**Universidad San Buenaventura**

**Mario Juan Sebastian Reyes Casas**

**Manueles de openrefine y talend**

**Bogotá D.C**

**14/05/20**

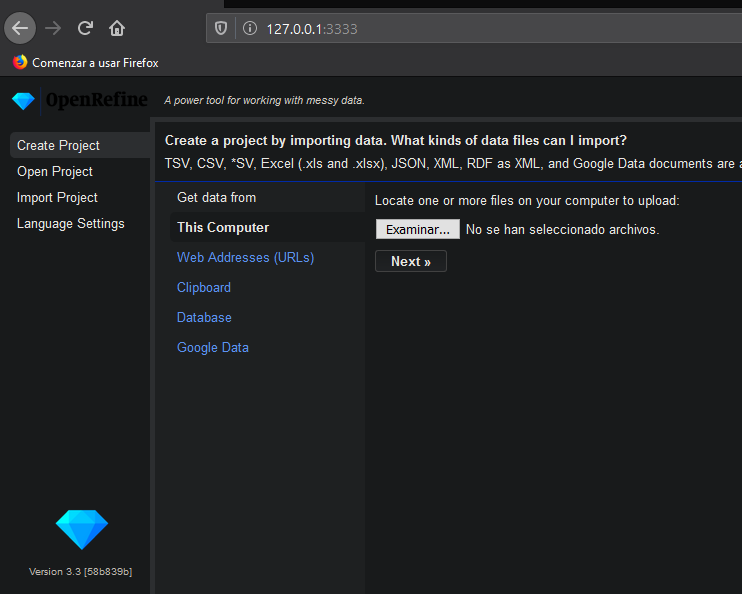
**Índice :**

**Open Refine**

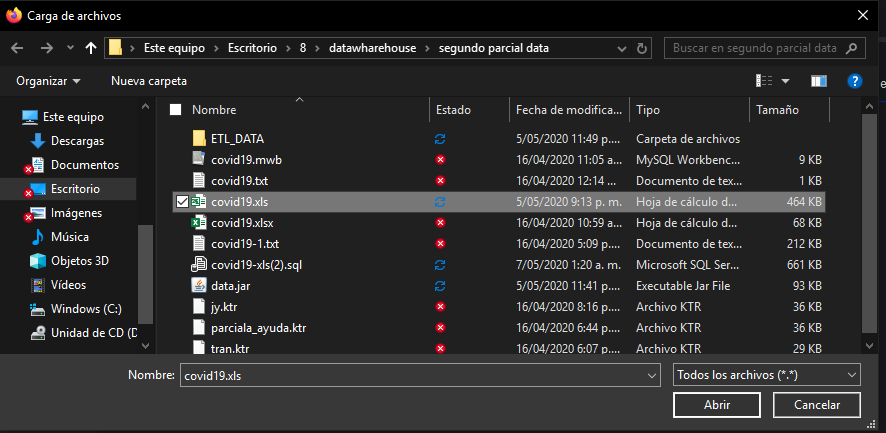
**Talend**

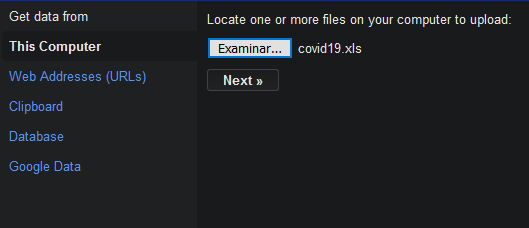
**Open Refine :**

Para abrir el software de openRefine simplemente lo descargamos y ejecutamos el .exe ,una vez abierto tendremos algo así :

