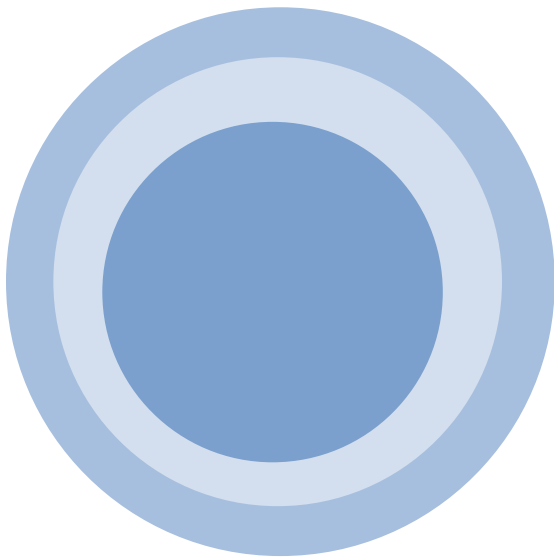


# PROCESO ETL (EXTRACCIÓN TRANSFORMACIÓN Y CARGA)



Autor: ing.Alejandro Ramos

## ➤ REQUISITOS PREVIOS A LA INSTALACIÓN

### Requisitos mínimos de hardware

- Procesador Procesador x64: AMD Opteron, AMD Athlon 64, Intel Xeon compatible con Intel EM64T, Intel Pentium IV compatible con EM64T
- 1 GB de memoria RAM
- Disco Duro con al menos 2 GB libres

### Requisitos de software

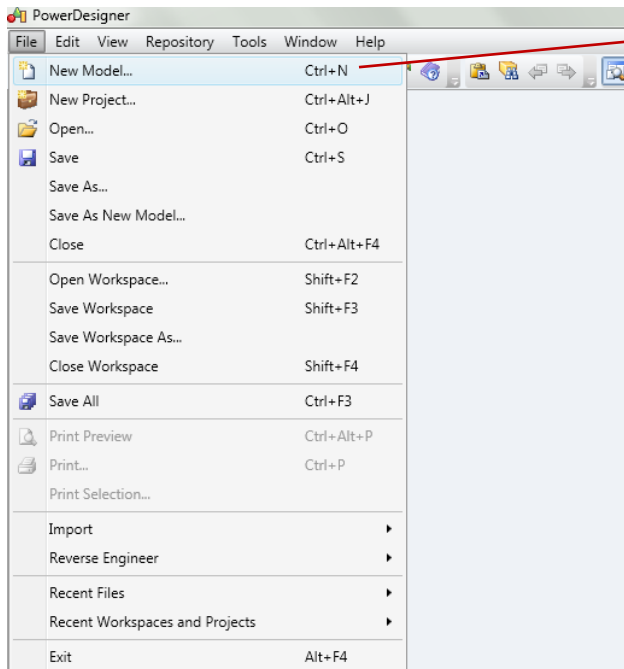
- Java run Time Enviroment 5 o posteriores
- MySQL version 5 o posteriores
- Pentahoo Data Integration(PDI)
- Power Designer

### Sistema Operativo

El sistema operativo que se ocupó para este desarrollo fue Windows, las herramientas tienen soporte para cualquier versión desde Windows 7 en adelante.

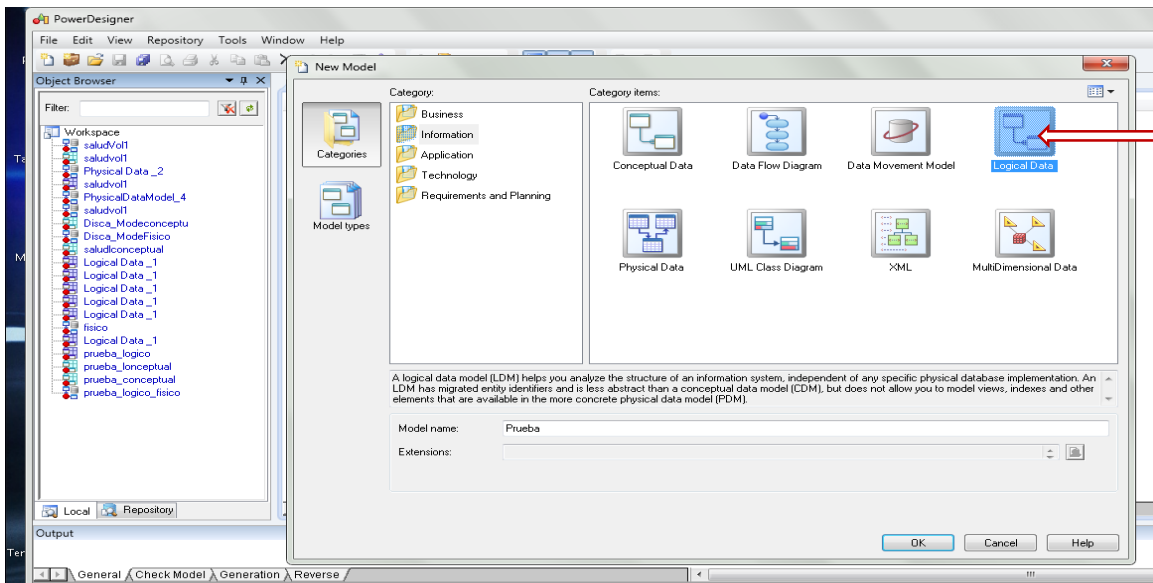
# DISEÑO DE LA BASE DE DATOS CON POWER DESIGNER

Ejecutamos Power Designer

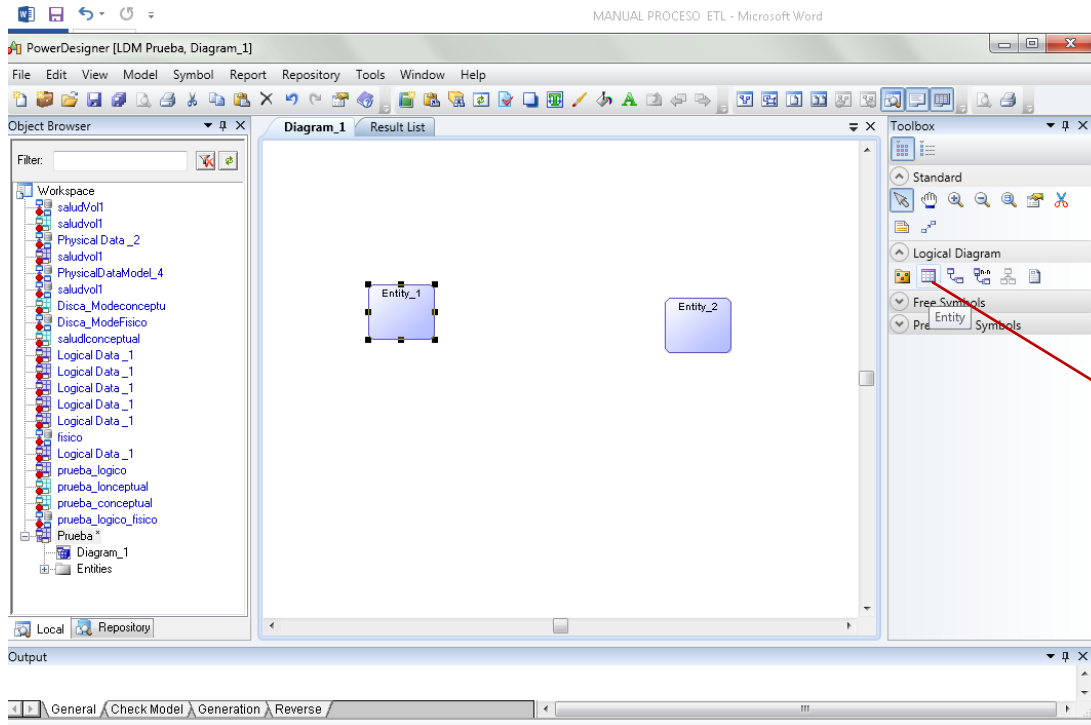


Abrimos File y seleccionamos New Model

A continuación seleccionamos crear el modelo lógico.

