CONSTRUCCIÓN DE UN DASHBOARD PARA LA IDENTIFICACIÓN DE LA PRODUCCIÓN CIENTÍFICA DE INVESTIGADORES RECONOCIDOS POR MINCIENCIAS EN INSTITUCIONES DE EDUCACIÓN SUPERIOR (IES) EN COLOMBIA.

Autor

Juan Fernando Corredor Niño

Director

Elio H. Cables Pérez, PhD

Universidad Antonio Nariño

Facultad de Ingeniería de Sistemas

Especialización en Gobierno de Datos

Bogotá D.C

2023

Índice de contenidos

Índice de figuras

Índice de tablas

Resumen

En la actualidad, existen diversas herramientas y metodologías para identificar y analizar datos provenientes de diversas fuentes. Los dashboard interactivos son herramientas que ofrecen una visualización clara y accesible de la información, y algunos utilizan algoritmos de minería de datos y aprendizaje automático para clasificar la información y descubrir patrones.

Sin embargo, aún no se ha desarrollado una herramienta que integre las bases de datos CvLAC y SNIES y permita visualizar el estado actual de la producción científica por cada IES en Colombia. Por lo tanto, se hace necesario el desarrollo de una herramienta que integre eficientemente estas bases de datos y facilite el análisis y la comprensión de la producción científica en las IES del país.

Palabras clave:

Dashboards, visualización, CvLAC, SNIES, producción científica, IES.

Abstract

Currently, there exist various tools and methodologies for identifying and analyzing data from diverse sources. Interactive dashboards are instruments that provide a lucid and accessible visualization of information, and some employ data mining algorithms and machine learning to classify data and uncover patterns.

However, a tool that integrates the CvLAC and SNIES databases and enables the visualization of the present state of scientific production for each higher education institution in Colombia has not yet been created. Consequently, it is essential to develop a tool that efficiently integrates these databases and facilitates the analysis and comprehension of scientific production in the country.

Keywords:

Dashboards, visualization, CvLAC, SNIES, scientific production, IES.

1. Introducción

En Colombia, la calidad de la educación superior es esencial para garantizar el desarrollo del país. Por esta razón, las instituciones de educación superior (IES) deben acreditar sus programas y facultades para demostrar su calidad y contribución a la nación. Uno de los factores que se consideran en los procesos de acreditación es la producción científica de los investigadores en las instituciones. Por lo tanto, es necesario identificar la producción científica de los investigadores de las IES para posiblemente ayudar en el proceso de acreditación según el decreto 1279 de 2002, este establece los criterios y procedimientos para la evaluación y acreditación de los programas que ofrecen las universidades en el país. Este decreto establece que la evaluación de los programas se realizará por parte del Ministerio de Educación Nacional y se llevará a cabo a través de un proceso de autoevaluación y evaluación externa. Entre los aspectos a evaluar se incluyen la calidad del cuerpo docente, el plan de estudios, la infraestructura y los recursos disponibles para la investigación, la calidad y la relevancia de las publicaciones y la producción científica de los investigadores, entre otros (Ministerio de Educación Nacional).

En este contexto, una herramienta clave para identificar la producción científica de los investigadores en Colombia es el CvLAC. El CvLAC es un sistema de información creado por Minciencias (antes Colciencias) para recolectar, procesar y difundir información sobre los investigadores y sus actividades de investigación. Este sistema permite a los investigadores colombianos registrar y difundir su información de investigación, lo que les permite tener mayor visibilidad y reconocimiento en la comunidad científica.

Por otro lado, el Ministerio de Ciencia, Tecnología e Innovación de Colombia, también conocido como Minciencias, es la entidad encargada de fomentar y coordinar el desarrollo científico y tecnológico del país. Minciencias establece políticas y programas para promover la investigación y la innovación en Colombia. Además, Minciencias es responsable de reconocer y apoyar a los investigadores y grupos de investigación destacados en el país.

Minciencias reconoció a un grupo de investigadores de Colombia por su producción científica. Estos investigadores son aquellos que han sido evaluados por Minciencias y han cumplido con los criterios de calidad en investigación. La identificación de estos investigadores es crucial para el análisis y la evaluación de la producción científica en las instituciones de educación superior.

La producción científica de los investigadores reconocidos por Minciencias en las instituciones de educación superior en Colombia se puede dividir en diferentes áreas de conocimiento y tipos de productos. Estos productos pueden incluir artículos científicos, libros, capítulos de libros y otros. Es importante tener en

cuenta el rango etario y el sexo de los investigadores para obtener una imagen completa de la producción científica en las IES.

Para concluir, Identificar el trabajo científico de los investigadores en Colombia es crucial para conocer el panorama actual de la producción científica de las instituciones y poder tomar decisiones informadas. Además, el ministerio juega un papel fundamental en el reconocimiento y apoyo a los investigadores destacados en el país. La Contribución científica de los investigadores puede incluir diferentes áreas de conocimiento y tipos de productos, es transcendental tener una perspectiva clara del estado actual y del impacto que la producción científica tiene en el país.

2. Descripción y formulación del problema

En Colombia, las Instituciones de Educación Superior (IES) deben acreditar sus programas para garantizar su calidad y su contribución al desarrollo del país. La producción científica de los investigadores es una de las variables que se consideran en algunos procesos de acreditación. Por lo tanto, es importante identificar dicha producción de los investigadores en Colombia en las instituciones de educación superior en Colombia, incluyendo sus áreas de conocimiento y tipos de productos. Esto permitiría a las instituciones de educación superior tener una mejor comprensión del panorama de producción científica y utilizar esta información para mejorar sus procesos de acreditación de alta calidad.

Actualmente, no existe un sistema de consulta para acceder a estos datos. Por lo tanto, las entidades encargadas tienen que realizar los cálculos de forma manual. Sería beneficioso para todas las partes involucradas contar con un sistema accesible que facilite la identificación y el análisis de la producción científica de los investigadores en las IES privadas y públicas de Colombia.

Formulación del problema

¿Cómo identificar la producción científica de los investigadores reconocidos por Minciencias y afiliados a las IES en Colombia utilizando las bases de datos CvLAC y SNIES?

2.1. Objetivo general

Desarrollar un dashboard interactivo que permita identificar y visualizar de manera clara y accesible la producción científica de los investigadores asociados a las IES en Colombia utilizando las bases de datos CvLAC y SNIES.

2.2. Objetivos específicos

1. Extraer los datos sobre la producción científica de los investigadores reconocidos por Minciencias en las instituciones de educación superior en Colombia.
2. Caracterizar la producción científica de los investigadores avalados por Minciencias en Colombia a partir de las bases de datos SNIES y CvLAC.
3. Diseñar un dashboard que permitan visualizar la producción científica de los investigadores reconocidos por Minciencias, en las instituciones de educación superior en Colombia, a través de diferentes elementos visuales.
4. Implementar el dashboard que permita a los usuarios navegar y explorar la información sobre la producción científica de los investigadores reconocidos por Minciencias en las IES en Colombia.

3. Marco referencial

3.1. Marco teórico

El proceso de evaluación de la producción científica de los investigadores es fundamental para medir el impacto de las Instituciones de Educación Superior (IES) en la sociedad y su contribución al desarrollo del país. La utilización de herramientas tecnológicas en este proceso puede mejorar la eficiencia y precisión en la identificación y análisis de la producción científica, como motores de bases de datos SQL y software ETL entre estas herramientas.

Para la búsqueda de información sobre la producción científica de investigadores en Colombia, resultan fundamentales las bases de datos CvLAC y SNIES. Estas fuentes de información ofrecen datos valiosos que permiten la identificación y análisis de la producción científica de los investigadores en el país.

CvLAC

El CvLAC es una plataforma que permite registrar los perfiles de las personas involucradas en actividades relacionadas con Ciencia, Tecnología e Innovación - CTeI. Los individuos pueden ser clasificados como investigadores si cumplen con ciertos requisitos. Administrada por Minciencias, esta base de datos es fundamental para gestionar y dar seguimiento a las actividades investigativas de científicos, tecnólogos e innovadores (CTI) en Colombia. Asimismo, representa una fuente valiosa para identificar la producción científica de investigadores y grupos de investigación en el país, incluyendo publicaciones, proyectos de investigación y patentes, entre otros aspectos relevantes. (CvLAC, s.f.)

El SCIENTI, un sistema nacional de Colombia recopila y organiza datos sobre los investigadores, grupos de investigación, proyectos y publicaciones científicas y tecnológicas del país. Los datos contenidos en las bases del SCIENTI incluyen perfiles de investigadores, información sobre grupos de investigación, proyectos de investigación y publicaciones científicas y tecnológicas, así como información sobre congresos y eventos científicos. Estos datos son utilizados por diversos sectores para analizar la producción científica del país, identificar oportunidades de financiamiento y colaboración, y realizar estudios en el área. (CvLAC, s.f.)

SNIES

La base de datos SNIES, gestionada por el Ministerio de Educación Nacional de Colombia, es una fuente de información abierta sobre las Instituciones de Educación Superior y sus programas académicos. (Qué es el SNIES, s.f.)

SQL Server

A través del lenguaje de consulta estructurado (SQL), es posible acceder y manipular los datos almacenados en las bases de datos de SQL Server, lo que facilita su gestión y utilización en diversas aplicaciones y sistemas. SQL Server es un sistema de gestión de bases de datos relacionales creado por Microsoft Corporation (Microsoft, 2021). Se trata de una plataforma informática robusta y escalable que permite almacenar y administrar grandes cantidades de datos de manera eficiente. (Microsoft, 2021).

PowerBI

Power BI es una herramienta de análisis de datos y visualización de información desarrollada por Microsoft Corporation (Microsoft, 2021). Con Power BI, los usuarios pueden conectar, modelar y visualizar datos desde una amplia variedad de fuentes, incluyendo bases de datos, archivos, servicios en la nube y otras aplicaciones empresariales. Además, Power BI ofrece diversas herramientas y funciones para el análisis y la presentación de datos, como gráficos, tablas, informes y paneles interactivos, que permiten a los usuarios obtener información valiosa y tomar decisiones informadas de manera más eficiente (Microsoft, 2021).

Talend Open Studio

Para el desarrollo de este proyecto se utilizará el software Talend Open Studio, es una herramienta de código abierto para la integración y transformación de datos. ETL es una abreviatura que significa "Extract, Transform, Load" (Extraer, Transformar, Cargar). Talend Open Studio es una herramienta de software que ayuda a los usuarios a extraer datos de diferentes fuentes, transformarlos de acuerdo con sus necesidades y cargarlos en una ubicación de destino. (Talen Open Studio, 2021),

Extraer, Transformar y Cargar - ETL

Según Kimball y Ross (2013), este proceso consiste en extraer los datos de las fuentes de origen, transformarlos para adaptarlos a un formato común y adecuado para el destino, y finalmente cargarlos en una base de datos de destino. Una ETL es un proceso utilizado en el ámbito de la gestión de datos para la integración y consolidación de información procedente de diferentes fuentes.

3.2. Estado del arte

La capacidad de las instituciones de educación superior para generar investigación de alta calidad es un indicador esencial para medir su impacto y contribución al desarrollo del país. Por lo tanto, resulta importante contar con una herramienta que facilite el análisis y la comprensión de la producción científica de los investigadores afiliados a las universidades en Colombia. Por esta razón, resulta fundamental tener una herramienta que permita analizar la producción científica de los investigadores asociados a las universidades del país.

Aunque existen diversas herramientas para analizar datos, aún no se ha desarrollado una herramienta que permita integrar y visualizar la información de las bases de datos SCIENTI y SNIES, para la identificación de la producción científica de investigadores las cuales contienen información valiosa sobre la producción científica en las IES colombianas.

El desarrollo de un dashboard interactivo que integre ambas bases de datos, permitiría una visualización clara y accesible de la producción científica por cada IES. Además, se podrían identificar patrones y tendencias en la producción científica.

3.3. Impacto

El desarrollo de un dashboard interactivo que integre las bases de datos CvLAC y SNIES permitirá mejorar la eficiencia y precisión en la identificación y análisis de la producción científica de las IES en Colombia. Buscando facilitar la evaluación de la calidad y la contribución de las IES al desarrollo del país, y permitirá una mejor toma de decisiones por parte de los distintos actores involucrados en el Sistema Nacional de Ciencia, Tecnología e Innovación.

3.4. Componente de innovación

Creación de un dashboard interactivo para la identificación de la producción científica de investigadores que integre las bases de datos CvLAC y SNIES para ofrecer una visualización clara y accesible de la producción científica de las IES en Colombia. Este dashboard también incluirá información detallada sobre las áreas de conocimiento, rango de edad, género y tipos de productos de los investigadores. Al ofrecer una evaluación más completa y detallada de la producción científica, este componente de innovación permitirá una mejor toma de decisiones y una mayor comprensión del panorama científico en las IES del país.

4. Metodología

A pesar de que existen distintas metodologías con enfoques y propuestas diferentes, elegir cuál es la más adecuada para implementar no es una tarea sencilla. En realidad, los requisitos específicos de cada proyecto son los que determinan la estructura que mejor se adapta al mismo. Además, la decisión también depende de los objetivos de la organización o del proyecto (Castañeda & García, 2021).

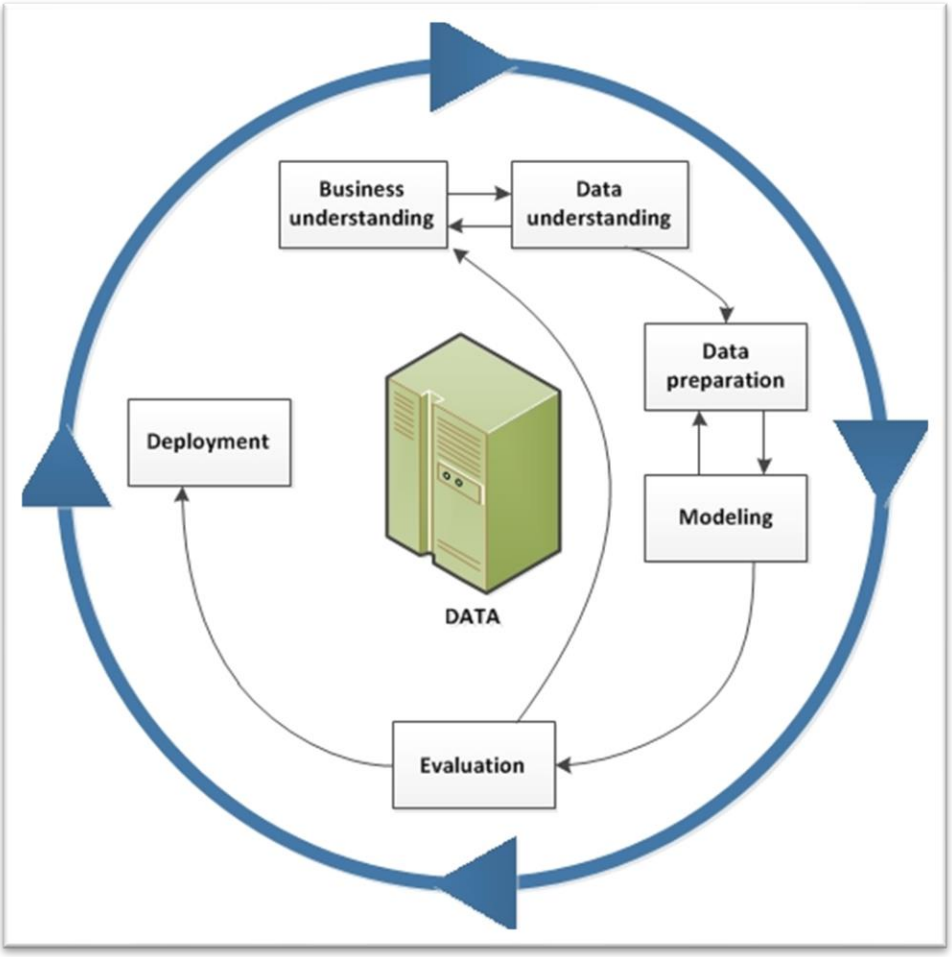
Para este proyecto se trabajará bajo la metodología CRISP-DM (Cross Industria Standard Process for Data Mining) es un proceso estructurado utilizado para llevar a cabo proyectos de datos.