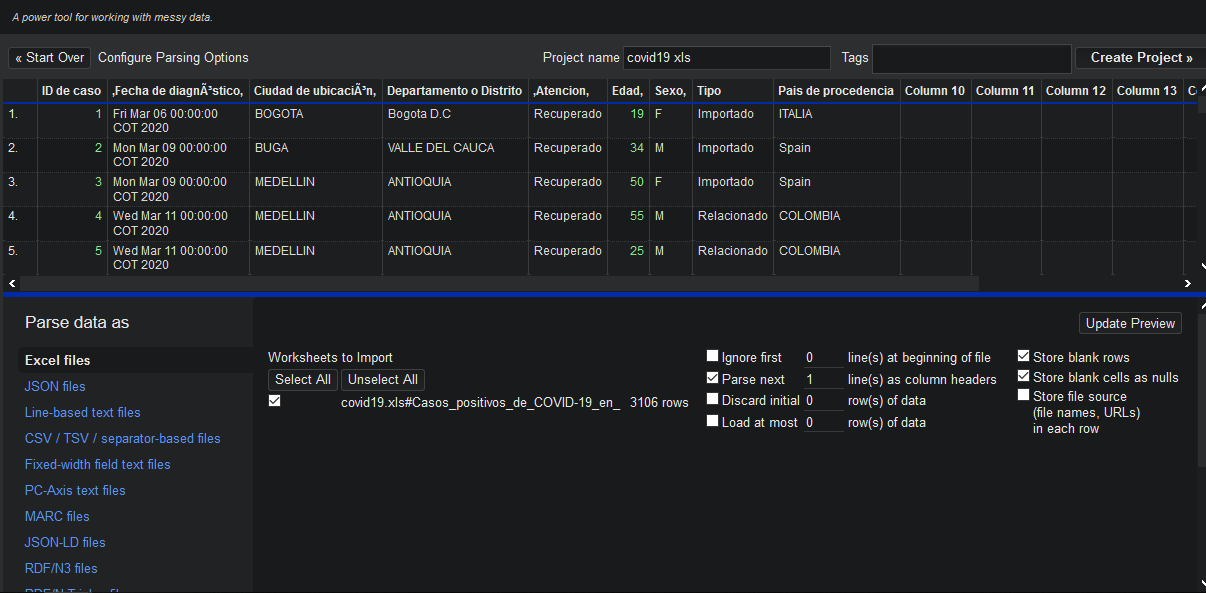


Se abre el explorador y nos abrirá el localhost en el puerto 3333 ,ahora vamos a dar en “examinar” para traer el xlsx





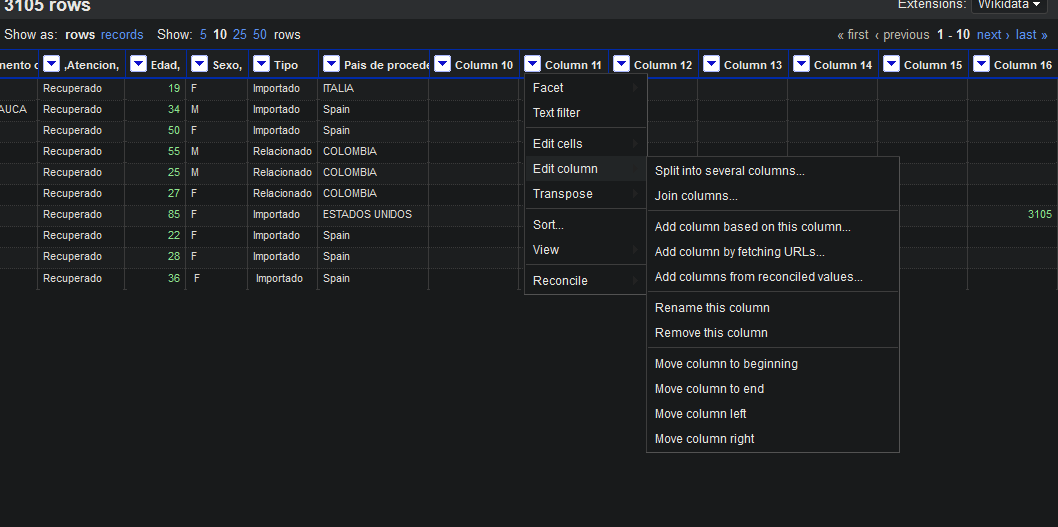
Damos en next



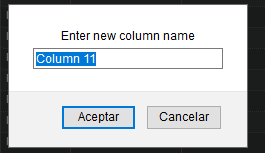
Y luego creamos el proyecto, podemos ver los datos que ya están aquí



Podemos ver cuantas filas hay , podemos entonces ahora empezar a realizar las transformaciones :



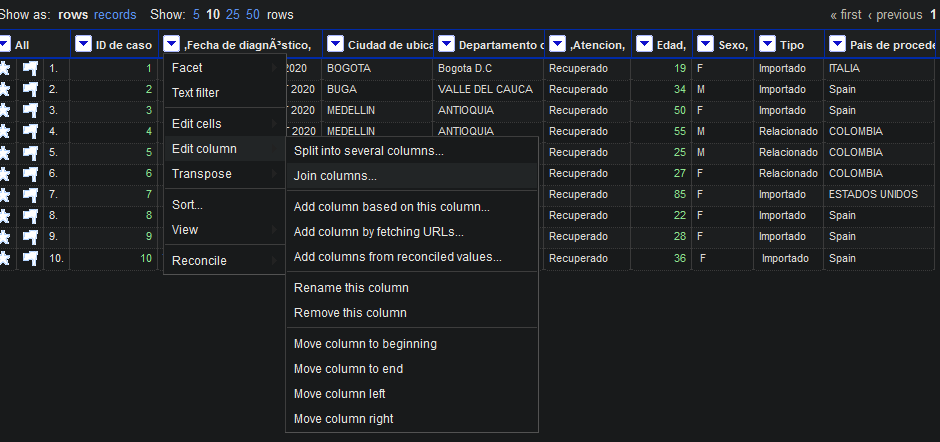
Vemos que hay 7 columnas que se agregaron sin haberlo deseado , podemos eliminarlas dando en edit colum>remove this columna



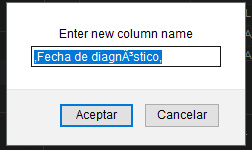
Luego nos pedirá el nombre de la columna a eliminar



Vemos que la fecha esta en un formato erroneo,entonces se puede editar el nombre de la columna



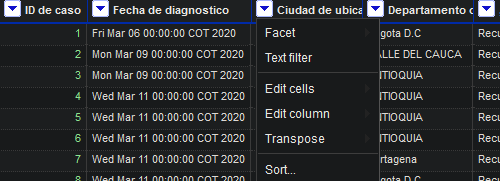
Damos click en las opciones para editar el nombre , edit columna>rename the columna



