



Data Management

KeepCoding

Fechas: 12 y 13 de Diciembre de 2019.

Horas: 8h

Instructora: Nerea Sevilla

Nombre Material: Ejercicios_Talend_DataIntegration

Talend Data Integration

Talend es una solución que se inició en el mercado con una solución de ETL (extracción, transformación y carga) pero poco a poco esta solución ha ido evolucionando, implementando nuevas aplicaciones, siempre concerniente al dato. Son un conjunto de aplicaciones que se centran en otra forma de extraer y transformar la información y ponerla a disposición de otras aplicaciones que puedan consumirlo.

Talend Data Integratoion dispone de una versión opensource gratuita, es una herramienta basada en eclipse (java) y está basado en los estándares. La ventaja de Talend Data Integration es que cualquier código que se desarrolla se puede exportar y ejecutar lo en cualquier ordenador que tenga una máquina virtual de java, es decir no necesitamos tener la aplicación para poder ejecutar los procesos ya implementados.

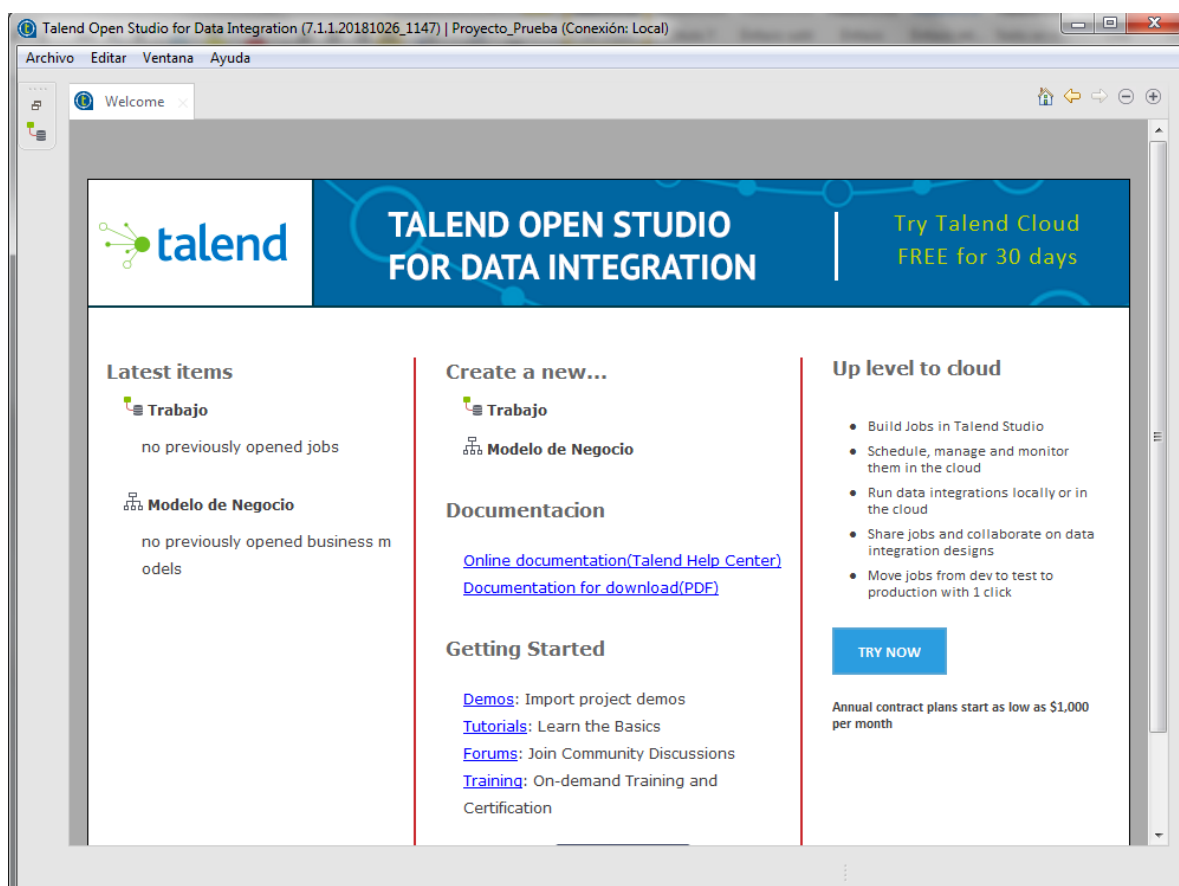
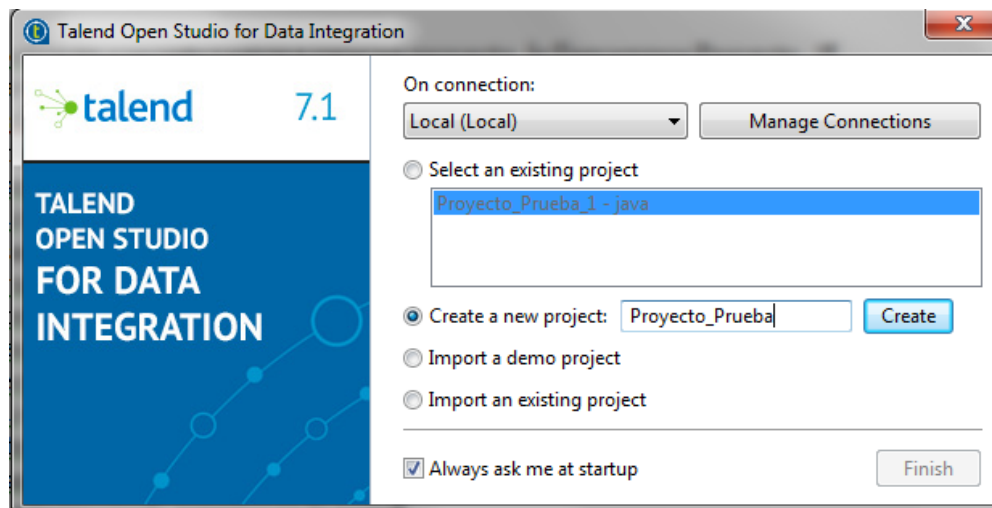
La versión gratuita de Talend solo permite ejecuciones de procesos en local, para poder ejecutar en servidores, habría que pasarse a una versión de pago por suscripción. Aunque los desarrollos realizados con la versión gratuita se pueden exportar para que puedan ser utilizarlos en otras plataformas.