La metodología CRISP-DM ofrece varias ventajas en la gestión de proyectos de datos. En primer lugar, su enfoque estructurado y bien definido permite a los equipos de proyecto establecer un marco claro de trabajo, lo que facilita la colaboración. Además, al dividir el proceso en fases distintas, se pueden identificar rápidamente los problemas y riesgos, lo que permite abordarlos de manera temprana y minimizar su impacto en el proyecto. Otro beneficio importante es que esta metodología se puede adaptar fácilmente a diferentes tipos de proyectos y entornos, lo que la convierte en una metodología versátil y flexible para la gestión de proyectos relacionados con datos (Castañeda & Garcia, 2021).

La metodología CRISP-DM es un proceso estructurado que consta de seis fases para el desarrollo de proyectos de minería de datos. En este proyecto, se tomarán en cuenta **cuatro fases de la metodología**, siendo estas: **Comprensión del negocio, Comprensión de los datos, Preparación de los datos** y **Despliegue**. (Ver ilustración 1).

De esta manera, se llevará a cabo un proceso riguroso y sistemático para la creación del dashboard interactivo que permitirá visualizar de manera clara y accesible la producción científica de los investigadores asociados a las IES en Colombia.

Ilustración 1 - Modelo CRISP-DM



Fuente: IBM conceptos CRISP-DM

A continuación, se describen las fases para el desarrollo del proyecto.

1. Comprensión del negocio
2. Comprensión de los datos
3. Preparación de los datos
4. Despliegue

En la primera fase del proceso, se dedica especial atención a comprender los objetivos y la forma en que se llevará a cabo el proyecto. Para lograr esto, se lleva a cabo un análisis de las fuentes de datos disponibles, y se define una hoja de ruta del proyecto. Este paso se conoce comúnmente como la fase de **comprensión del negocio**, ya que se busca tener un conocimiento del contexto en el que se desenvuelve el proyecto y las necesidades específicas que deben ser satisfechas.

De esta manera, se establece una base sólida para las fases siguientes del proyecto, lo que garantiza una mayor eficacia y eficiencia en la consecución de los objetivos establecidos.

La siguiente fase, conocida como **comprensión de los datos**, implica la recopilación y exploración de los datos relevantes para el proyecto. Durante esta etapa, se realiza una exploración inicial para comprender la estructura de los datos, identificar las variables y tablas más importantes y realizar una descripción sobre su uso. El objetivo principal es obtener una comprensión profunda de los datos y su contexto,

La tercera fase, conocida como **preparación de los datos**, implica seleccionar las variables que se utilizarán en el modelo, la limpieza de datos, la transformación de variables o procesos de ETL para preparar los datos para su uso en el análisis.

En la fase final del proceso, se llevará a cabo el **despliegue** del modelo de visualización seleccionado en un entorno operativo, lo que implica la implementación de este en producción. Asimismo, se generarán visualizaciones con el fin de comunicar los resultados obtenidos a los interesados.

5. Desarrollo de la propuesta

A continuación, se presentará el desarrollo del proyecto enfocado en el diseño de un dashboard interactivo que tiene como objetivo brindar una visualización clara y accesible de la producción científica de los investigadores asociados a las Instituciones de Educación Superior (IES) en Colombia. La información que se utilizará para tal fin provendrá de las bases de datos de CvLAC y SNIES, que contienen información relevante sobre los investigadores y sus productos académicos.

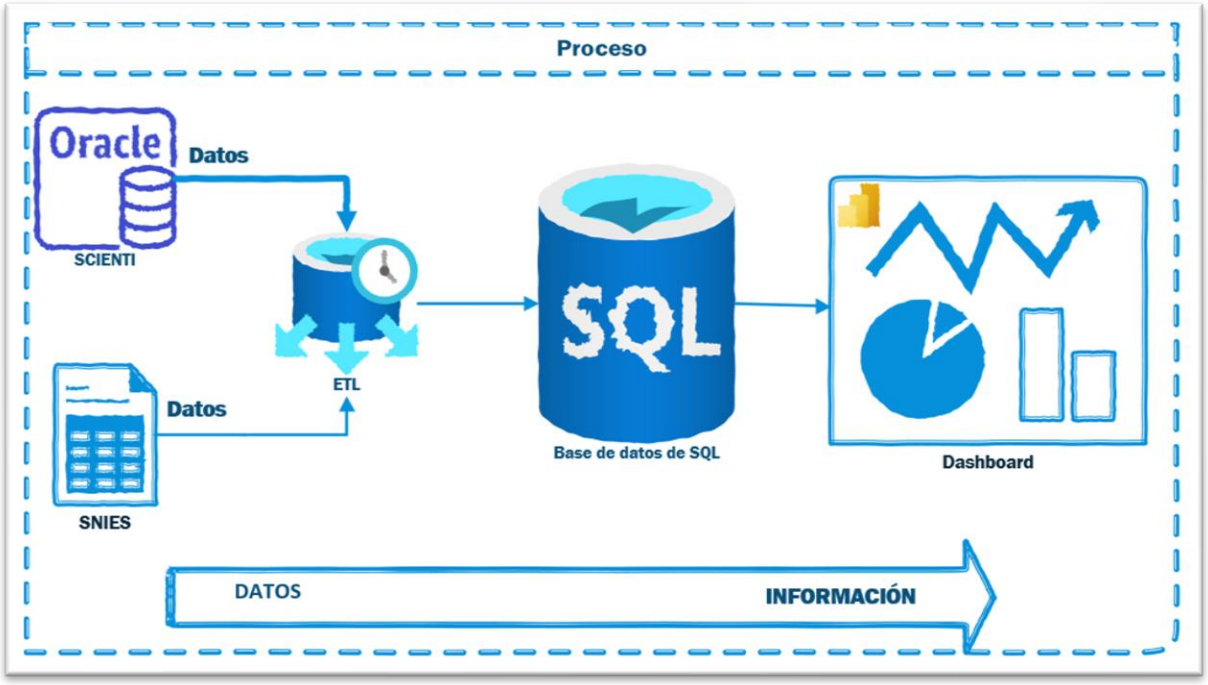
Para alcanzar los objetivos establecidos, se seguirá la metodología expresada anteriormente, la cual representa un enfoque estructurado y sistemático para la ejecución de proyectos de datos. Esto permitirá mantener un enfoque riguroso y eficiente durante todo el proceso de desarrollo del dashboard.

Comprensión del negocio

Para este proyecto, se han elegido fuentes de datos las bases de datos CvLAC (Acceso limitado a cierto público por solicitud) y SNIES (Acceso abierto), las cuales se han evaluado como fiables al provenir de fuentes oficiales y se cuenta con acceso autorizado para extraer la información necesaria. Asimismo, se han definido los criterios de análisis y la hoja de ruta que se utilizará para clasificar y presentar visualmente la producción científica de los investigadores.

Se llevará a cabo una extracción de datos de las fuentes de información, los cuales serán procesados mediante un proceso ETL y posteriormente almacenados en una base de datos SQL. Esta base de datos será conectada a un visualizador de dashboard para su posterior visualización y análisis, como se muestra en el siguiente gráfico (ver ilustración 2).

Ilustración 2 - Proceso



Fuente: Elaboración propia

Comprensión de los datos

La base de datos SCIENTI es una fuente rica y amplia de información, con un total de 155 tablas. En el contexto de este proyecto, se ha enfocado en las tablas que se relacionan directamente con los objetivos de este proyecto, lo cual ha permitido un análisis exhaustivo y riguroso de la información requerida. La selección cuidadosa de estas tablas permitirá en la siguiente fase una mayor eficiencia en la extracción de información y una mejor comprensión de los patrones y tendencias en la producción científica de los investigadores (ver tabla 1).

Tabla 1 - Tablas del SCIENTI:

EN_ACT_ADMINISTRACION	EN_PROD_EMPRESA_ID	RE_EVENTO_PROYECTO
EN_ACT_DOCENCIA	EN_PROD_NORMA	RE_EVENTO_SECTOR_APL
EN_ACT_INVESTIGACION	EN_PROD_OBRA_ARTISTICA	RE_HOMOLOGACION_AREAS
EN_ACTIVIDAD	EN_PROD_PARTITURA	RE_HOMOLOGACION_SECTORES
EN_ANOMALIA	EN_PROD_SOFTWARE	RE_INSTITUCION_EVENTO
EN_AREA_CON_SIR	EN_PROD_TECNICA	RE_INSTITUCION_PRODUCTO

EN_AREA_CONOCIMIENTO	EN_PROD_TECNOLOGICO	RE_INSTITUCION_RED
EN_AREA_OCDE	EN_PROD_VEGETAL	RE_LIBRO_IDIOMA
EN_CALIF_LIBRO_REF	EN_PRODUCTO	RE_LIBRO_LIBRO
EN_CALIF_REVISTA	EN_PROGRAMA_ACADEMICO	RE_LIBRO_REVISTA
EN_CALIF_REVISTA_2015	EN_PROYECTO	RE_LINEA_INV_AREA_CON
EN_CENTRO_INV	EN_RECLAMO	RE_LINEA_INV_PALABRA_CLAVE
EN_COMUNIDAD	EN_RECLAMO_BK	RE_LINEA_INV_SECTOR_APL
EN_DEPARTAMENTO	EN_RECONOCIMIENTO	RE_ORGANIZACION_EVENTO
EN_DETALLE_JOVEN	EN_RECURSO_HUMANO	RE_PARTICIPACION_EVENTO
EN_DIRECCION_RH	EN_RECURSO_HUMANO_OTRO	RE_PATENTE_CONVO
EN_DOCUMENTO	EN_RED	RE_PRODUCTO_AREA_CON
EN_EDITORIAL	EN_REGION	RE_PRODUCTO_COMUNIDAD
EN_EDITORIAL_OTRO	EN_REGISTRO	RE_PRODUCTO_CONVO
EN_EVENTO	EN_RELIGION	RE_PRODUCTO_EVENTO
EN_EXAMEN_IDIOMA	EN_REPORTE	RE_PRODUCTO_PALABRA_CLA
EN_FOTO	EN_RESUMEN	RE_PRODUCTO_PRODUCTO
EN_HISTORICO_ACT	EN_REVISTA	RE_PRODUCTO_RECONOCIMIENTO
EN_HISTORICO_ACT_2016_02_25	EN_REVISTA_CATEGORIA	RE_PRODUCTO_RECURSO_HUM_OTRO
EN_IDIOMA	EN_REVISTA_ISSN	RE_PRODUCTO_SECTOR_APL
EN_INSTITUCION_OTRA	EN_REVISTA_OTRA	RE_PRODUCTOS_MDCN
EN_LIBRO	EN_SECRETO_INDUSTRIAL	RE_PROYECTO_COMUNIDAD
EN_LIBRO_CATEGORIA	EN_SECTOR_ACTIVIDAD_ECONOMICA	RE_PROYECTO_CONVO
EN_LIBRO_OTRO	EN_SECTOR_APLICACION	RE_PROYECTO_INSTITUCION
EN_LIBRO_REF	EN_SERIAL	RE_PROYECTO_PRODUCTO
EN_LINEA_INV	EN_SIR	RE_PROYECTO_REC_HUMANO_OTRO
EN_LOG_ELIMINACION	EN_SUBTIPO_PRODUCTO	RE_PROYECTO_RED
EN_MUNICIPIO	EN_TESIS_ORIENTADA	RE_REC_HUMANO_AREA_CON
EN_NIVEL_FORMACION	EN_TIPO_CONSTANTE	RE_REC_HUMANO_EXAMEN_IDI
EN_NOMBRE_TABLA	EN_TIPO_DIVULGACION	RE_REC_HUMANO_IDIOMA
EN_PAIS	EN_TIPO_DOCUMENTO	RE_RED_COMUNIDAD
EN_PALABRA_CLAVE	EN_TIPO_EVENTO	RE_RED_CONVO
EN_PAR_EVALUADOR	EN_TIPO_FINANCIACION	RE_RED_SOCIAL_IDENT
EN_PARTICIPACION_COMITE	EN_TIPO_NACIONALIDAD	RE_REGISTRO_CONVO
EN_PATENTE	EN_TIPO_PATENTE	RE_RH_CONVO
EN_PATENTE_REGISTRO	EN_TIPO_PRODUCTO	RE_RH_CONVO_DATOS
EN_PROD_APROPIACION_SOCIAL_CON	EN_TIPO_PRODUCTO_MDCN	RE_RH_CONVO_INST
EN_PROD_ARTICULO	EN_TIPO_RECONOCIMIENTO	RE_RH_O_EVENTO
EN_PROD_ARTICULO2	EN_TIPO_VINCULACION	RE_RH_O_RED
EN_PROD_ARTISTICA	EN_TRAYECTORIA_ESCOLAR	RE_SECRETO_INDUSTRIAL_CONVO

EN_PROD_ARTISTICA_DETALLE	EN_TRAYECTORIA_PROFESIONAL	RE_TRAY_ESCOLAR_AREA_CON
EN_PROD_AUDIOVISUAL	EN_VARIABLES	RE_TRAY_ESCOLAR_PALABRA_CLA
EN_PROD_BIBLIO	RE_ACT_INV_LINEA_INV	RE_TRAY_ESCOLAR_REC_HUMANO
EN_PROD_CAP_MEMORIA	RE_AREAS_OCDE	RE_TRAY_ESCOLAR_SECTOR_APL
EN_PROD_CAPITULO_LIBRO	RE_ARTISTICA_DET_CONVO	RE_EVENTO_CONVO
EN_PROD_CURSO	RE_EVENTO_AREA_CON	RE_EVENTO_PALABRA_CLA
EN_PROD_DIVULGACION_CON	RE_EVENTO_COMUNIDAD	

Fuente: Base de datos SCIENTI - Minciencias

A continuación, se presentará una breve descripción de las tablas que se utilizarán en el proyecto. Cada tabla se identifica por un nombre y una clave primaria (PK). Además, se proporcionará una breve explicación de la funcionalidad de cada tabla en el contexto del proyecto. Esto permite una mejor comprensión de la estructura de la base de datos y cómo se relacionan las diferentes tablas entre sí para lograr los objetivos del proyecto.

EN_RECURSO_HUMANO: esta tabla almacena los datos básicos del investigador suscrito en el portal, como nombres, sexo, datos de nacimiento y complementarios. La llave primaria es PK_RH.

EN_CALIF_REVISTA: esta tabla almacena la calificación obtenida por cada revista. También almacena el año de obtención, el cuartil, el área de conocimiento y el ISSN de la revista. La llave primaria es PK_CALIF_REV.

EN_HISTORICO_ACT: esta tabla almacena el histórico de las actividades de registro y actualización de la información en el aplicativo CvLAC por el investigador. Esta tabla no tiene una llave primaria específica.

EN_LIBRO_REF: esta tabla almacena información sobre los libros referenciados y utilizados por el investigador, que no son de su autoría, pero se encuentran en la base nacional de productos. La llave primaria es PK_LIBRO_REF.

