

Herramienta para identificar las inconsistencias entre los Sistemas Financiero y Académico utilizados en la Universidad la Gran Colombia

Autores

Fabio Grandas Amado
Yennifer Alejandra Vargas Zabala

Director

Elio Higinio Cables Pérez, Ph.D

Universidad Antonio Nariño
Facultad de Ingeniería de Sistemas
Especialización en Gobierno de Datos
Bogotá D.C
2022

Índice de contenidos

Introducción	1
Descripción y formulación del problema	3
Objetivo General	4
Objetivos Específicos	4
Marco referencial	5
Marco teórico	5
Estado del Arte	10
Impacto	13
Componente de Innovación	13
Metodología	14
Etapas de la metodología	14
Desarrollo de la propuesta	17
Extracción de datos	17
Procesamiento de datos	38
Visualización de datos	40
Conclusiones	48
Referencias	49

Índice de Figuras

Figura 1. Funcionamiento del patrón modelo-vista-controlador	8
Figura 2. Metodología	14
Figura 3. Modelo relacional Sistema Académico	20
Figura 4. Modelo relacional Sistema Financiero	24
Figura 5. Creación conexión fuentes de origen.	25
Figura 6. Configuración de variables.	25
Figura 7. Visualización configuración base de datos.	26
Figura 8. Listado de tablas y campos.	27
Figura 9. Tarea de inicio.	28
Figura 10. Conexión base de datos.	28
Figura 11. Ejecución de consultas.	28
Figura 12. Mapeo de datos.	29
Figura 13. Definición de registros.	30
Figura 14. Consola - log.	30
Figura 15. Cierre de conexión.	30
Figura 16. Cierre de conexión.	30
Figura 17. Configuración componente de conexión	31
Figura 18. Creación de consulta Sistema Académico	32
Figura 19. Creación de tabla Sistema Financiero	32
Figura 20. Detalle de los campos Sistema Académico	33
Figura 21. Detalle de los campos Sistema Financiero	34
Figura 22. Mapeo de datos	35
Figura 23. Aplicación de reglas	36
Figura 24. Conversión de datos	36
Figura 25. Diseño final para extracción Sistema Financiero	37
Figura 26. Diseño final para extracción Sistema Académico	38
Figura 27. Modelo relacional nuevo modelo	39
Figura 28. Diseño inicio de sesión	42
Figura 29. Mensaje de credenciales inválidas	43
Figura 30. Mensaje de permisos	44
Figura 31. Visualización de datos	45
Figura 32. Visualización de datos estadísticos en gráfico circular	46
Figura 33. Visualización de datos estadísticos en gráfico de barras	47

Índice de tablas

Tabla 1. Campos tabla bas_tercero	17
Tabla 2. Campos tabla src_uni_academica	18
Tabla 3. Campos tabla src_pensum	18
Tabla 4. Campos tabla src_formulario	19
Tabla 5. Campos tabla persona	20
Tabla 6. Campos tabla cliente	21
Tabla 7. Campos tabla documento	21
Tabla 8. Campos tabla orden	22
Tabla 9. Campos tabla cct_ordenes_academico	23

Resumen

Con el presente trabajo de grado se pudo construir una herramienta que logra identificar las inconsistencias presentadas entre los Sistemas Académico y Financiero de la Universidad la Gran Colombia, con relación al pago de los recibos de inscripción para los aspirantes que desean ingresar a la institución. De acuerdo con las actividades realizadas, se logró obtener los datos de los formularios de inscripción y recibos generados entre los años 2021 a 2022 de los Sistemas de Información. Se utilizó una ETL que significa (extracción, transformación y carga) de los datos, extrayendo desde las fuentes de origen, realizando el correspondiente tratamiento a los datos y almacenándolos en una fuente de destino con un nuevo modelo. A partir del desarrollo de la herramienta se pretende implementarla para identificar los casos que presentan inconsistencias y así solventarlos de manera oportuna y ágil para evitar molestias o quejas durante el proceso. Además, esta aplicación permite ser adaptada e implementada en otros procesos o áreas de la institución y que a su vez permita mejorar el análisis de los datos y visualización de la información.

Palabras clave: Bases de datos, extracción de datos, modelo de datos, visualización de datos.

Abstract

With this degree work, it was possible to build a tool that manages to identify the inconsistencies presented between the Academic and Financial Systems of the Universidad la Gran Colombia, in relation to the payment of registration receipts for applicants who wish to enter the institution. According to the activities carried out, it will be possible to obtain the data from the registration forms and receipts generated between the years 2021 to 2022 from the Information Systems. The data extraction, transformation and loading was performed using an ETL, extracting this from the source sources, performing the corresponding data treatment and storing them in a destination source with a new model. From the development of the tool, it is intended to implement it to identify cases that present inconsistencies and thus solve them in a timely and agile manner to avoid inconveniences or complaints during the process. In addition, this application allows it to be adapted and implemented in other processes or areas of the institution and which in turn allows improving data analysis and information visualization.

Key words: Database, data extraction, data model, data visualization.

1. Introducción

Hoy en día, con el creciente aumento en el volumen de los datos las compañías han buscado maneras para explotarlos y generar valor a través de ellos. Por lo cual, se han utilizado diferentes técnicas y tecnologías que permitan efectuar un análisis e interpretar estos datos, tanto estructurados como no estructurados, de esta manera las empresas pueden emplearlos como base para tomar decisiones (Hitachi Vantara, 2021).

Existen diferentes formas para procesar los datos, sin embargo, es importante que el método a utilizar no genere problemas de rendimiento o disponibilidad sobre las fuentes de datos. Por tanto, muchas organizaciones optan por implementar nuevas arquitecturas para su almacenamiento. Lo cual, consiste en tomar los datos de la fuente original y cargarlos en una fuente de destino, generando un nuevo modelo de datos y de esta manera manipularlos de acuerdo con las necesidades (Hitachi Vantara, 2021).

Con la necesidad de aprovechar al máximo los datos, se han creado herramientas tecnológicas que permiten agilizar y automatizar diversas tareas, de esta manera se genera la visualización de estos, de acuerdo con las necesidades planteadas. Además, los datos pueden ser utilizados y elegidos para los diferentes propósitos.

Esta visualización se logra mediante la implementación de un tablero de datos y así, al realizar la gestión de la información facilitan el rastreo y cumplimiento de indicadores, la creación y seguimiento de métricas y otros puntos que pueden ser clave y relevantes en un negocio. Simplificando la visualización de los conjuntos de datos para proporcionar a los usuarios una vista del progreso actual de dicho proceso o problemática y realizar un análisis sobre los datos de forma más eficiente. Además, se puede realizar un diseño atractivo a estos tableros, de tal forma que las visualizaciones sean con cifras significativas y utilizables para cada oficina de la organización, todos los roles que intervienen tienen preguntas comerciales que deben responderse en una parte en particular del proceso. Identificando qué para la implementación de un buen tablero de control es importante tener en cuenta las siguientes características (Kaplan & Norton, 2001)

- Facilitar la modificación en su visualización, de forma que se puedan crear diferentes presentaciones para revisar las prioridades propias o las de la compañía.
- Tener un acceso útil y fácil desde cualquier dispositivo.

- Ser escalable para tomar una multitud de fuentes de datos y hacer crecer al negocio.

Al examinar el mercado en la actualidad, se han creado diversas herramientas para el análisis de los datos y tableros implementados para facilitar la gestión en las empresas, así como tecnologías que permiten la creación de aplicaciones con estas funcionalidades. Por tanto, es importante que la herramienta que se utilice para este propósito tenga un control integral y sea aplicable en las diferentes áreas de la compañía, de esta manera no se generan varias herramientas y los datos queden dispersos. Finalmente, estas herramientas deben estar pensadas no solo para la presentación de los datos, sino también para el soporte en la toma de decisiones en todos los niveles de la empresa.

Lo expresado anteriormente se utilizará en este proyecto para abordar una problemática de integración entre dos bases de datos, lo cual atenta al buen desempeño de la organización. Por tanto, se propone solucionar esta problemática, mediante la extracción de los datos, los cuales contarán con unas reglas de negocio, para que estos sean cargados a un nuevo modelo y así permitir la visualización de la información para el análisis, consumo, así como la toma de decisiones por parte de los usuarios finales.

2. Descripción y formulación del problema

La Universidad la Gran Colombia cuenta con un procedimiento para la inscripción, selección, admisión y matrícula para la vinculación de aspirantes, estudiantes de reingreso, transferencia externa, transferencia interna o egresados a los programas académicos de pregrado y postgrado para la sede Bogotá y la seccional de Armenia. Para este proceso, los interesados en ingresar a la institución deben inscribirse por medio de un formulario que se encuentra habilitado en el portal web de la Universidad, donde diligencian sus datos personales, posteriormente se genera el correspondiente recibo de pago de inscripción para continuar con la admisión. Esta información es almacenada en el Sistema de Gestión Académico.

Una vez que el aspirante, estudiante o egresado realiza su pago, este se guarda en el Sistema Financiero y a través de una integración con el Sistema Académico se valida el estado del pago para que el usuario interesado pueda continuar con la admisión. Sin embargo, se han encontrado inconsistencias en este paso, debido a que los pagos se liquidan en el Sistema Financiero, pero no se ven reflejados en el Sistema Académico.

Los motivos por los que esto sucede son los siguientes: los pagos en línea, ya que a partir del ingreso del pago debería liquidarse automáticamente y visualizarse en ambos Sistemas, pero esto no sucede, ya que el recibo presenta un número de cliente diferente al número de identificación del usuario, lo que provoca esta discrepancia, también se puede presentar que el recibo sea generado manualmente en el Sistema Financiero, sin haber registrado el formulario de inscripción, por tanto el pago nunca se va a visibilizar hasta que se complete todo el proceso en el Sistema Académico. Por último, el aspirante se puede registrar en varios tipos de admisión, por tanto, se crean diferentes recibos, liquidando el que no corresponde para su proceso, por lo cual, esta liquidación debe ser cruzada manualmente contra el recibo que corresponde.

Esto impide a los aspirantes continuar el siguiente paso, causando molestias y quejas, lo que también puede ocasionar que desistan del proceso y se marchen a otras instituciones, esto también genera que el tiempo de admisión sea más largo, cuando debe efectuarse en tiempos más reducidos. Finalmente, se produce que la calidad de los datos no sea óptima por presentar una falta de confiabilidad de los usuarios hacia estos sistemas, que se originan por no tener consistencia, exactitud e integridad.

Por tanto, la formulación del problema identificado es: ¿Cómo identificar los aspirantes que cuentan con el pago de su recibo en el Sistema Financiero, pero no en el Sistema académico y por tanto no pueden continuar con el proceso de su admisión?

2.1 Objetivo General

Elaborar una herramienta para la identificación de los aspirantes que están en proceso de admisión y registraron su pago en el Sistema Financiero con respecto al estado indicado en el Sistema Académico.

2.2 Objetivos Específicos

- Caracterizar las herramientas TIC que permitan gestionar el proceso de admisión del aspirante y el registro de los pagos efectuados.
- Elaborar un nuevo modelo de datos que soporte la información de las bases de datos de origen y permita hacer el estudio de estos datos.
- Implementar un dashboard que permita recuperar la información de los aspirantes que presentan la inconsistencia de pago entre el Sistema Financiero y el Sistema Académico.

3. Marco referencial

3.1. Marco teórico

Los sistemas de información donde se gestiona el proceso de admisión de aspirantes y generación de los recibos se encuentran en bases de datos Oracle 12C en la nube, lo que implica el uso de principios de calidad que garanticen que el proceso se ejecute de forma adecuada y esté encaminado a la solución de la problemática, por lo cual también es necesario utilizar un conjunto de herramientas TIC, para su extracción, tratamiento y almacenamiento, los cuales se describen a continuación.

Como punto de partida se establecen las normas y procesos que se aplicarán durante el desarrollo de la investigación, de acuerdo con las necesidades de la compañía y la implementación de buenas prácticas, las cuales se detallan en el siguiente apartado.

Gestión de la calidad de los procesos

Para (García P., M., Quispe A., C., & Ráez G., L., 2003), existen ocho principios en la gestión de la calidad, los cuales fueron planteados para que las directivas de una organización puedan implementarlos estableciendo el mejoramiento continuo en la compañía.

El primer principio es tener como enfoque al cliente, debido a que toda organización siempre depende de sus clientes y es por eso se debe estar atentos a sus requerimientos para las necesidades existentes y futuras y así satisfacer sus exigencias.

Después, se encuentra el liderazgo, ya que los líderes constituyen el propósito y la orientación de la organización, en donde se establece un ambiente interno, de forma tal que el personal logre involucrarse en el logro de los objetivos de la empresa.

Personal: la participación de cada uno los colaboradores, compromiso y habilidades en cada uno de sus niveles es la esencia de la organización, los cuales son usados para el beneficio de esta.

Enfoque a los procesos: el logro de resultados eficientes es alcanzado cuando las actividades y los recursos están definidos como procesos y se gestionan de esta forma.

Sistema enfocado hacia la gestión: Definir, determinar y gestionar los procesos conectados como un sistema, estableciendo la efectividad y eficacia de las empresas en el cumplimiento de sus logros (ISO 9001,2015).

Mejoramiento continuo: la aplicación de la mejora continua debe ser un objetivo permanente para el desempeño de la organización (ISO 9000, 2005).

Toma de decisiones: Estas decisiones ejecutadas de forma eficaz se realizan con base en el análisis de los datos y la información.

Relación recíproca y productiva con proveedores: La compañía y sus proveedores dependen recíprocamente del otro, siendo una relación mutuamente beneficiosa para ambos, permitiendo dar valor a ambas partes.

Una vez definidas las normas, se seleccionan las herramientas TIC a utilizar para cumplir los objetivos de este trabajo, cuales se detallan a continuación.

Talend Open Studio

Esta aplicación permite la integración de los datos al combinar diferentes fuentes en una sola vista para obtener datos significativos que puedan ayudar a diferentes procesos para obtener datos limpios, hacer transformaciones y preparar datos de calidad, consistencia, centralización, reutilización de componentes, lo que facilita mantener y crear procesos como análisis avanzado y toma de decisiones en aplicaciones de inteligencia artificial a los procesos nuevos o existentes.

Esta herramienta cuenta con una versión gratuita la cual tiene una gran cantidad de componentes que permiten hacer varios procesos, permitiendo de forma visual modelar transformaciones mediante un compilado del lenguaje de programación Java.

Después, se tiene en cuenta una serie de lenguajes que se aplicarán en el transcurso de la etapa de desarrollo, por lo cual se describe en este apartado los lenguajes de programación a utilizar para la creación de la herramienta.

