



Universidad  
**Inca Garcilaso de la Vega**

**Nuevos Tiempos. Nuevas Ideas**

Facultad de Ingeniería de Sistemas, Cómputo y Telecomunicaciones

Desarrollo de una solución ETL para la Carga Masiva de Registros de la  
Superintendencia de Banca, Seguros y AFP para el Área de Sistemas  
Administrativos de una Entidad Financiera.

(Development of an ETL solution for the mass loading of records from the  
Superintendencia de Banca, Seguros y AFP for the Administrative Systems  
department of a financial institution.)

### **TRABAJO DE SUFICIENCIA PROFESIONAL**

Para optar el título profesional de Ingeniero de Sistema y Cómputo

#### **AUTOR**

Sánchez Timoteo, Jorge Luis

<https://orcid.org/0009-0008-3510-9862>

#### **ASESOR**

Dr. Hilario Falcon, Francisco Manuel

<https://orcid.org/0000-0003-3153-9343>

**Lima – Perú- 2025**

# Turnitin Informe de Originalidad

Procesado el: 13-ago-2025 8:51 p. m. -05

Identificador: 2699622888

Número de palabras: 11724

Entregado: 12

Trabajo de Suficiencia-Recortado-  
Jorge\_Sanchez\_Timoteo.docx Por Jorge Luis  
Sánchez Timoteo

Índice de similitud	Similitud según fuente
9%	Fuentes de Internet: 8% Publicaciones: 1% Trabajos de estudiante: 3%

Coincidencia del 1% ( )

[Trujillo Pulecio, Oscar Mauricio. "Construcción de un data Mart para el cálculo de indicadores de calidad del servicio en el área de gestión operativa de la CHEC S.A. E.S.P.", 'Facultad De Ingeniería Universidad Del Zulia', 2018](#)

Coincidencia del < 1% (Internet desde 24-ene-2024)

[https://alicia.concytec.gob.pe/vufind/Record/UUNI\\_af57b762b3bedf2a8a319f182b2f802e](https://alicia.concytec.gob.pe/vufind/Record/UUNI_af57b762b3bedf2a8a319f182b2f802e)

Coincidencia del < 1% (Internet desde 02-jun-2025)

[https://alicia.concytec.gob.pe/vufind/index.php/Record/UNAP\\_4a88461f7b82ef874d67f85826b857b2/Details](https://alicia.concytec.gob.pe/vufind/index.php/Record/UNAP_4a88461f7b82ef874d67f85826b857b2/Details)

Coincidencia del < 1% (Internet desde 13-may-2025)

[https://alicia.concytec.gob.pe/vufind/index.php/Record/UTPD\\_e9be618f866aedeeadd952321f54b0e](https://alicia.concytec.gob.pe/vufind/index.php/Record/UTPD_e9be618f866aedeeadd952321f54b0e)

Coincidencia del < 1% (Internet desde 24-abr-2025)

[https://alicia.concytec.gob.pe/vufind/Record/AUTO\\_c3be7f202fce5ce727eb4efe5fc3447c/Details](https://alicia.concytec.gob.pe/vufind/Record/AUTO_c3be7f202fce5ce727eb4efe5fc3447c/Details)

Coincidencia del < 1% (Internet desde 25-abr-2025)

[https://alicia.concytec.gob.pe/vufind/Record/REVURP\\_f71ddfe7ae7146d1f8f2102fb3da3510/Details](https://alicia.concytec.gob.pe/vufind/Record/REVURP_f71ddfe7ae7146d1f8f2102fb3da3510/Details)

Coincidencia del < 1% (Internet desde 15-abr-2024)

<http://repositorio.uigv.edu.pe/bitstream/handle/20.500.11818/8080/TSP-ARO%20DE%20PERLACIOS%20NANCY%20FLOR%20-RP.pdf?isAllowed=y&sequence=1>

Coincidencia del < 1% (Internet desde 06-may-2024)

<http://repositorio.uigv.edu.pe/bitstream/handle/20.500.11818/8140/TSP-MONTES%20CHAVEZ%20HENRY%20ALBERTO%20-%20RP.pdf?isAllowed=y&sequence=1>

Coincidencia del < 1% (Internet desde 05-may-2024)

<http://repositorio.uigv.edu.pe/bitstream/handle/20.500.11818/7997/TSP-ENRIQUEZ%20MENDOZA%20WALTER%20ANDREI-RP.pdf?isAllowed=y&sequence=1>

Coincidencia del < 1% (Internet desde 17-mar-2024)

<http://repositorio.uigv.edu.pe/bitstream/handle/20.500.11818/7910/TSP-ZAMBRANO%20ESCALANTE%20FAUSTO%20BENITO-RP.pdf?isAllowed=y&sequence=12>

Coincidencia del < 1% (Internet desde 13-may-2024)

<http://repositorio.uigv.edu.pe/bitstream/handle/20.500.11818/8198/TSP-COAGUILA%20ANDIA%20DAVID%20PEDRO%20-RP-.pdf?isAllowed=y&sequence=1>

Coincidencia del < 1% (Internet desde 03-feb-2024)

<http://repositorio.uigv.edu.pe/bitstream/handle/20.500.11818/6864/TSP%20PALOMINO%20LEVANO%20MELISA%20PAOLA%20rpo.pdf?isAllowed=y&sequence=18>

Coincidencia del < 1% (Internet desde 03-oct-2023)

<http://repositorio.uigv.edu.pe/bitstream/handle/20.500.11818/7296/TSP%20Vanessa%20Salas%20Loayza%20rp.pdf?isAllowed=y&sequence=1>

Coincidencia del < 1% (Internet desde 19-dic-2024)

<https://www.coursehero.com/file/200785673/AC-U3/>

Coincidencia del < 1% (Internet desde 11-sept-2023)

<https://www.coursehero.com/file/193912116/TAREA-1docx/>

Coincidencia del < 1% (Internet desde 21-abr-2025)

<https://www.coursehero.com/file/86665736/art03pdf/>

Coincidencia del < 1% (Internet desde 11-dic-2024)

<https://www.coursehero.com/file/229385875/2270418-1pdf/>

Coincidencia del < 1% (Internet desde 25-mar-2025)

<https://www.coursehero.com/file/68913458/temasdocx/>

Coincidencia del < 1% ( )

[Martín Gutiérrez, Pablo. "Data Warehouse : marco de calidad", 2012](#)

Coincidencia del < 1% ( )

## DEDICATORIA

*A mis padres y hermanos, por ser mi soporte constante, por cada gesto de amor, por creer en mí incluso en los momentos en que yo dudaba. Su confianza ha sido la base sólida sobre la que construí este camino. A mis amigos, compañeros de ruta en esta etapa, por las risas compartidas, el aliento en los días difíciles y las memorias que quedarán impresas en cada línea de este trabajo.*

*Gracias por ser parte de esta red de afectos y aprendizajes que me sostuvo y me impulsó a seguir adelante.*

## AGRADECIMIENTOS

En primer lugar, doy gracias a Dios, cuya guía y fortaleza han iluminado cada paso de este camino académico. A mis padres, por su amor incondicional y por enseñarme que con disciplina y entrega todo es posible; a mis hermanos, por su aliento constante y su alegría, que hicieron más livianas las jornadas de estudio. A mis compañeros de estudio, por las discusiones enriquecedoras y el apoyo mutuo en las largas horas de análisis y preparación. Extiendo mi gratitud a los docentes de la Facultad, cuyas enseñanzas y consejos han moldeado mi formación profesional. Finalmente, un reconocimiento especial a mi asesor de tesis, el Dr. Francisco Manuel Hilario Falcon, por su paciencia y su valiosa guía en esta fase de mi desarrollo profesional.



## RESUMEN Y PALABRAS CLAVE

Las entidades financieras del Perú deben transformar los voluminosos datos regulatorios que la Superintendencia de Banca, Seguros y AFP (SBS) les remite (créditos, desembolsos y morosidad).

Actualmente, los procesos de extracción, limpieza, normalización y carga de los datos masivos enviados por la Superintendencia de Banca, Seguros y AFP (SBS) se realizan de manera manual y con un software obsoleto e incompatible con los sistemas operativos actuales. Para superar estas limitaciones, en el presente trabajo se describe el desarrollo de una solución ETL que automatiza la ingesta de esos datos, optimizando su procesamiento y permitiendo al área de Sistemas Administrativos de una entidad financiera explotarlos eficientemente,

Para el desarrollo de esta solución se empleó la herramienta de Inteligencia de Negocios de Microsoft SQL Server Integration Services, como base de datos se utilizó el motor Microsoft SQL Server 2022, se implementó bajo la metodología Extreme Programming (XP) para un desarrollo ágil y adaptativo.

Las pruebas internas demuestran que el nuevo proceso ETL automatiza la carga masiva de registros de la SBS. reduciendo drásticamente los tiempos de ejecución frente al método anterior y elimina la dependencia de software obsoleto, entregando a los usuarios información confiable y actualizada de forma inmediata, ayudando así a la toma de decisiones.

Palabras clave: ETL, datos, mejora, análisis, filtrado, carga masiva, toma de decisión.

Development of an ETL solution for the mass loading of records from the Superintendencia de Banca, Seguros y AFP for the Administrative Systems department of a financial institution

## ABSTRACT AND KEYWORDS

Peruvian financial institutions must process the large volumes of regulatory data supplied by the Superintendencia de Banca, Seguros y AFP (SBS) (credit records, disbursements and delinquency metrics).

At present, the extraction, cleansing, normalization, and loading processes for the large volumes of data sent by the Superintendencia de Banca, Seguros y AFP (SBS) are performed manually using outdated software that is incompatible with modern operating systems. To overcome these limitations, this work describes the development of an ETL solution that automates the ingestion of this data, optimizes its processing, and enables the Administrative Systems department of a financial institution to leverage it efficiently.

For the development of this solution, the Microsoft SQL Server Integration Services Business Intelligence tool was used, the Microsoft SQL Server 2022 as the database engine, and the implementation was carried out following the Extreme Programming (XP) methodology to ensure an agile and adaptive development process.

Internal tests demonstrate that the new ETL process automates the bulk loading of SBS records, drastically reduces execution times compared to the previous method, and eliminates reliance on obsolete software, delivering reliable, up-to-date information to users instantly and thereby supporting decision-making.

Keywords: ETL, data, improvement, analysis, filtering, bulk loading, decision making.

## INDICE GENERAL

<b>DEDICATORIA</b>	<b>2</b>
<b>AGRADECIMIENTOS</b>	<b>3</b>
<b>RESUMEN Y PALABRAS CLAVE</b>	<b>4</b>
<b>ABSTRACT AND KEYWORDS</b>	<b>5</b>
<b>ÍNDICE DE FIGURAS</b>	<b>8</b>
<b>INDICE DE TABLAS</b>	<b>9</b>
<b>INTRODUCCIÓN</b>	<b>10</b>
<b>CAPÍTULO I: MARCO TEÓRICO DE LA INVESTIGACIÓN</b>	<b>12</b>
1.1. BASES TEÓRICAS	12
1.2. MARCO LEGAL	27
1.3. ANTECEDENTES DEL ESTUDIO	28
1.4. MARCO CONCEPTUAL	31
<b>CAPÍTULO II: PLANTEAMIENTO DEL PROBLEMA</b>	<b>33</b>
2.1. DESCRIPCIÓN DE LA REALIDAD PROBLEMÁTICA	33
2.2. FORMULACIÓN DEL PROBLEMA GENERAL Y ESPECIFICOS	35
2.3. OBJETIVOS GENERAL Y ESPECIFICOS	35
<b>CAPÍTULO III: JUSTIFICACIÓN Y DELIMITACIÓN DE LA INVESTIGACIÓN</b>	<b>37</b>
3.1. JUSTIFICACIÓN E IMPORTANCIA DEL ESTUDIO	37
3.2. DELIMITACIÓN DEL ESTUDIO	41
<b>CAPÍTULO IV: FORMULACIÓN DE DISEÑO</b>	<b>42</b>
4.1. DISEÑO ESQUEMÁTICO	42
4.2. DESCRIPCIÓN DE LOS ASPECTOS BÁSICOS DEL DISEÑO	47
4.3. METODOLOGÍA DE DESARROLLO - XP	51
<b>CAPÍTULO V: PRUEBA DE DISEÑO</b>	<b>53</b>
5.1. APLICACIÓN DE LA PROPUESTA DE SOLUCIÓN	53
<b>CONCLUSIONES</b>	<b>65</b>
<b>RECOMENDACIONES</b>	<b>66</b>
<b>REFERENCIAS BIBLIOGRÁFICAS</b>	<b>67</b>
<b>ANEXOS</b>	<b>71</b>
<b>ANEXO A: ESTRUCTURA PRINCIPAL DE LOS ARCHIVOS SBS DATOS CLIENTES Y REGISTRO SALDOS</b>	<b>71</b>
<b>ANEXO B: MODELO DE BASES DE DATOS</b>	<b>73</b>



## ÍNDICE DE FIGURAS

<b>Figura 1.</b> ETL es la opción más escogida para la migración de datos (Elaboración propia).	18
<b>Figura 2.</b> Modelo Bill Inmon (Elaboración propia).	22
<b>Figura 3.</b> Modelo de Ralph Kimbal (Elaboración propia).	23
<b>Figura 4.</b> Proceso ETL (Elaboración propia).	26
<b>Figura 5.</b> Proceso ELT (Elaboración propia).	26
<b>Figura 6.</b> Flujo de carga actual desarrollado en Microsoft SQL Server Enterprise Manager.	34
<b>Figura 7.</b> Diseño esquemático solución ETL (Elaboración propia).	43
<b>Figura 8.</b> Diagrama general de Fases Carga SBS (Elaboración propia).	46
<b>Figura 9.</b> Componentes de la solución ETL Carga SBS (Elaboración propia).	47
<b>Figura 10.</b> Modelo de Bases de datos (Elaboración propia).	78
<b>Figura 11.</b> Calificación Matching (Data modificada).	79
<b>Figura 12.</b> Ahorros no validos e Instituciones faltantes (Data modificada).	80
<b>Figura 13.</b> Estadísticas Carga Ahorros y SBS (Resultado del proceso ETL).	80
<b>Figura 14.</b> Estadísticas Matching y Finales (Resultado del proceso ETL).	81

## INDICE DE TABLAS

<i>Tabla 1</i>	<i>Listado de Tecnologías de desarrollo e implementación empleadas.</i>	44
<i>Tabla 2</i>	<i>Componentes de la Solución ETL en Microsoft SQL Server Integration Services.</i>	48
<i>Tabla 3</i>	<i>Variables de Medición para la solución ETL.</i>	50
<i>Tabla 4</i>	<i>Detalle de las tablas BN_SBS.</i>	54
<i>Tabla 5</i>	<i>Detalle de los componentes de la solución ETL según flujo de Fases.</i>	58
<i>Tabla 6</i>	<i>Métricas y resultados obtenidos (Entorno Desarrollo).</i>	63
<i>Tabla 7</i>	<i>Interpretación de la cadena que contiene los registros de los Datos del Cliente.</i>	71
<i>Tabla 8</i>	<i>Interpretación de la cadena que contiene los registros de los Saldos del Cliente.</i>	72



## INTRODUCCIÓN

En el Perú, la Superintendencia de Banca, Seguros y AFP (SBS) entrega mensualmente a las entidades financieras volúmenes masivos de información en formato de texto plano. Esta data incluye, por un lado, los datos generales de cada usuario y, por otro, sus saldos y movimientos crediticios. Aunque fundamental para la gestión del riesgo, la toma de decisiones y el cumplimiento regulatorio, el formato irregular y la ausencia de procesos de limpieza o filtrado provocan que muchas entidades carguen estos registros tal como vienen en sus bases de datos, arrastrando duplicados, campos mal formateados y registros obsoletos.

Este proyecto solicitado por una entidad financiera que requiere una solución tecnológica moderna para la carga de dichos datos, debido a que el proceso actual se realiza de manera manual, utilizando software y paquetes obsoletos, haciendo que la calidad de los datos no sea confiable y que el tiempo de proceso sea prolongado, lo cual afecta a las áreas que requieren de esa información de manera oportuna.

En este trabajo se detallará el desarrollo de una solución ETL utilizando las herramientas de Microsoft SQL BI Integration Services, indicando los procesos de Extracción, Transformación y Carga que otorgará al Área de Sistemas Administrativos una data de calidad, de manera rápida y eficaz, lo cual facilitará la toma de decisiones, el cumplimiento regulatorio y fortalezca su competitividad en el mercado financiero.

Este trabajo se conforma de 5 capítulos:

**Capítulo I:** Se presentan los puntos teóricos en la que se basa este trabajo, además del marco legal, los antecedentes de estudios que sirvieron también de apoyo para el desarrollo de la solución ETL.

**Capítulo II:** En este apartado se describe la realidad problemática por la que pasa la entidad financiera y se formula las cuestiones principales y específicas, además de indicar los objetivos que se han propuesto para el desarrollo de esta solución tecnológica.

**Capítulo III:** Aquí se aborda la justificación para el desarrollo de esta solución tecnológica, indicando su importancia y las limitaciones de este trabajo.

**Capítulo IV:** Comprende el diseño de la solución ETL para la entidad financiera, indicando los componentes principales, el software empleado para el desarrollo, además de la metodología que se empleó para su desarrollo.

**Capítulo V:** En este apartado se muestra la aplicación de la solución ETL, indicando más a detalle cada componente que constituye la solución, el modelo de datos, la preparación del entorno, la ejecución de la carga de datos, métricas y resultados obtenidos.



## CAPÍTULO I: MARCO TEÓRICO DE LA INVESTIGACIÓN

### 1.1. BASES TEÓRICAS

#### 1.1.1 Teoría de la información

Tal como asevera Pérez (2024) es una tesis propuesta por el matemático e ingeniero eléctrico Claude Shannon y el biólogo Warren Weaver en los últimos años de la década de 1940, la cual detalla el proceso y la medición de los datos cuando se transmite la información, esta tesis también se le conoce como la teoría matemática de la comunicación.

#### 1. Componentes:

Pérez (2024) afirma que el proceso de comunicación contiene las siguientes partes fundamentales:

- El Mensaje: es el conjunto de datos que envía el emisor al receptor, esta se transporta por medio de canales de comunicación y la cual se descifra conforme a las normas de un código específico
- El Emisor: es quien envía el mensaje, tal como indica esta teoría, se le conoce como la fuente de información.
- El Receptor: es quien toma recepción el mensaje
- El Ruido: se refiere a las causas que hacen que el mensaje llegue de manera incorrecta al receptor y no pueda comprenderlo.

#### 2. Utilidad de la teoría de la información:

Según Piero (2021), tiene su utilidad en lo relacionado con la comprensión de los datos transferidos, en el reconocimiento de factores o causas que puedan distorsionar o cancelar la información enviada a los destinatarios o receptores, así como también la decodificación, codificación y la velocidad de la información enviada, además de tener como objetivo primordial el encontrar la forma más eficaz que facilite la trasmisión de un mensaje sin alteraciones.

### 1.1.2 La Limpieza de Datos

Para Noles (2024) la limpieza de datos es una fase importante que se tiene que realizar antes de cualquier proceso de análisis de información. En la limpieza de información se detectan, corrigen y eliminan los registros que son incorrectos, los cuales podrían dar resultados distorsionados. Para ello se apoya en herramientas que garanticen que los datos no presenten fallos.

#### 1. La importancia de una buena limpieza de datos

Como indica Arregui (2024) es fundamental por las siguientes razones:

- La precisión del análisis.
- La confiabilidad de los modelos analíticos y predictivos.
- Un procesamiento más eficiente.
- La toma de decisiones basada en datos.

#### 2. Pasos para la limpieza de datos

- Remover los datos irrelevantes.
- Remover los datos duplicados.
- Reparar errores estructurarles de los datos.
- Atender datos faltantes o brechas de información.
- Validar que los datos estén correctos.

### 1.1.3 La Calidad de Datos

Según Consoli (2022), se refiere a los datos organizados ubicados en bases relacionales. Para realizar su medición e impulsar la calidad de la información hay muchos marcos de referencia; por ejemplo, la norma ISO/IEC 25012, que indica que un dato de calidad

tiene que cumplir lo siguiente.:

- Completitud.
- Unicidad.
- Oportunidad
- Validez.
- Precisión.
- Consistencia.

## 1. Importancia de la calidad de datos

Según Bintelligenz (2023), tener datos de calidad es primordial para la gestión eficaz de una empresa. Cuando los datos son precisos y fiables, las empresas pueden respaldar sus decisiones en bases robustas, pero si la información presenta defectos o están incompletos, puede darse conclusiones poco efectivas o equivocadas.

Las consecuencias de no tener datos de calidad en las empresas abarcan desde las pérdidas de ventas y rentabilidad, impresión en el análisis de las herramientas que trabajan con esos datos, genera mala reputación y que obtengan sanciones legales (London Consulting Group, 2024)

### 1.1.4 Bases de datos (Data base)

En su definición puede decirse que las bases de datos son repositorios estructurados donde se guarda la información y se gestionan dentro de un sistema informático. Su finalidad es el manejo de datos ya sean en grandes volúmenes estructurados como también los que no cuentan con una organización definida, dando mecanismos y herramientas que facilitan tanto su almacenamiento como su búsqueda y tratamiento (Iglesias, 2023).

