

Se quiere realizar un estudio del padrón de una provincia de España, puedes elegir la provincia que desees.

- Datos del padrón por provincias: [INE \(Instituto Nacional de Estadística\)](#)
 - En este fichero hay filas de datos con totales que debes eliminar. Revisa la estructura y el contenido de los datos
 - NOTA: Descargar los ficheros en formato CSV
- Otros datos ya preparados obtenidos del INE y del CNIG (Centro Nacional de Información Geográfica).
 - Se pueden descargar de <https://doi.org/10.6084/m9.figshare.13505136>.
 - Un fichero con codificación de los rangos de habitantes
 - Un fichero con codificación de municipios

Tabla DatosCenso

- 1 de 7

Actividad:							
Pentaho Data Integration (PDI). Ejemplo INE							
Ciclo:	IABD	Modulo:	BDA	Curso:		Agrupación:	1
Alumno/a:						Grupo:	

Los pasos y acciones que debes realizar son:

1. Revisión y análisis de los datos de entrada. Recuerda que hay datos que sobran
2. Elabora un plan de trabajo definiendo jobs y transformaciones. Dibújalo si te facilita el trabajo.
3. Analiza las comprobaciones previas antes de empezar. Conexiones a BBDD, existencia de ficheros, etc... En caso de fallo en alguna de ellas se deberá notificar mediante correo electrónico al desarrollador de la ETL.
4. Incluye una tabla a modo de log en la que se vayan registrando los procesos que se realizan. La estructura de la tabla debe ser algo así:
 - Fecha y Hora (Timestamp)
 - Proceso
 - Estado (Inicio, Fin)
5. Ingesta de los datos y transformación
6. Carga de los datos
 - Carga los datos en la tabla con la estructura definida al inicio del enunciado de la actividad

Ten en cuenta que:

- Deberás crear tantas transformaciones y jobs como sean necesarios
- Crea etiquetas para facilitar la lectura de la ETL
- Definir una estructura de jobs y transformaciones clasificada y organizada
- Deberás realizar comprobaciones para evitar fallos en el proceso.

Entrega de la práctica:

- Los jobs y transformaciones
- Sitúa los datos originales, los jobs y las transformaciones de Pentaho en la misma carpeta, para facilitar la corrección
- Un volcado de la BD (utiliza una solo), para facilitar la corrección
- Un ejemplo de los 10 primeros registros de la tabla resultante.

CIPFP www.fpmislata.com							
Actividad:	Pentaho Data Integration (PDI). Ejemplo INE						
Ciclo:	IABD	Modulo:	BDA	Curso:		Agrupación:	1
Alumno/a:						Grupo:	