EN_LINEA_INV: esta tabla almacena información sobre la línea de investigación que posee el investigador. La llave primaria es PK_LINEA_INV.

EN_REVISTA: esta tabla almacena información sobre las revistas que se encuentran en la base nacional. La llave primaria es PK_REVISTA.

EN_RED: esta tabla almacena información sobre las redes de conocimiento especializado registradas dentro de los productos de Apropiación Social del Conocimiento. La llave primaria es PK_EN_RED.

EN_REGISTRO: esta tabla almacena información del Registro que puede ser asociado a un producto tecnológico. La llave primaria es PK_REGISTRO.

EN_ACT_ADMINISTRACION: esta tabla almacena información sobre la experiencia profesional relacionada con actividades de administración realizadas por el investigador. La llave primaria es PK_ACT_ADMINISTRACION.

EN_ACT_DOCENCIA: esta tabla almacena información sobre la experiencia profesional relacionada con actividades de enseñanza realizadas por el investigador. La llave primaria es PK_ACT_DOCENCIA.

EN_ACT_INVESTIGACION: esta tabla almacena información sobre la experiencia profesional relacionada con actividades de investigación realizadas por el investigador. La llave primaria es PK_ACT_INVESTIGACION.

EN_AREA_CONOCIMIENTO: esta tabla almacena información de las áreas del conocimiento relacionadas con un programa académico para el registro de una formación académica, complementaria o áreas de actuación del investigador. La llave primaria es PK_AREA_CONOCIMIENTO.

Preparación de los datos

Durante esta fase, se detallará el proceso de preparación de los datos.

Extrayendo la información

Fecha de nacimiento, sexo, código de identificación interno de los investigadores (llave), nombre de la institución y tipo de institución:

Para obtener una base de datos completa de los investigadores reconocidos por Minciencias en las IES, que incluya información como su rango de edad, sexo, nombre, tipo y ubicación de la institución en la que trabajan, se requiere la utilización puntualmente de las siguientes seis tablas de la base de datos SCIENTI:

- En_recurso_humano de la base de datos OCYT_CV (CVLAC).
- En_trayectoria_profesional de la base de datos OCYT_CV.
- En_institucion de la base de datos OCYT_REF (Dic_Referencia).
- EN_TIPO_INSTITUCION de la base de datos OCYT_REF.
- EN_MUNICIPIO de la base de datos OCYT_REF.
- EN_DEPARTAMENTO de la base de datos OCYT_REF.

La tabla en_recurso_humano de la base de datos OCyT_CV (CVLAC) contiene información importante, como la fecha de nacimiento, sexo, cédula y el código de identificación interno (COD_RH) de los investigadores de la base de datos. La tabla en_trayectoria_profesional, por su parte, proporciona información acerca de las instituciones en las que el investigador ha trabajado o tiene vínculo laboral. Las tablas en_institucion, EN_TIPO_INSTITUCION, en_municipio y en_departamento

ofrecen información adicional, como el nombre, tipo, municipio y departamento de las instituciones registradas en la base de datos SCIENTI.

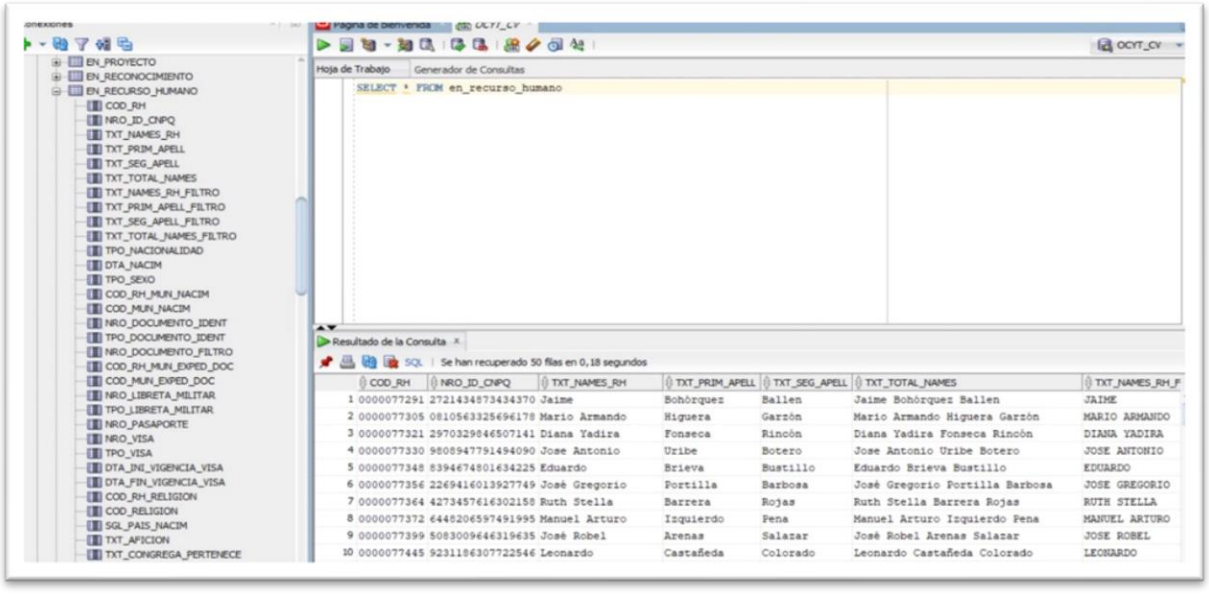
Edad, sexo e ID de la institución:

Para obtener una tabla de investigadores reconocidos en las IES, clasificados por edad, sexo e ID de la institución en la que actualmente trabajan, se utiliza la siguiente consulta SQL:

```
SELECT  a.nro_documento_ident, a.COD_RH, a.DTA_NACIM, a.TPO_SEXO, c.ID_INSTITUCION FROM en_recurso_humano a, en_trayectoria_profesional c WHERE a.cod_rh=c.cod_rh and c.STA_FILIACION_ACTUAL='T´
```

Para obtener una lista de todos los investigadores, junto con su COD_RH, cédula, fecha de nacimiento, sexo e ID de la institución en la que actualmente trabajan, se cruza la información de las tablas en_recurso_humano y en_trayectoria_profesional de la base de datos OCYT_CV. La columna STA_FILIACION_ACTUAL de la tabla en_trayectoria_profesional permite filtrar los resultados y mostrar solo las instituciones en las que el investigador tiene un vínculo laboral actualmente. Si STA_FILIACION_ACTUAL es igual a 'T', entonces el investigador trabaja en esa institución en la actualidad.

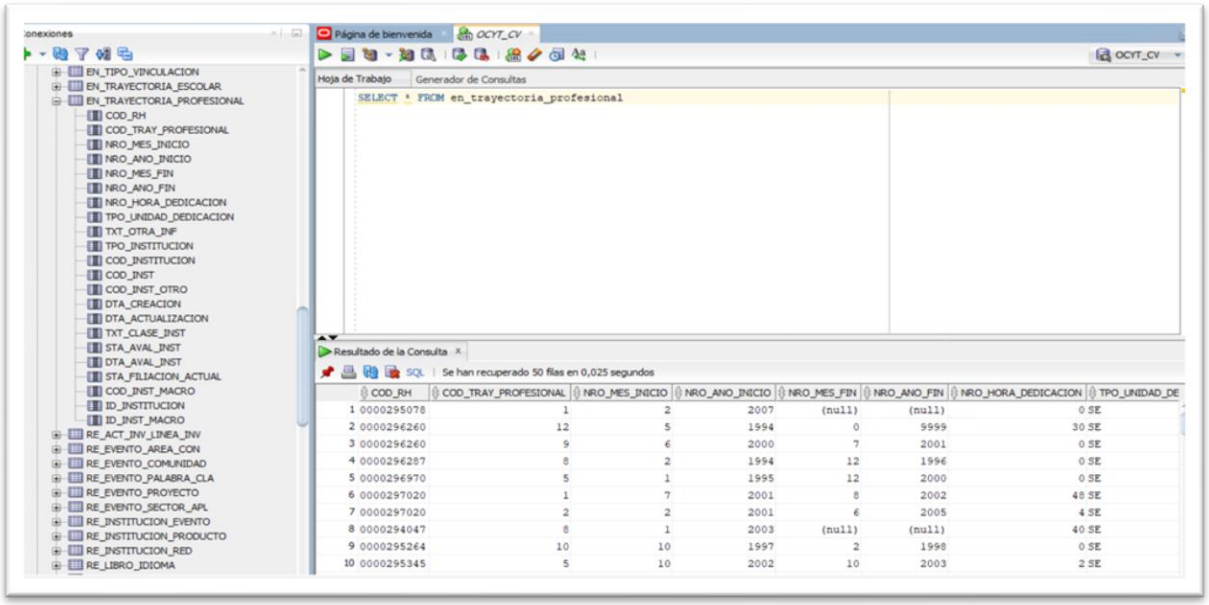
Ilustración 3 - Consulta a Tabla EN_RECURSO_HUMANO



Fuente: DBMS Oracle Sqldeveloper Versión 22.2.0.173

Es importante tener en cuenta que un investigador puede estar vinculado a varias instituciones, por lo que la base puede tener varias filas para el mismo investigador. Finalmente, organizando la consulta, se obtiene una nueva tabla con una sola fila por investigador que incluye el nombre de todas las instituciones en las que labora en una sola tabla. Esto permite obtener una lista clara y detallada de los investigadores reconocidos de las IES, clasificados por edad, sexo e ID de la institución.

Ilustración 4 - Tabla EN_TRAYECTORIA_PROFESIONAL



Fuente: DBMS Oracle Sqldeveloper Versión 22.2.0.173

Nombre, tipo y ubicación de la institución:

Para obtener el nombre de la institución en donde labora el investigador, se ejecuta la siguiente consulta en Oracle en la base OCYT_REF:

```
SELECT a.nme_inst, a.ID_INSTITUCION FROM en_institucion a.
```

Esta consulta proporciona una tabla con el nombre de la institución y su ID, lo que permite adicionar el nombre de la institución en una nueva columna en la base de datos de investigadores reconocidos.

Para consultar la información que indique si la institución es una IES y su carácter público, privado o mixto, se ejecuta la siguiente consulta SQL en la base OCYT_REF:

```
SELECT a. ID_TIPO_INSTITUCION, a.ID_INSTITUCION
```

Esta consulta proporciona dos ID: uno del tipo de institución y otro de la institución. De acuerdo con la tabla en_tipo_institucion de OCYT_REF, el primer número del "código tipo de institución" indica si la institución es un centro, una IES , una empresa o de otro tipo.

Por último, para agregar una columna que indique el departamento de la institución en donde labora el investigador, se deben realizar varios pasos:

Se debe extraer de la tabla "en_institucion" de la base de datos "OCYT_REF" una tabla que contenga los ID de las instituciones y los ID de los municipios donde estas están ubicadas mediante la consulta SELECT a.ID, a.id_municipio.

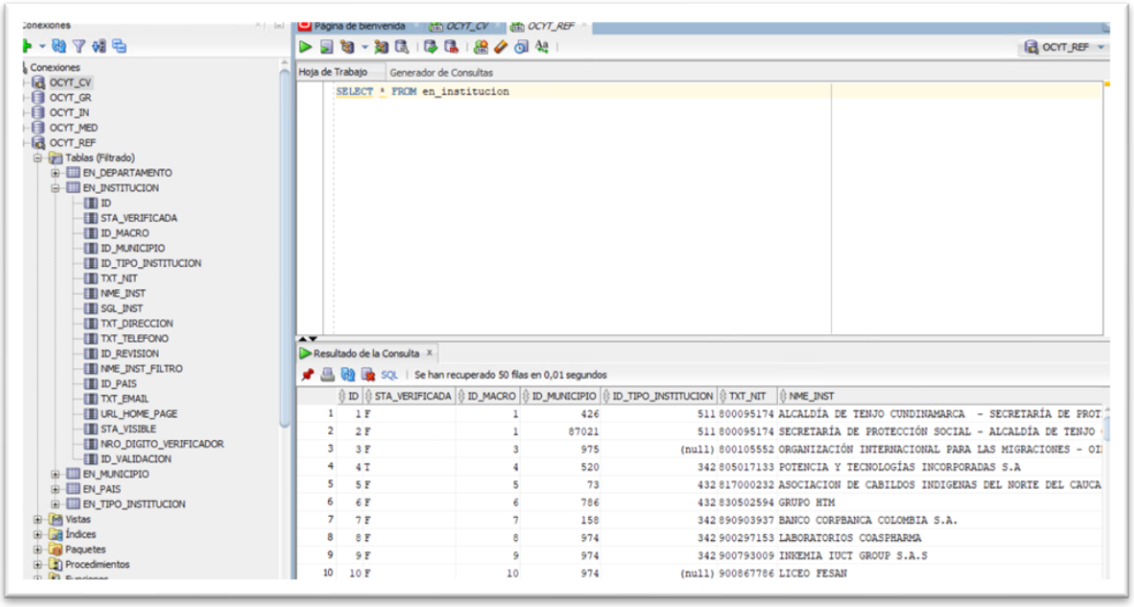
De igual manera, se debe extraer de la tabla "en_municipio" de la misma base de datos "OCYT_REF" una tabla que contenga los ID del municipio y el ID del departamento.

Con la ayuda de la segunda tabla, se procede a reemplazar los códigos ID del municipio por los códigos ID del departamento en la primera tabla.

Para llevar a cabo la siguiente etapa, se requiere extraer de la tabla "en_departamento" de la base de datos "OCYT_REF" el código ID del departamento y su nombre, de forma que se pueda realizar el reemplazo del código ID del departamento por su correspondiente nombre en la primera tabla, esto generará una tabla que contendrá el código de la institución y el nombre del departamento, lo que permitirá anexar la columna de departamento en la base de datos de investigadores reconocidos.

Finalmente, para obtener la base final de investigadores reconocidos en las IES, se filtra la columna de tipo de institución por IES públicas y privadas en la base de datos.

Ilustración 5 - Tabla EN_INSTITUCION



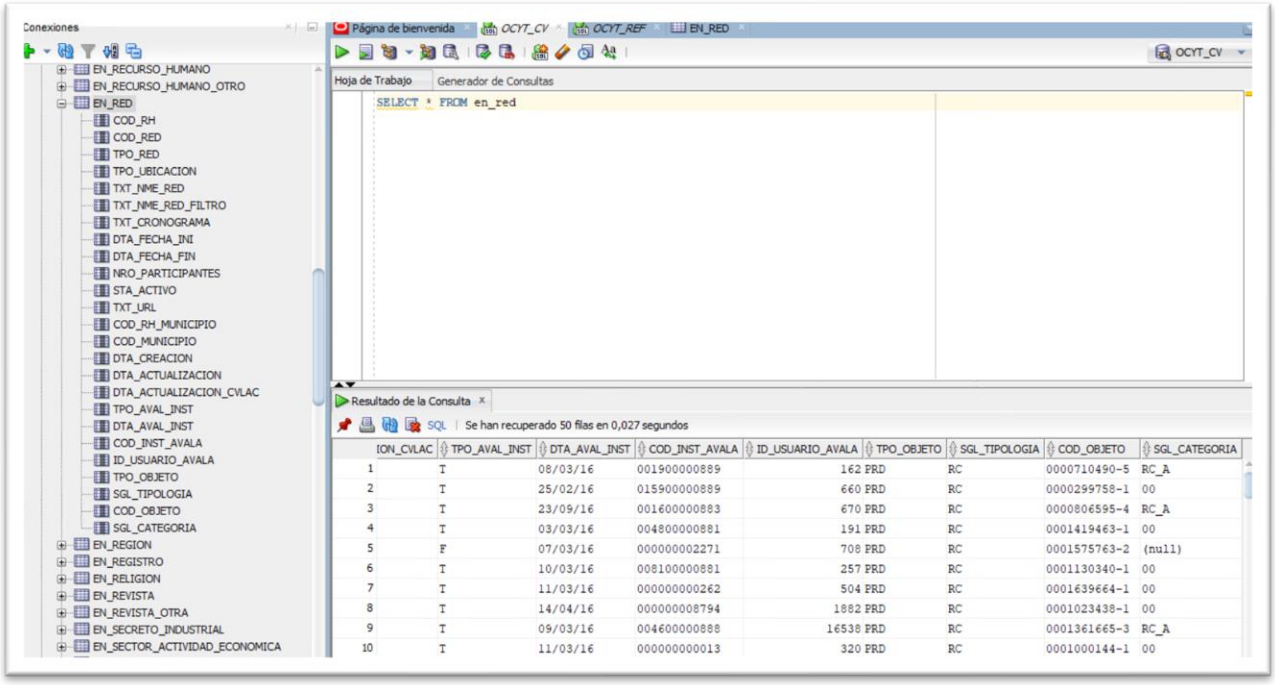
Fuente: DBMS Oracle Sqldeveloper Versión 22.2.0.173

Categorización de los productos de los investigadores de las IES

Se categorizan los productos en cuatro categorías: "Apropiación social del conocimiento y divulgación pública de la ciencia (ASC)", "Desarrollo tecnología e innovación (DTI)", "Formación del recurso humano (FRH)", y "Nuevo conocimiento (NC)".