Lenguaje PHP

Es un lenguaje de programación del lado del servidor, es decir, se utiliza para la comunicación y uso de consultas hacia la base de datos, su acrónimo significa Hypertext Preprocessor, este lenguaje de programación ejecuta código y permite el enmarcado de código HTML para crear aplicaciones y servicios, es de código abierto por lo que no hay restricciones de derechos y esta es constante actualización y mantenimiento (Recio, 2016).

HTML

Es un lenguaje de etiquetas y tags para describir páginas web, estructura el contenido con diferentes elementos que componen la página con bloques de marcado y texto, scripts y rutas de archivos y contenido multimedia, este es interpretado por el navegador e interactúan con JavaScript con el modelo de objeto de documento (DOM) mostrando al usuario final una vista grafica del texto cargado. Hay varias especificaciones de este lenguaje que están en uso actualmente y se encuentra en constante desarrollo la versión HTML5, el organismo encargado de esta estandarización es la W3C (Celaya, 2014).

JavaScript

JavaScript es un lenguaje interpretado de programación del lado del cliente, este ejecuta secuencias de comandos que permiten la creación e implementación de funciones en páginas web, por lo cual, cuando una página genera un evento este ejecuta acciones que hacen que la navegación sea más intuitiva y cómoda para el usuario, este lenguaje permite manipular el modelo de objeto de documento (DOM) del navegador y así mostrar información de forma dinámica, generando oportunas actualizaciones en el contenido, mapas interactivos, animaciones de Gráficos 2D y 3D, es compacto y muy flexible y se han escrito gran cantidad de librerías como JQuery (Recio, 2016).

CSS3

Es un lenguaje que permite elaborar páginas web atractivas, definiendo el diseño gráfico mediante el uso de un documento estructurado u hojas de estilo, este lenguaje

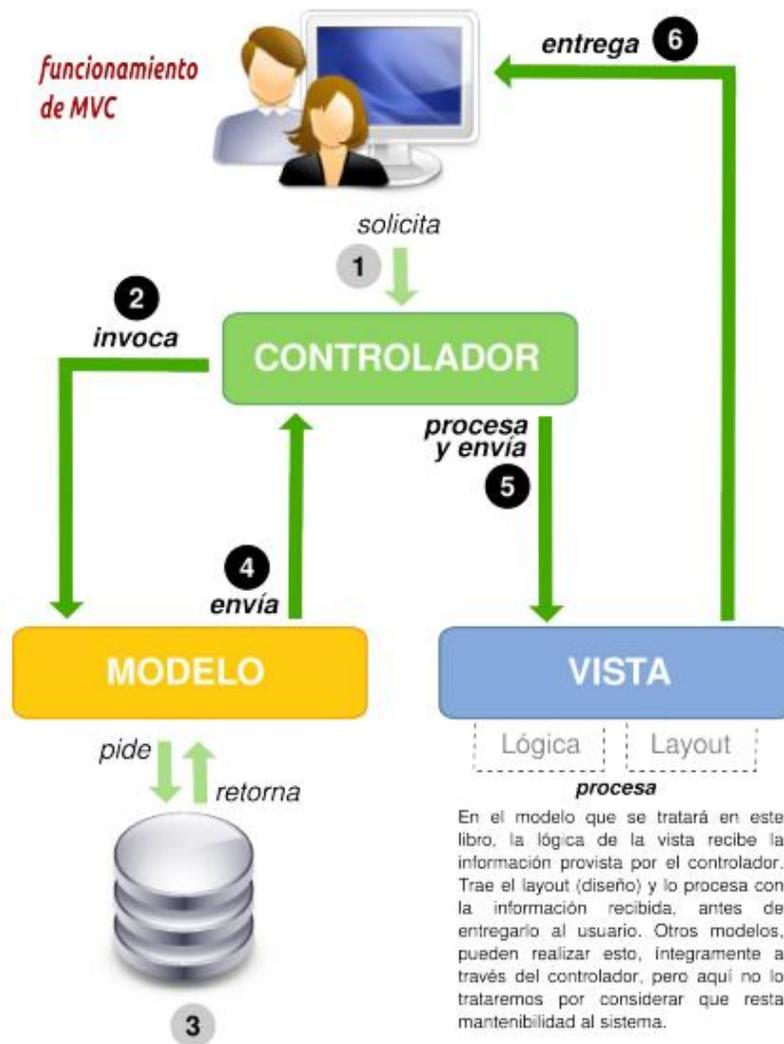
de marcado es usado para crear el diseño visual en los sitios web y las interfaces del usuario (Recio, 2016).

Para la programación de la aplicación se utiliza el siguiente patrón de arquitectura, el cual permite brindar un correcto funcionamiento.

Modelo Vista Controlador (MVC)

Es un patrón de arquitectura de software que permite desglosar el código para la aplicación, la interfaz de usuario, y la lógica de control en tres componentes diferentes, sus responsabilidades y las relaciones que existen entre ellos, se usa para desarrollar interfaces de usuario modernas, este modelo ha demostrado ser útil y apropiado durante varios años en diferentes aplicaciones, sobre diversos lenguajes y plataformas para desarrollo, cada componente se describe a continuación (Bahit, 2014), en la siguiente figura se describe el funcionamiento del modelo (Ver Figura 1).

Figura 1. Funcionamiento del patrón modelo-vista-controlador



Fuente: Extraído del libro POO y MVC en PHP.

Modelo: Comprende la representación de los datos que maneja el sistema, encargado del acceso de forma directa entre la base de datos y la aplicación.

Vista: Se encarga de generar los componentes de lenguaje de marcado que se enviarán al cliente o interfaz de acuerdo con la adaptación de la tecnología del usuario.

Controlador: Es el encargado de gestionar el flujo que se genera entre el modelo y la vista, realizar las transformaciones necesarias según las necesidades de la solicitud, es el intermediario entre el sistema y el cliente.

Para la implementación de la arquitectura descrita anteriormente, se utiliza el componente Smarty, es un motor de plantillas para el lenguaje PHP, por el cual se puede separar la parte lógica de la presentación. Esto permite tener un objeto escalable y manipulable, de acuerdo con las necesidades futuras. También favorece la depuración de código en la aplicación y de realizar el mantenimiento.

Finalmente, se tienen los siguientes marcos de trabajo, los cuales se establecen para la implementación de la herramienta, mediante los lenguajes de programación descritos anteriormente.

jQuery

Es una librería de JavaScript que simplifica, facilita la manipulación del Modelo Objeto de Documento (DOM) de HTML, facilitando grandes resultados en el desarrollo web en pocas líneas de código, permite la manipulación, modificación y creación de eventos con mayor facilidad con algoritmos de fusión y estructuras de datos (Recio, 2016).

Bootstrap

Es un framework CSS y JavaScript de código abierto que proporciona beneficios en el desarrollo web de forma ágil y sencilla, elaborado con el objetivo de ofrecer la creación de interfaces limpias y con un diseño adaptable, permitiendo la modificación de formularios ajustando elementos como: botones, tablas, menús desplegados, tipografías y mucho más. Además, brinda herramientas y funciones para que los usuarios puedan crear cualquier sitio web (Gallego, 2017).

3.2. Estado del Arte

Desarrollar una solución que permita construir un modelo para la extracción, análisis y visualización de la información entre dos sistemas de información implica conocer otras técnicas utilizadas para este fin, por tanto, a continuación, se describen los aportes más relevantes de autores sobre estos temas, ya que es fundamental contar con una base que permita abordar adecuadamente este tema:

Según la investigación de, “Desarrollo de un data warehouse para el proceso de denuncias de la defensoría del consumidor” (Hernández, O., Mejía, Y., Silva, H., Cuchillas, R., 2010), los autores tratan la necesidad de contar con un único acceso donde pueden generar reportes para el análisis y mejora en la toma de decisiones, ya que la información se encuentra consignada en dos sistemas de información diferentes. SARA, el cual se encarga del funcionamiento básico de la Defensoría del consumidor automatizando los procesos que llevan manualmente y ESPACIO, el cual brinda el apoyo al Sistema Nacional de Protección al Consumidor, permitiendo a las instituciones obtener el registro y seguimiento de denuncia de los consumidores en las áreas que les compete.

Por lo cual, debido a que no se encuentran integrados los datos, la solución a esta necesidad fue la creación de un data warehouse, para lo cual fue necesario realizar la caracterización, análisis, diseño y documentación del proceso. Para comprender el alcance de este proyecto se implementó la siguiente metodología:

- Se establecieron las herramientas para la recolección de los datos como: la entrevista, el cuestionario, la observación y apoyo de referencias bibliográficas.
- Se efectuó el diagnóstico y planteamiento del problema mediante un diagrama causa – efecto y el método de la caja negra.
- Se generó el modelo UML de los componentes a utilizar para la implementación de la Bodega de Datos.
- Posteriormente se definieron las fases para la implementación de la investigación:
 - Recolección y refinamiento de los requerimientos.
 - Diseño y construcción del prototipo.
 - Evaluación del prototipo por el cliente, prueba y revisión.
 - Refinamiento del prototipo.
 - Finalmente, se genera el plan de implantación y documentación del producto final.

Por otra parte, teniendo en cuenta la misma temática de problemas de integración de datos expuesta anteriormente, en el año 2016 la empresa OneLink BPO tuvo el

requerimiento de implementar una solución para el análisis de información y reportes. Sin embargo, para este fin la información se encontraba consignada en dos fuentes de información: Salesforce y una base de datos creciente en SQL (Lenguaje de consulta estructurada).

La metodología implementada para este proyecto fue la siguiente:

- Detallar los antecedentes del proceso.
- Definir el alcance de los reportes a implementar en el proyecto.

El resultado final de esta investigación proporcionó la implementación de un sistema de reportes en Power BI, el cual permitió reducir la cantidad de reportes al poder unificar la información, además de ejecutarse automáticamente, ya que antes tenían que realizar modificaciones manuales en los archivos reduciendo cerca del 90% de estos procesos manuales, generando que la probabilidad de error sea mínima y la integración y la calidad sea mejor y más alta. Adicionalmente, permitió a la organización comenzar la incorporación de otros procesos con esta solución para otros clientes e incluso para otros reportes corporativos (Sotaquira, 2017).

Complementando la investigación anterior, se encuentra el trabajo de grado “Diseño e implementación de prototipo BI utilizando una herramienta de Big Data para empresas Pymes distribuidoras de tecnología” (Acosta, J. & Flórez, D., 2015), debido a que, muchas de las empresas de Pymes tienen inconvenientes en la implementación de tecnologías de la información y por ser empresas pequeñas no quiere decir que no requieran organizar y gestionar sus datos para generar conocimiento, tener un control y tomar las decisiones correspondientes de su negocio. Para este documento se seleccionó la empresa SED Colombia, en la cual se identificó que manejan dos aplicaciones, un ERP para el manejo de actividades primarias y de apoyo y el sistema de reportes Cristal Reports, utilizando el apoyo de aplicaciones ofimáticas, bases de datos y consultas exportadas a Excel.

Esto generaba que la información se viera afectada, debido a que los archivos son planos y la información no estaba consolidada completamente, además de no contar

con filtros de búsqueda y también es muy lento el tiempo de análisis, dificultando el cruce de información con otros archivos recurriendo a usar Excel.

Por tanto, la metodología aplicada para esta investigación fue la siguiente:

- Definir el tipo de estudio, realizando análisis a la situación actual de la organización en cuanto a estrategias de BI y Big Data.
- Después, se definen las herramientas que pueden ser aplicadas para este proceso.
- Se genera el diseño del prototipo y finalmente su implementación en la empresa.

En conclusión, el diseño e implementación de este prototipo la compañía está en capacidad de saber en todo momento el estado de sus ventas, analizar la rentabilidad de sus productos y marcas. Además, de acuerdo con la necesidad que se presente en el momento podrán tener la posibilidad de visualizar como se encuentra la empresa y así establecer indicadores que suministren alertas en el rendimiento de cada proceso que la compañía quiera aplicar, con el propósito de crecer y operar adecuadamente en el mercado (Acosta, J. & Flórez, D., 2015).

3.3. Impacto

Para este proyecto se busca implementar una herramienta que permitirá identificar el estado de pago de los recibos de inscripción generados para los aspirantes nuevos o reintegros que se encuentran en proceso de admisión a la Universidad. De esta manera se puede agilizar el tiempo de admisión del aspirante, reduciendo las posibles quejas y deserciones del proceso. Además, con esta herramienta se promueve la disponibilidad y accesibilidad de los datos garantizando la integridad, seguridad, usabilidad y calidad de estos.

3.4. Componente de Innovación

Si bien existen en el mercado herramientas que permiten la visualización de la información, con el desarrollo de este proyecto se entregará una plataforma tecnológica propia, la cual además de contener lo propuesto para esta investigación, también servirá

como insumo para continuar implementando más proyectos y mejorar a los procesos en la institución.

4. Metodología

De acuerdo con el planteamiento del problema y los objetivos del proyecto, la metodología aplicada para se propone a partir de una serie de pasos, los cuales deben ejecutarse para llegar a la respectiva solución. Esta metodología permite abordar la implementación de la herramienta para recuperar la información requerida (Ver Figura 2).

Figura 2. Metodología



Fuente: Elaboración propia.

4.1. Etapas de la metodología

Para comprender mejor el desarrollo del trabajo en cada una de sus etapas, se describe de forma general con base a la bibliografía consultada, el planteamiento del

problema y los objetivos del documento. Se detalla cada una de las tareas elaboradas con el propósito de presentar los resultados obtenidos en la investigación.

Extracción de datos

Es el proceso por el cual se ajustan los datos desde las diferentes fuentes, donde se enriquecen los datos de forma tal que se genere información valiosa. Los aspectos para considerar son:

- Identificar los datos de las fuentes relacionadas.
- Extraer los datos eficazmente desde las diferentes fuentes.
- Brindar un mecanismo para la transferencia de los datos.
- Enriquecer los datos por medio de las fuentes, para una sencilla actualización y cambio que sea aplicable. (Mohanty, 2004).

Para esta etapa se utiliza una herramienta ETL, la cual permite extraer, transformar y cargar los datos desde las fuentes de datos de origen hacia la fuente de destino, con el propósito de manipular varios datos obteniendo un formato adecuado. Esta aplicación debe lograr los resultados deseados en un tiempo menor a métodos tradicionales. Algunos aspectos por considerar para la selección de la herramienta son los siguientes (Banerjee, 2004):

- Fácil de usar desde el punto de vista del mantenimiento.
- La ETL debe integrarse con el proceso de negocio.
- Soportar el procesamiento de grandes volúmenes de datos.
- Extraer datos de diversas fuentes heterogéneas.
- Debe soportar el procesamiento en paralelo.
- Tener la posibilidad de estandarizar los datos desde las diferentes fuentes.