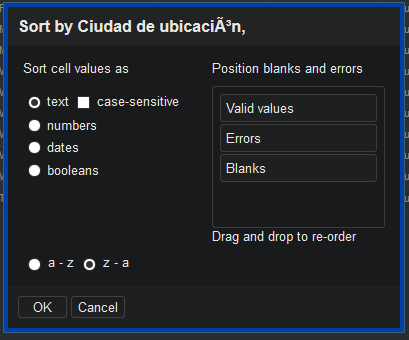
Podemos editarla



Ya hemos cambiado el nombre



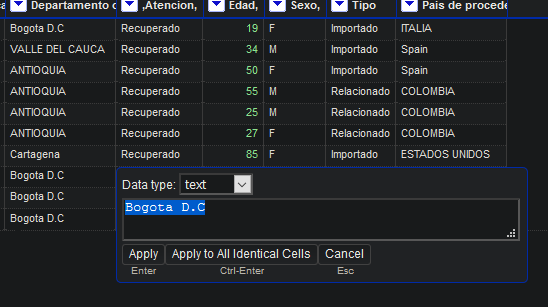
Ahora vamos a organizar la ciudad de ubicación editando sur orden para que se pueda ver ascendentemente o descendentemente , yo elijo que sea descendente



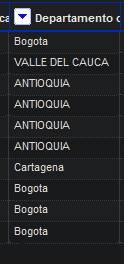
Ahora vemos como se ha editado



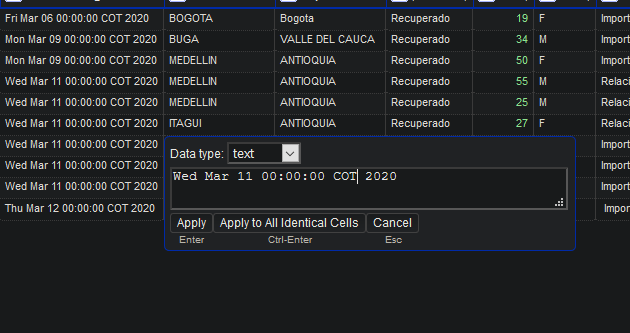
Ahora si queremos quitar el D.C de Bogotá simplemente vamos a edit en la celda :



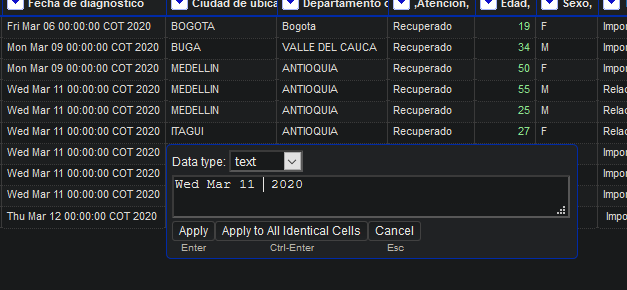
Ahora borramos el D.C y luego Apply to all identical Cells



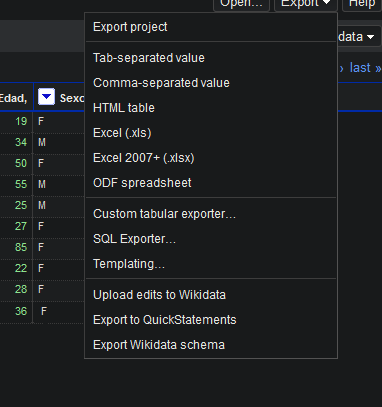
Ahora voy a cambiar el 0:00:00 que tienen algunas celdas



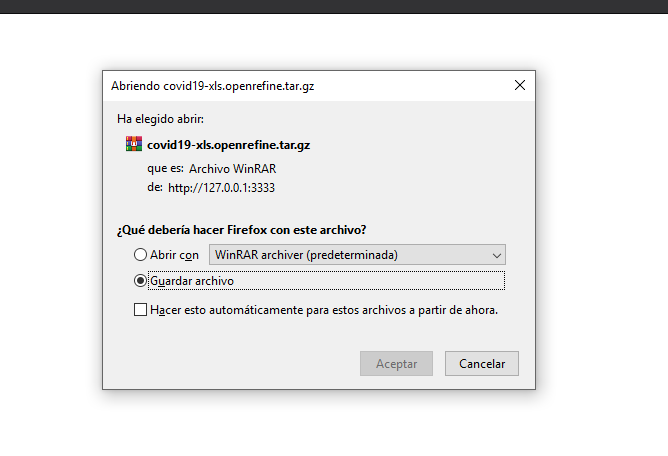
Simplemente elimino los 0 y el COT y listo



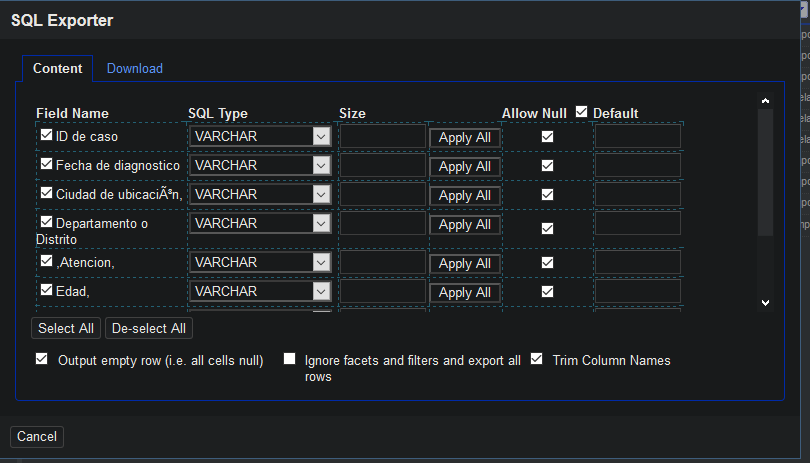
Vemos que han cambiado las celdas , ahora voy a realizar la carga en la base de datos , desafortunadamente la carga la debemos realizar nosotros descargando el .sql e importarlo en workbench o tu cliente de preferencia