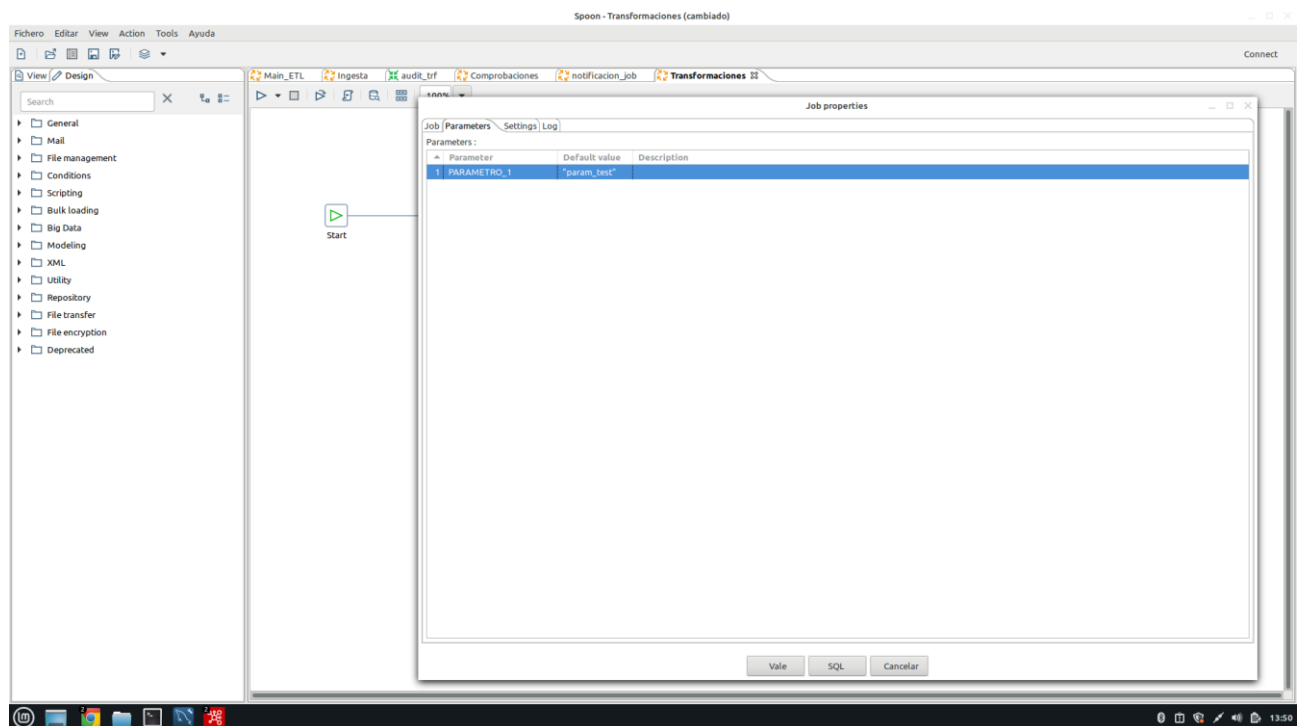
Para facilitaros un poco la tarea de notificar y auditar la ETL, os dejamos unos consejos que os pueden servir de ayuda.

Cómo pasar un parámetro de un job a una transformación. Utilizamos la actividad03 como ejemplo.

1.- Empezamos por definir unos parámetros globales al job desde el menú:

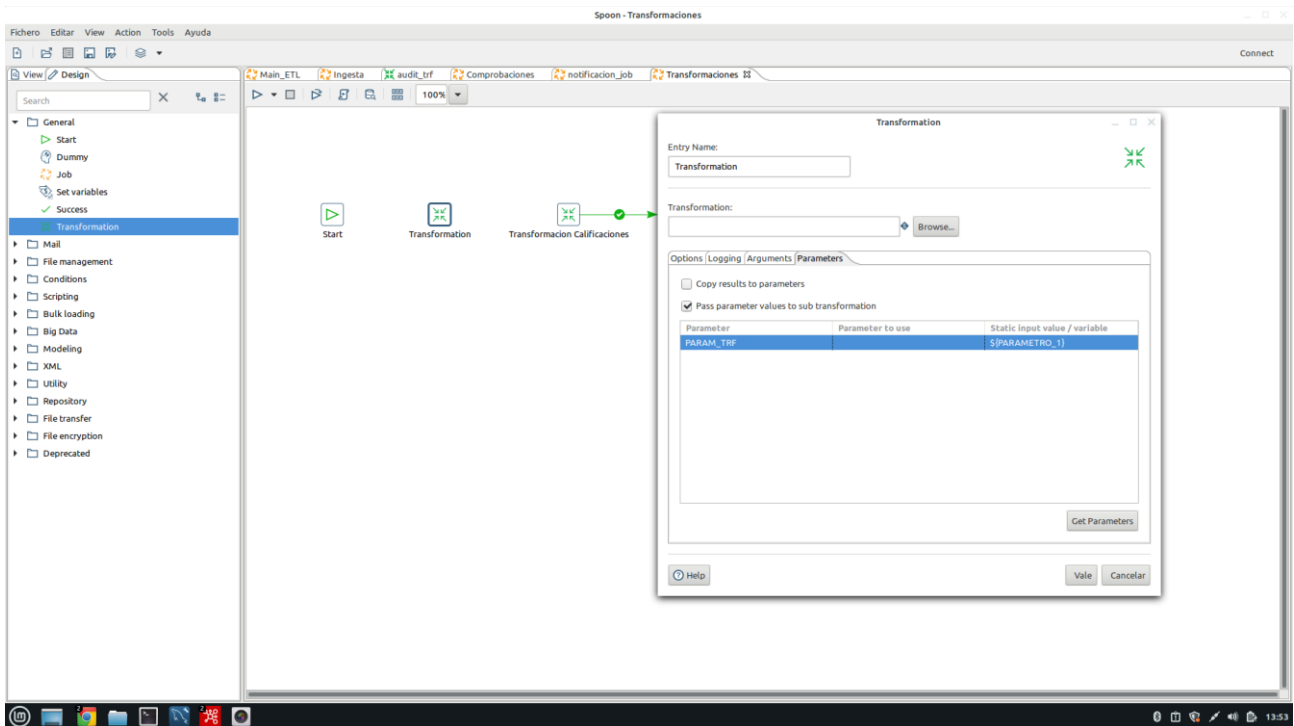
Editar ---> Configuración ---> Parámetros

