

# Skills Matrix

---

Ana Sousa Cabanilla

Curso 2021/2022

Grado Superior de Desarrollo de Aplicaciones Multiplataforma

I.E.S. El Majuelo

## ÍNDICE

<b>INTRODUCCIÓN</b>	<b>2</b>
<b>REQUISITOS HARDWARE Y SOFTWARE</b>	<b>3</b>
Hardware	3
Software	3
<b>GUÍA DE DESCARGA E INSTALACIÓN</b>	<b>4</b>
PostgreSQL	4
DBeaver	18
Talend Open Studio	25
<b>PUESTA EN MARCHA</b>	<b>29</b>
<b>EXPLICACIÓN DEL FUNCIONAMIENTO DE LA APLICACIÓN</b>	<b>31</b>
<b>EXPLICACIÓN DEL CÓDIGO O PROCESOS MÁS COMPLEJOS O INTERESANTES</b>	<b>38</b>
<b>MAYORES DIFICULTADES ENCONTRADAS</b>	<b>39</b>
<b>LICENCIA</b>	<b>40</b>

## INTRODUCCIÓN

Este proyecto está basado en un situación real de la empresa Deloitte. En uno de los departamentos, Artificial Intelligence and Data, se diseñó un excel donde se recogen datos relativos a los empleados del departamento (personales, skills, proyectos, etc.). Este archivo es modificado constantemente por los empleados y está ideado para que sea fácil de manejar para ellos.

La finalidad de este proyecto es diseñar una base de datos y realizar una ETL para, en un futuro, utilizarla para la visualización de los datos con Power BI.

## REQUISITOS HARDWARE Y SOFTWARE

### Hardware

Procesador: AMD Ryzen 5 2500U with Radeon Vega Mobile Gfx 2.00 GHz

RAM instalada: 8,00 GB (6,95 GB usable)

Tipo de sistema: Sistema operativo de 64 bits, procesador basado en x64

ROM: 256,00 GB SSD

### Software

Sistema operativo: Windows 10 Pro de 64 bits

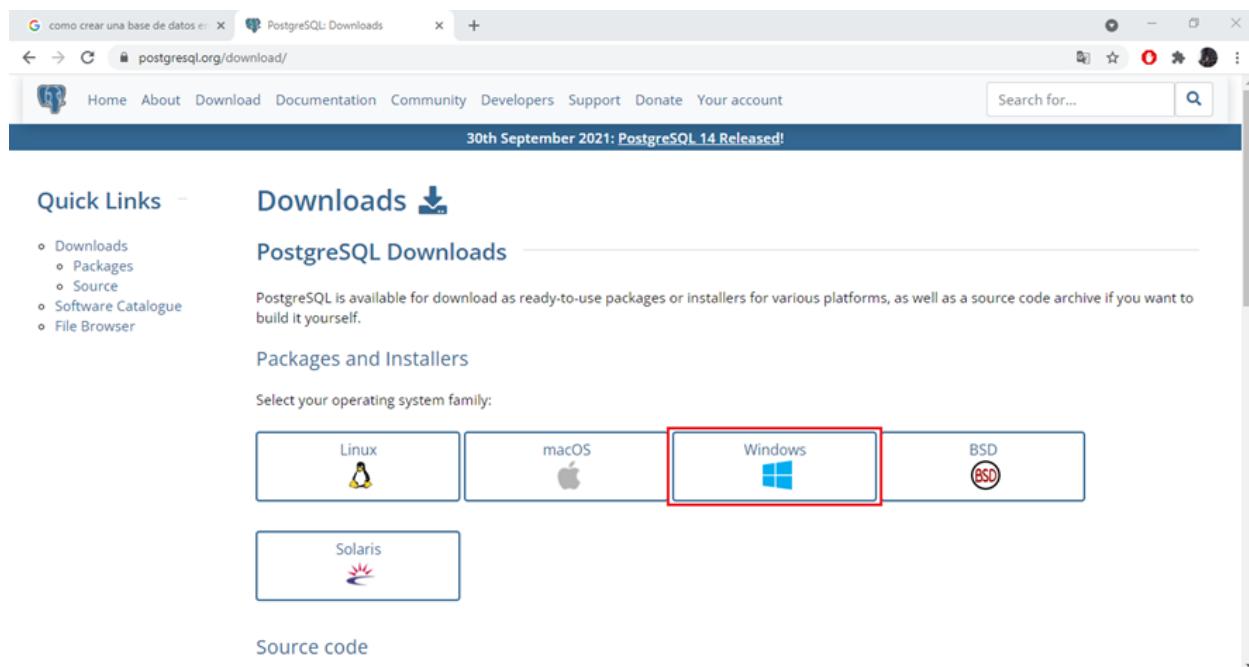
Sistema de gestión de bases de datos: PostgreSQL

Herramientas de administración de bases de datos: pgAdmin y DBeaver

Herramienta on premise para ETL: Talend Open Studio

## GUÍA DE DESCARGA E INSTALACIÓN

### PostgreSQL



The screenshot shows a web browser displaying the PostgreSQL Downloads page at [postgresql.org/download/](https://postgresql.org/download/). The page header indicates "30th September 2021: PostgreSQL 14 Released!". On the left, there's a "Quick Links" sidebar with links to Downloads, Packages, Source, Software Catalogue, and File Browser. The main content area is titled "Downloads" with a downward arrow icon. Below it is a section titled "PostgreSQL Downloads". A note states: "PostgreSQL is available for download as ready-to-use packages or installers for various platforms, as well as a source code archive if you want to build it yourself." Under "Packages and Installers", there's a "Select your operating system family:" dropdown. It contains five options: Linux (with a penguin icon), macOS (with an apple icon), Solaris (with a sun icon), Windows (with a blue square icon, which is highlighted with a red box), and BSD (with a circular logo). At the bottom, there's a link for "Source code".

Podemos descargar PostgreSQL desde su web oficial. Accederemos a través del siguiente [enlace](#) y haremos clic en el logo del sistema operativo que deseemos, en este caso Windows.

**Quick Links**

- Downloads
  - Packages
  - Source
- Software Catalogue
- File Browser

## Windows installers

### Interactive installer by EDB

[Download the installer](#) certified by EDB for all supported PostgreSQL versions.

This installer includes the PostgreSQL server, pgAdmin; a graphical tool for managing and developing your databases, and StackBuilder; a package manager that can be used to download and install additional PostgreSQL tools and drivers. Stackbuilder includes management, integration, migration, replication, geospatial, connectors and other tools.

This installer can run in graphical or silent install modes.

The installer is designed to be a straightforward, fast way to get up and running with PostgreSQL on Windows.

*Advanced users* can also download a [zip archive](#) of the binaries, without the installer. This download is intended for users who wish to include PostgreSQL as part of another application installer.

#### Platform support

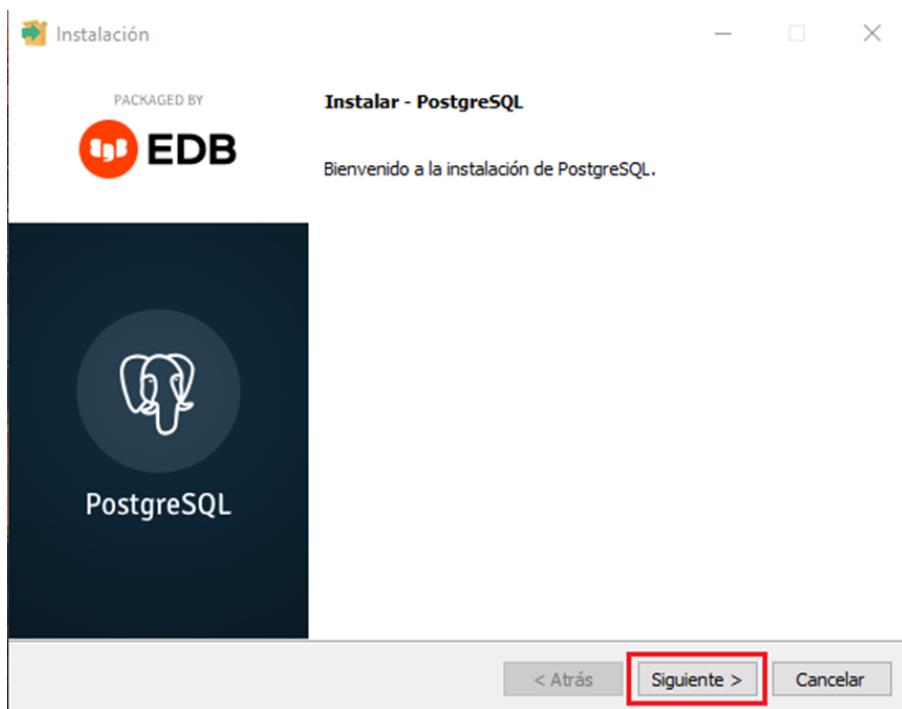
The installers are tested by EDB on the following platforms. They can generally be expected to run on other comparable versions:

PostgreSQL Version	64 Bit Windows Platforms	32 Bit Windows Platforms
13	2019, 2016	
12	2019, 2016, 2012 R2	

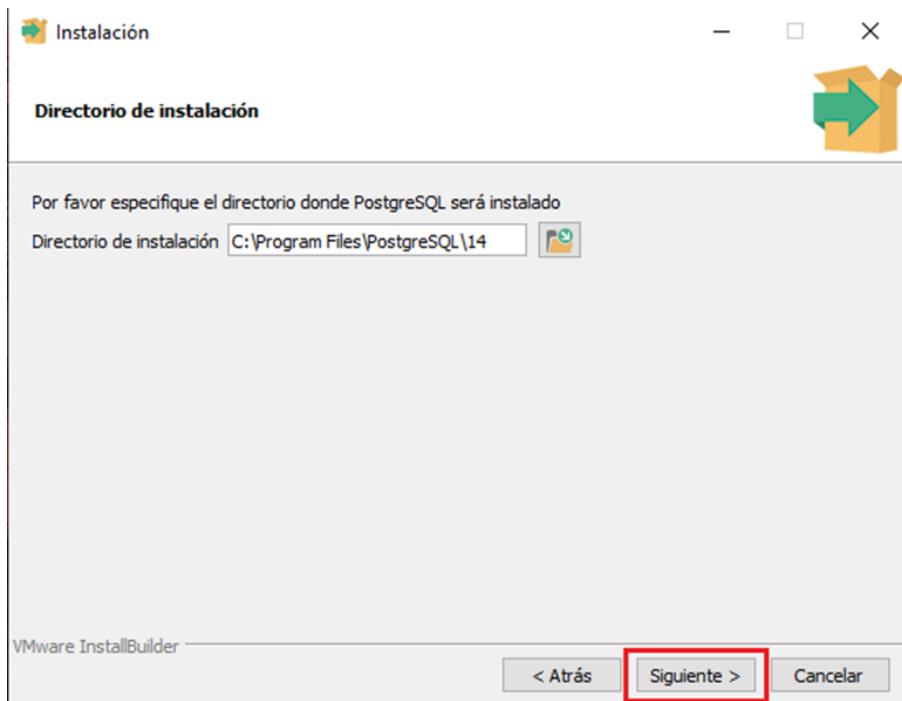
Haremos clic en “Download the installer”.

Version	Linux x86-64	Linux x86-32	Mac OS X	Windows x86-64	Windows x86-32
14	N/A	N/A	<a href="#">Download</a>	<a href="#">Download</a>	N/A
13.4	N/A	N/A	<a href="#">Download</a>	<a href="#">Download</a>	N/A
12.8	N/A	N/A	<a href="#">Download</a>	<a href="#">Download</a>	N/A
11.13	N/A	N/A	<a href="#">Download</a>	<a href="#">Download</a>	N/A
10.18	<a href="#">Download</a>				
9.6.23	<a href="#">Download</a>				

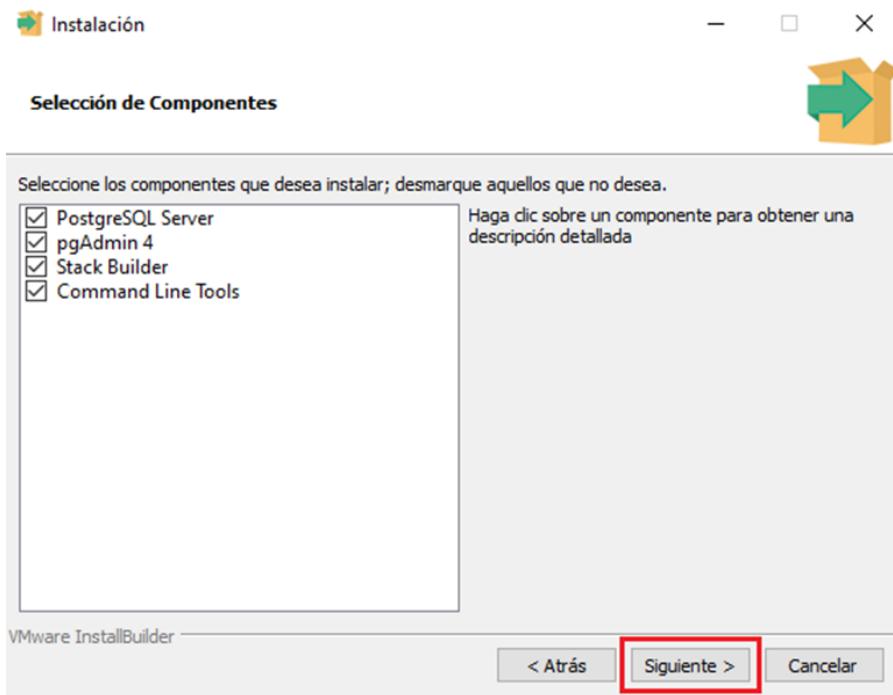
Clicaremos en el “Download” de la versión más reciente de la columna “Windows x86-64”.



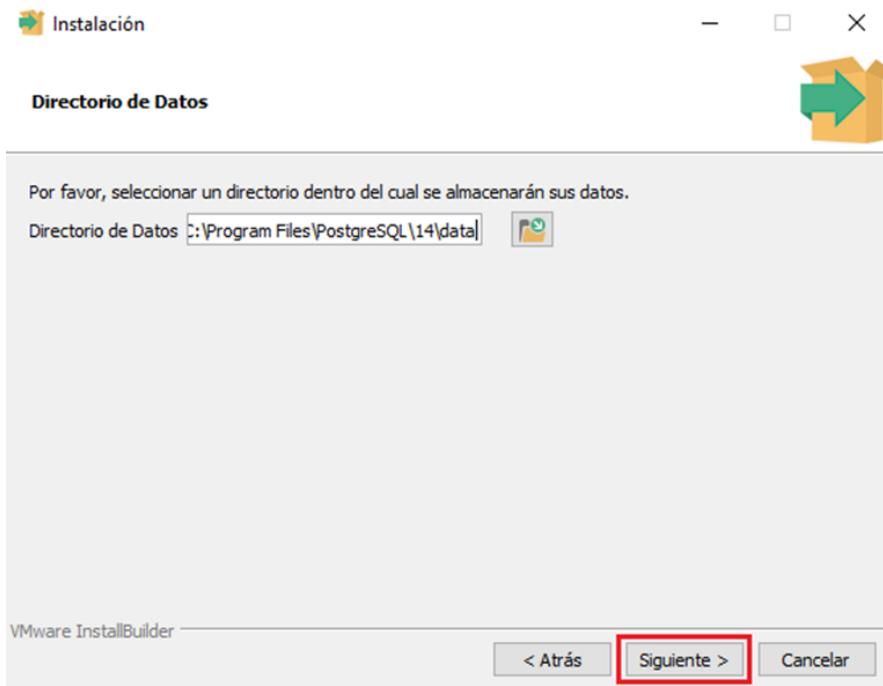
Una vez se haya descargado el archivo .exe lo ejecutaremos. Aparecerá una ventana como la anterior y haremos clic en "Siguiente".



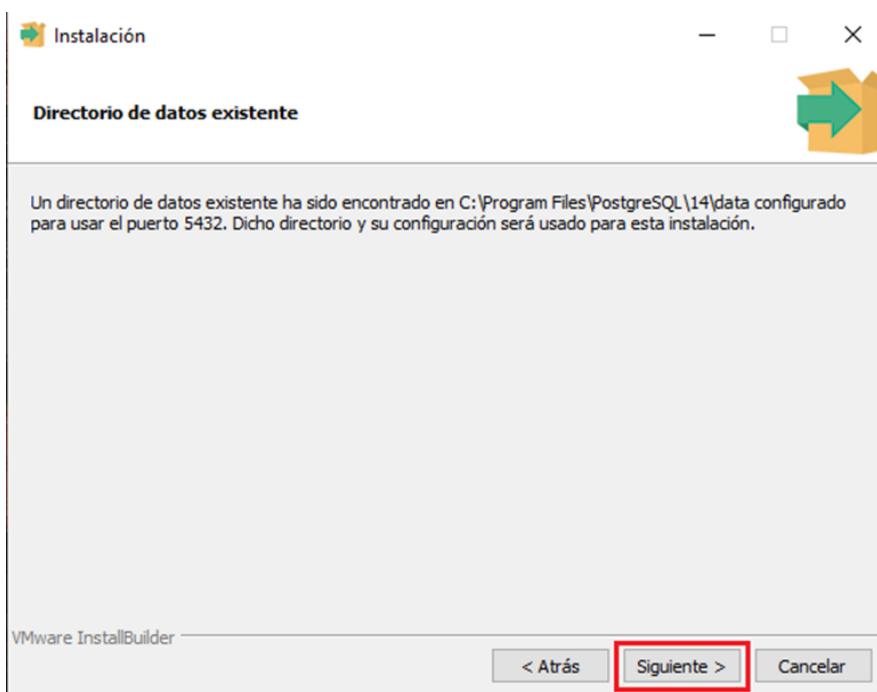
Dejaremos el directorio predeterminado y haremos clic en "Siguiente".



