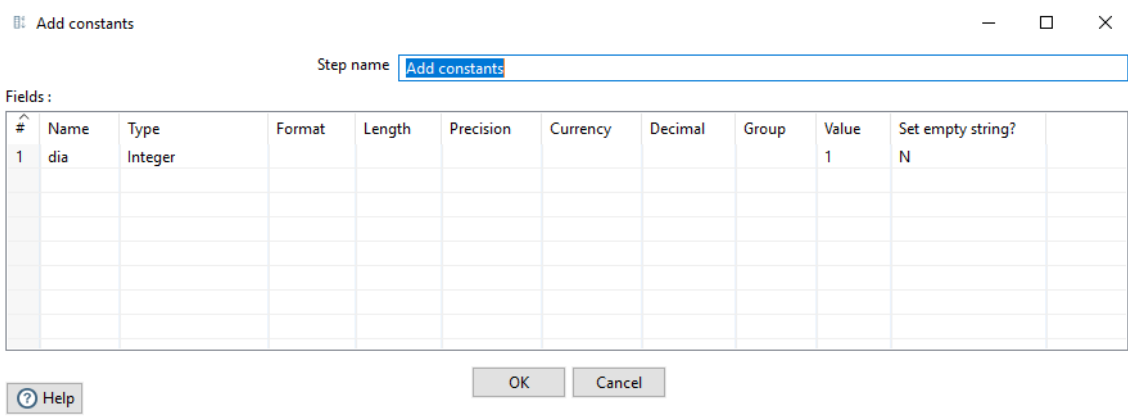


Ilustración 73 - Añadimos Constante IN_FECHAS.

Concatenación

Un aspecto a tener en cuenta es que respecto a “STG_Llamadas112” no tenemos una fecha como tal, sino que tenemos tres campos de tipo entero que nos indican el año, mes y día. Por lo tanto, lo primero que debemos de hacer es concatenar estos campos en un string:

Step name: Concat fields

Target Field Name: fecha_llamadas112

Length of Target Field: 0

Separator: /

Enclosure: "

#	Name	Type	Format	Length	Precision	Currency	Decimal	Group	Trim Type	Null
1	anio	String		4	0				both	
2	mes	String		2	0				both	
3	dia	String		2	0				both	

Buttons: Get Fields, Minimal width, Help, OK, Cancel

Ilustración 74 - Concatenación IN_FECHAS.

Conversión

Una vez que tenemos el string con el formato de la fecha, tenemos que convertir dicho campo a tipo date, para así poder almacenarlo en la base de datos:

Step name: Select values

Fields to alter the meta-data for:

#	Fieldname	Rename to	Type	Length	Precision	Binary t
1	fecha_llamadas112		Date			N

Buttons: Get fields to change, Help, OK, Cancel

Ilustración 75 - Conversión IN_FECHAS.

Guardado

Finalmente, guardamos todas las fechas de la tabla “STG_Llamadas112” en la tabla intermedia “STG_Fechas”, para ello asociamos el campo de la transformación con el de la tabla de la base de datos:

Table output

Step name: STG_Fechas

Connection: ctx

Target schema: dbo

Target table: STG_Fechas

Commit size: 1000

Truncate table: ☒

Ignore insert errors: ☐

Specify database fields: ☒

Main options / Database fields

Fields to insert:

#	Table field	Stream field
1	fecha	fecha_llamad...

Get fields

Enter field mapping

Help OK Cancel SQL

Ilustración 76 - Guardado IN_FECHAS.

Lectura

Al igual que hicimos con las llamadas al 112, tenemos que leer todas las fechas que hay en la tabla “STG_Denuncias_Infracciones”, para ello cargamos los datos a partir de la siguiente secuencia SQL:

Table input

Step name:

Connection:

SQL

```
SELECT fecha_final
FROM STG_Denuncias_Infracciones
```

Get SQL select statement...

Line 1 Column 0

☐ Store column info in step meta data

☐ Enable lazy conversion

☐ Replace variables in script?

Insert data from step:

☐ Execute for each row?

Limit size:

Ilustración 77 - Lectura IN_FECHAS.

Guardado

Como en este caso la tabla “STG_Denuncias_Infracciones” ya contiene todas las fechas de forma correcta las podemos guardar directamente en la base de datos, para ello asociamos los campos:

Table input

Step name:

Connection:

SQL:

Line 1 Column 0

☐ Store column info in step meta data

☐ Enable lazy conversion

☐ Replace variables in script?

Insert data from step

☐ Execute for each row?

Limit size

Ilustración 79 - Lectura IN_FECHAS.

Guardado

Al igual que en “STG_Denuncias_Infracciones” en “STG_Movilidad” no necesitamos realizar ninguna transformación, ya que todos los datos están de forma correcta. Debido a esto podemos almacenarlos directamente en la base de datos, y para ello asociamos los campos:

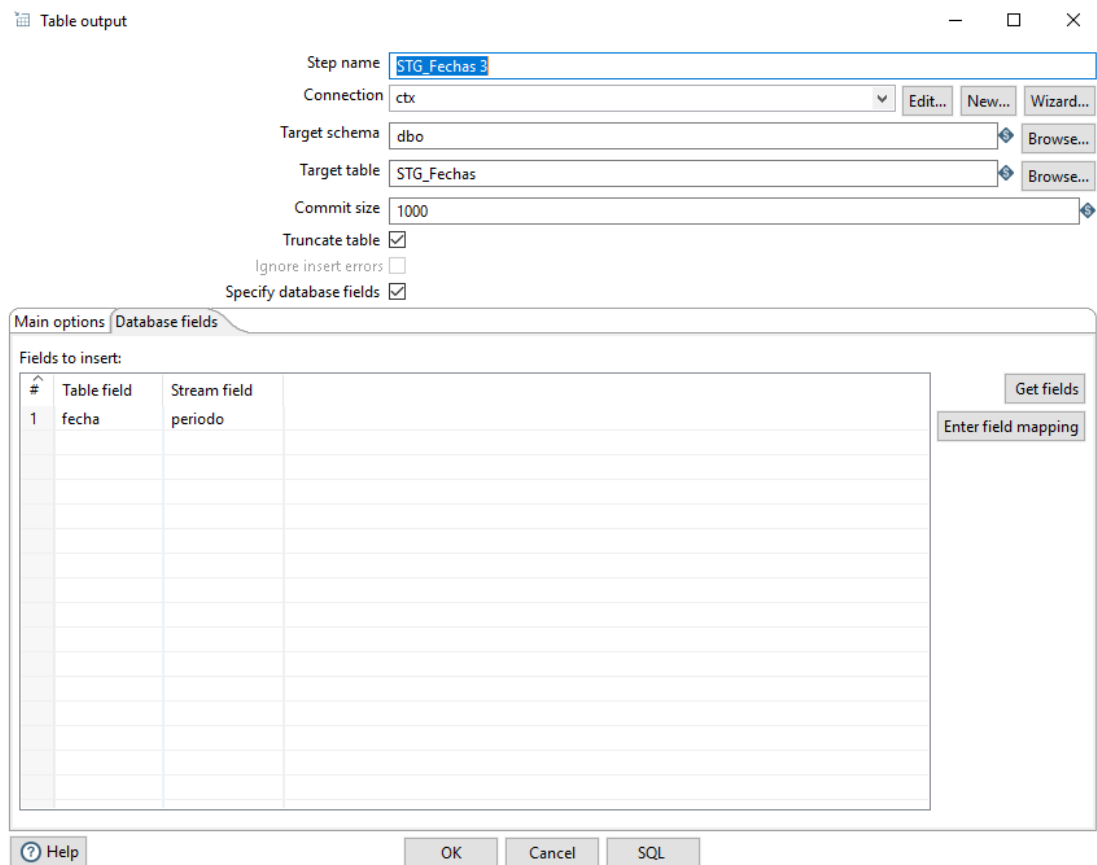


Ilustración 80 - Guardado IN_FECHAS.

Al ejecutar la anterior transformación obtenemos las siguientes métricas:

Execution Results

#	Stepname	Copynr	Read	Written	Input	Output	Updated	Rejected	Errors	Active	Time	Speed (r/s)	input/output
1	Borrado STG_Fechas	0	0	1	0	0	0	0	0	Finished	0.0s	-	-
2	Lectura STG_Llamadas112	0	0	82	82	0	0	0	0	Finished	0.1s	651	-
3	Add constants	0	82	82	0	0	0	0	0	Finished	0.1s	550	-
4	Concat fields	0	82	82	0	0	0	0	0	Finished	0.2s	366	-
5	Select values	0	82	82	0	0	0	0	0	Finished	0.2s	332	-
6	STG_Fechas	0	82	82	0	82	0	0	0	Finished	0.3s	293	-
7	Lectura STG_Denuncias_Infracciones	0	0	219	219	0	0	0	0	Finished	0.0s	16,846	-
8	STG_Fechas 2	0	219	219	0	219	0	0	0	Finished	0.2s	978	-
9	Lectura STG_Movilidad	0	0	4732	4732	0	0	0	0	Finished	0.0s	100,681	-
10	STG_Fechas 3	0	4732	4732	0	4732	0	0	0	Finished	0.4s	13,218	-

Ilustración 81 - Métricas IN_FECHAS.

De la anterior ejecución vemos que en “STG_Llamadas112” tenemos 82 fechas, en “STG_Denuncias_Infracciones” hay 219 fechas y en “STG_Movilidad” 4732. Estas fechas no significan que sean únicas, de hecho todo lo contrario, como veremos más adelante solo 170 fechas son diferentes.

3.2.9. Transformación IN_AMBITO_GEOGRAFICO

La última transformación que vamos a realizar respecto a este bloque es “IN_AMBITO_GEOGRAFICO”, su objetivo es leer y almacenar en

“STG_Ambito_Geografico” todas las localizaciones que existen en los ficheros proporcionados.

La transformación nos ha quedado de la siguiente forma:

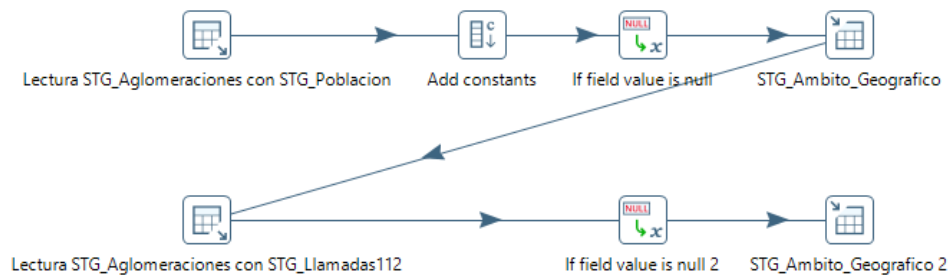


Ilustración 82 - IN_AMBITO_GEOGRAFICO.

Lectura

Lo primero que vamos a hacer es un join entre las tablas “STG_Evitar_Aglomeracion” y “STG_Poblacion”, de esta forma conseguimos obtener el código de la provincia, su nombre y su comunidad respecto a todas las provincias españolas. Como el campo en común entre las dos tablas es el nombre de la provincia, usamos dicho campos para hacer el join:

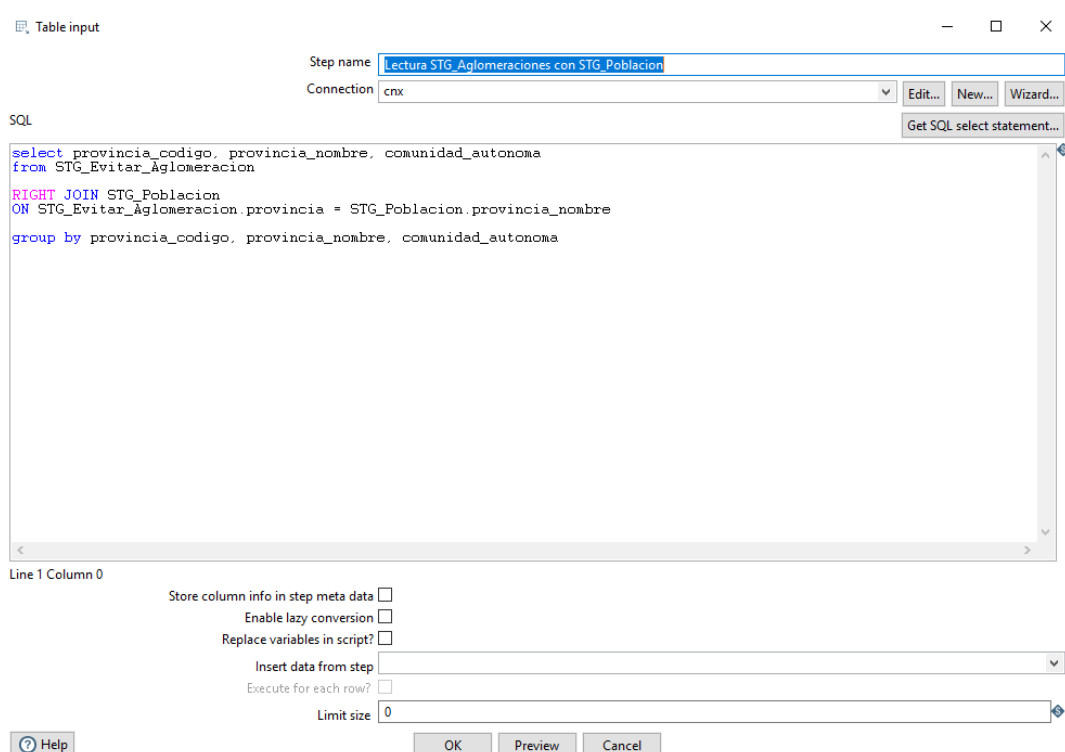


Ilustración 83 - Lectura IN_AMBITO_GEOGRAFICO.

Añadimos la comarca y el municipio

Debido a que en la dimensión ámbito geográfico tenemos que almacenar tanto la comarca y el municipio, debemos de crear dos atributos cuyos valores van a ser constantes “NA”. Esto se debe a que al hacer el join de las dos tablas anteriores carecemos de dichos campos, por lo tanto en Spoon nos queda de la siguiente forma:

[illegible]

Ilustración 84 - Constatnes Comarca Municipio IN_ AMBITO GEOGRAFICO.

Replace

Una vez que ya tenemos todos los datos puede darse el caso por ejemplo de que haya provincias sin comunidades, Ceuta y Melilla son ciudades autónomas pero no comunidades, por lo que vamos a reemplazar por “NA” aquellos valores que son nulos:

If field value is null

Step name

If field value is null

Replace Null for all fields

Replace by value

NA

Set empty string?

☐

Mask (Date)

Select fields

☐

Select value type

☐

Value types

#	Type	Replace by value	Conversion mask (Date)	Set empty string?

Fields

#	Field	Replace by value	Conversion mask (Date)	Set empty string?
1				

Help

OK

Get Fields

Cancel

Ilustración 85 - Replace IN_AMBITO_GEOGRAFICO.

Guardado

Finalmente, respecto a los datos de estas dos tablas los podemos guardar directamente a la base de datos, para ello indicamos la tabla “STG_Ambito_Geografico” y asociamos los valores:

[illegible]

Ilustración 86 - Guardado IN AMBITO GEOGRAFICO.

Lectura

Con los procesos anteriores conseguimos almacenar las provincias de España de forma general, sin embargo hay más localizaciones que se encuentran en las llamadas al 112, ya que en este fichero tenemos los campos comarca y municipio. Por lo tanto, hacemos un join entre “STG_Evitar_Aglomeracion”, “STG_Llamadas112” y “STG_Población”, de esta forma conseguimos el código de la provincia, su nombre, su comunidad autónoma, su comarca y su municipio:

Table input

Step name:

Connection:

SQL:

Get SQL select statement...

Line 1 Column 0

☐ Store column info in step meta data

☐ Enable lazy conversion

☐ Replace variables in script?

Insert data from step

☐ Execute for each row?

Limit size

Ilustración 87 - Lectura IN_AMBITO_GEOGRAFICO.

Replace

Al igual que antes puede darse el caso de que haya valores nulos en determinados atributos, esto lo solucionamos reemplazando los nulos por “NA”:

If field value is null

Step name

If field value is null 2

Replace Null for all fields

Replace by value

NA

Set empty string?

Mask (Date)

Select fields

Select value type

Value types

#	Type	Replace by value	Conversion mask (Date)	Set empty string?

Fields

#	Field	Replace by value	Conversion mask (Date)	Set empty string?
1				

Help

OK

Get Fields

Cancel

Ilustración 88 - Replace IN_AMBITO_GEOGRAFICO.