Abrimos Talend que ya tenemos instalado.

La primera pantalla de proyectos ofrece tres opciones:

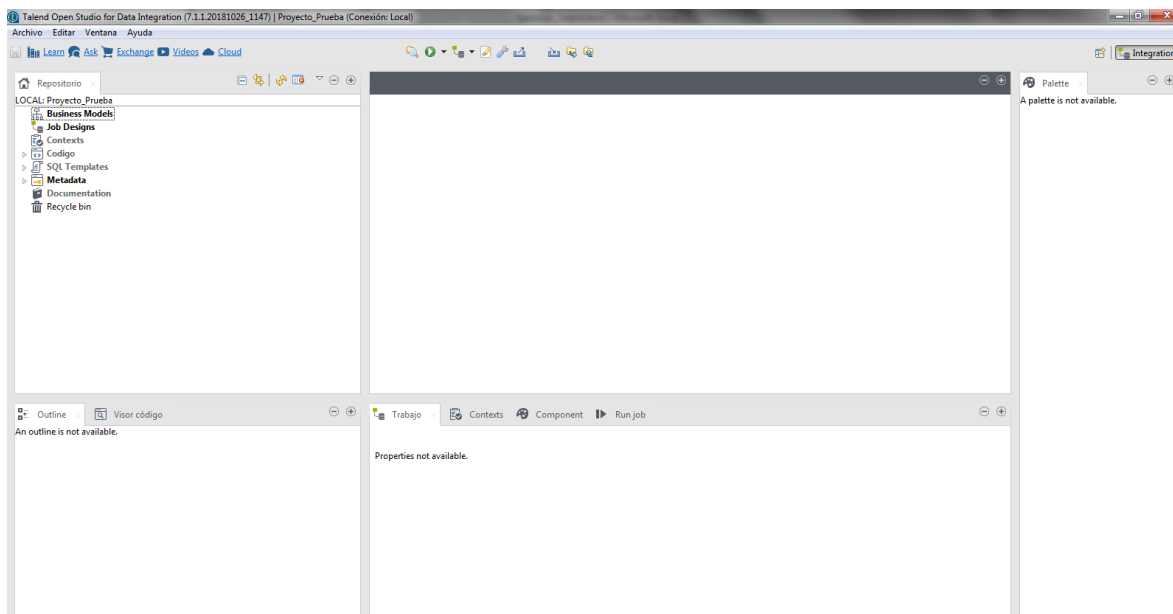
- Crear un proyecto nuevo
- Importar un proyecto de demo, estos demo project se descargan desde la plataforma de Talend. donde la comunidad pone a disposición de los desarrolladores distintos proyectos, Jobs, componentes que se pueden analizar y utilizar.
- Importar un proyecto existente.

Vamos a crear un proyecto, le llamaremos Proyecto_Prueba y pulsamos Finish, para empezar a cargar la interfaz del desarrollo.



Se muestra la pantalla de bienvenida donde nos da documentación, quick stars, demos, tutoriales, foros, cursos etcétera.

Iremos directamente a la parte de desarrollo, cerrando la pantalla de Welcome.



Ejercicio 1.-Creación de Conexiones a BDAs.