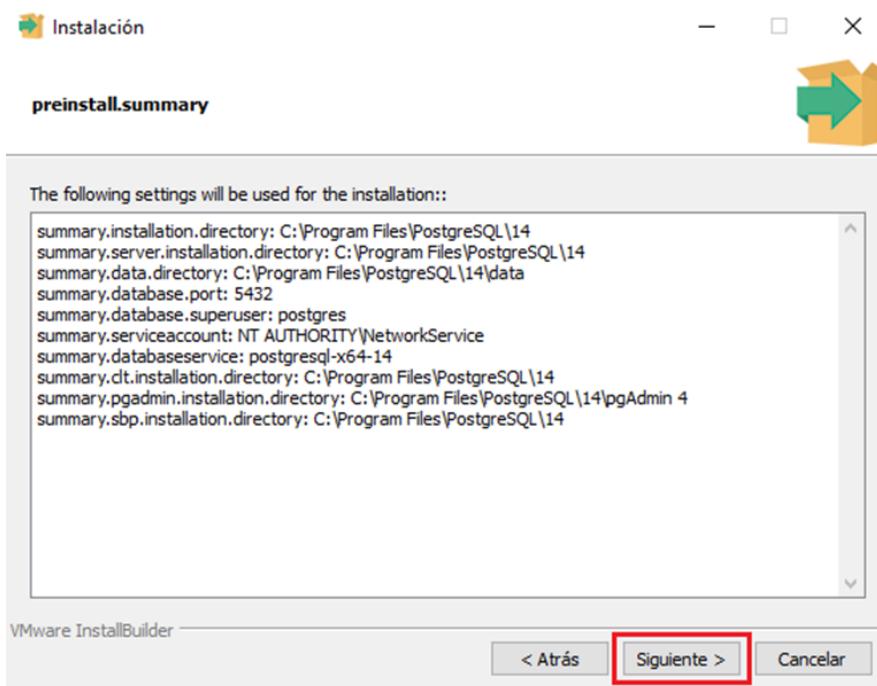
Comprobamos que estén seleccionados todos los componentes y clicaremos en "Siguiente".



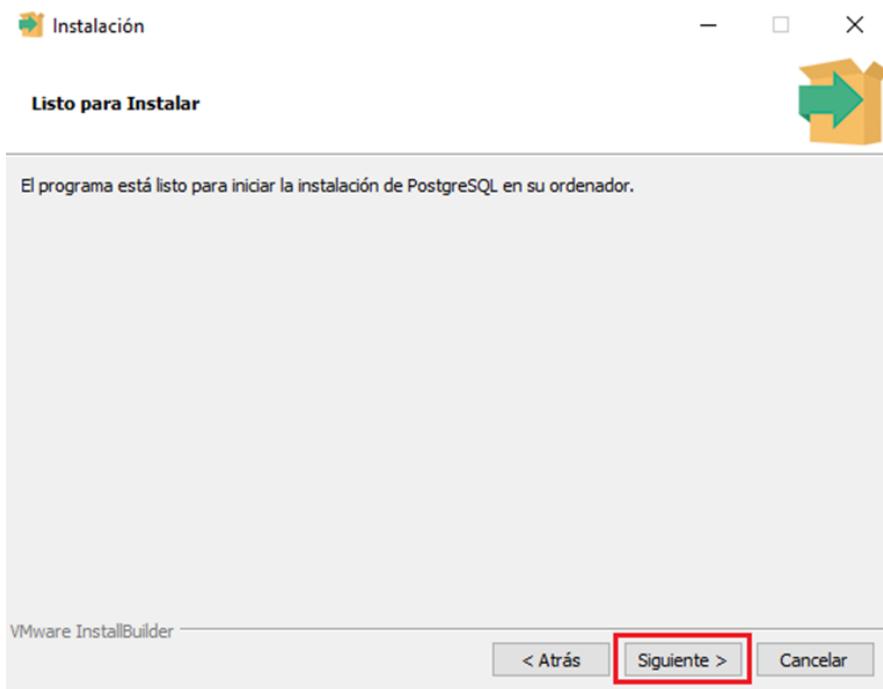
Dejaremos el directorio predeterminado y haremos clic en "Siguiente".



Clicaremos en "Siguiente".



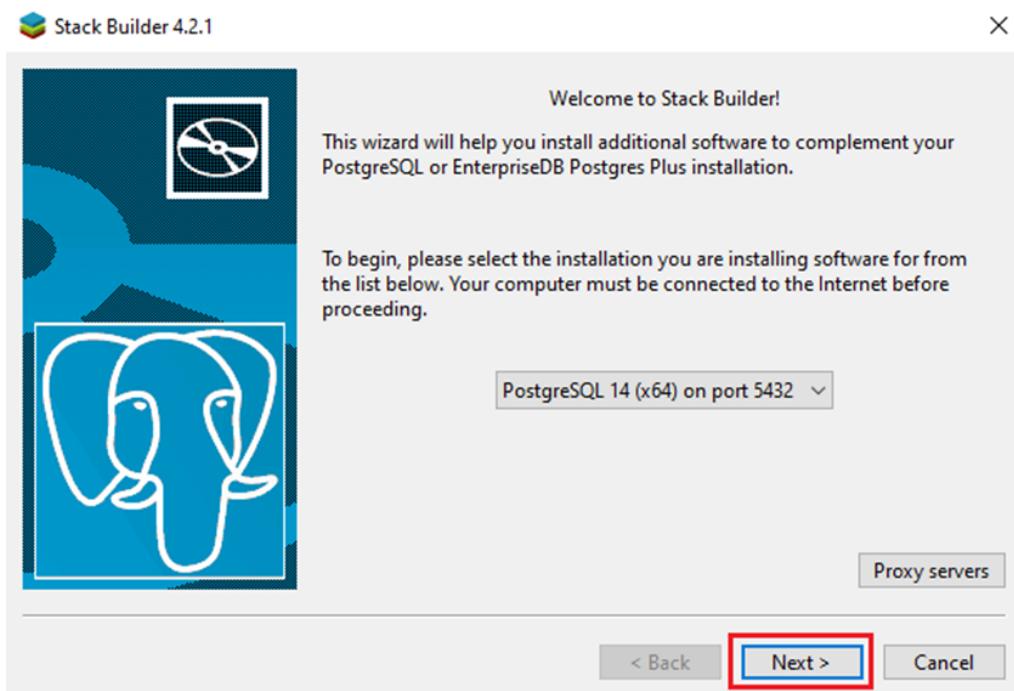
Volveremos a clicar en "Siguiente".



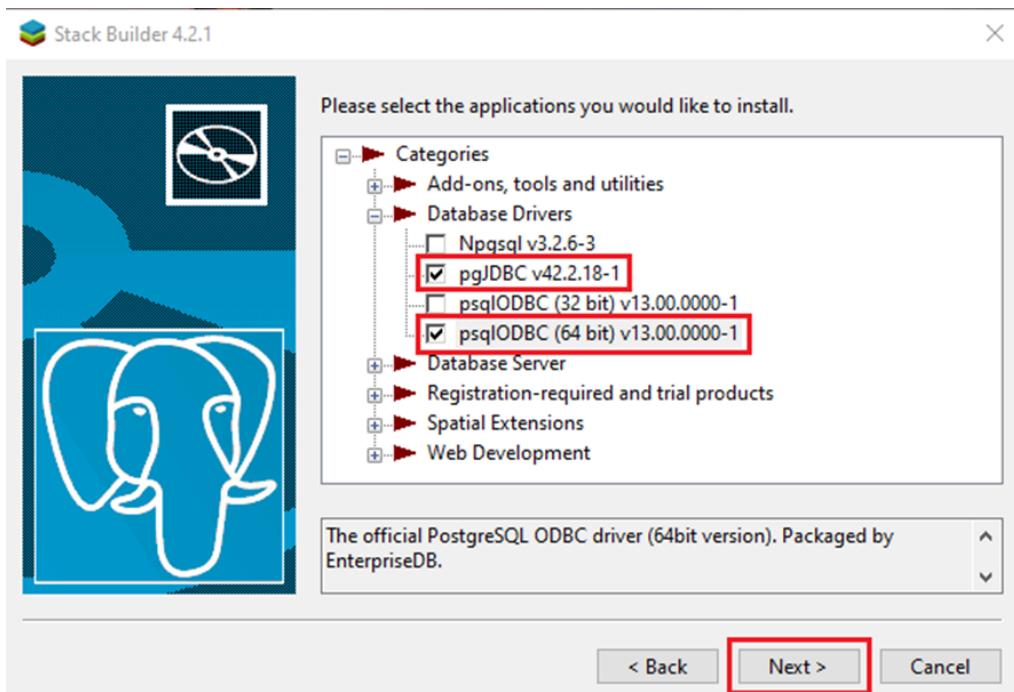
Haremos clic en "Siguiente" y comenzará la instalación.



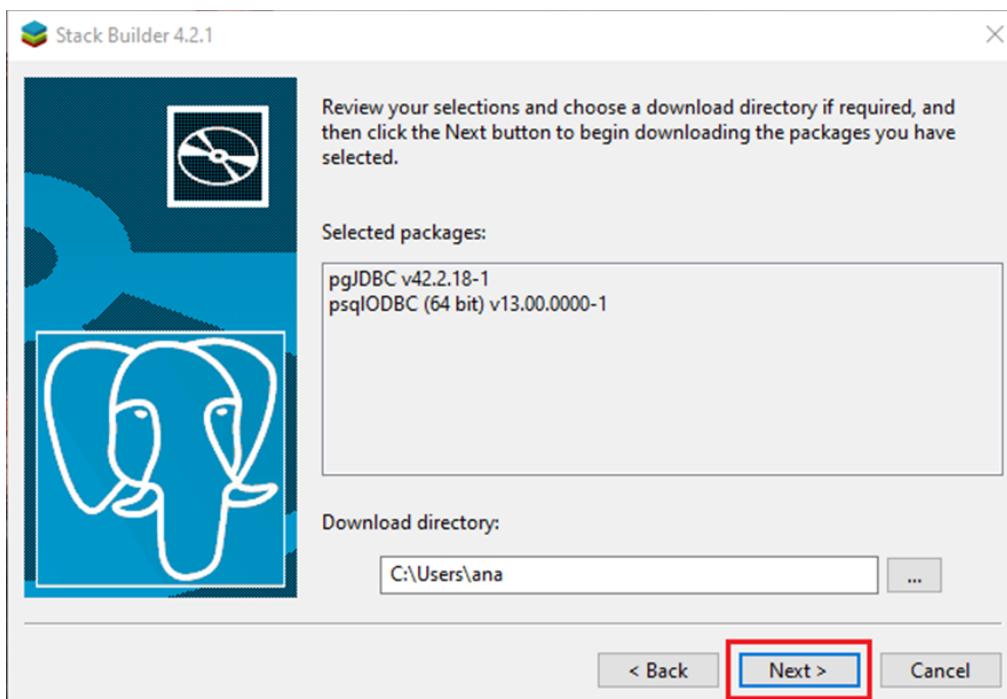
Una vez finalice la instalación haremos clic en "Finalizar".



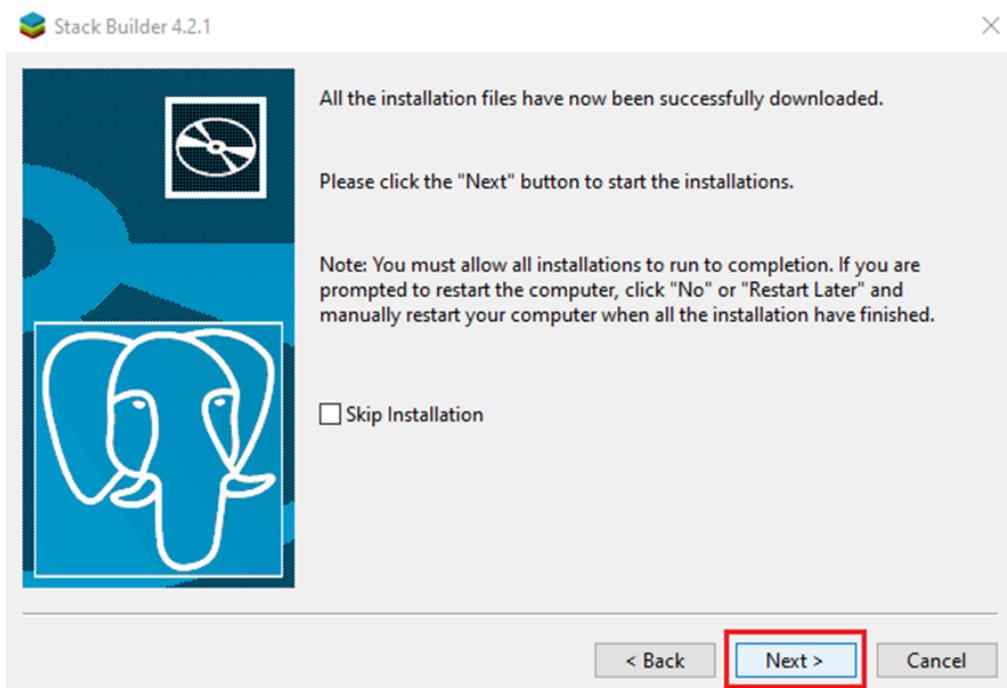
Se nos abrirá una nueva ventana y comenzaremos con la instalación de software adicional. Haremos clic en “Next”.



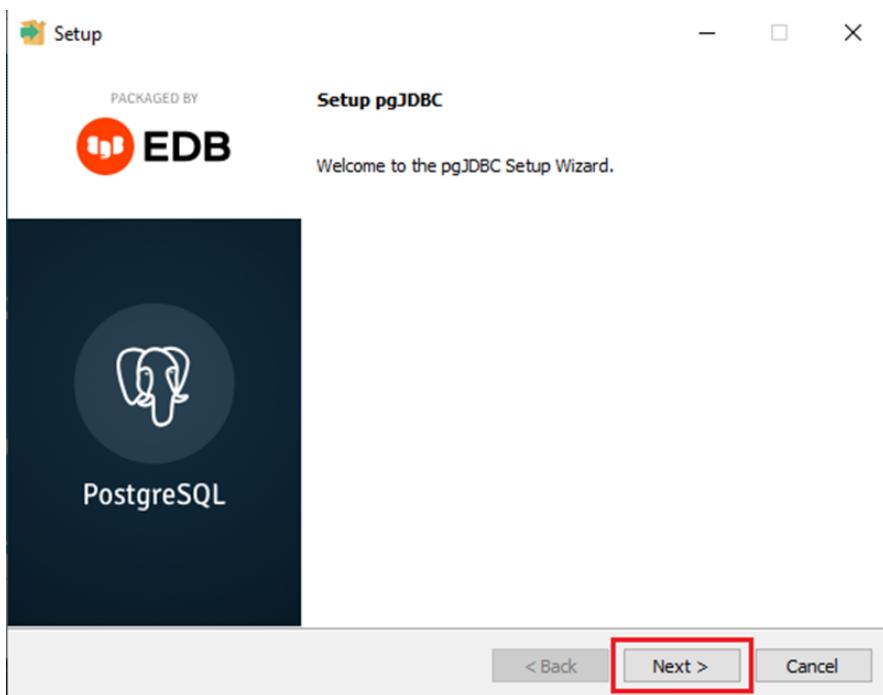
Seleccionaremos las aplicaciones que deseemos instalar, en este caso “pgJDBC v42.2.18-1” y “psqlODBC (64 bit) v13.00.0000-1”, y haréis clic en “Next”.



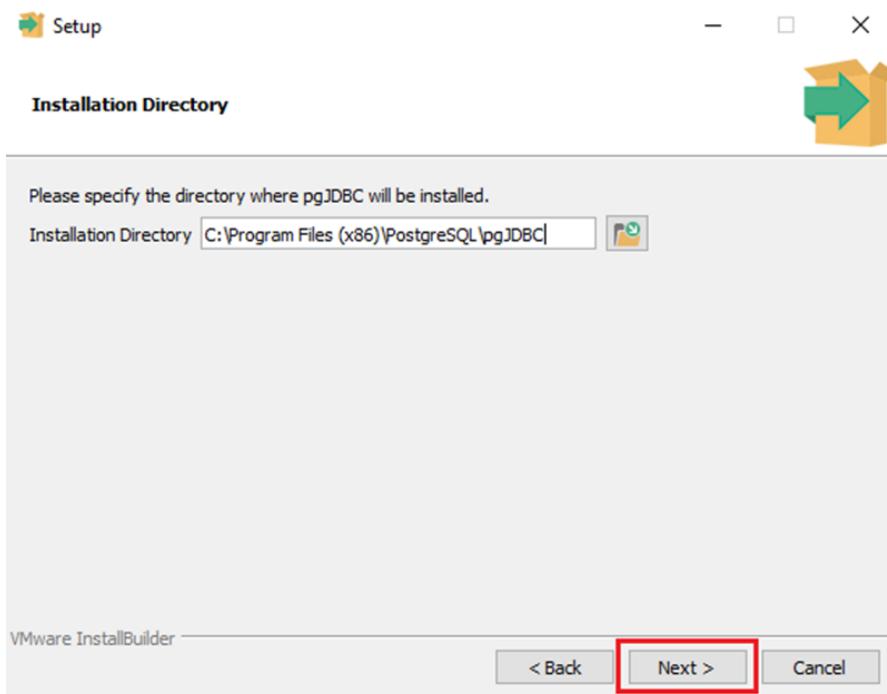
Clicaremos en "Next".



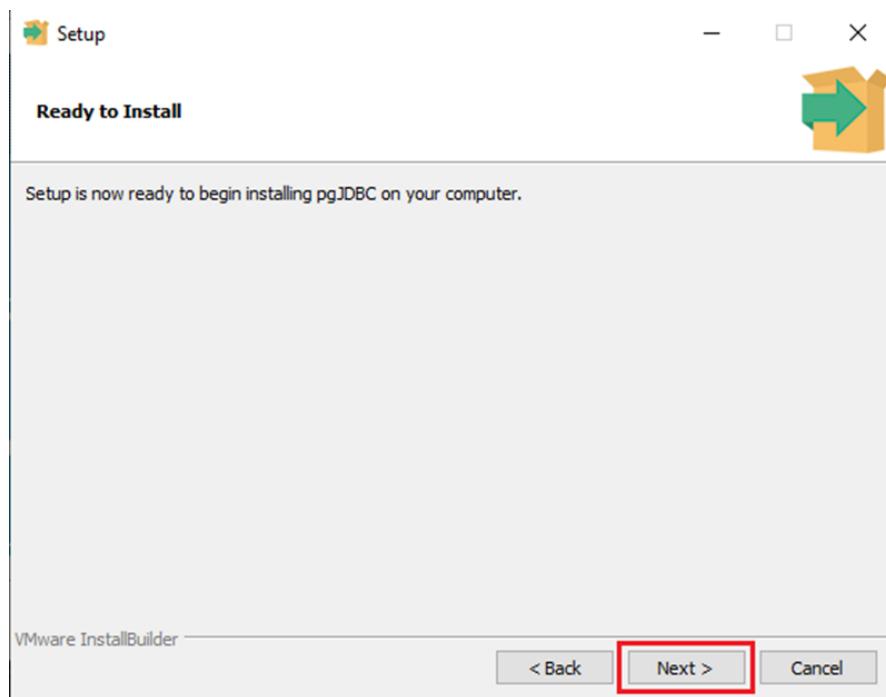
Volveremos a hacer clic en "Next" asegurándonos que la opción de "Skip Installation" esté desmarcada.