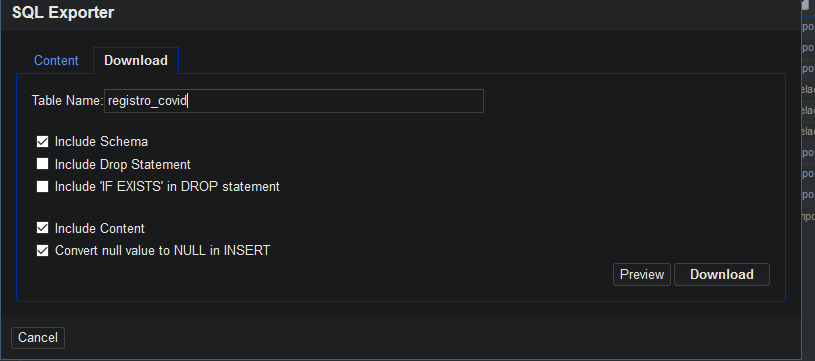
Elegimos SQL exporter , o exportar el proyecto si queremos volverlo a abrir



Esto para el proyecto ,ahora el .SQL



Ahora se pueden ver las celdas , podemos cambiar el SQL Type , elegiremos number para que no nos de error



Luego en la pestaña de Download damos en include ‘if exists’ para que borre si existe la tabla , luego cambiamos el nombre de la tabla para que sea mas sencillo importar el .sql

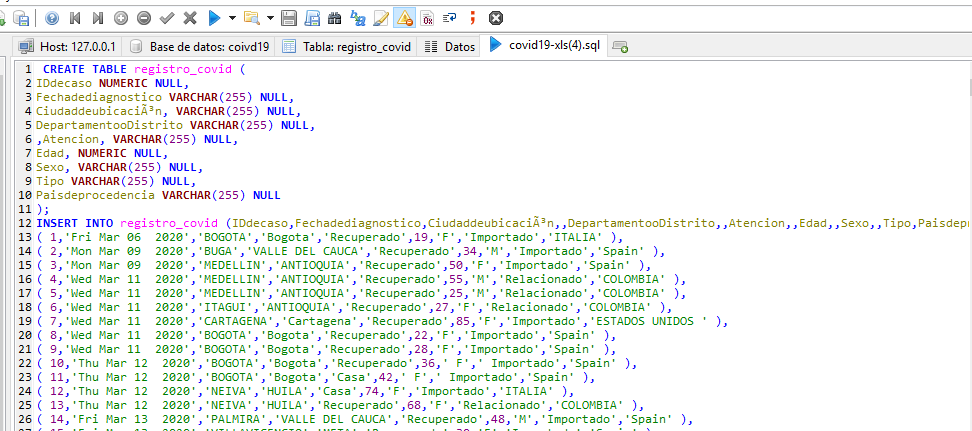
El archivo generado es :



Si lo abrimos en blog de notas:

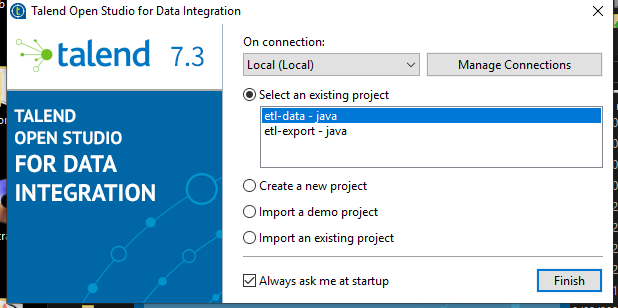


voy a usar heidy sql para importar los cambios

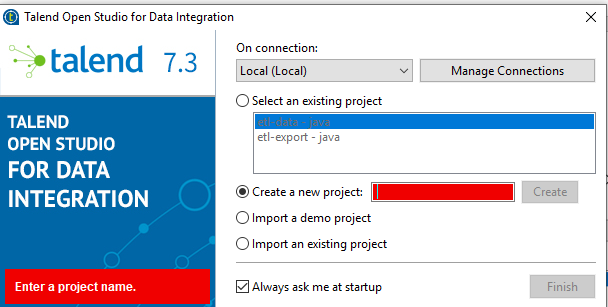


Ahora simplemente damos start y se llenará la base de datos

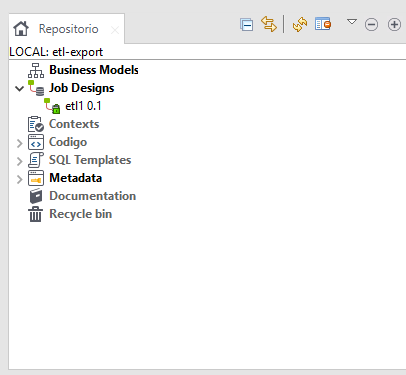
**Talend:**



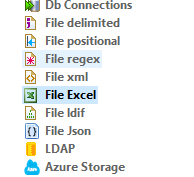
Lo primero es crear un nuevo proyecto , si ya tenemos uno simplemente elegimos el que es