Guardado

Finalmente, ya tenemos todos los datos de forma correcta y los podemos guardar en la tabla “STG_Ambito_Geografico”. Por lo tanto, asociamos los atributos y ejecutamos la transformación:

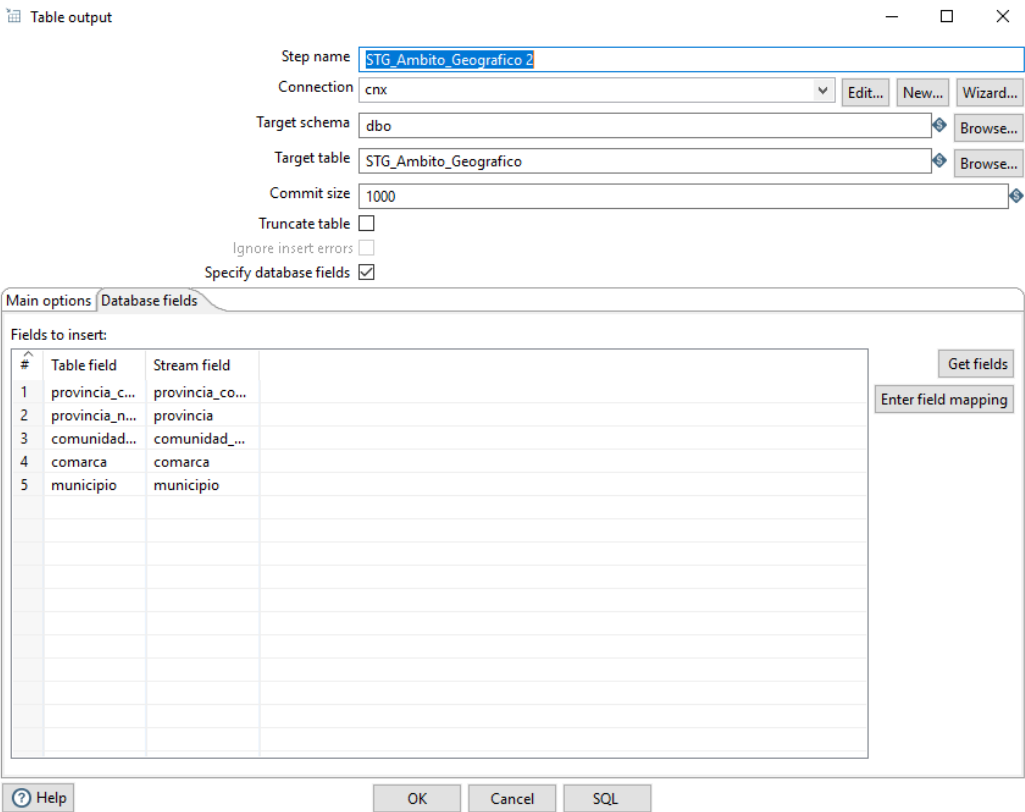


Ilustración 89 - Guardado IN_AMBITO_GEOGRAFICO.

Al ejecutar la anterior transformación obtenemos las siguientes métricas:

Execution Results												
Logging Execution History Step Metrics Performance Graph Metrics Preview data												
#	Stepname	Copynr	Read	Written	Input	Output	Updated	Rejected	Errors	Active	Time	Speed (r/s)
1	Lectura STG_Aglomeraciones con STG_Poblacion	0	0	52	52	0	0	0	0	Finished	0.0s	2,261
2	Add constants	0	52	52	0	0	0	0	0	Finished	0.0s	2,080
3	If field value is null	0	52	52	0	0	0	0	0	Finished	0.0s	1,625
4	STG_Ambito_Geografico	0	52	52	0	52	0	0	0	Finished	0.1s	553
5	Lectura STG_Aglomeraciones con STG_Llamadas112	0	0	971	971	0	0	0	0	Finished	3.3s	297
6	If field value is null 2	0	971	971	0	0	0	0	0	Finished	3.3s	296
7	STG_Ambito_Geografico 2	0	971	971	0	971	0	0	0	Finished	3.3s	292

Ilustración 90 - Métricas IN_AMBITO_GEOGRAFICO.

Como podemos apreciar de la primera lectura leemos 52 registros, uno por cada provincia española, y de la segunda 971 registros.

3.3. Bloque TR Dimensiones

Una vez que hemos almacenado toda la información en la base de datos gracias a las tablas intermedias, ahora vamos a hacer uso de estos datos para crear las diferentes dimensiones de nuestro modelo.

3.3.1. Transformación TR_DIM_GRUPO_EDAD

La primera transformación que vamos a realizar se llama “TR_DIM_GRUPO_EDAD”, su objetivo es almacenar los diferentes grupos de edad para así hacer uso de ellos en el hecho de mediciones, el resultado de esta transformación va a ser los datos almacenados en “DIM_Grupo_Edad”.

La transformación nos ha quedado de la siguiente forma:



Ilustración 91 - TR_DIM_GRUPO_EDAD.

Borrado

Lo primero que debemos de hacer es el borrado de los registros que contenía la dimensión, al estar relacionado con el hecho “FACT_Mediciones” tenemos que borrar este primero, para ello escribimos directamente la sentencia SQL y la ejecutamos:

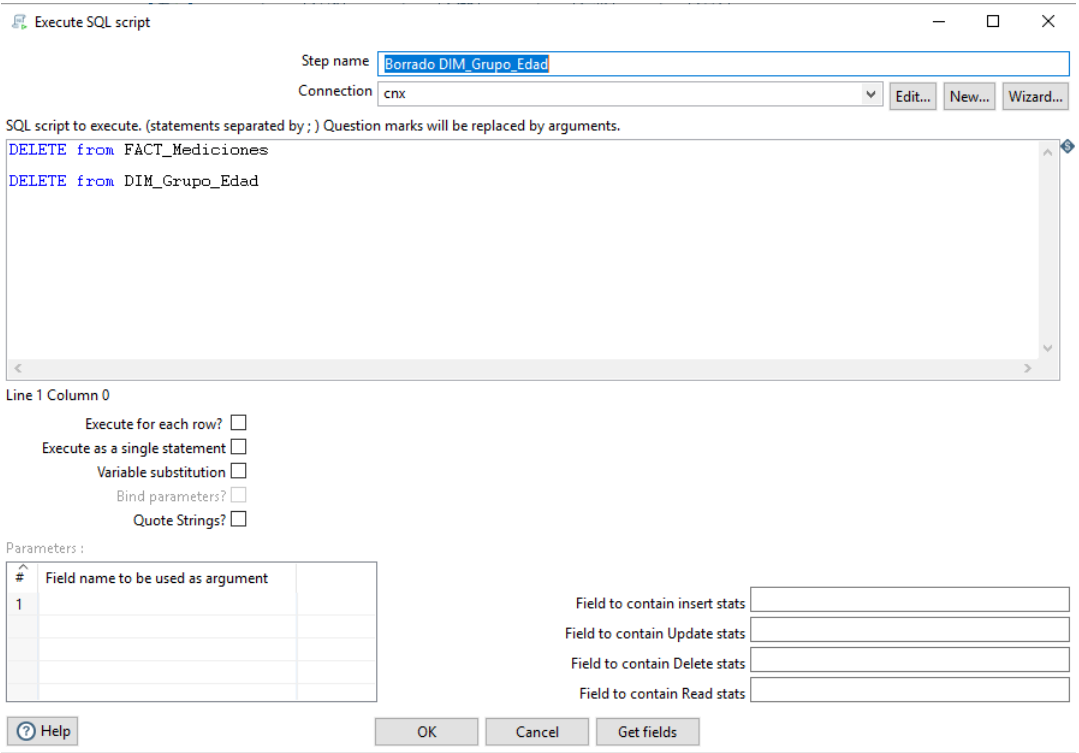


Ilustración 92 - Borrado TR_DIM_GRUPO_EDAD.

Grid

Puesto que la información de esta dimensión es fija y tiene tan solo 7 registros, nos resulta más fácil almacenar la información a partir de un grid (ya que en el enunciado de la práctica no se indica que no se pueda hacer uso de ellos), es por ello que hemos definido el siguiente grid:

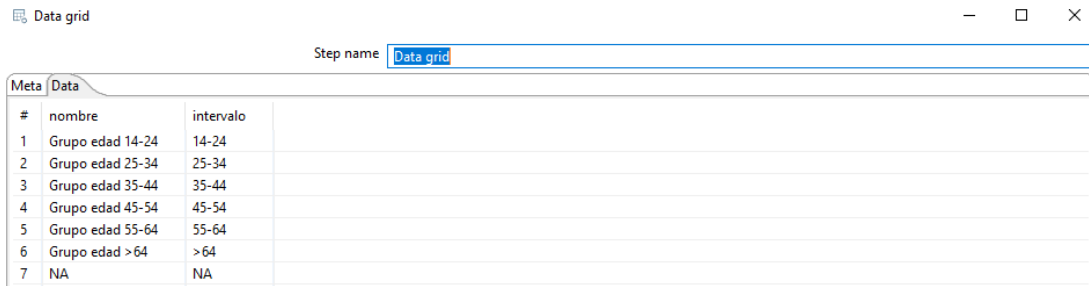
Data grid

Step name: Data grid

#	Name	Type	Format	Length	Precision	Currency	Decimal	Group	Set empty string?
1	nombre	String		20					N
2	intervalo	String		20					N

Ilustración 93 - Grid TR_DIM_GRUPO_EDAD.

Una vez definidos los campos, introducimos los registros de forma manual. Cabe destacar que vamos a tener un registro con valores “NA”, esto significa que esta dimensión no va a aplicar para calcular ciertas medidas:

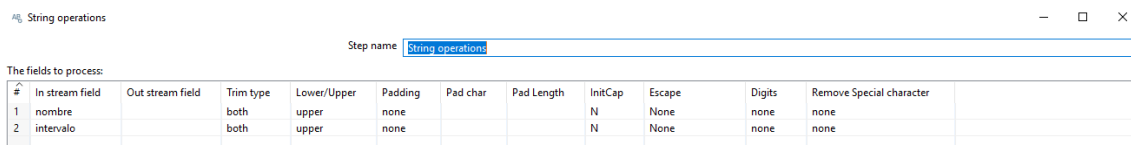


#	nombre	intervalo
1	Grupo edad 14-24	14-24
2	Grupo edad 25-34	25-34
3	Grupo edad 35-44	35-44
4	Grupo edad 45-54	45-54
5	Grupo edad 55-64	55-64
6	Grupo edad >64	>64
7	NA	NA

Ilustración 94 - Grid TR_DIM_GRUPO_EDAD.

Normalización

Normalizamos tanto el nombre como el intervalo para que estén en mayúsculas y no tengan espacios ni al principio ni al final:

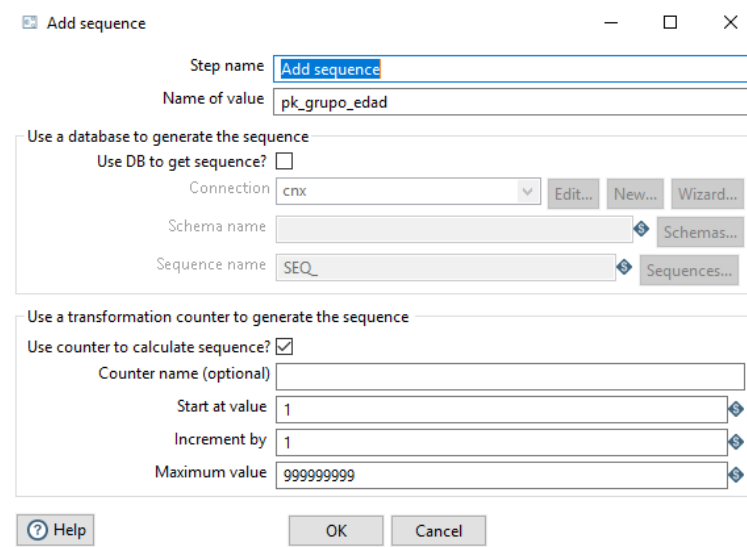


#	In stream field	Out stream field	Trim type	Lower/Upper	Padding	Pad char	Pad Length	InitCap	Escape	Digits	Remove Special character
1	nombre		both	upper	none			N	None	none	none
2	intervalo		both	upper	none			N	None	none	none

Ilustración 95 - Normalización TR_DIM_GRUPO_EDAD.

Secuenciación

Otro aspecto a destacar es que las dimensiones ya tienen claves primarias, por lo tanto vamos a definir la misma como un autonumérico incrementándose de uno en uno:



Step name: Add sequence

Name of value: pk_grupo_edad

Use a database to generate the sequence

Use DB to get sequence? ☐

Connection: cnx

Schema name:

Sequence name: SEQ_

Use a transformation counter to generate the sequence

Use counter to calculate sequence? ☒

Counter name (optional):

Start at value: 1

Increment by: 1

Maximum value: 99999999

Help OK Cancel

Ilustración 96 - Secuenciación TR_DIM_GRUPO_EDAD.

Guardado

Finalmente, realizamos el guardado en la dimensión indicando la tabla destino como "DIM_Grupo_Edad" y asociamos los atributos:

Table output

Step name:

Connection:

Target schema:

Target table:

Commit size:

Truncate table: ☐

Ignore insert errors: ☐

Specify database fields: ☒

Main options Database fields

Partition data over tables: ☐

Partitioning field:

Partition data per month: ☒

Partition data per day: ☐

Use batch update for inserts: ☒

Is the name of the table defined in a field? ☐

Field that contains name of table:

Store the tablename field: ☒

Return auto-generated key: ☐

Name of auto-generated key field:

Help OK Cancel SQL

Ilustración 97 - Guardado TR_DIM_GRUPO_EDAD.

[illegible]

Ilustración 98 - Guardado TR_DIM_GRUPO_EDAD.

Al ejecutar la anterior transformación obtenemos las siguientes métricas:

Execution Results

Logging Execution History Step Metrics Performance Graph Metrics Preview data

#	Stepname	Copynr	Read	Written	Input	Output	Updated	Rejected	Errors	Active	Time	Speed (r/s)	input/output
1	Borrado DIM_Grupo_Edad	0	0	1	0	0	0	0	0	Finished	0.0s	67	-
2	Data grid	0	0	7	0	0	0	0	0	Finished	0.0s	368	-
3	String operations	0	7	7	0	0	0	0	0	Finished	0.0s	292	-
4	Add sequence	0	7	7	0	0	0	0	0	Finished	0.0s	259	-
5	DIM_Grupo_Edad	0	7	7	0	7	0	0	0	Finished	0.1s	119	-

Ilustración 99 - Métricas TR_DIM_GRUPO_EDAD.

Como podemos observar generamos los 7 registros creados manualmente y guardamos todos en la base de datos.

3.3.2. Transformación TR_DIM_MEDICION

La segunda transformación que vamos a realizar se llama “TR_DIM_GRUPO_EDAD”, su objetivo es almacenar las diferentes medidas que vamos a usar en la tala de hechos mediciones, el resultado de esta transformación va a ser los datos almacenados en “DIM_Medicion”.

La transformación nos ha quedado de la siguiente forma:

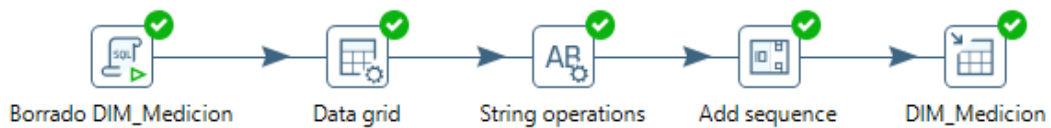


Ilustración 100 - TR_DIM_MEDICION.

Borrado

Lo primero que debemos de hacer es borrar todos los registros que hay en la tabla, al estar relacionado con “FACT_Mediciones” debemos de borrar los registros del hecho antes:

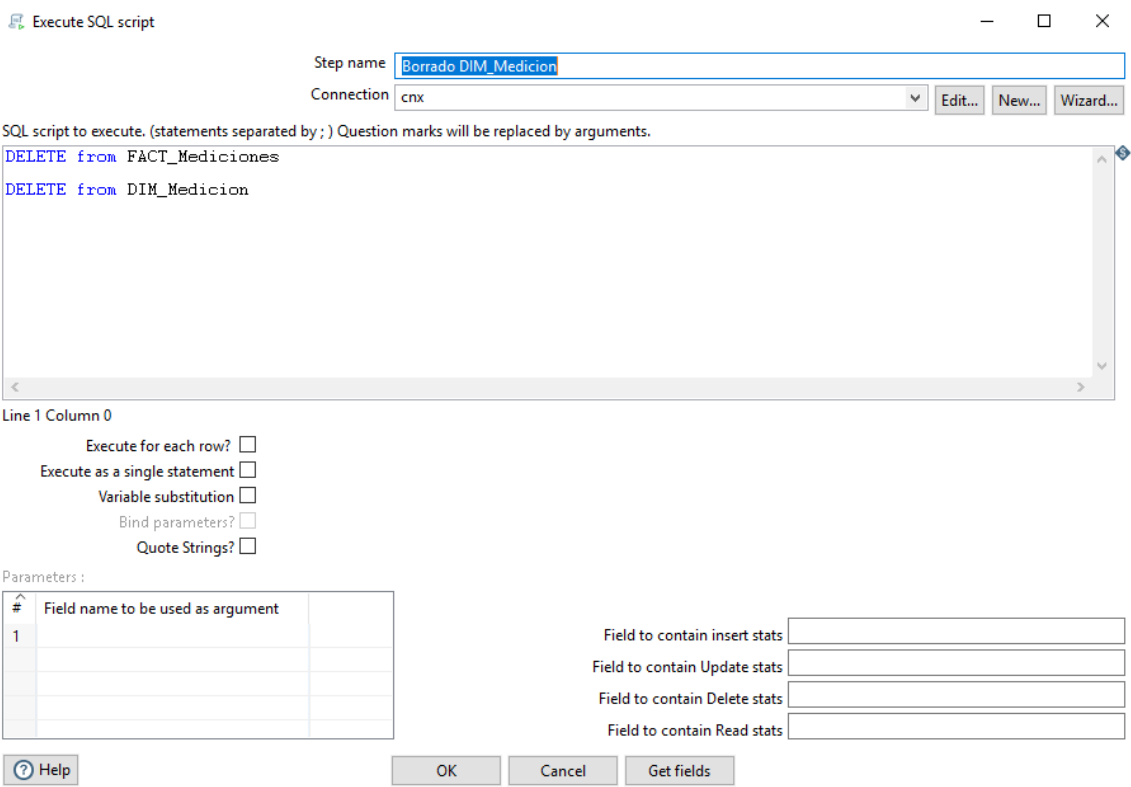


Ilustración 101 - Borrado TR_DIM_MEDICION.