Añadimos un nuevo parámetro llamado PARAMETRO_1



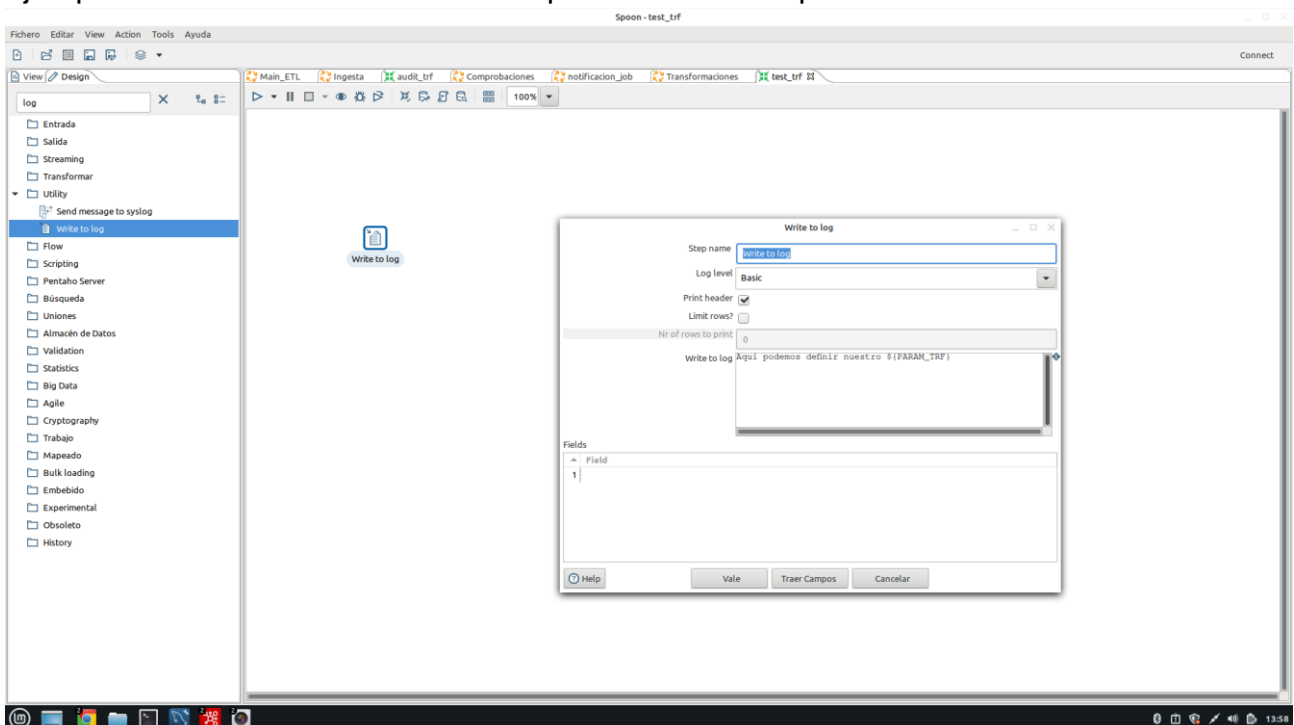
2.- Desde el job vamos a añadir una transformación, que posteriormente implementaremos, y le configuramos este parámetro para que se lo pase a la transformación que definamos.