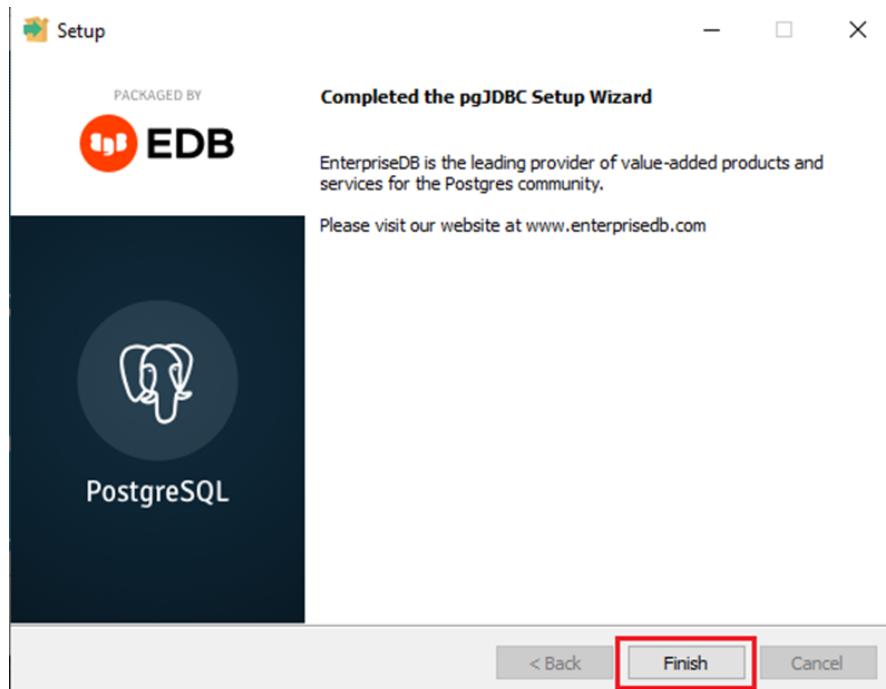
Se nos abrirá una nueva ventana y comenzaremos con la instalación de cada una de las aplicaciones adicionales que seleccionamos anteriormente. Haremos clic en "Next".



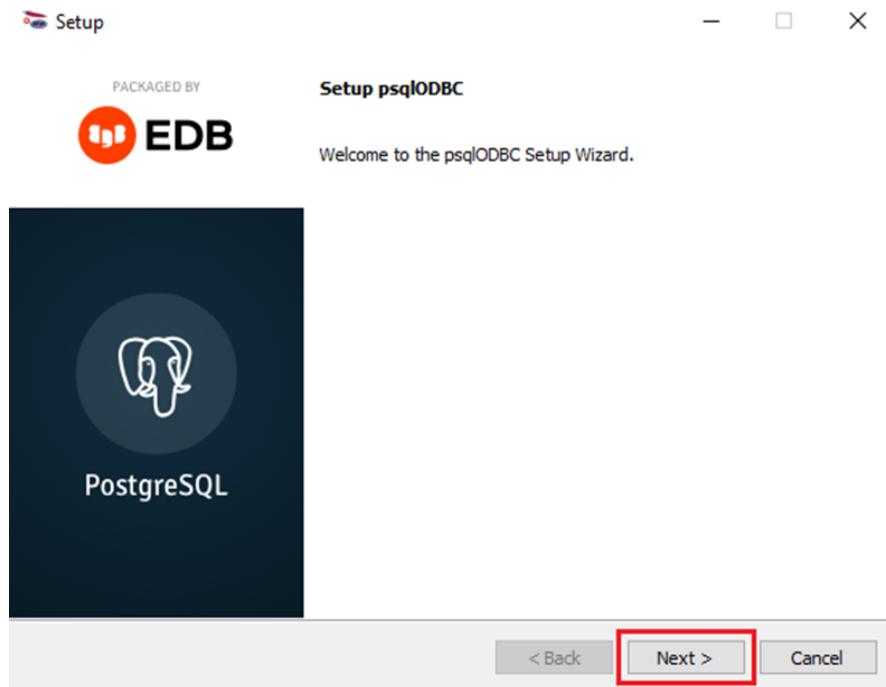
Haremos clic en "Next".



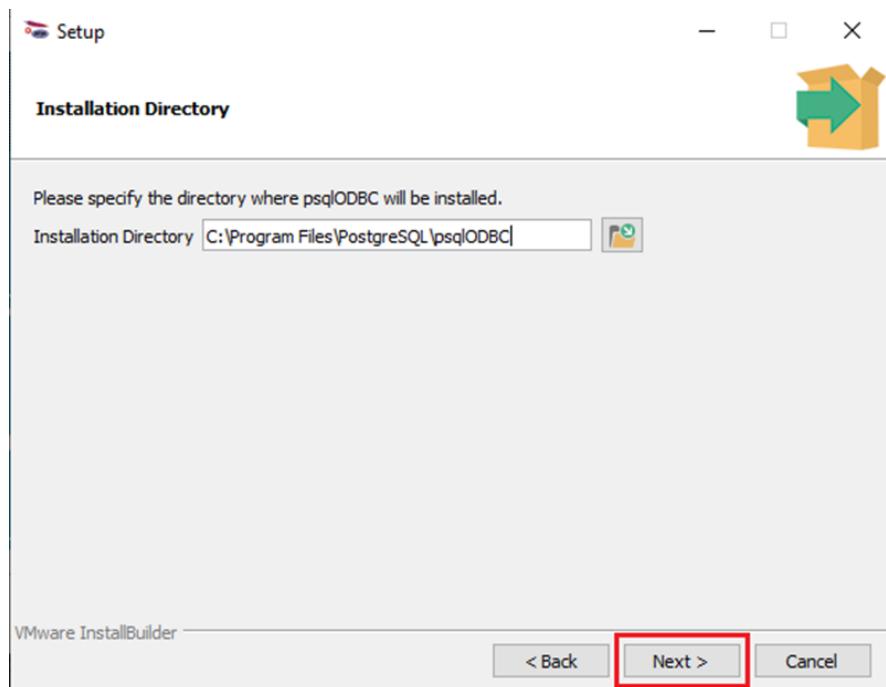
Volveremos a hacer clic en "Next" y comenzará la instalación.



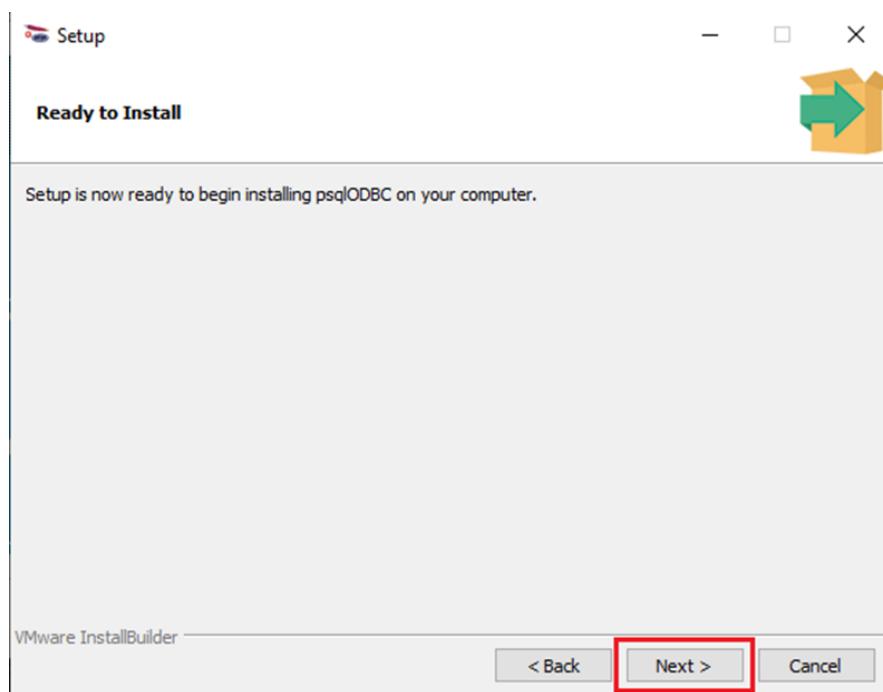
Una vez finalice el proceso clicaremos en "Finish".



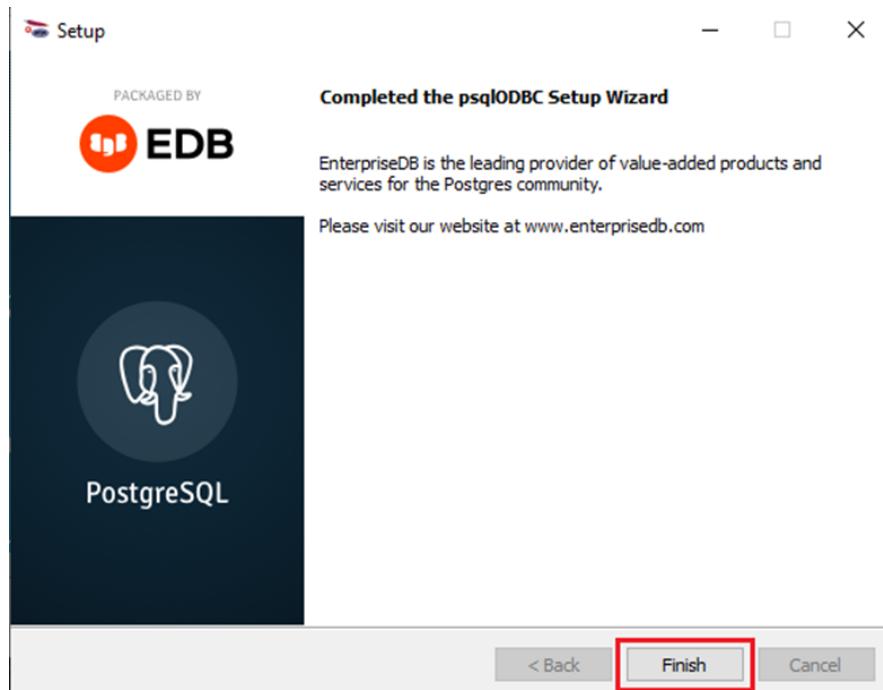
Se nos abrirá una nueva ventana para comenzar con la instalación de la siguiente aplicación. Haremos clic en "Next".



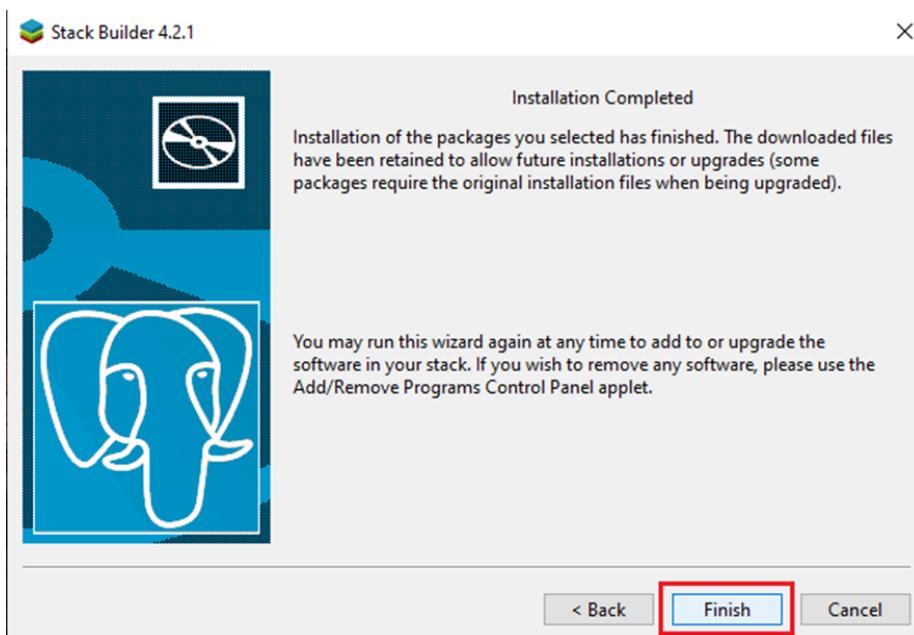
Clicaremos en "Next".



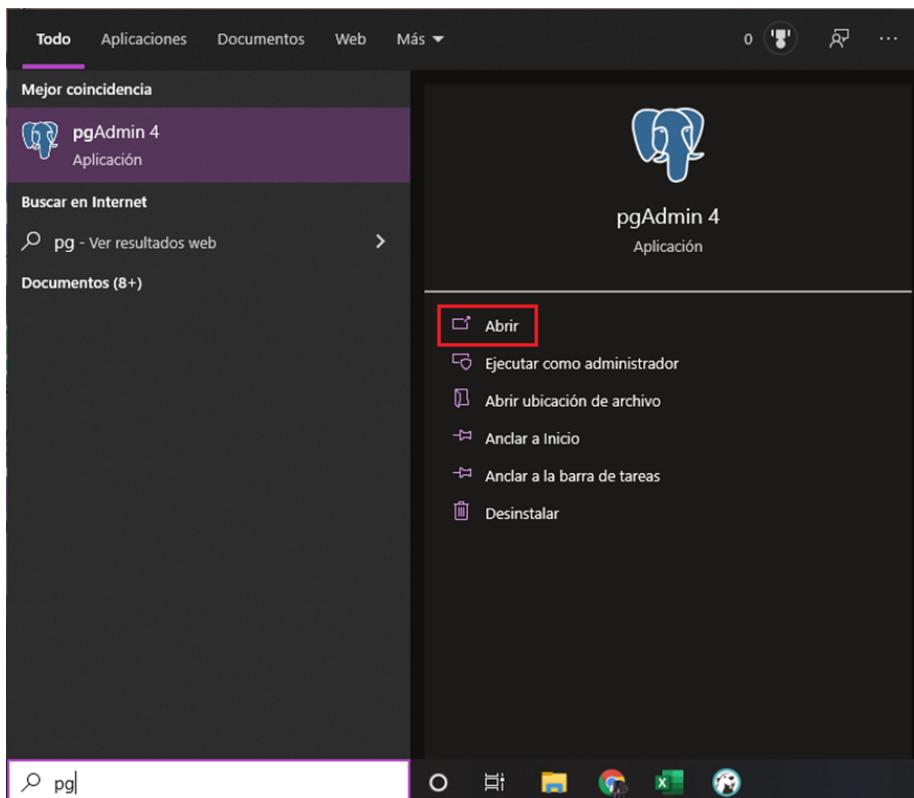
Haremos clic en "Next" y comenzará la instalación.



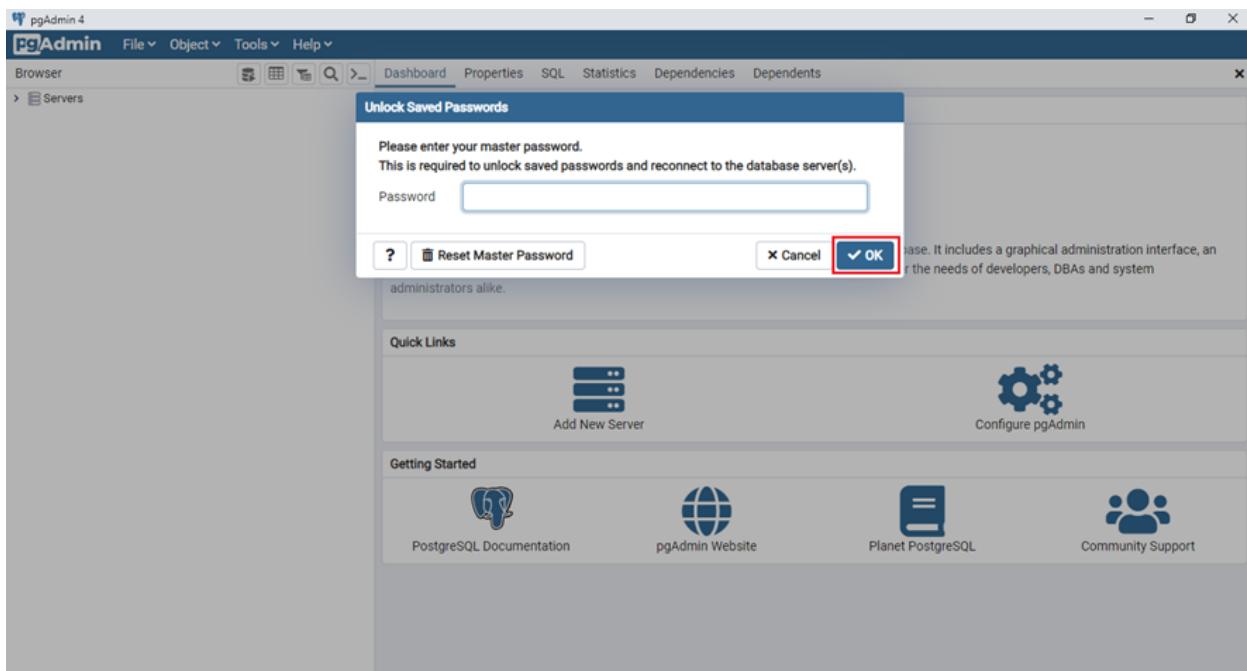
Una vez finalice el proceso clicaremos en "Finish".



En la ventana anterior haremos clic en "Finish" y ya habremos terminado la instalación.