Grid

Como los datos de esta dimensión no se encuentran en ningún fichero, la única solución que tenemos es introducirlos de forma manual, es por ello que hemos creado el siguiente grid:

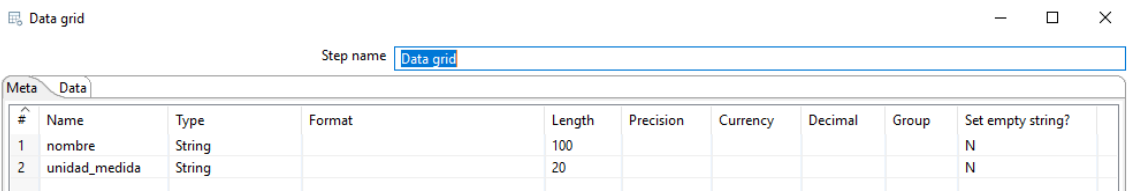


Ilustración 102 - Grid TR_DIM_DIM_MEDICION.

Una vez definidos los campos, introducimos los registros de forma manual:

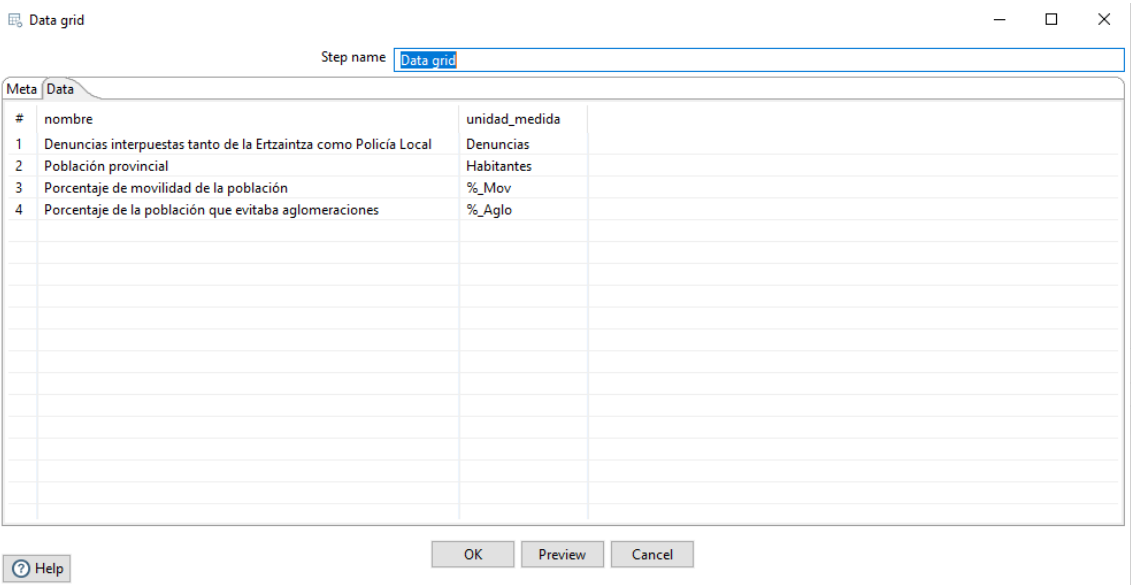


Ilustración 103 - Grid TR_DIM_MEDICION.

Normalización

Normalizamos tanto el nombre como la unidad de mediada, para que así todo esté en mayúsculas y sin espacios:

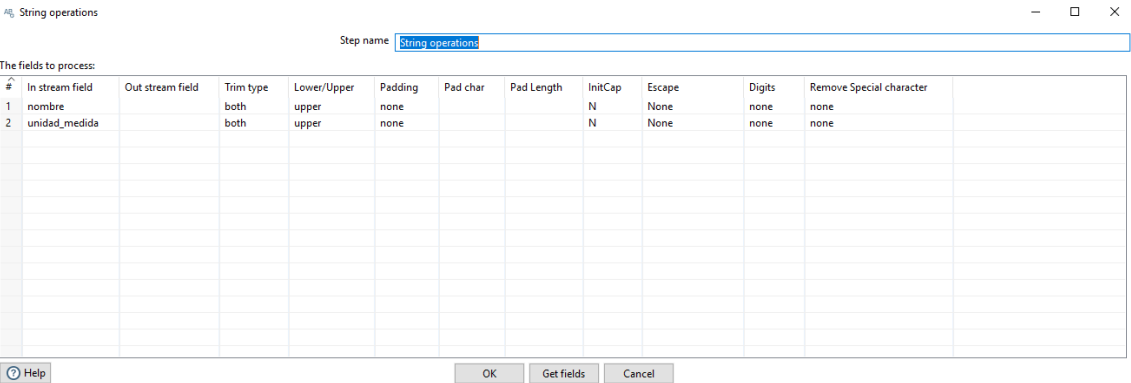


Ilustración 104 - Normalización TR_DIM_MEDICION.

Secuenciación

Al igual que sucedía antes, las tablas dimensiones ya tienen claves primarias, por lo que tenemos que definir la misma para esta dimensión:

Add sequence

Step name:

Name of value:

Use a database to generate the sequence

Use DB to get sequence? ☐

Connection:

Schema name:

Sequence name:

Use a transformation counter to generate the sequence

Use counter to calculate sequence? ☒

Counter name (optional):

Start at value:

Increment by:

Maximum value:

Ilustración 105 - Secuenciación TR_DIM_MEDICION.

Guardado

Una vez que ya tenemos todos los datos de forma correcta, procedemos a realizar el guardado en la tabla correspondiente, en nuestro caso “DIM_Medicion”:

Table output

Step name:

Connection:

Target schema:

Target table:

Commit size:

Truncate table: ☐

Ignore insert errors: ☐

Specify database fields: ☒

Main options **Database fields**

Partition data over tables: ☐

Partitioning field:

Partition data per month: ☒

Partition data per day: ☐

Use batch update for inserts: ☒

Is the name of the table defined in a field?: ☐

Field that contains name of table:

Store the tablename field: ☒

Return auto-generated key: ☐

Name of auto-generated key field:

Ilustración 106 - Guardado TR_DIM_MEDICION.

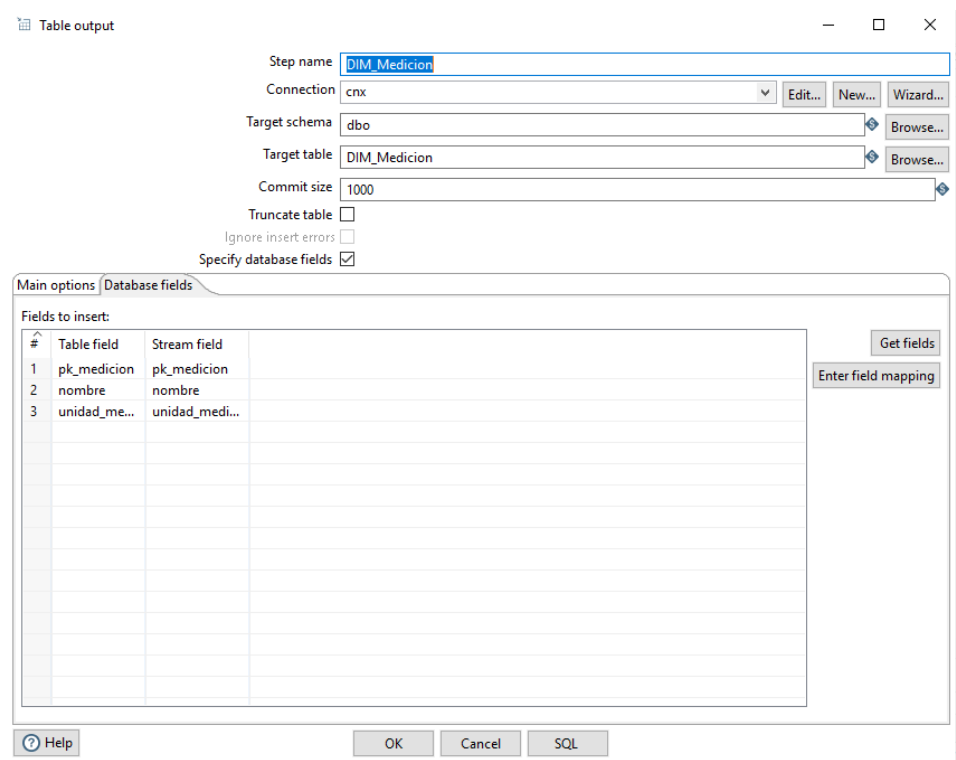


Ilustración 107 - Guardado TR_DIM_MEDICION.

Al ejecutar la anterior transformación nos proporciona las siguientes métricas:

Execution Results													
Logging Execution History Step Metrics Performance Graph Metrics Preview data													
#	Stepname	Copynr	Read	Written	Input	Output	Updated	Rejected	Errors	Active	Time	Speed (r/s)	input/output
1	Borrado DIM_Medicion	0	0	1	0	0	0	0	0	Finished	0.0s	33	-
2	Data grid	0	0	4	0	0	0	0	0	Finished	0.0s	143	-
3	String operations	0	4	4	0	0	0	0	0	Finished	0.1s	71	-
4	Add sequence	0	4	4	0	0	0	0	0	Finished	0.1s	59	-
5	DIM_Medicion	0	4	4	0	4	0	0	0	Finished	0.2s	16	-

Ilustración 108 - Métricas TR_DIM_MEDICIONES.

Como podemos observar generamos los 4 registros de forma manual y los guardamos perfectamente en la base de datos.

3.3.3. Transformación TR_DIM_TIPOLOGIA

La tercera transformación de este bloque se corresponde con “TR_DIM_TIPOLOGIA”, su objetivo es almacenar las diferentes tipologías en las llamadas al 112 en Cataluña, el resultado de esta transformación va a ser los datos almacenados en “DIM_Tipologia”.

La transformación nos ha quedado de la siguiente forma:



Ilustración 109 - TR_DIM_TIPOLOGIA.

Borrado

Al igual que en las transformaciones anteriores lo primero que debemos de hacer es el borrado de los registros que tenemos en la dimensión, pero antes hay que borrar los registros del hecho “FACT_Llamadas112”:

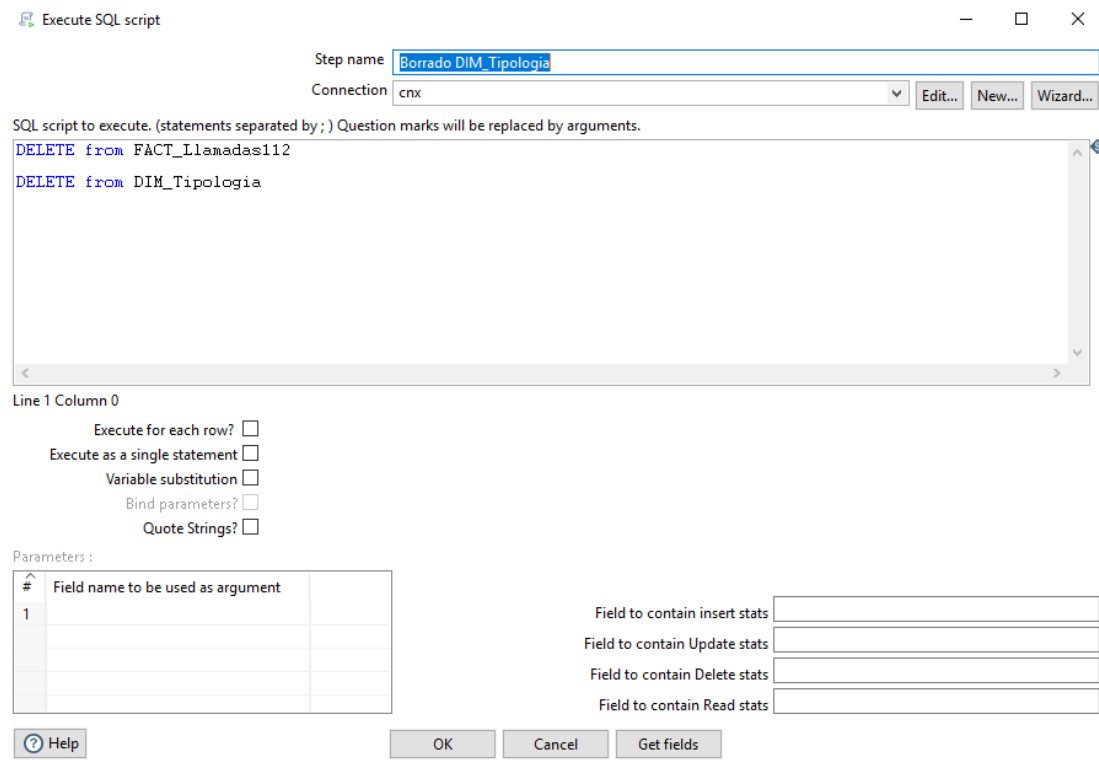


Ilustración 110 - Borrado TR_DIM_TIPOLOGIA.

Lectura

Aunque en este caso hay también pocas tipologías, para ser más exactos hay 10, podríamos haber usado un grid pero hemos considerado que lo mejor es hacer la lectura de la tabla intermedia “SGT_Llamadas112” porque la información no es fija, es decir, en un futuro pueden pasarnos tipologías nuevas y de no hacerlo así tendríamos que modificar la transformación.

Al hacer la lectura indicamos el campo “tipo”, la tabla “STG_Llamadas112” y con la función distinct nos quedamos con todas las tipologías diferentes:

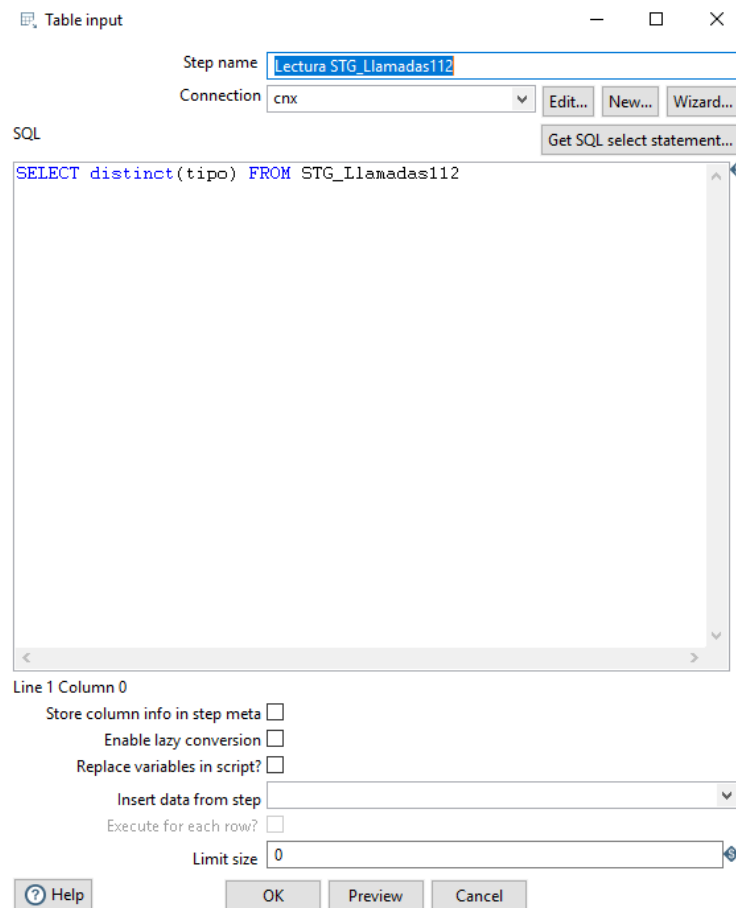


Ilustración 111 - Lectura TR_DIM_TIPOLOGIA.

Secuenciación

Al igual que en las transformaciones anteriores, definimos la clave primaria de “DIM_Tipologia” a partir de una secuencia numérica:

Add sequence

Step name: Add sequence

Name of value: pk_tipologia

Use a database to generate the sequence

Use DB to get sequence? ☐

Connection: cnx

Schema name:

Sequence name: SEQ_

Use a transformation counter to generate the sequence

Use counter to calculate sequence? ☒

Counter name (optional):

Start at value: 1

Increment by: 1

Maximum value: 999999999

Help OK Cancel

Ilustración 112 - Secuenciación TR_DIM_TIPOLOGIA.

Guardado

Finalmente, realizamos el guardado en la dimensión indicando la tabla destino como “DIM_Tipologia” y asociamos los campos:

Table output

Step name: DIM_Tipologia

Connection: cnx

Target schema: dbo

Target table: DIM_Tipologia

Commit size: 1000

Truncate table: ☐

Ignore insert errors: ☐

Specify database fields: ☒

Main options Database fields

Partition data over tables: ☐

Partitioning field:

Partition data per month: ☒

Partition data per day: ☐

Use batch update for inserts: ☒

Is the name of the table defined in a field? ☐

Field that contains name of table:

Store the tablename field: ☒

Return auto-generated key: ☐

Name of auto-generated key field:

Help OK Cancel SQL

Ilustración 113 - Guardado TR_DIM_TIPOLOGIA.