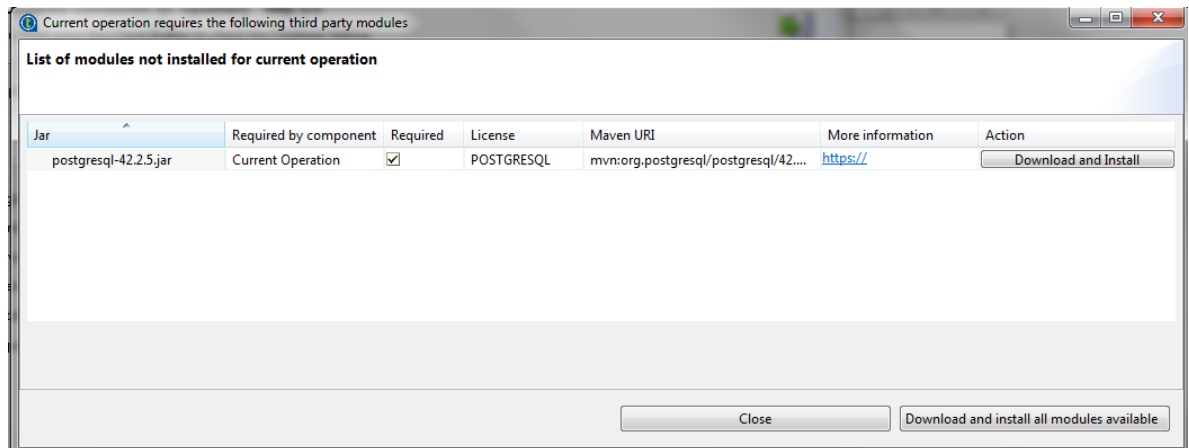
La conexión a base de datos la haremos desde Metadata, sección de Talend Studio donde se definen todas las conexiones tanto a orígenes como destinos de datos.

Las conexiones a bases de datos siempre se realizan por jdbc.

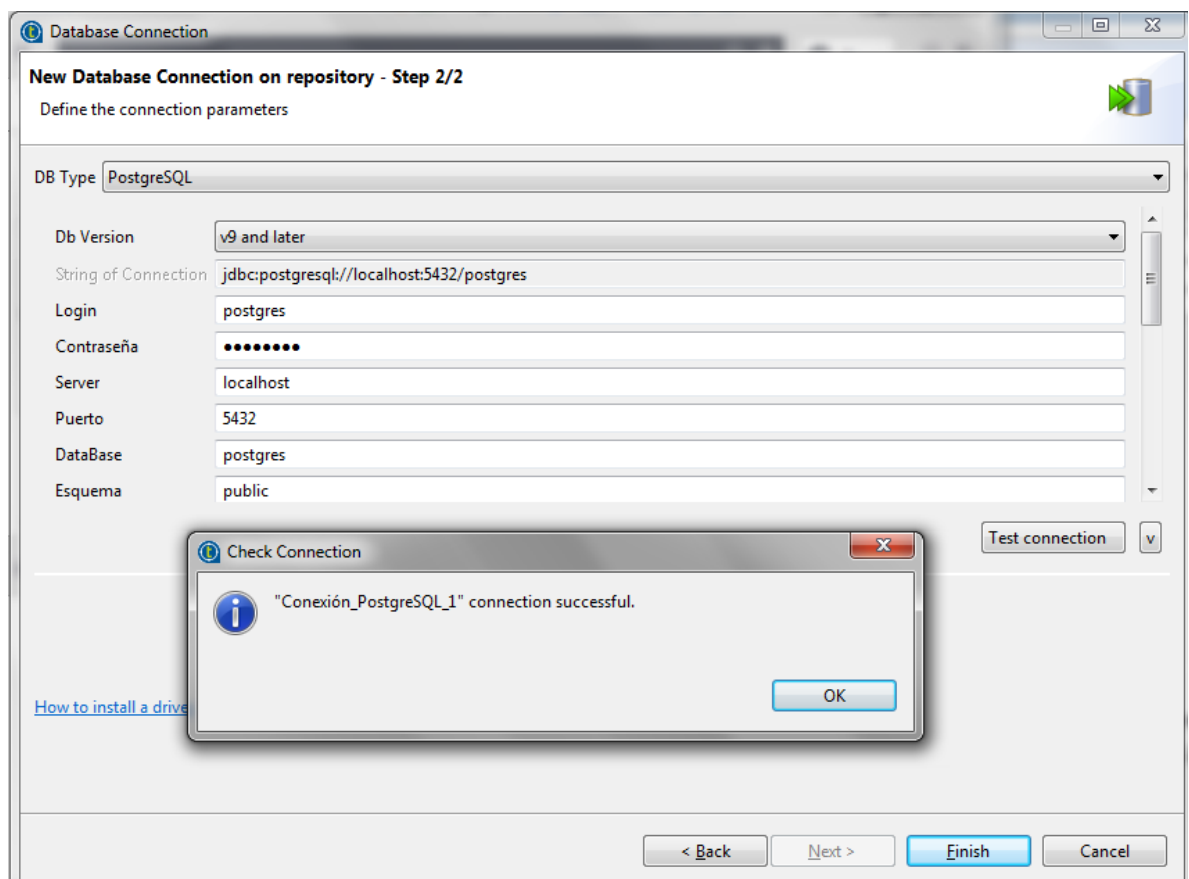
Vamos a crear una conexión al esquema "postgres" de la base de datos que tenemos instalada, que será el origen de los datos y otra conexión para el destino, que será el esquema stage de la misma base de datos.