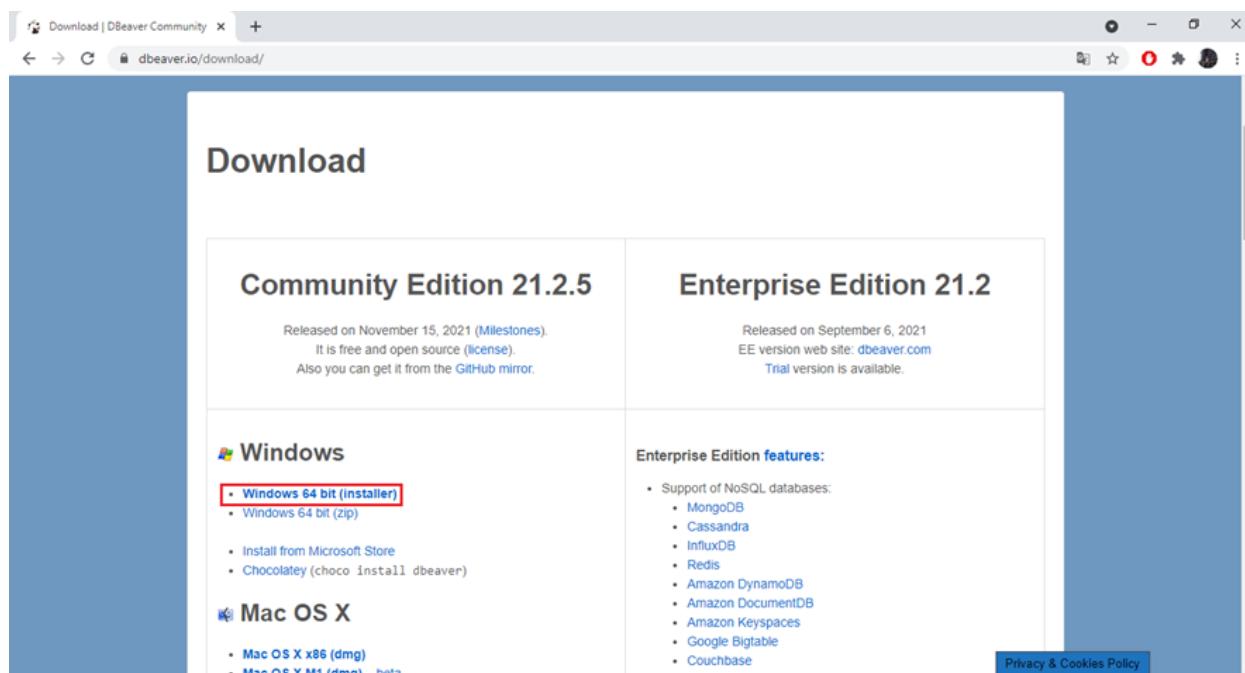


Ya podremos abrir el programa buscándolo en la barra de búsqueda en la esquina inferior izquierda de nuestra pantalla.

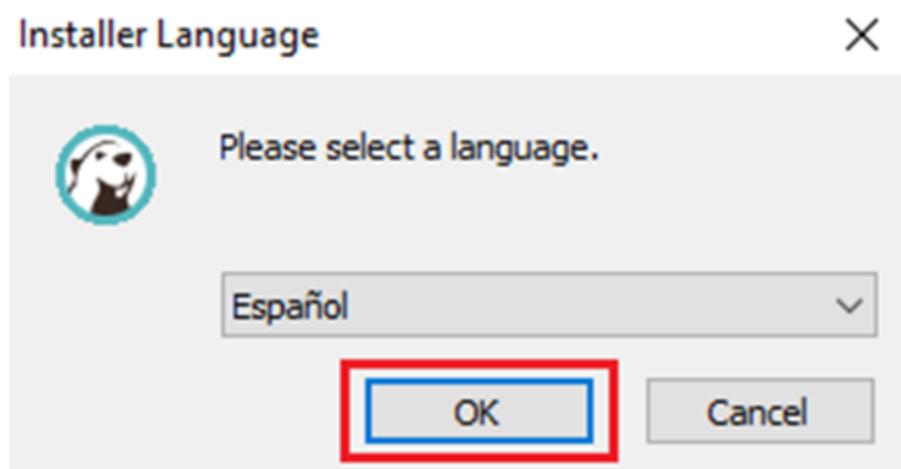


Nada más iniciarse nos solicitará que establezcamos una contraseña que será necesaria para ciertas acciones como, por ejemplo, acceder al servidor. Es importante no olvidarla.

## DBeaver



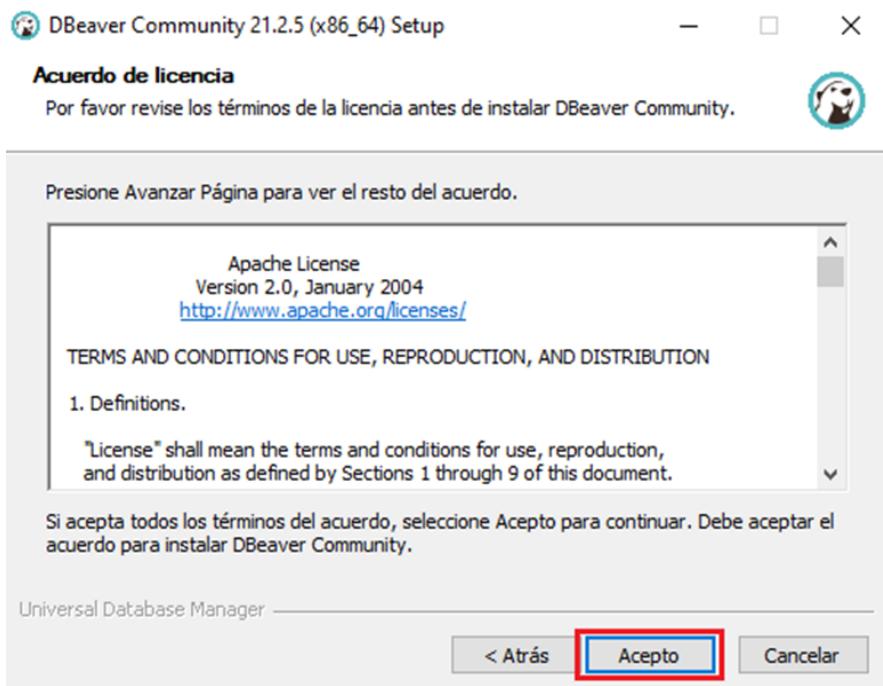
Podemos descargar DBeaver desde su web oficial. Accederemos a través del siguiente [enlace](#) y haremos clic en "Windows 64 bit (installer)".



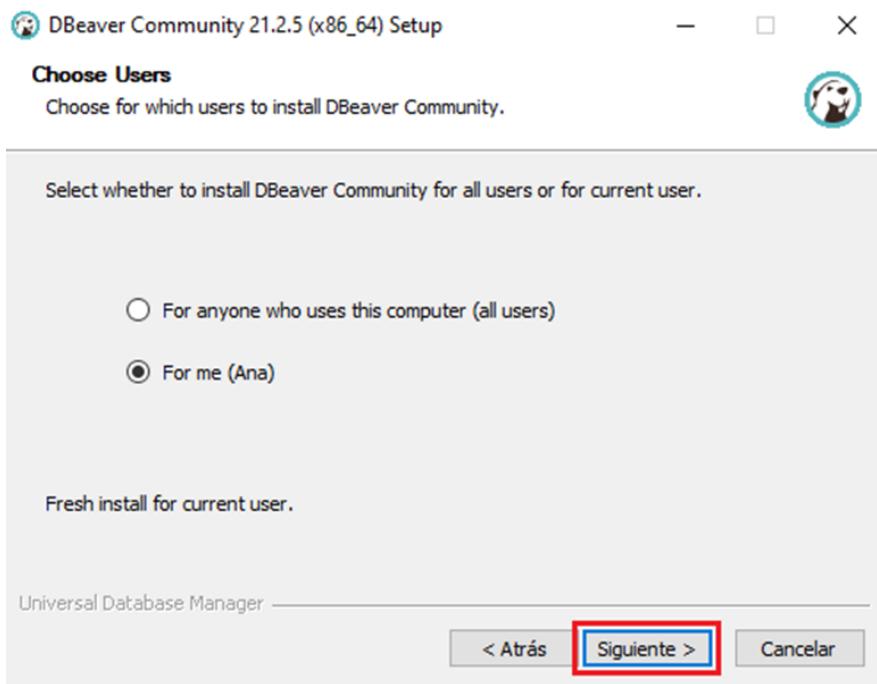
Una vez se haya descargado el archivo .exe lo ejecutaremos. Aparecerá una ventana como la anterior, seleccionaremos el idioma deseado y haremos clic en "OK".



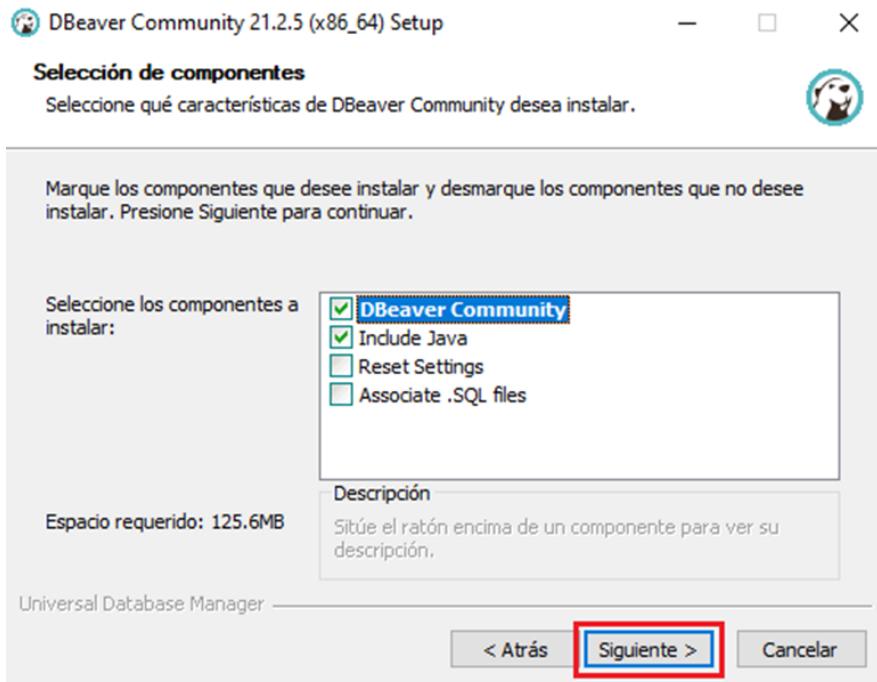
Clicaremos en "Siguiente".



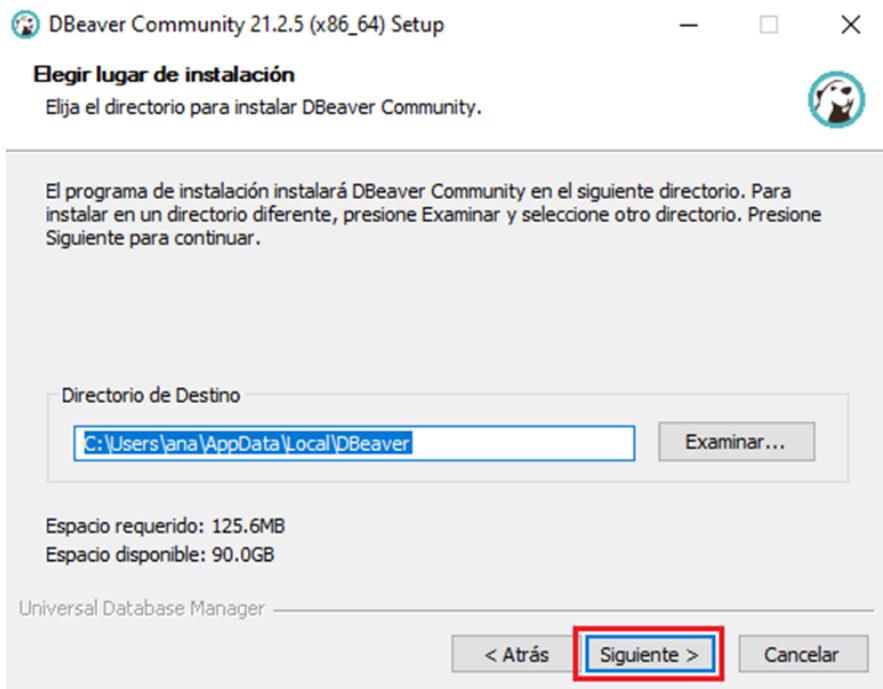
Aceptaremos la licencia haciendo clic en "Acepto".



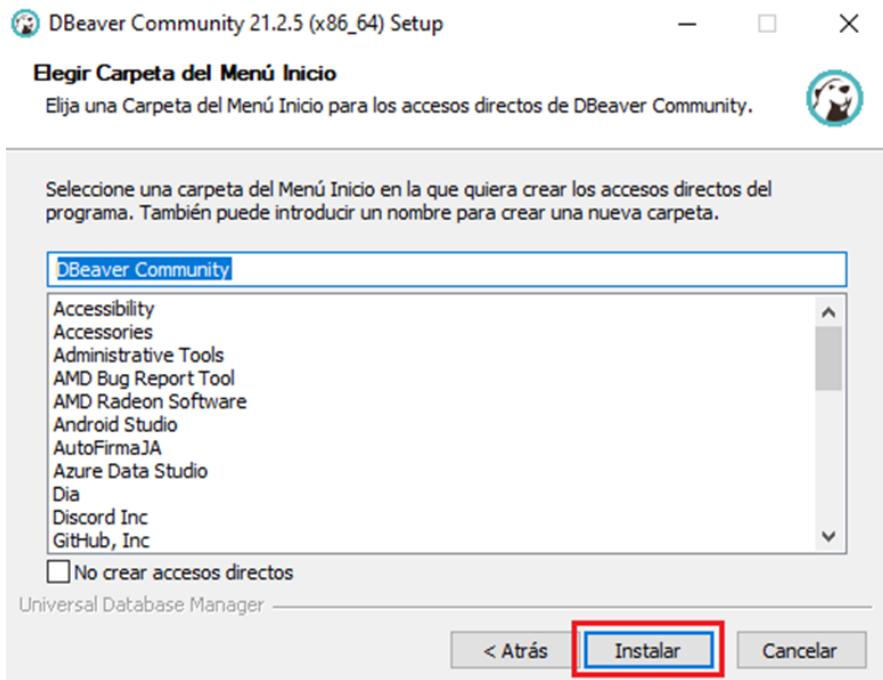
Elegiremos la opción “For me” y haremos clic en “Siguiente”.



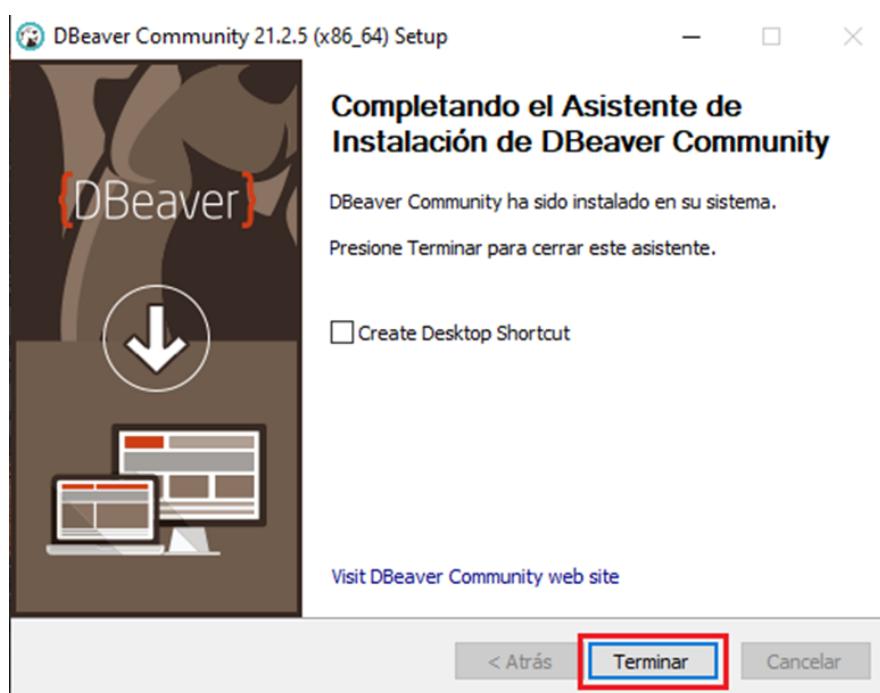
Nos aseguraremos de que están seleccionadas los componentes “DBeaver Community” y “Include Java” y clicaremos en “Siguiente”.



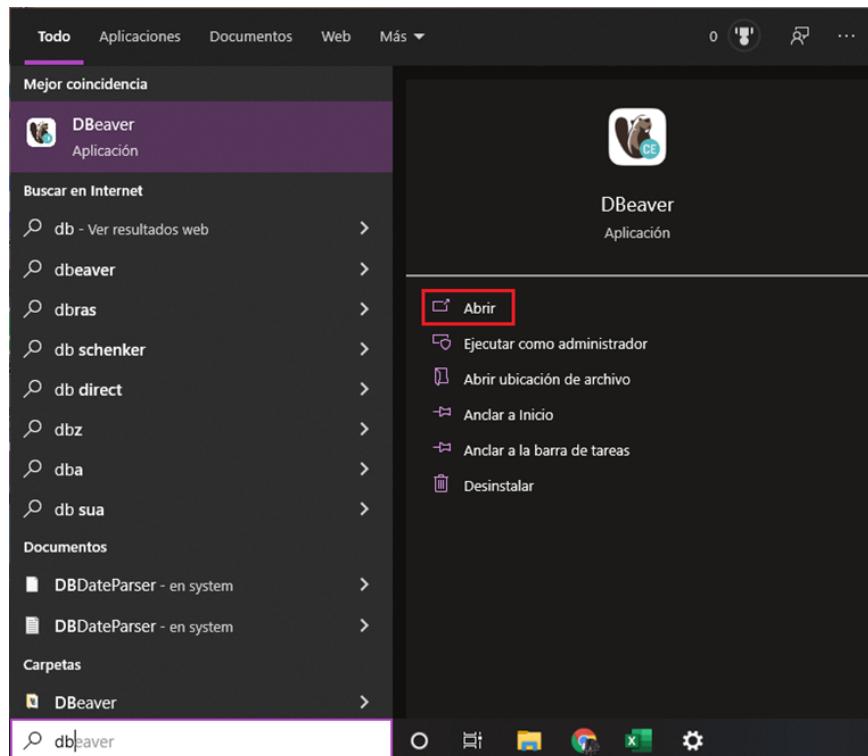
Elegiremos el directorio donde deseemos que se instale DBeaver y clicaremos en "Siguiente".