Actividad:	Pentaho Data Integration (PDI). Ejemplo INE						
Ciclo:	IABD	Modulo:	BDA	Curso:		Agrupación:	1
Alumno/a:						Grupo:	



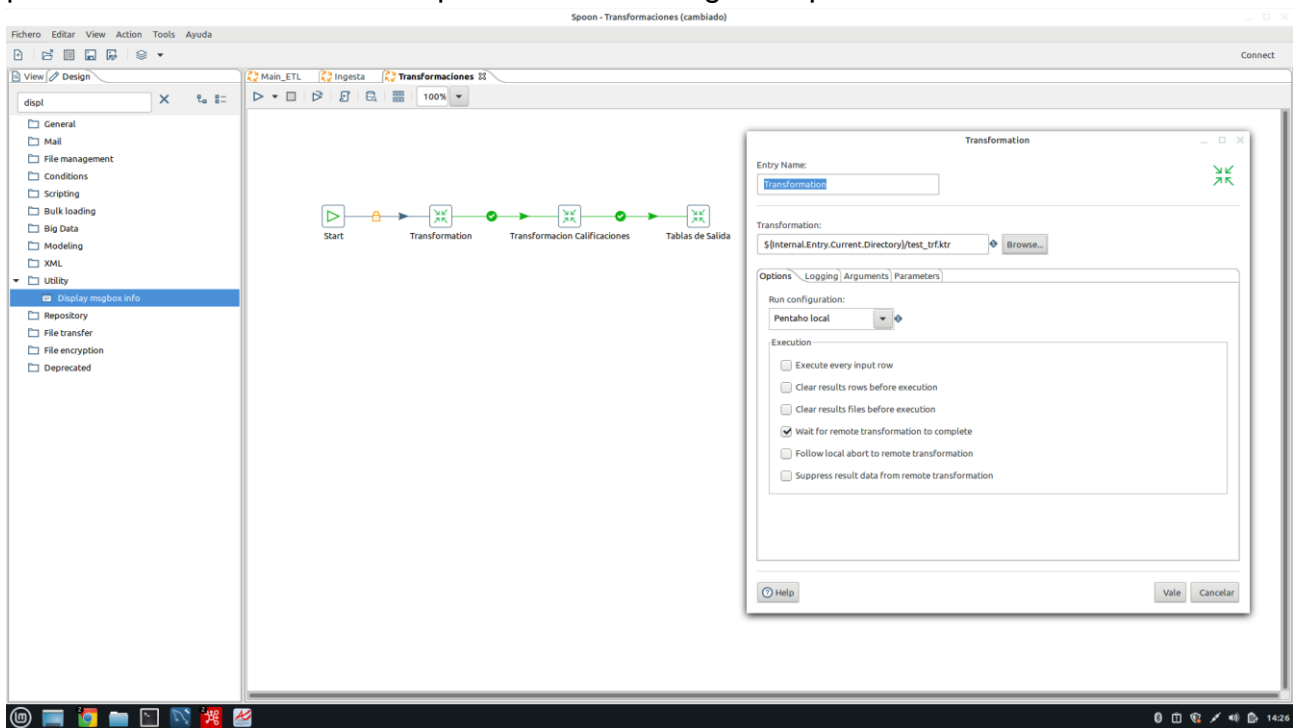
Fijaros que el parámetro de entrada sería `${PARAMETRO_1}` y que el parámetro de salida que navegaría hacia la transformación será `${PARAM_TRF}`

3.- Ahora añadimos la transformación y simplemente le agregamos una escritura a log como ejemplo de la lectura de esa variable que acabamos de pasarle.



CIPFP www.fpmislata.com							
Actividad:	Pentaho Data Integration (PDI). Ejemplo INE						
Ciclo:	IABD	Modulo:	BDA	Curso:		Agrupación:	1
Alumno/a:						Grupo:	

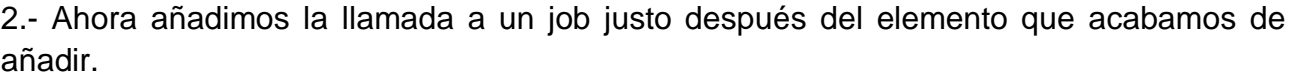
4.- Este paso sería el último, donde enlazamos la llamada a la transformación creada en el paso 2 con la transformación que escribe en el log en el paso 3.