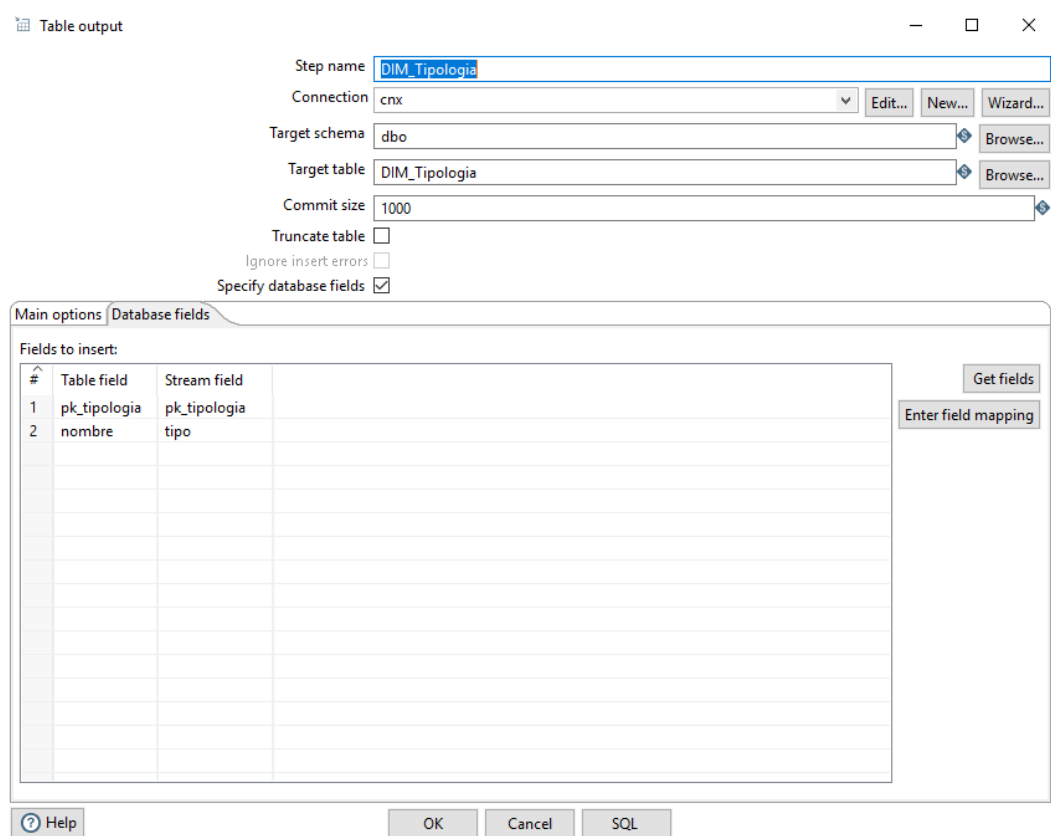


Ilustración 114 - Guardado TR_DIM_TIPOLOGIA.

Al ejecutar la anterior transformación obtenemos las siguientes métricas:

Execution Results

#	Stepname	Copynr	Read	Written	Input	Output	Updated	Rejected	Errors	Active	Time	Speed (r/s)	input/output
1	Borrado DIM_Tipologia	0	0	1	0	0	0	0	0	Finished	0.0s	33	-
2	Lectura STG_Llamadas112	0	0	10	10	0	0	0	0	Finished	0.4s	26	-
3	Add sequence	0	10	10	0	0	0	0	0	Finished	0.4s	25	-
4	DIM_Tipologia	0	10	10	0	10	0	0	0	Finished	0.4s	23	-

Ilustración 115 - Métricas TR_DIM_TIPOLOGIA.

3.3.4. Transformación TR_DIM_AMBITO_GEOGRAFICO

La cuarta transformación se corresponde con una dimensión compartida por ambos hechos, esta transformación se llama “TR_DIM_AMBITO_GEOGRAFICO” y se encarga de almacenar todos los datos geográficos recogidos de la fuentes proporcionadas, es decir, datos que se encuentran en las tablas intermedias.

Una vez que hemos leído todos los datos los vamos a almacenar a la tabla “DIM_Ambito_Geografico”, ya que es ésta la que se corresponde con la dimensión.

La transformación nos ha quedado de la siguiente forma:

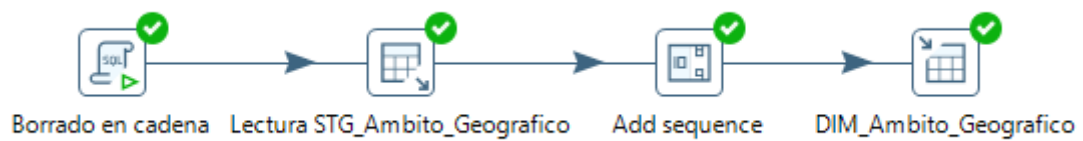


Ilustración 116 - TR_DIM_AMBITO_GEOGRAFICO.

Borrado

Lo primero que debemos de hacer es un borrado de los registros (si hay) de “DIM_Ambito_Geografico”, para ello escribimos directamente la sentencia SQL y la ejecutamos, cabe destacar que al estar relacionado con los hechos debemos primero eliminar los registros de los hechos:

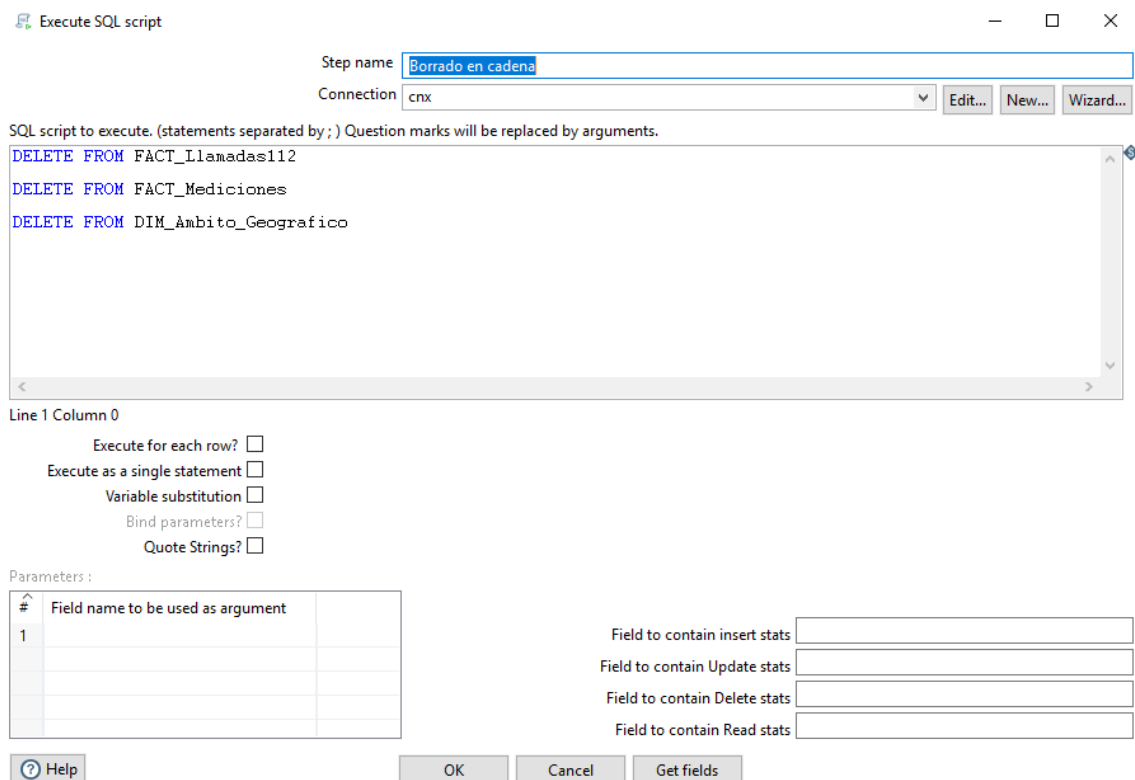


Ilustración 117 - Borrado TR_DIM_AMBITO_GEOGRAFICO.

Lectura

Una vez que hemos eliminado todos los registros los podemos volver a cargar, para ello tenemos que hacer una consulta a la tabla “STG_Ambito_Geografico”, y agrupamos para que no haya registros duplicados:

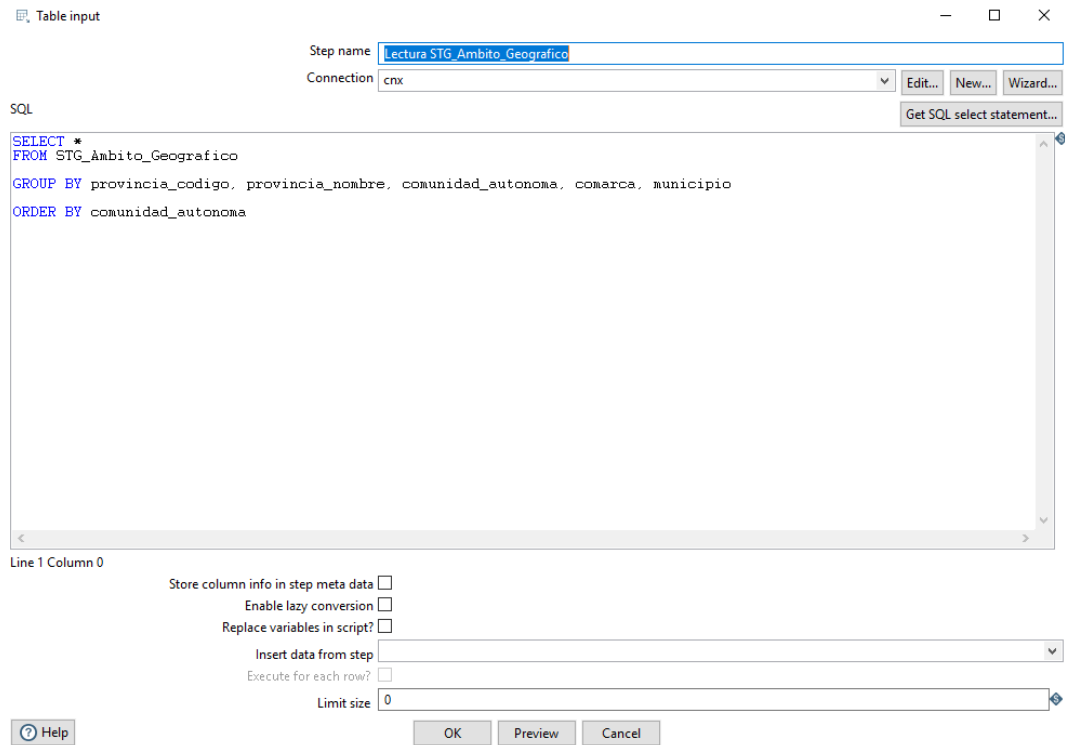


Ilustración 118 - Lectura TR_DIM_AMBITO_GEOGRAFICO.

Secuenciación

En este caso no normalizamos los datos porque ya lo hicimos al crear las tablas STG, de tal forma que todos los datos están en mayúsculas y sin espacios.

Otro aspecto a tener en cuenta es la creación de la clave primaria para esta dimensión, por lo que vamos a definir la misma como un autonumérico incrementándose de uno en uno:

Add sequence

Step name:

Name of value:

Use a database to generate the sequence

Use DB to get sequence? ☐

Connection:

Schema name:

Sequence name:

Use a transformation counter to generate the sequence

Use counter to calculate sequence? ☒

Counter name (optional):

Start at value:

Increment by:

Maximum value:

Ilustración 119 - Secuenciación TR_DIM_AMBITO_GEOGRAFICO.

Guardado

Finalmente, realizamos el guardado en la dimensión indicando la tabla destino como “DIM_Ambito_Geografico” y asociamos los atributos:

Table output

Step name:

Connection:

Target schema:

Target table:

Commit size:

Truncate table: ☐

Ignore insert errors: ☐

Specify database fields: ☒

Main options **Database fields**

Partition data over tables: ☐

Partitioning field:

Partition data per month: ☒

Partition data per day: ☐

Use batch update for inserts: ☒

Is the name of the table defined in a field?: ☐

Field that contains name of table:

Store the tablename field: ☒

Return auto-generated key: ☐

Name of auto-generated key field:

Ilustración 120 - Guardado TR_DIM_AMBITO_GEOGRAFICO.

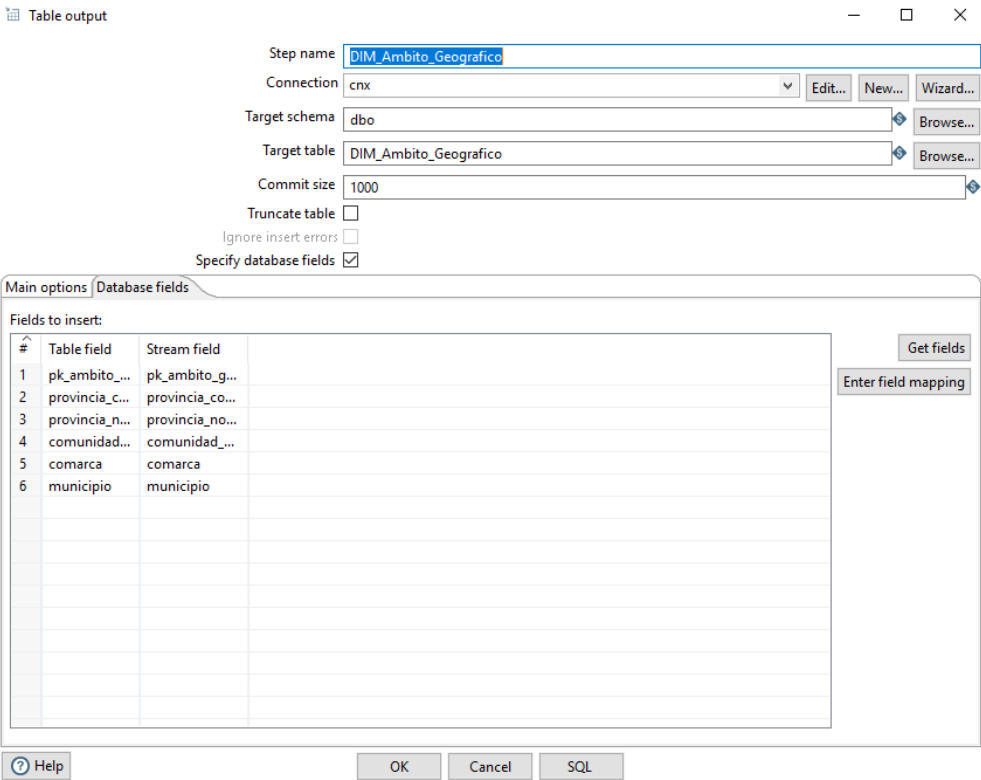


Ilustración 121 - Guardado TR_DIM_AMBITO_GEOGRAFICO.

El ejecutar la anterior transformación obtenemos las siguientes métricas:

Execution Results

Logging Execution History Step Metrics Performance Graph Metrics Preview data													
#	Stepname	Copynr	Read	Written	Input	Output	Updated	Rejected	Errors	Active	Time	Speed (r/s)	input/output
1	Borrado en cadena	0	0	1	0	0	0	0	0	Finished	0.0s	-	-
2	Lectura STG_Ambito_Geografico	0	0	1022	1022	0	0	0	0	Finished	0.0s	39,308	-
3	Add sequence	0	1022	1022	0	0	0	0	0	Finished	0.0s	21,745	-
4	DIM_Ambito_Geografico	0	1022	1022	0	1022	0	0	0	Finished	0.1s	10,758	-

Ilustración 122 - Métricas TR_DIM_AMBITO_GEOGRAFICO.

Como podemos observar leemos 1022 registros y almacenamos en la dimensión los mismos 1022 registros.

3.3.5. Transformación TR_DIM_FECHA

La última transformación respecto a las dimensiones es “TR_DIM_FECHA”, su objetivo es almacenar todas las fecha que se encuentran en las tablas intermedias y almacenarlas en “DIM_Fecha”.

La transformación nos ha quedado de la siguiente forma:

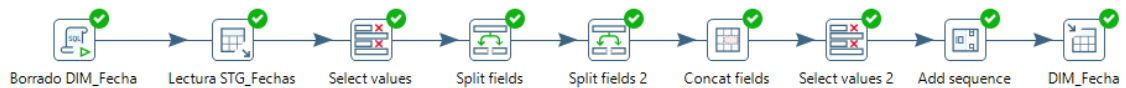


Ilustración 123 - TR_DIM_FECHA.