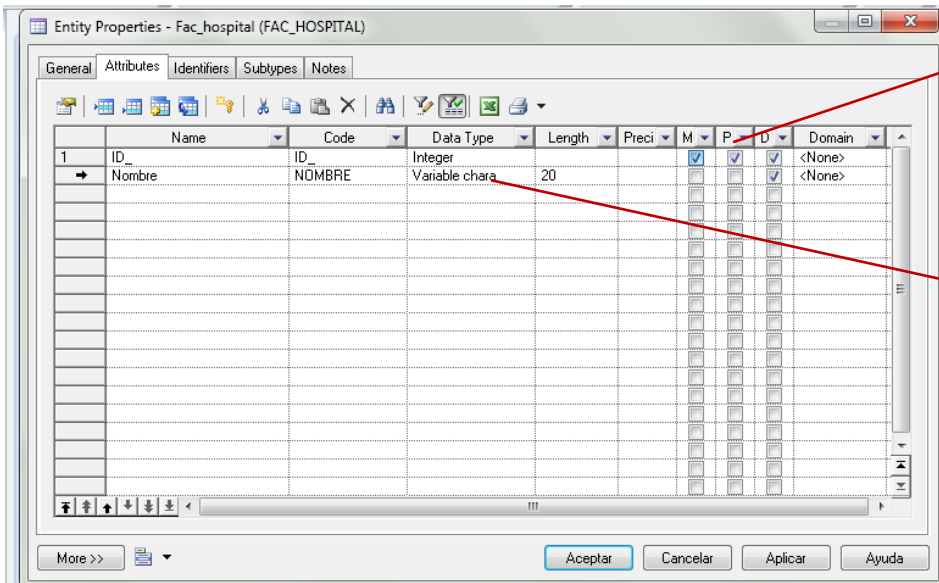


Se nos presenta una nueva ventana donde podemos empezar con el diseño lógico de la base de datos



Seleccionamos del cuadro de herramientas Entity y creamos las diferentes entidades

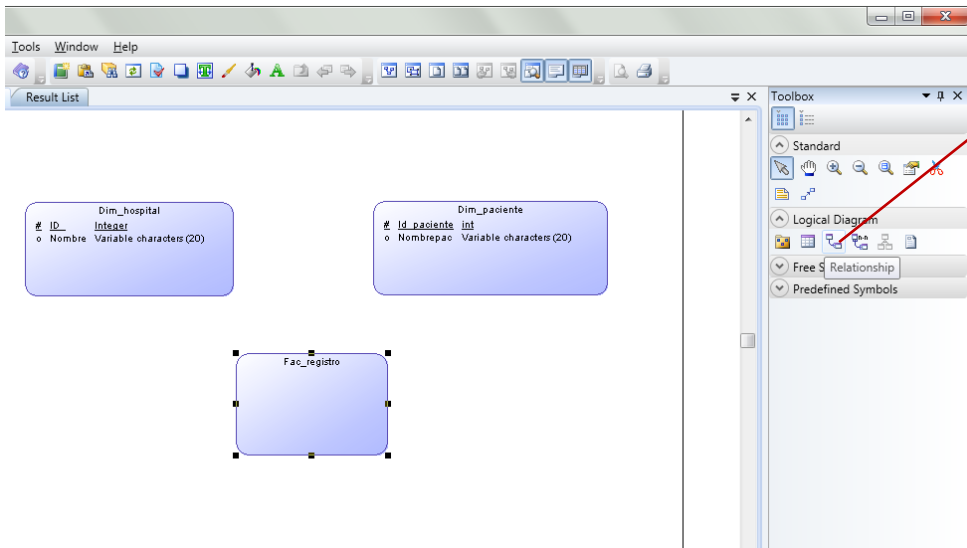
Damos doble click en la entidad para ingresar su nombre y atributos



Definir Clave primaria (PK)

Tipo de variable

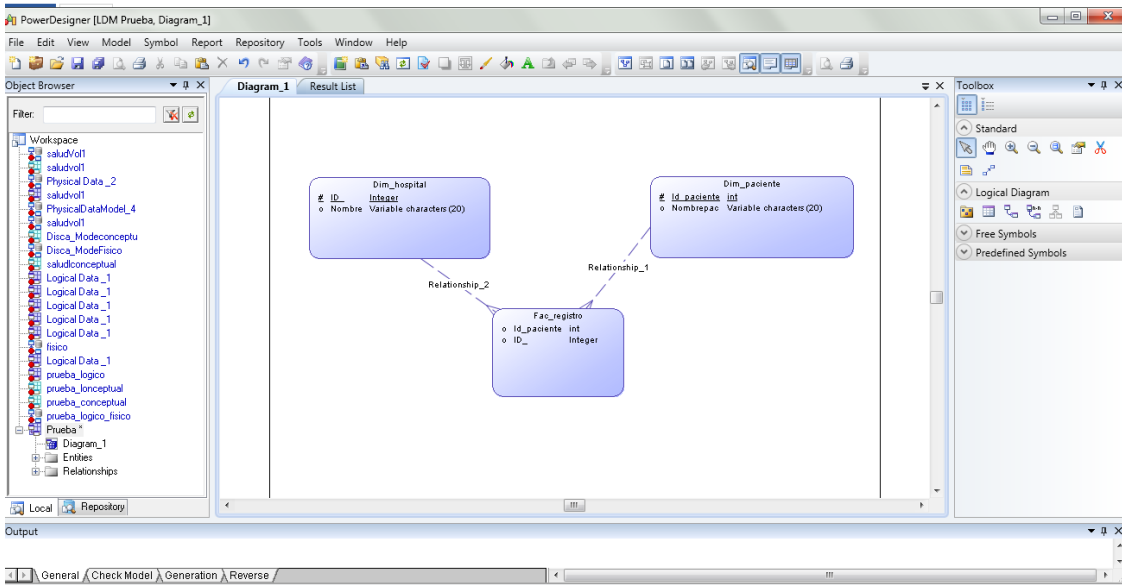
Una vez definimos las entidades y atributos procedemos a generar su relación.