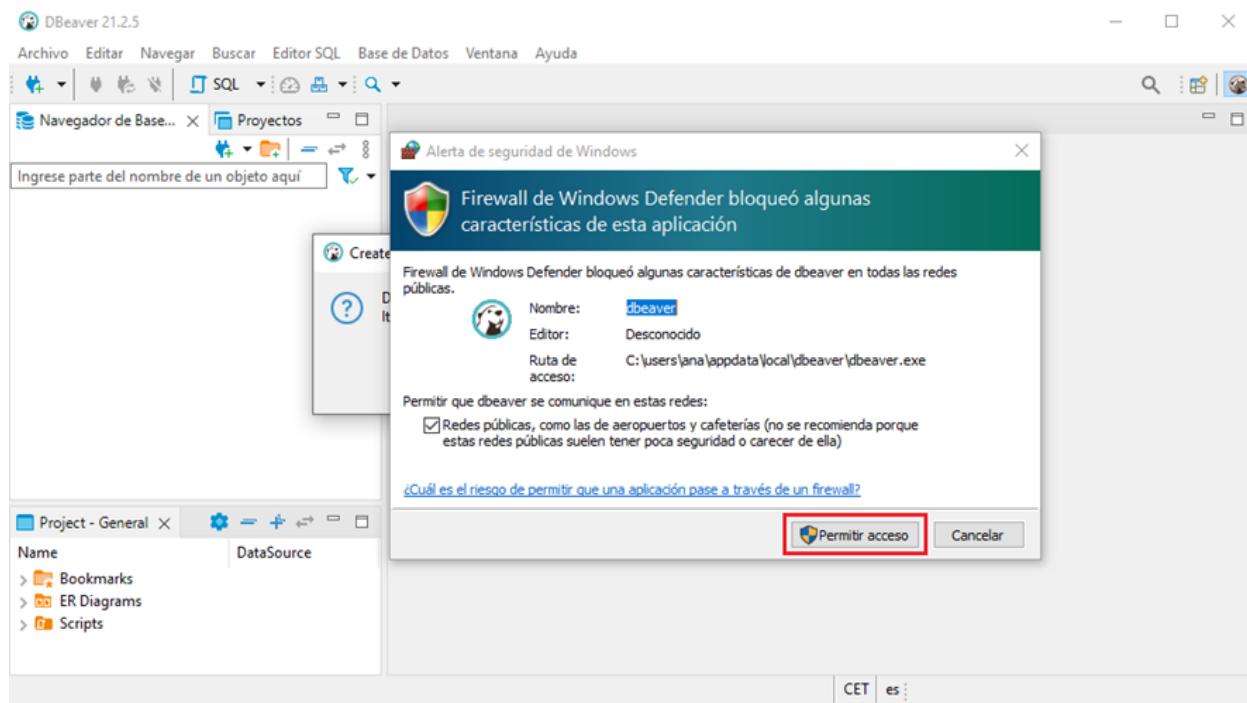
Haremos clic en "Instalar" para comenzar la instalación.



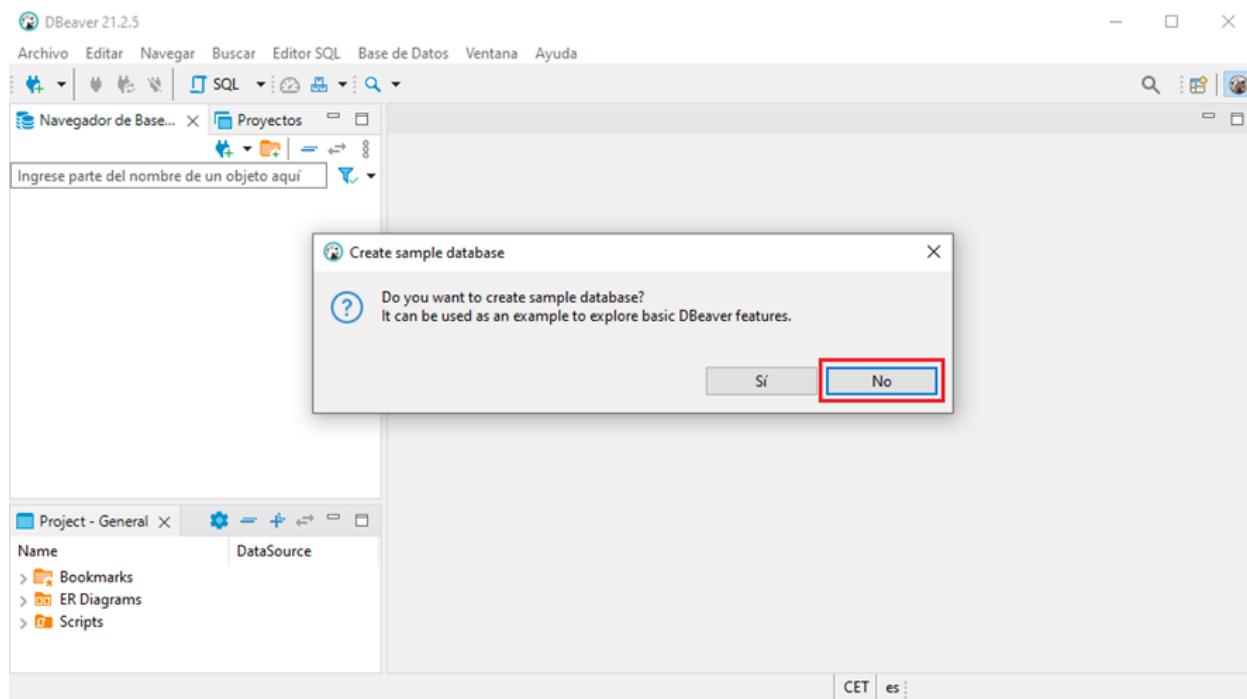
Una vez finalice el proceso haremos clic en “Terminar”.



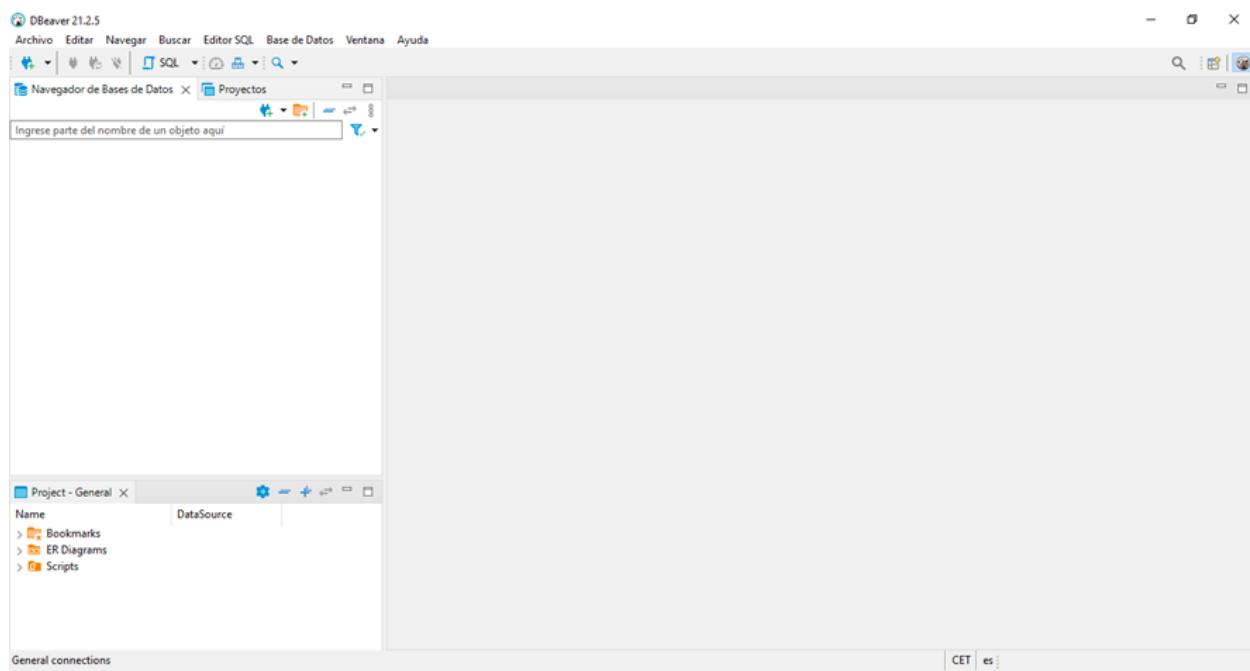
Ya podremos abrir el programa buscándolo en la barra de búsqueda en la esquina inferior izquierda de nuestra pantalla.



Si nos aparece una ventana como la anterior al abrirlo clicaremos en “Permitir acceso”.

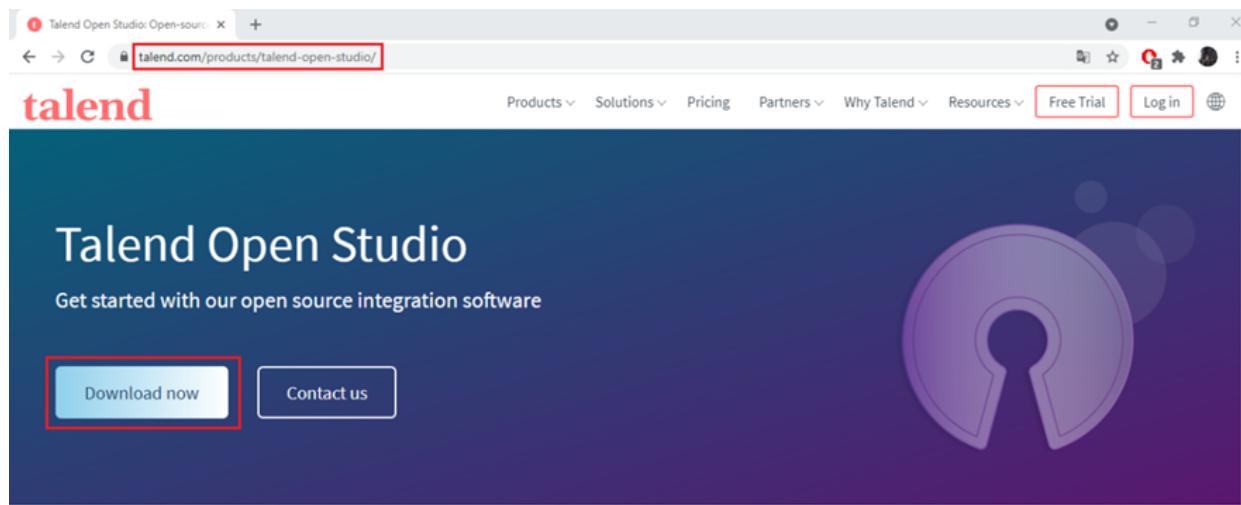


Denegaremos la opción de crear una base de datos de ejemplo haciendo clic en “No” de la ventana emergente.



DBeaver estará listo para comenzar a trabajar.

## Talend Open Studio



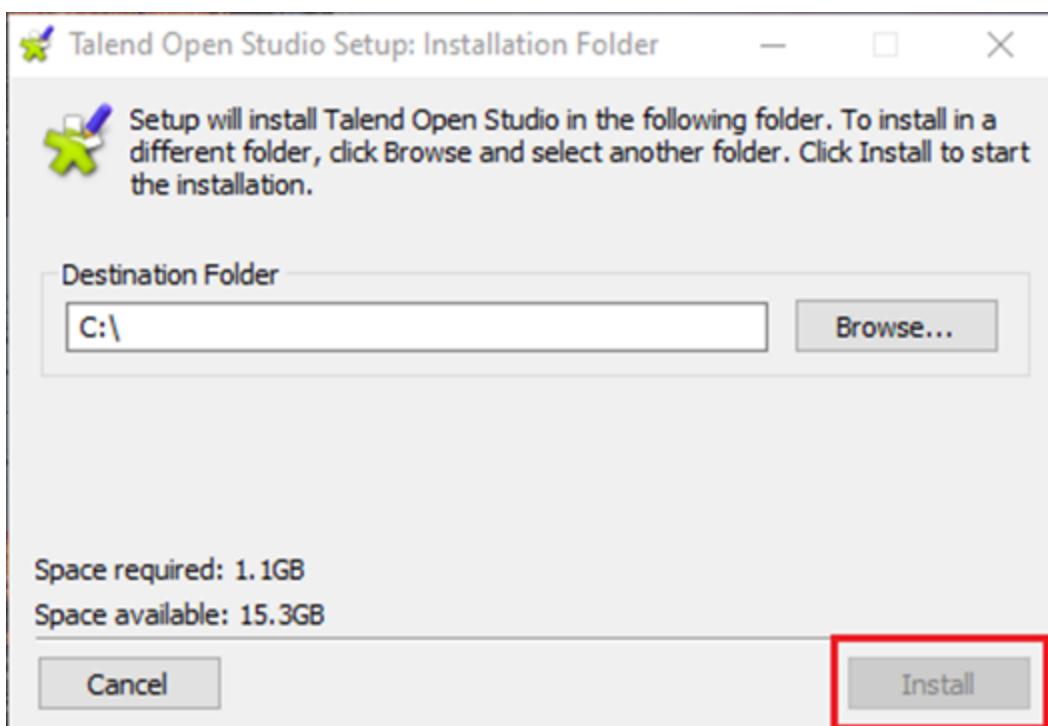
Take your data for a spin

With Talend Open Studio, you can begin building basic data pipelines in no time. Execute simple ETL and data integration tasks. [get graphical](#)

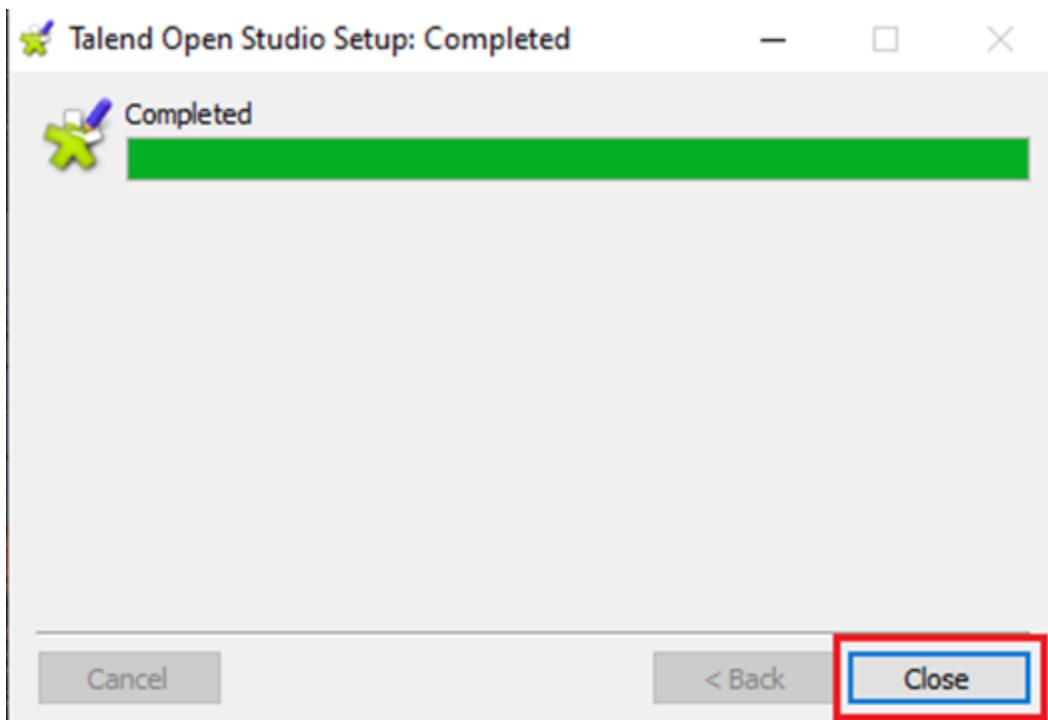
Podemos descargar Talend desde su web oficial. Accederemos a través del siguiente [enlace](#) y haremos clic en "Download now".

privacy policy.' There's also a 'Contact sales' button in the top right corner." data-bbox="115 538 875 852"/&gt;

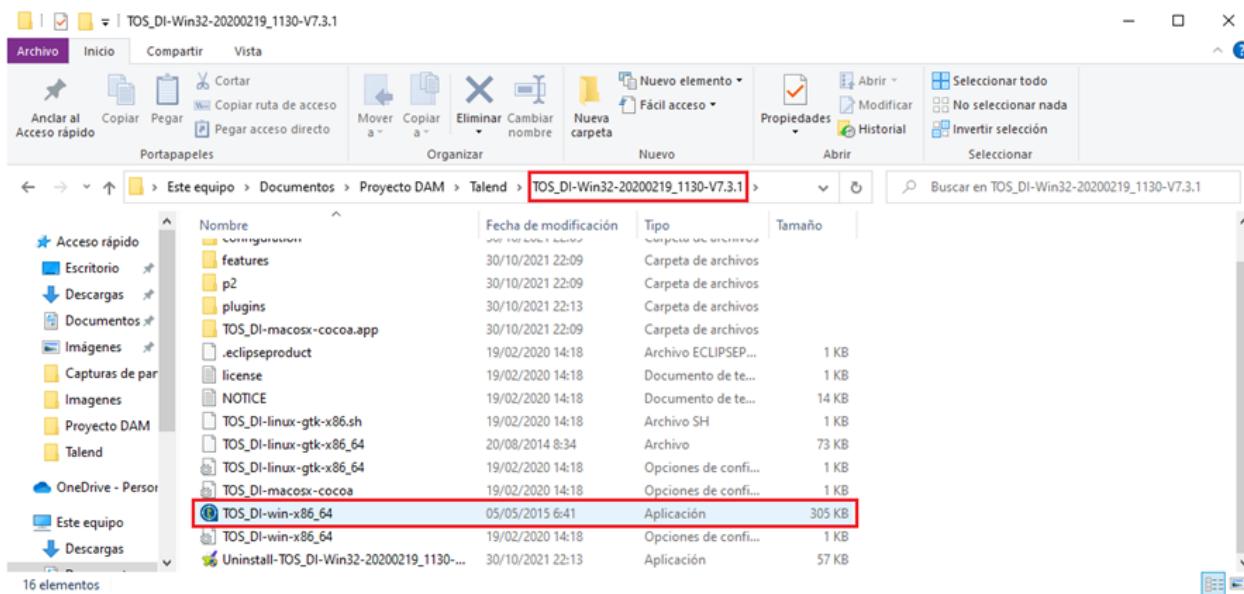
Rellenaremos el formulario y clicaremos en "Download Now".