Una vez extraídos los datos se trabajan en la siguiente etapa, correspondiente al procesamiento.

Procesamiento de datos

Este es un paso primordial, donde se normalizan y seleccionan los datos, si este paso no se realiza puede existir información errónea, irrelevante o inconsistente, generando resultados menos precisos.

Se analizan los atributos seleccionados durante la extracción, unificando los campos, cambiando los tipos de datos y asignando una etiqueta a cada atributo donde haya ocurrido una falla. Finalmente, la base de datos se normaliza, mediante la creación de un nuevo modelo de datos, de esta manera se disminuye el costo computacional, mejora el desempeño y la precisión del método de clasificación y visualización de los datos (González, W., Mora, J., Pérez, S., 2014).

Visualización de datos

Se realiza la representación gráfica de la información, así como la presentación de listados detallados. Lo cual se realiza con el propósito de interpretar y construir un significado mediante el análisis de los datos, descubriendo y comprendiendo la lógica detrás de los conjuntos de datos. Este diseño se realiza de acuerdo con unos principios formales que rigen la visualización de los datos.

Existen herramientas tradicionales para la visualización de los datos que poseen dos componentes: el de datos y el gráfico, los cuales utilizan una arquitectura en memoria para permitir a los usuarios interactuar con los datos obteniendo beneficios como:

- Análisis sofisticado y tipos de visualización como mapas de calor, histogramas y tipos de gráficos.
- Adherencia a mejores prácticas de visualización.
- Descubrimiento visual interactivo.

Esto posibilita a los usuarios a identificar de forma ágil los cambios que tienen los datos en el tiempo, consiguiendo información de elementos que están relacionados y entendiendo relaciones de causa y efecto antes de que ocurra un impacto (DAMA-DMBOK, 2017).

5. Desarrollo de la propuesta

Para el desarrollo de la propuesta, se siguieron los pasos descritos en el apartado de la metodología, la primera etapa que se realizó fue la siguiente:

5.1. Extracción de datos

Como primer aspecto para tener en cuenta en esta investigación será la identificación de las fuentes a utilizar, por lo cual será mediante las bases de datos del Sistema Académico y el Sistema Financiero, a su vez realizando una descripción de uso de cada una con las tablas que se intervendrán.

Sistema Académico: Esta aplicación fue implementada durante el año 2020 en una base de datos en Oracle versión 12C, cuenta con toda la información académica de los estudiantes y egresados de la institución en relación con procesos como: admisión de aspirante, registro de matrícula, historial académico, ceremonia de grado, entre otros.

Las tablas utilizadas en esta fuente son las siguientes:

Bas_tercero: Contiene los datos personales de los estudiantes y egresados. En el proceso de admisión, esta tabla es empleada para almacenar los datos de los aspirantes, los cuales son diligenciados durante el primer paso de su inscripción. Tiene un total de 112 campos, sin embargo, se utilizarán los que se describen a continuación (Ver Tabla 1).

Tabla 1. Campos tabla bas_tercero

Campo	Descripción
Id_tercero	Id único de cada persona
Tip_identificacion	Tipo de documento de identificación
Num_identificacion	Número de documento de identificación
Nom_tercero	Primer nombre

Seg_nombre	Segundo nombre
Pri_apellido	Primer apellido
Seg_apellido	Segundo apellido

Fuente: Elaboración propia.

Src_uni_academica: Esta destinada para almacenar la información de los programas académicos existentes en la Universidad. Tiene un total de 118 campos, sin embargo, se utilizarán los siguientes (Ver Tabla 2).

Tabla 2. Campos tabla src_uni_academica

Campo	Descripción
Cod_unidad	Código único por el cual se identifica cada programa académico
Nom_unidad	Nombre del programa académico

Fuente: Elaboración propia.

Src_pensum: Almacena la información de los planes de estudio que se encuentran asociados a los programas académicos. Tiene 95 campos, sin embargo, se utilizarán los que se describen a continuación (Ver Tabla 3).

Tabla 3. Campos tabla src_pensum

Campo	Descripción
Cod_pensum	Código único por el cual se identifica cada plan de estudios relacionado al programa académico
Nom_pensum	Nombre del plan de estudios

Fuente: Elaboración propia.

Src_formulario: Se almacenan los datos del proceso de admisión del aspirante. Tiene 61 campos, pero se utilizarán los más relevantes, los cuales son (Ver Tabla 4).

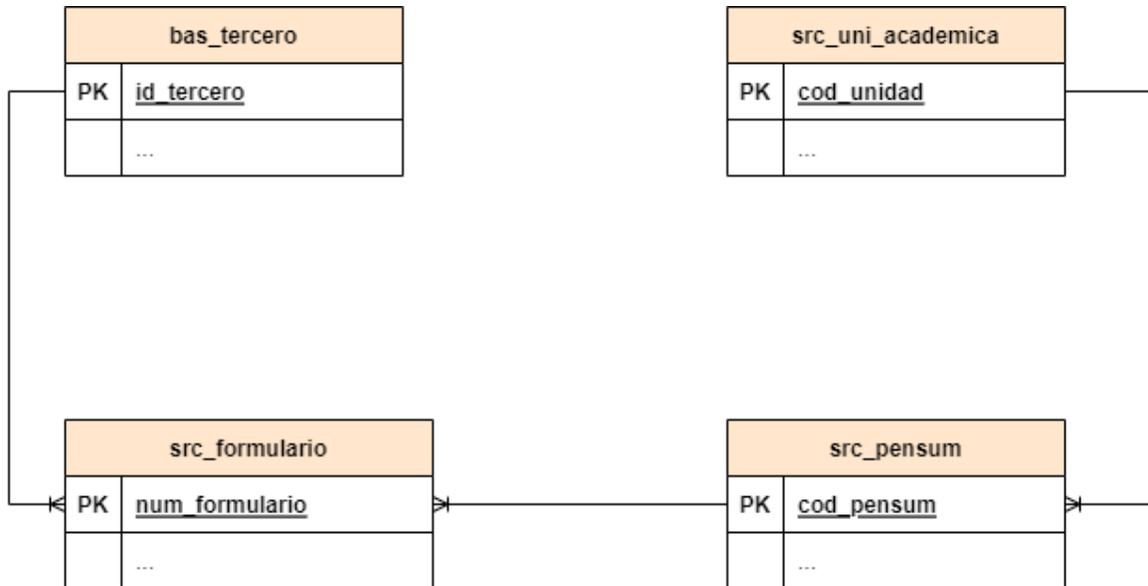
Tabla 4. Campos tabla src_formulario

Campo	Descripción
Id_tercero	Id único de cada persona
num_formulario	Número único del formulario donde se registra el aspirante
cod_modalidad	Modalidad de registro: Diurno, nocturno, virtual
num_identificacion	Número de documento de identificación
pri_apellido	Primer apellido
seg_apellido	Segundo apellido
num_recibo	Número del recibo de pago, el cual es generado después de registrar el formulario de inscripción
cod_unidad	Código único por el cual se identifica cada programa académico
est_formulario	Estado en que se encuentra el aspirante en su proceso de admisión
val_formulario	Valor del recibo de inscripción
obs_liquidacion	Observaciones registradas, después de que se liquida el pago del recibo
fec_pago	Fecha del registro de pago del recibo de inscripción
num_pago	Número del pago
cod_periodo	Periodo de inscripción del aspirante

Fuente: Elaboración propia.

A continuación, se presenta el modelo relacional de las tablas detalladas anteriormente del Sistema Académico (Ver Figura 3).

Figura 3. Modelo relacional Sistema Académico



Fuente: Elaboración propia.

Sistema financiero: Este sistema es utilizado desde el año 2014 y su última versión implementada en la nube se encuentra desde el año 2020 en una base de datos en Oracle versión 12C, está fuente cuenta con la información financiera de la institución con relación a procesos como: generación y pago de recibos, compras, inventario, entre otros.

Las tablas que se intervendrán son las siguientes:

Persona: Almacena los datos personales de los usuarios. Tiene un total de 41 campos, sin embargo, se emplearán 8 campos, los cuales se indican en la siguiente tabla (Ver Tabla 5).

Tabla 5. Campos tabla persona

Campo	Descripción
Secuencia_persona	Código único de la secuencia de la persona.

Tipo_identificacion	Tipo de documento de identificación
Identificación	Número de documento de identificación
Tipo_persona	Persona [N]atural/[J]uridica.
Nombre_razon_social	Primer y segundo nombre
Primer_apellido	Primer apellido
Segundo_apellido	Segundo apellido
Estado	Activo o Inactivo

Fuente: Elaboración propia.

Cliente: Cada usuario que realiza el pago de recibos como: inscripción, matrícula u otros conceptos, cuenta con un registro único, identificándose como cliente de la institución para conocer sus transacciones. Contiene 30 campos, pero solo se utilizarán los siguientes (Ver Tabla 6).

Tabla 6. Campos tabla cliente

Campo	Descripción
Cliente	Código único por el que se identifica cada persona, para la generación y pago de recibos
Codigo_anterior	Código anterior del cliente.

Fuente: Elaboración propia.

Documento: Almacena los tipos de órdenes que se manejan para la generación de los recibos. La tabla tiene 6 campos, sin embargo, solo se utilizarán 2 campos (Ver Tabla 7).

Tabla 7. Campos tabla documento

Campo	Descripción
-------	-------------

Documento	Código único del documento
Nombre_documento	Nombre del documento

Fuente: Elaboración propia.

Orden: Se encuentra la información del recibo que se genera para cada aspirante durante su inscripción. Tiene 44 campos, sin embargo, se emplearán 10 campos (Ver Tabla 8).

Tabla 8. Campos tabla orden

Campo	Descripción
Documento	Código único del documento
Orden	Número único de la orden de pago
Fecha_orden	Fecha en que se genera la orden
Estado	Estado de la orden
Cliente	Número del cliente a quien se asocia la orden
Fecha_vencimiento	Fecha máxima de pago del recibo
Valor_bruto	Valor del recibo
Valor_descuento	Valor con descuento
Valor_iva	Valor total + iva
Valor_total	Valor completo para pagar del recibo

Fuente: Elaboración propia.

cct_ordenes_academico: Está tabla se utiliza para almacenar los datos de la integración entre los recibos creados para la inscripción en el Sistema Financiero y los datos de la inscripción del Sistema Académico. En la tabla se tienen 41 campos de los cuales solo se utilizarán 10 como se detalla a continuación (Ver Tabla 9).

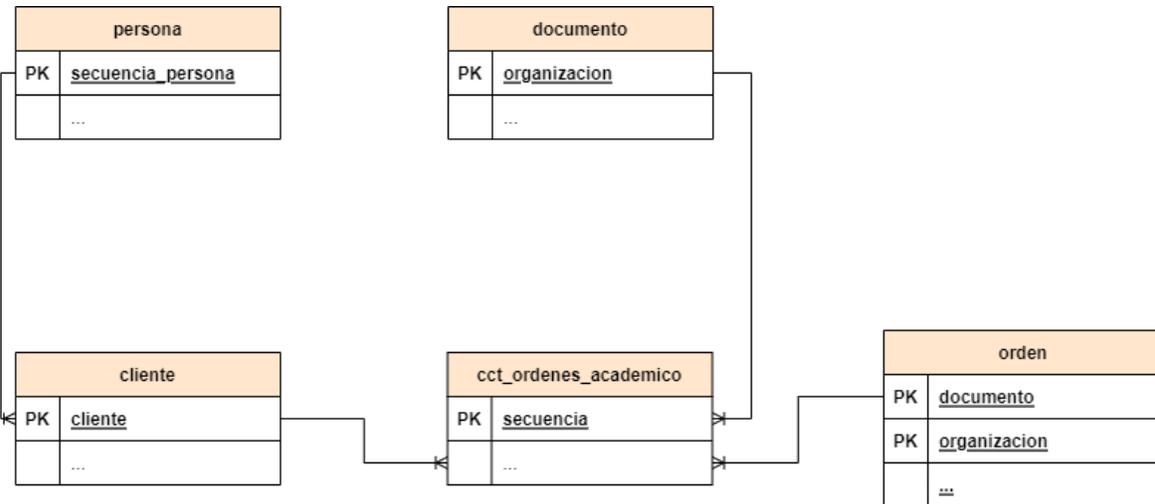
Tabla 9. Campos tabla cct_ordenes_academico

Campo	Descripción
Cliente	Número del cliente a quien se asocia la orden
Periodo	Periodo en que se genera la orden
Fecha_registro	Fecha de registro en la integración
Fecha_pago	Fecha de pago del recibo
Estado	Estado de pago del recibo
Programa_academico	Programa en que se inscribió el aspirante
Orden	Número único de la orden de pago
Valor_documento	Valor del recibo
Recibo_consignacion	Número del recibo
Identificación_aspirante	Número de identificación del aspirante

Fuente: Elaboración propia.

El modelo relacional de esta fuente de información y las tablas descritas para el Sistema Financiero son las siguientes (Ver Figura 4).

Figura 4. Modelo relacional Sistema Financiero



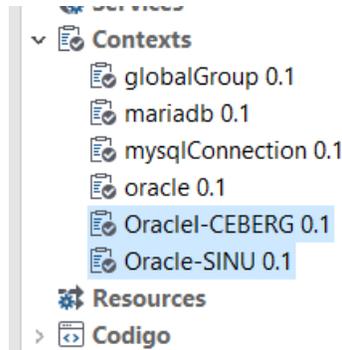
Fuente: Elaboración propia.

Una vez identificadas y detalladas las fuentes, se procede con la extracción de los datos, para este paso se seleccionó la siguiente herramienta:

Talend Open Source

Se utiliza la aplicación en la versión 7.3 como herramienta ETL en las siguientes dos etapas de este proyecto. Se procede a realizar la conexión para cada base de datos descrita anteriormente, mediante la configuración correspondiente. Estas conexiones se crean en una sección denominada contextos, por esa razón, se crea un nuevo contexto, asignando las respectivas variables y valores para configurar la conexión a las bases de datos de origen Oracle (Ver Figura 5).