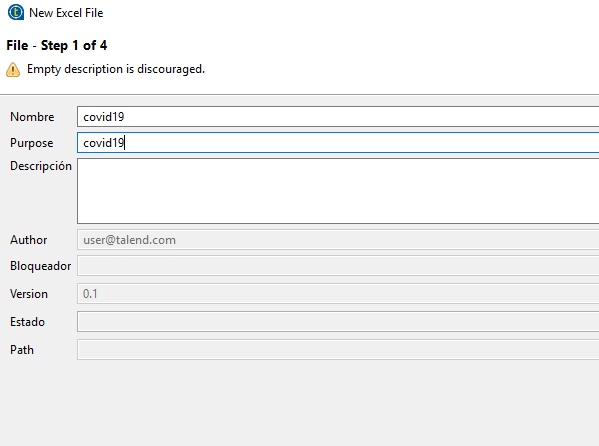
Ahí se pone el nombre del proyecto ,el que desee el usuario



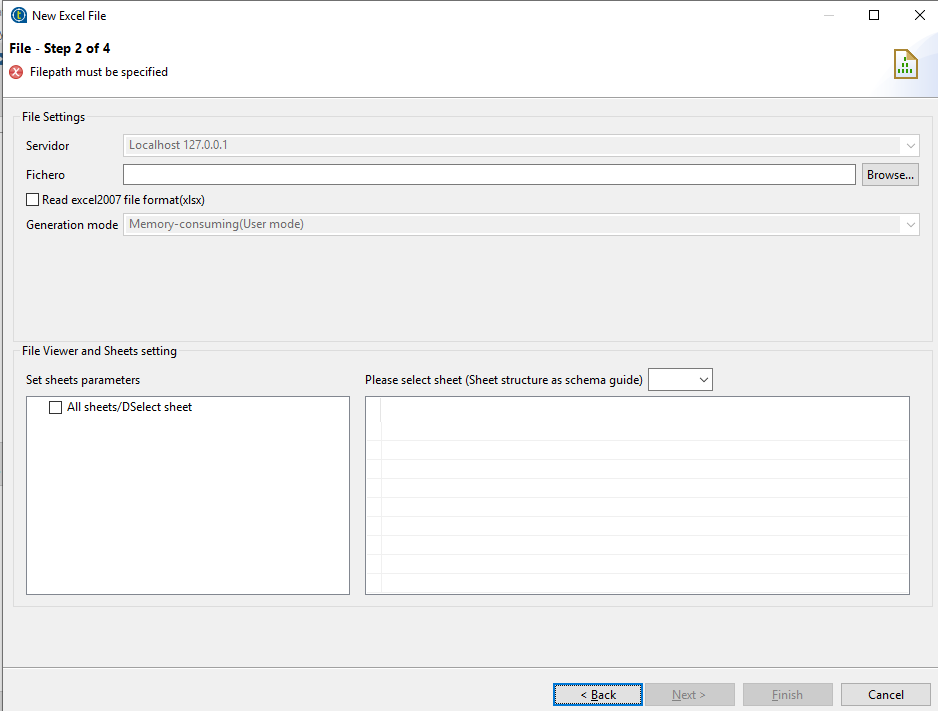
Al lado izquierdo se pueden encontrar los contextos ,los trabajos , la metadata etc. Lo mas indicado es que primero se de click izquierdo sobre “job designs” y luego se cree un nuevo trabajo luego de esto se debe crear un input



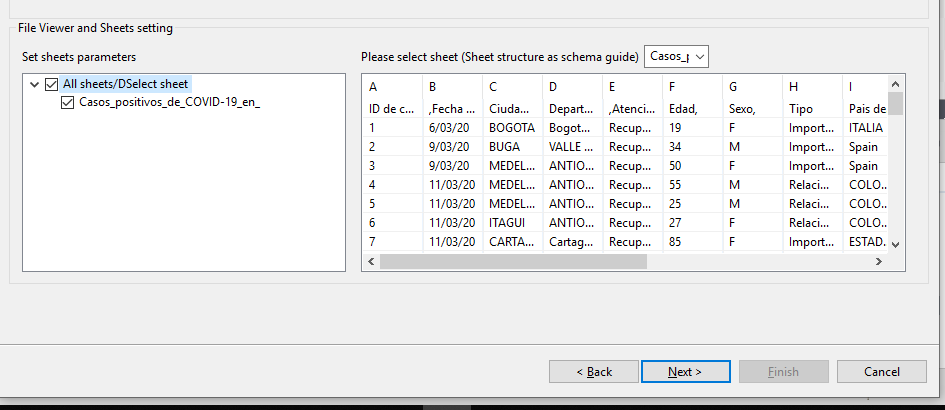
Damos click en metadata y luego creamos una conexión con Excel dando click izquierdo sobre “file Excel”