Son extraídos de la tabla en_productos de la base OCyT_CV (CVLAC), utilizando la columna "SGL_CATEGORIA" para filtrar y seleccionar los productos correspondientes. Sin embargo, hay productos que no se encontraron en esta tabla, llamados “SE” . El producto "SE" se halla en la tabla en_secreto_industrial de OCyT_CV, y para obtenerlo también se filtró por las siglas del producto en la columna "SGL_CATEGORIA.

Ilustración 6 - Tabla EN_RED



Fuente: DBMS Oracle Sqldeveloper Versión 22.2.0.173

Los productos PE, PF_A, PF_B, PIC_A, PIC_B, PIC_C, PID_A, PID_B y PID_C fueron obtenidos de la tabla en_proyecto de la base OCyT_CV (CVLAC). Por otro lado, los productos APO se extrajeron de la tabla en_evento de la misma base de datos. En cuanto a los demás productos, se tomaron de la tabla en_productos de la base OCyT_CV (CVLAC). Para obtenerlos de estas tablas, se realizó un filtro por las siglas del producto en la columna “SGL_CATEGORIA”.

Los productos NC corresponden a más de la mitad de ellos se toman de la tabla en_productos de la base OCyT_CV (CVLAC). Para ello se utiliza la columna "SGL_CATEGORIA" para filtrar y extraer los productos que pertenecen a la categoría NC. Sin embargo, existen algunos productos que no se encontraban en esta tabla, como AAD_A, AAD_A1, AAD_B, AAD_C, MA1, MA2, MA3, MA4, MB2, MB4, MB5, MC, PA1, PA2, PA3, PA4, PB1, PB2, PB3, PB4, PB5 y PC. En particular, los productos AAD_A, AAD_A1, AAD_B y AAD_C se encuentran en la tabla en_prod_artistica_detalle de OCyT_CV, mientras que los productos MA1, MA2, MA3, MA4, MB2, MB4, MB5, MC, PA1, PA2, PA3, PA4, PB1, PB2, PB3, PB4, PB5 y PC están en la tabla en_patente de OCyT_CV. Para obtenerlos, se filtra por las siglas del producto en la columna "SGL_CATEGORIA" de cada tabla correspondiente, mediante consultas SQL que permitan su filtrado.

Resumen del proceso:

Después de extraer y realizar las consultas necesarias a la base de datos, se generan dos tablas importantes. El primer reporte es un recuento total de los investigadores reconocidos por Minciencias que pertenecen a las IES, que es de **16310**. Este reporte contiene varios campos importantes que se utilizarán en análisis posteriores. Entre ellos se encuentran el código del investigador (PK), el nivel de formación, la IES a la que pertenece, el área de conocimiento OCDE y el rango de edad del investigador según su nivel de formación.

El segundo reporte, por su parte, es una descripción detallada de la producción científica de cada investigador, incluyendo el tipo de producto y categoría. Este reporte también incluye el campo del código del investigador, que será útil para cruzar los datos con el primer reporte y realizar análisis más específicos.

Estos dos reportes son esenciales para poder analizar la producción científica de las IES y sus investigadores, así como para hacer comparaciones entre diferentes instituciones y campos de conocimiento. A partir de ellos se pueden identificar tendencias y áreas de oportunidad para mejorar la producción científica en las IES del país.

Despliegue

En esta fase de despliegue, se describe el proceso de implementación del sistema en el entorno de producción.

Carga de datos

Para llevar a cabo la carga de datos al servidor SQL, se utilizó el software Open Talend Studio. Se ha diseñado un Job llamado **ETL_INV_CON_894_IES** que tiene como objetivo realizar la extracción, transformación y carga de los datos de los investigadores y su producción científica.

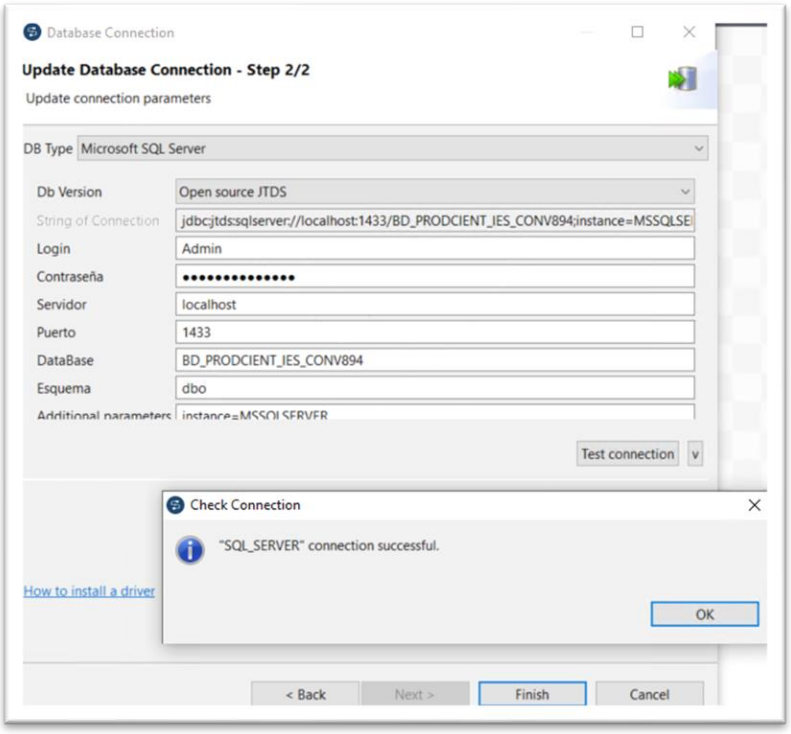
En la base de datos se encontraron datos de diversos tipos, incluyendo valores enteros, decimales y otros formatos. Para asegurar la calidad y consistencia de los datos, se llevó a cabo un proceso de ETL, el cual implicó la extracción, transformación y carga de los datos. En este proceso, los datos fueron estandarizados para que pudieran ser utilizados de manera coherente y fiable. La estandarización incluyó la conversión de los valores a un formato común y la eliminación de datos redundantes o irrelevantes. De esta manera, los datos se prepararon adecuadamente para su análisis y uso en la toma de decisiones.

Este Job **ETL_INV_CON_894_IES** permite mantener actualizada la información de los investigadores y su producción científica en el servidor SQL(Ver ilustración 8).

Parámetros de conexión a SQL SERVER desde Talend Open Studio:

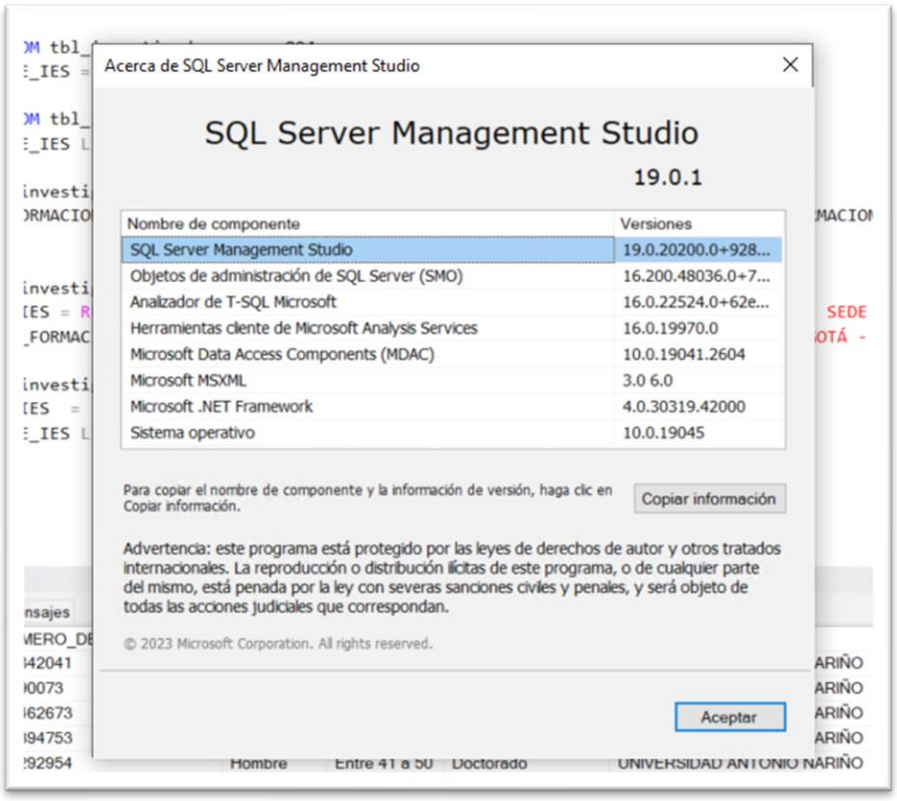
Para conectar a SQL Server Onpremise desde Talend, se debe ejecutar Talend y crear un nuevo proyecto. A continuación, en el panel "Repository", se debe hacer clic con el botón derecho del mouse y seleccionar "Create Connection". Se debe seleccionar el tipo de conexión "Microsoft SQL Server" y completar la información de conexión, incluyendo el nombre de la conexión, el servidor al que se quiere conectar, el puerto de conexión, la base de datos a la que se desea acceder y las credenciales de inicio de sesión del usuario. Una vez que se ha completado toda la información de conexión, se debe probar la conexión para verificar que se establece correctamente. Si la prueba es exitosa, se puede guardar la conexión y utilizarla para acceder a la base de datos de SQL Server desde Talend. (Ver ilustración 7).

Ilustración 7 - Configuración conexión hacia servidor SQL SERVER



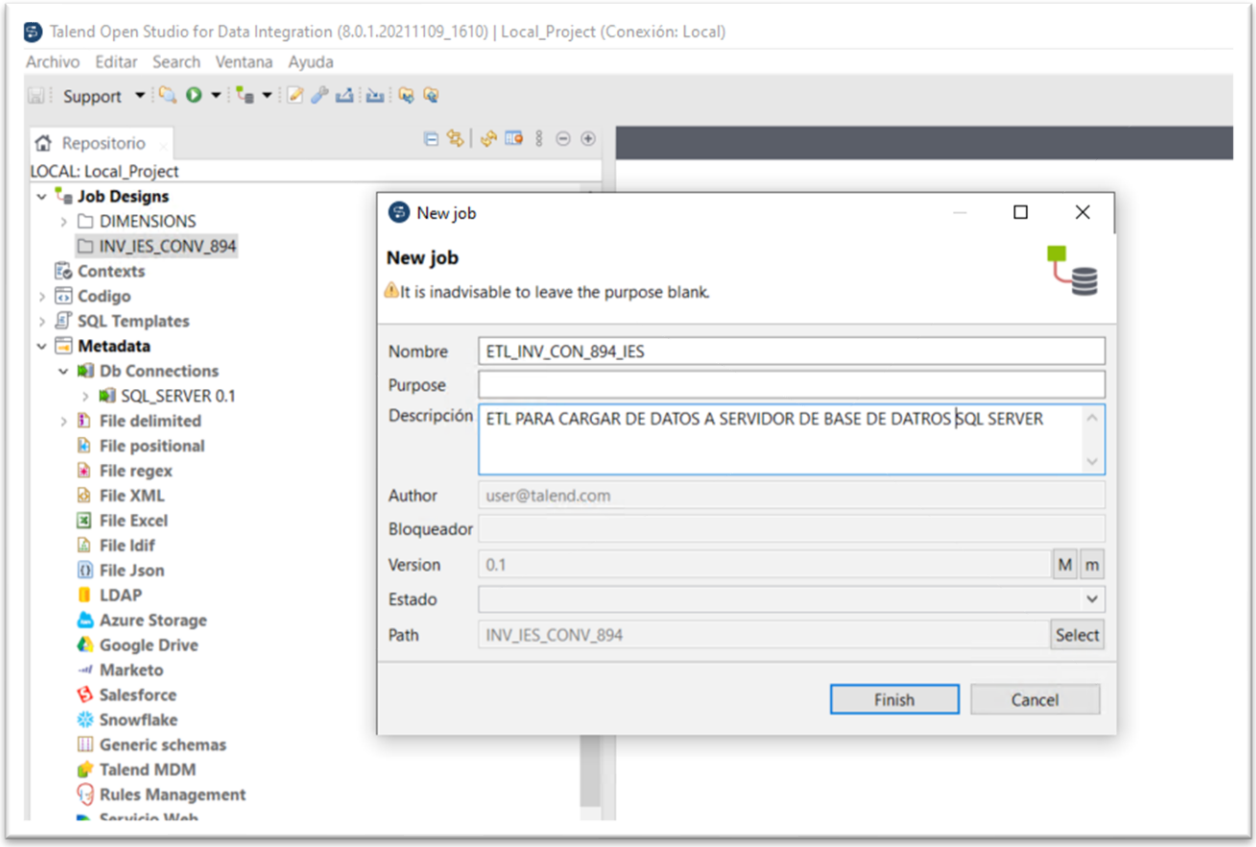
Fuente: Talend Open Studio for Data Integration - Version: 8.0.1