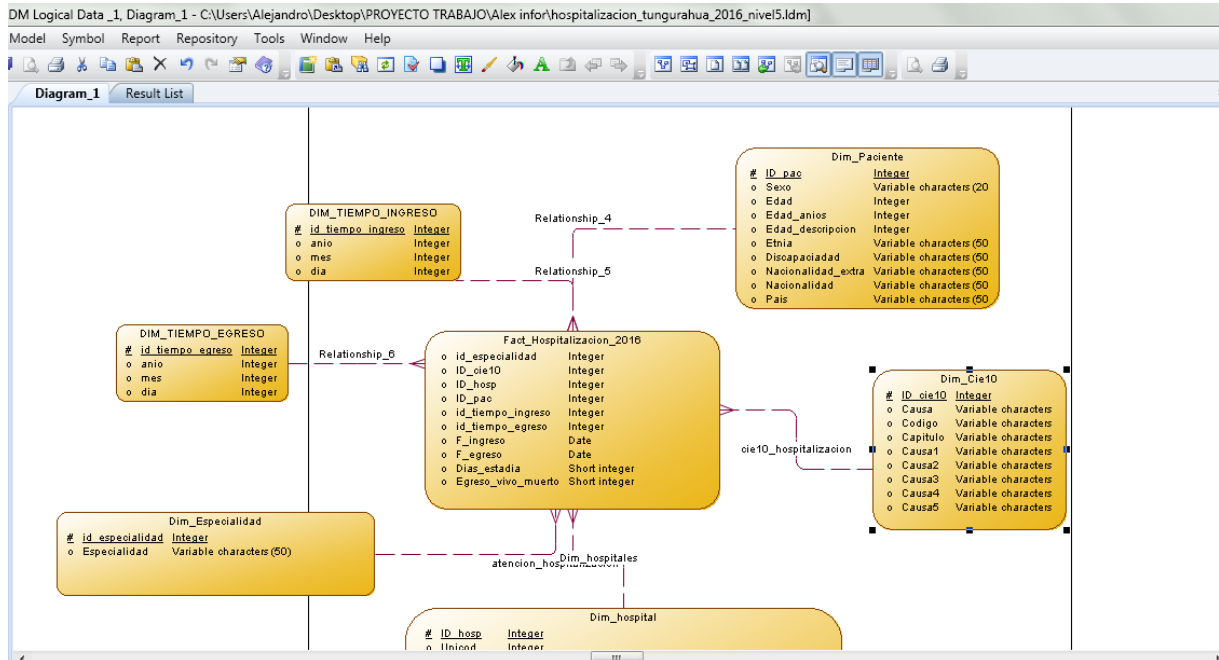
Seleccionamos la opción Relationship

Recordemos que en este caso es una base de datos Dimensional tipo estrella, la misma que cuenta con tablas Dim relacionadas a una tabla de hechos o tabla fac.

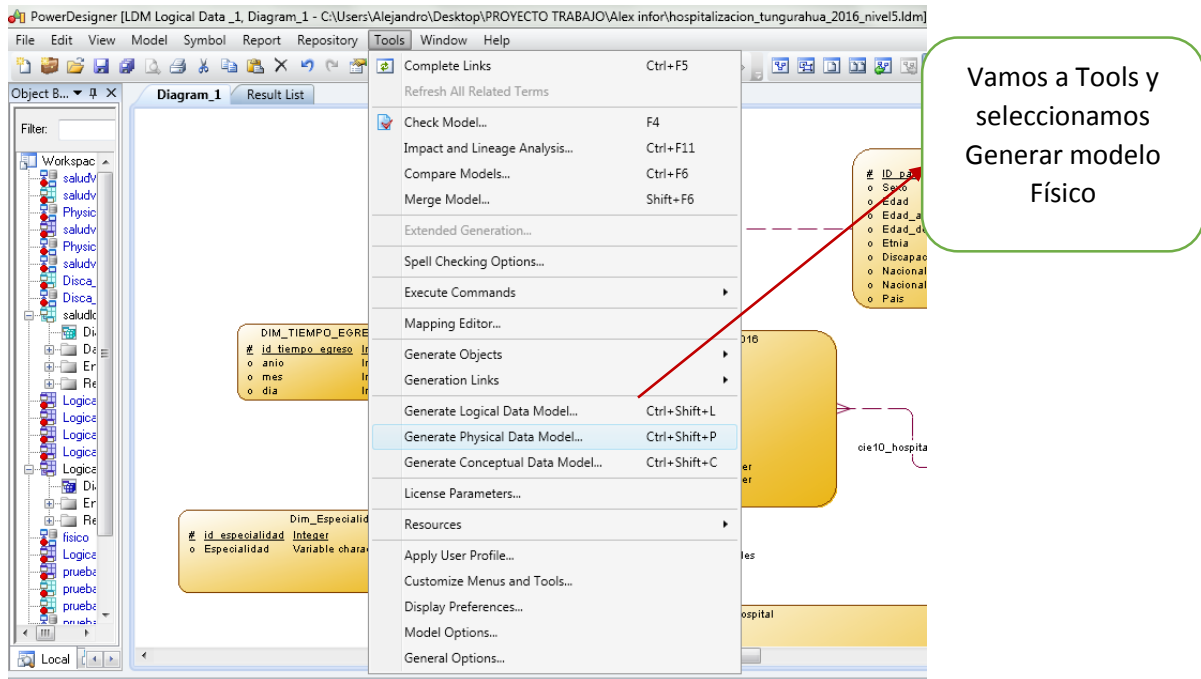
Ahora ya tenemos las relaciones entre las entidades Dim y la tabla Fac o tabla de hechos.



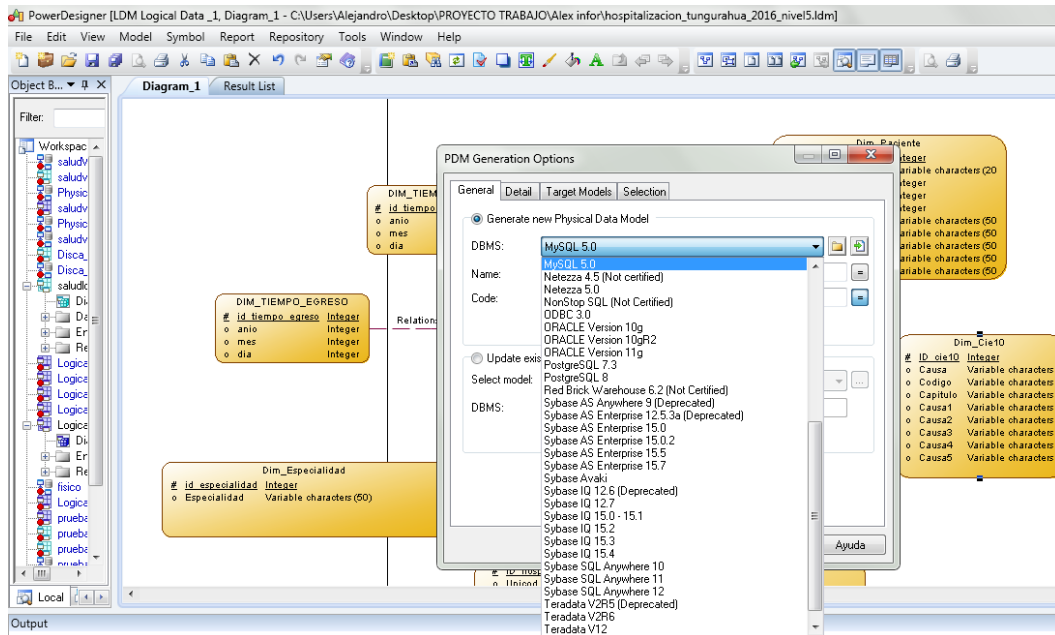
Luego de crear todas las entidades y relaciones obtendremos nuestro diseño de la base de datos dimensional.