Para crear la conexión de origen de los datos, en Metadata; Db Connections botón derecho y en el menú contextual, elegimos Crear Conexión. Para crearla, le daremos un nombre, Conexión_postgresql_1, y pulsamos Next, para pasar a especificar el tipo de base de datos para el vamos hacer la conexión, en nuestro caso PostgreSQL y la versión. También los parámetros de conexión, login:postgres y contraseña:postgres, server: localhost, puerto:5432, Database: postgres esquema: public

Se precisará instalar el driver postgresql para Talend, procederemos pulsando Download and Install. Tendremos que aceptar la licencia



Antes de finalizar (Finish) comprobar que la conexión es correcta haciendo el Test



Lo mismo para crear la conexión destino, Conexión_Stage, y los parámetros de conexión, login:stage y contraseña: stage, host: localhost, puerto:1521 y SID: orcl. Realizaremos el Test de Conexión y si funciona correctamente pulsaremos el botón Finish para crearla también.



Database Connection

New Database Connection on repository - Step 2/2

Define the connection parameters

DB Type: Oracle with SID

Db Version: Oracle 12

String of Connection: jdbc:oracle:thin:@localhost:1521:ord

Login: stage

Contraseña: •••••

Server: localhost

Puerto: 1521

Sid: ord

Esquema:

Additional parameters:

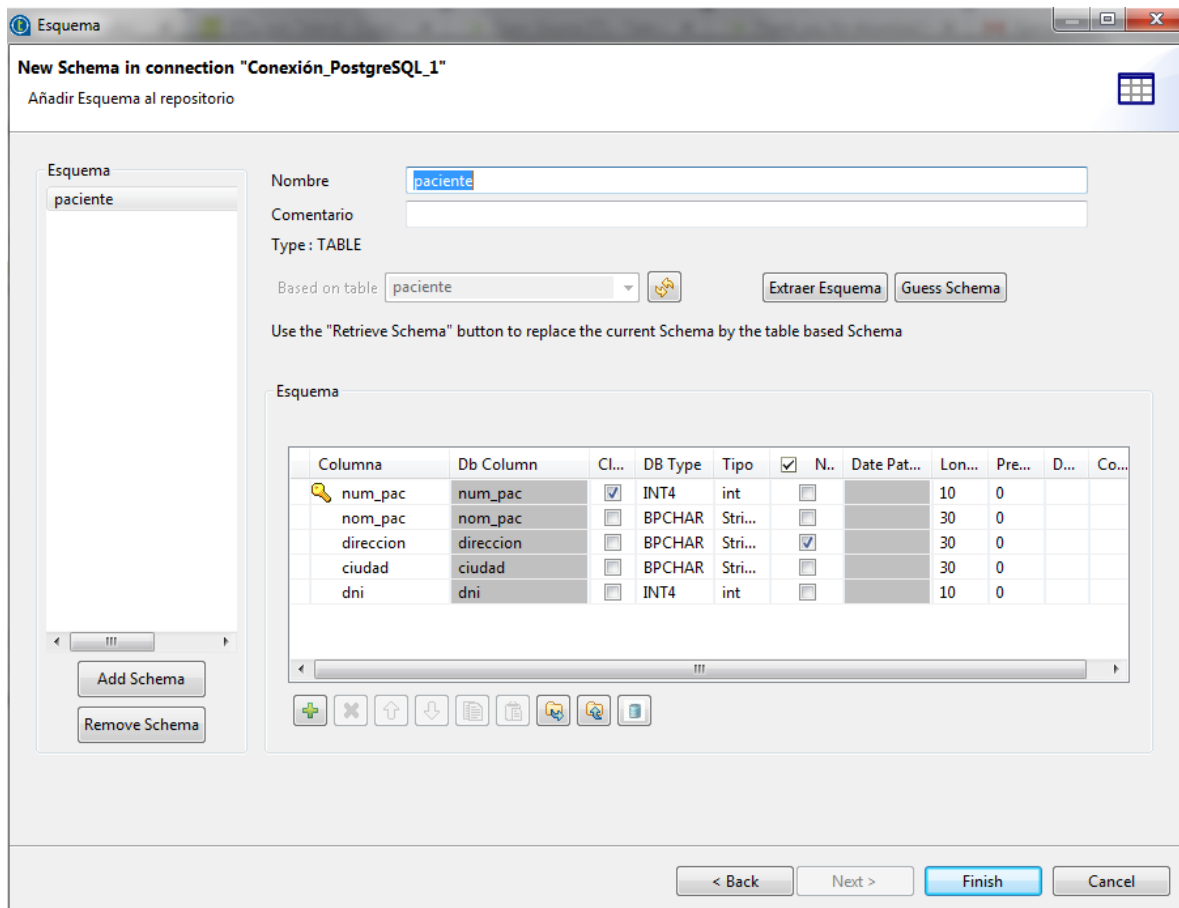
Test connection

Exportar como contexto Revertir Contexto

[How to install a driver](#)