## 1. Tipos de bases de datos

### a) Relacionales

Son las más comunes y se basan en la organización de los datos en tablas que representan diferentes tipos de entidades; cada una de ellas está comprendida por filas y columnas. Debido a esto los datos quedan organizados de tal forma que al momento de acceder a ellos resulta eficaz y flexible. Algunos motores de bases de datos de tipo relacional son Microsoft SQL Server, MySQL, Microsoft Azure SQL Server, PostgreSQL y MariaDB (Microsoft, 2021).

En general este tipo de base de datos usan el lenguaje SQL (Structured Query Language), como afirma Iglesias (2023) para realizar búsquedas y también para la manipulación de datos.

### b) No relacionales

Son las administran la información poco organizada o sin organización, sin acudir a la clásica disposición de filas y columnas como lo es en uno del tipo relacional. En vez esto, se ajustan a las características de cada tipo de dato. Este enfoque es ideal para la gestión de volúmenes masivos de datos distribuidos, actualizándolos y analizándolos de manera rápida. Algunos motores de bases de datos no relacionales: Azure Cosmos DB, MongoDB, Cassandra, Hbase y Redis (Microsoft, 2021).

## 2. Ventajas de las bases de datos

Según Piñeiro (2024) algunas ventajas son:

- Los datos son independientes a los tratamientos y viceversa.
- Los datos son consistentes siempre que se elimine o controlen su redundancia.
- La distribución de datos por todos los usuarios que tengan autorización.
- El valor informativo en su conjunto es mayor que la suma del valor informativo individual.

- Otorga un mejor acceso a los datos.
- La integridad de los datos mejora, esto se da gracias a las restricciones y reglas que se deben de cumplir.

### 3. Diseño de bases de datos

Como indica Postigo (2022) para el diseño de bases de datos se necesita pasar por las siguientes fases:

- La modelación de datos.
- El análisis de requisitos.
- Construcción del diseño conceptual.
- Construcción del diseño lógico.
- Construcción del diseño físico.
- Explotación de los datos.

Se pueden aplicar las buenas prácticas para la optimización de base de datos, tal como señalan Maquen, Terán, Castillo y Villón (2022), creación de índices en las tablas, el uso de columnas calculadas y de operadores correctos en las consultas que se realicen.

### 4. Lenguajes usados para la gestión de las bases de datos

Postigo (2022), señala los siguientes:

- **Lenguaje para la definición de datos (DDL)**, son los comandos para la construir objetos. Ejemplo: CREATE, DROP y ALTER.
- **Lenguaje para la manipulación de datos (DML)**, son los comandos utilizados para realizar operaciones. Ejemplo: UPDATE, INSERT y DELETE,
- **Lenguaje para el control de datos (DCL)**, son los comandos que se utilizan

para el control de acceso a los datos. Ejemplo: GRANT y REVOKE.

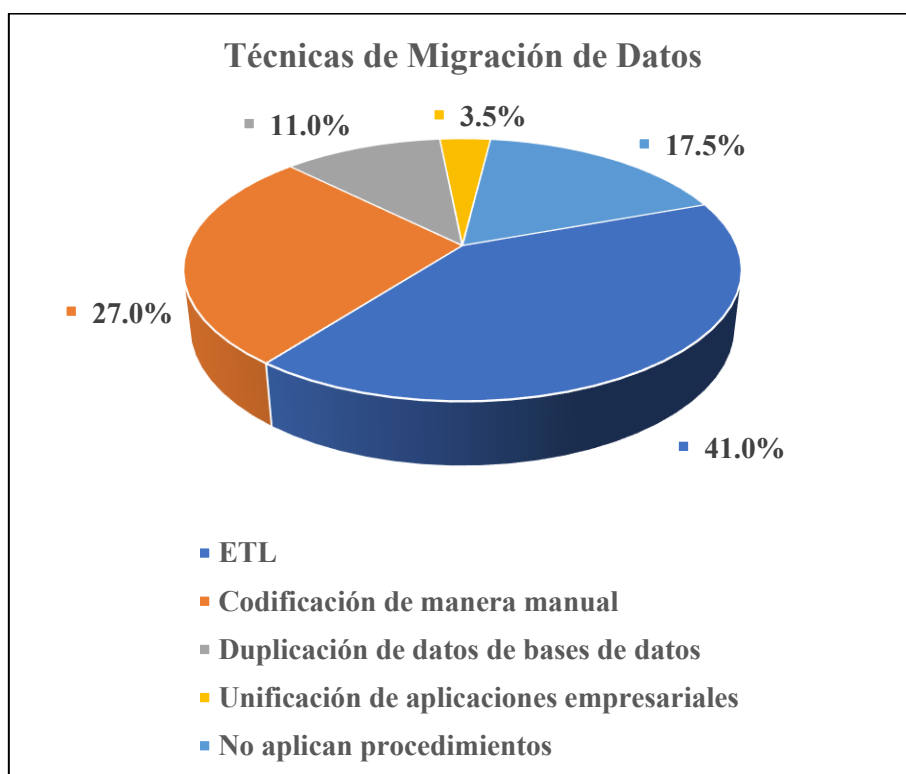
- **Lenguaje para el control de transacciones (TCL)**, son aquellos comandos que utilizan para gestionar la transacciones. Ejemplo: ROLLBACK, SET TRANSACTION Y COMMIT.

### 1.1.5 Migración de datos

Según Naeem (2025), se basa en mover la información a otros sistemas de almacenamiento, formato de archivo y estructuras de base de datos. Pero ese proceso no solo es la transferencia, también involucra el mapeo y la transformación de los datos de origen para adecuarlos a los requerimientos de destino cuando sus estructuras son diferentes.

#### 1. Técnicas de migración de datos

Según la encuesta realizada por TDWI (2025), citado por Naeem (2025), la técnica más empleada por las organizaciones es el ETL, con un 41% de aceptación. Esto se debe a el uso de herramientas ETL permite realizar los difíciles requisitos del proceso de migración. Las demás técnicas son la codificación manual con un 27%, la duplicación de datos entre bases de datos con un 11% y la unificación de aplicaciones empresariales 3.5%.



*Figura 1. ETL es la opción más escogida para la migración de datos (Elaboración propia).*

### 1.1.6 Business Intelligence / Inteligencia de Negocios

También conocida como inteligencia Empresarial como indica Oracle Corporation (2021), alude a las funcionalidades que permiten a las empresas mejorar su capacidad para tomar decisiones estratégicas y medidas fundamentadas, además de implementar procesos de negocio más eficientes. Entre sus funcionalidades permiten:

- La recopilación de datos.
- Mostrar los datos en formatos entendibles, como lo son gráficos y tablas.
- Entregar a los empleados los datos de una manera oportuna.

La inteligencia de negocios mantiene actualizado a las organizaciones y el éxito de ellas dependerá en mayor parte del conocimiento del mercado, que incluye saber el que, el quién, el cuándo, el dónde, el cómo y el por qué.

Las empresas que se basan en datos tienen como corazón la inteligencia de negocio, y esta tiene como objetivos finales el aumentar el impacto de una empresa u organización y volverla más eficiente.

Según las conclusiones a las que llegan Barón, García y Sánchez (2021), donde se aplica más la inteligencia de negocios y tecnologías relacionadas a estas, es en el sector empresarial, secundado por los sectores de salud, tecnología, educación y turismo. Y tal como indica Primak (2020), citado por Dos Santos y Gibertoni (2022), debido a que mejora la eficiencia en las tomas de decisiones de las empresas, el interés por la Inteligencia de Negocios ha crecido de manera significativa en estos últimos años.

## 1. Arquitectura de la Inteligencia de Negocios

Según Duque, Godinho y Vasconcelos (2021) quienes afirman que la arquitectura de la inteligencia de negocio está conformada por 4 capas que son:

- **Capa de Orígenes de Datos**, la cual se encuentran las fuentes de datos que pueden estar dentro o fuera de la organización, ser de diversos tipos y los datos son generados a partir del proceso operativo de la empresa y que se registran mediante los sistemas On-Line Transactional Processing (OLTP).
- **Capa de Movimientos de Datos (ETL)**, que posibilita la gestión de tratamiento de datos, eso involucra la extracción de diversas fuentes, la transformación que abarca la depuración y selección de los datos, finalizando por la carga de datos en estructuras dimensionales, las cuales podrán acceder los usuarios finales y aplicaciones del sistema.
- **Capa de Entorno Data Warehouse**, que es un almacén de datos estructurados, la cual se integra mediante el proceso ETL, y es la base fundamental para las tareas de estudio de datos. Su razón de ser es que se estructure la información de una manera multidimensional, logrando que sea

fácil su explotación.

- **Capa de Análisis del Negocio**, que permite el trabajar con el almacén de datos organizado, empleando muchas técnicas, como es la minería de datos y OLAP, con el fin de producir información relevante que respalde la toma de decisiones.

## 2. Beneficios de la aplicación de Inteligencia de negocios

Para Romagnano, Pantano y Becerra (2021) consideran los siguientes beneficios:

- Información precisa: Al aplicar BI se cuenta con la información sólida, fundamentada y bien organizada, la cual permite tomar acciones y decisiones en base al conocimiento generado por la misma empresa u órgano.
- Mejora la toma de decisiones: Aplicando BI permite un mejor aprovechamiento de los datos que son utilizados por la empresa, ya que otorga conclusiones sólidas para la toma de decisiones.
- Es una ayuda para el conocimiento del negocio: Se tiene un control óptimo con relación a las áreas que conforman la empresa. El tener la información centralizada para su explotación beneficia enormemente en lo relacionado con el tiempo y costos.
- Permite detectar residuos en el sistema: El aplicar BI permite detectar recursos que son irrelevantes y que no dan ningún valor agregado para las empresas.
- Facilita la consulta y el intercambio de información: Al centralizar la información por medio de BI, se logran resultados rápidos, a diferencia de otros sistemas que dependen de varios informes independientes.
- Baja los riesgos de cuello de botella: Una solución BI permite analizar los

datos de los clientes y en base a eso construir modelos predictivos que permitirían aumentar la rentabilidad de la empresa.

- Análisis en tiempo real: Con BI se cuenta con los datos centralizados y bien estructurados, de tal forma que su consulta y análisis es rápida, ahorrando tiempo.

### 1.1.7 Metodologías de Inteligencia de Negocios

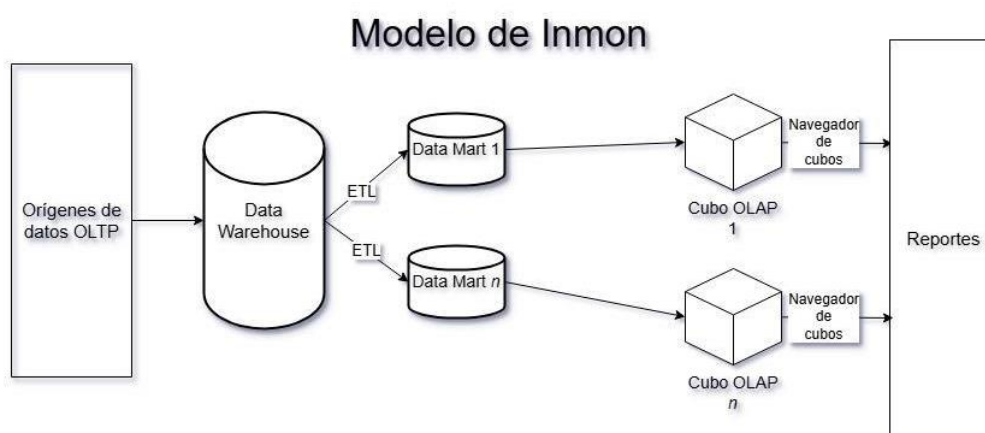
#### 1. Metodología de Bill Inmon:

Esta metodología propone un modelo que se basa en la empresa, con una arquitectura de muchos niveles, los Data Warehouse se alimentan con datos a un nivel atómico, y los Data Marts son dependientes.

Inmon aboga por una metodología del tipo Top-Down (Descendente) y considera mejor los datos corporativos. Luego de finalizar el Data Warehouse de una organización, los Data Marts se genera.

#### Características:

- Enfoque dirigido a temas: La información relacionada aun mismo tema tienen que organizarse de una manera que queden juntos entre si dentro del Data Warehouse.
- No presenta volatibilidad: la información no se elimina después su almacenamiento, convirtiéndose en un recurso de solo lectura la cual permanece accesible para consultas futuras.
- Integración: Las bases de datos tiene que ser consistentes.
- Variación en el tiempo: Los cambios realizados en los datos quedan guardados para que los reportes generados evidencien los cambios ocurridos (Guerrero & Rodas, 2022).



*Figura 2. Modelo Bill Inmon (Elaboración propia).*

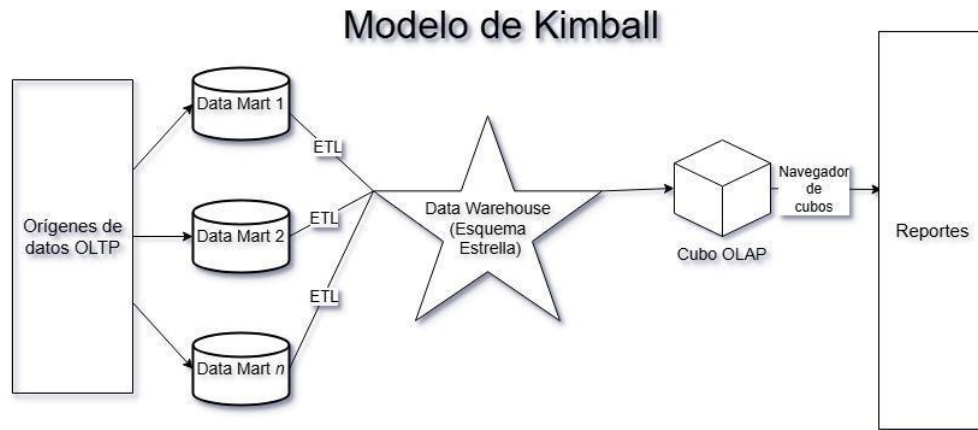
## 2. Metodología de Ralph Kimball:

Según Guerrero y Rodas-Silva (2022), esta metodología propuesta por Ralph Kimball, se basa en un esquema estrella, que con una arquitectura que se caracteriza en Data Marts y áreas de interés, que pueblan con datos atómicos resumizados a los Data Warehouse, siendo de un enfoque Bottom-Top (ascendente), que otorga una construcción de los almacenes de manera escalonada, incorpora artefactos y fases para lograr mejores resultados: análisis de requerimientos, sistemas transaccionales, modelo ETL, procesos ETL, el desempeño, soporte del Data Warehouse.

### Características:

- Enfoque dirigido al negocio: Comienza con identificar los requerimientos del negocio, con su valor relacionado, e implementar relaciones solidas con el negocio, utilizando esos esfuerzos.
- Una solución completa: Ofrecer todo lo necesario para que los usuarios del negocio consigan valor; como primer paso calve es tener un almacén de datos robusto, estructurado correctamente, de fácil acceso y con calidad.
- Infraestructura de información: Se busca establecer una base única de datos que sea combine la facilidad de uso con un alto desempeño, integrada el cual muestre la amplia variedad de requerimientos de negocio que fueron detectados en la organización.

- Hacer entregas por partes que aporten valor: Generar Data Warehouse en plazos de 6 a 12 meses. Para calcular el orden de aplicación de los incrementos se utiliza el valor de negocio de cada objeto detectado.



*Figura 3. Modelo de Ralph Kimball (Elaboración propia).*

### 1.1.8 Big Data

Según Arocha, Mejía y Suárez (2021), Big Data es el término más empleado para referirse a todo lo relacionado a los modelos de negocio, procesos y tecnologías en las que están basados los activos de información o datos, las cuales no pueden estudiarse por herramientas tradicionales, y que exigen a las organizaciones, buscar soluciones creativas y novedosas para la mejora de tomas de decisión.

#### 1. Características más notables

- Alto Volumen.
- Variedad.
- Velocidad.

En estos tiempos todo lo que se relaciona con Big Data, ya no se considera una tendencia, sino más bien una realidad que causa cambios radicales en todas las empresas del mundo, y será un factor decisivo para estas, el poder estudiar los volúmenes masivos de información de una forma rápida y adaptable.

A través del Big data se puede capturar, estudiar y gestionar la información con una mayor efectividad, por ejemplo, en el sector financiero, se recalca los relacionado al análisis de clientes, efectividad operativa, Estudio de riesgo y fraude, estudio de Internet, redes sociales y dispositivos móviles, nuevos modelos de negocios.

## 2. Procesamiento de datos y Big Data

Para Malla, López, Arévalo y Tóala (2023), dicho procesamiento comprende las etapas de recopilación de información, limpieza de datos, transformación y análisis de la información que procede de diversas fuentes; y esto está vinculado con Big Data, debido a los inmensos volúmenes de información, las empresas buscan nuevos métodos para poder explotarlas.

### 1.1.9 ETL (Extracción, Transformación y Carga)

Para Poppy (2023), ETL consiste en la recopilación de datos de orígenes diversos, transformarlos para darle una estructura en específica, según el destino y luego cargarlos en un almacén de datos centralizado.

#### 1. Procesos de ETL

IBM (2021) indica que el ETL consisten en 3 procesos que son:

##### a) Extracción (Extraction)

Durante la extracción de datos que no han sido procesados, se exportan o copian desde las ubicaciones de origen hacia un lugar temporal de almacenamiento. Las múltiples fuentes, tanto estructuradas como la que no lo están, son extraídas por los equipos que gestionan los datos, lo tipos de datos de datos pueden ser:

- Servidores SQL, NoSQL.
- Sistemas ERP y CRM.
- Archivos planos.

- XML, JSON.
- Páginas Web.
- Emails.

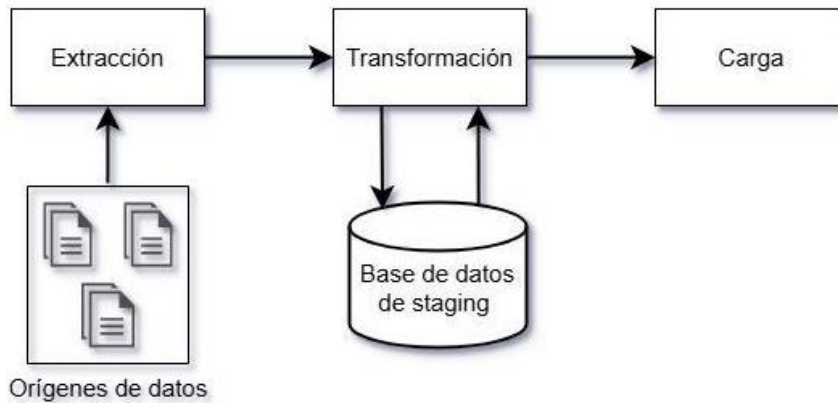
## **b) Transformación (Transformation)**

Los datos no procesados se procesan en el área de preparación. Para su uso analítico previsto se transforman y consolidan, este proceso puede incluir:

- Filtrado, limpieza, validación desduplicado y autenticación de datos.
- Se realizan auditorias para asegurar el cumplimiento y la calidad de los datos, además de calcular métricas.
- Realizar, calculo, resúmenes o traslaciones que se basan en los datos no procesados. Esto posiblemente tenga incluido la modificación de las cabeceras de columnas y filas con el objeto de tener coherencia, transformaciones como la traslación de monedas, unidades de medida o la edición de textos, etc.
- El formateo de los datos en tablas con el fin de que coincidan con la estructura del almacén de datos final.
- La eliminación, cifrado y protección de datos regulados por las autoridades gubernamentales y sectoriales.

## **c) Carga (Load)**

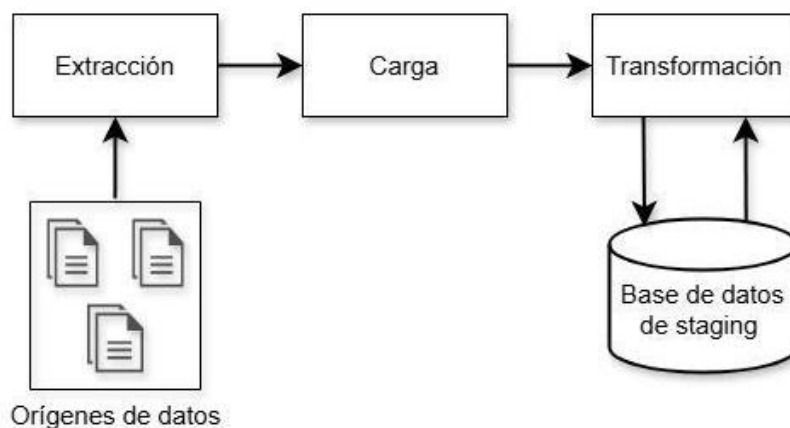
En este proceso, los datos que fueron transformados, filtrados y validados se envían del área de almacenamiento temporal a un repositorio de datos de destino. Habitualmente, que conlleva la carga inicial de los datos, que es secundada por la carga periódica de cambios incrementales, seguida, con poca frecuencia, de las actualizaciones para el borrado y reemplazo de los datos en el almacén. Generalmente el proceso ETL se desarrolla fuera del horario de laboral, cuando la actividad es baja en los sistemas de origen y en el almacén de datos.