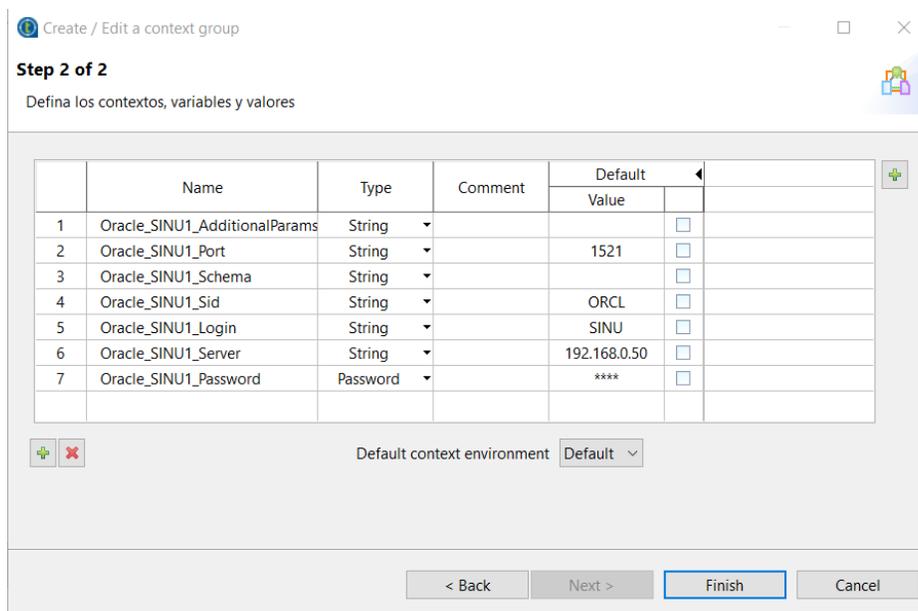
Figura 5. Creación conexión fuentes de origen.



Fuente: Extraído de la herramienta Talend Open Source.

Las variables y valores que se requieren para las conexiones son los siguientes: el nombre o IP del host, el puerto, el usuario y contraseña de ingreso y el nombre del servicio o SID (Ver Figura 6).

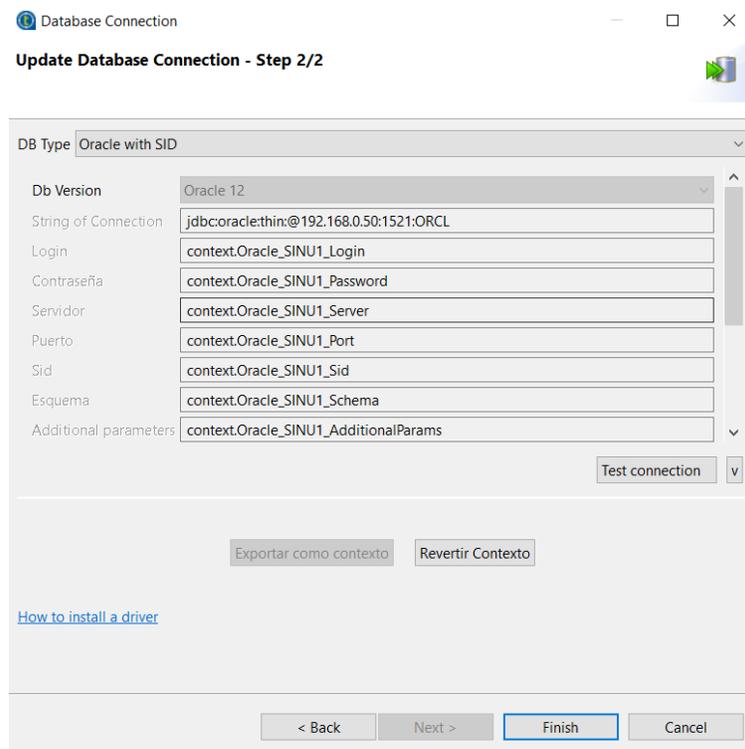
Figura 6. Configuración de variables.



Fuente: Extraído de la herramienta Talend Open Source.

Una vez digitados los datos para el contexto se verifica la configuración, de acuerdo con la versión de Oracle seleccionada, a continuación, se pueden visualizar estos componentes y se puede realizar una prueba de conexión (Ver Figura 7).

Figura 7. Visualización configuración base de datos.



Fuente: Extraído de la herramienta Talend Open Source.

Una vez terminada la configuración, se puede comprobar que la conexión ha sido exitosa al desplegar el listado de las tablas y campos que compone esta base de datos (Ver Figura 8).

Figura 8. Listado de tablas y campos.



Fuente: Extraído de la herramienta Talend Open Source.

Después de estructurada cada conexión, se seleccionan los elementos a utilizar para habilitar la conexión de las fuentes de datos de origen, extracción de datos y almacenamiento en la fuente de destino.

Componente que define el inicio de la tarea (Job), y produce otras acciones que son necesarias para realizar la tarea (Ver Figura 9).

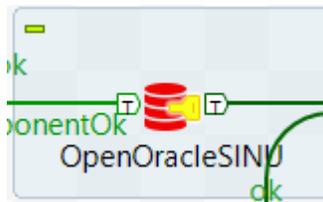
Figura 9. Tarea de inicio.



Fuente: Extraído de la herramienta Talend Open Source.

Componente que permite realizar la conexión a la base de datos, si este paso es exitoso, se puede continuar ejecutando otras tareas (Ver Figura 10).

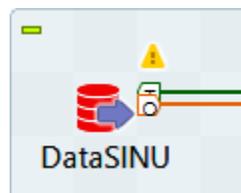
Figura 10. Conexión base de datos.



Fuente: Extraído de la herramienta Talend Open Source.

Componente que permite ejecutar las consultas de entrada de datos de una fuente, puede ser única a un objeto o compuesta de varios objetos, por ejemplo, compuesta de varias tablas (Ver Figura 11).

Figura 11. Ejecución de consultas.



Fuente: Extraído de la herramienta Talend Open Source.

Componente que permite ejecutar las operaciones de selección, limpieza, normalización, filtrado, iteración, unificación, este componente genera la entrada de varias fuentes y también permite definir varias salidas según la necesidad de la operación, estas entradas en motores de base de datos se conocen como JOIN.

Internamente se puede aplicar todo tipo de expresiones regulares, estas pueden ser creada mediante la programación por medio del lenguaje Java, creadas por el usuario o las que provee la herramienta (Ver Figura 12).

Figura 12. Mapeo de datos.



Fuente: Extraído de la herramienta Talend Open Source.

En este componente se definen las operaciones a realizar con los datos para la fuente de destino, las opciones son:

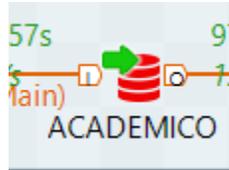
- Eliminar y crear tabla
- Crear tabla
- Crear tabla si no existe
- Eliminar tabla si existe y crear
- Limpiar tabla
- Truncar tabla
- Ninguno

También las opciones DDL de la base de datos:

- Insert
- Update
- Insert o update
- Update o insert
- Borrar

Finalmente, se puede definir, descartar, agregar y guardar todos los atributos de entrada, así mismo si se quiere cambiar algún tipo de dato (Ver Figura 13).

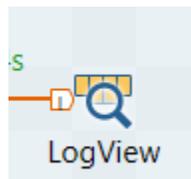
Figura 13. Definición de registros.



Fuente: Extraído de la herramienta Talend Open Source.

Con este componente se visualiza el flujo de registros por consola (Ver Figura 14).

Figura 14. Consola - log.



Fuente: Extraído de la herramienta Talend Open Source.

Componente que se utiliza para finalizar la ejecución de las tareas (Ver Figura 15).

Figura 15. Cierre de conexión.



Fuente: Extraído de la herramienta Talend Open Source.

Finalmente, con esta tarea se cierra la conexión abierta al inicio de la tarea (Ver Figura 16).

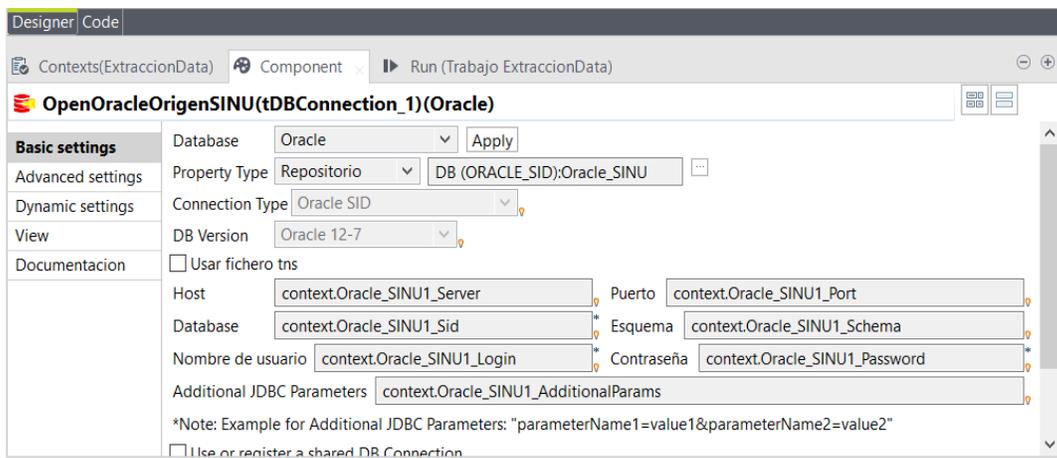
Figura 16. Cierre de conexión.



Fuente: Extraído de la herramienta Talend Open Source.

Una vez indicados cada uno de los componentes se establece la configuración o ajustes en cada uno. En el caso del componente de conexión a la BD, se debe seleccionar el contexto creado, de esta manera se podrán visualizar las variables que se crearon para en la figura 5, a continuación, se evidencia el detalle de esta configuración (Ver Figura 17).

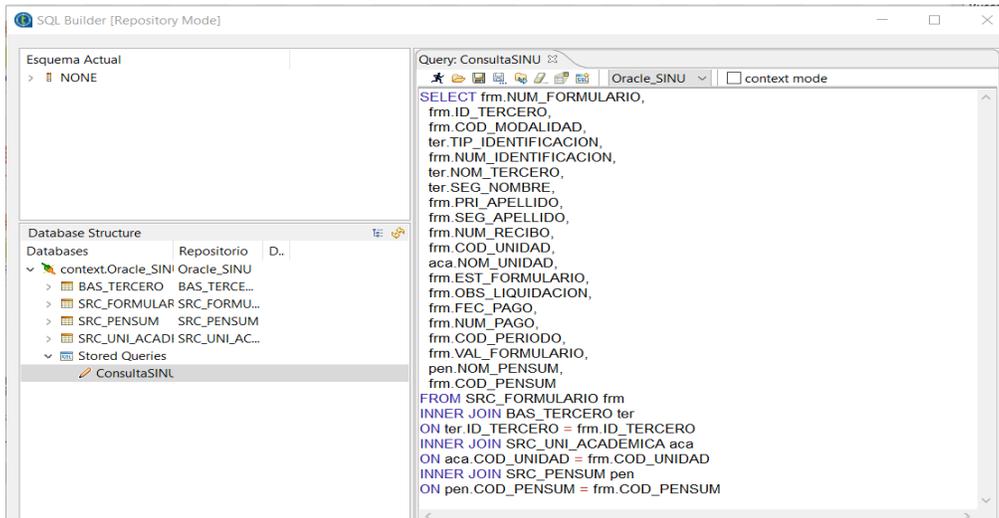
Figura 17. Configuración componente de conexión



Fuente: Extraído de la herramienta Talend Open Source.

Posteriormente, se selecciona el componente de ejecución de consultas. En la consola llamada “Query” se construye el script para seleccionar los campos que se requieren de cada una de las tablas descritas en la base de datos del Sistema Académico (Ver Figura 18).

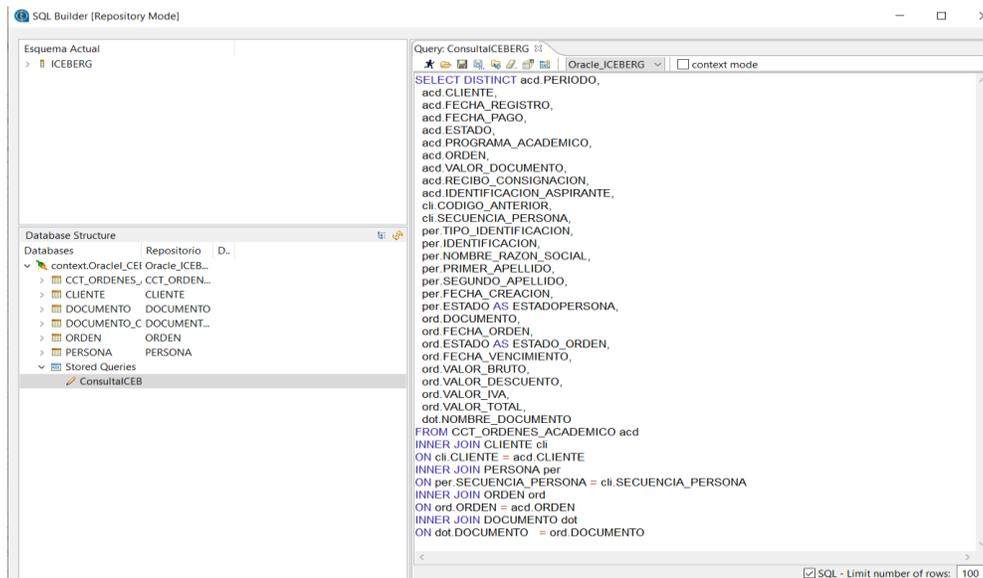
Figura 18. Creación de consulta Sistema Académico



Fuente: Extraído de la herramienta Talend Open Source.

Se realiza nuevamente el proceso indicado en el paso anterior, pero en la base de datos del Sistema Financiero (Ver Figura 19).

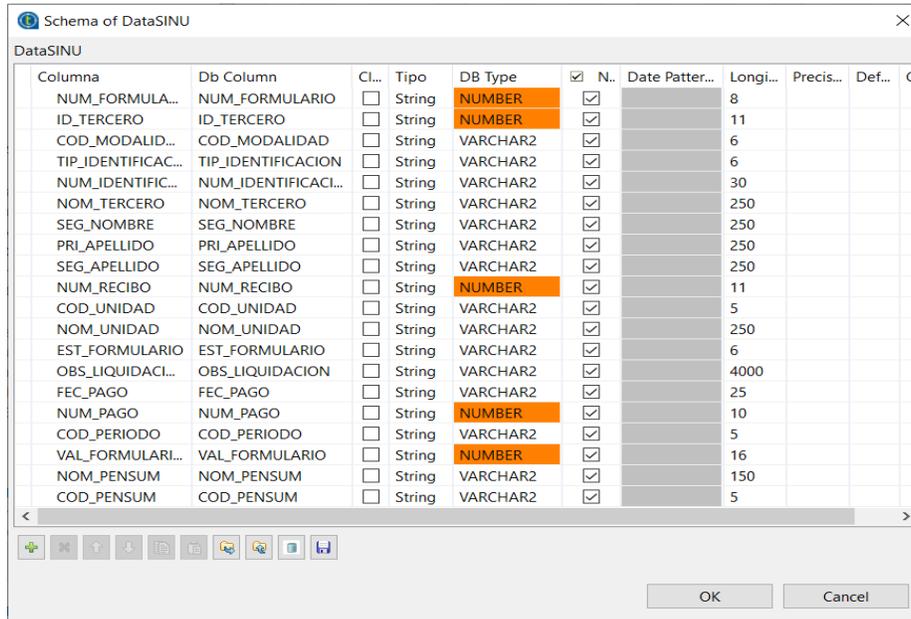
Figura 19. Creación de tabla Sistema Financiero



Fuente: Extraído de la herramienta Talend Open Source.