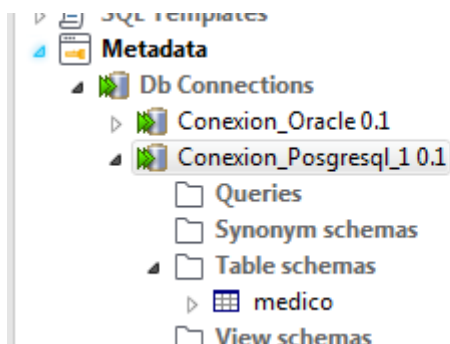
< Back Next > Finish Cancel

Una vez tenemos creadas las conexiones, lo que haremos es extraer los modelos de datos de los esquemas origen y destino de los datos, haciendo clic en las conexiones que acabamos de crear, botón derecho y elegir del menú contextual Extraer Esquema.



En el caso de la extracción del modelo de postgresQL, en la ventana que nos aparece pulsaremos Next, para que nos muestre todo, y elegiremos el esquema public y la tabla Medico que es la que queremos cargar. Nos muestra todos los campos con sus propiedades en una ventana. Pulsaremos Finish para extraer la estructura de la tabla médico del esquema public.

Se ha creado el modelo de datos MEDICO por debajo de la conexión Conexión_Postgresql_1.



Se ha creado el modelo de datos MEDICO por debajo de la conexión Conexión_Postgresql_1.

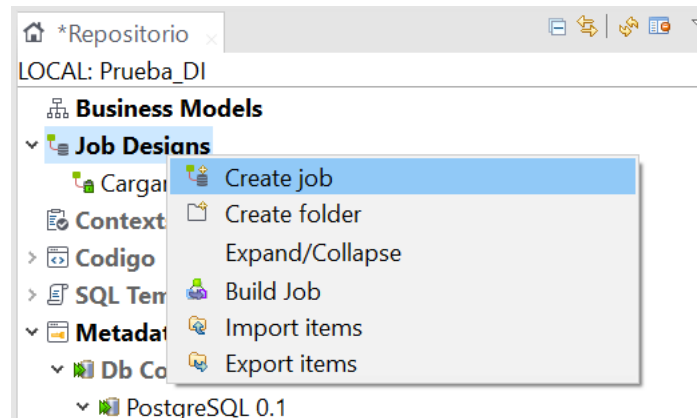
Lo mismo se hará para el modelo de datos del esquema stage de Oracle.



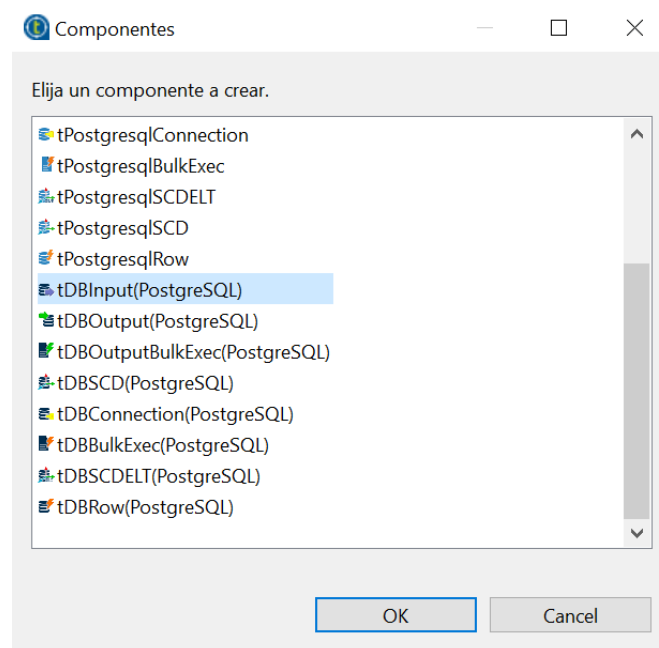
Ejercicio 2.-Creación de Job de Carga de datos.

En Talend el job es la unidad mínima de ejecución, es decir, es el primer componente que por sí solo va a tener poder de ejecución

Para crear un Job, nos situamos en Job Designs, botón derecho y en el menú contextual, elegimos Create Job, lo primero será asignarle un nombre: Carga_tabla_Medico_Stage y pulsamos Finish.