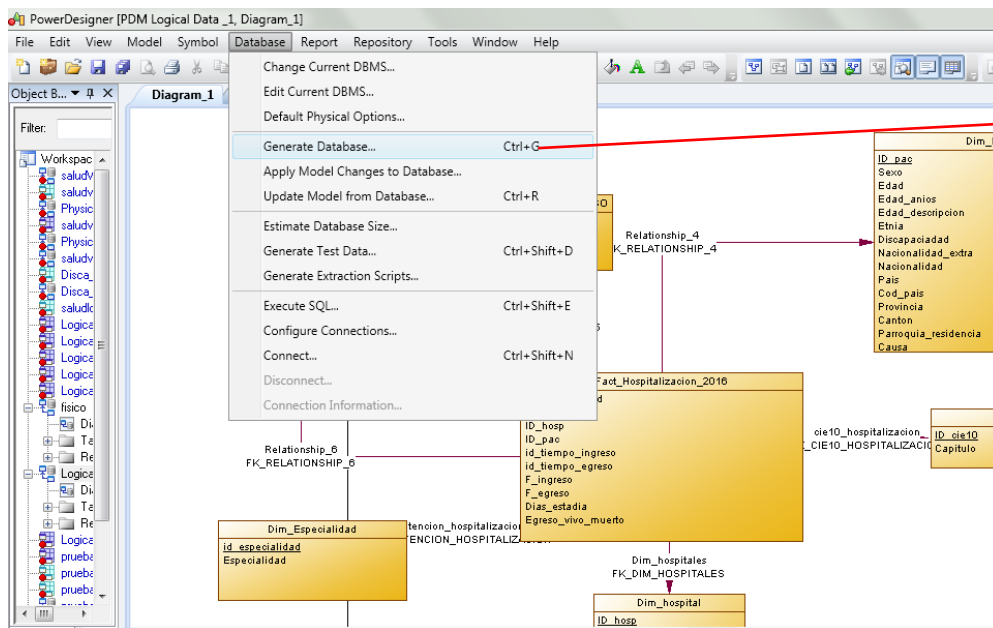
Una vez que tenemos todo el diseño de la base de datos, ya podemos generar el modelo físico.



A continuación escogemos el tipo de BDD, en nuestro caso seleccionamos Mysql 5.0

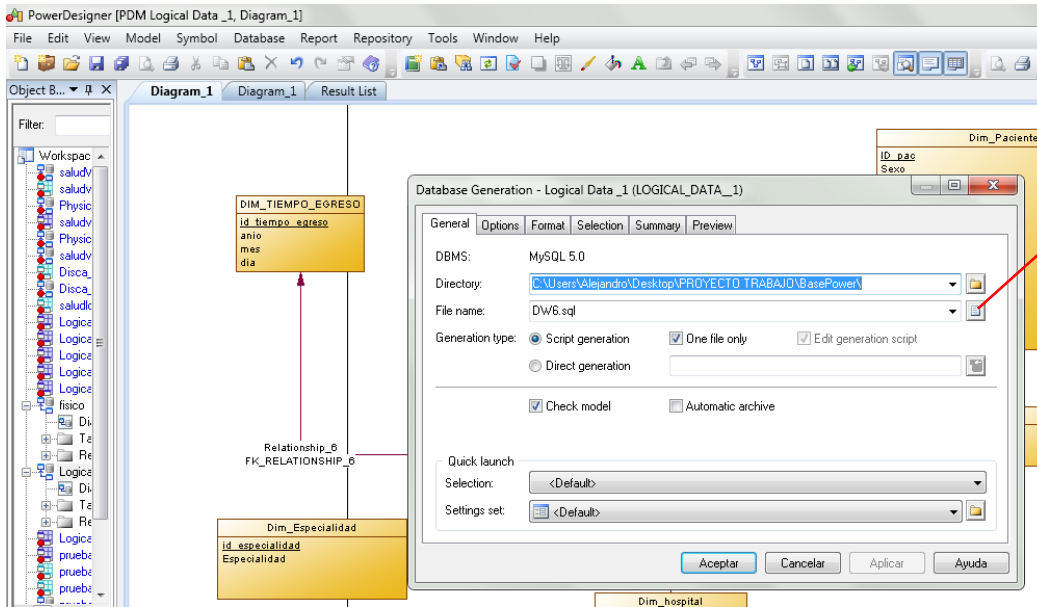


Con el modelo físico ya podemos generar el Script para la BDD



En la pestaña Database, seleccionamos Generate Database

A continuación nos presenta la pantalla de Generate Databases



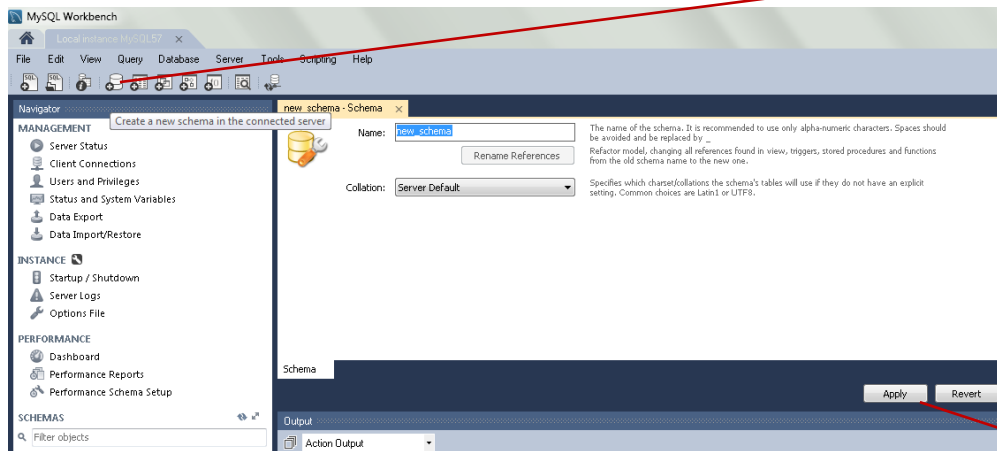
Indicamos la dirección para guardar nuestro Script



## CREACIÓN, BASE DE DATOS EN MYSQL

El gestor de base de datos que usaremos para subir el script anteriormente generado es Mysql Workbench 6.3.

Abrimos Mysql Workbench.



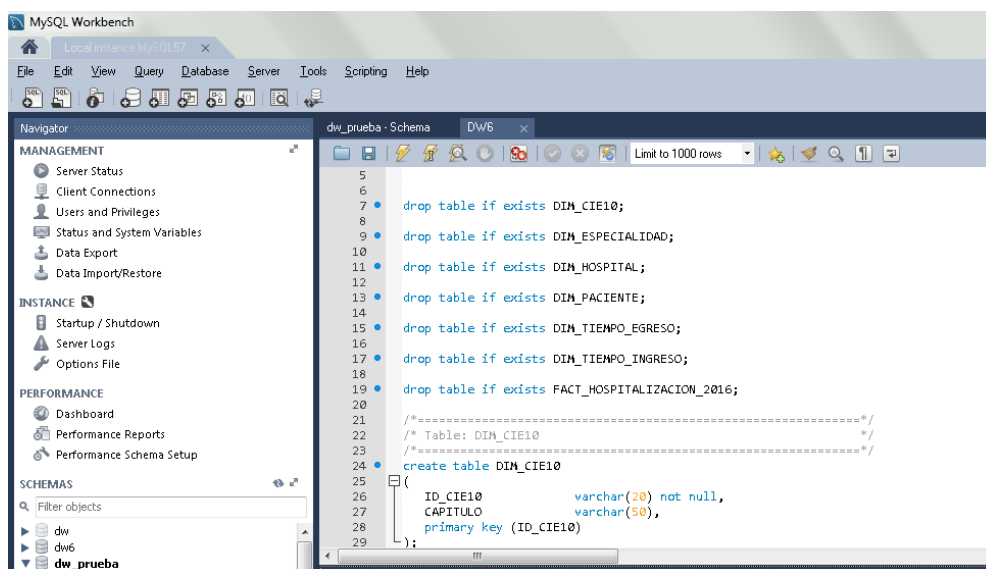
Damos click en la opción Create new schema. Para generar una base de datos.