**Figura 4.** *Proceso ETL (Elaboración propia).*

## 2. ETL y ELT diferencias

Según Zivkovic (2023), el modelo ELT invierte el orden que se conoce normalmente en un ETL. Primero se extraen los datos de sus respectivas fuentes y se cargan tal como están, en el almacenamiento de destino. Luego aplican la transformación de los datos. La ventaja es que mantienen la data original, eso incluye los errores e información faltante, evitando que se vuelva a configurar los complicados flujos ETL cada que cambian las reglas de transformación.



**Figura 5.** *Proceso ELT (Elaboración propia).*

## 3. Buenas prácticas en el proceso ETL

Banoth, Balasubramaniam, Kumar y Singh (2024) consideran que es clave la

planificación de un proyecto ETL desde el comienzo, definiendo detalladamente cómo circularan los datos y cómo se tendrá control del flujo de trabajo, para mantener el buen desempeño y así poder a futuro ampliar el sistema, de igual manera, incluir mecanismos de captura y control de errores, auditorías en plena ejecución y gestión de los metadatos que refuerce la estabilidad y la calidad de información.

## 1.2. MARCO LEGAL

Para la implementación del proceso ETL que extraerá, filtrará y estandarizará los datos que son otorgados a las entidades financieras peruanas por la Superintendencia de Banca, Seguros y AFP (SBS) el siguiente instrumento normativo se toma en consideración:

### **Reglamento de la Ley N° 29733, Ley de Protección de Datos Personales (Decreto Supremo N° 016-2024-JUS, 30 de noviembre de 2024)**

Reglamento que reemplaza al anterior (D.S. N° 003-2013-JUS) que entró en vigor desde el 30 marzo del 2025, el cual establece medidas más estrictas con relación a la organización y seguridad técnica para el manejo de datos personales.

1. Consentimiento: basado en los Artículos del 1-10 donde los datos solo serán para uso interno de la entidad financiera y siempre con los debidos permisos dados por los dueños de dichos datos, eso abarca las calificaciones crediticias, salud, etc.
2. Limitaciones al consentimiento: en base al Artículo 11 Se debe respetar la información en base a los principios de Ley y Reglamento; se respeta la información otorgada por la SBS y no se deberá modificar, salvo sea por correcciones indicadas por la misma entidad financiera.
3. Transferencia: en los artículos 12, 16, nos indica sobre la transferencia de información requiere consentimiento, incluso dentro de la misma organización que tenga los datos.

4. Medidas de seguridad dados en los artículos 46 al 56, sobre el acceso a la data, con respecto al desarrollo de la solución la data se trabaja dentro de las instalaciones de la entidad financiera, evitando la salida de datos por medios extraíbles.

### 1.3. ANTECEDENTES DEL ESTUDIO

#### 1.3.1. Internacionales

##### 1. Refinamiento del proceso ETL para la ingesta automatizada de datos financieros en un Data Lake (Rubio Gómez, J., 2024)

El autor en su tesis nos muestra como diseña e implementa la lógica de transformación ETL con el fin de cargar automáticamente los datos financieros en un Data Lake. Examinando diferentes estrategias y metodologías con objetivo de optimizar la calidad y el rendimiento del proceso mencionado, otorgando una visión global de las tecnologías y practicas más utilizadas. Así mismo, como ejecutarán el análisis y las pruebas extensas con el propósito de validar tanto la eficacia como la exactitud del algoritmo de transformación.

##### 2. Inteligencia de Negocios para la toma de decisiones estratégicas en una microempresa (Ortega Maldonado, G. L., 2024)

En estudio el autor tuvo como objetivo dotar a una empresa emergente, dedicada a la venta de artículos para bicicletas, de una solución BI (Inteligencia de Negocios) aplicando las herramientas ETL para análisis y exhibición de datos para fortalecer el proceso decisional. La eficiencia e inmediatez con respecto a la visualización de la data, hace que sea fácil la toma de decisiones estratégicas. Demostrando que el uso de BI como herramienta tiene como fin mejorar el desempeño de microempresas en el mercado.

##### 3. Módulo gerencial aplicando Business Intelligence (BI) para la Cooperativa de Ahorro y Crédito Mushuc Runa (Torres Bustos, A. D., 2021)

En este trabajo el autor nos comenta el desarrollo de su trabajo en la Cooperativa de Ahorro y Crédito Mushuc Runa, nos indica que actualmente emplea un módulo gerencia básica que no facilita el análisis de indicadores claves y dependen de archivos de hojas de cálculo, lo cual retrasa la generación de reportes. Por ello, esta investigación propuso el diseñar un módulo BI usando Pentaho Community para los procesos ETL, creación de cubos dimensionales, creación de informes, etc. El proyecto se baso en la metodología SCRUM y que además se utilizó Hefesto para la construcción de Data Marts en PostgreSQL. También se documentaron con notación BPMN los flujos ETL para hacer los procesos más comprensibles para los usuarios.

### 1.3.2. Nacionales

#### 1. Inteligencia de negocios para la toma de decisiones en una entidad financiera (Quispe Ponce, R., 2023)

El autor en este estudio analiza la toma de decisiones que son influenciadas de la Inteligencia de Negocios en una entidad financiera, que contempla sus procesos, su gestión de riesgos y alternativas estratégicas. Dicho estudio se toma como un proyecto de tecnología aplicada y explicativa, aplicando un enfoque inductivo-deductivo y un diseño experimental. En sus resultados se observan que, al emplear BI, la entidad financiera consigue la información de forma detallada y rápida para la toma de decisiones diarias, también optimiza sus reportes operativos y mejora sus estrategias internas para mantenerse en competitividad.

#### 2. Procesos ETL para generar información de créditos financieros como recurso para la toma de decisiones del área operativa de una entidad financiera, Lima, 2021 (Grande Flores, C. A., 2023)

El autor en esta tesis muestra los procesos ETL desarrollados en el área operativa de una entidad financiera, mejorando de una manera muy notable la generación de reportes crediticios, los flujos de caja, reportes operativos, estados de cuenta y transacciones. Aplicando la metodología mixta que se basa en las mejores prácticas del ciclo Russom, los procesos ETL dieron como resultado un aumento en la

velocidad en la generación de reportes, con una mayor precisión y calidad de los datos, presentando también una baja de costos en el manejo de la información. Por ello los procesos ETL dieron un impacto positivo, puesto que optimizó la creación de los informes y mejorando las decisiones de la entidad.

### **3. Implementación de data perfilada de clientes para la emisión de créditos en una entidad financiera (Miranda Mendoza, A. A., 2023)**

En este trabajo realizado en la Caja municipal de Ahorro y Crédito de Ica, consistió en el desarrollo de un ETL, denominado “Data Perfilada”, cuyo fin es identificar y clasificar los clientes especiales que tengan potencial de crédito. Para conseguirlo se procede a extraer los datos mensuales de crédito, se realizan los cálculos y se efectúan las reglas de negocio, dando como resultado la lista de clientes aptos para una preaprobación de crédito. Para lograr que el proceso se más rápido y fácil, se realizó la automatización por medio de ETL, y el almacenamiento se realizó en un servidor con medidas de seguridad robustas y un repositorio en Azure para la gestión de cambios.

### **4. Implementación de Business Intelligence para mejorar la toma de decisiones en el área de ventas de la empresa La Sangu (Saucedo Ruiz, J. R., 2022)**

El autor de esta tesis detectó múltiples problemas complican a la alta administración al momento de necesitar información precisa y detallada dentro de área de ventas. En base a esto procedieron con el método de Ralph Kimball, Que brinda una mejor comprensión de la información gracias al desarrollo de reportes, predicciones de ventas y apoyo en la optimización de decisiones estratégicas para la alta dirección. Se inició el desarrollo empleando Microsoft SQL Server 2018 Business Intelligence, en el sitio en el que está localizado el repositorio de datos de la empresa, se empleó Visual Studio 2017 para la implementación de los procesos ETL, además de la fabricación del cubo tipo OLAP que concederá un estudio confiable y exacto en las futuras tomas de decisión de la organización. Por último, conseguirse el uso estratégico de la información con la generación de informes específicos y gráficos de ventas por área, que serán útiles para las tomas de decisión de la empresa.

#### 1.4. MARCO CONCEPTUAL

1. **Datos:** Son cualquier tipo de información que registra y guarda relacionada a una entidad (persona, objetos, lugares, etc.) junto con sus características (atributos), todo lo que se recopila y se guarda para describir algo se le considera dato (Microsoft, 2021).
2. **Metadatos:** Son los datos que sirven de referencia o explicación de un dato, podría decirse también que es la información sobre los datos y son importantes dependiendo del contexto en las cuales se utilicen (Vargas, Velasco, Acosta & Camelo, 2021).
3. **Modelo de datos:** Son abstracciones conceptuales que representan las reglas, objetos y vínculos de un contexto concreto. A partir de un conjunto de requerimientos informacionales, se elabora un modelo lógico que orienta el diseño de un repositorio de datos. En este esquema, cada clase de objeto se define por sus propiedades y cada registro de datos corresponde a la asignación de valores concretos a dichas propiedades (Postigo, 2022).
4. **Orígenes de Datos:** También conocidos como Data Source, son fuentes de datos en bruto que utilizan las organizaciones, las cuales pueden ser internas o externas, pero con la característica de ser relevantes para dichas organizaciones, además de ser de diferentes tipos, formatos y estructuras (Vargas, Velasco, Acosta & Camelo, 2021).
5. **Sistema de gestión de base de datos:** Se denomina al sistema que otorga a las empresas almacenar su información, generalmente cuentan con una interfaz amigable, además proveen de seguridad por medio de roles y controles de acceso, utilizan lenguajes de programación estandarizados y avanzados (Vargas, Velasco, Acosta & Camelo, 2021).
6. **SSIS:** SQL Server Integration Services (SSIS) es la plataforma de Microsoft SQL Server, que se dedica a las tareas de integración de datos, permitiendo diseñar y

automatizar flujos ETL mediante un entorno visual con componentes preconfigurados y opciones de personalización, gracias a su robustez y flexibilidad, SSIS simplifica el traslado, la limpieza y la transformación de los datos, fortaleciendo la gestión y calidad de los datos en las organizaciones (Banoth, Balasubramaniam, Kumar & Singh, 2024).

7. **Toma de decisiones:** Es el proceso de apoyo a la gestión estratégica que se basa en flujos herramientas BI, por ejemplo, soluciones ETL para extraer, depurar y cargar datos en un Data Warehouse, de modo que se disponga de información limpia y estructurada que faciliten y fortalezcan las decisiones fundamentadas (Dos Santos & Gibertoni, 2022).



## CAPÍTULO II: PLANTEAMIENTO DEL PROBLEMA

### 2.1. DESCRIPCIÓN DE LA REALIDAD PROBLEMÁTICA

En Perú cada mes, la Superintendencia de Banca, Seguros y AFP (SBS), entrega a las entidades financieras, información crediticia de los usuarios, para su debida explotación. Dicha información es enviada generalmente en formato de texto plano, contiene aproximadamente 81 millones de registros, las cuales son los datos generales de cada usuario con sus correspondientes datos crediticios.

Dichos datos suelen estar internamente distribuidos en formatos de cadena irregular para el caso de los datos generales del usuario, y en formato de cadena regular para los datos crediticios. Pero estos no están filtrados, es decir, incluye datos de personas que ya no están vivas, tipo de documentos que ya no son vigentes (Libreta Electoral), caracteres extraños, espacios vacíos, etc.

Las entidades financieras para explotar debidamente esa información, y tomar las decisiones más efectivas para mantenerse en competitividad en el mercado financiero, requieren de la extracción de los datos, su limpieza, transformación y carga en sus almacenes de datos, los cuales serán la base para otros procesos internos. Dicho esto, cada entidad financiera explota la información según los requerimientos de cada área que las conforma.

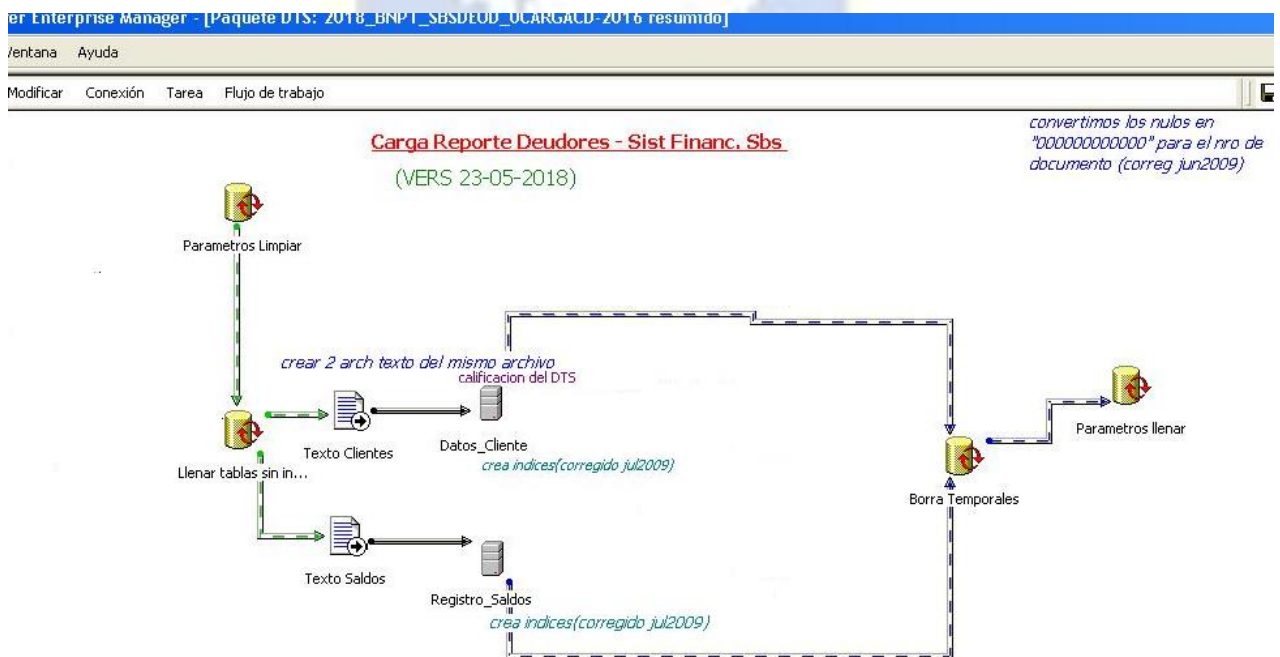
En este caso la entidad financiera es del tipo bancario y su área de Sistemas Administrativos realiza el tratamiento de dichos datos, cargándola tal cual, en base de datos Microsoft SQL Server 2008 R2, bajo el sistema operativo Microsoft Windows Server 2008 R2, las cuales ya no reciben actualizaciones, ni soporte.

Bajo esta situación, los datos pasan por un proceso de carga, pero no se realiza la limpieza y la filtración de datos, lo cual genera que dicho proceso tome más tiempo, la transformación directamente se realiza en el repositorio final, es decir, en aquellas bases de datos usadas para la explotación y el estudio de los datos, lo cual puede potencialmente incurrir a errores de resultados en las consultas, tales como información incompleta, equivocada, desactualizada, etc. Generando desconfianza en el área usuaria.

El área de Sistemas Administrativos tiene un proceso que es realizado generalmente en la madrugada, y consiste en ejecutar un paquete DTS que fue desarrollado con el software Microsoft SQL Server Enterprise Manager, que comienza con la carga de la información de la SBS al repositorio de datos BN\_SBS en dos bloques principales que son Datos Generales del Cliente y Datos de Saldos del Cliente, y se realiza de manera secuencial, luego de esa etapa, se lleva a cabo la limpieza y filtrado básico de los datos, para después dejar listo los datos para su explotación directa (para aplicativos antiguos) y su posterior migración a bases de datos más modernas (Microsoft SQL Server 2022).

Debido al uso de software obsoleto hace que prolongue el proceso de carga, además de mantener la dependencia de tecnologías desactualizadas.

Y tal como se mencionó, la limpieza y el filtrado son básicos, obligando a se vuelva a analizar los datos con intención de para evitar usar datos incompletos, o innecesarios, esto como indica Mucci (2024), descuida la calidad de los datos, dando lugar a análisis inexactos y una toma de decisiones incorrectas.



**Figura 6.** Flujo de carga actual desarrollado en Microsoft SQL Server Enterprise Manager.

## 2.2. FORMULACIÓN DEL PROBLEMA GENERAL Y ESPECIFICOS

### 2.2.1. Problema General

¿Cómo incide la ausencia de una solución ETL moderna en la eficiencia, calidad y confiabilidad en el proceso de la carga masiva de los registros enviados por la Superintendencia de Banca, Seguros y AFP (SBS) en el Área de Sistemas Administrativos de una Entidad Financiera?

### 2.2.2. Problemas Específicos

- 1) ¿En qué grado afecta la ausencia de una solución ETL moderna en el proceso actual de carga masiva de datos de la SBS, debido al uso de tecnologías obsoletas?
- 2) ¿Cómo afecta la ausencia de una solución ETL moderna con respecto a la limpieza y filtrado de los registros recibidos por la SBS a la calidad de los datos utilizados por el Área de Sistemas Administrativos?
- 3) ¿Qué riesgos se toman al no contar con una solución ETL moderna en la explotación de los datos, relacionados con la integridad y confiabilidad de estos?
- 4) ¿Cuáles son los beneficios de contar con una solución ETL moderna para el tratamiento de los registros de la SBS?

## 2.3. OBJETIVOS GENERAL Y ESPECIFICOS

### 2.3.1. Objetivo General

El objetivo principal de este estudio es el desarrollo de una solución ETL que permita mejorar el proceso de carga masiva de los registros enviados por la Superintendencia de Banca, Seguros y AFP (SBS), optimizando la eficiencia, limpieza y confiabilidad de los datos en el Área de Sistemas Administrativos de una Entidad Financiera.

### 2.3.2. Objetivo Específicos

- 1) Analizar el proceso actual de carga de datos de la SBS y ver sus limitaciones con

respecto al tiempo de ejecución y el uso de tecnologías obsoletas.

- 2) Detectar los problemas de calidad de datos causados por la ausencia de un proceso adecuado de limpieza y filtrado en los registros proporcionados por la SBS.
- 3) Evaluar cómo afecta la ausencia de una solución ETL moderna en la calidad de datos y su confiabilidad.
- 4) Diseñar e implementar una solución ETL moderna que automatice obtención, conversión y almacenamiento de datos proporcionados por la SBS, garantizando su eficiencia y calidad.



## CAPÍTULO III: JUSTIFICACIÓN Y DELIMITACIÓN DE LA INVESTIGACIÓN

### 3.1. JUSTIFICACIÓN E IMPORTANCIA DEL ESTUDIO

#### 3.1.1. Justificación del Estudio

En el actual trabajo que surge como solución a la necesidad de optimizar el proceso de carga y tratamiento de los datos entregados mensualmente por la Superintendencia de Banca, Seguros y AFP (SBS), los cuales son la base para las operaciones y toma de decisiones de las entidades financieras del Perú. En estos tiempos, el empleo de recursos tecnológicos obsoletos, como también la falta de procesos óptimos para el tratamiento de los datos, afecta considerablemente la eficiencia operativa y la confiabilidad de la información analizada por las áreas usuarias.

Como se indicó en la problemática, para el Área de sistemas Administrativos, el procesamiento de estos datos se realiza utilizando tecnologías ya obsoletas, como son como Microsoft SQL Server 2008 R2 y DTS, que no cuentan con soporte oficial ni actualizaciones. Esto da lugar a tiempos prolongados en la ejecución del proceso de carga, además de los riesgos relacionados a la transformación e integración de los datos.

En base a esto se propone una solución ETL moderna, que permitirá optimizar la carga masiva de registros, automatizando las actividades críticas de validación, limpieza y transformación de los datos antes de almacenarse. Esto se traduciría en la reducción del tiempo de procesamiento y la carga operativa del Área de Sistemas Administrativos; además de que optimizará la integridad y la calidad de los datos, para su uso en los demás sistemas.

Este trabajo se justifica bajo la necesidad crítica de modernización tecnológica y la mejora en el tratamiento de los datos, con un impacto directo en eficiencia operativa del Área de Sistemas Administrativos.

Además, al tratarse de una problemática común en múltiples instituciones financieras del Perú, los resultados alcanzados pueden ser empleados como material de consulta para el desarrollo de soluciones similares en otros tipos de contextos, generando un aporte al fortalecimiento general de la infraestructura de datos en el sector financiero.

### 3.1.2. Importancia del Estudio

Su importancia yace en su impacto multidimensional: técnico, operativo, organizacional y académico.

Desde el punto de vista técnico, este trabajo propone una solución concreta que reemplaza un proceso crítico obsoleto por uno actual, basándose en las buenas prácticas de ingeniería de datos y plataformas actualizadas. Esto reduciría el tiempo y los recursos requeridos para la preparación de los datos, y a su vez mejoraría la precisión y confiabilidad de estos.