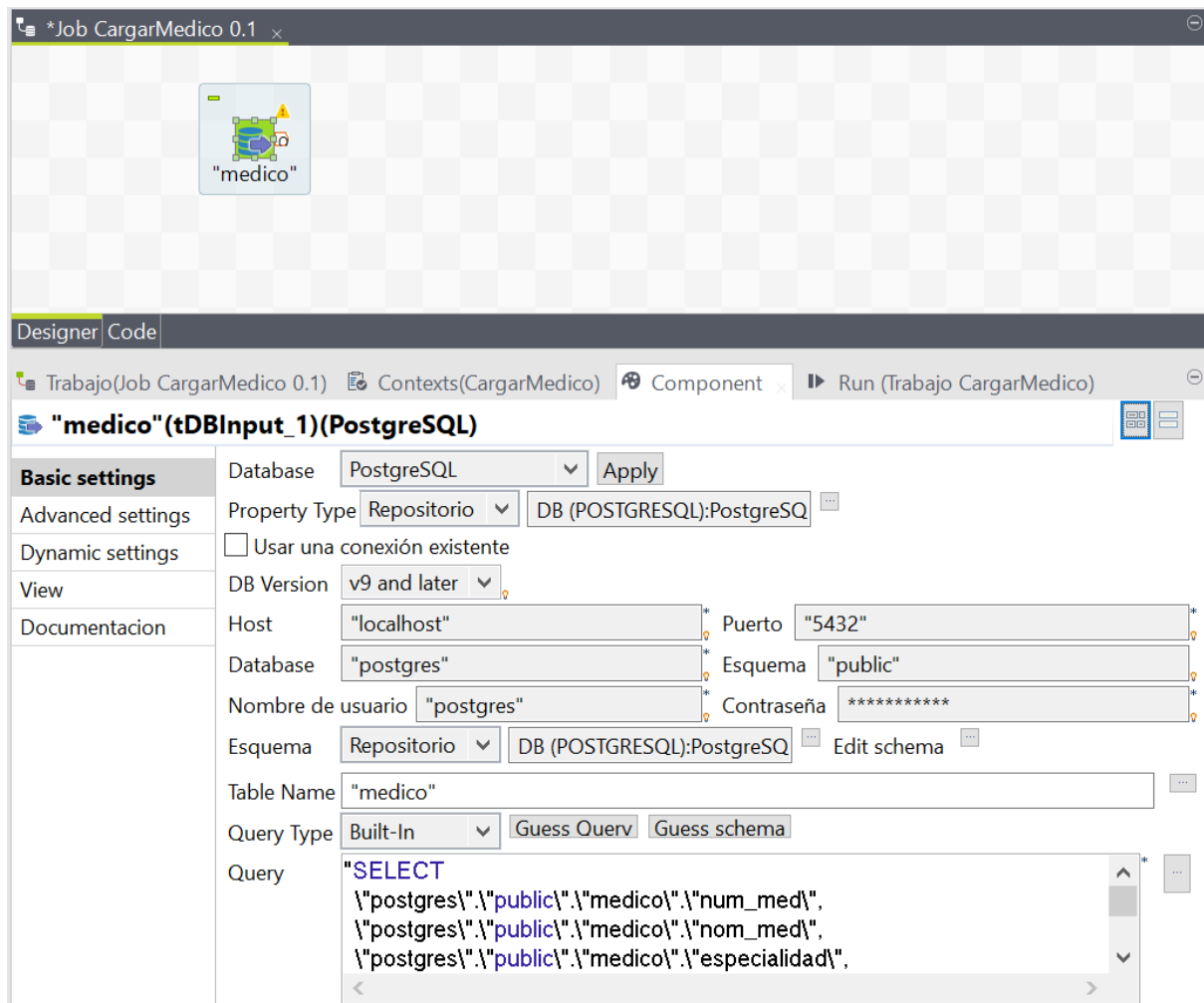


Arrastraremos el modelo de datos de la tabla Medico creado en el ejercicio anterior sobre el área de trabajo para que Talend nos ofrezca todos los componentes con los que se puede diseñar el flujo del proceso de carga. Elegiremos tDBInput(PostgreSQL), Lectura desde PostgreSQL.