Click en Apply

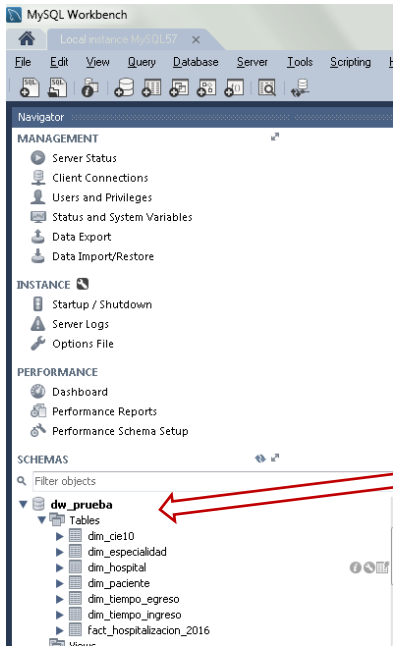
Listo ahora tenemos nuestra base de datos, creada con el nombre dw\_prueba.

En el siguiente proceso tenemos que cargar el archivo con el script generado en el diseño de la base de datos.

En la siguiente imagen observamos el script cargado y damos en ejecutar



Listo ahora tenemos la base de datos cargada con todas las tablas.

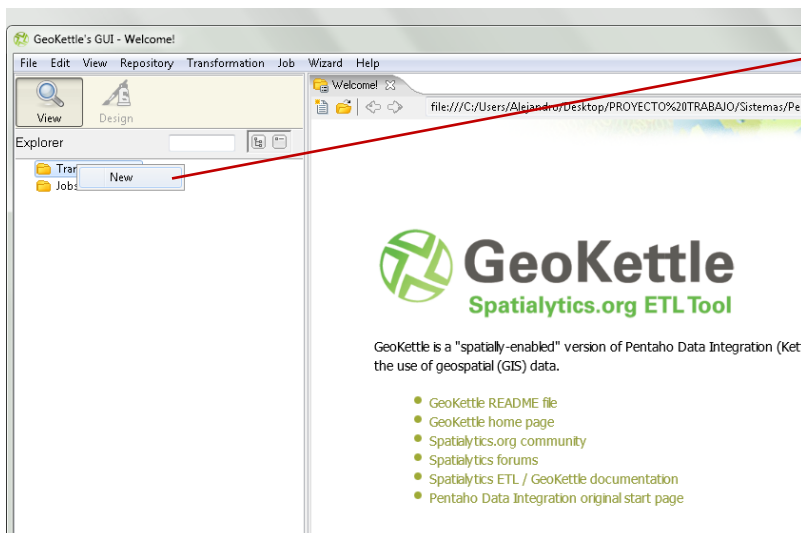


## CARGA DE LA BASE DE DATOS CON PENTAHO

Pentaho es una herramienta de Business Intelligence de software libre, es una plataforma compuesta de diferentes programas. En nuestro caso usaremos Pentaho Data Integration GeoKettle

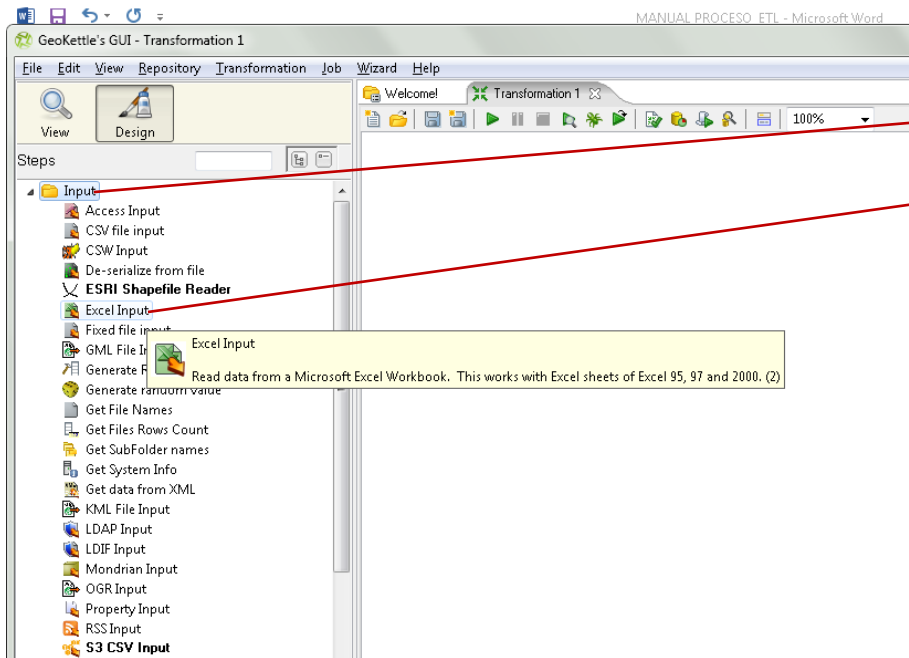
Link de descarga: <http://community.pentaho.com/>

Descomprimos el archivo y ejecutamos



En transformación damos click derecho y damos en NEW

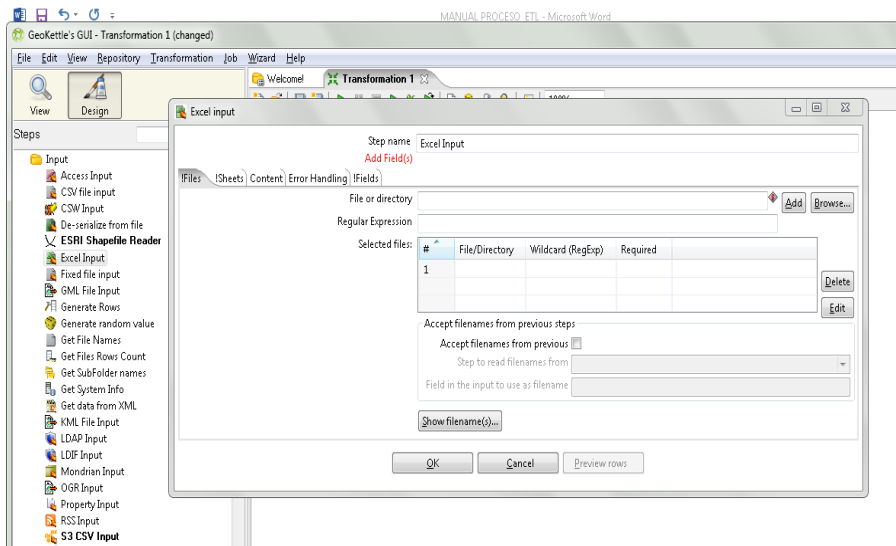
Ahora ya tenemos nuestro espacio para trabajar con nuestra primera transformación.