En la siguiente figura (Ver Figura 20) se visualiza el detalle de los campos seleccionados en la consulta del Sistema Académico, dónde se puede evidenciar sus nombres, tipo de datos, longitud y obligatoriedad.

Figura 20. Detalle de los campos Sistema Académico

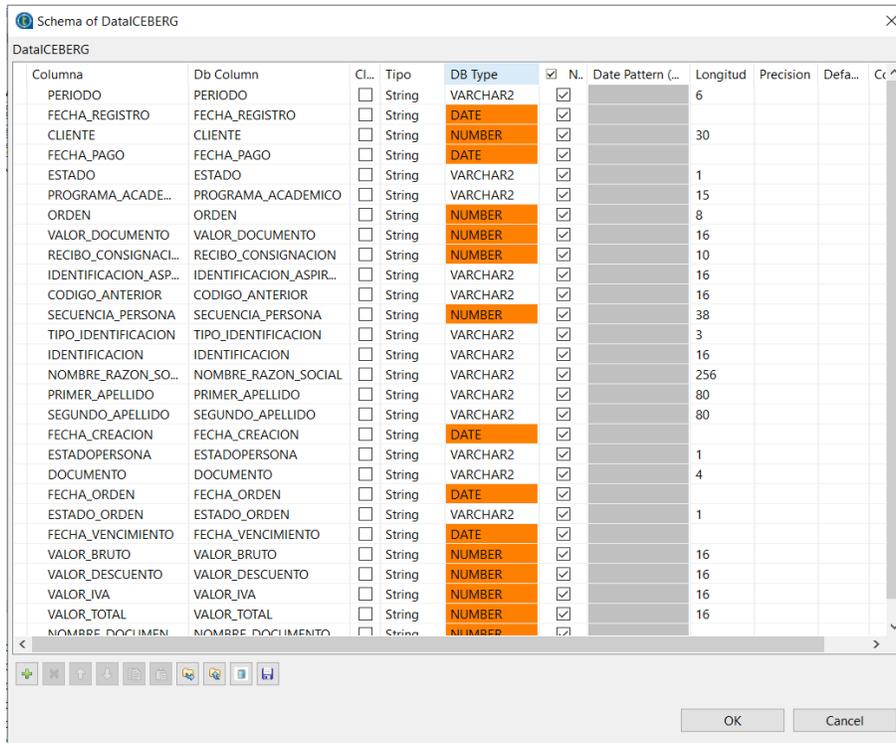


Columna	Db Column	Cl...	Tipo	DB Type	<input checked="" type="checkbox"/>	N..	Date Patter...	Longi...	Precis...	Def...	C
NUM_FORMULA...	NUM_FORMULARIO	<input type="checkbox"/>	String	NUMBER	<input checked="" type="checkbox"/>			8			
ID_TERCERO	ID_TERCERO	<input type="checkbox"/>	String	NUMBER	<input checked="" type="checkbox"/>			11			
COD_MODALID...	COD_MODALIDAD	<input type="checkbox"/>	String	VARCHAR2	<input checked="" type="checkbox"/>			6			
TIP_IDENTIFICAC...	TIP_IDENTIFICACION	<input type="checkbox"/>	String	VARCHAR2	<input checked="" type="checkbox"/>			6			
NUM_IDENTIFIC...	NUM_IDENTIFICACI...	<input type="checkbox"/>	String	VARCHAR2	<input checked="" type="checkbox"/>			30			
NOM_TERCERO	NOM_TERCERO	<input type="checkbox"/>	String	VARCHAR2	<input checked="" type="checkbox"/>			250			
SEG_NOMBRE	SEG_NOMBRE	<input type="checkbox"/>	String	VARCHAR2	<input checked="" type="checkbox"/>			250			
PRI_APELLIDO	PRI_APELLIDO	<input type="checkbox"/>	String	VARCHAR2	<input checked="" type="checkbox"/>			250			
SEG_APELLIDO	SEG_APELLIDO	<input type="checkbox"/>	String	VARCHAR2	<input checked="" type="checkbox"/>			250			
NUM_RECIBO	NUM_RECIBO	<input type="checkbox"/>	String	NUMBER	<input checked="" type="checkbox"/>			11			
COD_UNIDAD	COD_UNIDAD	<input type="checkbox"/>	String	VARCHAR2	<input checked="" type="checkbox"/>			5			
NOM_UNIDAD	NOM_UNIDAD	<input type="checkbox"/>	String	VARCHAR2	<input checked="" type="checkbox"/>			250			
EST_FORMULARIO	EST_FORMULARIO	<input type="checkbox"/>	String	VARCHAR2	<input checked="" type="checkbox"/>			6			
OBS_LIQUIDACI...	OBS_LIQUIDACION	<input type="checkbox"/>	String	VARCHAR2	<input checked="" type="checkbox"/>			4000			
FEC_PAGO	FEC_PAGO	<input type="checkbox"/>	String	VARCHAR2	<input checked="" type="checkbox"/>			25			
NUM_PAGO	NUM_PAGO	<input type="checkbox"/>	String	NUMBER	<input checked="" type="checkbox"/>			10			
COD_PERIODO	COD_PERIODO	<input type="checkbox"/>	String	VARCHAR2	<input checked="" type="checkbox"/>			5			
VAL_FORMULARI...	VAL_FORMULARIO	<input type="checkbox"/>	String	NUMBER	<input checked="" type="checkbox"/>			16			
NOM_PENSUM	NOM_PENSUM	<input type="checkbox"/>	String	VARCHAR2	<input checked="" type="checkbox"/>			150			
COD_PENSUM	COD_PENSUM	<input type="checkbox"/>	String	VARCHAR2	<input checked="" type="checkbox"/>			5			

Fuente: Extraído de la herramienta Talend Open Source.

En la siguiente figura se evidencia el detalle indicado anteriormente, ahora para los campos seleccionados en el Sistema Financiero (Ver Figura 21).

Figura 21. Detalle de los campos Sistema Financiero

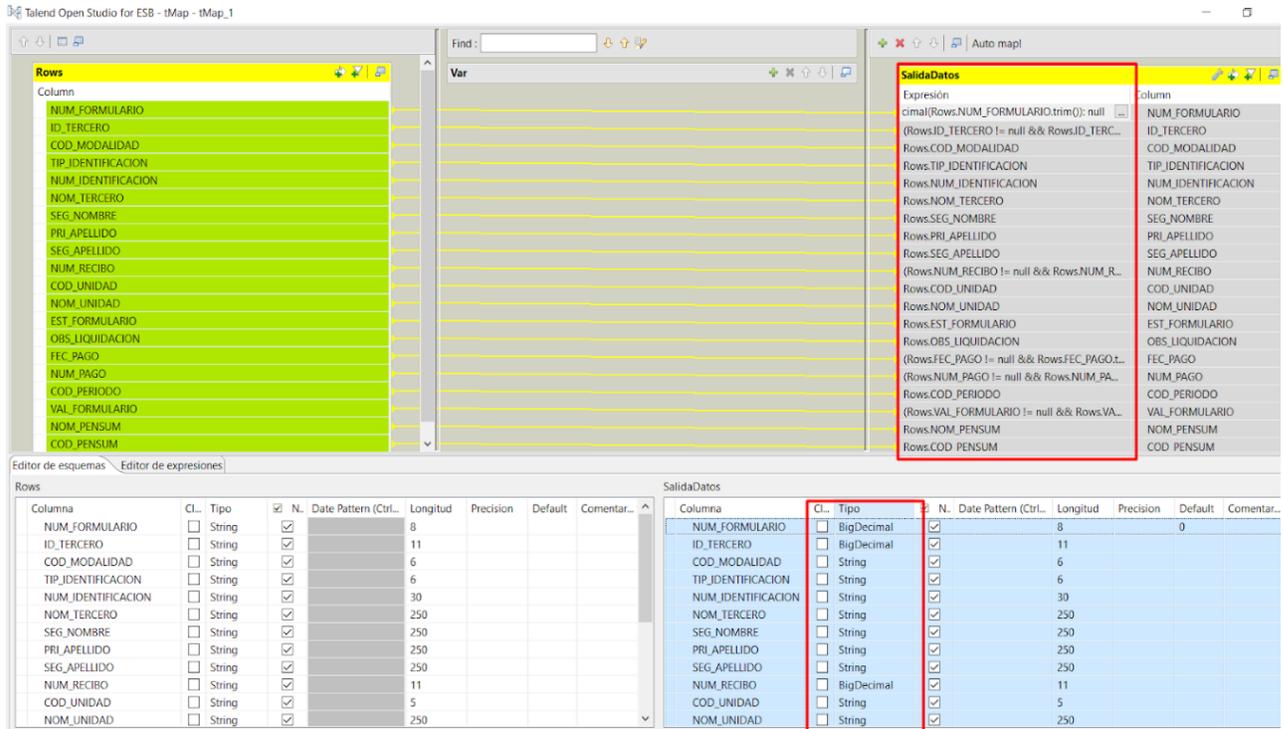


Columna	Db Column	Cl...	Tipo	DB Type	N.	Date Pattern (...)	Longitud	Precision	Defa...	Cr...
PERIODO	PERIODO	<input type="checkbox"/>	String	VARCHAR2	<input checked="" type="checkbox"/>		6			
FECHA_REGISTRO	FECHA_REGISTRO	<input type="checkbox"/>	String	DATE	<input checked="" type="checkbox"/>					
CLIENTE	CLIENTE	<input checked="" type="checkbox"/>	String	NUMBER	<input checked="" type="checkbox"/>		30			
FECHA_PAGO	FECHA_PAGO	<input type="checkbox"/>	String	DATE	<input checked="" type="checkbox"/>					
ESTADO	ESTADO	<input type="checkbox"/>	String	VARCHAR2	<input checked="" type="checkbox"/>		1			
PROGRAMA_ACADE...	PROGRAMA_ACADEMICO	<input type="checkbox"/>	String	VARCHAR2	<input checked="" type="checkbox"/>		15			
ORDEN	ORDEN	<input type="checkbox"/>	String	NUMBER	<input checked="" type="checkbox"/>		8			
VALOR_DOCUMENTO	VALOR_DOCUMENTO	<input type="checkbox"/>	String	NUMBER	<input checked="" type="checkbox"/>		16			
RECIBO_CONSIGNACL...	RECIBO_CONSIGNACION	<input type="checkbox"/>	String	NUMBER	<input checked="" type="checkbox"/>		10			
IDENTIFICACION_ASP...	IDENTIFICACION_ASPIR...	<input type="checkbox"/>	String	VARCHAR2	<input checked="" type="checkbox"/>		16			
CODIGO_ANTERIOR	CODIGO_ANTERIOR	<input type="checkbox"/>	String	VARCHAR2	<input checked="" type="checkbox"/>		16			
SECUENCIA_PERSONA	SECUENCIA_PERSONA	<input type="checkbox"/>	String	NUMBER	<input checked="" type="checkbox"/>		38			
TIPO_IDENTIFICACION	TIPO_IDENTIFICACION	<input type="checkbox"/>	String	VARCHAR2	<input checked="" type="checkbox"/>		3			
IDENTIFICACION	IDENTIFICACION	<input type="checkbox"/>	String	VARCHAR2	<input checked="" type="checkbox"/>		16			
NOMBRE_RAZON_SO...	NOMBRE_RAZON_SOCIAL	<input type="checkbox"/>	String	VARCHAR2	<input checked="" type="checkbox"/>		256			
PRIMER_APELLIDO	PRIMER_APELLIDO	<input type="checkbox"/>	String	VARCHAR2	<input checked="" type="checkbox"/>		80			
SEGUNDO_APELLIDO	SEGUNDO_APELLIDO	<input type="checkbox"/>	String	VARCHAR2	<input checked="" type="checkbox"/>		80			
FECHA_CREACION	FECHA_CREACION	<input type="checkbox"/>	String	DATE	<input checked="" type="checkbox"/>					
ESTADOPERSONA	ESTADOPERSONA	<input type="checkbox"/>	String	VARCHAR2	<input checked="" type="checkbox"/>		1			
DOCUMENTO	DOCUMENTO	<input type="checkbox"/>	String	VARCHAR2	<input checked="" type="checkbox"/>		4			
FECHA_ORDEN	FECHA_ORDEN	<input type="checkbox"/>	String	DATE	<input checked="" type="checkbox"/>					
ESTADO_ORDEN	ESTADO_ORDEN	<input type="checkbox"/>	String	VARCHAR2	<input checked="" type="checkbox"/>		1			
FECHA_VENCIMIENTO	FECHA_VENCIMIENTO	<input type="checkbox"/>	String	DATE	<input checked="" type="checkbox"/>					
VALOR_BRUTO	VALOR_BRUTO	<input type="checkbox"/>	String	NUMBER	<input checked="" type="checkbox"/>		16			
VALOR_DESCUENTO	VALOR_DESCUENTO	<input type="checkbox"/>	String	NUMBER	<input checked="" type="checkbox"/>		16			
VALOR_IVA	VALOR_IVA	<input type="checkbox"/>	String	NUMBER	<input checked="" type="checkbox"/>		16			
VALOR_TOTAL	VALOR_TOTAL	<input type="checkbox"/>	String	NUMBER	<input checked="" type="checkbox"/>		16			
NOMBRE DOCUMEN	NOMBRE DOCUMENTO	<input type="checkbox"/>	String	NUMBER	<input checked="" type="checkbox"/>					

Fuente: Extraído de la herramienta Talend Open Source.

Posteriormente, con el componente del mapeo se seleccionan los campos que serán enviados a la base de datos de destino, para esto, se realiza una conversión de los tipos de datos, como cambiar datos de texto a numéricos, es decir de String a BigDecimal (Ver Figura 22).