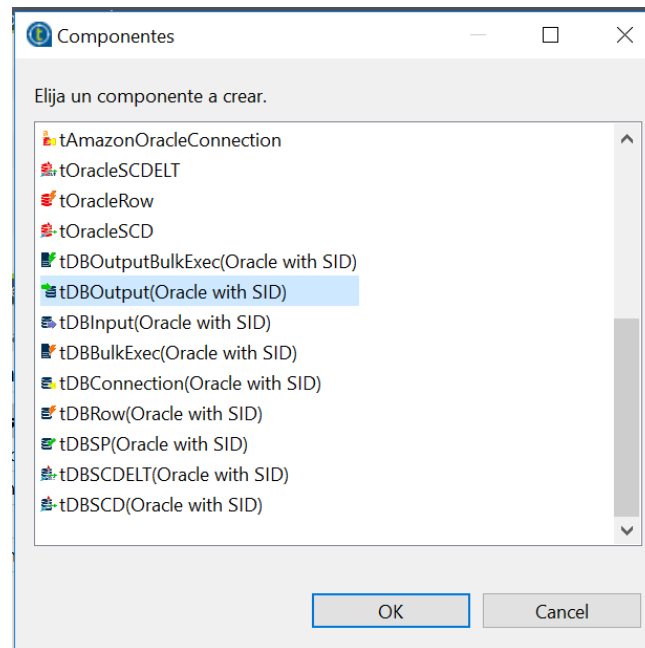




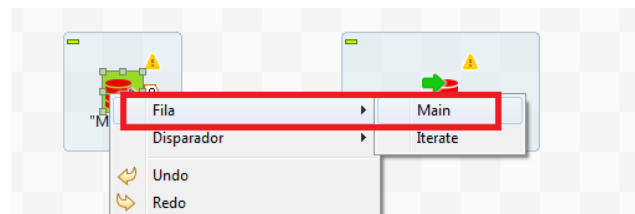
Como resultado veremos que el componente elegido que se aplicará sobre el modelo de datos MEDICO. Podemos ver las propiedades pulsando en Component de la ventana que se muestra en la parte inferior del área de trabajo.



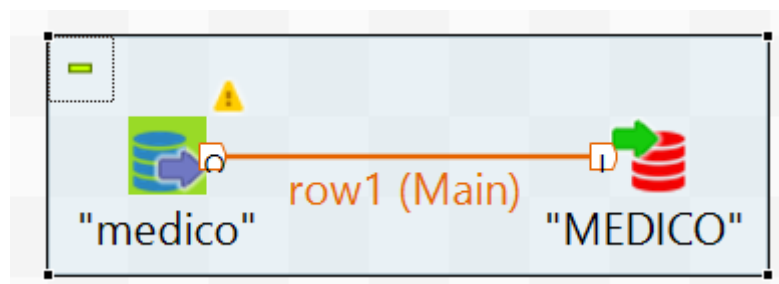
Ya tenemos en el Job, la parte correspondiente al origen de los datos del proceso, ahora vamos a añadir la parte destino de los datos. Para ello, arrastraremos la Conexión_stage al área de trabajo del Job. Nos aparecerá la pantalla de componentes de Talend, elegiremos tDBOutput (Oracle with SID), Escritura en Oracle.



Ahora faltaría conectar el flujo en el job, para unir el componente de lectura de la tabla con el componente de escritura, se haría, haciendo clic en el componente MEDICO, con botón derecho y elegir Fila; Main, soltando el ratón sobre el componte final del flujo.



El resultado del job Carga_tabla_Medico_Stage quedaría así:



Ejercicio 3.-Ejecución del Job de Carga de datos.

Para ejecutar el job que acabamos de diseñar, iremos a la pestaña Run (Trabajo Carga_tabla_Medico_Stage) que se encuentra debajo del área de trabajo del job.

Pulsaremos Run para Ejecutar

Si comprobamos en el SQLDeveloper veremos que se han cargado los datos a la tabla médico del stage

lab

oradba_sys

ord_sys

postgres

stage

Tablas (Filtrado)