Desde un enfoque operativo, este trabajo contribuirá a mejorar y fortalecer los procesos internos que lleva a cabo el área de Sistemas Administrativos, permitiéndole contar con datos correctos en un menor tiempo. Esto facilitará el trabajo de las áreas usuarias, quienes podrán tomar mejores decisiones y con mayor confianza en los datos disponibles.

Desde un punto de vista organizacional, este trabajo apunta a reducir la dependencia de tecnologías que ya no cuentan con soporte, reduciendo los riesgos de seguridad, mantenimiento y escalabilidad, lo cual se alinea con las estrategias de transformación digital que actualmente están incorporando las entidades financieras.

Para finalizar, desde la perspectiva académica y profesional, este trabajo realiza un aporte al campo de la ingeniería de sistemas al presentar un caso práctico de modernización tecnológica en el procesamiento masivo de datos críticos. Sirve como base para los futuros trabajos de investigación relacionadas con el ETL, calidad de datos, arquitectura de la información y análisis en tiempo real.

### 3.1.3. ¿Quiénes son los beneficiados?

El desarrollo de una solución ETL moderna traerá beneficios directos e indirecto a diferentes actores dentro y fuera de la entidad financiera, dichos beneficiados se hallan:

- **Área de Sistemas Administrativos:** Es la principal beneficiaria, ya que se optimizará sus procesos de carga y transformación de datos, bajando los tiempos operativos, cargas manuales y riesgos de error. Poseerá con una herramienta

más robusta automatizable y mantenible.

- **Áreas usuarias de información:** Tendrán acceso a datos más limpios, consistentes y oportuno, lo que aumentará la calidad del análisis y la confianza de los reportes que respaldan la toma de decisiones.
- **La propia Entidad Financiera:** Al optimizar la calidad de los datos y mejorar el manejo de la información crítica, se fortalecerá su capacidad de respuesta, cumplimiento regulatorio y competitividad en el mercado.
- **Otros profesionales y equipos TI:** Los cuales podrán reutilizar, o adaptar la solución ETL a sus propias necesidades dentro de ecosistema tecnológico de la entidad o incluso en otras similares del sistema financiero.
- **Comunidad académica y técnica:** Que hallará en este trabajo un caso real sobre la modernización de procesos ETL, útil para posteriores estudios, desarrollo de mejores prácticas y validación de modelos tecnológicos.

#### 3.1.4. Alcances del Estudio

El alcance de este trabajo se limita a los procesos de carga, limpieza, transformación y preparación de los registros enviados por la Superintendencia de Banca, Seguros y AFP (SBS), específicamente a relacionado con los siguientes bloques:

- Carga Calificación Ahorros.
- Carga Comparación Cuentas de Ahorro.
- Carga Datos del Cliente.
- Carga Registros de Saldos del Cliente.
- Carga Calificación Matching (Clientes vs Saldos).
- Informe de Comparaciones (Registros no existentes cuentas contables, instituciones).
- Informes de Estadísticas de Carga, Matching y Log.

- Informes de Calificaciones.

No incluye los procesos posteriores de explotación a análisis en otras áreas, ni contempla la migración total de la infraestructura tecnológica, aunque si sienta las bases para futuras mejoras en esos ámbitos.

### 3.1.5. Impacto del Estudio

El presente trabajo tiene un impacto directo en la entidad Financiera objeto de análisis, específicamente en el Área de Sistemas Administrativos, que enfrenta dificultades operativas en el tratamiento de los datos enviados por la Superintendencia de Banca, seguros y AFP (SBS).

- **Impacto Tecnológico dentro de la Entidad Financiera:** El desarrollo de una solución ETL moderna representa un cambio clave desde un entorno tecnológico desfasado, hacia uno que es actualizado, escalable y seguro. Esta mejora bajará notablemente los riesgos relacionados al uso de plataforma sin soporte, facilitará la integración con sistemas de análisis más actuales y otorgará una mejor gestión de los recursos tecnológicos del área.
- **Impacto Operativo en el Área de Sistemas Administrativos:** Con una nueva solución ETL, el Área de Sistemas Administrativos podrá automatizar y organizar correctamente estos procesos, lo que se traducirá en una reducción considerable de tiempo de procesamiento, menor carga de trabajo técnico y mayor control sobre la calidad de datos cargados.
- **Impacto en la calidad y veracidad de los datos:** La automatización del filtrado y la transformación asegura que la información que llega a las bases de explotación este depurada, actualizada y que sea confiable. Lo cual es fundamental para que las áreas usuarias, puedan basar sus decisiones en información correcta y oportuna, sin tener que volver a reprocesar o validar.
- **Impacto Organizacional a nivel Institucional:** Esta iniciativa se alinea con la visión de transformación digital. Disminuir la dependencia de soluciones legadas fortalece la capacidad operativa de la Entidad Financiera, mejora su

robustez tecnológica y abre caminos para la adopción de soluciones analíticas futuras.

## **3.2. DELIMITACIÓN DEL ESTUDIO**

### **3.2.1. Delimitación Espacial**

El desarrollo de la solución ETL se realiza en el contexto de una Entidad Financiera de tipo Bancario que opera en el Perú. El área específica involucrada es el Área de Sistemas Administrativos, responsable del procesamiento de los datos provenientes de la Superintendencia de Banca, Seguros y AFP (SBS).

### **3.2.2. Delimitación Temporal**

El estudio se enmarca entre los años 2024 y 2025, periodo durante el cual se analiza el proceso actual, se diseña la propuesta ETL y se desarrolla el prototipo de solución.

### **3.2.3. Delimitación Operativa**

El proyecto considera la automatización de la carga masiva de los datos proporcionados por la SBS principalmente y cálculos internos relacionados a estos.

### **3.2.4. Delimitación Funcional (Actores involucrados)**

Se considera la participación de los actores directamente vinculados con el proceso del tratamiento de los datos de la SBS dentro de la Entidad Financiera. Estos incluyen, el jefe, el analista a cargo pertenecientes al Área de sistemas Administrativos, el Administrador de base de datos y usuarios finales. No están considerados los actores externos a la Entidad Financiera, proveedores de software, ni decisiones externas al caso.

## **CAPÍTULO IV: FORMULACIÓN DE DISEÑO**

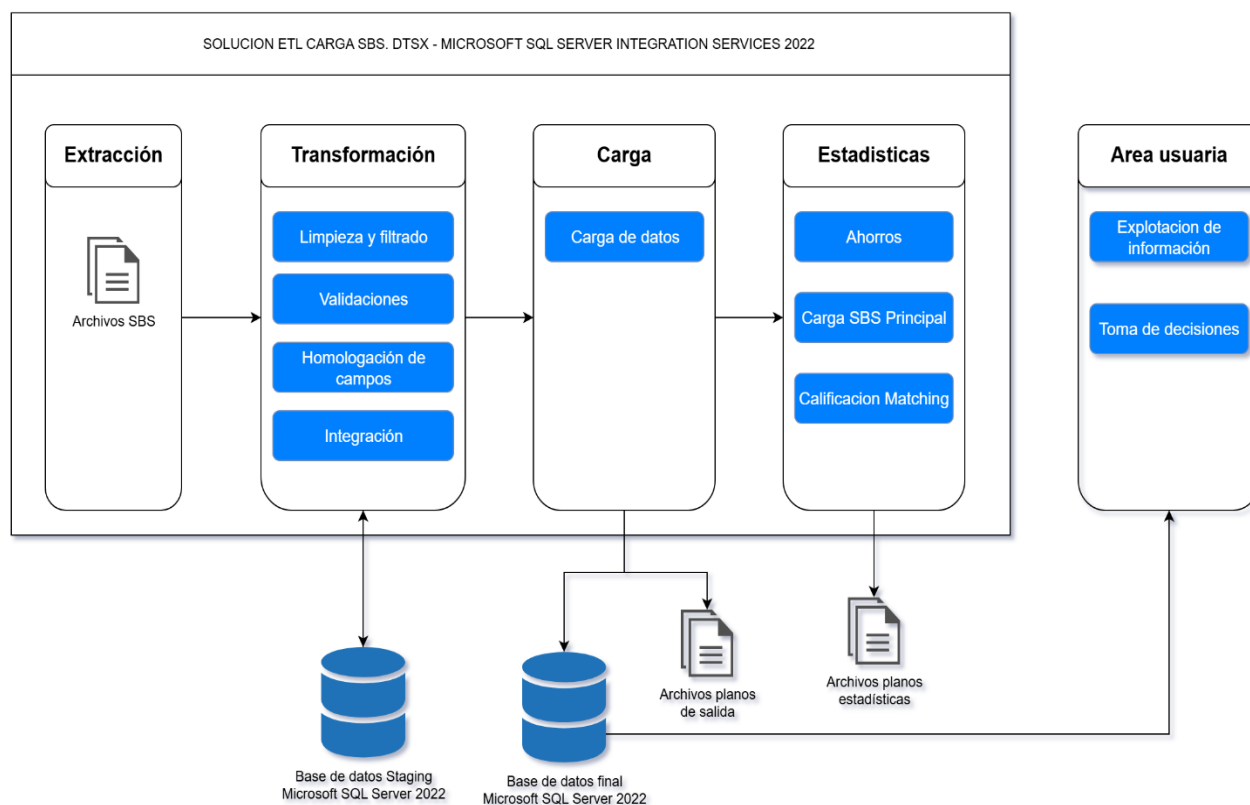
En este capítulo se narra el desarrollo de la solución ETL para la carga masiva de registros de la Superintendencia de Banca, Seguros y AFP. Se muestra la estructura del ETL, los componentes principales que la integran y la manera en que se ejecutan para realizar un mejor tratamiento de los datos recibidos por la SBS, y que asegura la eficiencia y calidad, para su posterior explotación y utilización en los sistemas finales y tomas de decisión.

El desarrollo fue realizado desde el punto de vista práctico, basándose en los procesos actuales de la carga de datos entregados por la SBS, realizados por el Área de Sistemas Administrativos de una Entidad Financiera. Se tomó como puntos importantes la eficiencia, calidad de los datos de salida, técnicas de limpieza y filtrado, además de dimensionar la solución de manera que permita adecuarse a las tecnologías actuales permitiendo su escalabilidad para futuras mejoras y/o modificaciones.

#### 4.1. DISEÑO ESQUEMÁTICO

El diseño esquemático muestra de forma gráfica los componentes principales de la solución ETL desarrollada, la cual se caracteriza en ser moderna y escalable, además de ser eficiente y brindando datos de calidad, para el Área de Sistemas Administrativos.

El diseño esquemático básico:



*Figura 7. Diseño esquemático solución ETL (Elaboración propia).*

#### 4.1.1. Componentes básicos

1. **Fuente de datos (Archivos SBS):** Son los archivos planos proporcionados mensualmente por la SBS, también los generados por la entidad financiera para sus cálculos internos del proceso ETL.
  - Datos Clientes.
  - Registros Saldos.
  - Calificación Ahorros.
2. **Módulo de Extracción:** Responsable de leer los archivos originales.
3. **Módulo de Transformación:** Encargado de limpiar, estandarizar, validar y estructurar los datos. Incluye procesos de:
  - Filtrado de registros inválidos o desactualizados.

- Conversión de formatos y homologación de campos.
- Validación de reglas de negocio.
- Enriquecimiento de datos.

**4. Módulo de Carga:** Inserta los datos ya transformados en la base de datos de explotación Microsoft SQL Server 2022, asegurando integridad referencial y trazabilidad.

#### 4.1.2. Tecnologías de Implementación:

**Tabla 1**

*Listado de Tecnologías de desarrollo e implementación empleadas.*

Componente	Tecnología / Plataforma	Justificación
<b>IDE de Desarrollo</b>	Visual Studio 2022	Permite el desarrollo de la solución ETL moderna.
<b>Manejador de Base de datos</b>	Microsoft SQL Management Studio 19	Permite realizar las consultas directamente en las bases de datos utilizadas para la carga ETL.
<b>Motor ETL</b>	Microsoft SQL Server Integration Services (SSIS)	Permite diseñar visualmente flujos ETL, gestionar conexiones, transformaciones y despliegues de forma centralizada y con amplio soporte corporativo.
<b>Base de datos Staging y destino</b>	Microsoft SQL Server 2022	Ofrece alto rendimiento, particionamiento y nuevas funcionalidades de seguridad y compresión para manejar masivos volúmenes de datos.
<b>Sistema operativo servidor</b>	Windows Server 2019	Plataforma estable y probada para entornos críticos, con soporte extendido y características de clustering y seguridad avanzadas.

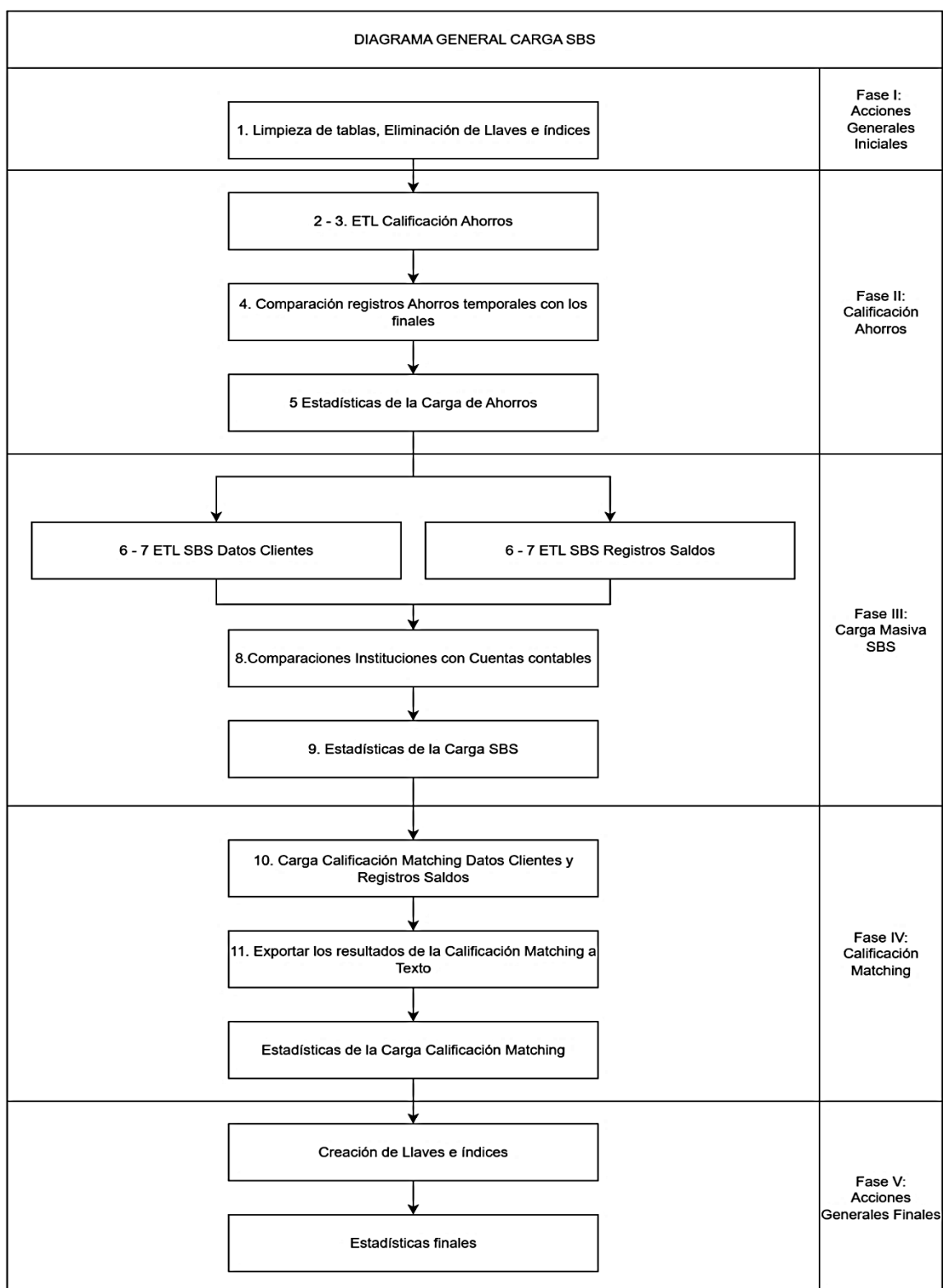
<b>Orquestación y scheduling</b>	SQL Server Agent	Para programar la ejecución nocturna de paquetes SSIS y automatizar tareas de pre y post-procesamiento.
<b>Scripting y transformaciones</b>	T-SQL	T-SQL para transformaciones complejas en la base de datos

#### 4.1.3. Fases la solución ETL desarrollada

Los componentes internos se distribuyen en 5 fases:

- 1. Fase I - Acciones generales iniciales:** Fase que contempla las acciones relacionadas a la limpieza, eliminación de llaves e índices de las tablas relacionadas a la carga SBS.
- 2. Fase II - Calificación Ahorros:** En esta fase se desarrolla la primera carga relacionada a la calificación de Ahorros en base al archivo de texto 'data\_ahorros.txt', además de la comparación de los registros del archivo de origen vs los registros cargados en la base de datos final, incluye también las estadísticas relacionadas a números de registros, hora de inicio y hora de fin del proceso.
- 3. Fase III - Carga Masiva SBS:** Esta fase se realiza la carga de los datos masivos de la SBS, relacionados a los datos de los clientes y los registros de los saldos de estos, se realiza la limpieza y filtrado, además de realizar la comparación de instituciones y cuentas contables con respecto a los registros saldos, incluye también las estadísticas relacionadas a números de registros, hora de inicio y hora de fin del proceso.
- 4. Fase IV - Calificación Matching:** Fase que realiza la carga de la calificación Matching entre Datos Clientes y Calificación Ahorros, además de exportar los resultados en un archivo de texto plano, incluye también las estadísticas relacionadas a números de registros, hora de inicio y hora de fin del proceso.
- 5. Fase V – Acciones generales finales:** Fase que contempla las acciones relacionadas a la creación de índices y llaves en las tablas relacionadas a la carga SBS, incluye

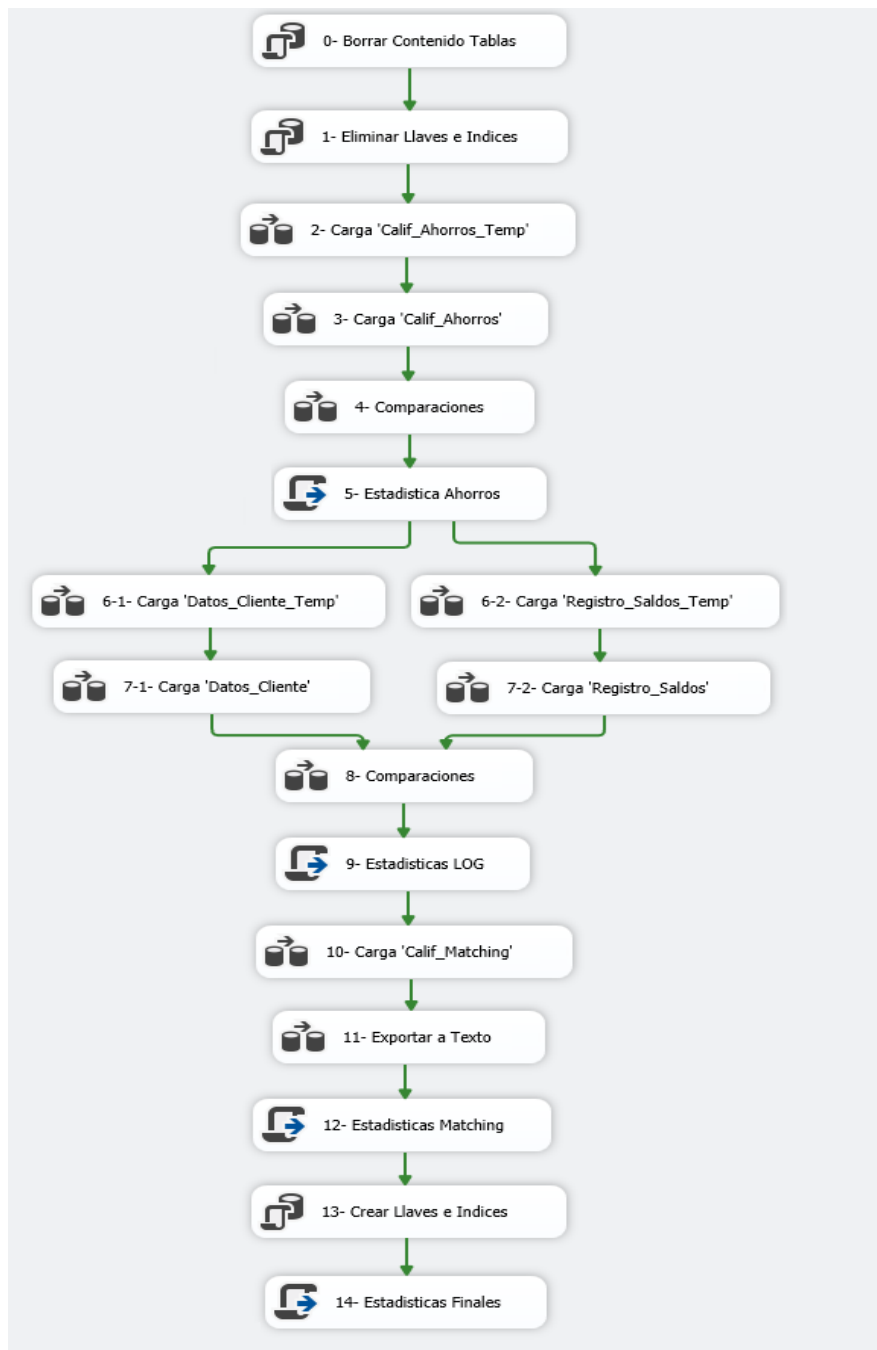
también las estadísticas relacionadas a la hora de inicio y hora de fin del proceso.