Borrado

Lo primero que debemos hacer es el borrado de los registros que contenía la dimensión, para ello escribimos directamente la sentencia SQL y la ejecutamos. Antes que nada debemos eliminar los registros de los hechos “FACT_Llamadas112” y “FACT_Mediciones”:

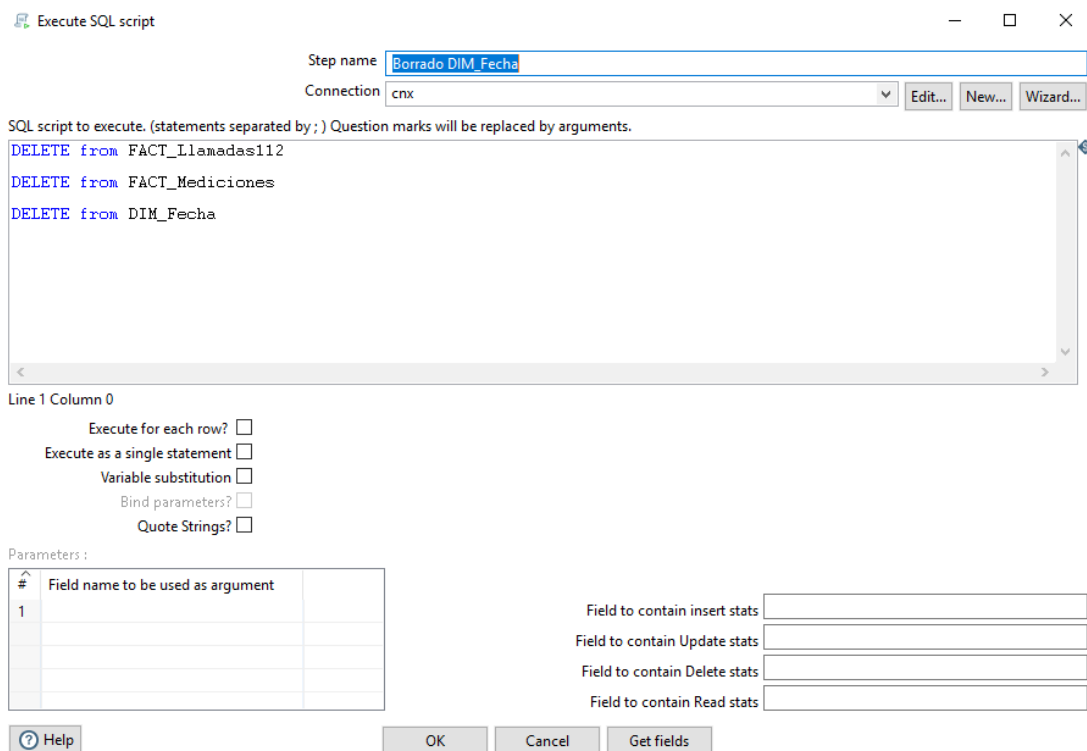


Ilustración 124 - Borrado TR_DIM_FECHA.

Lectura

Para hacer la lectura de las fechas tenemos que cargar todos los registros distintos de “STG_Fechas”, ya que es en esta tabla intermedia donde están almacenadas todas las fechas de todos los ficheros fuentes que nos han proporcionado.

Table input

Step name: Lectura STG_Fechas

Connection: cnx [Edit...] [New...] [Wizard...]

SQL

```
SELECT distinct (fecha)
FROM STG_Fechas
ORDER BY fecha
```

Get SQL select statement...

Line 1 Column 0

Store column info in step meta data ☐

Enable lazy conversion ☐

Replace variables in script? ☐

Insert data from step

Execute for each row? ☐

Limit size: 0

[Help] [OK] [Preview] [Cancel]

Ilustración 125 - Lectura TR_DIM_FECHA.

Conversión

Lo siguiente que debemos hacer es convertir el campo fecha que hemos leído en el paso anterior a string, ya que para esta dimensión no solo tenemos que guardar la fecha como tal, sino que también el día, mes y año por separado. Por lo tanto, convertimos a string la fecha:

Select values

Step name: Select values

Select & Alter Remove Meta-data

Fields to alter the meta-data for :

#	Fieldname	Rename to	Type	Length	Precision	Binary to Norm:
1	fecha		String			N

Get fields to change

[Help] [OK] [Cancel]

Ilustración 126 - Conversión String TR_DIM_FECHA.

Split

La fecha que tenemos en la base de datos nos proporciona también la hora, pero estos datos no nos interesan, es por ello que usamos split, le indicamos que divida la fecha a partir de un espacio en blanco:

Split fields

Step name

Split fields

Field to split

fecha

Delimiter

Enclosure

Fields

#	New field	ID	Remove ID?	Type	Length	Precision	Format	Group	Decimal	Currency	Nullif	Default	Trim type
1	fecha		N	String									both
2	time		N	String									both

Help

OK

Cancel

Ilustración 127 - Split TR_DIM_FECHA.

Split

Una vez que tenemos solamente la fecha en formato string, volvemos a hacer un split de los campos para obtener el día, mes y año en formato numérico:

Split fields

Step name

Split fields 2

Field to split

fecha

Delimiter

/

Enclosure

Fields

#	New field	ID	Remove ID?	Type	Length	Precision	Format	Group	Decimal	Currency	Nullif	Default	Trim type
1	año		N	Integer									none
2	mes		N	Integer									none
3	día		N	Integer									none

Help

OK

Cancel

Ilustración 128 - Split TR_DIM_FECHA.

Concatenación

Al hacer split sobre el campo fecha hemos perdido ese campo como tal, es decir, hemos perdido el campo que tenía tanto el mes, día y año con el formato “yyyy/MM/dd”, es por ello que a partir de los campos creados en el paso anterior volvemos a crear la fecha:

Step name: Concat fields

Target Field Name: fecha

Length of Target Field: 0

Separator: /

Enclosure:

Fields

#	Name	Type	Format	Length	Precision	Currency	Decimal	Group	Trim Type	Null
1	anyo	String							none	
2	mes	String							none	
3	dia	String							none	

Get Fields Minimal width

Help OK Cancel

Ilustración 129 - Concatenación TR_DIM_FECHA.

Conversión

Ese nuevo campo fecha está en formato string, pero en la dimensión necesitamos que sea de tipo date, es por ello que la convertimos a dicho tipo:

Step name: Select values 2

Select & Alter Remove Meta-data

Fields to alter the meta-data for:

#	Fieldname	Rename to	Type	Length	Precision	Binary to Normi
1	fecha		Date			N

Get fields to change

Help OK Cancel

Ilustración 130 - Conversión TR_DIM_FECHA.

Secuenciación

Al igual que sucedía con las dimensiones anteriores, necesitamos definir una clave primaria, es por ello que creamos un nuevo campo autonumérico que se va incrementado de uno en uno:

Add sequence

Step name:

Name of value:

Use a database to generate the sequence

Use DB to get sequence? ☐

Connection: Edit... New... Wizard...

Schema name: Schemas...

Sequence name: Sequences...

Use a transformation counter to generate the sequence

Use counter to calculate sequence? ☒

Counter name (optional):

Start at value: S...

Increment by: S...

Maximum value: S...

? Help OK Cancel

Ilustración 131 - Secuenciación TR_DIM_FECHA.

Guardado

Finalmente realizamos el guardado de todo este proceso en la “DIM_Fecha”, para ello asociamos los campos con la tabla de la base de datos:

[illegible]

Ilustración 132 - Guardado TR DIM FECHA.

Al ejecutar la anterior transformación obtenemos las siguientes métricas:

Execution Results

Logging Execution History Step Metrics Performance Graph Metrics Preview data

#	Stepname	Copynr	Read	Written	Input	Output	Updated	Rejected	Errors	Active	Time	Speed (r/s)	input/output
1	Borrado DIM_Fecha	0	0	1	0	0	0	0	0	Finished	0.0s	71	-
2	Lectura STG_Fechas	0	0	170	170	0	0	0	0	Finished	0.0s	6,296	-
3	Select values	0	170	170	0	0	0	0	0	Finished	0.0s	3,953	-
4	Split fields	0	170	170	0	0	0	0	0	Finished	0.1s	3,400	-
5	Split fields 2	0	170	170	0	0	0	0	0	Finished	0.1s	1,405	-
6	Concat fields	0	170	170	0	0	0	0	0	Finished	0.1s	1,338	-
7	Select values 2	0	170	170	0	0	0	0	0	Finished	0.1s	1,288	-
8	Add sequence	0	170	170	0	0	0	0	0	Finished	0.1s	1,141	-
9	DIM_Fecha	0	170	170	0	170	0	0	0	Finished	0.2s	867	-

Ilustración 133 - Métricas TR_DIM_FECHA.

Como podemos apreciar de la anterior ilustración, tenemos 170 fechas únicas entre todos los ficheros proporcionados, y éstas mismas 170 se almacenan en la dimensión de forma correcta.

3.4. Bloque TR Hechos

Una vez que hemos almacenado toda la información respectiva a las dimensiones, podemos hacer uso de estos datos junto con los de las tablas intermedias para crear las tablas de hechos.

3.4.1. Transformación TR_FACT_LLAMADAS112

La primera transformación que vamos a realizar se llama “TR_FACT_LLAMADAS112”, su objetivo es almacenar las métricas relativas a dicho hecho, el resultado de esta transformación va a ser los datos almacenados en “FACT_Llamadas112”.

La transformación nos ha quedado de la siguiente forma:

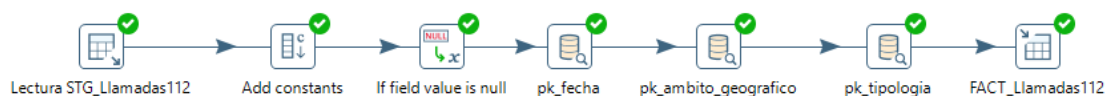


Ilustración 134 - TR_FACT_LLAMADAS112.

Lectura

Lo primero de todo es cargar la información que tenemos en la tabla intermedia “STG_Llamadas112”, en dicha tabla se encuentra las métricas de llamadas. Para cargar dicha información hacemos una consulta directamente a la base de datos:

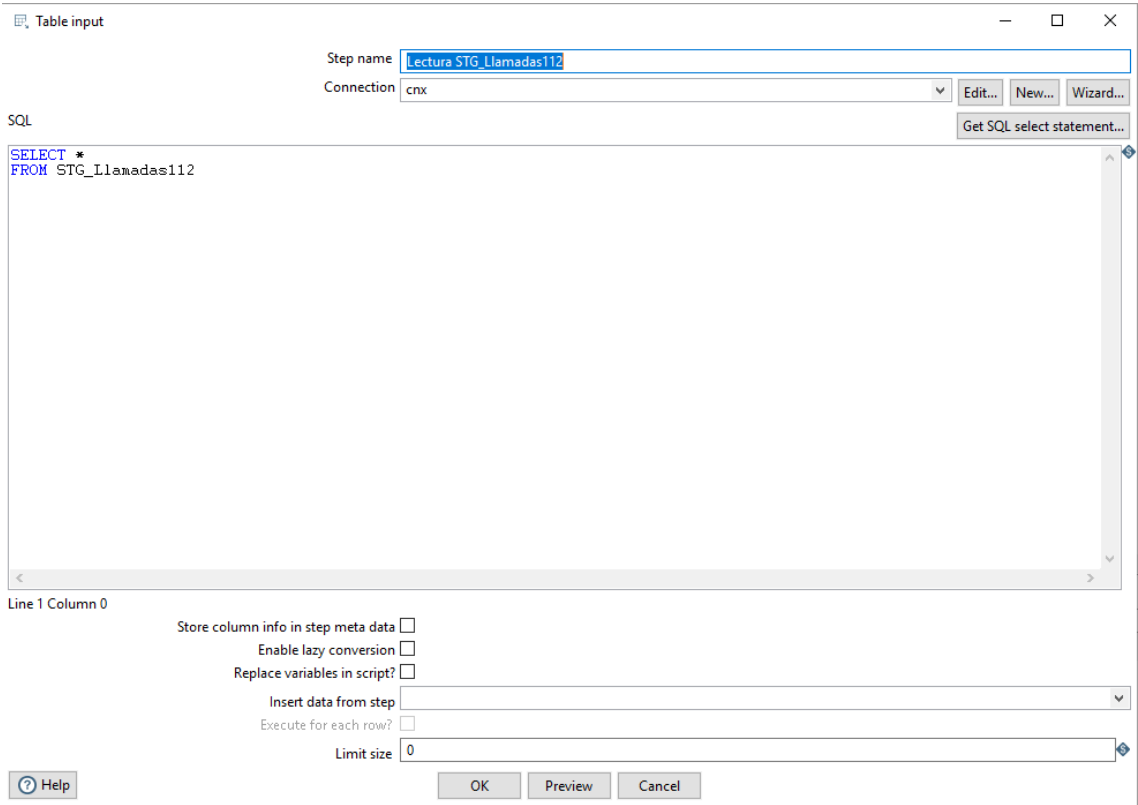


Ilustración 135 - Lectura TR_FACT_LLAMADAS112.

Añadimos el día

Aunque a priori no es necesario el día para obtener la fecha, ya que las llamadas se registran con el año y el mes, para evitar problemas en el futuro añadimos el día como una constantes con valor 1:

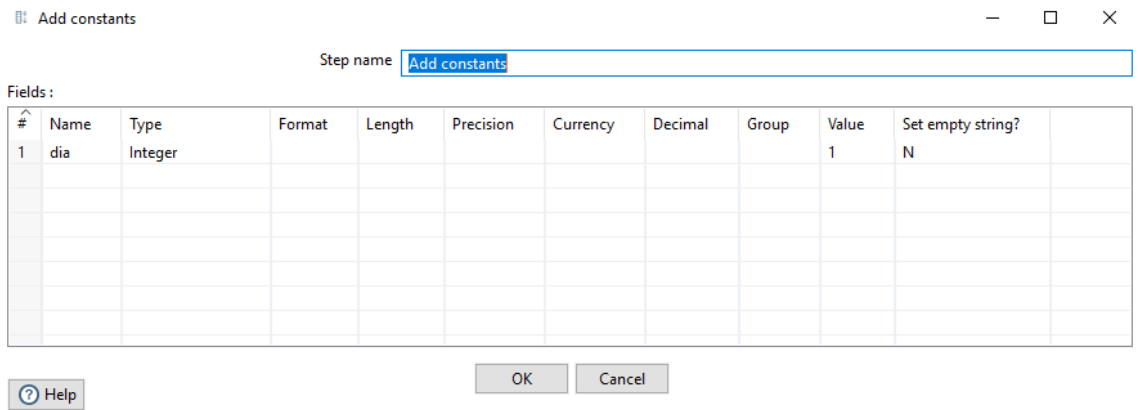


Ilustración 136 - Añadimos Día TR_FACT_LLAMADAS112.

Replace

El siguiente paso es hacer un replace de los valores nulos por “NA”, esto se debe a que en la tabla intermedia “STG_Llamadas112” almacenamos la información sin hacer

esta transformación (sustituir los valores nulos por “NA”), por ello indicamos que en los campos (provincia, comarca y municipio) si hay un valor nulo lo reemplace por “NA”:

If field value is null

Step name

If field value is null

Replace Null for all fields

Replace by value

Set empty string?

☐

Mask (Date)

Select fields

☒

Select value type

☐

Value types

#	Type	Replace by value	Conversion mask (Date)	Set empty string?

Fields

#	Field	Replace by value	Conversion mask (Date)	Set empty string?
1	provincia	NA		N
2	comarca	NA		N
3	municipio	NA		N

Help

OK

Get Fields

Cancel

Ilustración 137 - Replace TR_FACT_LLAMADAS112.

Búsqueda PK_Fecha

Una vez que ya tenemos todos los datos de forma correcta, pasamos a obtener las claves primarias que serán a su vez claves foráneas de la tabla del hecho. La primera clave primaria a buscar es la fecha, para ello la buscamos según su año, mes y día:

Database lookup

Step name

pk_fecha

Connection

cnx

Edit...

New...

Wizard...

Lookup schema

dbo

Browse...

Lookup table

DIM_Fecha

Browse...

Enable cache?

☐

Cache size in rows (0=cache)

0

Load all data from table

☐

The key(s) to look up the value(s):

#	Table field	Comparator	Field1	Field2
1	anyo	=	anio	
2	mes	=	mes	
3	dia	=	dia	

Values to return from the lookup table :

#	Field	New name	Default	Type
1	pk_fecha			None

Do not pass the row if the lookup fails

☐

Fail on multiple results?

☐

Order by

Help

OK

Cancel

Get Fields

Get lookup fields

Ilustración 138 - Búsqueda PK Fecha TR_FACT_LLAMADAS112.

Búsqueda PK_Ambito_Geografico

La segunda clave que tenemos que obtener está relacionada con la dimensión ámbito geográfico, para ello la buscamos a partir de su provincia, comarca y municipio:

Database lookup

Step name

pk_ambito_geografico

Connection

cnx

Edit...

New...

Wizard...

Lookup schema

dbo

Browse...

Lookup table

DIM_Ambito_Geografico

Browse...

Enable cache?

Cache size in rows (0=cache

0

Load all data from table

The key(s) to look up the value(s):

#	Table field	Comparator	Field1	Field2
1	provincia_nombre	=	provincia	
2	comarca	=	comarca	
3	municipio	=	municipio	

Values to return from the lookup table :

#	Field	New name	Default	Type
1	pk_ambito_geografico			None

Do not pass the row if the lookup fails

Fail on multiple results?

Order by

Help

OK

Cancel

Get Fields

Get lookup fields

Ilustración 139 - Búsqueda PK_Ambito_Geografico TR_FACT_LLAMADAS112.

Búsqueda PK_Tipología

La última clave foránea que tenemos en la tabla del hecho está relacionada con la clave primaria de la dimensión de tipologías. Para obtener dicha clave buscamos por el nombre de la tipología:

Database lookup

Step name

pk_tipologia

Connection

cnx

Edit...

New...

Wizard...

Lookup schema

dbo

Browse...

Lookup table

DIM_Tipologia

Browse...

Enable cache?