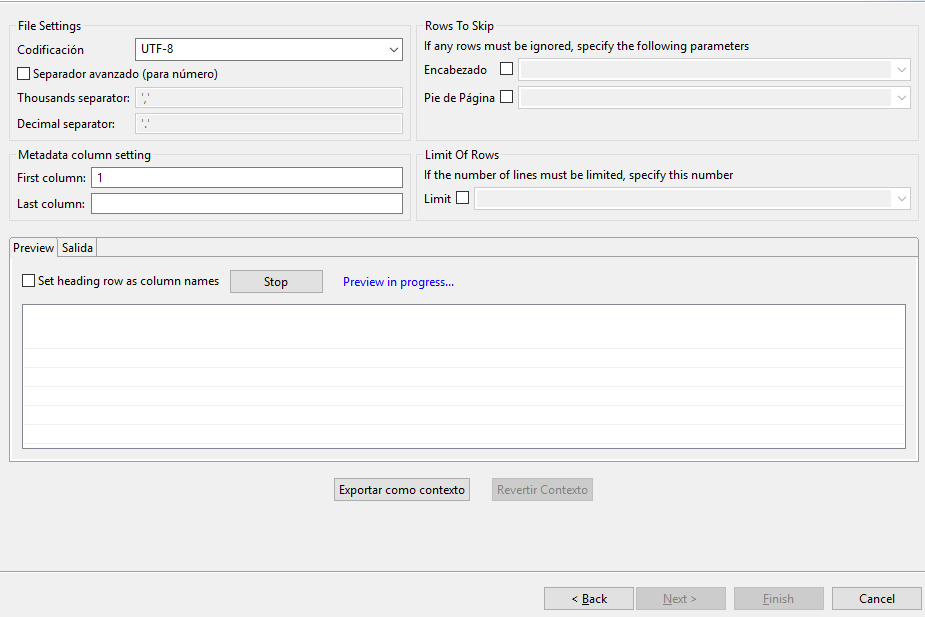
Aseguramos que este lleno el campo de nombre y propósito



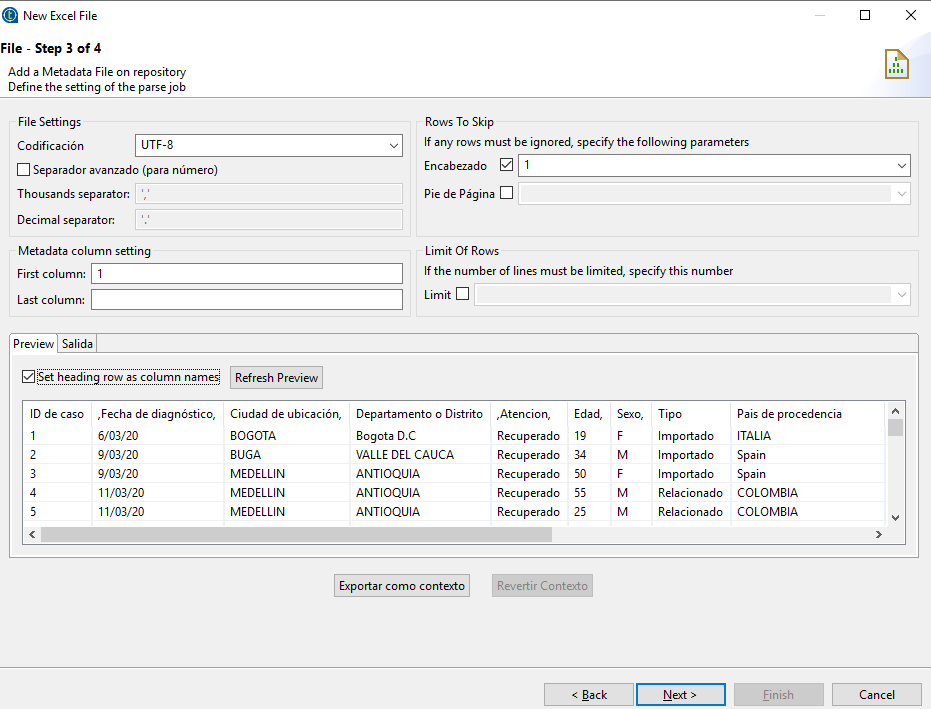
Localizamos el fichero



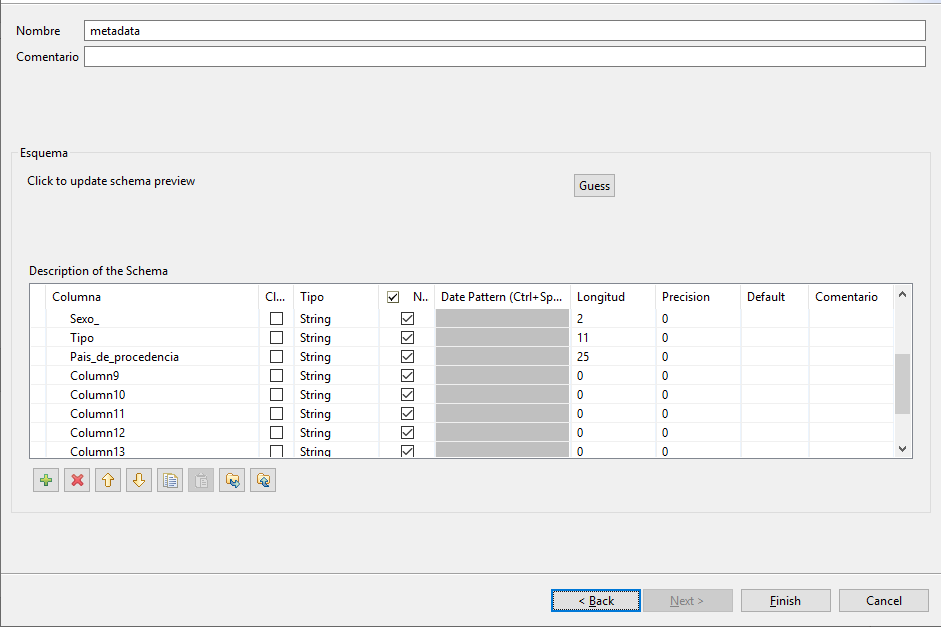
Elegimos el datasheet o la hoja de Excel que necesitamos ,en este caso solo tengo una y los datos se pueden visualizar a la derecha