En la carpeta Input seleccionamos Excel Input y arrastramos a nuestro espacio de trabajo

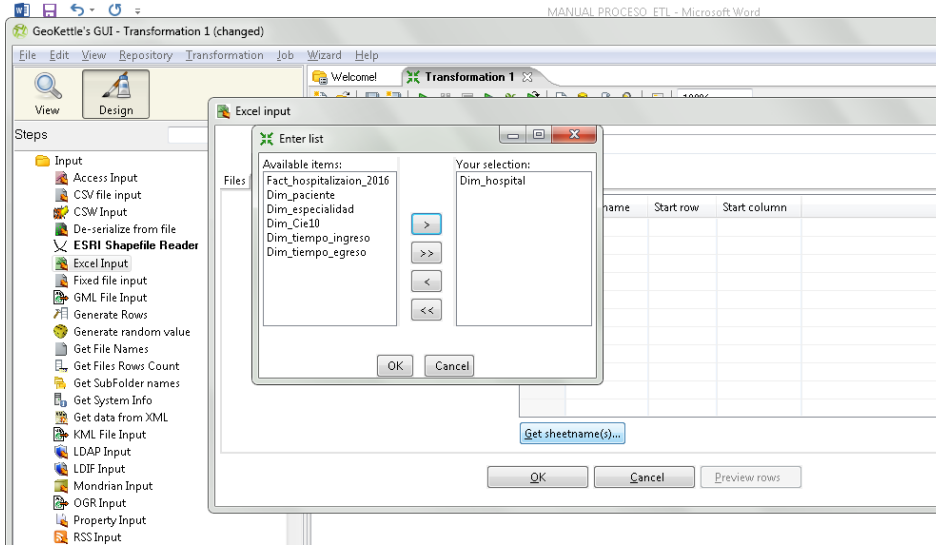


Damos doble click en Excel Input



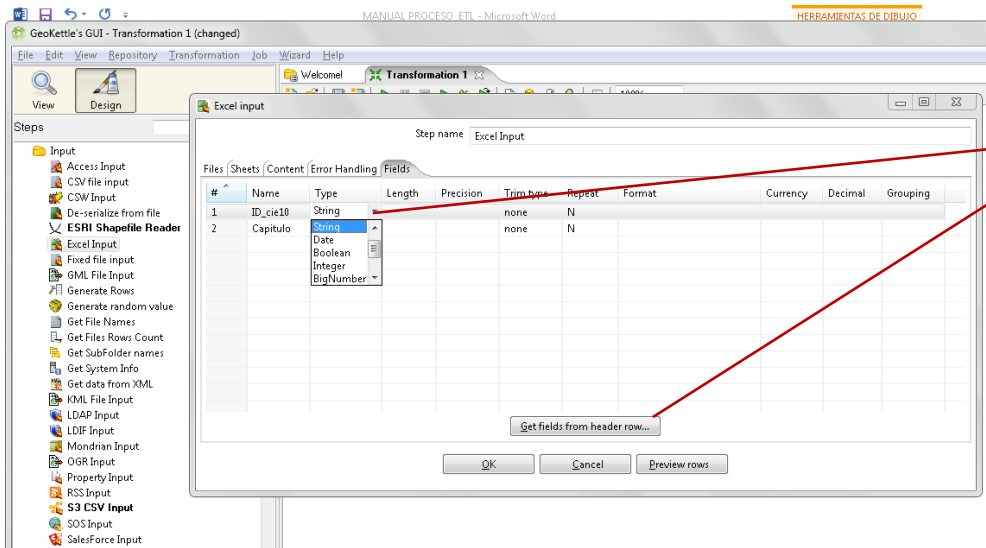
En la opción Browse seleccionamos nuestro archivo en formato xls y añadimos

## Marcamos la siguiente pestaña ¡Sheets



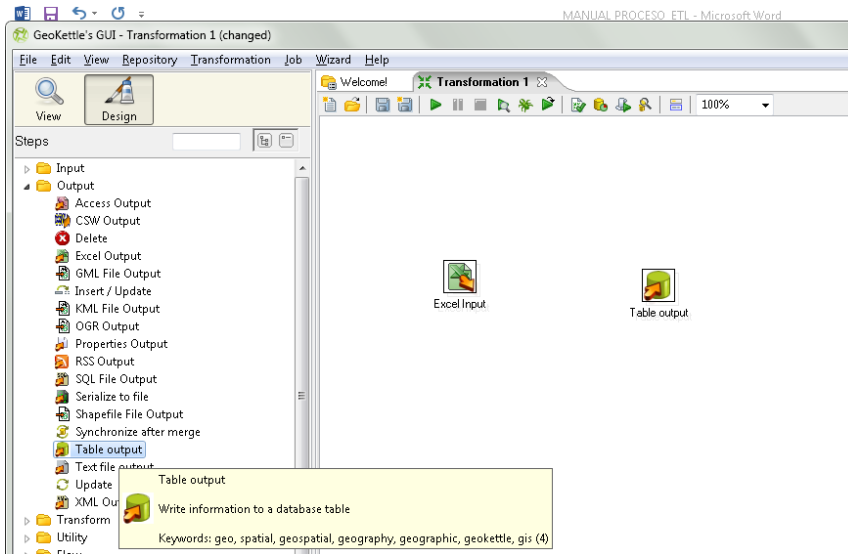
En la opción Get sheetname seleccionamos la hoja de trabajo para la transformación

## En la pestaña Fields cargamos los atributos de la tabla



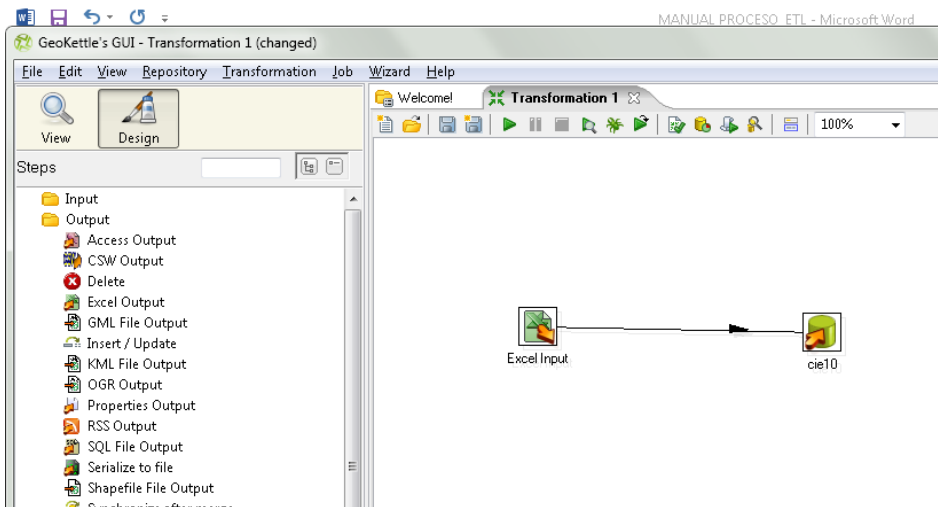
Cargamos los atributos de la tabla y seleccionamos el tipo de valor.


Una vez que tenemos la entrada de los datos, ahora seleccionamos el destino que en este caso será una base de datos.