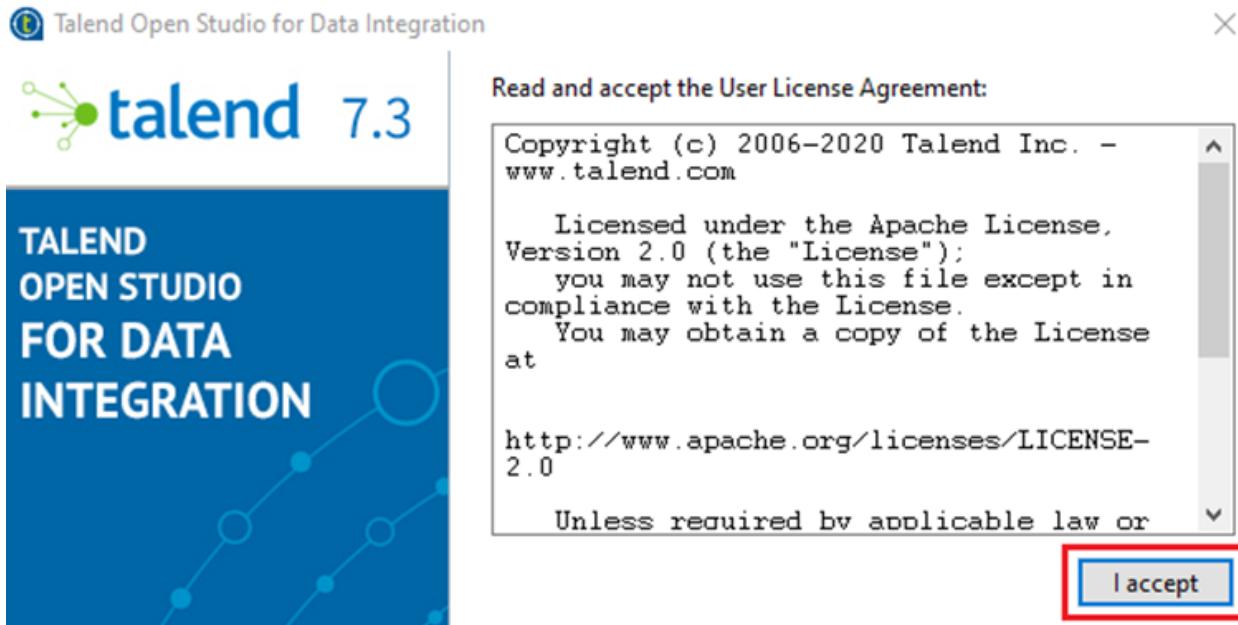
Una vez se haya descargado el archivo .exe lo ejecutaremos. Aparecerá una ventana como la anterior, seleccionaremos la carpeta de destino y haremos clic en "Install".



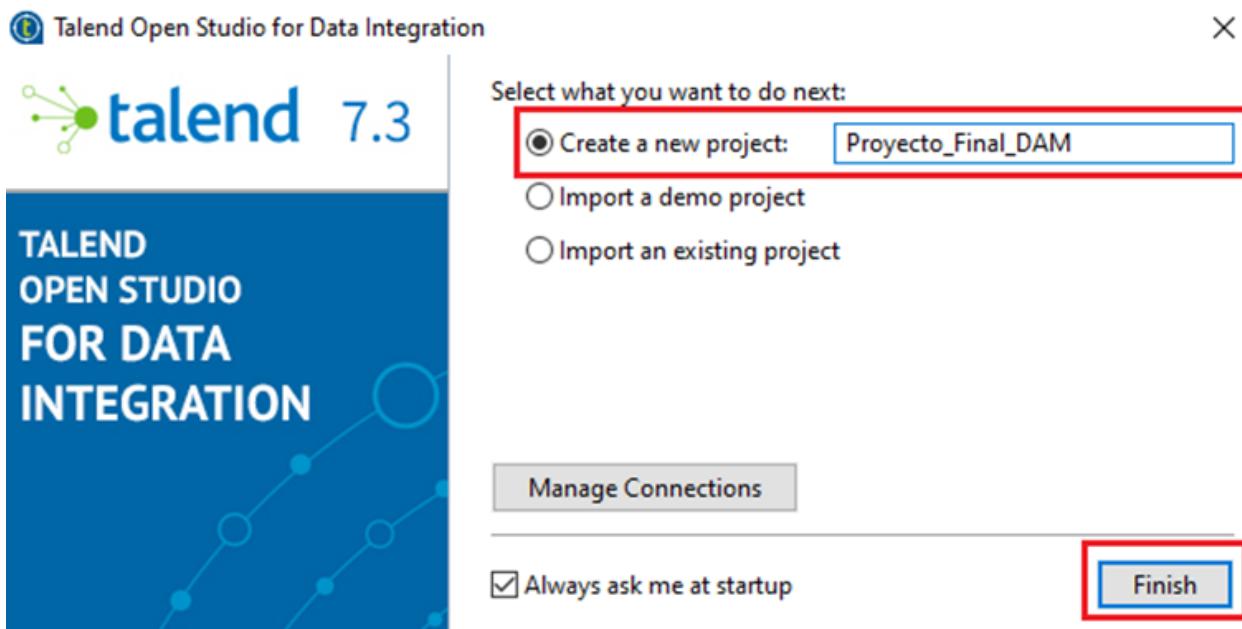
Cuando haya finalizado haremos clic en "Close".



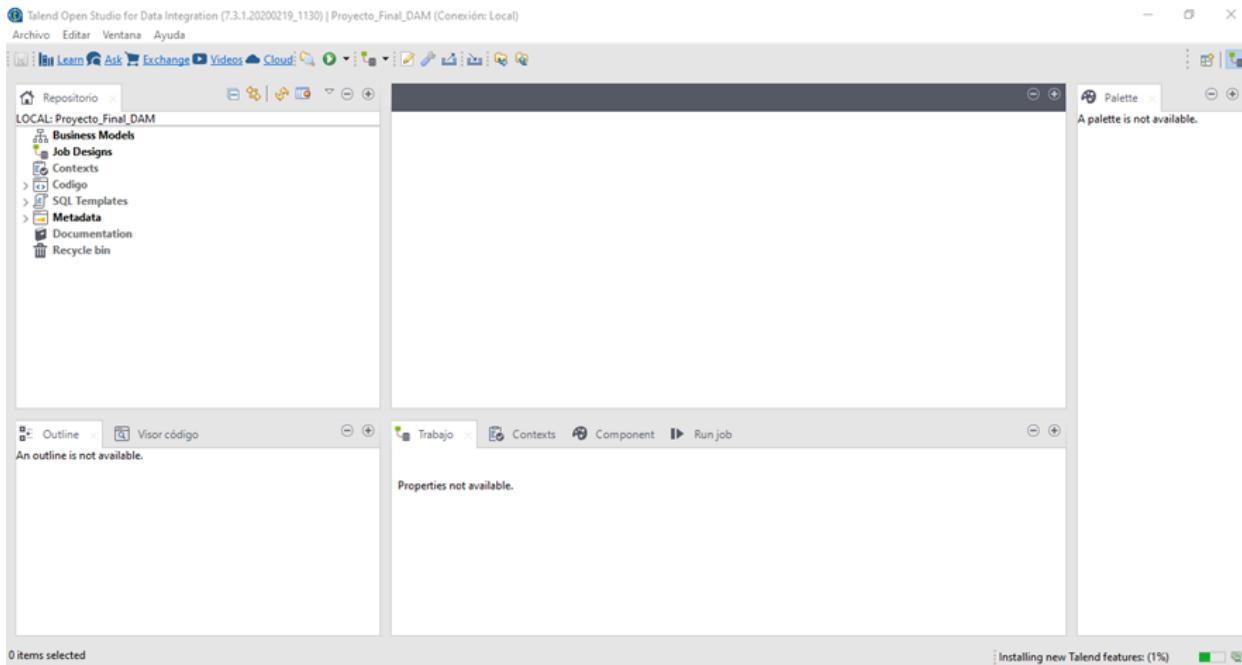
Se creará una carpeta en el directorio indicado. Dentro de esta se encuentra el archivo ejecutable con nombre "TOS\_DI-win-x86\_64.exe". Haremos doble clic sobre él para ejecutarlo. Esta operación hay que realizarla cada vez que queramos abrir Talend.



Nos aparecerá esta ventana y haremos clic en "I accept".



Seleccionaremos la opción “Create a new project” para crear un proyecto nuevo e introduciremos el nombre que deseemos para este. Una vez hecho esto haremos clic en “Finish”.

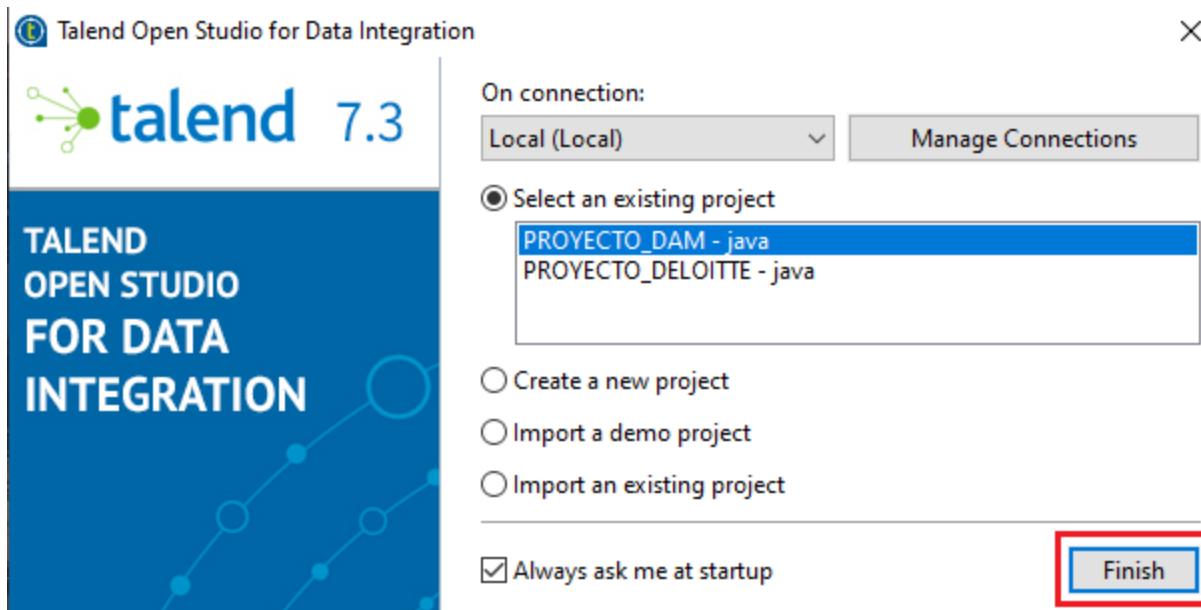


Se abrirá Talend listo para comenzar a trabajar.

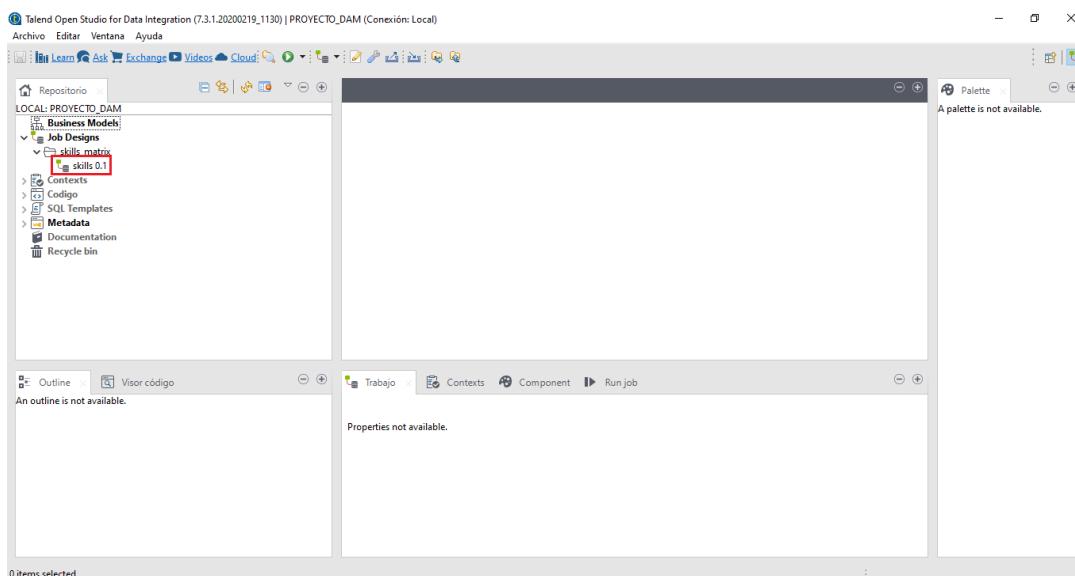
## PUESTA EN MARCHA

Para ejecutar la ETL tenemos dos opciones: si estamos desarrollandola deberemos abrir Talend y ejecutar el job y en caso de que tengamos la ETL lista utilizaremos el ejecutable.

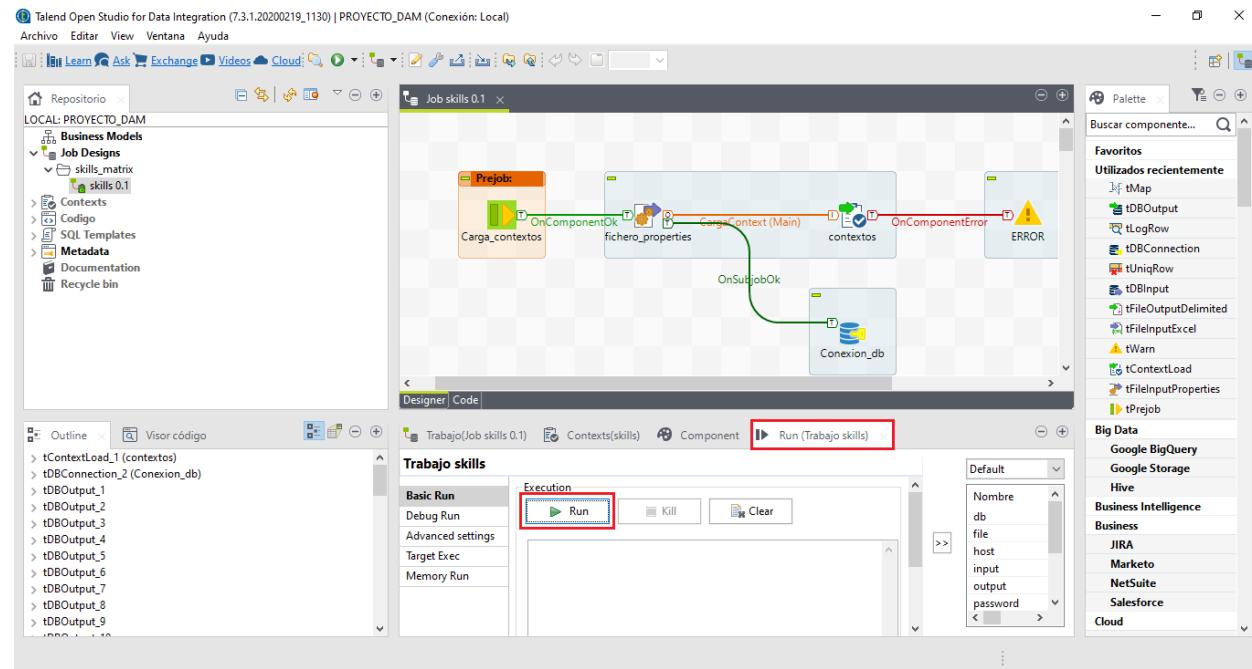
Si queremos realizar la primera abriremos Talend como está explicado en el punto anterior y en lugar de crear un proyecto nuevo, abriríamos el proyecto correspondiente.



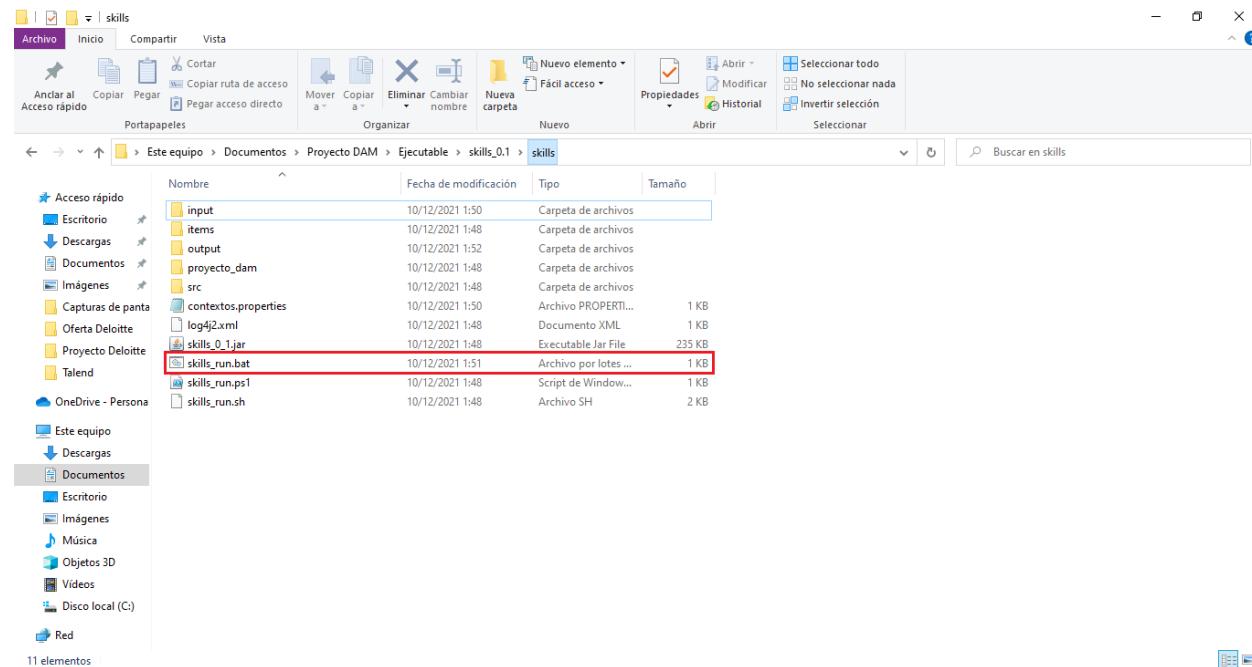
Una vez se abra el proyecto desplegaremos los jobs y haremos doble clic en la carpeta y job correspondientes.