☐

Cache size in rows (0=cache)

0

Load all data from table

☐

The key(s) to look up the value(s):

#	Table field	Comparator	Field1	Field2
1	nombre	=	tipo	

Values to return from the lookup table :

#	Field	New name	Default	Type
1	pk_tipologia			None

Do not pass the row if the lookup fails

☐

Fail on multiple results?

☐

Order by

Help

OK

Cancel

Get Fields

Get lookup fields

Ilustración 140 - Obtención PK_Tipología TR_FACT_LLAMADAS112.

Guardado

Finalmente podemos guardar los datos del hecho, es decir, las claves foráneas junto con su métrica “llamadas”, para ello asociamos los campos:

Ilustración 141 - Guardado TR_FACT_LLAMADAS112.

Al ejecutar la anterior transformación obtenemos las siguientes métricas:

Execution Results

Logging Execution History Step Metrics Performance Graph Metrics Preview data

#	Stepname	Copynr	Read	Written	Input	Output	Updated	Rejected	Errors	Active	Time	Speed (r/s)	input/output
1	Lectura STG_Llamadas112	0	0	340307	340307	0	0	0	0	Finished	2mn 13s	2,550	-
2	Add constants	0	340307	340307	0	0	0	0	0	Finished	2mn 18s	2,461	-
3	If field value is null	0	340307	340307	0	0	0	0	0	Finished	2mn 23s	2,377	-
4	pk_fecha	0	340307	340307	340307	0	0	0	0	Finished	2mn 27s	2,301	-
5	pk_ambito_geografico	0	340307	340307	340307	0	0	0	0	Finished	2mn 32s	2,230	-
6	pk_tipologia	0	340307	340307	340307	0	0	0	0	Finished	2mn 32s	2,229	-
7	FACT_Llamadas112	0	340307	340307	0	340307	0	0	0	Finished	2mn 32s	2,229	-

Ilustración 142 - Métricas TR_FACT_LLAMADAS112.

De la ilustración anterior podemos ver que almacenamos 340307 registros.

3.4.2. Transformación TR_FACT_MEDICIONES

La última transformación respecto a esta práctica es “TR_FACT_MEDICIONES”, su objetivo es almacenar ciertas mediciones (la movilidad, los datos de la población, los expedientes sancionadores aplicados y el grado de consciencia de los ciudadanos para evitar aglomeraciones).

La transformación nos ha quedado de la siguiente forma:



Ilustración 143 - TR_FACT_MEDICIONES.

Como podemos apreciar esta transformación a nivel visual es la más compleja, no por su complejidad en sí, sino por la cantidad de elementos que hemos tenido que introducir debido a los diferentes flujos de entrada.

Cabe destacar que la metodología para poder almacenar de forma correcta cada flujo de entrada es igual, es decir, leemos la tabla intermedia correspondiente, añadimos las constantes necesarias, buscamos las diferentes claves foráneas, añadimos la clave primaria y lo almacenamos en la base de datos. Para que esta transformación siguiera un orden hemos decidido que la salida de una fuera la entrada de otra.

Por lo tanto, vamos a explicar que hemos hecho en esta transformación en cada uno de sus componentes, pero al ser todos los componentes comunes no vamos a entrar en mucho detalle para no explicar todo el rato lo mismo.

Lectura

En esta fase hemos realizado la lectura de todas las tablas intermedias que contenían la información necesaria para cumplir con el hecho.

Respecto a “STG_Evitar_Aglomeracion”, hemos leído todos los campos ya que éstos nos iban a hacer falta:

Table input

Step name: Lectura STG_Evitar_Aglomeracion

Connection: cnx

SQL:

```
SELECT *
FROM STG_Evitar_Aglomeracion
```

Line 1 Column 0

Store column info in step meta data ☐

Enable lazy conversion ☐

Replace variables in script? ☐

Insert data from step

Execute for each row? ☐

Limit size: 0

Help OK Preview Cancel

Ilustración 144 - Lectura TR_FACT_MEDICIONES.

Respecto a “STG_Denuncias_Infracciones”, también hemos leído todos los campos, aunque en este caso solo son necesarios: la provincia, la fecha y las denuncias (tanto de la Ertzaintza como de la Policía Local).

Cabe destacar que según la solución de la práctica uno, el valor de denuncias tiene que ser la suma de las denuncias de la Ertzaintza más las denuncias de la policía local, es por ello que la lectura respecto a esta tabla nos ha quedado de la siguiente forma:

Table input

Step name: Lectura STG_Denuncias_Infracciones

Connection: cnx

SQL:

```
SELECT *. (denuncias_ertzaintza + denuncias_ppl1) as denuncias
FROM STG_Denuncias_infracciones
```

Line 1 Column 0

Store column info in step meta data ☐

Enable lazy conversion ☐

Replace variables in script? ☐

Insert data from step

Execute for each row? ☐

Limit size: 0

Help OK Preview Cancel

Ilustración 145 - Lectura TR_FACT_MEDICIONES.

En cuanto a “STG_Poblacion” leemos todos los campos:

Table input

Step name:

Connection:

SQL:

Get SQL select statement...

Line 1 Column 0

Store column info in step meta data ☐

Enable lazy conversion ☐

Replace variables in script? ☐

Insert data from step

Execute for each row? ☐

Limit size:

Ilustración 146 - Lectura TR_FACT_MEDICIONES.

Finalmente hacemos la lectura de “STG_Movilidad”, al igual que antes leemos todos los campos:

Table input

Step name:

Connection:

SQL:

Get SQL select statement...

Line 1 Column 0

Store column info in step meta data ☐

Enable lazy conversion ☐

Replace variables in script? ☐

Insert data from step

Execute for each row? ☐

Limit size:

Ilustración 147 - Lectura TR_FACT_MEDICIONES.

Añadimos los campos necesarios

Una vez que hemos hecho todas las lecturas, debemos de añadir los campos necesarios para buscar las claves foráneas de nuestra tabla del hecho.

Respecto a “STG_Evitar_Aglomeracion”, creamos el campo fecha cuyo valor va a ser siempre el “2020-01-01”, creamos el campo medida con el valor “%_AGLO”, e indicamos que la comarca y el municipio son “NA” debido a que no tenemos dichos valores:

Add constants

Step name: Add constants

Fields:

#	Name	Type	Format	Length	Precision	Currency	Decimal	Group	Value	Set empty string?
1	Fecha_Evitar_Aglomeracion	Date	yyyy-MM-dd						2020-01-01	N
2	Medida_Evitar_Aglomeracion	String							%_AGLO	N
3	Comarca_Evitar_Aglomeracion	String							NA	N
4	Municipio_Evitar_Aglomeracion	String							NA	N

Help OK Cancel

Ilustración 148 - Añadimos Campos TR_FACT_MEDICIONES.

Respecto a “STG_Denuncias_Infracciones”, indicamos que la medida es “DENUNCIAS”, su comarca/municipio y grupo de edad es “NA” (carecemos de dicha información y el grupo de edad no aplica según estos datos) e indicamos que la comunidad autónoma es “PAÍS VASCO”:

Add constants

Step name: Add constants 2

Fields:

#	Name	Type	Format	Length	Precision	Currency	Decimal	Group	Value	Set empty string?
1	Medida_Denuncias	String							DENUNCIAS	N
2	Comarca_Denuncias	String							NA	N
3	Municipio_Denuncias	String							NA	N
4	Comunidad_Denuncias	String							PAÍS VASCO	N
5	Grupo_Edad_Denuncias	String							NA	N

Help OK Cancel

Ilustración 149 - Añadimos Campos TR_FACT_MEDICIONES.

Respecto a “STG_Poblacion”, indicamos que la medida es “HABITANTES”, su comarca/municipio y grupo de edad es “NA” (por el mismo motivo que en el caso anterior) y la fecha va a ser “2020-01-01” (en dicha tabla la fecha viene dada como “1 DE ENERO DE 2020” para todos los valores):

Add constants

Step name: Add constants 3

Fields:

#	Name	Type	Format	Length	Precision	Currency	Decimal	Group	Value	Set empty string?
1	Medida_Poblacion	String							HABITANTES	N
2	Comarca_Poblacion	String							NA	N
3	Municipio_Poblacion	String							NA	N
4	Grupo_Edad_Poblacion	String							NA	N
5	Fecha_Poblacion	Date	yyyy-MM-dd						2020-01-01	N

Help OK Cancel

Ilustración 150 - Añadimos Campos TR_FACT_MEDICIONES.

En cuanto a “STG_Movilidad”, indicamos que la medida es “%_MOV”, e indicamos que la comarca/municipio y grupo de edad es “NA” por los mismos motivos que antes:

Step name: Add constants4

Fields:

#	Name	Type	Format	Length	Precision	Currency	Decimal	Group	Value	Set empty string?
1	Medida_Movilidad	String							%_MOV	N
2	Comarca_Movilidad	String							NA	N
3	Municipio_Movilidad	String							NA	N
4	Grupo_Edad_Movilidad	String							NA	N

Buttons: Help, OK, Cancel

Ilustración 151 - Añadimos Campos TR_FACT_MEDICIONES.

Búsqueda PK_Fecha

El siguiente paso es obtener cada una de las claves foráneas, en nuestro caso, la primera es la fecha.

Respecto a “STG_Evitar_Aglomeracion”, buscamos la fecha como tal y devolvemos la clave primaria asociada a dicha fecha, el procedimiento va a ser el mismo en todos los flujos de entrada:

Step name: pk_fecha

Connection: cnx

Lookup schema: dbo

Lookup table: DIM_Fecha

Enable cache? ☐

Cache size in rows (0=cache): 0

Load all data from table ☐

The key(s) to look up the value(s):

#	Table field	Comparator	Field1	Field2
1	fecha	=	Fecha_Evitar_Aglomeracion	

Values to return from the lookup table:

#	Field	New name	Default	Type
1	pk_fecha			None

Do not pass the row if the lookup fails ☐

Fail on multiple results? ☐

Order by:

Buttons: Help, OK, Cancel, Get Fields, Get lookup fields

Ilustración 152 - Búsqueda PK_Fecha TR_FACT_MEDICIONES.

Respecto a “STG_Denuncias_Infracciones”, realizamos lo mismo que en caso anterior pero con el campo fecha correspondiente:

Database lookup

Step namepk_fecha 2

ConnectioncnxEdit...New...Wizard...

Lookup schemadbobrowse...

Lookup tableDIM_Fechabrowse...

Enable cache?☐

Cache size in rows (0=cache)0

Load all data from table☐

The key(s) to look up the value(s):

#	Table field	Comparator	Field1	Field2
1	fecha	=	fecha_final	

Values to return from the lookup table:

#	Field	New name	Default	Type
1	pk_fecha			None

Do not pass the row if the lookup fails☐

Fail on multiple results?☐

Order by

HelpOKCancelGet FieldsGet lookup fields

Ilustración 153 - Búsqueda PK_Fecha TR_FACT_MEDICIONES.

En cuanto a “STG_Poblacion” realizamos lo mismo pero con su atributo correspondiente:

Database lookup

Step name

pk_fecha

Connection

cnx

Edit...

New...

Wizard...

Lookup schema

dbo

Browse...

Lookup table

DIM_Fecha

Browse...

Enable cache?

☐

Cache size in rows (0=cache)

0

Load all data from table

☐

The key(s) to look up the value(s):

#	Table field	Comparator	Field1	Field2
1	fecha	=	Fecha_Poblacion	

Values to return from the lookup table:

#	Field	New name	Default	Type
1	pk_fecha			None

Do not pass the row if the lookup fails

☐

Fail on multiple results?

☐

Order by

Help

OK

Cancel

Get Fields

Get lookup fields

Ilustración 154 - Búsqueda PK_Fecha TR_FACT_MEDICIONES.

Finalmente en “STG_Movilidad” hacemos lo mismo:

Database lookup

Step name

pk_fecha 4

Connection

cnx

Edit...

New...

Wizard...

Lookup schema

dbo

Browse...

Lookup table

DIM_Fecha

Browse...

Enable cache?

☐

Cache size in rows (0=cache)

0

Load all data from table

☐

The key(s) to look up the value(s):

#	Table field	Comparator	Field1	Field2
1	fecha	=	periodo	

Values to return from the lookup table:

#	Field	New name	Default	Type
1	pk_fecha			None

Do not pass the row if the lookup fails

☐

Fail on multiple results?

☐

Order by

Help

OK

Cancel

Get Fields

Get lookup fields

Ilustración 155 - Búsqueda PK_Fecha TR_FACT_MEDICIONES.

Búsqueda PK_Ambito_Geografico

Lo siguiente es obtener la clave primaria respecto al ámbito geográfico de queda flujo de entrada.

En cuanto a “STG_Evitar_Aglomeracion”, buscamos tanto por el nombre de la provincia, su comunidad autónoma, comarca y municipio:

Database lookup

Step name

pk_ambito_geografico

Connection

cnx

Edit...

New...

Wizard...

Lookup schema

dbo

Browse...

Lookup table

DIM_Ambito_Geografico

Browse...

Enable cache?

Cache size in rows (0=cache)

0

Load all data from table

The key(s) to look up the value(s):

#	Table field	Comparator	Field1	Field2
1	provincia_nombre	=	provincia	
2	comunidad_autonoma	=	comunidad_autonoma	
3	comarca	=	Comarca_Evitar_Aglomeracion	
4	municipio	=	Municipio_Evitar_Aglomeracion	

Values to return from the lookup table:

#	Field	New name	Default	Type
1	pk_ambito_geografico			None

Do not pass the row if the lookup fails

Fail on multiple results?

Order by

Help

OK

Cancel

Get Fields

Get lookup fields

Ilustración 156 - Búsqueda PK_Ambito_Geografico TR_FACT_MEDICIONES.

Respecto a “STG_Denuncias_Infracciones” buscamos la clave primaria a partir de los mismos datos que en caso anterior, es decir, de la provincia, comunidad autónoma, comarca y municipio:

Database lookup

Step name: pk_ambito_geografico 2

Connection: cnx [Edit...] [New...] [Wizard...]

Lookup schema: dbo [Browse...]

Lookup table: DIM_Ambito_Geografico [Browse...]

Enable cache? ☐

Cache size in rows (0=cache): 0

Load all data from table ☐

The key(s) to look up the value(s):

#	Table field	Comparator	Field1	Field2
1	provincia_nombre	=	provincia	
2	comunidad_autonoma	=	Comunidad_Denuncias	
3	comarca	=	Comarca_Denuncias	
4	municipio	=	Municipio_Denuncias	

Values to return from the lookup table:

#	Field	New name	Default	Type
1	pk_ambito_geografico			None

Do not pass the row if the lookup fails ☐

Fail on multiple results? ☐

Order by:

[?] Help [OK] [Cancel] [Get Fields] [Get lookup fields]

Ilustración 157 - Búsqueda PK_Ambito_Geografico TR_FACT_MEDICIONES.

En cuanto a “STG_Poblacion”, realizamos un cambio respecto a los casos anteriores, ahora no tenemos la comunidad autónoma, sin embargo sí que tenemos el código de la provincia, por lo tanto buscamos la clave primaria a partir de la provincia, su código, comarca y municipio:

Database lookup

Step name

pk_ambito_geografico 3

Connection

cnx

Edit...

New...

Wizard...

Lookup schema

dbo

Browse...

Lookup table

DIM_Ambito_Geografico

Browse...

Enable cache?

☐

Cache size in rows (0=cache)

0

Load all data from table

☐

The key(s) to look up the value(s):

#	Table field	Comparator	Field1	Field2
1	provincia_nombre	=	provincia_nombre	
2	provincia_codigo	=	provincia_codigo	
3	comarca	=	Comarca_Poblacion	
4	municipio	=	Municipio_Poblacion	

Values to return from the lookup table:

#	Field	New name	Default	Type
1	pk_ambito_geografico			None

Do not pass the row if the lookup fails

☐

Fail on multiple results?

☐

Order by

Help

OK

Cancel

Get Fields

Get lookup fields

Ilustración 158 - Búsqueda PK_Ambito_Geografico TR_FACT_MEDICIONES.

Finalmente “STG_Movilidad”, solo tenemos información sobre el nombre de la provincia, su comarca y municipio, por lo que buscamos por esos campos:

Database lookup

Step name:

Connection:

Lookup schema:

Lookup table:

Enable cache? ☐

Cache size in rows (0=cache):

Load all data from table ☐

The key(s) to look up the value(s):

#	Table field	Comparator	Field1	Field2
1	provincia_nombre	=	zonas_movilidad	
2	comarca	=	Comarca_Movilidad	
3	municipio	=	Municipio_Movilidad	

Values to return from the lookup table:

#	Field	New name	Default	Type
1	pk_ambito_geografico			None

Do not pass the row if the lookup fails ☐

Fail on multiple results? ☐

Order by:

Ilustración 159 - Búsqueda PK_Ambito_Geografico TR_FACT_MEDICIONES.

Búsqueda PK_Grupo_Edad

La siguiente clave foránea que tenemos que buscar es el grupo de edad.

Respecto a “STG_Evitar_Aglomeracion”, vamos a buscar que el intervalo de grupo de edad coincida para así obtener la clave, este proceso es el mismo para todos los flujos de entrada:

Database lookup

Step name

pk_grupo_edad

Connection

cnx

Edit...

New...

Wizard...

Lookup schema

dbo

Browse...

Lookup table

DIM_Grupo_Edad

Browse...

Enable cache?