Figura 22. Mapeo de datos



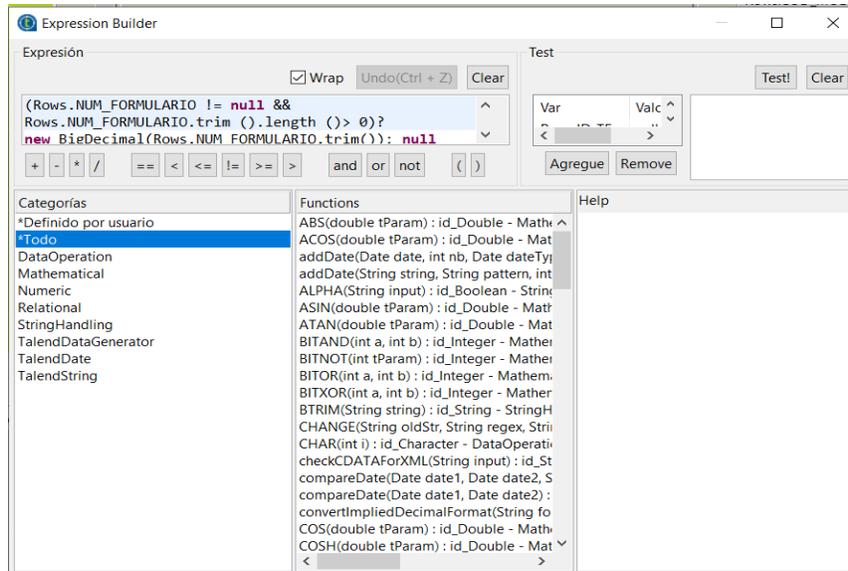
Fuente: Extraído de la herramienta Talend Open Source.

Además, se aplican reglas para la validación de los datos, creando condiciones para:

- Extraer únicamente los datos de los recibos que cuentan con un formulario registrado en el Sistema Académico.
- Obtener los recibos desde el año 2021 primer semestre en adelante.
- Obtener los recibos de correspondan a los conceptos de inscripción en las sedes Bogotá y Armenia (AOCC, BOCC).

De esta manera, se generan las especificaciones que deben cumplir los registros para garantizar que sean adecuados para su análisis y revisión (Ver Figura 23).

Figura 23. Aplicación de reglas



Fuente: Extraído de la herramienta Talend Open Source.

De esta manera se puede realizar la conversión final de los datos, para que sean almacenados en la base de datos de destino (Ver Figura 24).

Figura 24. Conversión de datos

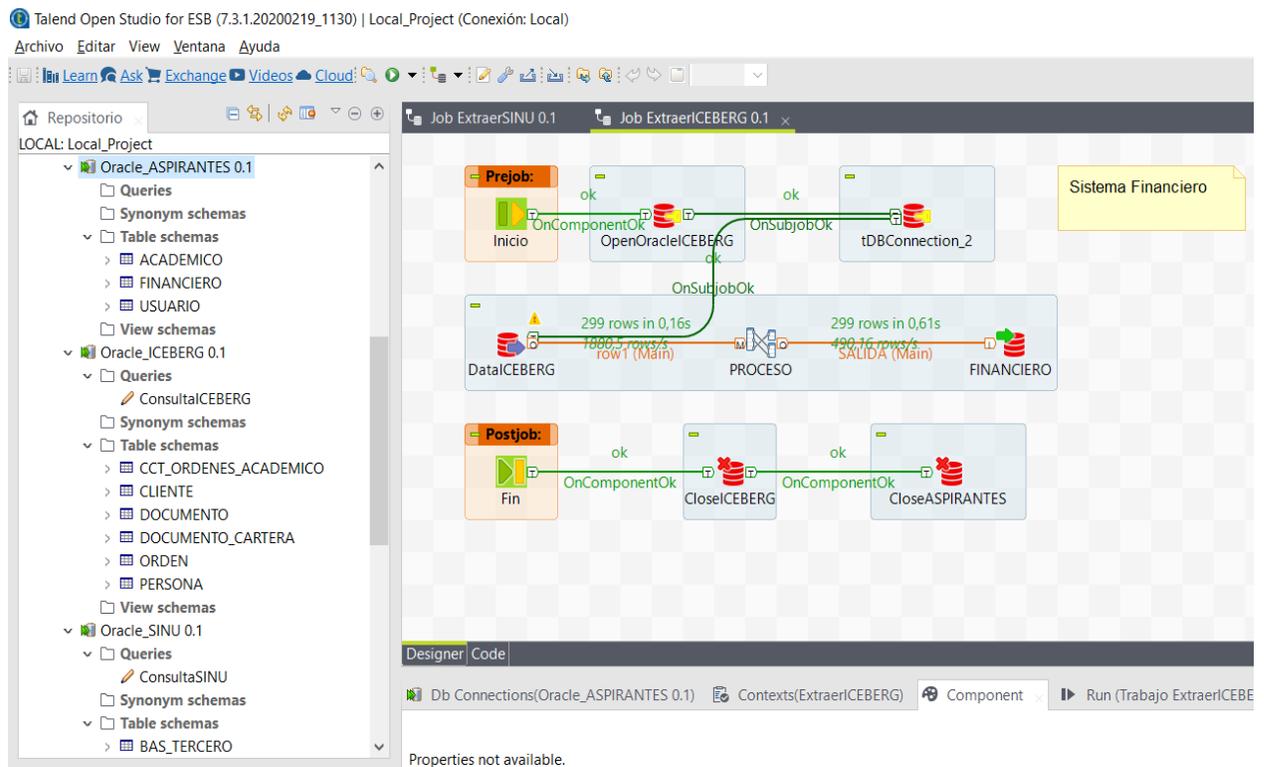
Columna	Cl...	Tipo	✓	N.	Date Patte...	Longi...	Preci...	Def...
NUM_FORMUL...		BigDecimal	✓			8		0
ID_TERCERO		BigDecimal	✓			11		
COD_MODALID...		String	✓			6		
TIP_IDENTIFICA...		String	✓			6		
NUM_IDENTIFI...		String	✓			30		
NOM_TERCERO		String	✓			250		
SEG_NOMBRE		String	✓			250		
PRI_APELLIDO		String	✓			250		
SEG_APELLIDO		String	✓			250		
NUM_RECIBO		BigDecimal	✓			11		
COD_UNIDAD		String	✓			5		
NOM_UNIDAD		String	✓			250		
EST_FORMULA...		String	✓			6		
OBS_LIQUIDACI...		String	✓			4000		
FEC_PAGO		String	✓			10		
NUM_PAGO		BigDecimal	✓			10		
COD_PERIODO		String	✓			5		
VAL_FORMULA...		BigDecimal	✓			16		
NOM_PENSUM		String	✓			150		
COD_PENSUM		String	✓			5		

Columna	Db Column	Cl.	Tipo	DB Type	✓	N.	Date P...	Lo...
NUM_FOR...	NUM_FORM...		BigDecimal	NUMBER	✓			8
ID_TERCE...	ID_TERCERO		BigDecimal	NUMBER	✓			11
COD_MO...	COD_MODALID...		String	VARCHAR2	✓			6
TIP_IDENT...	TIP_IDENTIFI...		String	VARCHAR2	✓			6
NUM_IDE...	NUM_IDENTI...		String	VARCHAR2	✓			30
NOM_TER...	NOM_TERCE...		String	VARCHAR2	✓			250
SEG_NOM...	SEG_NOMBRE		String	VARCHAR2	✓			250
PRI_APELL...	PRI_APELLIDO		String	VARCHAR2	✓			250
SEG_APELL...	SEG_APELLID...		String	VARCHAR2	✓			250
NUM_REC...	NUM_RECIBO		BigDecimal	NUMBER	✓			11
COD_UNI...	COD_UNIDAD		String	VARCHAR2	✓			5
NOM_UNI...	NOM_UNIDA...		String	VARCHAR2	✓			250
EST_FOR...	EST_FORMUL...		String	VARCHAR2	✓			6
OBS_LIQU...	OBS_LIQUID...		String	VARCHAR2	✓			40...
FEC_PAGO	FEC_PAGO		String	VARCHAR2	✓			10
NUM_PAGO	NUM_PAGO		BigDecimal	NUMBER	✓			10
COD_PERI...	COD_PERIO...		String	VARCHAR2	✓			5
VAL_FOR...	VAL_FORMU...		BigDecimal	NUMBER	✓			16
NOM_PEN...	NOM_PENSU...		String	VARCHAR2	✓			150
COD_PEN...	COD_PENSUM		String	VARCHAR2	✓			5

Fuente: Extraído de la herramienta Talend Open Source.

El resultado final del diseño para la extracción en la base de datos del Sistema Financiero es el siguiente (Ver Figura 25).

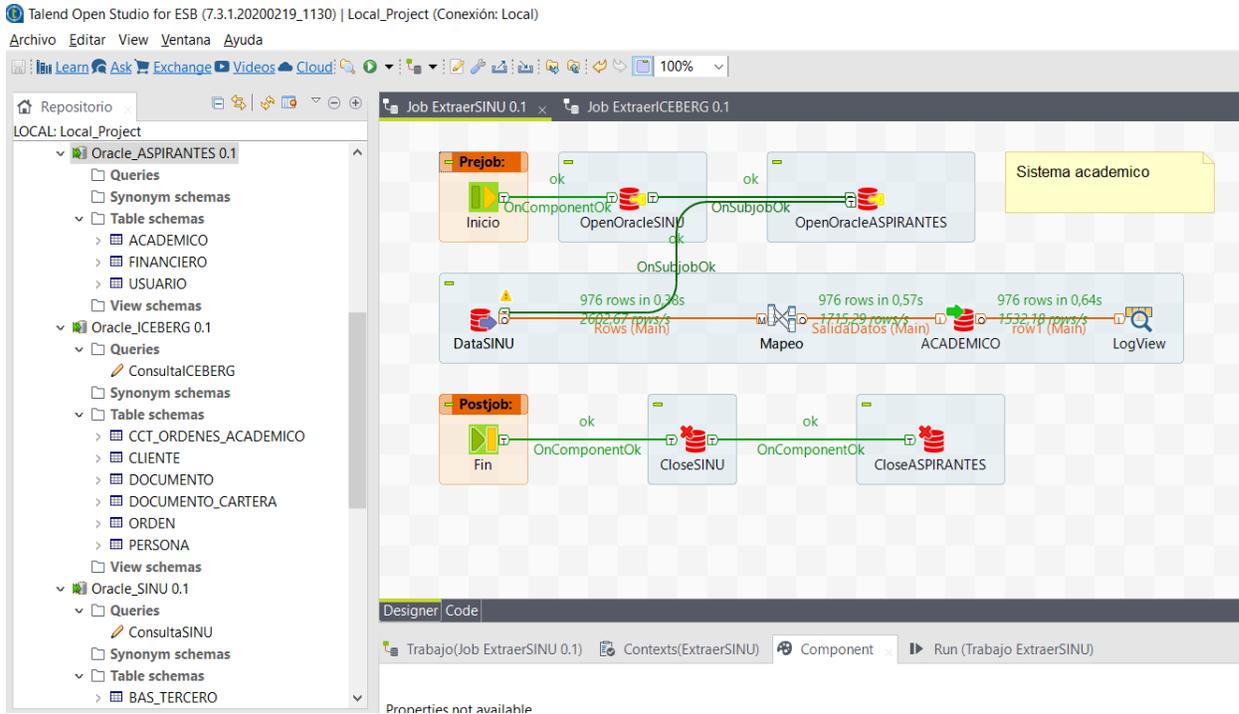
Figura 25. Diseño final para extracción Sistema Financiero



Fuente: Extraído de la herramienta Talend Open Source.

Y el resultado final del diseño para la extracción en la base de datos del Sistema Académico, es el siguiente (Ver Figura 26).

Figura 26. Diseño final para extracción Sistema Académico



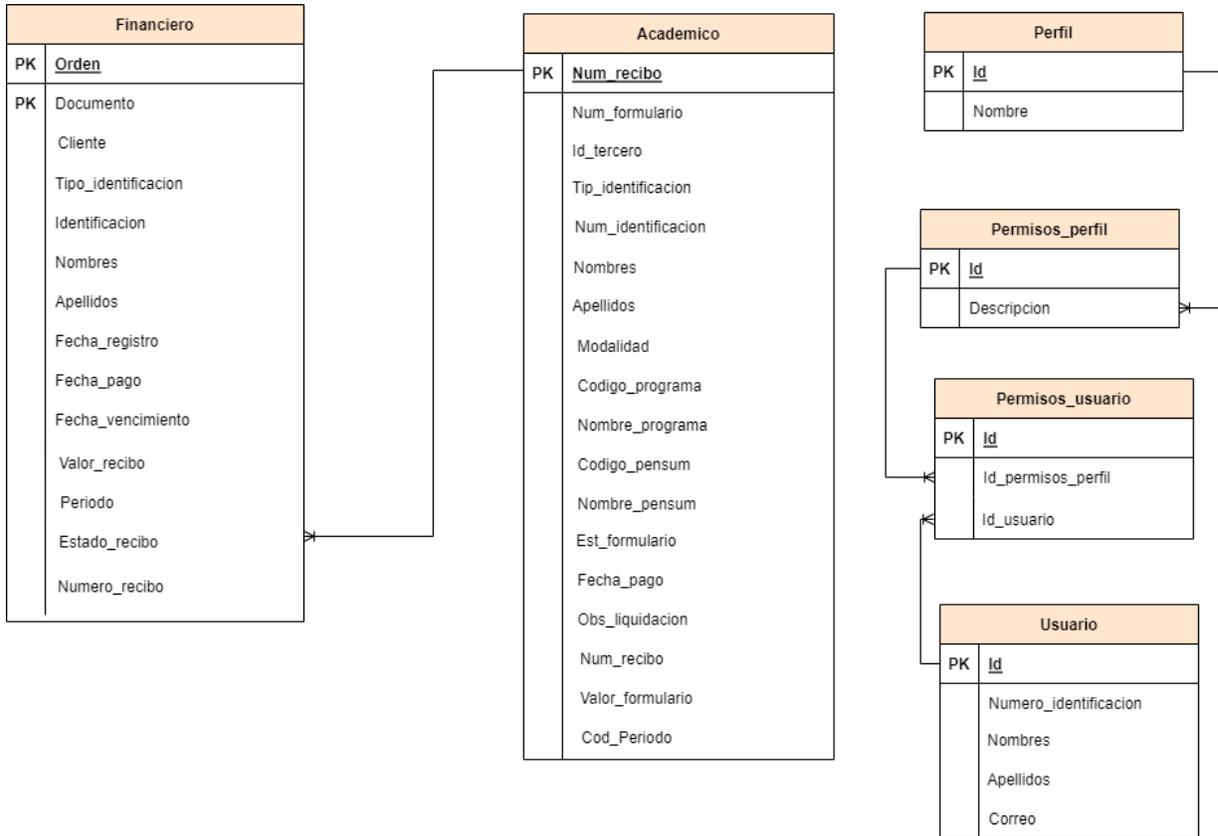
Fuente: Extraído de la herramienta Talend Open Source.