*Figura 8. Diagrama general de Fases Carga SBS (Elaboración propia).*

## 4.2. DESCRIPCIÓN DE LOS ASPECTOS BÁSICOS DEL DISEÑO

Para dar solución al problema identificado y cumplir con los objetivos mencionados en este trabajo, se ha seleccionado un diseño de aplicación orientado a ingeniería de datos. A continuación, se describen sus aspectos esenciales.



*Figura 9. Componentes de la solución ETL Carga SBS (Elaboración propia).*

#### 4.2.1. Componentes Principales de la Solución ETL

Teniendo en cuenta las fases generales se detalla en la siguiente tabla los componentes de la solución ETL en Microsoft SQL Server Integration Services.

**Tabla 2**

*Componentes de la Solución ETL en Microsoft SQL Server Integration Services.*

Fase	Componente	Tipo Tarea	Descripción
<b>I</b>	<b>Borrar Contenido Tablas</b>	Tarea Ejecutar SQL	Componente encargado de limpiar el contenido de las tablas Staging y destino.
	<b>Eliminar llaves e índices</b>	Tarea Ejecutar SQL	Componente encargado de eliminar las llaves e índices de las tablas destino.
<b>II</b>	<b>Carga Calif_Ahorros_Temp</b>	Tarea Flujo de Datos	Realiza la carga de la data del archivo Data_Ahorros.txt a la tabla Staging Calif_Ahorros_Temp.
	<b>Carga Calif_Ahorros</b>	Tarea Flujo de Datos	Realiza la carga y filtrado de los datos desde la tabla Staging Calif_Ahorros_Temp hacia la tabla destino Calif_Ahorros.
	<b>Comparaciones (Calif_Ahorros)</b>	Tarea Flujo de Datos	Compara los datos de la tabla Calif_Ahorros_Temp con los de la tabla Calif_Ahorros y exporta un archivo de texto relacionado a los ahorros no registrados (Ahorros_NoValido.txt).
	<b>Estadística Ahorros</b>	Tarea Script	Exporta los datos relacionados al inicio y fin del proceso además de la cantidad de registros.
<b>III</b>	<b>Carga Datos_Cliente_Temp</b>	Tarea Flujo de Datos	Realiza la carga, limpieza y filtrado de la data del archivo data_SBS_1.ope a la tabla Staging Datos_Cliente_Temp.

	<b>Carga Datos_Cliente</b>	Tarea Flujo de Datos	Realiza la carga y últimos cálculos de los datos desde la tabla Staging Datos_Cliente_Temp hacia la tabla destino Datos_Cliente.
	<b>Carga Registro_Saldos_Temp</b>	Tarea Flujo de Datos	Realiza la carga, limpieza y filtrado de la data del archivo data_SBS_2.ope a la tabla Staging Registro_Saldos_Temp.
	<b>Carga Registro_Saldos</b>	Tarea Flujo de Datos	Realiza la carga de los datos desde la tabla Staging Registro_Saldos_Temp hacia la tabla destino Registro_Saldos.
	<b>Comparaciones (Registro Saldos)</b>	Tarea Flujo de Datos	Realiza la comparación con la tabla Registro_Saldos y exporta las Cuentas contables no registradas (Ctas_Contables_Faltantes.txt) y las instituciones faltantes (Instituciones_Faltantes.txt).
	<b>Estadísticas LOG</b>	Tarea Script	Exporta los datos relacionados al inicio y fin de los procesos además de la cantidad de registros.
IV	<b>Carga Calif_Matching</b>	Tarea Flujo de Datos	Realiza la carga de la calificación Matching entre Datos Clientes y Calificación Ahorros.
	<b>Exportar a Texto</b>	Tarea Flujo de Datos	Exporta los resultados de la carga Calificación Matching en un archivo de texto plano.
	<b>Estadísticas Matching</b>	Tarea Script	Exporta los datos relacionados al inicio y fin del proceso además de la cantidad de registros.
V	<b>Crear Llaves e Índices</b>	Tarea Ejecutar SQL	Crea las llaves e índices a las tablas de destino

	<b>Estadísticas finales</b>	Tarea Script	Exporta los datos relacionados al inicio y fin de los procesos además de la cantidad de registros de cada carga.
--	-----------------------------	--------------	--

#### 4.2.2. Variables de Medición

**Tabla 3**

*Variables de Medición para la solución ETL.*

Variable	Definición	Indicador de Éxito
Tiempo de ejecución.	Duración total del proceso ETL (hh:mm:ss).	Reducción $\geq 40\%$ vs. proceso actual.
Porcentaje de registros válidos.	% de filas sin errores tras limpieza y filtrado.	$\geq 95\%$ .
Uso de recursos del servidor.	CPU y memoria promedio durante la ejecución.	No superar el umbral de 70% de CPU.
Confiabilidad (errores).	Número de fallos críticos detectados por lote de datos.	$\leq 1$ error crítico por cada 1 000 000 registros.

#### 4.2.3. Procesos de Validación

1. **Pruebas unitarias** sobre cada rutina de extracción y transformación.
2. **Pruebas de integración** para verificar el flujo completo en un entorno de staging.
3. **Pruebas de rendimiento** midiendo tiempos y uso de CPU/RAM en un volumen de 80 millones de registros.
4. **Validación funcional** con el Área de Sistemas Administrativos, mediante revisión de muestra de datos procesados y retroalimentación de los usuarios finales.

### 4.3. METODOLOGÍA DE DESARROLLO - XP

Para el desarrollo de la solución ETL se decidió emplear la metodología XP o también conocida como Extreme Programming, tal como indica Ben (2024), plantea un modelo de desarrollo de software que responde con rapidez a las necesidades del cliente, el cual evoluciona de manera ágil y se nutre de ciclos constantes de retroalimentación. Su esencia yace en la búsqueda de la excelencia técnica, la simplicidad y la flexibilidad para realizar cambios de forma inmediata.

#### 4.3.1. Principios y Valores XP

Según afirma Ben (2024) son los siguientes:

1. **Comunicación continua:** Reuniones diarias breves entre los desarrolladores y los analistas de sistemas administrativos.
2. **Simplicidad:** Centrarse en lo que se requiere en el momento, así evitando que se complique el desarrollo con características superfluas.
3. **Retroalimentación constante:** Usar el feedback de los clientes, usuarios y las pruebas continuas para dirigir el desarrollo y asegurar que la solución cumpla con los objetivos.
4. **Coraje:** Tener la fuerza de hacer siempre lo mejor para el desarrollo del proyecto o solución, incluso si eso conlleva cuestionar las opciones existentes.
5. **Respeto:** Cada miembro del equipo valora lo desarrollado por los demás, generando un ambiente productivo y positivo.

#### 4.3.2. Adaptación de la metodología XP

##### 1. Generación de Historias de Usuario y Planning Game

- Recoger las historias de usuario.
- Priorizar las historias que aporten más valor.

##### 2. Integración Continua

- Despliegue del paquete SSIS en entorno de desarrollo.
- Ejecución de Pruebas Unitarias.
- Informe de errores si falla alguna prueba.

### **3. Programación en Pareja**

- Al realizarse actividades complejas, el desarrollador y el analista cliente trabajan juntos, garantizando que el flujo sea como se indica las reglas del negocio.

### **4. Refactorización Constante**

- Tras cada entrega funcional, se realiza una revisión del diseño de los paquetes, para realizar ajustes o cambios que sirvan para mejorar o simplificar funciones.

### **5. Despliegues Frecuentes**

- Sprints de una semana, entregando un flujo ETL completo.

### **6. Cliente Presente**

- El analista del Área de Sistemas Administrativos realizará una revisión diaria de los datos de prueba, mostrando aprobación u observaciones, según el caso, cerrando el ciclo de retroalimentación

## CAPÍTULO V: PRUEBA DE DISEÑO

Este capítulo muestra el proceso de validación del diseño propuesto y desarrollado para la carga masiva de los registros enviados por la Superintendencia de Banca, Seguros y AFP para el Área de Sistemas Administrativos de una Entidad Financiera. La finalidad de este apartado es verificar que la solución ETL cumpla con lo solicitado por el área indicada, tanto en el tiempo de carga, la cantidad y la calidad de registros, además de la escalabilidad de la solución a nivel tecnológico.

La solución fue desarrollada en Microsoft SQL Server Integration Services, teniendo compatibilidad con bases de datos SQL Server 2019 – 2022, bajo un sistema operativo Windows 10, y su ejecución en Windows Server 2019.

Por medio de las pruebas constantes cuya finalidad es la obtención de la retroalimentación del usuario analista del Área de Sistemas Administrativos principalmente, además de indicada por los usuarios que ejecutan la carga y los usuarios finales de otras áreas.

Se toma en cuenta también la escalabilidad de la solución con respecto a futuras mejoras y modificaciones que requiera realizar la entidad financiera.

### 5.1. APLICACIÓN DE LA PROPUESTA DE SOLUCIÓN

Para verificar la viabilidad y eficacia de la solución ETL, se llevó a cabo un despliegue controlado en un entorno de pruebas que replica las condiciones reales del Área de Sistemas Administrativos. A continuación, se describen las fases y resultados de esta aplicación.

#### 5.1.1. Preparación del Entorno de Pruebas

##### 1. Infraestructura

- Servidor Windows Server 2019 configurado con SQL 2022 y Catalogo SSISDB.
- Bases de datos de staging y repositorio final (BN\_SBS).

## 2. Datos de Entrada

- Se utilizaron los archivos de texto plano con extensión OPE, para el caso de SBS y TXT para las demás cargas.
- La estructura de del archivo SBS de entrada se basa a lo indicado en el Anexo A, entregado por el Analista de Sistemas Administrativos.
- Se configuró un volumen de prueba experimental de 81 millones de registros aproximadamente para el caso de SBS, 13 millones aproximadamente de registros para Calificación Ahorros.

## 3. Configuración de SSIS

- Despliegue del paquete SSIS en Servidor SQL Server dedicado.
- Parámetros de ejecución: Rutas de los ficheros, conexiones de los repositorios de datos.

### 5.1.2. Modelo de Base de Datos

El diseño de la base de datos fue importante para poder estructurar la solución ETL (Ver Anexo B), el modelo, se consideró la creación de las siguientes tablas en conformidad con lo solicitado por el analista:

**Tabla 4**

*Detalle de las tablas BN\_SBS.*

Tabla	Columna	Tipo Dato	Longitud	Llave
<b>Calif_Ahorros</b>	Doc_Identidad	-Char-	10	
	Tip_Doc_Identidad	-Char-	1	
	Cta_Ahorros	-Char-	11	
<b>Calif_Ahorros_Temp</b>	Doc_Identidad	-Char-	10	
	Tip_Doc_Identidad	-Char-	1	
	Cta_Ahorros	-Char-	11	
<b>Calif_Matching</b>	Doc_Identidad	-Char-	10	

	Tip_Doc_Identidad_Sbs	-Char-	1	
	Nom_Cliente_Compl	-Varchar-	120	
	Cta_Ahorros	-Char-	11	
	Cal_Normal	-Numeric-	5,2	
	Cal_Cpp	-Numeric-	5,2	
	Cal_Deficiente	-Numeric-	5,2	
	Cal_Dudoso	-Numeric-	5,2	
	Cal_Perdido	-Numeric-	5,2	
	CalificacionFinal	-Char-	1	
	Doc_Identidad_Sbs_Corr eg	-Char-	11	
	Cod_deudor	-Char-	10	
	Tip_Doc_Identidad	-Char-	1	
	Nom_Cliente_Compl_70	-Char-	70	
<b>Calificacion</b>	Cod_Calif	-Char-	1	-Primaria-
	Det_Calif	-Char-	15	
	DetLargo_Calif	-Nvarchar-	2000	
<b>Cuentas_Contables</b>	Cod_cta_contable	-Decimal-	14,0	-Primaria-
	Desc_cta_contable	-Varchar-	255	
<b>Datos_Cliente</b>	Cod_Deudor	-Char-	10	-Primaria-
	Fec_Reporte	-Small datetime-		
	Tip_Doc_Trib	-Char-	1	
	Ruc	-Char-	11	
	Tip_Doc_Identidad_sbs	-Char-	1	
	Doc_Identidad_sbs	-Char-	12	
	Tip_Persona	-Char-	1	
	Tip_Empresa	-Char-	1	
	Num_Empresas	-Char-	3	
	Cal_Normal	-Numeric-	5,2	
	Cal_Cpp	-Numeric-	5,2	

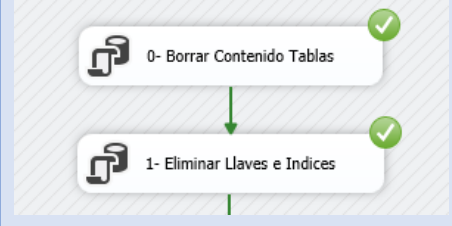
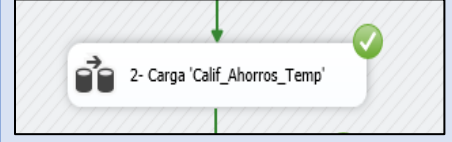
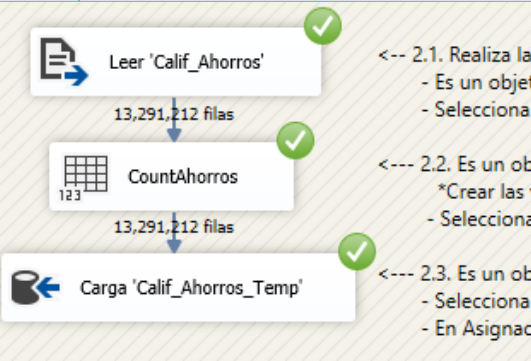
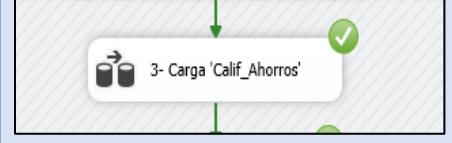
	Cal_Deficiente	-Numeric-	5,2	
	Cal_Dudoso	-Numeric-	5,2	
	Cal_Perdido	-Numeric-	5,2	
	Nom_Cliente_Compl	-Varchar-	120	
	CalificacionFinal_DTS	-Char-	1	
	Tip_doc_Identidad_sbs_	-Char-	1	
	Convert_ahorr			
	Doc_Identidad_sbs_Corr	-Char-	10	
	eg			
<b>Datos_Cliente_Temp</b>	Tip_registro	-Char-	1	
	Cod_deudor	-Char-	10	
	Fec_reporte	-Small datetime-		
	Tip_doc_trib	-Char-	1	
	Ruc	-Char-	11	
	Tip_doc_identidad	-Char-	1	
	Doc_identidad	-Char-	12	
	Tip_persona	-Char-	1	
	Tip_empresa	-Char-	1	
	Num_empresas	-Char-	3	
	Cal_normal	-Numeric-	5,2	
	Cal_cpp	-Numeric-	5,2	
	Cal_deficiente	-Numeric-	5,2	
	Cal_dudoso	-Numeric-	5,2	
	Cal_perdido	-Numeric-	5,2	
	Nom_Cliente_ApePat	-Varchar-	100	
	Nom_Cliente_ApeMat	-Varchar-	40	
	Nom_Cliente_ApeCas	-Varchar-	40	
	Nom_Cliente_PriNom	-Varchar-	40	
	Nom_Cliente_SegNom	-Varchar-	40	
	Nom_Cliente_Compl	-Varchar-	220	

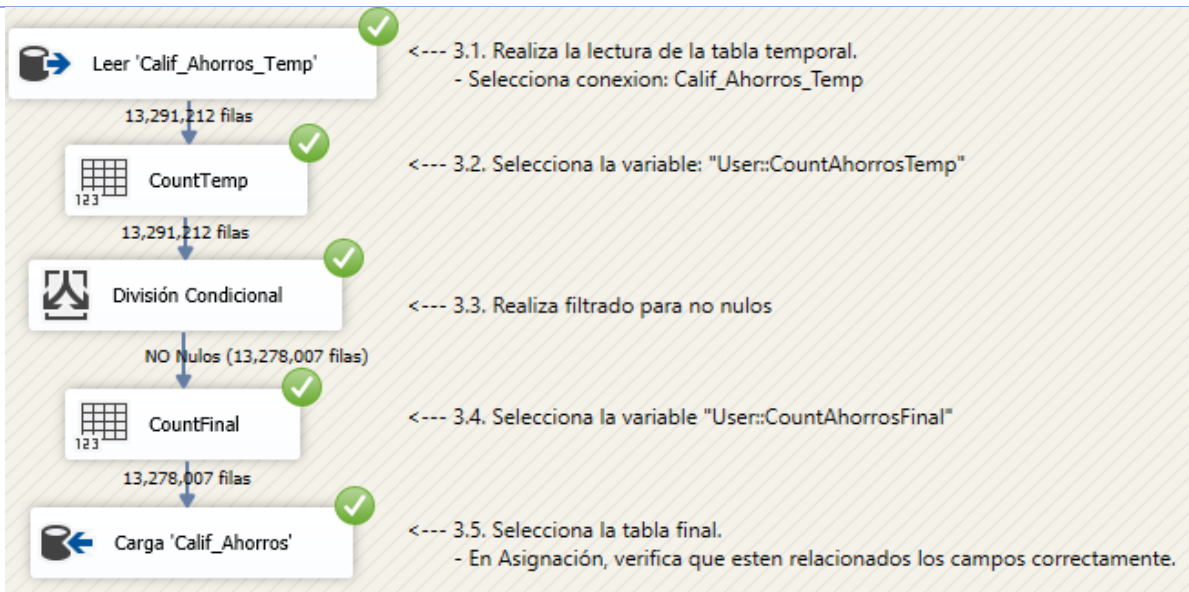
<b>Instituciones</b>	Cod_Institucion	-Char-	5	-Primaria-
	Tipo_Institucion	-Varchar-	30	
	Estado_Institucion	-Varchar-	30	
	Nom_Institucion	-Varchar-	150	
<b>Registro_Saldos</b>	Cod_Deudor	-Char-	10	-Primaria-
	Cod_Empresa	-Char-	5	-Primaria-
	Tip_Credito	-Char-	2	-Primaria-
	Cod_Cta_Contable	-Numeric-	14,0	-Primaria-
	Condicion	-Char-	4	-Primaria-
	Saldo	-Numeric-	15,2	
	Calif_Final_xSbs	-Char-	1	
<b>Registro_Saldos_Temp</b>	Tip_registro	-Char-	1	
	Cod_deudor	-Char-	10	
	Cod_empresa	-Char-	5	
	Tip_credito	-Char-	2	
	Cod_cta_contable	-Char-	14	
	Condicion	-Char-	4	
	Saldo	-Numeric-	18,2	
	Calif_Final_xSbs	-Char-	1	
<b>Tipo_Credito</b>	Tip_Credito	-Char-	2	-Primaria-
	Desc_Credito	-Varchar-	35	
	DescLarga_Credito	-Nvarchar-	2000	
<b>Tipo_Documento</b>	Tip_Doc_Identidad	-Char-	1	-Primaria-
	Desc_Documento	-Varchar-	40	
<b>Tipo_Persona</b>	Tip_Persona	-Char-	1	-Primaria-
	Desc_Persona	-Varchar-	125	

### 5.1.3. Detalle de los Componentes de la Carga de Datos SBS

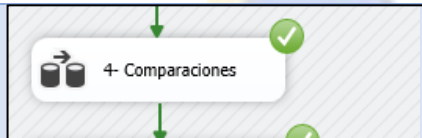
**Tabla 5**

*Detalle de los componentes de la solución ETL según flujo de Fases.*

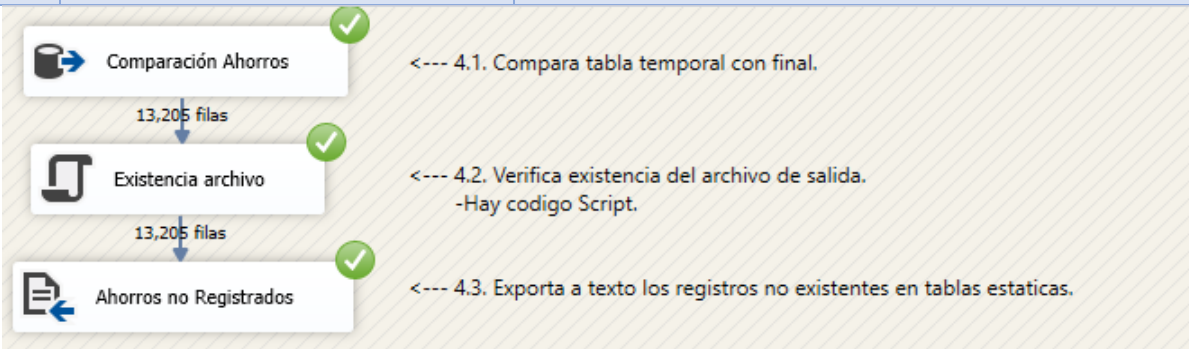
Fase 1: Acciones Generales Iniciales		
<b>1</b>		<p>1.1. Realiza la limpieza de tablas.</p> <p>1.2. Realiza la limpieza llaves e índices</p>
Fase 2: Carga Calificación Ahorros		
<b>2</b>		<p>2. La tarea de flujo de datos realiza la carga a la tabla staging 'Calif_Ahorro_Temp', contiene el siguiente proceso:</p>
		<p>&lt;-- 2.1. Realiza la lectura del archivo de entrada, - Es un objeto del Cuadro de herramientas, llamado "Origen de archivo plano". - Selecciona conexion:03_Calif_Ahorros.</p> <p>&lt;--- 2.2. Es un objeto del Cuadro de herramientas: "Recuento de filas". *Crear las variables: "CountAhorrosTXT", "CountAhorrosTemp", "CountAhorrosFinal"* - Selecciona la variable: "User::CountAhorrosTXT".</p> <p>&lt;--- 2.3. Es un objeto del Cuadro de herramientas: "Destino OLE DB". - Selecciona la tabla temporal. - En Asignaciones, verifica que estén relacionados correctamente.</p>
<b>3</b>		<p>3. La tarea de flujo de datos realiza la carga a la tabla final 'Calif_Ahorro', contiene el siguiente proceso:</p>