☐

Cache size in rows (0=cache)

0

Load all data from table

☐

The key(s) to look up the value(s):

#	Table field	Comparator	Field1	Field2
1	intervalo	=	grupo_edad	

Values to return from the lookup table :

#	Field	New name	Default	Type
1	pk_grupo_edad			None

Do not pass the row if the lookup fails

☐

Fail on multiple results?

☐

Order by

Help

OK

Cancel

Get Fields

Get lookup fields

Ilustración 160 - Búsqueda PK_Grupo_Edad TR_FACT_MEDICIONES.

Respecto a “STG_Denuncias_infracciones”:

Database lookup

Step name

pk_grupo_edad.2

Connection

cnx

Edit...

New...

Wizard...

Lookup schema

dbo

Browse...

Lookup table

DIM_Grupo_Edad

Browse...

Enable cache?

☐

Cache size in rows (0=cache)

0

Load all data from table

☐

The key(s) to look up the value(s):

#	Table field	Comparator	Field1	Field2
1	intervalo	=	Grupo_Edad_Denuncias	

Values to return from the lookup table:

#	Field	New name	Default	Type
1	pk_grupo_edad			None

Do not pass the row if the lookup fails

☐

Fail on multiple results?

☐

Order by

Help

OK

Cancel

Get Fields

Get lookup fields

Ilustración 161 - Búsqueda PK_Grupo_Edad TR_FACT_MEDICIONES.

En cuanto a “STG_Poblacion”:

Database lookup

Step name

pk_grupo_edad 3

Connection

cnx

Edit...

New...

Wizard...

Lookup schema

dbo

Browse...

Lookup table

DIM_Grupo_Edad

Browse...

Enable cache?

☐

Cache size in rows (0=cache)

0

Load all data from table

☐

The key(s) to look up the value(s):

#	Table field	Comparator	Field1	Field2	
1	intervalo	=	Grupo_Edad_Poblacion		

Values to return from the lookup table:

#	Field	New name	Default	Type	
1	pk_grupo_edad			None	

Do not pass the row if the lookup fails

☐

Fail on multiple results?

☐

Order by

Help

OK

Cancel

Get Fields

Get lookup fields

Ilustración 162 - Búsqueda PK_Grupo_Edad TR_FACT_MEDICIONES.

En cuanto a “STG_Movilidad”:

Database lookup

Step name

pk_grupo_edad 4

Connection

cnx

Edit...

New...

Wizard...

Lookup schema

dbo

Browse...

Lookup table

DIM_Grupo_Edad

Browse...

Enable cache?

☐

Cache size in rows (0=cache)

0

Load all data from table

☐

The key(s) to look up the value(s):

#	Table field	Comparator	Field1	Field2
1	intervalo	=	Grupo_Edad_Movilidad	

Values to return from the lookup table :

#	Field	New name	Default	Type
1	pk_grupo_edad			None

Do not pass the row if the lookup fails

☐

Fail on multiple results?

☐

Order by

Help

OK

Cancel

Get Fields

Get lookup fields

Ilustración 163 - Búsqueda PK_Grupo_Edad TR_FACT_MEDICIONES.

Búsqueda PK_Medicion

Finalmente, la última clava foránea que tenemos que buscar es la clave primaria respecto a la medición.

En cuanto a “STG_Evitar_Aglomeracion”, indicamos que busque la clave primaria a partir del nombre de la medición, este proceso es igual para todos los flujos de entrada:

Database lookup

Step name

pk_medicion

Connection

cnx

Edit...

New...

Wizard...

Lookup schema

dbo

Browse...

Lookup table

DIM_Medicion

Browse...

Enable cache?

Cache size in rows (0=cache)

0

Load all data from table

The key(s) to look up the value(s):

#	Table field	Comparator	Field1	Field2	
1	unidad_medida	=	Medida_Evitar_Aglomeracion		

Values to return from the lookup table:

#	Field	New name	Default	Type	
1	pk_medicion			None	

Do not pass the row if the lookup fails

Fail on multiple results?

Order by

Help

OK

Cancel

Get Fields

Get lookup fields

Ilustración 164 - Búsquedas PK_Medicion TR_FACT_MEDICIONES.

Respecto a “STG_Denuncias_Infracciones”:

Database lookup

Step name

pk_medicion2

Connection

cnx

Edit...

New...

Wizard...

Lookup schema

dbo

Browse...

Lookup table

DIM_Medicion

Browse...

Enable cache?

☐

Cache size in rows (0=cache)

0

Load all data from table

☐

The key(s) to look up the value(s):

#	Table field	Comparator	Field1	Field2	
1	unidad_medida	=	Medida_Denuncias		

Values to return from the lookup table:

#	Field	New name	Default	Type	
1	pk_medicion			None	

Do not pass the row if the lookup fails

☐

Fail on multiple results?

☐

Order by

Help

OK

Cancel

Get Fields

Get lookup fields

Ilustración 165 - Búsqueda PK_Medicion TR_FACT_MEDICIONES.

Respecto a “STG_Poblacion”:

Database lookup

Step name

pk_medicion 3

Connection

cnx

Edit...

New...

Wizard...

Lookup schema

dbo

Browse...

Lookup table

DIM_Medicion

Browse...

Enable cache?

☐

Cache size in rows (0=cache)

0

Load all data from table

☐

The key(s) to look up the value(s):

#	Table field	Comparator	Field1	Field2
1	unidad_medida	=	Medida_Poblacion	

Values to return from the lookup table :

#	Field	New name	Default	Type
1	pk_medicion			None

Do not pass the row if the lookup fails

☐

Fail on multiple results?

☐

Order by

Help

OK

Cancel

Get Fields

Get lookup fields

Ilustración 166 - Búsqueda PK_Medicion TR_FACT_MEDICIONES.

Respecto a “STG_Movilidad”:

Database lookup

Step name

pk_medicion 4

Connection

cnx

Edit...

New...

Wizard...

Lookup schema

dbo

Browse...

Lookup table

DIM_Medicion

Browse...

Enable cache?

Cache size in rows (0=cache)

0

Load all data from table

The key(s) to look up the value(s):

#	Table field	Comparator	Field1	Field2
1	unidad_medida	=	Medida_Movilidad	

Values to return from the lookup table :

#	Field	New name	Default	Type
1	pk_medicion			None

Do not pass the row if the lookup fails?

Fail on multiple results?

Order by

Help

OK

Cancel

Get Fields

Get lookup fields

Ilustración 167 - Búsqueda PK_Medicion TR_FACT_MEDICIONES.

Secuenciación

Una vez que ya tenemos todas las claves foráneas, tenemos que establecer la clave primaria de nuestra tabla del hecho, en este caso es una secuencia que se va autoincrementado de uno en uno.

Este componente es exactamente el mismo para todos los flujos de entrada, es por ello que solo lo vamos a indicar una vez el cómo está configurado:

Add sequence

Step name:

Name of value:

Use a database to generate the sequence

Use DB to get sequence? ☐

Connection:

Schema name:

Sequence name:

Use a transformation counter to generate the sequence

Use counter to calculate sequence? ☒

Counter name (optional):

Start at value:

Increment by:

Maximum value:

Ilustración 168 - Secuenciación TR_FACT_MEDICIONES.

Guardado

Cuando ya tenemos todos los datos podemos realizar el guardado en la tabla “FACT_Mediciones”, en ella vamos a almacenar toda la información necesaria respecto a las métricas.

En cuanto a “STG_Evitar_Aglomeracion”, al ser la primera transformación hacemos un truncate de la tabla por si hay algún registro. Por otro lado, indicamos el id, la fecha, el ámbito geográfico, el grupo de edad, la medición y la métrica (el porcentaje de la población que evitaba aglomeraciones):

[illegible]

Ilustración 169 - Guardado TR_FACT_MEDICIONES.

Respecto a “STG_Denuncias_Infracciones”, todos los campos son comunes con el caso anterior, a excepción de la métricas (en este caso el número de denuncias):

Table output

Step name: FACT_Mediciones 2

Connection: cnx

Target schema: dbo

Target table: FACT_Mediciones

Commit size: 1000

Truncate table: ☐

Ignore insert errors: ☐

Specify database fields: ☒

Main options Database fields

Fields to insert:

#	Table field	Stream field
1	pk_id	pk_id
2	fk_fecha	pk_fecha
3	fk_ambito_geografico	pk_ambito_geografico
4	fk_grupo_edad	pk_grupo_edad
5	fk_medicion	pk_medicion
6	valor	denuncias

Get fields

Enter field mapping

OK Cancel SQL Help

Ilustración 170 - Guardado TR_FACT_MEDICIONES.

En cuanto a “STG_Poblacion”, se difiere de los procesos anteriores por la métrica, en este caso la población por provincia que hay:

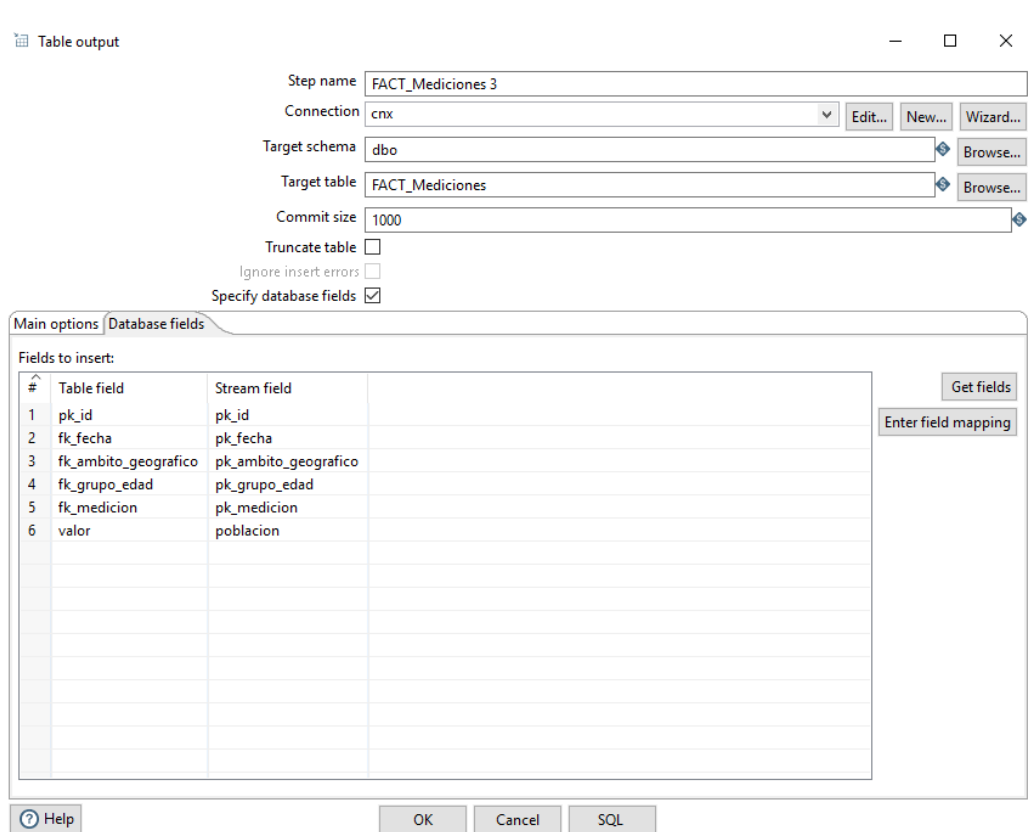


Ilustración 171 - Guardado TR_FACT_MEDICIONES.

Finalmente respecto a “STG_Movilidad”, indicamos el porcentaje de la población que se trasladó en el estado de alarma:

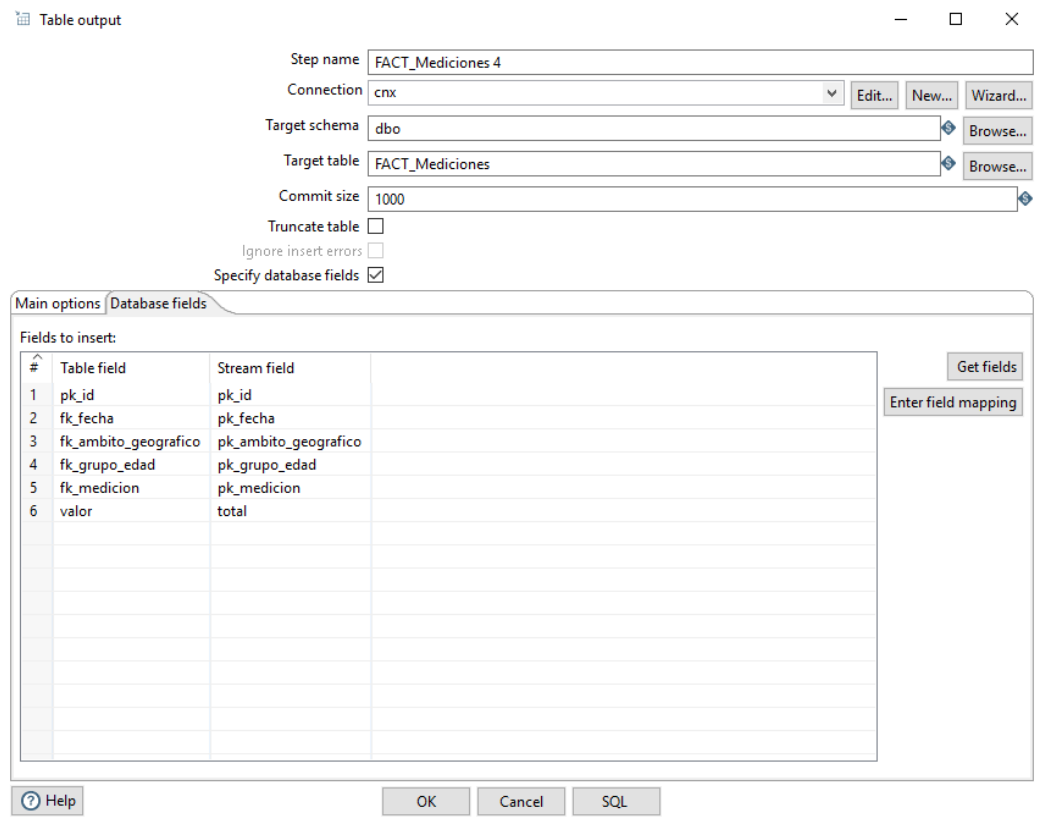


Ilustración 172 - Guardado TR_FACT_MEDICIONES.

Al ejecutar la anterior transformación obtenemos las siguientes métricas:

Execution Results

Logging Execution History Step Metrics Performance Graph Metrics Preview data													
#	Stepname	Copynr	Read	Written	Input	Output	Updated	Rejected	Errors	Active	Time	Speed (r/s)	input/output
1	Lectura STG_Evitar_Aglomeracion	0	0	300	300	0	0	0	0	Finished	0.0s	11,538	-
2	Add constants	0	300	300	0	0	0	0	0	Finished	0.2s	1,639	-
3	pk_fecha	0	300	300	300	0	0	0	0	Finished	0.4s	696	-
4	pk_ambito_geografico	0	300	300	300	0	0	0	0	Finished	0.6s	466	-
5	pk_grupo_edad	0	300	300	300	0	0	0	0	Finished	0.6s	488	-
6	pk_medicion	0	300	300	300	0	0	0	0	Finished	2.2s	138	-
7	Add sequence	0	300	300	0	0	0	0	0	Finished	2.2s	138	-
8	FACT_Mediciones	0	300	300	0	300	0	0	0	Finished	2.6s	114	-
9	Lectura STG_Denuncias_Infracciones	0	0	219	219	0	0	0	0	Finished	0.1s	1,698	-
10	Add constants 2	0	219	219	0	0	0	0	0	Finished	0.2s	1,197	-
11	pk_fecha 2	0	219	219	219	0	0	0	0	Finished	0.6s	370	-
12	pk_ambito_geografico 2	0	219	219	219	0	0	0	0	Finished	0.6s	367	-

Ilustración 173 - Métricas TR_FACT_MEDICIONES.

Execution Results

Execution Results													
Logging Execution History Step Metrics Performance Graph Metrics Preview data													
#	Stepname	Copynr	Read	Written	Input	Output	Updated	Rejected	Errors	Active	Time	Speed (r/s)	input/output
13	pk_grupo_edad 2	0	219	219	219	0	0	0	0	Finished	0.6s	358	-
14	pk_medicion 2	0	219	219	219	0	0	0	0	Finished	2.1s	102	-
15	Add sequence 2	0	219	219	0	0	0	0	0	Finished	2.1s	106	-
16	FACT_Mediciones 2	0	219	219	0	219	0	0	0	Finished	2.6s	83	-
17	Lectura STG_Poblacion	0	0	52	52	0	0	0	0	Finished	0.2s	284	-
18	Add constants 3	0	52	52	0	0	0	0	0	Finished	0.2s	284	-
19	pk_fecha 3	0	52	52	52	0	0	0	0	Finished	0.4s	126	-
20	pk_ambito_geografico 3	0	52	52	52	0	0	0	0	Finished	0.6s	91	-
21	pk_grupo_edad 3	0	52	52	52	0	0	0	0	Finished	0.6s	90	-
22	pk_medicion 3	0	52	52	52	0	0	0	0	Finished	0.6s	89	-
23	Add sequence 3	0	52	52	0	0	0	0	0	Finished	0.6s	92	-
24	FACT_Mediciones 3	0	52	52	0	52	0	0	0	Finished	2.6s	20	-

Ilustración 174 - Métricas TR_FACT_MEDICIONES.