Finalizado este paso, se prosigue a la segunda etapa de esta investigación, la cual corresponde al procesamiento de datos.

5.2. Procesamiento de datos

Se realiza el proceso de selección y normalización de los datos que son requeridos para el planteamiento de este proyecto, por lo cual, se lleva a cabo en esta etapa el diseño del nuevo modelo de datos (Ver Figura 27).

Figura 27. Modelo relacional nuevo modelo



Fuente: Elaboración propia.

Se realiza la implementación de este nuevo modelo en la base de datos de destino, la cual se encuentra instalada en un servidor local de la institución, esta base de datos se encuentra en el motor Oracle en la versión 11. El detalle de los campos creados para este modelo se puede consultar en el Anexo 1 (Diccionario de datos). Para cargar los datos, se realizó la creación de una tarea en la herramienta Talend, por lo cual, durante la etapa de la extracción se procesan los datos de la base de datos de origen y se almacenan en la de destino.

Finalmente, una vez se establece el nuevo modelo, se procede con el desarrollo de la aplicación para la visualización de los datos.

5.3. Visualización de datos

Para llevar a cabo la visualización de datos en esta etapa, no se utilizará una herramienta de las que existen en el mercado, en su lugar se desarrollará una página web para presentar esta información. Como primer paso se establecen los requerimientos funcionales y no funcionales que debe tener la aplicación para su correcto uso y aprovechamiento en la institución:

Requerimientos funcionales

- La herramienta permitirá crear 3 tipos de usuarios: administrador, usuario operativo y usuario gerencial.
- La aplicación debe permitir administrar los usuarios que ingresen con sus respectivos privilegios.
- El usuario de administrador debe tener la posibilidad de agregar nuevos usuarios y asignar los roles definidos.
- Tendrá un inicio de sesión que debe integrarse para la autenticación con el directorio activo LDAP, el cual es un protocolo ligero de acceso a directorio, con el propósito de utilizar el mismo usuario y contraseña de otras plataformas tecnológicas que tiene la institución.
- Debe permitir la búsqueda por diferentes filtros, de acuerdo con las necesidades de los usuarios.
- El usuario del sistema debe poder visualizar en detalle el listado de los aspirantes, así como su estado de pago en el Sistema Académico y su estado de pago en el Sistema Financiero.
- El usuario del sistema debe tener la posibilidad de visualizar mediante gráficas la cantidad de recibos que se encuentran pagos y no pagos.
- La aplicación debe permitir al usuario consultar la información almacenada, pero no modificarla o eliminarla.

Requerimientos no funcionales

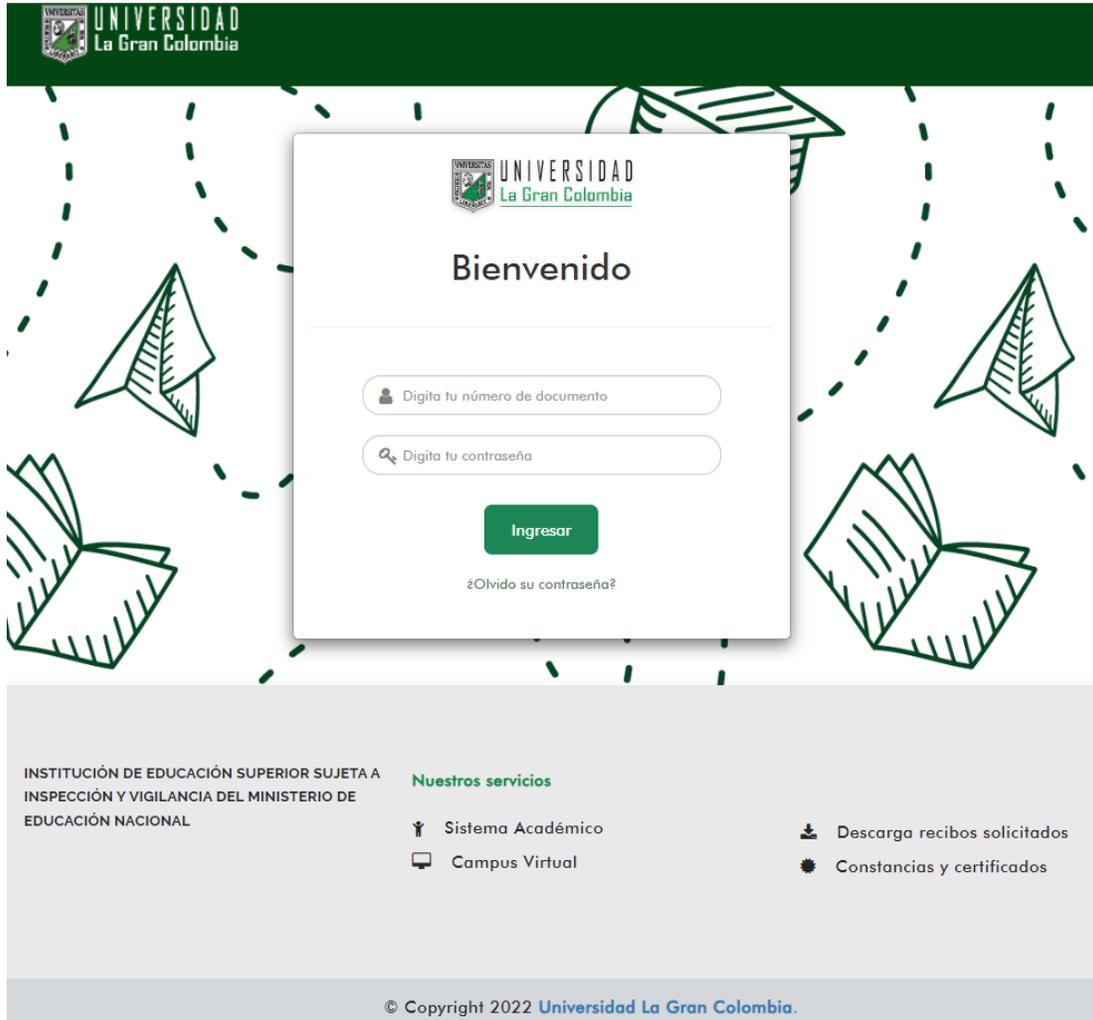
- El diseño de la aplicación debe realizarse con base a lo establecido en el manual de marca institucional.
- Debe tener ayudas visuales para el usuario, como colores para identificar los diferentes estados que se ven en pantalla.
- La herramienta debe ser entendible para cualquier usuario al que se proporcionen los permisos y credenciales de acceso, mediante una interfaz sencilla para que sea fácil de manejar.
- El tiempo de carga dentro de los filtros definidos no debe superar los 5 segundos de respuesta.
- La aplicación debe tener una disponibilidad continua, con un nivel de servicio para que los usuarios puedan ingresar y consultar la información siempre que sea requerido. En caso de presentarse alguna caída el servicio debe proporcionar un mensaje claro y adecuado a los usuarios.
- El acceso a los datos debe efectuarse de manera segura.
- La herramienta deberá ser escalable, con el propósito de que pueda continuar creciendo e incorporar nuevas funcionalidades.
- Debe ser funcional en diferentes navegadores, dispositivos y poder adaptarse en diferentes resoluciones de pantalla.
- La aplicación debe presentar mensajes de error entendibles para el usuario.

Teniendo en cuenta los requerimientos indicados anteriormente, se procedió con el desarrollo de la aplicación, generando el siguiente resultado:

Inicio de sesión

Se desarrolla la primera pantalla para el inicio de sesión, en la siguiente imagen se evidencia el diseño con los colores institucionales y la integración con el directorio activo (Ver Figura 28).

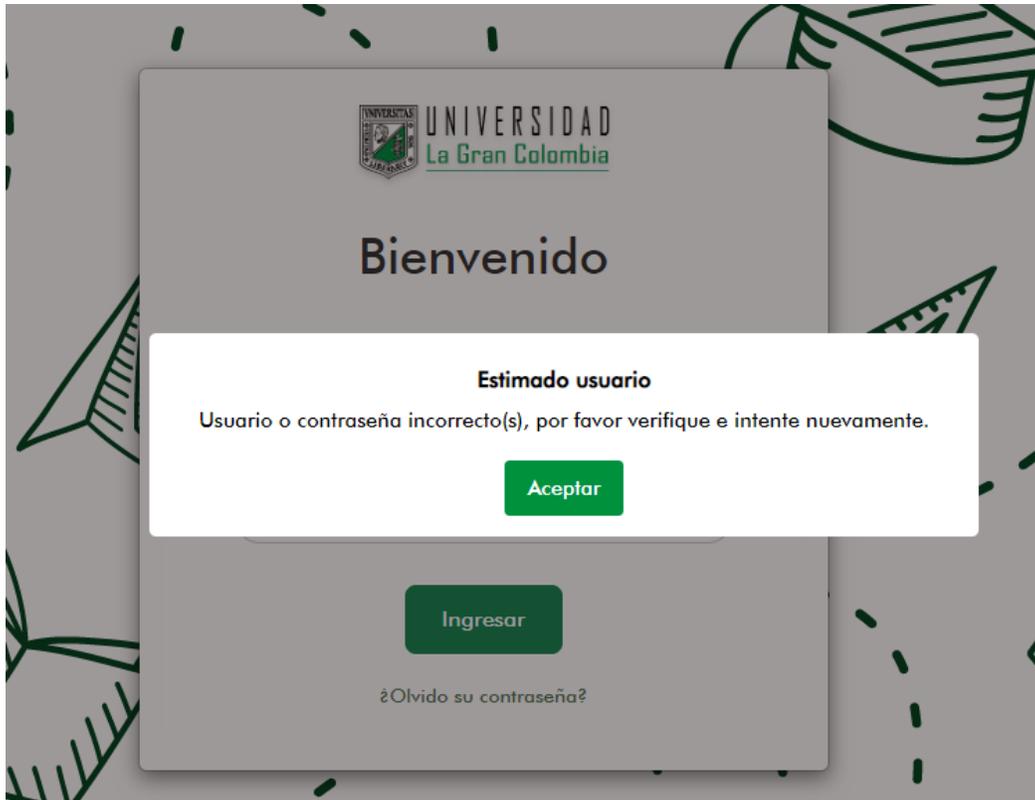
Figura 28. Diseño inicio de sesión



Fuente: Extraído aplicación web.

Se puede observar que si el usuario digita una contraseña diferente a la almacenada en el directorio activo o no existe aparece el siguiente mensaje (Ver Figura 29).

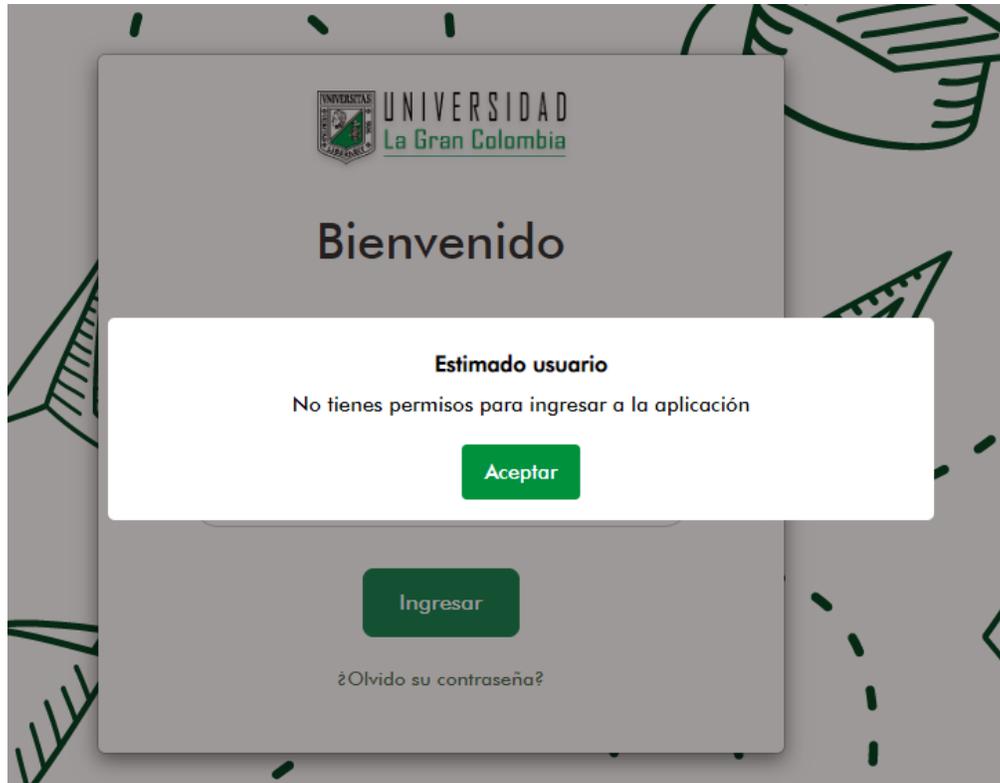
Figura 29. Mensaje de credenciales inválidas



Fuente: Extraído aplicación web.

En la siguiente figura, se puede validar el mensaje de alerta cuando el usuario no tiene permisos sobre la aplicación (Ver Figura 30).

Figura 30. Mensaje de permisos



Fuente: Extraído aplicación web.

Visualización de datos

Se crea la sección para la visualización de los datos, la cual tiene los siguientes elementos:

- Filtros modificables para la consulta de la información, se encuentra el filtro obligatorio del periodo y como filtros opcionales el programa y plan de estudios.
- Después de seleccionar los filtros, dar clic en el botón “Buscar”, a continuación, se genera la tabla para listar los registros, de acuerdo con los filtros aplicados.

El listado permite obtener los registros de los aspirantes que han realizado su proceso de admisión, así como su estado de pago en el Sistema Académico y su estado de pago en el Sistema Financiero (Ver Figura 31).