MEDICO

NUM_MED

NOM_MED

ESPECIALIDAD

DIRECCION

CIUDAD

TELEFONO

SUELDO

DNI

USUARIOS

Vistas

Vistas de Edición

Indice

Paquetes

Procedimientos

Funciones

Operadores

Colas

Tablas de Colas

Disparadores

Diagramas de Edición Gráfica

Moja de trabajo

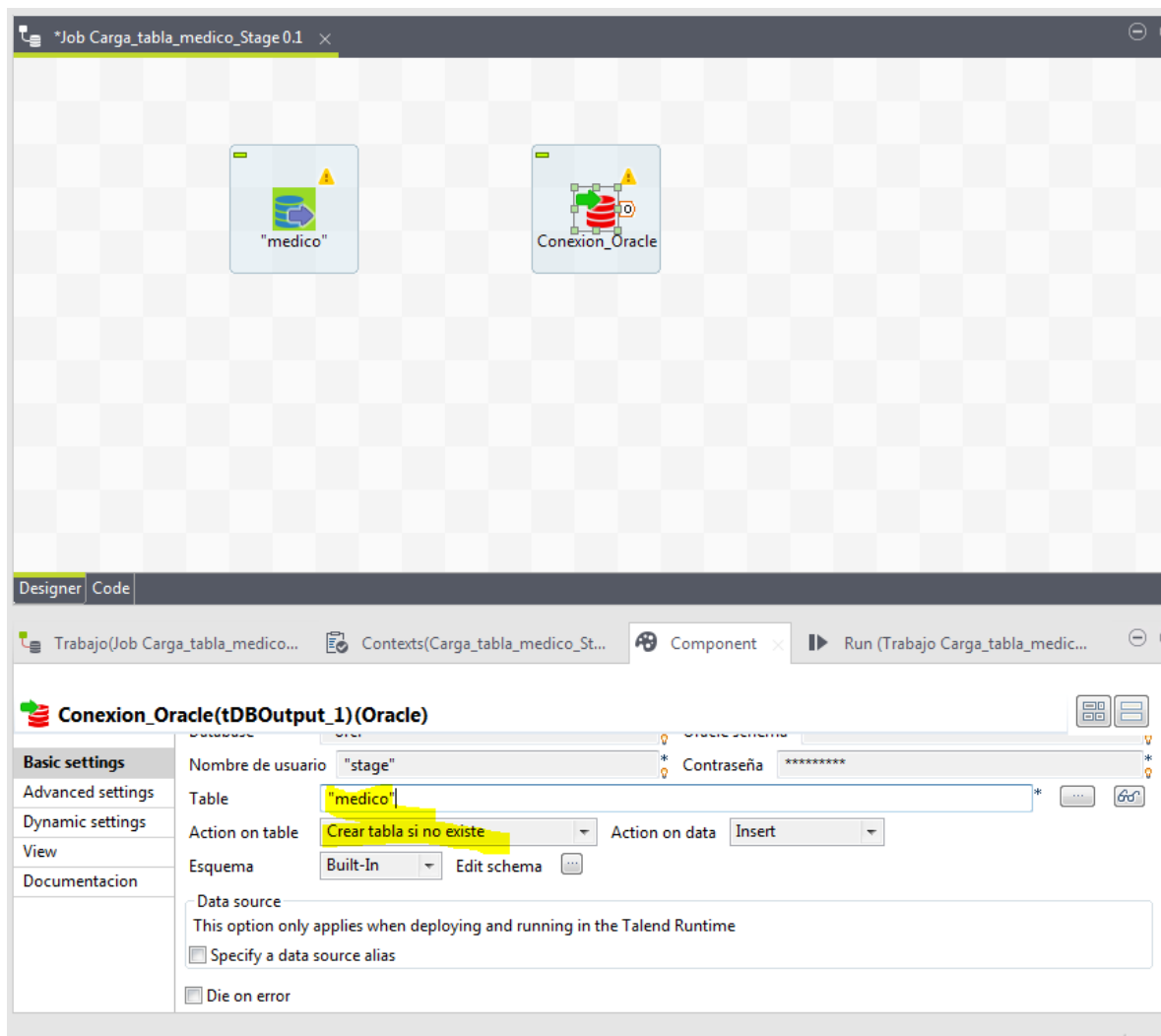
Generador de Consultas

SELECT * FROM MEDICO;

Resultado de la Consulta x

Todas las Filas Recuperadas: 14 en 0,145 segundos

	NUM_MED	NOM_MED	ESPECIALIDAD	DIRECCION	CIUDAD	TELEFONO	SUELDO	DNI
1	1	Severino Martin	Atención Primaria	(null)	Barcelona	938597565	45000	99658965
2	2	Mariaas Benin	Atención Primaria	Rambia Catalunya 27	Tarragona	658599714	35000	55632659
3	3	Pawel Pikorski	Cardiología	Paseo Cracovia 23	Girona	652356981	27000	55896655
4	4	Emilia Rodriguez	Radiología	El Pelin 189	Mataró	(null)	38000	88957854
5	5	Ana Maria López	Traumatología	(null)	Barcelona	(null)	98000	44125485
6	6	Conrado Pitarque	Cirugía Plástica	Las Hermosas 78-156	Tarragona	938567435	46000	99658412
7	7	Andrés Libertad	Atención Primaria	Avenida de las Marinas 2	Peníscola	(null)	38000	35663233
8	8	Celestino Posa	Oftalmología	(null)	Barcelona	654649235	77500	85569966
9	9	Elena Rodriguez	Atención Primaria	Las Sombras 56	L'Hospitalet	934526895	66500	74585221
10	10	Mateo Austriaco	Traumatología	Urb. Cancleras Apt 45	Fremiá de Mar	(null)	23000	96549785
11	11	Minerva Guerra	Oncología	Las Azores 90-6C	Barcelona	984657008	89400	78545485
12	12	Pedro Ramirez	Cardiología	C/Primer a la Derecha 3	Barcelona	(null)	59500	55896569
13	13	Roberto Juárez	Radiología	Las Cuartillas 36	Barcelona	634695216	64250	88599658
14	14	Fernando Fernández	Atención Primaria	Paseo Cracovia 56	Girona	653469523	36525	44125553



Antes debemos comprobar que la tabla no existe en el destino, si existe la borramos con la ejecutando desde SQL Developer, la siguiente instrucción

DROP TABLE "medico";

y visualizamos sus propiedades, para cambiar el nombre de la tabla y llamarle igual que la tabla origen, MEDICO. Además, indicaremos que la tabla no existe en destino y que hay que crearla al ejecutar el proceso.

Esto se hace cambiando las propiedades Table: MEDICO y Acción: Crear tabla si no existe.