25	Lectura STG_Movilidad	0	0	4732	4732	0	0	0	0	Finished	0.4s	11,485	-
26	Add constants 4	0	4732	4732	0	0	0	0	0	Finished	0.2s	25,858	-
27	pk_fecha 4	0	4732	4732	4732	0	0	0	0	Finished	5.2s	913	-
28	pk_ambito_geografico 4	0	4732	4732	4732	0	0	0	0	Finished	5.5s	857	-
29	pk_grupo_edad 4	0	4732	4732	4732	0	0	0	0	Finished	5.6s	844	-
30	pk_medicion 4	0	4732	4732	4732	0	0	0	0	Finished	5.7s	827	-
31	Add sequence 4	0	4732	4732	0	0	0	0	0	Finished	5.7s	827	-
32	FACT_Mediciones 4	0	4732	4732	0	4732	0	0	0	Finished	5.8s	820	-

Ilustración 175 - Métricas TR_FACT_MEDICIONES.

De la anterior ejecución podemos ver que de “STG_Evitar_Aglomeracion” almacenamos 300 registros, de “STG_Denuncias_Infracciones” 219 registros, de “STG_Poblacion” 52 registros y de “STG_Movilidad” 4732 registros. Esto hace un total de 5303 registros en la tabla “FACT_Mediciones”.

4. Implementación de jobs

Teniendo en cuenta los bloques de procesos implementados:

- Bloque IN: procesos ETL de transformación y carga al área intermedia.
- Bloque TR_DIM: procesos ETL de transformación y carga de dimensiones.
- Bloque TR_FACT: procesos ETL de transformación y carga de hechos.

Vamos a diseñar los trabajos (*jobs*) mediante PDI que van a permitir la ejecución secuencial de todos los procesos ETL incluidos en cada bloque.

4.1. JOB_IN

El trabajo “JOB_IN” ejecuta todas las transformaciones encargadas de almacenar la información en las tablas intermedias (*staging area*).

El diseño de este trabajo es el siguiente:



Ilustración 176 - JOB_IN.

Este trabajo está formado por los siguientes componentes:

- Start: nos indica el comienzo del trabajo, en él no hay que configurar nada.
- IN_DENUNCIAS_INFRACCIONES: transformación de IN_DENUNCIAS_INFRACCIONES, para realizar esta transformación hay que indicar el fichero en el que se encuentra la misma. Por otro lado, tenemos un total de 219 registros.

Transformation

Entry Name:

Transformation:

Ilustración 177 – Transformación IN_DENUNCIAS_INFRACCIONES.

- IN_POBLACION: transformación de IN_POBLACION. Hay un total de 52 registros.

Transformation

Entry Name:

Transformation:

Ilustración 178 - Transformación IN_POBLACION.

- IN_MOVILIDAD: transformación de IN_MOVILIDAD, tiene un total de 4732 registros.

Transformation

Entry Name:

Transformation:

Ilustración 179 - Transformación IN_MOVILIDAD.

- IN_AGLOMERACION: transformación de IN_AGLOMERACION, tiene un total de 300 registros.

✖ Transformation

Entry Name:

IN_AGLOMERACION

Transformation:

`${Internal.Entry.Current.Directory}/IN_AGLOMERACION.ktr` Browse...

Ilustración 180 - Transformación IN_AGLOMERACION.

- IN_LLAMADAS112: transformación de IN_LLAMADAS112, tiene un total de 340307 registros.

✖ Transformation

Entry Name:

IN_LLAMADAS112

Transformation:

`${Internal.Entry.Current.Directory}/IN_LLAMADAS112.ktr` Browse...

Ilustración 181 Transformación IN_LLAMADAS112.

- IN_FECHAS: transformación de IN_FECHAS, tiene un total de 814 registros.

✖ Transformation

Entry Name:

IN_FECHAS

Transformation:

`${Internal.Entry.Current.Directory}/IN_FECHAS.ktr` Browse...

Ilustración 182 – Transformación IN_FECHAS.

- IN_AMBITO_GEOGRAFICO: transformación de IN_AMBITO_GEOGRAFICO, tiene en total 1023 registros.

Transformation

Entry Name:

Transformation:

Ilustración 183 - Transformación IN_AMBITO_GEOGRAFICO.

- Success: nos indica que el trabajo ha finalizado.

Como resultado de ejecutar el anterior trabajo obtenemos las siguiente métricas:

Execution Results

Job / Job Entry	Comment	Result	Reason	Filename	Nr	Log date
Job: JOB_IN						
Start	Start of job execution		start			2021/05/18 13:43:46
Start	Job execution finished	Success			0	2021/05/18 13:43:46
IN_DENUNCIAS_INFRACCION	Start of job execution		Followed unconditional link	file:///F:/Mario/PRA2/IN_DENUNC...		2021/05/18 13:43:46
IN_DENUNCIAS_INFRACCION	Job execution finished	Success		file:///F:/Mario/PRA2/IN_DENUNC...	1	2021/05/18 13:43:51
IN_POBLACION	Start of job execution		Followed link after success	file:///F:/Mario/PRA2/IN_POBLACI...		2021/05/18 13:43:51
IN_POBLACION	Job execution finished	Success		file:///F:/Mario/PRA2/IN_POBLACI...	2	2021/05/18 13:43:51
IN_MOVILIDAD	Start of job execution		Followed link after success	file:///F:/Mario/PRA2/IN_MOVILID...		2021/05/18 13:43:52
IN_MOVILIDAD	Job execution finished	Success		file:///F:/Mario/PRA2/IN_MOVILID...	3	2021/05/18 13:43:52
IN_AGLOMERACION	Start of job execution		Followed link after success	file:///F:/Mario/PRA2/IN_AGLOME...		2021/05/18 13:43:53
IN_AGLOMERACION	Job execution finished	Success		file:///F:/Mario/PRA2/IN_AGLOME...	4	2021/05/18 13:43:53
IN_LLAMADAS112	Start of job execution		Followed link after success	file:///F:/Mario/PRA2/IN_LLAMAD...		2021/05/18 13:43:53
IN_LLAMADAS112	Job execution finished	Success		file:///F:/Mario/PRA2/IN_LLAMAD...	5	2021/05/18 14:19:24
IN_FECHAS	Start of job execution		Followed link after success	file:///F:/Mario/PRA2/IN_FECHAS...		2021/05/18 14:19:24
IN_FECHAS	Job execution finished	Success		file:///F:/Mario/PRA2/IN_FECHAS...	6	2021/05/18 14:19:25
IN_AMBITO_GEOGRAFICO	Start of job execution		Followed link after success	file:///F:/Mario/PRA2/IN_AMBITO_...		2021/05/18 14:19:25
IN_AMBITO_GEOGRAFICO	Job execution finished	Success		file:///F:/Mario/PRA2/IN_AMBITO_...	7	2021/05/18 14:19:28
Success	Start of job execution		Followed link after success			2021/05/18 14:19:28
Success	Job execution finished	Success			7	2021/05/18 14:19:28
Job: JOB_IN	Job execution finished	Success	finished		7	2021/05/18 14:19:28

Ilustración 184 - Métricas JOB_IN.

4.2. JOB_TR_DIM

El trabajo "JOB_TR_DIM" procesa todas las transformaciones relativas al almacenamiento en las dimensiones.

El diseño de este trabajo es el siguiente:



Ilustración 185 - JOB_TR_DIM.

Los pasos necesarios para cumplimentar dicho trabajo son:

- Start: nos indica el comienzo del trabajo.
- TR_DIM_GRUPO_EDAD: transformación de TR_DIM_GRUPO_EDAD, tiene en total 7 registros.

✖ Transformation

Entry Name:

TR_DIM_GRUPO_EDAD

Transformation:

\${Internal.Entry.Current.Directory}/TR_DIM_GRUPO_EDAD.ktr

Browse...

Ilustración 186 - Transformación TR_DIM_GRUPO_EDAD.

- TR_DIM_MEDICION: transformación de TR_DIM_MEDICION, tiene en total 4 registros.

✖ Transformation

Entry Name:

TR_DIM_MEDICION

Transformation:

\${Internal.Entry.Current.Directory}/TR_DIM_MEDICION.ktr

Browse...

Ilustración 187 - Transformación TR_DIM_MEDICION.

- TR_DIM_TIPOLOGIA: transformación de TR_DIM_TIPOLOGIA, tiene un total de 10 registros.

✖ Transformation

Entry Name:

TR_DIM_TIPOLOGIA

Transformation:

\${Internal.Entry.Current.Directory}/TR_DIM_TIPOLOGIA.ktr

Browse...

Ilustración 188 - Transformación TR_DIM_TIPOLOGIA.

- TR_DIM_AMBITO_GEOGRAFICO: transformación de TR_DIM_AMBITO_GEOGRAFICO, tiene en total 1022 registros.

✖ Transformation

Entry Name:

TR_DIM_AMBITO_GEOGRAFICO

Transformation:

#{Internal.Entry.Current.Directory}/TR_DIM_AMBITO_GEOGRAF... Browse...

Ilustración 189 - Transformación TR_DIM_AMBITO_GEOGRAFICO.

- TR_DIM_FECHA: transformación de TR_DIM_FECHA, tiene 170 registros.

✖ Transformation

Entry Name:

TR_DIM_FECHA

Transformation:

#{Internal.Entry.Current.Directory}/TR_DIM_FECHA.ktr Browse...

Ilustración 190 - Transformación TR_DIM_FECHA.

- Success: nos indica que el trabajo ha terminado.

Si ejecutamos el anterior trabajo obtenemos las siguientes métricas:

Execution Results

Job / Job Entry	Comment	Result	Reason	Filename	Nr	Log date
Job: JOB_TR_DIM	Start of job execution		start			2021/05/18 16:48:57
Start	Start of job execution		start			2021/05/18 16:48:57
Start	Job execution finished	Success			0	2021/05/18 16:48:57
TR_DIM_GRUPO_EDAD	Start of job execution		Followed unconditional link	file:///F:/Mario/PRA2/TR_DIM_GR...		2021/05/18 16:48:57
TR_DIM_GRUPO_EDAD	Job execution finished	Success		file:///F:/Mario/PRA2/TR_DIM_GR...	1	2021/05/18 16:48:57
TR_DIM_MEDICION	Start of job execution		Followed link after success	file:///F:/Mario/PRA2/TR_DIM_ME...		2021/05/18 16:48:57
TR_DIM_MEDICION	Job execution finished	Success		file:///F:/Mario/PRA2/TR_DIM_ME...	2	2021/05/18 16:48:58
TR_DIM_TIPOLOGIA	Start of job execution		Followed link after success	file:///F:/Mario/PRA2/TR_DIM_TIP...		2021/05/18 16:48:58
TR_DIM_TIPOLOGIA	Job execution finished	Success		file:///F:/Mario/PRA2/TR_DIM_TIP...	3	2021/05/18 16:48:58
TR_DIM_AMBITO_GEOGRAFIC	Start of job execution		Followed link after success	file:///F:/Mario/PRA2/TR_DIM_AM...		2021/05/18 16:48:58
TR_DIM_AMBITO_GEOGRAFIC	Job execution finished	Success		file:///F:/Mario/PRA2/TR_DIM_AM...	4	2021/05/18 16:48:59
TR_DIM_FECHA	Start of job execution		Followed link after success	file:///F:/Mario/PRA2/TR_DIM_FEC...		2021/05/18 16:48:59
TR_DIM_FECHA	Job execution finished	Success		file:///F:/Mario/PRA2/TR_DIM_FEC...	5	2021/05/18 16:48:59
Success	Start of job execution		Followed link after success			2021/05/18 16:48:59
Success	Job execution finished	Success			5	2021/05/18 16:48:59
Job: JOB_TR_DIM	Job execution finished	Success	finished		5	2021/05/18 16:48:59

Ilustración 191 - Métricas JOB_TR_DIM.

4.3. JOB_TR_FACT

El trabajo “JOB_TR_FACT” procesa todas las transformaciones relativas al almacenamiento de las tablas de los hechos.

El diseño de este trabajo es el siguiente:



Ilustración 192 - JOB_TR_FACT.

Los pasos para cumplimentar con dicho trabajo son:

- Start: nos indica que el trabajo comienza.
- TR_FACT_LLAMADAS112: transformación de TR_FACT_LLAMADAS112, tiene un total de 340307 registros.

✖ Transformation

Entry Name:

TR_FACT_LLAMADAS112

Transformation:

\${Internal.Entry.Current.Directory}/TR_FACT_LLAMADAS112.ktr

Browse...

Ilustración 193 - Transformación TR_FACT_LLAMADAS112.

- TR_FACT_MEDICIONES: transformación de TR_FACT_MEDICIONES, tiene un total de 5303 registros.

✖ Transformation

Entry Name:

TR_FACT_MEDICIONES

Transformation:

\${Internal.Entry.Current.Directory}/TR_FACT_MEDICIONES.ktr

Browse...

Ilustración 194 - Transformación TR_FACT_MEDICIONES.

- Success: nos indica que el trabajo ha terminado.

Del anterior trabajo obtenemos las siguientes métricas:

Execution Results						
Logging History Job metrics Metrics						
Job / Job Entry	Comment	Result	Reason	Filename	Nr	Log date
Job: JOB_TR_FACT						
Job: JOB_TR_FACT	Start of job execution		start			2021/05/18 16:51:39
Start	Start of job execution		start			2021/05/18 16:51:39
Start	Job execution finished	Success			0	2021/05/18 16:51:39
TR_FACT_LLAMADAS112	Start of job execution		Followed unconditional link	file:///F:/Mario/PRA2/TR_FACT_LL...		2021/05/18 16:51:39
TR_FACT_LLAMADAS112	Job execution finished	Success		file:///F:/Mario/PRA2/TR_FACT_LL...	1	2021/05/18 16:54:13
TR_FACT_MEDICIONES	Start of job execution		Followed link after success	file:///F:/Mario/PRA2/TR_FACT_M...		2021/05/18 16:54:13
TR_FACT_MEDICIONES	Job execution finished	Success		file:///F:/Mario/PRA2/TR_FACT_M...	2	2021/05/18 16:54:16
Success	Start of job execution		Followed link after success			2021/05/18 16:54:16
Success	Job execution finished	Success			2	2021/05/18 16:54:16
Job: JOB_TR_FACT	Job execution finished	Success	finished		2	2021/05/18 16:54:16

Ilustración 195 - Métricas JOB_TR_FACT.

4.4. JOB

El trabajo “JOB” es el encargado de ejecutar todos los trabajos mencionados anteriormente.

El diseño de este trabajo es el siguiente:

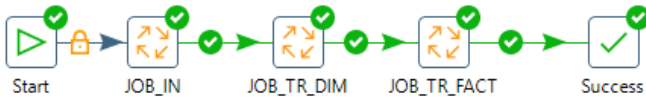


Ilustración 196 - JOB .

Los pasos que realiza este trabajo son:

- Start: comienza el trabajo.
- JOB_IN: trabajo de JOB_IN.



Entry Name:

JOB_IN


Job:

\${Internal.Entry.Current.Directory}/JOB_IN.kjb

Browse...

Ilustración 197 - Trabajo JOB_IN.

- JOB_TR_DIM: trabajo de JOB_TR_DIM.

 Job

Entry Name:

JOB_TR_DIM


Job:

#{Internal.Entry.Current.Directory}/JOB_TR_DIM.kjb

Browse...

Ilustración 198 - Trabajo JOB_TR_DIM.

- JOB_TR_FACT: trabajo de JOB_TR_FACT.

 Job

Entry Name:

JOB_TR_FACT

Job:

#{Internal.Entry.Current.Directory}/JOB_TR_FACT.kjb

Browse...

Ilustración 199 - Trabajo JOB_TR_FACT.

- Success: indica que el trabajo ha finalizado.

De la anterior ejecución obtenemos las siguientes métricas:

Execution Results						
Logging History Job metrics Metrics						
Job / Job Entry	Comment	Result	Reason	Filename	Nr	Log date
Job: JOB	Start of job execution		start			2021/05/18 16:57:39
Start	Start of job execution		start			2021/05/18 16:57:39
Start	Job execution finished	Success			0	2021/05/18 16:57:39
JOB_IN	Start of job execution		Followed unconditional link	file:///F:/Mario/PRA2/JOB_IN.kjb		2021/05/18 16:57:39
Job: JOB_IN						
JOB_IN	Job execution finished	Success		file:///F:/Mario/PRA2/JOB_IN.kjb	1	2021/05/18 17:33:00
JOB_TR_DIM	Start of job execution		Followed link after success	file:///F:/Mario/PRA2/JOB_TR_DIM...		2021/05/18 17:33:00
Job: JOB_TR_DIM						
JOB_TR_DIM	Job execution finished	Success		file:///F:/Mario/PRA2/JOB_TR_DIM...	2	2021/05/18 17:33:01
JOB_TR_FACT	Start of job execution		Followed link after success	file:///F:/Mario/PRA2/JOB_TR_FAC...		2021/05/18 17:33:01
Job: JOB_TR_FACT						
JOB_TR_FACT	Job execution finished	Success		file:///F:/Mario/PRA2/JOB_TR_FAC...	3	2021/05/18 17:35:35
Success	Start of job execution		Followed link after success			2021/05/18 17:35:35
Success	Job execution finished	Success			3	2021/05/18 17:35:35
Job: JOB	Job execution finished	Success	finished		3	2021/05/18 17:35:35

Ilustración 200 - Métricas JOB.