Figura 31. Visualización de datos



UNIVERSIDAD La Gran Colombia Inicio Reportes Dashboard

Selecciona los filtros que desees utilizar

Selecciona el periodo: 2021-11

Selecciona el programa (opcional): << Seleccione >>

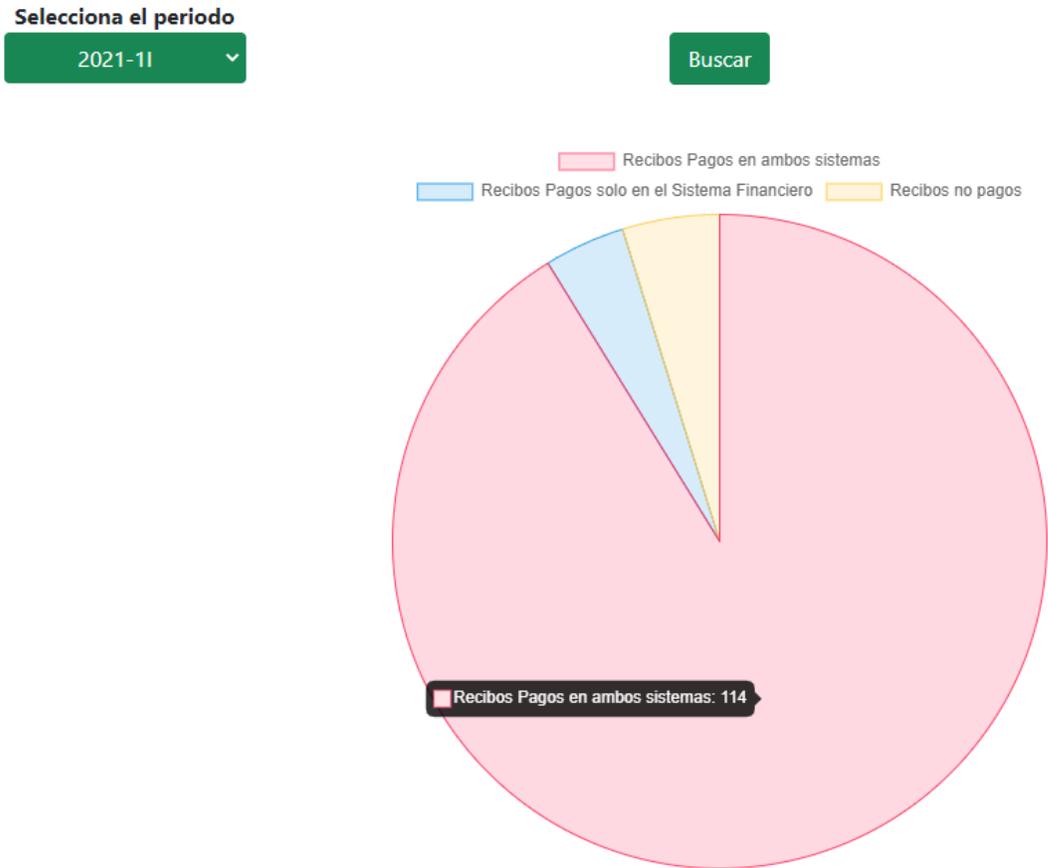
Buscar

FORMULARIO	IDENTIFICACION	NOMBRE COMPLETO	PROGRAMA	FECHA SGA	ORDEN	FECHA REGISTRO	FECHA SIAF	FECHA VENCIMIENTO	No RECIBO	VALOR	ESTADO
7437	649762	JAVIER ATERNINA	Ingenieria Civil		13810	29-APR-21		13-AUG-21	40738	67000	G
7410	861613	MILRO ROJAS	Administracion de Empresas	28-APR-21	13693	27-APR-21	28-APR-21	13-AUG-21	40292	67000	P
7327	819409	ALEJANDRA BAQUERO	Economia	12-MAY-21	13107	22-APR-21	11-MAY-21	04-AUG-21	39263	67000	R

Fuente: Extraído aplicación web.

Además, tiene la posibilidad de visualizar datos estadísticos, en la siguiente gráfica se obtiene la cantidad de recibos pagos en ambos sistemas, la cantidad de registros pagos en el Sistema Financiero, pero no en el Sistema Académico y la cantidad de registros que no tienen ningún pago (Ver Figura 32).

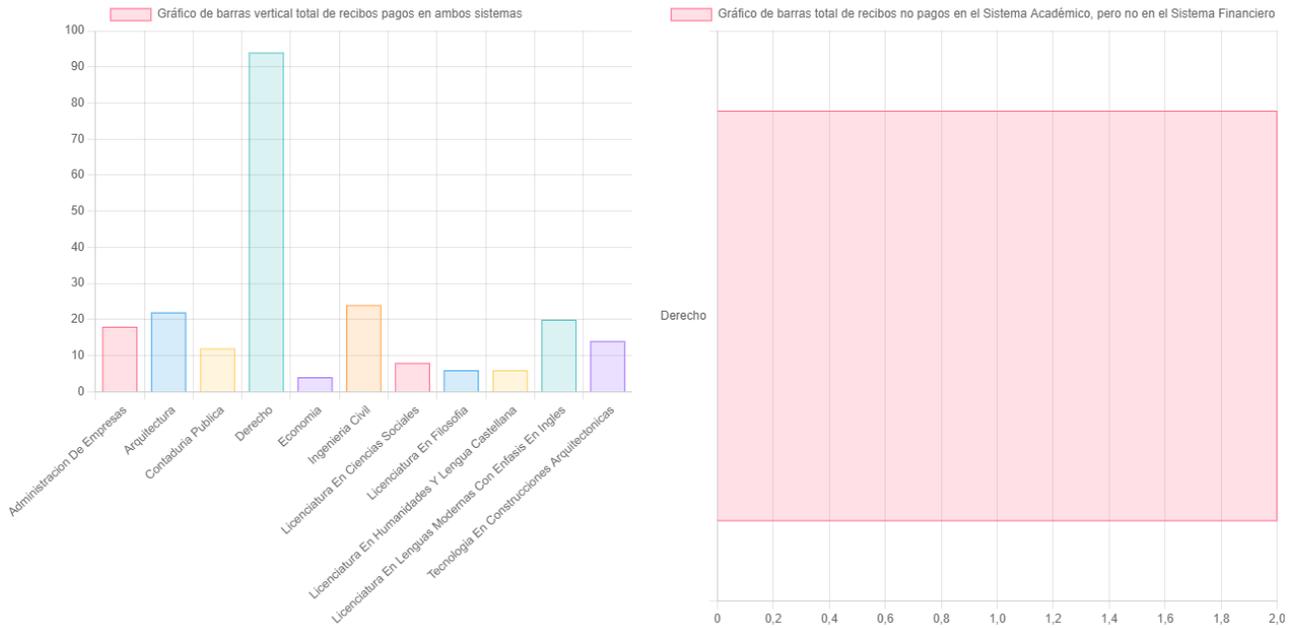
Figura 32. Visualización de datos estadísticos en gráfico circular.



Fuente: Extraído aplicación web.

En la siguiente gráfica, se obtiene la cantidad de recibos pagos en ambos sistemas, por cada uno de los programas y la cantidad de registros pagos en el Sistema Financiero, pero no en el Sistema Académico, por cada uno de los programas (Ver Figura 32).

Figura 33. Visualización de datos estadísticos en gráfico de barras.



Fuente: Extraído aplicación web.

6. Conclusiones

Es importante realizar la caracterización de las tecnologías utilizadas en cada proceso, ya que esto permite conocer en mayor detalle el flujo que manejan los datos, de igual forma se puede identificar los posibles motivos de la problemática y plantear una solución con un alcance adecuado que se relacione a las necesidades del negocio. Además, la caracterización permitió identificar los datos y tablas que interfieren en el proceso de admisión de aspirantes, así como el registro de los pagos, con el propósito de realizar validaciones en la extracción de los datos con la relación de tablas.

La creación del nuevo modelo de datos para el proceso de admisión y pagos permitió acortar la cantidad de tablas que se utilizan en las fuentes de datos de origen, de esta manera se reduce el consumo de datos durante las consultas. Además, con la generación de nuevas tablas se puede gestionar los permisos de los usuarios que podrán tener una visualización sobre la información consultada y de esta manera se mejora el rendimiento de la aplicación para la visualización de datos.

Con la implementación de la aplicación para las oficinas de Cartera y Admisiones se logró estructurar y recuperar la información consolidada en el nuevo modelo de datos de forma que sea concisa, pero de valor para el usuario, con esto se puede identificar a los aspirantes que presentan inconsistencia de pago entre los Sistemas Financiero y Académico de la institución.

7. Referencias

- Acosta, J. Flórez, D. (2015). Diseño e implementación de prototipo BI utilizando una herramienta de Big Data para empresas pymes distribuidoras de tecnologías. Recuperado de https://repository.ucatolica.edu.co/bitstream/10983/2543/1/PROYECTO%20FINAL_.pdf
- Bahit, E. (2014). POO y MVC en PHP. Recuperado de <https://bibliotecafacet.com.ar/wp-content/uploads/2014/12/eugeniabahitpooymvcenphp.pdf>
- Banerjee, A. (2004). Choose Your ETL Tool Carefully. Recuperado de <http://www.datawa-rehouse.com/article/?articleid=4881>.
- Celaya, A. (2014). Creación de páginas web: HTML 5.
- Cuchillas, R. Hernández, O. Mejía, Y. Silva, H. (2010). Desarrollo de un “data warehouse” para el proceso de denuncias de la defensoría del consumidor. Recuperado de <http://ri.ues.edu.sv/id/eprint/10307/1/Desarrollo%20de%20un%20Data%20Warehose%20para%20el%20proceso%20de%20denuncias%20de%20la%20Defensor%20C3%ADa%20del%20Consumidor.pdf>
- DAMA International, Early, B., Henderson, D., & Data Management Association. (2017). Data Management Body of Knowledge (DAMA-DMBOK guide), second edition. Technics Publications.
- García P., M., Quispe A., C., & Ráez G., L. (2003). Mejora Continua de la Calidad en los Procesos. Industrial Data, 6(1), 089–094. <https://doi.org/10.15381/idata.v6i1.5992>

- Gallego, A. (2017). Bootstrap 3: Introducción al diseño responsive.
- Hitachi Vantara. (2021). Data Fabric for Dummies.
- International Organization for Standardization, 2005. ISO 9000:2005: Sistemas de gestión de la calidad – Requisitos. Ginebra: ISO.
- International Organization for Standardization, 2015. ISO 9001:2015: Sistemas de gestión de la calidad – Requisitos. Ginebra: ISO.
- Kaplan, R. y Norton, D. (2001), “Cómo utilizar el cuadro de mando integral: para implementar y gestionar su estrategia”. Barcelona: Gestión 2000.
- Mohanty, S. (2004). Optimal ETL Processing – Changed Data Capture Strategy.
- Recio García, J. (2016). HTML, CSS3 y JQuery: Curso práctico.
- Sotaquirá, W. (2017). Power BI como herramienta de big data & business analytics para Onelink Colombia. Recuperado de https://repository.eafit.edu.co/bitstream/handle/10784/11767/SotaquiraAyala_WilliamJesus_2017.pdf?sequence=2&isAllowed=y
- González, W., Mora, J., Pérez, S., (2014). Análisis del procesamiento de los datos de entrada para un localizador de fallas en sistemas de distribución Analysis of the input data processing for fault location in power distribution systems. Redalyc.org. Recuperado el 9 de abril de 2022, de <https://www.redalyc.org/pdf/2570/257031319006.pdf>

Anexos

Anexo 1. Diccionario de datos.

Las columnas creadas para este diccionario son:

- ID: Número de columna
- Nombre del campo en la base de datos
- Formato: Tipo de campo que debe crearse para dicho campo.
- Descripción: Información de lo que almacena el campo.
- Llave: Obligatoriedad del campo.
 - PK: Llave primaria para cada tabla.
 - NN: Campo no nulo, es decir obligatorio.
 - Si se encuentra vacío el campo es nulo, es decir no obligatorio.

Se genera la lista de los elementos a utilizar dentro del nuevo flujo de los datos para la aplicación.

ID	Campo	Formato	Descripción	Llave
1	NUM_RECIBO	Numérico	Número del recibo de pago	PK
2	NUM_FORMULARIO	Numérico	Número del formulario de inscripción	NN
3	ID_TERCERO	Numérico	Identificador único del aspirante	NN
4	TIP_IDENTIFICACION	Texto	Tipo de identificación	NN
5	NUM_IDENTIFICACION	Texto	Número de identificación	NN
6	NOMBRES	Texto	Nombres	NN
7	APELLIDOS	Texto	Apellidos	NN
8	COD_UNIDAD	Texto	Código del programa	NN
9	NOM_UNIDAD	Texto	Nombre del programa	NN
10	COD_PENSUM	Texto	Código del plan de estudios	NN
11	NOM_PENSUM	Texto	Nombre del plan de estudios	NN
12	MODALIDAD	Texto	Modalidad de estudios	NN
13	EST_FORMULARIO	Texto	Estado del formulario de inscripción	NN

14	FEC_PAGO	Fecha	Fecha de pago	
15	VAL_FORMULARIO	Numérico	Valor del formulario	
16	OBS_LIQUIDACION	Texto	Observación del pago	
17	COD_PERIODO	Texto	Código del periodo	NN
18	ORDEN	Numérico	Número de la orden asociada al recibo	PK
19	DOCUMENTO	Texto	Concepto del tipo de recibo	NN
20	CLIENTE	Numérico	Identificador único del aspirante en el sistema financiero	NN
21	FECHA_REGISTRO	Fecha	Fecha de creación de la orden	NN
22	FECHA_PAGO	Fecha	Fecha de pago de la orden	
23	FECHA_VENCIMIENTO	Fecha	Fecha de vencimiento de la orden	NN
24	VALOR_RECIBO	Numérico	Costo del recibo	NN
25	ESTADO_RECIBO	Texto	Estado del recibo	NN
26	ID_PERFIL	Numérico	Identificador único del perfil del usuario	PK
27	NOMBRE_PERFIL	Texto	Nombre del perfil del usuario	NN
28	ID_PERMISOS_PERFIL	Numérico	Identificador de los permisos para ese perfil	PK
29	DESCRIPCION_PERMISOS_PERFIL	Texto	Descripción de los permisos de ese perfil	NN
30	ID_PERMISOS_USUARIO	Numérico	Identificador del permiso asociado al usuario	PK
31	ID_USUARIO	Numérico	Identificador único del usuario	PK
32	IDENTIFICACION_USUARIO	Texto	Número de identificación del funcionario	NN
33	NOMBRES_USUARIO	Texto	Nombres del funcionario	NN
34	APELLIDOS_USUARIO	Texto	Apellidos del funcionario	NN
35	CORREO_USUARIO	Texto	Correo del funcionario	NN