Ilustración 8 - Versión SQL SERVER



Fuente: SQL Server Management Studio - 19.0.20200.0+9286509b

Ilustración 9 - Creación Job Talend Open Studio



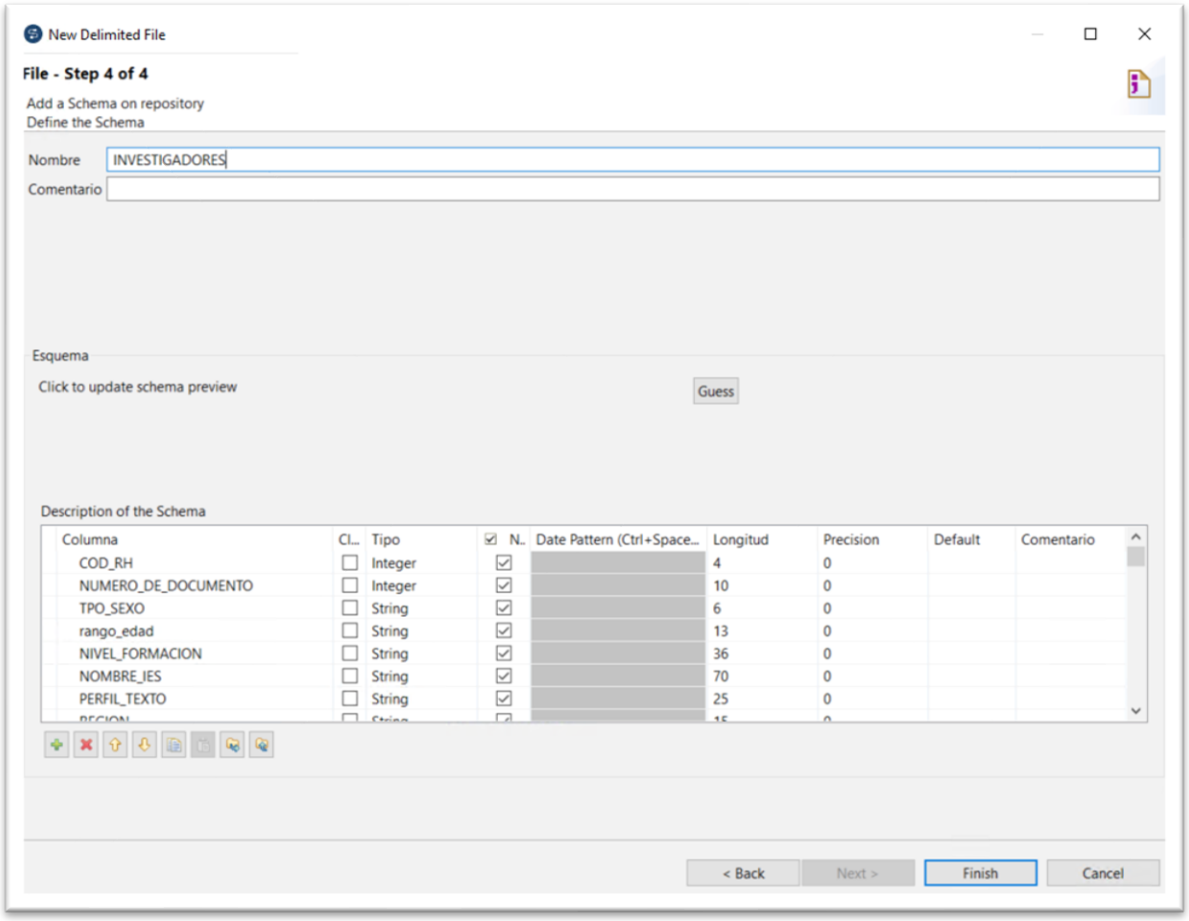
Fuente: Talend Open Studio for Data Integration - Version: 8.0.1

El proceso de conexión de un set de datos o metadata en Talend Open Studio puede variar dependiendo del tipo de archivo o base de datos que se desee utilizar (Ver ilustración 10).

Para este caso específico, se extrajeron dos tablas de Oracle y se cargaron en Talend Open Studio siguiendo los siguientes pasos:

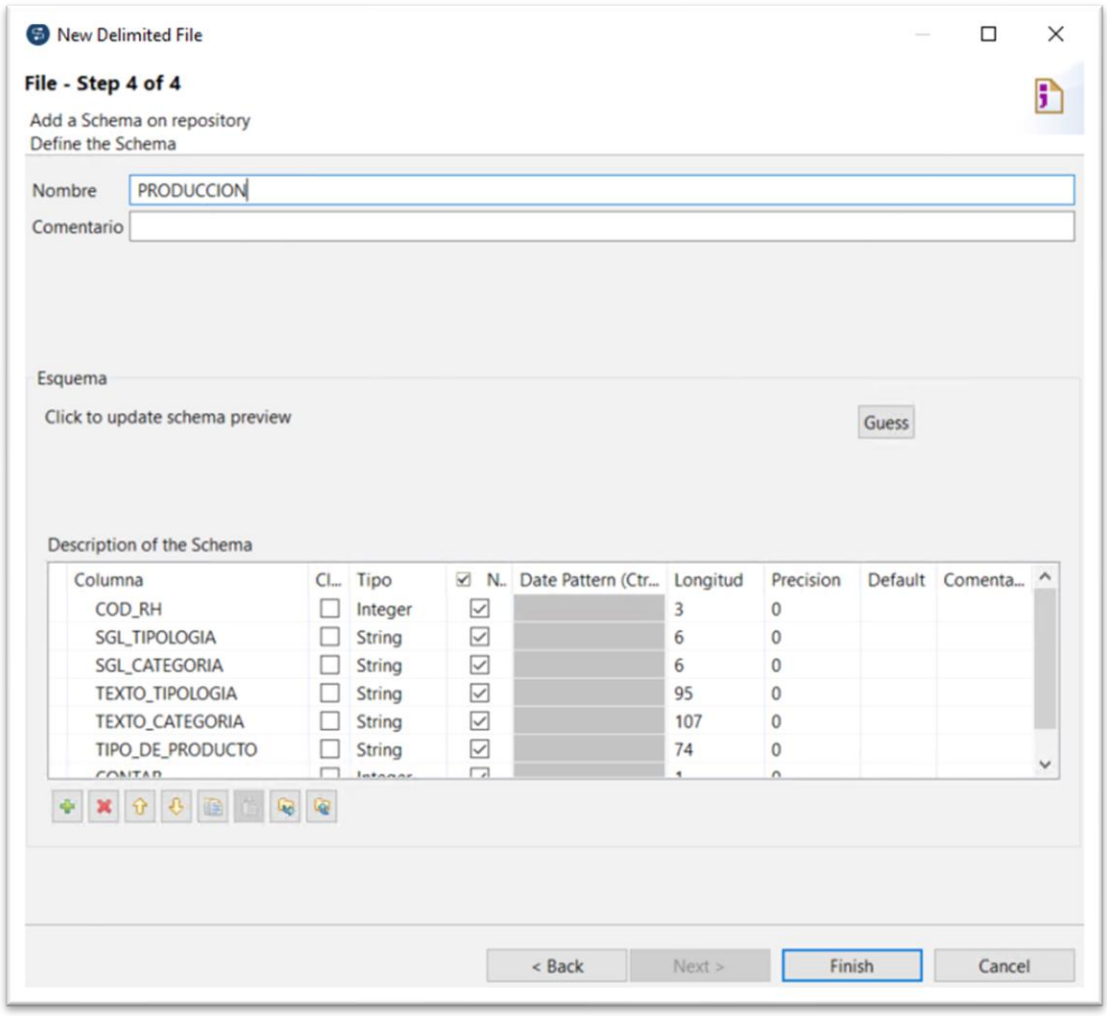
1. Seleccionar "Metadata" en el menú superior de Talend Open Studio y luego "Create a connection" para crear una nueva conexión.
2. Elegir el tipo de conexión deseado, como "Database" si se desea conectar a una base de datos, o "File" si se desea conectar a un archivo.
3. Ingresar los detalles de la conexión, como el nombre de la conexión, el tipo de base de datos o archivo, la dirección del servidor o la ruta del archivo, y las credenciales de acceso si es necesario.
4. Probar la conexión para asegurarse de que funciona correctamente.

Ilustración 10 - Carga set de datos Investigadores



Fuente: Talend Open Studio for Data Integration - Version: 8.0.1

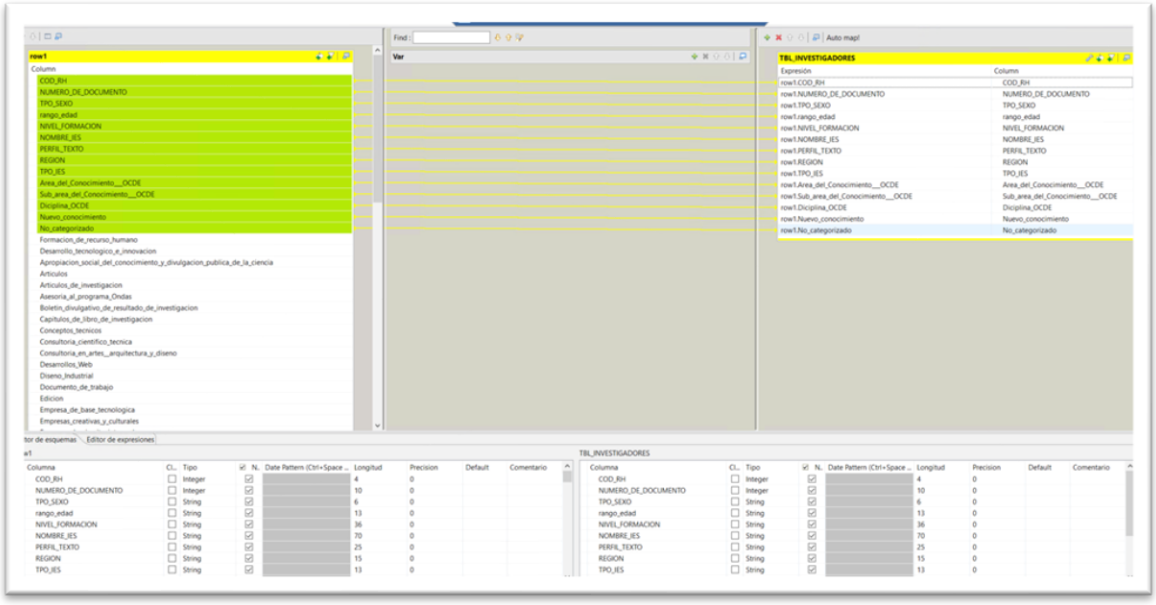
Ilustración 11 - Carga set de datos Producción Científica



Fuente: Talend Open Studio for Data Integration - Version: 8.0.1

De la tabla de Investigadores solo se extrajeron y cargaron los datos necesarios para la creación del reporte (Ver ilustración 12).

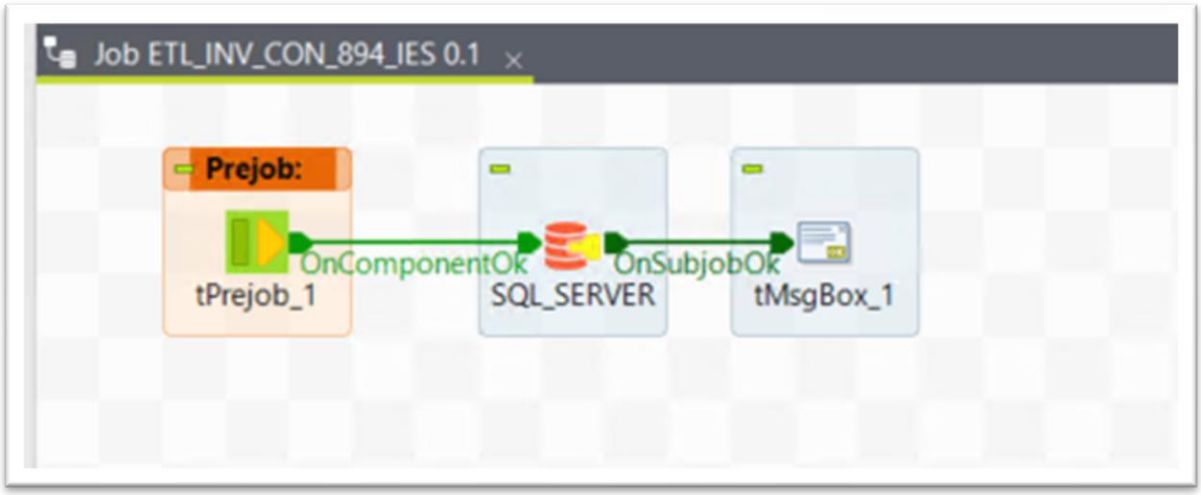
Ilustración 12 - Tmap Talend Open Studio



Fuente: Talend Open Studio for Data Integration - Version: 8.0.1

Antes de cargar los datos se verificó que la conexión al servidor SQL SERVER esté correctamente configurado, antes de comenzar la carga, para ello se utilizó en Talend Open Studio, el pre-job y el tMsgBox, que son componentes que se utilizan para llevar a cabo ciertas acciones antes de ejecutar un job principal (Ver ilustración 13).

Ilustración 13 - Verificación pre-job Talend Open Studio

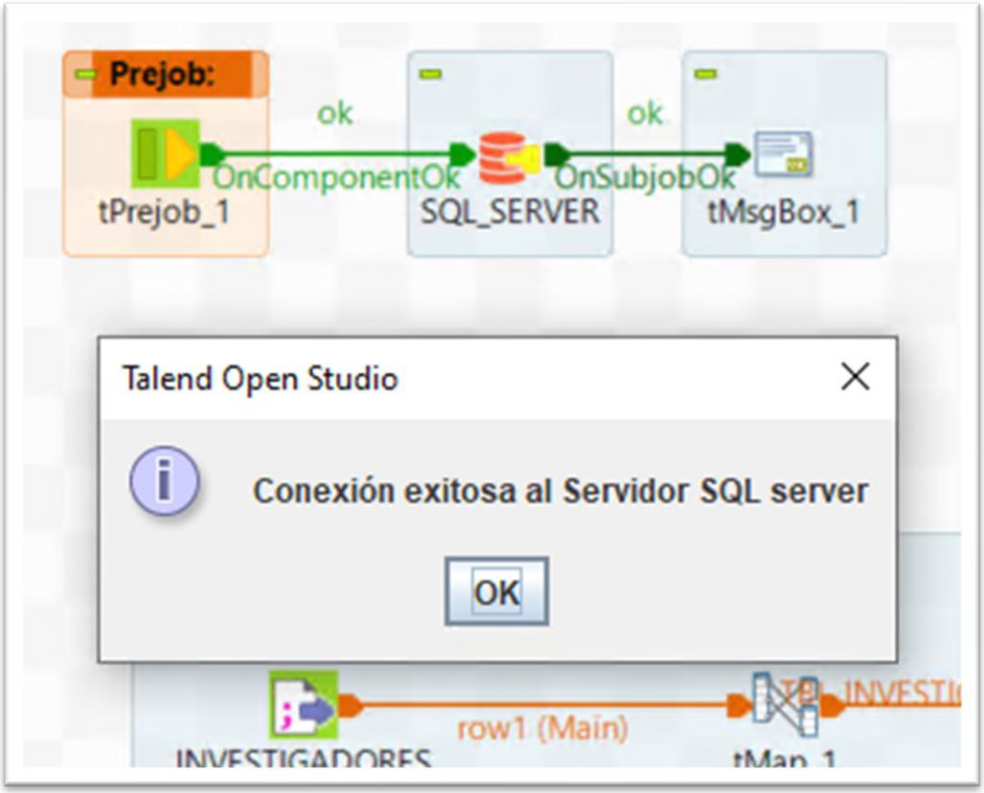


Fuente: Talend Open Studio for Data Integration - Version: 8.0.1

El pre-job es un componente que se utiliza para realizar tareas que deben ejecutarse antes de que comience el job principal. Por ejemplo, se puede utilizar para inicializar variables o realizar comprobaciones de conexión antes de procesar los datos.

El tMsgBox, por su parte, es un componente que se utiliza para mostrar mensajes de confirmación al usuario antes de ejecutar el job principal, para este caso se parametrizó el mensaje de salida ”Conexión exitosa al servidor SQL server”, por ejemplo, también se puede utilizar para pedir al usuario que confirme que desea procesar un gran volumen de datos o que ha realizado las copias de seguridad necesarias antes de iniciar el procesamiento (Ver ilustración 14).