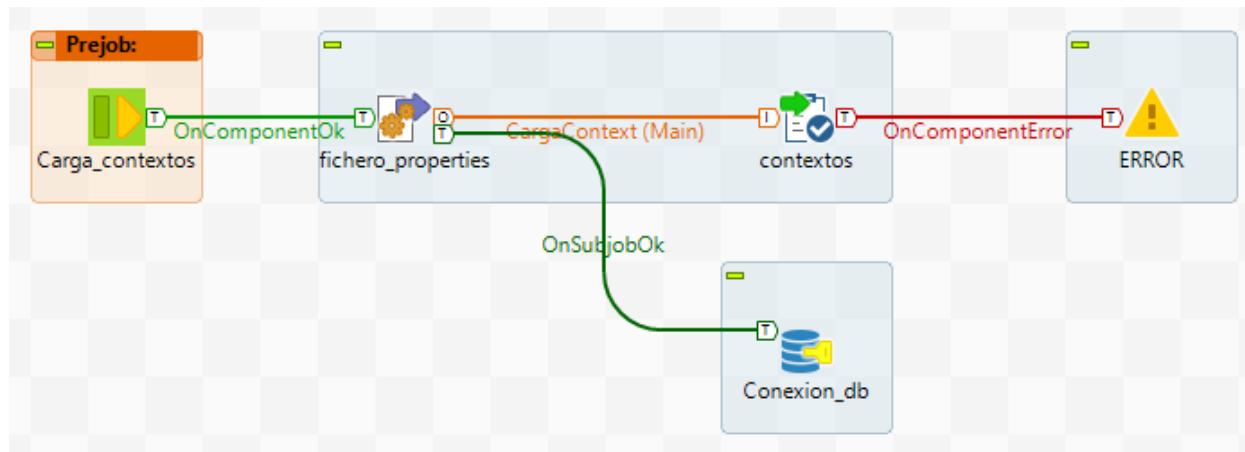
Cuando se haya abierto el job nos dirigiremos a la pestaña "Run" y haremos clic en "Run".



En el caso de que queramos utilizar el ejecutable lo único que tendremos que hacer es clicar dos veces sobre el archivo llamado "skills\_run.bat" cuya ruta es "C:\Users\ana\Documents\Proyecto DAM\Ejecutable\skills\_0.1\skills" o ejecutarlo desde la consola cmd.



## EXPLICACIÓN DEL FUNCIONAMIENTO DE LA APLICACIÓN



Comenzaremos por el componente "Prejob" cuya función es asegurar que primero de todo se ejecute la carga de las variables del fichero properties al contexto.

A continuación nos encontramos con el componente "File Input Properties". Este hace referencia al archivo .properties.

Imagen capturada de Talend Open Studio. Se muestra el diseño de un trabajo llamado "Job skills 0.1". En el centro, se ve el flujo de datos que incluye un "Prejob" (Carga\_contextos), un "fichero\_properties", un "CargaContext (Main)", un "contextos", un "Conexion\_db" y un "ERROR". En la parte inferior izquierda, el panel "Designer" muestra la configuración del componente "fichero\_properties(tFileInputProperties\_1)", donde se especifica el nombre del archivo como "C:/Users/ana/Documents/Proyecto DAM/Talend/contextos.properties". La paleta de componentes en la derecha incluye categorías como "Favoritos", "Utilizados recientemente" y "Big Data".

El contenido de este archivo es el siguiente:

```

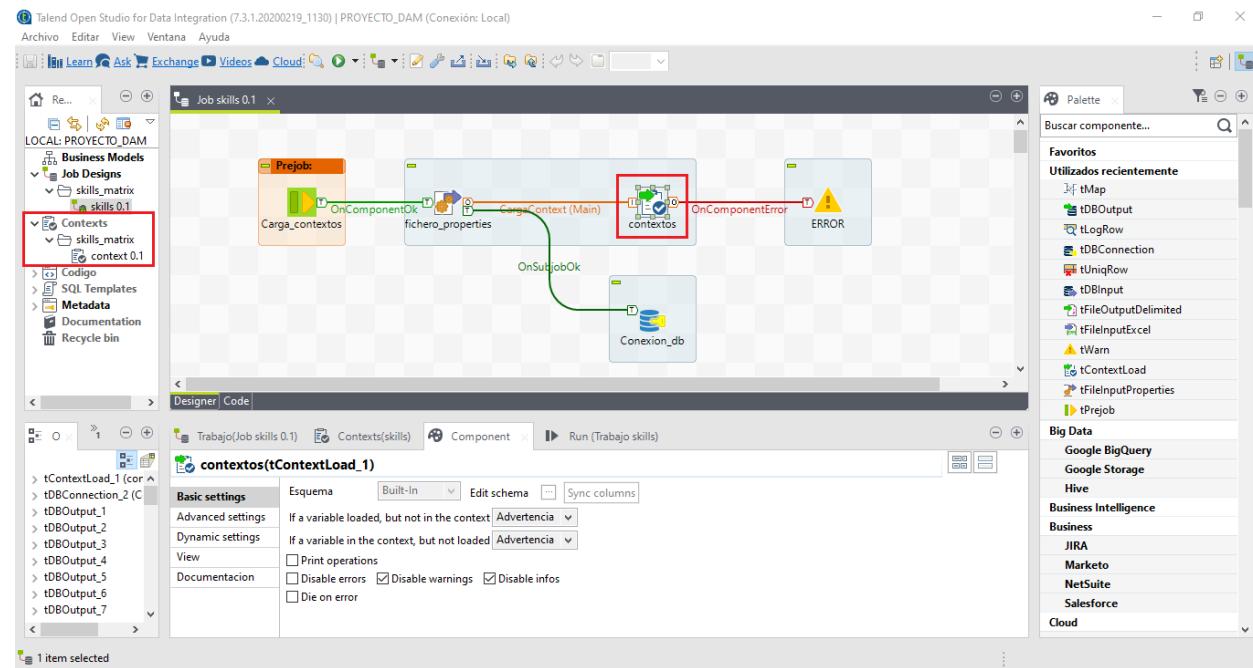
contextos.properties: Bloc de notas
Archivo Edición Formato Ver Ayuda
#Rutas archivos
input=C:/Users/ana/Documents/Proyecto DAM/Talend/input/
output=C:/Users/ana/Documents/Proyecto DAM/Talend/output/
file=Skills_Matrix-2.0.xlsx

#Conexiones Postgres (Confirmar detalles conexion "Donde tengas la BBDD")
db=db_DC_AID
user=postgres
password=admin
schema=skills_matrix
host=localhost
port=5432

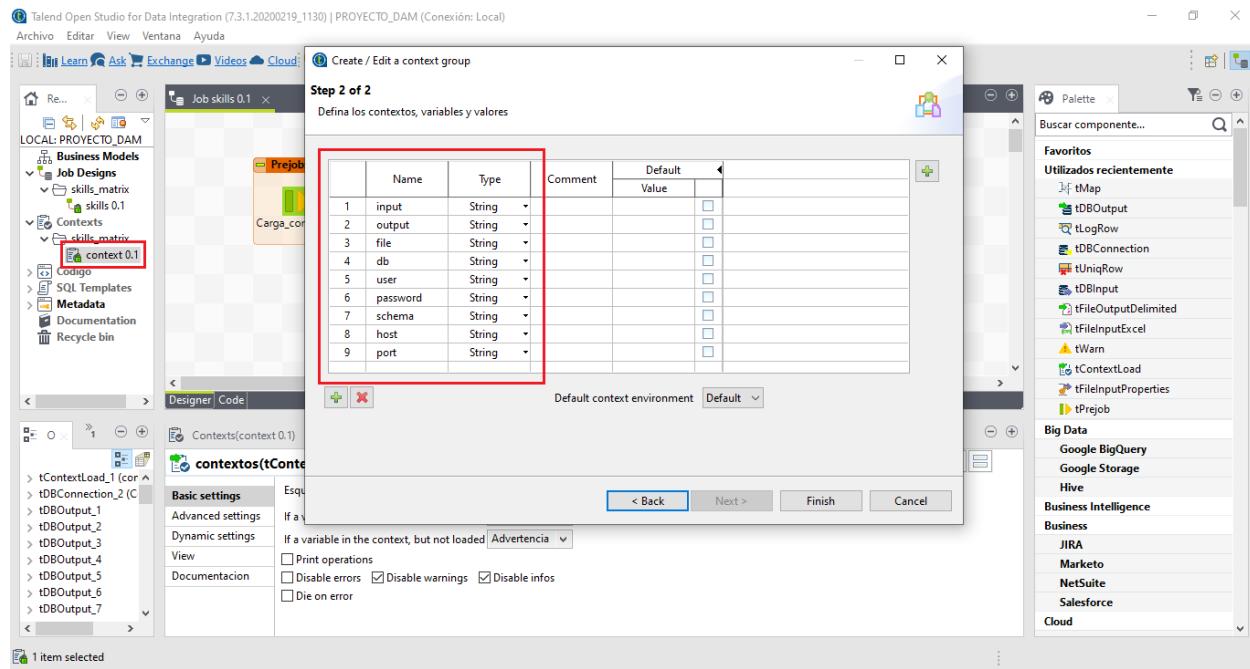
```

Como se puede observar, “fichero\_properties” está enlazado con dos componentes.

El primero es un “ContextLoad”. Este hace referencia al contexto que se ha creado con las variables correspondientes al fichero .properties. Y el enlace que tiene con “fichero\_properties” es para realizar la carga de los valores de las variables del fichero a las variables del contexto.



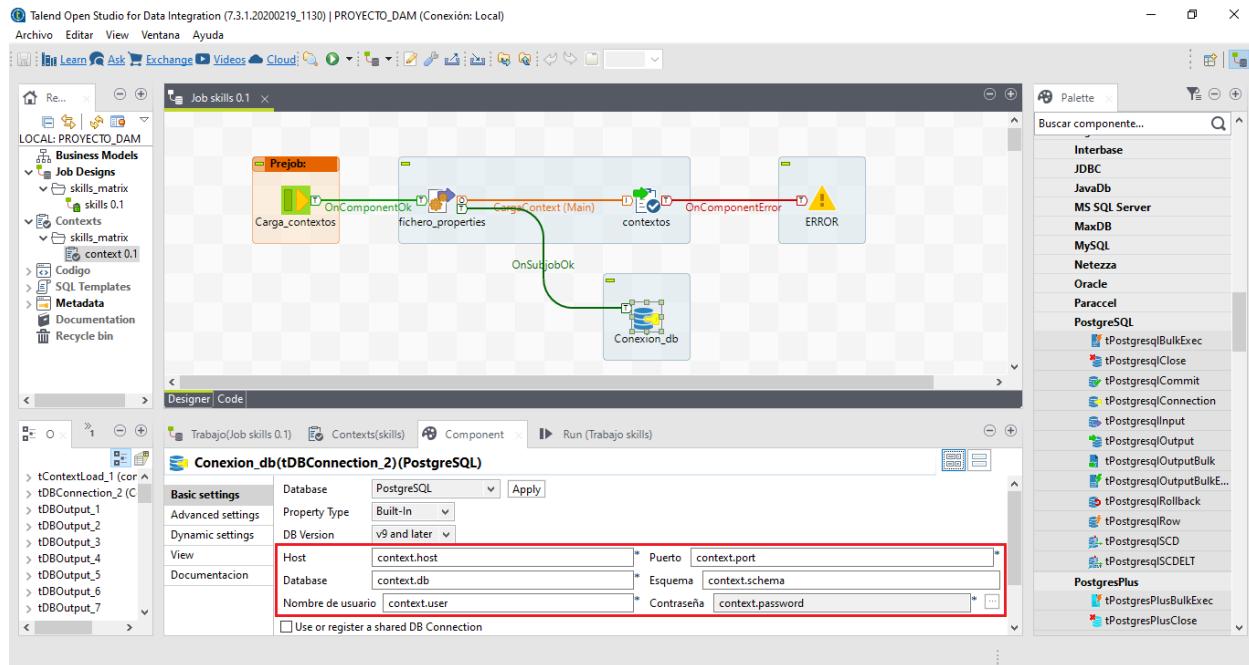
Haciendo doble clic en el contexto podemos observar las variables creadas:



El componente "ERROR" es de tipo "Warn". En este caso está enlazado con "contextos" de manera "OnComponentError" para que en el caso de que se produzca un error en la carga de las variables del contexto aparezca un mensaje.

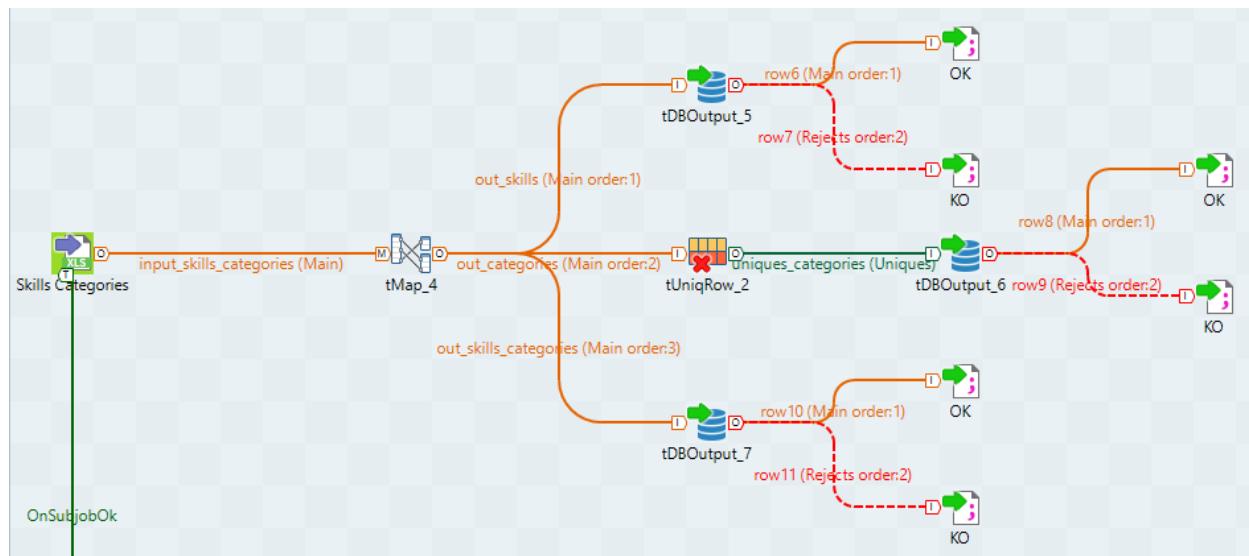
Una vez que se realice el proceso de carga de las variables, se procederá a la conexión con la base de datos. Para ello es necesario el componente "Conexion\_bd" que es de tipo "PostgresqlConnection".

En el componente para la conexión se introducen los valores necesarios (host, puerto, database...), que anteriormente hemos relacionado con las variables de contexto. Por ello, por cada valor haremos referencia a la variable correspondiente de la siguiente manera: nombredelcontexto.nombredelavariable.



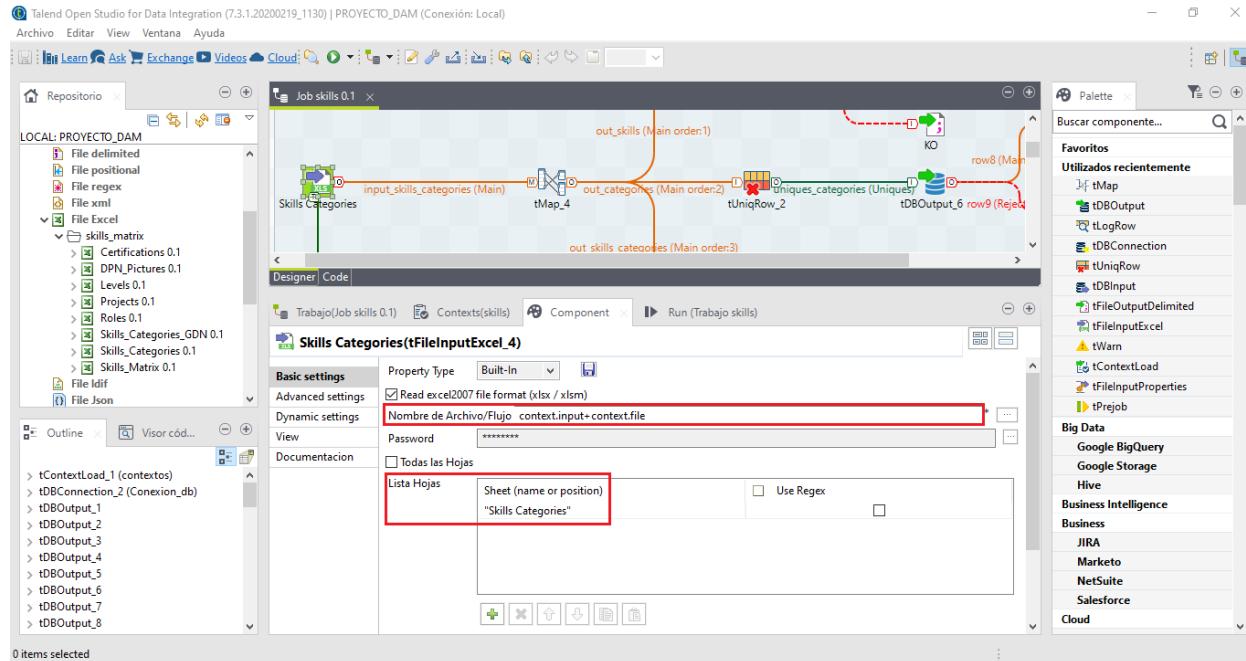
Una vez la conexión se haya realizado con éxito, comenzarán los procesos de extracción, transformación y carga de los datos.