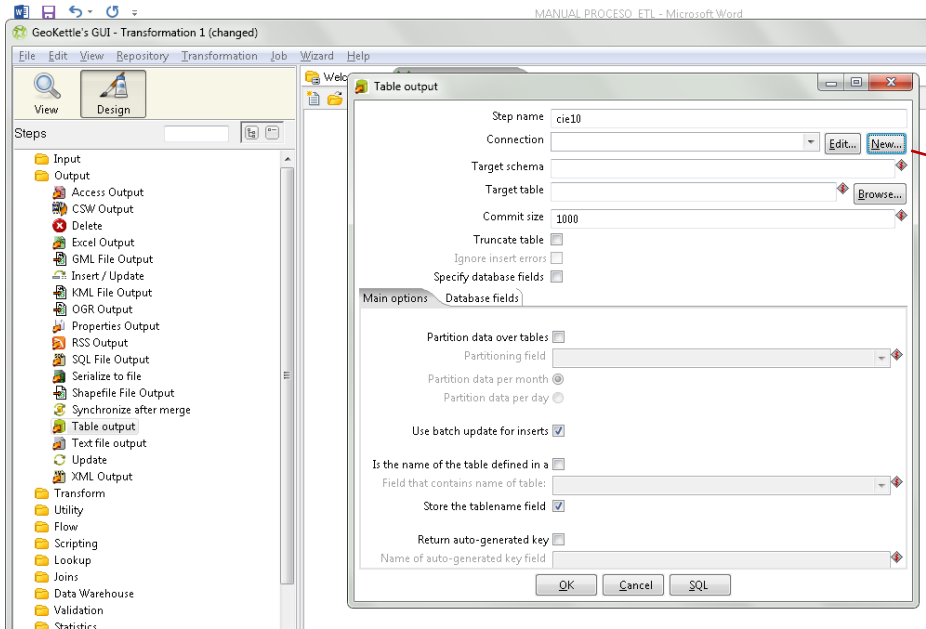


En la carpeta Output seleccionamos Table output y arrastramos a nuestra hoja de trabajo

Creamos la conexión con el archivo de entrada a la base de datos, mantenemos Ctrl y click izquierdo arrastramos a la salida de la bdd

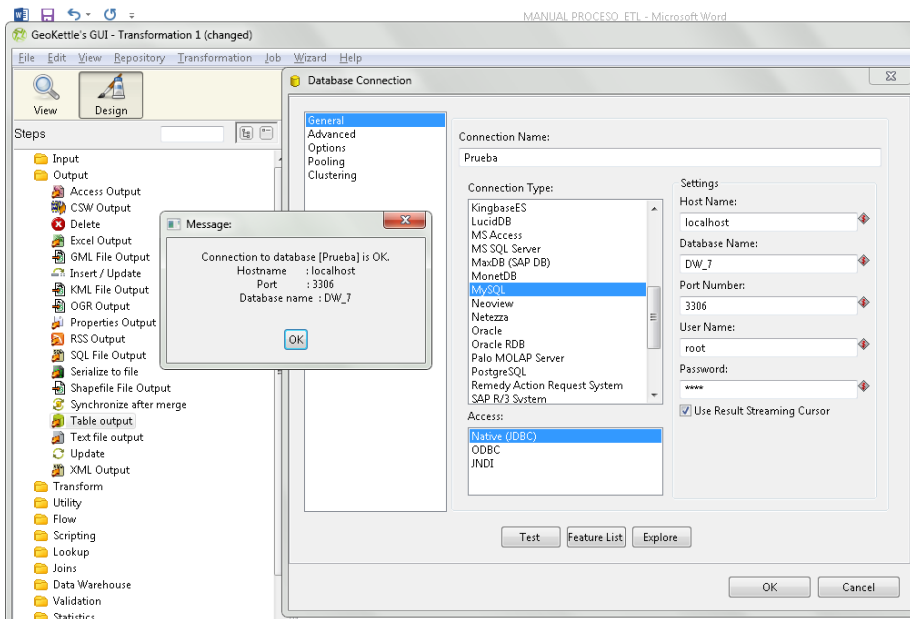


Ingresamos a las opciones  doble click

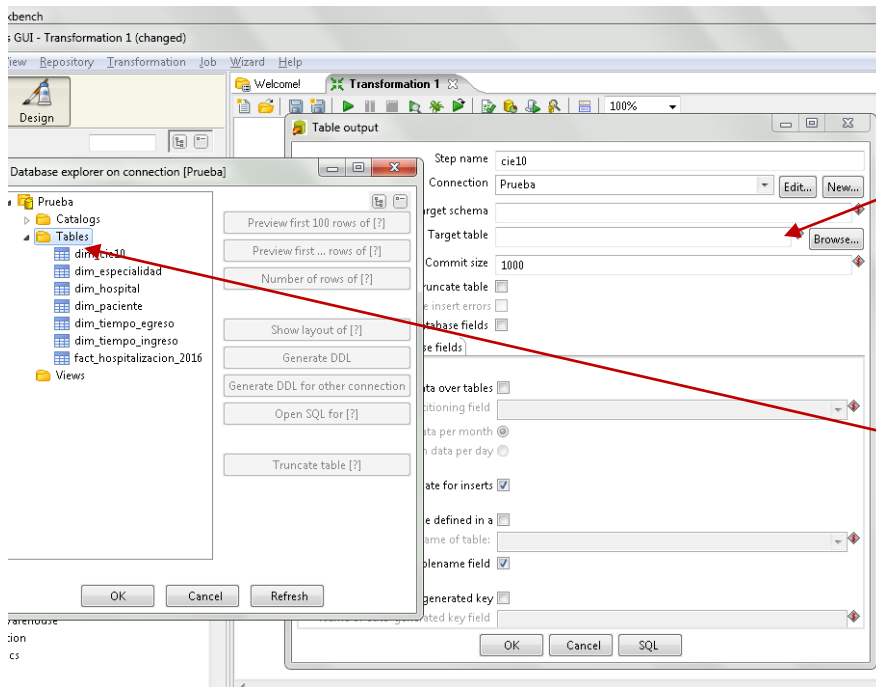


Creamos una nueva conexión

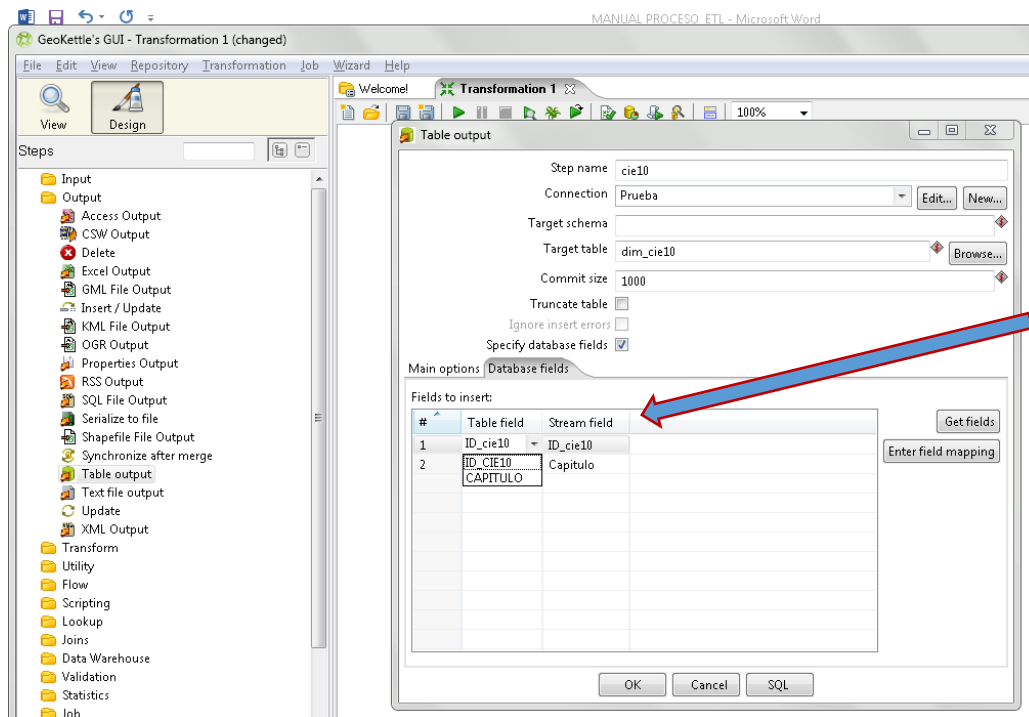
Buscamos Mysql para crear la conexión con la base de datos, configuramos la conexión y probamos con un Test