Ilustración 14 - Conexión exitosa a SQL SERVER



Fuente: Talend Open Studio for Data Integration - Version: 8.0.1

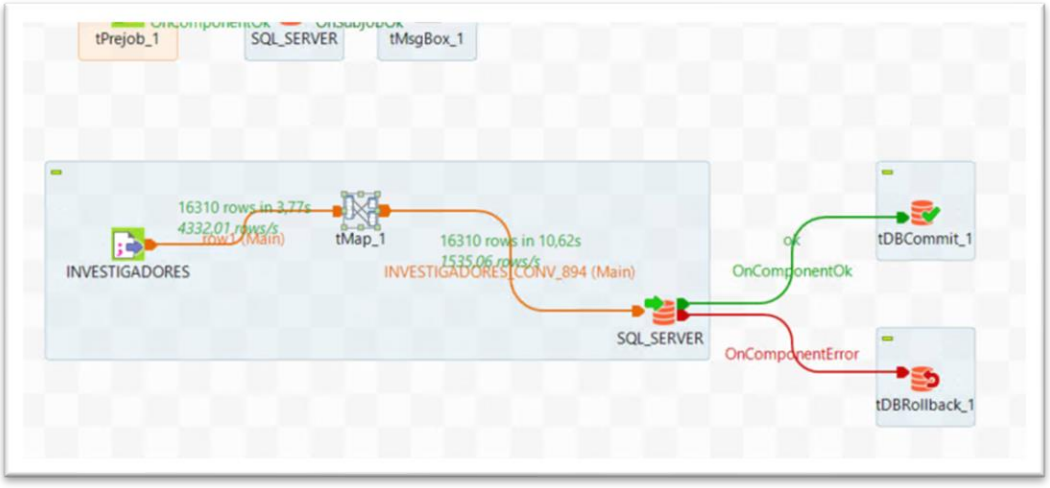
Ejecución de Jobs - ETL

Se ejecuta el job **ETL_INV_CON_894_IES** para cargar datos a la tabla "tbl_investigadores_conv894" en SQL SERVER. Se ha verificado que la ejecución del JOB fue exitosa y se han cargado los **16310** registros de investigadores reconocidos por Minciencias que pertenecen a las Instituciones de Educación Superior en Colombia - IES.

Se utiliza el componente TMap de Talen Open Studio, este es un componente de transformación de datos en Talend que permite a los usuarios transformar y enrutar datos entre fuentes y destinos diferentes. Al utilizar TMap, los usuarios pueden mapear los campos de entrada con los campos de salida, aplicar transformaciones y filtros a los datos y redirigir los registros a diferentes flujos de salida (Ver ilustración 15).

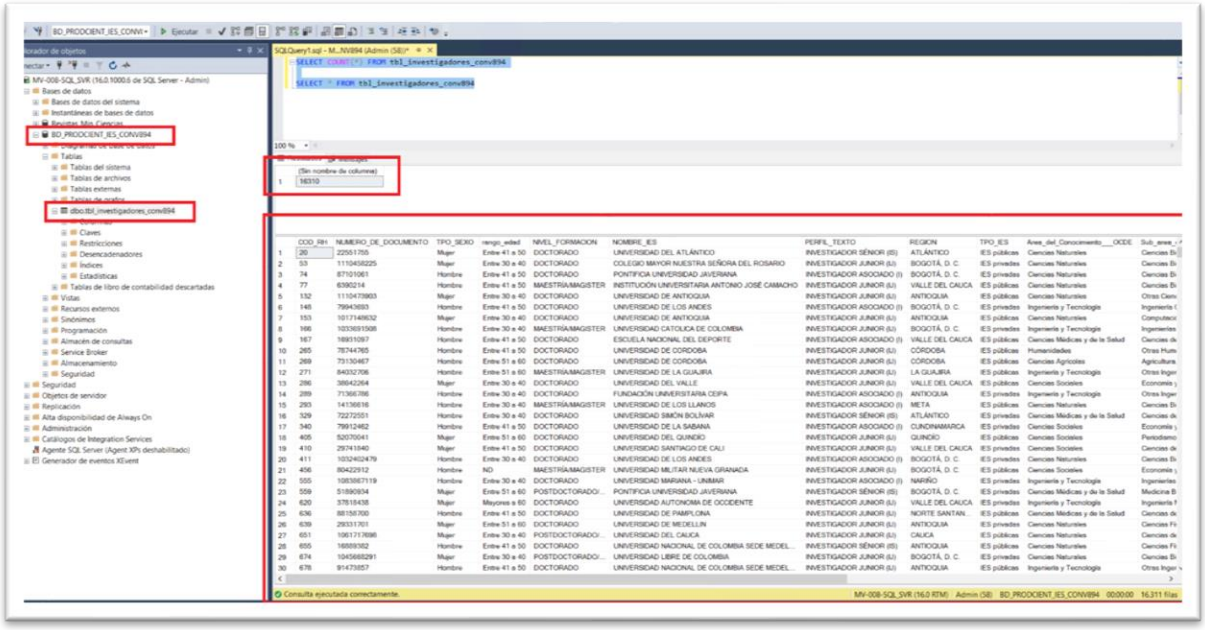
En el caso del JOB **ETL_INV_CON_894_IES**, el uso de TMap es fundamental para realizar la transformación de los datos de entrada, de forma que sean compatibles con la estructura y el formato de la tabla "tbl_investigadores_conv894". Además, el TMap permite filtrar los registros que no cumplan con los requisitos para ser cargados en la tabla, asegurando la integridad de los datos y evitando errores en la carga. Por lo tanto, la utilización de TMap sería esencial para llevar a cabo la tarea de carga de datos en la tabla "tbl_investigadores_conv894" de manera eficiente y efectiva (Ver ilustración 16).

Ilustración 15 - Ejecución Job 1 ETL



Fuente: Talend Open Studio for Data Integration - Version: 8.0.1

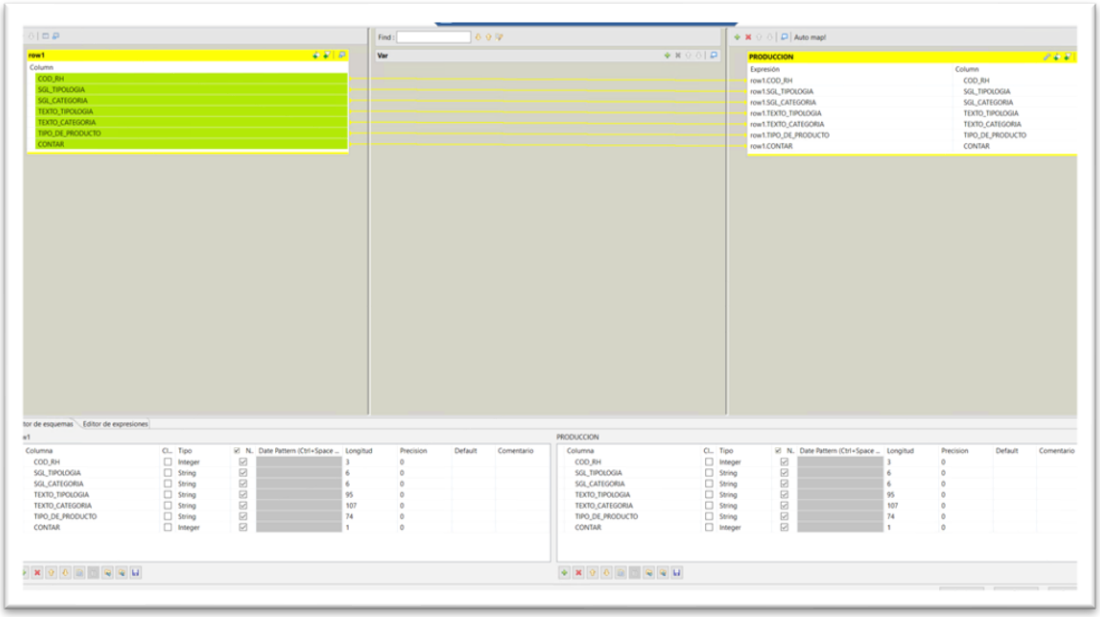
Ilustración 16 - Verificación de carga en SQL SERVER



Fuente: SQL Server Management Studio - 19.0.20200.0+9286509b

En un segundo Job **ETL_PROD_CONV_894_IES** se carga la tabla con la información de la producción científica de los investigadores (Ver ilustración 17).

Ilustración 17 - Tmap tabla producción

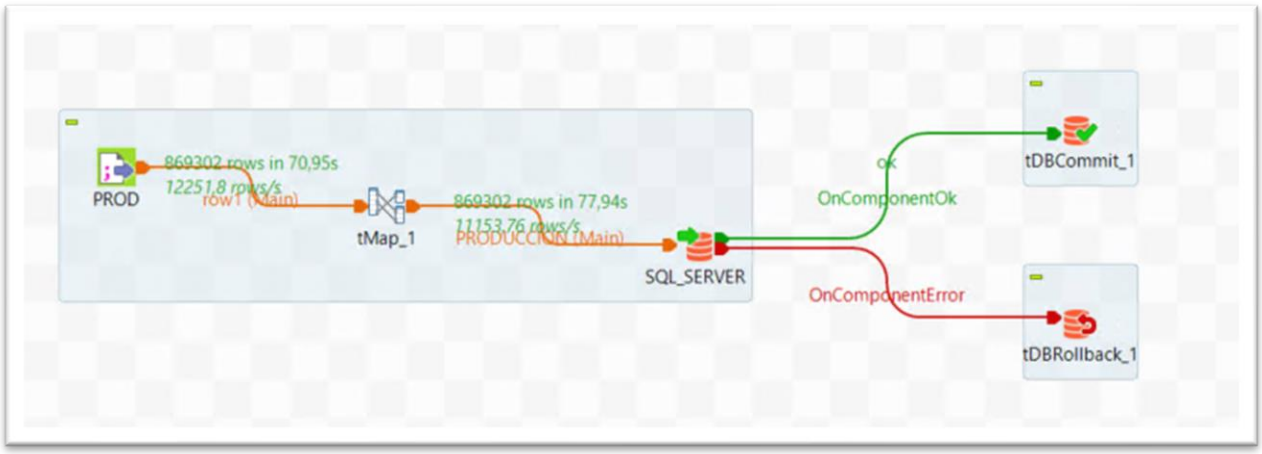


Fuente: Talend Open Studio for Data Integration - Version: 8.0.1

El JOB "ETL_PROD_CONV_894_IES" ha sido ejecutado para cargar datos en la tabla "tbl_produccion_conv894" en SQL SERVER. La ejecución del JOB fue verificada y se ha constatado que fue exitosa, ya que se cargaron los 869302 registros correspondientes a la producción científica de los investigadores reconocidos por Minciencias pertenecientes a las Instituciones de Educación Superior en Colombia (Ver ilustración 18).

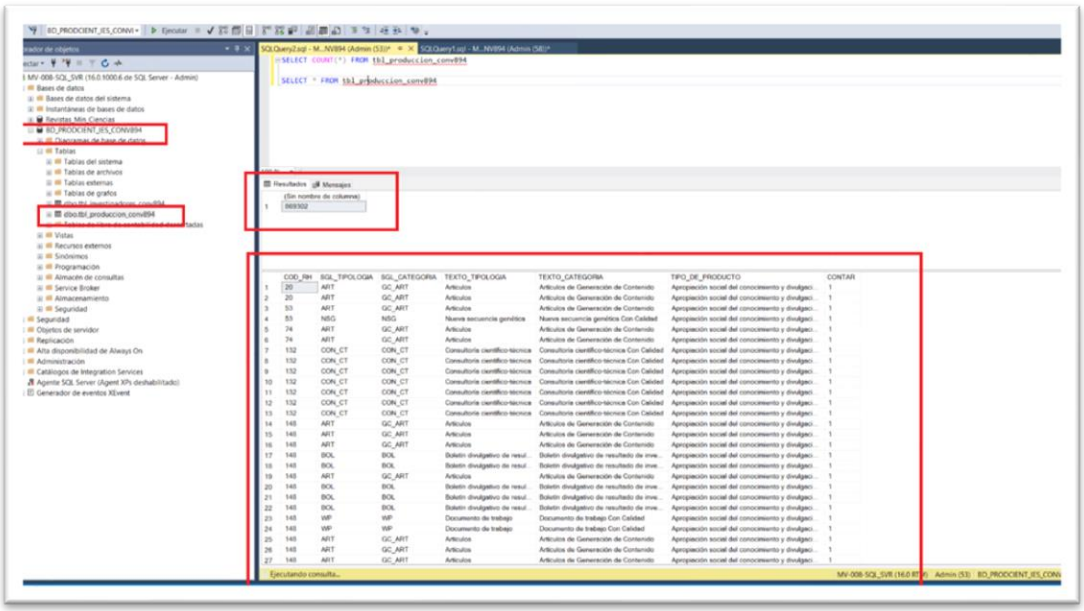
Para realizar esta tarea, se utilizó el componente TMap de Talend Open Studio. Este componente es esencial para la transformación de datos en Talend, ya que permite mapear los campos de entrada con los campos de salida, aplicar transformaciones y filtros a los datos, y redirigir los registros a diferentes flujos de salida. De esta manera, se puede asegurar que los datos de entrada sean compatibles con la estructura y el formato de la tabla de destino, y garantizar la integridad de los datos durante la carga, de aquí la importancia de la transformación de datos en el proceso de ETL (Ver ilustración 19).

Ilustración 18 - Ejecución job 2 ETL



Fuente: Talend Open Studio for Data Integration - Version: 8.0.1

Ilustración 19 - Verificación carga SQL server job 2



Fuente: Talend Open Studio for Data Integration - Version: 8.0.1

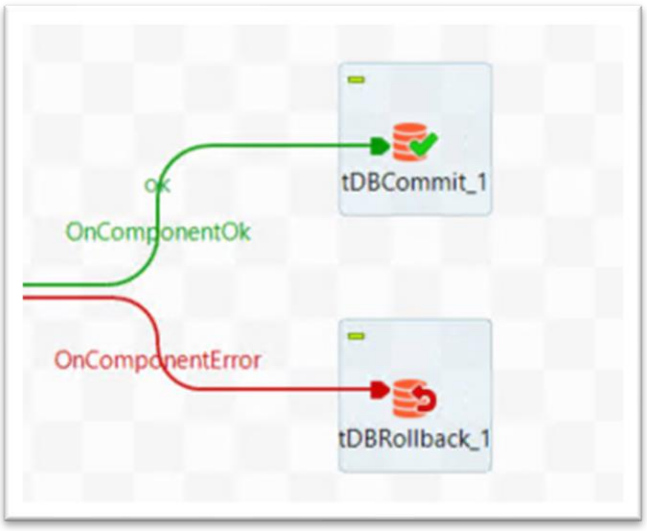
En los dos JOB se utilizaron los componentes **Commit** y **Rollback** que son empleados para controlar las transacciones de bases de datos durante la ejecución de un JOB.

El componente "Commit" permite confirmar los cambios realizados en la base de datos, es decir, asegura que los cambios realizados sean permanentes. Si se utiliza este componente en un JOB, se confirmarán los cambios realizados en la base de datos solo si se completa la ejecución del JOB correctamente. Si el JOB falla, los cambios realizados en la base de datos no serán confirmados.

Por otro lado, el componente "Rollback" se utiliza para deshacer los cambios realizados en la base de datos en caso de que el JOB falle. Es decir, si un JOB falla durante la ejecución, el componente Rollback asegura que todos los cambios realizados en la base de datos sean revertidos. Esto garantiza que la integridad de los datos en la base de datos sea mantenida.

Los componentes Commit y Rollback son esenciales para asegurar la integridad de los datos en una base de datos y permiten controlar las transacciones de bases de datos durante la ejecución de un JOB en Talend (Ver ilustración 20).