El orden establecido a la hora de insertar los datos en la base de datos es muy importante para respetar las claves foráneas.

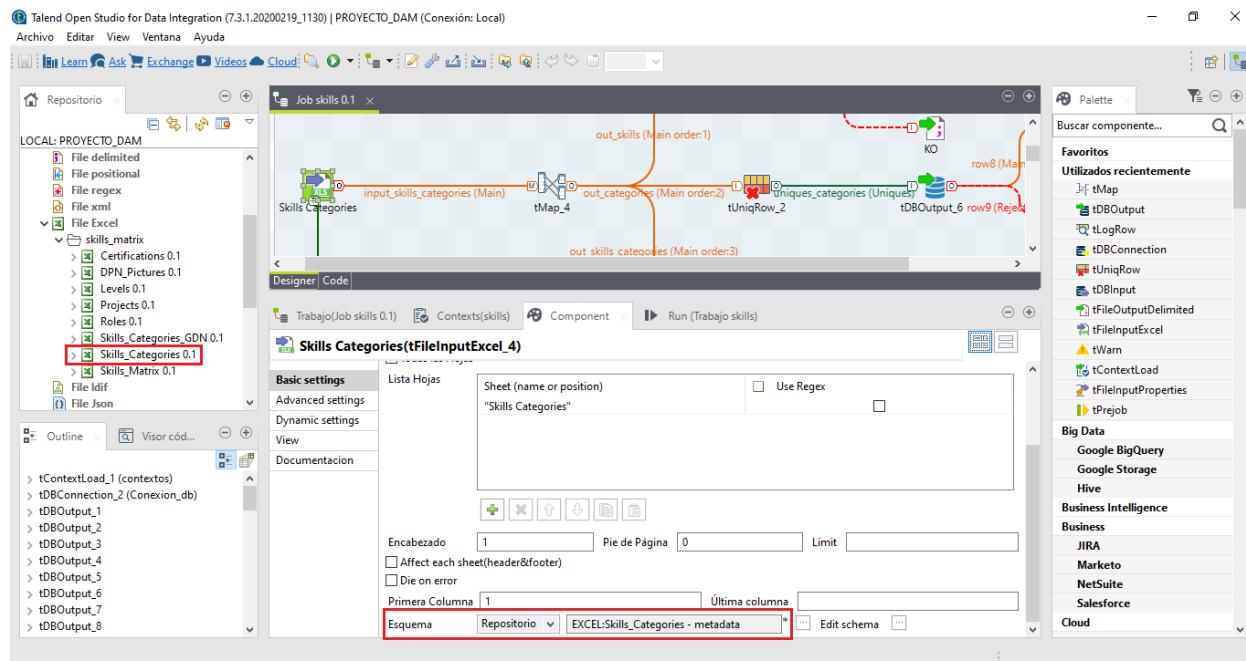


Se comienza por la hoja de Skills Categories. En este caso de esta hoja se extraen datos para tres tablas distintas. A continuación, analizaremos el proceso y los componentes.

El componente "Skills Categories" es de tipo "FileInputExcel" y hace referencia el archivo excel.



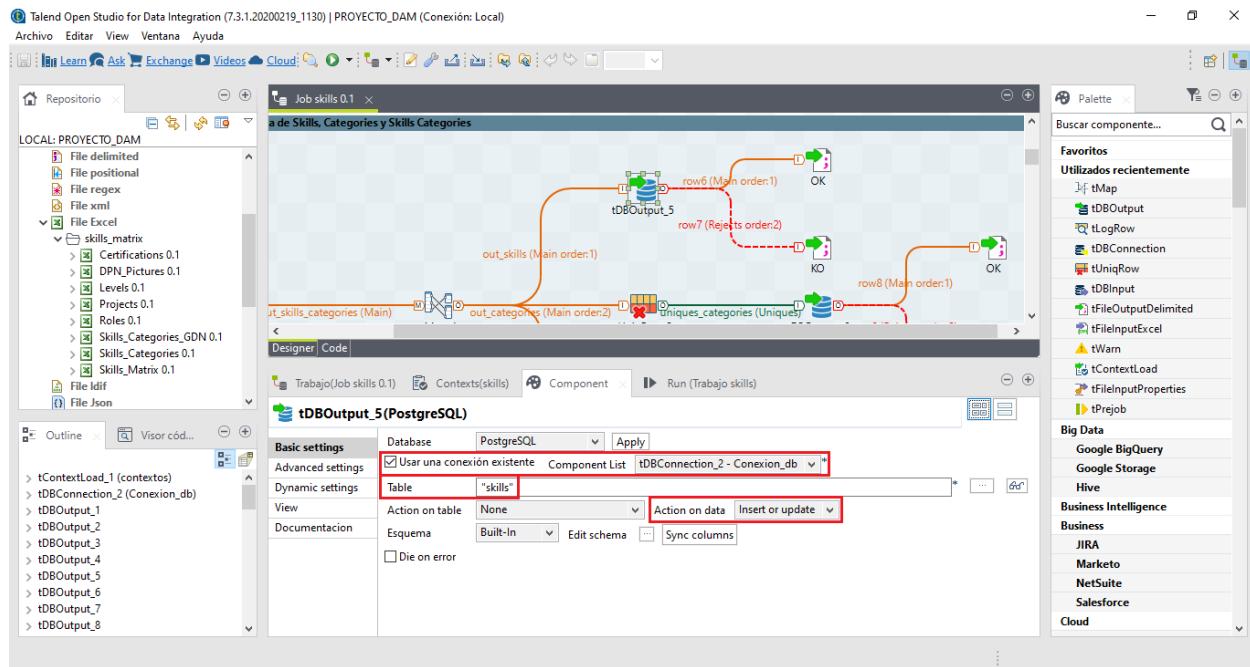
También se indica la hoja del excel correspondiente que hemos creado anteriormente en los metadatos.



A continuación nos encontramos con un componente de tipo “Map”. En el siguiente punto explicaremos con más detalle su función, pero a grandes rasgos sirve para realizar los cruces entre las diferentes columnas del excel y establecer los datos de salida a la base de datos.

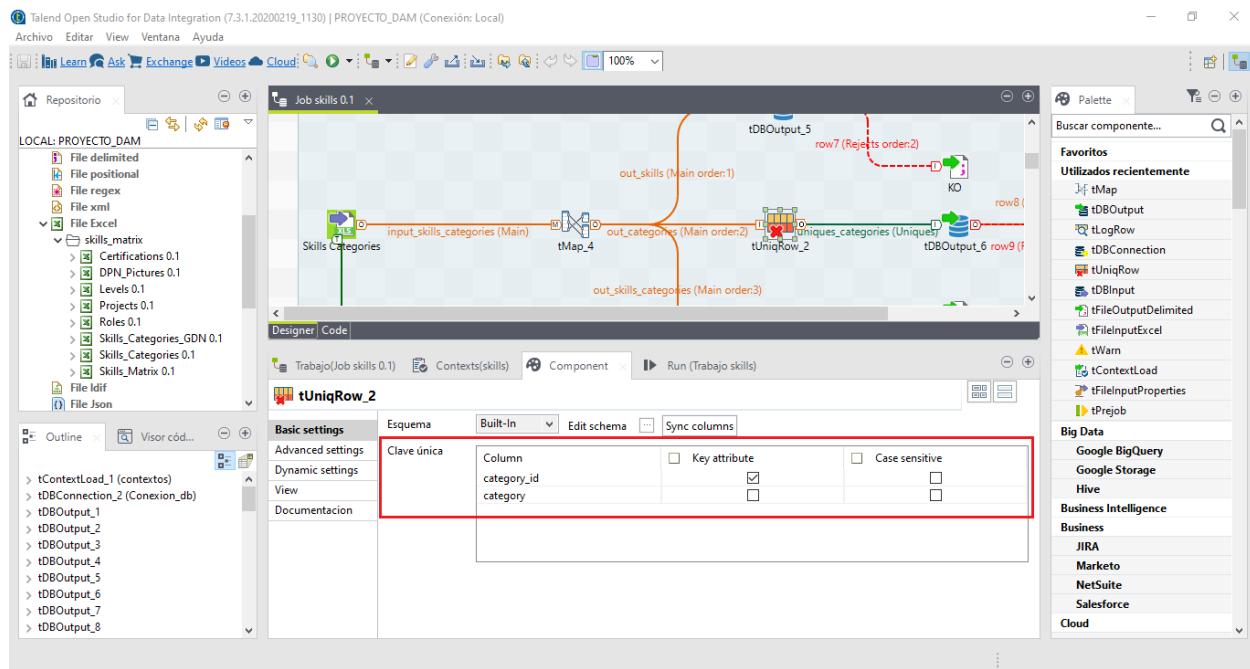
Depende de con que hoja del excel tratemos, nos encontraremos con uno o varios componentes tras mapeo. En el caso de la “Skills Categories” extraemos datos de esta para tres tablas distintas. Por ello, encontramos tres “ramas” diferenciadas.

Comenzaremos con la primera, que en cuanto al proceso, es igual que la tercera. Primero observamos el componente “DBOutput”. Este, utilizando la conexión creada anteriormente, realiza la inserción o actualización de los datos seleccionados en el mapeo en la tabla que indiquemos de la base de datos.



Tras el componente “DBOutput” nos encontramos con dos componentes de tipo “FileOutputDelimited”. Estos componentes son ficheros de salida donde quedan registrados, en el caso del “OK”, los datos insertados o actualizados correctamente, y en el caso del “KO”, los errores que hayan sucedido (por ejemplo errores de claves foráneas).

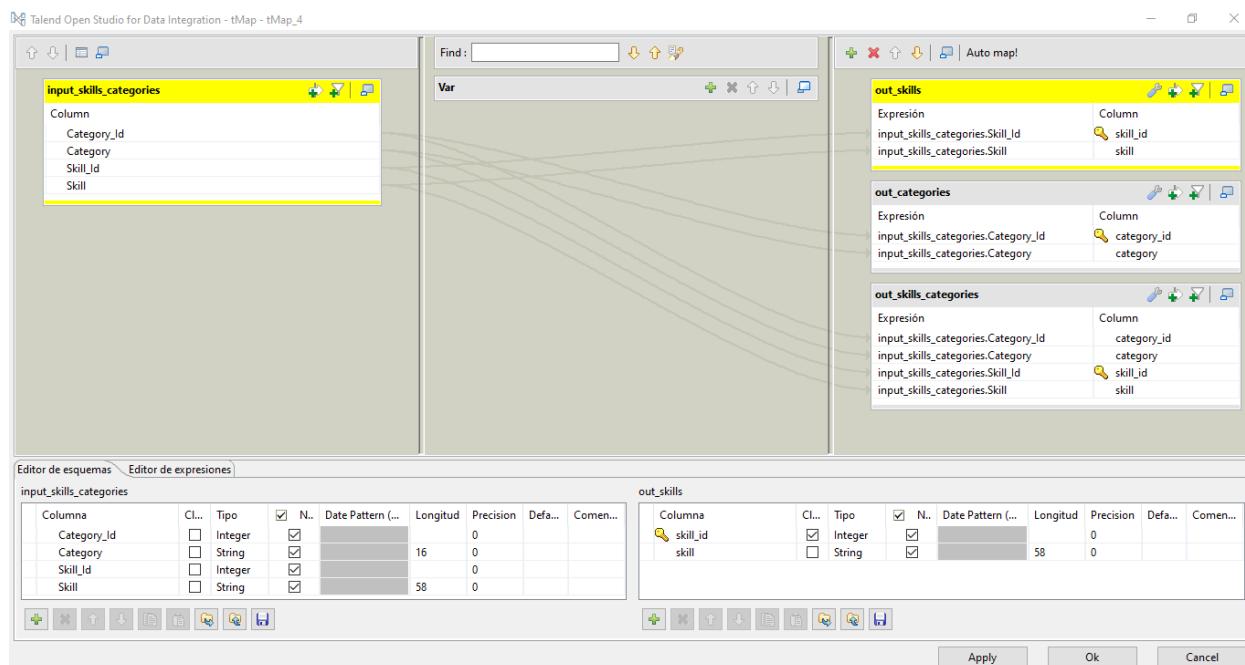
Pasamos a la segunda rama, la única diferencia con la primera y la tercera es que esta cuenta con un componente de tipo “UniqRow” entre el mapeo y la salida a la base de datos. Este componente se encarga de seleccionar las filas únicas por el campo que indiquemos.



Este proceso se repite con todas las hojas del excel (excepto “Template”). En el caso de la hoja “Skills Categories GDN”, de ella solo extraemos datos para una tabla, “skills\_categories\_gdn”. De “Roles”, extraemos datos para la tabla “roles”. De “Levels”, para la tabla “levels”. De “Skills Matrix”, para las tablas “practitioners” y “skills\_matrix”. De “Projects”, para “projects”. De “Certifications”, para “certifications”. Y de “DPN Pictures”, para “dpn\_pictures”.

## EXPLICACIÓN DEL CÓDIGO O PROCESOS MÁS COMPLEJOS O INTERESANTES

Consideramos que el proceso o procesos más complejos e interesantes son los mapeos. Como ejemplo vamos a utilizar el componente “Map” de la hoja “Skills Categories”.



Si hacemos doble clic en el componente nos aparecerá esta ventana. Podemos observar a la izquierda el esquema del input (hoja excel) y a la derecha el esquema de la o las salidas.

En el caso de la hoja “Skills Categories” podemos ver tres salidas distintas debido a que se realiza inserción y modificación de los datos en tres tablas diferentes. En este componente es donde establecemos que columnas queremos que vayan a que salidas o tablas y además indicamos las claves primarias.

Es importante que nos aseguremos que en la salida los nombres y los tipos de las columnas sean iguales a los nombres y tipos de estas en las tablas de la base de datos.

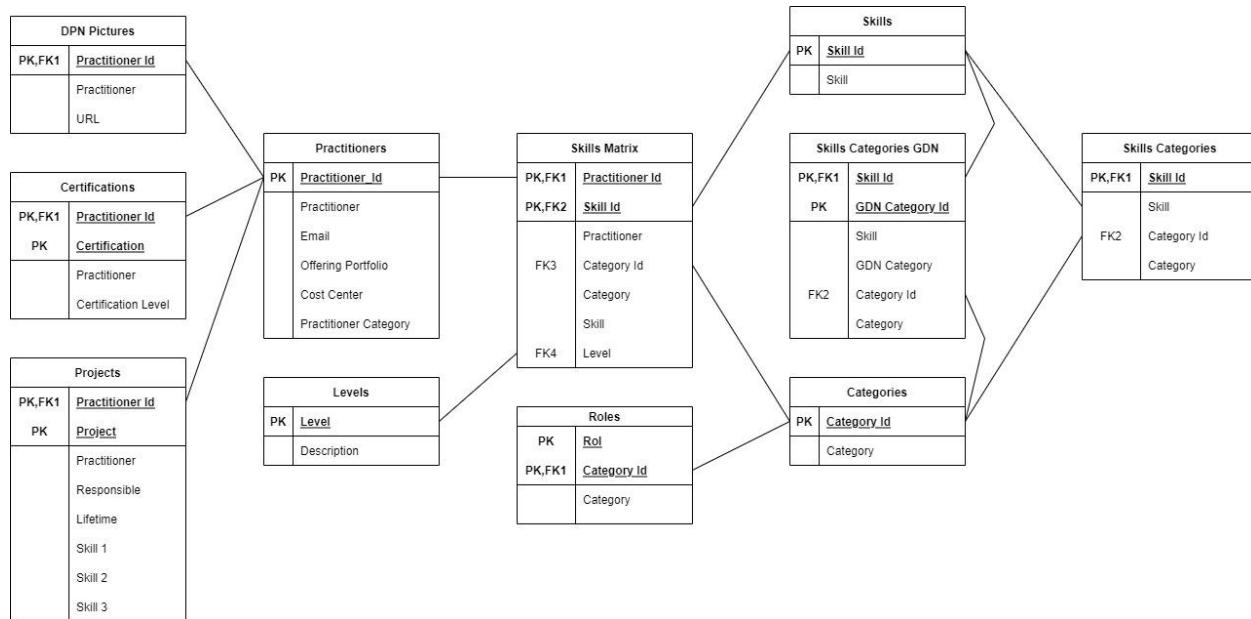
## MAYORES DIFICULTADES ENCONTRADAS

La mayor dificultad encontrada ha sido el diseño de la base de datos, ya que el archivo excel del cual extraemos los datos está ideado para que los empleados lo rellenen fácilmente y no está optimizado. Por ello, a la hora de diseñar la base de datos no era suficiente con crear una tabla por cada hoja del archivo. De esta manera encontraríamos inconvenientes con las claves primarias y foráneas de las tablas.

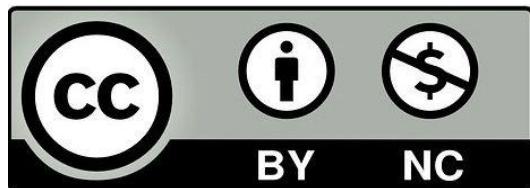
Para poder construir una base de datos óptima estudiamos los datos existentes y las necesidades de la empresa y planteamos los requisitos a cumplir para el futuro del proyecto. El modelo presenta duplicidad debido a que esta base de datos será utilizada para la visualización de los datos con alguna herramienta como Power BI y es necesario presentar los datos de esta manera.

En definitiva, la mayor dificultad encontrada ha sido diseñar una base de datos lo más optimizada posible pero que a la vez cumpliese todos los requisitos para su utilización en Power BI.

El modelo creado es el siguiente:



## LICENCIA



Esta obra está bajo una [Licencia Creative Commons Atribución-NoComercial 4.0 Internacional.](#)