Luego esperamos a que se cargue completamente los datos ,marcamos la opción en la que el encabezado de las columnas serán el nombre de las columnas



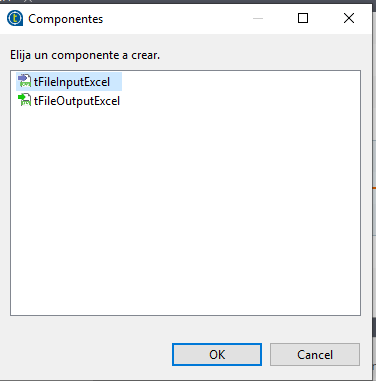
Los datos han sido cargados correctamente



Se puede seleccionar el tipo de dato de cada columna y eliminar columnas no deseadas ,como esas que dicen columna 9 a columna13



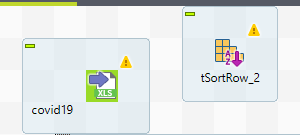
Si arrastramos este componente recibimos:

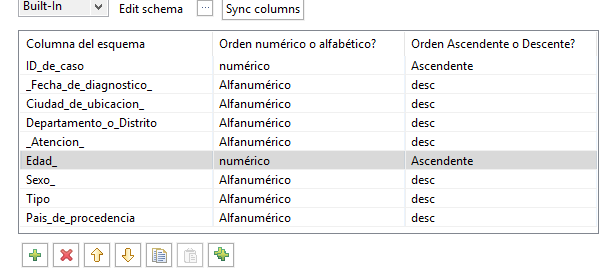


Podemos decidir si será un input o un output, seleccionamos un input ya que es la entrada de nuestros datos



Ahora vamos a traer una tarea que nos arregle las columnas ascendente o descendentemente

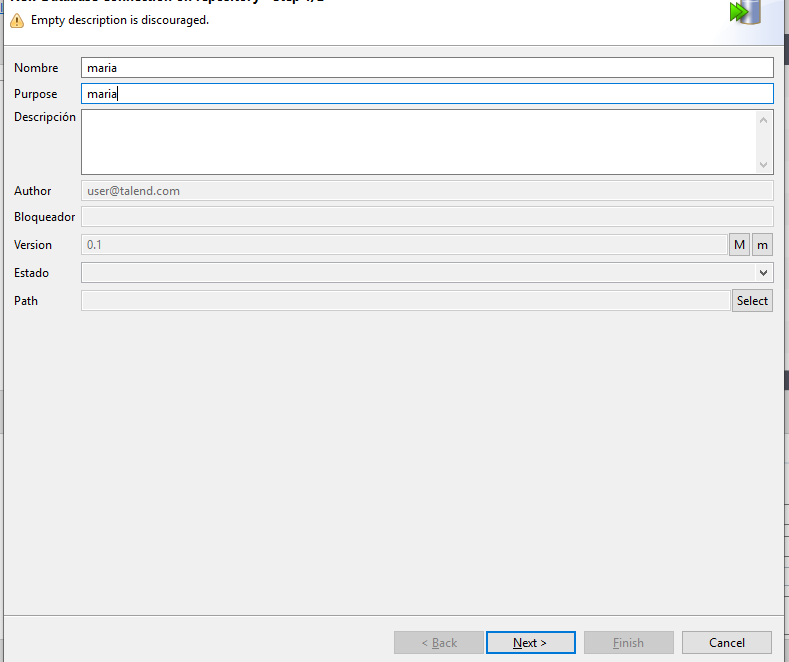




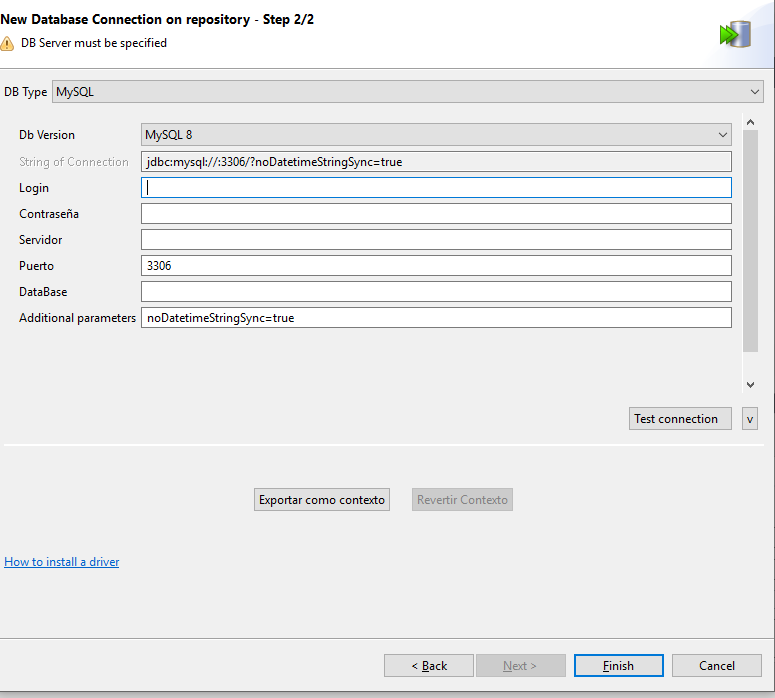
Traemos una función que nos arregla las columnas, ahí podemos observar las que nos arregla tanto ascendentemente como descendentemente y finalmente buscamos un output