Ilustración 20 – Commit y Rollback Talend Open Studio



Fuente: Talend Open Studio for Data Integration – Version: 8.0.1

Elaboración de dashboard

El dashboard se concibe como una herramienta que ofrecerá información detallada sobre las áreas de conocimiento, rango de edad, género y tipos de productos de los investigadores, lo que facilitará el análisis y la comprensión de la producción científica en las IES de Colombia.

Para la elaboración del dashboard se utilizó el software Power BI con licencia PRO. Esta herramienta de Business Intelligence permite conectar, analizar y visualizar datos de diversas fuentes para crear informes interactivos y paneles de control en tiempo real.

La versión PRO de Power BI ofrece funcionalidades adicionales, como la capacidad de publicar informes y paneles en línea, compartirlos con otros usuarios y colaborar en tiempo real en proyectos. Además, permite integrar el dashboard en otras aplicaciones y sistemas, como SharePoint o Microsoft Teams, y proporciona acceso a servicios en la nube para el almacenamiento y procesamiento de datos.

Para conectarse a una base de datos SQL Server desde Power BI, se deben seguir los siguientes pasos.

Conexión a base de datos

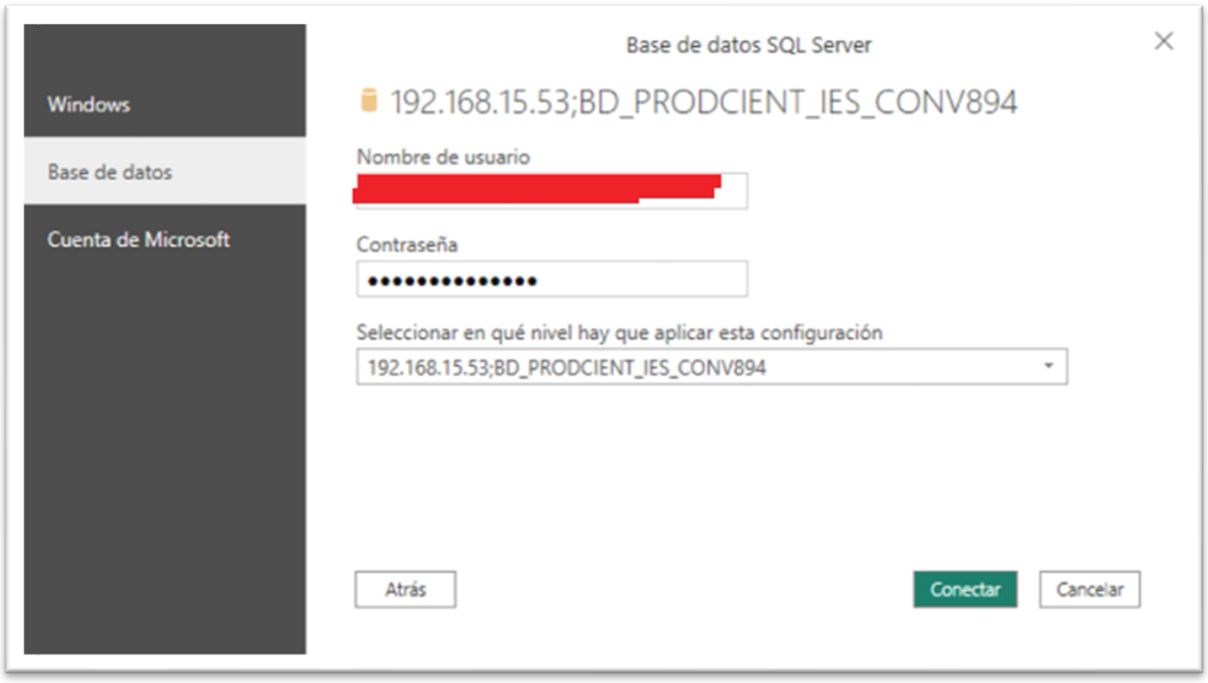
Desde Power BI Desktop se selecciona la opción “Obtener Datos” en la página de inicio.

En la ventana emergente que aparece, seleccionar la opción “SQL Server” en la lista de opciones disponibles y hacer clic en el botón “Conectar” (Ver ilustración 21).

En la siguiente ventana, ingresar el nombre del servidor y la base de datos a la que se desea conectarse.

En la sección de "Método de autenticación", el usuario puede elegir entre las opciones de autenticación de Windows o de SQL Server. Si se selecciona la autenticación de SQL Server, se deben ingresar las credenciales de autenticación necesarias para acceder a la fuente de datos (Ver ilustración 21).

Ilustración 21 - Conexión a base de datos SQL SERVER desde Power BI



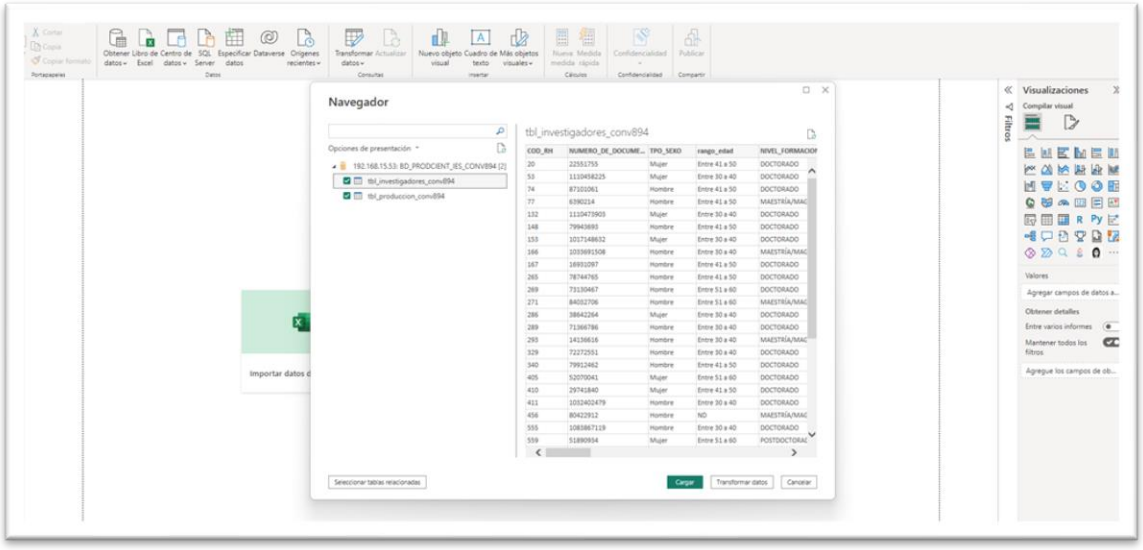
Fuente: Power BI - Versión: 2.115.663.0 64-bit (marzo de 2023)

Hacer clic en el botón "Conectar" para conectarse a la base de datos.

Una vez conectado, Power BI mostrará una lista de tablas y vistas disponibles en la base de datos. Se puede seleccionar las tablas que previamente se cargaron en SQL SERVER y vistas que desea utilizar para su análisis y visualización de datos.

También es posible realizar consultas personalizadas utilizando el lenguaje SQL para extraer y transformar los datos de la base de datos. Power BI ofrece una interfaz gráfica fácil de usar para crear visualizaciones y paneles de control basados en los datos extraídos de la base de datos SQL Server (Ver ilustración 22).

Ilustración 22 - Tablas para cargar en PowerBI



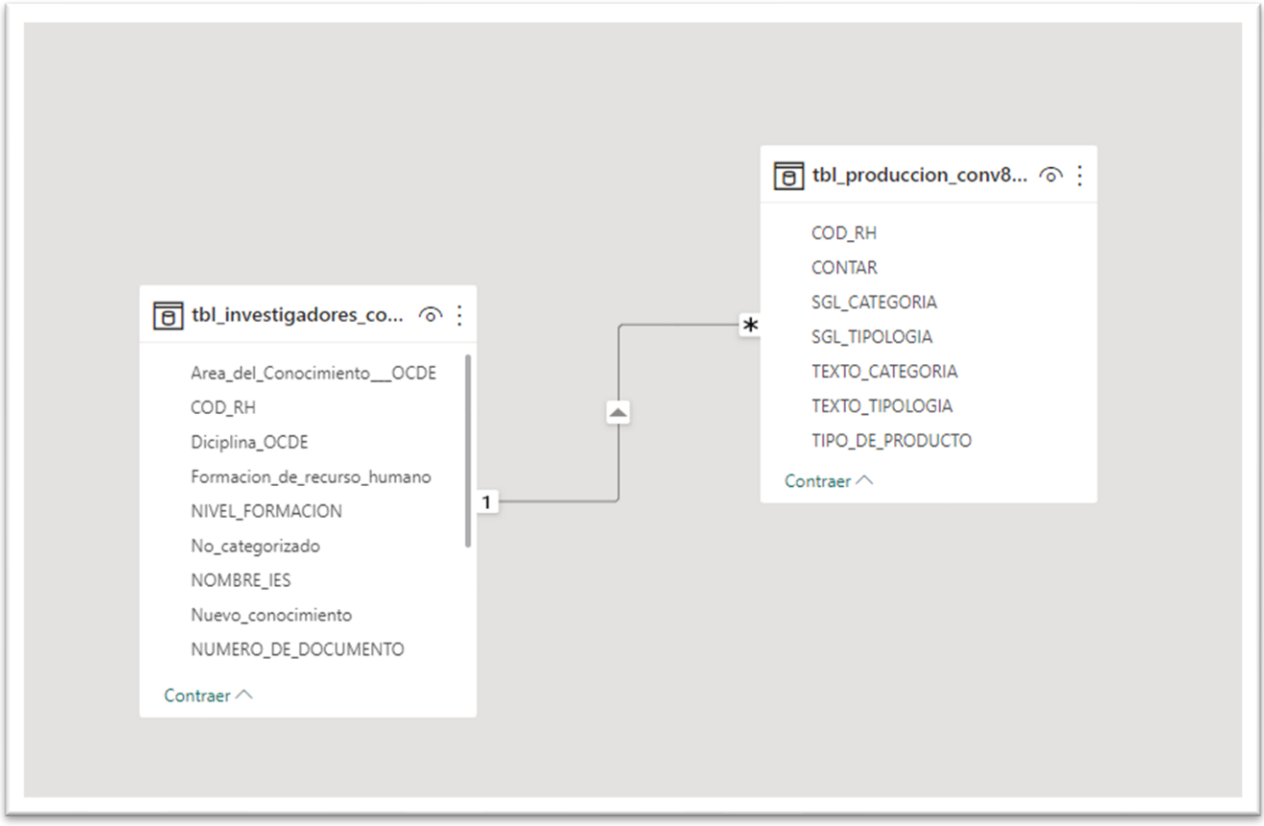
Fuente: Power BI - Versión: 2.115.663.0 64-bit (marzo de 2023)

Llaves:

Se realiza la unión de las tablas con llave primaria **(COD_RH)** esto permitirá combinar los datos de diferentes tablas (tbl_investigadores_conv894 y tbl_produccion_conv894). En muchas ocasiones, los datos se almacenan en múltiples tablas que están relacionadas entre sí, y la información completa sólo puede obtenerse al combinar estas tablas. La unión de tablas con llave primaria facilita la creación de informes y paneles de control que muestran una visión completa de los datos. Además, la unión de tablas también puede mejorar el rendimiento del informe al reducir la cantidad de datos que se deben cargar y procesar.

La llave primaria es esencial en la unión de tablas con Power BI, ya que garantiza que la información se combine correctamente. Al utilizar una llave primaria para unir las tablas, se asegura que cada registro se encuentre en la tabla correcta y se evita la duplicación de datos. Además, si se utilizan las relaciones de llave primaria-llave foránea, se puede establecer la cardinalidad de la relación entre las tablas, lo que permite una mayor precisión en los resultados de las consultas (Ver ilustración 23).

Ilustración 23 - Relación tablas

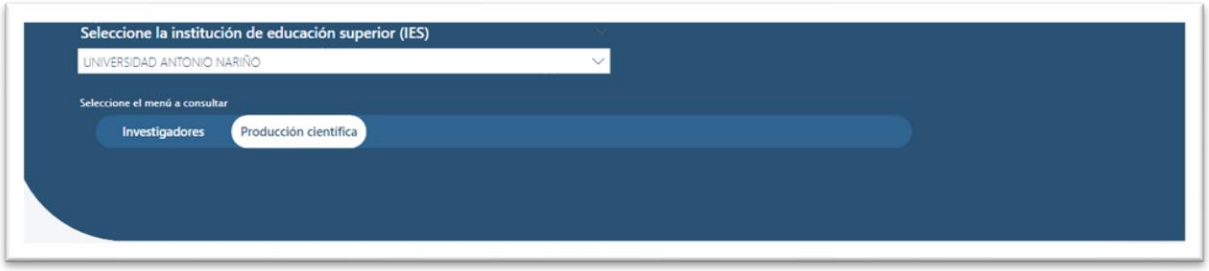


Fuente: Power BI - Versión: 2.115.663.0 64-bit (marzo de 2023)

Dashboard final

En el panel principal del dashboard, los usuarios pueden seleccionar la Institución de Educación Superior (IES) que desean consultar mediante un filtro específico (Ver ilustración 24).

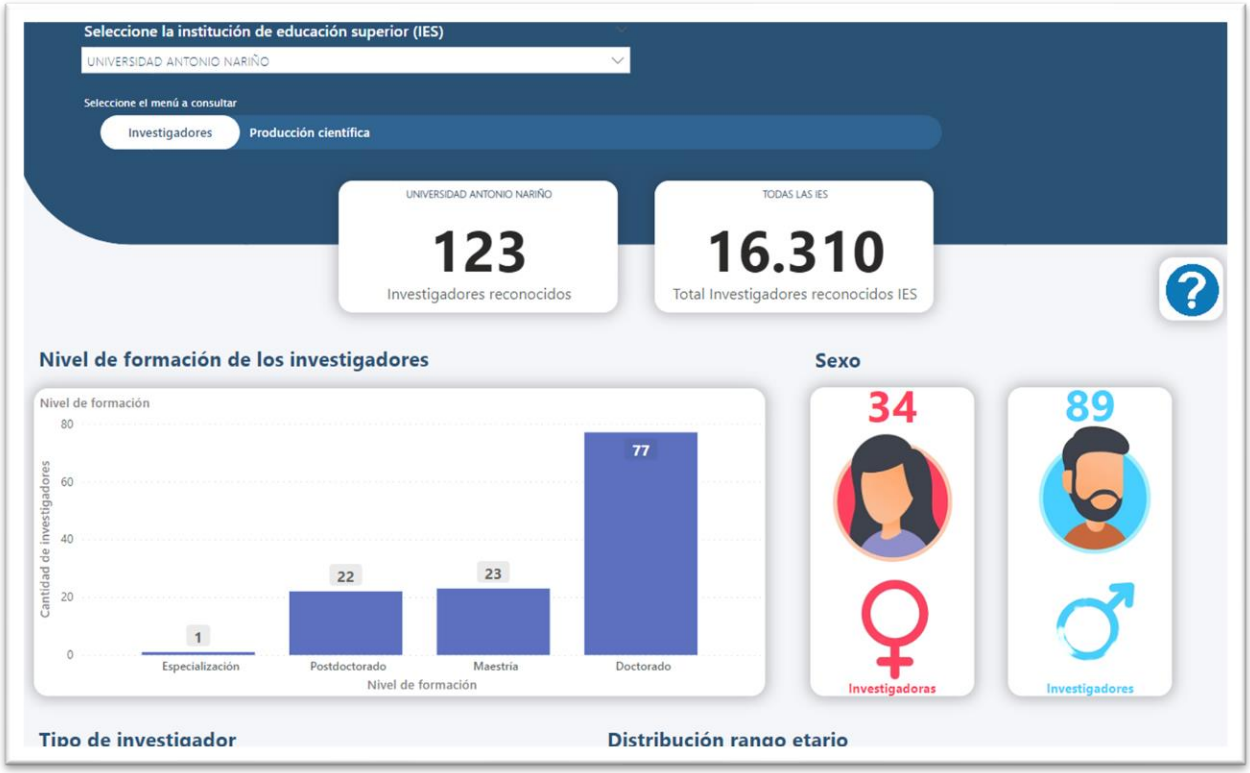
Ilustración 24 - Filtro IES



Fuente: Power BI - Versión: 2.115.663.0 64-bit (marzo de 2023)

Igualmente, se cuenta con un menú que permite una navegación fácil entre diferentes secciones del dashboard, como "INVESTIGADORES" y "PRODUCCIÓN CIENTÍFICA". A través de gráficos de barras y tarjetas informativas, el usuario puede explorar la información de manera intuitiva y visual. Además, se han incluido filtros que permiten a los usuarios ajustar la información de acuerdo con sus necesidades, lo que facilita el proceso de encontrar información específica de interés (Ver ilustración 25).