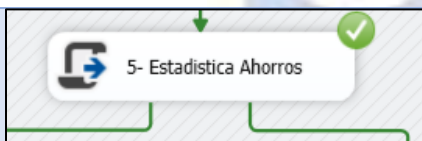
4



4. La tarea de flujo de datos realiza la Comparación de Calificación Ahorros, por medio del siguiente proceso:



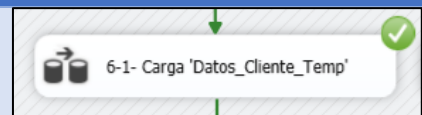
5



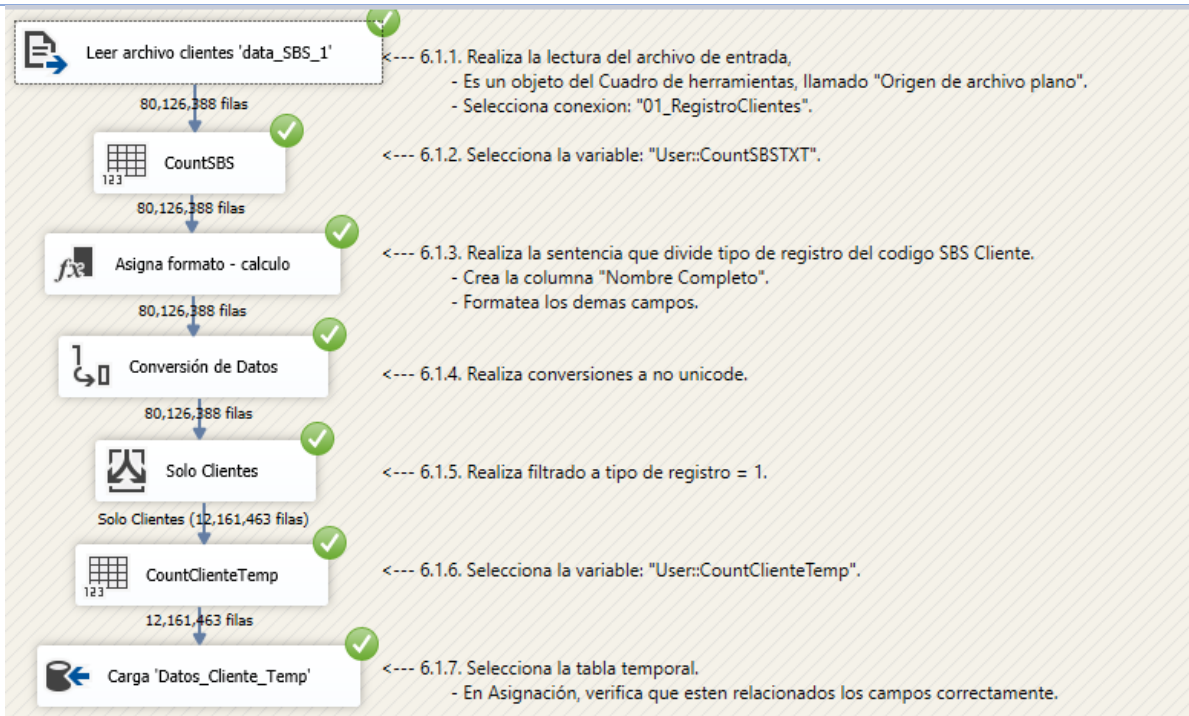
5. Se realiza la Estadística de Ahorros.

### Fase 3: Carga SBS

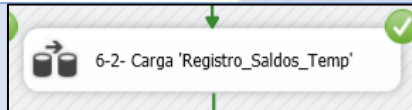
6.1.



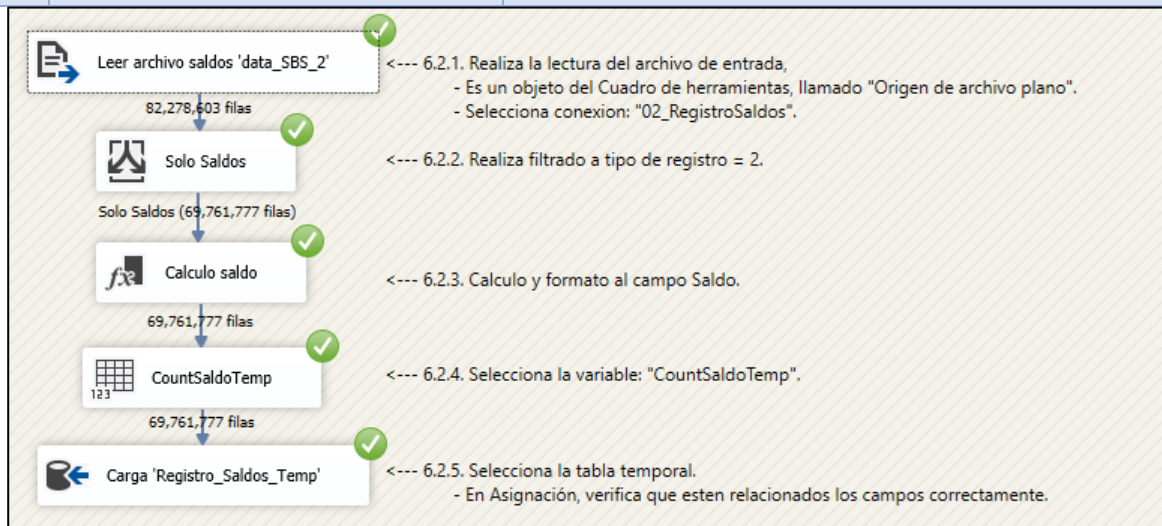
5.1. Se realiza la Carga de Datos Clientes a la tabla Staging 'Datos\_Cliente\_Temp', por medio del siguiente proceso:



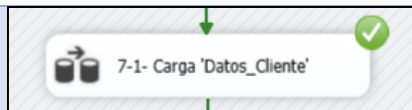
6.2.



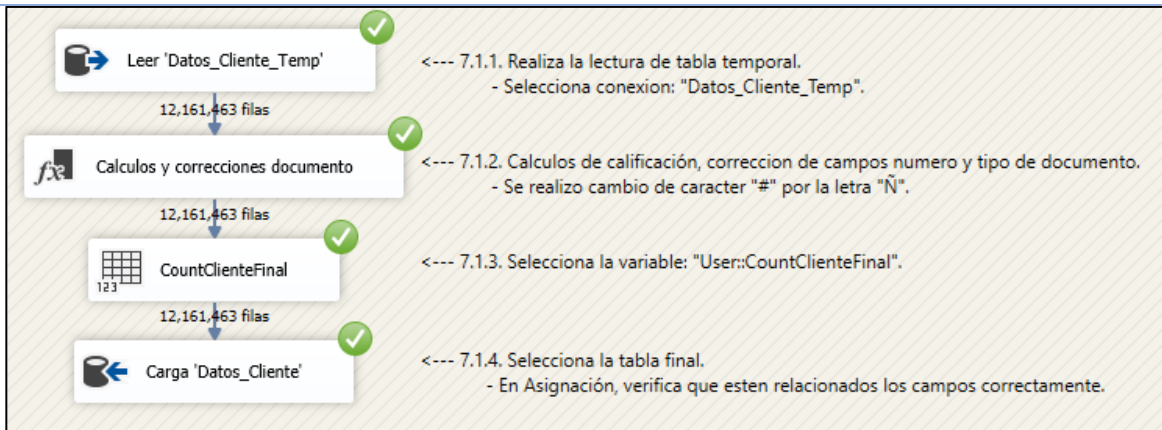
5.2. Se realiza la Carga de Registros Saldos a la tabla Staging 'Registro\_Saldos\_Temp', por medio del siguiente proceso:



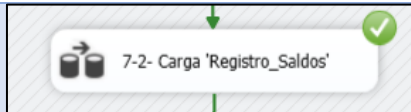
7.1.



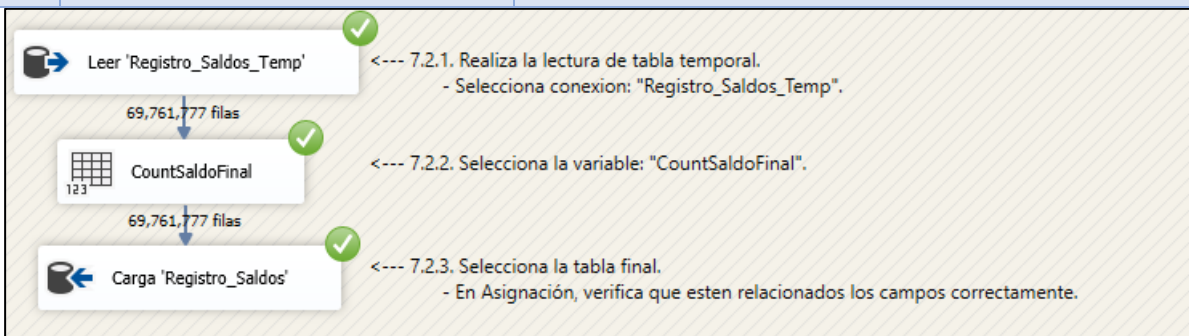
6.1. Se realiza la Carga de Datos Clientes de la tabla Staging 'Datos\_Cliente\_Temp' a la tabla final 'Datos\_Cliente', por medio del siguiente proceso:



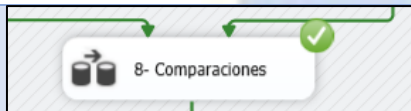
7.2.



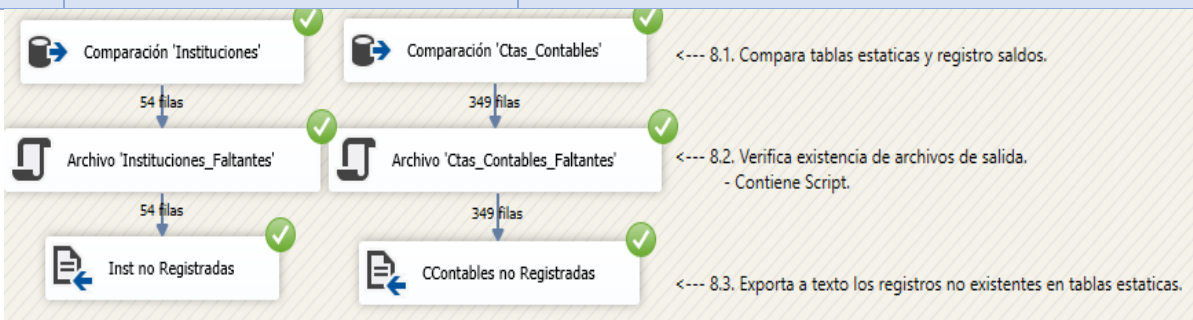
6.2. Se realiza la Carga de Datos Clientes de la tabla Staging 'Registro\_Saldos\_Temp' a la tabla final 'Registro\_Saldos', por medio del siguiente proceso:



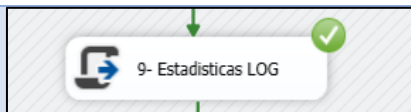
8



7. Se ejecutan las comparaciones de las instituciones faltantes y las cuentas contables no validas o faltantes y el resultado es exportado en un archivo de texto plano, por medio del siguiente proceso:




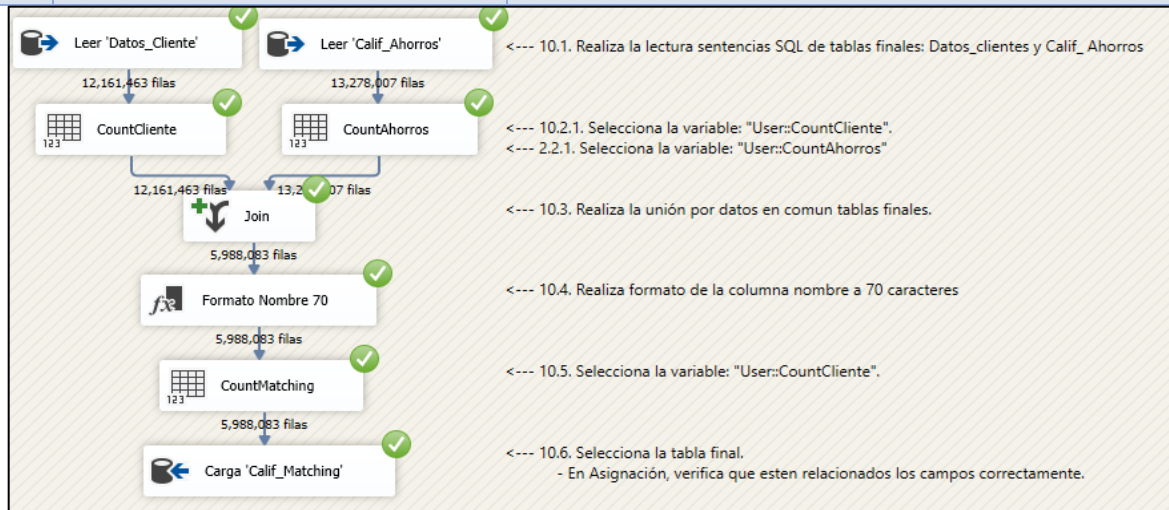
9

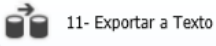


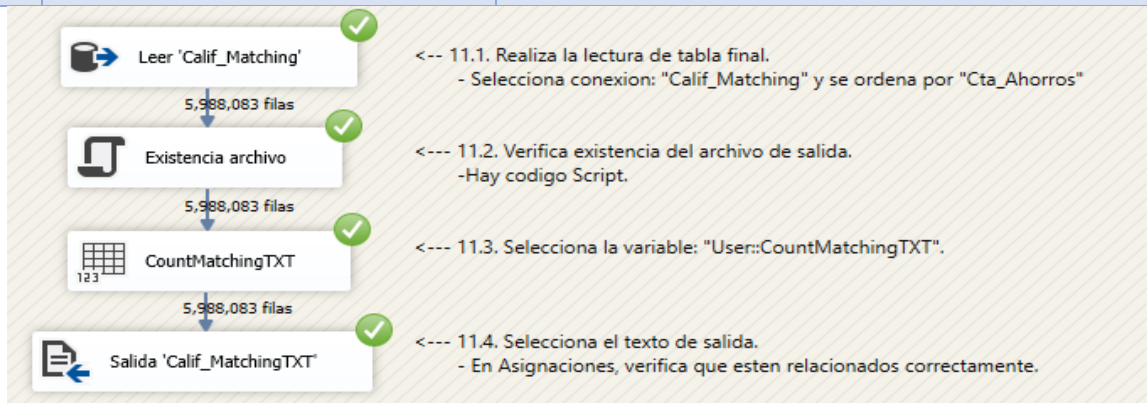
9. Se realiza la Estadística de la Carga SBS.


### Fase 4: Calificación Matching

<b>10</b>		<p>10. Se ejecuta la Carga de Datos Clientes vs Calificación Ahorros a la tabla final 'Calif_Matching', por medio del siguiente proceso:</p>
-----------	---	--





<b>11</b>		<p>11. Se exporta los resultados de la Calificación Matching, por medio del siguiente proceso:</p>
-----------	---	--



<b>12</b>		<p>12. Se realiza la Estadística de la Calificación Matching.</p>
-----------	---	---

### Fase 5: Acciones Generales Finales

<b>13</b>		<p>13. Se generan nuevamente las llaves e índices.</p>
<b>14</b>		<p>14. Dentro de la tarea Script se realiza las Estadísticas finales</p>

Ir al Anexo C para ver la estructura del documento de salida de las calificaciones, ahorros no validos e instituciones faltantes, además de las estadísticas.

#### 5.1.4. Ejecución del Flujo ETL

##### 1. Extracción

- Lectura de los registros de los archivos enviados por la SBS y el archivo de Calificación Ahorros.
- Validación de estructura: Archivos de salida con los registros no validos de Ahorro y estadísticas de carga.

##### 2. Transformación

- Aplicación en T-SQL y Flujo de datos (Data Flow): Limpieza de caracteres no válidos, filtrado de datos, columnas calculadas, complementación de caracteres, homologación de tipos de dato.

##### 3. Carga

- Inserción de lotes de 500 000 filas en la base BN\_SBS con transacciones controladas.
- Verificación de integridad y control de duplicados.

#### 5.1.5. Métricas y Resultados Obtenidos

Las métricas se realizaron en un entorno de desarrollo con las mismas características del entorno Producción, la carga se realizó en horario de oficina.

**Tabla 6**

*Métricas y resultados obtenidos (Entorno Desarrollo).*

Indicador	Antes (DTS/2008R2)	Después (SSIS/2022)	Mejora
Tiempo total de procesamiento	2.5 horas	0.75 – 1 hora	60 % – 70 %

% de registros válidos en staging	65 %	96 %	+31 p. p.
Número de errores críticos por lote	4 – 6	0 – 1	– 75 %
Uso promedio de CPU durante ejecución	85 %	55 %	– 30 p. p.
Memoria utilizada (pico)	12 GB	8 GB	– 33 %

### 5.1.6. Verificación y Validación Funcional

- Revisión del Analista de Sistemas Administrativos: Se entregaron muestras de los registros procesados para inspección manual. El analista confirmó la correcta aplicación de las reglas y ausencia de datos duplicados y datos incorrectos.
- Registro de trazabilidad: Cada ejecución de la solución ETL entregó estadísticas, permitiendo realizar las comparaciones de registros de origen vs los registros finales, además del tiempo de ejecución de cada proceso.

### 5.1.7. Observaciones y Ajustes

- Se realizaron ajustes en la carga de Calificación de Ahorros.
- Se realizaron ajustes en el filtrado de la carga de datos SBS (Datos\_Cliente, Registro\_Saldos).
- Se agregaron salidas de estadísticas finales en archivo de texto plano (TXT).
- Se modificaron los Buffer de lectura de datos para cada punto de carga.

## CONCLUSIONES

1. La solución ETL desarrollada redujo el tiempo de carga masiva de los registros de la SBS en un 60% promedio con respecto al proceso de carga actual. Pasando de ejecuciones que 2.5 horas solo en el proceso de carga principal a menos de 1 hora, lo que libera recursos, acortando la ventana del procesado nocturno.
2. Gracias a las reglas de limpieza y filtrado automatizadas, se minimizo los errores de duplicados y datos obsoletos que antes requerían revisiones manuales.
3. La implementación de un mecanismo de estadísticas para auditoria de cada proceso realizado que provee de información con respecto a la cantidad de registros, el tiempo de ejecución y la comparativa de los datos de entrada versus los datos filtrados de salida, la cual se puede utilizar para análisis posteriores y tomas de decisión.
4. La adopción de SSIS sobre SQL Server 2022 en Windows Server 2019 demostró ser una plataforma escalable y estable, que cuenta con la capacidad de adaptación a crecientes volúmenes de datos y nuevas reglas de negocio sin que se requiera rediseños mayores.
5. La aplicación de prácticas de Extreme Programming (entregas frecuentes, programación en pareja, etc.) facilito la rápida adaptación ante cambios en las especificaciones, mejorando la calidad del código ETL y asegurando entregas incrementales de alto valor.
6. Al disponer de datos más limpios, estructurados y disponibles, contribuye a optimizar la gestión de riesgo y mejorar la competitividad de la entidad financiera en el mercado.