Cómo pasar un parámetro de un job a un subjob. Utilizamos la actividad03 como ejemplo.

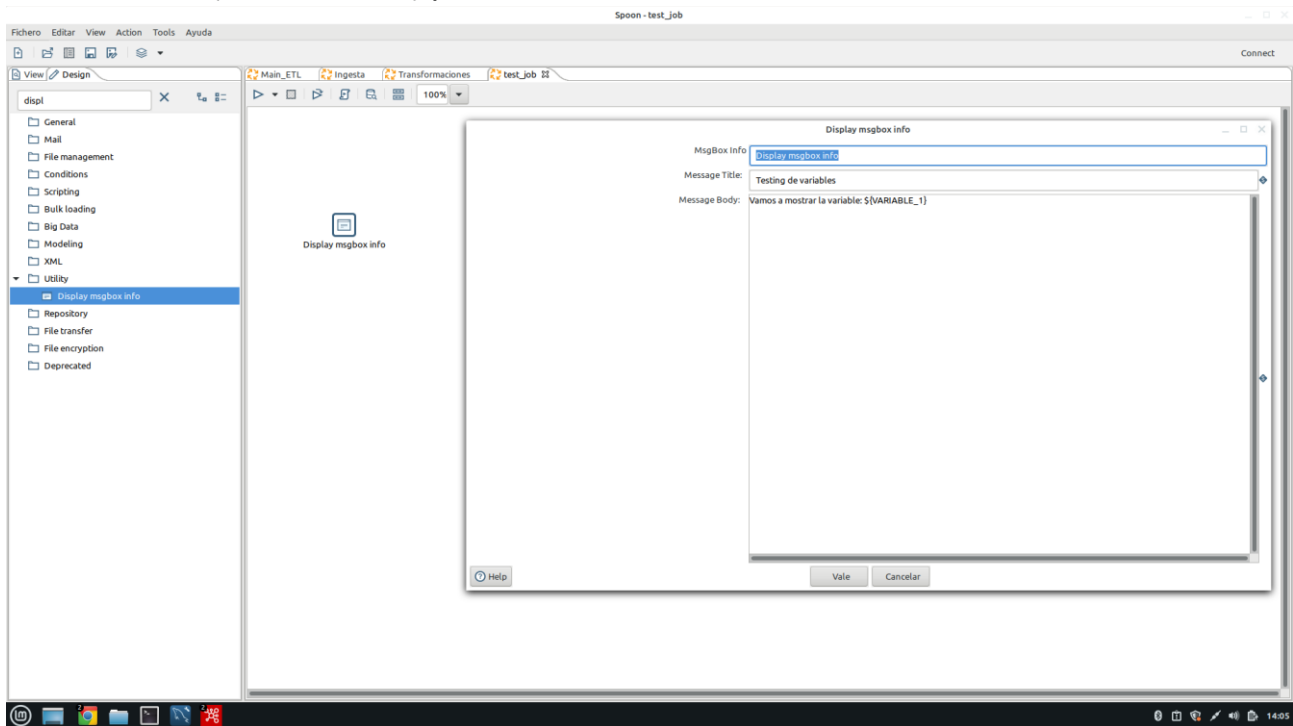
1.- Para este caso nos vamos a centrar en el uso de variables globales. Pensad que si algunas de ellas las tenéis claras pueden ir al principio del todo antes de las comprobaciones (en otro job, por supuesto).

Añadimos una variable nueva llamada VARIABLE_1 utilizando el elemento “Set Variables”



CIPFP www.fpmislata.com							
Actividad:	Pentaho Data Integration (PDI). Ejemplo INE						
Ciclo:	IABD	Modulo:	BDA	Curso:		Agrupación:	1
Alumno/a:						Grupo:	

3.- Creamos el job a enlazar con un elemento muy simple llamado “Display msg box” y utilizamos la `#{VARIABLE_1}` para mostrar su contenido.



4.- Este paso sería el último, donde enlazamos la llamada al job creado en el paso 2 con el job que muestra el mensaje del paso 3.