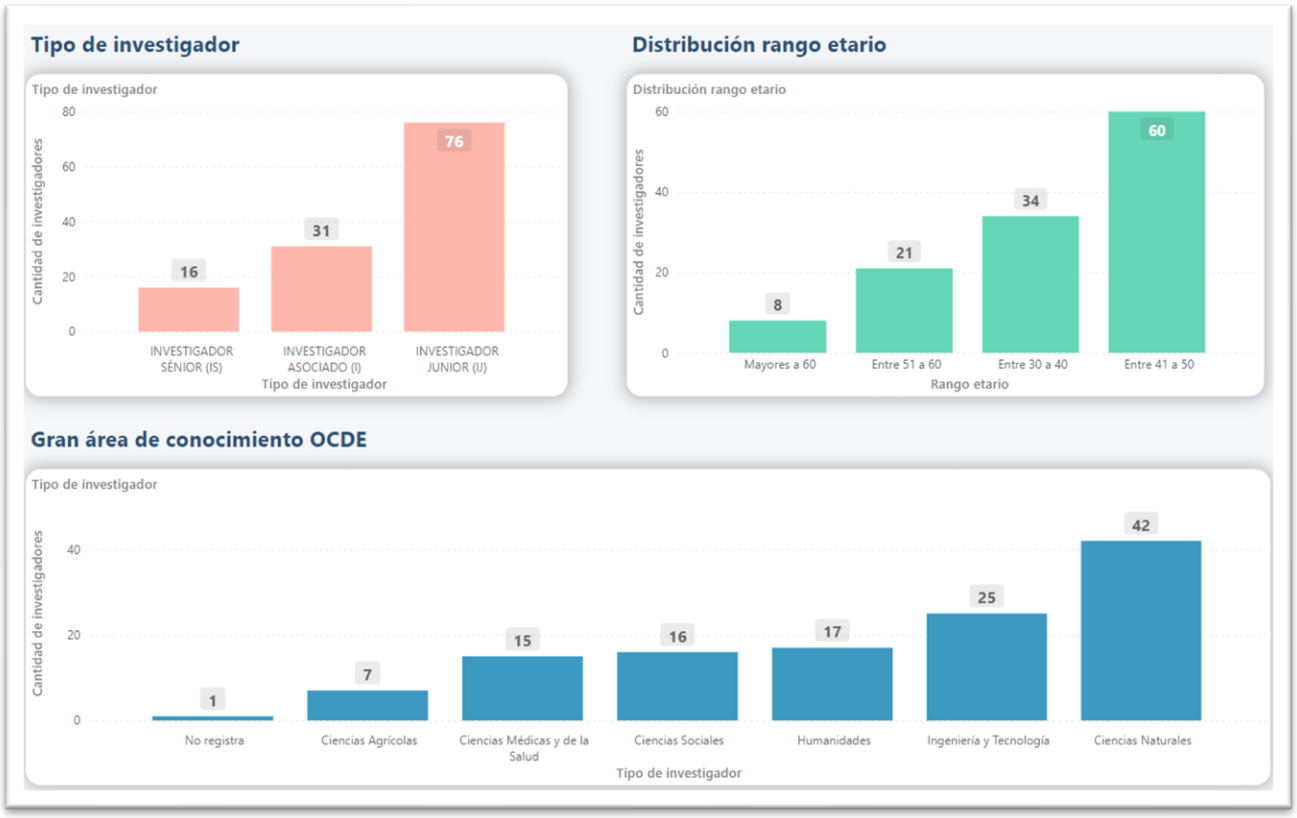
Ilustración 25 - Inicio dashboard



Fuente: Power BI - Versión: 2.115.663.0 64-bit (marzo de 2023)

Además, el dashboard también ofrece una sección para consultar la producción científica de los investigadores reconocidos, permitiendo filtrar por disciplina y sub-áreas para identificar a qué facultad podría pertenecer el investigador de la IES en cuestión. Esta información puede ser especialmente útil para la toma de decisiones en cuanto a la asignación de recursos y la elaboración de planes estratégicos de investigación en la institución (Ver ilustración 26).

Ilustración 26 - Desagregación



Fuente: Power BI - Versión: 2.115.663.0 64-bit (marzo de 2023)

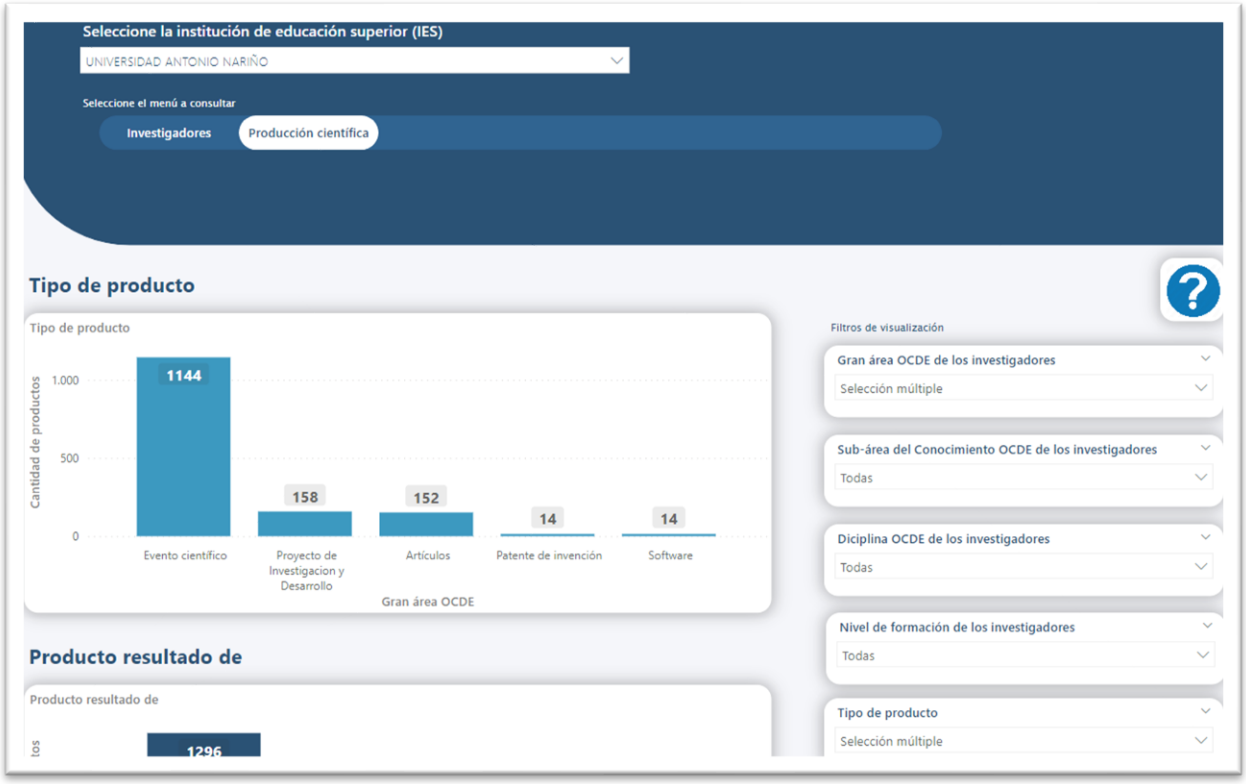
El panel de "PRODUCCIÓN CIENTÍFICA" presenta gráficos de barras con la cantidad de publicaciones por tipo de producto, permitiendo ver rápidamente cuántos artículos, libros, capítulos de libro, entre otros, han sido producidos por los investigadores. Además, se pueden visualizar tarjetas informativas con detalles adicionales sobre cada producto, como el título, los autores, el año de publicación y la editorial o revista donde se publicó (Ver ilustración 28).

También es posible hacer búsquedas específicas dentro de la sección de producción científica, por ejemplo, buscar todos los artículos publicados en el área de biología o todos los capítulos de libro escritos por investigadores de la con formación perteneciente a las Ciencias Naturales. Esto permite tener una visión detallada y personalizada de la producción científica en el campo de interés (Ver ilustración 27).

Ilustración 27 – Filtros de visualización

Fuente: Power BI – Versión: 2.115.663.0 64-bit (marzo de 2023)

Ilustración 28 – Producción científica



Fuente: Power BI – Versión: 2.115.663.0 64-bit (marzo de 2023)

En este panel, es posible acceder a una gran cantidad de información relevante acerca de la producción científica de los investigadores reconocidos por Minciencias en la Convocatoria 894 de 2021. En primer lugar, los usuarios pueden seleccionar la gran área OCDE a la que pertenecen los investigadores, lo que permite una categorización de estos y una visualización más clara de su producción científica por área de conocimiento. Además, el panel ofrece la posibilidad de filtrar los resultados por tipo de producto, lo que permite una exploración más específica de la producción de los investigadores.

Este análisis detallado de la producción científica puede resultar de gran utilidad para diversos usuarios, por ejemplo, para investigadores, quienes pueden identificar áreas de investigación que son relevantes en su campo y conocer los principales productores científicos en cada área. También puede ser útil para instituciones educativas y de investigación, quienes pueden analizar la producción científica de sus investigadores y diseñar estrategias para fomentar áreas de investigación en las que no tienen una gran presencia (Ver ilustración 29).

Ilustración 29 - Información producción científica

Fuente: Power BI - Versión: 2.115.663.0 64-bit (marzo de 2023)

6. Conclusiones

En el ámbito de la educación superior, es fundamental contar con una ventana de observación sobre la producción científica de los investigadores en las Instituciones de Educación Superior (IES) en Colombia. Para lograr esto, se ha realizado un procesamiento de las bases de datos del SCIENTI y SNIES. A partir de esta información, se pueden identificar áreas de oportunidad para fortalecer la producción científica de los investigadores en las IES. En este sentido, se han identificado diversas conclusiones que destacan la importancia de la visualización de datos y la utilización de bases de datos en la caracterización y análisis de la producción científica de los investigadores en Colombia. A continuación, se presentan estas conclusiones de manera más clara y precisa.

1. La utilización de un dashboard interactivo permite una visualización fácil y accesible de la producción científica de los investigadores en las IES de Colombia, utilizando las bases de datos CvLAC y SNIES.
2. Las bases de datos SNIES y CvLAC ofrecen información actualizada y detallada sobre la producción científica de los investigadores reconocidos por Minciencias en las IES de Colombia.
3. La caracterización de la producción científica de los investigadores avalados por Minciencias en Colombia es un paso fundamental para comprender el panorama de la investigación en las IES de Colombia.
4. La utilización de las bases de datos CvLAC y SNIES permite tener una amplia cobertura en cuanto a la información de la producción científica de los investigadores en Colombia.
5. La visualización de la producción científica de los investigadores por medio de un dashboard interactivo podría permitir una comprensión más fácil y rápida del panorama de la investigación en las IES de Colombia.
6. El dashboard podría contribuir a la identificación de áreas de oportunidad en las IES de Colombia para fortalecer la producción científica de sus investigadores.
7. La calidad de la educación superior en Colombia es esencial para el desarrollo del país, y uno de los factores considerados en el proceso de acreditación de las IES es la producción científica de los investigadores.
8. Conocer el estado actual de la producción científica en las IES puede ser útil para mejorar la calidad de la educación superior en Colombia, ya que permite identificar la producción científica de los investigadores y utilizar esa información en los procesos de acreditación de las IES.

7. Referencias

- Ministerio de Educación Nacional. (s.f.). Decreto número 1279 de 2002: Por el cual se reglamenta la Ley 30 de 1992 en lo relacionado con el funcionamiento de programas de posgrado [Archivo PDF]. Recuperado de https://www.mineducacion.gov.co/1621/articles-86434_Archivo_pdf.pdf
- Wanumen Silva, Luis Felipe, Edwin Rivas Trujillo, and Darin Jairo Mosquera Palacios. Bases De Datos En SQL Server. Primera Edición. ed. Bogotá, D.C: Ecoe Ediciones, 2018. Ingeniería Y Salud En El Trabajo. Informática. Web.
- CvLAC. (s.f.). Minciencias. Recuperado el 20 de marzo de 2023, de https://minciencias.gov.co/content/cvlac
- ¿Qué es el SNIES? (s.f.). Gov.co. Recuperado el 23 de marzo de 2023, de https://snies.mineducacion.gov.co/portal/EL-SNIES/Que-es-el-SNIES/
- (Microsoft data platform, s.f.). Microsoft data platform. (s.f.). Microsoft.com. Recuperado el 16 de marzo de 2023, de https://www.microsoft.com/es-es/sql-server/
- Visualización de datos. (s.f.). Microsoft.com. Recuperado el 10 de marzo de 2023, de https://powerbi.microsoft.com/es-es/
- (Talend). Talentopenstudio.com. Recuperado el 09 de marzo de 2023, de https://talentopenstudio.com/que-es-talent-open-studio/
- Castañeda, D. A. F., & Garcia, J. A. S. (2021). Introducción a la inteligencia de negocios basada en la metodología KIMBALL: Introduction to business intelligence based on KIMBALL Methodology. Tecnología Investigación y Academia, 9(1), 5–17. https://doi.org/10.15332/tia.2021.18082
- Shearer, C. (2000). El modelo CRISP-DM: el nuevo plan para la minería de datos. Journal of Data Warehousing, 5(4), 13-22. Recuperado el 17 de marzo de 2023, de https://doi.org/10.4018/jdw.2000100102
- IBM. (s.f.). IBM SPSS Modeler. Descripción general de la ayuda de CRISP-DM. Recuperado el 17 de marzo de 2023, de https://www.ibm.com/docs/es/spss-modeler/saas?topic=dm-crisp-help-overview
- Kimball, R., & Ross, M. (2013). El kit de herramientas de la bodega de datos: la guía definitiva para el modelado dimensional (3.a ed.). John Wiley & Sons.

- Danysoft. (s.f.). Integración de Talend Studio, SQL Server y Power BI para el análisis de datos. Recuperado el 27 de marzo de 2023, de https://www.danysoft.com/integracion-de-talend-studio-sql-server-y-power-bi-para-el-analisis-de-datos/
- Microsoft. (s.f.). Conexión a una base de datos SQL Server desde Power BI Desktop. Recuperado el 26 de marzo de 2023, de https://docs.microsoft.com/es-es/power-bi/connect-data/desktop-connect-sql-server-database