## RECOMENDACIONES

1. Organizar capacitaciones para el equipo del Área de Sistemas Administrativos en lo relacionado a SSIS, integración de paquetes DTSX en SQL Server 2022 y practicas XP.
2. Aprovechar la arquitectura modular para incorporar otros conjuntos de datos mensuales, reutilizando componentes de extracción y transformación.
3. Integrar los repositorios de datos resultantes a herramientas como Power BI o Azure Análisis Services con el objetivo de ofrecer cuadros de mando en tiempo real a las áreas de Riesgos, Cumplimiento y Planeamiento.
4. Establecer un plan semestral de revisión de las reglas de limpieza/filtrado y de actualización de las tecnologías (SSIS, SQL Server, Sistema Operativo) con el fin de asegurar que la solución ETL siga alineada con las necesidades regulatorias y de negocio
5. Planificar la adopción de contenedores o nube híbrida, es decir explorar la contenerización de paquetes SSIS con Docker o migrar gradualmente al servicio Azure (SSIS en Azure), esto facilitaría la escalabilidad bajo demanda y mejorara la capacidad de adaptación ante picos de carga.
6. Documentar de manera constante, es decir tener un repositorio de documentación técnica y de usuario: manuales de operación, diagramas actualizados, un registro de decisiones de diseño. Esto con el fin de bajar la curva de aprendizaje para los nuevos integrantes y mejorando la sostenibilidad del proyecto.

## REFERENCIAS BIBLIOGRÁFICAS

*En español:*

Arocha, J., Mejía, N., & Suárez, Y. (2021). Big data. Estrategia potenciadora de la competitividad en el sector financiero. *Revista científica global negotium*, 4(3), 155-168. *Recuperado el 03 de junio del 2025, de* <http://publishing.fgu.edu.com/ojs/index.php/RGN/article/view/202>

Consoli, E. (2022). Gestión de datos en organizaciones. Consejo Profesional de Ciencias Económicas de la Ciudad Autónoma de Buenos Aires. *Recuperado el 02 de junio del 2025, de* <https://archivo.consejo.org.ar/publicacionesedicon/Gestion-de-datos-en-Organizaciones..pdf>

Barón Ramírez, E., García Estrella, C. W. & Sánchez Gárate, S. K. (2021). La Inteligencia de negocios y la analítica de datos en los procesos empresariales. *Revista científica de sistemas e informática*, 1(2), 38-53. *Recuperado el 02 de junio del 2025, de* <https://revistas.unsm.edu.pe/index.php/rcsi/article/view/167>

Bintelligenz. (2023). La Importancia de la Calidad de Datos en la Gestión Empresarial. *Bintelligenz.com*. *Recuperado el 02 de junio del 2025, de* <https://bintelligenz.com/importancia-de-la-calidad-de-datos/>

De Arregui, M. (2025). Técnicas de data cleaning para garantizar datos de calidad. *Obsbusiness.school*. *Recuperado el 3 de junio de 2025, de* <https://www.obsbusiness.school/blog/tecnicas-de-data-cleaning-para-garantizar-datos-de-calidad>

Guerrero García, M. J. & Rodas Silva, J. (2022). Análisis comparativo de metodologías y herramientas tecnológicas para procesos de Business Intelligence orientado a la toma de decisiones. *Informática Y Sistemas*, 6(1), 53-62. *Recuperado el 02 de junio del 2025, de* <https://pdfs.semanticscholar.org/71ec/80ed57dab8140089aaa750855808819bbcf1.pdf>

Iglesias, M. J. F. (2023). Pequeña introducción a las bases de datos. *atlanTTic-Universidade de Vigo*, 3. *Recuperado el 02 de junio del 2025, de*

[https://www.researchgate.net/profile/Manuel-Jose-Fernandez-Iglesias/publication/369039369\\_Pequeña\\_introducción\\_a\\_las\\_bases\\_de\\_datos/links/60bd51510ca86798736157d/Pequeña-introducción-a-las-bases-de-datos.pdf](https://www.researchgate.net/profile/Manuel-Jose-Fernandez-Iglesias/publication/369039369_Pequeña_introducción_a_las_bases_de_datos/links/60bd51510ca86798736157d/Pequeña-introducción-a-las-bases-de-datos.pdf)

London Consulting Group. (2024). Importancia de la Calidad de los Datos en Business Intelligence. Londoncg.com. *Recuperado el 02 de junio del 2025, de* <https://londoncg.com/blog/importancia-de-la-calidad-de-los-datos-en-business-intelligenceimportancia-de-la-calidad-de-los-datos-en-business-intelligence>

Malla Valdiviezo, R. O., López Gorozabel, O. A., Arévalo Indio, J. A., & Tóala Briones, C. H. (2023). Mecanismos para el procesamiento de big data. Limpieza, transformación y análisis de Datos. Polo del Conocimiento, 8(4), 656-675. *Recuperado el 03 de junio del 2025, de* <https://www.polodelconocimiento.com/ojs/index.php/es/article/view/5457/13439>

Maquen Niño, G. L., Terán Santa Cruz, F. E., del Castillo Castro, C. I., & Villón Prieto, R. D. (2022). Buenas prácticas para la optimización de base de datos relacionales usando Microsoft SQL. Universidad Ciencia y Tecnología, 26(114), 29-38. *Recuperado el 3 de junio de 2025, de* <https://uctunexpo.autanabooks.com/index.php/uct/article/view/588>

Microsoft. (2021). ¿Qué son las bases de datos? *Recuperado el 03 de junio del 2025, de* <https://azure.microsoft.com/es-es/resources/cloud-computing-dictionary/what-are-databasesq=Dise%C3%B1o+de+bases+de+datos+relacionales&ots=Sdel9feNJx&sig=4Vm7nEut9LVBmxL05Ji-wOhR1ck>

Mucci, T. (2024). ¿Qué es la toma de decisiones basada en datos? IBM.com. *Recuperado el 14 de junio del 2025, de* <https://www.ibm.com/mx-es/think/topics/data-driven-decision-making>

Naeem, T. (2025). ¿Qué es la migración de datos? Conceptos y técnicas explicados. Astera.com. *Recuperado el 3 de junio de 2025, de* <https://www.astera.com/es/knowledge-center/data-migration-software/>

Oracle. (2021). ¿Qué es la inteligencia de negocio?. Oracle.com. *Recuperado el 02 de junio del 2025, de* <https://www.oracle.com/pe/what-is-business-intelligence/>

Peiró, R. (2021). Teoría de la Información. Economipedia.com *Recuperado el 02 de junio del 2025, de <https://economipedia.com/definiciones/teoria-de-la-informacion.html>*

Pérez Porto, J. (2024). Teoría de la Información. Definicion.de. *Recuperado el 02 de junio del 2025, de <https://definicion.de/teoria-de-la-informacion/>*

Piñero Gómez, J. (2024). Diseño de bases de datos relacionales. Ediciones Paraninfo, S.A. *Recuperado el 02 de junio del 2025, de <https://books.google.es/books?hl=es&lr=&id=pas0EQAAQBAJ&oi=fnd&pg=PR5&dq=Dise%C3%B1o+de+bases+de+datos+relacionales&ots=Sdel9feNJx&sig=4Vm7nEut9LVBmxL05Ji-wOhR1ck>*

Postigo Palacios, A. (2022). Gestión de bases de datos. Ediciones Paraninfo S.A. *Recuperado el 3 de junio de 2025, de [https://books.google.com/books?hl=es&lr=&id=4Yh3EAAAQBAJ&oi=fnd&pg=PR5&dq=Consultas+SQL&ots=aFIqOKRY\\_O&sig=oULT3SccliCWQWVOKj8bG1hE7bU](https://books.google.com/books?hl=es&lr=&id=4Yh3EAAAQBAJ&oi=fnd&pg=PR5&dq=Consultas+SQL&ots=aFIqOKRY_O&sig=oULT3SccliCWQWVOKj8bG1hE7bU)*

Romagnano, M., Pantano, J. C. & Becerra, M. C. (2021). El Rol de Business Intelligence y Business Analytics en Organizaciones Actuales. Revista Eletrônica Argentina-Brasil de Tecnologias da Informação e da Comunicação, 1(13). *Recuperado el 3 de junio de 2025, de <https://revistas.setrem.com.br/index.php/reabtic/article/view/388>*

*En ingles y portugués:*

Banoth, D. N., Balasubramaniam, V. S., Kumar, P., & Singh, N. (2024). The Role of SSIS In Data Transformation: Best Practices for ETL In SQL Server Environments. *Recuperado el 03 de junio del 2025, de [https://www.researchgate.net/profile/Dinesh-Nayak-Banoth/publication/390897552\\_The\\_Role\\_Of\\_Ssis\\_In\\_Data\\_Transformation\\_Best\\_Practices\\_For\\_ETL\\_In\\_SQL\\_Server\\_Environments/links/68020438df0e3f544f42a230/The-Role-Of-SSIS-In-Data-Transformation-Best-Practices-For-ETL-In-SQL-Server-Environments.pdf](https://www.researchgate.net/profile/Dinesh-Nayak-Banoth/publication/390897552_The_Role_Of_Ssis_In_Data_Transformation_Best_Practices_For_ETL_In_SQL_Server_Environments/links/68020438df0e3f544f42a230/The-Role-Of-SSIS-In-Data-Transformation-Best-Practices-For-ETL-In-SQL-Server-Environments.pdf)*

- Ben Salem, A. (2024). Extreme Programming (XP): A Beginner's Guide to the Agile Method. *Recuperado el 5 de julio de 2025, de <https://scrum-master.org/en/extreme-programming-xp-a-beginners-guide-to-the-agile-method/>*
- Dos Santos, V. L., & Gibertoni, D. (2022). Os impactos do business intelligence para tomada de decisões. *Revista Interface Tecnológica, 19(2)*, 258-269. *Recuperado el 3 de junio de 2025, de <https://revista.fatectq.edu.br/interfacetecnologica/article/view/1524>*
- Duque, J., Godinho, A., & Vasconcelos, J. (2022). Knowledge data extraction for business intelligence A design science research approach. *Procedia Computer Science, 204*, 131-139. *Recuperado el 3 de junio de 2025, de <https://www.sciencedirect.com/science/article/pii/S1877050922007542>*
- IBM. (2021). What is ETL?. IBM.com *Recuperado el 02 de junio del 2025, de <https://www.ibm.com/think/topics/etl>*
- Noles, C. (2024). Data Cleaning: Definition, Techniques & Best Practices. Technologyadvice.com. *Recuperado el 02 de junio del 2025, de <https://technologyadvice.com/blog/information-technology/data-cleaning/>*
- Poppy, D. (2023). Understanding ETL: Extract, Transform, Load. getdbt.com. *Recuperado el 7 de junio de 2025, de <https://www.getdbt.com/blog/extract-transform-load>*
- Vargas, J., Velasco, L. M., Acosta, A., & Camelo, J. P. (2021). Business Intelligence. Preprints.org. *Recuperado el 3 de junio de 2025, de [https://www.preprints.org/frontend/manuscript/c5a62f35b5c60681af41555117c255b2/download\\_pub](https://www.preprints.org/frontend/manuscript/c5a62f35b5c60681af41555117c255b2/download_pub)*
- Zivkovic, M. (2023). What is ETL and How Does it Work? The Definitive Guide to ETL. luzmo.com. *Recuperado el 7 de junio de 2025, de <https://www.luzmo.com/blog/what-is-etl>*

## ANEXOS

### ANEXO A: ESTRUCTURA PRINCIPAL DE LOS ARCHIVOS SBS DATOS CLIENTES Y REGISTRO SALDOS

#### A.1. Registro de Identificación del Cliente (Según Analista del Área de Sistemas Administrativos)

Tabla 7

Interpretación de la cadena que contiene los registros de los Datos del Cliente.

Nro.	Posición		Tamaño Bytes	Número decimales	Tipo campo	Nombre del campo	Descripción del campo
	Inicial	Final					
1.	1	1	1	0	N	TIPO DE REGISTRO	<b>1= Datos del cliente</b>
2.	2	D		0	N	CODIGO SBS DEL CLIENTE	Código del cliente, asignado por la SBS
3.		D		0	N	FECHA DEL REPORTE	Fecha del reporte de saldos ( <b>AAAAMDD</b> )
4.		D			C	TIPO DE DOCUMENTO TRIBUTARIO	2 = RUC de 8 dígitos 3 = RUC de 11 dígitos
5.		D			C	RUC	RUC
6.		D			C	TIPO DE DOCUMENTO DE IDENTIDAD	1 = Libreta Electoral / DNI 2 = Carné de Extranjería o Carta de Identidad 5 = Pasaporte
7.		D			C	DOCUMENTO DE IDENTIDAD	Código de documento de identidad
8.		D			C	TIPO DE PERSONA	1 = Persona Natural 2 = Persona Jurídica 3 = Personas Mancomunadas 4 = Patrimonios fideicometidos y vehículos de propósito especial. (De acuerdo con la definición del reglamento de titulización).
9.		D			C	TIPO DE EMPRESA	X = Empresa Extranjera E = Empresa del Estado M = Ministerios Blanco = Otro
10.		D		0	N	CANTIDAD EMPRESAS	Número de empresas en que ha sido reportado el cliente.
11.		D		2	N	DEUDA CALIFICACIÓN 0	Saldo de deuda en calificación Normal
12.		D		2	N	DEUDA CALIFICACIÓN 1	Saldo de deuda en calificación CPP
13.		D		2	N	DEUDA CALIFICACIÓN 2	Saldo de deuda en calificación Deficiente
14.		D		2	N	DEUDA CALIFICACIÓN 3	Saldo de deuda en calificación Dudoso
15.		D		2	N	DEUDA CALIFICACIÓN 4	Saldo de deuda en calificación Pérdida
16.		D			C	RAZÓN SOCIAL O APELLIDO PATERNO O NOMBRE DE MANCÓMUNO	Razón Social para personas jurídicas ó Apellido Paterno para personas naturales ó Nombre de Mancómuno para personas mancómunas.
17.		D			C	APELLIDO MATERNO	Apellido Materno de persona natural
18.		D			C	APELLIDO CASADA	Apellido de Casada de persona natural
19.		D			C	PRIMER NOMBRE	Primer Nombre de persona natural
20.		D			C	SEGUNDO NOMBRE	Segundo Nombre de persona natural
21.			2		C	DELIMITADOR	Delimitador de registro (hex <b>0D 0A</b> )
<b>LEYENDA</b>							
Tipo campo:		<b>C</b>	Carácter Alfanumérico				
		<b>N</b>	Numérico				
Posic.. Final		<b>D</b>	Delimitador =   (ASCII 124)				

## A.2. Registro de Saldos del Cliente (Según Analista del Área de Sistemas Administrativos)

**Tabla 8**

*Interpretación de la cadena que contiene los registros de los Saldos del Cliente.*

Nro.	Posición		Tamaño bytes	Número Decimales	Tipo campo	Nombre del campo	Descripción del campo
	Inicial	Final					
1.	1	1	1	0	N	TIPO DE REGISTRO	<b>2 = Saldos</b>
2.	2	11	10	0	N	CODIGO SBS DEL CLIENTE	Código del cliente, asignado por la SBS
3.	12	16	5	0	N	CÓDIGO EMPRESA	Código de la empresa del Sistema Financiero en la que registra deuda.
4.	17	18	2		C	TIPO DE CREDITO	01 = Créditos Soberanos 02 = Créditos a Entidades del sector público 03 = Créditos a Bancos multilaterales de desarrollo 04 = Créditos a Empresas del sistema financiero 05 = Créditos a Empresas de valores 06 = Créditos Corporativos 07 = Créditos a Grandes Empresas 08 = Créditos a Medianas Empresas 09 = Créditos a Pequeñas Empresas 10 = Créditos a Microempresas 11 = Créditos de Consumo revolventes 12 = Créditos de Consumo no revolventes 13 = Créditos Hipotecarios para vivienda
5.	19	32	14	0	N	CODIGO DE CUENTA	Código de la cuenta contable
6.	33	36	4	0	N	CONDICION	Condición en días
7.	37	54	18	2	N	SALDO	Saldo
8.	55	55	1		C	CLASIFICACION	Clasificación del cliente en la empresa : 0 = Normal 1 = Con problemas Potenciales (CPP) 2 = Deficiente 3 = Dudoso 4 = Pérdida
9.	56	57	2		C	DELIMITADOR	Delimitador de registro (hex <b>0D 0A</b> )
<b>LEYENDA</b>							
Tipo campo	<b>C</b>	Carácter Alfanumérico					
	<b>N</b>	Numérico					

## ANEXO B: MODELO DE BASES DE DATOS

### B.1. Script Base de datos

```

USE [BN_SBS]
GO

/***** Object: Table [dbo].[Calif Ahorros]*****/
SET ANSI_NULLS ON
GO
SET QUOTED_IDENTIFIER ON
GO
CREATE TABLE [dbo].[Calif Ahorros](
    [Doc Identidad] [char](10) NOT NULL,
    [Tip Doc Identidad] [char](1) NULL,
    [Cta Ahorros] [char](11) NULL
) ON [PRIMARY]
GO

/***** Object: Table [dbo].[Calif Ahorros Temp]*****/
SET ANSI_NULLS ON
GO
SET QUOTED_IDENTIFIER ON
GO
CREATE TABLE [dbo].[Calif Ahorros Temp](
    [Doc Identidad] [char](10) NULL,
    [Tip Doc Identidad] [char](1) NULL,
    [Cta Ahorros] [char](11) NULL
) ON [PRIMARY]
GO

/***** Object: Table [dbo].[Calif Matching]*****/
SET ANSI_NULLS ON
GO
SET QUOTED_IDENTIFIER ON
GO
CREATE TABLE [dbo].[Calif Matching](
    [Doc Identidad] [char](10) NULL,
    [Tip Doc Identidad Sbs] [char](1) NULL,
    [Nom Cliente Compl] [varchar](120) NULL,
    [Cta Ahorros] [char](11) NULL,
    [Cal Normal] [numeric](5, 2) NULL,
    [Cal Cpp] [numeric](5, 2) NULL,
    [Cal Deficiente] [numeric](5, 2) NULL,
    [Cal Dudoso] [numeric](5, 2) NULL,
    [Cal Perdido] [numeric](5, 2) NULL,
    [CalificacionFinal] [char](1) NULL,
    [Doc Identidad Sbs Correg] [char](11) NULL,
    [cod deudor] [char](10) NULL,
    [Tip Doc Identidad] [char](1) NULL,
    [Nom_Cliente_Compl_70] [char](70) NULL
) ON [PRIMARY]
GO

/***** Object: Table [dbo].[Calificacion]*****/
SET ANSI_NULLS ON
GO
SET QUOTED_IDENTIFIER ON
GO

```

```

CREATE TABLE [dbo].[Calificacion](
    [Cod_Calif] [char](1) NOT NULL,
    [Det_Calif] [char](15) NOT NULL,
    [DetLargo_Calif] [nvarchar](2000) NOT NULL,
    CONSTRAINT [PK_Calificacion_ñol] PRIMARY KEY CLUSTERED
(
    [Cod_Calif] ASC
)WITH (PAD_INDEX = OFF, STATISTICS_NORECOMPUTE = OFF, IGNORE_DUP_KEY = OFF,
ALLOW_ROW_LOCKS = ON, ALLOW_PAGE_LOCKS = ON, OPTIMIZE_FOR_SEQUENTIAL_KEY = OFF) ON
[PRIMARY]
) ON [PRIMARY]
GO

/***** Object: Table [dbo].[Cuentas Contables]*****/
SET ANSI_NULLS ON
GO
SET QUOTED_IDENTIFIER ON
GO
CREATE TABLE [dbo].[Cuentas Contables](
    [cod_cta_contable] [numeric](14, 0) NOT NULL,
    [desc_cta_contable] [varchar](255) NOT NULL,
    CONSTRAINT [PK_Cuentas Contables_ñol] PRIMARY KEY CLUSTERED
(
    [cod_cta_contable] ASC
)WITH (PAD_INDEX = OFF, STATISTICS_NORECOMPUTE = OFF, IGNORE_DUP_KEY = OFF,
ALLOW_ROW_LOCKS = ON, ALLOW_PAGE_LOCKS = ON, OPTIMIZE_FOR_SEQUENTIAL_KEY = OFF) ON
[PRIMARY]
) ON [PRIMARY]
GO

/***** Object: Table [dbo].[Datos Cliente]*****/
SET ANSI_NULLS ON
GO
SET QUOTED_IDENTIFIER ON
GO
CREATE TABLE [dbo].[Datos Cliente](
    [Cod_Deudor] [char](10) NOT NULL,
    [Fec_Reporte] [smalldatetime] NOT NULL,
    [Tip_Doc_Trib] [char](1) NULL,
    [Ruc] [char](11) NULL,
    [Tip_Doc_Identidad_sbs] [char](1) NULL,
    [Doc_Identidad_sbs] [char](12) NULL,
    [Tip_Persona] [char](1) NOT NULL,
    [Tip_Empresa] [char](1) NOT NULL,
    [Num_Empresas] [char](3) NOT NULL,
    [Cal_Normal] [numeric](5, 2) NOT NULL,
    [Cal_Cpp] [numeric](5, 2) NOT NULL,
    [Cal_Deficiente] [numeric](5, 2) NOT NULL,
    [Cal_Dudoso] [numeric](5, 2) NOT NULL,
    [Cal_Perdido] [numeric](5, 2) NOT NULL,
    [Nom_Cliente_Comp] [varchar](120) NOT NULL,
    [CalificacionFinal_DTS] [char](1) NOT NULL,
    [Tip_doc_Identidad_sbs_convert_ahorr] [char](1) NULL,
    [Doc_Identidad_sbs_Coneg] [char](10) NULL,
    CONSTRAINT [PK_Datos Cliente_ñol] PRIMARY KEY CLUSTERED
(
    [Cod_Deudor] ASC
)WITH (PAD_INDEX = OFF, STATISTICS_NORECOMPUTE = OFF, IGNORE_DUP_KEY = OFF,
ALLOW_ROW_LOCKS = ON, ALLOW_PAGE_LOCKS = ON, OPTIMIZE_FOR_SEQUENTIAL_KEY = OFF) ON
[PRIMARY]