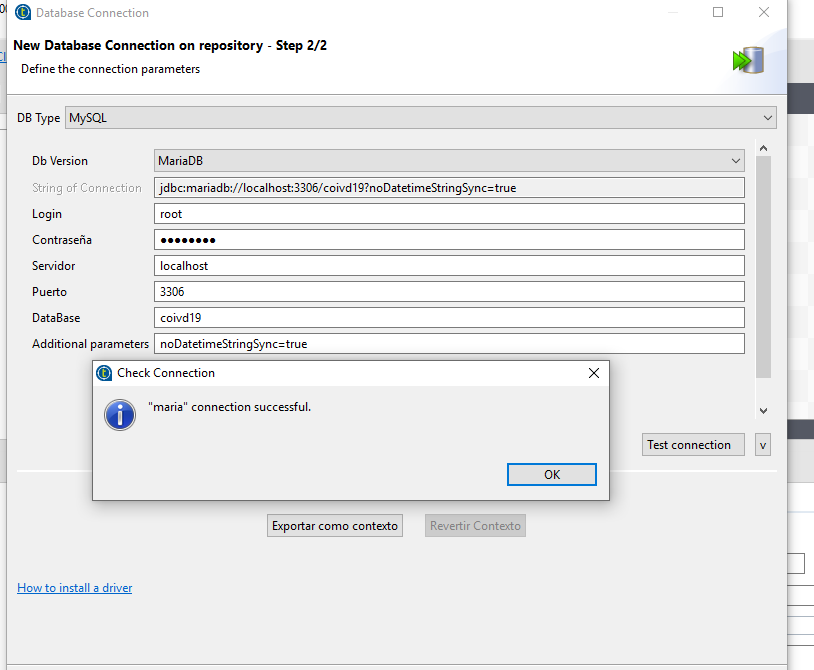
Vamos a crear una conexión en Db connections ,click izquierdo sobre ese componente y luego “crear conexión “



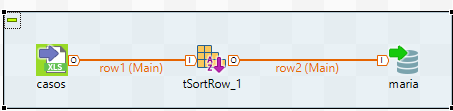
Garantizamos los parámetros para la conexión

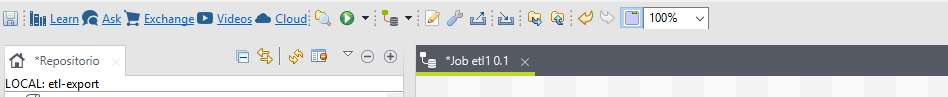


Debemos cambiar donde dice MYSQL8 por MARIADB

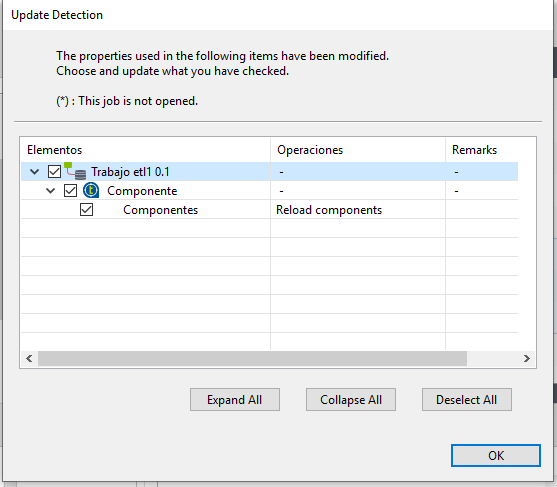


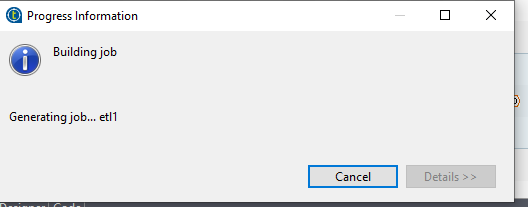
Una vez realizado esto, damos en ok->finish y luego unimos todo con las flechas ,para las flechas dar click izquierdo sobre el componente luego “fila-> main” y lo extendemos sobre el componente que queremos ,obteniendo:

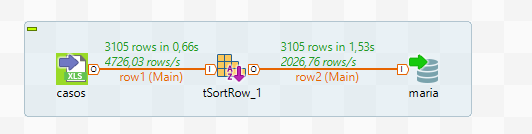




Damos en play







Consultando la base de datos se obtiene :