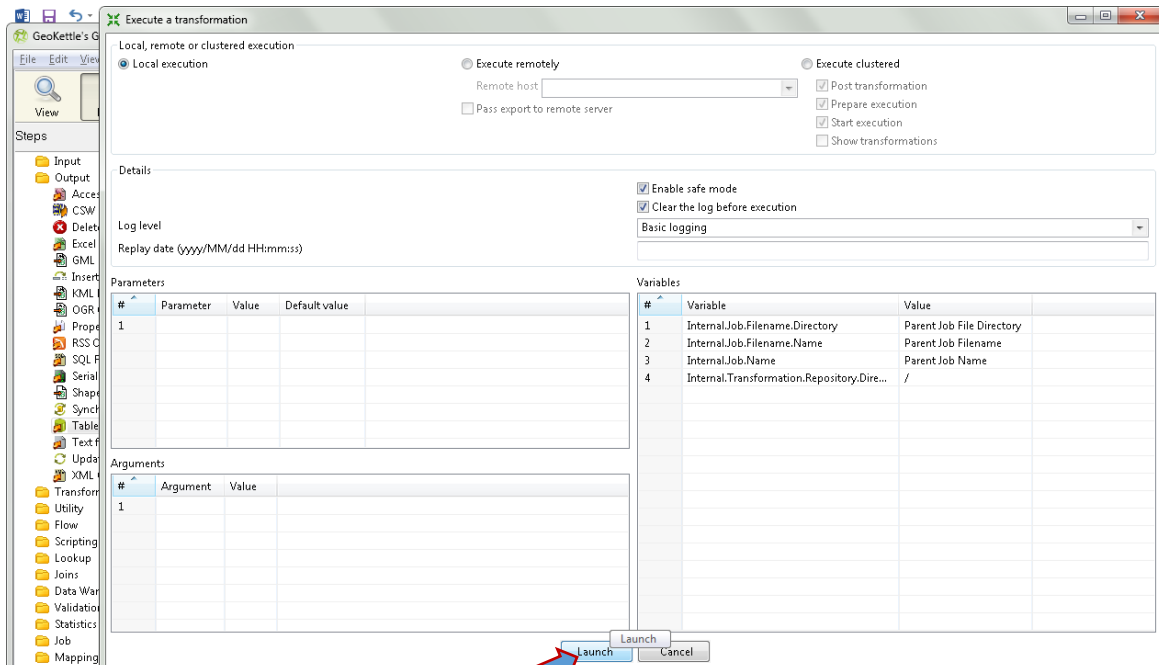
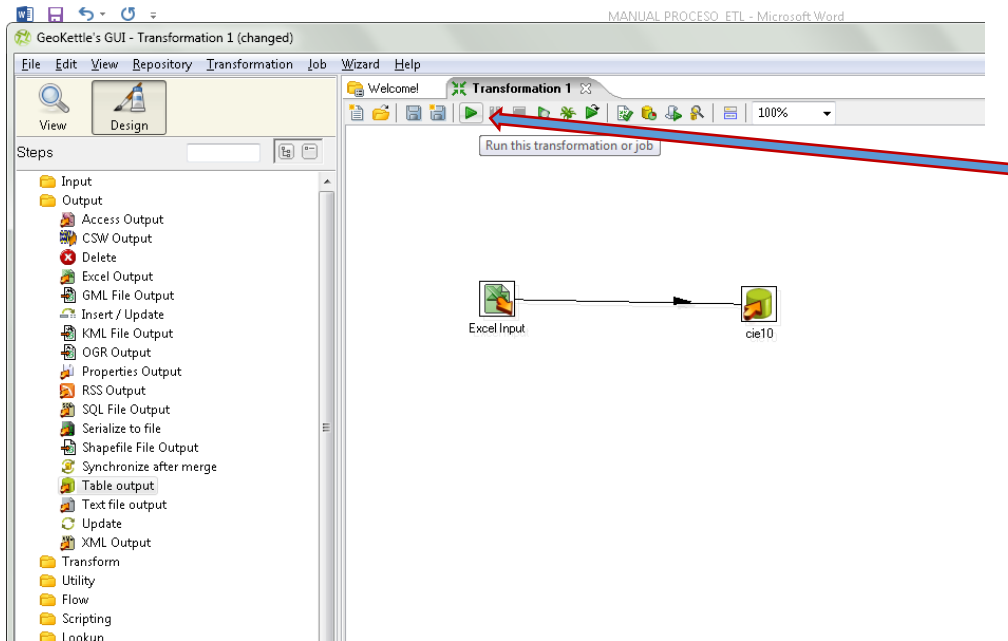
## Cargamos los campos de la base de datos



A lado izquierdo seleccionamos los campos de la base de datos, los campos de la derecha contienen datos de las tablas de Excel.

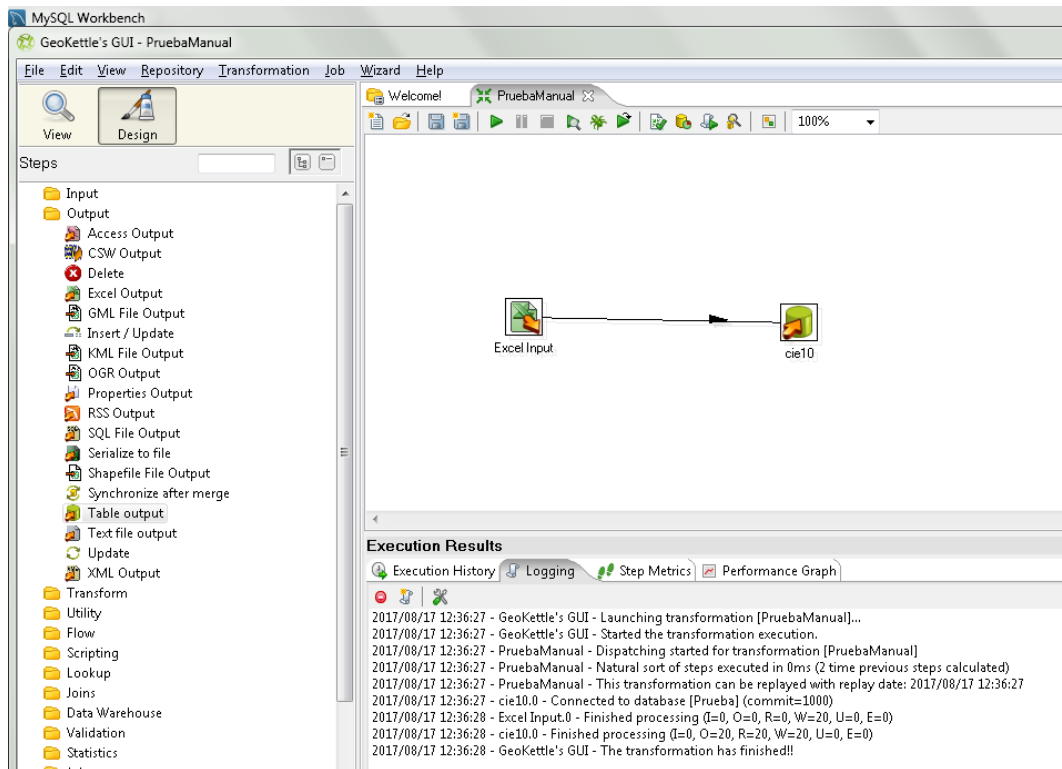


Para verificar que los datos están correctamente relacionados procedemos con ejecutar la transformación.





En la pestaña de resultados verificamos que no exista ningún error en la transformación.



A continuación verificamos los datos que se encuentran en la base de Mysql