```

```

) ON [PRIMARY]
GO

/***** Object: Table [dbo].[Datos_Cliente_Temp]*****/
SET ANSI_NULLS ON
GO
SET QUOTED_IDENTIFIER ON
GO
CREATE TABLE [dbo].[Datos_Cliente_Temp](
    [tip_registro] [char](1) NULL,
    [cod_deudor] [char](10) NULL,
    [fec_reporte] [smalldatetime] NULL,
    [tip_doc_trib] [char](1) NULL,
    [ruc] [char](11) NULL,
    [tip_doc_identidad] [char](1) NULL,
    [doc_identidad] [char](12) NULL,
    [tip_persona] [char](1) NULL,
    [tip_empresa] [char](1) NULL,
    [num_empresas] [char](3) NULL,
    [cal_normal] [numeric](5, 2) NULL,
    [cal_cpp] [numeric](5, 2) NULL,
    [cal_deficiente] [numeric](5, 2) NULL,
    [cal_dudoso] [numeric](5, 2) NULL,
    [cal_perdido] [numeric](5, 2) NULL,
    [Nom_Cliente_AnePat] [varchar](100) NULL,
    [Nom_Cliente_AneMat] [varchar](40) NULL,
    [Nom_Cliente_AneCas] [varchar](40) NULL,
    [Nom_Cliente_PriNom] [varchar](40) NULL,
    [Nom_Cliente_SegNom] [varchar](40) NULL,
    [Nom_Cliente_Comp] [varchar](220) NULL
) ON [PRIMARY]
GO

/***** Object: Table [dbo].[Instituciones]*****/
SET ANSI_NULLS ON
GO
SET QUOTED_IDENTIFIER ON
GO
CREATE TABLE [dbo].[Instituciones](
    [Cod_Institucion] [char](5) NOT NULL,
    [Tipo_Institucion] [varchar](30) NULL,
    [Estado_Institucion] [varchar](30) NULL,
    [Nom_Institucion] [varchar](150) NOT NULL,
    CONSTRAINT [PK_Instituciones] PRIMARY KEY CLUSTERED
(
    [Cod_Institucion] ASC
) WITH (PAD_INDEX = OFF, STATISTICS_NORECOMPUTE = OFF, IGNORE_DUP_KEY = OFF,
ALLOW_ROW_LOCKS = ON, ALLOW_PAGE_LOCKS = ON, OPTIMIZE_FOR_SEQUENTIAL_KEY = OFF) ON
[PRIMARY]
) ON [PRIMARY]
GO

/***** Object: Table [dbo].[Registro_Saldos]*****/
SET ANSI_NULLS ON
GO
SET QUOTED_IDENTIFIER ON
GO
CREATE TABLE [dbo].[Registro_Saldos](
    [Cod_Deudor] [char](10) NOT NULL,
    [Cod_Empresa] [char](5) NOT NULL,

```

```

[Tip_Credito] [char](2) NOT NULL,
[Cod_Cta_Contable] [numeric](14, 0) NOT NULL,
[Condicion] [char](4) NOT NULL,
[Saldo] [numeric](15, 2) NOT NULL,
[Calif_Final_xSbs] [char](1) NOT NULL,
CONSTRAINT [PK_Registro_Saldos_ño] PRIMARY KEY CLUSTERED
(
    [Cod_Deudor] ASC,
    [Cod_Empresa] ASC,
    [Tip_Credito] ASC,
    [Cod_Cta_Contable] ASC,
    [Condicion] ASC
)WITH (PAD_INDEX = OFF, STATISTICS_NORECOMPUTE = OFF, IGNORE_DUP_KEY = OFF,
ALLOW_ROW_LOCKS = ON, ALLOW_PAGE_LOCKS = ON, OPTIMIZE_FOR_SEQUENTIAL_KEY = OFF) ON
[PRIMARY]
) ON [PRIMARY]
GO

/***** Object: Table [dbo].[Registro Saldos Temp]*****/
SET ANSI_NULLS ON
GO
SET QUOTED_IDENTIFIER ON
GO
CREATE TABLE [dbo].[Registro Saldos Temp](
    [tip_registro] [char](1) NULL,
    [cod_deudor] [char](10) NULL,
    [cod_empresa] [char](5) NULL,
    [tip_credito] [char](2) NULL,
    [cod_cta_contable] [char](14) NULL,
    [condicion] [char](4) NULL,
    [saldo] [numeric](18, 2) NULL,
    [Calif_Final_xSbs] [char](1) NULL
) ON [PRIMARY]
GO

/***** Object: Table [dbo].[Tipo Credito]*****/
SET ANSI_NULLS ON
GO
SET QUOTED_IDENTIFIER ON
GO
CREATE TABLE [dbo].[Tipo Credito](
    [Tip_Credito] [char](2) NOT NULL,
    [Desc_Credito] [varchar](35) NOT NULL,
    [DescLarga_Credito] [nvarchar](2000) NULL,
    CONSTRAINT [PK_Tipo Credito_ño] PRIMARY KEY CLUSTERED
(
    [Tip_Credito] ASC
)WITH (PAD_INDEX = OFF, STATISTICS_NORECOMPUTE = OFF, IGNORE_DUP_KEY = OFF,
ALLOW_ROW_LOCKS = ON, ALLOW_PAGE_LOCKS = ON, OPTIMIZE_FOR_SEQUENTIAL_KEY = OFF) ON
[PRIMARY]
) ON [PRIMARY]
GO

/***** Object: Table [dbo].[Tipo Documento]*****/
SET ANSI_NULLS ON
GO
SET QUOTED_IDENTIFIER ON
GO
CREATE TABLE [dbo].[Tipo Documento](
    [Tip_Doc_Identidad] [char](1) NOT NULL,

```

```

[Desc Documento] [varchar](40) NOT NULL,
CONSTRAINT [PK_Tipo Documento ño] PRIMARY KEY CLUSTERED
(
    [Tip_Doc_Identidad] ASC
)WITH (PAD_INDEX = OFF, STATISTICS_NORECOMPUTE = OFF, IGNORE_DUP_KEY = OFF,
ALLOW_ROW_LOCKS = ON, ALLOW_PAGE_LOCKS = ON, OPTIMIZE_FOR_SEQUENTIAL_KEY = OFF) ON
[PRIMARY]
) ON [PRIMARY]
GO

/***** Object: Table [dbo].[Tipo Persona]*****/
SET ANSI_NULLS ON
GO
SET QUOTED_IDENTIFIER ON
GO
CREATE TABLE [dbo].[Tipo Persona](
    [Tip Persona] [char](1) NOT NULL,
    [Desc Persona] [varchar](125) NOT NULL,
    CONSTRAINT [PK_Tipo Persona ño] PRIMARY KEY CLUSTERED
(
    [Tip Persona] ASC
)WITH (PAD_INDEX = OFF, STATISTICS_NORECOMPUTE = OFF, IGNORE_DUP_KEY = OFF,
ALLOW_ROW_LOCKS = ON, ALLOW_PAGE_LOCKS = ON, OPTIMIZE_FOR_SEQUENTIAL_KEY = OFF) ON
[PRIMARY]
) ON [PRIMARY]
GO

```



## B.2. Modelo de Bases de Datos – SQL Server

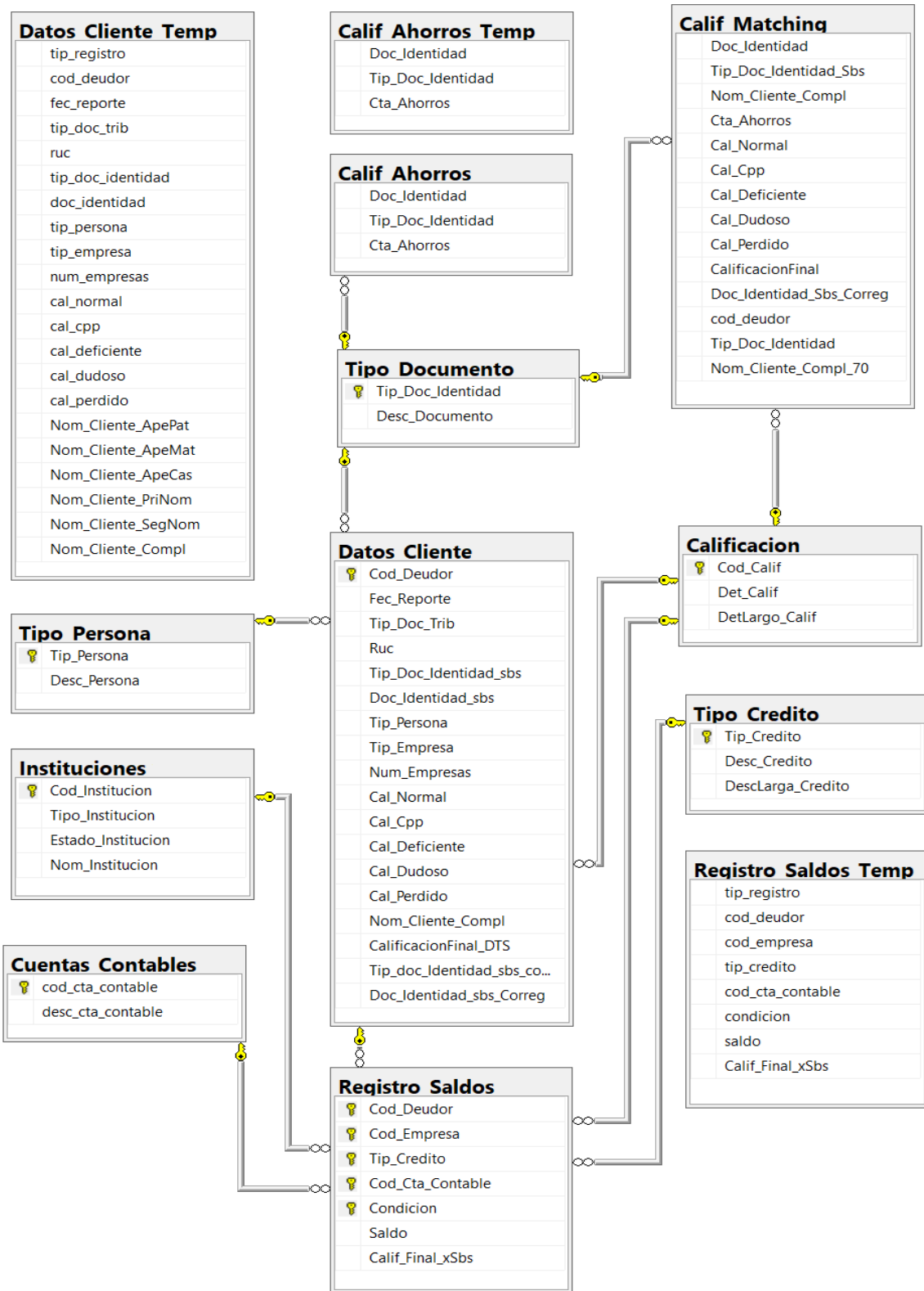
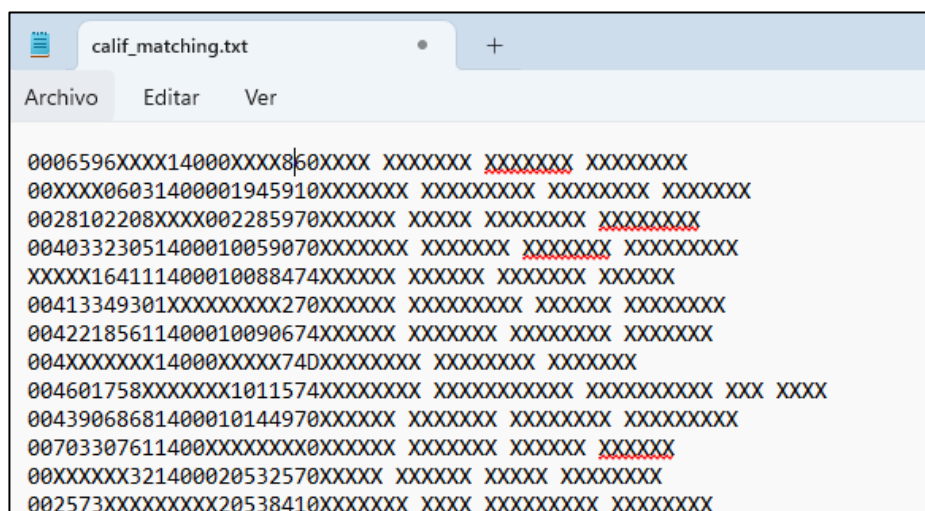


Figura 10. Modelo de Bases de datos (Elaboración propia).

## ANEXO C: DETALLE DE LOS ARCHIVOS DE SALIDA

### C.1. Archivo de Salida – Calificación Matching

A continuación, se muestran el archivo de salida de Calificación Matching, por motivos de seguridad se modificó la data sensible:



```
calif_matching.txt
Archivo  Editar  Ver

0006596XXXX14000XXXX860XXXX XXXXXXX XXXXXXX XXXXXXX
00XXXX06031400001945910XXXXXXX XXXXXXXXXX XXXXXXXXXX XXXXXXX
0028102208XXX002285970XXXXXXX XXXXX XXXXXXXXXX XXXXXXXXXX
00403323051400010059070XXXXXXX XXXXXXX XXXXXXX XXXXXXXXXX
XXXXX164111400010088474XXXXXXX XXXXXXX XXXXXXX XXXXXXX
00413349301XXXXXXXXX270XXXXXXX XXXXXXXXXX XXXXXXX XXXXXXXXXX
00422185611400010090674XXXXXXX XXXXXXX XXXXXXX XXXXXXX
004XXXXXXXX14000XXXXX74DXXXXXXXX XXXXXXX XXXXXXX
004601758XXXXXXXX1011574XXXXXXXX XXXXXXXXXXX XXXXXXXXXXX XXX XXXX
00439068681400010144970XXXXXXX XXXXXXX XXXXXXX XXXXXXXXXX
00703307611400XXXXXXXX0XXXXXXX XXXXXXX XXXXXXX XXXXXXX
00XXXXXX321400020532570XXXXXXX XXXXXXX XXXXX XXXXXXXXXX
002573XXXXXXXXXX20538410XXXXXXXX XXXX XXXXXXXXXX XXXXXXXXXX
```

*Figura 11. Calificación Matching (Data modificada).*

### C.2. Archivos de Salida – Ahorros No Validos e Instituciones Faltantes

A continuación, se muestran los archivos de salida de Ahorros no Validos e Instituciones Faltantes, por motivos de seguridad se modificó la data sensible:

Ahorros_NoValido.txt		Instituciones_Faltantes.txt	
Archivo	Editar Ver	Archivo	Editar Ver
xx01001518140272073977		00138	
31059498571xx011407407		00146	
4xx3728xxxx40158798377		00418	
4439471800140105568197		00419	
5032019101343762500707		00468	
53080xx86434749xx27097		00469	
CC531161507474xx537807		00470	
CH1GGGZM44740980408926		00471	
L. E. 140713xx7337		00472	
L. E. 00000 140002xx9607		00476	
L. E. 06xxx6140592000xx7		00477	
L. E. 072229140713388147		00478	
L. E. 09xxx5140002167xx7		00485	
L. E. 09942814000xx63787		00486	
L. E. 15xxx314331xx71167		00487	
N 1400x8012037		00488	
NN 1400x2184866		00496	
NN 1400x8715547		00497	
NN 142310164796		00498	
NN 143530036537		00499	
O070830057640194969407		00500	
RD49811188741940642717		00508	
RN24xxx702740458416567		00509	
S/D 1436100xx367		00510	
S/N 14005x000707		00511	
XX 1474x0085926		00512	

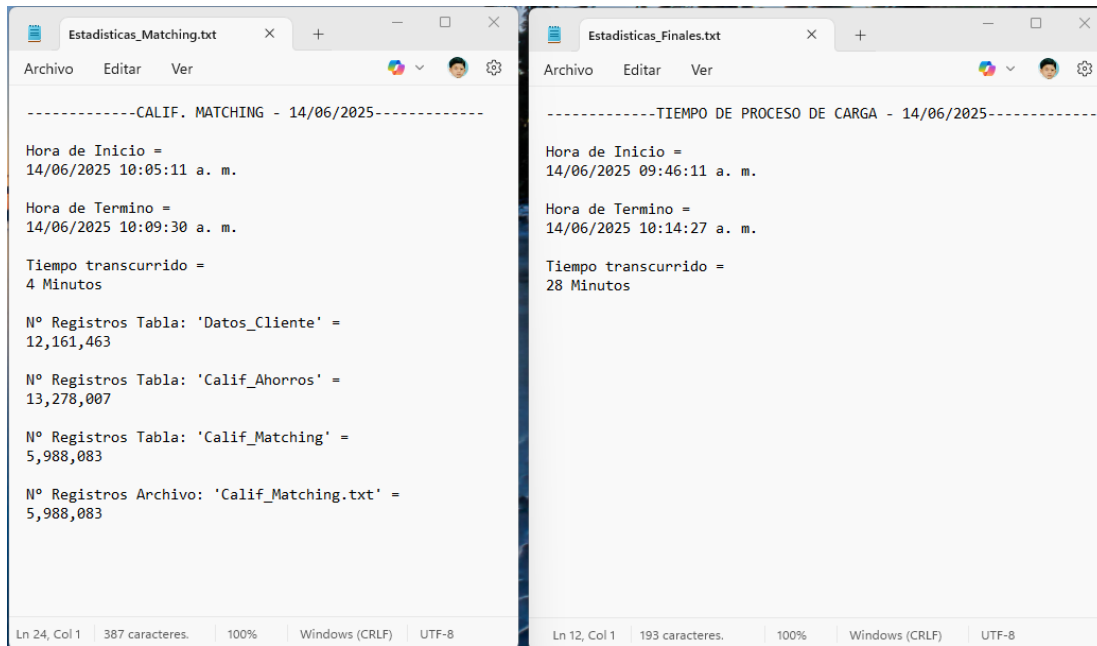
*Figura 12. Ahorros no validos e Instituciones faltantes (Data modificada).*

### C.3. Archivos de Salida – Estadísticas

A continuación, se muestran los Archivos de salida relacionados a las estadísticas:

Estadísticas_Ahorros.txt	Estadísticas_SBS.txt
-----CALIF. AHORROS - 14/06/2025-----	-----DATA SBS - 14/06/2025-----
Hora de Inicio = 14/06/2025 09:46:11 a. m.	Hora de Inicio = 14/06/2025 09:50:44 a. m.
Hora de Termino = 14/06/2025 09:50:44 a. m.	Hora de Termino = 14/06/2025 10:05:11 a. m.
Tiempo transcurrido = 4 Minutos	Tiempo transcurrido = 15 Minutos
Nº Registros Archivo: 'data_ahorros.txt' = 13,291,212	Nº Registros Archivo Input: 'data_SBS.txt' = 80,126,388
Nº Registros Tabla: 'Calif_Ahorros_Temp' = 13,291,212	Nº Registros Tabla: 'Datos_Cliente_Temp' = 12,161,463
Nº Registros Tabla: 'Calif_Ahorros' = 13,278,007	Nº Registros Tabla: 'Registro_Saldos_Temp' = 69,761,777
	Nº Registros Tabla: 'Datos_Cliente' = 12,161,463
	Nº Registros Tabla: 'Registro_Saldos' = 69,761,777
Ln 21, Col 1   340 caracteres.   100%   Windows (CRLF)   UTF-8	Ln 112, Col 1   2,226 caracteres.   100%   Windows (CRLF)   UTF-8

*Figura 13. Estadísticas Carga Ahorros y SBS (Resultado del proceso ETL).*



```
-----CALIF. MATCHING - 14/06/2025-----  
Hora de Inicio =  
14/06/2025 10:05:11 a. m.  
Hora de Termino =  
14/06/2025 10:09:30 a. m.  
Tiempo transcurrido =  
4 Minutos  
N° Registros Tabla: 'Datos_Cliente' =  
12,161,463  
N° Registros Tabla: 'Calif_Ahorros' =  
13,278,007  
N° Registros Tabla: 'Calif_Matching' =  
5,988,083  
N° Registros Archivo: 'Calif_Matching.txt' =  
5,988,083  
Ln 24, Col 1 | 387 caracteres. | 100% | Windows (CRLF) | UTF-8
```

```
-----TIEMPO DE PROCESO DE CARGA - 14/06/2025-----  
Hora de Inicio =  
14/06/2025 09:46:11 a. m.  
Hora de Termino =  
14/06/2025 10:14:27 a. m.  
Tiempo transcurrido =  
28 Minutos  
Ln 12, Col 1 | 193 caracteres. | 100% | Windows (CRLF) | UTF-8
```

*Figura 14. Estadísticas Matching y Finales (Resultado del proceso ETL).*

